Review

HPV and RNA Binding Proteins: What We Know and What Remains to Be Discovered

Sheila V. Graham

MRC-University of Glasgow Centre for Virus Research, School of Infection and Immunity, College of Medical Veterinary and Life Sciences, University of Glasgow, Glasgow G61 1QH, UK; sheila.graham@glasgow.ac.uk; Tel.: +44-1413306256

Abstract: Papillomavirus gene regulation is largely post-transcriptional due to overlapping open reading frames and the use of alternative polyadenylation and alternative splicing to produce the full suite of viral mRNAs. These processes are controlled by a wide range of cellular RNA binding proteins (RPBs), including constitutive splicing factors and cleavage and polyadenylation machinery, but also factors that regulate these processes, for example, SR and hnRNP proteins. Like cellular RNAs, papillomavirus RNAs have been shown to bind many such proteins. The life cycle of papillomaviruses is intimately linked to differentiation of the epithelial tissues the virus infects. For example, viral late mRNAs and proteins are expressed only in the most differentiated epithelial layers to avoid recognition by the host immune response. Papillomavirus genome replication is linked to the DNA damage response and viral chromatin conformation, processes which also link to RNA processing. Challenges with respect to elucidating how RBPs regulate the viral life cycle include consideration of the orchestrated spatial aspect of viral gene expression in an infected epithelium and the epigenetic nature of the viral episomal genome. This review discusses RBPs that control viral gene expression, and how the connectivity of various nuclear processes might contribute to viral mRNA production.

Keywords: Papillomaviruses; RNA binding proteins; RNA processing; epithelial differentiation; life cycle

1. Introduction

Papillomaviruses (PVs) infect mainly epithelial cells to cause benign disease, i.e., warts or papillomas [1]. PVs infect a wide range of animal species and can be divided into those that infect cutaneous and those that infect mucosal epithelia. Under certain conditions, PVs can cause lesions that progress to cancer [2]. This was first investigated in cotton tail rabbit papillomavirus (CRPV), now called *Sylvilagus floridanus* papillomavirus 1 (SfPV1), infections. SfPV1 benign lesions can progress to cancer in wild rabbits, but lesions more frequently become malignant in domestic rabbits. Bovine papillomaviruses (BPVs) infect fibroblasts as well as epithelial cells, causing cutaneous papillomas but also alimentary fibropapillomas and urinary bladder tumours in cattle exposed to bracken fern in their diet and to sarcoids, a non-invasive skin tumour of horses. Over 227 human papillomaviruses (HPVs) have been identified [3]. The majority are cutaneous epithelium-infective and cause common warts or verrucas, but a subset can infect mucosal epithelia in the anogenital region and oropharynx and can cause lesions that may progress to cancer [4]. Around 5% of all human cancers are caused by infection with oncogenic HPVs. Thirteen HPVs are defined as oncogenic by the International Agency for Research on Cancer (IARC). The most prevalent HPV worldwide is HPV genotype 16 (HPV16), which causes around 55% of cancer cases. Prophylactic vaccines have been developed against this and other oncogenic HPVs [5]. These vaccines are highly efficacious, but as they are prophylactic, there is an older global population, particularly in the developing world, at risk of disease.

Understanding the molecular biology of these viruses is key to developing novel antivirals and new approaches to vaccine design.

2. Papillomavirus Life Cycle

Papillomaviruses (PVs) have double-stranded DNA genomes of around 8 kb (Figure 1A). They replicate in the nucleus, where they use the cellular DNA replication machinery to synthesise progeny viral genomes. The nuclear environment also provides the RNA processing machinery and associated regulatory factors required to synthesise viral mRNAs. PV life cycles require a differentiating epithelium for replication (Figure 1B) [6,7]. Infection occurs specifically in the dividing cells of the basal layer of the epithelium. Upon cell entry, the virus travels to the nuclear periphery and transfers its circular (episomal) genome into the host cell nucleus. The viral genome is replicated and maintained in the basal cell nuclei at around 50–100 genome copies, which are tethered to host chromosomes via the PV E2 protein and cellular BRD4 [4]. Division of an infected basal cell can lead to a daughter cell, which begins the differentiation process by moving into upper epithelial layers where vegetative viral genome replication is initiated. Replicated genomes become packaged in the nucleus once the two viral capsid proteins, L1 and L2, are synthesised. Fully formed virions are released during the disintegration of terminally differentiated keratinocytes at the epithelial surface [6,8].
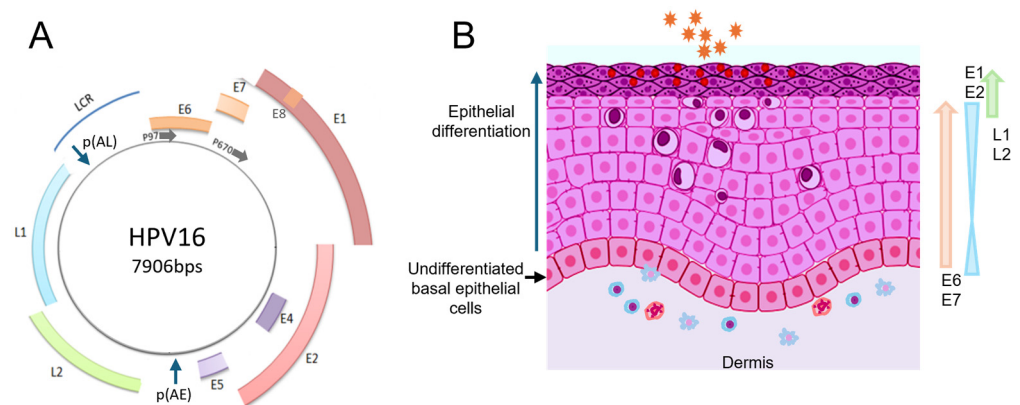

Figure 1. The HPV16 genome and viral life cycle. (**A**) Diagram of the HPV16 genome. Coloured arcs represent the viral open reading frames. LCR, long control region. Promoters are indicated with the letter P followed by a number and a horizontal arrow. p(AE), early polyadenylation site. p(AL), late polyadenylation site. (**B**) Diagram of an infected epithelium. Undifferentiated cells are coloured red/pink, differentiated cells are coloured pink/purple, and terminally differentiated cells at the top of the epithelium are coloured purple. Cells showing purple enlarged nuclei surrounded with halos represent koilocytes, a cytopathic effect of HPV infection. Virions (small orange stars) are produced in, and released from, the outermost layer of the epithelium. Arrows and triangles to the right-hand side indicate expression profiles of viral oncoproteins E6 and E7, replication factors E1 and E2, and capsid proteins L1 and L2. Cartoons of immune cells coloured red and blue are shown in the dermis. Partially created with Biorender.

During the virus life cycle, a carefully orchestrated programme of viral protein expression is tightly linked to keratinocyte differentiation (Figure 1B) [6,7]. The life cycles of the cancer-causing, or "high risk", human alpha papillomaviruses (HR-HPVs) are best understood, and the following discussion relates to the most prevalent HR-HPV, HPV16. Upon infection and insertion into the nucleus of a basal, undifferentiated epithelial cell, the viral genome expresses its replication factors E1, a DNA helicase which engages the host cell DNA replication machinery, and E2, which binds the viral origin of replication and interacts with E1 [9]. E2 also acts as the viral transcription factor. Its main role in the early stages of the virus life cycle is to repress activity of the viral early promoter to maintain low level expression of the viral E6 and E7 oncoproteins. E6 and E7 enhance cell cycle

progression [10,11] to allow HPVs to replicate in differentiated keratinocytes, which do not normally undergo cell division. Unscheduled cell division should induce apoptosis, but low levels of E6 inhibit this event by degrading the apoptosis activator, p53. These are the main early proteins of the viral life cycle. Viral proteins E4 and E5 are expressed in the early and late phases of the life cycle. Amongst other possible roles, E4 causes reassortment of the filamentous structure of differentiated keratinocytes, presumably to allow release of virus particles, while E5 regulates the host cell antiviral response [12,13]. The viral late proteins are L1, the major virus capsid protein, and L2, the minor capsid protein [14].

3. HPV and Cancer Progression

Persistent infection of basal epithelial cells, in concert with predisposing factors such as smoking, oral contraceptive use, and parity, can lead to cancer progression [2]. E6 and E7 are known as oncoproteins because high levels of expression induce tumour progression in cell culture, and animal, models, while raised E6 and E7 expression is the hallmark of HPV-induced tumours. This increase in oncoprotein expression may be due to increased transcription or post-transcriptional events. However, in human cells, integration of the viral episomal genome into the host genome often results in loss of the E2 coding region, meaning that E2 is no longer expressed to repress E6 and E7 expression. In addition, the location of insertion of the episomal genome into the host genome can lead to expression from a highly active cellular promoter or entry into a region of the genome with open chromatin conformation, leading to upregulated expression of the oncoproteins [15]. Finally, the oncogenes may end up under control of a heterologous 3′ UTR, resulting in mRNAs with increased stability [16]. Not all viral genomes are integrated during HPV-associate cancer progression; some remain episomal. How E6 and E7 expression is increased in this situation is less explored, but must involve changes in viral chromatin.

4. RNA Processing

4.1. Splicing

Upon transcription by RNA polymerase II (RNAPII), a nascent messenger RNA (mRNA) (pre-mRNA) emerges from the exit channel of the enzyme. The carboxy terminal domain (CTD) of RNAPII acts as a landing pad for factors that control the processes of capping, splicing, and polyadenylation of the pre-mRNA. Splicing is the process whereby introns in a pre-mRNA are removed, and exons are spliced together to form the mature mRNA. Splicing is catalysed by the spliceosome, a massive RNA/multi-protein complex which recognises splice sites and other signal sequences in the nascent RNA. The spliceosome comprises five small nuclear ribonucleoproteins particles or snRNPs: U1, U2, U4, U5, and U6 snRNP. These components of the spliceosome are reorganised at different stages of the splicing process, resulting in a sequential cleavage of the RNA at the 5′splice site, the formation of a lariat structure in the downstream intron, cleavage at the 3′ splice site, joining of the 5′ and 3′ exons, and degradation of the intron lariat.

The spliceosome recognises 5′ splice sites through U1 snRNP binding, while U2 snRNP is brought in and tethered at the so-called intron branch point through interaction with splicing factor U2AF. Then, snRNPs U4, U5, and U6 locate and interact with U1, resulting in the release of U1 and U4 snRNPs and allowing U6 snRNP to bind the 5′ splice site and interact with U2 downstream to form the intron lariat. U5 positions the 5′ and 3′ exons in the splicing reaction. Recognition of splice sites is controlled by positive and negative splicing factors. Exons contain cis-acting sequences, called exonic sequence enhancers (ESEs), while introns contain intronic sequence enhancers (ISEs), which influence splicing efficiency by binding splicing enhancing proteins. The best-studied activators of splicing are the serine-arginine rich (SR) proteins, which possess at least one RNA recognition motif (RRM) and a serine-arginine-rich domain (RS domain) which mediates interaction with other proteins [17–19]. Conversely, exonic sequence silencers (ESSs) or intronic sequence silencers (ISSs) bind splicing repressor proteins, the largest group of which are the hnRNP proteins. These RBPs also have RRMs and protein-binding motifs and can act in antagonism

to SR proteins [20]. Splicing enhancers bind proteins that activate the use of splice sites by attracting in the splicing machinery, by stabilising components of the spliceosome, and by preventing access of splicing silencers to silencing elements through steric hinderance. Splicing silencers act by competing with splicing enhancing factors, by blocking their access, or by forming RNA loops that block splicing of certain exons. SR proteins define exons by forming a bridge between spliceosome components at 3′ and 5′ splice sites across an exon, while hnRNP proteins block these interactions.

4.2. Alternative Splicing

Alternative splicing is the process in cells which allows expression of more than one mRNA from a single gene by mutually exclusive exon selection, or skipping or inclusion, of exons/introns in a pre-mRNA [21]. In addition, choice of alternative splice sites can alter exon size. Although consensus sequences have been determined, 5′- and 3′-splice sites are known to be degenerate, and sometimes poorly recognised by the splicing machinery. Ambiguity in splice site recognition can lead to multiple choices of splice sites within a complex pre-mRNA [22,23]. However, the presence of ESEs/ISEs and the binding of SR proteins can enhance recognition of splice sites by the splicing machinery, while ESS/ISS and hnRNP binding can repress splice site recognition leading to exon skipping. Interestingly, the "kinetic coupling of transcription and splicing" model suggests that SR/hnRNP protein recognition of cognate sequences is regulated by RNAPII transcription rate and opportunities for recruitment to the nascent RNA [24]. Alternative splicing is also controlled by the order in which exons emerge from RNA polymerase II during transcription, the pattern of RNA binding proteins (RBPs) on the pre-mRNA, and cell signalling in response to the cellular environment [25].

4.3. Polyadenylation

Cleavage and polyadenylation, or mRNA 3′ end formation, is the final event during transcription of a gene. The polyadenylation machinery consists of two complexes, CPSF (composed of Fip1, CPSF-30, CPSF-73, CPSF-100, and CPSF-160), which binds an AAUAAA signal sequence in the 3′ untranslated region (3′UTR) of the pre-mRNA around 10–30 nucleotides upstream of the eventual site of mRNA cleavage, and CstF (composed of CstF-50, CstF-64, and CstF-77), which binds a GU-rich sequence located just downstream of the cleavage site. Symplekin protein acts to bring CPSF and CstF complexes together to form an RNA loop, where cleavage will occur. Polyadenylation complexes are also associated with the CTD of RNAPII through CstF-70, and are therefore connected to upstream events such as splicing. For some genes, mRNA 3′ end formation is controlled by upstream or downstream sequence elements, their RBPs, post-translation modifications of these proteins, and the protein complex formed in and around these elements [26]. Such protein–protein–RNA interaction can also regulate mRNA stability [26].

5. Nuclear Connectivity

There is now clear evidence that many nuclear processes are highly connected (Figure 2). For example, transcription initiation and elongation by RNAPII regulates splicing and polyadenylation [24], and the DNA damage response links to transcription, alternative splicing, and polyadenylation [27–29]. These links will be explored in more detail as regards HPV gene expression later in this article. The kinetic coupling model has emerged to explain the link between transcription and alternative splicing. Generally, an elongating RNAPII will pause sequentially at splice sites along the pre-mRNA, as it is generated, to allow spliceosome assembly. This means that each intron should be spliced out in order. However, alternative exon skipping or inclusion suggests that such ordered splicing does not always occur. Instead, the rate of elongation of RNAPII determines alternative splicing outcomes: a slow progressive RNAPII can lead to inclusion of alternative exons that display weak 3′ splice sites, since splicing factors and their regulators have more time to find the splice sites and form complexes on the nascent RNA. In contrast, fast elongation

can result in exon skipping due to poor recognition of a weak splice site. Alternatively, RNA secondary structure formation over alternative exons/introns following transcription can inhibit access to splicing regulatory proteins [24]. Instead of co-transcriptional splicing, around 30% of introns are spliced out following transcription, and alternatively spliced introns may be the last to be spliced, perhaps to allow optimal recognition by RBPs [30]. There is also good evidence that changes in the recruitment of splicing factors in response to changes in the phosphorylation of the CTD of RNAPII during transcriptional elongation can regulate alternative splicing [31]. Chromatin conformation, nucleosome placement, and histone modifications can also regulate alternative splicing through changing the RNAPII elongation rate, but also by differential recruitment of splicing factor [32].
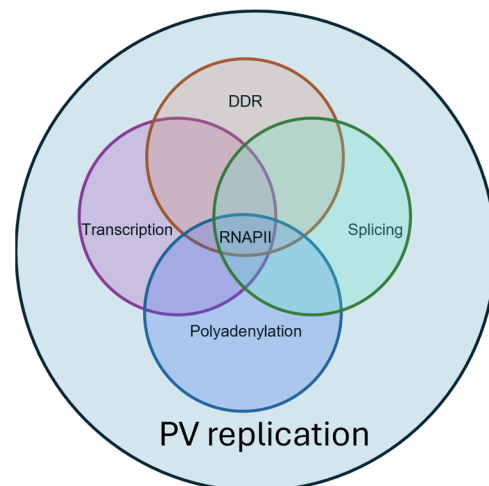

Figure 2. Diagram of the interrelated nuclear processes required for papillomavirus (PV) replication. Lilac circle: transcription. Blue circle: polyadenylation. Green circle: splicing. Orange circle: DNA damage response (DDR). RNAPII, RNA polymerase II.

Next, different nuclear RNA processing steps are tightly linked through RNAPII and co-transcriptional RNA processing and by associated RBPs, which can be shared by several different steps in mRNA biogenesis [33,34]. For example, splicing proteins can regulate nascent RNA cleavage and polyadenylation both positively and negatively, and cleavage and polyadenylation can control splicing of upstream exons [35]. SR proteins are a good example of RBPs that participate in more than one nuclear event, because they can regulate splicing, polyadenylation, nuclear export, mRNA stability and translation, as well as genome-related functions such as RNAPII elongation, chromatin remodeling, and genome stability [16,36].

The DNA damage response (DDR) can inhibit the progression of RNAPII along the genome to avoid a molecular conflict between transcription and damage repair. Once a DNA break is detected, RNA processing is altered, and DDR-related factors are recruited to the repair site [37]. DDR reprogramming of gene expression is reliant on alternative splicing controlled by splicing regulatory factors. Some of these splicing regulatory factors have been shown to have non-splicing roles in the DDR, while other studies have demonstrated that DDR-related proteins are actually splicing factors. In summary, there is a very strong connection between DDR and alternative splicing.

6. RNA Binding Proteins

RNA binding proteins (RBPs) are key regulators of gene expression at every level in the RNA biogenesis and protein translation pathways. Over 1500 RBPs exist. They regulate transcription elongation, splicing, polyadenylation, nuclear export, RNA localisation, and RNA stability and translation, as well as regulating microRNA processing. In virus-infected host cells, cellular RBPs bind viral mRNAs to regulate viral gene expression. Thus, RBPs are essential components in virus replication and key contributors to viral pathogenesis.

RBPs are important players in cellular homeostasis. Therefore, it is unsurprising that many are known to be dysregulated during cancer progression, either by altered expression levels or changed subcellular location. These changes can impact the gene expression of factors involved in key cellular processes involved in tumourigenesis, such as apoptosis, angiogenesis, invasion, and metastasis.

6.1. RBPs and Papillomavirus Life Cycles

All papillomaviruses studied to date express multiple viral mRNAs generated by alternative splicing and differential polyadenylation. Analysis is complicated by the presence of overlapping open reading frames (ORFs), meaning that some ORFs contain introns, and the fact that most PV splice signals are suboptimal [38]. Viral RNA processing is regulated according to whether viral gene expression is in the early or late phase, which is controlled by the differentiation stage of the infected epithelium (Figure 1B). Therefore, the cellular environment and differentiation status are key regulators of the viral life cycle. There is evidence that expression of RBPs changes during epithelial differentiation and during papillomavirus-associated cancer progression [39–41].

6.2. RBPs and Splicing Regulation

The involvement of RBPs in PV life cycles and tumourigenesis was first studied for BPV RNA splicing. Early studies using in vitro splicing assays, UV crosslinking, and RNA pull-down experiments showed that two ESEs upstream of the early 3′UTR bound SRSFs 1, 2, 4, and 6, and an ESS bound PTB, U2AF, and unidentified SR proteins [42,43]. Another element bound SRSF3 in a keratinocyte differentiation-dependent manner [44]. Binding of these proteins controlled the switch between viral early and late gene expression, because the elements they bound regulated inclusion or exclusion of the viral region containing the early polyadenylation site. Enhancement of a proximal splice donor site gave rise to late mRNAs that read through the early polyadenylation site to express the L2 minor capsid protein, while enhanced selection of the distal splice donor site produced mRNAs which skipped the early polyadenylation region and the L2 ORF, leading to mRNAs encoding the major capsid protein, L1.

Several approaches have been taken to identify proteins involved in regulating HPV splicing/alternative splicing. Most studies have used transfection of plasmids expressing parts of wild-type or mutated HPV genomes in HeLa cells as a model for cervical cancer or basal epithelial, cells. RNA pull-down, UV crosslinking, and immunoprecipitation studies have been used to demonstrate RBP interactions using overexpression or siRNA depletion of specific RNA binding proteins. In recent years, a Cre-Lox system to generate recombinant episomal HPV genomes in transfected cervical cell lines has also been used.

Most studies have focused on HPV16, since it is the most prevalent HPV. The HPV16 E6E7 gene region gives rise to at least four alternatively spliced mRNAs (Figure 3). The most abundant splice isoform is E6*I, where an intron from nts 226 to 409 is spliced out. This mRNA is thought to encode E7 [45]. Exon inclusion leads to a readthrough mRNA that terminates at the early polyadenylation site from which E6 is expressed. In experiments using keratinocytes transfected with a bicistronic E6E7-expressing plasmid in conjunction with siRNA depletion, splicing regulators Brm and Sam 68 were found to promote inclusion of exon 226–409 in an EGF-dependent manner, i.e., leading to expression of E6, while hnRNP A1 and A2 promoted exon exclusion which would lead to expression of E7 (Figure 4A) [46]. However, these factors were not formally proven to bind elements in the HPV16 mRNAs. Other studies on HPV16 have shown that hnRNP G and hnRNP D bind a sequence in the E7 ORF to repress production of the spliced E6*I mRNA, resulting in production of E6E7 readthrough mRNAs [47,48]. On the other hand, TRAP150/THRAP3 binds an enhancer in the E6 ORF, leading to production of the E6*I spliced RNAs from HPV16, 18, and 31 [49]. An exonic sequence silencer has been defined via mutational analysis in the HPV18 E7 ORF that inhibits the splicing event that leads to expression of E6*I by binding hnRNP A1 [50] (Figure 3). Given that relative levels of the E6- and E7-encoding mRNAs change

during cancer progression [51], it will be important to understand the relative levels and contributions of these proteins to expression of the viral oncoproteins in normal versus cancer cells.
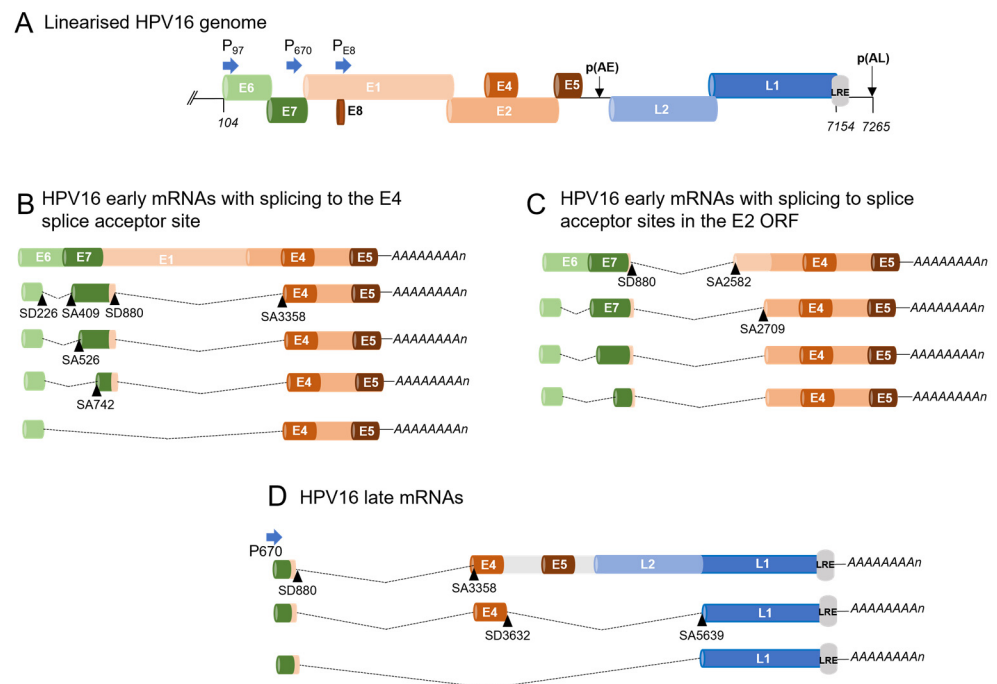

Figure 3. Diagram showing the splice sites used by selected HPV16 mRNAs. (**A**) Diagram of a linearised HPV16 genome showing coloured open reading frames. Promoters are indicated with blue horizontal arrows and designated "P" followed by a number. LRE, late regulatory element. p(AE), early polyadenylation site. p(AL), late polyadenylation site. (**B**) Diagrams of selected early mRNAs showing the readthrough mRNA (top diagram) and four mRNAs utilising the splice acceptor at the 5′ end of the E4 open reading frame. (**C**) Diagrams of early mRNAs utilising one of two splice acceptors located in the E2 open reading frame. Although only shown for splice acceptor 2709, it is assumed that mRNAs using each E2 splice acceptor can include any of the E6E7 region splice isoforms. (**D**) Diagrams of the main HPV16 late mRNAs. SD, splice donor site. SA, splice acceptor site. A(n), poly(A) tail. Exons are indicated with horizontal cylinders. Introns are indicated with dotted black lines. Splice donor and acceptor sites are indicated with black triangles.

HPV E1 and E2-encoding mRNAs are expressed at low levels. Therefore, analysis of RBPs that regulate their expression during a normal infection is problematic [52]. In most viral mRNAs, most of the HPV E1 ORF is spliced out as an intron between the 5′ splice site at the beginning of the E1 ORF and the 3′ splice site at the 5′ end of the E4 ORF (Figure 3). A recent study demonstrated that hnRNP D regulates E1 inclusion as an intron in viral early mRNA; however, a binding site for this protein was not identified in the E1 region [48]. Instead, as discussed above, hnRNP D binds in the E7 ORF and could exert regulatory pressure to ignore the splice donor site at the start of the E1 gene region (Figure 4A). E2-encoding mRNAs utilise the same 5′ splice site, with splicing to one of two 3′ splice sites in the middle of the E1 gene region (Figure 4B). A recent study has revealed a splicing enhancer located in the E2 ORF, which activates splicing to the splice acceptor site at its 5′ end to allow expression of E2. hnRNP A1, hnRNP A2, hnRNP G, hnRNP M, RBM14, and RBM15 were found to bind this element. Of these, hnRNP G had an enhancing effect on splicing, while hnRNP A1 and A2 inhibited splicing [47]. Elucidating the relative contributions of these factors, and any additional RBP interactions in these ORFs, is crucial to understanding the viral life cycle since the mRNAs produced encode E1 and E2 proteins, which regulate viral transcription and replication [4].

Splicing to the 3′ splice site at the start of the E4 ORF is a very common event in both early and late viral mRNA production (Figure 3) [38]. However, like other HPV splice sites, it is of weak consensus [53]. Some seminal studies showed that SRSFs1, 3, and 9 bind to a strong ESE within the HPV16 E4 ORF to enhance recognition of the 3′ splice site located at its 5′ end (Figure 4C) [54–56]. The E4 ORF does not contain an ATG start codon. Its start codon is acquired from the very start of the E1 ORF through splicing from the splice donor near the beginning of the E1 gene region to the E4 splice acceptor site. E4 is the most highly expressed viral protein, suggesting that splicing to produce E4 mRNAs is tightly regulated as it is essential for the virus life cycle.

PV capsid protein expression is restricted to differentiated keratinocytes. The major capsid protein, L1, is thought to be expressed mainly from mRNAs spliced from the 5′ splice site at the 3′ end of the E4 coding region to a splice acceptor at the 5′ end of the L1 ORF. This splicing event removes, as a large intron, the early 3′ UTR and polyadenylation site, as well as the ORF encoding the L2 minor capsid protein. If this splice event does not take place, and there is readthrough of the early polyadenylation site (see below), mRNAs expressing the minor capsid protein L2 are expressed (Figure 3D). Splicing repressors hnRNP L, hnRNP A1, hnRNP A2/B1, and SAM68 have been shown to bind to elements in the L1 coding region to inhibit capsid expression, possibly by inhibiting binding of U2AF and formation of an early splicing complex at the 3′ splice site at the 5′ end of the L1 coding region (Figure 4C) [57–59]. PTB (hnRNP I) also controls capsid mRNA expression by negatively regulating splice inhibitory signals around the splice donor site at the end of the E4 ORF. This regulates production of the viral capsid mRNAs together with splicing silencers in the L1 ORF (Figure 4C) [60]. When cellular levels are high, PTB may alleviate repression (by other RPBs, see below) of the splice donor site at the end of the E4 ORF, leading to splicing to the splice acceptor at the 5′ end of the L1 ORF and synthesis of L1-encoding mRNAs. Alternatively, if splicing silencers are present in the L1 ORF (Figure 4C), readthrough mRNAs expressing L2 would be synthesised.
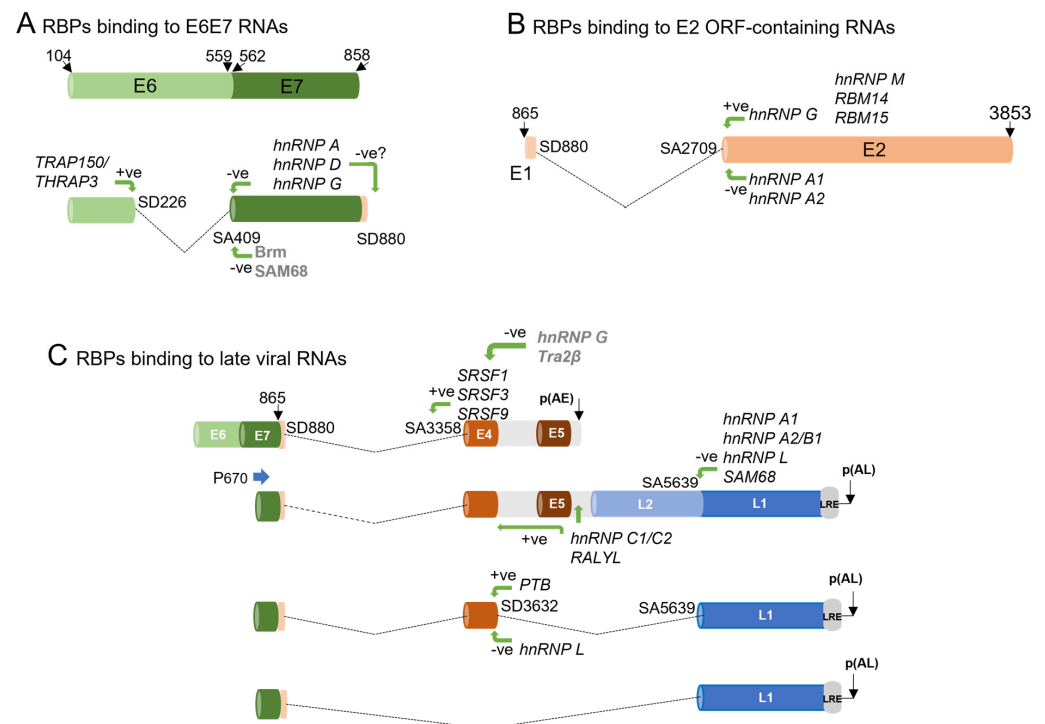

Figure 4. Diagrams of regions of the HPV16 RNAs shown to interact with RBPs. (**A**) RBPs interacting with exonic sequence motifs in the E6 and E7 open reading frames. (**B**) RBPs interacting with exonic

sequence motifs in the E2 open reading frame. (**C**) RBPs interacting with exonic sequence motifs and the early 3′ UTR in viral late RNAs. RBPs not formally shown to bind the RNAs are shown in grey type. The location of the RBP names indicates their approximate binding position on the RNAs. Promoters are indicated with blue horizontal arrows and designated "P" followed by a number. LRE, late regulatory element. p(AE), early polyadenylation site. p(AL), late polyadenylation site. Exons are indicated with horizontal cylinders. Introns are indicated with dotted black lines. Green arrows indicate the positive (+ve) or negative (−ve) effect of various RNA binding proteins on splicing. The approximate binding positions of each of these proteins can be obtained in the review by Kajitani and Schwartz [61].
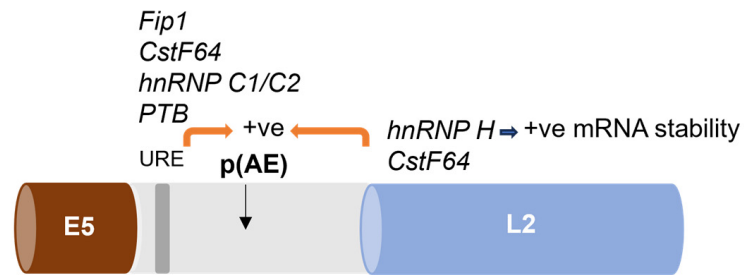
Splicing regulators RALYL and hnRNP C1/C2 bind mRNAs downstream of the HPV16 early polyadenylation site in the early 3′ UTR (Figure 4C) [62]. Overexpression of both proteins results in the production of mRNAs expressing the L1 capsid proteins. hnRNP C1 has been shown to enhance recognition of the 5′ splice site at the end of the E4 ORF, resulting in splicing out of the early 3′UTR and polyadenylation site. hnRNP L can bind to sequences around the splice donor site at the 3′ end of the E4 ORF within the early polyadenylation region, but also near the splice acceptor site at the 5′ end of the L1 ORF (Figure 4C) [63]. hnRNP L may antagonise hnRNP C1 activation of the E4 splice donor site while repressing the L1 splice acceptor site, resulting in readthrough mRNAs which do not splice out the E5 gene region and the early region 3′UTR [63]. In addition, hnRNP G and Tra2β can antagonise SR protein binding to the E4 ESE, leading to repression of expression of E4-encoding mRNAs [64].

6.3. RBPs Controlling HPV Early Polyadenylation

HPV early gene expression terminates at the early polyadenylation site (Figure 3). HPVs possess weak consensus early pAs p(AE), requiring enhancement of recognition by the cleavage and polyadenylation machinery. The HPV16 p(AE) is the best studied to date. It has an upstream regulatory element which binds polyadenylation-enhancing RBPs such as human Fip1, cleavage stimulation factor 64 kDa subunit (CstF-64), heterogenous nuclear ribonucleoprotein (hnRNP) C1/C2, and polypyrimidine tract binding protein (PTB) Figure 5A) [65]. In addition, sequences downstream of the p(AE) in the HPV16 and HPV31 L2 ORFs (ORFs) bind splicing regulatory factor hnRNP H and CstF64 to enhance cleavage and polyadenylation (Figure 5A) [66–69].

The HPV E2 protein has three domains, a DNA binding and a transactivation domain, typical of a transcription factor, but also, similar to SR proteins, a serine-arginine-rich central hinge region [70]. E2 has been shown to have RNA binding properties [71]. It can be found in nuclear speckles, the home of SR proteins in the nucleus, and can be phosphorylated, suggesting that it may have SR protein-like properties in gene regulation [72–74]. E2 can bind SR proteins in HPV-infected cells, and has been shown to be part of an RNA processing interactome [75,76]. No conclusive evidence exists to show that E2 can regulate splicing, but through its transactivation domain and hinge region, E2 can control repression of the HPV early polyadenylation site by interacting with the CPSF 30 kDa subunit [77]. It remains to be seen if E2 can bind directly to viral RNAs.
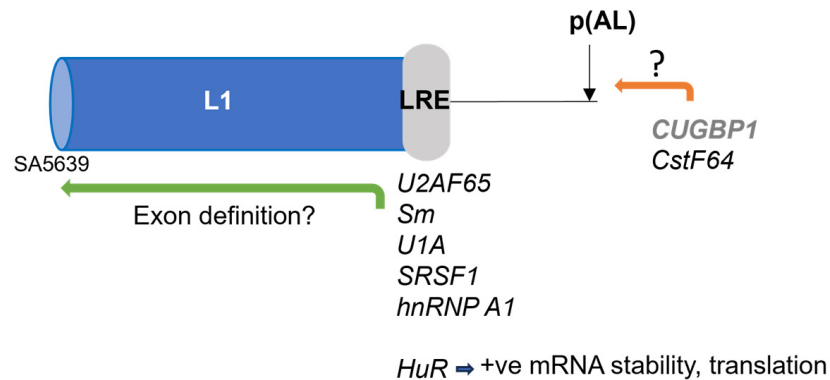
Figure 5. RBPs interacting with HPV16 3′ UTRs and polyadenylation sites. (**A**) RBPs that bind the HPV16 genome around the early polyadenylation site. Only the region of viral RNAs between the start of the E5 ORF and the end of the L2 ORF are shown in the diagram, which is not to scale. Open reading frames are shown by horizontal cylinders. Light grey region: early 3′UTR and polyadenylation site. Grey bar: upstream regulatory element in the early 3′ UTR. p(AE), early polyadenylation site. (**B**) RBPs that bind the late regulatory element (LRE: grey lozenge) in the HPV16 late 3′UTR. Only the L1 ORF is shown (not to scale). Green arrows, splicing regulation. Orange arrows, regulation of polyadenylation. RBPs not formally shown to bind the RNAs are shown in grey type.

6.4. RBPs Controlling Late Polyadenylation

BPV1—Late Regulatory Element (LRE)

Early experiments revealed a repressive cis-acting RNA element in the BPV1 late 3′UTR which inhibited expression of the mRNAs encoding the viral capsid proteins [78]. Having ruled out inhibition of transcription, polyadenylation, and mRNA decay to explain this phenomenon, it was suggested that the repressive element may mimic a 5′ splice site [79] through binding of a U1 snRNP complex [80]. However, the U1 70 kDa component of U1 snRNP was found to bind to poly (A) polymerase to inhibit polyadenylation [80]. It is possible that the element may have a dual function in BPV1 capsid RNA production; regulation of splicing of the L1 terminal exon, and inhibition of mRNA poly(A) addition in cells non-permissive for capsid protein expression, for example, undifferentiated epithelial cells.

HPV1—Late Regulatory Element

HPV1 also contains a cis-acting 57 nt element in its late 3′UTR that represses gene expression in HeLa cells [81]. The HPV1 element binds cellular proteins to repress late gene expression, potentially by restricting export to the cytoplasm [81], inducing mRNA

decay [82], and inhibiting translation [83]. RBPs that interact with the element include HuR [81], a key mRNA stability regulator also known to regulate translation initiation [84], hnRNP C1/C2 [85], and poly(A) binding protein (PABP) [83], which controls polyadenylation in the nucleus and translation in the cytoplasm [86]. Experimental evidence has suggested that HuR binding to the HPV1 element was able to control translation from a CAT reporter construct because mutation of the HuR-interacting sequences reversed inhibition of translation.

HPV16—Late Regulatory Element

The HPV16 79 nt element overlaps the 3′ end of the L1 ORF and extends into the late 3′UTR (Figure 5B) [87–89]. It contains four weak 5′ splice site sequences in its 5′ portion of 47 nts and a 32 nt G/U-rich sequence in its 3′ portion. Both portions are required for repression of gene expression in HeLa cells [90]. Using mutation analysis, the weak 5′ splice sites in the 5′ portion were shown to bind Sm proteins, U1A, and U1 snRNA, but, unlike the BPV1 3′UTR element, U1 70 K binding was not detected. Importantly, these in vitro RBP binding experiments were carried out using cell extracts from undifferentiated W12 cells, which are cervical epithelial cells naturally infected with HPV16, thus taking into account the natural cellular environment of early HPV16 infection. The first RBP shown to interact with the element was U2AF 65 kDa [89,91,92], which bound the G/U-rich 3′ portion of the repressive element [89]. Binding of a U1 snRNP-like complex to the 5′ portion and U2AF to the 3′ portion was suggestive of a cross-intron splicing complex. Further work revealed the presence of SRSF1 bound to U2AF in the complex [92].

During splicing, exon–intron junction complexes at 5′ and 3′ splice sites define the exons for RNA processing [93]. Papillomavirus terminal exons are extremely long; for example, the HPV16 L1 ORF is 1.5 kb in length. The splicing complex at the 3′ splice site located at the 5′ end of the L1 ORF should interact with the cleavage and polyadenylation complex downstream of the ORF to define the exon, and the late regulatory elementwith its associated RBP complex could mimic an early splicing complex to aid cleavage and polyadenylation and enhance L1 exon definition.

Other proteins shown to bind the element in its GU-rich 3′ portion include CstF-64, although binding is relatively weak [90,91], the splicing regulator hnRNP A1 [41], and the mRNA stability regulator HuR [91,94]. The 3′ part of the element has homology to a class III HuR binding site. HuR was first shown to bind the element using HeLa cell extracts, but was later shown to bind more strongly in extracts from differentiated W12 cells compared to undifferentiated W12 cell extracts. Working on the hypothesis that HuR could stabilise late gene transcripts, the depletion of HuR in differentiated W12 cells resulted in reduced capsid protein expression, while overexpression in undifferentiated cells induced capsid protein expression [94]. These results show that HuR is a key differentiation stage-related player in HPV late gene expression by upregulating viral late RNA stability. Finally, CUG binding protein 1 (CUGBP1) was shown to repress translation of mRNAs expressed from a Renillla reporter gene containing the GU-rich 3′ portion of the HPV16 LRE in HeLa cells [95]. However, given that HPV late mRNAs have not been shown to be present in the cytoplasm of undifferentiated epithelial cells, it is not clear how this mechanism would relate to the viral life cycle.

6.5. RBPs Controlling HPV mRNA Stability and Translation

Capsid protein expression is tightly restricted to differentiated keratinocytes, which occupy an immune privileged site at the surface of the epithelium. They are not normally expressed in undifferentiated epithelial cells, because to do so would activate the adaptive immune response to clear the infection. The mechanism of repression of HPV16 capsid protein expression in undifferentiated epithelial cells seems to be multifactorial. As well as repression via polyadenylation and splicing, an RNA regulatory element, located at the 5′ end of the HPV16 L2 ORF, can inhibit cytoplasmic late mRNA stability [96]. This region of the L2 ORF binds hnRNP H to regulate early polyadenylation, but it is possible that hnRNP H may also control stability of the L2 readthrough RNA [67,69,97]. A second

RNA regulatory element at the 3′ end of the L2 ORF has been shown to bind hnRNP K and poly(rC) binding proteins 1 and 2 to inhibit translation of L2 mRNA in vitro [98].

7. Epithelial Differentiation and RBPs

A key aspect of the PV life cycle is the close linkage between viral gene regulation events and epithelial differentiation. Of particular importance is the restriction of viral replication and virion assembly to the upper epithelial layers, where there is minimal immune surveillance. If the highly orchestrated patterns of viral gene expression were disrupted, it could lead to unscheduled expression of the immunogenic capsid proteins, and this would result in immune activation and clearance of the infection. Cellular differentiation-specific proteins must play an essential role in the viral life cycle by restricting capsid protein expression to the most differentiated epithelial layers. It is highly likely that cellular RBPs are expressed in a differentiation stage-specific manner, since it is known that keratinocyte differentiation changes functional groups of proteins involved in gene expression [99], and RBP levels are altered by B- and T-cell differentiation [100]. Importantly, there is evidence that HPV infection can alter RBP levels in a differentiation-specific manner. For example, SRSF1, 2, and 3 are transcriptionally upregulated by the E2 transcription factor of HPV16 and HPV31 in differentiated cervical keratinocytes [101–103]. This is also true of the kinase SRPK1, which controls SR protein activity by phosphorylation [104]. Therefore, it is important to establish how the levels of various RBPs known to bind and regulate PV RNA expression change during differentiation. Most studies examining RBPs interacting with or regulating HPV mRNAs have used transfections into, or cell extracts from, cervical cancer cells such as HeLa or C33a cells, or HEK293 cells, an embryonic kidney cell line, all of which have lost the capacity to differentiate. Experiments in these cells are useful as they may replicate the environment of undifferentiated epithelial cells, and can allow investigation of how RBPs regulate viral early mRNA production or repress late mRNA expression. However, since it is well recognised that many RBPs are upregulated in cancer cell lines because they control mis-regulation of cellular homeostasis [105], these cell lines probably do not express RBP levels representative of normal epithelial cells. On the other hand, results from such studies are very pertinent to HPV gene expression during tumour progression. This is particularly relevant to expression of the viral E6 and E7 oncoproteins, since their expression is directly regulated by alternative splicing, and increased levels of the oncoproteins is the direct cause of HPV-associated cancer.

8. Nuclear Connectivity, Papillomavirus mRNA Production and RBPs

Papillomavirus genomes are episomal, and the episomes are tightly packaged into chromatin [106]. As mentioned above, chromatin epigenetic modification and conformation and RNAPII processivity can regulate splicing and recruitment and access of RBPs to the pre-mRNA within the RNAPII "transcription factory". The rate of RNAPII progression around this small chromatinised episomal genome is unknown. Determining RNAPII occupancy along the viral genome at early and late stages of the viral life cycle would be helpful to identify any pauses potentially linked to the various splicing events and choice of polyadenylation sites. It would also be interesting to determine if viral RNA splicing is co- or post-transcriptional, since several viral mRNAs are products of readthrough of poor consensus splice junctions, while most result from alternative splicing events. Linked to this, evidence shows that CCCTC-binding factor (CTCF)-mediated chromatin looping of HPV episomes can regulate viral alternative splicing [52]. CTCF loops can create a roadblock to RNAPII to enhance recognition of suboptimal splice sites, leading to inclusion of poorly defined exons [107]. CTCF chromatin looping changes between undifferentiated and differentiated HPV-infected epithelial cells to allow epigenetic remodelling of viral episomes and activation of viral late gene expression [108]. Linking these changes in chromatin conformation to changes in RBP interaction with HPV RNAs may reveal new insights into how viral capsid encoding RNAs are expressed.

In the past, many of the model systems used to understand HPV mRNA production and the involvement of RBPs have used plasmids expressing PV subgenomic regions under control of a heterologous promoter, often a very active promoter such as the CMV promoter. This type of approach cannot account for the natural low activity of the viral promoters, or the chromatin conformation of the viral genome. Therefore, although these experiments have revealed many RBPs that interact with HPV RNAs, how the data relate to viral mRNA production from PV episomes requires further investigation.

In infected differentiated epithelial cells, PV replication mimics DNA damage. DNA damage repair (DDR) pathways are activated, allowing recombination-mediated replication of viral genomes [109]. Therefore, DDR is key to late events in the viral life cycle. There are clear connections between DDR and alternative splicing [37]. DDR results in cellular reprogramming mediated by changes in alternative splicing. On the other hand, DDR-activated proteins are known to be splicing regulators [24,27]. In agreement with a role of DDR in late events in the PV life cycle, one study revealed that DDR inhibited use of the HPV16 early polyadenylation signal, leading to production of the late HPV16 L1 and L2 proteins [110]. Activation of DDR was found to alter HPV16 late mRNA processing from episomal genomes and induce binding of DDR factors BCLAF1 and BRCA1 and splicing factors U2AF65 and hnRNP C to viral DNA, and U2AF65 and hnRNP C to viral RNAs. Once again, the involvement of DDR in PV replication in differentiated epithelial cells underscores the need to define RBP interactions with PV RNAs during an active infection in the correct cellular environment.

9. Conclusions

Ultimately, a full picture of RBPs and their involvement in PV life cycles will only be obtained once CLIP-Seq data for each of the viral RNAs expressed from episomal genomes is obtained. Taking viral genome replication and the differentiation-dependent nature of the life cycle into account will be key in figuring out the various regulatory pathways to full expression of the viral transcriptome. Studies such as these could be pivotal in designing novel therapies against PV infections. For example, understanding the involvement of RBPs in restricting viral capsid protein expression to differentiated epithelial cells could lead to therapies to stimulate expression of these immunogenic proteins in the lowest epithelial layers, where they would activate patrolling immune cells (Figure 1B) to clear the infection. Alternatively, understanding RBPs essential for HPV oncogene mRNA expression in cancer cells could lead to development of a therapeutic to inhibit expression and halt the reliance of HPV-associated tumour progression on E6 and E7 expression.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Rector, A.; Van Ranst, M. Animal papillomaviruses. *Virology* **2013**, *445*, 213–223. [CrossRef]
2. zur Hausen, H. Papillomaviruses in the causation of human cancers—A brief historical account. *Virology* **2009**, *384*, 260–265. [CrossRef]
3. Van Doorslaer, K.; Li, Z.; Xirasagar, S.; Maes, P.; Kaminsky, D.; Liou, D.; Sun, Q.; Kaur, R.; Huyen, Y.; McBride, A.A. The Papillomavirus Episteme: A major update to the papillomavirus sequence database. *Nucleic Acids Res.* **2017**, *45*, D499–D506. [CrossRef]
4. McBride, A.A. Human papillomaviruses: Diversity, infection and host interactions. *Nat. Rev. Microbiol.* **2022**, *20*, 95–108. [CrossRef]
5. Harper, D.M.; DeMars, L.R. HPV vaccines—A review of the first decade. *Gynecol. Oncol.* **2017**, *146*, 196–204. [CrossRef]

6. Doorbar, J.; Egawa, N.; Griffin, H.; Kranjec, C.; Murakami, I. Human papillomavirus molecular biology and disease association. *Rev. Med. Virol.* **2015**, *25*, 2–23. [CrossRef]
7. Graham, S.V. The human papillomavirus replication cycle, and its links to cancer progression: A comprehensive review. *Clin. Sci.* **2017**, *131*, 2201–2221. [CrossRef]
8. Kirk, A.; Graham, S.V. The human papillomavirus late life cycle and links to keratinocyte differentiation. *J. Med. Virol.* **2024**, *96*, e29461. [CrossRef]
9. McBride, A.A. The papillomavirus E2 proteins. *Virology* **2013**, *445*, 57–79. [CrossRef]
10. Vande Pol, S.B.; Klingelhutz, A.J. Papillomavirus E6 oncoproteins. *Virology* **2013**, *445*, 115–137. [CrossRef]
11. Roman, A.; Munger, K. The papillomavirus E7 proteins. *Virology* **2013**, *445*, 138–168. [CrossRef]
12. Wang, Q.; Griffin, H.M.; Southern, S.; Jackson, D.; Martin, A.; McIntosh, P.; Davy, C.; Masterson, P.J.; Walker, P.A.; Laskey, P.; et al. Functional analysis of the human papillomavirus type 16E1ˆE4 protein provides a mechanism for in vivo and in vitro keratin filament reorganization. *J. Virol.* **2004**, *78*, 821–833. [CrossRef]
13. DiMaio, D.; Petti, L.M. The E5 proteins. *Virology* **2013**, *445*, 99–114. [CrossRef]
14. Buck, C.B.; Cheng, N.; Thompson, C.D.; Lowy, D.R.; Steven, A.C.; Schiller, J.T.; Trus, B.L. Arrangement of L2 within the papillomavirus capsid. *J. Virol.* **2008**, *82*, 5190–5197. [CrossRef]
15. Groves, I.J.; Coleman, N. Human papillomavirus genome integration in squamous carcinogenesis: What have next-generation sequencing studies taught us? *J. Pathol.* **2018**, *245*, 9–18. [CrossRef]
16. Jeon, S.; Lambert, P.F. Integration of human papillomavirus type 16 DNA into the human genome leads to increased stability of E6 and E7 mRNAs: Implications for cervical carcinogenesis. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 1654–1658. [CrossRef]
17. Howard, J.M.; Sanford, J.R. The RNAissance family: SR proteins as multifaceted regulators of gene expression. *Wiley Interdiscip. Rev. RNA* **2015**, *6*, 93–110. [CrossRef]
18. Busch, A.; Hertel, K.J. Evolution of SR protein and hnRNP splicing regulatory factors. *Wiley Interdiscip. Rev. RNA* **2012**, *3*, 1–12. [CrossRef]
19. Martinez-Contreras, R.; Cloutier, P.; Shkreta, L.; Fisette, J.F.; Revil, T.; Chabot, B. hnRNP proteins and splicing control. *Adv. Exp. Med. Biol.* **2007**, *623*, 123–147.
20. Eperon, I.C.; Makarova, O.V.; Mayeda, A.; Munroe, S.H.; Caceres, J.F.; Hayward, D.G.; Krainer, A.R. Selection of alternative 5′ splice sites: Role of U1 snRNP and models for the antagonistic effects of SF2/ASF and hnRNP A1. *Mol. Cell Biol.* **2000**, *20*, 8303–8318. [CrossRef]
21. Graham, S.V.; Faizo, A.A.A. Control of human papillomavirus gene expression by alternative splicing. *Virus Res.* **2017**, *231*, 83–95. [CrossRef]
22. Barash, Y.; Calarco, J.A.; Gao, W.; Pan, Q.; Wang, X.; Shai, O.; Blencowe, B.J.; Frey, B.J. Deciphering the splicing code. *Nature* **2010**, *465*, 53–59. [CrossRef]
23. Roca, X.; Krainer, A.R.; Eperon, I.C. Pick one, but be quick: 5′ splice sites and the problems of too many choices. *Genes Dev.* **2013**, *27*, 129–144. [CrossRef]
24. Giono, L.E.; Kornblihtt, A.R. Linking transcription, RNA polymerase II elongation and alternative splicing. *Biochem. J.* **2020**, *477*, 3091–3104. [CrossRef]
25. Dvinge, H. Regulation of alternative mRNA splicing: Old players and new perspectives. *FEBS Lett.* **2018**, *592*, 2987–3006. [CrossRef]
26. Mayr, C. Regulation by 3′-Untranslated Regions. *Annu. Rev. Genet.* **2017**, *51*, 171–194. [CrossRef]
27. Shkreta, L.; Chabot, B. The RNA splicing response to DNA damage. *Biomolecules* **2015**, *5*, 2935–2977. [CrossRef]
28. Murphy, M.R.; Kleiman, F.E. Connections between 3′ end processing and DNA damage response: Ten years later. *Wiley Interdiscip. Rev. RNA* **2020**, *11*, e1571. [CrossRef]
29. Silva, E.; Ideker, T. Transcriptional responses to DNA damage. *DNA Repair* **2019**, *79*, 40–49. [CrossRef]
30. Choquet, K.; Baxter-Koenigs, A.R.; Dülk, S.L.; Smalec, B.M.; Rouskin, S.; Churchman, L.S. Pre-mRNA splicing order is pre-determined and maintains splicing fidelity across multi-intronic transcripts. *Nat. Struct. Mol. Biol.* **2023**, *30*, 1064–1076. [CrossRef]
31. Kornblihtt, A.R.; Schor, I.E.; Alló, M.; Dujardin, G.; Petrillo, E.; Muñoz, M.J. Alternative splicing: A pivotal step between eukaryotic transcription and translation. *Nat. Rev. Mol. Cell Biol.* **2013**, *14*, 153–165. [CrossRef]
32. Muniz, L.; Nicolas, E.; Trouche, D. RNA polymerase II speed: A key player in controlling and adapting transcriptome composition. *EMBO J.* **2021**, *40*, e105740. [CrossRef]
33. Bentley, D.L. Coupling mRNA processing with transcription in time and space. *Nat. Rev. Genet.* **2014**, *15*, 163–175. [CrossRef]
34. Maniatis, T.; Reed, R. An extensive network of coupling among gene expression machines. *Nature* **2002**, *416*, 499–506. [CrossRef]
35. Martinson, H.G. An active role for splicing in 3′-end formation. *WIREs RNA* **2011**, *2*, 459–470. [CrossRef]
36. Das, S.; Krainer, A.R. Emerging functions of SRSF1, splicing factor and oncoprotein, in RNA metabolism and cancer. *Mol. Cancer Res.* **2014**, *12*, 1195–1204. [CrossRef]
37. Mikolaskova, B.; Jurcik, M.; Cipakova, I.; Kretova, M.; Chovanec, M.; Cipak, L. Maintenance of genome stability: The unifying role of interconnections between the DNA damage response and RNA-processing pathways. *Curr. Genet.* **2018**, *64*, 971–983. [CrossRef]

38. Yu, L.; Majerciak, V.; Zheng, Z.M. HPV16 and HPV18 genome structure, expression, and post-transcriptional regulation. *Int. J. Mol. Sci.* **2022**, *23*, 4943. [CrossRef]

39. Mole, S.; McFarlane, M.; Chuen-Im, T.; Milligan, S.G.; Millan, D.; Graham, S.V. RNA splicing factors regulated by HPV16 during cervical tumour progression. *J. Pathol.* **2009**, *219*, 383–391. [CrossRef]

40. Fay, J.; Kelehan, P.; Lambkin, H.; Schwartz, S. Increased expression of cellular RNA-binding proteins in HPV-induced neoplasia and cervical cancer. *J. Med. Virol.* **2009**, *81*, 897–907. [CrossRef]

41. Chuen-Im, T.; Zhang, J.; Milligan, S.G.; McPhillips, M.G.; Graham, S.V. The alternative splicing factor hnRNP A1 is up-regulated during virus-infected epithelial cell differentiation and binds the human papillomavirus type 16 late regulatory element. *Virus Res.* **2008**, *131*, 189–198.

42. Zheng, Z.M.; He, P.J.; Baker, C.C. Structural, functional, and protein binding analyses of bovine papillomavirus type 1 exonic splicing enhancers. *J. Virol.* **1997**, *71*, 9096–9107. [CrossRef]

43. Zheng, Z.-M.; Huynen, M.; Baker, C.C. A pyrimidine-rich exonic splicing suppressor binds multiple RNA splicing factors and inhibits spliceosome assembly. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 14088–14093. [CrossRef]

44. Jia, R.; Liu, X.; Tao, M.; Kruhlak, M.; Guo, M.; Meyers, C.; Baker, C.C.; Zheng, Z.M. Control of the papillomavirus early-to-late switch by differentially expressed SRp20. *J. Virol.* **2009**, *83*, 167–180. [CrossRef]

45. Tang, S.; Tao, M.; McCoy, J.P.; Zheng, Z.-M. The E7 oncoprotein is translated from spliced E6*I transcripts in high-risk human papillomavirus type 16- or type 18-positive cervical cancer cell lines via translation reinitiation. *J. Virol.* **2006**, *80*, 4249–4263. [CrossRef]

46. Rosenberger, S.; Arce, J.D.-C.; Langbein, L.; Steenbergen, R.D.M.; Rösl, F. Alternative splicing of human papillomavirus type-16 E6/E6* early mRNA is coupled to EGF signaling via Erk1/2 activation. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 7006–7011. [CrossRef]

47. Hao, C.; Zheng, Y.; Jönsson, J.; Cui, X.; Yu, H.; Wu, C.; Kajitani, N.; Schwartz, S. hnRNP G/RBMX enhances HPV16 E2 mRNA splicing through a novel splicing enhancer and inhibits production of spliced E7 oncogene mRNAs. *Nucleic Acids Res.* **2022**, *50*, 3867–3891. [CrossRef]

48. Cui, X.; Hao, C.; Gong, L.; Kajitani, N.; Schwartz, S. HnRNP D activates production of HPV16 E1 and E6 mRNAs by promoting intron retention. *Nucleic Acids Res.* **2022**, *50*, 2782–2806. [CrossRef]

49. Jönsson, J.; Wang, L.; Kajitani, N.; Schwartz, S. A novel HPV16 splicing enhancer critical for viral oncogene expression and cell immortalization. *Nucleic Acids Res.* **2024**, *52*, 316–336. [CrossRef]

50. Ajiro, M.; Jia, R.; Zhang, L.; Liu, X.; Zheng, Z.-M. Intron definition and a branch site adenosine at nt 385 vontrol RNA Splicing of HPV16 E6*I and E7 expression. *PLoS ONE* **2012**, *7*, e46412. [CrossRef]

51. McFarlane, M.; MacDonald, A.I.; Stevenson, A.; Graham, S.V. Human papillomavirus 16 oncoprotein expression is controlled by the cellular splicing factor SRSF2 (SC35). *J. Virol.* **2015**, *89*, 5276–5287. [CrossRef]

52. Ferguson, J.; Campos-León, K.; Pentland, I.; Stockton, J.D.; Günther, T.; Beggs, A.D.; Grundhoff, A.; Roberts, S.; Noyvert, B.; Parish, J.L. The chromatin insulator CTCF regulates HPV18 transcript splicing and differentiation-dependent late gene expression. *PLoS Pathog.* **2021**, *17*, e1010032. [CrossRef]

53. Milligan, S.G.; Veerapraditsin, T.; Ahamat, B.; Mole, S.; Graham, S.V. Analysis of novel human papillomavirus type 16 late mRNAs in differentiated W12 cervical epithelial cells. *Virology* **2007**, *360*, 172–181. [CrossRef]

54. Somberg, M.; Li, X.; Johansson, C.; Orru, B.; Chang, R.; Rush, M.; Fay, J.; Ryan, F.; Schwartz, S. Serine/arginine-rich protein 30c activates human papillomavirus type 16 L1 mRNA expression via a bimodal mechanism. *J. Gen. Virol.* **2011**, *92*, 2411–2421. [CrossRef]

55. Rush, M.; Zhao, X.; Schwartz, S. A splicing enhancer in the E4 coding region of human papillomavirus type 16 is required for early mRNA splicing and polyadenylation as well as inhibition of premature late gene expression. *J. Virol.* **2005**, *79*, 12002–12015. [CrossRef]

56. Li, X.; Johansson, C.; Cardoso Palacios, C.; Mossberg, A.; Dhanjal, S.; Bergvall, M.; Schwartz, S. Eight nucleotide substitutions inhibit splicing to HPV-16 3′-splice site SA3358 and reduce the efficiency by which HPV-16 increases the life span of primary human keratinocytes. *PLoS ONE* **2013**, *8*, e72776. [CrossRef]

57. Collier, B.; Oberg, D.; Zhao, X.; Schwartz, S. Specific inactivation of inhibitory sequences in the 5′ end of the human papillomavirus type 16 L1 open reading frame results in production of high levels of L1 protein in human epithelial cells. *J. Virol.* **2002**, *76*, 2739–2752. [CrossRef]

58. Li, X.; Johansson, C.; Glahder, J.; Mossberg, A.-K.; Schwartz, S. Suppression of HPV-16 late L1 5′-splice site SD3632 by binding of hnRNP D proteins and hnRNP A2/B1 to upstream AUAGUA RNA motifs. *Nucleic Acids Res.* **2013**, *41*, 10488–10508. [CrossRef]

59. Zhao, X.; Fay, J.; Lambkin, H.; Schwartz, S. Identification of a 17-nucleotide splicing enhancer in HPV-16 L1 that counteracts the effect of multiple hnRNP A1-binding splicing silencers. *Virology* **2007**, *369*, 351–363. [CrossRef]

60. Somberg, M.; Zhao, X.; Frõhlich, M.; Evander, M.; Schwartz, S. Polypyrimidine tract binding protein induces human papillomavirus type 16 late gene expression by interfering with splicing inhibitory elements at the major late 5′ splice site, SD3632. *J. Virol.* **2008**, *82*, 3665–3678. [CrossRef]

61. Kajitani, N.; Schwartz, S. The role of RNA-binding proteins in the processing of mRNAs produced by carcinogenic papillomaviruses. *Semin. Cancer Biol.* **2022**, *86*, 482–496. [CrossRef]

62. Dhanjal, S.; Kajitani, N.; Glahder, J.; Mossberg, A.-K.; Johansson, C.; Schwartz, S. Heterogeneous nuclear ribonucleoprotein C proteins interact with the human papillomavirus type 16 (HPV16) early 3′-untranslated region and alleviate suppression of HPV16 late L1 mRNA splicing. *J. Biol. Chem.* **2015**, *290*, 13354–13371. [CrossRef]

63. Kajitani, N.; Glahder, J.; Wu, C.; Yu, H.; Nilsson, K.; Schwartz, S. hnRNP L controls HPV16 RNA polyadenylation and splicing in an Akt kinase-dependent manner. *Nucleic Acids Res.* **2017**, *45*, 9654–9678. [CrossRef]

64. Yu, H.; Gong, L.; Wu, C.; Nilsson, K.; Li-Wang, X.; Schwartz, S. hnRNP G prevents inclusion on the HPV16 L1 mRNAs of the central exon between splice sites SA3358 and SD3632. *J. Gen. Virol.* **2018**, *99*, 328–343. [CrossRef]

65. Zhao, X.; Oberg, D.; Rush, M.; Fay, J.; Lambkin, H.; Schwartz, S. A 57-nucleotide upstream early polyadenylation element in human papillomavirus type 16 interacts with hFip1, CstF-64, hnRNP C1/C2 and polypyrimidine tract binding protein. *J. Virol.* **2005**, *79*, 4270–4288. [CrossRef]

66. Terhune, S.S.; Milcarek, C.; Laimins, L.A. Regulation of human papillomavirus type 31 polyadenylation during the differentiation-dependent life cycle. *J. Virol.* **1999**, *73*, 7185–7192. [CrossRef]

67. Terhune, S.S.; Hubert, W.G.; Thomas, J.T.; Laimins, L.A. Early polyadenylation signals of human papillomavirus type 31 negatively regulates capsid gene expression. *J. Virol.* **2001**, *75*, 8147–8157. [CrossRef]

68. Öberg, D.; Collier, B.; Zhao, X.; Schwartz, S. Mutational inactivation of two distinct negative RNA elements in the human papillomavirus type 16 L2 coding region induces production of high levels of L2 in human cells. *J. Virol.* **2003**, *77*, 11674–11684. [CrossRef]

69. Öberg, D.; Fay, J.; Lambkin, H.; Schwartz, S. A downstream polyadenylation element in human papillomavirus type 16 L2 encodes multiple GGG motifs and interacts with hnRNP H. *J. Virol.* **2005**, *79*, 9254–9269. [CrossRef]

70. Hegde, R.S. The papillomavirus E2 proteins. *Annu. Rev. Biophys. Biomol. Struct.* **2002**, *31*, 343–360. [CrossRef]

71. Bodaghi, S.; Jia, R.; Zheng, Z.-M. Human papillomavirus type 16 E2 and E6 are RNA-binding proteins and inhibit in vitro splicing of pre-mRNAs with suboptimal splice sites. *Virology* **2009**, *386*, 32–43. [CrossRef]

72. Sekhar, V.; Reed, S.C.; McBride, A.A. Interaction of the betapapillomavirus E2 tethering protein with mitotic chromosomes. *J. Virol.* **2010**, *84*, 543–557. [CrossRef]

73. Sekhar, V.; McBride, A.A. Phosphorylation regulates binding of the human papillomavirus type 8 E2 protein to host chromosomes. *J. Virol.* **2012**, *86*, 10047–10058. [CrossRef]

74. Lai, M.-C.; Teh, B.H.; Tarn, W.-Y. A human papillomavirus E2 transcriptional activator. *J. Biol. Chem.* **1999**, *274*, 11832–11841. [CrossRef]

75. Muller, M.; Jacob, Y.; Jones, L.; Weiss, A.; Brino, L.; Chantier, T.; Lotteau, V.; Favre, M.; Demeret, C. Large scale genotype comparison of human papillomavirus E2-host interaction networks provides new insights for E2 molecular functions. *PLoS Pathog.* **2012**, *8*, e1002761. [CrossRef]

76. Jang, M.K.; Anderson, D.E.; van Doorslaer, K.; McBride, A.A. A proteomic approach to discover and compare interacting partners of papillomavirus E2 proteins from diverse phylogenetic groups. *Proteomics* **2015**, *15*, 13. [CrossRef]

77. Johansson, C.; Somberg, M.; Li, X.; Backström Winquist, E.; Fay, J.; Ryan, F.; Pim, D.; Banks, L.; Schwartz, S. HPV-16 E2 contributes to induction of HPV-16 late gene expression by inhibiting early polyadenylation. *EMBO J.* **2012**, *31*, 3212–3227. [CrossRef]

78. Furth, P.A.; Baker, C.C. An element in the bovine papillomavirus late 3′ untranslated region reduces polyadenylation cytoplasmic RNA levels. *J. Virol.* **1991**, *65*, 5806–5812. [CrossRef]

79. Furth, P.A.; Choe, W.-T.; Rex, J.H.; Byrne, J.C.; Baker, C.C. Sequences homologous to 5′ splice sites are required for the inhibitory activity of papillomavirus late 3′ untranslated regions. *Mol. Cell Biol.* **1994**, *14*, 5278–5289.

80. Gunderson, S.I.; Polycarpou-Schwarz, M.; Mattaj, I.W. U1 snRNP inhibits pre-mRNA polyadenylation through a direct interaction between U1 70K and poly(A) polymerase. *Mol. Cell* **1998**, *1*, 255–264. [CrossRef]

81. Tan, W.; Felber, B.K.; Zolotukhin, A.S.; Pavlakis, G.N.; Schwartz, S. Efficient expression of the human papillomavirus type 16 L1 protein in epithelial cells by using rev and the rev-response element of human immunodeficiency virus or the cis-acting transactivation element of simian retrovirus type 1. *J. Virol.* **1995**, *69*, 5607–5620. [CrossRef]

82. Sokolowski, M.; Furneaux, H.; Schwartz, S. The inhibitory activity of the AU-rich RNA element in the human papillomavirus type 1 late 3′ untranslated region correlates with its affinity for the elav-like HuR protein. *J. Virol.* **1999**, *73*, 1080–1091. [CrossRef]

83. Wiklund, L.; Sokolowski, M.; Carlsson, A.; Rush, M.; Schwartz, S. Inhibition of translation by UAUUUAU and UAUUUUUAU motifs of the AU-rich RNA instability element in the HPV-1 late 3′ untranslated region. *J. Biol. Chem.* **2002**, *277*, 40462–40471. [CrossRef]

84. Srikantan, S.; Gorospe, M. HuR function in disease. *Front. Biosci.* **2012**, *17*, 189–205. [CrossRef]

85. Sokolowski, M.; Schwartz, S. Heterogeneous nuclear ribonucleoprotein C binds exclusively to the functionally important UUUUU-motifs in the human papillomavirus type-1 AU-rich inhibitory element. *Virus Res.* **2001**, *73*, 163–175. [CrossRef]

86. Smith, R.W.; Gray, N.K. Poly(A)-binding protein (PABP): A common viral target. *Biochem. J.* **2010**, *426*, 1–12. [CrossRef]

87. Kennedy, I.M.; Haddow, J.K.; Clements, J.B. Analysis of human papillomavirus type 16 late mRNA 3′ processing signals in vitro and in vivo. *J. Virol.* **1990**, *64*, 1825–1829. [CrossRef]

88. Kennedy, I.M.; Haddow, J.K.; Clements, J.B. A negative regulatory element in the human papillomavirus type 16 genome acts at the level of late mRNA stability. *J. Virol.* **1991**, *65*, 2093–2097. [CrossRef]

89. Dietrich-Goetz, W.; Kennedy, I.M.; Levins, B.; Stanley, M.A.; Clements, J.B. A cellular 65kDa protein recognizes the negative regulatory element of human papillomavirus late mRNA. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 163–168. [CrossRef]

90. Cumming, S.A.; McPhillips, M.G.; Veerapraditsin, T.; Milligan, S.G.; Graham, S.V. Activity of the human papillomavirus type 16 late negative regulatory element is partly due to four weak consensus 5′ splice sites that bind a U1 snRNP-like complex. *J. Virol.* **2003**, *77*, 5167–5177. [CrossRef]

91. Koffa, M.D.; Graham, S.V.; Takagaki, Y.; Manley, J.L.; Clements, J.B. The human papillomavirus type 16 negative regulatory element interacts with three proteins that act at different posttranscriptional levels. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 4677–4682. [CrossRef]

92. McPhillips, M.G.; Veerapraditsin, T.; Cumming, S.A.; Karali, D.; Milligan, S.G.; Boner, W.; Morgan, I.M.; Graham, S.V. SF2/ASF binds the human papillomavirus type 16 late RNA control element and is regulated during epithelial differentiation. *J. Virol.* **2004**, *78*, 10598–10605. [CrossRef]

93. De Conti, L.; Baralle, M.; Buratti, E. Exon and intron definition in pre-mRNA splicing. *Wiley Interdiscip. Rev. RNA* **2013**, *4*, 49–60. [CrossRef]

94. Cumming, S.A.; Chuen-Im, T.; Zhang, J.; Graham, S.V. The RNA stability regulator HuR regulates L1 protein expression in vivo in differentiating cervical epithelial cells. *Virology* **2009**, *383*, 142–149. [CrossRef]

95. Goraczniak, R.; Gunderson, S. The regulatory element in the 3′-untranslated region of human papillomavirus 16 inhibits expression by binding CUG-binding protein 1. *J. Biol. Chem.* **2008**, *283*, 2286–2296. [CrossRef]

96. Sokolowski, M.; Tan, W.; Jellne, M.; Schwartz, S. mRNA instability elements in the human papillomavirus type 16 L2 coding region. *J. Virol.* **1998**, *72*, 1504–1515. [CrossRef]

97. Khan, M.I.; Zhang, J.; Liu, Q. HnRNP F and hnRNP H1 regulate mRNA stability of amyloid precursor protein. *NeuroReport* **2021**, *32*, 824–832. [CrossRef]

98. Collier, B.; Goobar, L.; Sokolowski, M.; Schwartz, S. Translational inhibition in vitro of human papillomavirus type 16 L2 mRNA mediated through interaction with heterogeneous ribonucleoprotein K and poly (rC)-binding proteins 1 and 2. *J. Biol. Chem.* **1998**, *273*, 22648–22656. [CrossRef]

99. Jones, M.; Dry, I.R.; Frampton, D.; Singh, M.; Kanda, R.K.; Yee, M.B.; Kellam, P.; Hollinshead, M.; Kinchington, P.R.; O'Toole, E.A.; et al. RNA-seq analysis of host and viral gene expression highlights interaction between varicella zoster virus and keratinocyte differentiation. *PLoS Pathog.* **2014**, *10*, e1003896. [CrossRef]

100. Zandhuis, N.D.; Nicolet, B.P.; Wolkers, M.C. RNA-binding protein expression alters upon differentiation of human B cells and T cells. *Front. Immunol.* **2021**, *12*, 717324. [CrossRef]

101. Mole, S.; Milligan, S.G.; Graham, S.V. Human papillomavirus type 16 E2 protein transcriptionally activates the promoter of a key cellular splicing factor, SF2/ASF. *J. Virol.* **2009**, *83*, 357–367. [CrossRef]

102. Mole, S.; Veerapraditsin, T.; McPhillips, M.; Graham, S.V. Regulation of splicing-associated SR proteins by HPV16. *Biochem. Soc. Trans.* **2006**, *34*, 1145–1147. [CrossRef]

103. Klymenko, T.; Hernandez-Lopez, H.; MacDonald, A.I.; Bodily, J.M.; Graham, S.V. Human papillomavirus E2 regulates SRSF3 (SRp20) to promote capsid protein expression in infected differentiated keratinocytes. *J. Virol.* **2016**, *90*, 5047–5058. [CrossRef]

104. Mole, S.; Faizo, A.A.A.; Hernandez-Lopez, H.; Griffiths, M.; Stevenson, A.; Roberts, S.; Graham, S.V. Human papillomavirus type 16 infection activates the host serine arginine protein kinase 1 (SRPK1)—Splicing factor axis. *J. Gen. Virol.* **2020**, *101*, 523–532. [CrossRef]

105. Kang, D.; Lee, Y.; Lee, J.S. RNA-binding proteins in cancer: Functional and therapeutic perspectives. *Cancers* **2020**, *12*, 2699. [CrossRef]

106. Stünkel, W.; Bernard, H.U. The chromatin structure of the long control region of human papillomavirus type 16 represses viral oncoprotein expression. *J. Virol.* **1999**, *73*, 1918–1930. [CrossRef]

107. Shukla, S.; Kavak, E.; Gregory, M.; Imashimizu, M.; Shutinoski, B.; Kashlev, M.; Oberdoerffer, P.; Sandberg, R.; Oberdoerffer, S. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **2011**, *479*, 74–79. [CrossRef]

108. Pentland, I.; Campos-León, K.; Cotic, M.; Davies, K.J.; Wood, C.D.; Groves, I.J.; Burley, M.; Coleman, N.; Stockton, J.D.; Noyvert, B.; et al. Disruption of CTCF-YY1-dependent looping of the human papillomavirus genome activates differentiation-induced viral oncogene transcription. *PLoS Biol.* **2018**, *16*, e2005752. [CrossRef]

109. Sakakibara, N.; Chen, D.; McBride, A.A. Papillomaviruses use recombination-dependent replication to vegetatively amplify their genomes in differentiated cells. *PLoS Pathog.* **2013**, *9*, e1003321. [CrossRef]

110. Nilsson, K.; Wu, C.; Kajitani, N.; Yu, H.; Tsimtsirakis, E.; Gong, L.; Winquist, E.B.; Glahder, J.; Ekblad, L.; Wennerberg, J.; et al. The DNA damage response activates HPV16 late gene expression at the level of RNA processing. *Nucleic Acids Res.* **2018**, *46*, 5029–5049. [CrossRef]