



Epigenetic regulation of human papillomavirus transcription in the productive virus life cycle

Megan Burley¹ · Sally Roberts¹ · Joanna L. Parish¹

Received: 7 October 2019 / Accepted: 8 November 2019 / Published online: 9 January 2020
© The Author(s) 2020

Abstract

Human papillomaviruses (HPV) are a large family of viruses which contain a circular, double-stranded DNA genome of approximately 8000 base pairs. The viral DNA is chromatinized by the recruitment of cellular histones which are subject to host cell-mediated post-translational epigenetic modification recognized as an important mechanism of virus transcription regulation. The HPV life cycle is dependent on the terminal differentiation of the target cell within epithelia—the keratinocyte. The virus life cycle begins in the undifferentiated basal compartment of epithelia where the viral chromatin is maintained in an epigenetically repressed state, stabilized by distal chromatin interactions between the viral enhancer and early gene region. Migration of the infected keratinocyte towards the surface of the epithelium induces cellular differentiation which disrupts chromatin looping and stimulates epigenetic remodelling of the viral chromatin. These epigenetic changes result in enhanced virus transcription and activation of the virus late promoter facilitating transcription of the viral capsid proteins. In this review article, we discuss the complexity of virus- and host-cell-mediated epigenetic regulation of virus transcription with a specific focus on differentiation-dependent remodelling of viral chromatin during the HPV life cycle.

Keywords Papillomavirus · Epigenetics · Life cycle · Chromatin structure · Transcription

Introduction

At the time of writing, the *Papillomaviridae* (PV) family of viruses is composed of over 450 distinct types of human papillomavirus (HPV) and over 200 animal papillomaviruses (<http://pave.niaid.nih.gov>). Each PV type is defined by a > 2% difference in sequence from any other known type. PV types are arranged into distinct genera that share > 60% identity in the L1 open reading frame (ORF). HPV types are phylogenetically arranged in five genera; alpha, beta, gamma, mu and nu [1]. All known HPV types occupy a tightly defined niche; they exclusively replicate in keratinocytes within squamous epithelia of either the cutaneous or mucosal

surfaces of the human body. Infection with the vast majority of HPV types results in benign disease that is often sub-clinical, but can develop into the growth of papillomas or warts at the epithelial surface. A subset of HPV types (HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59 and 66) are the causative agent of cancers of the anogenital and oropharyngeal tracts and defined as group I carcinogens by the World Health Organization [2]. Due to their association with cancer development, these so-called high-risk HPV types have been most widely studied and will therefore be the focus of this review.

HPV genome structure

The genome of all HPV types has a similar arrangement characterized by an approximately 8000 base pair circular double-stranded DNA genome encased in a non-enveloped icosahedral capsid of about 55 nm in diameter [1]. The viral genome contains 7–9 open reading frames (ORF) divided into early (E1, E2, E4, E5, E6, E7 and E8, although E5 and E8 ORF are not present in the genomes of all HPV types) and late (L1 and L2) genes (Fig. 1a). The core proteins, E1 and E2, have key roles in viral DNA replication and amplification, and

This article is a contribution to the special issue on Infection-induced epigenetic changes and the pathogenesis of diseases - Guest Editor: Nicole Fischer

✉ Joanna L. Parish
j.l.parish@bham.ac.uk

¹ College of Medical and Dental Sciences, Institute of Cancer and Genomic Sciences, University of Birmingham, B152TT, Birmingham, UK

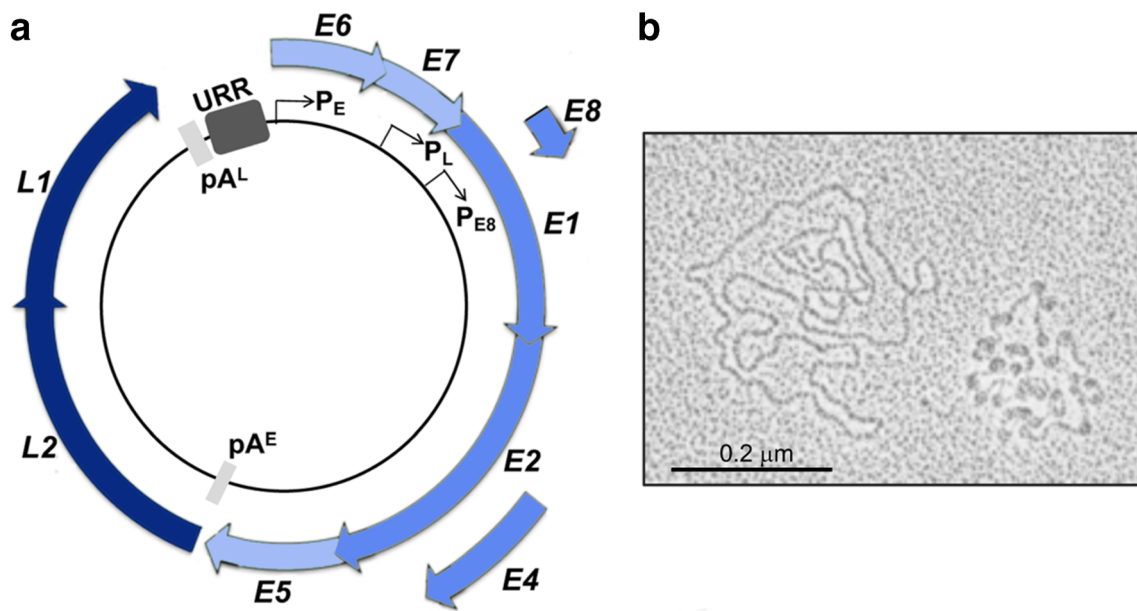


Fig. 1 HPV genome organization. **a** The circular, double-stranded HPV genome is about 8000 base pairs and divided into three regions: the early, late and non-coding upstream regulatory region (URR). The early region contains open-reading frames (ORF), some overlapping of E1, E2, E4, E5, E6, E7 and E8. The late region contains L1 and L2 ORF—the capsid proteins. Viral transcription is controlled by the HPV E2 protein and host factors binding sequences within the URR. The main promoters are the early P_E , the late P_L and the E8 promoter P_{E8} , and viral transcripts

terminate at the early pA^E or late pA^L poly-adenylation sites. The URR also contains the origin of replication. **b** In a study in the late 1970s [3], electron microscopy of metal-shadowed HPV genomes isolated from plantar warts showed naked HPV DNA molecules (left) and nucleoprotein-DNA complexes (right) revealing an intricate ‘beads on a string’ conformation of nucleosomes. Reproduced with permission from the American Society for Microbiology

regulating viral transcription, and the L1 and L2 proteins form the capsid, as well as L2 having a role in delivery of the viral genome to the nucleus upon infection and viral genome encapsidation during capsid assembly. Accessory proteins include E4, E5, E6 and E7 and these proteins facilitate the different stages of the vegetative virus life cycle primarily by forming virus-host interactions to alter the environment of the keratinocyte to support viral replication and enable evasion of host anti-viral defences. For the high-risk HPV types, the key players in oncogenesis are the oncoproteins E5, E6 and E7. A non-coding region referred to as the upstream regulatory region (URR; also known as the long control region (LCR)) is situated upstream of the early region (Fig. 1a). This region contains binding sites for a plethora of transcription and regulatory factors that either activate or repress the early (E) and late (L) promoters (P_E : P97 - HPV16, P105 - HPV18; P_L : P670 - HPV16; P811 - HPV18), the origin of replication to which the E1 protein binds, as well as multiple binding sites for the viral E2 protein. Relevant to this review was a study in the 1970s that showed that the HPV genome does not exist in a naked state in an productive lesion but as a nucleoprotein complex containing cellular histones [3](Fig. 1b).

HPV transcription

Several promoters within the HPV genome have been identified, but for the purposes of this review, we will mention here

only the early promoter (P_E) that is active in undifferentiated basal keratinocytes, the late promoter (P_L) that lies within the E7 ORF and becomes active in differentiated keratinocytes, and the P_{E8} promoter present within the E1 ORF which controls the expression of an E8^{E2} protein that regulates viral transcription and viral DNA replication. Alternative RNA splicing leads to the production of multicistronic viral messenger RNAs transcribed from one strand of the genome. Early transcripts initiate from P_E and terminate at the early polyadenylation (PolyA) site (pA^E) situated at the end of the E5 ORF (Fig. 1a). Transcripts from the differentiation-regulated P_L also utilize the early PolyA site but those that encode the structural proteins terminate at the late PolyA motif (pA^L) present in the URR. The P_{E8} promoter is active throughout the infectious cycle and unlike P_E and P_L , constitutive activation of P_{E8} is not controlled by the viral enhancer elements situated within the URR [4].

The programme of HPV transcription is intimately linked to the physiology of the keratinocyte and all stages of RNA metabolism are regulated during the virus life cycle, including promoter usage, polyadenylation, splice site usage, mRNA stability and translation (reviewed in [5]). The overall effect of this complex, differentiation-specific programme ensures that low levels of those early proteins necessary for initial amplification and establishment of the viral genome are expressed in basal cells. As differentiation occurs and the life cycle switches to the vegetative cycle, the expression levels of

these proteins rise along with E4, E5, E6 and E7 to alter the keratinocyte milieu to enable viral DNA amplification and restrict expression of the structural proteins necessary for virion assembly in the upper most differentiated cells. One key aspect of this programme is that it avoids high expression levels of the viral proteins in basal keratinocytes and thereby avoids activation of host immune pathways. Integral to this control process of HPV transcription is epigenetic modification of the viral chromatin.

HPV life cycle

HPVs infect basal keratinocytes, the proliferative compartment of squamous epithelia, through wounds and micro-abrasions in the epithelium (Fig. 2). Keratinocyte infection is a lengthy multi-step cascade of host factor binding and protease-induced capsid conformational changes initiated following virion binding to heparin sulphate proteoglycans on the basal lamina. Upon mitosis of the infected keratinocyte, the incoming viral genome in complex with the minor capsid protein L2 is bound to the condensed chromatin. Following an initial phase of viral DNA amplification, the episomal genome is established at a copy number of approximately 50 to 100 copies per cell. The early proteins E1 and E2 along with host replication factors including DNA polymerase α /primase, replication protein A and topoisomerase I facilitate viral DNA replication; E1 functioning as an ATP-dependent DNA helicase to unwind the double-stranded DNA and E2 acting

as a sequence-specific DNA binding protein to load E1 helicase onto the viral origin of replication in the URR (Fig. 3a).

The maintenance phase of HPV DNA replication occurs in synchrony with the host DNA which ensures equal partitioning of the viral episomes to daughter cells. E2 plays a central role in this process by tethering the viral genomes to host chromatin to ensure efficient inheritance of the viral DNA. The viral genomes are tethered to different regions of mitotic chromosomes and this involves interaction of E2 with different host factors, including the epigenetic reader bromodomain protein Brd4, the DNA helicase chromosome loss-related 1 (ChlR1) and topoisomerase binding protein 1 (TopBP1) (reviewed in [6]). E2-mediated tethering via Brd4 association also appears to be important for positioning viral genomes to host genomic fragile sites that contain large and actively transcribed genes in epigenetically active chromatin [7].

E2 protein function is also central to virus transcription control. There are four conserved palindromic E2 binding sites within the URR of high-risk alpha-HPVs with the consensus sequence ACCG(N)₄CGGT that each binds an E2 dimer [8] (Fig. 3a). Binding of E2 to these sites can either activate or, more commonly, repress transcription depending on the sequence context of the occupied E2 sites within the URR. The mechanism of E2-dependent transcriptional repression is both through steric hindrance of transcriptional activators such as TATA binding protein (TBP) and specificity protein 1 (SP1)

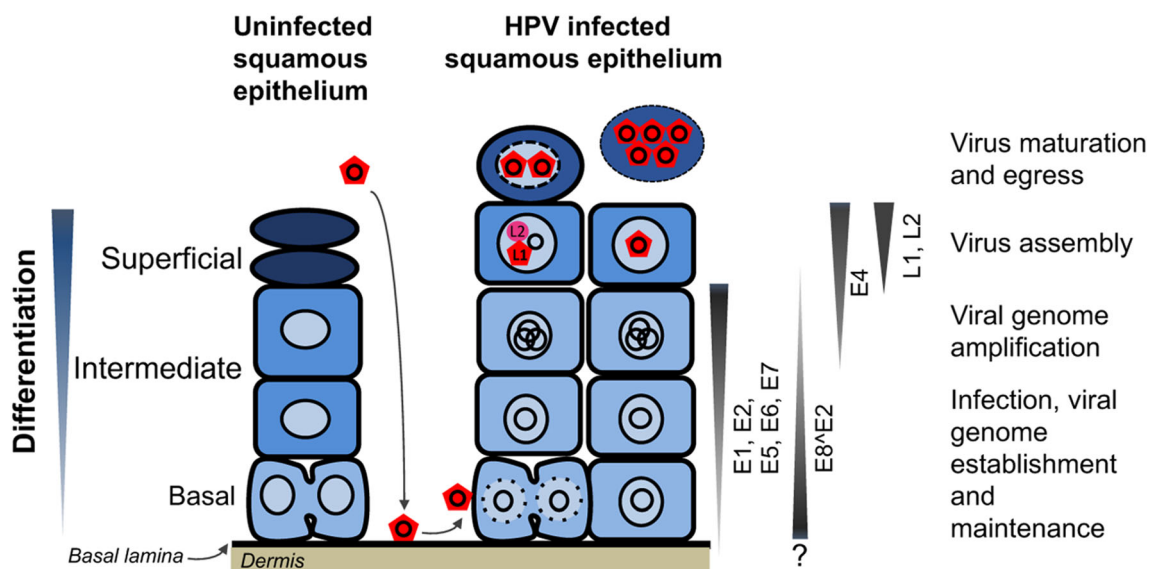


Fig. 2 HPV life cycle organization. The HPV life cycle is dependent on the terminal differentiation of the keratinocyte. The virus life cycle begins by the viral particle (red pentagon) gaining access to the basal lamina and then infecting the mitotically active basal keratinocyte. The viral genome (black circle) is established as an extra-chromosomal replicon and maintained in basal cells until the cell differentiates and HPV early protein expression increases. Differentiating cells are pushed back into

cell cycle and the viral genome amplifies to high copy number. Finally, the cell completes differentiation, expresses the viral late structural proteins L1 and L2 enabling virion assembly and egress. A viral regulator E8[^]E2 regulates viral transcription and replication and can also inhibit its own promoter P_{E8} suggesting that levels of the regulator may be finely tuned during the life cycle

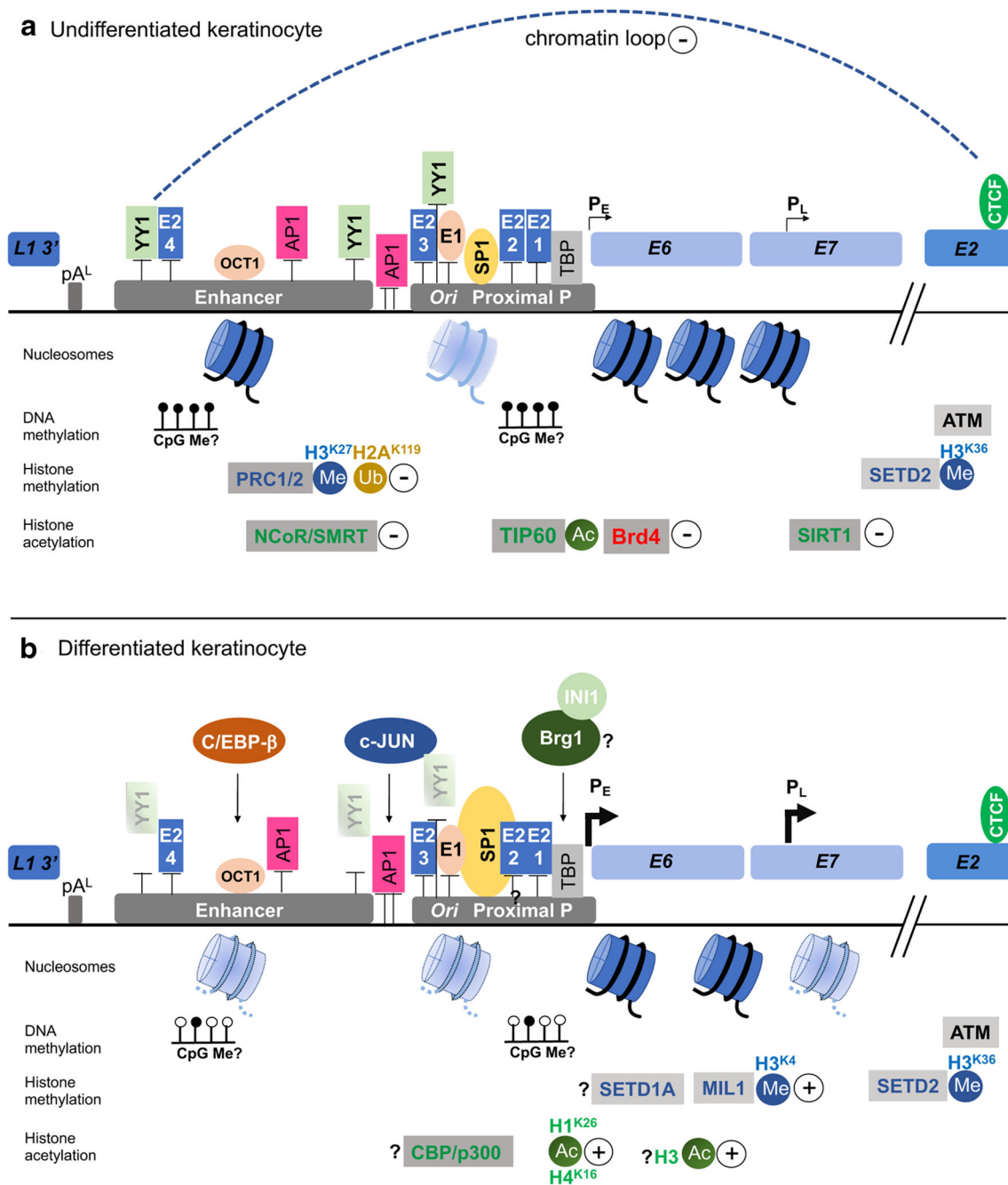


Fig. 3 Epigenetic regulation of the HPV transcription during the virus life cycle. Binding of nucleosomes to the HPV URR (contains the enhancer, origin of replication (*Ori*) and proximal promoter P_E), epigenetic modifications (DNA methylation (CpG Me), histone methylation (Me) and histone acetylation (Ac)) and some of the major host epigenetic modifiers involved are shown in undifferentiated (**a**) and differentiated keratinocytes (**b**). Details are given in the text. Strongly bound

nucleosomes are shown in dark blue, weakly bound in faint blue, and nucleosomes that are lost or repositioned upon differentiation are shown in faint blue and with dotted lines. Question marks (?) indicate uncertainty of the epigenetic modification occurring in differentiated cells. Plus and minus signs indicate stimulation or repression of viral transcription. Host factors known to bind the URR that are relevant to this review are shown only, alongside E1 and E2 binding

to proximal promoter elements [9–13] or by facilitating recruitment of cellular factors that mediate epigenetic regulation of the viral chromatin [14–16].

The E8^{E2} protein is encoded by a transcript that originates from the P_{E8} promoter situated within the E1 ORF of many but not all animal and human papillomavirus types. The

E8^{E2} product is formed from an alternative exon within the E1 ORF (E8) spliced into the 3' exon of the E2 gene such that E8^{E2} protein contains a novel N-terminal domain fused to the DNA-binding domain of E2 [17]. As such, E8^{E2} is unable to bind to the E1 helicase and many of the cellular transcriptional regulators bound by the N-terminal domain of full

length E2 but retain the ability to homo- and hetero-dimerize with full length E2 and bind to E2 DNA binding sites in the viral URR [18]. E8^ΔE2 is therefore a strong repressor of HPV replication by excluding E1 from the origin of replication, and of E2-dependent transcription activation by preventing the recruitment of E2-associated transcriptional activators to the URR. Although P_{E8} is constitutively active and independent of enhancer elements with the HPV URR, E8^ΔE2 is able to strongly inhibit its own promoter but E2 weakly activates it. Such fine tuning of E8^ΔE2 expression requires E2/E8^ΔE2 binding within the URR and is thought to represent a mechanism whereby HPV controls viral copy number [4].

Once the infected cell moves from the basal lamina, the normal entry into differentiation is delayed and early gene expression increases with stimulation of P_E and P_L activities. E5, E6 and E7 protein functions synergize to promote cell cycle re-entry and proliferation, and survival of post-mitotic keratinocytes to orchestrate viral DNA replication competent cells. They do so by targeting the components of key growth control pathways including p53 and retinoblastoma, as well as pathways that enable the virally infected cells to avoid immune detection [19]. The virus also deregulates epigenetic control of host gene expression contributing to the proliferative state and survival of the infected cell [20]. To avoid competing with host DNA replication during S phase, viral DNA amplification occurs in the G2 phase of the cell cycle, and by activating the ATM DNA damage repair pathways, HPV utilizes the repair factories that form to replicate its genome during this phase [21]. Once viral DNA amplification is completed, E2 binds to sites within the URR to repress the expression of early proteins and cell cycle activity ceases, enabling keratinocyte differentiation and the expression of the late structural proteins L1 and L2. This late productive stage encompasses viral genome encapsidation, maturation of progeny virus and the sloughing off of superficial cells packed with infectious new progeny from the epithelial surface. The vegetative phase is accompanied by expression of large quantities of E4, a viral protein of ill-defined function but one that contributes to the efficiency of this phase of the virus life cycle [22].

Epigenetic chromatin organization

Histone modification

DNA in the nucleus of eukaryotic cells is wrapped around octameric complexes of proteins called histones, composed of dimers of H2A, H2B, H3 and H4. Each histone core is wrapped by ~147 base pairs of DNA [23] to form nucleosomes that create a classical ‘beads-on-a-string’ conformation known as chromatin. Chromatin structure is central to the control of gene transcription as the arrangement of

nucleosomes on regulatory units controls accessibility to transcription factors. Histones can be covalently modified on the N-terminal tails that protrude from the core histone complexes by a series of enzymes known as writers including DNA methyltransferases (DNMTs), histone lysine methyltransferases, protein arginine methyltransferases and histone acetyltransferases (HATs). Such modifications include mono- (Me1), di- (Me2) and tri- (Me3) methylation, hydroxy-methylation and acetylation (Ac) which are added to a plethora of arginine (R) and lysine (K) residues within the protruding N-terminal tails of histones. In altering the affinity of histones for DNA, specific covalent modification of histones can differentially recruit or disrupt the binding of factors termed readers that further regulate chromatin structure and function, such as remodelling enzymes that drive repositioning of histones (for review see [24]).

Through epigenetic modification, chromatin exists in different physical states to regulate transcription. Tightly packed, closed chromatin known as heterochromatin is transcriptionally repressed, whereas open chromatin known as euchromatin is permissive for gene transcription as it contains much less densely packed and dynamically associated histones allowing transcription factors to access regulatory elements and drive transcription. These general forms of chromatin are dictated by post-translational modification (PTM) of histones or by direct, covalent methylation of DNA. Heterochromatin is generally characterized by enrichment of repressive epigenetic marks such as H2K4Me2/3, H3K9Me1, H3K27Me2/3 and H4K20Me3. Acetylation of histones decreases their affinity to DNA and as such open euchromatin is enriched in epigenetic marks including H3K4Ac and H3K27Ac [25]. The diversity of histone PTMs that regulate the epigenome creates a gradient of chromatin structure ranging from heterochromatin to repressed but permissive chromatin that can be rapidly activated to constitutively active chromatin. Reversible changes to chromatin that influence gene expression were first hypothesised in 1942 [26]. Evolution of this initial hypothesis over many decades of research has now defined epigenetics as heritable changes occurring in the genome that regulate gene expression patterns without affecting the underlying DNA sequence. Epigenetic regulation of gene expression is crucial in cellular programming during development and in the regulation of cellular processes and response to environmental stimuli without altering the underlying genetic code.

DNA methylation

DNA can be directly methylated on the 5' position of the cytosine pyrimidine ring creating 5-methylcytosine (5mC). This covalent modification most often occurs on cytosines preceding a guanine (CpG), and is catalysed by DNA methyltransferases (DNMTs) which catalyse either maintenance or de novo DNA methylation (reviewed by [27]). While CpG

methylation occurs globally across the genome, there are large clusters of these sites, known as CpG islands [28]. CpG islands are important in regulating chromatin structure and gene expression control. Up to 60% of gene promoters contain CpG islands in which methylation blocks transcription initiation. However, methylation within gene bodies can also enhance transcription and alter gene splicing [29]. DNA methylation regulates gene silencing via a number of mechanisms. It can mediate the direct inhibition of essential protein-DNA interactions and reduce chromatin accessibility [30]. CpG methylation is also known to recruit methyl-CpG binding proteins (MeCPs), resulting in further alteration of chromatin structure [31]. Cytosine methylation is mediated by three key members of the DNMT family which possess methyltransferase activity. The activity of DNMT1 is preferential for hemi-methylated DNA and is often referred to as a maintenance methyltransferase while the DNMT3 family (DNMT3A, DNMT3B) can also catalyse de novo DNA methylation.

Epigenetic regulation of HPV transcription

HPV chromatin structure

The association of histone complexes with encapsidated HPV DNA was first described by Favre and colleagues in 1977 [3]. Electrophoresis of highly purified HPV virions revealed association with proteins of similar molecular mass to the canonical histone complex, H2A, H2B, H3 and H4, and it was estimated that these histone-like proteins constituted 87% of the total DNA-associated protein. Nucleated HPV DNA was analysed by electron microscopy which revealed an intricate ‘beads on a string’ conformation with each nucleosome measuring 12 nm in diameter corresponding to canonical nucleosomal formation. Up to 32 nucleosome complexes were observed on the complete HPV genome and interestingly, the interconnecting DNA was of variable length indicating sequence- or regulatory element-dependent positioning of nucleosomes (Fig. 1b). The precise arrangement of nucleosomes on the viral enhancer and promoter elements is likely to be fundamental to virus transcription regulation. Nucleosome mapping demonstrates that at least two nucleosomes are located within the URR in HPV16 and 18, one overlapping with the viral enhancer and a second overlapping with the E1 binding site within the replication origin and SP1 binding site in the early promoter [32] (Fig. 3a). The nucleosome positioning at the early promoter functions to repress virus transcription by excluding SP1 recruitment [32]. However, the replication origin and early promoter have been shown to have weaker affinity for histones than other areas of the viral genome [33], suggesting that this nucleosome is easily displaced to activate transcription and/or replication. Increased SP1 concentration

can displace this nucleosome in vitro [32]. The E1 and E2 proteins have also been shown to induce a change in nucleosomal positioning [33, 34] suggesting that nucleosome arrangement is dictated by DNA sequence as well as the binding of host and viral factors. Further three nucleosomes are positioned at the late promoter within the E6 ORF and 5' end of the E7 ORF [32, 35] (Fig. 3a). Interestingly, significant remodeling of chromatin structure with the E7 ORF occurs upon keratinocyte differentiation to increase accessibility and activation of the late promoter [35] (Fig. 3b).

Histone acetylation

HPV has been shown to interact with several HAT and HDAC family members to regulate viral transcription. CREB-binding protein (CBP) and its paralogue p300 are transcriptional coactivators that bind DNA-bound transcriptional regulators and acetylated histones. Once bound to a promoter, CBP/p300 recruit the basal transcription machinery to activate transcription. CBP/p300 also have intrinsic HAT activity and can acetylate histones [36], thereby causing relaxation of DNA at transcriptional promoters, and basal transcription factors to further activate transcription [37]. Numerous studies have demonstrated a role for p300 in maintaining the high expression of E6/E7 in cervical cancer cells. The E2, E6 and E7 proteins from various HPV types have all been shown to bind to p300 [38–41]. HPV E2 and p300 cooperate to activate the HPV early promoter cloned into transcriptional reporter constructs [42] and the interaction between E7 and p300 may be an important feedback loop as E7 abrogates CBP/p300-mediated E2 transactivation [41]. CBP/p300 can also bind to the HPV18 URR in the absence of E2 as recruitment has been demonstrated in E2-negative cervical cancer cells [43]. CBP/p300-dependent E6/E7 transcription activation is associated with acetylation of H3 at the HPV URR providing evidence that CBP/p300 activates HPV transcription by altering the epigenetic status of the viral enhancer/promoter [44]. Increased histone acetylation by CBP/p300 results in enhanced recruitment of the SWI/SNF chromatin remodelling complex catalytic subunit, the Brahma-related gene-1 (Brg1), to the URR which is required for efficient RNA polymerase II recruitment [45]. Interestingly, chemical inhibition of p300 HAT activity inhibits E6/E7 mRNA expression and induces apoptosis cervical cancer cells [46], suggesting that CBP/p300 inhibition may be an effective anti-HPV strategy. While these studies demonstrate a role for CBP/p300 in the sustained E6/E7 expression in HPV-driven cancer, the function of HAT activity in the productive virus life cycle is not understood although increased histone acetylation has been detected at the URR and late promoter following host cell differentiation.

In the context of an HPV infection, E2 functions to repress E6/E7 transcript production. In an siRNA screen designed to

identify cellular factors that contribute to E2-mediated repression of the HPV18 URR, EP400, a component of the NuA4/TIP60 histone acetylase complex was identified [15]. Acetylation of histones in the HPV URR by TIP60 results in the recruitment of bromodomain containing protein Brd4 [47]. Brd4 is a strong corepressor of E2-dependent HPV transcription [15, 16]. Therefore, rather than functioning as a coactivator of transcription as is the canonical function of TIP60, recruitment of TIP60 to the URR results in Brd4 recruitment and strong transcriptional repression.

Sirtuins (SIRT1–SIRT7) are a protein family of class III HDACs that function in DNA damage repair and apoptosis. The stable maintenance of HPV16 and HPV31 episomes within human foreskin keratinocytes results E6/E7-dependent elevation of SIRT1 expression. This increase is maintained within differentiated keratinocytes [48]. SIRT1 promotes HPV episome replication in undifferentiated keratinocytes and genome amplification upon differentiation and is important for late transcription production in differentiated cultures [48, 49]. In undifferentiated cells, SIRT1 binds to the HPV31 URR and deacetylates histone 1 at Lys26 (H1K26Ac) and histone 4 at Lys16 (H4K16Ac), enabling repression of late gene transcription. SIRT1 also stimulates the recruitment of Werner helicase (WRN) to enhance E1–E2-dependent replication fidelity [50]. Following differentiation, SIRT1 binding to HPV episomes is significantly reduced resulting in the hyperacetylation of histone-1 (Lys26) and enhanced late gene expression [48]. Interestingly, SIRT1 knock-out results in reduced E2 protein acetylation suggesting that E2 is a direct target for SIRT1 [49]. Further epigenetic repression of the viral URR is mediated by E8^E2-mediated recruitment of the HDAC3-containing NCoR/SMRT transcriptional repressor complex [51].

Histone methylation

The viral episome in undifferentiated keratinocytes exists in a repressed chromatin state in part by the recruitment of polycomb repressor complexes 1 and 2 (PRC1/2) which catalyse deposition of repressive H3K27Me3 and H3K119Ub [52]. While this is likely to be important for the productive virus life cycle, integration of viral DNA and upregulation of viral oncogene E6/E7 expression have been shown to correlate with enrichment of open chromatin at the HPV16 LCR and early promoter, mediated by chromatin remodelling enzymes Brg1 and INI1 (hSNF5/SmadB1) [53]. This disease-associated alteration of the epigenetic status of HPV chromatin increases the accessibility of positive transcriptional regulators including c-Jun and histone lysine methyltransferases, including SETD1A and MIL1, which catalyse deposition of transcriptionally active histone marks, including H3K4me3 creating a favourable landscape for RNA polymerase II recruitment which drives HPV16 oncogene transcription from

the early promoter [53]. Whether this is important in the productive virus life cycle has yet to be determined.

The histone methyltransferase SETD2 is a writer of trimethylation of histone 3 lysine 36 (H3K36me3), a mark of active transcription. High-risk E7 mediates the post-transcriptional stabilization of SETD2 resulting in increased levels in HPV31 and HPV16 containing human foreskin keratinocytes. SETD2-dependent H3K36me3 deposition is apparent throughout the viral genome and enriched at the 3' end of the early gene region in both undifferentiated and differentiated keratinocytes and is essential for both maintenance and productive viral replication [54]. Interestingly, the DNA damage kinase enzyme ataxia-telangiectasia mutated (ATM) is required for maintenance of H3K36Me3 on viral chromatin presumably through inhibition of the demethylases KDM2A and/or KDM4A suggesting that ATM not only facilitates recruitment of DNA damage repair factors to the viral genome but also influences epigenetic status [54]. Conversely, HPV E7 has been shown to enhance cellular expression of the H3K27Me3 demethylase KDM6A, resulting in derepression of host genes [55] but the consequences of KDM6A upregulation on the epigenetic status of the viral genome have not been studied.

CpG DNA methylation

The first evidence of epigenetic modification of HPV DNA was in the form of covalent methylation of CpG dinucleotides on HPV1 DNA [56, 57]. It was initially demonstrated that CpG methylation of integrated HPV18 DNA in tumourigenic and non-tumourigenic cell lines has an inverse correlation with virus transcript levels. In addition, treatment of HeLa cells with the DNA methylation inhibitor 5-azacytidine resulted in reduced HPV mRNA expression [58]. Purified HPV18 DNA can be CpG methylated in vitro resulting in attenuation of activity of transfected HPV transcription reporters [59]. Differentially methylated CpG dinucleotides are present within consensus E2 binding sites in the URR and CpG methylation at these sites inhibits E2 binding, alleviating E2-mediated repression of E6/E7 oncogenes [60]. CpG methylation changes that are initiated by cellular differentiation may influence E2-dependent virus transcription during the HPV life cycle although this is not understood. Studies in HPV16-episome containing W12 cells derived from a naturally occurring low-grade cervical lesion [61] demonstrated that the viral LCR is enriched in methylated CpG dinucleotides in poorly differentiated cells and become hypomethylated upon cellular differentiation [62]. Although it has also been noted that episomal HPV DNA in premalignant biopsy material is unmethylated suggesting that de novo methylation may occur after integration of HPV DNA into the host to attenuate production of viral transcripts, which could result in viral latency [59]. Transcriptionally silent HPV integrants can be found in the

healthy cervixes of older women suggesting that such a mechanism of HPV latency may be at play [63].

Modulation of CpG methylation of HPV DNA is important during carcinogenesis. Several studies have demonstrated a correlation between increased CpG methylation within the late gene region of integrated viral sequences and disease progression [64–67]. High-grade cervical intraepithelial neoplasia (CIN2+) cases show significantly higher methylation compared with HPV DNA clearance controls and this was found to be largely associated with the L1 and L2 ORFs [65]. Interestingly, a correlation between increased methylation status of E2 binding sites in the URR in the presence of an intact E2 ORF and disease severity has been reported in oropharyngeal squamous cell carcinomas (OPSCC) [68]. Since the DNA binding affinity of E2 is reduced by CpG methylation, this is likely to explain why CpG methylation of the HPV URR correlates with increased E6/E7 expression compared with tumours with undetectable methylation [60, 68], and that high-grade OPSCC frequently have high E2 expression combined with high E6/E7 expression [69].

Differentiation-dependent regulation of the HPV epigenome

Stimulation of keratinocyte differentiation results in an increase in transcripts that originate from the early promoter and the appearance of transcripts that are initiated from within the E6/E7 ORFs around the late promoter [35, 70, 71]. Sequences within the URR and E6/E7 gene regions are required for late promoter activation [72] and it was proposed that differentiation-dependent regulation of HPV transcription was initiated by changes to chromatin structure in these regions enhancing accessibility to host transcriptional regulators. DNase I footprinting experiments identified a region within the E7 ORF that was depleted of histone proteins following differentiation of HPV31 episome harbouring keratinocytes [35]. Differentiation-induced enhancement of chromatin accessibility was shown to be due to alterations in epigenetic status of the viral chromatin including dramatic enhancement of H3 and H4 acetylation and H3K4Me2 at both the early and late promoters [73]. These changes in chromatin structure are co-incident with enhanced binding of cellular transcriptional activators such as C/EBP- β and c-Jun to the keratinocyte specific enhancer within the URR [73].

YY1 is a strong repressor of the HPV keratinocyte specific enhancer [74, 75] and functions as a transcriptional repressor by facilitating the recruitment of the Polycomb group (PcG) of proteins to chromatin [76–78]. PcG proteins are epigenetic writers that assemble into multimeric complexes including the Polycomb repressive complexes 1 and 2 (PRC1, PRC2). PRC1 functions as a ubiquitin ligase which specifically ubiquitinylates H2A lysine 119 (H2AK119Ub) [79]. PRC2

contains the methyltransferase Enhancer of Zeste homologue 2 (Ezh2) which catalyses transcriptionally repressive H3K27Me3 deposition [80]. PRC1 and PRC2 are enriched on HPV18 URR in undifferentiated keratinocytes coincident with the enrichment of H3K27Me3 and H2K199Ub, and repression of virus transcription [52]. Studies have shown that expression of E6 and E7 from HPV types 16, 18 and 38 induces increased Ezh2 protein levels, which was shown to be required for the proliferation of HPV-positive tumour cells although a concomitant increase in H3K27Me3 levels was not observed [81–83]. This apparent disconnect is thought to be due to enhanced phosphorylation of Ezh2 at serine 21 induced by E6/E7 expression [82], which inhibits the enzymatic activity of Ezh2 [84]. It would be interesting to determine whether this represents a positive feedback mechanism of HPV transcription activation resulting from differentiation-induced enhancement of E6/E7 expression.

Chromatin binding architectural proteins such as the zinc-finger CCCTC-binding factor CTCF are fundamental in the three-dimensional organization of chromatin. CTCF is a ubiquitously expressed DNA-binding protein that binds tens of thousands of sites in the human genome [85] and functions as an epigenetic boundary insulator, transcriptional activator and repressor [86]. By facilitating the formation of chromatin loops at sites that are also enriched in cohesin [87], CTCF is important in the maintenance of long-range chromatin interactions [88]. Interestingly, the majority of genomic chromatin loops are stabilized by CTCF bound to inverted cognate sites [89] and inversion of specific CTCF sites has profound effects on chromatin loop formation [90]. The specificity of CTCF binding site orientation has been central to the hypothesis that chromatin loops are formed by extrusion of the DNA through cohesin rings that are blocked by CTCF ‘anchors’ at specific genomic loci. Whether there is a specific motor complex that drives chromatin loop extrusion is yet to be decided, but theoretical modelling suggests that loop extrusion may occur via diffusive motion within the nucleus rather than an ATP-dependent motor protein [91].

CTCF has been demonstrated to regulate transcription of several large DNA viruses including the γ -herpesviruses Kaposi sarcoma-associated herpesvirus (KSHV) and Epstein-Barr virus (EBV) [92–96]. Similarly, multiple CTCF binding sites have been identified within the genomes of several HPV types. These include a conserved cluster within the late gene region found in over 80% of 125 types screened (high- and low-risk) and also sites within the E2 ORF, present in less than 20% of HPV types analysed and which appears to be conserved in high-risk HPV types only [97, 98]. Using HPV31 episome-containing cells as a model system, the Laimins group showed that CTCF was predominantly recruited to the late gene region and that depletion of CTCF, or mutation of the L2 binding site cluster resulted in reduced episome copy number and failure of episomal establishment

[97]. Somewhat in contrast to these findings, our laboratory has shown that HPV18 genomes have enriched CTCF binding at the high-risk HPV-specific E2 ORF with an absence of binding in the late gene region, suggesting that different high-risk HPV types have evolved different strategies of gene expression control [52, 98]. Mutation of the single E2-CTCF binding site in HPV18 had no effect on replication or maintenance of HPV18 episomes, but resulted in increased early transcript production and a concomitant increase in E6 and E7 protein expression and cellular hyperproliferation [98]. Importantly, we showed that CTCF-mediated repression of HPV early gene transcription via the stabilization of a chromatin loop formed between the E2 ORF and the URR. Rather than being formed between two convergent CTCF binding sites, as has been shown in the host genome, the CTCF-dependent chromatin loop in HPV18 episomes is formed between CTCF bound at the E2 ORF and a second transcription factor, Yin Yang 1 (YY1), bound at the viral URR [52] (Fig. 3a). CTCF and YY1 have been previously shown to directly interact and co-operate in the stabilization of chromatin loops between distant loci in the human genome, providing an alternative mechanism of chromatin organization [99, 100].

It is feasible that loop formation in the HPV genome occurs via loop extrusion between the CTCF and YY1 anchor points but whether this loop is stabilized by cohesin is currently unknown although phosphorylated SMC1 (pSMC1), a structural component of cohesin rings, binds to HPV31 episomes that appears to be important for viral genome amplification [97]. Whether SMC1 also plays a role in HPV transcription and/or whether cohesin is required for chromatin loop stabilization has yet to be resolved. While YY1 is abundantly expressed in undifferentiated keratinocytes and recruited to the HPV URR, stimulation of keratinocyte differentiation causes a dramatic reduction in YY1 protein expression and a loss of YY1 recruitment to the URR (Fig. 3b). This causes a loss of repressive chromatin loop formation in the viral genome, stimulating increased early gene transcription [52]. Whether the loss of chromatin loop formation in HPV18 genomes during keratinocyte differentiation is directly responsible for late promoter activation is presently unknown.

Interestingly, ChIP-Seq data available from the ENCODE project [101] for HeLa cells have been analysed to define histone and transcription factor occupancy on the integrated HPV18 locus. While the CTCF binding site within the E2 ORF is maintained in the integrated HPV18 DNA, CTCF protein was not detected at this site [102]. This may indicate that the repression of viral transcription via CTCF-YY1 chromatin organization is abrogated in HPV-driven cancer cells, a hypothesis we are currently testing. Since CTCF binding sites have been shown to be a major hotspot for mutation in a variety of cancers [103], the loss of CTCF binding to viral DNA in tumours may represent a similar driving event in cancer development.

Conclusions

Epigenetic regulation of HPV transcription is necessary for episome establishment, genome maintenance and completion of the productive HPV life cycle. The complex interplay of positive and negative epigenetic regulation of HPV transcription is inextricably linked to the differentiation status of the infected cell; the viral genome exists in an epigenetically repressed state in the undifferentiated basal cells with low-level gene expression such that the episome can replicate but prevent immune activation. As cells enter a programme of differentiation, epigenetic repression of the viral genome is alleviated and the viral chromatin structure is maintained in an active state, resulting in increased expression of viral replication proteins and activation of the late promoter and capsid protein production. This complex transcriptional programme, requiring a plethora of host cell epigenetic regulators, appears to be disrupted in HPV-induced carcinogenesis providing the possibility of new therapeutic strategies against HPV-induced disease.

Funding information The authors are funded by the Medical Research Council (MRC; MR/N023498/1 and MR/R022011/1) and a Cancer Research UK (CRUK) PhD studentship training award.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Van Doorslaer K, Chen Z, Bernard HU, Chan PKS, DeSalle R, Dillner J, Forslund O, Haga T, McBride AA, Villa LL, Burk RD, Ictv Report C (2018) ICTV virus taxonomy profile: Papillomaviridae. *J Gen Virol* 99(8):989–990. <https://doi.org/10.1099/jgv.0.001105>
2. IARC WGoTEoCRtH (2007) Human papillomaviruses. *IARC Monogr Eval Carcinog Risks Hum* 90:1–636
3. Favre M, Breitburd F, Croissant O, Orth G (1977) Chromatin-like structures obtained after alkaline disruption of bovine and human papillomaviruses. *J Virol* 21(3):1205–1209

4. Straub E, Fertey J, Dreer M, Iftner T, Stubenrauch F (2015) Characterization of the human papillomavirus 16 E8 promoter. *J Virol* 89(14):7304–7313. <https://doi.org/10.1128/JVI.00616-15>
5. Graham SV (2017) Keratinocyte differentiation-dependent human papillomavirus gene regulation. *Viruses* 9(9). <https://doi.org/10.3390/v9090245>
6. Feeney KM, Parish JL (2009) Targeting mitotic chromosomes: a conserved mechanism to ensure viral genome persistence. *Proc Biol Sci* 276(1662):1535–1544. <https://doi.org/10.1098/rspb.2008.1642>
7. Jang MK, Shen K, McBride AA (2014) Papillomavirus genomes associate with BRD4 to replicate at fragile sites in the host genome. *PLoS Pathog* 10(5):e1004117. <https://doi.org/10.1371/journal.ppat.1004117>
8. Androphy EJ, Lowy DR, Schiller JT (1987) Bovine papillomavirus E2 trans-activating gene product binds to specific sites in papillomavirus DNA. *Nature* 325(6099):70–73. <https://doi.org/10.1038/325070a0>
9. Bernard BA, Bailly C, Lenoir MC, Darmon M, Thierry F, Yaniv M (1989) The human papillomavirus type 18 (HPV18) E2 gene product is a repressor of the HPV18 regulatory region in human keratinocytes. *J Virol* 63(10):4317–4324
10. Dong G, Broker TR, Chow LT (1994) Human papillomavirus type 11 E2 proteins repress the homologous E6 promoter by interfering with the binding of host transcription factors to adjacent elements. *J Virol* 68(2):1115–1127
11. Hou SY, Wu SY, Zhou T, Thomas MC, Chiang CM (2000) Alleviation of human papillomavirus E2-mediated transcriptional repression via formation of a TATA binding protein (or TFIID)-TFIIB-RNA polymerase II-TFIIF preinitiation complex. *Mol Cell Biol* 20(1):113–125
12. Stubenrauch F, Leigh IM, Pfister H (1996) E2 represses the late gene promoter of human papillomavirus type 8 at high concentrations by interfering with cellular factors. *J Virol* 70(1):119–126
13. Tan SH, Leong LE, Walker PA, Bernard HU (1994) The human papillomavirus type 16 E2 transcription factor binds with low cooperativity to two flanking sites and represses the E6 promoter through displacement of Sp1 and TFIID. *J Virol* 68(10):6411–6420
14. Schweiger MR, Ottinger M, You J, Howley PM (2007) Brd4-independent transcriptional repression function of the papillomavirus e2 proteins. *J Virol* 81(18):9612–9622. <https://doi.org/10.1128/JVI.00447-07>
15. Smith JA, White EA, Sowa ME, Powell ML, Ottinger M, Harper JW, Howley PM (2010) Genome-wide siRNA screen identifies SMCX, EP400, and Brd4 as E2-dependent regulators of human papillomavirus oncogene expression. *Proc Natl Acad Sci U S A* 107(8):3752–3757. <https://doi.org/10.1073/pnas.0914818107>
16. Wu SY, Lee AY, Hou SY, Kemper JK, Erdjument-Bromage H, Tempst P, Chiang CM (2006) Brd4 links chromatin targeting to HPV transcriptional silencing. *Genes Dev* 20(17):2383–2396. <https://doi.org/10.1101/gad.1448206>
17. Wang X, Meyers C, Wang HK, Chow LT, Zheng ZM (2011) Construction of a full transcription map of human papillomavirus type 18 during productive viral infection. *J Virol* 85(16):8080–8092. <https://doi.org/10.1128/JVI.00670-11>
18. Kurg R, Uusen P, Vosa L, Ustav M (2010) Human papillomavirus E2 protein with single activation domain initiates HPV18 genome replication, but is not sufficient for long-term maintenance of virus genome. *Virology* 408(2):159–166. <https://doi.org/10.1016/j.virol.2010.09.010>
19. Moody CA, Laimins LA (2010) Human papillomavirus oncoproteins: pathways to transformation. *Nat Rev Cancer* 10(8):550–560. <https://doi.org/10.1038/nrc2886>
20. Soto D, Song C, McLaughlin-Drubin ME (2017) Epigenetic alterations in human papillomavirus-associated cancers. *Viruses* 9(9). <https://doi.org/10.3390/v9090248>
21. Moody CA, Laimins LA (2009) Human papillomaviruses activate the ATM DNA damage pathway for viral genome amplification upon differentiation. *PLoS Pathog* 5(10):e1000605. <https://doi.org/10.1371/journal.ppat.1000605>
22. Egawa N, Wang Q, Griffin HM, Murakami I, Jackson D, Mahmood R, Doorbar J (2017) HPV16 and 18 genome amplification show different E4-dependence, with 16E4 enhancing E1 nuclear accumulation and replicative efficiency via its cell cycle arrest and kinase activation functions. *PLoS Pathog* 13(3):e1006282. <https://doi.org/10.1371/journal.ppat.1006282>
23. Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389(6648):251–260. <https://doi.org/10.1038/38444>
24. Biswas S, Rao CM (2018) Epigenetic tools (the writers, the readers and the erasers) and their implications in cancer therapy. *Eur J Pharmacol* 837:8–24. <https://doi.org/10.1016/j.ejphar.2018.08.021>
25. Cavalli G, Heard E (2019) Advances in epigenetics link genetics to the environment and disease. *Nature* 571(7766):489–499. <https://doi.org/10.1038/s41586-019-1411-0>
26. Waddington CH (1942) Canalization of development and the inheritance of acquired characters. *Nature* 150:563–565. <https://doi.org/10.1038/1831654a0>
27. Moore LD, Le T, Fan G (2013) DNA methylation and its basic function. *Neuropsychopharmacology* 38(1):23–38. <https://doi.org/10.1038/npp.2012.112>
28. Bird A, Taggart M, Frommer M, Miller OJ, Macleod D (1985) A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell* 40(1):91–99. [https://doi.org/10.1016/0092-8674\(85\)90312-5](https://doi.org/10.1016/0092-8674(85)90312-5)
29. Jones PA (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* 13(7):484–492. <https://doi.org/10.1038/nrg3230>
30. Keshet I, Lieman-Hurwitz J, Cedar H (1986) DNA methylation affects the formation of active chromatin. *Cell* 44(4):535–543. [https://doi.org/10.1016/0092-8674\(86\)90263-1](https://doi.org/10.1016/0092-8674(86)90263-1)
31. Boyes J, Bird A (1991) DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. *Cell* 64(6):1123–1134. [https://doi.org/10.1016/0092-8674\(91\)90267-3](https://doi.org/10.1016/0092-8674(91)90267-3)
32. Stunkel W, Bernard HU (1999) The chromatin structure of the long control region of human papillomavirus type 16 represses viral oncoprotein expression. *J Virol* 73(3):1918–1930
33. Swindle CS, Engler JA (1998) Association of the human papillomavirus type 11 E1 protein with histone H1. *J Virol* 72(3):1994–2001
34. Lefebvre O, Steger G, Yaniv M (1997) Synergistic transcriptional activation by the papillomavirus E2 protein occurs after DNA binding and correlates with a change in chromatin structure. *J Mol Biol* 266(3):465–478. <https://doi.org/10.1006/jmbi.1996.0807>
35. del Mar Pena LM, Laimins LA (2001) Differentiation-dependent chromatin rearrangement coincides with activation of human papillomavirus type 31 late gene expression. *J Virol* 75(20):10005–10013. <https://doi.org/10.1128/JVI.75.20.10005-10013.2001>
36. Bannister AJ, Kouzarides T (1996) The CBP co-activator is a histone acetyltransferase. *Nature* 384(6610):641–643. <https://doi.org/10.1038/384641a0>
37. Imhof A, Yang XJ, Ogryzko VV, Nakatani Y, Wolffe AP, Ge H (1997) Acetylation of general transcription factors by histone acetyltransferases. *Curr Biol* 7(9):689–692. [https://doi.org/10.1016/S0960-9822\(06\)00296-x](https://doi.org/10.1016/S0960-9822(06)00296-x)

38. Lee D, Lee B, Kim J, Kim DW, Choe J (2000) cAMP response element-binding protein-binding protein binds to human papillomavirus E2 protein and activates E2-dependent transcription. *J Biol Chem* 275(10):7045–7051
39. 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(18):5061–5072. <https://doi.org/10.1093/emboj/18.18.5061>
40. 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(8):6209–6219
41. Bernat A, Avvakumov N, Mymryk JS, Banks L (2003) Interaction between the HPV E7 oncoprotein and the transcriptional coactivator p300. *Oncogene* 22(39):7871–7881. <https://doi.org/10.1038/sj.onc.1206896>
42. Kruppel U, Muller-Schiffmann A, Baldus SE, Smola-Hess S, Steger G (2008) E2 and the co-activator p300 can cooperate in activation of the human papillomavirus type 16 early promoter. *Virology* 377(1):151–159. <https://doi.org/10.1016/j.virol.2008.04.006>
43. Bouallaga I, Teissier S, Yaniv M, Thierry F (2003) HMG-I(Y) and the CBP/p300 coactivator are essential for human papillomavirus type 18 enhanceosome transcriptional activity. *Mol Cell Biol* 23(7):2329–2340. <https://doi.org/10.1128/mcb.23.7.2329-2340.2003>
44. Valencia-Hernandez A, Cuevas-Bennett C, Garrido E (2007) Transcriptional regulation of human papillomavirus type 18 P105 promoter by the co-activator CBP. *Intervirology* 50(6):418–425. <https://doi.org/10.1159/000112917>
45. He H, Luo Y (2012) Brg1 regulates the transcription of human papillomavirus type 18 E6 and E7 genes. *Cell Cycle* 11(3):617–627. <https://doi.org/10.4161/cc.11.3.19115>
46. He H, Lai Y, Hao Y, Liu Y, Zhang Z, Liu X, Guo C, Zhang M, Zhou H, Wang N, Luo XG, Huo L, Ma W, Zhang TC (2017) Selective p300 inhibitor C646 inhibited HPV E6-E7 genes, altered glucose metabolism and induced apoptosis in cervical cancer cells. *Eur J Pharmacol* 812:206–215. <https://doi.org/10.1016/j.ejphar.2017.06.005>
47. Jha S, Vande Pol S, Banerjee NS, Dutta AB, Chow LT, Dutta A (2010) Destabilization of TIP60 by human papillomavirus E6 results in attenuation of TIP60-dependent transcriptional regulation and apoptotic pathway. *Mol Cell* 38(5):700–711. <https://doi.org/10.1016/j.molcel.2010.05.020>
48. Langsfeld ES, Bodily JM, Laimins LA (2015) The deacetylase sirtuin 1 regulates human papillomavirus replication by modulating histone acetylation and recruitment of DNA damage factors NBS1 and Rad51 to viral genomes. *PLoS Pathog* 11(9):e1005181. <https://doi.org/10.1371/journal.ppat.1005181>
49. Das D, Smith N, Wang X, Morgan IM (2017) The deacetylase SIRT1 regulates the replication properties of human papillomavirus 16 E1 and E2. *J Virol* 91(10). <https://doi.org/10.1128/JVI.00102-17>
50. Das D, Bristol ML, Smith NW, James CD, Wang X, Pichierri P, Morgan IM (2019) Werner helicase control of human papillomavirus 16 E1-E2 DNA replication is regulated by SIRT1 deacetylation. *Mbio* 10(4):e00263–e00219. <https://doi.org/10.1128/mBio.01635-19>
51. Dreer M, Fertey J, van de Poel S, Straub E, Madlung J, Macek B, Iftner T, Stubenrauch F (2016) Interaction of NCOR/SMRT repressor complexes with papillomavirus E8^{E2C} proteins inhibits viral replication. *PLoS Pathog* 12(4):e1005556. <https://doi.org/10.1371/journal.ppat.1005556>
52. Pentland I, Campos-Leon K, Cotic M, Davies KJ, Wood CD, Groves IJ, Burley M, Coleman N, Stockton JD, Noyvert B, Beggs AD, West MJ, Roberts S, Parish JL (2018) Disruption of CTCF-YY1-dependent looping of the human papillomavirus genome activates differentiation-induced viral oncogene transcription. *PLoS Biol* 16(10):e2005752. <https://doi.org/10.1371/journal.pbio.2005752>
53. Groves IJ, Knight EL, Ang QY, Scarpini CG, Coleman N (2016) HPV16 oncogene expression levels during early cervical carcinogenesis are determined by the balance of epigenetic chromatin modifications at the integrated virus genome. *Oncogene* 35(36):4773–4786. <https://doi.org/10.1038/ncr.2016.8>
54. Gautam D, Johnson BA, Mac M, Moody CA (2018) SETD2-dependent H3K36me3 plays a critical role in epigenetic regulation of the HPV31 life cycle. *PLoS Pathog* 14(10):e1007367. <https://doi.org/10.1371/journal.ppat.1007367>
55. Soto DR, Barton C, Munger K, McLaughlin-Drubin ME (2017) KDM6A addiction of cervical carcinoma cell lines is triggered by E7 and mediated by p21(CIP1) suppression of replication stress. *Plos Pathogens* 13(10):e1006661. <https://doi.org/10.1371/journal.ppat.1006661>
56. Danos O, Katinka M, Yaniv M (1980) Molecular cloning, refined physical map and heterogeneity of methylation sites of papilloma virus type 1a DNA. *Eur J Biochem* 109(2):457–461. <https://doi.org/10.1111/j.1432-1033.1980.tb04815.x>
57. Burnett TS, Sleeman JP (1984) Uneven distribution of methylation sites within the human papillomavirus 1a genome: possible relevance to viral gene expression. *Nucleic Acids Res* 12(23):8847–8860. <https://doi.org/10.1093/nar/12.23.8847>
58. Rosl F, Durst M, Zur Hausen H (1988) Selective suppression of human papillomavirus transcription in non-tumorigenic cells by 5-azacytidine. *EMBO J* 7(5):1321–1328
59. Rosl F, Arab A, Klevenz B, zur Hausen H (1993) The effect of DNA methylation on gene regulation of human papillomaviruses. *J Gen Virol* 74(Pt 5):791–801. <https://doi.org/10.1099/0022-1317-74-5-791>
60. Thain A, Webster K, Emery D, Clarke AR, Gaston K (1997) DNA binding and bending by the human papillomavirus type 16 E2 protein - recognition of an extended binding site. *J Biol Chem* 272(13):8236–8242. <https://doi.org/10.1074/jbc.272.13.8236>
61. Stanley MA, Browne HM, Appleby M, Minson AC (1989) Properties of a non-tumorigenic human cervical keratinocyte cell line. *Int J Cancer* 43(4):672–676
62. 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(23):12450–12459. <https://doi.org/10.1128/jvi.77.23.12450-12459.2003>
63. Leonard SM, Pereira M, Roberts S, Cuschieri K, Nuovo G, Athavale R, Young L, Ganesan R, Woodman CB (2016) Evidence of disrupted high-risk human papillomavirus DNA in morphologically normal cervixes of older women. *Sci Rep* 6:20847. <https://doi.org/10.1038/srep20847>
64. Kalantari M, Calleja-Macias IE, Tewari D, Hagmar B, Lie K, Barrera-Saldana HA, Wiley DJ, Bernard HU (2004) Conserved methylation patterns of human papillomavirus type 16 DNA in asymptomatic infection and cervical neoplasia. *J Virol* 78(23):12762–12772. <https://doi.org/10.1128/JVI.78.23.12762-12772.2004>
65. Mirabello L, Schiffman M, Ghosh A, Rodriguez AC, Vasiljevic N, Wentzensen N, Herrero R, Hildesheim A, Wacholder S, Scibior-Bentkowska D, Burk RD, Lorincz AT (2013) Elevated methylation of HPV16 DNA is associated with the development of high grade cervical intraepithelial neoplasia. *Int J Cancer* 132(6):1412–1422. <https://doi.org/10.1002/ijc.27750>
66. Mirabello L, Sun C, Ghosh A, Rodriguez AC, Schiffman M, Wentzensen N, Hildesheim A, Herrero R, Wacholder S, Lorincz A, Burk RD (2012) Methylation of human papillomavirus type 16 genome and risk of cervical precancer in a Costa Rican

- population. *J Natl Cancer Inst* 104(7):556–565. <https://doi.org/10.1093/jnci/djs135>
67. Sun C, Reimers LL, Burk RD (2011) Methylation of HPV16 genome CpG sites is associated with cervix precancer and cancer. *Gynecol Oncol* 121(1):59–63. <https://doi.org/10.1016/j.ygyno.2011.01.013>
 68. Reuschenbach M, Huebbers CU, Prigge ES, Bermejo JL, Kalteis MS, Preuss SF, Seuthe IM, Kolligs J, Speel EJ, Olthof N, Kremer B, Wagner S, Klusmann JP, Vinokurova S, von Knebel DM (2015) Methylation status of HPV16 E2-binding sites classifies subtypes of HPV-associated oropharyngeal cancers. *Cancer* 121(12):1966–1976. <https://doi.org/10.1002/cncr.29315>
 69. Anayannis NV, Schlecht NF, Ben-Dayan M, Smith RV, Belbin TJ, Ow TJ, Blakaj DM, Burk RD, Leonard SM, Woodman CB, Parish JL, Prystowsky MB (2018) Association of an intact E2 gene with higher HPV viral load, higher viral oncogene expression, and improved clinical outcome in HPV16 positive head and neck squamous cell carcinoma. *PLoS One* 13(2):e0191581. <https://doi.org/10.1371/journal.pone.0191581>
 70. Ozbun MA, Meyers C (1997) Characterization of late gene transcripts expressed during vegetative replication of human papillomavirus type 31b. *J Virol* 71(7):5161–5172
 71. Ozbun MA, Meyers C (1998) Temporal usage of multiple promoters during the life cycle of human papillomavirus type 31b. *J Virol* 72(4):2715–2722
 72. Bodily JM, Meyers C (2005) Genetic analysis of the human papillomavirus type 31 differentiation-dependent late promoter. *J Virol* 79(6):3309–3321. <https://doi.org/10.1128/JVI.79.6.3309-3321.2005>
 73. Wooldridge TR, Laimins LA (2008) Regulation of human papillomavirus type 31 gene expression during the differentiation-dependent life cycle through histone modifications and transcription factor binding. *Virology* 374(2):371–380. <https://doi.org/10.1016/j.virol.2007.12.011>
 74. O'Connor MJ, Tan SH, Tan CH, Bernard HU (1996) YY1 represses human papillomavirus type 16 transcription by quenching AP-1 activity. *J Virol* 70(10):6529–6539
 75. Bauknecht T, Angel P, Royer HD, zur Hausen H (1992) Identification of a negative regulatory domain in the human papillomavirus type 18 promoter: interaction with the transcriptional repressor YY1. *EMBO J* 11(12):4607–4617
 76. Srinivasan L, Atchison ML (2004) YY1 DNA binding and PcG recruitment requires CtBP. *Genes Dev* 18(21):2596–2601. <https://doi.org/10.1101/gad.1228204>
 77. Wilkinson F, Pratt H, Atchison ML (2010) PcG recruitment by the YY1 REPO domain can be mediated by Yaf2. *J Cell Biochem* 109(3):478–486. <https://doi.org/10.1002/jcb.22424>
 78. Wilkinson FH, Park K, Atchison ML (2006) Polycomb recruitment to DNA in vivo by the YY1 REPO domain. *Proc Natl Acad Sci U S A* 103(51):19296–19301. <https://doi.org/10.1073/pnas.0603564103>
 79. Wang H, Wang L, Erdjument-Bromage H, Vidal M, Tempst P, Jones RS, Zhang Y (2004) Role of histone H2A ubiquitination in Polycomb silencing. *Nature* 431(7010):873–878. <https://doi.org/10.1038/nature02985>
 80. Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, Jones RS, Zhang Y (2002) Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* 298(5595):1039–1043. <https://doi.org/10.1126/science.1076997>
 81. Holland D, Hoppe-Seyler K, Schuller B, Lohrey C, Maroldt J, Durst M, Hoppe-Seyler F (2008) Activation of the enhancer of zeste homologue 2 gene by the human papillomavirus E7 oncoprotein. *Cancer Res* 68(23):9964–9972. <https://doi.org/10.1158/0008-5472.CAN-08-1134>
 82. Hyland PL, McDade SS, McCloskey R, Dickson GJ, Arthur K, McCance DJ, Patel D (2011) Evidence for alteration of EZH2, BMI1, and KDM6A and epigenetic reprogramming in human papillomavirus type 16 E6/E7-expressing keratinocytes. *J Virol* 85(21):10999–11006. <https://doi.org/10.1128/JVI.00160-11>
 83. Saidj D, Cros MP, Hernandez-Vargas H, Guarino F, Sylla BS, Tommasino M, Accardi R (2013) Oncoprotein E7 from beta human papillomavirus 38 induces formation of an inhibitory complex for a subset of p53-regulated promoters. *J Virol* 87(22):12139–12150. <https://doi.org/10.1128/JVI.01047-13>
 84. Cha TL, Zhou BHP, Xia WY, Wu YD, Yang CC, Chen CT, Ping B, Otte AP, Hung MC (2005) Akt-mediated phosphorylation of EZH2 suppresses methylation of lysine 27 in histone H3. *Science* 310(5746):306–310. <https://doi.org/10.1126/science.1118947>
 85. Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD, Zhang MQ, Lobanenko VV, Ren B (2007) Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* 128(6):1231–1245. <https://doi.org/10.1016/j.cell.2006.12.048>
 86. Braccioli L, de Wit E (2019) CTCF: a Swiss-army knife for genome organization and transcription regulation. *Essays Biochem* 63(1):157–165. <https://doi.org/10.1042/EBC20180069>
 87. Wendt KS, Yoshida K, Itoh T, Bando M, Koch B, Schirghuber E, Tsutsumi S, Nagae G, Ishihara K, Mishiro T, Yahata K, Imamoto F, Aburatani H, Nakao M, Imamoto N, Maeshima K, Shirahige K, Peters JM (2008) Cohesin mediates transcriptional insulation by CCTC-binding factor. *Nature* 451(7180):796–801. <https://doi.org/10.1038/nature06634>
 88. Splinter E, Heath H, Kooren J, Palstra RJ, Klous P, Grosveld F, Galjart N, de Laat W (2006) CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev* 20(17):2349–2354. <https://doi.org/10.1101/gad.399506>
 89. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159(7):1665–1680. <https://doi.org/10.1016/j.cell.2014.11.021>
 90. de Wit E, Vos ES, Holwerda SJ, Valdes-Quezada C, Verstegen MJ, Teunissen H, Splinter E, Wijchers PJ, Krijger PH, de Laat W (2015) CTCF binding polarity determines chromatin looping. *Mol Cell* 60(4):676–684. <https://doi.org/10.1016/j.molcel.2015.09.023>
 91. Brackley CA, Johnson J, Michieletto D, Morozov AN, Nicodemi M, Cook PR, Marenduzzo D (2018) Extrusion without a motor: a new take on the loop extrusion model of genome organization. *Nucleus* 9(1):95–103. <https://doi.org/10.1080/19491034.2017.1421825>
 92. Pentland I, Parish JL (2015) Targeting CTCF to control virus gene expression: a common theme amongst diverse DNA viruses. *Viruses* 7(7):3574–3585. <https://doi.org/10.3390/v7072791>
 93. Kang H, Wiedmer A, Yuan Y, Robertson E, Lieberman PM (2011) Coordination of KSHV latent and lytic gene control by CTCF-cohesin mediated chromosome conformation. *PLoS Pathog* 7(8):e1002140. <https://doi.org/10.1371/journal.ppat.1002140>
 94. Tempera I, Lieberman PM (2010) Chromatin organization of gammaherpesvirus latent genomes. *Biochim Biophys Acta* 1799(3–4):236–245. <https://doi.org/10.1016/j.bbagr.2009.10.004>
 95. Chau CM, Zhang XY, McMahon SB, Lieberman PM (2006) Regulation of Epstein-Barr virus latency type by the chromatin boundary factor CTCF. *J Virol* 80(12):5723–5732. <https://doi.org/10.1128/JVI.00025-06>
 96. Chen HS, Martin KA, Lu F, Lupey LN, Mueller JM, Lieberman PM, Tempera I (2014) Epigenetic deregulation of the LMP1/LMP2 locus of Epstein-Barr virus by mutation of a single

- CTCF-cohesin binding site. *J Virol* 88(3):1703–1713. <https://doi.org/10.1128/JVI.02209-13>
97. Mehta K, Gunasekharan V, Satsuka A, Laimins LA (2015) Human papillomaviruses activate and recruit SMC1 cohesin proteins for the differentiation-dependent life cycle through association with CTCF insulators. *PLoS Pathog* 11(4):e1004763. <https://doi.org/10.1371/journal.ppat.1004763>
98. Paris C, Pentland I, Groves I, Roberts DC, Powis SJ, Coleman N, Roberts S, Parish JL (2015) CCCTC-binding factor recruitment to the early region of the human papillomavirus 18 genome regulates viral oncogene expression. *J Virol* 89(9):4770–4785. <https://doi.org/10.1128/JVI.00097-15>
99. Beagan JA, Duong MT, Titus KR, Zhou L, Cao Z, Ma J, Lachanski CV, Gillis DR, Phillips-Cremins JE (2017) YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. *Genome Res*. <https://doi.org/10.1101/gr.215160.116>
100. Donohoe ME, Zhang LF, Xu N, Shi Y, Lee JT (2007) Identification of a Ctfc cofactor, Yy1, for the X chromosome binary switch. *Mol Cell* 25(1):43–56. <https://doi.org/10.1016/j.molcel.2006.11.017>
101. Rosenbloom KR, Dreszer TR, Long JC, Malladi VS, Sloan CA, Raney BJ, Cline MS, Karolchik D, Barber GP, Clawson H, Diekhans M, Fujita PA, Goldman M, Gravell RC, Harte RA, Hinrichs AS, Kirkup VM, Kuhn RM, Learned K, Maddren M, Meyer LR, Pohl A, Rhead B, Wong MC, Zweig AS, Haussler D, Kent WJ (2012) ENCODE whole-genome data in the UCSC genome browser: update 2012. *Nucleic Acids Res* 40(Database issue):D912–D917. <https://doi.org/10.1093/nar/gkr1012>
102. Johannsen E, Lambert PF (2013) Epigenetics of human papillomaviruses. *Virology* 445(1–2):205–212. <https://doi.org/10.1016/j.virol.2013.07.016>
103. Katainen R, Dave K, Pitkanen E, Palin K, Kivioja T, Valimaki N, Gylfe AE, Ristolainen H, Hanninen UA, Cajuso T, Kondelin J, Tanskanen T, Mecklin JP, Jarvinen H, Renkonen-Sinisalo L, Lepisto A, Kaasinen E, Kilpivaara O, Tuupanen S, Enge M, Taipale J, Aaltonen LA (2015) CTCF/cohesin-binding sites are frequently mutated in cancer. *Nat Genet* 47(7):818–821. <https://doi.org/10.1038/ng.3335>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.