UNIVERSITY OF TARTU
Faculty of Science and Technology
Institute of Technology

Sofiya Babok

Characterization of the high-risk human papillomavirus 18 genome replication and transcription in different cellular models

Bachelor's Thesis (12 ECTS)

Curriculum Science and Technology

Supervisor(s):
Assoc. Prof., PhD. Marko Piirsoo
Assoc. Prof., PhD. Alla Piirsoo

Tartu, 2023

Characterization of the high-risk human papillomavirus 18 genome replication and transcription in different cellular models

Abstract:

Human papillomavirus (HPV) is responsible for approximately 30% of infection-caused cancers worldwide. The development of antiviral drugs requires a thorough understanding of the viral life cycle and its interactions with host cells. Replication of HPV genomes is restricted to a very limited number of human cell lines, and the reasons for this limitation are unknown. Human primary keratinocytes and osteosarcoma U2OS cells are widely used cellular model systems in HPV-related research. Conversely, it has been established that viral proteins, namely helicase E1 and transcription factor E2, can trigger the replication from the origin of replication if expressed from separate plasmids in almost all other cell types studied so far. This suggests that a deficiency in at least one of these viral factors may be a reason for the lack of replication of the viral genome in most cell lines. To test this hypothesis, the transcription, translation, and replication of the oncogenic HPV type 18 genome were studied in two different types of cells: embryonic kidney cells 293FT, which are not permissive for HPV replication, and well-described HPV host U2OS cells. The results obtained in the present thesis show that expression levels of all early viral genes, including E1 and E2, are very low in 293FT cells, and the absence of the E1 and E2 proteins causes the lack of viral replication. Additionally, it has been shown that in epithelial cells, such as keratinocytes HaCaT and squamous cervical cancer C33A cells, replication of HPV18 may be restored in the absence of the E8^E2 viral transcriptional repressor, which is expressed from an alternative viral promoter and counteracts E2 activities. However, this effect was not observed in 293FT cells and only exogenously added E1 protein was able to trigger replication of the HPV18 genome in these cells. These results suggest that different cell type-specific mechanisms may be employed to restrict replication of the HPV genome.

Keywords: replication, inhibition, transcription, human papillomavirus (HPV).

CERCS: B230 Microbiology, bacteriology, virology, mycology.

Kõrge riskiga inimese papilloomiviiruse 18 genoomi replikatsiooni ja transkriptsiooni uurimine erinevates rakumudelites

Lühikokkuvõte:

Inimese papilloomiviirus (HPV) vastutab ligikaudu 30% maailmas esinevate nakkusest põhjustatud vähkkasvajate eest. Viirusevastaste ravimite arendamine nõuab põhjalikku arusaamist viiruse elutsüklist ja selle seotusest peremeesrakkudega. HPV genoomid on võimelised paljunema väga piiratud hulgas inimese rakutüüpides ning selle põhjused on teadmata. Inimese primaarseid keratinotsüüte ja osteosarkoomi U2OS rakke kasutatakse laialdaselt HPV-ga seotud uurimistes rakumudelitena. Samas on näidatud, et kolm viirusefaktorit, nimelt helikaas E1, transkriptsioonifaktor E2 ja replikatsiooni alguspunkt, on vajalikud ja piisavad HPV replikatsiooni käivitamiseks kõigis uuritud raku tüüpides, kui need faktorid via rakkudesse eraldi plasmiididena. See viitab sellele, et vähemalt ühe nendest viirusefaktoritest puudumine võib olla viiruse genoomi replikatsiooni puudumise põhjuseks. Selle hüpoteesi kontrollimiseks uuriti onkogeense HPV tüübi 18 genoomi transkriptsiooni, translatsiooni ja replikatsiooni kahte erinevat tüüpi rakkudes: embrüonaalses neerurakuliinis 293FT, mis ei võimalda HPV replikatsiooni, ning hästi kirjeldatud HPV peremeesrakkudes U2OS. Käesolevas väitekirjas saadud tulemused näitavad, et kõigi viiruse varajaste geenide, sealhulgas E1 ja E2, ekspressioonitase on väga madal 293FT rakkudes ning E1 ja E2 valkude puudumine põhjustab viiruse replikatsiooni puudumist. Lisaks on näidatud, et epiteelirakkudes, näiteks keratinotsüütides HaCaT ja lamerakulises emakakaelavähis C33A, võib HPV18 replikatsioon taastuda ilma E8^E2 viirusliku transkriptsioonirepressori olemasoluta, mis avaldub alternatiivsest viiruslikust promootorist ja mõjutab E2 aktiivsust. Siiski ei täheldatud seda efekti 293FT rakkudes ning ainult väliselt lisatud E1 valk suutis käivitada HPV18 genoomi replikatsiooni nendes rakkudes. Need tulemused viitavad sellele, et erinevad rakutüübi spetsiifilised mehhanismid võivad olla kasutusel HPV genoomi replikatsiooni piiramiseks.

Võtmesõnad: replikatsioon, pärssimine, transkriptsioon, inimese papillomaviirus (HPV).

CERCS: B230 Mikrobioloogia, bakterioloogia, viroloogia, mükoloogia.

TABLE OF CONTENTS

TERMS, ABBREVIATIONS AND NOTATIONS

cDNA - complementary DNA

DBD - DNA binding domain

dNTPs - deoxynucleoside triphosphates

DTT - dithiothreitol

EDTA - ethylenediaminetetraacetic acid

Flag-tag - peptide protein tag with DYKDDDDK sequence

GAPDH - glyceraldehyde 3-phosphate dehydrogenase

HA-tag - human influenza hemagglutinin peptide protein tag

Hepes - 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HPV - human papillomavirus

HR - high-risk

LR - low-risk

miRNAs - microRNA

ND10 - Nuclear Domain 10

NES - nuclear export signals

NLS - nuclear localization signals

ORF - open reading frame

p53 - tumor suppressor protein 53

PBS - phosphate-buffered saline

PVDF - polyvinylidene fluoride

PVs - papillomaviruses

RT - room temperature or reverse transcriptase

SDS - sodium dodecyl sulphate

SSC - saline-sodium citrate

TAE - tris-acetate-EDTA

URR - upstream regulatory region

wt - wild-type

INTRODUCTION

Papillomaviruses are DNA viruses that can cause various types of epithelial malignancies in humans, including several types of cancers. Human papillomavirus (HPV) is the most common sexually transmitted virus that is responsible for the majority of anogenital and oropharyngeal cancers worldwide, being the leading cause of cervical cancer. HPV-related cancers are primarily caused by HPV16 and HPV18 types. While HPV vaccination serves as a highly effective preventive measure against HPV infections, there is currently no cure for ongoing HPV infections. Therefore, the development of therapeutic agents for the treatment of established HPV infections remains a critical area of investigation.

A major challenge in HPV-related research is the shortage of cells capable of supporting HPV replication that can be used as suitable cellular models for studying the viral life cycle. Along with human primary keratinocytes, U2OS osteosarcoma cells have been described as one of a few cell lines permissive to the replication of different HPV types. In all other cells studied so far, HPV replication can be triggered by using simultaneously three exogenous viral factors, namely proteins E1 and E2, and the origin of replication.

In this thesis, I aimed to investigate the lifecycle of the oncogenic HPV type 18 in two different cell types, namely human embryonic kidney 293FT cells and U2OS cells, with the further goal of revealing potential factors that may restrict or support HPV18 replication in different cells. HPV18 as the second most prevalent HPV type associated with cancer was chosen due to its known oncogenic potential and well-described mode of replication in U2OS cells. The widely used 293FT cell line was chosen because of its ease of handling and high transfection efficiency, which makes it a desirable model cell line. I examined the viral genome functioning on three molecular levels: replication, transcription, and translation. I focused on studying the major regulators of viral replication and transcription, E1 and E2 proteins, along with the repressor E8^E2 protein. This study was performed in the molecular virology research group, Institute of Technology, University of Tartu.

1 LITERATURE REVIEW

1.1 General description of human papillomaviruses

Papillomaviruses (PVs) are small double-stranded DNA viruses that infect mucosal or cutaneous keratinocytes of stratified squamous epithelia. PVs can infect various vertebrates: birds, reptiles, marsupials, and other mammals (Jara & Escobar, 2019). There are over 450 HPV genotypes identified so far (Burley *et al.*, 2020). There is no evidence that PVs can be transmitted between different species, with the exception of bovine PV (Handisurya *et al.*, 2009). Some HPVs cause benign epithelial lesions (warts, papillomas), while others give rise to lesions that can progress to cancer (Handisurya *et al.*, 2009). Most HPV infections are asymptomatic and go unnoticed unless the immune system is compromised.

HPVs can be grouped into high- and low-risk (HR and LR, respectively) types based on their oncogenicity. Five HPV genera have been identified based on nucleotide sequence comparison: alpha, beta, mu, nu, and gamma (**Figure 1**). The major viral capsid protein gene L1 is used for HPV classification, distinct types of HPV must be at least 10% divergent from other types in their L1 sequence (Bernard *et al.*, 2010). Alpha papillomaviruses can be divided into LR cutaneous, LR mucosal, and HR (light-brown, yellow, and pink on Figure 1). The alpha genus is the most thoroughly studied due to its association with clinically relevant diseases. HR mucosal alpha HPVs cause 4.5% of cancers worldwide (de Martel *et al.*, 2017; Alhamlan *et al.*, 2021).

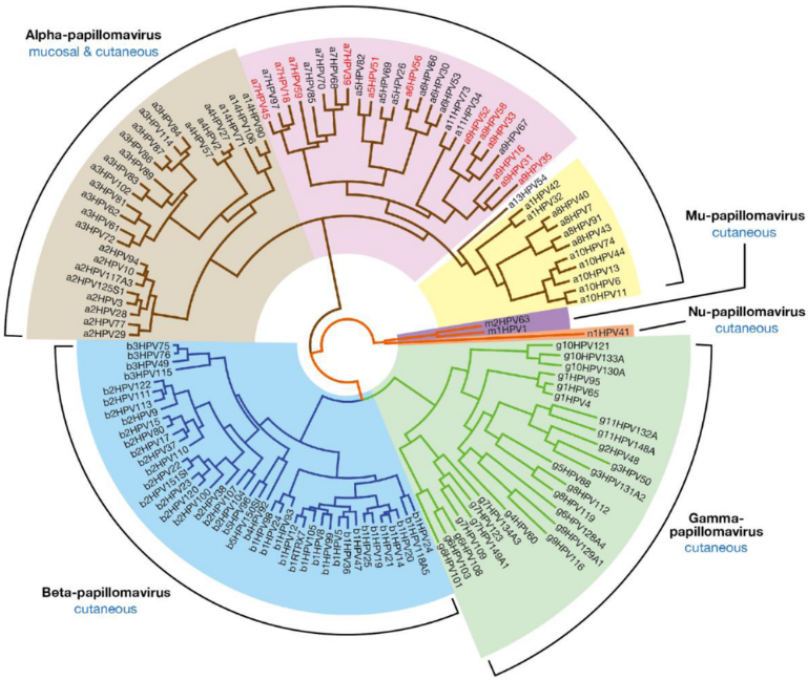

Figure 1. Phylogenetic tree of Human Papillomaviruses (Egawa *et al.*, 2015)

1.2 HPV life cycle

The HPV life cycle is keratinocyte differentiation-dependent and consists of five stages: HPV entry, initial amplification of the viral genome, persistent infection of the basal epithelium, productive infection in the upper layers of the stratified epithelium, and release of the virus (**Figure 2**).

The virus accesses the basal keratinocytes of the stratified epithelium through microabrasions (Handisurya *et al.*, 2009; Belnap *et al.*, 1996). The viral capsid consists of L1 and L2 viral proteins. L1 protein binds the primary attachment receptors, heparan sulfate proteoglycans, on the cell surface and extracellular matrix of the epithelial basement membrane (Liu & Thorp, 2002; Kines *et al.*, 2009). This causes conformational changes in the capsid that result in the exposure of the N-terminus of the minor capsid protein L2, followed by its cleavage (Kines *et al.*, 2009). Then HPV virion binds to an uncharacterized secondary receptor and enters the host cell by endocytosis, resulting in the formation of HPV-containing endocytic vesicles (DiGiuseppe *et al.*, 2017; Schelhaas *et al.*, 2012). HPV is uncoated as it moves through the endosomal compartment, and only L2 capsid protein remains associated with the viral DNA (Bienkowska-Haba *et al.*, 2012; DiGiuseppe *et al.*, 2017). Using the Golgi network, the virus is delivered to the nuclear membrane, where it waits for cell division to occur (Aydin *et al.*, 2014). During cell division, the host cell nuclear membrane breaks down, and the L2-genome complex gains access to the nucleus (DiGiuseppe *et al.*, 2017, Aydin *et al.*, 2014). There is evidence that the L2-genome complex is transported to the nucleus in a vesicle and binds to cellular condensed host chromosomes, and remains there until mitosis is finished (DiGiuseppe *et al.*, 2017).

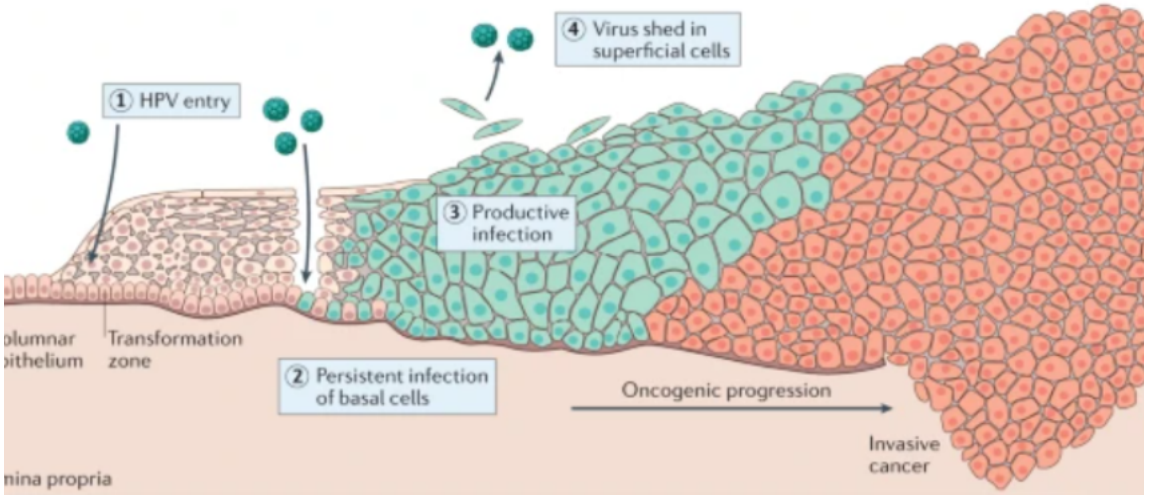

Figure 2. Infectious cycle of high-risk alpha HPV in cervical tissue (McBride, 2021)

The L2-genome complex is protected within these vesicles. After the completion of mitosis, the L2-genome complex becomes localized next to Nuclear Domain 10 (ND10), nuclear structures distributed throughout the nucleoplasm (Day *et al.*, 2004). These bodies are considered important for antiviral defence, but many viruses localize there during the early stages of infection, disrupting the bodies or some of their components. L2 protein reorganizes proteins within ND10, making the viral genome suitable for the establishment of transcription and replication of viral DNA (Florin *et al.*, 2002; Becker *et al.*, 2003). HPV uses cellular factors to initiate the synthesis of early viral transcripts since it lacks transcriptional regulatory proteins inside the virion (Ozbun, 2002). Early viral transcripts encode E1 and E2 proteins that are essential for the initiation of replication. During the initial amplification, the viral genome copy number reaches around 200 copies per cell (McBride, 2008).

After that, the stage of persistent infection starts, and the viral genome is replicated in concert with the host genome and is maintained at a constant copy number in the dividing basal cells of the epithelium (McBride, 2008). Viral genome replication occurs only once per cell cycle (Liblekas *et al.*, 2021). During persistent infection, HPV limits cellular DNA damage response caused by unscheduled DNA synthesis by limiting the presence of viral replication factors E1 and E2 in the nucleus. The E1 protein is transported to the cytoplasm when it is not used for replication (Deng *et al.*, 2004). The activity of the E2 protein is limited by the expression of one of its isoforms, E8^E2, which represses viral DNA replication (Dreer *et al.*, 2017). Host cellular factors such as p53 and cellular miRNAs can also limit HPV replication by targeting E1 and E2 coding regions (Porter *et al.*, 2017).

Productive infection begins as the infected keratinocytes undergo differentiation. The viral genome is amplified at much higher levels, and late genes L1 and L2 are expressed (Conway & Meyers, 2009). Capsid proteins production is essential for future virion capsid assembly. This stage is also characterized by increased expression of E1 and E2 proteins (Graham, 2017; Mac & Moody, 2020). This transition relies on viral E1 and E7 proteins. They promote DNA damage response which leads to the recruitment of cellular factors that can be hijacked by the virus to allow viral DNA replication outside of the S-phase of cell cycle (Gautam & Moody, 2016). Once the DNA is packaged inside the capsid, viruses are released from the surface of the epithelium during shredding (McBride, 2021). This completes the HPV life cycle.

If infected cells are not cleared by the immune system, the infection can become chronic and eventually undergoes oncogenic progression (Handisurya *et al.*, 2008). Cells are unable to undergo differentiation and do not produce viruses, and in a small number of cases, they can progress into invasive cancer (McBride, 2021). It has been shown that on the molecular level,

increased expression of E6 and E7 viral oncoproteins is responsible for the cancer progression. These proteins re-activate the cell cycle and proliferation of the differentiating cells while deregulating cellular proliferation and apoptotic machinery, which results in the accumulation of DNA damage (Mac & Moody, 2020, Graham, 2017).

1.3 Medical importance of HPV

HPV infects stratified squamous epithelial cells of skin and mucosa (Hadisurya *et al.*, 2008). Sites of infection of HPVs include skin, oropharynx, larynx, oral cavity, and anogenital tract: cervix, vulva, vagina, penis, anus (Alhamlan *et al.*, 2021; Li & Xu, 2017). The typical mode of transmission for the virus involves direct contact with the infected skin or mucosal membranes with friction. Additionally, there is evidence that cutaneous HPVs can be indirectly transmitted via contact with surfaces that have been contaminated (Hadisurya *et al.*, 2008). Diseases caused by HPVs vary from skin and genital warts to neoplasias and cervical cancer. Warts and papillomas are dome-shaped skin papules that can appear on different body parts, they are usually caused by LR HPVs (Cubie, 2013; Hadisurya *et al.*, 2008).

Intraepithelial neoplasia, head, neck, and cervical cancers are associated with HR HPVs (Cubie, 2013). HPV16 and HPV18 types, both alpha HR PV, are responsible for the majority of HPV-related cancers (Hadisurya *et al.*, 2008; Egawa *et al.*, 2015). HPV is linked to over 90% of cases of anal and cervical cancers, approximately 70% of vaginal and vulvar cancers, around 70% of oropharyngeal cancers, more than 60% of penile cancers, and over 10% of cancers in the oral cavity (Arbyn *et al.,* 2012).

Cervical cancer is the fourth most common cancer in women and the most common cause of cancer-related death in women under the age of 35 (Arbyn *et al.*, 2020). HPV16 causes approximately 61% of cervical cancers, HPV18 causes approximately 10%, and the remainder is caused by other high-risk HPV (Schiffman & Kjaer, 2003). Regular cervical cancer screenings have resulted in a more than 50% decline in cervical cancer incidence since 1999 (Islami *et al.*, 2019). Similar to the cervix, the anal cavity epithelium is highly susceptible to HPV infection that can result in anal intraepithelial neoplasia and anal carcinoma (Lin *et al*., 2018; Li & Xu, 2017). In recent years, the number of instances of anal cancer have been increasing (Li & Xu, 2017). The same is true for oropharyngeal cancers associated with oral HPV infections, although the prevalence of HPV-related head and neck cancer is still comparatively lower than that of the genital tract (Li & Xu, 2017). The prognosis for the HPV-related oropharyngeal cancer is generally better than for non-HPV-related oropharyngeal

cancer, patients tend to be younger and have a better response to treatment (Leemans *et al.*, 2011).

Although there is no direct cure for HPV, several HPV vaccines are available. They are capable of preventing infection and related dysplasia associated with the specific types of HPV covered by the vaccine, with an efficacy rate of up to 100% (Hadisurya *et al.*, 2008). The number of targeted HPV types varies from 2 to 9. All HPV vaccines protect against HPV16 and HPV18, since these types are the most dangerous (Li & Xu, 2017). A significant limitation of existing vaccines is their high cost. This issue is particularly important in developing countries that lack the necessary resources for efficient screening programs, where about 80% of all cervical carcinomas arise (Hadisurya *et al.*, 2008).

1.4 HPV genome and proteins

The typical HPV genome size is approximately 8 kb in length. A schematic circular view of HPV18 genome used in this study is presented in **Figure 3**. HPV genome can be divided into three functional sections: early region, late region, and upstream regulatory region (URR).

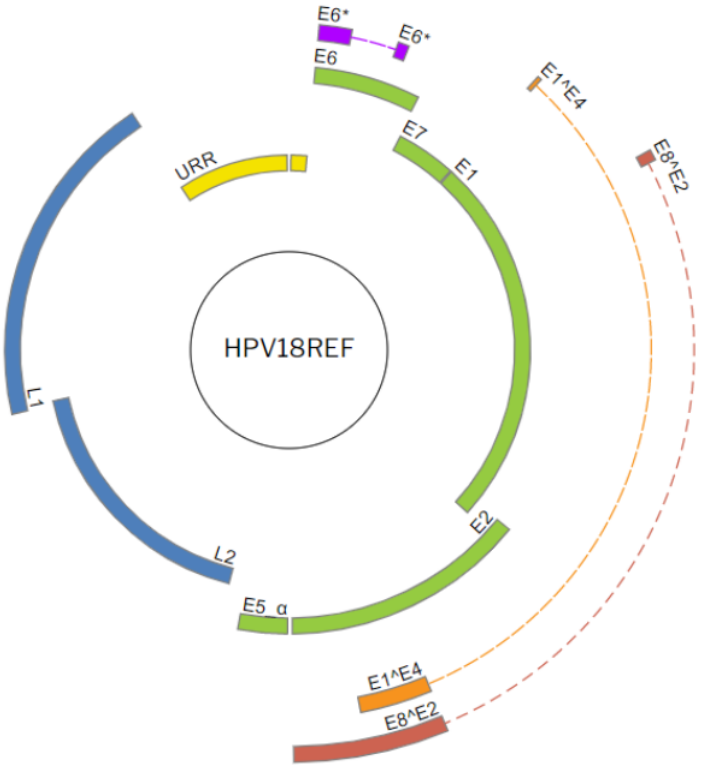

Figure 3. HPV18 genome organization. Early ORFs (E6, E7, E1, E2, E5) are shown in green, and late ORFs (L1, L2) are shown in blue, URR is colored yellow. Additionally, truncated protein isoforms are shown (E6*, E1^E4, E8^E2). Obtained from Papillomavirus Episteme database.

The early region produces at least seven viral proteins (translated from ORFs E1-E8) that regulate viral genome transcription (E2, E8/E2), replication (E1, E2, E8/E2), and host cell response (all proteins) of the infected epithelial cells, while the late region produces two viral structural capsid proteins: L1, L2 (McBride, 2021; Bergvall *et al*., 2013; Graham, 2010). Additionally, some HPV transcripts undergo alternative splicing, creating truncated versions of proteins (E1^E4, E8^E2, E6*). The main functions of HPV proteins are summarized in **Table 1.** URR region contains replication origin, transcriptional enhancers and promoters, and sequences required for genome maintenance (McBride, 2008). All viral RNAs are transcribed from one stand and undergo substantial alternative splicing (Graham, 2017).

Table 1. Main functions of HPV proteins (adopted from McBride, 2021)

Protein	Role in HPV infection
E1	initiation of viral replication and recruitment of cellular replication machinery
E2	regulation of viral transcription, support of replication, secure viral DNA next to host chromatin
E5	reduces immune detection, promotes cell proliferation and productive stages of infection
E6	prevents growth arrest and apoptosis, promotes genome replication and maintenance
E7	reduces immune detection, promotes cell proliferation and genome amplification, induces DNA damage response
E8^E2	represses viral transcription and replication to maintain low-level persistent infection
E1^E4	promotes viral genome amplification, disrupts and reorganizes keratin filaments to promote viral release, inhibits G2-to-M transition
L1	major capsid protein
L2	minor capsid protein, viral genome chaperone

Initiation of HPV replication requires E1 and E2 viral proteins, and replication origin. It contains several E2 binding sites, an E1 binding site, and a region rich in adenine and thymine also known as A/T-rich region (Ustav *et al.*, 1991). Typically, at least one E2 binding site is required for genome replication, E1 binding site can be absent and substituted by two E2

binding sites (Ustav *et al.*, 1991; McBride, 2017). Several E2 binding sites enhance the initiation of replication (McShan & Wilson, 1997).

1.5 The E1 protein

The E1 protein is a 70-kD ATP-dependent helicase that initiates viral replication by binding to the replication origin (McBride, 2008; Bergvall *et al.*, 2013, Clertant & Seif, 1984). E1 has four domains: N-terminal domain, DNA binding domain (DBD), helicase domain, and oligomerization domain (Bergvall *et al.*, 2013). The N-terminal domain, which contains approximately 200 amino acids, is important for intracellular localization; it contains both nuclear export signals (NES) and nuclear localization signals (NLS) (Bergvall *et al.*, 2013). The DBD domain binds to E1 binding sites in the replication origin (Auster & Joshua-Tor, 2004). The helicase domain has short duplex DNA unwinding activity, it is also able to bind DNA but with lower specificity (White *et al.*, 2001; McBride, 2008). E2 assists in loading the E1 helicase and increases the specificity of E1 DNA binding, E1 trimers and E2 dimers cooperatively bind to their neighbouring sites at the viral replication origin, the helicase domain of E1 interacting with the N-terminal domain of E2 (McBride, 2008; Sanders & Stenlund, 1998). At higher concentrations, E1 protein loading is less dependent on E2 protein (McBride, 2017).

Then E2 dissociates from the complex, E1 encircles the DNA as a double hexameric ring helicase (Sanders & Stenlund, 1998) During that change, the replication origin is melted, and single-stranded DNA is released through the hexamer ring. Additionally, E1 participates in recruitments of host cell replication machinery that is used to establish HPV replication (Bergvall *et al.*, 2013).

1.6 The E2 protein

E2 is a major transcriptional regulator that can act both as activator and repressor, its presence is required for viral DNA replication (McBride, 2008).

It consists of two domains that are connected with a hinge region: the N-terminal (transactivation) domain and the C-terminal (DNA binding) domain. The N-terminal domain is approximately 200 amino acids long, it is responsible for transcriptional activation, replication, and interaction with E1 (Antson *et al.*, 2000; McBride, 2013). The C-terminal domain is approximately 85-100 amino-acids. It is responsible for E2 dimerization and its association with DNA allows DNA binding and dimerization (McBride *et al.*, 1988). The E2 hinge region doesn’t have a stable structure and acts as a link between transactivation and DNA binding

domains (Gauthier *et al.*, 1991). Additionally, there is evidence that the hinge region can interact with host cell proteins (McBride, 2013)

1.7 E8^E2 protein

HPV genome encodes an E2-derived protein E8^E2 that represses both transcription and replication of the viral genome and helps maintain persistent infection (McBride, 2021; McBride, 2017). E8^E2 is a truncated version of the E2 protein that arises from splicing between E8 and E2 genes. It suppresses viral genome replication during persistent infection stage of the viral life cycle (Kurg *et al.*, 2010). E8^E2 interacts with cellular NCoR/SMRT corepressor complexes through E8 part of the protein, resulting in repression of viral transcription and replication (Dreer *et al.*, 2017). Additionally, it is also able to repress viral replication by counteracting with E1 and E2 proteins (McBride, 2008).

E8^E2 is expressed from a separate promoter, and the cellular factors controlling its expression are unknown (Dreer *et al.*, 2017). E8^E2 inhibits its own promoter, while E2 protein slightly activates it (Straub *et al.*, 2015). In various cell types including immortalized human keratinocytes, normal human keratinocytes, and the U2OS osteosarcoma cell line, HPV genomes deficient in E8^E2 expression replicate at significantly higher levels, ranging from 10 to 100-fold, compared to the respective wildtype genomes (Dreer *et al.*, 2017).

1.8 HPV18 replication in different cell lines

A small number of cells are known to support wild type HPV18 genome replication efficiently. These include U2OS osteosarcoma cells and primary keratinocytes. However, if E1 and E2 expression vectors are co-transfected with the plasmid containing the viral origin of replication, replication can be observed in many cell lines. Among these is the 293FT cell line, which has been widely used in HPV-related studies. 293FT cells were isolated from the embryonic human kidney of a fetus. These cells are fast-growing, and highly transfectable. HPV18 genome does not replicate in these cells, but the origin-dependent replication, if E1 and E2 are provided *in trans*, is very efficient (Chiang *et al.*, 1992).

2 THE AIMS OF THE THESIS

The general aim of this thesis is to characterize and compare the HPV18 genome lifecycle in 293FT and U2OS cells through replication, transcription, and translation analysis with further goal to reveal cellular factors restricting replication of HPV18 in 293FT cells. To accomplish this goal, a series of objectives were established:

- analysis of replication efficiency of the HPV18 genome in the presence of E1 or/and E2 overexpressed proteins in two different cell lines, U2OS and 293FT.
- analysis of different viral transcript levels in U2OS and 293FT cells.
- analysis of the E1 and E2 protein expression in U2OS and 293FT cells.
- analysis of replication efficiency of the wt and E8^E2-deficient HPV18 genomes in four different cell lines: 293FT, C33A, MCF7 and HaCaT.

3 EXPERIMENTAL PART

3.1 MATERIALS AND METHODS

3.1.1 DNA constructs

- HPV18 m.c. — HPV18 wild-type genome (Orav *et al.*, 2013).

- HPV18E1- m.c. — HPV18 genome with a point mutation in the first ATG of E1 open reading frame (ORF), lacks E1 expression (Orav *et al.*, 2013).

- HPV18E2- m.c. — HPV18 genome with a point mutation in the first ATG of E2 ORF, lacks E2 expression (Orav *et al.*, 2013).

- HPV18E8- m.c.— HPV18 genome with a point mutation in the first ATG into ACG of E8 viral protein, amino acids remain the same in overlapping E1 gene, lacks E8/E2 expression (Orav *et al.*, 2013).

- HPV18-E1HA-Nluc-E2Flag m.c.— HPV18 genome containing the HA tag encoding sequence in the frame of E1 and Flag tag encoding sequence in the frame of E2. Nluc encoding sequence is inserted in the frame of E2 after the E1 transcription termination codon (Piirsoo *et al.*, 2020).

- E1 e.v. — expression vector of HPV18 E1 protein (Kadaja *et al.*, 2007).

- E2-Flag c.o. e.v. — expression vector for the Flag-tagged E2 of HPV18, codon-optimized for human cells (obtained from the laboratory).

- E2-Flag e.v. — expression vector of HPV18 E2 viral protein that contains the Flag tag (Piirsoo *et al.*, 2020).

3.1.2 Cell lines

- Human osteosarcoma U2OS cell line
- Human embryonic kidney 293FT cell line
- Human breast adenocarcinoma MCF7 cell line
- Human cervical carcinoma C33A cell line
- Immortalized human keratinocytes HaCaT cell line

All cell lines used in the present study were purchased from American Type Culture Collection (ATCC).

3.1.3 Cell lines handling and transfection

293FT, MCF7, and HaCaT cells were grown in Dulbecco's Modified Eagle Media (Corning Inc.), U2OS and C33A cells were grown in Iscove's Modified Dulbecco's Media (Corning Inc.). Cell media were supplemented with 10% fetal calf serum (Corning Inc.) and 1% penicillin/streptomycin (Sigma-Aldrich). All cells were grown at 37 ℃ and 5% CO2.

All cell lines were transfected by electroporation (975 uF, and 200 V for 293FT, HaCaT, and MCF7, or 220 V for U2OS, or 180 V for C33A) using Gene Pulser XCell system (Bio-Rad Laboratories). Cells were transfected the next day after splitting. For splitting, cells were grown on 100 mm plates (Corning Inc.). First, the medium was removed. Cells were washed with 4 ml of Phosphate-Buffered Saline (PBS; 137 mM NaCl, 2.7 mM KCl, 8 mM Na_2HPO_4, and 2 mM KH_2PO_4), detached using 0.25% Trypsin-EDTA solution (Thermo Fisher Scientific Inc.), and plated onto new plates with corresponding growth media (10 ml of medium per 100 mm plate). For the transfection, the medium was removed, cells were washed with PBS, and detached with 2 ml of Trypsin-EDTA. Then cells were transferred into 6 ml of the corresponding medium and centrifuged at 1000 rpm for 5 min using Eppendorf Centrifuge 5810R (Thermo Fisher Scientific Inc.). The medium was removed, and cells were resuspended in 250 µl of medium per transfection.

The following amounts of DNA were used for transfection for southern blot experiments (per time point): H18 m.c - 500 ng, HPV18E8- - 500 ng, E1 e.v. - 40 ng, E2 e.v. - 100 ng.

The following amounts of DNA were used for transfection for reverse transcription-quantitative polymerase chain reaction (RT-qPCR) experiments: HPV18E1- - 1000 ng, HPV18E2- - 1000 ng, E2 c.o. e.v. - 375 ng.

The following amounts of DNA were used for transfection for western blot experiments: HPV18 m.c. - 2500 ng, HPV18E1HAE2Flag m.c. - 2500 ng, E2-Flag e.v - 2500 ng.

Also, 5 µg of salmon sperm DNA (ssDNA, used as carrier DNA) per transfection was added to the cells. The mixture of cells, viral DNA, and carrier DNA was resuspended with a pipette and transferred into an electroporation cuvette (4 mm gap size, Bio-Rad Laboratories). After the electroporation, cells were immediately transferred into growth medium and plated. Cells were grown for different periods of time depending on the experimental goal.

3.1.4 Total DNA extraction

Transfected cells were grown on 6 cm plates for 2, 3, or 4 days prior to total DNA extraction. Cells were washed with PBS and lysed in 500 µl of Sol IV buffer (20 mM Tris 7.5 pH, 100

mM NaCl, 10 µM EDTA, 0.2% SDS). Lysates were incubated for 5 min at room temperature (RT), transferred to a 1.5 ml microtube, and homogenized using 1 ml syringe with 26G needle to ensure the efficiency of proteinase treatment. Then samples were supplemented with 50 µg of Proteinase K and incubated at 56 °C overnight.

Next day, the lysates were mixed with 500 µl of a phenol-chloroform (1:1) mixture (Sigma Aldrich), vortexed for approximately 10 sec, and centrifuged at RT 13000 rpm for 3 min (Heraeus Biofuge Pico, Heraeus instruments). Approximately 450 µl of the upper aqueous phase containing DNA was transferred to a new 1.5 ml microtube, and 1 ml of 96% ethanol was added to each sample for DNA precipitation. Samples were vortexed for approximately 10 sec, incubated at -20 ℃ for 30 min, and then centrifuged at 4 ℃ 14800 rpm for 15 min (Fresco™ 21 Microcentrifuge, Thermo Fisher Scientific Inc.). Supernatant was discarded, 150 µl of ice-cold 70% ethanol was added to the pellet, followed by additional 10 mins of centrifugation. Then ethanol was removed, and the pellet was air-dried at RT and dissolved in 50 µl of TE buffer (MACHEREY-NAGEL) containing 1 µg of RNase A (Thermo Fisher Scientific Inc.). Samples were incubated at 37 ℃ for 1 hour.

DNA was then re-precipitated: first, 100 µl of 96% ethanol and 3 µl of 5 M NaCl were added, and samples were vortexed for approximately 10 sec, followed by 10 min centrifugation at 4 ℃ 14800 rpm. Supernatant was removed, and 150 µl of ice-cold 70% ethanol was added to the pellet, followed by additional 4 min of centrifugation. Then ethanol was removed, and the pellet was air-dried at RT and then dissolved in 50 µl of nuclease-free water. DNA concentrations were measured using a Nanodrop-1000 spectrophotometer (Thermo Fisher Scientific Inc.).

3.1.5 RNA extraction

Transfected cells were grown on 6 cm plates for 48 hours before total RNA extraction. Cells were washed with PBS. Total RNA was isolated using the Quick-RNA Miniprep kit (Zymo Research). Cells were lysed in 600 µl of RNA lysis buffer. After 5 min incubation at RT, cell lysate was transferred onto the Spin-Away column (yellow). The collection tube with the filter was centrifuged at 13000 rpm for 1 min to remove the majority of genomic DNA. Then 600 µl of 96% ethanol was added to the flowthrough, mixed with a pipette, and the mix was transferred into a Zymo-Spin IIICG column (green) in a collection tube. The tubes were centrifuged at 13000 rpm for 30 sec, flowthrough was discarded. Next, 400 µl of RNA Prep buffer was added to the column, followed by centrifugation at 13000 rpm for 30 sec, then 700 µl of RNA Wash buffer (ethanol was previously added according to the kit's instructions) was

added to the column, which was then again centrifuged at the same settings. Flowthrough was discarded, 400 µl of RNA Wash buffer was added, and the column was centrifuged again. After discarding the flowthrough, the column was centrifuged at 13000 rpm for 1 min to ensure that the Wash buffer was removed completely. Then the column was transferred into a labelled microtube, 50 µl of DNase/RNase-Free water was added to the column matrix, and microtubes were centrifuged at 13000 rpm for 1 min to elute RNA. RNA concentrations were measured using a Nanodrop-1000 spectrophotometer.

Approximately 8 to 9 µg of total RNA was treated with TurboDNase (Invitrogen) for 2 hours at 37 ℃. The reaction mix included RNA, 3 µl of TurboDNase, 5 µl of TurboDNase buffer, and DNase/RNase-Free water up to 50 µl. Then EDTA was added to each sample to a final concentration of 15 mM, and the mix was incubated at 75 ℃ for 10 min to inactivate DNase. RNA was precipitated by the addition of 32.5 µl of 7 M LiCl/15 mM EDTA and centrifugation at 4 ℃ 13000 rpm for 20 min. The supernatant was removed, and 100 µl of RNA-grade 70 % ethanol was added, followed by another 4 min of centrifugation. Ethanol was removed, the pellet was air-dried and dissolved in 11 µl of water.

3.1.6 Complementary DNA synthesis

Complementary DNA (cDNA) was synthesized using approximately 3 µg of total RNA and a RevertAid First Strand cDNA synthesis kit (Thermo Fisher Scientific Inc.). Water was added to the RNA up to 11 µl volume. Next, 1 µl of 100 µM oligo(dT) primer, that anneals to poly(A) mRNA sequences, was added, samples were incubated at 65 ℃ for 5 min, then spun down and placed on ice. For the cDNA synthesis, 4 µl of 5x reaction buffer, 2 µl of 10 mM deoxynucleoside triphosphates (dNTPs), 1 ul of Ribolock RNase Inhibitor enzyme (20 U/µl), and 1 ul of RevertAid Reverse Transcriptase (200 U/µl) were added. The mixture was incubated at 42 ℃ for 1 hour, then enzymes were inactivated at 70 ℃ for 5 min.

Additionally, Reverse Transcriptase minus (RT-) negative controls were made using the same procedure, but water was added instead of RevertAid Reverse Transcriptase. These controls allow to assess for contamination of the cDNA samples with genomic and viral input DNA.

After cDNA synthesis, a PCR was performed to confirm the efficiency of cDNA synthesis and DNase treatment.

3.1.7 Polymerase chain reaction

The PCR was performed using 5x HOT FIREPol Blend Master Mix RTL with 12.5 mM $MgCl_2$ (Solis Biodyne), the full reaction composition and PCR program are described in

Table 2 and **Table 3**. Applied Biosystems ProFlex PCR System was used to run the reactions. GAPDH and E2 genes oligonucleotide mixes from the laboratory collection were used for PCR reactions.

Table 2. PCR mix per reaction

Component	Volume (µl)
cDNA	1
Oligonucleotide mix 5µM	2
5×HOT FIREPol Blend Master Mix	2
Water	Up to 40

Table 3. PCR program

Operation	Temperature	Time	Number of cycles
Initial denaturation	95 ℃	12 min	1
Denaturation	95 ℃	15 sec	30
Annealing	56 ℃	30 sec	
Extension	72 ℃	1 min	
Final extension	72 ℃	10 min	1

After the reaction, samples were loaded onto 0.8% Agarose 1xTAE (40 mM Tris-acetate, 1 mM EDTA) gel containing 0.3 µg of ethidium bromide. The presence of PCR products was evaluated under a UV transilluminator.

3.1.8 Quantitative polymerase chain reaction (qPCR)

The qRT-PCR was performed using a 5×HOT FIREPol EvaGreen qPCR Mix Plus (ROX) (Solis Biodyne) and LightCycler 480 II (Roche Diagnostics) machine. The full reaction composition and qPCR program are described in **Table 4** and **Table 5**. Gene expression levels were measured in triplicates and adjusted by the levels of the housekeeping gene GAPDH. Before running the program, a sealed 384-well qPCR plate (Roche Diagnostics) was centrifuged at 1200 rpm for 5 min to spin down the samples.

Table 4. qPCR mix per reaction

Component	Volume (µl)
cDNA	1
Oligonucleotide mix 5µM	1
5×HOT FIREPol Blend Master Mix	2
Water	6

Table 5. qPCR program. Continuous acquisition mode for the third stage of melting curve operation means that fluorescence data is acquired continuously, and temperature increases at the rate of 0,11 ℃/s.

Operation		Temperature	Time	Number of cycles	Analysis mode
Preincubation		95 ℃	12 min	1	None
Amplification	Denaturation	95 ℃	15 sec	45	Quantification
	Annealing	59 ℃	15 sec		
	Extension	72 ℃	20 sec		
Melting curve		95 ℃	5 sec	1	Melting curve
		65 ℃	1 min		
		97 ℃	Continuous Acquisition (0,11 ℃/s)		
Cooling		40 ℃	30 sec	1	None

Absolute Quantification analysis using the Second Derivative Maximum method was performed on the data acquired from the qPCR. The cycle numbers at which fluorescent signal crosses the threshold (CT values) were obtained from the program and used for calculations. Calculations were performed in Microsoft Excel.

First, an average of triplicate values was calculated for all samples. Then viral gene CT values were adjusted by GAPDH values by subtraction of averages. U2OS samples (CT_{U2OS}) were used as a reference for 293FT cells (CT_{293FT}). Final values for 293FT cells were calculated as follows: $2^{(CT_{U2OS} - CT_{293FT})}$. Error bars were based on standard deviations, which were calculated using the following formula:

$s = \sqrt{\frac{\sum(x-\mu)^2}{n-1}}$, where μ is the mean value of a set of data, n is the total number of measurements.

3.1.9 Protein extraction and immunoprecipitation (IP)

Prior to protein extraction, transfected cells were grown on 100 mm plates for 48 or 72 hours, and washed with PBS.

For the E1 immunoblotting, cells were detached using Trypsin-EDTA and collected into a microtube, then centrifuged at 1000 rpm for 5 min (Sigma Zentrifugen 1-14 microfuge). Liquid was aspirated, cell pellet was resuspended in 1 ml of PBS and centrifuged again. Then cell pellet volume was estimated visually, and cells were resuspended in 5 pellet volumes of cold Whole Cell Extract buffer (20 mM Hepes-KOH pH 7.9, 400 mM KCl, 1 mM EDTA, 5% glycerol, 10 mM dithiothreitol (DTT), protease inhibitors mix to 1x). Samples were snap-frozen in liquid nitrogen and then quickly thawed three times. Next, samples were centrifuged at 14000 rpm 4 ℃ for 30 min (Fresco™ 21 Microcentrifuge, Thermo Fisher Scientific Inc.). Supernatant was used for E1 and GAPDH detection using western blotting.

For E2 detection, 1.5 ml of cold lysis buffer (⅓ RIPA buffer (50 mM Tris HCl, 150 mM NaCl, 1.0% NP-40, 0.5% Sodium Deoxycholate, 1.0 mM EDTA, 0.1% SDS, 0.01% sodium azide pH 7.4), ⅔ Lysis Buffer (50 mM Tris HCl pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100 and Protease inhibitor cocktail (Roche))) was added to each plate, and the plates were incubated on ice for 5 min. Then cell lysate was transferred into 15 ml tubes, 1.5 ml of Lysis buffer was added, followed by incubation at 4 ℃ for 1 h on slow end-to-end rotation. Next, samples were centrifuged at 10000 rpm 4 ℃ for 5 min in the Refrigerated Centrifuge 5810R (Eppendorf). 5 µl of supernatant was transferred into a microtube with 25 µl of PBS for GAPDH detection. The rest of supernatant was transferred into new tubes for immunoprecipitation. 100 µl of washed agarose beads conjugated with mouse-α-Flag antibody (Sigma-Aldrich) was added, and samples were incubated at 4 ℃ overnight at slow rotation. Next day, samples were centrifuged at 5000 rpm 4 ℃ for 1 min. Supernatant was removed, and the pellet was washed twice with ice-cold Lysis buffer and resuspended in approximately 10 µl of Lysis buffer. These samples were used for E2 detection.

3.1.10 Western blot

A loading buffer (3x Laemmli buffer containing 330 mM DDT) was added to the protein samples in a volume ratio of 1:3. The samples were denatured at 100 ℃ for 5 min, and 17 µl of each sample was loaded onto an acrylamide gel for SDS-PAGE. Gels containing 12% or 10% of acrylamide:bisacrylamide (1:37.5) (Sigma-Aldrich) were used to analyze the E2 and GAPDH proteins or the E1 protein, respectively. Gels recipe is presented in **Table 6**. First, separation gel was made and poured into a gel cassette (Bio-Rad Laboratories), 1 ml of 96% ethanol was carefully added over the gel to prevent bubble formation. Once separating gel hardened, ethanol was removed, then the stacking gel was poured, and 1 mm teeth were inserted.

Gels were run at constant mA (23 mA per a gel) for approximately 1.5 h. Then proteins were transferred onto PVDF membrane (Millipore) using wet electrotransfer method. PVDF membrane was wetted in methanol. The stacking gel was cut off. Transfer sandwich was assembled in the following order (from black to transparent supports): sponge layer, 2 pieces of whatman paper, gel, PVDF membrane, 2 pieces of whatman paper, sponge. 1 L of cold Transfer buffer (14.4 g/L glycine, 3.5 g/L Tris, 5% ethanol in water) was used for the transfer. The transfer was performed at 400 mA for 2 hours.

After the transfer, the membrane was blocked in 5% non-fat dry milk in PBS containing 0.1% Tween (PBS-T) at RT for 1 h. Then membrane was incubated with primary antibodies in a 2.5% milk PBS-T solution overnight at 4 °C. The following antibodies were used: α-GAPDH-HRP 1:8000 (Invitrogen), α-HA-HRP 1:2000 (Sigma-Aldrich), α-Flag-HRP 1:3000 (Sigma-Aldrich). Next day, the membrane was washed three times with PBS-T for 15 min. SuperSignal West Dura Extended Duration Substrate kit (Pierce) was used for chemiluminescent signal detection. Enhancer and Stable Peroxide Buffer were mixed 1:1. Approximately 1 ml of the mixture was evenly distributed over the membrane, and incubated for 1 min. Next, membrane was rinsed in PBS-T. Membrane was placed into a film cassette and, while in the dark room, X-ray film (Agfa-Gevaert N.V.) was placed over the membrane. Afterwards, the film was developed and fixed in Developer and Fixer solutions (Agfa-Gevaert N.V.). Images were processed using Adobe Photoshop software (Adobe Inc.).

Table 6. Solutions for preparing separation and stacking gels for Tris-glycine SDS-Polyacrylamide Gel Electrophoresis. Components are added in the same order as they are listed below.

Type of the gel	Component	Volume (ml)
10% separation gel (5 ml)	water	1.9
	30% acrylamide/bis solution, 37.5:1 (SERVA Electrophoresis GmbH)	1.7
	1.5 M Tris buffer (pH 8.8)	1.3
	10% SDS	0.05
	10% ammonium persulfate	0.05
	1,2-di-(dimethylamino)ethane (TEMED, Fisher Scientific International Inc.)	0.002
12% separation gel (5 ml)	water	1.6
	30% acrylamide/bis solution	2.0
	1.5 M Tris buffer (pH 8.8)	1.3
	10% SDS	0.05
	10% ammonium persulfate	0.05
	TEMED	0.002
Stacking gel (1 ml)	water	0.68
	30% acrylamide/bis solution	0.17
	1.0 M Tris buffer (pH 6.8)	0.13
	10% SDS	0.01
	10% ammonium persulfate	0.01
	TEMED	0.001

3.1.11 Southern blot

First, approximately 5-7 µg of total DNA per sample was digested with FastDigest BglI enzyme to linearize the genome DNA and DpnI enzyme to digest bacterially methylated input DNA (Thermo Fisher Scientific Inc.). The reaction mix composition is described in **Table 7**. Samples were incubated at 37 °C for 3 h.

Table 7. Restriction reaction mix per sample

Component	Amount/Volume
DNA	5 - 7 µg
10x FastDigest Green Buffer (Thermo Fisher Scientific Inc.)	2 µl
BglI restriction enzyme	2 µl
DpnI restriction enzyme	1 µl
Water	Up to 20 µl

After restriction, gel electrophoresis was performed. Samples were loaded onto 0.8% agarose gel together with a GeneRuler 1 kb DNA Ladder or GeneRuler 1 kb plus DNA Ladder (Thermo Fisher Scientific Inc.).

DNA was visualized under the UV transilluminator. Then, the gel was incubated in Solution A (0.5 M NaOH, 1.5 M NaCl) at RT for 60 min, followed by incubation in Solution B (1 M Tris pH 8.0, 1.5 M NaCl) for 40 min. After washing, DNA was transferred onto a nylon membrane (GVS) in 10x saline-sodium citrate buffer (1.5 M NaCl, 150 mM $Na_3C_6H_5O_7$) for 16 h. The transfer sandwich was set up as depicted in **Figure 4** (Karcher, 1991).

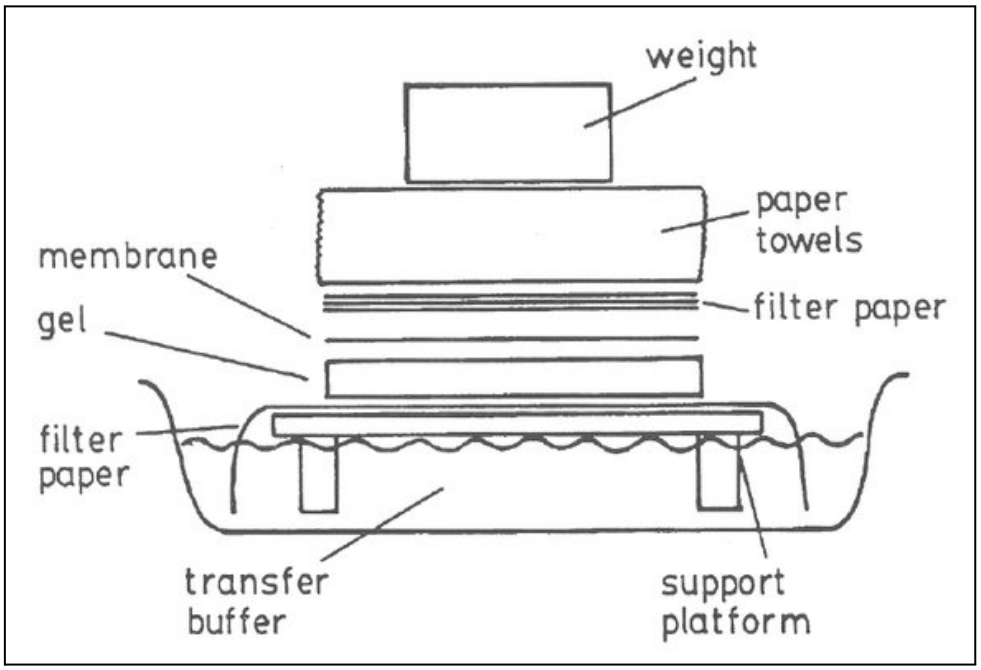

Figure 4. Schematic of southern blot transfer assembly

After the transfer, the membrane was air-dried. DNA was crosslinked using Stratalinker UV crosslinker 1800 (Stratagene) for approximately 40 sec.

To prevent non-specific binding and reduce background, the membrane was incubated with warm pre-hybridization mix (**Table 8**) at 67 ℃ for approximately 1 h.

The hybridization probe was made by mixing Random priming premix containing random hexamer oligonucleotides with 150 ng of linearized HPV18 m.c. DNA, and water up to 40 µl. This mix was incubated at 100 °C for 5 min, then chilled on ice and spun down. Then, 3 µl of dNTPs mix without dATP, 4 µl of α-P33-dATP isotope (Hartmann Group), and 1 µl of Exo-Klenow DNA polymerase (Thermo Fisher Scientific Inc.) were added, followed by 15 min incubation at 37 ℃. Finally, 4 µl of dNTP mix was added, and the mixture was incubated for 15 min at 37 °C. Then the probe was denatured at 100 ℃ for 5 min and added to the prehybridization solution.

The membrane was incubated overnight on rotation in a roller bottle at 67 ℃ in the hybridization oven (Bio-Techne). Next day, the membrane was washed with preheated washing solutions at 67 ℃: Solution I (1x SSC, 0.1% SDS) twice for 5 min, then Solution II (0.5x SSC, 0.1% SDS) once for 15 min, then Solution III (0.1x SSC, 0.1% SDS) twice for 10 min. Then the membrane was dried, covered in saran wrap, and exposed to the screen in a film cassette. Next day, the radioactive signals were visualized using Amersham Typhoon Biomolecular Imager (GE Healthcare) and scanned using a phosphor imaging plate to detect ^{33}P radioisotopes. The images were processed using Adobe Photoshop software (Adobe Inc.).

Table 8. Prehybridization mix: salmon sperm DNA was denatured at 100 ℃ for 10 min and then added to the rest of warm components, the mix was heated up to 65 ℃.

Component	**Volume (ml)**
20x SSC	9
50x Denhardt's Solution (1% Bovine serum albumin, 1% Ficoll, 1% Polyvinylpyrrolidone)	3
10% Sodium Dodecyl Sulfate (SDS)	1.5
Salmon Sperm DNA (10 mg/ml)	0.6
Water	15.9

3.2 RESULTS

3.2.1 Replication of HPV18 genome in 293FT cells is restored by co-transfection with E1 expression vector

It has been previously established that the HPV18 genome and other HPV types do not replicate in 293FT cells (Geimanen *et al.*, 2011). For many cell lines, including 293FT, co-transfection of the E1 and E2 expression vectors with the plasmid containing viral replication origin results in the restoration of viral replication (Chiang *et al.*, 1992). A generally accepted model suggests that both E1 and E2 proteins are required for viral genome replication. First, I investigated if overexpression of E1 and/or E2 proteins could restore the replication of the HPV18 genome in 293FT cells.

For this experiment, 293FT cells were transfected through electroporation with the following DNA combinations:

- HPV18 m.c.
- HPV18 m.c. and E1 e.v. (+ E1)
- HPV18 m.c. and E2 e.v. (+ E2)
- HPV18 m.c., E1 e.v., and E2 e.v. (+E1/E2)

Additionally, the HPV18E8- genome was transfected. This genome is deficient in the expression of the E8^E2 repressor protein. It has been described that a lack of the E8^E2 repressor potentiates replication efficiency of the viral genome up to 100 times in U2OS cells (Geimanen *et al.*, 2011). Therefore, elevated replication efficiency of the HPV18E8- genome was expected in 293FT cells.

Total DNA was extracted 48 h and 72 h after transfection. It was treated with BglI and DpnI restriction enzymes in order to digest input DNA used for transfection and linearise the viral genome, respectively. This experiment was repeated three times in order to confirm the results.

After hybridization with radioactively labelled HPV18 genome, no replication of the wt or HPV18E8- genomes was detected in 293FT cells. Bands running at around 8 kb were observed for co-transfections with E1 e.v. and both E1 and E2 e.v. (**Figure 5A**). The size of these bands corresponds to the linearized HPV18 genome marker loaded on the right side of the gel. These bands represent a linearised HPV18 genome (7.85 kb) that has replicated in the 293FT cells at least once. It is visible that the replication signal increased from 48 h to 72 h, while the amount of DpnI-sensitive DNA decreased, which in turn indicates the efficient

proliferation of the transfected cells. In order to control if the HPV18 genome used in these experiments is functional, we transfected it into U2OS cells and showed its capability to replicate efficiently (**Figure 5B**). This result indicates that replication of HPV18 genome in 293FT cells is restored in the presence of the overexpressed E1 protein.

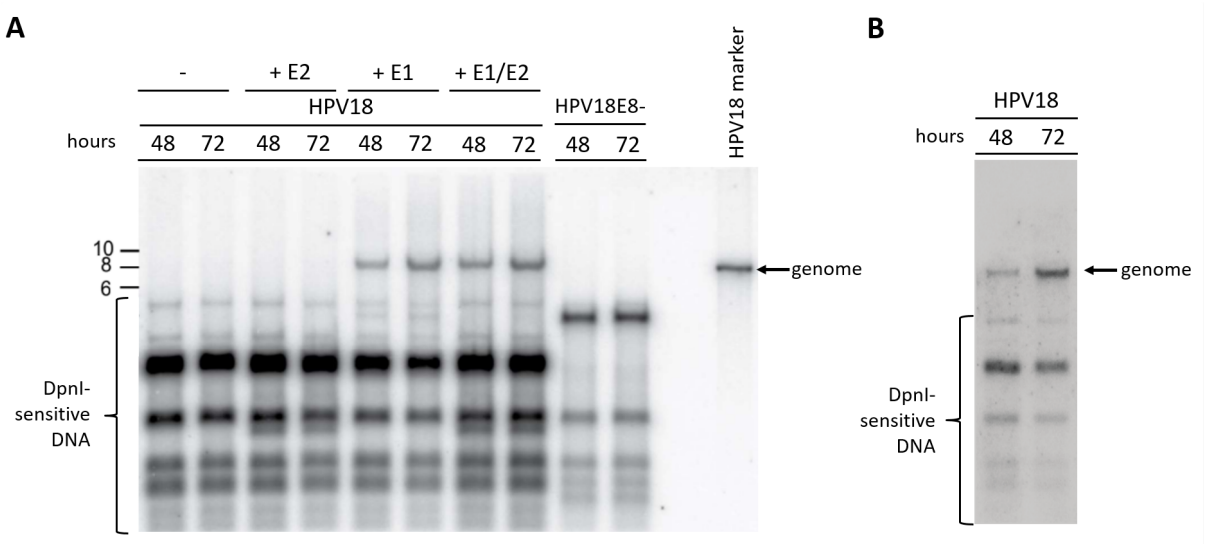

Figure 5. Replication of the HPV18 genome in 293FT cells is restored by the overexpressed E1 protein. HPV18 genome replication in 293FT and U2OS cells was evaluated through the Southern Blot 48 h and 72 h post-transfection. (A) 293FT cells were transfected with the HPV18 genome (-), with the addition of E1 e.v. (+E1), with the addition of E2 e.v. (+E2), or with the addition of both E1 and E2 (+E1/E2). The second part of this experiment included HPV18E8- genome transfection. 500 pg of linearized HPV18 genome was used as a marker. DNA ladder markings on the left side of the gel correspond to 10, 8, and 6 kb. (B) U2OS cells were transfected with the HPV18 genome to confirm its functionality. The signal corresponding to the replicated viral DNA increased from 48 to 72 h.

3.2.2 The HPV18 E1 and E2 proteins are not expressed from the transfected viral genomes in 293FT cells

E1 and E2 proteins are essential for the replication of HPV genomes, and E2 additionally functions as the main transcriptional regulator for the viral genome (McBride, 2021; McBride, 2008). As I established that the HPV18 genome is not able to replicate in 293FT cells (**Figure 5A**), I set out to analyze if E1 and E2 are expressed in these cells. I aimed to detect E1 and E2 proteins from U2OS and 293FT cells transfected with either the HPV18 genome or its derivative expressing the HA-tagged E1 and Flag-tagged E2 proteins. These modifications enable the detection of endogenous E1 and E2 proteins using specific commercially available anti-HA and anti-Flag antibodies. H18 m.c. served as a negative control, since the genome doesn't contain the Flag or HA tags that were targeted by the antibodies in this experiment.

For E2 detection, the following DNAs were used for transfections of both cell types:

- H18E1HANlucE2Flag m.c. (labelled as 1 in **Figure 6A**)
- HPV18 m.c. (labelled as 2 in **Figure 6A**)
- E2-Flag e.v. (labelled as 3 in **Figure 6A**)

E2-Flag e.v. was used as a positive control. On **Figure 6A**, the same gel developed at different exposure times is presented. The results of immunoblotting show that the endogenous Flag-tagged E2 is expressed from HPV18 genome in U2OS cells (**Figure 6A** longer exposure), while in 293FT cells, it was not detected.

For E1 detection, the transfections were performed with the following DNAs:

- HPV18 m.c.
- H18E1HANlucE2Flag m.c. (labeled as HPV18HA-E1 on **Figure 6B**)

The band of the correct size was present only for HPV18HA-E1 in U2OS, confirming that the E1 protein is not expressed in 293FT (**Figure 6B**). Both experiments were repeated twice to confirm the results. Additionally, the GAPDH protein was detected as a loading control in both experiments.

In conclusion, these results establish that the HPV18 genome is not able to replicate in 293FT cells because the essential viral replication proteins E1 and E2 are not expressed.

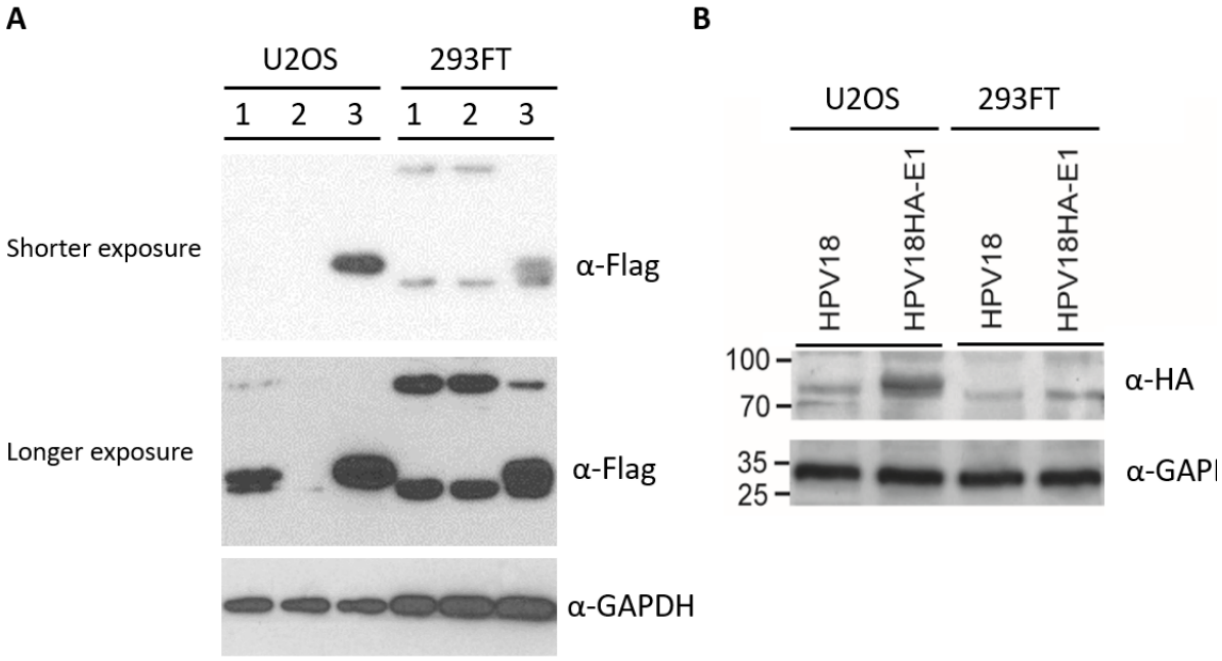

Figure 6. HPV18 E1 and E2 proteins are not expressed from the transfected viral genomes in 293FT cells. The expression of the E1 and E2 proteins was evaluated using Western Blot. Cells were

grown for 48 h prior to protein extraction. (A) E2 protein was immuno-precipitated from the cell lysate using agarose beads conjugated with α-Flag antibody, all of the immuno-precipitate was loaded on the gel. The supernatant was used for GAPDH detection. Labels 1, 2, and 3 correspond to the samples transfected with H18E1HANlucE2Flag m.c.; HPV18 m.c.; and E2-Flag e.v. respectively. Two exposure times are presented for better visualization of E2-Flag e.v. sample in 293FT cells. (B) E1 and GAPDH proteins were detected from the whole cell lysate using immunoblotting.

3.2.3 Levels of viral transcripts in 293FT cells are significantly lower than in U2OS cells

In the previous section, I established that E1 and E2 proteins are not detectable in transfected 293FT cells, as opposed to similarly transfected U2OS cells. This absence of the E1 and E2 proteins in 293FT cells could be attributed to either inefficiency of the viral genome transcription or translation or to the instability of E1 and E2 proteins in these cells. In order to analyze if the lack of E1 and E2 proteins in 293FT cells is the result of defective transcription of the viral genome, I performed RT-qPCR analysis of different viral transcripts in U2OS and 293FT cells. I used the replication-deficient HPV18 genomes HPV18E1- and HPV18E2- that lack expression of either E1 or E2 proteins, respectively. The use of these genomes enabled to account for the variation in viral genome copy numbers in the analyzed cells as U2OS cells support HPV18 replication and therefore exhibit significantly higher viral genome copy numbers. Also, these genomes enabled to estimate levels of the E2-dependent viral transcription.

For this experiment, cells were transfected with HPV18E1- or HPV18E2- genomes deficient in expression of E1 or E2, respectively. Primer sets for the following genes were used for RT-qPCR:

- E1
- E2
- E1^E4
- E8^E2
- E6

Before conducting RT-qPCR, samples underwent PCR analysis to ensure there was no contamination with genomic and viral input DNA. This is especially important, as E1, E2, E6 viral transcripts analyzed with RT-qPCR are unspliced, and therefore even minute genomic DNA contamination can result in misleading results. GAPDH and E2 primer sets were used for PCR. Samples that underwent cDNA synthesis in the absence of reverse transcriptase (RT)

were used as a negative control. The resulting PCR products were visualized on an agarose gel, and the presence of product bands was observed only in samples where RT was added (**Figure 7**).

For HPV18E1- genome, mRNA expression levels of all genes were 5 to 100-fold lower in 293FT cells, compared to U2OS cells (**Figure 8A**). The difference was the smallest for E8^E2 mRNAs. Based on the error bars, it was concluded that the differences observed in viral gene expression levels in 293FT and U2OS cells were significant. Similar differences were observed in the case of the HPV18E2- genome **(Figure 8B**).

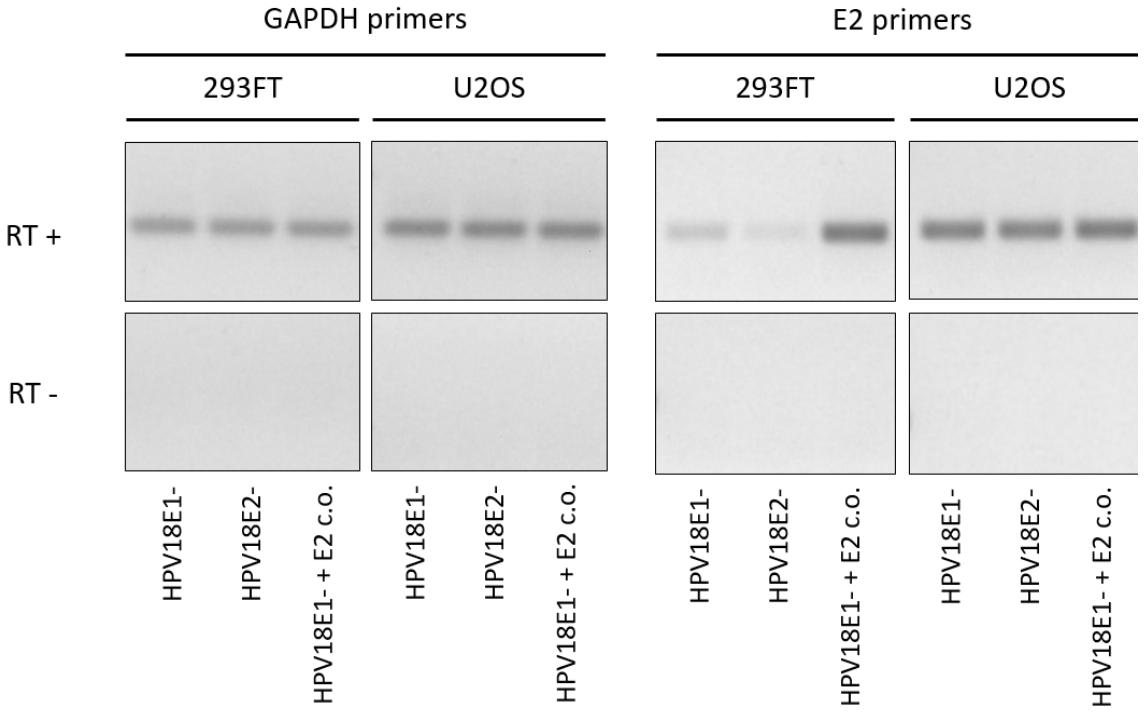

Figure 7. Synthesized cDNAs are not contaminated with genomic DNA. PCR reaction was performed with GAPDH and E2 primers to assess contamination of the samples with genomic and viral input DNA. PCR products were loaded onto 0.8% agarose-TAE gel and visualized under UV light. The lanes labelled as RT+ correspond to samples where reverse transcriptase was added during the cDNA synthesis, while the RT- lanes indicate samples where reverse transcriptase was absent. No visible bands are observed in the RT- lanes, confirming the purity of the samples.

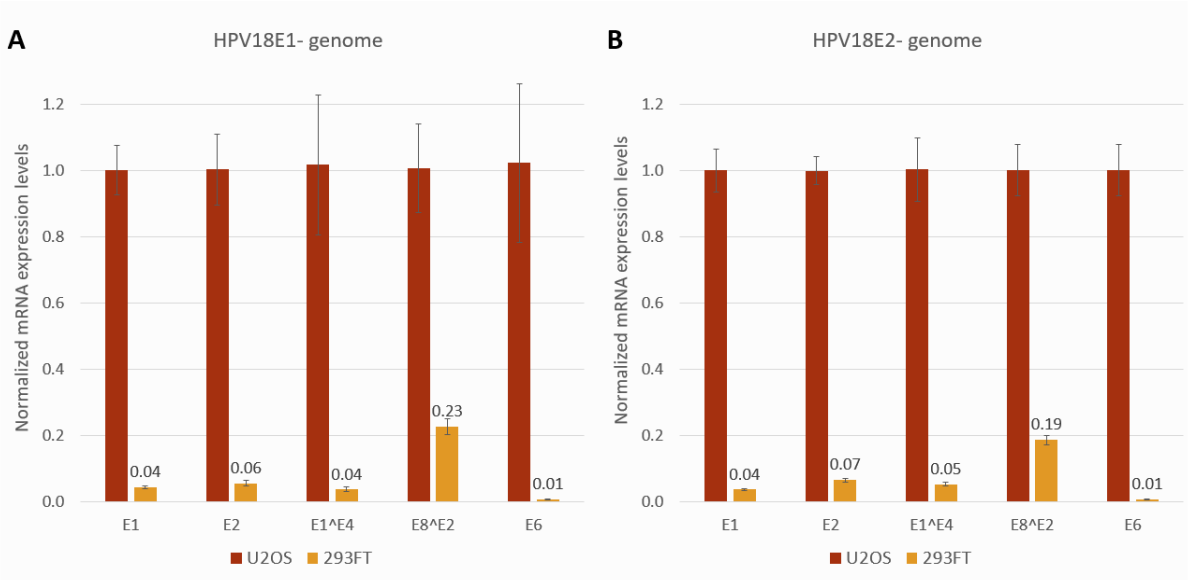

Figure 8. Levels of viral transcripts in 293FT cells are 5 to 100-fold lower than in U2OS cells. Cells were collected at 48 h after transfection. Transcription levels of HPV18E1- and HPV18E2-genomes were evaluated through the RT-qPCR method. The mRNA expression levels were normalized by GAPDH expression levels. Expression levels of the viral genes in U2OS cells were taken as 1 for easier comparison between cell lines. The data is presented as an average mean of two independent experiments +/- standard deviation.

This set of experiments establishes that HPV18 genomes are not able to replicate in 293FT cells because of the defect in the viral genome transcription.

3.2.4 E2 overexpression results in higher transcription levels of E2 and E8^E2 in 293FT cells than in U2OS cells

In the viral transcript analysis described in the previous section, there was one transcript that exhibited significantly higher levels than others in 293FT cells. This transcript codes for the viral E8^E2 protein that overlaps with the E2 protein in the DNA binding and dimerization domain and hinge region but lacks the N-terminal transactivation domain. E8^E2 protein inhibits transcription and replication of the viral genome (McBride, 2021; McBride, 2017). Basically, it counteracts the transcriptional activator function of the E2 protein by binding competitively to the same site in DNA (Lace *et al.*, 2008). Therefore, I studied if the overexpressed codon-optimized E2 can activate E2 and E8^E8 transcription from the viral genome.

To investigate this hypothesis, the following transfections of 293FT and U2OS cells were performed:

- HPV18E1-
- HPV18E2-
- HPV18E1- + E2 c.o. e.v.

Two primer sets, E2 and E8^E2, were used in this experiment. The E2 and E8^E2 mRNA expression levels were normalized to the GAPDH mRNA expression and compared within cell lines. The results indicate that the viral mRNA expression levels for HPV18E2- are significantly lower compared to HPV18E1- in both cell lines (**Figures 9A and 9C**). This result was expected since E2 protein plays a crucial role in transcription, hence its absence decreases transcription levels. Additionally, this result indicates that some amount of the E2 protein is also expressed in 293FT cells, although its expression remains under the detection limit in the immunoblotting assay (**Figure 6A**).

Compared to HPV18E1- transcripts, overexpression of the codon-optimized E2 resulted in a strong increase in mRNA levels of E2 and E8^E2 in both cell types. In U2OS cells, expression of E2 and E8^E2 increased approximately 7-fold in response to the overexpressed E2 protein, whereas in 293FT cells, E2 mRNA levels were over 3000-fold higher, and E8^E2 mRNA levels were increased by over 400-fold relative to HPV18E1- (**Figure 9A and 9B**).

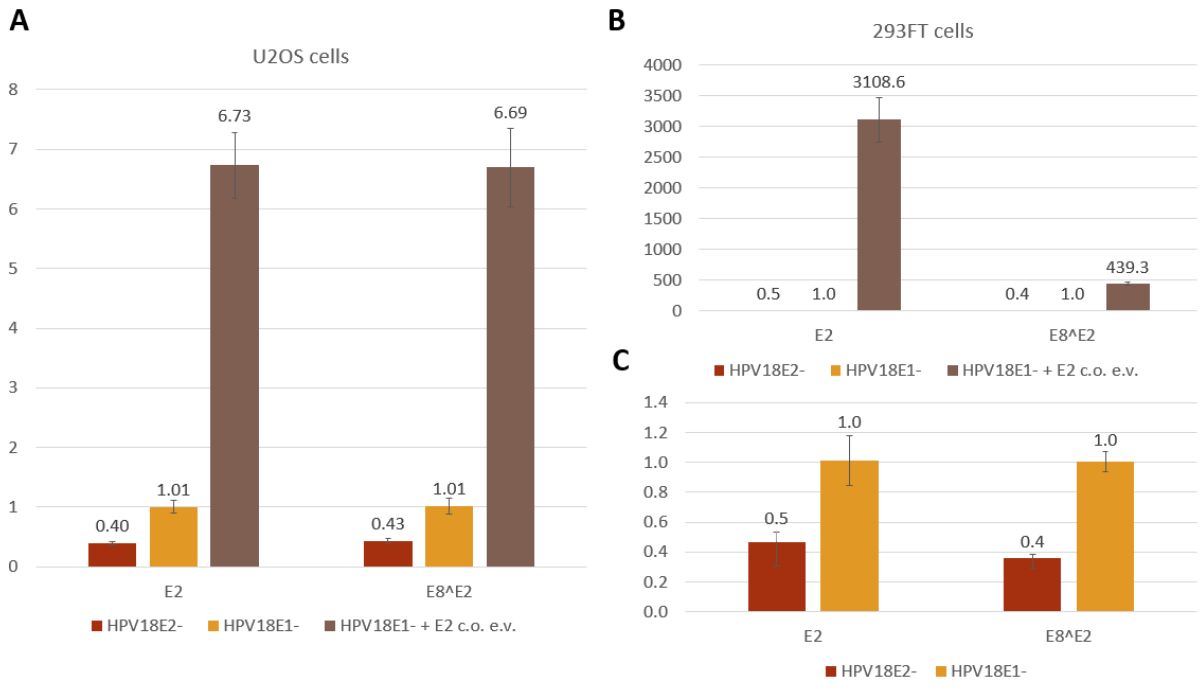

Figure 9. Expression of E2 and E8^E2 is triggered by the overexpressed E2 protein in 293FT and U2OS cells. Cells were collected 48 h after transfection. Viral gene mRNA expression levels were measured using RT-qPCR. The y-axis represents mRNA expression levels normalized to the GAPDH expression levels. Error bars represent standard deviations. Comparison of HPV18E1-, HPV18E2- and

HPV18E1- + E2 c.o. e.v. transfection was done within each cell type. HPV18E1- mRNA levels were taken as 1. (A) Comparison for U2OS cells. (B) Comparison for 293FT cells. (C) Comparison for 293FT for only HPV18E1- and HPV18E2, presented in order to illustrate error bars for these samples.

3.2.5 Replication of HPV18 and HPV18E8- genomes in different cell types

E8^E2 protein that arises from splicing between E8 and E2 genes represses both transcription and replication of viral genome (McBride, 2021; McBride, 2017). HPV genomes with an E8^E2 knock-out mutation have been observed to replicate at significantly higher levels in several cell types (Dreer *et al.*, 2017). The results of my previous experiments showed that E8^E2 mRNA expression level was one of the highest in 293FT cells (**Figure 9B**). These observations led to a hypothesis that elevated E8^E2 expression might account for the deficiency of the HPV18 replication and transcription in different cell lines that usually are not permissive for HPV replication. On the other hand, analysis of the HPV18E8- genome replication in 293FT cells suggested that deficiency in E8^E2 repressor could not restore HPV18 replication (**Figure 5A**). In order to clarify this issue, I analyzed replication efficiency of the HPV18 and HPV18E8- genomes in three other cell types: HaCaT, C33A, and MCF7. U2OS cells were excluded from this experiment due to the known higher replication levels of HPV18E8- genome compared to its wild-type counterpart in these cells (Geimanen *et al.,* 2011).

Cells were transfected with similar amounts of the HPV18 and HPV18E8- genomes, and total DNA was extracted 2, 3, and 4 days post-transfection. As it was expected, none of the studied cells were able to support the wt HPV18 genome replication. However, it was determined that lack of E8^E2 enabled replication of the viral genome in HaCaT and C33A cells that represent cells of epithelial origin (**Figure 10** HaCaT and C33A panels). In HaCaT cells, there was a noticeable increase in signals from day 2 to day 4. While in C33A cells, the signal increased from day 2 to day 3 but weakened on day 4.

However, in MCF7 cells, HPV18E8- was unable to replicate, as it was also observed in 293FT cells (**Figure 10** MCF panel and **Figure 5A**). These results indicate that lack of E8^E2 repressor activity may enable replication of the HPV18 genome in a cell type-specific manner.

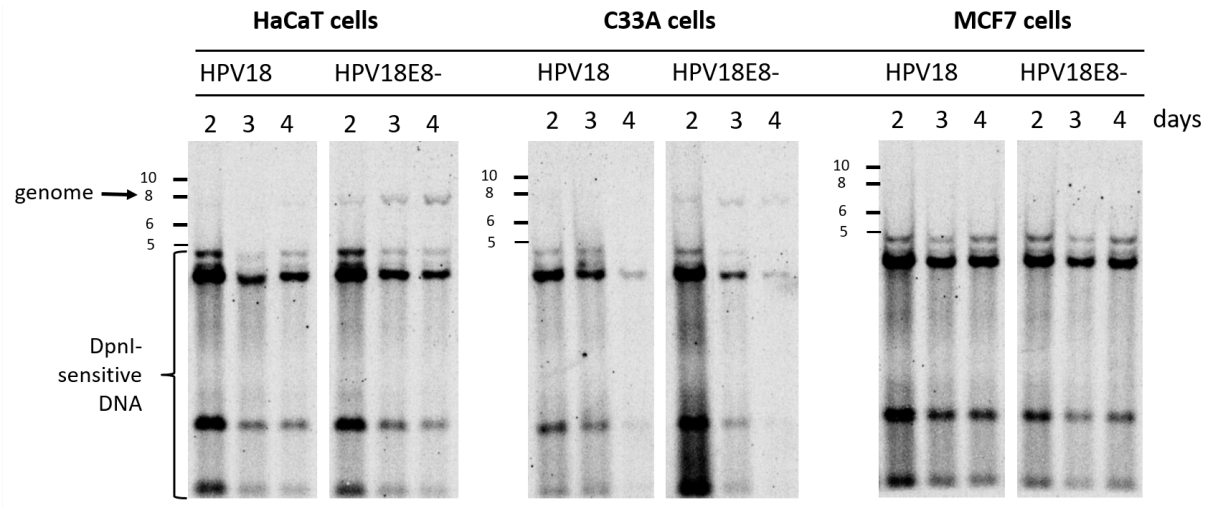

Figure 10. Replication of HPV18 and HPV18E8- genomes in different cell types. Replication of HPV18 and HPV18E8- genomes was evaluated using Southern Blot at 2, 3, and 4 days post-transfection. Cells were transfected with similar amounts of HPV18 m.c. and HPV18E8- m.c.. Total DNA was treated with DpnI and BglI restriction enzymes. It is visible that the amount of DpnI-sensitive DNA decreases from day 2 to day 4 time point. DNA ladder markings in kb are presented on the left side of each gel.

3.3 DISCUSSION

The importance of HPV is underscored by its etiological role in various types of epithelial malignancies, including cervical cancer. HPV16 and HPV18 are the primary cause behind the HPV-related cancers. While vaccination against HPV is highly effective in preventing infections, there is currently no cure for ongoing HPV infections. Therefore, antiviral research and drug development are crucial issues. However, HPV-related research faces challenges due to the limited availability of suitable cellular models for studying the viral life cycle. In this thesis conducted at the University of Tartu, the lifecycle of HPV type 18 was investigated in different cell types, aiming to identify factors that either support or restrict HPV18 replication. The study focused on understanding viral replication, transcription, and translation, with a particular emphasis on the viral E1 and E2 proteins, along with the repressor E8^E2 protein.

In the present thesis, it was established that 293FT, HaCaT, C33A, and MCF7 cells do not support replication of the wt HPV18 genome. However, knock-down of the viral E8^E2 transcriptional repressor led to restoration of HPV18 replication in cells of epithelial origin C33A and HaCaT. These results suggest that in these cells, either elevated expression or high activity of the E8^E2 protein may be at least one of the reasons behind absence of the HPV18 replication. As a next step, it would be desirable to investigate this causal relationship in C33A and HaCaT cells by conducting qPCR analysis of viral transcription patterns.

However, lack of the E8^E2 transcriptional repressor could not restore replication of the viral genome in other cell types studied, such as 293FT and MCF7 cells, suggesting that other factors inhibit genome replication in these cell types.

Further analysis showed that transcription levels of all early viral genes were extremely low in 293FT cells. Comparison of the viral gene expression levels in the HPV18 genomes deficient in either E1 or E2 expression indicated expression of some amount of E2 protein, as the levels of the viral transcripts were lower in the case of the E2-deficient genome compared with its E1-deficient counterpart. Although, neither E2 nor E1 were detectable through immunoblotting in 293FT cells. And since the E8^E2-deficient HPV18 genome doesn't replicate in 293FT cells, it can be concluded that the mechanism of replication repression in 293FT cells is not dependent on the E2 and E8^E2 viral proteins.

However, the overexpression of the E1 protein restored replication in 293FT cells. It has been previously shown that the function of the E1 helicase depends on its interaction with E2. E2 assists proper loading of E1 on the viral DNA to proceed with replication. My results suggest that the E2 protein expressed in 293FT cells is biologically active and capable of supporting

replication of the viral genome. However, the E1 protein, in addition to its very low expression levels, may either undergo degradation or localize outside the nucleus in these cells that in turn results in lack of viral genome replication. Additionally, it is possible that the overexpressed E1 protein does not require E2-mediated assistance to bind viral DNA. It has been shown that too high levels of the E1 protein may induce multiple DNA damage events and trigger DNA damage response pathways that in turn allows viral replication to occur via alternative non-conventional mechanism. The effect of E1 protein overexpression on MCF7, C33A, and HaCaT cells currently remains unknown.

Taken together, the results suggest that different mechanisms may be utilized to prevent viral replication in different cells. This study brings us closer to understanding the mechanism underlying the inhibition of replication and identifying the cellular factors responsible for restricting HPV18 replication in different cell models.

SUMMARY

In this thesis, I aimed to characterize HPV18 genome replication and transcription in different cellular models, focusing on the comparison of the HPV18 genome lifecycle in 293FT and U2OS cells. Additionally, replication of the E8^E2 deficient genome in C33A, HaCaT and MCF7 cells was analyzed.

The study resulted in the following conclusions:

- Wild type HPV18 genome does not replicate in 293FT, HaCaT, MCF7 and C33A cells
- Transcription levels of all early viral genes are 5 to 100-fold lower in 293FT cells, compared to U2OS cells
- The E1 and E2 proteins are not detectable through immunoblotting in 293FT cells
- Lack of the E2 protein expression has negative effect on transcription in 293FT
- Absence of the E8^E2 repressor protein restores replication in C33A and HaCaT cells, but has no effect on 293FT and MCF7 cells
- Replication of the HPV18 genome is restored by overexpression of E1 protein in 293FT cells

REFERENCES

Alhamlan, F. S., Alfageeh, M. B., Al Mushait, M. A., Al-Badawi, I. A., & Al-Ahdal, M. N. (2021). Human Papillomavirus-Associated Cancers. *Advances in experimental medicine and biology*, 1313, 1–14. https://doi.org/10.1007/978-3-030-67452-6_1

Antson, A. A., Burns, J. E., Moroz, O. V., Scott, D. J., Sanders, C. M., Bronstein, I. B., Dodson, G. G., Wilson, K. S., & Maitland, N. J. (2000). Structure of the intact transactivation domain of the human papillomavirus E2 protein. *Nature*, 403(6771), 805–809. https://doi.org/10.1038/35001638

Arbyn, M., de Sanjosé, S., Saraiya, M., Sideri, M., Palefsky, J., Lacey, C., Gillison, M., Bruni, L., Ronco, G., Wentzensen, N., Brotherton, J., Qiao, Y. L., Denny, L., Bornstein, J., Abramowitz, L., Giuliano, A., Tommasino, M., & Monsonego, J. (2012). EUROGIN 2011 roadmap on prevention and treatment of HPV-related disease. *International journal of cancer,* 131(9), 1969–1982. https://doi.org/10.1002/ijc.27650

Arbyn, M., Weiderpass, E., Bruni, L., de Sanjosé, S., Saraiya, M., Ferlay, J., & Bray, F. (2020). Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis. *The Lancet. Global health*, 8(2), e191–e203. https://doi.org/10.1016/S2214-109X(19)30482-6

Auster, A. S., & Joshua-Tor, L. (2004). The DNA-binding domain of human papillomavirus type 18 E1. Crystal structure, dimerization, and DNA binding. *The Journal of biological chemistry,* 279(5), 3733–3742. https://doi.org/10.1074/jbc.M311681200

Aydin, I., Weber, S., Snijder, B., Samperio Ventayol, P., Kühbacher, A., Becker, M., Day, P. M., Schiller, J. T., Kann, M., Pelkmans, L., Helenius, A., & Schelhaas, M. (2014). Large scale RNAi reveals the requirement of nuclear envelope breakdown for nuclear import of human papillomaviruses. *PLoS pathogens,* 10(5), e1004162. https://doi.org/10.1371/journal.ppat.1004162

Becker, K. A., Florin, L., Sapp, C., & Sapp, M. (2003). Dissection of human papillomavirus type 33 L2 domains involved in nuclear domains (ND) 10 homing and reorganization. *Virology*, *314*(1), 161-167. https://doi.org/10.1016/S0042-6822(03)00447-1

Belnap, D. M., Olson, N. H., Cladel, N. M., Newcomb, W. W., Brown, J. C., Kreider, J. W., Christensen, N. D., & Baker, T. S. (1996). Conserved features in papillomavirus and

polyomavirus capsids. *Journal of molecular biology,* 259(2), 249–263. https://doi.org/10.1006/jmbi.1996.0317

Bernard, H. U., Burk, R. D., Chen, Z., van Doorslaer, K., zur Hausen, H., & de Villiers, E. M. (2010). Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. *Virology*, 401(1), 70–79. https://doi.org/10.1016/j.virol.2010.02.002

Bienkowska-Haba, M., Williams, C., Kim, S. M., Garcea, R. L., & Sapp, M. (2012). Cyclophilins facilitate dissociation of the human papillomavirus type 16 capsid protein L1 from the L2/DNA complex following virus entry. *Journal of virology,* 86(18), 9875–9887. https://doi.org/10.1128/JVI.00980-12

Burley, M., Roberts, S. & Parish, J.L. Epigenetic regulation of human papillomavirus transcription in the productive virus life cycle. *Semin Immunopathol* 42, 159–171 (2020). https://doi.org/10.1007/s00281-019-00773-0

Clertant, P., & Seif, I. (1984). A common function for polyoma virus large-T and papillomavirus E1 proteins?. *Nature,* 311(5983), 276–279. https://doi.org/10.1038/311276a0

Cubie H. A. (2013). Diseases associated with human papillomavirus infection. *Virology,* 445(1-2), 21–34. https://doi.org/10.1016/j.virol.2013.06.007

Day, P. M., Baker, C. C., Lowy, D. R., & Schiller, J. T. (2004). Establishment of papillomavirus infection is enhanced by promyelocytic leukemia protein (PML) expression. *Proceedings of the National Academy of Sciences of the United States of America,* 101(39), 14252–14257. https://doi.org/10.1073/pnas.0404229101

de Martel, C., Plummer, M., Vignat, J., & Franceschi, S. (2017). Worldwide burden of cancer attributable to HPV by site, country and HPV type. *International journal of cancer,* 141(4), 664–670. https://doi.org/10.1002/ijc.30716

Deng, W., Lin, B. Y., Jin, G., Wheeler, C. G., Ma, T., Harper, J. W., Broker, T. R., & Chow, L. T. (2004). Cyclin/CDK regulates the nucleocytoplasmic localization of the human papillomavirus E1 DNA helicase. *Journal of virology,* 78(24), 13954–13965. https://doi.org/10.1128/JVI.78.24.13954-13965.2004

DiGiuseppe, S., Bienkowska-Haba, M., Guion, L. G., & Sapp, M. (2017). Cruising the cellular highways: How human papillomavirus travels from the surface to the nucleus. *Virus Research,* 231, 1-9. https://doi.org/10.1016/j.virusres.2016.10.015

DiGiuseppe, S., Bienkowska-Haba, M., Guion, L. G., & Sapp, M. (2017). Cruising the cellular highways: How human papillomavirus travels from the surface to the nucleus. *Virus research,* 231, 1–9. https://doi.org/10.1016/j.virusres.2016.10.015

Dreer, M., van de Poel, S., & Stubenrauch, F. (2017). Control of viral replication and transcription by the papillomavirus E8^E2 protein. *Virus research,* 231, 96–102. https://doi.org/10.1016/j.virusres.2016.11.005

Egawa, N., Egawa, K., Griffin, H., & Doorbar, J. (2015). Human Papillomaviruses; Epithelial Tropisms, and the Development of Neoplasia. *Viruses,* 7(7), 3863–3890. https://doi.org/10.3390/v7072802

Florin, L., Schäfer, F., Sotlar, K., Streeck, R. E., & Sapp, M. (2002). Reorganization of nuclear domain 10 induced by papillomavirus capsid protein l2. *Virology,* 295(1), 97–107. https://doi.org/10.1006/viro.2002.1360

Gautam, D., & Moody, C. A. (2016). Impact of the DNA Damage Response on Human Papillomavirus Chromatin. *PLoS pathogens,* 12(6), e1005613. https://doi.org/10.1371/journal.ppat.1005613

Gauthier, J. M., Dillner, J., & Yaniv, M. (1991). Structural analysis of the human papillomavirus type 16-E2 transactivator with antipeptide antibodies reveals a high mobility region linking the transactivation and the DNA-binding domains. *Nucleic acids research,* 19(25), 7073–7079. https://doi.org/10.1093/nar/19.25.7073

Geimanen, J., Isok-Paas, H., Pipitch, R., Salk, K., Laos, T., Orav, M., Reinson, T., Ustav, M., Jr, Ustav, M., & Ustav, E. (2011). Development of a cellular assay system to study the genome replication of high- and low-risk mucosal and cutaneous human papillomaviruses. *Journal of virology*, 85(7), 3315–3329. https://doi.org/10.1128/JVI.01985-10

Handisurya, A., Schellenbacher, C., & Kirnbauer, R. (2009). Diseases caused by human papillomaviruses (HPV). *Journal der Deutschen Dermatologischen Gesellschaft = Journal of the German Society of Dermatology* : JDDG, 7(5), 453–467. https://doi.org/10.1111/j.1610-0387.2009.06988.x

Islami, F., Fedewa, S. A., & Jemal, A. (2019). Trends in cervical cancer incidence rates by age, race/ethnicity, histological subtype, and stage at diagnosis in the United States. *Preventive medicine,* 123, 316–323. https://doi.org/10.1016/j.ypmed.2019.04.010

Jara, M., & Escobar, L. E. (2019). Papillomavirus in Wildlife. *Frontiers in Ecology and Evolution,* 7. https://doi.org/10.3389/fevo.2019.00406

Kadaja, M., Sumerina, A., Verst, T., Ojarand, M., Ustav, E. and Ustav, M. (2007), Genomic instability of the host cell induced by the human papillomavirus replication machinery. *The EMBO Journal,* 26: 2180-2191. https://doi.org/10.1038/sj.emboj.7601665

Karcher, S.J. (1991). Non-Radioactive DNA Hybridization Experiments for the Undergraduate Laboratory: The Southern Blot Analysis.

Kines, R. C., Thompson, C. D., Lowy, D. R., Schiller, J. T., & Day, P. M. (2009). The initial steps leading to papillomavirus infection occur on the basement membrane prior to cell surface binding. *Proceedings of the National Academy of Sciences of the United States of America,* 106(48), 20458–20463. https://doi.org/10.1073/pnas.0908502106

Kurg, R., Uusen, P., Võsa, L., & Ustav, M. (2010). Human papillomavirus E2 protein with single activation domain initiates HPV18 genome replication, but is not sufficient for long-term maintenance of virus genome. *Virology,* 408(2), 159–166. https://doi.org/10.1016/j.virol.2010.09.010

Laaneväli, A., Ustav, M., Ustav, E., & Piirsoo, M. (2019). E2 protein is the major determinant of specificity at the human papillomavirus origin of replication. *PloS one,* 14(10), e0224334. https://doi.org/10.1371/journal.pone.0224334

Lace, M. J., Anson, J. R., Thomas, G. S., Turek, L. P., & Haugen, T. H. (2008). The E8--E2 gene product of human papillomavirus type 16 represses early transcription and replication but is dispensable for viral plasmid persistence in keratinocytes. *Journal of virology,* 82(21), 10841–10853. https://doi.org/10.1128/JVI.01481-08

Leemans, C. R., Braakhuis, B. J., & Brakenhoff, R. H. (2011). The molecular biology of head and neck cancer. *Nature reviews. Cancer,* 11(1), 9–22. https://doi.org/10.1038/nrc2982

Li, Y., & Xu, C. (2017). Human Papillomavirus-Related Cancers. *Advances in experimental medicine and biology,* 1018, 23–34. https://doi.org/10.1007/978-981-10-5765-6_3

Lin, C., Franceschi, S., & Clifford, G. M. (2018). Human papillomavirus types from infection to cancer in the anus, according to sex and HIV status: a systematic review and meta-analysis. *The Lancet. Infectious diseases,* 18(2), 198–206. https://doi.org/10.1016/S1473-3099(17)30653-9

Liu, J., & Thorp, S. C. (2002). Cell surface heparan sulfate and its roles in assisting viral infections. *Medicinal research reviews,* 22(1), 1–25. https://doi.org/10.1002/med.1026

Mac, M., & Moody, C. A. (2020). Epigenetic Regulation of the Human Papillomavirus Life Cycle. *Pathogens,* 9(6), 483. https://doi.org/10.3390/pathogens9060483

McBride A. A. (2008). Replication and partitioning of papillomavirus genomes. *Advances in virus research,* 72, 155–205. https://doi.org/10.1016/S0065-3527(08)00404-1

McBride A. A. (2013). The papillomavirus E2 proteins. *Virology,* 445(1-2), 57–79. https://doi.org/10.1016/j.virol.2013.06.006

McBride, A. (2017). Mechanisms and strategies of papillomavirus replication. *Biological Chemistry,* 398(8), 919-927. https://doi.org/10.1515/hsz-2017-0113

McBride, A. A., Schlegel, R., & Howley, P. M. (1988). The carboxy-terminal domain shared by the bovine papillomavirus E2 transactivator and repressor proteins contains a specific DNA binding activity. *The EMBO journal,* 7(2), 533–539. https://doi.org/10.1002/j.1460-2075.1988.tb02842.x

McBride, A.A. (2022) Human papillomaviruses: diversity, infection and host interactions. *Nat Rev Microbiol* 20, 95–108. https://doi.org/10.1038/s41579-021-00617-5

McShan, G. D., & Wilson, V. G. (1997). Reconstitution of a functional bovine papillomavirus type 1 origin of replication reveals a modular tripartite replicon with an essential AT-rich element. *Virology,* 237(2), 198–208. https://doi.org/10.1006/viro.1997.8793

Orav, M., Henno, L., Isok-Paas, H., Geimanen, J., Ustav, M., & Ustav, E. (2013). Recombination-dependent oligomerization of human papillomavirus genomes upon transient DNA replication. *Journal of virology,* 87(22), 12051–12068. https://doi.org/10.1128/JVI.01798-13

Ozbun M. A. (2002). Human papillomavirus type 31b infection of human keratinocytes and the onset of early transcription. *Journal of virology,* 76(22), 11291–11300. https://doi.org/10.1128/jvi.76.22.11291-11300.2002

Piirsoo, A., Kala, M., Sankovski, E., Ustav, M., & Piirsoo, M. (2020). Uncovering the Role of the E1 Protein in Different Stages of Human Papillomavirus 18 Genome Replication. *Journal of virology*, 94(20), 74-20. https://doi.org/10.1128/JVI.00674-20

Porter, S. S., Stepp, W. H., Stamos, J. D., & McBride, A. A. (2017). Host cell restriction factors that limit transcription and replication of human papillomavirus. *Virus research,* 231, 10–20. https://doi.org/10.1016/j.virusres.2016.11.014

Schelhaas, M., Shah, B., Holzer, M., Blattmann, P., Kühling, L., Day, P. M., Schiller, J. T., & Helenius, A. (2012). Entry of human papillomavirus type 16 by actin-dependent, clathrin- and lipid raft-independent endocytosis. *PLoS pathogens,* 8(4), e1002657. https://doi.org/10.1371/journal.ppat.1002657

Schiffman, M., & Kjaer, S. K. (2003). Chapter 2: Natural history of anogenital human papillomavirus infection and neoplasia. *Journal of the National Cancer Institute. Monographs,* (31), 14–19. https://doi.org/10.1093/oxfordjournals.jncimonographs.a003476

Straub, E., Fertey, J., Dreer, M., Iftner, T., & Stubenrauch, F. (2015). Characterization of the Human Papillomavirus 16 E8 Promoter. *Journal of virology,* 89(14), 7304–7313. https://doi.org/10.1128/JVI.00616-15

Ustav, E., Ustav, M., Szymanski, P., & Stenlund, A. (1993). The bovine papillomavirus origin of replication requires a binding site for the E2 transcriptional activator. *Proceedings of the National Academy of Sciences of the United States of America,* 90(3), 898–902. https://doi.org/10.1073/pnas.90.3.898

Ustav, M., Ustav, E., Szymanski, P., & Stenlund, A. (1991). Identification of the origin of replication of bovine papillomavirus and characterization of the viral origin recognition factor E1. *The EMBO journal,* 10(13), 4321–4329. https://doi.org/10.1002/j.1460-2075.1991.tb05010.x

White, P. W., Pelletier, A., Brault, K., Titolo, S., Welchner, E., Thauvette, L., Fazekas, M., Cordingley, M. G., & Archambault, J. (2001). Characterization of recombinant HPV6 and 11 E1 helicases: effect of ATP on the interaction of E1 with E2 and mapping of a minimal helicase domain. *The Journal of biological chemistry,* 276(25), 22426–22438. https://doi.org/10.1074/jbc.M101932200

Sanders, C. M., & Stenlund, A. (1998). Recruitment and loading of the E1 initiator protein: an ATP-dependent process catalysed by a transcription factor. *The EMBO journal,* 17(23), 7044–7055. https://doi.org/10.1093/emboj/17.23.7044

NON-EXCLUSIVE LICENCE TO REPRODUCE THESIS AND MAKE THESIS PUBLIC

I, Sofiya Babok

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright, my thesis

“Characterization of the high-risk human papillomavirus 18 genome replication and transcription in different cellular models” supervised by Marko Piirsoo and Alla Piirsoo

2. I grant the University of Tartu a permit to make the work specified in p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work from ***24/05/2023*** until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in p. 1 and 2.

4. I certify that granting the non-exclusive licence does not infringe other persons’ intellectual property rights or rights arising from the personal data protection legislation.

Sofiya Babok

24/05/2023