Regulation of Human Papillomavirus Type 16 Early and Late Gene Expression

Yu, Haoran

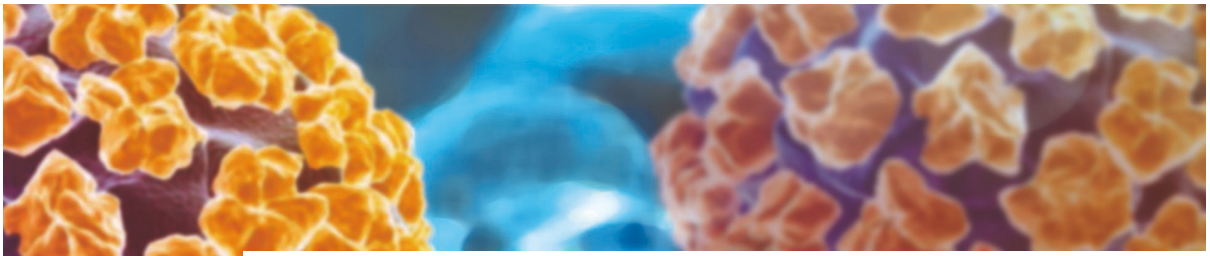
2019

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

Regulation of Human Papillomavirus Type 16 Early and Late Gene Expression

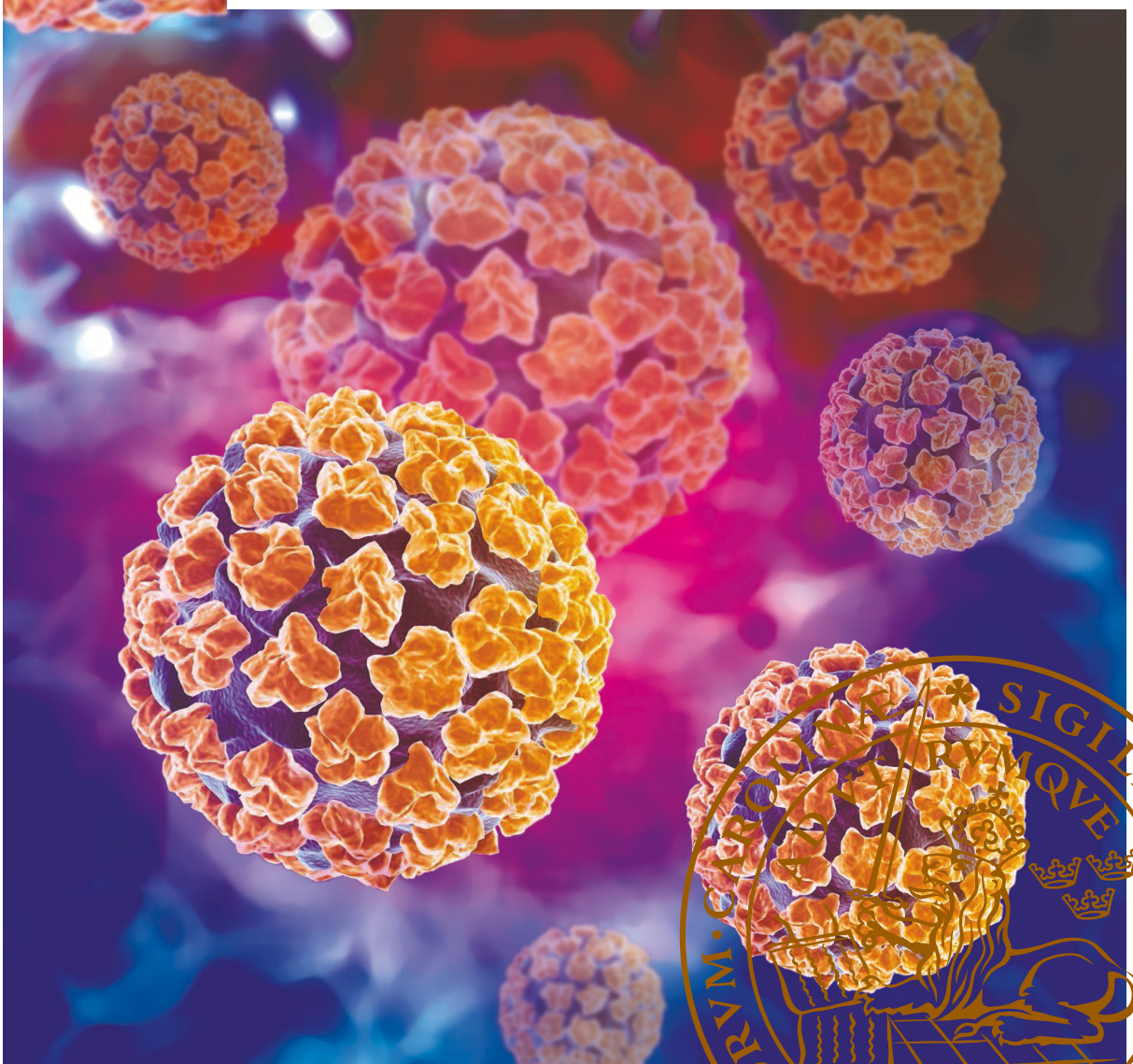
Yu, Haoran; Schwartz, Stefan

2019



Regulation of Human Papillomavirus Type 16 Early and Late Gene Expression

HAORAN YU
FACULTY OF MEDICINE | DEPARTMENT OF LABORATORY MEDICINE | LUND UNIVERSITY



Regulation of Human Papillomavirus Type 16 Early and Late Gene Expression

Haoran Yu

LUND
UNIVERSITY

DOCTORAL DISSERTATION
by due permission of the Faculty of Medicine, Lund University, Sweden.
To be defended at BMC I1345 on 26th, September at 9:00 am.

Faculty opponent
Professor Magnus Evander, Department of Clinical Microbiology,
Umeå University

Supervisor: Professor Stefan Schwartz
Co-supervisor: Professor Patrik Medstrand

Organization	Document name
LUND UNIVERSITY	DOCTORAL DISSERTATION
Department of Laboratory Medicine	**Date of issue:** 26th Spetember, 2019
Author(s) Haoran Yu	Sponsoring organization

Title and subtitle: Regulation of Human Papillomavirus Type16 Early and Late Gene Expression

Abstract

Human papillomavirus type 16 (HPV16) is a cancer-causing virus that contributes to at least 70% of cervical cancer as well as to other anogenital cancers and head and neck cancer. Understanding HPV16 gene regulation is important to enhance our understanding of HPV16 and may contribute to development of antiviral drugs. In this study, we show that nucleosides/nucleoside analogues have the ability to regulate HPV16 gene expression and we identify cellular splicing factor that control HPV16 mRNA processing.

Nucleosides play a significant role in cell differentiation and tumorigenesis and have been used in antiviral and anticancer treatments. We report that adenosine activates HPV16 late gene expression in a dose-and time-dependent manner, but only in the presence of guanosine. Induction of HPV16 late gene expression by adenosine and guanosine was mainly causing read-through at the HPV16 early polyadenylation signal to the HPV16 late region and resulted in HPV16 late L2 mRNA production. This effect was dependent on ENT1 nucleoside transporter that transports adenosine into cells where it is metabolized further by adenosine kinase (ADK) or adenosine deaminase (ADA). Adenosine and guanosine also increased the nuclear export of the cellular HuR protein, which suggested that HuR contributed to nucleoside-mediated induction of HPV16 late gene expression. In addition to adenosine and guanosine, we found that nucleoside analogue cordycepin could also induce HPV16 late gene expression. In this case by increasing the production of polyadenylation factor NUDT21/CPSF5, resulted in inhibition of the early HPV16 polyadenylation site and HPV16 late L2 mRNAs production. We conclude that nucleosides adenosine and guanosine as well as nucleoside analogue cordycepin can induce HPV16 late gene expression, suggesting that nucleosides may play an important role in the control of HPV16 gene expression.

The levels of the cellular splicing factor heterogeneous nuclear ribonucleoprotein G (hnRNP G) was also affected by nucleosides and was studied further. Over-expressing hnRNP G caused inhibition of a small-exon between HPV16 splice sites SA3358 and SA3632 on the L1 mRNA by interacting with an 8-nucleotide enhancer located within this exon. This interaction resulted in increased HPV16 late L1i mRNA production. Furthermore, we found that overexpression of hnRNP G strongly inhibited splicing of the HPV16 E6/E7 oncogene mRNAs. This splicing inhibitory effect resulted in production of unspliced mRNAs encoding the E6 protein at the expense of HPV16 E7 (E6*I) mRNA, indicating that hnRNP G controls expression of both HPV16 E6 and E7 oncogenes. Since hnRNP G affected splicing of both E6/E7 oncogene and late gene L1 mRNAs, the levels of hnRNP G may contribute to the final outcome of the HPV16 infections. In summary, this thesis shows that nucleosides/nucleoside analogues can affect HPV16 gene expression at the level of RNA processing and identifies hnRNP G as a major regulator of expression of the HPV16 oncogenes E6 and E7, and the HPV16 late L1 gene.

Key words HPV16, Gene Regulation, Cervical Cancer, Nucleosides, Splicing, Polyadenylation

Classification system and/or index terms (if any)

Supplementary bibliographical information		**Language** English
ISSN and key title 1652-8220		**ISBN** 978-91-7619-796-7
Recipient's notes	**Number of pages 92**	Price
	Security classification	

Signature  Date 2019-08-21

Regulation of Human Papillomavirus Type 16 Early and Late Gene Expression

Haoran Yu

LUND
UNIVERSITY

Coverphoto by Haoran Yu

Faculty of Medicine

Department of Laboratory Medicine

ISBN 978-91-7619-796-7

ISSN 1652-8220

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

To my supporting and loving parents

"We fall. We break. We fail …

But then, we rise, we heal, we overcome."

—— Kiana Azizian

仁者不忧，知者不惑，勇者不惧。

——《论语·宪问》

我就是我，是颜色不一样的烟火。

天空海阔，要做最坚强的泡沫。

—— 张国荣

Table of Contents

List of Papers

Yu, H., Wu, C., Nilsson, K., Kajitani, N, Schwartz, S. 2018. Adenosine causes read-through into the late region of the HPV16 genome in a guanosine-dependent manner. **Virology, 521:1-19**.

Yu, H., Nilsson, K., Wu, C., Schwartz, S. 2019. Role of nucleoside analogues on HPV16 late gene expression. **Manuscript.**

Yu, H., Gong, L., Wu, C., Nilsson, K., Wang-Li, 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.**

Yu, H., Hao, C., Jönsson, J., Gong, L., Wu, C., Schwartz, S. 2019. hnRNP G is a potent inhibitor of HPV16 oncogene mRNA splicing and promotes production of full-length E6 mRNA. **Manuscript.**

Related Manuscripts

Johansson C, Jamal Fattah T, **Yu H**, Nygren J, Mossberg AK, Schwartz S. Acetylation of intragenic histones on HPV16 correlates with enhanced HPV16 gene expression. Virology. 2015 Aug; 482:244-59.

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 Sep 19;45(16):9654-9678.

Nilsson N, Wu C, Kajitani N, **Yu H**, Tsimtsirakis E, Gong L, Ellenor B, Winquist J, Glahder L, Ekblad J, Schwartz S. The DNA damage response activates HPV16 late gene expression at the level of RNA processing. Nucleic Acids Res. 2018 Jun 1;46(10):5029-5049.

Wu C, Nilsson K, Zheng Y, Ekenstierna C, Sugiyama N, Forslund O, **Yu H,** Wennerberg J, Ekblad L, Schwartz S. 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 Jan 15;144(2):297-310

Abstract

Human papillomavirus type 16 (HPV16) is a cancer-causing virus that contributes to at least 70% of cervical cancer as well as to other anogenital cancers and head and neck cancer. Understanding HPV16 gene regulation is important to enhance our understanding of HPV16 infections and may contribute to development of antiviral drugs. In this study, we show that nucleosides/nucleoside analogues have the ability to regulate HPV16 gene expression and we identify cellular splicing factor that control HPV16 mRNA processing.

Nucleosides play a significant role in cell differentiation and tumorigenesis and have been used in antiviral and anticancer treatments. We report that adenosine activates HPV16 late gene expression in a dose-and time- dependent manner, but only in the presence of guanosine. Induction of HPV16 late gene expression by adenosine and guanosine was mainly causing read-through at the HPV16 early polyadenylation signal to the HPV16 late region and resulted in HPV16 late L2 mRNA production. This effect was dependent on ENT1 nucleoside transporter that transports adenosine into cells where it is metabolized further by adenosine kinase (ADK) or adenosine deaminase (ADA). Adenosine and guanosine also increased the nuclear export of the cellular HuR protein, which suggested that HuR contributed to nucleoside-mediated induction of HPV16 late gene expression. In addition to adenosine and guanosine, we found that nucleoside analogue cordycepin could also induce HPV16 late gene expression. In this case by increasing the production of polyadenylation factor NUDT21/CPSF5, resulted in inhibition of the early HPV16 polyadenylation site and HPV16 late L2 mRNAs production. We conclude that nucleosides adenosine and guanosine as well as nucleoside analogue cordycepin can induce HPV16 late gene expression, suggesting that nucleosides may play an important role in the control of HPV16 gene expression.

The levels of the cellular splicing factor heterogeneous nuclear ribonucleoprotein G (hnRNP G) was also affected by nucleosides and was studied further. Overexpression of hnRNP G caused inhibition of a small-exon between HPV16 splice sites SA3358 and SA3632 on the L1 mRNA by interacting with an 8-nucleotide enhancer located within this exon. This interaction resulted in increased HPV16 late L1i mRNA production. Furthermore, we found that overexpression of hnRNP G strongly inhibited splicing of the HPV16 E6/E7 oncogene mRNAs. This splicing inhibitory effect resulted in production of unspliced mRNAs encoding the E6 protein at the expense of HPV16 E7 (E6*I) mRNA, indicating that hnRNP G controls expression of both HPV16 E6 and E7 oncogenes. Since hnRNP G affected splicing of both E6/E7 oncogene and late gene L1 mRNAs, the levels of hnRNP G may contribute to the final outcome of the HPV16 infections. In summary, this thesis shows that nucleosides/nucleoside analogues can affect HPV16 gene expression at the level of RNA processing and identifies hnRNP G as a major regulator of expression of the HPV16 oncogenes E6 and E7, and the HPV16 late L1 gene.

Populärvetenskaplig Sammanfattning

Humant papillomvirus typ 16 (HPV16) är ett cancerframkallande virus som orsakar minst 70% av all livmoderhalscancer samt till andra anogenitala cancrar och huvud- och halscancer. Att förstå HPV16-genreglering är viktigt för att förbättra vår kunskap om HPV16 vilken kan bidra till utveckling av antivirala läkemedel. I denna studie visar vi att nukleosider / nukleosidanaloger har förmågan att reglera HPV16-genuttryck och vi identifierar ett cellulärt RNA-bindande protein som reglerar uttryck av HPV16 virusets gener.

Nukleosider spelar en betydande roll i celldifferentiering och tumörutveckling och har använts som antivirala medel och som cancerbehandlingar. Här rapporterar vi att adenosin kan aktivera uttryck av HPV16 virusets sena gener som kodar för virus strukturprotein. Detta sker på ett och tidsberoende och concentrationsberoende sätt, men förvånande nog, endast i närvaro av guanosin. Adenosin och guanosin orsakade huvudsakligen genomläsning vid den tidiga polyadenyleringssignalen på HPV16-genomet vilket resulterade i produktion av HPV16 L2-mRNA som kodar för HPV16 kapsidproteinet L2. Denna effekt var beroende av den cellulära ENT1-nukleosidtransportören som transporterar in adenosin i cellen där det metaboliseras av de cellulära enzymerna adenosinkinas (ADK) och/eller adenosindeaminas (ADA). Dessa processer leder till att det cellulära RNA bindandet proteinet binder till HPV16 mRNA. Vi spekulerar att HuR inhiberar HPV16 tidig polyadenylering och ger upphov till HPV16 L2 mRNA som HuR därefter följer ut till cytoplasman. Förutom adenosin och guanosin fann vi att nukleosidanalogen cordycepin också kunde inducera HPV16 sena geners uttryck. I detta fall genom att inducera produktion av den cellulära polyadenyleringsfaktorn NUDT21 / CPSF5, vilket resulterade i hämning av den tidiga HPV16-polyadenyleringssignalen och produktion av HPV16 L2-mRNA. Detta är intressant för att cordycepin är en viktigt komponent i traditionell kinesisk medicin. Sammanfattningsvis konstaterar vi att nukleosiderna adenosin och guanosin samt nucleosidanalogen cordycepin kan inducera HPV16 sent genuttryck, att guanosin kan öka effekten på HPV16 genuttryck av adenosin och andra nukleosidanaloger. Nukleosider kan bidraga till kontroll av HPV16 genuttryck och skulle kunna användas som mål för antiviral terapi. Vidare spekulerar vi att våra resultat tyder på att kombinationer av olika nukleosider skulle kunna öka effekten av olika nukleosidanaloger som används för behandlingar av bla virusinfektioner. Adenosin och guanosin påverkade också nivåerna av det cellulära RNA bindande proteinet hnRNP G som vi studerade vidare. Överuttryck av hnRNP G påverkade RNA-splitsning och därmed genuttrycket av både tidiga och sena HPV16 mRNA. Fram för allt fann vi att överuttryck av hnRNP G starkt inhiberade RNA splitsning av de mRNA som kodar för HPV16 onkogenerna E6 och E7. Effekten av hnRNP G resulterade i produktion av det osplitsade mRNA som kodar för E6-proteinet på bekostnad av det splitsade

mRNA som kodar för HPV16 E7. Eftersom både HPV16 E6 och E7 behövs för utveckling och fortsatt proliferation av en cancercell, spekulerar vi att hnRNP G nivåerna i en cell bidrar till HPV16 virusets cancerframkallande effekt. Sammanfattningsvis visar denna avhandling att nukleosider / nukleosidanaloger kan påverka HPV16 virusets genuttryck på RNA-nivå, sannolikt via det cellulära RNA-bindandet proteinet HuR. Vi visar också att det cellulära RNA-bindande proteinet hnRNP G, som också påverkas av adenosin och guanosin, kontrollerar uttryck av HPV16 virusets onkogener E6 och E7, samt kapsidproteinet L1 på RNA nivå. Slutligen visar vi att guanosin kan förstärka effekten av adenosin och av olika nukleosidanaloger på HPV16 virusets genuttryck, vilket skulle kunna vara av betydelse för behandling av virusinfektioner med antivirala medel som utgörs av nukleosidanaloger.

摘要

由人类乳突病毒（HPV）引起的癌症占病毒引起的癌症总数的百分之十五，每年有大约 600，000 起由人类乳突状病毒引起的癌症病例被确认。其中高危类人类乳突状病毒十六型病毒在超过百分之五十的病例中被检测到。人类乳头病毒是一种嗜上皮性病毒，它的基因表达与被感染细胞的分裂和分化有着密不可分的联系。大多数的人类乳突病毒感染可在感染后的十八到二十四个月自行消失，但如果持续性的感染高危型的人类乳突病毒将会有极大的可能引起宫颈癌变。人类乳突病毒的早期基因在感染基层细胞就早已表达，而其具有高度免疫原性的晚期病毒结构蛋白只在已经分化的上层表皮细胞表达。这种滞后的高免疫原性蛋白的表达可以帮助病毒逃脱宿主自身免疫系统的攻击，我们推测如果使具有高度免疫原性的晚期基因在早期感染中提早表达，可能为宫颈癌早期诊断和鉴定患病危险度的生物标记物提供重要的依据。

人类乳突病毒十六型病毒利用选择性剪切信使核糖核酸来调控病毒的早期和晚期基因表达。同时，病毒和细胞的调控因子也参与控制了病毒基因的表达。我们鉴定出腺苷和鸟苷同时可以共同激活人类乳突病毒十六型的晚期基因表达。腺苷的类似物虫草素也具有激活人类乳突病毒十六型晚期基因表达的特性。我们同时还鉴定和分析出人异质性胞核核糖核蛋白 G 参与了人类乳突病毒十六型的早期和晚期基因的表达。该因子可以间接的抑制人类乳突病毒十六型早期致癌基因六和七的信使核糖核酸的剪切同时也抑制了晚期基因一的外显子增加，从而可以影响致癌基因六和七对细胞生成周期的调控，降低了人类乳突病毒使宿主细胞永生化和癌变的能力。这些研究结果确定了人类乳突病毒十六型对信使核糖核酸的剪接和调控与其致病性有较高的关联性。

Abbreviations

3'RACE	3' Rapid Amplification of cDNA Ends
3'UTR	3' Untranslated Region
ADA	Adenosine Deaminase
ADK	Adenosine Kinase
AR	Adenosine Receptors
ALKBH5	AlkB Homologue
AML	Acute Myeloid Leukemia
AMP	Adenosine-5'monophosphate
ATM	Ataxia-Telangiectasia Mutated
ATP	Adenosine-5'triphosphate
CBC	Cap Binding Complex
CF	Cleavage Factor
CFIm	Human Cleavage Factor Im
CIN	Cervical Intraepithelial Neoplasia
CMV	Cytomegalovirus
CNT	Concentrative Nucleoside Transporters
CPEB1	Cytoplasmic Polyadenylation Element-Binding Protein1
CPSF	Cleavage and Polyadenylation Specificity Factor
CstF	Cleavage Stimulation Factor
DBD	DNA Binding Domain
DNA	Deoxyribonucleic Acid
EGF	Epidermal Growth Factor
EGF-R	Epidermal Growth Factor-Receptor
ENT	Equilibrative Nucleoside Transporters
ER	Endoplasmic Reticulum
ESE	Exonic Splicing Enhancers
ESS	Exonic Splicing Silencers
EV	Epidermodysplasia Verruciformis
FTO	Fat Mass and Obesity-Associated Protein
GTase	Guanylyltransferase
HLA-I	Human Leukocyte Antigen-I
hnRNA	Heterogenous Nuclear RNA
hnRNP	Heterogeneous Nuclear Ribonucleoproteins
HCV	Hepatitis C Virus
HPV	Human Papillomavirus
HR	High Risk Type
HR	Homologous Recombination
HS	Heparan Sulfate
HSPG	Heparan Sulfate Proteoglycans

IMPDH	Inosine Monophosphate Dehydrogenase
IRES	Internal Ribosomal Entry Site
ISE	Intronic Splicing Enhancers
ISS	Intronic Splicing Silencers
JNK	C-Jun NH2-Terminal Kinase
KH	K-Homology
LCR	Long Control Region
LR	Low Risk
MAPK/ERK	Mitogen-Activated Protein Kinase/extracellular Signal-Regulated Kinase
METTL	RNA Methyltransferases-Like Protein
MHC	Major Histocompatibility Complex
NHEJ	Non-Homologous End-Joining
NLS	Nuclear Localization Signal
ORF	Open Reading Frame
pAE	HPV Early Polyadenylation Site
pAL	HPV Late Polyadenylation Site
PARP-1	Poly (ADP-Ribose) Polymerase 1
PI3K	Phosphoinositide 3-Kinase (PI3K)
pRB	Tumor Suppressor Retinoblastoma-Associated Protein
RGG boxes	Repeats of Arg-Gly-Gly Tripeptides
RRM	RNA Recognition Motif
RS domain	Serine/Arginine-Rich Domain
RSV	Respiratory Syncytial Virus
SAH	S-Adenosyl Homocysteine
SCC	Squamous Cell Carcinomas
sLuc	Secreted Luciferase
snRNPs	Uridine-Rich Small Nuclear Ribonucleoproteins
SRSF1	Serine- and Arginine-Rich Splicing Factor 1
HNSCC	Head and Neck Squamous Cell Carcinoma
TBS	Tris-Buffered Saline
TAD	Transactivating Domain
TPase	RNA Triphosphatase
VLP	Virus-Like Particle
YTHDF	YTH Domain-Binding Protein

1 Introduction

1.1 Human papillomavirus (HPV) and cancer

Human Papillomaviruses (HPV) are double stranded, non-enveloped DNA tumor viruses. The viral icosahedral capsid is about 55nm in diameter, containing the virus genome of approximately 8,000 base pairs [1, 2]. HPVs are highly strictly epitheliotropic and infect only cutaneous epithelium and mucosal epithelium [3]. The mucosal HPV types are divided into low risk (LR) or high risk (HR) according to their association with malignant carcinoma progression [4]. HPV infections usually can be cleared by immune system within 12-24 months [5], however there is about 10% of HPV infections cannot be cleared and result in persistent long term infections [6]. Persistent infections by HPV high risk types may cause pre-malignant lesions and if left untreated they may finally develop to cancer [3], which is a rare event.

Virus induced cancers contribute to at least 15% human cancers. HPV accounts for nearly half of virus induced cancers. Each year approximately 600,000 cases of HPV associated cancers of cervix, anal, vulva and oropharynx cancers are diagnosed worldwide in addition to benign infections including genital warts and respiratory papillomatosis [7]. HPV infections in genital tract area are the most commonly sexually transmitted infections [8]. Almost every sexually active individual will be infected by HPV at least once or even multiple times during their lifetime. Most HPV infections are remaining asymptomatic and are cleared within one- or two-years after infection, but in rare cases persistent infections can occur and cause malignant lesions called cervical intraepithelial neoplasia (CIN) [3]. The CIN is graded into three levels depending on the proportion of abnormal cells in the epithelium layer. CIN1 is a low-grade squamous intraepithelial lesion, and is most likely to regress back to normal conditions, while CIN2 and CIN3 are high-grade lesions with high risk of progressing to cancer **(Figure 1.1)** [3]. Cervical cancer is the third most common cancer in women worldwide. The number of new cases of cervical cancer in 2020 is estimated to be approximately 665,000 cases [7]. HPV was detected in 99% of all cervical cancer cases and HPV16 is one of the most common HPV types. HPV16 and HPV18 together are responsible for at least 70% of all cervical cancer cases [9, 10]. Beside cervical cancer, there is an increasing trend of HPV-caused head and neck squamous cell carcinoma (HNSCC) arising

from various sites. Among these, HPV16 is the predominant type and attributes for 90% of HPV-positive HNSCC [11, 12] .

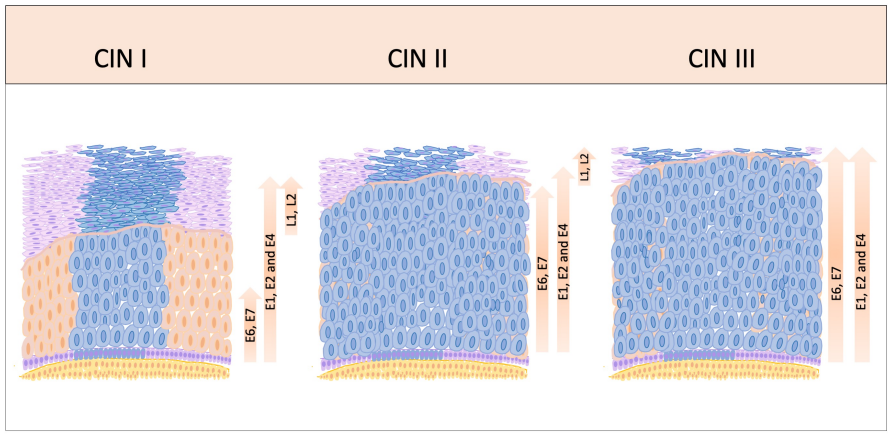


Figure 1.1 Changes in the HPV16 life cycle during the cervical cancer development.
The normal HPV life cycle is lost during the progression of Cervical Intraepithelial Neoplasia grade I (CIN1) to grade 3 (CIN3). CIN1 is the productive infection that has virus synthesis and releases from the differentiated upper layer. CIN2 and CIN3 are high grade lesion with deregulated viral gene expression.

To date vaccination is the only effective method in preventing HPV infections. All the available vaccinations are composed of recombinant HPV capsid proteins L1 that self-assemble into virus-like particles (VLP) [13]. There are three vaccines available on the market to date. Gardasil was the first vaccine approved and provides protection against the low risk type HPV6 and 11 as well as the high risk type 16 and 18. In recent years, Gardasil lunched another new vaccine Gardasil 9, which targets HPV16, 18, 6, 11, 31, 33, 45, 52, and 58. The new vaccine provides protection for extra 15-20% of cervical cancer cases [14]. In addition, Cervarix has a bivalent vaccine targeting 16 and 18 only. Unfortunately, none of the above vaccines are able to clear existing infections, and there is no medicine available on the market to treat HPV infection neither. Therefore, understanding of HPV infection mechanisms including HPV gene regulation is urgent for developing antiviral and anticancer treatments for HPV infections and HPV derived cancers.

1.2 HPV classifications

Human papillomavirus belongs to the family of *Papillomaviridae*, and there are over 200 HPVs has been classified to date [15]. HPVs may be grouped into cutaneous HPV or mucosal HPV based on the location of infection site. Cutaneous

HPVs usually cause unapparent or asymptomatic infections and can be cleared within months by immune system [16]. Some cutaneous HPV types contribute to cancer development such as squamous cell carcinomas (SCC) in cooperation with UV radiation in immunosuppressed individuals [17]. Mucosal HPVs mainly infect epithelial cells in genital or oral mucosa and maybe subdivided into high risk (HR) and low risk (LR) groups based on their relative risk of causing malignancy in the infected tissue [18].

Characterization of new types of HPVs are based on a defined sequence in the capsid protein L1 open reading frame (ORF) region that is highly conserved. Each HPV type differs from another by at least 10% in L1 ORF gene. Any HPV that differs less than 10% is defines as a subtype, or a variant if the difference is less than 2% [19]. All the HPVs are grouped into five genotypes: *alpha, beta, gamma, nu* and *mu* [19-21]. The largest group is *alphapapillomavirus*, which is found in both genital and mucosal regions. This group includes HPVs that induce benign skin warts as well as the ones causing cancer such as HPV16 [22]. The second largest group is the *betapapillomavirus* that is found in cutaneous regions and was first identified in flat warts that cause widespread asymptomatic infections in the human population [23]. Some types within this genus, like HPV5 or 8, are defined as possibly carcinogenic since they could promote keratinocyte carcinoma development in immunosuppressed patients and in patients suffering from epidermodysplasia verruciformis (EV) [24]. The other three groups *Gamma*, *Mu* and *Nu-papillomaviruses* mostly cause cutaneous warts including hand, plantar and flat warts [23]. The latter two groups only contain three HPV types HPV1, HPV63 (*mupapillomavirus*) and HPV41 (*nupapillomavirus*) [19].

1.3 HPV infections and life cycle

HPV infections generally start by infection of the undifferentiated basal layer of the mucosal epithelium cells through microlesions or wounds. Basal layer cells are the only actively dividing cells in the epithelium and consists of stem cells and transit amplifying cells. Infection of these cells is found to be necessary in order to establish persistent infections [25, 26]. The life cycle of HPV is tightly regulated by cell differentiation through both transcriptional and post-transcriptional mechanisms **(Figure 1.2)**.

Virions initially bind to the heparan sulfate proteoglycans (HSPGs) on the cell surface that induce conformational change of the capsid [27]. This conformational change allows the virions to bind to the unknown entry receptors [28]. The virus is then endocytosed by the cell and uncoating starts in the acidified endosomes in cell cytoplasm. The entry of the viral genome into the cell nucleus requires the recipient

cells to enter M phase undergoing mitosis [28, 29]. After the infection established, viral genome is stably maintained in the basal epithelium cell nucleus in episomal form. HPV viral genome replication can be divided into three stages: establishment, maintenance and productive replication [30]. At this stage the virus starts to replicate its genome along with the host cell genome once per cell cycle. This allows the viral genome copy number to be maintained at a very low level, about 50-100 copies per cell, thereby assuring low level of gene expression and making sure that the infected cell can escape from the immune surveillance [18]. During this phase, E1, E2, E6 and E7 are the first viral genes expressed. E1 and E2 initiate viral DNA replication by bringing the cellular DNA polymerase to viral genome. E7 binds to the tumor suppressor retinoblastoma-associated protein (pRB) and prevents binding of pRb to its partner E2F. The liberation E2F in turn activates the expression of DNA replication factors controlled by E2F [31]. E6 inactivates and degrades the tumor suppressor protein p53 that results in blocking of cell apoptosis. The E6 and E7 together modulate cell cycle regulators [18, 32]. All these four HPV proteins E1, E2, E6 and E7 are necessary for stimulating cell cycle progression and maintaining persistence of viral genome in basal layer cells.

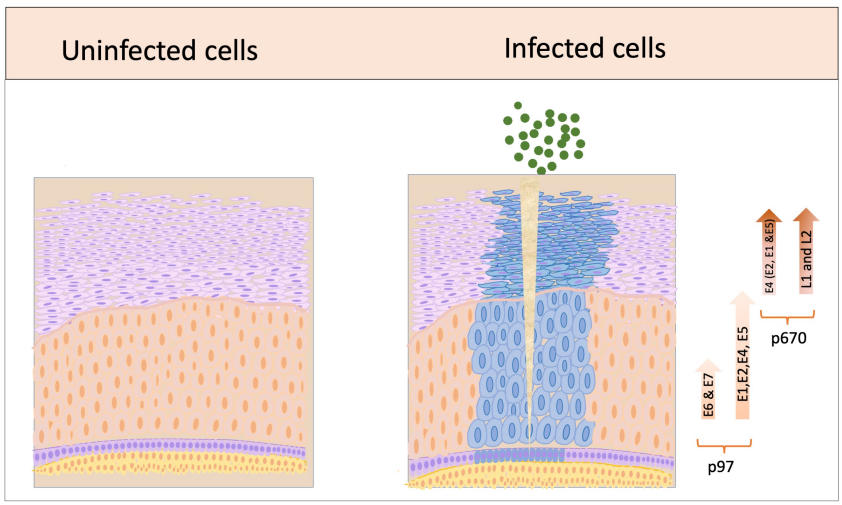


Figure 1.2 The HPV life cycle in epithelium cells.
HPV infects basal layer cells via microlesions. Upon infection, the viral genomes establish in the cell nucleus and are replicated in synchrony with the host DNA replication. When basal layer divides, one daughter cell migrates up from the basal layer and undergoes differentiation. Only early genes expressed from the early promoter p97 is expressed at this stage. When the cells reach the superficial layer and undergo final differentiation, both early and late gene from the late promoter are expressed.

When the basal layer cells start to divide, one daughter cell is pushed up from the basal layer to the supra-basal layer and starts to differentiate, while the other daughter cell remains in the basal layer and continues to proliferate. Once the

infected cell leaves basal layer, the viral genome starts to replicate to high copy number by forcing cell to remain active in the cell cycle by HPV viral early proteins [28]. The high expression levels of viral early proteins will eventually inhibit the viral early promoter which indirectly causes cell differentiate that actives the late viral promoter, thereby allow the cells to prepare for the production of the HPV late proteins [33, 34]. The viral capsid proteins L1 and L2 are produced in the terminal differentiation stage, leading to virus assembly and production in the superficial layer of the epithelium. These virions are then released to the environment and are capable of infecting new host or new sites.

Most HPV infections can be cleared within 12 to 48 months, but high-risk HPV infections may in rare cases lead to persistent infections and eventually cause pre-malignant cervical lesions. These lesions could progress to cancer if left untreated. The high-risk HPV caused malignant cancers are frequently containing integrated HPV viral genome into cellular chromosomes. Such events usually occur within the E2 ORF [31], resulting in loss of all viral gene expression except for the oncogenes E6 and E7 under control of the viral early promoter. This integration event allows high levels expression of E6 and E7, thereby promoting malignant progression [35]. However, in some high-risk HPV-caused cancer cases including cervical cancer. HPV is found episomal form in several cancer cases [36, 37], suggesting that integration is not essential for cancer development.

1.4 HPV16 virion genome structure and proteins

All HPVs contain double stranded circular DNA and can be divided into three regions: the early, the late and the long control region (LCR) **(Figure 1.3)**. These three regions are separated by the two viral promoters and the two polyadenylation sites [38, 39]. The early promoter p97 controls all the early genes including E6 and E7 and expressed mRNAs are polyadenylated at the early polyadenylation site (pAE). As the cells start to differentiate, the early promoter p97 is deactivated and resulting in inhibition of E6 and E7 oncoprotein expression. The late promoter p670 is then activated by cell differentiation and produces HPV early proteins E1 E2, E4 and E5 as well as all HPV late proteins. L1 and L2 are produced from the late promoter p670 and polyadenylated at the late polyadenylation site (pAL) [2, 39, 40]. The LCR region is about 850bp in size and starts right after the L1 stop codon and ends up right before the E6 start codon. The LCR region does not encode for any proteins, but it is important for viral gene regulations as it bears the origin of the viral replication and multiple transcription factor binding sites [2, 41, 42].

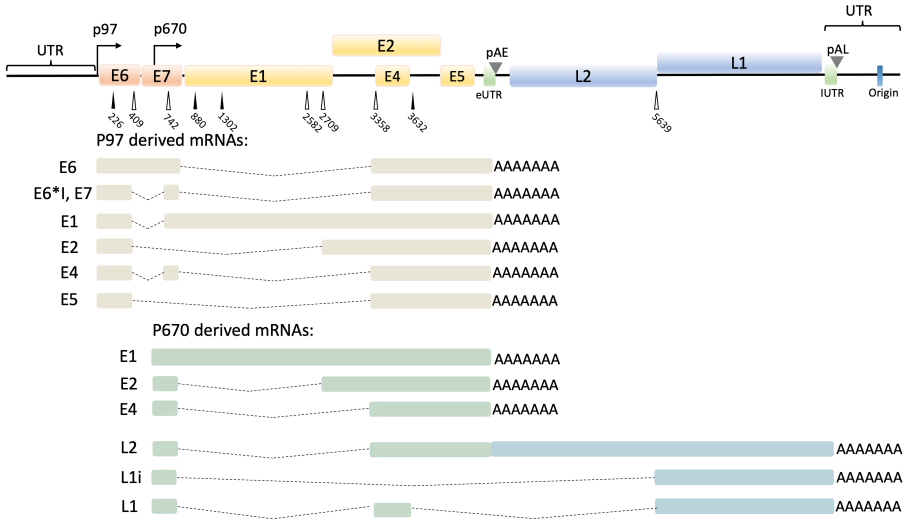


Figure 1.3 The HPV16 genome structure.
Oncogene E6 and E7 are indicated in orange, early genes including E1, E2, E4 and E5 are indicated in yellow and late gene L1, L2 are indicated in blue. The two promoters p97 and p670 and two polyadenylation sites pAE and pAL are also shown. The open triangles indicate splice acceptors and the black triangles indicate splice donors. The major early and late mRNAs produced by HPV16 are shown in the figure.

1.4.1 Oncoprotein E6 and E7

Both E6 and E7 in high risk HPVs play important roles in maintaining HPV infections and cellular transformations. They are also the first genes expressed after infection and are important for virus replication [3, 44]. They are also responsible for the cancer progression in HR HPV caused cervical lesions. E6 and E7 together interfere with cellular tumor suppressor proteins pRb and p53, driving cell immortalization and preventing apoptosis and terminal cell differentiation [39, 43].

The E7 protein from all identified HPV types is found to bind the members of tumor suppressor protein pRb family, but only the high-risk HPV E7 contributes to cancer progression. E7 is a protein of about 100 amino acids and locates in the nucleus. However recent studies suggest that E7 may also have a role in the cytoplasm [45]. E7 consists of three conserved regions: CR1, CR2 and CR3 [46]. The CR1 and CR2 are located in the N-terminal domain, which is associated with pRb destabilization through completive the binding to the transcription factor E2F [47]. CR3 is located at the C-terminal zinc finger domain of the E7 protein and has high affinity for pRb, as well as other proteins [48].

The E7 activates the infected cell cycle by binding and releasing/degrading the p105 (pRb) as well as other two pocket proteins, p107 and p130 [3, 49]. The tumor suppressor pRb has a critical role in controlling the cell cycle transition from G1 to S phase. In normal cells, the pRb is unphosphorylated in G1 phase, then is gradually phosphorylated towards S phase. The unphosphorylated pRb interacts with the transcription factor E2F and the E2F-pRB complex in turn suppress the expression of cell cycle dependent transcriptional regulators [50]. Binding of E7 protein to pRb will prevent pRb from binding to transcription factor E2F, in turn liberating E2F that activates various cell cycle-related genes, especially S-phase specific genes, thereby facilitating cell cycle transition into S phase [3, 43]. In high-risk HPV types, E7 can further degrade the pRb to completely deactivate its function and contribute to cancer development, while the E7 in low-risk HPVs do not have this function [51]. Besides pRb, E7 is reported to inactivate the cyclin dependent kinase inhibitors p21 and p27, which are responsible for cell growth arrest [51]. E7 also binds to various cellular transcription factors such as the DNA damage sensor ATM (ataxia-telangiectasia mutated) [52], E2F6 [53] and p600huh [54].

Due to the highly efficient abrogation of pRb, cells will stay in S-phase and promote cell proliferation, resulting in more stabilized tumor suppressor protein p53. This unscheduled cell proliferation event and high levels of p53 would eventually trigger apoptosis. However, high-risk HPVs express E6 proteins that degrade p53 to avoid apoptosis to occur [43, 55]. The E6 protein is a protein of approximately 150 amino acids in size and contains two zinc-finger domains. E6 protein is well known for binding to the E3 ubiquitin ligase, E6-associated protein (E6AP) [56] and tumor suppressor protein p53 [57]. E6AP together with E6 and its target protein p53 forms a trimeric complex that leads to ubiquitination of p53, subsequently followed by proteome mediated degradation of p53 [58]. The E6 binding to p53 also interferes with p53 DNA-binding ability thereby inhibiting transcription [59]. Even though both LR and HR HPV E6 can bind to p53, only HR HPV E6 can degrade p53. LR HPV E6 only binds to the C-terminus of p53, while the HR HPV E6 binds to both the C-terminus and the core region of p53. Only binding to the core region of p53 is essential for p53 degradation [60, 61]. Apart from directing degradation of p53, E6 also has the ability to inactive p53 through other mechanisms, including interacting with CBP/p300, which is known for acetylating p53. E6 can inhibit p53 acetylation by direct binding to p300, and in turn, down regulate the p53 activity [62, 63]. Another mechanism that has been proposed for E6 is to inhibit p53 activity by aberrating p53 localization. It has been hypothesized that by binding to p53, E6 might mask the p53 nuclear localization signal (NLS) in C-terminal region and enhance its nuclear export to cytoplasm. As a consequence, p53 signaling pathway is abolished by aberrant localization, even though the p53 remains intact [53, 64].

1.4.2 E1 and E2 proteins

The E1 and E2 proteins together play a key role in initiating viral genome replication. These two proteins are believed to be the very first proteins produced at the early stage of the virus life cycle. E1 is a protein about 68 kDa in size and consists of three domains: a N-terminal regulatory domain, a DNA binding domain (DBD), and a C-terminal ATPase/ helicase domain [65-67]. E1 protein is expressed at very low levels in both infected or transformed cells, and it has rarely been visualized in its natural context. E1 binds to its binding sites within LCR region with low affinity [68] and E2 stabilizes E1 binding by binding to DNA sequences adjacent to the origin of replication. With the help of E2, E1 increases its affinity for HPV DNA and binds efficiently to the origin of the viral genome. In the form of two hexamers, E1 acts as bi-directional helicase to unwind the viral DNA and recruit the host replication machinery to the site of the origin [69, 70].

The E2 protein is approximately 50kDa in size and is required for both DNA replication and transcription through the early and intermediate stages of the viral life cycle [71, 72]. E2 protein localizes primarily in the nucleus. However, one study observed an nuclear-cytoplasmic shuttling of E2 in high risk HPV types [73]. E2 protein consists of three domains, a conserved N-terminus is transactivating domain (TAD), a hinge region and C-terminal DNA binding domain (DBD) that also harbors the nuclear localization signal (NLS) [72]. The N-terminal responsible for protein-protein interactions, while the C-terminal region is responsible for DNA binding [74]. The DNA binding sequences $ACC(N)_6GGT$ for E2 are highly conserved in at least 9 different genera of HPV [75]. Notably, E2 binding sites include more than one CpG dinucleotide, suggesting that methylation might be involved in the regulation of E2 activity [76, 77]. There are four binding sites for E2 in the LCR region of the HPV genome, and their positions are highly conserved among all high-risk HPVs [74]. Three out of four binding sites are located proximal to the viral early promoter p97 and are required for the E1 mediated activation of viral DNA replication.

E2 is a multifunctional protein that is involved in viral replication, transcription, initiation of viral replication, maintenance and partitioning of extrachromosomal viral genomes [78]. E2 can either activate or repress viral transcription in a dose-dependent manner [79-82]. During the early stage, E2 binds to the HPV promoter region and enhances transcription. E2 also binds to E1 cooperatively and forms a dimer of hexamers that binds to the viral replication origin that contains both binding site for E1 and E2 [75]. After E2 has helped to load E1 onto the replication origin, E1 is converted to a double-hexametric helicase and recruits the cellular DNA replication machinery [75, 83]. E2 is released from its binding site [83, 84]. The E2 protein also can function as transcription repressor by binding to the viral

DNA at the late stage and repressing the early promoter p97 by inhibiting the transcription factors interacting with the early promoter p97. As a consequence, high expression levels of E2 will eventually lead to the inhibition of the early promoter p97, but not to the late promoter p670 [85, 86]. E6 and E7 expression will then shut down due to the promoter switch, and increased levels of early proteins E1, E2, E4 as well as late proteins L1 and L2 will be produced. In high grade lesions, the HPV viral genome is usually found integrated into the host chromosome at E2 region. This integration does not show any apparent hotspot in chromosomes, but often occurs near naturally occurring common fragile sites in the genome instable region [87, 88]. The E2 expression is abolished by the integration event, which leads to the deregulated expression of viral early genes including E2, upregulated expression of oncoproteins E6 and E7 and increased viral proliferative capacity, finally progress to cancer [89].

In infected basal cells, in order to ensure that HPV genome is segregated evenly in daughter cells, the viral DNA is replicated together with cellular genome through tethering viral genomes to host chromosomes by E2 protein [74]. E2 protein binds to the binding sites within LCR region of the HPV genome through its DNA binding domain, and tethers the HPV genome to the mitotic chromosomes via protein-protein interactions through the N-terminus domain [75]. The most studied tethering targets for E2 protein is the chromatin adapter protein, Brd4 [90-92]. Although all papillomavirus E2 proteins interact with Brd4 protein, but the binding strength varies between different papillomaviruses groups [93]. The E2 protein from *alpha papillomavirus* binds to Brd4 relatively weaker than other groups and cannot be detected easily on mitotic chromosomes [93, 94]. Studies showed that Brd4 is recruited to the HPV replication foci formed by E1 and E2 proteins in keratinocytes, suggesting that Brd4 might play a role in viral replication [52, 95, 96]. But the exact mechanism of E2-Brd4 chromosomal binding and tethering in *alpha papillomavirus* still remains elusive. Several other tethering targets have been suggested for tethering *alpha papillomavirus* viral genome, including mitotic spindle [97], an ATP-dependent DNA helicase ChIR1 [98]and mitotic kinesin like protein, MKlp2 [99].

During the initial viral replication, the viral genome is limited to approximately 50 to 200 copies per cell [18]. This limited replication is controlled by an E2 isoform called E8^E2. In order to avoid activation of the immune system. E8^E2 is encoded by a short E8 exon spliced to the splice acceptor in the middle of the E2 ORF [100]. The E8^E2 protein function as a transcriptional repressor for viral DNA replication, and it limits the viral copy number during early stage via the cellular NCoR/SMRT complex [101].

1.4.3 E4 and E5 proteins

The E4 protein is the most abundant HPV protein in different layers of HPV infected cells. The E4 ORF overlaps with the E2 ORF, but the E4 gene does not have a start codon. This obstacle is overcome by splicing of the HPV16 mRNAs to join the first five amino acid coding sequence of E1 onto E4 [72, 102, 103], thereby forming the E1^E4 protein (hereafter referred as E4 protein). The E4 protein is highly abundant and can be easily detected in lesions caused by different HPV types, so it could serve as a biomarker for viral infections as well as disease severity in high-risk HPV types [24, 104].

The E4 protein expression occurs late compared to other early proteins. It correlates with the onset of viral genome amplification, prior to the expression of the late capsid proteins [105, 106]. The E4 protein is believed to have many different functions, including modulating viral genome amplification [107], virus synthesis [108], virus release and/or transmission by interacting with HPV L1 and L2 protein [106]. The expression of HPV16 E4 protein can cause host cell cycle arrest in the G2 phase by sequestrating cytoplasmic kinase Cyclin B/Cdk1, and contribute to viral genome amplification efficiency [107, 109, 110]. In high risk HPV types such as HPV16, the G2 arrest by E4 protein might potentially inhibit the E6/E7 mediated cell proliferation in the middle layer of epithelial cells [106]. The E4 protein can also bind directly to type 1 keratins and trigger the cellular stress response with activation of different stress associated kinases including p38 and pJNK. This is also believed to contribute to viral escape from immune surveillance during infections [111]. The E4 protein is believed to have more functions that are important for virus gene regulations. However, the exact mechanisms are less understood and requires further investigation. These include interactions with cytoplasmic CyclinA/Cdk2 [112], associations with E2 [113] and effect on cellular kinases [111, 114].

The E5 protein is about 83 amino acids in size, and is a membrane-bound, hydrophobic protein that is associated with the nuclear membrane, the endoplasmic reticulum (ER) and the Golgi apparatus [115]. The E5 protein consists of three hydrophobic regions, the N-terminal, the C-terminal and the central region. Deletions in the N-terminal region results in alteration in E5 localization to ER and Golgi apparatus, suggesting that the N-terminus is probably responsible for E5 localization [116, 117]. The E5 protein is a viral replication protein along with E1, E2 and E4 proteins and helps in viral episomal DNA replication [39, 104]. The E5 protein has been purposed to be an oncoprotein due to the fact that E5 may accelerate carcinogenesis via the stimulation of type I IFN response in the early stage of tumorigenesis, as E5 is deleted later on when the HPV genome is integrated into the chromosome in some of the cancer cases [118-121]. E5 also interacts with epidermal growth factor-receptor (EGF-R) pathway and causes the increased

expression of EGF-R on the cell surface that leads to enhanced EGF signaling cascades including Ras-ERK1/2 and PI3K/Akt pathways [122]. Overexpression of EGF-R has been reported in numerous cancer cases, and E5 mediated overexpression of EGF-R makes cells more sensitive to EGF, and could be potentially serve as an anti-cancer drugs target [123]. E5 can also enhance cell cycle entry into the S phase in an EGF-R-dependent manner [124], but the exact mechanism remains unclear. E5 can help the virus-infected cell to escape from immune responses in different ways. E5 can reduce the expression of the major histocompatibility (MHC) class I and class II proteins on the cell surface in keratinocytes [125, 126]. E5 also down-regulates the human leukocyte antigen-I (HLA-I) by direct binding to its heavy chain to decrease its expression on the cell surface [127]. Thus, E5 helps the infected cell from immune surveillance and avoids elimination of the virus infected cells from functional immune system.

1.4.4 L1 and L2 late proteins

The late L1 and L2 genes encode the late L1 and L2 proteins, which are the structural components of the HPV viral capsid. The HPV viral capsid consists of 72 capsomeres and each capsid has 360 copies of L1 protein and 12 copies of L2 protein [128, 129]. Expression of these two proteins is highly regulated by cell differentiations. L1 and L2 proteins are not detectable during the early stage of HPV life cycle, and only expressed in the upper most layers of the squamous epithelium at the late stage of infections [40, 130].

The L1 protein is 55kDa in size and is the most conserved papillomavirus protein. L1 protein functions as the major capsid protein and can self-assemble into a virus-like particles (VLP), which is used for the HPV vaccines [8]. The initial interaction of the HPV viral capsid with host cell is largely attributed to L1 protein. L1 protein interacts with heparan sulfate (HS) carbohydrates that are displayed on the proteoglycans on the host cell surface [131,132] and induces a subtle conformational change of the viral capsid. This conformational change causes exposure of the amino terminus of the minor capsid protein L2 to the surface of viral capsid. The cellular furin protease cleaves the L2 protein and induces a secondary conformational change of viral capsid. This conformational change will eventually allow the virus to bind to an undefined secondary entry receptor on the cell surface [133-135].

The L2 protein is approximately 55kDa in size and has been reported to have functions in virus entry [136], nuclear entry and transport [137]. L2 cannot form VLPs but L2 can be incorporated into the capsid and highly increases the DNA encapsidation efficiency of the viral capsids [138, 139]. There are two highly

conserved cysteine residues in L2 across all HPV types. Mutations in either or both of these two cysteine residues will result in production of non-infectious virions [140, 141].

1.5 Post-transcriptional regulation of gene expression

1.5.1 5'capping

The process of 5' cap is an important step in all eukaryotic mRNAs and is the first RNA modification for newly transcribed mRNA. 5'capping starts soon after transcription initiated by adding an N7-methylated guanosine linked to the first nucleotide of the nascent RNA. The 5'cap structure does not only determine the mRNA splicing, polyadenylation, translation and stability [120, 121], but also function as a protecting group from 5' to 3' exonuclease activity for nascent mRNA. Once the newly transcribed mRNA is 25-30 nucleotides in length [142, 143], the 5'capping starts to take place in three steps. First, the RNA triphosphatase (TPase) removes the γ-phosphate from the nascent RNA to from 5'diphosphate RNA. Then, RNA guanylyltransferase (GTase) transfers a GMP group to the 5'diphosphate end and forms a 5'triphosphate. Last, guanine-N7 methyltransferase (guanine-N7 MTase) adds a methyl group to the N7 position of the guanosine to form the 5'cap [144]. The 5'cap structure is recognized by the cap binding complex (CBC) once the 5'cap is formed in the nucleus. This binding of 5'cap to CBC is required in many processes including spliceosome assembly, 3' processing, RNA nuclear export as well as the RNA nonsense mediated decay [145, 146].

1.5.2 Polyadenylation

Polyadenylation is part of the gene expression process that produces mature mRNAs after the transcription terminates. Polyadenylation is the non-template addition of adenosine residue to the precursor mRNA and results in a 200-300 nucleotide long tail of adenosines. This process is essential for mRNA stability, nuclear transport and the initiation of translation [147].

First, the Cleavage/Polyadenylation Specificity Factor (CPSF) binds to the conserved key polyadenylation signal sequence AAUAAA located near the 3'-end of the newly produced pre-mRNA [148]. The Cleavage stimulation Factor (CstF) binds to the GU-rich sequence located downstream of the polyadenylation signal [149]. The CPSF and CstF protein complexes further recruit Cleavage Factor (CF) I and II, poly (A) polymerase (PAP) and nuclear poly(A) binding protein (PABP)

to form the functional complex that cleaves the mRNA at the cleavage site 5'-CA-3' **(Figure 1.4A)**. The polyadenylation starts right after the pre-mRNA cleavage. PAP catalyzes the addition of adenosine monophosphate to the RNA [150] while PABP binds to the short newly synthesized poly (A) tail to increase its affinity for PAP. Once the poly(A) tail reaches a certain length, enzymes can no longer bind to CPSF and the polyadenylation stops [151] **(Figure 1.4B)**.

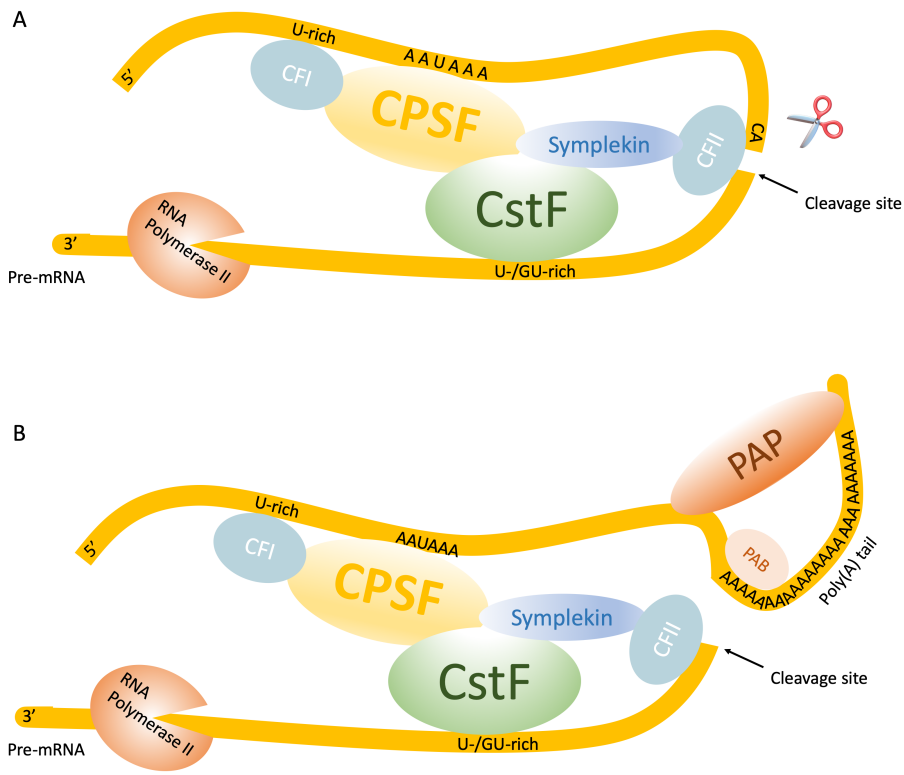


Figure 1.4 Schematic representation of the cleavage and polyadenylation process.
A. The polyadenylation complex recgonize and cut the cleavage site. **B.** After cleavage, PAP starts to catalyze the Poly(A) tail addition.

1.5.3 Alternative splicing

Splicing is a process that transfers the precursor mRNA into matured mRNA. During this process, the introns are excised to bring the exons together. Pre-mRNA splicing was first discovered in adenovirus late mRNAs, and was later found to be a common event in most mRNAs in eukaryotes [152]. There are two types of splicing processes: the constitutive splicing and the alternative splicing. The constitutive splicing removes introns and joins the exons together to produce mature mRNA, whereas in alternative splicing, the exons and introns are alternatively included or excluded to form different mature mRNAs from the same coding gene [153].

Alternative splicing is one of the RNA splicing mechanisms that allows a single gene to express multiple forms of encoded protein. During alternative splicing, the same exon may either be included or excluded on the final mRNA, which in turn is translated into different proteins with different functions [153]. In human, Over 90% mRNAs are alternatively spliced and give rise to around 4-5 fold more proteins variants than the protein-coding genes in the human genome [154].

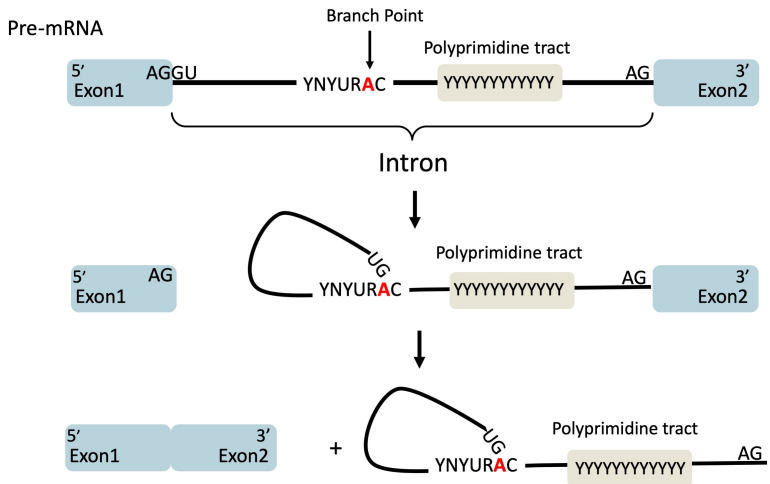


Figure 1.5 Schematic representation of pre-mRNA structure and mechanism of pre-mRNA splicing reaction. Exons are represented as blu boxes, intron, branch point and polyprimidine tract are indicatedd. Y represents U or C, R reprents G or A, and N represents any of four nucleotide.

Both introns and exons have multiple cis-acting regulatory elements for splicing **(Figure 1.5)**. There are four cis-acting elements are absolutely necessary for splicing process: the 5' splice site, the 3'splice site, the branch point (BP) and the polypyrimidine tract [153]. The splice site is a special sequence at the junctions of introns and exons. The 5' splice site marks the exon/intron junctions at 5'end of the intron and is recognized by the consensus sequence AG|GURAGU, where the

underlined sequence referred sequence that is highly conserved dinucleotide at the 5’ border of the intron [153]. The 3’ splice site is recognized by the consensus sequence YAG|RNNN and is located at the extreme 3’ end of the intron. The underlined AG refers to the highly conserved dinucleotide [153]. The branchpoint is located at 18-40 nucleotides upstream of the 3’ splice site. It is less conserved and has the sequence YYURAC, in which the underlined A is an adenosine that is conserved in all genes [155]. The U and C rich polypyrimidine tract locates between the branchpoint and the 3’ splice site.

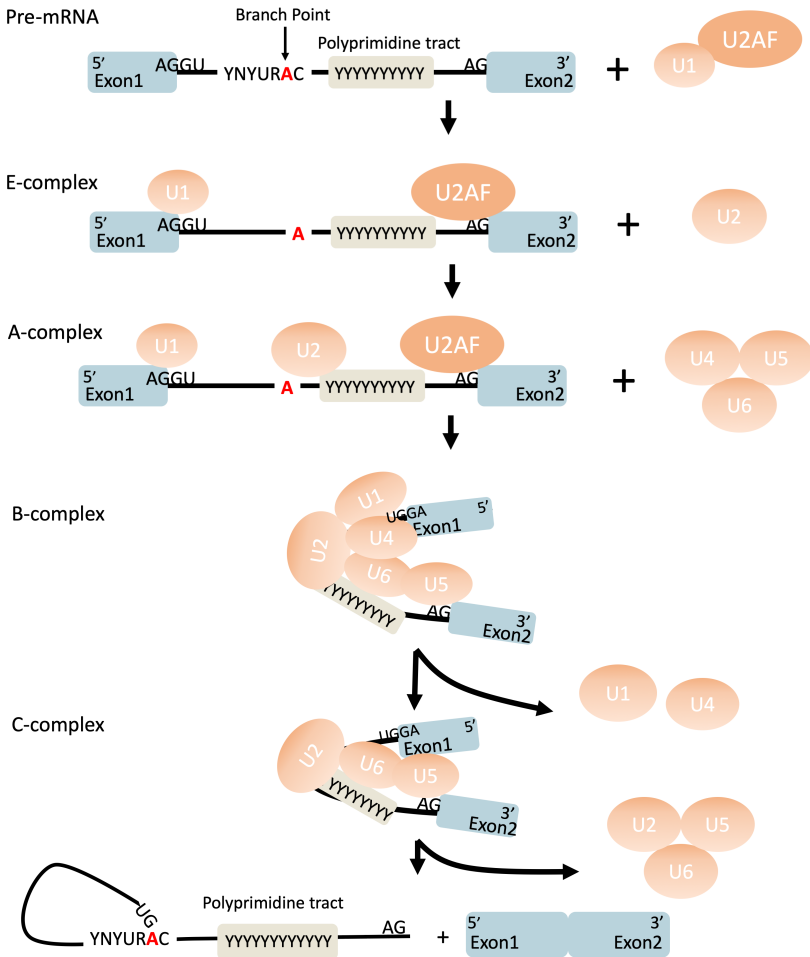


Figure 1.6 Schematic representation of pre-mRNA splicing and spliceosome assembly.
The model demonsteated that the spliceosome assemble is initatiated by snRNP U1 and U2 snRNP Auxiliary Factor (U2AF) subunit 65 recoginze the 5’- and 3’- splice site from the intron and forming the E-complex. The A-complex formed by snRNP U2 binds to the branche point, followed by additon of the snRNP U4/U6,U5 to form the B-complex. Re-modelling of B-complex will release the snRNP U1, U4 to form the splicing-active C complex.

Splicing is conducted by a large macromolecular complex called spliceosome through two transesterification steps **(Figure 1.6)**. During the first step, the highly conserved adenosine (A) at the branchpoint attacks the 5' splice site and causes the cleavage of the 5' exon from the intron. The second step is the detached 5'end of the intron attacks of the 3' end, then the two exons are ligated to each other and leave the intron in the form of a lariat [156]. The spliceosome assembly requires numerous proteins and five Uridine-rich small nuclear ribonucleoproteins (snRNPs): U1, U2, U4/U6 and U5. During assembly, the U1 snRNP first binds to the 5'splice site through the RNA-RNA interactions, while the 3'splice site is bound by the BP binding protein SF1 and the U2 snRNP Auxiliary Factor (U2AF) subunit 65 that binds to the polypyrimidine tract. The other subunit of U2AF, U2AF35 binds to the AG dinucleotide at the 3' splice site [157]. The U1 and U2AF together with two intron ends form the earliest complex called the E (early) complex. Then the U2 snRNP binds to the branch point through RNA-RNA interactions in an ATP-dependent manner to form the A complex [158]. B complex is formed when the U4/U6 and U5 binds to the A complex with the U2/U6 helix II formed by U2 and U6 snRNPs. Then the B complex undergo a re-modelling to release both U1 and U4 snRNPs to form complex C, which is the active spliceosome that conduct the 2-step splicing reaction [159].

1.6 Splicing and polyadenylation factors

1.6.1 Serine/arginine (SR) rich proteins

It is also important to note that splicing site cannot be recognized without adjacent RNA elements to enhance splice site strength. Regulation of alternative splicing event plays an important role in gene regulation. There is always competition between different splice sites, and the efficiency of splicing site usage is different under different conditions. The trans-acting factors and cis-acting elements are therefore used to regulate splice site activity. The cis-acting elements include exonic splicing enhancers (ESE), exonic splicing silencers (ESS), intronic splicing enhancers (ISE) and intronic splicing silencers (ISS) [160]. Generally, there are two groups of trans-acting splicing factors involved in the alternative splicing: the serine/arginine (SR) rich proteins and the heterogeneous nuclear ribonucleoproteins (hnRNPs) **(Figure 1.7)** [161-164]. Although many SR proteins function as splicing activators while many the hnRNPs are splicing repressors, it is not always the case. Both SR proteins and hnRNPs can function as splicing activators or repressors in position-dependent manner [165].

Figure 1.7 Schematic representation of general model of SR proteins and hnRNP proteins in splicing regulation.
ISE, intronic splicing enhancer; ESE, exonic splicing enhancer; ESS, exonic splicing sliencer; ISS, intronic splicing sliencer.

The SR proteins are a family of proteins that are involved in many cellular processes including mRNA exportation [166], translation [167, 168], RNA stability, activation or inhibition of constitutive and alternative splicing [169-171]. SR proteins are highly conserved and consist of at least one RNA recognition motif (RRM) and a serine/arginine-rich domain (RS domain) [171]. The RS domain is unique for SR proteins and responsible for targeting them to nuclear speckles [172, 173]. SR proteins usually recognize and bind to the cis-acting RNA sequence called exonic splicing enhancers (ESEs) and function as activators of splicing. They also contribute to a process called exon definition. This process is done either by interacting with other splicing factors and promoting the recruitment of spliceosomal components to the 3' splice site [174], or by contacting the pre-mRNA within the functional spliceosome through the RS domain of SR proteins [175, 176]. Despite that SR proteins generally act as activators of splicing, studies have shown that SR proteins can function as negative regulators of splicing by interfering with the functional assembly of the spliceosomes [177].

1.6.2 Heterogeneous nuclear ribonucleoproteins (hnRNPs)

Heterogeneous nuclear ribonucleoproteins (hnRNPs) are a large family of RNA binding proteins that common in many cell types. hnRNPs have complex functions based on the fact of the structural variation of hnRNPs in terms of different domain combinations and arrangements. hnRNPs are involved in numerous nuclear and cytoplasmic mechanisms such as regulation of constitutive and alternative splicing [178, 179], regulation of mRNA stability [180, 181], activation or inhibition of mRNA translation [182, 183] and mRNA transportation [163]. It has been reported that the expression levels of many hnRNPs are altered in various types of cancer, suggesting that hnRNPs might play a role in tumorigenesis. Despite cancer, hnRNPs have been linked to different neurodegenerative diseases including spinal muscular atrophy (SMA), Alzheimer's disease (AD) and amyotrophic lateral sclerosis (ALS) [184].

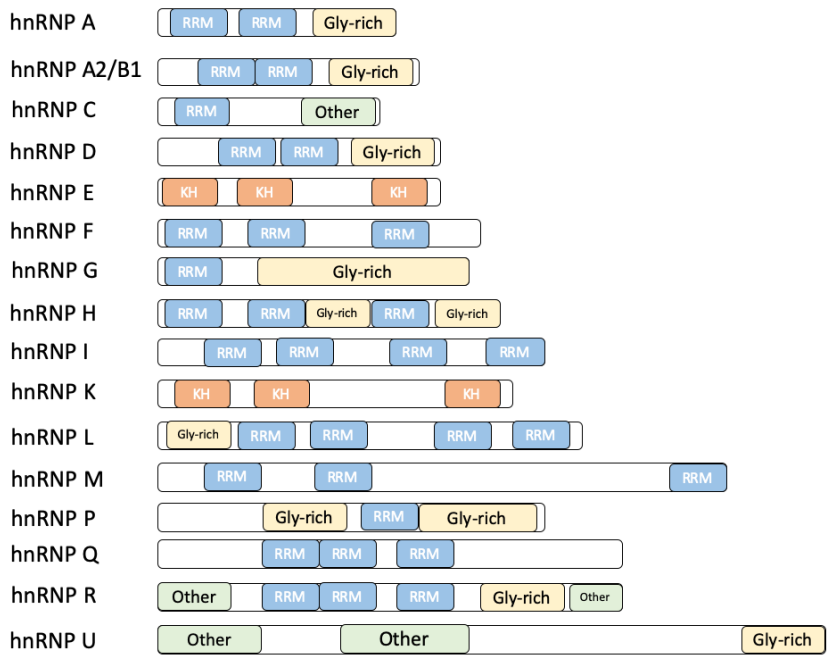


Figure 1.8 Schematic representation of the members in hnRNP family.
RRM, RNA recognition motif; KH, K homology domain; gly-rich, glycine rich auxiliary domain.

The hnRNPs generally consists of multiple domains linked by the linker sequence at various length. However, the domain composition is highly diverse in the hnRNP family **(Figure 1.8)**. There is not a single domain that is presented across the whole family [184]. There are four types of RBDs found in hnRNPs: the RNA recognition motif (RRM), the glycine-rich domain with RGG box, the quasi-RRM and the KH domain [184]. RRM is the most abundant domain and consists of two degenerate consensus sequences named RNP-1 and RNP-2, that are responsible for direct interactions with RNA. The RRM can also bind to single stranded nucleic acids of various length in a sequence independent manner, while the linker region, and the C- and N-terminal regions can increase the specificity of the RNA-binding to specific sequences. Thus, RRMs can both interact with nucleic acid specifically and non-specifically [179, 184, 185]. Some hnRNPs contain RGG boxes (repeats of Arg-Gly-Gly tripeptides) or auxiliary domains such as glycine-rich, acidic or proline-rich domains [163]. In hnRNP U, the RGG box responsible for its binding to mRNA and ssDNA [186]. The quasi-RRM was originally found in hnRNP F/H which contains an extra beta loop and lacks the RNP consensus sequences [187]. hnRNP E/K bind to RNA through a unique domain named KH (K-Homology) domain[188].

hnRNPs are among the most abundant proteins in the nucleus. Most hnRNPs contain a conventional nuclear localization signal (NLS) and locate predominantly in the nucleus, but there are exceptions like hnRNP A1, which shuttles between the nucleus and the cytoplasm [163]. It is also important to know that hnRNPs frequently undergo post-translational modifications including methylation, phosphorylation and ubiquitination that may affect their functions and ability to bind to RNA. These modifications often result in changes of hnRNPs biological activity and subcellular localization [189].

1.6.2.1 hnRNP G

hnRNP G, also known as RBMX, which refers to its position on the X chromosome. hnRNP G is ubiquitously expressed in various tissues but differs in concentrations. It has a homolog called RBMY, which is located on the Y chromosome, and is primarily expressed in testis [190, 191]. In addition, there is another testis-specific gene called hnRNP G-T. unlike RBMX and RBMY, hnRNP G-T is autosomally encoded by chromosome 11 [192]. Both RBMY and hnRNP G-T are necessary for male spermatogenesis [192].

The hnRNP G protein contains an RRM at its N-terminal and an auxiliary RNA binding domain at its glycine rich C-terminus [193]. hnRNP G does not have a classic nuclear localization signal (NLS), but has a novel transcript targeting domain (NTD) that is required for both hnRNP G nuclear localization as well as targeting of hnRNP G to nascent transcripts [194]. hnRNP G is a strictly nuclear protein and can regulate the mRNA alternative splicing [195]. Studies have showed that hnRNP G can affect alternative splice site selection mainly through two pathways. One is through protein-protein interaction with other splicing regulatory factors through its C terminus, the other way is by direct binding to RNA through RRM [195]. Other studies showed that hnRNP G can act as a splicing factor to regulate inclusion or exclusion of the alternative exons. hnRNP G can active the inclusion of the exon 7 in the SMN gene [196] and inhibit the inclusion of the exon 10 in the *tau* gene [197]. Depending on the situation, hnRNP G can also act synergistically or antagonistically with splicing factor Tra2β [196, 198]. In oral squamous cell carcinomas, hnRNP G can also suppress the growth of the cancerous cell, while mutated hnRNP G lost this function, suggesting that hnRNP G could function as a tumor suppressor [199].

Recent studies also showed that hnRNP G plays an important role in DNA damage response through either the homologous recombination (HR) pathway or non-homologous end-joining pathway (NHEJ) [200, 201]. For the HR pathway, hnRNP G promotes the DNA damage response by accumulating at the site of the DNA lesions in a poly (ADP-ribose) polymerase 1 (PARP-1)-dependent manner, and by promoting the expression of repair protein BRAC2 [200]. For the NHEJ pathway, hnRNP G binds to the DNA ends through its RRM domain and protects the DNA

ends by slowing down the exonuclease activity. Mutations in RRM domain resulted in loss of the protective ability [201].

1.6.3 Human antigen R (HuR)

The RNA binding protein HuR is a well-studied mRNA stabilizing factor encoded by the *Elavl 1* gene and it shuttles between nucleus and cytoplasm [202]. HuR contains three RNA binding domains (RRMs) connected with a hinge region **(Figure 1.9)**. The two N-terminal domains bind to AU-rich elements, where the C-terminus domain responsible for poly(A) tail binding and is believed involved in protein-protein interaction to increase the stability against degradation [203, 204]. The hinge region responsible for protein shuttling between nucleus and cytoplasm.

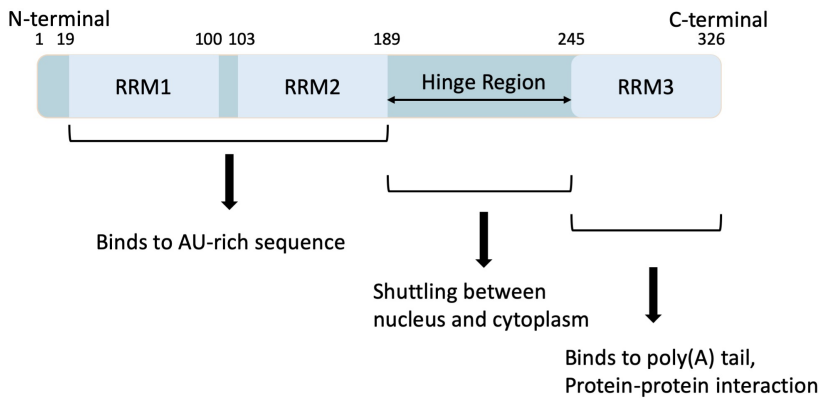


Figure 1.9 Schematic representation of protein HuR.
HuR consisits of three RNA binding motif (RRM), the RRM1 and RRM2 responsible for binding to AU-rich sequence. RRM3 responsibles for binding to poly(A) tail and protein-protein interaction. A hinge region locates between RRM2 and RRM3, responisble for shuttling between nucleus and cytoplasm.

HuR is also involved in various biological processes including RNA splicing, RNA exportation, cell cycle progression, DNA damage response, apoptosis and oncogenesis [205-207]. HuR levels are usually increased in tumors compared to normal cells due to the accumulation of HuR protein in the cytoplasm. HuR can act as an alternative mRNA splicing regulator by interfering with exon definition through direct inhibition of association of U2AF65 with 3' splice site [208]. Besides, HuR has been found overexpressed in several tumors and can affect alternative polyadenylation. Overexpression of HuR in HeLa cells selectively block

polyadenylation at poly(A) site in U-rich containing sequences. The study showed that HuR can block the interaction between RNA and poly(A) factors by binding to RNA and poly(A) factors such as CstF64 simultaneously and in turn block the polyadenylation [209].

1.6.4 Cleavage and polyadenylation factor 5 (CPSF5)

Cleavage and polyadenylation factor 5 (CPSF5) is a component of the cleavage factor Im (CFIm) complex, also known as NUDT21. CPSF5/ NUDT21 is an RNA binding protein that activates the pre-mRNA 3'-end cleavage and polyadenylation process [210]. CPSF5/ NUDT21 together with CPSF6 and/or CPSF7 form the CFIm complex and binds to the consensus sequence 5'-UGUA-3' to enhance the pre-mRNA 3'-end processing [211]. CPSF5/NUDT21 also contributes to recruit multiple protein complexes onto the cleavage and polyadenylation signals [212].

1.7 Regulation of HPV16 gene expression

HPV16 gene expression is tightly regulated by cell differentiation. There are two promoters in the HPV16 genome: the early promoter P97 and the late promoter p670. HPV16 lacks its own replication machinery and is totally dependent on the host for replication and gene expression. In the basal layer of epithelium, HPV16 uses the host cellular transcription machinery to activate viral transcription from the early p97 promoter. When cells start to differentiate in the upper layers, the differentiation-dependent late promoter p670 is activated and the early promoter p97 is inhibited [213]. As mentioned earlier, the HPV16 genome is separated into an early and a late region by the early polyadenylation site pAE. All the early mRNAs encoded from either p97 or p670 are polyadenylated at pAE, whereas the late mRNAs such as L1 and L2 are polyadenylated at pAL. In order to switch HPV16 gene expression from early to late, there are three requirements that need to be fulfilled. First, inhibition of early promoter p97 and activation of the late promoter p670. Second, inhibition of the early polyadenylation signal pAE and activation of the late polyadenylation site pAL. Third, activation of the L1 specific splice sites [44].

1.7.1 Regulation of HPV16 early and late polyadenylation

The HPV16 early polyadenylation site pAE controls all early gene expression and prevents premature expression of the late gene expression at the early stage of the HPV life cycle. Inhibition of the pAE activates other cryptic polyadenylation sites

locate upstream of pAE, demonstrating the existence of strong polyadenylation regulatory elements around pAE that force early polyadenylation to occur [44]. The polyadenylation regulatory elements are found both upstream and downstream of pAE. The HPV16 early 3'UTR (eUTR) region stimulates the efficiency of the early polyadenylation [214]. Deletion of the 3'UTR causes readthrough into the late region encoding L1 and L2 proteins. This regulatory effect is probably mediated by different regulatory proteins including polyadenylation factor FIP1, hnRNP C1/C2, polyadenylation pyrimidine-tract-binding protein 1 (PTB or hnRNP I) and cytoplasmic polyadenylation element-binding protein1 (CPEB1) [214]. Another regulatory element located downstream of pAE in the L2 coding frame, interacts with cleavage stimulation factor 64kDa subunit (CstF 64), which in turn stimulates polyadenylation at pAE [215].

Some other proteins are also shown to regulate the pAE efficiency. hnRNP H binds to GGGs motif in the L2 region downstream of pAE and stimulates polyadenylation at pAE [215]. Many DNA viruses including HPV, encode their own viral factors that regulate the switch from the early stage to the late stage of virus life cycle. In HPV16, E2 can inhibit early polyadenylation of pAE by preventing the assembly of the CPSF polyadenylation complex, thereby inducing the HPV16 late gene expression [216]. Simultaneously, E2 also shuts down the HPV16 early promoter. Together these two effects of E2 pushes HPV16 into the late stage of life cycle.

Regulatory elements for HPV16 pAL are located in the late 3'UTR (lUTR), and consists of multiple copies of the (A/G)U_{3-5}(A/G) motifs. lUTR inhibits the HPV16 gene expression either by decreasing the stability of newly synthesized mRNAs or by inhibiting late polyadenylation and mRNA translation [217, 218]. lUTR also contains multiple weak U1snRNA binding sites that can further inhibit HPV16 late polyadenylation, mRNA export and translation. However, this inhibitory effect can possibly overcome by late mRNAs interacting with shuttling proteins such as HuR [219], SRSF1 [220] and hnRNP A1 [221] to enhance the mRNA exportation.

1.7.2 Regulation of HPV16 splice sites.

The oncogenic activities in HPV16 are mediated by oncoprotein E6 and E7. E6 and E7 mRNAs are derived from the same polycistronic pre-mRNA but undergo different alternative splicing, the transcriptional and translation efficiency are also regulated by alternative splicing. There are 9 isoforms of spliced E6/E7 transcripts have been identified in HPV16 [2, 102, 222, 223], E6 and E6*I/E7 believed are the major isoforms that are translated into E6 and E7 oncoprotein respectively. The difference between E6 and E6*I/E7 mRNA by including or excluding the intron between SD226 and SA409. Inclusion of the intron results in E6 mRNA while

exclusion of the same intron results in E6*I/E7 mRNA, thus regulation of splicing between SD226 and SA409 is critical for oncogene E6/E7 balance. Studies showed that treatment with epidermal growth factor (EGF) in HPV16 infected cells results in the intron inclusion and give rise to E6 mRNA production while the other study showed depletion of EGF shifted the balance and promoted the E6*I/E7 mRNA splicing [224]. Besides, the splicing factor hnRNP A1 and hnRNP A2/BI also showed the ability to stimulate the splicing between SD226 and SA409 [224].

SA3358 is the most commonly used 3' splice site of HPV16. It is used to produce most HPV early mRNAs including oncogene E6/E7 mRNAs as well as late mRNAs [225]. Inactivation of SA3358 induces the premature expression of HPV16 L1i mRNAs. However, SA3358 is a poor 3' splice site and is totally dependent on the splicing enhancer downstream of SA3358 [226]. This splicing enhancer contains multiple binding sites for serine-and arginine-rich splicing factor 1 (SRSF1 or ASF/SF2) [220, 227]. SRSF1 binds to the enhancer and enhances the splicing efficiency of SA3358. Inactivation of these binding sites resulted in skipping of SA3358 [226]. SRSF1 has been identified as a proto-oncogene, high levels of SRSF1 enhances the expression of E6/E7 mRNAs and induces cell hyperproliferation in HPV16 infected cells [44]. However, SA3358 is also needed for late mRNAs production, study showed moderate levels of SRSF1 might induce premature late gene expression [220]. A splicing silencer is located downstream of the SRSF1 binding sites and can interact with SRSF3 (also known as SRp20) and SRSF6 (also known as SRp55) to inhibit the splicing activity of SA3358 [228, 229].

HPV16 splice sites SD3632 and SA5639 are uniquely used for L1 mRNAs production. These two sites are highly regulated in a cell differentiation-dependent manner to prevent premature HPV16 late mRNA expression. Therefore SD3632 and SA5639 are actively used in terminally differentiated cells and are strongly suppressed in mitotic cell and cervix cancer cells, The two HPV16 splice sites are suppressed by splicing silencers that interact with various cellular factors [44]. SD3632 is normally suppressed by the splicing silencers located both upstream and downstream of the splice site [226]. The suboptimal 3' splice site SA5639 is under control of a 17-nucletotide splice enhancer downstream of SA5639, but more importantly suppressed by splicing silencer elements is followed by purine-rich sequence that interact with hnRNP A1 and hnRNP H [230]. This purine-rich sequence acts as splice silencer that suppresses the splicing activity of SA3358 and overrides the upstream splicing enhancer [230].

1.8 Nucleosides and their analogues in antiviral and anticancer treatments

1.8.1 Nucleosides

Numerous studies have shown the importance of the purine nucleoside adenosine in various medical conditions. Adenosine provides protection against brain and spinal cord [231] as well as ischemia-reperfusion-injuries in the heart [232]. Adenosine is also known to participate in several neurodegenerative disorders including Alzheimer's disease, Huntington's disease and schizophrenia [233-235]. In addition, adenosine has been shown to play an important role in cancer. For example, adenosine presents and accumulates chronically in the tumor microenvironments. The high levels of adenosine together with other mechanisms result in an immunosuppressed environment that favors tumor growth [236-238] and actives the immune suppression pathways to escape from immune surveillance [239-241].

Adenosine is present in all body fluids and tissues. The concentrations of extracellular adenosine are rather constant under normal conditions. But the concentrations can rapidly increase in response to stressful conditions such as hypoxia, inflammation and trauma [242]. The high levels of extracellular adenosine in turn initiate signaling cascades by interacting with one or more of four different adenosine receptors (AR), AR1, AR2A, AR2B and AR3 **(Figure 1.10)**. Since adenosine can trigger a tremendous and broad range of downstream effects, adenosine is also called "the signal of the life" [243]. Adenosine signaling is one of the key metabolic pathways that regulate tumor immunity [244], which makes adenosine a perfect target for anti-cancer therapy. Depending on the tumor types and the receptors presented on the cell surface, the tumor cell can either respond to adenosine by increasing survival and proliferation or by apoptosis [245].

Due to the fact that adenosine initiates signaling cascades for numerous pathways, the concentrations of adenosine needs to be tightly regulated [246]. Adenosine can be transported into cells and be utilized as building blocks for ATP. Two types of nucleoside transporters are known to transport adenosine in and out of cells: the concentrative nucleoside transporters (CNT) and equilibrative nucleoside transporters (ENT) **(Figure 1.10)**. Once adenosine enters the cells, two major enzymes are responsible for adenosine metabolism: adenosine kinase (ADK) and adenosine deaminase (ADA) **(Figure 1.10)**. ADK phosphorylates the adenosine into AMP followed by metabolized into inosine while ADA deaminizes adenosine directly into inosine [247, 248]. Adenosine can also be metabolized by S-Adenosyl Homocysteine (SAH) -hydrolase, which hydrolyzes the SAH to adenosine mainly

in heart and is a reversible reaction that usually favors the SAH hydrolyzations since adenosine is normally quickly metabolized by other enzymes.

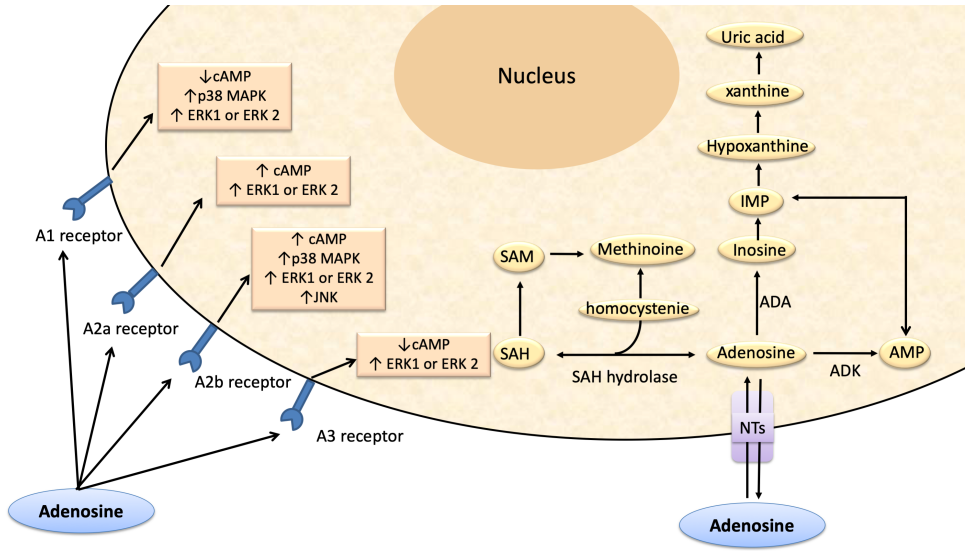


Figure 1.10 Schematic representation of adenosine metabolism and signalling pathway.
Extracellular adenosine can bind to four different adenosine receptors and regulate the cAMP activity. Extracellular adenosine can be transported into cells via nulceoside transporters (NT), followed by deaminased to inosine and AMP by adenosine deaminase (ADA) and adenosine kinse (ADK) respectively.

Guanosine is also a purine nucleoside that exerts protective effects in degenerative diseases and cell metabolism. Guanosine can also function as an extracellular signaling molecule that regulates the glutamatergic activity [249, 250]. Guanosine also shows protective effects against glutamate mediated cell damage by activating the phosphoinositide 3-kinase (PI3K) and mitogen-activated protein kinase/extracellular signal-regulated kinase (MAPK/ERK) pathways [251, 252]. Even though the exact guanosine receptors have not yet been identified, guanosine has been reported could potentially interact with adenosine receptors [252].

1.8.2 Nucleoside analogues

During the past decades, there are numerous reports of nucleoside analogues in anticancer and antiviral treatments. However, the cure for cancer and antiviral treatments still largely elusive despite the fact that various antiviral and chemotherapeutics agents are available. Therefore, intensive studies of nucleoside analogues are needed not only for understanding the mechanism of cancer but also

for developing novel therapeutic strategies for both anticancer and antivirus treatments.

Nucleoside analogues are a big diverse pharmacological family that includes antiviral agents as well as cytotoxic compounds for cancer treatment [253]. Nucleoside analogues are antimetabolites that can interfere with the synthesis of nucleic acids by incorporating into DNA and RNA to cause chain termination, or by interfering with enzymes that are involved in nucleic acid synthesis, or by modifying the nucleoside metabolism [253].

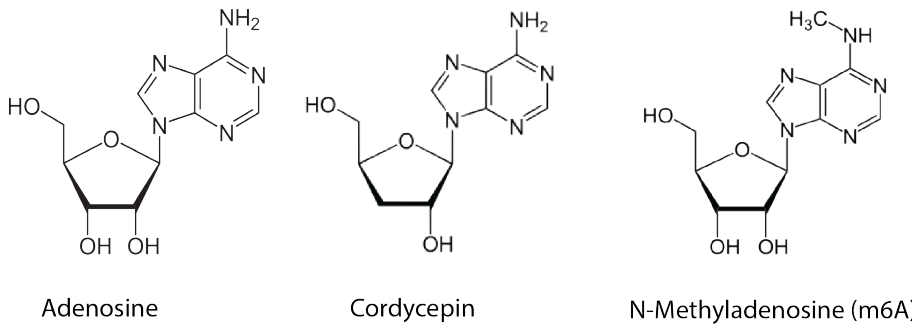
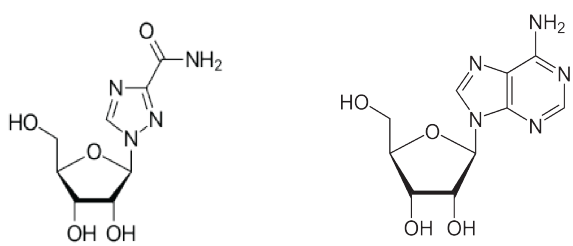


Figure 1.11 Chemical structure of adenosine and adenosine analogues: cordycepin and N-Methyladenosine (m6a), guanosine and guanosine analogue ribavirin.

Cordycepin, also known as 3′-deoxyadenosine, was first isolated from *Cordyceps miltaris* and has been used as Chinese traditional medicine for centuries **(Figure 1.11)**. Cordycepin is an adenosine analogue that lacks a 3' hydroxyl group and it is known to function through multiple mechanisms including inhibition of RNA biosynthesis [254], interfering with mTOR signaling pathway [255] and induction of apoptosis [256]. During transcription, some enzymes mistake cordycepin for adenosine and incorporate cordycepin into the transcript, resulting in premature termination of the mRNA due to cordycepin the lack of the 3'hydroxl group [254]. Recent studies showed cordycepin has apoptotic effects by inhibiting mRNA polyadenylation of apoptotic mRNAs such as the Cdkn1a mRNA [255]. Cordycepin can also inhibit mRNA polyadenylation thereby affecting mRNA stability and nuclear exportation [257]. Cordycepin also been suggested to act as an anticancer agent as it can induce apoptosis via multiple pathways. Cordycepin induces apoptosis through mitogen-activated protein kinases (MAPK) signaling pathway and increases the expression levels of c-Jun NH2-terminal kinase (JNK) p38, and Bcl-2 pro-apoptotic proteins in human colorectal cancer cells [258]. In NB-4 cells, cordycepin leads to release of mitochondrial cytochrome c to cytosol and results in activation of caspase-9 and caspase-3, leading to apoptotic cell death [259].

Cordycepin is also known have anti-inflammatory activity. Cancer cells can express various cytokines that are involved in inflammatory response. Cordycepin can suppress NF-κB activation and Akt- and p38- phosphorylation, which suppress the expression levels of the inflammatory mediators CO-2 and TNF-α [259]. Furthermore, it has been shown that cordycepin has ability to reduce tumor formation in mice [260]. Therefore, cordycepin has the potential to act as an anticancer agent for cancer treatment. However, further studies are needed to give in-depth understanding about cordycepin-mediated effects.

N6-methyladenosine (m6A) is an adenosine analogue that has a methyl group at the N6 position of adenosine **(Figure 1.11)**. The m6A modification is one of the most studied mRNA modifications and is a critical regulator of gene expression. The m6A modification was first identified in mammals in early 1970s, and was later found to be an abundant mRNA modification present in most eukaryotes and some viruses [261-263]. The m6A methylation affects numerous steps of RNA metabolism including splicing, translation, translocation and decay. The m6A modification is catalyzed by writers such as RNA methyltransferases-like protein 3 (METTL3), METTL14 and METTL16 and interacts with m6A readers including YTH domain-binding protein 1 (YTHDF1), YTHDF2 and YTHDF3 and is removed by m6A easers like alkB homologue 5 (ALKBH5) and fat mass and obesity-associated protein (FTO). Recent studies showed that m6A modification also plays an important role in cancer and tumor progression including cervical cancer [264]. However, emerging studies recently suggested that m6A modification is a double-edged sword for cancer. Deregulation of m6A in some cancers such as endometrial cancer could lead to activation of Akt signaling pathway and promote tumor progression [265]. But in AML, enhanced m6A modification of certain mRNAs including PTEN and MYC increased their stability and translation, leading to tumorigenesis and progression to AML [266, 267].



Guanosine Rlbavirin

Figure 1.12 Chemical structure guanosine and guanosine analogue ribavirin.

Ribavirin is a guanosine analogue **(Figure 1.12)** that has been used as an antiviral agent to treat numerous virus infections including Respiratory Syncytial Virus (RSV) and Hepatitis C Virus (HCV) [268]. Ribavirin can be transported into cells by nucleoside transporters and metabolized to triphosphate nucleotide by cellular kinase including adenosine kinase (ADK) [269], which is the active form of ribavirin that confers most of the antiviral activity [270]. There are few mechanisms of action proposed for ribavirin. First, ribavirin can be recognized by viral polymerases and cause inhibition of mRNA elongation by chain termination or by preventing binding to the polymerase of other nucleotides that are necessary for viral genome replication [271, 272]. A second mechanism by which ribavirin may inhibit viral replication is by RNA mutagenesis. Ribavirin can be incorporated into viral genomes and pair with both cytosine and uracil. Therefore, ribavirin can potentially increase the mutagenesis rate during viral genome replication, but this mechanism still remains controversial [273]. The other mechanism is inhibition of inosine monophosphate dehydrogenase (IMPDH). IMPDH is an essential enzyme for guanosine nucleotide *de novo* synthesis. Recent studies have shown that ribavirin can inhibit IMPDH and decrease the levels of GTP, which is needed for viral genome replication [274]. Besides the anti-viral effects, ribavirin has anti-cancer effects, at least in acute myeloid leukemia (AML) patients by inhibiting the activity of the oncogene and eukaryotic translation initiation factor eIF4E [275, 276]. The expression levels of eIF4E are increased in approximately 30% cancers, and they are elevated in nearly 100% of all head and neck cancers [276], suggesting that ribavirin could be beneficial beyond AML.

2 Aim of the Thesis

The general aim of this thesis was to investigate human papillomavirus type 16 early and late gene regulations.

The specific aims were:

1. To investigate if nucleosides including adenosine and guanosine can affect HPV16 gene expression.

2. To investigate the effect of nucleoside analogues on HPV16 late gene expression.

3. To investigate the role of splicing factor hnRNP G in regulation of HPV16 late L1 gene expression.

4. To investigate the role of splicing factor hnRNP G in the regulation of HPV16 E6 and E7 oncogene expression.

3 Material and Methods

3.1 Cell lines

The reporter cell line C33A2 is derived from the HPV16 negative cervical cancer cell line C33A. C33A2 has the sub-genomic HPV16 plasmid pBelsLuc stably integrated into chromosome [229]. The pBelsLuc plasmid contains HPV16 genome except the oncogenes E6 and E7, under control of the Cytomegalovirus (CMV) promoter. The late L1 gene was partially replaced by the poliovirus type 2A internal ribosomal entry site (IRES) followed by the secreted luciferase reporter gene (sLuc). Under normal conditions the C33A2 cells produce spliced early mRNAs including E4, whereas the late genes are only expressed upon induction. The C33A2 and HeLa cell lines were cultured in Dulbecco's Modified Eagle Medium supplemented with 5% heat inactivated bovine calf serum and penicillin/streptomycin from GE Healthcare Life Sciences HyClone Laboratories.

The LU-HNSCC-26 (HN26) cell line was isolated from a 48-year-old patient that diagnosed with HPV16 positive squamous cell tonsillar carcinoma [230]. The cell line contains episomal HPV16 genome and were cultured in RPMI-1640 medium supplemented with 10% heat inactivated bovine calf serum with 1nmo/Lsodium pyruvate, 1x MEM non-essential amino acids and 20µg/mL gentamicin from GE Healthcare Life Sciences HyClone Laboratories.

3.2 Plasmids

The following plasmids have been previously described: pHPV16AN [277, 278], pC97EL [230], pBelsLuc [230, 277], pBEL [230], p3*sLuc [277], pBELMsLuc [230, 277], pBspliceM [278], PT7 [226], PT8 [226], pCMVSRp30c [229], pCMVSRp30DRS [229], pCMVhnRNP G [216], pCHuR [216], pC086 [279], pNUDT21 (RC200936) was originate purchased from OriGene Technologies.

The plasmid pCMVsLuc was constructed by digestion of pBELsLuc plasmid with restriction enzymes SalI and MluI, followed by re-ligation to CMV promoted empty plasmid that generates secreted luciferase. The p7xWTsLuc and p7xMUTsLuc were generated by cloning of two annealed oligos with either seven copies of the wild

type 8-nucleotide element (p7xWTsLuc) or the mutant 8-nucleotide element (p7xMUTsLuc) into the plasmid pBELMsLuc, which contains a BssHII site at nucleotide position 3400 and HpaI site at nucleotide position 3521. To construct plasmids R3/pflag-hnRNP G, pGR1, pGR5, pGR6 and pGR7, the hnRNP G cDNA were amplified by PCR with primers: RRMS and RBDAS; RRMS and NTDAS; RRMS and RBDstartAS; RRMendS and RBDAS; and RRMendS and RBDAS respectively. For plasmids pX556A, pX856F, pX960A, pX1060A and pX1200A. The full HPV16 DNA sequence were amplified by PCR with primers B97S and X556AF, X856AF, X960A, X1060A or X1200A, respectively. The PCR products were cloned into pCR2.1-TOPO (Life technologies), followed by digestion with restriction enzymes PteI and XhoI, then subcloned into pC806 between CMV promoter and HPV16 polyadenylation signal.

3.3 Transfections

All transfections were carried out using Turbofect transfection reagents according to manufacturer's instructions (Fermentas). Briefly, the desired amount of DNA was diluted in 200µl of serum free medium, for every 1µg of DNA, 2µl of transfection reagents were added into the mixture of DNA/medium and the mixture was incubated at room temperature for 20 minutes before added into 60mm plates. Cells were harvested after 24 hours post transfection.

3.4 Secreted luciferase assay

The cell culture medium was harvested at desired time point. The level of *Metridia. longa* secreted luciferase (sLuc) secreted by cell was measured by using Ready-To-Glow secreted luciferase reporter assay (Clontech) according to the manufacture's protocol.

3.5 Chemicals

Chemicals used in this thesis are listed in the **Table 3.1**

Tabel 3.1.
List of chemicals

Name	Catalogue Number	Company
3′-Deazaneplanocin (SAH)	SML0305	Sigma
5’-azacitydine		
5′,2′-Dideoxyadenosine	D7408	Sigma
6-Thioguanosine	SML1296	Sigma
8-Aminoadenosine	SML0628	Sigma
Adenosine	A9251	Sigma
Alloxazine	A28651	Sigma
Caffeine	C0750	Sigma
CAMP	A9501	Sigma
CGS21680	C141	Sigma
Cordycepin	C3394	Sigma
N6-Cyclopentyladenosine (CPA)	C1530	Sigma
8-Cyclopentyltheophylline (CPT)	C102	Sigma
Cytidine	C4654	Sigma
Dihydrochloride Hydrate	V4265	Sigma
Dilazep Di-hydrochloride	D5294	Sigma
EHNA	E114	Sigma
Guanosine	G6752	Sigma
IB-MECA	I146	Sigma
Inosine	I4215	Sigma
Sb203580	S8307	Sigma
N6-Methyladenosine	S3190	Selleckchem
NBMPR	N2255	Sigma
NECA	E2387	Sigma
Oxypurinol	O6881	Sigma
Uracil	U0756	Sigma
VUF5574	V5888	Sigma

3.6 RNA extraction, RT-PCR and Real Time qPCR

The total RNAs were extracted using TRI Reagent (Sigma Aldrich) and Direct-zol RNA mini prep (Zymo Research) according to the manufacturer's instructions. 1μg of total RNA was subjected to cDNA synthesis using M-MLV Reverse Transcriptase (Invitrogen) together with random primers (Invitrogen) according to the manufacturer's protocol. 1μL of cDNA was used for RT-PCR or qPCR. Real time qPCR was performed by using MiniOpticon (Bio-Rad) with Sso Advanced SYBR Green Supermix (Bio-Rad) according to the manufacturer's protocol. All cDNA quantitations were normalized to house keep gene GAPDH. The primers used in this thesis is listed in **Table 3.2**.

3.7 3'Rapid amplification of cDNA ends (3'RACE)

3′ RACE was performed on total RNA extracted using TRI Reagent (SIGMA Aldrich Life Science) and Direct-zol RNA MiniPrep (ZYMO Research) according to manufacturer's protocols. 500ng of cytoplasmic RNA were subjected to reverse transcription in a 20μl reaction at 37°C by using M-MLV Reverse Transcriptase (Invitrogen) and oligo-dT primer followed by PCR amplification with desired primers.

3.8 *In vitro* Polyadenylation assay

The *in vitro* polyadenylation assay was performed by using the *Escherichia.coli* poly(A) polymerase (PAP). PAP catalyzes the addition of adenosine-5'monophosphate (AMP) from adenosine-5'triphosphate (ATP) to the 3' end of RNA in a template-independent manner. The synthetic RNA L3 served as template [216]. 3'-deoxyadenosine-5'-triphosphate (3'dATP) was included as positive control. 3'dATP, cordycepin and 8-aminoadenosine were used at the final concentration of 2mM. The mixture of PAP, ATP, RNA template and 3'dATP/ cordycepin/ 8-aminoadenosine were incubated at 37°C for 10 minutes, followed by heating at 70°C for 5 minutes then quickly chilled on ice to disrupt the RNA secondary structure.

Table 3.2. List of primers

Primers	Sequences	Primers	Sequences
757S	CGGTTGTGCGTACAAAGCACACACG	GAPDH-F	5'-ACC CAG AAG ACT GTG GAT GG-3´
773S	GCACACACGTAGACATTCGTACTTTG	GAPDH-R	5'-TTC TAG ACG GCA GGT CAG GT -3´
880AS	5'-GATCAGCCATGGTAGATTATGGTTTC-3´	L1AM-F	5'-CTT CTA CAA CCC CGA CAC-3´
880S	5'-GAAACCATAATCTACCATGGCTGATC-3´	L1AM´R	5'-TCC ATA GCA CCA AAG CCA -3´
97S	5'-GTCGACCTGCAATGTTTCAGGACCC-3´	L1AS	5'-GCAACATATTCATCCGTGCTTACAACC-3´
actin-F	5'-AACAAGAGGCCACACAAATAGG-3´	L2	5'-SAME AS ABOVE-3´
Actin-R	5'-CAGATGTACAGGAATAGCCTCCG -3´	NTDAS	5'-CTCGAGCTATGGTGGTGCATAATCTCTAGT-3'
ARA1 F	5'-TCCATCTCAGCTTTCCAGGC-3´	pAE-F	5'-CTTACTATTTTTTCTTTTTTATTTTCATA-3´
ARA1 R	5'-CTCGAACTCGCACTTGATCAC-3´	pAE-R	5'-TAGCCGATGCACGTTTTG-3´
ARA2A F	5'-ACCTGCAGAACGTCACCAAC-3´	pAL-F	5'-GCACAACATCAAGGGCATG-3´
ARA2A R	5'-CACAGGACGCGCTGTACGTG-3´	pAL-R	5'-GCAAATATAGTTTATATACAATGAATAACC-3´
ARA2B F	5'-TTCTGTGCAGTTGTTGGTGG-3´	PBGD-F	5'-GAAGATCTATGAGAGTGATTCGCGTGGGTACC-3'
ARA2B R	5'-AACGTGCTGGTCATCTGCGTGGTC-3´	PBGD-R	5'-GGAATTCTTAATGGGCATCGTTAAGCTGCCG-3'
ARA3 F	5'-GTAGTCCATTCTCATGACGGAAAC-3´	RBDAS	5'-CTCGAGCTAGTATCTGCTTCTGCC-3'
ARA3 R	5'-CTG CTT CAG CTG TCG TCG CG-3´	RBDstartAS	5'- CTCGAGCTAACCACTTGAGTAGAGATC-3'
B97S	5'-GGCGCGCAACTGCAATGTTTCAGGACCC-3'	RPLP0-F	5'-TGGTCATCCAGCAGGTGTTCGA-3´
CNT3 F	5'-GGT TCC CTA GGA ATC GTG AT-3´	RPLP0-R	5'-ACAGACACTGGCAACATTGCGG-3´
CNT3 R	5'-CCA ATT GAG TTC AGC GTT GGT-3´	RRMendS	5'-GCGCGCCACCATGGACTACAAAGACGATGACGACAAGGAACAAGCCACCAAACCA-3'
E4AS	5'-TGCTGCCTAATAGTTTCAGGAGAGG-3´	RRMS	5'-GCGCGCCACCATGGACTACAAAGACGATGACGACAAGGGAAAGCTCTTCATTGGTG-3'
E4S	5'-CCTCTCCTGAAACTATTAGGCAGCA-3'	X1060A	5'-GCTCGAGTCTCTGTTTCTGCCTGTGTTAAATAATCATTATC-3'
E4S-var	5'-CCTCTCCTGAAATTATTAGGCAGCG-3'	X1200A	5'-GCTCGAGAATCTAGGACTAATATTATTGTCTACACATCC-3'
ENT1 F	5'-GCCCACCAATGAAAGCCACTCTATC-3´	X556AF	5'-GCTCGAGTTACTTGTCGTCATCGTCTTTGTAGTCCAGCTGGGTTTCTCTACGTGTT-3'
ENT1R	5'-TGCCAGACCCAGACACAGGAAGAA-3´	X856AF	5'-GCTCGAGTTACTTGTCGTCATCGTCTTTGTAGTCTGGTTTCTGAGAACAGATGGGGCAC-3'
ENT2 F	5'-CTGGTCTGCCTGCGGTTCCTG-3´	X960A	5'-GCTCGAGAGCATCCCCTGTTTTTTTTTCCACTAC-3'
ENT2 R	5'-TTTCTTCCCCGCAGCACTCCA-3´		

3.9 Western blot

For total cell lysate, the cells were lysed with RIPA buffer (50 mM Tris, pH 7.8, 150 mM NaCl, 1% sodium deoxycholate, 0.1% SDS, 1% Triton X-100, 1mM DTT and protease inhibitors) followed by SDS-PAGE separation. Proteins then transferred onto nitrocellulose membranes, blocked with 5% skim milk in Tris-buffered saline (TBS) containing 0.1% Tween 20. The membranes then incubated with desired antibodies and secondary antibodies. The proteins were detected by using the Clarity Western ECL substrate (Bio Rad) or the Super Signal West Femto Chemiluminescence substrate (Pierce).

3.10 ssDNA/RNA-protein pull down assay

Nuclear extracts were prepared according to the procedure described previously [280]. Briefly, the cells were lysed by using lysis buffer A (10mM HEPES,1.5mM $MgCl_2$, 10mM KCl, 0.5mM DTT, 0.05% NP40 pH 7.9 and protease inhibitors) for cytoplasmic proteins, then using buffer B (5mM HEPES, 1.5mM $MgCl_2$, 0.2mM EDTA, 0.5mM DTT, 26% glycerol, pH7,9 and protease inhibitor) for nuclear proteins. The nuclear extracts were mixed with streptavidin-coated magnetic beads (Dynabeads M-280 Streptavidin, Invitrogen) bound with biotin-labelled single stranded DNA or RNA oligos in binding buffer (10 mM Tris, pH 7.4, 150 mM NaCl, 2.5 mM $MgCl_2$, 0.5% Triton X-100). The mixture was incubated at room temperature with rotation for 1 hour, followed by five times washing with wash buffer.

3.11 siRNA transfections

siRNA knock downs were carried out using DharmaFECT transfection reagents according to the manufacturer's instructions. Briefly, the siRNA was diluted to a final concentration of 40nM in 200µl serum free medium, and the mixture was added to 200µl of serum free medium with 5µl transfection reagent. The mixture was incubated at room temperature for 20 minutes before added to 6-well plate with sub-confluent cells. The secreted luciferase activity was measured at 48 hours post-transfections.

3.12 UV cross-linking and immunoprecipitation of RNA-protein complexes

The cells were cultured in 10cm dish and washed with ice-cold PBS, followed by UV-irradiated at 150 mJ/ cm2 in a Bio-link cross-linker (Biometra). Cells were lysed with 1ml of RIPA buffer and incubated on ice for 30 min with occasional vertexing to help lyse the cells. For immunoprecipitations, 1µg of the desired antibodies or normal mouse IgG were incubated for 2 hours at 4 °C in 0.5 ml of total cell lysate. 20µl (0.6mg) of Dynabeads Protein G (Invitrogen) were blocked with 1% BSA in RIPA buffer for 0.5 hour, followed by washing three times with RIPA buffer, then the beads were mixed with antibody-protein mixture followed by incubation at 4 °C for 1 hour. The beads were washed three times with RIPA buffer and RNA was eluted by phenol/chloroform extraction. The RNA was ethanol precipitated and dissolved in 20µl of ddH_2O. 10µl of immunoprecipitated RNA was reverse transcribed using M-MLV reverse transcriptase (Invitrogen) and random primers (Invitrogen) according to the manufacturers' protocol. 2µl of cDNA were subjected to PCR amplification.

4 Results

Nucleosides play a significant role in cell differentiation and tumorigenesis and have been used in antiviral and anticancer treatments. We screened a series of nucleosides and nucleoside analogues that either are used in cancer treatment or been reported to have anti-tumor effect. We tested these chemicals in our reporter cell line C33A2 to determine if they could affect HPV16 gene expression. The C33A2 cell line contains the sub-genomic reporter plasmid pBelsLuc stably integrated into the chromosome. Under normal conditions, only HPV16 early genes are expressed, but when HPV16 late gene is induced, the secreted luciferase is expressed and secreted into the cell culture medium. By measuring the enzyme activity of secreted luciferase, we could determine the HPV16 late gene expression levels.

Within this series of chemicals, there was one cytidine analogue named 5-azacytidine which is a DNA methyltransferase inhibitor that has been used as chemotherapeutic against acute myelogenous leukemia [281, 282]. When we treated the C33A2 cells with the 5-azacytidine (Batch 1), HPV16 late gene expression was induced up to 23-fold **(Figure 4.1A)**. However, this tremendous effect was completely gone when we tested the same compound but with different batch numbers from the same company (Batch 2 and Batch 3) or if we ordered the same substance from a different company. **(Figure 4.1B)**. The mass spectrometry analysis of Batch 1 "5-azacytidine" indicated that there were no detectable levels of azacytidine in that sample. Instead an unknown substance was present that was not detected in Batch 2, Batch 3 or in azacytidine purchased from another company. This suggested that the unknown substance in Batch 1 could potentially be the active compound that induced HPV16 gene expression rather than 5-azacytidine. From the mass spectrometry analysis, we were able to get the composition and molecular weight of the unknown compound. Mass spectrometry was performed on a Q-TOF MS at Stenhagen Analyslab AB, Mölndal, Sweden. Based on the MS/MS data, furanose seems to be present in the molecule in all batches, but the nitrogenous base appears different in the molecule in Batch 1. It could be either $C_8H_9NO_2$ or $C_6H_7N_4O$, but since azacytidine contains four nitrogen and one oxygen, the latter seems more likely, suggesting that the composition of the unknown molecule might be $C_{11}H_{15}N_4O_5$, with molecular weight of 283 g/mol. This substance could be methyl inosine, but this substance was not sold by the company in question. We went through the various nucleosides or nucleoside analogues sold by the company and

we found that only guanosine had exactly same molecular weight as the unknown substance but with different composition $C_{10}H_{13}N_5O_5$ **(Table 4.1)**. Due to lack of the exact structure of the unknown substance, we investigated if guanosine could have effect on HPV16 late gene expression.

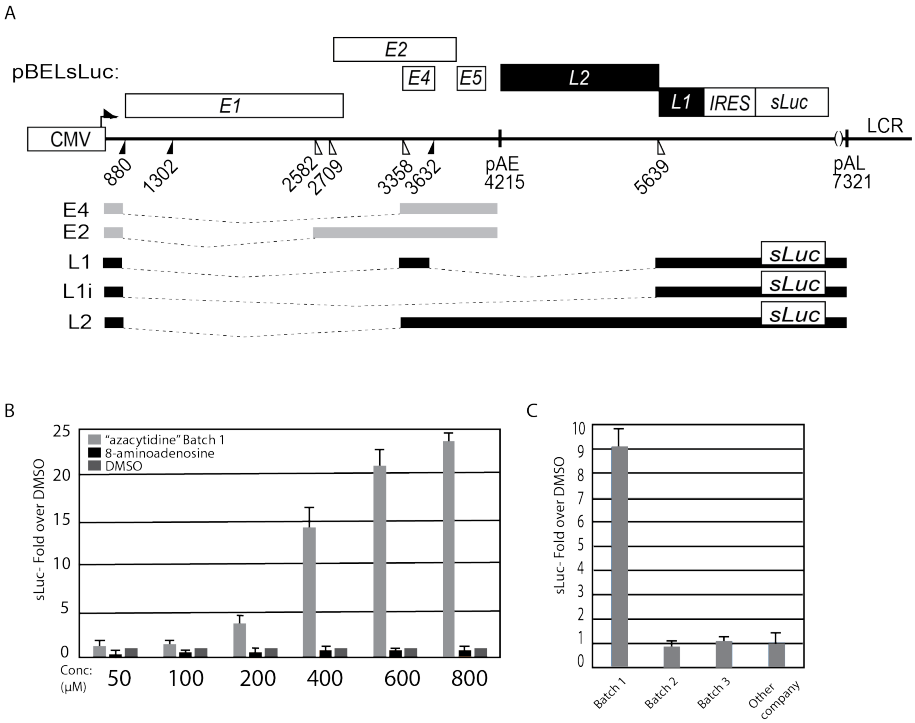
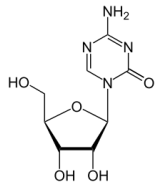
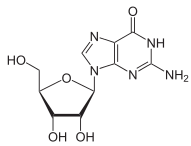
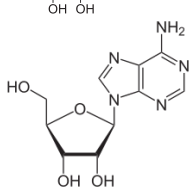
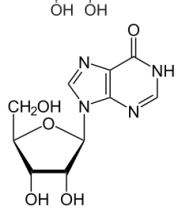
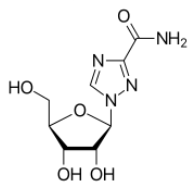


Figure 4.1 A. Schematic representation of the pBELsLuc reporter plasmids.

A. Transcription of the HPV16 sequences in the pBELsLuc driven by the human cytomegalovirus promoter (CMV). The sLuc gene inserted into the L1 region is preceded by the poliovirus 2A internal ribosome entry site (IRES). HPV16 E2 and E4 mRNAs produced by pBELsLuc plasmid are indicated in grey and HPV16 late mRNAs encoding sLuc are indicated in black. **B.** Secreted luciferase enzyme activity (sLuc) in the cell culture medium of reporter cell line C33A2 treated with various concentrations of "azacitidine" batch 1, 8-aminoadenosine and DMSO. 8-aminoadenosine serves as negative control. **C**. Secreted luciferase enzyme activity (sLuc) in the cell culture medium of reporter cell line C33A2 treated with azacitidine with different batches and other company.

Tabel 4.1.
List of relative chemicals including chemical strucutre, composition and molecular weight.

Chemical	Structure	Composition	Molecular Weight
Unknown substance	Unknown	$C_{12}H_{17}N_3O_5$	283g/mol
5-azacytidine		$C_8H_{12}N_4O_5$	244g/mol
Guanosine		$C_{10}H_{13}N_5O_5$	283g/mol
Adenosine		$C_{10}H_{13}N_5O_4$	267g/mol
Inosine		$C_{10}H_{12}N_4O_5$	268g/mol
Ribavirin		$C_8H_{12}N_4O_5$	244g/mol

4.1 Nucleosides and their analogues affect HPV16 late gene expression

4.1.1 Adenosine and guanosine induce HPV16 late gene expression in a dose- and time-dependent manner

In order to test if guanosine was the unknown compound in the batch 1 "5-azayitidine", we tested guanosine and other nucleosides in our in-house reporter cell line C33A2 at different concentrations. Our data suggested that adenosine, cytidine, uracil and 6-thioguannosine had no effect on HPV16 late gene expression. Even though guanosine shares the same molecular weight with the unknown substance, we did not observe same effect with guanosine in inducing the HPV16 late gene expression. Guanosine could only induce HPV16 late gene expression up to 2-fold, whereas the unknown substance could induce HPV16 late gene expression up to 23-fold. But to our surprise, adenosine could activate HPV16 late gene expression up to 9-fold in a concentration dependent manner, but only in the presence of guanosine **(Figure 4.2)**. This effect was observed as early as 6 hours post-treatment and was only observed when guanosine was combined with adenosine and not detectable when guanosine was combined with uracil or cytidine.

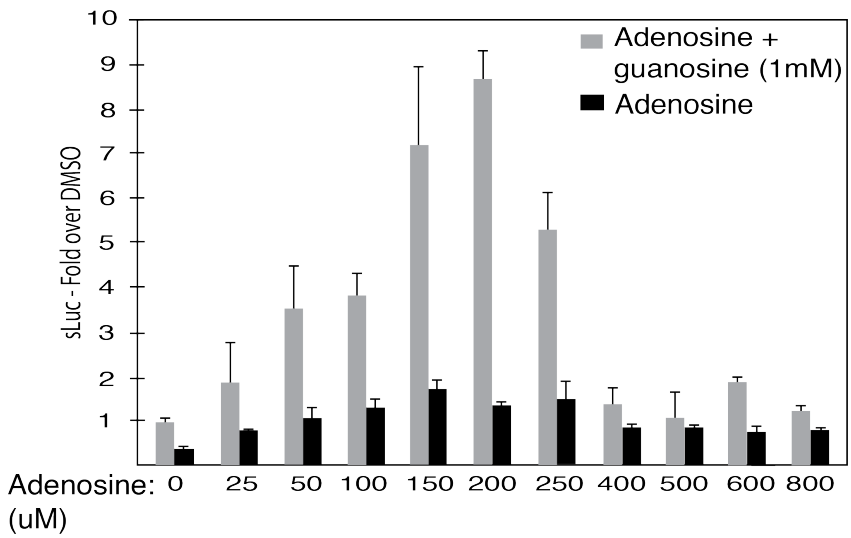


Figure 4.2 Secreted luciferase enzyme activity (sLuc) in the cell culture medium of reporter cell line C33A2 treated with adenosine alone or adenosine with 1mM guanosine for 20 hours.
The results showed that adenosine could induce HPV16 late gene expression in a dose-dependent manner (here measured as sLuc in the reporter cell line C33A2), but only in the presence of guanosine. These results showed that guanosine could stimulate adenosine to induce HPV16 late gene expression, demonstrating that mixtures of nucleosides could have effects on biological processes that are different from the effects of single, individual nucleoside.

The RT-PCR results indicated that guanosine and adenosine (G+A) mainly induced the HPV16 late L2 mRNAs up to 15-20-fold as well as the early mRNAs E2 and E4. This increase in mRNA occurred as early as 3 hours after addition of guanosine and adenosine, then the mRNA levelled off over time and the effect was almost entirely gone 14 to 24 hours after addition of guanosine and adenosine. Due to the fact that adenosine and guanosine induced primarily HPV16 late L2 mRNA, the results indicated that the L2 mRNA induction was the result of inhibition of HPV16 early polyadenylation site and caused read-through into the late polyadenylation site, thereby generating HPV16 the L2 mRNA.

Adenosine can function as a signaling molecule through the adenosine receptors presented on the cell surface **(Figure 1.10)**. We tested both antagonists and agonists for each adenosine receptor to investigate if receptors are involved in the induction of the HPV16 late gene expression. Our results showed that antagonists and agonists did not alter the adenosine and guanosine effect, suggesting that adenosine receptors did not play a major role in the induction of HPV16 late gene expression by guanosine and adenosine. However, the adenosine and guanosine effect was compromised when the nucleoside transporter ENT1 was knocked down or inhibited, suggesting that the adenosine and/or adenosine and guanosine together are transported into the cell through nucleoside transporter ENT1 to induce HPV16 late gene expression. Adenosine is metabolized by adenosine deaminase (ADA) and adenosine kinase (ADK) to AMP inosine and AMP respectively once adenosine enter the cells. Adenosine can also be metabolized to S-adenosyl homocysteine (SAH) through the reversible conversion by SAH hydrolase. In order to investigate which route was responsible for the guanosine and adenosine effect, we tested different inhibitors targeting ADA, ADK and SAH hydrolase respectively. The results showed that the induction of HPV16 late gene expression was significantly reduced when either ADA or ADK was inhibited. Interestingly, inhibition of SAH hydrolase itself increased the HPV16 late gene expression. Taken together, our results suggested that adenosine or adenosine and guanosine were transported into cells through the nucleoside transporter ENT1 followed by deamination by ADA and phosphorylation by ADK as the first steps towards activation of HPV16 late gene expression. In order to elucidate how guanosine and adenosine induced HPV16 late gene expression, we monitored levels of different RNA binding proteins that could potentially affect HPV16 gene expression. Western blot results indicated that levels of hnRNP G changed in the presence of guanosine and adenosine, but we also observed that the localization of the cellular HuR protein changed at 3h after the addition of guanosine and adenosine. The CLIP results showed an increase in association of the HuR protein with HPV16 mRNAs in an adenosine-dependent manner. It was reasonable to speculate that HuR inhibited HPV16 early polyadenylation signal and in turn induced HPV16 late L2 mRNA production.

4.1.2 Nucleoside analogues induce HPV16 late gene expression in a dose-dependent manner

As mentioned before, ribavirin has been recently gained attention by its anti-cancer effect in addition to the antiviral activity. A previous study showed that ribavirin can inhibit the activity of the oncogene elF4E to prevent cancer progression. The m6A is also an important RNA modification that plays a part in tumor progression. Cordycepin is also interesting due to its multiple anti-cancer effects and by its ability to induce apoptosis. Besides, ribavirin, m6A and cordycepin are also involved in the purinergic pathway as purine analogues which has been shown to be important for cancer treatment [283]. Therefore, we investigated if ribavirin, N6-methyladenosine (m6A) and cordycepin could affect HPV16 late gene expression. The in-house reporter cell line C33A2 cells were treated with different molecules either alone or with guanosine. The sLuc results revealed that ribavirin could induce HPV16 late gene expression to 6-fold in the presence of guanosine but did not show any effect alone by itself. While the m6A with guanosine induced HPV16 late gene expression as high as 22-fold and had less cytotoxicity to the cells compared to guanosine and adenosine at low concentrations. RNA analysis on cells treated with m6A with guanosine showed that m6A and guanosine primarily induced HPV16 late L2 mRNA as well as the early E4 mRNA, suggesting that m6A and guanosine together caused the inhibition of the early poly(A) signal and resulted in production of L2 mRNA. In addition, m6A and guanosine might increase the global transcription of viral genome to produce more E4 mRNA, but further analysis is needed to reveal the exact mechanism.

Unlike ribavirin and m6A, cordycepin alone could induce HPV16 late gene expression up to 4-fold in a dose-dependent manner **(Figure 4.3)**. RT-PCR results showed that cordycepin mainly induced HPV16 late L2 mRNAs with no measurable effect on L1 or L1i mRNAs. 3'RACE results suggested that cordycepin inhibited the HPV16 early polyadenylation site and caused readthrough into the late region of the HPV16 genome to produce HPV16 late L2 mRNA. As adenosine analogue, cordycepin lacks a 3'-OH group and can act as a mRNA chain terminator. In order to investigate if cordycepin induced HPV16 late gene expression by terminating mRNA elongation, we tested two other chain terminators: 3'ATP and 8-aminoadenosine. Although all of them could terminate the RNA elongation, but only cordycepin could induce HPV16 late gene expression. In an attempt to investigate how cordycepin induce HPV16 late gene expression, we monitored if different protein levels changed after addition of cordycepin. Western blots indicated that there was an increase in the protein levels of polyadenylation factor NUDT21, also known as CPSF5. Oligonucleotide-mediated pull-down using oligos that covered the early polyadenylation site demonstrated that

NUDT21 binds to sequences both upstream and downstream of the HPV16 early polyadenylation site.

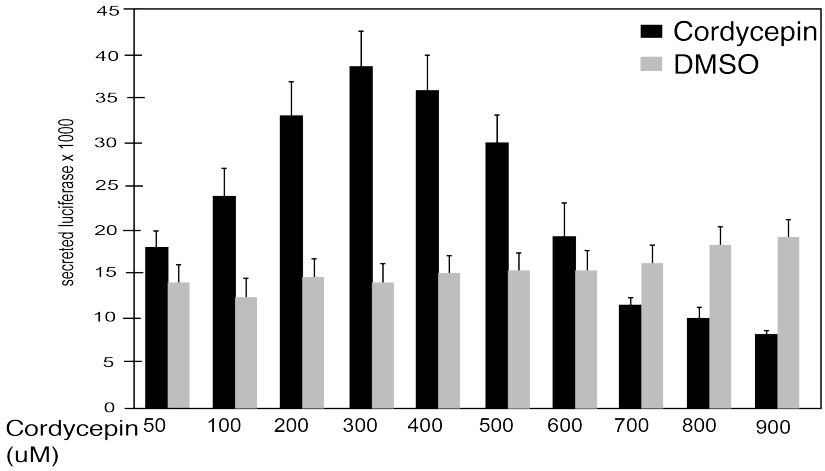


Figure 4.3 Secreted luciferase enzyme activity (sLuc) in the cell culture medium of reporter cell line C33A2 treated with cordycepin with different concentrations for 20 hours, DMSO as negative control.
The results showed that cordycpein could induce HPV16 late gene expression in a dose-dependent manner.

Taken all results together, we concluded that cordycepin induced HPV16 late gene expression by increasing the expression levels of the polyadenylation factor NUDT21. We speculated that the increased NUDT21 protein levels resulted in increased binding to pAE and inhibited pAE activity, in turn re-directing polyadenylation to the late polyadenylation signal, thereby producing late HPV16 L2 mRNA. Overexpression of NUDT21 with HPV16 sub-genomic plasmids pBelMsLuc and pBelMsLuc caused induction of HPV16 late gene in both plasmids by 7- and 5-fold respectively. RT-PCR results revealed that overexpression of NUDT21 mainly induced L2 mRNA. These data further supported the idea that cordycepin induced HPV16 late gene expression by increasing NUDT21 levels and inhibiting the early polyadenylation site, in turn resulting in HPV16 L2 mRNA production.

4.2 Splicing factor hnRNP G affects HPV16 early and late gene expression

4.2.1 hnRNP G inhibits small exon inclusion on the HPV16 L1 mRNA

In the experiments described above, in which we demonstrated that nucleosides could induce HPV16 late gene expression, we observed that treatment of C33A2 cells with nucleosides altered the levels of the RNA binding protein hnRNP G. Furthermore, we have shown that activation of the DNA damage response (DDR) can activate HPV16 late gene expression [284], and hnRNP G is known to play an active role in the DDR. We therefore wished to investigate the role of hnRNP G in the control of HPV16 gene expression. The HPV16 splice site SA3358 is the most commonly used splice site on HPV16 genome and is used to generate HPV16 E6, E7, L2 and L1 mRNA. There are two isoforms of the L1 mRNA: L1 and L1i. L1i differs from L1 by exclusion of the small exon from SA3358 to SD3632. Previously we had identified an 8-nucleotide RNA element (ACCGAAGAA) located downstream of SA3358. In order to further investigate the function of the 8-nucleotide element, we constructed mutated plasmid p3*sLuc with mutations in this 8-nucleotide sequence from ACCGAAGAA to GUGUGUGCA. The wild type plasmid pBelsLuc and the mutant plasmid p3*sLuc were transfected into HeLa cells separately and the total RNA was extracted and analyzed. The RT-PCR results indicated that the mutated plasmid p3*sLuc produced more L1i mRNA compared to wild type plasmid, suggesting that this 8-nucleotide element controlled HPV16 splice site SA3358 activity and caused inclusion of the HPV16 small exon between SA3358 to SD3632. We further constructed two more plasmids that contain either 7 copies of wild type 8-nucleotdie element (p7xWTsLuc) or 7 copies of mutated element (p7xMUTsLuc). Transfection of these two plasmids into HeLa cells showed a more than 20 times higher production of L1i mRNA from p7xMUTsLuc compared to p7xWTsLuc. This could be explained by low efficiency splicing to SA3358 in p7xMUTsLuc and therefore splicing was redirected to SA5639, generating high levels of L1i mRNA. We concluded that this 8-nucleotide element acted as a splicing enhancer to promote splicing to SA3358 for inclusion of the exon between SA3358 to SD3632 on the L1 mRNAs.

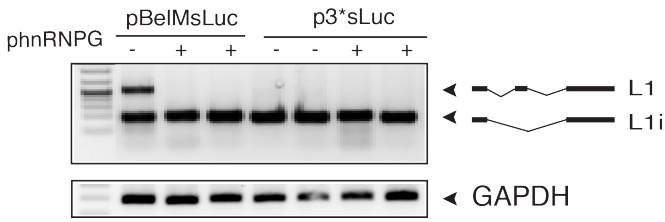


Figure 4.4 RT-PCR results of L1 and L1i mRNAs on RNA extracted from HeLa cells transfected with pBelMsLuc and P3*sLuc in the absence or presence of hnRNP G plasmid.
These results showed that overesrepssion of hnRNP G could inhibit the inclusion of the small exon between SA3358 and SD3632 on the L1 mRNA and result in increased expression leves of L1i mRNAs.

RNA-mediated pull-down with the wild type and mutant 8-nucleotide RNA oligos 3*wt and 3*mut suggested that hnRNP G and Tra2ß interacted specifically with the 8-nulceotide sequence. Overexpression of hnRNP G with pBelMsLuc plasmid induced HPV16 late L1i mRNA but not with p3*sLuc plasmid. RNA analysis further demonstrated that hnRNP G inhibited the inclusion of the small exon between SA3358 and SD3632 **(Figure 4.4)**. Taking these results together, we concluded that hnRNP G could induce HPV16 late gene expression and strongly inhibited inclusion of the exon between SA3358 and SD3632 on the L1 mRNA through the 8-nucleotide element. To our surprise, we monitored an inhibition of E4 mRNA levels in the mutated 8-nucleotide plasmid P3*sLuc plasmid compare to the wild type 8-nucleotide plasmid pBelMsLuc, suggesting that hnRNP G could inhibit splicing to SA3358 even in the absence of the 8-nucleotide element. It is reasonable to speculate that hnRNP G is competing with other positive splicing factors to bind to the 8-nucleotide element, but when the 8-nucleotide element was mutated or deleted, other hnRNP G binding site increased its significance and hnRNP G could bind and inhibit the splicing to SA3358. We concluded that hnRNP G could inhibit the inclusion of the exon between SA3358 and SD3632, primarily through interfering with the 8-nucleotide element and with other splicing factors.

4.2.2 hnRNP G and its mutant inhibit HPV16 oncogene E6/E7 splicing

Previously we showed that hnRNP G could affect the inclusion of the small exon in HPV16 late L1 mRNA splicing. We next investigated if hnRNP G could affect HPV16 oncogene E6/E7 mRNA splicing. For these experiments we used reporter plasmid pC97ELsLuc rather than pBelsLuc since pC97ELsLuc encodes HPV16 oncogene E6 and E7 mRNA. By co-transfecting hnRNP G expression plasmid together with the HPV16 sub-genomic plasmid pC97ELsLuc into HeLa cells, there was a strong inhibition of E6/E7 mRNA splicing **(Figure 4.5)**. Overexpression of

hnRNP G resulted in increased levels of unspliced E6 mRNAs at the expense of E6*I/E7 mRNAs.

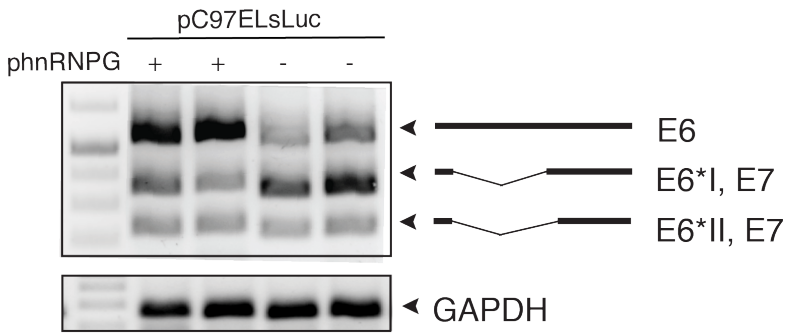


Figure 4.5 RT-PCR results of E6 and E7 mRNAs on RNA extracted from HeLa cells transfected with pC97ELsLuc in the absence or presence of hnRNP G plasmid.
These results showed that overexpression of hnRNP G could inhibt the splicing to E7 mRNA and favore the production of full-length E6 mRNA, suggested that hnRNP G had inhibitory effect on HPV16 oncogene E6/E7 mRNA splicing.

Further analysis with various sub-genomic HPV16 plasmids spanning the HPV16 E6/E7 region and extending into E1 region revealed that hnRNP G interacted with HPV16 sequences between nucleotide positions 960 to 1060. However, the exact sequence has not yet been identified. We further investigated which domain of hnRNP G that was responsible for the inhibitory effect on HPV16 E6/E7 splicing. To this end, we constructed various hnRNP G-deletion plasmids and co-transfected them together with plasmid pC97ELsLuc. The RT-PCR results indicated that deletion of the RNA-binding domains C-RBD or N-RRM one at a time did not abolish the hnRNP G inhibitory effect, simultaneous deletion of both domains did. Further extended deletion in C-terminal of hnRNP G resulted in alleviated effect of E6/E7 mRNA splicing inhibition, suggesting that the C-RBD was more necessary for the inhibitory effect of the E6/E7 mRNA splicing. On the other hand, co-transfection of N-RRM deleted plasmid with pBelsLuc plasmid showed increased inclusion of the small exon between SA3358 to SD3632 in L1 mRNA production compare to C-RBD deletion plasmid. It is reasonable to speculate that hnRNP G might inhibit E6/E7 splicing and exon inclusion in L1 mRNA through different domains.

5 Discussion and Future Perspectives

The HPV16 genome is highly compact and almost all transcripts from papillomavirus are polycistronic. These polycistronic mRNAs undergo extensive alternative splicing to generate mature mRNAs. HPV16 gene regulation is therefore very important in controlling HPV16 gene expression for virus life cycle. HPV16 gene expression is tightly regulated by cell differentiation. The highly immunogenic structural proteins L1 and L2 are only expressed when cells undergo final differentiation at the superficial layer, this delayed expression of immunogenic late proteins allows the virus to escape from host immune system [44]. Therefore, inducing HPV16 late gene expression prior to cell differentiation might be a valuable mechanism for developing HPV antiviral treatment.

In this study, we screened nucleosides/nucleosides analogues and splicing factors for their effect on HPV16 early and late gene expression. The purpose was to identify small molecules that could potentially induce HPV16 late gene expression and to enhance our understanding of HPV16 gene regulation. Although we identified a substance with a remarkable ability to induce HPV16 late gene expression in our C33A2 reporter cell line, the identity of the substance remains largely unknown, despite mass spectrometry analysis, which revealed that the substance is a nucleoside or nucleoside analogue. We are unfortunately still unable to identify the unknown substance. One speculation is that the unknown substance could be an intermediate substance synthesized during production, and that might be the reason for the absence of a similar product on the market. Alternatively, one might speculate that the unknown substance was a “mixture” of several substances, most likely nucleosides or nucleoside analogues. This led us to investigate the effects of nucleosides, nucleoside analogues or mixtures on HPV16 late gene expression.

Adenosine contributes to various medical conditions including cancer and it is believed that levels of adenosine are high in the hypoxic microenvironment of tumors [239, 240]. Adenosine is known to have direct effects on cancer proliferation and apoptosis signaling through one or more of the four different receptors: AR1, AR2A, AR2B and AR3. The exact effect of adenosine depends on which receptors that are presented on the cancer cell [242, 245]. However, inhibition of adenosine receptors did not compromise the guanosine and adenosine mediated HPV16 late gene expression in our case. Interestingly, activation of A2A and A2B receptors

increased guanosine and adenosine effect around 2-fold, while activation of A1 and A3 receptors inhibited the guanosine and adenosine effects, suggesting that adenosine receptors signaling is involved in the adenosine and guanosine effect but only to a minor extent. Our data did not support the idea of adenosine receptors played a major role in this adenosine and guanosine effect. Thus, we concluded that adenosine did not mainly function as a signaling molecule for the induction of HPV16 gene expression.

On the contrary, inhibition of nucleoside transporters inhibited the effect of guanosine and adenosine, indicating that adenosine or adenosine together with guanosine was transported into cells through the nucleoside transporters. It is well known that after transportation into the cell, most adenosine molecules are either phosphorylated by adenosine kinase (ADK) to AMP [247, 285], or deaminated by adenosine deaminase (ADA) to inosine [286]. Since AMP can be further converted to inosine, and inosine from both pathways can be further processed to hypoxanthine, xanthine and uric acid. One may speculate that inosine or any of the downstream metabolites could potentially be the inducer for HPV16 late gene expression. However, inosine alone could only trigger the HPV16 late gene expression to approximately 2-fold. Since xanthine oxidase catalyzes the conversion of hypoxanthine to xanthine, and xanthine to uric acid, inhibition of xanthine oxidase will lead to accumulation of hypoxanthine and/or xanthine. However, inhibition of xanthine oxidase did not result in any inhibition on HPV16 late gene expression. On the contrary, the effect was increased to 16-fold. Therefore, hypoxanthine or xanthine could potentially be the active substance that responsible for HPV16 late gene expression. Due to the fact that both hypoxanthine and xanthine are not soluble in water or other organic solvent, we could not test these two compounds directly.

Alternatively, Western blot results showed that adenosine and guanosine affected several RNA binding protein levels including HuR and hnRNP G. Adenosine and guanosine could also affect HPV16 late gene expression by controlling those proteins. HuR protein has previously been shown to have effect on HPV16 late mRNA levels in HPV16 infected cells by shuttling HuR between nucleus and cytoplasm [219]. In addition, previous studies showed that HuR can affect alternative polyadenylation. Overexpression of HuR in HeLa cells selectively block polyadenylation at poly(A) site [287]. In this study we have shown that HuR bound to the HPV16 early UTR and in turn inhibited the HPV16 early polyadenylation site to produce L2 mRNA. Re-localization of HuR from nucleus to cytoplasm was detected at 3 hours after addition of adenosine and guanosine, which correlates to the rapid production of L2 mRNAs. There were also increased expression levels of protein HuR in the cytoplasm in response to adenosine and guanosine, suggesting that guanosine and adenosine could potentially increase the HuR-mediated HPV16 mRNA exportation from the nucleus to cytoplasm, in turn the high level HuR protein inhibited the HPV16 early polyadenylation site and generated L2 mRNAs.

In contrast to adenosine, the receptors and metabolism mechanism of guanosine have not been well studied. However, it is known that guanosine can act as an extracellular signal molecule and interact with adenosine receptors mainly through A1R and A2AR [288]. Inhibition of ENT transporters did not inhibit the effect of guanosine alone on HPV16 late gene expression but did inhibit the effect of guanosine and adenosine together, suggesting that guanosine was not transported inside of the cells and might function as signaling molecule, but further investigation is needed.

Despite the incomplete knowledge of guanosine, a "adenosine-guanosine mechanism" has been reported previously [289]. These investigators showed that guanosine and adenosine inhibited cell proliferation and extra cellular guanosine interfered with the disposition of extracellular adenosine [289]. However, in their study, they ruled out the involvements of adenosine deaminase, adenosine kinase and ENT transportation. They proposed that extracellular guanosine regulated the extracellular adenosine levels both *in vitro* and *in vivo* [289, 290]. Our results indicated that nucleoside transporters are absolutely necessary for the induction of HPV16 late gene expression. One can speculate that the exact mechanism for guanosine and adenosine might be cell type specific. Even though the exact mechanism for the "adenosine and guanosine effect" still remains unclear, my study provided insights the role of guanosine in adenosine-mediated pathways as well as our understanding of the uptake and function of various nucleoside-based drugs.

Guanosine exerts a neuroprotective effect against glutamate toxicity and oxidative stress and prevents neuroinflammation [291]. Besides, guanosine has previously been suggested to have antitumor effect in Ehrlich carcinoma, guanosine treatment caused a 30% reduction of the tumor weight [292, 293]. A previous study also showed that combination of guanosine and chemotherapeutic agent 5'-doxy-5-flurouridine had an increased antitumor effect in leukemia models [294]. Besides, co-treatment with guanosine and acriflavine *in vivo* enhanced the antitumor effect by decreasing 96% of the total tumor weight [293]. My study showed that guanosine could increase the effect of adenosine on induction of HPV16 late gene expression. Taking these results together suggested that guanosine could potentially increase the effect of other nucleosides or nucleoside analogues, especially those are used as antiviral or anticancer drugs.

We next investigated if guanosine could increase the effect of other nucleoside analogues such as ribavirin and m6A on HPV16 late gene expression. Nucleoside analogues are antimetabolites that interfere with the nucleic acid synthesizes and have been widely used for anti-cancer and anti-viral treatments. Ribavirin is a guanosine analogue that was first discovered as antiviral treatment for RNA viruses including RSV and HCV infections [268, 270]. Ribavirin can be incorporated into RNA to pair with either uracil or cytosine and cause mutagenesis during viral

replication [273]. Ribavirin is also used to against DNA viruses, but the exact mechanism still remains a mystery [295]. In recent years ribavirin has also been suggested as anticancer treatment for AML patients [276]. The m6A is an adenosine analogue and m6A modification has drawn increasing attention to its involvement in cancer carcinogenesis [267]. A recent study also showed that methylation of RNA rather than DNA could potentially play a role in regulating HPV16 late gene expression.

The sLuc results from our sLuc-based reporter cell line for HPV16 late gene expression showed that both ribavirin and m6A can induce HPV16 late gene expression in the presence of guanosine. In particular, m6A with guanosine induced HPV16 late gene expression around 20-fold compared to adenosine and guanosine. Cytotoxicity analysis showed that m6A with guanosine showed less cell cytotoxicity compared to adenosine and guanosine, suggesting that m6A with guanosine might be the better inducer for HPV16 late gene expression. RNA analysis on m6A treated C33A2 cells indicated that m6A and guanosine induced HPV16 late gene expression through late L2 mRNA production, suggesting that m6A inhibited HPV16 the early polyadenylation signal, thereby causing readthrough to the late polyadenylation site. Further investigation is needed to determine how m6A and guanosine induced HPV16 late gene expression.

Even though the exact mechanism for the "guanosine and nucleoside/nucleoside analogue effect" still remains unclear, my study provided insights of the role for guanosine in nucleoside-mediated pathways as well as our understanding of the uptake and function of various nucleoside-based drugs. My study showed that guanosine could increase the effect of adenosine, m6A and ribavirin on induction of HPV16 late gene expression.

In addition to m6A and ribavirin, we also identified an adenosine analogue named cordycepin that has effects on HPV16 late gene expression alone. Cordycepin has been used as traditional Chinese medicine for centuries. It is well known for its multiple effects in inflammation, apoptosis and tumor formation [256, 296, 297]. In this study, we showed that cordycepin could induce HPV16 late gene expression in a dose-dependent manner. The RT-PCR results showed that cordycepin mainly increased HPV16 L2 mRNA production, suggesting that cordycepin mainly inhibited the HPV16 early polyadenylation. Further analysis with western blot revealed an increase of cellular protein NUDT21 after cordycepin treatment. NUDT21, also known as Cleavage and Polyadenylation Specificity Factor subunit 5 (CPSF5), it is a polyadenylation regulator and contributes the recruitment of different proteins to the cleavage and polyadenylation site. Overexpression of NUDT21 induced expression of HPV16 late L2 mRNA, ssDNA mediated pull-down showed that NUDT21 interacted with sequences at the HPV16 early polyadenylation site. Taking these results together, cordycepin induced HPV16 late

gene expression by increasing the expression levels of NUDT21 in the cell. In turn, NUDT21 inhibited the HPV16 early polyadenylation site and caused readthrough into the late region which resulted in L2 mRNA production. Cordycepin has been reported to have anti-tumor effects that are mediated by inducing apoptosis. However, we did not observe any signs of apoptosis by cordycepin in the C33A2 cells used here. It is worth to mention that C33A2 cells are p53-negative, which means that apoptosis is unlikely to happen in C33A2 cells. Cells appeared unaffected by the addition of cordycepin, at least for the 20 hours incubation period used here. Cordycepin is also well-known as an RNA chain terminator. However, our *in vitro* polyadenylation experiments indicated that even if cordycepin as well as two other chain terminators could inhibit RNA elongation *in vitro*, only cordycepin was capable of inducing the HPV16 late gene expression in the C33A2 cells. In summary, our results supported the conclusion that cordycepin acts through a mechanism that is distinct from termination of RNA elongation.

To summarize, my study suggested that nucleosides and mixture of nucleosides/nucleoside analogues could induce HPV16 late gene expression in our reporter cell line C33A2. These results suggested novel therapies for HPV16 infections in which one could possibly induce HPV16 late L1 and L2 gene expression at an early stage of infection to evoke host immune responses to HPV16 followed by clearance of the infection. However, there are several challenges still remain. First, the exact role of guanosine in this "guanosine and adenosine/ nucleoside analogue effect" still largely unknown. Experiments are needed to investigate if guanosine is involved in the purinergic signaling pathway and if guanosine acts through adenosine receptors. Second, the effect of guanosine and adenosine/nucleoside analogues are only observed in our cell culture studies. It is still left unanswered whether the effect of guanosine and adenosine/nucleoside analogues can occur in other cancer cells or even *in vivo*. Future studies will focus on these aspects and will determine if this guanosine and adenosine/nucleoside analogues effect could be therapeutically significant. Third, we showed that the adenosine signaling pathway stimulated by activation of the adenosine receptors did not play a major role in our case. Instead we showed that adenosine is transported into cells to induce HPV16 late gene expression. It is possible that guanosine and adenosine act through a different mechanism than other nucleoside analogues that have been used as antiviral or anticancer drugs, which may be of interest for development of novel antiviral/anticancer agents. There is a need to develop new agents that have mechanism of action that differ from the existing agents and provide improved bioavailability to overcome the resistance for existing agents.

In our previous study with guanosine and adenosine, we observed that the expression levels of hnRNP G changed after guanosine and adenosine treatment. hnRNP G is an RNA-binding protein that is ubiquitously expressed in different tissues. Some studies suggested that hnRNP G might play a role in cancer

development. In endometrial cancer, high expression levels of hnRNP G predicted to have a better clinical outcome [298]. Another study showed a correlation between high expression levels of hnRNP G in cervical cancer and a higher 5-year survival rate [299]. It is therefore of interest to investigate if hnRNP G could affect HPV16 gene expression. We had previously identified an 8-nucleotide RNA element that acts as a splicing enhancer and promotes inclusion of the exon between SA3358 and SD3632 in HPV16 late L1 mRNA. RNA-mediated oligo pull-down with the 8-nucleotide oligo followed by silver staining showed that several proteins including cellular protein hnRNP G are binding to this enhancer either to enhance or to inhibit its function. However, it remains unclear if hnRNP G bound to this enhancer directly or through protein-protein interactions. Overexpression of hnRNP G resulted in exon skipping and produced more L1i mRNAs at the expense of L1 mRNA, indicating that hnRNP G could function as an inhibitor for exon inclusion. Since the 8-nucleotide element promotes exon inclusion and hnRNP G inhibits the exon inclusion, the final effect is determined by the relative abundance of different proteins. It is reasonable to speculate that hnRNP G might compete with other enhancer-binding proteins that interact with the 8-nucleotide element. The balance between the various proteins was interrupted when hnRNP G was overexpressed and resulted in a predominant inhibitory effect of exon inclusion.

Both L1 and L1i mRNAs are capable of producing functional L1 protein, and both mRNAs have been detected in HPV16 positive cells [103], but respective translation efficiency is unknown. The significance of the regulation of the small exon between SA3358 and SD3632 lies in its role in splicing site activation for late mRNAs production. The splice site SA3358 is most commonly used throughout the HPV16 life cycle to produce both early and late mRNAs. One can speculate that the cellular proteins such as hnRNP G contributed in regulation of mRNA splicing and acted as a switch for HPV16 from the early to late gene expression by inducing HPV16 late mRNA splicing.

Expression of both HPV16 E6 and E7 is required to induce and maintain cancers caused by HPV16. The balance between E6 and E7 oncogene expression is important for cancer development. Since too little E6 will lead to cell apoptosis, while too little E7 will halt cancer cell proliferation. Both E6 and E7 mRNAs are produced from the same HPV16 pre-mRNA but undergo different alternative splicing. Controlling the splicing between E6 and E7 mRNA therefore is really important in order to achieve a perfect balance in production of the E6 and E7 proteins. Previously we showed that hnRNP G affects the splicing of late L1 mRNA by inhibiting splicing to splice site SA3358, it would be of interest to investigate if hnRNP G could affect splicing of E6 and E7 mRNA.

Overexpression hnRNP G together with sub-genomic HPV16 plasmid pBelsLuc in HeLa cells resulted in inhibition of E6/E7 pre-mRNA splicing and produced more

unspliced E6 mRNAs at the expense of the E7 mRNAs. Further analysis showed that hnRNP G targets sequences between nucleotide position 990 to 1060. There are two RNA binding domains reported in hnRNP G: the classic RNA recognition motif (RRM) and RNA binding domain (RBD). These two domains appear to have different sequence specificity for RNA binding. The RRM seems to prefer CCC and CCA [300], AAGU[198] or AAN, while the RBD seems to prefer a long and purine rich sequence GGAAA [301]. We showed that the RRM domain in hnRNP G is not absolutely necessary for E6/E7 mRNA splicing. Overexpression of mutant hnRNP G with RRM deletion could still inhibit splicing to E6/E7 mRNAs but with lower efficiency. On the contrary, overexpression of mutant hnRNP G with RRM deletion resulted in increased exon inclusion on the L1 mRNAs and produced more L1 mRNA. These results suggested that the RRM domain is needed for exon inclusion on the L1 mRNAs, but not essential for E6/E7 splicing inhibition. However, further study is needed to confirm whether hnRNP G regulated HPV16 E6/E7 mRNA splicing by direct binding or by indirect protein-protein interactions.

hnRNP G is widely expressed in different tissues including the HPV targeting tissues cervical and tonsillar epithelium. Studies showed that hnRNP G is involved in DNA damage response (DDR) for both homologous recombination (HR) and end joining (EJ) pathways. The hnRNP G in both pathways accumulates at DNA breakage site to protect the DNA from exonuclease activity [200, 201]. It is worth to mention that HPV infections activates DDR and HPV hijacks the DNA repair machinery for viral replication [302]. It is reasonable to speculate that hnRNP G might be recruited by DDR to the HPV16 genome and interfered with HPV16 life cycle by regulating the mRNA alternative splicing including oncogene E6/E7 and late L1 mRNAs.

In addition, high expression levels of hnRNP G in cervical cancer resulted in a better five-year survival rate [298]. There is an absence or large decrease in hnRNP G protein levels in malignant oral lesions and head and neck SCC cancer cells compares to normal oral cells [199]. These data suggest that hnRNP G might play an important role as tumor suppressor in carcinogenesis. In my study, overexpression of hnRNP G strongly inhibited the E6 splicing to E7 mRNA, which result in lower expression level of E7 mRNA. Since the balance between E6 and E7 is essential for cancer progression, decreased E7 expression levels will liberate the pRb protein and halt cancer cell proliferations. This finding suggests that hnRNP G is a key regulator of HPV16 oncogene E6 and E7 expression and it is in agreement with previous study that hnRNP G might have a role in carcinogenesis [199]. E6/E7 splicing mechanism is a possible target for treatment of HPV16 infections and HPV16 caused high grade lesions or cancers. Targeting hnRNP G could be one potential way in order to achieve the goal of affecting HPV16 oncogene expression.

In conclusion, initial experiments performed by me during my doctoral studies identified a small molecule that strongly induced HPV16 late gene expression. Although this substance was not the azacytidine substance that we had ordered from the vendor, and the identity of this substance is still unknown, these initial experiments resulted in the unexpected finding that combinations of nucleosides (primarily adenosine and guanosine) significantly affect HPV16 gene expression. Furthermore, our results showed that guanosine could enhance the effect of nucleoside analogues on HPV16 gene expression, which may have ramifications for treatments of virus infections with nucleoside analogues. This project also resulted in the identification of hnRNP G as a novel regulator of HPV16 E6/E7 oncogene- and L1 late-gene expression.

6 Acknowledgement

I would like to take this opportunity to give my sincerest gratitude to all the people who helped and supported me during my PhD study. This work cannot be accomplished without their help.

To my supervisor professor **Stefan Schwartz**, thank you for giving me this opportunity and guiding me into the world of molecular virology. Your vision, motivation and creativity has always been so inspiring. I will always remember in every lab meeting, you would repeat thousands of times of "Good!", "Well done!", "Excellent!". You are such a positive and cheerful supervisor as one can imagine. To my co-supervisor professor **Patrik Medstrand**, thank you for your helpful discussion and being supportive for the whole time.

To former members in the group, **Anki Mossberg**, thank you so much for being there for me especially in the beginning of my PhD life. You are like a family to me, you are always so passionate, patient and whenever I have a question, you have an answer. I am also so inspired by your way of living. I wish I could be like you somedays. **Chengjun Wu**, can't thank you enough for helping me with my PhD study, I learned so many things from you. You were almost like "Wikipedia" to me, my PhD study would not be this smooth without your continuous help. Wish you the best best luck in Dalian. **Naoko Kajitani**, you were the role model to me. I really admire your passion to science. I wish I could have half of your strength and passion. Best wishes for your future career. **Xiaoze Li-Wang**, it was such a lucky thing to meet you, I couldn't be here doing my PhD without your help in the beginning. Best wishes for you and your little daughter. **Cecilia Johansson**, thank you for your expertise and skillful knowledge. **Efthymios Tsimtsirakis**, you are always so kind and helpful, I wish you best luck for your PhD.

To current members, I am very much grateful for your support and all the great times we spent together. **Kersti Nilsson**, thank you for your encouragement whenever I feel low, you always can cheer me up. I hope you get well soon and get championship together with your sister in CrossFit one day. Dream big! **Xiaoxu Cui**, thank you for all the laughter we shared, you are really a good listener. I loved our weekly gossip dinner so much and we should definitely keep the tradition. I wish you best for you PhD study, you are going to rock. **Johanna Jönsson**, thank you for your help with my Swedish abstract. Your smile always so bright and

cheerful. Remember I am just one e-mail away if you need any help. **Yunji Zheng**, thank you for all the good times and dinners, you really have a good sense of humor.

To **Marianne Jansson**, **Joakim Esbjörnsson**, thank you so much for helping me with my application, and all the interesting topics we discussed during the joint meeting. To **Ram Ajore**, **Ellinor Johnsson**, thank you so much for your constant help, whenever I need help your answer would always be "Yes!". To **all the former and current members in B13**, thank you for the great environment you have created, thank you for all the great time we shared and all the help you provided.

To my friends in Sweden, Xu Yan, Siyuan Hu, Kun Xue, Wen Li, Rachel Cheong, Yuan Ouyang, Qimin Zhou, Dingyuan Hu, Di Liu, Qi Shi, Xin Liu, Haiyue Gong, Juanzi Shi, Rui Liu, Yiheng Tong, Wubin Weng, Mengshu Hao, Xiaowen Chen, Yihong Chen and many others. You made my life in Lund so colorful and unforgettable. I couldn't be such a happy and positive person without constant encouragements from you guys.

To my dear boyfriend **Hampus Åström**, thank you for being such a considerate and supportive boyfriend, you make my life so wonderful. To my family in Sweden, **Kerstin, Mats, Jan, Eleonore** and **Erik**, thank you so much for your caring and kindness, you make me feel home again even when I am far away from it.

感谢我的大学闺蜜**王沫霖**和**张若琪**，在马来西亚和你们的日子是我这辈子最难忘的，最喜欢每次回国和你们一起聊天到深夜，我何其有幸有你们这样的闺蜜，永远爱你们。感谢我的初中闺蜜**董媛媛**，这么多年没有见面但是你总是想着我，希望你和宝宝都健健康康平平安安。

最后我要感谢我的父母，谢谢他们这么多年以来对我的支持和理解，尽管他们心中万千不舍，但总是选择让我勇敢去飞。感谢我的弟弟在我不在父母身边的时候替我照顾父母，姐姐很遗憾这么多年没能经常陪在你身边看着你长大，时光荏苒，希望未来的路我们可以一起相互扶持同手同脚的走下去。还要感谢我其他的家人，谢谢你们在我不在的时候帮我照顾父母，让我在异国他乡的十多年可以没有后顾之忧。

7 References

1. zur Hausen, H., *Papillomaviruses in the causation of human cancers - a brief historical account.* Virology, 2009. **384**(2): p. 260-5.
2. Zheng, Z.-M. and C.C. Baker, *Papillomavirus genome structure, expression, and post-transcriptional regulation.* Frontiers in bioscience : a journal and virtual library, 2006. **11**: p. 2286-2302.
3. zur Hausen, H., *Papillomaviruses and cancer: from basic studies to clinical application.* Nat Rev Cancer, 2002. **2**(5): p. 342-50.
4. Parkin, D.M. and F. Bray, *Chapter 2: The burden of HPV-related cancers.* Vaccine, 2006. **24 Suppl 3**: p. S3/11-25.
5. Richardson, H., et al., *The natural history of type-specific human papillomavirus infections in female university students.* Cancer Epidemiol Biomarkers Prev, 2003. **12**(6): p. 485-90.
6. Stanley, M., *Immunobiology of HPV and HPV vaccines.* Gynecol Oncol, 2008. **109**(2 Suppl): p. S15-21.
7. Arbyn, M., et al., *EUROGIN 2011 roadmap on prevention and treatment of HPV-related disease.* Int J Cancer, 2012. **131**(9): p. 1969-82.
8. Barr, E. and G. Tamms, *Quadrivalent human papillomavirus vaccine.* Clin Infect Dis, 2007. **45**(5): p. 609-7.
9. de Sanjose, S., et al., *Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study.* Lancet Oncol, 2010. **11**(11): p. 1048-56.
10. Li, N., et al., *Human papillomavirus type distribution in 30,848 invasive cervical cancers worldwide: Variation by geographical region, histological type and year of publication.* Int J Cancer, 2011. **128**(4): p. 927-35.
11. Gillison, M.L., et al., *Evidence for a causal association between human papillomavirus and a subset of head and neck cancers.* J Natl Cancer Inst, 2000. **92**(9): p. 709-20.
12. Begum, S., et al., *Tissue distribution of human papillomavirus 16 DNA integration in patients with tonsillar carcinoma.* Clin Cancer Res, 2005. **11**(16): p. 5694-9.
13. Knipe, D., et al., *Fields virology, Lippincott Williams & Wilkins.* Philadelphia, PA, 2007.
14. Joura, E.A., et al., *A 9-valent HPV vaccine against infection and intraepithelial neoplasia in women.* N Engl J Med, 2015. **372**(8): p. 711-23.

15. Van Doorslaer, K., et al., *The Papillomavirus Episteme: a major update to the papillomavirus sequence database.* Nucleic acids research, 2017. **45**(D1): p. D499-D506.
16. Antonsson, A., et al., *The ubiquity and impressive genomic diversity of human skin papillomaviruses suggest a commensalic nature of these viruses.* Journal of virology, 2000. **74**(24): p. 11636-11641.
17. Orth, G., *Epidermodysplasia verruciformis: a model for understanding the oncogenicity of human papillomaviruses.* Ciba Found Symp, 1986. **120**: p. 157-74.
18. Bodily, J. and L.A. Laimins, *Persistence of human papillomavirus infection: keys to malignant progression.* Trends in microbiology, 2011. **19**(1): p. 33-39.
19. de Villiers, E.M., et al., *Classification of papillomaviruses.* Virology, 2004. **324**(1): p. 17-27.
20. Bzhalava, D., et al., *A systematic review of the prevalence of mucosal and cutaneous human papillomavirus types.* Virology, 2013. **445**(1-2): p. 224-31.
21. de Villiers, E.M., *Cross-roads in the classification of papillomaviruses.* Virology, 2013. **445**(1-2): p. 2-10.
22. Lorincz, A.T., et al., *Human papillomavirus infection of the cervix: relative risk associations of 15 common anogenital types.* Obstet Gynecol, 1992. **79**(3): p. 328-37.
23. Howley, P.M. and H.J. Pfister, *Beta genus papillomaviruses and skin cancer.* Virology, 2015. **479-480**: p. 290-6.
24. Borgogna, C., et al., *Characterization of beta papillomavirus E4 expression in tumours from Epidermodysplasia Verruciformis patients and in experimental models.* Virology, 2012. **423**(2): p. 195-204.
25. Kaur, P. and A. Li, *Adhesive properties of human basal epidermal cells: an analysis of keratinocyte stem cells, transit amplifying cells, and postmitotic differentiating cells.* J Invest Dermatol, 2000. **114**(3): p. 413-20.
26. Egawa, K., *Do human papillomaviruses target epidermal stem cells?* Dermatology, 2003. **207**(3): p. 251-4.
27. Schafer, G., M.J. Blumenthal, and A.A. Katz, *Interaction of human tumor viruses with host cell surface receptors and cell entry.* Viruses, 2015. **7**(5): p. 2592-617.
28. Day, P.M. and M. Schelhaas, *Concepts of papillomavirus entry into host cells.* Curr Opin Virol, 2014. **4**: p. 24-31.
29. Pyeon, D., et al., *Establishment of human papillomavirus infection requires cell cycle progression.* PLoS Pathog, 2009. **5**(2): p. e1000318.
30. McBride, A.A., *Mechanisms and strategies of papillomavirus replication.* Biol Chem, 2017. **398**(8): p. 919-927.
31. Munger, K. and P.M. Howley, *Human papillomavirus immortalization and transformation functions.* Virus Res, 2002. **89**(2): p. 213-28.

32. Doorbar, J., et al., *The biology and life-cycle of human papillomaviruses.* Vaccine, 2012. **30 Suppl 5**: p. F55-70.
33. Ruesch, M.N. and L.A. Laimins, *Human papillomavirus oncoproteins alter differentiation-dependent cell cycle exit on suspension in semisolid medium.* Virology, 1998. **250**(1): p. 19-29.
34. Peh, W.L., et al., *Life cycle heterogeneity in animal models of human papillomavirus-associated disease.* J Virol, 2002. **76**(20): p. 10401-16.
35. Duensing, S. and K. Munger, *Mechanisms of genomic instability in human cancer: insights from studies with human papillomavirus oncoproteins.* Int J Cancer, 2004. **109**(2): p. 157-62.
36. Matsukura, T., S. Koi, and M. Sugase, *Both episomal and integrated forms of human papillomavirus type 16 are involved in invasive cervical cancers.* Virology, 1989. **172**(1): p. 63-72.
37. Vinokurova, S., et al., *Type-dependent integration frequency of human papillomavirus genomes in cervical lesions.* Cancer Res, 2008. **68**(1): p. 307-13.
38. Andersen, F.F., et al., *The RNA splicing factor ASF/SF2 inhibits human topoisomerase I mediated DNA relaxation.* J Mol Biol, 2002. **322**(4): p. 677-86.
39. Doorbar, J., *The papillomavirus life cycle.* J Clin Virol, 2005. **32 Suppl 1**: p. S7-15.
40. Johansson, C. and S. Schwartz, *Regulation of human papillomavirus gene expression by splicing and polyadenylation.* Nat Rev Microbiol, 2013. **11**(4): p. 239-51.
41. Kämmer, C., et al., *Sequence analysis of the long control region of human papillomavirus type 16 variants and functional consequences for P97 promoter activity*. Vol. 81. 2000. 1975-81.
42. Lopez-Saavedra, A., et al., *Functional implication of sequence variation in the long control region and E2 gene among human papillomavirus type 18 variants.* Arch Virol, 2009. **154**(5): p. 747-54.
43. Moody, C.A. and L.A. Laimins, *Human papillomavirus oncoproteins: pathways to transformation.* Nat Rev Cancer, 2010. **10**(8): p. 550-60.
44. Johansson, C. and S. Schwartz, *Regulation of human papillomavirus gene expression by splicing and polyadenylation.* Nature Reviews Microbiology, 2013. **11**: p. 239.
45. McLaughlin-Drubin, M.E. and K. Münger, *The human papillomavirus E7 oncoprotein.* Virology, 2009. **384**(2): p. 335-344.
46. Phelps, W.C., et al., *The human papillomavirus type 16 E7 gene encodes transactivation and transformation functions similar to those of adenovirus E1A.* Cell, 1988. **53**(4): p. 539-47.
47. Jones, R.E., et al., *Specific N-methylations of HPV-16 E7 peptides alter binding to the retinoblastoma suppressor protein.* J Biol Chem, 1992. **267**(2): p. 908-12.

48. Patrick, D.R., A. Oliff, and D.C. Heimbrook, *Identification of a novel retinoblastoma gene product binding site on human papillomavirus type 16 E7 protein.* J Biol Chem, 1994. **269**(9): p. 6842-50.
49. Johannsen, E. and P.F. Lambert, *Epigenetics of human papillomaviruses.* Virology, 2013. **445**(1-2): p. 205-12.
50. Kato, J., et al., *Direct binding of cyclin D to the retinoblastoma gene product (pRb) and pRb phosphorylation by the cyclin D-dependent kinase CDK4.* Genes Dev, 1993. **7**(3): p. 331-42.
51. Munger, K., et al., *Biological activities and molecular targets of the human papillomavirus E7 oncoprotein.* Oncogene, 2001. **20**(54): p. 7888-98.
52. Moody, C.A. and L.A. Laimins, *Human papillomaviruses activate the ATM DNA damage pathway for viral genome amplification upon differentiation.* PLoS Pathog, 2009. **5**(10): p. e1000605.
53. McLaughlin-Drubin, M.E., K.W. Huh, and K. Munger, *Human papillomavirus type 16 E7 oncoprotein associates with E2F6.* J Virol, 2008. **82**(17): p. 8695-705.
54. Huh, K.W., et al., *Association of the human papillomavirus type 16 E7 oncoprotein with the 600-kDa retinoblastoma protein-associated factor, p600.* Proc Natl Acad Sci U S A, 2005. **102**(32): p. 11492-7.
55. Martinez-Zapien, D., et al., *Structure of the E6/E6AP/p53 complex required for HPV-mediated degradation of p53.* Nature, 2016. **529**(7587): p. 541-5.
56. Huibregtse, J.M., M. Scheffner, and P.M. Howley, *A cellular protein mediates association of p53 with the E6 oncoprotein of human papillomavirus types 16 or 18.* Embo j, 1991. **10**(13): p. 4129-35.
57. Howie, H.L., R.A. Katzenellenbogen, and D.A. Galloway, *Papillomavirus E6 proteins.* Virology, 2009. **384**(2): p. 324-334.
58. Scheffner, M., et al., *The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53.* Cell, 1993. **75**(3): p. 495-505.
59. Lechner, M.S. and L.A. Laimins, *Inhibition of p53 DNA binding by human papillomavirus E6 proteins.* J Virol, 1994. **68**(7): p. 4262-73.
60. Crook, T., J.A. Tidy, and K.H. Vousden, *Degradation of p53 can be targeted by HPV E6 sequences distinct from those required for p53 binding and trans-activation.* Cell, 1991. **67**(3): p. 547-56.
61. Li, X. and P. Coffino, *High-risk human papillomavirus E6 protein has two distinct binding sites within p53, of which only one determines degradation.* J Virol, 1996. **70**(7): p. 4509-16.
62. Patel, D., et al., *The E6 protein of human papillomavirus type 16 binds to and inhibits co-activation by CBP and p300.* Embo j, 1999. **18**(18): p. 5061-72.
63. Zimmermann, H., et al., *Interaction with CBP/p300 enables the bovine papillomavirus type 1 E6 oncoprotein to downregulate CBP/p300-mediated transactivation by p53.* J Gen Virol, 2000. **81**(Pt 11): p. 2617-23.

64. Mantovani, F. and L. Banks, *Inhibition of E6 induced degradation of p53 is not sufficient for stabilization of p53 protein in cervical tumour derived cell lines.* Oncogene, 1999. **18**(22): p. 3309-15.
65. Hughes, F.J. and M.A. Romanos, *E1 protein of human papillomavirus is a DNA helicase/ATPase.* Nucleic acids research, 1993. **21**(25): p. 5817-5823.
66. Sedman, J. and A. Stenlund, *The papillomavirus E1 protein forms a DNA-dependent hexameric complex with ATPase and DNA helicase activities.* J Virol, 1998. **72**(8): p. 6893-7.
67. Titolo, S., et al., *Identification of domains of the human papillomavirus type 11 E1 helicase involved in oligomerization and binding to the viral origin.* Journal of virology, 2000. **74**(16): p. 7349-7361.
68. Sun, S., et al., *Identification of a 68-kilodalton nuclear ATP-binding phosphoprotein encoded by bovine papillomavirus type 1.* J Virol, 1990. **64**(10): p. 5093-105.
69. Del Vecchio, A.M., et al., *Transient replication of human papillomavirus DNAs.* J Virol, 1992. **66**(10): p. 5949-58.
70. Masterson, P.J., et al., *A C-terminal helicase domain of the human papillomavirus E1 protein binds E2 and the DNA polymerase alpha-primase p68 subunit.* J Virol, 1998. **72**(9): p. 7407-19.
71. Moody, C.A. and L.A. Laimins, *Human papillomavirus oncoproteins: pathways to transformation.* Nature Reviews Cancer, 2010. **10**: p. 550.
72. Graham, S.V. and A.A.A. Faizo, *Control of human papillomavirus gene expression by alternative splicing.* Virus Research, 2017. **231**: p. 83-95.
73. Blachon, S., et al., *Nucleo-cytoplasmic Shuttling of High Risk Human Papillomavirus E2 Proteins Induces Apoptosis.* Journal of Biological Chemistry, 2005. **280**(43): p. 36088-36098.
74. Hegde, R.S., *The papillomavirus E2 proteins: structure, function, and biology.* Annu Rev Biophys Biomol Struct, 2002. **31**: p. 343-60.
75. McBride, A.A., *The Papillomavirus E2 proteins.* Virology, 2013. **445**(1): p. 57-79.
76. Thain, A., et al., *CpG methylation directly inhibits binding of the human papillomavirus type 16 E2 protein to specific DNA sequences.* Journal of Virology, 1996. **70**(10): p. 7233.
77. Kim, K., et al., *Methylation Patterns of Papillomavirus DNA, Its Influence on E2 Function, and Implications in Viral Infection.* Journal of Virology, 2003. **77**(23): p. 12450.
78. McBride, A.A. and M.K. Jang, *Current understanding of the role of the Brd4 protein in the papillomavirus lifecycle.* Viruses, 2013. **5**(6): p. 1374-94.
79. Fujii, T., et al., *High and Low Levels of Cottontail Rabbit Papillomavirus E2 Protein Generate Opposite Effects on Gene Expression.* Journal of Biological Chemistry, 2001. **276**(2): p. 867-874.

80. Bouvard, V., et al., *Characterization of the human papillomavirus E2 protein: evidence of trans-activation and trans-repression in cervical keratinocytes.* Embo j, 1994. **13**(22): p. 5451-9.
81. Steger, G. and S. Corbach, *Dose-dependent regulation of the early promoter of human papillomavirus type 18 by the viral E2 protein.* J Virol, 1997. **71**(1): p. 50-8.
82. Thierry, F. and M. Yaniv, *The BPV1-E2 trans-acting protein can be either an activator or a repressor of the HPV18 regulatory region.* The EMBO Journal, 1987. **6**(11): p. 3391-3397.
83. Sanders, C.M. and A. Stenlund, *Recruitment and loading of the E1 initiator protein: an ATP-dependent process catalysed by a transcription factor.* The EMBO journal, 1998. **17**(23): p. 7044-7055.
84. Sanders, C.M. and A. Stenlund, *Transcription factor-dependent loading of the E1 initiator reveals modular assembly of the papillomavirus origin melting complex.* J Biol Chem, 2000. **275**(5): p. 3522-34.
85. Smith, J.A., et al., *SMCX and components of the TIP60 complex contribute to E2 regulation of the HPV E6/E7 promoter.* Virology, 2014. **468-470**: p. 311-321.
86. Bernard, H.U., *Gene expression of genital human papillomaviruses and considerations on potential antiviral approaches.* Antivir Ther, 2002. **7**(4): p. 219-37.
87. Smith, P.P., et al., *Viral integration and fragile sites in human papillomavirus-immortalized human keratinocyte cell lines.* Genes Chromosomes Cancer, 1992. **5**(2): p. 150-7.
88. Ziegert, C., et al., *A comprehensive analysis of HPV integration loci in anogenital lesions combining transcript and genome-based amplification techniques.* Oncogene, 2003. **22**(25): p. 3977-84.
89. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer.* Cell, 2000. **100**(1): p. 57-70.
90. You, J., et al., *Interaction of the Bovine Papillomavirus E2 Protein with Brd4 Tethers the Viral DNA to Host Mitotic Chromosomes.* Cell, 2004. **117**(3): p. 349-360.
91. Baxter, M.K., et al., *The Mitotic Chromosome Binding Activity of the Papillomavirus E2 Protein Correlates with Interaction with the Cellular Chromosomal Protein, Brd4.* Journal of Virology, 2005. **79**(8): p. 4806.
92. Ilves, I., et al., *Brd4 Is Involved in Multiple Processes of the Bovine Papillomavirus Type 1 Life Cycle.* Journal of Virology, 2006. **80**(7): p. 3660.
93. Oliveira, J.G., L.A. Colf, and A.A. McBride, *Variations in the association of papillomavirus E2 proteins with mitotic chromosomes.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(4): p. 1047.

94. Donaldson, M.M., W. Boner, and I.M. Morgan, *TopBP1 Regulates Human Papillomavirus Type 16 E2 Interaction with Chromatin.* Journal of Virology, 2007. **81**(8): p. 4338.
95. Gillespie, K.A., et al., *Human Papillomaviruses Recruit Cellular DNA Repair and Homologous Recombination Factors to Viral Replication Centers.* Journal of Virology, 2012. **86**(17): p. 9520.
96. Fradet-Turcotte, A., et al., *Nuclear Accumulation of the Papillomavirus E1 Helicase Blocks S-Phase Progression and Triggers an ATM-Dependent DNA Damage Response.* Journal of Virology, 2011. **85**(17): p. 8996.
97. Dao, L.D., et al., *Dynamic Localization of the Human Papillomavirus Type 11 Origin Binding Protein E2 through Mitosis While in Association with the Spindle Apparatus.* Journal of Virology, 2006. **80**(10): p. 4792.
98. Parish, J.L., et al., *ChlR1 Is Required for Loading Papillomavirus E2 onto Mitotic Chromosomes and Viral Genome Maintenance.* Molecular Cell, 2006. **24**(6): p. 867-876.
99. Yu, T., Y.-C. Peng, and E.J. Androphy, *Mitotic Kinesin-Like Protein 2 Binds and Colocalizes with Papillomavirus E2 during Mitosis.* Journal of Virology, 2007. **81**(4): p. 1736.
100. Stubenrauch, F., et al., *The E8^E2C Protein, a Negative Regulator of Viral Transcription and Replication, Is Required for Extrachromosomal Maintenance of Human Papillomavirus Type 31 in Keratinocytes.* Journal of Virology, 2000. **74**(3): p. 1178.
101. Dreer, M., S. van de Poel, and F. Stubenrauch, *Control of viral replication and transcription by the papillomavirus E8^E2 protein.* Virus Res, 2017. **231**: p. 96-102.
102. Doorbar, J., et al., *Detection of novel splicing patterns in a HPV16-containing keratinocyte cell line.* Virology, 1990. **178**(1): p. 254-262.
103. Milligan, S.G., et al., *Analysis of novel human papillomavirus type 16 late mRNAs in differentiated W12 cervical epithelial cells.* Virology, 2007. **360**(1): p. 172-181.
104. Middleton, K., et al., *Organization of human papillomavirus productive cycle during neoplastic progression provides a basis for selection of diagnostic markers.* J Virol, 2003. **77**(19): p. 10186-201.
105. Peh, W.L., et al., *The Viral E4 Protein Is Required for the Completion of the Cottontail Rabbit Papillomavirus Productive Cycle In Vivo.* Journal of Virology, 2004. **78**(4): p. 2142.
106. Doorbar, J., *The E4 protein; structure, function and patterns of expression.* Virology, 2013. **445**(1-2): p. 80-98.
107. Wilson, R., et al., *The full-length E1^E4 protein of human papillomavirus type 18 modulates differentiation-dependent viral DNA amplification and late gene expression.* Virology, 2007. **362**(2): p. 453-460.

108. Nakahara, T., et al., *Human Papillomavirus Type 16 E1[^]E4 Contributes to Multiple Facets of the Papillomavirus Life Cycle.* Journal of Virology, 2005. **79**(20): p. 13150.
109. Davy, C.E., et al., *Identification of a G₂ Arrest Domain in the E1^E4 Protein of Human Papillomavirus Type 16.* Journal of Virology, 2002. **76**(19): p. 9806.
110. Davy, C.E., et al., *Human Papillomavirus Type 16 E1[^]E4-Induced G₂ Arrest Is Associated with Cytoplasmic Retention of Active Cdk1/Cyclin B1 Complexes.* Journal of Virology, 2005. **79**(7): p. 3998.
111. McIntosh, P.B., et al., *E1^E4-mediated keratin phosphorylation and ubiquitylation: a mechanism for keratin depletion in HPV16-infected epithelium.* Journal of Cell Science, 2010. **123**(16): p. 2810.
112. Davy, C.E., et al., *HPV16 E1^E4 protein is phosphorylated by Cdk2/cyclin A and relocalizes this complex to the cytoplasm.* Virology, 2006. **349**(1): p. 230-244.
113. Davy, C., et al., *A novel interaction between the human papillomavirus type 16 E2 and E1^E4 proteins leads to stabilization of E2.* Virology, 2009. **394**(2): p. 266-275.
114. Wang, Q., et al., *Phosphorylation of the Human Papillomavirus Type 16 E1^E4 Protein at T57 by ERK Triggers a Structural Change That Enhances Keratin Binding and Protein Stability.* Journal of Virology, 2009. **83**(8): p. 3668.
115. Conrad, M., V.J. Bubb, and R. Schlegel, *The human papillomavirus type 6 and 16 E5 proteins are membrane-associated proteins which associate with the 16-kilodalton pore-forming protein.* J Virol, 1993. **67**(10): p. 6170-8.
116. Thomsen, P., et al., *The HPV16 E5 oncogene inhibits endocytic trafficking.* Oncogene, 2000. **19**(52): p. 6023-32.
117. Lewis, C., et al., *The first hydrophobic region of the HPV16 E5 protein determines protein cellular location and facilitates anchorage-independent growth.* Virol J, 2008. **5**: p. 30.
118. Bouvard, V., et al., *The human papillomavirus type 16 E5 gene cooperates with the E7 gene to stimulate proliferation of primary cells and increases viral gene expression.* Virology, 1994. **203**(1): p. 73-80.
119. Oelze, I., et al., *Human papillomavirus type 16 E5 protein affects cell-cell communication in an epithelial cell line.* Journal of virology, 1995. **69**(7): p. 4489-4494.
120. Valle, G.F. and L. Banks, *The human papillomavirus (HPV)-6 and HPV-16 E5 proteins co-operate with HPV-16 E7 in the transformation of primary rodent cells.* J Gen Virol, 1995. **76 (Pt 5)**: p. 1239-45.
121. Stoppler, M.C., et al., *The E5 gene of HPV-16 enhances keratinocyte immortalization by full-length DNA.* Virology, 1996. **223**(1): p. 251-4.

122. Straight, S.W., B. Herman, and D.J. McCance, *The E5 oncoprotein of human papillomavirus type 16 inhibits the acidification of endosomes in human keratinocytes.* J Virol, 1995. **69**(5): p. 3185-92.
123. Dannenberg, A.J., et al., *Cyclooxygenase-2 and epidermal growth factor receptor: pharmacologic targets for chemoprevention.* J Clin Oncol, 2005. **23**(2): p. 254-66.
124. Pedroza-Saavedra, A., et al., *The human papillomavirus type 16 E5 oncoprotein synergizes with EGF-receptor signaling to enhance cell cycle progression and the down-regulation of p27(Kip1).* Virology, 2010. **400**(1): p. 44-52.
125. Marchetti, B., et al., *The bovine papillomavirus oncoprotein E5 retains MHC class I molecules in the Golgi apparatus and prevents their transport to the cell surface.* Oncogene, 2002. **21**(51): p. 7808-7816.
126. Zhang, B., et al., *The E5 protein of human papillomavirus type 16 perturbs MHC class II antigen maturation in human foreskin keratinocytes treated with interferon-gamma.* Virology, 2003. **310**(1): p. 100-8.
127. Ashrafi, G.H., et al., *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): p. 2105-12.
128. Buck, C.B. and B.L. Trus, *The Papillomavirus Virion: A Machine Built to Hide Molecular Achilles' Heels*, in *Viral Molecular Machines*, M.G. Rossmann and V.B. Rao, Editors. 2012, Springer US: Boston, MA. p. 403-422.
129. Buck, C.B., et al., *Arrangement of L2 within the Papillomavirus Capsid.* Journal of Virology, 2008. **82**(11): p. 5190.
130. Wang, J.W. and R.B. Roden, *L2, the minor capsid protein of papillomavirus.* Virology, 2013. **445**(1-2): p. 175-86.
131. Giroglou, T., et al., *Human papillomavirus infection requires cell surface heparan sulfate.* J Virol, 2001. **75**(3): p. 1565-70.
132. Johnson, K.M., et al., *Role of heparan sulfate in attachment to and infection of the murine female genital tract by human papillomavirus.* J Virol, 2009. **83**(5): p. 2067-74.
133. Day, P.M. and J.T. Schiller, *The role of furin in papillomavirus infection.* Future Microbiology, 2009. **4**(10): p. 1255-1262.
134. Kines, R.C., et al., *The initial steps leading to papillomavirus infection occur on the basement membrane prior to cell surface binding.* Proceedings of the National Academy of Sciences, 2009. **106**(48): p. 20458.
135. Day, P.M., D.R. Lowy, and J.T. Schiller, *Heparan Sulfate-Independent Cell Binding and Infection with Furin-Precleaved Papillomavirus Capsids.* Journal of Virology, 2008. **82**(24): p. 12565.
136. Kämper, N., et al., *A Membrane-Destabilizing Peptide in Capsid Protein L2 Is Required for Egress of Papillomavirus Genomes from Endosomes.* Journal of Virology, 2006. **80**(2): p. 759.

137. Kondo, K., et al., *Nuclear location of minor capsid protein L2 is required for expression of a reporter plasmid packaged in HPV51 pseudovirions.* Virology, 2009. **394**(2): p. 259-265.
138. Roden, R.B., et al., *In vitro generation and type-specific neutralization of a human papillomavirus type 16 virion pseudotype.* J Virol, 1996. **70**(9): p. 5875-83.
139. Stauffer, Y., et al., *Infectious human papillomavirus type 18 pseudovirions.* J Mol Biol, 1998. **283**(3): p. 529-36.
140. Campos, S.K. and M.A. Ozbun, *Two Highly Conserved Cysteine Residues in HPV16 L2 Form an Intramolecular Disulfide Bond and Are Critical for Infectivity in Human Keratinocytes.* PLOS ONE, 2009. **4**(2): p. e4463.
141. Campo, M.S., et al., *A Peptide Encoding a B-Cell Epitope from the N-Terminus of the Capsid Protein L2 of Bovine Papillomavirus-4 Prevents Disease.* Virology, 1997. **234**(2): p. 261-266.
142. Shatkin, A.J. and J.L. Manley, *The ends of the affair: capping and polyadenylation.* Nat Struct Biol, 2000. **7**(10): p. 838-42.
143. Moteki, S. and D. Price, *Functional coupling of capping and transcription of mRNA.* Mol Cell, 2002. **10**(3): p. 599-609.
144. Ahola, T. and L. Kääriäinen, *Reaction in alphavirus mRNA capping: formation of a covalent complex of nonstructural protein nsP1 with 7-methyl-GMP.* Proceedings of the National Academy of Sciences of the United States of America, 1995. **92**(2): p. 507-511.
145. Topisirovic, I., et al., *Cap and cap-binding proteins in the control of gene expression.* Wiley Interdiscip Rev RNA, 2011. **2**(2): p. 277-98.
146. Hocine, S., R.H. Singer, and D. Grunwald, *RNA processing and export.* Cold Spring Harb Perspect Biol, 2010. **2**(12): p. a000752.
147. Wu, X. and G. Brewer, *The regulation of mRNA stability in mammalian cells: 2.0.* Gene, 2012. **500**(1): p. 10-21.
148. Bienroth, S., W. Keller, and E. Wahle, *Assembly of a processive messenger RNA polyadenylation complex.* The EMBO journal, 1993. **12**(2): p. 585-594.
149. Beaudoing, E., et al., *Patterns of variant polyadenylation signal usage in human genes.* Genome research, 2000. **10**(7): p. 1001-1010.
150. Balbo, P.B. and A. Bohm, *Mechanism of poly(A) polymerase: structure of the enzyme-MgATP-RNA ternary complex and kinetic analysis.* Structure (London, England : 1993), 2007. **15**(9): p. 1117-1131.
151. Wahle, E., *Poly(A) tail length control is caused by termination of processive synthesis.* J Biol Chem, 1995. **270**(6): p. 2800-8.
152. Berget, S.M., C. Moore, and P.A. Sharp, *Spliced segments at the 5' terminus of adenovirus 2 late mRNA.* Proceedings of the National Academy of Sciences of the United States of America, 1977. **74**(8): p. 3171-3175.
153. Black, D.L., *Mechanisms of Alternative Pre-Messenger RNA Splicing.* Annual Review of Biochemistry, 2003. **72**(1): p. 291-336.

154. Hallegger, M., M. Llorian, and C.W. Smith, *Alternative splicing: global insights.* Febs j, 2010. **277**(4): p. 856-66.
155. Taggart, A.J., et al., *Large-scale mapping of branchpoints in human pre-mRNA transcripts in vivo.* Nat Struct Mol Biol, 2012. **19**(7): p. 719-21.
156. Moore, M.J. and P.A. Sharp, *Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing.* Nature, 1993. **365**(6444): p. 364-8.
157. Modrek, B. and C. Lee, *A genomic view of alternative splicing.* Nat Genet, 2002. **30**(1): p. 13-9.
158. Das, R., Z. Zhou, and R. Reed, *Functional association of U2 snRNP with the ATP-independent spliceosomal complex E.* Mol Cell, 2000. **5**(5): p. 779-87.
159. Nilsen, T.W., *A Parallel Spliceosome.* Science, 1996. **273**(5283): p. 1813.
160. Wang, Z., et al., *General and specific functions of exonic splicing silencers in splicing control.* Mol Cell, 2006. **23**(1): p. 61-70.
161. Hertel, K.J. and T. Maniatis, *Serine-arginine (SR)-rich splicing factors have an exon-independent function in pre-mRNA splicing.* Proc Natl Acad Sci U S A, 1999. **96**(6): p. 2651-5.
162. Graveley, B.R., K.J. Hertel, and T. Maniatis, *SR proteins are 'locators' of the RNA splicing machinery.* Curr Biol, 1999. **9**(1): p. R6-7.
163. Han, S.P., Y.H. Tang, and R. Smith, *Functional diversity of the hnRNPs: past, present and perspectives.* Biochem J, 2010. **430**(3): p. 379-92.
164. Busch, A. and K.J. Hertel, *Evolution of SR protein and hnRNP splicing regulatory factors.* Wiley Interdiscip Rev RNA, 2012. **3**(1): p. 1-12.
165. Lim, E., et al., *Using positional distribution to identify splicing elements and predict pre-mRNA processing defects in human genes*. Vol. 108. 2011. 11093-8.
166. Caceres, J.F., G.R. Screaton, and A.R. Krainer, *A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm.* Genes Dev, 1998. **12**(1): p. 55-66.
167. Michlewski, G., J.R. Sanford, and J.F. Caceres, *The splicing factor SF2/ASF regulates translation initiation by enhancing phosphorylation of 4E-BP1.* Mol Cell, 2008. **30**(2): p. 179-89.
168. Bedard, K.M., S. Daijogo, and B.L. Semler, *A nucleo-cytoplasmic SR protein functions in viral IRES-mediated translation initiation.* Embo j, 2007. **26**(2): p. 459-67.
169. Long, J.C. and J.F. Caceres, *The SR protein family of splicing factors: master regulators of gene expression.* Biochem J, 2009. **417**(1): p. 15-27.
170. Sanford, J.R., J. Ellis, and J.F. Caceres, *Multiple roles of arginine/serine-rich splicing factors in RNA processing.* Biochem Soc Trans, 2005. **33**(Pt 3): p. 443-6.
171. Shepard, P.J. and K.J. Hertel, *The SR protein family.* Genome Biol, 2009. **10**(10): p. 242.

172. Caceres, J.F., et al., *Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity.* J Cell Biol, 1997. **138**(2): p. 225-38.
173. Hedley, M.L., H. Amrein, and T. Maniatis, *An amino acid sequence motif sufficient for subnuclear localization of an arginine/serine-rich splicing factor.* Proc Natl Acad Sci U S A, 1995. **92**(25): p. 11524-8.
174. Graveley, B.R., *Sorting out the complexity of SR protein functions.* Rna, 2000. **6**(9): p. 1197-211.
175. Shen, H., J.L. Kan, and M.R. Green, *Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly.* Mol Cell, 2004. **13**(3): p. 367-76.
176. Shen, H. and M.R. Green, *A pathway of sequential arginine-serine-rich domain-splicing signal interactions during mammalian spliceosome assembly.* Mol Cell, 2004. **16**(3): p. 363-73.
177. Kanopka, A., O. Mühlemann, and G. Akusjärvi, *Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA.* Nature, 1996. **381**(6582): p. 535-538.
178. Martinez-Contreras, R., et al., *hnRNP proteins and splicing control.* Adv Exp Med Biol, 2007. **623**: p. 123-47.
179. Dreyfuss, G., et al., *hnRNP proteins and the biogenesis of mRNA.* Annu Rev Biochem, 1993. **62**: p. 289-321.
180. Zhao, T.T., et al., *hnRNP A1 regulates UV-induced NF-kappaB signalling through destabilization of cIAP1 mRNA.* Cell Death Differ, 2009. **16**(2): p. 244-52.
181. Fahling, M., et al., *Heterogeneous nuclear ribonucleoprotein-A2/B1 modulate collagen prolyl 4-hydroxylase, alpha (I) mRNA stability.* J Biol Chem, 2006. **281**(14): p. 9279-86.
182. Alkan, S.A., K. Martincic, and C. Milcarek, *The hnRNPs F and H2 bind to similar sequences to influence gene expression.* Biochem J, 2006. **393**(Pt 1): p. 361-71.
183. Bomsztyk, K., O. Denisenko, and J. Ostrowski, *hnRNP K: one protein multiple processes.* Bioessays, 2004. **26**(6): p. 629-38.
184. Geuens, T., D. Bouhy, and V. Timmerman, *The hnRNP family: insights into their role in health and disease.* Human genetics, 2016. **135**(8): p. 851-867.
185. Gorlach, M., et al., *Interaction of the RNA-binding domain of the hnRNP C proteins with RNA.* Embo j, 1992. **11**(9): p. 3289-95.
186. Kiledjian, M. and G. Dreyfuss, *Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box.* Embo j, 1992. **11**(7): p. 2655-64.
187. Dominguez, C. and F.H. Allain, *NMR structure of the three quasi RNA recognition motifs (qRRMs) of human hnRNP F and interaction studies with Bcl-x G-tract RNA: a novel mode of RNA recognition.* Nucleic Acids Res, 2006. **34**(13): p. 3634-45.

188. Siomi, H., et al., *The pre-mRNA binding K protein contains a novel evolutionarily conserved motif.* Nucleic Acids Res, 1993. **21**(5): p. 1193-8.
189. Chaudhury, A., P. Chander, and P.H. Howe, *Heterogeneous nuclear ribonucleoproteins (hnRNPs) in cellular processes: Focus on hnRNP E1's multifunctional regulatory roles.* RNA (New York, N.Y.), 2010. **16**(8): p. 1449-1462.
190. Elliott, D.J., et al., *An evolutionarily conserved germ cell-specific hnRNP is encoded by a retrotransposed gene.* Hum Mol Genet, 2000. **9**(14): p. 2117-24.
191. A. Lingenfelter, P., et al., *Expression and conservation of processed copies of the RBMX gene*. Vol. 12. 2001. 538-45.
192. Elliott, D.J., *The role of potential splicing factors including RBMY, RBMX, hnRNPG-T and STAR proteins in spermatogenesis.* Int J Androl, 2004. **27**(6): p. 328-34.
193. Soulard, M., et al., *hnRNP G: sequence and characterization of a glycosylated RNA-binding protein.* Nucleic acids research, 1993. **21**(18): p. 4210-4217.
194. Kanhoush, R., et al., *Novel domains in the hnRNP G/RBMX protein with distinct roles in RNA binding and targeting nascent transcripts.* Nucleus (Austin, Tex.), 2010. **1**(1): p. 109-122.
195. Heinrich, B., et al., *Heterogeneous nuclear ribonucleoprotein G regulates splice site selection by binding to CC(A/C)-rich regions in pre-mRNA.* The Journal of biological chemistry, 2009. **284**(21): p. 14303-14315.
196. Hofmann, Y. and B. Wirth, *hnRNP-G promotes exon 7 inclusion of survival motor neuron (SMN) via direct interaction with Htra2-beta1.* Hum Mol Genet, 2002. **11**(17): p. 2037-49.
197. Maziuk, B.F., et al., *RNA binding proteins co-localize with small tau inclusions in tauopathy.* Acta Neuropathologica Communications, 2018. **6**(1): p. 71.
198. Nasim, M.T., et al., *HnRNP G and Tra2β: opposite effects on splicing matched by antagonism in RNA binding.* Human Molecular Genetics, 2003. **12**(11): p. 1337-1348.
199. Shin, K.H., et al., *Heterogeneous nuclear ribonucleoprotein G shows tumor suppressive effect against oral squamous cell carcinoma cells.* Clin Cancer Res, 2006. **12**(10): p. 3222-8.
200. Adamson, B., et al., *A genome-wide homologous recombination screen identifies the RNA-binding protein RBMX as a component of the DNA-damage response.* Nature cell biology, 2012. **14**(3): p. 318-328.
201. Shin, K.H., et al., *p53 promotes the fidelity of DNA end-joining activity by, in part, enhancing the expression of heterogeneous nuclear ribonucleoprotein G.* DNA Repair (Amst), 2007. **6**(6): p. 830-40.

202. Abdelmohsen, K. and M. Gorospe, *Posttranscriptional regulation of cancer traits by HuR.* Wiley Interdisciplinary Reviews: RNA, 2010. **1**(2): p. 214-229.
203. Brennan, C.M. and J.A. Steitz, *HuR and mRNA stability.* Cell Mol Life Sci, 2001. **58**(2): p. 266-77.
204. Chen, C.-Y.A., N. Xu, and A.-B. Shyu, *Highly Selective Actions of HuR in Antagonizing AU-Rich Element-Mediated mRNA Destabilization.* Molecular and Cellular Biology, 2002. **22**(20): p. 7268.
205. Ghosh, M., et al., *Essential role of the RNA-binding protein HuR in progenitor cell survival in mice.* J Clin Invest, 2009. **119**(12): p. 3530-43.
206. Hinman, M.N. and H. Lou, *Diverse molecular functions of Hu proteins.* Cell Mol Life Sci, 2008. **65**(20): p. 3168-81.
207. Gupta, S., et al., *Molecular Determinants of Retinoic Acid Sensitivity in Pancreatic Cancer.* Clinical Cancer Research, 2012. **18**(1): p. 280.
208. Izquierdo, J.M., *Hu antigen R (HuR) functions as an alternative pre-mRNA splicing regulator of Fas apoptosis-promoting receptor on exon definition.* J Biol Chem, 2008. **283**(27): p. 19077-84.
209. Zhu, H., et al., *Hu Proteins Regulate Polyadenylation by Blocking Sites Containing U-rich Sequences.* Journal of Biological Chemistry, 2007. **282**(4): p. 2203-2210.
210. Ruegsegger, U., K. Beyer, and W. Keller, *Purification and characterization of human cleavage factor Im involved in the 3' end processing of messenger RNA precursors.* J Biol Chem, 1996. **271**(11): p. 6107-13.
211. Gaudet, P., et al., *Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium.* Brief Bioinform, 2011. **12**(5): p. 449-62.
212. Zhu, Y., et al., *Molecular Mechanisms for CFIm-Mediated Regulation of mRNA Alternative Polyadenylation.* Mol Cell, 2018. **69**(1): p. 62-74.e4.
213. Grassmann, K., et al., *Identification of a differentiation-inducible promoter in the E7 open reading frame of human papillomavirus type 16 (HPV-16) in raft cultures of a new cell line containing high copy numbers of episomal HPV-16 DNA.* J Virol, 1996. **70**(4): p. 2339-49.
214. Zhao, X., et al., *A 57-nucleotide upstream early polyadenylation element in human papillomavirus type 16 interacts with hFip1, CstF-64, hnRNP C1/C2, and polypyrimidine tract binding protein.* Journal of virology, 2005. **79**(7): p. 4270-4288.
215. Oberg, D., et al., *A downstream polyadenylation element in human papillomavirus type 16 L2 encodes multiple GGG motifs and interacts with hnRNP H.* J Virol, 2005. **79**(14): p. 9254-69.
216. Johansson, C., et al., *HPV-16 E2 contributes to induction of HPV-16 late gene expression by inhibiting early polyadenylation.* The EMBO journal, 2012. **31**(14): p. 3212-3227.

217. Graham, S.V., *Papillomavirus 3′ UTR regulatory elements.* Front Biosci, 2008. **13**: p. 5646-63.
218. Jeon, S., B.L. Allen-Hoffmann, and P.F. Lambert, *Integration of human papillomavirus type 16 into the human genome correlates with a selective growth advantage of cells.* J Virol, 1995. **69**(5): p. 2989-97.
219. Cumming, S.A., et al., *The RNA stability regulator HuR regulates L1 protein expression in vivo in differentiating cervical epithelial cells.* Virology, 2009. **383**(1): p. 142-149.
220. Somberg, M. and S. Schwartz, *Multiple ASF/SF2 sites in the human papillomavirus type 16 (HPV-16) E4-coding region promote splicing to the most commonly used 3'-splice site on the HPV-16 genome.* Journal of virology, 2010. **84**(16): p. 8219-8230.
221. Cheunim, T., et al., *The alternative splicing factor hnRNP A1 is up-regulated during virus-infected epithelial cell differentiation and binds the human papillomavirus type 16 late regulatory element.* Virus research, 2008. **131**(2): p. 189-198.
222. Ajiro, M. and Z.M. Zheng, *E6^E7, a novel splice isoform protein of human papillomavirus 16, stabilizes viral E6 and E7 oncoproteins via HSP90 and GRP78.* MBio, 2015. **6**(1): p. e02068-14.
223. Ajiro, M., et al., *Intron definition and a branch site adenosine at nt 385 control RNA splicing of HPV16 E6*I and E7 expression.* PLoS One, 2012. **7**(10): p. e46412.
224. Rosenberger, S., et al., *Alternative splicing of human papillomavirus type-16 E6/E6* early mRNA is coupled to EGF signaling via Erk1/2 activation.* Proc Natl Acad Sci U S A, 2010. **107**(15): p. 7006-11.
225. Schmitt, M., et al., *Diagnosing cervical cancer and high-grade precursors by HPV16 transcription patterns.* Cancer Res, 2010. **70**(1): p. 249-56.
226. Rush, M., X. Zhao, and S. Schwartz, *A splicing enhancer in the E4 coding region of human papillomavirus type 16 is required for early mRNA splicing and polyadenylation as well as inhibition of premature late gene expression.* J Virol, 2005. **79**(18): p. 12002-15.
227. Cartegni, L., et al., *ESEfinder: A web resource to identify exonic splicing enhancers.* Nucleic Acids Res, 2003. **31**(13): p. 3568-71.
228. Jia, R., et al., *Control of the papillomavirus early-to-late switch by differentially expressed SRp20.* J Virol, 2009. **83**(1): p. 167-80.
229. Somberg, M., et al., *Serine/arginine-rich protein 30c activates human papillomavirus type 16 L1 mRNA expression via a bimodal mechanism.* J Gen Virol, 2011. **92**(Pt 10): p. 2411-21.
230. Zhao, X., M. Rush, and S. Schwartz, *Identification of an hnRNP A1-dependent splicing silencer in the human papillomavirus type 16 L1 coding region that prevents premature expression of the late L1 gene.* J Virol, 2004. **78**(20): p. 10888-905.

231. Phillis, J.W. and H.G. Goshgarian, *Adenosine and neurotrauma: therapeutic perspectives.* Neurol Res, 2001. **23**(2-3): p. 183-9.
232. Headrick, J.P. and R.D. Lasley, *Adenosine receptors and reperfusion injury of the heart.* Handbook of experimental pharmacology, 2009(193): p. 189-214.
233. Amadio, S., et al., *Purinergic signalling at the plasma membrane: a multipurpose and multidirectional mode to deal with amyotrophic lateral sclerosis and multiple sclerosis.* J Neurochem, 2011. **116**(5): p. 796-805.
234. Cunha, R.A., *Adenosine as a neuromodulator and as a homeostatic regulator in the nervous system: different roles, different sources and different receptors.* Neurochem Int, 2001. **38**(2): p. 107-25.
235. Boison, D., et al., *Adenosine hypothesis of schizophrenia--opportunities for pharmacotherapy.* Neuropharmacology, 2012. **62**(3): p. 1527-43.
236. Hasko, G. and P. Pacher, *Regulation of macrophage function by adenosine.* Arterioscler Thromb Vasc Biol, 2012. **32**(4): p. 865-9.
237. Ryzhov, S., et al., *Host A(2B) adenosine receptors promote carcinoma growth.* Neoplasia (New York, N.Y.), 2008. **10**(9): p. 987-995.
238. Spychala, J., *Tumor-promoting functions of adenosine.* Pharmacol Ther, 2000. **87**(2-3): p. 161-73.
239. Chan, E.S. and B.N. Cronstein, *Adenosine in fibrosis.* Mod Rheumatol, 2010. **20**(2): p. 114-22.
240. Karmouty-Quintana, H., Y. Xia, and M.R. Blackburn, *Adenosine signaling during acute and chronic disease states.* J Mol Med (Berl), 2013. **91**(2): p. 173-81.
241. Sitkovsky, M.V., et al., *Hypoxia-adenosinergic immunosuppression: tumor protection by T regulatory cells and cancerous tissue hypoxia.* Clin Cancer Res, 2008. **14**(19): p. 5947-52.
242. Hasko, G., et al., *Adenosine receptors: therapeutic aspects for inflammatory and immune diseases.* Nat Rev Drug Discov, 2008. **7**(9): p. 759-70.
243. Engler, R.L., *Adenosine. The signal of life?* Circulation, 1991. **84**(2): p. 951-4.
244. Leone, R.D. and L.A. Emens, *Targeting adenosine for cancer immunotherapy.* Journal for immunotherapy of cancer, 2018. **6**(1): p. 57-57.
245. Chen, J.-F., H.K. Eltzschig, and B.B. Fredholm, *Adenosine receptors as drug targets — what are the challenges?* Nature Reviews Drug Discovery, 2013. **12**: p. 265.
246. Deussen, A., M. Borst, and J. Schrader, *Formation of S-adenosylhomocysteine in the heart. I: An index of free intracellular adenosine.* Circ Res, 1988. **63**(1): p. 240-9.
247. Boison, D., *Adenosine Kinase: Exploitation for Therapeutic Gain.* Pharmacological Reviews, 2013. **65**(3): p. 906.

248. Pak, M.A., et al., *Inhibition of adenosine kinase increases endogenous adenosine and depresses neuronal activity in hippocampal slices.* Neuropharmacology, 1994. **33**(9): p. 1049-53.
249. Schmidt, A.P., D.R. Lara, and D.O. Souza, *Proposal of a guanine-based purinergic system in the mammalian central nervous system.* Pharmacol Ther, 2007. **116**(3): p. 401-16.
250. Lanznaster, D., et al., *Guanosine: a Neuromodulator with Therapeutic Potential in Brain Disorders.* Aging Dis, 2016. **7**(5): p. 657-679.
251. Molz, S., et al., *Neuroprotective effect of guanosine against glutamate-induced cell death in rat hippocampal slices is mediated by the phosphatidylinositol-3 kinase/Akt/ glycogen synthase kinase 3beta pathway activation and inducible nitric oxide synthase inhibition.* J Neurosci Res, 2011. **89**(9): p. 1400-8.
252. Dal-Cim, T., et al., *Guanosine controls inflammatory pathways to afford neuroprotection of hippocampal slices under oxygen and glucose deprivation conditions.* J Neurochem, 2013. **126**(4): p. 437-50.
253. Galmarini, C.M., J.R. Mackey, and C. Dumontet, *Nucleoside analogues and nucleobases in cancer treatment.* Lancet Oncol, 2002. **3**(7): p. 415-24.
254. Holbein, S., et al., *Cordycepin interferes with 3' end formation in yeast independently of its potential to terminate RNA chain elongation.* Rna, 2009. **15**(5): p. 837-49.
255. Wong, Y.Y., et al., *Cordycepin inhibits protein synthesis and cell adhesion through effects on signal transduction.* J Biol Chem, 2010. **285**(4): p. 2610-21.
256. Tuli, H.S., et al., *Cordycepin: a bioactive metabolite with therapeutic potential.* Life Sci, 2013. **93**(23): p. 863-9.
257. Penman, S., M. Rosbash, and M. Penman, *Messenger and heterogeneous nuclear RNA in HeLa cells: differential inhibition by cordycepin.* Proc Natl Acad Sci U S A, 1970. **67**(4): p. 1878-85.
258. He, W., et al., *Cordycepin induces apoptosis by enhancing JNK and p38 kinase activity and increasing the protein expression of Bcl-2 pro-apoptotic molecules.* J Zhejiang Univ Sci B, 2010. **11**(9): p. 654-60.
259. Chen, Y., et al., *Cordycepin induces apoptosis of C6 glioma cells through the adenosine 2A receptor-p53-caspase-7-PARP pathway.* Chem Biol Interact, 2014. **216**: p. 17-25.
260. Nakamura, K., et al., *Effect of cordycepin (3'-deoxyadenosine) on hematogenic lung metastatic model mice.* In Vivo, 2005. **19**(1): p. 137-41.
261. Lence, T., et al., *Mechanistic insights into m(6)A RNA enzymes.* Biochim Biophys Acta Gene Regul Mech, 2019. **1862**(3): p. 222-229.
262. Lavi, S. and A.J. Shatkin, *Methylated simian virus 40-specific RNA from nuclei and cytoplasm of infected BSC-1 cells.* Proc Natl Acad Sci U S A, 1975. **72**(6): p. 2012-6.

263. Dominissini, D., et al., *Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq.* Nature, 2012. **485**(7397): p. 201-6.
264. Zhou, S., et al., *FTO regulates the chemo-radiotherapy resistance of cervical squamous cell carcinoma (CSCC) by targeting β-catenin through mRNA demethylation.* Molecular Carcinogenesis, 2018. **57**(5): p. 590-597.
265. Liu, J., et al., *m(6)A mRNA methylation regulates AKT activity to promote the proliferation and tumorigenicity of endometrial cancer.* Nat Cell Biol, 2018. **20**(9): p. 1074-1083.
266. Vu, L.P., et al., *The N6-methyladenosine (m6A)-forming enzyme METTL3 controls myeloid differentiation of normal hematopoietic and leukemia cells.* Nature Medicine, 2017. **23**: p. 1369.
267. Barbieri, I., et al., *Promoter-bound METTL3 maintains myeloid leukaemia by m6A-dependent translation control.* Nature, 2017. **552**: p. 126.
268. Rodriguez, W.J. and R.H. Parrott, *Ribavirin aerosol treatment of serious respiratory syncytial virus infection in infants.* Infect Dis Clin North Am, 1987. **1**(2): p. 425-39.
269. Willis, R.C., D.A. Carson, and J.E. Seegmiller, *Adenosine kinase initiates the major route of ribavirin activation in a cultured human cell line.* Proc Natl Acad Sci U S A, 1978. **75**(7): p. 3042-4.
270. Bougie, I. and M. Bisaillon, *Initial binding of the broad spectrum antiviral nucleoside ribavirin to the hepatitis C virus RNA polymerase.* J Biol Chem, 2003. **278**(52): p. 52471-8.
271. Fernandez-Larsson, R., et al., *Molecular analysis of the inhibitory effect of phosphorylated ribavirin on the vesicular stomatitis virus in vitro polymerase reaction.* Antimicrob Agents Chemother, 1989. **33**(10): p. 1668-73.
272. Maag, D., et al., *Hepatitis C virus RNA-dependent RNA polymerase (NS5B) as a mediator of the antiviral activity of ribavirin.* J Biol Chem, 2001. **276**(49): p. 46094-8.
273. Crotty, S., et al., *The broad-spectrum antiviral ribonucleoside ribavirin is an RNA virus mutagen.* Nat Med, 2000. **6**(12): p. 1375-9.
274. Franchetti, P. and M. Grifantini, *Nucleoside and non-nucleoside IMP dehydrogenase inhibitors as antitumor and antiviral agents.* Curr Med Chem, 1999. **6**(7): p. 599-614.
275. Kentsis, A., et al., *Ribavirin suppresses eIF4E-mediated oncogenic transformation by physical mimicry of the 7-methyl guanosine mRNA cap.* Proc Natl Acad Sci U S A, 2004. **101**(52): p. 18105-10.
276. Borden, K.L.B. and B. Culjkovic-Kraljacic, *Ribavirin as an anti-cancer therapy: acute myeloid leukemia and beyond?* Leukemia & lymphoma, 2010. **51**(10): p. 1805-1815.
277. Li, X., et al., *Eight Nucleotide Substitutions Inhibit Splicing to HPV-16 3'-Splice Site SA3358 and Reduce the Efficiency by which HPV-16 Increases*

the Life Span of Primary Human Keratinocytes. PLOS ONE, 2013. **8**(9): p. e72776.
278. Li, X., et al., *Suppression of HPV-16 late L1 5'-splice site SD3632 by binding of hnRNP D proteins and hnRNP A2/B1 to upstream AUAGUA RNA motifs.* Nucleic Acids Res, 2013. **41**(22): p. 10488-508.
279. Collier, B., et al., *Specific Inactivation of Inhibitory Sequences in the 5′ End of the Human Papillomavirus Type 16 L1 Open Reading Frame Results in Production of High Levels of L1 Protein in Human Epithelial Cells.* Journal of Virology, 2002. **76**(6): p. 2739.
280. Dignam, J.D., R.M. Lebovitz, and R.G. Roeder, *Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei.* Nucleic acids research, 1983. **11**(5): p. 1475-1489.
281. Sorm, F., et al., *5-Azacytidine, a new, highly effective cancerostatic.* Experientia, 1964. **20**(4): p. 202-3.
282. Cihak, A., *Biological effects of 5-azacytidine in eukaryotes.* Oncology, 1974. **30**(5): p. 405-22.
283. Burnstock, G. and F. Di Virgilio, *Purinergic signalling and cancer.* Purinergic signalling, 2013. **9**(4): p. 491-540.
284. Nilsson, K., et al., *The DNA damage response activates HPV16 late gene expression at the level of RNA processing.* Nucleic Acids Res, 2018. **46**(10): p. 5029-5049.
285. Park, J. and R.S. Gupta, *Adenosine kinase and ribokinase--the RK family of proteins.* Cell Mol Life Sci, 2008. **65**(18): p. 2875-96.
286. Whitmore, K.V. and H.B. Gaspar, *Adenosine Deaminase Deficiency – More Than Just an Immunodeficiency.* Frontiers in Immunology, 2016. **7**(314).
287. Al-Ahmadi, W., et al., *Alternative polyadenylation variants of the RNA binding protein, HuR: abundance, role of AU-rich elements and auto-Regulation.* Nucleic acids research, 2009. **37**(11): p. 3612-3624.
288. Lanznaster, D., et al., *Guanosine: a Neuromodulator with Therapeutic Potential in Brain Disorders.* Aging and disease, 2016. **7**(5): p. 657-679.
289. Jackson, E.K., et al., *Extracellular guanosine regulates extracellular adenosine levels.* American Journal of Physiology-Cell Physiology, 2012. **304**(5): p. C406-C421.
290. Jackson, E.K. and Z. Mi, *The guanosine-adenosine interaction exists in vivo.* The Journal of pharmacology and experimental therapeutics, 2014. **350**(3): p. 719-726.
291. Di Liberto, V., et al., *The Guanine-Based Purinergic System: The Tale of An Orphan Neuromodulation.* Frontiers in Pharmacology, 2016. **7**(158).
292. Shin, D.H., et al., *Pharmacokinetics of guanosine in rats following intravenous or intramuscular administration of a 1:1 mixture of guanosine and acriflavine, a potential antitumor agent.* Arch Pharm Res, 2008. **31**(10): p. 1347-53.

293. Kim, S.G., et al., *Enhanced anti-tumour effects of acriflavine in combination with guanosine in mice.* J Pharm Pharmacol, 1997. **49**(2): p. 216-22.
294. Iigo, M. and A. Hoshi, *Effect of guanosine on antitumor activity of fluorinated pyrimidines against P 388 leukemia.* Cancer Chemotherapy and Pharmacology, 1984. **13**(2): p. 86-90.
295. Graci, J.D. and C.E. Cameron, *Mechanisms of action of ribavirin against distinct viruses.* Rev Med Virol, 2006. **16**(1): p. 37-48.
296. Wu, W.C., et al., *The apoptotic effect of cordycepin on human OEC-M1 oral cancer cell line.* Cancer Chemother Pharmacol, 2007. **60**(1): p. 103-11.
297. Chen, Y.-H., et al., *Cordycepin enhances cisplatin apoptotic effect through caspase/MAPK pathways in human head and neck tumor cells.* OncoTargets and therapy, 2013. **6**: p. 983-998.
298. Hirschfeld, M., et al., *HNRNP G and HTRA2-BETA1 regulate estrogen receptor alpha expression with potential impact on endometrial cancer.* BMC Cancer, 2015. **15**: p. 86.
299. Uhlén, M., et al., *Tissue-based map of the human proteome.* Science, 2015. **347**(6220): p. 1260419.
300. Heinrich, B., et al., *Heterogeneous nuclear ribonucleoprotein G regulates splice site selection by binding to CC(A/C)-rich regions in pre-mRNA.* J Biol Chem, 2009. **284**(21): p. 14303-15.
301. Moursy, A., F.H. Allain, and A. Clery, *Characterization of the RNA recognition mode of hnRNP G extends its role in SMN2 splicing regulation.* Nucleic Acids Res, 2014. **42**(10): p. 6659-72.
302. Spriggs, C.C. and L.A. Laimins, *Human Papillomavirus and the DNA Damage Response: Exploiting Host Repair Pathways for Viral Replication.* Viruses, 2017. **9**(8): p. 232.

Printed by Media-Tryck, Lund 2019 NORDIC SWAN ECOLABEL 3041 0903

FACULTY OF MEDICINE

Faculty of Medicine
Department of Laboratory Medicine

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2019:67
ISBN 978-91-7619-796-7
ISSN 1652-8220