LUND UNIVERSITY

Regulation of HPV16 Early Gene Expression

Zheng, Yunji

2021

Regulation of HPV16 Early Gene Expression

YUNJI ZHENG

DEPARTMENT OF LABORATORY MEDICINE | FACULTY OF MEDICINE | LUND UNIVERSITY

FACULTY OF
MEDICINE

LUND
UNIVERSITY

Regulation of HPV16 Early Gene Expression

Yunji Zheng

DOCTORAL DISSERTATION

by due permission of the Faculty of Medicine, Lund University, Sweden.
To be defended at BMC on 25th, March at 9:00 am.

Faculty opponent
Professor Reet Kurg, Institute of Technology, Tartu University, Estonia

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

Organization LUND UNIVERSITY	Document name Doctoral Dissertation
	Date of issue 25th, March, 2021
Author(s) Yunji Zheng	Sponsoring organization

Title and subtitle Regulation of HPV16 Early Gene Expression

Abstract

Human papillomavirus type 16 (HPV16) is the most common high-risk type and is associated with more than 50% cervical cancer cases as well as a number of head and neck cancer cases. The life cycle of HPV16 is strictly linked to cell differentiation in the squamous epithelium with expression of the HPV16 early genes. The oncoproteins E6 and E7 are essential for preventing apoptosis and inducing proliferation of HPV16 infected cells. As cell differentiation proceeds, viral protein E2 suppresses early promoter p97 and results in downregulation of E6/E7 proteins and activation of HPV16 late gene expression. It is not surprising that the HPV16 E2 gene is inactivated in many HPV16-driven cancer cells. Therefore, the study of HPV16 early gene regulation especially regulation of E1, E2, E6 and E7 expression is important to understand HPV16 infection. These studies may contribute to development of antiviral drugs to HPV16.

The HPV16 E6 protein is translated from unspliced mRNAs, whereas the E7 protein is translated from the mRNAs spliced from HPV16 5'-splice site SD226 to 3'-splice site SA409. The balanced expression of unspliced and spliced mRNAs is important to produce the E6 and E7 proteins that affect the HPV16 life cycle progression and pathogenesis. We have investigated the regulation of splicing from SD226 to SA409. We determined that cellular splicing factors hnRNP A1 and hnRNP A2B1 function as inhibitors of HPV16 E7 expression. Both hnRNP A1 and hnRNP A2 inhibited splicing to SA409 through direct and specific interaction with a C-less RNA element located between HPV16 nucleotide position 594 and 604. Even though inhibition SA409 exerted by hnRNP A1 and hnRNP A2 had different effects. Overexpression of hnRNP A1 inhibited SA409 and resulted in enhanced expression of unspliced E6 mRNAs at the expense of the E7 mRNAs. Overexpression of hnRNP A2 inhibited SA409 but resulted in alternative splicing to 3'-splice site SA742. The splice site SA742 is used for the production of the HPV16 E6^E7, E1, and E4 mRNAs.

Since HPV16 E2 is a transcriptional regulator indirectly downregulates HPV16 E6 and E7 expression via suppressing of HPV16 early promoter p97. HPV16 produces various E2 mRNAs that initiate at either early promoter p97 or late promoter p670 and several splice sites are used. Our study demonstrated that the most efficiently translated E2 mRNA initiates at late promoter p670 spliced from 5'-splice site SD880 to 3'-splice site SA2709. An alternative 3'-splice site named SA2582 and located upstream of the E2 ATG could also produce E2 mRNAs, but it is less efficiently to be translated into E2 protein. These results suggest that the splicing enhancer may be located downstream of SA2709 to regulate E2 mRNA expression. To further understand how HPV16 regulates 3'-splice site SA2709, we introduced a number of deletions and mutations to HPV16 subgenomic plasmids. We identified a 19-nucleotide RNA element consisting of three "AC(A/G)AGG" repeats located downstream of SA2709.This RNA element acted as a splicing enhancer. This splicing enhancer sequence interacted with cellular RNA binding protein hnRNP G to enhance splicing to SA2709 and interacted with hnRNP A1 and A2B1 to inhibit splicing to SA2709. Deletion of the splicing enhancer reduced mRNA splicing to SA2709, and redirected splicing to the upstream 3'-splice site SA2582. The later mRNA has poor ability to produce E2 protein, but it may produce a small protein called E1C. We also to establish a bio-assay that can monitor expression of the various E1 mRNAs and the alternatively spliced E2 mRNAs using either SA2582 or SA2709. We constructed a number of subgenomic HPV16 plasmids containing green fluorescent protein, secreted luciferase, neomycin resistance gene, or thymidine kinase reporter genes. These reporter plasmids produce either E1-fusion proteins or E2-fusion proteins. We observed that the balance between unspliced E1 mRNAs and spliced E2 mRNAs was affected by overexpression of RNA binding proteins SRp30c, Tra2b, hnRNP A1, and hnRNP G. However, further study is required to optimize these reporter plasmids for quantitative analysis of HPV16 E1 and E2 mRNA splicing. This bioassay could be used to enhance our understanding pf HPV16 gene regulation and to identify small molecules that interfere with HPV16 mRNA splicing. Such substances could potentially be used as antiviral drugs to HPV16 infections and HPV16-driven cancers.

Key words HPV16, splicing, mRNA, hnRNP A1, hnRNP G, E6, E7, E1, E2

Classification system and/or index terms (if any)

Supplementary bibliographical information	Language English

ISSN and key title 1652-8220	ISBN 978-91-8021-033-1

Recipient's notes	Number of pages 90	Price
	Security classification	

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature _[signature]_   Date 2021-02-17

Regulation of HPV16 Early Gene Expression

Yunji Zheng

LUND
UNIVERSITY

送给我最爱的父母，妻子以及女儿

Table of Contents

List of Papers

Zheng, Y., Jönsson, J., Hao, C., Shoja, Chaghervand., S., Cui, X., Kajitani, N., Gong, L., Wu, C., Schwartz, S. hnRNP A1 and hnRNP A2 inhibit splicing to HPV16 splice site SA409 through a UAG-containing sequence in the E7 coding region. **J. Virol**., 94: October 2020 Volume 94 Issue 20 e01509-20.

Zheng, Y., Cui, X., Nilsson, K., Yu, H., Gong, L., Wu, C., Schwartz, S. 2020. Efficient production of HPV16 E2 protein from HPV16 late mRNAs spliced from SD880 to SA2709. **Virus Res**. 2020 Aug,2 85: 198004, doi:10.1016/j.virusres.2020.198004

Zheng, Y., et al, 2021. "hnRNP G promotes splicing to the HPV16 E2 mRNA splice site SA2709 through a splicing enhancer consisting of an AC(A/G)AGG-triplicate", **manuscript.**

Zheng, Y., et al, 2021. "A bio-assay for HPV16 E1/E2 mRNA splicing", **manuscript**.

Abbreviations

HPV   Human Papillomavirus
HR   High-Risk
LR   Low-Risk
VLPs   Virus-Like Particles
ORF   Open Reading Frame
pRB   Retinoblastoma-Associated Protein
LCR   Long Control Region
pAE   Early Polyadenylation Signal
pAL   Late Polyadenylation Signal
TAD   Transactivation Domain
CR   Conserved Region
ATM   Ataxia-Telangiectasia Mutated.
DBD   DNA Binding Domain
pre-mRNA   Precursor Messenger RNA
GMP   Guanosine Monophosphate
CPSF   Cleavage/Polyadenylation Specificity Factor
CstF   Cleavage Stimulation Factor
XRN-2   Exo-Ribonulease Enzyme
PAP   Poly-A-Polymerase
PABP   Poly-A-Polymerase Binding Protein
SRE   Splicing Regulatory Element
5'-ss   5'-Splice Site
SD   Splice Donor
3' ss   3'-Splice Site
SA   Splice Acceptor
BPS   Branch Point Sequence
PPT   Polypyrimidine Tract

snRNA	Small Nuclear RNA
snRNP	Small Nuclear Ribonucleoprotein
SF1	Splicing Factor 1
RBP	RNA Binding Protein
ESE	Exonic Splicing Enhancer
ESS	Exonic Splicing Silencer
ISE	Intonic Splicing Enhancer
ISS	Intronic Splicing Silencer
SR protein	Serine/Arginine-Rich Protein
hnRNP	Heterogeneous Nuclear Ribonucleoprotein
RBM	RNA Binding Motif
RRM	RNA Recognition Motif
RS	Serine/Arginine-Rich
CTD	C-Terminal Domain
qRRM	Quasi-RNA Recognition Motif
RGG	Arg-Gly-Gly
RNP	Ribonucleoprotein
NLS	Nuclear Localization Signal
m6A	N6-Methyladenosine
EGF	Epidermal Growth Factor
3'-UTR	3'-Untranslated Region
CSTF64	Cleavage Stimulation Factor-64
eIF	Eukaryotic Initiation Factor
RT-PCR	Reverse Transcription-PCR
sLuc	Secreted Luciferase

Abstract

Human papillomavirus type 16 (HPV16) is the most common high-risk type and is associated with more than 50% cervical cancer cases as well as a number of head and neck cancer cases. The life cycle of HPV16 is strictly linked to cell differentiation in the squamous epithelium with expression of the HPV16 early genes. The oncoproteins E6 and E7 are essential for preventing apoptosis and inducing proliferation of HPV16 infected cells. As cell differentiation proceeds, viral protein E2 suppresses early promoter p97 and results in downregulation of E6/E7 proteins and activation of HPV16 late gene expression. It is not surprising that the HPV16 E2 gene is inactivated in many HPV16-driven cancer cells. Therefore, the study of HPV16 early gene regulation especially regulation of E1, E2, E6 and E7 expression is important to understand HPV16 infection. These studies may contribute to development of antiviral drugs to HPV16.

The HPV16 E6 protein is translated from unspliced mRNAs, whereas the E7 protein is translated from the mRNAs spliced from HPV16 5'-splice site SD226 to 3'-splice site SA409. The balanced expression of unspliced and spliced mRNAs is important to produce the E6 and E7 proteins that affect the HPV16 life cycle progression and pathogenesis. We have investigated the regulation of splicing from SD226 to SA409. We determined that cellular splicing factors hnRNP A1 and hnRNP A2B1 function as inhibitors of HPV16 E7 expression. Both hnRNP A1 and hnRNP A2 inhibited splicing to SA409 through direct and specific interaction with a C-less RNA element located between HPV16 nucleotide position 594 and 604. Even though inhibition SA409 exerted by hnRNP A1 and hnRNP A2 had different effects. Overexpression of hnRNP A1 inhibited SA409 and resulted in enhanced expression of unspliced E6 mRNAs at the expense of the E7 mRNAs. Overexpression of hnRNP A2 inhibited SA409 but resulted in alternative splicing to 3'-splice site SA742. The splice site SA742 is used for the production of the HPV16 E6^E7, E1, and E4 mRNAs.

Since HPV16 E2 is a transcriptional regulator indirectly downregulates HPV16 E6 and E7 expression via suppressing of HPV16 early promoter p97. HPV16 produces various E2 mRNAs that initiate at either early promoter p97 or late promoter p670 and several splice sites are used. Our study demonstrated that the most efficiently translated E2 mRNA initiates at late promoter p670 spliced from 5'-splice site SD880 to 3'-splice site SA2709. An alternative 3'-splice site named SA2582 and located upstream of the E2 ATG could also produce E2 mRNAs, but it is less

efficiently to be translated into E2 protein. These results suggest that a splicing enhancer may be located downstream of SA2709 to regulate E2 mRNA expression. To further understand how HPV16 regulates 3'-splice site SA2709, we introduced a number of deletions and mutations to HPV16 subgenomic plasmids. We identified a 19-nucleotide RNA element consisting of three "AC(A/G)AGG" repeats located downstream of SA2709.This RNA element acted as a splicing enhancer. This splicing enhancer sequence interacted with cellular RNA binding protein hnRNP G to enhance splicing to SA2709 and interacted with hnRNP A1 and A2B1 to inhibit splicing to SA2709. Deletion of the splicing enhancer reduced mRNA splicing to SA2709, and redirected splicing to the upstream 3'-splice site SA2582. The later mRNA has poor ability to produce E2 protein, but it may produce a small protein called E1C. We also wish to establish a bio-assay that can monitor expression of the various E1 mRNAs and the alternatively spliced E2 mRNAs using either SA2582 or SA2709. We constructed a number of subgenomic HPV16 plasmids containing green fluorescent protein, secreted luciferase, neomycin resistance gene, or thymidine kinase reporter genes. These reporter plasmids produce either E1-fusion proteins or E2-fusion proteins. We observed that the balance between unspliced E1 mRNAs and spliced E2 mRNAs was affected by overexpression of RNA binding proteins SRp30c, Tra2b, hnRNP A1, and hnRNP G. However, further study is required to optimize these reporter plasmids for quantitative analysis of HPV16 E1 and E2 mRNA splicing. This bioassay could be used to enhance our understanding of HPV16 gene regulation and to identify small molecules that interfere with HPV16 mRNA splicing. Such substances could potentially be used as antiviral drugs to HPV16 infections and HPV16-driven cancers.

Populärvetenskaplig Sammanfattning

Humant papillomvirus typ 16 (HPV16) är det vanligaste papillomviruset av hög-risk typ och är associerat med mer än 50% av all livmoderhalscancer samt ett antal fall av huvud- och hals-cancer. Virusets livscykel är helt beroende av cellens differentiering i det infekterade skivepiteliet. HPV16 proteinerna E6 och E7 förhindrar att den infekterade cellen går i apoptos och inducerar celldelning i de HPV16-infekterade cellerna. Detta skapar förutsättningar för att HPV16 genomet skall kunna replikeras. När celldifferentiering fortskrider inhiberar HPV16 E2 proteinet den tidiga promotorn p97 vilket nedreglerar uttryck av E6/E7-proteinerna. Celldifferentiering aktiverar även uttryck av de sena HPV16 generna L1 och L2. Det är därför inte förvånande att HPV16 E2-genen är inaktiverad i många HPV16-drivna cancerceller. På grund av deras viktiga roll i virus replikationscykel och vid uppkomsten av cancer, är det viktigt att förstå hur uttryck av HPV16 proteinerna E2, E6 och E7 är reglerat. Dessa studier kan bidra till identifiering av nya mål för antivirala läkemedel mot HPV16 och HPV16-inducerad cancer.

HPV16 E6-proteinet translateras från osplitsat mRNA, medan E7-proteinet translateras från mRNA som splitsats från HPV16 "5'-splice donor" SD226 till "3'-splice acceptor" SA409. En perfekt balans i splitsningseffektivitet är därför nödvändig för att producera tillräcklig av både E6- och E7-protein. Denna balans är viktig i virus replikationscykel, men är även nödvändigt för överlevnad av de cancerceller som HPV16 infektionen kan ge upphov till. Ökad förståelse av E6/E7 mRNA splitsning är därför viktig för att förstå hur HPV16-infektionen fortskrider men även för att förstå hur HPV16 orsakar cancer. Vi har undersökt regleringen av mRNA-splitsning från HPV16 SD226 till SA409. Vi fann att de cellulära splitsningsfaktorerna hnRNP A1 och hnRNP A2 fungerar som hämmare av HPV16 mRNA splitsning och därmed av HPV16 E7-produktion. Både hnRNP A1 och hnRNP A2 inhiberade splitsning till SA409 genom en direkt och specifik interaktion med ett RNA-element beläget mellan HPV16-nukleotidpositionerna 594 och 604. Överuttryck av hnRNP A1 inhiberade SA409 och resulterade i ökat uttryck av icke-splitsat E6-mRNA på bekostnad av splitsat E7-mRNA. Överuttryck av hnRNP A2 hämmade också SA409 men resulterade i alternativ splitsning till SA742 som används för produktion av HPV16 E6^E7-, E1- och E4-mRNA.

Eftersom HPV16 E2 nedreglerar uttryck av HPV16 onkogenerna E6 och E7 genom att inhibera den tidiga HPV16 promotorn p97, är det av intresse att studera även hur uttryck av HPV16 E2 regleras. HPV16 producerar olika E2-mRNA som initieras

vid antingen den tidiga promotorn p97 eller den sena promotorn p670. Dessa mRNA kan splitsas på flera olika sätt. Våra resultat visar att det mRNA som translateras mest effektivt till E2 protein, initieras vid den sena promotorn p670 och är splitsat från SD880 till SA2709. Ett alternativt "3'-splice site" med namnet SA2582, som är beläget uppströms om E2, kan också producera E2-mRNA, men detta mRNA translateras mindre effektivt till E2-protein. Våra resultat antydde att en så kallad "splicing enhancer"-sekvens fanns nedströms om SA2709 och at denna RNA sekvens var nödvändig för produktion av E2-mRNA. För att förstå hur HPV16 reglerar SA2709 introducerade vi ett antal deletioner och mutationer nedströms om SA2709 i subgenomiska HPV16 plasmider. Vi identifierade på så sätt ett 19-nukleotider långt RNA-element bestående av tre "AC (A/G) AGG" sekvensmotiv nedströms om SA2709. Detta RNA-element fungerade som en "splicing enhancer". Denna RNA sekvens interagerade med det cellulära, RNA-bindande proteinet hnRNP G som stimulerade splitsning till SA2709. Det interagerade även med hnRNP Al och A2 som båda hämmade splitsning till SA2709. Deletion av denna RNA sekvens reducerade mRNA-splitsning till SA2709 och omdirigerades splitsning till SA2582. Detta mRNA translaterades ineffektivt till E2-protein, men kan potentiellt producera ett litet HPV16 protein som kallas E1C. Till sist genererade vi en "bio-assay" som kan användas för att studera splitsning av HPV16 mRNA som kodar för E1, E1C eller E2. Denna "bio-assay" var baserad på ett antal subgenomiska HPV16-plasmider med reportergener som antingen uttrycker grönt fluorescerande protein (GFP), "secreted luciferase" (sLuc), neomycinresistens eller tymidinkinas. Dessa reporterplasmider producerar antingen El-fusionsproteiner eller E2-fusionsproteiner. Vi visar att att balansen mellan osplitsat E1-mRNA och splitsat E2-mRNA påverkades av överuttryck av RNA-bindande proteiner såsom SRp30c, Tra2b, hnRNP A1 och hnRNP G. Ytterligare studier krävs dock för att optimera dessa reporterplasmider för kvantitativ analys. Denna "bio-assay" kan användas dels för att öka vår förståelse av HPV16-genreglering, dels för att identifiera små molekyler som inhiberar HPV16-mRNA-splitsning. Sådana substanser skulle potentiellt kunna användas som antivirala läkemedel mot HPV16-infektioner och mot HPV16-beroende cancrar såsom livmoderhalscancer och tonsillcancer.

1 Introduction

1.1 Cancer associated human papillomavirus

Human papillomaviruses (HPV) are non-enveloped small double-stranded DNA virus (**Figure 1.1**), which genome size is about 8000 nucleotides [1, 2]. HPVs infect cutaneous epithelium or mucosal epithelium. Concerning cancer progression, mucosal HPV types typically divide into high-risk (HR) and low-risk (LR) [3, 4]. HPV infection is one of the most common sexually transmitted viral infection, almost all the sexually active individuals are infected by HPVs. Most the HPV infections are cleared by the host immune system within 1-2 years [5]. In some rare cases, however, HPV infections will turn to persistent infections [6], which may develop to precancerous lesions and further progress to cancer if left untreated [3, 6].
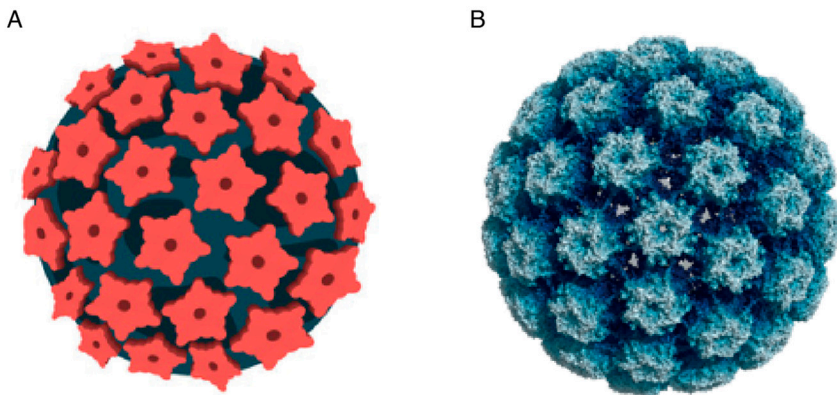
Figure 1.1 Human papillomavirus **(A)**. assembled virion. **(B)**. Cryo-electron microscopy structure of the HPV type 16 viral capsid protein (Adapted from www.wikipedia.com).

About 15% of human cancers are related to viral infection, among which nearly half of the human cancers associated with viral infections are induced by HR HPVs. Each year, approximately 600,000 cases of HPVs infected patients are diagnosed with cancers include cervical, vulva, anal, penile, and oropharynx cancers. Genital cancers, especially cervical cancer, more than 99% cases are highly related to HPVs infections. Moreover, about 95% of anal cancer cases and 80% of head and neck

cancer cases are associated with HPVs infections [7]. The World Health Organization reflects that cervical cancer is one of the greatest threats to women's health; there are more than 300,000 deaths in women each year [8]. More than 99% of cases of all cervical cancer are associated with HR HPVs. HR HPV-16 and HPV-18 are the most common HPV types. HPV16 accounts for more than 50% of cervical cancer cases, while together with HPV18 accounts for approximately 70% of cases [9, 10]. Moreover, more oropharynx cancer cases are reported to be associated with HPVs in recent years, especially HPV-16 type.

At present, vaccination is the most effective way for humans to stay away from the infection of HPVs [11]. All the HPV vaccines are developed based on the purified virus-like particles (VLPs) of the major capsid L1 proteins of different HPVs [12]. One of the HPV vaccines in the market is Gardasil 9, which targets HPV6, 11, 16, 18, 31, 33, 45, 52, and 58 [13]. Even though vaccines are not able to clear existed HPV infections. Thus, HPV infection mechanism, especially HPV gene expression and regulation, is necessary for anti-viral and anti-cancer treatment.

1.2 HPV classification

HPV belongs to the *Papillomaviridae* family, an ancient taxonomic family of DNA viruses [14]. Hundreds of papillomaviruses types identified to infect mammals, birds, snakes, turtles, and fish. To date, there are more than 200 HPVs have been identified and classified [12, 13]. According to the location where HPVs infection occurs, HPV can be grouped into cutaneous HPV types and mucosal HPV types [3, 4]. Cutaneous HPVs commonly cause asymptomatic infections and are cleared by the host immune system. However, some cutaneous HPVs such as HPV5, 9 lead squamous cell carcinomas (SCC) UV radiation in immunosuppressed patients [15]. Mucosal HPV infects epithelium cells mainly, depends on the risk of progressing to cancer in infected cells, mucosal HPV types can be divided into High-risk (HR) and Low-risk (LR) [16].

Characterization of PVs is determined by the capsid protein L1 open reading frame (ORF) region [17]. To date, all HPVs are classified into five genotypes: *alpha*, 65 types including HPV16, 18, 31, etc; *beta*, 53 types including HPV5, 9, etc; *gamma*, 98 types including HPV4, 48, etc; *nu,* one type HPV41; and *mu,* 3 types including HPV1, 63, 204 [17-19]. *Alpha* group, is also known as *alphapapillomavirus,* is the largest group include HPVs that induce benign skin warts, such as HPV6, and the HPVs causing cancer such as HPV16, 18. Alpha-papillomaviruses are the most studied group since they are identified to relate to 5% of cancers worldwide. Alphapapillomavirus are further sub-divided into 13 subgroups (α1-α13) [20]. Most of the HR HPVs belong to α7, such as 18, and α9, such as HPV16. HPVs in betapapillomavirus are found mainly in cutaneous regions and mainly cause

asymptomatic infections in human. However, HPV5 and HPV8 in beta genus are defined as possibly carcinogenic since they can promote the development of keratinocyte carcinoma in immunosuppressed individuals. The HPVs gamma, mu, and nu normally cause cutaneous warts such as hand, flat and plantar warts [21, 22].

1.3 HPV-16 life cycle

HPV infection usually starts at the basal layer of the mucosal epithelium through lesions or wounds. Basal layer cells are actively dividing cells and stem cells alike. HPVs could not replicate their DNA [23, 24]. Therefore, HPV gains access to these basal layer cells by binding to heparan sulfate proteoglycans on the cell surface [25]. The cellular entry mechanism is still remaining to investigate; however, endocytosis is required for virus entry into the cytoplasm while mitosis is required for viral genomic DNA entry into the nucleus [26, 27]. Once infection is established, the HPV viral genome starts to replicate [30]. In the beginning, the viral genome replicates at a low-level with the help of viral protein E1 and E2 by bringing the host cellular DNA replication machinery to the viral genome [28, 29]. During this phase, viral oncoprotein E6 and E7 genes are expressed either. E6 degrades the tumor suppressor protein p53 causing the inhibition of cell apoptosis [16, 30]. E7 binds to tumor suppressor retinoblastoma-associated protein (pRB) and blocks the binding between pRB and its partner E2F, resulting in the activation of the expression of DNA replication factor [29]. Accumulation of E2 will then suppress the expression of E6 and E7 by inhibiting HPV early promoter, which commits the cells start to differentiate and activates late viral promoter 670 [31, 32, 34], consequently viral late capsid protein L1 and L2 are produced. The production of L1 and L2 assembles into progeny of virions and release them in the superficial layer (**Figure 1.2**).

Regularly, HPV infections will be eliminated by the host immune system for up to 2 years. However, in some rare cases of HR HPV infections may lead to persistent infections and cause pre-cancerous lesions to the cervix, which could eventually progress to cervical cancer. Majority of HR HPVs caused malignant tumors or high-grade lesions are accompanied by integration of the HPV viral genome into human cellular chromosomes. Such integration is conducted by disrupting the E2 open reading frame (ORF) [29] and results in high and persistent expression of viral oncogene E6 and E7 but loss of other gene expression. E6*I, E6*II and E6^E7 may also be expressed but their functions are largely unknown [35]. High expression of oncogene E6 and E7 caused by the integration of viral genome may result in a selective growth advantage compared to cells that maintain HR HPV genomes as episomes, which may promote the cancer progression. However, both episomal and integrated forms of HPVs are found in invasive cervical cancer [36, 37].
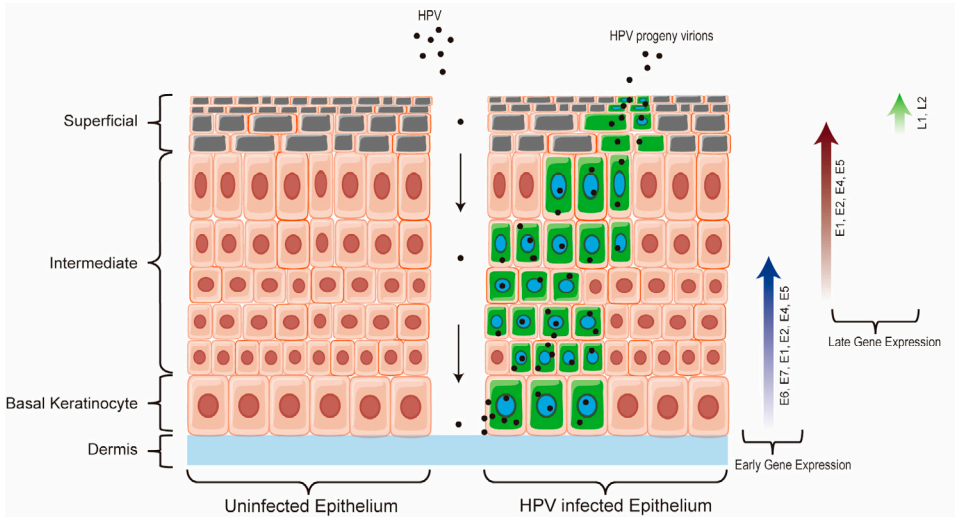
Figure 1.2 Structure of the cervical epithelium and human papillomavirus life cycle.

1.4 Overview of HPV-16 genome and replication

HR HPV-16 is one of the most common HPV types associated with cancers. HPV-16 genome is about 8kb and consists of three regions: the early region, the late region, and the long control region (LCR) [38, 39]. The early promoter p97 located in LCR produces mRNAs encoding all of the early genes (E6, E7, E1, E2, E4 and E5). These mRNAs are polyadenylated at the early polyadenylation signal (pAE) [34]. Once the infected cell differentiates, the late promoter p670 is activated while the early promoter p97 is deactivated resulted in repression of E6 and E7 genes and induces high expression of E1, E2, and E4 mRNAs. The termination of the differentiation decreases the activity of pAE, the late viral protein L1 and L2 and followed by polyadenylation at the late polyadenylation signal (pAL) [2, 39, 40]. LCR is a non-coding region but plays a significant role in DNA replication and gene regulation as it holds the origin of the viral replication and multiple transcription factor binding sites (**Figure 1.3**) [2,41,42].
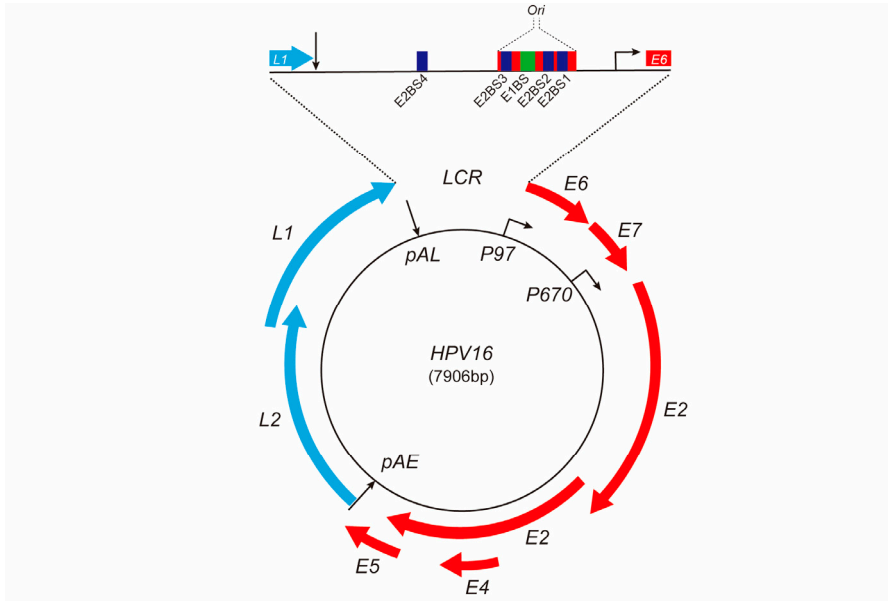
Figure 1.3 Schematic presentation of the HPV16 genome. Ori: The replication origin, LCR: Long control region, E1BS: E1 protein binding site, E2BS: E2 protein binding site, p97: early promoter 97, pAE: Early polyadenylation signal, pAL: Late polyadenylation signal.

So far, it is widely accepted that papillomavirus (PVs) DNA replication, encompasses three phases. Firstly, the entry of PVs to the basal cell layer causes infection of keratinocyte and triggers the initial replication of PV DNA to a low copy number. Secondly, maintenance of the PV genome and the genome is replicated at a constant copy number in proliferating cells. Finally, vegetative amplification of the viral DNA to a high copy number in differentiated cells followed by the production of progeny virions.

1.4.1 Replication initiation

The known significant knowledge of early stage of PVs genome replication cycle mainly originated from the study on the bovine papillomavirus type 1 (BPV-1). Initiation of PVs genome needs the presence of replication origin, E1 and E2 viral proteins. In HPV16, the origin of replication (*Ori*) locates in the non-coding long control region (LCR). This short *Ori* region contains one E1 binding site and three E2 binding sites, and an additional E2 binding site is located further upstream. The E2 binding site sequence in HPV16 *Ori* has been identified as $ACCN_6GGT$. However, the numbers, locations, and the binding sequence of E1 and E2 binding sites within the *Ori* region varies in different PVs. Once the PV infection occurs, the early promoter is activated that initiates the expression of the E1, E2 and other early proteins [43-51]. The E1 viral protein is not only a primary replication protein but also an ATP-

dependent helicase. It melts and unwinds the double helix during replication fork heading to 3' prime to allow binding of the cellular DNA polymerase. However, the E1 protein has poor binding activity and the helicase domain is sequence non-specific [52-54]. The previous study determined that the binding activity of E1 protein needs aid from E2 protein. The E2 protein is a transcription activator that acts as a specificity factor for E1 to promote the binding activity of E1 to *Ori* by inhibiting the non-specific DNA binding activity of the E1 helicase domain. Once E1 and E2 proteins bind to each other, E2 loads E1 protein to *Ori* then replication starts. Afterward, E2 is liberated from E1 by recruiting the second E1 to *Ori.* The released E2 protein will be used by the next replication cycle [55-57].

Initiation of the replication phase of PVs depends on the levels of E1 and E2 protein that is regulated at the transcriptional and translational level. Both E1 and E2 transcripts produce truncated products that potentially inhibit replication, such as E8^E2C protein produced and highly conserved in *alpha*-HPVs. This short protein is known as a transcriptional and replication repressor inhibiting replication via the cellular NCoR/SMART complex or antagonizing to E2 binding site at viral origin to compete for binding activity with E2 protein [58-62]. Moreover, apoptosis mediator p53 is also involved in the regulation PV DNA replication through its oligomerized DNA binding domain. The p53 mediated replication inhibition is through direct binding of p53 to HPV16 E2 protein or p53 competitively binding to the adjacent site on origin overlaps the E2 binding site [63, 64]. Furthermore, the initiation of replication is also regulated by number of cellular factors such as TATA-binding protein, Yin-yang 1 [65].

1.4.2 Stable maintenance replication

After the initiation of replication, the copy number of PVs can range from 50-200 per cell *in vitro*. It is believed that the first stage is rapid and transient. The genome must be maintained at a constant copy number in the dividing basal cells to persistent [66, 67].

It is widely accepted that the viral proteins E1 and E2 are essential in long-term replication maintenance [68-70]. However, it can be argued that E1 may not be necessary for the maintenance of replication [71]. Some studies suggest that the DNA replication can be initiated by host cellular proteins during maintenance stage and that viral proteins are only needed for retention and partitioning of the viral genomes. Previous studies found that the E1 protein is needed for viral genome establishment but is not required later on maintenance phase since these genomes could be maintained as extrachromosomal elements. In addition to E1, it has been reported fully functional viral proteins E6 and E7 are essential for the formation of the extrachromosomal DNA that suggests these oncoproteins may work together to prevent the checkpoints blocking the long-term retention of extrachromosomal DNAs [71, 72].

As for E2 protein, it is clear that the E2 protein is a crucial player in the stable maintenance of replication. The first evidence of the important function of E2 in genome partitioning was reported in the study of BPV1. It showed that E2 binds mitotic chromosomes via protein-protein interactions that are conducted by the E2 transactivation domain (TAD) [73]. The E2 DBD binds to the viral LCR region and tethers it to the condensed mitotic chromosome [73, 74]. Many of the tethering targets for E2 protein have been identified, among which, the Brd4 is well studied [75]. Moreover, The E2 truncated products E8^E2 were reported as replication repressor. Some studies indicated that the viral DNA copy number is up to 20-fold higher in the transformed cells that were transfected by mutants of BPV1 plasmid, unable to produce E8^E2. These results suggest that E8^E2 act as a repressor in viral DNA replication, but they are needed for stable replication [44, 59, 60].

1.4.3 Vegetative replication

The final stage of viral replication is vegetative replication that is taking place only in the differentiated cells in the epithelium. During this stage the viral genomes are amplified to a high copy number destined to be packaged in progeny virions' capsids. However, the mode of vegetative replication is less known, but one can speculate the occurrence of vegetative replication may be resulted from the accumulation of E1 and E2 viral replication protein. Several studies monitored greatly increased levels of the E2 protein in the BPV1 infected and differentiated cancer cells [76,]. Immunological detection of the E1 protein still invalid up to date. Therefore, it is hard to determine the role of E1 protein within the vegetative replication stage [77].

1.5 Viral proteins

1.5.1 E6 and E7

The HPV E6 and E7 proteins are among the earliest HPV proteins and are both essential for viral replication. In addition, they also terminate cell differentiation and contribute to initiation and maintenance of cancer due to their interactions with the cellular tumor suppressor proteins p53 and pRB [39, 40, 78].

The E6 viral protein consists of 150 amino acids and contains a zinc-finger domain in C-terminus and N-terminus. To adapt and to survive in the presence of the host immune system, HR HPV E6 proteins bind to tumor suppressor p53 and causes p53 degradation [78-80] (**Figure 1.4A**). P53 is a transcription factor that possesses DNA binding activity and activates genes involved in DNA damage repair and apoptosis. The E6 protein binds to p53 together with the E6-associated protein (E6AP) causing

ubiquitination of p53 followed by its degradation (**Figure 1.5**). Moreover, E6 directly binds to p53 and negatively interferes with the p53 DNA binding activity leading to transcription inhibition. E6 also interferes with p53 function by inhibiting p300/CBP, which is a histone acetyltransferase, especially for p53 acetylation which determines p53 transactivational function. The inhibition of p300/CBP is caused by E6 indirectly by its downregulation of p53. Furthermore, E6 has been suggested to mask the p53 nuclear localization signal in the C-terminus of p53. LR HPV E6 can bind to p53 but do not cause p53 degradation [81-88].
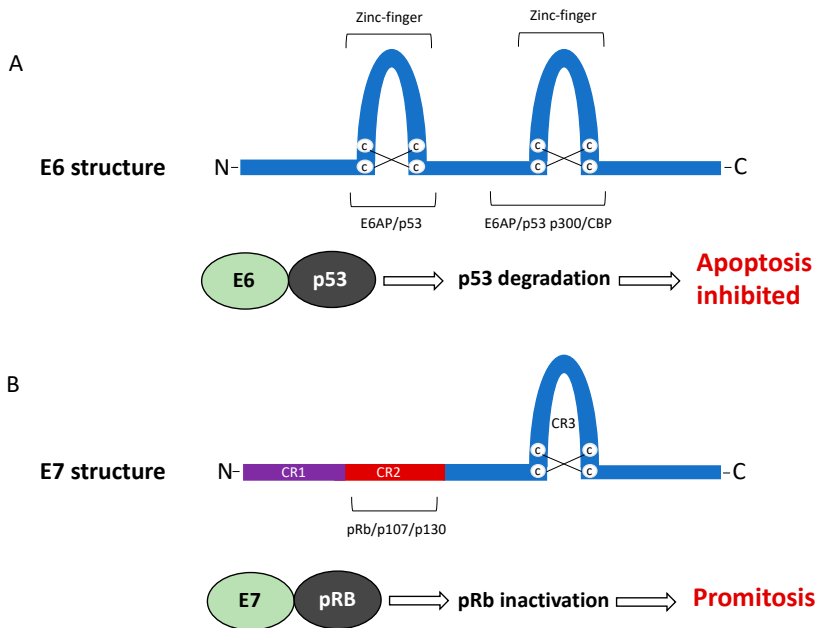
Figure 1.4 Schematic structure of **(A)**. E6 protein, cellular proteins binding sites are indicated. **(B)**. E7 protein, cellular proteins binding sites are indicated.

E7 is a protein of approximately 100 amino acids that is located mainly in the nucleus. E7 consists of three highly conserved regions (CRs): CR1 and CR2 that are located in the N-terminus and CR3 that is located in the C-terminus. The N-terminal CR1 and CR2 interact with the transcription factor E2F and cause destabilization of pRb. C-terminal CR3 is a protein binding domain that has a high binding affinity for pRb [89-91] (**Figure 1.4B**). The tumor suppressor protein pRb plays an essential role in regulating the cell cycle, especially the transition from G1 to S phase. During the G1 phase, pRb is unphosphorylated and suppresses transcription of a number of genes by interacting with transcription factor E2F. As pRb gradually becomes phosphorylated, the cell cycle is forwarded to S phase [92]. The binding of pRB and E2F is prevented by E7 binding to pRB, thereby releasing E2F and activating the S phase- specific genes expression (**Figure 1.5**). HR HPV E7 can degrade pRb and

the other two pocket proteins p107 and p130, which together contribute to progression to cancer in persistent HR HPV infections [93, 94]. Furthermore, E7 has also been reported to bind to protein kinase such as ataxia-telangiectasia mutated (ATM) and interfere with the DNA damage repair, thereby possibly increasing genetic instability of the infected cell [95].
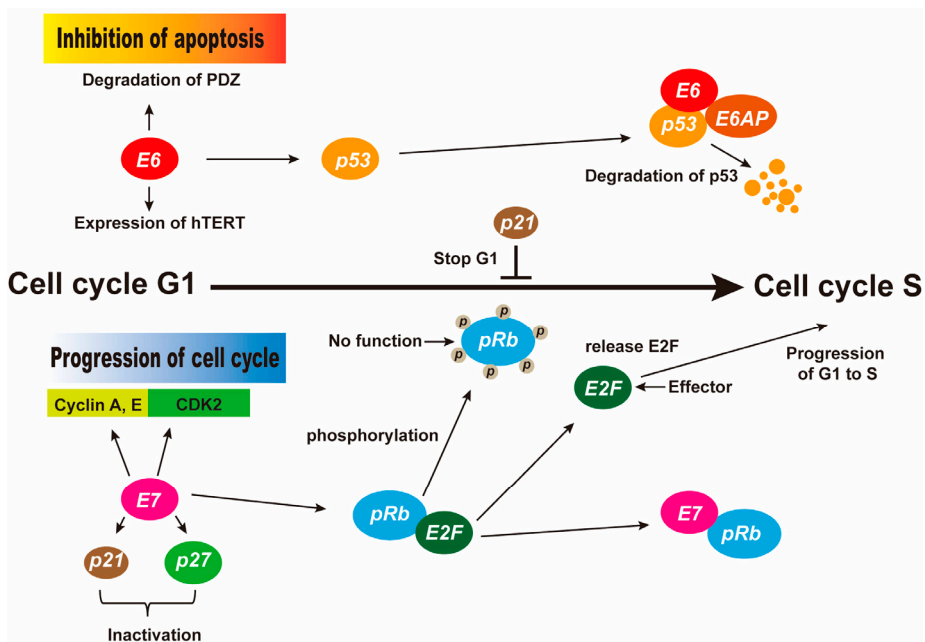
Figure 1.5 Schematic presentation of HPV16 oncoprotein E6 and E7 function of apoptosis signal degradation and cell growth progression.

In general, both viral oncoprotein E6 and E7 play a key role in HR HPV infection and the occurrence of cancers. A balance between E6 and E7 expression determines the infected cellular fate to infection or immortalization and is regulated by post-transcriptional events, mostly alternative RNA splicing. This feature may be an interest for future clinical therapeutic study.

1.5.2 E1 and E2 viral proteins

The viral proteins E1 and E2 are expressed at the early stage of the HPV life cycle, together with E6 and E7, but can also be expressed from the HPV late promoter when E6 and E7 have been shut down. E1 protein together with E2 is necessary for the initiation of viral genome replication. E1 is a 68kD protein, and it locates in both nucleus and cytoplasmic fraction. E1 includes three domains. The N-terminal region contains a regulatory domain and followed with a DNA binding domain (DBD).

DBD can recognize the origin of replication within LCR. Moreover, an ATPase/helicase domain is found in the C-terminal region [52, 96, 97]. E1 is hard to detect due to it is low-level expression. Once viral genome replication initiates, E1 binds to HPV LCR but with a low affinity [98]. E2 binds to the element adjacent to the origin of the replication (**Figure 1.6A**), which increases and stabilizes E1 binding to the origin of replication. With the aid of E2, E1 assembles into double-hexamers and acts as a helicase, which unwinds the viral DNA and brings down host cellular replication machinery associated proteins such as DNA polymerase to the origin of replication [99, 100].

The E2 protein is about 50kDa in size and localizes mainly to the nucleus. The E2 protein is divided into three parts: a conserved N-terminal transactivation domain involved in protein-protein interactions, a C-terminal DNA binding domain, and a so-called "hinge" region by which both domains are linked. The hinge region differs in length and nucleotide composition among different genera of HPVs [101-103]. Previous studies indicated that the HPV16 E2 binding sequence is $ACCN_6GGT$ (**Figure 1.6B**), which is highly conserved in at least nine different HPVs. The various E2 binding sites all contain at least one CpG dinucleotide. Methylation of CpG motifs is one of the major epigenetic modifications in cells and is highly related to transcriptional repression, suggesting that methylation in E2 binding sites may interfere with E2 binding to HPV DNA and its ability to regulate transcription. Several clinical studies indicate the methylation of E2 binding sites represses the HPV early promoter. Unmethylated HPV DNA is required for active replication and transcription during cell differentiation, while integrated HPV DNA may be methylated [104-106].
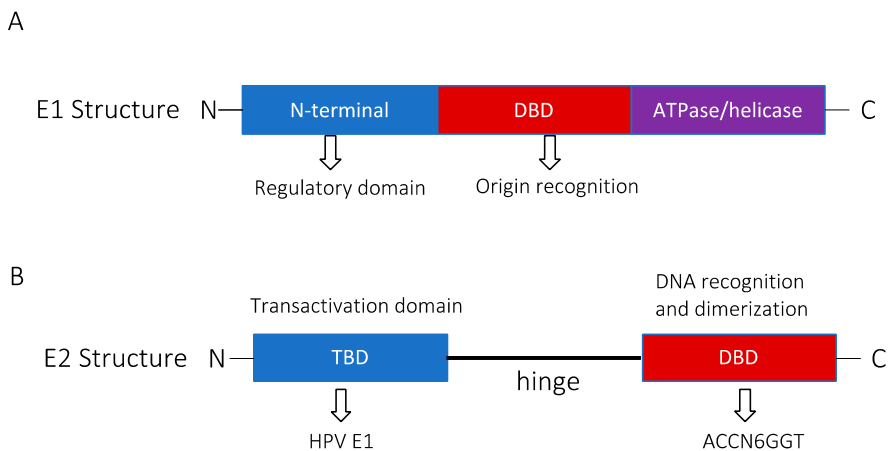

Figure 1.6 Schematic presentation of HPV16 **(A)**. E1 protein, the function of each domain. **(B)** E2 protein, the function of TBD and DBD. DBD: DNA binding domain, TBD: Transcavtivation domain.

E2 protein binds to HPV DNA and acts as a transcription repressor by repressing the early promoter p97. The inhibition of the early promoter p97 leads to downregulation of E6 and E7 expression, eventually leading to cell differentiation and activation of the HPV late promoter p670 [107, 108]. Therefore, the levels of E1, E2, E4, and E5 increase, later on, followed by production of the HPV late proteins L1 and L2. When HPV infected cells progress to cancer, the expression of viral E2 gene is often downregulated or abolished, and it has been speculated that this downregulation increases the expression of E6 and E7 [109]. E2, together with E1, is involved in HPV genome replication as mentioned previously. E2 is also required to partition of HPV genomes to the dividing cells. Although the detailed mechanism remains to be determined, as E2 interacts with cellular chromatin adapter protein Brd4 and tethers viral genome to host chromosomes [110-112]. Early studies suggested that E2 is also involved in the packaging of viral genomes into virions. The role of E2 during this process needs further investigation [113]. E2 also promotes HPV late L1 and L2 gene expression by inhibiting the HPV early polyadenylation signal pAE, thereby resulting in read-through into the late genes [114]. The E2 protein also appears to interacts with multiple cellular RNA-binding proteins and one may speculate that E2 is involved in the regulation of other RNA processing events such as alternative mRNA splicing [115, 116].

1.5.3 E4 and E5 proteins

The HPV E4 protein is the most abundant protein in HPV infected cells, and it is widely expressed in the infected mucosal epithelium. Therefore, E4 may serve as a biomarker for HPV infection. The functions of E4 has been suggested to include modulation of viral genome amplification, virion assembly and release of virions via interaction with cytokeratins to reorganize keratin network in cells. It has been also shown that E4 smodulates cell cycle to arrest at G2/M Although there are many functions ascribed to E4, its role in the HPV life cycle is poorly understood [117-119].

The HPV E5 protein is a hydrophobic membrane-bound protein that it is approximately 83 amino acids in size and may localize at endoplasmic reticulum, Golgi apparatus, perinuclear regions and plasma membrane [120]. E5 promotes the activity of the epidermal growth factor-receptor (EGF-R) pathway via interfering EGFR-recycling and increasing its amount on cell surface, thereby contributing to transformation in murine and human cell culture system [121, 122]. However, in contrast to HPV E6 and E7, the contribution of E5 for the carcinogenesis has been shown as described beforehand though it for the cancer maintenance has been controversial because E5 ORF gets deleted after the integration event in a sub-set of cervical cancers [123-126]. In addition, it has been reported that E5 downregulates MHC class I trafficking to the cell surface from the Golgi apparatus, thereby preventing HPV antigen surface presentation. Presumably to protect the HPV infected cell from the immune response of the host [127, 128].

1.5.4 L1 and L2 proteins

Both L1 and L2 late viral capsid proteins have a calculated molecular weight of about 55kDa. L1 and L2 proteins are produced only at the late stages of the HPV infection since expression of the L1 and L2 mRNAs occurs at transcription level and at RNA processing level with differentiation-dependent manner [129, 130]. L1 is the major capsid protein and L2 is minor capsid protein of the HPV particle. When HPV infection starts, L1 interacts with host cells by binding to heparan sulfate proteoglycan on the cell surface, which leads to a conformational change of HPV capsid. This conformational change allows exposure of L2 amino terminus to the surface of viral capsid. Afterward, L2 is cleaved by a cellular furin protease resulting in a secondary conformational change, which allows virus binding to the cell surface and entry into the cell [131, 132].

The L1 protein can self-assemble into a virus-like particles (VLP) in eukaryotic cells. Production of HPV L1 in yeast or insect cells assemble into VLPs, a property of HPV L1 that has been exploited to generate HPV type specific prophylactic vaccines [133, 134].

1.6 Regulation of HPV16 gene expression

Post-transcriptional modification plays an important role during HPV16 gene expression, which is required for not only viral mRNA stabilization and proteins translation, but also creating viral proteins diversity to accomplish HPV16 replication cycle. The post-transcriptional modification of HPV16 pre-mRNA includes 5' capping, polyadenylation, alternative splicing regulation and so on and is executed by cellular RNA processing machineries.

1.6.1 Pre-mRNA 5' Capping

In all eukaryotes, 5' capping is an important post-transcriptional modification by altering nucleotide on the 5' end of primary transcripts, especially precursor messenger RNA (pre-mRNA). 5' capping is important and required for polyadenylation, pre-mRNA splicing, mature mRNA generation, stabilization, and translation [125, 126]. 5' capping takes place after transcription initiation and formed in three steps. Firstly, the γ-phosphate of 5' triphosphate RNA (5' PPPNPN) is removed by RNA triphosphatase and forming a biphosphate (5' PPNPN); secondly, guanylyltransferase brings a guanosine monophosphate (GMP) residue to 5' biphosphate forming a 5' guano-triphosphate (5' GPPPNPN) [135]; thirdly, guanine-N7 is methylated by mRNA guanine-*N7*-methyltransferase to form 5' cap structure (*m7*GPPPNPN) [136, 137].

1.6.2 Polyadenylation

Polyadenylation is a part of the process for mature mRNA production by adding a poly(A) tail to mRNAs. The length of poly(A) tail can get up to approximately 250 adenine bases in mammalian cells. If 5' capping is a "start -checkpoint" for other biological processing such as translation, the 3' polyadenylation is an "end-checkpoint". Polyadenylation can be conducted in both nuclear and cytosol. Polyadenylation is important to mature mRNA stabilization, mRNA export to cytoplasmic for the following translation in the ribosome [138].
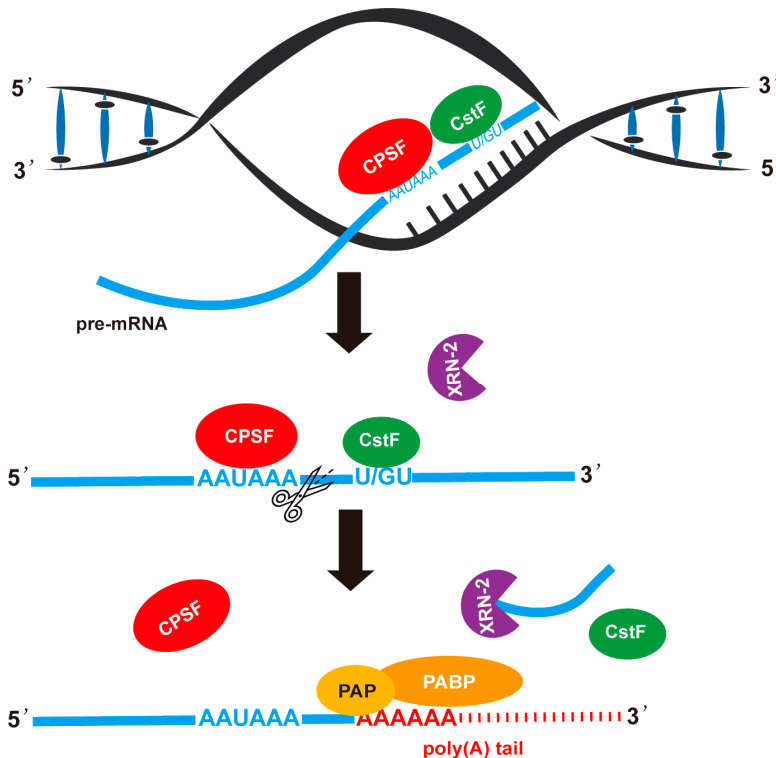
Figure 1.7 Schematic presentation of the pre-mRNA cleavage and polyadenylation process.

In nuclear, the mature 3' ends of mRNAs are created by cleavage then followed by polyadenylation. Pre-mRNAs cleavage/polyadenylation site is normally flanked by two *cis*-acting elements, AAUAAA element and GU-rich region. Upstream AAUAAA *cis*-acting signal is usually located 11 to 30 nucleotides from the cleavage site [139, 140]. Cleavage/Polyadenylation Specificity factors (CPSF), CPSF-160 binds directly to signal AAUAAA followed by CPSF-73 binding which cleaves mRNAs just downstream of AAUAAA signal. CPSFs' activity needs to be stimulated by Cleavage stimulator factor (CstF) which associates with CFI and

CFII. CstF binds to the other *cis*-acting element, U- or GU-rich region downstream of cleavage site. Once the pre-mRNAs are cleaved, the Exo-Ribonulease Enzyme (XRN-2) degrades leftover uncapped mRNA from 5' to 3'. Polyadenylation initiates right away after pre-mRNA cleavage. During polyadenylation, poly-A-polymerase (PAP) is required to catalyze the addition of adenosine monophosphate to mRNA 3' end. Pap's activity needs to be stimulated by nuclear poly-A-polymerase binding protein (PABP) through protein-protein interaction. Once the poly (A) tail reaches certain length polyadenylation stops [141, 142] (**Figure 1.7**).

1.6.3 Alternative splicing

Pre-mRNA splicing is an RNA processing that brings exons together in different combinations by removing introns from pre-mRNA to generate mature mRNAs consisting of exons and encode functional proteins (**Figure 1.8A**). As a common splicing mechanism, alternative splicing creates about 95% multiexonic genes in human and generates several mRNAs from one gene. These alternatively spliced mRNAs may produce an alternative version of the same protein that have altered function, thereby contributing to evolution. Furthermore, alternative splicing is a mechanism that ensures a simple genome to express many different proteins, but alternative splicing is also used to control gene expression. Recent research demonstrated globally altering splicing profiles were found in different cancer pathogenesis, which is originated from direct or/and indirect splicing deregulation, such as abnormal expression of splicing factors, mutation of splicing regulatory elements (SRE) on pre-mRNA and so on. Alternative splicing is a tightly regulated RNA processing event. So far, 5 types of alternative splicing were found including intron retention, exon skipping, mutually exclusive exons, alternative 5'-ss selection, and alternative 3' ss selection [143-145] (**Figure 1.8B**).
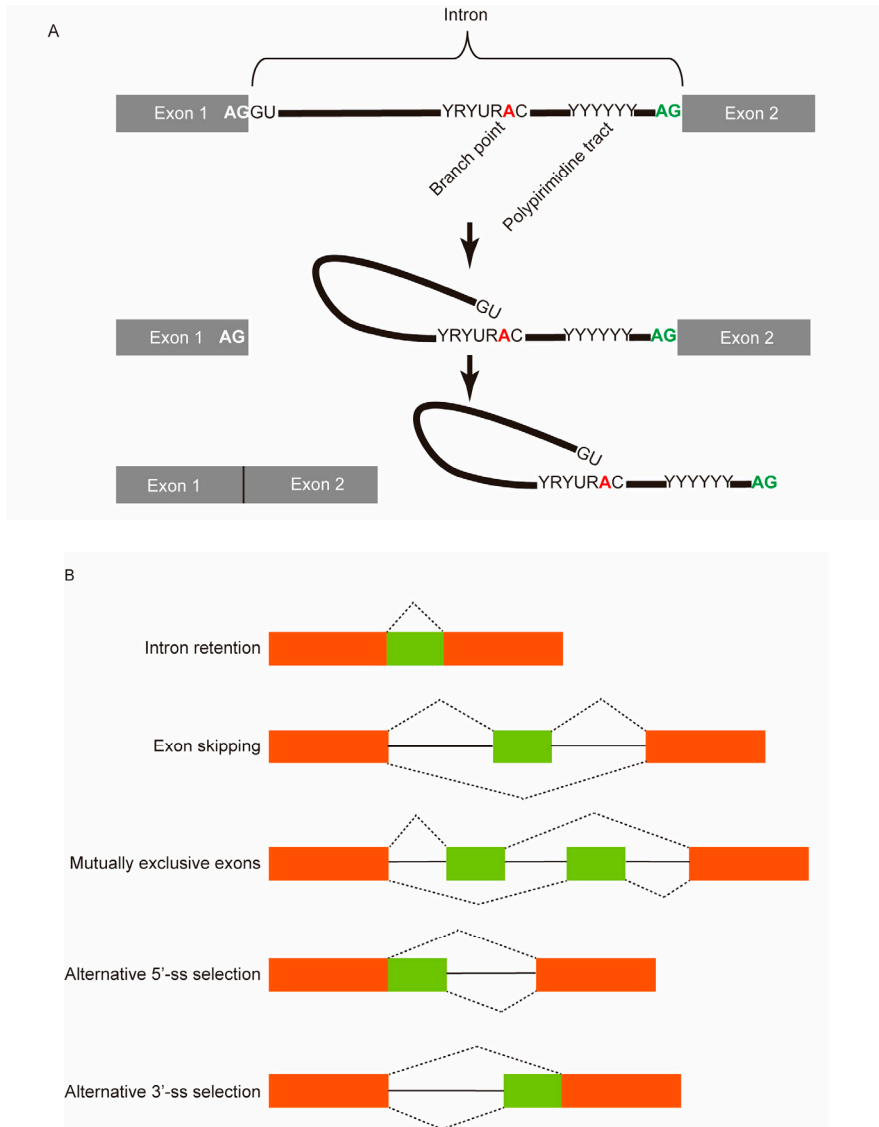
Figure 1.8 (A). Schematic presentation of pre-mRNA structure and mechanism of pre-mRNA splicing. Exons, intron, branch point, and polypyrimidine tract are indicated. Y represents U or C, R represents G or A, N represents any of four nucleotide. **(B).** Schematic presentation of five-type alternative splicing.

1.6.3.1 Structures and elements required for alternative splicing

Junctions of the non-coding intronic region and exonic coding regions are flanked by a 5'-splice site (5'-ss) with a highly conserved invariable GU di-nucleotide, also called splice donor (5' SD), and a 3'-splice site (3'-ss) with a highly conserved invariable AG di-nucleotide, also called splice acceptor (3' SA). In addition to the splice sites, a

branch point sequence (BPS) contains a conserved adenine in all genes is located at 18-40 nucleotides upstream of the 3' ss, and a polypyrimidine tract (PPT) are also required for splicing (**Figure 1.8A and Figure 1.9**). The activity of each splice site is controlled by multiple *cis*-acting regulatory elements in both exons and introns. These RNA elements are termed splicing enhancer or splicing silencers depending on their effect on the splice site they control and function by binding to cellular RNA-binding proteins that interact with the splicing machinery (**Figure 1.9**). Two prominent families of cellular RNA binding proteins that bind splicing enhancers and silencers are the SR protein- and the hnRNP protein-families [144-146].
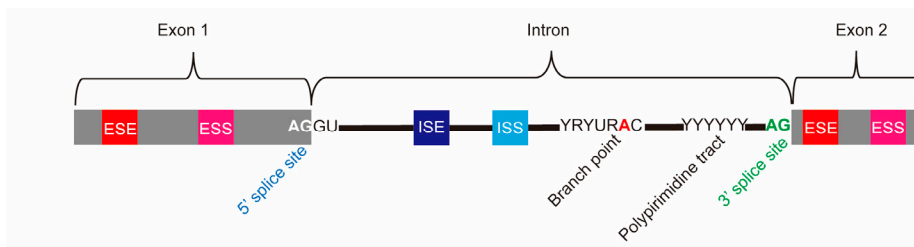
Figure 1.9 Schematic presentation of *cis*-elements and splice site in splicing regulation. ESE: Exonic splicing enhancer, ESS: Exonic splicing silencer, ISE: Intronic splicing enhancer, ISS: Intronic splicing silencer.

1.6.3.2 Brief overview of splicesome assembling and activity

Splicing as well as alternative splicing is catalyzed by an enormous macromolecular complex known as spliceosome. Spliceosome is composed of five small nuclear RNAs (snRNA) U1, U2, U4, U5, and U6. These snRNAs can combine with uridine-rich cellular factors forming RNA-protein complexes called small nuclear ribonucleoproteins (snRNPs) [149]. Several complicated procedures conduct the formation of spliceosome complex included numerous cellular proteins, here showing a brief overview.

Firstly, the U1 snRNP complex binds to 5'-ss GU di-nucleotide, splicing factor 1 (SF1) binds to the BPS while U2AF35 binds to 3'-ss, and U2AF65 binds to PPT. The E early complex then formed [148]. Followed by U2 snRNP complex binding to BPS to release SF1 and ATP is hydrolyzed, pre-spliceosome also known as complex A formed [149]. Tri-snRNP composed of U4, U5, and U6 is recruited by complex A to form pre-catalytic spliceosome, also known as complex B. In complex B, U5 snRNP binds to 3'-ss, while U6 snRNP binds to U2 snRNP, forming helix II. The formation of tri-snRNP is still unclear but may be mediated by protein-protein interaction. Afterward, U1 and U4 snRNP are released from complex B to form complex C, which is a catalytic active spliceosome. The active spliceosome catalyzes transesterification leading the 5'-ss of the intron ligate to conserved adenine in BPS and form a lariat and 5'-ss cleaved. Finally, the U2, U5, and U6 snRNP keep bind to the lariat, but the detached 5'-end of the intron attacks 3'-ss, causing cleavage of 3'-ss, and exons are

ligated. Lariat released and degraded, snRNPs liberated for recycling [150]. Such a mechanism of splicing is termed canonical splicing or lariat pathway. The mechanism conducts more than 99% of splicing. However, the canonical splicing strictly requires that the intronic region is flanked by "GU" di-nucleotide in 5'-ss and "AG" di-nucleotide in 3'-ss (**Figure 1.10**).

Figure 1.10 Schematic presentation of pre-mRNA splicing and spliceosome assembly.

1.7 Molecular basis of splicing regulation

1.7.1 *Cis*-acting elements

The splice site selection of alternative splicing is tightly regulated by short nucleotide sequences are called *cis*-acting elements [151]. *Cis*-acting elements are locat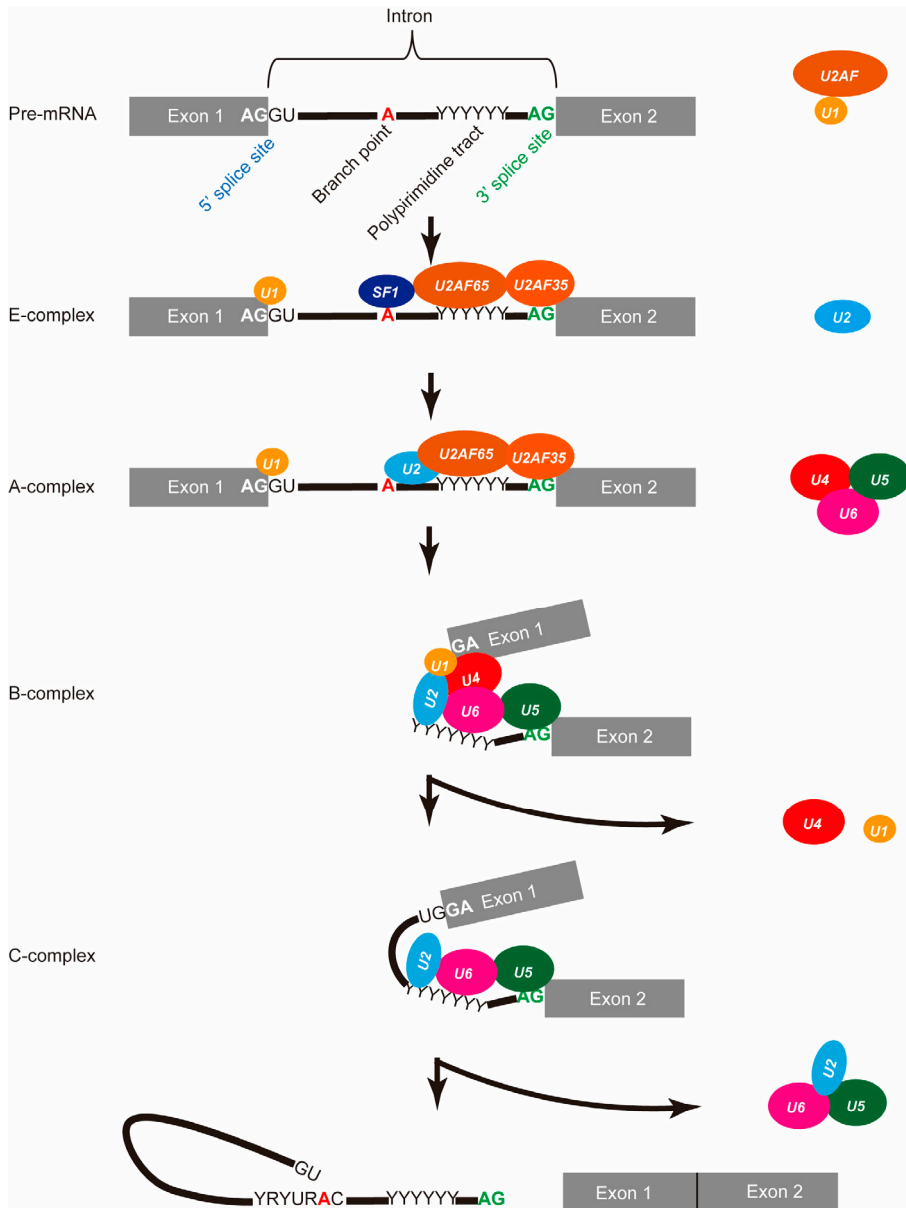ed close to the splice site in both exons and introns of the pre-mRNA and contain binding sites for recruiting RNA binding proteins (RBPs) such as *trans*-acting factors and aiding spliceosome assembly. Regulatory *cis*-acting elements are crucial for spliceosome components recognition and splicing catalysis. These elements located in exons to recruit splicing factors and spliceosome to execute splicing are called exonic splicing enhancers (ESEs), whereas those that regulate splicing negatively are called exonic splicing silencers (ESSs). Similarly, the *cis*-acting elements that are located in introns can be classified as intronic splicing enhancers (ISEs) or intronic splicing silencers (ISSs). *Cis*-acting RNA element could potentially be both enhancer or silencer if it can bind to proteins with either positive or negative effects on splicing [152]. These proteins may be expressed to different levels in different tissues or may compete for binding to the RNA element. Up to date, many RBPs were identified that interact with these *cis*-acting regulatory RNA elements [153]. The perturbation of these elements was reported resulting in different disease and cancer progression (**164**) (**Figure 1.11**).

1.7.2 *Trans*-acting factors

The RBPs or so-called splicing factors that bind to the pre-mRNA *cis*-acting elements close to either exons or introns are called *trans*-acting factors. The *trans*-acting factors together with *cis*-acting elements recruit the spliceosome components to the splice site for further regulating splicing. Two prominent families of splicing factors are known to compose *trans*-acting factors to enhance or repress splicing. Serine/arginine-rich proteins family (SR proteins) bind to both ESE and ISE function as splicing activators, while heterogeneous nuclear ribonucleoproteins (hnRNPs) bind to ESS and ISS function as splicing repressors. However, this division is not strict [154-156]. The molecular mechanisms of several members in both families have been revealed, even though SR proteins and hnRNPs can function oppositely in splicing regulation depending on the positions of splicing regulatory elements relative to the 5'-ss [157]. In addition, other RBPs families are reported involving splicing regulation such as RNA binding motif proteins (RBMs) and RNA helicases [158] (**Figure 1.11**).
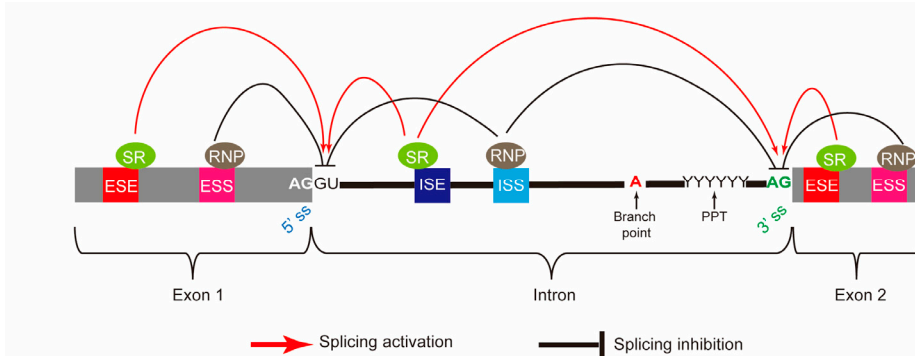
Figure 1.11 Schematic presentation of general model of SR proteins and RNP proteins in splicing regulation.

1.7.3 SR proteins regulate pre-mRNA splicing

SR proteins are multiple functional splicing regulatory factors controlling the constitutive and alternative splicing through protein-protein interaction, and interacting with RNA. Moreover, SR proteins are involved in other cellular processes including RNA exportation [159], translation [160, 161] and mRNA stabilization [162-164]. The structures of SR proteins are highly conserved include at least one RNA recognition motif (RRM), and a serine/arginine-rich domain (RS domain). The RS domain is unique for SR proteins, which makes SR proteins distinguishable from other RBPs [164-166].

In regulating splicing, SR proteins firstly bind to hyperphosphorylated c-terminal domain (CTD) of RNA polymerase II through the RS domain. Once the transcription elongation finished, the SR proteins will move to newly produced pre-mRNA transcripts and stimulate the formation of spliceosome by promoting the binding of U1 and U2AF snRNPs to the pre-mRNA [167]. It will assist U2 snRNP in recognizing and binding to BPS of the intron that is to be cleaved. Furthermore, SR proteins play an important role in alternative splicing splice site selection. They mainly bind to ESEs [168, 169]. However, SR proteins are also capable of binding to ESSs such as SRSF10. hnRNPs and SR proteins may have antagonistic effects on splicing by competing for ESE or ESSs binding in a concentration-dependent manner [170-172].

In HPV16 alternative splicing, SRSF1, SRSF3 and SRSF9 act as splicing regulators of the HPV16 3' ss SA3358 selection that is important to the production of HPV16 E4 protein and in the activation of late gene L1 and L2 expression [173-175]. However, the roles of SR proteins in the regulation of expression of other HPV16 genes remain to be further studied .

1.7.4 hnRNPs regulate pre-mRNA splicing

hnRNPs proteins belong to a large family of RBPs participating in multiple biological activities, including regulation of transcription, translation, alternative splicing [176, 177], and mRNA stabilization [178, 179]. Most of the hnRNPs possess similar features, but the variation in composition results in different functional properties. Some neurodegenerative diseases such as Alzheimer's disease (AD), spinal muscular atrophy (SMA), amyotrophic lateral sclerosis (ALS), and fronto-temporal lobe dementia (FTLD) are related to hnRNPS [180]. Altered expression levels of hnRNPs monitored in cancer patients suggest a role of hnRNPs in cancer pathogenesis. Investigating the role of hnRNPs in carcinogenesis could be meaningful for future cancer treatment development [180].
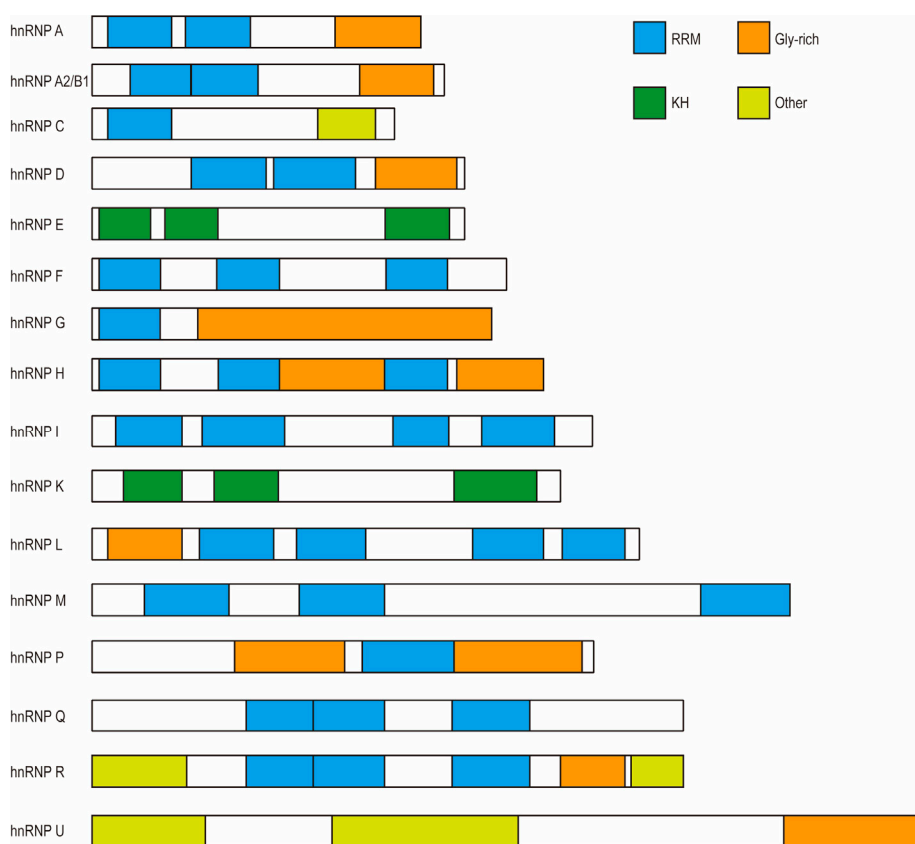

Figure 1.12 Schematic presentation of the proteins in hnRNP family. The blue box represents RRM: RNA recognition motif, Orange box represents gly-rich: glycine-rich domain, Green box represents KH: K homology domain.

Many hnRNPs contain distinct domains, for example RRM, glycine-rich domain, and KH domain [180]. The composition of these domains is not highly conserved and many of the domains are different in size through all hnRNPs, which contributes to functional variation. The RNA binding domain (RBD) in hnRNPs allows RNA binding of the hnRNPs, which further contributes to regulating RNA processing. Up to date, four RBDs were identified, including RRMs, quasi-RRM (qRRM), Arg-Gly-Gly (RGG) tripeptides box containing glycine-rich domain, and K-Homology (KH) domain [180]. The RRMs are the most common RBDs in the hnRNP family, containing two ribonucleoprotein domains (RNP) RNP-1 and RNP-2. They can help hnRNPs to directly interact with RNA. qRRM was found in hnRNP F and H. Unlike the other RRMs in other hnRNPs, qRRM does not have RNP. The glycine-rich domain contains RGG box, which binds to both RNA and ssDNA. KH domain was identified in hnRNP K and F but also found in other RBPs. Furthermore, a nuclear localization signal (NLS) exists in most of hnRNPs, which helps hnRNPs to locate to the nucleus [181-184] (**Figure 1.12**). However, hnRNP A1 may be an exception, that can shuttle in between cytoplasm and nucleus. Post-translational modification such as methylation, phosphorylation, ubiquitination, and sumoylation commonly occurs on hnRNPs, modulating functions and localization of hnRNPs [185]. At present, some structures or domains located on the hnRNPs are poorly studied, which may allow researchers opportunity to detect more functions of hnRNPs in the regulation of gene expression or other physiological processes.

During HPVs gene expression regulation, hnRNPs play an important role by regulating both early and late expression. Based on the previous studies known gene expression regulatory functions of hnRNPs as are the following: hnRNP A1 inhibits the expression of HPV18 oncoprotein E7 by repressing the alternative splicing to 3' ss SA416 [186], hnRNP D, DL, AB, and A2/B1 function as alternative splicing repressors to inhibit HPV16 pre-mRNA splicing to 3' ss SD3632 in the E4 open-reading frame (ORF) [187], hnRNP H stimulates polyadenylation at pAE [188, 189]. Concerning my doctoral research, I will focus more on the introduction of hnRNP A1/A2 and hnRNP G in the text below.

1.7.4.1 hnRNP A1/A2B1

Both hnRNP A1 and A2B1 belong to hnRNP A/B subtype. They are globally and widely expressed hnRNPs. hnRNP A1/A2B1 have a highly conserved N-terminal region that contains two RRMs, and a less conserved c-terminal glycine-rich domain that contains an RGG box and the M9 nuclear targeting signal [190, 191]. The RRMs can affect alternative splicing through RNA-protein interactions. The RGG box offers hnRNP A1/A2B1 both protein and RNA binding abilities to further regulate alternative splicing, but how the RGG box regulates splicing is unclear. The M9 nuclear targeting signal is different from other NLS since it also mediates shuttling of hnRNP A1/A2B1 between nuclear and cytoplasm [190-194].

Posttranslational modifications can modify the function of hnRNP A1, including phosphorylation on Ser4-6 in N-terminus; sumoylation on Lys183 located in the end of RRM2, and the methylation on Arg218/225/232 located in the RGG box [195]. These posttranslational modifications modulate hnRNP A1 RNA binding activity. Similarly, the function of hnRNP A2B1 may be modified by such posttranslational modifications. As the best studied hnRNP, hnRNP A1 has been reported to be highly expressed in different cancers, for example lung cancer, and Burkitt lymphoma [196, 197]. Moreover, hnRNP A1 regulates gene expression and translation of some important factors related to cancer progression, such as the cellular proto-oncogene cMYC [198]. Furthermore, hnRNP A1 was reported to participate the maintenance of telomere length via activation of telomerase, resulting in malignant transformation initiation. hnRNP A2/B1, similarly to hnRNP A1, is also involved in cancer progression through its expression changes [199-201].

In general, both hnRNP A1 and A2B1 have multiple functions, including control of RNA processing, such as alternative splicing, mature mRNA stabilization, translation [202], and N^6-methyladenosine (m6A) modification [203].

1.7.4.2 hnRNP G

The hnRNP G, also called RBMX refers to the hnRNP G-encoding gene of hnRNP G that is located on the X chromosome, while the hnRNP G homolog located on the Y chromosome is called RBMY. hnRNP G is widely expressed in many organs and tissues but varies in concentration. In cervix tissue, the expression of hnRNP G in grandular cells and squamous epithelia cells are high. In contrast, RBMY is mainly expressed in testis and is required for sperm development [204, 205].

The structure of hnRNP G includes only one RRM at its N-terminus, an RBD in the glycine-rich C-terminus, which is seen as an auxiliary structure [206]. hnRNP G does not have a typical RGG box structure, but it has an RGG or RG rich region in the middle part and in the RBD. Furthermore, instead of the classical NLS, a novel transcript targeting domain (NTD) is located in the middle part of hnRNP G. This NTD helps hnRNP G in nuclear targeting and binding to nascent transcripts [207]. The strict nuclear localization of hnRNPG is necessary for its effect on alternative splicing and control of splice site selection [208].

The pathways that hnRNP G adopted in splicing regulation includes indirectly interacting with other cellular splicing factors through its C-terminus or directly interacting with RNA as a splicing regulatory factor. hnRNP G binds to RNA via RRM and RGG motifs, followed by interacting with the phosphorylated C-terminal of RNA polymerase II through the RGG motifs on the hnRNP G low-complexity region. Previous studies indicated that hnRNP G can inhibit or promote exons inclusion, suggesting hnRNP G's role in regulation of splicing is not only as repressor, but also as enhancer [209, 210]. The splicing factor Tra2β was reported to interact with hnRNP G synergistically or antagonistically, resulting in opposite

effects in splicing [209, 211]. Moreover, hnRNP G acts as an m6A reader protein that further affects pre-mRNA processing. m6A modifications that are close to splice sites in pre-mRNA may affect hnRNP G binding, further affecting the RNA polymerase II occupancy patterns and enhancing exon inclusion [212]. In addition to the regulation of splicing, hnRNP G is also crucial to the DNA damage response through two pathways. In the homologous recombination pathway, hnRNP G is enriched at the site where the DNA damage occurs with the help of poly (ADP-ribose) polymerase 1 (PARP-1), resulting in upregulation of repair protein BRAC2 [213]. Then the DNA damage repair starts. Another pathway is dependent on none-homologous end-joining in which hnRNP G RRM binds to the DNA protecting DNA by weakening exonuclease activity [214].

1.8 Regulation of HPV16 gene expression

The regulation of HPV16 gene expression is highly linked to the HPV16 life cycle. The HPV16 life cycle can be divided into early and late stages in which the early promoter p97 is active during the early phase and the late promoter p670 at the late phase. The HPV16 infection starts by HPV16 gaining entry to the host cell's nucleus, in an epithelial cell located at the basal layer of the mucosal epithelium. The cellular transcription machinery is brought down to the HPV16 genome and transcription from the HPV16 from early promoter p97 is activated [23]. With the occurrence of cell differentiation, the HPV16 E2 protein accumulates and shuts down the HPV16 early promoter p97. Consequently, HPV16 E6 and E7 expression are shut off, the cells differentiate and the HPV16 differentiation-dependent late promoter p670 is activated [214]. All the early mRNAs produced by p97 and p670 are polyadenylated at pAE , but high levels of HPV16 E2 protein inhibits HPV16 pAE and contributes to activation of HPV16 late L1 and L2 gene expression. The late mRNAs produced by HPV16 late promoter p670 are polyadenylated at pAL [40].

1.8.1 HPV16 early promoter p97 mediates expression of the early genes

The HPV16 genome contains a large number of splice sites and expresses a large number of alternatively spliced mRNAs to produce all HPV16 proteins in a controlled manner during the HPV16 life cycle. In HPV16, all mRNA isoforms are strictly regulated by alternative splicing. mRNAs encoding the HPV16 oncoproteins E6 and E7 are obtained from the same polycistronic pre-mRNA that gives rise to either unspliced E6 mRNAs or spliced E7 mRNAs by alternative splicing. Several isoforms of spliced E6 and E7 mRNAs are produced by HPV16 (at least 9 isoforms have been identified) [2, 215-217]. The two major E6 and E7 mRNAs are unspliced E6 mRNA and spliced E6*I/E7 mRNAs (**Figure 1.13**). They differ in the inclusion

or exclusion of an intron located between HPV16 splice sites 5'-ss SD226 and 3'-ss SA409. The regulation of alternative splicing of 5'-ss SD226 and 3'-ss SA409 is essential for balancing the production of E6 and E7, in particular the relative levels of these two proteins [218].
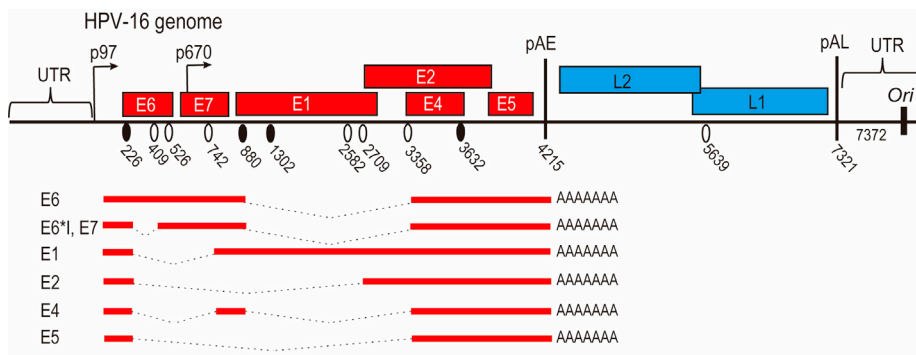
Figure 1.13 Schematic presentation of HPV16 genome and early promoter p97 derived mRNAs

Previous studies have suggested that the epidermal growth factor (EGF), that induces activation of the ERK1-ERK2 pathway inhibits splicing between HPV16 5'-ss SD226 and 3'-ss SA409, thereby giving rise to E6 mRNA production. In contrast, inhibition of EGF increased the splicing and generated E6*I/E7 mRNAs, which primarily encode E7 protein [218]. Furthermore, study indicated that hnRNP A1 and A2/B1 both inhibited splicing between 5'-ss SD226 and 3'-ss SA409 on HPV18, which suggested a similar functional effect may be applied on HPV16 [218].

The HPV16 3'-ss SA3358 is the most commonly used splice site in the HPV16 genome SA3358 is used to generate early mRNAs encoding E4, E6, E7, and possibly E5 (**Figure 1.13**). SA3358 is a weak 3'-splice site which splicing activity is mainly decided by the enhancer downstream of 3'-ss SA3358 [219, 220]. This enhancer is a serine-arginine-rich protein (SR protein) binding sites enriched region, especially serine-arginine-rich factor 1 (SRSF1 is also known as ASF/SF2) [173, 221]. Inactivation of these SRSF1 protein binding sites leads to skipping of the 3'-ss SA3358, suggesting the splicing of 3'-ss SA3358 requires the participation of SRSF1. Moreover, SRSF1 has been reported to be a proto-oncogene. Since SA3358 is used by the majority of the E6 and E7 mRNAs, high levels of SRSF1 upregulates the expression of E6 and E7 and induces cell hyperproliferation and possibly cancer. Adjacent elements downstream of the SRSF1-rich splicing enhancer contains binding sites for SRSF3 and SRSF9. Both SRSF3 and SRSF9 interact with this element resulting in inhibition of SA3358. Nevertheless, the role of SRSF1 in regulating of 3'-ss SA3358 site is probably bigger, because SA3358 is needed also for the production of L1 and L2 mRNAs. Intracellular levels of SRSF1 are increased

by HPV16 E2 expression. High-level of E2 results in high-level of SRSF1 and SRSF1 promotes splicing to Since SA3358 competes with the E2 mRNA-specific 3'-ss SA2709, high-level of SRSF1 may eventually shut down E2 expression [173-175].

1.8.2 HPV16 late promoter p670 mediates early and late genes expression

When HPV16 infected cells step into a differentiation stage, late HPV16 differentiation-dependent promoter p670 is activated. Late promoter p670 can efficiently express E1, E2, and E4 mRNAs, all of which stop at pAE (**Figure 1.14**) [108]. High-level of E2 suppresses the expression of E6/E7 and enhances expression of E1, E2 and E4 by inhibiting early promoter p97, but not p670. This step is also required for the production of L1 and L2. E2 can be detected in infected epithelium primarily in the upper layers of epithelium, whereas E1 is hard to detect [35]. Since both E1 and E2 proteins are required for replication of the HPV16 genome, both E1 and E2 mRNAs must be present in sufficient quantities [220, 221]. While E1 is expressed from unspliced mRNA with a retained intron, E2 is produced from the spliced version of the same pre-mRNAs as E1. Similarly to E6 and E7 mRNAs, a perfect balance in splicing is required to produce both unspliced E1 mRNAs and spliced E2 mRNAs. The ratio of E1 and E2 proteins can affect their ability to replicate the HPV16 DNA genome. Alternative splicing to 3'-ss SA2709 produces E2 mRNAs, while splicing to 3'-ss SA3358 generates E4 mRNAs. Considering both of splice site are used in a mutually exclusive manner, SA2709 and SA3358 are competing. In general, more research is needed to understand the regulation of HPV16 E2 splice site 3'-ss SA2709 [222]. This research would also pave the way for a better understanding of the control of HPV16 E1, E2 and E4 expression.
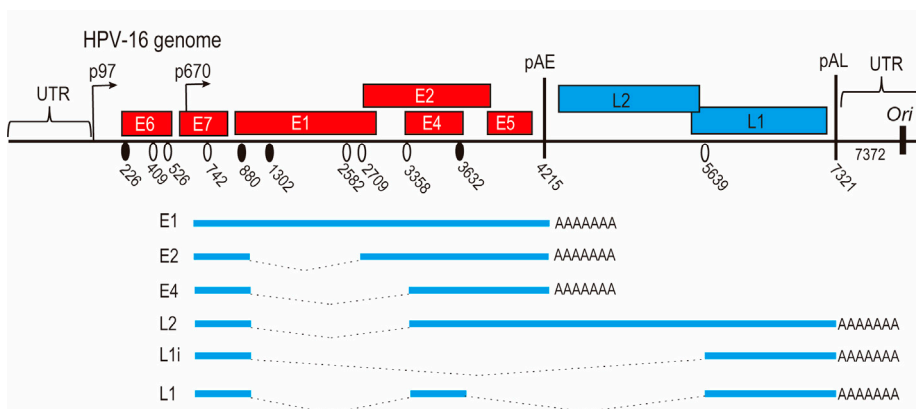

Figure 1.14 Schematic presentation of HPV16 genome and late promoter p670 derived mRNAs

The HPV16 5'-ss SD3632 splicing to 3'-ss SA5639 is required uniquely for L1 mRNA production (**Figure 1.14**). To prevent premature production of the late pre-mRNAs, both splice sites are strictly regulated in a cell differentiation-dependent manner. Hence, active splice sites SD3632 and SA5639 can be found primarily in differentiated cells while the same splice sites are suppressed in cervical cancer cells and mitotic cells. HPV16 5'-ss SD3632 is suppressed through the interactions between various splicing factors and the splicing silencers located adjacent to this splice site. Proteins that suppress HPV16 SD3632 include hnRNP D and hnRNP A2B1, while 3'-ss SA5639 is suppressed by splicing factors hnRNP A1 and hnRNP H [220, 40, 223].

1.8.3 Regulation of HPV16 early and late polyadenylation

The HPV16 early polyadenylation site pAE divides the HPV16 genome into an early region and a late region. The HPV16 early polyadenylation signal (pAE) controls all the early gene expression derived from p97 and p670 and blocks the expression of the late gene at the early stage of HPV life cycle. There are multiple strong polyadenylation regulatory elements located both upstream and downstream of pAE that ensure that early polyadenylation occurs. Deletion of HPV16 early 3' untranslated region (3'-UTR or eUTR) region leads to readthrough into the late region, demonstrating that the eUTR stimulates early polyadenylation. This eUTR regulatory function may be mediated by cellular polyadenylation factors such as FIP1 and hnRNPC1/C2. The vital regulatory RNA elements downstream of pAE are located in the L2 coding region and interact with cleavage stimulation factor 64kDa (CSTF64) and hnRNP H. Moreover, the HPV16 E2 protein acts as an "on/off protein" that contributes to the switch from HPV16 early to late gene expression by inhibiting pAE [224, 189].

The late polyadenylation signal pAL is located downstream of the 3' late UTR (IUTR). IUTR contains several repeats of the $A/GU_{3-5}A/G$ motif and inhibits the HPV16 late gene expression by reducing HPV16 mRNA stability and/or translation. Furthermore, IUTR interacts with U1snRNp or CUGBP to suppress HPV16 late mRNA polyadenylation, as well as mRNA exportation and translation. This inhibitory effect can be reversed by some shuttling proteins such as hnRNP A1 and SRSF1 [173, 225-227].

1.9 HPV mRNA translation

HPVs are obligate intracellular parasites that are completely dependent on the host cell for mRNA translation since HPVs lack their own translation machinery. Translation is a universal process that happens in ribosomes. During translation,

mRNA is translated into a protein in the 5'-3' direction. In general, many cellular factors and mechanisms are involved in the process of translation that can be divided into three steps: initiation, elongation, and termination. Translation initiation is an extremely complex process that includes three steps. Firstly, the Met-tRNA binds to 40s ribosomal subunit forming 43s preinitiation complex. Secondly, this complex binds to the 5' end of the template mRNA with the help of eukaryotic initiation factors (eIFs) and moves along the mRNA in a 5'-3' direction, scanning the mRNA for the first start codon (normally AUG) in an optimal position for translation initiation [235]. An optimal AUG has a G at position +4 (in relation to the "A" in the AUG) and an A/G (purine) at position -3. Sometimes, if the first "AUG" start codon the complex encounters in a weak context (lacking A/G at -3 and/or a G at +4), the weak AUG may be bypassed by the ribosome, that continues to scan the mRNA until it finds another AUG in an optimal context further down. In this case, initiation of the translation may occur also at a downstream AUG and two proteins can be translated from the same mRNA. Thirdly, 60s and 80s ribosomes bind to this complex at start codon on the mRNA and translation starts [235]. Elongation of translation can be simply described as the process in which the ribosome moves along the mRNA, reading codons on the mRNA and bringing the right aminoacyl tRNA to assemble the protein according to the template. Once the ribosome meets a stop codon (UAG, UAA, or UGA), the translation terminates [235].

Both HPV16 early and late proteins are produced from a variety of alternatively spliced mRNAs expressed by the early and late promoters p97 and p670. The structures of many of the HPV16 mRNAs are known and it is clear that all HPV16 mRNAs are polycistronic, encoding multiple ORFs [236]. However, it is not clear what each HPV16 mRNA is translated into. The mRNAs produced from the HPV16 early p97 promoter all contain the E6 AUG at the immediate 5'-end of the mRNA. The E7 ORF immediately follows the E6 ORF. It has been suggested that E7 is translated by ribosomes that miss the E6 AUG and continue to scan the mRNA until they reach the E7 ATG, where translation starts [237]. Alternatively, E7 is produced from an alternatively spliced mRNA in which the E6/E7 pre-mRNA is spliced between SD226 and SA409 in the E6 ORF, thereby shortening the upstream E6 ORF and favoring translation of E7. In the latter case, alternative splicing determines the ratio between HPV16 E6 and E7 expression [236, 237].

Similarly, a recent study indicated that HPV16 E5 is produced from an mRNA initiating at the early promoter p97, followed by splicing from SD226 to SA3358. Since this mRNA contains only short ORFs upstream of E5 (that normally do not interfere with the translation machinery) these results establish that E5 is produced from an early mRNA and therefore primarily is an early protein that cannot be made in significant quantities after the promoter switch in the HPV16 life cycle. In contrast to the E5 mRNA, the E2 protein might be produced from two mRNAs spliced to either 3'-splice site SA2582 or to 3'-splice site SA2709. Both 3'-splice sites are located immediately upstream of the translational start codon of E2. We

wish to identify the splice site that generates the best HPV16 E2 mRNA and determine how this splice site is controlled [231].

Taken together, the main goals of my PhD studies are: to identify the HPV16 mRNAs that are translated into E2 in human cells, to determine how the HPV16 E2 mRNAs are generated and regulated by cellular mechanisms and how splicing of E6 and E7 mRNAs is regulated by cellular factors.

Aim of the thesis

The general aim of my thesis was to investigate the regulation of human papillomavirus type 16 early gene expression at the level of mRNA splicing, in particular expression of the HPV16 replication/transcription factor E2 and the two HPV16 oncogenes E6 and E7.

The specific aims were:

1. To determine how hnRNPA1 and hnRNPA2B1 control expression of the HPV16 oncogenes E6 and E7 at the level of mRNA splicing.

2. To determine the efficiency by which each alternatively spliced HPV16 E2 mRNA is translated into E2 protein.

3. To identify HPV16 *cis*-acting splicing enhancers and cellular *trans*-acting splicing factors that control HPV16 E2 mRNA production at the level of splicing.

4. To establish a bioassay for E1/E2 mRNA splicing that can be used to identify small molecules that could potentially be used as antiviral drugs to treat HPV16 infections.

2 Results

Paper I

hnRNP A1 and hnRNP A2B1 inhibit splicing to HPV16 splice site SA409 through a UAG-containing sequence in the E7 coding region.

HPV16 viral proteins E6 and E7 are necessary for the HPV16 replication cycle. Both E6 and E7 are also needed to induce and maintain malignancy caused by HPV16 persistent infection. The E7 mRNAs are generated from the 5'-ss SD226 to 3'-ss SA409, while the unspliced mRNAs are translated into E6. Therefore, the competition of spliced and unspliced mRNAs are important for balancing the production of E6 and E7 proteins. If more mRNAs are spliced to 3'-ss SA409, the level of the E7 protein will increase while the E6 protein will decrease. High enough level of E7 will result in reduced pRb expression. Levels of E6 mRNAs below the threshold level would result in less E6 protein. Perhaps insufficient for p53 degradation. Cellular factors working at the posttranscriptional level must regulate this balance between E6 and E7 mRNAs. Previous studies had identified hnRNP A1 and hnRNP A2B1 as regulators of splicing of HPV16 late mRNAs and HPV18 E6 and E7 mRNAs. We then hypothesized that hnRNP A1 and A2B1 might affect HPV16 E6 and E7 mRNA splicing. To study the mechanism of HPV16 oncoproteins E6 and E7 mRNA splicing regulation is important for future potential development of new drugs for anti-viral or anti-cancer treatment.

hnRNP A1/A2B1 inhibited HPV16 mRNAs splicing to 3'-ss SA409 and resulted in the promotion of different alternatively spliced or unspliced mRNAs
We started with the transfection of the HPV16 subgenomic reporter plasmid pC97ELsL. The transfection of pC97ELsL into HeLa cells and the reverse transcription-PCR (RT-PCR) results indicated that various alternatively spliced HPV16 transcript isoforms encoding both early and late genes can be detected by using specific primer pairs. To monitor the effect of hnRNP A1/A2B1 on E6 and E7 mRNAs, both hnRNP A1 and hnRNP A2B1 plasmids were serially diluted and cotransfected with pC97ELsL. The RT-PCR results generated by using specific primers for E6 and E7 region indicated that overexpression of hnRNP A1 and A2B1 inhibited splicing to 3'-ss SA409 and SA526. However, the high-level of hnRNP A1 resulted in a remarkable increase in unspliced E6 mRNA. In contrast, the high-

level of hnRNP A2 resulted in a significant increase in mRNA splicing from 5'-ss SD226 to 3'-ss SA742 (**Figure 2.1**).
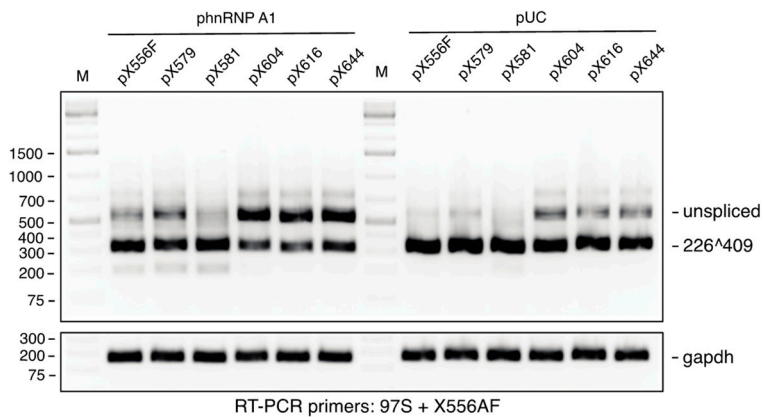
Figure 2.1 Mapping of the hnRNP A1 target site in the HPV16 E7 coding region to a splicing-inhibitory HPV16 RNA sequence located between HPV16 genomic nucleotide positions 579 and 604. RT-PCR on RNA extracted from HeLa cells transfected with plasmid pX556F, pX579, pX581, pX604, pX616, or pX644 in the absence (pUC) or presence of phnRNP A1. The bands representing unspliced HPV16 E6 mRNAs and HPV16 mRNAs spliced from SD226 to SA409 are indicated to the right. cDNAs were amplified by RT-PCR primers 97S and X556AF.

hnRNP A1 and hnRNP A2B1 act on "UAG"-containing splicing inhibitory RNA sequences in the HPV16 E7 coding region

To identify the target site on HPV16 E6 and E7 mRNAs for hnRNP A1 and hnRNP A2B1 binding, we firstly constructed several subgenomic HPV16 expression plasmids from pC97ELsL, including pX1200, pX1060, pX960, pX856F, pX644, pX616, pX604, pX581, pX579, and pX556F. The digits in each deletion plasmid name represent the nucleotide position (according to HPV16 reference sequence) of the endpoint of HPV16 sequences in the plasmids. Cotransfection of these deletion plasmids with hnRNP A1 into HeLa cells and RT-PCR results showed that overexpression of hnRNP A1 did not affect the splicing from plasmid pX581, but inhibited the splicing of pX604 further indicating that the splicing inhibitory RNA elements are located between HPV16 nucleotides 581 to 604. This inhibitory RNA element contains a UAG motif that has been reported as an hnRNP A1 binding motif. Thus, we introduced mutations to this UAG motif in pX604 to generate pM604. The RT-PCR analysis of co-transfection pX604 or pM604 with or without the presence of hnRNP A1 or hnRNP A2B1 demonstrated that the UAG motif was needed for splicing inactivating further suggesting that hnRNP A1 and A2B1 may recognize UAG as the binding motif. We then designed five biotinylated single stranded RNA oligos as shown (**Figure 2.2A**). Oligo 604BS1 does not contain UAG motif, while oligo 604AS1 is a UAG positive control. Point mutations were introduced into UAG or upstream of UAG motif (604AM1, AM2, and AM3) as

shown. RNA oligo pulldown of proteins from HeLa nuclear extract followed by Western-blotting with hnRNP A1/A2B1 antibodies demonstrated both hnRNP A1 and A2B1 binding to wild type UAG motif but not to mutations (**Figure 2.2B**). In addition, more assays were applied to further confirm UAG is the binding site for hnRNP A1 and A2B1, but not shown here (details in Paper I).
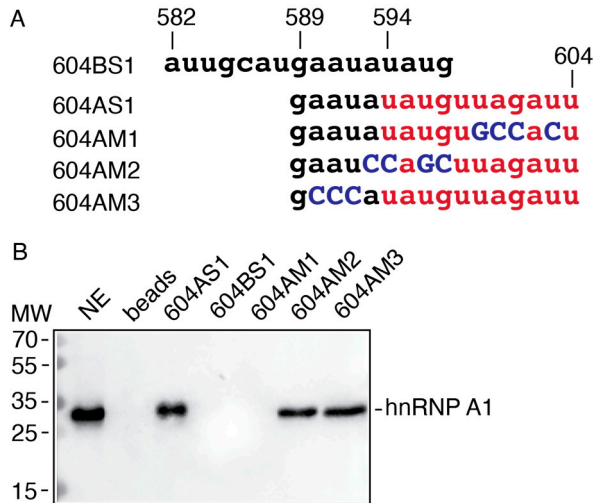
Figure 2.2 Nucleotide substitutions abolish pulldown of hnRNP A1 by HPV16 RNA oligonucleotides. **(A)**. Sequences of biotinylated HPV16 RNA oligonucleotides used in RNA-mediated protein pulldowns of proteins from HeLa cell nuclear extracts. **(B)**. Wester-blotting of hnRNP A1 pulldown by Streptavidin-coated beads. Filter was stained by hnRNP A1 antibody.

hnRNPA1 C-terminus is needed for inhibition of 3'-ss SA409 and production of unspliced E6 mRNAs

hnRNP A1 and hnRNP A2B1 originate from the same hnRNP AB family. Therefore, they are structurally similar to each other. N-terminus of hnRNP A1 and A2B1 are highly conserved, containing two RRMs, while the glycine-rich C-terminus is less conserved, harboring an RGG box and an NLS. Considering that both hnRNP A1 and A2B1 have the common function in affecting HPV16 E6 and E7 mRNA splicing, but with different results, we wished to identify the protein domains that were involved in the control of HPV16 mRNA splicing. We hybridized hnRNP A1 N-terminus to hnRNP A2B1 C-terminus or opposite to generate pA1/A2 or pA2/A1 (**Figure 2.3A**). Both hybrid plasmids were HA-tagged and confirmed by Western-blotting (**Figure 2.3B**). We next cotransfected these hybrid plasmids with subgenomic HPV16 plasmid pC97ELsL. The mRNA levels from different HPV16 mRNA splicing patterns were analyzed by RT-PCR. The hnRNP A1 N-terminus hybridized to hnRNP A2 C-terminus inhibited the mRNAs spliced from SD226 to SA409, but the unspliced E6 mRNA increased, although the efficiency was less than

with wild type hnRNP A1. Replacement of hnRNP A1 C-terminus with hnRNP A2 C-terminus reduced the inhibition of splicing, pointing to the importance of hnRNP A1 C-terminus in splicing inhibition and production of unspliced E6 mRNA. On the other hand, when N-terminus derived from hnRNP A2B1 was hybridized to hnRNP A1 C-terminus, the inhibition of mRNAs spliced from SD226 to SA409 and the promotion of production of unspliced E6 mRNA were similar to the effect of wild type hnRNP A1, further confirming that the C-terminus of hnRNP A1 is necessary to balancing unspliced E6 mRNA and E7 mRNA spliced from SD226 to SA409.
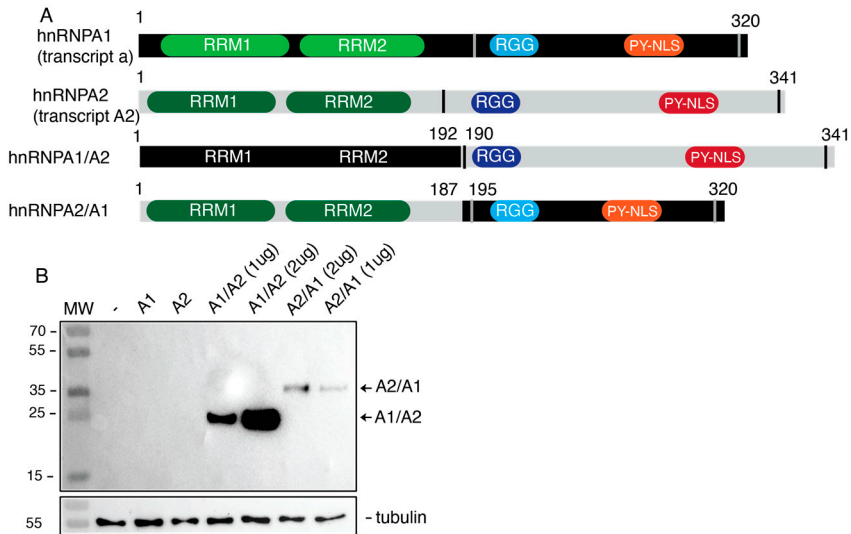
Figure 2.3 C-terminus of hnRNP A1 is required for efficient inhibition of splicing between HPV16 5'-splice site SD226 and 3'-splice site SA409 in the E6 coding region. **(A)**. Schematic representation of the hnRNP A1 and A2B1 proteins and the A1/A2 and A2/A1 hybrids. RNA recognition motifs 1 and 2 (RRM1 and RRM2), RGG, and pyrimidine-rich nuclear/cytoplasmic localization sequence (PY-NLS) (also termed M9) are indicated. hnRNP A1 consists of 320 amino acids and hnRNP A2 of 341 amino acids. The break points of the A1/A2 and A2/A1 hybrids are indicated. **(B)**. Western blotting on extracts from HeLa cells transfected with plasmids expressing the hnRNP A1/A2 and hnRNP A2/A1 hybrids. Blots are stained with anti-HA tag antibodies that detect the HA-tagged hnRNP A1/A2 and hnRNP A2/A1 hybrids but not the untagged hnRNP A1 and hnRNP A2B1 proteins. MW, molecular weight marker.

Overexpression of hnRNP A1 inhibits splicing from SD2226 to SA409 from episomal HPV16 DNA and knock down of hnRNP A1 has opposite effect.

To study if overexpression of hnRNP A1 and hnRNP A2B1 could inhibit splicing from SD226 to SA409 and promote the production of unspliced E6 mRNA in the context of the complete episomal HPV16 genome, plasmid pHPV16AN was used. pHPV16AN contains and encodes the whole HPV16 genome flanked by two lox sites. The presence of lox sites can be cleaved by CRE enzyme and release the HPV16 genome that forms the episomal form. We next cotransfected pHPV16AN with CRE-expressing plasmid with or without the participation of hnRNP A1 or

hnRNP A2B1 expression plasmids. The overexpression of hnRNP A1 and hnRNP A2B1 inhibited the E7 mRNAs splicing from SD226 to SA409 and upregulated the unspliced E6 mRNAs (**Figure 2.4A**). On the contrary, when the hnRNP A1 knock-down was applied, the increase of the mRNAs splicing from SD226 to SA409 was observed. Furthermore, we applied siRNA-mediated knockdown assay on HPV16 positive cancer cell line SiHa and found that the E7 mRNAs spliced from SD226 to SA409 increased and so did E7 protein levels.
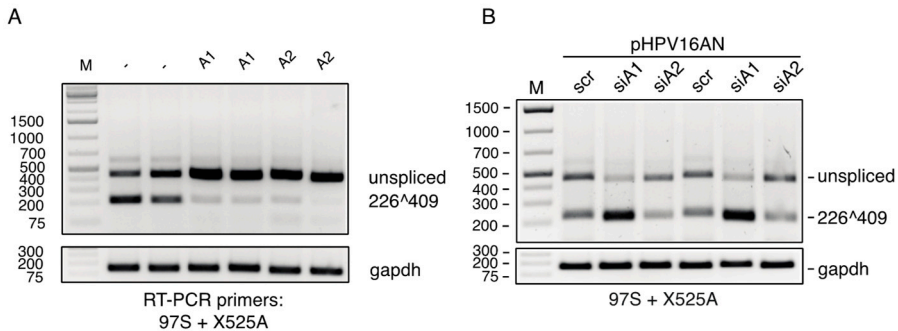
Figure 2.4 hnRNP A1 and A2 inhibit splicing of HPV16 E6 and E7 mRNAs produced from the episomal HPV16 DNA genome. **(A)**. PCR with primers 16S and 16A on Hirt DNA extracted from HeLa cells transfected with the indicated HPV16 plasmids in the presence or absence (-) of the cre-expressing plasmid pCAGGSnlscre. **(B)**. pHPV16AN (40) and pCAGGS-nlscre were cotransfected in duplicates with scrambled (scr) siRNAs or siRNAs to hnRNP A1 (siA1) or hnRNP A2B1 (siA2), RNA was extracted, and RT-PCR was performed with primers 97S and X525A.

Taken together, we concluded that hnRNP A1 and hnRNP A2B1 inhibit splicing to the HPV16 splice site SA409 through a UAG-containing sequence in the E7 coding region.

Paper II

Efficient production of HPV16 E2 protein from HPV16 late mRNAs spliced from SD800 to SA2709

HPV16 produces plenty of transcripts by selecting different splice sites. A number of alternatively spliced transcripts have the potential to be translated into E2 proteins. These mRNAs are mainly divided into two groups. One group starts at early promoter p97 spliced from 5'-ss SD880 to 3'-ss SA2582 or SA2709. Another group starts at late promoter p670 and are spliced from 5'-ss SD880 to 3'-ss SA2582 or SA2709. The E2 viral protein plays animportant role in the entire HPV16 life cycle, including regulation of DNA replication, transcription and HPV16 early and late gene expression. Little is known about the regulation of HPV16 E2 gene expression, primarily since there are many alternatively spliced mRNAs that encode E2. It is hard to predict how much E2 protein each mRNA actually produces. Thus, we wished to establish E2 cDNA reporter plasmids with secreted luciferase (sLuc) in place of the E2 ORF to identify the most efficiently translated E2 mRNA.

The most highly expressed HPV16 E2 mRNAs are spliced from 5'-ss SD880 to 3'-ss SA2709

The HPV16 generates several mRNAs that are potentially translated into E2 proteins. These mRNAs share either 5'-ss SD226 or SD880 and are spliced to either 3'-ss SA2582 or SA2709. To identify the most highly expressed HPV16 E2 mRNA, we transfected HeLa cells with pC97EL or pHPV16 with the presence of the CRE-expressing plasmid. The total RNA was extracted from HeLa cells and analyzed by RT-PCR. Primers 773S and E2A were used to detect one strong band representing splicing from SD880 to SA2709 and a weaker band locating higher up representing the mRNA splicing from SD880 to SA2582. All PCR bands were extracted, purified and followed by sequencing. Consistent results were obtained from the HPV16 positive tonsillar cancer cell line HN26.
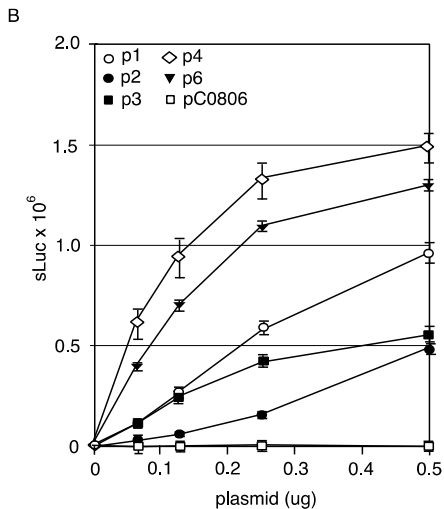
Figure 2.5 (A). Schematic representation of the HPV16 E2 cDNA expression plasmids. The HPV16 sequences present on the cDNAs in each plasmid are indicated. Names of cDNA expression plasmids are indicated to the right. E2/sLuc represents the fusion of the secreted luciferase coding sequence in frame with the E2 ATG. The open reading frames upstream of the E2 ATG are indicated. The HPV16 E6*I-, E6*IV- and E6^E7-orfs all have the E6 ATG and are generated by splicing from HPV16 SD226 to SA2709 (E6*IV), SD226 to SA742 (E6^E7) or SD226 to SA409 (E6*I). The HPV16 E1* orf has the HPV16 E1 ATG and is generated by splicing from SD880 to SA2709. Two HPV16 short upstream orfs with ATGs (uORF1 and uORF2) on the mRNA produced form the p4 cDNA are indicated. The full-length E7 orf is also indicated. **(B)**. Secreted luciferase enzyme activity (sLuc) in the cell culture medium at 24 h after triplicated transfections of HeLa cells with the indicated plasmids.

The most efficiently translated HPV16 E2 mRNA initiates at late promoter p670 and is spliced from 5'-ss SD880 to 3'-ss SA2709

To identify the HPV16 mRNAs most efficiently translated into E2 protein, we constructed CMV-driven cDNA expression plasmids with the most common E2 mRNAs as shown in **Figure 2.5A**. For the convenience of monitoring all E2 mRNA isoforms at the translational level, the sLuc encoding gene replaces the E2 ORF immediately downstream of E2 start codon (ATG). The constructed plasmids were transfected to HeLa cells in comparison to CMV-driven sLuc negative plasmid pC0806 and CMV-driven sLuc positive plasmid p6. The sLuc level in the cell culture medium was monitored 24hrs post transfection and are shown in **Figure 2.5B**. The results are indicating that the mRNAs that initiate at late promoter p670 and are spliced from SD880 to SA2709 is the most efficient translated. To confirm the sLuc assay reflected the translation of the entire E2 ORF, we constructed two E2 expressing plasmids p97E2 and p670E2. These plasmids were transfected into HeLa cells and protein was extracted and analyzed by Western-blotting. As shown, p670E2 produced higher level of E2 protein than p97E2 (**Figure 2.6A**), while mRNA levels produced from both plasmids were in similar level (**Figure 2.6B**).
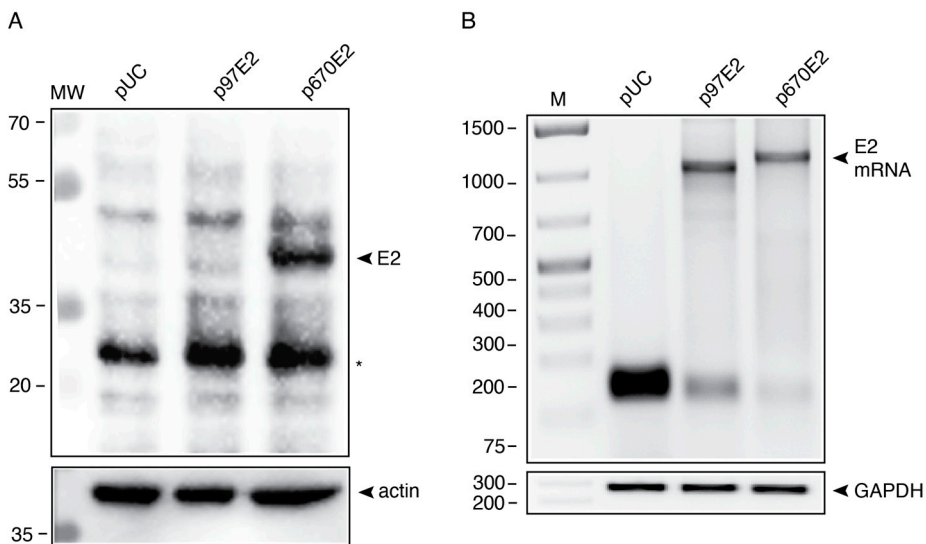
Figure 2.6 (A) Western blotting for E2 protein on cell extracts obtained from HeLa cells transfected with the indicated plasmids, using anti-HPV16 E2 specific antibody as described in Materials and Methods. **(B)** RT-PCR with primers CSENSE and PANTI on total RNA extracted from HeLa cells transfected with the indicated plasmids.

Furthermore, the 3'-ss SA2582 is a competitive splice site to SA2709 and can potentially be translated into E2 To compare mRNAs spliced to SA2582 with mRNAs spliced to SA2709, we constructed a cDNA expressing plasmids initiating at p670 and was spliced from SD880 to SA2582 (**Figure 2.7A**). sLuc assay was applied to monitor the sLuc activity. Results indicated that the HPV16 mRNA spliced from SD880 to SA2709 was more efficiently translated into E2 protein than the mRNA splicing from SD880 toSA2582 (**Figure 2.7B**).
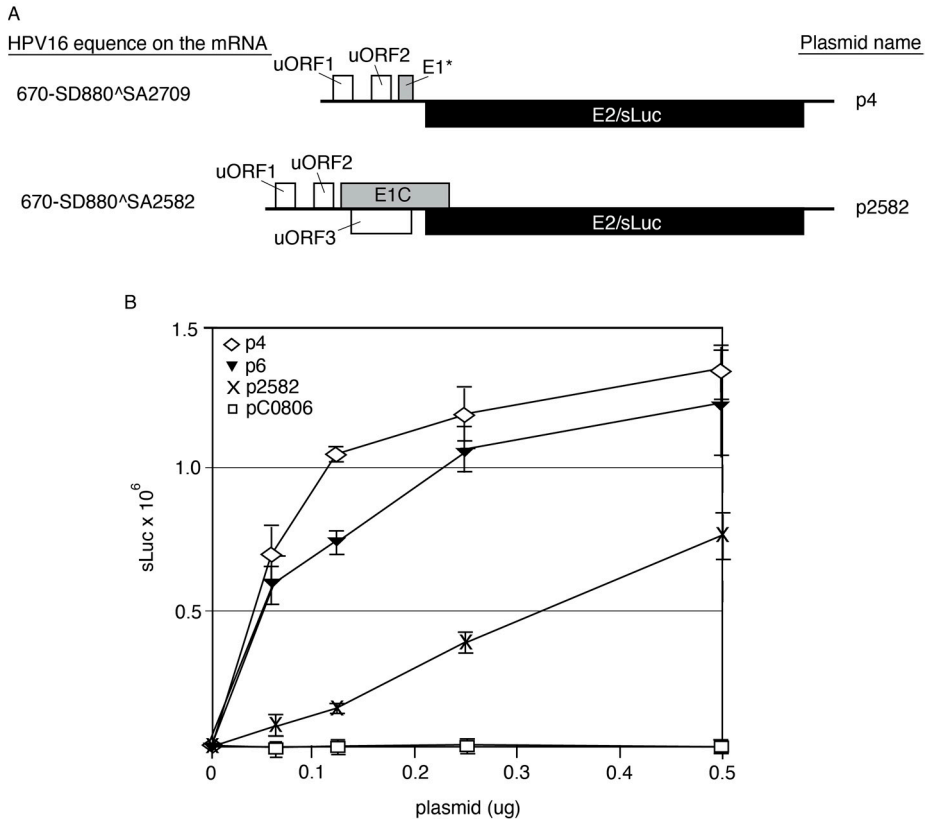
Figure 2.7 (A). Schematic representation of the HPV16 E2 cDNA expression plasmids p4 and p2582. The HPV16 sequences present on the cDNAs in each plasmid are indicated. Names of cDNA expression plasmids are indicated to the right. E2/sLuc represents the fusion of the secreted luciferase coding sequence in frame with the E2 ATG. The open reading frames upstream of the E2 ATG are indicated. The HPV16 E1* orf on plasmid p4 has the HPV16 E1 ATG and is generated by splicing from SD880 to SA2709. The HPV16 E1C orf on plasmid p2582 has the HPV16 E1 ATG and is generated by splicing from SD880 to SA2582. Two HPV16 short upstream orfs with ATGs (uORF1 and uORF2) on the mRNA produced form the p4 and the p2582 cDNAs are indicated. HPV16 uORF3 is a short orf with ATG that spans the E1 orf and is unique to plasmid p2582. **(B)**. Secreted luciferase enzyme activity (sLuc) in the cell culture medium at 24 h after transfection in triplicates into HeLa cells of serially diluted cDNA plasmids (0.5-, 0.25ug, 0.125ug and 0.0625ug of each indicated plasmid were transfected in triplicates).

Finally, we concluded that the most translated HPV16 E2 mRNA initiates at late promoter p670 and is spliced from 5'-ss SD880 to 3'-ss SA2709

Paper III

hnRNP G promotes splicing to the HPV16 E2 mRNA splice site SA2709 through a splicing enhancer consisting of an AC(A/G)AGG-triplicate

The high-risk (HR) HPVs infect mucosal epithelia cells in basal layer causing persistent infection that may progress to cancer if left untreated. Various cancers are associated with HR HPVs, especially HPV16. The HPV16 life cycle is strictly linked to cell differentiation. The HPV16 viral protein E2 functions as replication and transcription factor and plays an important role in switching HPV16 early gene expression to late gene expression by downregulating HPV16 early promoter and inhibiting the pAE signal to induce late gene expression. Our previous study on E2 expression demonstrated that the most efficiently translated E2 mRNA initiates at the HPV16 late promoter and is spliced from 5'-ss SD880 to 3'-ss SA2709. Interestingly, 3'-ss SA2582 is a competitive splice site could potentially be translated into E2 protein. We identified SA2709 as a better candidate for E2 production, suggesting the existence of post-transcriptional regulation for E2 splice site selection. We speculated that HPV16 E2 splice site SA2709 was modulated by RNA elements through interactions with cellular splicing factors. Therefore, we wished to identify the mechanism of regulation of HPV16 E2 mRNA splicing.

Three AC(G/A)AGG repeats downstream of 3'-ss SA2709 are required for production of the most efficiently translated HPV16 E2 mRNA
To investigate how production of the E2 mRNA spliced from SD880 to SA2709 is regulated, we used a subgenomic HPV16 reporter plasmid called pBEL. pBEL produces unspliced E1 mRNA and the E2 mRNAs alternatively spliced to SA2582 or SA2709 that can be measured by RT-PCR. We then introduced a number of deletions and mutations downstream of SA2709 on pBEL followed by transfection on HeLa cells. The total RNA was extracted from HeLa cells and analyzed by RT-PCR with the primers as indicated. We concluded that a 19-nucleotide RNA sequence that contains three AC(G/A)AGG repeats and is located immediately downstream of 3'-ss SA2709 and upstream of E2 start codon functions as a 3'-ss SA2709 splicing enhancer (**Figure 2.8A**). Each repeat of AC(G/A)AGG was respectively marked as motif 1, 2 and 3 (M1,2,3). When we mutated M1 or M3 or all three (**Figure 2.8A**), the splicing to SA2709 was inhibited significantly. However, it was noticed that mutation of M2 alone or M2 together with M3 did not significantly inhibit the enhancer (**Figure 2.8B**). Taken together, these results demonstrated that three AC(G/A)AGG repeats act together as a splicing enhancer and are required for regulation of E2 mRNA splicing. The results indicated that M1 and M3 are more important motifs in this enhancement.
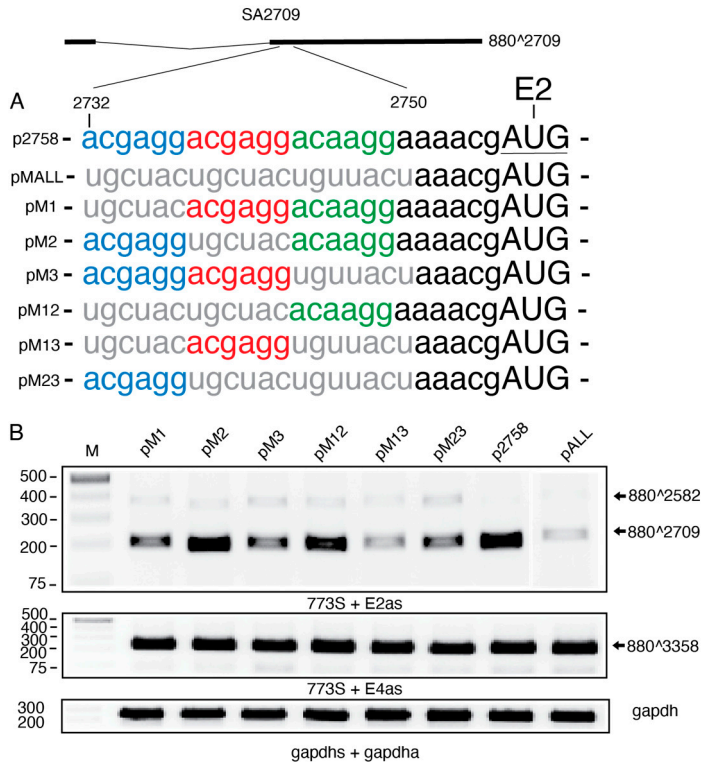
Figure 2.8 (A). The sequence of the 19-nucleotide RNA enhancer is displayed and the three AC(G/A)AGG - repeats are indicated. Various mutant enhancer sequences are shown and names of the plasmids are indicated to the left. Numbers refer to the HPV16 reference strain HPV16R. The E2 AUG is indicated. **(B)**. RT-PCR with primers 773s and E2as on RNA extracted from HeLa cells transfected with the indicated plasmids. mRNAs spliced between SD880 and SA2582 or SA2709 are indicated to the right.

We wished to further investigate the significance of the identified enhancer sequence in regulation of splice site SA2709 and exclude that other HPV16 sequences in pBEL plasmid affected splicing to SA2709. We generated plasmid pBELEN and pBELENdEN from pBEL by introducing deletions of HPV16 sequences downstream of E2 AUG. In plasmid pBELENdEN, the 19-nucleotide enhancer was also deleted either. The HPV16 splice sites in these plasmids include 5'-ss SD880 and SD1302, and 3'-ss SA2582 and SA2709, but the cryptic splice site SD1302 is inefficiently used in these plasmids. Transfection of these plasmids to HeLa cells and analysis of extracted RNA by RT-PCR indicated that major mRNA products generated by pBEL and pBELEN were spliced from SD880 to SA2709. However, plasmid pBELENdEN without 19-nucleotide enhancer significantly produced lower levels of the mRNA spliced to SA2709, but favor to produce unspliced mRNA or mRNAs redirected spliced to the competing 3'-ss SA2582. Similarly, to exclude possible effects from the E1 coding region on splice site

SA2709, we removed the majority of E1 coding sequnce from plasmid pBELEN and generated plasmid pBELENdE1. RT-PCR analysis of mRNA produced by pBELENdE1 indicated that the E1 coding region did not affect the function of the splicing enhancer downstream of SA2709.

Furthermore, we noticed that once the 19-nucleotide enhancer was destroyed or mutated, splicing to SA2709 was remarkably reduced and altered to the production of unspliced E1 mRNA or mRNA spliced to competitive splice site SA2582. The mRNA spliced to SA2582 shares part of the E1 ORF and could potentially be translated into a small E1C protein. However, very little information is available on E1C.

hnRNPG was identified as an activator of the 19-nucleotide enhancer of HPV16 E2 splice site SA2709

To identify the cellular splicing factors that interact with the identified E2 splicing enhancer, we used biotinylated RNA oligos containing the complete splicing enhancer (2758WT) or a single repeat (3WT) motif and their mutants, MALL and 3M, respectively. These RNA oligos bound to streptavidin coated beads were used for pull-down of proteins from nuclear and cytoplasmic extracts prepared from HeLa cells. Afterwards, the proteins were released from beads and analyzed on silver-stained polyacrylamide gels. In comparison to mutants oligos, multiple bands were specifically pulled down by wild type RNA oligos. These bands were selected, purified, and sent for mass spectrometry analysis. Multiple proteins were identified. We selected proteins with high coverage and multiple peptides identified and focused on RNA binding proteins for Western blotting to further confirm specificity of binding activity. Many candidates were highlighted according to specificity and efficiency, including hnRNP-A1, -A2B1, -G, -M, RBM-14, -15 and Tra2b. In contrast, hnRNP L and hnRNP U were not pulled down suggesting they were pulled down based on the protein-protein interaction with the RNA binding proteins and therefore were loosely associated with complexes or were present in small quantities.

To investigate if these candidates activated splicing to SA2709 through interaction with the identified HPV16 E2 enhancer, we cotransfected HPV16 reporter plasmid pBEL or pBELENdE1 with plasmids expressing hnRNP-A1, -A2B1, -G, -M, -L, -U, RBM14, RBM15, or Tra2b (**Figure 2.9**). As can be seen, overexpression of hnRNP G significantly promoted the splicing to SA2709 in both pBEL and pBELENdE1 and hnRNP A1 and hnRNP A2B1 strongly inhibited splicing to SA2709, while the other proteins had weaker or no effect on SA2709. Interestingly, the enhancing effect of hnRNP G resulted in decrease of unspliced E1 mRNA, while the inhibition effect of hnRNP A1 and A2B1 resulted in increase of unspliced E1 mRNA. This result suggested that the ratio of unspliced E1 mRNA to E2 mRNA spliced to either SA2582 or SA2709 was determined by binding sites selection by different splicing factors. Moreover, Tra2b inhibited E2 mRNA splicing to SA2709 but promoted E2 mRNA splicing to SA2582.
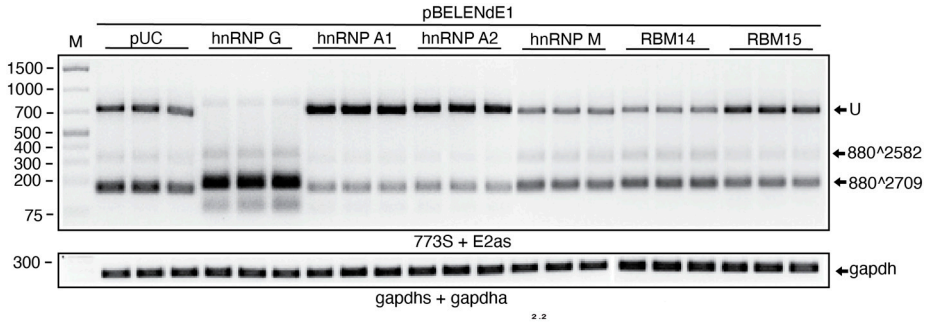
Figure 2.9 RT-PCR with primers 773s and E2as on RNA extracted from HeLa cells transfected with HPV16 subgenomic plasmid pBELENdE1 in the absence or present of the indicated protein expressing plasmids. Unspliced (U) mRNA and mRNAs spliced between SD880 and SA2582 or SA2709 (primer pair 773S and E2as) are indicated to the right.

In contrast to overexpression of hnRNP G, knock down of hnRNP G inactivated the production of E2 mRNA spliced to SA2709, whereas the unspliced E1 mRNA production was promoted (**Figure 2.10A, B, and C**). This result together with previous evidence further supported our conclusion, we established that hnRNP G enhances splicing to HPV16 E2-specific splice site SA2709.
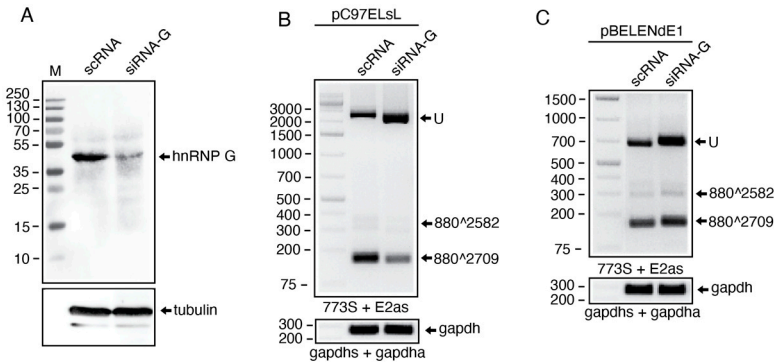
Figure 2.10 Knock-down of hnRNP G reduces splicing to HPV16 SA2709. **(A).** Knock-down of hnRNP G in HeLa cells transfected with siRNAs to hnRNP G monitored by Western blotting of hnRNP G. **(B).** HeLa cells were transfected with HPV16 subgenomic expression plasmid pC97ELsL and scrambled siRNA (scRNA) or siRNAs to hnRNP G, (siRNA-G), RNA was extracted and RT-PCR was performed with primers 773S and E2as. Unspliced (U) mRNA and mRNAs spliced between SD880 and SA2582 or SA2709 (primer pair 773S and E2as) are indicated to the right. **(C).** HeLa cells were transfected with HPV16 subgenomic expression plasmid pBELENdE1 and scrambled siRNA (scRNA) or siRNAs to hnRNP G, (siRNAG), RNA was extracted and RT-PCR was performed with primers 773S and E2as. Unspliced (U) mRNA and mRNAs spliced between SD880 and SA2582 or SA2709 are indicated to the right.

Paper IV

A bio-assay for HPV16 E1/E2 mRNA splicing

HPV16 is the most common high-risk HPV type and is therefore of particular medical interest. During the early stage HPV16 life cycle, E1 protein with the help of E2 protein bind to the origin of replication initiation (*Ori*) and bring down host DNA polymerase to *Ori* to initiate HPV16 DNA replication. E2 protein has multiple functions in regulation of HPV16 gene expression, for example shutting down early gene expression and activation of late gene expression. When HPV16 infected cells progress to cancer, E2 protein expression is commonly repressed while E6 and E7 oncoproteins are upregulated. Moreover, E2 protein also interacts with cellular RBPs, such as hnRNPs and SRs, presumably to affect alternative splicing of viral and cellular mRNAs. Our previous study demonstrated that the HPV16 mRNA initiating at late promoter p670 and spliced from SD880 to SA2709 is most efficiently translated into E2 protein. The selection of major splice site SA2709 is strictly regulated by a downstream splicing enhancer through the interaction with hnRNP G. However, unspliced E1 mRNA or the mRNA spliced to competitive splice site SA2582 is upregulated if the major splice site SA2709 break down or is mutated. Unspliced E1 mRNA is used for E1 production, while the mRNA spliced to SA2582 could potentially be translated into E1C protein. The role of E1C in HPV16 life-cycle is not well studied. These results suggest that the balance of E1, E1C and E2 proteins is highly regulated by the enhancer located downstream of SA2709 that interacts with multiple splicing factors. We therefore wished to establish a bio-assay by introducing various reporter genes to an HPV16 subgenomic plasmid that could monitor the balance between unspliced E1 mRNAs and spliced E2 mRNAs. This bio-assay could potentially be applied for anti-HPV16 drug development.

Generation of reporter genes diagnostic for E1 and E2

We reported that the HPV16 subgenomic plasmid pBELEN could produce three mRNAs including unspliced E1 mRNA and the mRNAs spliced to either SA2709 or SA2582. The latter mRNA has the potential to produce an E1-derived protein called E1C. In order to construct plasmids that express reporter genes for identification for E1, E1C and E2 mRNAs, we firstly generated pE1BAM and pE2BAM in which the BamHI restriction site was inserted in frame with either E1/E1C or E2 using a naturally occurring CsiI site located in the overlap between E1/E1C and E2 ORFs in the pBELEN plasmid. Four reporter genes were used including secreted luciferase (sLuc), green fluorescent protein (GFP), neomycin resistance (neo) and thymidine kinase (tk). They were fused in frame with E1 or E2 ORFs through the polylinker downstream of BamHI (**Figure 2.11 A and B**). All four reporter genes fused to E1 ORF were expected to produce reporter proteins from unspliced mRNA (E1 fusion proteins) or from mRNA spliced to SA2582 (E1C

fusion protein), but not from mRNAs that are spliced to SA2709. Similarly, the reporter genes under the E2 ORF were expected to express E2 fusion proteins (reporter proteins utilizing the E2 AUG) from mRNA spliced to SA2709, but not from unspliced E1 mRNA or E1C mRNA spliced to SA2582.
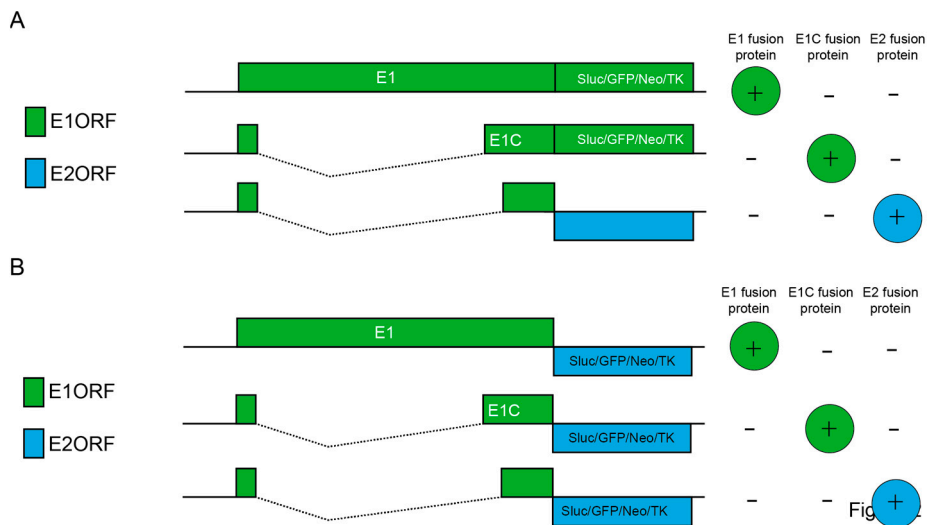
Figure 2.11 (A). Predicted mRNAs are shown below the plasmid representations and the predicted production of the various E1 fusion proteins or E2 are indicated to the right. **(B)**. Structures of the various pE2BAM-derived plasmids carrying reporter gens fused in frame with the E2 ATG are shown. Predicted mRNAs are shown below the plasmid representations and the predicted production of the two E1 proteins or E2-fusion proteins are indicated to the right.

Analysis of reporter plasmids in transfected cells.

These reporter plasmids were then transfected to HeLa cells. The sLuc activity in the cell culture medium was detected 24hrs posttransfection. sLuc levels produced by pE2sLuc were significantly higher than produced by pE1sLuc. As expected, the GFP levels produced by pE2GFP was significantly higher than pE1GFP produced (**Figure 2.12A and B**). The transfection of pE1TK or pE2TK was followed by ganciclovir (GCV) treatment. MTT-assay was applied for viability calculation. As can be seen (**Figure 2.12C**), pE2tk transfected HeLa cells were more sensitive to ganciclovir treatment than the cells transfected by pE1tk or negative control pUC. The transfections of pE1Neo and pE2Neo were followed by G418 treatment. Viability of transfected cells were assessed by MTT-assay. Since negative control, pUC transfected cells were unaffected by G418, G418 treatment of the transfected cells needs to be optimized.

Figure 2.12 (A). Secreted luciferase enzyme activity (Sluc) in the cell culture medium at 24h posttransfection of HeLa cells transfected in triplicates with pE1sLuc or pE2sLuc. Mean values and standard deviations are shown. **(B)**. GFP expression was monitored by fluorescence microscope at 24h and 48h after transfection of HeLa cells with pE1GFP or pE2GFP. **(C)**. MTT based cell viability assay on HeLa cells transfected with pUC, pE1tk or pE2tka for 48hrs and treated with indicated concentrations of ganciclovir for another 48hrs.

Our previous study demonstrated that E2 mRNA spliced to SA2709 was promoted by hnRNP G and inhibited by hnRNP A1. We therefore cotransfected E2-sLuc reporter to HeLa with the presence of hnRNP G and hnRNP A1. sLuc levels in medium were measured and indicated overexpression of hnRNP G significantly promoted the production of E2-sLuc, while overexpression of hnRNP A1 repressed the E2-sluc production (**Figure 2.13**). This result agrees with our previous study and demonstrated that the E2BAM reporter plasmid could be used as a reporter assay to identify factors or substances that interfere with E2 mRNA splicing. Furthermore, SR protein SRSF9 and Tra2b have been shown to affect HPV16 mRNA splicing. Therefore, we cotransfected SRSF9 and Tra2b expression plasmids with sLuc and GFP reporters to HeLa cells. The transcriptional and translational expression level

were respectively monitored. As shown, SRSF9 and Tra2b inhibited alternative splicing to SA2709 results in splicing to SA2582 in posttranscriptional level but did not reflect in translational level.
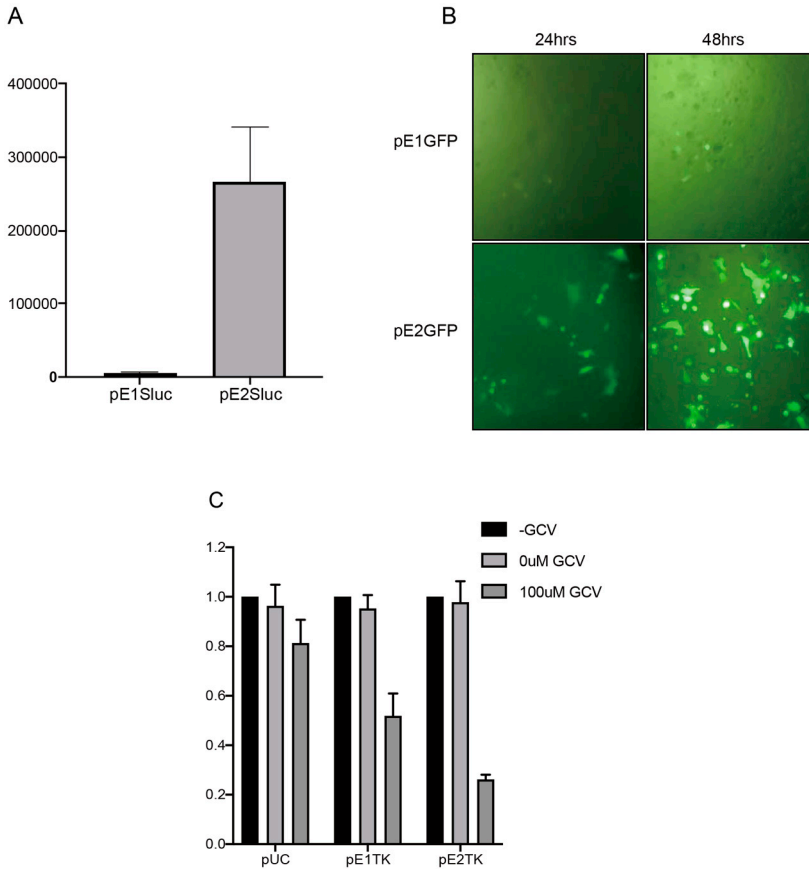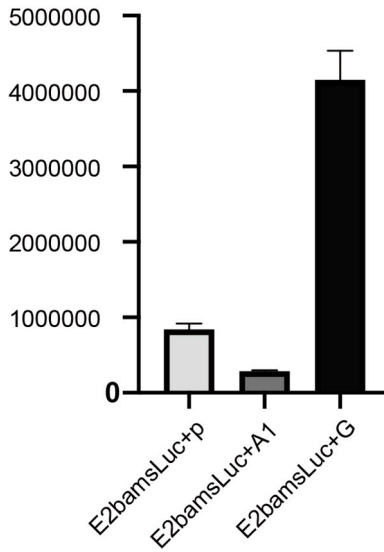

Figure 2.13 Secreted luciferase enzyme activity (Sluc) in the cell culture medium at 24h posttransfection of HeLa cells cotransfected in triplicates with pE2sLuc with or without hnRNP G or hnRNP A1 or pUC. Mean values and standard deviations are shown.

3 Conclusions and future perspectives

The results presented in this thesis highlighted the importance of regulation of human papillomavirus 16 early gene expression. It especially reflected the importance of alternative splicing regulation, including splice site selection through various RNA *cis*-element interacting with numbers of cellular splicing factors.

The HPV16 oncoprotein E7 is translated from mRNA spliced from SD226 to SA409. The HPV16 3'-ss SA409 is a high-risk HPVs specific landscape, highly associated with cancer progression. We determined that both cellular splicing factors hnRNP A1 and hnRNP A2B1 inhibited the production of major E6/E7 mRNA products spliced from 5'-ss SD2226 to 3'-ss SA409. However, overexpression of hnRNP A1 resulted in the upregulation of unspliced E6 mRNA. The mechanism of hnRNP A1 upregulation of the unspliced E6 mRNA through inhibition of SA409 remains to be studied. Interestingly, overexpression of hnRNP A2B1 resulted in alternative splicing to SA742. Further investigation is needed to determine how these two proteins affect the splicing of E6/E7 mRNA in such an opposed way, but we speculated the differential effect is mainly caused by differences in the less conserved C-terminus of both proteins. In agreement with the previous study, we mapped the hnRNP A1 binding site in HPV16 to the 5' - UAUGU**UAG**AUU -3'-sequence. The motif "UAG" has been determined as hnRNP A1 binding site by multiple previously published studies [232, 233]. Moreover, this 11-nucleotide sequence locates in E7 coding region. Previous study indicated that E7 is highly conserved. We therefore performed sequence alignments. We reported that sequence UAGAU is strictly conserved in all identified high-risk HPVs in the HPV *alpha*-9 group. This sequence is less conserved in the HPV *alpha*-7 group. The HPV18 is a prototype in *alpha*-7 group. Previous study has shown that hnRNP A1 inhibits the splice site in the HPV18 E6 region [216], and the reported binding site is similar to what we identified and mapped in HPV16. Therefore, we speculate that hnRNP A1 is important in controlling of HPV16 and HPV18 E6 and E7 expression. HPV16 E6 protein induces degradation of apoptosis mediator p53 and both E6 and E7 play key roles in HPV16-driven cancers. Moreover, we have shown that the levels of hnRNP A1 increased in cervical cancer. Taken together, it is reasonable to speculate that hnRNP A1 plays a vital role in HPV16 life cycle and HPV16-driven carcinogenesis. In addition to splicing regulation, we speculate that functional hnRNP A1 requires posttranslational modification, such as phosphorylation, to be active, which is an area of future research.

Various HPV16 mRNAs are produced for E2 translation. These E2 mRNAs initiate at either early promoter p97 or late promoter p670 and many splice sites at both 5'- and 3'-ends are involved. In line with the previous study, we demonstrated that the E2 mRNAs initiate at early promoter p97 and are spliced from SD226 to SA2709 are more efficiently produced than the mRNAs spliced from SD226^SA409-SD880. However, the expression efficiency of E2 mRNAs initiate at late promoter p670 in comparison to prior early promoter initiated E2 mRNAs have not been investigated. We demonstrated that the most efficiently expressed E2 mRNA initiated at p670 and was spliced from SD880 to SA2709. Interestingly, the E2 mRNAs have two competitive 3' alternative splice site SA2582 and SA2709, the E2 mRNAs spliced to either splice site has potential to produce E2 protein. The sLuc levels demonstrated that the E2 mRNAs spliced to SA2709 were significantly higher than the mRNAs spliced to SA2582. The E2 mRNAs spliced to SA2582 is in E1 ORF and capable to produce a protein named E1C. However, the significance of 3'-ss SA2582 in HPV16 is unknown, and the benefit of keeping a second E2 splice site is unclear as is the generation of a suboptimal E2 mRNA.

We investigated the regulation of 3'-splice site SA2709. Our results demonstrated that the 19-nucleotide in length RNA element downstream of splice site SA2709 functions as a splicing enhancer. This identified splicing enhancer consists of three AC(G/A)AGG repeats. Repeat 1 and repeat 3 are more important, while repeat 2 is less important. However, if all three repeats are mutated or deleted, splicing to SA2709 is inhibited, further indicating that 19-nucleotide sequence is needed for E2 production. This enhancer sequence was aligned with other HPVs or all HPV16 sub types and lineages indicated that this short enhancer region is strictly conserved in all HPV16 subtypes and lineages; well-conserved in *alpha*-9 HPVs; but less conserved in *alpha*-7 HPVs. In addition to three repeats of AC(G/A)AGG motifs, alternative motifs are noticed and needed to be determined in the future. We demonstrated that cellular splicing factor hnRNP G enhances the expression of E2 mRNA spliced to SA2709. However, if the enhancement of E2 mRNA is caused by hnRNP G interacting directly with the enhancer RNA downstream of SA2709 or indirectly interacting with other RNA binding factors is still unknown. We speculate that hnRNP G promotes the E2 mRNA spliced to SA2709 through direct interaction with identified enhancer downstream of SA2709 since the known binding sites of hnRNP G are purine-rich sequences. To determine the mechanism how hnRNP G upregulates E2 mRNA spliced to SA2709, we introduced deletions to hnRNP G plasmid. According to our present results, we concluded that C-terminal RBD of hnRNP G was not required for enhancement of splicing to HPV16 SA2709. Our results also suggested that hnNRP G interacts with the E2 splicing enhancer through its N-terminal RRM. The previous study determined that hnRNP G is also a part of the DNA damage response (DDR) machinery [213]. It is reasonable to speculate that hnRNP G is used by HPV16 for other duties as well, not just as E2 mRNA splicing activator. Our previous study identified that hnRNP G controls HPV16 late L1 mRNA splicing [234], in addition to

E2 mRNA splicing. It appears that hnRNP G plays significant role in the regulation of HPV16 RNA processing. It will be our future interest to investigate how hnRNP G responds to HPV16 infection and to cell differentiation and carcinogenesis. Besides the splicing activator hnRNP G, we found splicing factor Tra2b inhibited the splicing to SA2709 and enhanced splicing to HPV16 splice site SA2582. The E2 mRNA spliced from SD880 to SA2582 could potentially be translated into E1C protein. The function and the role of E1C protein in HPV16 gene expression is still unclear. Therefore, the study on E1C may be of interest in the future. Furthermore, we observed that the mRNA ratio of E1-, E1C-, and E2-mRNAs is controlled by the 19-nucleotide splicing enhancer downstream of HPV16 E2 splice site SA2709. Both E1 and E2 mRNAs originate from the same unspliced pre-mRNA. The balance in expression between E1 and E2 is necessary for the HPV16 life cycle as well as carcinogenesis. During the HPV16 life cycle, the E2 protein assists E1 in conducting DNA replication, followed by shutting down HPV16 early promoter and repressing the expression of E6 and E7, which is required for cell differentiation and activation of HPV16 late gene expression. The ability to regulate splice site SA2709 and altering the ratio between E1- and E2- mRNAs may be crucial to the HPV16 life cycle. We speculate that altering the E1/E2 ratio is required for switching from a state of high HPV16 DNA replication to a state of virus production. Besides suppressing early promoter, the E2 protein inhibits the HPV16 early polyadenylation signal to allow for reading-through into the late region of the HPV16 genome followed by expression of L1 and L2 mRNAs and proteins and formation of virus particles. Therefore, the future studies on E2 enhancer downstream of HPV16 splice site SA2709 would be meaningful in order to further our understanding of the HPV16 replication cycle.

Finally, we wish to investigate ratio altering between E1 and E2 mRNAs and proteins by establishing a bio-assay to monitor both mRNA and protein levels of E1, E2 and the small poorly studied viral protein E1C. We successfully generated a number of reporter plasmids encoding sLuc, GFP, neomycin resistant gene, or thymidine kinase. These reporters were respectively fused to either the E1 ORF or the E2 ORF. As we demonstrated previously, the dominant E2 mRNA spliced from SD880 to SA2709 is the best E2 protein producer while the unspliced mRNA is translated into E1 protein. However, the E1 expression is normally too low to be detected. The E2 mRNA spliced to SA2582 is speculated to be translated into a small protein E1C. Our results indicated E2 fusion proteins were generally higher than E1 fusion proteins that is in line with our previous study. However, these reporter plasmids need to be further optimized, and continue search of alternative and better suited reporter genes are necessary. Finally, the results in my thesis describe novel findings of *cis*- and *trans*-acting factors that control HPV16 gene expression at the level of RNA processing. These results further our understanding of HPV16 gene expression. Furthermore, we developed a bio-assay for HPV16 E2 mRNA splicing that could be used to identify small molecules that potentially interfere with HPV16-infections and HPV16-driven cancers.

4 Methods

Plasmids

The following plasmids were described previously: pC97ELsL (107), pBEL (223), pBELsL (228), pCRE (229), pHPV16AN (107), pC0806 (230), p6 (231), pTra2b (cDNA purchased from Origene (SC322178), pSRSF1 (173), and pSRp30c (174). Plasmids phnRNP A1 and phnRNP A2 express the hnRNP A1 or A2 cDNAs from the CMV promoter. The other plasmids were detailed described respectively in Paper *I, II, III* and *IV*.

Cells

HeLa cells, HEK293T, C33A2, and HPV16-positive SiHa cells were cultured in Dulbecco's modified Eagle medium (GE Healthcare Life Science HyClone Laboratories) with 10% bovine calf serum (GE Healthcare Life Science HyClone Laboratories) and 1% penicillin-streptomycin (Gibco Thermo Fisher Science). The reporter cell line C33A2 is derived from the HPV16 negative cervical cancer cell line C33A. C33A2 has the sub-genomic HPV16 plasmid pBelsLuc stably integrated into chromosome. The pBelsLuc plasmid contains HPV16 genome except the oncogenes E6 and E7, under control of the Cytomegalovirus (CMV) promoter. The HN26 cells are derived from a tumor of a 48-year old nonsmoking man with non-keratinizing, HPV16-positive tonsil oral squamous cell carcinoma, stage T2N0M0. The HN26 cells contain episomal HPV16 DNA and have an intact p53 gene. HeLa cells and HN26 cells were cultured in RPMI-1640 supplemented with 10 % bovine calf serum and 1% penicillinstreptomycin.

Transfections

Transfections of HeLa cells were carried out using TurboFect according to the manufacturer's instructions (Thermo Fisher Science). TurboFect was mixed with plasmid DNA and incubated at room temperature for 15 min prior to dropwise addition to 60-mm plates with subconfluent HeLa cells. Cells were harvested at 20 h posttransfection. Each plasmid was transfected in triplicate in a minimum of two independent experiments. For analysis of episomal HPV16, plasmid pHPV16AN was cotransfected with plasmid pCAGGS-nlscre (generously provided by Andras Nagy at University of Toronto), which expresses the cre recombinase that releases the HPV16 genome from the plasmid at two flanking lox sites.

siRNA transfections

siRNA knockdowns were carried out using DharmaFECT-DUO transfection reagents according to the manufacturer's instructions. Briefly, the siRNA was diluted to 40 nM final concentration in 250 ul serum-free medium, and the mixture was added to 250 ul of serum-free medium with 5 ul transfection reagent. The mixture was incubated at room temperature for 20 min prior to addition to a 60-mm plate with subconfluent HeLa or SiHa or C33A2 cells. siRNA to hnRNPA1 was ON-TARGET plus SMART pool human hnRNPA1 (L-008221-00-0020; Dharmacon) and to hnRNPA2B1 ON-TARGET plus SMART pool human hnRNPA2B1 (L-011690-01-0020; Dharmacon). The scrambled control (scr) was siGENOME control pool nontargeting number 2 (D-001206-14-20; Dharmacon). siRNA to hnRNP G was ON-TARGET plus SMART pool human hnRNP G.

RNA extraction and RT-PCR

Total RNA was extracted using TRI Reagent and a Direct-zol RNA MiniPrep kit (ZYMO Research) according to the manufacturer's protocol. One microgram of total RNA was reverse transcribed in a 20 ul reaction mixture at 37°C by using M-MLV reverse transcriptase (Invitrogen) and random primers (Invitrogen) according to the protocol of the manufacturer. One microliter of cDNA was subjected to PCR amplification. HPV16 mRNAs spliced from HPV16 5'-splice site SD226 to 3'-splice site SA409 were amplified with RT-PCR primers 97S and 438A and from HPV16 5=-splice site SD226 to 3'-splice sites SA409, SA526, and SA742 with RT-PCR primers 97S and 880A. HPV16 mRNAs spliced from 5'-splice site SD880 to 3'-splice site SA2709 or SA3358 were amplified with RT-PCR primers 773S and E2A or E4A. HPV16 late L1 mRNAs were amplified with RT-PCR primers 773S and L1A. Glyceraldehyde-3phosphate dehydrogenase (GAPDH) cDNA was amplified with primers GAPDH-F and GAPDH-R. To monitor recombination at the loxP sites in pHPV16AN, PCR was performed with primers 16S and 16A on DNA extracted from the transfected cells (this PCR yields a 366-nucleotide PCR fragment that is diagnostic for recombination at the LoxP sites). Sequences of the RT-PCR primers are available on request. Examples of control PCR experiments performed on RNA samples in the absence of reverse transcriptase are shown in the various figures.

Secreted luciferase assay

The Metridia longa secreted luciferase activity in the culture medium of the HeLa cells was monitored with the help of the Ready To Glow secreted luciferase reporter assay according to the instructions of the manufacturer (Clontech Laboratories). Briefly, 50 μL of cell culture medium were mixed with 5 μL of secreted luciferase substrate in reaction buffer and luminescence was monitored in a Tristar LB941 luminometer (Berthold Technologies).

Protein extraction and Western immunoblotting

Proteins for Western blotting were extracted from cells using the radioimmunoprecipitation assay (RIPA) buffer (50 mM Tris, pH 7.4, 500 mM NaCl, 1% Na-DOC, 0,1% SDS, 1% Triton X-100) with 30 min of incubation on ice and occasional vortexing. Western blotting was performed as described previously. The following antibodies were used: anti-hnRNP A1 (04-1469; Millipore), anti-hnRNP A2B1 (ab227465; Abcam), anti-HPV16 E7 (GTX133411; GeneTex), antibeta tubulin (T9026; Sigma-Aldrich), anti-actin (SC-1616; Santa Cruz), anti-HA tag (SC7392; Santa Cruz), and anti-GST (A5800; Invitrogen) antibody. Secondary antibodies conjugated with horseradish peroxidase were used, and proteins were detected using the Clarity Western ECL substrate (Bio-Rad) or the Super Signal West Femto chemiluminescence substrate (Pierce).

ssRNA-mediated protein pulldown assay

Nuclear extracts were prepared according to the procedure described previously. Briefly, the cells were lysed using lysis buffer A (10 mM HEPES, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM dithiothreitol, 0.05% NP-40, pH 7.9, and protease inhibitors) to release cytoplasmic proteins. The pelleted nuclei were resuspended in buffer B (5 mM HEPES, 1.5 mM MgCl2, 0.2 mM EDTA, 0.5 mM dithiothreitol, 26% glycerol, pH 7.9, and protease inhibitor) to release nuclear proteins. The nuclear extracts were mixed with streptavidin-coated magnetic beads (Dynabeads M-280 Streptavidin; Invitrogen) bound to biotin-labeled single-stranded RNA (ssRNA) oligonucleotides (Sigma-Aldrich) in binding buffer (10 mM Tris, pH 7.4, 150 mM NaCl, 2.5 mM $MgCl_2$, 0.5% Triton X-100). Sequences of all biotinylated ssRNA oligonucleotides are available on request. The mixtures were incubated at room temperature with rotation for 1 h, followed by washing five times with 1 ml wash buffer. Proteins were eluted by boiling of the beads in SDS-PAGE loading buffer and subjected to SDS-PAGE followed by Western blotting with the indicated antibodies. GST and GST-hnRNP A1 were purified from Escherichia coli transformed with pGEX-derived plasmids using GS beads.

RNA immunoprecipitation

HeLa cells were transfected with pC97ELsL in the presence of hnRNP A1 or hnRNP A2. After 24 h, cells were lysed in RIP buffer (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 10% glycerol, 1% NP-40, 2 mM EDTA, pH 8.0, 400 U/ml Ribolock RNase inhibitor [ThermoFisher], protease inhibitor cocktail [Sigma]) and rotated for 1 h at 4°C. Samples were then centrifuged for 15 min at 14,000 g at 4°C, and the supernatants were transferred to new tubes and incubated with 4 ug/ml anti-hnRNPA1 antibody (ab5832; Abcam), anti-hnRNPA2 antibody (ab227465; Abcam), or IgG overnight at 4°C. Each RNA-IP was incubated with 50 _ l Dynabeads protein G (Life Technologies) for 1.5 h at 4°C, followed by washing four times in RIP buffer supplemented with 300 mM NaCl. The immunoprecipitated RNA then was extracted as described previously and subjected to RT-PCR by HPV16-specific primers 773s and E42AS, which detect HPV16 mRNA spliced

between HPV16 splice sites SD880 and SA3358. Input represents RT-PCR with the same primers on RNA extracted from 5% of the input in the immunoprecipitation incubation.

Quantitations
The software used to determine band intensity in Western blots and RT-PCR gels was Image Lab 6.0.1, and quantitations were performed with the software Prism GraphPad 8.4.0.

Sequence alignment
The software for alignments was Jalview 2.0.5, and the alignment method was Muscle.

Acknowledgements

Since I started to write my thesis, I have been looking forward to this part. When I get to this moment, thousands of words with those familiar faces are flashing in my mind, but I don't know where to start.

First and foremost, I would like to express my deep and sincere gratitude to my research supervisor, Professor **Stefan Schwartz**, for offering me the opportunity to do research and provide invaluable guidance throughout my research. I am also extremely grateful for your spiritual and financial support during my Ph.D. study. Your dynamism, vision, sincerity, and motivation have deeply inspired me. I admire your attitude and passion for scientific research that will motivate me all the time. I benefit a lot from each discussion and meeting we ever had. Every time I had problems in the study, I would never forget every time you always encouraged me and helped me find a way out. You let me know how a good supervisor should be.

I am very much grateful to **Chinese Scholarship Council (CSC)**. Thanks to offer me four years scholarship to do my research at Lund.

I am sincerely grateful to our former member Professor **Chengjun, Wu**. Thank you so much for teaching me so many things. You also taught me how to work independently. You are the most helpful and considerate person I ever met. Whenever I need your help, you always stand behind me and support me. **Naoko Kajitani**, thank you for proposing those smart ideas in our lab meeting. Whenever we ask you questions, your answer is always knowledgeable and precise. Your passion for research and patience with us will inspire me all the time.

To all former members of the team, **Lijing**, **Efthymios**, **Mandy**, **Julia**, **Kersti**, with a special thanks to **Haoran, Yu**, you are a caring confident friend, thank you for all the fun times we had. I wish you to be a future big star on "Little Red Book".

To current lab members, with a special thanks to **Johanna Jonsson**, we discussed not only research but everything, bitterness, and happiness in life. **Xiaoxu, Cui,** whenever we were struggling in cloning, your jokes always make me feel better. Thanks for your magical "golden finger" to point out the right colony I should pick. **Shirin Shoja Chaghervand,** you are fighting with a challenging project that consumes a lot of energy and time. I wish you all the luck. **Chengyu, Hao,** thanks for sharing the office with me. Your positive attitude to life always encourages me.

I am so lucky to have you, fantastic guys, throughout my Ph.D. More than colleagues, we are friends and a family.

To **Patrick Medstrand**, **Marianne Jansson**, **Joakim Esbjörnsson**, thank you for inviting lots of researchers in different areas to give us amazing presentations that expanded our visions. To **Ram**, **Caterina**, **Ines, Daniel, Manoj, Aftab,** thank you for sharing reagents and chemicals. Your selfless ensure our corridor a friendly working environment.

To my dear wife **Aodeng, Gaowa**, I am very much grateful to your unconditional support, you have never complained me because of my job. You work hard for our family and our daughter. Marrying you is my greatest luck. To my dear daughter **Shuyue, Zheng,** you are my best gift, just like your name, "moon", a "moon light" can always melt my heart and motivate me moving forward.

To all my dear friends, **Fangyuan**, **Di, Wang**, **Xuan**, **Ruiyu**, **Lu Chen**, **Hong Chang**, **Jingbei**, **Kena**, **Jixing**, **Zhengdong**, **Ouyang**, **Li Zheng**, thanks for every wonderful moment you spent with me. With special thanks to **Minjun, Yang**, I learned a lot of basis of bioinfomatic knowledge from you. **Zsófia Szojka**, thanks for your greatest sense of humor and those relaxing conversation. I am so glad you come back to Sweden to start your new era. To apologise to all the ones I did not nominate, it is difficult to include all those friends appeared in my life and brought me with happiness.

最后，我要真挚的感谢我亲爱的父母。谢谢你们为我成长道路上所付出的辛劳和汗水。你们的包容，理解以及支持只为成就我的梦想，让我飞的更高更远。这里短短几行字无法表达我对你们的爱与感激，不远的将来儿子即将学成归国，唯愿以后的每个节日都能与你们相伴，希望你们健康，长寿。我也要感谢我们大家庭里的每个成员，当我不在父母身边的时候，是你们对我父母的照顾与关怀才让我没有后顾之忧。我也要诚挚的感谢我的岳父和岳母，格格才出生的时候，你们不远万里来到异国他乡的瑞典帮我们度过了一段艰难的时光，你们的无私帮助让我们的小家庭更加稳固。

References

1. zur Hausen, H., *Papillomaviruses in the causation of human cancers - a brief historical account.* Virology, 2009. **384**(2): p. 260-5.

2. Zheng, Z.-M. and C.C. Baker, *Papillomavirus genome structure, expression, and post-transcriptional regulation.* Frontiers in bioscience: a journal and virtual library, 2006. **11**: p. 2286.

3. Zur Hausen, H., *Papillomaviruses and cancer: from basic studies to clinical application.* Nature reviews cancer, 2002. **2**(5): p. 342-350.

4. Parkin, D.M. and F. Bray, *The burden of HPV-related cancers.* Vaccine, 2006. **24**: p. S11-S25.

5. Richardson, H., et al., *The natural history of type-specific human papillomavirus infections in female university students.* Cancer Epidemiology and Prevention Biomarkers, 2003. **12**(6): p. 485-490.

6. Stanley, M., *Immunobiology of HPV and HPV vaccines.* Gynecologic oncology, 2008. **109**(2): p. S15-S21.

7. Arbyn, M., et al., *EUROGIN 2011 roadmap on prevention and treatment of HPV-related disease.* International journal of cancer, 2012. **131**(9): p. 1969-1982.

8. Gellin, B., et al., *Quadrivalent human papillomavirus vaccine.* Clinical Infectious Diseases, 2007. **45**(5): p. 609-617.

9. De Sanjose, S., et al., *Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study.* The lancet oncology, 2010. **11**(11): p. 1048-1056.

10. Li, N., et al., *Human papillomavirus type distribution in 30,848 invasive cervical cancers worldwide: Variation by geographical region, histological type and year of publication.* International journal of cancer, 2011. **128**(4): p. 927-935.

11. Gillison, M.L., et al., *Evidence for a causal association between human papillomavirus and a subset of head and neck cancers.* Journal of the National Cancer Institute, 2000. **92**(9): p. 709-720.

12. Nathanson, N., *Viral pathogenesis and immunity*. 2007: Elsevier.

13. Joura, E.A., et al., *A 9-valent HPV vaccine against infection and intraepithelial neoplasia in women.* New England Journal of Medicine, 2015. **372**(8): p. 711-723.

14. Begum, S., et al., *Tissue distribution of human papillomavirus 16 DNA integration in patients with tonsillar carcinoma.* Clinical Cancer Research, 2005. **11**(16): p. 5694-5699.

15.   Van Doorslaer, K., et al., *The Papillomavirus Episteme: a major update to the papillomavirus sequence database.* Nucleic acids research, 2017. **45**(D1): p. D499-D506.

16.   Antonsson, A., et al., *The ubiquity and impressive genomic diversity of human skin papillomaviruses suggest a commensalic nature of these viruses.* Journal of virology, 2000. **74**(24): p. 11636-11641.

17.   Orth, G., *Epidermodysplasia verruciformis: a model for understanding the oncogenicity of.* Papillomaviruses, 2008. **120**: p. 157.

18.   Bodily, J. and L.A. Laimins, *Persistence of human papillomavirus infection: keys to malignant progression.* Trends in microbiology, 2011. **19**(1): p. 33-39.

19.   De Villiers, E.-M., et al., *Classification of papillomaviruses.* Virology, 2004. **324**(1): p. 17-27.

20.   Bzhalava, D., et al., *A systematic review of the prevalence of mucosal and cutaneous human papillomavirus types.* Virology, 2013. **445**(1-2): p. 224-231.

21.   de Villiers, E.-M., *Cross-roads in the classification of papillomaviruses.* Virology, 2013. **445**(1-2): p. 2-10.

22.   Lorincz, A.T., et al., *Human papillomavirus infection of the cervix: relative risk associations of 15 common anogenital types.* Obstetrics and gynecology, 1992. **79**(3): p. 328-337.

23.   Howley, P.M. and H.J. Pfister, *Beta genus papillomaviruses and skin cancer.* Virology, 2015. **479**: p. 290-296.

24.   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.

25.   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.* Journal of Investigative Dermatology, 2000. **114**(3): p. 413-420.

26.   Egawa, K., *Do human papillomaviruses target epidermal stem cells?* Dermatology, 2003. **207**(3): p. 251-254.

27.   Schäfer, G., M.J. Blumenthal, and A.A. Katz, *Interaction of human tumor viruses with host cell surface receptors and cell entry.* Viruses, 2015. **7**(5): p. 2592-2617.

28.   Day, P.M. and M. Schelhaas, *Concepts of papillomavirus entry into host cells.* Current opinion in virology, 2014. **4**: p. 24-31.

29.   Pyeon, D., et al., *Establishment of human papillomavirus infection requires cell cycle progression.* PLoS Pathog, 2009. **5**(2): p. e1000318.

30.   McBride, A.A., *Mechanisms and strategies of papillomavirus replication.* Biological chemistry, 2017. **398**(8): p. 919-927.

31.   Münger, K. and P.M. Howley, *Human papillomavirus immortalization and transformation functions.* Virus research, 2002. **89**(2): p. 213-228.

32.   Doorbar, J., et al., *The biology and life-cycle of human papillomaviruses.* Vaccine, 2012. **30**: p. F55-F70.

33. 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.

34. Peh, W.L., et al., *Life cycle heterogeneity in animal models of human papillomavirus-associated disease.* Journal of virology, 2002. **76**(20): p. 10401-10416.

35. Duensing, S. and K. Münger, *Mechanisms of genomic instability in human cancer: insights from studies with human papillomavirus oncoproteins.* International Journal of Cancer, 2004. **109**(2): p. 157-162.

36. Matsukura, T., S. Koi, and M. Sugase, *Both episomal and integrated forms of human papillomavirus type 16 are involved in invasive cervical cancers.* Virology, 1989. **172**(1): p. 63-72.

37. Vinokurova, S., et al., *Type-dependent integration frequency of human papillomavirus genomes in cervical lesions.* Cancer research, 2008. **68**(1): p. 307-313.

38. Andersen, F.F., et al., *The RNA splicing factor ASF/SF2 inhibits human topoisomerase I mediated DNA relaxation.* Journal of molecular biology, 2002. **322**(4): p. 677-686.

39. Doorbar, J., *The papillomavirus life cycle.* Journal of clinical virology, 2005. **32**: p. 7-15.

40. Johansson, C. and S. Schwartz, *Regulation of human papillomavirus gene expression by splicing and polyadenylation.* Nature reviews Microbiology, 2013. **11**(4): p. 239-251.

41. Kämmer, C., et al., *Sequence analysis of the long control region of human papillomavirus type 16 variants and functional consequences for P97 promoter activity.* Journal of General Virology, 2000. **81**(8): p. 1975-1981.

42. López-Saavedra, A., et al., *Functional implication of sequence variation in the long control region and E2 gene among human papillomavirus type 18 variants.* Archives of virology, 2009. **154**(5): p. 747-754.

43. Chiang, C.-M., et al., *Viral E1 and E2 proteins support replication of homologous and heterologous papillomaviral origins.* Proceedings of the National Academy of Sciences, 1992. **89**(13): p. 5799-5803.

44. Kurg, R., et al., *Characterization of the functional activities of the bovine papillomavirus type 1 E2 protein single-chain heterodimers.* Journal of virology, 2006. **80**(22): p. 11218-11225.

45. Russell, J. and M.R. Botchan, *cis-Acting components of human papillomavirus (HPV) DNA replication: linker substitution analysis of the HPV type 11 origin.* Journal of virology, 1995. **69**(2): p. 651-660.

46. Stenlund, A., *Initiation of DNA replication: lessons from viral initiator proteins.* Nature reviews Molecular cell biology, 2003. **4**(10): p. 777-785.

47. Sverdrup, F. and S.A. Khan, *Replication of human papillomavirus (HPV) DNAs supported by the HPV type 18 E1 and E2 proteins.* Journal of Virology, 1994. **68**(1): p. 505-509.

48.   Ustav, E., et al., *The bovine papillomavirus origin of replication requires a binding site for the E2 transcriptional activator.* Proceedings of the National Academy of Sciences, 1993. **90**(3): p. 898-902.

49.   Ustav, M. and A. Stenlund, *Transient replication of BPV-1 requires two viral polypeptides encoded by the E1 and E2 open reading frames.* The EMBO journal, 1991. **10**(2): p. 449-457.

50.   Remm, M., R. Brain, and J.R. Jenkins, *The E2 binding sites determine the efficiency of replication for the origin of human papillomavirus type 18.* Nucleic acids research, 1992. **20**(22): p. 6015-6021.

51.   Ustav, M., et al., *Identification of the origin of replication of bovine papillomavirus and characterization of the viral origin recognition factor E1.* The EMBO journal, 1991. **10**(13): p. 4321-4329.

52.   Sedman, J. and A. Stenlund, *The papillomavirus E1 protein forms a DNA-dependent hexameric complex with ATPase and DNA helicase activities.* Journal of virology, 1998. **72**(8): p. 6893-6897.

53.   Yang, L., et al., *The E1 protein of bovine papilloma virus 1 is an ATP-dependent DNA helicase.* Proceedings of the National Academy of Sciences, 1993. **90**(11): p. 5086-5090.

54.   Stenlund, A., *E1 initiator DNA binding specificity is unmasked by selective inhibition of non-specific DNA binding.* The EMBO journal, 2003. **22**(4): p. 954-963.

55.   Chen, G. and A. Stenlund, *Sequential and ordered assembly of E1 initiator complexes on the papillomavirus origin of DNA replication generates progressive structural changes related to melting.* Molecular and cellular biology, 2002. **22**(21): p. 7712-7720.

56.   Enemark, E.J., A. Stenlund, and L. Joshua-Tor, *Crystal structures of two intermediates in the assembly of the papillomavirus replication initiation complex.* The EMBO journal, 2002. **21**(6): p. 1487-1496.

57.   Sanders, C.M. and A. Stenlund, *Recruitment and loading of the E1 initiator protein: an ATP-dependent process catalysed by a transcription factor.* The EMBO journal, 1998. **17**(23): p. 7044-7055.

58.   McBride, A.A., J.C. Byrne, and P.M. Howley, *E2 polypeptides encoded by bovine papillomavirus type 1 form dimers through the common carboxyl-terminal domain: transactivation is mediated by the conserved amino-terminal domain.* Proceedings of the National Academy of Sciences, 1989. **86**(2): p. 510-514.

59.   Lambert, P., B. Monk, and P. Howley, *Phenotypic analysis of bovine papillomavirus type 1 E2 repressor mutants.* Journal of virology, 1990. **64**(2): p. 950-956.

60.   Riese, D.d., et al., *Bovine papillomavirus E2 repressor mutant displays a high-copy-number phenotype and enhanced transforming activity.* Journal of virology, 1990. **64**(2): p. 944-949.

61.   Powell, M.L., et al., *NCoR1 mediates papillomavirus E8^ E2C transcriptional repression.* Journal of virology, 2010. **84**(9): p. 4451-4460.

62. Ilves, I., M. Kadaja, and M. Ustav, *Two separate replication modes of the bovine papillomavirus BPV1 origin of replication that have different sensitivity to p53.* Virus research, 2003. **96**(1-2): p. 75-84.

63. Brown, C., et al., *P53 represses human papillomavirus type 16 DNA replication via the viral E2 protein.* Virology journal, 2008. **5**(1): p. 1-9.

64. Akgül, B., et al., *Dual role of tumor suppressor p53 in regulation of DNA replication and oncogene E6-promoter activity of epidermodysplasia verruciformis-associated human papillomavirus type 8.* Virology, 2003. **308**(2): p. 279-290.

65. Hartley, K.A. and K.A. Alexander, *Human TATA binding protein inhibits human papillomavirus type 11 DNA replication by antagonizing E1-E2 protein complex formation on the viral origin of replication.* Journal of virology, 2002. **76**(10): p. 5014-5023.

66. Flores, E.R., et al., *Establishment of the human papillomavirus type 16 (HPV-16) life cycle in an immortalized human foreskin keratinocyte cell line.* Virology, 1999. **262**(2): p. 344-354.

67. Meyers, C., T.J. Mayer, and M.A. Ozbun, *Synthesis of infectious human papillomavirus type 18 in differentiating epithelium transfected with viral DNA.* Journal of virology, 1997. **71**(10): p. 7381-7386.

68. Dvoretzky, I., et al., *A quantitative in vitro focus assay for bovine papilloma virus.* Virology, 1980. **103**(2): p. 369-375.

69. Law, M.-F., et al., *Mouse cells transformed by bovine papillomavirus contain only extrachromosomal viral DNA sequences.* Proceedings of the National Academy of Sciences, 1981. **78**(5): p. 2727-2731.

70. Rabson, M.S., et al., *Bovine papillomavirus type 1 3'early region transformation and plasmid maintenance functions.* Journal of virology, 1986. **60**(2): p. 626-634.

71. Kim, K. and P.F. Lambert, *E1 protein of bovine papillomavirus 1 is not required for the maintenance of viral plasmid DNA replication.* Virology, 2002. **293**(1): p. 10-14.

72. Thomas, J.T., et al., *Human papillomavirus type 31 oncoproteins E6 and E7 are required for the maintenance of episomes during the viral life cycle in normal human keratinocytes.* Proceedings of the National Academy of Sciences, 1999. **96**(15): p. 8449-8454.

73. Li, R., et al., *Specific recognition nucleotides and their DNA context determine the affinity of E2 protein for 17 binding sites in the BPV-1 genome.* Genes & Development, 1989. **3**(4): p. 510-526.

74. Skiadopoulos, M.H. and A.A. McBride, *Bovine papillomavirus type 1 genomes and the E2 transactivator protein are closely associated with mitotic chromatin.* Journal of virology, 1998. **72**(3): p. 2079-2088.

75. Barbera, A.J., et al., *The nucleosomal surface as a docking station for Kaposi's sarcoma herpesvirus LANA.* Science, 2006. **311**(5762): p. 856-861.

76.   Burnett, S., et al., *Induction of bovine papillomavirus E2 gene expression and early region transcription by cell growth arrest: correlation with viral DNA amplification and evidence for differential promoter induction.* Journal of virology, 1990. **64**(11): p. 5529-5541.

77.   Ozbun, M.A. and C. Meyers, *Human papillomavirus type 31b E1 and E2 transcript expression correlates with vegetative viral genome amplification.* Virology, 1998. **248**(2): p. 218-230.

78.   Moody, C.A. and L.A. Laimins, *Human papillomavirus oncoproteins: pathways to transformation.* Nature Reviews Cancer, 2010. **10**(8): p. 550-560.

79.   Martinez-Zapien, D., et al., *Structure of the E6/E6AP/p53 complex required for HPV-mediated degradation of p53.* Nature, 2016. **529**(7587): p. 541-545.

80.   Huibregtse, J.M., M. Scheffner, and P.M. Howley, *A cellular protein mediates association of p53 with the E6 oncoprotein of human papillomavirus types 16 or 18.* The EMBO journal, 1991. **10**(13): p. 4129-4135.

81.   Howie, H.L., R.A. Katzenellenbogen, and D.A. Galloway, *Papillomavirus E6 proteins.* Virology, 2009. **384**(2): p. 324-334.

82.   Scheffner, M., et al., *The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53.* Cell, 1993. **75**(3): p. 495-505.

83.   Lechner, M.S. and L.A. Laimins, *Inhibition of p53 DNA binding by human papillomavirus E6 proteins.* Journal of virology, 1994. **68**(7): p. 4262-4273.

84.   Crook, T., K.H. Vousden, and J.A. Tidy, *Degradation of p53 can be targeted by HPV E6 sequences distinct from those required for p53 binding and trans-activation.* Cell, 1991. **67**(3): p. 547-556.

85.   Li, X. and P. Coffino, *High-risk human papillomavirus E6 protein has two distinct binding sites within p53, of which only one determines degradation.* Journal of virology, 1996. **70**(7): p. 4509-4516.

86.   Patel, D., et al., *The E6 protein of human papillomavirus type 16 binds to and inhibits co-activation by CBP and p300.* The EMBO journal, 1999. **18**(18): p. 5061-5072.

87.   Zimmermann, H., et al., *Interaction with CBP/p300 enables the bovine papillomavirus type 1 E6 oncoprotein to downregulate CBP/p300-mediated transactivation by p53.* Journal of General Virology, 2000. **81**(11): p. 2617-2623.

88.   Mantovani, F. and L. Banks, *Inhibition of E6 induced degradation of p53 is not sufficient for stabilization of p53 protein in cervical tumour derived cell lines.* Oncogene, 1999. **18**(22): p. 3309-3315.

89.   Phelps, W.C., et al., *The human papillomavirus type 16 E7 gene encodes transactivation and transformation functions similar to those of adenovirus E1A.* Cell, 1988. **53**(4): p. 539-547.

90.   Jones, R., et al., *Specific N-methylations of HPV-16 E7 peptides alter binding to the retinoblastoma suppressor protein.* Journal of Biological Chemistry, 1992. **267**(2): p. 908-912.

91.   Patrick, D., A. Oliff, and D. Heimbrook, *Identification of a novel retinoblastoma gene product binding site on human papillomavirus type 16 E7 protein.* Journal of Biological Chemistry, 1994. **269**(9): p. 6842-6850.

92.   Kato, J.-Y., et al., *Direct binding of cyclin D to the retinoblastoma gene product (pRb) and pRb phosphorylation by the cyclin D-dependent kinase CDK4.* Genes & development, 1993. **7**(3): p. 331-342.

93.   Johannsen, E. and P.F. Lambert, *Epigenetics of human papillomaviruses.* Virology, 2013. **445**(1-2): p. 205-212.

94.   MuÈnger, K., et al., *Biological activities and molecular targets of the human papillomavirus E7 oncoprotein.* Oncogene, 2001. **20**(54): p. 7888-7898.

95.   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.

96.   Hughes, F.J. and M.A. Romanos, *E1 protein of human papillomavirus is a DNA helicase/ATPase.* Nucleic acids research, 1993. **21**(25): p. 5817-5823.

97.   Titolo, S., et al., *Identification of domains of the human papillomavirus type 11 E1 helicase involved in oligomerization and binding to the viral origin.* Journal of virology, 2000. **74**(16): p. 7349-7361.

98.   Sun, S., et al., *Identification of a 68-kilodalton nuclear ATP-binding phosphoprotein encoded by bovine papillomavirus type 1.* Journal of Virology, 1990. **64**(10): p. 5093-5105.

99.   Del Vecchio, A.M., et al., *Transient replication of human papillomavirus DNAs.* Journal of Virology, 1992. **66**(10): p. 5949-5958.

100.   Masterson, P.J., et al., *A C-terminal helicase domain of the human papillomavirus E1 protein binds E2 and the DNA polymerase α-primase p68 subunit.* Journal of virology, 1998. **72**(9): p. 7407-7419.

101.   Graham, S.V. and A.A.A. Faizo, *Control of human papillomavirus gene expression by alternative splicing.* Virus research, 2017. **231**: p. 83-95.

102.   Blachon, S., et al., *Nucleo-cytoplasmic shuttling of high risk human Papillomavirus E2 proteins induces apoptosis.* Journal of Biological Chemistry, 2005. **280**(43): p. 36088-36098.

103.   Hegde, R.S., *The papillomavirus E2 proteins: structure, function, and biology.* Annual review of biophysics and biomolecular structure, 2002. **31**(1): p. 343-360.

104.   McBride, A.A., *The papillomavirus E2 proteins.* Virology, 2013. **445**(1-2): p. 57-79.

105.   Thain, A., et al., *CpG methylation directly inhibits binding of the human papillomavirus type 16 E2 protein to specific DNA sequences.* Journal of Virology, 1996. **70**(10): p. 7233-7235.

106.   Kim, K., et al., *Methylation patterns of papillomavirus DNA, its influence on E2 function, and implications in viral infection.* Journal of virology, 2003. **77**(23): p. 12450-12459.

107.   Smith, J.A., et al., *SMCX and components of the TIP60 complex contribute to E2 regulation of the HPV E6/E7 promoter.* Virology, 2014. **468**: p. 311-321.

108. Bernard, H.-U., *Gene expression of genital human papillomaviruses and considerations on potential antiviral approaches.* Antiviral therapy, 2002. **7**(4): p. 219-237.

109. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer.* cell, 2000. **100**(1): p. 57-70.

110. Baxter, M.K., et al., *The mitotic chromosome binding activity of the papillomavirus E2 protein correlates with interaction with the cellular chromosomal protein, Brd4.* Journal of virology, 2005. **79**(8): p. 4806-4818.

111. Ilves, I., et al., *Brd4 is involved in multiple processes of the bovine papillomavirus type 1 life cycle.* Journal of virology, 2006. **80**(7): p. 3660-3665.

112. Oliveira, J.G., L.A. Colf, and A.A. McBride, *Variations in the association of papillomavirus E2 proteins with mitotic chromosomes.* Proceedings of the National Academy of Sciences, 2006. **103**(4): p. 1047-1052.

113. Zhao, K.-N., et al., *BPV1 E2 protein enhances packaging of full-length plasmid DNA in BPV1 pseudovirions.* Virology, 2000. **272**(2): p. 382-393.

114. Johansson, C., et al., *HPV-16 E2 contributes to induction of HPV-16 late gene expression by inhibiting early polyadenylation.* The EMBO journal, 2012. **31**(14): p. 3212-3227.

115. Muller, M., et al., *Large scale genotype comparison of human papillomavirus E2-host interaction networks provides new insights for e2 molecular functions.* PLoS Pathog, 2012. **8**(6): p. e1002761.

116. Jang, M.K., et al., *A proteomic approach to discover and compare interacting partners of papillomavirus E2 proteins from diverse phylogenetic groups.* Proteomics, 2015. **15**(12): p. 2038-2050.

117. Doorbar, J., *The E4 protein; structure, function and patterns of expression.* Virology, 2013. **445**(1-2): p. 80-98.

118. Wilson, R., et al., *The full-length E1^ E4 protein of human papillomavirus type 18 modulates differentiation-dependent viral DNA amplification and late gene expression.* Virology, 2007. **362**(2): p. 453-460.

119. Nakahara, T., et al., *Human papillomavirus type 16 E1∧E4 contributes to multiple facets of the papillomavirus life cycle.* Journal of virology, 2005. **79**(20): p. 13150-13165.

120. Conrad, M., V. Bubb, and R. Schlegel, *The human papillomavirus type 6 and 16 E5 proteins are membrane-associated proteins which associate with the 16-kilodalton pore-forming protein.* Journal of virology, 1993. **67**(10): p. 6170-6178.

121. Thomsen, P., et al., *The HPV16 E5 oncogene inhibits endocytic trafficking.* Oncogene, 2000. **19**(52): p. 6023-6032.

122. Lewis, C., et al., *The first hydrophobic region of the HPV16 E5 protein determines protein cellular location and facilitates anchorage-independent growth.* Virology journal, 2008. **5**(1): p. 1-10.

123. Bouvard, V., et al., *The human papillomavirus type 16 E5 gene cooperates with the E7 gene to stimulate proliferation of primary cells and increases viral gene expression.* Virology, 1994. **203**(1): p. 73-80.

124. Oelze, I., et al., *Human papillomavirus type 16 E5 protein affects cell-cell communication in an epithelial cell line.* Journal of Virology, 1995. **69**(7): p. 4489-4494.

125. Valle, G.F. and L. Banks, *The human papillomavirus (HPV)-6 and HPV-16 E5 proteins co-operate with HPV-16 E7 in the transformation of primary rodent cells.* Journal of general virology, 1995. **76**(5): p. 1239-1245.

126. Stöppler, M.C., et al., *The E5 gene of HPV-16 enhances keratinocyte immortalization by full-length DNA.* Virology, 1996. **223**(1): p. 251-254.

127. Marchetti, B., et al., *The bovine papillomavirus oncoprotein E5 retains MHC class I molecules in the Golgi apparatus and prevents their transport to the cell surface.* Oncogene, 2002. **21**(51): p. 7808-7816.

128. Zhang, B., et al., *The E5 protein of human papillomavirus type 16 perturbs MHC class II antigen maturation in human foreskin keratinocytes treated with interferon-γ.* Virology, 2003. **310**(1): p. 100-108.

129. Buck, C.B. and B.L. Trus, *The papillomavirus virion: a machine built to hide molecular Achilles' heels.* Viral molecular machines, 2012: p. 403-422.

130. Buck, C.B., et al., *Arrangement of L2 within the papillomavirus capsid.* Journal of virology, 2008. **82**(11): p. 5190-5197.

131. Giroglou, T., et al., *Human papillomavirus infection requires cell surface heparan sulfate.* Journal of virology, 2001. **75**(3): p. 1565-1570.

132. Johnson, K.M., et al., *Role of heparan sulfate in attachment to and infection of the murine female genital tract by human papillomavirus.* Journal of virology, 2009. **83**(5): p. 2067-2074.

133. Roden, R., et al., *In vitro generation and type-specific neutralization of a human papillomavirus type 16 virion pseudotype.* Journal of virology, 1996. **70**(9): p. 5875-5883.

134. Stauffer, Y., et al., *Infectious human papillomavirus type 18 pseudovirions.* Journal of molecular biology, 1998. **283**(3): p. 529-536.

135. Ahola, T. and L. Kääriäinen, *Reaction in alphavirus mRNA capping: formation of a covalent complex of nonstructural protein nsP1 with 7-methyl-GMP.* Proceedings of the National Academy of Sciences, 1995. **92**(2): p. 507-511.

136. Topisirovic, I., et al., *Cap and cap-binding proteins in the control of gene expression.* Wiley Interdisciplinary Reviews: RNA, 2011. **2**(2): p. 277-298.

137. Hocine, S., R.H. Singer, and D. Grünwald, *RNA processing and export.* Cold Spring Harbor perspectives in biology, 2010. **2**(12): p. a000752.

138. Wu, X. and G. Brewer, *The regulation of mRNA stability in mammalian cells: 2.0.* Gene, 2012. **500**(1): p. 10-21.

139. Bienroth, S., W. Keller, and E. Wahle, *Assembly of a processive messenger RNA polyadenylation complex.* The EMBO journal, 1993. **12**(2): p. 585-594.

140. Beaudoing, E., et al., *Patterns of variant polyadenylation signal usage in human genes.* Genome research, 2000. **10**(7): p. 1001-1010.

141.   Balbo, P.B. and A. Bohm, *Mechanism of poly (A) polymerase: structure of the enzyme-MgATP-RNA ternary complex and kinetic analysis.* Structure, 2007. **15**(9): p. 1117-1131.

142.   Wahle, E., *Poly (A) tail length control is caused by termination of processive synthesis.* Journal of Biological Chemistry, 1995. **270**(6): p. 2800-2808.

143.   Berget, S.M., C. Moore, and P.A. Sharp, *Spliced segments at the 5' terminus of adenovirus 2 late mRNA.* Proceedings of the National Academy of Sciences, 1977. **74**(8): p. 3171-3175.

144.   Black, D.L., *Mechanisms of alternative pre-messenger RNA splicing.* Annual review of biochemistry, 2003. **72**(1): p. 291-336.

145.   Hallegger, M., M. Llorian, and C.W. Smith, *Alternative splicing: global insights.* The FEBS journal, 2010. **277**(4): p. 856-866.

146.   Taggart, A.J., et al., *Large-scale mapping of branchpoints in human pre-mRNA transcripts in vivo.* Nature structural & molecular biology, 2012. **19**(7): p. 719.

147.   Moore, M.J. and P.A. Sharp, *Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing.* Nature, 1993. **365**(6444): p. 364-368.

148.   Modrek, B. and C. Lee, *A genomic view of alternative splicing.* Nature genetics, 2002. **30**(1): p. 13-19.

149.   Das, R., Z. Zhou, and R. Reed, *Functional association of U2 snRNP with the ATP-independent spliceosomal complex E.* Molecular cell, 2000. **5**(5): p. 779-787.

150.   Nilsen, T.W., *A parallel spliceosome.* Science, 1996. **273**(5283): p. 1813-1813.

151.   Manley, J.L., et al., *SR proteins and splicing control 1569.* Genes & development, 1996.

152.   Wang, Z., et al., *General and specific functions of exonic splicing silencers in splicing control.* Molecular cell, 2006. **23**(1): p. 61-70.

153.   Zhou, Z. and X.-D. Fu, *Regulation of splicing by SR proteins and SR protein-specific kinases.* Chromosoma, 2013. **122**(3): p. 191-207.

154.   Hertel, K.J. and T. Maniatis, *Serine–arginine (SR)-rich splicing factors have an exon-independent function in pre-mRNA splicing.* Proceedings of the National Academy of Sciences, 1999. **96**(6): p. 2651-2655.

155.   Graveley, B.R., K.J. Hertel, and T. Maniatis, *SR proteins are 'locators' of the RNA splicing machinery.* Current Biology, 1999. **9**(1): p. R6-R7.

156.   Han, S.P., Y.H. Tang, and R. Smith, *Functional diversity of the hnRNPs: past, present and perspectives.* Biochemical Journal, 2010. **430**(3): p. 379-392.

157.   Lim, K.H., et al., *Using positional distribution to identify splicing elements and predict pre-mRNA processing defects in human genes.* Proceedings of the National Academy of Sciences, 2011. **108**(27): p. 11093-11098.

158.   Elliott, D.J., et al., *Dynamic changes in the subnuclear organisation of pre-mRNA splicing proteins and RBM during human germ cell development.* Journal of Cell Science, 1998. **111**(9): p. 1255-1265.

159. Cáceres, J.F., G.R. Screaton, and A.R. Krainer, *A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm.* Genes & development, 1998. **12**(1): p. 55-66.

160. Michlewski, G., J.R. Sanford, and J.F. Cáceres, *The splicing factor SF2/ASF regulates translation initiation by enhancing phosphorylation of 4E-BP1.* Molecular cell, 2008. **30**(2): p. 179-189.

161. Bedard, K.M., S. Daijogo, and B.L. Semler, *A nucleo-cytoplasmic SR protein functions in viral IRES-mediated translation initiation.* The EMBO journal, 2007. **26**(2): p. 459-467.

162. Long, J.C. and J.F. Caceres, *The SR protein family of splicing factors: master regulators of gene expression.* Biochemical Journal, 2009. **417**(1): p. 15-27.

163. Sanford, J., J. Ellis, and J. Caceres, *Multiple roles of arginine/serine-rich splicing factors in RNA processing.* Biochemical Society Transactions, 2005. **33**(3): p. 443-446.

164. Shepard, P.J. and K.J. Hertel, *The SR protein family.* Genome biology, 2009. **10**(10): p. 1-9.

165. Cáceres, J.F., et al., *Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity.* The Journal of cell biology, 1997. **138**(2): p. 225-238.

166. Hedley, M.L., H. Amrein, and T. Maniatis, *An amino acid sequence motif sufficient for subnuclear localization of an arginine/serine-rich splicing factor.* Proceedings of the National Academy of Sciences, 1995. **92**(25): p. 11524-11528.

167. Graveley, B.R., *Sorting out the complexity of SR protein functions.* Rna, 2000. **6**(9): p. 1197.

168. Shen, H., J.L. Kan, and M.R. Green, *Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly.* Molecular cell, 2004. **13**(3): p. 367-376.

169. Shen, H. and M.R. Green, *A pathway of sequential arginine-serine-rich domain-splicing signal interactions during mammalian spliceosome assembly.* Molecular cell, 2004. **16**(3): p. 363-373.

170. Sanford, J.R., et al., *A novel role for shuttling SR proteins in mRNA translation.* Genes & development, 2004. **18**(7): p. 755-768.

171. Wang, Z. and C.B. Burge, *Splicing regulation: from a parts list of regulatory elements to an integrated splicing code.* Rna, 2008. **14**(5): p. 802-813.

172. Tacke, R. and J.L. Manley, *Determinants of SR protein specificity.* Current opinion in cell biology, 1999. **11**(3): p. 358-362.

173. Somberg, M. and S. Schwartz, *Multiple ASF/SF2 sites in the human papillomavirus type 16 (HPV-16) E4-coding region promote splicing to the most commonly used 3'-splice site on the HPV-16 genome.* Journal of virology, 2010. **84**(16): p. 8219-8230.

174. Somberg, M., et al., *Serine/arginine-rich protein 30c activates human papillomavirus type 16 L1 mRNA expression via a bimodal mechanism.* Journal of general virology, 2011. **92**(10): p. 2411-2421.

175. Jia, R., et al., *Control of the papillomavirus early-to-late switch by differentially expressed SRp20.* Journal of virology, 2009. **83**(1): p. 167-180.

176. Martinez-Contreras, R., et al., *8 hnRNP proteins and splicing control.* Advances in Experimental Medicine & Biology, 2008. **623**: p. 123.

177. Dreyfuss, G., et al., *hnRNP proteins and the biogenesis of mRNA.* Annual review of biochemistry, 1993. **62**(1): p. 289-321.

178. Zhao, T., et al., *hnRNP A1 regulates UV-induced NF-κ B signalling through destabilization of cIAP1 mRNA.* Cell Death & Differentiation, 2009. **16**(2): p. 244-252.

179. Fähling, M., et al., *Heterogeneous nuclear ribonucleoprotein-A2/B1 modulate collagen prolyl 4-hydroxylase, α (I) mRNA stability.* Journal of Biological Chemistry, 2006. **281**(14): p. 9279-9286.

180. Geuens, T., D. Bouhy, and V. Timmerman, *The hnRNP family: insights into their role in health and disease.* Human genetics, 2016. **135**(8): p. 851-867.

181. Görlach, M., et al., *Interaction of the RNA-binding domain of the hnRNP C proteins with RNA.* The EMBO journal, 1992. **11**(9): p. 3289-3295.

182. Kiledjian, M. and G. Dreyfuss, *Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box.* The EMBO journal, 1992. **11**(7): p. 2655-2664.

183. Dominguez, C. and F.H.-T. Allain, *NMR structure of the three quasi RNA recognition motifs (qRRMs) of human hnRNP F and interaction studies with Bcl-x G-tract RNA: a novel mode of RNA recognition.* Nucleic acids research, 2006. **34**(13): p. 3634-3645.

184. Siomi, H., et al., *The pre-mRNA binding K protein contains a novel evolutionary conserved motif.* Nucleic acids research, 1993. **21**(5): p. 1193-1198.

185. Chaudhury, A., P. Chander, and P.H. Howe, *Heterogeneous nuclear ribonucleoproteins (hnRNPs) in cellular processes: Focus on hnRNP E1's multifunctional regulatory roles.* Rna, 2010. **16**(8): p. 1449-1462.

186. 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.* Journal of virology, 2016. **90**(20): p. 9138-9152.

187. 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 research, 2013. **41**(22): p. 10488-10508.

188. Öberg, D., et al., *Mutational inactivation of two distinct negative RNA elements in the human papillomavirus type 16 L2 coding region induces production of high levels of L2 in human cells.* Journal of virology, 2003. **77**(21): p. 11674-11684.

189. Öberg, D., et al., *A downstream polyadenylation element in human papillomavirus type 16 L2 encodes multiple GGG motifs and interacts with hnRNP H.* Journal of virology, 2005. **79**(14): p. 9254-9269.

190. Jean-Philippe, J., S. Paz, and M. Caputi, *hnRNP A1: the Swiss army knife of gene expression.* International journal of molecular sciences, 2013. **14**(9): p. 18999-19024.

191. Kenan, D.J., C.C. Query, and J.D. Keene, *RNA recognition: towards identifying determinants of specificity.* Trends in biochemical sciences, 1991. **16**: p. 214-220.

192. Dreyfuss, G., L. Philipson, and I.W. Mattaj, *Ribonucleoprotein particles in cellular processes.* The Journal of cell biology, 1988. **106**(5): p. 1419-1425.

193. Guil, S., J.C. Long, and J.F. Cáceres, *hnRNP A1 relocalization to the stress granules reflects a role in the stress response.* Molecular and cellular biology, 2006. **26**(15): p. 5744-5758.

194. Roy, R., et al., *hnRNPA1 couples nuclear export and translation of specific mRNAs downstream of FGF-2/S6K2 signalling.* Nucleic acids research, 2014. **42**(20): p. 12483-12497.

195. Roy, R., et al., *Emerging roles of hnRNPA1 in modulating malignant transformation.* Wiley Interdisciplinary Reviews: RNA, 2017. **8**(6): p. e1431.

196. Pino, I., et al., *Altered patterns of expression of members of the heterogeneous nuclear ribonucleoprotein (hnRNP) family in lung cancer.* Lung cancer, 2003. **41**(2): p. 131-143.

197. Brockstedt, E., et al., *Identification of apoptosis-associated proteins in a human Burkitt lymphoma cell line: cleavage of heterogeneous nuclear ribonucleoprotein A1 by caspase 3.* Journal of Biological Chemistry, 1998. **273**(43): p. 28057-28064.

198. Shi, Y., et al., *Cyclin D1 and c-myc internal ribosome entry site (IRES)-dependent translation is regulated by AKT activity and enhanced by rapamycin through a p38 MAPK-and ERK-dependent pathway.* Journal of Biological Chemistry, 2005. **280**(12): p. 10964-10973.

199. Le, P.N., et al., *TERRA, hnRNP A1, and DNA-PKcs interactions at human telomeres.* Frontiers in oncology, 2013. **3**: p. 91.

200. Sui, J., et al., *DNA-PKcs phosphorylates hnRNP-A1 to facilitate the RPA-to-POT1 switch and telomere capping after replication.* Nucleic acids research, 2015. **43**(12): p. 5971-5983.

201. Tahara, H., et al., *G-Quadruplex stabilization by telomestatin induces TRF2 protein dissociation from telomeres and anaphase bridge formation accompanied by loss of the 3' telomeric overhang in cancer cells.* Oncogene, 2006. **25**(13): p. 1955-1966.

202. He, Y. and R. Smith, *Nuclear functions of heterogeneous nuclear ribonucleoproteins A/B.* Cellular and molecular life sciences, 2009. **66**(7): p. 1239-1256.

203. Villarroya-Beltri, C., et al., *Sumoylated hnRNPA2B1 controls the sorting of miRNAs into exosomes through binding to specific motifs.* Nature communications, 2013. **4**(1): p. 1-10.

204. Elliott, D.J., et al., *An evolutionarily conserved germ cell-specific hnRNP is encoded by a retrotransposed gene.* Human Molecular Genetics, 2000. **9**(14): p. 2117-2124.

205. Lingenfelter, P.A., et al., *Expression and conservation of processed copies of the RBMX gene.* Mammalian Genome, 2001. **12**(7): p. 538-545.

206. Soulard, M., et al., *hnRNP G: sequence and characterization of a glycosylated RNA-binding protein.* Nucleic acids research, 1993. **21**(18): p. 4210-4217.

207. Kanhoush, R., et al., *Novel domains in the hnRNP G/RBMX protein with distinct roles in RNA binding and targeting nascent transcripts.* Nucleus, 2010. **1**(1): p. 109-122.

208. Heinrich, B., et al., *Heterogeneous nuclear ribonucleoprotein G regulates splice site selection by binding to CC (A/C)-rich regions in pre-mRNA.* Journal of Biological Chemistry, 2009. **284**(21): p. 14303-14315.

209. Hofmann, Y. and B. Wirth, *hnRNP-G promotes exon 7 inclusion of survival motor neuron (SMN) via direct interaction with Htra2-β1.* Human molecular genetics, 2002. **11**(17): p. 2037-2049.

210. Maziuk, B.F., et al., *RNA binding proteins co-localize with small tau inclusions in tauopathy.* Acta neuropathologica communications, 2018. **6**(1): p. 1-14.

211. Nasim, M.T., et al., *HnRNP G and Tra2β: opposite effects on splicing matched by antagonism in RNA binding.* Human molecular genetics, 2003. **12**(11): p. 1337-1348.

212. Zhou, K.I., et al., *Regulation of co-transcriptional pre-mRNA splicing by m6A through the low-complexity protein hnRNPG.* Molecular cell, 2019. **76**(1): p. 70-81. e9.

213. Adamson, B., et al., *A genome-wide homologous recombination screen identifies the RNA-binding protein RBMX as a component of the DNA-damage response.* Nature cell biology, 2012. **14**(3): p. 318-328.

214. Shin, K.-H., et al., *p53 promotes the fidelity of DNA end-joining activity by, in part, enhancing the expression of heterogeneous nuclear ribonucleoprotein G.* DNA repair, 2007. **6**(6): p. 830-840.

215. Doorbar, J., et al., *Detection of novel splicing patterns in a HPV16-containing keratinocyte cell line.* Virology, 1990. **178**(1): p. 254-262.

216. 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).

217. 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.

218. Rosenberger, S., et al., *Alternative splicing of human papillomavirus type-16 E6/E6* early mRNA is coupled to EGF signaling via Erk1/2 activation.* Proceedings of the National Academy of Sciences, 2010. **107**(15): p. 7006-7011.

219. Grassmann, K., et al., *Identification of a differentiation-inducible promoter in the E7 open reading frame of human papillomavirus type 16 (HPV-16) in raft cultures of a new cell line containing high copy numbers of episomal HPV-16 DNA.* Journal of virology, 1996. **70**(4): p. 2339-2349.

220. Deng, W., et al., *mRNA splicing regulates human papillomavirus type 11 E1 protein production and DNA replication.* Journal of virology, 2003. **77**(19): p. 10213-10226.

221. Männik, A., et al., *Induction of the bovine papillomavirus origin "onion skin"-type DNA replication at high E1 protein concentrations in vivo.* Journal of virology, 2002. **76**(11): p. 5835-5845.

222. Hubert, W.G. and L.A. Laimins, *Human papillomavirus type 31 replication modes during the early phases of the viral life cycle depend on transcriptional and posttranscriptional regulation of E1 and E2 expression.* Journal of virology, 2002. **76**(5): p. 2263-2273.

223. 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.* Journal of virology, 2004. **78**(20): p. 10888-10905.

224. Zhao, X., et al., *A 57-nucleotide upstream early polyadenylation element in human papillomavirus type 16 interacts with hFip1, CstF-64, hnRNP C1/C2, and polypyrimidine tract binding protein.* Journal of virology, 2005. **79**(7): p. 4270-4288.

225. Graham, S.V., *Papillomavirus 3'UTR regulatory elements.* Frontiers in bioscience: a journal and virtual library, 2008. **13**: p. 5646-5663.

226. Jeon, S., B.L. Allen-Hoffmann, and P.F. Lambert, *Integration of human papillomavirus type 16 into the human genome correlates with a selective growth advantage of cells.* Journal of virology, 1995. **69**(5): p. 2989-2997.

227. Cheunim, T., et al., *The alternative splicing factor hnRNP A1 is up-regulated during virus-infected epithelial cell differentiation and binds the human papillomavirus type 16 late regulatory element.* Virus research, 2008. **131**(2): p. 189-198.

228. Dhanjal, S., et al., *Heterogeneous nuclear ribonucleoprotein C proteins interact with the human papillomavirus type 16 (HPV16) early 3'-untranslated region and alleviate suppression of HPV16 Late L1 mRNA splicing.* Journal of Biological Chemistry, 2015. **290**(21): p. 13354-13371.

229. Nagy, A., *Cre recombinase: the universal reagent for genome tailoring.* genesis, 2000. **26**(2): p. 99-109.

230. 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.* Journal of virology, 2002. **76**(6): p. 2739-2752.

231. Nilsson, K., et al., *HPV16 E5 is produced from an HPV16 early mRNA spliced from SD226 to SA3358.* Virus research, 2018. **244**: p. 128-136.

232. Bruun, G.H., et al., *Global identification of hnRNP A1 binding sites for SSO-based splicing modulation.* BMC biology, 2016. **14**(1): p. 1-19.

233. Huelga, S.C., et al., *Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins.* Cell reports, 2012. **1**(2): p. 167-178.

234.   Yu, H., et al., *hnRNP G prevents inclusion on the HPV16 L1 mRNAs of the central exon between splice sites SA3358 and SD3632.* Journal of General Virology, 2018. **99**(3): p. 328-343.

235.   Kozak, M., *Regulation of translation in eukaryotic systems.* Annual review of cell biology, 1992. **8**(1): p. 197-225.

236.   Luukkonen, B., W. Tan, and S. Schwartz, *Efficiency of reinitiation of translation on human immunodeficiency virus type 1 mRNAs is determined by the length of the upstream open reading frame and by intercistronic distance.* Journal of virology, 1995. **69**(7): p. 4086-4094.

237.   Alloul, N. and L. Sherman, *The E2 protein of human papillomavirus type 16 is translated from a variety of differentially spliced polycistronic mRNAs.* Journal of general virology, 1999. **80**(1): p. 29-37.