Digital Comprehensive Summaries of Uppsala Dissertations
from the Faculty of Medicine 1967

The balance of splicing

A novel insight into the splicing regulation of high-risk HPV E6 and E7 oncogenes.

JOHANNA JÖNSSON

ACTA UNIVERSITATIS
UPSALIENSIS
2023

ISSN 1651-6206
ISBN 978-91-513-1867-7
urn:nbn:se:uu:diva-508478

UPPSALA
UNIVERSITET

Dissertation presented at Uppsala University to be publicly examined in A1:111, Biomedical Science Center, Husargatan 3, 751 23, Uppsala, Thursday, 28 September 2023 at 09:00 for the degree of Doctor of Philosophy (Faculty of Medicine). The examination will be conducted in English. Faculty examiner: Docent Francesca Aguilo (Umeå University, Wallenberg Centre for Molecular Medicine (WCMM)).

Abstract
Jönsson, J. 2023. The balance of splicing. A novel insight into the splicing regulation of high-risk HPV E6 and E7 oncogenes.. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine* 1967. 119 pp. Uppsala: Acta Universitatis Upsaliensis.
ISBN 978-91-513-1867-7.

HPV is associated with several cancers. The genome consists of a long control region, early (E1, E2, E4, E5, E6 and E7) and late (L1 and L2) genes. The E6 and E7 proteins prevent cells from entering apoptosis and regulate the cell cycle. A deregulated expression of these can result in malignant transformations. Therefore, a deeper understanding of their regulation is important. HPV gene expression regulation occurs mainly through alternative pre-mRNA splicing with many splice events being mutually exclusive. This is the case with E6 and E7. E6 is expressed from the intron-retained mRNA while E7 is expressed from a spliced mRNA. In this thesis it was aimed to understand the alternative splice events of HPV16 E6 and E7 mRNAs by identifying regulatory elements controlling these splice events. A strong enhancer, downstream of SA409, was identified. It consists of a perfect bipartite repeat and mutations in the element disrupts SA409 splicing. Trans-acting factors were determined to TRAP150 and BCLAF1 (Paper I). Downstream of this is another cis-element. It consists of a GGGG-motif with a silencing effect on SA409 splicing. Two additional cis-elements, one at the end of the E6 ORF and the other at the start of the E7 ORF, were additionally found. These are suggested to hybridize forming an internal-loop structure when analyzed in silico. The cis-element at the end of the E6 ORF is context dependent, functioning as an enhancer or silencer depending on if the E7 cis-element is included or not. It was identified as ATCATCA (Paper III). The cis-element at the start of the E7 ORF was a silencer and consisted of an AUG-rich element. Two trans-acting factors interacted with it, hnRNP A1 and hnRNP A2, and prevented SA409 splicing. However, hnRNP A1 increased the intron-retained E6 mRNA, while hnRNP A2 redirected splicing to a downstream acceptor, SA742 (Paper IV). SA742 was additionally found to be regulated by another cis-element, within the E1 ORF. It consisted of three GGG/GGGG motifs and the integrity of these and of SD880 were important for SA742 usage, indicating the importance of regulatory factors in the modulation of the splicing modes: intron definition and exon definition. The trans-acting factor was hnRNP H (Paper II). In this thesis five cis-elements and five trans-acting factors affecting splicing regulation within the E6 and E7 ORFs were identified. With this knowledge to build upon several important targets to change or disrupt splicing were identified.

Keywords: HPV, splice donor, splice acceptor, cis-element, trans-acting factor, pre-mRNA splicing, splicing regulation.

Johanna Jönsson, Department of Medical Biochemistry and Microbiology, Box 582, Uppsala University, SE-75123 Uppsala, Sweden.

ISSN 1651-6206
ISBN 978-91-513-1867-7
URN urn:nbn:se:uu:diva-508478 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-508478)

To my family.

List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I. Johanna Jönsson, Lianqing Wang, Naoko Kajitani, Stefan Schwartz (2023). ***A novel HPV16 splicing enhancer critical for viral oncogene expression and cell immortalization***. (Submitted to Nucleic Acids Research and under revision)

II. Johanna Jönsson, Qiaoli Zhai, Naoko Kajitani, Stefan Schwartz (2023) ***hnRNP H controls alternative splicing of human papillomavirus type 16 E1, E6, E7 and E6^E7 mRNAs via GGG motifs.*** (Submitted to Journal of Virology)

III. Johanna Jönsson, Lianqing Wang, Naoko Kajitani, Stefan Schwartz (2023). ***A Janus-faced RNA element controls splicing of the HPV16 E6 and E7 oncogene mRNAs***. (Manuscript).

IV. Yunji Zheng, Johanna Jönsson, Chengyu Hao, Shirin Shoja Chaghervand, Xiaoxu Cui, Naoko Kajitani, Lijing Gong, Chengjun Wu, Stefan Schwartz (2020). ***Heterogeneous Nuclear Ribonucleoprotein A1 (hnRNP A1) and hnRNP A2 Inhibit Splicing to Human Papillomavirus 16 Splice Site SA409 through a UAG-Containing Sequence in the E7 Coding Region***. Journal of Virology. Sep 29;94(20):e01509-20. doi: 10.1128/JVI.01509-20.

Articles not included in the thesis

I. Hao C, Gong L, Cui X, Jönsson J, Zheng Y, Wu C, Kajitani N, Schwartz S. ***Identification of heterogenous nuclear ribonucleoproteins (hnRNPs) and serine- and arginine-rich (SR) proteins that induce human papillomavirus type 16 late gene expression and alter L1 mRNA splicing.*** Arch Virol. 2022 Feb;167(2):563-570. doi: 10.1007/s00705-021-05317-2. Epub 2021 Dec 3. PMID: 34860285; PMCID: PMC8843915.

II. Hao C, Zheng Y, Jönsson J, Cui X, Yu H, Wu C, Kajitani N, Schwartz S. ***hnRNP G/RBMX enhances HPV16 E2 mRNA splicing through a novel splicing enhancer and inhibits production of spliced E7 oncogene mRNAs.*** Nucleic Acids Res. 2022 Apr 22;50(7):3867-3891. doi: 10.1093/nar/gkac213. PMID: 35357488; PMCID: PMC9023273.

Contents

Abbreviations

AdE1A	Adenovirus E1A
ASOs	Antisense oligonucleotides
ATM	Ataxia-telangiectasia mutated
ATR	ATM and Rad3-related
BPS	Branch point sequence
Btf	Bcl-2 associated transcription factor
CBC	Cap binding complex
cDNA	Complementary DNA
CDS	Coding sequence
CFI and CFII	Cleavage factors I and II
CIN	Cervical intraepithelial neoplasia
CMV	Cytomegalovirus
CPSF	Cleavage and polyadenylation factor
CstF	Cleavage and stimulation factor
DBD	DNA binding domain
DDR	DNA damage response
DSB	Double strand break
E6AP	E6 associated protein
EJC	Exon junction complex
ESE	Exonic splicing enhancer
ESS	Exonic splicing silencer
FBS	Fetal bovine serum
HDACs	Histone deacetylases
hnRNPs	Heterogenous ribonucleoprotein family
HSPG	Heparin sulfate proteoglycans
HPV	Human papillomavirus
IARC	International Association for Research on Cancer
INFs	Interferons
ISE	Intronic splicing enhancer
ISS	Intronic splicing silencer
IRES	Internal ribosome entry site
KSHV	Kaposi's sarcoma-associated herpesvirus
LCR	Long control region
NLS	Nuclear localization signal
NCR	Non coding region
nHFK	Normal Human Foreskin Keratinocytes

ORF	Open reading frame
Ori	Origin of replication
pAE	Early polyadenylation signal
pAL	Late polyadenylation signal
PCR	Polymerase chain reaction
PID	PSF-interacting domain
pRb	Retinoblastoma protein
PAP	Polyadenylation polymerase
PV	Papillomavirus
PPT	Polypyrimidine tract
qRRM	Quasi-RRM
RIPA	Radioimmunoprecipitation assay
RBP	RNA binding protein
RRM	RNA recognition motif
sLuc	Secreted luciferase
snRNPs	Small nuclear ribonucleoproteins
SR-proteins	Serine/arginine-rich protein family
SRSFs	Serine/arginine-rich splicing factors
SV40 T	Simian vacuolating virus 40 large tumor antigen
THRAP3	Thyroid Hormone Receptor Associated Protein 3
TRAP150	Thyroid hormone receptor-associated protein complex 150kDa component
VLPs	Virus-like particles
WST	Water soluble tetrasodium
UUR	Upstream regulatory region

Populärvetenskaplig sammanfattning på svenska.

Virus är idag ett begrepp som de flest känner till, men vad ett virus är och vad det gör är det desto färre som vet. Något som alltid har varit fascinerande för mig är att virus inte klassas som levande, eftersom de inte kan föröka sig själva utan behöver en värdcell till detta. En virusfamilj som idag består av hundratals olika medlemmar är humant papillomvirus (HPV) och det är denna familj som undersökts här. I HPV familjen ingår det runt 400 medlemmar men 20 av dessa sticker ut. Detta eftersom en långvarig infektion med någon av dessa 20 typer kan resultera i cancer. Flera olika typer av cancer har länkats till HPV men vad alla dessa har gemensamt är att de sker på ställen med slemhinnor, så som halsen och underlivet. Idag är en HPV infektion stark associerat med livmoderhalscancer och årligen insjuknar runt 604 000 kvinnor världen över. Men trots detta är det bara i ytterst få fall som en HPV infektion faktiskt resulterar i cancer. I normala fall läker infektionen ut av sig själv inom ett år efter att man smittats.

Vid en virusinfektion behöver viruset ta sig in i cellen och väl här börjar viruset att uttrycka olika gener. Detta för att kapa cellens egna system till sin egen vinning. I normal fall ringer det ett alarmsystem i cellen vid denna tidpunkt men HPV viruset går in och stänger av detta. Denna avstängning görs utav två olika gener som kallas E6 och E7. I normal fall regleras E6 och E7 generna så att de hålls i balans på en stadig nivå men i sällsynta fall rubbas balansen. När balansen rubbas sker det förändringar inom cellen och om dessa inte stoppas kan cancer bildas.

I denna avhandling har uttrycket av E6 och E7 undersökts. Vad man sedan tidigare visste är att E6 och E7 uttrycks från samma ”för-läsram”, men för att E7 skall kunna uttryckas behöver den klippas och klistras ihop på ett annat sätt än ursprungsformen. Om läsramen förblir i sin första form kommer E6 att uttryckas men om den klipps kommer istället E7 att uttryckas. Vad som också är viktigt att veta är att E6 och E7 inte kan uttryckas samtidigt från samma ”för-läsram”, utan antingen uttrycks E6 eller så klipps den och E7 uttrycks. Det är detta event som undersökts här. Dvs. vad är det som bestämmer att ”för-läsramen” skall klippas eller inte och hur regleras detta.

Här har vi hittat fem olika delar inom ”för-läsramen” samt fem yttre faktorer, inom cellen, som påverkar hur klippningen och klistringen sker. Intressant är också att vissa av dessa har en stimulerande effekt medan vissa har en

inhibitorisk effekt och det är summan av alla dessa som resulterar i slutresultatet. Dvs. klippning eller ingen klippning, E6 eller E7 uttryck.

Identifieringen av dessa delar och faktorer är viktiga. Detta eftersom de ger oss en djupare förståelse för hur E6 och E7 uttrycks och regleras i cellen. Genom att förstå detta kan också svaga punkter i denna reglering identifieras och dessa kan användas för att konstruera läkemedel som ändrar uttrycket eller snarare balansen av E6 och E7 i cellen. Genom att rubba balansen i cellerna kan man också ändra cellernas tillstånd och på så sätt kanske hitta en bättre behandling för HPV relaterad cancer.

Introduction

Human papillomaviruses: classification and disease.

Papillomaviruses (PVs) are a large group of viruses constituting the *Papillomaviridae* family. In this family many different host species are found. One species infected by PVs is us humans and the papillomaviruses infecting us are called human papillomaviruses (HPVs) [1]. Within this group over 400 different types are found, out of which 220 have been completely characterized [2]. The HPV genome is a DNA genome which is double stranded and circular and its size is around 8kb. The virus is not enveloped and has an icosahedral capsid of around 55nm is diameter [3]. However, despite the small genome size the virus should not be underestimated since it has a very complex molecular biology [1].

The HPVs can be classified and grouped in several ways and one these is how much they differ in their sequence. This is done by looking at the most conserved gene of the genome, the L1 gene. The sequence similarity is also used to identify new HPV types. If part of the L1 gene differ by 10% or more from the other identified types this is considered to be a new HPV type. However, if the difference is less than 10% but more than 2% this is a subtype and sequences with less than 2% difference is considered variants. It has also been stated that HPV types with 60% or more identity in the complete L1 region is grouped into the same genus. Today there are five different genera in which HPVs are classified. These are the alpha-, beta-, gamma-, mu- and nu-genera. Within the beta-, gamma-, mu- and nu-genera the HPV types infecting cutaneous tissues are found. These infect the skin and form asymptomatic infections or common skin warts [1]. Within one of these genera there are two HPV types which stand out. These are found in the beta-genus and are HPV type 5 and 8. For the main part these cause normal infections of the skin, however, in patients with Epidermodysplasia verruciformis they have been linked to the formation of squamous cell carcinomas [4]. In the alpha-genus on the other hand HPV types infecting both cutaneous and mucosal tissues are found. Within this genus the mucosal HPV types 6 and 11 are found. These are known to cause benign genital warts and infections with these are non-harmful to the host. Found in the same genus is HPV16 and 18 which have throughout history been linked with a strong association to cancer, mainly cervical cancer [1], [5]. Due to the association of some HPV types with the formation of cancer another grouping or classification of HPV types is also used. This

classifies the HPVs into high- and low-risk, where high-risk are linked to pre-malignant and malignant lesions while the low-risk have no association with malignant transformation in healthy individuals. Interestingly, the high-risk HPVs are only found in one genus, namely the alpha-genus, and today there are evidence for 20 different HPV types and malignant transformations [5], [3]. By the International Association for Research on Cancer (IARC) these twenty types are further divided into group 1, group 2a and group 2b carcinogens (Table 1). Group 1 carcinogens have been proven to be directly linked to cancer while 2a and 2b are probably and possibly carcinogenic, respectively [6].

Table 1. Showing the IARC classification of high-risk HPV types.

Phylogenetic group	Group 1 carcinogens (High-risk)	Group 2A carcinogen (probably high-risk)	Group 2B carcinogen (possibly high-risk)
α5	51		26, 82
α6	56		53, 66
α7	18, 39, 45, 59	68	70
α9	16, 31, 33, 35, 52, 58		67
α11			73

HPV is today the most commonly sexually transmitted disease amongst the human population. It is estimated that the majority of sexually active men and women will be infected at least once during their lifetime, however, most will never develop any symptoms of the infection [7], [8]. Infections with low-risk HPVs such as HPV6 and 11 can result in the formation of papillomas in the genitalia, however, these will not progress into malignancy [1]. Another example of disease caused by low-risk HPVs is recurrent respiratory papillomatosis (RRP). This is usually a benign disease which occurs in the aero-digestive tract of children. Majority of cases are caused by HPV6 and 11 and infection predominantly occurs during delivery of mothers harboring genital warts. The infected children may in many cases need to undergo surgery to remove the papillomas to secure safe airways and proper speech. Interestingly an association has been made between incidence and socioeconomic status where lower education levels and socioeconomic status have a high incidence [9].

Infections with high-risk HPVs usually results in clearance of infection, however if these infections persist, they can result in malignancies. Today it is found that HPV results in 4.5% of human malignancies and a strong association between high-risk HPVs and cancer has been made. In cervical cancer the majority of cases have been linked to HPV infection and for several other cancers a strong association between HPV infection and progression to malignancy also is seen [5]. Some of these cancers are anal cancer (88%), vulvar

cancer (43%), vaginal cancer (70%), penile cancer (50%) and oropharyngeal cancer (37%), where the number in brackets indicate the percentage of cases caused by HPV [7]. However, today the association between oropharyngeal carcinomas and HPV is strongly increasing and in United States and Western Europe this association is now 70-80% [10]. And it is rapidly replacing oropharyngeal carcinomas caused by tobacco consumption and alcohol intake, which previously were the two main contributors [11], [12].

The main cancer type formed from HPV infections is squamous cell carcinoma. This since the cell type foremost infected by HPVs is immature squamous cells. In the cervix these cells are mainly found at the transformation zone which is where the squamous epithelium of the outer cervix meets the columnar cells of the endocervix. However, infections of HPV can also occur at other places in the mucosal tissue. Several risk factors have been identified and linked to progression into high-grade dysplasia and cancer. Some of these factors are persistent HPV infection, age over 30-years, multiple HPV infections simultaneously, high-risk HPV infection, tobacco use and immunosuppression [13].

Today it is estimated that around 604,000 new cases of cervical cancer are diagnosed every year. It has also been found that HPV is responsible for 3.3% of cancer-related deaths amongst women reaching a total number of 325,000 deaths per year [14]. However, the number of infections and deaths is not evenly distributed across the world. When looking at this distribution it is seen that 85% of these deaths occur in developing countries. The explanation for this can be several, however, some important factors are less implementation of cervical screening programs and less vaccination coverage of the HPV vaccines [7]. By implementing screening programs such as cytological screening programs it has been shown that the mortality and incidence of cervical cancer can be reduced [15].

There are effective vaccines against HPV. What all these have in common is their target. These vaccines are constructed for the immune system to target the L1 major capsid protein and elicit an immunological response. More specifically the vaccine consists of L1 capsid proteins which assembles into virus like particles resembling the actual virions and this elicits the immunological response in the body [16]. On the market are three different vaccines Cevarix, Gardasil and Gardasil-9. These are bivalent, quadrivalent and nonavalent, respectively. What these have in common is that they target the two most common cancer-causing HPV types, namely HPV16 and HPV18, however in addition to these Gardasil and Gardasil-9 have protection against 2 or 7 other HPV types, respectively. Today it is recommended to vaccinate both boys and girls and vaccination should start at 11-12 years of age [17].

HPV16 genome and life cycle

HPV genome

The HPV genome is around 8kb in size and is normally divided into three distinct regions and contains eight open reading frames (ORFs) (Figure 1 A and B). One of these regions is the long control region (LCR) or upstream regulatory region (UUR) or noncoding region (NCR) as it is also called. This region contains many important regulatory sites such as transcription control sequences and the origin of DNA replication. Within this region the early promoter is also found. Following the LCR is the early region. This contains a total of six ORFs (E1, E2, E4, E5, E6 and E7) and all genes expressed from this region are given the letter E to designate early and a number. The expression of the early genes is transcribed from the early promoter, p97 in HPV16. Transcription from this promoter has the ability to express all early genes. Following the early promoter is an additional weaker promoter called the late promoter, for HPV16 p670. This is mainly active during the vegetative replication and transcription from this promoter allows for expression of the late region. ORFs from the late region are designated using the letter L and, in this region, the two capsid proteins, L1 and L2, are located [18]. Additional to the two promoters, two polyadenylation signals are also found within the HPV16 genome which also regulate the gene expression. Just like for the promoter these are divided into the early and the late polyadenylation signals, pAE and pAL, respectively. pAE is as indicated active early in the early phase of infection and results in polyadenylation of the early transcripts. As the cells differentiate and the viral lifecycle progresses to the late phase, the polyadenylation from the late polyadenylation site is upregulated allowing for expression of the capsid proteins, L1 and L2 [19]. Normally the HPV genome is covered with human histones which also regulate the expression from the early and late region depending on histone modifications. The phosphorylation, acetylation, SUMOylation and several other modifications of these allow for regulation of the HPV genome [20].

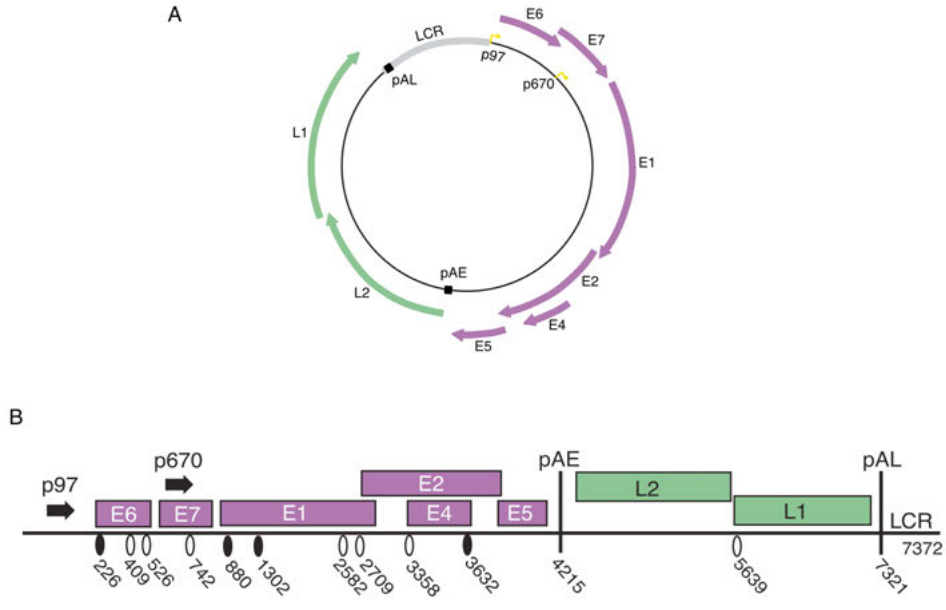

Figure 1. Schematic representation of the HPV16 genome. **A.** The circular form of the HPV16 genome. The LCR, early and late regions with the respective ORFs are seen as well as the two polyadenylation signals, pAE and pAL. **B.** A linearized form of the HPV16 genome seen in A. The two promoters, p97 and p670 are indicated. Indicated are also the splice donors and acceptors using black and white ovals, respectively.

The natural HPV lifecycle

An HPV infection is not a lytic infection. Instead, the viral genome is maintained for a long time and the viral life cycle progresses dependent on the terminal differentiation of the infected epithelial cells [18]. The natural HPV life cycle can be divided into three distinct replication phases (Figure 2). These include initial (establishment), stable (maintenance) and vegetative (productive) viral DNA replication. Following these stages is the assembly and the maturation of new virions, which can infect new individuals by skin to skin or sexual contact. Normally, an infection with HPV is cleared after 12-24 months with little to no detection by the immune system. However, in rare occasions the infection is not cleared but becomes persistent and may progress to malignancies [21].

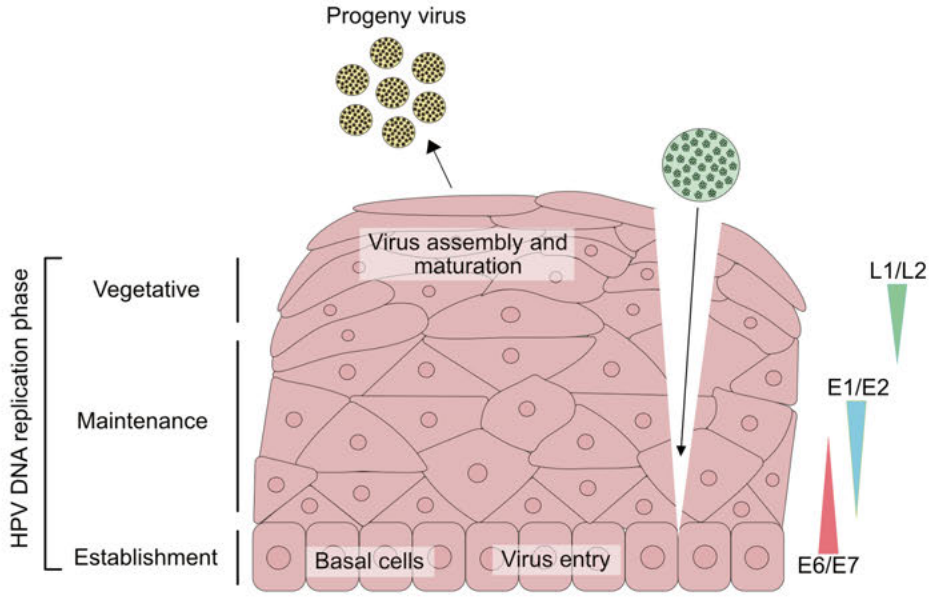

Figure 2. Displaying an HPV infection of an epithelium. At the side of the epithelium the E6/E7, E1/E2 and L1/L2 expression of an abortive infection is seen.

Infection and initiation of viral DNA replication

Primary target cells infected by HPV sit in the basal layer of the squamous epithelium [22]. The virus gains access to these cells in two possible ways. One is through micro abrasions in the epithelium giving access to the basal epithelial cells and the other is by reaching cells in the squamocolumnar transformation zone between the ecto- and endo-cervix, an area which is especially sensitive to HPV infections [18], [23]. It is necessary for the virus to target the cells in the basal layer since the yet unidentified infection receptors may be expressed specifically in these cells. They are also so-called stem cells and are therefore amplifying and actively dividing cells. By targeting these cells, the virus assures a stable infection [22].

In order for the virus to infect cells of the basal epithelium, the L1 protein in the viral capsid (Figure 3) binds to the heparin sulfate proteoglycans (HSPGs) on the cell membrane. Once bound to the receptor it induces several conformational changes in the virion facilitating the virus to bind to another, yet unidentified (secondary) receptor(s), and loose its connection with HSPGs. Following this the L2 protein is displayed on the surface which allows for a furin protease cleavage and the virus is thereafter internalized in an endosome. Within the endosome several changes occur in the capsid and viral DNA, as it matures. The endosome travels through the trans Golgi network and eventually the complex of viral DNA and L2 protein ends up at the cell nuclei.

Here the HPV genome is internalized as the infected cell undergoes mitosis. Once the viral DNA has gained access to the cell nuclei it can be replicated and transcribed [18].

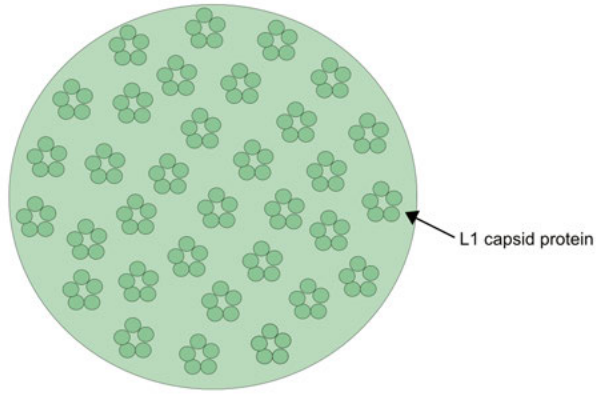

Figure 3. The HPV virion. Indicated using an arrow is the pentamer of L1 proteins formed on the outer part of the virion.

Stable maintenance replication

After gaining access to the nuclei of the basal cells, the transient genome amplification of HPV is initiated. Two of the first proteins to be expressed is the E1 and E2 proteins. These proteins are needed for viral DNA replication and with the help of these and the cellular DNA replication machinery, the HPV genome copy number reaches a level of 50-200 copies per cell [24]. After reaching this level, the viral genomes are replicated once per cell cycle. In this way the viral genome is maintained in the basal cell layer but also in the suprabasal layer. This phase in the HPV infections lifecycle is known as maintenance replication or “latent” infection. Following this phase is the vegetative replication [18].

Vegetative replication

When cells of the basal epithelium divide it results in two new cells, one which remains in the basal layer, and one which travels up through the epithelium to the suprabasal layer where it starts to differentiate. It is within the cells in the suprabasal layers that the vegetative HPV replication occurs. Normally, these differentiating cells would not allow for DNA synthesis since they have exited the cell cycle. However, due to the expression of viral genes the cellular DNA replication machinery remains active allowing for replication and expression of HPV genes [3], [18], [25].

Two HPV proteins which are expressed early in the infection and are crucial for HPV genome replication are the E1 and the E2 proteins. The E1

protein is an ATP dependent helicase and is therefore the only enzyme within the HPV genome [26]. Within the LCR region is the origin of replication (Ori). Here several AT-rich sequences are located to which the E1 protein binds. This binding is weak but the binding and oligomerization of the E1 protein is stabilized by the E2 protein which binds to sequences surrounding the E1 binding site. The binding of these two proteins allows for recruitment of the cellular DNA replication machinery and replication of the viral DNA can commence [27]. Another two proteins which are expressed during this time are the E6 and E7 proteins. The expression of these two proteins is crucial during this stage of infection. This since the two proteins forces the cells to stay in the cell cycle and not reach terminal differentiation. Within the cell the E7 protein has many functions, however, one of the most important is to bind to the retinoblastoma protein (pRb) family and target these proteins for degradation. The high-risk E7 protein targets several proteins in this family. The p107 and p130, which are involved in control of cell cycle entry, are degraded by E7, resulting in the re-entry into the cell cycle in the upper epithelial layers [28], [29], [30]. The degradation of pRb proteins results in a release of the E2F proteins, which is a family of transcription factors. The release of these results in an upregulated gene expression, mainly of genes which control DNA synthesis and proliferation of cells. Additional to the effects of E7, the HPV E6 protein also have multiple functions within the infected cell. One of these is to target the p53 protein for degradation. This is done by interacting with E3 ubiquitin ligase, also known as E6 associated protein, and target p53 for proteasomal degradation. With lowered p53 levels, pro-apoptotic and anti-proliferation effects are now reduced [31]. Additional to the effect on p53, the high-risk HPV E6 proteins have also been shown to increase the telomerase activity in the infected cells. In this way the integrity of the telomeres is kept intact even though there is an increased cell division of these cells [32].

As the infected cells travels up through the epithelium several changes occur within the cells and these changes are also displayed on HPV16 gene expression. One major change is the switch in promoters. The expression from the early promoter is turned off and the expression from the late promoter is now upregulated [33], [34]. This allows for a shutdown of E6 and E7 gene expression and an extensive expression of the E1, E2, E4 and E5 genes [35], [36]. The high expression of these genes and elevated levels of E1 and E2 protein leads to a vast replication of the viral genome bringing the genome copy number in these cells to several thousands of copies per cell [37]. As the cells migrate even further up in the epithelium there is a read through of the early polyadenylation site allowing for expression of the viral capsid proteins L1 and L2. Once the L1 and L2 proteins have been synthesized concomitant with massively replicated viral genomes it allows for assembly of the virions. This occurs in terminally differentiated epithelial cells. This since the environment within these cells have become oxidizing and the oxidizing environment will form strong disulfide bonds between the L1 proteins in the viral

capsid. Once the virions are formed, they shed together with desquamated cells. This allows for viral shedding without necrosis or apoptosis and therefore the cellular immune system will remain inactivated [38], [18].

Deregulated HPV life cycle

Normally an HPV infection is cleared within several months by the human immune system. However, in rare cases the infection persists and over time can result in changes of viral gene expression followed by changes in the infected cell [39], [40]. In this instance the otherwise transient viral infection becomes persistent and the viral HPV life cycle deregulated. If this is not cleared by the immune system at this stage the risk for malignant progression arises [40]. If the HPV infection occurs in the cervix, it first progresses to cervical disease and if not cleared by the immune system or treated it can progress to cervical cancer [41]. In order to prevent cervical cancer, cytological testing is used and a classification system to grade the severity of infection. The system is divided into three distinct stages 1, 2 and 3 and named cervical intraepithelial neoplasia (CIN) [42]. It is well known that a clear succession can be seen from the lowest stage, CIN1, to the higher CIN3 stage. If this is not treated or cleared by the immune system at this point the next stage is cervical cancer in situ [43]. However, solely because an HPV infection has resulted in a CIN stage all hopes are not lost. Analysis shows that around 80-90% of CIN1 infections revert and are cleared by the immune system. However, the higher on the CIN scale, the more difficult to clear the infection, nevertheless around 60% of high-grade CIN3 lesions can lapse without treatment [44].

In order for an HPV infection to become a deregulated non-productive infection it must start in the basal cells of the epithelia [45], [46]. However, a persistent HPV infection is not enough to result in cancer, but there must also be a dysregulated expression of both viral and cellular genes [39], [40]. Two viral genes necessary for malignant transformation are the E6 and E7 genes. The expression of these results in increased cell growth, instability of the cellular chromosomes and inhibition of cell differentiation, all events which can result in malignancies and tumor progression. In 70-80% of malignant transformations the increased expression of E6 and E7 oncogenes is due to integration of the HPV genome into the cellular genome [47]. Such an integration is an abnormal part of the viral life cycle and is a main contributor to the deregulated expression of E6 and E7. In many cases this integration occurs within the E2 ORF resulting in disruption of the ORF and therefore the protein expression. If the E2 protein is no longer expressed, the downregulation of the early promoter p97 and therefore downregulation of E6 and E7 production does not occur [45]. However, other studies have also shown that the E6 and E7 mRNAs can become more stable as a result of being transcribed together with a cellular ORF. This can also affect the translation efficiency of the

mRNA [48]. Worth noting is that in nearly 15% of cervical cancers the HPV genome remains unintegrated and in its naturally episomal form and the E2 ORF is therefore intact. How these cases result in malignancies is still not completely clarified [21].

Regulation of HPV life cycle: the DNA damage response

Within a human cell the DNA is not always intact or copied in the correct way. In order to prevent that incorrect or damaged genetic material is kept and transferred to the new daughter cells a phenomenon known as DNA damage response (DDR) exists within the cells. This is a system that reacts to abnormal DNA, resulting in a cascade of actions which repairs it back to its former self. Within the DNA damage response there are two important kinases known as ataxia-telangiectasia mutated (ATM) and ATM and Rad3-related (ATR). These are found to be activated as a result of DNA damage or replication stress and resolve it. However, importantly the ATM and ATR work in two different ways resolving different types of DNA damages within the cell. The ATM is known to act in repones of double strand breaks (DSB). As a response to these, it phosphorylates several factors such as BRCA1 and Chk2 which thereafter have downstream effects resulting in a solution to the DSB [49]. The ATR pathway, on the other hand, reacts to single strand breaks as well as stalled replication forks [50]. This pathway has several proteins to resolve the DNA break or stalled fork. One of these proteins is the Chk1 protein which is phosphorylated by ATR in order to have further downstream effects [51].

HPV was in 2009 found to use the DNA damage response to its advantage. This was done using HPV31 and it was found that it uses the ATM-dependent pathway for its viral replication and is dependent on the differentiation [52]. In 2011 two other high-risk HPVs, namely 16 and 18, were also found to use the ATM-dependent DDR for its replication [53]. It was also found that a 70% reduction of HPV16 viral copy number at maintenance phase was seen when inhibiting the ATM-dependent DDR in HPV infected cells [54]. Furthermore, the Chk2 protein was also found to be imperative. This since the Chk2 protein activates caspase-7 which in turn has downstream effects on HPV E1 protein activity and the vegetative replication [52], [55]. Two viral proteins which have found to activate the ATM-driven DDR are E1 and E7 [56], [53]. When E1 is co-expressed with E2 these form replication foci at which the viral DNA replication occurs. To these foci several DDR factors are recruited such as Rad51, BRCA1 and phosphorylated Chk2 [53], [57]. The HPV E7 protein functions in other ways when it comes to regulating the DDR. Studies have shown that it binds to and increase the half-life of several DDR factors. The part of the E7 protein responsible for this is suggested to be the pRb binding domain [58]. HPV E7 has additionally been found to activate STAT-5 to modulate ATM activation. STAT-5 is a protein which is involved in the innate

immune response and additionally been shown crucial for the vegetative HPV replication [59].

HPV has moreover been found to use the ATR-branch for its viral replication. The activation of the ATR pathway was likewise found to be dependent on the HPV E1 and E7 proteins [57], [60]. Unlike the ATM-dependent DDR, the ATR-dependent DDR is activated throughout the HPV life cycle [60]. At the initial replication phase, the activation of ATR is E1 dependent possibly due to the replication stress caused by E1 activity [57]. ATR is also activated at the maintenance and productive replication phases. This is believed to be through a direct effect of the HPV E6 and E7 proteins. This since they hijack the cellular replication system and forces the cell into a replication stage when it otherwise would have remained inactive [61], [62].

Two cellular factors which have been linked to the DDR are Bcl-2 associated transcription factor (Btf, otherwise known as BCLAF1) and Thyroid hormone receptor-associated protein complex 150kDa component (TRAP150). These have been shown to regulate the formation of several mRNA variants coding for proteins involved in the DDR. In cells with no TRAP150 and BCLAF1 expression, effects are seen on several steps in the RNA maturation process. Affected pathways are mRNA splicing but also mRNA export resulting in changes in protein expression levels of these DDR factors [63]. Additionally, an interaction between BCLAF1 and the well-known DDR factor BRCA1 has been identified. Somatic mutations of BCLAF1 have also been identified in several types of cancers [64]. Furthermore, it was demonstrated that TRAP150 was phosphorylated by DDR factors upon DDR induction, indicating a link between DDR and the functions of TRAP150 and BCLAF1. In fact, previously we have demonstrated that DDR dependent TRAP150 regulation of HPV16 late gene expression occurs both on cotranscriptional and RNA processing levels. This is evidenced by an increased association between TRAP150 and HPV16 DNA as well as TRAP150 and U2AF65 which has been seen as a response to the DNA alkylating agent melphalan [65]. On the other hand, the effect of TRAP150 and/or BCLAF1 on HPV16 early mRNA splicing have not been studied. In this thesis, we wished to address the regulatory mechanism of HPV16 E6 and E7 mRNA splicing mediated by TRAP150. These results show an increase in E6*I/E7 mRNA production when TRAP150 and BCLAF1 are overexpressed. Additionally, when TRAP150 and/or BCLAF1 levels are reduced a reduction on E6*I/E7 mRNA production is additionally seen (Paper I). Combining this knowledge, it could be interesting to study the TRAP150/BCLAF1 effect on DDR in HPV infected cells.

Viral proteins

E6 and E7

The E6 and E7 genes of high-risk HPVs are transcribed as a bi-cistronic pre-mRNA which is either kept intact allowing for E6 protein expression or spliced within the E6 ORF allowing for E7 protein expression [66]. Interestingly, these two proteins are continuously expressed in HPV-related cancer cells and a continuous expression of these are needed to keep the cancer properties of the cells. The expression of E6 and E7 allows for the six hallmarks of cancer. These hallmarks are as following, continuous proliferation signaling, facilitation of metastasis and invasion, production of angiogenesis factors, defiance of cell death, hinderance of growth suppression factors and replicative immortality. These are the results of the E6 and E7 proteins targeting and interacting with a row of cellular proteins, resulting in either degradation or up- or down-regulation of these [67].

The HPV E6 protein consist of around 150 amino acids. When looking at the different parts of the protein it is seen that there are two zinc finger domains and one PDZ-binding motif (Figure 4A). The zinc finger domains are specific for the E6 protein and are not found within human proteins. It is known that these are the responsible domains for the oncogenicity of the protein. The PDZ-binding motif, on the other hand, is responsible for binding and interacting with several human proteins containing a PDZ motif [68]. The HPV E6 protein has been found to interact with several proteins within the cell, however, the by far most important and well-studied is the p53 protein. This is a tumor suppressor protein which reacts to cellular stress and as a result of this binds to the promoters of cellular stress responsive genes and upregulates the expression of these resulting in apoptosis or cell cycle arrest. Since apoptosis would result in an abortive infection, the virus wants to overcome this and it does so by interacting with a cellular factor called the E6 associated protein (E6AP). When the E6 and the E6AP interact with each other these can bind to the p53 protein and ubiquitinate it, resulting in its proteasomal degradation (Figure 4B). And if the p53 protein is no longer present, the upregulation of stress responsive genes cannot occur and the cell will not enter apoptosis or cell cycle arrest via this pathway [69].

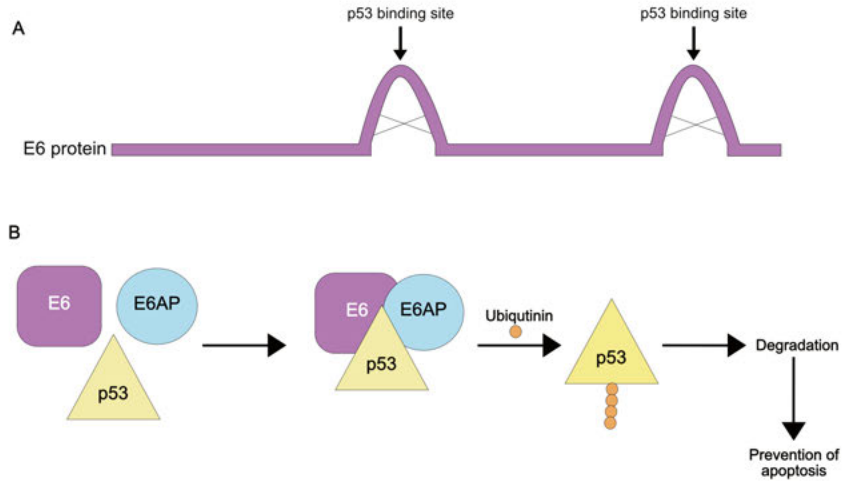

Figure 4. The E6 protein. **A.** Schematic representation of the E6 protein structure. Indicated ate the two p53 binding sites. **B.** Interaction between E6, E6AP and p53 displaying the downstream effects on p53 degradation.

In addition to the p53 pathway, HPV E6 also blocks apoptosis by targeting other p53-independent pathways. One of these is the Bax/Bak pathway where the Bak protein is targeted for degradation by the E6 protein [70]. If the Bak protein is degraded it can no longer bind to the mitochondria and result in cytochrome c release. It also binds to procaspase 8 and prevents it from further downstream effects in the apoptotic cascade [69].

HPV E6 proteins are further known to interact with several histone acetyl transferases (HATs) within the cells. These are targeted for degradation which results in downstream effects on transcription. One of these proteins is p300 which is a transcriptional activator of pro-apoptotic genes. However, in the case of p300, E6 binding only inhibits the functions and does not result in its degradation [69]. The high-risk E6 protein has furthermore been proven to modulate factors involved in keratinocyte differentiation [71]. One of these factors is Notch, which is a main factor in the signaling pathway regulating the differentiation development, and E6 represses the signaling pathway controlled by this protein [69].

The E6 protein also repress the expression of several cellular microRNAs. Some of these are miR-34a [72], miR-218 [73] and miR-23b [74] which are involved in cell cycle control and cell migration. How these are regulated by the E6 protein is, however, still unknown [69].

Beside from a number of factors which are downregulated by E6, a factor which instead is upregulated by HPV E6 expression is NF-κB. This upregulation is believed to have an impact on the malignant transformation of cells [75].

In 90% of cancer and immortalized cells the telomerase activity is upregulated and this is believed to be essential for immortalization and cancer formation. The upregulated telomerase activity results in a sustained telomere length and therefore a prevention of chromosomal instability resulting in senescence. In HPV infected cells the telomer length and telomerase activity is likewise upregulated. The protein responsible for this is the E6 protein. Studies using this protein has shown that E6 upregulates the expression of hTERT, which is the catalytic part of the telomerase activity. The expression of hTERT RNA was shown to be upregulated by E6 binding to a regulatory element within the hTERT promoter [76].

The E7 protein consists of around 100 amino acids and when compared to cellular proteins none are found to have a similar sequence (Figure 5A). However, the conserved LXCXE motif is found in human proteins. Interestingly, other viral proteins share sequence similarities with the E7 protein, e.g., the adenovirus E1A (AdE1A) protein and the simian vacuolating virus 40 large tumor antigen (SV40 T). It is believed that the E6 and E7 proteins might come from a common ancestor, since some papillomaviruses have one of the two proteins but not the other one, i.e., some has E6 but not E7 while other have E7 but not E6 [77].

Today many different factors have been identified to interact with the HPV E7 proteins resulting in various effects in the infected cells. The most well-known and well-studied interaction is that of the human retinoblastoma protein (pRb) family. This is a protein family consisting of many proteins, however, the pRb protein has been well studied in connection with HPV E7. The pRb protein is under normal conditions known to bind to transcription factors of the E2F family. This family consists of both positively and negatively regulating transcription factors. pRb binds several of the positively, upregulating transcription factors which are involved in G1 to S-phase progression. Under normal conditions the pRb protein is bound to E2F factors, however, as the cells approaches the G1 to S-phase transition the pRb is phosphorylated resulting in the release of E2F and an upregulation of S-phase responsive genes. As previously described, cells in the lowest layers of the epithelium have an active cell cycle, however, as the cells progress up in the epithelium, the cell cycle is turned off and the cells start to differentiate. The shutdown of the cell cycle is to some extent regulated by pRb binding to E2F factors. In HPV infected cells the E7 protein bind pRb. This binding results in a conformational change releasing the E2F transcription factors and the cells can reenter the cell cycle and upregulate the transcription of S-phase responsive genes (Figure 5B). The high-risk E7 proteins also target the pRb protein for degradation via the proteasomal pathway by binding several ubiquitin monomers to the protein. Additionally, two other factors of the pRb family have shown to be affected by HPV E7 proteins, i.e., p107 and p130 [78], [45].

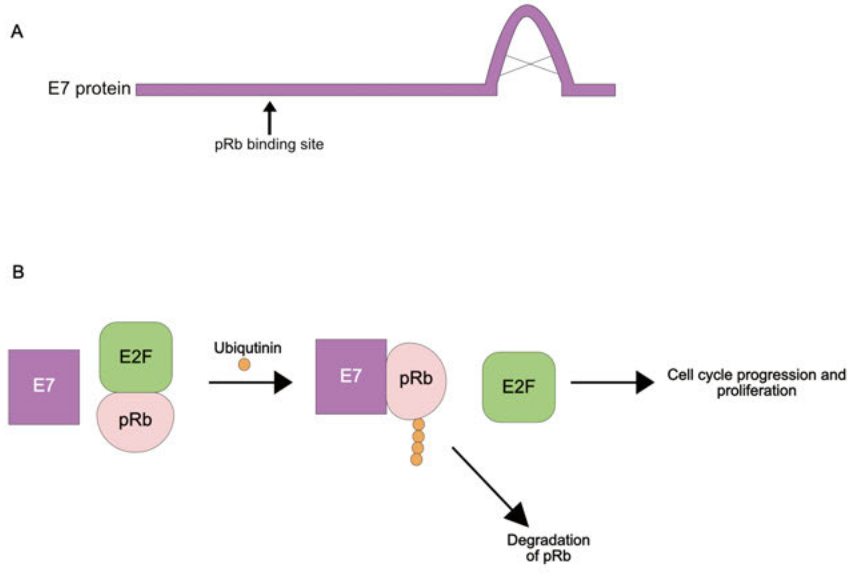

Figure 5. The E7 protein. **A.** Schematic drawing of the E7 protein structure. **B.** Interaction between E7 and the pRb protein resulting in release of E2F and upregulation of S-phase genes.

HPV E7 also interacts with factors of the innate immune response, i.e., interferon responsive genes (IRFs). These regulate the expression of interferons in response to pathogenic infections. HPV E7 prohibits these factors from binding to promoters and upregulate the expression of several immune responsive genes such as MHC I and interferons (IFNs). It has additionally been shown to bind to the STING proteins and block the increased expression of IFNs. STAT1 functions are also blocked and its nuclear translocation prevented which has downstream effects on the expression of several immune response genes [78].

The HPV E7 protein has also been found to modulate the expression of several cellular miRNAs by either suppressing, upregulating or downregulating the expression of these. One of these mRNAs is miR-145, which has been found to impact HPV genome amplification negatively and is therefore suppressed by E7. Further, miR-203, which stimulates keratinocyte differentiation, is also inhibited by E7 [78].

In addition to the above-mentioned abilities, the E7 protein can also change other actions in the cells such as epigenetic regulation. This is done by interacting with HATs, such as CBP and p300, which attaches acetyl groups on histones. Histone deacetylases (HDACs) which instead removes acetyl groups are also found to be regulated by the E7 protein. Other epigenetic changes such as DNA methylation is furthermore changed by the E7 protein and has

been found to increase at promoters of genes with negative effects on HPV infection [78].

E1 and E2

The E1 and E2 proteins are thought to be expressed very early in the infection process and throughout most of the viral life cycle. This due to their involvement in the replication of the HPV genome [79].

The E1 protein is the only enzyme expressed from the papillomavirus genome and is an ATPase dependent helicase. Like other HPV proteins it shares homology with other viral proteins with similar functions such as the large T antigen of SV40. The size of E1 is approximately 650 amino acids and the protein is usually divided into three separate parts based on the functionality of each part. These are the N-terminal domain, the central region and the C-terminal domain [26]. The N-terminal domain constitutes the first 200 amino acids of the protein and in this region the nuclear import and export signals are located [80]. Within the central region is the DNA binding domain (DBD) and this is the part of the protein which binds to the E1 binding sites within the LCR of the HPV genome. The C-terminal domain is the domain with ATPase activity and it is needed to form the hexameric structures necessary for the functional helicase activity of the protein [26].

The E1 function can be modulated via post-translational modifications. Methylation of the protein impacts the shuttling and SUMOylation and ubiquitination has also been detected, where the latter one results in degradation of the E1 protein [26]. Besides from its role in replication E1 affects the expression of the immune response genes. Such genes are interferons, e.g, IFNβ and IFNλ, and interferon stimulated genes (ISGs), which are all downregulated by E1 [81].

The E1 and E2 proteins are both needed to replicate the viral genome in vivo. E1 binds to several binding sites at the origin of replication, however, to gain access here it also needs to interact with the E2 protein which helps to stabilize its binding. The E2 protein is divided into several regions with discrete functions. These are the N-terminal domain, which is also called the transactivation domain and binds directly to the E1 protein. The C-terminal domain is on the other hand is called the DNA binding domain and this domain is responsible for E2 binding within the LCR [26]. There are four E2 binding sites within the LCR on the HPV genome. These sites are overlapping with the binding sites of cellular transcription factors that regulates the early promoter and binding of E2 at these sites regulates the promoter activity by competing with the binding of transactivators or transrepressors. During mitosis E2 binds to the HPV genome and interacts with cellular proteins called chromatin adapter proteins. Through this interaction the viral genome ends up in both daughter cells [27], [82].

The viral replication occurs in nuclear foci and it has been shown that E2 is necessary for the formation of these. Within the foci other viral proteins are also found which induce the DDR and allows for viral replication [27].

E4 and E5

The functions of the E4 and E5 proteins are less well known than those of other HPV proteins. The E4 protein has several functional motifs which are linked to distinct functions. These motifs are the N-terminal domain, the central region and the C-terminal regions, where the first one is rich in leucine residues and the central one has a high proline content [83]. The N-terminal region mediates keratin binding, whereas the C-terminal region is involved in self association, forming amyloid-like structures. The length of the central region varies among different HPVs, contributing to structural modifications. It has been suggested that high-risk HPV E4s are crucial for viral DNA replication in differentiated cells but not in undifferentiated cells although low-risk HPV E4 provides conflicting results [84], [83].

A high E4 protein expression has been seen during the vegetative replication phase. During this time the E4 protein is thought to bind cellular proteins such as Cyclin B and Cdk1. By binding to these proteins, their nuclear translocation and phosphorylation are inhibited. This may result in a G2 to M phase cell cycle arrest in the epithelial cells, which is considered to contribute to the induction of vegetative viral DNA replication at differentiated cells although the evidence is still missing [84]. Furthermore, in the upper epithelial layers, where virus assembly occurs, the E4 protein is expressed abundantly and may interact with keratin filaments and weaken the keratin cytoskeleton networks, which is assumed to help in virus release [3], [85].

Modifications such as phosphorylation has been identified on the HPV E4 protein, however these have been found to be transient modifications and changes throughout the cell cycle. The expression of E4 protein is assumed efficient if it is transcribed from the late promoter and is detected excessively in the late phase of viral life cycle compared to other early viral genes [84].

The E5 protein is a small protein consisting of 83 amino acids [86]. Interestingly, not all HPV types contain the E5 ORF, but it is found within all high-risk HPVs [87]. Just like other proteins the E5 protein has several important domains such as the N-terminal, central, and C-terminal. All these three domains have hydrophobic regions which forms three transmembrane domains and due to this the E5 protein is mainly found in cellular membranes [88]. It is mainly located to the endoplasmic reticulum, however during high expression it is also found to be localized to other membranes such as the Golgi network, plasma membrane and nuclear envelope [87]. Like the E6 and E7 proteins, E5 also has transforming activities and even though these are considered weaker it is still regarded as an oncogene. One important cellular receptor interacting with the E5 protein is the transmembrane receptor named

epidermal growth factor receptor (EGFR). The E5 protein increases the number of receptors at the cellular surface, which in turn makes the cells more susceptible and sensitive to epidermal growth factors. This increase is caused by an upregulated recycling of receptors towards the cell surface [89], [87]. Additional, to the transforming activities, HPV E5 is also responsible for immune evasion by changing MHC class I receptors at the cell surface. MHC class I interaction with E5 keeps these receptors at the Golgi and ER and prevents translocation to the cell membrane. Additional effects are also seen on MHC class II where processing and maturation is changed by E5 expression [87]. Furthermore, E5 interacts with STAT1 which changes interferon stimulating genes (ISGs) expression [86].

L1 and L2

The L1 and L2 proteins are the two late genes expressed from the HPV genome. These make up the viral capsid and are necessary for the formation of new virions. Due to their high immunogenicity, they are expressed late in the infection process to avoid detection by the immune system.

The L1 protein is the major capsid protein with a size of roughly 55 kDa. This protein can self-assemble into structures resembling the viral particle, called virus-like particles (VLPs). Since these VLPs are highly immunogenic they have been used to produce vaccines against HPVs. However, the L1 structure differ substantially between different HPV types and therefore neutralizing antibodies against one HPV type functions poorly against other closely related types. Therefore, a vaccine needs to be designed against each L1 particle it aims to target [90].

In each virion the L1 particles are connected via disulfide bonds and assembles into pentamers. Each virion is made up of 72 pentamers meaning that a total of 360 L1 particles make up the HPV virion [91], [90].

The L1 protein have other important functions besides from creating the protective virion. One of these is during the infection of new cells. In this process the L1 capsid protein is the first to make contact with the non-infected epithelial cells. This is done by binding to cellular receptors and when bound to this receptor a conformational change occurs in the virion which allows for the L2 protein to be displayed and make contact with a yet unidentified receptor. The virus is thereafter taken up in the cell via the endosomal pathway and the infection has begun [90], [18].

The L2 protein is the minor capsid protein consisting of approximately 500 amino acids. The size of the protein is similar to that of the L1. The L2 protein assembles with the virion after the L1 proteins have formed their associations and today it is known that the number of L2 can differ, however, the maximum number is 72 rendering the L2 to L1 ratio 1:5 [91], [92], [93]. In the infection process the L2 capsid protein has been found to be cleaved by furin. This cleavage is an essential event in the infection process. L2 is also necessary for

the HPV genome to exit the endosome and enter the cellular nuclei, where the genome is replicated and transcribed [92]. Before entering the nucleus, the endosome firstly matures from an early to a late endosome, the L2 proteins thereafter make sure that the viral DNA travels through the trans Golgi network and thereafter reaches the nucleus. When the cell undergoes mitosis, the nuclear membrane become more porous allowing for the L2 proteins bound to the viral DNA to tether the viral DNA to the cellular DNA undergoing mitosis, followed by the formation of nuclear membranes. In this way the nuclei of both new cells will harbor the viral DNA [91], [92]. The L1 and L2 proteins are also important in the formation of new virions. As the cells reach the upper layers of the epithelium the L1 and L2 expression is high and virion assembly takes place. When the HPV genome is ready to be packed the L1 and L2 proteins translocate to the nucleus where the genome is packed into new virions [92].

Regulation of gene expression and mRNA maturation.

To regulate the HPV gene expression several processes are used. All the processes are administered by cellular machineries. Therefore, the general mechanism of cellular gene expression is discussed in this section. Some of these mechanisms are alternative splicing and alternative polyadenylation. For an mRNA to be recognized as mature and functional, a cap structure must also be added at the 5'-end of the mRNA [19], [66], [94]. These mechanisms and how HPV16 is regulated using these are described below.

5'Cappling of pre-mRNA

As a part of the maturation process towards a mature mRNA a 5'-cap is added at the 5'-end of the pre-mRNA [95]. The 5'-cap is attached to the first nucleotide of the pre-mRNA and consists of a N7 methylated guanosine. The modification is linked to the transcription process and is the first modification to occur on the pre-mRNA. Three cellular proteins are involved in this process namely an RNA guanylyltransferase, an RNA triphosphatase and a guanine-N7 methyltransferase. Once the 5'-cap is added to the mRNA a cap binding complex (CBC) binds to it and this binding is necessary for many processes in which the mRNA is involved [96]. One of these processes is translation and the cap-structure has proven to be of major importance for this. It is needed to initiate translation and an interaction between the 5'-cap and the 3'-polyA tail has been identified during the translation [95]. Additional to its role in mRNA translation the 5'-cap has shown a major importance for RNA stability as well as involvement in pre-mRNA splicing and polyadenylation. If the 5'-cap is removed from the mRNA it is degraded via exonucleases. The addition of the

cap is also needed for the mRNA to be recognized as a fully mature mRNA and allow for nuclear export [96].

Polyadenylation

To mature into a mRNA the 3'-end of the pre-mRNA must also be modified. This is done by attaching a polyA tail and as the name implies it consists of a stretch of adenosine residues. The purpose of this polyA-tail is many but one of the most important ones is mRNA stability. By attaching the tail at the 3'-end of the mRNA, the stability and half-life is increased. However, in addition to this it also aids in nuclear export and is important for mRNA translation [97]. The length of the polyA tail has been shown to determine its stability. If the tail is short the translation efficiency is poor, which can result in decapping of the mRNA following mRNA decay. The process in which the polyA tail becomes shorter and shorter is known as deadenylation and this is done by exonucleases [95]. For an mRNA to become polyadenylated several cellular proteins and sequences on the pre-mRNA are involved. The sequences are in total three and consists of a polyadenylation signal, a polyadenylation site and a U or GU-rich element downstream of the polyadenylation site. The polyadenylation signal has a consensus sequence of AAUAAA. Following these three sequences is also a cleavage site at which the pre-mRNA is cleaved during transcription [97], [98]. In many pre-mRNAs more than one polyadenylation signal can be found allowing for further mRNA regulation and formation of different mRNAs from the same reading frame [97]. As for the proteins or rather protein families there are at least four involved. These are the cleavage and polyadenylation factor (CPSF), cleavage and stimulation factor (CstF), cleavage factors I and II (CFI and CFII) and polyadenylation polymerase (PAP). CPSF is the protein which senses and binds to the polyadenylation site. Once bound here it interacts with and pulls down other proteins, such as CstF and PAP, resulting in both cleavage and addition of a poly-A tail to the pre-mRNA [98].

Splicing and alternative splicing

When a mRNA is transcribed, it is called a pre-mRNA, since it needs to go through several modifications to transition into its mature mRNA form. One of these modifications is splicing. In the splicing process the introns, which are classically defined as the non-coding parts, are removed and the coding exons are joined together. Since this is a very delicate affair it needs to be done in the correct manner to maintain the integrity of the mRNA and therefore several sequences are found on the pre-mRNA to help direct the splicing machinery. These sequences are called the 5' splice site, also named splice donor (SD), the 3' splice site, or splice acceptor (SA), the branch point sequence (BPS) and polypyrimidine tract (PPT) (Figure 6). The 5' splice site is marked

by a GU sequence and designates the start of the intron, while the 3' splice site is defined by an AG sequence and marks the end of the intron to be excised out [99]. In addition to these there are additional sequences surrounding the splice sites which highlights them and facilitates for the splicing machinery. For both the 5' and 3' splice sites are so called consensus sequences which renders a strong splice site which will be used constitutively. For the 5' splice site the consensus sequence is (C/A)AG_**GU**(A/G)AGU and for the 3' splice site it is (C/U)nX(C/U)**AG**_(A/G) [100]. Additional to the 5' and 3' splice sites are the BPS and PP. These are located within the intron, upstream of the 3' splice site and are needed to define the end of the intron. The BPS is rich in adenosines and the PPT is as the name denotes rich in pyrimidines [99].

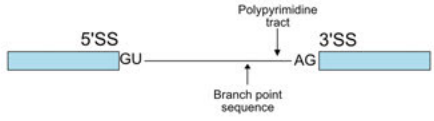

Figure 6. Elements needed to for splicing. Displayed are 5' and 3' splice sites and the branch point sequence and polypyrimidine tract.

To carry out the splicing a multiprotein complex called the spliceosome is used. This consists of over 100 different proteins and several small nuclear ribonucleoproteins (snRNPs) [101]. The spliceosome can recognize the introns and exons in two different ways, depending on their length. In the human genome the classical way is long introns and short exons and if this is the case the spliceosome assembly occurs over the exon and it is titled exon definition. If the spliceosome assembly occurs over an intron, it is instead named intron definition, however, this only occurs in the presence of short introns. This event is uncommon in the human genome but more common in lower eukaryotes [102].

When an mRNA is spliced, it is done so in a step-by-step process (Figure 7), where different spliceosomal factors bind to the pre-mRNA and form different complexes. These complexes are designated using letters e.g., E and A. In the first complex, named complex E, the U1 snRNP binds to the 5' splice site and two factors bind and defines the 3' splice site and BPS. The SF1 protein binds to the BPS and the U2AF bind to the 3' splice site. Following this, the SF1 is displaced by the U2 snRNP which bind to the BPS and complex A is formed. Thereafter three more snRNPs are introduced, U4, U5 and U6 which bind to the 5' splice site forming the pre-B complex. Once U1 and U4 snRNP are released, the activated B complex is formed. In the last step, complex C, the second step of the splicing reaction takes place and the two exons are joined [102].

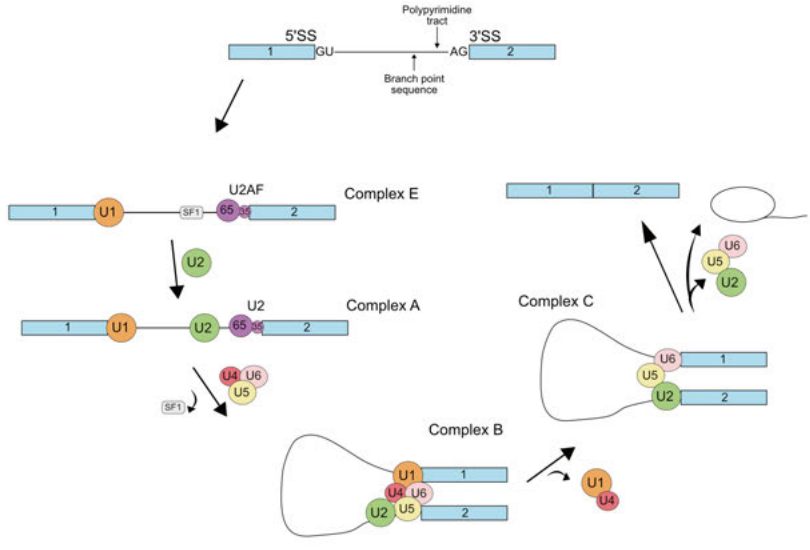

Figure 7. The different complexes in spliceosome assembly.

To complicate matters even more the sequences surrounding the 5' and 3' splice sites can differ from that of the consensus sequence. How well these coincide with the consensus sequence determines the strength of them and therefore also how well their usage can be regulated. A strong splice site corresponds well with the consensus sequence while a weaker splice site only partially overlaps with it. If the splice site is strong, it mainly results in constitutive splicing while a weaker splice site can allow for more regulation in the form of alternative splicing. In alternative splicing the same sequence can be included or excluded in the mature mRNA depending on the surrounding circumstances [99], [102]. This is the case for HPV where all splice sites are so called suboptimal splice sites, which do not completely correspond to the consensus sequence. This allows for regulation of the sites through alternative splicing [100].

Today there are 5 main types of alternative splicing (Figure 8). One of these is mutually exclusive exons, where two exons are included or excluded to the same extent. Additional to this, splicing events using alternative 5' or 3' splice sites are also seen as well as exon skipping and intron retention. All these events change the mRNA sequence and therefore also the coding sequence (CDS) of the protein which results in the production of different proteins or protein isoforms. On occasion the alternative splicing event can also result in the insertion of a stop codon which terminates the translation prematurely [101]. In recent years a new term has been stated for short protein-coding intron sequences located within the protein-coding exon and which can be

alternatively spliced. These are named exitrons. Even if retained in the mRNA such exitron-retained mRNAs are exported from the nucleus and translated. This produces a full-length protein whereas if skipped they can result in the production of a truncated protein. Splicing of exitrons could additionally cause frameshifts and the insertion or deletion of a premature stop codon [99], [103].

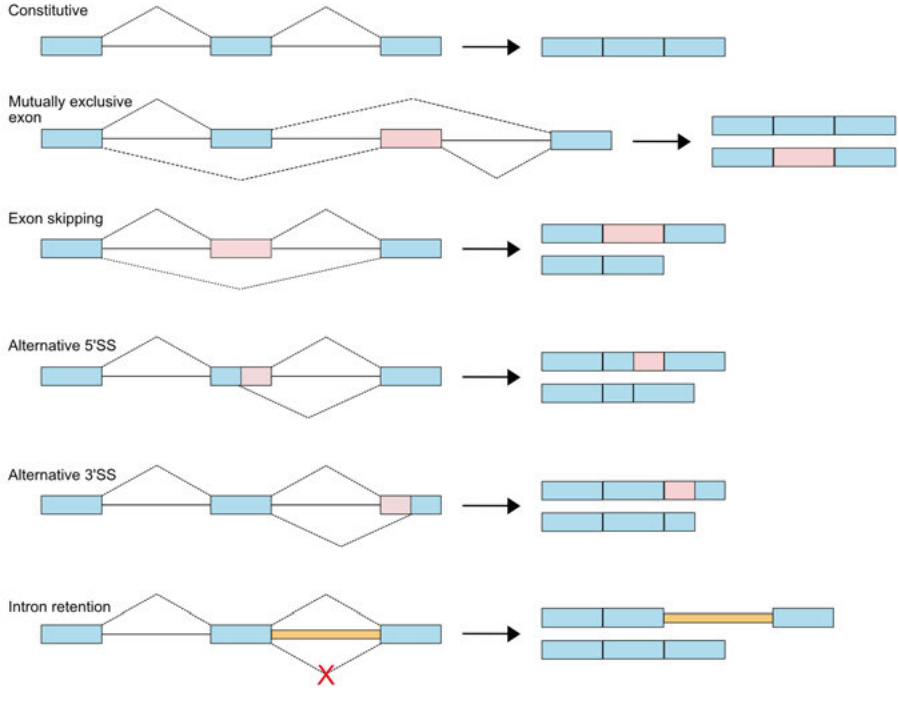

Figure 8. Different types of alternative splicing. The five different alternative splicing events are seen, mutually exclusive exons, exon skipping, alternative 5' and 3' splice sites and intron retention.

As an additional layer of alternative splicing regulation cis- and trans-elements exist within the pre-mRNA and it is the cumulative event of these which determines the alternative splicing outcome [102]. These elements are further explained below.

Cis-acting elements

Cis-acting elements are splicing regulatory elements located within the pre-mRNA and regulate the alternative splicing action. They are mainly active in alternative splicing and help to regulate suboptimal splice sites either by directing or preventing the splicing machinery to these sites. Depending on their location within the mRNA as well as their function they are given different names (Figure 9). A cis-element which positively stimulates splicing is called

an enhancer and depending on if it is located within the exon or intron, it is either named exonic splicing enhancer (ESE) or intronic splicing enhancer (ISE), respectively. If the cis-element on the other hand suppresses splicing it is deemed a silencer and therefore called exonic splicing silencer (ESS) or intronic splicing silencer (ISS) [102].

In many cases the alternative splicing event is a cumulative event occurring from the collective regulation of several cis-elements regulating the splicing of the same splice site [99]. The secondary and tertiary structure of the mRNA also affects regulation of the cis-elements. This since they can be displayed or concealed depending on the structure which will therefore impact the end result [99], [102]. To the cis-elements trans-acting factors bind and further regulates the splicing event [102].

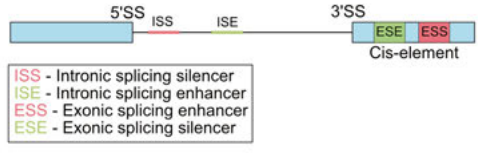

Figure 9. Cis-elements involved in pre-mRNA splicing. Both exonic and intronic enhancers and silencers are indicated.

Trans-acting factors

A trans-acting factor is an RNA binding protein (RBP) which binds to a cis-element and regulates the splicing machinery (Figure 10). They can be both positive and negative. Today there are two classical families of RBPs involved in splicing regulation and these are called serine/arginine-rich protein family (SR-proteins) and heterogenous ribonucleoprotein family (hnRNPs). Classically, SR-proteins are said to have a positive stimulation on alternative splicing, while hnRNPs have a negative regulation, however, contradictory evidence to this exists. It should also be noted that there are many more RBPs which have been found to regulate alternative splicing events [102]. Such proteins are described below. Since the RBPs are proteins, their function can be regulated via post-translational modifications, e.g., phosphorylation, SUMOylation, acetylation and so on. These modifications can change the activity and binding of the RBPs and therefore also the splicing outcome [99], [101]. It can also change their cellular localization as well as their stability [104].

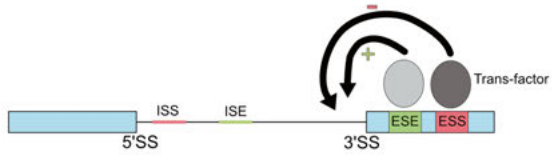

Figure 10. Trans-acting factors involved in pre-mRNA splicing.

SR-proteins

The serine/arginine-rich protein family are a family consisting of 12 different proteins which are called serine/arginine-rich splicing factors (SRSFs). Within this protein family there are several sequences or sequence motifs which are conserved. These are the RS or SR domain, which have given the family its name, and which consists of a stretch of arginine and serine residues. Additional there is also an RNA recognition motif (RRM) [105]. SR-proteins are known to positively stimulate splicing, however contradictory evidence also exists where SR-proteins negatively impact exon inclusion [106].

To be classified as an SR-protein there are two functional domains which are needed within the protein. These are the RRM and RS-domains (Figure 11). The RS-domain has been found to be conserved in other RBPs as well, however, since these do not fulfill the rest of the SR-protein criteria they do not belong to this protein family. They are instead known to be SR-related or SR-like proteins. The RRM domain of SR-proteins interacts with RNA like the RRM domains of other RNA binding proteins. The SR-domain is in turn responsible for protein binding. Within this domain is also the NLS signal which allows for SR-proteins to shuttle between the nucleus and cytoplasm. Within the nucleus the SR-proteins locate to so called nuclear speckles and it is the RS-domains which make sure they end up here. Additionally, the proteins can be post-translationally modified by phosphorylating serines in the RS-domain. This PTM facilitates the shuttling function and thereby also the ability to export mRNA to the cytosol. Furthermore, SR-protein have also been shown to affect the translation of mRNA [107].

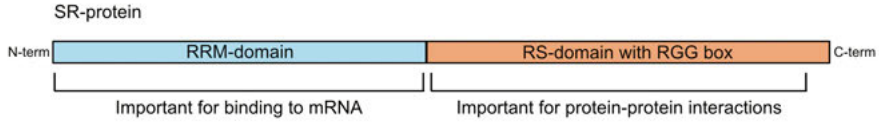

Figure 11. Schematic overview of SR-proteins.

hnRNPs

hnRNPs is a family of classical RBPs. These are predominantly involved in pre-mRNA splicing [105], however, involvement in nuclear export and translation have also been shown [108]. Normally hnRNPs are said to be inhibitory of splicing, however, several studies are today showing a different more context dependent picture. A total of 30 hnRNPs have today been identified and they are all categorized using letters, e.g. A to U [109].

The proteins in the hnRNP family share several conserved domains. These are an RNA recognition motif or in some cases a quasi-RRM or K homology (KH-) domain and an RGG-domain as well as sequences rich in glycine or proline or acidic amino acids (Figure 12). The RRM is as the name denotes a region which recognizes and binds to RNA. Each of these RRMs have a specific RNA sequence which it recognizes and binds to. For each hnRNP these sequences are different allowing them to bind to different sites on the pre-mRNA. The RGG-domain consist of consecutive repeats of arginine-glycine-glycine and it is this domain which interacts with other proteins [108], [105]. Many hnRNPs have a nuclear localization signal (NLS) which allows them to shuttle between the nucleus and the cytoplasm. However, this shuttling as well as other functions can be modulated and regulated via PTMs and today several PTMs have been identified on hnRNPs. Such modifications are SUMOylation, phosphorylation, methylation and ubiqutination which all affect the function of the proteins. It has also been shown that hnRNP expression is altered or disrupted in many human diseases and cancers, proving an important regulatory role for these in healthy tissue [108].

In this thesis there are three hnRNPs which are of particular interest and these are summarized below.

Figure 12. hnRNP protein family. Important regions are highlighted.

hnRNP H

The hnRNP H protein have three separate RRMs, however due to their non-consistency with classical RRMs these are called quasi-RRMs (qRRM). Two of the qRRMs are found in the N-terminal part of the protein, while the third is found in the C-terminal part interspaced between the two glycine-rich domains, GYR and GY [110]. The binding motif of the qRRMs have been determined to GGG- or GGGG-motifs, also called G-tracts. The glycine-rich domains are the two domains responsible for protein-protein interactions and within the GYR-domain a NLS have been identified which interacts with the nuclear import receptor Transportin 1 [108]. In the hnRNP H subfamily several proteins are found. These are hnRNP H, H' and F. The hnRNP H and H' are two different isoforms of the hnRNP H protein and have a sequence homology of 96%. The hnRNP F protein on the other hand is expressed from another gene but still has a 78% homology to hnRNP H and it also binds to similar RNA sequences [111]. Besides from its evident role in pre-mRNA splicing other functions have also been transcribed to hnRNP H and are connected to the maturation process of pre-mRNAs. Namely mRNA capping and polyadenylation as well as mRNA export and translation have been shown to be regulated by hnRNP H [110]. It has also been found to bind to pre-mRNAs

and pull down the spliceosome factor U2AF65, pointing to an important function in splicing [112]. hnRNP H has been shown to interact with the human telomerase complex. In the 5'-end of hTERC several GGG structures are found to which hnRNP H binds. When bound here it interacts with the telomerase holoenzyme and affect the telomerase function and length. This in turn affects the life span of cells and previous studies have pointed to a role in proliferation of cancer cells as well as metastasis formation and resistance towards apoptosis. Knock-down of hnRNP H prevented all these characteristics [113]. Previous interactions between hnRNP H and HPV16 are also reported. In this study binding of hnRNP H in the L2 coding region regulated polyadenylation at the early polyadenylation signal. In the same study it was also seen that the hnRNP H expression changes throughout the epithelium with an extensive expression in the lower layers while it is strongly reduced in the upper layers [114]. Like many other RBPs the function of hnRNP H can change with PTMs. Reported PTMs for hnRNP H are phosphorylation, methylation and SUMOylation [110].

hnRNP A1 and A2/B1

hnRNP A1 and A2/B1 are well studied hnRNPs. They have both been linked to roles in splicing and mRNA translation [108].

The hnRNP A1 protein contains two classical RRMs and one glycine-rich domain in which an RGG box is located [109]. The binding motif of the RRMs is identified as UAG-rich, however other similar motifs are also believed to be able to be bound by the protein [115]. Just like for other RBPs the glycine-rich domain is responsible for protein-protein interactions but the RGG-box has also proven possible to interact with RNA. In this region another important motif is also located, the M9 motif. This motif has been shown to harbor an NLS which in turn interacts with Transportin 1 and 2 and allows for nuclear and cytoplasmic shuttling. Several reports of PTMs also exist. Methylation has been shown to affect splicing and translation while phosphorylation affects the shuttling properties of the protein. Additionally, SUMOylation has proven important for translocation to the nucleus and its ability to bind RNA. Further effects on polyadenylation, 5'-capping, mRNA export, stability and translation are also linked to hnRNP A1 [109]. hnRNP A1 interactions with HPV pre-mRNA is previously reported. In HPV 18 it affects the formation of the E6*I/E7 mRNA [116]. And in HPV16 hnRNP A1 binds to a silencer sequence within the L1 coding region facilitating the expression of the L1 mRNA [117].

The hnRNP A2/B1 protein contains two RRMs which are found in the N-terminal part of the protein and the C-terminal part consists, like many other hnRNPs, of a glycine-rich region. In this region the RGG box is found as well as the nuclear localization signal. The RRMs have been found to bind to specific nucleotide sequences in the mRNA, namely AAG and UAG motifs. The RGG-box and glycine rich region is important for protein-protein interaction.

Within the cell the hnRNP A2/B1 functions are plenty. Involvement in transcription and RNA maturation such as splicing and export have been identified and hnRNP A2/B1 has also been shown to affect the stability and translation of the mRNA. The protein is additionally linked to DNA repair and has been found to negatively regulate the repair of DBS. It has a similar binding motif to that of hnRNP A1, which is not unexpected since they share 70% sequence homology. Like many other hnRNPs the hnRNP A2/B1 expression is deregulated in cancer [118]. It has previously been shown to affect alternative splicing of HPV16 mRNAs. Namely, it has been shown to prevent splicing of the late L1 mRNA by binding to regulatory sequences upstream of the L1 splice site [119].

TRAP150 and BCLAF1

TRAP150/THRAP3 (Thyroid Hormone Receptor-Associated Protein, 150 KDa Subunit/Thyroid Hormone Receptor Associated Protein 3) is a protein consisting of 955 amino acids (Figure 13). In the N-terminal domain a serine-arginine- (RS-) rich domain has been identified, however to this day no RNA recognition motif has been recognized in the protein. It belongs to the BCLAF1/TRAP150 protein family and shares a large sequence similarity with the BCLAF1/Btf (BCL2 Associated Transcription Factor 1/Bcl-2-associated transcription factor) protein. When comparing the amino acid composition, it is seen that they are 39% identical and show a 66% similarity in their sequences [120] and when comparing the C-terminal domain of the two proteins they show a 48% identity [121].

TRAP150 is post-translationally modified and phosphorylation of five different seine residues is found [122] as well as PARylation of the protein [123]. It is associated with the spliceosome pointing to an involvement in splicing [124] and an increased interaction between TRAP150 and the 3'splice factor U2AF65 is seen as a response to melphalan treatment, which is a DNA damage agent inducing the DDR [65].

TRAP150 is predominantly located to nuclear speckles together with other splicing factors and knockdown of the protein has a detrimental effect on splicing. By deleting functional parts of the protein, the N-terminal part with the RS-domain was shown indispensable for splicing. If this part was deleted splicing was affected. Deletions of the C-terminal domain resulted in no loss of splicing. TRAP150 is associated with the exon junction complex (EJC) as well as the TAP protein, pointing to a role in mRNA export [125].

In addition to the N-terminal domain another part of the protein between amino acids 685 and 775 is deemed important. This region interacts with the nuclear protein PSF/SFPQ (PTB-associated Splicing Factor/Splicing Factor Proline-Glutamine rich) and prevents it from binding to cellular mRNAs and regulate splicing of these. This region is now called the PSF-interacting domain (PID) [126].

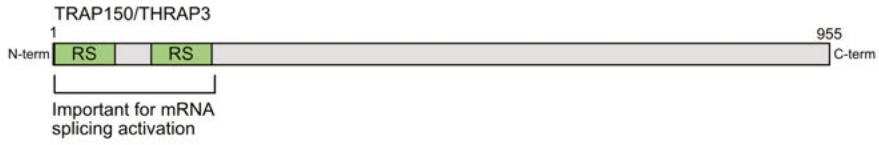

Figure 13. Schematic drawing of the TRAP150 protein.

BCLAF1 is encoded into several different isoforms with the largest of these being 920 amino acids. Within the protein there are several important functional domains such as the RS-domain, the bZIP and the MYB DNA-binding domain (Figure 14) [127], [128]. The RS-domain is located at the N-terminal part of the protein and has been linked to several functions such as pre-mRNA splicing and establishment of the ribonucleoprotein [129], [130], [131]. The RS-domain of the protein interacts with spliceosome factors and snRNPs such as U1-70K and U2AF [132], [133]. More specifically interactions with both U2AF factors, U2AF65 and U2AF35, is seen [64], [65]. The bZIP and the MYB DNA binding domain regulate the proteins transcriptional affinity [134], [135]. Both BCLAF1 and TRAP150 are involved in the spliceosome. However, while TRAP150 is a component of complex B, BCLAF1 has been linked to complex C. In this complex it is associated with proteins of the exon junction complex [136], [137], [124]. Studies involving both TRAP150 and BCLAF1 have shown that knockdown of one protein results in an upregulation of the other, pointing to an overlapping or compensatory function of the proteins [120].

In addition to the role in pre-mRNA splicing BCLAF1 is also linked to the DDR [64], apoptosis [127], proliferation and differentiation of muscle cells [138], [139] and the cell cycle [140]. In the DDR it interacts with BRCA1, a key protein in DDR, and truncations indicates that the C-terminal is responsible for this action [64].

Previous links between BCLAF1 and viral infections are reported. In these infections the Kaposi's sarcoma-associated herpesvirus (KSHV) reduces BCLAF1 protein by expressing a viral micro-RNA [141]. However, in studies using HPV it looks different. Here BCLAF1 increases in response to melphalan treatment and this increase results in an upregulation of HPV16 replication and late gene expression. An interaction between BCLAF1 and U2AF65 was also seen [142].

Figure 14. Schematic drawing of the BCLAF1 protein.

Regulation of HPV16 gene expression

From the HPV16 genome many polycistronic mRNAs are formed which code for several viral proteins [94]. More specifically, more than 20 mRNAs are formed and the formation of these is regulated on multiple levels where a predominant one is alternative splicing (Figure 15 A and B). As described previously two promoters are found within the HPV16 genome and as a general rule of thumb these are not active at the same time. Generally, the early promoter is active early in the viral life cycle and as the cells travels up in the epithelium there is a switch in promoter usage and therefore also a switch in gene expression. Since the E6 and E7 genes are only expressed from the early promoter it leads to a shutdown of their expression once the late promoter, p670, becomes active. Additionally, there are two polyadenylation sites, the early and the late, pAE and pAL, which further add on to the regulation [66]. These are also active at different timepoints of the life cycle. Surrounding pAE, there are hnRNP H binding sites suppressing readthrough at this region. Since the hnRNP H expression levels are lowered as the cells differentiate and reach terminal differentiation this allows for a readthrough of the early polyadenylation signal and an upregulation of pAL usage and late gene expression [114].

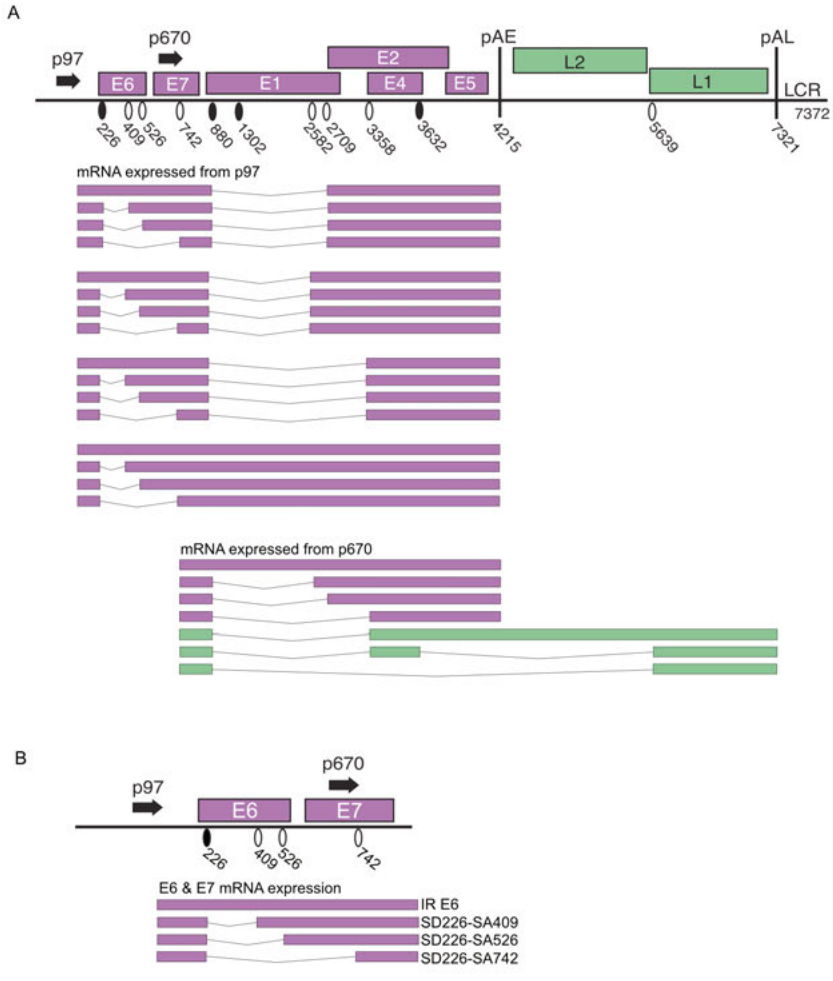

Figure 15. A map of different HPV16 mRNAs formed through alternative splicing. **A.** mRNAs expressed from the early promoter, p97, and late promoter, p670, and spliced using splice donors SD226, SD880 and/or SD3632 and splice acceptors SA409, SA526, SA742, SA2582, SA2709, SA3358 or SA5639. **B.** Focus of the splice pattern within the E6 and E7 region. Seen are the four different mRNAs.

However, one of the most important regulations in HPV16 gene expression is alternative splicing. As mentioned more than 20 mRNAs can be formed from the HPV16 genome through alternative splicing [66] and interestingly over 90% of the HPV16 genome consists of protein coding regions [19]. Since the HPV lacks its own splicing and polyadenylation machinery it therefore relies on the cellular system. None of the splice sites in HPV16 is consistent with the consensus sequences rendering them weak and prone to regulation by cis-elements and RBPs. There is always a competition between the splicing events

and splice sites, since one splice variant excludes the formation of the remaining mRNAs. To form the different mRNAs splice donors, or 5' splice sites, and splice acceptors, or 3' splice sites, are involved and today at least four splice donors (SD) and seven splice acceptors (SA) have been identified within the HPV16 genome. These are splice donors SD226, SD880, SD1302 and SD3632 and splice acceptors SA409, SA526, SA742, SA2582, SA2709, SA3358 and SA5639 [66]. In addition to these splice sites, two additional splice donors are also found within the E6 ORF, namely SD191 and SD221, however these are regarded as cryptic splice sites and their function remain fairly unknown [143]. To understand the regulation of all these splice sites would open up for potential antiviral targets and today several cis-elements and trans-acting factors regulating some of these splice sites are identified [66]. Which opens up for new potential antiviral targets.

Within the E6 and E7 region one splice donor (SD226) and three splice acceptors (SA409, SA526 and SA742) are found resulting in the formation of four different mRNAs in this region. Three of these are spliced and one still has the intron retained. The intron-retained mRNA is the only mRNA which can produce the E6 protein and is therefore of major importance. The three splice variants are all produced using the same splice donor but different splice acceptors, i.e SD226-SA409, SD226-SA526 and SD226-SA742. The SD226-SA409 splice variant has been found to be the major E7 producing mRNA and therefore it could be of major interest to identify the regulation of this mRNA [66]. Interestingly, this is also the most common of the four mRNA variants in HPV infected cells and it is regarded as a high-risk marker since this can only be formed in high-risk HPV types [21], [66]. Today one cis-element and four trans-acting factors regulating these splicing events have been identified and published. A silencer element located at the start of the E7 ORF binding both hnRNP A1 and hnRNP A2 has been found to affect the splicing outcome in this region. When hnRNP A1 binds it results in an upregulation of the intron-retained E6 mRNA and when hnRNP A2 binds it instead redirects the splicing machinery to SA742 [144]. Furthermore, hnRNP G [145] and hnRNP D40 [146] have additionally been found to inhibit splicing within the E6 and E7 region. Additional cis-elements and trans-acting factors have also been identified affecting the splicing events within the E6 and E7 region. These are found to mainly affect the formation of the E6*I/E7 mRNA as well as the SD226-SA742 splicing (unpublished, manuscripts I, II and III of this thesis). The SD226-SA526 splicing still remains a mystery, since the function of this mRNA remains unknown. What is known is that it forms an E6*II/E7 mRNA, however the expression level of this mRNA is fairly low in HPV infected cells [66]. Due to this it could be speculated if this is the result of splicing evolution or splicing noise [147].

Following the E6 and E7 region is the E1 and E2 region. In this region several splice donors and acceptors are located, namely SD880, SD1302, SD3632 and SA2582, SA2709, SA3358. Amongst the two first splice donors SD880 is much more extensively used compared to SD1302. This is used to

form both the E4 mRNA (SD880-SA3358) and the E2 mRNA (SD880-SA2709), however the usage of SD880 also needs to be regulated to allow for the formation of the intron-retained E1 mRNA. How SD880 usage is regulated is not well established, however, other regulatory elements/factors in this region have been identified [66]. hnRNP D40 was found to have an inhibitory effect on splicing and increasing the intron-retained E1 mRNA, however the cis-element involved in this process remains to be identified [146]. Additionally, hnRNP G was found to bind to a cis-element located just upstream of the E2 ORF and affecting splicing to SA2709 [145] and thereby increasing the formation of the major E2 protein producing mRNA. Regulation of SA3358 have also been identified. The usage of this splice site and SD880-SA3358 results in an E4 mRNA. Today three SR-proteins have been found which regulate usage of SA3358. These are SRSF1, SRSF3 and SRSF9 [66]. Within this region there is another splice acceptor with a relatively unknown function, SA2582. What is known is that in combination with SD880 it forms an mRNA called E1C, however if an E1C protein is expressed in HPV infected cells remains to be determined [100]. Interestingly, though, is that this mRNA is found in HPV infected cells and a connection between high-grade lesions and the expression of this mRNA has been made. It has been found that this mRNA as well as the E6*I mRNA increases in high-grade HPV lesions and therefore it would be of interest to further study the role of this mRNA [148].

To express the late genes there needs to be a readthrough of the early polyadenylation signal followed by splicing. The two splice sites which are used to form the L1 mRNA are SD3632 and SA5639. The SD3632 is located upstream of the pAE. However, to form the L2 mRNA this splicing event cannot occur since the L2 ORF would then be spliced out. Trans-acting factors involved in this splicing regulation has been determined to hnRNP D, DL, AB and A2/B1 which all have a negative impact on splicing by preventing splicing to the splice donor. hnRNP A1 and L have also been found to regulate SA5639 [66].

Since the HPV16 mRNAs are produced by the cellular RNA polymerase II (RNA pol II) it is also believed to be involved in the regulation of mRNA transcripts as a rate limiting factor. Histones covering the DNA is believed to slow down the RNA pol II and this is believed to affect the splice site usage [100]. Also, the secondary structure of the pre-mRNA can potentially mask or enhance cis-elements and slice sites. Additionally, RNA modifications by attaching molecular groups, like m6A, are believed to have an additive effect on the splicing outcome [66].

As discussed so far, the pathogenesis of HPV16 infections largely stem from the dysregulated expression of two viral oncogene products, the E6 and E7 proteins. These are produced from two different mRNAs, which are generated in a mutually exclusive manner and the understanding of the mechanism by which HPV16 E6 and E7 mRNA variants are regulated remain poor.

The regulation of the HPV16 mRNA splicing is complicated for several reasons. First, the genomic structure and regulation of expression of HPV16 genes

differ significantly from that of cellular genes. Furthermore, comprehensive and reliable information about the interaction between cis-RNA elements and trans-acting factors critical for alternative splicing regulation is still limited.

In this thesis, it was aimed to identify novel regulatory factors for HPV16 oncogene encoding mRNAs and gain insight into how high-risk HPVs have coevolved and adapted to the host system to expand the diversity of the viral transcriptome. The results obtained from studies in this thesis are adding to the information and the accumulating knowledge in the fields of RNA processing regulation and tumor-virus associated oncogenesis development.

HPV16 plasmids of distinct interest

Several HPV16 plasmid constructs have been extensively used in the papers included in this thesis. These have been used to study and map cis-elements regulating HPV16 alternative splicing as well as determine the effect of trans-acting factors on this splice pattern. The following plasmids pC97ELsL, pX478, pX550 and pHPV16AN are further described below.

pC97ELsL

The pC97ELsL plasmid is a subgenomic HPV16 plasmid (Figure 16). It contains all HPV16 genes and therefore also all HPV16 splice sites. In addition, it also contains both polyadenylation signals, i.e., pAE and pAL. The early promoter, p97, has been removed and replaced by a cytomegalovirus promoter (CMV) and part of the L1 gene is replaced by an internal ribosome entry site (IRES) and a secreted luciferase (sLuc) gene. By replacing the L1 gene with secreted luciferase, changes in the late gene expression can be measured by changes in sLuc and this can be measured in the growth medium of the cells. The plasmid has been indispensable to identify elements used in HPV16 gene regulation and mRNA processing. It has been used in all four manuscripts and papers. It has furthermore been an important foundation when constructing mutational plasmids, resulting in mutations of the crucial cis-element downstream of SA409 (Paper I), the GGG motifs located downstream of SD880 (Paper II), the janus-faced cis-element (Paper III) and the silencer element at the start of the E7 ORF (Paper IV). This followed the formation of five new plasmids, pC97M1M4, pC97ELsLAA, pC97ELsLGTG, pC97ELEh2M4, and pC97ELsLM1. It has moreover been used when constructing the immortalized keratinocyte cell line, JJM9721, which is further explained below.

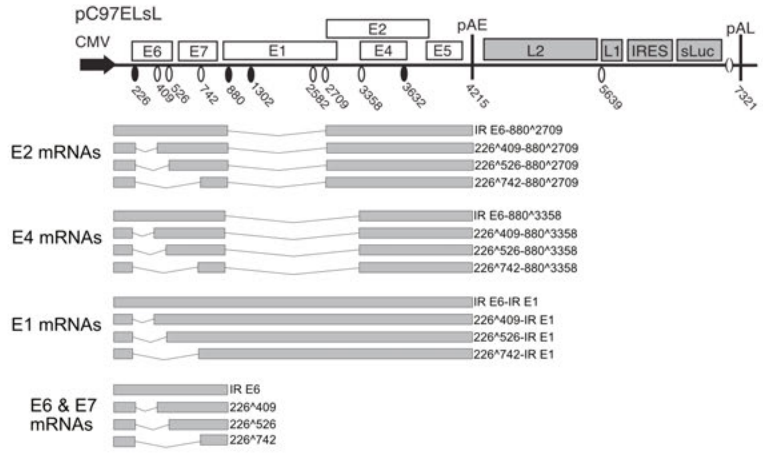

Figure 16. Schematic representation of the HPV16 pC97ELsL plasmid. Below the plasmid construct different HPV16 early mRNAs fromed from this plasmid are seen.

pX478

The pX478 plasmid is a short subgenomic HPV16 plasmid containing part of the E6 ORF (Figure 17). It starts at nucleotide 97 and as the name denotes end at nucleotide 478. Within this plasmid the identified cis-element regulating SA409 splicing is found (Paper I). The plasmid has been extensively used to study the alternative splicing event between SD226 and S409. Several plasmids carrying mutations in the identified cis-element have been constructed (Paper I). In these the cis-element is partly or completely mutated. The plasmid is a good tool when studying intron-retained E6 mRNA and SD226-SA409 splicing, since these are the two mRNAs expressed from this plasmid.

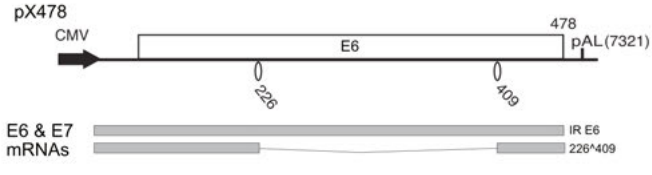

Figure 17. Schematic drawing of the HPV16 plasmid, pX478. Seen are the two mRNAs which can be formed.

pX550

The pX550 plasmid is a small subgenomic HPV16 plasmid containing almost the complete E6 ORF (Figure 18). Just as for pX478, this plasmid starts at nucleotide 97 but end at nucleotide 550, as the name implies and the expression is controlled from a CMV promoter. This plasmid has been an important

tool when studying the janus-faced cis-element, named ENH-2 (Paper III). Within this plasmid SD226, SA409 and SA526 are found which can result in the formation of three mRNAs, one intron-retained and two splice variants SD226-SA409 and SD226-SA526. It has mainly been used to study the alternative splicing event between SD226 and SA409.

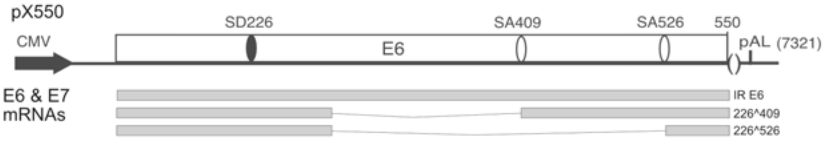

Figure 18. Schematic drawing of the HPV16 plasmid, pX550. Seen are the three mRNAs which can be formed from the plasmid.

pHPV16AN

The pHPV16AN is a large HPV16 plasmid containing the complete genome of HPV16 (Figure 19). Additional to the early and late regions this plasmid also contains the LCR region. Included in the plasmid are two LoxP sites and when co-transfecting the plasmid with pCAGGS-NLS-Cre, the plasmid recombinates between the two LoxP sites and the transfected plasmid now mimics an episomal HPV infection. Since the plasmid contains the entire HPV16 genome it therefore contains both promoters, p97 and p670, from which the gene expression occurs. Several mutations have been made in the plasmid construct (Paper I and Paper II), which results in destruction of the cis-element downstream of SA409 or the GGG-motif downstream of SD880. The new names of these plasmids are pHPV16ANM1M4 for the plasmid with mutations in the cis-element (Paper I) and pHPV16AN AA and GTG for mutations in the GGG-motif (Paper II).

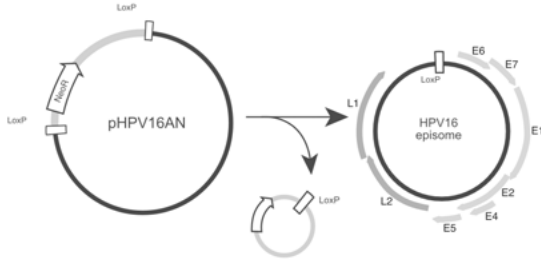

Figure 19. The pHPV16AN plasmid. Seen is the transfected form and the episomal form which is formed after cotransfection with a Cre-lox plasmid digesting the plasmid at the LoxP sites resulting in a plasmid mimicking the episomal HPV16 infection.

Aim of thesis

The overall aim of the thesis was to identify cis-acting RNA elements and trans-acting factors which regulate splicing of the HPV16 E6 and E7 oncogene mRNAs. This to get a better understanding of the HPV E6 and E7 splicing regulation.

Paper I:

The aim of this paper was to determine the effect the identified enhancer element on SD226-SA409 splicing and identify the trans-acting factors binding to the enhancer and regulating this splicing event.

Paper II:

The aim of this paper was to determine the effect of the GGG-cluster downstream of SD880 on E6/E7 splicing and SD880 usage.

Paper III:

The aim of this paper was to determine the effect of the identified cis-elements located between nt478-502 and between nt525-550 on E6 and E7 splicing.

Paper IV:

The aim of this paper was to determine the effect and sequence of the cis-element at the start of the E7 ORF and identify the trans-acting factors binding this element as well as the splicing outcomes of these.

Materials and methods

Cells

HeLa cells is an established cell line and was one of the first cell lines to be established. It originates from a woman, 31 years old, whom had developed adenocarcinoma in her cervix as a result of an HPV infection. The cells contain HPV18, which has integrated in the cellular genome. It is an adherent cell line [149].

SiHa cells originates from a 55-year-old woman diagnosed with squamous cell carcinoma of the cervix. The cells contain the HPV16 genome, which is integrated and it is estimated that 1-2 copies of the HPV16 genome is found in each cell. Like HeLa, SiHa cells are an adherent cell line [150].

293T cells originates from human embryonic kidney cells. The cell line contains the SV40 T-antigen. It is described as an epithelial like cell line and has a high transfection efficiency. It is further an adherent cell line, however, ever so loosely attached to the growth surface [151].

Cells used throughout the experiments are HeLa, SiHa and 293T cells. These are grown in Dulbecco's modified Eagle medium (DMEM) (HyClone) which is supplemented with 10% fetal bovine serum (HyClone) and 1% Penicillin-Streptomycin (Gibco).

Normal Human Foreskin Keratinocytes (nHFK, from ATCC, PCS-200-010) and JJM9721 cells (Paper I) were grown in EpiLife (Gibco) supplemented with 1% human keratinocyte growth supplement (HKGS, Gibco) and 0.2% Amphotericin/Gentamicin (Gibco). In short, JJM9721 cells is an in house established cell line immortalized using the pC97ELsL plasmid and DMEM selection according to Sherman and Schlegel [152].

Plasmids

Previously describe plasmids are pHPV16AN [119], pC97elsLuc [153], pBEL [119], pX765 [145], pX656 [146], pX478 [146], pCL086 [154], phnRNPK [155], pRBM15b [145], phnRNPM [145], pBCLAF1 [125], pRSVneo [153], pCAGGS-NLS-Cre [156], pTRAP150 [125].

The hnRNP H plasmid was constructed using polymerase chain reaction (PCR) followed by subcloning into the pCL086 backbone. In this plasmid the

hnRNP H expression is driven using a CMV promoter. As a template the pET-15B-hnRNP H plasmid was used. This was kindly given by Dr. D Black [157].

The construction of pX438, pX502, pX525, pX550, pX478AW+M, pX478AM+W, pX478M+M, pX478AM1, pX478AM2, pX478AM3, pX478AM4, pX478AM5, pX478sM1, pX478sM4, pX478midM, pC97M1M4, pHPV16ANM1M4, pC16HAE6E7F, pC6bHAE6E7F, pC11HAE6E7F, pC31HAE6E7F, pC18HAE6E7F, pC45HAE6E7F, pC.Enh1.16to6b, pC.Enh1.16to31, pC.Enh1.16to18, pC.Enh1.16to45 are described in paper I.

The construction of pX1200GTG, pX1200M3, pX1200M8, pX920, pX920MG123, pX920AA, pX920TC, pHPV16ANAA, pHPV16ANGTG, pC97ELsLAA, pC97ELsLGTG are described in paper II.

The construction of pX534, pX541, pX550comp, pX550C1, pX550C2, pX550C3, pX550M1, pX550M2, pX550M3, pX550M4, pX550M5, pX550M6, pX550M7, pX550M8, pC97ELsLEh2M4, pX581Eh2M4, pX604Eh2M4, pX656Eh2M4, pX856Feh2M4, pX502MA, pX502CC, pX502m63, pX502m63-1, pX502m63-2, pX502m63-3, pX502TT, pX502TG, pX502GT, pX502AG, pX502TC, pX502GC are described in paper III.

The construction of pX1200, pX1060, pX960, pX865AF, pX644, pX616, pX604, pM604A, pX581, pX579, pX556AF, pA1NA2C, pA2NA1C are described in paper IV.

To construct the high-risk HPV E6/E7 expressing plasmids separate primers for each HPV type were constructed. These amplified the region from the ATG of E6 to the stop codon of E7 of each high-risk HPV type, respectively. On the 5'-primer, a PteI site was included and on the 3'-primer an XhoI site which were used to clone the fragments into the correct backbone. The requirements for the 5'-primer were the following, a PteI site followed by 5 nucleotides of the HPV sequence leading up to the ATG of E6 followed by 20-25 nucleotides of the HPV sequence. The 3'-primer had the following requirements, an XhoI site at the 5'-end followed by the E7 stop codon and 20-25 nucleotides of the HPV sequence. The melting temperature of the primer should be around 60ºC. The E6E7 fragments were amplified using PCR and thereafter TA-cloned into a TA-vector (Thermo Fisher Scientific) and transformed. Following colony formation an overnight culture and miniprep was done on the colonies. The miniprep was digested using PteI and XhoI and subcloned onto the pCL086 vector. The E6E7 expression for each respective HPV type was now driven by a CMV promoter.

Transfections

HeLa cells were transfected using Turbofect (Thermo Fisher Scientific). This was done according to the manufacturer's protocol. In short, plasmids DNA

and DMEM without serum were mixed. To the mixture Turbofect was added with a ratio of 1:2 of plasmids DNA in µg to Turbofect in µl. The mixture was incubated for 15-20 min at room temperature before being added dropwise to the subconfluent cells.

HeLa, SiHa and 293T cells were transfected using Fugene6 (Promega). In short, DMEM without serum and plasmid DNA was mixed. Fugene6 was added to the mixture with a ratio of 1:2 or 1:3, e.g. to 1µg DNA to 2 or 3µl of Fugene6. The mixture was incubated at room temperature for 15 minutes before being added dropwise to subconfluent cells.

To construct the JJM9721 cell line Normal Human Foreskin Keratinocytes (nHFK, PCS-200-010, ATCC) were transfected using Xfect (TaKaRa) according to manufacturer's protocol. In short, plasmid DNA and Xfect were mixed with a ratio of 1:0.3 in the Xfect reaction buffer. After adding the reagents, the mixture was mixed by vortexing and incubated for 10 min at room temperature before briefly being centrifuged and thereafter dropwise added to subconfluent cells. 6 hrs post transfection the culture medium was changed.

siRNA transfections

SiHa and JJM9721 cells were transfected with ON-TARGET SMARTpool siRNA from Dharmacon using Dharmafect 1 (Dharmacon), according to the manufacturer's protocol. In short, siRNA was diluted to a final concentration of 50nM or 100nM in 200µl serum free medium. This was mixed with an additional 200µl serum free medium containing 6µl Dharmafect1. The mixture was incubated for 20min at room temperature before addition to 6-well plates containing subconfluent cells growing in 1600ul medium. siRNAs used were (si-scr) was siGENOME Control pool non-targeting #2 (D-001206-14-20, Dharmacon™). siRNA to TRAP150 was ON-TARGETplus Human THRAP3 (GeneID: 9967) siRNA-SMARTpool (L-019907-00-0020) and siRNA against BCLAF1 was ON-TARGETplus Human BCLAF1 (GeneID: 9774) siRNA SMARTpool (L-020734-00-0005).

Nuclear and cytoplasmic fraction extraction

To extract nuclear and cytoplasmic fractions the NE-PER Nuclear and Cytoplasmic Extraction Reagents (Thermo Fisher Scientific) kit was used. The instructions from the manufacturer were followed. To shortly describe, 60mm plates with HeLa cells were transfected and 24hpt the cells were harvested by scraping and pelleting. The cell pellets were resuspended in CERI buffer supplemented with protease inhibitor (Sigma Aldrich) and RNase inhibitor (Ribolock, Thermo Fisher Scientific). After resuspension the cells were vortexed

and incubated on ice for 10min followed by addition of ice cold CERII buffer. The samples were thereafter vortexed on the highest speed and incubated on ice for a minute. The samples were centrifuged at maximum speed for 5 minutes to obtain the supernatant consisting of the cytoplasmic extraction. The nuclear pellet was washed using ice cold PBS and collected as nuclear extracts. From the nuclear and cytoplasmic extractions RNA was extracted according to the protocol below. The integrity of fractionation was evaluated by RT-PCR, using primers for spliced and unspliced actin.

RNA extraction, RT-PCR and Real time qPCR.

Total RNA was extracted by lysing cells in TRI Reagent (Sigma Aldrich) and using the Direct-zol RNA Miniprep kit (ZYMO Research), according to manufacturer's protocol. To increase the RNA quality during extraction, chloroform was added to the samples prior to extraction. The chloroform-TRI reagent suspension was shaken and incubated before centrifugation at 12,000g for 15 minutes to isolate the two phases. The upper phase was used for RNA extraction. The RNA quality was routinely evaluated by the ratio between 28S and 18s rRNA determined by RNA electrophoresis on an agarose gel. Additionally, the ratio of 260/280 and 260/230nm using a Nanodrop (Thermo Scientific) was looked upon.

To synthesize complementary DNA (cDNA) the M-MLV reverse transcriptase kit (Invitrogen) was used. To this, random hexamers (Invitrogen) were added and the manufacturers protocol followed. The cDNA was synthesized in a volume of 20µl and both positive (+RT) and negative (-RT) reverse transcriptase reactions were made as control. To the -RT no reverse transcriptase was added. Both the +RT and -RT reactions were used in the PCR reaction to evaluated the contamination of transfected DNA contamination in RNA samples that may result in preposterous results in determining the intron-retained viral mRNA detection.

The PCR was made using DreamTaq polymerase (Thermo Fisher Scientific) and the adjoining protocol was followed. In summary, DreamTaq buffer and polymerase, ddH_2O, forward and revers primers (Eurofins Genomic) and dNTP (Thermo Fisher Scientific) were mixed to a total volume of 25µl for each sample and 1µl cDNA was added prior to PCR. The analysis of the PCR was made on a 2% agarose gel (SeaKem).

qPCR was done using the SsoAdvanced SYBR Green Supermix (Bio-Rad) according to the manufacturer's instructions. 1µl cDNA and a total volume of 10µl was used in the analysis in a 384-well plate with the StepOnePlus cycler (Appliced Biosystems). Expression levels of each mRNA was determined based on the threshold cycle (Ct) and relative expression levels (fold) of each

sample was calculated using $2^{-\Delta\Delta Ct}$. To normalize each sample the expression levels GAPDH or actin were used.

Western blot

Total protein was extracted from cells 48-72hpt by lysing cells in radioimmunoprecipitation assay (RIPA) buffer (50mM Tris pH 8.0, 150mM NaCl, 1.0% NP-40, 0.5% Sodium deoxycholate, 0.1% SDS, 2x Protease inhibitor and 1mM DTT). The cells were incubated at end-over rotation for 30min and 4ºC. To remove cell debri the lysate was centrifuged at max speed for 15min and supernatant collected. Proteins were denatured using Laemmli buffer and boiling. After denaturing the samples, they were run on an SDS-PAGE gel and transferred to a nitrocellulose membrane. This was blocked using 5% non-fat dry milk in PBS containing 0.1% Tween-20 or Intercept (PBS) blocking buffer (LICOR) following incubation with primary antibody. Thereafter incubation using secondary antibody was done. Secondary antibodies were either HRP-conjugated or IR-conjugated for Super Signal West Femto chemiluminescence substrate (Thermo Fisher Scientific) and ChemiDoc MP (Bio-Rad) detection or Odyssey CLx Imager (LICOR) detection, respectively.

ssRNA oligo pull down

Prior to oligo pull down HeLa cells were fractionated into cytoplasmic and nuclear fractions according to the protocol from Kajitani et al. [158]. In short, HeLa cells were lysed using buffer I (10mM Hepes pH 7.9, 10mM KCl, 0.1mM EGTA, 1mM DTT, 0.2% NP-40 and protease inhibitor). The lysate was centrifuged and supernatant stored as cytoplasmic extract. The nuclear pellet was resuspended and lysed using buffer II (10mM Hepes pH 7.9, 400mM NaCl, 1mM EGTA, 1mM EDTA, 1mM DTT, protease inhibitor) to obtain nuclear proteins.

Dynabeads M280 (Invitrogen) were incubated with biotinylated ssRNA oligos (Sigma). Following incubation these were mixed with nuclear extract and incubated at room-temperature for 40 minutes before washing using ice-cold binding buffer (10mM Hepes pH 7.9, 2.5mM MgCl2, 0.5% Triton-X, 150mM NaCl, 1mM DTT). The proteins were eluted using 2x Laemmli buffer and boiling for 5 minutes. The samples were run on and SDS-PAGE gel followed by Western blotting.

Co-immunoprecipitation of protein complexes

293T cells were transfected and harvested by scraping 48hpt. The cell pellets were lysed in 1ml mild lysis buffer (20mM Tris HCl pH 8.0, 137mM NaCl, 1% NP-40, 2mM EDTA and 2x Protease inhibitor). And the lysis mixture was incubated at end-over rotation at 4ºC for 20min followed by centrifugation to pellet debri. The cell lysate was precleared using 10µl Protein G Dynabeads (Thermo Fisher Scientific) at 4ºC and end-over rotation for 30min. Protein G Dynabeads were washed and incubated with 1µg antibody followed by three washings. 5 percent of the cell extract was taken as input sample and the rest was divided into control and pull-down sample. The bead-antibody-cell lysate mixture was incubated at 4ºC overnight followed by washing for six times. The samples were eluted using 2x Laemelli buffer and an SDS-PAGE and Western blotting was done to analyze the samples.

RNA immunoprecipitation of RNA-protein complexes.

293T cells were transfected and harvested 48h post-transfection. Following harvest, the cells were resuspended in 1ml mild lysis buffer (20mM Tris HCl pH 8.0, 137mM NaCl, 1% IGEPAL CA-630, 2mM EDTA, RNase inhibitor and 2x Protease inhibitor). A 19-gauge needle was used to sheer the DNA and obtain a better lysate. The lysis step was incubated at 4ºC for 30min under constant rotation. After lysis, the samples were centrifuged and pre-cleared using 25µl Protein G Dynabeads (Thermo Fisher Scientific). Five percent were taken as input samples and 500ul of each lysate was mixed with 1µg of antibody under constant rotation for 2h at 4ºC. The samples were thereafter incubated with 50µl pre-washed Protein G Dynabeads at 4ºC for 2h before washing. The beads were washed using wash buffer (20mM Tris HCl pH 8.0, 137mM NaCl, 1% IGEPAL CA 630, 2mM EDTA). The pull-downed RNA was extracted with TriReagent and DNasel (Sigma). RNA was precipitated using phenol-chloroform (Sigma) and ethanol. The samples were resuspended in 22µl ddH2O and 10µl was reverse transcribed into cDNA using the M-MLV reverse transcriptase kit (Invitrogen) and random hexamers (Thermo Fisher Scientific) according to manufacturer's instructions. 1µl of cDNA was used in PCR to detect binding to HPV16 mRNA.

Water soluble tetrasodium (WST)-1 cell viability assay

Cells were seeded in 96-well plates. Depending on the cell type different cell numbers were seeded. 24h post seeding, cells were transfected with control (pUC19 or siCtrl) or increasing amounts of pTRAP150 or siTRAP150 and/or siBCLAF1, according to the transfection conditions mentioned above. The

cell viability was measured 24-, 48- and 72-hours post transfection, according to the manufacturers protocol. In short, two hours prior to measuring viability the medium was change to fresh medium. Following these two hours the WST-1 assay reagent (Roche) was added to each well. Control wells, containing only medium, were also prepared and the end concentration of the WST-1 reagent was 10%. Following addition of the reagent cell viability was measured at 30min, 1, 2, 3 and 4h post addition. In between the measurements cells were incubated at 37ºC and 5% CO_2. The formation of formazan product was measured in a TriStar Plate Reader (Berthold Tech) at 450nm.

Immortalization of nHFK using HPV16 DNA.

To immortalize Normal Human Foreskin Keratinocytes (nHFKs) a method by Sherman and Schlegel [152] was used. In this method cells are selected based on calcium- and serum-dependent differentiation and only cells which are resistant to this will survive. The cells were selected by transfecting subgenomic HPV16 plasmids. Cells at subconfluency were transfected using either pC97ELsL or pC97M1M4 with Xfect transfection reagent. These cells were growing on 100mm plates in EpiLife medium and were at the time of transfection at passage 8. The day after transfection cells reached confluency and were transferred onto 150mm plates and again grown till confluency. Once confluency was reached medium was changed to DMEM (HyClone) containing 10% bovine calf serum (HyClone) and 1% Penicillin-Streptomycin (Gibco). This selected the cells and only differentiation resistant cells will survive. The selection preceded for eight weeks with medium changed twice per week. Following selection only pC97ELsL cells had survived. These cells were transferred back into EpiLife medium and a population doubling curve was made by growing the cells alongside untransfected nHFKs. At this point the selected cells had reached passage 15 and nHFKs passage 4. Population doubling level (PDL) fo the two cell types was calculated using PDL = 3.32(log(total viable cells at harvest/total viable cells at seed)). Cells were also harvested at passages 0, 2, 4 and 9 to determine hTERT expression using RT-qPCR. The pC97ELsL immortalized cell line was named JJM9721. It was expanded and stocks frozen in liquid nitrogen.

A similar experiment was made using the pHPV16AN plasmids. In short, subconfluent nHFKs on 100mm plates were transfected with the following plasmids: pUC19, RSVneo, pHPV16AN, pHPV16ANM1M4, pHPV16AN+ pCAGGS-NLS-Cre or pHPV16ANM1M4+ pCAGGS-NLS-Cre, resectively. At time of transfection, they were on passage 5. Following transfection, the cells were split 1:4 onto new 100mm plates and selected using Neomycin (G418) at a concentration of 40µg/ml. This selection preceded for three days following another three days of selection at 10µg/ml. Post-selection, the cells

were expanded and stocks frozen in liquid nitrogen. Cells were no at passage 8. The cells were defrosted and recovered on 100mm plates before seeded in 6-well plates and new 100mm plates. Cells growing in 6-well plates were selected using DMEM supplemented with 10% BCS. Following 11 weeks of selection colony formation was looked upon by fixing cells and staining using crystal violet solution (0.05% w7v Crystal Violet (Sigma), 1% Formaldehyde, 1% Methanol and PBS) for 20 min. The cells in 100mm plates were used for hTERT analysis at passage 0, 2 and 4.

Quantitation

To quantitate and determine band intensity of RT-PCR and Western blot images. Image Lab 6.0.1 was used.

Statistical analysis.

Student t-tests (unpaired, two-tailed test) and quantitative analysis was achieved using GraphPad Prism 9. All experiments were carried out at with three independent experiments and reproducible results were obtained at each occasion.

Predictions of splice sites withing the E6 and E7 ORFs of high-risk HPVs

To predict the splice sites of high-risk HPVs an *in silico* analysis was used. In this, the different HPV types were aligned with either HPV16 or HPV18 depending on which it shared the most relatedness with. To try to identify the splice sites the consensus sequences for the donor and acceptor sites were also looked for, (C/A)AGGU_(A/G)AGU and (C/U)AG_, respectively.

Results

Paper I. A novel HPV16 splicing enhancer critical for viral oncogene expression and cell immortalization.

The HPV E6 and the E7 proteins are well known oncogenes and are associated with cancer formation in high-risk HPV infected cells [45]. It is known that the two genes are expressed as a bicistronic pre-mRNA which thereafter matures into one of four potential mRNAs. These mRNAs are produced in a mutually exclusive manner and the regulation of these is not well known. What is known is that there must be a balance between the intron-retained E6 mRNA and the mRNA spliced between SD226-SA409, which is the major producer of E7 protein [66]. However, how this balance is maintained and how the splicing regulation of SD226-SA409 is regulated is today not well-know and therefore this article aims to understand this process. In this article a critical cis-element regulating SD226-SA409 splicing was identified as well as novel RBPs in HPV16 splice site regulation.

A cis-element located downstream of SA409 is critical for SD226-SA409 splicing

The E6 and E7 expression of HPV16 is controlled from the early promoter and regulated through alternative splicing. The intron-retained E6 producing mRNA and the E6*I/E7 mRNA, spliced between SD226-SA409 are produced in an exclusive manner, where the formation of one excludes the formation of the other [66]. How the regulation of SD226-SA409 splicing is regulated is disperse and it was therefore aimed to highlight and identify the splicing regulation forming this mRNA. In order to identify cis-elements regulating SD226-SA409 splicing, several plasmids containing increasing lengths of the E6, E7 and E1 ORF were constructed to detect changes in the E6 and E7 splice pattern (Figure 20 A, B and C). These deletion plasmids started at the same nucleotide, however ended at different nucleotides, continuously increasing the length of the E6, E7 and E1 region of HPV16. A strong enhancer sequence was identified through RT-PCR of these plasmids. It is located downstream of SA409 within in the 40 nucleotides between nt438 and nt478 and increased splicing to SA409 by almost 80% (Figure 20 D and E). Other cis-elements

were also identified, however not looked upon in this paper. The analyses of these are found in paper III and IV.

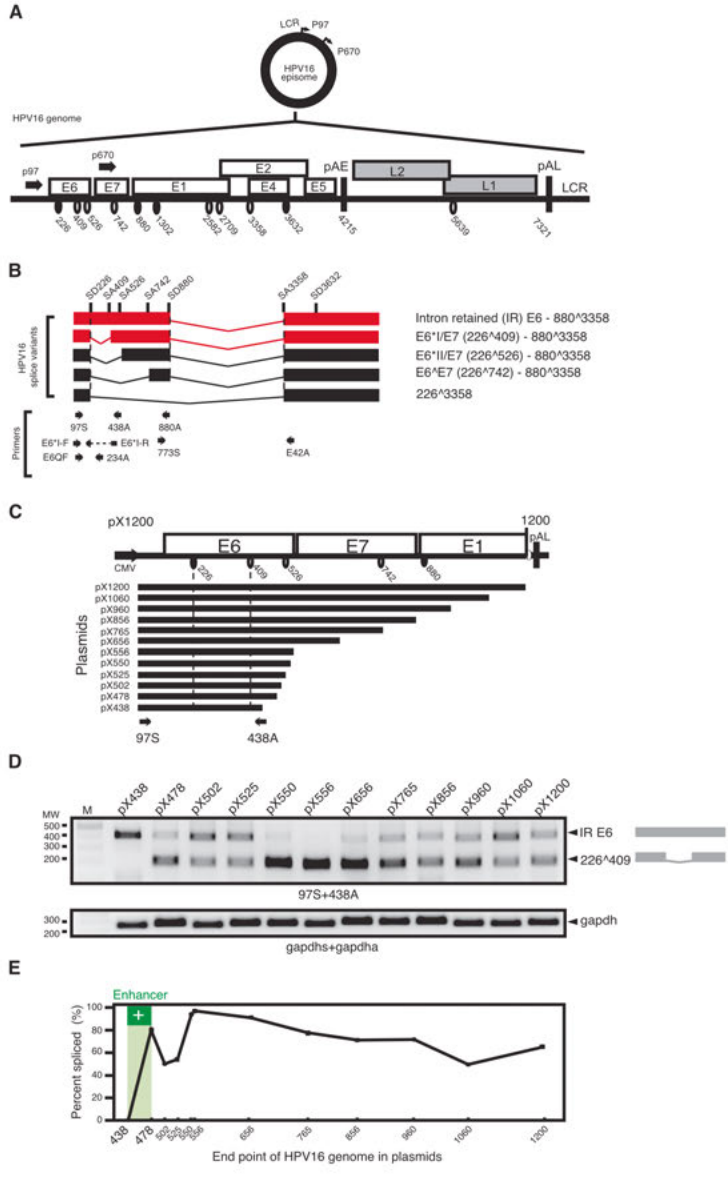

Figure 20. Deletion plasmids used to map cis-elements regulating SD226-SA409 splicing. **A.** Schematic representation of the HPV16 genome. **B.** mRNAs which can be formed by splicing within the E6 and E7 region as well as splicing using downstream splice donors and acceptors. **C.** Shematic representation of the constructed deletion plasmids of HPV16 consisting of parts of the E6, E7 and E1 ORFs. **D.** RT-PCR results of the plasmids in C. **E.** Plotting of the percent spliced mRNA (SD226-SA409) forming from these plasmids.

The cis-element consist of a bipartite perfect repeat and is crucial for E6*I/E7 mRNA formation and E7 protein production.

The sequence between nucleotide 438 and 478 was looked upon and a sequence consisting of AAAAGCAAAGA was found and repeated twice. Since this did not seem uncoincidental one or both of the 11 nucleotide repeats were mutated and analyzed using RT-PCR (Figure 21 A, B and C). When one or both of the repeats were mutated splicing to SA409 was lost in the pX478 construct. To establish if all 22 nucleotides were important for the splicing event point mutations of two or three nucleotides in both repeats were constructed and analyzed by RT-PCR. In all constructs splicing to SA409 was lost, proving an involvement of all 22 nucleotides in the splicing event. In a larger plasmid construct, pC97ELsL, a more subtle mutation was introduced in the identified enhancer element which also abolished splicing to SA409 (Figure 21 D and E). This plasmid was named pC97M1M4 and when looking at the E6 and E7 protein production by Western blot it was seen that the E7 protein expression was strongly reduced compared to the wild-type (Figure 21 F).

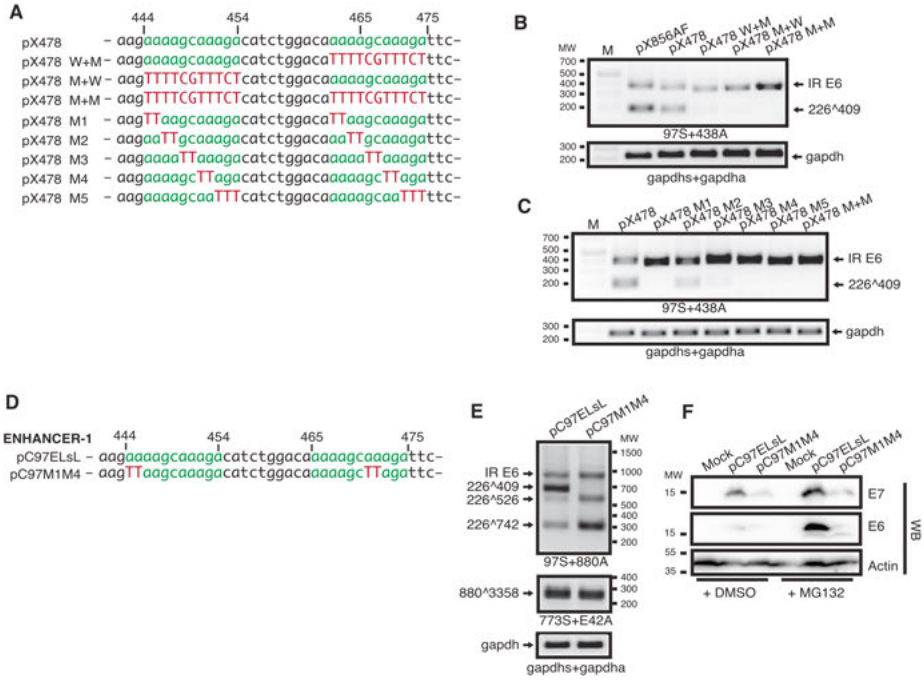

Figure 21. Identification of the cis-element regulating SD226-SA409 splicing. **A.** The mutations constructed in the cis-element in the pX478 plasmid. **B and C.** RT-PCR results of the mutations displayed in A. **D.** The wild-type and mutant sequence in pC97ELsL and pC97M1M4. **E.** RT-PCR of the two plasmids from D. **F.** Western blot looking at the E6 and E7 expression from the two plasmids in D.

The trans-acting factor responsible for SD226-SA409 splicing is identified as TRAP150/THRAP3

To identify the trans-acting factor which binds to the identified cis-element, an RNA oligo pull-down was done in which TRAP150/THRAP3 was identified (Figure 22 A and B). When TRAP150 was overexpressed in transfections with pC97ELsL or the M1M4 mutant and analyzed using RT-PCR it was seen to only increase SD226-SA409 splicing in the wild-type construct and not in the mutant (Figure 22 C). This increase in SD226-SA409 splicing also resulted in an increased E7 protein production (Figure 22 D), further emphasizing the importance of the cis-element and now identified trans-acting factor on SD226-SA409 splicing.

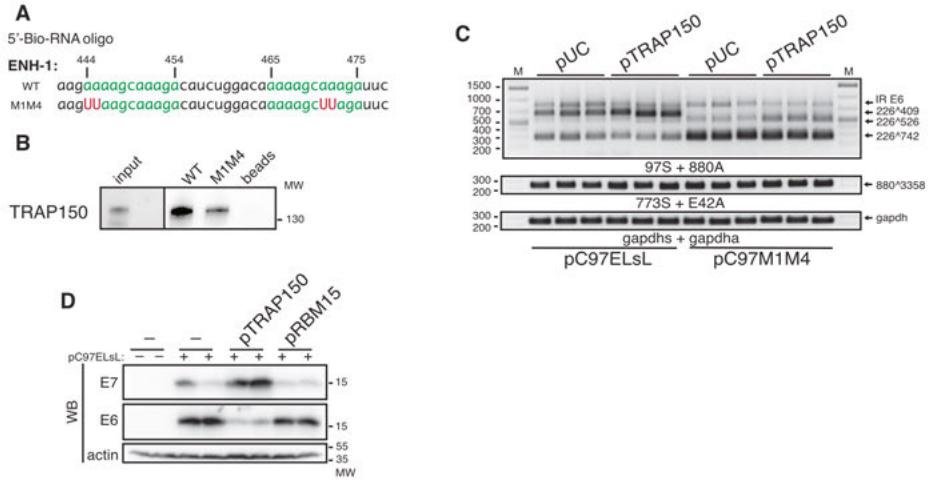

Figure 22. Identification of the trans-factor. **A.** The biotinylated RNA oligos used to detect protein binding to the cis-element. **B.** Western blot analysis of the RNA pull-down analysis using oligos in A. **C.** RT-PCR on TRAP150 overexpression on pC97ELsL or pC97M1M4. **D.** Western blot analysis of TRAP150 and RBM15 overexpression on pC97ELsL. The E6 and E7 protein levels were determined.

TRAP150 interacts with the 3'-splice factor U2AF65 and HPV16 mRNA.

To investigate if TRAP150 interacts with wild type E6/E7 mRNA but not mRNA expressed from the M1M4 mutant, an RNA in vitro pull-down assay was done. In this it was confirmed that TRAP150 interacts with wild-type E6/E7 mRNA but cannot interact with mRNA from the mutant (Figure 23 A). Since TRAP150 was shown to regulate splicing to the 3'-splice site SA409 and previous interactions with the 3'-splice factor U2AF65 had previously been reported by our group [65] we wanted to determine if TRAP150 and

U2AF65 interacted with each other in this model as well. Co-immunoprecipitation experiments with immunoprecipitation of either FLAG-tagged TRAP150 or endogenous U2AF65 showed an interaction between the two proteins (Figure 23 B and C). These and the above results rose to a proposed model (Figure 23 D). A cis-element located downstream of SA409 pulls down the trans-acting factor TRAP150, which in turn interacts with the 3'-splice factor U2AF65 and results in splicing to SA409. The interaction between the identified novel cis-element and TRAP150 is indispensable to balance the inhibition of apoptosis and the upregulation of proliferation in HPV infected cells. To test this proposed model, the significance of the novel cis-element on the viral pathogenesis in the development of immortalization of human primary keratinocytes is determined next.

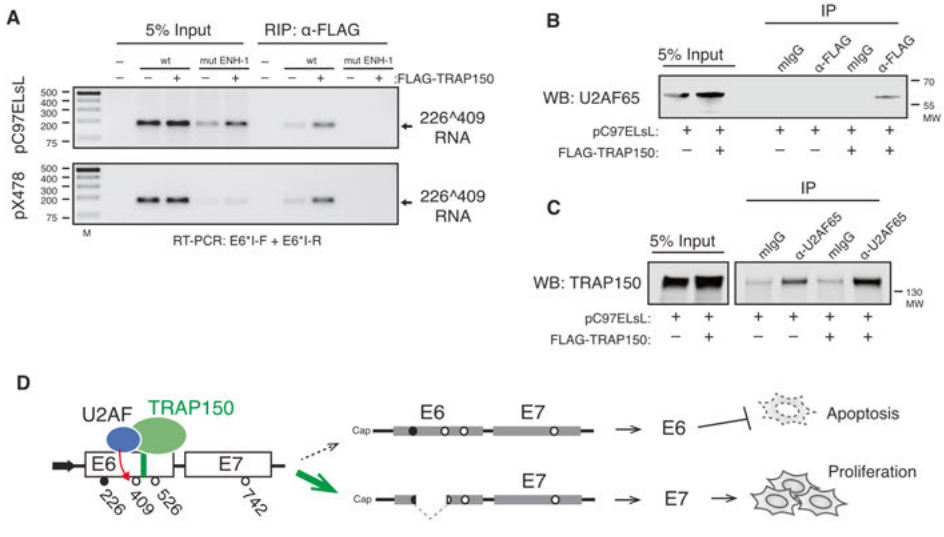

Figure 23. TRAP150 interacts with HPV16 mRNA and the 3'-splice factor U2AF65. **A.** RIP analysis of TRAP150 binding to HPV16 mRNA. **B and C.** Co-immuno precipitation of TRAP150 and U2AF65. **D.** A proposed model for the cis-element, TRAP150 and U2AF65 function on SD226-SA409 splicing.

The cis-RNA element is crucial for immortalization of primary keratinocytes.

The wild-type pC97ELsL and mutant M1M4 plasmids were transfected into subconfluent normal human foreskin keratinocytes (nHFKs). Once these cells reached confluency the medium was changed from the normal primary keratinocyte medium (Epilife) to DMEM supplemented with fetal bovine serum (Figure 24 A). This supplemented DMEM medium worked as a selection agent since the calcium in the medium will force the cells to differentiate and only cells resistant to differentiation will survive selection. After eight weeks

of selection the control cells had died and the selection was deemed successful. Interestingly, and surprisingly, only the wild-type transfected cells survived and were able to form colonies on the plate (Figure 24 B). The mutant cells did not survive the selection. The immortalized, wild-type transfected cells were transferred back into the primary cell medium, EpiLife, and the hTERT expression was looked upon and compared to nHFK. As is seen, the hTERT expression in the immortalized cells (JJM) increases for each passage while it for the nHFK decreases (Figure 23 C). The increase in hTERT in the immortalized cells is explained by HPV16 E6 protein expression.

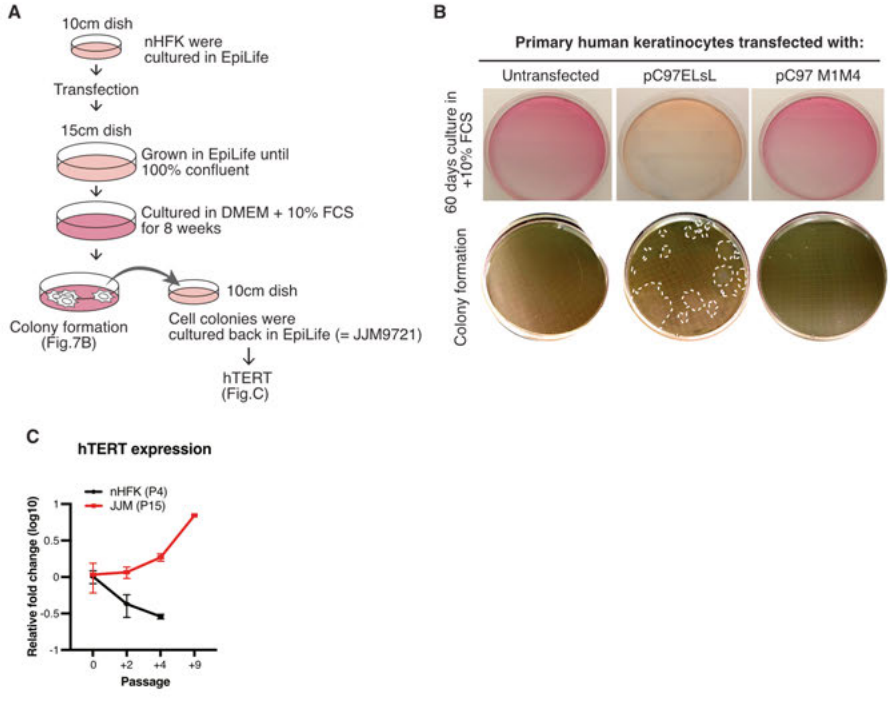

Figure 24. Immortalization assay of normal human foreskin keratinocytes. **A.** A schematic overview of the experiment layout. **B.** The results of the colony formation once the experiment was stopped. C. hTERT using qRT-PCR expression from the immortalized cells (JJM) and nHFK.

Conclusion

In this manuscript a novel cis-element located downstream of SA409 was found to positively regulate splicing to this acceptor. The cis-element was mapped to a perfect bi-partite repeat between nucleotides 444-475. To the cis-element the trans-acting factor TRAP150 binds and regulates the splicing machinery. Interaction with the 3'-splice factor U2AF65 was found and further emphasized the importance for TRAP150 in this splice event. In an

immortalization assay using wild-type and mutated enhancer constructs it was found that expression of only the wild-type and not mutant HPV16 mRNA could immortalize the nHFK cells further emphasizing the crucial importance of the identified cis-element.

Paper II. hnRNP H controls alternative splicing of human papillomavirus type 16 E1, E6, E7 and E6^E7 mRNAs via GGG motifs.

Downstream of the splice acceptor SA742 and the E7 ORF is a 78 nucleotide stretch which have proven important for HPV16 mRNA stability. If these are removed the E7 mRNA becomes very unstable and is degraded fast [159]. Within this nucleotide stretch several important elements are found which are believed to be of importance. One of these is the splice donor 880 (SD880). Also, a cluster of GGG motifs are located in this region and found directly downstream of SD880.

hnRNP H is an RNA binding protein belonging to the hnRNP family and the binding motif for this protein is determined to GGG or GGGG motifs. Since it has previously been shown that GGG-motifs are often found in close connection to 5'-splice sites [115] the aim of this article was to investigate the GGG-motifs and hnRNP H function on SD880 usage as well as splicing regulation in the upstream E6 and E7 ORFs.

A GGG/G-motif directly downstream of HPV16 splice donor SD880 is well conserved in high-risk HPVs.

GGG motifs in mRNA sequences have previously been identified as splicing regulators. Since these GGG motifs also exist in the HPV16 genome we aimed to identify these and see if clusters of GGGs were located in the HPV16 genome. To do so, a sequence analysis looking at the number of GGG-motifs per 200 nucleotides was done (Figure 25 A and B). In this analysis several clusters were identified and one of them was located directly downstream of SD880 in the E1 ORF. When comparing this sequence with the same sequence of other high-risk and low-risk HPVs it was seen that a GGG cluster of three GGGs or GGGGs were found in high-risk HPVs, however, this was not conserved in low-risk HPVs where only two GGG/GGGGs were located (Figure 25 C). This pointed to a regulatory role for the GGG-motifs in high-risk HPVs and it was therefore further investigated.

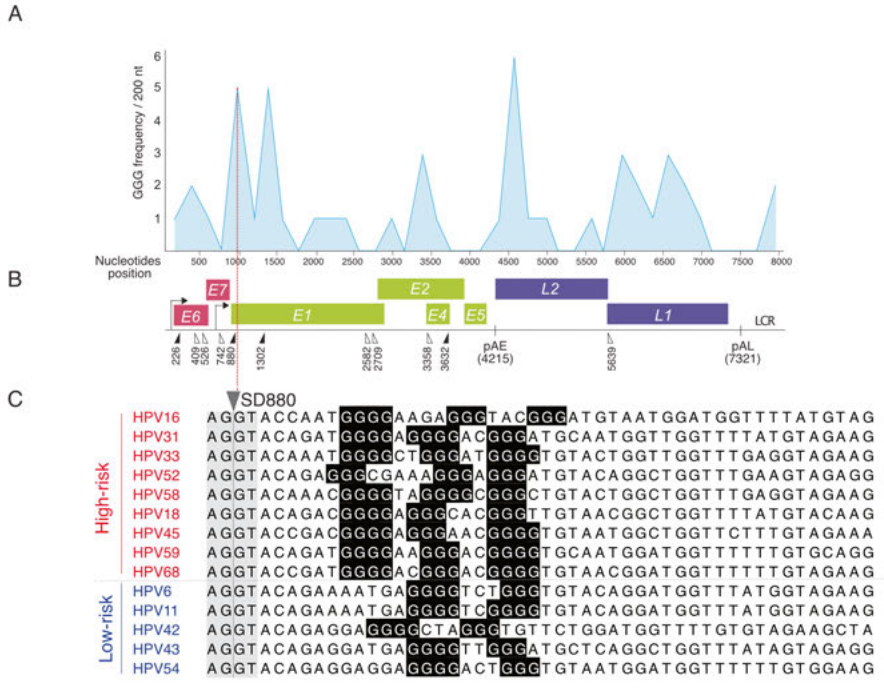

Figure 25. Analysis of GGG-motifs in the HPV16 region and comparison with other HPV types. **A.** A graph showing the number of GGG motifs per 200 nucleotides in the HPV16 genome. **B.** A schematic drawing of the HPV16 genome. An overlap between a GGG-cluster and SD880 is indicated. **C.** Sequence comparison of the GGG-cluster downstream of SD880 in different HPV types. High-risk in red and low-risk in blue.

Splicing of E6 and E7 mRNAs and usage of SD226 is highly affected by mutations in the GGG-motifs.

To determine the effect of the GGG-cluster downstream of SD880 two mutations disrupting the GGGs were constructed. In one, which is named AA, the protein coding sequence remained intact but the GGGs were disrupted and in the other mutant the GGGs were mutated to GTGs disrupting both the RNA sequence as well as changing the protein coding sequence (Figure 26 A). These two mutations were constructed in the pC97ELsL plasmid and named pC97ELsL AA and pC97ELsL GTG, respectively. When investigating changes in the splice pattern using RT-PCR several changes were detected. In the E6 and E7 region splicing between SD226-SA742 were almost completely lost in the two mutants, proving an important effect of the GGG-cluster on this splice pattern (Figure 26 B and C). Splicing from SD226 to SA409 was also reduced in these mutant plasmids and the intron-retained E6 mRNA was markedly increased (Figure 26 B and C). These results indicated that the GGG-cluster had an effect on SD226 function as well. Since splicing to the

splice acceptors in the E6 and E7 region was reduced, it was also aimed to investigate if this was redirected to downstream splice sites located in other parts of the HPV16 genome. This analysis showed that splicing to downstream acceptors increased when the GGG-cluster was mutated. Namely splicing increased from SD226 to SA2709 or SA3358, resulting in an increase of the E2 and E4 mRNAs, respectively.

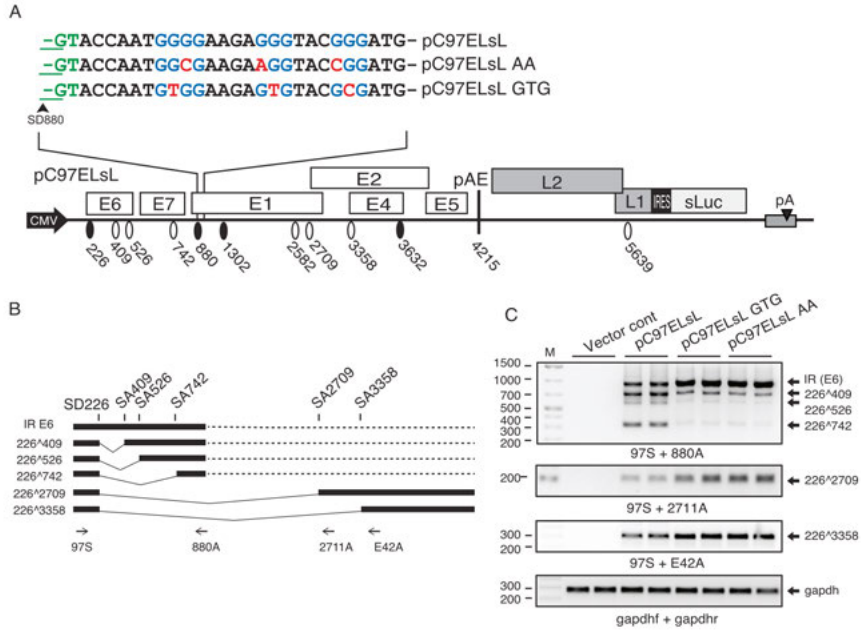

Figure 26. Mutations in of the GGG-motifs affect splicing in the E6 and E7 ORFs and redirects splicing to downstream acceptors. **A.** A schematic drawing of the pC97ELsL plasmid. Highlighted are the GGG motifs downstream of SD880 and the mutations of these are shown. **B.** The mRNAs which can be formed from the early region through alternative splicing events involving SD226. **C.** RT-PCR results on pC97ELsL and the two mutants GTG and AA.

The GGG/G-motifs directly downstream of SD880 are inapt for splicing from SD880 to downstream SAs when the E6 and E7 region is not present.

Since the GGG-motifs are located directly downstream of SD880 it was also of great value to determine the effect of the GGG cluster on SD880 usage. To do so, two different plasmid constructs were used, pC97ELsL and pBELsL. The pBELsL plasmid is exactly the same as the pC97ELsL plasmid, except that it is lacking the E6 and E7 region of HPV16 and therefore the effect on SD880 usage without the upstream E6 and E7 region could be looked upon. SD880 splicing to two downstream splice acceptors (SA2709 and SA3358)

was looked upon using RT-PCR (Figure 27). The SD880-SA2709 splicing is marginally inhibited in the pC97ELsL construct when the GGG-motifs are mutated and no inhibitory splicing effect is seen on the pBELsL constructs. However, when looking at SD880-SA3358 a more enhanced effect is seen. In the pC97ELsL constructs splicing between SD880-SA3358 was markedly reduced when the GGG-cluster was destroyed, however, no effect on splicing was seen on the pBELsL constructs. When combining these results with those of changes in the E6 and E7 splice pattern and SD226 usage, it becomes clear that the GGG-cluster downstream of SD880 is very important for SD226 splicing function and usage.

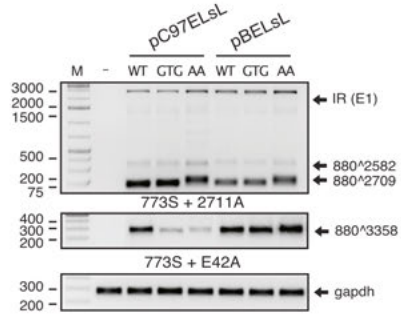

Figure 27. RT-PCR results of pC97ELsL and pBELsL constructs.

The GGG-motifs and SD880 are needed for the upstream splicing events in the E6 and E7 region.

Previous results indicate that the GGG-motifs and SD880 are of great importance for the upstream splicing within the E6 and E7 region, however, the sequences downstream of the GGG-motifs had little to no effect on the E6 and E7 splice patten. To test this, two plasmids were constructed which lacked the downstream sequences, pX960 and pX856F. In the pX856F, which ends at the E7 stop codon, the SD880 and GGG-motifs are also excluded. The RT-PCR results of these plasmids indicate that the SD880 and GGG-motifs are important for SD226-SA742 splicing and once removed splicing between the donor and acceptor is inhibited (Figure 28 A and B). Which further indicates the important role of the GGG-motifs and SD880. Since other GGG-motifs are also found within the E6 and E7 ORFs, the role of these on E6 and E7 alternative splicing was also investigated. Either all GGGs or the GGGs upstream or downstream of SA742 were mutated resulting in three new plasmids. The splice pattern of these was looked upon using RT-PCR and it was seen that SD226-SA742 splicing was lost when all GGGs or the GGGs downstream of SA742 were mutated. When the GGGs upstream of SA742 were mutated splicing to SA742 was only affected to a small extent (Figure 28 C

and D). This further strengthens the role for the GGG-motifs downstream of SD880 on SA742 splicing.

So far, the role of SD880 on splicing in the upstream E6 and E7 region had not been investigated. To determine the effect of SD880 on the regulation of this splice pattern it was mutated. When mutating SD880, splicing from SD226-SA742 was disrupted indicating a role of SD880 in splicing of the upstream sequences (Figure 28 E and F).

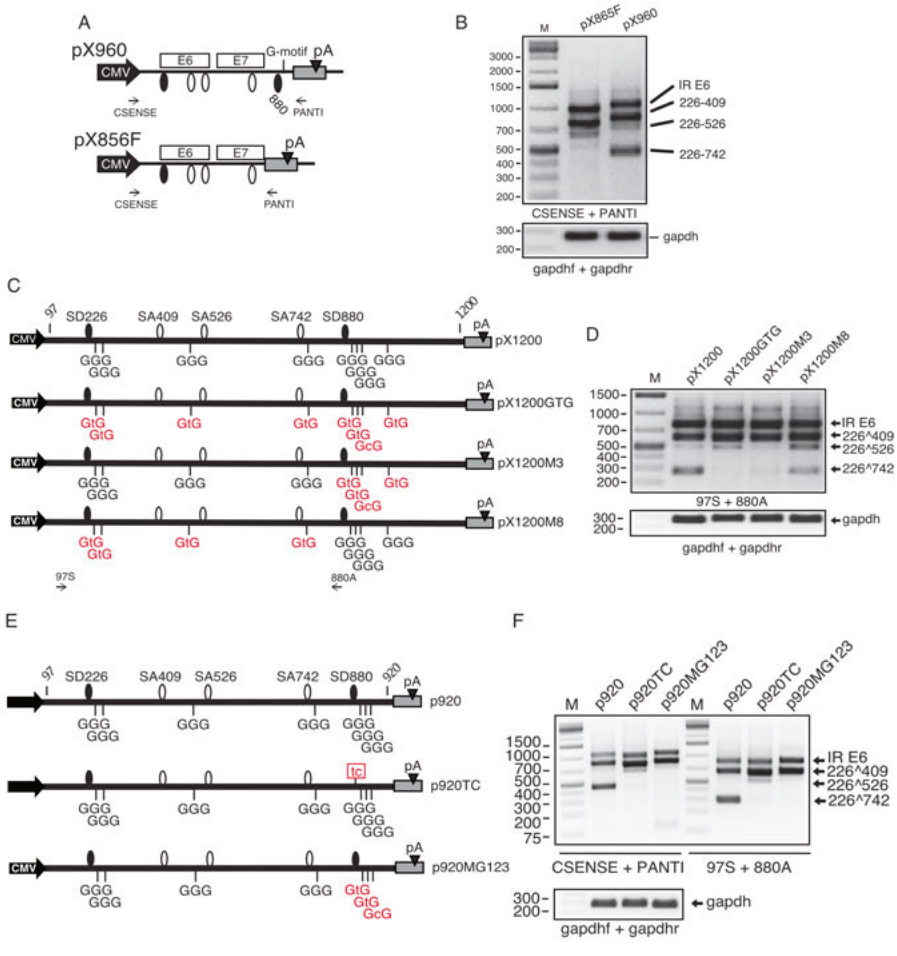

Figure 28. Determining the role of SD880 and GGG motifs within the E6 and E7 ORF on E6 and E7 splicing. **A.** Schematic overview on pX856F and pX960 constructs. **B.** RT-PCR analysis of plasmids displayed in A. **C.** Shematic overview on pX1200 plasmids, both wild-type and plasmids harbouring mutations in the GGG-motifs within the E6 and E7 ORFs. **D.** RT-PCR of the plasmids seen in C. **E.** Schematic overview of pX920 plasmids. Seen is wild-type of mutation in SD880 or GGGs downstream of SD880. **F.** RT-PCR analysis of plasmids in E.

hnRNP H binds to the GGG-motifs surrounding SD880.

Since hnRNP H is known to bind GGG-motifs and the splicing patten of HPV16 mRNA changes when these are mutated it was aimed to see how hnRNP H binds to wild-type RNA oligos as well as oligos containing mutated GGG-motifs or SD880 site. This was analyzed using an RNA oligo pull down and Western blot. As seen hnRNP H binds well to the wild type oligo, but binding is lost when the GGG-motifs are mutated (Figure 29). Interestingly, binding of hnRNP H is also reduced when SD880 is destroyed, providing valuable information about SD880 and GGG-motif integrity on hnRNP H binding. Since SD880 is a splice donor it should interact with factors of the spliceosome and when investigating binding of U1C and U1-70K, which are part of the U1 snRNP binding to 5' splice sites, both of these interact with the wild-type oligo, however the interaction is disrupted or lost when SD880 or GGG-motifs are mutated. Interestingly 3'splice factors U2AF35 and 65 also showed interactions with the RNA oligo in vitro.

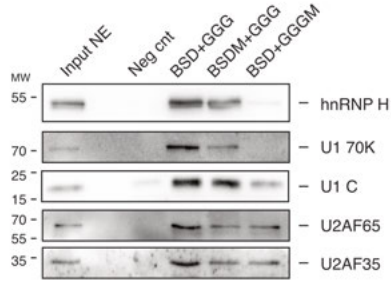

Figure 29. RNA oligo pull-down using wild-type and mutated RNA oligos. Binding of proteins is detected using Western blot.

hnRNP H overexpression changes the splice pattern of E6, E7 and E1 mRNAs.

Since hnRNP H showed interactions with the GGG-motifs in vitro and is a well-established splicing regulator we wanted to investigate how the HPV16 splice pattern changes when hnRNP H is overexpressed. This overexpression was done on wild-type or mutant pC97ELsL plasmids to determine the role of the GGG-motifs. When analyzing changes in E6 and E7 splice patterns through RT-PCR it is seen that hnRNP H increased splicing between SD226-SA742 in the wild-type plasmid (Figure 30 A). A small increase is also seen when using the mutant plasmid, however, it is not as enhanced as in the wild-type and is believed to be the result of hnRNP H interaction with the remaining GGG-motifs in the E6 and E7 region. When determining changes in protein expression using Western blot, it is seen that both E6 and E7 protein production is reduced by hnRNP H overexpression, which is consistent with the

changes in splicing (Figure 30 B). Interestingly, the E6 protein production for the GTG mutant is reduced compared to control, even though an increase was seen in intron-retained mRNA. This decrease is believed to be the result of changes in RNA processing or e.g. mRNA export.

When examining other splice pattern changes as a result of hnRNP H overexpression it is seen that the intron-retained E1 mRNA increase dramatically. A decrease in SD880 usage to the two downstream acceptors SA2709 and SA3358 is also seen (Figure 30 A). It therefore seems like hnRNP H has multifactorial effects on the HPV16 splice pattern. It increases splicing within the E6 and E7 region resulting in an improved formation of the E6^E7 mRNA and inhibits splicing involving the downstream SD880 resulting in an increased E1 mRNA.

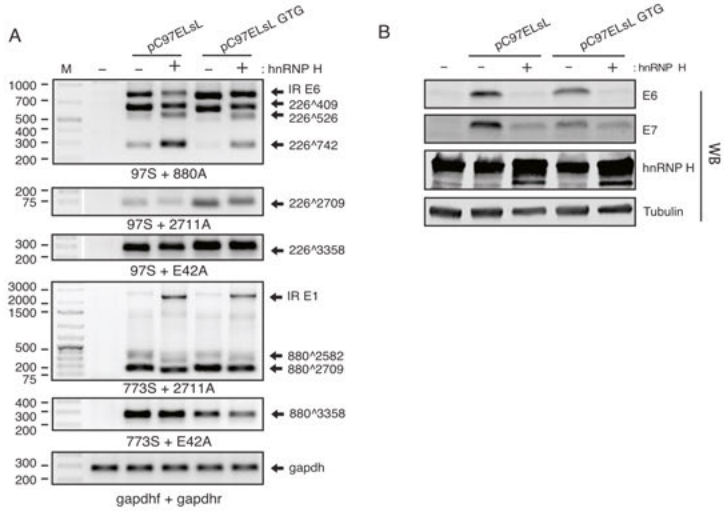

Figure 30. RT-PCR and Western blot analysis of hnRNP H overexpression on wild-type and mutant plasmids. **A.** RT-PCR analysis of pC97ELsL and pC97ELsL GTG mRNA expression with endogenous hnRNP H or overexpression. **B.** Western blot analysis of E6 and E7 protein expression from pC97ELsL and pC97ELsL GTG with endogenous hnRNP H or overexpression.

Conclusion

In this manuscript a GGG-cluster downstream of SD880 was identified. This cluster and SD880 was found to affect SA742 usage in pre-mRNA splicing. hnRNP H was identified as the binding factor and overexpression increased SA742 usage as well as reduced splicing downstream of SD880. Mutation of these GGG-motifs also increased SD226 usage to the downstream SA2709 and SA3358 splice sites. A schematic representation of the function of the

GGG-motifs and hnRNP H on HPV16 splicing of the early region is seen below (Figure 31).

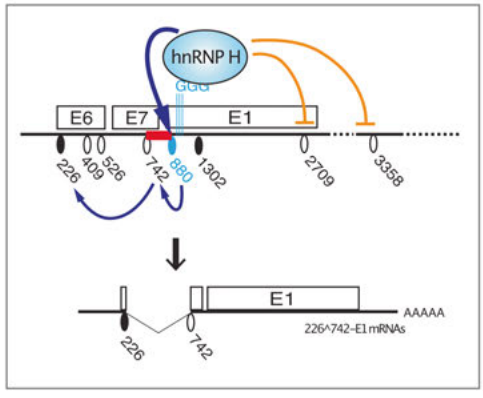

Figure 31. A representation of GGG motifs, SD880 and hnRNP H function on splicing in the early region of HPV16.

Paper III. A Janus-faced RNA element controls splicing of the HPV16 E6 and E7 oncogene mRNAs.

To regulate alternative splicing events, cis-elements and trans-acting factors are used. It is also well known that several cis-elements and trans-acting factors affect the outcome of one alternative splicing event and it is the collective regulation of these which determines the outcome [99], [102]. Since two such cis-elements and at least three trans-factors have previously been identified which regulate the outcome of the HPV16 SD226-SA409 splicing event it was aimed to see if further cis-elements regulating this splice event could be identified. These were looked for within the E6 ORF.

Two new cis-elements regulating E6 and E7 mRNA formation were identified.

Within the HPV16 E6 and E7 region one splice donor and three splice acceptors are found, which can result in the formation of four different mRNAs. The intron-retained mRNA codes for the E6 protein while two splice variants SD226-SA409 and SD226-SA526, produces two E6 truncated variants, E6*I and E6*II, and codes for the E7 protein. The formation of the E6*I/E7 mRNA is of great importance since this is the most abundant of the E6/E7 mRNAs and it is also seen as a high-risk marker [66]. To study the regulation of the intron-retained E6 and SD226-SA409 mRNA formation, deletion mutants

with increasing lengths downstream of SA409 were constructed and the formation of the intron-retained E6 and the SD226-SA409 mRNAs were looked upon using RT-PCR (Figure 32). Using these plasmids several cis-elements controlling the formation of the SD226-SA409 mRNA were found. Two of these, one positive and one negative element, have previously been reported (Paper I and Paper IV of this thesis). Two additional elements regulating SD226-SA409 splicing were found in this paper (Figure 32). These are located between nucleotides 478-502 and 525-550.

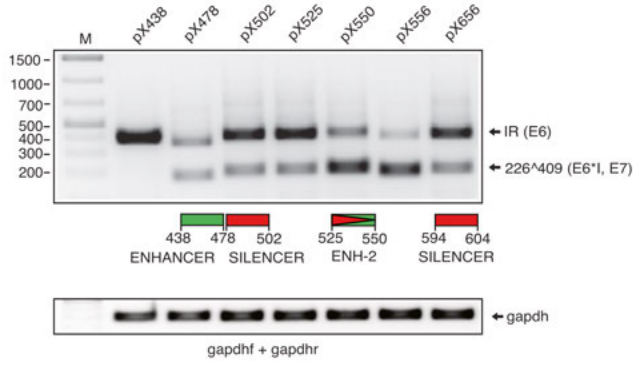

Figure 32. RT-PCR analysis of deletions mutants to identify cis-elements controlling formation of the E6*I/E7 mRNA spliced between SD226-SA409.

The cis-element between nucleotides 478 to 502 is inhibitory of SA409 splicing.

When analyzing the sequence between nucleotides 478 to 502 a motif consisting of a GGGG-sequence was found. To analyze if this was the inhibitory sequence of SD226-SA409 splicing, mutations in the GGGG motif were introduced. In addition to this, plasmids harboring mutations next to the element were also constructed. When analyzing changes in splicing using RT-PCR it is seen that only mutations in the GGGG-motif results in changes of the splice pattern (Figure 33 A and B). More specifically, splicing is increased when the GGGG-motif is disrupted further strengthening the role of the GGGG-motif as a silencer. Further mutations in the motif were introduced and mutations in the last two guanosines of the GGGG-motif resulted in more splicing between SD226 and SA409, proving an inhibitory role on SD226-SA409 splicing of the GGGG-motif in this region (Figure 33 C and D).

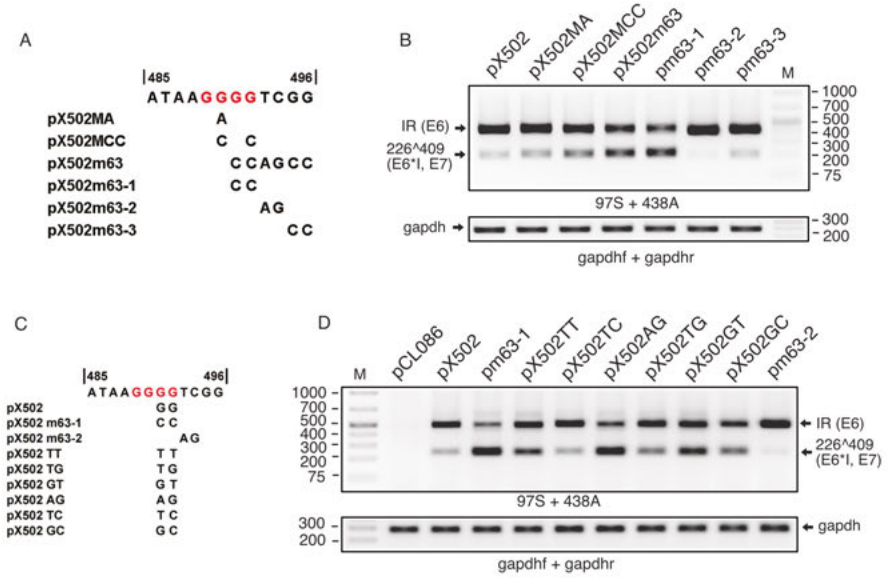

Figure 33. Analyzing the GGGG-motif between nucleotides 478 and 502. **A.** Displaying the mutations made in and next to the GGGG-motif. **B.** RT-PCR of plasmids harboring mutations shown in A. **C.** Mutations of the last two Gs in the GGGG motif and the resulting amino acid change. **D.** RT-PCR analysis of plasmids harboring mutations in C.

The cis-element at position 525 to 550 increases splicing to SA409.

A cis-element which enhanced SD226-SA409 splicing was located at nucleotides 525 to 550. To further map the enhancing element several plasmids were constructed (Figure 34 A). These were transfected and the splicing pattern was determined using RT-PCR. Splicing is inhibited in two plasmids, pX550C1 and pX550comp, when compared to the pX550 control (Figure 34 B). This proves that the enhancing element must consist of the ATCATCA sequence located between nucleotides 526 and 532. This sequence was therefore denoted as ENH-2.

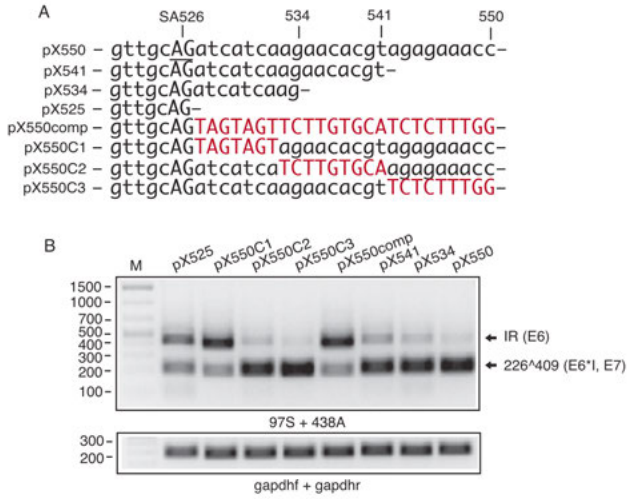

Figure 34. Analysis of the cis-element between nucleotide 525 to 550. **A.** Representation of sequences used to construct new plasmids to further map the element. **B.** RT-PCR using the plasmids denoted in A.

Further identification and characterization of the enhancing cis-element at nucleotides 525-550.

To further characterize the element and determine if all nucleotides in Figure 34 were of equal importance, mutations of two nucleotides at a time were constructed in the pX550 plasmid (Figure 35 A). When analyzing these mutations through RT-PCR it was seen that splicing between SD226-SA409 was reduced mainly in pX550M4 and M5 plasmids further pointing to an important role of the sequence identified above (Figure 35 B). The main inhibitory effect was seen in the pX550M4 plasmid and pointing to a key involvement of the two nucleotides 530 and 531 (TC).

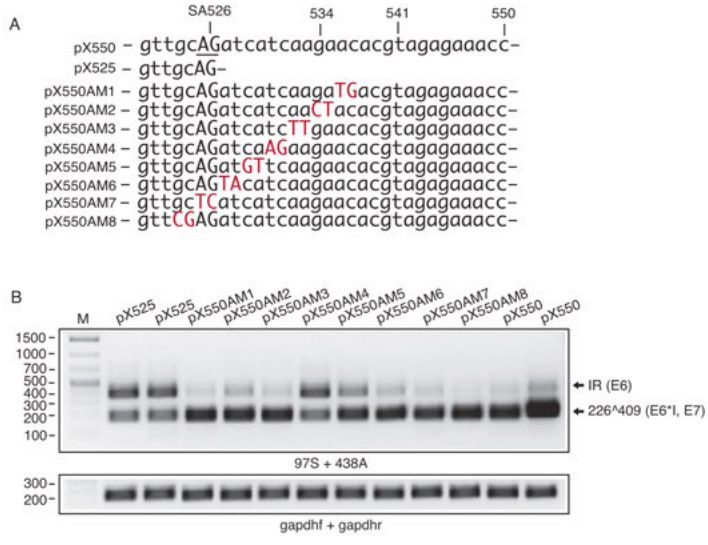

Figure 35. Analysis of plasmids harboring point mutations in the identified element and surrounding sequences. **A.** Schematic representation of mutations constructed in the pX550 plasmid. **B.** RT-PCR analysis of plasmids in A.

Changes in the cis-element behavior in pC97ELsL plasmid.

Since the M4 mutant had the most enhanced inhibitory effect this mutation was introduced in the pC97ELsL plasmid to determine the effect on splicing in the setting of all HPV16 splice sites and genes. This resulted in a new plasmid named pC97ELEh2M4. However, when introducing the mutation in this plasmid construct the results were unexpected. The M4 mutation in the pC97ELsL construct resulted in an enhancement of SD226-SA409 splicing (Figure 36 A and B), which was opposite to the results obtained from the pX550 constructs.

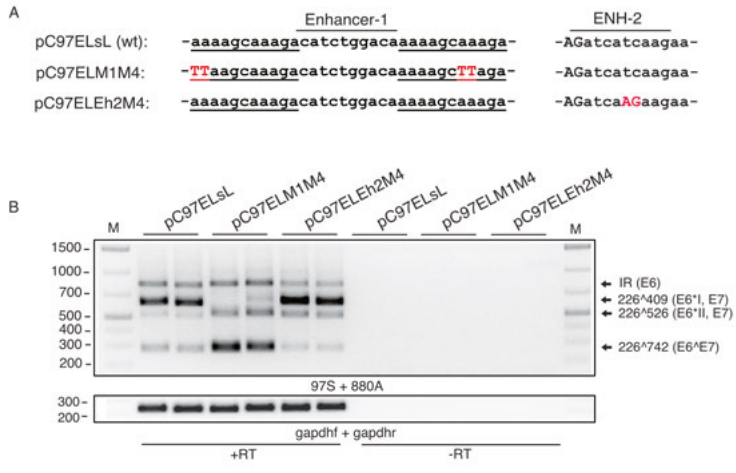

Figure 36. Introduction of the M4 mutation in the pC97ELsL produced unexpected results. **A.** Sequences showing mutations of either Enh-1 (paper I) or ENH-2 identified in this paper. **B.** RT-PCR of the plasmids from A.

Secondary structure predictions of RNA suggest an interaction with a previously identified cis-element.

To try to explain the conflicting results of the M4 mutation when using two different plasmid constructs, pX550 and pC97ELsL, predictions of the RNA secondary structure were made using *in silico* analysis. Surprisingly, the results of these predictions suggests that the identified ENH-2 sequence interacts with a silencer sequence which was previously identified and located at nucleotides 594-604 (Figure 37). This could therefore explain the conflicting effects of ENH-2. If the silencer sequence is included in the plasmid, the silencer and ENH-2 sequences pairs according to the *in silico* predictions and an internal-loop structure of the two cis-elements are formed. This internal-loop structure may expose or hide the primary sequence of those cis-elements and could therefore facilitate for RBPs to recognize and interact with the sequence which would result in downstream effects on splicing.

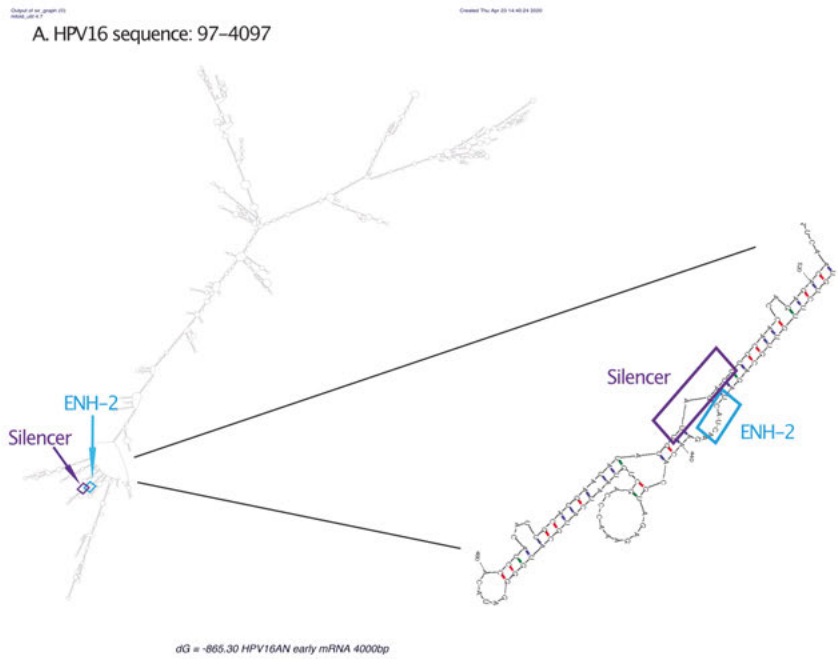

Figure 37. *In silico* predictions of RNA secondary structures.

Conclusion

In this paper two cis-elements regulating SD226-SA409 splicing were identified. The first one, is a silencer and consists of a GGGG-motif located between nucleotides 485-496. The other one, which has been named ENH-2, is found at nucleotides 526-532 and the role of this is more complex. It has been found that this cis-element has a sequence dependent effect. In a small plasmid construct containing almost the whole E6 ORF the sequence functions as an enhancer of splicing, however, in a bigger plasmid construct, pC97ELsL, it functions as a silencer. Mutational analysis and *in silico* analysis points to an interaction with a previously identified silencer located between nucleotides 594-604. The trans-acting factor binding the ENH-2 sequence still remains to be identified. A model proposing an interaction between TRAP150, the Enh-1 sequence, ENH-2 and the 594-604 silencer is displayed below (Figure 38).

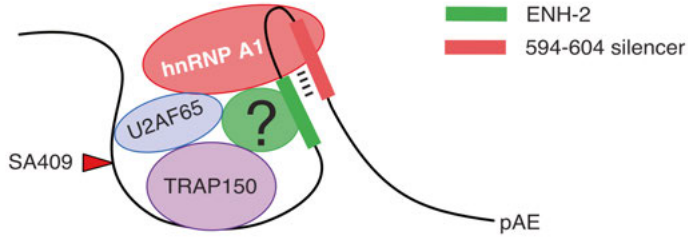

Figure 38. A predicted model for how Enh-1, TRAP150, ENH-2 and the 594-604 silencer interacts.

Paper IV. Heterogeneous Nuclear Ribonucleoprotein A1 (hnRNP A1) and hnRNP A2 Inhibit Splicing to Human Papillomavirus 16 Splice Site SA409 through a UAG-Containing Sequence in the E7 Coding Region.

Transcripts from HPV16 are exposed to lots of processing to allow for expression of all the HPV16 proteins. This since the transcripts are densely packed with several ORFs closely to each other. To facilitate for the translation machinery these mRNAs undergo extensive processing which facilitates translation. One of these processing events is the alternative splicing within the E6 and E7 region. In this region one splice donor (SD226) and three splice acceptors (SA409, SA526 and SA742) are found allowing for the formation of four mRNAs. Of these four mRNAs there is one which stands out and that is the mRNA spliced from SD226-SA409. This forms an mRNA coding for a shortened version of E6 called E6*I and has been found to be the major E7 protein producing mRNA [66]. Since the formation of this mRNA is formed through alternative splicing it is believed that there are several elements, both in cis and trans, which affect the splicing outcome and this article aimed to shed some light on this regulation.

hnRNP A1 and A2 changes the E6 and E7 splice pattern and reduces E7 protein expression.

hnRNP A1 has previously been shown to affect alternative splicing of HPV16 mRNAs [117]. Further, hnRNP A1 has previously also been reported to affect alternative splicing of HPV18 E6 and E7 mRNA [116]. To determine the effect of hnRNP A1 and A2 on E6 and E7 alternative splicing, these were overexpressed together with the pC97ELsL plasmid and the results were analyzed using RT-PCR. When analyzing the effects of the respective proteins it is seen

that they both affect E6 and E7 mRNA splicing, however, not in the same way (Figure 39 A). It is seen that hnRNP A1 inhibits splicing resulting in the formation of more intron-retained E6 mRNA, while hnRNP A2 stimulates splicing between SD226 and SA742. Both of these results occur at the expense of the remaining mRNA formation within the E6 and E7 frame. Since the mRNA expression changes when hnRNP A1 or A2 are overexpressed it would also make sense if there is a change in protein expression. Due to the marked reduction in mRNA spliced between SD226-SA409 there would naturally also be a decrease in E7 protein production. This decrease in E7 protein is seen both when hnRNP A1 or hnRNP A2 are overexpressed (Figure 39 B).

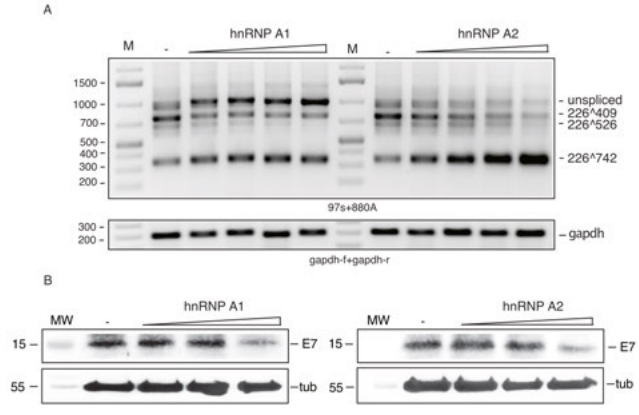

Figure 39. Determining the effect of hnRNP A1 and A2 on E6 and E7 splicing and protein expression. A. RT-PCR using pC97ELsL with either hnRNP A1 or A2 overexpression. B. Western blot analysis determining changes in E7 protein expression when hnRNP A1 or A2 are overexpressed.

hnRNP A1 and A2 work on sequences downstream of nt556 and upstream of nt856.

After having established that hnRNP A1 and A2 affected splicing within the E6 and E7 region we aimed to determine the region where the two RBPs act. To do so several deletion mutants containing varying lengths of the E6, E7 and E1 ORFs were constructed and transfected (Figure 40 A). The plasmids were either co-transfected using an empty vector or hnRNP A1 or A2 and the results were analyzed using RT-PCR. When comparing to the empty vector it is seen that hnRNP A1 and A2 inhibit splicing in all these plasmids, except for pX556F, indicating that the region where hnRNP A1 or A2 bind is downstream of nucleotide 556 (Figure 40 B).

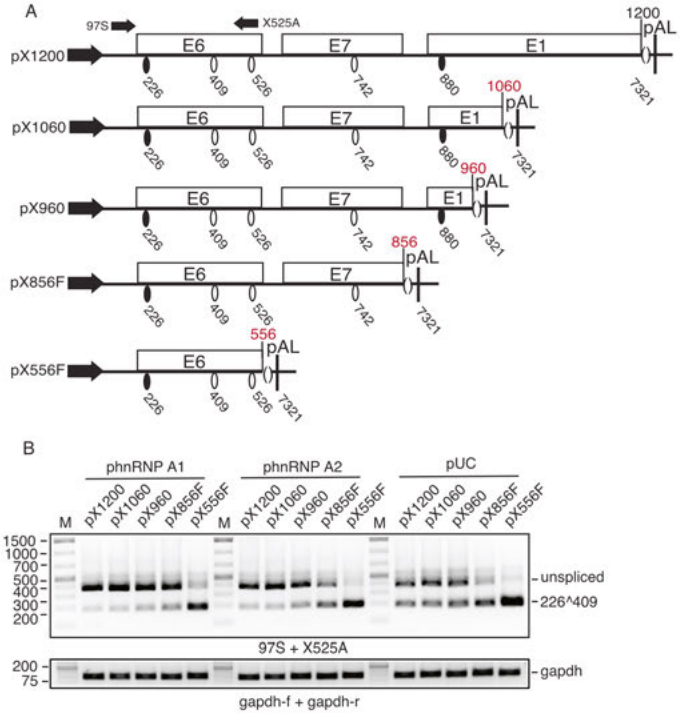

Figure 40. Narrowing down the hnRNP A1 and A2 binding sites. **A.** Plasmid maps of the plasmids used to narrow down the hnRNP A1 or A2 binding site. **B.** RT-PCR analysis of the plasmids seen in A with overexpression of either an empty vector or hnRNP A1 or A2.

The inhibitory sequences are located upstream of nt 604.

To further map the region, several new deletion plasmids were constructed and analyzed. These were either co-transfected with the empty vector or with hnRNP A1 and the results analyzed through RT-PCR. The results of these indicate that the region of interest is downstream of nucleotide 581 but upstream of nucleotide 604 (Figure 41 A). When taking a closer look at the nucleotide composition within this region an AUG-motif is found here (Figure 41 B). Since both hnRNP A1 and A2 are known to bind to AUG-rich motifs this was mutated and the effect of hnRNP A1 and A2 overexpression on the wild type and mutant plasmids, pX604 and pM604, were analyzed through RT-PCR. When looking at these results it is seen that the effect of both hnRNP A1 and A2 is reduced or lost when the AUG-motif is mutated (Figure 41 C).

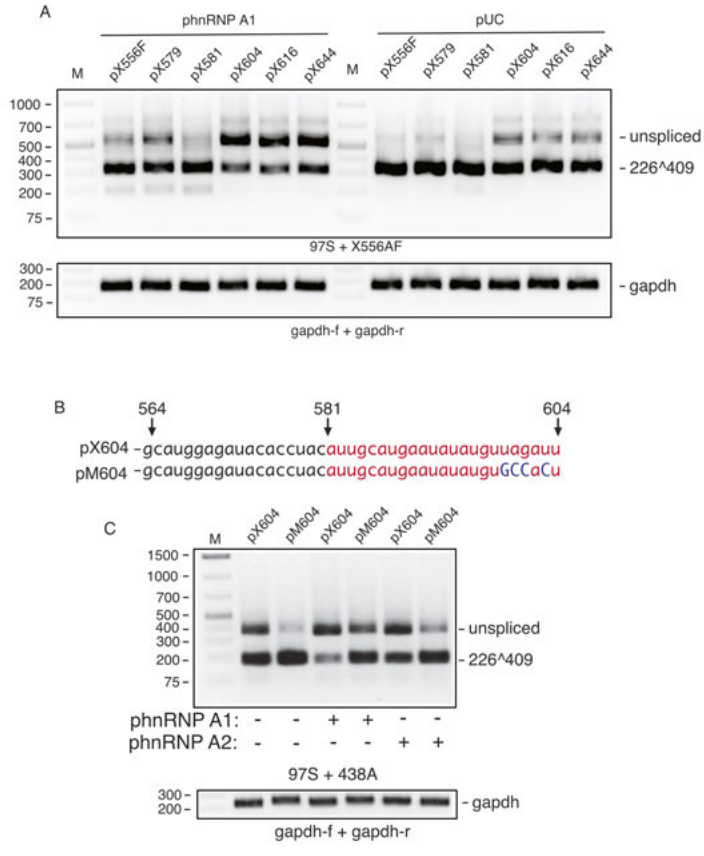

Figure 41. The hnRNP A1 and A2 interact with an AUG-motif between nucleotides 581-604. **A.** RT-PCR of further deletion mutants used to narrow down the binding site. **B.** Sequence analysis and mutation of an AUG-motif. **C.** RT-PCR of plasmids harboring the wild-type and mutant sequence seen in B.

hnRNP A1 and hnRNP A2 bind to a UAGAU-element at nt 594-604.

To confirm the results above and determine the importance of the AUG-element for hnRNP A1 and A2 binding, an RNA oligo pull-down assay was made using wild-type and mutant RNA oligos (Figure 42 A). The results were analyzed using Western blot. The results of this pull-down assay shows that both hnRNP A1 and A2 bind to the wild-type oligo containing an intact AUG-binding motif (oligo 604S1). However, if this AUG-motif is excluded or mutated, binding of both hnRNP A1 and A2 is lost (oligo 604BS1 and 604AM1) (Figure 42 B and C). With the remaining two oligos containing mutations outside of the AUG motif hnRNP A2 binding is affected using oligo 604AM2. This

indicates that nucleotides upstream of the AUG-motif may also be essential for hnRNP A2 binding.

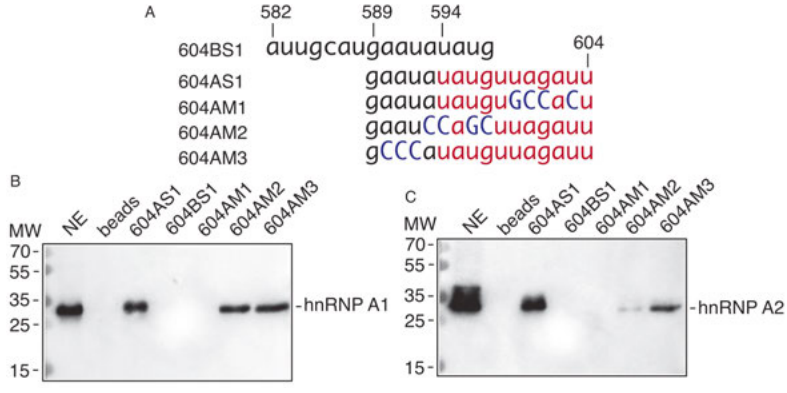

Figure 42. hnRNP A1 and A2 binding in RNA oligo pull-down assay. **A.** RNA oligos used in the RNA oligo pull down assay. **B.** Western blot analysis of the RNA oligo pull down assay.

Both the N- and C-terminal part is important for hnRNP A1 and hnRNP A2 splicing outcome.

An RNA-binding protein can generally be divided into a N-terminal and a C-terminal part. In the case of hnRNP proteins the RNA recognition motifs (RRMs) are generally found in the N-terminal part while the C-terminal domain is responsible for interaction with other proteins. It has also been seen that the N-terminal domain of hnRNPs tend to be more conserved in its sequence while the C-terminal interaction domain is more variable [108], [105]. In order to determine which domain is important for hnRNP A1 and A2 function chimeric proteins were constructed containing either the N-terminal domain from A1 or A2 and the C-domain from the counterpart (Figure 43 A). These chimeric proteins and the wild-type proteins were overexpressed in transfections using the pC97ELsL plasmid and changes in splice pattern analyzed using RT-PCR. From these results it is seen that both the N- and C-domains are needed for hnRNP A1 to exert its proper function (Figure 43 B). If the N-terminal domain of A1 is coupled to the C-terminal domain of A2 splicing inhibitory functions resulting in increasing levels of intron-retained E6 is still seen, however, the effect is reduced compared to the wild type protein. When doing the opposite, and coupling the N-terminal domain of A2 to the C-terminal domain of A1, inhibition of SD226-SA409 splicing is still seen. However, compared to the wild-type A2 which redirects splicing to SA742, the chimeric A2/A1 protein also redirects splicing to SA742 but additionally inhibits splicing resulting in the formation of more intron-retained E6 mRNA.

By compiling all these results, it is seen that both the N- and C-terminal domains of hnRNP A1 are needed for splicing.

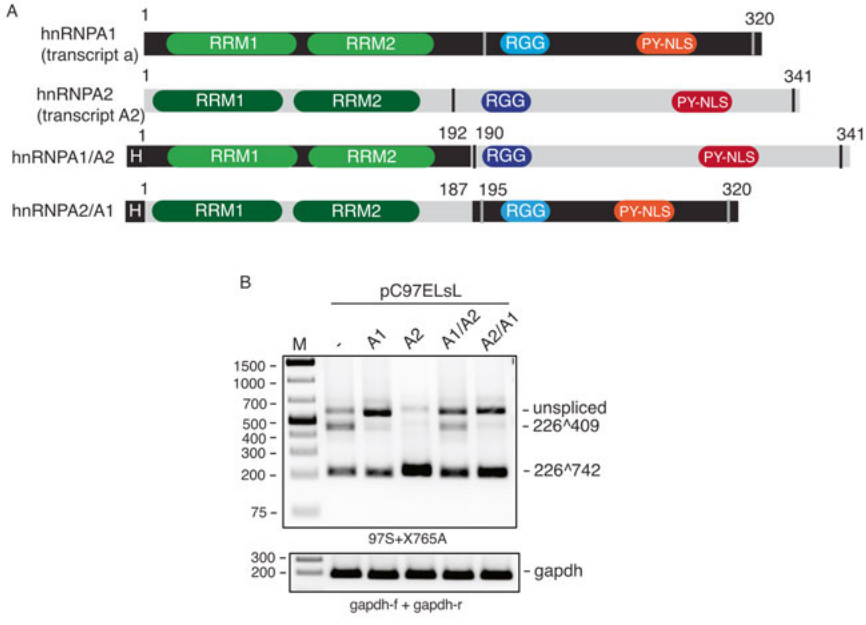

Figure 43. Chimeric hnRNP proteins and their function onE6 and E7 splicing. **A.** Chimeric hnRNP A1 and A2 proteins were constructed. The schematic representations of these are displayed. **B.** RT-PCR results of E6 and E7 splicing with overexpression of either wilt-type and chimeric hnRNP A1 and A2 plasmids.

Lowering hnRNP A1 levels in SiHa cells changes the E6 and E7 mRNA distribution and protein concentrations.

So far, we have looked at and determined the effect of hnRNP A1 and A2 overexpression on E6 and E7 splicing but have not looked at changes in splicing when these are knocked down. Therefore, hnRNP A1 or A2 were knocked down using siRNA in the HPV16 immortalized cell line SiHa cells. The changes in splicing and E7 protein expression was looked upon using RT-PCR and Western blotting. When looking at the results of hnRNP A1 knock down it is seen that more mRNA spliced between SD226-SA409 is formed compared to control (Figure 44 A). This change in mRNA splicing is also detected on a protein level with increased E7 protein expression as a result of hnRNP A1 knock down (Figure 44 B). However, for hnRNP A2 the results are more elusive where no significant changes were seen on mRNA splicing or E7 protein production.

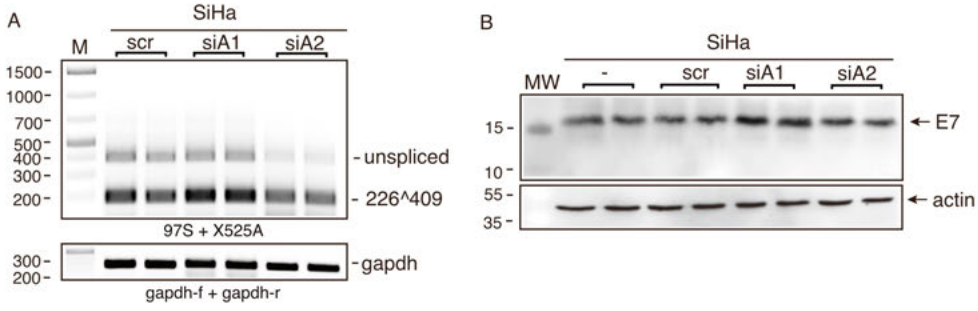

Figure 44. siRNA knock down in SiHa cells. **A.** Changes in E6 and E7 splicing was looked upon using RT-PCR when hnRNP A1 or A2 was knocked down. **B.** Changes in E7 protein expression in SiHa cells transfected with siRNA against hnRNP A1 or A2 was determined using Western blot.

Conclusion

In this paper an UAG-binding motif located at nucleotides 594-604 was identified and found to regulate HPV16 E6 and E7 mRNA splicing. To the UAG-motif hnRNP proteins A1 and A2 bind and regulate the splicing in two different ways. Overexpression of hnRNP A1 resulted in splicing inhibition and the formation of more intron-retained E6 mRNA while hnRNP A2 overexpression re-directed the splicing machinery to SA742, resulting in an increased SD226-SA742 mRNA formation. Construction of chimeric proteins proved that both the N- and C-terminal part of hnRNP A1 were important for its splicing function. Knock-down of the same protein resulted in an increase in SD226-SA409 splicing and E7 protein production.

Determining the splice pattern of high-risk HPV E6/E7 oncogenes.

Today 400 different HPV types have been identified of which 220 have been completely characterized [2]. Of these, 20 types have been associated with cancer and are by the International Association for Research on Cancer classified as group 1, group 2a and 2b carcinogens. Group 1 carcinogens have been shown to have a direct association with malignancy and cancer formation but group 2a and 2b are probably and possibly carcinogenic, respectively [6]. These HPV types are therefore called high-risk HPVs since an infection can in rare cases progress to malignancy. A hallmark for high-risk HPV types is the formation of an mRNA which is spliced within the E6 ORF creating an E6* mRNA called E6*I. This E6*I is furthermore type specific for each individual HPV type [66]. Amongst all high-risk HPV types, the splice pattern of

HPV16 and HPV18 is well characterized. Within the HPV16 E6/E7 ORF, three different E6* mRNA variants can be formed by the usage of one splice donor and three different splice acceptors. For HPV18 two such E6* mRNA variants have been identified [19]. Since the splice pattern within the E6/E7 ORF for the remaining HPV types remains elusive it was aimed to characterize the splice pattern of the remaining 18 high-risk types.

Until present, a number of studies have been conducted to determine the detection of high-risk HPVs E6* mRNAs from patient biopsies and/or from HPVE6E7 gene expressing cell culture systems (Table 2). However, it has been poorly tested to determine the E6* expression of all high-risk HPVs at the same time. Furthermore, previous studies are predominantly focusing on E6*I detection, discounting the detection of possible other truncated E6 mRNA variants. Therefore, we aimed to analyze the mechanism to produce E6* mRNAs of all high-risk HPVs in our laboratory system.

Table 2. Compilation of previously identified splice sites within the E6 and E7 ORF of high-risk HPVs.

HR-HPV	SD1	SA	Reference
16	226	409	[160] [161] [162]
		526	[162]
		742	[163] [164]
18	233	416	[160] [165]
		791	[166]
26	177	402	[160] [167]
31	210	413	[160] [168]
33	231	509	[161] [168] [166]
		785	[166]
35	232	415	[160] [168]
39	235	418	[160] [169]
45	230	413	[169]
51	177	402	[160] [170]
52	224	502	[168]
53	236	419	[160] [167]
56	161	416	[160] [168]
58	232	510	[168]
59	183	582	[168]
66	161	416	[160] [169]
67	224	502	[167]
68b	232	415	[169]
70	235	418	[160] [167]
73	227	410	[160] [167]
82	182	407	[160] [167]

All high-risk HPV types produces one E6* mRNA besides HPV16, HPV35 and HPV53

To characterize the oncogene splice pattern of all high-risk HPVs and the E6* formation, subgenomic plasmids containing the E6/E7 ORF of each HPV type were constructed. These plasmids start at the ATG of E6 and end at the E7 stop codon. Following construction, the subgenomic plasmids were transfected into HeLa cells, total RNA extracted and the splice pattern looked upon through RT-PCR (Figure 45 and 46). The identified fragments were thereafter purified from the agarose gel and sent for sequence analysis and the splice donor/acceptor mapped through sequence alignment with a reference genome of each HPV type (Table 3).

All twenty expression plasmids, except for HPV16, HPV35 and HPV53, produced one intense band and in most cases also a weaker one of a bigger size. The intense band represents the spliced E6* mRNA and was sent for sequence analysis. The splice donor and acceptor for the majority of these 20 high-risk E6* mRNAs were mapped (Table 2). For HPV16, 35 and 53 several bands were observed in the RT-PCR and all these were therefore sequenced to verify the splice pattern. However, due to insufficient material the splice pattern of lower bands expressed from HPV35 and 53 could not be determined and the splice donor and acceptor of these therefore remains elusive. For HPV39 the splice pattern likewise remains undetermined. This due to a miss-annealing of the RT-PCR primer, where half of the primer binds at the midregion of the E6/E7 ORF and therefore only detect part of the intron-retained fragment. To overcome this, the melting temperature was increased in the PCR, however to little material could be recovered for sequence analysis.

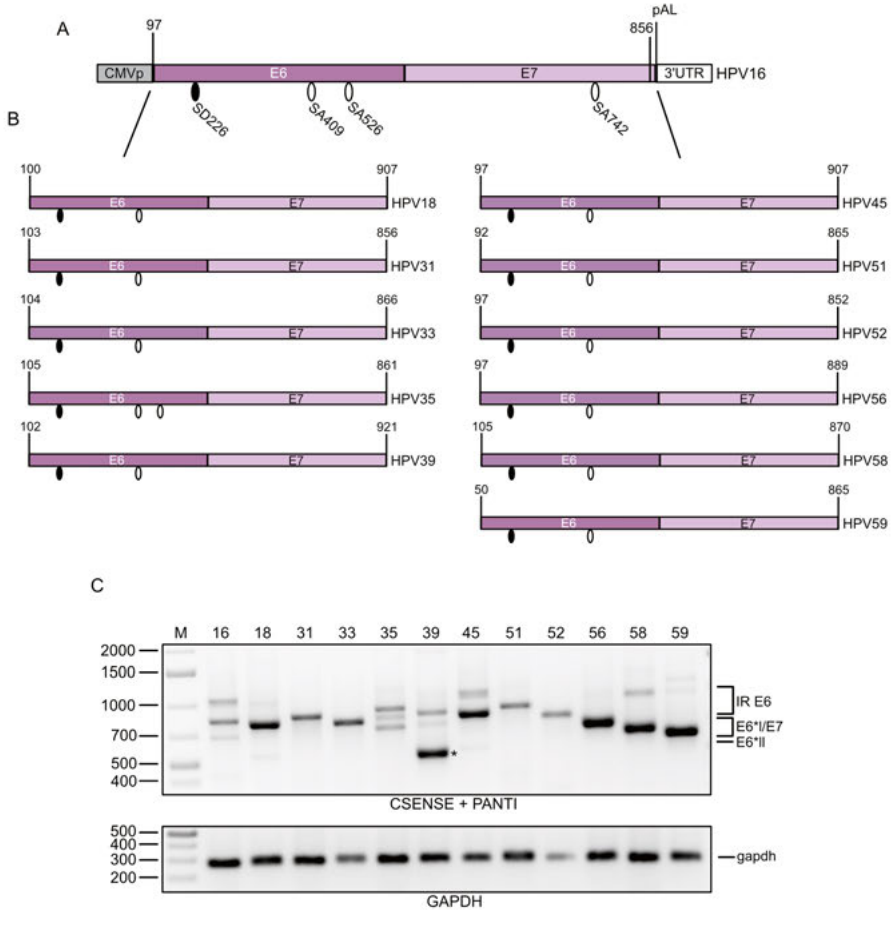

Figure 45. Determining the splice pattern of HPVs belonging to group 1 carcinogens. **A.** A schematic overview of the HPV16 E6 and E7 plasmid. Indicated are start and stop of the E6 and E7 frame and the established splice donor and acceptors. **B.** Schematic overviews of the E6 and E7 region of HPVs belonging to Group 1 carcinogens. The start and stop of E6 and E7, respectively is indicated. **C.** The RT-PCR of plasmids shown in A and B.

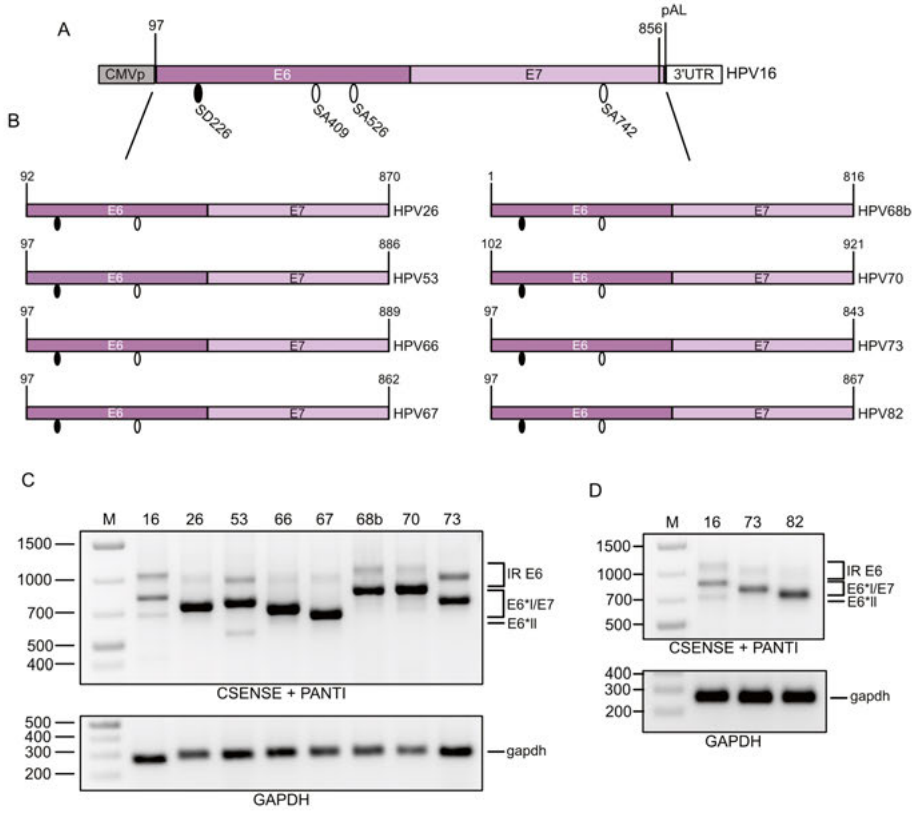

Figure 46. Determining the splice pattern of HPVs belonging to group 2a and 2b carcinogens. **A.** A schematic overview of the HPV16 E6 and E7 plasmid. Indicated are start and stop of the E6 and E7 frame and the established splice donor and acceptors. **B.** Schematic overviews of the E6 and E7 region of HPVs belonging to group 2a and 2b carcinogens. The start and stop of E6 and E7, respectively is indicated. **C.** The RT-PCR of plasmids shown in A and B.

Table 3. A summary of the start of E6 and end of E7 in each of the subgenomic expression plasmids. Indicated is also the identified splice donor and splice acceptor. For HPV35, 39 and 53 the splice donor and/or splice acceptor could not be identified and hence these are marked with N.D. (not determined). HPV82 is missing in the table since it was not available at the time of analysis.

HPV type	Start of E6	End of E7	Splice donor	Splice acceptor	Splice acceptor	Splice acceptor
16	97	856	226	409	526	742
18	100	907	233	416		
26	92	870	177	402		
31	103	856	210	413		
33	104	866	230	507		
35	105	861	232	415	N.D.	
39	102	921	N.D.	N.D.		
45	97	907	230	413		
51	92	865	177	402		
52	97	852	224	502		
53	97	886	236	419	N.D.	
56	97	889	161	416		
58	105	870	232	510		
59	50	865	183	582		
66	97	889	161	416		
67	97	862	224	502		
68b	1	816	130	312		
70	102	921	235	419		
73	97	843	227	410		
82	97	867	N.D.	N.D.		

When looking at the phylogenetic relatedness of the different high-risk HPVs, HPV31 is one of the closest relatives to HPV16. Knowing this one could assume that these two would also have the most similar splice pattern within the E6 and E7 ORFs. However, when analyzing the splice patterns of the different HPVs, HPV35 showed most similarity to that of HPV16.

For HPV16 three PCR products were detected in the RT-PCR. These were the intron-retained E6 mRNA, the E6*I mRNA, spliced between SD226-SA409, and the E6*II mRNA, spliced between SD226-SA526. The smallest mRNA spliced between SD226-SA742 could not be detected. One reason for this can be the length of the E6/E7 ORF within the subgenomic plasmid which ranges from nt97-856. Previously a GGG-motif located directly downstream of SD880 has been found which positively regulate SD226-SA742 splicing. In this subgenomic plasmid the GGG-motif is excluded since it is located downstream of the E7 ORF (Paper II of this thesis). Two additional splice donors have previously been identified in the E6 ORF of HPV16, SD191 and SD221. In this analysis no splicing using these splice sites could be identified.

These two splice donors are so called cryptic splice sites and are mostly used when SD226 function is impaired [143].

In previous publications the splice pattern of HPV18 and 33 within the E6/E7 bi-cistronic mRNA have been characterized. In these publications they have identified two splicing events for both HPV18 and HPV33. For HPV18 one donor, SD223, and two acceptors, SA416 and SA791 are used to created mRNAs spliced between SD223-SA416 and SD223-SA791. For HPV33 a donor named SD231 and two acceptors named SA509 and SA785 are used to form SD231-SA509 and SD231-SA785 mRNAs [116], [171], [172], [166]. However, in our analysis only one splice event for both HPV types were identified, SD223-SA416 and SD231-SA509 for HPV18 and HPV33, respectively. The reason for not detecting SD223-SA791 and SD231-SA785 splicing can be multifactorial. One explanation could be the regulation of the splice sites. The regulation of E6*I formation could be so strong in these plasmid constructs that it outshines other splicing events leaving mRNAs spliced to SA791 or SA785 below detection level. Or it could be that regulatory elements controlling the splice events are deleted in these constructs. The GGG-cluster (identified in paper II of this thesis) was found to be conserved in high-risk HPV types. Therefore, one can speculate that these GGG-motifs would also regulate SD223-SA791 and SD231-SA785 splicing for HPV18 and HPV33, respectively.

The splice pattens for the group 2b carcinogens HPV26, 35, 51, 56, 66, and 70 is consistent with that of previous data, however, the splice donor and acceptor differ by few nucleotides [160]. This could be due to a discrepancy in mapping of the sequences in the sequence analysis. For the group 2a carcinogen, HPV68b, the identified splice sites are inconsistent with previous findings. When comparing these with the splice donor and acceptor identified in this analysis they differ with around 100 nucleotides. This could be caused due to a discrepancy in experimental setup or a variance in the numbering of the HPV genome [167]. For HPV35 we detected two different splice variants. To our knowledge only one splice variant has previously been reported for HPV35 indicating that a novel splice variant could be detected in this paper. However, to confirm this further analysis of the splice pattern within the E6 and E7 region of HPV35 is needed.

The phylogenetic tree created using the E6 and E7 ORF is not consistent with the classical one.

The norm when characterizing newly identified HPV types is to compare part of the L1 gene. If the sequences differ by more than 10% it is considered a new HPV type. These sequences are also used when constructing a phylogenetic tree to determine the kinship between the different HPVs. In the classical phylogenetic tree HPV31 and 35 are grouped together and the closest relative to these is HPV16 [1].

In addition to the phylogenetic analysis based on the L1 gene, several studies conducted the analysis based on the E6, or combined E6-E7-L1 genes to obtain a higher informative comparison reflecting biological properties. Nevertheless, the results from such studies established similar or often identical relationships to that based on the L1 gene except for the relatedness among HPV16, 31 and 35. Phylogenetic analysis based on the E6 gene have demonstrated the phylogenetically closer relatedness between HPV16 and 35 than that of 31 and 35, which may qualify the unpreceded pattern in HPV35 splicing [173], [174].

To obtain a further informative phylogenetic comparison to understand high-risk specific E6* splicing pattern we aimed to see if a phylogenetic tree based on the E6 and E7 ORF of all high-risk HPV types would differ from that of the classical one. The phylogenetic tree was constructed using the software from Dereeper *et al.* [175]. When studying the newly constructed phylogenetic tree it resembles the classical one to a large extent but some differences exist (Figure 47). One cluster which differs in particular is the one containing HPV16, 31 and 35. In the E6 and E7 based tree HPV16 and HPV35 are the most closely related and share a common ancestor. These two in turn share an earlier ancestor with HPV31. If this can explain the E6 and E7 splice pattern remains to be seen, however, it sheds a light on the necessity for more research and sequence-based analysis of high-risk HPV types.

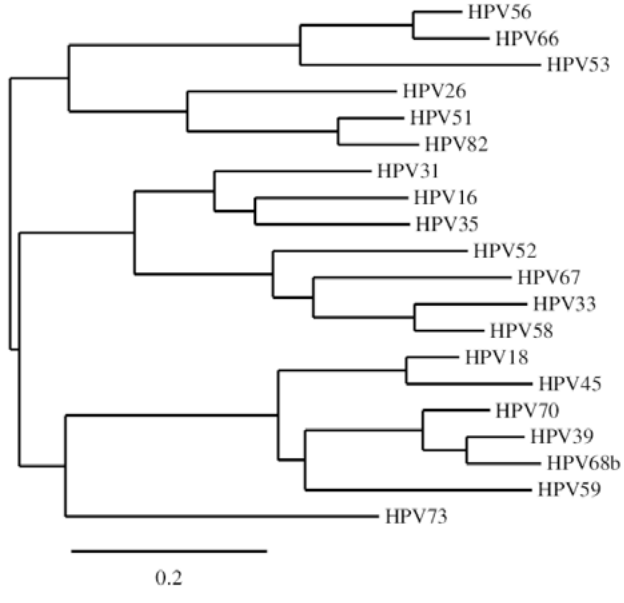

Figure 47. A phylogenetic tree based of the E6/E7 ORF of the high-risk HPV types.

The predicted splice sites, identified splice sites and reported splice sites concur

Before constructing E6 and E7 expressing plasmids to determine the E6 and E7 splice pattern, it was predicted using in silico sequence alignment. This was done by aligning the different HPV types with the HPV type they were most closely related with. HPV types sharing a recent ancestor with HPV16 were aligned with this sequence and the ones with a recent ancestor with HPV18 were aligned with it. The results of this alignments and predicted splice sites were thereafter compared with the splice sites identified using E6 and E7 expression plasmids (Table 4). For some HPV types, e.g., HPV56 and 59 either one or both of the predicted sites did not overlap well with the sites identified through sequencing of RT-PCR bands. But for others the predicted and identified sites overlapped well. This was the case for HPV31, 45 and 73. For HPV39 the comparison between predicted and identified splice sites could not be obtained due to problems in the RT-PCR analysis.

The reported splice sites of high-risk HPVs in previous studies were additionally compared with the splice sites identified using the E6 and E7 expression plasmids (Table 5). Most of the HPV types showed that the reported and identified sites overlapped well, validating our experimental system. For HPV39 the comparison between splice sites could not be obtained due to problems with the sequencing.

Table 4. Table displaying the predicted and identified splice sites of high-risk E6 and E7 ORFs. N.D. stands for not determined.

HPV plasmid	Predicted donor/ Identified donor	Predicted acceptor/ Identified acceptor
26	217/177	383/402
31	210/210	413/413
33	229/230	489/507
35	230/232	405/415
39	233/N.D.	408/N.D.
45	228/230	406/413
51	217/177	392/402
52	222/224	482/502
53	227/236	409/419
56	231/161	406/416
58	230/232	490/510
59	181/183	310/582
66	159/161	406/416
67	222/224	479/502
68b	127/130	302/312
70	233/235	408/419
73	225/227	403/410

Table 5. Table displaying the reported splice sites and identified splice sites of high-risk E6 and E7 ORFs. N.D. stands for not determined.

HPV type	Reported donor/Identified donor	Reported acceptor/Identified acceptor
26	177/177	402/402
31	210/210	413/413
33	231/230	509/507
35	232/232	415/415
39	235/N.D.	418/N.D.
45	230/230	413/413
51	177/177	402/402
52	224/224	502/502
53	236/236	419/419
56	161/161	416/416
58	232/232	510/510
59	183/183	582/582
66	161/161	416/416
67	224/224	502/502
68b	232/130	415/312
70	235/235	418/419
73	227/227	410/410

Detection of E6 and E7 mRNA using APTIMA from Hologic

In Region Skåne an APTIMA machine (Hologic) is used to screen cytology samples for HPV mRNA in the cervical screening program. This machine should, according to the manufacturer, be able to detect HPV mRNA from 14 out of 20 high-risk HPV types. These 14 types are HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66 and 68. The manufacturer also forewarn for false positive results with HPV26, 67, 70 and 82.

To put the APTIMA machine up to the test, a collaboration was started with Ola Forslund, whom is responsible for the cervical screening program in Region Skåne. The aim was to see if the mRNA expressed from the subgenomic HPV plasmids could be detected using the APTIMA machine and if the machine detected all 14 high-risk types the manufacturer claimed. To test this, the E6 and E7 expressing plasmids were transfected into 293T cells, which is an HPV negative cell line. The cells were thereafter harvested and resuspended in Thinprep (Hologic), which is used when taking cytological samples. Once the cells are resuspended in this liquid it keeps the mRNA in the cells stable by which the samples can be analyzed. The samples were thereafter handed to the microbiology department at Region Skåne where they were analyzed. In each round of testing, a positive and negative control was included to verify the results. These were the HPV16 E6/E7 expressing plasmid and pcl086, which is an HPV negative plasmid and thereby functioned as the negative control. Due to time limitations only 15 of the 20 plasmids were tested. An S/CO value above 0 indicates a positive detection of HPV mRNA. It is seen that the APTIMA machine can detect all high-risk HPV types listed by the manufacturer and a false positive result is given for HPV26, which is consistent with the manufacturer's instructions (Table 6). This analysis proves a safe screening of the cytology samples in the HPV screening program. However, worth noting is that the machine can only detect 14 high-risk types with a false positive discovery of an additional four HPVs. This brings the total types detected by the machine to 18, leaving two high-risk HPVs undetected by the machine, i.e., HPV53 and 73. This is worth noting since women with an abnormal cytology caused by HPV53 and/or HPV73 could be missed in such a screening program.

Table 6. Analysis results using the APTIMA machine to detect high-risk HPV mRNA.

Plasmid (Control/ HPV type)	S/CO value in analysis.	Detection result based on S/CO value.
pcl0806	0, 0, 0	Negative
16	10.81, 18.12, 10.82	Positive
18	15.94	Positive
26	23.58	Positive
31	21.23	Positive
33	13.78	Positive
35	31.32	Positive
39	8.91	Positive
45	29.76	Positive
51	10.1	Positive
52	28.61	Positive
56	21.04	Positive
58	15.62	Positive
66	21.48	Positive
68	15.23	Positive
73	0	Negative

Conclusion

In this project the E6 and E7 splice pattern of high-risk HPV types have been determined. All high-risk HPV types were found to form one intron-retained E6 mRNA and one spliced E6*I mRNA, except for HPV16, 35 and 53. The splice pattern of the E6*I mRNAs were analyzed using sequencing to identify the splice donor and acceptor used to form the respective E6*I mRNAs. However, the splice pattern of HPV39 could not be determined due to a missannealing of the primer. When comparing the identified splice sites to that of the predicted splice sites these were in the majority of cases very consistent. The phylogenetic tree constructed using the E6 and E7 region is consistent with that of the classical one, based on L1, however the branch containing HPV16, 31 and 35 differs and could also be an explanation to the splice pattern.

Analysis of the E6 and E7 plasmids using the APTIMA machine proves it can detect mRNA from all HPV types the manufacturer claims, however false positive detection of HPV26 is seen. HPV53 and 73 cannot be detected by the machine pointing to a risk that these samples remain undetected.

Establishment of an immortalized primary keratinocyte cell line using HPV16 DNA.

Within the HPV16 genome there are a total of 8 genes expressed. Of these there are two genes which have the ability to transform cells and these are the E6 and E7 genes [176]. These genes code for the E6 and E7 proteins which are also known as oncoproteins. The two proteins are expressed from two mutually exclusive mRNAs, which are formed either through intron-retention or through splicing in the E6 ORF. Interestingly, both the E6 and E7 oncoproteins are needed to immortalize human cells since they work in a cooperative manner and the functions of one protein complements that of the other. The two oncoproteins are mainly located in the nucleus, however, the E6 protein has also been found to be located in the cytoplasm [176], [177]. When conducting experiments to determine the immortalization ability of the respective proteins, it was found that the E7 protein can to a small extent immortalize primary cells but at a very low rate, while the E6 protein completely lacked this transforming ability. When combining the two proteins the immortalization ability was greatly enhanced proving a need for both E6 and E7 proteins to immortalize primary cells [178], [179]. A key aspect to remember, is that majority of people infected with high-risk HPVs and expression of E6 and E7 oncoproteins do not develop malignancies and carcinomas. This due to the additional need of changes in the infected cells. When the E6 and E7 proteins are expressed, they have many functions but two key actions are proliferation and prevention of apoptosis. When these two actions are active over time, damages to the cellular DNA will occur and this will result in mutations of the DNA which will eventually result in cancer development due to HPV [45].

Since both E6 and E7 protein expression is needed to immortalize primary human keratinocytes, we aimed to determine the immortalization ability of our wild type pC97ELsL plasmid and the pC97M1M4 mutant (for results see Paper I). It was also aimed to establish a primary human foreskin keratinocyte derived cell line which could be used as a reporter cell line to screen for late gene expression as well as study the HPV16 gene regulation.

Establishment of an immortalized primary keratinocyte derived cell line and its characterization.

The pC97ELsL plasmid is a subgenomic HPV16 plasmid controlled by a CMV promoter instead of the early HPV16 promoter p97 (Figure 48 A and B). It contains all HPV16 genes as well as all splice donors and acceptors and both the early and late polyadenylation sites, pAE and pAL, respectively. Worth indicating, however, is that part of the L1 gene is replaced by an internal ribosome entry site (IRES) and a secreted luciferase gene. In this way the late gene expression can be determined by measuring the secreted luciferase

of the growth medium and cells stably transfected with this plasmid can be screened for increased or reduced late gene expression when treated with various chemicals.

To establish an immortalized primary keratinocyte cell line several steps were taken (Figure 48 C). The exact procedure of this is explained in the materials and methods section. By using this immortalization method, cells are forced to differentiate and only differentiation resistant cells can survive. As previously mentioned, HPV E6 and E7 genes change the proliferation and differentiation stages of the infected cells and therefore only cells expressing these genes can survive the selection. Following selection, the surviving cells were transferred back into the normal keratinocyte medium and the gene expression from the newly established cell line was determined. It was given the name JJM9721. Total RNA was extracted from the cells and looked upon through RT-PCR to characterize the cell line.

The E6 and E7 splice pattern can be can be detected from these cells (Figure 48 D). A faint band is seen for the intron-retained E6 mRNA and a much stronger band for the SD226-SA409 mRNA, which is the major E7 coding mRNA. Two additional bands, are also seen, however these are very faint. These represent the SD226-SA526 and SA226-SA742 splicing.

The E1 mRNA expression was also looked at, using primers 97S and E1as and the intron-retained E1 mRNA as well as the intron-retained mRNA with the upstream splicing in the E6 ORF was detected (Figure 48 E). The E2 mRNA expression was also determined using primers 97S and E2qas. By using this primer pair splicing within the E6 ORF and downstream splicing events resulting in the E2 mRNA formation can be looked upon. All splice pattern within the E6 ORF with an additional downstream splice event between the major HPV16 splice donor, SD880, and splice acceptor SA2709 is detected (Figure 48 F). The splicing event between SD880 and SA2709 creates the major E2 mRNA. Since the E2 mRNA can be detected in this cell line it could indicate that the plasmid is still in its episomal form and has yet to be integrated in the genome or has integrated in another part of the plasmid. Worth noting is that in around 80% of high-grade lesions with integrated genome, the genome has integrated within the E2 ORF [45]. For E4 splicing all upstream splice events within the E6 ORF can be detected (Figure 48 G). When analyzing these results, it is seen that the strongest splice event for the E4 coding mRNA is spliced from SD226-SA742 and SD880-SA3358.

In an RT-PCR several different primer pair can be used to analyze the same target. In this characterization, two different primer pairs have been used for this request. One pair is 97S and an antisense primer binding to either the E7, E1, E2 or E4 ORF. By using such a primer pair several splice events can be looked upon at the same time, e.g., splicing within the E6 ORF and E1, E2 or E4 respectively. In this way it can be determined which is the most common splice event for a specific downstream splicing event, i.e., for E4 splicing SD226-SA742 is the most common splicing for the downstream SD880-

SA3358 splice event. In this way the mRNA production from the early promoter can also be determine, however in this study it is replaced by a CMV promoter.

By using other primer pairs mRNA production from the late promoter and splicing from the splice donor SD880 can be determined. This pair is 773S paired with either E2, E4 or L1 antisense primers. Splicing between SD880 and the two downstream splice acceptors SA2582 and SA2709 was looked upon (Figure 48 H). This showed that splicing between SD880-SA2709 is much stronger than SD880-SA2582 splicing in these cells. The E4 mRNA which is spliced between SD880 and SA3358 is also identified using 773S and E42as primers (Figure 48 I). The L1 mRNA expression from the JJM9721 cells could not be detected in this RT-PCR analysis (Figure 48 J). From previous experiments in our group, a treatment with calcium has been shown necessary to induce L1 expression and could explain the results (unpublished). Following selection and establishment of the cell line, photographs were also taken of the cells to display the characteristic features of the cell line (Figure 48 K). When comparing the phenotype of the JJM9721 cells to the keratinocytes they are derived from it is seen that a small phenotypic change has occurred (comparable image not shown).

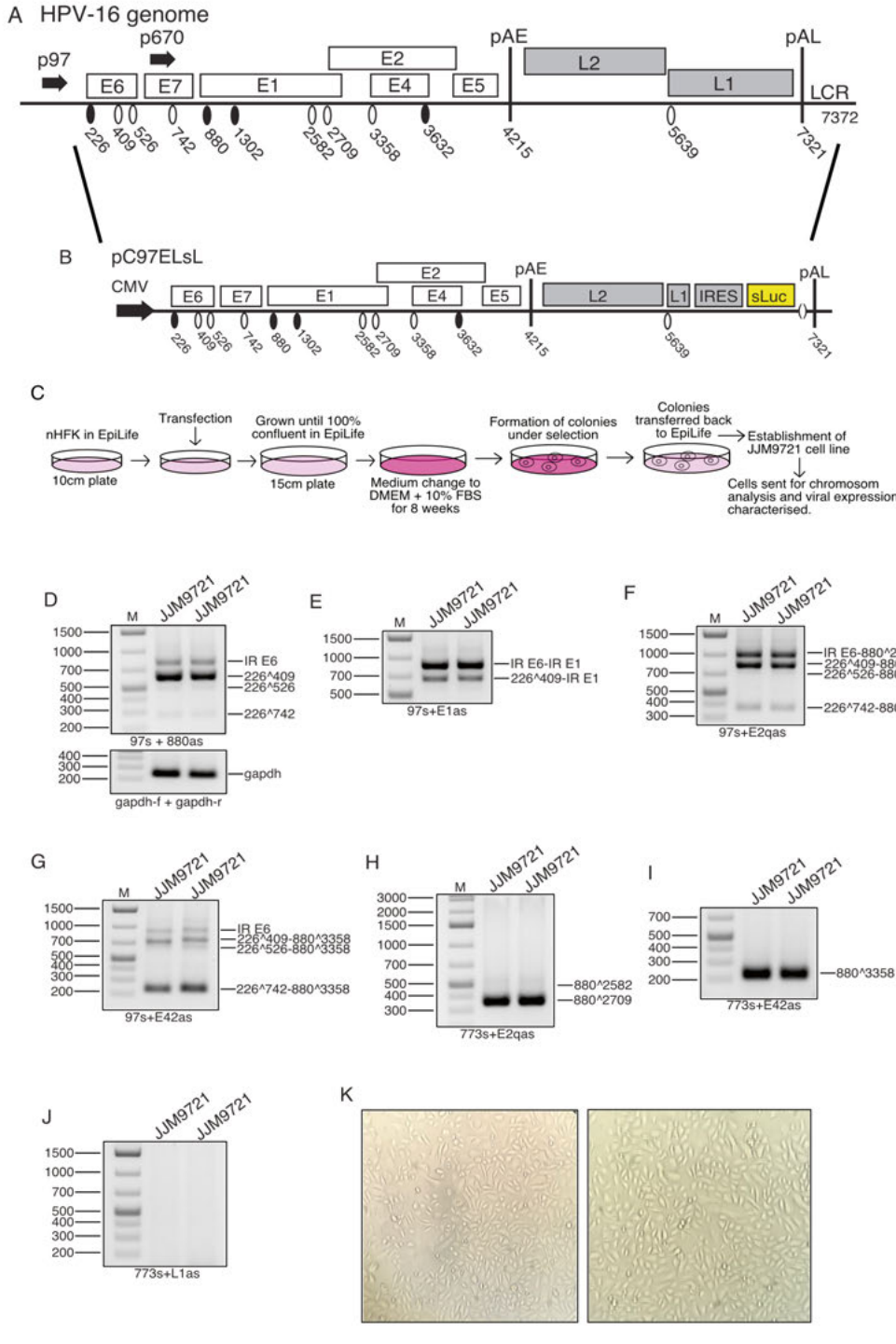

Figure 48. Establishment and characterization of the JJM9721 cell line.

Conclusion

In summary, a primary keratinocyte derived cell line, immortalized using the HPV16 subgenomic plasmid pC97ELsL, was established. This newly derived cell line was characterized by looking at mRNA expression using total RNA from two individual cell pellets. Detection of HPV16 E6 and E7, E1, E2 and E4 mRNA could be detected from these cells. Additionally, this cell line can be used to screen for an upregulation in late gene expression by measuring secreted luciferase and can therefore be used to screen for drugs inducing late gene expression.

Concluding remarks and future perspectives

In this thesis the E6 and E7 splicing of HPV16 and other high-risk HPV types have been looked upon. Several findings are made, which shed an important light on this splicing regulation. A total of four cis-elements regulating SD226-SA409 splicing in HPV16 were identified. The first of these, which we call Enh-1 (Paper I) is a strong enhancer located downstream of SA409. It consists of a perfect bipartite repeat (AAAAGCAAAGA) and mutations disrupting this element results in a marked reduction of SD226-SA409 splicing and E7 protein expression. Using an immortalization assay it was found that HPV16 plasmids harboring mutations in the cis-element were unable to immortalize primary keratinocytes. This points to a key role of the cis-element in keeping the balance of E6 and E7 mRNAs, protein expression and immortalization ability. Importantly, this cis-element is well conserved in high-risk HPV types pointing to a need for this element in high-risk splicing regulation. Further cis-elements regulating the SD226-SA409 splice event was also identified. One silencer was located between nt485-495 and it was found to consist of a GGGG-motif (Paper III). When mutated the silencing effect on SD226-SA409 splicing was alleviated proving its function as a silencer. Another silencer at nt594-604 was further identified (Paper IV). Mutations of this silencer increased SD226-SA409 splicing further indicating its silencing effects on this splice pattern. The lastly identified cis-element controlling SD226-SA409 splicing in this thesis turned out to be context dependent (Paper III). The motif of this element was narrowed down to ATCATCA, where the second TC was shown to be critical. However, depending on the plasmid construct the results of this motif was either enhancing or silencing. In a small construct, mutations of the element resulted in splicing inhibition, however, in larger constructs with all HPV16 genes, mutations enhanced splicing. The conflicting results of this cis-element may lay in the secondary structure of the mRNA, where an *in silico* analysis showed that it interacts with the downstream silencer identified in paper IV.

In addition to identifying cis-elements, trans-acting factors binding to these and regulating the splicing outcome were also identified. In Paper I two trans-acting factors increasing SD226-SA409 splicing, through binding to the Enh-1 element were identified. These factors are TRAP150 and BCLAF1 and have previously been linked to alternative splicing as well as RNA export of cellular mRNAs. Furthermore, hnRNP A1 and A2 were identified as trans-acting

factors (Paper IV). However, these interacted with the cis-element identified in paper IV and depending on the trans-acting factor the splicing outcome was different. hnRNP A1 inhibited splicing, resulting in more of the intron-retained E6 mRNA while hnRNP A2 redirected splicing to the downstream acceptor SA742. The trans-acting factors interacting with the cis-elements in Paper III remains to be identified.

In Paper II several important discoveries were also made. A GGG-cluster downstream of SD880 was identified. This was shown to be conserved in high-risk HPVs. Mutation analysis proved its importance in SD226-SA742 splicing. Furthermore, the integrity of SD880 was also proven to be important for this splicing event. The trans-acting factor was determined to hnRNP H.

Furthermore, the E6 and E7 splicing of all high-risk HPVs were identified. All high-risk HPVs form at least one spliced mRNA expressed from the E6 and E7 ORFs.

A primary keratinocyte immortalized cell line was furthermore established using HPV16 DNA. This can be used to screen for chemicals which changes the late gene expression by measuring secreted luciferase. Furthermore, changes in the HPV16 splice pattern can be looked upon using this cell line.

By compiling all this data, a new understanding of HPV16 E6 and E7 splicing regulation has been obtained. Now the challenge remains to build upon this to gain an even deeper understanding of this regulation. To continue on the research topic of the thesis it would be desirable to continue the studies of cis-elements and trans-acting factors of this thesis and understand the regulation of these on a deeper level. It would be interesting to understand how this regulation could be disrupted and thereby disrupt the E6 and E7 splicing and potentially induce apoptosis of the infected cells.

There are several ways to manipulate the alternative splicing events and change the splicing outcomes. To highlight the main findings in this thesis two possible paths could be taken. One is to understand the regulation of the trans-acting factors. As mentioned above RNA binding proteins undergo post-translational modifications which change their behavior and function within the cell [104]. By understanding these modifications and the effect of these, they can be targeted to change the protein function. By disrupting its function, the splicing outcome could thereby also be changed. TRAP150, is an important regulator of SD226-SA409 splicing, and it is a relatively unexplored protein. Therefore, it would be interesting to learn more about it. One way to do so could be to investigate the different domains of the protein and their effect on the splicing outcome. PTMs could be another interesting target and by understanding how this is regulated it could be manipulated using inhibitors.

Another path to take is to directly target the cis-elements. This could be possible by using antisense oligonucleotides (ASOs). An ASO is a short oligonucleotide which is complementary to the mRNA sequence and it generally consists of 15-25 nucleotides. When these bind to the mRNA, the cis-element could be masked preventing the trans-acting factor from binding and exerting

its effect on splicing. Thus, the alternative splicing event would not occur [180]. Binding of ASOs could also result in degradation of the mRNA. This degradation is controlled by RNases which recognizes double stranded RNA [104]. If such an ASO would be synthesized and target Enh-1, it should result in inhibition of SD226-SA409 splicing of the E6 and E7 pre-mRNA. This results in a loss of the major E7 producing mRNA and therefore also a loss of E7 protein production and if no E7 is expressed in the cells they enter senescence. However, before such conclusions can be made further research is needed to validate the effects of ASOs targeting the strong enhancer sequence. It must also be validated that the ASO will not affect any critical alternative splicing events of cellular mRNAs.

Additionally, the cell line constructed in this thesis (JJM9721) could be used to screen for small chemicals inducing late gene expression by measuring changes in luciferase expression. By identifying chemicals inducing late gene expression this could force HPV infected cells to prematurely express the late genes L1 and L2. Since these are highly immunogenic it could result in detection of the immune system and an attack of this on HPV infected cells [66]. The cell line could additionally be used to screen for chemicals disrupting the alternative splicing events of HPV16 mRNA, in this case primarily of E6 and E7 mRNAs.

To conclude, the findings in this thesis have opened up to a deeper understanding of HPV16 E6 and E7 mRNA splicing. Even though more research is required, the findings here can be built upon to one day develop better treatments for HPV infected patients.

Acknowledgements

I would like to say a huge thank you to Dr. **Naoko Kajitani**, my main supervisor. You came when I had already started my PhD but without you, I would not be where I am today. You showed me so many things in the lab and I have looked up to you and taken a lot of inspiration from you when it comes to structuring and planning my work. Without you the keratinocyte work and the JJM9721 cells would not have been possible and for this I will be forever thankful. You have inspired me in many ways and without you I would not be the person I am in the lab today.

To Professor **Stefan Schwartz**, you were the one whom accepted me to the PhD position and my main supervisor for the first years and thereafter my co-supervisor. For this I will be forever thankful. I also appreciate how you convinced me to never give up and continue pushing. When the cloning did not work in the beginning of my PhD you told me that one day it will work, which it always did. It might have been on the eight or tenth try but it did work in the end.

Dr**. Ola Forslund**, my co-supervisor, thank you for giving us the high-risk HPV plasmids. These have really help me with my experiments during my PhD journey. It was nice to collaborate and to see if "your" machine could detect the E6/E7 expressing plasmids or not and see how well this coincided with the manufacturer's instructions.

Dr. **Tanel Punga**. I want to thank you for all the help and support I received from you with the re-registration and move to Uppsala. Your help and the fact that you were never more than an e-mail away whenever I had a question has helped me so much. Thank you for always helping me whenever I felt unaccustomed to the new PhD system and all that came with it. You're a great scientist and I'm amazed by the knowledge you have.

To the current members in our group. I am thankful for all the support and nice feedback along the way. To **Lourenco Cardoso**, you are excellent at what you do! Keep on the good work and your future will be bright. I wish you the best of luck for everything you take on in the future. **Yani Zhao**, thank you for all the support during my time in Uppsala. The smile on your face

always put a smile on mine. I wish you the best of luck with your cell line. To **Xiaoli Zhai** and **Lianqing Wang**, you came to our lab in Uppsala shortly after our move here and quickly became acquainted with the lab and our group. To the both of you I want to say a big thank you for all the help and input on my work. Seeing you work have given me lots of inspiration.

To previous members of our group. To **Xiaoxu Cui**, your passion and your knowledge have always inspired me. I looked up to you in the lab and I'm forever thankful for the help and support you gave me during our time in Lund. Whenever I had questions or was confused about something you always took the time to help me out and explain whatever I could not comprehend. To **Chengyu Hao**, I loved the discussions we used to have in our office. We could talk about work and the struggles with it but also about life and all that comes with it. To **Ye Zhu**, I hope you're continuing to spread the positive energy you have. You always had the ability to make others happy. To **Yuniji Zheng**, my previous office mate. I enjoyed our discussions about life and all there to come. To **Shirin Shoja Chagervand,** you were always happy in the lab and I wish to always have your optimistic view on life. To **Chengjung Wu**, thank you for the invitation to visit your lab. If time was longer, I would have happily come to visit you and your lab in Dalian. To **Haoran Yu**, the girl whom always smiled. Thank you for all the help I received from you when I came to the lab and thank you for being my lab supervisor when I first took the Virology course back in 2015. This inspired me to continue with my interest in virology. To **Kersti Nilsson**, thank you for all the help and support I received from you when I first came to the lab. I learnt a lot from you! To **Lijing Gong**, even though our meeting was brief you inspired me. Thank you for the help and support I received at the start of my research. To **Mandy Do**, the year you spent in the lab was a great year. I appreciated that we talked about everything and am happy that our friendship still lasts today. To **Neele Bergemann**, your ability to continue working and never give up inspired me. I wish you all the happiness and success.

I would like to thank everyone attending our Friday seminar group at BMC in Uppsala. The presentations and talks we had were very inspiring to me. I enjoyed hearing about other viruses and widening my horizons. I would like to give a special thank you to Professor **Göran Magnusson**, Professor **Göran Akusjärvi**, Professor **Catharina Svensson**, Dr **Anders Bergqvist** and Dr **Tanel Punga** for all the feedback on my projects. This helped me develop as a researcher.

My office members and friends in Uppsala. We might be in different research fields, virology, bacteriology and immunology, but this only widened my horizons. Thank you for the work discussions but also for the discussions we had about life outside of work. I really appreciated our Friday game nights. To

Xinran Zhao for going to the gym with me at 6am and inspiring me to continue going to the gym even during the stressful times, when I otherwise would have not. To **Erik Shubert** for always being someone I could talk to and for understanding. And thank you for the nice contributions when I was deciding about my thesis title. To **Chris von Beek**, may the honorary title of this thesis always be hot and splicy.

To **Mamma, Pappa and Josefin**. Thank you for always being there and being my greatest support. We did this journey together and I would not have been able to do it without you by my side. Jag älskar er! And to **Erik**, thank you for always showing interest in what I do and always asking how it is going along the way. Even though we are in completely different line of work you have always showed interest and encouragement in what I do.

To **Ashish**, my number one fan and partner, you where there when it started and you are there when it ends. Thank you for always being there to pick me up when I fall. Thank you for all the love, patience, support and encouragement you have shown me throughout this journey. It made us stronger. Now we'll see what the future has in stored.

My friends. For being there to talk and ventilate about life whenever I needed to. And for being a break in my day-to-day life and giving me many things to look forward to. Even though we all live far away from each other I'm always looking forward to when we meet or Skype and you have always given me much needed energy after seeing each other.

Reference list

1. de Villiers, E.M., et al., *Classification of papillomaviruses.* Virology, 2004. **324**(1): p. 17-27.
2. *Pave*. [cited 2023 2023-03-31].
3. Doorbar, J., et al., *The biology and life-cycle of human papillomaviruses.* Vaccine, 2012. **30 Suppl 5**: p. F55-70.
4. 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.
5. McBride, A.A., *Human papillomaviruses: diversity, infection and host interactions.* Nat Rev Microbiol, 2022. **20**(2): p. 95-108.
6. Cancer., L.F.I.A.f.r.o., *IARC working Group on the Evaluation of Carcinogenic Risks to Humans. Biological agents.* , in *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, No. 100B.* 2012, IARC: Lyon, France. .
7. Kombe Kombe, A.J., et al., *Epidemiology and Burden of Human Papillomavirus and Related Diseases, Molecular Pathogenesis, and Vaccine Evaluation.* Front Public Health, 2020. **8**: p. 552028.
8. Chesson, H.W., et al., *The estimated lifetime probability of acquiring human papillomavirus in the United States.* Sex Transm Dis, 2014. **41**(11): p. 660-4.
9. Fortes, H.R., et al., *Recurrent respiratory papillomatosis: A state-of-the-art review.* Respir Med, 2017. **126**: p. 116-121.
10. Saraiya, M., et al., *US assessment of HPV types in cancers: implications for current and 9-valent HPV vaccines.* J Natl Cancer Inst, 2015. **107**(6): p. djv086.
11. *Tobacco habits other than smoking; betel-quid and areca-nut chewing; and some related nitrosamines. IARC Working Group. Lyon, 23-30 October 1984.* IARC Monogr Eval Carcinog Risk Chem Hum, 1985. **37**: p. 1-268.
12. Cancer, I.A.f.R.o., *Alcohol Drinking. IARC Monographs on the Evaluation of the Carcinogenic Risks to Humans. Volume 44.* . 1988.
13. Brianti, P., E. De Flammineis, and S.R. Mercuri, *Review of HPV-related diseases and cancers.* New Microbiol, 2017. **40**(2): p. 80-85.
14. IARC. *Cervical cancer*. [cited 2023 07-28]; Available from: https://www.iarc.who.int/cancer-type/cervical-cancer/.
15. Parkin, D.M. and F. Bray, *Chapter 2: The burden of HPV-related cancers.* Vaccine, 2006. **24 Suppl 3**: p. S3/11-25.
16. Ma, B., R. Roden, and T.C. Wu, *Current status of human papillomavirus vaccines.* J Formos Med Assoc, 2010. **109**(7): p. 481-3.
17. (CDC)., C.f.D.C.a.P. *Human Papillomavirus (HPV) Vaccination: What Everyone Should Know.* . 2021; Available from: https://www.cdc.gov/vaccines/vpd/hpv/public/index.html.
18. Harden, M.E. and K. Munger, *Human papillomavirus molecular biology.* Mutat Res Rev Mutat Res, 2017. **772**: p. 3-12.

19. Yu, L., V. Majerciak, and Z.M. Zheng, *HPV16 and HPV18 Genome Structure, Expression, and Post-Transcriptional Regulation.* Int J Mol Sci, 2022. **23**(9).
20. Porter, S.S., et al., *Histone Modifications in Papillomavirus Virion Minichromosomes.* mBio, 2021. **12**(1).
21. Graham, S.V., *The human papillomavirus replication cycle, and its links to cancer progression: a comprehensive review.* Clin Sci (Lond), 2017. **131**(17): p. 2201-2221.
22. Kaur, P. and A. Li, *Adhesive properties of human basal epidermal cells: an analysis of keratinocyte stem cells, transit amplifying cells, and postmitotic differentiating cells.* J Invest Dermatol, 2000. **114**(3): p. 413-20.
23. Herfs, M., et al., *A discrete population of squamocolumnar junction cells implicated in the pathogenesis of cervical cancer.* Proc Natl Acad Sci U S A, 2012. **109**(26): p. 10516-21.
24. Maglennon, G.A., P. McIntosh, and J. Doorbar, *Persistence of viral DNA in the epithelial basal layer suggests a model for papillomavirus latency following immune regression.* Virology, 2011. **414**(2): p. 153-63.
25. Bodily, J. and L.A. Laimins, *Persistence of human papillomavirus infection: keys to malignant progression.* Trends Microbiol, 2011. **19**(1): p. 33-9.
26. Bergvall, M., T. Melendy, and J. Archambault, *The E1 proteins.* Virology, 2013. **445**(1-2): p. 35-56.
27. McBride, A.A., *The papillomavirus E2 proteins.* Virology, 2013. **445**(1-2): p. 57-79.
28. Barrow-Laing, L., W. Chen, and A. Roman, *Low- and high-risk human papillomavirus E7 proteins regulate p130 differently.* Virology, 2010. **400**(2): p. 233-9.
29. Roman, A., *The human papillomavirus E7 protein shines a spotlight on the pRB family member, p130.* Cell Cycle, 2006. **5**(6): p. 567-8.
30. Klingelhutz, A.J. and A. Roman, *Cellular transformation by human papillomaviruses: lessons learned by comparing high- and low-risk viruses.* Virology, 2012. **424**(2): p. 77-98.
31. Munger, K. and P.M. Howley, *Human papillomavirus immortalization and transformation functions.* Virus Res, 2002. **89**(2): p. 213-28.
32. Gewin, L. and D.A. Galloway, *E box-dependent activation of telomerase by human papillomavirus type 16 E6 does not require induction of c-myc.* J Virol, 2001. **75**(15): p. 7198-201.
33. Middleton, K., et al., *Organization of human papillomavirus productive cycle during neoplastic progression provides a basis for selection of diagnostic markers.* J Virol, 2003. **77**(19): p. 10186-201.
34. Ruesch, M.N. and L.A. Laimins, *Human papillomavirus oncoproteins alter differentiation-dependent cell cycle exit on suspension in semisolid medium.* Virology, 1998. **250**(1): p. 19-29.
35. Ozbun, M.A. and C. Meyers, *Characterization of late gene transcripts expressed during vegetative replication of human papillomavirus type 31b.* J Virol, 1997. **71**(7): p. 5161-72.
36. Ozbun, M.A. and C. Meyers, *Temporal usage of multiple promoters during the life cycle of human papillomavirus type 31b.* J Virol, 1998. **72**(4): p. 2715-22.
37. Bedell, M.A., et al., *Amplification of human papillomavirus genomes in vitro is dependent on epithelial differentiation.* J Virol, 1991. **65**(5): p. 2254-60.
38. Buck, C.B., et al., *Maturation of papillomavirus capsids.* J Virol, 2005. **79**(5): p. 2839-46.
39. Stanley, M., *Pathology and epidemiology of HPV infection in females.* Gynecol Oncol, 2010. **117**(2 Suppl): p. S5-10.

40. Egawa, N., et al., *Human Papillomaviruses; Epithelial Tropisms, and the Development of Neoplasia.* Viruses, 2015. **7**(7): p. 3863-90.
41. Berman, T.A. and J.T. Schiller, *Human papillomavirus in cervical cancer and oropharyngeal cancer: One cause, two diseases.* Cancer, 2017. **123**(12): p. 2219-2229.
42. Baldwin, P., R. Laskey, and N. Coleman, *Translational approaches to improving cervical screening.* Nat Rev Cancer, 2003. **3**(3): p. 217-26.
43. Doorbar, J., et al., *Human papillomavirus molecular biology and disease association.* Rev Med Virol, 2015. **25 Suppl 1**(Suppl Suppl 1): p. 2-23.
44. Martin, C.M. and J.J. O'Leary, *Histology of cervical intraepithelial neoplasia and the role of biomarkers.* Best Pract Res Clin Obstet Gynaecol, 2011. **25**(5): p. 605-15.
45. Moody, C.A. and L.A. Laimins, *Human papillomavirus oncoproteins: pathways to transformation.* Nat Rev Cancer, 2010. **10**(8): p. 550-60.
46. Groves, I.J. and N. Coleman, *Pathogenesis of human papillomavirus-associated mucosal disease.* J Pathol, 2015. **235**(4): p. 527-38.
47. Pett, M. and N. Coleman, *Integration of high-risk human papillomavirus: a key event in cervical carcinogenesis?* J Pathol, 2007. **212**(4): p. 356-67.
48. Jeon, S. and P.F. Lambert, *Integration of human papillomavirus type 16 DNA into the human genome leads to increased stability of E6 and E7 mRNAs: implications for cervical carcinogenesis.* Proc Natl Acad Sci U S A, 1995. **92**(5): p. 1654-8.
49. Marechal, A. and L. Zou, *DNA damage sensing by the ATM and ATR kinases.* Cold Spring Harb Perspect Biol, 2013. **5**(9).
50. Zeman, M.K. and K.A. Cimprich, *Causes and consequences of replication stress.* Nat Cell Biol, 2014. **16**(1): p. 2-9.
51. 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.
52. 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.
53. Sakakibara, N., R. Mitra, and A.A. McBride, *The papillomavirus E1 helicase activates a cellular DNA damage response in viral replication foci.* J Virol, 2011. **85**(17): p. 8981-95.
54. Edwards, T.G., et al., *Human papillomavirus episome stability is reduced by aphidicolin and controlled by DNA damage response pathways.* J Virol, 2013. **87**(7): p. 3979-89.
55. Moody, C.A., et al., *Human papillomaviruses activate caspases upon epithelial differentiation to induce viral genome amplification.* Proc Natl Acad Sci U S A, 2007. **104**(49): p. 19541-6.
56. 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.
57. Reinson, T., et al., *Engagement of the ATR-dependent DNA damage response at the human papillomavirus 18 replication centers during the initial amplification.* J Virol, 2013. **87**(2): p. 951-64.
58. 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.
59. Hong, S. and L.A. Laimins, *The JAK-STAT transcriptional regulator, STAT-5, activates the ATM DNA damage pathway to induce HPV 31 genome*

amplification upon epithelial differentiation. PLoS Pathog, 2013. **9**(4): p. e1003295.
60. 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.
61. Bester, A.C., et al., *Nucleotide deficiency promotes genomic instability in early stages of cancer development.* Cell, 2011. **145**(3): p. 435-46.
62. Spardy, N., et al., *HPV-16 E7 reveals a link between DNA replication stress, fanconi anemia D2 protein, and alternative lengthening of telomere-associated promyelocytic leukemia bodies.* Cancer Res, 2008. **68**(23): p. 9954-63.
63. Vohhodina, J., et al., *The RNA processing factors THRAP3 and BCLAF1 promote the DNA damage response through selective mRNA splicing and nuclear export.* Nucleic Acids Res, 2017. **45**(22): p. 12816-12833.
64. Savage, K.I., et al., *Identification of a BRCA1-mRNA splicing complex required for efficient DNA repair and maintenance of genomic stability.* Mol Cell, 2014. **54**(3): p. 445-59.
65. Nilsson, K., et al., *The DNA damage response activates HPV16 late gene expression at the level of RNA processing.* Nucleic Acids Res, 2018. **46**(10): p. 5029-5049.
66. Kajitani, N. and S. Schwartz, *The role of RNA-binding proteins in the processing of mRNAs produced by carcinogenic papillomaviruses.* Semin Cancer Biol, 2022. **86**(Pt 3): p. 482-496.
67. Hoppe-Seyler, K., et al., *The HPV E6/E7 Oncogenes: Key Factors for Viral Carcinogenesis and Therapeutic Targets.* Trends Microbiol, 2018. **26**(2): p. 158-168.
68. Pal, A. and R. Kundu, *Human Papillomavirus E6 and E7: The Cervical Cancer Hallmarks and Targets for Therapy.* Front Microbiol, 2019. **10**: p. 3116.
69. Vande Pol, S.B. and A.J. Klingelhutz, *Papillomavirus E6 oncoproteins.* Virology, 2013. **445**(1-2): p. 115-37.
70. Thomas, M. and L. Banks, *Human papillomavirus (HPV) E6 interactions with Bak are conserved amongst E6 proteins from high and low risk HPV types.* J Gen Virol, 1999. **80 (Pt 6)**: p. 1513-1517.
71. Duffy, C.L., S.L. Phillips, and A.J. Klingelhutz, *Microarray analysis identifies differentiation-associated genes regulated by human papillomavirus type 16 E6.* Virology, 2003. **314**(1): p. 196-205.
72. Wang, X., et al., *Oncogenic HPV infection interrupts the expression of tumor-suppressive miR-34a through viral oncoprotein E6.* RNA, 2009. **15**(4): p. 637-47.
73. Martinez, I., et al., *Human papillomavirus type 16 reduces the expression of microRNA-218 in cervical carcinoma cells.* Oncogene, 2008. **27**(18): p. 2575-82.
74. Au Yeung, C.L., et al., *Human papillomavirus type 16 E6 induces cervical cancer cell migration through the p53/microRNA-23b/urokinase-type plasminogen activator pathway.* Oncogene, 2011. **30**(21): p. 2401-10.
75. James, M.A., J.H. Lee, and A.J. Klingelhutz, *Human papillomavirus type 16 E6 activates NF-kappaB, induces cIAP-2 expression, and protects against apoptosis in a PDZ binding motif-dependent manner.* J Virol, 2006. **80**(11): p. 5301-7.
76. Veldman, T., et al., *Transcriptional activation of the telomerase hTERT gene by human papillomavirus type 16 E6 oncoprotein.* J Virol, 2001. **75**(9): p. 4467-72.

77. Roman, A. and K. Munger, *The papillomavirus E7 proteins.* Virology, 2013. **445**(1-2): p. 138-68.
78. 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.
79. Della Fera, A.N., et al., *Persistent Human Papillomavirus Infection.* Viruses, 2021. **13**(2).
80. Deng, W., et al., *Cyclin/CDK regulates the nucleocytoplasmic localization of the human papillomavirus E1 DNA helicase.* J Virol, 2004. **78**(24): p. 13954-65.
81. Castro-Munoz, L.J., et al., *The Human Papillomavirus (HPV) E1 protein regulates the expression of cellular genes involved in immune response.* Sci Rep, 2019. **9**(1): p. 13620.
82. Graham, S.V., *Human Papillomavirus E2 Protein: Linking Replication, Transcription, and RNA Processing.* J Virol, 2016. **90**(19): p. 8384-8.
83. Nakahara, T., et al., *Human papillomavirus type 16 E1circumflexE4 contributes to multiple facets of the papillomavirus life cycle.* J Virol, 2005. **79**(20): p. 13150-65.
84. Doorbar, J., *The E4 protein; structure, function and patterns of expression.* Virology, 2013. **445**(1-2): p. 80-98.
85. Doorbar, J., *The papillomavirus life cycle.* J Clin Virol, 2005. **32 Suppl 1**: p. S7-15.
86. Basukala, O. and L. Banks, *The Not-So-Good, the Bad and the Ugly: HPV E5, E6 and E7 Oncoproteins in the Orchestration of Carcinogenesis.* Viruses, 2021. **13**(10).
87. DiMaio, D. and L.M. Petti, *The E5 proteins.* Virology, 2013. **445**(1-2): p. 99-114.
88. Ganguly, N., *Human papillomavirus-16 E5 protein: oncogenic role and therapeutic value.* Cell Oncol (Dordr), 2012. **35**(2): p. 67-76.
89. Venuti, A., et al., *Papillomavirus E5: the smallest oncoprotein with many functions.* Mol Cancer, 2011. **10**: p. 140.
90. Buck, C.B., P.M. Day, and B.L. Trus, *The papillomavirus major capsid protein L1.* Virology, 2013. **445**(1-2): p. 169-74.
91. Ozbun, M.A. and S.K. Campos, *The long and winding road: human papillomavirus entry and subcellular trafficking.* Curr Opin Virol, 2021. **50**: p. 76-86.
92. Wang, J.W. and R.B. Roden, *L2, the minor capsid protein of papillomavirus.* Virology, 2013. **445**(1-2): p. 175-86.
93. Buck, C.B., et al., *Arrangement of L2 within the papillomavirus capsid.* J Virol, 2008. **82**(11): p. 5190-7.
94. Zheng, Z.M. and C.C. Baker, *Papillomavirus genome structure, expression, and post-transcriptional regulation.* Front Biosci, 2006. **11**: p. 2286-302.
95. Passmore, L.A. and J. Coller, *Roles of mRNA poly(A) tails in regulation of eukaryotic gene expression.* Nat Rev Mol Cell Biol, 2022. **23**(2): p. 93-106.
96. Ramanathan, A., G.B. Robb, and S.H. Chan, *mRNA capping: biological functions and applications.* Nucleic Acids Res, 2016. **44**(16): p. 7511-26.
97. Zhang, Y., et al., *Alternative polyadenylation: methods, mechanism, function, and role in cancer.* J Exp Clin Cancer Res, 2021. **40**(1): p. 51.
98. Chan, S., E.A. Choi, and Y. Shi, *Pre-mRNA 3'-end processing complex assembly and function.* Wiley Interdiscip Rev RNA, 2011. **2**(3): p. 321-35.
99. Ramanouskaya, T.V. and V.V. Grinev, *The determinants of alternative RNA splicing in human cells.* Mol Genet Genomics, 2017. **292**(6): p. 1175-1195.

100. Kajitani, N. and S. Schwartz, *Role of Viral Ribonucleoproteins in Human Papillomavirus Type 16 Gene Expression.* Viruses, 2020. **12**(10).
101. K. Lynch., L.M.-M., *Alternative Splicing*, in *Encyclopedia of biological chemistry*, M.D.L. William J. Lenarz, Editor. 2013, Academic Press. p. 75-80.
102. De Conti, L., M. Baralle, and E. Buratti, *Exon and intron definition in pre-mRNA splicing.* Wiley Interdiscip Rev RNA, 2013. **4**(1): p. 49-60.
103. Staiger, D. and G.G. Simpson, *Enter exitrons.* Genome Biol, 2015. **16**(1): p. 136.
104. Black, A.J., J.R. Gamarra, and J. Giudice, *More than a messenger: Alternative splicing as a therapeutic target.* Biochim Biophys Acta Gene Regul Mech, 2019. **1862**(11-12): p. 194395.
105. 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.
106. Fisette, J.F., et al., *hnRNP A1 and hnRNP H can collaborate to modulate 5' splice site selection.* RNA, 2010. **16**(1): p. 228-38.
107. Shepard, P.J. and K.J. Hertel, *The SR protein family.* Genome Biol, 2009. **10**(10): p. 242.
108. Geuens, T., D. Bouhy, and V. Timmerman, *The hnRNP family: insights into their role in health and disease.* Hum Genet, 2016. **135**(8): p. 851-67.
109. Siculella, L., et al., *A comprehensive understanding of hnRNP A1 role in cancer: new perspectives on binding with noncoding RNA.* Cancer Gene Ther, 2023. **30**(3): p. 394-403.
110. Van Dusen, C.M., et al., *A glycine-rich domain of hnRNP H/F promotes nucleocytoplasmic shuttling and nuclear import through an interaction with transportin 1.* Mol Cell Biol, 2010. **30**(10): p. 2552-62.
111. Honore, B., et al., *Heterogeneous nuclear ribonucleoproteins H, H', and F are members of a ubiquitously expressed subfamily of related but distinct proteins encoded by genes mapping to different chromosomes.* J Biol Chem, 1995. **270**(48): p. 28780-9.
112. Ortuno-Pineda, C., et al., *Binding of hnRNP H and U2AF65 to respective G-codes and a poly-uridine tract collaborate in the N50-5'ss selection of the REST N exon in H69 cells.* PLoS One, 2012. **7**(7): p. e40315.
113. Xu, C., et al., *HnRNP F/H associate with hTERC and telomerase holoenzyme to modulate telomerase function and promote cell proliferation.* Cell Death Differ, 2020. **27**(6): p. 1998-2013.
114. 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.
115. Han, K., et al., *A combinatorial code for splicing silencing: UAGG and GGGG motifs.* PLoS Biol, 2005. **3**(5): p. e158.
116. Ajiro, M., et al., *Serine/Arginine-Rich Splicing Factor 3 and Heterogeneous Nuclear Ribonucleoprotein A1 Regulate Alternative RNA Splicing and Gene Expression of Human Papillomavirus 18 through Two Functionally Distinguishable cis Elements.* J Virol, 2016. **90**(20): p. 9138-52.
117. 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.
118. Liu, Y. and S.L. Shi, *The roles of hnRNP A2/B1 in RNA biology and disease.* Wiley Interdiscip Rev RNA, 2021. **12**(2): p. e1612.

119. 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.
120. Varia, S., et al., *Btf and TRAP150 have distinct roles in regulating subcellular mRNA distribution.* Nucleus, 2013. **4**(3): p. 229-40.
121. Ito, M., et al., *Identity between TRAP and SMCC complexes indicates novel pathways for the function of nuclear receptors and diverse mammalian activators.* Mol Cell, 1999. **3**(3): p. 361-70.
122. Beli, P., et al., *Proteomic investigations reveal a role for RNA processing factor THRAP3 in the DNA damage response.* Mol Cell, 2012. **46**(2): p. 212-25.
123. Jungmichel, S., et al., *Proteome-wide identification of poly(ADP-Ribosyl)ation targets in different genotoxic stress responses.* Mol Cell, 2013. **52**(2): p. 272-85.
124. Wahl, M.C., C.L. Will, and R. Luhrmann, *The spliceosome: design principles of a dynamic RNP machine.* Cell, 2009. **136**(4): p. 701-18.
125. Lee, K.M., W. Hsu Ia, and W.Y. Tarn, *TRAP150 activates pre-mRNA splicing and promotes nuclear mRNA degradation.* Nucleic Acids Res, 2010. **38**(10): p. 3340-50.
126. Yarosh, C.A., et al., *TRAP150 interacts with the RNA-binding domain of PSF and antagonizes splicing of numerous PSF-target genes in T cells.* Nucleic Acids Res, 2015. **43**(18): p. 9006-16.
127. Kasof, G.M., L. Goyal, and E. White, *Btf, a novel death-promoting transcriptional repressor that interacts with Bcl-2-related proteins.* Mol Cell Biol, 1999. **19**(6): p. 4390-404.
128. McPherson, J.P., et al., *Essential role for Bclaf1 in lung development and immune system function.* Cell Death Differ, 2009. **16**(2): p. 331-9.
129. Philipps, D., et al., *Arginine/serine repeats are sufficient to constitute a splicing activation domain.* Nucleic Acids Res, 2003. **31**(22): p. 6502-8.
130. Schaal, T.D. and T. Maniatis, *Multiple distinct splicing enhancers in the protein-coding sequences of a constitutively spliced pre-mRNA.* Mol Cell Biol, 1999. **19**(1): p. 261-73.
131. Smith, C.W. and J. Valcarcel, *Alternative pre-mRNA splicing: the logic of combinatorial control.* Trends Biochem Sci, 2000. **25**(8): p. 381-8.
132. Wu, J.Y. and T. Maniatis, *Specific interactions between proteins implicated in splice site selection and regulated alternative splicing.* Cell, 1993. **75**(6): p. 1061-70.
133. Kohtz, J.D., et al., *Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors.* Nature, 1994. **368**(6467): p. 119-24.
134. Sarras, H., S. Alizadeh Azami, and J.P. McPherson, *In search of a function for BCLAF1.* ScientificWorldJournal, 2010. **10**: p. 1450-61.
135. Wen, Y., et al., *Bclaf1 promotes angiogenesis by regulating HIF-1alpha transcription in hepatocellular carcinoma.* Oncogene, 2019. **38**(11): p. 1845-1859.
136. Merz, C., et al., *Protein composition of human mRNPs spliced in vitro and differential requirements for mRNP protein recruitment.* RNA, 2007. **13**(1): p. 116-28.
137. Bessonov, S., et al., *Isolation of an active step I spliceosome and composition of its RNP core.* Nature, 2008. **452**(7189): p. 846-50.
138. Meinke, P., T.D. Nguyen, and M.S. Wehnert, *The LINC complex and human disease.* Biochem Soc Trans, 2011. **39**(6): p. 1693-7.

139. Haraguchi, T., et al., *Emerin binding to Btf, a death-promoting transcriptional repressor, is disrupted by a missense mutation that causes Emery-Dreifuss muscular dystrophy.* Eur J Biochem, 2004. **271**(5): p. 1035-45.
140. Lee, Y.Y., et al., *BCLAF1 is a radiation-induced H2AX-interacting partner involved in gammaH2AX-mediated regulation of apoptosis and DNA repair.* Cell Death Dis, 2012. **3**(7): p. e359.
141. Ziegelbauer, J.M., C.S. Sullivan, and D. Ganem, *Tandem array-based expression screens identify host mRNA targets of virus-encoded microRNAs.* Nat Genet, 2009. **41**(1): p. 130-4.
142. Nilsson, K., C. Wu, and S. Schwartz, *Role of the DNA Damage Response in Human Papillomavirus RNA Splicing and Polyadenylation.* Int J Mol Sci, 2018. **19**(6).
143. Ajiro, M., et al., *Intron definition and a branch site adenosine at nt 385 control RNA splicing of HPV16 E6*I and E7 expression.* PLoS One, 2012. **7**(10): p. e46412.
144. Zheng, Y., et al., *Heterogeneous Nuclear Ribonucleoprotein A1 (hnRNP A1) and hnRNP A2 Inhibit Splicing to Human Papillomavirus 16 Splice Site SA409 through a UAG-Containing Sequence in the E7 Coding Region.* J Virol, 2020. **94**(20).
145. Hao, C., et al., *hnRNP G/RBMX enhances HPV16 E2 mRNA splicing through a novel splicing enhancer and inhibits production of spliced E7 oncogene mRNAs.* Nucleic Acids Res, 2022. **50**(7): p. 3867-3891.
146. Cui, X., et al., *HnRNP D activates production of HPV16 E1 and E6 mRNAs by promoting intron retention.* Nucleic Acids Res, 2022. **50**(5): p. 2782-2806.
147. Melamud, E. and J. Moult, *Stochastic noise in splicing machinery.* Nucleic Acids Res, 2009. **37**(14): p. 4873-86.
148. Schmitt, M., et al., *Diagnosing cervical cancer and high-grade precursors by HPV16 transcription patterns.* Cancer Res, 2010. **70**(1): p. 249-56.
149. ATCC. *HeLa*. 2023 [cited 2023 30/07]; Available from: https://www.atcc.org/products/crm-ccl-2.
150. ATCC. *SiHa*. 2023 [cited 2023 30/07]; Available from: https://www.atcc.org/products/htb-35.
151. ATCC. *293T*. 2023 [cited 2023 30/07]; Available from: https://www.atcc.org/products/crl-3216.
152. Sherman, L.S., R., *Serum- and calcium-induced differentiation of human keratinocytes is inhibited by the E6 oncoprotein of human papillomavirus type 16.* J. Virol, 1996. **70(5)**: p. 3269-79.
153. 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.
154. 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.
155. Hao, C., et al., *Identification of heterogenous nuclear ribonucleoproteins (hnRNPs) and serine- and arginine-rich (SR) proteins that induce human papillomavirus type 16 late gene expression and alter L1 mRNA splicing.* Arch Virol, 2022. **167**(2): p. 563-570.
156. Nagy, A., *Cre recombinase: the universal reagent for genome tailoring.* Genesis, 2000. **26**(2): p. 99-109.

157. Chou, M.Y., et al., *hnRNP H is a component of a splicing enhancer complex that activates a c-src alternative exon in neuronal cells.* Mol Cell Biol, 1999. **19**(1): p. 69-77.
158. Kajitani, N., et al., *hnRNP L controls HPV16 RNA polyadenylation and splicing in an Akt kinase-dependent manner.* Nucleic Acids Res, 2017. **45**(16): p. 9654-9678.
159. Belaguli, N.S., M.M. Pater, and A. Pater, *Nucleotide 880 splice donor site required for efficient transformation and RNA accumulation by human papillomavirus type 16 E7 gene.* J Virol, 1992. **66**(5): p. 2724-30.
160. Mesplede, T., et al., *p53 degradation activity, expression, and subcellular localization of E6 proteins from 29 human papillomavirus genotypes.* J Virol, 2012. **86**(1): p. 94-107.
161. Olmedo-Nieva, L., et al., *The Role of E6 Spliced Isoforms (E6*) in Human Papillomavirus-Induced Carcinogenesis.* Viruses, 2018. **10**(1).
162. Smotkin, D., H. Prokoph, and F.O. Wettstein, *Oncogenic and nononcogenic human genital papillomaviruses generate the E7 mRNA by different mechanisms.* J Virol, 1989. **63**(3): p. 1441-7.
163. Zheng, Z.M., et al., *Splicing of a cap-proximal human Papillomavirus 16 E6E7 intron promotes E7 expression, but can be restrained by distance of the intron from its RNA 5′ cap.* J Mol Biol, 2004. **337**(5): p. 1091-108.
164. Tang, S., et al., *The E7 oncoprotein is translated from spliced E6*I transcripts in high-risk human papillomavirus type 16- or type 18-positive cervical cancer cell lines via translation reinitiation.* J Virol, 2006. **80**(9): p. 4249-63.
165. Schneider-Gadicke, A. and E. Schwarz, *Different human cervical carcinoma cell lines show similar transcription patterns of human papillomavirus type 18 early genes.* EMBO J, 1986. **5**(9): p. 2285-92.
166. Snijders, P.J., et al., *Human papillomavirus type 33 in a tonsillar carcinoma generates its putative E7 mRNA via two E6* transcript species which are terminated at different early region poly(A) sites.* J Virol, 1992. **66**(5): p. 3172-8.
167. Halec, G., et al., *Biological activity of probable/possible high-risk human papillomavirus types in cervical cancer.* Int J Cancer, 2013. **132**(1): p. 63-71.
168. Nakagawa, S., et al., *Ubiquitous presence of E6 and E7 transcripts in human papillomavirus-positive cervical carcinomas regardless of its type.* J Med Virol, 2000. **62**(2): p. 251-8.
169. Sotlar, K., et al., *Detection of high-risk human papillomavirus E6 and E7 oncogene transcripts in cervical scrapes by nested RT-polymerase chain reaction.* J Med Virol, 2004. **74**(1): p. 107-16.
170. Lungu, O., C.P. Crum, and S. Silverstein, *Biologic properties and nucleotide sequence analysis of human papillomavirus type 51.* J Virol, 1991. **65**(8): p. 4216-25.
171. Ajiro, M. and Z.M. Zheng, *E6;E7, a novel splice isoform protein of human papillomavirus 16, stabilizes viral E6 and E7 oncoproteins via HSP90 and GRP78.* mBio, 2015. **6**(1): p. e02068-14.
172. Toots, M., et al., *The transcription map of human papillomavirus type 18 during genome replication in U2OS cells.* PLoS One, 2014. **9**(12): p. e116151.
173. Van Ranst, M., J.B. Kaplan, and R.D. Burk, *Phylogenetic classification of human papillomaviruses: correlation with clinical manifestations.* J Gen Virol, 1992. **73 (Pt 10)**: p. 2653-60.
174. Chan, S.Y., et al., *Analysis of genomic sequences of 95 papillomavirus types: uniting typing, phylogeny, and taxonomy.* J Virol, 1995. **69**(5): p. 3074-83.

175. Dereeper, A., et al., *BLAST-EXPLORER helps you building datasets for phylogenetic analysis.* BMC Evol Biol, 2010. **10**: p. 8.
176. McLaughlin-Drubin, M.E. and K. Munger, *Oncogenic activities of human papillomaviruses.* Virus Res, 2009. **143**(2): p. 195-208.
177. Howie, H.L., R.A. Katzenellenbogen, and D.A. Galloway, *Papillomavirus E6 proteins.* Virology, 2009. **384**(2): p. 324-34.
178. Hawley-Nelson, P., et al., *HPV16 E6 and E7 proteins cooperate to immortalize human foreskin keratinocytes.* EMBO J, 1989. **8**(12): p. 3905-10.
179. Munger, K., et al., *The E6 and E7 genes of the human papillomavirus type 16 together are necessary and sufficient for transformation of primary human keratinocytes.* J Virol, 1989. **63**(10): p. 4417-21.
180. Le, K.Q., et al., *Alternative splicing as a biomarker and potential target for drug discovery.* Acta Pharmacol Sin, 2015. **36**(10): p. 1212-8.

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 1967

Editor: The Dean of the Faculty of Medicine

A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-508478

ACTA UNIVERSITATIS
UPSALIENSIS
2023