DISSERTATION

submitted to the

Combined Faculties for the Natural Sciences and for Mathematics

of the Ruperto-Carola University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

presented by

Diplom-Biologin Daniela Höfler

born in Frankfurt am Main/Germany

Oral examination:……

HPV16 RNA patterns as diagnostic marker for cervical cancer precursor lesions: Validation by newly developed high-throughput RT-qPCR

Referees: Prof. Dr. Ralf Bartenschlager

Prof. Dr. Martin Müller

Declarations according to § 8 of the doctoral degree regulations:

I hereby declare that I have written the submitted dissertation myself and in this process have used no other sources or materials than those expressly indicated. I hereby declare that I have not applied to be examined at any other institution, nor have I used the dissertation in this or any other form at any other institution as an examination paper, nor submitted it to any other faculty as a dissertation.

Heidelberg, Daniela Höfler

Acknowledgements

The thesis was performed under the supervision of Dr. Michael Pawlita and Dr. Markus Schmitt from December 2009 to October 2013 in the German Cancer Research Center (DKFZ, Heidelberg), Research Program Infection and Cancer, Division of Genome Modifications and Carcinogenesis.

For reviewing this thesis I would like to thank Prof. Dr. Ralf Bartenschlager and Prof. Dr. Martin Müller.

I would like to thank all members of my committee, Prof. Dr. Ralf Bartenschlager, Prof. Dr. Martin Müller and Prof. Dr. Lutz Gissmann for the assistance they provided at all levels of the research project.

Thanks to Michael for giving me the idea of the project and opportunity to work in his lab. Moreover, I am most grateful for the stimulating discussions, the critical input regarding the data, the help and guidance with my presentations, for corrections and much more. I thank him for holding me to a high research standard and enforcing strict validations for each research result, thereby teaching me how to do quality research. Last but not least, I thank him for the opportunity to participate in international conferences in order to gain ideas and give interesting talks.

I thank Markus for the far-reaching, professional and kind supervision at every phase of this thesis. Your expertise, understanding and patience added considerably to my graduate experience and your constructive criticism were thought-provoking and helped me focus my ideas. Thank you for always being there to listen and give adivce. Also, I appreciate your effort regarding my publications and the possibility to attend conferences and to get to know other researchers. Moreover, I thank you for the amicable working atmosphere which I greatly enjoyed not only during my PhD but already for the diploma thesis work.

Prof. Dr. Lutz Gissmann and Dr. Gerd Böhmer (routine colposcopy clinic in Bad Münder, Germany) enabled the validation of the HPV16 RNA patterns with so many cervical samples. Many thanks for that. As part of this cooperation, thanks also go to Dr. Heinrich Neumann und Prof. Dr. v. Wasielewski who re-examined the histological slides besides the routine histological diagnosis.

Dr. Werner Nicklas and Petra Mauter enabled the further development of three multiplex assays detecting DNA-viruses, RNA-viruses and bacteria in laboratory rodents that I started

during my diploma thesis. With your help two of the assays are now an integral part of laboratory animal quality assurance program of the DKFZ. Thanks.

To Taxiarchis Katsinelos, Elmar Seiberlich, Jennifer Pajio and André Leischwitz for their experimental assistance during this thesis I would like to extend my thanks.

Thanks to all my collegues that made our lab a convivial place to work. In particular, I would like to thank Gordana Halec for her friendship, geniality and help in the past years. Many thanks also to all others, including Angelika Michel, Martina Willhauck-Fleckenstein, Lorena Baboci, Ute Koch, Julia Butt, Dana Holzinger, Monika Oppenländer, Tim Waterboer, Kristina Michael, and the ones I forgot to mention who have inspired me in research and in and outside of the lab.

My family has been a constant source of love, concern, support and strength all these years. Thanks.

Finally, I want to thank Matze who listened “voluntarily” to all my presentations and who gave me all the laughter throughout the last four years.

Table of Contents

I Summary

Cervical cancer (CxCa) is the second most common cancer among women world-wide. DNA of 14 high-risk human papillomavirus (hrHPV) types is found in almost all cervical cancers (>96.3%), with HPV16 being the most prevalent type. Cervical cells can be screened for abnormalities using the cytological Papanicolaou (Pap) test and for the presence of HPV DNA and HPV E6/E7 full-length mRNA. However, all of the available tests show considerable drawbacks exhibiting either poor clinical sensitivity or specificity. Reduced clinical specificity leads to over-treatment, additional costs and enormous anxiety for women concerned. A highly specific identification of high-grade lesions is desirable since those require surgical therapy contrary to early lesions that are likely to regress spontaneously. Recently described HPV16 RNA patterns comprising the upregulated transcripts E6*II and/or E1C, and the downregulated transcripts E1^E4 and/or L1 in HPV16-transformed cervical cells showed potential to substantially improve HPV-based cervical precancer screening: in a small cytology-based pilot study these patterns identified severe HPV16-induced cervical lesions with a clinical sensitivity of 74% and a specificity of 84%. However, singleplex nucleic acid sequence-based amplification (NASBA) assays used in the previous work was labour, time and cost intensive.

In this PhD thesis, novel quantitative multiplex high-throughput reverse transcription PCR were developed for the detection of the HPV16 RNA patterns with detection limits below 10^2 transcript copies per PCR. Cross-reactivity with unspliced HPV16 sequences and cellular background DNA/RNA was excluded. The comparison of the newly developed assay and the established singleplex NASBA assays using RNA from 32 HPV16-positive fresh-frozen oropharyngeal squamous cell carcinomas indicated a good quantitative correlation. The concordance of HPV16 RNA patterns between both methods was 100%. In order to validate HPV16 RNA patterns as diagnostic marker for cervical cancer and its severe precursor lesions 165 single HPV16 DNA-positive cervical cell samples were analysed. The sensitivity for CIN3 and CxCa was 88% and the specificity 84%, even in archived specimens with low RNA quality. Several HPV16 RNA patterns "false-positive" CIN1 lesions in follow-up had progressed to CIN2 or CIN3 lesions, raising the possibility that HPV RNA patterns could be an earlier marker for developing severe lesions than histology. Poor quality of RNA extracted from formalin-fixed paraffin embedded tissues strongly limits HPV16 RNA patterns analysis. To extend the HPV RNA patterns to the further seven most frequent hrHPV types 18, 31, 33, 35, 45, 52 and 58, unknown splice junctions were identified. For HPV18 singleplex assays were developed. The detection limit ranged between 10^1 E6*I and E1C and 10^2 E1^E4 copies per PCR and will be improved in future experiments.

II Zusammenfassung

Das Zervixkarzinom (CxCa) ist weltweit die zweithäufigste Krebserkrankung bei Frauen. An dessen Entstehung sind humane Hochrisiko-Papillomviren, insbesondere HPV16 beteiligt. Zervixabstriche können sowohl zytologisch mit Hilfe des Papanicolaou (Pap) Tests, als auch durch den Nachweis von HPV DNA und HPV E6/E7 volle-Länge mRNA auf Abnormitäten untersucht werden. Alle bisher in der Zervixvorsorgeuntersuchung verfügbaren Methoden weisen einen Mangel an klinischer Sensitivität oder Spezifität auf. Beispielsweise kann eine geringe/moderate klinische Spezifität zur unnötigen Behandlung von Patientinnen, zu zusätzlichen Kosten und zur immensen Sorge der Betroffenen führen. Um dem vorzubeugen ist eine hochspezifische Identifizierung von hochgradigen Läsionen erstrebenswert, da nur diese Frauen chirurgischer Therapie bedürfen. Frühe Läsionen regredieren im Gegensatz dazu voraussichtlich spontan. Kürzlich wurden diagnostische HPV16 RNA Muster identifiziert, die die Hochregulierung der E6*II und/oder E1C Transkripte und die Herunterregulierung der E1^E4 und/oder L1 Transkripte in HPV16-transformierten zervikalen Zellen beinhalten. Diese Muster weisen das Potential auf, bisherige HPV-basierte Tests deutlich zu verbessern: In einer kleinen, Zytologie basierten Pilotstudie konnten mittels der Muster mit einer klinischen Sensitivität von 74% und Spezifität von 84% zwischen leichten und schweren Läsionen unterschieden werden, jedoch mit dem Nachteil, dass die verwendeten singleplex Nukleinsäuresequenz-basierten Amplifikations-Reaktionen (NASBA) arbeits-, zeit- und kostenaufwendig sind.

In der vorliegenden Arbeit wurde eine neue quantitative Hochdurchsatz-PCR für die Detektion der HPV16 RNA Muster entwickelt. Die analytische Sensitivität erstreckte sich zwischen 10^1 und 10^2 Trankriptkopien pro PCR. Kreuzreaktionen mit ungespleißten HPV16 Sequenzen und mit zellulärer Hintergrund-DNA/RNA wurden ausgeschlossen. Der Vergleich der neuen Methode und der etablierten NASBA, anhand von RNA aus 32 HPV16-positiven frisch-gefrorenen Karzinomen des Oropharynx, deutet sowohl eine gute quantitative Korrelation als auch eine 100%ige qualitative Konkordanz an. Um die HPV16 RNA Muster als diagnostische Marker zur Erkennung von CxCa und den hochgradigen Vorstufen zu validieren, wurden 165 –ausschließlich HPV16 DNA-positive– in PreservCyt gelagerte Zervixabstriche analysiert und eine klinische Spezifität von 84% und eine Sensitivität von 88% berechnet, selbst unter der Verwendung archivierten Materials mit geringer RNA Qualität. Mehrere durch HPV16 RNA Muster „falsch-positive“ CIN1 Läsionen schreiten in Verlaufsbeobachtungen zu CIN2 oder CIN3 fort, was ein Hinweis darauf sein könnte, dass die HPV16 RNA Muster ein früherer Marker für die Entwicklung schwerer Läsionen sind als histologische Untersuchungen. Schlechte Qualität extrahierter RNA aus Formalin-fixiertem, Paraffin-eingebetteten Gewebe limitiert die HPV16 RNA Muster Analyse stark. Um die HPV RNA Muster auf die häufigsten sieben Hochrisiko-HPV Typen zu übertragen, wurden alle bisher unbekannten Spleiß-Junctions identifiziert und bereits für die Entwicklung von singleplex PCR für HPV18 RNA Muster genutzt, deren analytische Sensitivität sich zwischen 10^1 und 10^2 Trankriptkopien pro PCR erstreckt und in zukünftigen Experimenten noch verbessert wird.

1 Introduction

1.1 Papillomaviruses

Papillomaviruses (PV) form their own family, *Papillomaviridae*, and are small, non-enveloped viruses. Their icosahedral ~55 nm diameter capsid is composed of 72 pentameric capsomers (Figure 1.1).

PV infect the skin and mucosal epithelium of mammals but also birds. Some PV types cause benign tumours, e.g. condylomata acuminata, and other types can transform cells leading to malignant tumours. Several mucosal human papillomaviruses (HPV) are associated with the development of cancer, including cervical, vaginal, vulvar, penile, anal and oropharyngeal cancer.

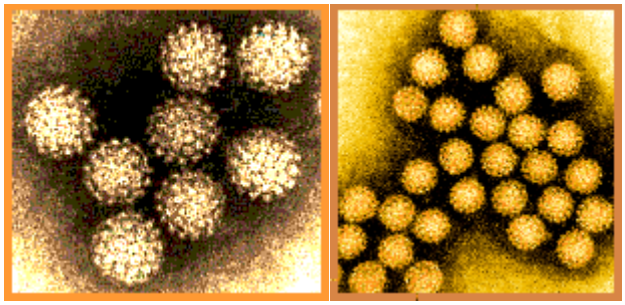


Figure 1.1. **Structure of HPV**. Electron microscopic pictures of HPV[1].

1.2 Human papillomaviruses

1.2.1 HPV genome organisation

All PV contain closed-circular double-stranded circular DNA which replicates as an extrachromosomal plasmid in the nucleus of infected keratinocytes. Although the viral genome can vary slightly in size between different HPV types, it typically contains around 8,000 basepairs (bp), and encodes up to nine open reading frames (ORF) from the same DNA strand. Three functional areas have been identified: the long control region (LCR), also called upstream regulatory region (URR), the “early” and the “late” transcription regions (Figure 1.2). The URR of about 800 bp contains no ORF but numerous regulatory sequence elements, such as transcription factor binding sites, viral transcription factor E2

binding sites, and the origin of viral DNA replication with binding sites for the viral replication protein E1. Furthermore, it contains silencer as well as enhancer sequences and harbours the p97 core promoter upstream of the E6 ORF. The "early" region consists of the E1, E2, E4-E7 ORF that are necessary for the replication of viral DNA. The "late" region encodes the L1 and L2 protein that are necessary for the assembly of newly produced virus particles[2-4].

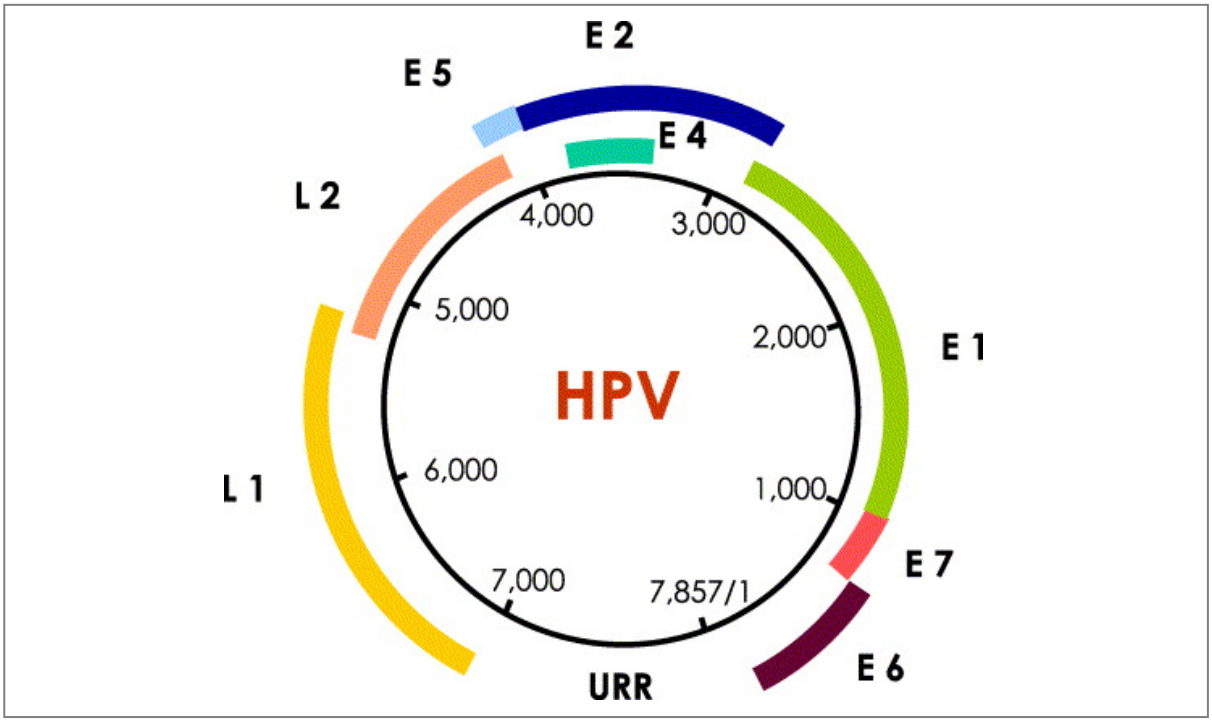


Figure 1.2. **HPV genome showing the arrangement of early (E) and late genes (L) and the upstream regulatory region (URR)[5].** In yellow and orange: late structural proteins; in blue, green, red and purple: early proteins; in red and purple: viral oncogenes.

1.2.2 HPV transcription

HPV generates mRNAs that can express than one functional protein through independent, tandemly arranged ORF. Since HPV16 has a focal role within this work, the transcription mechanisms are exemplarily presented for this type (Figure 1.3).

The expression of HPV16 is a complex process using at least two promoters, multiple splice donor and acceptor sites and two polyadenylation signals. The process also includes post-transcription mechanisms. All transcripts are poly-cistronic.

The promoter (p97) for the early genes is located in the URR and the late promoter (p670) for the late genes in the E7 region. The activity of p670 is suppressed in undifferentiated cells and activated in terminally differentiating keratinocytes[6].

Depending on splicing events, up to three different reading frames are utilised for translation. Maturation of viral RNA involves complex splicing. The HPV16 transcriptome exhibits several splice donor (nt 226, 880, 1302 and 3632) and acceptor (nt 409, 526, 742, 2582, 2709, 3358 and 5639) sites creating at least 11 different splice junctions[4,7-12]. The functions of the various spliced transcripts are still under investigation. Spliced transcripts may allow more effective translation of a downstream ORF[13-15]. One mechanism to achieve this is leaky scanning in which upstream start codons when lying in an unfavourable context are ignored[16,17]. Moreover, the translation of spliced transcripts can generate truncated proteins with functions important for the regulation of the viral life cycle[18,19]. HPV transcript patterns and splice junctions have been investigated in HPV16-infected cervical lesions[20], cell lines[11,21-24] and keratinocytes immortalised by HPV16[10,20].

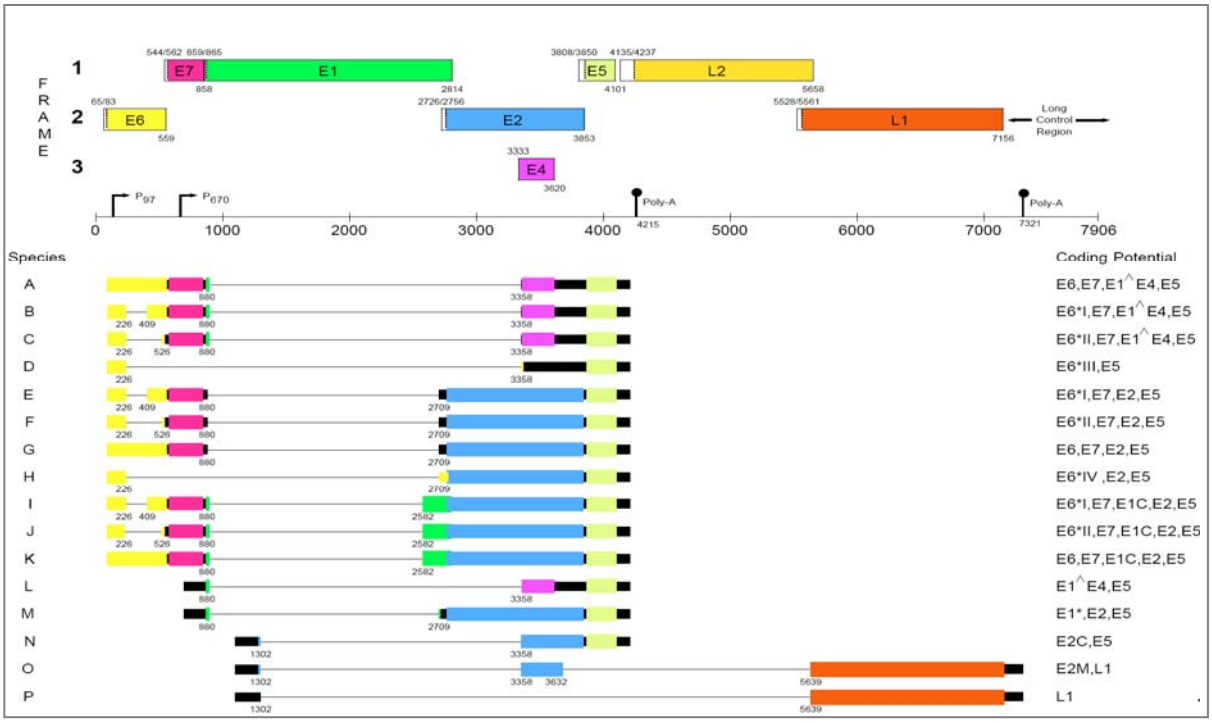


Figure 1.3. **Transcription map of HPV16**. ORF are shown as colored rectangles in their proper reading frames (top of the figure). The first number at the upper left end of the rectangles matches to the nucleotide (nt) position of the ORF start, id est (i.e.) the first nt following a stop codon. The second number is the nt position of the first ATG, also indicated by a dotted line within the rectangles. The position of the last nt in the stop codon of each ORF is written at the lower right corner of the rectangles. Located below the genome scale

are diagrams of spliced mRNA species with their coding potential at the right. Untranslated exons are illustrated by black rectangles, while intervening introns are indicated by black lines. Colored rectangles indicate in-frame sequences that can potentially code for proteins. Numbers printed below the lines indicate the 5' and 3' splice junctions. The promoter for the last three transcripts i.e. species N-P has not been mapped. Transcripts encoding full length (fl) E1 and L2 protein are not shown. Potential truncated gene products of E6 and E1 are indicated by asterisks (*). The fusion product of the E1 and E4 protein is indicated as E1^E4[25]. Source [26].

1.2.3 Major HPV16 proteins and their transcripts

The fl E1 protein is encoded by the unspliced E1 ORF-containing transcript. It is required for viral DNA replication, has a helicase activity and binds to cellular proteins including replication protein A (RPA) and DNA polymerase α primase[27-30]. E1 with a truncated N-terminus (E1C) is translated from 880^2582 spliced mRNA (Figure 1.6, species I-K) and acts as trans-activator of the URR[18,31].

The E2 protein is mainly encoded by spliced 226^2709 (E6*IV) mRNA (species H) but also to a lower extent by the 880^2709 mRNA (species E-G) and 880^2582 mRNA (species I-K) initiating mostly from the p97 promoter[17]. Viral DNA replication and viral segregation require E2[32]. E2 is a DNA-binding protein recognizing a palindromic motif in the non-coding region of the viral genome (four such motifs in HPV16). Via protein-protein interaction, E2 recruits E1[33]. E2 can also be active as a transcription factor that regulates the viral early promoter p97 causing auto-regulation of E2 expression and regulation of the expression of viral oncogenes (E6 and E7). At low level, E2 is a transcriptional activator, whereas at high level, E2 represses oncogene expression[34]. A shorter form of E2 (E2C), translated from the spliced 1302^3358 mRNA (species N), is presumably a replication and transcription repressor that inhibits the function of fl E2[35]. The presumed E2M protein with unknown function could be expressed from double-spliced transcripts (species O).

The E4 ORF, located in the early region, overlaps with the E2 ORF in a different reading frame. Its gene product, E1^E4 translated from 880^3358 spliced mRNA (species A-C, L) is only expressed in differentiated, upper epithelial layers and plays a role by binding to proteins of the keratin cytoskeleton during virus assembly and virus release[36].

The E5 protein is translated from an unspliced E2/E5 transcript (species E-G, M), but not from the E1^E4/E5 transcript (species A-C, L)[37]. It is a transmembrane protein that remains predominantly in the endoplasmatic reticulum (ER), by associating with the vacoular proton, it can delay the process of endosomal acidification[38-40]. This leads by affecting the recycling of growth factor receptor on the cell surface to an increase in epidermal growth factor-mediated (EGF) receptor signalling and the maintenance of a replication competent environment in the upper epithelial layer[41]. The E5 expression is observed to be lost during the integration of HPV DNA into the cellular genome, seen during progression to cervical cancer[42]. E5 is therefore not obligatory in late events of HPV-mediated carcinogenesis.

The E6 protein is solely encoded by mRNA containing the fl E6 ORF (including species A, G, K), promoting proliferation of infected cells and thus leading to their resistance to apoptosis. The high-risk (hr) E6 protein forms a tripartite complex with p53 and the cellular ubiquitin ligase E6-associated protein (E6AP), which leads to proteosome-mediated degradation of p53[43]. On the other hand, E6 proteins from low-risk HPV (lrHPV) types do not bind p53 at detectable levels and have no effect on p53 stability *in vitro*[44]. The hrHPV E6 proteins have a C-terminal PDZ ligand domain that activates the degradation of several cellular targets, thought to be involved in the regulation of cell growth and attachment[45]. Another important characteristic of E6 is its ability to stimulate the catalytic subunit of telomerase, which adds hexamer repeats to the telomeric ends of chromosome[46]. Such an activity may predispose to long-term infection. However, in case of E7 absence, $p16^{INK4A}$ would inhibit these E6 functions[47]. Additionally to the fl E6, spliced RNA species can generate four either truncated (E6*I and *II) or fused E6 forms (E6*III or *IV). However, the functions of the proteins generated from the different transcripts are unclear. Alloul et al. reported that E6*I trans-activated and E6*IV trans-represses the viral URR[18], other researchers argued that these protein fragments are not themselves important, and that the splicing event is responsible for a shift in favour of E7 translation[13]. Recently, an E6^E7 fusion transcript, spliced from 226^742 and of unknown function has been described[48].

E7 is translated from alternatively spliced E6*I mRNA (species B, E, I) that all use the same splice donor (nt 226) and acceptor (nt 409) sites[13-15], or as others suggest, from the full-length (fl) E6/E7 mRNA[11,16,49]. The E7 protein is mainly responsible for maintaining differentiating cells in S-phase[50]. It binds to the retinoblastoma tumour suppressor protein (pRb) and disrupts the association between pRb and the E2F family of external growth

factors. E2F then trans-activates cellular proteins, including $p16^{INK4A}$ and enables rapid entry of the cell into S-phase. In addition, E7 targets a wide range of proteins involved in cell proliferation, as histone deacetylases, components of the transcription complex AP1 and the cyclin-dependent kinase inhibitors p21 and p27. It also stimulates the S-phase genes cyclin A and E. These interactions induce multiple cellular responses including inhibition of $p16^{INK4A}$ and stabilisation of the p53 protein that normally counteract the stimulated cell replication by induction of apoptosis. However, as stated above, only hrHPV E6 increases degradation of p53 and thereby blocks the cellular response. Conversely, E7 acts synergistically with E6 by rescuing it from $p16^{INK4A}$ inhibition (reviewed in[51]).

L1 is encoded in transcripts initiated at p670, either unspliced or spliced between splice donor nt 1302 and splice acceptor nt 5639 (species P) (described in W12 cells[10]) or nt 3632 (species O). Studies addressing cell binding with non-infectious virus-like particles (VLP) that self-assembled from overexpressed L1 or both L1 and L2 showed similar binding characteristics, implying that L1 contains the major determinants for initial attachment[2,52,53].

L2 protein is expressed only from unspliced late transcripts. During late stages of the productive infection the minor capsid protein, L2, and the major capsid protein, L1, are expressed in differentiated cells close to the surface of the epithelium. They form the viral capsids in the granular cell layer and virions are believed to be released in a non-cytolytic manner[3]. L1 protein has been found only in the upper layers of the epithelium. However, L1 and L2 RNA can also be detected in lower epithelial layers[54,55] indicating that their expression is also regulated post-transcriptionally[4,31].

1.2.4 HPV life cycle

HPV replicates solely in keratinocytes of the stratified epithelium of skin and mucosa. Mucosal HPV are mainly sexually transmitted and thus people with multiple sexual partners are at increased risk for acquiring an infection. The productive life cycles of all HPV are organised in a similar way (Figure 1.4) and are tightly linked to the differentiation program of the host keratinocytes.

The cycle is initiated by infectious particles reaching the basal layer of the epithelium, where they bind and enter into cells through small breaks[5]. Receptors involved in viral entry

are thought to be heparin sulphate proteoglycans playing a role in initial binding and/or virus uptake[56]. The presence of secondary receptors seems to be required for an efficient HPV infection, such as α6 integrin[57]. For HPV16, papillomavirus pseudovirions are taken into the cell by clathrin-coated endocytosis[58].

Following virus binding, entry and transport to the nucleus, the viral genome establishes itself as stable low copy-number episome in the basal cells (approximately 20-100 copies per cell)[32,59,60]. A stable pool of infected cells is maintained in the basal epithelial layer, while infected daughter cells detach and migrate into the suprabasal layer. Normally, suprabasal cells exit the cell cycle and begin to differentiate. During this process, nuclei are mostly degraded. However, in HPV-infected keratinocytes, the restraint on cell-cycle progression is lost by stimulating the G1-to-S phase through the expression of E6 and E7. As the cells move up through the epithelium, the activation of the late, differentiation-dependent viral promoter (p670 for HPV16) and as consequence an increase of viral transcription lead to a productive replication of the viral genome to thousands of copies per cell.

The up-regulation of the differentiation-dependent promoter is critical for the onset of late events and directs two set of transcripts; one set terminates at the early poly-A site while the second set terminates at the late poly-A site downstream of L1. The first group of late transcripts encodes E1^E4, E5, E1 and E2 and the second encodes the capsid proteins L1 and L2. The first protein expressed in the late phase of the viral life cycle is E1^E4. It is detected in the spinous and granular cell layers and has several functions[4,31,61]. Virus capsid proteins L1 and L2, are expressed only in cells of the granular layer with viral particle assembly taking place in the cornified layer. Infected cells are scaled off from the epithelial surface and may be transmitted directly to other individuals[31].

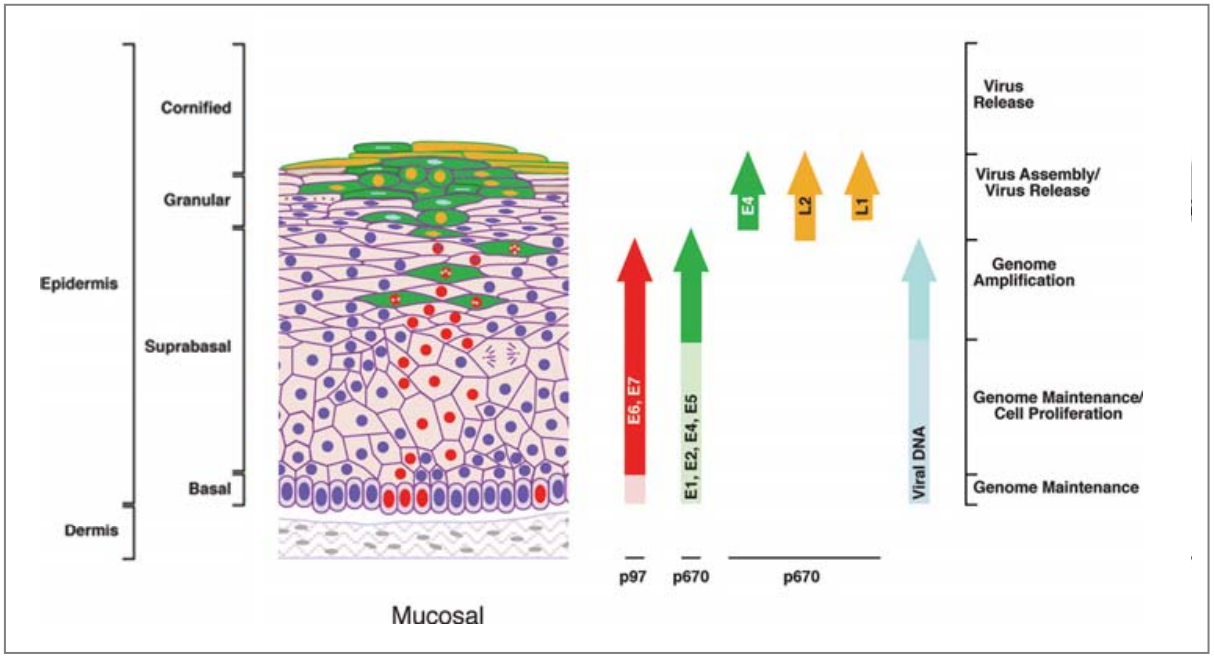


Figure 1.4. **HPV life cycle organisation and viral protein expression.** Modified from [59]. Cell layers of the epithelium are described and plotted on the left. Red nuclei indicate expression of E6 and E7 and green cells the expression of E1-E5. The expression of L1 and L2 (yellow) occurs in viral DNA containing cells in the upper epithelial layers.

1.3 Cervical cancer

Cervical cancer (CxCa) is the second most common cancer among women world-wide, with 530,000 estimated new cases and nearly 250,000 deaths every year[62,63]. About 86% of the cases occur in developing countries[64]. Worldwide, mortality rates of cervical cancer are substantially lower than incidence rates with a relative ratio of 52%[65]. In Germany, the annual number of new invasive CxCa cases in 2008 was 4,440 and the annual number of deaths due to CxCa was 2,018.

Approximately 80-85% of CxCa are squamous cell carcinomas (SCC) which are generated from dividing keratinocytes in the squamous epithelium of the ectocervix. Further 10-15% are adenocarcinomas (ADC) which arise from glandular cells located in the endocervix and adenosquamous cell carcinomas (ADSCC)[66]. Other CxCa histologies such as neuroendocrine tumours are rare[4].

1.4 HPV and its association with cervical cancer

Thus far, more than 150 HPV types have been identified of which 51 types infect the genital mucosa. The genital HPV types are divided into three groups based on their epidemiological association with CxCa: 14 hrHPV (types 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66 and 68), 6 putative hrHPV (26, 53, 67, 70, 73 and 82) and 31 lrHPV (e.g. 6, 11, 40, 42, 43, 44, 71)[67,68]. DNA of one of the 14 hrHPV types is found in almost all CxCa (>96.3%), with HPV16 being the most prevalent type when squamous cell carcinoma is diagnosed[69,70] (Figure 1.5). A milestone for public health which was awarded with the Nobel Prize in 2008 (Harald zur Hausen) was the discovery that CxCa is a consequence of an infection with some mucosal hrHPV types and has been designated as the first ever identified necessary infectious cause for a human cancer.

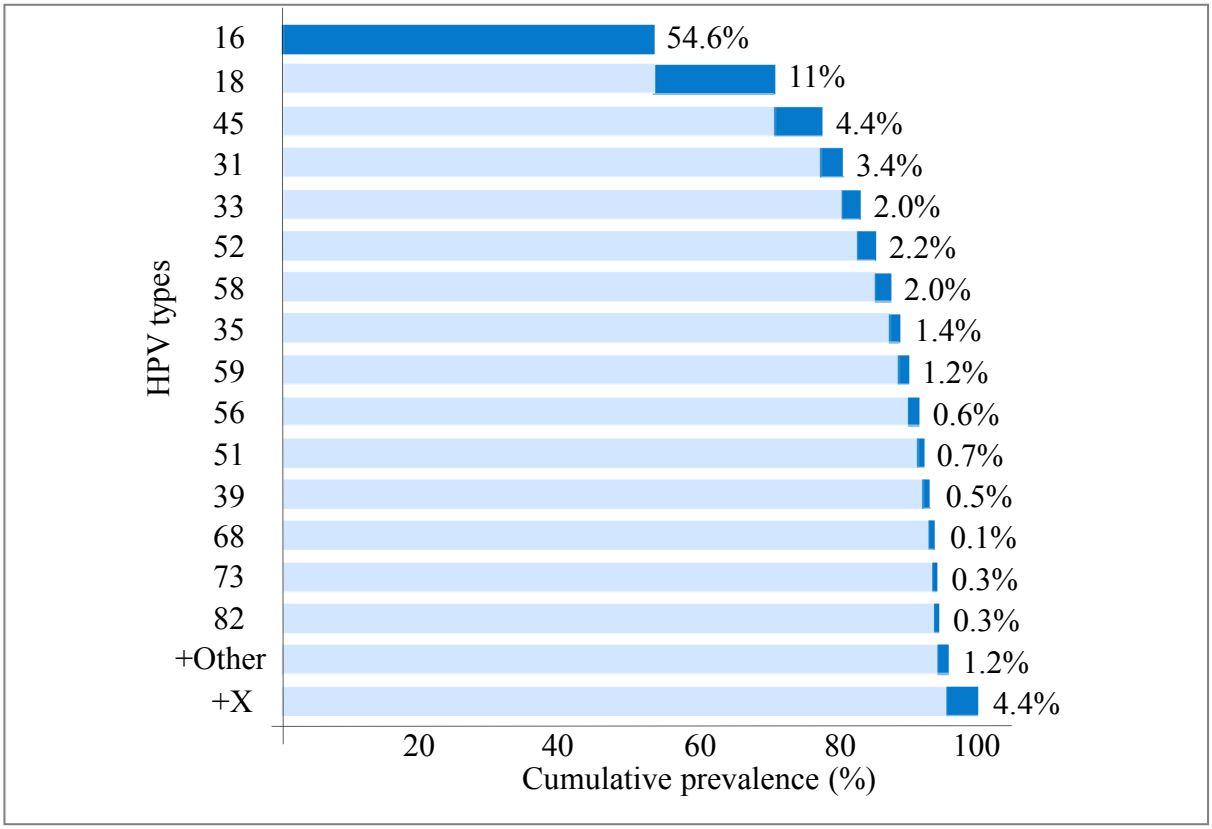


Figure 1.5. **Prevalence of HPV DNA in cervical cancer cases.** Modified from [71].

Most women are infected with HPV soon after beginning their first sexual relationship[72], with the highest prevalence in women aged 25 to 30 years. Thereafter, prevalence decrease rapidly. A second peak has been seen in older women[73,74]. The progression from an initial infection to carcinoma is a slow process over years to decades and is characterised by pre-malignant phases that can be diagnosed by cytological examination of exfoliated cervical

cells and confirmed by histological examination of cervical biopsies (Figure 1.6). Cytology distinguishes between the "no intraepithelial lesion or malignancy" (NIL/M), "atypical squamous cells of undetermined significance" (ASC-US), "low-grade squamous intraepithelial lesions" (LSIL) and "high-grade squamous intraepithelial lesions" (HSIL). Precursor lesions are defined histologically as cervical intraepithelial neoplasia (graded CIN1 to CIN3).

While 60-80% of women worldwide will be infected at least once per life time with genital HPV only 1-4% develop CxCa[75]. It has been assessed that the majority of CIN1 lesions (up to 57%) regresses spontaneously in healthy individuals, probably mediated by cellular immunity[75]. However, CIN1 lesions may remain (32%) or progress (11% to 20%) to CIN2[59,75]. CIN2 lesions still display a high regression rate (43% to 55%), while 22% progress to CIN3 and the rest persists[76-78]. Finally, CIN3 lesions may regress (32%), persist (56%) or progress (>12%) to CxCa[59,75,79].

Persisting infection with a hrHPV type increases the probability of developing CIN2, CIN3 and CxCa (≥CIN2)[31,80]. The transforming potential of hrHPV is induced by the deregulation of the viral oncogenes E6 and E7, together with additional alterations of viral and cellular genes and pathways[81,82]. This results ultimately in chromosomal instability and the accumulation of mutations. The underlying mechanisms for deregulation are diversely and integration, methylation as well as the involvement of HPV16 E1C protein are discussed: Integration of the HPV genome is a characteristic step in cervical cancer development and correlates with progression and for HPV16 DNA occurs in ~55% of cervical cancer cases during carcinogenesis[83]. This event often leads to a loss of E2 expression and, thus, to an elevated expression of the viral oncogenes E6 and E7. In CxCa HPV DNA integration has occurred into various regions of the human genome, with certain preferences. Recently, indications of a non-random integration and an accumulation in hotspot regions like the cytogenetic bands 3q28, 8q24.21 and 13q22.1 have been published[84]. However, the integration breakpoints of the viral and cellular genomes are different in all CxCa[85]. Integration though an indicator of progressed lesions, is not a necessity for transformation[31,83,84,86]. Thus, in CxCa containing episomal HPV genome methylation of the viral URR may affect regulatory features, such as the repressive function of E2, that control transcription and replication of the viral genome[82]. Fundamental changes in the methylation profile of transcription factor binding sites in the HPV16 URR may trigger neoplastic

transformation[82]. Furthermore, it has been described that the overexpression of HPV16 E1C protein may contribute to the transformation of cervical cells[87]. While the exact role of E1C remains to be elucidated, Alloul et al.[17] described this protein as possible trans-activator of the viral URR that could upregulate expression of E6 and E7.

Other risk factors increasing the risk of cancer development in addition to hrHPV infection are: deficiency of cellular immunity such as mediated by adjunctive HIV infection, pharmacological immunosuppression as well as long-term use of oral contraceptives and smoking[88-94], age at first intercourse, number of life-time partners, co-infection with chlamydia trachomatis, parity, age at first birth, and genetic predisposition[4,95-99].

During the progression to CxCa, the distribution of HPV16 proteins E4, L1 and E7 changes[59]. LSIL or CIN1 caused by HPV are similar to productive infection in the patterns of viral gene expression, and viral coat proteins can usually be detected in cells at the epithelial surface[59] (see chapter 1.2.4). HSIL or CIN2 or CIN3 lesions have a more extensive proliferative phase, with a retarded expression of viral coat proteins (Figure 1.6)[31]. In CxCa, the expression of viral coat proteins is entirely lost.

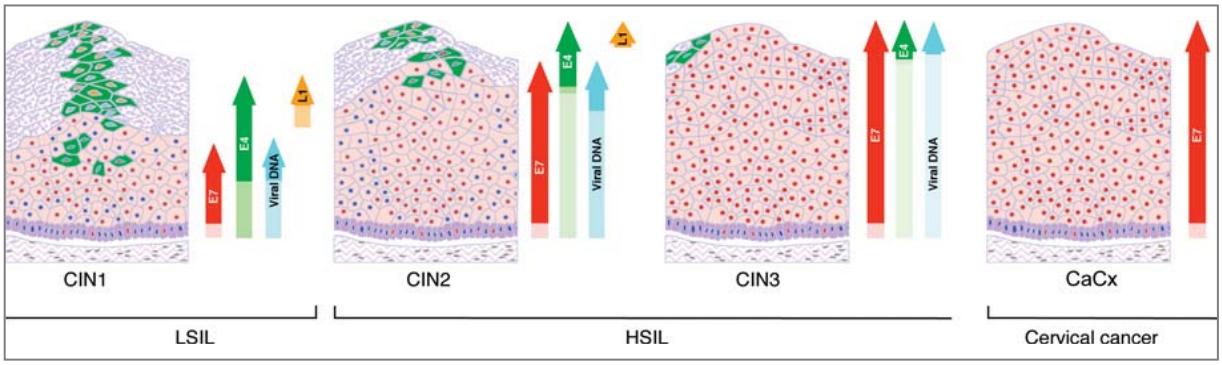


Figure 1.6. **HPV-mediated progression to cervical cancer.** Adapted from [59]. Red nuclei indicate expression of E7, green cells the expression of E4 and yellow nuclei the expression of L1.

1.5 Cervical cancer precursor screening

Most transient infections and mild lesions (CIN1 or LSIL) regress spontaneously, while low number of severe lesions (CIN2, CIN3 or HSIL) regress and high number progress to CxCa (Figure 1.7)[59,75,77-79]. Effective screening needs to distinguish between infections or mild lesions versus severe lesions and, thus, enable treatment of HSIL, CIN2 and CIN3 to reduce the risk of CxCa development. Several effective strategies for secondary cervical cancer

prevention by precursor screening have been identified, including either conventional or liquid-based cytology (LBC) detecting abnormal cells in the endocervical canal, hrHPV DNA tests looking for the presence of an infection, hrHPV DNA high viral load tests and hrHPV RNA testing (Figure 1.7) identifying presence of any lesion. These mentioned screening tests are described in detail in the following chapters.

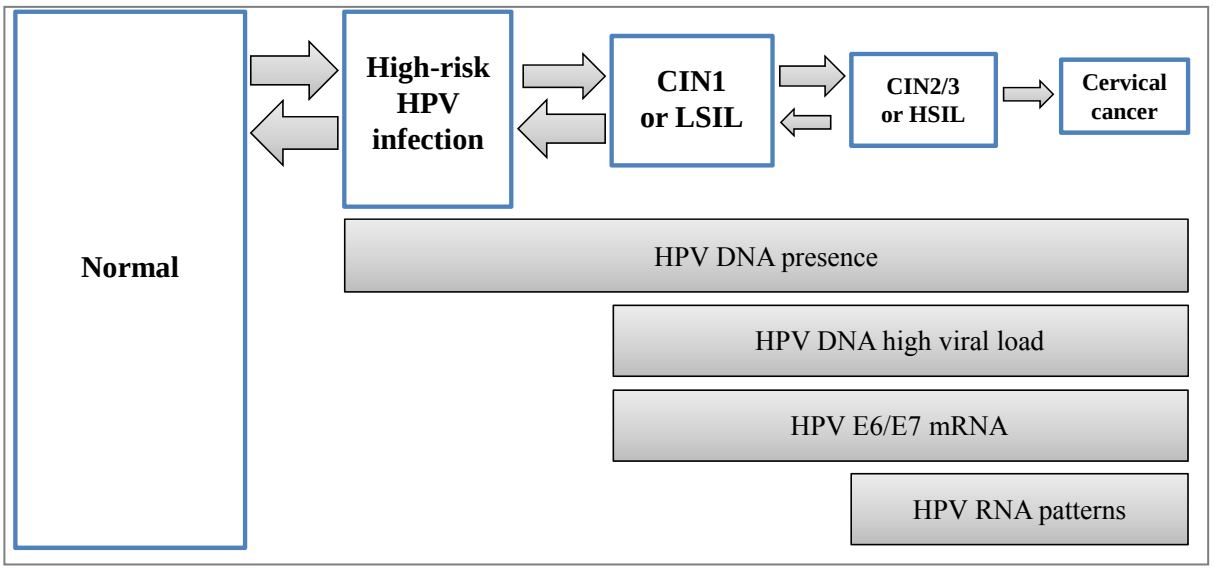


Figure 1.7. **Overview of spectrum of progression stages towards CxCa and their detection by HPV-based screening methods.** Modified from [100]. The sizes of blue-framed boxes indicate the fraction of women in the various groups. The size and direction of arrows indicate the relative frequencies of spontaneous regressions as well as progressions, respectively. The HPV-based screening tests are shown with the grey boxes indicating the range of CxCa progression stages reacting positive in the various assays. HPV DNA cannot discriminate between infection and high-grade lesion, HPV DNA high viral load as well as HPV E6/E7 mRNA does not discriminate any grade of CIN. HPV RNA patterns potentially discriminate between ≤CIN1 and ≥CIN2.

1.5.1 Screening using cytology and histology

Although it has never been tested in a controlled, randomised trial, the effectiveness of Papanicolaou (Pap) smear screening in reducing cervical cancer mortality is nearly universally accepted. Most developed countries saw dramatic reductions in the incidence and death rates from cervical cancer after the implementation of organised screening programs. The Pap test is a microscopical search for abnormal cells among exfoliated cells collected from the opening of the cervix during a pelvic examination and smeared and finally stained on a glass slide. However, the Pap test for cervical screening has severe limitations. One is the dependence on subjective judgement of the degree of mitotic activity

and nuclear atypia making lesion categorisation only moderately reproducible[101,102]. Another one is the poor sensitivity which implies a high failure rate in detecting cervical abnormalities in approximately 50% (10 to 70%) of women compared to colposcopy[103]. The moderate specificity of Pap smear cytology results in a high number of patients that are unnecessarily treated for a non-existing disease[31]. A newer method called LBC, in which the swab is suspended in preservative fluid, the cell layer applied by machines to the glass slide is more homogenous, allows computer-assisted reading and has logistical and operational advantages (interpretation at higher speed, lower rate of unsatisfactory smears and possibility of ancillary molecular testing using remnant cell suspension), but is more expensive and neither more sensitive nor more specific than conventional cytology[104-106].

Colposcopy-directed biopsies followed by histology are the clinical reference standard for diagnosing and grading of precancerous lesions. Advancing grades of CIN (grades 1–3) are distinguished mainly according to the extent of the vertical extension of abnormal cells in the cervical epithelium. Abnormal cells confined to the basal third are designated CIN1, abnormal cells restricted to the basal two-thirds are designated CIN2, and full-thickness extension of abnormal cells are designated CIN3. This division is arbitrary. A significant lack of reproducibility in CIN classifications has been described which can lead to inadequate treatment decisions[107-109].

1.5.2 Screening using HPV DNA tests

The combination of cytological results and HPV testing reaches a better sensitivity, with only slight reduction of specificity[110-112]. Cohort studies over five to ten years have shown that the combination of normal cytological result and negative HPV provide better long-term protection against CIN3 than cytological testing alone[113-117]. Women hrHPV DNA-positive in a single specimen and those with two or more positive specimens over time, were 16 and 216 times more likely to develop CIN than women who were HPV DNA-negative[31,118-120].

Despite the high sensitivity of HPV DNA testing, HPV infection is widespread and as many as 90% of women positive for HPV DNA subsequently test negative within 6-24 months[121-123]. Consequently, HPV DNA tests recognise also clinically not relevant infections or regressing lesions. Therefore, the positive predictive value (PPV) of a single hrHPV DNA-

positive result for the presence of or risk of developing an advanced cervical lesion is low: The resulting high proportion of test-positive but disease-negative diagnoses cause considerable anxiety for women concerned, over-treatment and unnecessary costs[124,125].

According to the guidelines of the "Deutsche Gesellschaft für Gynäkologie und Geburtshilfe" (DGGG) in 2010, HPV DNA testing is still not a part of cervical cancer precursor screening in Germany[112]. However, in the Netherlands the Health Council recommends that hrHPV testing should replace cytology as the primary screening method in women above 30 years in five to ten years intervals. To prevent unnecessary colposcopy referrals, hrHPV-positive women should not be offered colposcopy immediately but should be further stratified by means of triage testing. It is appropriate to use cytology for this purpose[126].

There are several tests in routine use for HPV DNA detection. The first-generation HPV DNA test, FDA proven in 1995, was the hybrid capture tube (HCT) (Qiagen, Hilden) and the second generation hybrid capture 2 (HC2) (Qiagen, Hilden), FDA proven in 1999. Both methods use direct signal amplification for the detection of eight hrHPV types (16, 18, 31, 33, 35, 45, 51, 56) without distinguishing which type(s) are present. HC2 includes additionally types 39, 52, 58, 59 and 68. The sensitivity range of HC2 has been recorded at about 80 to 90% and the specificity between 57 and 89%[127]. Both, the sensitivity (65% to 80-90%, Table 1.1) as well as the specificity (60% to 57-87%, Table 1.1) was increased compared to HCT through decreasing the detection limit (50,000 to 5,000 viruses/sample) and the cutoff, respectively[128]. Apart from HPV testing for carcinogenic types collectively, specific genotyping has recently been raised as one of the options to improve triage and screening[129]. One commercially available genotyping assay, the Linear Array HPV Genotyping (LA) (Roche Molecular Systems, Basel) is a qualitative PCR technique detecting 37 most prevalent (low, intermediate, and carcinogenic) HPV types: 6, 11, 16, 18, 26, 31, 33, 35, 39, 40, 42, 45, 51, 52, 53, 54, 55, 56, 58, 59, 61, 62, 64, 66, 67, 68, 69, 70, 71, 72, 73, 81, 82, 83, 84, IS39, and CP6108. However, LA tests exhibited greater sensitivity (91-97%), but lower specificity (47-51%), than HC2 for detecting ≥CIN2 (Table 1.1)[130,131].

Table 1.1. **Sensitivity and specificity of HPV-based tests**[127,130,131,128,139,140,132].

Diagnostic test	Detecting	Sensitivity in detecting ≥CIN2	Specificity in detecting ≥CIN2
Hybrid Capture Tube (HCT)	HPV DNA	~65%	~60%
Hybrid Capture 2 (HC2)	HPV DNA	80-90%	56-89%
Linear Array HPV Genotyping (LA)	HPV DNA	91-97%	47-51%
PreTect™ HPV-Proofer	HPV RNA	72%	79%
APTIMA HPV assay	HPV RNA	99%	47%

Besides the commercial tests, several in-house HPV DNA tests had been developed. Since the BSGP5+/6+-PCR/MPG assay was used in this thesis, it is described briefly: The assay comprises the homogeneous amplification of all known genital HPV types and a Multiplex HPV genotyping (MPG) assay with bead-based xMAP Luminex suspension array technology[133]. The Luminex analyser measures a hybridisation product of bead-coupled probes and streptavidin-R-phycoerythrin stained PCR-amplicons. The beads contain two spectrally distinct fluorochromes. A precise ratio of these fluorochromes creates 100 different beads with different spectra enabling a simultaneous detection (multiplexing) of up to 100 beads per reaction[134]. BSGP5+/6+-PCR/MPG determines viral load as this is a more clinically useful marker with a relatively simple technical feasibility of measurements and interpretation. An excellent clinical sensitivity and specificity for CIN3 and CxCa (≥CIN3) over 95% has been described[134]. However, it cannot discriminate between different grades of lesion, but between lesion and no lesion (Figure 1.7) and consequently, reduces the number of false-positive women with no lesion[134].

1.5.3 Screening using HPV RNA tests

Unlike HPV DNA testing, RNA detection allows the identification and analysis of transcriptionally active viruses. The introduction of preservation media for cervical smears that, apart from DNA and cell morphology, also conserve RNA, enhanced the application of HPV RNA detection methods. Using quantitative assays, high expression of E6/E7 fl and E6*I was found to correlate with cytologically diagnosed severe cervical dysplasia[135,136]. Thus, detection and quantitation of mRNA from E6 and E7 genes of hr types could be a more specific marker of the presence of high-grade disease and cancer similar to HPV DNA

tests measuring viral load (Figure 1.7). Commercial tests targeting HPV E6 and E7 mRNA are the PreTect™ HPV-Proofer developed by NorChip/BioMérieux, Inc. (Klokkarstua, Hurum, Norway) and the APTIMA® HPV assay by Gen-Probe (San Diego, CA, USA). The PreTect™ HPV-Proofer detects early fl mRNA by NASBA targeting E6 and E7 sequences (E6/E7) from hrHPV types 16, 18, 31, 33 and 45. Since a simultaneous DNA detection by this test has been described[137], a false-positive test result for a substantial number of HPV DNA-positive women with normal cytology is not surprising[138-140]. Overall, the sensitivity of the HPV-proofer for the detection of ≥CIN2 is 72.8% and the specificity 78.8% (Table 1.1)[140,141]. The APTIMA® HPV assay is an isothermal E6/E7 mRNA amplification method for 14 HPV types detecting ≥CIN2 with a sensitivity of 98.5% and a specificity of 46.9% (Table 1.1)[132]. In summary, like DNA-tests, E6/E7 transcript-based RNA tests will lead to considerable anxiety for women concerned, over-treatment and additional unnecessary costs.

1.6 $p16^{INK4a}$ as diagnostic marker for development of CIN

As discussed, alterations in the viral gene expression pattern mark the progression of a productive to a transforming infection. $p16^{INK4a}$ is a cellular correlate of the increased expression of oncogenic E6/E7 mRNA: $p16^{INK4a}$ overexpression has been shown in the vast majority of cervical precancers and cancers, whereas in normal tissue, $p16^{INK4a}$ expression is found only rarely[142,143]. Consequently, an overexpression of $p16^{INK4a}$ is a promising immunohistochemical biomarker for HPV-related cancers[144]. Immunochemical staining for $p16^{INK4a}$ offers a diagnostic adjunct in the evaluation of cervical histology specimens as well as cytological samples. However, in cytological diagnosis a dual staining with proliferation marker Ki-67 can provide a high sensitivity level for detecting underlying CIN2 and CIN3, whereas the specificity using this morphology independent dual biomarker approach may be further improved compared to specificity rates observed when single immunoreactivity for $p16^{INK4a}$ is applied[145].

Several studies describe an overexpression of $p16^{INK4a}$ in CIN1 (54-72.3%), CIN2 (86-91%), CIN3 (96-98.3%) and CxCa (98.5%) using $p16^{INK4a}$ histology and in LSIL (37%), HSIL (93%) and CxCa (99%) using $p16^{INK4a}$ cytology[142]. For these approaches, a commercial assay, the CINtec® $p16^{INK4a}$ (Roche mtm laboratories), has been introduced.

Because a single protein target would detect CIN associated with any carcinogenic HPV type, the cost of a $p16^{INK4a}$ assay is likely to be lower than direct HPV tests which require multiplexed reagents to reach equivalent sensitivity[100].

1.7 HPV16 RNA patterns

Recently, diagnostic HPV16 RNA patterns were identified with the potential to improve the clinical specificity of HPV-based tests substantially[31,87,146].

The expression levels of ten spliced HPV16 RNA sequences, 226^409 (encoding the E6*I protein), 226^526 (E6*II), 226^3358 (E6*III), 226^2709 (E6*IV), 880^2582 (E1C), 880^2709 (E2, E5), 880^3358 (E1^E4), 1302^3358 (E2C, E2M), 1302^5639, 3632^5639 (L1), five fl sequences, E6 fl, E7 fl, E1 fl, E5 fl, L1 fl and the cellular $p16^{INK4A}$ had been determined in NIL/M, LSIL, HSIL and CxCa by singleplex nucleic acid sequence-based amplification (NASBA)[146 31]. NASBA is a one-step isothermal process for amplifying RNA and consists of avian myeloblastosis virus reverse transcriptase (AMV-RT), reverse transcriptase (RT), T7 RNA polymerase and RNAse H with two oligonucleotide primers. NASBA quantified the viral transcripts using external standard curves.

Of all analysed HPV16 RNA sequences, five spliced transcripts E6*I, E6*II, E1^E4, E1C and L1 were identified to be highly diagnostic for discriminating mild from severe lesions. The prevalence of E6*I, E6*II and E1C gradually increased from NIL/M to CxCa while the prevalence of spliced E1^E4 and L1 encoding transcripts decreased from LSIL to CxCa. Whereas the qualitative presence of the spliced transcript E1C was already a highly specific marker for severe lesions, the prevalence of E6*I, E6*II, E1^E4 and L1 alone could not discriminate the lesion grade.

Based on the upregulation of HPV16 transcripts E6*II and/or E1C and the downregulation of E1^E4 and/or L1 transcripts in HPV16- transformed cells[146], two HPV16 RNA patterns were defined, each comprising a ratio of two quantified spliced HPV16 transcripts: The amount of E6*II normalised to the amount of E1^E4 (pattern 1) and the amount of E1C normalised to the amount of L1 (pattern 2). Although both E6*II and E6*I could be used for discriminating between mild and severe lesions, E6*II was upregulated more strongly in CxCa than E6*I and included in pattern 1.

The computation of ratios of two viral transcripts allowed normalising for variable quantities of HPV-infected cells in cervical smears, since it has to be kept in mind that by far not every cell in a cervical smear is HPV-infected. On average, a LBC slide contains 23,000 cells (range 1,000 to 100,000). Among NIL/M specimens, on average 0.2 HPV-positive cells per slide (1 case with a total of 6 HPV-positive cells), in LSIL approximately 127 (range 0-1,000; 0.56%) and in HSIL 450 (range 0-6,250; 1.95%) has been described[147]. Although the computation of ratios can compensate the variable number of HPV-infected cells in cervical smears, a high analytical sensitivity but also the presence of HPV-infected cells in the cervical smear is necessary in order to reduce the number of false-negative diagnosis.

In a pilot study comprising 77 cytologically diagnosed HPV16 DNA-positive samples, the two ratios of the four marker transcripts were determined. In total, 100% of CxCa (7 out of 7) and 67% of HSIL (16 out of 24) samples were correctly classified as severe and 74% of LSIL (17 out of 23) and 92% of NIL/M (21 out of 23) samples as mild lesions[87] (Figure 1.8). The specificity of the HPV16 RNA patterns for discriminating severe and mild cytological lesions was 83%, the specificity was 74%.

However, the pilot study was limited by small sample numbers defined by cytology only and the labour, time and cost intensity of singleplex NASBA assays not suitable for routine diagnostic test. In addition, the quantification of transcripts worked only in ten-fold steps over four to five magnitudes and was at best semi-quantitative. The slope of the transcript standard curves differed, being almost linear for E1^E4, E6*I and L1 but polynomal for E1C complicating the accurate quantification (Figure A8.1). Reproducibility of quantifying viral RNA in HPV16 DNA-positive cell lines was only moderate showing the highest variation for quantification of E1C[148].

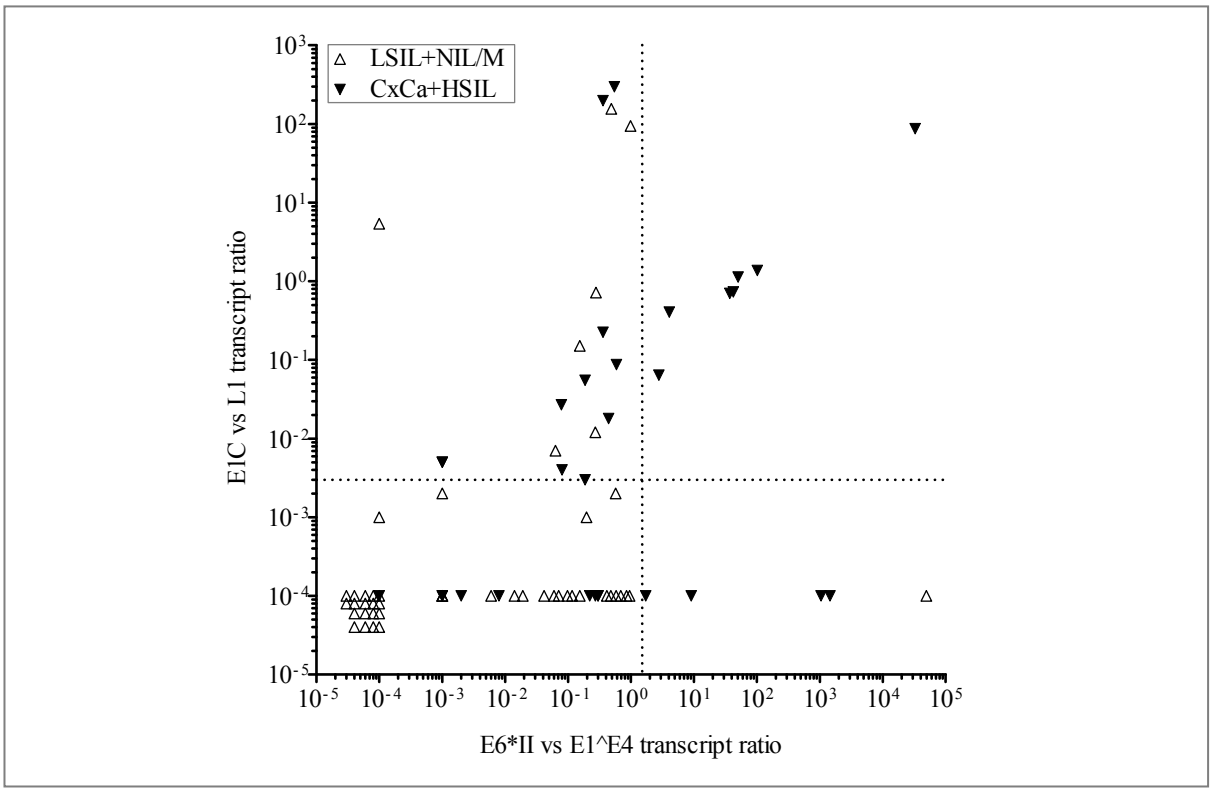


Figure 1.8. **Distribution of HPV16 RNA patterns in exfoliated cervical cells of women with HPV16 DNA-positive mild (NIL/M and LSIL, open triangles) and severe lesions (HSIL and CxCa, filled triangles).** Pattern 1 (x-axis) and pattern 2 (y-axis) (data of ref [87], kindly provided by Dr. Markus Schmitt (DKFZ)). Cutoff (1.5 for E6*II/E1^E4 ratio and 0.003 for the E1C/L1 ratio) is shown as dotted line. If E1C and L1 were positive in NASBA, an E1C/L1 ratio was calculated; if one transcript was negative, its value was set to 0.001 to calculate the E1C/L1 ratio; if both transcripts were negative, the E1C/L1 ratio was set below 0.0001. If E6*II and E1^E4 were positive in NASBA, an E6*II/E1^E4 ratio was calculated; if one transcript was negative, its value was set to 0.001 to calculate the E6*II/E1^E4 ratio; if both transcripts were negative, the E6*II/E1^E4 ratio was set to 0.0001.

1.8 Aim of the thesis

The association of hrHPV infection and cervical cancer development has been unequivocally shown. This is particularly true for HPV16 causing about 55% of CxCa cases. While HPV DNA tests exhibit excellent sensitivity and a very high negative predictive value (NPV), they show deficiencies in specificity and PPV for advanced lesions. As a consequence, they lead to overtreatment of patients, and may not be suited for stand-alone solutions in screening programs. The impetus for new screening technologies in the developed world is predominantly driven by the need to increase the PPV and reduce overmanagement of low-grade and often transient abnormalities. Detection of HPV16 RNA

patterns has been identified as a novel biomarker with potential to sensitively and specifically differentiate mild from severe cervical precancerous lesions. However, RNA patterns detection has been based on laborious, cost-intensive singleplex NASBA reactions and had been validated with a small number and only cytologically defined specimens.

The aim of this thesis was the design, development and validation of multiplex reverse transcriptase quantitative PCR (RT-qPCR) for the detection of the spliced HPV16 transcripts (E6*II, E1^E4, E1C, L1) and the cellular housekeeping gene ubiquitin C (ubC). This project was part of a collaborative research agreement with Roche Molecular Systems, Inc.. With the novel RT-qPCR an important step towards high-throughput analyses of large studies and diagnostic applications would be done. Furthermore, the clinical sensitivity and specificity of HPV16 RNA patterns should be determined more accurately using a large number of histologically defined cervical samples.

Another aim was provide the basis to diagnostic HPV RNA patterns analysis to lesions caused by the other prevalent hrHPV types 18, 31, 33, 35, 45, 52 and 58.

2 Materials

2.1 Reagents

2-(N-Morpholino) ethanesulfonic acid (MES)	Sigma-Aldrich, Steinheim
20% Sarcosyl (N-Lauroylsarcosine) RNase-free	Sigma-Aldrich, Steinheim
5-Bromo-4-chloro-3-indoyl-β-D-glactopyranoside (X-gal)	Carl Roth, Karlsruhe
Agarose	GIBCO Life Technologies, Paisley (Scotland)
Ampicillin	Roche, Mannheim
Bacto-Agar	Becton Dickinson, Sparks (MD, USA)
DMEM	Sigma, Deisenhofen
EDTA-Disodium, RNase-free	AppliChem, Darmstadt
Ethanol (EtOH) (abs.)	Riedel-de Haën, Seelze
Ethidium bromide (EtBr)	Sigma-Aldrich, Steinheim
Fetal Bovine Serum (FCS)	Invitrogen, Carlsbad (CA, USA)
Glycerol 100%, waterfree	Carl Roth, Karlsruhe
Isopropanol	J.T Baker, Deventer (The Netherlands)
N-(3-dimethylaminopropyl)- N-ethylcarbodiimide (EDC)	Pierce, Thermo, Rockford (IL, USA)
Natriumhypochloride (NaClO)	Carl Roth, Karlsruhe
Phenol/Chloroform	Carl Roth, Karlsruhe
PreserveCytTM solution	Cytyc Corp., Boxborough (MA, USA)
RNA ladder „low range“	NEB, Frankfurt am Main
RNA, MS2	Roche, Applied Science, Mannheim
RNase away spray	Molecular Bioproducts, San Diego (CA, USA)
Smart ladder	NEB, Frankfurt am Main
Sodium acetate (NaAc)	J.T. Baker, Deventer (The Netherlands)
Sodium dodecyl sulfate (SDS)	Serva Electrophoresis, Heidelberg
Streptavidin-R-Phycoerythrin (Strep-PE)	Molecular Probes, Eugene (OR, USA)
TE buffer, RNase-free, 1x (10 mM Tris-HCl, 1 mM EDTA, pH 8.0)	Acros Organics, Geel (Belgium)
Tetramethylammonium chloride (TMAC)	Sigma-Aldrich, Steinheim
Tris-HCl, pH 8.0, RNase-free, 1 M	Jena Bioscience, Jena

Trypsin-EDTA (0.25%)	Life-Technologies, Karlsruhe
Tween20®, RNase-free	Serva Electrophoresis, Heidelberg
Water, DNase/RNase-free	Invitrogen, Carlsbad (CA, USA)
xMAP^{TM} Sheath Fluid	Luminex Corp., Austin (TX, USA)
Xylene cyanol	Merck, Darmstadt

2.2 General buffers and solutions

10 x NTP stock solution	mixture of 2 mM of each NTP
ATP, TTP, CTP, GTP in H2O	storage at -20°C
50 x Electrophoresis-buffer (TAE)	2 M Tris, pH 7.8
	0.25 M NaAc (water free)
	0.05 M EDTA
	H_2O ad 8 L
	storage at RT
50% Glycerol	25 mL glycerol (100%)
	25 mL H2O
	autoclave
DNA-loading buffer	0.02% xylene cyanol,
	40% (w/v) sucrose in H2O
	storage at -20°C
Ethidium bromide stock solution	10 mg/mL in H_2O
	storage at 4°C in the dark
LB-Agar	LB medium
	1.5% (w/v) Bacto-Agar
	autoclave , storage at 4°C
LB medium	10 g (1% (w/v)) Bacto-Trypton
	5 g (0.5% (w/v)) Bacto-yeast extract
	10 g (1% (w/v)) NaCl
	H_2O bidest ad 1 L
	adjust to pH 7.5 with NaOH (5 M)
	autoclaved, storage at 4°C
PBS (Phosphate buffered saline),	124 mM NaCl
1 x, pH 7.4	22 mM Na_2HPO_4
storage at RT	10 mM KH_2PO_4

2.3 Luminex buffers

Detection solution Luminex DNA/RNA	2 M TMAC 75 mM Tris-HCl, pH 8.0 6 mM EDTA, pH 8.0 1.5% (w/v) Sarcosyl
DNA hybridisation solution	0.15 M TMAC 75 mM Tris-HCl, pH 8.0 6 mM EDTA, pH 8.0 1.5% (w/v) Sarcosyl
Hybridisation wash buffer Luminex	0.02% Tween 1 x PBS, pH 7.4
0.1 M MES coupling buffer	4.88 g MES H_2O ad 250 mL adjust to pH 4.5 with NaOH (5 M)
Wash buffer I (coupling)	50 µL Tween H_2O ad 250 mL
Wash buffer II (coupling)	2.5 mL of SDS (10%) H_2O ad 250 mL

2.4 Enzymes and inhibitors

Proteinase K	Sigma-Aldrich, Steinheim
Restriction enzyme: EcoRV	Fermentas, St. Leon-Rot
RiboLock RNase Inhibitor	Fermentas, St. Leon-Rot
T7 RNA Polymerase	Fermentas, St. Leon-Rot

2.5 Commercial kits and reagents

DH5α competent cells	Invitrogen, Carlsbad (CA, USA)
Gelextraction QIAquick Kit	QIAGEN, Hilden
High Pure PCR Template Preparation Kit	Roche Applied Science, Mannheim
Light Cycler 480 RNA	Roche Applied Science, Mannheim
Master Hydrolysis Probe Kit	

MagNA Pure 96 Cellular RNA Large Volume Kit	Roche Applied Science, Mannheim
MagNA Pure 96 DNA and viral NA Large Volume Kit	Roche Applied Science, Mannheim
Mini-preparation QIAprep®Spin Miniprep Kit	QIAGEN, Hilden
Pure-Link FFPE Total RNA Isolation Kit	Invitrogen, Carlsbad (CA, USA)
Qiagen Multiplex PCR Kit	QIAGEN, Hilden
QIAsymphony DSP Virus/Pathogen Kit	QIAGEN, Hilden
RNAeasy MinElute Cleanup Kit	QIAGEN, Hilden
StrataClone Blunt PCR Cloning Kit	Agilent Technologies, Santa Clara (CA, USA)

2.6 Commercial plasmid vector

Bluescript M13-KS vector	Agilent Technologies, Santa Clara (CA, USA)
pSC-B-amp/kan vector	Agilent Technologies, Santa Clara (CA, USA)

2.7 Consumables

Tubes, graduated, flat-base, 5ml	QIAGEN, Hilden
96 PCR plate	Sorenson, Bioscience Inc., Utah, USA
Cell culture flask (75, 150 cm)	Greiner, Frickenhausen
Conductive Filtered Tips (50 µL), QIAgility	QIAGEN, Hilden
Disposable protective coats-Foliodress	Steinbrenner Laborsysteme, Wiesenbach
Falcon-tubes (15 mL)	TPP®, Trasadingen (Switzerland)
Filter sterile (0.2 µm)	Renner, Dannstadt
Filter tips 0.5-10 µL	nerbe plus GmbH, Winsen
Rainin filter tips LTS, 1,000 µL	Steinbrenner Laborsysteme, Wiesenbach
Filter tips: Extended 0.1 to 10 µL (S1120-3810) 2 to 20 µL (S1120-1810) 20 to 200 µL (S1120-8810) 200 to 1000 µL (S1126-7810)	star lab GmbH, Ahrensburg

Filter tips, 5,000 µL	Eppendorf, Hamburg
LightCycler®480 Multiwell Plate 96, white	Roche Applied Science, Mannheim
LightCycler®480 Sealing Foil	Roche Applied Science, Mannheim
Litter plastic bags	Greiner Bio-one, Frickenhausen
Low-retention LR tubes	G. Kisker GbR, Steinfurt
Multiscreen 96-well wash plates	Millipore, Bedford (MA, USA)
Petri dishes (Ø 8,5 cm)	Greiner bio-one, Frickenhausen
Pipette tips for motor pipette	Biozym, Hessisch Oldendorf
Polypropylen protective sheet	HJ-Bioanalytik GmbH, Mönchenglattbach
Reagent Reservoir 4870 (50 mL)	Corning Incorporated, Corning (NY, USA)
SeroMap™ beads	Luminex Corp., Austin (TX, USA)
Syringe (50 mL)	Terumo, Leuven, Belgium

2.8 Laboratory devices

8-channel pipette 0.5–10 µL	Brand, Roskilde (Denmark)
8-channel pipette 20-200 µL	Brand, Roskilde (Denmark)
8-channel-motor pipette Precision®	Biozym Diagnostik, Hessisch Oldendorf
Agarose gel electrophoresis chamber including gel tray and combs	Renner, Dannstadt
Agilent 2100 bioanalyzer	Agilent, Santa Clara (CA, USA)
Bunsen burner	Buddeberg, Mannheim
Centrifuge (Minifuge GL)	Heraeus-Christ, Osterode
Cobas z480	Roche, Applied Science, Mannheim
Drigalski spatula	Buddeberg, Mannheim
Eppendorf Research pipette 5,000µL	Eppendorf, Hamburg
Freezer -20°C	Liebherr, Bulle (Schweiz) und Bosch
Freezer -80°C	Forma Scientific (OH, USA)
Fridge 4°C	Liebherr, Bulle (Schweiz) und Bosch
Gilson pipettes (2 µL, 20 µL, 200 µL, 1,000 µL)	Gilson-Abimed, Düsseldorf
Heating block Thermomixer compact	Eppendorf, Hamburg
Ice machine (automatic ice maschines AF30)	Scotsman, Mailand (Italy)
Image Master ® VDS	Pharmacia Biotech, Freiburg
Laboratory Oven (120°C)	Bachofer, Reutlingen

Light Microscope	Leica Microsystems, Wetzlar
Luminex 100 Analyser	Luminex Corp., Austin (TX, USA)
MagNA Pure 96 System	Roche Applied Science, Mannheim
Mastercycler Eppendorf	Eppendorf, Hamburg
Mega Centrifuge 40R	Heraeus-Christ, Osterode
Micro wave	Mikromat AEG, Nürnberg
Mini-centrifuge	Labnet, Woodbridge (NJ, USA)
Mini-centrifuge	Qualitron Inc., Karachi (Pakistan)
NanoDrop 1000	Thermo Scientific, Wilmington (DE,USA)
Neubauer Zählkammer	Carl Roth, Karlsruhe
Pasteur Pipettes	Sigma-Aldrich, Steinheim
pH-Meter	inoLab®, WTW, Weilheim
Pipetboy	Integra Biosciences, Fernwald
QIAgility	QIAGEN, Hilden
QIAsymphony SP instrument	QIAGEN, Hilden
Rainin Multichannel pipette 10-1,200 µL LTS	Steinbrenner Laborsysteme, Wiesenbach
Shaker at 37°C	INFORS AG, Bottmingen (Switzerland)
Shaker at 37°C SM25	Edmund Bühler, Tübingen
Shaker at RT UNIMAX 1010	Heidolph, Schwabach
Table Centrifuge 5415 D	Eppendorf, Hamburg
Table Centrifuge Sigma 2K15	Sigma, Osterode/Harz
Vacuum-wash station	Millipore, Bedford (MA, USA)
Vortex Genie	Bender & Hobein, Zürich (Switzerland)
Vortex R LABIN	Kurt Migge, Heidelberg
Vortex Reax top	Heidolph, Schwabach

2.9 Oligonucleotides

All primers and probes, designed for the HPV16 RNA patterns assay were synthesised by Roche Molecular Diagnostics, Pleasanton, CA and primers and probes, and by Sigma-Aldrich, St.Louis (USA) in HPLC purification quality. Detailed sequence information of primers and TaqMan probes are not presented for reason of intellectual property (IP) protection, but can be made available upon request with appropriate IP protection agreement.

2.10 Cloned plasmids containing spliced HPV16 transcripts

All plasmids containing spliced HPV16 transcripts were kindly provided by Dr. Schmitt (DKFZ) (Table 2.1).

Table 2.1. **Overview of plasmids.**

Name	Accession number	Cloned region from reference genome	Cloning vector
E6*I	HPV16R[a]	83-795	Bluescript M13-KS vector
E6*II	HPV16R	83-795	Bluescript M13-KS vector
E1C	HPV16R	83-4210	Bluescript M13-KS vector
E1^E4	HPV16R	804-4210	Bluescript M13-KS vector
L1	HPV16R	3368-5882	Bluescript M13-KS vector
HPV16 fl	HPV16R	1-7904	Bluescript M13-KS vector

[a] HPV16 sequence and base positions are numbered according to the 1996 sequence database (Los Alamos National Laboratory). http://hpv-web.lanl.gov/COMPENDIUM_PDF/95PDF/1/A1-9.pdf

2.11 Clinical specimens

A convenience sampling strategy was used to include patients in the study. Any patient who visited the routine colposcopy clinic in Bad Münder in the period between October 2010 and March 2012 was eligible to participate. A patient was referred to the routine colposcopy clinic in Bad Münder after suspicious cytological finding by the women´s gynaecologists. A total of 691 exfoliated cervical cell samples were collected and stored in PreservCyt at 4°C for up to 2 years (Table 2.2). The patients' age ranged from 20 to 65 years, with a medium

of 31 years of age. Furthermore, one colposcopy-directed biopsy was taken from each patient. Hematoxylin and eosin (HE) and p16 stained sections from formalin-fixed paraffin-embedded (FFPE) biopsy tissues were read independently by two histologists, one of them, following standard clinical practice, was aware of the patient's HPV DNA status determined in the cytology sample by LA HPV genotyping test (Roche Molecular Diagnostics). In a period of 24 months 114 patients were followed-up at least once after abnormal cytology. The study was enriched by HPV16 DNA-positive PreservCyt cervical cytology samples (stored at 4°C for 7 years before RNA extraction) from CxCa patients (n=27)[12] and NIL/M cases (n=71)[149] from HPV prevalence studies conducted in Ulaanbaatar, Mongolia in 2005 (Table 2.2). From all CxCa patients also FFPE biopsies were available. Fixation and processing of FFPE blocks followed the same protocol. The FFPE biopsies were previously characterised by histology, HPV DNA and ultra-short E6*I mRNA RT-PCR assays (see methods 3.5)[12]. HPV16 E6*I mRNA positive patients (n=27) were included in this study.

Furthermore, 24 FFPE samples from the Barcelona study collected from the worldwide Catalan Institute of oncology (ICO) in nine different countries (Portugal, Peru, Colombia, Lebanon, Paraguay, Venezuela, Philippines, Greece and France) between year 1929 and 2006 were included in this thesis to examine the applicability of HPV16 RNA patterns to these materials (Table 2.2)[4,150]. The FFPE samples had been determined as singly HPV16 RNA-positive by ultra-short E6*I mRNA RT-PCR assays and singly HPV16 DNA-positive by BSGP5+/6+-PCR/MPG (see methods 3.2)[4] (Halec et al., submitted to JNCI[151]).

Additionally, 32 RNA specimens isolated from sections of fresh-frozen oropharyngeal squamous cell carcinomas (OPSCC)[152] obtained from the Ear, Nose and Throat Department of the University Hospital Heidelberg, Germany, collected between 1990 and 2008, were used for a comparison of the established NASBA assays with the newly developed RT-qPCR assays (Table 2.2). All OPSCC had been determined to be HPV16 DNA-positive by the BSGP5+/6+-PCR/MPG. NASBA data of the OPSCC have recently been described[152].

Table 2.2. **Overview of different studies included in this PhD thesis.**

Study	Material	Number	Storage time (in months)
Bad Münder	PreservCyt samples	691	9-24
Mongolia	PreservCyt samples	98	84
Mongolia	FFPE material	27	60
Barcelona	FFPE material	24	48-92
Heidelberg	Fresh-frozen OPSCC samples	32	0

2.12 Ethical clearance

The Mongolian population-based HPV prevalence study was approved by ethical review boards of the Mongolian Health Ministry. All patients gave written informed consent and completed standard WHO questionnaires[149]. The OPSCC study as well as the Bad Münder were approved by the Ethics Committee of the Medical Faculty of the University of Heidelberg, study codes 176/2002[152] and S-004/2010, respectively.

2.13 Computer programs

Luminex 100 IS 2.3 SP1 Software	Luminex Corp., Austin (TX, USA)
Microsoft Office	Microsoft Corp., Unterschleißheim
Microsoft Windows XP	Microsoft Corp., Unterschleißheim
GraphPad Prism® 5	GraphPad Software, Inc. La Jolla, (CA, USA)
SAS 9.3	SAS Institute Inc., Cary (NC, USA)
LightCycler Probe Design software 2.0	Roche Molecular Diagnostics (CA, USA)

2.14 Online software

Blast NCBI	http://www.ncbi.nlm.nih.gov/BLAST/
EMBL-EBI ClustalW2	http://www.ebi.ac.uk/Tools/msa/clustalw2/
FastPCR	http://www.biocenter.helsinki.fi/bi/Programs/fastpcr.htmL
Nucleotide NCBI	http://www.ncbi.nlm.nih.gov/nuccore/

2.15 Services

Multiplexion GmbH	http://www.multiplexion.de/
GATC Biotech	http://www.gatc-biotech.com/de/index.html

3 Methods

3.1 DNA and RNA isolation and quality determination

3.1.1 DNA extraction from exfoliated cells

DNA was released from 800 µL exfoliated cervical cell samples from Bad Münder using the QIAsymphony DSP Virus/Pathogen Midi Kit on the platform of the QIAsymphony SP instrument in combination with the Complex800_V5_DSP protocol. DNA was eluted in 60 µL elution buffer (Qiagen, Hilden, Germany) according to the manufacturer´s instructions. DNA from Mongolian samples was extracted as described previously[149]. DNA from the 32 fresh frozen OPSCC samples had been isolated using Qiagen's QIAamp DNA Mini Kit as described[152].

3.1.2 RNA extraction

For RNA extraction, 10^6 cells per cell line, including CaSki[153], MRI-H196[154] and MRI-H186[154], SiHa[22] and exfoliated cervical cell samples from Bad Münder and Mongolia were subjected to an automated extraction on the MagNA Pure 96 in combination with the RNA LV 2.0 protocol. According to the manufacturer´s recommendations, the *MagNA Pure 96 Cellular RNA Large Volume Kit* was applied with some modifications. Briefly, 4,000 µL and 1,000 µL cervical sample volume, respectively, were centrifuged for 10 min at 4,700 x g. The pellet was resuspended in 200 µL of PreservCyt medium, applied to the MagNA pure 96 and RNA eluted in 50 µL elution volumes and stored at -80°C until use. Similarly, 10^6 cultured cells were resuspended in 200 µL PBS and applied to the instrument. Cell equivalents were computed based on the assumption that each cell contains about 17 pg of total RNA[148]. Additionally, RNA from exfoliated cervical cell samples was extracted with another kit (*MagNA Pure 96 DNA and viral NA Large Volume Kit*) under same conditions as described before.

Total RNA from the 27 FFPE was prepared using the Pure-Link FFPE Total RNA isolation kit as described[12] and stored at -20°C until use. Same extraction procedure was performed with the 24 FFPE samples from the Barcelona study. RNA from the 32 fresh frozen OPSCC samples had been isolated using Qiagen RNeasy Mini kit as described[152].

3.1.3 RNA quality determination

The concentration of all RNA extracts was measured with the NanoDrop 2000. RNA samples were also run on an Agilent 2100 bioanalyzer, which automatically calculates the RNA Integrity Number (RIN). Integrity is no longer determined by the ratio of the ribosomal bands, but the entire electrophoretic profile of the RNA sample. An RNA profile is separated according to the molecular weight and subsequent detection via laser-induced fluorescence in channels of microfabricated chips. The algorithm that underlies the RIN calculation divides the entire RNA profile into nine different regions (pre-, marker-, 5S-, fast-, 18S-, inter-, 28S-, precursor- and post-region) and applies a statistic to these giving a continuous value from 10 down to 1 defining the extend of RNA degradation (10=intact, 1=totally degraded)[155].

3.2 HPV DNA analysis of clinical specimens

The BSGP5+/6+-PCR/MPG assay comprises the BSGP5+/6+-PCR, which homogenously amplifies all known genital HPV types generating biotinylated amplimers of ~150 bp from the L1 region[156] and the MPG assay with bead-based xMAP Luminex suspension array technology, which is able to simultaneously detect 51 HPV types and the ß-globin (bg) gene[31,133,157]. Briefly, amplification was performed using the Multiplex PCR Kit (Qiagen, Hilden, Germany) using 0.2–0.5 µM of each BSGP5+ and 5´-biotinylated BSGP6+ primers and 0.15 µM of each ß-globin primer MS3 and 5´-biotinylated MS10. Following PCR amplification, 10 µL of each reaction mixture is hybridised to bead coupled probes as described before[156]. Results were measured with the Luminex 100 analyser. The median reporter fluorescence intensity (MFI) of at least 100 beads was computed for each bead set in the sample. As described earlier[158], the cutoff value (5 net MFI) to define HPV positivity was applied.

3.3 HPV viral load analysis

Quantification of HPV signals was implemented with the MFI values obtained by the BSGP5+/6+-PCR/MPG assay. In detail, for each positive reaction, the relative HPV MFI

signal (%) was computed by dividing the measured HPV MFI value with the maximum value detected of this HPV type using PCR product where DNA, available in form of colonies, was directly applied. Finally, the relative MFI (%) was divided by the measured ß-globin MFI value to form a non-descriptive viral load value (%HPV MFI/ß-globin MFI). High viral load was assessed for all HPV types by a HPV type-independent high viral load cutoff (0.0007 units) correlating to 0.46 HPV copies per cell, as recently described[133,134].

3.4 NASBA assays

Singleplex NASBA assays detecting E6*II, E1^E4, E1C and L1 HPV16 transcripts in OPSCC samples were carried out as described[31,87,152]. Briefly, purified mRNA was quantified using the principle of competitive NASBA, which is based on the simultaneous co-amplification of calibrator RNA (primer binding sites and amplicon size identical to target RNA, but unique probe binding site), combined with subsequent hybridisation to oligonucleotide probes coupled to Luminex beads.

3.5 Ultra-short E6*I mRNA RT-PCR assays

The ultra-short E6*I mRNA RT-PCR assays designed for the analysis of FFPE samples comprised duplex PCR amplifying the E6*I mRNA of one type plus ubC mRNA, followed by Luminex hybridisation as described previously[12]. In total, the presence of 12 high-risk HPV types and 8 possibly high-risk HPV types can be determined in 20 duplex PCR.

3.6 Identification of splice junctions for seven hrHPV types

So far, splice junctions for E6*I of the seven hrHPV types 18, 31, 33, 35, 45, 52 and 58 were described before[159,160]. Also the E1^E4 splice junction of HPV18 (929^3434), HPV31 (887^3295)[161], HPV33 (894^3351), the E1C splice junction of HPV31 (877^2646)[162] and the L1 splice junction of HPV31 (3832^555)[161] have been identified. For the identification of the remaining unkown HPV16 equivalent splice junctions E1C, E1^E4 and L1, 72 exfoliated cervical cell samples harboring DNA of hrHPV types 18, 31, 33, 35, 45, 52 and

58 were selected from participants of a Mongolian population-based HPV prevalence study. The criterion for sample selection was presence of hrHPV types of interest as single or multiple infections. Of 72 exfoliated samples, 5 harbored HPV18, 9 HPV31, 6 HPV33, 13 HPV35, 13 HPV45, 13 HPV52 and 14 HPV58. One samples harbored combination of the two types of interest. Total RNA was extracted using the RNeasy Mini Kit (Qiagen, Hilden, Germany). DNase treatment (RNase-Free DNase Set, Qiagen, Hilden, Germany) was performed to remove potentially co-extracted DNA. Using Fast PCR programme, type specific primers (Sigma-Aldrich, Hamburg, Germany) were designed for amplification of E1C, E1^E4 and L1 spliced transcripts. For the E1^E4 amplification of HPV58, 52, 35 and 45 primers were re-designed two times as well as for the L1 amplification of HPV35, 33 and 18 primers. Presence of spliced transcripts was verified by 2% agarose gel electrophoresis. Amplimers with expected size (271-400 bp for spliced transcripts) were excised and extracted using the QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany). Cloning followed the protocol of the StrataClone Blunt PCR Cloning Kit (Stratagene, Agilent Technologies, California, USA). Cloned fragments were sequenced at GATC (Konstanz, Germany). HPV types were identified by BLAST search.

3.7 Agarose gel electrophoresis

DNA/RNA fragments can be separated by size in an agarose gel using gel electrophoresis and become visible under UV light (260 nm) after ethidium bromide staining. Depending on the DNA/RNA size to be examined, agarose concentrations differed between 1% for separation of 1-10 kb fragments and 2% for analysis of small PCR products and in vitro transcribed RNA.

For a 2% gel, 3 g of agarose and 150 mL of 1 x TAE electrophoresis buffer were briefly boiled in a microwave oven, supplemented with 15 µL of 10 mg/mL ethidium bromide stock solution and poured into a gel chamber to solidify at room temperature. PCR products and digested plasmids were supplemented with 1/5 volume of 6 x DNA loading buffer and separated electrophoretically at 90 to 120 V in 1 x TAE buffer. The Smart-Ladder DNA marker was used to determine size and concentration of DNA fragments in agarose gels. For RNA analyses, RNA was supplemented with equal quantities of 2 x RNA-loading buffer.

The RNA ladder “low range” marker was used for size determination. Digital photographs were taken using Image Master VDS.

3.8 Design of RT-qPCR primers and probes

The aim was to design specific primers and probes binding all isolates from HPV16. Sequences available at public nucleotide database of the NCBI homepage were aligned using ClustalW2. Primers were selected flanking the HPV16 splice junctions 226^409 (coding potential E6*I), E6*II, E1C, E1^E4 and L1 and, as an internal RNA quality control, the spliced ubiquitin C (ubC) transcript (5455^6267) analogously to recently described NASBA primers[31,87]. To be splice-site specific, TaqMan probes were positioned covering the splice junctions (Figure 3.1). Splice junctions of HPV16 have already been published[7,14,163]. Probes were labelled with different fluorochromes to enable multiplex detection of PCR products. To ensure the specificity of primers and splice-site specificity of probes, prior to practice, an *in silico* analysis was done. For this purpose, BLAST at the NCBI homepage was performed for the comparison of the amplicons with homologous sequences. In order to exclude unspecific amplification of other HPV types, at most nine consecutive nucleotides within the primer and probe region was tolerated. All primers and probes were assessed for their annealing temperature by FastPCR. Primers and probes match the following criteria: Melting temperature of primers should be around 50°C–60°C and the probe should be approximately 10°C higher, amplicon length should be ideally maximal 150 bp and as few degenerate nucleotides as possible should be present. All primers and probes were tested in single- and multiplex quantitative reverse transcriptase PCR (RT-qPCR).

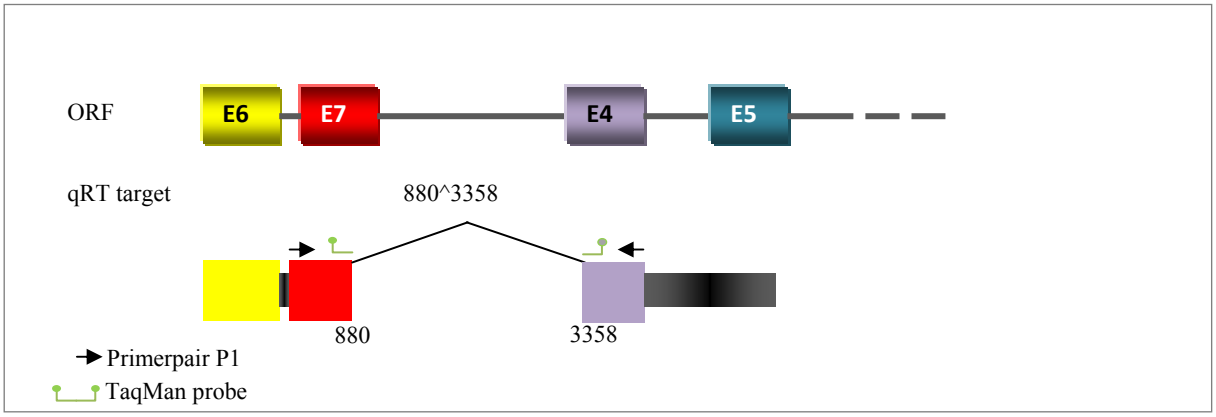


Figure 3.1. **Position of primers and TaqMan probes.** While primers for the detection of spliced transcripts bound in close proximity to a splice junction (black arrows), exemplarily shown for the 880^3358 transcript, TaqMan probes directly covered the splice junction (green lines).

3.9 Preparation of RT-qPCR controls

3.9.1 Production of *in vitro* transcripts

HPV16 i*n vitro* transcripts were generated from Bluescript M13-KS vector (Stratagene) containing spliced E6*I, E6*II, E1^E4, E1C, L1 and ubC cDNA, respectively, in 2007 by Dr. Markus Schmitt as described earlier[31,87]. HPV18 *in vitro* transcripts were generated from pSC-B-amp/kan vector (Stratagene) containing spliced E6*I, E1^E4, E1C or L1. Absence of residual DNA was confirmed by DNA-directed PCR (see chapter 3.2). The viral copy number per unit mass was calculated by assuming that 1 bp weighs about 340 Da. Assuming an *in vitro* transcript size of 2,000 bp, this corresponds to a mass of 9.8 x 10^{-10} ng per transcript molecule. Knowledge of the concentration of the purified RNA preparations allowed computing the number of *in vitro* transcripts per µL.

3.9.2 Cervical cancer cell lines

Cell lines CaSki, MRI-H196, MRI-H186 and SiHa were cultured as monolayers in cell culture flasks. In order to quantify cell numbers prior to RNA isolation, DMEM medium was aspirated. Cells were washed with PBS. PBS was removed and cells were treated with 2 mL of trypsin/EDTA for 2 to 5 min. After cells detached from the flask, 8 mL of medium containing 10% FCS was added to inactivate trypsin. Subsequently, a coverslip wetted with

exhaled breath was slid over the Neubauer counting chamber back and forth using slight pressure until Newton's refraction rings appeared. Under sterile conditions one drop was added to one side of the chamber viewed under a light microscope using x 20 magnification[31]. The number of visible living cells was counted in all four main squares (each comprising 16 small squares). The cell number in the suspension was calculated as follows:

- Total number of counted cells of all four main chambers divided by two
- multiplied by 10,000 (correction factor) and by 10.0 (volume of cell suspension).

Each cell line was authenticated using Multiplex Cell Authentication by Multiplexion (Heidelberg, Germany)[164]. The SNP profiles matched known profiles. The purity of cell lines was validated using the Multiplex cell Contamination Test by Multiplexion (Heidelberg, Germany)[165]. No *Mycoplasma*, SMRV or interspecies contamination was detected.

3.9.3 Negative controls

As PCR negative controls, DNA extracted from human placenta and genomic HPV16 DNA[69] were included. Human placenta DNA had been extracted as recently described[31] and was provided by Dr. Markus Schmitt (DKFZ).

3.10 Quantitative reverse transcriptase PCR (RT-qPCR)

3.10.1 Principle of quantitative PCR

In the quantitative PCR, signals were expressed as crossing point (Cp) values that indicate the cycle at which the increase of fluorescence is highest. The TaqMan system used in the newly developed RT-qPCR contains probes labelled with a fluorescence reporter and a fluorescence quencher, in close proximity to each other (Figure 3.2). When the probe is intact, the quencher dye is close enough to the reporter dye to suppress the reporter fluorescent signal. During PCR, the 5´nuclease activity of the polymerase cleaves the hydrolysis probe, separating the reporter and quencher. In the cleaved probe, the reporter is

no longer quenched and can emit a fluorescence signal when excited. There is a correlation between Cp and concentration: the higher the concentration of target RNA in starting material, the sooner a significant increase in fluorescent signal will be observed, yielding in a lower Cp (Figure 3.2). This correlation between amount of template and value of Cp facilitates quantitative analysis.

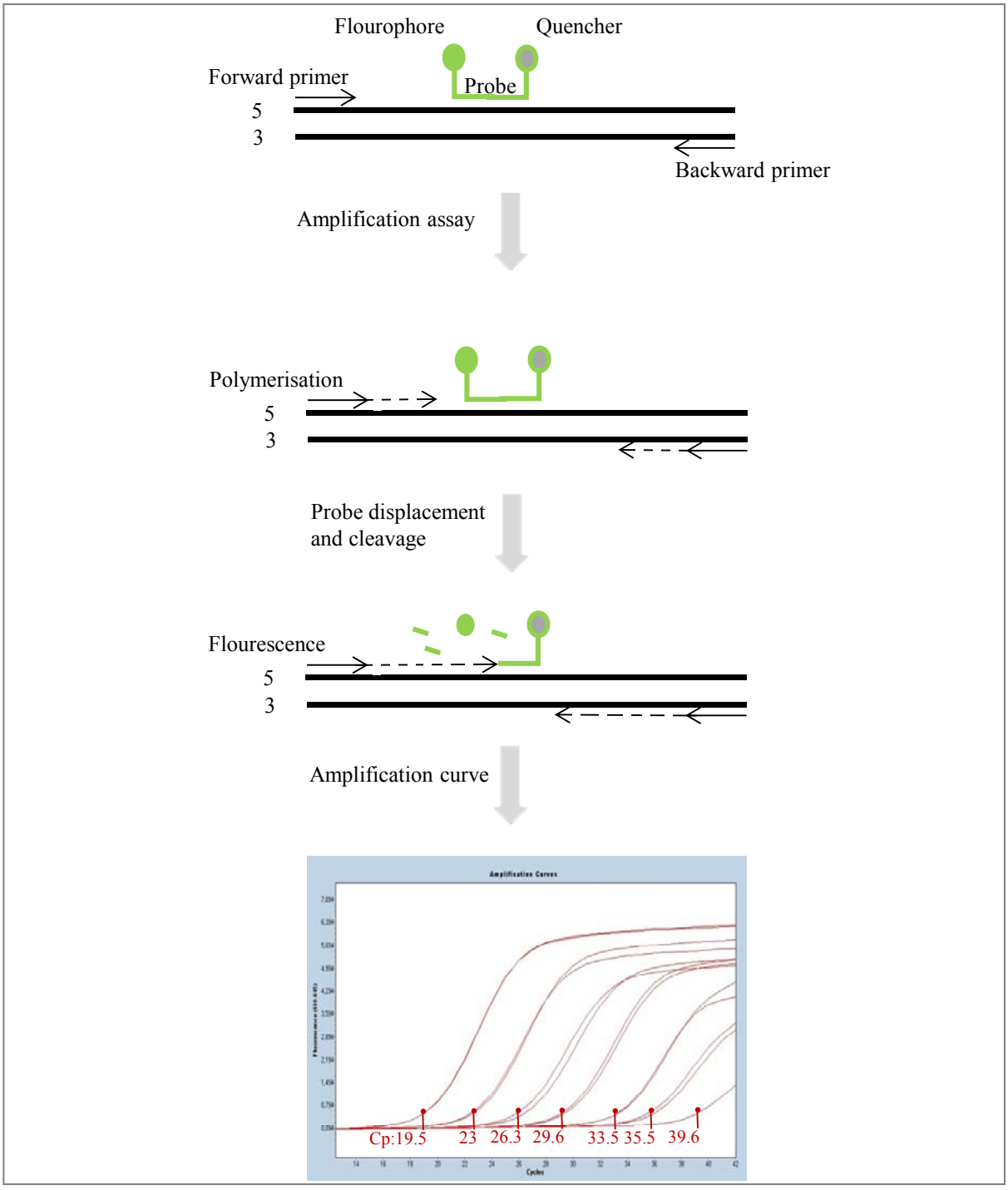


Figure 3.2. **Schematic principle of quantitative PCR using TaqMan probes and the sigmoidal profile of fluorescence accumulation across cycles during amplification.**

3.10.2 Singleplex RT-qPCR

Mastermix of the *Light Cycler 480 RNA Master Hydrolysis Probes Kit* was prepared in a laminar flow hood equipped with UV light. In order to reduce contamination, all pipettes, reagents and machines were tidied up before. The singleplex RT-qPCR were performed in a final volume of 20 µL comprising 1x Light Cycler®480 RNA Master Hydrolysis Probes (Roche Applied Science, Germany), 1x Enhancer, 3.25 mM $Mn(OAc)_2$ and the novel primers and TaqMan probes. The primer and probe concentration was chosen as depicted in table 3.1 ranging between 0.25 and 0.5 µM. In the singleplex RT-qPCR, one to two in case of E1C primer pairs and one TaqMan probe were applied per mastermix. The mastermix and 1 µL of purified RNA or *in vitro* transcripts were pipetted by QIAgility (Qiagen). In all runs, tubes that contained all PCR components but without template RNA were used to ensure that the reagents were free of contamination. Every RNA sample was tested in duplicates.

Because PCR is so sensitive, strong laboratory discipline and specialised conditions to prevent contamination of samples by PCR products or sample-to sample carry-over were required. Physically separated hoods and machines for serial dilution preparation, PCR preparation, and PCR analysis were used. During the whole procedure sterile disposable plates and tubes, pipette filter tips and gloves had to be used.

The PCR plate were transferred to Cobas z480 (Roche Molecular Diagnostics, Pleasanton), which was located in a separated laboratory. A 3 min reverse transcription step at 63°C was followed by a 30 s denaturation step at 95°C. The amplification was performed in 50 cycles including a denaturation step at 95°C for 10 s, an annealing step at 60°C for 30 s and an extension step at 72°C for 1 s. Cooling was performed at 40°C for 10 s.

Table 3.1. **Primer and probe concentrations**.

Transcript	HPV type	Primer and probe name	Primer concentration in SP[a] [µM]	Probe concentration in SP [µM]
E6*I	16	526S, bw1_409, p1_409	0.5	0.25
		fw_526, bw2_409, p1_490	0.5	0.25
E6*II	16	fw_526, bw1_526, p_526m	0.5	0.25
		526S, 526A, p_526m	0.5	0.25
E1C	16	2582A, 2582S, 2582F, 2582R2, p2_2582[b]	0.5	0.35
		2582A, 2582S, p1_2582	0.5	0.35
		2582R, 2582S, p2_2582	0.45	0.35
		fw2_2582,bw1_2582,p2_2582	0.5	0.35
E1C_sj[c]	16	S8_2582,2582R2,p3_2582	0.5	0.25
		S6_2582,2582R2,p3_2582	0.5	0.35
E1^E4	16	2582S, bw1_3358, p1_3358	0.5	0.25
		2582S, bw2_3358, p1_3358	0.5	0.25
		fw3_3358, 3358R, p1_3358	0.5	0.25
L1	16	fw3_5639, bw3_5639, p_5639	0.5	0.35
		fw2_5639, bw3_5639, p3_5639	0.5	0.35
		fw2_5639, bw4_5639, p3_5639	0.5	0.35
ubC		fw2_ubC, bw2_ubC, p2_LubC	0.5	0.25
		fw2_ubC, bw1_ubC, p2_LubC	0.5	0.25
		fw5_ubC, ubCR, p5_LubC	0.5	0.25
E6*I	18	fw1_e6I, bw2_e6I, HPV18_short	0.25	0.35
		f_e6I, bw2_e6I, HPV18_short	0.125	0.35
E1C	18	S_e1c, A_e1c, p_e1c	0.5	0.35
		S_e1c, R_e1c, p_e1c	0.5	0.35
E1^E4	18	fw1_e1^e4, E1E4R, p_e1e4	0.25	0.35
		f_e1^e4, E1E4A, p_e1e4	0.125	0.35
L1	18	S_18, A_18, p_L1_18	0.5	0.35
		S_18, R_18, p_L1_18	0.5	0.35

[a] singleplex (SP), [b] four primers amplifying E1C (seminested), [c] forward primer covers splice junction

3.10.3 Changed parameters for E1C singleplex RT-qPCR$_{E1C}$

The HPV16 singleplex E1C RT-qPCR (RT-qPCR$_{E1C}$) using 0.5µM of the primers 2582A, 2582S, 2582F and 2582R2 and 0.35µM of p2_2582 probe applied to *Light Cycler 480 RNA Master Hydrolysis Probes Kit* was performed as singleplex RT-qPCR assays with some

modifications: 2 µL purified RNA were used and 3.0 mM $Mn(OAc)_2$; the reverse transcription time was increased to 5 min and the annealing time to 40 s.

3.10.4 Fiveplex RT-qPCR

The procedure was performed as for the singleplex RT-qPCR. The single difference was that only one mastermix of the *Light Cycler 480 RNA Master Hydrolysis Probes Kit* was prepared, including all primers and probes. The fiveplex RT-qPCR was optimised with varying concentrations of the following reagents: primers and probes, $Mn(OAc)_2$ and an enhancer. Primers and probes were tested in concentration ranging from 0.125 µM to 0.5 µM (Table 3.2), $Mn(OAc)_2$ was titrated ranging from 3.0 mM to 3.5 mM and enhancer was included and excluded, respectively.

Table 3.2. **Concentration of primers and probes in fiveplex RT-qPCR.**

Transcript	Primer and probe name	Concentration combination 1 [µM]		Concentration combination 2 [µM]		Concentration combination 3 [µM]	
		Primer	Probe	Primer	Probe	Primer	Probe
E6*I	526S, bw2_409, p1_409	0.25	0.35	0.25	0.35	0.125	0.35
E1C	2582R, 2582S, p2_2582	0.5	0.35	0.313	0.35	0.313	0.35
E1^E4	2582S, bw1_3358, p1_3358	0.25	0.35	0.25	0.35	0.19	0.35
L1	fw3_5639, bw3_5639, p_5639	0.25	0.35	0.5	0.35	0.5	0.35
ubC	fw2_ubC, bw2_ubC, p2_LubC	0.25	0.25	0.25	0.25	0.125	0.25

Furthermore, the fiveplex RT-qPCR was also performed with mastermix of *HawkZ Kit* in a final volume of 20 µL comprising 1x HawkZ05 Fast Master Mix (Roche Applied Science, Germany) and 1.5 mM $Mn(OAc)_2$ and the novel primers and TaqMan probes concentration combination 2 (Table 3.2). PCR run protocol with the Cobas z480 (Roche Molecular Diagnostics, Pleasanton) were adapted to the *HawkZ Kit*: A 5 min reverse transcription step at 63°C was followed by a 5 s denaturation step at 92°C. The amplification was performed in 50 cycles including a denaturation step at 95°C for 5 s, an annealing step at 60°C for 40 s and an extension step at 72°C for 1 s.

3.10.5 Colour compensation

To compensate for spectral overlap of the probes, a colour compensation template was generated according to the LightCycler 480 Instrument Operator's Manual: three replicates of spliced transcripts were applied to the singleplex RT-qPCR as well as one blank RT-

qPCR without primers, probes and template. The temperature profile used mimics the typical PCR as described in 3.10.2.

3.10.6 Triplex RT-qPCR

The triplex RT-qPCR (RT-qPCR$_{TP}$) was performed according to singleplex RT-qPCR and quantified the high-abundance viral transcripts E6*I, E1^E4 and ubC. The RT-qPCR$_{TP}$ was performed in a final reaction volume of 20 µL comprising 1x Light Cycler®480 RNA Master Hydrolysis Probes (Roche Applied Science, Germany), 1x Enhancer, 3.0 mM $Mn(OAc)_2$, 1 µL of purified RNA or *in vitro* transcript, and optimised with different primer and probe concentrations ranging from 0.05 to 0.125 µM of each primer and 0.25 to 0.35 µM of each probe (Table 3.3).

Table 3.3. **Different primer and probe concentration in RT-qPCR$_{TP}$.**

Transcript	Primer and probe name	Concentration combination 1 [µM]		Concentration combination 2 [µM]		Concentration combination 3 [µM]	
		Primer	Probe	Primer	Probe	Primer	Probe
E6*I	526S, bw2_409, p1_409	0.125	0.35	0.2	0.35	0.125	0.35
E1^E4	2582S, bw1_3358, p1_3358	0.125	0.35	0.125	0.35	0.125	0.35
ubC	fw2_ubC, bw2_ubC, p2_LubC	0.1	0.25	0.05	0.25	0.05	0.25

3.10.7 Duplex RT-qPCR

The duplex RT-qPCR (RT-qPCR$_{DP}$) quantified the low-abundance E1C and L1 transcripts. The RT-qPCR$_{DP}$ was performed using *Light Cycler 480 RNA Master Hydrolysis Probes Kit* (Roche Applied Science, Germany). The procedure of mastermix preparation and PCR program was performed as for the RT-qPCR$_{TP}$ with the single difference of used primers and probes (Table 3.4).

Table 3.4. **Different primer and probe concentration in RT-qPCR$_{DP}$.**

Transcript	Primer and probe name	Concentration [µM]	
		Primer	Probe
L1	fw3_5639, bw3_5639, p_5639	0.25	0.35
E1C	2582A, 2582S, p1_2582	0.5	0.35

3.11 Standard curves and quantification of transcripts

Standard curves for all transcripts were obtained by amplification of a dilution series of 10^6 to 10^1 copies of *in vitro* transcripts in the background of 20 ng/µL MS2 RNA (Roche Applied Science, Germany). Copy numbers were plotted against the measured Cp values. Absolute quantification of copy numbers was achieved by linear regression analysis comparing the Cp value of the unknown sample against the standard.

3.12 Definition and statistics

PCR efficiency (E), a quantitative expression of the quality of the PCR process, was calculated from the slope of the standard curves according to the equation E=10^[-1/slope][166]. The highest quality PCR run with an efficiency of 2, meaning that the number of target molecules doubles with every PCR cycle. The correlation between Cp and the log of transcript copy number per PCR was described by the coefficient of determination, denoted R^2 (1.00 indicated perfect correlation). The reproducibility was described by the coefficient of variation (CV). Statistical analyses were performed with the SAS software, version 9.3 (SAS Institute). Pearson correlation analysis was run on comparison of transcript quantification by NASBA and RT-qPCR$_{TP}$ and RT-qPCR$_{E1C}$ and on comparison of transcripts in repeated RNA extraction. Pairwise comparison of ubC, E6*I, E1^E4 and E1C copy numbers were done by Wilcoxon-signed-rank test. Comparisons of the HPV16 RNA patterns in all histological or cytological grades were performed by the non-parametric Mann-Whitney test. The agreement of histological or cytological examinations and HPV16 RNA patterns was monitored by kappa statistic (κ), where a value of 1 represents complete agreement, 0 represents no agreement. All tests were two-sided, and a p value <0.05 was considered significant. Where appropriate, 95% confidence intervals (CI) were computed. To calculate the difference of RNA quality in fresh-frozen MRI-H196 and in archived exfoliated cervical cell samples, the median ubC RNA copies per cell detected in the archived and in the fresh-frozen samples were divided.

A clinical RNA sample was classified as valid if >80 ubC or >200 E6*I copies per PCR were detected. UbC as the longest amplicon with 238 bp is most susceptible to RNA

degradation and thus used for monitoring validity. The classification included additionally the E6*I copies in order to increase number of valid samples.

RNA samples were applied in duplicates to the RT-qPCR$_{TP}$ and RT-qPCR$_{E1C}$ and scored as positive if both duplicates showed a cutoff Cp value below 42. E6*I/E1^E4 ratios were calculated as: mean E6*I copies per PCR over mean E1^E4 copies per PCR. If E6*I or E1^E4 was negative, value was set to 1 copy per PCR and E6*I/E1^E4 ratio calculated. With the positive controls run on every PCR plate, a correction factor for the E6*I/E1^E4 ratio was calculated and used in order to reduce plate differences. The cutoff$_{E6*I/E1^E4}$ was determined to be 0.095 units for cervical RNA samples (see chapter 4.1.7.7) and 45 units for OPSCC RNA samples. An OPSCC sample with the E6*I/E1^E4 ratio and E1C below cutoff was defined as CxCa-like RNA patterns-negative OPSCC (CxCaRNA$^-$). Tumours with either the E6*I/E1^E4 ratio or E1C above cutoff were defined as CxCa-like RNA patterns-positive OPSCC (CxCaRNA$^+$)[152].

Histological diagnosis were summarised in severe (≥CIN3) and mild (NIL/M, suspicious cytology but normal histology (CIN0), CIN1: ≤CIN1) lesions, respectively. CIN2 was categorised as intermediate and not classified within the severe lesions.

An exfoliated cervical cell sample was classified as mild lesion if the E6*I/E1^E4 ratio and E1C were below cutoff and classified as severe lesion if at least one of the E6*I/E1^E4 ratio or E1C was above cutoff. Clinical samples negative for the 3 viral transcripts but positive for ubC were automatically scored as mild lesion.

4 Results

HPV16 RNA patterns (see 1.7) so far detected by singleplex NASBA assays may have the potential to replace current cervical cancer screening tests such as cytological examinations and HPV DNA-based tests (see 1.5). Large studies to confirm HPV16 RNA patterns as possibly clinically useful biomarker are still needed, but singleplex NASBA assays quantify only 10-fold differences, are laborious and cost-intensive and not applicable as routine screening test with high sample numbers.

The aim of this thesis was to transfer the singleplex NASBA assays to multiplex RT-qPCR and to validate the HPV16 RNA patterns with a large number of exfoliated cervical cell samples in comparison to cytology and/or histology (Figure 4.1).

For the development of RT-qPCR assays, primers and probes were designed to detect sensitively and specifically the viral transcripts E6*II, E1^E4, E1C and L1, as well as the housekeeping transcript ubC. In addition, a RT-qPCR for HPV16 E6*I was co-developed. This was necessary since other hrHPV types do not express E6*II but only E6*I encoding transcripts. With regard to extending the HPV16 RNA patterns to other hrHPV types, it could be beneficial to use uniform target sequences. Furthermore, essential PCR parameters were optimised comprising combinations and concentrations of primers and TaqMan probes, PCR annealing and reverse transcription times, and the amount of purified RNA applied per reaction. Moreover, the maximal multiplexing degree of single RT-qPCR was explored. In addition, the most efficient RNA isolation method as well as the reproducibility of RNA isolation and RT-qPCR were determined.

In a further step, the newly developed RT-qPCR assays were compared to the previously developed singleplex NASBA assays. Additionally, the HPV16 RNA patterns were validated using 360 HPV16 DNA-positive exfoliated cervical cell samples and compared to cytology and histology as well as to viral load for predicting the presence of cervical lesions.

Finally, the applicability of RT-qPCR in FFPE biopsies was explored. As a first step towards extending the HPV16 RNA patterns analyses to other prevalent hrHPV types -id est (i.e.) HPV 18, 31, 33, 35, 45, 52 and 58- HPV16-like splice junctions were identified for these seven additional hrHPV types.

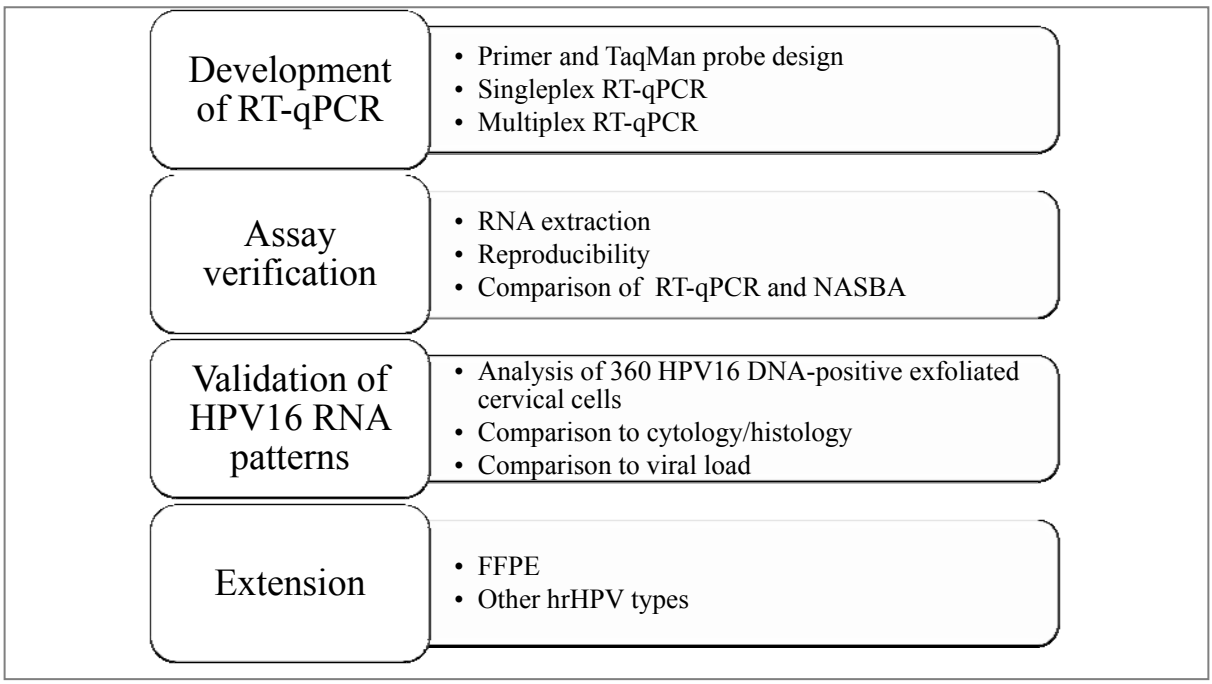


Figure 4.1. **Structure of result sections**.

4.1 HPV16 RNA patterns

4.1.1 Design of primers and TaqMan probes

RT-qPCR uses different enzymatic conditions than NASBA and consequently requires different primers and probes including different length and melting temperatures. All primers and probes with the potential to sensitively and specifically detect the spliced HPV16 transcripts E6*I, E6*II, E1C, E1^E4 and L1 as well as the housekeeping transcript ubC had to be systematically re-designed.

For each target sequence several forward (fw) and backward (bw) primers were designed (Table 4.1). Primers were between 18 and 26 nucleotides long with theoretical annealing temperatures of 51.8°C - 60.5°C as calculated by the FastPCR program. Final amplicon sizes varied between 108 and 238 bp.

Primers were designed to flank the splice junctions of the six transcripts and TaqMan probes to cover the splice junctions with one exception: For the E1C transcript the same splice donor is used as for the E1^E4 and thus the two transcripts compete for the same forward

primer. Therefore an alternative strategy was explored with the E1C forward primers -rather than the probes- spanning the splice junction.

Primers and probes were located within sequence regions conserved among all HPV16 isolates available in public Nucleotide data bases at the NCBI homepage. To exclude HPV16 primers and probes binding to sequences of other HPV types only a maximum of eight consecutive nucleotide matches with non-HPV16 sequences was allowed.

Table 4.1. **Overview of primers and probes tested for the quantification of HPV16 and ubC.**

Transcript splice site	Coding potential	Accession number	fw[a] primers	bw[b] primers	N of theoretical combinations	N of combinations tested[c]	TaqMan probes[d]	Fluorochrome	Excitation and emission filters
226^409	E6*I	FJ610152.1	3	2	6	2	2	JA270	610-645
226^526	E6*II	FJ429103.1	3	5	15	12	3	JA270	610-645
880^2582	E1C	HM057182.1	5	4	20	8	3	FAM	495-521
880^2582	E1C_sj[e]	HM057182.1	8	1	8	8	2	FAM	495-521
880^3358	E1^E4	K02718.1	2	4	8	10	3	HEX	540-580
3632^5639	L1	FJ610152.1	4	4	16	7	3	Coumarin	435-470
5455^6267	ubC	NG_027722.1	7	7	49	27	10	Cy 5.5	680-700

[a] number (N) of forward (fw) primers designed, [b] number of backward (bw) primers designed, [c] tested in singleplex RT-qPCR, [d] number of TaqMan probes designed, [e] primers covering splice junction (sj)

TaqMan probes were between 26 and 33 nucleotides long with theoretical melting temperatures of 56.1°C-63.9°C. For later multiplexing, they contained a transcript-specific fluorescent reporter dye attached to the 5´-end and a BHQ-2 quencher at the 3´-end (Table 4.1).

As this thesis was performed in a collaborative research agreement with Roche Molecular Systems, Inc., detailed sequence information of primers and TaqMan probes are not presented for reason of intellectual property (IP) protection, but can be made available upon request with appropriate IP protection agreement.

In total, 32 fw primer, 27 bw primer and 26 TaqMan probe sequences were selected and synthesised. In the next chapter the functional characteristics of primers and probes established in singleplex RT-qPCR are described.

4.1.2 Singleplex RT-qPCR

Primers and TaqMan probes were tested in singleplex PCR. *In vitro* transcripts, generated from Bluescript M13-KS vector containing spliced E6*I, E6*II, E1^E4, E1C, L1 and ubC cDNA, respectively, were added to the RT-qPCR mix. At first, detection limits (DL) of different primer and TaqMan probe combinations were determined using ten-fold and/or three-fold dilution series of the *in vitro* transcripts. The specificity of primers and TaqMan probes was assessed with all *in vitro* transcripts as well as full-length HPV16 DNA and human placenta DNA as templates. Water was used as negative control template.

4.1.2.1 Transcript detection limits in singleplex RT-qPCR

Serial dilutions of the *in vitro* transcripts E6*I, E6*II, E1C, E1^E4, L1 and ubC with copies ranging from 2 to 10^6 copies per µL were applied in duplicates to singleplex RT-qPCR using different combinations of primers and TaqMan probes. For each transcript up to three combinations are exemplarily shown in table 4.2.

For the E6*I, both primer and TaqMan probe combinations detected 10 copies per PCR with a median Cp value of 36 (Table 4.2). The coefficient of determination (R^2) obtained by both E6*I combinations indicated a perfect linear correlation between crossing point (Cp) and the log of transcript copy number per PCR. The PCR efficiency was slightly better for combination 2. One million E6*I copies were detected at cycle 21 using combination 1 and at cycle 20 using combination 2. The same analyses are depicted also for the other HPV16 transcripts and for ubC in table 4.2. The E1C singleplex RT-qPCR combination 1 also included two primer pairs in order to increase the detection limit (refer to 5.3.2 for discussion).

Table 4.2. **Singleplex detection limit for spliced transcripts.**

Transcript	Primer and probe name	Combi-nation	DL[a] (copies per PCR)	Median Cp at DL	R^{2} [b]	E[c]	Median Cp at 10^6 copies per PCR	Used in further experiments [d]
E6*I	526S, bw1_409, p1_409	1	10	36	0.99	2.20	21	no
	526S, bw2_409, p1_490	2	10	36	1.00	2.09	20	yes
E6*II	fw_526, bw1_526, p_526m	1	1,000	33	0.95	1.98	23	no
	526S, 526A, p_526m	2	100	38	1.00	1.77	22	no
E1C	2582A, 2582S, 2582F, 2582R2, p2_2582	1	2	40	0.98	2.02	20	yes
	2582A, 2582S, p1_2582	2	10	38	0.99	2.14	23	no
	2582R, 2582S, p2_2582	3	10	41	0.99	2.18	35	no
	fw2_2582,bw1_2582,p2_2582	4	10,000	41	0.99	2.18	35	no
E1C_sj[e]	S8_2582,2582R2,p3_2582	1	6	37	0.99	2.07	20	yes
	S6_2582,2582R2,p3_2582	2	10	38	0.98	1.95	21	no
E1^E4	2582S, bw1_3358, p1_3358	1	10	40	1.00	2.08	24	yes
	2582S, bw2_3358, p1_3358	2	1,000	35	0.97	2.70	27	no
	fw3_3358, 3358R, p1_3358	3	10,000	36	0.96	4.60	34	no
L1	fw3_5639, bw3_5639, p_5639	1	10	40	1.00	2.07	21	yes
	fw2_5639, bw3_5639, p3_5639	2	100	43	1.00	1.78	27	no
	fw2_5639, bw4_5639, p3_5639	3	1,000	39	0.95	1.69	26	no
ubC	fw2_ubC, bw2_ubC, p2_LubC	1	10	38	1.00	2.07	21	yes
	fw2_ubC, bw1_ubC, p2_LubC	2	100	32	1.00	1.78	27	no
	fw5_ubC, ubCR, p5_LubC	3	100	40	0.95	1.69	26	no

[a] detection limit (DL) defined as lowest copy number tested with signals above background in duplicates, [b] coefficient of determination (R^2) describes the correlation between crossing point (Cp) and the log of transcript copy number per PCR, [c] PCR efficiency (E); under optimal conditions PCR run with an efficiency of 2, meaning that the number of target molecules doubles with every PCR cycle, [d] the combination with lowest DL but at most 10 copies per PCR was selected for further experiments; in case both had same DL, the one with E closer to 2 was chosen, [e] primers covering splice junction (sj)

E6*II singleplex RT-qPCR lacked good detection limits in all 12 tested combinations and was replaced by E6*I RT-qPCR in the subsequent developments.

After testing multiple primer and TaqMan probe combinations, a sensitive quantification of all spliced transcripts was achieved with DL ranging from 2 to 10 copies per PCR.

4.1.2.2 Specificity of singleplex RT-qPCR

Specificity of singleplex RT-qPCR is mediated by (A) HPV16 type-specific primers and (B) by splice-site and type-specific TaqMan probes. In addition, specific detection of splice-

sites further provides an advantage because it cannot be affected by co-amplified contaminating DNA.

HPV16 transcript specificity of the RT-qPCR was evaluated using the *in vitro* transcripts E6*I, E1^E4, E1C, L1 and ubC. Genomic HPV16 DNA with copy numbers ranging from 10^4 to 10^6 copies per µL and human placenta DNA with concentrations ranging from 0.1 to 100 ng/µL were used to demonstrate RNA specificity of the RT-qPCR and to exclude cross-hybridisation with unspliced HPV16 sequences and cellular background DNA, respectively. Templates were applied in duplicates to singleplex RT-qPCR using finally selected primer and TaqMan probe combinations determined to be most sensitive (see 4.1.2.1, primer and probe combination 2 for E6*I and combination 1 for the other transcripts).

With one exception, no unspecific PCR products were detected for the transcripts as well as for 10^6 HPV16 genome copies and 100 ng/µL human placenta DNA (corresponding to about 1.5 x 10^4 cellular genome equivalents). The exception was the E1C_sj RT-qPCR which detected 10^5 genomic HPV16 DNA copies with a Cp of 36 (Figure 4.2, d). Even by shortening the splice junction covering forward E1C_sj primer by six nucleotides the unspecific amplification of genomic DNA remained (data not shown). Consequently, the E1C_sj primer and TaqMan probe combination was not used in future experiments.

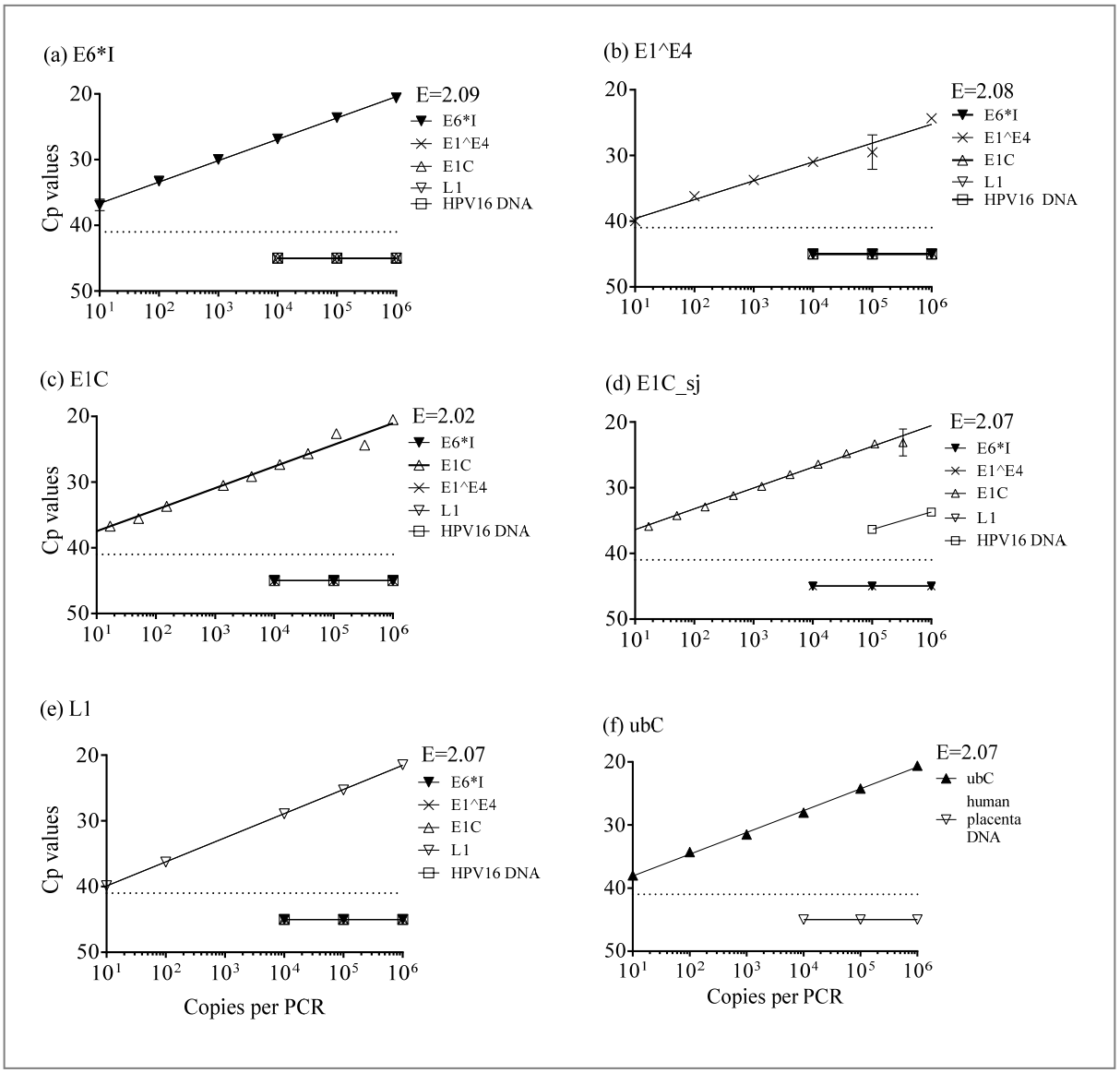


Figure 4.2. **Specificity of singleplex RT-qPCR with finally selected primer and probe combinations.** Dilution series of *in vitro* transcripts, HPV16 DNA and human placenta DNA were used as templates in E6*I (a), E1^E4 (b), E1C (c), E1C_sj (d), L1 (e) and ubC (f) RT-qPCR using the finally selected primer and probe combinations. All templates were applied in duplicates, error bars show standard deviation. Most duplicates were nearly identical and error bars too small to be visible. Negative results, i.e. no detectable RT-qPCR signal within 42 cycles, are indicated as Cp 45 below the dotted Cp 42 line. E values are indicated.

None of the other RT-qPCR gave rise to unspecific signals proving specificity of the assays.

4.1.3 Fiveplex RT-qPCR

Multiplexing with simultaneous presence of all primers and probes in one reaction can potentially result in higher detection limits for the transcripts, for example through dimer

formations of primers and probes. Furthermore, the presence of all transcripts in one reaction can lead to interference in template amplification and product detection through competition for PCR reagents. At least theoretically multiplexing could also result in reduced specificity due to new amplification products from combinations of primers designed for different targets.

To achieve simultaneous sensitive and specific detection of the four HPV16 transcripts and the ubC transcript in a fiveplex RT-qPCR, variations of primer and probe concentrations, PCR reagents including various concentrations of $Mn(OAc)_2$ and the presence/absence of an enhancer were tested with *in vitro* transcripts and with RNA from HPV16-transformed cervical cancer cell lines. The salt concentrations ($Mn(OAc)_2$) can affect the annealing temperature and the enhancer, supplied in the *Light Cycler 480 RNA Master Hydrolysis Probe Kit* (*LC480*) (Roche, Mannheim), can improve the amplification of difficult, e.g. GC-rich targets and increase the fluorescence intensity when using low RNA concentrations.

The fiveplex reaction contained five probes each labelled with a different fluorochrome (Table 4.1). As the wavelengths of light emitted by the different fluorescent dyes can overlap and thus can cause one detector channel to pick up signals from more than one fluorochrome, a colour compensation template was generated according to the Light Cycler 480 Instrument Operator's Manual.

4.1.3.1 Variation of primer and probe concentration for fiveplex RT-qPCR

Three different concentrations of chosen primer and probe combinations, determined as most sensitive in singleplex RT-qPCR (Table 3.2), were titrated. The final primer concentration was titrated between 0.125 and 0.5 µM, the probe concentrations between 0.35 and 0.25 µM and the concentration of $Mn(OAc)_2$ between 3.25 and 3.5 mM. The enhancer was either in- or excluded. Under all these various conditions, DL was determined applying *in vitro* transcripts E6*I, E1C, E1^E4, L1 and ubC ranging from 10^1 to 10^6 copies per PCR as single template.

The DL for each of the five *in vitro* transcripts ranged between 10^1 and 10^3 (Figure 4.3). Most sensitive transcript detection ranging from 10^1 to 10^2 was achieved with primer and probe concentration 2: The R^2 varied between 0.99 and 1 indicating a perfect linear

correlation of the data points with the standard curves. PCR efficiency ranged from 1.93 (E6*I) to 2.32 (E1C).

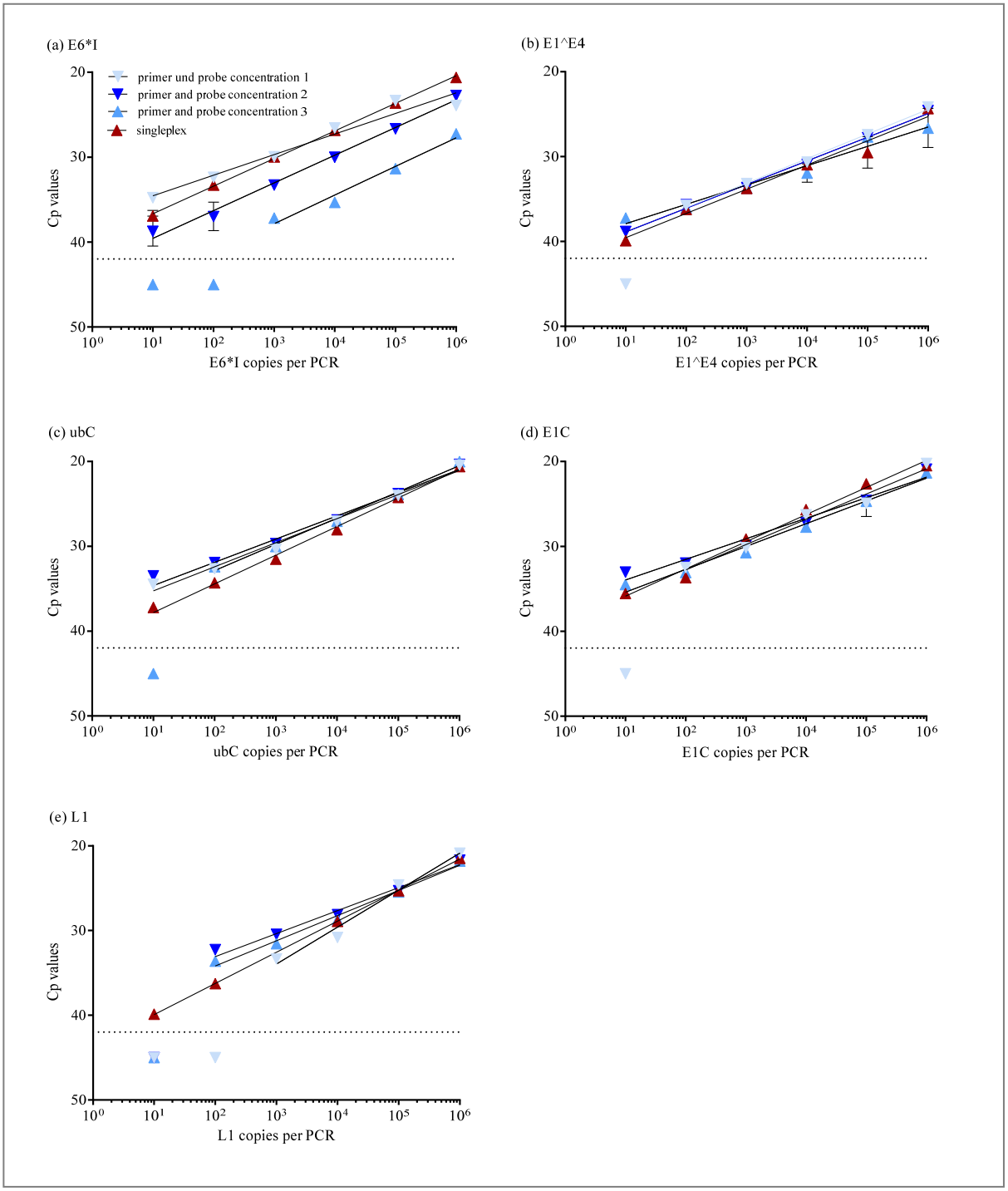


Figure 4.3. **Effect of primer and probe concentrations on Cp values in fiveplex RT-qPCR.** Standard titration curves of E6*I (a), E1^E4 (b), ubC (c), E1C (d) and L1 (e) *in* vitro transcripts with three different primer/probe concentrations (light, middle and dark blue triangles), $Mn(OAc)_2$ concentration of 3.0 mM and the inclusion of enhancer. Data of singleplex RT-qPCR from figure 4.2 are given for comparison (red triangles). All templates were applied in duplicates, error bars show standard error of the mean. Most

duplicates were nearly identical and error bars too small to be visible. Negative results, i.e. no detectable RT-qPCR signal within 42 cycles, are indicated as Cp 45 below the dotted Cp 42 line.

Attempts to improve the DL of the RT-qPCR with primer and probe concentration 2, by the exclusion of the enhancer or by changing the $Mn(OAc)_2$ concentration to 3.25 or 3.5 mM were unsuccessful and resulted in DL for the five *in vitro* transcripts between 10^2 and 10^3, 10^2 and 10^3 or 10^2 and 10^4, respectively (data not shown).

4.1.3.2 Co-amplification of transcripts in HPV16-positive cell lines

The next step was to test the co-amplification of all transcripts present in HPV16-transformed cervical cancer cell lines in the five singleplex and the fiveplex RT-qPCR using the identified sensitive conditions, namely primer and probe concentration 2, a $Mn(OAc)_2$ concentration of 3.0 mM and enhancer included. Total RNA (33 ng, corresponding to approximately 2,200 cells) extracted from cell lines MRI-H186 and CaSki was used. Both cell lines had also been analysed previously by NASBA methods[148] and are compared to RT-qPCR results in chapter 4.1.6.1.

The abundant transcripts E6*I, E1^E4 and ubC were efficiently quantified in both cell lines with the fiveplex RT-qPCR setting (Figure 4.4). Statistical analyses to determine differences between fiveplex and singleplex RT-qPCR were not performed due to small number of data points. Compared to singleplex RT-qPCR, similar mean copy numbers per PCR of E6*I, E1^E4 and ubC were identified in both cell lines also by the fiveplex RT-qPCR.

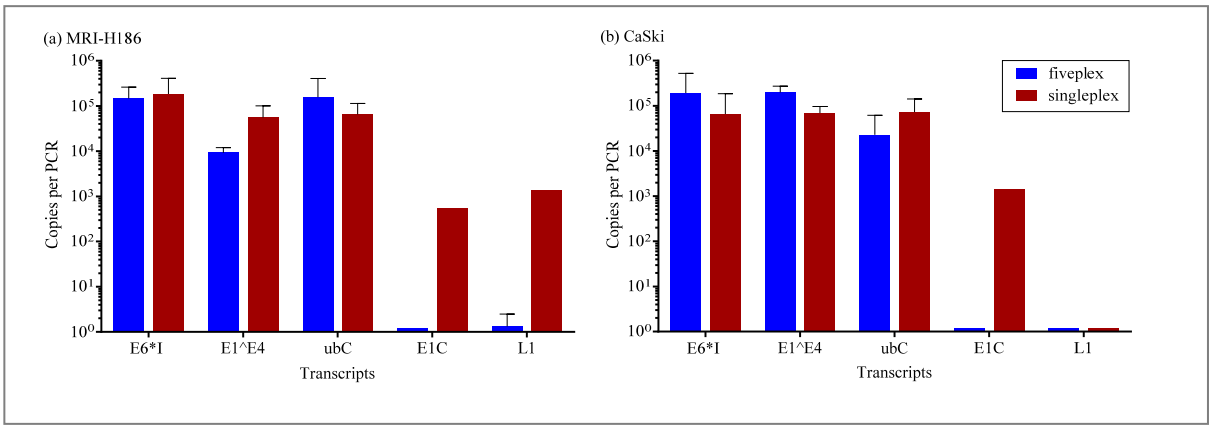


Figure 4.4. **Effect of co-amplification of transcripts in fiveplex RT-qPCR.** Copies per PCR of E6*I, E1^E4, ubC, E1C and L1 in 33 ng RNA from HPV16-tranformed cervical cancer cell lines MRI-H186 (a) and CaSki (b) quantified in fiveplex (blue bars) and singleplex (red bars) RT-qPCR. All templates were analysed in

duplicates and represented as means with standard deviation. If duplicates are nearly identical, error bars cannot be recognised. Negatives were set to 1.2 copies per PCR.

However, sensitivity for the low-abundance transcript E1C (negative in fiveplex versus 500 copies in MRI-H186 and 1,500 in CaSki in singleplex) and L1 (600-fold reduced copies in MRI-H186 fiveplex) was strongly reduced (Figure 4.4). In the fiveplex RT-qPCR the simultaneous amplification of high-abundance transcripts appeared to inhibit concurrent amplification of low-abundance transcripts much more severely than that of other high-abundance transcripts.

Despite the high sensitivity optimised on the basis of *in vitro* transcript dilution series, an accurate detection of all transcripts known to be present in cell lines could not be achieved by the fiveplex RT-qPCR, requiring further protocol development.

4.1.3.3 Variation of fiveplex RT-qPCR protocols

For a better detection of the low-abundance viral transcripts E1C and L1 in the fiveplex RT-qPCR, the *HawkZ Kit* was explored and compared to the standard *Light Cycler 480 RNA Master Hydrolysis Probes Kit* (*LC480 RNA*) used before.

In duplicates, 33 ng RNA extracted from MRI-H186 and CaSki were applied to two fiveplex RT-qPCR containing the *HawkZ* and the *LC480 RNA* reagents, respectively.

In MRI-H186, the mean copy number of detected E6*I, E1^E4 and ubC was up to 35-fold higher using the *LC480 RNA* (Table 4.3). Statistical analyses were not performed due to small number of data points. Likewise, the mean copy number of the quantified E6*I and ubC in CaSki was up to 10-fold higher using the *LC480 RNA* but nearly identical for E1^E4.

The detection of L1 and E1C was not improved in the tested cell lines by using the *HawkZ Kit* with both transcripts remaining negative (Table 4.3).

Table 4.3. **Influence of two different one-step RT kits on the quantification of transcripts in HPV16-transformed cervical cancer cell lines.**

Transcript	Dupli-cate	MRI-H186		CaSki	
		LC480 RNA[a] (copies/PCR)	HawkZ[b] (copies/PCR)	LC480 RNA (copies/PCR)	HawkZ (copies/PCR)
E6*I	1	223,152	10,100	288,214	31,861
	2	202,143	2,015	217,959	17,923
	Mean	**212,648**	**6,058**	**253,086**	**24,892**
E1^E4	1	11,142	5,072	250,094	305,914
	2	10,522	2,482	147,037	115,163
	Mean	**10,832**	**3,777**	**198,566**	**210,538**
ubC	1	443,296	19,206	67,794	5,294
	2	100,559	16,743	8,670	2,716
	Mean	**271,928**	**17,975**	**38,232**	**4,005**
E1C	1	0[c]	0	0	0
	2	0	0	0	0
	Mean	**0**	**0**	**0**	**0**
L1	1	5	0	0	0
	2	1	0	0	0
	Mean	**3**	**0**	**0**	**0**

[a] *Light Cycler 480 RNA Master Hydrolysis Probes Kit* (*LC480 RNA*), [b] *HawkZ Kit*, [c] negative results, i.e. no detectable RT-qPCR signal within 42 cycles

As no L1 and E1C amplification was achieved with neither the *HawkZ* nor the *LC480 RNA* using the fiveplex RT-qPCR, a triplex RT-qPCR (RT-qPCR$_{TP}$) for the high-abundance transcripts E6*I, E1^E4, and ubC and a duplex RT-qPCR (RT-qPCR$_{DP}$) for the low-abundance viral transcripts E1C and L1 were explored using the *LC480 RNA*.

4.1.3.4 Performance of triplex RT-qPCR (RT-qPCR$_{TP}$) for high-abundance transcripts

Three concentrations of primers and TaqMan probes (Table 3.3) were applied to the RT-qPCR$_{TP}$. DL was determined using serial dilutions of *in vitro* transcripts E6*I, E1^E4 and ubC ranging from 10^1 to 10^6 copies per PCR each as single templates and in parallel 33 ng RNA extracted from MRI-H186 and CaSki were applied to quantify all three transcripts in one reaction.

The DL for the three *in vitro* transcripts in all three primer and TaqMan probe concentrations ranged between 10^1 and 10^3 (data not shown). Most sensitive transcript detection with a DL of <10 copies per PCR was achieved for all three targets with primer and probe concentration 3 (Figure 4.5, a). The R^2 varied between 0.98 and 1 indicating a

perfect linear correlation of data points with the standard curves. The PCR efficiency value E ranged from 1.92 (ubC) to 2.03 (E6*I) indicating optimal PCR conditions.

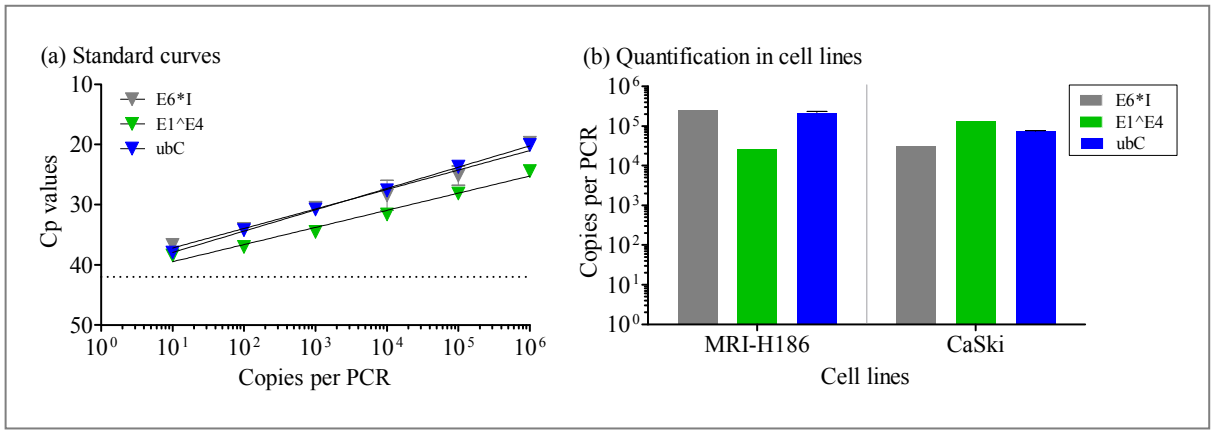


Figure 4.5. **Standard curves and quantification of transcripts in the RT-qPCR$_{TP}$ with the finally selected primer and probe concentration 3.** Standard titration curves of the *in vitro* transcript E6*I (grey), E1^E4 (green) and ubC (blue) (a). E6*I (grey), E1^E4 (green) and ubC copies (blue) quantified in 33 ng of RNA from MRI-H186 and CaSki (b). All templates were applied in duplicates as indicated by the error bars (standard deviation). If duplicates are nearly identical, error bars cannot be recognised.

With the finally selected primer and probe combination the quantity of three highly abundance-transcripts ranged between 25,399 and 291,678 copies in MRI-H186 and CaSki (Figure 4.5, b).

Overall, RT-qPCR$_{TP}$ appeared to be highly sensitive for the simultaneous detection and quantification of the two viral transcripts and of ubC and was used further in the validation of HPV16 RNA patterns in clinical samples (see chapter 4.1.7).

4.1.3.5 Performance of duplex RT-qPCR (RT-qPCR$_{DP}$) for low-abundance transcripts

The performance of duplex RT-qPCR$_{DP}$ (Table 3.4) was analysed using serially diluted *in vitro* transcripts and RNA from cell lines MRI-H186 and MRI-H196. Additionally, in order to examine a potential competition during PCR when E1C and L1 are present at the same time, serial dilutions of each transcript ranging from 10^1 to 10^6 copies per PCR in a background of 10^4 copies per PCR of the other transcript were applied. In these analyses CaSki was replaced by MRI-196 cells because it is known to contain both E1C and L1[148].

The DL for the *in vitro* transcripts E1C and L1 was 10 copies per PCR in the background of MS2RNA as well as when the other viral transcript was present with 10^4 copies per PCR

(Figure 4.6, a). The Cp values at the dilution steps between 10^3 and 10^6 copies per PCR were almost identical in the absence but also presence of co-amplification of the other viral RNA. In contrast, for 10^2 to 10^1 copies per PCR Cp values increased with simultaneous co-amplification of the other viral transcript. At a 100-fold excess of the other viral transcript, 10^1 E1C and L1 copies were only borderline positive (Figure 4.6, a).

In 33 ng of MRI-H186 and MRI-H196 RNA, singleplex RT-qPCR detected in mean 15 and 643 E1C and 961 and 11 L1 copies, respectively. In MRI-H186, RT-qPCR$_{DP}$ detected no E1C, but 914 copies of the L1 (Figure 4.6, b). In MRI-H196, RT-qPCR$_{DP}$ detected 670 EIC transcript copies but no L1. Consequently, if L1 were more frequent than E1C in one sample, E1C detection was reduced and vice versa, though it should be kept in mind that E1C (in MRI-H186) and L1 (in MRI-H196) were only borderline positive in the singleplex RT-qPCR.

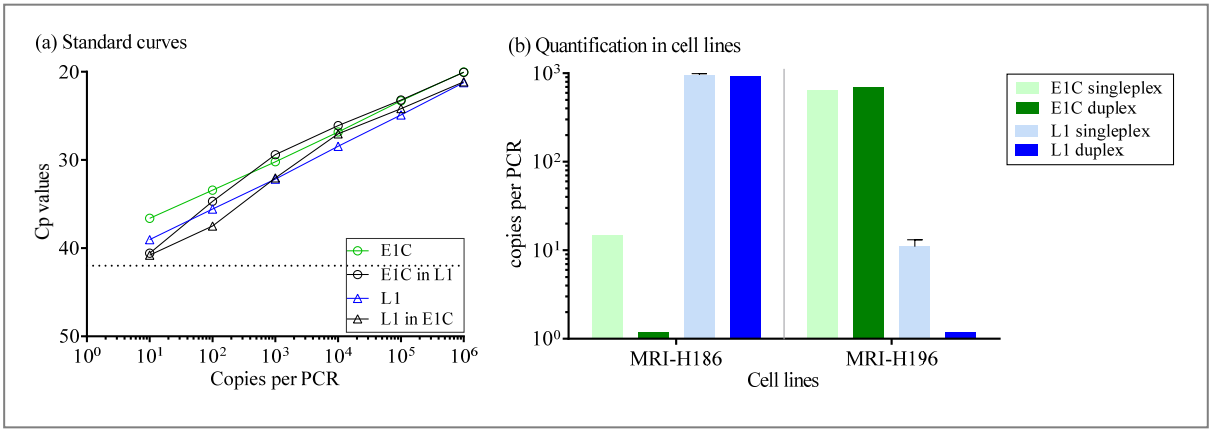


Figure 4.6. **Titration and quantification of transcripts in duplex RT-qPCR$_{DP}$.** (a) Standard titration curves of the *in vitro* transcript E1C and L1 standard curves without (coloured) and with (black) 10^4 copies of the L1 or E1C as background RNA, respectively. Cutoff (Cp=42) is indicated by the dotted line. (b) E1C (green) and L1 (blue) quantified in 33 ng of RNA from MRI-H186 and MRI-H196 in RT-qPCR$_{DP}$ and in singleplex RT-qPCR using same primer and TaqMan probe combination and concentrations. Negative results, i.e. no detectable RT-qPCR signal within 42 cycles, are indicated as 1.2 copies per PCR. All templates were applied in duplicates as indicated by the bars (standard deviation). If duplicates are nearly identical, error bars cannot be recognised.

A reduced detection of E1C and L1 was observed once both transcripts were co-amplified simultaneously and suppressed by the presence of high-abundance transcripts in one sample. As a consequence, reliable detection of these transcripts appeared to be possible in singleplex reactions only.

4.1.3.6 Comparison of RT-qPCR$_{DP}$ and singleplex RT-qPCR$_{E1C}$ in clinical specimens

As mentioned before, E1C is an important component of the HPV16 RNA patterns for the identification of severe lesions. Since it is known to be a low-abundance transcript, detection should be as sensitive as possible. To decide whether E1C detection is more sensitive in singleplex RT-qPCR$_{E1C}$ than in duplex RT-qPCR$_{DP}$, a comparison was performed using clinical specimens.

RNA extracted from HPV16 DNA single-positive exfoliated cervical cell samples defined as NIL/M (n=10), normal histology (CIN0, n=6), CIN1 (n=3), CIN2 (n=24), CIN3 (n=17) and CxCa (n=18) (Chapter 2.11) were analysed by singleplex RT-qPCR$_{E1C}$ (Chapter 3.10.3) and by RT-qPCR$_{DP}$ (Chapter 3.10.7).

Despite its high analytical sensitivity with *in vitro* transcripts (chapter 4.1.3.5), the duplex RT-qPCR$_{DP}$ never simultaneously identified L1 and E1C in any of the cervical cell samples (Figure 4.7). E1C were found only in one CxCa and one NIL/M sample with 174 and 51 copies, respectively. E1C was negative in all CIN0, CIN1, CIN2 but detected in one CIN3 with 28 copies. L1 were not detected in any of the CxCa and of the NIL/M samples. L1 copies ranging from 267 to 1,982 (mean 862) were detected in one CIN1, six CIN2 and two CIN3.

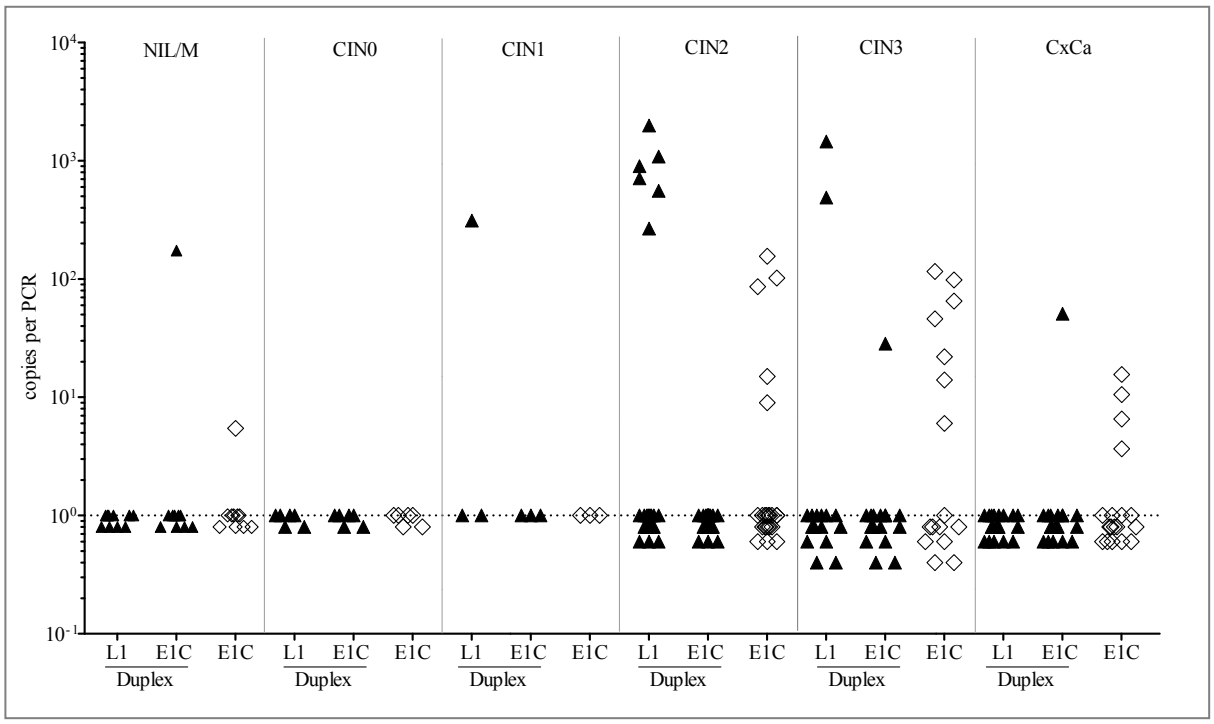


Figure 4.7. **Influence of L1 co-quantification on E1C detection in different cytological or histological stages.** The x-axis compares duplex detection of E1C and L1 (filled triangles) with singleplex E1C detection only (open rhombi). The y-axis shows copies per PCR. All templates were applied in duplicates and the mean Cp value was used for calculation of copies per PCR. Samples were set below 1 copy per PCR if transcripts showed negative RT-qPCR results, i.e. no detectable RT-qPCR signal within 42 cycles.

In singleplex RT-qPCR$_{E1C}$, one CIN2 and two CIN3 being positive with 488 to 1,983 L1 copies in the duplex only, turned E1C-positive which demonstrated the higher sensitivity of the singleplex reaction. Likewise, a significant proportion of CxCa (n=3) and CIN3 (n=6) samples was additionally E1C-positive only by RT-qPCR$_{E1C}$ (Figure 4.7), while at the same time, no additional NIL/M, CIN0 or CIN1 samples tested positive by RT-qPCR$_{E1C}$.

Overall, a low number of E1C copies per PCR was observed leading to <1 copy per cell. However, it should be kept in mind that not every cell in a cervical smear is HPV-infected[147]. The low number of E1C copies in clinical specimens had been already noted in the pilot NSABA study[87]. Nonetheless, to exclude potential errors in the quantification, the *in vitro* transcription of E1C as well as the titration curve of the transcript dilutions was repeated and quantification of the RT-qPCR$_{E1C}$ was confirmed to be correct and reproducible.

In conclusion, an adequate and highly sensitive detection of E1C was possible in singleplex RT-qPCR$_{E1C}$ reactions only.

4.1.3.7 Critical examination of L1 contribution to HPV16 RNA patterns classification

As already introduced in chapter 1.7, the qualitative and quantitative detection of viral transcripts E6*I, E1^E4 and L1 alone cannot discriminate the grade of a cervical lesion: A low number of cancerous cells in a cervical smear, exemplarily, would result in a low number of E6*I copies and consequently in a false-negative classification. The computation of ratios of two viral transcripts allowed normalising for variable quantities of HPV-infected cells in cervical smears assuming that the transcript ratio is similar in the majority of HPV-positive cells in a given sample. Thereby, false classification can be substantially reduced. In contrast to the mentioned transcripts above, the qualitative presence of the spliced transcript E1C is already a highly specific marker for severe lesions as already shown in the pilot study[87]. Assuming that three RT-qPCR for the detection of four viral transcripts are not viable for a routine diagnostic test, the aim of this chapter -based on data from the NASBA pilot study- was to explore whether L1 detection could be omitted from the HPV16 RNA patterns classification.

In the NASBA pilot study, RNA extracted from HPV16-infected DNA-positive exfoliated cervical cell samples defined as NIL/M (n=25), LSIL (n=23), HSIL (n=24) and CxCa (n=7) were analysed by the E1C and L1 singleplex NASBA. If E1C and L1 were positive in NASBA, an E1C/L1 ratio was calculated; if one transcript was negative, its value was set to 0.001 to calculate the E1C/L1 ratio; if both transcripts were negative, the E1C/L1 ratio was set to 0.00001. A cervical sample above the E1C/L1 cutoff of 0.003 was classified as severe lesion[87].

Overall, 1 of 25 NIL/M, 6 of 23 LSIL, 10 of 24 HSIL and 4 of 7 CxCa were defined as severe lesion either due to the detection of E1C alone or due to a high E1C/L1 ratio which resulted from high E1C and low L1 signals (Figure 4.8, blue and red triangles). For these cervical samples, the normalisation of E1C with L1 showed no benefit over E1C detection alone. Moreover, none of the E1C-positive NIL/M or LSIL were negative using the E1C/L1 ratio indicating that a normalisation with L1 did not increase the test specificity. However, three HSIL (12.4%) and one CxCa (14.3%) were above E1C/L1 ratio cutoff and thus classified as severe lesion by the presence of very low L1 copies alone (E1C-negative) that would be missed by the exclusion of L1. However, identifying a high-grade lesion or cancer just by the presence of few L1 copies is questionable.

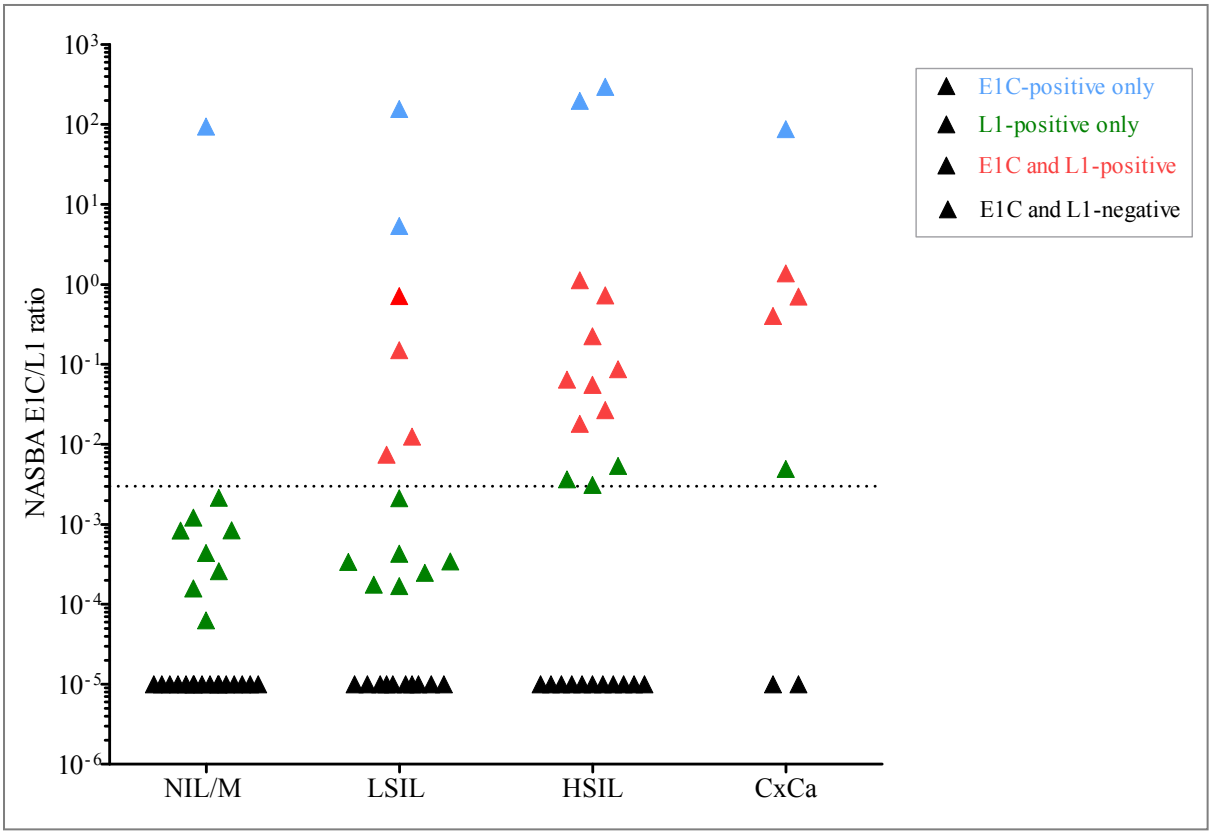


Figure 4.8. **Influence of L1 on HPV16 RNA patterns classification in different cytological stages (x-axis, categories).** The y-axis shows ratio of E1C wildtype (wt)/calibrator (q) divided by L1 wildtype (wt)/calibrator (q) signal. Cervical sample with detection of E1C (blue) and L1 (green) only and with simultaneous detection of E1C and L1 (red) are colour-coded. A sample was set to a ratio of 0.00001 if both transcripts showed negative NASBA results (black). Cutoff is indicated by dotted line.

In summary, in the NASBA pilot study the normalisation of E1C signals with L1 did not increase the test specificity and contributed little to the test sensitivity, justified to exclude L1 from further analyses.

4.1.4 RNA extraction

Automated RNA isolation enables high-throughput and highly reproducible extraction of purified RNA. Despite automation, the choice of RNA extraction kits and extraction volumes can influence the RNA yield. This will be described in the following two chapters.

4.1.4.1 Influence of RNA extraction kits on RNA yield

Two MagNA Pure96 kits were tested to determine which one was best suited for the RNA extraction from clinical specimens.

For RNA extraction, 1,000 µL PreservCyt volume of HPV16 DNA-positive patients from Mongolia cytologically defined as NIL/M (n=3), LSIL (n=4), HSIL (n=4) and CxCa (n=4) were processed with the *MagNA Pure96 DNA and Viral NA Large Volume Kit (Viral NA)* extracting total nucleic acids and with the *MagNA Pure96 Cellular RNA Large Volume Kit (Cell RNA)* that includes a DNaseI treatment extracting total RNA. One µL of RNA was applied to the E6*I and ubC singleplex RT-qPCR.

The mean nucleic acid concentrations among all cervical samples were 34.4 ng/µL (range 4.6-80.2) using *Viral NA* (extraction of DNA and RNA) and 14.2 ng/µL (range 2.9-37.7) using *Cell RNA* (extraction of RNA only) (Table 4.4).

E6*I copies could not be detected in the *Viral NA* extracts. In contrast, the mean E6*I copy number per PCR quantified in the *Cell RNA* extracts was 223 (range 97 to 448). The *Cell RNA* was also superior in extracting amplifiable ubC as depicted by a mean copy number per PCR of 740 (*Viral NA*) versus 2,128 (*Cell RNA*). The reduced number of RNA in the *Viral NA* extracts might be an explanation for the insufficient E6*I detection. Furthermore, a strong variation of ubC was observed in the *Viral NA* extracts ranging between 4 and 4,854 copies. Differences of ubC quantification in comparison to *cell RNA* were highest in the CxCa samples (up to 2,000-times reduced) but also present in HSIL, LSIL and NIL/M samples (median of 2-times, range 1-24). A PCR inhibition by the co-existence of DNA might be an explanation.

Table 4.4. **E6*I and ubC copies per PCR detected in RNA samples extracted by two MagNA Pure96 kits.**

Cytology	Nucleic acid concentration [ng/µL]		E6*I				ubC			
			Mean Cp[a]	copies /PCR	Mean Cp	copies /PCR	Mean Cp	copies/ PCR	Mean Cp	copies/ PCR
	Viral NA[b]	*Cell RNA*[c]	*Viral NA*		*Cell RNA*		*Viral NA*		*Cell RNA*	
CxCa	15.6	4.7	>42	0	33.7	245	36.5	11	29.4	1,528
CxCa	9.2	3.1	>42	0	33.6	284	35.4	24	29.9	1,058
CxCa	15.2	2.9	>42	0	33.8	226	37.0	7	29.3	1,609
CxCa	4.6	6.8	>42	0	33.0	448	37.9	4	27.4	6,681
HSIL	26.4	6.9	>42	0	34.3	157	31.2	504	28.8	2,333
HSIL	14.7	26.9	>42	0	34.2	164	31.6	365	30.3	809
HSIL	11.6	4.0	>42	0	33.9	215	33.7	83	29.1	1,970
HSIL	70.3	28.5	>42	0	33.6	265	28.1	4,854	27.1	8,270
LSIL	37.9	21.7	>42	0	34.0	195	31.3	471	30.0	1,005
LSIL	19.6	8.3	>42	0	34.3	156	31.3	482	29.1	1,948
LSIL	77.2	17.3	>42	0	>42	0	30.2	1,016	29.9	1,039
LSIL	80.2	15.9	>42	0	>42	0	29.5	1,764	29.0	2,105
NIL/M	42.5	37.7	>42	0	34.8	97	32.1	251	31.5	327
NIL/M	65.4	18.2	>42	0	>42	0	30.0	1,228	30.1	893
NIL/M	25.6	9.5	>42	0	>42	0	34.7	38	31.4	347
Mean:	34.4	14.2	>42	0	33.9[d]	223[e]	31.6	740	29.4	2,128

[a] all templates were applied in duplicates and mean Cp values were used for the calculation of copies per PCR, [b] *MagNA Pure96 DNA and Viral NA Large Volume Kit (Viral NA)*, [c] *MagNA Pure96 Cellular RNA Large Volume Kit (Cell RNA)*, [d] >42 were excluded for calculating the mean, [e] 0 copies per PCR were excluded from the calculation of mean

Although the total nucleic acid concentration was higher in the *Viral NA* extracts (in part at least due to co-extracted DNA), quantification of ubC copies per PCR was at least 1.7-times higher and quantification of E6*I only possible using the *Cell RNA*. Consequently, the latter will be used in future analyses.

4.1.4.2 Influence of extraction volume on HPV16 RNA quantification

The amount of PreservCyt volume used for RNA extraction can influence the RNA yield and consequently the quantification of HPV16 RNA.

For RNA extraction, 1,000 µL and 4,000 µL PreservCyt volume of HPV16 DNA-positive cervical samples from Mongolian women cytologically defined as NIL/M (n=44) and CxCa (n=27) were processed with the *MagNA Pure96 Cellular RNA Large Volume Kit*. Elution

volume was constantly 50 µL. RNA was applied to RT-qPCR$_{TP}$ and to RT-qPCR$_{E1C}$. Subsequently, ubC copies, the E6*I/E1^E4 ratio and E1C copies were calculated.

A median of 1,219 and 328 ubC copies were quantified in 4,000 µL and 1,000 µL extracts, respectively ($p<0.0001$, Wilcoxon matched-pairs signed rank test). This was also observed with the mean RNA concentration measured in 16 samples: 18.4 ng/µL in 4,000 µL and 6.6 ng/µL in the 1,000 µL extracts. Likewise, the E6*I ($p<0.0001$, Wilcoxon matched-pairs signed rank test) and E1^E4 ($p=0.031$, Wilcoxon matched-pairs signed rank test) copy numbers were significantly higher in 4,000 µL than in the 1,000 µL extracts. The slopes (m) of the ubC, E6*I and E1^E4 regression lines were between 1.06 and 2.63 confirming higher amount of copy numbers detected in 4,000 µL extracts (Figure 4.9, a-c). Statistical analyses for E1C were not performed due to small number of positive samples (n=9).

The qualitative comparison of E6*I detected in 1,000 µL and in the 4,000 µL extracts showed concordantly 23 E6*I-positive, 35 negative and 13 discordant samples (4 positive in 1,000 µL and 9 in 4,000 µL extracts only) ($\kappa=0.6$, CI.95=0.44-0.81) (Figure 4.9, b). The qualitative comparison of E1^E4 detected in 1,000 µL and in the 4,000 µL extracts showed 17 concordant positive samples, 47 concordant negative and 7 discordant samples resulting in a $\kappa=0.8$ (CI.95=0.59-0.93) (Figure 4.9, c). One of the three E1^E4-positive NIL/M samples positive in 1,000 µL extracts only was also E6*I-positive only. Equally, the one CxCa and one of the three NIL/M positive in 4,000 µL extracts only, were the same samples also E6*I-positive only.

To exclude, that the additional positive samples detected in the 4,000 µL extraction volume only were attributed to a worse reproducibility of RNA extraction or RT-qPCR, this will be focused in the following chapter (4.1.5).

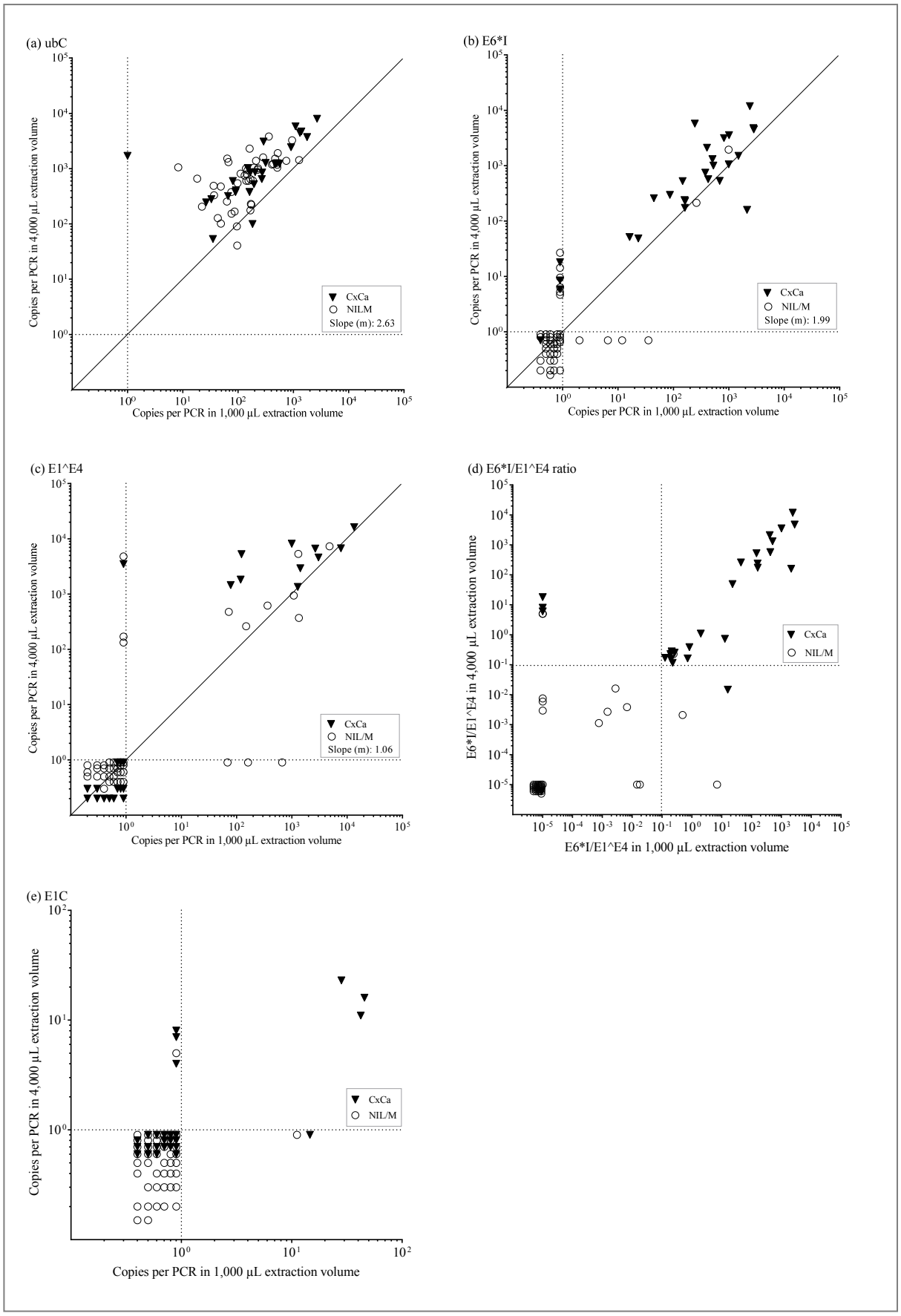


Figure 4.9. **Influence of extraction volume on HPV16 RNA patterns analyses**. UbC (a), E6*I (b), E1^E4 (c), E6*I/E1^E4 ratio (d) and E1C (e) detected in 1,000 µL extraction volume (x-axis) are plotted against

4,000 µL extraction volume (y-axis). The diagonal is indicated in (a-c). Slope (m) of the regression line is depicted, a value of 1 should indicate the same copy number per PCR detected with both extraction volumes. CxCa samples are displayed as filled triangles and NIL/M samples as empty circles. All templates were applied in duplicates and mean Cp value was used for the calculation of copies per PCR. Negative results, i.e. no detectable RT-qPCR signal within 42 cycles, are indicated below 1 copy per PCR under the dotted cutoff line. An E6*I/E1^E4 ratio was set to 0.00001 below the dotted cutoff line if both transcripts showed negative RT-qPCR results.

The concordance of E6*I/E1^E4 classification as negative and positive between 4,000 µL and 1,000 µL extracts was κ=0.8 (CI.95=0.64-0.94) with 22 concordantly positive, 42 concordantly negative and 7 discordant samples (Figure 4.9, d). Three CxCa and one NIL/M E6*I/E1^E4 ratio-positive in 4,000 µL extracts only were same samples E6*I-positive only in 4,000 µL extracts and were borderline ubC-positive in 1,000 µL extracts with less than 100 copies. The one CxCa, E6*I/E1^E4 ratio-positive in 1,000 µL extracts only was the same sample, E1^E4-positive in 4,000 µL extracts only and thus above E6*I/E1^E4 ratio cutoff. The two discordant NIL/M samples were either below 100 ubC copies in 1,000 µL or in the 4,000 µL extracts.

For E1C, three cervical samples were concordantly positive, 63 concordantly negative and six discordant (κ=0.45, CI.95=0.04-0.87) (Figure 4.9, e). However, three CxCa samples analysed by RT-qPCR$_{E1C}$ were positive only in 4,000 µL while only one was discordantly positive in 1,000 µL extracts.

Since the number of CxCa samples above cutoff increased using 4,000 µL extraction volumes while at the same time the number of NIL/M samples above cutoff did not increase, the larger volume was used further for validating HPV16 RNA patterns.

4.1.5 Reproducibility of RT-qPCR and RNA extraction

4.1.5.1 Reproducibility of RT-qPCR$_{TP}$ and RT-qPCR$_{E1C}$

Intra- and inter-plate variations of the final RT-qPCR$_{TP}$ and RT-qPCR$_{E1C}$ were determined with ten-fold dilution series of the *in vitro* transcripts E6*I, E1^E4, E1C and ubC in duplicates on the same plate and on two different plates.

For E6*I *in vitro* transcripts the mean intra-plate coefficient of variation (CV) was 1.07% (range 0.17 to 2.38%) (Table 4.5, a) and the mean inter-plate CV was 1.73 % (range 0.57 to 3.05%) (Table 4.5, b). The mean intra-plate CV of the other HPV16 transcripts and of ubC ranged between 0.57 to 0.87% and the inter-plate CV between 0.73 to 1.62% (Table 4.5). As expected, inter-plate was higher than intra-plate CV for all viral transcripts and ubC. Variation increased with decreasing RNA dilutions, i.e. the closer the template concentrations to DL the higher the CV.

Table 4.5. **Intra (a) - and inter (b) -plate reproducibility of RT-qPCR$_{TP}$ and RT-qPCR$_{E1C}$.**

(a) Intra-plate reproducibility

RNA	E6*I		E1^E4		ubC		E1C	
dilutions	mean Cp	CV (%)[a]	mean Cp	CV (%)	mean Cp	CV (%)	mean Cp	CV (%)
10	37	2.37	> 42		> 42		35	2.02
100	33	2.27	36	0.88	34	0.12	34	0.13
1,000	30	0.16	34	0.54	31	1.41	31	0.30
10,000	27	0.18	31	0.18	28	0.52	28	0.10
100,000	24	0.48	28	0.38	24	1.56	25	0.68
1,000,000	21	0.96	24	0.84	21	0.72	21	0.30
Mean[b]:	29	1.07	31	0.57	28	0.87	29	0.59

(b) Inter-plate reproducibility

RNA	E6*I		E1^E4		ubC		E1C	
dilutions	mean Cp	CV (%)	mean Cp	CV (%)	mean Cp	CV (%)	mean Cp	CV (%)
10	36	2.46	> 42		> 42		36	3.09
100	33	1.40	36	0.78	34	0.24	34	2.91
1,000	30	0.57	34	1.08	31	1.22	31	0.82
10,000	27	1.25	31	0.48	28	0.63	28	0.78
100,000	23	1.64	28	0.26	24	1.14	25	1.46
1,000,000	20	3.05	24	1.03	21	0.51	21	0.65
Mean:	28	1.73	31	0.73	28	0.75	29	1.62

[a] coefficient of variation (CV) calculated as standard deviation/mean*100 with 0% indicating no difference between Cp values, [b] Cp >42 was not included in calculation of the mean

A reduced DL as observed for E1^E4 and ubC in this experiment might be explained by RNA degradation of the *in vitro* transcript dilution series over time (Table 4.5). Therefore dilution series were periodically renewed.

Overall, high inter- and intra-plate reproducibility of the RT-qPCR$_{TP}$ and RT-qPCR$_{E1C}$ was observed with mean CV from 0.59 to 2.27% and 0.75 to 1.94%, respectively.

4.1.5.2 Reproducibility of RNA extraction

The reproducibility of RNA extraction was determined by extracting twice the RNA from 48 HPV16 DNA-positive exfoliated cervical cell samples from Bad Münder (CIN0 (n=5), CIN1 (n=8), CIN2 (n=18), and CIN3 (n=17)). All RNA extracts were applied in duplicates to the RT-qPCR$_{TP}$.

Only 1 out of 48 exfoliated cervical cell samples was discordantly ubC-positive, however, quantified with 10^2 copies numbers only in extraction 2 (Figure 4.10, a). In total, 42 exfoliated cervical cell samples were concordantly positive or negative for all viral transcripts and six were discordant, being negative once either for E6*I or E1^E4. The three discordant E6*I cases were once borderline positive (below 12 copies per PCR) (Figure 4.10, b). The three discordant E1^E4 cases were quantified with copies between 10^2 and 10^3 per PCR (Figure 4.10, c), however, showed in either both or one of the extracts low ubC copies per PCR (<100 copies per PCR).

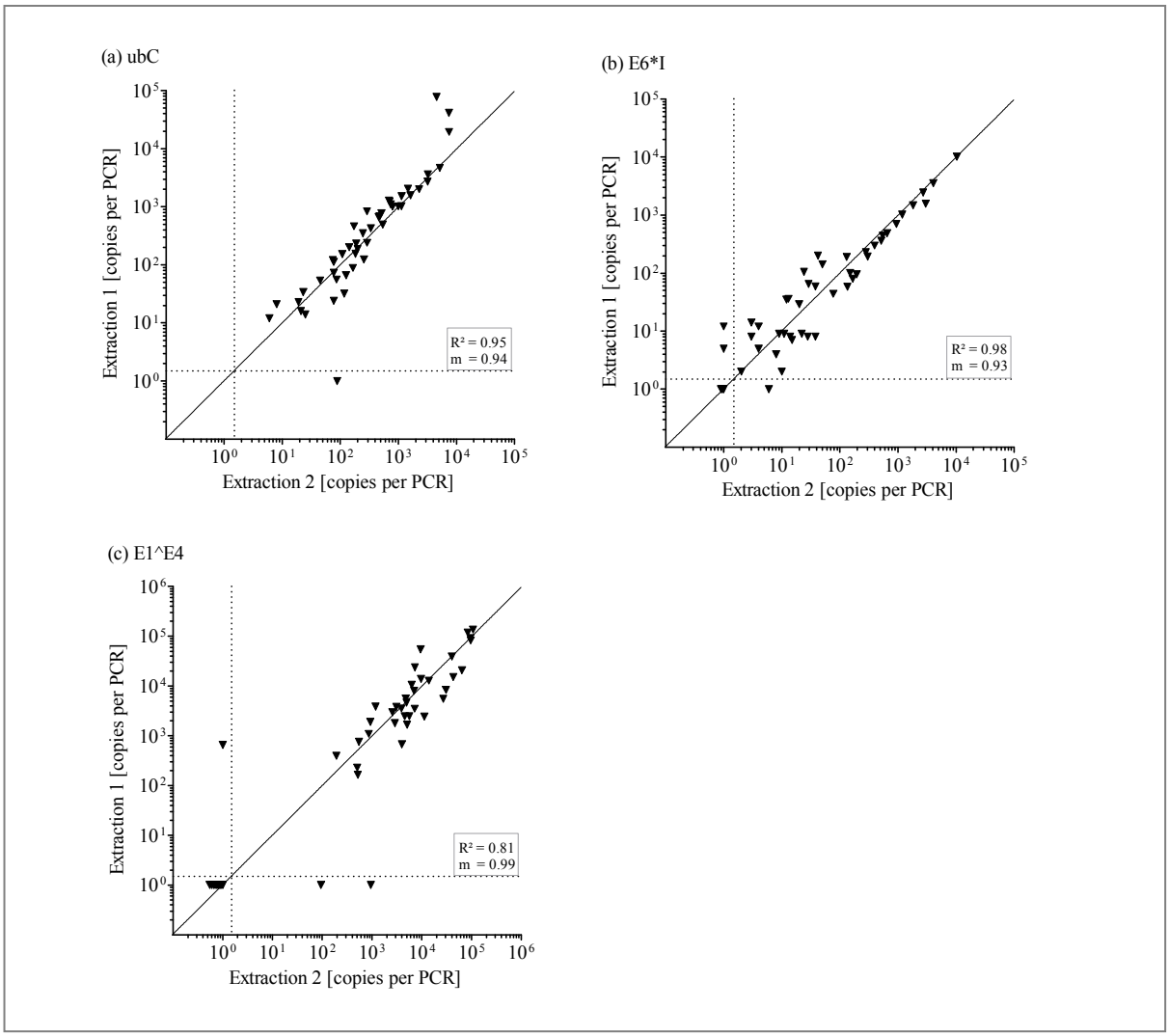


Figure 4.10. **Reproducibility of the RNA extraction**. UbC (a), E6*I (b) and E1^E4 copies per PCR (c) are shown in extraction 1 (y-axis) compared to extraction 2 (x-axis). The diagonal is indicated. Slope (m) of the regression line is depicted, a value of 1 indicates the absence of systematic differences between the extractions. Coefficient of determination (R^2) describes the correlation between copies per PCR of extraction 1 and 2, with value of 1 being perfect. All templates were applied in duplicates and mean Cp value was used for the calculation of copies per PCR. Negative results, i.e. no detectable RT-qPCR signal within 42 cycles, are indicated below the dotted cutoff line.

Across E6*I, E1^E4 and ubC, the coefficient of determination (R^2) ranged between 0.81 to 0.98, $p<0.0001$, indicating a highly significant linear correlation. The slopes (m) of the regression lines were between 0.93 and 0.99 confirming a robust reproducibility of the RNA extraction (Figure 4.10). The Cp values <42 measured in exfoliated cervical cell samples from extraction 1 and extraction 2 varied with a mean CV of 1.96% (0-16%) for E6*I, 2.24% (0-18.3%) for E1^E4 and 1.41% (0-18.3%) for ubC. The reproducibility of the

repeated extraction was slightly weaker compared to the intra-/inter-plate CV of RT-qPCR$_{TP}$ (Table 4.5).

Overall, a high reproducibility of RNA extraction was observed.

4.1.6 Comparison of RT-qPCR and singleplex NASBA

The performance of RT-qPCR$_{TP}$ and RT-qPCR$_{E1C}$ for the detection of E6*I, E1^E4, E1C and ubC was compared to the established NASBA method detecting E6*II, E1^E4, E1C, L1 and ubC. The comparison was performed using cervical cancer cell lines (chapter 4.1.6.1) and followed by oropharyngeal squamous cell carcinomas (OPSCC) (chapter 4.1.6.2). RNA from the cervical cancer cell lines originated from different passages, but the RNA extracts from OPSCC were the same samples applied to both methods.

4.1.6.1 Transcript quantification in cervical cancer cell lines

A comparison of RT-qPCR$_{TP}$ and RT-qPCR$_{E1C}$ and the established NASBA method was done using RNA from cervical cancer cell lines MRI-H186, SiHa, CaSki and MRI-H196.

The NASBA data was based on two experiments run on two different days using the same cell line RNA without intra-assay duplicates showing very high variations. MRI-H186 was known to contain between 7.5 and 7,500 E6*I, between 0.75 and 750 E1^E4, between 0.08 and 0.75 E1C, between 0.08 and 7.5 L1 and between 0.75 and 750 ubC copies per cell[148] (Figure 4.11, a). Variations observed in SiHa, CaSki and MRI-H196 were shown in figure 4.11, b-d.

Total RNA (33 ng, corresponding to approximately 2,200 cells) extracted from cell lines MRI-H186, SiHa, CaSki and MRI-H196 was applied 2 to 12 (in case of MRI-H186)-times to RT-qPCR$_{TP}$ and RT-qPCR$_{E1C}$.

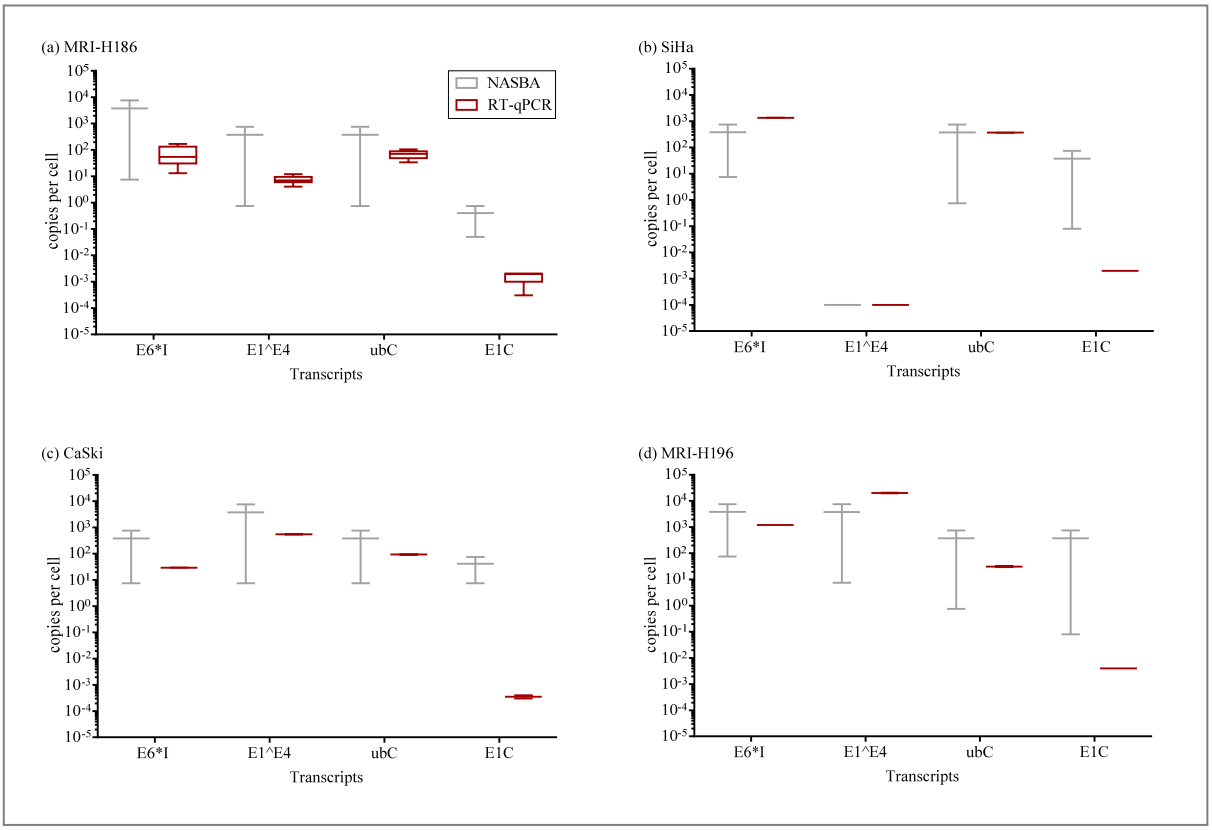


Figure 4.11. **Comparison of RT-qPCR and singleplex NASBA.** Copies per cell of E6*I, E1^E4, ubC and E1C in RNA from HPV16-tranformed cervical cancer cell lines MRI-H186 (a), SiHa (b), CaSki (c) and MRI-H196 (d) quantified in NASBA[148] (grey bars) and RT-qPCR (red bars). Calculation of copies per cell based on the assumption that 5 ng of RNA correspond to 330 cells. A value below 1 indicates that the transcript is not present in every cell. All templates were analysed in duplicates and in RT-qPCR in dodecuplicate in case of MRI-H186 and presented as boxplots with whiskers from minimum to maximum. If duplicates are nearly identical, whiskers cannot be recognised. If transcript was not detected in any duplicate, value was set to 0.0001.

The mean copies per cell of all high-abundance transcripts quantified by RT-qPCR$_{TP}$ were either identical (ubC in SiHa), between 3- (E6*I in MRI-H196) and 50-times (E6*I and E1^E4 in MRI-H186) lower or between 4- (E6*I in SiHa) and 5-times (E1^E4 in MRI-H196) higher than by NASBA (Figure 4.11). In all cell lines, the mean E1C copies per cell were lower by RT-qPCR$_{E1C}$ (refer to 5.2.5 for discussion).

However, CV of NASBA duplicates were high: between 139 and 141%. Therefore, the difference of mean copies per cell was not significant. In contrast, the quantified number of transcripts per cell by RT-qPCR$_{TP}$ and RT-qPCR$_{E1C}$ was highly reproducible with CV between 0% (E1C) and 72% (E6*I).

To exclude mistakes of the RT-qPCR$_{TP}$ and RT-qPCR$_{E1C}$ quantification, the *in vitro* transcription of all transcripts as well as the titration curve of the transcript dilutions was repeated (data not shown). The DL as well as RT-qPCR$_{TP}$ and RT-qPCR$_{E1C}$ efficiency was confirmed, excluding the possibility that mistakes in RT-qPCR$_{TP}$ and RT-qPCR$_{E1C}$ standard curves caused the discrepancy to NASBA (refer to 5.2.5 for discussion).

4.1.6.2 Transcript quantification in oropharyngeal squamous cell carcinomas

RNA samples had been extracted from 32 fresh-frozen OPSCC biopsies and the copies of E6*II, E1^E4, E1C and ubC had been determined by singleplex NASBA performed by Dr. Dana Holzinger[152]. The same RNA samples were now analysed by RT-qPCR$_{TP}$ and RT-qPCR$_{E1C}$. A comparison of E6*I with E6*II could be performed as their transcript expression highly correlates in HPV infected cells, though the E6*I expression is around ten-fold higher than for E6*II (see introduction 1.7)[136,146,167].

The qualitative comparison of E6*I and E6*II (E6*I/*II) detected by RT-qPCR$_{TP}$ and NASBA, respectively, showed 24 concordantly E6*I/*II positive samples, 1 concordantly negative and 7 discordant samples resulting in an overall agreement of 78.1% (Figure 4.12, a). The discordant samples that were only positive with RT-qPCR$_{TP}$ were attributed to the higher level of E6*I than E6*II in HPV16-positive cells. A very good quantitative correlation of E6*I/*II detected by both methods was indicated by Pearson correlation (r_p) of 0.94 ($p<0.0001$) (Figure 4.12, a).

The qualitative comparison of E1^E4 detected by RT-qPCR$_{TP}$ and NASBA, respectively, showed 21 concordant positive samples, 5 concordant negative and 6 discordant samples resulting in an overall agreement of 81.3% (Figure 4.12, b). Three were borderline positive with RT-qPCR$_{TP}$ only and three with NASBA. A good quantitative correlation of both methods for E1^E4 was observed with an r_p of 0.69 ($p<0.0001$) (Figure 4.12, b).

A 100% concordance of OPSCC samples defined as positive or negative by the E6*I/E1^E4 ratio in RT-qPCR$_{TP}$ and by the E6*II/E1^E4 ratio in NASBA, respectively, was observed (Figure 4.12, d).

E1C detection was concordantly positive in 16, concordantly negative in 14 and discordant in 2 OPSCC cases that were detected by RT-qPCR$_{E1C}$ only resulting in an overall agreement

of 93.8%. Thus, E1C detection by RT-qPCR$_{E1C}$ was slightly more sensitive than with NASBA. A quantitative correlation of E1C was not observed (r_p=0.15, p=0.41) (Figure 4.12, c) which can be explained by the fact that NASBA was at best semi-quantitative for E1C while RT-qPCR$_{E1C}$ enabled a quantification over 5 logs (see Introduction 1.7).

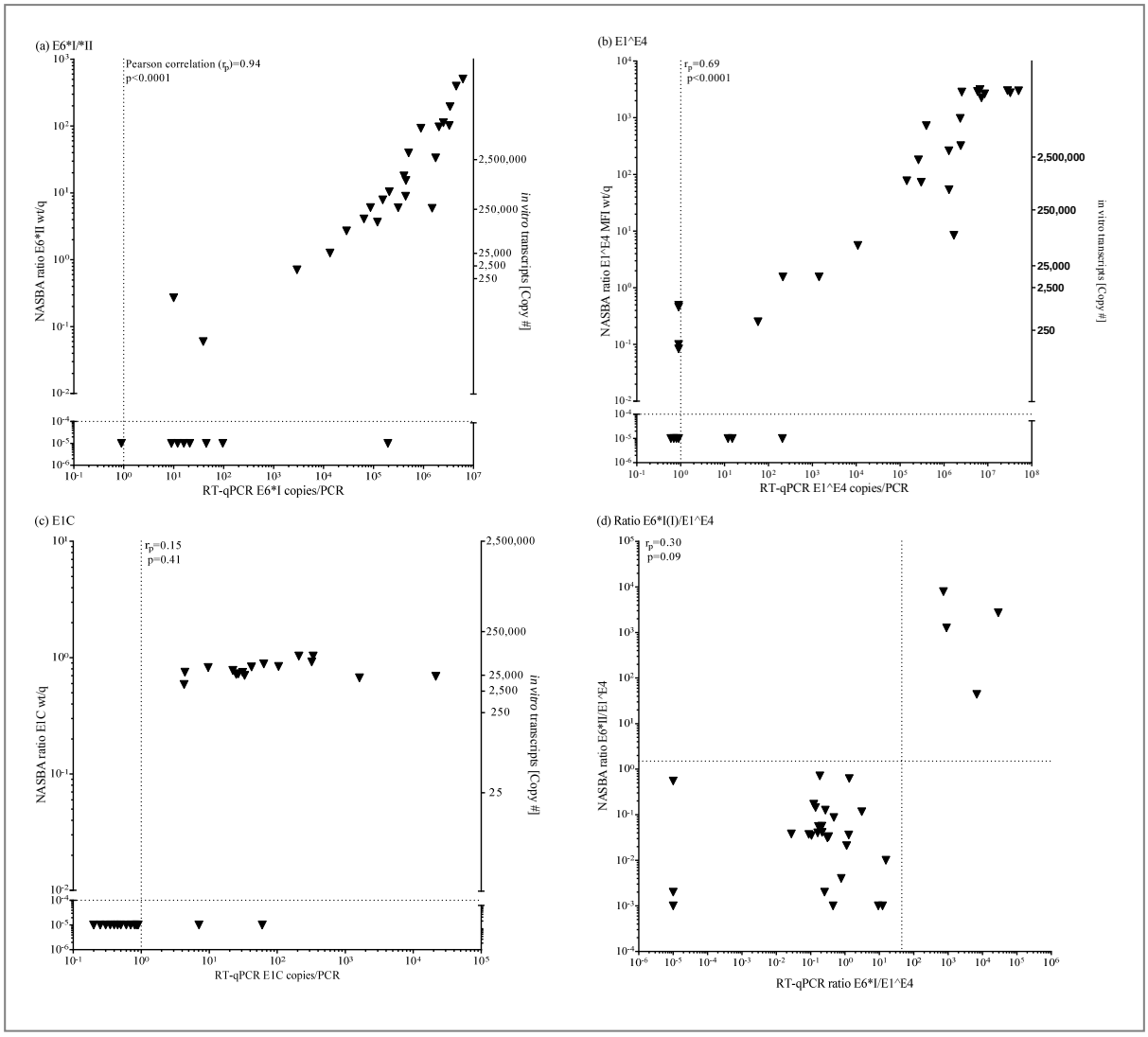


Figure 4.12. **Comparison of transcript quantification by RT-qPCR (copies per PCR on x-axis) and by NASBA (ratio of wildtype (wt)/calibrator (q) signal on y-axis) as well as comparison of E6*I/E1^E4 ratio by RT-qPCR$_{TP}$ and E6*II/E1^E4 ratio by NASBA (d).** Dotted lines indicate the cutoffs. In RT-qPCR, all templates were applied in duplicates and mean Cp value was used for the calculation of copies per PCR. Negative values are set below 1 copy per RT-qPCR, set to 0.00001 for the NASBA ratio wt/q (a-c), set to 0.001 for the NASBA ratio E6*II/E1^E4 and to 0.00001 for the RT-qPCR ratio E6*I/E1^E4 (d).

The overall concordance of the HPV16 RNA patterns (E6*I(I)/E1^E4 ratio and E1C) was 100%: In total, 21 OPSCC cases were concordantly classified as CxCa-like RNA patterns-positive (CxCaRNA$^+$) and 11 concordantly as CxCa-like RNA patterns-negative

(CxCaRNA$^-$). The 11 tumours not displaying viral RNA patterns exhibited very low signals for all viral transcripts in NASBA and RT-qPCR$_{TP}$ and RT-qPCR$_{E1C}$.

4.1.7 Clinical validation of the HPV16 RNA patterns

The RT-qPCR-based HPV16 RNA patterns test to detect advanced (≥CIN2) cervical lesions was clinically validated with exfoliated cervical cells fixed in PreservCyt buffer. The samples were obtained from women cytologically and/or histologically diagnosed as NIL/M, CIN0, CIN1, CIN2, CIN3 and CxCa in studies from Bad Münder and Mongolia. The presence of HPV DNA was determined by BSGP5+/6+-PCR/MPG. Subsequently, in the singly infected HPV16 DNA-positive cervical cell samples the spliced HPV16 transcripts E6*I, E1^E4 and E1C and the cellular transcript ubC were quantified by RT-qPCR$_{TP}$ and RT-qPCR$_{E1C}$.

The HPV16 RNA patterns were compared with the cytological or histological diagnoses near the time of sampling and at follow-up if available. Slides for liquid-based cytology were prepared using a Thin Prep 2000 processor and diagnosed as described elsewhere[149]. Histological diagnoses of colposcopy-directed biopsies were based on independent blinded readings of two pathologists with knowledge of cytological, HPV DNA and p16^{INK4a} staining results in the Bad Münder study. Discordant readings were discussed to reach a final consensus diagnosis. In the Mongolia study, CxCa histology diagnoses were based on agreement of three study pathologists[12].

4.1.7.1 Ascertainment of histological case diagnoses

The accuracy of histological diagnoses in cervical biopsies is critical for adequate patient management and also for evaluation of new screening technologies for cervical cancer precursors such as the HPV16 RNA patterns test.

This chapter summarises the agreement of two pathologists on the diagnosis of 104 cervical biopsies with HPV16 single-infected RNA-positive exfoliated cervical cells from the Bad Münder study.

Overall agreement of initial readings was 58% (60 out of 104). CIN3 had the highest class-specific agreement, with 78% concordance of all initially diagnosed CIN3 cases. Concordant CIN1 diagnosis was present in only 15% of all initially diagnosed CIN1 cases (Table 4.6).

Table 4.6. **Overall agreement of initial histological diagnoses by two pathologists in the Bad Münder study** [a].

		Pathologist 2				
Pathologist 1		CIN0	CIN1	CIN2	CIN3	Total
CIN0						
	N	**7**	5	3	0	15
	%	7	5	3	0	14
	Row %	47	33	20	0	
	Column %	70	56	8	0	
CIN1						
	N	0	**2**	8	3	13
	%	0	2	8	3	13
	Row %	0	15	62	23	
	Column %	0	22	21	6	
CIN2						
	N	1	2	**20**	13	36
	%	1	2	19	13	35
	Row %	3	6	56	36	
	Column %	10	22	53	28	
CIN3						
	N	2	0	7	**31**	40
	%	2	0	7	30	38
	Row %	5	0	18	78	
	Column %	20	0	18	66	
Total						
	N	10	9	38	47	104
	%	10	9	37	45	100

[a] Shaded diagonals represent concordant interpretations. In the shaded diagonals, values in boldface indicate the frequencies (N) of concordance.

Pathologist 2 tended to give more severe diagnoses than pathologist 1. After rediscussion of discordant cases, the diagnosis of pathologist 2 was taken as final diagnosis in 21% of cases, that of pathologist 1 in 17% and in 4% a new diagnosis was given.

The moderate intraobserver reproducibility in initial diagnosis in part may reflect the subjectivity of histological diagnoses. This should be kept in mind when histology, especially derived from a single pathological viewing, is used as gold standard in validation of novel diagnostic assays.

4.1.7.2 Type-specific HPV DNA prevalence in validation samples

DNA was extracted from 691 exfoliated cervical cell samples from Bad Münder and subsequently analysed by BSGP5+/6+-PCR/MPG (Table 4.7). The data from 98 cervical samples from Mongolia have been described before[149].

Table 4.7. **HPV16 DNA prevalence in HPV DNA-positive exfoliated cervical cell samples from Bad Münder and Mongolia**.

Histological or cytological stage	Origin	HPV DNA-positive	HPV16 DNA-positive		Multiple HPV16 DNA-positive[a]		Single HPV16 DNA-positive	
		N	N	%	N	%	N	%
CIN0	Bad Münder	116	27	23	17	15	10	9
CIN1	Bad Münder	129	34	26	19	15	15	12
CIN2	Bad Münder	190	92	48	51	27	41	22
CIN3	Bad Münder	170	109	64	65	38	44	26
NIL/M	Mongolia[b]	71	71	100	34	48	37	52
CxCa	Mongolia[b]	27[c]	27	100	9	33	18	67
Total		703	360	51	195	54	165	46

[a] HPV16 plus any mucosal HPV type, [b] pre-selected for HPV16 DNA-positivity, [c] selected for HPV16 E6*I mRNA positivity by the analysis of corresponding FFPE biopsies using ultra-short HPV E6*I mRNA assays[12]

Of 789 exfoliated cervical cell samples, 703 (89%) were HPV DNA-positive with 360 (51%) containing HPV16 DNA, of these 165 (46%) as single mucosal HPV type. Viral RNA patterns were analysed in HPV16 single-infected samples (chapter 4.1.7.8), in HPV16 multiple-infected samples (chapter 4.1.7.10), and in all HPV DNA negative samples. No false-positive detection of viral transcripts in the HPV negative samples (n=86) was observed (data not shown).

4.1.7.3 RNA quality assessment of validation samples

RNA quality might affect the sensitivity of the HPV16 RNA patterns analyses. In particular storage time and storage conditions of cervical cell samples might be critical for RNA quality.

RNA quality was assessed in RNA samples extracted from the cell line MRI-H196 as well as from a subset of 12 archived exfoliated cervical cell samples stored at 4°C in PreservCyt for 9 to 84 months. The RNA integrity number (RIN) as well as the ubC copies per cell were determined. The RIN is calculated by using the entire RNA profile that is separated according to the molecular weight giving a continuous value from 10 down to 1 defining the extend of RNA degradation (10=intact, 1=totally degraded).

A RIN value of 8.7 measured in MRI-H196 indicated a good RNA quality (Table 4.8). However, the RIN values in the archived cervical cell samples of only 1.0 to 5.2 (median 2.3) indicated partial up to severe degradation of the RNA.

Table 4.8. **RNA quality in extracts from fresh frozen cell line and archived validation samples.**

Source	Material	Storage	Archived at 4°C (months)	Number of samples	Median RNA concentration [ng/µL]	Median RIN (range)[a]	Mean ubC RNA copies/PCR[b] (range)	Median ubC RNA copies/cell[c] (range)
Heidelberg	MRI-H196	Fresh-frozen	0	1	19	8.7	22,485 (19,055-28,489)	16 (11-23)
Bad Münder	Cervical smear	PreservCyt	9-12	5	16	2.1 (1-2.4)	6,955 (53-69,614)	2 (1-3)
Bad Münder	Cervical smear	PreservCyt	13-24	5	45	3.44 (2-5.2)	1,687 (81-15,757)	2.8 (1-6)
Mongolia	Cervical smear	PreservCyt	84	2	24	1.95 (1.6-2.2)	3,661 (1,521-5,800)	3 (1-5)

[a] median RNA integrity number (RIN) where 1 indicates totally degraded and 10 completely intact RNA,

[b] mean ubC RNA copies per PCR in MRI-H186 was calculated from same RNA extract applied 5-times to RT-qPCR$_{TP}$ and in cervical smear from 5 and 2 different samples, respectively, applied in duplicates to RT-qPCR$_{TP}$, [c] ubC RNA copies per cell were calculated using the measured RNA concentration per µL, the quantified ubC RNA copies per PCR and the assumption that 5 ng of RNA correspond to 330 cells

The calculated ubC copies per cell were up to 8-times higher in the fresh-frozen cell line than in the archived cervical cell samples (Table 4.8). However, no correlation of ubC copy number with RIN was observed (r_p=0.25, p=0.4, Figure A8.2).

RNA quality was reduced in archived cervical cell samples compared to the fresh-frozen cell line sample.

4.1.7.4 Qualitative prevalence of single transcripts in validation samples

The previous NASBA pilot study had shown that the qualitative transcript analyses cannot discriminate mild from severe lesions[87]. The aim of this chapter was to reproduce these findings by using the RT-qPCR$_{TP}$ and RT-qPCR$_{E1C}$.

RNA was extracted from 4 mL of the single HPV16 DNA-positive cervical cell samples from Bad Münder (n=110) and Mongolia (n=55). Subsequently, RNA samples were analysed in duplicates by RT-qPCR$_{TP}$ and RT-qPCR$_{E1C}$.

Of the 165 RNA samples, 158 (96%) were valid (>80 ubC or >200 E6*I copies per PCR, chapter 3.12). Overall prevalence of any viral transcript increased from NIL/M (8%) to CIN0 (50%) and reached 95-100% in the cervical lesions (Table 4.9).

Table 4.9. **Prevalence of individual transcripts in 158 HPV16 single-infected cervical cell samples with valid RNA by cytological or histological lesion stage.**

Transcript	All		Cytological or histological stage											
			NIL/M		CIN0		CIN1		CIN2		CIN3		CxCa	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%
RNA invalid	7	4	1	3	0	0	4	30	0	0	2	5	0	0
RNA valid														
E6*I-positive	113	72	3	8	4	40	9	82	39	95	40	95	18	100
E1^E4-positive	96	61	2	6	3	30	10	91	39	95	36	86	6	33
E1C-positive	34	22	1	3	0	0	0	0	7	17	22	52	4	22
Any viral transcript-positive	117	74	3	8	5	50	11	100	40	98	40	95	18	100
ubC-positive	158	100	36	100	10	100	11	100	41	100	42	100	18	100

In agreement with the previous NASBA pilot study, E6*I prevalence increased from NIL/M (8%) and CIN0 (40%) to CIN1 (82%) and remained between 95% and 100% for CIN2, CIN3 and CxCa. Likewise, E1^E4 prevalence increased as well, reaching a plateau around 90% in CIN1, CIN2 and CIN3, but decreased substantially in CxCa (33%). In contrast, the prevalence of E1C was below 11% until CIN2, peaked in CIN3 (52%) and decreased again in CxCa (22%). The decrease of the E1C prevalence in CxCa was not in agreement with the pilot study, where 57% of CxCa samples were E1C-positive (refer to 5.3.3 for discussion).

While viral transcripts were more frequent in cervical abnormalities compared to healthy cytology, the qualitative expression of E6*I and E1^E4 cannot discriminate ≤CIN1 from ≥CIN2 lesions. In contrast, a qualitative detection of E1C appears to be a specific marker for ≥CIN2 lesions.

4.1.7.5 Quantitative expression levels of single transcripts in 158 validation samples

The previous NASBA pilot study showed that E6*I and E1C encoding transcripts were strongly upregulated in severe lesions, whereas spliced E1^E4 encoding transcript was markedly downregulated[146]. The aim of this chapter was to reproduce these findings by using the RT-qPCR$_{TP}$ and RT-qPCR$_{E1C}$.

The median ubC copy number in the single HPV16 DNA-positive cervical cell samples (Table 4.7) was 901 (range 54 to 69,614). The copy number distribution showed no significant difference neither between the different cytological and histological stages (Figure 4.13, d) nor between samples from Bad Münder and Mongolia (p=0.7, Mann-Whitney-Test) (Figure A8.3).

Among the E6*I-positives, the median copies per PCR increased from CIN0 (30 copies per PCR), CIN1 (75), CIN2 (381), CIN3 (894) to CxCa (1,180). E6*I showed a significant upregulation in ≥CIN3 compared to ≤CIN1 (p=0.0039, Mann-Whitney-Test) (Figure 4.13, a).

Among the E1^E4-positives, the median copies per PCR were equivalent in NIL/M (4,114 copies per PCR), CIN0 (2,972) and CIN1 (2,243), increased from CIN2 (6,679) to CIN3 (13,082) and declined in the CxCa samples (5,973), however, without significant differences between the different stages (p=0.91, 1-way-Anova) (Figure 4.13, b).

A wide range of copies (between 2 and 841,438) of the high-abundance transcripts E6*I and E1^E4 was observed.

The mean E1C copies per PCR were nearly equivalent in CIN2 (40 copies per PCR) and CIN3 (16) and significantly reduced in CxCa (9) in comparison to CIN2 (p=0.0424, Mann-Whitney-Test) (refer to 5.3.3 for discussion) (Figure 4.13, c).

Compared to ≤CIN1, in ≥CIN2 transcript levels of E6*I (p=0.0088), E1^E4 (p=0.0267) and E1C (significance not possible to calculate due to low number of positives among ≤CIN1) were significantly augmented. However, in further analyses CIN2 was categorised as intermediate and not grouped within the severe lesions since CIN2 classification has been reported to have the highest intraobserver variation among pathologists[107].

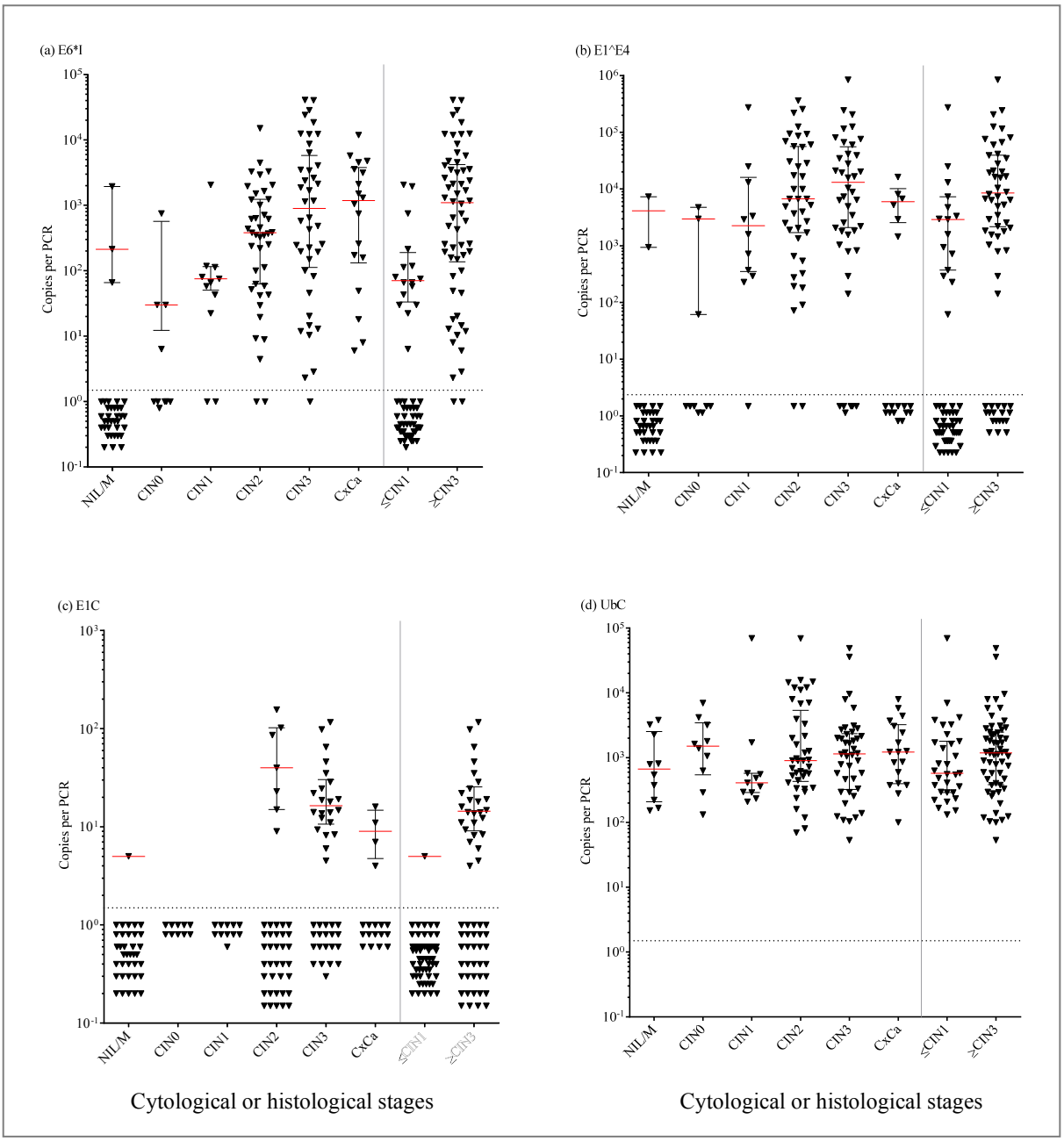


Figure 4.13. **Copy number per PCR (y-axis) of E6*I (a), E1^E4 (b), E1C (c) and ubC (d) transcripts in 158 cervical cell samples of different cytological or histological lesion stages (x-axis, categories).** NIL/M (n=36), CIN0 (n=10) and CIN1 (n=11) are combined in the ≤CIN1 (n=57) category and CIN3 (n=42) and CxCa (n=18) in the ≥CIN3 (n=60) category. Negative values are indicated below 1 copy per PCR. Cutoff is shown as dotted line, red lines represent the arithmetic median with interquartile range. All templates were applied in duplicates and mean Cp value was used for the calculation of copy number.

The spliced E6*I and E1^E4 changed the most in different grades of lesion. However, the E6*I and E1^E4 copy numbers alone cannot discriminate ≤CIN1 from ≥CIN2 lesions. Only very low E1C copy numbers were detected in E1C-positive samples, which may hint at a regulatory function of the E1C protein (refer to 5.3.3 for further discussion).

4.1.7.6 Influence of cell number on viral transcript quantification

Supposing that all cells in the cervical smear are HPV-infected, a positive correlation between the quantified ubC and the viral transcript copies per PCR is expected. Furthermore, a negative correlation is expected if a high copy number of ubC inhibits the quantification of viral transcripts.

The viral transcript copy numbers (described in chapter 4.1.7.5) were directly compared to ubC copy numbers (Figure 4.14).

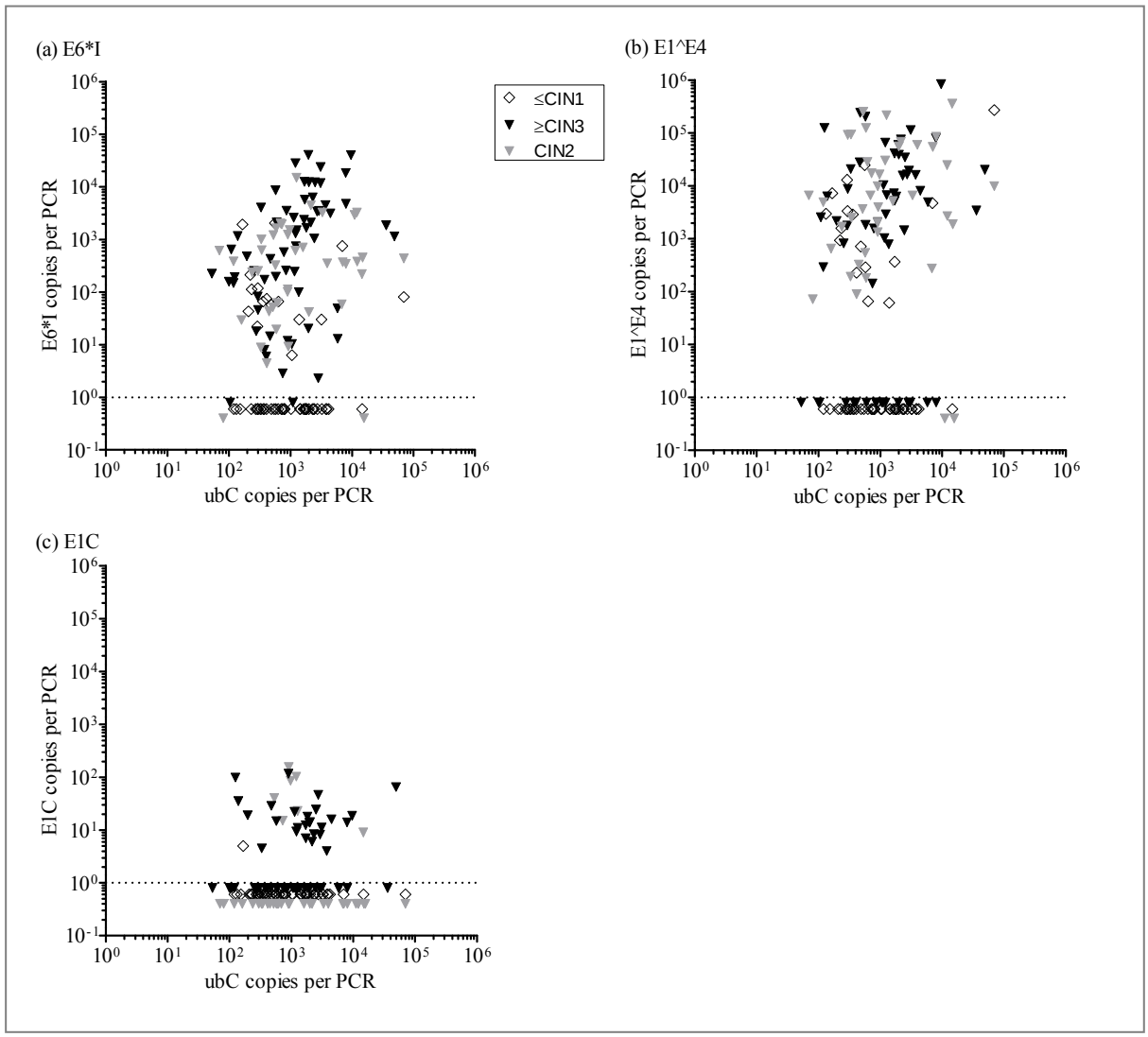


Figure 4.14. **Correlation of viral transcript copies per PCR (y-axis) and ubC transcript copies (x-axis) in 158 cervical cell samples of different cytological or histological lesion stages**. NIL/M (n=36), CIN0 (n=10) and CIN1 (n=11) are combined as ≤CIN1 (n=57) and CIN3 (n=42) and CxCa (n=18) as ≥CIN3 (n=60). Lesions ≥CIN3 are depicted as black triangles, CIN2 as grey triangles and ≤CIN1 as open rhombi. All templates were applied in duplicates and mean Cp value was used for the calculation of copies per PCR. Negative values are indicated below 1 copy per PCR. Dotted lines indicate cutoffs.

No correlation of ubC and E6*I copy numbers was observed neither in all samples, nor in ≥CIN3 or in ≤CIN1 as indicated by the low Pearson correlation of r_p=0.02 (p=0.8), r_p=0.06 (p=0.7) and r_p=-0.1 (p=0.9), respectively. Only 5 out of 16 samples with ubC <200 copies per PCR were E6*I-negative. Thus, a low number of ubC copies did not correlate with a low number of E6*I copies. As already shown above (Figure 4.13), high E6*I copy numbers are mainly present in lesions diagnosed as ≥CIN2.

Likewise, no correlation of ubC and E1^E4 copy numbers was observed neither in all samples (r_p=0.2, p=0.1), nor in ≥CIN3 (r_p=0.06, p=0.7) nor in ≤CIN1 (r_p=0.04, p=0.8). Low ubC copy numbers did not correlate with the quantification of E1^E4 since only 5 of 16 (31%) samples with ubC <200 copies and 57 out of 142 (40%) with ubC >200 copies were E1^E4-negative.

Similarly, no correlation of ubC and E1C copy numbers was observed neither in ≥CIN3 (r_p=0.2, p=0.3), nor in ≤CIN1 (cannot be calculated due to small number of data points) nor in all samples (r_p=0.02, p=0.8) (Figure 4.14, c). E1C was mostly detected in high-grade lesions with copies ranging between 4 and 156. Eight cervical cell samples diagnosed as ≥CIN3 had ubC copies per PCR <200, however two of those were E1C-positive with 35 and 100 copies per PCR, respectively. As not all ≥CIN3 lesions were expected to be E1C-positive, low ubC copies did not correlate necessarily with low E1C copies.

Consequently, the assumption that all cells are HPV-infected and that ubC influences viral transcript quantification can be rejected. Cervical smear samples contain variable amounts of HPV-positive cells.

4.1.7.7 Cutoff definition of HPV16 RNA patterns in HPV16 single-infected validation samples

As mentioned in the previous chapter, the relative and total number of HPV-positive cells per cervical smear can vary. Thus, a copy number ratio of E6*I and E1^E4 was computed allowing the normalisation for variable quantities of HPV-infected cells in cervical smears assuming that the transcript ratio is similar in the majority of HPV-positive cells in a given sample. This chapter summarises the cutoff definition of the E6*I/E1^E4 ratio.

To obtain a clinically useful cutoff for the E6*I/E1^E4 ratio in detecting ≥CIN3, a panel of 117 HPV16 single-infected samples was analysed by RT-qPCR$_{TP}$ and RT-qPCR$_{E1C}$. The 41 CIN2 samples were excluded due to the low diagnostic accuracy of this histological category[107]. Cutoff for E6*I/E1^E4 (cutoff$_{E6*I/E1^E4}$) was determined by Receiver Operating Characteristic (ROC) curves comparing E1C-negative samples ≤CIN1 with E1C-negative ≥CIN3.

The area under the curve (AUC) of 0.9 indicated that all calculated cutoffs were better than random (Figure 4.15, a). Sum of sensitivity and specificity was highest with a cutoff$_{E6*I/E1^E4}$ of 0.095 with 80.56% and 84.21%, respectively. When E1C RNA positivity was added to the definition of advanced lesions the sensitivity for ≥CIN3 increased to 88%, without loosing specificity.

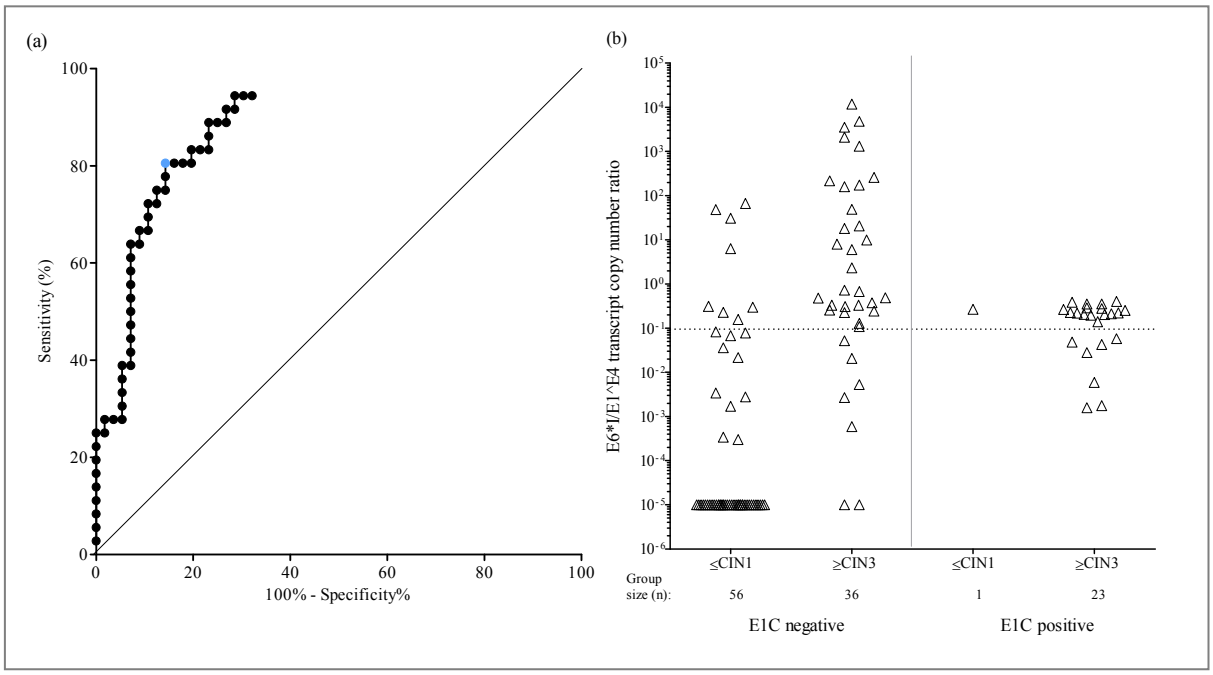


Figure 4.15. **Determination of E6*I/E1^E4 cutoff**. (a) Receiver operating characteristic (ROC) curve calculates sensitivity and specificity for discriminating ≤CIN1 and ≥CIN3 lesions. Combination of highest sensitivity and specificity is marked in blue. The diagonal divides the ROC space, points above indicate good (better than random) and points below poor (worse than random) classification. (b) E6*I/E1^E4 ratio (y-axis) in NIL/M, CIN0, CIN1 are combined in the ≤CIN1 category and CIN3 and CxCa in the ≥CIN3 category with no E1C detection (left) and E1C detection (right). The cutoff line (0.095) is shown as dotted line. All templates were applied in duplicates and mean Cp value was used for the calculation of copy numbers.

The cutoff$_{E6*I/E1^E4}$ of 0.095 divided the cervical samples in three groups: E1C-positive only, E1C and E6*I/E1^E4 double positive and E6*I/E1^E4 ratio-positive only (Figure 4.15, b) (refer to chapter 4.1.7.8 and 5.3.5 for discussion).

In conclusion, a cutoff$_{E6*I/E1^E4}$ of 0.095 was used in all further analyses.

4.1.7.8 HPV16 RNA patterns in HPV16 single-infected validation samples

Next, I analysed the presence of transformation-specific HPV16 RNA patterns (copy number ratio of E6*I/E1^E4 >0.095 and/or presence of E1C) in the same HPV16 single-infected cervical lesions used for cutoff definition before (Figure 4.16).

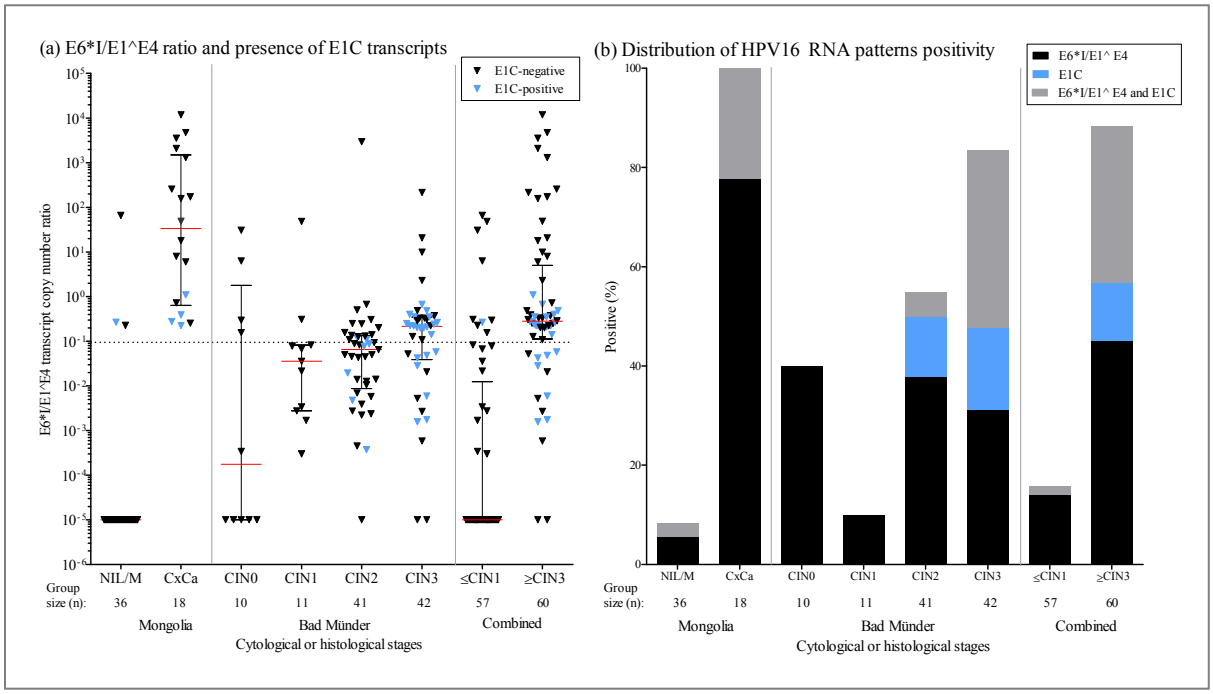


Figure 4.16. **HPV16 RNA patterns in 158 cervical lesions of different cytological and histological stages (x-axis, categories)**. (a) Transcript copy number ratio of E6*I/E1^E4 (y-axis) and presence of E1C (marked in blue) in NIL/M (n=36), CIN0 (n=10) and CIN1 (n=11) are combined in the ≤CIN1 (n=57) category and CIN3 (n=42) and CxCa (n=18) in the ≥CIN3 (n=60) category. Cutoff (0.095 for E6*I/E1^E4 ratio) is shown as dotted line, red lines represent the median with interquartile range. E6*I and E1^E4-negative samples are indicated arbitrarily by a ratio of 0.00001. E1C-positive samples were detected with Cp <42. (b) Black bars in the bar diagram summarise the only E6*I/E1^E4 ratio-positives, blue bars the only E1C-positives and grey bars the E6*I/E1^E4 ratio and E1C double positives.

The frequency of E6*I/E1^E4 ratio-positive samples increased from NIL/M (8%) to CIN1 (18%), CIN3 (67%) and increased in CxCa (100%) (Figure 4.16). The frequency in CIN0 with 40% was elevated but this is based on 4 of 10 data points only (Figure 4.16, a). Compared to mild lesions (≤CIN1), the increase was significantly higher in CxCa ($p<0.0001$, Wilcoxon-signed-rank test), in CIN3 ($p=0.0005$) and in CIN2 ($p=0.0135$) as well

as in the group of severe lesions (≥CIN3) ($p<0.0001$). All samples with an E6*I/E1^E4 ratio above 2, were E1^E4 negative. The frequency of E1C-positives rose from NIL/M (3%) to CIN3 (52%) and dropped again in CxCa (22%) (Figure 4.16). In CIN2 lesions reported to have low diagnostic accuracy[107], 37% were E6*I/E1^E4 ratio- and 17% E1C-positive, as expected for the intermediated group between CIN1 and CIN3 diagnosed lesions. Overall, the frequency of samples positive for at least one of the HPV16 RNA patterns increased from NIL/M (8%), CIN0 (40%), CIN1 (19%) to CIN3 (83%) and to CxCa (100%). Thus, 48 out of 57 (84%) HPV16 single-infected cervical cell samples with NIL/M, CIN0, CIN1 were correctly classified as mild and 53 out of 60 (88%) CIN3 and CxCa were correctly classified as severe lesion.

4.1.7.9 HPV16 load in HPV16 single-infected validation samples

High HPV16 load is a more precise marker for predicting the presence of cervical lesions than HPV DNA positivity alone. Viral loads can reliably discriminate between no lesion and any lesion[134]. For a comparison between the two markers (HPV16 RNA patterns and high HPV16 load (defined as >0.0007 %HPV MFI/ß-globin MFI)), the prevalence of high HPV16 loads in 158 single-infected HPV16 DNA-positive cervical cell samples was examined.

The HPV16 loads in cervical cells of single infected CIN0 ($p=0.0002$, Mann-Whitney-Test), CIN1 ($p=0.0004$), CIN2 ($p<0.0001$), CIN3 ($p<0.0001$) and CxCa ($p=0.0018$) lesions were significantly increased compared to NIL/M (Figure 4.17). However, insignificant viral load difference was observed between CIN0 and CIN1 ($p=0.5$, Mann-Whitney-Test), CIN2 ($p=0.9$), or CIN3 ($p=0.6$). The frequency of high HPV16 load increased from NIL/M (8%) to CIN0 (100%), CIN1 (100%), CIN2 (97%), CIN3 (98%) and CxCa (100%).

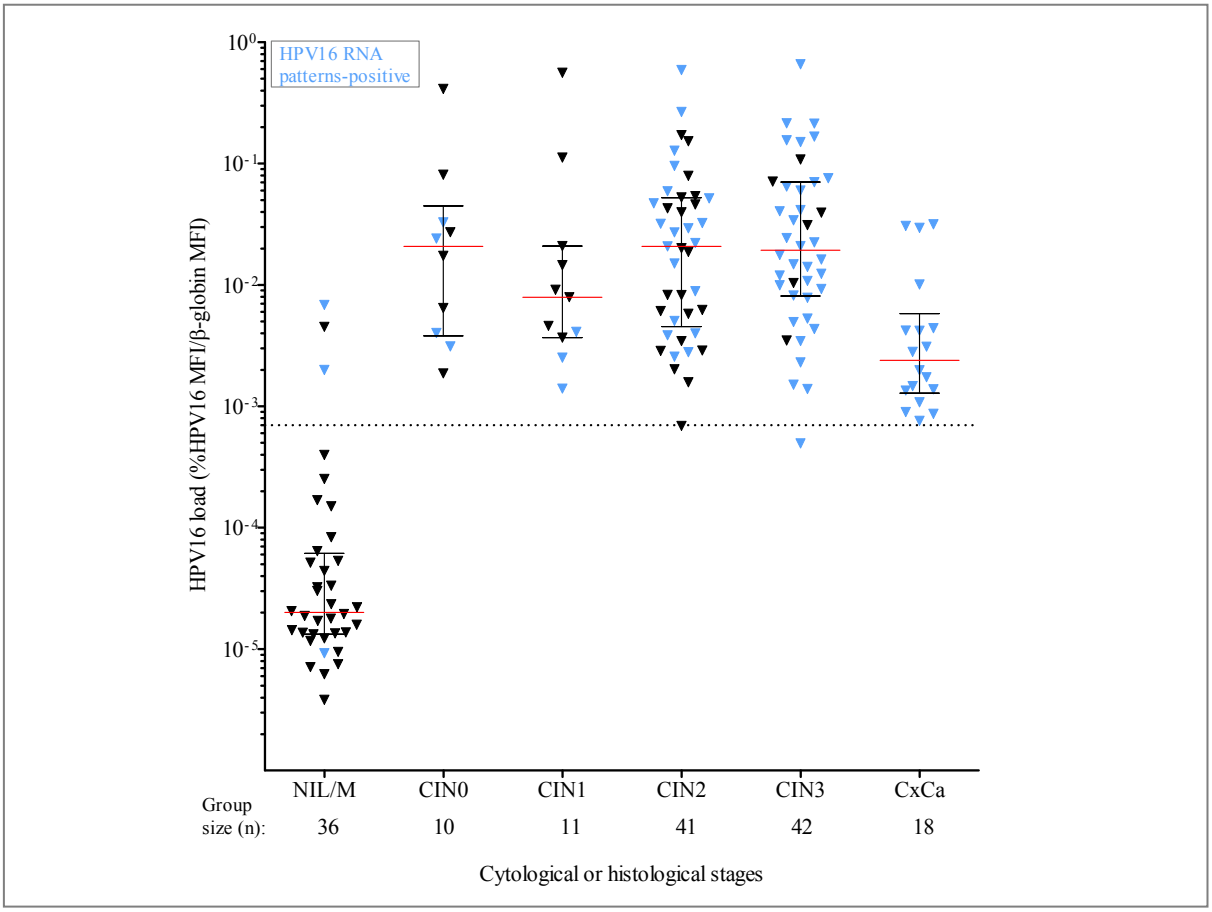


Figure 4.17. **Distribution of viral load among cervical lesions of different grades in single-infected HPV16 DNA-positive cervical cell samples**. HPV16 loads (y-axis) are plotted for groups of cervical lesions defined by cytology or histology (x-axis, categories) including NIL/M (n=36), CIN0 (n=10), CIN1 (n=11), CIN2 (n=41), CIN3 (n=42) and CxCa (n=18). Cutoff (0.0007 for high viral load) is shown as dotted line, red lines represent the median with interquartile range in black. Samples with transformation-specific HPV16 RNA patterns are marked in blue.

Of interest, two NIL/M samples had high HPV16 loads and showed transformation-specific HPV16 RNA patterns (Figure 4.17). One additional NIL/M sample showed high HPV16 load only. Moreover, all CIN0 and CIN1 had a high HPV16 load, while only 4 of 10 CIN0 and 3 of 11 CIN1 were HPV16 RNA patterns-positive. In contrast to the HPV16 RNA patterns, 41 of 42 CIN3 showed high HPV16 load. All CxCa showed high HPV16 load and were HPV16 RNA patterns-positive.

In concordance to the literature[133,134], high viral load cannot discriminate between CIN0, CIN1, CIN2 and CIN3 while HPV16 RNA patterns can specifically identify 84% of the HPV16 single-infected ≤CIN1 as negative for transformation-specific HPV16 RNA patterns.

4.1.7.10 HPV16 RNA patterns and HPV16 load in multiple-infected cervical cell samples

Next, I compared the presence of transformation-specific HPV16 RNA patterns with the cytological or histological classification of cervical cell samples that in addition to HPV16 contained also at least one additional mucosal HPV type (Table 4.7). Identifying the "driving" HPV type for cervical neoplasia is difficult because multiple HPV types are frequently detected within the epithelium of the uterine cervix[134,158]. In addition, multiple cervical lesions within an individual patient may be caused by different types and a precursor lesion, caused by a specific carcinogenic type, can be surrounded by transient HPV infections[168]. In this chapter, HPV16 E6*I-negative samples with HPV16 not being the driving force to carcinogenesis were excluded. HPV16 E6*I-positive samples were additionally analysed for viral load.

RNA was extracted from 195 HPV16 DNA-positive cervical cell samples with multiple HPV infections identified (see chapter 4.1.7.2) in the series from Bad Münder (n=152) and Mongolia (n=43) using an extraction volume of 4 mL exfoliated cervical cell sample. RNA was applied in duplicates to RT-qPCR$_{TP}$ and RT-qPCR$_{E1C}$. Viral loads of all mucosal types were determined by the quantitative BSGP5+/6+-PCR/MPG assay.

Out of the 195 RNA samples, 181 (93%) were RNA-valid (ubC >80 or E6*I >200 copies per PCR, chapter 3.12) and composed of NIL/M (n=34), CIN0 (n=11), CIN1 (n=19), CIN2 (n=48), CIN3 (n=60) and CxCa (n=9). The median ubC copies was 958 ranging from 53 to 70,887 with insignificant difference in the cervical cell samples from Bad Münder und Mongolia (p=0.06, Mann-Whitney-Test) (data not shown).

Of the 64 ≤CIN1, 24 were E6*I-positive including NIL/M (5 out of 34, 15%), CIN0 (7 out of 11, 64%) and CIN1 (12 out of 19, 36%). The frequency of HPV16 RNA patterns-positive samples increased from NIL/M (2 out of 5, 6%), CIN0 (4 out of 7, 36%) to CIN1 (8 out of 12, 42%) (Figure 4.18, a). However, these false-positive cervical cell samples (transformation-specific HPV16 RNA patterns but ≤CIN1 by histology) will be focused in the next chapter (4.1.7.11).

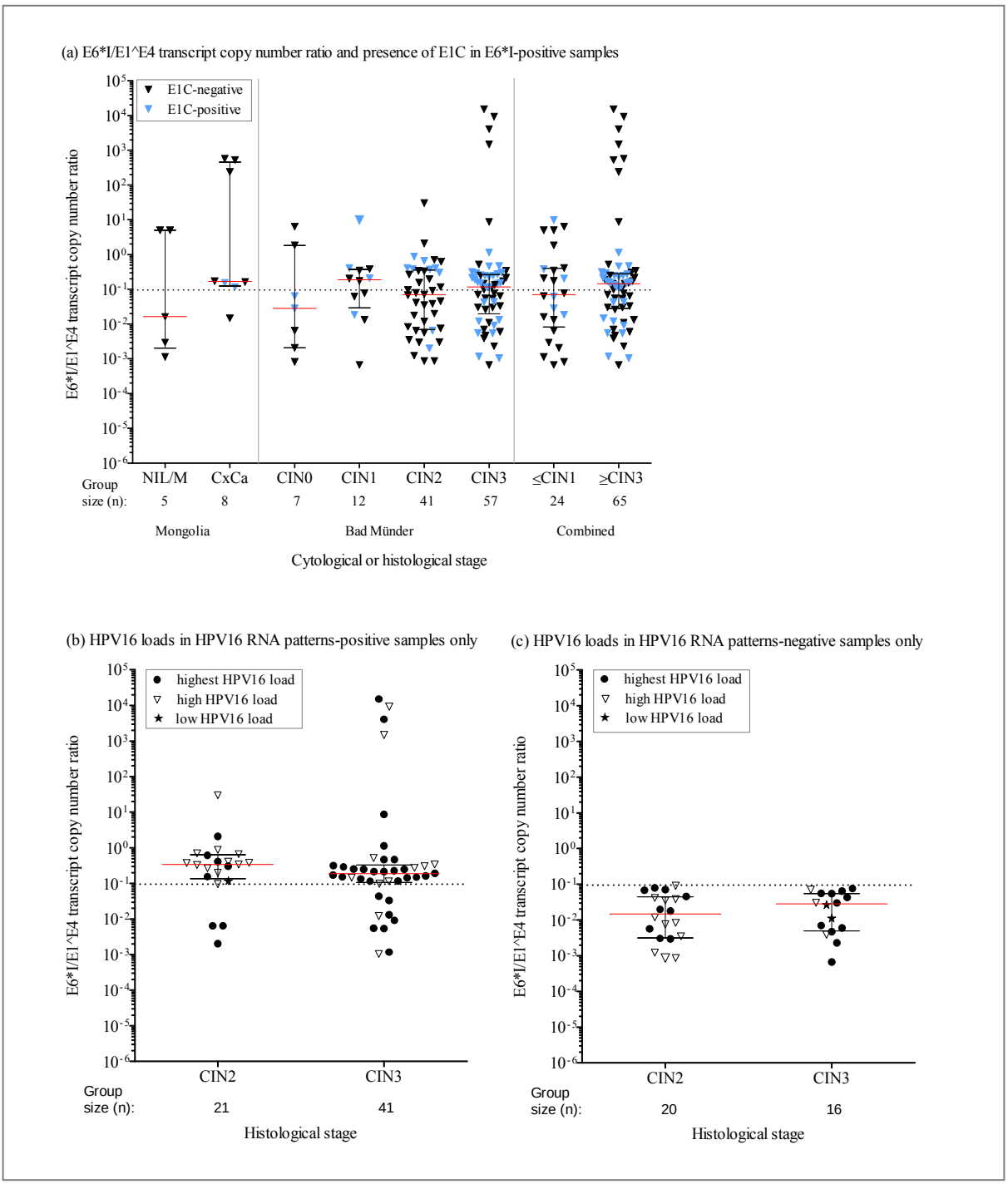


Figure 4.18. **HPV16 RNA patterns in multiple HPV16 DNA-positive cervical lesions of different stages selected for HPV16 E6*I-positivity**. (a) Transcript copy number ratio of E6*I/E1^E4 (y-axis) and presence of E1C (marked in blue) in cervical lesions defined by cytology or histology (x-axis, categories). NIL/M (n=5), CIN0 (n=7) and CIN1 (n=12) are combined in the ≤CIN1 (n=24) category and CIN3 (n=57) and CxCa (n=8) in the ≥CIN3 (n=65) category. Cutoff (0.095 for E6*I/E1^E4 ratio) is shown as dotted line, red lines represent the median with interquartile range in black. E1C-positive samples were detected with Cp <42. (b) Subgroup of E6*I/E1^E4 ratio-positive (y-axis) and/or E1C-positive CIN2 (n=21) and CIN3 (n=41) (x-axis, categories) with low HPV16 load (filled stars), high HPV16 load (open triangles) and highest HPV16 load (filled circles). (c) Subgroup of E6*I/E1^E4 ratio-negative (y-axis) and E1C-negative CIN2 (n=20) and CIN3 (n=16) (x-axis,

categories) with low HPV16 load (filled stars), high HPV16 load (open triangles) and highest HPV16 load (filled circles).

Of the 117 ≥CIN2, 106 were E6*I-positive including CIN2 (n=41), CIN3 (n=57) and CxCa (n=8). The frequency of HPV16 RNA patterns-positive samples increased from CIN2 (n=21, 51%), CIN3 (n=41, 72%) to CxCa (n=7, 88%) (Figure 4.18, a). In addition, 20 of the 21 (95%) HPV16 RNA patterns-positive CIN2 had high HPV16 load and 8 (38%) highest HPV16 load (Figure 4.18, b). Of the 20 HPV16 RNA patterns-negative CIN2, all (100%) had high HPV16 load and 9 (45%) highest HPV16 load (Figure 4.18, c). However, high HPV16 load could also be caused by temporarily productive HPV16 infection without any significance and another HPV type could be responsible for lesion. Confirmatively, 12 CIN2 (60%) had high load with at least one additional hrHPV type. Of the 41 HPV16 RNA patterns-positive CIN3, all (100%) had high HPV16 load and 30 (73%) highest HPV16 load (Figure 4.18, b). However, of the 16 HPV16 RNA patterns-negative CIN3, 13 (82%) had high load with at least one additional hrHPV type (Figure 4.18, c).

As mentioned before, all CxCa samples had been pre-selected for HPV16 E6*I mRNA positivity[12], thus for CxCa samples another HPV type being the driving force to carcinogenesis was excluded. However, FFPE biopsies corresponding to CxCa cervical smear were also analysed by HPV16 RNA patterns (Chapter 4.1.8.1) and the one cervical smear falsely patterns-negative was also patterns-negative in the corresponding FFPE biopsy (refer to 5.3.2 for further discussion).

In summary, the HPV16 RNA patterns positivity was reduced in multiple-infected cervical cell samples selected for E6*I-positivity: Among the CIN2 samples, the number of pattern-positives decreased from 56% (23 out of 41) in HPV16-single to 51% (21 out of 41) in HPV16-positive but multiple-infected samples and among the CIN3 from 83% (35 out of 42) to 72% (41 out of 57) demonstrating the need to extend the HPV16 RNA patterns by other hrHPV types (see chapter 4.2).

4.1.7.11 The predictive value of HPV16 RNA patterns

Samples diagnosed as false-positives by the HPV16 RNA patterns (transformation-specific HPV16 RNA patterns but ≤CIN1 by histology) could indicate transformed cells not yet

identified by histology and thus could represent a diagnostic improvement with an earlier detection of progressing lesions and thus indicate higher sensitivity.

Thirteen HPV16 RNA patterns false-positive samples with histological follow-up data were available, eight multiple infected HPV16 DNA-positive CIN0 and CIN1, and five single infected HPV16 DNA-positive CIN1.

Of the eight patients with multiple HPV infections, two CIN0 lesions remained CIN0 and three CIN0 lesions and three CIN2 lesions, respectively, progressed to CIN2 and CIN3 (Figure 4.19). Four patients with single HPV16 infections progressed from CIN0 and CIN1, respectively, to CIN2 and CIN3, while one CIN1 lesions remained CIN1.

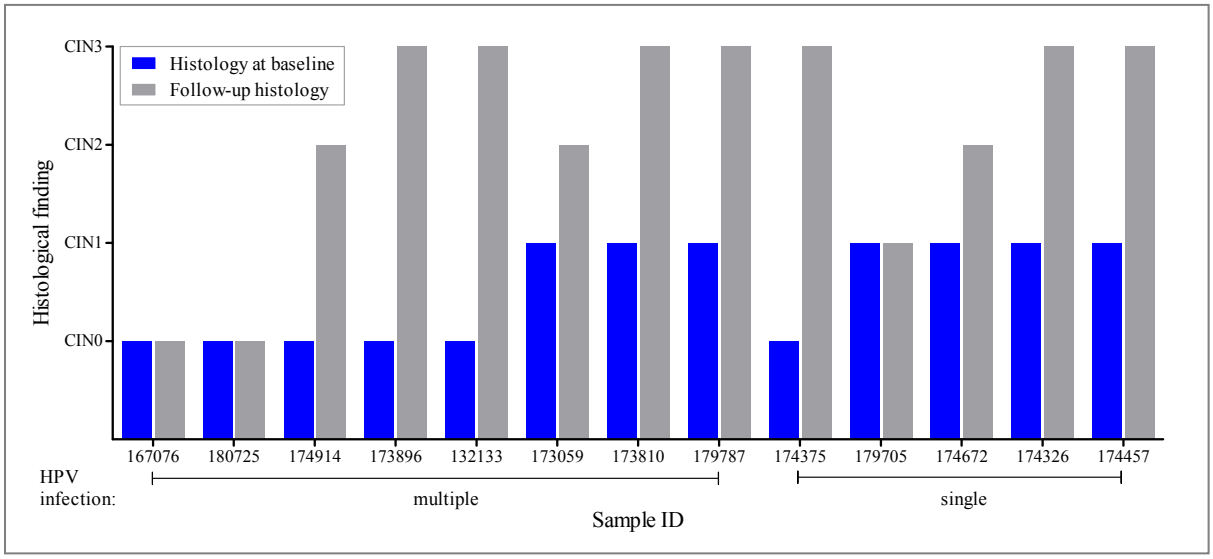


Figure 4.19. **Histological follow-up in HPV16 RNA patterns false-positive patients**. Histological diagnosis at base line (blue) and histological diagnosis at 5-24 months follow-up (grey) of two multiple- and two single-infected HPV16 DNA-positive patients with HPV16 RNA patterns indicating presence of severe lesions.

Thus in eleven of the thirteen patients with base line false-positive classification by RNA patterns, the mild histological lesion persisted in one case (follow-up after 5 month) and progressed in ten cases (follow up after 9-24 month). Longitudinal studies with higher sample numbers are required to further verify the increased predictive value of HPV16 RNA patterns in cervical cancer precursor screening.

4.1.8 HPV16 RNA patterns in FFPE biopsies

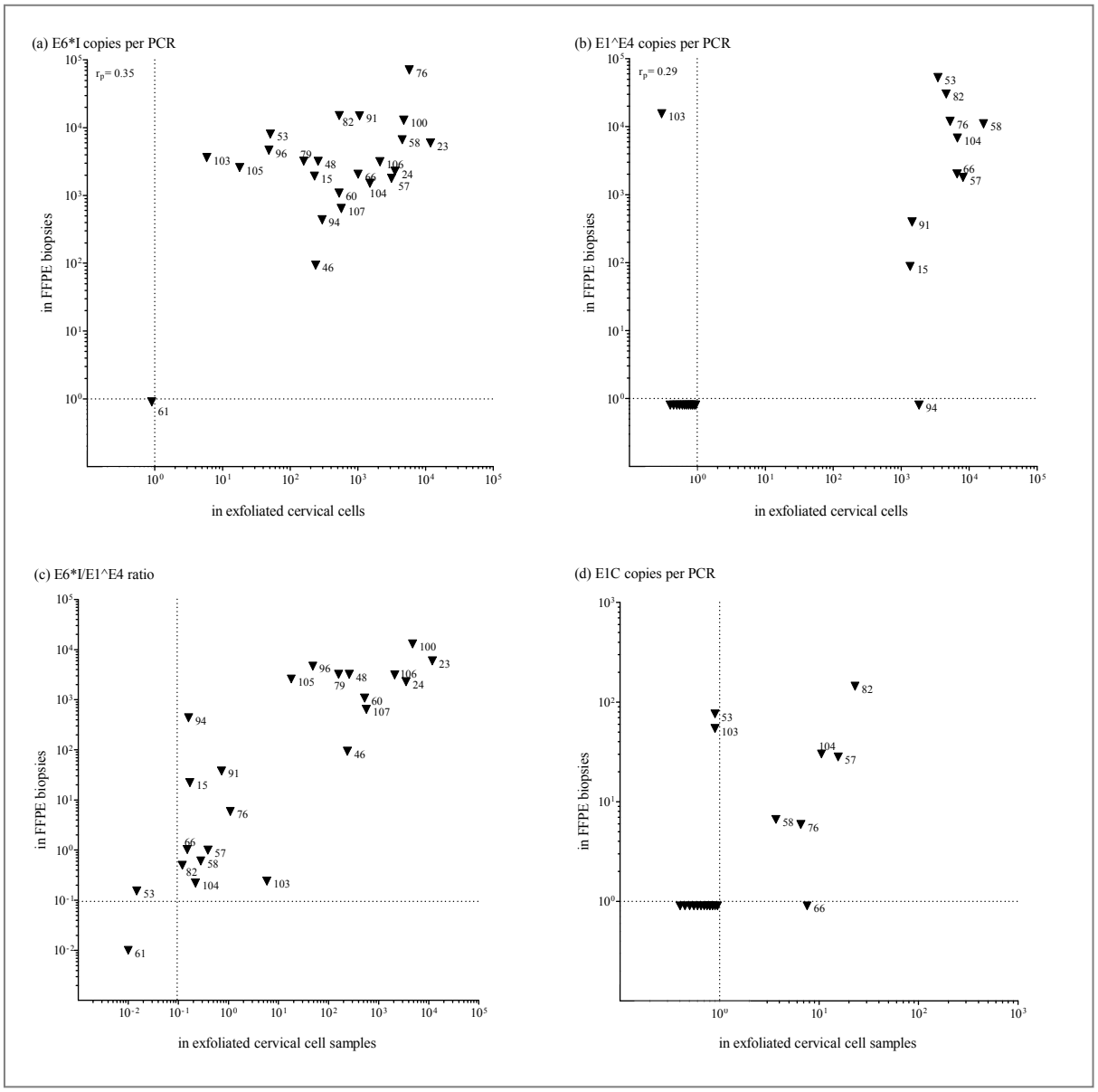
The aim of this chapter was to examine whether HPV16 RNA patterns analyses is also applicable to RNA extracted from formalin-fixed, paraffin-embedded (FFPE) tissue blocks, which would allow use of these tissue samples collected and processed in clinical and pathological standard procedures and also would give access to large sample banks for retrospective studies.

4.1.8.1 HPV16 RNA patterns in paired FFPE tumour biopsies and exfoliated cervical cells

For 27 Mongolian CxCa patients (18 singly HPV16-infected and 9 multiple HPV infected), RNA extracted from exfoliated cervical cells and FFPE biopsies as described in 3.1.2 were analysed by RT-qPCR$_{TP}$ and RT-qPCR$_{E1C}$ for transformation-specific HPV16 RNA patterns. Formalin fixation and size of biopsies followed a strict protocol as described elsewhere[169].

RNA analysis was valid (>80 ubC or >200 E6*I copies per PCR, chapter 3.12) in both exfoliated cervical cell and FFPE samples of 23/27 (85%) patients. UbC numbers were significantly lower ($p<0.0001$, Mann-Whitney-Test) in the FFPE biopsies than in the exfoliated cell samples with a median ubC copies per PCR of 54 (8 to 885 copies per PCR) and 863 (53 to 7,976 copies per PCR), respectively.

For E6*I and E1^E4 detection a moderate quantitative correlation was observed between paired exfoliated cervical cells and FFPE biopsies (r_p=0.35 and 0.29, respectively, Figure 4.20, a and b). The correlation was good for E1C (r_p= 0.73, Figure 4.20, d). In contrast to the quantification of single transcripts, the quantitative correlation of the E6*I/E1^E4 ratio was good (r_p=0.61, $p=0.0019$), indicating a high robustness of this normalisation for variable HPV16-positive cell numbers obtained from different specimens from the same patient.



(Continued on the following page)

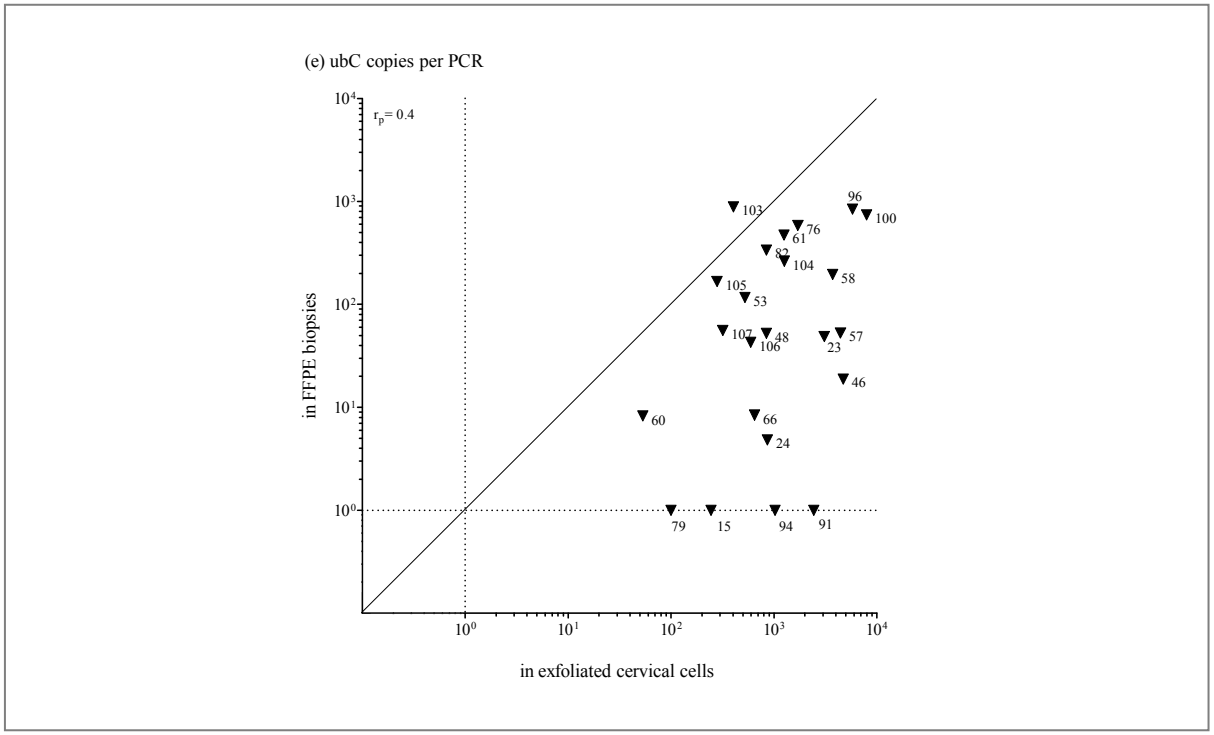


Figure 4.20. **E6*I (a), E1^E4 (b), E6*I/E1^E4 ratio (c), E1C (d) and ubC (e) copies per PCR in exfoliated cervical cells (x-axis) and in FFPE biopsies (y-axis) from the same CxCa patients**. All templates were applied in duplicates and mean Cp value was used for the calculation of copies per PCR. Negative values are indicated below 1 copy per PCR. E6*I and E1^E4-negative samples are indicated arbitrarily by a ratio of 0.00001. Dotted lines indicate cutoffs (0.095 for E6*I/E1^E4 ratio and Cp <42 for E6*I, E1^E4, E1C and ubC).

For the E6*I/E1^E4 ratio 22 RNA samples (96%) were concordantly positive (n=21) or negative (n=1) and one discordantly positive in FFPE biopsy only (κ=0.65; CI.95=0-1) (Figure 4.20, c) which might be due to a higher number of HPV-transformed cells in this sample compared to the smear. E1C was concordantly positive in 5 pairs and concordantly negative in 15 pairs. One sample was E1C-positive only in the cervical smear, which may be explained by degraded RNA in the corresponding FFPE sample containing 8 ubC copies only. Another two samples were E1C-positive only in the FFPE. One of those was irreproducible E1C-positive in smear (Figure 4.20, d-ID53) indicating lower number of HPV-transformed cells. This resulted in an E1C agreement of κ=0.68; CI.95=0.34-1 (Figure 4.20, d).

In total, 21 of 23 CxCa patients were concordantly grouped as having transformation-specific HPV16 RNA patterns, while one (ID61) was concordantly diagnosed as negative, despite high ubC RNA copy numbers in both samples (refer to 5.3.2 for further discussion). The remaining patient (ID53) was HPV16 RNA patterns-positive (E6*I/E1^E4 ratio plus

E1C) only in the FFPE sample resulting in a total concordance of 96% between exfoliated cervical cells and FFPE biopsies.

A good correlation of E6*I/E1^E4 ratio and E1C detected in FFPE biopsies and PreservCyt stored exfoliated cervical cells was observed. However, one CxCa samples was HPV16 RNA patterns-positive only with FFPE, suggesting that the corresponding cervical cell sample contained to little HPV-transformed cells. Consequently, FFPE biopsies obtained under same conditions regarding fixation and tissue size used in this Mongolian study can be applied to the HPV16 RNA patterns analyses.

4.1.8.2 HPV16 RNA patterns in FFPE biopsies of different storage histories

The RNA quality in FFPE biopsies depends among other things strongly on storage times of FFPE blocks. In this chapter, in addition to the Mongolian, FFPE biopsies from the archived Barcelona study including routine biopsies from 38 countries all over the world (Chapter 2.11) collected between 1929 and 2006[170] (Halec et al., submitted to JNCI[151]) were analysed for HPV16 RNA patterns. Criteria for including Barcelona biopsies in this thesis were positivity with the HPV16 E6*I mRNA assay and single HPV16-positivity by the BSGP5+/6+-PCR/MPG assay -consequently good DNA quality due to the long amplicon size of HPV. Besides the varying storage time, Barcelona biopsies differ in size and fixation time.

RNA extracted from 51 FFPE tissues from CxCa patients from Mongolia (n=27) and Barcelona (n=24) were applied in duplicates to RT-qPCR$_{TP}$ and RT-qPCR$_{E1C}$. While the FFPE tissues from the Barcelona study were collected between years 1929 and 2006, FFPE tissues from Mongolia were collected in 2005.

Only 13% of the Barcelona study (all stored since 1997-2006) and 78% from Mongolia were ubC-positive with a median of 1 and 117 copies per PCR, respectively (Table 4.10). In general, the prevalence of all transcripts was lower in samples from Barcelona (58%) compared to Mongolia (89%). Only 50% of FFPE tissues from the Barcelona study and 89% from Mongolia were E6*I-positive. Among these E6*I-positive FFPE tissues from Barcelona, 33% were stored since 1929 to 1993 and 66% since 1997 to 2006.

Table 4.10. **Frequency of single transcripts and HPV16 RNA patterns in CxCa FFPE biopsies from Barcelona and Mongolia.**

Positivity of		FFPE samples		
		Barcelona	Mongolia	Total
ubC				
	N	3/24	21/27	24/51
	%	13%	78%	47%
Any viral transcript				
	N	14/24	24/27	38/51
	%	58%	89%	75%
E6*I				
	N	12/24	24/27	36/51
	%	50%	89%	71%
E1^E4				
	N	7/24	11/27	18/51
	%	29%	41%	35%
E1C				
	N	0/24	7/27	7/51
	%	0%	26%	14%
HPV16 RNA patterns				
	N	2/3[a]	22/23[a]	23/26[a]
	%	67%	97%	89%

[a] samples with valid RNA only (ubC >80 or E6*I >200 copies per PCR)

Among the RNA valid FFPE tissues, the percentage of HPV16 RNA patterns-positives was 97% in Mongolian samples and 67% in Barcelona samples.

This data indicates that the HPV16 RNA patterns analyses cannot be applied to routine FFPE biopsies as long as no consistent protocol is used similar to that used in the Mongolian study.

4.2 Other hrHPV types

After HPV16, the hrHPV types 18, 31, 33, 35, 45, 52 and 58 most prevalent in CxCa account for about 28% of cervical cancer cases[71].

In preparation for RNA patterns analyses for these HPV types, in this chapter, the identification of splice junctions analogous to the HPV16 E1^E4, E1C and L1 splice junctions was performed, followed by first singleplex HPV18 RT-qPCR experiments.

4.2.1 Identification of splice junctions

While some splice junction had already been described in the literature (Table 4.11) E1^E4, E1C and L1 splice donor and acceptor site for four hrHPV types 35, 45, 52 and 58 as well as the E1C and L1 splice donor and acceptor site for two hrHPV types 33 and 18 were unknown.

Seventy-two exfoliated cell samples were selected from the Mongolian population-based study harbouring types of interest as single or multiple infections. The criterion for selection was HPV DNA MFI values for the type of interest in BSGP5+/6+-PCR/MPG assay above 120.

For all the seven hrHPV types (HPV18, 31, 33, 45, 35, 52, 58) the locations of the detected or confirmed E1C, E1^E4 and L1 splice sites were homologous to those of HPV16 (Table 4.11).

Table 4.11. **Position of splice junctions of selected hrHPV types**.

HPV type	E1C splice junction	E1^E4 splice junction	L1 splice junction
16	880^2582	880^3358	3632^5639
18	929^2779[a]	929^3434[b]	3696^5613[a]
31	877^2646[b]	877^3295[b]	3590^5552[c]
33	894^2702[a]	894^3351[b]	3589^5597[a]
45	929^2737[a]	929^3392[a]	3360^5608[a]
35	883^2649[a]	883^3298[a]	3575^5574[a]
52	879^2696[a]	879^3345[a]	3625^5643[a]
58	898^2706[a]	898^3355[a]	3608^5647[a]

[a] newly identified, [b] confirmed as described in the literature[161,162,171], [c] described in the literature without confirmation in this thesis[162,161]

Sequencing results demonstrated that all seven hrHPV types of interest undergo splicing of their E1, E4 and L1 gene.

4.2.2 Singleplex HPV18 RT-qPCR

In the following experiments, designed primers and TaqMan probes detecting the HPV18 transcripts E6*I, E1C, E1^E4 and L1 were applied in singleplex RT-qPCR to test specificity and DL.

Serial dilutions of the HPV18 *in vitro* transcripts E6*I, E1C, E1^E4 and L1 with copies per µL ranging from 1 to 10^6 as well as 10^6 for genomic HPV18 DNA were applied in duplicates to singleplex RT-qPCR using different combinations of primers and TaqMan probes.

For all transcripts, primers had to be relocated relative to the HPV16 targets to reach a sensitive, efficient and specific detection (Table 4.12). For E6*I, one of seven combinations tested, detected 10 copies per PCR with a median Cp value of 41. The R^2 induced by this E6*I was 0.99 indicating a perfect linear correlation. In addition, E was 1.91. The detection of 10^5 E6*I copies per PCR was observed from cycle 26 onwards. The same parameters are depicted for the other HPV18 transcripts (Table 4.12).

Table 4.12. **Detection limit and specificity of spliced HPV18 transcripts.**

Transcript	fw primer	bw primer	TaqMan probe	Nr of combinations tested	DL[a] (copies per PCR)	Median Cp at DL	R^2 [b]	E[c]	Median Cp at 10^5 copies per PCR	Sensitive& efficient[d]	Specific[e]
E6*I	1/3[f]	1/3	1/1	7	10	40.96	0.99	1.91	26.7	yes	yes
E1C	1/2	1/2	1/1	5	10	41.73	1.00	2.02	29.35	yes	yes
E1^E4	1/3	1/2	1/1	6	100	34.89	1.00	2.11	25.69	no	yes
L1	1/1	1/2	1/1	2	100	36.67	0.98	2.07	28.46	no	yes

[a] detection limit (DL) defined as lowest copy number tested with signals above background in duplicates, [b] coefficient of determination (R^2) describes the correlation between crossing point (Cp) and the log of transcript copy number per PCR, [c] PCR efficiency (E); under optimal conditions PCR run with an efficiency of 2, meaning that the number of target molecules doubles with every PCR cycle, [d] defined as sensitive and efficient if DL at most 10 copies per PCR, [e] defined as specific if no cross-reaction with fl HPV18 was observed, [f] tested primer/efficient primer and probe, e.g. three different forward primer and three different backward had to be designed and tested for the amplification of E6*I und only one forward and one backward primer worked sensitively, efficiently and specifically

After relocation of primers, two out of four transcripts were sensitively, efficiently and specifically detected with DL of 10 copies per PCR. For E1^E4 and L1 more experiments are required in order to improve DL.

5 Discussion

5.1 Rationale

A major limitation of current HPV-based cervical cancer precursor screening programs is that they detect infection rather than disease. This leads to poor clinical specificity and low positive predictive value for precancerous lesion resulting for many women in unnecessary anxiety, unnecessary referral to costly follow-up and/or overtreatment[138-140].

This thesis aimed at transferring a recently described HPV16 RNA patterns assay[87] on a new technological platform (RT-qPCR) and validating its potential for substantially improving the clinical specificity of HPV-based tests by discriminating between the presence of mild (cytological NIL/M and LSIL and histological CIN0 and CIN1) and severe cervical neoplastic lesions (cytological HSIL and histological CIN3 and CxCa). CIN2 samples representing rather an intermediate between mild and severe frequently were excluded from test qualification analyses. Furthermore, the various RNA patterns observed in severe lesions in the context of HPV-DNA status and transformation mechanism is discussed. The thesis also included preparatory work to extend the HPV RNA patterns analysis to hrHPV types 18, 31, 33, 35, 45, 52 and 58.

The HPV16 RNA patterns analysis comprises the quantification by RT-qPCR (Chapter 4.1.2.1) of the spliced viral transcripts E6*I, E1^E4, E1C and initially comprised also L1 (modification will be discussed in 5.2.3) that display different expression levels in mild versus severe lesion grades. An important event in cervical carcinogenesis is the deregulation of the oncogene E6 and E7 expression[81,82,172]. However, the underlying mechanisms for the deregulation are diverse and genome integration, viral DNA methylation as well as the involvement of HPV16 E1C protein are discussed: Integration of HPV DNA often leads to a loss of E2 expression and the subsequent deregulation of oncogene E6 and E7 expression[21,173,174]. Thus, expression of E6 and E7 from the p97 promoter increases in the absence of high amounts of E2, which can act as a transcriptional repressor. The integration status of HPV16 might be assessed by the ratio of the amount of E6*I copy numbers and the amount of E1^E4 copy numbers[87,146]. For these cases with intact E2, accounting for ~60% of CxCa[83], other mechanisms for the deregulation of the E6 and E7 oncogenes must exist. Methylation of the viral URR might be such an alternative mechanism that is intensively studied at the moment[82,175,176-178]. Chaiwongkot and colleagues described that the methylation

level of the E2 binding sites (E2BS) in the HPV16 URR was higher in all lesions with only episomal HPV16 genomes compared with lesions displaying single integrated copies[178]. In addition, samples with multiple integrated HPV16 copies displayed high methylation levels for all E2BS suggesting that the majority of the copies were silenced by extensive methylation. However, recent data from our laboratory indicates that also a substantial fraction of CxCa samples with multiple HPV copies per cell did not show any methylation of the E2BS (Holzinger et al., manuscript in preparation). Thus, at least a third mechanism should exist to deregulate E6 and E7 oncogenes. The third mechanism might be seen in an upregulation of the viral transcript E1C that is hardly detectable in mild lesions but frequently detected in severe lesions[87,146,179]. E1 with a truncated N-terminus (E1C) is translated from 880^2582 spliced mRNA and has been described to be a potential trans-activator of the viral URR, promoting the upregulation of E6 and E7[17]. Data from our laboratory indicated that the fraction of CxCa samples with multiple HPV copies per cell which did not show methylation of the E2BS were E1C-positive (Holzinger et al., manuscript in preparation). Consequently, in addition to methylation, E1C might be responsible for an E2 deletion-independent mechanism of HPV16-mediated transformation of cervical cells.

5.2 HPV16 assay development

5.2.1 RT-qPCR versus NASBA

RT-qPCR greatly enables simple amplicon recognition by using fluorescent reporter molecules to monitor the level of amplicons after each PCR cycle. Consequently, no post-PCR processing is required and the analysis is performed in a closed system which can then be disposed of without a contamination risk for the laboratory environment. In addition, the wide dynamic range allows the analysis of samples differing in target abundance by several orders of magnitude. Furthermore, there is little inter-assay variation, which helps generating reliable and reproducible results[180] (see chapter 4.5).

RT-qPCR also allows an although limited degree of multiplexing, an important step towards high-throughput analyses in large studies and diagnostic applications. Ergo, large numbers of samples can be screened within a short period of time and at rather low costs, therefore, leading to better patient management.

In contrast, the NABSA assay comprised two separate reactions, amplification plus hybridisation and multiplexing could not be implemented despite extensive efforts (personal communication Dr. Markus Schmitt). Additional disadvantages of NASBA were the high reagents cost per sample (~20€ per NASBA reaction) and the enormous amount of time and labour required for conducting the assay (for 90 samples 7-8 hours per single NASBA with subsequent hybridisation).

In this thesis, multiplex RT-qPCR were developed using the Cobas z480 platform as promising robust, sensitive, specific and cost-efficient technology for detecting the HPV16 RNA patterns.

5.2.2 Design of RT-qPCR primers and TaqMan probes

For a specific and sensitive amplification of the spliced HPV16 transcripts via RT-qPCR, primers and probes were newly designed (see 4.1.1) and sequences were altered to optimise DL and specificity. Primer and probe sequences were designed according to strict rules[181], however, for E6*II a DL better than 10^3 copies could not be reached, although 12 different primer/probe combinations were tested. With NASBA detection the E6*I and E6*II transcript numbers had been strongly correlated and a replacement of E6*II by E6*I in the E6*II/E1^E4 ratio led to almost identical results. In addition, the E6*I expression is around ten-fold higher than for E6*II, which may be beneficial when analysing poor quality samples[146,136,167]. Finally, with regard to extending RNA patterns analysis to other hrHPV types that express only E6*I but no E6*II, the E6*II RT-qPCR was replaced by E6*I RT-qPCR.

To improve DL of splice site-specific detection of E1C, two strategies were tested: either the forward primer or the probe covered the splice junction. Primers covering the splice junction led to unspecific amplification and detection of genomic HPV16 DNA. This could be reduced but not eliminated by shortening the primer 3´ end. The shortening reduced the number of matches between HPV16 DNA and primer from 12 to 5 nucleotides. Still, 10^5 genomic HPV16 DNA copies were detectable (see chapter 4.1.2.2). In the final assay, probes were included covering the splice junction. The sensitivity could be increased by a semi-nested detection of E1C which included two primer pairs in one reaction, one outer and one inner primer pair, both flanking the splice junction.

As mentioned above, E2 deletion and, thus, HPV integration was monitored using the E1^E4 transcript as surrogate marker. The E1^E4 splice donor is located at nt 880 within the E1 ORF, while the splice acceptor is located at nt 3358 within the E2 and E4 ORF. Breakpoints of integrated HPV16 DNA are located within the E1 ORF (59%), the E2 ORF (31%), the E1-E2 overlap region (3%), the E5 ORF (7%) and the L1 ORF (1%)[85,182]. Consequently, 92% of viral integrates show E2 ORF deletions. The remaining 8% of CxCa with intact E2 ORF despite integration must depend on different mechanisms of E6 and E7 upregulation (see discussion 5.1). Using RT-qPCR$_{TP}$, 94% of the E2 deletions were recognised as having E2 deletions. This was made possible by placing the backward primer binding within the E2 and E4 ORF region as close to the E5 ORF as possible. However, to cover all cases with E2 deletions, the backward primer would have to be moved downstream for an additional 230 nucleotides. The present E1^E4 amplicon size of 199 bp exceeded already the ideal amplicon length in RT-qPCR, which is around 150 bp. Longer amplicons are less efficiently amplified and are prone to false-negative results in case of RNA template degradation. Consequently, the current E1^E4 RT-qPCR represents a compromise between detecting as many HPV integrates with E2 deletions as possible and the technical limitations of the RT-qPCR.

To make the assay as sensitive as possible, potential mismatches of known HPV16 variants with primer and probe sequences were minimized in order to reduce number of false-negative results. Placing the primers in conserved regions, identical between HPV16 variants, minimized the possibility of false-negative results. To this end, alignments of target regions were performed with all variants of HPV16, available in public data bases. Degenerate primers or probes, i.e. a pool of oligonucleotides which differ slightly in sequences were ordered if more than 1% of HPV16 isolates had sequence diversity at a specific position. Thus, two E1C primers and the E1C and E1^E4 TaqMan probes were designed with one degenerate position each. E6*I primers and probes were not designed with degenerate positions since sequence variations were present in only 0.3% of HPV16 isolates.

To test the analytical specificity of the assay, undesired amplification of genomic HPV16 DNA, human placenta DNA and human RNA as well as cross-reactivity among the transcripts were examined. Unspecific detection was successfully eliminated by redesigning primers and probes as it was described for E1C (Chapter 4.1.2.2). Cross-reactivity with other HPV types was excluded by *in silico* analysis using BLAST at the NCBI homepage (data not shown). As expected[183], the identity within the E1^E4, E1C and E6*I transcript regions

between HPV35 and HPV16 was 95%, and 80% between HPV31 and HPV16. Primers and probes had 2 to 10 mismatches within the primer and between 4 and 6 mismatches within the probe regions. However, the degree of cross-reactivity between these highly related HPV types still needs to be tested experimentally in the future.

5.2.3 Fiveplex RT-qPCR

Development of a fiveplex RT-qPCR enabling the simultaneous quantitative detection of the four HPV16 transcripts E6*I, E1^E4, E1C and L1 and of the cellular ubC transcript was the initial goal of this thesis. A fiveplex RT-qPCR would have allowed stratifying clinical HPV16 DNA-positive specimens into mild or severe lesions by a single reaction.

By changing concentrations of PCR reagents, the final DL for each of the *in vitro* transcripts analysed individually in fiveplex RT-qPCR ranged between 10 and 10^2 which was comparable to singleplex RT-qPCR. Nevertheless, the presence of multiple amplifiable transcripts in one sample, led to reduced detection of low-abundance transcripts E1C and L1 (see 4.1.3.2). An explanation might be found in the fact that co-amplification of highly expressed transcripts, such as E6*I and E1^E4 reduced the amount of available PCR reagents, to the detriment of these low-abundance transcripts. As a consequence the fiveplex was split into two separate RT-qPCR, the RT-qPCR$_{TP}$ for E6*I, E1^E4 and ubC and the RT-qPCR$_{DP}$ for E1C and L1. However, a reciprocal interference of the low-abundance transcripts was still observed once one transcript was present in higher copy number than the other (see 4.1.3.5).

In the NASBA pilot study[87] as described in chapter 4.1.3.7, L1 had contributed only little to the cervical precursor diagnosis: Applying the E1C only algorithm instead of the E1C/L1 ratio, one out of seven CxCa and three out of twenty-four HSIL samples (13%) would have been classified falsely as mild lesion (Figure 4.8). Consequently, these four samples were classified as severe lesions only by the presence of very low amounts of L1 and the absence of E1C. L1 is the major protein forming the viral capsid and is not actively involved in cancer progression. Thus, it was questionable whether the absence of L1 alone should be used for diagnosing severe lesions. However, it is tempting to speculate that the approximately 12-fold higher sensitivity of the RT-qPCR$_{E1C}$ might have identified these 4 samples as EIC-positive. Consequently, L1 was omitted in the RT-qPCR$_{DP}$ resulting in final RT-qPCR$_{E1C}$. A normalisation with another transcript was not necessary since E1C was nearly exclusively

detected in CIN2, CIN3 and CxCa (Chapter 4.1.7.4), meaning once E1C is detected a severe lesion can be diagnosed independent of the number of copies.

As a consequence, in the final setting E6*I, E1^E4 and ubC were quantified in RT-qPCR$_{TP}$ and E1C in RT-qPCR$_{E1C}$.

5.2.4 RNA extraction

Sensitivity of the RT-qPCR is also dependent on the RNA quality in the tested samples and by the presence of inhibitors. Primers and probes detecting ubC were integrated in the RT-qPCR for controlling the functionality of the assay and the quality of the RNA template. In clinical samples, a cutoff of 80 ubC copies per PCR was applied.

Different extraction methods of RNA including *MagNA Pure96 DNA and Viral NA Large Volume Kit* and *MagNA Pure96 Cellular RNA Large Volume Kit* as well as different extraction volumes (1 mL and 4 mL) were compared to achieve an optimal extraction of RNA. Best data was obtained with RNA extracted from 4 mL sample volume using the *MagNA Pure96 Cellular RNA Large Volume Kit*. The ubC quantification was 1.7- and 4-times higher under these conditions compared to 1 mL and *MagNA Pure96 DNA and Viral NA Large Volume Kit*, respectively (Chapter 4.1.4).

5.2.5 Comparison of RT-qPCR and singleplex NASBA in cell lines and OPSCC samples

To compare the previously established singleplex NASBA assays with the newly developed RT-qPCR assays, RNA from cervical cancer cell lines MRI-H186, MRI-H196, SiHa and CaSki and from 32 OPSCC analysed previously by NASBA was here additionally analysed by RT-qPCR. NASBA data previously published for these samples were kindly provided by Dr. Markus Schmitt (DKFZ)[31,148] and by Dr. Dana Holzinger (DKFZ)[152].

When interpreting differences in quantification by NASBA and by RT-qPCR, it has to be kept in mind that the RT-qPCR was excellently reproducible even with RNA extracts obtained at different dates (Chapter 4.1.5) while NASBA showed large copy number variation (Chapter 4.1.6.1).

In cell lines, the mean transcript copy number per cell for E6*I, E1^E4 and ubC by NASBA appeared to be up to fifty-fold and for E1C up to 10^5-fold higher than by RT-qPCR.

However, the E6*I, E1^E4 and ubC copies per cell quantified by RT-qPCR were within the copy range of the NASBA data. In addition, the quantification of transcripts by NASBA used ten-fold dilution steps over four to five orders of magnitude complicating the direct comparison to RT-qPCR (Chapter 1.7). The dose-response curve of NASBA standards that were almost linear for E1^E4 and E6*I but polynomal for E1C complicating the accurate quantification and might be an explanation for the huge difference especially for E1C quantification in NASBA and RT-qPCR.

Consequently, the RT-qPCR appears to be more reliable than the previous NASBA data.

MRI-H186 cells have been described to contain around 26 integrated full-length as well as truncated HPV16 genome copies[184]. Consistently, RT-qPCR quantified a low amount of E1^E4 (mean 8 copies per cell) and high E6*I (mean 76 copies per cell) and the calculated E6*I/E1^E4 ratio was above cutoff (ratio-positive). E1C was also detected with 0.002 copies per cell.

SiHa cells contain one integrated HPV16 genome with disruption of the E2 region and deletion of bases nt 3132 to nt 3384[22]. Consistently, we found no transcripts containing the splice acceptor at nt 3358 but we found the E1C transcript having the splice acceptor at nt 2582.

In contrast, CaSki cells contain around 600 truncated but also full-length HPV16 genomes at multiple integration sites[22,185]. Here, all three viral transcripts were detected, with a mean of 546 E1^E4 copies and 29 E6*I copies per cell, thus E6*I/E1^E4 ratio-positive. E1C was detected with 0.0004 copies per cell.

Taken together, the quantitative HPV16 expression data correlated well with the known integration status of the HPV16 genome.

The conspicuously low ubC copy number per cell (RT-qPCR: median 90 copies per cell, range 30-400; NASBA: median 375, range between 0.75-750) was found by both methods. Unfortunately, no reference information was found in the literature.

For defining OPSCC as CxCa-like RNA patterns-positive ($CxCaRNA^+$) and as CxCa-like RNA patterns-negative ($CxCaRNA^-$), a 100% concordance was achieved between NASBA and RT-qPCR.

However, the concordance for single transcripts was lower: E6*I positivity detected by RT-qPCR and E6*II positivity by NASBA, respectively, were concordantly classified in 25 OPSCC samples (78%), seven further samples were E6*I positive by $RT\text{-}qPCR_{TP}$ only (κ=0.2, CI.95=0-0.7). This could be attributed to the higher level of E6*I than E6*II in HPV16-positive cells[136,167]. For E1^E4, 26 OPSCC samples were concordantly classified and 6 were discordantly, although only borderline positive by RT-qPCR or NASBA (κ=0.4, CI.95=0.1-0.8). Nevertheless, the quantitative correlation of these two transcripts detected by both methods with an r_p=0.94 for E6*I was excellent and moderate with an r_p=0.69 for E1^E4.

The quantitative correlation of E1C was low (r_p=0.15) and as mentioned above can be explained by the fact that NASBA is at best semi-quantitative for E1C while $RT\text{-}qPCR_{E1C}$ enables a precise quantification over 5 logs (Figure 4.12). By NASBA and by RT-qPCR E1C copy numbers were much lower than those of E6*I or E1^E4. However, the transcript copies per reaction for E1C by NASBA ranged between 2,500 and 250,000 (median 25,000) and by RT-qPCR between 5 and 21,000 (median 33). Since the initial E1C copy numbers detected by RT-qPCR were so low, a new batch of reference *in vitro* transcripts as well as the reference titration curve of the transcript dilutions were repeated and the validity and accuracy of the E1C transcript quantification by RT-qPCR was confirmed. Besides the semi-quantitative detection of E1C, the low reproducibility of NASBA could be an explanation for the quantitative discrepancy between the two methods: E1C quantification of the same samples in two independent experiments by NASBA showed a thousand-fold difference ranging between 0.08-0.75 and 75-750. Similarly, applying the E1C standard curve to NASBA reaction twice had coefficients of variation (CV) for the six dilution steps between 2 to 140%. Nonetheless, between both methods the concordance of classifying OPSCC samples as E1C-positive or E1C-negative was very high (κ=0.9, CI.95=0.71-1).

5.3 HPV16 RNA patterns in clinical specimens

5.3.1 Cell concentration and RNA quality in validation samples

Of all three viral transcripts E1C shows the lowest abundance frequently close to the detection limit and thus is most sensitive to low concentration of reverse-transcribable and amplifiable RNA due to low number of HPV-infected cells or RNA degradation.

In conventional cytological diagnoses the number of abnormal cells in the sample is also critical for true-positive diagnosis[186,187]. Cell numbers in exfoliated cervical cell samples determined by qPCR of human β-globin show huge variation.

Twenty mL of PreservCyt-stored cervical cell samples from Bad Münder (n=40) contained only a median of 76,969 (mean of 155,562 with a minimum of 314 and a maximum of 1,056,806) cells while samples from Mongolia (n=20) contained a median 5,602,338 (mean of 10,213,686 with a minimum of 275,246 and a maximum of 50,659,658) cells (personal communication Dr. Markus Schmitt, DKFZ). No reference information for absolute cell content of exfoliated cervical cell samples was found in the literature. With this data, the median number of cells applied to a RT-qPCR assay was 385 cells in the Bad Münder and 28,012 in the Mongolian samples.

Quantification of ubC was integrated into the RT-qPCR to monitor RNA quantity and thus to control the sum of cell content, RNA extraction, RT-qPCR performance, and RNA quality since ubC is the longest amplicon in the assay (238 bp).

The ubC transcripts observed in the 158 single-infected HPV16 RNA-positive cervical cell samples ranged between 54 and 69,615 copies per PCR. Surprisingly, despite the different cell concentrations in the samples, differences in the ubC quantification between PreservCyt-stored exfoliated cervical cells from Bad Münder and from Mongolia were insignificant (p=0.7, Mann-Whitney-Test). This indicated strongly increased RNA degradation of the Mongolian samples archived for about 84 months compared to the Bad Münder samples archived for 9-24 months. Low RNA quality of all archived PreservCyt-stored exfoliated cervical cell samples was also confirmed by a median RIN factor of only 2.3 (see 4.1.7.3). It has been described that a reduction of RIN from 9 to 2 resulted in a reduced quantification by a factor of approximately 20[188,189]. However, the relative to the Bad Münder samples further

increased RNA degradation in the Mongolian samples did not result in even further reduction of the RIN factor.

The median number of ubC transcripts per cell (ubC copies per PCR/number of cells per PCR) was 4 in the Bad Münder and 0.31 in the Mongolian exfoliated cervical cell samples. However, since the median RIN of the Bad Münder samples was 2.6, the ubC quantification, as mentioned above, was presumably 20-times reduced, thus, theoretically, 80 ubC transcript copies per cell could be present and would be similar to the number of ubC copies detected in cell lines (Chapter 5.2.5). However, these calculations can only give a rough idea about the RNA quality of cervical samples.

Still, the same cutoff of 80 ubC copies per PCR as marker of minimal absolute-number of amplifiable RNA can be used for Bad Münder as well as for Mongolian samples. In this study comprising highly degraded RNA samples, the number of E6*I copies (>200) was additionally included for defining a sample as valid. With it, the number of analysable samples was increased by two HPV16 RNA patterns-positive CIN3 and CxCa but also by one HPV16 RNA patterns-negative CIN2. The definition for validity could also have been the following: >80 ubC copies or E6*I/E1^E4 ratio positivity with E6*I >200 copies or E1C positivity. Retrospectively, this definition would include another HPV16 RNA pattern-positive CIN3. However, using non-archived material probably needs another definition for validity and probably does not need the inclusion of E6*I.

In summary, poor sampling and a too long storage period of cells even in PreservCyt buffer can both lead to false-negative results and this has to be kept in mind when interpreting data resulting from archival exfoliated cervical cells. Furthermore, in future studies RNA quality in samples stored for shorter time periods has to be assessed to define the maximum storage period suitable for HPV RNA patterns analyses.

5.3.2 HPV16 RNA patterns in FFPE biopsies

Expanding the application of the HPV16 RNA patterns to formalin-fixed paraffin-embedded FFPE would allow use of the tissue samples collected and processed in clinical and pathological standard procedures and also would give access to large sample banks for

retrospective studies. The major limitation of the use of FFPE samples is potential RNA degradation and cross-linking.

While 85% of the Mongolian FFPE samples collected 7 years ago under strict study conditions were RNA valid, FFPE biopsies collected between 1929 and 2006 from nine different countries under clinical and pathological standard procedures (not using the same protocol for fixation-time and size of biopsy) were RNA valid in only 13%.

Nevertheless, HPV16 RNA patterns classification of RNA extracted from FFPE biopsies and from exfoliated cervical cells of the same single and multiple HPV16 DNA-positive Mongolian CxCa patients was compared. A total of 21 of 23 CxCa patients were concordantly diagnosed as having severe lesions, while one (ID61) was concordantly diagnosed as mild lesion and one sample (ID53) was positive in the FFPE biopsy only. In ID61 HPV involvement is questionable in the absence of $p16^{INK4a}$ upregulation and p53 and pRb downregulation (immunohistochemical data of Dr. Gordana Halec, DKFZ[12]). The discordant sample (ID53) was diagnosed as severe lesion by the E1C presence in the FFPE biopsy. In the cervical smear RNA EIC was detected in only one of the duplicates and with only 12 copies, indicating a probably insufficient number of cancerous cells in the smear.

Adequately fixed PreservCyt samples remain the targets of HPV16 RNA patterns in primary screening. However, to expand the application to FFPE samples and thus, to enable the amplification of partially degraded and cross-linked RNA, a complete re-design of the RT-qPCR with shortened amplicon sizes would be necessary. Nevertheless, as discussed above, shortening the E1^E4 amplicon, i.e. moving the backward primer upstream led to a reduced detection of HPV integrates and consequently, to a reduced analytical and thus also clinical sensitivity of the HPV16 RNA patterns test.

5.3.3 Prevalence and amount of HPV16 transcripts in validation samples

The prevalence and expression of the three transcripts E6*I, E1^E4 and E1C quantified by the novel RT-qPCR in 158 single-infected HPV16 RNA-positive cervical cell samples were compared to previously described prevalence and expression quantified by singleplex NABSA assays in a study of 77 HPV16 RNA-positive cervical cell samples[146]. Discrepancies are summarized and discussed.

As a quick reminder: E6*I and E1^E4 prevalence was between 82 and 95% in CIN1-3, increased in CxCa to 100% for E6*I but decreased to 33% for E1^E4 (Table 4.9). This is in concordance to Sherman et al., 1992[20] and Schmitt et al., 2011[146], where the prevalence of E6*I and E1^E4 was also not specifically associated with lesion stage (CIN1, CIN2 or CIN3), indicating the need for E6*I and E1^E4 quantification.

In contrast, E1C was observed almost exclusively in CIN2 (17%), CIN3 (52%) and CxCa (22%) patients in this thesis (Table 4.9). As a consequence, the qualitative detection of E1C appeared to be a specific albeit alone not a sensitive marker for severe lesions. The reduced E1C prevalence in CxCa patients compared to CIN3 is contradicting the NASBA data where 57% of CxCa were E1C-positive[87,146]. One explanation might be that NASBA data included CxCa samples from France with a short storage time, while CxCa samples in this study were collected from Mongolian women 7 years before RNA extraction. Thus, RNA degradation could have caused the lower sensitivity of E1C detection in the 7-years old CxCa samples. Another explanation could be a low number of cancer cells in the Mongolian cervical smear compared to France and Bad Münder due to sampling device or expertise and consequently a lack of E1C-positive cells. Supporting the assumption that less E1C-positive cells were present in Mongolian samples, in Mongolian CxCa a reduced mean of E1C copies per PCR was observed compared to CIN2 and CIN3 from Bad Münder. Nevertheless, only 4 out of 18 Mongolian CxCa samples were E1C-positive with copy numbers similar to the lower quartile of E1C-positive CIN2 and CIN3.

E6*I, E1^E4 and E1C were quantified by RT-qPCR. As a quick reminder: E6*I and E1C encoding transcripts were strongly upregulated in severe lesions, whereas the level of E1^E4 was similar in mild and severe lesions (Figure 4.13) which is in concordance with the NASBA data[146]. In contrast to the E6*I and E1^E4 transcripts, the E1C copy number was always low and ranged between 4 and 156 (median=16). Thus, less than one copy per cell was present under the incorrect assumption that all analysed cells were HPV-infected. However, from the literature it is known that in a cervical smear diagnosed as HSIL lesion only up to 2% of cells are HPV-infected[147]. Under the assumption that 4 ubC transcript copies are present per cell (Chapter 5.3.1), a median of 425 cells were analysed in all E1C-positive CIN2, CIN3 and CxCa samples (calculated as median 1,701 ubC copies per PCR/4 ubC copies per cell). With 2% of cells HPV-infected (8 out of 425 cells) the number of E1C per

HPV-infected cell would range between 1 and 20 (median=2). That underlines the importance of a high quality cervical smear in order to make a correct classification that cannot be compensated by RT-PCR.

5.3.4 Sensitivity and specificity of HPV16 RNA patterns

In order to validate the HPV16 RNA patterns as diagnostic marker for severe cervical neoplastic lesions, RNA was analysed from the 158 single-infected HPV16 RNA-positive, histologically and/or cytologically defined exfoliated cervical cell samples already presented above.

In total, RT-qPCR-based HPV16 RNA patterns analysis correctly classified 88% of ≥CIN3 as severe lesions (sensitivity) and 84% of ≤CIN1 as mild lesions (specificity). CIN2 was categorised as intermediate stage with 49% HPV16 RNA patterns-positive and not classified within the severe lesions.

The complexities of HPV16 RNA patterns in severe lesions are discussed in 5.3.5 and discordant samples with either false-negative or false-positive lesion classification in 5.3.6.

In conclusion, RT-qPCR-based HPV16 RNA patterns analysis was confirmed as diagnostic tool to distinguish mild from severe neoplastic cervical lesions.

In multiple-infected exfoliated cervical cell samples (n=185), i.e. samples containing HPV16 and at least one other hrHPV type, the fraction of correctly classified severe lesions decreased from 88 to 70% and the fraction of correctly classified mild lesions from 84 to 79%. The reduced sensitivity probably is due to cases transformed by the other hrHPV type rather than by HPV16.

5.3.5 Complexity of HPV16 RNA patterns in single-infected samples concordantly classified as severe lesions

HPV16 RNA patterns analysis may indicate the viral genome integration status by the ratio of E6*I versus E1^E4 and the presence of HPV16 E1C. Loss of E1^E4 was used as surrogate

marker for integration and consequently, loss of E2 repressor, at least in 94% of cases covered by the RT-qPCR (see discussion 5.2.2).

Exfoliated cervical cell samples defined by HPV16 RNA patterns as severe lesion fall in 3 groups: group 1 is only E1C-positive (Cp<42), in group 2 the E6*I/E1^E4 ratio is positive (above cutoff$_{E6*I/E1^E4}$=0.095) and in group 3 both, E1C and the E6*I/E1^E4 ratio are positive. This chapter focuses on these three groups with regard to the HPV-DNA status and proposes a possible explanation of how lesions developed.

HPV-DNA status can only be assumed from transcriptional active DNA, the additional presence of transcriptionally inactive episomal and integrated full-length HPV16 genomes cannot be excluded.

Of the 158 single-infected HPV16 RNA-positive ≥CIN3 12% were group 1, 45% were group 2 and 32% were group 3 (Chapter 4.1.7.8).

In group 1 E1C is present as well as E1^E4, the latter leading to a low E6*I/E1^E4 ratio. This indicates that at least the region upstream of splice acceptor nt 3358 must be intact (Figure 5.1). So, E1C is expressed from full-length integrated or episomal HPV or, if truncation occurred, it must be downstream of nt 3358. With this DNA-status, also E2 could be expressed. Expressed E2 could counteract E1C. I have observed higher E1C copies in group 1 with a mean of 51 as in group 3 with a mean of 21 (p=0.0183, Mann-Whitney-Test). This could indicate that higher E1C copy numbers in group 1 are needed to counteract E2 and thus further supports a function of E1C in maintenance of transformation.

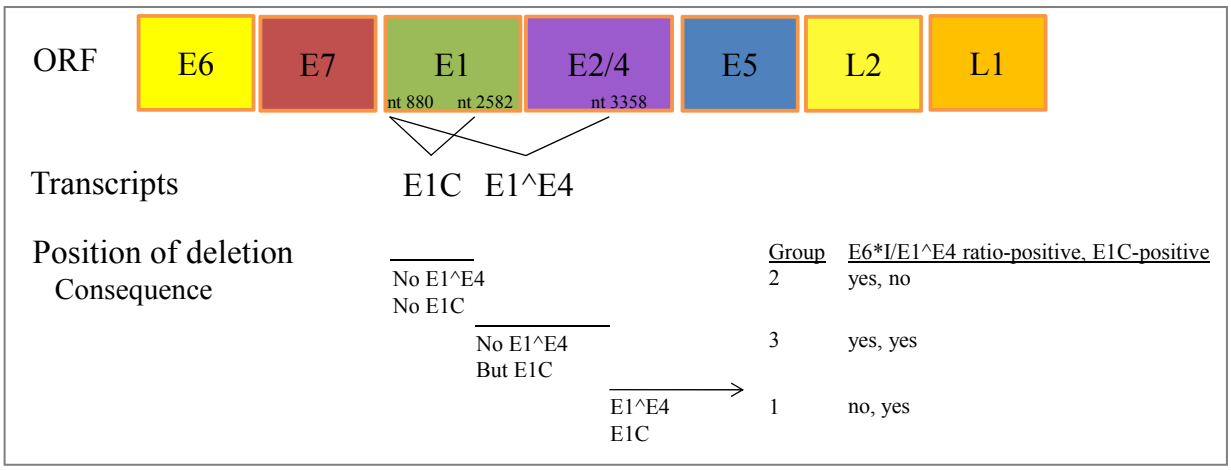


Figure 5.1. **Position of viral sequence deletion due to integration and the consequences for presence of E1C and E1^E4 and for E6*I/E1^E4 and/or E1C positivity**.

Group 2 contains only E6*I/E1^E4 ratio-positives indicating the absence of splice acceptor nt 3358. The absence of E1C in group 2 could be due to either loss of E1C splice acceptor or absence of the transcription mechanism that is active in E1C-positive severe lesions. In the case of loss of E1C splice acceptor, truncation would be upstream of nt 2582. Group 2 was observed in 45% of all ≥CIN3 in this thesis being in consistence with the frequency of HPV16 breakpoints identified upstream of nt 2582 in the literature[85]. The loss of E1C expression in group 2 could still be compatible with transformation because concomitantly the E2 repressor is absent due to the truncation.

Group 3 contains samples double positive for E6*I/E1^E4 ratio and E1C, indicating either transcriptionally active episomes as well as truncated integrates or truncated integrated forms only with integration upstream of nt 3358 but downstream of nt 2582. In the literature, breakpoints within this region have been described in 35% of CxCa samples analysed[85]. E2 repressor expression is reduced or absent and the reduced E1C copy numbers are sufficient for maintenance of transformation.

So in summary, in all three groups E2 repressor is either undetectable or if present could be out-titrated by E1C. However, in view of E1C absence in mild lesions, the activation of E1C in severe lesions still remains to be explained.

5.3.6 False-negative and false-positive lesion classification by HPV16 RNA patterns analysis in single-infected HPV16 DNA-positive exfoliated cervical cells

Overall, 35 out of 42 CIN3 (83%) were correctly classified as severe lesions. The 7 false-negative CIN3 (17%) may be explained by (i) non-optimal storage conditions, (ii) wrong histological diagnosis or (iii) poor sampling (as described in 5.3.1) and ergo lack of abnormal cells. In addition, when a cell sample contains a mixture of non-transformed but HPV16-infected cells and others exhibiting severe lesion, the HPV16 RNA patterns analysis may give equivocal results if EIC is not expressed in the severe lesion and E1^E4 from the non-transformed, HPV16-infected cells leads to a low E6*I/E1^E4 ratio. RNA degradation in the up to two years old samples may have contributed to E1C negativity. Future studies should include freshly collected, non-archived cervical cell samples which also better represent the potential application of the HPV16 RNA patterns assay as clinical diagnostic test.

In this study, the sensitivity of HPV16 RNA patterns analysis increased by including samples in the E6*I/E1^E4 ratio computation (described in detail in chapter 3.12) with E6*I copy numbers lower than 10 and simultaneously negative for E1^E4. Thereby, two CxCa and one CIN3 were classified as severe lesion with E6*I/E1^E4 ratio above cutoff while at the same time only a single CIN0 became false-positive. The ubC copies per PCR of all four samples ranged between 377 and 2,831. However, when using the HPV16 RNA patterns analysis as routine diagnostic test, samples with low E6*I together with absence of E1^E4 should be classified as invalid and a repeat sample should be analysed.

Furthermore, 48 out of 57 (84%) single-infected HPV16 DNA-positive exfoliated cervical cell samples were correctly classified as mild lesion demonstrating a high specificity of the HPV16 RNA patterns analysis for detection of severe lesions. The 16% false-positive samples (8% of NIL/M, 40% of CIN0 and 18% of CIN1) may be explained by incorrect biopsy or histological misclassification. Of the false-positive CIN1, none was E1C-positive but E6*I/E1^E4 ratio-positive. An upcoming progression as an explanation for false-positive samples might be rejected, since -as mentioned above- a mixed form of non-transformed but HPV16-infected cells and others exhibiting severe lesion could give equivocal results by the HPV16 RNA patterns analysis when E1C is negative. However, follow-up data were available for seven HPV16 RNA pattern-positive CIN1 samples: Within 9-24 months six progressed to CIN2 and CIN3, respectively. Furthermore, one HPV16 RNA patterns-negative CIN2 was diagnosed as CIN0 at follow-up 12 months later. This may indicate a potential predictive value of the HPV16 RNA patterns in mild lesions. Larger longitudinal studies are required to further investigate the predictive value of HPV16 RNA patterns in cervical cancer precursor screening.

5.4 Intraobserver reproducibility of histological CIN diagnosis

In most studies histological diagnosis of CIN lesions has been used as gold standard. However, only a moderate reproducibility of histological CIN diagnosis has been reported between different pathologists[108,191]. Dalla Palma et al., 2009[109] described that 30% of cases with a negative diagnosis, 63% of CIN1, 62% of CIN3 and 48% of CIN2 were agreed on by 2 reviewers. Similarly, Stolar et al., 2001[108] described concordant histological diagnosis in 43% of CIN1 and in 77% of ≥CIN2.

Similar variation was also observed in this study with 44 out of 104 CIN0 to CIN3 samples (42%) discordantly diagnosed. 50% of CIN0 and CIN1 samples diagnosed by pathologist 1 were diagnosed as CIN2 or worse by pathologist 2 and 7% of CIN2 and CIN3 as CIN0 or CIN1. This is surprising since for histological diagnosis also immunohistochemistry for $p16^{INK4A}$ in a parallel tissue section was available.

HPV16 RNA patterns analysis correctly classified 48 out of 57 (84%) single-infected HPV16 DNA-positive ≤CIN1 samples as mild lesion and 53 out of 60 (88%) ≥CIN3 samples as severe lesion when the consensus histological diagnoses were used as reference. However, using the diagnoses of pathologist 1 the number of correctly classified mild lesions by HPV16 RNA patterns was reduced to 49 out of 64 (77%) and that of correctly classified severe lesions decreased to 50 out of 58 (86%). Similarly, with pathologist 2 diagnoses the corresponding numbers for mild lesions were reduced to 45 out of 55 (82%) and to 52 out of 65 (80%) for severe lesions.

These observations underline that the validation of novel diagnostic tools for cervical lesions, such as the novel HPV16 RNA patterns test, is complicated by variability of the histological diagnoses used as gold standard.

5.5 Alternative molecular markers for cervical lesion staging

HPV DNA tests recognise HPV-induced cervical lesions but also clinically not relevant infections. The determination of high viral load is a clinically more useful marker as it can reliably discriminate between no lesion and any lesion[134]. In this thesis 98% of ≥CIN3 were classified as severe lesion by high viral load analysis, but also 100% of CIN0 and CIN1. It has to be kept in mind that patients diagnosed as histologically normal CIN0 in the colposcopy clinic in Bad Münder had preceding suspicious cytology findings as reason for referral. However, only 30% of NIL/M from another study[149] were classified as high viral load-positive.

As alternative modalities, novel biomarkers for stages of cervical carcinogenesis are emerging based on methylation of viral but also of human sequences[175,192]. With human genes such as AM3, EPB41L3 and TERTC13ORF18, CADM1 and MAL as marker, sensitivities for hrHPV positive ≥CIN3 patients[193,194] between 54% and 84%[193,194] and specificities for ≤CIN1 of

69%[193] and 53%[194] were reported. However, the methylation frequencies even for the same human gene varied widely between studies[192]. Methylation of HPV sequences seems to be a more promising biomarker. Alterations of the HPV-methylome in high-grade precancer and invasive cancer were observed particularly in the URR and L2 and L1 genes[82]. In general, methylation of viral DNA distinguishes women with ≥CIN2 from those without lesion or a lesion <CIN2[175]. A sensitivity for ≥CIN2 of 60% and a specificity of 91% have been reported when methylation of a combination of three CG sites (in L1, L2 and E2-E4) was used[195]. Whether HPV methylation is suited as biomarker for triage of screening HPV-positive women requires further evaluation[175].

5.6 Potential consequences of HPV RNA patterns analysis for cervical cancer screening

Hitherto, discrimination of mild and severe lesions in exfoliated cervical cells by cytology as primary screening tool is of limited sensitivity and specificity. As mentioned above, HPV-DNA tests and even high viral load determine mild lesions (≤LSIL and ≤CIN1) as test-positive. It is therefore necessary to develop new modalities that are sensitive and allow differentiation of mild (clinically less relevant) and severe lesions (clinically more relevant).

The HPV16 RNA patterns test could be used as triage test for secondary cervical cancer prevention in HPV16 DNA (high viral load)-positive cases for the sensitive discrimination of mild (wait and follow-up) versus severe (intervention) lesions which could allow extension of follow-up intervals. At a follow-up visit, the HPV16 RNA patterns test exclusively could be sufficient as diagnostic tool and if (i) negative for HPV16 E6*I RNA, the patient could return to normal routine screening, (ii) positive for HPV16 E6*I RNA, but HPV16 RNA patterns-negative follow-up would be necessary and (iii) HPV16 RNA patterns-positive intervention would be necessary. Such sequential combinations of HPV DNA and HPV16 RNA patterns tests could be more reliable and particularly useful for limiting extensive examinations to patients identified at high risk only.

At present, the limitation of the HPV RNA patterns test is the restriction to HPV16 only which covers only about future 55% of cervical cancer cases. If successfully extended to further hrHPV types, RNA patterns test could also be used in primary screening.

6 Outlook

In this thesis, HPV16 RNA patterns analysis has been transferred from cumbersome NASBA technology to robust and much simpler RT-qPCR. HPV16 RNA patterns analysis of large numbers of clinical samples by the new RT-qPCR confirmed the suitability of this biomarker to sensitively and specifically discriminate severe from mild cervical lesions. Furthermore, 6 out of 7 HPV16 RNA patterns-positive CIN1 progressed to ≥CIN2 in the follow-up histology after 9-24 months. This observation requires larger follow-up studies in order to confirm the predictive value of the HPV16 RNA patterns.

In future experiments care should also be taken to extract RNA within days after sampling since I observed loss of amplifiable RNA after extended storage of exfoliated cell samples despite use of PreservCyt buffer. Establishing the logistics of fast sample processing would also bring the analysis closer to clinical routine.

From the literature it is known that HPV-DNA detection works excellent in self-collected lavage samples[196-198]. Such samples, provided that they can be stored in a RNA-protecting buffer and after careful test validation, could also represent a suitable sample source for HPV RNA patterns analysis.

Extension of the HPV16 RNA patterns to other hrHPV types appears doable with reasonable effort and well justified. Together with HPV16, the 7 hrHPV types 18, 31, 33, 35, 45, 52 and 58 most prevalent in CxCa account for about 83% of cervical cancer cases[71]. All splice sites of these seven hrHPV types necessary for RNA patterns analyses have been identified and/or confirmed in this thesis. With the experience gained with HPV16 the technical development of HPV RNA patterns analysis to these 7 types is expected to take approximately another 2 years. The extension to the second most prevalent hrHPV type 18 has already been started. For the validation of non-HPV16 RNA patterns tests a substantial number of clinical samples has already been collected and characterised in the course of this thesis. In total, 718 cervical cell samples positive for at least one of the seven other hrHPV types are available ranging from a minimum of 5 HPV58-positive CxCa samples to a maximum of 108 HPV31-positive CIN samples. RNA of all the samples has been already extracted.

For other cancer entities with HPV involvement, extension of RNA patterns testing beyond HPV16 is of lower importance. HPV16 causes more than 90% of all HPV-driven

OPSCC[152,199]. The HPV16 RNA patterns may be used in order to select the primary treatment modalities for patients with oropharyngeal squamous cell carcinomas as demonstrated before[152]: It could be shown that tumours being HPV16 RNA patterns-positive have a better overall survival and progression free survival. In situations like in Germany, with heterogeneity among the HPV DNA-positive tumour patients, precise definition of HPV activity by the usage of the HPV16 RNA patterns test could help in selection of the primary treatment modality[152].

Of all HPV DNA-positive anal cancer cases, HPV16 DNA is found in 86%[200]. Squamous cell carcinoma of the anus, like CxCa, is thought to be preceded by a spectrum of intraepithelial changes, anal intraepithelial neoplasia (AIN) of varying severity. High-grade AIN can be identified using high-resolution anascopy (HRA) or colposcopy of the anus and perianus after application of acetic acid followed by biopsy and histology. However, in comparison to the number of cervical colposcopists, there are currently far fewer trained experts in HRA. Since there are no recommendations for screening, many patients are either diagnosed serendipitously during surgery for benign anal conditions or occasionally during colonoscopy or if they present with anal symptoms[201]. A potential use of HPV16 RNA patterns for the identification of high-grade AIN and anal SCC needs to be examined in future studies.

HPV related vulvar intraepithelial lesions (VIN) and invasive vulvar cancer is mainly (73%) associated to HPV16[202]. Examination of the vulva is part of the gynecologic evaluation and a suspicious lesion needs to undergo a biopsy for further diagnosis. A potential use of HPV16 RNA patterns for the identification of high-grade VIN and vulvar carcinoms needs to be examined in future studies.

HPV16 RNA patterns can also potentially be used for the identification of vaginal intraepithelial neoplasia (VAIN) and vaginal cancer where HPV16 was most prevalent subtype[203,204].

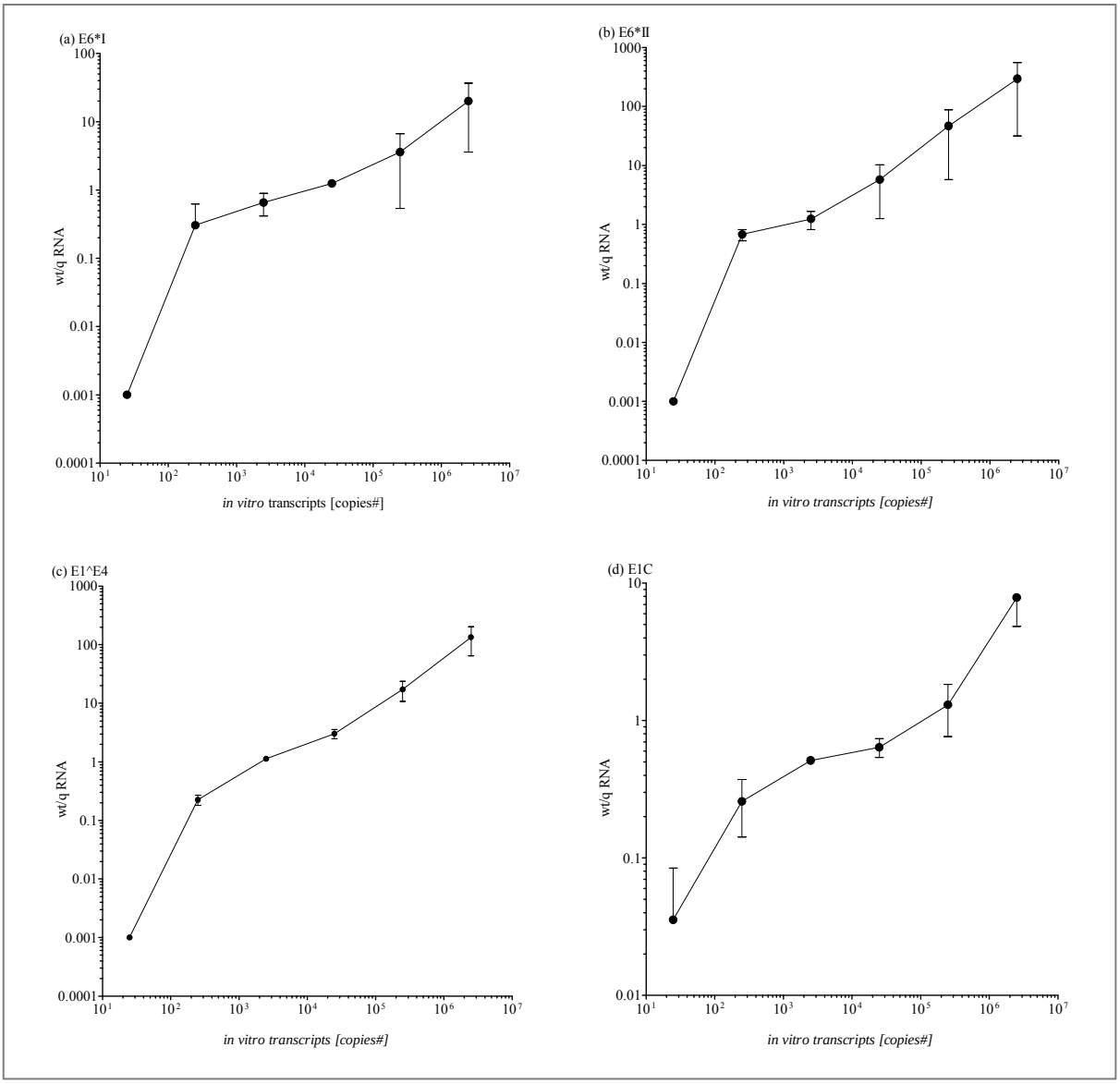
7 Abbreviations

ADC	Adenocarcinoma
ADSCC	Adenosquamous cell carcinoma
AIN	Anal intraepithelial lesion
AMV-RT	Avian myeloblastosis virus reverse transcriptase
ASC-US	Atypical squamous cells of undetermined significance
AUC	area under the curve
Bg	betaglobin
bp	base pair
BS	broad-spectrum
bw	backward
°C	degree Celsius
Cell RNA	MagNA Pure96 Cellular RNA Large Volume Kit
CI	confidence interval
CIN	cervical intraepithelial neoplasia
Cp	crossing point
CV	coefficient of variation
CxCa	cervical cancer
DGGG	Deutsche Gesellschaft für Gynäkologie und Geburtshilfe
DKFZ	Deutsches Krebsforschungszentrum
DL	detection limit
DNA	Desoxyribonucleic acid
DP	duplex
E	efficiency
e.g.	example gratia
E1C	N-terminal truncated E1
E6AP	E6-associated protein
EGF	epidermal growth factor
ELISA	enzyme-linked immunosorbent assay
ER	endoplasmatic reticulum
FCS	fetal calf serum
FELASA	Federation of European Laboratory Animal Science Association
FFPE	formalin-fixed paraffin embedded

fl	full-length
fw	forward
HC2	Hybrid Capture 2
HCT	Hybrid Capture Tube
HE	hematoxylin and eosin
HPV	Human papillomavirus(es)
hr	high-risk
HRA	high-resolution anascopy
HSIL	high-grade squamous intraepithelial lesion
i.e.	id est
ICO	Catalan Institute of oncology
IP	intellectual property
κ	kappa
LA	Linear Array HPV Genotyping
LBC	Liquid-based cytology
LC480 RNA	Light Cycler 480 RNA Master Hydrolysis Probes Kit
LCR	long control region
lr	low-risk
LSIL	low-grade squamous intraepithelial lesion
m	slope
mM	millimolar
MFI	median fluorescence intensity
MPG	multiplex HPV genotyping
mRNA	messenger-RNA
µL	micro liter
N	number
NASBA	nucleic acid sequence-based amplification
NIL/M	negative for intraepithelial lesion or malignancy
NPV	negative predictive value
nt	nucleotide
OPSCC	oropharynx squamous cell carcinomas
ORF	Open reading frame
Pap	Papanicolaou
PBS	Phosphate buffered saline

PCR	Polymerase chain reaction
PPV	positive predictive value
pRb	retionoblastoma tumor suppressor protein
PV	Papillomavirus(es)
q	quantitative
R^2	coefficient of determination
rBFF	rodent bacteria and fungus finder
rDVF	rodent DNA virus finder
RIN	RNA integrity number
RNA	Ribonucleic acid
ROC	Receiver operating characteristic
r_p	Pearson correlation
RPA	replication protein A
rRVF	rodent RNA virus finder
RT	reverse transcriptase
RT-qPCR	reverse transcriptase quantitative PCR
SCC	squamous cell carcinomas
sj	splice junction
SP	singleplex
TP	triplex
ubC	ubiquitin C
URR	upstream regulatory region
VAIN	vaginal intraepithelial neoplasia
VIN	vulvar intraepithelial neoplasia
Viral NA	MagNA Pure96 DNA and Viral NA Large Volume Kit
VLP	virus-like particle
WHO	World Health Organization
wt	wildtype

8 Appendix



(Continued on the following page)

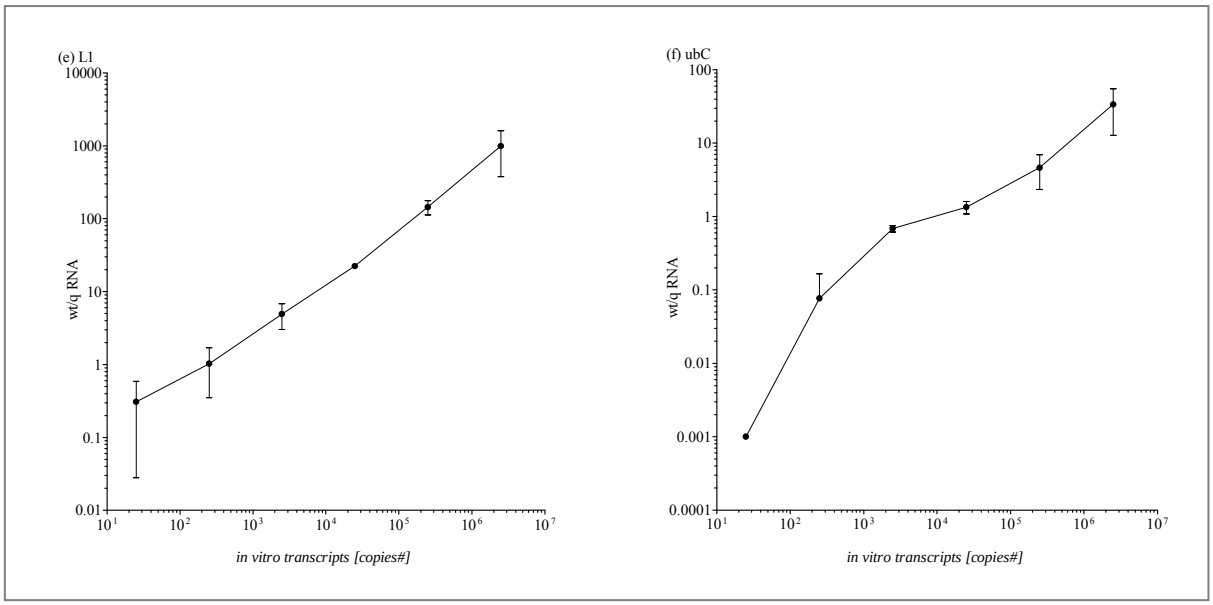


Figure A8.1. **Dose-response curves of *in vitro* transcript copies and NASBA wt/q signal**. NASBA signals (ratio of wildtype (wt)/calibrator (q) signal on y-axis) for dilution series of *in vitro* transcribed E6*I (a), E6*II (b), E1^E4 (c), E1C (d), L1 (e) and ubC (f) RNA (data were kindly provided by Dr. Schmitt (DKFZ)). All templates were applied in duplicates, error bars show standard error.

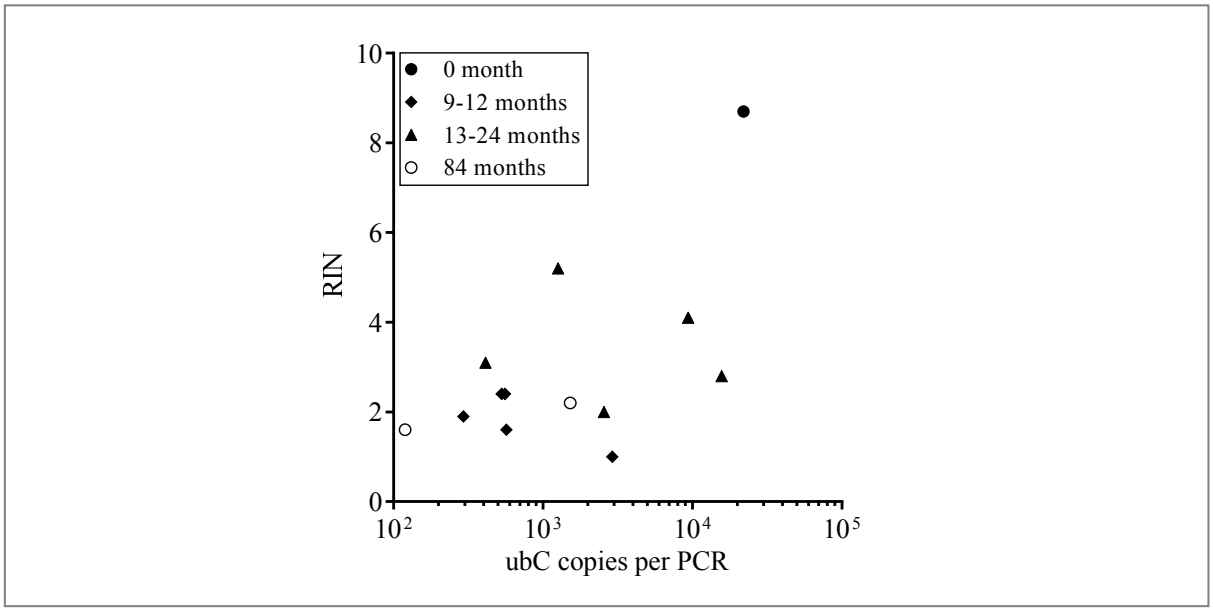


Figure A8.2. **Correlation of ubC copies per PCR and RIN factor in cervical cell samples of different storage times.** UbC copies per PCR (x-axis) and RIN factor of cervical cell samples archived between 9-12 months (rhombi), 13-24 months (triangles), 84 months (open circles) and MRI-H186 cell line (filled circle) are blotted. All templates were applied in duplicates and the mean Cp value was used for calculation of copies per PCR.

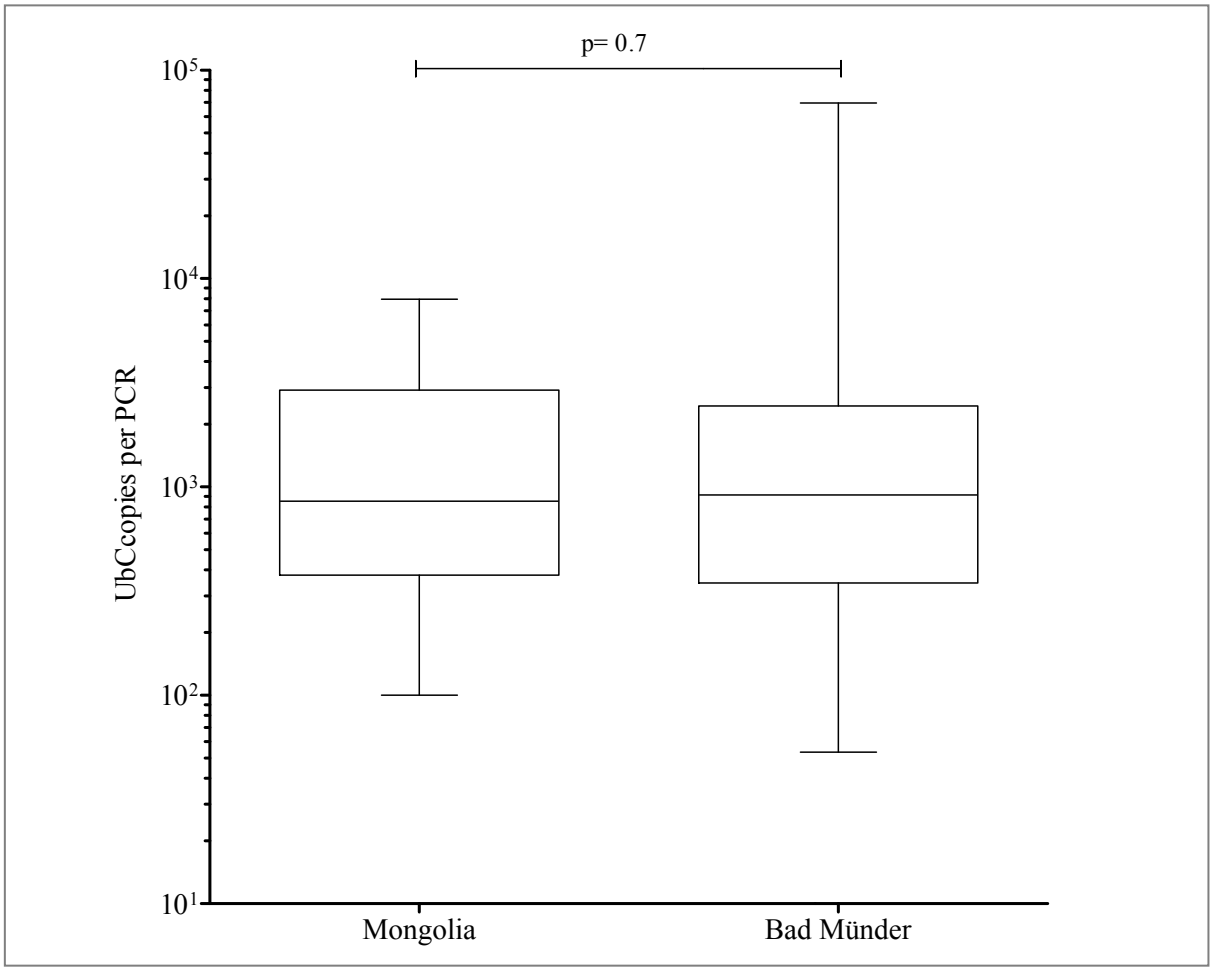


Figure A8.3. **RNA quality of HPV16 single-infected cervical cell samples from Bad Münder (n=104) and Mongolia (n=54).** UbC copies per PCR (whiskers min to max) (y-axis) in exfoliated cervical cells from Mongolia and Bad Münder (x-axis). The lower edge of the box plot is the first 25th quartile, the upper edge the 75th percentile.

9 Literature

1 http://web.uct.ac.za/depts/mmi/stannard/papillo.html. <http://web.uct.ac.za/depts/mmi/stannard/papillo.html> (2008).

2 Modis, Y., Trus, B. L. & Harrison, S. C. Atomic model of the papillomavirus capsid. *The EMBO journal* **21**, 4754-4762 (2002).

3 Buck, C. B. *et al.* Arrangement of L2 within the papillomavirus capsid. *Journal of virology* **82**, 5190-5197, doi:10.1128/JVI.02726-07 (2008).

4 Halec, G. Molecular evidence for transforming activity of rare and probable/possible high-risk human papillomavirus types in cervical cancer. *Ruperto-Carola University of Heidelberg* **Dissertation** (2012).

5 Munoz, N., Castellsague, X., de Gonzalez, A. B. & Gissmann, L. Chapter 1: HPV in the etiology of human cancer. *Vaccine* **24 Suppl 3**, S3/1-10, doi:10.1016/j.vaccine.2006.05.115 (2006).

6 Sato, K. *et al.* Human papillomavirus type 16 P670 promoter is negatively regulated by CCAAT displacement protein. *Virus genes* **35**, 473-481, doi:10.1007/s11262-006-0074-8 (2007).

7 Sherman, L. & Alloul, N. Human papillomavirus type 16 expresses a variety of alternatively spliced mRNAs putatively encoding the E2 protein. *Virology* **191**, 953-959 (1992).

8 Rohlfs, M., Winkenbach, S., Meyer, S., Rupp, T. & Durst, M. Viral transcription in human keratinocyte cell lines immortalized by human papillomavirus type-16. *Virology* **183**, 331-342 (1991).

9 Zheng, Z. M., Tao, M., Yamanegi, K., Bodaghi, S. & Xiao, W. Splicing of a cap-proximal human Papillomavirus 16 E6E7 intron promotes E7 expression, but can be restrained by distance of the intron from its RNA 5' cap. *Journal of molecular biology* **337**, 1091-1108 (2004).

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

11 Smotkin, D. & Wettstein, F. O. Transcription of human papillomavirus type 16 early genes in a cervical cancer and a cancer-derived cell line and identification of the E7 protein. *Proceedings of the National Academy of Sciences of the United States of America* **83**, 4680-4684 (1986).

12 Halec, G. *et al.* Biological activity of probable/possible high-risk human papillomavirus types in cervical cancer. *Int J Cancer* **132**, 63-71, doi:10.1002/ijc.27605 (2012).

13 Sedman, S. A. *et al.* The full-length E6 protein of human papillomavirus type 16 has transforming and trans-activating activities and cooperates with E7 to immortalize keratinocytes in culture. *J Virol* **65**, 4860-4866 (1991).

14 Smotkin, D., Prokoph, H. & Wettstein, F. O. Oncogenic and nononcogenic human genital papillomaviruses generate the E7 mRNA by different mechanisms. *Journal of virology* **63**, 1441-1447 (1989).

15 Tang, S., Tao, M., McCoy, J. P., Jr. & Zheng, Z. M. The E7 oncoprotein is translated from spliced E6*I transcripts in high-risk human papillomavirus type 16- or type 18-positive cervical cancer cell lines via translation reinitiation. *Journal of virology* **80**, 4249-4263 (2006).

16 Stacey, S. N. *et al.* Leaky scanning is the predominant mechanism for translation of human papillomavirus type 16 E7 oncoprotein from E6/E7 bicistronic mRNA. *Journal of virology* **74**, 7284-7297 (2000).

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

18 Alloul, N. & Sherman, L. Transcription-modulatory activities of differentially spliced cDNAs encoding the E2 protein of human papillomavirus type 16. *The Journal of general virology* **80 (Pt 9)**, 2461-2470 (1999).

19 Shirasawa, H. *et al.* Transcription-modulatory activity of full-length E6 and E6*I proteins of human papillomavirus type 16. *Virology* **203**, 36-42, doi:S0042-6822(84)71452-8 [pii] 10.1006/viro.1994.1452 (1994).

20 Sherman, L., Alloul, N., Golan, I., Durst, M. & Baram, A. Expression and splicing patterns of human papillomavirus type-16 mRNAs in pre-cancerous lesions and carcinomas of the cervix, in human keratinocytes immortalized by HPV 16, and in cell lines established from cervical cancers. *Int J Cancer* **50**, 356-364 (1992).

21 Schwarz, E. *et al.* Structure and transcription of human papillomavirus sequences in cervical carcinoma cells. *Nature* **314**, 111-114 (1985).

22 Baker, C. C. *et al.* Structural and transcriptional analysis of human papillomavirus type 16 sequences in cervical carcinoma cell lines. *J Virol* **61**, 962-971 (1987).

23 Shirasawa, H., Tanzawa, H., Matsunaga, T. & Simizu, B. Quantitative detection of spliced E6-E7 transcripts of human papillomavirus type 16 in cervical premalignant lesions. *Virology* **184**, 795-798 (1991).

24 McNicol, P., Guijon, F., Wayne, S., Hidajat, R. & Paraskevas, M. Expression of human papillomavirus type 16 E6-E7 open reading frame varies quantitatively in biopsy tissue from different grades of cervical intraepithelial neoplasia. *Journal of clinical microbiology* **33**, 1169-1173 (1995).

25 Halec, G. *et al.* Biological activity of probable/possible high-risk human papillomavirus types in cervical cancer. *Int J Cancer* **132**, 63-71, doi:10.1002/ijc.27605 (2013).

26 Baker, C., and C. Calef. Maps of papillomavirus mRNA transcripts. (1996).

27 Loo, Y. M. & Melendy, T. Recruitment of replication protein A by the papillomavirus E1 protein and modulation by single-stranded DNA. *Journal of virology* **78**, 1605-1615 (2004).

28 Masterson, P. J., Stanley, M. A., Lewis, A. P. & Romanos, M. A. A C-terminal helicase domain of the human papillomavirus E1 protein binds E2 and the DNA polymerase alpha-primase p68 subunit. *Journal of virology* **72**, 7407-7419 (1998).

29 Conger, K. L., Liu, J. S., Kuo, S. R., Chow, L. T. & Wang, T. S. Human papillomavirus DNA replication. Interactions between the viral E1 protein and two subunits of human dna polymerase alpha/primase. *The Journal of biological chemistry* **274**, 2696-2705 (1999).

30 Han, Y., Loo, Y. M., Militello, K. T. & Melendy, T. Interactions of the papovavirus DNA replication initiator proteins, bovine papillomavirus type 1 E1 and simian virus 40 large T antigen, with human replication protein A. *Journal of virology* **73**, 4899-4907 (1999).

31 Schmitt, M. Detection of nucleic acids from human alpha papillomaviruses in the uterine cervix. *Ruperto-Carola University of Heidelberg* **Dissertation** (2008).

32 Donaldson, M. M. *et al.* An Interaction between Human Papillomavirus 16 E2 and TopBP1 Is Required for Optimum Viral DNA Replication and Episomal Genome Establishment. *Journal of virology* **86**, 12806-12815, doi:10.1128/JVI.01002-12 JVI.01002-12 [pii] (2012).

33 Melendy, T., Sedman, J. & Stenlund, A. Cellular factors required for papillomavirus DNA replication. *Journal of virology* **69**, 7857-7867 (1995).

34 Bouvard, V., Storey, A., Pim, D. & Banks, L. Characterization of the human papillomavirus E2 protein: evidence of trans-activation and trans-repression in cervical keratinocytes. *The EMBO journal* **13**, 5451-5459 (1994).

35 Milligan, S. G., Veerapraditsin, T., Ahamet, B., Mole, S. & Graham, S. V. Analysis of novel human papillomavirus type 16 late mRNAs in differentiated W12 cervical epithelial cells. *Virology* **360**, 172-181 (2007).

36 Doorbar, J., Ely, S., Sterling, J., McLean, C. & Crawford, L. Specific interaction between HPV-16 E1-E4 and cytokeratins results in collapse of the epithelial cell intermediate filament network. *Nature* **352**, 824-827 (1991).

37 Johnsen, C. K., Stanley, M. & Norrild, B. Analysis of human papillomavirus type 16 E5 oncogene expression in vitro and from bicistronic messenger RNAs. *Intervirology* **38**, 339-345 (1995).

38 Hwang, E. S., Nottoli, T. & Dimaio, D. The HPV16 E5 protein: expression, detection, and stable complex formation with transmembrane proteins in COS cells. *Virology* **211**, 227-233, doi:10.1006/viro.1995.1395 (1995).

39 Disbrow, G. L., Hanover, J. A. & Schlegel, R. Endoplasmic reticulum-localized human papillomavirus type 16 E5 protein alters endosomal pH but not trans-Golgi pH. *Journal of virology* **79**, 5839-5846, doi:10.1128/JVI.79.9.5839-5846.2005 (2005).

40 Straight, S. W., Herman, B. & McCance, D. J. The E5 oncoprotein of human papillomavirus type 16 inhibits the acidification of endosomes in human keratinocytes. *Journal of virology* **69**, 3185-3192 (1995).

41 Crusius, K., Rodriguez, I. & Alonso, A. The human papillomavirus type 16 E5 protein modulates ERK1/2 and p38 MAP kinase activation by an EGFR-independent process in stressed human keratinocytes. *Virus genes* **20**, 65-69 (2000).

42 Kiselev, F. L. *et al.* [Status of the human DNA papillomavirus in cervical tumors]. *Molekuliarnaia biologiia* **35**, 470-476 (2001).

43 Huibregtse, J. M., Scheffner, M. & Howley, P. M. Cloning and expression of the cDNA for E6-AP, a protein that mediates the interaction of the human papillomavirus E6 oncoprotein with p53. *Molecular and cellular biology* **13**, 775-784 (1993).

44 Werness, B. A., Levine, A. J. & Howley, P. M. Association of human papillomavirus types 16 and 18 E6 proteins with p53. *Science* **248**, 76-79 (1990).

45 Zeitler, J., Hsu, C. P., Dionne, H. & Bilder, D. Domains controlling cell polarity and proliferation in the Drosophila tumor suppressor Scribble. *The Journal of cell biology* **167**, 1137-1146, doi:10.1083/jcb.200407158 (2004).

46 Klingelhutz, A. J., Foster, S. A. & McDougall, J. K. Telomerase activation by the E6 gene product of human papillomavirus type 16. *Nature* **380**, 79-82, doi:10.1038/380079a0 (1996).

47 Zerfass, K. *et al.* Sequential activation of cyclin E and cyclin A gene expression by human papillomavirus type 16 E7 through sequences necessary for transformation. *J Virol* **69**, 6389-6399 (1995).

48 Ordonez, R. M., Espinosa, A. M., Sanchez-Gonzalez, D. J., Armendariz-Borunda, J. & Berumen, J. Enhanced oncogenicity of Asian-American human papillomavirus 16 is associated with impaired E2 repression of E6/E7 oncogene transcription. *The Journal of general virology* **85**, 1433-1444 (2004).

49 Stacey, S. N. *et al.* Translation of the human papillomavirus type 16 E7 oncoprotein from bicistronic mRNA is independent of splicing events within the E6 open reading frame. *J Virol* **69**, 7023-7031 (1995).

50 Bodily, J. M., Mehta, K. P., Cruz, L., Meyers, C. & Laimins, L. A. The E7 open reading frame acts in cis and in trans to mediate differentiation-dependent activities in the human papillomavirus type 16 life cycle. *Journal of virology* **85**, 8852-8862, doi:10.1128/JVI.00664-11 (2011).

51 zur Hausen, H. Papillomaviruses and cancer: from basic studies to clinical application. *Nature reviews* **2**, 342-350, doi:10.1038/nrc798 (2002).

52 Roden, R. B., Kirnbauer, R., Jenson, A. B., Lowy, D. R. & Schiller, J. T. Interaction of papillomaviruses with the cell surface. *J Virol* **68**, 7260-7266 (1994).

53 Schiller, J. T., Day, P. M. & Kines, R. C. Current understanding of the mechanism of HPV infection. *Gynecol Oncol* **118**, S12-17, doi:S0090-8258(10)00290-8 [pii] 10.1016/j.ygyno.2010.04.004 (2010).

54 Crum, C. P., Nuovo, G., Friedman, D. & Silverstein, S. J. Accumulation of RNA homologous to human papillomavirus type 16 open reading frames in genital precancers. *J Virol* **62**, 84-90 (1988).

55 Stoler, M. H., Wolinsky, S. M., Whitbeck, A., Broker, T. R. & Chow, L. T. Differentiation-linked human papillomavirus types 6 and 11 transcription in genital condylomata revealed by in situ hybridization with message-specific RNA probes. *Virology* **172**, 331-340 (1989).

56 Shafti-Keramat, S. *et al.* Different heparan sulfate proteoglycans serve as cellular receptors for human papillomaviruses. *Journal of virology* **77**, 13125-13135 (2003).

57 Evander, M. *et al.* Identification of the alpha6 integrin as a candidate receptor for papillomaviruses. *Journal of virology* **71**, 2449-2456 (1997).

58 Day, P. M., Lowy, D. R. & Schiller, J. T. Papillomaviruses infect cells via a clathrin-dependent pathway. *Virology* **307**, 1-11 (2003).

59 Doorbar, J. Molecular biology of human papillomavirus infection and cervical cancer. *Clin Sci (Lond)* **110**, 525-541, doi:10.1042/CS20050369 (2006).

60 Stubenrauch, F. & Laimins, L. A. Human papillomavirus life cycle: active and latent phases. *Seminars in cancer biology* **9**, 379-386, doi:10.1006/scbi.1999.0141 (1999).

61 Doorbar, J. *et al.* Characterization of events during the late stages of HPV16 infection in vivo using high-affinity synthetic Fabs to E4. *Virology* **238**, 40-52, doi:10.1006/viro.1997.8768 (1997).

62 Jemal, A. *et al.* Cancer statistics, 2003. *CA: a cancer journal for clinicians* **53**, 5-26 (2003).

63 Bosch, F. X., Lorincz, A., Munoz, N., Meijer, C. J. L. M. & Shah, K. V. The causal relation between human papillomavirus and cervical cancer. *J Clin Pathol* **55**, 244-265 (2002).

64 Centre)., W. I. I. C. o. H. a. C. C. H. I. *Human Papillomavirus and Related Cancers in Germany.* , 2010).

65 IARC. *THE GLOBOCAN PROJECT*, 2008).

66 Pirog, E. C. *et al.* Prevalence of human papillomavirus DNA in different histological subtypes of cervical adenocarcinoma. *The American journal of pathology* **157**, 1055-1062, doi:10.1016/S0002-9440(10)64619-6 (2000).

67 zur Hausen, H. Papillomaviruses in human cancers. *Proceedings of the Association of American Physicians* **111**, 581-587 (1999).

68 Schiffman, M., Clifford, G. & Buonaguro, F. M. Classification of weakly carcinogenic human papillomavirus types: addressing the limits of epidemiology at the borderline. *Infectious agents and cancer* **4**, 8, doi:10.1186/1750-9378-4-8 (2009).

69 Durst, M., Gissmann, L., Ikenberg, H. & zur Hausen, H. A papillomavirus DNA from a cervical carcinoma and its prevalence in cancer biopsy samples from different geographic regions. *Proceedings of the National Academy of Sciences of the United States of America* **80**, 3812-3815 (1983).

70 Bosch, F. X. *et al.* Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. International biological study on cervical cancer (IBSCC) Study Group. *Journal of the National Cancer Institute* **87**, 796-802 (1995).

71 Munoz, N. *et al.* Epidemiologic classification of human papillomavirus types associated with cervical cancer. *The New England journal of medicine* **348**, 518-527, doi:10.1056/NEJMoa021641 (2003).

72 Collins, S. *et al.* High incidence of cervical human papillomavirus infection in women during their first sexual relationship. *BJOG : an international journal of obstetrics and gynaecology* **109**, 96-98 (2002).

73 Castle, P. E. *et al.* A prospective study of age trends in cervical human papillomavirus acquisition and persistence in Guanacaste, Costa Rica. *The Journal of infectious diseases* **191**, 1808-1816, doi:10.1086/428779 (2005).

74 Franceschi, S. *et al.* Variations in the age-specific curves of human papillomavirus prevalence in women worldwide. *Int J Cancer* **119**, 2677-2684, doi:10.1002/ijc.22241 (2006).

75 Ostor, A. G. Natural history of cervical intraepithelial neoplasia: a critical review. *International journal of gynecological pathology : official journal of the International Society of Gynecological Pathologists* **12**, 186-192 (1993).

76 Syrjanen, K. *et al.* Natural history of cervical human papillomavirus lesions does not substantiate the biologic relevance of the Bethesda System. *Obstetrics and gynecology* **79**, 675-682 (1992).

77 Moore, K. *et al.* Adolescent cervical dysplasia: histologic evaluation, treatment, and outcomes. *American journal of obstetrics and gynecology* **197**, 141 e141-146, doi:10.1016/j.ajog.2007.03.029 (2007).

78 Fuchs, K., Weitzen, S., Wu, L., Phipps, M. G. & Boardman, L. A. Management of cervical intraepithelial neoplasia 2 in adolescent and young women. *Journal of pediatric and adolescent gynecology* **20**, 269-274, doi:10.1016/j.jpag.2007.04.012 (2007).

79 Hoskins, W. J., Perez, C.A., Young R.C., Barakat, R., Markman, M., Randall, M. Principles and practice of gynecologic oncology. *Lippincott Williams & Wilkins* **Fourth Edition** (2005).

80 Schiffman, M., Castle, P. E., Jeronimo, J., Rodriguez, A. C. & Wacholder, S. Human papillomavirus and cervical cancer. *Lancet* **370**, 890-907, doi:10.1016/S0140-6736(07)61416-0 (2007).

81 Bodily, J. & Laimins, L. A. Persistence of human papillomavirus infection: keys to malignant progression. *Trends in microbiology* **19**, 33-39, doi:10.1016/j.tim.2010.10.002 (2011).

82 Vinokurova, S. & von Knebel Doeberitz, M. Differential methylation of the HPV 16 upstream regulatory region during epithelial differentiation and neoplastic transformation. *PloS one* **6**, e24451, doi:10.1371/journal.pone.0024451 (2011).

83 Vinokurova, S. *et al.* Type-dependent integration frequency of human papillomavirus genomes in cervical lesions. *Cancer research* **68**, 307-313, doi:10.1158/0008-5472.CAN-07-2754 (2008).

84 Schmitz, M., Driesch, C., Jansen, L., Runnebaum, I. B. & Durst, M. Non-random integration of the HPV genome in cervical cancer. *PloS one* **7**, e39632, doi:10.1371/journal.pone.0039632 (2012).

85 Xu, B. *et al.* Multiplex Identification of Human Papillomavirus 16 DNA Integration Sites in Cervical Carcinomas. *PloS one* **8**, e66693, doi:10.1371/journal.pone.0066693 (2013).

86 Kraus, I. *et al.* The majority of viral-cellular fusion transcripts in cervical carcinomas cotranscribe cellular sequences of known or predicted genes. *Cancer research* **68**, 2514-2522, doi:10.1158/0008-5472.CAN-07-2776 (2008).

87 Schmitt, M. *et al.* Diagnosing cervical cancer and high-grade precursors by HPV16 transcription patterns. *Cancer research* **70**, 249-256, doi:10.1158/0008-5472.CAN-09-2514 (2010).

88 Plummer, M. *et al.* Smoking and cervical cancer: pooled analysis of the IARC multi-centric case--control study. *Cancer causes & control : CCC* **14**, 805-814 (2003).

89 Collins, S., Rollason, T. P., Young, L. S. & Woodman, C. B. Cigarette smoking is an independent risk factor for cervical intraepithelial neoplasia in young women: a longitudinal study. *Eur J Cancer* **46**, 405-411, doi:10.1016/j.ejca.2009.09.015 (2010).

90 Berrington de Gonzalez, A., Sweetland, S. & Green, J. Comparison of risk factors for squamous cell and adenocarcinomas of the cervix: a meta-analysis. *British journal of cancer* **90**, 1787-1791, doi:10.1038/sj.bjc.6601764 (2004).

91 Parkin, D. M. The fraction of cancer attributable to lifestyle and environmental factors in the UK in 2010. *Br J Cancer* **105 Suppl 2**, S2-5, doi:10.1038/bjc.2011.474 bjc2011474 [pii] (2011).

92 Zondervan, K. T., Carpenter, L. M., Painter, R. & Vessey, M. P. Oral contraceptives and cervical cancer--further findings from the Oxford Family Planning Association contraceptive study. *Br J Cancer* **73**, 1291-1297 (1996).

93 Acladious, N. N. *et al.* Persistent human papillomavirus infection and smoking increase risk of failure of treatment of cervical intraepithelial neoplasia (CIN). *Int J Cancer* **98**, 435-439, doi:10.1002/ijc.10080 [pii] (2002).

94 Appleby, P. *et al.* Cervical cancer and hormonal contraceptives: collaborative reanalysis of individual data for 16,573 women with cervical cancer and 35,509 women without cervical cancer from 24 epidemiological studies. *Lancet* **370**, 1609-1621, doi:10.1016/S0140-6736(07)61684-5 (2007).

95 Munoz, N. *et al.* Role of parity and human papillomavirus in cervical cancer: the IARC multicentric case-control study. *Lancet* **359**, 1093-1101, doi:10.1016/S0140-6736(02)08151-5 (2002).

96 Hildesheim, A. *et al.* HPV co-factors related to the development of cervical cancer: results from a population-based study in Costa Rica. *Br J Cancer* **84**, 1219-1226, doi:10.1054/bjoc.2001.1779 S0007092001917793 [pii] (2001).

97 Calil, L. N., Igansi, C. N., Meurer, L., Edelweiss, M. I. & Bozzetti, M. C. Chlamydia trachomatis and human papillomavirus coinfection: association with p16INK4a and Ki67 expression in biopsies of patients with pre-neoplastic and neoplastic lesions. *Braz J Infect Dis* **15**, 126-131, doi:S1413-86702011000200006 [pii] (2011).

98 Deluca, G. D. *et al.* Chlamydia trachomatis as a probable cofactor in human papillomavirus infection in aboriginal women from northeastern Argentina. *Braz J Infect Dis* **15**, 567-572, doi:S1413-86702011000600011 [pii] (2011).

99 Smith, J. S. *et al.* Herpes simplex virus-2 as a human papillomavirus cofactor in the etiology of invasive cervical cancer. *J Natl Cancer Inst* **94**, 1604-1613 (2002).

100 Gravitt, P. E. *et al.* New technologies in cervical cancer screening. *Vaccine* **26 Suppl 10**, K42-52, doi:10.1016/j.vaccine.2008.05.002 (2008).

101 Stoler, M. H. HPV for cervical cancer screening: is the era of the molecular pap smear upon us? *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society* **49**, 1197-1198 (2001).

102 Fahey, M. T., Irwig, L. & Macaskill, P. Meta-analysis of Pap test accuracy. *American journal of epidemiology* **141**, 680-689 (1995).

103 Baldwin, P., Laskey, R. & Coleman, N. Translational approaches to improving cervical screening. *Nature reviews* **3**, 217-226 (2003).

104 Davey, E. *et al.* Effect of study design and quality on unsatisfactory rates, cytology classifications, and accuracy in liquid-based versus conventional cervical cytology: a systematic review. *Lancet* **367**, 122-132 (2006).

105 Ronco, G. *et al.* Accuracy of liquid based versus conventional cytology: overall results of new technologies for cervical cancer screening: randomised controlled trial. *BMJ (Clinical research ed* **335**, 28 (2007).

106 Cuzick, J. *et al.* Overview of human papillomavirus-based and other novel options for cervical cancer screening in developed and developing countries. *Vaccine* **26 Suppl 10**, K29-41, doi:10.1016/j.vaccine.2008.06.019 (2008).

107 Carreon, J. D. *et al.* CIN2 is a much less reproducible and less valid diagnosis than CIN3: results from a histological review of population-based cervical samples. *International journal of gynecological pathology : official journal of the International Society of Gynecological Pathologists* **26**, 441-446, doi:10.1097/pgp.0b013e31805152ab (2007).

108 Stoler, M. H. & Schiffman, M. Interobserver reproducibility of cervical cytologic and histologic interpretations: realistic estimates from the ASCUS-LSIL Triage Study. *JAMA : the journal of the American Medical Association* **285**, 1500-1505 (2001).

109 Dalla Palma, P. *et al.* The reproducibility of CIN diagnoses among different pathologists: data from histology reviews from a multicenter randomized study. *American journal of clinical pathology* **132**, 125-132, doi:10.1309/AJCPBRK7D1YIUWFP (2009).

110 Kjaer, S. *et al.* The absolute risk of cervical abnormalities in high-risk human papillomavirus-positive, cytologically normal women over a 10-year period. *Cancer research* **66**, 10630-10636, doi:10.1158/0008-5472.CAN-06-1057 (2006).

111 Cuzick, J. *et al.* Overview of the European and North American studies on HPV testing in primary cervical cancer screening. *Int J Cancer* **119**, 1095-1101, doi:10.1002/ijc.21955 (2006).

112 (DGGG), L. d. D. G. f. G. u. G. Prävention, Diagnostik und Therapie der HPV-Infektion und präinvasiver Läsionen des weiblichen Genitale. *AWMF online* (2010).

113 Bulkmans, N. W., Rozendaal, L., Voorhorst, F. J., Snijders, P. J. & Meijer, C. J. Long-term protective effect of high-risk human papillomavirus testing in population-based cervical screening. *British journal of cancer* **92**, 1800-1802 (2005).

114 Hoyer, H. *et al.* Cumulative 5-year diagnoses of CIN2, CIN3 or cervical cancer after concurrent high-risk HPV and cytology testing in a primary screening setting. *International journal of cancer* **116**, 136-143 (2005).

115 Bulkmans, N. W. *et al.* Human papillomavirus DNA testing for the detection of cervical intraepithelial neoplasia grade 3 and cancer: 5-year follow-up of a randomised controlled implementation trial. *Lancet* **370**, 1764-1772 (2007).

116 Katki, H. A. *et al.* Cervical cancer risk for women undergoing concurrent testing for human papillomavirus and cervical cytology: a population-based study in routine clinical practice. *The lancet oncology* **12**, 663-672, doi:10.1016/S1470-2045(11)70145-0 (2011).

117 Sherman, M. E. *et al.* Baseline cytology, human papillomavirus testing, and risk for cervical neoplasia: a 10-year cohort analysis. *Journal of the National Cancer Institute* **95**, 46-52 (2003).

118 Lee, K. R. *et al.* Comparison of conventional Papanicolaou smears and a fluid-based, thin-layer system for cervical cancer screening. *Obstetrics and gynecology* **90**, 278-284 (1997).

119 Remmink, A. J. *et al.* The presence of persistent high-risk HPV genotypes in dysplastic cervical lesions is associated with progressive disease: natural history up to 36 months. *Int J Cancer* **61**, 306-311 (1995).

120 Ho, G. Y. *et al.* Persistent genital human papillomavirus infection as a risk factor for persistent cervical dysplasia. *J Natl Cancer Inst* **87**, 1365-1371 (1995).

121 Franco, E. L. *et al.* Epidemiology of acquisition and clearance of cervical human papillomavirus infection in women from a high-risk area for cervical cancer. *The Journal of infectious diseases* **180**, 1415-1423, doi:10.1086/315086 (1999).

122 Woodman, C. B. *et al.* Natural history of cervical human papillomavirus infection in young women: a longitudinal cohort study. *Lancet* **357**, 1831-1836, doi:10.1016/S0140-6736(00)04956-4 (2001).

123 Wright, T. C., Jr. Natural history of HPV infections. *The Journal of family practice* **58**, S3-7 (2009).

124 Cancer, I. A. f. R. o. *Cervix Cancer Screening*. (IARC Press, 2005).

125 Cuzick, J. *et al.* New technologies and procedures for cervical cancer screening. *Vaccine* **30 Suppl 5**, F107-116, doi:10.1016/j.vaccine.2012.05.088 (2012).

126 Netherlands, H. C. o. t. Population screening for cervical cancer. (2011).

127 Cuzick, J. *et al.* A systematic review of the role of human papillomavirus testing within a cervical screening programme. *Health Technol Assess* **3**, i-iv, 1-196 (1999).

128 Malloy, C., Sherris, J., Herdman, C. HPV DNA Testing: Technical and Programmatic Issues for Cervical Cancer Prevention in Low-Resource Settings. *Path* (2000).

129 Meijer, C. J., Snijders, P. J. & Castle, P. E. Clinical utility of HPV genotyping. *Gynecologic oncology* **103**, 12-17, doi:10.1016/j.ygyno.2006.07.031 (2006).

130 Monsonego, J. *et al.* Detection of human papillomavirus genotypes among high-risk women: a comparison of hybrid capture and linear array tests. *Sexually transmitted diseases* **35**, 521-527, doi:10.1097/OLQ.0b013e318164e567 (2008).

131 Stevens, M. P. *et al.* Comparison of the Digene Hybrid Capture 2 assay and Roche AMPLICOR and LINEAR ARRAY human papillomavirus (HPV) tests in detecting high-risk HPV genotypes in specimens from women with previous abnormal Pap smear results. *Journal of clinical microbiology* **45**, 2130-2137, doi:10.1128/JCM.02438-06 (2007).

132 Castle, P. E. *et al.* A cross-sectional study of a prototype carcinogenic human papillomavirus E6/E7 messenger RNA assay for detection of cervical precancer and cancer. *Clin Cancer Res* **13**, 2599-2605 (2007).

133 Schmitt, M. *et al.* Prevalence and viral load of 51 genital human papillomavirus types and three subtypes. *Int J Cancer*, doi:10.1002/ijc.27891 (2012).

134 Schmitt, M. *et al.* Viral load of high-risk human papillomaviruses as reliable clinical predictor for the presence of cervical lesions. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, doi:10.1158/1055-9965.EPI-12-1067 (2013).

135 Kosel, S., Burggraf, S., Engelhardt, W. & Olgemoller, B. Increased levels of HPV16 E6*I transcripts in high-grade cervical cytology and histology (CIN II+) detected by rapid real-time RT-PCR amplification. *Cytopathology* **18**, 290-299 (2007).

136 Forslund, O., Lindqvist, P., Haadem, K., Czegledy, J. & Hansson, B. G. HPV 16 DNA and mRNA in cervical brush samples quantified by PCR and microwell hybridization. *J Virol Methods* **69**, 209-222 (1997).

137 Boulet, G. A. *et al.* Nucleic acid sequence-based amplification assay for human papillomavirus mRNA detection and typing: evidence for DNA amplification. *Journal of clinical microbiology* **48**, 2524-2529, doi:10.1128/JCM.00173-10 (2010).

138 Cuschieri, K. S., Whitley, M. J. & Cubie, H. A. Human papillomavirus type specific DNA and RNA persistence--implications for cervical disease progression and monitoring. *Journal of medical virology* **73**, 65-70, doi:10.1002/jmv.20062 (2004).

139 Molden, T., Kraus, I., Karlsen, F., Skomedal, H. & Hagmar, B. Human papillomavirus E6/E7 mRNA expression in women younger than 30 years of age. *Gynecologic oncology* **100**, 95-100, doi:10.1016/j.ygyno.2005.07.108 (2006).

140 Molden, T. *et al.* Comparison of human papillomavirus messenger RNA and DNA detection: a cross-sectional study of 4,136 women >30 years of age with a 2-year follow-up of high-grade squamous intraepithelial lesion. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **14**, 367-372, doi:10.1158/1055-9965.EPI-04-0410 (2005).

141 Szarewski, A. *et al.* Comparison of predictors for high-grade cervical intraepithelial neoplasia in women with abnormal smears. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **17**, 3033-3042, doi:10.1158/1055-9965.EPI-08-0508 (2008).

142 Cuschieri, K. & Wentzensen, N. Human papillomavirus mRNA and p16 detection as biomarkers for the improved diagnosis of cervical neoplasia. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **17**, 2536-2545, doi:10.1158/1055-9965.EPI-08-0306 (2008).

143 Khleif, S. N. *et al.* Inhibition of cyclin D-CDK4/CDK6 activity is associated with an E2F-mediated induction of cyclin kinase inhibitor activity. *Proc Natl Acad Sci U S A* **93**, 4350-4354 (1996).

144 Lesnikova, I., Lidang, M., Hamilton-Dutoit, S. & Koch, J. p16 as a diagnostic marker of cervical neoplasia: a tissue microarray study of 796 archival specimens. *Diagnostic pathology* **4**, 22, doi:10.1186/1746-1596-4-22 (2009).

145 Schmidt, D., Bergeron, C., Denton, K. J. & Ridder, R. p16/ki-67 dual-stain cytology in the triage of ASCUS and LSIL papanicolaou cytology: results from the European equivocal or mildly abnormal Papanicolaou cytology study. *Cancer Cytopathol* **119**, 158-166, doi:10.1002/cncy.20140 (2011).

146 Schmitt, M. *et al.* The HPV16 transcriptome in cervical lesions of different grades. *Molecular and cellular probes* **25**, 260-265, doi:10.1016/j.mcp.2011.05.003 (2011).

147 Algeciras-Schimnich, A. *et al.* Evaluation of quantity and staining pattern of human papillomavirus (HPV)-infected epithelial cells in thin-layer cervical specimens using optimized HPV-CARD assay. *Cancer* **111**, 330-338, doi:10.1002/cncr.22946 (2007).

148 Schmitt, M. & Pawlita, M. The HPV transcriptome in HPV16 positive cell lines. *Molecular and cellular probes* **25**, 108-113, doi:10.1016/j.mcp.2011.03.003 (2011).

149 Dondog, B. *et al.* Human papillomavirus infection in Ulaanbaatar, Mongolia: a population-based study. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **17**, 1731-1738, doi:17/7/1731 [pii] 10.1158/1055-9965.EPI-07-2796 (2008).

150 de Sanjose, S. *et al.* Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study. *The lancet oncology* **11**, 1048-1056, doi:10.1016/S1470-2045(10)70230-8 (2010).

151 Halec, G., Alemany, L., Lloveras, B., Schmitt, M., Alejo, M., Bosch, F.X., Tous, S., Klaustermeier, J.E., Guimera, N., Grabe, N., Lahrmann, B., Gissmann, L., Quint, W., Bosch, F.X., Sanjose, S., Pawlita, M. Pathogenic role of human papillomavirus (HPV) types 26, 53, 66, 67, 68, 70, 73 and 82 in cervical cancer. *submitted to JNCI* (2013).

152 Holzinger, D. *et al.* Viral RNA Patterns and High Viral Load Reliably Define Oropharynx Carcinomas with Active HPV16 Involvement. *Cancer research* **72**, 1-11, doi:10.1158/0008-5472.CAN-11-3934 (2012).

153 Pattillo, R. A. *et al.* Tumor antigen and human chorionic gonadotropin in CaSki cells: a new epidermoid cervical cancer cell line. *Science* **196**, 1456-1458 (1977).

154 Baldus, S. E. *et al.* Smad4 deficiency in cervical carcinoma cells. *Oncogene* **24**, 810-819, doi:10.1038/sj.onc.1208235 (2005).

155 Schroeder, A. *et al.* The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC molecular biology* **7**, 3, doi:10.1186/1471-2199-7-3 (2006).

156 Schmitt, M., Dondog, B., Waterboer, T. & Pawlita, M. Homogeneous amplification of genital human alpha papillomaviruses by PCR using novel broad-spectrum GP5+ and GP6+ primers. *Journal of clinical microbiology* **46**, 1050-1059, doi:10.1128/JCM.02227-07 (2008).

157 Schmitt, M. *et al.* Bead-based multiplex genotyping of human papillomaviruses. *Journal of clinical microbiology* **44**, 504-512 (2006).

158 Schmitt, M. *et al.* Abundance of multiple high-risk human papillomavirus (HPV) infections found in cervical cells analyzed by use of an ultrasensitive HPV genotyping assay. *Journal of clinical microbiology* **48**, 143-149, doi:10.1128/JCM.00991-09 (2010).

159 Pim, D., Massimi, P. & Banks, L. Alternatively spliced HPV-18 E6* protein inhibits E6 mediated degradation of p53 and suppresses transformed cell growth. *Oncogene* **15**, 257-264, doi:10.1038/sj.onc.1201202 (1997).

160 Sotlar, K. *et al.* Detection of high-risk human papillomavirus E6 and E7 oncogene transcripts in cervical scrapes by nested RT-polymerase chain reaction. *Journal of medical virology* **74**, 107-116, doi:10.1002/jmv.20153 (2004).

161 Hummel, M., Lim, H. B. & Laimins, L. A. Human papillomavirus type 31b late gene expression is regulated through protein kinase C-mediated changes in RNA processing. *Journal of virology* **69**, 3381-3388 (1995).

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

163 Zheng, Z. M. & Baker, C. C. Papillomavirus genome structure, expression, and post-transcriptional regulation. *Frontiers in bioscience : a journal and virtual library* **11**, 2286-2302 (2006).

164 Castro, F. V. *et al.* 5T4 oncofetal antigen is expressed in high risk of relapse childhood pre-B acute lymphoblastic leukemia and is associated with a more invasive and chemotactic phenotype. *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, U.K* **26**, 1487-1498, doi:10.1038/leu.2012.18 (2012).

165 Schmitt, M. & Pawlita, M. High-throughput detection and multiplex identification of cell contaminations. *Nucleic Acids Res* **37**, e119, doi:gkp581 [pii] 10.1093/nar/gkp581 (2009).

166 Nybo, K. Molecular biology techniques Q&A. *BioTechniques* **51**, 235, 237, doi:10.2144/000113747 (2011).

167 De la Rosa-Rios, M. A., Martinez-Salazar, M., Martinez-Garcia, M., Gonzalez-Bonilla, C. & Villegas-Sepulveda, N. The intron 1 of HPV 16 has a suboptimal branch point at a guanosine. *Virus research* **118**, 46-54, doi:10.1016/j.virusres.2005.11.010 (2006).

168 Sherman, M. E. *et al.* Determinants of human papillomavirus load among women with histological cervical intraepithelial neoplasia 3: dominant impact of surrounding low-grade lesions. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **12**, 1038-1044 (2003).

169 Dondog, B. Human Papilloma Virus Infection and Cervical Cancer in Mongolia. *Ruperto-Carola University of Heidelberg* **Dissertation** (2008).

170 Halec, G. *et al.* Biological evidence for a causal role of HPV16 in a small fraction of laryngeal squamous cell carcinoma. *British journal of cancer*, doi:10.1038/bjc.2013.296 (2013).

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

172 Pett, M. & Coleman, N. Integration of high-risk human papillomavirus: a key event in cervical carcinogenesis? *The Journal of pathology* **212**, 356-367, doi:10.1002/path.2192 (2007).

173 Cullen, A. P., Reid, R., Campion, M. & Lorincz, A. T. Analysis of the physical state of different human papillomavirus DNAs in intraepithelial and invasive cervical neoplasm. *J Virol* **65**, 606-612 (1991).

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

175 Clarke, M. A. *et al.* Human papillomavirus DNA methylation as a potential biomarker for cervical cancer. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **21**, 2125-2137, doi:10.1158/1055-9965.EPI-12-0905 (2012).

176 Kalantari, M., Lee, D., Calleja-Macias, I. E., Lambert, P. F. & Bernard, H. U. Effects of cellular differentiation, chromosomal integration and 5-aza-2'-deoxycytidine treatment on human papillomavirus-16 DNA methylation in cultured cell lines. *Virology* **374**, 292-303, doi:10.1016/j.virol.2007.12.016 (2008).

177 Kim, K., Garner-Hamrick, P. A., Fisher, C., Lee, D. & Lambert, P. F. Methylation patterns of papillomavirus DNA, its influence on E2 function, and implications in viral infection. *Journal of virology* **77**, 12450-12459 (2003).

178 Chaiwongkot, A. *et al.* Differential methylation of E2 binding sites in episomal and integrated HPV 16 genomes in preinvasive and invasive cervical lesions. *Int J Cancer* **132**, 2087-2094, doi:10.1002/ijc.27906 (2013).

179 Schmitt, M. *Multiplex HPV Genotyping- Simultaneous nucleic acid detection of multiple HPV types using bead technology* Diploma thesis, Ruperto-Carola University of Heidelberg, (2005).

180 Bustin, S. A. & Mueller, R. Real-time reverse transcription PCR (qRT-PCR) and its potential use in clinical diagnosis. *Clin Sci (Lond)* **109**, 365-379, doi:10.1042/CS20050086 (2005).

181 http://de-de.invitrogen.com/site/de/de/home/References/Ambion-Tech-Support/rtpcr-analysis/general-articles/top-ten-pitfalls-in-quantitative-real-time-pcr-primer.html. (2013).

182 Thorland, E. C., Myers, S. L., Gostout, B. S. & Smith, D. I. Common fragile sites are preferential targets for HPV16 integrations in cervical tumors. *Oncogene* **22**, 1225-1237, doi:10.1038/sj.onc.1206170 (2003).

183 Chen, Z. *et al.* Evolution and taxonomic classification of human papillomavirus 16 (HPV16)-related variant genomes: HPV31, HPV33, HPV35, HPV52, HPV58 and HPV67. *PloS one* **6**, e20183, doi:10.1371/journal.pone.0020183 (2011).

184 Häfner, N. *Physikalischer Status des HPV16 Genoms und die virale Onkogenexpression - Bedeutung für die Zervixkarzinogenese*, Friedrich-Schiller Universität (2007).

185 Mincheva, A., Gissmann, L. & zur Hausen, H. Chromosomal integration sites of human papillomavirus DNA in three cervical cancer cell lines mapped by in situ hybridization. *Medical microbiology and immunology* **176**, 245-256 (1987).

186 Leung, K. M., Lam, K. K., Tse, P. Y., Yeoh, G. P. & Chan, K. W. Characteristics of false-negative ThinPrep cervical smears in women with high-grade squamous intraepithelial lesions. *Hong Kong medical journal = Xianggang yi xue za zhi / Hong Kong Academy of Medicine* **14**, 292-295 (2008).

187 O'Sullivan, J. P. *et al.* A case-control study of true-positive versus false-negative cervical smears in women with cervical intraepithelial neoplasia (CIN) III. *Cytopathology : official journal of the British Society for Clinical Cytology* **9**, 155-161 (1998).

188 Fleige, S. & Pfaffl, M. W. RNA integrity and the effect on the real-time qRT-PCR performance. *Molecular aspects of medicine* **27**, 126-139, doi:10.1016/j.mam.2005.12.003 (2006).

189 Fleige, S. *et al.* Comparison of relative mRNA quantification models and the impact of RNA integrity in quantitative real-time RT-PCR. *Biotechnology letters* **28**, 1601-1613, doi:10.1007/s10529-006-9127-2 (2006).

190 Saunier, M. *et al.* Analysis of human papillomavirus type 16 (HPV16) DNA load and physical state for identification of HPV16-infected women with high-grade lesions or

cervical carcinoma. *Journal of clinical microbiology* **46**, 3678-3685, doi:10.1128/JCM.01212-08 (2008).

191 Stoler, M. H. *et al.* The interplay of age stratification and HPV testing on the predictive value of ASC-US cytology. Results from the ATHENA HPV study. *American journal of clinical pathology* **137**, 295-303, doi:10.1309/AJCPGW1V2BBWMOCX (2012).

192 Wentzensen, N., Sherman, M. E., Schiffman, M. & Wang, S. S. Utility of methylation markers in cervical cancer early detection: appraisal of the state-of-the-science. *Gynecologic oncology* **112**, 293-299, doi:10.1016/j.ygyno.2008.10.012 (2009).

193 Eijsink, J. J. *et al.* A four-gene methylation marker panel as triage test in high-risk human papillomavirus positive patients. *Int J Cancer* **130**, 1861-1869, doi:10.1002/ijc.26326 (2012).

194 Hesselink, A. T. *et al.* Combined promoter methylation analysis of CADM1 and MAL: an objective triage tool for high-risk human papillomavirus DNA-positive women. *Clinical cancer research : an official journal of the American Association for Cancer Research* **17**, 2459-2465, doi:10.1158/1078-0432.CCR-10-2548 (2011).

195 Mirabello, L. *et al.* Methylation of human papillomavirus type 16 genome and risk of cervical precancer in a Costa Rican population. *Journal of the National Cancer Institute* **104**, 556-565, doi:10.1093/jnci/djs135 (2012).

196 Snijders, P. J. *et al.* High-risk HPV testing on self-sampled versus clinician-collected specimens: a review on the clinical accuracy and impact on population attendance in cervical cancer screening. *Int J Cancer* **132**, 2223-2236, doi:10.1002/ijc.27790 (2013).

197 Gok, M. *et al.* HPV testing on self collected cervicovaginal lavage specimens as screening method for women who do not attend cervical screening: cohort study. *BMJ* **340**, c1040, doi:10.1136/bmj.c1040 (2010).

198 Igidbashian, S. *et al.* Self-collected human papillomavirus testing acceptability: comparison of two self-sampling modalities. *J Womens Health (Larchmt)* **20**, 397-402, doi:10.1089/jwh.2010.2189 (2011).

199 Jung, A. C. *et al.* Biological and clinical relevance of transcriptionally active human papillomavirus (HPV) infection in oropharynx squamous cell carcinoma. *Int J Cancer* **126**, 1882-1894, doi:10.1002/ijc.24911 (2010).

200 Skamperle, M., Kocjan, B. J., Maver, P. J., Seme, K. & Poljak, M. Human papillomavirus (HPV) prevalence and HPV type distribution in cervical, vulvar, and anal cancers in central and eastern Europe. *Acta dermatovenerologica Alpina, Panonica, et Adriatica* **22**, 1-5 (2013).

201 Stanley, M. A., Winder, D. M., Sterling, J. C. & Goon, P. K. HPV infection, anal intra-epithelial neoplasia (AIN) and anal cancer: current issues. *BMC cancer* **12**, 398, doi:10.1186/1471-2407-12-398 (2012).

202 Silvia de Sanjose´, L. A., Jaume Ordi, Sara Tous, Maria Alejo, *et al.* Worldwide human papillomavirus genotype attribution in over 2000 cases of intraepithelial and invasive lesions of the vulva. *accepted at European Journal of Cancer* (2013).

203 Bjorge, T. *et al.* Prospective seroepidemiological study of role of human papillomavirus in non-cervical anogenital cancers. *BMJ* **315**, 646-649 (1997).

204 Dittmer, C., Katalinic, A., Mundhenke, C., Thill, M. & Fischer, D. Epidemiology of vulvar and vaginal cancer in Germany. *Archives of gynecology and obstetrics* **284**, 169-174, doi:10.1007/s00404-011-1850-9 (2011).

10 Publications

Ehehalt D, Lener B, Pircher H, Dreier K, Pfister H, Kaufmann AM, Frangini S, Ressler S, Müller-Holzner E, Schmitt M, Höfler D, Rostek U, Kaiser A, Widschwendter A, Zwerschke W, Jansen-Dürr P. Detection of human papillomavirus type 18 E7 oncoprotein in cervical smears: a feasibility study. ***J Clin Microbiol.*** **2012 Feb; 50(2):246-57. doi: 10.1128/JCM.01108-11. Epub 2011 Nov 30.**

Schmitt M, Höfler D, Koleganova N, Pawlita M. Human polyomaviruses and other human viruses in neuroendocrine tumors. ***Cancer Epidemiol Biomarkers Prev.*** **2011 Jul; 20(7):1558-61. doi: 10.1158/1055-9965.EPI-11-0424. Epub 2011 May 25.**

Kostareli E, Holzinger D, Bogatyrova O, Hielscher T, Wichmann G, Keck M, Lahrmann B, Grabe N, Flechtenmacher C, Schmidt CR, Seiwert T, Dyckhoff G, Dietz A, Höfler D, Pawlita M, Benner A, Bosch FX, Plinkert P, Plass C, Weichenhan D and Hess J. HPV-related methylation signature correlates with survival in oropharyngeal squamous cell carcinomas. ***The Journal of clinical investigation*** **123, 2488-2501, doi:10.1172/JCI67010 (2013).**

Halec G, Holzinger D, Schmitt M, Flechtenmacher C, Dyckhoff G, Lloveras B, Höfler D, Bosch FX, Pawlita M. Biological evidence for causal role of HPV16 in a small fraction of laryngeal squamous cell carcinoma. ***British journal of cancer,*** **doi:10.1038/bjc.2013.296 (2013).**

Niebler M, Qian X, Höfler D, Kaufmann A, Ly R, Zawatzky R, Rösl F, Rincon-Orozco B. Degradation of IL-1ß is E6AP/proteasome mediated in HPV induced carcinogenesis affected a proper inflammasome function. **Accepted at PLOS pathogens.**

Höfler D, Böhmer G, Wasielewski R, Neumann H, Halec G, Holzinger D, Gissmann L, Pawlita M, Schmitt M. Validity of HPV16 RNA patterns as diagnostic marker for cervical cancer precursor lesions. **In preparation.**

One side project that had been started during my diploma thesis was the further development of three multiplex assays detecting DNA-viruses (rDVF), RNA-viruses (rRVF), bacteria and one fungus (rBFF) in laboratory rodents.

Briefly, conventional methods used to date for the detection of pathogens as enzyme-linked immunosorbent assay (ELISA), immunofluorescence and singleplex PCR assays are expensive, time-consuming, not standardised or lack high-throughput abilities. We developed multiplex assays based on multiplex PCR and multiplex detection of amplimers by Luminex for the simultaneous identification of pathogens in mice and rats, transplantable tumours, embryos, cell lines and other biological materials.

The rDVF and rRVF are already integral part of laboratory animal quality assurance program of the DKFZ, supplementing traditional serology, virology and pathology techniques as it is time- and cost-efficient, and reduces the number of animals needed.

Following publications of this project are in preparation:

Höfler D, Nicklas W, Mauter P, Klaessen E, Pawlita M., Schmitt M. A bead-based multiplex assay for the detection of DNA viruses in laboratory rodents and biological material. **Submitted to Journal of Virological Methods.**

Höfler D, Nicklas W, Mauter P, Pawlita M., Schmitt M. A bead-based multiplex assay for the detection of bacteria and fungus in laboratory rodents and biological material. **In preparation.**

Höfler D, Nicklas W, Mauter P, Klaessen E, Pawlita M., Schmitt M. A bead-based multiplex assay for the detection of RNA viruses in laboratory rodents and biological material. **In preparation.**