



LUND UNIVERSITY

The Regulation of Human Papillomavirus Type16 Early Gene Expression

Cui, Xiaoxu

2022

Document Version:

Publisher's PDF, also known as Version of record

[Link to publication](#)

Citation for published version (APA):

Cui, X. (2022). *The Regulation of Human Papillomavirus Type16 Early Gene Expression*. [Doctoral Thesis (compilation), Department of Laboratory Medicine]. Lund University, Faculty of Medicine.

Total number of authors:

1

General rights

Unless other specific re-use rights are stated the following general rights apply:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

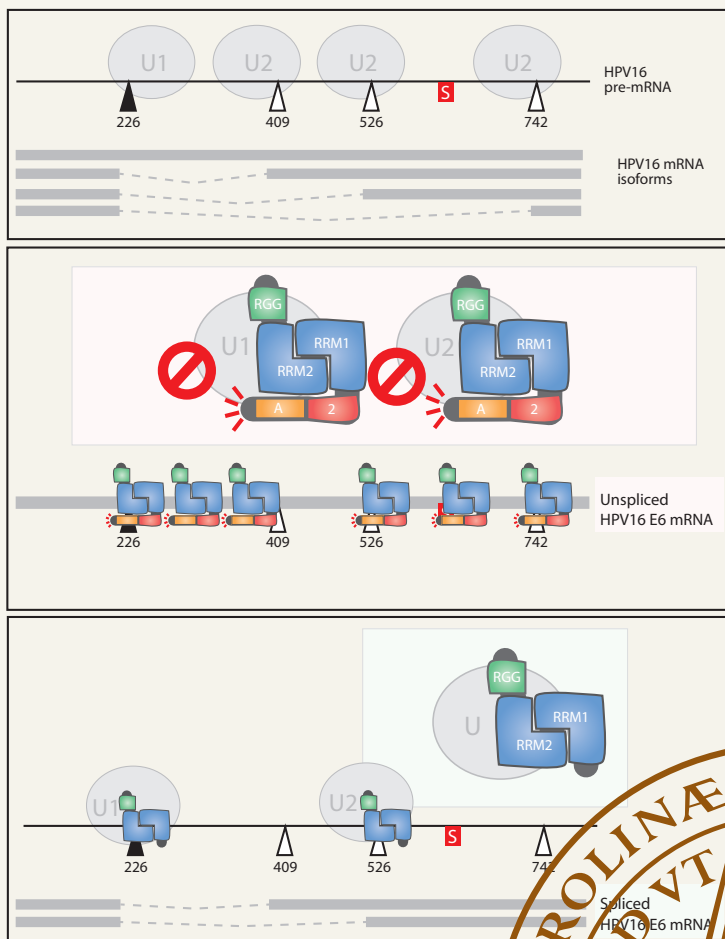
LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

The Regulation of Human Papillomavirus Type 16 Early Gene Expression

XIAOXU CUI

DEPARTMENT OF LABORATORY MEDICINE | FACULTY OF MEDICINE | LUND UNIVERSITY



The Regulation of Human Papillomavirus Type16 Early Gene Expression

The Regulation of Human Papillomavirus Type16 Early Gene Expression

Xiaoxu Cui



LUND
UNIVERSITY

DOCTORAL DISSERTATION

by due permission of the Faculty of Medicine, Lund University, Sweden.

To be defended at Segerfalksalen on 1st, June at 9:00 AM

Faculty opponent

Docent Jonas Klingström, Department of Medicine, Huddinge,
Karolinska Institutet

Supervisor: Professor Stefan Schwartz

Co-supervisors: Docent Ola Forslund and Dr. Naoko Kajitani

Organization LUND UNIVERSITY Department of Laboratory Medicine Author(s) Xiaoxu Cui	Document name DOCTORAL DISSERTATION	
	Date of issue: 1st June,2022	
	Sponsoring organization	
Title and subtitle The Regulation of Human Papillomavirus Type16 Early Gene Expression		
Abstract Human papillomaviruses (HPVs) are small non enveloped viruses that contain a double-stranded circular DNA genome of approximately 8kb in size. It has been estimated that approximately 5% of all cancers are caused by HPV infection. Cervical cancer represents the majority of HPV-associated anogenital cancers worldwide. HPV type16 (HPV16) accounts for 65% of cervical cancer and approximately 90% of the other HPV-associated cancers, such as head-and-neck-squamous cell carcinomas (HNSCC). Persistent infection is the critical risk factor for HPV16-associated cancer progression. Dysregulation of HPV16 gene expression, especially oncogenes E6/E7, viral replication and transcription factors E1/E2, and the highly pathogenic major structural protein L1, may contribute to establishment of persistence that may, in the end, result in cancer. In this study, we demonstrated that cellular splicing factor hnRNP D and m6A modification of HPV16 mRNA play significant roles in the regulation of HPV16 gene expression. In addition, the long control region (LCR) of the HPV16 genome appears to control E6/E7 mRNA splicing, indicating that the LCR DNA may contain splicing regulatory elements. Cellular RNA binding protein (RBP) hnRNP D acts as a splicing inhibitor of HPV16 E1/E2 and E6/E7 mRNAs, generating intron-containing E1 and E6 mRNAs. N- and C-termini of hnRNP D contributed to HPV16 mRNA splicing control differently. The N- terminus of hnRNP D played the core inhibitory function. Furthermore, the RGG domain of the C- terminus contributed to splicing inhibition by interacting with the splicing machinery. Also, hnRNP D intensively interacted with HPV16 mRNAs in an RRM1 dependent manner. As a result, the cytoplasmic levels of intron-retained HPV16 mRNAs were increased in the presence of hnRNP D. In addition, we detected direct binding of hnRNP D to HPV16 mRNAs in an HPV16-driven tonsillar cancer cell line and in HPV16-immortalized human keratinocytes. Furthermore, knockdown of hnRNP D in HPV16-driven cervical cancer cells enhanced the production of the HPV16 E7 oncoprotein. Our results suggest that hnRNP D plays significant roles in the regulation of HPV gene expression and HPV-associated cancer development. HPV16 mRNA m6A modification also affected alternative splicing. Overexpression of m6A demethylase ALKBH5 promoted E6 mRNA production and altered L1 splicing by regulating exon skipping. Overexpression of methyltransferase METTL3 induced E1 mRNA production and altered L1 oppositely to ALKBH5. YTHDC1 worked as an m6A reader that could induce E6 mRNA production. Our results suggest that m6A-modifications of HPV16 mRNAs contribute to the control of HPV16 gene expression. In summary, in this thesis we report that the cellular hnRNP D protein plays a major role in the control of production of the HPV16 E1 mRNA as well as in the production of the E7 oncogene mRNA. We also found yet unidentified splicing regulatory elements in the HPV16 E6 coding region that affected E6/E7 mRNA splicing. In addition, non-transcribed sequences in the LCR may contribute to E6/E7 splicing regulation by altering the association of RNA polymerase with RNA binding proteins required for HPV16 mRNA splicing during a co-transcriptional splicing process. Finally, we found that m6A modification of HPV16 mRNAs contribute to control of HPV16 gene expression at the level of RNA processing. Taken together, these results enhance our understanding of the carcinogenic human papillomavirus HPV16.		
Key words Human papillomavirus, splicing, hnRNP D, m6A, LCR		
Classification system and/or index terms (if any)		
Supplementary bibliographical information		Language English
ISSN and key title 1652-8220		ISBN 978-91-8021-240-3
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

Xiaoxu Cui

Date 2022-04-26

The Regulation of Human Papillomavirus Type16 Early Gene Expression

Xiaoxu Cui



LUND
UNIVERSITY

Coverphoto by Xiaoxu Cui

Copyright pp 1-90 Xiaoxu Cui

Paper 1 © Xiaoxu Cui

Paper 2 © Xiaoxu Cui

Paper 3 © Yunji Zheng

Paper 4 © Xiaoxu Cui (Manuscript unpublished)

Faculty of Medicine
Department of Laboratory Medicine


ISBN 978-91-8021-240-3

ISSN 1652-8220

Printed in Sweden by Media-Tryck, Lund University
Lund 2022



Media-Tryck is a Nordic Swan Ecolabel
certified provider of printed material.
Read more about our environmental
work at www.mediatryck.lu.se

MADE IN SWEDEN 

To my loved ones

Yesterday's the past, tomorrow's the future, but today is a gift. That's why it's called the present.

Bil Keane

静水流深

Table of Contents

List of articles and manuscripts	10
Articles not included in the thesis.....	10
Abstract.....	11
Populärvetenskaplig Sammanfattning.....	13
Abbreviations	15
1 Introduction	17
1.1 Human Papillomavirus classifications and diseases	17
1.2 HPV16 genome and proteins	18
1.2.1 Long Control Region	19
1.2.2 E6 and E7 proteins	20
1.2.3 E1 and E2 proteins	21
1.2.4 E4 proteins	22
1.2.5 E5 proteins	22
1.2.6 L1 and L2 proteins	22
1.3 HPV16 life cycle	23
1.3.1 Virus entry.....	24
1.3.2 Viral replication	25
1.4 HPV16 gene expression and regulation	26
1.4.1 mRNA Splicing.....	28
1.4.2 Alternative mRNA splicing in HPV16	30
1.4.3 Polyadenylation.....	31
1.4.4 Polyadenylation in HPV16.....	32
1.4.5 m6A RNA methylation	32
1.5 hnRNP family	34
1.5.1 hnRNP D.....	36
1.6 HPV16 plasmids of special interest	37
1.6.1 Subgenomic reporter plasmid pC97ELsLuc.....	37
1.6.2 Subgenomic reporter plasmid pBELsLuc	38
1.6.3 Genomic plasmid pHPV16AN.....	39
2 Aim of the Thesis	40
3 Material and Methods.....	41
3.1 Cells	41

3.2	Plasmids	41
3.3	Transfections.....	42
3.4	Nuclear and Cytoplasmic Extraction	42
3.5	RNA extraction, RT-PCR and Real-time quantitative PCR (qPCR)	43
3.6	Secreted luciferase assay.....	43
3.7	Western blotting.....	44
3.8	ssRNA oligo pull down.....	44
3.9	Co-Immunoprecipitation.....	44
3.10	UV-crosslinking and immunoprecipitation (CLIP)	45
3.11	Ribonucleoprotein (RNP) immunoprecipitation (RIP) analysis	45
3.12	siRNA library and siRNA transfections.....	45
3.13	Lentiviral based shRNA for knockdown	46
3.14	<i>In vitro</i> translation assay	46
3.15	<i>In vitro</i> RNA syntheses.....	47
3.16	RNA preparation for MeRIP-seq.....	47
3.17	Quantitations	47
4	Results.....	53
4.1	HnRNP D activates production of HPV16 E1 and E6 mRNA by promoting intron retention.....	53
4.1.1	Both N- and C-termini of hnRNP D40 contribute to hnRNP D40 splicing control ability, but in different ways.....	53
4.1.2	hnRNP D40 increases the levels of HPV16 intron-retained E6 and E1 mRNAs in the cytoplasm	56
4.1.3	hnRNP D40 interacts with HPV16 mRNAs in an HPV16-positive cancer cell line and in HPV16-immortalized human keratinocytes	58
4.2	m6A modifications on HPV16 mRNA regulate alternative splicing	60
4.2.1	m6A “eraser” ALKBH5, “writer” METTL3 and “reader” YTHDC1 alter HPV16 mRNA splicing.....	60
4.2.2	HPV16 mRNAs are m6A-methylated in HPV16-positive tonsillar cancer cell line HN26	63
4.3	Identification of nucleotide substitutions in the 5'-end of HPV16 early mRNAs and in the non-transcribed long control region that affect E6 and E7 mRNA splicing.....	67
5	Discussion and Future Perspectives.....	70
6	Acknowledgements	74
7	References	76

List of articles and manuscripts

1. **Xiaoxu Cui**, Chengyu Hao, Lijing Gong, Naoko Kajitani and Stefan Schwartz. HnRNP D activates production of HPV16 E1 and E6 mRNAs by promoting intron retention. *Nucleic Acids Research*.2022 Mar 2; **gkac**. doi: [10.1093/nar/gkac132](https://doi.org/10.1093/nar/gkac132). PMID: **35234917**.
2. **Xiaoxu Cui**, Kersti Nilsson, Naoko Kajitani, Stefan Schwartz. Overexpression of m6A-factors METTL3, ALKBH5 and YTHDC1 alters HPV16 mRNA splicing. *Virus Genes*.2022 Feb 21; doi: [10.1007/s11262-022-01889-6](https://doi.org/10.1007/s11262-022-01889-6). PMID: **35190939**.
3. Yunji Zheng, **Xiaoxu Cui**, Kersti Nilsson, Haoran Yu, Lijing Gong, Chengjun Wu, Stefan Schwartz. 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](https://doi.org/10.1016/j.virusres.2020.198004).
4. **Xiaoxu Cui**, Naoko Kajitani and Stefan Schwartz. Identification of nucleotide substitutions in the 5'-end of HPV16 early mRNAs and in the non-transcribed long control region that affect E6 and E7 mRNA splicing. **Manuscript**.

Articles not included in the thesis

1. Yunji Zheng, Johanna Jönsson, Chengyu Hao, Shoja Chaghervand S, **Xiaoxu Cui**, Naoko Kajitani, Lijing Gong, Chengjun Wu, Stefan Schwartz. Heterogeneous Nuclear Ribonucleoprotein A1 (hnRNP A1) and hnRNP A2 Inhibit Splicing to Human Papillomavirus 16 Splice Site SA409 through a UAG-Containing Sequence in the E7 Coding Region. *J Virol*.2020 Sep 29;94(20): e01509-20. doi: [10.1128/JVI.01509-20](https://doi.org/10.1128/JVI.01509-20).
2. Chengyu Hao, Lijing Gong, **Xiaoxu Cui**, Johanna Jönsson, Yunji Zheng, Chengjun Wu, Naoko Kajitani, Stefan Schwartz. Identification of heterogenous nuclear ribonucleoproteins (hnRNPs) and serine- and arginine-rich (SR) proteins that induce human papillomavirus type 16 late gene expression and alter L1 mRNA splicing. *Arch Virol*. 2021 Dec 3. doi: [10.1007/s00705-021-05317-2](https://doi.org/10.1007/s00705-021-05317-2).
3. Chengyu Hao, Yunji Zheng, Johanna Jönsson, Xiaoxu Cui, Haoran Yu, Chengjun Wu, Naoko Kajitani, Stefan Schwartz. hnRNP G/RBMX enhances HPV16 E2 mRNA splicing through a novel splicing enhancer and inhibits production of spliced E7 oncogene mRNAs. *Nucleic Acids Res*. 2022 Apr 22;50(7):3867-3891. doi: [10.1093/nar/gkac213](https://doi.org/10.1093/nar/gkac213). PMID: **35357488**.

Abstract

Human papillomaviruses (HPVs) are small non enveloped viruses that contain a double-stranded circular DNA genome of approximately 8kb in size. It has been estimated that approximately 5% of all cancers are caused by HPV infection. Cervical cancer represents the majority of HPV-associated anogenital cancers worldwide. HPV type16 (HPV16) accounts for 65% of cervical cancer and approximately 90% of the other HPV-associated cancers, such as head-and-neck-squamous cell carcinomas (HNSCC). Persistent infection is the critical risk factor for HPV16-associated cancer progression. Dysregulation of HPV16 gene expression, especially oncogenes E6/E7, viral replication and transcription factors E1/E2, and the highly pathogenic major structural protein L1, may contribute to establishment of persistence that may, in the end, result in cancer. In this study, we demonstrated that cellular splicing factor hnRNP D and m6A modification of HPV16 mRNA play significant roles in the regulation of HPV16 gene expression. In addition, the long control region (LCR) of the HPV16 genome appears to control E6/E7 mRNA splicing, indicating that the LCR DNA may contain splicing regulatory elements.

Cellular RNA binding protein (RBP) hnRNP D acts as a splicing inhibitor of HPV16 E1/E2 and E6/E7 mRNAs, generating intron-containing E1 and E6 mRNAs. N- and C-termini of hnRNP D contributed to HPV16 mRNA splicing control differently. The N- terminus of hnRNP D played the core inhibitory function. Furthermore, the RGG domain of the C- terminus contributed to splicing inhibition by interacting with the splicing machinery. Also, hnRNP D intensively interacted with HPV16 mRNAs in an RRM1 dependent manner. As a result, the cytoplasmic levels of intron-retained HPV16 mRNAs were increased in the presence of hnRNP D. In addition, we detected direct binding of hnRNP D to HPV16 mRNAs in an HPV16-driven tonsillar cancer cell line and in HPV16-immortalized human keratinocytes. Furthermore, knockdown of hnRNP D in HPV16-driven cervical cancer cells enhanced the production of the HPV16 E7 oncoprotein. Our results suggest that hnRNP D plays significant roles in the regulation of HPV gene expression and HPV-associated cancer development.

HPV16 mRNA m6A modification also affected alternative splicing. Overexpression of m6A demethylase ALKBH5 promoted E6 mRNA production and altered L1 splicing by regulating exon skipping. Overexpression of methyltransferase METTL3 induced E1 mRNA production and altered L1 oppositely to ALKBH5.

YTHDC1 worked as an m6A reader that could induce E6 mRNA production. Our results suggest that m6A-modifications of HPV16 mRNAs contribute to the control of HPV16 gene expression.

In summary, in this thesis we report that the cellular hnRNP D protein plays a major role in the control of production of the HPV16 E1 mRNA as well as in the production of the E7 oncogene mRNA. We also found yet unidentified splicing regulatory elements in the HPV16 E6 coding region that affected E6/E7 mRNA splicing. In addition, non-transcribed sequences in the LCR may contribute to E6/E7 splicing regulation by altering the association of RNA polymerase with RNA binding proteins required for HPV16 mRNA splicing during a co-transcriptional splicing process. Finally, we found that m6A modification of HPV16 mRNAs contribute to control of HPV16 gene expression at the level of RNA processing. Taken together, these results enhance our understanding of the carcinogenic human papillomavirus HPV16.

Populärvetenskaplig Sammanfattning

Humana papillomvirus (HPVs) är en grupp små icke-höljebärande virus som har ett dubbelsträngat cirkulärt DNA-genom på ungefär åtta tusen baspar. Det har uppskattats att cirka 5 % av alla cancerfall orsakas av en HPV-infektion. Livmoderhalscancer representerar majoriteten av de HPV-orsakade fallen av anogenital cancer över hela världen. HPV typ16 (HPV16) orsakar ca 65 % av all livmoderhalscancer och cirka 90 % av andra HPV-associerade cancerformer, såsom huvud-hals cancer (HNSCC). Persisterande infektioner med cancer-associerade HPV typer såsom tex HPV16 är den kritiska riskfaktorn för uppkomst av HPV16-associerad cancer. Förändringar i uttrycket av HPV16-generna, särskilt onkogenerna E6 och E7, de virala replikations- och transkriptionsfaktorerna E1 och E2, och det immunogena strukturproteinet L1, kan bidra till etablering av persisterande HPV16 infektioner, och därmed i slutändan också till cancer.

I denna studie visar vi att det cellulära RNA-bindande proteinet hnRNP D och m6A-modifiering av HPV16-mRNA spelar betydande roller i regleringen av virus genuttryck. Dessutom visar vi att den så kallade långa kontrollregionen (LCR) i HPV16-genomet också påverkar splitsning av E6/E7-mRNA, vilket indikerar att LCR kan innehålla splitsningsreglerande DNA element som kan påverka HPV16-genreglering. Det cellulära, RNA-bindande protein hnRNP D fungerade som en splitsningsinhibitor av HPV16 E1/E2 och E6/E7 mRNA. hnRNP D bidrar till att generera intron-innehållande E1 och E6 mRNA. N- och C-terminala delar av hnRNP D bidrog till HPV16-mRNA-splitsningskontroll på olika sätt. Den N-terminala delen av hnRNP D spelade en betydande roll vid inhibition and HPV16 mRNA splitsning. Samtidigt bidrog även RGG-domänen i hnRNP D proteinets C-terminala del till inhibition av HPV16 mRNA splitsning. Vidare visade vi att hnRNP D interagerade effektivt med HPV16-mRNA och att denna interaktion var starkt beroende av den RNA-bindande domänen RRM1. Ett resultat av interaktionerna mellan hnRNP D och HPV16 mRNA var en ökning av nivåerna av osplitsat HPV16 E1 mRNA i cytoplasman. Vi upptäckte också en direkt bindning av hnRNP D till HPV16 mRNA i en HPV16-driven tonsillcancer cellinje och i HPV16-immortaliserade humana epitelceller (keratinocyter). Dessutom ökade "knock-down" av hnRNP D produktionen av HPV16 E7-onkoproteinet i HPV16-drivna livmoderhalscancer celler. Våra resultat tyder på att hnRNP D spelar en betydande roll i regleringen av HPV-genuttryck och HPV-associerad cancerutveckling. Vi undersökte även m6A-metylering av HPV16 mRNA och fann att m6A modifiering

av HPV16 mRNA påverkade alternativ splitsning av olika HPV16 mRNA. Överuttryck av m6A-demetylas ALKBH5 främjade E6-mRNA-produktion och förändrade splitsning av HPV16 L1-mRNA genom att orsaka "exon skipping". Överuttryck av metyltransferas METTL3 inducerade E1-mRNA-produktion och förändrade splitsning av HPV16 L1 mRNA på ett motsatt än ALKBH5. YTHDC1 som "läser av" m6A-modifierade mRNA, påverkade HPV16 mRNA, fram för allt genom att inducera E6-mRNA-produktion genom intron retention.

Sammanfattnings, fann vi att det cellulära RNA-bindande proteinet hnRNP D spelar en central roll vid splitsning av HPV16 mRNA. Vidare fann vi att m6A-metylering av HPV16 mRNA påverkar splitsning av HPV16 mRNA och att det finns ett oidentifierat RNA element i de E6-kodande delarna som reglerar splitsning av HPV16 mRNA. Vidare fann vi att icke-transkriberade sekvenser i HPV16 LCR påverkar splitsning av HPV16 E6/E7-mRNA. Resultaten som presenteras i avhandlingen ökar vår förståelse av hur det cancerassocierade HPV16 viruset reglerar uttryck av sina gener.

Abbreviations

3'ss, SA	3' splice site, Splice Acceptor
5'ss, SD	5' splice site, Splice Donor
ALKBH5	Alkylated DNA repair protein AlkB homolog 5
ATM	Ataxia Telangiectasia Mutated
ATR	Ataxia Telangiectasia and Rad3-related
BPS	Branch Point Sequence
BPV	Bovine Papillomavirus
CMV	Cytomegalovirus
CLIP	Crosslinking and Immunoprecipitation
CTD	C-Terminal Domain
DBD	DNA Binding Domain
DDR	DNA Damage Response
E2BS	E2-binding site
ER	Endoplasmic Reticulum
ESE	Exonic Splicing Enhancer
ESS	Exonic Splicing Silencer
E6AP	E6-associated protein
FTO	Fat Mass and Obesity-Associated Protein
hnRNP	Heterogeneous Nuclear Ribonucleoprotein
HNSCC	Head and Neck Squamous Cell Carcinoma
HPV	Human Papillomavirus
HR	High Risk
HSPG	Heparin Sulphate Proteoglycans

IRES	Internal Ribosomal Entry site
ISE	Intronic Splicing Enhancer
ISS	Intronic Splicing Silencer
KH	hnRNP K-Homology
LCR	Long Control Region
LR	Low Risk
m6A	Methyl-N6-adenosine
METTL	Methyltransferases-like protein
MHC	Major Histocompatibility Complex
NLS	Nuclear Localization Signal
ORF	Open Reading Frame
pAE	HPV Early Polyadenylation Site
pAL	HPV Late Polyadenylation Site
PCR	Polymerase Chain Reaction
pRB	Retinoblastoma protein
RGG	Arg-Gly-Gly tripeptides repeats
RRM	RNA Recognition Motif
YTH	YT521-B Homology
Sluc	Secreted luciferase gene
snRNP	small nuclear Ribonucleoprotein
SR	Serine Arginine rich protein
UTR	Untranslated Region
VLP	Virus-like particles
WTAP	Wilms Tumor 1 Associated Protein

1 Introduction

1.1 Human Papillomavirus classifications and diseases

Human papillomaviruses (HPVs), which belong to the *Papillomaviridae* family, are epitheliotropic small DNA viruses that are prevalent in the human population. HPVs are characterized by genotype (termed “types”), defined as $> 10\%$ L1 DNA sequence divergence from other known types, and numbered in order of discovery [1]. Until now, more than 200 types of HPVs have been identified and classified into five different genotype groups (α , β , γ , μ , and ν) (**Figure 1.1**) [2]. These viruses may be transmitted via skin-to-skin contact or sexually and infect human cutaneous or mucosal epithelial cells. Most HPV types within the β -genus and γ -genus infect cutaneous epithelia and only cause benign skin warts, usually on the hands or feet [3, 4]. Some key β -HPV types are associated with an increased risk of non-melanoma squamous cell carcinomas [3]. α -genus contains 64 HPVs classified as either “low-risk” (LR) or “high-risk” (HR) according to their clinical characteristic and mainly infect mucosal epithelia. About 40 LR types of HPV are trophic for anogenital epithelium, cause common sexually transmitted infections like genital warts. For example, HPV6 and HPV11 cause condyloma acuminatum. LR HPV may also cause laryngeal papillomatosis, which is a rare benign proliferative disease, that may narrow the airway. Due to the recurrent nature of the virus, repeated treatments usually are needed [5]. Approximately 15 α -HPVs are highly pathogenic, so-called HR HPVs, that may cause anogenital cancer and head and neck squamous cell carcinoma (HNSCC). In fact, nearly all cases of cervical cancer, the fourth most common cancer in women, are the result of high-risk HPV infection [6]. The latest report from World Health Organization (WHO) shows that in 2018, an estimated 570,000 women were diagnosed with cervical cancer and about 311,000 women died from the disease. Of all the cervical cancer cases, HPV16 accounts for approximately 65%, and HPV18 accounts for 20% [7]. HR HPVs can also lead to other anogenital cancers in both women and men. HPV16 is the most prevalent HPV type detected in these cancers [8]. Regarding HNSCC, there is an increasing number of cases. Of all the HPV-positive cases of HNSCC, about 90% are positive for HPV16 [9, 10].

Until now, no medicine that can cure any type of HPV infection. However, the Cervarix 2vHPV vaccine, which is based on L1 Virus-like particles (VLPs), may help protect against two of the most common high-risk HPVs, HPV16 and HPV18.

Additionally, the Gardasil 4vHPV vaccine also protects against two of the most prevalent low-risk HPVs, HPV6 and HPV11, which cause genital warts. A 9vHPV vaccine, Gardasil 9, containing five more VLPs from HPV31, HPV33, HPV45, HPV52, and HPV58 is now available. It protects against an additional 15–20% of all cervical cancer cases [11]. As a result of the total three approved vaccines, the prevention of cervical cancer cases should reach around 90% [12]. Notably, these vaccinations have not been approved for the protection of HPV-associated HNSCC.

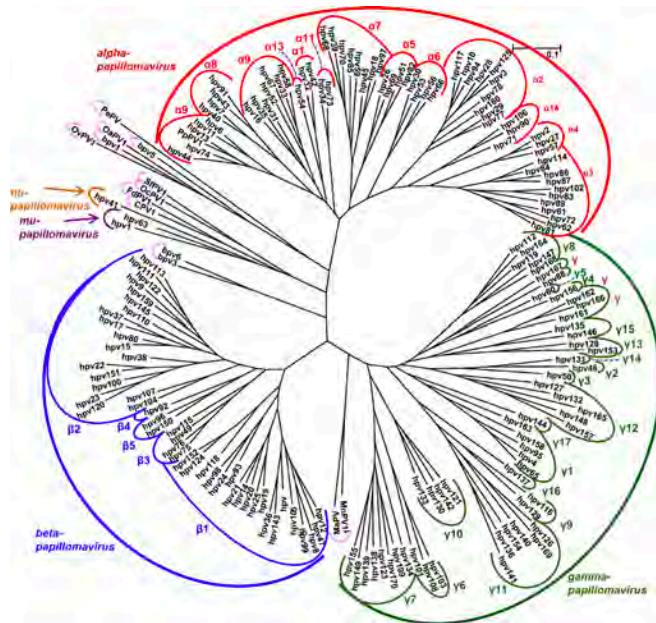


Figure 1.1 Phylogenetical tree based on the L1 ORF sequences of 170 HPV types, as well as single animal papillomaviruses. Reprinted from de Villiers, Cross-roads in the classification of papillomaviruses, in *Virology* 2013; 445:2-10, with permission from Elsevier.

1.2 HPV16 genome and proteins

All HPVs are small non enveloped viruses with an approximately 55 nm capsid in diameter and have a double-stranded circular DNA genome in episomal form (**Figure 1.2 A**) [13]. The approximately 8 kb genome includes three regions: early region, late region, and long control region (LCR) (**Figure 1.2 B, C**). Regarding HPV16, the early region encodes six non-structure proteins, including E6 and E7 oncoproteins and E1, E2, E4, and E5. Two genes encode the structural proteins L1 and L2 that package HPV DNA to form the viral capsid. LCR, also known as upstream regulatory region (URR), contains *cis*-acting elements that can control viral replication and transcription [14]. HPV16 has two promoters named P97 and

P670, which are early and late promoters respectively [15]. The early promoter P97 is located in LCR. It appears to be constitutively active throughout the virus replication cycle. In contrast, activation of later promoter P670, located in the E7 open reading frame, is a marker of the late phase of the viral life cycle. In addition, 3' end formation of virus mRNAs may occur either at the early or late polyadenylation sites. The early polyadenylation site located downstream of the E5 gene. Two late polyadenylation sites are in the late 3' untranslated region just downstream of the L1 open reading frame. They conform well to the consensus polyadenylation signals and HPV16 mRNAs terminating at each site have been detected in infected cervical epithelial cells [16].

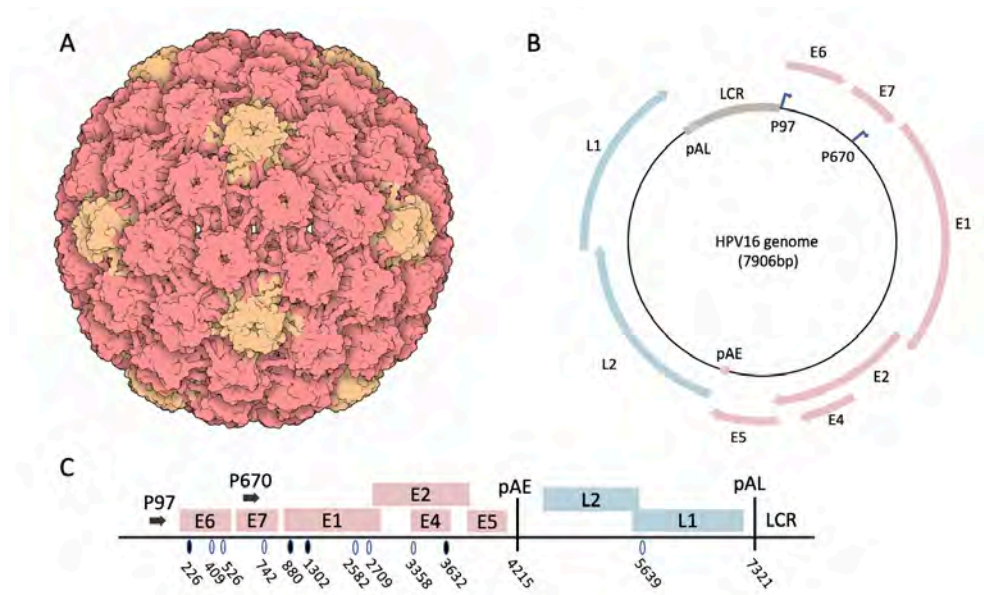


Figure 1.2 Schematic structures of HPV.

A. HPV capsid from cryoelectron microscopy. Capsomeres in orange are surrounded by five other capsomeres, and the ones in red are surrounded by six neighbors. Adapted from <https://pdb101.rcsb.org/>. **B.** Schematic representation of the HPV16 circular genome organisation. **C.** Schematic representation of the HPV16 genome in linearized form. Early genes including oncogenes E6 and E7, also E1, E2, E4 and E5 are indicated in pink. Late genes L1 and L2 are indicated in blue. LCR: long control region. P97: early promoter P97. P670: late promoter P670. pAE: early polyadenylation site. pAL: late polyadenylation site. Known splice sites are indicated in linearized HPV16 genome. Black oval: splice donor. Hollow oval: splice acceptor. Numbers refer to the HPV16 reference strain GeneBank: K02718.1.

1.2.1 Long Control Region

LCR, an 850 bp genomic segment in HPV16, covering about 11% of the whole genome, regulates the replication and gene expression of the virus. It contains numerous regulatory binding motifs for both viral and cellular factors [14, 17, 18]. Nucleotide variability in LCR is relatively high compared to other regions of the

HPV16 genome. However, no substantial evidence shows association between specific genome alterations and the tumorigenic capacity [19]. LCR does not contain open reading frames, and it starts from the stop codon of the L1 gene (5' LCR) and ends at the transcription start of the E6 gene (3' LCR). These two parts are flanking the “central LCR”. HPV16 early promoter P97, SP1 binding sites, and two of four E2 binding sites (E2BS) are located in the 3' LCR. In addition, several other cellular transcription factors, such as AP1 [20] and YY1 [21], which could regulate P97, have been shown to bind to specific LCR positions.

1.2.2 E6 and E7 proteins

The association between HPV16 infection and cervical cancer has been deeply studied for decades. The evidence that HPV E6 and E7 are oncoproteins playing critical roles in oncogenesis is for sure [22, 23]. E6 and E7 are small proteins with estimated 18 KDa and 15 KDa molecular weight, respectively. HPV16 oncoproteins E6 and E7 work together to drive cell cycle, allowing HPV genome amplification in the mid to upper layers of epithelium [24, 25].

In HPV-infected cells, E7 proteins target several cellular proteins including the tumor suppressor retinoblastoma protein (pRB), which has a critical role in G1/S transition of the cell cycle and is the best understood interaction [26, 27]. E7 binds phosphorylated pRB, followed by ubiquitin/proteasome-dependent degradation of pRB. Without the inhibition of pRB, E2F transcription factors are released, resulting in the constitutive activation of E2F dependent transcription, such as the expression of cyclin A and cyclin E [28, 29]. E2F-binding sites often exist in the promoters of genes that are responsible for regulation of cell cycle, differentiation, mitosis, and apoptosis [30, 31]. In addition to targeting pRB, HR E7 could also target other proteins that control the cell cycle, such as CDK inhibitors p21 and p27, with a much higher efficiency than LR E7 [32, 33]. p21 and p27 have been reported to be crucial regulators in epithelial differentiation [34].

As a result of unscheduled cell proliferation activated by E7, tumor suppressor p53 expression levels increased to block cell growth and promote apoptosis [35, 36]. To reduce these effects, HR HPV E6 proteins inactivate the p53 by directing it for proteasomal degradation via the E3 ubiquitin ligase E6-associated protein (E6AP) [37, 38]. Also, E6 could bind to p53 and suppress transcription directly without the association of E6AP [39]. Taken together, E6 actions inhibit p53 anti-proliferative and apoptotic properties in response to DNA damage and cellular stress induced by abnormal S-phase entrance mediated by E7. In LR HPV infection, the E6-E6AP complex was also detected, but this interaction does not lead to p53 degradation [40]. E6 counteracts p53 in several other ways. For example, E6 inhibit p300 and CREB-binding proteins (CBP) which are known for p53 acetylation, thereby causing destabilization of p53 [41, 42]. In addition to targeting p53, HR HPV E6 proteins target cellular PDZ domain containing proteins leading to maintenance of

a suitable cell environment for HPV DNA replication and episome maintenance [43, 44].

1.2.3 E1 and E2 proteins

E1 and E2 proteins are responsible for replication of the HPV DNA genome. The HPV genome is small and E1 is the only one DNA replication enzyme [45, 46]. Apart from E1 and the help of E2, HPV genome replication is totally dependent on the cellular DNA synthesis machinery. E1 consists of three domains: a N-terminal regulatory region, a center DNA binding domain (sometime referred to as the origin-binding domain) and a C-terminal helicase domain, which allows E1 to act as an ATP-dependent DNA helicase that unwinds the double-stranded DNA at the origin of the viral episome to initiate genome replication [45, 47, 48]. Much of the understanding of E1 protein structure and function originates from bovine papillomavirus type 1 (BPV1), and though the E1 helicase basic principle should be well conserved, our understanding of HPV16 E1 function is limited.

The HPV E2 protein which possesses one DNA-binding domain and one protein-binding domain linked by a flexible hinge region is a critical factor in the HPV replication cycle and is multifunctional [49]. E2 contributes to viral DNA replication through binding to E1 [50]. E1 recognizes and binds the origin of HPV genome through its DNA binding domain with low affinity. E2 could increase and direct E1 binding to the origin through its ability that binds to E1 and E2 binding site present in replication origin with high affinity [51]. Further, with the help of E2, E1 assemble as a dimer of hexamers with unwinding activity [52] and recruits the cellular DNA replication machinery. There are four E2BSs located in the HPV LCR. Three of these are located adjacent to the origin of replication and are essential for E1- activated viral replication. In addition, E2 could act as viral transcriptional regulator and directs the expression of E6 and E7 by regulating the early promoter both positively and negatively depending on its expression level. In the late stage of productive HPV infection, E2 represses early promoter P97 as it accumulates, thereby decreasing the expression levels of E6 and E7 [53, 54]. A truncated form of E2 transcribed from an internal ATG and the E8^{E2} fusion protein in several types of HPV could repress E2 function [49]. HPV16 splice site SD1302 is believed to be used to produce E8^{E2} RNA [55].

E2 also facilitates genome partitioning during mitosis [49] and plays a role in the post-transcriptional control of HPV gene expression [56]. A number of HPV transcripts, initiated from early and late promoters encode full-length E2 mRNAs that are with multicistronic and encode several open reading frames (ORFs) have been identified. Recently, it has been demonstrated that the HPV16 mRNA with the strongest potential to produce E2 is a late mRNA initiated at late promoter P670 and spliced from HPV16 5' splice donor SD880 to HPV16 3' splice acceptor SA2709 [57].

1.2.4 E4 proteins

The E4 coding region is contained within the E2 coding region and the E4 protein is generated from an in-frame fusion starting at E1 ATG and the first five amino acids of E1 to the E4 ORF starting at splice site SA3358 (HPV16), resulting in an E1^{E4} fusion protein [58]. Although E4 ORF was classified as an early viral gene, E4 protein expression level is relatively high at the late stage of the viral life cycle, contributing to HPV DNA replication and virus synthesis. The precise mechanisms by which E4 contributes to genome amplification have not been clearly studied, but it is reasonable to speculate from the known E4 function. E4 in proliferating epithelial cells causes G2 arrest by sequestering cellular Cyclin B/Cdk1, which could contribute to HPV genome amplification. Also, E4 binds to cytokeratin filaments and disrupts their structure and reorganizes the cytokeratin filament network, thereby having an essential role in virion egress [59]. With its high expression levels, E4 protein may serve as a biomarker for active virus infection, and in the case of high-risk types also disease severity [60, 61].

1.2.5 E5 proteins

The E5 proteins are short, hydrophobic transmembrane proteins encoded by many but not all HPVs. E5 has transforming activity but poorly defined roles in the productive HPV infection. E5 is believed to have transforming activity by modulating the activity of many cellular proteins [62]. In HPV16, E5 is an 83 amino acid protein likely to be expressed more at a late stage of the viral life cycle, in the upper layers of the epithelium. At the cellular level, HPV16 E5 localizes primarily to the membranes of endoplasmic reticulum (ER) and Golgi apparatus, plasma membrane, and nuclear envelope. HPV16 E5 has recently been shown to be produced primarily from mRNAs spliced from SD226 to SA3358 [63].

HPV16 E5 could interact with and activates receptor tyrosine kinase receptors like EGFR, resulting in increased EGFR levels at the cell surface, thereby sensitizing cells to EGF. HPV E5 proteins have also been reported to play a role in apoptosis and in evasion of the immune response by down-regulating major histocompatibility complex I (MHC-I) [62].

1.2.6 L1 and L2 proteins

HPVs encode two structural or capsid proteins, major L1 and minor L2 that are synthesized late in the infectious cycle [64]. The L1 and L2 capsid proteins could self-assemble into an icosahedral shell that encapsidates viral double-stranded circular DNA genome, resulting in the production of progeny virions in the superficial layer of the epithelium [65].

L1 is the major HPV16 capsid protein that is highly regulated by cell differentiation and produced exclusively in the terminally differentiated cells [66]. Each viral particle is composed of 72 L1 pentamers that stabilize the capsid. L1 proteins are highly immunogenic and recombinant L1 pentamers self-assemble into VLPs that are the source of the existing highly effective prophylactic vaccine for HPV [67, 68].

L2, a largely internal protein, is synthesized prior to L1 and is imported into the nucleus. The L2 amount or the exact ratio of L1/L2 in the virion is not clear in different types of HPV. Studies based on cryoelectron microscopy of native BPV1 suggested one L2 protein occludes the center of each pentavalent capsomere for a total of 12 L2 per virion [69]. Studies based on SDS-PAGE of native HPV1, biochemical analysis of HPV11, and studies in HPV16 pseudovirions suggested that HPV virus particles can contain up to 36 and as much as 72 L2 protein per particle which means each pentamer contains one L2 protein [70-72]. While L1 is sufficient to produce synthetic VLP, L2 has been shown to affect the final structure of the virion, in addition to enhancing infectivity and DNA encapsidation of the particle [73, 74].

1.3 HPV16 life cycle

One unique character of the HPV life cycle is that it relies on infecting the basal layer of epithelium that undergoes mitosis and differentiation processes. The life cycle of HPV16 is tightly linked to differentiation of the stratified squamous epithelium, which is the exclusive cell type it infects (**Figure 1.3**). There is a differential spatial and temporal pattern of HPV gene expression in the infected epithelium. In brief, HPV genome replicates in undifferentiated cells at low levels and to high levels in differentiated cells. HPV16 infects dividing basal epithelial cells through micro-abrasions or by entering cells of the single-layered squamous cellular junction between the endo and ectocervix [13, 75]. During the healing process or normal cell division process the viral episomal dsDNA genome enters the nuclei of the keratinocytes. Once an infected daughter cell begins the process of differentiation, it triggers a tightly orchestrated pattern of viral gene expression required to complete the viral life cycle. Terminal cell differentiation is regained for HPV late gene expression and production of virus particles.

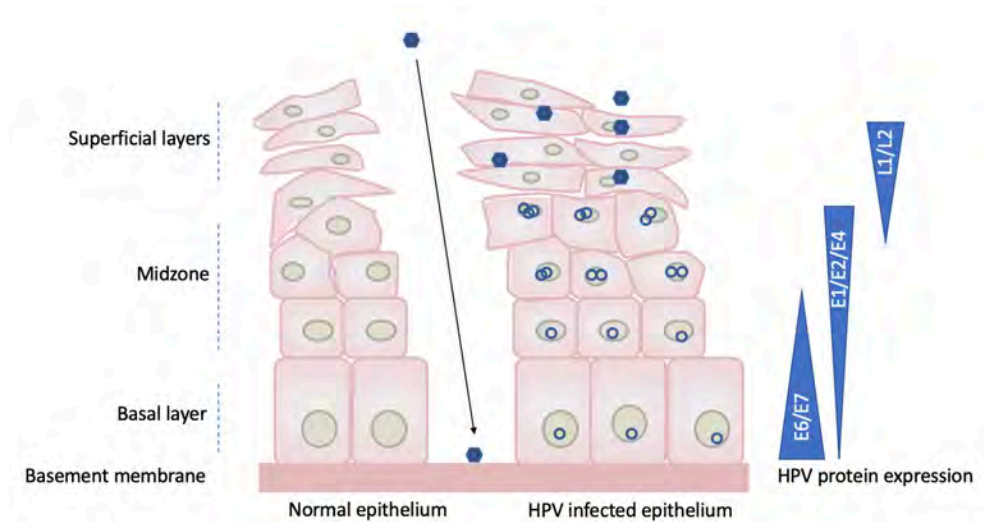


Figure 1.3 Life cycle of HPV16 in cervical epithelium.

HPV16 infect stratified squamous epithelium, usually cervical epithelium, through micro-abrasions. HPV16 virions access to the basement membrane and infect basal layer cells. The viral genome in infected basal cells is maintained as episome form in low copy number. In normal epithelium, as basal cells dividing, the daughter cells migrate from basal layer to superficial layers and undergo cell differentiation. In HPV16 infected epithelium, due to the expression of E6 and E7 oncoprotein, cells in the lower layers of midzone re-entry into the cell cycle instead of undergoing differentiation. This event is necessary for HPV16 genome vegetative amplification and E1, E2 and E4 production in mid to up layer cells. Once the infected cells reach differentiated superficial layers of epithelium, the late promoter is activated resulting in capsid proteins L1 and L2 production and virions packaging, allowing virions to be released.

1.3.1 Virus entry

HPV viral particles interact with the basement membrane when the L1 capsid protein binds to the primary cellular receptor heparin sulfate proteoglycans (HSPG) on the basement membrane and basal layer cells [76, 77]. L1 and L2 capsid proteins undergo conformational changes. L1 conformational change allows the virus to move from the cell surface HSPGs binding to the unidentified secondary or uptake receptor. The N-terminus of L2 is exposed and undergoes further cleavage that is essential for L2 to perform its function in helping the viral genome to escape from the endosome [78]. Several proteins could be secondary receptors like EGFR, integrin, or laminin. Alternatively, more than one secondary receptor works together in one infection [79-81]. This process needs more investigation. Binding to the secondary receptor trigger endocytosis that allows HPV to enter cells [82]. Overall, the secondary receptor with the virus bound to it ends up in endosomes, in which the pH gradually drops as the endosome matures. This pH drop will cause the virus to be cleaved off from the secondary receptor, and capsid starts to break down. After these events, HPV travels through membrane-bound cytoplasmic components and the trans-Golgi network and even endoplasmic reticulum network [83, 84]. Finally, the viral genome complexes that end up in the endoplasmic reticulum enter the nucleus during basal cell division.

1.3.2 Viral replication

Upon HPV entry into the dividing cells of the basal layer of the squamous epithelium, HPV early transcription is activated. Replication of HPV DNA relies on cell DNA-replication machinery, and HPV establishes a low number of episomal genomes [85], which express low levels of viral proteins escape the host immune system [86]. On this stage, the viral transcription factor E2 that plays an important role. E2 can bind E1, the DNA helicase, as a dimer of hexamers to the viral origin of replication and recruits the cellular DNA replication machinery [87]. At the same time, E2 works as a repressor of early HPV promoter P97 by inhibiting other transcription activators, thereby limiting E6 and E7 expression [88]. HPV DNA is maintained in daughter cells upon segregation by attaching virus genomes to host chromosomes via E2 [89]. Infection may persist in the infected stem cell and its daughter cells in the basal layer of the epithelium over a long period, up to several years. Since HPV hijacks the cellular DNA synthesis machinery to replicate its own genome in actively dividing cells, virus needs to find a strategy to delay basal epithelium differentiation and keep cells in proliferation. E6 and E7 are the most well understood viral oncoproteins that contribute to the infected basal cell disorder in the epithelium differentiation.

E7 activates the cell cycle of infected, differentiating cells by binding and releasing, and finally degrading pRb to liberate the cellular E2F protein and keep cells in G1 to S-phase to synthesize S phase genes [25]. So, the S phase genes will all be activated, and therefore, the virus will be able to replicate its genome at a fantastic great rate. However, cells respond to any unscheduled cell proliferation event by inducing apoptosis. Therefore, HPV E7 activity might be expected to induce apoptosis. To avoid this, HR-HPVs express the E6 protein, which binds p53, a key cellular regulator of apoptosis, and targets it for degradation [90]. The well-organized E6 and E7 oncoproteins target p53 and pRb, respectively, and play essential and complicated roles together to maintain cell proliferation and inhibit apoptosis and epithelial differentiation [91-93].

The combined work of E6 and E7 gives a suitable environment for HPV to replicate its genome which could amplify to yield several thousands of copies. During this phase, not only E1 and E2 but also E4 and E5 expression increases. E4 is the most abundant protein in the HPV late replication phase. It has multiple functions like regulating cell cycle in G2 arrest [94] and cellular kinases activation [95], but the most important one is to facilitate viral particle release [96]. E5 protein function is not as clear as other HPV proteins. Not so much research is available. One primary function of E5 is to repress the MHC presentation of virus peptides which results in immune escape [97]. The interaction of E5 and EGF-receptor and PDGF receptor results in activation of several cell signal pathways, increasing host cell proliferation [62]. The late stage of the viral life cycle is marked by activation of the viral major late promoter during epithelial differentiation [98]. After efficient viral episomal

DNA amplification, cell differentiation is reactivated, resulting in synthesis of HPV L1 and L2 capsid protein [99]. L1 proteins assembled into pentameric capsomeres in the cytoplasm, then transferred into the nucleus for further virion packaging [100-102]. L2 molecules tethered to the viral genome finally assemble into 72 pentamers of L1 capsids with one L2 molecule incorporated in the centre of each pentamer [103]. In the end, the virus life cycle is completed in the uppermost epithelial layers, where newly free virions are released.

HPV infections also trigger the cellular DNA damage response (DDR) and utilize the DDR to replicate HPV genome by activating Ataxia telangiectasia mutated (ATM) and Ataxia telangiectasia and Rad3-related (ATR), two major DDR regulators. In brief, ATM is required for double strand DNA break repair [104] and ATR is responsible for single strand DNA damage repair [105]. It has been shown that HPV early proteins are sufficient to induce both ATM and ATR activity in differentiating cells [106]. Elicitation of the DDR create the cellular environment which support for viral genome replication outside of the cell cycle. Except for facilitating HPV DNA replication, a recent study demonstrated that activating DDR can recruit RNA-binding proteins to HPV16, altering its RNA splicing and polyadenylation in promoting late gene expression [107].

From the HPV life cycle, we could see that HPV is strictly dependent on and closely interacts with host epithelium cell proliferation and differentiation, both in time and space. Usually, the HPV infection is cleared by the host immune system within one year. Cancer formation does not result from this type of infection [108]. However, the longer the virus-infected cells persist, the more chances of HPV genome integration. Transient infections that become persistent, keep expressing viral proteins, except for L1 and L2. Therefore, virions are not produced [99]. HPV DNA integration in cellular chromosomes is a bad consequence for both virus and host. Since integration breaks virus genome, it results in failure to complete the HPV life cycle and virion particles can no longer be produced. In addition, HPV genome is preferentially integrated in E2 gene, though other regions are broadly affected as well [44, 109]. The disrupted E2 ORF cannot produce E2, which is a transcriptional suppressor of E6/E7 oncogenes [110]. Despite HPV integration, enhanced expression and stabilization of viral oncogene E6 and E7 transcripts is observed [111]. Regarding host genome, the integration causes rearrangements, deletions, and/or translocations. The establishment of persistent infections contributes a lot to the progression of cancer.

1.4 HPV16 gene expression and regulation

The HPV16 life cycle in infected keratinocytes tightly depends on host cell differentiation, as described above. HPV viral proteins are produced from

polycistronic mRNAs that undergo extensive alternative RNA splicing (**Figure 1.4**), which allowing HPV to encode multiple proteins in a very compact genome and take advantage of alternative splicing to differentially express HPV proteins in a cell differentiation-dependent and temporal manner [112, 113]. HPV16 contains two promoters, P97 and P670, that initiate early and late transcription, respectively. Early promoter P97 is responsible for the expression of early proteins at early stage of viral life cycle. The late stage of the life cycle is marked by activation of the viral major late promoter P670. The completed viral life cycle is dependent on the precise activation and switching of promoters [15, 114] which is regulated by both cellular proteins and viral E2 protein [115, 116]. The regulation of HPV16 gene expression is mainly at transcriptional and post-transcriptional levels including splicing and polyadenylation due to the limited number of promoters [14, 112, 117].

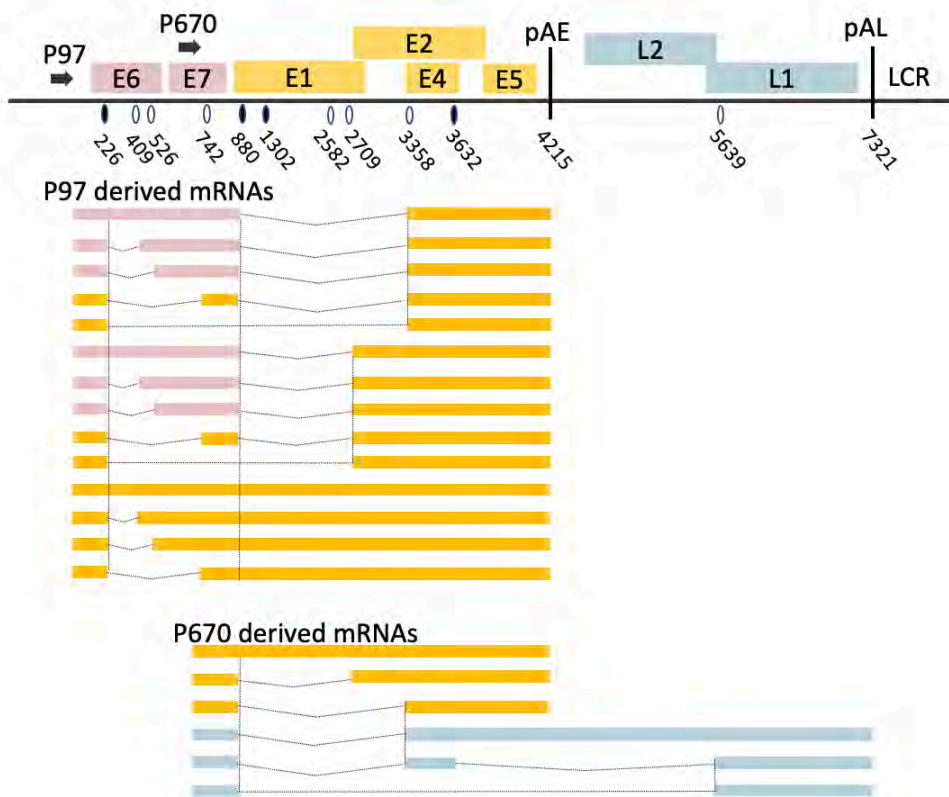


Figure 1.4 Schematic map of linearized HPV16 genome and representative mRNAs.

The linearized HPV16 genome is indicated on the top. A subset of HPV16 mRNAs generated by alternative splicing is indicated: HPV16 mRNAs derived from P97 and polyadenylated at pAE are listed in middle part. HPV16 mRNAs derived from P670 and polyadenylated at either pAE or pAL are listed at bottom. Early oncogenes E6 and E7 are indicated in pink; Early genes E1, E2, E4 and E5 are indicated in yellow; Late genes L1 and L2 are in blue. P97: early promoter P97. P670: late promoter P670. pAE: early polyadenylation site. pAL: late polyadenylation site. Known splice sites are indicated. Black oval: splice donor. Hollow oval: splice acceptor.

1.4.1 mRNA Splicing

In a pre-mRNA, two neighboring exons (5' exon and 3' exon) are separated by an intron, which contains three conserved sequence elements: the 5' splice site (5'ss), also called 5' splice donor (5' SD), the branch point sequence (BPS), and the 3' splice site (3'ss), also called 3' splice acceptor (3' SA) [118]. Splicing is a two-step process. The 5'ss is involved in both steps of splicing. In the first step, the 2'-hydroxyl group (OH) of the branchpoint adenosine attacks the phosphodiester bond (p) at the 5'ss and displaces the 5' exon; in the second step, the 3'-hydroxyl group of the 5' exon attacks the phosphodiester bond at the 3'ss and displaces the lariat intron (**Figure 1.5 A,B**) [119].

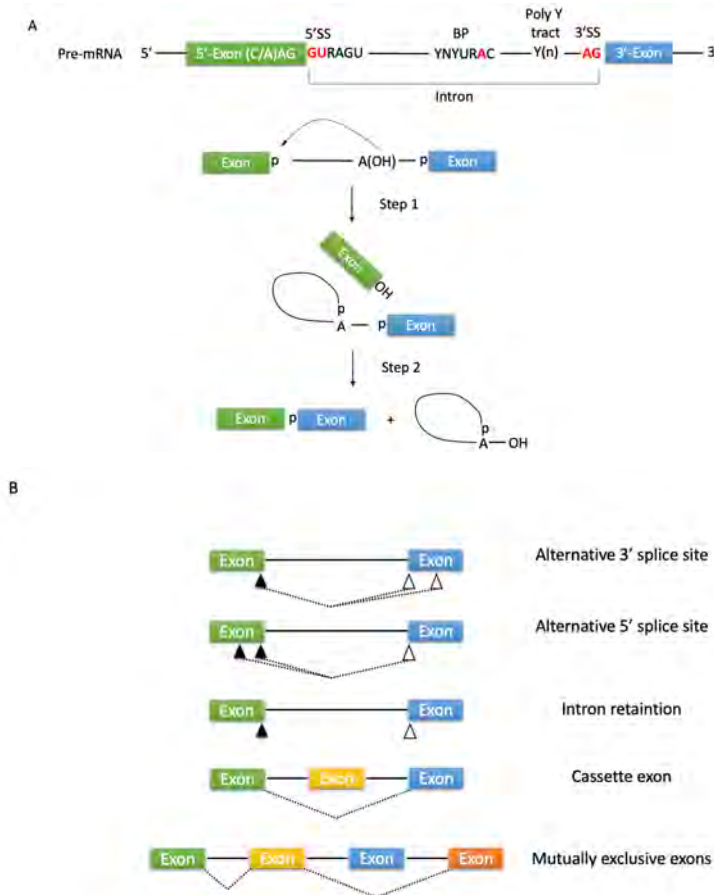


Figure 1.5 Schematic representation of pre-mRNA splicing and alternative splicing.

A. pre-mRNA structure and the mechanism of two-step reaction of pre-mRNA splicing. Boxes represent the exons. Solid lines or loops represent introns. Consensus sequences at 5'ss, BP and 3'ss are indicated with nucleotides in red letter. Y represents pyrimidine, R represents purine and N represents any nucleotide. Y(n) represents polypyrimidine tract. **B.** Splicing products that generated by alternative splicing in various ways. Dot lines represent alternative splicing mechanisms. Splice donors are indicated as black triangle, splice acceptors are indicated as hollow triangle.

Pre-mRNA splicing is executed by spliceosome [120] which consists of five small nuclear ribonucleoproteins (snRNPs) (U1, U2, U4, U5 and U6) [121] and a substantial number of proteins [122-124]. During the two-step splicing reaction, the conformation and composition of the spliceosome complex are highly dynamic. Each snRNP and splicing factor follows strict assembly and rearrangement for the specific within the spliceosome (**Figure 1.6**). Briefly, U1 snRNP recognizes and binds to pre-mRNA 5'ss through RNA-RNA interaction, U2 auxiliary factor (U2AF) with subunits of U2AF65 and U2AF35 binds to 3'ss adjacent poly pyrimidine tract (Yn) [125, 126], forming the early complex (E complex). Next, U2 snRNP binds to BPS to form a pre-spliceosome A complex. Subsequently, pre-assembled U4.U6.U5 tri-snRNPs binds to A complex, generating B complex. Later, B complex undergoes rearrangements in RNA and protein interactions and disassembles with U1 and U4 snRNPs, leading to generating B* complex with catalytic activity for step 1 reaction followed by yield C complex with catalytic activity for step 2 [127]. The spliceosome then dissociates and releases mature mRNA.

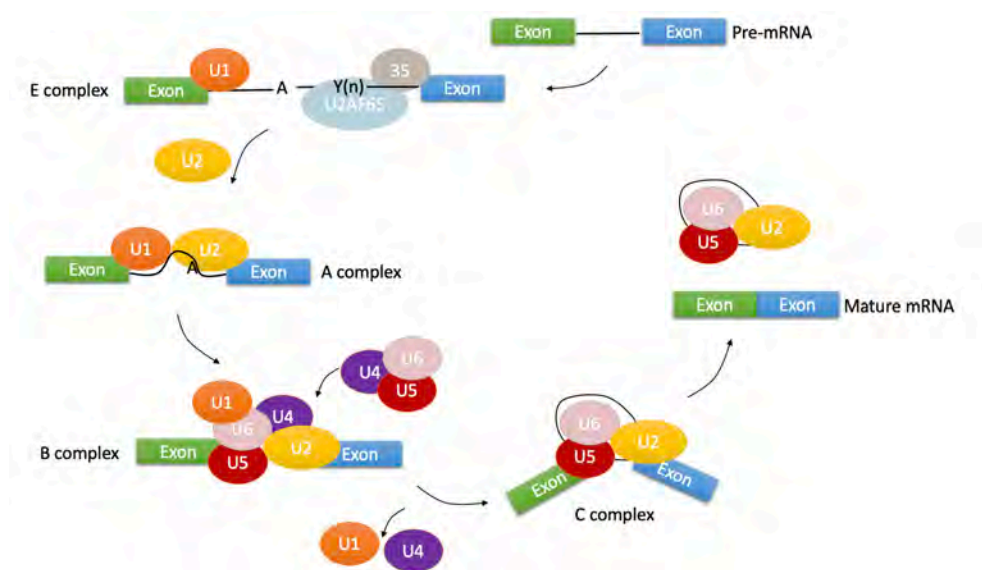


Figure1.6 Schematic representation of pre-mRNA splicing major steps and major protein composition of spliceosome.

Pre-mRNA splicing is initiated from 5'SS being recognized by U1 snRNP and 3' end of intron interacting with U2AF by Y(n) , forming the E complex. Next, U2 snRNP is recruited to branch point, forming the A complex. In a subsequent step, U4/U6.U5, a pre-assembled snRNPs, interacts with A complex, generating B complex. Then, U1 and U4 snRNPs are released, leading to C complex formation. After the two catalytic steps, the spliceosome is disassembled, leading to the mature mRNA being released.

1.4.2 Alternative mRNA splicing in HPV16

During HPV16 infection cycle, more than twenty different mRNAs are expressed, most of which are alternative splicing products. There is extensive splicing in the E6E7 region of the pre-mRNAs since there are three splicing acceptors that can be selected and generate at least four alternative spliced mRNA isoforms (**Figure 1.7**). E6 full-length mRNA is unspliced, intron-retained transcript that includes the E6 and E7 open reading frames believed to be translated into E6 protein. E6*I/E7 (226^409) mRNA, which is responsible for E7 protein expression [128-131], E6*II/E7 (226^526), and E6^E7 (226^742) are mRNAs alternatively spliced from one 5' splice site SD226 to one of alternative 3' splice sites SA526, and SA742, respectively, in the primary transcript. All four isoforms are produced in a mutually exclusive way to affect E6 and E7 protein expression levels, but less is known about the coding potential of the last two isoforms [132, 133]. In LR HPVs, the E7 transcript is generated from a unique promoter, not by AS like HPV16 and other HR HPVs [134].

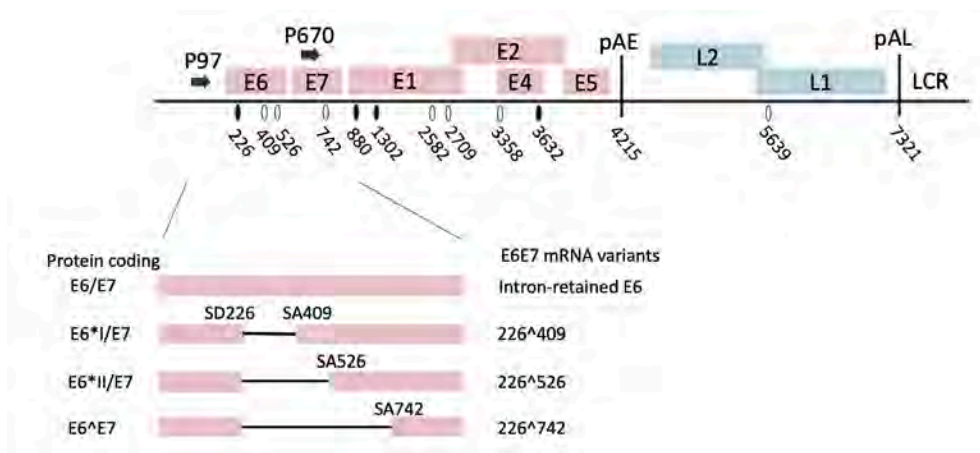


Figure 1.7 Splice sites used for HPV16 alternative splicing in E6E7 coding region.

Schematic representation of linearized HPV16 genome is shown on the top. Four isoforms of alternative spliced mRNAs generated from this region are listed. Splice sites (SD226, SA409, SA526 and SA742) used by spliced mRNA are indicated in each transcript and on the right side as well (226^409, 226^526, 226^742). Unspliced mRNA indicated as intron-retained E6. Protein potentially produced from each mRNA is listed on the left.

Splice factor hnRNPA1 could bind to *cis*-element located in the E7 coding region that acts as an ESS, leading to inhibition of E6E7 mRNA splicing and increased intron retained E6 mRNA production at the expense of the spliced HPV mRNAs [57, 135]. In contrast, hnRNPA2, a close relative of hnRNP A1 bound to the same ESS, promoted production of the alternatively spliced 226^742 mRNA [129]. Also, hnRNP D40 promoted intron-retained E6 mRNA production at the expense of all the other spliced isoforms in the E6E7 region. The interaction of hnRNP D40 and

this particular ESS may play an essential role in the splicing inhibitory function [135]. Taken together, hnRNPs contribute to E6/E7 splicing regulation, leading to balancing the production of all E6/E7 mRNAs.

There are more complicated spliced isoforms produced from the E1/E2 region of the genome since three 5' splice donors (SD880, SD1302 and SD3632) and three 3' acceptors (SA2582, SA2709 and SA3358) are located in this region (**Figure 1.8**) [136]. SD880 is much more active than SD1302. It is utilized in the production of E2 mRNA (880[^]2709) and E4 mRNA (880[^]3358) and the two acceptors compete for SD880. E1 mRNAs are generated by retention of the E1 coding intron located downstream of SD880. E1/E2 mRNA production has a similar mutually exclusive way as E6/E7 mRNA alternative splicing. SA2709 is better known than SA2582 since mRNAs spliced to SA2709 are more efficiently translated than mRNAs spliced from SD880 to SA2582 [57]. SD3632 is specific for L1 mRNA production, which will not be described in this part. The regulation of E1/E2 splicing is barely known, but it has been shown that SD880 is regulated in an Akt kinase-dependent way and hnRNP D40 inhibits E2 mRNA production and increases E1 mRNA, but a more detailed mechanism is needed for further understanding [135, 137].

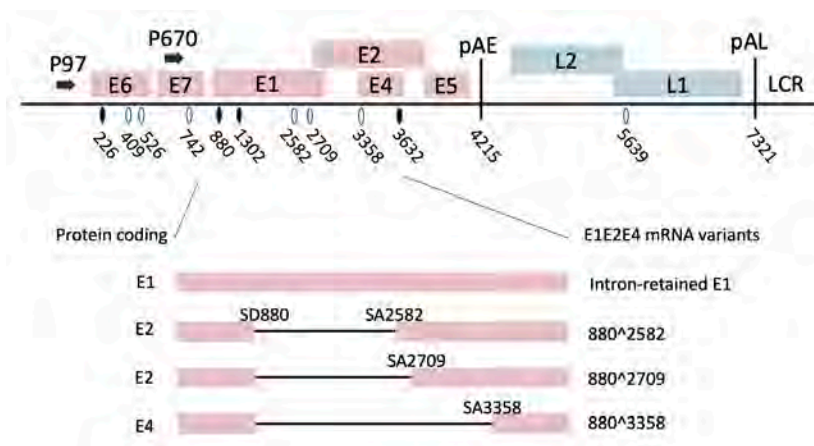


Figure 1.8 Splice sites used for HPV16 alternative splicing in E1E2 coding region

Schematic representation of linearized HPV16 genome is shown on the top. Four isoforms of alternative spliced mRNAs generated from this region are listed. 5' splice sites (SD880, SA2582, SA2709 and SA3358) used by spliced mRNA are indicated in each transcript and on the right side as well (880[^]2582, 880[^]2709, 880[^]3358). Unspliced mRNA indicated as intron-retained E1. Protein potentially produced from each mRNA is listed on the left.

1.4.3 Polyadenylation

Pre-mRNAs undergo maturation before being exported from the nucleus to the cytoplasm. The maturation of mRNA at 3' ends is a two-step process that involves

endonucleolytic cleavage of the pre-mRNA followed by synthesis of a poly (A) tail on the 3' end of the product by poly (A) polymerase (PAP) [138]. These two coupled reactions, collectively called polyadenylation, are intimately linked to transcription termination and highly regulated. The key protein factors and consensus RNA elements responsible for polyadenylation are well understood. The polyadenylation site is defined by an upstream conserved hexamers AAUAAA element and a downstream less conserved GU- and U-rich region located close to the 3' ends of the nascent transcripts in eukaryotic cells [139]. Firstly, cellular protein cleavage and polyadenylation specificity factor (CPSF) which is the core protein involved in 3' end processing recognizes the AAUAAA signal and mediates the endonucleolytic cleavage. Meanwhile, cleavage stimulation factor (CstF) that recognizes GU- and U-rich region, cleavage factor I (CFI) and CFII are also responsible for cleavage event, which help to determine the site of cleavage [140, 141]. Once the 3' end is cleaved, polyadenylation step that is catalyzed by PAP starts [142, 143]. Poly (A) binding protein 1 (PABPN1) binds to the growing poly (A) tail, disrupting the interaction between CPSF and the poly (A) polymerase when the tail is around 250 nucleotides and thus finishes the polyadenylation process [144, 145].

1.4.4 Polyadenylation in HPV16

The two polyadenylation sites, HPV early polyadenylation site (pAE) and HPV late polyadenylation site (pAL) are essential for HPV16 mRNA maturation. HPV16 transcripts in undifferentiated host cells stop at pAE, resulting in HPV16 early mRNAs maturation. With the host cell differentiation, the activity of pAE is decreased, leading to the transcription reading through and stopping at pAL allowing viral late protein L1 and L2 expression [146]. Switching of the pAE to pAL is another important part of HPV16 gene regulation, though with unclear mechanism. GGG-motifs located downstream of pAE interaction with hnRNP H and increased the activity of pAE [147, 148]. Early 3' untranslated region (3' UTR) is also involved in the regulation of pAE but in a different manner by interacting with multiple cellular proteins. The interaction of hnRNP C and RALYL [149] with U-rich elements upstream of pAE inhibits polyadenylation activity while Fip-1 activates it [150]. Multiple polyadenylation signals were present at pAL [151]. Late 3' UTR contains *cis*-elements that interact with multiple cellular proteins and inhibit HPV gene expression by destabilizing late mRNAs or by affecting translation [152-154].

1.4.5 m6A RNA methylation

Reversible methylation is the most prevalent internal mRNA modification. N6-methyladenosine (m6A) plays a great role in post-transcriptional regulation of gene expression(**Figure 1.9**) [155]. m6A is widely conserved among eukaryotic cells and

viruses with a nuclear phase. The conserved motif of m6A is DRACH which is [G/A/U][G>A]m6AC[U>A>C]. It is estimated that each mRNA contains 3-5 m6A sites on average and that it is enriched around stop codons, 3' UTR, and in internal long exons [156]. This dynamic and reversible RNA modification is controlled by different groups of proteins. m6A mRNA methylation is catalyzed by a core methyltransferase complex, so-called “writers”, which contains methyltransferases-like (METTL) protein METTL3, METTL14, and wilms tumor 1 associated protein (WTAP). Fat mass and obesity-associated protein (FTO) and alkB homolog 5 (ALKBH5) are efficient demethylation proteins that could remove m6A and are called “erasers”. These two groups of proteins make great contributions to dynamic and reversible m6A mRNA modulation [157]. For m6A sites to have a specific biological function, they are recognized by proteins reading m6A called “readers”. The most prominent group is YTH21-B homology (YTH) domain-containing family. In the nucleus, YTHDC1 regulates splicing by binding to pre-mRNAs, while in the cytoplasm, YTHDF2 causes mRNA decay and YTHDF1 or 3 and YTHDC2-control translation [156]. Furthermore, accumulating evidence has confirmed that some members of hnRNP family also have reading capabilities. The binding of hnRNPG and hnRNPC to m6A sites is dependent on m6A-mediated structural alteration of mRNAs [158, 159].

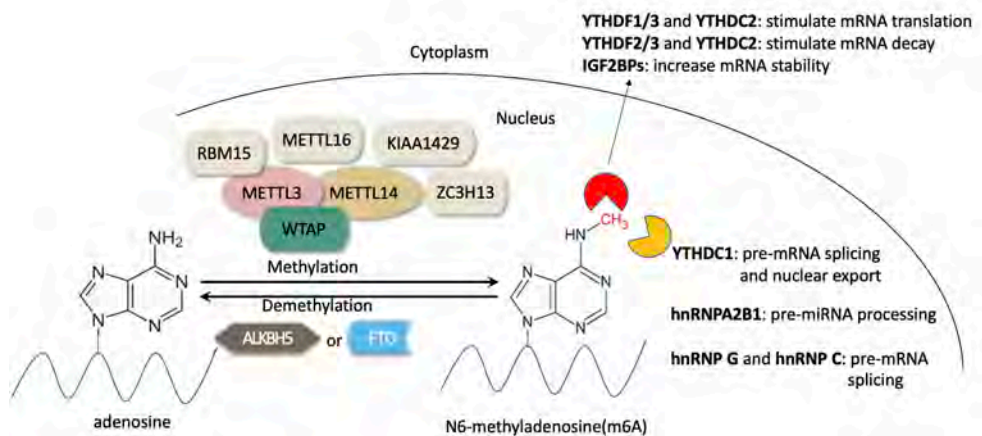


Figure 1.9 An overview of the m6A modification.

The reversible and dynamic m6A modification on mRNA influences mRNA metabolism in many steps including splicing, nuclear export, decay and translation. At least three groups of proteins are involved in the m6A modification. Methyltransferases proteins, also called writers, install m6A on mRNA. METTL3, METTL14 and WTAP core complex is the most well-known m6A writers. Other factors such as METTL16, RBM15, KIAA1429 and ZC3H13 are also reported as m6A writers, recently. ALKBH5 and FTO are identified as demethylases, also called erasers, which could remove m6A methylation. The protein group that impacts the fate of m6A-containing mRNA is m6A readers. YTH family members, including YTHDC1/2, YTHDF1/2/3 could recognize m6A and regulate mRNA metabolism in different ways. In addition, hnRNPs (hnRNPC, hnRNPA2B1 and hnRNPG) and IGF2BP are also report as m6A readers.

Viral RNAs, such as Rous sarcoma virus [160] and simian virus 40 [161] are known to contain m6A modifications for quite a long time, but the precise m6A landscape and the function in particular of post-transcriptional regulation of this modification needs to be further explored. Influenza virus was found that gene expression and replication was enhanced by cis-action of m6A-modified residues [162, 163]. Research on HIV showed that m6A modification is important for HIV-1 replication and its interaction with the host immune system. m6A in HIV genome could regulate viral infection and Gag protein expression [164, 165]. Moreover, m6A cluster in HIV mRNAs was enriched in splicing regulatory sequences [166]. And recently, it has been found that m6A modifications on adenovirus mRNAs are necessary for efficient splicing [167].

1.5 hnRNP family

The hnRNPs constitute an RNA-binding protein family that comprises 20 major RNA binding proteins (**Figure 1.10**). Their main biological function is to regulate nucleic acid metabolism, including alternative splicing. Various hnRNPs interact with different RNA motifs. The hnRNPs share “core” structures that include RNA recognition motifs (RRM) or quasi-RRM domains. In addition to this, other motifs like a K Homology (KH) domain or an arginine/glycine-rich RGG domain is usually included in hnRNPs. Other auxiliary domains like as glycine- and proline-rich domains may also be present in some hnRNPs. Regarding RGG, its main function is to interact with other hnRNPs [168].

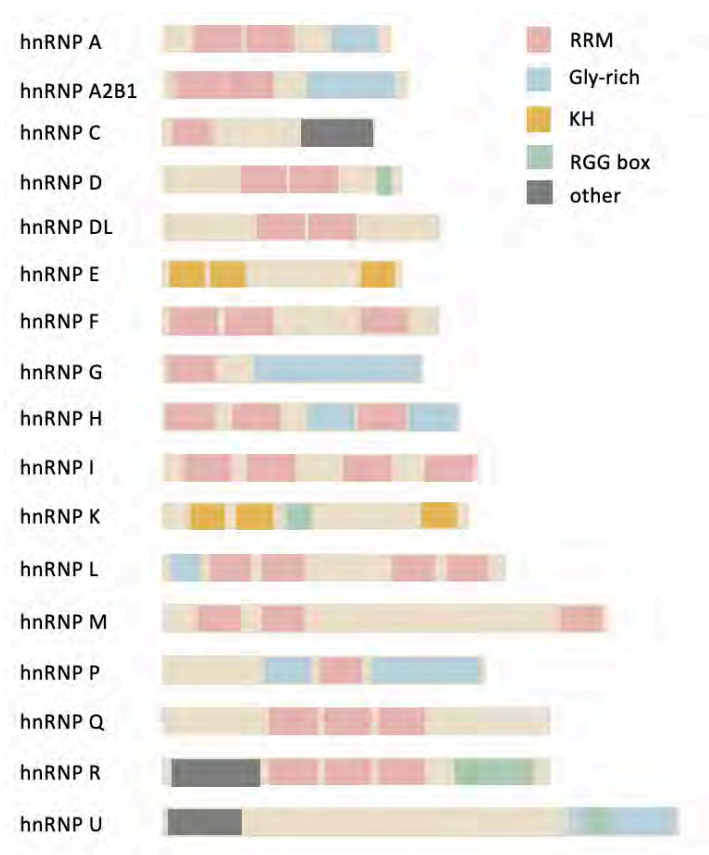


Figure 1.10 Schematic representation of the structure of hnRNP family.

hnRNP family members, named alphabetically from A to U, share several core structure domains including RRM, KH, Gly-rich domain and RGG box. In addition, other structures like qRRM, Gly-rich domain, Ala-rich domain, NLS, M9, Zinc finger domain also present in different hnRNPs. The combination of different structures and motifs increases the functional diversity of hnRNPs. The RNA binding capacity of hnRNPs is mediated by RNA binding domains like RRM, qRRM, RGG and KH domain. RRM: RNA recognition motif. RGG: Arg-Gly-Gly rich domain. KH: hnRNPK-homology domain. NLS: nuclear localization signal.

The canonical function of hnRNPs in alternative splicing is inhibition of splicing reactions through their binding to cognate sites, also known as exonic splicing silencers (ESSs) and intronic splicing silencers (ISSs), repressing either the assembly of the spliceosome complex on the 5' splice site (5'SS) and 3' splice site (3'SS) or the recruitment of SR proteins on exonic splicing enhancers (ESEs) and intronic splicing enhancers (ISEs) following multimerization along exons or by looping out entire exons (Figure 1.11) [169]. On the other hand, some hnRNPs have been shown to promote splicing through their interaction with ISE sequences containing GGG motifs.

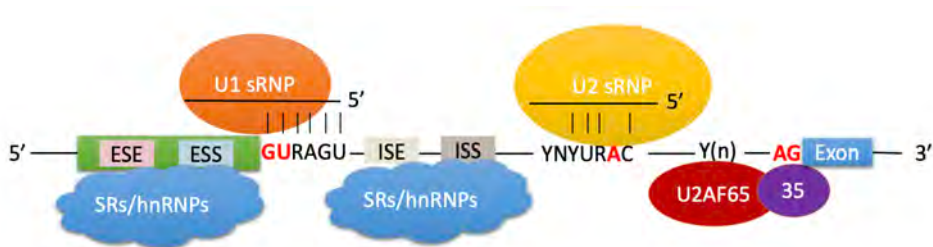


Figure 1.11 Schematic representation of the regulation of splicing by hnRNPs and SRs.

Pre-mRNA splicing is catalyzed by spliceosome, which is a process that initiates from exon and intron definition. U1 snRNPs recognize 5' splice sites and U2 snRNPs with the help of U2AF65/35 recognize 3' splice sites, respectively. Additional interactions of splicing enhancers and/or silencers cis-acting elements in exons and/or introns with trans-acting factors SRs/hnRNPs could enhance the efficiency of exon and intron definition. The splicing regulation depends on the composition of cis-elements and the cognate trans-activators. ESS: exonic splicing silencer. ESE: exonic splicing enhancer. ISE: intronic splicing enhancer.

1.5.1 hnRNP D

One hnRNP that is of particular interest for this thesis is hnRNP D. hnRNP D, also known as AUF1, was originally identified as binding specifically to single-stranded DNA d(TTAGGG)_n of the human telomeric DNA repeats [170]. More research on hnRNP D revealed that it is involved in multiple biological processes. For instance, it could act as a transcription factor and as a protein that regulates mRNA decay by binding AU-rich RNA domains [171]. hnRNP D was also reported to have functions on RNA stabilization and translation through the interaction most frequently with 3' UTR of target mRNAs [172]. hnRNP D has four isoforms, including hnRNP D37, hnRNP D40, hnRNP D42, and hnRNP D45, generated by alternative splicing of the pre-mRNA made from the hnRNP D gene (**Figure 1.12**). The differences between the four mRNA isoforms are exon 2 and exon 7 that encode 19 and 49 amino acid inserts near the N- and C-terminus, respectively. hnRNP D37 has neither exon 2 nor exon 7, hnRNP D40 contains exon 2 but not exon 7, hnRNP D42 contains exon 7 and hnRNP D45 contains both [173]. All four isoforms have a similar structure, including two tandemly arranged, non-identical RRM domains, one alanine-rich domain at N-terminus and a glutamine-rich domain followed by an RGG box at C-terminus after the RRM2 [174]. At the very end of the C-terminus, hnRNP D has a proline-tyrosine nuclear localization signal (PY-NLS) [175]. Unlike hnRNP A1, its most closely related member of the hnRNPs family, hnRNP D subcellular distribution is not so clear. For most cell types, the two larger isoforms hnRNP D42 and hnRNP D45 appear to be largely nuclear, while the smaller two hnRNP D37 and hnRNP D40 reside both in nuclear and cytoplasm [175].

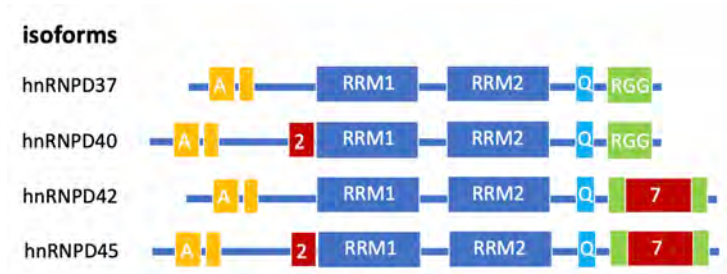


Figure 1.12 Schematic representation of hnRNP D isoforms.

The four hnRNP D isoforms are derived by alternative splicing. The exclusion or inclusion of exon 2 and/or exon 7 is responsible for the differences between isoforms. RGG box in hnRNP D42 and D45 is interrupted by exon 7. A: alanine-rich motif. Q: glutamine-rich motif.

1.6 HPV16 plasmids of special interest

1.6.1 Subgenomic reporter plasmid pC97ELsLuc

The subgenomic reporter plasmid pC97ELsLuc [176] contains all known genes as well as splice sites and polyadenylation sites of HPV16 genome, therefore can produce all alternatively spliced HPV16 mRNAs. Early promoter P97 in pC97ELsLuc was replaced by CMV promoter. Secreted luciferase gene (Sluc) is inserted in L1 gene following the internal ribosomal entry site (IRES) (**Figure 1.13**). The pC97ELsLuc plasmid plays a vital role in researching the regulation of HPV16 mRNA processing in the articles listed in this thesis. A subset of alternatively spliced HPV16 mRNAs produced by pC97ELsLuc is shown in Figure 1.13. Splicing at SD226 generates four splice variants in the E6E7 region by the selection of splice acceptors (SA409, SA526 and SA742). Each E6E7 mRNA variant can be alternatively spliced from downstream SD880 to generate E4 mRNAs (SA3358), E2 mRNAs (SA2709), or intron-retained E1 mRNAs by unspliced at SD880.

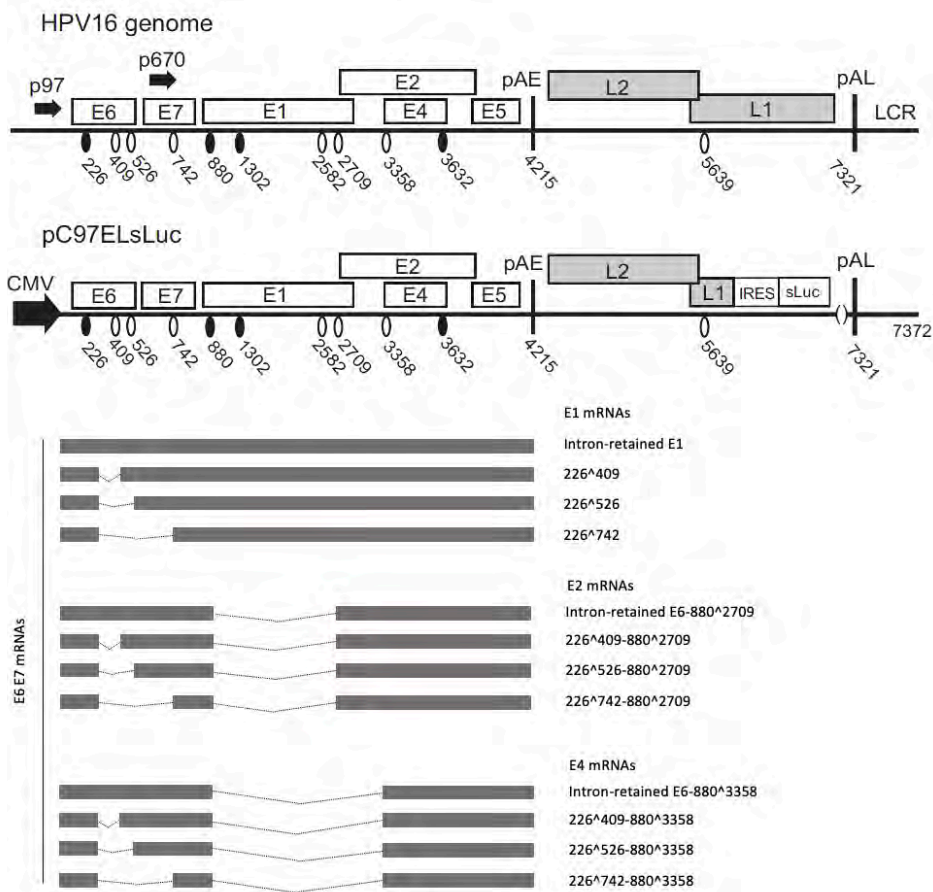


Figure 1.13 Schematic representation of the HPV16 genome, subgenomic plasmid pC97ELsLuc and representative mRNAs.

1.6.2 Subgenomic reporter plasmid pBELsLuc

The subgenomic reporter plasmid pBELsLuc [177] encodes HPV16 genes except for oncogene E6 and E7. Late promoter P670 in pBELsLuc was replaced by CMV promoter. Also, the sluc gene was integrated in L1 following the IRES in pBELsLuc (Figure 1.14). The absence of E6 and E7 coding region allowed pBELsLuc to primarily produce mRNAs spliced from SD880, such as E2 mRNAs (880^2582, 880^2709) and E4 mRNAs (880^3358) or intron-retained E1 mRNAs. The late genes L1 and L2 expression can be induced in pBELsLuc. A subset of alternatively spliced HPV16 mRNAs produced by pBELsLuc is listed in Figure 1.14.

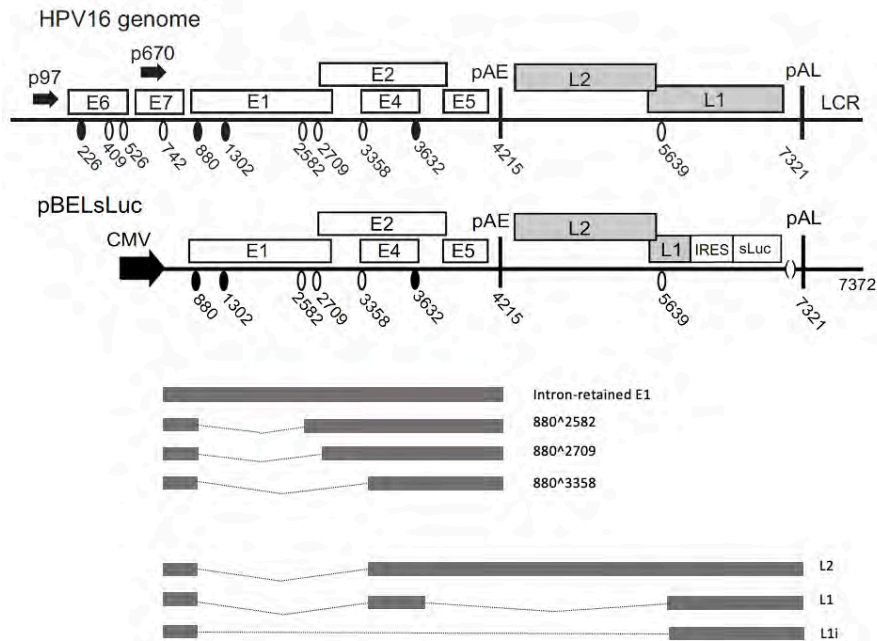


Figure 1.14 Schematic representation of the HPV16 genome, subgenomic plasmid pBELsLuc and representative mRNAs.

1.6.3 Genomic plasmid pHPV16AN

The genomic plasmid pHPV16AN [177] encodes full-length HPV16 genome and can generate episomal HPV16 by Cre-loxP transfection system. The long control region of HPV16 is flanked by loxP-sites. Co-transfection of pHPV16AN with plasmid pCRE results in excision and circulation of the viral DNA at the loxP sites to generate episomal HPV16 genome (Figure 1.15).

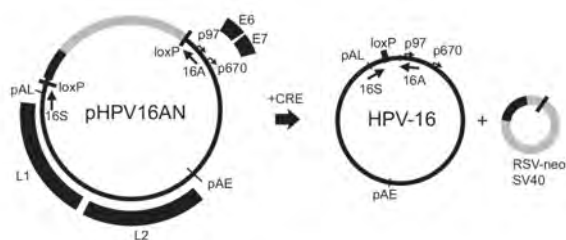


Figure 1.15 Schematic structure of the HPV16 genomic plasmid pHPV16AN and Cre-loxP transfection system.

2 Aim of the Thesis

General aim of all projects is to investigate human papillomavirus type 16 gene regulation and mRNA processing.

Specific aims for each project:

Project I: hnRNP D activates production of human papillomavirus type 16 E1 and E6 mRNAs by promoting intron retention.

This project aims to investigate the role of cellular factor hnRNP D in the regulation of E1 and E6 mRNA expression.

Project II: Overexpression of m6A-factors METTL3, ALKBH5 and YTHDC1 alters HPV16 mRNA splicing.

This project aims to investigate the role of m6A-RNA modifications in HPV16 gene expression at the level of RNA processing.

Project III: Efficient production of HPV16 E2 protein from HPV16 late mRNAs spliced from SD880 to SA2709.

This project aims to identify the most efficiently translated HPV16 E2 mRNA.

Project IV: Identification of nucleotide substitutions in the 5' end of HPV16 early mRNAs and in the non-transcribed long control region that affect E6 and E7 mRNA splicing-manuscript.

This project aims to investigate the link between the HPV16 long control region and E6/E7 mRNA splicing regulation.

3 Material and Methods

3.1 Cells

HeLa, 293T, SiHa and C33A2 cells were cultured in Dulbecco's modified Eagle medium (DMEM) (HyClone) with 10% bovine calf serum (HyClone) and penicillin/streptomycin (Gibco). C33A2 cell line has been described previously [176]. Briefly, C33A2 is an in-house cell line derived from C33A cells stably transfected with HPV16 reporter plasmid pBELsLuc [177]. HPV16-positive tonsillar cancer cell line HN26 has been described previously [178]. Briefly, the HN26 cells are derived from a tumor of a 48-year-old non-smoking 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. HN26 cells were cultured in RPMI 1640 medium (HyClone) with 10% iron-supplemented bovine calf serum (HyClone), 5% MEM Non-essential Amino Acid Solution (Sigma Aldrich) and 5% sodium pyruvate (Sigma Aldrich). Neonatal Primary Normal Human Foreskin Keratinocytes (nHFK, purchased from Thermo Fisher Scientific) and 3310 cells were cultured in EpiLife medium (Gibco) supplemented with 1% human keratinocyte growth supplement (HKGS, Gibco) and 0.2% Gentamicin/Amphotericin (Gibco). Differentiation of nHFK was induced by addition of CaCl₂ at a final concentration of 2.4 mM in the keratinocyte culture medium for 24 hrs. The HPV16-immortalized keratinocyte cell line 3310 cell line has been described previously and was generated by stable transfection of nHFK with HPV16 genome plasmid pHPV16ANE2fs [177].

3.2 Plasmids

The following plasmids have been described previously: pHPV16AN [177], pC97ELsLuc [176], pBELsLuc [177], pX856F [129] and pHPV16R [179]. Construction of phnRNP F, phnRNP I, phnRNP A2 and phnRNP Q have been described previously [180] and so has phnRNP A1 [129] and phnRNP C1 [149]. Plasmids Flag-p37, Flag-p40, Flag-p42 and Flag-p45 were kindly given by Dr R. J. Schneider [173], hnRNP G plasmid by Dr. I. C. Eperon [181], histidine and myc-tagged hnRNP L plasmid by Dr. S. Guang [182] and histidine and myc-tagged hnRNP R plasmid by Dr. P. Xu [183]. phnRNP AB encoding myc-tagged hnRNP

AB transcript variant 1(RC204360) and phnRNP DL encoding hnRNP DL transcript variant 2 (RC204064) were purchased from OriGene Technologies. phnRNP H contains the hnRNP H open reading frame driven by a Cytomegalovirus (CMV) promoter. It was constructed by polymerase chain reaction (PCR)-amplification of hnRNP H coding sequence from pET-15B-hnRNP H (generously provided by Dr. D. Black) [184] followed by subcloning into pCL086 [185]. PcDNA3-FLAG-HA-hYTHDC1(#85167), pcDNA3/Flag-METTL3 (#53739), pcDNA3/Flag-METTL14 (#53740), pcDNA3/Flag-WTAP (#53741), pFRT/TO/HIS/FLAG/HA-ALKBH5 (#38073) were purchased from Addgene, pcDNA3.1+/C-(K)DYK-FTO was purchased from GenScript.

The construction of the following plasmids pX656, pX478, pXH856F, pXH856SDmF, p16E1-3xF, p16E1SDm-3xF, pBELEN, pBELENdE1, pFLAG-hnRNPd40, pD1, pD2, pD3, pD4, pD5, pD7, pD8, pD9, pQ6A, pAGG, pD1-AGG, pD1-Q6A, pD2-AGG, pD2-Q6A, pEGFP-D40, pEGFP-D1, pEGFP-D2, pEGFP-D3, pEGFP-D4, pEGFP-D5, pEGFP-D7, pEGFP-D8, pEGFP-D9, pET32a-HA-E6SDm, pET32a-HA-E6SDm and pcDNA3.1(+)-E1SDm-3Xflag have been described in Paper 1.

The plasmids pLM16R, pLM16AN, pLM16(R+AN), pLM16(AN+R), pLL16AN, pLMS16AN1, pLMS16AN2, pLS16AN and pLSS16AN have been described in Paper 4.

3.3 Transfections

Transfections of HeLa cells and 293T cells were performed with Turbofect according to the manufacturer's protocol (Thermo Fisher Scientific). Briefly, a mixture of Turbofect: DNA ratio of 2:1 (μ l reagent: μ g DNA) for HeLa cells or 4:1 for 293T cells and DMEM without serum was incubated at room temperature for 20 min prior to dropwise addition to subconfluent cells. Transfections of nHFK were performed with ViaFect according to the manufacturer's protocol (Promega). In brief, a mixture of ViaFect: DNA ratio of 3:1 and EpiLife without serum was incubated at room temperature for 20 min prior to dropwise addition to subconfluent cells. Fluorescence images of EGFP set of plasmids transfected HeLa cells were acquired using Olympus CKX53 inverted microscope.

3.4 Nuclear and Cytoplasmic Extraction

Nuclear and cytoplasmic extracts were prepared from HeLa cells grown in 6 cm dishes at 24 h post-transfection. Cells were harvested by using NE-PER Nuclear and

Cytoplasmic Extraction Reagents (Thermo Fisher Scientific) according to the manufacturer's protocol. In brief, cell pellets were resuspended in ice cold buffer CER I with protease inhibitors (Sigma Aldrich) and vortexed vigorously prior to incubation on ice for 10 minutes. Ice-cold buffer CER II was added to the samples that were vortexed vigorously and incubated on ice for one minute. After five minutes of maximum speed centrifugation, the supernatants were collected as cytoplasmic extracts. The pellets were washed once by PBS and collected as nuclear extracts.

3.5 RNA extraction, RT-PCR and Real-time quantitative PCR (qPCR)

Total RNA was extracted from transfected cells using TRI Reagent (Sigma Aldrich) and Direct-zol RNA MiniPrep (ZYMO Research) according to the manufacturer's protocols. Reverse transcription (RT) was performed in a 20 μ l reaction using random hexamers (Invitrogen) and reverse transcriptase (Invitrogen). One μ l of cDNA was subjected to PCR-amplification. The control PCR experiments are performed on RNA samples in the absence of reverse transcriptase.

qPCR was performed in a final reaction volume of 20 μ L with 1 μ l of cDNA prepared as described above in a MiniOpticon (Bio-Rad) using the SsoAdvanced SYBR Green Supermix (Bio-Rad) according to the manufacturer's instructions. The expression levels of the mRNAs were determined from the threshold cycle (Ct), and the relative expression levels were calculated using the $2^{-\Delta\Delta C_t}$ method. Results were normalized to GAPDH. Primers used in Paper 1, 2, and 4 were list in **Table 1**.

3.6 Secreted luciferase assay

The *Metridia longa* secreted luciferase (sluc) activity in the cell culture medium of transfected cells was monitored with the help of the Ready-To-Glow Secreted Luciferase Reporter assay according to the instructions of the manufacturer (Clontech) as described previously [177]. In brief, 50 μ l of cell culture medium was added to 5 μ l of 0.5X Secreted Luciferase substrate/Reaction buffer in a 96-well plate and luminescence was determined in a Tristar LB941 Luminometer.

3.7 Western blotting

Cell extracts for Western blotting were obtained by resuspending transfected cells in radioimmunoprecipitation assay (RIPA) buffer consisting of 50 mM Tris HCl pH 7.4, 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, 2 mM EDTA, 1 mM DTT and protease inhibitor (Sigma Aldrich), followed by centrifugation at full speed for 20 minutes and collection of the supernatants. Proteins were denatured by boiling in Laemmli buffer. After SDS-PAGE, the proteins on the gels were transferred onto nitrocellulose membranes, blocked with 5% nonfat dry milk in PBS containing 0.1% Tween 20, and stained with specific primary antibodies to the indicated proteins followed by incubation with secondary antibody conjugated with horseradish peroxidase and detection with chemiluminescence reagents. Antibodies used in Paper 1 and 2 were list in **Table 2**.

3.8 ssRNA oligo pull down

Whole cell lysates of HeLa cells were prepared using cell lysis buffer consisting of 25 mM Tris HCl pH 7.4, 150 mM NaCl, 1% NP-40, 1 mM EDTA, 1 mM DTT, 200 units of RiboLock RNase Inhibitor (Thermo Fisher Scientific), Proteinase inhibitor (Sigma Aldrich) and 5% glycerol. Biotin-labeled ssRNA oligos were purchased from Sigma Aldrich. DynabeadsM-280 Streptavidin magnetic beads (Invitrogen) were bound to biotin-labeled ssRNA oligonucleotides in 200 μ l of binding buffer (10 mM Tris, pH 7.4, 150 mM NaCl, 2.5 mM MgCl₂, 0.5% Triton X-100) in strips of PCR tubes incubated at RT for 20 minutes. To pull down proteins, 15 μ g HeLa whole cell lysate was mixed with beads bound with ssRNA oligos and incubated with rotation for 1 h at room temperature followed by washing of beads 10 times with binding buffer using DynaMag 96 Side magnetic plate (Invitrogen). Proteins were eluted by boiling of the beads in Laemmli Buffer. Samples were subjected to SDS-PAGE followed by Western blot analysis with the indicated antibodies. ssRNA oligos used in Paper 1 were list in **Table 3**.

3.9 Co-Immunoprecipitation

Transfected Hela cells were lysed in cell lysis buffer as described above under “ssRNA oligo pull down”. For immunoprecipitation, anti-flag antibody (M2, Sigma Aldrich) or IgG was added to Dynabeads protein G and incubated with cell lysates under gentle rocking at 4°C overnight. The complexes were washed six times using cell lysis buffer and eluted by boiling in Laemmli buffer. Samples were subjected to SDS-PAGE followed by Western blotting with specific primary antibodies.

3.10 UV-crosslinking and immunoprecipitation (CLIP)

Transfected HeLa cells grown in 10-cm dishes were washed by ice-cold PBS followed by crosslinking twice with 0.4 J cm^{-2} UV light (254 nm) in a bio-link cross-linker (Biometra). Cytoplasmic extracts were prepared as described above. Whole cell lysates were prepared by resuspending cells in one ml of RIPA buffer and incubated on ice for 30 min with occasional vortexing to lyse cells. For immunoprecipitations, 2 μg of the anti-flag antibody (M2, Sigma Aldrich) or mouse IgG was incubated at 4°C overnight in 0.5ml of cell lysate. 20 μl of Dynabeads Protein G (10004D, Invitrogen) and 20 μl Dynabeads Protein A (10001D, Invitrogen) were blocked with 1% BSA for 0.5 h, washed three times in RIPA buffer and then added to the antibody-protein mixtures followed by incubation for 1h at 4°C . The beads were washed three times with buffer I (50 mM Tris HCl pH 7.4, 300 mM NaCl, 0.5% NP-40, 1 mM EDTA, 1 mM DTT), three times with buffer II (50 mM Tris HCl pH 7.4, 800 mM NaCl, 0.5% NP-40, 1 mM EDTA, 1 mM DTT) and three times with buffer III (50 mM Tris HCl pH 7.4, 800 mM NaCl, 250 mM LiCl, 0.5% NP-40, 1 mM EDTA, 1 mM DTT). RNA was eluted by phenol/chloroform extraction and ethanol-precipitated and dissolved in 20 μl of water. 10 μl of immunoprecipitated RNA was reverse transcribed using M-MLV reverse transcriptase (Invitrogen) and random primers (Invitrogen) according to the protocol of the manufacturer. Two microliters of cDNA were subjected to PCR amplification.

3.11 Ribonucleoprotein (RNP) immunoprecipitation (RIP) analysis

For immunoprecipitation of endogenous ribonucleoprotein (RNP) complexes (RIP analysis) from whole cell extracts, HN26 cells or 3310 cells were lysed in cell lysis buffer. The supernatants were incubated with anti-AUF1 antibody (Cell signalling) or IgG (Millipore) overnight at 4°C . Dynabeads Protein G (10004D, Invitrogen) + 20 μl Dynabeads Protein A (10001D, Invitrogen) were added to the antibody-protein mixtures followed by incubation for 1h at 4°C . After washing of the beads six times using cell lysis buffer, RNA was eluted using Tri reagent and incubated with 20U of RNase-free DNase I for 1 h at 37°C and subjected to RT-PCR.

3.12 siRNA library and siRNA transfections

ON-TARGETplus human hnRNP D siRNA SMARTpool consists of four siRNAs to hnRNP D (L-004079-00-0010) (Dharmacon). A scrambled negative control

siRNA (D-001810-10-05) was also purchased from Dharmacon. Transfections were conducted with DharmaFECT1 (Dharmacon) according to the instructions of the manufacturer. The siRNA SMARTpool to hnRNP D or scrambled control siRNAs were transfected in duplicates into C33A2 or SiHa cells grown in 12-well plates for RNA extraction and RT-qPCR or in 6-well plates for protein extraction and Western blotting. Cells were harvested at 48 h post-transfection for RNA extraction and RT-qPCR or 72 h post-transfection for protein extraction and Western blotting performed as described above.

3.13 Lentiviral based shRNA for knockdown

Lentivirus for the short hairpin RNA (shRNA)-mediated knockdown of ALKBH5 was generated by co-transfection of HEK293T cells with a pLKO.1 vector or pLKO.1 vector carrying specific shRNA together with the packaging vector pMISSION-GAG-POL and a vesicular stomatitis virus G protein expressing vector pMISSION-VSV-G. pLKO.1 was purchased from Sigma-Aldrich (SHC001). Two days post transfection, lentivirus-containing supernatants were harvested, centrifuged to remove cellular debris, and filtered with a 0.45- μ m filter. Lentivirus production efficiency was determined in parallel using a GFP overexpression lentivirus vector. C33A2 cells were inoculated with stocks of recombinant lentiviruses by centrifugation at 2,000 g for 2 h at room temperature in the presence of 10 μ g/ml polybrene (Fisher Scientific). Empty pLKO.1 vector was used as negative control. Cells were then resuspended and grown in normal RPMI media for 2 days, after which transduced cells were selected in the presence of puromycin (1 μ g/ml). Cells were either harvested for Western blotting or for RNA extraction and RT-PCR.

3.14 *In vitro* translation assay

In vitro translation was carried out with TNT^(R) Quick Coupled Transcription/Translation Systems (Promega) according to the instructions of the manufacturer. In brief, 100 ng pET32a-HA-E6SDm or 200 ng pcDNA3.1(+)-E1SDm-3XFlag plasmid was translated in the absence or presence of 1 μ M recombinant hnRNP D protein (EUPROTEIN) or 1 μ M BSA. The 25 μ L reactions were incubated for 90 min at 30°C. The translation reactions were analyzed by Western blotting as described above. The Luciferase control RNA was also translated in the absence or presence of 1 μ M hnRNP D or 1 μ M BSA. Luciferase activity in the translation reactions were monitored according to the instructions of the manufacturer using Tristar LB941 Luminometer. Recombinant hnRNP D and

BSA were separated on SDS-PAGE followed by staining with Colloidal Blue Staining Kit (Invitrogen).

3.15 *In vitro* RNA syntheses

In vitro RNA was synthesized using the TranscriptAid T7 High Yield Transcription Kit (#K0441, Thermo Scientific). Three reactions were made in parallel in which 0-, 1- or 10% of the ATP-pool were substituted for the N6-methyladenosine base analogue m6A (S3190, Selleckchem).

3.16 RNA preparation for MeRIP-seq

To establish MeRIP-seq libraries, total RNA extracted (Direct-zol RNA miniprep, Zymo Research) from 5×10^7 HN26 cells. Total RNA quality (RNA Nano Chips, Agilent Technologies) was measured by bioanalyzer. Poly-A selection of the mRNA by oligo-dT beads (Dynabeads Oligo(dT), Thermo Fisher). mRNA quality was measured (RNA Nano Chips, Agilent Technologies) by bioanalyzer. mRNA was fragmented using RNaseA to a size between 30-50nt. 0.2 ng RNaseA (10mg/ml, Thermo Fisher) for 1 μ g mRNA at 37°C/5min then add 1 μ l ribolock (RNase inhibitor, 40 U/ μ l, Thermo Fisher) incubate 5 min in RT then put samples on ice. The size and quality of fragmented mRNA (Small RNA Chips, Agilent Technologies) was measured by bioanalyzer. Immunoprecipitation of fragmented mRNA with m6A antibody (Rb pAb m6A, abcam) and IgG (R IgG, Cell Signaling). Extract the IP RNA (phenol-chloroform/precipitation) then resuspended by 10 μ l H₂O and check the final quality of RNA by bioanalyzer (Small RNA Chips, Agilent Technologies).

3.17 Quantitations

The software used to determine band intensity in Western blots and RT-PCR gels is "Image Lab 6.1.0" and quantitations were performed with the software "Prism GraphPad 8.4.0".

Table 1. List of primers

Amplified region	primers	Sequence 5'-3'
E6/E7	97S	GTCGACCTGCAATGTTTCAGGACCC
	880AS	GAAACCATAATCTACCATGGCTGATC
	438AS	GCTCGAGGGAATCTTTGCTTTTTGTCCAGATGTCT
	757AS	CGTGTGTGCTTTGTACGCACAACCG
	TotalE6F	TGTTTCAGGACCCACAGGAG
	TotalE6R	TTGCTTGCAGTACACACATTC
	234AS	TGCATAAATCCCGAAAAGCAAAGT
E2	773S	GCACACACGTAGACATTCGTACTTTG
	E2AS	GTCCAGATTAAGTTTGCACGAGGAC
	E2QAS	CAGCCAGCGTTGGCACCACCT
E4	773S	GCACACACGTAGACATTCGTACTTTG
	E4AS	CCTCTCCTGAAATTATTAGGCAGCA
E1	773S	GCACACACGTAGACATTCGTACTTTG
	E1AS	CCATCCATTACATCCCGTACC
	1302S	CTGAAGTGGAACTCAGCAGATGTTAC
	2293AS	TAATGATTTACCTGTGTTAGCTGCACC
	880S	GAAACCATAATCTACCATGGCTGATC
	F-E1-1	AGTAGAGCTGCAAAAAGGAGATTA
	PstI-SD880mE1-F	GATCCTGCAGCTACCAATGGGGAAGAGGGTAC
E2Xba		AACCAGGTTCTAGACAAACTTAATCTGGACCACG
		AACCAGGTTCTAGACAAACTTAATCTGGACCACG
LoxP	16S	TATGTATGGTATAATAAACACGTGTGTATGTG
	16A	GCAGTGCAGGTCAGGAAAACAGGGATTTGGC
hnRNPD isoforms	2S	GAAGATTGACGCCAGTAAGAACG
	7A	TACCTTCCCATAACCACTCTGCT
pX478	B97S	GCGCGCTGCAATGTTTCAGGACCC
	X478A	GCTCGAGGGAATCTTTGCTTTTTGTCCAGATGTCT
pX656	B97S	GCGCGCTGCAATGTTTCAGGACCC
	X656A	CCTCGAGCTGTCAATTAATGCTCATAACAG
pXH856F	sense	GACAGCGCGCACCATGTACCCATACGATGTTCCAGATTACGCTATGTTCAGGACCCACAG
	anti sense	TTACTCGAGCTACTTATCGTCGCATCCTTGTAACTGGTTTCTGAGACACAGAT
pXH856SDmF	sense	CTGCGACGTGAGGCCTATGACTTTGC
	anti sense	GCAAAGTCATAGGCCTCACGTCGCAG
p16E1-3xF	sense	ATCAGCGCGCCACCATGGCTGATCCTGCAGGTAC
	anti sense	TAATCGATGTCATGATCTTTATAATCACCGTCATGGTCTTTGTAGTCTAATGTGTTAGTATTTG
	anti sense	TCCCTCGAGCTACTTGTGCATCGTCATCCTTGTAACTCGATGTCATGATC
p16E1SDm-3xF	sense	GATCCTGCAGCTACCAATGGGGAAGAGGGTAC
	anti sense	GGCGCCCTTCTAGCTGTAACATCTGCTGAGT
	anti sense	TACAGCTAGAAGGGCGCCATGAGACTG
pflag-D40	sense	GTTGCGCGGTTATCATGGACTACAAAGACGATGACGACAAG
	anti sense	CCTCGAGTTAGTATGGTTTGTAGCTATTTTGTAGACCACC
pD1.pD1-AGG and pD1-Q6A	sense	GGCGCGGCCACCATGGACTACAAAGACGATGACGACAAGATGTTTATAGGAGGCCTTAGCTGGG
	anti sense	CCTCGAGTTAGTATGGTTTGTAGCTATTTTGTAGACCACC

pD2,pD2-AGG and pD2-Q6A	sense	GGCGCGCGCCACCATGGACTACAAAGACGATGACGACAAGATGGTT AAAAAAAAATTTTGTGGTGGCC
	anti sense	CCTCGAGTTAGTATGGTTTGTAGCTATTTTGATGACCACC GTTGCGCGCGTTATCATGGACTACAAAGACGATGACGACAAGGAA GCGGAGCCGGGAC
pD5	sense	CCTCGAGTTAGTATGGTTTGTAGCTATTTTGATGACCACC
	anti sense	GTTGCGCGCGCCACCATGGACTACAAAGACGATGACGACAAG
pD3	sense	CCGCTCGAGTTAGGCTTTGGCCCTTTTAGGATC
	anti sense	GTTGCGCGCGCCACCATGGACTACAAAGACGATGACGACAAG
pD4	sense	CCGCTCGAGTTACATGGCTACTTTTATTTACATTTAC
	anti sense	GTTGCGCGCGCCACCATGGACTACAAAGACGATGACGACAAG
pD7	sense	CTTCTCGAGTTAAGATCCCCTGTTGCTGTTGCTGATATTG
	anti sense	GGCGCGCGCCACCATGGACTACAAAGACGATGACGACAAGATGTT ATAGGAGGCCTTAGCTGGG
pD8	sense	CCGCTCGAGTTACATGGCTACTTTTATTTACATTTAC
	anti sense	GTTAAGCTTGTATCATGGACTACAAAGACGATGACGACAAG
pEGFP-D40	sense	CTTGGATCCTTAGTATGGTTTGTAGCTATTTTGATGACCACC
	anti sense	GTTAAGCTTGTATCATGGACTACAAAGACGATGACGACAAG
pEGFP-D1	sense	CTTGGATCCTTAGTATGGTTTGTAGCTATTTTGATGACCACC
	anti sense	GTTAAGCTTGTATCATGGACTACAAAGACGATGACGACAAG
pEGFP-D2	sense	CTTGGATCCTTAGTATGGTTTGTAGCTATTTTGATGACCACC
	anti sense	GTTAAGCTTGTATCATGGACTACAAAGACGATGACGACAAG
pEGFP-D5	sense	CTTGGATCCTTAGTATGGTTTGTAGCTATTTTGATGACCACC
	anti sense	GTTAAGCTTGTATCATGGACTACAAAGACGATGACGACAAG
pEGFP-D3	sense	CTTGGATCCTTAGTATGGTTTGTAGCTATTTTGATGACCACC
	anti sense	GTTAAGCTTGTATCATGGACTACAAAGACGATGACGACAAG
pEGFP-D4	sense	CTTGGATCCTTAGGCTTTGGCCCTTTTAGGATC
	anti sense	GTTAAGCTTGTATCATGGACTACAAAGACGATGACGACAAG
pEGFP-D7	sense	CTTGGATCCTTACATGGCTACTTTTATTTACATTTAC
	anti sense	GTTAAGCTTGTATCATGGACTACAAAGACGATGACGACAAG
pEGFP-D8	sense	CTTGGATCCTTAAAGATCCCCTGTTGCTGTTGCTGATATTG
	anti sense	GTTAAGCTTGTATCATGGACTACAAAGACGATGACGACAAG
pEGFP-D9	sense	CTTGGATCCTTACATGGCTACTTTTATTTACATTTAC
	anti sense	GTTAAGCTTGTATCATGGACTACAAAGACGATGACGACAAG
GAPDH	GAPDH	ACCCAGAAGACTGTGGATGG
	GAPDHR	TTCTAGACGGCAGGTCAGGT
spliced actin	actin-s	TGAGCTGCG TGTGGCTCC
	actin-a	GGCATGGGGGAGGGCATAACC
unspliced actin	actin-s1	CCAGT GGCTTCCCAGTG
	actin-a	GGCATGGGGGAGGGCATAACC
Involucrin	INV_S	CTCTGCCTCAGCCTTACTGTGA
	INV_R	GCTCCTGATGGGTATTGACTGG
pET32a-HA-E6SDm	sense	TATGAATTCATGTACCCATACGATGTTCCAGAT
	antisense	TACTCGAGCTATGTTTCTGAGAACAGATGGGG GCTCCTGATGGGTATTGACTGG
pLM16AN	H3FM	GGCAAGCTTAATAAACTTATTGTTTCAACACC
	SBFA	CATTGGTACCTGCAGGATCAGCCAT
pLM16(AN+R)	H3FM	GGCAAGCTTAATAAACTTATTGTTTCAACACC
	SBFA	CATTGGTACCTGCAGGATCAGCCAT
pLM16(R+AN)	overlap S	CGAAATCGGTTGAACCGAAACCGGTTAGTATAAAAGCAGA

	overlap A	TCTGCTTTTATACTAACC GGTTTCGGTTCAACCGATTTCG
pLL16AN	H3F	GGCAAGCTTGTATTGTATGTATGTTGAATTAGTGTTG
	SBFA	CATTGGTACCTGCAGGATCAGCCAT
pLMS16AN1	H3FMS1	GGCAAGCTTCATTGTATATAAACTATATTTGCTACATCC
	SBFA	CATTGGTACCTGCAGGATCAGCCAT
pLMS16AN2	H3FMS2	GGCAAGCTTTATATATACTATATTTTTGTAGCGCCAG
	SBFA	CATTGGTACCTGCAGGATCAGCCAT
pLS16AN	H3FSPH	GGTAAGCTTGCATGCTTTTTGGCACAAAATGT
	SBFA	CATTGGTACCTGCAGGATCAGCCAT
pLSS16AN	7620HS	GGCAAGCTTGGGACATGCATGCTATGTGCAACTACTGAATCACT
	SBFA	CATTGGTACCTGCAGGATCAGCCAT
pLM16AN	H3FM	GGCAAGCTTAATAAACTTATTGTTTCAACACC
	SBFA	CATTGGTACCTGCAGGATCAGCCAT
pLM16(AN+R)	H3FM	GGCAAGCTTAATAAACTTATTGTTTCAACACC
	SBFA	CATTGGTACCTGCAGGATCAGCCAT
pLM16(R+AN)	overlap S	CGAAATCGGTTGAACCGAAACCGTTAGTATAAAAAGCAGA
	overlap A	TCTGCTTTTATACTAACC GGTTTCGGTTCAACCGATTTCG
pLL16AN	H3F	GGCAAGCTTGTATTGTATGTATGTTGAATTAGTGTTG
	SBFA	CATTGGTACCTGCAGGATCAGCCAT
pLMS16AN1	H3FMS1	GGCAAGCTTCATTGTATATAAACTATATTTGCTACATCC
X556A	antisense	GCTCGAGCAGCTGGGTTTCTCTACGTGTT
MALAT1	MALAT1s	CGTAGACCAGAACCAATTTAGAAG
	MALAT1as	CATATTGCCGACCTCACGGAT
	MALAT1asQ	AGCACCTGGGTCAGCTGTCAAT
RPLP0	RPL0s	ACCTGGAAGTCCAACCTACTTCCTTA
	RPL0as	GATCTCAGTGAGGTCCTCCTTG
T7s	in vitro RNA	ACGTTAAGGGATTTTGGTCATGAGA
T7as	in vitro RNA	TCAAATATGTATCCGCTCATGAGA

Tabel 2. List of antibodies

	Cat.	Company
AUF1/hnRNPD	12382S	Cell Signaling
hnRNPM	ab177957	abcam
hnRNPU	ab10297	abcam
Flag,M2	F3165/F1804	Sigma Aldrich
U1 70K	ab83306	abcam
U2AF65	sc-53942	Santa Cruz
U2AF35	ab172614	abcam
HPV16 E6	GTX132686	Genetex
HPV16 E7	GTX133411	Genetex
β -actin	A5441	Sigma Aldrich
Tubulin	T9026	Sigma Aldrich
Histone	4620S	CST
involucrin	sc-21748	Santa Cruz
HA	sc-7392	Santa Cruz
Lamin B	ab16048	abcam
SRSF1	ab38017	abcam
SRSF2	ab204916	abcam
SF3b	ab172634	abcam
PABP-C1	ab6125	abcam
veriblot	ab131366	abcam
Normal Rabbit IgG	2729S	Cell Signaling
Normal Mouse IgG	12-371	Millipore
anti-Mouse IgG-HRP	A9044	Sigma Aldrich
anti-Rabbit IgG-HRP	A9169	Sigma Aldrich
SF3b	ab172634	abcam
PABP-C1	ab6125	abcam
veriblot	ab131366	abcam
m6A	ab151230	abcam
ALKBH5	ab174124	abcam

Tabel 3. List of ssRNA oligos

HPV16 pull down region	Sequence 5'-3'
1, 178-214	UAUAAUAAUAGAAUGUGUGUACUGCAAGCAACAGUUA
2, 196-233	AGUUACUGCGACGUGAGGUAAUAGACUUUGCUUUUCGG
3, 229-266	UUCGGGAUUUAUGCAUAGUAUAUAGAGAUGGGAAUCCA
4, 276-313	AUCCAUAUGCUGUAUGUGAUAAAUGUUUAAAGUUUUU
5, 309-346	UUUAUUCUAAAAUUAGUGAGUAUAGACAUAUUAUGUUU
6, 342-379	GUUAUAGUUUGUAUGGAACAACAUAUAGAACAGCAAUAC
7, 375-412	AAUACAACAACCGUUGUGUGAUUUUGUUAAUUAGGUGU
8, 408-445	GGUGUAUUAAACUGUCAAAAAGCCACUGUCCUGAAGAA
9, 441-478	AAGAAAAGCAAAGACAUCUGGACAAAAAGCAAAGAUUC
10, 474-498	GAUCCAUAUUUAUAAAGGGGUCGGUG
11, 492-516	GUCGGUGGACCGGUCGAUGUAUGUC
12, 506-530	CGAUGUAUGUCUUGUUGCAGAUCAU
13, 521-545	UGCAGAUCAUCAAGAACACGUAGAG
14, 536-560	ACACGUAGAGAAACCCAGCUGUAAU
15, 549-573	CCCAGCUGUAAUC AUGCAUGGAGAU
16, 564-588	GCAUGGAGAUACACCUACA UUGCAU
17, 579-604	UACA UUGCAUGAAUUAUGUUAGAUU
18, 594-620	UAUGUUAGAUUUGCAACCAGAGACAAC
19, 611-635	CAGAGACAACUGAUCUCUACUGUUA
20, 631-665	UGUUAUGAGCAAUUAUAAUGACAGCUCAGAGGAGGA
21, 661-695	GAGGAGGAUGAAUUAAGAUUGGUCCAGCUGGACAAGC
22, 691-725	CAAGCAGAACCGGACAGAGCCAUUACAUAUUGU
23, 721-755	AUUGUAACCUUUUGUUGCAAGUGUGACUCUACGCU
24, 751-785	ACGCUUCGGUUGUGCGUACAAAGCACACACGUAGA
25, 781-815	GUAGACAUUCGUACUUUGGAAGACCUGUUAAUGGG
26, 811-845	AUGGGCACACUAGGAAUUGUGUGCCCCAUCUGUUC
27, 841-875	UGUUCUCAGAAACCAUAAUCUACCAUGGCUGAUCC
28, BSD+GGG	CUGAUCCUGCAGGUACCAAUGGGGAAGAGGGUACGGGAUGUA

4 Results

4.1 HnRNP D activates production of HPV16 E1 and E6 mRNA by promoting intron retention

To better understand how hnRNPs regulate HPV16 early gene expression, we determined the effects of different hnRNPs on E6/E7 mRNAs, E1/E2 mRNAs, and E4 mRNAs splicing. HnRNP D became an intriguing candidate since it dramatically inhibits both E6/E7 mRNAs and E1/E2 mRNAs splicing, leading to E6 and E1 intron retention enhancement. Furthermore, we tested all four isoforms of hnRNP D effects on E6/E7 mRNAs and E1/E2 mRNAs splicing. The results showed that all hnRNP D isoforms inhibited HPV16 early mRNA alternative splicing to some extent, and hnRNP D40 has a stronger inhibitory effect than the other three isoforms. Since there is not much known about the role of hnRNP D in HPV16 early mRNA splicing, we reasoned that this research project would also contribute to the understanding of the function of hnRNP D.

4.1.1 Both N- and C-termini of hnRNP D40 contribute to hnRNP D40 splicing control ability, but in different ways.

To investigate how hnRNP D40 exerted inhibitory effect on E1/E2, E6/E7 mRNA splicing, we identified the functional domains of hnRNP D40 first. hnRNP D40 and its deletion mutants were transfected with HPV16 plasmid pC97ELsLuc. The effect on splicing was determined by RT-PCR. The results revealed that wild-type hnRNP D40 inhibited E1/E2, E6/E7 mRNA splicing and thus promoted the production of intron-retained E1- and E6-encoding mRNAs (**Figure 4.1**), respectively.

All of the hnRNP D40 mutants with an N-terminal deletion (D1, D2, D5, and D8) have a decreased ability to induce the production of intron-retained E1 and E6 mRNAs (**Figure 4.1A, B**). D5 lost N-terminal Alanine-rich domain, which reduced the capacity of hnRNP D40 to promote production of intron-retained E6 mRNAs only. D1, further lost exon-2 coding region of hnRNP D40, reduced induction of both intron-retained E6 and E1 mRNAs, and unexpectedly activated splicing from SD226 to SA526. In D2, deletion of the RRM1 domain eliminated induction of both intron-retained E6 and E1 mRNAs, and activated production of isoform spliced from SD226 to SA409. Mutant D8 consisted of only RRM1 and RRM2 domains

and had a minor effect on E1/E2 or E6/E7 mRNA splicing. We concluded that the N-terminus of hnRNP D40 significantly inhibited HPV16 early mRNA splicing and intron-retained E1 and E6 mRNA production. Specifically, the RRM1 domain was essential.

C-terminal deletion mutants D3, D4, D7 and D9 enhanced production of intron-retained E1 and E6 mRNAs, although less efficiently than wild-type hnRNP D40 (Figure 4.1C, D). With increasing C-terminal deletions, the splicing inhibition was gradually eliminated. Taken together, these results supported the idea that N-terminus of hnRNP D40 was critical in inhibiting HPV16 E1/E2 and E6/E7 mRNA splicing and that this effect was increased by the C-terminus, implying that the C-terminus may interact with the splicing machinery.

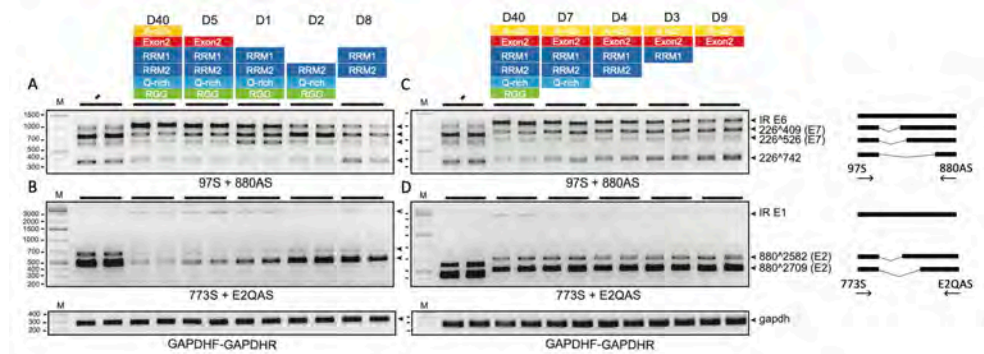


Figure 4.1 Functional domains of hnRNP D40 in E1/E2, E6/E7 mRNA splicing.

RT-PCR results showed that hnRNP D40 inhibited E1/E2, E6/E7 mRNA splicing. RRM1 domain of hnRNP D40 was critical for HPV16 E6 and E1 intron retention whereas N-terminal and C-terminal domains contributed to modification of HPV16 early alternative mRNA splicing. Schematic representation of amplified mRNAs and RT-PCR primers are indicated to the right.

To confirm the interaction of hnRNP D40 with splicing machinery, we first investigated if hnRNP D40 and two N-terminal D1, D2 deletion mutants co-immunoprecipitated spliceosome components. The results showed that wild-type hnRNP D40 co-immunoprecipitated cellular U1 snRNP components U1-70K and U2 auxiliary factors U2AF65 and U2AF35 (Figure 4.2A). Deletion N-terminal region reduced interactions of hnRNP D40 with cellular spliceosome, which was reflected by D1 and D2 results. Thus, we concluded that hnRNP D40 interacts with spliceosome complex A components at its C-terminus and that the N-terminus contributed to the efficiency of these interactions.

Also, we used UV-crosslinking and Immunoprecipitation (CLIP) to detect the interaction of hnRNP D40 and HPV16 RNA in HeLa cells that were co-transfected with hnRNP D40 or D1, D2 mutants and pC97ELsLuc plasmids. The results revealed wild-type hnRNP D40 and D1 interacted with HPV16 mRNA but D2 did

not, which indicates that the RRM1 domain of hnRNP D40 plays a key role in this direct interaction (**Figure 4.2B**).

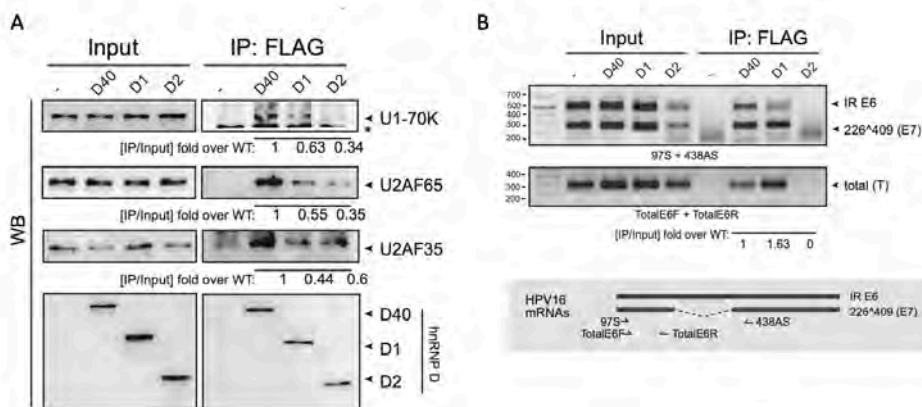


Figure 4.2 hnRNP D40 contributes to HPV16 mRNA control via interacting with spliceosome factors and HPV16 mRNAs.

A. Co-IP results showed that hnRNP D40 C-terminus interacts with spliceosome components U1-70K, U2AF65 and U2AF35. **B.** CLIP assay showed that RRM1 domain mediates the interaction of hnRNP D40 and HPV16 mRNAs. Schematic representation of amplified mRNAs and RT-PCR primers are indicated to the right bottom.

To further identify HPV16 sequences that were targeted by hnRNP D40, we used RNA-mediated protein pull-down assay. Interactions between sequences in the E6/E7 coding region was detected by RNA-pull down of hnRNP D40 with overlapping oligos (**Figure 4.3**). One obvious result revealed that all RNA oligos spanning E6-encoding intron between SD226 and SA409 pulled down hnRNP D40. In addition, a previously identified splicing silencer also pulled down hnRNP D40. These interactions correlated with the inhibition ability of hnRNP D40 on E6/E7 splicing and the promotion of the intron retention.

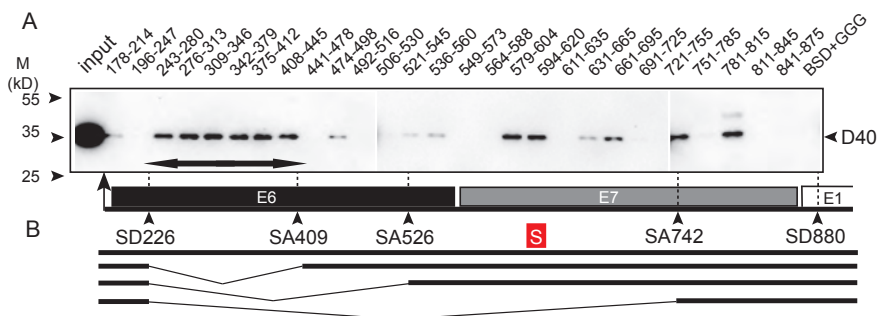


Figure 4.3 hnRNP D40 interacts with E6 intronic region and E6 splicing silencer.

A. RNA oligo pull-down results showed that hnRNP D40 interacts with E6 intronic region (SD226-SA409) and E6 splicing silencer. **B.** Schematic representation of RNA splicing in the HPV16 E6 and E7 coding region.

4.1.2 hnRNP D40 increases the levels of HPV16 intron-retained E6 and E1 mRNAs in the cytoplasm

Additionally, our findings indicated that hnRNP D40 should increase E6 protein levels and decrease E7 protein levels correspondingly since E6 and E7 mRNAs were upregulated or down regulated as the results of splicing inhibition function of hnRNP D40. Similar results were obtained from pC97ELsLuc and shorter plasmid pXH856F which encodes only E6 and E7. Surprisingly, hnRNP D40 inhibited both E6 and E7 protein expression levels (**Figure 4.4**), not only E7 reduction suggested by the mRNA results. To investigate if this inhibition of protein expression correlated with the regulation function of hnRNP D40 on E6/E7 mRNA splicing, 5' splice site mutation plasmid pXH856SDmF was also analyzed. Both E6 and E7 protein expression was suppressed by hnRNP D40, which was the same results as its wild-type version pXH856F. These results suggested that inhibition of E6 and E7 protein production was independent of the splicing process. hnRNP D40 may also have an effect on other RNA processing processes, which would explain the decrease in E6 protein levels.

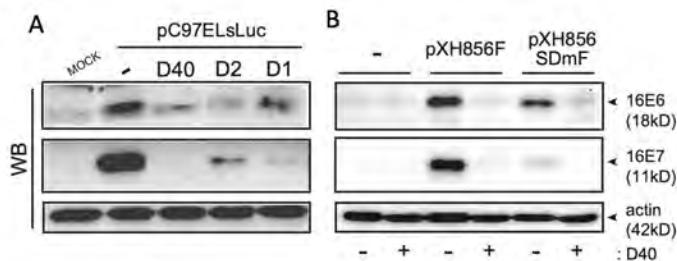


Figure 4.4 hnRNP D40 inhibits HPV16 E6 and E7 protein production.

WB results showed that the inhibition of E6 and E7 protein production was independent of E6/E7 mRNA splicing process. E6 and E7 protein produced from subgenomic plasmid pC97ELsLuc, plasmid encodes only E6/E7 region pXH856F and its 5' splice site mutant pXH856SDmF were reduced by hnRNP D40.

One possibility was that hnRNP D40 restricted intron-retained E6 mRNA export to the cytoplasm, thus affecting protein production. But the subcellular distribution of intron-retained E6 mRNAs were increased in cytoplasm of pC97ELsLuc in presence of hnRNP D40 (**Figure 4.5A, B**). Taken together, we concluded that hnRNP D40 had multiple impacts and affected various steps of HPV16 E6 and E7 mRNA processing including inhibition of E6/E7 mRNA splicing and increasing levels of intron-retained E6 mRNA in cytoplasm. However, hnRNP D40 also significantly reduced E6 protein produced from the intron-retained E6 mRNAs, implying that hnRNP D40 may inhibit mRNA translation. Meanwhile, intron-retained E1 mRNAs levels were also increased in both nucleus and cytoplasm (**Figure 4.5C**), whereas the levels of HPV16 spliced E2 mRNAs (880[^]2709) were reduced in the presence of hnRNP D40, and were located primarily in the cytoplasmic fraction.

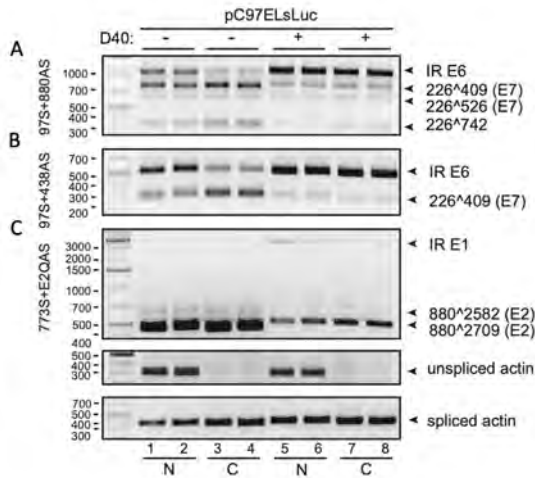


Figure 4.5 The levels of intron- retained E6 and E1 mRNA in cytoplasm were increased in the presence of hnRNP D40.

RT-PCR results of nuclear and cytoplasmic fractions showed that subcellular distribution of intron-retained E6 and E1 mRNAs was affected by hnRNP D40. In the presence of hnRNP D40, primarily intron-retained E6 and E1 mRNA were increased in the cytoplasm compared to the spliced mRNAs.

Finally, we attempted to investigate if the increase of intron-retained E1 mRNAs in the cytoplasm caused by hnRNP D40 also resulted in increased E1 protein levels. In fact, E1 protein levels were decreased which is consistent with the inhibitory effect of hnRNP D40 on E6 mRNA and protein production (**Figure 4.6**).

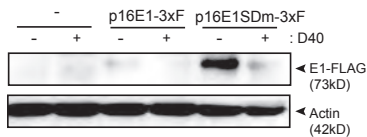


Figure 4.6 hnRNP D40 inhibits HPV16 E1 protein production.

WB results showed that the inhibition of E1 protein production was independent of E1 mRNA splicing process. E1 protein produced from E1 expression plasmid p16E1-3xF and its splice donors mutant p16E1SDm-3xF was reduced by hnRNP D40.

The results presented above suggested that hnRNP D40 not only inhibited HPV16 mRNA splicing, but also might have the ability to accompany the HPV16 mRNAs to the cytoplasm. Western blot results showed that hnRNP D40 produced from expression plasmid localized to both nuclear and cytoplasmic fractions (**Figure 4.7A**). Furthermore, hnRNP D40 in cytoplasm associated with intron-retained HPV16 E1 mRNAs (**Figure 4.7B**), suggesting that hnRNP D40 followed with

HPV16 mRNAs from nuclear to cytoplasm and regulated intron-retained E1 and E6 mRNA translation negatively.

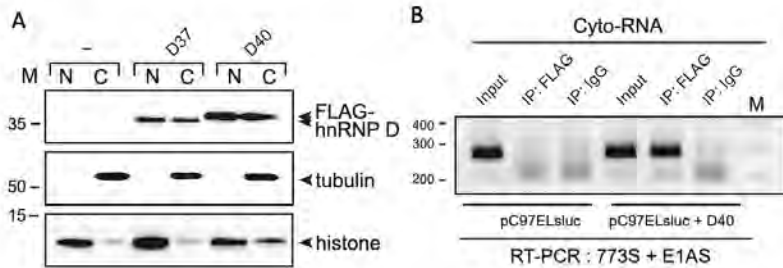


Figure 4.7 hnRNP D40 interacts with HPV16 mRNAs in cytoplasm.

A. WB results of nuclear and cytoplasmic fractionation showed that hnRNP D40 produced from expression plasmid could be detected in both nuclear and cytoplasmic. **B.** CLIP assay showed that hnRNP D40 interacts with intron-retained E1 mRNA in cytoplasm.

4.1.3 hnRNP D40 interacts with HPV16 mRNAs in an HPV16-positive cancer cell line and in HPV16-immortalized human keratinocytes

The ability to promote the production of intron-retained E6 mRNAs and E1 mRNAs of hnRNP D40 was also confirmed in human primary keratinocytes which contain integrated HPV16 DNA (**Figure 4.8A**). Due to calcium treatment, human primary keratinocytes undergo differentiation, thereby increasing the levels of intron retained E1 mRNAs. This effect was reinforced by hnRNP D40. Furthermore, the association of HPV16 early mRNAs and endogenous hnRNP D was detectable in HPV16 positive tonsillar cancer cell line HN26 and in-house HPV16-immortalized keratinocyte cell line 3310 with integrated HPV16 DNA (**Figure 4.8B**). Knockdown hnRNP D in HPV16 positive cell line C33A2 and in cervical cancer cell line SiHa reduced the levels of intron-retained E6 and E1 mRNAs, meanwhile increasing the E7 protein levels (**Figure 4.8C, D**). We concluded that the levels of hnRNP D had a direct impact on production of the HPV16 E7 oncoprotein in cervical cancer cells. These results reflected the physiological relevance of hnRNP D and HPV16 early gene expression as well as its contribution to E7 protein in cervical cancer cells.

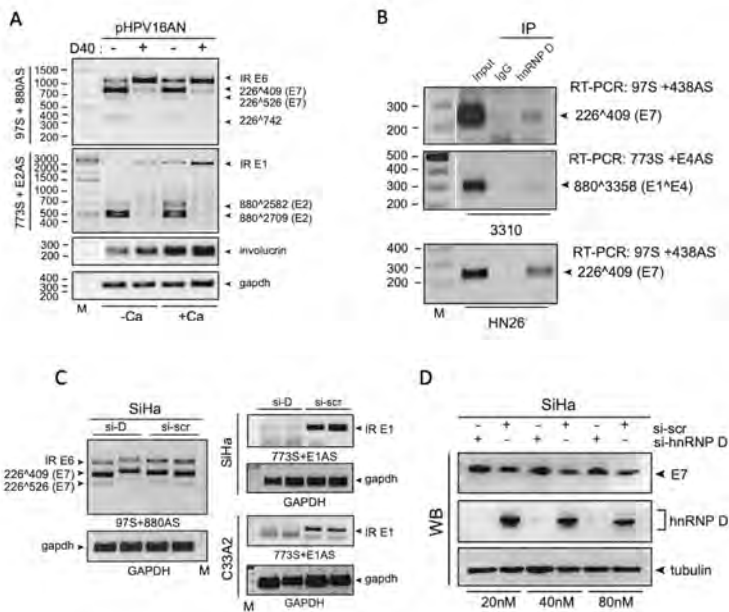


Figure 4.8 hnRNP D associates with E6/E7 and E1/E2 mRNAs in HPV16-immortalized human keratinocytes and HPV16 positive cancer cell line.

A. RT-PCR results showed that hnRNP D40 promotes the levels of intron-retained E6 and E1 mRNAs produced from episomal HPV16 DNA in human primary keratinocytes transfected with genomic plasmid pHPV16AN and CRE plasmid. hnRNP D40 reinforced the increasement of intron-retained E1 mRNA induced by calcium treatment. **B.** RIP results showed that endogenous hnRNP D interacts with HPV16 mRNAs in HPV16-immortalized keratinocyte cell line 3310 and HPV16-positive cancer cell line HN26. **C.** Knock down of hnRNP D in HPV16-positive cancer cell line SiHa reduced the intron-retained E6 and E1 mRNAs. Knock down of hnRNP D in pBELsLuc stable cell line C33A2 reduced intron-retained E1 mRNAs as well. **D.** WB results showed that Knock down of hnRNP D in SiHa increased E7 oncoprotein production.

Summary and Significance

- We have elucidated how the elusive, intron-retained HPV16 E6 and E1 mRNAs are generated.
- We have shown that expression of both E6 and E7 oncogenes are affected by the hnRNP D levels.
- We found that reducing hnRNP D levels could enhance E7 mRNA and protein production at the expense of E6.
- Our results contribute to understanding the function of hnRNP D protein, particularly its role in splicing regulation.
- hnRNP D regulated HPV16 mRNA splicing could potentially be a target for therapy to HPV16 infections and cancer.

4.2 m6A modifications on HPV16 mRNA regulate alternative splicing

The importance of m6A modification, the prevalent mammalian mRNA internal modification, has been deeply investigated in viral gene expression regulation in recent years. In this study, we started by discovering the core m6A modification “eraser,” “writer,” and “reader” proteins and its reversible dynamic effects on HPV16 mRNA post-transcriptional regulation, especially alternative splicing. The results indicated that the effects of m6A “eraser,” “writer,” and “reader” proteins on HPV16 mRNA alternative splicing are mediated by m6A sites on HPV16 transcripts. Later, we wished to map m6A sites on HPV16 mRNAs by RNA-seq. The sequencing results yielded peaks representing promising m6A sites on HPV16 mRNA, though precise interpretation and confirmation are needed.

4.2.1 m6A “eraser” ALKBH5, “writer” METTL3 and “reader” YTHDC1 alter HPV16 mRNA splicing.

To investigate if m6A-erasers (ALKBH5, FTO), m6A-writers (METTL3, METTL14, WTAP) and m6A-readers (YTHDC1) affected HPV16 mRNA splicing, HPV16 reporter plasmid pC97ELsLuc was co-transfected into Hela cells with expression plasmids pALKBH5, pFTO, pMETTL3, pMETTL14, pWTAP or pYTHDC1. First, the results revealed that ALKBH5 inhibits HPV16 E6/E7 mRNA splicing (**Figure 4.9**).

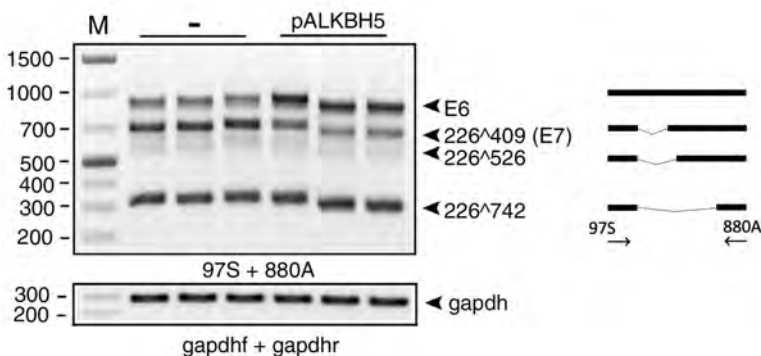


Figure 4.9 m6A eraser ALKBH5 inhibits HPV16 E6/E7 mRNA splicing.

RT-PCR showed that ALKBH5 inhibited E6/E7 mRNA splicing. Intron-retained E6 mRNAs was increased at the expense of spliced mRNA 226^Δ409. Schematic representation of amplified mRNAs and RT-PCR primers are indicated to the right.

Given that ALKBH5 has the ability to "remove" m6A modification, we examined whether overexpression of ALKBH5 decreased m6A-methylation of HPV16 mRNAs. The m6A-IP results showed that the HPV16 mRNAs produced by pC97ELsLuc are methylated in transfected HeLa cells (**Figure 4.10A**). Furthermore, the amount of HPV16 mRNAs immunoprecipitated by m6A antibody was reduced by overexpression of ALKBH5 (**Figure 4.10B**).

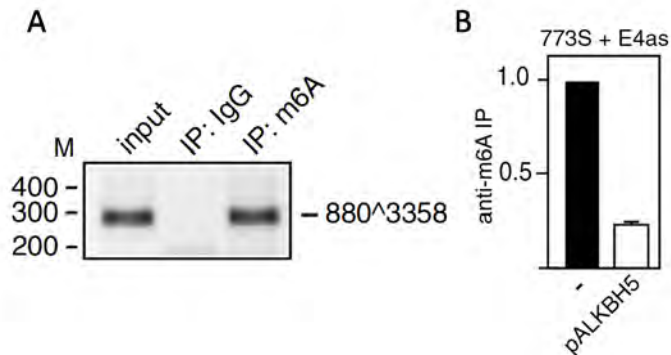


Figure 4.10 HPV16 mRNA contains m6A-methylation.

A. RT-PCR results of the CLIP assay showed that m6A antibody immunoprecipitated HPV16 mRNAs transcribed from pC97ELsLuc. **B.** qPCR results of the CLIP assay showed that co-transfection of pC97ELsLuc together with ALKBH5 expression plasmid reduced the m6A-methylated mRNA level of HPV16.

RT-PCR results of HPV16 late gene L1 revealed that METTL3 and ALKBH5 had opposite effect on L1 alternative splicing (**Figure 4.11C**). Overexpression of METTL3 enhanced inclusion of the central exon on the L1 mRNA and inhibited splicing directly from SD880 to SA5639, thereby promoting production of HPV16 L1 mRNAs. Overexpression of ALKBH5 altered HPV16 L1 mRNA splicing by causing skipping of the central exon of the L1 mRNAs and enhancing splicing from SD880 to SA5639, thereby promoting production of HPV16 L1i mRNAs. Also, we unexpectedly found that ALKBH5 inhibited E2 mRNA splicing and promoted production of intron retained E1 like METTL3 (**Figure 4.11A**). Furthermore, knockdown ALKBH5 in C33A2 cell line resulted in enhancement of E2 mRNA production and L1 mRNA production which were the opposite effects of ALKBH5 overexpression (**Figure 4.12**).

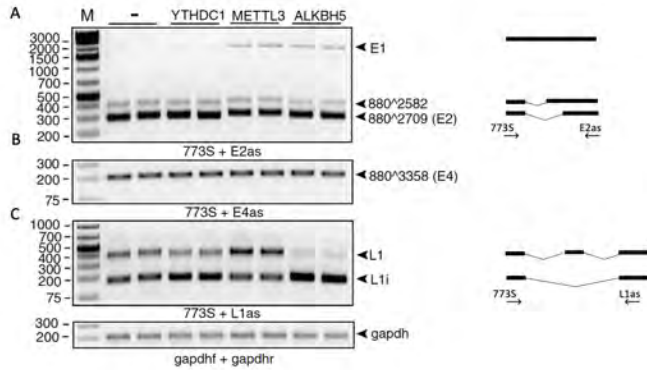


Figure 4.11 METTL3 and ALKBH5 alter HPV16 late L1 mRNA splicing

A. RT-PCR results showed that METTL3 and ALKBH5 inhibit E1/E2 mRNA splicing. Both METTL3 and ALKBH5 increase production of intron-retained E1. **B.** RT-PCR results showed that E4 mRNA level was not affected by m6A regulatory factors YTHDC1, METTL3 and ALKBH5. **C.** RT-PCR results showed that METTL3 and ALKBH5 had opposite effect on L1 alternative splicing.

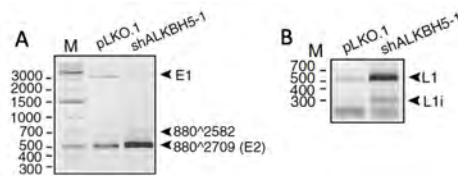


Figure 4.12 Knock down ALKBH5 enhanced the production of L1 mRNAs and splice E2 mRNAs.

RT-PCR results showed that Knock down of ALKBH5 in pBELsLuc stable cell line C33A2 enhanced L1 mRNAs and splice E2 mRNAs..

Since overexpression of the m6A “writers” and “erasers” METTL3 and ALKBH5 affected alternative splicing of various HPV16 mRNAs, one may speculate that a nuclear “reader” YTHDC1 might impact HPV16 mRNA splicing as well. RT-PCR results showed that overexpression of YTHDC1 altered HPV16 E6/E7 mRNA splicing only. YTHDC1 primarily enhanced intron retention to promote production of E6-encoding mRNAs (226^409) (**Figure 4.13A**). The results showed that YTHDC1 acted directly on the E6/E7 coding region. Finally, we analyzed the effect of YTHDC1 on the full-length HPV16 genome by using plasmid pHPV16AN which could release episomal form of the HPV16 genome with the help of Cre-loxP system (**Figure 1.15**). Results showed that YTHDC1 inhibited E6/E7 mRNA splicing and promoted retention of the E6-encoding intron at the expense of the spliced E7 mRNA (**Figure 4.13B**).

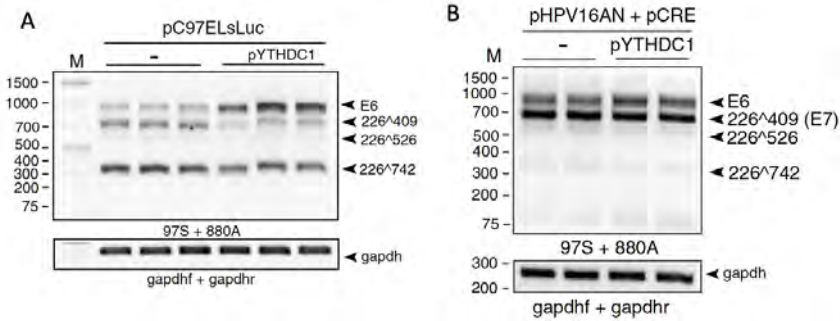


Figure 4.13 YTHDC1 promotes the intron retention of the E6 mRNAs.

RT-PCR results showed that intron-retained E6 mRNAs produced from subgenomic plasmid pC97ELsLuc and episomal HPV16 DNA generated from plasmid pHPV16AN and CRE plasmid were increased by YTHDC1.

To investigate if YTHDC1 binds directly to HPV16 mRNAs, we performed a CLIP assay on YTHDC1 on HeLa cells transfected with pC97ELsLuc with or without pYTHDC1. The results revealed that YTHDC1 and flag antibody could immunoprecipitate HPV16 mRNAs. Therefore, both endogenous and overexpressed YTHDC1 interacted with HPV16 mRNAs in the transfected HeLa cells (**Figure 4.14**).

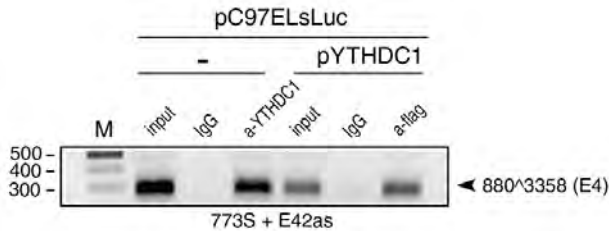


Figure 4.14 YTHDC1 interacts with HPV16 mRNAs.

RT-PCR results of CLIP assay showed that both endogenous and overexpressed YTHDC1 interacts with HPV16 mRNAs.

4.2.2 HPV16 mRNAs are m6A-methylated in HPV16-positive tonsillar cancer cell line HN26

In the end, we wished to determine if HPV16 mRNAs are m6A-methylated in HPV16-driven cancer cells. We therefore extracted RNA from the HPV16-positive tonsillar cancer cell line HN26 and subjected RNA to immunoprecipitation with m6A specific monoclonal antibody. RT-PCR results showed that mRNA encoding HPV16 E2, E4, E6 and E7 were immunoprecipitated by the m6A antibody which confirmed that HPV16 mRNAs are m6A-methylated in cancer cells which supported the idea that m6A-methylation plays an important role in the control of HPV16 gene expression (**Figure 4.15**).

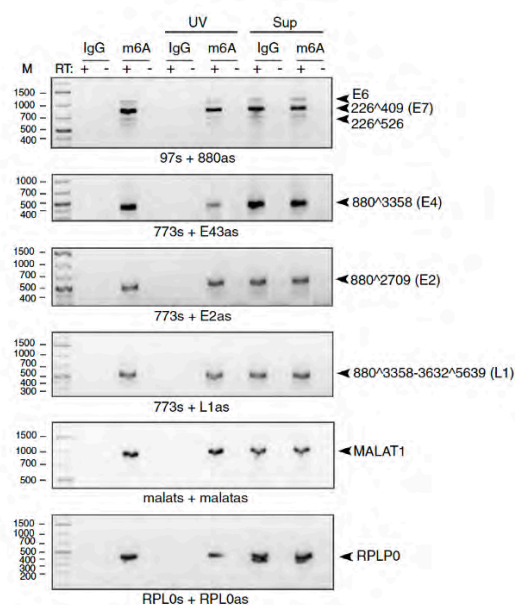


Figure 4.15 HPV16 mRNAs are m6A methylated in HN26

RT-PCR results of CLIP assay showed that mRNAs extracted from HPV16-positive HN26 cells were immunoprecipitated by m6A antibody.

To address the question of where the m6A sites are located on HPV16 mRNAs, meRIP-seq on two times oligo d(T) selected mRNA from HN26 cells was performed. Sequencing was performed on fragmented mRNA in 30-50nt size, immunoprecipitated by m6A antibody or IgG. The IP groups contain four biological replicates. Peaks represent potential m6A sites on cellular transcriptomes and HPV16 mRNAs. The analysis of cellular transcriptomes revealed m6A-IP over IgG peaks contained sequences similar to the m6A consensus motif, indicating that the MeRIP-seq strategy was successful (**Figure 4.16 A**). However, unexpectedly, m6A peaks of m6A IP replicates (black curves) overlapped with the IgG group (red curve) in HPV16 mRNA m6A analysis for unknown reasons (**Figure 4.16 B**). It was difficult to go further with interpretation and investigation. We concluded that the MeRIP experiment would have to be repeated, but lack of time prevented us from doing so within the timeframe of this thesis project.

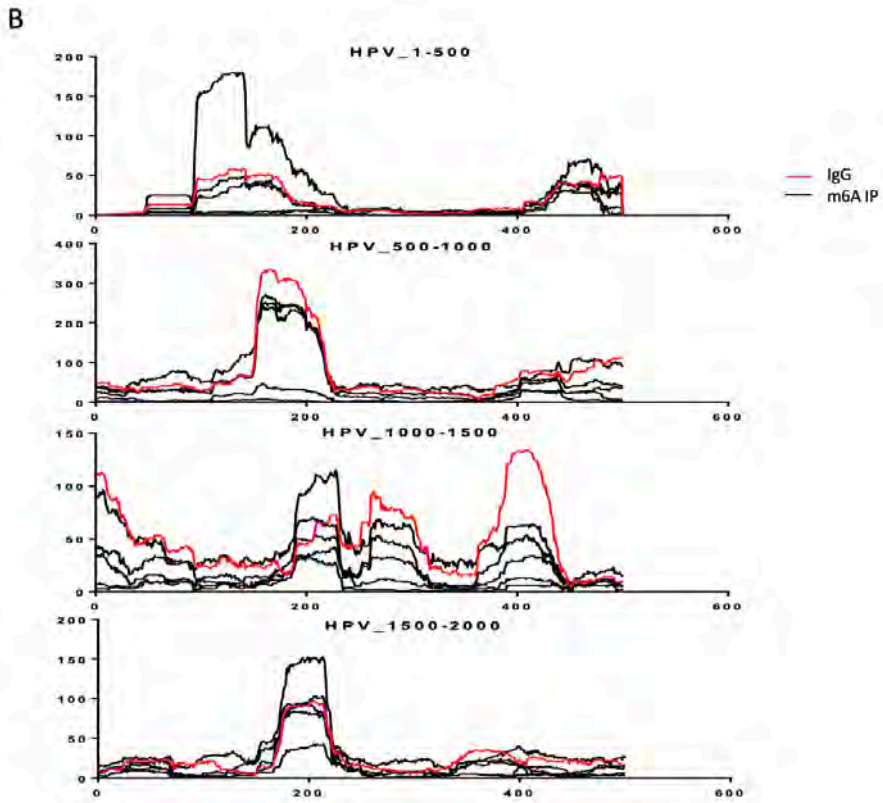
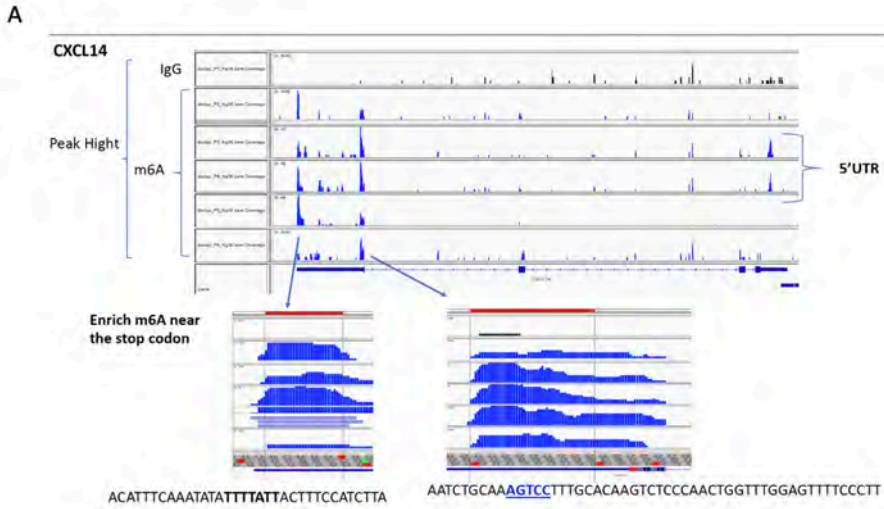


Figure 4.16 MeRIP-seq analysis of the m6A methylations of HPV16 mRNAs in HN26.

A. Peaks represent potential m6A sites on cellular transcriptomes B. Peaks represent potential m6A sites on HPV16 mRNAs in the first 2000 nucleotides.

Summary and Significance

- We found that overexpression of m6A eraser ALKBH5 promoted production of intron-retained HPV16 E6 mRNAs.
- ALKBH5 eraser and m6A writer METTL3 oppositely altered L1 mRNA splicing.
- METTL3 also induced production of intron-retained HPV16 E1 mRNAs.
- The nuclear m6A reader YTHDC1 promoted intron retention of E6 mRNAs over E7 mRNAs produced from the episomal HPV16 genome.
- We provide direct evidence that altering m6A methylation of HPV16 mRNAs affects HPV16 mRNA splicing.
- We demonstrate that HPV16 mRNA are m6A methylated in HPV16-positive cancer cell line HN26.

4.3 Identification of nucleotide substitutions in the 5'-end of HPV16 early mRNAs and in the non-transcribed long control region that affect E6 and E7 mRNA splicing

To investigate if sequence variation in HPV16 genome affects HPV16 mRNA splicing, we performed a sequence comparison on two HPV16 molecular clones HPV16AN and HPV16R. The alignment results revealed multiple sequence differences. The nucleotide substitutions in LCR at positions 7447, 7528, 7612, 12 and 13 or in the first 880 nucleotides at positions 131, 350 and 570 are particularly interesting since they have the potential to affect HPV16 oncogenes E6 and E7 transcription and post-transcription including alternative splicing. Firstly, we investigated if the sequence heterogeneity affected HPV16 E6/E7 mRNA splicing. Results showed that HPV16 E6/E7 mRNAs derived from HPV16AN are poorly spliced, whereas splicing pattern in HPV16R is typical (**Figure 4.17**).

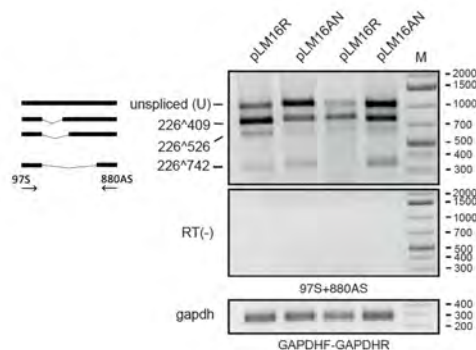


Figure 4.17 HPV16 E6/E7 mRNA splicing efficiency of subgenomic plasmids containing nucleotide substitutions in HPV16 LCR and 5' end of early mRNAs.

RT-PCR results of HeLa cells transfected with pLM16R or pLM16AN showed that splicing of E6/E7 mRNAs was affected by nucleotide substitution in LCR and 5' end of early mRNAs. Subgenomic plasmid pLM16AN derived from HPV16AN generated E6/E7 mRNAs with inefficient splicing.

Due to the surprising different splicing pattern caused by the eight nucleotide positions only, we narrow it down into two parts, five (7447, 7528, 7612, 12, and 13) in the non-transcribed part of the LCR and three (131, 350, and 570) in the transcribed HPV16 region downstream of the initiation site of the early promoter P97. We generated two hybrid plasmids with different combination of these two parts. Results revealed that sequences containing positions 131, 350 or 570 originated from HPV16R- conferred efficient splicing to E6/E7 mRNAs (**Figure 4.18**), leading to speculation that nucleotide positions 131, 350 or 570 overlapped cis-acting elements that controlled HPV16 early mRNA splicing.

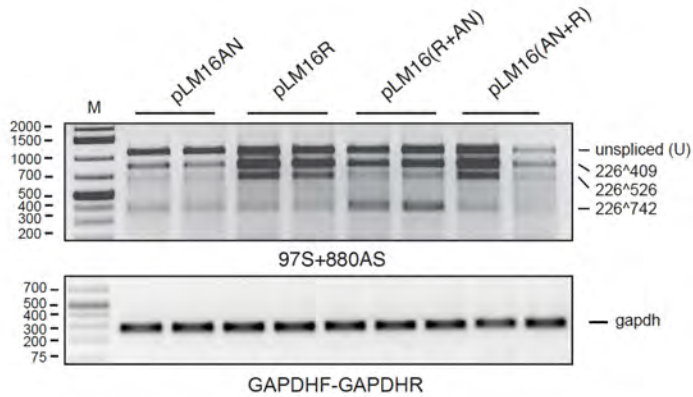


Figure 4.18 Sequences containing positions 131, 350 or 570 in 5' end of HPV16 early mRNA originated from HPV16R confer efficient splicing to E6/E7 mRNAs.

RT-PCR results showed that splicing efficiency of E6/E7 mRNAs generated from sequence hybridization subgenomic plasmids pLM16(R+AN) and pLM16(AN+R) was mainly determined by 5' end of HPV16 early mRNA sequence derived from HPV16R.

Another assumption is LCR originated from HPV16AN might contain *cis*-elements that potentially inhibit E6/E7 mRNA splicing. To verify this idea, plasmid pHPV16AN was used. pHPV16AN could generate HPV16 episomal with the help of Cre-loxP system. Results showed that transfection of pHPV16AN alone yielded poor splicing as we expected, but episomal form of HPV16 had efficient E6/E7 splicing ability (**Figure 4.19**). It appears that vector between loxP sites interfere the sequence in LCR that contribute to efficient splicing of E6/E7 mRNAs.

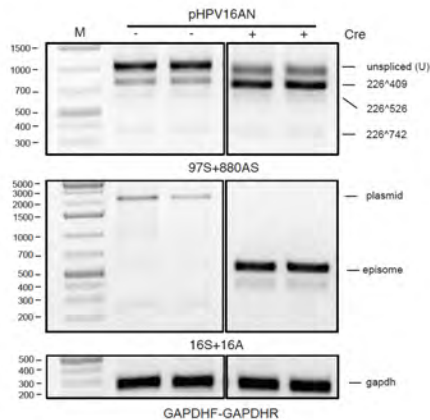


Figure 4.19 Intact LCR is needed for efficient E6/E7 mRNA splicing.

RT-PCR results showed that transfect genomic plasmid pHPV16AN with or without pCre results in different E6/E7 mRNA splicing efficiency. The E6/E7 mRNAs generated from episomal HPV16 genome released by Cre had high splicing efficiency.

The episomal results also indicated that reconstitution of the complete LCR also compensated for the poor splicing of HPV16 E6/E7 mRNAs. To confirm this finding, we generated HPV16 subgenomic expression plasmids driven by LCR sequences of various length. RT-PCR results showed that LCR sequences extended to L1 stop codon help efficient splicing, whereas splicing efficiency was diminished if LCR start from pAL and even more poorly with shorter LCR (**Figure 4.20**).



Figure 4.20 The E6/E7 mRNA splicing efficiency of subgenomic plasmids containing various length of HPV16 LCR sequence.

The RT-PCR results showed that LCR sequences extended to L1 stop codon help efficient splicing, whereas splicing efficiency was diminished if LCR start from pAL and even more poorly with shorter LCR.

Summary and Significance

- We found that sequences variation, which eight nucleotide positions only in HPV16 non-transcribed region of LCR and 5' end of the early mRNAs, affects HPV16 oncogenes E6/E7 mRNA splicing.
- Our results suggest that potential splicing regulatory RNA elements identified by the sequence variation in 5' end of the early mRNAs.
- HPV16 LCR 5' non-transcribed regions of the HPV16 genome may affect mRNA processing including transcription and splicing.

5 Discussion and Future Perspectives

The biological importance of alternative splicing has been shown in many normal physiological processes like cell differentiation and tissue maturation, therefore dysregulation of alternative splicing is the typical issue in human diseases, particularly, malignancies [186]. Alternative splicing is a tightly regulated process that depends on splicing regulatory elements (*cis*-elements) and splicing regulator factors (trans-factors). Alterations in splicing regulatory factors levels are common characteristics in cancers, for instance, hnRNP family members hnRNP A1 and hnRNP A2/B1, and SR protein member SRSF1 levels are frequently upregulated in cancer [187]. Altered expression of hnRNP D protein has been investigated in different cancers as well [188], which indicates hnRNP D may participate in carcinogenesis. HR HPV, primarily HPV16, is the causative agent of cervical cancer and head-and-neck cancer. The regulation of RNA processing, especially alternative pre-mRNA splicing, is particularly important for the regulation of HPV16 gene expression since all HPV16 genes are expressed from a plethora of alternatively spliced viral mRNAs.

In **Paper 1**, we demonstrated that hnRNP D inhibited HPV16 E1/E2 and E6/E7 mRNA splicing followed by the subsequent steps of HPV mRNA processing, including a potential increase in mRNA stability and nuclear export of intron-retained HPV16 E1 and E6 mRNA, but not the spliced E2 and E7 mRNA. Since E1 and E2 proteins regulate viral DNA replication and transcription, E6 and E7 oncoproteins are tightly functioning in cancer cell maintenance and malignant progression. Meanwhile, E6/E7 mRNAs and E1/E2 mRNAs are generated in a mutually exclusive way via alternative splicing. Thus, we believe hnRNP D contributes to controlling the HPV16 life cycle via regulating alternative splicing and other RNA processing events.

Except for *cis*-element and trans-factors, chemical modifications, especially m6A on mRNAs, also contributes to the regulation of mRNA alternative splicing. In **paper 2** we demonstrated m6A demethylase or “eraser” ALKBH5, methyltransferase or “writer” METTL3, and recognizer or “reader” YTHDC1 could affect HPV16 mRNA alternative splicing. Both ALKBH5 and YTHDC1 overexpression could inhibit E6/E7 mRNA splicing, while METTL3 did not affect it. One explanation for this result could be that E6/E7 mRNAs were highly m6A modulated, so that no more potential adenosine could be modified by METTL3, but ALKBH5 could remove m6A, thereby reducing the ability of a yet undisclosed

cellular “reader” and splicing factor to bind the mRNA and enhance splicing, while YTHDC1 may bind to m6A sites on the same mRNA and prevent the splicing factor to reach this position. In addition, we demonstrated that HPV16 mRNA are m6A-methylated and that m6A regulators METTL3, ALKBH5, and YTHDC1 affect HPV16 mRNA splicing. We concluded that m6A modification of HPV16 mRNAs contributes to the regulation of HPV16 gene expression at the level of mRNA splicing and that it may contribute to the pathogenic properties of HPV16. In the original plan, we further wanted to map the m6A sites on HPV16 transcripts, which are shown only in Project II of this thesis, not Paper 2, since the interpretation of sequencing results was difficult. Sequencing and preliminary alignment yielded peaks representing the potential m6A modified position of both HN26 cell transcripts and HPV16 episome transcripts, especially in the first 2000nt, including the 3' LCR and E6E7 coding region and a large part of the E1 coding region. We got eight peaks in this region by their rough positions A100, A450, A650, A1000, A1200, A1250, A1400 and A1700. Quality controls of the experiment were done, and the results of mapping the reads to the human genome showed that sequences under the peaks reflected the expected consensus motif for m6A modification (GGACU). The shallow analysis of human genome m6A enriched peaks showed a promising outcome of HPV16 analysis. However, a problem in the m6A mapping of the human genome was that the control IgG sample gave a lot of reads which surprised us since the amount of fragmented mRNA of IgG quality control data before sequencing was significantly lower than the m6A IP sample replicates. The unusual high reads did not affect human genome analysis very much but caused a big problem with the outcome of the HPV16 results. Even though HPV16 mapping results gave eight identity peaks between 0-2000nt of HPV16, the IgG group showed no difference with m6A IP replicates. The unusual high background of HPV16 mapping results made it difficult to interpret. One of the most possible reasons is sample concentration of the cDNA library of IgG was adjusted to the same as m6A IP replicates, resulting in a fake high reads outcome. Compared to HPV16, human genome m6A IP results contain millions more reads, and HPV16 m6A IP replicates only occupied a very small percentage, so HPV16 mapping was affected much more than the human genome. Another piece of information that made us believe the HPV16 mapping results is that m6A peaks are not random but have potential overlap with functional regulatory elements. A100 is close to the P97 promoter, A650 is located close to a splicing silencer, and A1000 is close to the AUAGUA motif, which is the hnRNP D binding site. In addition, A100, A450, and A650 correspond to the m6A estimated position from a prediction server named SRAMP. Taken together, m6A RNA-seq results gave us the perspective of m6A modulation sites on mRNA of HPV16 but not solid evidence. The RNA-Seq experiment needs to be repeated and results validated before the precise sites of m6A on HPV16 mRNA can be determined.

LCR is the non-coding region that contains most regulatory elements for HPV16 replication and transcription. LCR locates between the stop codon of L1 ORF and

the start codon of the E6 ORF. If the *cis*-elements and the corresponding trans factors in this region that regulate replication and transcription could potentially affect HPV16 mRNA splicing, or if *cis*-elements interact with splicing factor directly, have not been investigated. In **Paper 4**, we found that non-transcribed LCR sequences upstream of position 7470 could affect E6/E7 mRNA splicing, producing poorly spliced HPV16 mRNAs. Since this is preliminary data, we cannot speculate on the exact mechanism, but in combination with the phenomenon that transcription efficiency also decreased by the LCR sequence upstream of 7470, one possible explanation is that DNA sequences in the promoter-encoding region could potentially affect splicing by the recruitment of DNA-binding transcription factors that execute their function by recruiting RNA-binding proteins to the DNA-dependent RNA polymerase, thereby loading the polymerase with RNA-binding proteins that may swiftly interact with RNA elements on de novo synthesized mRNA. Furthermore, it has been shown that HPV18 mRNA splicing may be affected by long-range interactions of DNA-binding proteins that interact with the HPV genome [189-191].

hnRNP D show multifunction in **Paper 1**. In addition to the inhibitory effect on E6/E7 and E1/E2 mRNA splicing, resulting in increasing levels of E6 and E1 mRNAs, hnRNP D also accompanied these mRNAs to the cytoplasm. However, E6 and E1 protein levels were not enhanced, but decreased instead. In this study, we did not dig out the possible mechanism, but the results showed that hnRNP D and E6 E1 mRNA were binding to each other in the cytoplasm. These results indicate that translation of mRNAs to E6 and E1 proteins potentially needed other factors to replace hnRNP D. It is also possible that hnRNP D affects the degradation of E6 and E1 mRNA in cytoplasm, but more detailed experiments need to be done to uncover the fate of E6 and E1 mRNA.

Another intriguing result of the hnRNP D project is that deletion mutants D1 and D2 activated 226⁵²⁶ and 226⁴⁰⁹ splicing, respectively. This result indicated that the inhibitory effect of hnRNP D is located in N-terminal, A-rich domain and exon-2, while the RNA binding ability was exerted by both RRM1 and RRM2. Spliceosome binding was performed by the N-terminus and C-terminus of hnRNP D. Thus, D1 or D2 could bind target mRNAs and spliceosome factors but reduce splicing inhibitory function.

Taken together, we have established the role of hnRNP D in HPV16 gene expression, mRNA splicing and shown that hnRNP D levels affect E7 mRNA splicing and E7 oncogene production. In the future, one can investigate the role of hnRNP D in cancer progression, target for therapy for HPV16 infection and/or cancer, or hnRNP D as a biomarker for progression of HPV16 infections to high-grade lesion or cancer. Similarly, m6A methylation of HPV16 mRNAs is reversible, m6A added and removed by enzymes. Inhibiting these enzymes may alter HPV16 mRNA splicing and E6 and E7 oncogene expression. Maybe HPV16 infection of

HPV16 driven cancers would be sensitive to inhibitors of the m6A enzymes ALKBH5 or METTL3.

6 Acknowledgements

I appreciate everyone who contributed to my four years of study and life in Lund.

First, I would like to thank my supervisor **Prof. Stefan Schwartz**. Without your guidance, support, and encouragement, the projects and my graduation would not have been possible. You are always being there without any hesitation whenever I need help, making me brave when facing challenges and difficulties. You are also a great example to me that showing how hard-working, professional, open-minded, and optimistic are important for being both a scientist and supervisor. Last but not most important, you show me straightforward when making discussions not only in my career but also in life. I am very grateful to have been a PhD student of yours.

I would also like to thank my co-supervisors **Docent Ola Forslund** and **Dr. Naoko Kajitani**. Thanks for all the discussion, ideas, and kindness Ola offered to me, you are always so nice. Naoko is the best, I have learned so much from you, not only about academic knowledge but also about the persistence and seriousness of work. You are never miserly with time and energy for helping us of troubleshooting and resolve the confusion. I really admire your passion and patience and I believe you will be an excellent scientist and supervisor, hope you all the best in your career and life.

To my current colleagues in the lab, **Johanna Jönsson, Chengyu Hao, and Ye Zhu**. Thank you so much for spending time and staying with me together. For **Johanna**, you are always nice and generous for help. You are a great example of how an excellent Swedish girl makes a work-life balance through good organizing and deliberate planning. It is a pity that your exchange study to China was delayed, but you are always warmly welcome to China for both research and life. I will be there if you need any help. For **Chengyu**, thanks for showing me that never say “it is impossible”. It is not easy to take care of a kid and pursue PhD study, but you did it. You will also have your PhD dissertation soon, wish you good luck! For **Ye**, you are so warm and cheerful, here is more laughter when you were in the lab. Without your company, the difficulty would have been doubled when I in my paper under revision time.

My previous colleagues in the lab **Haoran Yu, Kersti Nilsson, Yunji Zheng, Shirin Shoja Chaghervand** and **Chengjun Wu**, I would also say thank you so much for all the help and support you gave to me. It is so lucky to have Haoran as both lab mate and friend. You are always clear mind about life. You gave me a lot

of support when I came to Sweden, solving practical problems, introducing new friends, etc. I will never forget the happy trip we spent together. Wish you have a happy life and work forever! Kersti, it is good to have you as a co-worker in one project. You show your cleverness, encourage and strength so well. Yunji (Tiger), thanks for your generous suggestions and help, you are the one who picked up me to the lab when I came to BMC the first time. I was also touched by your responsibility to work and family. Shirin, you are a good example of an independent and brave person who work abroad, wish you all well. Dr. Cheng jun Wu, you are Prof. Wu now! We did not have that much time overlap in lab, but it didn't an obstacle since you are so friendly. Thank you for your suggestion for my study in Lund, wish you have a great career in China.

Thank you my m6A project collaborators **Göran Akusjärvi**, **Shady Younis** and **Bo Segerman** and all the members who contribute to this project in Uppsala University. I believe we could do a great project together if I have more time. Like Stefan always said, this is life, but I still sincerely appreciate all your contributions and support. Big thanks to **Marianne Jansson** and all the members in HIV group for supporting us with lab spaces. I really enjoy the lab meetings about a new research field with all your discussion and critical thinking.

I would also thank my dear friends Ouyang Yuan, Yang Liu (BMC A11), Yang Liu (BMC C13), Rachel Cheong, Qimin Zhou, Julia ortis Sunyer, Wen Li, Siyuan Cang , Aodeng Gaowa, Qinyu Zhang, Chang Li, Hong Yan, Peng Huang, Ping Li and many many others. We spent a precious time together with fun, happiness, tears, and struggling... You all made my four years in Lund unforgettable.

In the end, I would like to conclude by thanking again all the people contributing to my study and the dissertation, particularly CHINA SCHOLARSHIP COUNCIL(CSC), that sponsored me for the past four years of study.

感谢我在中国的父母，你们总是尊重支持我的选择，你们平淡健康的生活是对我在瑞典这四年最好的支持，再多的言语都无法表达对你们的感恩，希望以后可以给你们更好的生活和陪伴。

在中国受到的教育和工作过的单位都让我成为了更加成熟全面的人，正是因为这些过往让我在瑞典的学习工作和生活顺利，坚定，满足。

最后感谢自己不负时光。

7 References

1. de Villiers EM, Fauquet C, Broker TR, Bernard HU, zur Hausen H: **Classification of papillomaviruses**. *Virology* 2004, **324**(1):17-27.
2. Kajitani N, Schwartz S: **The role of RNA-binding proteins in the processing of mRNAs produced by carcinogenic papillomaviruses**. *Semin Cancer Biol* 2022.
3. McLaughlin-Drubin ME: **Human papillomaviruses and non-melanoma skin cancer**. *Semin Oncol* 2015, **42**(2):284-290.
4. Tommasino M: **The biology of beta human papillomaviruses**. *Virus Res* 2017, **231**:128-138.
5. Andratschke M, Betz C, Leunig A: **[Laryngeal papillomatosis: etiology, diagnostics and therapy]**. *HNO* 2008, **56**(12):1190-1196.
6. Bray F, Ren JS, Masuyer E, Ferlay J: **Global estimates of cancer prevalence for 27 sites in the adult population in 2008**. *Int J Cancer* 2013, **132**(5):1133-1145.
7. Walboomers JM, Jacobs MV, Manos MM, Bosch FX, Kummer JA, Shah KV, Snijders PJ, Peto J, Meijer CJ, Muñoz N: **Human papillomavirus is a necessary cause of invasive cervical cancer worldwide**. *J Pathol* 1999, **189**(1):12-19.
8. Gillison ML, Chaturvedi AK, Lowy DR: **HPV prophylactic vaccines and the potential prevention of noncervical cancers in both men and women**. *Cancer* 2008, **113**(10 Suppl):3036-3046.
9. Chaturvedi AK, Engels EA, Pfeiffer RM, Hernandez BY, Xiao W, Kim E, Jiang B, Goodman MT, Sibug-Saber M, Cozen W *et al*: **Human papillomavirus and rising oropharyngeal cancer incidence in the United States**. *J Clin Oncol* 2011, **29**(32):4294-4301.
10. Forslund O, Sugiyama N, Wu C, Ravi N, Jin Y, Swoboda S, Andersson F, Bzhalava D, Hultin E, Paulsson K *et al*: **A novel human in vitro papillomavirus type 16 positive tonsil cancer cell line with high sensitivity to radiation and cisplatin**. *BMC Cancer* 2019, **19**(1):265.
11. Joura EA, Giuliano AR, Iversen OE, Bouchard C, Mao C, Mehlsen J, Moreira ED, Ngan Y, Petersen LK, Lazcano-Ponce E *et al*: **A 9-valent HPV vaccine against infection and intraepithelial neoplasia in women**. *N Engl J Med* 2015, **372**(8):711-723.

12. Beavis AL, Levinson KL: **Preventing Cervical Cancer in the United States: Barriers and Resolutions for HPV Vaccination.** *Front Oncol* 2016, **6**:19.
13. D.M. Knipe, Howley PM: **Fields Virology**, 6th ed. edn. Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins Health; 2013.
14. Bernard HU: **Regulatory elements in the viral genome.** *Virology* 2013, **445**(1-2):197-204.
15. Grassmann K, Rapp B, Maschek H, Petry KU, Iftner T: **Identification of a differentiation-inducible promoter in the E7 open reading frame of human papillomavirus type 16 (HPV-16) in raft cultures of a new cell line containing high copy numbers of episomal HPV-16 DNA.** *J Virol* 1996, **70**(4):2339-2349.
16. Schwartz S: **Papillomavirus transcripts and posttranscriptional regulation.** *Virology* 2013, **445**(1-2):187-196.
17. Sichero L, Sobrinho JS, Villa LL: **Identification of novel cellular transcription factors that regulate early promoters of human papillomavirus types 18 and 16.** *J Infect Dis* 2012, **206**(6):867-874.
18. Tan SH, Baker CC, Stünkel W, Bernard HU: **A transcriptional initiator overlaps with a conserved YY1 binding site in the long control region of human papillomavirus type 16.** *Virology* 2003, **305**(2):486-501.
19. Bletsas G, Zagouri F, Amoutzias GD, Nikolaidis M, Zografos E, Markoulatos P, Tsakogiannis D: **Genetic variability of the HPV16 early genes and LCR. Present and future perspectives.** *Expert Rev Mol Med* 2021, **23**:e19.
20. Cripe TP, Alderborn A, Anderson RD, Parkkinen S, Bergman P, Haugen TH, Pettersson U, Turek LP: **Transcriptional activation of the human papillomavirus-16 P97 promoter by an 88-nucleotide enhancer containing distinct cell-dependent and AP-1-responsive modules.** *New Biol* 1990, **2**(5):450-463.
21. Dong XP, Stubenrauch F, Beyer-Finkler E, Pfister H: **Prevalence of deletions of YY1-binding sites in episomal HPV 16 DNA from cervical cancers.** *Int J Cancer* 1994, **58**(6):803-808.
22. McLaughlin-Drubin ME, Münger K: **Oncogenic activities of human papillomaviruses.** *Virus Res* 2009, **143**(2):195-208.
23. Scarth JA, Patterson MR, Morgan EL, Macdonald A: **The human papillomavirus oncoproteins: a review of the host pathways targeted on the road to transformation.** *J Gen Virol* 2021, **102**(3).
24. Vande Pol SB, Klingelhutz AJ: **Papillomavirus E6 oncoproteins.** *Virology* 2013, **445**(1-2):115-137.
25. Roman A, Munger K: **The papillomavirus E7 proteins.** *Virology* 2013, **445**(1-2):138-168.
26. Dyson N, Howley PM, Münger K, Harlow E: **The human papilloma virus-16 E7 oncoprotein is able to bind to the retinoblastoma gene product.** *Science* 1989, **243**(4893):934-937.

27. Felsani A, Mileo AM, Paggi MG: **Retinoblastoma family proteins as key targets of the small DNA virus oncoproteins.** *Oncogene* 2006, **25**(38):5277-5285.
28. Dyson N: **The regulation of E2F by pRB-family proteins.** *Genes Dev* 1998, **12**(15):2245-2262.
29. Zerfass K, Schulze A, Spitkovsky D, Friedman V, Henglein B, Jansen-Dürr P: **Sequential activation of cyclin E and cyclin A gene expression by human papillomavirus type 16 E7 through sequences necessary for transformation.** *J Virol* 1995, **69**(10):6389-6399.
30. Lammens T, Li J, Leone G, De Veylder L: **Atypical E2Fs: new players in the E2F transcription factor family.** *Trends Cell Biol* 2009, **19**(3):111-118.
31. DeGregori J, Johnson DG: **Distinct and Overlapping Roles for E2F Family Members in Transcription, Proliferation and Apoptosis.** *Curr Mol Med* 2006, **6**(7):739-748.
32. He W, Staples D, Smith C, Fisher C: **Direct activation of cyclin-dependent kinase 2 by human papillomavirus E7.** *J Virol* 2003, **77**(19):10566-10574.
33. Nguyen CL, Münger K: **Direct association of the HPV16 E7 oncoprotein with cyclin A/CDK2 and cyclin E/CDK2 complexes.** *Virology* 2008, **380**(1):21-25.
34. Shin MK, Balsitis S, Brake T, Lambert PF: **Human papillomavirus E7 oncoprotein overrides the tumor suppressor activity of p21Cip1 in cervical carcinogenesis.** *Cancer Res* 2009, **69**(14):5656-5663.
35. Jones DL, Thompson DA, Münger K: **Destabilization of the RB tumor suppressor protein and stabilization of p53 contribute to HPV type 16 E7-induced apoptosis.** *Virology* 1997, **239**(1):97-107.
36. Eichten A, Rud DS, Grace M, Piboonniyom SO, Zacny V, Münger K: **Molecular pathways executing the "trophic sentinel" response in HPV-16 E7-expressing normal human diploid fibroblasts upon growth factor deprivation.** *Virology* 2004, **319**(1):81-93.
37. Scheffner M, Werness BA, Huibregtse JM, Levine AJ, Howley PM: **The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53.** *Cell* 1990, **63**(6):1129-1136.
38. Scheffner M, Huibregtse JM, Vierstra RD, Howley PM: **The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53.** *Cell* 1993, **75**(3):495-505.
39. Lechner MS, Laimins LA: **Inhibition of p53 DNA binding by human papillomavirus E6 proteins.** *J Virol* 1994, **68**(7):4262-4273.
40. Li X, Coffino P: **High-risk human papillomavirus E6 protein has two distinct binding sites within p53, of which only one determines degradation.** *J Virol* 1996, **70**(7):4509-4516.
41. Patel D, Huang SM, Baglia LA, McCance DJ: **The E6 protein of human papillomavirus type 16 binds to and inhibits co-activation by CBP and p300.** *EMBO J* 1999, **18**(18):5061-5072.

42. Zimmermann H, Degenkolbe R, Bernard HU, O'Connor MJ: **The human papillomavirus type 16 E6 oncoprotein can down-regulate p53 activity by targeting the transcriptional coactivator CBP/p300.** *J Virol* 1999, **73**(8):6209-6219.
43. Delury CP, Marsh EK, James CD, Boon SS, Banks L, Knight GL, Roberts S: **The role of protein kinase A regulation of the E6 PDZ-binding domain during the differentiation-dependent life cycle of human papillomavirus type 18.** *J Virol* 2013, **87**(17):9463-9472.
44. Akagi K, Li J, Broutian TR, Padilla-Nash H, Xiao W, Jiang B, Rocco JW, Teknos TN, Kumar B, Wangsa D *et al*: **Genome-wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability.** *Genome Res* 2014, **24**(2):185-199.
45. Bergvall M, Melendy T, Archambault J: **The E1 proteins.** *Virology* 2013, **445**(1-2):35-56.
46. Hughes FJ, Romanos MA: **E1 protein of human papillomavirus is a DNA helicase/ATPase.** *Nucleic Acids Res* 1993, **21**(25):5817-5823.
47. Chen G, Stenlund A: **The E1 initiator recognizes multiple overlapping sites in the papillomavirus origin of DNA replication.** *J Virol* 2001, **75**(1):292-302.
48. Titolo S, Brault K, Majewski J, White PW, Archambault J: **Characterization of the minimal DNA binding domain of the human papillomavirus e1 helicase: fluorescence anisotropy studies and characterization of a dimerization-defective mutant protein.** *J Virol* 2003, **77**(9):5178-5191.
49. McBride AA: **The papillomavirus E2 proteins.** *Virology* 2013, **445**(1-2):57-79.
50. Ustav M, Stenlund A: **Transient replication of BPV-1 requires two viral polypeptides encoded by the E1 and E2 open reading frames.** *EMBO J* 1991, **10**(2):449-457.
51. Berg M, Stenlund A: **Functional interactions between papillomavirus E1 and E2 proteins.** *J Virol* 1997, **71**(5):3853-3863.
52. Sanders CM, Stenlund A: **Recruitment and loading of the E1 initiator protein: an ATP-dependent process catalysed by a transcription factor.** *EMBO J* 1998, **17**(23):7044-7055.
53. Parish JL, Kowalczyk A, Chen HT, Roeder GE, Sessions R, Buckle M, Gaston K: **E2 proteins from high- and low-risk human papillomavirus types differ in their ability to bind p53 and induce apoptotic cell death.** *J Virol* 2006, **80**(9):4580-4590.
54. Blachon S, Bellanger S, Demeret C, Thierry F: **Nucleo-cytoplasmic shuttling of high risk human Papillomavirus E2 proteins induces apoptosis.** *J Biol Chem* 2005, **280**(43):36088-36098.
55. Dreer M, van de Poel S, Stubenrauch F: **Control of viral replication and transcription by the papillomavirus E8^{E2} protein.** *Virus Res* 2017, **231**:96-102.

56. Johansson C, Somberg M, Li X, Backström Winquist E, Fay J, Ryan F, Pim D, Banks L, Schwartz S: **HPV-16 E2 contributes to induction of HPV-16 late gene expression by inhibiting early polyadenylation.** *EMBO J* 2012, **31**(14):3212-3227.
57. Zheng Y, Cui X, Nilsson K, Yu H, Gong L, Wu C, Schwartz S: **Efficient production of HPV16 E2 protein from HPV16 late mRNAs spliced from SD880 to SA2709.** *Virus Res* 2020, **285**:198004.
58. Doorbar J, Parton A, Hartley K, Banks L, Crook T, Stanley M, Crawford L: **Detection of novel splicing patterns in a HPV16-containing keratinocyte cell line.** *Virology* 1990, **178**(1):254-262.
59. Doorbar J: **The E4 protein; structure, function and patterns of expression.** *Virology* 2013, **445**(1-2):80-98.
60. Doorbar J, Cubie H: **Molecular basis for advances in cervical screening.** *Mol Diagn* 2005, **9**(3):129-142.
61. Griffin H, Wu Z, Marnane R, Dewar V, Molijn A, Quint W, Van Hoof C, Struyf F, Colau B, Jenkins D *et al*: **E4 antibodies facilitate detection and type-assignment of active HPV infection in cervical disease.** *PLoS One* 2012, **7**(12):e49974.
62. DiMaio D, Petti LM: **The E5 proteins.** *Virology* 2013, **445**(1-2):99-114.
63. Nilsson K, Norberg C, Mossberg AK, Schwartz S: **HPV16 E5 is produced from an HPV16 early mRNA spliced from SD226 to SA3358.** *Virus Res* 2018, **244**:128-136.
64. Chen XS, Garcea RL, Goldberg I, Casini G, Harrison SC: **Structure of small virus-like particles assembled from the L1 protein of human papillomavirus 16.** *Mol Cell* 2000, **5**(3):557-567.
65. Doorbar J: **The papillomavirus life cycle.** *J Clin Virol* 2005, **32** Suppl 1:S7-15.
66. Chow LT, Broker TR, Steinberg BM: **The natural history of human papillomavirus infections of the mucosal epithelia.** *APMIS* 2010, **118**(6-7):422-449.
67. Roden RBS, Stern PL: **Opportunities and challenges for human papillomavirus vaccination in cancer.** *Nat Rev Cancer* 2018, **18**(4):240-254.
68. Kirnbauer R, Booy F, Cheng N, Lowy DR, Schiller JT: **Papillomavirus L1 major capsid protein self-assembles into virus-like particles that are highly immunogenic.** *Proc Natl Acad Sci U S A* 1992, **89**(24):12180-12184.
69. Trus BL, Roden RB, Greenstone HL, Vrhel M, Schiller JT, Booy FP: **Novel structural features of bovine papillomavirus capsid revealed by a three-dimensional reconstruction to 9 Å resolution.** *Nat Struct Biol* 1997, **4**(5):413-420.
70. Doorbar J, Gallimore PH: **Identification of proteins encoded by the L1 and L2 open reading frames of human papillomavirus 1a.** *J Virol* 1987, **61**(9):2793-2799.

71. Finnen RL, Erickson KD, Chen XS, Garcea RL: **Interactions between papillomavirus L1 and L2 capsid proteins.** *J Virol* 2003, **77**(8):4818-4826.
72. Buck CB, Cheng N, Thompson CD, Lowy DR, Steven AC, Schiller JT, Trus BL: **Arrangement of L2 within the papillomavirus capsid.** *J Virol* 2008, **82**(11):5190-5197.
73. Kirnbauer R, Taub J, Greenstone H, Roden R, Dürst M, Gissmann L, Lowy DR, Schiller JT: **Efficient self-assembly of human papillomavirus type 16 L1 and L1-L2 into virus-like particles.** *J Virol* 1993, **67**(12):6929-6936.
74. Kondo K, Ishii Y, Ochi H, Matsumoto T, Yoshikawa H, Kanda T: **Neutralization of HPV16, 18, 31, and 58 pseudovirions with antisera induced by immunizing rabbits with synthetic peptides representing segments of the HPV16 minor capsid protein L2 surface region.** *Virology* 2007, **358**(2):266-272.
75. Herfs M, Yamamoto Y, Laury A, Wang X, Nucci MR, McLaughlin-Drubin ME, Mürger K, Feldman S, McKeon FD, Xian W *et al*: **A discrete population of squamocolumnar junction cells implicated in the pathogenesis of cervical cancer.** *Proc Natl Acad Sci U S A* 2012, **109**(26):10516-10521.
76. Raff AB, Woodham AW, Raff LM, Skeate JG, Yan L, Da Silva DM, Schelhaas M, Kast WM: **The evolving field of human papillomavirus receptor research: a review of binding and entry.** *J Virol* 2013, **87**(11):6062-6072.
77. Schiller JT, Day PM, Kines RC: **Current understanding of the mechanism of HPV infection.** *Gynecol Oncol* 2010, **118**(1 Suppl):S12-17.
78. Kines RC, Thompson CD, Lowy DR, Schiller JT, Day PM: **The initial steps leading to papillomavirus infection occur on the basement membrane prior to cell surface binding.** *Proc Natl Acad Sci U S A* 2009, **106**(48):20458-20463.
79. Surviladze Z, Dziduszko A, Ozbun MA: **Essential roles for soluble virion-associated heparan sulfonated proteoglycans and growth factors in human papillomavirus infections.** *PLoS Pathog* 2012, **8**(2):e1002519.
80. Abban CY, Meneses PI: **Usage of heparan sulfate, integrins, and FAK in HPV16 infection.** *Virology* 2010, **403**(1):1-16.
81. Culp TD, Budgeon LR, Marinkovich MP, Meneguzzi G, Christensen ND: **Keratinocyte-secreted laminin 5 can function as a transient receptor for human papillomaviruses by binding virions and transferring them to adjacent cells.** *J Virol* 2006, **80**(18):8940-8950.
82. DiGiuseppe S, Bienkowska-Haba M, Guion LG, Sapp M: **Cruising the cellular highways: How human papillomavirus travels from the surface to the nucleus.** *Virus Res* 2017, **231**:1-9.

83. Zhang W, Kazakov T, Popa A, DiMaio D: **Vesicular trafficking of incoming human papillomavirus 16 to the Golgi apparatus and endoplasmic reticulum requires γ -secretase activity.** *mBio* 2014, **5**(5):e01777-01714.
84. DiGiuseppe S, Luszczek W, Keiffer TR, Bienkowska-Haba M, Guion LG, Sapp MJ: **Incoming human papillomavirus type 16 genome resides in a vesicular compartment throughout mitosis.** *Proc Natl Acad Sci U S A* 2016, **113**(22):6289-6294.
85. Maglennon GA, McIntosh P, Doorbar J: **Persistence of viral DNA in the epithelial basal layer suggests a model for papillomavirus latency following immune regression.** *Virology* 2011, **414**(2):153-163.
86. Westrich JA, Warren CJ, Pyeon D: **Evasion of host immune defenses by human papillomavirus.** *Virus Res* 2017, **231**:21-33.
87. Sanders CM, Stenlund A: **Transcription factor-dependent loading of the E1 initiator reveals modular assembly of the papillomavirus origin melting complex.** *J Biol Chem* 2000, **275**(5):3522-3534.
88. Smith JA, Haberstroh FS, White EA, Livingston DM, DeCaprio JA, Howley PM: **SMCX and components of the TIP60 complex contribute to E2 regulation of the HPV E6/E7 promoter.** *Virology* 2014, **468-470**:311-321.
89. You J, Croyle JL, Nishimura A, Ozato K, Howley PM: **Interaction of the bovine papillomavirus E2 protein with Brd4 tethers the viral DNA to host mitotic chromosomes.** *Cell* 2004, **117**(3):349-360.
90. Martinez-Zapien D, Ruiz FX, Poirson J, Mitschler A, Ramirez J, Forster A, Cousido-Siah A, Masson M, Vande Pol S, Podjarny A *et al*: **Structure of the E6/E6AP/p53 complex required for HPV-mediated degradation of p53.** *Nature* 2016, **529**(7587):541-545.
91. Mantovani F, Banks L: **The human papillomavirus E6 protein and its contribution to malignant progression.** *Oncogene* 2001, **20**(54):7874-7887.
92. James MA, Lee JH, Klingelhutz AJ: **Human papillomavirus type 16 E6 activates NF-kappaB, induces cIAP-2 expression, and protects against apoptosis in a PDZ binding motif-dependent manner.** *J Virol* 2006, **80**(11):5301-5307.
93. White EA, Kramer RE, Tan MJ, Hayes SD, Harper JW, Howley PM: **Comprehensive analysis of host cellular interactions with human papillomavirus E6 proteins identifies new E6 binding partners and reflects viral diversity.** *J Virol* 2012, **86**(24):13174-13186.
94. Davy CE, Jackson DJ, Wang Q, Raj K, Masterson PJ, Fenner NF, Southern S, Cuthill S, Millar JB, Doorbar J: **Identification of a G(2) arrest domain in the E1 wedge E4 protein of human papillomavirus type 16.** *J Virol* 2002, **76**(19):9806-9818.
95. Davy CE, Jackson DJ, Raj K, Peh WL, Southern SA, Das P, Sorathia R, Laskey P, Middleton K, Nakahara T *et al*: **Human papillomavirus type**

- 16 E1 E4-induced G2 arrest is associated with cytoplasmic retention of active Cdk1/cyclin B1 complexes.** *J Virol* 2005, **79**(7):3998-4011.
96. Wang Q, Griffin H, Southern S, Jackson D, Martin A, McIntosh P, Davy C, Masterson PJ, Walker PA, Laskey P *et al*: **Functional analysis of the human papillomavirus type 16 E1=E4 protein provides a mechanism for in vivo and in vitro keratin filament reorganization.** *J Virol* 2004, **78**(2):821-833.
97. Ashrafi GH, Haghshenas M, Marchetti B, Campo MS: **E5 protein of human papillomavirus 16 downregulates HLA class I and interacts with the heavy chain via its first hydrophobic domain.** *Int J Cancer* 2006, **119**(9):2105-2112.
98. Kadaja M, Silla T, Ustav E, Ustav M: **Papillomavirus DNA replication - from initiation to genomic instability.** *Virology* 2009, **384**(2):360-368.
99. Middleton K, Peh W, Southern S, Griffin H, Sotlar K, Nakahara T, El-Sherif A, Morris L, Seth R, Hibma M *et al*: **Organization of human papillomavirus productive cycle during neoplastic progression provides a basis for selection of diagnostic markers.** *J Virol* 2003, **77**(19):10186-10201.
100. Becker KA, Florin L, Sapp C, Sapp M: **Dissection of human papillomavirus type 33 L2 domains involved in nuclear domains (ND) 10 homing and reorganization.** *Virology* 2003, **314**(1):161-167.
101. Merle E, Rose RC, LeRoux L, Moroianu J: **Nuclear import of HPV11 L1 capsid protein is mediated by karyopherin alpha2beta1 heterodimers.** *J Cell Biochem* 1999, **74**(4):628-637.
102. Darshan MS, Lucchi J, Harding E, Moroianu J: **The L2 minor capsid protein of human papillomavirus type 16 interacts with a network of nuclear import receptors.** *J Virol* 2004, **78**(22):12179-12188.
103. Florin L, Sapp C, Streeck RE, Sapp M: **Assembly and translocation of papillomavirus capsid proteins.** *J Virol* 2002, **76**(19):10009-10014.
104. Ciccia A, Elledge SJ: **The DNA damage response: making it safe to play with knives.** *Mol Cell* 2010, **40**(2):179-204.
105. Zeman MK, Cimprich KA: **Causes and consequences of replication stress.** *Nat Cell Biol* 2014, **16**(1):2-9.
106. Anacker DC, Moody CA: **Modulation of the DNA damage response during the life cycle of human papillomaviruses.** *Virus Res* 2017, **231**:41-49.
107. Nilsson K, Wu C, Kajitani N, Yu H, Tsimitsirakis E, Gong L, Winquist EB, Glahder J, Ekblad L, Wennerberg J *et al*: **The DNA damage response activates HPV16 late gene expression at the level of RNA processing.** *Nucleic Acids Res* 2018, **46**(10):5029-5049.
108. Stanley M: **Pathology and epidemiology of HPV infection in females.** *Gynecol Oncol* 2010, **117**(2 Suppl):S5-10.
109. Thierry F, Yaniv M: **The BPV1-E2 trans-acting protein can be either an activator or a repressor of the HPV18 regulatory region.** *EMBO J* 1987, **6**(11):3391-3397.

110. Bernard BA, Bailly C, Lenoir MC, Darmon M, Thierry F, Yaniv M: **The human papillomavirus type 18 (HPV18) E2 gene product is a repressor of the HPV18 regulatory region in human keratinocytes.** *J Virol* 1989, **63**(10):4317-4324.
111. Jeon S, Lambert PF: **Integration of human papillomavirus type 16 DNA into the human genome leads to increased stability of E6 and E7 mRNAs: implications for cervical carcinogenesis.** *Proc Natl Acad Sci U S A* 1995, **92**(5):1654-1658.
112. Zheng ZM, Baker CC: **Papillomavirus genome structure, expression, and post-transcriptional regulation.** *Front Biosci* 2006, **11**:2286-2302.
113. Kajitani N, Schwartz S: **Role of Viral Ribonucleoproteins in Human Papillomavirus Type 16 Gene Expression.** *Viruses* 2020, **12**(10).
114. Smotkin D, Wettstein FO: **Transcription of human papillomavirus type 16 early genes in a cervical cancer and a cancer-derived cell line and identification of the E7 protein.** *Proc Natl Acad Sci U S A* 1986, **83**(13):4680-4684.
115. Romanczuk H, Thierry F, Howley PM: **Mutational analysis of cis elements involved in E2 modulation of human papillomavirus type 16 P97 and type 18 P105 promoters.** *J Virol* 1990, **64**(6):2849-2859.
116. Sousa R, Dostatni N, Yaniv M: **Control of papillomavirus gene expression.** *Biochim Biophys Acta* 1990, **1032**(1):19-37.
117. Ribeiro AL, Caodaglio AS, Sichero L: **Regulation of HPV transcription.** *Clinics (Sao Paulo)* 2018, **73**(suppl 1):e486s.
118. Yan C, Wan R, Shi Y: **Molecular Mechanisms of pre-mRNA Splicing through Structural Biology of the Spliceosome.** *Cold Spring Harb Perspect Biol* 2019, **11**(1).
119. Roca X, Krainer AR, Eperon IC: **Pick one, but be quick: 5' splice sites and the problems of too many choices.** *Genes Dev* 2013, **27**(2):129-144.
120. Brody E, Abelson J: **The "spliceosome": yeast pre-messenger RNA associates with a 40S complex in a splicing-dependent reaction.** *Science* 1985, **228**(4702):963-967.
121. Friendewey D, Keller W: **Stepwise assembly of a pre-mRNA splicing complex requires U-snRNPs and specific intron sequences.** *Cell* 1985, **42**(1):355-367.
122. Wahl MC, Will CL, Lührmann R: **The spliceosome: design principles of a dynamic RNP machine.** *Cell* 2009, **136**(4):701-718.
123. Will CL, Lührmann R: **Spliceosome structure and function.** *Cold Spring Harb Perspect Biol* 2011, **3**(7).
124. Shi Y: **Mechanistic insights into precursor messenger RNA splicing by the spliceosome.** *Nat Rev Mol Cell Biol* 2017, **18**(11):655-670.
125. Sickmier EA, Frato KE, Shen H, Paranawithana SR, Green MR, Kielkopf CL: **Structural basis for polypyrimidine tract recognition by the essential pre-mRNA splicing factor U2AF65.** *Mol Cell* 2006, **23**(1):49-59.

126. Kielkopf CL, Rodionova NA, Green MR, Burley SK: **A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer.** *Cell* 2001, **106**(5):595-605.
127. Zhang Y, Rahmani RS, Yang X, Chen J, Shi T: **Integrative expression network analysis of microRNA and gene isoforms in sacred lotus.** *BMC Genomics* 2020, **21**(1):429.
128. Zheng ZM, Tao M, Yamanegi K, Bodaghi S, Xiao W: **Splicing of a cap-proximal human Papillomavirus 16 E6E7 intron promotes E7 expression, but can be restrained by distance of the intron from its RNA 5' cap.** *J Mol Biol* 2004, **337**(5):1091-1108.
129. Zheng Y, Jönsson J, Hao C, Shoja Chaghervand S, Cui X, Kajitani N, Gong L, Wu C, Schwartz S: **Heterogeneous Nuclear Ribonucleoprotein A1 (hnRNP A1) and hnRNP A2 Inhibit Splicing to Human Papillomavirus 16 Splice Site SA409 through a UAG-Containing Sequence in the E7 Coding Region.** *J Virol* 2020, **94**(20).
130. Tang S, Tao M, McCoy JP, Zheng ZM: **The E7 oncoprotein is translated from spliced E6*I transcripts in high-risk human papillomavirus type 16- or type 18-positive cervical cancer cell lines via translation reinitiation.** *J Virol* 2006, **80**(9):4249-4263.
131. Stacey SN, Jordan D, Williamson AJ, Brown M, Coote JH, Arrand JR: **Leaky scanning is the predominant mechanism for translation of human papillomavirus type 16 E7 oncoprotein from E6/E7 bicistronic mRNA.** *J Virol* 2000, **74**(16):7284-7297.
132. Ajiro M, Zheng ZM: **E6^ΔE7, a novel splice isoform protein of human papillomavirus 16, stabilizes viral E6 and E7 oncoproteins via HSP90 and GRP78.** *mBio* 2015, **6**(1):e02068-02014.
133. Pim D, Tomaic V, Banks L: **The human papillomavirus (HPV) E6* proteins from high-risk, mucosal HPVs can direct degradation of cellular proteins in the absence of full-length E6 protein.** *J Virol* 2009, **83**(19):9863-9874.
134. Smotkin D, Prokoph H, Wettstein FO: **Oncogenic and nononcogenic human genital papillomaviruses generate the E7 mRNA by different mechanisms.** *J Virol* 1989, **63**(3):1441-1447.
135. Cui X, Hao C, Gong L, Kajitani N, Schwartz S: **HnRNP D activates production of HPV16 E1 and E6 mRNAs by promoting intron retention.** *Nucleic Acids Res* 2022, **50**(5):2782-2806.
136. Van Doorslaer K, Tan Q, Xirasagar S, Bandaru S, Gopalan V, Mohamoud Y, Huyen Y, McBride AA: **The Papillomavirus Episteme: a central resource for papillomavirus sequence data and analysis.** *Nucleic Acids Res* 2013, **41**(Database issue):D571-578.
137. Kajitani N, Glahder J, Wu C, Yu H, Nilsson K, Schwartz S: **hnRNP L controls HPV16 RNA polyadenylation and splicing in an Akt kinase-dependent manner.** *Nucleic Acids Res* 2017, **45**(16):9654-9678.
138. Richard P, Manley JL: **Transcription termination by nuclear RNA polymerases.** *Genes Dev* 2009, **23**(11):1247-1269.

139. Hu J, Lutz CS, Wilusz J, Tian B: **Bioinformatic identification of candidate cis-regulatory elements involved in human mRNA polyadenylation.** *RNA* 2005, **11**(10):1485-1493.
140. Mandel CR, Kaneko S, Zhang H, Gebauer D, Vethantham V, Manley JL, Tong L: **Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease.** *Nature* 2006, **444**(7121):953-956.
141. Ryan K, Calvo O, Manley JL: **Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease.** *RNA* 2004, **10**(4):565-573.
142. Martin G, Doublé S, Keller W: **Determinants of substrate specificity in RNA-dependent nucleotidyl transferases.** *Biochim Biophys Acta* 2008, **1779**(4):206-216.
143. Kühn U, Wahle E: **Structure and function of poly(A) binding proteins.** *Biochim Biophys Acta* 2004, **1678**(2-3):67-84.
144. Eckmann CR, Rammelt C, Wahle E: **Control of poly(A) tail length.** *Wiley Interdiscip Rev RNA* 2011, **2**(3):348-361.
145. Gruber AJ, Zavolan M: **Alternative cleavage and polyadenylation in health and disease.** *Nat Rev Genet* 2019, **20**(10):599-614.
146. Wu C, Kajitani N, Schwartz S: **Splicing and Polyadenylation of Human Papillomavirus Type 16 mRNAs.** *Int J Mol Sci* 2017, **18**(2).
147. Oberg D, Fay J, Lambkin H, Schwartz S: **A downstream polyadenylation element in human papillomavirus type 16 L2 encodes multiple GGG motifs and interacts with hnRNP H.** *J Virol* 2005, **79**(14):9254-9269.
148. Oberg D, Collier B, Zhao X, Schwartz S: **Mutational inactivation of two distinct negative RNA elements in the human papillomavirus type 16 L2 coding region induces production of high levels of L2 in human cells.** *J Virol* 2003, **77**(21):11674-11684.
149. Dhanjal S, Kajitani N, Glahder J, Mossberg AK, Johansson C, Schwartz S: **Heterogeneous Nuclear Ribonucleoprotein C Proteins Interact with the Human Papillomavirus Type 16 (HPV16) Early 3'-Untranslated Region and Alleviate Suppression of HPV16 Late L1 mRNA Splicing.** *J Biol Chem* 2015, **290**(21):13354-13371.
150. Kaufmann I, Martin G, Friedlein A, Langen H, Keller W: **Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase.** *EMBO J* 2004, **23**(3):616-626.
151. Kennedy IM, Haddow JK, Clements JB: **Analysis of human papillomavirus type 16 late mRNA 3' processing signals in vitro and in vivo.** *J Virol* 1990, **64**(4):1825-1829.
152. Goracznik R, Gunderson SI: **The regulatory element in the 3'-untranslated region of human papillomavirus 16 inhibits expression by binding CUG-binding protein 1.** *J Biol Chem* 2008, **283**(4):2286-2296.

153. Cumming SA, Chuen-Im T, Zhang J, Graham SV: **The RNA stability regulator HuR regulates L1 protein expression in vivo in differentiating cervical epithelial cells.** *Virology* 2009, **383**(1):142-149.
154. Cumming SA, McPhillips MG, Veerapraditsin T, Milligan SG, Graham SV: **Activity of the human papillomavirus type 16 late negative regulatory element is partly due to four weak consensus 5' splice sites that bind a U1 snRNP-like complex.** *J Virol* 2003, **77**(9):5167-5177.
155. Yue Y, Liu J, He C: **RNA N6-methyladenosine methylation in post-transcriptional gene expression regulation.** *Genes Dev* 2015, **29**(13):1343-1355.
156. Fu Y, Dominissini D, Rechavi G, He C: **Gene expression regulation mediated through reversible m⁶A RNA methylation.** *Nat Rev Genet* 2014, **15**(5):293-306.
157. Xiao W, Adhikari S, Dahal U, Chen YS, Hao YJ, Sun BF, Sun HY, Li A, Ping XL, Lai WY *et al*: **Nuclear m(6)A Reader YTHDC1 Regulates mRNA Splicing.** *Mol Cell* 2016, **61**(4):507-519.
158. Liu N, Zhou KI, Parisien M, Dai Q, Diatchenko L, Pan T: **N6-methyladenosine alters RNA structure to regulate binding of a low-complexity protein.** *Nucleic Acids Res* 2017, **45**(10):6051-6063.
159. Liu N, Dai Q, Zheng G, He C, Parisien M, Pan T: **N(6)-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions.** *Nature* 2015, **518**(7540):560-564.
160. Kane SE, Beemon K: **Precise localization of m6A in Rous sarcoma virus RNA reveals clustering of methylation sites: implications for RNA processing.** *Mol Cell Biol* 1985, **5**(9):2298-2306.
161. Finkel D, Groner Y: **Methylations of adenosine residues (m6A) in pre-mRNA are important for formation of late simian virus 40 mRNAs.** *Virology* 1983, **131**(2):409-425.
162. Krug RM, Morgan MA, Shatkin AJ: **Influenza viral mRNA contains internal N6-methyladenosine and 5'-terminal 7-methylguanosine in cap structures.** *J Virol* 1976, **20**(1):45-53.
163. Courtney DG, Kennedy EM, Dumm RE, Bogerd HP, Tsai K, Heaton NS, Cullen BR: **Epitranscriptomic Enhancement of Influenza A Virus Gene Expression and Replication.** *Cell Host Microbe* 2017, **22**(3):377-386.e375.
164. Gokhale NS, Horner SM: **RNA modifications go viral.** *PLoS Pathog* 2017, **13**(3):e1006188.
165. Tirumuru N, Zhao BS, Lu W, Lu Z, He C, Wu L: **N(6)-methyladenosine of HIV-1 RNA regulates viral infection and HIV-1 Gag protein expression.** *Elife* 2016, **5**.
166. Lichinchi G, Gao S, Saletore Y, Gonzalez GM, Bansal V, Wang Y, Mason CE, Rana TM: **Dynamics of the human and viral m(6)A RNA methylomes during HIV-1 infection of T cells.** *Nat Microbiol* 2016, **1**:16011.

167. Price AM, Hayer KE, McIntyre ABR, Gokhale NS, Abebe JS, Della Fera AN, Mason CE, Horner SM, Wilson AC, Depledge DP *et al*: **Direct RNA sequencing reveals m.** *Nat Commun* 2020, **11**(1):6016.
168. Cartegni L, Maconi M, Morandi E, Cobianchi F, Riva S, Biamonti G: **hnRNP A1 selectively interacts through its Gly-rich domain with different RNA-binding proteins.** *J Mol Biol* 1996, **259**(3):337-348.
169. Ule J, Blencowe BJ: **Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution.** *Mol Cell* 2019, **76**(2):329-345.
170. Ishikawa F, Matunis MJ, Dreyfuss G, Cech TR: **Nuclear proteins that bind the pre-mRNA 3' splice site sequence r(UUAG/G) and the human telomeric DNA sequence d(TTAGGG)n.** *Mol Cell Biol* 1993, **13**(7):4301-4310.
171. Fialcowitz EJ, Brewer BY, Keenan BP, Wilson GM: **A hairpin-like structure within an AU-rich mRNA-destabilizing element regulates trans-factor binding selectivity and mRNA decay kinetics.** *J Biol Chem* 2005, **280**(23):22406-22417.
172. Yoon JH, De S, Srikantan S, Abdelmohsen K, Grammatikakis I, Kim J, Kim KM, Noh JH, White EJ, Martindale JL *et al*: **PAR-CLIP analysis uncovers AUF1 impact on target RNA fate and genome integrity.** *Nat Commun* 2014, **5**:5248.
173. Sarkar B, Lu JY, Schneider RJ: **Nuclear import and export functions in the different isoforms of the AUF1/heterogeneous nuclear ribonucleoprotein protein family.** *J Biol Chem* 2003, **278**(23):20700-20707.
174. Meyer A, Golbik RP, Sanger L, Schmidt T, Behrens SE, Friedrich S: **The RGG/RG motif of AUF1 isoform p45 is a key modulator of the protein's RNA chaperone and RNA annealing activities.** *RNA Biol* 2019, **16**(7):960-971.
175. Suzuki M, Iijima M, Nishimura A, Tomozoe Y, Kamei D, Yamada M: **Two separate regions essential for nuclear import of the hnRNP D nucleocytoplasmic shuttling sequence.** *FEBS J* 2005, **272**(15):3975-3987.
176. Li X, Johansson C, Cardoso Palacios C, Mossberg A, Dhanjal S, Bergvall M, Schwartz S: **Eight nucleotide substitutions inhibit splicing to HPV-16 3'-splice site SA3358 and reduce the efficiency by which HPV-16 increases the life span of primary human keratinocytes.** *PLoS One* 2013, **8**(9):e72776.
177. Li X, Johansson C, Glahder J, Mossberg AK, Schwartz S: **Suppression of HPV-16 late L1 5'-splice site SD3632 by binding of hnRNP D proteins and hnRNP A2/B1 to upstream AUAGUA RNA motifs.** *Nucleic Acids Res* 2013, **41**(22):10488-10508.
178. Wu C, Nilsson K, Zheng Y, Ekenstierna C, Sugiyama N, Forslund O, Kajitani N, Yu H, Wennerberg J, Ekblad L *et al*: **Short half-life of HPV16 E6 and E7 mRNAs sensitizes HPV16-positive tonsillar cancer**

- cell line HN26 to DNA-damaging drugs. *Int J Cancer* 2019, **144**(2):297-310.
179. Dürst M, Gissmann L, Ikenberg H, zur Hausen H: **A papillomavirus DNA from a cervical carcinoma and its prevalence in cancer biopsy samples from different geographic regions.** *Proc Natl Acad Sci U S A* 1983, **80**(12):3812-3815.
180. Somberg M, Zhao X, Fröhlich M, Evander M, Schwartz S: **Polypyrimidine tract binding protein induces human papillomavirus type 16 late gene expression by interfering with splicing inhibitory elements at the major late 5' splice site, SD3632.** *J Virol* 2008, **82**(7):3665-3678.
181. Nasim MT, Chernova TK, Chowdhury HM, Yue BG, Eperon IC: **HnRNP G and Tra2beta: opposite effects on splicing matched by antagonism in RNA binding.** *Hum Mol Genet* 2003, **12**(11):1337-1348.
182. Guang S, Felthouser AM, Mertz JE: **Binding of hnRNP L to the pre-mRNA processing enhancer of the herpes simplex virus thymidine kinase gene enhances both polyadenylation and nucleocytoplasmic export of intronless mRNAs.** *Mol Cell Biol* 2005, **25**(15):6303-6313.
183. Huang J, Li SJ, Chen XH, Han Y, Xu P: **hnRNP-R regulates the PMA-induced c-fos expression in retinal cells.** *Cell Mol Biol Lett* 2008, **13**(2):303-311.
184. Chou MY, Rooke N, Turck CW, Black DL: **hnRNP H is a component of a splicing enhancer complex that activates a c-src alternative exon in neuronal cells.** *Mol Cell Biol* 1999, **19**(1):69-77.
185. Collier B, Oberg D, Zhao X, Schwartz S: **Specific inactivation of inhibitory sequences in the 5' end of the human papillomavirus type 16 L1 open reading frame results in production of high levels of L1 protein in human epithelial cells.** *J Virol* 2002, **76**(6):2739-2752.
186. Baralle FE, Giudice J: **Alternative splicing as a regulator of development and tissue identity.** *Nat Rev Mol Cell Biol* 2017, **18**(7):437-451.
187. Urbanski LM, Leclair N, Anczuków O: **Alternative-splicing defects in cancer: Splicing regulators and their downstream targets, guiding the way to novel cancer therapeutics.** *Wiley Interdiscip Rev RNA* 2018, **9**(4):e1476.
188. Zucconi BE, Wilson GM: **Modulation of neoplastic gene regulatory pathways by the RNA-binding factor AUF1.** *Front Biosci (Landmark Ed)* 2011, **16**:2307-2325.
189. Ferguson J, Campos-León K, Pentland I, Stockton JD, Günther T, Beggs AD, Grundhoff A, Roberts S, Noyvert B, Parish JL: **The chromatin insulator CTCF regulates HPV18 transcript splicing and differentiation-dependent late gene expression.** *PLoS Pathog* 2021, **17**(11):e1010032.
190. Pentland I, Campos-León K, Cotic M, Davies KJ, Wood CD, Groves IJ, Burley M, Coleman N, Stockton JD, Noyvert B *et al*: **Disruption of**

CTCF-YY1-dependent looping of the human papillomavirus genome activates differentiation-induced viral oncogene transcription. *PLoS Biol* 2018, **16**(10):e2005752.

191. Paris C, Pentland I, Groves I, Roberts DC, Powis SJ, Coleman N, Roberts S, Parish JL: **CCCTC-binding factor recruitment to the early region of the human papillomavirus 18 genome regulates viral oncogene expression.** *J Virol* 2015, **89**(9):4770-4785.

Paper I



HnRNP D activates production of HPV16 E1 and E6 mRNAs by promoting intron retention

Xiaoxu Cui¹, Chengyu Hao¹, Lijing Gong^{1,2}, Naoko Kajitani^{1,3,*} and Stefan Schwartz^{1,3}

¹Department of Laboratory Medicine, Lund University, BMC-B13, 221 84 Lund, Sweden, ²China Institute of Sport and Health Sciences, Beijing Sport University, Haidian District, Beijing 100084, China and ³Department of Medical Biochemistry and Microbiology (IMBIM), Uppsala University, BMC-B9, 751 23 Uppsala, Sweden

Received September 19, 2021; Revised January 26, 2022; Editorial Decision February 09, 2022; Accepted February 11, 2022

ABSTRACT

Human papillomavirus type 16 (HPV16) E1 and E6 proteins are produced from mRNAs with retained introns, but it has been unclear how these mRNAs are generated. Here, we report that hnRNP D act as a splicing inhibitor of HPV16 E1/E2- and E6/E7-mRNAs thereby generating intron-containing E1- and E6-mRNAs, respectively. N- and C-termini of hnRNP D contributed to HPV16 mRNA splicing control differently. HnRNP D interacted with the components of splicing machinery and with HPV16 RNA to exert its inhibitory function. As a result, the cytoplasmic levels of intron-retained HPV16 mRNAs were increased in the presence of hnRNP D. Association of hnRNP D with HPV16 mRNAs in the cytoplasm was observed, and this may correlate with unexpected inhibition of HPV16 E1- and E6-mRNA translation. Notably, hnRNP D40 interacted with HPV16 mRNAs in an HPV16-driven tonsillar cancer cell line and in HPV16-immortalized human keratinocytes. Furthermore, knockdown of hnRNP D in HPV16-driven cervical cancer cells enhanced production of the HPV16 E7 oncoprotein. Our results suggest that hnRNP D plays significant roles in the regulation of HPV gene expression and HPV-associated cancer development.

INTRODUCTION

Human papillomaviruses (HPV) contain a double-stranded circular DNA genome of approximately 8 kb in size (1). HPV infections cause a range of disease from benign warts to invasive cancers, for example cervical cancer and tonsillar cancer (2). HPV type 16 (HPV16) is responsible for around 55% of all cervical cancers while the remainder is caused by other high risk (HR) HPV types (3). Cancer progression is due to an increased continuous expression of HPV oncoproteins E6 and E7 that inactivate tumor sup-

pressor proteins p53 and pRb (4), respectively. E6 and E7 activate the cell cycle, inhibit apoptosis and cause genomic instability (5–9). The HPV16 E1 and E2 proteins are key factors during replication of HPV16 genomic DNA. E1 functions as DNA helicase whereas E2 has a multifunctional role including transcriptional regulation, initiation of HPV16 DNA replication, facilitation of HPV16 genome partitioning during mitosis and post-transcriptional control of HPV16 gene expression (10–13). In contrast to E6 and E7, E2 has pro-apoptotic properties and is frequently inactivated when the HPV16 genome integrates in cellular chromosomes, a process that possibly enhances carcinogenesis (11,14). The HPV16 E4 and E5 proteins are essential for completion of the HPV16 replication cycle and E5 may contribute to carcinogenesis (15,16).

Since the HPV16 genome has two promoters only, alternative mRNA splicing plays a major role in the regulated expression of all HPV16 genes (17–22). A complex pattern of alternatively spliced and polyadenylated HPV16 mRNAs is observed during the HPV16 life cycle. Therefore, it is not surprising that a number of cis-acting regulatory RNA elements and their cognate trans-acting factors control the HPV16 alternative splicing and polyadenylation. In addition, it has been shown that the levels of various RNA-binding proteins are altered during the progression of HPV16-infected cells to cervical cancer through a series of premalignant cervical intraepithelial lesions (23,24). Thus, it is of interest to identify cellular RNA-binding proteins that control HPV16 gene expression.

Heterogeneous nuclear ribonucleoproteins (hnRNPs) represent a large family of RNA-binding proteins (RBPs). The association of hnRNP proteins with pre-mRNAs is initiated co-transcriptionally at the nascent transcripts. Many of the RNA-binding proteins remain bound to the resulting mRNAs all the way to the ribosomes and shuttle back and forth between the nucleus and the cytoplasm, demonstrating that RNA-binding proteins are important determinants of pre-mRNA processing during the entire mRNA pathway including mRNA splicing, localization, translation and stability (25). hnRNPs are massively involved in alternative splicing and the canonical function of hnRNPs

*To whom correspondence should be addressed. Tel: +46 18 4714444; Fax: +46 18 4714673; Email: naoko.kajitani@imbim.uu.se

is performed through their binding to RNA elements adjacent to splice sites, thereby either repressing or supporting the assembly of the spliceosome complex on splice donor (SD) or splice acceptor (SA) sites. Alternatively, they affect the recruitment of other RNA-binding proteins such as Serine/Arginine (SR) rich proteins to exonic or intronic splicing enhancers or silencers (26). Furthermore, many hnRNPs participate in more than one of these processes (27), e.g. regulation of both alternative splicing- and poly(A)-site usage by hnRNP H/F or L, alternative splicing and translation by hnRNP A1, or alternative splicing and RNA stability by hnRNP D.

The hnRNP family comprises at least 20 major RNA-binding proteins originally named alphabetically from A1 to U. These proteins share modular structures that include RNA recognition motifs (RRM) or quasi-RRMs, other motifs such as the K Homology (KH) domain and the arginine/glycine-rich RGG motif and other RNA-binding domains (RBD) present on a subset of hnRNPs (28). In some hnRNPs other auxiliary domains like glycine rich domains and proline-rich domains may be present (25,29). Although all hnRNPs display RNA-binding activity with a certain degree of specificity, an hnRNP protein may not bind exclusively to high-affinity binding sites. The RNA-binding specificity of hnRNPs is strongly influenced by the type and number of RNA-binding domains on the hnRNP, which in turn generates both general and specific interactions with nucleic acids, as well as the primary and secondary structures of the target RNA (30). HnRNPs also possess nuclear localization sequences (NLSs) that bring the proteins to the nucleus, while other hnRNPs possess nucleocytoplasmic shuttling domains which allow hnRNPs to shuttle between nuclear and cytoplasmic compartments (28). Given the important role of hnRNPs in mRNA metabolism, we speculate that HPV16 early gene expression is under control of a range of hnRNPs. Here we determined the effect of 13 different hnRNPs on HPV16 early mRNA splicing and we found that a number of the hnRNPs affected HPV16 early mRNA splicing, but in different ways. We focused our attention on the members of the cellular hnRNP D protein family. hnRNP D, also known as AU-rich element binding factor 1 (AUF1) promotes the decay of many target mRNAs (31), but it was also reported to enhance the stability and affect translation of target transcripts. As a result, hnRNP D is involved in multiple cellular processes, including miRNA biogenesis (32), translational regulation (33,34), telomere maintenance (35,36), cell cycle control (37), apoptosis (38) and inflammatory responses (39) since hnRNP D target mRNAs encode proteins implicated in these processes. Reports on hnRNP D and mRNA splicing are relatively scarce but a regulatory function of hnRNP D proteins in alternative splicing has been suggested (34). One example is that hnRNP D proteins as well as hnRNP DL control their own expression by auto- or cross-alternative splicing regulation (40). Another example suggests that hnRNP D together with neuronal members of the ELAVL protein family (nELAVLs) induce neuron-specific alternative splicing of the amyloid precursor protein (APP) (41).

In this manuscript, we show that hnRNP D proteins inhibit splicing and promote retention of the E1- and E6-

encoding introns on the HPV16 early mRNAs, thereby specifically stimulating production of the partially spliced HPV16 E1- and E6-encoding mRNAs. Furthermore, hnRNP D facilitated export of the partially spliced HPV16 E1 and E6 mRNAs to the cytoplasm. However, despite the fact that the HPV16 E1 and E6 mRNAs reached the cytoplasm in association with hnRNP D, HPV16 mRNAs were poorly translated. In conclusion, our results suggest that hnRNP D proteins inhibit HPV16 early mRNA splicing and enhance production of the partially spliced, intron-containing HPV16 E1 and E6 mRNAs, thereby playing an important role at the initial steps of the biogenesis of HPV16 E1 and E6 mRNA.

MATERIALS AND METHODS

Cells

HeLa, 293T, SiHa and C33A2 cells were cultured in Dulbecco's modified Eagle medium (DMEM) (HyClone) with 10% bovine calf serum (HyClone) and penicillin/streptomycin (Gibco). C33A2 cell line has been described previously (42). Briefly, C33A2 is an in-house cell line derived from C33A cells stably transfected with HPV16 reporter plasmid pBELsLuc (43). HPV16-infected tonsillar cancer cell line HN26 has been described previously (44). Briefly, 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. HN26 cells were cultured in RPMI 1640 medium (HyClone) with 10% iron-supplemented bovine calf serum (HyClone), 5% MEM non-essential amino acid solution (Sigma Aldrich) and 5% sodium pyruvate (Sigma Aldrich). Neonatal Primary Normal Human Foreskin Keratinocytes (nHFK, purchased from Thermo Fisher Scientific) and 3310 cells were cultured in EpiLife medium (Gibco) supplemented with 1% human keratinocyte growth supplement (HKGS, Gibco) and 0.2% Gentamicin/Amphotericin (Gibco). Differentiation of nHFK was induced by addition of CaCl₂ at a final concentration of 2.4 mM in the keratinocyte culture medium for 24 h. The HPV16-immortalized keratinocyte 3310 cell line has been described previously and was generated by stable transfection of nHFK with HPV16 genome plasmid pHPV16ANE2fs (43).

Plasmids

The following plasmids have been described previously: pHPV16AN (43), pC97ELsluc (42) and pBELsluc (43). Construction of phnRNP F, phnRNP I, phnRNP A2 and phnRNP Q has been described previously (45) and so has phnRNP A1 (46) and phnRNP C1 (47). Plasmids Flag-p37, Flag-p40, Flag-p42 and Flag-p45 were kindly given by Dr R. J. Schneider (48), hnRNP G plasmid by Dr I. C. Eperon (49), histidine and myc-tagged hnRNP L plasmid by Dr S. Guang (50) and histidine and myc-tagged hnRNP R plasmid by Dr P. Xu (51). phnRNP AB encoding myc-tagged hnRNP AB transcript variant 1 (RC204360) and phnRNP DL encoding hnRNP DL transcript variant 2 (RC204064) were purchased from OriGene Technologies. phnRNP H

contains the hnRNP H open reading frame driven by a CMV promoter. It was constructed by PCR-amplification of hnRNP H coding sequence from pET-15B-hnRNP H (generously provided by Dr D. Black) (52) followed by subcloning into pCL086 (53).

To construct HPV16 subgenomic plasmids pX656 and pX478, primer B97S (Supplementary Table S1) was used in combination with primer X656A or X478A to amplify HPV16 sequences that were digested with PteI and XhoI and inserted between the PteI and XhoI sites in CMV promoter driven empty vector pCL086 (53). HPV16 subgenomic plasmid pXH856F encoded HPV16 E6 and E7 genes (HPV16 nucleotides positions 104-855) (HPV16 nucleotide positions refer to the HPV16 reference sequence HPV16R (GeneBank: K02718.1)). E6 was fused to an HA tag sequence at the 5'-end and E7 to a flag tag sequence at the 3'-end. pXH856F was generated by insertion of DNA fragment PCR-amplified with pXH856F sense and anti-sense primer (Supplementary Table S1) and subcloned into pCL086 at PteI and XhoI sites. The 5'-splice site SD226 in pXH856F was mutated (T228C, A229C) creating pXH856SDmF. The mutations in SD226 were introduced by PCR mutagenesis with pXH856SDmF sense and antisense primers (Supplementary Table S1) in a two-step PCR amplification reaction using overlapping followed by subcloning into pXH856F. Plasmid p16E1-3xF encodes the HPV16 E1 gene (HPV16 nucleotides positions 865-2811) fused with a 3xFLAG tag sequence at the 3'-end and was constructed by PCR amplification using p16E1-3xF inverse sense and anti-sense primers (Supplementary Table S1) and subcloned into pCL086. HPV16 5'-splice sites SD880 and SD1302 on p16E1SDm-3xF were mutated (G881C and G1303C) to create p16E1SDm-3xF. The mutations were conducted by p16E1SDm-3xF sense and antisense primers (Supplementary Table S1) in two-step PCR amplification.

To construct HPV subgenomic plasmids pBELEN, pBELEndE1, pBELsluc plasmid was cut with restriction enzymes CsiI and XhoI, followed by blunting of sticky DNA ends with Klenow fragment and re-ligation, resulting in pBELEN. To construct pBELEndE1, pBELEN was cut with restriction enzymes AdeI and NsiI to delete sequences in the E1 coding region, followed by blunting of sticky DNA ends with Klenow fragment and re-ligation.

To construct plasmid pFLAG-hnRNP40, the hnRNP40 sequence was fused to a FLAG-tag sequence at the 5'-end by PCR amplification with primers pflag-D40 sense and anti-sense (Supplementary Table S1) using plasmid FLAG-p40 (48) as template. To construct hnRNP40 deletion mutants pD1, pD2 and pD5, hnRNP40 sequences fused to flag tag at the 5'-end were first PCR-amplified with primers 'pD1-, pD2- or pD5-sense' and 'pflag-D40 anti-sense' (Supplementary Table S1) from plasmid pflag-hnRNP40. To construct hnRNP40 deletion mutants pD3, pD4 and pD7, hnRNP40 sequences fused to a flag tag at the 5'-end were first PCR-amplified with primers 'pflag-D40 sense' and 'pD3-, pD4- or pD7- anti-sense' (Supplementary Table S1) from plasmid pFLAG-hnRNP40. To construct hnRNP40 deletion mutant pD8, hnRNP40 sequences fused to a FLAG tag at the 5'-end were first PCR-amplified with primers 'pD1 sense' and 'pD4 anti-sense' (Supplementary Table S1) from plas-

mid pFLAG-hnRNP40. The PCR-fragments mentioned above were digested with restriction enzymes PteI and XhoI and subcloned into plasmid pCL086 cut with PteI and XhoI. pD9 was constructed by blunt end re-ligation of pFLAG-hnRNP40 cut by StuI and XhoI. pflag-hnRNP40 and deletion mutants are shown schematically in Figure 4A. To construct hnRNP40 substitution mutants pQ6A and pAGG, hnRNP D40 sequences (BglII and XhoI) encoding hnRNP40 mutant fragments were synthesized by Eurofins Genomics. These fragments were digested with restriction enzymes BglII and XhoI and subcloned into plasmid pFLAG-hnRNP40 digested with BglII and XhoI. To construct pD1 and pD2 substitution mutants pD1-AGG, pD1-Q6A, pD2-AGG and pD2-Q6A, plasmids pAGG or pQ6A were digested with BglII and XhoI and subcloned into plasmid pD1 and pD2 digested with BglII and XhoI. Substitution mutants are shown schematically in Figure 5A. To generate hnRNP D40-plasmids fused to GFP, primer 'pEGFP-D40 sense' was first used in combination with primer 'pEGFP-D40-, pEGFP-D3-, pEGFP-D4-, pEGFP-D7- or pEGFP-D9-anti-sense' (Supplementary Table S1) to amplify full-length hnRNP40 or deletion mutant sequences of hnRNP40 from pflag-hnRNP40, pD1, pD2, pD5, pD3, pD4, pD8, pD7 or pD9, respectively. The PCR-fragments were digested with HindIII and BamHI and inserted between the HindIII and BamHI sites in the pEGFP-C3 vector (Clontech) resulting in pEGFP-D40, pEGFP-D1, pEGFP-D2, pEGFP-D5, pEGFP-D3, pEGFP-D4, pEGFP-D8, pEGFP-D7 and pEGFP-D9.

To construct the E6 protein encoding plasmid with T7 promotor pET32a-HA-E6SDm, E6 sequences were PCR amplified from pXH856SDmF using pET32a-HA-E6SDm sense and antisense primers (Supplementary Table S1) followed by subcloning into pET32a using EcoRI and XhoI sites. To construct the E1 protein encoding plasmid with T7 promotor pcDNA3.1(+)-E1SDm-3XFlag, E1 sequences were released from plasmid p16E1SDm-3xF by HindIII and XhoI digestion followed by subcloning into pcDNA3.1(+) digested with HindIII and XhoI.

Transfections

Transfections of HeLa cells and 293T cells were performed with Turbofect according to the manufacturer's protocol (Thermo Fisher Scientific). Briefly, a mixture of Turbofect:DNA ratio of 2:1 (μ l reagent: μ g DNA) for HeLa cells or 4:1 for 293T cells and DMEM without serum was incubated at room temperature for 20 min prior to dropwise addition to subconfluent cells. Transfections of nHFK were performed with ViaFect according to the manufacturer's protocol (Promega). In brief, a mixture of ViaFect:DNA ratio of 3:1 and EpiLife without serum was incubated at room temperature for 20 min prior to dropwise addition to subconfluent cells. Fluorescence images of EGFP set of plasmids transfected HeLa cells were acquired using Olympus CKX53 inverted microscope.

Nuclear and cytoplasmic extraction

Nuclear and cytoplasmic extracts were prepared from HeLa cells grown in 6 cm dishes at 24 h post-transfection. Cells

were harvested by using NE-PER Nuclear and Cytoplasmic Extraction Reagents (Thermo Fisher Scientific) according to the manufacturer's protocol. In brief, cell pellets were resuspended in ice cold buffer CER I with protease inhibitors (Sigma Aldrich) and vortexed vigorously prior to incubation on ice for 10 min. Ice-cold buffer CER II was added to the samples that were vortexed vigorously and incubated on ice for one minute. After 5 min of maximum speed centrifugation, the supernatants were collected as cytoplasmic extracts. The pellets were washed once by PBS and collected as nuclear extracts.

RNA extraction and RT-PCR

Total RNA was extracted from transfected cells using TRI Reagent (Sigma Aldrich) and Direct-zol RNA MiniPrep (ZYMO Research) according to the manufacturer's protocols. Reverse transcription (RT) was performed in a 20 μ l reaction using random hexamers (Invitrogen) and reverse transcriptase (Invitrogen). One microliter of cDNA was subjected to PCR-amplification. cDNA representing HPV16 mRNAs spliced from HPV16 5'-splice site SD226 to 3'-splice site SA409 was amplified with RT-PCR primers 97S and 438AS (Supplementary Table S1) and cDNA representing HPV16 mRNAs spliced from 5'-splice site SD226 to 3'-splice sites SA409, SA526, or SA742 with RT-PCR primers 97S and 880AS (Supplementary Table S1). cDNA representing HPV16 mRNAs spliced from 5'-splice site SD880 to 3'-splice site SA2709 or SA3358 was amplified with RT-PCR primers 773S and E2AS/E2QAS or E4AS (Supplementary Table S1). cDNA representing HPV16 mRNAs spliced from 5'-splice site SA2709 or SA3358 were amplified with RT-PCR primers 97S and E2AS/E2QAS or E4AS, respectively (Supplementary Table S1). Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) cDNA was amplified with primers GAPDHF and GAPDHR (Supplementary Table S1). Spliced actin and unspliced actin cDNAs were amplified with primers actin-s or actin-s1 and actin-a (Supplementary Table S1). cDNA representing mRNAs encoding each of the four isoforms of hnRNP D were amplified with hnRNP D mRNA-specific primers 2S and 7A (Supplementary Table S1). For location of these mRNAs see Figure 3A. 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). Examples of control PCR experiments performed on RNA samples in the absence of reverse transcriptase are shown in various figures. Primer pairs used for RT-PCR and RT-qPCR (described in next section) are summarized in Supplementary Table S4.

Real-time quantitative PCR (qPCR)

qPCR was performed in a final reaction volume of 20 μ l with 1 μ l of cDNA prepared as described above in a MiniOpticon (Bio-Rad) using the SsoAdvanced SYBR Green Supermix (Bio-Rad) according to the manufacturer's instructions. To quantitate intron retained E6 mRNAs by RT-qPCR primers TotalE6F and 234AS were used (Supplementary Table S1). For primer location, see Figure 1C and

Supplementary Figure S5. Spliced E2 mRNAs were quantitated by RT-qPCR using primers 773S and E2AS (Supplementary Table S1, for primer location, see Figure 1). For quantitation of E1 mRNA by RT-qPCR, primers 773S and E1AS were used (Supplementary Table S1, for primer location, see Figure 1). The expression levels of the mRNAs were determined from the threshold cycle (C_t), and the relative expression levels were calculated using the $2^{-\Delta\Delta C_t}$ method. Results were normalized to GAPDH mRNA levels determined with primers GAPDHF and GAPDHR (Supplementary Table S1). mRNA quantification was performed in triplicates, and negative controls were included in each reaction. Melting curves were analyzed in each reaction.

Secreted luciferase assay

The *Metridia longa* secreted luciferase (sLuc) activity in the cell culture medium of transfected cells was monitored with the help of the Ready-To-Glow Secreted Luciferase Reporter assay according to the instructions of the manufacturer (Clontech) as described previously (43). In brief, 50 μ l of cell culture medium was added to 5 μ l of 0.5X Secreted Luciferase substrate/Reaction buffer in a 96-well plate and luminescence was determined in a Tristar LB941 Luminometer.

Western blotting

Cell extracts for western blotting were obtained by resuspending transfected cells in radioimmunoprecipitation assay (RIPA) buffer consisting of 50 mM Tris HCl pH 7.4, 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, 2 mM EDTA, 1 mM DTT and protease inhibitor (Sigma Aldrich), followed by centrifugation at full speed for 20 min and collection of the supernatants. Proteins were denatured by boiling in Laemmli buffer. After SDS-PAGE, the proteins on the gels were transferred onto nitrocellulose membranes, blocked with 5% nonfat dry milk in PBS containing 0.1% Tween 20 and stained with specific primary antibodies (Supplementary Table S2) to the indicated proteins followed by incubation with secondary antibody (Supplementary Table S2) conjugated with horseradish peroxidase and detection with chemiluminescence reagents.

ssRNA oligo pull down

Whole cell lysates of HeLa cells were prepared using cell lysis buffer consisting of 25 mM Tris HCl pH 7.4, 150 mM NaCl, 1% NP-40, 1 mM EDTA, 1 mM DTT, 200 units of RiboLock RNase Inhibitor (Thermo Fisher Scientific), Proteinase inhibitor (Sigma Aldrich) and 5% glycerol. Biotin-labeled ssRNA oligos were purchased from Sigma Aldrich (Supplementary Table S3). DynabeadsM-280 Streptavidin magnetic beads (Invitrogen) were bound to biotin-labeled ssRNA oligonucleotides in 200 μ l of binding buffer (10 mM Tris, pH 7.4, 150 mM NaCl, 2.5 mM MgCl₂, 0.5% Triton X-100) in strips of PCR tubes incubated at RT for 20 min. To pull-down proteins, 15 μ g HeLa whole cell lysate was mixed with beads bound with ssRNA oligos and

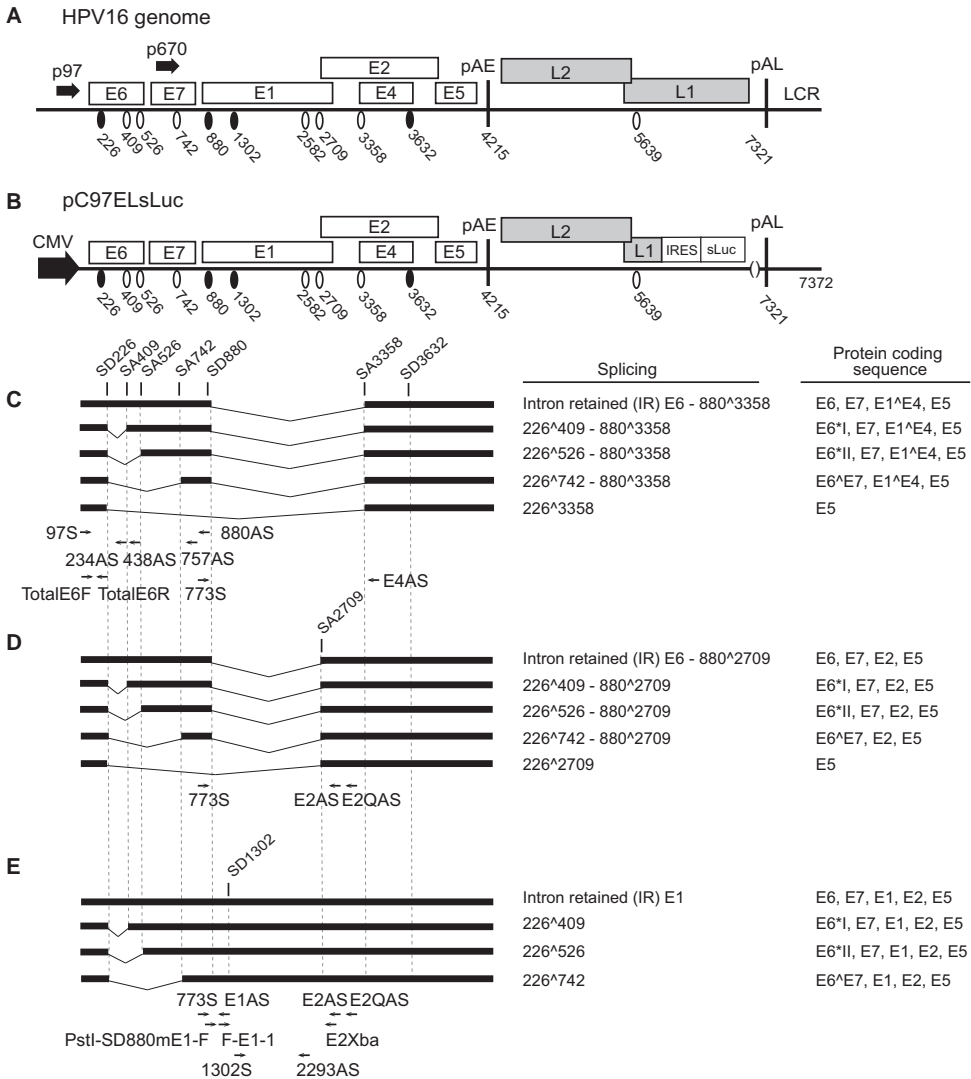


Figure 1. Schematic representation of the HPV16 genome, HPV16 subgenomic plasmid pC97ELsLuc and representative HPV16 mRNAs. (A) Linearized HPV16 genome (numbers refer to the HPV16 reference strain GeneBank: K02718.1). Early and late genes are indicated. P97: HPV16 early promoter. P670: HPV16 late promoter. Black oval: splice donor. White oval: splice acceptor. pAE: HPV16 early polyadenylation site. pAL: HPV16 late polyadenylation site. LCR: HPV16 long control region. (B) HPV16 subgenomic plasmid pC97ELsLuc encodes all HPV16 genes. HPV16 early promoter p97 was replaced by human cytomegalovirus immediate early promoter (CMV). Secreted luciferase (sLuc) gene was integrated in the L1 gene following the poliovirus 2A internal ribosomal entry site (IRES) sequence. (C–E) Schematic structures of HPV16 early transcripts produced from pC97ELsLuc. Splicing at SD226 occurs independently of splicing at the downstream SD880 and generates splice variants in the E6E7-coding region concomitantly with alternative splicing from SD880, i.e. E4 (880^Δ3358) (C), E2 (880^Δ2709) (D) or mRNAs unspliced at SD880 (E). The protein coding sequences present on each mRNA are indicated to the right. Arrows: Annealing positions of HPV16 RT-PCR primers. Primer sequences are available in Supplementary Table S1.

incubated with rotation for 1 h at room temperature followed by washing of beads 10 times with binding buffer using DynaMag 96 Side magnetic plate (Invitrogen). Proteins were eluted by boiling of the beads in Laemmli Buffer. Samples were subjected to SDS-PAGE followed by western blot analysis with the indicated antibodies (Supplementary Table S2).

Co-immunoprecipitation

Transfected HeLa cells were lysed in cell lysis buffer as described above under 'ssRNA oligo pull down'. For immunoprecipitation, anti-flag antibody (M2, Sigma Aldrich) or IgG was added to Dynabeads protein G and incubated with cell lysates under gentle rocking at 4°C overnight. The complexes were washed six times using cell lysis buffer and eluted by boiling in Laemmli buffer. Samples were subjected to SDS-PAGE followed by western blotting with specific primary antibodies (Supplementary Table S2).

UV-crosslinking and immunoprecipitation (CLIP)

Transfected HeLa cells grown in 10 cm dishes were washed by ice-cold PBS followed by crosslinking twice with 0.4 J cm⁻² UV light (254 nm) in a bio-link cross-linker (Biometra). Cytoplasmic extracts were prepared as described above. Whole cell lysates were prepared by resuspending cells in one ml of RIPA buffer and incubated on ice for 30 min with occasional vortexing to lyse cells. For immunoprecipitations, 2 µg of the anti-flag antibody (M2, Sigma Aldrich) or mouse IgG was incubated at 4°C overnight in 0.5 ml of cell lysate. About 20 µl of Dynabeads Protein G (10004D, Invitrogen) and 20 µl Dynabeads Protein A (10001D, Invitrogen) were blocked with 1% BSA for 0.5 h, washed three times in RIPA buffer and then added to the antibody-protein mixtures followed by incubation for 1 h at 4°C. The beads were washed three times with buffer I (50 mM Tris HCl pH 7.4, 300 mM NaCl, 0.5% NP-40, 1 mM EDTA, 1 mM DTT), three times with buffer II (50 mM Tris HCl pH 7.4, 800 mM NaCl, 0.5% NP-40, 1 mM EDTA, 1 mM DTT) and three times with buffer III (50 mM Tris HCl pH 7.4, 800 mM NaCl, 250 mM LiCl, 0.5% NP-40, 1 mM EDTA, 1 mM DTT). RNA was eluted by phenol/chloroform extraction and ethanol-precipitated and dissolved in 20 µl of water. About 10 µl of immunoprecipitated RNA was reverse transcribed using M-MLV reverse transcriptase (Invitrogen) and random primers (Invitrogen) according to the protocol of the manufacturer. Two microliters of cDNA were subjected to PCR amplification using HPV16-specific primer pairs 773S and E1AS, 773S and 438AS or E6F and E6R (Supplementary Table S1) as described above.

Ribonucleoprotein (RNP) immunoprecipitation (RIP) analysis

For immunoprecipitation of endogenous ribonucleoprotein (RNP) complexes (RIP analysis) from whole cell extracts, HN26 cells or 3310 cells were lysed in cell lysis buffer. The supernatants were incubated with anti-AUF1 antibody (Cell Signalling) or IgG (Millipore) (Supplementary Table

S2) overnight at 4°C. Dynabeads Protein G (10004D, Invitrogen) + 20 µl Dynabeads Protein A (10001D, Invitrogen) were added to the antibody-protein mixtures followed by incubation for 1 h at 4°C. After washing of the beads six times using cell lysis buffer, RNA was eluted using Tri reagent and incubated with 20U of RNase-free DNase I for 1 h at 37°C and subjected to RT-PCR using HPV16-specific primer pairs 773S and E4AS or 97S and 438AS (Supplementary Table S1) as described above.

siRNA library and siRNA transfections

ON-TARGETplus human hnRNP D siRNA SMARTpool consists of four siRNAs to hnRNP D (L-004079-00-0010) (Dharmacon). A scrambled negative control siRNA (D-001810-10-05) was also purchased from Dharmacon. Transfections were conducted with DharmaFECT1 (Dharmacon) according to the instructions of the manufacturer. The siRNA SMARTpool to hnRNP D or scrambled control siRNAs were transfected in duplicates into C33A2 or SiHa cells grown in 12-well plates for RNA extraction and RT-qPCR or in 6-well plates for protein extraction and western blotting. Cells were harvested at 48 h post-transfection for RNA extraction and RT-qPCR or 72 h post-transfection for protein extraction and western blotting performed as described above.

In vitro translation assay

In vitro translation was carried out with TNT(R) Quick Coupled Transcription/Translation Systems (Promega) according to the instructions of the manufacturer. In brief, 100 ng pET32a-HA-E6SDm or 200 ng pcDNA3.1(+)-E1SDm-3XFlag plasmid was translated in the absence or presence of 1 µM recombinant hnRNP D protein (EUPROTEIN) or 1 µM BSA. The 25 µl reactions were incubated for 90 min at 30°C. The translation reactions were analyzed by western blotting as described above. The Luciferase control RNA was also translated in the absence or presence of 1 µM hnRNP D or 1 µM BSA. Luciferase activity in the translation reactions were monitored according to the instructions of the manufacturer using Tristar LB941 Luminometer. Recombinant hnRNP D and BSA were separated on SDS-PAGE followed by staining with Colloidal Blue Staining Kit (Invitrogen).

Quantifications

The software used to determine band intensity in western blots and RT-PCR gels is 'Image Lab 6.1.0' and quantifications were performed with the software 'Prism GraphPad 8.4.0'.

RESULTS

hnRNP A1, hnRNP D and hnRNP I inhibit splicing of HPV16 early mRNAs

To enhance our understanding of the regulated expression of the HPV16 early genes, we wished to identify hnRNP proteins that control splicing of HPV16 mRNAs. We therefore utilized subgenomic HPV16 reporter

plasmid pC97ELsLuc that encodes all HPV16 genes and can produce all alternatively spliced HPV16 mRNAs. The schematic representation of the HPV16 genome (Figure 1A), pC97ELsLuc (Figure 1B), the structures of the HPV16 alternatively spliced mRNAs it produces (Figure 1C–E) and the location of the HPV16 RT-PCR primers are displayed in Figure 1. HPV16 plasmid pC97ELsLuc was co-transfected individually with plasmids expressing either hnRNP A1, hnRNP A2, hnRNP AB, hnRNP C1, hnRNP D, hnRNP DL, hnRNP F, hnRNP G, hnRNP H, hnRNP I, hnRNP L, hnRNP Q and hnRNP R, RNA was extracted and the various spliced HPV16 mRNAs were monitored by RT-PCR (Figure 2A–C). Quantitation of E6- and E7-encoding mRNA isoforms obtained in Figure 2A or the E1- and E2-encoding mRNA isoforms in Figure 2C are shown in Figure 2D,E, respectively. We considered three observations to be of particular interest: (i) hnRNP A1 and hnRNP A2 affected splicing of E6/E7 mRNAs as previously described with different effects on splicing of E6 and E7 mRNAs. hnRNP A1 promoted production of intron-retained E6 mRNAs, while hnRNP A2 enhanced production of 226 \wedge 742-mRNAs (46) (Figure 2A); (ii) hnRNP D, hnRNP DL, hnRNP G and hnRNP I inhibited E6/E7-mRNA splicing and enhanced production of intron-retained E6 mRNAs (Figure 2A) and (iii) hnRNP A1, hnRNP D, hnRNP DL and hnRNP I reduced the levels of spliced E2 mRNAs and promoted production of intron-retained E1-encoding mRNAs (Figure 2C,E). hnRNPs that inhibited E2 mRNA splicing also appeared to inhibit splicing between SD880 and SA3358 (E1 \wedge E4), with the exception of hnRNP I (Figure 2B). This may be explained by the presence of a long, uninterrupted polypyrimidine tract; eleven consecutive pyrimidine upstream of E2 splice site SA2709, while the E1 \wedge E4 splice site SA3358 has no more than four consecutive pyrimidine-nucleotides. Thus, HPV16 SA3358 would be a poor target for polypyrimidine tract binding protein/hnRNP I compared to HPV16 SA2709. All transfections have been repeated at least three times and the indicated RT-PCR products have been cut out from gels and subjected to sequencing to confirm their identities. The significant effects of hnRNP D on the majority of the HPV16 mRNAs warranted further studies on the effects of hnRNP D on HPV16 mRNA splicing.

All four hnRNP D variants promote intron retention of HPV16 E1 and E6 mRNAs

We transfected plasmids expressing all four isoforms of hnRNP D (hnRNP D37, D40, D42 and D45) (Figure 3A) with HPV16 plasmid pC97ELsLuc and determined the effect on HPV16 mRNA splicing by RT-PCR (Figure 1). All four isoforms of hnRNP D are ubiquitously expressed in a variety of cell lines including normal primary human foreskin keratinocytes (nHK) at the level of RNA (Supplementary Figure S1F). The transfection of each hnRNP D isoform expressing plasmid revealed equal levels of individual isoform protein expression with an increase of 7-fold compared to the endogenous protein level (Supplementary Figure S1G and H). The results revealed that all hn-

RNP D proteins inhibited HPV16 E6/E7 mRNA splicing (Figure 3B). All hnRNP D isoforms promoted production of intron-retained mRNAs encoding E1 at the expense of the spliced E2 mRNAs (880 \wedge 2709) (Figure 3C). hnRNP D proteins also had an inhibitory effect on HPV16 E6/E7 mRNA splicing, which resulted in production of intron-retained E6-encoding mRNAs at the expense of the spliced E7 mRNAs (226 \wedge 409 and 226 \wedge 526) (Figure 3B). Furthermore, RT-PCR amplification with primer pair 97S+E1AS revealed the hnRNP D also enhanced production of an HPV16 mRNA that was intron-retained in both E6/E7- and E1/E2-regions (Supplementary Figure S1A and B), suggesting that hnRNP D affected early steps in the splicing reaction or independently inhibited splice sites in the E1- and E6-coding regions. Finally, the splicing inhibitory effect of hnRNP D on the HPV16 E6/E7 mRNAs (Supplementary Figure S1C) and the E1/E2 mRNAs (Supplementary Figure S1D), and to some extent E4 mRNAs (Supplementary Figure S1E), was dependent on the concentration of transfected hnRNP D plasmid. We concluded that hnRNP D promoted production of intron-retained HPV16 mRNAs encoding E1 or E6.

Next, we investigated if the effect of hnRNP D proteins on HPV16 intron-retained E1 mRNA production could be reproduced independently of upstream splice sites located in E6/E7 gene region. To this end, we used pBELsLuc (Figure 3E) that encodes all HPV16 genes present in pC97ELsLuc except for the E6 and the E7 genes. Since the RT-PCR products representing intron-retained E1 mRNAs (Figure 3C) could potentially originate from plasmid DNA that contaminated the RNA preparations, we performed RT-PCR in the absence or presence of reverse transcription (RT). As can be seen in Figure 3F, the bands representing intron-retained E1-encoding mRNAs were not detected in the absence of RT, nor were they easily detected in the absence of hnRNP D protein (Figure 3F). Furthermore, primers located entirely within the E1 coding region (primers 1302S+2293AS) yielded similar results as the 773S and E2AS primers regarding detection of the intron-retained E1-encoding mRNAs (Figure 3G). RT-PCR with primers 773S+E1AS (located immediately downstream of SD880 and detecting intron-retained mRNAs, Figure 3E) also showed an increase in the presence of hnRNP D (Figure 3H). It also appeared that hnRNP D40 had a stronger splicing-inhibitory effect on the HPV16 mRNAs than the remaining members of the hnRNP D family (Figure 3B–D). Thus, hnRNP D40 was used in the majority of the subsequent experiments. The effect of hnRNP D40 on intron-retained E1 mRNA production was determined by RT-qPCR with primers 773S+E1AS revealing an increase of intron-retained E1 mRNAs with 4-fold in the presence of the hnRNP D40 (Figure 3I). Furthermore, the ratio between intron-retained E1 mRNAs and spliced E2 mRNAs revealed increases of 8-fold from pC97ELsLuc (Figure 3D) and 6-fold from pBELsLuc (Figure 3J) in the presence of hnRNP D40. The enhancing effect of hnRNP D40 on production of intron-retained E1 mRNAs was further confirmed by comparisons between pC97ELsLuc and pBELsLuc (Supplementary Figure S2A and B) in the absence or presence of hnRNP D40 (Supplementary Figure S2C;

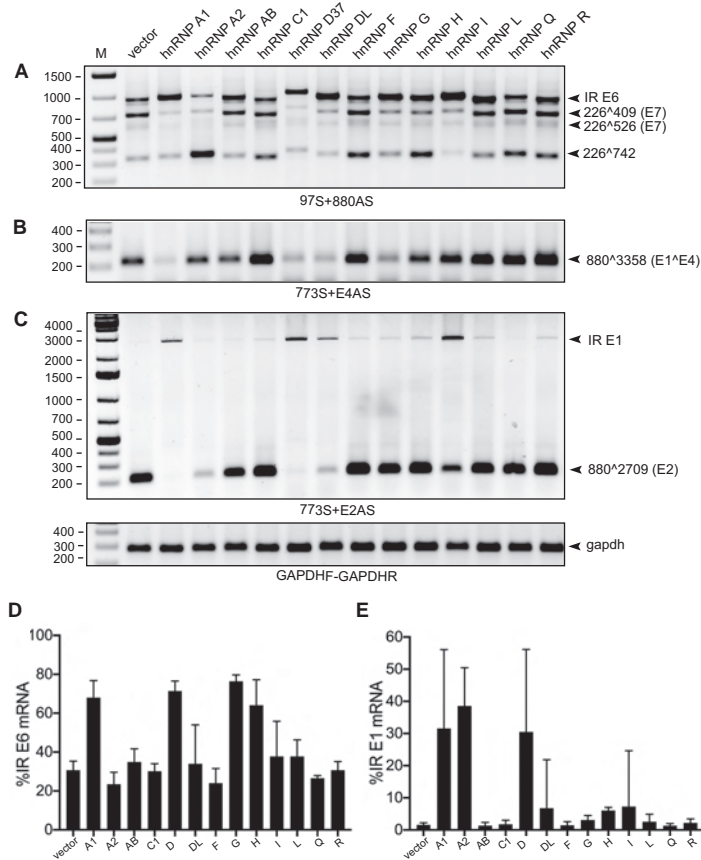


Figure 2. Identification of hnRNP D as a strong suppressor of HPV16 E6E7 and E1E2 mRNA splicing. (A–C) The effect of 13 different hnRNPs on HPV16 alternative mRNA splicing was determined by RT-PCR with different pairs of primers. HPV16 subgenomic plasmid pC97ELSLuc (Figure 1) was cotransfected individually with each of 13 different hnRNP expressing plasmid into HeLa cells. p37 isoform of hnRNP D (hnRNP D37) was used. Total RNA was extracted at 24 hpt and the various spliced HPV16 mRNAs were monitored by RT-PCR using HPV16-specific RT-PCR primer pairs: (A) E6/E7 mRNAs: 97S+880AS, (B) E4 mRNA: 773S+E4AS and (C) E1/E2 mRNAs: 773S+E2AS. Resulting RT-PCR products were electrophoresed on agarose gels and gel images are displayed. The schematic structures of the HPV16 spliced mRNAs and the locations of the HPV16-specific RT-PCR primers are indicated in Figure 1. Primer sequences are available in Supplementary Table S1. Each band was verified by sequencing and HPV16 spliced mRNA isoforms are indicated to the right of each gel. Validation of RT-PCR was also evaluated by reverse transcriptase negative samples using corresponding RNA samples. M: DNA size marker. Representative gel images of results repeated at least three-times are shown. (D) The effect of each hnRNP protein on the enhancement of HPV16 intron-retained (IR) E6 mRNA production shown in (A) was evaluated. A band intensity of each spliced isoform in (A) was quantitated described in Materials and Methods. The percentage of intron-retained E6 mRNA over a total sum of all four isoforms (intron-retained (IR) E6-, 226^409-, 226^526-, and 226^742-mRNA) was calculated. (E) The effect of each hnRNP protein on the enhancement of HPV16 intron-retained E1 mRNA production shown in (C) was evaluated as described in (D).

773S+E2AS) or by transfection with serially diluted hnRNP D40 plasmid (Supplementary Figure S2D). Taken together, we concluded that hnRNP D proteins inhibited splicing of HPV16 early mRNAs and promoted retention of introns encoding E1 and E6, respectively. Retention of the E1 intron occurred independently of intron retention in the E6 coding region.

The N-terminal domain of hnRNP D40 contributes to induction of intron-retained HPV16 E1 and E6 mRNAs

In an effort to elucidate how hnRNP D40 inhibits HPV16 mRNA splicing, we investigated the effect of deletion mutations in hnRNP D40 (Figure 4A). Analysis of FLAG-tagged, wild-type and mutant hnRNP D40-proteins revealed that all mutants were expressed in the transfected

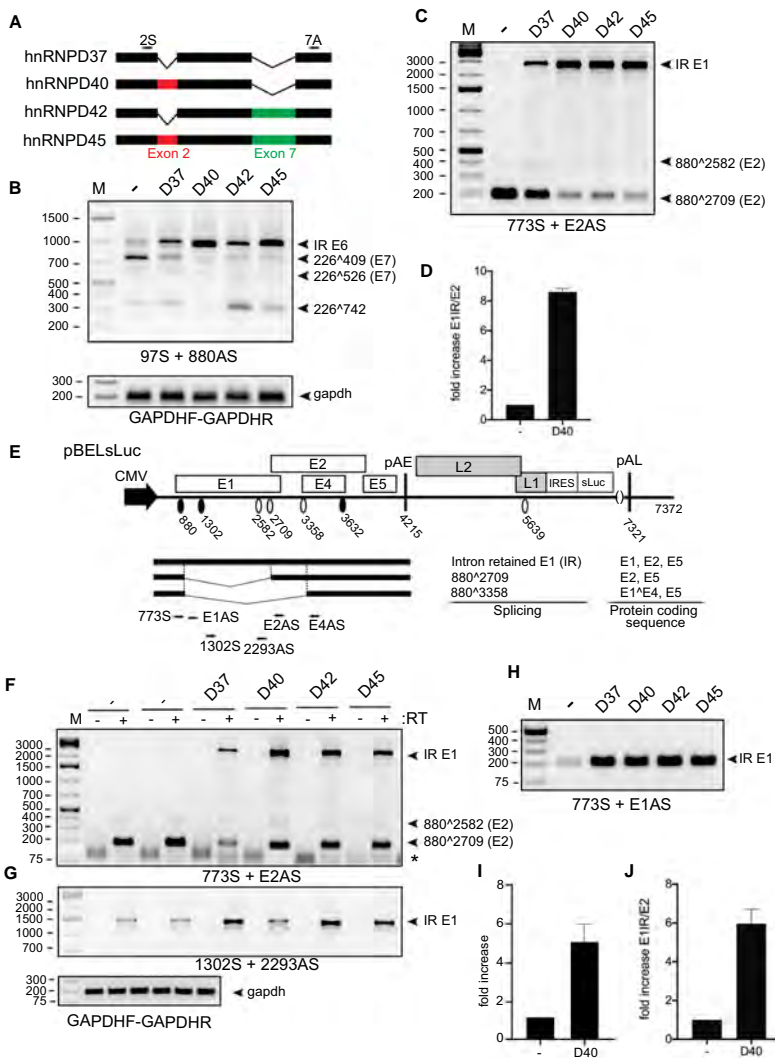


Figure 3. hnRNP D40 enhanced the production of intron-retained E6 and E1 mRNA more potently than the other hnRNP D isoforms. (A) Schematic structures of the hnRNP D protein isoforms. Each isoform differs by the inclusion of Exon 2 and/or Exon 7, respectively. Arrows: Primers' position used for Supplementary Figure S1F. (B and C) Effect of indicated hnRNP D isoforms on HPV16 alternative mRNA splicing was investigated by HPV16 RT-PCR on RNA extracted from HeLa cells transfected with pC97ELSLuc in the absence or presence of plasmids expressing either of the hnRNP D isoforms. HPV16 RT-PCR primer pairs: (B) 97S+880AS and (C) 773S+E2AS. (D) RT-qPCR to quantitate intron-retained E1 mRNAs and spliced E2 mRNAs from pC97ELSLuc in the absence or presence of hnRNP D40 was performed. Fold change of intron-retained E1 mRNAs over spliced E2 mRNAs is shown. (E) Schematic representation of HPV16 subgenomic plasmid pBELSLuc that encodes all HPV16 genes except E6 and E7 and is driven by the CMV promoter. HPV16 mRNAs polyadenylated at pAE are shown below pBELSLuc. Arrows: Annealing positions of HPV16 RT-PCR primers. Primer sequences are available in Supplementary Table S1. (F–H) Effect of indicated hnRNP D isoforms on HPV16 mRNA splicing was further investigated with plasmid pBELSLuc. pBELSLuc was cotransfected with indicated hnRNP D isoform expressing plasmids into HeLa cells. Total RNA was extracted and subjected to RT-PCR using different HPV16 RT-PCR primers pairs. RT-PCR was performed in the absence (–) or presence (+) of reverse transcriptase: (F) 773S+E2AS, (G) 1302S+2293AS and (H) 773S+E1AS. Only RT (+) samples for gapdh RT-PCR are shown. (I) RT-qPCR with primer pairs 773S+E1AS, normalized to GAPDH levels. (J) RT-qPCR to quantitate intron-retained E1 mRNAs over spliced E2 mRNAs from pBELSLuc in the absence or presence of hnRNP D40. Fold change of intron-retained E1 mRNAs over spliced E2 mRNAs is shown.

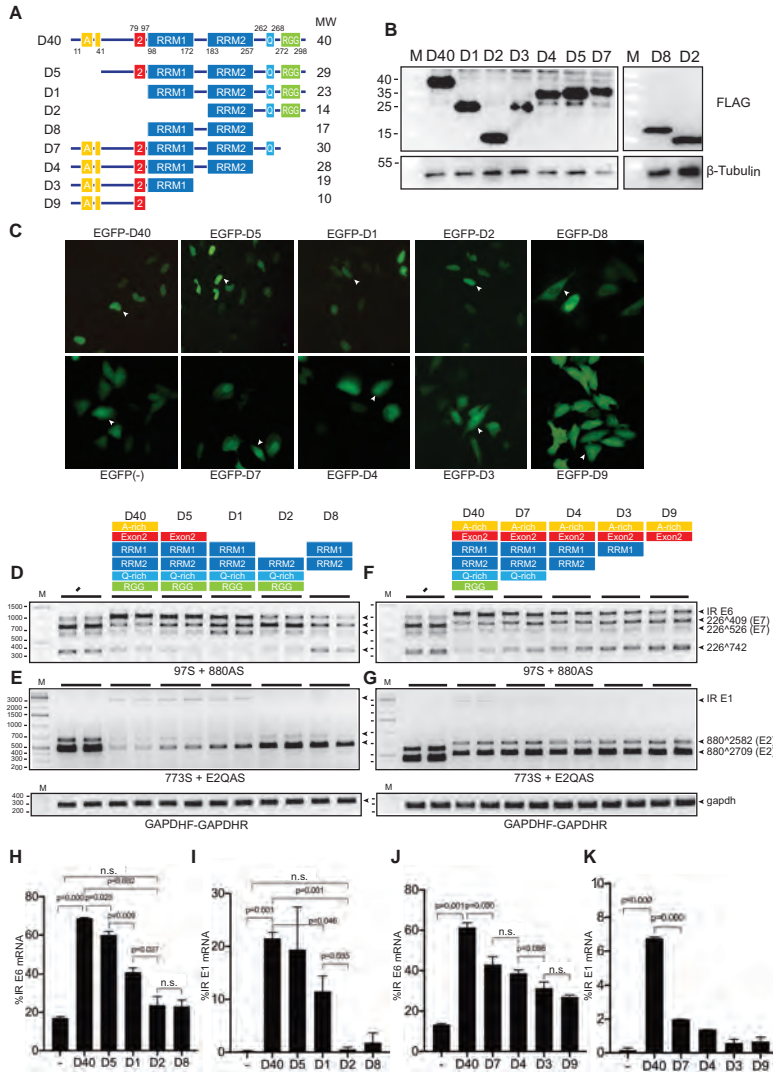


Figure 4. RRM1 domain of hnRNP D40 was required for HPV16 E6 and E1 intron retention whereas N-terminal and C-terminal domains contributed to modification of HPV16 early alternative mRNA splicing. (A) Schematic structures of wild-type hnRNP D40 and hnRNP D40 deletion mutants. All mutants were fused with FLAG-tag at its N-terminus. Box with A: Alanine(A)-rich region, box with 2: exon 2 region, box with RRM: RNA recognition motif (RRM) domain, box with Q: Glutamine(Q)-rich region and box with RGG: Arginine-Glycine/Arginine-Glycine-Glycine (RG/RGG) motif region. The names of each deletion mutant are indicated to the left and the predicted molecular weight of each protein is indicated to the right. (B) Western blotting against the FLAG epitope demonstrating protein expression of the hnRNP D40 deletion mutants depicted in (A). M: molecular weight marker. (C) Subcellular localization of indicated deletion mutants of hnRNP D40. Each mutant was fused to the C-terminus of EGFP. Representative cells expressing each mutant are highlighted by white arrowheads and shown in Supplementary Figure S3 with higher magnification. (D–G) Effect of FLAG-hnRNP D40 deletion mutants on HPV16 mRNA splicing was monitored by HPV16 RT-PCR on RNA extracted from HeLa cells transfected in the absence (–) or presence of hnRNP D40 or deletion mutants thereof. Representative gel images from experiments independently repeated three times are shown. (H and J) Percentage of intron-retained E6 mRNA over all four E6/E7 alternatively spliced isoforms quantified from (D) or (F). (I and K) Percentage of intron-retained E1 mRNA percentage over all three E1/E2 alternatively spliced isoforms quantified from (E) or (G). Proportions among all four isoforms is shown in Supplementary Figures S3B and S3C. Student's *t*-test was executed and obtained *P* values were displayed; n.s., no significance.

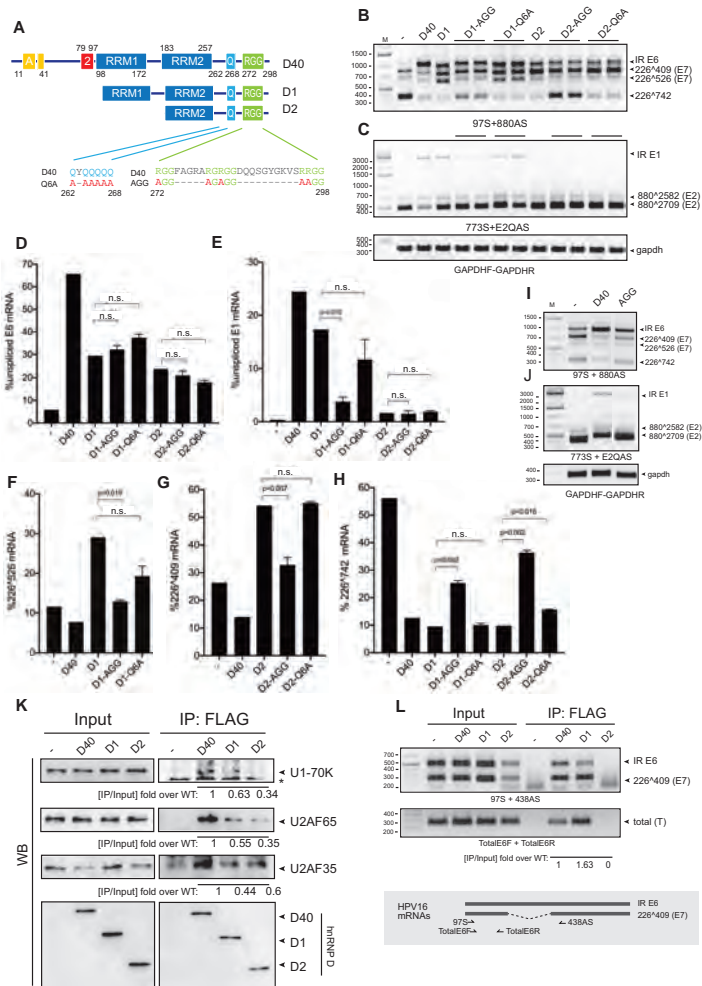


Figure 5. hnRNP D40 C-terminus contributes to HPV16 mRNA splicing control differently from its N-terminus. (A) Schematic structures of hnRNP D40 and various hnRNP D40 mutants with N-terminal deletions and C-terminal amino acid substitutions. Amino acid substitutions in the Q-rich region or the RG/RGG-motif region are indicated. Q: Glutamine, A: Alanine and R: Arginine. '-' indicates wild-type sequence. (B and C) Effect of hnRNP D40 or hnRNP D40 mutants on HPV16 E6/E7 mRNA splicing (B) or E1/E2 mRNA splicing (C) was monitored by RT-PCR with indicated primer pairs. (D–H) Percentage of an indicated HPV16 mRNA splice variant among all spliced isoforms of HPV16 E6/E7 or E1/E2 mRNAs was quantitated from (B) and (C) as described in Figure 2D and E: (D) percentage intron-retained E6 mRNAs, (E) percentage intron-retained E1 mRNAs, (F) percentage 226^409 mRNAs, (G) percentage 226^526 mRNAs and (H) percentage 226^742 mRNAs. Student's *t*-test was executed and obtained *P* values were displayed. n.s., no significance. (I and J) Effects of amino acid substitutions in the RG/RGG region of full-length hnRNP D40 on HPV16 mRNA splicing was monitored by HPV16 RT-PCR as described in (B). (K) Interactions between hnRNP D40 and cellular spliceosome factors U1-70K, U2AF65 and U2AF35 were investigated by co-immunoprecipitation (Co-IP) assay. Indicated FLAG-tagged, wild-type or mutant hnRNP D40 expression plasmid was transfected into HeLa cells. Whole cell extracts were subjected to IP by anti-FLAG antibody to purify the FLAG-protein interactome complex, followed by western blotting to detect the presence of endogenous cellular spliceosome factors in the complex. Band intensity was quantitated and IP efficiency was calculated as IP band intensity over corresponding Input band intensity. Fold change over WT D40 of IP efficiency is shown under each western blot image. (L) Interactions between hnRNP D40 and HPV16 mRNAs were investigated by UV-crosslinking and immunoprecipitation (CLIP) assay. HPV16 subgenomic plasmid pC97ELsLuc was cotransfected with FLAG-tagged wild-type or mutant hnRNP D40 expression plasmid into HeLa cells. UV-crosslinked whole cell extracts were subjected to IP by anti-FLAG antibody followed by extraction of RNA from the FLAG-protein interactome complex and RT-PCR using HPV16-specific RT-PCR primers, 97S+438AS or TotalE6F+TotalE6R. The schematic depiction of the location of the HPV16-specific RT-PCR primers is shown below the gel image. Fold change of CLIP efficiency of wild-type and mutant hnRNP D40s over wild-type D40 was calculated as described in (K).

HeLa cells (Figure 4B). To determine the subcellular localization of these hnRNP D40 proteins, a set of plasmids in which the hnRNP D40 open reading frame was fused to EGFP was also created. All hnRNP D40 proteins were localized to the nucleus, although mutant D8 differed from other mutants by showing stronger cytoplasmic than nuclear localization (Figure 4C and Supplementary Figure S3A). Wild-type hnRNP D40 and the N-terminally deleted proteins (D1, D2 and D5) were present primarily in the nucleus, whereas C-terminally deleted proteins (D3, D4, D7 and D9) also showed substantial cytoplasmic staining (Figure 4C and Supplementary Figure S3A). The increased cytoplasmic localization of the C-terminally deleted proteins was a result of deletion of the RGG-domain as substitutions of the arginines in four RGG-motifs for alanine had the same effect (Figure 5A and Supplementary Figure S4B). We concluded that the nuclear localization of the mutant hnRNP D40 proteins allowed us to determine the effect of these proteins on HPV16 mRNA splicing.

Co-transfection of HPV16 reporter plasmid pC97ELsLuc with hnRNP D40 or either of the N-terminally deleted hnRNP D40 plasmids (D1, D2, D5 and D8) (Figure 4D) revealed that all deletions reduced the ability of hnRNP D40 to promote production of HPV16 intron-retained E6 mRNAs (Figure 4D,H) and intron-retained E1 mRNAs (Figure 4E,I). Quantitated percent of all isoforms is shown in Supplementary Figure S3B and C. Longer exposed images of Figure 4E,G are available in Supplementary Figure S3D. Quantitation showed that the reduction in splicing was statistically significant (Figure 4H,I). More specifically, we found that deletion of the N-terminal Alanine-rich region of hnRNP D40 (mutant D5) reduced the ability of hnRNP D40 to promote production of intron-retained HPV16 E6 mRNAs (Figure 4D,H). Deletion of the exon-2 coding region, in addition to the alanine-rich region, as in hnRNP D40 mutant D1 further reduced induction of intron-retained E6 mRNA (Figure 4D,H) and intron-retained E1 mRNA (Figure 4E, I) and unexpectedly activated splicing from SD226 to SA526 (226 \wedge 526) (Figure 4D and Supplementary Figure S3B). Deletion also of RRM1 domain, as in hnRNP D40 mutant D2, abolished induction of both intron-retained E6 mRNAs (Figure 4D,H) and intron-retained E1 mRNAs (Figure 4E, I) and unexpectedly activated splicing from SD226 to SA409 (226 \wedge 409) (Figure 4D and Supplementary Figure S3B). Overexpression of a mutant hnRNP D40 protein consisting only of RRM1 and RRM2 domains (hnRNP D40 mutant D8) had only a minor effect on HPV16 mRNA splicing (Figure 4D,E), which in this case may be explained by the preferential localization of D8 to the cell cytoplasm (Figure 4C). We concluded that the N-terminus of hnRNP D40 contributed significantly to the inhibition of HPV16 early mRNA splicing and production of E1- and E6-mRNAs and that RRM1 domain was of particular importance.

Interestingly, we found that hnRNP D40 mutant D1 unexpectedly gained the ability to activate splicing to HPV16 3'-splice site SA526 (226 \wedge 526) (Figure 4D), while retaining some of its ability to inhibit splicing of both E6/E7 mRNAs and E1/E2 mRNAs (Figure 4D,E). These results suggested that deletion of the N-terminal, alanine-rich region and the

exon 2 coding sequence weakened the ability of hnRNP D40 to inhibit splicing, while retaining interactions with the splicing machinery, of which the latter was documented by the enhanced splicing to SA526. Hence, interactions of hnRNP D40 with the splicing machinery and its ability to inhibit splicing could be separated. This interpretation was reinforced by the phenotype of mutant D2 that lost much of its inhibitory effect on the E6/E7 mRNA splicing (Figure 4D,H) and all of its inhibitory effect on E1 mRNA splicing (Figure 4E,I) but gained the ability to promote splicing to HPV16 SA409 (226 \wedge 409) (Figure 4D). To determine if similar results were obtained in cells other than HeLa cells, we co-transfected pC97ELsLuc plasmid into 293T cells in the presence of hnRNP D40 or the two N-terminal deletions of hnRNP D40 D1 and D2. Similar results were obtained in the two cell lines (Supplementary Figure S4A): D1 and D2 had lost much of their splicing inhibitory function and activated alternative splicing of the HPV16 E6/E7 mRNAs (Supplementary Figure S4A). RRM1 and RRM2 apparently played a decisive role in the selection of HPV16 splice site since the presence of RRM1 and RRM2 in D1 activated splicing to HPV16 SA526 (226 \wedge 526) (Figure 4D), whereas RRM2 alone in D2 activated splicing to HPV16 SA409 (226 \wedge 409) (Figure 4D). We concluded that hnRNP D40 RRM1 was essential for efficient splicing inhibition and induction of intron-retained HPV16 E1 and E6 mRNAs and that the N-terminal alanine-rich region and the exon-2 region of hnRNP D40 contributed to splicing inhibition possibly via different mechanisms.

Nested C-terminal deletions in hnRNP D gradually reduced ability of hnRNP D to inhibit HPV16 mRNA splicing

In contrast to N-terminal deletions of hnRNP D40, either of the C-terminally deleted hnRNP D40 plasmids D3, D4, D7 and D9 (Figure 4A) promoted production of intron-retained E6 mRNAs (Figure 4F, J), although less efficiently than wild-type hnRNP D40. Inhibition of splicing to SA409 and in particular to SA742 was gradually reduced with larger C-terminal deletions in hnRNP D40 (Figure 4F and Supplementary Figure S3B). Even plasmid D9 that contained only the N-terminal alanine-rich region and the exon 2 coding region retained some splicing inhibitory activity (Figure 4F,J). Similarly, splicing inhibition of the HPV16 E1/E2 mRNAs was gradually lost with increasing size of hnRNP D40 C-terminal deletions (Figure 4G,K). Similar results were obtained with the two-C-terminal deletions D3 and D4 in other cell line (Supplementary Figure S4A), corroborating the observation that C-terminal deletions reduced the inhibitory effect on alternative splicing of HPV16 E6/E7 mRNAs. Taken together, these results supported the idea that the N-terminal region of hnRNP D40 played an important role in inhibition of HPV16 E1/E2 and E6/E7 mRNA splicing and indicated that this effect was enhanced by the C-terminus, suggesting that the C-terminus of hnRNP D40 may interact with the splicing machinery.

The RGG-domain of hnRNP D40 contributes to its ability to inhibit HPV16 mRNA splicing

In an effort to understand how hnRNP D40 interacted with the splicing machinery, we utilized hnRNP D40 mutants

D1 and D2 for further mutagenesis. D1 and D2 have N-terminal deletions but retain either both RRM1 and RRM2 (D1), or only RRM2 (D2). Both D1 and D2 contain an intact C-terminus with the Q-rich region and the RG/RGG-rich 'RGG'-region (Figure 5A). We separately introduced point mutations in the glutamine-rich region (Q) and in the RGG-region in D1 and D2 (Figure 5A). Substitutions of arginine for alanine in four RG/RGG motifs in the RGG-region in D1 (D1-AGG) did not affect the ability of D1 to promote production of intron-retained E6 mRNA (Figure 5B,D) but abrogated the splicing enhancing effect to HPV16 SA526 (226 \wedge 526) (Figure 5B,F) and reduced the ability of D1 to promote production of intron-retained E1 mRNAs (Figure 5C, E). In contrast, substitutions of all six glutamines in the Q-rich region for alanine (D1-Q6A), did not affect the ability of D1 to enhance splicing to HPV16 SA526 (Figure 5B,F), nor did they affect the ability of D1 to promote production of intron-retained E1 mRNAs (Figure 5C,E). We concluded that the hnRNP D40 RGG-region was important for enhancement of E6 mRNA splicing to SA526, as well as for inhibition of E1 mRNA splicing. These results were supported by the analysis of hnRNP D40-D2 with the same mutations. Mutations in the RGG domain in D2 (D2-AGG) (Figure 5A) alleviated splicing enhancement of SA409 (Figure 5B,G). Interestingly, both D1-AGG and D2-AGG restored HPV16 E6 mRNA splicing to HPV16 SA742 (226 \wedge 742) compared to their corresponding parental mutants, D1 and D2, respectively (Figure 5B,H), which is in line with the observation in Figure 4F that sequential C-terminal deletions restored 226 \wedge 742. The D2-AGG mutations did not substantially affect production of intron-retained E1 mRNAs (Figure 5C,E) or intron-retained E6 mRNAs (Figure 5B,D). Similarly to D1-Q6A and D1, D2-Q6A and D2 had similar phenotype (Figure 5F, G, H). Our results demonstrated that the RGG-domain contributed to the ability of D1 and D2 to activate splicing to alternative HPV16 splice sites in the absence of the splicing inhibitory part in the N-terminus of full-length hnRNP D40. We speculated that the C-terminus of hnRNP D40 interacted with the splicing machinery and that the inhibition of splicing and the interactions with the splicing machinery were two separate properties of hnRNP D40. Finally, substitutions of arginine for alanine in the RGG-region in the context of full-length hnRNP D40 showed a reduction of the splicing inhibitory effect of hnRNP D40 on E6 and E1 mRNAs (Figure 5I,J). The effects in the context of the full-length protein were relatively subtle but reproducible and significant (Supplementary Figure S4C). Taken together, our results suggested that the C-terminal RGG-domain of hnRNP D40 contributed to the splicing inhibitory function of hnRNP D40 by interacting with the splicing machinery, while the inhibitory effect on splicing was mediated by the N-terminus of hnRNP D40.

hnRNP D40 interacts with complex A components of the splicing machinery

To determine if hnRNP D40 interacted with the splicing machinery, we investigated if hnRNP D40 co-immunoprecipitated components of the cellular spliceosome. HeLa cells were transfected with plasmids expressing

full-length (D40) or D1 or D2 mutants of hnRNP D40. Cell extracts were prepared and subjected to immunoprecipitation by anti-FLAG antibody followed by western blotting with antibodies to spliceosome components U1-70K, U2AF65 or U2AF35, or with anti-FLAG antibody. Our results revealed that wild-type hnRNP D40 co-immunoprecipitated cellular U1 snRNP component U1-70K and U2 auxiliary factors U2AF65 and U2AF35 (Figure 5K). The two N-terminal mutants of hnRNP D40, D1 and D2, retained the interactions with cellular U1 snRNP components U1-70K and U2AF65 and U2AF35, albeit less efficiently than full-length hnRNP D40 (Figure 5K). Furthermore, the branch-point adenosine recognizing complex component, splicing factor 3b (SF3b) and one of the translation regulators, cytosolic poly(A)-binding protein-1 (PABP-C1) interacted with full-length hnRNP D40 as well as with the N-terminal mutants D1 and D2 (Supplementary Figure S4D). These results not only demonstrated that the C-terminus of hnRNP D40 was sufficient for interactions with these spliceosomal complex A components as well as with downstream RNA-processing regulators but also revealed that the N-terminus of hnRNP D40 contributed to the efficiency of these interactions. We concluded that hnRNP D40 protein interacted with complex A components of the splicing machinery.

hnRNP D40 interacts with HPV16 mRNAs in an hnRNP D40 RRM1-dependent manner

Next, we investigated if hnRNP D40 or deletion mutants thereof (D1 and D2) interacted with HPV16 RNA. Plasmids expressing FLAG-tagged hnRNP D40, D1 or D2 were co-transfected with HPV16 reporter plasmid pC97ELsLuc into HeLa cells followed by UV-crosslinking and Immunoprecipitation (CLIP) with affinity purification of FLAG tagged hnRNP D40-RNA complexes. FLAG-antibody-immunoprecipitated UV-crosslinked RNA was extracted and subjected to RT-PCR with HPV16 specific primers (Figure 5L). The results revealed that immunoprecipitation of wild-type hnRNP D40 also immunoprecipitated HPV16 RNA detected by HPV16 primer pair TotalE6F+TotalE6R that amplifies cDNA representing all HPV16 mRNAs transcribed from pC97ELsLuc (Figure 5L). Analysis of the HPV16 RNA with primer pair 97S+438AS that discriminates between intron-retained and spliced HPV16 E6 mRNAs revealed that hnRNP D40 interacted with both intron-retained and spliced E6 mRNAs (Figure 5L). Deletion of N-terminal alanine-rich and exon-2 region (D1 mutant) did not affect the interactions with HPV16 RNA but deletion of RRM1 did (D2 mutant) (Figure 5L). We concluded that hnRNP D40 interacts with HPV16 mRNAs directly in an hnRNP D40 RRM1-dependent manner.

hnRNP D40 interacts with the entire intronic region of the E6 coding region as well as with a previously identified HPV16 splicing silencer in E7 region

Having established that hnRNP D40 interacts with HPV16 mRNAs (Figure 5L) and with components of the splicing machinery (Figure 5K) thereby inhibiting HPV16 early

mRNA splicing and promoting intron retention and production of intron-retained HPV16 E1 and E6 mRNAs, we wished to identify the HPV16 sequences that were targeted by hnRNP D40. For this purpose, we focused on the E6/E7 coding region as it is shorter than the E1 coding region. To this end, we performed RNA-mediated pull downs of hnRNP D40 with overlapping, biotinylated RNA oligonucleotides that covered the E6- and E7-coding regions from HPV16 nucleotide position 178 to 875 (Figure 6A,B). This experimental approach was chosen since hnRNP D40 may not necessarily bind directly to HPV16 mRNAs. Cell extracts were prepared from HeLa cells transfected with FLAG-tagged hnRNP D40. These extracts were subjected to RNA-mediated protein pull downs with the RNA oligos indicated in Figure 6A,B, followed by western blotting with anti-FLAG antibody (Figure 6A,B). The results revealed that hnRNP D40 was specifically pulled down by a subset of the RNA oligos. Of particular interest was that all RNA oligos spanning the E6-encoding intron between HPV16 5'-splice site SD226 and 3'-splice site SA409 pulled down hnRNP D40 (Figure 6A,B, highlighted with a double-headed arrow). Furthermore, RNA oligos 579-604 and 594-620 that spanned a previously identified splicing silencer (46) also pulled down hnRNP D40 (Figure 6A,B). Additional RNA oligos that pulled down hnRNP D40 did not coincide with known RNA elements except for oligos 721-755 that encoded HPV16 3'-splice site SA742 (Figure 6A,B). We concluded that interactions of hnRNP D40 with a previously identified splicing silencer located between HPV16 nucleotide positions 594 and 604, as well as with the intronic sequences in the E6 coding region (243-445), correlated with the ability of hnRNP D40 to inhibit splicing between SD226 and SA409 and to promote retention of the intron between the same two splice sites.

hnRNP D40 promotes intron retention of HPV16 E6 mRNAs independently of the HPV16 E1 sequences

Since hnRNP D40 interacted with sequences in the E6 coding region, we investigated if hnRNP D40 could inhibit E6 mRNA splicing and induce intron retention of the E6 mRNAs in the absence of E1 sequences. We had previously showed that hnRNP D40 inhibited splicing and induced intron retention of the E1/E2 mRNAs in the absence of E6 and E7 sequences (see effect of hnRNP D40 on pBELsLuc (Figure 3E–J, Supplementary Figure S2)). To this end we co-transfected hnRNP D40 plasmid with HPV16 plasmid pX656 that starts at HPV16 nucleotide position 97 and ends at HPV16 position 656 and does not overlap with the previously used HPV16 plasmid pBELsLuc (Supplementary Figure S5A and B). hnRNP D40 inhibited splicing of HPV16 E6/E7 mRNAs produced from pX656, albeit less efficiently than it inhibited splicing of E6/E7 mRNAs produced from pC97ELsLuc (Figure 6C,D), demonstrating that hnRNP D40 can inhibit splicing in the HPV16 E6/E7-coding region independently of E1 sequences. The pX656 contained a previously identified splicing silencer located between HPV16 positions 594 and 604 in addition to the E6-encoding intron with HPV16 splice sites SD226, SA409 and SA526, but the pull-down experiments displayed in Figure 6A demonstrated that the major sites of interactions be-

tween hnRNP D40 and HPV16 E6/E7 mRNAs were within the E6 intron between SD226 and SA409. Thus, our results predicted that a smaller plasmid containing only a part of the E6 coding region encompassing SD226, SA409 and the E6 intron (without splicing silencer in the E7 coding region) should also respond to hnRNP D40 overexpression. HPV16 plasmid pX478 that contains HPV16 sequences from HPV16 nucleotide position 97 to position 478 (Supplementary Figure S5C) was transfected into HeLa cells in the absence or presence of hnRNP D40 expression plasmid. RT-PCR analysis of HPV16 mRNAs revealed that hnRNP D40 inhibited splicing and promoted production of intron-retained HPV16 mRNAs from this minimal HPV16 pX478 expression plasmid (Figure 6C,D). As expected, hnRNP D40 deletion mutant D1 also inhibited splicing of the HPV16 mRNAs produced from plasmid pX478 but less efficiently than wild-type hnRNP D40 (Figure 6C,D). We concluded that hnRNP D40 promoted intron retention of HPV16 E6 mRNAs independently of HPV16 E1 sequences.

hnRNP D40 increases the levels of HPV16 intron-retained E6 mRNAs in the cytoplasm

Our results also suggested that hnRNP D40 should increase E6 protein production at the expense of E7 protein production as a result of its splicing inhibitory function of HPV16 E6/E7 mRNAs. Analysis of HPV16 E6 and E7 protein production from pC97ELsLuc transfected in the absence or presence of hnRNP D40 expression plasmid unexpectedly revealed that overexpression of hnRNP D40 suppressed production of both E6 and E7 protein (Figure 6E). Similar results were obtained with hnRNP D40 mutant D1 (Figure 6E). hnRNP D40 also inhibited E6 and E7 protein production from a smaller HPV16 expression plasmid named pXH856F that encodes only E6 and E7 (Figure 6F and Supplementary Figure S5D). This effect was independent of the splicing process itself since overexpression of hnRNP D40 inhibited E6 and E7 protein production also from a version of the pXH856F plasmid in which the major HPV16 5' splice site SD226 was mutationally inactivated, plasmid pXH856SDmF (Figure 6F and Supplementary Figure S5E). These results suggested that hnRNP D40 was associated with HPV16 mRNAs also in the absence of RNA splicing and as such, could potentially affect other RNA processing steps, thereby explaining the reduction in E6 protein levels despite the increase in the levels of intron-retained E6 mRNAs.

To determine if hnRNP D40 affected export of intron-retained E6 mRNAs, we analyzed the subcellular distribution of the intron-retained E6 mRNAs produced from the splicing deficient plasmid pXH856SDmF in the absence or presence of hnRNP D40. The intron-retained E6 mRNAs produced from pXH856SDmF were present in the nucleus and in the cytoplasm (Figure 6G). Unexpectedly, the levels and subcellular distribution of E6 mRNAs were similar in the absence or presence of hnRNP D40 (Figure 6G,H). RT-PCR of intron-retained and spliced actin RNA served as a control for proper fractionation of the transfected cells (Figure 6G). We concluded that the nuclear export of intron-retained E6 mRNAs transcribed from the splicing deficient plasmid pXH856SDmF was not consid-

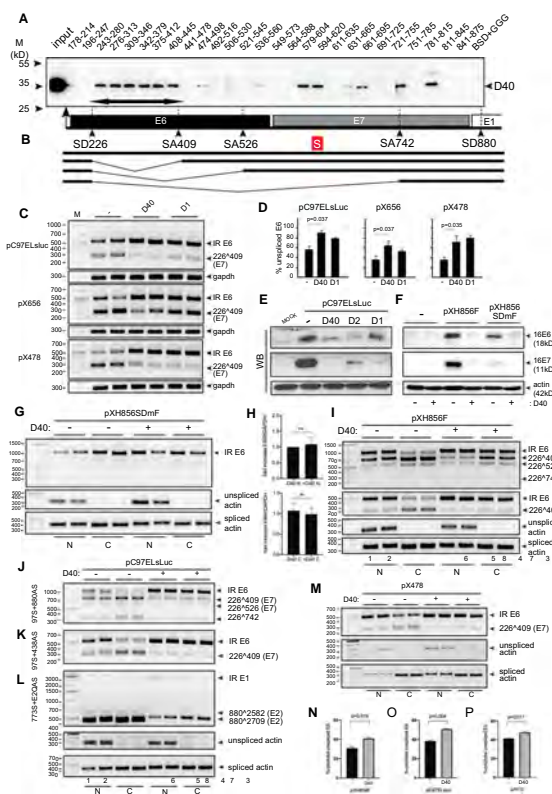


Figure 6. hnRNP D40 interacted with HPV16 E6/E7 RNA sequences *in vitro* and influenced HPV16 E6/E7 mRNA processing. (A) RNA oligonucleotide-mediated pull-down assay using sequential short RNA oligos spanning the entire HPV16 E6- and E7- coding regions. The numbers in each RNA-oligo name represent HPV16 nucleotide positions and refer to nucleotide numbers in the reference HPV16 genome (GenBank: K02718.1). Thick double-headed arrow: intronic region of E6/E7 mRNAs located between SD226 and SA409. (B) Schematic representation of RNA splicing in the HPV16 E6 and E7 coding region. Intron-retained E6 mRNA is depicted as well as spliced HPV16 mRNAs 226^Δ409, 226^Δ526 and 226^Δ742. Box with S: previously identified E6/E7 mRNA splicing silencer element (46). RNA oligo sequences are available in Supplementary Table S3. (C) The effect of hnRNP D40 on the splicing of HPV16 E6/E7 mRNA produced from pC97ELsLuc as well as shorter plasmids pX656 and pX478 was monitored by RT-PCR. The schematic maps of pX656 and pX478 plasmids and the positions of RT-PCR primers are shown in Supplementary Figure S5. (D) The percentage of intron-retained E6 mRNA over a total sum of all spliced isoform quantitated from (C) are shown. (E) HPV16 E6 and E7 proteins expressed in HeLa transfected with pC97ELsLuc together with indicated wild-type or mutant hnRNP D40. Western blotting was performed using anti-HPV16 E6 or E7 specific antibodies. (F) HPV16 E6 and E7 proteins produced in 293T cells transfected with pXH856F, which encodes intact HPV16 E6- and E7- coding regions, or with pXH856SDmF, which contains a mutationally inactivated SD226 splice site, in the absence (-) or presence (+) of cotransfected hnRNP D40 expression plasmid. Western blotting was performed with anti-HPV16 E6 or E7 specific antibodies. Wild-type SD226 (GAGIGUUAUGA; a vertical line indicates 5' splice site) was changed to (GAGIGCCUAUGA; underline indicates changed nucleotide) in SD226 mutant. Schematic maps of pXH856F and pXH856SDmF are shown in Supplementary Figure S5. (G) Subcellular distribution of intron-retained E6 mRNAs produced from pXH856SDmF in the absence (-) or presence (+) of hnRNP D40 plasmid. Nuclear and cytoplasmic fractions were prepared from the transfected cells and RNA was extracted and subjected to HPV16 RT-PCR using primers TotalE6F and 757AS (for primer location, see Figure 1C and Supplementary Figure S5E). Cell fractionation was validated by RT-PCR analysis of unspliced actin RNA located exclusively in the nuclear fraction while spliced actin mRNAs found in both fractions. (H) qPCR using cDNA samples from (G) to evaluate levels of intron-retained E6 mRNAs in nuclear and cytoplasmic fractions in the absence or presence of hnRNP D40. Primers TotalE6F and 234AS were used (for primer location, see Figure 1C and Supplementary Figure S5E). (I) Nuclear (N) and cytoplasmic (C) distribution of HPV16 alternatively spliced mRNAs produced from pXH856F in the absence (-) or presence (+) of hnRNP D40 plasmid. RT-PCR was performed with primers: 97S+757AS (top panel) or 97S+438AS (second panel). Primer locations are shown in Figure 1C and Supplementary Figure S5D. (J-L) Nuclear (N) and cytoplasmic (C) distribution of HPV16 alternatively spliced mRNAs produced from pC97ELsLuc in the absence (-) or presence (+) of cotransfected hnRNP D40 plasmid. RT-PCR was performed with the following primers: (J) 97S+880AS, (K) 97S+438AS or (L) 773S+E2QAS. (M) Nuclear (N) or cytoplasmic (C) distribution of HPV16 alternatively spliced mRNAs produced from HPV16 plasmid pX478 in the absence (-) or presence (+) of cotransfected hnRNP D40 plasmid, followed by RT-PCR using primers, 97S+438AS. Schematic map of plasmid pX478 and location of HPV16 RT-PCR primers are shown in Supplementary Figure S5C. (N-P) Percentage of cytoplasmic intron-retained E6 mRNAs over the total sum of nuclear and cytoplasmic intron-retained E6 mRNAs in the absence (-) or presence of hnRNP D40 quantitated on (N) pXH856F from (I), (O) pC97ELsLuc from (K) or (P) pX478 from (M). Student t-test was executed and obtained *P* values were displayed. n.s., no significance.

erably affected by hnRNP D40 (Figure 6G,H). Thus, inefficient nuclear mRNA export of intron-retained E6 mRNA in the presence of hnRNP D40 could not explain the significant inhibition of E6 protein production exerted by hnRNP D40 (Figure 6F). To exclude the possibility that the mutation of HPV16 5'-splice site SD226 in pXH856SDmF affected the RNA export regulation mediated by hnRNP D40, the effect of hnRNP D40 on the subcellular distribution of E6 and E7 mRNAs was investigated using pXH856F encoding wild-type SD226. The inhibitory effect of hnRNP D40 on HPV16 E6/E7 mRNA splicing was observed however the effect appeared to be mainly against 226 \wedge 409 (Figure 6J). We observed that the intron-retained E6 mRNAs were to a large extent restricted to the nuclear fraction while the levels of the spliced, HPV16 226 \wedge 409-mRNAs were higher in the cytoplasm than in the nucleus (Figure 6I, top panel, lanes 1–4). In the presence of hnRNP D40, primarily intron-retained E6 mRNAs were detected both in the nucleus and cytoplasm (Figure 6I, top panel, lanes 5–8). The results were confirmed using a primer pair that only detected intron-retained E6 mRNAs and mRNAs spliced between splice sites SD226 and SA409 (226 \wedge 409) (Figure 6I, second panel), demonstrating a statistically significant increase of cytoplasmic intron-retained E6 distribution in the presence of hnRNP D40 (Figure 6N). These results indicated that hnRNP D40 did not inhibit the nuclear export of intron-retained E6 mRNAs. Thus, inhibition of nuclear export of intron-retained E6 mRNAs could not explain the reduction of E6 protein levels in the presence of hnRNP D40 observed in Figure 6E and F. This observation was consistent with results obtained using HPV16 subgenomic plasmid pC97ELsLuc (Figure 6J, K and O). Furthermore, we showed that hnRNP D40 enhanced the levels of cytoplasmic intron-retained E6 mRNAs also from the smaller HPV16 subgenomic plasmid pX478 (Figure 6M). This increase was statistically significant (Figure 6P). Taken together, we concluded hnRNP D40 had multiple impacts and affected various steps of HPV16 E6 and E7 mRNA processing: hnRNP D40 interacted with E6/E7 mRNAs at multiple positions, it inhibited E6/E7 mRNA splicing independently of the E1 region, and increased levels of intron-retained E6 mRNAs in the cytoplasm. However, hnRNP D40 also significantly reduced the levels of E6 protein produced from the intron-retained E6 mRNAs.

hnRNP D40 increased the levels of HPV16 intron-retained E1 mRNAs in the cytoplasm

In addition to the intron-retained E6 mRNAs produced from pC97ELsLuc, intron-retained E1 mRNAs transcribed from pC97ELsLuc were also increased in both nucleus and cytoplasm (Figure 6L). To confirm this effect also in the absence of E6/E7 sequences on the mRNAs, we used HPV16 subgenomic reporter plasmid pBELsLuc that lacks E6/E7 genes (Figure 3E). pBELsLuc was transfected into HeLa cells in the absence or presence of hnRNP D40 plasmid followed by fractionation into nuclear and cytoplasmic fractions prior to RNA extraction and RT-PCR. As can be seen in Figure 7A, spliced E2 mRNA (880 \wedge 2709) levels were reduced in the presence of hnRNP D40, as expected, but they were present in both nuclear and cytoplasmic fractions, al-

though predominantly in the cytoplasm. In sharp contrast to the spliced E2 mRNAs, the intron-retained E1 mRNAs were detected primarily in the presence of hnRNP D40 (Figure 7A). In the presence of hnRNP D40, the intron-retained E1 mRNAs were present in both the nuclear and cytoplasmic fractions (Figure 7A), demonstrating that hnRNP D40 not only inhibited splicing of the HPV16 E1/E2 mRNAs and promoted intron retention to generate intron-retained E1 mRNAs, it also allowed these mRNAs to be exported to the cytoplasm. A similar effect was observed with hnRNP A1 (Figure 7A) an hnRNP protein evolutionarily close to hnRNP D protein family (54) and initially shown to inhibit HPV16 E6/E7 and E1/E2 mRNA splicing (Figure 2A,C). Actin mRNAs served as controls for cellular fractionation: unspliced actin mRNAs were present primarily in the nuclear fraction, whereas spliced actin mRNAs were detected in both fractions (Figure 7B). Nuclear restricted Lamin B and SRSF2 and the shuttling SRSF1 protein served as additional controls for cellular fractionation (Supplementary Figure S6). We also performed RT-qPCR with primers 773s and E1AS on the cytoplasmic fractions, a primer pair that is specific for intron-retained HPV16 E1 mRNAs (Figure 1C). The levels of HPV16 intron-retained E1 mRNAs in the cytoplasm increased more than four-fold in the presence of hnRNP D40 (Figure 7C). We concluded that hnRNP D40 inhibited HPV16 E1/E2 mRNA splicing, promoted intron retention to generate intron-retained E1 mRNAs and enhanced the appearance of the E1 mRNAs in the cytoplasm independently of E6/E7 sequences on the mRNAs.

The HPV16 subgenomic plasmid pBELEN encodes only the HPV 16 E1 gene including all splice sites in the E1 coding region (Supplementary Figure S7B). pBELEN responded to hnRNP D40 by enhanced production of intron-retained E1 mRNA (Supplementary Figure S7D). RT-PCR analysis of HPV16 mRNAs produced from pBELEN in the absence or presence of hnRNP D40, revealed that the level of HPV16 intron-retained E1 mRNA increased in both nucleus and cytoplasm, whereas the levels of HPV16 spliced E2 mRNAs (880 \wedge 2709) were reduced in the presence of hnRNP D40 (Supplementary Figure S7E). This result indicated that redistribution of intron-retained E1 mRNA to the cytoplasm was enhanced in the presence of hnRNP D40 (Supplementary Figure S7F). For further confirmation, we generated a plasmid that produced a shorter intron-retained E1 mRNA that would be easier to detect and quantitate by RT-PCR since the predicted size of the entire intron-retained E1 mRNA produced from pBELEN is relatively big (>2 kb). We therefore introduced a deletion in the E1 coding region in pBELEN, between the HPV16 major splice sites SD880 and SA2709, to generate a smaller plasmid named pBELENdE1 (Supplementary Figure S7C). Detection of intron-retained 'E1' mRNAs was considerably improved when pBELENdE1 was used (Figure 7D lanes 1–4, compared to Figure 7A). RT-PCR analysis of HPV16 mRNAs produced from pBELENdE1 in the absence or presence of hnRNP D40, revealed that the levels of HPV16 intron-retained mRNAs were increased in both nucleus and cytoplasm (Figure 7D), an increase that was statistically significant (Figure 7E), whereas the levels of HPV16 spliced E2 mRNAs (880 \wedge 2709) were reduced in the presence of hnRNP D40, and were located primarily in the cytoplasm.

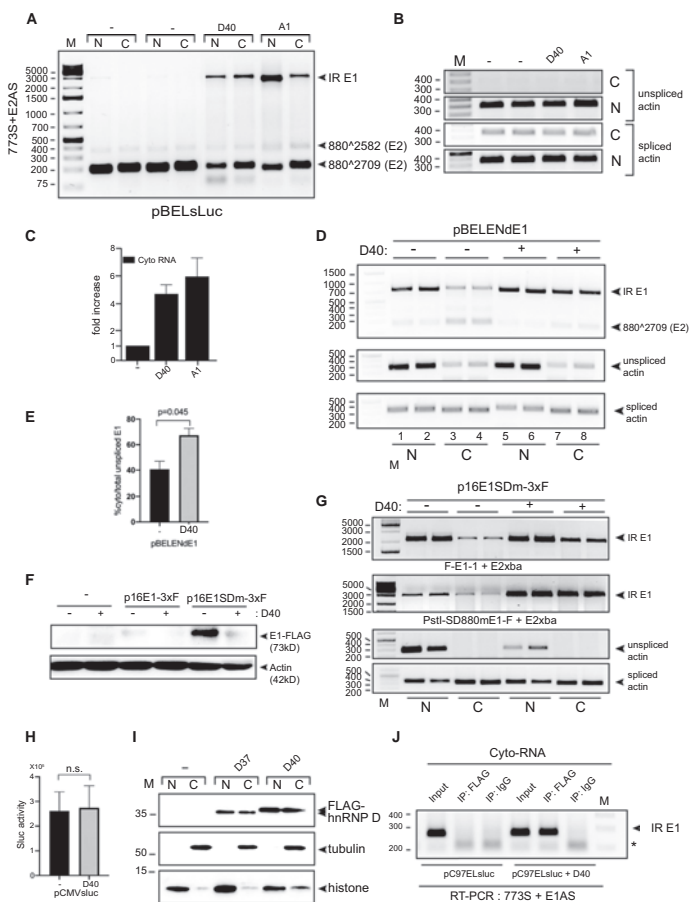


Figure 7. hnRNP D40 upregulated levels of HPV16 intron-retained E1 mRNAs in the cytoplasm and interacts with these mRNAs in the cytoplasmic fraction. (A) HeLa cells were transfected with pBEsLuc plasmid in the absence (–) or presence of plasmids expressing either hnRNP D40 or hnRNP A1. Nuclear (N) and cytoplasmic (C) fractions were prepared from the transfected cells and RNA was extracted and subjected to HPV16-specific RT-PCR using primer pair 773S+E2AS. (B) Cellular fractionation was validated by analysis by RT-PCR of unspliced and spliced actin mRNAs. (C) Levels of HPV16 intron-retained E1 mRNAs in the cytoplasmic fraction were determined by RT-PCR with primers 773S+E1AS. Quantification of electrophoresed RT-PCR products were performed as described and normalized to RT-PCR products representing cytoplasmic spliced actin mRNA. Fold over control (–) is shown. (D) Effect hnRNP D40 on subcellular localization of HPV16 intron-retained E1 mRNAs and spliced E2 mRNAs produced from HPV16 subgenomic plasmid pBELENdE1 (for schematic representation of pBELENdE1, see Supplementary Figure S7C). pBELENdE1 was transfected into HeLa cells in the absence (–) or presence (+) of hnRNP D40 plasmid. RT-PCR was performed on nuclear or cytoplasmic RNA using HPV16-specific primers pair 773S+E2Xba that detects HPV16 intron-retained E1 mRNAs as well as spliced mRNAs (880^2709). (E) Percentage cytoplasmic intron-retained HPV16 E1 mRNAs over a total sum of nuclear and cytoplasmic intron-retained E1 mRNAs in absence (–) or presence of hnRNP D40 from (D). (F) Effect of hnRNP D40 on HPV16 E1 protein levels was determined by western blotting on extracts from HeLa cells transfected with HPV16 E1-FLAG expressing plasmid p16E1-3xF (wild type) or p16E1SDm-3xF (harboring mutations at splice donors SD880 and SD1302 in the E1 gene thereby abolishing E2 mRNA splicing), in the absence (–) or presence (+) of hnRNP D40 expressing plasmid. Western blotting using anti-FLAG antibody (M2). (G) The levels of HPV16 intron-retained E1 mRNAs produced by p16E1SDm-3xF in nuclear and cytoplasmic fractions, in the absence (–) or presence (+) of hnRNP D40 expression plasmid, was determined by HPV16-specific RT-PCR primers primer pairs F-E1-1+E2Xba and PstI-SD880mE1-F+E2Xba (primer positions are displayed in Supplementary Figure S7). (H) Overexpression of hnRNP D40 does not affect production of sLuc from a CMV-promoter driven sLuc gene. (I) HeLa cells were transfected with indicated, FLAG-tagged hnRNP D expression plasmids. Nuclear and cytoplasmic extracts were subjected to Western blotting with anti-FLAG antibody or anti-tubulin or anti-histone antibody to control for subcellular fractionation. (J) Association between HPV16 E1 mRNAs and hnRNP D40 protein in the cytoplasm demonstrated by CLIP assay on cytoplasmic extracts from HeLa cells transfected with HPV16 subgenomic plasmid pC97ELsLuc in the absence or presence of FLAG-hnRNP D40 expressing plasmid. Cytoplasmic extracts were subjected to immunoprecipitation with anti-FLAG antibody to purify FLAG-hnRNP D40: RNA complexes. RNA in the ribonucleoprotein (RNP) complex was extracted and subjected to RT-PCR using primers 773S and E1AS that specifically detect HPV16 intron-retained E1 mRNAs.

mic fraction (Figure 7D). Of particular interest was that the proportion of cytoplasmic intron-retained HPV16 mRNAs was significantly increased in the presence of hnRNP D40 (Figure 7E). These results confirmed that hnRNP D40 inhibited HPV16 mRNA splicing and promoted intron retention to generate intron-retained HPV16 E1 mRNAs that were exported to the cytoplasm.

Finally, we wished to determine if the increase of intron-retained E1 mRNAs in the cytoplasm caused by hnRNP D40 also resulted in increased E1 protein levels. We generated an HPV16 E1 expression plasmid in which the E1 gene was fused with three-times FLAG sequence at its 3'-end named p16E1-3xF (Supplementary Figure S7G). This plasmid contained intact HPV16 5'-splice sites SD880 and SD1302 and had the potential to produce an intron-retained mRNA encoding a 73kD, 3xFLAG-tagged HPV16 E1 protein, as well as mRNAs spliced to HPV16 3'-splice site SA2709. In a second version of this plasmid, SD880 and SD1302 were mutationally inactivated to generate plasmid p16E1SDm-3xF (Supplementary Figure S7H). As expected, HPV16 E1 protein production was greatly improved from p16E1SDm-3xF compared to p16E1-3xF (Figure 7F), but overexpression of hnRNP D40 decreased E1 protein levels from both plasmid p16E1-3xF and p16E1SDm-3xF (Figure 7F). Analysis of the mRNAs produced by p16E1-3xF and p16E1SDm-3xF revealed that intron-retained HPV16 E1 mRNA levels transcribed from p16E1SDm-3xF were increased in the cytoplasm in the presence of hnRNP D40 (Figure 7G). The results were confirmed by using different sets of primer pairs (the primer positions are indicated in Supplementary Figure S7H). The results revealed that the levels of intron-retained E1 mRNAs increased in the presence of hnRNP D40 even though HPV16 5'-splice sites SD880 and SD1302 had been inactivated (Figure 7G), which indicated that hnRNP D40 also functioned independently of splicing. This effect was not due to transcriptional activation by hnRNP D40 since hnRNP D40 does not affect transcription from CMV-promoter driven plasmids (Figure 7H). Taken together, these results were reminiscent of the effect of hnRNP D40 on the HPV16 E6 mRNA and E6 protein levels (Figure 6). Our results indicated that hnRNP D40, in addition to its splicing inhibitory function, interacted with HPV16 E1 mRNAs also in the absence of RNA splicing and promoted the appearance of intron-retained HPV16 E1 mRNAs in the cytoplasm. However, these mRNAs were poorly translated into E1 protein and the results suggested that hnRNP D40 negatively interfered with mRNA translation. Indeed, translation *in vitro* of both E6 and E1 mRNAs into E6 and E1 proteins was inhibited by recombinant hnRNP D protein but not by bovine serum albumin (BSA) (Supplementary Figure S8A, B, D and E). This translational inhibition was specific for HPV16 E6 and E1 mRNAs since luciferase mRNA translation was unaffected by the presence of recombinant hnRNP D protein (Supplementary Figure S8C).

hnRNP D40 is associated with HPV16 mRNAs in the cytoplasm

The results presented above suggested that hnRNP D not only inhibited HPV16 mRNA splicing but also might have

the ability to accompany the HPV16 mRNAs to the cytoplasm. As already shown in Figure 4C, overexpressed hnRNP D40-EGFP was primarily located in the nucleus. However, hnRNP D is known to be a shuttling protein and small amounts of hnRNP D40 could potentially be present in the cytoplasm. We therefore transfected HeLa cells with a plasmid expressing a FLAG-tagged hnRNP D40 protein, fractionated the cells and analyzed nuclear and cytoplasmic fractions for FLAG-tagged hnRNP D40 by western blotting. The results revealed that the FLAG-tagged hnRNP D40 produced from the expression plasmid could be detected in both nuclear and cytoplasmic compartments (Figure 7I), whereas tubulin was detected primarily in the cytoplasmic- and histone in the nuclear-fractions (Figure 7I). To determine if hnRNP D40 was associated with the intron-retained HPV16 E1 mRNAs in the cytoplasm, we performed a CLIP assay on cytoplasmic extracts from HeLa cells transfected with HPV16 reporter plasmid pC97ELsLuc in the absence or presence of transfected pFLAG-hnRNP D40 plasmid. As can be seen from the results, intron-retained HPV16 mRNAs could be specifically amplified from the UV-crosslinked RNA-protein complexes immunoprecipitated with anti-FLAG antibody (Figure 7J). These results demonstrated that HPV16 intron-retained E1 mRNAs were associated with hnRNP D40 in the cell cytoplasm, suggesting that hnRNP D40 accompanied the intron-retained HPV16 mRNAs from the nucleus to the cytoplasm where it potentially interfered with HPV16 mRNA translation in a negative manner.

hnRNP D40 promotes intron retention and production of E1 and E6 mRNAs produced from episomal HPV16 genomes in human primary keratinocytes

Finally, we wished to determine if hnRNP D40 was actively inhibiting splicing also of mRNAs produced from complete, episomal HPV16 genomes. We therefore used plasmid pHPV16AN that encodes the entire HPV16 genome (Figure 8A) (43). The pHPV16AN plasmid was cotransfected with cre-recombinase-expressing plasmid to release the HPV16 genome between the loxP sites and generate the episomal form of the HPV16 genome (Figure 8A). Transfections were performed in the absence or presence of hnRNP D40 expression plasmid. As can be seen in Figure 8B, hnRNP D40 inhibited splicing of the E6/E7 mRNAs and promoted production of intron-retained E6 mRNAs (Figure 8B). A relatively minor inhibitory effect on splicing from SD880 to SA3358 was observed (Figure 8C). Importantly, the inhibitory effect on the E2 mRNAs was readily observed, followed by intron retention and generation of intron-retained E1 mRNAs (Figure 8D). Since transfection of cells with pHPV16AN is relatively inefficient, the larger, intron-retained E1 mRNAs may be relatively inefficiently amplified with primer pair 773S and E2AS. Therefore, we also monitored the levels of intron-retained E1 mRNAs with primers 773S and E1AS (Figure 1E) that also detected an increase in the levels of E1 mRNAs in the presence of hnRNP D40 (Figure 8E). These results were reproduced by transfection of pHPV16AN in the absence or presence of hnRNP D40 plasmid in human primary keratinocytes. As can be seen, hnRNP D40 inhibited splic-

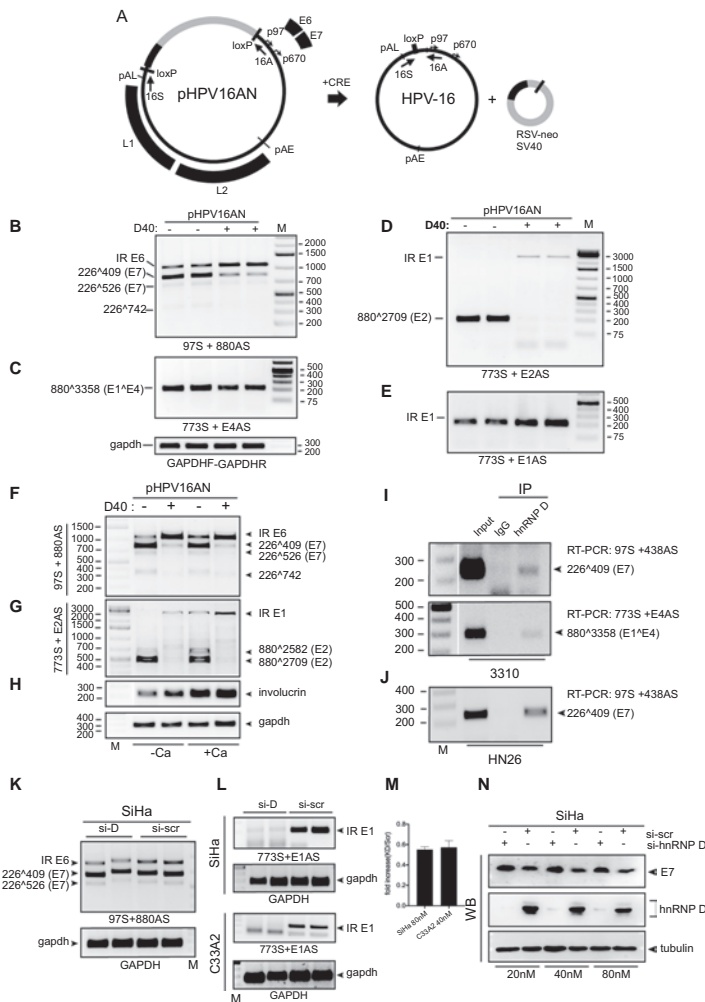


Figure 8. hnRNP D40 inhibits splicing of E6E7 and E1E2 mRNAs transcribed from episomal HPV16 DNA in primary human keratinocytes. (A) Schematic representation of the pHPV16AN plasmid and episomal HPV16 DNA production using Cre-loxP transfection system. (B-E) RT-PCR on RNA extracted from HeLa cells transfected with pHPV16AN together with Cre recombinase expressing plasmid in the absence (-) or presence (+) of hnRNP D40 expressing plasmid. RNA was extracted and HPV16 RNA splicing was monitored by RT-PCR using primers (B) 97S+880AS, (C) 773S+E4AS, (D) 773S+E2AS and (E) 773S+E1AS. (F-H) hnRNP D40 effect on HPV16 mRNA splicing from episomal HPV16 DNA in human primary keratinocytes with or without induction of differentiation. pHPV16AN was transfected into primary normal human foreskin keratinocyte (nHFk) together with Cre recombinase expressing plasmid, in the absence (-) or presence (+) of hnRNP D40 expression plasmid. CaCl₂ at a final concentration of 2.4 mM was added at 12 h post transfection (+Ca) for 24 h to induce differentiation. RNA was extracted and HPV16 RNA splicing was monitored by RT-PCR using primer pairs indicated to the left of the gel images. (I and J) Interactions between endogenously expressed HPV16 mRNAs and cellular hnRNP D proteins were analyzed by RIP assay on extracts from HPV16-immortalized human keratinocyte cell line 3310 (56) (I) or HPV16-positive tonsillar cancer cell line (55) HN26 cells (J). Whole cell extracts were subjected to immunoprecipitation with anti-hnRNP D antibody followed by RNA extraction and RT-PCR with HPV16-specific primers indicated to the right of the gel images. (K and L) SiHa or C33A2 cells were transfected with siRNAs to hnRNP D (si-hnRNP D) or scrambled siRNA (si-scr). 80 nM siRNA for SiHa cells and 40 nM siRNA for C33A2 cells were transfected. RNA was extracted at 24 h post-transfection to monitor intron-retained E6 mRNAs (K) or at 48 hrs post-transfection to monitor intron-retained E1 mRNAs (L), followed by RT-PCR with indicated primer pairs. (M) RT-qPCR to evaluate hnRNP D KD effect on HPV16 intron-retained E1 mRNAs was performed with cDNA from (L). (N) WB analysis to evaluate hnRNP D knockdown on HPV16 E7 protein production in SiHa cells transfected with siRNA to hnRNP D (si-hnRNP D) or with scrambled siRNA (si-scr) for 72 h.

ing and promoted production of intron-retained E6 mRNAs (Figure 8F) and E1 mRNAs (Figure 8G). Notably, induction of calcium-dependent differentiation of the transfected primary keratinocytes increased the levels of intron-retained E1 mRNAs (Figure 8G), an effect that was further enhanced by hnRNP D40 overexpression (Figure 8G). Induction of differentiation of the primary keratinocytes by calcium was confirmed by the increase in the levels of involucrin mRNAs (Figure 8H). Furthermore, we wished to obtain support for the physiological relevance of the interactions between hnRNP D protein and HPV16 mRNAs. To this end we monitored interactions between hnRNP D and HPV16 early mRNAs by RIP assay in the HPV16-positive head-and-neck cancer derived HN26 cells recently established from a tonsillar cancer patient at Lund University hospital (55), and from the in-house HPV16-immortalized keratinocyte cell line 3310 with integrated HPV16 DNA (56). The interactions between hnRNP D and HPV16 early E6/E7 mRNAs were readily detected by RT-PCR in the immunoprecipitated hnRNP D-RNA complexes obtained from either the tonsillar cancer cell line HN26 (Figure 8I) or the HPV16-immortalized human keratinocyte 3310 cell line (Figure 8J). Finally, the effect on HPV16 mRNA splicing and protein expression of small-interfering RNA (siRNA) mediated knockdown (KD) of hnRNP D proteins was determined using HPV16 positive cancer cell line SiHa and the in-house generated cell line C33A2 (42) (Figure 8K–N). Briefly, the C33A2 cell line was derived from C33A cells stably transfected with HPV16 reporter plasmid pBELsLuc plasmid (Figure 3E). KD of hnRNP D revealed a reduction in intron retention on E6 mRNAs in SiHa cells (Figure 8K) as well as a reduction in intron retention on E1 mRNAs both in SiHa and C33A2 cells (Figure 8L,M). Although the effect of hnRNP D knockdown on E6 mRNAs in SiHa cells was weaker than the effect on E1 mRNAs, this decrease was statistically significant ($P = 0.003$). Furthermore, KD of hnRNP D resulted in an increase of HPV16 E7 protein levels in SiHa cells (Figure 8N), which is consistent with an increase in spliced 226^409 mRNAs that are dedicated for E7 protein production (Figure 8K). Taken together, we concluded that hnRNP D40 promoted production of HPV16 intron-retained E1 and E6 mRNAs with retained introns also on mRNAs expressed from the episomal form of the HPV16 genome suggesting that the interactions of between HPV16 early mRNAs and hnRNP D are of significance also in HPV16 positive cancer cell lines. Furthermore, knockdown of hnRNP D enhanced production of the HPV16 E7 protein in HPV16-driven cervical cancer cells.

DISCUSSION

Previously, we have shown that HPV16 5'- splice site SD3632 that is exclusively utilized during late stages of the HPV16 life cycle, is suppressed in mitotic cells by binding of hnRNP D proteins to AUAGUA motifs upstream of SD3632 (43). In our current manuscript, we present results that indicate that hnRNP D proteins inhibit HPV16 E1/E2- and E6/E7-mRNA splicing. The effect of hnRNP D on the HPV16 mRNA splicing process (Figure 3) was followed by effects on subsequent steps of HPV16 mRNA processing including a potential increase in mRNA stabil-

ity (Figures 6 and 7) and enhanced nuclear export (Figures 6 and 7) of intron-retained HPV16 mRNAs. Consequently, intron-retained HPV16 E1 and E6 mRNA levels in the cytoplasm increased in the presence of hnRNP D. The regulation of RNA processing including alternative pre-mRNA splicing is particularly important for the regulation of HPV16 gene expression since all HPV16 genes are expressed from a plethora of alternatively spliced viral mRNAs. Of note, alternative RNA splicing of E1 and E2 mRNAs and E6 and E7 mRNAs are of particular interest since E1 and E2, and E6 and E7 are produced from mRNAs that are produced in a mutually exclusive manner: while E1 and E6 are produced from intron-retained mRNAs, E2 and E7 are produced from the spliced variants of the same pre-mRNAs, respectively. Since E1 and E2 are functionally linked, as are E6 and E7, it is of utmost importance that the balance between production of intron-retained E1 and E6 mRNAs and spliced E2 and E7 mRNAs is optimal to meet the demand for optimal E1/E2- and E6/E7-protein ratios. A tightly tuned balance between the expression of high-risk HPV E6 and E7 oncoproteins is essential for cancer cell maintenance and malignant progression, while the balance between the expression of HPV E1 and E2 is essential during the viral replication cycle. Thus, hnRNP D may contribute to HPV16 gene expression during the viral life cycle as well as during progression to cancer and maintenance of the malignant stage of the HPV16 infected cells but that remains to be investigated.

The hnRNP D protein family consists of four members named p37, p40, p42 and p45 distinguished by the presence or absence of exon 2 and exon 7 (Figure 3A). Presence or absence of these alternatively spliced exons confers distinct biological properties to individual hnRNP D isoforms. Presence of exon 7 interrupts the C-terminal domain (CTD) of hnRNP D that has been known to affect nuclear-cytoplasmic distribution of hnRNP D (48,57). An uninterrupted CTD (p37 and p40) increased nuclear uptake, whereas an interrupted CTD (p42 and p45) resulted in cytoplasmic accumulation. Furthermore, presence of exon 7 is known to suppress ubiquitination of p42 and p45 (58). On the other hand, absence of exon 2 (p37 and p42) enhances high affinity binding to target RNAs (59). In addition, C-terminal RGG/RG motifs of hnRNP D have been shown to exert RNA chaperon and RNA annealing activities (60). Our results demonstrated that C-terminal deletions and substitutions in RGG motifs altered the subcellular distribution of hnRNP D (Figure 4C and Supplementary Figure S4B), supporting the idea that the C-terminus and/or RGG contribute to localization of hnRNP D but also revealed that neither the C-terminus nor the C-terminal RGG-motifs were essential for nuclear localization. However, the only deletion that displayed a nuclear-excluded phenotype was D8, which effectively encoded only RRM1 and RRM2 and had deletions in both N- and C-termini, suggesting that amino acid sequences in the N- and C-termini of hnRNP D collaborate to bring hnRNP D to the nucleus.

All deletion mutants of hnRNP D40 displayed in this manuscript had reduced splicing inhibitory effect on HPV16 E6/E7 or E1/E2 mRNA splicing, albeit to a different extent (Figures 4 and 5). Obviously, hnRNP D40

deletion mutants D2 and D7 lost their inhibitory effect on E1/E2 mRNA splicing as well as on E6/E7 mRNA splicing (Figure 4H–K). Surprisingly, D2 appeared to activate HPV16 226 \wedge 409-splicing while D1 activated HPV16 226 \wedge 526-splicing (Figure 4D), indicating that the splicing inhibitory function of hnRNP D is located in its N-terminus and that D1 and D2 were in contact with the splicing machinery despite their N-terminal deletions. Indeed, both D1 and D2 interacted with U1-70K, U2AF65 and U2AF35, albeit at a decreased efficiency compared to wild-type hnRNP D40 (Figure 5K). In contrast to D1 that retained its ability to associate with HPV16 mRNAs, D2 displayed a reduced association with HPV16 mRNAs (Figure 5L). These results combined, we propose a model for the effect of hnRNP D on HPV16 mRNA splicing shown in Figure 9. For the interaction of hnRNP D with HPV16 RNA, both RRM1 and RRM2 domains are essential, while for the interaction with cellular spliceosome factors, U1-70K and U2AF65/35, both hnRNP D N- and C-termini regions contribute. The interaction of hnRNP D with spliceosome factors may cause functional inhibition of the splicing machinery and this inhibition appears to be mediated by hnRNP D N-terminal A-rich and exon 2-encoding region. Without the inhibitory N-terminus, hnRNP D may function as a mediator of splicing that interacts with target transcripts and recruits spliceosome factors to it, but in contrast to full-length hnRNP D, N-terminally deleted hnRNP D activates mRNA splicing. Since D1, D2 or RGG mutants differed in their inhibitory effects on HPV16 226 \wedge 526-, 226 \wedge 409- and 226 \wedge 742-RNA splicing, respectively, it appeared as if other domains than RRM1/RRM2 on hnRNP D contributed to the selection of target RNA sequences, i.e. splice site selection. Alternatively, hnRNP D40 mutants without inhibitory N-terminus (D1 and D2) functioned as decoys for full-length endogenous hnRNP D proteins and formed hetero- or homo-oligomers that may possess altered target RNA specificity or splicing inhibitory functions. In conclusion, full-length hnRNP D40 appeared to target mainly 226 \wedge 409.

In addition to the hnRNP D functional analysis, we explored hnRNP D target sites on HPV16 E6/E7 coding RNA sequence using RNA-mediated, protein pull-down assay and found that hnRNP D was associated with the E6 intronic region (nt243–445) as well as with previously identified HPV16 E6/E7 splicing silencer element located between HPV16 nt579–620 (46). In addition, hnRNP D was associated with several spots adjacent to alternative HPV16 splice sites SA526 and SA742. hnRNP D was originally identified as an RNA-binding protein that associates with the 5'- and 3'-untranslated regions of target mRNAs that contained AU-rich RNA elements (AREs) (61) to affect stability, translation or subcellular localization of target transcripts. hnRNP D/AUF1 was shown to bind directly to the AU-rich RNA elements (UAUUA) (27) and recent PAR-CLIP experiments revealed that hnRNP D/AUF1 primarily recognized U-/GU-rich sequences (e.g. UUUAGA, AGUUU, GUUUG, UUUU, UUAGUU and AGUUU) (34). These interactions affected the fate of the target transcripts in different ways, primarily by lowering steady-state levels of numerous mRNAs or by promoting translation of numerous other mRNAs. Surprisingly, hn-

RNP D also enhanced the steady-state levels of several target RNAs. The majority of the PAR-CLIP-identified binding sites for hnRNP D were present on multiple positions in the E6-E7 coding regions (nt97–856) and in the E1 coding region (865–2814). The previously identified HPV16 E6/E7 mRNA splicing silencer UUAGAUU (46) was also found at multiple positions in the E6/E7E1 coding region. This (UUAGAUU) partially overlaps with one of the binding sites for hnRNP D identified by PAR-CLIP (UUUAGA). We did not investigate if hnRNP D does bind to the multiple HPV16 RNA sequences directly, but our CLIP experiment showed that hnRNP D binds directly to HPV16 mRNAs in transfected cells. However, hnRNP D probably interacts with HPV16 mRNAs both directly and indirectly via other RNA-binding proteins. It is still unknown if each hnRNP D isoform differ in binding to specific RNA sequences. It would be of interest to use PAR-CLIP to determine potential binding sites for hnRNP D on the HPV16 mRNAs.

We showed that hnRNP D40 inhibited HPV16 early mRNA splicing and increased intron retention on 16E6 and 16E1 mRNAs and accompanied these mRNAs to the cytoplasm. This appears similar to the function of the human immunodeficiency virus (HIV) Rev protein that induces nuclear export of unspliced HIV-1 transcripts to the cytoplasm, thereby upregulating HIV Gag, Pol and Env protein levels (62–64). Here we found that hnRNP D inhibited splicing of HPV16 mRNAs and increased intron retention on HPV16 E6 and E1 mRNAs but did not enhance protein levels of either 16E6 or 16E1. As a matter of fact, the E1 and E6 protein levels were decreased in the presence of hnRNP D40. Therefore, hnRNP D may function differently from the HIV Rev protein, at least after the nuclear export of the HPV16 transcripts. It remains unclear whether hnRNP D directly affects export of intron-retained HPV16 mRNAs to the cytoplasm and it remains to be determined how these intron-retained mRNAs are exported to the cytoplasm. However, since our results demonstrate that the translation of 16E6 and 16E1 mRNA is suppressed by hnRNP D40 still sitting on the HPV16 mRNAs in the cytoplasm, we speculate that hnRNP D has a functional role in the cytoplasm. To allow E6 or E1 mRNAs to enter a productive pathway in the cytoplasm, it may be necessary to replace hnRNP D40 with unknown factor(s) on these mRNAs once they enter the cytoplasm. It is also reasonable to speculate that interactions of HPV16 mRNAs with hnRNP D affects the half-life of the intron-retained HPV16 mRNAs but that remains to be determined. To uncover the fate of the intron-retained HPV16 E1 and E6 mRNAs in the cytoplasm, a better understanding of hnRNP D and its role in the cytoplasm is needed.

Interactions between hnRNP D and various viral RNAs have been described previously, both with nuclear-restricted viral RNAs such as those produced by Epstein-Barr virus (EBV) (65) and cytoplasm-restricted viral RNAs such as those produced by viruses of the Picornaviridae family, including poliovirus, Coxsackievirus B3 (CVB3), Human Rhinovirus (HRV) and Enterovirus EV71 (66,67) or of the Flaviviridae family, including West Nile virus (WNV) and Dengue virus (DENV) (68,69), as well as with HIV of the Retroviridae family (70). The consequence of the interactions between hnRNP D and the various viral RNAs dif-

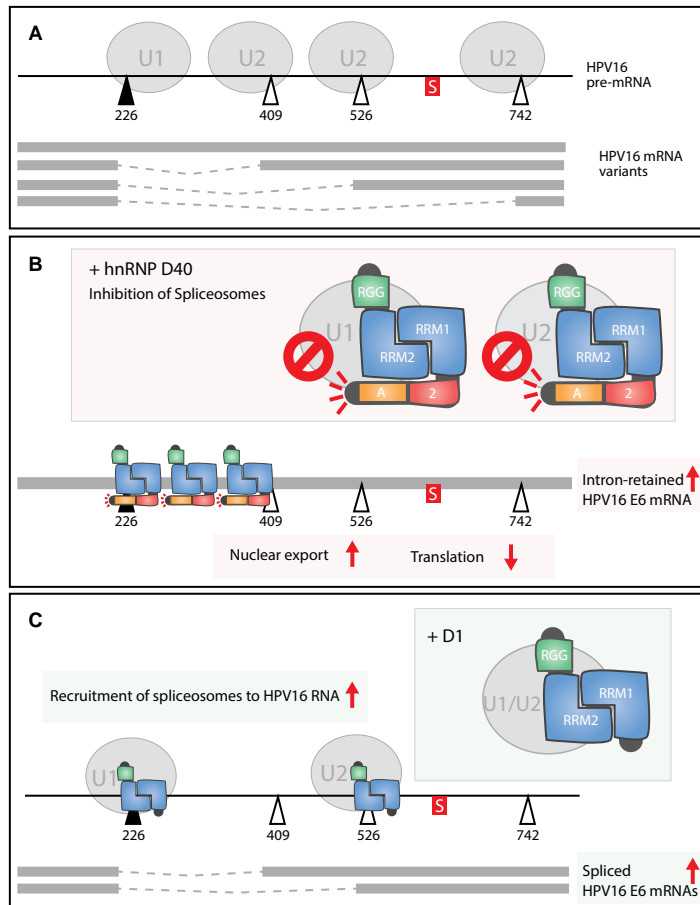


Figure 9. Model how hnRNP D regulates HPV16 E6/E7 mRNA splicing. (A) In regular condition, four splice isoforms of HPV16 E6/E7 mRNAs are produced. This is regulated by fine-tuned recruitment of cellular U1 spliceosome factor to HPV16 E6/E7 splice donor SD226 and that of U2 spliceosome factors to HPV16 E6/E7 splice acceptors SA409, SA526 or SA742. (B) hnRNP D40 binds to both U1 spliceosome factor and U2 spliceosome factor at its C-terminus and with some extent at its N-terminus, which inhibits spliceosome factor activity. This inhibitory function is mediated by N-terminus of hnRNP D. At the same time, hnRNP D interacts with several sites in HPV16 E6/E7 coding RNAs at its RRM1/RRM2 domain. As a result, hnRNP D mainly inhibits 226/409 splicing on the HPV16 E6/E7 mRNAs. The intron-retained E6 mRNAs are accompanied by hnRNP D to the cytoplasm. However, hnRNP D association on intron-retained E6 mRNA suppresses translation of E6 protein by unknown mechanism. To enhance translation of HPV16 E6 protein from intron-retained E6 mRNA, hnRNP D40 must need to be replaced from intron-retained E6 mRNA. (C) Deletion mutant of hnRNP D40, D1 that doesn't possess N-terminal A-rich and exon 2 but harbors intact RRM1/RRM2 domain and C-terminus, still has potential to interact with spliceosome factors and with HPV16 RNA though it loses inhibitory effect on spliceosome factors. As a result, D1 mutant mediates recruitment of spliceosome factors to HPV16 splice sites, thereby increasing HPV16 E6/E7 mRNA splicing.

fers. Interactions of hnRNP D and EBV EBER1 noncoding RNAs are believed to modulate cellular alternative RNA splicing, although direct evidence is lacking at this point (65). Interactions between hnRNP D and Picornaviral genomic RNAs at the internal ribosome entry site (IRES) of the 5'-non-coding region (NCR) or at the 3'-NCR, resulted in inhibition of translation (71) or decreased viral genomic

RNA levels (67), respectively. Interactions between hnRNP D and flaviviral genomic RNAs, either at the 5'-NCR or at the 3'-NCR contributed to a shift in the RNA structure from a linear to a cyclic viral genome of which the latter is favored as a template for viral RNA replication rather than translation (68,69). Interactions between different hnRNP D isoforms and HIV-1 exon 3-RNA altered the relative lev-

els of spliced and unspliced HIV-1 mRNAs (72). In conclusion, hnRNP D plays a major role in replication and gene expression of multiple viruses, including DNA viruses and RNA viruses, affecting nuclear as well as cytoplasmic functions of these viral RNAs. In particular, hnRNP D appears to inhibit translation of mRNAs produced by several different viruses, which is in line with our results presented here.

In contrast to the RNA viruses described above that replicate to high levels and cause acute infections, HPVs establish persistent infections that normally last for 12–24 months, thereby requiring elaborate immune evasion mechanisms, suggesting that the inhibitory effect of hnRNP D on the translation of the E1 and E6 mRNAs may contribute to persistence in the presence of a functional immune system of the host. At the regulatory level, the ratio of E1 to E2, produced from intron-retained and spliced mRNAs respectively, may affect various steps in the HPV life cycle. While higher levels of E1 than E2 is required for genomic DNA replication, increased E2 levels will eventually shut down the early HPV promoter to allow for a re-entry into the cellular differentiation program and completion of the HPV life cycle. Regarding E6 and E7 proteins that are produced from intron-retained and spliced mRNAs respectively, an optimal ratio of E6 and E7 is required to achieve an optimal balance between antiapoptotic and promitotic functions exerted by E6 and E7, respectively. Although it remains to be determined how the intron-retained HPV E1 and E6 mRNAs are translated, hnRNP D may contribute to the modulation of HPV16 mRNA translation and to the control of the delicate balance of the E1/E2- and E6/E7 protein ratios during HPV16 infection.

Finally and importantly, the inhibitory effect of hnRNP D40 on HPV16 E6 or E1 mRNA splicing was reproduced with episomal genomic HPV16 DNA in human primary keratinocytes. This is of significance since the HPV16 genome normally exists as an episome during its life cycle in infected cells *in vivo*. Furthermore, we demonstrated an association between endogenous HPV16 RNA and hnRNP D proteins in HN26 cells, a cell line recently isolated from a male patient with HPV16-positive tonsillar cancer, and in the previously described HPV16-immortalized human keratinocyte cell line 3310 (Figure 8). Interestingly, induction of keratinocyte differentiation in primary keratinocytes appeared to promote production of intron-retained E1 mRNAs, an effect that was enhanced by overexpression of hnRNP D40, further supporting a role for hnRNP D in the differentiation-dependent HPV16 gene expression program. The hnRNP D proteins are expressed in cervical epithelium as well as in tonsillar epithelial cells *in vivo* and appears to be overexpressed in cervical and tonsillar cancer (<https://www.proteinatlas.org/ENSG00000138668-HNRNP-D>), indicating that hnRNP D is available to HPV16 in its natural environment as well as during progression to cancer. It has been shown that the levels of multiple cellular RNA-binding proteins are altered or differ in functional activity through cervical cancer development caused by HPV infection (73), underscoring the importance of RNA-binding proteins during progression of HPV-infected cells to cancer. In fact, altered expression of hnRNP D protein has been observed in different cancers (74) supporting the idea of a role for hnRNP D in car-

cinogenesis. Modulation of HPV16 gene expression by hnRNP D suggests that hnRNP D may be a future target for anti-HPV and anti-cancer treatment. This conclusion is reinforced by our results presented here, demonstrating that knockdown of hnRNP D in HPV16-driven cervical cancer cells enhanced production of the HPV16 E7 oncoprotein.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We are grateful to Dr. R. J. Schneider, H. Leffers, D. Black, I. C. Eperon, C. Gooding, K. Bomsztyk, S. Guang, N. Funk and P. Xu for generously providing hnRNP plasmids and to Júlia Ortís Sunyer, Johanna Jönsson, Yunji Zheng, Chengjun Wu, Haoran Yu, Kersti Nilsson and Shirin Shoja Chaghervand for contributing to some experiments and discussions.

FUNDING

Swedish Research Council [VR2019-01210 to S.S.]; Swedish Cancer Society [CAN2018/702 to S.S.]; China Scholarship Council [201809120016 to X.C., 201706170061 to C.H., 201606525004 to L.G.]. Funding for open access charge: Swedish Research Council [VR2019-01210 to S.S.].
Conflict of interest statement. None declared.

REFERENCES

- Chow, L.T., Broker, T.R. and Steinberg, B.M. (2010) The natural history of human papillomavirus infections of the mucosal epithelia. *APMIS*, **118**, 422–449.
- McBride, A.A. (2021) Human papillomaviruses: diversity, infection and host interactions. *Nat. Rev. Microbiol.*, **20**, 95–108.
- Bouvard, V., Baan, R., Straif, K., Grosse, Y., Secretan, B., El Ghissassi, F., Benbrahim-Talaa, L., Guha, N., Freeman, C., Galichet, L. *et al.* (2009) A review of human carcinogens—Part B: biological agents. *Lancet Oncol.*, **10**, 321–322.
- Moody, C.A. and Laimins, L.A. (2010) Human papillomavirus oncoproteins: pathways to transformation. *Nat. Rev. Cancer*, **10**, 550–560.
- Akagi, K., Li, J., Broutian, T.R., Padilla-Nash, H., Xiao, W., Jiang, B., Rocco, J.W., Teknos, T.N., Kumar, B., Wangsa, D. *et al.* (2014) Genome-wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability. *Genome Res.*, **24**, 185–199.
- Martinez-Zapien, D., Ruiz, F.X., Poirson, J., Mitschler, A., Ramirez, J., Forster, A., Cousido-Siah, A., Masson, M., Vande Pol, S., Podjarny, A. *et al.* (2016) Structure of the E6/E6AP/p53 complex required for HPV-mediated degradation of p53. *Nature*, **529**, 541–545.
- Mantovani, F. and Banks, L. (2001) The human papillomavirus E6 protein and its contribution to malignant progression. *Oncogene*, **20**, 7874–7887.
- James, M.A., Lee, J.H. and Klingelutz, A.J. (2006) Human papillomavirus type 16 E6 activates NF- κ B, induces cIAP-2 expression, and protects against apoptosis in a PDZ binding motif-dependent manner. *J. Virol.*, **80**, 5301–5307.
- White, E.A., Kramer, R.E., Tan, M.J., Hayes, S.D., Harper, J.W. and Howley, P.M. (2012) Comprehensive analysis of host cellular interactions with human papillomavirus E6 proteins identifies new E6 binding partners and reflects viral diversity. *J. Virol.*, **86**, 13174–13186.
- Bergvall, M., Melendy, T. and Archambault, J. (2013) The E1 proteins. *Virology*, **445**, 35–56.
- McBride, A.A. (2013) The papillomavirus E2 proteins. *Virology*, **445**, 57–79.

12. Burley, M., Roberts, S. and Parish, J.L. (2020) Epigenetic regulation of human papillomavirus transcription in the productive virus life cycle. *Semin. Immunopathol.*, **42**, 159–171.
13. Kurg, R., Uusen, P., Vösa, L. and Ustav, M. (2010) Human papillomavirus E2 protein with single activation domain initiates HPV18 genome replication, but is not sufficient for long-term maintenance of virus genome. *Virology*, **408**, 159–166.
14. Kadaja, M., Isok-Paas, H., Laos, T., Ustav, E. and Ustav, M. (2009) Mechanism of genomic instability in cells infected with the high-risk human papillomaviruses. *PLoS Pathog.*, **5**, e1000397.
15. DiMaio, D. and Petti, L.M. (2013) The E5 proteins. *Virology*, **445**, 99–114.
16. Doorbar, J. (2013) The E4 protein: structure, function and patterns of expression. *Virology*, **445**, 80–98.
17. Johansson, C. and Schwartz, S. (2013) Regulation of human papillomavirus gene expression by splicing and polyadenylation. *Nat. Rev. Microbiol.*, **11**, 239–251.
18. Kajitani, N. and Schwartz, S. (2020) Role of Viral Ribonucleoproteins in Human Papillomavirus Type 16 Gene Expression. *Viruses*, **12**, 1110.
19. Ajiro, M. and Zheng, Z.M. (2014) Oncogenes and RNA splicing of human tumor viruses. *Emerg. Microbes Infect.*, **3**, e63.
20. Zheng, Z.M. (2010) Viral oncogenes, noncoding RNAs, and RNA splicing in human tumor viruses. *Int. J. Biol. Sci.*, **6**, 730–755.
21. Graham, S.V. and Faizo, A.A.A. (2017) Control of human papillomavirus gene expression by alternative splicing. *Virus Res.*, **231**, 83–95.
22. Graham, S.V. (2017) Keratinocyte differentiation-dependent Human Papillomavirus gene regulation. *Viruses*, **9**, 245.
23. Fay, J., Kelehan, P., Lambkin, H. and Schwartz, S. (2009) Increased expression of cellular RNA-binding proteins in HPV-induced neoplasia and cervical cancer. *J. Med. Virol.*, **81**, 897–907.
24. Mole, S., McFarlane, M., Chuen-Im, T., Milligan, S.G., Millan, D. and Graham, S.V. (2009) RNA splicing factors regulated by HPV16 during cervical tumour progression. *J. Pathol.*, **219**, 383–391.
25. Dreyfuss, G., Kim, V.N. and Kataoka, N. (2002) Messenger-RNA-binding proteins and the messages they carry. *Nat. Rev. Mol. Cell Biol.*, **3**, 195–205.
26. Chen, M. and Manley, J.L. (2009) Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat. Rev. Mol. Cell Biol.*, **10**, 741–754.
27. Van Nostrand, E.L., Freese, P., Pratt, G.A., Wang, X., Wei, X., Xiao, R., Blue, S.M., Chen, J.Y., Cody, N.A.L., Dominguez, D. et al. (2020) A large-scale binding and functional map of human RNA-binding proteins. *Nature*, **583**, 711–719.
28. Geuens, T., Bouhy, D. and Timmerman, V. (2016) The hnRNP family: insights into their role in health and disease. *Hum. Genet.*, **135**, 851–867.
29. Cartegni, L., Maconi, M., Morandi, E., Cobianchi, F., Riva, S. and Biamonti, G. (1996) hnRNP A1 selectively interacts through its Gly-rich domain with different RNA-binding proteins. *J. Mol. Biol.*, **259**, 337–348.
30. Jankowsky, E. and Harris, M.E. (2015) Specificity and nonspecificity in RNA-protein interactions. *Nat. Rev. Mol. Cell Biol.*, **16**, 533–544.
31. Gratacós, F.M. and Brewer, G. (2010) The role of AUF1 in regulated mRNA decay. *Wiley Interdiscip. Rev. RNA*, **1**, 457–473.
32. Abdelmohsen, K., Tominaga-Yamanaka, K., Srikantan, S., Yoon, J.H., Kang, M.J. and Gorospe, M. (2012) RNA-binding protein AUF1 represses Dicer expression. *Nucleic Acids Res.*, **40**, 11531–11544.
33. Liao, B., Hu, Y. and Brewer, G. (2007) Competitive binding of AUF1 and TIAR to MYC mRNA controls its translation. *Nat. Struct. Mol. Biol.*, **14**, 511–518.
34. Yoon, J.H., De, S., Srikantan, S., Abdelmohsen, K., Grammatikakis, I., Kim, J., Kim, K.M., Noh, J.H., White, E.J., Martindale, J.L. et al. (2014) PAR-CLIP analysis uncovers AUF1 impact on target RNA fate and genome integrity. *Nat. Commun.*, **5**, 5248.
35. Eversole, A. and Maizels, N. (2000) In vitro properties of the conserved mammalian protein hnRNP D suggest a role in telomere maintenance. *Mol. Cell Biol.*, **20**, 5425–5432.
36. Pont, A.R., Sadri, N., Hsiao, S.J., Smith, S. and Schneider, R.J. (2012) mRNA decay factor AUF1 maintains normal aging, telomere maintenance, and suppression of senescence by activation of telomerase transcription. *Mol. Cell*, **47**, 5–15.
37. Torrisani, J., Unterberger, A., Tendulkar, S.R., Shikimi, K. and Szyf, M. (2007) AUF1 cell cycle variations define genomic DNA methylation by regulation of DNMT1 mRNA stability. *Mol. Cell Biol.*, **27**, 395–410.
38. Lapucci, A., Donnini, M., Papucci, L., Witort, E., Tempestini, A., Bevilacqua, A., Nicolin, A., Brewer, G., Schiavone, N. and Capaccioli, S. (2002) AUF1 is a bcl-2 A + U-rich element-binding protein involved in bcl-2 mRNA destabilization during apoptosis. *J. Biol. Chem.*, **277**, 16139–16146.
39. Lu, J.Y., Sadri, N. and Schneider, R.J. (2006) Endotoxic shock in AUF1 knockout mice mediated by failure to degrade proinflammatory cytokine mRNAs. *Genes Dev.*, **20**, 3174–3184.
40. Kemmerer, K., Fischer, S. and Weigand, J.E. (2018) Auto- and cross-regulation of the hnRNPs D and DL. *RNA*, **24**, 324–331.
41. Fragkouli, A., Koukouraki, P., Vlachos, I.S., Paraskevopoulou, M.D., Hatzigeorgiou, A.G. and Doxakis, E. (2017) Neuronal ELAVL proteins utilize AUF-1 as a co-partner to induce neuron-specific alternative splicing of APP. *Sci. Rep.*, **7**, 44507.
42. Li, X., Johansson, C., Cardoso Palacios, C., Mossberg, A., Dhanjal, S., Bergvall, M. and Schwartz, S. (2013) Eight nucleotide substitutions inhibit splicing to HPV-16 3'-splice site SA3358 and reduce the efficiency by which HPV-16 increases the life span of primary human keratinocytes. *PLoS One*, **8**, e72776.
43. Li, X., Johansson, C., Glahder, J., Mossberg, A.K. and Schwartz, S. (2013) Suppression of HPV-16 late L1 5'-splice site SD3632 by binding of hnRNP D proteins and hnRNP A2/B1 to upstream AUAGUA RNA motifs. *Nucleic Acids Res.*, **41**, 10488–10508.
44. Wu, C., Nilsson, K., Zheng, Y., Ekenstierna, C., Sugiyama, N., Forslund, O., Kajitani, N., Yu, H., Wennerberg, J., Ekblad, L. et al. (2019) Short half-life of HPV16 E6 and E7 mRNAs sensitizes HPV16-positive tonsillar cancer cell line HN26 to DNA-damaging drugs. *Int. J. Cancer*, **144**, 297–310.
45. Somborg, M., Zhao, X., Fröhlich, M., Evander, M. and Schwartz, S. (2008) Polypyrimidine tract binding protein induces human papillomavirus type 16 late gene expression by interfering with splicing inhibitory elements at the major late 5' splice site, SD3632. *J. Virol.*, **82**, 3665–3678.
46. Zheng, Y., Jönsson, J., Hao, C., Shoja Chaghervand, S., Cui, X., Kajitani, N., Gong, L., Wu, C. and Schwartz, S. (2020) Heterogeneous nuclear ribonucleoprotein A1 (hnRNP A1) and hnRNP A2 Inhibit Splicing to Human Papillomavirus 16 Splice Site SA409 through a UAG-containing sequence in the E7 coding region. *J. Virol.*, **94**, e01509-20.
47. Dhanjal, S., Kajitani, N., Glahder, J., Mossberg, A.K., Johansson, C. and Schwartz, S. (2015) Heterogeneous nuclear ribonucleoprotein C proteins interact with the Human Papillomavirus Type 16 (HPV16) early 3'-untranslated region and alleviate suppression of HPV16 Late L1 mRNA splicing. *J. Biol. Chem.*, **290**, 13354–13371.
48. Sarkar, B., Lu, J.Y. and Schneider, R.J. (2003) Nuclear import and export functions in the different isoforms of the AUF1/heterogeneous nuclear ribonucleoprotein protein family. *J. Biol. Chem.*, **278**, 20700–20707.
49. Nasim, M.T., Chernova, T.K., Chowdhury, H.M., Yue, B.G. and Eperon, I.C. (2003) HnRNP G and Tra2beta: opposite effects on splicing matched by antagonism in RNA binding. *Hum. Mol. Genet.*, **12**, 1337–1348.
50. Guang, S., Felthouser, A.M. and Mertz, J.E. (2005) Binding of hnRNP L to the pre-mRNA processing enhancer of the herpes simplex virus thymidine kinase gene enhances both polyadenylation and nucleocytoplasmic export of intronless mRNAs. *Mol. Cell Biol.*, **25**, 6303–6313.
51. Huang, J., Li, S.J., Chen, X.H., Han, Y. and Xu, P. (2008) hnRNP-R regulates the PMA-induced c-fos expression in retinal cells. *Cell. Mol. Biol. Lett.*, **13**, 303–311.
52. Chou, M.Y., Rooke, N., Turck, C.W. and Black, D.L. (1999) hnRNP H is a component of a splicing enhancer complex that activates a c-src alternative exon in neuronal cells. *Mol. Cell Biol.*, **19**, 69–77.
53. Collier, B., Oberg, D., Zhao, X. and Schwartz, S. (2002) Specific inactivation of inhibitory sequences in the 5' end of the human papillomavirus type 16 L1 open reading frame results in production of high levels of L1 protein in human epithelial cells. *J. Virol.*, **76**, 2739–2752.
54. Busch, A. and Hertel, K.J. (2012) Evolution of SR protein and hnRNP splicing regulatory factors. *Wiley Interdiscip. Rev. RNA*, **3**, 1–12.

55. Forslund, O., Sugiyama, N., Wu, C., Ravi, N., Jin, Y., Swoboda, S., Andersson, F., Bzhalava, D., Hultin, E., Paulsson, K. *et al.* (2019) A novel human in vitro papillomavirus type 16 positive tonsil cancer cell line with high sensitivity to radiation and cisplatin. *BMC Cancer*, **19**, 265.
56. Johansson, C., Jamal Fattah, T., Yu, H., Nygren, J., Mossberg, A.K. and Schwartz, S. (2015) Acetylation of intragenic histones on HPV16 correlates with enhanced HPV16 gene expression. *Virology*, **482**, 244–259.
57. Arai, Y., Kuriyama, R., Kayama, F. and Kato, S. (2000) A nuclear matrix-associated factor, SAF-B, interacts with specific isoforms of AUF1/hnRNP D. *Arch. Biochem. Biophys.*, **380**, 228–236.
58. Laroia, G. and Schneider, R.J. (2002) Alternate exon insertion controls selective ubiquitination and degradation of different AUF1 protein isoforms. *Nucleic Acids Res.*, **30**, 3052–3058.
59. Wagner, B.J., DeMaria, C.T., Sun, Y., Wilson, G.M. and Brewer, G. (1998) Structure and genomic organization of the human AUF1 gene: alternative pre-mRNA splicing generates four protein isoforms. *Genomics*, **48**, 195–202.
60. Meyer, A., Golbik, R.P., Sanger, L., Schmidt, T., Behrens, S.E. and Friedrich, S. (2019) The RGG/RG motif of AUF1 isoform p45 is a key modulator of the protein's RNA chaperone and RNA annealing activities. *RNA Biol.*, **16**, 960–971.
61. Zhang, W., Wagner, B.J., Ehrenman, K., Schaefer, A.W., DeMaria, C.T., Crater, D., DeHaven, K., Long, L. and Brewer, G. (1993) Purification, characterization, and cDNA cloning of an AU-rich element RNA-binding protein, AUF1. *Mol. Cell. Biol.*, **13**, 7652–7665.
62. Felber, B.K., Hadzopoulou-Cladaras, M., Cladaras, C., Copeland, T. and Pavlakis, G.N. (1989) rev protein of human immunodeficiency virus type 1 affects the stability and transport of the viral mRNA. *Proc. Natl. Acad. Sci. U.S.A.*, **86**, 1495–1499.
63. Felber, B.K., Zolotukhin, A.S. and Pavlakis, G.N. (2007) Posttranscriptional control of HIV-1 and other retroviruses and its practical applications. *Adv. Pharmacol.*, **55**, 161–197.
64. Ostermann, P.N., Ritchie, A., Ptok, J. and Schaal, H. (2021) Let It Go: HIV-1. *J. Virol.*, **95**, e0034221.
65. Lee, N., Pimienta, G. and Steitz, J.A. (2012) AUF1/hnRNP D is a novel protein partner of the EBER1 noncoding RNA of Epstein-Barr virus. *RNA*, **18**, 2073–2082.
66. Cathcart, A.L. and Semler, B.L. (2014) Differential restriction patterns of mRNA decay factor AUF1 during picornavirus infections. *J. Gen. Virol.*, **95**, 1488–1492.
67. Wong, J., Si, X., Angeles, A., Zhang, J., Shi, J., Fung, G., Jagdeo, J., Wang, T., Zhong, Z., Jan, E. *et al.* (2013) Cytoplasmic redistribution and cleavage of AUF1 during coxsackievirus infection enhance the stability of its viral genome. *FASEB J.*, **27**, 2777–2787.
68. Friedrich, S., Schmidt, T., Geissler, R., Lilie, H., Chabierski, S., Ulbert, S., Liebert, U.G., Golbik, R.P. and Behrens, S.E. (2014) AUF1 p45 promotes West Nile virus replication by an RNA chaperone activity that supports cyclization of the viral genome. *J. Virol.*, **88**, 11586–11599.
69. Friedrich, S., Engelmann, S., Schmidt, T., Szczepankiewicz, G., Bergs, S., Liebert, U.G., Kummerer, B.M., Golbik, R.P. and Behrens, S.E. (2018) The host factor AUF1 p45 supports flavivirus propagation by triggering the RNA switch required for viral genome cyclization. *J. Virol.*, **92**, e01647-17.
70. Lund, N., Milev, M.P., Wong, R., Sanmuganatham, T., Woolaway, K., Chabot, B., Abou Elela, S., Moulant, A.J. and Cochrane, A. (2012) Differential effects of hnRNP D/AUF1 isoforms on HIV-1 gene expression. *Nucleic Acids Res.*, **40**, 3663–3675.
71. Ullmer, W. and Semler, B.L. (2018) Direct and indirect effects on viral translation and RNA replication are required for AUF1 restriction of enterovirus infections in human cells. *mBio*, **9**, e01669-18.
72. Hillebrand, F., Peter, J.O., Brillen, A.L., Otte, M., Schaal, H. and Erkelenz, S. (2017) Differential hnRNP D isoform incorporation may confer plasticity to the ESSV-mediated repressive state across HIV-1 exon 3. *Biochim. Biophys. Acta Gene Regul. Mech.*, **1860**, 205–217.
73. Cerasuolo, A., Buonaguro, L., Buonaguro, F.M. and Tornesello, M.L. (2020) The role of RNA splicing factors in cancer: regulation of viral and human gene expression in Human Papillomavirus-related cervical cancer. *Front. Cell Dev. Biol.*, **8**, 474.
74. Zucconi, B.E. and Wilson, G.M. (2011) Modulation of neoplastic gene regulatory pathways by the RNA-binding factor AUF1. *Front. Biosci. (Landmark Ed.)*, **16**, 2307–2325.

SUPPLEMENTARY FIGURES LEGENDS

Supplementary Figure S1

(A) Schematic structures of pC97ELsLuc and representative HPV16 early transcripts produced from pC97ELsLuc and spliced or not spliced at SD880. Plasmid pC97ELsLuc encodes all HPV16 genes. Early promoter P97 was replaced by CMV promoter. Secreted luciferase (sLuc) gene was integrated in the L1 gene following the poliovirus 2A internal ribosomal entry site (IRES) sequence. Numbers refer to the HPV16 reference strain GeneBank: K02718.1. Early and late genes are indicated. Black oval: splice donor. White oval: splice acceptor. pAE: HPV16 early polyadenylation site. pAL: HPV16 late polyadenylation site. Arrows indicate primer annealing positions. **(B)** Indicated hnRNP D isoform-encoding plasmids were cotransfected with pC97ELsLuc into HeLa cells. RNA was extracted and subjected to RT-PCR with primers 97S and E1AS. This RT-PCR reaction detected mRNAs that were unspliced at SD880. The authenticity of intron-retained E6E7E1 mRNA band (intron-retained, IR) was determined by sequencing and confirmed by the reverse transcriptase negative control assay on the same RNA samples. **(C-E)** Dose-dependent effect of hnRNP D40 was monitored by cotransfecting pC97ELsLuc with serially diluted hnRNP D40 encoding plasmid. RNA was extracted and subjected to RT-PCR with HPV16 RT-PCR primers 97S+880AS **(C)**, 773S+E2AS **(D)** or 773S+E4AS **(E)**. **(F)** The four mRNAs isoforms of hnRNP D in indicated cell lines were analyzed by RT-PCR. **(G)** Exogenous expression of hnRNP D40 (lane 3) was determined to compare with endogenous hnRNP D (lane 1). **(H)** The exogenous expression of each isoform of hnRNP D protein in HeLa cells were analyzed by Western-blotting using anti-FLAG antibody.

Supplementary Figure S2

Determination of the consistent hnRNP D effect on HPV16 early mRNA splicing both from pC97ELsLuc and pBELsLuc. (A and B) Schematic representation of HPV16 subgenomic plasmid pC97ELsLuc, pBELsLuc and representative HPV16 transcripts investigated in (C) and (D) are shown. Plasmid pC97ELsLuc encodes all HPV16 genes. Early promoter P97 was replaced by human cytomegalovirus immediate early promoter (CMV). Plasmid pBELsLuc encodes all HPV16 genes except E6 and E7 and is driven by the CMV promoter. Secreted luciferase (sLuc) gene was integrated in the L1 gene following the poliovirus 2A internal ribosomal entry site (IRES) sequence. Early and late genes are indicated. Black oval: splice donor. White oval: splice acceptor. pAE: HPV16 early polyadenylation site. pAL: HPV16 late polyadenylation site. Numbers refer to the HPV16 reference strain GeneBank: K02718.1. Schematic representation of HPV16 alternatively spliced mRNAs produced from pBELsLuc are shown below pBELsLuc. Arrows indicate annealing positions of HPV16 RT-PCR primers. **(C)** hnRNP D37 or hnRNP D40 encoding plasmid was cotransfected with pC97ELsLuc or pBELsLuc, RNA was extracted and HPV16 E1/E2 mRNA splicing or E4 mRNA splicing was determined by RT-PCR with primer pairs 773S+E2AS or 773S+E4AS, respectively. **(D)** Serially diluted hnRNP D40 plasmid was cotransfected with pC97ELsLuc or pBELsLuc. RNA was extracted and HPV16 E1/E2 mRNA splicing was determined by RT-PCR with primer pair 773S+E2AS. intron-retained (IR) E1 mRNAs and spliced 880^Δ2709 mRNAs are indicated.

Supplementary Figure S3

Subcellular localization of hnRNP D40 and hnRNP D40-mutants fused to EGFP. (A) Higher magnifications of microscopic images adapted from **Fig. 4C**. A cell highlighted by white arrowhead in each image is the same cell that is highlighted in **Fig. 4C**. Enhanced green fluorescent protein gene was fused in frame with the entire hnRNP D40 open reading frame or mutants thereof as indicated. EGFP(-), plasmid expressing EGFP not fused to any protein. **(B and C)** Proportion of each alternatively spliced HPV16 E6E7 mRNA isoforms (intron-retained E6, 226^Δ409, 226^Δ526 and 226^Δ742) or E1E2 mRNA isoforms (intron-retained E1, 880^Δ2582 and 880^Δ2709) over all spliced isoforms displayed in **Fig. 3D or 3E**. The proportion of a spliced isoform was calculated as a percentage of an isoform band intensity over the sum of all spliced isoform band intensities in each lane. Percentage bar graphs of all isoforms were built with mean values of three independent replicates. **(D)** Longer exposures of gel images of HPV16 E1E2 mRNA RT-PCR in **Fig. 3D and E** (773S+E2QAS) to enhance detection of the longer and inefficiently amplified cDNAs representing intron-retained E1 mRNAs.

Supplementary Figure S4

(A) hnRNP D40 and hnRNP D40 mutants had similar effect on HPV16 mRNA splicing in 293T cells (upper panel) and HeLa cells (lower panel). Plasmids expressing hnRNP D40 or the indicated hnRNP D40-mutants were individually cotransfected with pC97ELsLuc into 293T cells or HeLa cells, RNA was extracted and HPV16 E6E7 mRNA-splicing was analyzed by RT-PCR with HPV16 RT-PCR primers 97S+880AS. GAPDH, gapdh RT-PCR with primers GAPDHF and GAPDHR. **(B)** Subcellular localization of the EGFP-AGG mutant that in which "R" in the RG/RGG-motifs is replaced by "A" as described in **Fig. 5A**. Higher magnification image of EGFP-AGG was shown as "EGFP-AGG (H)". The same cells were highlighted with white arrowhead. **(C)** Proportion of the indicated spliced isoforms of HPV16 mRNAs quantitated from **Fig. 5I and J**. Percentage of the indicated spliced isoform over the sum of all spliced isoforms in each lane was calculated as described in **Fig. 4H-K**. Significance was calculated as described in Materials and Methods. **(D)** The coimmunoprecipitation of hnRNP D40 protein and SF3b or PABP-C1 analyzed as described in Figure 5K.

Supplementary Figure S5

(A) HPV16 subgenomic plasmid pC97ELsLuc encodes all HPV16 genes. Early promoter P97 was replaced by human cytomegalovirus immediate early promoter (CMV). Secreted luciferase (sLuc) gene was integrated in L1 gene following the poliovirus 2A internal ribosomal entry site (IRES) sequence. Early and late genes are indicated. Black opal: splice donor. White opal: splice acceptor. pAE: HPV16 early polyadenylation site. pAL: HPV16 late polyadenylation site. Numbers refer to the HPV16 reference strain GeneBank: K02718.1. "S" in red box represents a previously identified splicing silencer. **(B-E)** Schematic representations of HPV16 subgenomic expression plasmids pX656 **(B)**, pX478 **(C)**, pXH856F **(D)** and pXH856SDmF **(E)** used in **Figure 6**. HA-tags and FLAG-tags fused to HPV16 E6 and E7, respectively, in pXH856F and pX856SDmF are indicated. Wild type SD226 (GAG|GUUAUAUGA : a

vertical line indicates 5' splice site) in pXH856F was changed to (GAG|GCCUAUGA: underline indicates changed nucleotide) for SD226 inactivation in pXH856SDmF. RT-PCR primers are indicated. The various alternatively spliced mRNAs produced by these plasmids are indicated. CMV: CMV promoter.

Supplementary Figure S6

Quality controls of nuclear/cytoplasmic fractionation for Figure 7A by Western-blotting analysis using anti-Lamin B, anti-SRSF1 or anti-SRSF2 antibody.

Supplementary Figure S7

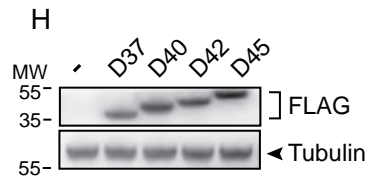
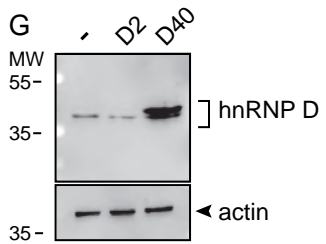
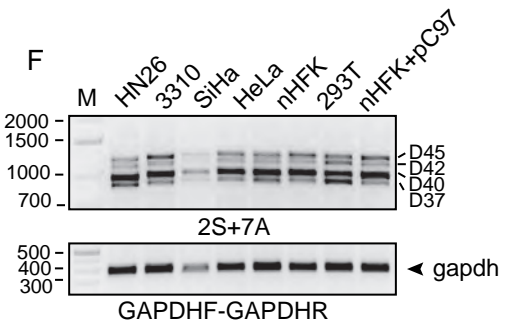
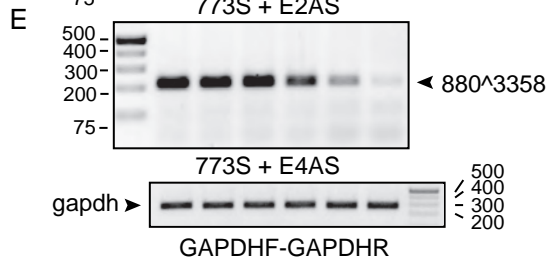
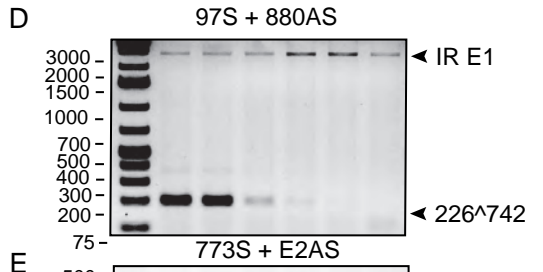
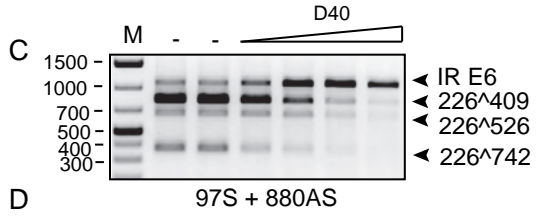
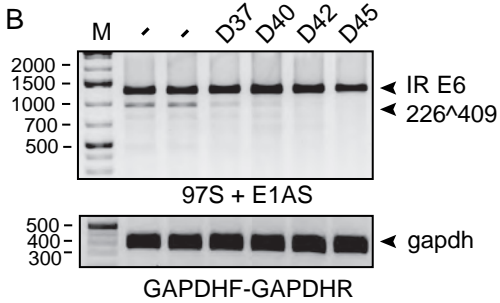
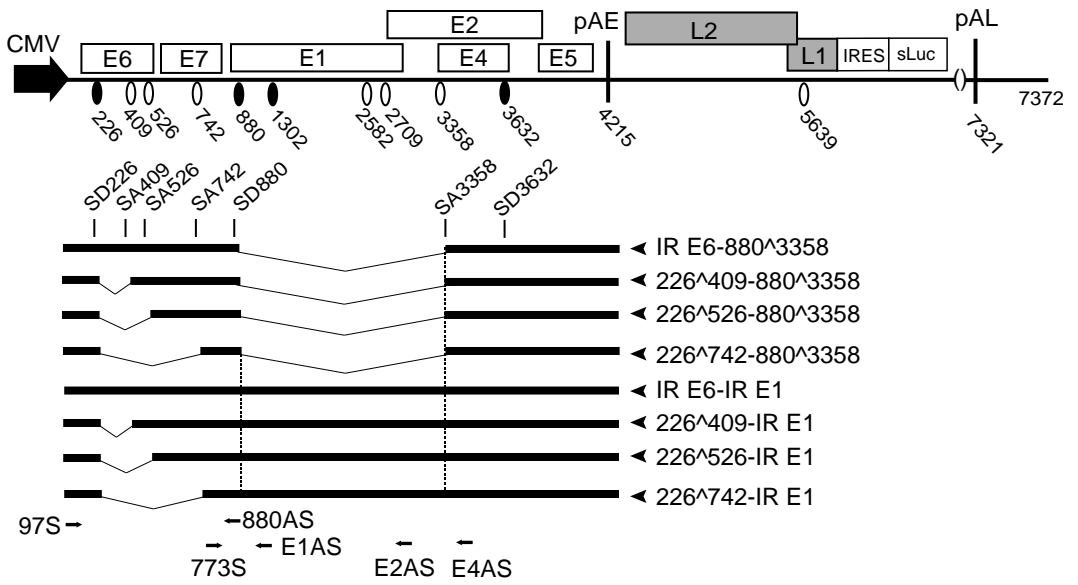
(A) HPV16 subgenomic plasmid pBELsLuc encodes all HPV16 genes except E6 and E7 and is driven by the CMV promoter. Numbers refer to the HPV16 reference strain GeneBank: K02718.1. Early and late genes are indicated. Black oval: splice donor. White oval: splice acceptor. pAE: HPV16 early polyadenylation site. pAL: HPV16 late polyadenylation site. Secreted luciferase (sLuc) gene was integrated in L1 gene following the poliovirus 2A internal ribosomal entry site (IRES) sequence. **(B and C)** Schematic drawings of HPV16 subgenomic plasmids pBELEN and pBELEndE1 of which the latter harbors a deletion inside the E1 gene (Δ nt1150-2481). Arrows indicate RT-PCR primer annealing positions. **(D)** Effect of hnRNP D40 overexpression on HPV16 E1/E2 mRNA splicing using HPV16 reporter plasmid pBELEN. pBELEN was transfected into HeLa cells in the absence (-) or presence (+) of hnRNP D40 expression plasmid. RNA was extracted and RT-PCR was performed with HPV16 primers 773S+E2Xba. intron-retained E1 mRNAs and spliced 880[^]2709 mRNAs are indicated. **(E)** HPV16 subgenomic plasmid pBELEN was transfected into HeLa cells in the absence (-) or presence (+) of hnRNP D40 expression plasmid. The transfected cells were fractionated into nuclear and cytoplasmic fractions, RNA was extracted and RT-PCR was performed with HPV16 primers 773S+E2Xba. intron-retained E1 mRNAs and spliced 880[^]2709 mRNAs are indicated. Unspliced and spliced actin mRNAs were detected by RT-PCR to control for subcellular fractionation. N, nuclear fraction; C, cytoplasmic fraction. **(F)** Quantitation of the RT-PCR products. The percentage of cytoplasmic intron-retained E1 mRNA over the sum of cytoplasmic and nuclear levels were calculated as described in **Fig. 6D**. **(G and H)** Schematic representations of HPV16 subgenomic plasmids p16E1-3xF and p16E1SDm-3xF encoding HPV16 E1 tagged with three FLAG-tags. Wild type SD880 (CAG|GUACCA: a vertical line indicates 5' splice site) and SD1302 (CAG|GUAGAA) in p16E1-3xF were changed to (CAG|CUACCA: underline indicates a changed nucleotide) for SD880 inactivation or (CAG|CUAGAA) for SD1302 inactivation in p16E1SDm-3xF, respectively. Numbers refer to the HPV16 reference strain GeneBank: K02718.1. Black oval: splice donor. White oval: splice acceptor. "x" represents mutational inactivation of HPV16 5'-splice sites SD880 and SD1302.

Supplementary Figure S8

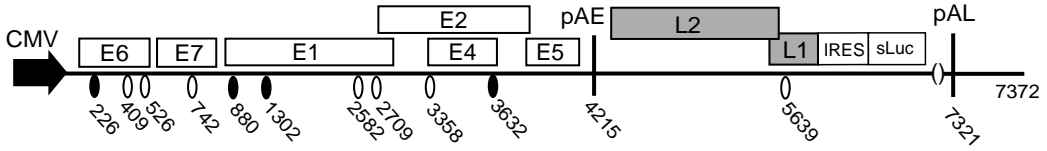
(A and B) Inhibitory effect of hnRNP D on HPV16 E6 or E1 mRNA translation was demonstrated by in vitro transcription and translation assay in the presence or absence of recombinant hnRNP D protein. T7 promoter driven tagged-HPV16 E6 expressing plasmid (A) or FLAG-E1 expressing plasmid (B) was added in vitro transcription and in vitro translation reaction mixture in the absence (-) or presence (+) of

recombinant hnRNP D protein or BSA. Resulted proteins were analyzed by Western-blotting using anti-E6 antibody (A) or anti-FLAG antibody (B). Splice donor sites of HPV16 E6 and E1 were mutated as pXH856SDmF or p16E1SDm-3xF, respectively. HPV16 E6 protein was fused with TrxA-6xHis-S-Tag-HA tag at the N-terminus to increase the molecular weight (predicted size is 36 kD), separating from the unspecific bands detected in Western-blotting. **(C)** Effect of hnRNP D on translation of luciferase control mRNA was analysed by luciferase assay. Instead of E6 or E1 expressing plasmid, luciferase expressing plasmid was used. **(D)** The protein amount of recombinant hnRNP D and BSA used for the reaction indicated in (A) and (B) was confirmed by Coomassie staining (final 0.5uM of proteins in a reaction). **(E)** The quality and identity of recombinant hnRNP D protein was confirmed by Western-blotting using anti-hnRNP D antibody.

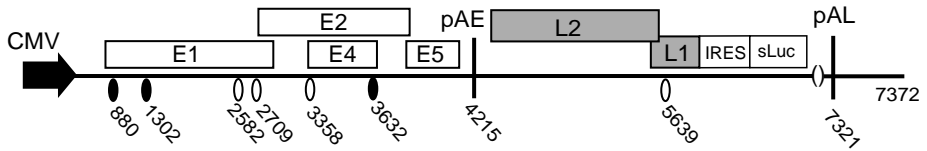
A pC97ELsLuc



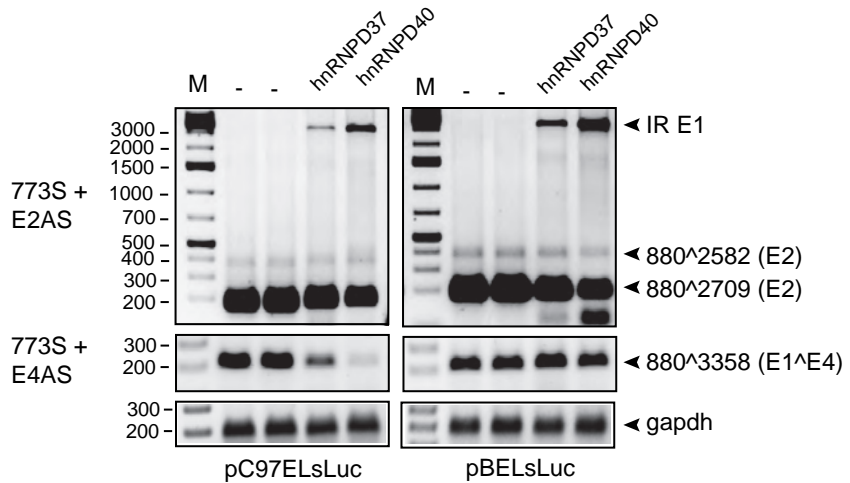
A pC97ELsLuc



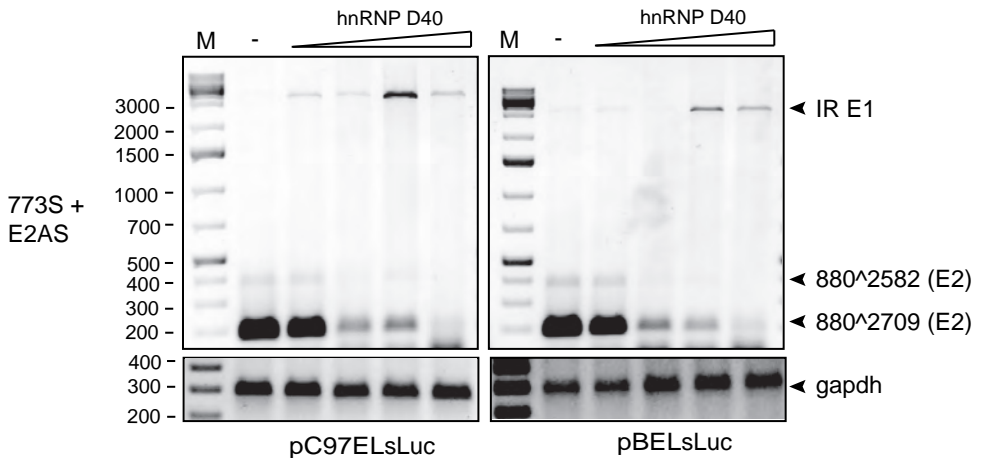
B pBELsLuc



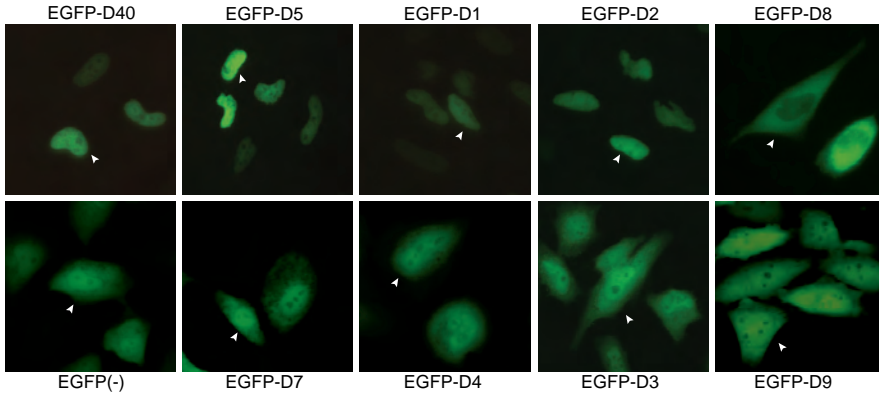
C



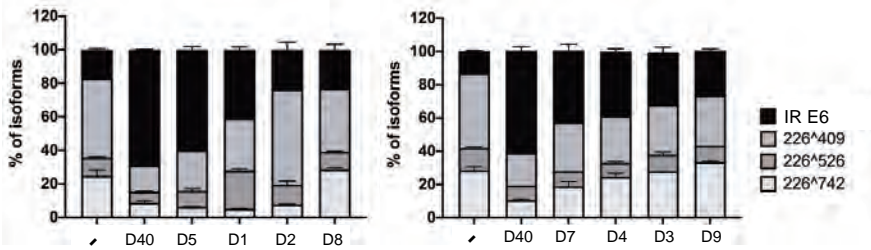
D



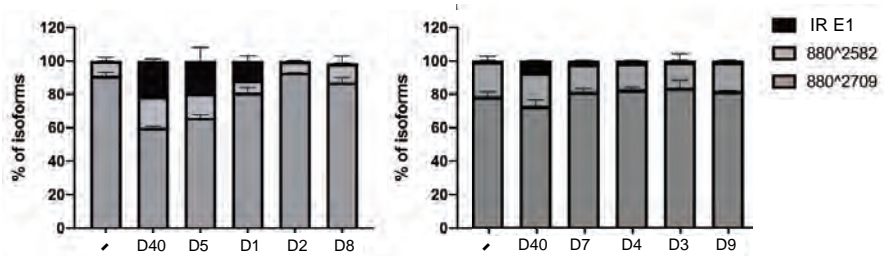
A



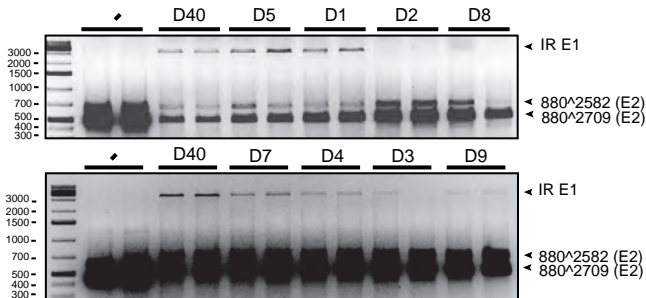
B

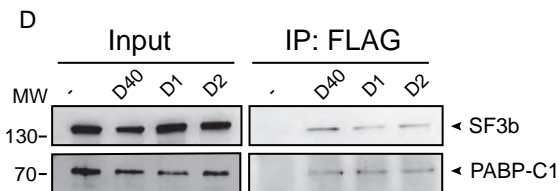
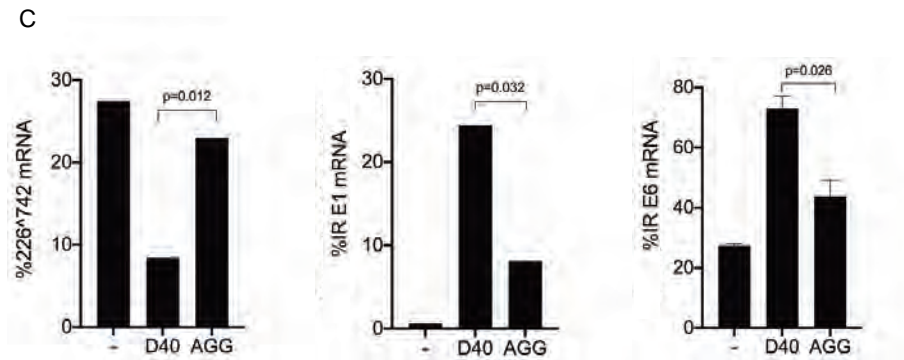
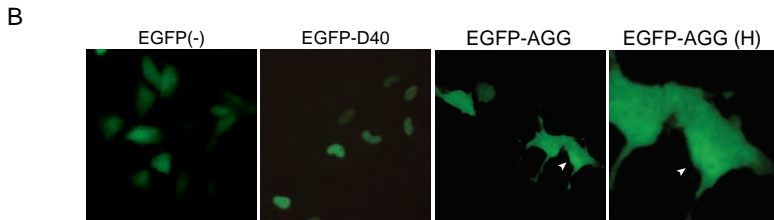
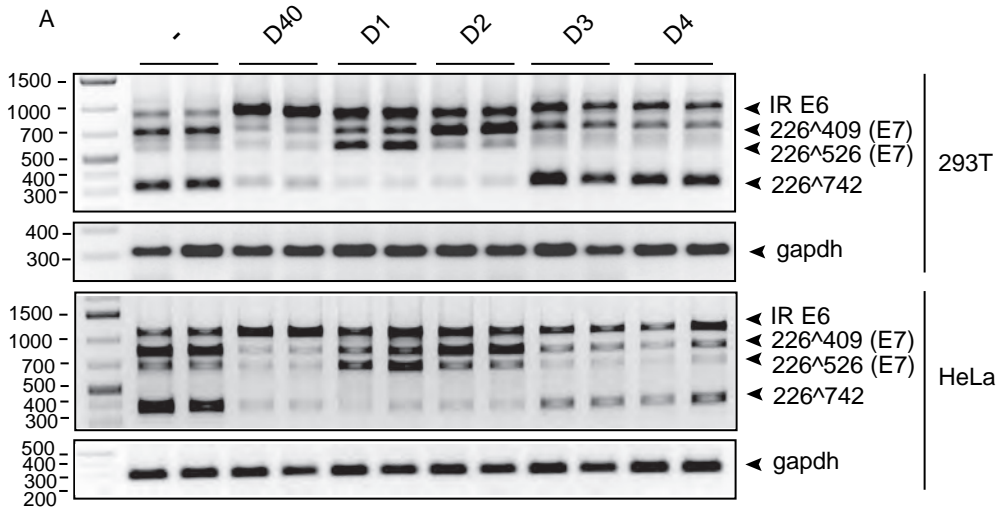


C



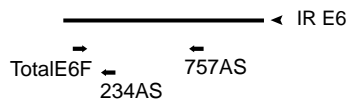
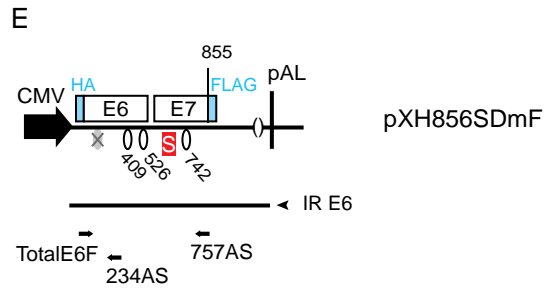
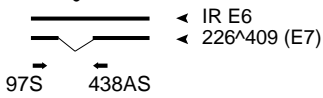
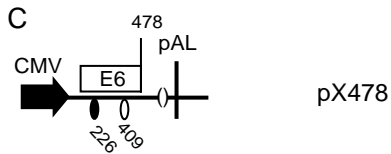
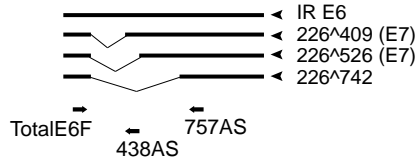
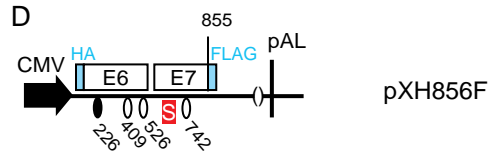
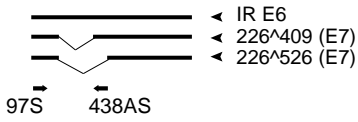
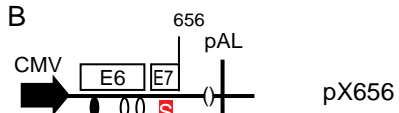
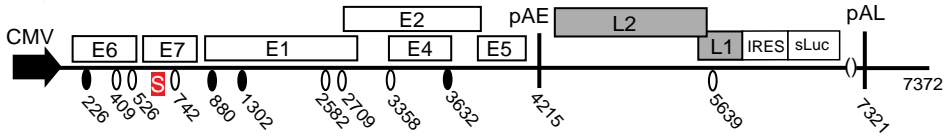
D





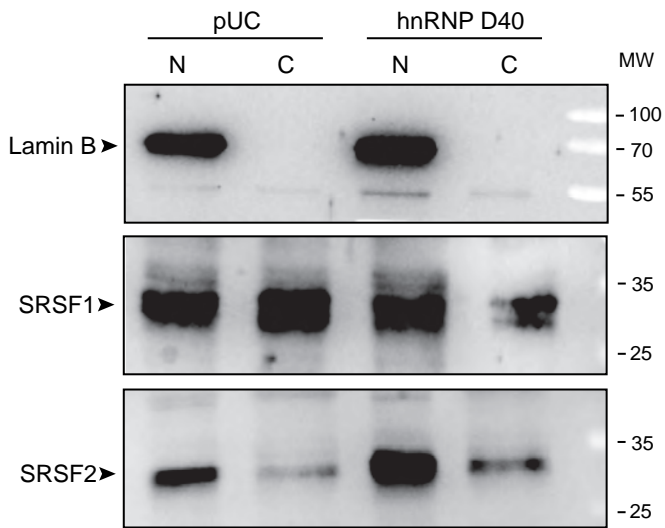
Supplementary Figure S5

A pC97ELsLuc

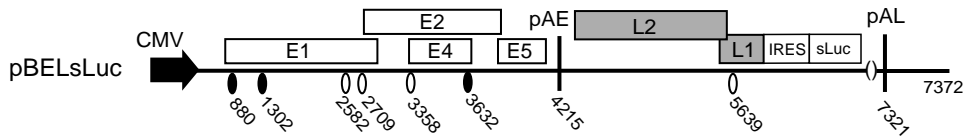


SD226	GAG GUAUAUGA
SD226m	GAG GCCUAUGA

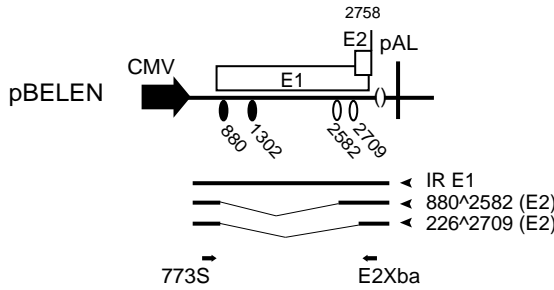
Supplementary Figure S6



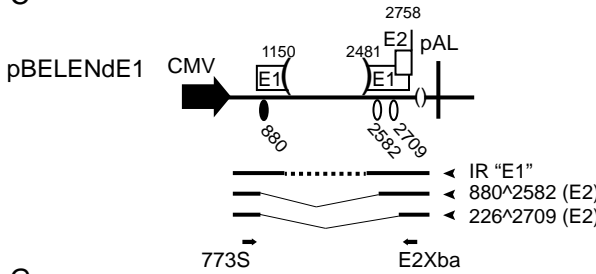
A



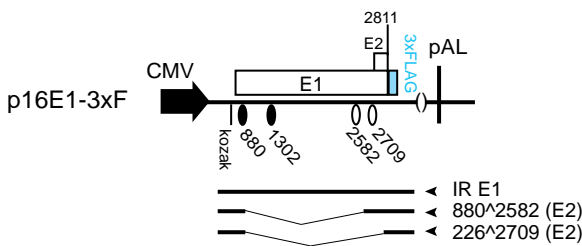
B



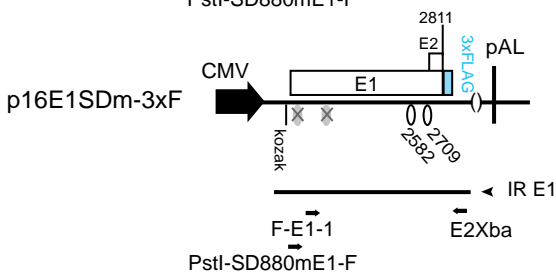
C



G



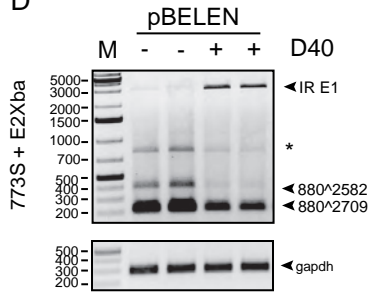
H



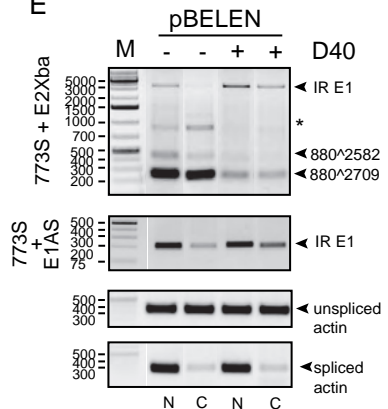
SD880 GAG|GUACCA
SD880m GAG|CUACCA

SD1302 GAG|GUAGAA
SD1302m GAG|CUAGAA

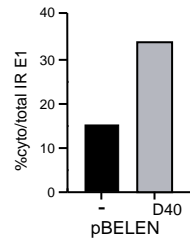
D

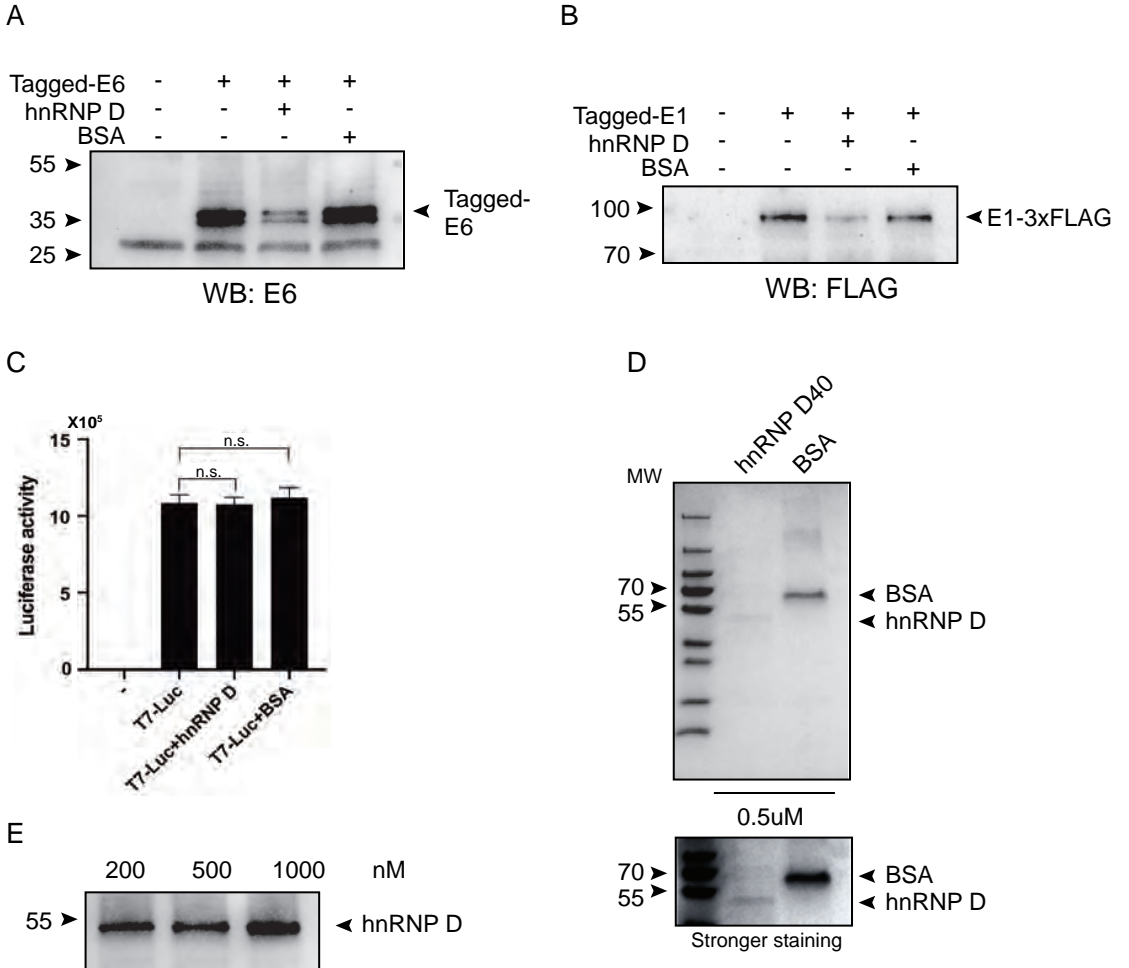


E



F





Supplementary Table 1. Sequences of PCR primers

Amplified region	primers	Sequence 5'-3'
E6/E7	97S	GTCCACCTGCAATGTTTCAGGACCC
	880AS	GAAACCATAATCTACCATGGCTGATC
	438AS	GCTCGAGGGAATCTTTGCTTTTTGTCCAGATGTCT
	757AS	CGTGTGTGCTTTGTACGCACAACCG
	TotalE6F	TGTTTCAGGACCCACAGGAG
	TotalE6R	TTGCTTGCAGTACACACATTC
234AS	TGCATAAATCCCGAAAAGCAAATG	
E2	773S	GCACACACGTAGACATTCGTACTTTG
	E2AS	GTCCAGATTAAGTTTGCCACGAGGAC
	E2QAS	CAGCCAGCGTTGGCACCACCT
E4	773S	GCACACACGTAGACATTCGTACTTTG
	E4AS	CCTCTCCTGAAATATTAGGCAGCA
E1	773S	GCACACACGTAGACATTCGTACTTTG
	E1AS	CCATCCATTACATCCCCTACC
	1302S	CTGAAGTGAAACTCAGCAGATGTTAC
	2293AS	TAATGATTTACCTGTGTTAGCTGCACC
	880S	GAAACCATAATCTACCATGGCTGATC
	F-E1-1	AGTAGAGCTGCAAAAAGGAGATTA
	PstI-SD880mE1-F	GATCCTGCAGCTACCAATGGGGAAGAGGGTAC
E2Xba	AACCAGGTTCTAGACAAACTTAATCTGGACCACG	
LoxP	16S	TATGTATGGTATAATAAACACGTGTGTATGTG
	16A	GCAGTGCAGGTCAGGAAAACAGGGATTGGC
hnRNPD isoforms	2S	GAAGATTGACGCCAGTAAGAAGC
	7A	TACCTTCCATAAACCCTCTGCT
pX478	B97S	GCGCCGTGCAATGTTTCAGGACCC
	X478A	GCTCGAGGGAATCTTTGCTTTTTGTCCAGATGTCT
pX656	B97S	GCGCCGTGCAATGTTTCAGGACCC
	X656A	CCTCGAGCTGTCAATTAATGCTATAACAG
pXH856F	sense	GACAGCGCGCACCATGTACCATACGATGTTCCAGATTACGCTATGTTTCAGGACCCACAG
	anti sense	TTACTCGAGCTACTTATCGTGTGCTCCTTGTAATCTGGTTTCTGAGAACAGAT
pXH856SDmF	sense	CTGCGACGTGAGGCCATGACTTTTGC
	anti sense	GCAAAGTCATAGGCCCTACGCTCGAC
p16E1-3xF	sense	ATCAGCGCGCCACCATGGCTGATCCTGCAGGTAC
	anti sense	TAATCGATGTCATGATCTTTAATCACCCTCATGCTCTTTGTAGTCTAATGTGTAGTATTTTG
	anti sense	TCCCCTCGAGCTACTTGTGTCATGCTCCTTGTAATCGATGTCATGATC
p16E1SDm-3xF	sense	GATCCTGCAGCTACCAATGGGGAAGAGGGTAC
	anti sense	GGCGCCCTTCTAGCTGTAACATCTGCTGAGT
	anti sense	TACAGCTAGAAAGGGCGCCATGAGACTG
pflaq-D40	sense	GTTGCGCGCCTTATCATGGACTACAAGACGATGACGACAAG
	anti sense	CCTCGAGTTAGTATGGTTTGTAGCTATTTTGATGACCACC
pD1.pD1-AGG and pD1-Q6A	sense	GGCGCGCCACCACCTGGACTACAAGACGATGACGACAAGATGTTTATAGGAGGCCCTTAGCTGGG
	anti sense	CCTCGAGTTAGTATGGTTTGTAGCTATTTTGATGACCACC
pD2.pD2-AGG and pD2-Q6A	sense	GGCGCGCCACCACCTGGACTACAAGACGATGACGACAAGATGTTTAAAAAATTTTTGTTGGTGGCC
	anti sense	CCTCGAGTTAGTATGGTTTGTAGCTATTTTGATGACCACC
pD5	sense	GTTGCGCGCCTTATCATGGACTACAAGACGATGACGACAAGGGAAGCGGAGCCGGGAC
	anti sense	CCTCGAGTTAGTATGGTTTGTAGCTATTTTGATGACCACC
pD3	sense	GTTGCGCGCCACCACCTGGACTACAAGACGATGACGACAAG
	anti sense	CCGCTCGAGTTAGGCTTTGGCCCTTTAGGATC
pD4	sense	GTTGCGCGCCACCACCTGGACTACAAGACGATGACGACAAG
	anti sense	CCGCTCGAGTTACATGGCTACTTTTATTTACATTTAC
pD7	sense	GTTGCGCGCCACCACCTGGACTACAAGACGATGACGACAAG
	anti sense	CTTCTCGAGTTAAGTACCCCACTGTTGCTGTTGCTGATATTG
pD8	sense	GGCGCGCCACCACCTGGACTACAAGACGATGACGACAAGATGTTTATAGGAGGCCCTTAGCTGGG
	anti sense	CCGCTCGAGTTACATGGCTACTTTTATTTACATTTAC
pEGFP-D40	sense	GTTAAGCTTGTATCATGGACTACAAGACGATGACGACAAG
	anti sense	CTTGGATCCTTAGTATGGTTTGTAGCTATTTTGATGACCACC
pEGFP-D1	sense	GTTAAGCTTGTATCATGGACTACAAGACGATGACGACAAG
	anti sense	CTTGGATCCTTAGTATGGTTTGTAGCTATTTTGATGACCACC
pEGFP-D2	sense	GTTAAGCTTGTATCATGGACTACAAGACGATGACGACAAG
	anti sense	CTTGGATCCTTAGTATGGTTTGTAGCTATTTTGATGACCACC
pEGFP-D5	sense	GTTAAGCTTGTATCATGGACTACAAGACGATGACGACAAG
	anti sense	CTTGGATCCTTAGTATGGTTTGTAGCTATTTTGATGACCACC
pEGFP-D3	sense	GTTAAGCTTGTATCATGGACTACAAGACGATGACGACAAG
	anti sense	CTTGGATCCTTAGGCTTTGGCCCTTTAGGATC
pEGFP-D4	sense	GTTAAGCTTGTATCATGGACTACAAGACGATGACGACAAG
	anti sense	CTTGGATCCTTAGTATGGCTACTTTTATTTACATTTAC
pEGFP-D7	sense	GTTAAGCTTGTATCATGGACTACAAGACGATGACGACAAG
	anti sense	CTTGGATCCTTAGTATGGCTACTTTTATTTACATTTAC
pEGFP-D8	sense	GTTAAGCTTGTATCATGGACTACAAGACGATGACGACAAG
	anti sense	CTTGGATCCTTAGTATGGCTACTTTTATTTACATTTAC

pEGFP-D9	sense	GTTAAGCTTGTTATCATGGACTACAAAGACGATGACGACAAG
	anti sense	CTTGGATCCTTACCTCCTATAAACATTTTCC
GAPDH	GAPDH	ACCCAGAAGACTGTGGATGG
	GAPDHR	TTCTAGACGGCAGGTCAGGT
spliced actin	actin-s	TGAGCTGCG TGTGGCTCC
	actin-a	GGCATGGGGGAGGGCATAACC
unspliced actin	actin-s1	CCAGT GGCTTCCCAGTG
	actin-a	GGCATGGGGGAGGGCATAACC
Involucrin	INV_S	CTCTGCCTCAGCCTTACTGTGA
	INV_R	GCTCCTGATGGGTATTGACTGG
pET32a-HA-E6SDm	sense	TATGAATTCATGTACCCATACGATGTTCCAGAT
	anti sense	TTACTCGAGCTATGGTTTCTGAGAACAGATGGGG

Supplementary Table 2. Antibodies used in WB, CLIP, RIP and Co-IP

	Cat.	Company
AUF1/hnRNP	12382S	Cell Signaling
hnRNPM	ab177957	abcam
hnRNPU	ab10297	abcam
Flag,M2	F3165/F1804	Sigma Aldrich
U1 70K	ab83306	abcam
U2AF65	sc-53942	Santa Cruz
U2AF35	ab172614	abcam
HPV16 E6	GTX132686	Genetex
HPV16 E7	GTX133411	Genetex
β -actin	A5441	Sigma Aldrich
Tubulin	T9026	Sigma Aldrich
Histone	4620S	CST
involucrin	sc-21748	Santa Cruz
HA	sc-7392	Santa Cruz
Lamin B	ab16048	abcam
SRSF1	ab38017	abcam
SRSF2	ab204916	abcam
SF3b	ab172634	abcam
PABP-C1	ab6125	abcam
veriblot	ab131366	abcam
Normal Rabbit IgG	2729S	Cell Signaling
Normal Mouse IgG	12-371	Millipore
anti-Mouse IgG-HRP	A9044	Sigma Aldrich
anti-Rabbit IgG-HRP	A9169	Sigma Aldrich

Supplementary Table 3. Sequences of oligos for pull down assay

HPV16 pull down region	Sequence 5'-3'
1, 178-214	UAUAAUAAUAGAAUGUGUGUACUGCAAGCAACAGUUA
2, 196-233	AGUUACUGCGACGUGAGGUAAUAGACUUUGCUUUUCGG
3, 229-266	UUCGGGAUUUAUGCAUAGUAUUAJAGAGAUAGGAAUCCA
4, 276-313	AUCCAUAUGCUGUAUGUGAUAAAUGUUUAAAGUUUUUAU
5, 309-346	UUUAAUUCUAAAAUUAGUGAGUAUAGACAUUAAUUGUUUAU
6, 342-379	GUUAAUGUUUGUAUGGAACAACAUAAGAACAGCAAUAC
7, 375-412	AAUACAACAACCGUUGUGUGAUUUGUUAAUUAGGUGU
8, 408-445	GGUGUAUUAACUGUCAAAAAGCCACUGUGUCCUGAAGAA
9, 441-478	AAGAAAAGCAAAGACAUCUGGACAAAAAGCAAAGAUUC
10, 474-498	GAUUCCAUAUAUAAGGGGUCGGUG
11, 492-516	GUCGGUGGACCGGUCGAUGUAUGUC
12, 506-530	CGAUGUAUGUCUUGUUGCAGAUCAU
13, 521-545	UGCAGAUCAUCAAGAACACGUAGAG
14, 536-560	ACACGUAGAGAAACCCAGCUGUAAU
15, 549-573	CCCAGCUGUAAUCAUGCAUGGAGAU
16, 564-588	GCAUGGAGAUACACCUACAUUGCAU
17, 579-604	UACAUUGCAUGAAUUAUUGUUAGAUU
18, 594-620	UAUGUUAGAUUUGCAACCAGAGACAAC
19, 611-635	CAGAGACAACUGAUCUCUACUGUUA
20, 631-665	UGUUAUGAGCAAUUAAAUGACAGCUCAGAGGAGGA
21, 661-695	GAGGAGGAUGAAAUAGAUGGUCCAGCUGGACAAGC
22, 691-725	CAAGCAGAACCGGACAGAGCCCAUUAACAUAUUGU
23, 721-755	AUUGUAACCUUUUGUUGCAAGUGUGACUCUACGCU
24, 751-785	ACGCUUCGGUUGUGCGUACAAAGCACACACGUAGA
25, 781-815	GUAGACAUUCGUACUUUGGAAGACCUGUAAAUGGG
26, 811-845	AUGGGCACACUAGGAUUUGUGUGCCCAUCUGUUC
27, 841-875	UGUUCUCAGAAACCAUAAUCUACCAUGGCUGAUCC
28, BSD+GGG	CUGAUCCUGCAGGUACCAAUGGGGAAGAGGGUACGGGAUGUA

Supplementary Table 4. Primer pair information

RT-PCR				
Amplified region	Sense primer	Antisense primer	Primer location	Shown in following Figures:
E6/E7	97S	880AS	Fig. 1C	Fig. 2A, 3B, 4D, 4F, 5B, 5I, 6C, 8B, 8F, 8K
	97S	438AS	Fig. 1C, 5L, S5B, S5C	Fig. 5L, 6C, 6I, 6M, 8I, 8J
	TotalE6F	Total E6R	Fig. 1C, Fig 5L	Fig. 5L
	TotalE6F	757AS	Fig. 1C, S5D, S5E	Fig. 6C
E1/E2	773S	E2AS	Fig. 1E, 3E	Fig. 2A, 3C, 3F, 7A, 8D, 8G
	773S	E2QAS	Fig. 1E	Fig. 4E, 4G, 5C, 5J
	773S	E2Xba	Fig. 1E, S7C	Fig. 7D
	F-E1-1	E2Xba	Fig. 1E, S7H	Fig. 7G
	PstI-SD880mE1-F	E2Xba	Fig. 1E, S7H	Fig. 7G
	1302S	2293AS	Fig. 1E, 3E	Fig. 3G
	773S	E1AS	Fig. 1E, 3E	Fig. 3H, 7J, 8E, 8K, 8L
E4(880*3358)	773S	E4AS	Fig. 1C	Fig. 2B, 8C, 8I

RT-qPCR				
Amplified region	Sense primer	Antisense primer	Primer location	Shown in following Figures:
Spliced E2 mRNA	773S	E2AS	Fig. 1E, 3E	Fig. 3D, 3J
IR E1 mRNA	773S	E1AS	Fig. 1E, 3E	Fig. 3D, 3J
IR E6 mRNA	TotalE6F	234AS	Fig. 1C, Fig S5	Fig. 6H

Paper II





Overexpression of m6A-factors METTL3, ALKBH5, and YTHDC1 alters HPV16 mRNA splicing

Xiaoxu Cui² · Kersti Nilsson² · Naoko Kajitani^{1,2} · Stefan Schwartz^{1,2}

Received: 4 November 2021 / Accepted: 12 February 2022 / Published online: 21 February 2022
© The Author(s) 2022

Abstract

We report that overexpression of the m6A-demethylase alkB homolog 5 RNA demethylase (ALKBH5) promoted production of intron retention on the human papillomavirus type 16 (HPV16) E6 mRNAs thereby promoting E6 mRNA production. ALKBH5 also altered alternative splicing of the late L1 mRNA by an exon skipping mechanism. Knock-down of ALKBH5 had the opposite effect on splicing of these HPV16 mRNAs. Overexpression of the m6A-methylase methyltransferase-like protein 3 (METTL3) induced production of intron-containing HPV16 E1 mRNAs over spliced E2 mRNAs and altered HPV16 L1 mRNA splicing in a manner opposite to ALKBH5. Overexpression of the nuclear m6A-“reader” YTH domain-containing protein 1 (YTHDC1), enhanced retention of the E6-encoding intron and promoted E6 mRNA production. We also show that HPV16 mRNAs are bound to YTHDC1 in human cells and that YTHDC1 affected splicing of HPV16 E6/E7 mRNAs produced from the episomal form of the HPV16 genome. Finally, we show that HPV16 mRNAs are m6A-methylated in tonsillar cancer cells. In summary, HPV16 mRNAs are methylated in HPV16-infected tonsillar cancer cells and overexpression of m6A-“writer” METTL3, m6A-“eraser” ALKBH5 and the m6A-“reader” YTHDC1 affected HPV16 mRNA splicing, suggesting that m6A plays an important role in the HPV16 gene expression program, at least in cancer cells.

Keywords Human papillomavirus · m6A · Splicing · YTHDC1 · METTL3 · ALKBH5

Introduction

Human Papillomaviruses (HPVs) are small DNA viruses associated with 99% of all cervical cancers, and a growing number of head and neck cancers [1, 2]. The most commonly cancer associated HPV type is HPV16 that is present in > 50% of the diagnosed HPV-caused cancers [3]. HPV16 replicates in the nucleus of keratinocytes and viral gene expression is linked to the differentiation program of the cell [4, 5]. The HPV16 genome contains two promoters (P97 and P670) and two polyadenylation signals (pAE and pAL) that separate early and late HPV genes [6] (Fig. 1A). The HPV16 E2 protein E2 regulates HPV gene expression

by inhibiting the HPV early promoter and the early polyadenylation signal, thereby inducing a switch from early to late gene expression [7–9]. The pre-mRNAs produced from either HPV16 promoter are subject to RNA processing resulting in a plethora of alternatively spliced and polyadenylated mRNAs [10]. HPV16 mRNA splicing is regulated by multiple cis-acting regulatory RNA elements on the HPV16 mRNAs and their cognate cellular, trans-acting splicing factors [10–13]. HPV16 is totally dependent on the cellular splicing machinery for production of the mature mRNAs that produce HPV16 proteins.

Cellular mRNAs carry various chemical modifications that contribute to the regulation of RNA processing. m6A is believed to be the most common modification of mRNAs in eukaryotic cells. The m6A modification is reversible and is catalyzed by a methyltransferase complex consisting of methyltransferase-like-3 (METTL3), methyltransferase-like-14 (METTL14), and additional factors such as Wilms tumor 1-associated protein (WTAP), RNA-binding motif protein 15 (RBM15), KIAA1429, and zinc-finger CCCH-type 13 containing (CZ3H13) [14, 15]. Of these, METTL3 is the catalytic subunit, while other factors such as

Edited by Hartmut Hengel.

✉ Stefan Schwartz
stefan.schwartz@imbim.uu.se

¹ Department of Medical Biochemistry and Microbiology, Uppsala University, BMC-B9, 751 23 Uppsala, Sweden

² Department of Laboratory Medicine, Lund University, BMC-B13, 221 84 Lund, Sweden

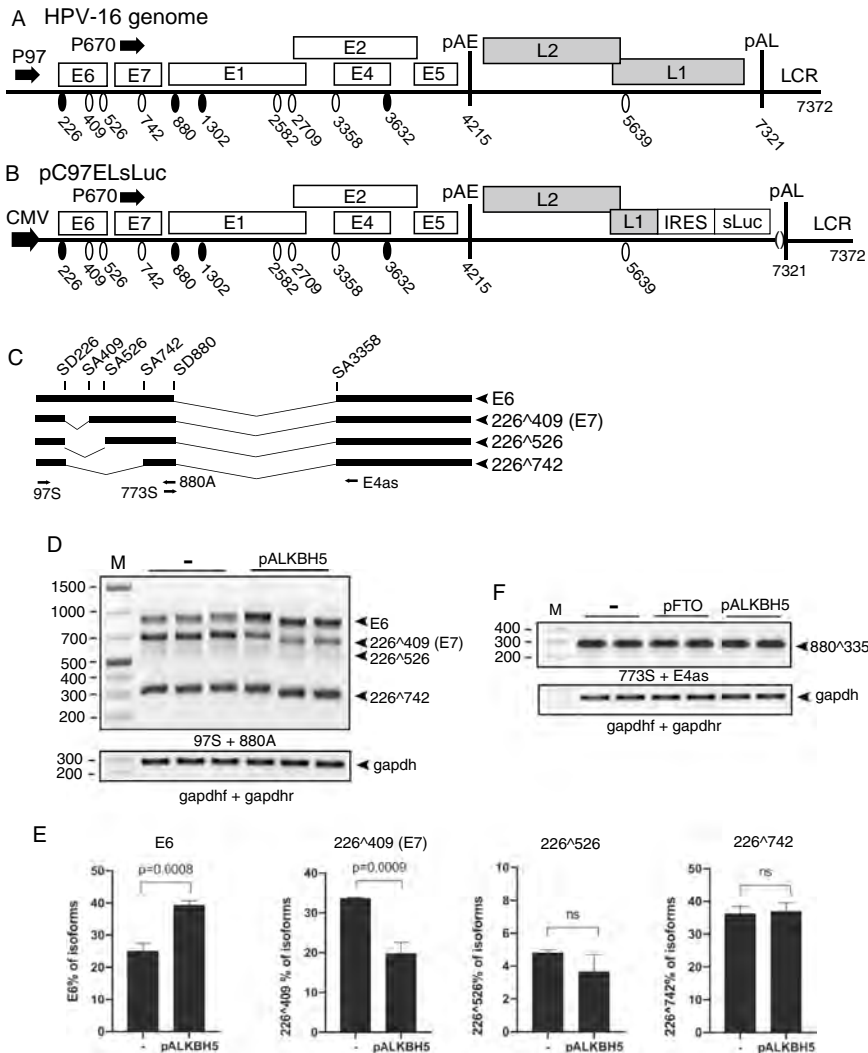


Fig. 1 **A** Linearized HPV16 genome (numbers refer to the HPV16 reference strain GeneBank: K02718.1). Early and late genes are indicated. P97: HPV16 early promoter. P670: HPV16 late promoter. Black oval: splice donor. White oval: splice acceptor. pAE: HPV16 early polyadenylation site. pAL: HPV16 late polyadenylation site. LCR: HPV16 long control region. **B** HPV16 subgenomic plasmid pC97ELsLuc encodes all HPV16 genes. HPV16 early promoter P97 was replaced by human cytomegalovirus immediate early promoter (CMV). Secreted luciferase (sLuc) gene was integrated in the L1 gene following the poliovirus 2A internal ribosomal entry site (IRES) sequence. **C** Schematic structures of HPV16 early transcripts produced from pC97ELsLuc. Splicing at SD226 occurs independently

of splicing at the downstream SD880 and generates splice variants in the E6- and E7-coding region. Arrows indicate HPV16 RT-PCR primers. **D** Effect of ALKBH5 on HPV16 E6/E7 mRNA splicing was monitored by RT-PCR with indicated primer pair on RNA extracted from triplicate transfections of pC97ELsLuc with empty pUC plasmid (-) or pALKBH5. **E** Densitometric quantification of the results in E. The percentage of each indicated cDNA isoform of all cDNAs in a lane in the absence of ALKBH5 (-) or in the presence of ALKBH5 overexpression is displayed. **F** Effect of FTO or ALKBH5 on HPV16 E4 mRNA splicing was monitored by RT-PCR with indicated primer pair on RNA extracted from HeLa cells transfected with pC97ELsLuc and empty pUC plasmid (-), pFTO, or pALKBH5

METTL14 plays other essential roles in RNA recognition and maintenance of complex stability. Two demethylases have been identified: alkylated DNA repair protein AlkB homolog 5 (ALKBH5) and fat mass and obesity-associated protein (FTO), that both remove m6A, but with different and independent, enzymatic mechanisms [16, 17]. Thus, m6A-methylation of mRNAs is a reversible process. Taken together, it appears that m6A-methylation plays an important regulatory role in mRNA splicing.

It has been shown that m6A modifications can directly interfere with interactions between RNA and RNA-binding proteins [18, 19]. For example, secondary structures on mRNAs may be affected by the presence of m6A. Thus, m6A may affect binding of splicing factors hnRNPC and hnRNPG to their RNA target sequences [20, 21]. Proteins with an YTH-domain bind m6A-modified RNAs and affect RNA-localization, stability, polyadenylation, and translation [14, 22, 23]. The nuclear, m6A-binding YTH-protein YTHDC1 has been shown to affect splicing by specific recruitment of SR-proteins [24]. YTHDC1 also interacts with hnRNPA2 to mediate miRNA splicing [25]. Knock-down of METTL3 or deficiency in either of the demethylases ALKBH5 or FTO has been shown to affect alternative splicing of thousands of cellular mRNAs.

Since the discovery of m6A, its presence on viral RNAs has been demonstrated for a wide range of viruses, including Flaviviruses that replicate in the cytoplasm [26]. Studies aimed at unraveling the biological significance of viral RNA-methylation suggested effects on the human herpesvirus type 8 (HHV8/KSHV) lytic cycle [27], replication capacity of HIV-1 [28, 29], SV40 replication [30], and adenovirus replication and mRNA splicing [31]. HPV16 was recently shown to produce circular E6/E7 mRNAs that were m6A-methylated [32], suggesting that m6A-methylation may contribute to regulation of HPV16 gene expression at the level of RNA splicing. Since HPV-mRNAs are processed by the cellular RNA processing machinery [11], it is reasonable to speculate that m6A “writers”, “reader”, and “erasers” could affect HPV16 mRNA splicing. We therefore investigated if the m6A “writers” METTL3, METTL14 and WTAP, the m6A “erasers” ALKBH5 and FTO and the nuclear, m6A “reader” YTHDC1 could affect HPV16 mRNA splicing. We found that ALKBH5 and YTHDC1 overexpression promoted intron retention in the HPV16 early E6-coding region thereby enhancing E6 mRNA production. Furthermore, METTL3 overexpression enhanced intron retention in the HPV16 E1-coding region, thereby promoting production of E1 mRNAs over spliced E2 mRNAs. On the HPV16 late mRNAs, ALKBH5 overexpression caused skipping of the internal exon in the HPV16 E4-coding region of the HPV16 late L1 mRNAs, suggesting that m6A also control late gene expression of HPV16. We concluded that various m6A “writers”,

“erasers”, and “readers” altered HPV16 mRNA splicing, suggesting an important role for these factors in the HPV16 gene expression program.

Materials and methods

Cells

HeLa and C33A2 cells were cultured in Dulbecco’s modified Eagle medium (DMEM) (GE Healthcare Life Science HyClone Laboratories) with 10% heat-inactivated fetal bovine calf serum (GE Healthcare Life Sciences HyClone Laboratories) and penicillin/streptomycin (Gibco Thermo Fisher Science). HN26 cells were cultured in RPMI-1640 medium (GE Healthcare Life Science HyClone Laboratories) supplemented with 10% Bovine Calf Serum (GE Healthcare Life Science HyClone Laboratories), 1% Sodium Pyruvate solution (Sigma-Aldrich), 1% Non-Essential Amino Acid (Sigma-Aldrich), and Gentamycin. The C33A2 cell line originates from C33A and has the subgenomic HPV16 plasmid pBELsLuc (Fig. 2A, B) stably integrated into the genome [33]. In the pBELsLuc construct, part of the L1 coding region have been exchanged for a poliovirus 2A internal ribosome entry site (IRES) in front of the *Metridia longa* secreted luciferase (sLuc) gene. Induction of HPV16 late gene expression results in the secretion of sLuc into the cell culture medium. The HN26 cell line has previously been described and is a tonsillar cancer cell line with episomal HPV16 DNA [34, 35]. Briefly, 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.

Plasmids and transfections

The following plasmids have been described previously: pC97ELsLuc [36, 37], pBELsLuc [36, 37], pHPV16AN [36, 37], and pX856F [38]. pcDNA3-FLAG-HA-hYTHDC1 (#85167), pcDNA3/Flag-METTL3 (#53739), pcDNA3/Flag-METTL14 (#53740), pcDNA3/Flag-WTAP (#53741), pFRT/TO/HIS/FLAG/HA-ALKBH5 (#38073) were purchased from Addgene. pcDNA3.1⁺/C-(K)DYK-FTO was purchased from GenScript. Transfections were made with Turbofect according to the manufacturer’s protocol (Fermentas). Briefly, a mixture of 2 μ l Turbofect per 1 μ g DNA and 100 μ l of DMEM without serum was incubated at room temperature for 25 min prior to dropwise addition to 60-mm plates with subconfluent HeLa cells.

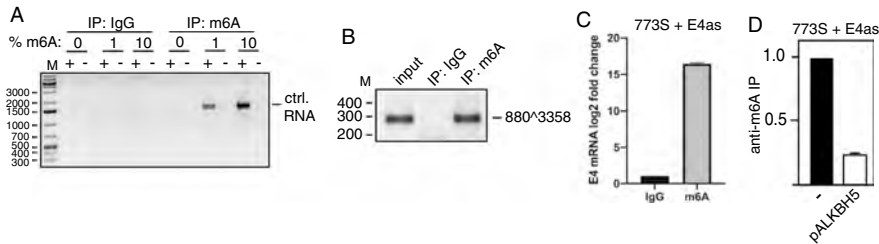


Fig. 2 **A** Vector RNA was synthesized *in vitro* (Supplementary Fig. S1) in the presence of 0%, 1%, or 10% m6A-nucleotide and subjected to immunoprecipitation with either IgG or anti-m6A antibody followed by RT-PCR with T7 forward and reverse primers (Supplementary Table 1). **B** HeLa cells transfected with pC97ELSLuc were subjected to CLIP assay. The RNA–protein complexes were immunoprecipitated with IgG or by anti-m6A antibody followed by RT-PCR with primers 773S and E4as on the extracted RNA. **C** HeLa cells transfected with pC97ELSLuc were subjected to CLIP assay. The RNA–protein complexes were immunoprecipitated with IgG

or by anti-m6A antibody followed by RT-qPCR with primers 773S and E4as on the extracted RNA. Fold difference between RT-qPCR on RNA immunoprecipitated with anti-m6A antibody over IgG is shown. **D** HeLa cells transfected by pC97ELSLuc in the absence or presence of pALKBH5 were subjected to CLIP assay. The RNA–protein complexes were immunoprecipitated with IgG or anti-m6A antibody followed by RT-qPCR with primers 773S and E4as on the extracted RNA. Fold difference between RT-qPCR on RNA immunoprecipitated with anti-m6A antibody over IgG is shown

In vitro RNA syntheses

In vitro RNA was synthesized using the TranscriptAid T7 High Yield Transcription Kit (#K0441, Thermo Scientific). The provided control template DNA was used according to the manufacturer's instruction, the sequence of the control DNA is listed in Supplementary Fig. S1. Three reactions were made in parallel in which 0-, 1-, or 10% of the ATP-pool were substituted for the N6-methyladenosine base analogue m6A (S3190, Selleckchem).

RNA extraction from transfections and RT-PCR

Total RNA was extracted 24 h and/or 48 h post-transfection using TRI Reagent (SIGMA Aldrich Life Science) and Direct-zol RNA MiniPrep (ZYMO Research) according to the manufacturer's protocols. cDNA was synthesized from 500 ng RNA in a 20 μ l reaction at 37 $^{\circ}$ C with M-MLV Reverse Transcriptase (Invitrogen) and random primers (Invitrogen) according to the protocol of the manufacturer. One microliter of cDNA was subjected to polymerase chain reaction (PCR) amplification.

UV-crosslinking and immunoprecipitation with m6A antibody

C33A2 and HN26 cells were grown in 10-cm dishes until reaching 80% confluency. Total RNA was then harvested using TriReagent (Invitrogen), followed by DNase I (Thermo Scientific) treatment for 1 h at 37 $^{\circ}$ C according to manufacturer's protocols. RNA was extracted by phenol/chloroform extraction followed by ethanol precipitation

and resuspension in 20 μ l H₂O. 20 of μ g RNA was then diluted in 450 μ l immunoprecipitation (IP) buffer (50 mM Tris pH 7.4, 100 mM NaCl, 0.05% NP40), and 5 μ g of either IgG or Anti-m6A antibody (Abcam rabbit polyclonal (ab151230)) and 5 μ l RI Ribolock (Thermo Scientific). In some experiments, the solutions were transferred into 3 cm cell culture dishes and crosslinked twice with 0.15 J cm² UV light (254 nm) in a Stratalinker (Agilent). Samples were incubated overnight on a rotating wheel at 4 $^{\circ}$ C. Thirty microliters (0.6 mg) of Dynabeads Protein G (#10004D) (Invitrogen) were added to the antibody-RNA mixture followed by incubation for 2 h, at 4 $^{\circ}$ C. The beads were washed twice on a magnetic stand with high-salt buffer (50 mM Tris pH7.4, 1 M NaCl, 1 mM EDTA, 1% NP40, 0.1% SDS), and 4 times with IP buffer. RNA was eluted by phenol/chloroform extraction. All immunoprecipitated antibody–RNA complexes were extracted directly from the immunoprecipitations, purified, and analyzed by RT-PCR and/or qRT-PCR. RNA was dissolved in 20 μ l of water and ten microliters were reverse transcribed using random primers (Invitrogen) and M-MLV reverse transcriptase for 50 min at 37 $^{\circ}$ C (Invitrogen). The remaining ten microliters served as negative controls. 1 μ l cDNA was subjected to RT-PCR with primers indicated in the figures (see Supplementary Table 1 for primer sequences). For quantification of immunoprecipitated RNA, RT-qPCR was performed on 1 μ l of cDNA synthesized as described above. RT-qPCR was performed in a MiniOpticon (Bio-Rad) using the Sso Advanced SYBR Green Supermix (Bio-Rad) according to the manufacturers protocol. For primer sequences, see Supplementary Table 1. To calculate relative expression levels of each gene, the C_T values

of the target gene from treated cells were normalized to the levels in DMSO treated cells.

Western blotting

Cell extracts for Western blotting were obtained by resuspending transfected cells in radioimmunoprecipitation assay (RIPA) buffer consisting of 50 mM Tris HCl pH 7.4, 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, 2 mM EDTA, 1 mM DTT, and protease inhibitor (Sigma Aldrich), followed by centrifugation at full speed for 20 min and collection of the supernatants. Proteins were denatured by boiling in Laemmli buffer. After SDS-PAGE, the proteins on the gels were transferred onto nitrocellulose membranes, blocked with 5% nonfat dry milk in PBS containing 0.1% Tween 20, and stained with specific primary antibodies (anti-Flag antibody M2 F1804 (Sigma Aldrich), anti-beta-actin antibody A5441 (Sigma Aldrich) or rabbit anti-ALKBH5 antibody #ab174124 (Abcam)) to the indicated proteins followed by incubation with secondary antibody conjugated with horseradish peroxidase and detection with chemiluminescence reagents.

Lentiviral based shRNA for knockdown

Lentivirus for the short hairpin RNA (shRNA)-mediated knockdown of ALKBH5 was generated by co-transfection of HEK293T cells with a pLKO.1-vector or pLKO.1-vector carrying specific shRNA together with the packaging vector pMISSION-GAG-POL and a vesicular stomatitis virus G protein expressing vector pMISSION-VSV-G. pLKO.1 was purchased from Sigma-Aldrich (SHC001). Two days post-transfection, lentivirus-containing supernatants were harvested, centrifuged to remove cellular debris, and filtered with a 0.45-µm filter. Lentivirus production efficiency was determined in parallel using a GFP overexpression lentivirus vector. C33A2 cells were inoculated with stocks of recombinant lentiviruses by centrifugation at 2000 g for 2 h at room temperature in the presence of 10 µg/ml polybrene (Fisher Scientific). Empty pLKO.1-vector was used as negative control. Cells were then resuspended and grown in normal RPMI media for 2 days, after which transduced cells were selected in the presence of puromycin (1 µg/ml). Cells were either harvested for Western blotting or for RNA extraction and RT-PCR.

UV-crosslinking and immunoprecipitation (CLIP)

Transfected HeLa cells grown in 10-cm dishes were washed by ice-cold PBS followed by crosslinking twice with 0.4 J/cm² UV light (254 nm) in a bio-link cross-linker (Biometra). Cytoplasmic extracts were prepared as described above. Whole cell lysates were prepared by resuspending cells in

one ml of RIPA buffer and incubated on ice for 30 min with occasional vortexing to lyse cells. For immunoprecipitations, 2 µg of the anti-flag antibody (M2, Sigma Aldrich) or mouse IgG was incubated at 4 °C overnight in 0.5 ml of cell lysate. 20 µl of Dynabeads Protein G (10004D, Invitrogen) and 20 µl Dynabeads Protein A (10001D, Invitrogen) were blocked with 1% BSA for 0.5 h, washed three times in RIPA buffer, and then added to the antibody–protein mixtures followed by incubation for 1 h at 4 °C. The beads were washed three times with buffer I (50 mM Tris HCl pH 7.4, 300 mM NaCl, 0.5% NP-40, 1 mM EDTA, 1 mM DTT), three times with buffer II (50 mM Tris HCl pH 7.4, 800 mM NaCl, 0.5% NP-40, 1 mM EDTA, 1 mM DTT), and three times with buffer III (50 mM Tris HCl pH 7.4, 800 mM NaCl, 250 mM LiCl, 0.5% NP-40, 1 mM EDTA, 1 mM DTT). RNA was eluted by phenol/chloroform extraction and ethanol-precipitated and dissolved in 20 µl of water. 10 µl of immunoprecipitated RNA was reverse transcribed using M-MLV reverse transcriptase (Invitrogen) and random primers (Invitrogen) according to the protocol of the manufacturer. Two microliters of cDNA were subjected to PCR amplification using HPV16-specific primers.

Quantitations

The software used to determine band intensity in Western blots and RT-PCR gels is “Image Lab 6.1.0” and quantitations were performed with the software “Prism GraphPad 8.4.0”.

Results

Overexpression of the m6A “eraser” ALKBH5 alters HPV16 alternative mRNA splicing

To investigate if m6A-“erasers” (ALKBH5, FTO), m6A-writers (METTL3, METTL14, WTAP) and m6A-readers (YTHDC1) affected HPV16 mRNA splicing, HPV16 reporter plasmid pC97ELsLuc (Fig. 1B) was cotransfected into HeLa cells with plasmids expressing ALKBH5, FTO, METTL3, METTL14, WTAP, or YTHDC1 (Supplementary Fig. S2A). First we analyzed the effect of FTO and ALKBH5 on alternatively spliced HPV16 E6/E7 mRNAs (Fig. 1C). The major 5'-splice site in the E6- and E7-coding region named SD226 may be spliced to either SA409, SA526, or SA742 (Fig. 1C). Splicing between HPV16 splice sites SD226 and SA409 generates the E7 mRNA, whereas the significance of SA526 and SA742 is less clear. Retention of the E6-encoding intron downstream of SD226 generates the E6-encoding mRNA. First, pC97ELsLuc was cotransfected with plasmids expressing flag-tagged ALKBH5 or FTO m6A eraser proteins (Supplementary

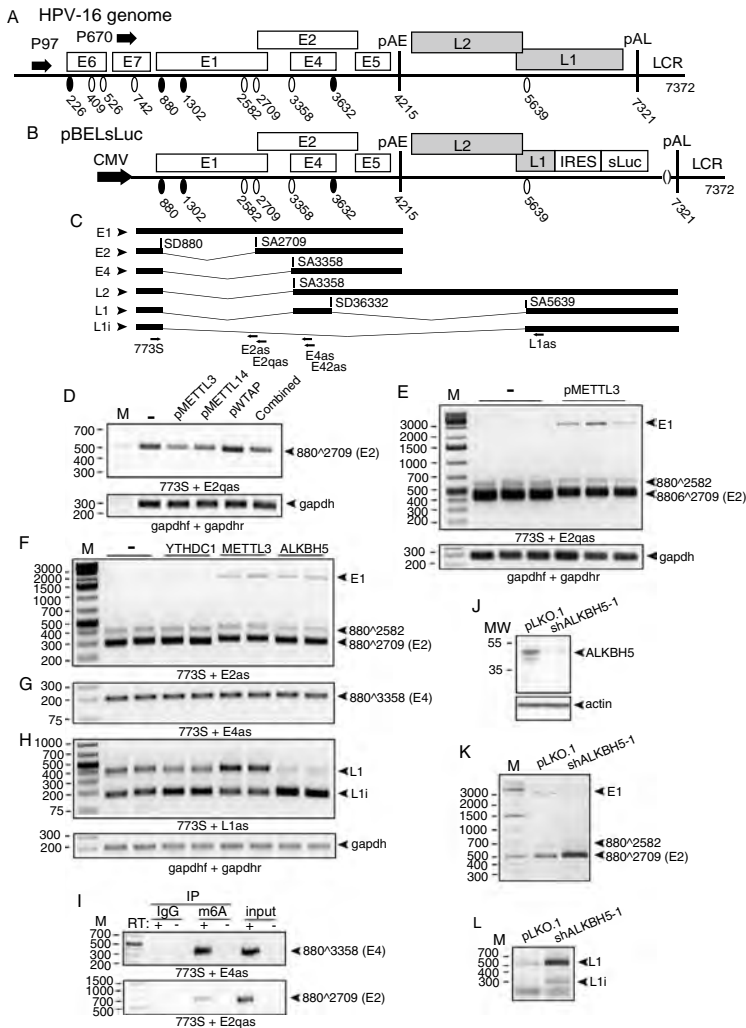


Fig. S2A). Triplicate transfections of ALKBH5 with pC97ELsLuc (Fig. 1D) followed by quantitation revealed that the splicing inhibitory effect of ALKBH5 on the HPV16 E6/E7 mRNAs was subtle, but reproducible (Fig. 1E). Splicing from SD880 to SA3358 appeared unaffected by FTO and ALKBH5 (Fig. 1F). FTO did not affect HPV16 E6/E7 mRNA splicing (data not shown). In conclusion, overexpression of ALKBH5 inhibited production of the HPV16 mRNAs spliced between 226 and 409 and promoted retention of the E6-encoding intron thereby favoring production of E6 mRNAs over E7 mRNAs.

Immunoprecipitation of HPV16 mRNAs by anti-m6A antibody.

Since ALKBH5 is an m6A “eraser”, we investigated if ALKBH5 overexpression reduced m6A-methylation of HPV16 mRNAs by immunoprecipitating m6A-methylated HPV16 mRNAs in the absence or presence of ALKBH5 overexpression. First we determined the specificity of the m6A antibody for m6A-methylated RNA. A cloning vector RNA sequence (Supplementary Fig. S1) was synthesized in vitro using nucleotide mixtures consisting of both

Fig. 3 **A** Linearized HPV16 genome (numbers refer to the HPV16 reference strain GeneBank: K02718.1). Early and late genes are indicated. P97: HPV16 early promoter. P670: HPV16 late promoter. Black oval: splice donor. White oval: splice acceptor. pAE: HPV16 early polyadenylation site. pAL: HPV16 late polyadenylation site. LCR: HPV16 long control region. **B** Schematic representation of HPV16 subgenomic plasmid pBELsLuc. Transcription from pBELsLuc is driven by the human cytomegalovirus immediate early promoter (CMV). Secreted luciferase (sLuc) gene was integrated in the L1 gene following the poliovirus 2A internal ribosomal entry site (IRES) sequence. **C** Schematic structures of a subset of HPV16 alternatively spliced mRNAs produced from pBELsLuc. Arrows indicate HPV16 RT-PCR primers. **D** Effect of METTL3, METTL14, WTAP, or all three combined on HPV16 E2 mRNA splicing was monitored by RT-PCR with indicated primer pair on RNA extracted from HeLa cells transfected with pBELsLuc and pMETTL3, pMETTL14, pWTAP or pMETTL3, pMETTL14 and pWTAP combined. **E** Effect of METTL3 on HPV16 E2 mRNA splicing was monitored by RT-PCR with indicated primer pair on RNA extracted from HeLa cells transfected with pBELsLuc and empty pUC plasmid (–) or pMETTL3 in triplicates. **F** Effect of METTL3 on HPV16 E2 mRNA splicing was monitored by RT-PCR with indicated primer pair on RNA extracted from HeLa cells transfected with pBELsLuc and pUC (–) or pMETTL3 in triplicates. The RT-PCR was optimized for detection also of the larger E1 mRNAs with retained E1-encoding intron (E1). Effect of YTHDC1, METTL3, or ALKBH5 on HPV16 E2- (**F**), E4- (**G**), or L1- (**H**) mRNA splicing was monitored by RT-PCR with indicated primer pair on RNA extracted from HeLa cells transfected with pBELsLuc and empty pUC plasmid (–), pYTHDC1, pMETTL3, or pALKBH5 in duplicates. **I** RNA immunoprecipitation of RNA extracted from C33A2 cells using IgG or anti-m6A antibody followed by RNA extraction and RT-PCR with indicated primers. Input: RT-PCR on RNA extracted from C33A2 cells in the absence of immunoprecipitation. **J** Western blot with anti-ALKBH5 or actin antibody on proteins extracted from C33A2 cells infected with lentivirus pLKO.1-vector or pLKO.1 with shRNA to ALKBH5 (shAKBH5-1). **K** Effect of ALKBH5 knock-down on HPV16 E2 mRNA splicing (880²⁷⁰⁹) was monitored by RT-PCR with RT-PCR primers 773S and E2qas on RNA extracted from C33A2 cells infected with pLKO.1-vector or pLKO.1 with shRNA to ALKBH5 (shALKBH5-1). The RT-PCR was optimized for detection also of the larger E1 mRNAs with retained E1-encoding intron (E1). **L** Effect of ALKBH5 knock-down on HPV16 L1/L1 mRNA splicing was monitored by RT-PCR with RT-PCR primers 773S and L1as on RNA extracted from C33A2 cells infected with pLKO.1-vector or pLKO.1 with shRNA to ALKBH5 (shALKBH5-1)

adenosine and m6A. RNA sequences with mixtures containing 0%, 1%, or 10%-m6A nucleotides were subjected to immunoprecipitation with Abcam anti-m6A antibody followed by RT-PCR. As can be seen, only RNA containing m6A was immunoprecipitated by the m6A antibody while none of the RNAs were immunoprecipitated by IgG under the experimental conditions used here (Fig. 2A). Next, RNA extracted from HeLa cells transfected with HPV16 reporter plasmid pC97ELsLuc and empty pUC plasmid (–) or pALKBH5 expression plasmid were subjected to immunoprecipitation with anti-m6A antibody and RT-PCR on RNA extracted after immunoprecipitation. The 773 s-E4as primer pair detects all HPV16 early mRNAs spliced between SD880 and SA3358 (Fig. 1C). The results revealed that the

anti-m6A antibody efficiently immunoprecipitated HPV16 mRNAs compared to IgG, demonstrating that the HPV16 mRNAs produced by pC97ELsLuc are m6A-methylated in the transfected cells (Fig. 2B). The results were confirmed by RT-qPCR (Fig. 2C). Next we immunoprecipitated RNA with anti-m6A antibody from cells transfected with pC97ELsLuc and empty pUC plasmid or ALKBH5 expressing plasmid and performed RT-qPCR. As can be seen, overexpression of ALKBH5 reduced the amount of HPV16 mRNAs immunoprecipitated by the m6A antibody (Fig. 2D). These results demonstrated that ALKBH5 overexpression reduced the levels of m6A-methylation on HPV16 mRNAs and that this reduction correlated with retention of the E6-encoding intron on the HPV16 mRNAs. Taken together, our results indicated that a reduction of m6A-methylation of HPV16 mRNAs inhibited HPV16 mRNA splicing and promoted intron retention, suggesting that increased m6A-methylation of HPV16 mRNAs would enhance splicing of the HPV16 E6/E7 mRNAs and promoted production of HPV16 E7 mRNAs.

Overexpression of the m6A “writer” METTL3 alters HPV16 alternative mRNA splicing

Next, we cotransfected HPV16 reporter plasmid pC97ELsLuc with plasmids that overexpressed the m6A-“writers” METTL3, METTL14, and WTAP (Supplementary Fig. S2A). None of these proteins appeared to significantly affect HPV16 E6/E7 mRNA splicing (Supplementary Fig. S2B) nor did they appear to affect splicing of HPV16 mRNAs between SD880 and SA3358 (data not shown). However, METTL3 overexpression exerted an inhibitory effect on HPV16 E2 mRNAs (880²⁷⁰⁹) produced from pC97ELsLuc (Supplementary Fig. S2C). This inhibitory effect of HPV16 E2 mRNA splicing was specific for METTL3 and was not observed with METTL14 or WTAP, the other two m6A writers analyzed here (Supplementary Fig. S2C). We decided to investigate the effect of METTL3 on E2 mRNA splicing further using HPV16 reporter plasmid pBELsLuc that lacks the E6/E7-coding region but encodes E1 and E2 (Fig. 3A, B). The pBELsLuc plasmid has the potential to produce HPV16 E1, E2, and E4 mRNAs as well as L2 and two alternatively spliced L1 mRNAs named L1i and L1 (Fig. 3C). Cotransfection of the m6A “writers” with HPV16 reporter plasmid pBELsLuc reproduced the effect on the E2 mRNAs previously observed with HPV16 reporter plasmid pC97ELsLuc: METTL3 inhibited HPV16 E2 mRNA splicing, whereas WTAP did not (Fig. 3D). However, METTL14 inhibited production of spliced E2 mRNA from pBELsLuc, but not from pC97ELsLuc (Fig. 3D). This may be explained by the fact that pBELsLuc produces relatively few alternatively spliced HPV16 mRNAs compared to pC97ELsLuc and therefore is likely to be a more

sensitive reporter plasmid for E1/E2 mRNA splicing than pC97ElsLuc. In addition, METTL14 does play a significant role in RNA recognition and stability of the methyltransferase complex. Since the effect of METTL14 overexpression was HPV16-plasmid specific, we did not pursue these results further. Performing RT-PCR under conditions that allowed detection also of the larger, HPV16 E1-intron encoding mRNAs in addition to the spliced E2 mRNAs confirmed that METTL3 inhibited production of the spliced E2 mRNAs (Fig. 3E) and revealed that METTL3 promoted production of the E1-intron containing mRNAs (Fig. 3E). Unexpectedly, overexpression of ALKBH5 also inhibited E2 mRNA splicing and promoted retention of the E1-encoding intron, but to a lesser extent than METTL3 (Fig. 3F). Splicing between SD880 and SA3358 appeared unaffected by both METTL3 and ALKBH5 (Fig. 3G). Taken together, these results supported the idea that m6A-methylation modulate splicing of HPV16 early E1/E2 and E6/E7 mRNAs.

METTL3 and ALKBH5 alter HPV16 late L1 mRNA splicing

We also monitored the effect of the m6A “writers” and “erasers” on the two alternatively spliced, HPV16 late L1 mRNAs here named L1 and L1i (Fig. 3C). Overexpression of METTL3 enhanced inclusion of the central exon on the L1 mRNAs and inhibited splicing directly from SD880 to SA5639, thereby promoting production of HPV16 L1 mRNAs over L1i mRNAs (Fig. 3H). Overexpression of ALKBH5 altered HPV16 L1 mRNA splicing by causing skipping of the central exon of the L1 mRNAs and enhancing direct splicing from SD880 to SA5639, thereby promoting production of HPV16 L1i mRNAs over L1 mRNAs (Fig. 3H). FTO overexpression did not affect splicing of the HPV16 L1 mRNAs (Supplementary Fig. S2D). To further support a role for METTL3 and ALKBH5 in the splicing of E1/2- and L1/L1i mRNAs, we attempted to knock-down these two proteins in the C33A2 cell line by using lentiviral vectors carrying shRNA to either METTL3 or ALKBH5. The C33A2 cell line carries the HPV16 reporter plasmid pBELsLuc stably integrated into its genome (Fig. 3B) and constitutively express the depicted HPV16 E1 and E2 mRNAs (Fig. 3C). We first confirmed that HPV16 mRNA are m6A-methylated in the C33A2 cell line by immunoprecipitation of the HPV16 mRNAs by anti-m6A antibody. The results revealed that anti-m6A antibody, but not IgG, immunoprecipitated HPV16 mRNAs extracted from C33A2 cells, indicating that they were m6A-methylated (Fig. 3I). We were unable to reproducibly knock-down METTL3, but ALKBH5 was efficiently knocked down after infection with a lentivirus carrying shRNA to ALKBH5 (Fig. 3J). Knock-down of ALKBH5 caused an increase in E2 mRNA splicing

(Fig. 3K) and enhanced inclusion of the central exon on the L1 mRNAs (Fig. 3L). As expected, knock-down of ALKBH5 had the opposite effect of ALKBH5 overexpression on the splicing of E1/E2 mRNAs (compare Fig. 3F and K) and L1/L1i mRNAs (compare Fig. 3H and L). We concluded that the m6A “eraser” ALKBH5 and the m6A “writer” METTL3 affected alternative splicing of HPV16 early and late mRNAs, further supporting the idea that m6A plays an important role in the control of HPV16 alternative mRNA splicing and gene expression.

Overexpression of the m6A “reader” YTHDC1 inhibits HPV16 E6/E7 mRNA splicing, thereby promoting production of full-length E6 mRNAs

Since overexpression of the m6A “writers” and “erasers” METTL3 and ALKBH5 affected alternative splicing of various HPV16 mRNAs, one may speculate that a nuclear “reader” of m6A-containing mRNAs might alter HPV16 mRNA splicing as well. The most well known nuclear m6A “reader” is YTHDC1. Plasmid pYTHDC1 produces flag-tagged YTHDC1 (Supplementary Fig. S2A) and was co-transfected with HPV16 reporter plasmid pC97ElsLuc (Fig. 1B). Overexpression of YTHDC1 altered HPV16 E6/E7 mRNA splicing (Fig. 4A), but did not affect HPV16 mRNAs spliced from SD880 to SA2709 (E2) (Fig. 3F), SD880 to SA3358 (E4) (Fig. 3G), or L1/L1i mRNA splicing (Fig. 3H). Quantitation of the effect of YTHDC1 on E6/E7 mRNA splicing in triplicate transfections (Fig. 4B) demonstrated that YTHDC1 primarily enhanced intron retention to promote production of E6-encoding mRNAs, and reduced production of spliced HPV16 mRNAs (226*409) (Fig. 4B). Next we investigated if YTHDC1 acted directly on the E6/E7-coding region by co-transfecting YTHDC1 in triplicates with the smaller HPV16 reporter plasmid pX856F [38] that encodes only E6 and E7 (Fig. 4C). Overexpression of YTHDC1 promoted retention of the E6-encoding exon at the expense of the E7 mRNA spliced from SD226 to SA409 (Fig. 4D, E). These results indicated that YTHDC1 acted directly on the E6/E7-coding region. To investigate if YTHDC1 binds directly to HPV16 mRNAs, we performed a CLIP assay on YTHDC1 on HeLa cells transfected with pC97ElsLuc, either in the presence of empty pUC plasmid (–) or pYTHDC1 that produced flag-tagged YTHDC1. The results revealed that anti-YTHDC1 antibody immunoprecipitated HPV16 mRNAs from pC97ElsLuc and pUC transfected cells (Fig. 4F) and that anti-flag antibody immunoprecipitated HPV16 mRNAs from cells cotransfected with pC97ElsLuc and flag-tagged pYTHDC1 (Fig. 4F). Thus both endogenous and overexpressed YTHDC1 interacted with HPV16 mRNAs in the transfected HeLa cells.

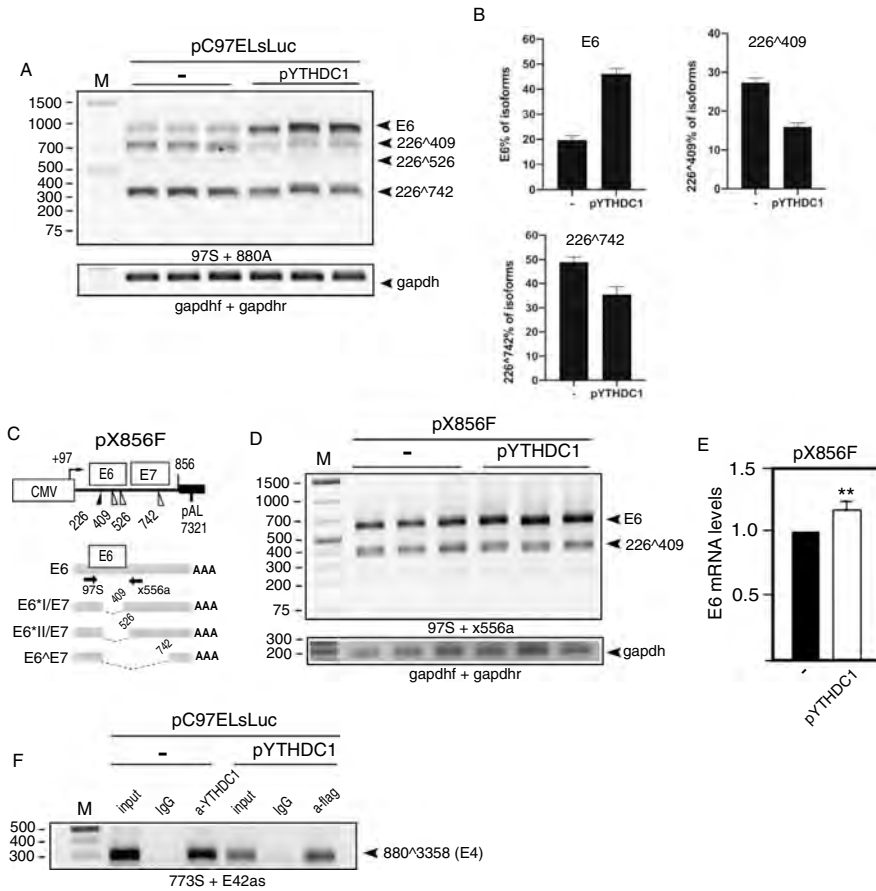


Fig. 4 **A** RT-PCR with primers 97S and 880A on RNA extracted from HeLa cells transfected in triplicates with pC97ELsLuc and empty pUC plasmid (–) or pYTHDC1 plasmid. **B** Densitometric quantitation of gel image in (A). The percentage of each indicated cDNA isoform of all cDNAs in a lane in the absence of YTHDC1 (–) or in the presence of YTHDC1 overexpression is displayed. **C** Schematic representation of the HPV16 E6- and E7-encoding reporter plasmid pX856F. The human cytomegalovirus promoter (CMV), the HPV16 E6 and E7 open reading frames, the HPV16 polyadenylation signal (pAL), and splice sites (226, 409, 526, and 742) are indicated. Alternatively spliced mRNAs produced by pX856F are

displayed. Arrows represent RT-PCR primers 97S and x556a. **D** RT-PCR with primers 97S and x556a on RNA extracted from HeLa cells transfected in triplicates with pX856F and empty pUC plasmid (–) or pYTHDC1 plasmid. **E** Densitometric quantitation of gel image in (D). **F** CLIP assay was performed on cell extracts from HeLa cells transfected with pC97ELsLuc and empty pUC plasmid (–) or pYTHDC1 that produces flag-tagged YTHDC1. Immunoprecipitation of RNA–protein complexes was performed with anti-YTHDC1 antibody (a-YTHDC1) or with anti-flag antibody (a-flag). RT-PCR was performed with HPV16 primers 773S and E42as. For location of RT-PCR primers see Fig. 3C

YTHDC1 inhibits splicing of HPV16 mRNAs produced from episomal HPV16 genome

Next, we analyzed the effect of YTHDC1 on the full-length HPV16 genome by using plasmid pHPV16AN (Fig. 5A). This full-length HPV16 genome is flanked by loxP sites and co-transfection with a cre-expressing plasmid releases

the episomal form of the HPV16 genome (Fig. 5A). The cre/lox reaction is controlled with PCR with primers 16S and 16A that discriminate between the transfected pHPV16AN plasmid DNA and the cre/loxed HPV16 episomal DNA (Fig. 5B). Analyses of the E6/E7 mRNAs produced by the episomal form of the HPV16 genome in the absence or presence of overexpressed YTHDC1 revealed that YTHDC1

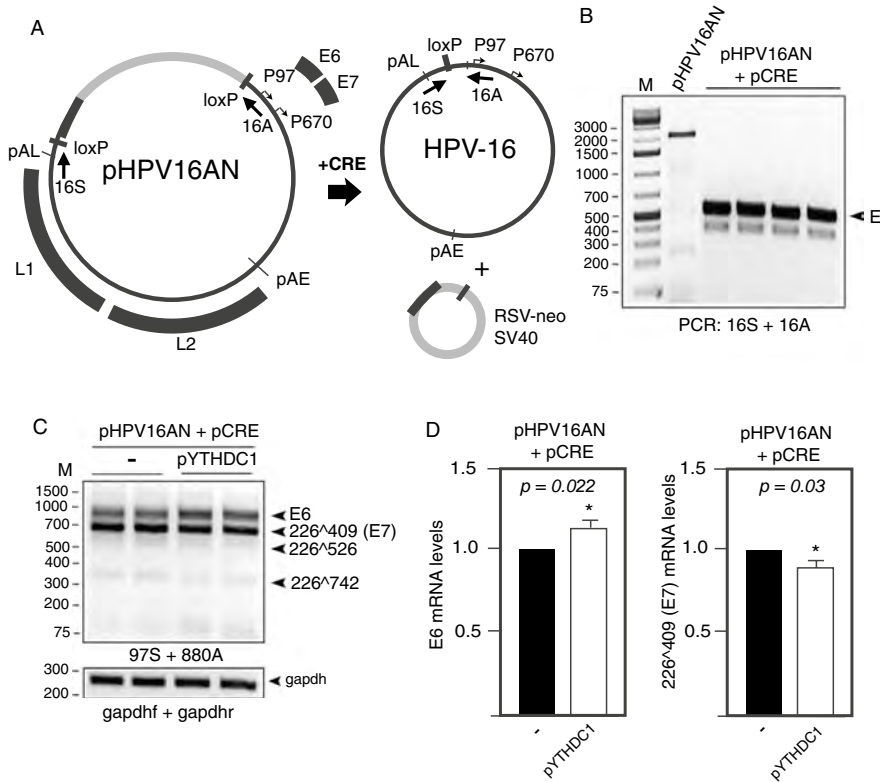


Fig. 5 **A** Schematic representation of the pHPV16AN plasmid that carries the full-length HPV16 genome flanked by loxP sites. Co-transfection of pHPV16AN and pCRE results in excision and circularization of the viral DNA at the loxP sites to generate episomal HPV16 DNA. Positions of PCR primers 16A and 16S used to monitor HPV16 genome excision and circularization are indicated. **B** Gel shows PCR with primers 16A and 16S on DNA harvested 48 h post-transfection. E, episomal HPV16 DNA. **C** RT-PCR on RNA extracted

from HeLa cells transfected with pHPV16AN and pCRE with empty pUC plasmid (–) or pYTHDC1. Transfections were performed in duplicates. RT-PCR was performed with the indicated RT-PCR primers. **D** Densitometric quantitation of gel images of RT-PCR products shown in (C). The RT-PCR band representing E6 mRNAs as well as the band representing spliced (226⁴⁰⁹) E7 mRNAs were quantified and fold difference in the absence or presence of YTHDC1 is shown in the graph. *p*-values are indicated

inhibited E6/E7 mRNA splicing and promoted retention of the E6-encoding intron at the expense of the spliced E7 mRNAs (226⁴⁰⁹) (Fig. 5C). However, quantitation revealed that the significance was relatively low (*p* < 0.05) (Fig. 5D). We concluded that YTHDC1 enhanced production of E6 mRNAs by inhibiting HPV16 mRNA splicing.

HPV16 mRNAs are m6A-methylated in HPV16-infected tonsillar cancer cell line HN26

Finally, we wished to determine if HPV16 mRNAs are m6A-methylated in HPV16-driven cancer cells. We therefore extracted RNA from the HPV16-positive tonsillar

cancer cell line HN26 [34]. This cell line was recently isolated 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 have an intact p53 gene and contain a complete episomal HPV16 DNA genome (Fig. 6A). A selected set of the alternatively spliced HPV16 mRNAs produced in this cell line is depicted (Fig. 6B). We subjected RNA extracted from the HN26 cells to immunoprecipitation with m6A-specific monoclonal antibody. RT-PCR on RNA extracted from these immunoprecipitations revealed that mRNAs encoding HPV16 E2, E4, E6, and E7 were immunoprecipitated by the m6A antibody, but not by IgG (Fig. 6C).

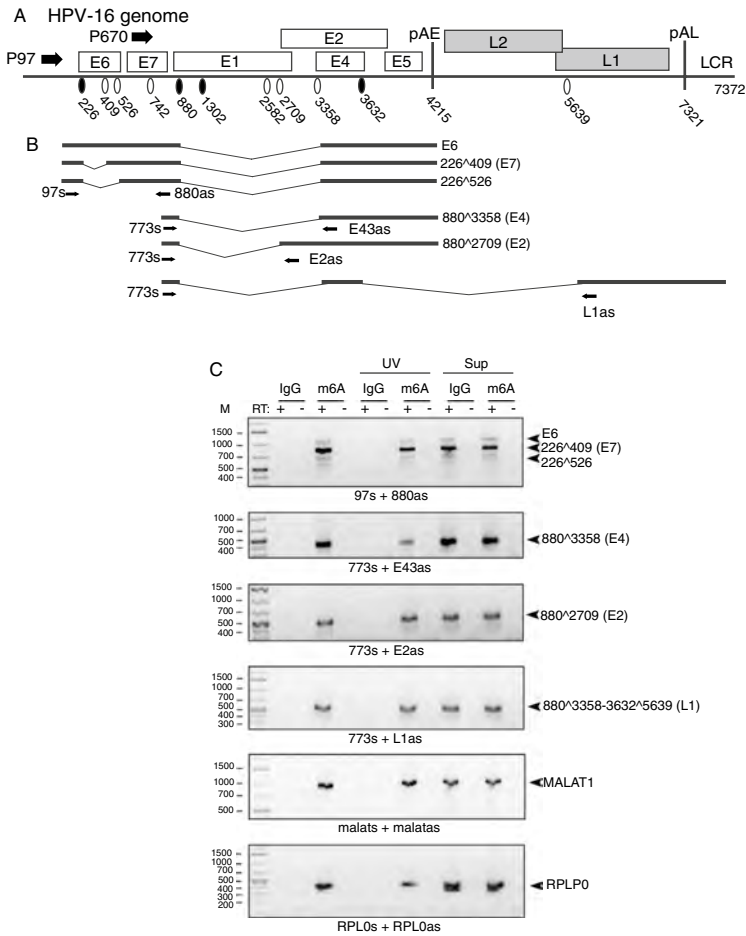


Fig. 6 **A** Full-length HPV16 genome (numbers refer to the HPV16 reference strain GeneBank: K02718.1). Early and late genes are indicated. P97: HPV16 early promoter. P670: HPV16 late promoter. Black oval: splice donor. White oval: splice acceptor. pAE: HPV16 early polyadenylation site. pAL: HPV16 late polyadenylation site. LCR: HPV16 long control region. **B** Schematic representations of a subset of HPV16 alternatively spliced mRNAs produced from the HPV16 genome. Arrows indicate positions of RT-PCR primers. **C** RNA was extracted from HN26 cells and subjected to immunoprecipitation with either IgG or anti-m6A antibody (m6A) followed by RT-PCR on RNA extracted from the immunoprecipitated antibody–RNA

complexes. HPV16 RT-PCR primers are indicated. RT-PCR primers used for amplification of HPV16 mRNAs, cellular RPLP0 mRNAs, and cellular MALAT RNAs are listed in Supplementary Table 1. The spliced HPV16 mRNAs represented by the various DNA fragments are indicated to the right. UV, immunoprecipitation samples treated with UV light prior to washing of the antibody–RNA complexes. Sup, RT-PCR on the supernatants obtained after immunoprecipitation of the HPV16 mRNAs. + reverse transcription in the presence of reverse transcriptase (RT); – reverse transcription in the absence of RT

Antibody–RNA complexes were also stabilized by UV cross linking prior to addition of protein G beads, with similar results (Fig. 6C). The hypermethylated MALAT1 lncRNA was also readily detected by the m6A antibody (Fig. 6C) as was the mRNA of the RPLP0 house keeping

gene (Fig. 6C). Taken together, our results confirmed that HPV16 mRNAs are m6A-methylated in cancer cells which supported the idea that m6A-methylation plays an important role in the control of HPV16 gene expression.

Discussion

Adenosines are more likely to be m6A-methylated if located in the so called DRACH motif, i.e., (A/G)AC(ACU). However, RNA-sequencing data have identified m6A modifications on other adenosines as well and has allowed the development of prediction servers for m6A sites, such as SRAMP [39]. We analyzed the HPV16 genome for the presence of potential m6A sites using SRAMP. Unlike the general pattern of m6A on cellular mRNAs, which clusters around start codons and in the 3'-UTR [19, 40], m6A on HPV16 mRNAs are predicted to cluster around HPV16 splice sites (Supplementary Fig. S3). The vast majority of all predicted m6A-methylated adenosine residues clustered downstream of the major HPV16 E6/E7 splice site SA409 and downstream of the HPV16 E2 splice site SA2709 (Supplementary Fig. S3). A few sites were predicted to be present in the small HPV16 exon located between splice sites SA3358 and SD3632 as well (Supplementary Fig. S3). Three predicted m6A sites immediately downstream of SA3358 coincide with a previously identified splicing enhancer, suggesting that m6A may modulate SA3358 through this splicing enhancer [41]. Interestingly, these are the HPV16 splice sites that were affected by the overexpression of ALKBH5, METTL3, and/or YTHDC1. In particular, overexpression of the ALKBH5 “eraser” inhibited splicing of the E6/E7 mRNAs and so did overexpression of the m6A “reader” YTHDC1. To reconcile these results, one may speculate that ALKBH5 removed m6A-adenosines thereby reducing the ability of a yet undisclosed cellular “reader” and splicing factor to bind to the mRNAs and enhance splicing, while YTHDC1 may bind to m6A sites on the same mRNAs and prevent the said cellular splicing factor from binding. In addition to these observations, METTL3 overexpression inhibited splicing to SA2709 and the production of E2 mRNAs, while at the same time promoting retention of the E1-encoding intron and production of the E1 mRNAs. These results were obtained also with the HPV16 pBEL reporter plasmid on which the E6/E7 region is absent, suggesting that m6A modifications in the E1/E2 region controlled splicing in the E1/E2 region, i.e., HPV16 3'-splice site SA2709. However, it remains to be determined if these predicted m6A sites are indeed methylated in cells and how this methylation potentially affects binding of cellular splicing factors to the HPV16 mRNAs.

The combined functions of the high-risk papillomavirus E6 and E7 proteins are required for the establishment of a productive viral infection and for transformation of HPV16-infected cells to cancer cells or for efficient immortalization of primary keratinocytes *in vitro* [42, 43]. E6 and E7 are both expressed in cancers that are caused by

HPV16, i.e., various anogenital cancers, primarily cervical cancer, and head and neck cancer primarily tonsillar cancer. While E7 targets the cell cycle, E6 prevents apoptotic cell death. Therefore, a balanced expression of E6 and E7 is pivotal. The E7 mRNA (226⁴⁰⁹) is believed to be the major E7 producing mRNA while the unspliced mRNA mainly produces E6 [44, 45]. It is therefore of paramount importance for HPV16 to maintain a well-balanced splicing efficiency to allow for production of sufficient quantities of both E6 and E7. Control of E6/E7 mRNA splicing is crucial for HPV16 replication but also for induction and maintenance of cancers driven by HPV16. It has been shown that EGF-signaling can regulate the balance between unspliced E6 and spliced E7 mRNAs (226⁴⁰⁹) in HPV16-infected cells, with EGF-signaling promoting production of unspliced E6 mRNAs [46]. Recently, METTL3 was shown to have a positive effect on EGFR levels in cancer cells through translational regulation [47]. On the other hand, our results demonstrated that overexpression of the demethylase ALKBH5 increased the levels of unspliced E6 mRNA whereas METTL3 did not, which suggested that a reduction of m6A modifications inhibited E6/E7 mRNA splicing. It remains to be determined how METTL3 and ALKBH5 are involved in the control of HPV16 E6/E7 mRNA splicing in HPV16-infected cells.

Unexpectedly, overexpression of the reader YTHDC1 had a similar effect on HPV16 E6/E7 splicing as the eraser ALKBH5—it caused retention of the E6-encoding intron and promoted production of E6 mRNAs. Our results demonstrated that YTHDC1 physically associated with HPV16 mRNAs by the use of CLIP assay. These results may therefore suggest that YTHDC1 either interacts preferentially with m6A-containing mRNAs to prevent factors with a splicing enhancing function to bind to HPV16 mRNAs that YTHDC1 interacts with the HPV16 E6/E7 mRNAs and inhibit adjacent splicing factors or that YTHDC1 recruits factors with splicing inhibitory function. It appears that YTHDC1 can regulate splicing by recruiting various splicing factors. This has been described for a subset of cellular mRNAs to which YTHDC1 recruited SRSF3 to cause exon inclusion while blocking access to SRSF10 that promoted exon exclusion [24]. YTHDC1 was found to interact with SRSF3 also during mouse oocyte development together with SRSF7 and polyadenylation factor CPSF3, affecting both alternative splicing and alternative polyadenylation [48]. Interestingly, SRSF3 has been shown to control HPV16 mRNA splicing [49]. These results indicate that YTHDC1 can affect splicing in discrete ways. It remains to be determined how YTHDC1 promotes intron retention of HPV16 E6/E7 mRNAs.

The binding to RNA of other RNA-binding proteins than YTHDC1 may also be affected by m6A-methylation, for example members of the hnRNP-family of splicing factors.

Some hnRNPs may be direct readers of m6A [19], e.g., hnRNPA2B1 that has been functionally linked to m6A-mediated splicing [25]. Furthermore, the binding of HuR [18], hnRNP G, and hnRNP C [20, 21] to regulatory RNA elements is altered in response to distortions of RNA- secondary structures by the presence of m6A modifications. All three proteins have previously been shown to bind RNA elements on HPV16 mRNAs and regulate HPV16 mRNA splicing [37, 38, 50, 51] and may therefore potentially be affected by m6A-methylation of HPV16 mRNAs. Recently, a splicing silencer RNA element located in the HPV16 E7-coding region has been identified. It acts by reducing production of the spliced E7 mRNA (226⁴⁰⁹) [38]. A similar sequence is present in HPV18 [49]. These RNA elements interacted with hnRNP A1, but it is currently unknown if the interactions between hnRNP A1 and the HPV16 RNA silencers are affected by m6A-methylation or by YTHDC1. Taken together, our results suggest that m6A-methylation of HPV16 mRNAs may function as an additional layer in the control of HPV16 mRNA splicing.

Limitations of the study

These experiments do not allow us to conclude whether the effect on HPV16 mRNA splicing is a direct effect due to m6A-methylation of the HPV16 mRNAs or an indirect effect of m6A-methylation of cellular transcripts, e.g., those encoding splicing regulatory factors. In this respect, it may be of interest to note that high confidence m6A sites are present on mRNAs encoding SRSF2, a protein that may control HPV16 mRNA splicing. Since we have not mapped the exact sites of m6A-methylation on the HPV16 mRNAs, we do not know if HPV16 mRNAs produced from integrated HPV16 genomes found in the majority of the cancers caused by HPV16 are m6A-methylated.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11262-022-01889-6>.

Acknowledgements This work was supported by the Swedish Research Council-Medicine (Vetenskapsrådet) [Grant No. VR2019-01210 to S.S.]; the Swedish Cancer Society (Cancerfonden) [Grant No. CAN2018/702 to S.S.]; and the China Scholarship Council [Grant No. 201809120016 to X.C.].

Author contributions XC, KN, NK, SS: Conceptualisation, XC, KN: Methodology, XC, KN: Formal analysis and investigation, XC, KN, SS: Writing—original draft preparation, XC, KN, NK, SS: Writing—reviewing and editing, XC, SS: Funding acquisition, XC, KN, NK: Resources, NK, SS: Supervision.

Funding Open access funding provided by Uppsala University.

Declarations

Conflict of interest The authors declare no conflicts of interest.

Research involving human and animal participants Not applicable.

Informed consent Not applicable.

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

References

- Walboomers JM, Jacobs MV, Manos MM, Bosch FX, Kummer JA, Shah KV, Snijders PJ, Peto J, Meijer CJ, Munoz N (1999) Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J Pathol* 189:12–19
- zur Hausen H (2002) Papillomaviruses and cancer: from basic studies to clinical application. *Nat Rev Cancer* 2:342–350
- Bouvard V, Baan R, Straif K, Grosse Y, Secretan B, El Ghissassi F, Benbrahim-Tallaa L, Guha N, Freeman C, Galichet L, Coglianò V, on WIAfR, Group CMW (2009) A review of human carcinogens—Part B: biological agents. *Lancet Oncol* 10:321–322
- Schiffman M, Doorbar J, Wentzensen N, de Sanjose S, Fakhry C, Monk BJ, Stanley MA, Franceschi S (2016) Carcinogenic human papillomavirus infection. *Nat Rev Dis Primers* 2:16086
- Chow LT, Broker TR, Steinberg BM (2010) The natural history of human papillomavirus infections of the mucosal epithelia. *APMIS* 118:422–449
- Van Doorslaer K, Tan Q, Xirasagar S, Bandaru S, Gopalan V, Mohamoud Y, Huyen Y, McBride AA (2013) The papillomavirus episteme: a central resource for papillomavirus sequence data and analysis. *Nucleic Acids Res* 41:D571–D578
- Thierry F (2009) Transcriptional regulation of the papillomavirus oncogenes by cellular and viral transcription factors in cervical carcinoma. *Virology* 384:375–379
- McBride AA (2013) The papillomavirus E2 proteins. *Virology* 445:57–79
- Johansson C, Somberg M, Li X, Backstrom Winquist E, Fay J, Ryan F, Pim D, Banks L, Schwartz S (2012) HPV-16 E2 contributes to induction of HPV-16 late gene expression by inhibiting early polyadenylation. *EMBO J* 31:3212–3227
- Johansson C, Schwartz S (2013) Regulation of human papillomavirus gene expression by splicing and polyadenylation. *Nat Rev Microbiol* 11:239–251
- Kajitani N, Schwartz S (2020) Role of viral ribonucleoproteins in human papillomavirus Type 16 gene expression. *Viruses*. <https://doi.org/10.3390/v12101110>
- Graham SV, Faizo AA (2017) Control of human papillomavirus gene expression by alternative splicing. *Virus Res* 231:83–95

13. Jia R, Zheng ZM (2009) Regulation of bovine papillomavirus type 1 gene expression by RNA processing. *Front Biosci* 14:1270–1282
14. Fu Y, Dominissini D, Rechavi G, He C (2014) Gene expression regulation mediated through reversible m(6)A RNA methylation. *Nat Rev Genet* 15:293–306
15. Liu J, Yue Y, Han D, Wang X, Fu Y, Zhang L, Jia G, Yu M, Lu Z, Deng X, Dai Q, Chen W, He C (2014) A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nat Chem Biol* 10:93–95
16. Jia G, Fu Y, Zhao X, Dai Q, Zheng G, Yang Y, Yi C, Lindahl T, Pan T, Yang YG, He C (2011) N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat Chem Biol* 7:885–887
17. Zheng G, Dahl JA, Niu Y, Fedorcsak P, Huang CM, Li CJ, Vagbo CB, Shi Y, Wang WL, Song SH, Lu Z, Bosmans RP, Dai Q, Hao YJ, Yang X, Zhao WM, Tong WM, Wang XJ, Bogdan F, Furu K, Fu Y, Jia G, Zhao X, Liu J, Krokkan HE, Klungland A, Yang YG, He C (2013) ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol Cell* 49:18–29
18. Wang X, He C (2014) Reading RNA methylation codes through methyl-specific binding proteins. *RNA Biol* 11:669–672
19. Dominissini D, Moshitch-Moshkovitz S, Schwartz S, Salmon-Divon M, Ungar L, Osenberg S, Cesarkas K, Jacob-Hirsch J, Amariglio N, Kupiec M, Sorek R, Rechavi G (2012) Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* 485:201–206
20. Liu N, Dai Q, Zheng G, He C, Parisien M, Pan T (2015) N(6)-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature* 518:560–564
21. Liu N, Zhou KI, Parisien M, Dai Q, Diatchenko L, Pan T (2017) N6-methyladenosine alters RNA structure to regulate binding of a low-complexity protein. *Nucleic Acids Res* 45:6051–6063
22. Wang X, Lu Z, Gomez A, Hon GC, Yue Y, Han D, Fu Y, Parisien M, Dai Q, Jia G, Ren B, Pan T, He C (2014) N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* 505:117–120
23. Zhu T, Roundtree IA, Wang P, Wang X, Wang L, Sun C, Tian Y, Li J, He C, Xu Y (2014) Crystal structure of the YTH domain of YTHDF2 reveals mechanism for recognition of N6-methyladenosine. *Cell Res* 24:1493–1496
24. Xiao W, Adhikari S, Dahal U, Chen YS, Hao YJ, Sun BF, Sun HY, Li A, Ping XL, Lai WY, Wang X, Ma HL, Huang CM, Yang Y, Huang N, Jiang GB, Wang HL, Zhou Q, Wang XJ, Zhao YL, Yang YG (2016) Nuclear m(6)A reader YTHDC1 regulates mRNA splicing. *Mol Cell* 61:507–519
25. Alarcon CR, Goodarzi H, Lee H, Liu X, Tavazoie S, Tavazoie SF (2015) HNRNPA2B1 Is a mediator of m(6)A-dependent nuclear RNA processing events. *Cell* 162:1299–1308
26. Lichinchi G, Zhao BS, Wu Y, Lu Z, Qin Y, He C, Rana TM (2016) Dynamics of human and viral RNA methylation during Zika virus infection. *Cell Host Microbe* 20:666–673
27. Ye F (2017) RNA N(6)-adenosine methylation (m(6)A) steers epitranscriptomic control of herpesvirus replication. *Inflamm Cell Signal*. <https://doi.org/10.14800/ics.1604>
28. Tirumuru N, Zhao BS, Lu W, Lu Z, He C, Wu L (2016) N(6)-methyladenosine of HIV-1 RNA regulates viral infection and HIV-1 Gag protein expression. *Elife*. <https://doi.org/10.7554/eLife.15528>
29. Lichinchi G, Gao S, Saletore Y, Gonzalez GM, Bansal V, Wang Y, Mason CE, Rana TM (2016) Dynamics of the human and viral m(6)A RNA methylomes during HIV-1 infection of T cells. *Nat Microbiol* 1:16011
30. Tsai K, Courtney DG, Cullen BR (2018) Addition of m6A to SV40 late mRNAs enhances viral structural gene expression and replication. *PLoS Pathog* 14:e1006919
31. Price AM, Hayer KE, McIntyre ABR, Gokhale NS, Abebe JS, Della Fera AN, Mason CE, Horner SM, Wilson AC, Depledge DP, Weitzman MD (2020) Direct RNA sequencing reveals m(6)A modifications on adenovirus RNA are necessary for efficient splicing. *Nat Commun* 11:6016
32. Zhao J, Lee EE, Kim J, Yang R, Chamseddin B, Ni C, Gusho E, Xie Y, Chiang CM, Buszczak M, Zhan X, Laimins L, Wang RC (2019) Transforming activity of an oncoprotein-encoding circular RNA from human papillomavirus. *Nat Commun* 10:2300
33. Nilsson K, Wu C, Kajitani N, Yu H, Tsimtsirakis E, Gong L, Winquist EB, Glahder J, Ekblad L, Wennerberg J, Schwartz S (2018) The DNA damage response activates HPV16 late gene expression at the level of RNA processing. *Nucleic Acids Res* 46:5029–5049
34. Forslund O, Sugiyama N, Wu C, Ravi N, Jin Y, Swoboda S, Andersson F, Bzhalava D, Hultin E, Paulsson K, Dillner J, Schwartz S, Wennerberg J, Ekblad L (2019) A novel human in vitro papillomavirus type 16 positive tonsil cancer cell line with high sensitivity to radiation and cisplatin. *BMC Cancer* 19:265
35. Wu C, Nilsson K, Zheng Y, Ekenstierna C, Sugiyama N, Forslund O, Kajitani N, Yu H, Wennerberg J, Ekblad L, Schwartz S (2019) Short half-life of HPV16 E6 and E7 mRNAs sensitizes HPV16-positive tonsillar cancer cell line HN26 to DNA-damaging drugs. *Int J Cancer* 144:297–310
36. Li X, Johansson C, Cardoso-Palacios C, Mossberg A, Dhanjal S, Bergvall M, Schwartz S (2013) Eight nucleotide substitutions inhibit splicing to HPV-16 3'-splice site SA3358 and reduce the efficiency by which HPV-16 increases the life span of primary human keratinocytes. *PLoS ONE* 8:e72776
37. Li X, Johansson C, Glahder J, Mossberg AK, Schwartz S (2013) Suppression of HPV-16 late L1 5'-splice site SD3632 by binding of hnRNP D proteins and hnRNP A2/B1 to upstream AUAGUA RNA motifs. *Nucleic Acids Res* 22:10488–10508
38. Zheng Y, Jonsson J, Hao C, Shoja Chaghervand S, Cui X, Kajitani N, Gong L, Wu C, Schwartz S (2020) Heterogeneous nuclear ribonucleoprotein A1 (hnRNP A1) and hnRNP A2 inhibit splicing to human papillomavirus 16 splice site SA409 through a UAG-containing sequence in the E7 coding region. *J Virol*. <https://doi.org/10.1128/JVI.01509-20>
39. Zhou Y, Zeng P, Li YH, Zhang Z, Cui Q (2016) SRAMP: prediction of mammalian N6-methyladenosine (m6A) sites based on sequence-derived features. *Nucleic Acids Res* 44:e91
40. Meyer KD, Saletore Y, Zumbo P, Elemento O, Mason CE, Jaffrey SR (2012) Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* 149:1635–1646
41. Rush M, Zhao X, Schwartz S (2005) A splicing enhancer in the E4 coding region of human papillomavirus type 16 is required for early mRNA splicing and polyadenylation as well as inhibition of premature late gene expression. *J Virol* 79:12002–12015
42. Vande Pol SB, Klingelutz AJ (2013) Papillomavirus E6 oncoproteins. *Virology* 445:115–137
43. Roman A, Munger K (2013) The papillomavirus E7 proteins. *Virology* 445:138–168
44. Tang S, Tao M, McCoy JP Jr, Zheng ZM (2006) The E7 oncoprotein is translated from spliced E6*1 transcripts in high-risk human papillomavirus type 16- or type 18-positive cervical cancer cell lines via translation reinitiation. *J Virol* 80:4249–4263
45. Stacey SN, Jordan D, Williamson AJ, Brown M, Coote JH, Arrand JR (2000) Leaky scanning is the predominant mechanism for translation of human papillomavirus type 16 E7 oncoprotein from E6/E7 bicistronic mRNA. *J Virol* 74:7284–7297
46. Rosenberger S, De-Castro Arce J, Langbein L, Steenbergen RD, Rosl F (2010) Alternative splicing of human papillomavirus type-16 E6/E6* early mRNA is coupled to EGF signaling via Erk1/2 activation. *Proc Natl Acad Sci USA* 107:7006–7011
47. Lin S, Choe J, Du P, Triboulet R, Gregory RI (2016) The m(6)A methyltransferase METTL3 promotes translation in human cancer cells. *Mol Cell* 62:335–345
48. Kasowitz SD, Ma J, Anderson SJ, Leu NA, Xu Y, Gregory BD, Schultz RM, Wang PJ (2018) Nuclear m6A reader YTHDC1

- regulates alternative polyadenylation and splicing during mouse oocyte development. *PLoS Genet* 14:e1007412
49. Ajiro M, Tang S, Doorbar J, Zheng ZM (2016) Serine/arginine-rich splicing factor 3 and heterogeneous nuclear ribonucleoprotein A1 regulate alternative RNA splicing and gene expression of human papillomavirus 18 through two functionally distinguishable cis elements. *J Virol* 90:9138–9152
 50. Yu H, Gong L, Wu C, Nilsson K, Li-Wang X, Schwartz S (2018) hnRNP G prevents inclusion on the HPV16 L1 mRNAs of the central exon between splice sites SA3358 and SD3632. *J Gen Virol* 99:328–343
 51. Dhanjal S, Kajitani N, Glahder J, Mossberg AK, Johansson C, Schwartz S (2015) Heterogeneous nuclear ribonucleoprotein C proteins interact with the human papillomavirus type 16 (HPV16) early 3'-untranslated region and alleviate suppression of HPV16 late L1 mRNA splicing. *J Biol Chem* 290:13354–13371

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

SUPPLEMENTARY FIGURE LEGENDS

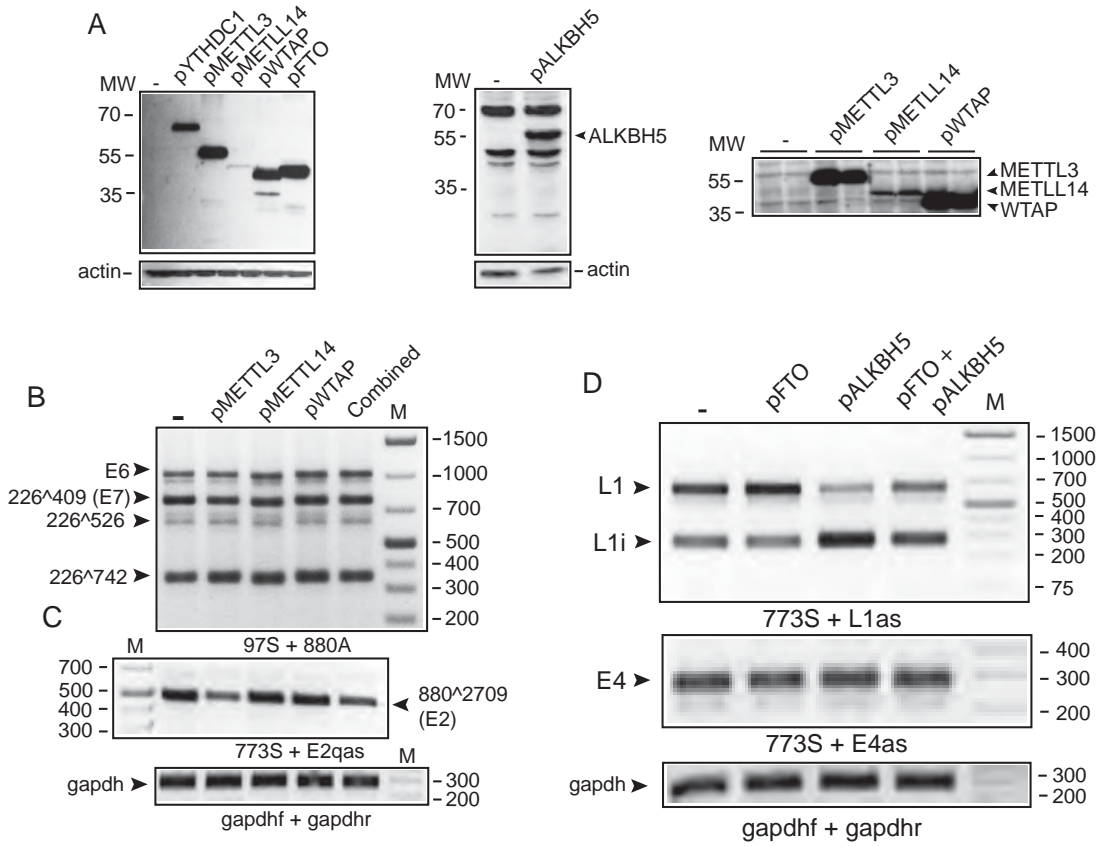
Supplementary Figure S1 Sequence of in vitro synthesized, m6A-containing RNA.

Supplementary Figure S2 (A) Left panel: Western blot with anti-Flag antibody on proteins harvested from HeLa cells transfected with empty pUC plasmid (-), pYTHDC1, pMETTL3, pMETTL14, pWTAP or pFTO. Middle panel: Western blot with anti-ALKBH5 antibody on proteins harvested from HeLa cells transfected with empty pUC plasmid (-) or pALKBH5. Right panel: Long exposure of Western blot with anti-Flag antibody on proteins harvested from HeLa cells transfected in duplicates with empty pUC plasmid (-), pMETTL3, pMETTL14 or pWTAP. **(B)** Effect of METTL3, METTL14, WTAP or all three combined on HPV16 E6/E7 mRNA splicing was monitored by RT-PCR with indicated primer pair on RNA extracted from HeLa cells transfected with pC97ELsLuc and empty pUC plasmid (-), pMETTL3, pMETTL14, pWTAP or pMETTL3, pMETTL14 and pWTAP combined. **(C)** Effect of METTL3, METTL14, WTAP or all three combined on HPV16 E2 mRNA splicing was monitored by RT-PCR with indicated primer pair on RNA extracted from HeLa cells transfected with pC97ELsLuc and empty pUC plasmid (-), pMETTL3, pMETTL14, pWTAP or pMETTL3, pMETTL14 and pWTAP combined. **(D)** Effect of FTO and ALKBH5 overexpression on HPV16 L1- and E4- mRNA splicing was monitored by RT-PCR with indicated primer pairs on RNA extracted from pBELsLuc with empty pUC plasmid (-), pFTO or pALKBH5.

Supplementary Figure S3 (A) m6A site prediction on the HPV16 genome performed using the online software SRAMP (<http://www.cuilab.cn/sramp/>) (Zhou). Selected HPV16 splice sites SD226, SA409, SD880, SA2709, SA3358 and SD3632 are indicated. Numbers refer to the HPV16 reference strain (GeneBank: K02718.1). HPV16 early and late polyadenylation signals pAE and pAL, respectively, are indicated. **(B)** A subset of alternatively spliced HPV16 mRNAs are depicted.

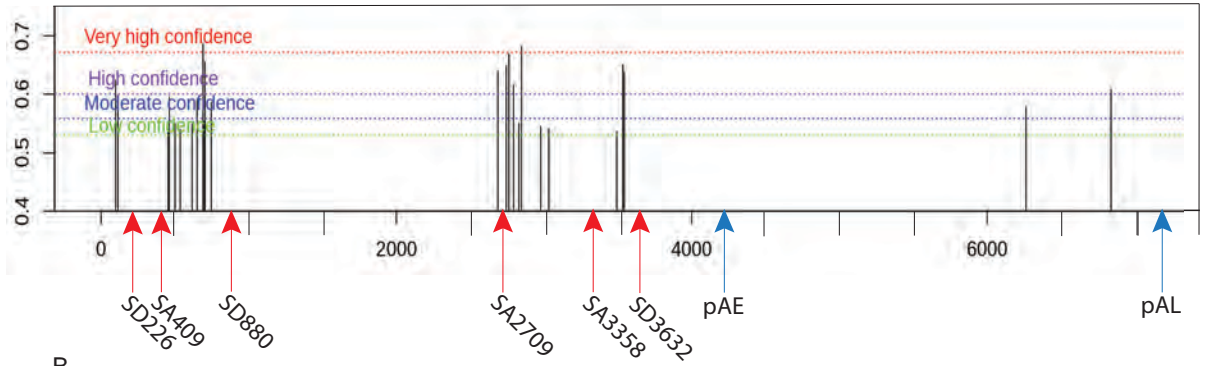
Supplementary Figure S1. *In Vitro* transcribed cloning vector RNA sequence.

AGGATTTTCGAGCGTGGGTCACTGAGCTCAGTGGACGAAAACCTCACGTTAAGGGATTTT
GGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAATAAATGAAGTTTT
AAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGT
GAGGCACCTATCTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTTGCCTGACTCCCCGT
CGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCCTGCAATGATA
CCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAAACCAGCCAGCCGGAA
GGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGT
TGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCAT
TGCTACAGGCATCGTGGTGTACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTT
CCCAACGATCAAGGCGAGTTACATGATCCCCATGTTGTGCAAAAAAGCGGTTAGCTCT
TTCGGTCTCCGATCGTTGTCAGAAGTAAGTTGGCCGAGTGTATCACTCATGGTTAT
GGCAGCACTGCATAATTCTTACTGTGATGCCATCCGTAAGATGCTTTTCTGTGACTG
GTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGC
CCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCAT
TGGAAAACGTTCTTCGGGACGAAAACCTCAAGGATCTTACCGCTGTTGAGATCCAGTT
CGATGTAACCCACTCGTGCACCCAACCTGATCTTCAGCATCTTTTACTTTACCCAGCGTTT
CTGGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACA
CGAAAATGTTGAATACTCATACTCTTCCTTTTCAATATTATTGAAGCATTATCAGGGTT
ATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAACAAATAGGGGTTCC
GCCGCAATTTTCCCCAAAATGCCACCTGAACTTCAAGAAAC

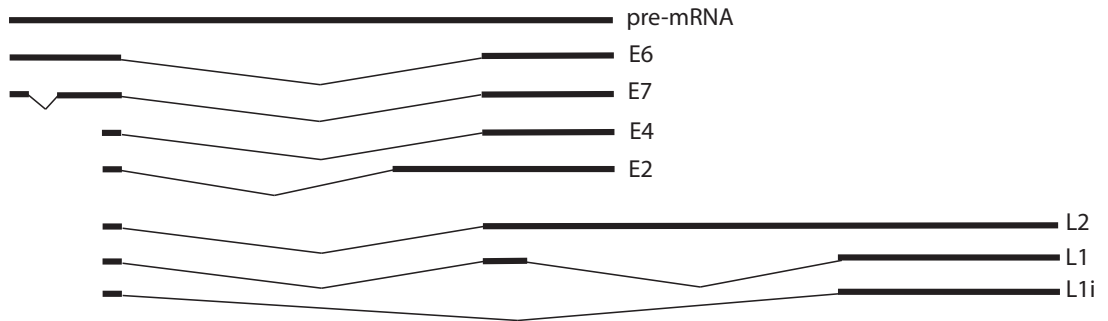


A

m6A predictions in the HPV16 genome



B



Supplementary Table 1. Primer list.

RT-PCR primer name	Amplified sequences	Sequence
97s	HPV16	GTCGACCTGCAATGTTTCAGGACCC
880as	HPV16	GAAACCATAATCTACCATGGCTGATC
X556A	HPV16	GCTCGAGCAGCTGGGTTTCTCTACGTGT
773s	HPV16	GCACACACGTAGACATTCGTACTIONG
E4as	HPV16	TGCTGCCTAATAGTTTCAGGAGAGG
E43as	HPV16	GTGGTGTGGCAGGGGTTTCCGGTGTG
E42as	HPV16	CGGTGCCCAAGGCGACGGCTTTGG
L1as	HPV16	GCAACATATTCATCCGTGCTTACAACC
E2as	HPV16	CAGCCAGCGTTGGCACCACCT
E2qas	HPV16	CAGCCAGCGTTGGCACCACCT
16S	HPV16	TATGTATGGTATAATAAACACGTGTGTATGTG
16A	HPV16	GCAGTGCAGGTCAGGAAAACAGGGATTTGGC
MALAT1s	MALAT1	CGTAGACCAGAACCAATTTAGAAG
MALAT1as	MALAT1	CATATTGCCGACCTCACGGAT
MALAT1asQ	MALAT1	AGCACCTGGGTCAGCTGTCAAT
RPL0s	RPLP0	ACCTGGAAGTCCAACACTTCCTTA
RPL0as	RPLP0	GATCTCAGTGAGGTCCTCCTTG
gapdhf	GAPDH	ACCCAGAAGACTGTGGATGG
gapdhr	GAPDH	TTCTAGACGGCAGGTCAGGT
T7s	<i>In vitro</i> RNA	ACGTTAAGGGATTTTGGTCATGAGA
T7as	<i>In vitro</i> RNA	TCAAATATGTATCCGCTCATGAGA

Paper III





Efficient production of HPV16 E2 protein from HPV16 late mRNAs spliced from SD880 to SA2709



Yunji Zheng^a, Xiaoxu Cui^a, Kersti Nilsson^a, Haoran Yu^a, Lijing Gong^{a,b}, Chengjun Wu^{a,c,d,*}, Stefan Schwartz^{a,*}

^a Department of Laboratory Medicine, Lund University, BMC-B13, 221 84, Lund, Sweden

^b China Institute of Sport and Health Sciences, Beijing Sport University, Xinxu Road 48, Haidian District, 100084, Beijing, China

^c The First Affiliated Hospital of Shandong First Medical University, No.16766 Jingshi Road, Jinan, Shandong Province, 250014, China

^d Institute of Basic Medicine, Shandong First Medical University & Shandong Academy of Medical Sciences, No.18877 Jingshi Road, Jinan, Shandong Province, 250062, China

ARTICLE INFO

Keywords:
HPV16
E2
Splicing
Translation

ABSTRACT

Human papillomaviruses (HPVs) produce a large number of alternatively spliced mRNAs, including a number of differently spliced mRNAs with the potential to produce E2 protein. To identify the alternatively spliced HPV16 mRNA with the highest ability to produce E2 protein, we have generated E2 cDNA expression plasmids representing the most common, alternatively spliced E2 mRNAs, and assessed their translational potential. Our results revealed that an mRNA initiated at the HPV16 late promoter p670 and spliced from the HPV16 5'-splice site SD880 to the HPV16 3'-splice site SA2709, located immediately upstream of the E2 ATG, produced higher levels of E2 than any of the other alternatively spliced, E2-encoding mRNAs. Utilization of a known, alternative 3'-splice site located upstream of the E2 ATG named SA2582, generated mRNAs with lower ability to produce E2 than mRNAs spliced to SA2709. Finally, analysis of HPV16 mRNA splicing demonstrated that SA2709 was more efficiently spliced to the upstream 5'-splice site SD880 than to the upstream 5'-splice site SD226. In conclusion, the HPV16 mRNA with the greatest ability to produce E2 protein is generated from the HPV16 late promoter and is spliced between HPV16 5'-splice site SD880 and HPV16 3'-splice site SA2709.

1. Introduction

Human papillomavirus (HPV) is a large family of small DNA viruses with epithelial tropism. Some of these HPV types prefer mucosal cells and at least 12 of those have been deemed carcinogenic by the International Agency for Research on Cancer (Bouvard et al., 2009). HPV16 is the most common cancer-associated HPV type and is associated with a number of anogenital cancers (Bouvard et al., 2009; Chow et al., 2010; Schiffman et al., 2016; Walboomers et al., 1999) primarily cervical cancer in women, as well as oropharyngeal cancer, primarily tonsillar cancer.

The HPV16 life cycle is intimately linked to cellular differentiation and commences with an early stage with low level replication of the HPV16 episomal DNA genome. This stage is followed by a “pre-late” stage characterized by highly efficient HPV16 DNA replication orchestrated by the HPV16 early proteins E1 and E2 and finally entering a true late stage in which the HPV16 late L1 and L2 structural proteins are synthesized and virions are produced (Chow et al., 2010; Doorbar

et al., 2012; Kajitani et al., 2012; Mighty and Laimins, 2014). It is presumably during the “pre-late” stage that the HPV16 late promoter is activated, thereby generating mRNAs that bypass the upstream E6 and E7 open reading frames (orfs) (Bernard, 2013; Thierry, 2009). This promoter switch could therefore potentially generate mRNAs with greater ability to translate the downstream open reading frames E1, E2 and E4. In addition to the promoter switch, abundant alternative splicing of the HPV16 mRNAs contribute to the structure of the final mRNA products (Graham and Faizo, 2017; Jia and Zheng, 2009; Johansson and Schwartz, 2013; Schwartz, 2013). Little is known about the translational potential of HPV16 mRNAs initiated at the early and late HPV16 promoters and processed by alternative mRNA splicing.

The HPV16 E2 protein is a key factor in the HPV16 replication cycle (Kadaja et al., 2009; McBride, 2013; Thierry, 2009). E2 is a truly multifunctional protein. Among all, E2 specifically binds HPV16 DNA and functions as the main viral transcriptional regulator of the HPV16 genes, it contributes to formation of nuclear foci in which the HPV16 DNA genome replicates, E2 contributes to initiation of HPV16 DNA

* Corresponding authors.

E-mail addresses: wj5532@dlut.edu.cn (C. Wu), Stefan.Schwartz@med.lu.se (S. Schwartz).

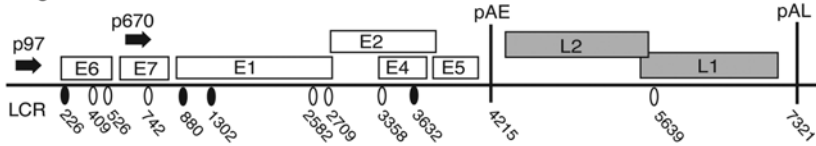
<https://doi.org/10.1016/j.virusres.2020.198004>

Received 16 April 2020; Received in revised form 29 April 2020; Accepted 29 April 2020

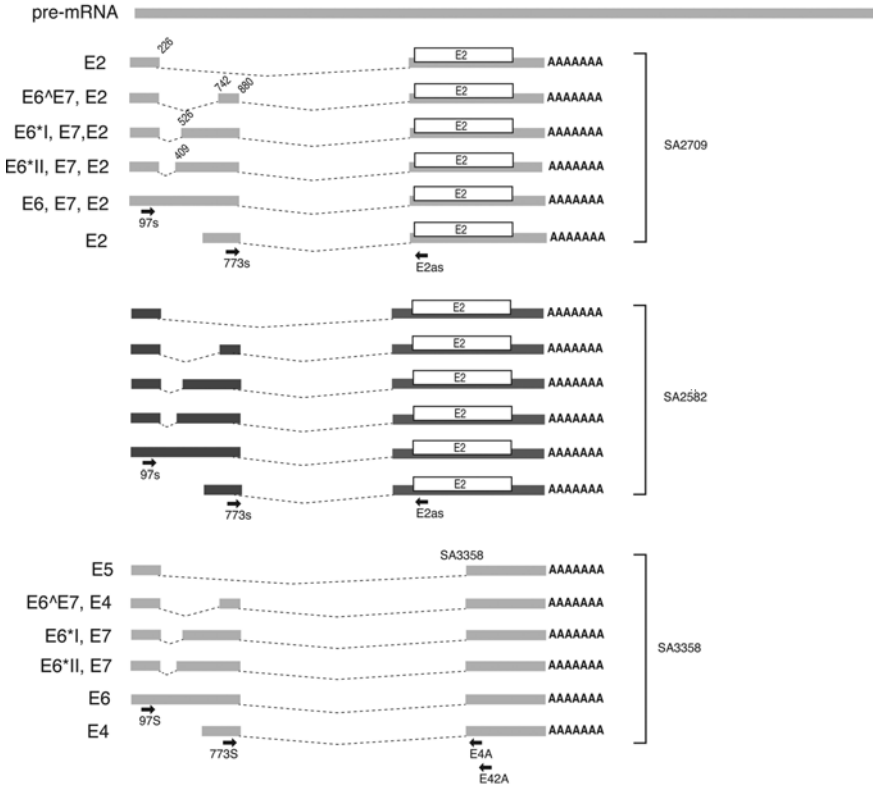
Available online 05 May 2020

0168-1702/ © 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

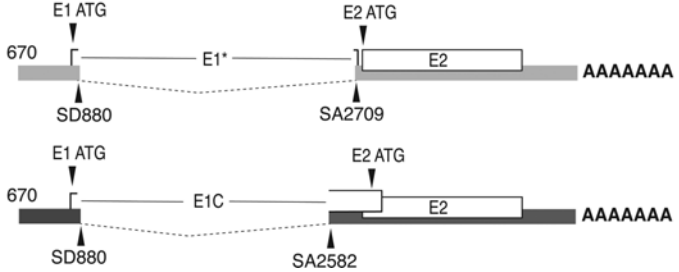
A HPV16 genome



B



C



(caption on next page)

Fig. 1. (A) Schematic representation of the HPV16 genome. are indicated. HPV16 early and late promoters p97 and p670 and splice sites are indicated. pAE, HPV16 early polyadenylation signal; pAL, HPV16 late polyadenylation signal. (B) Schematic representation of HPV16 mRNAs that encode full-length E2 and are utilizing HPV16 3'-splice site SA2709 or SA2582. Below these mRNAs, schematic representations of HPV16 mRNAs utilizing alternative 3'-splice site SA3358. These mRNAs spliced to SA3358 do not encode E2. RT-PCR primers 97S, 773S, E2A, E4A and E42A are indicated. Nucleotide positions refer to HPV16R (Van Doorslaer et al., 2013). (C) Schematic representation of HPV16 mRNAs that encode full-length E2 and are utilizing HPV16 3'-splice site SA2709 or SA2582. The open reading frames E1* created by splicing from SD880 to SA2709 and E1C created by splicing from SD880 to AA2582 are indicated.

replication, it tethers the HPV16 genome to cellular chromatin during mitosis to facilitate partitioning and E2 contributes to packaging of the genomic DNA into virus particles (Kadaja et al., 2009; McBride, 2013; Thierry, 2009). In addition to these functions, E2 plays a role in the posttranscriptional control of HPV16 gene expression (Johansson et al., 2012). The E2 protein paves the way for HPV16 late gene expression by shutting down the HPV16 early promoter p97 (McBride, 2013), and by inhibiting the HPV16 early polyadenylation signal pAE to allow for read-through into the late region of the HPV16 genome (Johansson et al., 2012). The HPV16 E2 gene may be inactivated when HPV16 occasionally integrates in the cellular genome (Kadaja et al., 2009; McBride, 2013; Thierry, 2009). Restoration of E2 in such cells leads to cellular growth arrest and senescence. Thus, inactivation of E2 may contribute to the ability of HPV16 to cause cancer. In the HPV16 life cycle, high levels of E2 shut down the HPV16 early promoter, which down-regulates expression of the growth promoting E6 and E7 genes, thereby offering a chance for the infected cell to proceed to terminal differentiation, induce late gene expression and to produce virus particles. It appears that E2 is required through the entire HPV16 life cycle, suggesting that E2 could be produced from both HPV16 early and late mRNAs.

The HPV16 early and late proteins are produced from a number of alternatively spliced transcripts expressed from the HPV16 early and late promoter p97 and p670 (Graham and Faizo, 2017; Jia and Zheng, 2009; Johansson and Schwartz, 2013; Schwartz, 2013). Although the structures of many HPV16 mRNAs are known, less is known about their translational potential. A number of HPV16 mRNAs encoding full-length E2 have been identified, some of which initiate at the HPV16 early promoter and some at the HPV16 late promoter (Schwartz, 2013; Van Doorslaer et al., 2013). However, since all HPV16 mRNAs are multicistronic and encode several orfs, it is often difficult to predict what they actually produce during translation. Here we have generated E2 cDNAs with the secreted luciferase (sLuc) reporter gene in place of E2 to determine how well a number of alternatively spliced E2 mRNAs translate the E2 orf. Our results demonstrate that a late mRNA initiating at HPV16 nucleotide position 670 (the late promoter in the HPV16 genome), and being spliced from HPV16 5'-splice site SD880 to HPV16 3'-splice site SA2709, located immediately upstream of E2, is the HPV16 mRNA with the highest ability to produce E2.

2. Materials and methods

2.1. Plasmids

The following plasmids have been described previously: plasmid pHPV16AN contains a full-length HPV16 genome flanked by loxP sites that can be recombined by cre-enzyme to release the episomal form of the HPV16 genome in transfected cells (Johansson et al., 2012; Li et al., 2013b) (Fig. 2A), plasmid pC97EL contains HPV16 sequences from HPV16 nucleotide position 97–7372 and is driven by the human cytomegalovirus immediate early promoter (CMV) (Li et al., 2013a) (Fig. 2B), plasmid pC97ELsLuc is a derivative of pC97EL in which a BamHI-XhoI fragment in the HPV16 L1 gene has been replaced by the poliovirus 2A internal ribosome entry site followed by the secreted luciferase (sLuc) gene (Li et al., 2013a), plasmid p3* encodes an HPV16 cDNA representing an HPV16 mRNA that is initiated at HPV16 nucleotide position 97 and spliced from HPV16 5'-splice site SD226 to HPV16 3'-splice site SA742 and from HPV16 5'-splice site SD880 to HPV16 3'-splice site

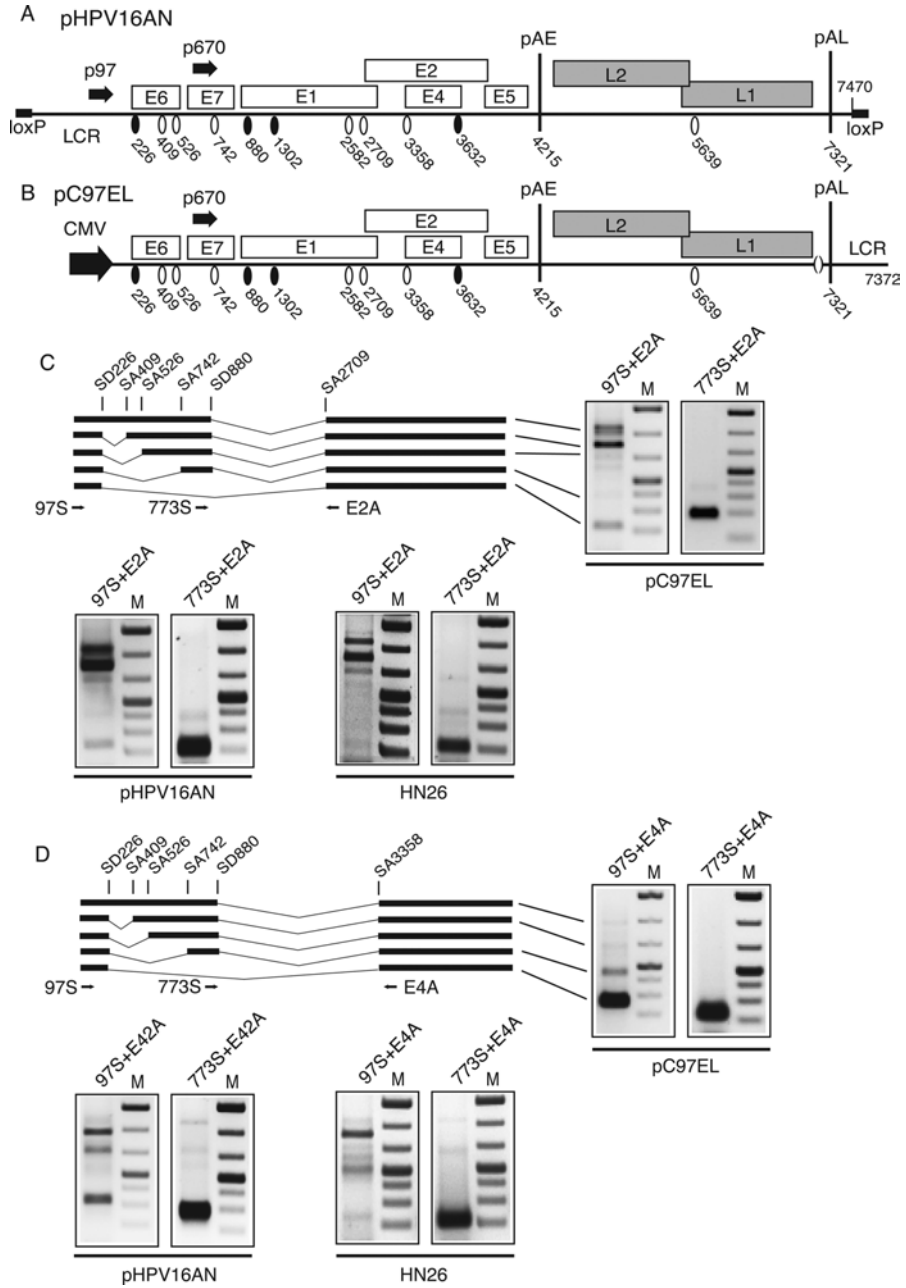
SA3358 (97-SD226*SA742-SD880*SA3358) (Nilsson et al., 2018), plasmid p5 encodes an HPV16 cDNA representing an HPV16 mRNA that is initiated at HPV16 nucleotide position 97 and spliced from HPV16 5'-splice site SD226 to HPV16 3'-splice site SA409 and from HPV16 5'-splice site SD880 to HPV16 3'-splice site SA3358 (97-SD226*SA409-SD880*SA3358) (Nilsson et al., 2018), plasmid p6 contains the sLuc gene under control of the CMV promoter (Nilsson et al., 2018), plasmid pSRSF1 contains the SRSF1 cDNA under control of the CMV promoter (Somberg and Schwartz, 2010), plasmid pC0806 contains the CMV promoter followed by a polylinker and the HPV16 late polyadenylation signal pAL (Collier et al., 2002). To generate p1 and p2 (Fig. 3A), PCR was first carried out on p3* and p5, respectively, with oligonucleotides 97SALS (5'-ggctcgcagagaaactgcaatgtttcaggaccacag-3') and 880–2709AS (5'-aacgcgtgtttcttctctgctctgctcgtgcaaaacttaactggaccactgctcgcagatcagcattgtag-3'). To generate p3 and p4 (Fig. 3A), PCR was performed on pC97ELsLuc with primers 97SALS and 226–2709AS (5'-aacgcgtgtttcttctctgctctgctcgtgcaaaacttaactggaccactgctcgcagatcagcattgtag-3'), 670SALS (5'-ggctcgcagaaatagatggtcagctggacaagc-3') and 880–2709AS. The amplified p1, p2, p3 and p4 fragments were subcloned into pCR4-TOPO (ThermoFisher). The TOPO plasmids containing p1, p2, p3, and p4 fragments were digested with Sall (FD0644, ThermoFisher) and MluI (FD0564, ThermoFisher) followed by agarose electrophoresis and purification with a Gel-Extraction Kit (CAT#28706, QIAGEN). The fragments were ligated to the vector created by p3* (Nilsson et al., 2018) digested with Sall and MluI. To generate p2582, we ordered synthesis of a DNA sequence containing a Sall site followed by an HPV16 sequences starting at HPV16 nucleotide position 670 and continuing to position 880 followed by HPV16 sequences from position 2582 to an MluI site immediately upstream of the E2ATG at position 2756 (Eurofins). This DNA fragment was released with Sall and MluI and insert and ligated to Sall- and MluI-digested p3* vector (Nilsson et al., 2018). Nucleotide positions refer to HPV16R (Van Doorslaer et al., 2013).

2.2. Cells and transfection

The HN26 cells are derived from a tumor of a 48-year old non-smoking 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 (Forslund et al., 2019; Wu et al., 2019). HeLa cells and HN26 cells were cultured in Dulbecco's modified Eagle medium with 10% bovine calf serum and 1% penicillin-streptomycin. Transfections of HeLa cells were carried out using Turbofect according to the manufacturer's instructions (Fermentas). Turbofect was mixed with plasmid DNA and incubated at room temperature for 15 min prior to drop-wise addition to 60-mm plates with subconfluent HeLa cells. Cells were harvested at 20 h post transfection. Each plasmid was transfected in triplicate, in a minimum of two independent experiments. For analysis of episomal HPV16, plasmid pHPV16AN was co-transfected with plasmid pCAGGS-nlscr (generously provided by Dr. Andras Nagy at University of Toronto) (Nagy, 2000), which expresses the cre recombinase that releases the HPV16 genome from the plasmid at two flanking lox-sites.

2.3. RNA extraction and RT-PCR

Total RNA was extracted using TRI Reagent and Direct-zol RNA MiniPrep kit (ZYMO Research) according to the manufacturer's



(caption on next page)

Fig. 2. (A) Schematic representation of the HPV16 genome and (B) the HPV16 subgenomic pC97EL reporter plasmid (Li et al., 2013a). Transcription of the HPV16 sequences in the pC97EL plasmid is driven by the human cytomegalovirus immediate early promoter (CMV). (C) Schematic representation of HPV16 mRNAs spliced to SA3358. RT-PCR primers 97 s, 773 s and E4a are indicated. Gel pictures of RT-PCR reactions performed with 97s or 773 s and E4a on total RNA extracted from HeLa cells transfected with HPV16 subgenomic expression plasmid pC97EL (Li et al., 2013a), cre-loxod full-length HPV16 genomic plasmid pHPV16AN (Li et al., 2013b) or from HPV16-containing tonsillar cancer cell line HN26 (Forsslund et al., 2019; Wu et al., 2019). (D) Schematic representation of HPV16 mRNAs spliced to SA2709. RT-PCR primers 97 s, 773 s and E2a are indicated. Gel pictures of RT-PCR reactions performed with 97s or 773 s and E2a on total RNA extracted from HeLa cells transfected with HPV16 subgenomic expression plasmid pC97EL, cre-loxod full-length HPV16 genomic plasmid pHPV16AN or from HPV16-containing tonsillar cancer cell line HN26. Nucleotide positions refer to HPV16R (Van Doorslaer et al., 2013).

protocol. 1 µg of total RNA was reverse transcribed in a 20 µL reaction at 37 °C 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 were amplified with RT-PCR primers 97S (5'-ggcgcgaactgcaatgtttcaggacc-3') and E4A (5'-tgctcctaatagtttcaggagagg-3'), E42A (5'-cgtgccaagc-gacgcttgg-3') or E2A (5'-cctgaccaccgcatgaacttc-3'). HPV16 mRNAs spliced from HPV16 5'-splice site SD880 were amplified with RT-PCR primers 773S (5'-gcacacacgtagacattgctactttg-3') and E4A, E42A or E2A. Alternatively, HPV16 cDNAs generated from mRNAs produced from the E2 expression plasmids were PCR-amplified with primers CSENSE (5'-ctgaattgggtgctgacgt-3') and PANTI (5'-tgttgcacacacatagtactc-3'). The HPV16 RT-PCR primers used here are specific for cDNAs produced from HPV16 mRNAs and do not amplify cDNA from untransfected HeLa cells. GAPDH cDNA was amplified with primers F-GAPDH (5'-accaga-gaagactgtggaugg-3') and R-GAPDH (5'-ttctagacgcgagtcaggt-3').

2.4. Protein extraction and Western blotting

Western blotting was performed as described previously (Kajitani et al., 2017). The following antibodies were used: anti-HPV16 E2 antibody TVG261 (ab-17185, Abcam) and anti-actin antibody (sc-1616, Santa Cruz). Filters were thereafter incubated with rabbit anti-mouse IgG-horse radish peroxidase antibody (A9044, Sigma Aldrich), and proteins were detected using the Clarity Western ECL Substrate (Bio Rad) or the Super Signal West Femto chemiluminescence substrate (Pierce).

2.5. Secreted luciferase assay

The *Metridia longa* secreted luciferase activity (Markova et al., 2004) in the culture medium of the C33A2 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).

3. Results

3.1. The most common HPV16 E2 mRNA is generated by splicing from SD880 to SA2709

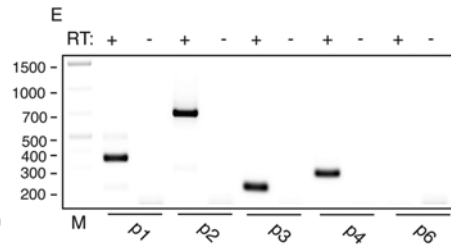
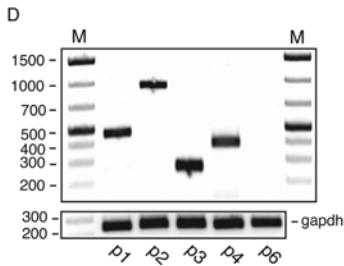
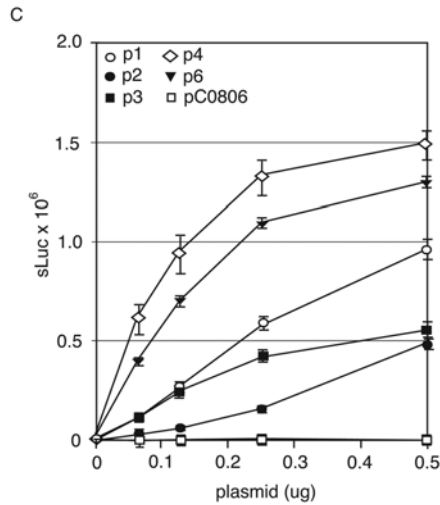
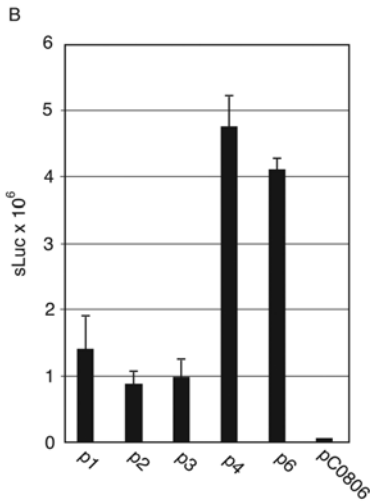
The HPV16 genome produces a number of mRNAs encoding the E2 open reading frame (ORF) (Fig. 1 A and B) (Graham, 2016; Jia and Zheng, 2009; Johansson and Schwartz, 2013; Schwartz, 2013; Van Doorslaer et al., 2013). These mRNAs are generated by splicing from HPV16 5'-splice site SD226 or SD880 to the HPV16 3'-splice site SA2709 located immediately upstream of the E2 ORF (Fig. 1A and B). These splicing events generate a number of alternatively spliced mRNAs that encode E2 and could potentially produce E2 protein. An additional HPV16 3'-splice site named SA2582 is located 175 nucleotides upstream of the E2 ATG (Fig. 1A and B). The significance of SA2582 is unclear and this mRNA could potentially be translated into E2 protein. However, it also encodes a shorter E1-orf named E1C that could

potentially be produced (Fig. 1C), but expression and function of E1C have not been studied. This is in contrast to HPV16 mRNAs that utilize SA2709, on which the shorter E1-orf named E1* stops prior to the E2 ATG (Fig. 1C). If HPV16 5'-splice sites SD226 and/or SD880 fail to splice to SA2709 or SA2582, they splice to the downstream HPV16 3'-splice site SA3358 (Fig. 1A and B). The latter mRNAs do not encode full-length E2.

To identify the most common alternatively spliced HPV16 E2 mRNA, we extracted total RNA from cells transfected with the HPV16 subgenomic expression plasmid pC97EL which is driven by the CMV promoter (Li et al., 2013a) (Fig. 2B) or with pHPV16AN (Johansson et al., 2012; Li et al., 2013b) (Fig. 2A) that encodes the full-length HPV16 genome. The latter plasmid contains a loxP site on each side of the genome (Fig. 2A), and was co-transfected with cre-expressing plasmid to release HPV16 episomes in the transfected cells as described previously (Johansson et al., 2012; Li et al., 2013b). RNA was extracted from HeLa cells transfected with either of these two plasmids and the RNA samples were analyzed by RT-PCR. Primers 773S and E2A (Fig. 2C) detected primarily one band representing splicing from SD880 to SA2709 (Fig. 2C) and at times a weaker band running higher up in the gel, representing an mRNA spliced from SD880 to SA2582 (Fig. 2C). All RT-PCR bands were sequenced to confirm their identity. Similar results were obtained with RNA extracted from the HPV16-infected tonsillar cancer cell line HN26 (Forsslund et al., 2019; Wu et al., 2019) (Fig. 2C). Interestingly, when RNA samples were analyzed with primers 97S and E2A, the results revealed that splicing from SD226 to E2 splice site SA2709 was relatively rare in pC97- and pHPV16AN-transfected cells and in HPV16-carrying tonsillar cancer cell line HN26 (Fig. 2C). Splicing to the competing 3'-splice site SA3358 often occurred from SD880 (Fig. 2D) and splicing from SD226 to SA3358 was relatively more common than splicing from SD226 to SA2709 (Fig. 2C and D). Note that band intensities between the different gels are not compared, only relative differences between bands in each lane. We concluded that the most common HPV16 E2 mRNA is generated by splicing from SD880 to SA2709 in HPV16-transfected cells as well as in HPV16-infected cells.

3.2. The most efficiently translated HPV16 E2 mRNA is generated by splicing from SD880 to SA2709

To determine the translational potential of the various alternatively spliced E2 mRNAs, we generated cDNA expression plasmids with the most common alternatively spliced E2 mRNAs under control of the human cytomegalovirus immediate early promoter (CMV) (Fig. 3A). The E2 ORF was replaced by the secreted luciferase gene downstream of the E2 ATG for easy quantitation of translation initiation at the E2 ATG. The generated plasmids were individually transfected into HeLa cells in parallel with empty CMV-vector (pC0806) (Collier et al., 2002) and CMV-driven sLuc positive control plasmid p6 (Nilsson et al., 2018). sLuc activity in the cell culture medium was monitored at 24 h post transfection, mean values from triplicate transfections determined and standard deviation calculated (Fig. 3B). The results revealed that the cDNA representing an E2 mRNA initiated at position 670 (late promoter) and spliced from SD880 directly to SA2709 (plasmid p4) produced more sLuc than the other cDNAs that represented mRNAs initiated at HPV16 nucleotide 97 (HPV16 early promoter) and were



(caption on next page)

spliced to SA2709 directly from SD226 (plasmid p3) or via SA409 and SD880 (plasmid p2), or SA742 and SD880 (plasmid p1) (Fig. 3B). Transfection of serially diluted cDNA expression plasmids yielded

similar results (Fig. 3C). The mRNA levels were similar in all transfections and only one band was detected from all plasmids (Fig. 3D). The RT-PCR bands were not detected in the absence of reverse

Fig. 3. (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*1-, 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*1). 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 from 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. (C) 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). (D) RT-PCR with primers CSENSE and E2A on total RNA extracted from the HeLa cells transfected with the indicated plasmids. (E) RT-PCR with primers CSENSE and E2A on total RNA extracted from the HeLa cells transfected with the indicated plasmids. RT-PCR was performed in the absence (-) or presence (+) of reverse transcription (RT).

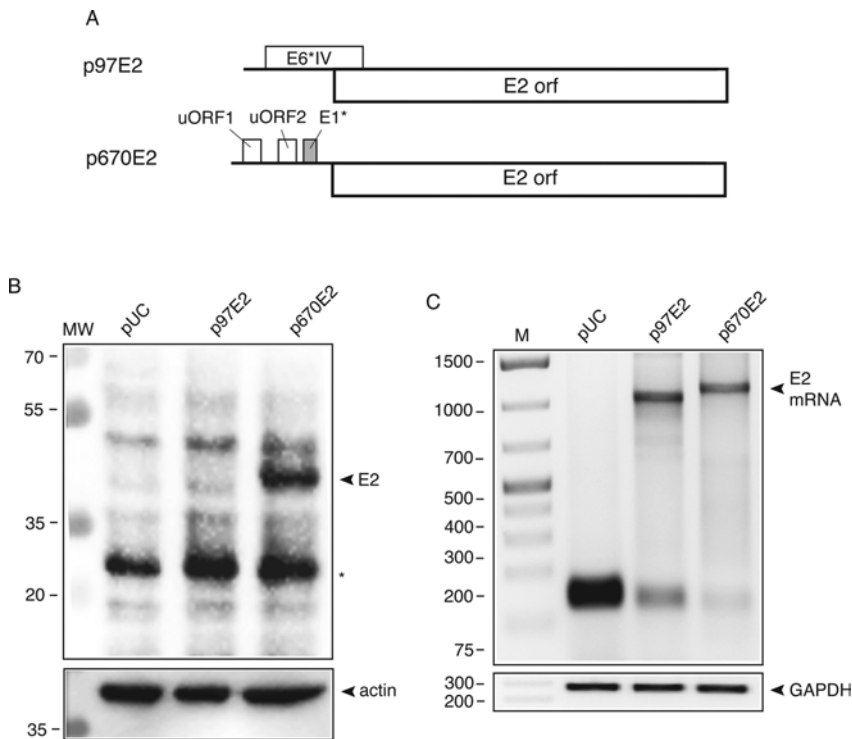


Fig. 4. (A) Schematic representation of the HPV16 E2 cDNA expression plasmids p97E2 and p670E2. p97E2 encodes a cDNA representing an HPV16 E2 mRNA initiating at nucleotide position 97 (early promoter P97 position) and spliced from SD226 directly to SA2709, and p670E2 represents an mRNA initiating at nucleotide position 670 (late promoter P670 position) and spliced from SD880 to SA2709. The HPV16 E6*IV-orf has the E6 ATG and is generated by splicing from HPV16 SD226 to SA2709 and 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 from the p670E2 cDNA are indicated. (B) 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. (C) RT-PCR with primers CSENSE and PANTI on total RNA extracted from HeLa cells transfected with the indicated plasmids.

transcription (Fig. 3E). We concluded that the HPV16 mRNA that was initiated at HPV16 late promoter-position at nucleotide 670 and was spliced from HPV16 5'-splice site SD880 to HPV16 3'-splice site SA2709 showed that highest translation initiation at the E2 ATG.

To confirm that the mRNAs encoding sLuc reflected the translation of the entire E2 orf, we generated two E2 expression plasmids with the same leader sequences as in plasmids p3 and p4 (Fig. 3A and Fig. 4A). In contrast to p3 and p4, these plasmids encoded the entire E2 orf and were named p97E2 and p670E2 (Fig. 4A). Plasmids p97E2 and p670E2 were separately transfected into HeLa cells and the levels of HPV16 E2

protein were monitored by Western blotting. As can be seen, plasmid p670E2 produced higher levels of E2 protein than plasmid p97E2 (Fig. 4B), while mRNA levels produced from the two plasmids were similar (Fig. 4C). Therefore, these results reproduced the results obtained with the sLuc-plasmids. We concluded that the HPV16 mRNA with the highest ability to produce E2 protein was an mRNA initiating at nucleotide position 670 (HPV16 late promoter) followed by splicing from HPV16 5'-splice site SD880 to HPV16 3'-splice site SA2709.

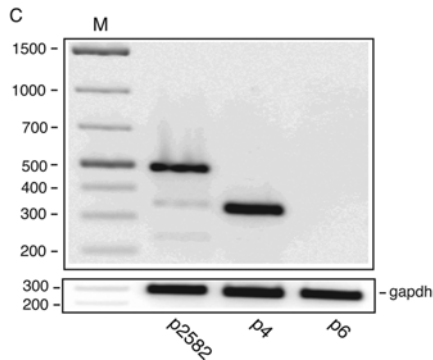
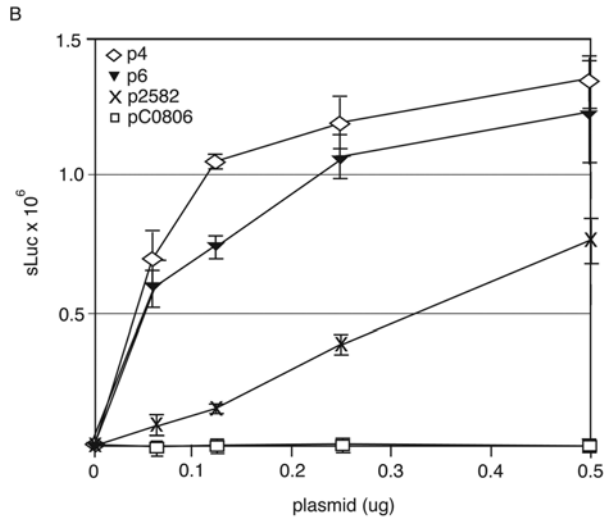
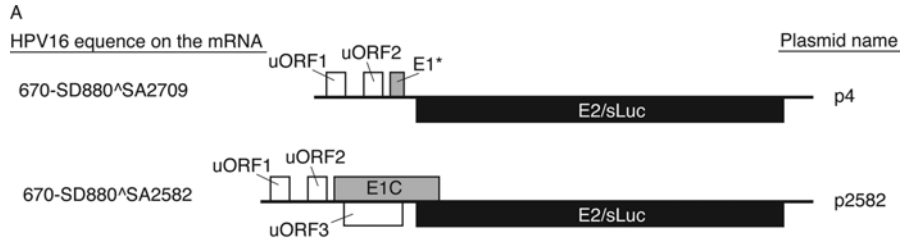


Fig. 5. (A) Schematic representation of the HPV16 E2 cDNA expression plasmids p4 and p2582. The HPV16 sequences present on the cDNAs in each plasmid are indicated. Names of cDNA expression plasmids are indicated to the right. E2/sLuc represents the fusion of the secreted luciferase coding sequence in 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 from 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). **(C)** RT-PCR with primers CSENSE and E2A on total RNA extracted from HeLa cells transfected with the indicated plasmids.



Fig. 6. Schematic representation of the 5'-end of the late HPV16 mRNA starting at nucleotide position 670 (at the late HPV16 promoter p670) followed by splicing from HPV16 SD880 to HPV16 SA2709 and continuing to the E2 ATG. Three open reading frames with ATGs are located upstream of the E2 ATG and they are indicated in the figure (uORF1, uORF2 and uORF3). None of these uORFs span the E2 ATG. The sequence at each ATG is indicated. uORF1 encodes a 16-amino acid peptide, uORF2 a 15-amino acid peptide and uORF3 a 10-amino acid peptide. SJ, splice junction between HPV16 splice sites SD880 and SA2709.

Table 1

Context of ATGs of small upstream open reading frames on various HPV16 E2 mRNAs.

ORF name	ATG context	orf length (aa)
"optimal ATG"	CCRCATGG	–
uORF1	aatagATGg	16
uORF2	tgtaaATGg	15
uORF3	tctcaATGg	30
E1*/E1C (E1 ATG)	ctaccATGg	10
E2	aacgATGg	365

3.3. The HPV16 mRNA that is spliced between HPV16 SD880 and SA2709 is more efficiently translated into E2 than the mRNA spliced from HPV16 SD880 to SA2582

Two 3'-splice sites are located immediately upstream of the HPV16 E2 ATG: SA2582 and SA2709 (Van Doorslaer et al., 2013) (Alloul and Sherman, 1999a, b) (see also Fig. 1A, B and C). Although both of these splice sites are utilized in HPV16 infected cells, it appears that SA2709 dominates (Van Doorslaer et al., 2013) (Alloul and Sherman, 1999a) (Fig. 2C). The significance of HPV16 SA2582 in the HPV16 life cycle is currently unknown. We investigated the translation efficiency of mRNAs spliced from SD880 to SA2582 compared to mRNAs spliced from SD880 to SA2709. cDNAs representing mRNAs initiated at HPV16 position 670 (late promoter) and spliced from SD880 to SA2582 (p2582) or to SA2709 (p4) were cloned downstream of the CMV promoter and the E2 orf was replaced by the sLuc orf (Fig. 5A). These two plasmids and negative control plasmid (pC0806) and positive control plasmid p6 were serially diluted and transfected into HeLa cells in triplicates. The results revealed that mRNA spliced from SD880 to SA2582 initiated translation at the E2 ATG much less efficiently than mRNAs spliced from SD880 to SA2709 (Fig. 5B). As expected, mRNAs spliced from SD880 to SA2709 produced similar levels of sLuc as the positive control plasmid p6. Cells transfected with empty pUC plasmid did not produce sLuc-levels above background (Fig. 5B). Plasmids p2582 and p6 produced similar mRNA levels in the transfected HeLa cells (Fig. 5C). In conclusion, our results demonstrated that initiation of translation at the E2 ATG occurred much less efficiently on mRNAs spliced between SD880 and SA2582 than on mRNAs spliced from SD880 to SA2709. We concluded that the major HPV16 E2 mRNA initiated at late promoter p670 and was spliced from SD880 to SA2709 (Fig. 6).

4. Discussion

The structure of the major HPV16 E2 mRNA initiated at late promoter 670 and spliced from SD880 to SA2709 is shown in Fig. 6. This mRNA contains three ATGs upstream of the E2 ATG. These upstream

ATGs head open reading frames designated uORF1, uORF2 and uORF3 that could potentially be translated into peptides of 16-, 15- or 10-amino acids, respectively (Table 1). A comparison to the Kozak sequence for optimal translation initiation which predicts a "G" nucleotide at position +4 downstream of the ATG and a purine (R) at position -3 upstream of the ATG (Kozak, 1992), revealed that only the E1* has an optimal ATG (apart from the E2 ATG itself) (Table 1). uORF3 is actually headed by the E1 ATG, which is an optimal ATG, but this orf runs into a stop codon very quickly on the mRNAs that are spliced from SD880 to SA2709 (Table 1). The short E1* orf is predicted to encode a peptide of 10 amino acids and could as such be recognized by scanning ribosomes. However, since short upstream orfs have a lower ability than long upstream orfs to affect translation of downstream orfs (Luukkonen et al., 1995), it is reasonable to assume that the E1* is too short to pose a threat to E2 translation, despite its optimal ATG. In addition, all upstream orfs end prior to the E2 ATG and none of them overlap the E2 ATG. Therefore, scanning ribosomes that recognize these upstream orfs have the potential to re-initiate at the E2 ATG. This is in sharp contrast to ribosomes that translate upstream orfs that overlap a downstream ATG (Kozak, 1992). Such upstream ATGs may have a major impact on translation initiation of downstream orfs if the upstream orf encodes a protein of significant size (Kozak, 1992) (Luukkonen et al., 1995). Since the sLuc production from the E2 mRNA spliced from SD880 to SA2709 was equally high as the sLuc production from the control plasmid p6 in which the sLuc gene is located immediately downstream of the CMV promoter (Fig. 3 and 4), we concluded that the short upstream orfs on the E2 mRNAs that were spliced from SD880 to SA2709 did not significantly affect translation initiation at the E2 ATG.

Translation of HPV16 E2 mRNAs has been addressed previously in an article in which E2 cDNAs were in vitro translated as well as transfected into COS cells to determine how much E2 each of the mRNAs produced (Alloul and Sherman, 1999a). These investigators analyzed three E2 mRNAs, all initiated at the early promoter, two that were spliced to SA2709 and one that was spliced to SA2582 (Alloul and Sherman, 1999a). Their results revealed that of the E2 mRNAs starting at nucleotide position 97 (HPV16 early promoter p97), the mRNA spliced from SD226 to SA2709 produced more E2 than a mRNA spliced from SD226-SA409-SD880-SA2709 (Alloul and Sherman, 1999a). This is in total agreement with our results. These authors also concluded that a control plasmid that only encoded the E2 orf produced more E2 than their E2 cDNA expression plasmids (Alloul and Sherman, 1999a). This is also in agreement with our results. However, these investigators did not analyze translation of an E2 mRNA initiating at HPV16 nucleotide 670 (representing the HPV16 late promoter p670). Our results revealed that this HPV16 mRNA that starts at nucleotide position 670 and is spliced from SD880 to SA2709 was the mRNA that produced most E2 protein. As a matter of fact, it was translated equally

well as our sLuc control plasmid that lacked HPV16 sequences upstream of the E2 ATG. Regarding technical differences between our experiments and those published by Alloul and Sherman that could potentially affect experimental outcome, it should be noted that Alloul and Sherman used plasmid vector pJS55 in which they inserted a cDNA encoding a C-terminally tagged E2 protein under the control of Simian Virus 40 (SV40) promoter and with SV40 polyadenylation signal. This vector also contains a beta-globin intron sequence with functional splice sites located between the SV40 promoter and the E2 cDNA, as well as an SV40 origin of replication that allows for plasmid replication in transfected COS cells. In contrast, our E2 cDNA expression plasmids are driven by the CMV promoter, they do not have additional splice sites and they do not replicate in transfected cells. Furthermore, Alloul et al. monitored protein production from the various mRNAs by immunoprecipitation of labeled E2 protein, while we monitored sLuc production as a marker for E2 production. Finally, Alloul et al. used CV-1-derived, SV40-immortalized simian cell line COS that produces SV40 large-T antigen that replicates transfected plasmids with SV40 origin of replication, while we used HeLa cancer cells containing HPV18. We believe that HeLa cells are suitable for these studies since they can be easily and reproducibly transfected. Since they are derived from the uterine cervix and contain integrated HPV18, they are derived from cells that are permissive for HPV. They are transformed malignant cells and as such do not fully represent keratinocytes that are targeted by HPV16 at the site of infection, but since translation is a fundamental process, we believe that our results are representative for translation in keratinocytes. However, transfection of primary keratinocytes is required to confirm our results. Taken together, our results confirmed previous studies on HPV16 early mRNAs encoding E2 and extended them to show that the most efficiently translated E2 mRNA is a late E2 mRNA initiated at p670 and being spliced from SD880 to SA2709.

An alternative E2 3'-splice site has been mapped to sequences 175 nucleotides upstream of the E2 ATG and 128 nucleotides upstream of SA2709 in HPV16 (Van Doorslaer et al., 2013) (Alloul and Sherman, 1999a, b). For all other HPV types on which transcript-mapping has been performed, it appears that only one 3'-splice site is present immediately upstream of E2. The significance of the alternative E2 splice site SA2582in HPV16 is unknown and the benefit of having a second E2 splice site is unclear generating a suboptimal E2 mRNA is unclear. Our results clearly demonstrate that an mRNA initiated at position 670 (representing late promoter p670) and is spliced from SD880 to SA2582 is less well translated into E2 than mRNAs spliced from SD880 to SA2709. However, a previous analysis of HPV16 E2 mRNA translation compared E2 production from two mRNAs initiated at nucleotide 97 (HPV16 early promoter p97), followed by splicing from SD226 to SA409 and SD880 to either SA2709 or SA2582 (Alloul and Sherman, 1999a). Their experiments suggested that the early E2 mRNAs spliced to SA2582 produced slightly more E2 than the mRNAs spliced to SA2709 (Alloul and Sherman, 1999a). This is opposite to the results we obtained with late HPV16 E2 mRNAs. We concluded that HPV16 late E2 mRNAs spliced to SA2582 are suboptimal compared to E2 mRNAs spliced to SA2709. The big difference in E2 translation, suggested that SA2582 does not play a significant role in the late stage of the HPV16 infection, whereas at the early stage, it may contribute to E2 production. The inefficient translation of E2 from the mRNA spliced from SD880 to SA2582 is likely due to the fact that the upstream strong E1 ATG generates a short orf named E1C that overlaps the E2 ATG (Van Doorslaer et al., 2013), leaving these mRNAs no chance to translate E2 by reinstitution of translation. This is in contrast to the other E2 mRNAs that use SA2709 on which the strong upstream E1 ATG heads a short open reading frame named E1* that stops prior to the E2 ATG (Van Doorslaer et al., 2013), thereby allowing ribosomes that recognize the short E1* orf to reinstitute translation at the E2 ATG. In addition to the E1C orf with an optimal ATG, these mRNAs also contain an additional orf with ATG (uORF3) that overlaps the E1C orf and encodes 30 amino acids and could potentially reduce translation initiation at the E2 ATG.

However, this is less likely since the uORF3 orf is endowed with a suboptimal ATG (Table 1). However, to provide experimental results that support this idea would require evaluation of ribosome binding and protein synthesis rates, which is beyond the scope of this project. Since the role of the mRNAs utilizing HPV16 3'-splice site SA2582 is unknown, one may speculate that SA2582 is used for modulation of E2 production. Alternatively, the short upstream E1C orf produces a functional protein.

We have previously identified the major HPV16 E5 mRNA (Nilsson et al., 2018). In contrast to the HPV16 E2 mRNA, the HPV16 E5 mRNA initiated at nucleotide position 670 was a very poor producer of E5 since it was preceded by the E1*E4 orf. This translational block was caused by the strong E1 ATG of the E1*E4 orf (Nilsson et al., 2018). Although the same E1 ATG is located upstream of the E2 ATG on the E2 mRNA identified here as the best producer of E2, it heads a very short orf on these alternatively spliced E2 mRNAs and as such has very little influence on E2 translation (Fig. 6). In contrast, the major E5 mRNA was spliced from SD226 directly to SA3358 thereby bypassing the E1 ATG and the upstream E1*E4 orf (Nilsson et al., 2018). We concluded that HPV16 E5 is primarily produced from an early mRNA initiated at nucleotide position 97 (HPV16 early promoter) and spliced from SD226 to SA3358, while, in contrast, E2 is primarily produced from a late mRNA initiated at nucleotide position 670 (HPV16 late promoter). These results are consistent with a major role for the E2 protein at the late, or pre-late stage of the HPV16 replication cycle, when the late promoter p670 is activated by cell differentiation and HPV16 E2 shuts down the early promoter p97, when E2 contributes to HPV16 DNA replication and when E2 inhibits early mRNA polyadenylation to induce HPV16 late gene expression.

5. Conclusion

We have shown that the HPV16 mRNA with the greatest ability to produce E2 protein is generated from the HPV16 late promoter and is spliced between HPV16 5'-splice site SD880 and HPV16 3'-splice site SA2709.

CRedit authorship contribution statement

Yunji Zheng: Methodology, Validation, Formal analysis, Investigation, Visualization, Funding acquisition, Writing - original draft. **Xiaoxu Cui:** Validation, Methodology, Investigation, Project administration. **Kersti Nilsson:** Resources, Visualization, Project administration. **Haoran Yu:** Resources, Project administration. **Lijing Gong:** Investigation, Resources, Supervision. **Chengjun Wu:** Conceptualization, Supervision, Project administration. **Stefan Schwartz:** Conceptualization, Supervision, Project administration, Writing - review & editing, Funding acquisition.

Acknowledgements

This work was supported by the Swedish Research Council-Medicine [grant VR2019-01210] (Schwartz), by the Swedish Cancer Society [CAN2018/702] (Schwartz) and by the China Scholarship Council [Grant File No. 201606525004] (Zheng), [Grant File No. 201809120016] (Cui) and [Grant File No. 201606525004] (Gong).

References

- Alloul, N., Sherman, L., 1999a. The E2 protein of human papillomavirus type 16 is translated from a variety of differentially spliced polycistronic mRNAs. *J. Gen. Virol.* 80 (Pt 1), 29–37.
- Alloul, N., Sherman, L., 1999b. Transcription-modulatory activities of differentially spliced cDNAs encoding the E2 protein of human papillomavirus type 16. *J. Gen. Virol.* 80 (Pt 9), 2461–2470.
- Bernard, H.U., 2013. Regulatory elements in the viral genome. *Virology* 445, 197–204.
- Bouvard, V., Baan, R., Straif, K., Grosse, Y., Secretan, B., El Ghissassi, F., Benbrahim-

- Tallaa, L., Guha, N., Freeman, C., Galichet, L., Cogliano, V., on, W.I.A.F.R., Group, C.M.W., 2009. A review of human carcinogens—Part B: biological agents. *Lancet Oncol.* 10, 321–322.
- Chow, L.T., Broker, T.R., Steinberg, B.M., 2010. The natural history of human papillomavirus infections of the mucosal epithelia. *APMIS* 118, 422–449.
- Collier, B., Oberg, D., Zhao, X., Schwartz, S., 2002. Specific inactivation of inhibitory sequences in the 5' end of the human papillomavirus type 16 L1 open reading frame results in production of high levels of L1 protein in human epithelial cells. *J. Virol.* 76 (6), 2739–2752.
- Doorbar, J., Quint, W., Banks, L., Bravo, I.G., Stoler, M., Broker, T.R., Stanley, M.A., 2012. The biology and life-cycle of human papillomaviruses. *Vaccine* 30 (Suppl 5), F55–70.
- Forslund, O., Sugiyama, N., Wu, C., Ravi, N., Jin, Y., Swoboda, S., Andersson, F., Bzhalava, D., Hultin, E., Paulsson, K., Dillner, J., Schwartz, S., Wennerberg, J., Ekblad, L., 2019. A novel human in vitro papillomavirus type 16 positive tonsil cancer cell line with high sensitivity to radiation and cisplatin. *BMC Cancer* 19 (1), 265.
- Graham, S.V., 2016. Human papillomavirus E2 protein: linking replication, transcription, and RNA processing. *J. Virol.* 90 (19), 8384–8388.
- Graham, S.V., Faizo, A.A., 2017. Control of human papillomavirus gene expression by alternative splicing. *Virus Res.* 231, 83–95.
- Jia, R., Zheng, Z.M., 2009. Regulation of bovine papillomavirus type 1 gene expression by RNA processing. *Front. Biosci.* 14, 1270–1282.
- Johansson, C., Schwartz, S., 2013. Regulation of human papillomavirus gene expression by splicing and polyadenylation. *Nature Rev. Microbiol.* 11, 239–251.
- Johansson, C., Somberg, M., Li, X., Backström Winquist, E., Fay, J., Ryan, F., Pim, D., Banks, L., Schwartz, S., 2012. HPV-16 E2 contributes to induction of HPV-16 late gene expression by inhibiting early polyadenylation. *EMBO J.* 31, 3212–3227.
- Kadaja, M., Silla, T., Ustav, E., Ustav, M., 2009. Papillomavirus DNA replication - from initiation to genomic instability. *Virology* 384 (2), 360–368.
- Kajitani, N., Satsuka, A., Kawate, A., Sakai, H., 2012. Productive lifecycle of human papillomaviruses that depends upon squamous epithelial differentiation. *Front. Microbiol.* 3, 152.
- Kajitani, N., Glahder, J., Wu, C., Yu, H., Nilsson, K., Schwartz, S., 2017. hnRNP L controls HPV16 RNA polyadenylation and splicing in an Akt kinase-dependent manner. *Nucleic Acids Res.* 45 (16), 9654–9678.
- Kozak, M., 1992. Regulation of translation initiation in eucaryotic systems. *Annu. Rev. Cell Biol.* 8, 197–225.
- Li, X., Johansson, C., Cardoso-Palacios, C., Mossberg, A., Dhanjal, S., Bergvall, M., Schwartz, S., 2013a. Eight nucleotide substitutions inhibit splicing to HPV-16 3'-splice site SA3358 and reduce the efficiency by which HPV-16 increases the life span of primary human keratinocytes. *PLoS One* 8, e72776.
- Li, X., Johansson, C., Glahder, J., Mossberg, A.K., Schwartz, S., 2013b. Suppression of HPV-16 late L1 5'-splice site SD3632 by binding of hnRNP D proteins and hnRNP A2/B1 to upstream AUAGUA RNA motifs. *Nucleic Acids Res.* 41, 10488–10508.
- Luukkonen, B.G., Tan, W., Schwartz, S., 1995. 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 interchromatin distance. *J. Virol.* 69 (7), 4086–4094.
- Markova, S.V., Golz, S., Frank, L.A., Kalthof, B., Vysotski, E.S., 2004. Cloning and expression of cDNA for a luciferase from the marine copepod *Metridia longa*. *J. Biol. Chem.* 279, 3212–3217.
- McBride, A.A., 2013. The papillomavirus E2 proteins. *Virology* 445, 57–79.
- Mighty, K.K., Laimins, L.A., 2014. The role of human papillomaviruses in oncogenesis. *Recent Results Cancer Res.* 193, 135–148.
- Nagy, A., 2000. Cre recombinase: the universal reagent for genome tailoring. *Genesis* 26, 99–109.
- Nilsson, K., Norberg, C., Mossberg, A.K., Schwartz, S., 2018. HPV16 E5 is produced from an HPV16 early mRNA spliced from SD226 to SA3358. *Virus Res.* 244, 128–136.
- Schiffman, M., Doorbar, J., Wentzensen, N., de Sanjose, S., Fakhry, C., Monk, B.J., Stanley, M.A., Franceschi, S., 2016. Carcinogenic human papillomavirus infection. *Nat. Rev. Dis. Primers* 2, 16086.
- Schwartz, S., 2013. Papillomavirus transcripts and posttranscriptional regulation. *Virology.* 445, 187–196.
- Somberg, M., Schwartz, S., 2010. Multiple ASF/SF2 sites in the human papillomavirus type 16 (HPV-16) E4-coding region promote splicing to the most commonly used 3'-splice site on the HPV-16 genome. *J. Virol.* 84 (16), 8219–8230.
- Thierry, F., 2009. Transcriptional regulation of the papillomavirus oncogenes by cellular and viral transcription factors in cervical carcinoma. *Virology* 384, 375–379.
- Van Doorslaer, K., Tan, Q., Xirasagar, S., Bandaru, S., Gopalan, V., Mohamoud, Y., Huyen, Y., McBride, A.A., 2013. The Papillomavirus Episteme: a central resource for papillomavirus sequence data and analysis. *Nucleic Acids Res.* 41 (Database issue), D571–578.
- Walboomers, J.M., Jacobs, M.V., Manos, M.M., Bosch, F.X., Kummer, J.A., Shah, K.V., Snijders, P.J., Peto, J., Meijer, C.J., Munoz, N., 1999. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J. Pathol.* 189 (1), 12–19.
- Wu, C., Nilsson, K., Zheng, Y., Ekenstierna, C., Sugiyama, N., Forslund, O., Kajitani, N., Yu, H., Wennerberg, J., Ekblad, L., Schwartz, S., 2019. Short half-life of HPV16 E6 and E7 mRNAs sensitizes HPV16-positive tonsillar cancer cell line HN26 to DNA-damaging drugs. *Int. J. Cancer* 144 (2), 297–310.

Paper IV



**Identification of nucleotide substitutions in the 5'-end of
HPV16 early mRNAs and in the non-transcribed
long control region that affect
E6 and E7 mRNA splicing**

Xiaoxu Cui¹, Naoko Kajitani^{1,2} and Stefan Schwartz*^{1,2}

¹ Department of Laboratory Medicine, Lund University, BMC-B13, 221 84 Lund, Sweden.

² Department of Medical Biochemistry and Microbiology (IMBIM), Uppsala University, BMC-B9, 751 23 Uppsala, Sweden.

* To whom correspondence should be addressed.

Email: Stefan.Schwartz@med.lu.se

Running title: sequence variability of HPV16 affects HPV16 E6/E7 mRNA splicing

ABSTRACT

To investigate if sequence variation in HPV16 genomes could affect HPV16 mRNA splicing, we compared sequences of two HPV16 clones available to us. We identified multiple sequence differences between the two genomes but focused on variability in the long control region (LCR), the HPV16 early promoter and the first 880 nucleotides that are transcribed from the early promoter. We identified a set of three nucleotide positions within the first 570 nucleotides of the transcribed sequences that affected splicing of the HPV16 E6/E7 mRNAs. These were nucleotide positions 131, 350 and 570 (numbering according to the HPV16 reference genome HPV16R). With 131G, 350G and 570G rather than 131A, 350T and 570A, splicing of the HPV16 E6/E7 mRNAs was reduced. Our results suggested that at least one of these nucleotide positions affected a yet unidentified splicing regulatory RNA element within the first 570 nucleotides of the HPV16 E6/E7 mRNAs. Surprisingly, if transcription of the HPV16 E6/E7 genes occurred from plasmids with the full LCR (starting at HPV16 genomic position 7153) present upstream of the HPV16 transcription start site, rather than from plasmids with a shorter LCR starting at HPV16 genomic position 7470, the splicing efficiency was restored. Thus, a DNA sequence in the very 5'-end of the LCR compensated for the splicing reduction of HPV16 E6/E7 mRNAs caused by one or all of the three mutant positions within the first 570 nucleotides of the transcribed region. In summary, our results suggested that splicing regulatory elements may be present in both transcribed and non-transcribed regions of the HPV16 genome.

INTRODUCTION

Papillomaviruses are small DNA viruses with a diameter of approximately 50nm and a circular DNA genome of around eight kilobases. The life-cycle of HPV is strictly linked to cell differentiation in the squamous epithelium with expression of the HPV early genes in the lower and middle layers of the stratified epithelium (1-4). Essential early proteins include E6 and E7 that prevent apoptosis and induce proliferation of the HPV-infected cell (5,6). Simultaneous expression of the two HPV DNA replication factors E1 and E2 ensures replication of the HPV DNA genome (7-10) (7-10). The E6 and E7 proteins are notorious in the sense that they are responsible for induction and maintenance of cancer caused by a subset of the HPVs termed high-risk HPVs (HR-HPV) with tropism for mucosal epithelium (11). HPV16 is the most common high risk HPV type and is associated with various anogenital cancers as well as head and neck cancer (11,12).

High risk HPVs including HPV16 produce E6 and E7 from two alternatively spliced mRNAs generated from the same pre-mRNA in a mutually exclusive fashion (13). The E6 mRNA is generated by retention of an intron that encodes E6, while the E7 mRNA is generated by splicing and removal of the upstream E6-encoding intron. Splicing generates an mRNA on which the E7 open reading frame is preceded by a shorter form of E6 named E6*I that is less of an obstacle for E7-translation than the full length E6 ORF (14-16). The spliced mRNA named the E6*I or E7 mRNA is diagnostic for HR-HPV types (13) and is the major E7-producing mRNA (14-16). The major splice sites used for production of the spliced E6*/E7 mRNA are named SDS226 and SA409 in HPV16 (13). To maintain an optimal ratio between intron retention and splicing between SD226 and SA409 for production of sufficient quantities of both E6 and E7 mRNAs, this splicing event must be strictly controlled. Previous research has shown that some HPV16 splice sites are controlled by adjacent cis-acting RNA elements that either enhance or reduce splicing to a specific splice site, so called splicing enhancer and silencer elements (17-19). Splicing silencer RNA elements interacting with hnRNP A1 and hnRNP A2 have been identified (14,16). Furthermore, hnRNP D plays an important role in promoting intron retention in the E6-coding region to enhance production of E6 encoding mRNAs over spliced E7 mRNAs (20). Thus, RNA elements that interact with cellular RNA-binding proteins control the utilization efficiency of each splice site, thereby playing a major role in the regulated expression of the HPV genes at the level of RNA processing (21,22). Recent results have shown that the HPV16 mRNAs contain hot spots for RNA-binding proteins that are often located near HPV16 splice sites (21). Here we report that sequence polymorphism within the first 570 nucleotides of the HPV16 early mRNAs affected HPV16 mRNA splicing. Surprisingly, non-transcribed sequences in the LCR could compensate for the inefficient splicing of the E6/E7 mRNAs observed with some HPV16 sequences.

MATERIALS AND METHODS

Plasmids

pHPV16AN (23) and pHV16R (24) have been described previously. To construct HPV16 plasmid pLM16R in which transcription is under control of the LCR with the early P97 promoter from HPV16R, an HPV16 fragment containing the HPV16 LCR and the E6/E7 coding region was synthesized by Eurofins Genomics, digested with restriction enzymes HindIII and SbfI and subcloned into plasmid pC97ELsLuc to replace the CMV promoter and flanking sequences between HindIII and SbfI. To construct HPV16 plasmid pLM16AN in which transcription is under control of LCR with the early P97 promoter from HPV16AN, primers H3FM and SBFA ([Supplementary Table 1](#)) were used in combination to first amplify an HPV16 sequence containing LCR and E6/E7. pHPV16AN plasmid was digested with SphI, re-ligated and used as DNA template for PCR amplification. PCR products were digested with HindIII and SbfI and subcloned into pC97ELsLuc thereby replacing the CMV promoter and flanking sequences between HindIII and SbfI with HPV16 sequences. Hybrid plasmids pLM16(R+AN) and pLM16(AN+R) were generated by insertion of hybrid DNA fragments PCR-amplified from pLM16R and pLM16AN into pC97ELsLuc. Overlapping PCR primers named overlap S and overlap A ([Supplementary Table 1](#)) located between HPV16 nucleotide positions 37 and 76 were used in combination with primers H3FM and SBFA in a two-step PCR amplification reaction followed by subcloning into pC97ELsLuc with HindIII and SbfI as described above. To construct HPV16 subgenomic expression plasmid pLL16AN, primers H3F and SBFA ([Supplementary Table 1](#)) were used to amplify sequences containing the LCR, E6 and E7. To this end, pHPV16AN plasmid was digested with SphI, re-ligated and used as DNA template for PCR amplification. The resulting PCR products were digested with HindIII and SbfI and subcloned into pC97ELsLuc, thereby replacing the CMV promoter and flanking sequences between HindIII and SbfI. HPV16 subgenomic expression plasmids pLMS16AN1, pLMS16AN2, pLS16AN and pLSS16AN contain LCR/P97-promoter fragments of various length and were generated by PCR on pLL16AN using sense primer H3FMS1, H3FMS2, H3FSPH or 7620HS ([Supplementary Table 1](#)) (all containing a HindIII site) in combination with antisense primer SBFA and subcloned into pC97ELsLuc at HindIII and SbfI, as described above.

Cells

HeLa cells were cultured in Dulbecco's modified Eagle medium (DMEM)(HyClone) with 10% bovine calf serum (HyClone) and penicillin/streptomycin (Gibco).

Transfections

Transfections of HeLa cells were performed with Turbofect according to the manufacturer's protocol (Thermo Fisher Scientific). Briefly, a mixture of Turbofect and plasmid DNA with a Turbofect:DNA ratio of 3:1 (ul reagent:ug DNA) in DMEM without serum was incubated at room temperature for 20 min prior to dropwise addition to subconfluent for HeLa cells.

RNA extraction and RT-PCR

Total RNA was extracted from transfected cells using TRI Reagent (Sigma Aldrich) and Directzol RNA MiniPrep (ZYMO Research) according to the manufacturer's protocols. Reverse transcription (RT) was performed in a 20 ul reaction using random hexamers (Invitrogen) and reverse transcriptase (Invitrogen). One ul of cDNA was subjected to PCR amplification. cDNA representing HPV16 mRNAs spliced from HPV16 5'-splice site SD226 to 3'-splice sites SA409, SA526 or SA742 were monitored by RT-PCR with primers 97S and 880AS ([Supplementary Table 1](#)). Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) cDNA was amplified with primers GAPDHF and GAPDHR ([Supplementary Table 1](#)). To monitor recombination at the loxP sites in pHPV16AN, PCR was performed with primers 16S and 16A ([Supplementary Table 1](#)) on DNA extracted from the transfected cells (this PCR yields a 366-nucleotide PCR fragment that is diagnostic for recombination at the LoxP sites). Examples of control PCR experiments performed on RNA samples in the absence of reverse transcriptase are shown in various figures. Primer pairs used for RT-PCR are summarised in [Supplementary Table 1](#).

Sequence alignment

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

RESULTS AND DISCUSSION

HPV16 nucleotide positions 131, 350 and/or 570 in the 5'-end of the early mRNAs affect HPV16 E6/E7 mRNA splicing

The HPV16 genome may be divided into three regions. The early region encoding E1, E2, E4, E5, E6 and E7, the late region encoding L1 and L2 and the non-protein coding long control region (LCR) located between the L1 stop codon and the E6 start codon (Fig. 1A and B). A sequence comparison between two molecular clones of HPV16, HPV16AN and HPV16R, revealed multiple sequence differences (Supplementary Fig. 1A). We focused on nucleotide positions in the long control region (LCR) and the 5'-end of the HPV16 early mRNAs as they could potentially affect transcription and/or splicing of the E6- and E7-oncogene encoding mRNAs (Fig. 1B and C). We identified sequence variability at positions 7447, 7528, 7612, 12 and 13 in the non-transcribed region of the LCR or positions 131, 350 and 570 in the first 880 nucleotides of the HPV16 early mRNAs (Fig. 2A and B) (numbering refers to the HPV16 reference genome HPV16R) (13). To investigate if the sequence heterogeneity affected HPV16 E6/E7 mRNA splicing, we generated HPV16 plasmids under control of the LCR from either HPV16AN or HPV16R (Fig. 3A). In essence, the CMV promoter of the HPV16 subgenomic plasmid pC97EL was replaced with the HPV16 LCR and the early promoter derived from either the HPV16R or the HPV16AN sequence resulting in plasmids pLM16AN and pLM16R, respectively (Fig. 3A). Transfection of these plasmids into HeLa cells followed by RNA extraction and RT-PCR of the HPV16 E6 and E7 mRNAs using primers that detected the various alternatively spliced HPV16 E6/E7 mRNAs was performed (for RT-PCR primer locations, see Figure 1C). The results revealed that splicing of HPV16 E6/E7 mRNAs derived from HPV16AN were inefficient compared to mRNAs generated from HPV16R derived plasmids (Fig. 3B). Since they differed at eight nucleotide positions only, five in the non-transcribed part of the LCR and three in the transcribed HPV16 region downstream of the initiation site of the early promoter P97 at nucleotide position 97, we next generated two hybrid plasmids, one containing HPV16AN-sequences upstream of promoter P97 and HPV16R downstream of P97, and one containing HPV16R-sequences upstream of promoter P97 and HPV16AN downstream of P97 (Fig. 4A). Transfection of these plasmids into HeLa cells followed by RNA extraction and RT-PCR of the HPV16 E6 and E7 mRNAs revealed that HPV16R-sequences downstream of HPV16 early promoter P97 conferred efficient splicing to the HPV16 E6/E7 mRNAs (Fig. 4B). Thus, a "G" at either or all nucleotide positions 131, 350 or 570, as in HPV16AN, reduced splicing efficiency compared with the 131A, 350T or 570A, as in HPV16R (Fig. 4A and B). These results suggested that the nucleotide positions 131, 350 or 570 overlapped cis-acting elements that controlled HPV16 early mRNA splicing. However, since these positions were all located downstream of the transcriptional start site of the HPV16 early promoter, they could potentially be active as either RNA or DNA elements.

Since these sequences are transcribed and the region contains splice sites in the vicinity of the mutations, it is reasonable to speculate that at least one of these mutations are affecting positive or negative, cis-acting splicing regulatory RNA elements termed splicing enhancers or splicing silencers, respectively (25). Such sequences bind to cellular RNA-binding proteins, e.g. serine and arginine rich (SR) proteins or heterogeneous nuclear ribonucleoproteins (hnRNP) that interact with each other or with the cellular splicing machinery, either to enhance splicing or inhibit splicing (26,27). Cellular RNA-binding proteins play a major role in the control of HPV16 gene expression (22). Indeed, cis-acting splicing silencer elements have been identified in this region of HPV16 (16) and in HPV18 (14). Sequences located between HPV16 positions 594 and 604 interact with hnRNP A1 and A2 to inhibit splicing between major HPV16 splice sites SD226 and SA409 (16). Thus, hnRNP A1 and A2 inhibit production of the major HPV16 E7 encoding mRNAs (226-409-splicing (Fig. 1C)). Furthermore, hnRNP D appears to interact with the intronic sequences between SD226 and SA409, thereby inhibiting splicing and promoting production of intron retained, E6-encoding mRNAs (20). The phenotype of the 131A, 350T or 570A (HPV16R) to 131G, 350G or 570G (HPV16AN) substitutions indicate that cis-acting splicing regulatory RNA elements may be present on at least one of these locations. Presumably, such elements would be splicing enhancer elements rather than splicing silencer elements. Further experiments are required to determine if all or a subset of the substitutions at HPV16 nucleotide positions 131, 350 and 570 are required for splicing inhibition.

Non-transcribed HPV16 sequences in the HPV16 long control region compensate for impaired splicing caused by a G-nucleotide in positions 131, 350 and/or 570

To determine if E6/E7 mRNAs produced from pHPV16AN were poorly spliced also in the context of the full-length genome, plasmid pHPV16AN (Fig. 5A) was used. Plasmid pHPV16AN has a loxP sites inserted at either site of the unique Sph1 site (nucleotide position 7470) in the LCR of the HPV16 genome (Fig. 5A) and may therefore be released into an episomal form of the HPV16 genome by cotransfection with a cre-recombinase expressing plasmid (Fig. 5A). As can be seen, transfection of pHPV16AN alone, yielded poor splicing (Fig. 5B), reminiscent of the results displayed above (Fig. 3B). Surprisingly, cotransfection of pHPV16 with the pCRE plasmid that produces the cre-recombinase that releases the episomal form of the HPV16 genome from pHPV16AN, resulted in production of HPV16 E6/E7 mRNAs that were efficiently spliced (Fig. 5B). Since the loxP sites had been inserted in position 7470 in the HPV16 LCR, excision of the HPV16 genome to form the episomal form of the HPV16 genome, also brought together LCR sequences from each side of position 7470 in the LCR of the HPV16 genome (Fig. 5A). Thus, it appeared that excision of the plasmid vector that had been inserted into the Sph1 site of the HPV16 genome in plasmid pHPV16AN, brought together

LCR sequences from upstream of position 7470 with the HPV16 P97 early promoter and start of transcription. This rearrangement of the HPV16 LCR affected splicing of the HPV16 early E6/E7 mRNAs. These results suggested that non-transcribed LCR sequences upstream of position 7470 could affect splicing of HPV16 early E6/E7 mRNAs. Furthermore, reconstitution of the complete LCR also compensated for the poor splicing of the HPV16 early E6/E7 mRNAs conferred by the “G”-nucleotide in positions 131, 350 and/or 570.

To confirm this finding, we generated HPV16 subgenomic expression plasmids driven by HPV16 sequences of various length (Fig. 6A). They were generated by replacing the CMV promoter in the HPV16 subgenomic expression plasmid pC97EL with LCR sequences of various length from the poorly spliced pHPV16AN (Fig. 6A). Transfection of these plasmids into HeLa cells followed by RNA extraction and RT-PCR of the HPV16 E6 and E7 mRNAs revealed that the HPV16-plasmid with the long LCR (pLL), that included sequences upstream of HPV16 position 7470 that extended to the L1 stop codon at position 7156 were efficiently spliced (226⁴⁰⁹ mRNAs dominate) (Fig. 6B), whereas a plasmid with the shorter form of the LCR which started at the Sph1 site at HPV16 nucleotide position 7470 (pLS), produced poorly spliced E6/E7 mRNAs (unspliced (U) mRNAs dominate) (Fig. 6B). Plasmid pLM, that contained LCR sequences that started between pLL and pLS, at HPV16 nucleotide position 7321, also produced poorly spliced HPV16 mRNAs, although less so than pLL (Fig. 6B). These observations were confirmed with a bigger set of LCR deletion mutants (Fig. 6C). We observed that transcription from plasmids driven by the HPV16 P97 promoter was inherently inefficient and required transfection with relatively high amounts of plasmid DNA (Fig. 6D). These results confirmed that non-transcribed HPV16 LCR sequences upstream of position 7470 affected splicing of HPV16 early E6/E7 mRNAs.

DNA sequences in the promoter-encoding region of genes could potentially affect splicing by the recruitment of DNA-binding transcription factors that execute their function by recruiting RNA-binding proteins to the DNA-dependent RNA polymerase, thereby loading the polymerase with RNA-binding proteins that may swiftly interact with RNA elements on de novo synthesized mRNAs (28). Thus, a promoter sequence could potentially affect mRNA splicing. It is well established that transcription and mRNA splicing are coupled processes (29). Indeed, it has been shown that HPV18 mRNA splicing may be affected by long-range interactions of DNA-binding proteins that interact with the HPV-genome (30). The LCR is the most highly variable region of the HPV genome and contains multiple regulatory DNA elements (31). Sequence polymorphism within HPV16 may therefore potentially affect the function of the LCR, including transcription and mRNA processing. In addition, the LCR is also very different between HPV types. For example, LCR of mucosal and cutaneous HPV types differ

significantly with respect to LCR length (13). All the results combined, further studies are warranted to elucidate the interactions between the HPV16 non-transcribed LCR region or the 5'-end of the HPV16 early mRNAs and splicing control of the HPV16 mRNAs encoding the E6 and E7 oncogenes.

ACKNOWLEDGEMENT

We are grateful to members of the research group for discussions.

FUNDING

This work was supported by the Swedish Research Council-Medicine [VR2019-01210 to S.S.]; the Swedish Cancer Society [CAN2018/702 to S.S.]; and the China Scholarship Council [201809120016 to X.C.].

REFERENCES

1. Kajitani, N., Satsuka, A., Kawate, A. and Sakai, H. (2012) Productive Lifecycle of Human Papillomaviruses that Depends Upon Squamous Epithelial Differentiation. *Front Microbiol*, **3**, 152.
2. Hong, S. and Laimins, L.A. (2013) Regulation of the life cycle of HPVs by differentiation and the DNA damage response. *Future Microbiol*, **8**, 1547-1557.
3. Chow, L.T., Broker, T.R. and Steinberg, B.M. (2010) The natural history of human papillomavirus infections of the mucosal epithelia. *APMIS*, **118**, 422-449.
4. Schiffman, M., Doorbar, J., Wentzensen, N., de Sanjose, S., Fakhry, C., Monk, B.J., Stanley, M.A. and Franceschi, S. (2016) Carcinogenic human papillomavirus infection. *Nat Rev Dis Primers*, **2**, 16086.
5. Roman, A. and Munger, K. (2013) The papillomavirus E7 proteins. *Virology*, **445**, 138-168.
6. Vande Pol, S.B. and Klingelutz, A.J. (2013) Papillomavirus E6 oncoproteins. *Virology*, **445**, 115-137.
7. Bergvall, M., Melendy, T. and Archambault, J. (2013) The E1 proteins. *Virology*, **445**, 35-56.
8. Kadaja, M., Silla, T., Ustav, E. and Ustav, M. (2009) Papillomavirus DNA replication - from initiation to genomic instability. *Virology*, **384**, 360-368.
9. McBride, A.A. (2013) The papillomavirus E2 proteins. *Virology*, **445**, 57-79.
10. Thierry, F. (2009) Transcriptional regulation of the papillomavirus oncogenes by cellular and viral transcription factors in cervical carcinoma. *Virology*, **384**, 375-379.
11. Walboomers, J.M., Jacobs, M.V., Manos, M.M., Bosch, F.X., Kummer, J.A., Shah, K.V., Snijders, P.J., Peto, J., Meijer, C.J. and Munoz, N. (1999) Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J Pathol*, **189**, 12-19.
12. Bouvard, V., Baan, R., Straif, K., Grosse, Y., Secretan, B., El Ghissassi, F., Benbrahim-Tallaa, L., Guha, N., Freeman, C., Galichet, L. et al. (2009) A review of human carcinogens--Part B: biological agents. *Lancet Oncol.* , **10**, 321-322. .
13. Van Doorslaer, K., Tan, Q., Xirasagar, S., Bandaru, S., Gopalan, V., Mohamoud, Y., Huyen, Y. and McBride, A.A. (2013) The Papillomavirus Episteme: a central resource for papillomavirus sequence data and analysis. *Nucleic Acids Res*, **41**, D571-578.
14. Ajiro, M., Tang, S., Doorbar, J. and Zheng, Z.M. (2016) Serine/Arginine-Rich Splicing Factor 3 and Heterogeneous Nuclear Ribonucleoprotein A1 Regulate Alternative RNA Splicing and Gene Expression of Human Papillomavirus 18 through Two Functionally Distinguishable cis Elements. *J Virol*, **90**, 9138-9152.
15. Tang, S., Tao, M., McCoy, J.P., Jr. and Zheng, Z.M. (2006) The E7 oncoprotein is translated from spliced E6*1 transcripts in high-risk human papillomavirus type 16- or

- type 18-positive cervical cancer cell lines via translation reinitiation. *J Virol*, **80**, 4249-4263.
16. Zheng, Y., Jonsson, J., Hao, C., Shoja Chaghervand, S., Cui, X., Kajitani, N., Gong, L., Wu, C. and Schwartz, S. (2020) Heterogeneous Nuclear Ribonucleoprotein A1 (hnRNP A1) and hnRNP A2 Inhibit Splicing to Human Papillomavirus 16 Splice Site SA409 through a UAG-Containing Sequence in the E7 Coding Region. *J Virol*, **94**.
 17. Johansson, C. and Schwartz, S. (2013) Regulation of human papillomavirus gene expression by splicing and polyadenylation. *Nature Rev. Microbiol.*, **11**, 239-251.
 18. Jia, R. and Zheng, Z.M. (2009) Regulation of bovine papillomavirus type 1 gene expression by RNA processing. *Front. Biosci.*, **14**, 1270-1282.
 19. Graham, S.V. and Faizo, A.A. (2017) Control of human papillomavirus gene expression by alternative splicing. *Virus Res*, **231**, 83-95.
 20. Cui, X., Hao, C., Gong, L., Kajitani, N. and Schwartz, S. (2022) hnRNP D activates production of HPV16 E1 and E6 mRNAs by promoting intron retention. *Nucleic Acids Res*, **in press**.
 21. Kajitani, N. and Schwartz, S. (2020) Role of Viral Ribonucleoproteins in Human Papillomavirus Type 16 Gene Expression. *Viruses*, **12**.
 22. Kajitani, N. and Schwartz, S. (2022) The role of RNA-binding proteins in the processing of mRNAs produced by carcinogenic papillomaviruses. *Semin Cancer Biol*, **in press**.
 23. Li, X., Johansson, C., Glahder, J., Mossberg, A.K. and Schwartz, S. (2013) Suppression of HPV-16 late L1 5'-splice site SD3632 by binding of hnRNP D proteins and hnRNP A2/B1 to upstream AUAGUA RNA motifs. *Nucleic Acids Res*, **22**, 10488-10508.
 24. Durst, M., Gissmann, L., Ikenberg, H. and zur Hausen, H. (1983) A papillomavirus DNA from a cervical carcinoma and its prevalence in cancer biopsy samples from different geographic regions. *Proc Natl Acad Sci U S A*, **80**, 3812-3815.
 25. Cartegni, L., Chew, S.L. and Krainer, A.R. (2002) Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet*, **3**, 285-298.
 26. Shepard, P.J. and Hertel, K.J. (2009) The SR protein family. *Genome Biol*, **10**, 242.
 27. Han, S.P., Tang, Y.H. and Smith, R. (2010) Functional diversity of the hnRNPs: past, present and perspectives. *Biochem J*, **430**, 379-392.
 28. Kolathur, K.K. (2021) Role of promoters in regulating alternative splicing. *Gene*, **782**, 145523.
 29. Tellier, M., Maudlin, I. and Murphy, S. (2020) Transcription and splicing: A two-way street. *Wiley Interdiscip Rev RNA*, **11**, e1593.

30. Ferguson, J., Campos-Leon, K., Pentland, I., Stockton, J.D., Gunther, T., Beggs, A.D., Grundhoff, A., Roberts, S., Noyvert, B. and Parish, J.L. (2021) The chromatin insulator CTCF regulates HPV18 transcript splicing and differentiation-dependent late gene expression. *PLoS Pathog*, **17**, e1010032.
31. Bernard, H.U. (2013) Regulatory elements in the viral genome. *Virology* . , **445**, 197-204.

FIGURE LEGENDS.**Figure 1. Schematic representation of the HPV16 genome, long control region and early E6/E7 encoding mRNAs.**

(A) Linearized HPV16 genome (numbers refer to the HPV16 reference strain GeneBank: K02718.1) (HPV16R). Early and late genes are indicated. P97: HPV16 early promoter. P670: HPV16 late promoter. Black oval: splice donor. White oval: splice acceptor. pAE: HPV16 early polyadenylation site. pAL: HPV16 late polyadenylation site. LCR: HPV16 long control region. (B) Schematic structures of long control region (LCR) and HPV16 E6/E7-encoding sequences up to HPV16 5'-splice site SD880. (C) Schematic representation of alternatively spliced HPV16 E6 and E7 encoding mRNAs. Splicing at SD226 generates splice variants in the E6/E7-coding region. Arrows indicate HPV16 RT-PCR primers 97S and 880A. Primer sequences are available in [Supplementary Table 1](#).

Figure 2. Sequence variability from the non-transcribed region of the LCR to HPV16 E6/E7 coding region between HPV16AN and HPV16R.

(A) Alignment of HPV16AN and HPV16R sequences spanning the LCR and the E6/E7 coding region up to HPV16 nucleotide 880. Nucleotide variability is indicated by red highlights. (B) Sequence variability at different positions from the non-transcribed region of the LCR to the HPV16 E6/E7 coding region is indicated by black triangles, "-" indicates nucleotide deletion, numbers refer to the HPV16 reference strain GeneBank: K02718.1 (HPV16R).

Figure 3. HPV16 sequence variation in LCR and E6/E7 coding region affect HPV16 E6/E7 mRNA splicing

(A) HPV16 subgenomic plasmid pC97EL encodes all HPV16 genes. HPV16 early promoter P97 is replaced by human cytomegalovirus immediate early promoter (CMV) (pC97EL). pL97EL changes CMV back to LCR derived from either HPV16AN or HPV16R. (B) Sequence variation in LCR/P97-sequences and E6/E7 coding region affect HPV16 E6/E7 mRNA splicing. HPV16 E6/E7 alternative mRNA splicing was investigated by performing RT-PCR on RNA extracted from HeLa cells transfected with indicated plasmids in duplicates. RT-PCR primer pair: 97S+880AS. RT-PCR was validated by reverse transcriptase negative samples using corresponding RNA samples. Only RT positive samples for gapdh RT-PCR are shown. Primer sequences are available in [Supplementary Table 1](#). M: DNA size marker.

Figure 4. HPV16R sequences downstream of HPV16 early promoter P97 conferred efficient HPV16 mRNA splicing

(A) Schematic representation of HPV16 genomic region between the L1 stop codon and the 5'-splice site SD880. HPV16 sequences derived from HPV16R, HPV16AN or hybrids thereof

are displayed. Sequence variability at different positions is indicated by black triangles, “-” indicates nucleotides deletion. Nucleotides derived from HPV16AN are shown in green, while nucleotides derived from HPV16R are shown in red. **(B)** HPV16R sequences downstream of HPV16 early promoter P97 conferred efficient splicing. HPV16 E6/E7 alternative mRNA splicing was investigated by RT-PCR on RNA extracted from HeLa cells transfected with indicated plasmids. RT-PCR primer pairs: 97S+880AS. gapdh RT-PCR is shown as internal control. Primer sequences are available in [Supplementary Table 1](#). M: DNA size marker.

Figure 5. Non-transcribed HPV16 sequences in the HPV16 long control region compensate for maimed splicing in full-length genomic plasmid pHPV16AN.

(A) Schematic representation of the pHPV16AN plasmid and episomal HPV16 DNA production using Cre-loxP transfection system. LoxP sites located at the unique SphI site (nucleotide position 7470 in the LCR of the HPV16 genome) are flanked by the plasmid backbone. HPV16 genome is released into an episomal form by cotransfection with a cre-recombinase expressing plasmid (pCAGSS-nlscre). **(B)** RT-PCR on RNA extracted from HeLa cells transfected with pHPV16AN with or without Cre recombinase expressing plasmid. RNA was extracted and HPV16 RNA splicing was monitored by RT-PCR using primers 97S+880AS. PCR with primers 16S and 16A on DNA extracted from the transfected cells yields a shorter PCR fragment that is diagnostic for recombination at the LoxP sites(+cre) in plasmid pHPV16AN. A larger band is amplified from plasmid DNA that has not recombined (-cre). gapdh RT-PCR are shown as internal control. Primer sequences are available in [Supplementary Table 1](#). M: DNA size marker.

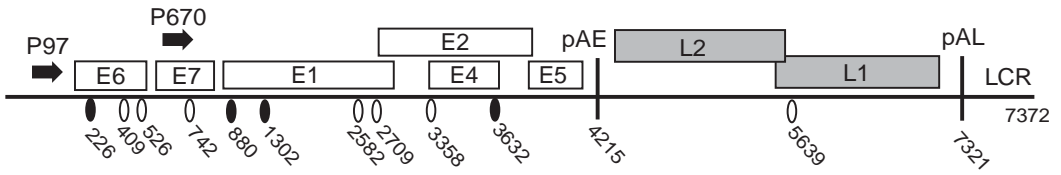
Figure 6. Non-transcribed HPV16 LCR sequences upstream of HPV16 nucleotide position 7470 affect splicing of HPV16 early E6/E7 mRNAs

(A) Schematic representation of HPV16 subgenomic expression plasmids in which transcription is driven by HPV16 sequences of various length. HPV sequences in the subgenomic plasmids pLL16AN, pLM16AN, pLMS16AN1, pLMS16AN2, pLS16AN and pLSS16AN start at HPV16 nucleotide positions 7156, 7321, 7366, 7407, 7465 and 7620, respectively. **(B)** HPV16 E6/E7 alternative mRNA splicing was investigated by RT-PCR on RNA extracted from HeLa cells transfected with plasmids pLL16AN, pLM16AN and pLS16AN. RT-PCR primer pair: 97S+880AS. gapdh RT-PCR is shown as internal control. Primer sequences are available in [Supplementary Table 1](#). M: DNA size marker. **(C)** HPV16 E6/E7 alternative mRNA splicing was investigated by RT-PCR on RNA extracted from HeLa cells transfected with plasmids pLL16AN, pLM16AN, pLMS16AN1, pLMS16AN2, pLS16AN and pLSS16AN in duplicates. RT-PCR primer pairs: 97S+880AS. gapdh RT-PCR is shown as internal control. Primer sequences are available in [Supplementary Table 1](#). M: DNA size

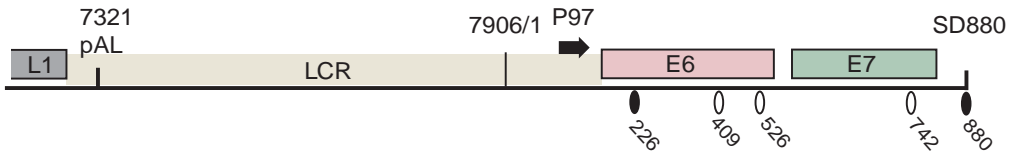
marker. **(D)** HPV16 E6/E7 alternative mRNA splicing was investigated by RT-PCR on RNA extracted from HeLa cells transfected with pLL16AN, pLM16AN and pLS16AN plasmids using different amounts of plasmid DNA for transfection as indicated. RT-PCR primer pairs: 97S+880AS. gapdh RT-PCR are shown for normalization. Primer sequences are available in [Supplementary Table 1](#). M: DNA size marker.

A HPV-16 genome

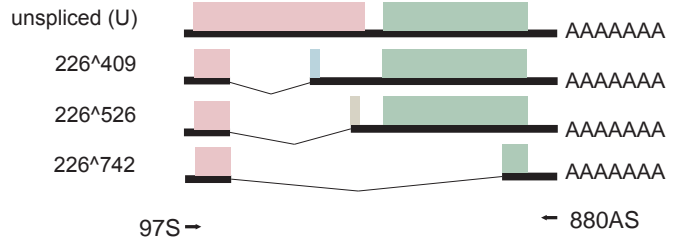
Figure 1



B



C

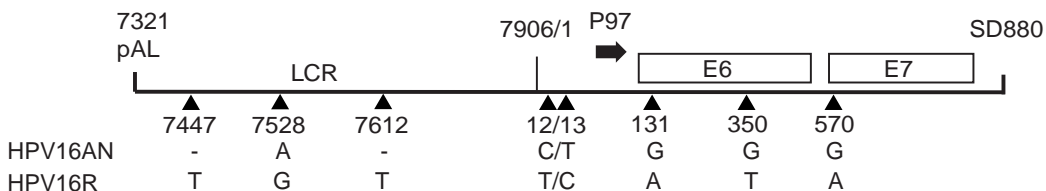


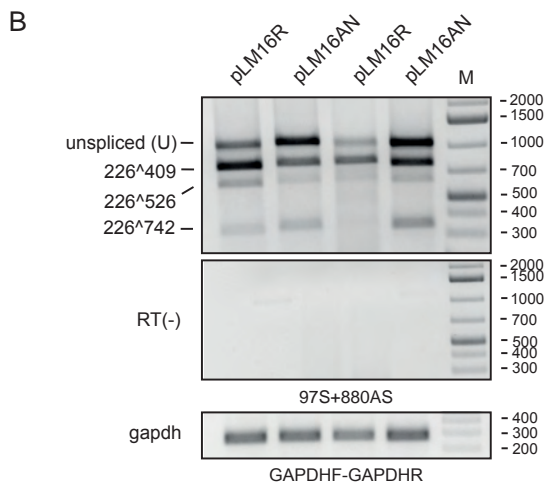
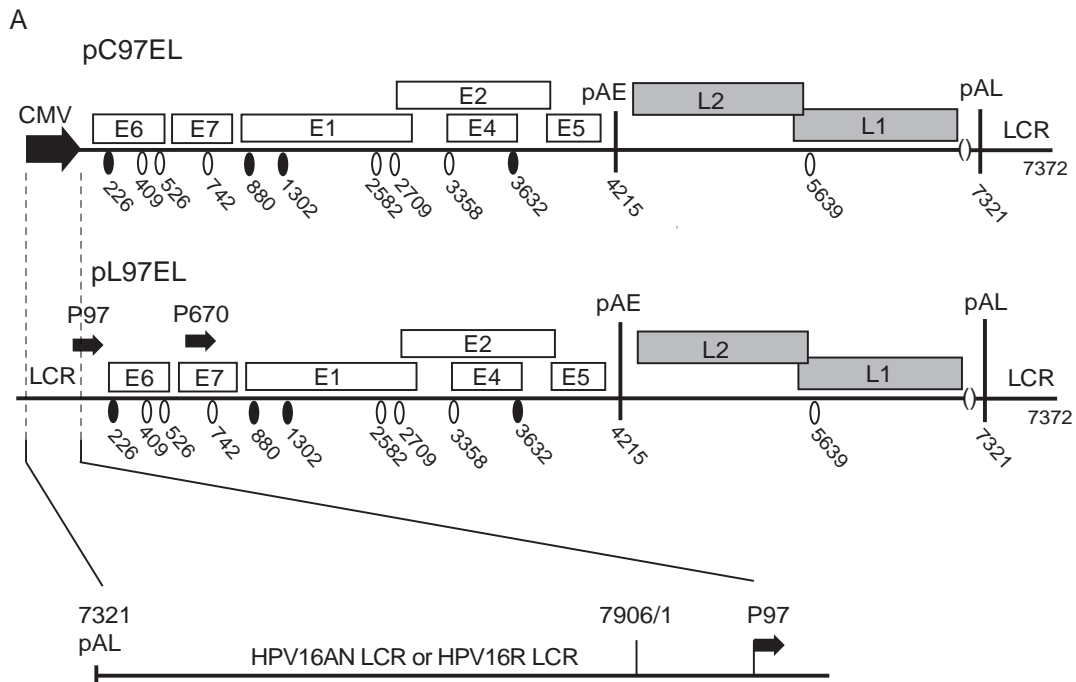
A

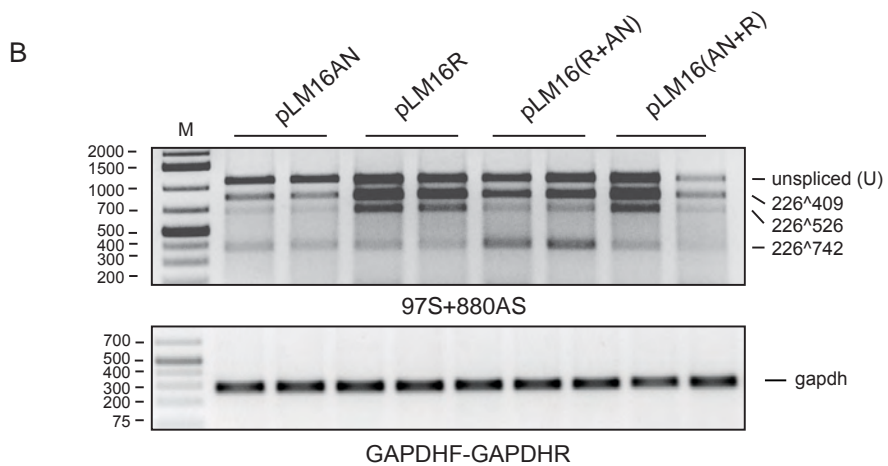
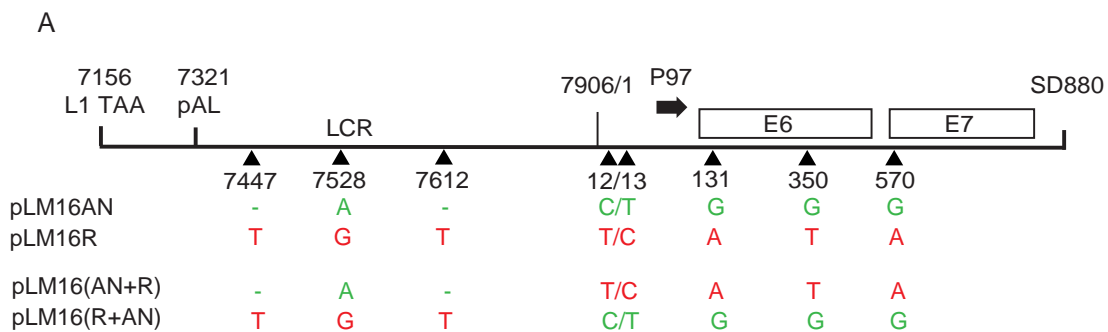
```

HPV16AN 7260 TATGGTATAATAAACACGCTGTGTATGTTTTAAATGCTTGTGTAACATTTGTGTCATGCAACATAAAATAAAGCTTATTGTTTCAACACCTACTAATTGTGTGGTTA 7369
HPV16R 7254 TATGGTATAATAAACACGCTGTGTATGTTTTAAATGCTTGTGTAACATTTGTGTCATGCAACATAAAATAAAGCTTATTGTTTCAACACCTACTAATTGTGTGGTTA 7363
HPV16AN 7370 TTCATTGTATATAAACTATATTTGCTACATCCTGTTTTGTTTTATATATACTATATTTTGTAGCGCCAGCGGCCA-TTTGTAGCTTCAACCGAATTCCGTTGCATGCTT 7478
HPV16R 7364 TTCATTGTATATAAACTATATTTGCTACATCCTGTTTTGTTTTATATATACTATATTTTGTAGCGCCAGCGGCCA-TTTGTAGCTTCAACCGAATTCCGTTGCATGCTT 7473
HPV16AN 7479 TTTGGCACAAAATGTGTTTTTTTTAAATAGTTCTATGTCAGCAACTATGTTTTAAACTTGTACGTTTTCTGCTTGGCATGGGTGCCAAATCCCTGTTTTCTGACCTGCAC 7588
HPV16R 7474 TTTGGCACAAAATGTGTTTTTTTTAAATAGTTCTATGTCAGCAACTATGTTTTAAACTTGTACGTTTTCTGCTTGGCATGGGTGCCAAATCCCTGTTTTCTGACCTGCAC 7583
HPV16AN 7589 TGCTTGCCAACCATTCACCTG-TTTTTACACTGCACTATGTGCAACTACTGAATCACTATGTACATTTGTGCATATAAAATAAATCACTATGGCCAAACCGCTTACATAC 7697
HPV16R 7584 TGCTTGCCAACCATTCACCTG-TTTTTACACTGCACTATGTGCAACTACTGAATCACTATGTACATTTGTGCATATAAAATAAATCACTATGGCCAAACCGCTTACATAC 7693
HPV16AN 7698 CGCTGTAGGCACATATTTTTGGCTTGTTTAACTAACCTAATTGCATATTTGGCATAAAGTTTTAAACTTCTAAGGCCAACTAAATGTCACCCTAGTTACATACGAAC 7807
HPV16R 7694 CGCTGTAGGCACATATTTTTGGCTTGTTTAACTAACCTAATTGCATATTTGGCATAAAGTTTTAAACTTCTAAGGCCAACTAAATGTCACCCTAGTTACATACGAAC 7803
HPV16AN 7808 GTGTAAGGTTAGTCATACATTTGTCATTTGTA AAACTGCACATGGGTGTGTCGCAACCGTTTTGGGTTACACATTTACAAGCAACTATATAATAATACTAA 7910
HPV16R 7804 GTGTAAGGTTAGTCATACATTTGTCATTTGTA AAACTGCACATGGGTGTGTCGCAACCGTTTTGGGTTACACATTTACAAGCAACTATATAATAATACTAA 7906
HPV16AN 1ACTACAATAATCTATGTA AAACTAAGCGCTAACCGAAATCGGTTGAACCGAAACCGGTTAGTATAAAGCAGACATTTTGCACCAAAAAGAGAAGCTCCAATGTTTC 110
HPV16R 1ACTACAATAATCTATGTA AAACTAAGCGCTAACCGAAATCGGTTGAACCGAAACCGGTTAGTATAAAGCAGACATTTTGCACCAAAAAGAGAAGCTCCAATGTTTC 110
HPV16AN 111 AGGACCCACAGGAGCGACCCGAAAAGTTACACAGTTATGCACAGAGCTGCAAAACACTATACATGATATAAATATTAGAATGTGTACTGCAAGCAACAGTTACTGCGA 220
HPV16R 111 AGGACCCACAGGAGCGACCCGAAAAGTTACACAGTTATGCACAGAGCTGCAAAACACTATACATGATATAAATATTAGAATGTGTACTGCAAGCAACAGTTACTGCGA 220
HPV16AN 221 CGTGAGGTATATGACTTTGCTTTTCGGGATTTATGCATAGTATATAGAGATGGGAATCCATATGCTGTATGTGATAAATGTTAAAGTTTTATTCTAAAATTAGTGAGTA 330
HPV16R 221 CGTGAGGTATATGACTTTGCTTTTCGGGATTTATGCATAGTATATAGAGATGGGAATCCATATGCTGTATGTGATAAATGTTAAAGTTTTATTCTAAAATTAGTGAGTA 330
HPV16AN 331 TAGACATATTTGTTATAGTTGTATGGAACAACATAGAACAGCAATACAACAAACCGTTGCTGATTTGTTAATTAGGTGATTAAGTGTCAAAAAGCCACTGTGTCCTG 440
HPV16R 331 TAGACATATTTGTTATAGTTGTATGGAACAACATAGAACAGCAATACAACAAACCGTTGCTGATTTGTTAATTAGGTGATTAAGTGTCAAAAAGCCACTGTGTCCTG 440
HPV16AN 441 AAGAAAAGCAAAAGACATCTGGACAAAAGCAAAGATTCCATAAATATAAGGGGTCGGTGGACCGGTCGATGTATGCTTGTGGCAGATCATCAAGAACACGTAGAGAAACC 550
HPV16R 441 AAGAAAAGCAAAAGACATCTGGACAAAAGCAAAGATTCCATAAATATAAGGGGTCGGTGGACCGGTCGATGTATGCTTGTGGCAGATCATCAAGAACACGTAGAGAAACC 550
HPV16AN 551 CAGCTGAATCATCCATGGCATAACCTACATTCATGAATATATGTTAGATTTGCAACAGACAACTGATCTCTACTGTTATGAGCAATTAATGACAGCTCAGAG 660
HPV16R 551 CAGCTGAATCATCCATGGCATAACCTACATTCATGAATATATGTTAGATTTGCAACAGACAACTGATCTCTACTGTTATGAGCAATTAATGACAGCTCAGAG 660
HPV16AN 661 GAGGAGGATGAAATAGATGGTCCAGCTGGACAAGCAGAACCGGACAGGCCATTACAATATTGTAACCTTTTTGTTGCAAGTGTGACTCTACGCTTCGGTTGTCGGTACA 770
HPV16R 661 GAGGAGGATGAAATAGATGGTCCAGCTGGACAAGCAGAACCGGACAGGCCATTACAATATTGTAACCTTTTTGTTGCAAGTGTGACTCTACGCTTCGGTTGTCGGTACA 770
HPV16AN 771 AAGCACACACGTAGACATTCGTACTTTGGAAGACCTGTTAATGGGCACACTAGGAATTTGTGCCCCATCTGTTCTCAGAAAACCAATACTACCATGGCTGATCCTGCAG 880
HPV16R 771 AAGCACACACGTAGACATTCGTACTTTGGAAGACCTGTTAATGGGCACACTAGGAATTTGTGCCCCATCTGTTCTCAGAAAACCAATACTACCATGGCTGATCCTGCAG 880
    
```

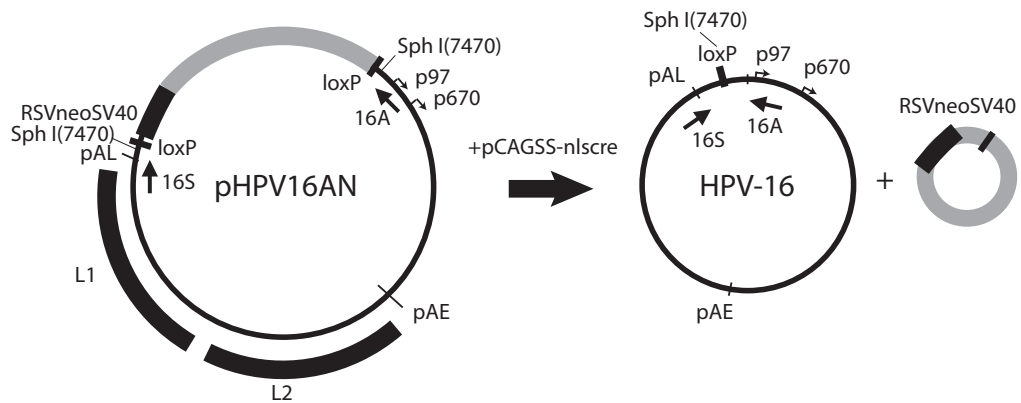
B



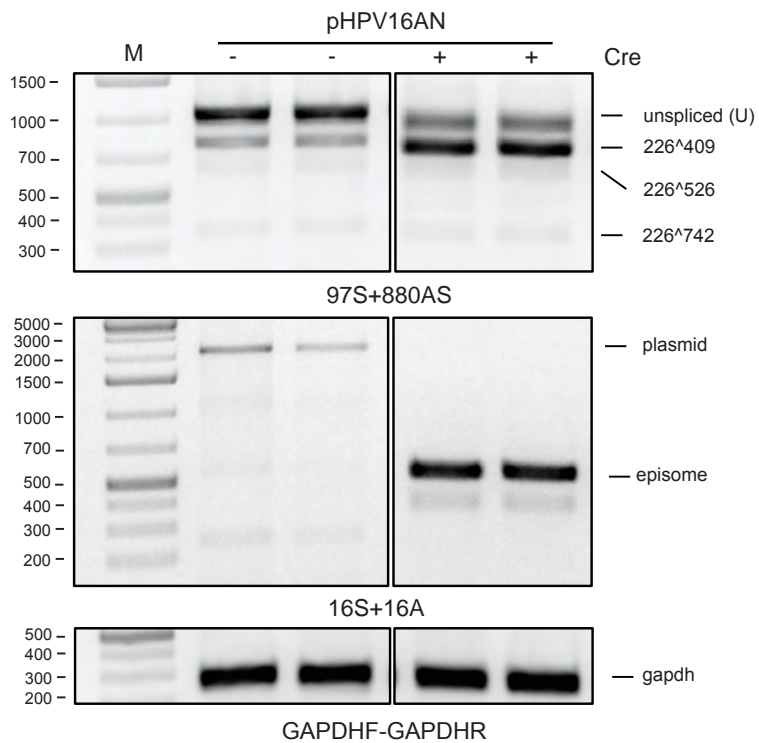




A

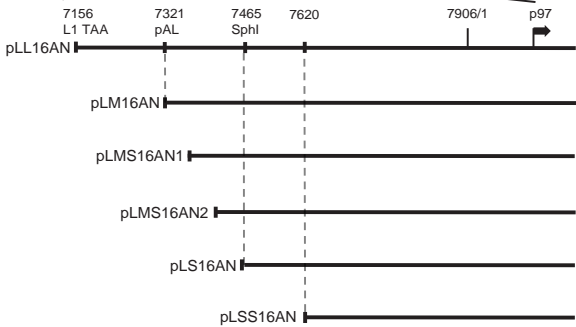
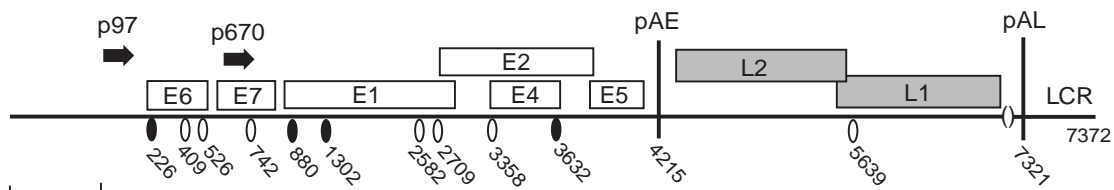


B

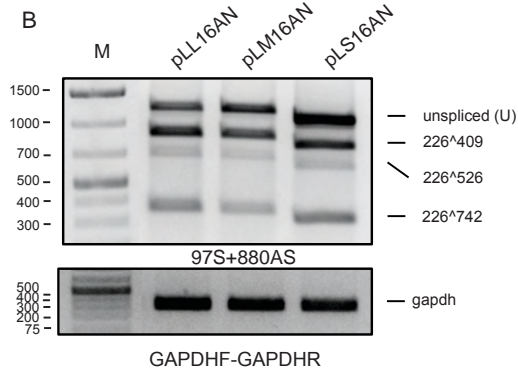


A

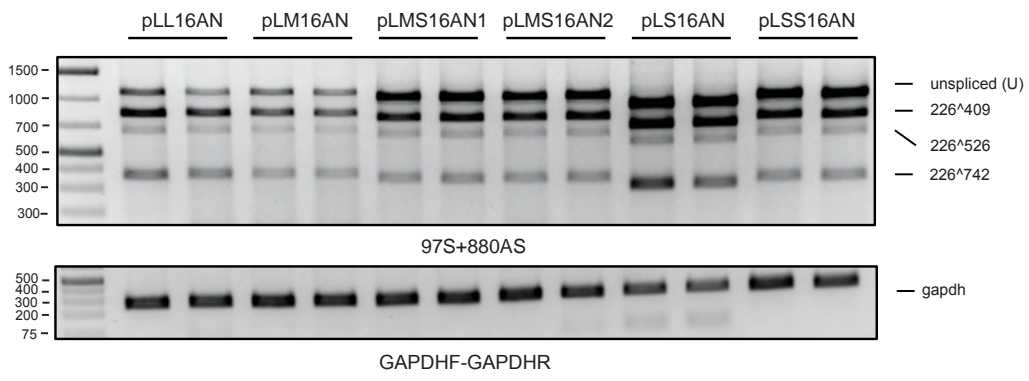
pL97EL



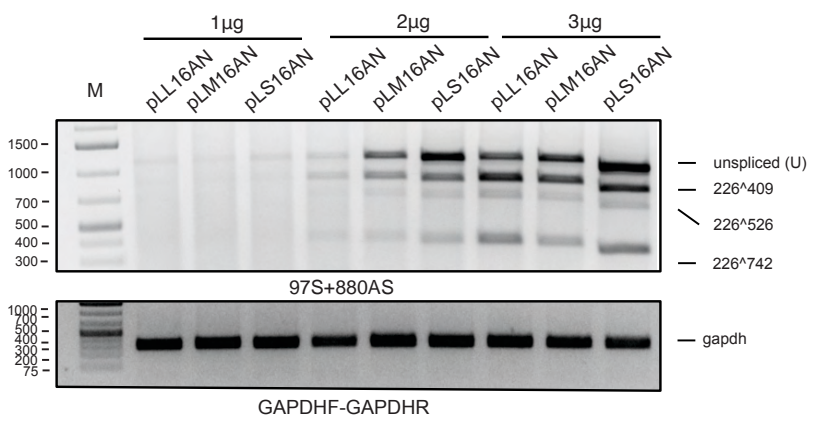
B



C



D



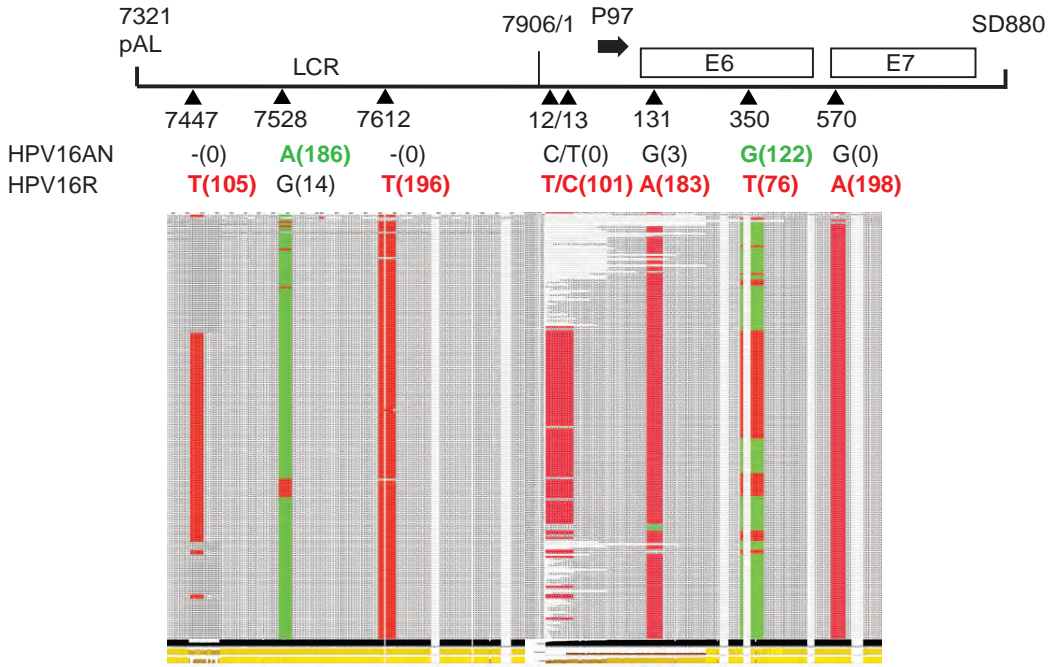
Supplementary Table 1. Sequences of PCR primers

RT-PCR Primers		
Amplified region	primers	Sequence 5'-3'
E6/E7	97S	GTCGACCTGCAATGTTTCAGGACCC
	880AS	GAAACCATAATCTACCATGGCTGATC
GAPDH	GAPDHF	ACCCAGAAGACTGTGGATGG
	GAPDHR	TTCTAGACGGCAGGTCAGGT
LoxP	16S	TATGTATGGTATAATAAACACGTGTGTATGTG
	16A	GCAGTGCAGGTCAGGAAAACAGGGATTGGC
Cloning Primers		
Amplified region	primers	Sequence 5'-3'
pLM16AN	H3FM	GGCAAGCTTAATAAACTTATTGTTCAACACC
	SBFA	CATTGGTACCTGCAGGATCAGCCAT
pLM16(AN+R)	H3FM	GGCAAGCTTAATAAACTTATTGTTCAACACC
	SBFA	CATTGGTACCTGCAGGATCAGCCAT
pLM16(R+AN)	overlap S	CGAAATCGGTTGAACCGAAACCGTTAGTATAAAAGCAGA
	overlap A	TCTGCTTTTATACTAACCGGTTTCGGTTCAACCGATTTCG
pLL16AN	H3F	GGCAAGCTTGATTGTATGTATGTTGAATTAGTTG
	SBFA	CATTGGTACCTGCAGGATCAGCCAT
pLMS16AN1	H3FMS1	GGCAAGCTTCATTGTATATAAACTATATTGCTACATCC
	SBFA	CATTGGTACCTGCAGGATCAGCCAT
pLMS16AN2	H3FMS2	GGCAAGCTTTATATATACTATATTTGTAGCGCCAG
	SBFA	CATTGGTACCTGCAGGATCAGCCAT
pLS16AN	H3FSPH	GGTAAGCTTGCAATGCTTTTTGGCACAAAATGT
	SBFA	CATTGGTACCTGCAGGATCAGCCAT
pLSS16AN	7620HS	GGCAAGCTTGGGACATGCATGCTATGTGCAACTACTGAATCACT
	SBFA	CATTGGTACCTGCAGGATCAGCCAT

SUPPLEMENTARY FIGURE LEGEND

Figure 1 (A) Sequence alignment of 200 subtypes of HPV16 including HPV16AN and HPV16R with sequences spanning the LCR and the E6/E7 coding region. Sequence variability at nucleotide positions 7447, 7528, 7612, 12 and 13 in the non-transcribed region of the LCR and at positions 131, 350 or 570 in the HPV16 E6/E7 coding region are highlighted. Numbering refers to the HPV16 reference genome HPV16R. Sequences with same nucleotide(s) as HPV16AN are shown in green, while sequences having the same nucleotide(s) as HPV16R are shown in red. Numbers in parenthesis refer to the number of HPV16 sequences (out of the 200) that carries the indicated nucleotide.

A





**FACULTY OF
MEDICINE**

Department of Laboratory Medicine

Lund University, Faculty of Medicine

Doctoral Dissertation Series 2022:79

ISBN 978-91-8021-240-3

ISSN 1652-8220

