LUND UNIVERSITY

The role of cellular RNA processing functions in Human papillomavirus type 16 gene regulation

Nilsson, Kersti

2019

Document Version:
Publisher's PDF, also known as Version of record

The role of cellular RNA processing functions in Human papillomavirus type 16 gene regulation

Kersti Nilsson

LUND
UNIVERSITY

DOCTORAL DISSERTATION

by due permission of the Faculty Faculty of Medicine, Lund University, Sweden.
To be defended at BMCI1345 on 19th, October at 9.00 am.

Faculty opponent
Assistant Professor Nelson Gekara, Department of Molecular Biology, Umeå University

Supervisor: Professor Stefan Schwartz
Co-supervisor: Associate Professor Ola Forslund

Organization LUND UNIVERSITY	Document name Doctoral dissertation
	Date of issue: 19th October, 2019
Author(s) Kersti Nilsson	Sponsoring organization

Title and subtitle: The role of cellular RNA processing functions in Human papillomavirus type 16 gene regulation

Abstract

Infections with Human papillomavirus type 16 (HPV16) is the most the most common, sexually transferred and is responsible for genital warts, cervical cancers and a growing number of head and neck cancers. Knowledge about how HPV16 interacts with the infected cell to regulate its gene expression is essential for therapeutic development.

Here, we show that the viral E5 protein which may contribute to carcinogenesis can only be expressed efficiently during the early infection. E5 is the last open reading frame (ORF) on the early pre-mRNAs. Therefore, its expression requires removal of upstream, inhibitory E7 and E1 ORFs by alternative splicing.

Alternative splicing is also required for HPV16 later gene expression together with an inhibition of the early polyadenylation signal. We report that induction of the DNA damage response (DDR) by alkylating agent melphalan caused both an activation of late gene splicing and inhibition of the early polyadenylation signal. The connection between induction of DDR and HPV16 late gene expression by melphalan was dependent on ATM signaling and caused an accumulation of phosphorylated BRCA1 on HPV16 DNA. BRAC1 also interacted with splice factors U2AF65 and hnRNPC that were recruited to HPV16 mRNAs to generate a expression of late, L1 and L2 HPV16 mRNAs. The ATM-branch of the DDR is hijacked by HPV for productive viral replication prior to late gene expression. Therefore, we suggest that HPV16 also has evolved to utilize the connection between DDR and RNA-processing for late gene activation.

Finally, we investigated the presence of methylated N6-adenosine (m6A) on HPV16 mRNAs in connection to regulation of alternative splicing. Here, we report that HPV16 mRNAs undergoes methylation and that the proteins associated with m6A affects viral splicing. In particular, splicing of mRNAs that express the viral oncoproteins E6 and E7 were affected in a mutually exclusive manner. The same pattern was also observed for mRNAs expressing viral replication proteins E1 and E2. Further investigations will reveal the true regulatory importance of m6A for splicing of these mRNAs. In conclusion, this thesis identifies the main E5 mRNA, connects HPV16 late gene expression to the DDR and suggests a regulatory function of m6A in HPV16 alternative splicing.

Key words: HPV16, Gene regulation, Cervical cancer, Splicing, Polyadnylation, Influenza A, Dna damage response

Classification system and/or index terms (if any)

Supplementary bibliographical information		**Language** English
ISSN and key title		**ISBN**
Recipient's notes	**Number of pages** 61	Price
	Security classification	

Signature Date 2019-09-16

The role of cellular RNA processing functions in Human papillomavirus type 16 gene regulation

Kersti Nilsson

LUND
UNIVERSITY

Faculty of Medicine
Department of Laboratory Medicine
Lund University, Faculty of Medicine Doctoral Dissertation Series 2019:92
ISBN 978-91-7619-821-6
ISSN 1652-8220

Printed in Sweden by Media-Tryck, Lund University, Lund 2019

Media-Tryck is an environmentally certified and ISO 14001:2015 certified provider of printed material. Read more about our environmental work at www.mediatryck.lu.se

MADE IN SWEDEN

To my loving family who always pick me up when I fall

Table of Contents

List of papers

This thesis is based on the following papers and manuscripts that are referred to by their roman numerical throughout.

I. **Kersti Nilsson**, Christopher Norberg, Ann-Kristin Mossberg, Stefan Schwartz. HPV16 E5 is produced from an HPV16 early mRNA spliced from SD226 to SA3358.Virus Research. 2018. 244. p. 128-136.

II. **Kersti Nilsson**, Chengjun Wu, Naoko Kajitani, Haoran Yu, Efthymios Tsimtsirakis, Lijing Gong, Ellinor B. Winquist, Jacob Glahder, Lars Ekblad, Johan Wennerberg, Stefan Schwartz. The DNA damage response activates HPV16 late gene expression art the level of RNA processing. Nucleic Acids Res. 2018; 5029-5049.

III. **Kersti Nilsson**, Chengjun Wu, Stefan Schwartz. Role of the DNA damage response in Human papillomavirus RNA splicing and polyadenylation. International Juornal of Molecular Sciences. 2018; 19(6).

IV. **Kersti Nilsson**, Xiaoxu Cui, Stefan Schwartz. HPV16 mRNAs are m6A-methylated and their splicing is modulated by the m6A writers, erasers and readers METTL3, ALKBH5 and YTHDC1.

V. **Kersti Nilsson**, Samir Abdurham, Stefan Schwartz. Influenza virus segment 8 mRNA from H1N1 and H3N2 interact with different hnRNP proteins. Manuscript.

Populärvetenskaplig sammanfattning

Två virus som frekvent orsakar både mild och allvarlig sjukdom är Humant papillomvirus (HPV) och Influensa A (IVA). HPV typ 16 (HPV16) är den vanligaste sexuellt överförbara virusinfektionen. Förutom kondylom orsakar HPV16 ca 70% av all livmoderhalscancer. Influensa orsakar inte bara årliga epidemier utan har även potentialen att skapa pandemier med allvarliga konsekvenser som de 50 miljoner liv skördade utav spanska sjukan. HPV och IVA är två helt olika virus som förutom att dom är starkt associerade med sjukdom, utnyttjar cellulär splitsling för att uttrycka sina gener i våra celler. Med mer kunskap om hur HPV och IVA interagerar med cellulära proteiner för att reglera uttrycket av de virala generna kan vi utveckla bättre behandlingsstrategier.

Kroniska HPV16 infektioner riskerar att utvecklas till cancer därför att de virala proteinerna E6 och E7 driver celldelning och inhiberar apoptos. Även ett tredje protein, E5, tros ligga bakom cancerutvecklingen men E5 är svårt att upptäcka och därmed placera i HPVs livscykel. Här beskriver vi hur E5 produceras från ett tidigt mRNA som är alternativ splitsat. Detta placerar E5 i den tidiga delen av infektionen. Dessutom kan identifieringen av E5-mRNA tillhandahålla ett sätt att detektera uttryck av E5 i patientmaterial. Vidare visas här hur splitsning av sena HP16 mRNAn som kodar för kapsidproteinerna kan kopplas till det signalnätverk som reparerar skador på DNA (DDR). Aktivering av DDR ledde till att DDR-faktorer som BRCA1 och BARD1 associerade med HPV16 DNA och positionerade splitsfaktorerna hnRNPC och U2AF65 på HPV16 mRNAn. Detta gav specifik produktion av enkom L1 och L2 mRNA genom aktivering av splitsning och inhibering av polyadenylation. Det som gör detta extra intressant är att HPV använder DDR för att replikera sitt DNA, ett steg som direkt följs av L1 och L2 produktion. Därmed föreslår vi här en modell i vilken HPV16 inte bara kapar DDR för replikering utan även att de sena generna regleras utav de splitsfaktorer som är knutna till DDR. I anknytning till regleringen av splitsning i HPV16 visar vi att det förekommer metylering av HPV16 mRNA. Överuttryck av metylerande/de-metylerande proteinerna METTL3 och ALKBH5 samt YTHDC1 som binder specifikt till metylerat RNA påverkar HPV16 splitsning så att mRNA nivåerna av både E6 vs. E7 och E1 vs. E2 förändras. Slutligen rapportera vi att splitsfaktorer som HuR, hnRNPA1, PTB och hnRNPG binder olika till segment 8 från IVA H1N1 (spanska sjukan) och H3N2. Detta kunde delvis förklara skillnaden i hur

mycket splitsade mRNA som produceras från H1N1 segment 8 jämfört med H3N2 segment 8. Skillnader i splitsning mellan olika IVA arter kan ha betydelse för deras patogena egenskaper.

Sammanfattningsvis så har vi visat att endast ett splitsat tidigt HPV16 mRNA kan stödja uttryck utav E5. Vi har också påvisat en potentiell koppling mellan uttryck av de sena HPV16 generna och DDR som kan ha stor betydelse för hur HPV16 kopplar sin livscykel till differentieringen av sin värdcell. Dessutom rapporterar vi att HPV16 mRNA metyleras och att detta ytterligare kan finjustera regleringen av viral alternativ splitsning. Slutligen beskriver vi hur sekvenspolymorfism mellan H1N1 och H3N2 segment 8 påverkar hur de interagerar med cellulära proteiner och därmed hur effektivt mRNA från segment 8 splitsas.

Abbreviations

3´ss	3´splice site, splice acceptor
5´ss	5´splice site, splice donor
ALKBH5	Alkylated DNA repair protein AlkB homolog 5
ATM	Ataxia telangiectasia mutated kinase
ATR	Ataxia telangiectasia and Rad3-related protein FRAP-related protein
BPS	Branch point sequence
BRCA1	Breast cancer type 1 susceptibility protein
CFI	Cleavage factors I
CFII	Cleavage factors II
CODON	Nucleotide triplet
CPSF	Cleavage and polyadenylation specificity factors
CstF	Cleavage stimulation factor
DDR	DNA Damage Response
DSB	Double stranded breaks
FTO	Fat mass and obesity-associated protein
H1N1	A/Brevig Mission/1918/1
H3N2	A/Netherlands/178/95
hnRNP	Heterogeneous ribonucleoprotein family
HNSCC	head and neck squamous cell carcinomas
HPV	Human papillomavirus
HR	High risk
HR	Homologous repair
LCR	Long control region
LR	Low risk
m6A	Methyl-N6-adenosine
METTL3	Methyltransferase complex containing methylatransfereaselike-3
MRN	MRE11, Rad50 and NBS1 complex
mRNA	messenger RNA

ORF	open reading frame
pAE	Early polyadenylation signal
pAL	Late polyadenylation signal
PAP	Poly(A) polymerase
PPT	Polypyrimidine tract
pre-mRNA	Precursor mRNA
RNA	Ribonucleic acid
sLuc	Secreted luciferase
snRNP	Uridine-rich small nuclear ribonucleoprotein
SR	Serine arginine rich protein family
ssDNA	single stranded DNA
UTR	Untranslated region

Introduction

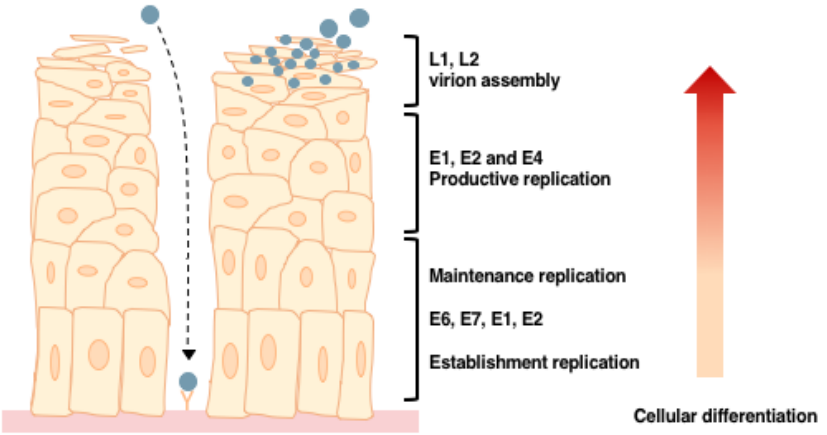
1.1 Human papillomaviruses (HPVs) and cancer

Human papillomavirus (HPV) are small DNA viruses that only infect keratinocytes of mucosal or cutaneous epithelia [1]. The almost 200 different types of HPVs generally cause asymptomatic infections that are cleared spontaneously within 2 years [2, 3]. Only in rare cases do some HPV infections persist to develop into warts or malignant lesions that, if left untreated, may progress into cancer [1, 4].

HPVs are classified into five genotypes called *alpha*, *beta*, *gamma*, *mu* and *nu* [5, 6]. The *beta* group contains cutaneous HPV types that are omnipresent and mostly cause asymptotic infections or warts[7]. In rare cases, they may contribute to melanoma and non-melanoma skin cancers[8]. However, *beta* HPVs only appears to sensitize the cell to carcinogenesis rather than driving and maintaining it. The *alpha* HPVs infect mucosal epithelia and most of them are sexually transmitted[9]. They represent the most common sexually transferred infections among humans[10]. It is persistent infections with the mucosal *alpha* HPVs that cause the majority of the HPV-induced human cancers, such as 99% of all cervical cancer and 80% of the rapidly increasing head and neck squamous cell carcinomas (HNSCC) [11]. HPV can cause other types of anogenital cancers as well, like penile and vulvar cancer, and 95% of all anal cancer[11]. Depending on their association with malignancies, *alpha* HPVs are divided into high risk- (HR, and low risk types (LR)[12]. Persistent infections with LR HPVs may cause warts, but are not associated with pre-cancers lesions and cancer. The most common HR type found in cancers and genital warts, as well as in the human population, is HPV type 16 (HPV16)[11]. HPV18 is also frequently found in cervical cancers, whereas almost all HPV-induced HNSCCs are caused by HPV16 [13-15]. Research into how HPV16 interacts with the infected cell to regulate its gene expression is crucial to understand why some infections persist to cause cancer while others do not.

1.2 The life cycle of HPV

The HPV life cycle is intimately linked to the differentiation of its host keratinocyte cell[16]. As HPV has no means of replicating its own DNA, the infection starts in the actively dividing cells of the basal layer [17](**Figure 1.1**). HPV gains access to these cells through micro abrasions in the dermis and is initially bound by heparan sulfate proteoglycans in the extracellular matrix[18]. The exact mechanism of cellular entry is unknown but involves endocytosis and requires mitosis for entry into the nucleus[19]. HPV replicates strictly in the nucleus where the early viral proteins E6 and E7 are expressed. Together, they delay cell differentiation and keep the cell cycle of their host cell active to give HPV access to cellular functions for replication and gene expression[16, 20]. E7 releases the pro-cell cycle E2F transcription factor from pRb, while E6 inhibits p53-mediated apoptosis of the infected cell by promoting p53 degradation[21, 22]. Low level expression of viral replication proteins E1 and E2 is also initiated upon entry. E1 and E2 amplify the HPV genome and maintain a low viral genome copy number. Together with the cellular DNA polymerase, they ensure that the viral genome replicates trough out the early and mid-stages of the infection[23]. As the E2 protein accumulates towards mid-infection, it inhibits expression of E6 and E7 thereby allowing the cell to resume differentiation. The transcription inhibition of E6/E7 by E2 results in increased expression of E1, E2 and E4 that all support productive replication of the HPV genome together with DNA damage response factors[23]. Immediately after productive genome amplification, the late genes L1 and L2 are expressed. L1 and L2 encode the highly immunogenic viral capsid proteins. Therefore, timing their expression and productive genome amplification to differentiated cells is pivotal to avoid detection of HPV by the immune response. Allowing the cell to becoming terminally differentiated also means that the HPV-infected cells are pushed upward against the surface of the dermis by the actively dividing cells below. In this way, HPV can hitch-hike to the dermal surface where the new virions can be released.



.

Figure 1.1 Schematic presentation of Human papillomavirus (HPV) life cycle. HPV gains entry to undifferentiated keratinocytes through micro wounds in the dermis and then migrates to the dermal surface as cells differentiate.

1.3 The genomic structure of HPV

HPV16 is an 8kb double stranded DNA, circular genome that is associated with cellular histones in a manner similar to our chromatin[24, 25]. The long control region (LCR) contains the viral origin of replication and multiple binding sites for viral proteins and cellular transcription factors that regulate the early promoter p97 that is located immediately upstream of the HPV E6 gene. In contrast, the late promoter p670 is located within the E7 coding region[24, 26]. The HPV genome also encodes two polyadenylation signals, pAE and pAL, and at least 8 proteins (**Figure 1.2**)[27]. All early (E) genes are expressed from p97 and are polyadenylated at pAE[28]. E1, E2 and E4 can also expressed from mRNAs initiated at the late p670 promoter and polyadenylated at the early polyadenylation site during mid-infection. The late HPV16 genes L1 and L2 are expressed from the late promoter and polyadenylated at pAL[28]. The use of two promoters and two polyadenylation signals divides gene expression into an early-, mid- and late phases. This division ensures a complete shut-down of E6/E7 to allow cell differentiation at late stage of infection, and prevents premature expression of the immunogenic

L1 and L2 by read-through into the late coding region at the early stage. Upstream of both polyadenylation sites are untranslated regions (UTR) that harbor multiple binding sites for cellular factors that regulate polyadenylation and stability of the HPV16 mRNAs[29].

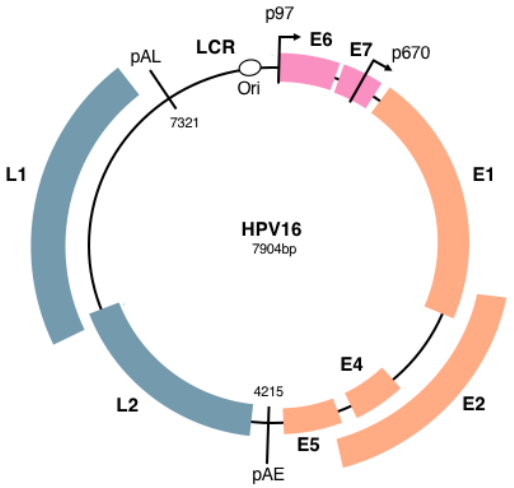


Figure 1.2 Schematic presentation of the HPV16 genome. The long control region (LCR) that contains the origin of replication (Ori) and regulatory elements for transcription factors are indicated. The early promoter p97, late promoter p670, and early and late polyadenylation signals, pAE and pAL respectively, are also indicated. The protein coding genes are highlighted in different colours to indicate their expression during the viral life cycle. White proteins are expressed in the early infection, light grey in early- to mid-infection and dark grey are restricted to the late stage of infection.

1.4 The HPV proteins

There are eight HPV proteins that have been functionally characterized. However, due to HPVs extensive use of alternative splicing, there may be several isoforms of these with slightly different, regulatory functions.

1.4.1 E6 and E7

The most well characterized function of E6 from HR HPVs is the inhibition of p53-mediated apoptosis. E6 binds simultaneously to both p53 and E3 ubiquitin ligase E6-associated protein (E6AP) to promote ubiquitination and subsequent proteosomal degradation of p53[30, 31]. E6 of HR HPVs can therefore drive immortalization of the infected cell, although less efficiently than E7. E6 is about 150 amino acids in size and has a Zinc domain in both the C-, and N-terminus that are required for p53 and E6AP interaction[32, 33]. Low risk E6 are unable to cause degradation of p53. Instead, they seem to not activate p53 apoptosis as strongly as HR HPVs and block the transcriptional activity of p53, possibly through the acetylation and cytoplasmic sequestering described above[34]. E6 itself is protected from degradation while bound to E6AP.

E6 can also inactivate p53 functions trough other mechanisms like sequestering of p53 in the cytoplasm by masking its nuclear import signal[35]. In addition, E6 can bind to CBP/p300 to inhibit acetylation of p53, a protein modification that is required for its DNA binding ability[36, 37]. Degradation, cytoplasmic sequestering and acetylation can all block p53´s function as a transcriptional activator of genes involved in the regulation of apoptosis, DNA damage response, cell cycle progression and telomerase maintenance. Shorter isoforms of E6 have been reported to have the opposite effects and may play a part in regulating the balance between the functions of E6 and E7 proteins[38]. The E6/E7 balance is pivotal for a successful infection and cell immortalization as E6 alone cannot drive the cell cycle re-entry, and E7´s elicitation of cell cycle re-entry will induce apoptosis. Thus, offsetting this balance may be a future therapeutic interest.

E7 is mainly a nuclear protein that forces the HPV infected cell to re-enter S-phase of the cell cycle. E7 has three conserved regions (CR); CR1, CR2 and CR3. They interact with the retina blastoma tumor suppressor protein (pRb or p105) and similar pocket proteins, e.g. p107 and p130[39, 40]. These pocket proteins bind to transcription factors like E2F to form a repression complex in

G1. Phosphorylation then leads to the release of E2F to promote transcription of proteins that stimulates entry into the S-phase, like cyclin A and E[41]. HPV E7 competes with E2F binding to pRb, thereby releasing E2F to activate the cell cycle[42]. The release of E2F also stimulates the expression genes that controls cellular growth. In addition, E7 mediates degradation of pRb and cyclin dependent kinase inhibitors p21 and p27[43].

HR E7 has also been shown to interfere with a number of additional transcription factors that regulate the DNA damage response and DNA repair, cell proliferation, inflammatory responses and immune responses. These transcription factors include E2F6, c-Myc, c-Jun, c-Fos, SMAD2/3, NF-κB, TBP and STAT1 and STAT[44]. Many of those are upregulated in non-HPV induced human cancers as well. E7 also contributes to productive HPV replication by activating the DDR-signaling cascade, and possibly aid in the positioning of DDR factors on the viral origin of replication (Ori) as E7 can interact directly with cellular DDR factor Nsb1[45].

1.4.2 E5

E5 is a transmembrane protein that associates with the endoplasmatic reticulum (ER), Golgi and nuclear membranes[46]. It appears to form pH-dependent ion channels by oligomerization into hexameric rings[47]. It mainly affects the infected cells by interfering with cell signaling and apoptosis, and by modulating of the immune responses. The HPV16 E5 protein has week effects on cell immortalization alone, but primarily acts to enhance E6- and E7-driven cell transformation in vitro. The function of E5 appears to be dependent on expression of epidermal growth factor receptor (EGFR) on the cellular surface[48, 49]. EGFR signaling does not only promote cell growth and is often upregulated in cancers, but has also been shown to enhance E6 expression[50]. In addition, E5 can suppress apoptosis by down regulating Fas and Bcl2 expression[51, 52]. E5 can contribute to immune evasion by reducing the expression of major histocompatibility (MHC) antigen I and II on the cellular surface [53-55]. E5 is often lost upon viral integration during carcinogenesis. In combination with technical difficulties in the detection of E5, its exact importance for the viral life cycle and carcinogenesis remains unanswered.

1.4.3 E1, E2 and E4

The HPV replication proteins E1 and E2 are involved in all stages of viral genome amplification[23]. E1 has three functional domains; an N-terminal regulatory domain, a DNA binding domain (DBD) and a C-terminal, ATP-dependent helicase domain[56, 57]. E1 is a DNA helicase that opens the viral origin of replication (Ori) to initiate genome amplification. The E1 DNA-binding domain is non-specific while E2 binds highly specifically to sites within the HPV LCR[58]. Therefore, E1 and E2 acts cooperatively to efficiently replicate the HPV DNA. After loading of E1 onto Ori, E2 is released and E1 forms the active, double hexamer helicase[59-61]. E1 interacts with cellular DNA polymerase alpha primase, topoisomerase I and replication protein A (RPA) to form a replication complex that is responsible for establishment and maintenance of replication[59]. During productive DNA replication in differentiated cells, the HPV Ori is nucleated with DNA damage response factors together with E1 and E2[45].

E2 is a 50kDa protein that, in addition to replication, has important functions in transcription during the viral life cycle. It has a conserved transactivation domain at the N-terminus and a DBD at the C-terminus[62]. The E2 DBD specifically recognizes a highly conserved 12 nucleotide sequence in the HPV LCR[60]. There are four E2 binding sites in the LCR and they span CpG islands, suggesting that methylation can regulate HPV replication and/or transcription[63]. E2 tethers the episomal viral genome to chromosomes through interactions between the transactivation domain and chromatin adapter proteins. E2 interacts with adapter protein Brd4, and Brd4 is found at the HPV Ori[64]. However, the interaction is week and there may be other tethering proteins like ChIR1[60].

E2 regulates transcription from the early, p97 promoter and thereby controls the switch from early to late HPV gene expression. Cellular transcription factors activate transcription from p97 in the immediate infection to express low levels of E2. E2 further stimulates the use of p97 by binding to a few of its LCR-binding sites. As the levels of E2 rise towards mid-infection, E2 will occupy all binding sites in the LCR and sterically block access to the TATA-box, effectively shutting down early promoter p97 [65, 66]. In addition, E2 seems to affect cellular splicing and polyadenylation throughout the viral infection. E2 is often lost during the progression to cancer as the viral genome integrates into the host chromosomes which may further accelerate the process.

The small E4 protein is the most abundantly expressed HPV protein from the onset of productive viral replication [12]. E4 has many functions such as G2 arrest of the cell cycle by sequestering Cyclin B/Cdk1 in the cytoplasm[67, 68]. Other studies have shown that E4 is involved in productive HPV genome amplification[69]. The cell cycle arrest in G2 may be an important contribution to productive DNA replication as it mitigates E6/E7 functions, allowing differentiation and giving access to DNA damage response factors[70]. Other described functions of E4 include virion assembly, RNA processing, late gene activation and release of the virions. Yet, its exact functions warrant further investigations.

1.4.4 L1 and L2

The HPV genome is packaged in a non-enveloped, icosahedral capsid made by late proteins L1 and L2. Each capsid is built of 360 copies of HPV L1 and 12 copies of L2[71]. Both L1 and L2 are about 55kD in size. The L1 protein is the most well conserved protein among HPVs and its sequence is used for classification of HPVs[5]. L2 is less conserved, apart from two cysteine residues that are required for viral transmission. L1 can self-assemble into virus like particles in the absence of both viral DNA and L2, a property that has been exploited for generation of HPV-vaccines[10]. The positively charged L1 protein interacts with the negatively charged heparan sulfate (HS) proteoglycans on the extra cellular matrix (ECM)[72]. This causes a conformational change in the viral capsid that results in cleavage of L2 by Furin. The exact mechanism of cellular entry is unknown but involves endocytosis.

1.5 Cellular RNA processing

1.5.1 Alternative splicing

Splicing is a reaction that brings sequences called exons together by removing sequences called introns from the precursor RNA (pre-mRNA) (**Figure 1.3 I**). Most of the cellular protein-encoding genes consists of small segments of coding sequences interspersed among non-coding sequences like beads on a string[73, 74]. This arrangement makes it possible to combine exons during mRNA splicing, a process termed alternative splicing. This, it is possibly to

create different mRNAs and protein products from the same gene, depending on the spatial and temporal needs of the cell. Exon/intron junctions are flanked by small sequence elements known as 5´- and a 3´-splices. The 5´ss has the consensus sequence AGGURAGU with the underlined GU as the invariable first di-nucleotide of the intron[73]. The 3´ss has a shorter consensus sequence, YAGRNNN, in which the underlined invariable AG di-nucleotide marks the extreme end of the intron. The branch point sequence (BPS) contains a conserved adenosine that plays a crucial part in the splicing reaction and is located 18-40 nucleotides upstream of the 3´ss[73, 75]. The BPS is followed by a polypyrimidine tract (PPT) in front of the 3´-splice site that plays a key role in 3´ss-recognition and regulation. These sequence elements are recognized by the spliceosome, a large RNA-protein complex that catalyzes the splicing reaction. The core of the spliceosome consists of five Uridine-rich small nuclear ribonucleoproteins (snRNPs), U1, U2, U4/U6, and U5, that carry out the splicing reaction with two transesterification reactions (**Figure 1.3 II**)[75, 76]. The 5´ss is bound by U1, and the 3´ss is recognized by U2 auxiliary factors SF1 that binds to the BPS, and U2AF65 and U2AF35 that binds to the PPT and the 3´ss AG respectively[74]. U2 then binds to the BPS to form the A-complex by ATP-dependent RNA-RNA interactions. Next, the preformed tri-snRNP U4/U5/U6 is recruited to form the B-complex in which U2 and U6 forms the first of three helical structures that align sequences and stabilize the spliceosome[76]. Additional ATP-dependent rearrangements cause releases of U1 and U4 to form the catalytic, C-complex, in which both transesterification reactions are catalyzed. First, the adenosine of the BPS attacks the 5´ss to cleave the 5´exon from the intron. Then, the 3´ss is cleaved and the two exons are ligated with the aid of U5[75]. The lariat structure formed by the removed intron is released with the snRNPs and degraded.

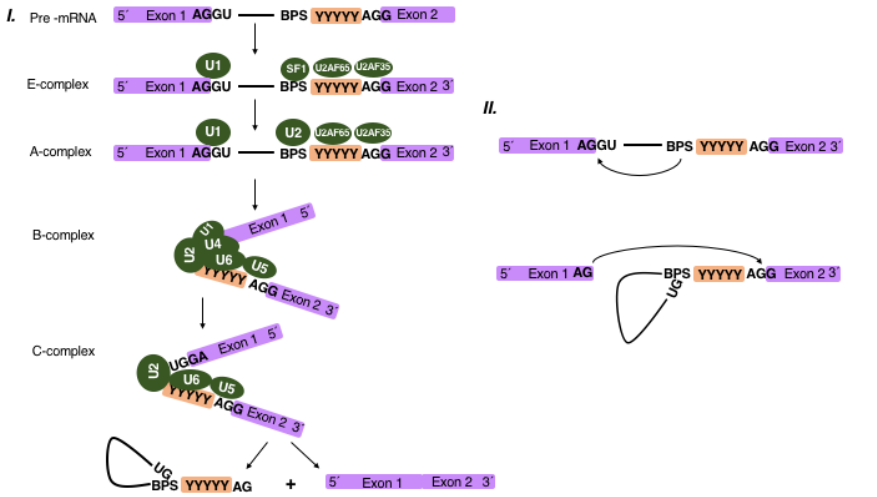


Figure 1.3 Schematic representation of the pre-mRNA splicing process and spliceosome assembly. *I.* Exons are represented as purple boxes, the intron is represented by a black line and the BPS, PPT, 3´ss and 5´ss are all indicated. A model of how the spliceosome assembles on the pre-mRNA. First, the U1 and U2 snRNPs recognize the 3´- and 5´ss with the help of auxiliary factors U2AF and SF1. Then, the U4/U6 and U5 snRNPs joins to form the active C-complex that catalyses the splicing reaction. *II.* The splicing reaction is carried out in two enzymatic steps within the spliceosome.

1.5.2 Regulation of alternative splicing

The choice of splice sites in alternative splicing is regulated by RNA binding proteins that bind to sequence elements on the pre-mRNA[77-79]. These regulatory RNA elements are usually located in the vicinity of the splice sites or overlapping them. Many of the RNA binding proteins that bind these elements belong to the Serine arginine rich protein family (SR) and Heterogeneous ribonucleoprotein family (hnRNP)[80-82]. Apart from binding to the mRNA, SR-proteins and hnRNPs can also interact with other proteins, including components of the spliceosome like U2AF65 that recognizes the 3´ss[83, 84]. The net result is to either increase or decrease the use of given splice site. In addition, secondary structures on the mRNA can affect the accessibly of both SR proteins, hnRNPs and the spliceosome. Thus, hnRNPs, SR proteins and secondary RNA structures can regulate the inclusion of exons on the mRNA.

1.5.3 Polyadenylation

Transcription of a gene ends when it passes a polyadenylation signal in eukaryotes. The maturation of the generated pre-mRNA then involves the cleavage of the 3´end at the polyadenylation signal followed by the addition of a poly(A) tail by the poly(A) polymerase (PAP) (**Figure 1.4**). The generated adenosine tail is required for translation, mRNA stability and nuclear export. The polyadenylation signal consists of the 5´-AAUAAA-3´ sequence element that is recognized by a protein complex of cleavage and polyadenylation specificity factors (CPSFs). Cleavage stimulation factors (CstFs) bind to a UG- or U-rich sequence downstream of AAUAAA, and the Cleavage factor I (CFI) binds to a UGUA element upstream of the AAUAAA[85]. This results in the formation of a large protein complex, which recruits the Cleavage factor II (CFII) that cuts the pre-mRNA. Then, PAP adds 200-300 adenosine residues to the 3´end[86]. The polyadenylation complex assembles co-transcriptionally through interactions with the C-terminal domain of RNA polymerase II[87]. Many polyadenylation sites only have two of the three described sequence elements which may be combined in any manner, still resulting in the recruitment of CFII and PAP[85]. Similar to alternative splicing, several polyadenylation signals may exist and may be regulated to generate different mRNA molecules[88]. Indeed, alternative polyadenylation is estimated to create different 3´ends on 70% of all cellular mRNAs[89, 90].

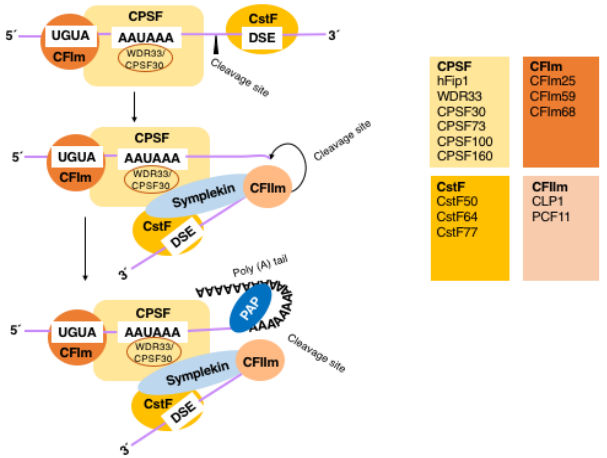


Figure 1.4 Schematic model of polyadenylation. The polyadenylation complex forms around the sequence elements that constitutes the polyadenylation signal. After cleavage by CFIIm, PAP catalyses the addition of the poly(A) tail. The components of each protein complex that is involved in recognition of polyadenylation signals are indicated to the right.

1.5.4 Translation

Translation is the process in which the mRNA becomes translated into a protein in the ribosomes. Within the ribosomes, the nucleotides of the mRNA are read in triplets (a codon) and matched with the corresponding amino acid to build the protein[91]. The mRNA is feed into the ribosome with its 5´end first and the ribosome scans the mRNA for the first codon of a protein encoding sequence referred to as an open reading frame (ORF). As there are more possible combinations of nucleotide triplets than amino acids, more than one codon can signal the start of an ORF. According to the ruling model for translation initiation, some codons create a stronger start signal than others[92, 93]. If the ribosome first encounters a week start codon, the ribosome will bypass this codon the majority of times and translation has a chance to start at the next, strong start codon instead. The ribosome translates the ORF until it reaches a stop codon where it will fall off. If a start and a stop codon slightly overlap the ribosome can also stay on the mRNA to translate a second ORF. If there is space between the stop codon of the one ORF and the start codon of the next, the ribosome can re-initiate translation at the downstream ORF[91]. There are also translation mechanisms that are independent of the mRNA 5´cap that allows to the ribosome to directly bind elements within the mRNA molecule[94].

1.6 *N*6-methylation of mRNA

There are over 100 chemical modifications on cellular RNA that could add an additional layer of regulation to mRNA biology. Among them, the *N*6 methylation of adenosine (m6A) is the most common, reversible modification[95]. M6A is also present on RNA belonging to viruses like Zika, Influenza, and HIV[96-99]. The functional significance of m6A remains elusive for cellular and viral RNA alike, but recent studies suggest regulatory functions in RNA processing, stability and translation[100, 101]. There is one known writer of m6A, a methyltransferase complex containing methyltransfereaselike-3 (METTL3), and two erasers fat mass and obesity-associated protein (FTO) and alkylated DNA repair protein AlkB homolog 5 (ALKBH5) that appears to remove m6A with different mechanisms [102-104](**Figure 1.5**). Members of the YTH-domain family bind to m6A-modified RNA to relay biological function. The nuclear YTH-protein YTHDC1 links m6A to splicing[105], while cytoplasmic YTH-proteins communicates effects

on mRNA degradation (YTHDF2)[106] and translation (YTHDF1/3)[107]. The presence of a methylation group could also result in mRNA degradation by blocking binding of mRNA-stability protein HuR[101], and promote translation by direct, cap-independent interactions with eukaryotic translation initiation factor 3 (eIF3)[108]. In addition, m6A can affect the access of RNA binding proteins to RNA molecules by disrupting secondary structures. [109, 110]. It is possible that the *N*6-methylation of adenosine could play a fine-tuning role in many cellular processes.

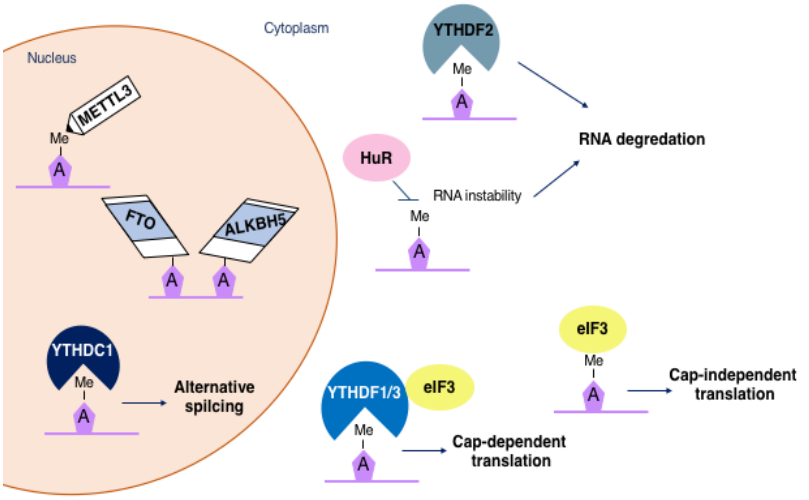


Figure 1.5 Schematic model of described writers, erasers and readers of *N*6-metyhlated adenosine (m6A) on mRNA. METTL3 adds the methyl group, while FTO and ALKBH5 functions independently to remove m6A. Readers from the YTH-domain family ties m6A to regulation of alternative splicing (YTHDC1), RNA degradation (YTHDF2), and translation (YTHDF1/3). M6A can also lead to RNA degradation through HuR, and translation initiation through direct interaction with eIF3.

1.7 HPV gene expression and regulation

1.7.1 HPV gene expression

HPV16 gene expression is highly regulated to produce each viral protein in an order that links the viral life cycle to the differentiation program of the host cell. This is accomplished both at the level of transcription initiation, and RNA processing through polyadenylation and alternative splicing[28, 111]. Any combination of promoter and polyadenylation signal produces polycistronic pre-mRNAs with multiple, and sometimes overlapping ORFs. HPV16 therefore depends on alternative splicing to generate mRNAs that can express each viral protein efficiently. Indeed, the HPV genome is riddled with splice sites and a myriad of alternatively spliced mRNAs are produced during its life cycle (**Figure 1.6**) [112]. As for cellular mRNAs, alternative splicing is carried out by the spliceosome and regulated by multiple RNA elements on the HPV16 mRNAs and their cognate, cellular splicing factors. In addition, the viral E2 protein inhibits both the early promoter and early polyadenylation signal to promote late gene expression[113]. HPV DNA is associated with histones that make it conceivable that HPV gene expression also is affected by epigenetics, although this remains to be determined[114, 115].

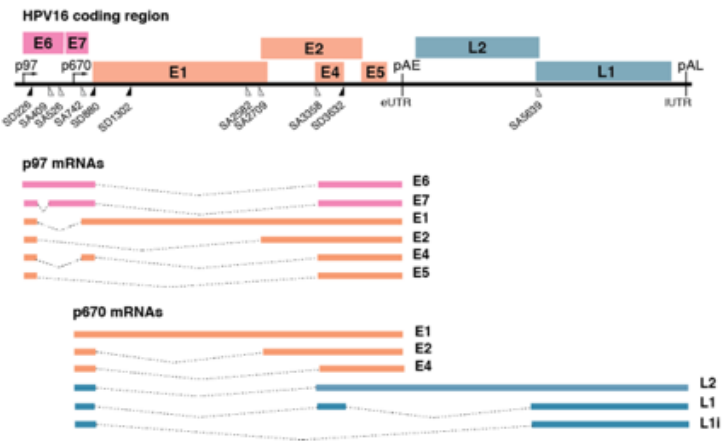


Figure 1.6 Schematic representation of the HPV16 protein coding region. The two promoters p97 and p670, and polyadenylation sites pAE and pAL are indicated. The ORF of immediate early proteins E6 and E7 are shown in pink, E1, E2, E4 and E5 are shown in orange and late L1 and L2 ORFs in blue. Splice acceptors are indicated with an empty triangle and splice donors with a full triangle. The picture shows only the major mRNA species for each viral protein.

1.7.2 Regulation of HPV late gene expression

Expression of the late, immunogenic L1 and L2 proteins is under several layers of control and requires a switch in promoter and inhibition of the early polyadenylation signal to allow transcription into the late coding region. In addition, activation of the late HPV16 splice sites is required for L1 expression[28]. The HPV E2 protein plays a key role in the transition between early and late gene expression, and HPV gene regulation depends on SR- and hnRNP proteins for correct and timely processing of the HPV mRNAs [111, 116].

During the early stage of HPV infection, cellular transcription factors bind to HPV LCR to activate the early p97 promoter. At mid-infection, the levels of E2 increase and E2 will bind to several sites in the LCR to eventually shut down the early promoter. As less E6 and E7 are expressed, the cell can resume differentiation and the late, differentiation dependent promoter p670 is activated[117]. The differentiation signal alone has been shown sufficient for activation of p670, but the exact mechanism remains unknown. Thus, both promoters are active at mid-infection, but polyadenylation still occurs at the early polyadenylation site until E2 shuts it down as well by preventing formation of the polyadenylation complex, possibly through CPSF30[113].

All mRNA transcribed from p97 are polyadenylated at the early polyadenylation site, pAE. Inactivation of pAE triggers cryptic polyadenylation in the vicinity, underlining its importance for functional HPV gene expression[28]. HPV pAE consists of the elements typical for a cellular polyadenylation site (Fig.4). They are bound by polyadenylation factors Fip1, CstF64 and cytoplasmic polyadenylation element-binding protein 1 (CPEB1) to promote polyadenylation[118]. In addition, binding of hnRNPH to a downstream sequence stimulates polyadenylation[119]. Binding of hnRNPC1/C2 and polypyrimidine tract binding protein (PTB) to the early 3´UTR inhibit pAE and activates SD3632[118]. The pAL also has an AAUAAA element but is unsensitive to E2 interference of CPSF assembly. Little is known about the regulation of pAL, but a strong, negative regulatory RNA-element is located in the late 3´UTR. This element is conserved among various HPV types and probably serves to further ensure efficient suppression of HPV late gene expression at the early stage of infection[28]. U1 snRNP and CUGBP binds to this negative RNA element [120].

The splice acceptor (SA) 3358 is the most commonly used splice acceptor throughout the viral life cycle[112]. It is used to connect early mRNAs to the

early polyadenylation signal, and it is used for inclusion of the central E4-exon between SA3358 and SD3632 in the late, L1 mRNA. Inactivation of SA3358 will generate the shorter L1 mRNA, L1i[116]. Binding of hnRNPG downstream of SA3358 will also inhibit E4-exon inclusion to promote the L1i mRNA[121]. The 3´splice site SA3358 has low homology to a canonical 3´ss and relies on multiple SRSF1-binding sites in a downstream regulatory element for spliceosome recognition[116]. Low levels of SRF1 will still promote SA3358 splicing, but to produce late mRNAs[122]. Further downstream, there is a negative regulatory element to which SRSF3 and SRSF6 bind to inhibit SA3358 splicing[123, 124]. As levels of both hnRNPH and SRSF1 are high in undifferentiated cells, it is likely that they together promote splicing at SA3358 in connection with cleavage at pAE to generate early HPV mRNAs. When their levels drop upon differentiation, proteins like PTB and hnRNPC can inhibit pAE and connect SA3358 to late HPV mRNA production instead.

The late splice sites SD3632 and SA5639 are activated in response to cellular differentiation to generate L1 and L1i mRNAs in terminally differentiated cells. SD3632 is flanked by two RNA elements that strongly suppress its use in undifferentiated keratinocytes[28, 116]. Overexpression of hnRNPC seems to interfere with SD3632 suppression and contributes to activation of SD3632[125]. Splicing to late 3´splice site SA5639 is dependent on an immediate downstream, 17nt sequence with unknown binding partner[126]. A major splicing silencer sequence are positioned adjacent to the 17-nucleotide enhancer. These sequences contain binding sites for hnRNPH and hnRNPA1 that efficiently suppress usage of SA5639 in undifferentiated cells. As described for hnRNPH and SRSF1, levels of hnRNPA1 decrease with differentiation, thereby alleviating SA5639 suppression.

1.8 HPV genome replication

1.8.1 The DNA damage response and ATM and ATR signalling pathways

The DNA damage response (DDR) comprises a large protein network that detects and signals the presence of genotoxic stress to either activate repair or apoptosis when repair is not possible. Each type of damage has its unique sensor that passes the information to kinases like Ataxia telangiectasia mutated kinase (ATM) and Ataxia telangiectasia and Rad3-related protein FRAP-related protein (ATR)[127]. The signal is then relayed mainly through phosphorylation of downstream factors and additional kinases, like Chk1 and Chk2, to organize the appropriate responses such as DNA synthesis to restore the DNA molecule after the damage has been removed[128, 129]. This feature is exploited by some DNA viruses like Herpes-, Polyoma- and Papillomavirus for the syntheses of their own genomic DNA.

ATM is activated by double stranded breaks (DSB) [130] while ATR is activated by the presence of single stranded DNA[131, 132]. Abnormal stretches of ssDNA can arise from several types of DNA insults when these DNA damages interfere with DNA replication or transcription. ssDNA-stretches can also be created during repair processes such as the repair of DSBs. Therefore, there is a slight overlap in the ATM and ATR signaling pathways, but they are not redundant[128, 131]. Signaling cascades from both ATM and ATR kinases can lead to large scale activation of processes such as damage-repair, cell cycle progression, senescence, apoptosis and/or cell proliferation[133].

The cell is alerted to the presence of double stranded breaks (DSBs) by the MRN complex, (Mre11, Rad50, Nbs1) (**Figure 1.7**)[128, 130]. The MRN complex activates ATM through acetylation by the acetyltransferase Tip60. ATM passes the signal to targets such as Chk2, p38, p53, SMC1 and Breast cancer type 1 susceptibility protein (BRCA1) through a series of phosphorylation events. ATM also phosphorylates the histone variant H2A.X (γH2A.X) over the area surrounding the DSB. The newly formed γH2A.X is then recognized by Mcd1, which will lead to the recruitment of homologous repair factors (HR) such as BRCA1 and Rad51 to the site of damage.

ssDNA is bound and stabilized by replication protein A (RPA). RPA attracts ATRIP, the 9-1-1 complex (Rad9-Hus1-Rad1) and TopBP1, and all three of

these components are required for proper activation of ATR[127, 131]. Unlike ATM, ATR causes phosphorylation of target proteins in complex with other proteins to create regulatory modules. Another example is Claspin that mediates the activation of downstream kinase Chk1 to activate downstream events similar to ATM.

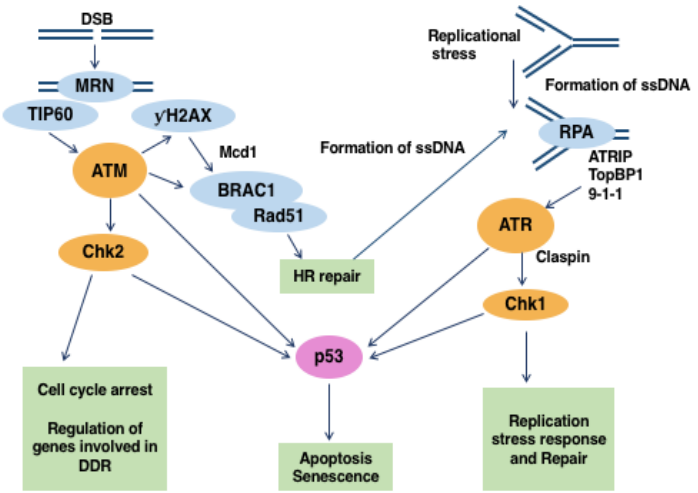


Figure 1.7 Schematic representation of DNA-damage response (DDR) signalling through ATM and ATR signalling. ATM is activated in response to double stranded breaks (DSB) which are detected by the MRN complex. ATR is activated by ssDNA detected by RPA.

1.8.2 HPV genome amplification

Amplification of the HPV genome follows three phases; establishment-, maintenance- and productive-replication[23]. As the infection originates from actively dividing cells, HPV proteins E1 and E2 can support establishment replication by recruiting the existing DNA replication machinery. E2 nucleates the HPV origin of replication with viral E1 DNA helicase that interact with cellular replication factors to yield about 50-100 copies of the viral genome[60, 134]. To delay cell cycle exit and terminal differentiation of the infected keratinocyte, HPV expresses the E6 and E7 proteins mainly during early stages

of infection. E7 pushes the infected cells back into S-phase by alleviating suppression of cellular transcription factor E2F. Elicitation of cell cycle re-entry would induce apoptosis which is why the E6 protein inhibits by degrading p53. Meanwhile, the viral genome copy number is maintained at about 50-100 copies per new daughter cells as cells divide during the maintenance phase[23].

Because the E7-induced S-phase is a forced situation, the DNA replication machinery will coexist with differentiation factors. As E2 starts to inhibit p97 promoter at mid-infection, these differentiation factors activate the late, differentiation dependent promoter. When the levels of E1, E2 and E4 are high enough to support productive replication, there is a burst in viral genome copy number[23]. This step only occurs in differentiated cells high up in the epithelium. It is generally believed that this delay in productive replication and HPV late gene expression contributes to the ability of HPV to avoid immune detection. Since differentiation results in entry into G2 phase, productive replication is dependent on the DNA damage response.

1.8.3 HPV replication and the DNA damage response

Elicitation of the DNA damage response provides a means of conducting DNA replication outside of the cell cycle, as in differentiated keratinocytes. HPV activates both ATR and ATM, and factors from both signaling branches are required for amplification of the viral genome[45, 135]. HPV E1 and E7 proteins activate ATR and downstream Chk1 by unknown mechanisms (**Figure 1.8**)[136-138]. E1 also appears to activate ATM and by randomly inducing DSB on cellular DNA due to its unspecific DNA helicase activity[137]. Several mechanisms for activating and maintaining ATM and ATR- mediated DDR have been described for E7 and include; protein stabilization, transcription activation through factors like E2F and STAT5 that increase the levels of DDR factors, and activation of the Tip60 acetyltransferase[45, 139]. Stimulation of signal transducer and trans-activator 5 protein (STAT5) has been linked to ATR activation, and Tip60 acetylation activates ATM[138, 140]. However, the ATR and ATM branches crosstalk and there are variations in the activation and dependence of them among HPV types[141].

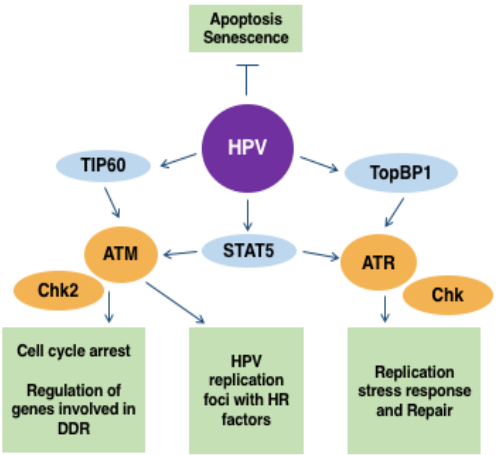


Figure 1.8 Schematic model of known interactions between HPV and the DNA damage response during productive genome amplification in differentiated cells. HPV elicits the Ataxia-Telangiectasia Mutated (ATM) and ATM and Rad3-related (ATR) signaling pathways, resulting in the recruitment of Homologous repair (HR) factors to replication foci.

For HPV16, ATR is active throughout the infection, but weather this is required for all stages of replication or merely is a consequence of a replicating virus in the cell remains unclear. The TopBP1 topoisomerase that responds to the presence of single stranded DNA (ssDNA) upstream of ATR activation is, however, required for all phases of HPV replication[45]. TopBP1 is recruited to RPA-filaments on cellular ssDNA. As RPA is a component of the HPV replication complex already during the early stages of the infection, it is likely that RPA is responsible for TopBP1 recruitment and that subsequent ATR activation is consequential rather than required for productive replication.

As replication moves into the productive phase, large foci of HPV DNA and DDR factors form. Homologous repair (HR) factors Rad51, BRCA1 and the MRN (MRE11, Rad50 and NBS1) complex are bound to the HPV origin of replication together with E1 and E2, all of which are required for productive replication[135, 142]. ATM and downstream kinases Chk2, p38 and MK2 are activated by this complex[136]. The H2A.X histones associated with HPV DNA are also phosphorylated to generate the DSB-signaling hallmark γH2AX, which further aids in the recruitment of HR-factors[45]. HR is a repair mechanism that re-synthesize a large stretch of DNA by using the homologous sequence of the sister chromatid as a template. Therefore, HPV may have

evolved to specifically use HR-factors as they offer high-fidelity replication of long stretches of DNA and are available during G2.

1.9 Influenza virus

All viruses are constrained by host adaptation and the limited coding capacity of small genomes. Therefore, a chimeric of commonly employed strategies to replicate within the host cell can be found across the spectra of viral diversity. Although a single stranded RNA virus, Influenza A (IVA) replicates in the nucleus where it hijacks the spliceosome for processing two of the eight viral mRNAs (segments 7 and 8)[143]. Similar to HPV, cellular proteins are believed to take part in the regulation of IVA splicing[143]. However, the identity and importance of such cellular proteins are largely unknown for segment 8 mRNA splicing. Recently, it we have shown that there are differences in the splicing efficiency between different Influenza virus A species [144]. Therefore, we wished to identify which splicing regulatory proteins that bind to Influenza segment 8 mRNAs, and investigate if their finding could explain the observed differences in splicing between highly pathogenic A/Brevig Misson/1918/1 (H1N1) and low pathogenic A/Netherlands/178/95 (H3N2).

1.9.1 Influenza virus A pathogenesis

Influenza viruses belongs to the Orthomyxoviridae family and this family is divided into type A, B, C, or D. Influenza A (IVA) causes global, seasonal epidemics and occasional pandemics of respiratory disease[145]. Seasonal flu virus causes influenza while pandemic flu can have a more severe toll on human life. For example, the 1918 Spanish flu virus that killed 50 million people [146]. However, severe disease and death is most commonly attributed to secondary bacterial infections that cause pneumonia[145]. Influenza A can infect humans, pigs and birds, and are restricted by the type of sialic acid displayed on the cell surface. Aquatic birds create a reservoir for IVA where IVA causes asymptomatic or mild gastrointestinal infections. Influenza has a segmented genome of eight ssRNA segments with negative polarity that encodes its own polymerase. A segmented genome allows for re-assortment of segments from different IVA species, referred to as genetic shift, while the sloppy IVA polymerase generates a rapid accumulation of mutations,

generating genetic drift. Host and tissue tropism can both change in response to either drift or shift as novel IVA species emerges[145]. It is therefore difficult to predict the severity of pandemics, as well as to develop vaccines.

The pathogenicity of an influenza species is determined by a combination of virus properties. They include the influenza species capacity to replicate to high levels within the cell and the ability to replicate in the lower respiratory tract. In addition, the modulation of the host immune response is a key factor. Virulence may be determined by how well the influenza species inhibits the innate immune response and the number of cytokines the infection induces[147]. Sequence variations in the surface proteins haemagglutinin (HA) and neuraminidase (NA) can allow infection of the lower respiratory tract. Genetic drift of the viral polymerase genes may change the replicative ability of the virus[147]. Influenza NS1 protein interferes with the innate immune response and sequence variations in NS1 may contribute to immune evasion and cytokine production[148]. NS1 is expressed from the unspliced IVA segment 8 mRNA. Variations in the ratio between spliced and unspliced segment 8 have been shown to occur between different IVA species [144]. Therefore, it is conceivable that also the levels of NS1 produced by different IVA species may contribute to pathogenicity of IVA.

1.9.2 The life cycle of human Influenza A

The Influenza A genome is organized into eight ssRNA segments of negative polarity. They are associated with numerus copies of the viral nucleoprotein (NP) that create rod-like capsid structures[149]. Each RNA is also bound to one copy of the viral polymerase complex which consists of PB1, PB2 and PA. The protein bound RNA structures (vRNPs) are then packaged into an enveloped capsid containing the viral surface antigens hemagglutinin (HA) and neuraminidase (NA)[149]. Infections starts by HA binding to sialic acid receptors on the cellular surface which triggers endocytosis (**Figure 1.9**). Inside the cell, the vRNPs are transferred to the nucleus with the help of a nuclear import signal[150, 151]. IVA transcription and replication are carried out by the viral polymerase. The viral polymerase also adds a smaller polyadenosine tail to the viral mRNAs and a 5´cap by cap-snatching from cellular mRNAs[152]. The generated mRNAs are then either spliced (segment 7 and 8) or exported directly to the cytoplasm for translation of the viral proteins. Segment 8 encodes the NS1 protein that is translated from the unspliced mRNA and the NS2 protein that is translated from the spliced mRNA. HA and NA mRNAs are translated by ribosomes on the endoplasmatic

reticulum for plasma membrane incorporation. New vRNPs are then assembled in the cytoplasm and the new virion can bud from the host cell through cleavage by NA[150].

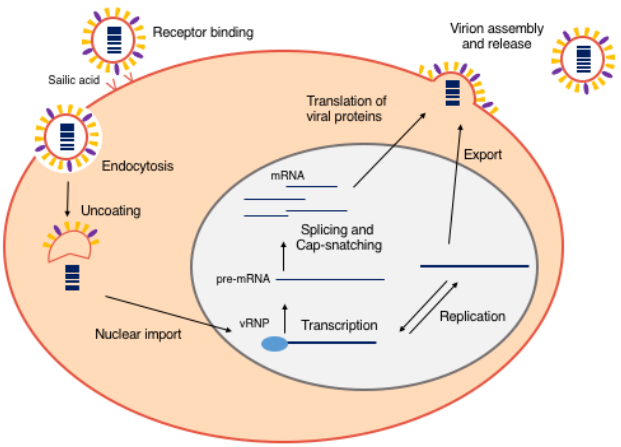


Figure 1.9 Schematic presentation of the life cycle of Influenza A. The influenza virion enters the cell through endocytosis an the vRNPs are imported to the nucleus for transcription and replication. Two of the influenza segemnts are splicied prior to export. The viral proteins are then translated and a new virion can assemble.

1.9.3 Influenza virus A splicing

Influenza mRNA splicing is dependent on the cellular splicing machinery[153, 154]. The IVA splice sites shows strong homology to canonical cellular splicing elements and are highly conserved between Influenza species[143]. At least two proteins are expressed from influenza segment 7: M1 and M2 (**Figure 1.10**). M1 is expressed from unspliced mRNAs while M2 is expressed from mRNAs spliced between SD51 and SA740. The presence of a third transcript spliced from an alternative 5´ss to SA740 have been described for several Influenza virus species, and most of what is known about influenza A mRNA splicing come from studies on 5´ss selection in segment 7 splicing[143]. The second 5´ss lies only about 11 nucleotides into segment 7

and binding of the IVA polymerase to the RNA would therefore block this 5´ss. The IVA NS1 protein has been shown to regulate segment 7 mRNA splicing in a manner dependent on the NS1 RNA-binding domain[155, 156]. Other studies contradict a regulatory role for NS1 in influenza virus A splicing of either segment 7, or 8 mRNAs [157, 158]. The RNA sequence over the 3´ss SA740 form extensive secondary structures that regulate splicing of segment 7[159-161]. The open conformation creates a hairpin structure that expose the BPS, the 3´ss and binding sites for SRSF1 that recruit the spliceosome[162]. The closed conformation buries the BPS, 3´ss and the SRSF1 binding sequence in a pseudoknot which prevents splicing.

The IVA segment 8 generates unspliced NS1 mRNAs and spliced NS2 mRNAs[153]. It is generally believed that segment 8 splicing is less complex than that of segment 7 mRNAs. There is one 5´ss SD54, and one 3´ss SA526 on the influenza virus A segment 8. A recent study identified a novel 5´ss located further into the NS1 coding sequence in a few IVA species, suggesting adaptive evolution[163]. Similar to segment 7, the formation of extensive secondary structures over the 3´ss have been described for segment 8, but no mechanistic studies have connected them to regulation of splicing[159-161]. There is a long-standing theory that segment 8 splicing is kept constant throughout the IVA infection and that changes in NS1/NS2 protein levels are due changes in mRNA export[157]. However, we have reported variation in splicing efficiency of both segment 7 and 8 between IVA isolates[144].

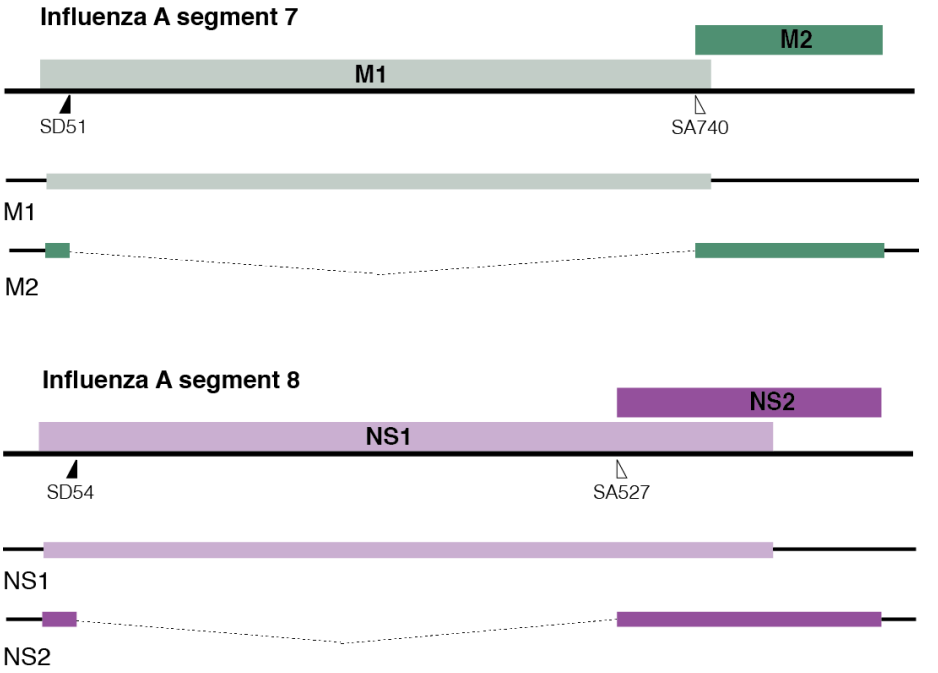


Figure 1.10 schematic presentation of splicing from the Influenza virus A segemnts 7 and 8. Proteincoding sequences are dispalyed as coloured boxen. The position of 3´ss and 5´ss on each genomic RNA segemnt an their major mRNAs are indicated.

Introduction to papers

2.1 Paper 1

HPV16 E5 is produced from an HPV16 early mRNA spliced from SD226 to SA3358

The HPV16 E5 protein is generally described as an oncogene. However, detection of the E5 protein in cervical lesions has been difficult which makes the true contribution of E5 to pathogenesis somewhat of an enigma. An alternative method to protein detection could be identification and quantification of the E5 mRNA levels in infected cells and patient samples.

Due to the location of the E5 open reading frame on the HPV16 genome, E5 could potentially be expressed from the majority of the plethora of alternatively spliced mRNAs produced by HPV16 during its life cycle. As the identity of the E5 mRNA is unknown, we investigated the E5 coding capacity of the major HPV16 mRNAs species using cDNA plasmids. By replacing the E5 ORF with the reporter gene secreted luciferase (sLuc), we could quantitate sLuc as a measurement of how well E5 was translated from each mRNA.

Our results showed that mRNAs with E1 and E7 AUGs upstream of E5 were inefficiently translated to E5 protein. Only the early HPV16 mRNA spliced from SD226 to SA3358 on which neither the E1, nor the E7 AUG was present produced high levels of E5. These results were confirmed with experiments on the full length HPV16 genome. Identification of the major E5 mRNA could aid in future research on the role of E5 in the HPV16 life cycle and pathogenesis.

2.2 Paper 2

The DNA damage response activates HPV16 late gene expression at the level of RNA-processing

HPV16 hijacks components of the cellular DNA-damage response to replicate the HPV DNA genome in differentiating cells. This step is immediately followed by expression of the HPV16 late genes L1 and L2. Therefore, we speculated that these two events are connected through the DNA damage response.

We show that activation of the DNA-damage response by alkylating cancer drug melphalan induced HPV16 late gene expression. DDR activation inhibited the early polyadenylation signal which resulted in read-through into the late coding region of the HPV16 genome and production of the unspliced late L2 mRNA. In parallel, the late splice sites SD3632 and SA5639 were activated to generate late L1 mRNAs. These changes in HPV16 gene expression were dependent on ATM- and Chk1/2-signaling.

At the same time, the association of phosphorylated BRCA1, BARD1, BCLAF1 and TRAP150 with HPV16 DNA increased, while the RNA processing factors U2AF65 and hnRNPC increased in their association with HPV16 mRNAs. Thus, the DDR-factors that assembled on HPV16 DNA recruited RNA processing factors that interacted with HPV16 mRNAs and induced HPV16 late gene expression.

2.3 Paper 3

The role of the DNA damage response in Human papillomavirus RNA splicing and polyadenylation

The HPV infection elicits the cellular DNA-damage response and the major DDR-signaling branches activated by ATM and ATR are required for HPV genome amplification in differentiated cells. Immediately after the HPV genome has been replicated, HPV16 late gene expression is activated and we hypothesized that these two events are connected.

Recently, it has been shown that the DDR also directs RNA processing events of a subset of cellular mRNAs to which the spliceosome was directly recruited by DDR factors. We found that activation of the DDR with DNA-alkylating drug melphalan increased the association of DNA damage repair factor with HPV DNA and activated HPV16 late gene expression by recruitment of splice factors.

Here, we propose a model in which HPV16 has evolved to utilize the RNA processing events connected to DDR for the activation of HPV16 late gene expression. The DDR factors already assembled on HPV16 chromatin recruit RNA binding proteins that inhibit the early polyadenylation signal and activate the HPV16 late splice sites to produce L2 and L1 mRNAs.

2.4 Paper 4 (manuscript)

HPV16 mRNAs are m6A-methylated and their splicing is modulated by the m6A writers, erasers and readers METTL3, ALKBH5 and YTHDC1

The presence of a reversible methylation on N6 of adenosine (m6A) have long been known for many viruses other than HPV. However, its function has remained an enigma for decades. Recently, m6A has been connected to splicing of both cellular mRNAs and non-coding RNAs. As HPV replicates in the nucleus where it produces a myriad of alternatively spliced mRNAs, we wished to investigate if m6A is present on HPV16 mRNAs and if this modification affects HPV RNA processing.

We found that HPV16 mRNAs were m6A methylated in both transfected and HPV16-infected cells using immunoprecipitation with m6A-specific antibody. Furthermore, overexpression of m6A-writer METTL3 and m6A-eraser ALKBH5 had the opposite effects on splicing of E6/E7 mRNAs. Overexpression of ALKBH5 also caused the exclusion of an exon in the L1 mRNA, generating the shorter isoform L1i. The nuclear m6A-reader YTHDC1 also caused changes in both E6/E7 and E2 splicing by inhibiting it.

In conclusion, we found that HPV16 mRNAs are subject to m6A modification and that proteins that modulate this modification affect HPV16 mRNA splicing.

2.5 Paper 5 (manuscript)

Influenza virus segment 8 from H1N1 and H3N2 interacts with different hnRNP proteins

We previously observed variations in splicing efficiency between segment 8 of different influenza isolates. Therefore, we wished to investigate if sequence polymorphism in segment 8 from various influenza isolates affected the binding of cellular splice factors to viral mRNAs. We speculated that such differences may account for the observed differences in segment 8 splicing efficiency.

Here, we report that segment 8 mRNAs from influenza A/Brevig Mission/1918/1 (H1N1) and A/Netherlands/178/95 (H3N2) interact with several members of the heterogeneous nuclear ribonucleoprotein family (hnRNP). We identified several "hot-spots" for binding of hnRNPs to the Influenza virus mRNAs. Interestingly, sequence variation between H1N1 and H3N2 segment 8 resulted in specific interaction of hnRNPs with either H1N1 or H3N2. Mutational replacement of H1N1 sequences with corresponding H3N2 sequences over regions in which difference in hnRNP binding were observed resulted in altered splicing of H1N1 segment 8 mRNAs. We conclude that the lower efficiency of splicing observed for H1N1 segment 8 mRNAs compared to H3N2 mRNA in part could be explained by differential binding of hnRNPs to these mRNAs.

Segment 8 encodes two proteins: NS1 that is expressed from the unspliced segment 8 mRNA and NS2 that is expressed from the spliced mRNA. The viral NS1 protein contributes to influenza pathogenicity by interfering with innate immune responses during the IVA infection. Therefore, sequence variations among influenza species may not only affect pathogenicity by affecting the NS1 protein sequence and function, but also affect its expression levels through mRNA splicing.

Concluding remarks and future perspectives

Collectively, the results presented in this thesis underlines the importance of RNA processing events for both HPV16 and Influenza H1N1 and H3N2 gene expression, and the impact that cellular factors have on the control of viral mRNA processing.

We found that HPV16 E5 protein is only expressed in high levels from one of all alternatively spliced HPV16 mRNAs, transcribed from the p97 promoter and spliced between SD226 and SA3358. Splicing between SD226 and SA3358 removes upstream ORFs that otherwise would inhibit E5 translation. Identification of the major E5 mRNA supports an early functional role for E5 in the life cycle of HPV16. It may also assist in the detection of E5 expression in cell cultures and patient samples in future investigations into the role of E5 in the HPV16 life cycle and carcinogenesis.

We found that the DNA damage response activated by the DNA alkylating cancer drug melphalan induced HPV16 late gene expression. The activation of late genes was dependent on ATM signaling and lead to the phosphorylation of BRCA1. The phosphorylated form of BRCA1 interacted with BCLAF1 and splicing factors U2AF65 and hnRNPC. Phosphorylated BRCA1 also associated with HPV DNA, which allowed BRCA1 to position the splicing factors onto HPV16 mRNAs. Consequently, the HPV16 early polyadenylation site was inhibited and late splice sites SD3632 and SA5639 activated to produce HPV16 late L1 and L2 mRNAs.

BRCA1 has been shown to connect the DNA damage response to splicing of a subset of cellular mRNAs. During productive replication of the HPV genome, BRCA1 and associated homologues repair proteins nucleate the HPV origin of replication. These proteins allow the virus to replicate in differentiated cells. Therefore, this thesis proposes a model in which components of the DNA damage repair already situated on the HPV genome for productive amplification, also activate HPV late gene expression (**Figure 3.1**). To truly

test this model, experimental system that can illustrate the complete life cycle of HPV16 during cellular differentiation are necessary.

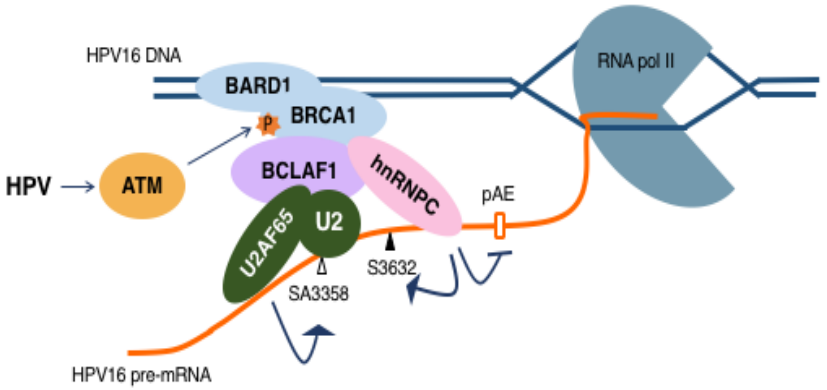


Figure 3.1 Schematic model of how the DNA damage response could activate HPV16 late gene expression through recruitment of splice factors. The DNA damage respons cpuld activly recruit the spliceosome and regulatory proteins to HPV16 mRNAs to activate late gene expression following genome amplification.

In addition, we found that HPV16 mRNAs undergo m6A-methylation and that the proteins that creates, removes and reads this modification can affect the splicing of early HPV16 mRNAs encoding E6/E7, E2 and E1. Little is known about the regulation of these mRNAs. Therefore, this project may provide additional information on how splicing is regulated in the early stages of the HPV infection. Interestingly, studies on cellular mRNA splicing have made connections between m6A and hnRNPG, hnRNPC and hnRNPA2B1[104, 109, 164]. All of these splicing regulatory proteins have previously been shown to affect HPV16 splicing[29]. Prediction software place the modified adenosine residues around the splice sites located in the early coding region, but confirming the exact positions of the modified adenosines will be crucial in order the truly decipher the role of m6A in HPV16 mRNA processing.

Also splicing of Influenza segment 8 is affected by cellular proteins. There are many hot spots for binding of hnRNPs on segment 8 mRNAs from both A/Brevig Misson/1918/1 (H1N1) and influenza virus A/Netherlands/178/95

(H3N2). Strikingly, small sequence variations had a large impact on both the occurrence of hot-spots for hnRNP binding and the efficiency of segment 8 mRNA splicing. The observation that IVA species can interact with different cellular proteins to generate variable amounts of NS1 and NS2 mRNAs may increase our understanding of Influenza pathogenicity.

Method summery

4.1 Plasmid construction

The HPV16 cDNA expression plasmids were the E5 open reading frame was replaced by the coding sequence of Meritida longa secreted luciferase (sLuc) was generated by introducing PCR fragments into the previously described cloning vector pC086[165]. pCMVSEAP and pHPV16AN was previously described [166, 167]. pHPV16E5sL contains the full HPV16 genome with the E5 ORF replaced by sLuc and was generated from the previously described pHPV16AN plasmid[168]. Inactivation of splice sites and start codons on the pHPV16E5sL was conducted by either site directed mutagenesis or site directed ligase independent mutagenesis (SLIM) [169]. A detailed version of the cloning and mutational procedure can be found in paper *I*.

The following plasmids have been described previously: pBELsLuc, pBEL, pHPV16ANsL, hnRNPC1[125]. PcDNA3-FLAG-HA-hYTHDC1(#85167), pcDNA3/FlagMETTL3 (#53739), pcDNA3/Flag-METTL14 (#53740), pcDNA3/FlagWTAP (#53741), pFRT/TO/HIS/FLAG/HA-ALKBH5 (#38073) were all bought at Addagene while pcDNA3.1+ /C-(K)DYK-FTO was purchased from GenScript. Plasmids Flag-TRAP150, FI3-BARD1/pIRESpuro, pBCLAF1 and pCU2AF65 were gifts from Drs Woan-Yuh Tarn (Institute of Biomedical Sciences, Taiwan), Richard Bear (Columbia University, USA), Jun Tang (Anhui Agricultural University, China) and Jesus Valdes Flores (CINVESTAV, Mexico), respectively.

The plasmids pcH1N1 and pcH3N2 have been described previously[144]. All other plasmids created for Paper *V* was constructed using the SLIM method [169] and a detailed description of the plasmid construction ca be fund in Paper *V*.

4.2 Cells, transfections and siRNA transfections

The following cell lines were used; lung cancer cell line A549, HPV-positive cervical cancer cell line HeLa, HPV-negative cervical cancer cell line C33A, kidney cancer cell line 293T and neonatal human epidermal keratinocytes HEKn. In addition, three in house cell lines were used, C33A2, C33A6 and HN26. C33A2 and C33A6 are derived from C33A cells that were stably transfected with the subgenomic HPV16 plasmid pBELsLuc. In the pBELsLuc construct, part of the L1 coding region have been exchanged for a poliovirus 2A internal ribosome entry site (IRES) in front of the Metridia longa secreted luciferase (sLuc) gene. Expression of the late HPV16 genes will therefore result in sLuc in the cell culture medium. The HN26 cell line originates from a tonsillar cancer and contains episomal HPV16 DNA. All cells except HN26 and HEKn were cultured in Dulbecco's modified Eagle medium with 10% bovine calf serum and 1% penicillinstreptomycin. HN26 cells were propagated in RPMI-1640 medium (GE Healthcare Life Science HyClone Laboratories) supplemented with 10% Bovine Calf Serum (GE Healthcare Life Science HyClone Laboratories), 1% Sodium Pyruvate solution (Sigma-Aldrich),1% Non-Essential Amino Acid (Sigma-Aldrich) and Gentamycin. HEKn cells were cultured in EpiLife Medium supplemented with Human keratinocyte Growth Supplement (Gibco Invitrogen Cell Culture). All plasmid transfections were made in 60mm cell culture dishes using either Fugene (Promega), or Turbofect (Fermentas) according to the manufacturer's instructions. The amount of plasmid used and time of harvest post transfection is indicated in each projects Material and Methods sections. siRNA in the form of siRNA SMARTpools from GE Healthcare Dharmacon were transfected with DharmaFECT1 (GE Healthcare Dharmacon) and cells or cell culture medium were harvested 72hrs post transfection.

4.3 sLUC-, SEAP-, CAT-ELISA- and MTT-assays

Meritida longa secreted luciferase (sLuc) was measured in cell culture medium with the ready to Ready To Glow Secreted Luciferase Reporter assay according to the instructions of the manufacturer (Clontech) and luminescence was monitored in a Tristar LB941 Luminometer. The activity of secreted embryonic alkaline phosphatase (SEAP) was measured after heat inactivation of the bovine calf serum in cell culture media for 5 min at 65°C. 100µl of

1mg/ml of phosphatase substrate/dietanolamine (SIGMA) and 20µl cell media was then added to each well of a clear, 96-well plate and incubated in dark, at room temperature, until the mixture became yellow and the absorbance was measured at 405 nM in a Tristar LB941 Luminometer. Chloramphenicol acethyltransferase (CAT) enzyme-linked immunosorbent assay (ELISA) kit (Roche), was used on transfected cells according the protocol previously described [166]. The 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenthyletrazolium bromide (MTT) assay was performed according to the supplier's instructions (Sigma-Aldrich).

4.5 Drugs and drug administration

All drugs were dissolved in dimethyl sulfoxide (DMSO), and DMSO alone served as a control in all experiments. Drugs were administrated on sub confluent cells by replacing the old culture medium with medium containing the indicated drug, in indicated concentrations. A detailed account of all drugs, concentrations and incubation times are made under each papers Material and Methods.

4.6 RNA extractions, RT-PCR, qRT-PCR and 3´RACE

Cytoplasmic RNA was extracted as described previously (Tan and Schwartz, 1995). Total RNA was extracted either with TRI Reagent (SIGMA Aldrich Life Science) and Direct-zol RNA MiniPrep (ZYMO Research), or Qiagen RNeasy Mini Kit (Qiagen). All were used according to the manufacturer's protocols followed by 1h DNase1 treatment (Thermo Scientific or Sigma Aldrich) at 37ºC. Following phenol/chloroform extraction, the RNA was ethanol-precipitated and dissolved in 20µl of water. For RT-PCR and Real time quantitative PCR (qPCR), 500-300ng of RNA were reverse transcribed in a total volume of 20µl with M-MLV Reverse Transcriptase (Invitrogen) and random hexamers (Invitrogen) according to the protocol of the manufacturer. One microliter of cDNA was used for all PCR amplifications. qPCR was performed in a MiniOpticon (BioRad) with Sso Advanced SYBR Green Supermix (BioRad) according the manufacturers' instructions. All cDNA

quantitation's were normalized to GAPDH mRNA levels unless otherwise stated. In Paper II, 3´ -RACE was performed on 500ng TotalRNA that was reverse transcribed in a 20µl reaction at 37◦C using M-MLV Reverse Transcriptase (Invitrogen) and an oligo-dT primer, P3-17dT. The following PCR was then performed with the P3-17dT and indicated, mRNA-specific forward primers.

4.7 Protein extraction, immunoprecipitation and western blotting

Proteins for Western blotting and Immunoprecipitations were extracted from cells using the radioimmunoprecipitation assay (RIPA) buffer (50 mM Tris pH 7.4, 500 mM NaCl, 1% Na-DOC, 0,1% SDS, 1% Triton X) with 30min incubation on ice and occasional vortexing. The Co-Immunoprecipitation has previously been described (39) and 400ng of cell extract was incubated over-night with Dynabeads (Invitrogen) and 1–2µg of antibody, followed by washing three times with washing buffer. Western blotting was performed as described previously [113].

4.8 Chromatin immunoprecipitation (CHIP)

Chromatin immunoprecipitations (ChIPs) were performed using the SimpleChIPs Enzymatic Chromatin IP Kit (Cell Signaling) according to the manufacturer's instructions with the exception of optimizations previously described[114]. The subsequent qPCR was performed as described above and Ct values obtained from control IgG were subtracted from Ct values obtained from immunoprecipitations with specific antibodies. These values were then used to calculate the amount of each immunoprecipitation relative to the input with the following formula: Relative amount = $2^{-Ct(\text{specific antibody-IgG})}/2^{-Ct(\text{input-IgG})}$, for each PCR amplicon. For Melphalan treatments, the fold change compared to DMSO-treated samples was calculated according to Fold change = $(2^{-Ct(\text{specific antibody-IgG})}/2^{-Ct(\text{input-IgG})})_{drug}/(2^{-Ct(\text{specific antibody-IgG})}/2^{-Ct(\text{input-IgG})})_{DMSO}$. For ChIP analysis of histone acetylation, values were normalized to those obtained from unmodified histones.

4.9 DNA immunoprecipitation (DIP)

Chromosomal DNA was sonicated to obtain DNA fragments between 200- and 800-nt in size. 500ng of the sonicated DNA was then used for immunoprecipitation followed by DNA extraction and qPCR.

4.10 UV-crosslinking and immunoprecipitation (CLIP)

Cells were grown in 10cm dishes and transfected with Turbofect (Thermo Fisher Scientific) according to the manufactures' instructions. 24hrs post transfection, cells were washed with ice-cold phosphate-buffered saline and UV-irradiated at 150 mJ/cm^2 in a Bio-link cross-linker (Biometra). Cells were then lysed in 1ml of radioimmunoprecipitation assay (RIPA) buffer (described above) and incubated on ice for 30 min with occasional vortexing. 0.5ml of the cell lysate was incubated for 2h, at 4◦C, with either IgG or protein-specific antibody. 0.6mg Dynabeads Protein G (#10004D, Invitrogen) were blocked with 1% BSA in RIPA buffer for 0.5h and washed three times with RIPA buffer. The beads were then added to the antibody-cell lysate samples and incubated for 1h at 4◦C. Beads were washed three times with RIPA buffer on a magnetic stand and RNA was eluted by phenol/chloroform extraction and ethanol-precipitated. No RNA degradation was used. Instead, the eluted RNA was dissolved in 20µl water and 10µl was directly reverse transcribed (as described above) followed by PCR amplification.

4.11 *In vitro* RNA synthesis

In vitro RNA was synthesized using the TranscriptAid T7 High Yield Transcription Kit according to manufactures instruction using the protocol control DNA as a template (#K0441, Thermo Scientific). Three reactions were with either 0- , 1- or 10% of the ATP-pool substituted for the N6-methyladenosine base analogue m6A (S3190, Selleckchem). The invitro RNA was then used for immunoprecipitation according the method outlined below.

4.12 UV-crosslinking and immunoprecipitation with m6A-antibody

Total RNA was extracted from C33A2 and HN26 cells with TriReagent (Invitrogen), followed by DNase1 (Thermo Scientific) treatment for 1hr at 37°C according to manufacturer's protocols. RNA was then extracted with phenol/chloroform followed by ethanol precipitation and resuspension in 20 µl H2O. 20 µg RNA were diluted in 450µl immunoprecipitation (IP) buffer (50 mM Tris pH 7.4, 100 mM NaCl, 0.05% NP40), together with 5ug of either IgG or Anti-m6A antibody (Abcam rabbit polyclonal (ab151230)), and 5µl RI Ribolock (Thermo Scientific). Following over-night incubation at 4°C, the antibody-RNA samples were transferred into a single well each in a 12-well plate and crosslinked twice with 0.15 J cm−2 UV light (254 nm) in a Stratalinker (Agilent). 0.6mg of Dynabeads (Invitrogen) were added to the antibody-RNA mixture and incubated for 2h, at 4°C. Beads were then washed twice on a magnetic stand with high-salt buffer (50 mM Tris pH7.4, 1M NaCl, 1 mM EDTA, 1% NP40, 0.1% SDS), and 4 times with IP buffer. RNA was eluted by phenol/chloroform extraction and resuspended in 20µl water. 10µl were reverse transcribed using random primers (Invitrogen) and M-MLV reverse transcriptase for 50min at 37°C (Invitrogen). 1µl cDNA was subjected to RT-PCR. Quantification of immunoprecipitated RNA from DAA treated cells was performed in a MiniOpticon (Bio-Rad) using the Sso Advanced SYBR Green Supermix (Bio-Rad) with 1µl cDNA. CT values of the target gene from treated cells were normalized to the levels of the same mRNA in DMSO treated cells.

4.13 Nuclear extracts, DNA- and RNA-oligo pulldown

Nuclear extracts were prepared from sub confluent cells as previously described[170] . Biotinylated ssDNA oligos were purchased from Eurofin genomics and ssRNA oligos from either Eurofin genomics or SIGMA Aldrich. Specific cell type and sequences for each oligo are listed separately under material and methods in each paper. The pulldown assay has previously been described[125]. Briefly, one mg Streptavidin-coated magnetic beads (Dynabeads M-280 Streptavidin) from Thermo Fisher were first coated with

200 pmol of biotinylated ssDNA/ssRNA oligo. Beads were blocked with 1% BSA in IxWB buffer (5mM Hepes pH 7.9, 0,5mM EDTA, 1M NaCl) for 40 min. 60µg of nuclear extract were diluted in NE Buffer III (20mM Hepes pH 7.9, 3mM MgCl2 10% Glycerol) to reach a NaCl concentration of 150mM, and incubated with the magnetic beads for 40min. The protein coated oligos attached to the beads were washed 10 times with binding buffer (10mM HEPES pH7.9, 2.5mM MgCl2, 0.5% Triton-X,150mM NaCl, 1mM DTT) on a magnetic stand. Elution of proteins were done by boiling the beads in sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) loading buffer followed by SDS-PAGE. Antibodies used for western blot analysis are listed separately in each paper.

4.14 Array and bioinformatics

RNA was extracted from subconfluent C33A2 incubated with either Melphalan or DMSO for 24hrs with Qiagen RNeasy Mini Kit (Qiagen) according to suppliers' instructions. Five samples of each treatment were analyzed on Affymatrix GeneChip Human Transcriptome array 2.0 at SCIBLU Genomics (Lund University, Lund, Sweden). Protein coding genes were analyzed in respect to total mRNA levels and changes in splicing. Genes with a minimum of a 2-fold change in mRNA levels between melphalan and DMSO treated cells, were sorted in the Transcriptome Analysis Console (TAC) from Thermo Fisher. Genes with a minimum of a 2-fold change in the use of a splice junction or exon inclusion/exclusion event between melphalan and DMSO treatment were sorted similarly in TAC. Following sorting, the TAC-generated lists of genes were exported to PANTHER version 13.1, Gene List Analysis tool (Available online: http://pantherdb.org) for overrepresentation tests based on their biological function.

4.15 Statistics

Data in graphs are presented as mean values with standard deviation of the mean as error bars. Standard deviations and student T-tests were all conducted in Excel and P values < 0.05 were considered significant.

Acknowledgement

I am sincerely grateful towards all the people who helped and supported me during the completion of my thesis: colleges, friends and family.

First, and foremost I would like to thank my supervisor, professor Stefan Schwartz for all the invaluable knowledge you have given me during these years. But most of all, I would like to thank you for your fatherly patience. I have not only completed my PhD under your guidance, but also battled against my health and I am deeply grateful and well aware that none of it would have been possible without your kindness and flexibility.

To recent and former members of Schwartz group; thank you sincerely for all your advice, support and sharing of experiences. Thank you all for brightening my day, and all the delicious fika. A special thank you to Haoran who walked this road alongside me, gave me support and a shoulder to cry on when my health was collapsing. To Anki Mossberg for teaching me how to wheel my pipette towards success, you truly are a wizard of skills. To my co-supervisor associate professor Ola Forslund; thank you for all your help.

I would also like to thank Marianne Jansson and Patrik Medstrand for the knowledge you have passed on during our joint meetings and seminars, and for all the encouraging lunchroom chitty chats that have brighten my day.

Sist men inte minst vill jag tacka min familj för att ni alltid finns där, stöttandes, hejandes, styrkegivande och kärleksfulla. Detta hade aldrig gått utan er, ni är mitt allt.

References

1. zur Hausen, H., *Papillomaviruses and cancer: from basic studies to clinical application.* Nat Rev Cancer, 2002. **2**(5): p. 342-50.
2. Richardson, H., et al., *The natural history of type-specific human papillomavirus infections in female university students.* Cancer Epidemiol Biomarkers Prev, 2003. **12**(6): p. 485-90.
3. Van Doorslaer, K., et al., *The Papillomavirus Episteme: a major update to the papillomavirus sequence database.* Nucleic Acids Res, 2017. **45**(D1): p. D499-D506.
4. Stanley, M., *Immunobiology of HPV and HPV vaccines.* Gynecol Oncol, 2008. **109**(2 Suppl): p. S15-21.
5. de Villiers, E.M., et al., *Classification of papillomaviruses.* Virology, 2004. **324**(1): p. 17-27.
6. Bzhalava, D., et al., *A systematic review of the prevalence of mucosal and cutaneous human papillomavirus types.* Virology, 2013. **445**(1-2): p. 224-31.
7. Howley, P.M. and H.J. Pfister, *Beta genus papillomaviruses and skin cancer.* Virology, 2015. **479-480**: p. 290-6.
8. Borgogna, C., et al., *Characterization of beta papillomavirus E4 expression in tumours from Epidermodysplasia Verruciformis patients and in experimental models.* Virology, 2012. **423**(2): p. 195-204.
9. Lorincz, A., *Detection of human papillomavirus DNA without amplification: prospects for clinical utility.* IARC Sci Publ, 1992(119): p. 135-45.
10. Barr, E. and G. Tamms, *Quadrivalent human papillomavirus vaccine.* Clin Infect Dis, 2007. **45**(5): p. 609-7.
11. Arbyn, M., et al., *EUROGIN 2011 roadmap on prevention and treatment of HPV-related disease.* Int J Cancer, 2012. **131**(9): p. 1969-82.
12. Parkin, D.M. and F. Bray, *Chapter 2: The burden of HPV-related cancers.* Vaccine, 2006. **24 Suppl 3**: p. S3/11-25.
13. Li, N., et al., *Human papillomavirus type distribution in 30,848 invasive cervical cancers worldwide: Variation by geographical region, histological type and year of publication.* Int J Cancer, 2011. **128**(4): p. 927-35.

14. de Sanjose, S., et al., *Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study.* Lancet Oncol, 2010. **11**(11): p. 1048-56.
15. Gillison, M.L., et al., *Evidence for a causal association between human papillomavirus and a subset of head and neck cancers.* J Natl Cancer Inst, 2000. **92**(9): p. 709-20.
16. Doorbar, J., et al., *The biology and life-cycle of human papillomaviruses.* Vaccine, 2012. **30 Suppl 5**: p. F55-70.
17. Egawa, K., *Do human papillomaviruses target epidermal stem cells?* Dermatology, 2003. **207**(3): p. 251-4.
18. Schafer, G., M.J. Blumenthal, and A.A. Katz, *Interaction of human tumor viruses with host cell surface receptors and cell entry.* Viruses, 2015. **7**(5): p. 2592-617.
19. Pyeon, D., et al., *Establishment of human papillomavirus infection requires cell cycle progression.* PLoS Pathog, 2009. **5**(2): p. e1000318.
20. Day, P.M. and M. Schelhaas, *Concepts of papillomavirus entry into host cells.* Curr Opin Virol, 2014. **4**: p. 24-31.
21. Munger, K. and P.M. Howley, *Human papillomavirus immortalization and transformation functions.* Virus Res, 2002. **89**(2): p. 213-28.
22. Bodily, J. and L.A. Laimins, *Persistence of human papillomavirus infection: keys to malignant progression.* Trends Microbiol, 2011. **19**(1): p. 33-9.
23. McBride, A.A., *Mechanisms and strategies of papillomavirus replication.* Biol Chem, 2017. **398**(8): p. 919-927.
24. Zheng, Z.M. and C.C. Baker, *Papillomavirus genome structure, expression, and post-transcriptional regulation.* Front Biosci, 2006. **11**: p. 2286-302.
25. Favre, M., et al., *Chromatin-like structures obtained after alkaline disruption of bovine and human papillomaviruses.* J Virol, 1977. **21**(3): p. 1205-9.
26. Kammer, C., et al., *Sequence analysis of the long control region of human papillomavirus type 16 variants and functional consequences for P97 promoter activity.* J Gen Virol, 2000. **81**(Pt 8): p. 1975-81.
27. Doorbar, J., *The papillomavirus life cycle.* J Clin Virol, 2005. **32 Suppl 1**: p. S7-15.
28. Johansson, C. and S. Schwartz, *Regulation of human papillomavirus gene expression by splicing and polyadenylation.* Nat Rev Microbiol, 2013. **11**(4): p. 239-51.
29. Kajitani, N. and S. Schwartz, *RNA Binding Proteins that Control Human Papillomavirus Gene Expression.* Biomolecules, 2015. **5**(2): p. 758-74.

30. Moody, C.A. and L.A. Laimins, *Human papillomavirus oncoproteins: pathways to transformation.* Nat Rev Cancer, 2010. **10**(8): p. 550-60.
31. Martinez-Zapien, D., et al., *Structure of the E6/E6AP/p53 complex required for HPV-mediated degradation of p53.* Nature, 2016. **529**(7587): p. 541-5.
32. Crook, T., J.A. Tidy, and K.H. Vousden, *Degradation of p53 can be targeted by HPV E6 sequences distinct from those required for p53 binding and trans-activation.* Cell, 1991. **67**(3): p. 547-56.
33. Li, X. and P. Coffino, *High-risk human papillomavirus E6 protein has two distinct binding sites within p53, of which only one determines degradation.* J Virol, 1996. **70**(7): p. 4509-16.
34. Howie, H.L., R.A. Katzenellenbogen, and D.A. Galloway, *Papillomavirus E6 proteins.* Virology, 2009. **384**(2): p. 324-34.
35. Mantovani, F. and L. Banks, *Inhibition of E6 induced degradation of p53 is not sufficient for stabilization of p53 protein in cervical tumour derived cell lines.* Oncogene, 1999. **18**(22): p. 3309-15.
36. Patel, D., et al., *The E6 protein of human papillomavirus type 16 binds to and inhibits co-activation by CBP and p300.* EMBO J, 1999. **18**(18): p. 5061-72.
37. Zimmermann, H., et al., *Interaction with CBP/p300 enables the bovine papillomavirus type 1 E6 oncoprotein to downregulate CBP/p300-mediated transactivation by p53.* J Gen Virol, 2000. **81**(Pt 11): p. 2617-23.
38. Olmedo-Nieva, L., et al., *The Role of E6 Spliced Isoforms (E6*) in Human Papillomavirus-Induced Carcinogenesis.* Viruses, 2018. **10**(1).
39. Patrick, D.R., A. Oliff, and D.C. Heimbrook, *Identification of a novel retinoblastoma gene product binding site on human papillomavirus type 16 E7 protein.* J Biol Chem, 1994. **269**(9): p. 6842-50.
40. Johannsen, E. and P.F. Lambert, *Epigenetics of human papillomaviruses.* Virology, 2013. **445**(1-2): p. 205-12.
41. Kato, J., et al., *Direct binding of cyclin D to the retinoblastoma gene product (pRb) and pRb phosphorylation by the cyclin D-dependent kinase CDK4.* Genes Dev, 1993. **7**(3): p. 331-42.
42. Jones, R.E., et al., *Specific N-methylations of HPV-16 E7 peptides alter binding to the retinoblastoma suppressor protein.* J Biol Chem, 1992. **267**(2): p. 908-12.
43. Munger, K., et al., *Biological activities and molecular targets of the human papillomavirus E7 oncoprotein.* Oncogene, 2001. **20**(54): p. 7888-98.

44. Songock, W.K., S.M. Kim, and J.M. Bodily, *The human papillomavirus E7 oncoprotein as a regulator of transcription.* Virus Res, 2017. **231**: p. 56-75.
45. Moody, C.A. and L.A. Laimins, *Human papillomaviruses activate the ATM DNA damage pathway for viral genome amplification upon differentiation.* PLoS Pathog, 2009. **5**(10): p. e1000605.
46. Conrad, M., V.J. Bubb, and R. Schlegel, *The human papillomavirus type 6 and 16 E5 proteins are membrane-associated proteins which associate with the 16-kilodalton pore-forming protein.* J Virol, 1993. **67**(10): p. 6170-8.
47. Wetherill, L.F., et al., *High-risk human papillomavirus E5 oncoprotein displays channel-forming activity sensitive to small-molecule inhibitors.* J Virol, 2012. **86**(9): p. 5341-51.
48. Pedroza-Saavedra, A., et al., *The human papillomavirus type 16 E5 oncoprotein synergizes with EGF-receptor signaling to enhance cell cycle progression and the down-regulation of p27(Kip1).* Virology, 2010. **400**(1): p. 44-52.
49. Straight, S.W., B. Herman, and D.J. McCance, *The E5 oncoprotein of human papillomavirus type 16 inhibits the acidification of endosomes in human keratinocytes.* J Virol, 1995. **69**(5): p. 3185-92.
50. Rosenberger, S., et al., *Alternative splicing of human papillomavirus type-16 E6/E6* early mRNA is coupled to EGF signaling via Erk1/2 activation.* Proc Natl Acad Sci U S A, 2010. **107**(15): p. 7006-11.
51. Auvinen, E., A. Alonso, and P. Auvinen, *Human papillomavirus type 16 E5 protein colocalizes with the antiapoptotic Bcl-2 protein.* Arch Virol, 2004. **149**(9): p. 1745-59.
52. Kabsch, K., et al., *The HPV-16 E5 protein inhibits TRAIL- and FasL-mediated apoptosis in human keratinocyte raft cultures.* Intervirology, 2004. **47**(1): p. 48-56.
53. Marchetti, B., et al., *The bovine papillomavirus oncoprotein E5 retains MHC class I molecules in the Golgi apparatus and prevents their transport to the cell surface.* Oncogene, 2002. **21**(51): p. 7808-16.
54. Zhang, B., et al., *The E5 protein of human papillomavirus type 16 perturbs MHC class II antigen maturation in human foreskin keratinocytes treated with interferon-gamma.* Virology, 2003. **310**(1): p. 100-8.
55. Ashrafi, G.H., et al., *Down-regulation of MHC class I is a property common to papillomavirus E5 proteins.* Virus Res, 2006. **120**(1-2): p. 208-11.
56. Hughes, F.J. and M.A. Romanos, *E1 protein of human papillomavirus is a DNA helicase/ATPase.* Nucleic Acids Res, 1993. **21**(25): p. 5817-23.

57. Sedman, J. and A. Stenlund, *The papillomavirus E1 protein forms a DNA-dependent hexameric complex with ATPase and DNA helicase activities.* J Virol, 1998. **72**(8): p. 6893-7.
58. Sun, S., et al., *Identification of a 68-kilodalton nuclear ATP-binding phosphoprotein encoded by bovine papillomavirus type 1.* J Virol, 1990. **64**(10): p. 5093-105.
59. Masterson, P.J., et al., *A C-terminal helicase domain of the human papillomavirus E1 protein binds E2 and the DNA polymerase alpha-primase p68 subunit.* J Virol, 1998. **72**(9): p. 7407-19.
60. McBride, A.A., *The papillomavirus E2 proteins.* Virology, 2013. **445**(1-2): p. 57-79.
61. Sanders, C.M. and A. Stenlund, *Recruitment and loading of the E1 initiator protein: an ATP-dependent process catalysed by a transcription factor.* EMBO J, 1998. **17**(23): p. 7044-55.
62. Hegde, R.S., *The papillomavirus E2 proteins: structure, function, and biology.* Annu Rev Biophys Biomol Struct, 2002. **31**: p. 343-60.
63. Thain, A., et al., *CpG methylation directly inhibits binding of the human papillomavirus type 16 E2 protein to specific DNA sequences.* J Virol, 1996. **70**(10): p. 7233-5.
64. McBride, A.A. and M.K. Jang, *Current understanding of the role of the Brd4 protein in the papillomavirus lifecycle.* Viruses, 2013. **5**(6): p. 1374-94.
65. Bernard, H.U., *Gene expression of genital human papillomaviruses and considerations on potential antiviral approaches.* Antivir Ther, 2002. **7**(4): p. 219-37.
66. Smith, J.A., et al., *SMCX and components of the TIP60 complex contribute to E2 regulation of the HPV E6/E7 promoter.* Virology, 2014. **468-470**: p. 311-321.
67. Davy, C.E., et al., *Human papillomavirus type 16 E1 E4-induced G2 arrest is associated with cytoplasmic retention of active Cdk1/cyclin B1 complexes.* J Virol, 2005. **79**(7): p. 3998-4011.
68. Davy, C.E., et al., *Identification of a G(2) arrest domain in the E1 wedge E4 protein of human papillomavirus type 16.* J Virol, 2002. **76**(19): p. 9806-18.
69. Wilson, R., et al., *The full-length E1E4 protein of human papillomavirus type 18 modulates differentiation-dependent viral DNA amplification and late gene expression.* Virology, 2007. **362**(2): p. 453-60.
70. Doorbar, J., *The E4 protein; structure, function and patterns of expression.* Virology, 2013. **445**(1-2): p. 80-98.

71. Buck, C.B. and B.L. Trus, *The papillomavirus virion: a machine built to hide molecular Achilles' heels.* Adv Exp Med Biol, 2012. **726**: p. 403-22.
72. Johnson, K.M., et al., *Role of heparan sulfate in attachment to and infection of the murine female genital tract by human papillomavirus.* J Virol, 2009. **83**(5): p. 2067-74.
73. Black, D.L., *Mechanisms of alternative pre-messenger RNA splicing.* Annu Rev Biochem, 2003. **72**: p. 291-336.
74. Modrek, B. and C. Lee, *A genomic view of alternative splicing.* Nat Genet, 2002. **30**(1): p. 13-9.
75. Moore, M.J. and P.A. Sharp, *Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing.* Nature, 1993. **365**(6444): p. 364-8.
76. Das, R., Z. Zhou, and R. Reed, *Functional association of U2 snRNP with the ATP-independent spliceosomal complex E.* Mol Cell, 2000. **5**(5): p. 779-87.
77. Wang, Z., et al., *General and specific functions of exonic splicing silencers in splicing control.* Mol Cell, 2006. **23**(1): p. 61-70.
78. Shen, H., J.L. Kan, and M.R. Green, *Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly.* Mol Cell, 2004. **13**(3): p. 367-76.
79. Busch, A. and K.J. Hertel, *Evolution of SR protein and hnRNP splicing regulatory factors.* Wiley Interdiscip Rev RNA, 2012. **3**(1): p. 1-12.
80. Hertel, K.J. and T. Maniatis, *Serine-arginine (SR)-rich splicing factors have an exon-independent function in pre-mRNA splicing.* Proc Natl Acad Sci U S A, 1999. **96**(6): p. 2651-5.
81. Graveley, B.R., K.J. Hertel, and T. Maniatis, *SR proteins are 'locators' of the RNA splicing machinery.* Curr Biol, 1999. **9**(1): p. R6-7.
82. Dreyfuss, G., et al., *hnRNP proteins and the biogenesis of mRNA.* Annu Rev Biochem, 1993. **62**: p. 289-321.
83. Martinez-Contreras, R., et al., *hnRNP proteins and splicing control.* Adv Exp Med Biol, 2007. **623**: p. 123-47.
84. Graveley, B.R., *Sorting out the complexity of SR protein functions.* RNA, 2000. **6**(9): p. 1197-211.
85. Shi, Y. and J.L. Manley, *The end of the message: multiple protein-RNA interactions define the mRNA polyadenylation site.* Genes Dev, 2015. **29**(9): p. 889-97.
86. Bienroth, S., W. Keller, and E. Wahle, *Assembly of a processive messenger RNA polyadenylation complex.* EMBO J, 1993. **12**(2): p. 585-94.
87. Hirose, Y. and J.L. Manley, *RNA polymerase II and the integration of nuclear events.* Genes Dev, 2000. **14**(12): p. 1415-29.

88. Beaudoing, E., et al., *Patterns of variant polyadenylation signal usage in human genes.* Genome Res, 2000. **10**(7): p. 1001-10.
89. Di Giammartino, D.C., K. Nishida, and J.L. Manley, *Mechanisms and consequences of alternative polyadenylation.* Mol Cell, 2011. **43**(6): p. 853-66.
90. Elkon, R., A.P. Ugalde, and R. Agami, *Alternative cleavage and polyadenylation: extent, regulation and function.* Nat Rev Genet, 2013. **14**(7): p. 496-506.
91. Jackson, R.J., C.U. Hellen, and T.V. Pestova, *The mechanism of eukaryotic translation initiation and principles of its regulation.* Nat Rev Mol Cell Biol, 2010. **11**(2): p. 113-27.
92. Kozak, M., *Circumstances and mechanisms of inhibition of translation by secondary structure in eucaryotic mRNAs.* Mol Cell Biol, 1989. **9**(11): p. 5134-42.
93. Hinnebusch, A.G., *The scanning mechanism of eukaryotic translation initiation.* Annu Rev Biochem, 2014. **83**: p. 779-812.
94. de la Parra, C., et al., *A widespread alternate form of cap-dependent mRNA translation initiation.* Nat Commun, 2018. **9**(1): p. 3068.
95. Rottman, F.M., R.C. Desrosiers, and K. Friderici, *Nucleotide methylation patterns in eukaryotic mRNA.* Prog Nucleic Acid Res Mol Biol, 1976. **19**: p. 21-38.
96. Ye, F., E.R. Chen, and T.W. Nilsen, *Kaposi's Sarcoma-Associated Herpesvirus Utilizes and Manipulates RNA N(6)-Adenosine Methylation To Promote Lytic Replication.* J Virol, 2017. **91**(16).
97. Krug, R.M., M.A. Morgan, and A.J. Shatkin, *Influenza viral mRNA contains internal N6-methyladenosine and 5'-terminal 7-methylguanosine in cap structures.* J Virol, 1976. **20**(1): p. 45-53.
98. Tirumuru, N., et al., *N(6)-methyladenosine of HIV-1 RNA regulates viral infection and HIV-1 Gag protein expression.* Elife, 2016. **5**.
99. Lichinchi, G., et al., *Dynamics of Human and Viral RNA Methylation during Zika Virus Infection.* Cell Host Microbe, 2016. **20**(5): p. 666-673.
100. Fu, Y., et al., *Gene expression regulation mediated through reversible m(6)A RNA methylation.* Nat Rev Genet, 2014. **15**(5): p. 293-306.
101. Wang, Y., et al., *N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells.* Nat Cell Biol, 2014. **16**(2): p. 191-8.
102. Zheng, G., et al., *ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility.* Mol Cell, 2013. **49**(1): p. 18-29.
103. Jia, G., et al., *N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO.* Nat Chem Biol, 2011. **7**(12): p. 885-7.

104. Liu, J., et al., *A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation.* Nat Chem Biol, 2014. **10**(2): p. 93-5.
105. Xiao, W., et al., *Nuclear m(6)A Reader YTHDC1 Regulates mRNA Splicing.* Mol Cell, 2016. **61**(4): p. 507-519.
106. Wang, X., et al., *N6-methyladenosine-dependent regulation of messenger RNA stability.* Nature, 2014. **505**(7481): p. 117-20.
107. Shi, H., et al., *YTHDF3 facilitates translation and decay of N(6)-methyladenosine-modified RNA.* Cell Res, 2017. **27**(3): p. 315-328.
108. Meyer, K.D., et al., *5′ UTR m(6)A Promotes Cap-Independent Translation.* Cell, 2015. **163**(4): p. 999-1010.
109. Liu, N., et al., *N6-methyladenosine alters RNA structure to regulate binding of a low-complexity protein.* Nucleic Acids Res, 2017. **45**(10): p. 6051-6063.
110. Liu, N., et al., *N(6)-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions.* Nature, 2015. **518**(7540): p. 560-4.
111. Graham, S.V. and A.A.A. Faizo, *Control of human papillomavirus gene expression by alternative splicing.* Virus Res, 2017. **231**: p. 83-95.
112. Schmitt, M., et al., *Diagnosing cervical cancer and high-grade precursors by HPV16 transcription patterns.* Cancer Res, 2010. **70**(1): p. 249-56.
113. Johansson, C., et al., *HPV-16 E2 contributes to induction of HPV-16 late gene expression by inhibiting early polyadenylation.* EMBO J, 2012. **31**(14): p. 3212-27.
114. Johansson, C., et al., *Acetylation of intragenic histones on HPV16 correlates with enhanced HPV16 gene expression.* Virology, 2015. **482**: p. 244-59.
115. Soto, D., C. Song, and M.E. McLaughlin-Drubin, *Epigenetic Alterations in Human Papillomavirus-Associated Cancers.* Viruses, 2017. **9**(9).
116. Rush, M., X. Zhao, and S. Schwartz, *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**(18): p. 12002-15.
117. Grassmann, K., et al., *Identification of a differentiation-inducible promoter in the E7 open reading frame of human papillomavirus type 16 (HPV-16) in raft cultures of a new cell line containing high copy numbers of episomal HPV-16 DNA.* J Virol, 1996. **70**(4): p. 2339-49.
118. Zhao, X., et al., *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**(7): p. 4270-88.
119. Oberg, D., et al., *A downstream polyadenylation element in human papillomavirus type 16 L2 encodes multiple GGG motifs and interacts with hnRNP H.* J Virol, 2005. **79**(14): p. 9254-69.
120. Graham, S.V., *Papillomavirus 3′ UTR regulatory elements.* Front Biosci, 2008. **13**: p. 5646-63.
121. Yu, H., et al., *hnRNP G prevents inclusion on the HPV16 L1 mRNAs of the central exon between splice sites SA3358 and SD3632.* J Gen Virol, 2018.
122. Somberg, M. and S. Schwartz, *Multiple ASF/SF2 sites in the human papillomavirus type 16 (HPV-16) E4-coding region promote splicing to the most commonly used 3′-splice site on the HPV-16 genome.* J Virol, 2010. **84**(16): p. 8219-30.
123. Somberg, M., et al., *Serine/arginine-rich protein 30c activates human papillomavirus type 16 L1 mRNA expression via a bimodal mechanism.* J Gen Virol, 2011. **92**(Pt 10): p. 2411-21.
124. Jia, R., et al., *Control of the papillomavirus early-to-late switch by differentially expressed SRp20.* J Virol, 2009. **83**(1): p. 167-80.
125. Dhanjal, S., et al., *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**(21): p. 13354-71.
126. Zhao, X., M. Rush, and S. Schwartz, *Identification of an hnRNP A1-dependent splicing silencer in the human papillomavirus type 16 L1 coding region that prevents premature expression of the late L1 gene.* J Virol, 2004. **78**(20): p. 10888-905.
127. Zhou, B.B. and S.J. Elledge, *The DNA damage response: putting checkpoints in perspective.* Nature, 2000. **408**(6811): p. 433-9.
128. Marechal, A. and L. Zou, *DNA damage sensing by the ATM and ATR kinases.* Cold Spring Harb Perspect Biol, 2013. **5**(9).
129. Reinhardt, H.C., et al., *p53-deficient cells rely on ATM- and ATR-mediated checkpoint signaling through the p38MAPK/MK2 pathway for survival after DNA damage.* Cancer Cell, 2007. **11**(2): p. 175-89.
130. Lavin, M.F., *Ataxia-telangiectasia: from a rare disorder to a paradigm for cell signalling and cancer.* Nat Rev Mol Cell Biol, 2008. **9**(10): p. 759-69.
131. Jazayeri, A., et al., *ATM- and cell cycle-dependent regulation of ATR in response to DNA double-strand breaks.* Nat Cell Biol, 2006. **8**(1): p. 37-45.
132. Branzei, D. and M. Foiani, *Regulation of DNA repair throughout the cell cycle.* Nat Rev Mol Cell Biol, 2008. **9**(4): p. 297-308.

133. Matsuoka, S., et al., *ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage.* Science, 2007. **316**(5828): p. 1160-6.
134. Bergvall, M., T. Melendy, and J. Archambault, *The E1 proteins.* Virology, 2013. **445**(1-2): p. 35-56.
135. Anacker, D.C., et al., *Productive replication of human papillomavirus 31 requires DNA repair factor Nbs1.* J Virol, 2014. **88**(15): p. 8528-44.
136. Moody, C., *Mechanisms by which HPV Induces a Replication Competent Environment in Differentiating Keratinocytes.* Viruses, 2017. **9**(9).
137. Fradet-Turcotte, A., et al., *Nuclear accumulation of the papillomavirus E1 helicase blocks S-phase progression and triggers an ATM-dependent DNA damage response.* J Virol, 2011. **85**(17): p. 8996-9012.
138. Hong, S., et al., *STAT-5 Regulates Transcription of the Topoisomerase IIbeta-Binding Protein 1 (TopBP1) Gene To Activate the ATR Pathway and Promote Human Papillomavirus Replication.* MBio, 2015. **6**(6): p. e02006-15.
139. Johnson, B.A., H.L. Aloor, and C.A. Moody, *The Rb binding domain of HPV31 E7 is required to maintain high levels of DNA repair factors in infected cells.* Virology, 2017. **500**: p. 22-34.
140. Hong, S., A. Dutta, and L.A. Laimins, *The acetyltransferase Tip60 is a critical regulator of the differentiation-dependent amplification of human papillomaviruses.* J Virol, 2015. **89**(8): p. 4668-75.
141. Hong, S.Y., *DNA damage response is hijacked by human papillomaviruses to complete their life cycle.* J Zhejiang Univ Sci B, 2017. **18**(3): p. 215-232.
142. Anacker, D.C. and C.A. Moody, *Modulation of the DNA damage response during the life cycle of human papillomaviruses.* Virus Res, 2017. **231**: p. 41-49.
143. Dubois, J., O. Terrier, and M. Rosa-Calatrava, *Influenza viruses and mRNA splicing: doing more with less.* MBio, 2014. **5**(3): p. e00070-14.
144. Backstrom Winquist, E., et al., *Inefficient splicing of segment 7 and 8 mRNAs is an inherent property of influenza virus A/Brevig Mission/1918/1 (H1N1) that causes elevated expression of NS1 protein.* Virology, 2012. **422**(1): p. 46-58.
145. Cox, N.J. and K. Subbarao, *Global epidemiology of influenza: past and present.* Annu Rev Med, 2000. **51**: p. 407-21.

146. Tumpey, T.M., et al., *Characterization of the reconstructed 1918 Spanish influenza pandemic virus.* Science, 2005. **310**(5745): p. 77-80.
147. Basler, C.F. and P.V. Aguilar, *Progress in identifying virulence determinants of the 1918 H1N1 and the Southeast Asian H5N1 influenza A viruses.* Antiviral Res, 2008. **79**(3): p. 166-78.
148. Kochs, G., A. Garcia-Sastre, and L. Martinez-Sobrido, *Multiple anti-interferon actions of the influenza A virus NS1 protein.* J Virol, 2007. **81**(13): p. 7011-21.
149. Bouvier, N.M. and P. Palese, *The biology of influenza viruses.* Vaccine, 2008. **26 Suppl 4**: p. D49-53.
150. Josset, L., E. Frobert, and M. Rosa-Calatrava, *Influenza A replication and host nuclear compartments: many changes and many questions.* J Clin Virol, 2008. **43**(4): p. 381-90.
151. Hutchinson, E.C. and E. Fodor, *Transport of the influenza virus genome from nucleus to nucleus.* Viruses, 2013. **5**(10): p. 2424-46.
152. Resa-Infante, P., et al., *The influenza virus RNA synthesis machine: advances in its structure and function.* RNA Biol, 2011. **8**(2): p. 207-15.
153. Lamb, R.A. and P.W. Choppin, *Segment 8 of the influenza virus genome is unique in coding for two polypeptides.* Proc Natl Acad Sci U S A, 1979. **76**(10): p. 4908-12.
154. Lamb, R.A. and C.J. Lai, *Spliced and unspliced messenger RNAs synthesized from cloned influenza virus M DNA in an SV40 vector: expression of the influenza virus membrane protein (M1).* Virology, 1982. **123**(2): p. 237-56.
155. Robb, N.C. and E. Fodor, *The accumulation of influenza A virus segment 7 spliced mRNAs is regulated by the NS1 protein.* J Gen Virol, 2012. **93**(Pt 1): p. 113-8.
156. Lu, Y., X.Y. Qian, and R.M. Krug, *The influenza virus NS1 protein: a novel inhibitor of pre-mRNA splicing.* Genes Dev, 1994. **8**(15): p. 1817-28.
157. Robb, N.C., et al., *Splicing of influenza A virus NS1 mRNA is independent of the viral NS1 protein.* J Gen Virol, 2010. **91**(Pt 9): p. 2331-40.
158. Salvatore, M., et al., *Effects of influenza A virus NS1 protein on protein expression: the NS1 protein enhances translation and is not required for shutoff of host protein synthesis.* J Virol, 2002. **76**(3): p. 1206-12.
159. Moss, W.N., S.F. Priore, and D.H. Turner, *Identification of potential conserved RNA secondary structure throughout influenza A coding regions.* RNA, 2011. **17**(6): p. 991-1011.

160. Moss, W.N., et al., *The influenza A segment 7 mRNA 3' splice site pseudoknot/hairpin family.* RNA Biol, 2012. **9**(11): p. 1305-10.
161. Priore, S.F., W.N. Moss, and D.H. Turner, *Influenza A virus coding regions exhibit host-specific global ordered RNA structure.* PLoS One, 2012. **7**(4): p. e35989.
162. Moss, W.N., et al., *The 3' splice site of influenza A segment 7 mRNA can exist in two conformations: a pseudoknot and a hairpin.* PLoS One, 2012. **7**(6): p. e38323.
163. Selman, M., et al., *Adaptive mutation in influenza A virus non-structural gene is linked to host switching and induces a novel protein by alternative splicing.* Emerg Microbes Infect, 2012. **1**(11): p. e42.
164. Alarcon, C.R., et al., *HNRNPA2B1 Is a Mediator of m(6)A-Dependent Nuclear RNA Processing Events.* Cell, 2015. **162**(6): p. 1299-308.
165. Collier, B., et al., *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**(6): p. 2739-52.
166. Li, X., et al., *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**(9): p. e72776.
167. Li, X., et al., *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**(22): p. 10488-508.
168. Markova, S.V., et al., *Cloning and expression of cDNA for a luciferase from the marine copepod Metridia longa. A novel secreted bioluminescent reporter enzyme.* J Biol Chem, 2004. **279**(5): p. 3212-7.
169. Chiu, J., et al., *Site-directed, Ligase-Independent Mutagenesis (SLIM): a single-tube methodology approaching 100% efficiency in 4 h.* Nucleic Acids Res, 2004. **32**(21): p. e174.
170. Dignam, J.D., R.M. Lebovitz, and R.G. Roeder, *Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei.* Nucleic Acids Res, 1983. **11**(5): p. 1475-89.