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-Biologe Bo Xu

born in: Penglai, Shandong Province, P. R. China

Oral-examination:

Integration of Human Papillomavirus Type 16 DNA in Cervical Carcinogenesis: Design of a novel strategy for HPV16 integration site determination in cervical scrapes and analysis of HPV16-induced *c-myc* insertional mutagenesis

Referees: **Prof. Dr. Elisabeth Schwarz**

Prof. Dr. Gabriele Petersen

Acknowledgements

My very special gratitude goes to my supervisor Prof. Dr. Elisabeth Schwarz for giving me the opportunity to pursue my PhD thesis in her group at DKFZ (F030), for her excellent guidance and constant support during the past four years, and for the critical reading of my dissertation manuscript.

My sincere thanks also go to Prof. Dr. Gabriele Petersen (University of Heidelberg) for kindly being my second supervisor and to PD. Dr. Stefan Wiemann (DKFZ) for acting as a member of my thesis advisory committee. I feel very grateful to both of them for participating actively in my progress reports.

Within the Cancéropôle du Grand Est-DKFZ cooperation project, I would particularly like to thank Dr. Véronique Dalstein and Prof. Dr. Christine Clavel in Reims (Université de Reims Champagne-Ardenne) as well as Dr. Maëlle Saunier and Dr. Jean-Luc Prétet in Besançon (Université de Franche-Comté) for providing clinical DNA samples and performing the E2-E6 quantitative PCR.

Essential for the development of ASP16 strategy, I would like to thank Sasithorn Chotewutmontri in our laboratory for writing the bioinformatics programs and for her large contributions in the pyrosequence data analysis. I would also like to thank Ursula Klos in our laboratory for technical assistance, and Ilona Braspenning-Wesch for introduction into the Western blot technique.

I would like to thank Andreas Hunziker for Sanger sequencing and Dr. Gerald Nyakatura for performing GS-FLX pyrosequencing at the DKFZ Genomics and Proteomics Core Facility.

I would sincerely like to thank Dr. Jérôme Couturier for performing FISH and karyotyping assays, and Dr. Martine Peter for performing DIPS-PCR and *c-myc* quantitative PCR at the Institut Curie, Paris.

Finally, I wish to thank Prof. Dr. Frank Rösl and all the members at the Division of Viral Transformation Mechanisms (F030) for offering a very comfortable working atmosphere and plentiful scientific events.

Zusammenfassung

Persistierende Infektionen mit einem humanen Papillomvirus (HPV) der Hochrisiko-Gruppe, meistens mit HPV16, bilden die notwendige Voraussetzung zur Entstehung von Gebärmutterhalskrebs (Zervixkarzinom). Bei der Tumorprogression kommt es häufig zur Integration der HPV-DNA in das Genom der Wirtszellen an unterschiedlichen Zielorten. Die Integration von HPV-DNA kann durch Insertionsmutagenese zur Veränderung flankierender zellulärer Gene führen und auch auf diese Weise zum Mehrstufenprozess der Karzinogenese beitragen. Das primäre Ziel dieser Arbeit war die Entwicklung einer neuen effizienten Strategie, die es ermöglicht, HPV16-DNA-Integrationsstellen in ausgewählten Proben von präkanzerösen und kanzerösen Zervixabstrichen auf Sequenzebene zu bestimmen.

Bei Verwendung von Restriktionsstellen-PCR zur HPV16-DNA-Integrationsanalyse stellte sich die Zervixkarzinom-Zelllinie MRI-H186 als ein sehr interessanter Fall für HPV16-induzierte Insertionsmutagenese des *c-myc* Proto-Onkogens heraus. Deshalb wurden im Rahmen der Arbeit detaillierte Untersuchungen durchgeführt. Assoziiert mit der HPV16-DNA-Integration im *c-myc*-Locus wurde eine Überexpression von *c-myc* auf mRNA- und Proteinebene festgestellt. Verschiedene Varianten integrierter HPV16-DNA wurden auf Sequenzebene charakterisiert. Eine dieser Varianten zeigte eine starke Expression als *myc*-HPV16 Hybridtranskript sowie als MycHPV16E2 Fusionsprotein. Beide *myc*-HPV16-Hybridmoleküle sind ein besonderes Charakteristikum von MRI-H186. Im MycHPV16E2-Fusionsprotein ist der aminoterminale Teil der MYC-Transaktivierungsdomäne mit der Linker- und DNA-Bindungsdomäne des HPV16 E2-Proteins fusioniert. Die Ergebnisse für MRI-H186 verstärken die Vermutung, dass Insertionsmutagenese, in diesem Fall die Aktivierung/Mutation von *c-myc*, bei der Zervixkarzinom-Entstehung eine Rolle spielen kann.

Für die HPV16-DNA-Integrationsanalyse in klinischen Proben, von denen genomische DNA nur in Nanogramm-Mengen und häufig degradiert zur Verfügung steht, wurde eine neue Strategie entwickelt und ausgetestet, die die gleichzeitige Sequenzierung vieler DNA-Proben ermöglicht. Die Strategie erhielt die Bezeichnung ASP16, da sie Gesamtgenom-Amplifikation (A), HPV16-DNA-Selektion (S) und Hochdurchsatz-Pyrosequenzierung (P) gefolgt von bioinformatischer Datenanalyse miteinander kombiniert. Zwei ASP16-Experimente wurden durchgeführt, mit denen die Ausführbarkeit der neuen Strategie bewiesen werden konnte. Die bekannte 3' viral-zelluläre Verbindungsstelle der integrierten HPV16-DNA in MRI-H186 konnte in beiden Experimenten nachgewiesen werden. HPV16-DNA-Integrationsstellen wurden in solchen klinischen Proben identifiziert, die einen hohen Anteil an integrierter HPV16 DNA enthielten. Krebsbezogene zelluläre Gene an den Integrationsstellen, besonders *Gli1* und *Grem2*, sind interessante Kandidaten für eine HPV16-induzierte Insertionsmutagenese. Zukünftige Optimierungen der ASP16-Methode sollten hauptsächlich eine Erhöhung der durchschnittlichen Sequenzlängen zum Ziel haben, damit alle möglichen viralen DNA-Bruchpunkte in der HPV16 E1/E2-Region vollständig erfasst werden können.

Die ASP16-Strategie bietet erstmalig die Möglichkeit, in klinischen Proben von Zervixabstrichen die HPV16-DNA-Integrationsstellen in einem Multiplex-Ansatz systematisch zu bestimmen, unter besonderer Berücksichtigung der geringen Menge und Qualität der Proben-DNA. Dieses besondere Merkmal der ASP16-Strategie wird eine Eingliederung der HPV16-DNA-Integrationsanalyse in Zervix-Screeningprogramme ermöglichen. Die Verwendung der ASP16-Methode wird dazu beitragen, progressive und hochgradig veränderte Zervix-Läsionen eindeutig zu identifizieren sowie weitere Einblicke in die molekularen Mechanismen der Zervixkarzinom-Entstehung zu erhalten.

Abstract

Persistent infection with high-risk human papillomavirus (HPV) types, most frequently HPV16, is a necessary cause of cervical cancer. Correlated with tumor progression, the circular HPV DNA usually becomes integrated into the host cell genome with diverse target sites. HPV DNA integration may induce alterations of flanking cellular genes by insertional mutagenesis that contribute to the multistep process of carcinogenesis. The primary goal of this study was the design of a novel strategy to determine HPV16 DNA integration sites at the sequence level in selected pre-cancerous and cancerous lesions obtained from cervical scrapes.

While testing restriction-site PCR for HPV16 DNA integration analysis, cell line MRI-H186 was identified as an interesting case of HPV16-induced insertional mutagenesis of the *c-myc* proto-oncogene. Associated with HPV16 DNA integration into the *c-myc* locus, over-expression of *c-myc* was found at both mRNA and protein levels. In addition, several HPV16 integration variants were characterized at the sequence level. One of these variants is strongly expressed as *myc*-HPV16 hybrid transcript and MycHPV16E2 fusion protein both unique for MRI-H186. In MycHPV16E2, the amino-terminal part of the c-MYC trans-activation domain is fused to the linker and DNA-binding domain of HPV16 E2. These data strengthen the assumption that insertional mutagenesis, in this case *c-myc* activation/mutation, contributes to cervical carcinogenesis.

For HPV16 DNA integration analysis in clinical samples from which genomic DNA is available only in nanogramme amounts and often degraded, a novel strategy enabling the simultaneous sequencing of multiple DNA samples was developed and examined in the present study. This strategy was named ASP16 because it combines whole genome amplification (A), HPV16 DNA selection (S) and high-throughput pyrosequencing (P) followed by bioinformatics data analysis. Two ASP16 experiments were performed, which have proven the feasibility of the novel strategy. The known 3' viral-cellular junction sequence of integrated HPV16 DNA in MRI-H186 was identified in both experiments. HPV16 DNA integration sites were determined in clinical samples harbouring high percentages of integrated HPV16 DNA. Cancer-related cellular genes near or at the integration sites, namely *Gli1* and *Grem2*, hold great potentials as candidates for HPV16-induced insertional mutagenesis. Future optimizations of the ASP16 method will focus primarily on increasing the average sequence length for a complete coverage of all possible viral DNA breakpoints in the HPV16 E1/E2 region.

Taken together, the ASP16 strategy offers for the first time the opportunity to systematically explore HPV16 DNA integration sites in a multiplex manner in clinical samples obtained from cervical scrapes, with particular respect to low quantity and poor quality of the template DNA. This unique feature is of great interest for the incorporation of HPV16 DNA integration analysis into routine cervical screening programs. Application of APS16 will contribute unequivocally to identify lesions of progressive or highly advanced disease and to obtain further insights into the underlying molecular mechanisms of cervical carcinogenesis.

Abbreviations

AA	amino acid
APOT	amplification of papillomavirus oncogene transcripts
ASP	amplification-selection-pyrosequencing strategy
bp	base pair(s)
cDNA	complementary DNA
CGE	Cancéropôle du Grand-Est
CIN	cervical intraepithelial neoplasia
DIPS-PCR	detection of integrated papillomavirus sequences PCR
DKFZ	Deutsches Krebsforschungszentrum
dsDNA	double-stranded DNA
FISH	fluorescence *in situ* hybridization
GPUA	GenomePlex universal adapter
GSP	gene-specific primer
HPV	human papillomavirus
kb	kilo base pairs
LCR	long control region
Mb	mega base pairs
MDA	multiple displacement amplification
miRNA	microRNA
nt	nucleotides
ORF	open reading frame
PCR	polymerase chain reaction
qPCR	quantitative PCR
RA	Roche-A
RB	Roche-B
RACE	rapid amplification of cDNA ends
RS-PCR	restriction-site PCR
RT	room temperature
RT-PCR	reverse transcription-PCR
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
SIL	squamous intraepithelial lesion
ssDNA	single-stranded DNA
URR	upstream regulatory region
WGA	whole genome amplification

1. Introduction

1.1 Human papillomaviruses and cervical cancer

Cancer of cervix uteri is the second most frequent malignant disease diagnosed in women all over the world (Parkin *et al.*, 2005). It has been estimated that in 2002 there were about 493000 new cases of cervical cancer and 274000 deaths attributed to this sexually transmitted disease worldwide (Parkin *et al.*, 2005). The human population comprises about 2329 million females currently at the age of 15 or older, who are at the risk of developing cervical cancer and its precursors (Castellsagué *et al.*, 2007).

According to the cell type involved, cervical cancer can be histologically divided into three categories: squamous cell carcinoma, adenocarcinoma and adenosquamous carcinoma. Most reported cases of cervical cancer are squamous cell carcinomas, while only about 11.4% are ascribed to adenocarcinomas or adenosquamous carcinomas (Vizcaino *et al*., 1998). The role of human papillomavirus (HPV) as etiologic agent for cervical carcinogenesis was initially hypothesized in mid-1970s (zur Hausen *et al*., 1974; zur Hausen, 1976 and 1977). In early 1980s, zur Hausen and co-workers reported the isolation of HPV16 and HPV18 DNA directly from cervical carcinoma biopsies for the first time (Dürst *et al*., 1983; Boshart *et al*., 1984). Meanwhile, HPV DNA has been detected in as much as 99.7% of invasive cervical squamous cell carcinomas worldwide (Bosch *et al*., 1995; Walboomers *et al*., 1999). Analysis of pooled data from international case-control studies has also shown that about 93% of invasive cervical adenocarcinomas and adenosquamous carcinomas are infected with HPV (Castellsagué *et al.*, 2006). Although other risk factors are also contributive (Castellsagué and Muñoz, 2003), the causal relation between HPV infection and cervical carcinogenesis has been undoubtedly recognized and accepted (zur Hausen, 2002; Bosch *et al*., 2002), based on lines of molecular epidemiologic evidence from retrospective case-control and prospective cohort studies (Koutsky *et al*., 1992; Muñoz *et al*., 1992; Schiffman *et al*., 1993; Bosch *et al*., 1995; Kjær *et al*., 1996; Ho *et al*., 1998; Walboomers *et al*., 1999; Liaw *et al*., 1999; Woodman *et al*., 2001). As a consequence of this well-established concept, recently a variety of prophylactic vaccines against HPV infection have been developed and applied, in order to prevent the global burden of cervical cancer fundamentally (Villa *et al*., 2005; Harper *et al*., 2006; Future II Study Group, 2007; Garland *et al*., 2007).

Over 100 distinct types of HPV have been identified so far (de Villiers *et al*., 2004), causing a wide spectrum of proliferative epithelial lesions from benign warts to malignant carcinomas (zur Hausen, 2002). Among them, about 40 types are classified as mucosotropic (or genital) HPVs, which all infect mucosal epithelial cells of the genital tract. Other HPV types belong to the category of cutaneous HPVs, which are generally found in common skin warts, *Epidermodysplasia verruciformis* (EV) lesions or immune-suppressed individuals (de Villiers *et al*., 2004). Based on the prevalence in cervical squamous cell carcinomas, a subset of at least 15 genital types is epidemiologically classified as high-risk HPVs, which are significantly associated with cervical carcinogenesis (Muñoz *et al*., 2003). This group contains HPV types 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68, 73 and 82. By contrast, 12 low-risk genital types (especially HPV6 and HPV11) are common in genital warts, but rarely found in cervical cancer. Two high-risk HPV types, HPV16 and HPV18, are regarded as major contributors to cervical cancer, which account for about 70% (HPV16, 55%; HPV18, 15%) of the invasive cervical carcinomas reported worldwide (Clifford *et al*., 2003; Smith *et al*., 2007).

1.2 HPV genome organization and transcription map

The genomic information of HPV is compacted into circular double-stranded DNA with a size of about 8 kb, enclosed by a non-enveloped icosahedral capsid shell. The HPV16 DNA genome encodes eight open reading frames (ORFs) and can be divided into three subgenomic regions: early, late and non-coding region (*Figure 1-1*). The early region contains six ORFs, E1, E2, E4, E5, E6 and E7, which are expressed in the early stage of HPV16 life cycle. The late region contains two ORFs L1 and L2, which are expressed in the late stage. Between the L1 and E6 ORFs, there is a segment (about 1 kb) of non-coding DNA named upstream regulatory region (URR) or long control region (LCR), which harbours the viral DNA replication origin and regulatory elements (promoter and enhancer) essential for viral gene expression.

Two major promoters, p97 and p670, are responsible for HPV16 gene expression (Zheng and Baker, 2006). The common early primary transcript starts from the early promoter p97 and is terminated at the early polyadenylation site at position 4215. It is alternatively spliced into various polycistronic mRNA species with different coding potentials for the six early proteins including E6, E7, E1, E2, E4 and E5 (*Figure 1-2*; species A-N). Because the E4 ORF does not have an own start-codon, E4 is

translated as a fusion-protein named E1^E4 together with the five N-terminal amino acids from E1. The major late primary transcript starts from the late promoter p670 (within ORF E7) and is terminated at the late polyadenylation site at position 7321. It is also spliced into different mRNA species important for the expression of the two late proteins L1 and L2 (*Figure 1-2*; species R-T).

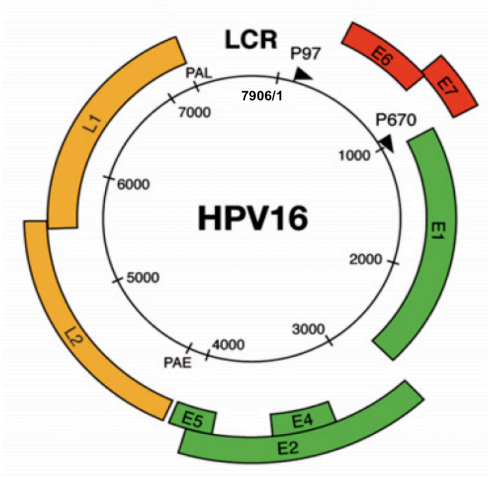


Figure 1-1: The genomic organization of HPV16 DNA. The HPV16 DNA genome (7906 bp; circular double-stranded DNA) and approximate positions of the eight HPV16-encoded ORFs are schematically presented. The long control region (LCR) is located between L1 and E6 ORFs. The HPV16 early (p97) and late (p670) promoters are marked as black arrows. The HPV16 early (PAE) and late (PAL) polyadenylation sites are also shown. The figure was taken from Doorbar (2006).

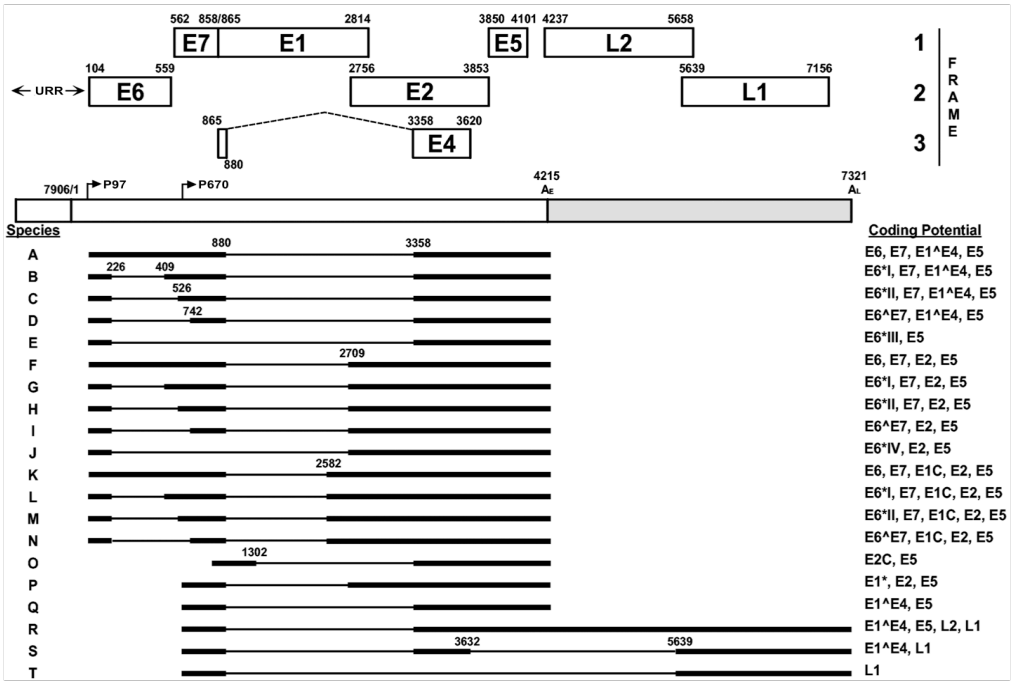


Figure 1-2: The HPV16 transcription map and coding potential of individual mRNA species. The HPV16 genome is shown in linear form, starting with the upstream regulatory region (URR) and followed by early and late regions. Positions for early (p97) and late (p670) promoters as well as for early (A_E) and late (A_L) polyadenylation sites are shown. The eight HPV16 ORFs are illustrated with positions of start and stop codons. The ORF E4 is spliced to the 5'-end of E1 at position 880. Various mRNA species derived from different usage of promoters and alternative splicing sites are shown together with their protein-coding potentials. For each species, exons (heavy lines), introns (thin lines) and positions of alternative splicing sites are demonstrated. The figure was taken from Zheng and Baker (2006).

1.3 HPV productive life cycle

The natural host cells of HPV are human keratinocytes of the stratified squamous epithelium (zur Hausen, 2002). HPV infection is thought to begin with the access of viral particles to cells of the basal layer (*Figure 1-3*), which are able to proliferate and differentiate (Fehrmann and Laimins, 2003). HPV DNA replication is strictly dependent on the cellular DNA replication machinery, but also requires the viral proteins E1 and E2. E1 is a DNA helicase highly conserved among different HPV types, which binds to a consensus motif in the viral DNA replication origin, but with relatively weak affinity (Hughes and Romanos, 1993; Dixon *et al*., 2000). The E2 protein as a replication factor helps to recruit E1 on its target site and facilitates the initial binding between them (Mohr *et al*., 1990).

At the latent stage of infection, the HPV DNA genome is moderately replicated in synchrony with cellular DNA synthesis (*Figure 1-3*). The subsequent productive stage of HPV infection is tightly coupled with epithelial differentiation (Doorbar, 2006). During the upward differentiation from basal cells to differentiated keratinocytes in the upper layers (*Figure 1-3*), the vegetative viral DNA amplification occurs, which is followed by the synthesis of capsid proteins and assembly of infectious virion particles (Doorbar, 2006). In normal epithelia, daughter basal cells exit from cell cycle and undergo terminal differentiation, as soon as they migrate into the suprabasal layers (Madison, 2003). However, in HPV-infected cells the differentiation process is retarded, the S-phase entry is activated, but the cell cycle is arrested at the G2/M boundary, thereby providing an environment permissive for both viral and genomic DNA replication (Cheng *et al*., 1995; Sherman *et al*., 1997; Davy and Doorbar, 2007). The re-entry of host cells into S-phase is believed to result from the transforming activities of viral oncoproteins E6 and E7, but the underlying mechanism for G2/M-arrest remains unclear. E7 promotes cell cycle progression, through direct association with the tumor suppressor protein pRB and the concurrent release of oncogenic transcription factor E2F (Dyson *et al*., 1989; Münger *et al*., 1989a; Chellappan *et al*., 1992; Boyer *et al*., 1996). E6 binds to and induces the ubiquitin-mediated degradation of tumor suppressor protein p53, thereby disrupting the p53-mediated apoptotic pathways (Werness *et al*., 1990; Scheffner *et al*., 1990; Scheffner *et al*., 1993). Expression analysis suggested that E4 and E5 proteins may also contribute to viral genome amplification, but less is known about their functions (Doorbar, 2006).

As cells move to the upper epithelial layers (*Figure 1-3*), capsid proteins L1 and L2 are expressed from late transcripts driven by the late promoter (p670 for HPV16), whose activation is believed to depend on cellular differentiation (Bedell *et al*., 1991; Higgins *et al*., 1992). Progeny virions are released from the tissue by regular desquamation, after the infected cells arrive at the epithelial surface (Fehrmann and Laimins, 2003).

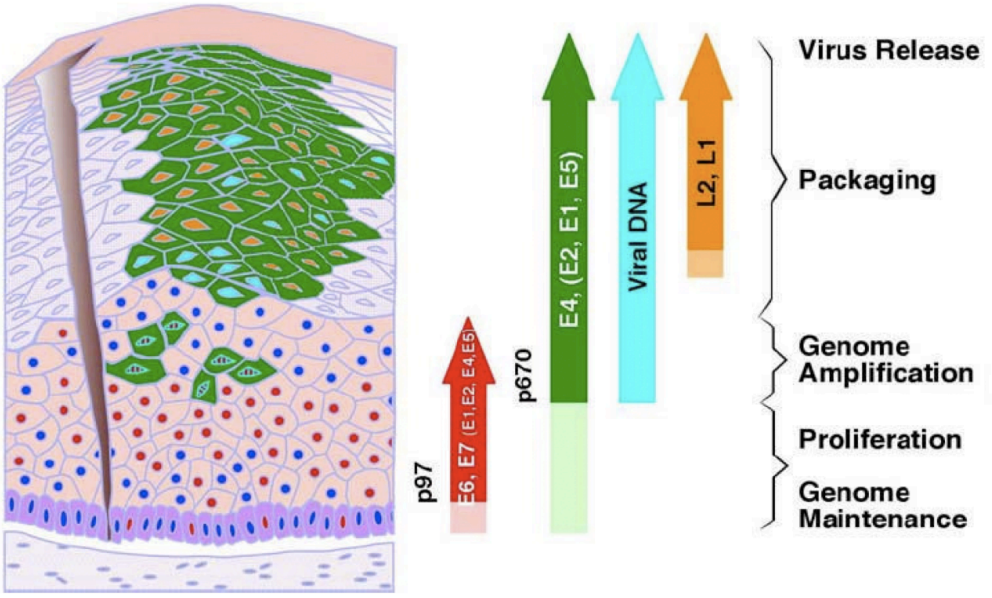


Figure 1-3: HPV16 productive life cycle and the molecular events involved. The stratified epithelium with different layers is schematically demonstrated on the left. Featured molecular events during the productive HPV16 infection are shown as arrows in the middle. Different stages of HPV16 life cycle are shown on the right. E6 and E7 (red circles) expressed from the early promoter p97 are present in the lower layers. Dark blue circles indicate the nuclei of normal, uninfected epithelial cells. Late events including viral DNA amplification and expression of capsid proteins (L1 and L2) are driven by the use of late promoter p670. The expression of E4 gene (green) is upregulated at this stage. The figure was taken from Middleton *et al* (2003).

1.4 The natural history of HPV infection

It has been estimated that up to 80% of the general female population carry HPV at a certain time of life (Brown *et al*., 2005). Fortunately, most infections are transient and not destined to develop cervical cancer (Ho *et al*., 1998; Moscicki *et al*., 1998; Woodman *et al*., 2001). Cervical carcinogenesis is a multi-stage process (Snijders *et al*., 2006), which generally takes over 10 years from the first infection with high-risk HPVs to invasive carcinomas (Wallin *et al*., 1999). Cervical cancer develops from stepwise, morphologically distinguishable pre-cancerous lesions (Holowaty *et al*., 1999). There are two nomenclature systems widely applied to define the different

grades of cervical pre-cancerous lesions (*Figure 1-4*): the cervical intraepithelial neoplasia (CIN) terminology defined by histopathological features (Richart, 1967), and the squamous intraepithelial lesion (SIL) system based on cytological abnormality (Solomon *et al*., 2002).

In general, about 80% of HPV infections are effectively cleared by host immune response (Ho *et al*., 1998; Franco *et al*., 1999; Baseman and Koutsky, 2005). From the other 20% of persistent infections, only few may progress to invasive cervical carcinomas ultimately (Ostör, 1993; Holowaty *et al*., 1999). It seems that low-grade lesions (CIN1 or low-grade SIL) represent the cytopathology of productive HPV infection (Peh *et al*., 2002; Middleton *et al*., 2003), and they often regress spontaneously (Cox *et al*., 2003; Moscicki *et al*., 2004). In contrast, high-grade lesions (CIN2/CIN3 or high-grade SIL) are considered as critical precursors to cervical cancer (Schiffman *et al*., 2007). However, only 10-20% of high-grade cervical pre-cancerous lesions will eventually progress to malignancy for unclear reasons (Nobbenhuis *et al*., 2001). In order to avoid over- and under-treatment in clinical practice, it is of particular interest to distinguish between cervical lesions with low and high risk for neoplastic progression.

Persistent infection with high-risk HPVs is the necessary risk factor for cervical cancer and its precursors (Remmink *et al*., 1995; Moscicki *et al*., 1998; Nobbenhuis *et al*., 1999; Schlecht *et al*., 2001; Kjær *et al*., 2002). The two oncoproteins E6 and E7 of high-risk HPVs are obligatory for the induction and maintenance of transformed phenotypes in cervical cancer cells (von Knebel Doeberitz *et al*., 1992; Münger and Howley, 2002). However, E6 and E7 activities are not sufficient for the development of malignant transformation (zur Hausen, 2000 and 2002). Although *in vitro* studies have shown the efficient immortalization of primary human keratinocytes by high-risk HPV E6 and E7 (Münger *et al*., 1989b; Hawley-Nelson *et al*., 1989; Hudson *et al*., 1990; Halbert *et al*., 1991), the HPV-immortalized cells are not tumorigenic in nude mice, unless the co-expression of additional cellular oncogenes (DiPaolo *et al*., 1989; Dürst *et al*., 1989; Pei *et al*., 1993; Narisawa-Saito *et al*., 2008). Therefore, it is suggested that in addition to the transforming activities of high-risk HPV oncogenes, the accumulation of yet unclear genetic and/or epigenetic alterations of host cells is also required for cervical malignant transformation.

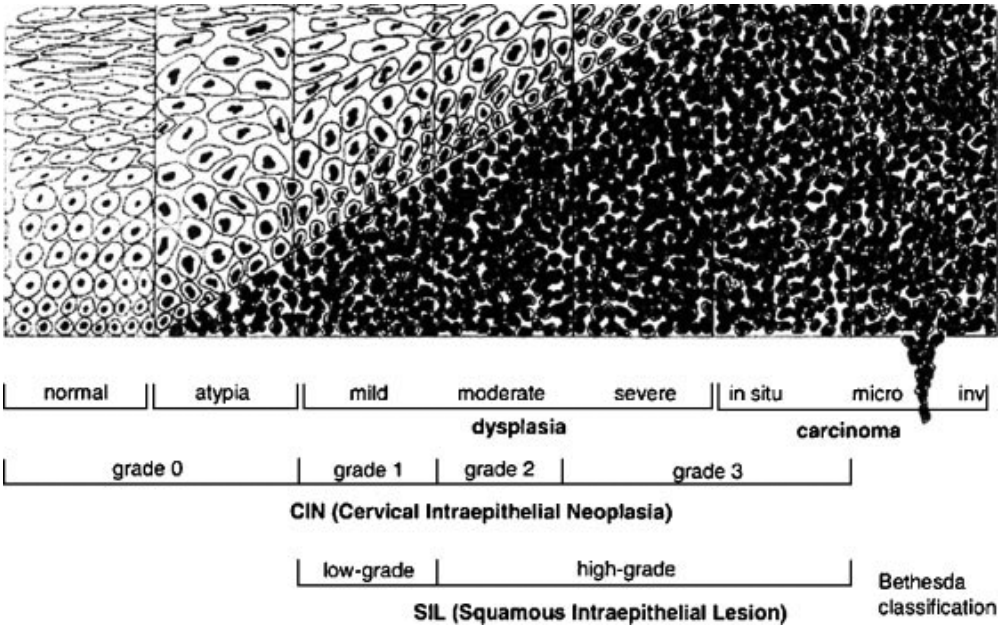


Figure 1-4: Classification of cervical pre-cancerous lesions. The morphological abnormalities observed in cervical pre-malignant lesions (dysplasia and carcinoma in situ) and invasive carcinomas are schematically demonstrated. The correlation between CIN and SIL nomenclature systems for the classification of cervical pre-cancerous lesions is shown. (micro = microinvasive; inv = invasive). The figure was taken from Snijders *et al* (2006).

1.5 HPV DNA integration in cervical carcinogenesis

HPV DNA integration into the host cell genome is a frequent event during cervical carcinogenesis (Pett and Coleman, 2007), which has been associated with neoplastic progression from pre-malignant lesions to invasive carcinomas (Dürst *et al*., 1985; Cullen *et al*., 1991; Daniel *et al*., 1995; Klaes *et al*., 1999; Tonon *et al*., 2001; Hopman *et al*., 2004; Hudelist *et al*., 2004). Integrated HPV DNA could be found in up to 88% of HPV16-positive and almost 100% of HPV18-positive cervical carcinomas (Cullen *et al*., 1991; Klaes *et al*., 1999; Kalantari *et al*., 2001). In clinical practice, HPV DNA integration has been shown to correlate with poor prognosis of cervical cancer and low-level of disease-free survival (Unger *et al*., 1995; Vernon *et al*., 1997; Kalantari *et al*., 1998).

The integration of HPV into cellular DNA is thought to be mediated by non-homologous DNA recombination mechanisms (Pett and Coleman, 2007). Upon integration, the episomal HPV DNA becomes linearized and inserted into human chromosomes (Schwarz *et al*., 1985). The breakpoints of HPV DNA are located in the E1 or E2 ORF from the early region, and in the L1 or L2 ORF from the late region, respectively (*Figure 1-5*). HPV DNA integration often results in disruption or

deletion of the E2 ORF, while the URR region together with E6 and E7 oncogenes are always retained (Schwarz *et al*., 1985). The transcription of integrated HPV DNA is activated by enhancer elements in the URR and started from the early promoter (*Figure 1-5*). Since the canonical HPV early polyadenylation site is missing, the transcription is terminated at the next possible polyadenlyation site in downstream cellular sequences. Finally, viral-cellular hybrid transcripts mainly for the expression of E6 and E7 oncoproteins are produced (Schneider-Gädicke and Schwarz, 1986; Baker *et al*., 1987).

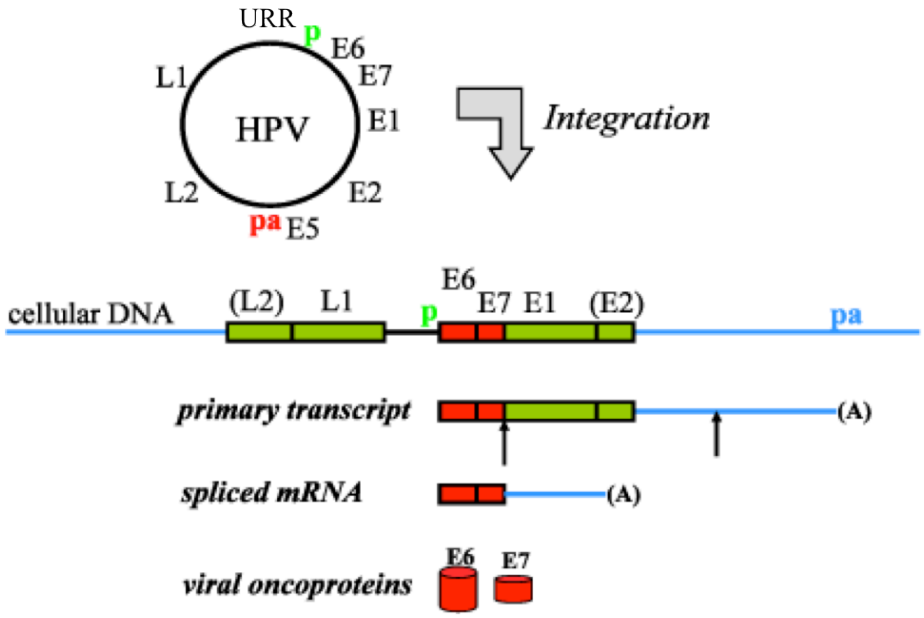


Figure 1-5: Structure and expression of integrated HPV DNA. Upon integration, the circular HPV DNA is broken in ORF E1 or E2 from the early region, and in ORF L1 or L2 from the late region. Since the URR, promoter and E6/E7 region are always retained, transcription can start from the viral early promoter (p), continue into flanking cellular sequences and stop at an appropriate cellular polyadenylation site (pa). The viral-cellular hybrid transcript is used for the expression of viral oncoproteins E6 and E7. The figure was kindly provided by Elisabeth Schwarz (DKFZ, Heidelberg).

Previous data have suggested that integrated HPV DNA provides a selective growth advantage for the host cells, in comparison with episomal DNA (Jeon *et al*., 1995). HPV DNA integration leads to the permanent incorporation of HPV DNA into the human genome, thereby enabling the constitutive expression of E6 and E7 oncogenes for cellular transformation. The multifunctional viral E2 protein acts as a repressor of HPV early gene expression (Bernard *et al*., 1989; Romanczuk *et al*., 1990). Since the E2 ORF is often disrupted in the integrated HPV DNA, HPV DNA integration may induce the over-expression of HPV oncogenes (Romanczuk and Howley, 1992). It has also been shown that the viral-cellular hybrid transcripts from

integrated HPV DNA have longer half-lives than the normal HPV early transcripts from episomal DNA (Jeon and Lambert, 1995).

The HPV DNA integration sites identified so far are dispersed all over the human genome (Wentzensen *et al*., 2004; Kraus *et al*., 2008). There are no specific target sequences on human chromosomes for HPV DNA integration. In general, HPV DNA integration occurs preferentially in transcriptionally active regions and common fragile sites (Choo *et al*., 1996; Wentzensen *et al*., 2002; Thorland *et al*., 2003). As a consequence, the expression of cellular genes at or near the HPV integration sites could be influenced via insertional mutagenesis. In the context of cervical carcinogenesis, it may involve the activation of cellular proto-oncogenes or the inactivation of tumor suppressor genes. One well-documented example of insertional mutagenesis in cervical cancer is the down-regulation of the potential tumor suppressor gene *APM-1 (ZBTB7C)*, which is induced by DNA integration of the rare high-risk HPV type 68 (Reuter *et al*., 1998). Another remarkable example is the over-expression of *c-myc* proto-oncogene in several cervical carcinoma cell lines, which is associated with HPV DNA integration into the *c-myc* locus (Peter *et al*., 2006). It has been estimated that the *myc* locus is the most recurrent HPV integration site, which has been observed in about 10% of cervical carcinomas analysed until 2006 (Peter *et al*., 2006). Thus, HPV-induced insertional mutagenesis of particular cellular genes might enhance malignant progression in at least a subset of cervical cancers.

1.6 Methods for HPV DNA integration analysis

The determination of HPV integration sites can be achieved at the levels of chromosomes, genomic DNA and cDNA. At the chromosomal level, the integrated HPV DNA can be localized in specific chromosomal regions (cytogenetic bands) by *in situ* hybridization (Mincheva *et al*., 1987; Popescu *et al*., 1987; Couturier *et al*., 1991). However, limited by the low resolution of this assay, it is not possible to identify which gene is located directly at or near the integration site. Therefore, in order to analyse the genetic context of integrated HPV DNA, integration sites have to be characterized at the DNA or cDNA sequence level.

Since the integrated HPV DNA is co-transcribed together with downstream cellular DNA as a hybrid mRNA, the cellular sequences from this featured structure can be used to determine the HPV integration sites. Viral-cellular junction sequences were

obtained from cDNA libraries by cloning, hybridization and sequencing (Schneider-Gädicke and Schwarz, 1986). In later years, a method termed *amplification of papilloma virus oncogene transcripts* (APOT) has been applied to amplify the 3' viral-cellular junction sequences by RT-PCR (Klaes *et al*., 1999). However, the cDNA-based assays have the common inherent limitation that the exact 3' viral-cellular DNA junctions cannot be identified, because they are spliced out as intron during RNA processing (see *Figure 1-5*). Therefore, the most accurate determination of HPV integration sites should come directly from DNA sequence analysis. Earlier studies in this field were done with genomic DNA libraries, hybridization and sequencing (Baker *et al*., 1987). Later, several PCR-based methods for the identification of HPV integration sites have been published. *Detection of integrated papillomavirus sequences* (DIPS-PCR) is a ligation-mediated assay including restriction digestion, adapter ligation, linear amplification with HPV primer and PCR using nested HPV primer together with the adapter (Luft *et al*., 2001). Another ligation-mediated assay called *restriction enzyme cleavage, self-ligation and inverse PCR* (*rli*PCR) was also developed (Kalantari *et al*., 2001). Unlike DIPS-PCR and *rli*PCR, the restriction-site PCR (RS-PCR) is a semi-nested PCR assay, which does not require the digestion and ligation procedure (Sarkar *et al*., 1993; Thorland *et al*., 2000). Details about the RS-PCR will be described in *Section 2.1*.

Instead of determining HPV integration sites, an assay called E2-E6 quantitative PCR was developed to rapidly estimate the physical state of HPV16 DNA (episomal, integrated or a mixture of both) (Nagao *et al*., 2002; Peitsaro *et al*., 2002). As mentioned earlier, the percentages of integrated HPV DNA are positively correlated with pathologic grades of cervical lesions (Cullen *et al*., 1991; Daniel *et al*., 1995; Klaes *et al*., 1999; Tonon *et al*., 2001; Hopman *et al*., 2004; Hudelist *et al*., 2004). Accordingly, clinical pre-cancerous lesions often harbour a mixed form of episomal and integrated HPV DNA. Based on the assumption that ORF E2 is deleted or disrupted upon integration, the E2/E6 ratio obtained by quantitative PCR is thought to represent the percentage of integrated *versus* episomal HPV16 DNA (*Figure 1-6*), thereby implicating the physical state of HPV16 DNA in a given sample (Nagao *et al*., 2002; Peitsaro *et al*., 2002). This method has been used in large-scale investigations to assess HPV DNA integration as a progression marker for cervical cancer (Arias-Pulido *et al*., 2006; Briolat *et al*., 2007; Saunier *et al*., 2008). However, for accurate interpretation it is necessary to validate the HPV DNA integration status (after E2-E6 quantitative PCR) at the DNA sequence level.

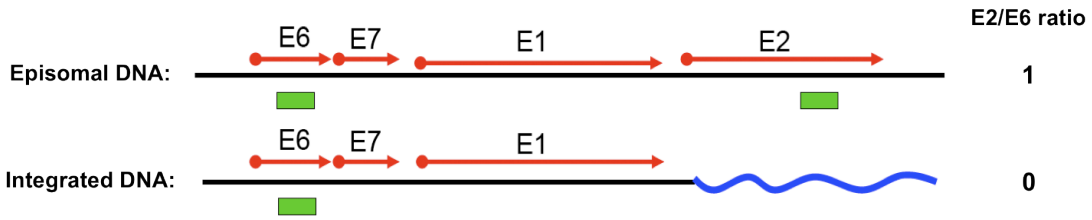


Figure 1-6: The principle of E2-E6 quantitative PCR. The structural difference between episomal and integrated HPV DNA is schematically represented, with the assumption that ORF E2 is deleted or disrupted in the integrated DNA. For better demonstration, the episomal HPV DNA is also shown in linear form. The blue line represents cellular DNA flanking the HPV integration site. Red arrows show the ORFs of E6, E7, E1 and E2. The E6 and E2 amplicons used in the quantitative PCR are marked in green. The E2/E6 ratio of 1 indicates purely episomal DNA, and the ratio of 0 indicates fully integrated DNA. E2/E6 ratios between 0 and 1 are interpreted as mixed HPV16 DNA. The figure was kindly provided by Elisabeth Schwarz (DKFZ, Heidelberg).

1.7 CGE-DKFZ co-operation project

The present study is part of a French-German co-operation project between Cancéropôle du Grand-Est (CGE) and DKFZ, aimed to improve the knowledge about HPV-induced cervical carcinogenesis. As a sub-project of this co-operation, our group in Heidelberg has been working on HPV16 DNA integration analysis in clinical cervical lesions together with two groups in France, one in Reims and the other in Besançon.

Clinical samples were collected from women who participate in routine cervical screening programs carried out in Reims and Besançon. The great benefit of cervical screening is to discover cervical lesions at an early stage before malignant conversion, through a cytological procedure called Pap smear test (Papanicolaou and Traut, 1941). Since 1990s, the conventional Pap smear test has been replaced by liquid-based monolayer cytology systems (Hutchinson *et al.*, 1991). The widely applied liquid-based ThinPrep Pap Test is also used in Reims and Besançon for primary screening (Clavel *et al.*, 2001; Bory *et al.*, 2002; Briolat *et al.*, 2007; Saunier *et al.*, 2008). With this method, samples are collected from the cervix using a cytobrush. After rinsing in the PreservCyt fixative, cells are fixated and preserved in the medium. For cytological test, an aliquot of liquid-fixed cells is spread on slide, stained and observed under a microscope, to detect cells with abnormal morphology (Hutchinson *et al.*, 1991).

Besides the liquid-based cytology, additional analyses are performed in Reims and Besançon. For samples containing abnormal cells, the same slide preparations are also used to detect DNA aneuploidy, one of the hallmarks of cancer cells, using DNA image cytometry (Lorenzato *et al*., 2001). In order to increase the screening accuracy, HPV DNA testing for high-risk types has been applied in conjunction with cytological test (Clavel *et al*., 2001). The HPV testing is mainly done with the Hybrid Capture II assay, which is based on the specific binding of a mixture of RNA probes with high-risk HPV DNA from cell lysates (Lörincz, 1996). Because Hybrid Capture II cannot distinguish between individual HPV types, HPV genotyping is further performed using the Linear Array Test from Roche (Briolat *et al*., 2007). For those samples, which are classified as high-grade lesions by cytology and contain at least one of the high-risk HPV types, DNA and RNA samples are isolated for detailed molecular analysis. For HPV16 integration analysis, HPV16-positive DNA samples from high-grade lesions are sent to our group at DKFZ. In parallel, the estimation of HPV16 physical states by E2-E6 quantitative PCR is performed in Reims and Besançon (Saunier *et al*., 2008).

1.8 Aims and outline of the thesis work

The intention of our co-operation project was to evaluate HPV DNA integration as predictor for high-grade cervical pre-cancerous lesions and as neoplastic progression marker for cervical cancer development. The primary goal of this thesis work was to determine HPV16 DNA integration sites at the sequence level (viral DNA breakpoints and human genome target sequences) in selected pre-cancerous and cancerous lesions obtained from cervical scrapes within routine screening programs. For this purpose, two prerequisites had to be fulfilled. Firstly, a whole genome amplification (WGA) method tolerant to template DNA degradation was required to amplify the tiny amounts of often partially degraded clinical DNA samples derived from cervical scrapes. Secondly, it was intended to establish an adequate procedure that can identify viral-cellular junction sequences against a background of episomal HPV16 DNA, especially for cervical pre-cancerous lesions.

In the present study, it was initially planned to combine the phi29-based multiple displacement amplification (MDA) for WGA (Dean *et al*., 2002) with restriction-site PCR (RS-PCR) for junction sequence determination. However, by applying these

methods to DNA from HPV16-positive cell lines and clinical samples, it became apparent that phi29-MDA failed to amplify degraded DNA frequently isolated from cervical scrapes and RS-PCR seemed to be inappropriate for clinical samples containing episomal HPV16 DNA. Therefore, during the thesis work a novel alternative strategy for HPV16 integration analysis was developed and tested with cervical carcinoma cell lines as well as clinical DNA samples from cervical pre-cancerous and cancerous lesions. The novel strategy, named ASP16, includes the robust PCR-based GenomePlex whole genome amplification (A), an HPV16 DNA enrichment-selection step (S) and high-throughput pyrosequencing (P) followed by bioinformatics analysis of the sequence data. Unique among methods for determination of HPV16 integration sites, ASP16 allows the simultaneous analysis of multiple DNA samples in one sequencing experiment.

Based on the sequence data of HPV16 DNA integration, insertional mutagenesis of candidate cellular genes at or near the integration sites was investigated in the context of cervical carcinogenesis. Using RS-PCR, the HPV16 integration site in the cell line MRI-H186 was determined closely upstream of the *c-myc* proto-oncogene. Thus, cell line MRI-H186 turned out to be a remarkable example of HPV-induced insertional mutagenesis. Consequently, over-expression and mutations of the *c-myc* proto-oncogene associated with HPV16 DNA integration into the *c-myc* locus in MRI-H186 were intensively further characterized in this work. For clinical samples analysed by ASP16, the cellular proto-oncogenes and tumor suppressor genes, whose expression might be directly affected by integrated HPV16 DNA, were predicted. In the context of the co-operation project, it is intended to interpret the HPV16 integration data in correlation with the molecular and clinical data assembled in Reims and Besançon to further validate HPV16 integration as marker for women with high-grade precursor lesions/cancer or at high risk for malignant progression and to further deepen the molecular understanding of cervical carcinogenesis.

2. Results

The experimental results of this thesis are divided into two parts. The first one describes the identification of HPV16 DNA integration sites with restriction-site PCR (RS-PCR) in the cervical carcinoma cell line MRI-H186, which turned out to be a remarkable example of insertional mutagenesis of the *c-myc* proto-oncogene. The second part is concerned with HPV16 DNA integration analysis in clinical DNA samples obtained from cervical scrapes, for which a novel strategy termed amplification-selection-pyrosequencing for HPV16 (ASP16) was developed.

2.1 HPV16 DNA integration analysis with restriction-site PCR in cervical carcinoma cell lines

At the beginning, it was planned to determine HPV16 DNA integration sites in clinical samples using restriction-site PCR (RS-PCR). As illustrated in *Figure 2-1*, this method includes two rounds of PCR (Sarkar *et al*., 1993; Thorland *et al*., 2000). For the first round, an HPV16 L1/L2 or E1/E2 primer is used in combination with eight restriction-site oligonucleotides (RSOs) in different reactions to amplify the 5' or 3' viral-cellular junction. The RSO primer contains the recognition sequence of an HPV16 non-cutting restriction enzyme (*Figure 2-1* and *Section 4.12.3*). The second round is a semi-nested PCR to increase amplification specificity, for which a nested HPV16 primer works together with the same RSOs. Before dealing with clinical samples, genomic DNA from cell lines was used as template to establish the RS-PCR. This involved the HPV16-positive cervical carcinoma cell lines SiHa, CaSki, MRI-H186 and MRI-H196. Among them, the chromosomal integration site of HPV16 DNA in SiHa had been reported previously, and both the 5' and 3' viral-cellular junction sequences were determined (Baker *et al*., 1987; Meissner, 1999). For the other three cell lines, HPV16 DNA integration sites had not been characterized before at the DNA sequence level.

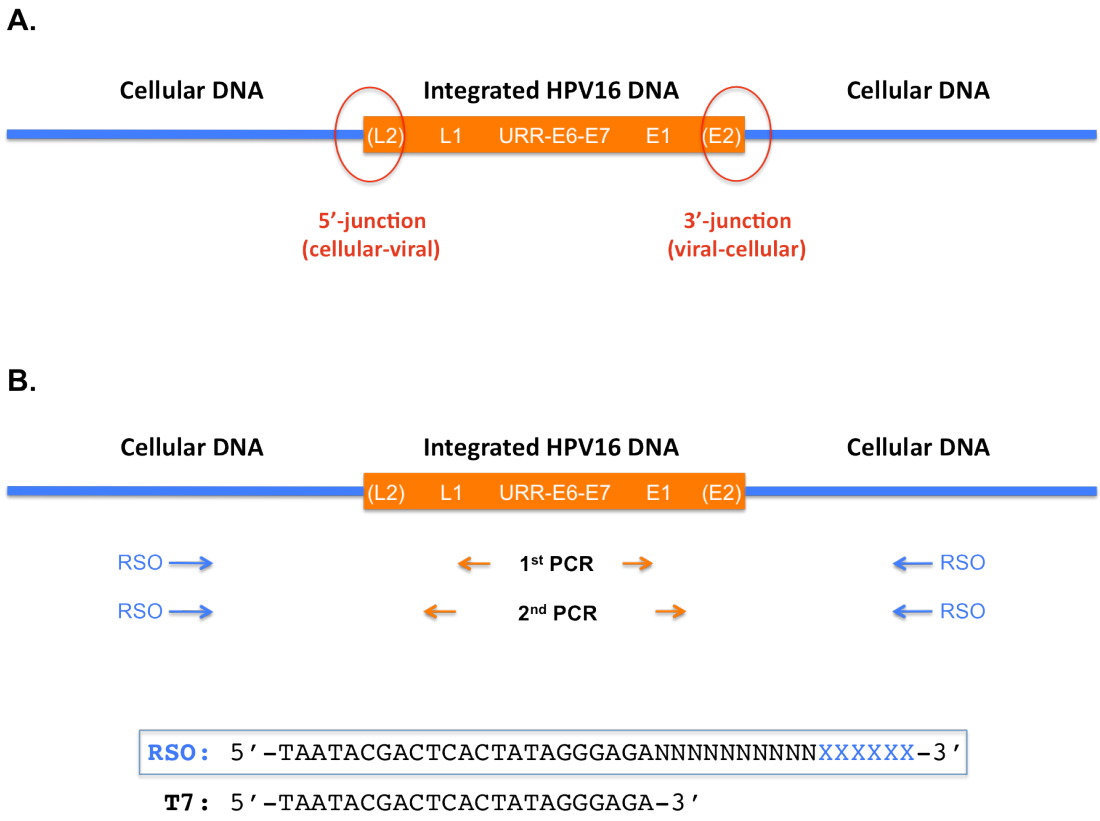


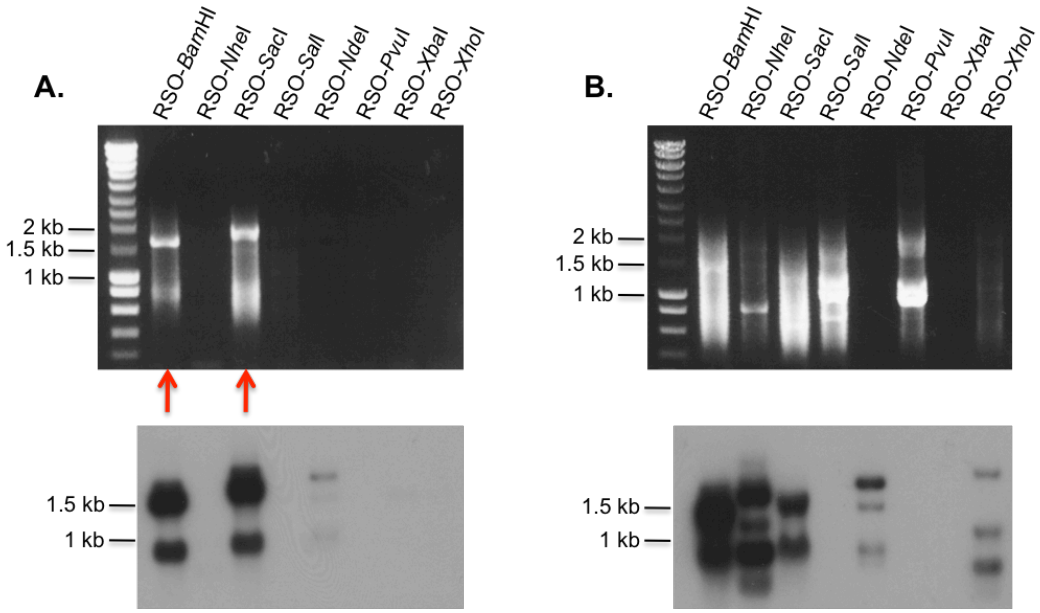
Figure 2-1: Schematic demonstration of viral-cellular junctions and restriction-site PCR (RS-PCR). Definition of the 5' and 3' viral-cellular junctions is shown in panel **A**. The procedure of RS-PCR is outlined in panel **B**. The RSO primer contains the T7-adapter sequence, a random sequence of ten nucleotides (N_{10}), and the recognition sequence (Xs in blue) of an HPV16 non-cutting restriction enzyme.

2.1.1 Identification of the 3' viral-cellular junction sequence in SiHa

Previous studies had shown that only one integration site of HPV16 DNA is present in SiHa cells, which is located in chromosomal region *13q22.1* (Baker *et al.*, 1987; Meissner, 1999). The breakpoints of integrated HPV16 DNA are located at position 3133 at the 3' terminus and position 3385 at the 5' terminus (Baker *et al.*, 1987). For establishment of the RS-PCR, it was tested whether the 3' viral-cellular junction sequence of SiHa can be re-identified. The first-round PCR was performed using primer HPV16-1546-26D in combination with eight individual RSO-primers in separate reactions, respectively (Thorland *et al.*, 2000). For the second-round PCR, the nested primer HPV16-1588-26D was used to increase PCR specificity of the

HPV16 part (Thorland *et al*., 2000). In parallel, applicability of the unique T7-adapter sequence instead of the full-length RSO-primers in the 2nd PCR was also examined. The 2nd-PCR products were checked by gel electrophoresis, and then hybridized with the HPV16-complete DNA probe (*Figure 2-2A and B*). The final PCR products of two positive reactions, RSO-*Bam*HI and RSO-*Sac*I (red arrows in *Figure 2-2A*), were cloned. After the selection by Southern hybridization, nine clones positive for HPV16 were sequenced completely. As a result, all clones contained the same 3' viral-cellular junction sequence of SiHa DNA (*Figure 2-2C and D*), which is identical with the published data (Baker *et al*., 1987; Meissner, 1999). The consensus sequence derived from the nine clones is shown in *Appendix 1*.

In conclusion, the 3' viral-cellular junction sequence of SiHa DNA (*Figure 2-2D*) could be re-identified with RS-PCR. The cellular parts of the RS-PCR amplicons (*Figure 2-2C*) were overall long enough for database searching to localize the junction sequences on human chromosomes. Concerning the specificity of priming in the 2nd PCR, the T7-adapter (*Figure 2-2A*) exhibited more reliability than the full-length RSO-primers (*Figure 2-2B*), and therefore was used in successive experiments.



C.

Clone Number	RSO Primer	Sequence Length	HPV16 Part (start – end)	Length of HPV16 Part	Cellular Part (start – end)	Length of Cellular Part
1	*Bam*HI	1592 bp	1588 – 3133	1546 bp	74087558 – 74087513	46 bp
2	*Bam*HI	1587 bp	1588 – 3133	1546 bp	74087558 – 74087518	41 bp
3	*Bam*HI	1630 bp	1588 – 3133	1546 bp	74087558 – 74087475	84 bp
4	*Bam*HI	1587 bp	1588 – 3133	1546 bp	74087558 – 74087518	41 bp
5	*Sac*I	1797 bp	1588 – 3133	1546 bp	74087558 – 74087308	251 bp
6	*Sac*I	1949 bp	1588 – 3133	1546 bp	74087558 – 74087156	403 bp
7	*Sac*I	1797 bp	1588 – 3133	1546 bp	74087558 – 74087308	251 bp
8	*Sac*I	1797 bp	1588 – 3133	1546 bp	74087558 – 74087308	251 bp
9	*Sac*I	1949 bp	1588 – 3133	1546 bp	74087558 – 74087156	403 bp

D.

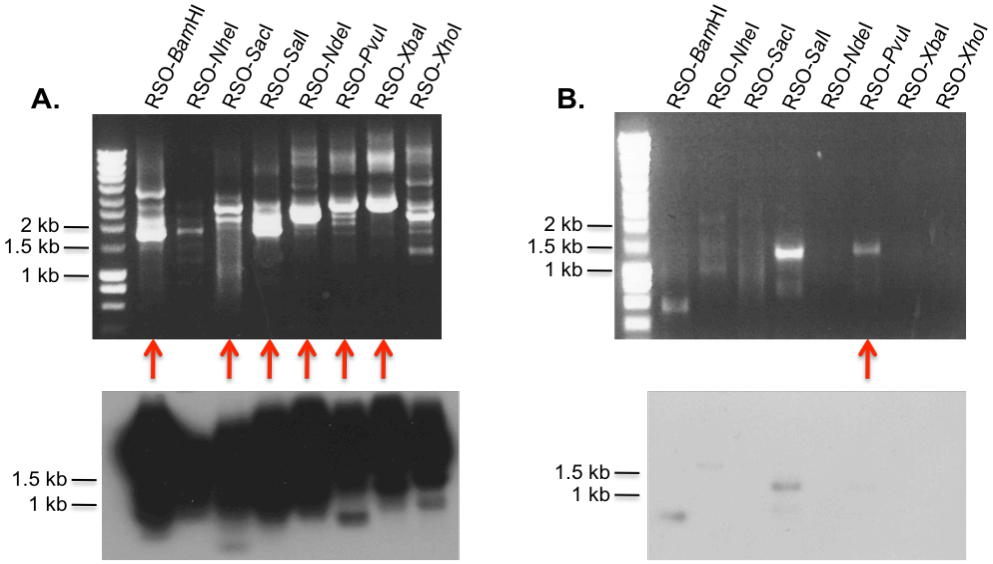
Figure 2-2: Identification of the 3' viral-cellular junction sequence in SiHa by RS-PCR. RS-PCR was performed using either T7-primer (panel **A**) or RSO (panel **B**) in combination with HPV16-1588-26D for the 2nd-round PCR reaction. In both cases, gel analysis (top) of the 2nd-PCR products and Southern hybridization (bottom) with the HPV16-complete DNA probe are shown. The SmartLadder was used as size marker. Nine clones (panel **C**) were obtained from the 2nd-PCR products of RSO-*Bam*HI and RSO-*Sac*I (red arrows in panel **A**) and were sequenced. The structure of the viral-cellular junction identified by RS-PCR is schematically demonstrated in panel **D**. The dashed arrows show the orientation of the viral and cellular DNA involved.

2.1.2 Identification of HPV16 DNA integration sites in cell lines MRI-H186, MRI-H196 and CaSki

MRI-H186:

It was shown in previous experiments that the HPV16 DNA is integrated in chromosomal region *8q24* near the *c-myc* proto-oncogene in MRI-H186 (Wanschura, 1992). However, the exact junction sequences were unknown. For determining the 3' viral-cellular junction sequence of MRI-H186 DNA with RS-PCR, the HPV16 primers used for the two-round reaction were exactly the same as in SiHa (see *Section 2.1.1*). To obtain the 5' junction sequence, primers HPV16-6688-25U and HPV16-5885-27U were used for the 1st and 2nd PCR, respectively (Thorland *et al*., 2000). As a result, both 3' and 5' viral-cellular junction sequences could be identified in MRI-H186 (*Figure 2-3*). Altogether 15 clones (*Figure 2-3C*) containing the 3' or 5' viral-cellular junction sequence of MRI-H186 were obtained from various 2nd-PCR products (red arrows in *Figure 2-3A and B*). The consensus sequences for both junctions derived from the 15 clones are shown in *Appendix 1*.

As shown in *Figure 2-3D*, the breakpoints of HPV16 DNA are located at position 2754 at the 3' terminus and position 4871 at the 5' terminus. The HPV16-E2 gene (position 2756 to 3853) is completely deleted in this DNA architecture. By database searching, the HPV16 integration site of MRI-H186 DNA was mapped to chromosomal band *8q24.21*. A small deletion of 8 bp was observed in the flanking cellular sequences around the integration site. Since the *c-myc* gene starts at position 128817498 of chromosome 8 (NC_000008.9), it became clear that this human proto-oncogene is located directly downstream and in close proximity to the integrated HPV16 DNA (see more details in *Section 2.2*).



C.

Clone Number	3' or 5' Junction	RSO Primer	Sequence Length	HPV16 Part (start – end)	Length of HPV16 Part	Cellular Part (start – end)	Length of Cellular Part
1	3'	*Bam*HI	1760 bp	1588 – 2754	1167 bp	128815788 – 128816380	593 bp
2	3'	*Sac*I	2700 bp	1588 – 2754	1167 bp	128815788 – 128817320	1533 bp
3	3'	*Sal*I	1651 bp	1588 – 2754	1167 bp	128815788 – 128816271	484 bp
4	3'	*Sal*I	2298 bp	1588 – 2754	1167 bp	128815788 – 128816918	1131 bp
5	3'	*Nde*I	2331 bp	1588 – 2754	1167 bp	128815788 – 128816951	1164 bp
6	3'	*Nde*I	1399 bp	1588 – 2754	1167 bp	128815788 – 128816019	232 bp
7	3'	*Nde*I	1399 bp	1588 – 2754	1167 bp	128815788 – 128816019	232 bp
8	3'	*Nde*I	1399 bp	1588 – 2754	1167 bp	128815788 – 128816019	232 bp
9	3'	*Pvu*I	3304 bp	1588 – 2754	1167 bp	128815788 – 128817924	2137 bp
10	3'	*Xba*I	1850 bp	1588 – 2754	1167 bp	128815788 – 128816470	683 bp
11	3'	*Xba*I	1850 bp	1588 – 2754	1167 bp	128815788 – 128816470	683 bp
12	3'	*Xba*I	1850 bp	1588 – 2754	1167 bp	128815788 – 128816470	683 bp
13	3'	*Xba*I	1850 bp	1588 – 2754	1167 bp	128815788 – 128816470	683 bp
14	5'	*Pvu*I	2006 bp	5885 – 4871	1015 bp	128815779 – 128814789	991 bp
15	5'	*Pvu*I	2006 bp	5885 – 4871	1015 bp	128815779 – 128814789	991 bp

D.

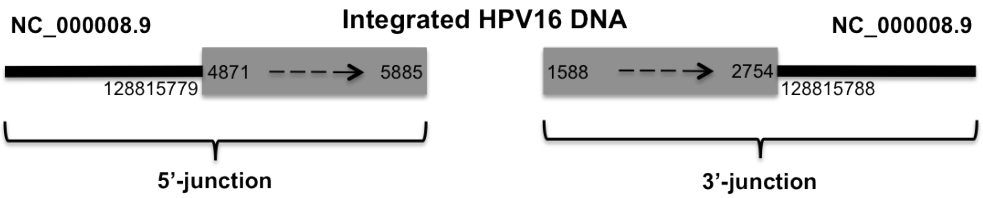


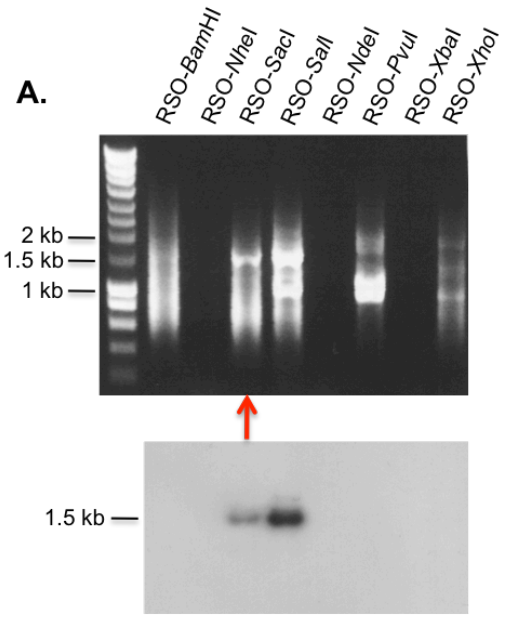
Figure 2-3: Identification of the 3' and 5' viral-cellular junction sequences in MRI-H186 by RS-PCR. The 2nd-PCR products and Southern hybridization with the HPV16-complete probe are shown for the 3'-junction (panel **A**) and 5'-junction RS-PCR (panel **B**). The SmartLadder was used as size marker. Fifteen clones (panel **C**) obtained from the final products of seven PCR reactions (red arrows in panel **A** and **B**) were sequenced. Panel **D** shows the structures of the 5' and 3' junctions based on the sequence data.

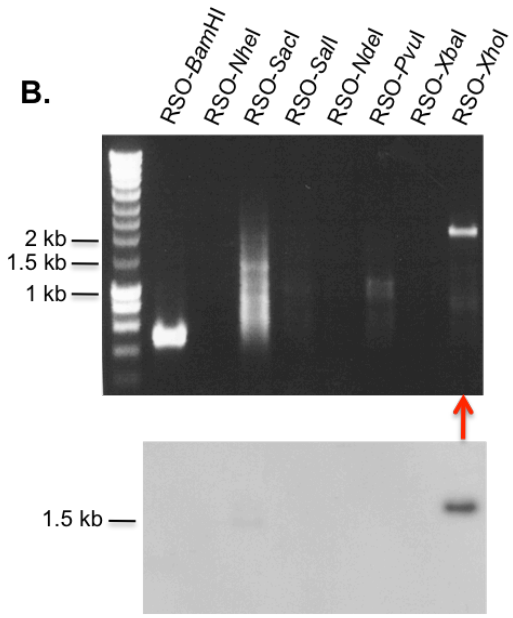
MRI-H196:

For MRI-H196, RS-PCR for the 3' viral-cellular junction was carried out using primers HPV16-1546-26D and HPV16-2387-25D sequentially. RS-PCR for the 5' junction was primed by HPV16-6688-25U and HPV16-6653-25U (Thorland *et al*., 2000). Altogether four clones (*Figure 2-4C*) from two 2nd-PCR products (red arrows in *Figure 2-4A and B*) were sequenced. The cloned 3'-junction was sequenced completely, and the breakpoint of HPV16 DNA is located at position 3858 (*Figure 2-4D*). This leaves the HPV16-E2 gene (position 2756 to 3853) intact. A small DNA insertion "TTAC" was found at the junction site between HPV16 (position 3858) and cellular DNA (position 47967861). The consensus sequence for the 3' junction of MRI-H196 is shown in *Appendix 1*. By contrast, the cloned 5'-junction could not be sequenced in full length for unknown reasons. Therefore, it was not possible to determine the exact breakpoints of HPV16 and the cellular DNA for this junction. The gap is estimated to be about 400 bp (*Figure 2-4D*). The partial sequences of the 5' junction are shown in *Appendix 1*. The HPV16 integration site of MRI-H196 DNA was mapped to chromosomal band *11p11.2*. By comparing the cellular sequence positions at the 5' and 3' junctions, a DNA deletion of about 1 Mb in the cellular target sequence was detected.

CaSki:

RS-PCR failed to determine any viral-cellular junction sequence of CaSki DNA. Only pure HPV16 amplicons from the E1/E2 region were obtained, which were caused by false priming of RSO primers onto the HPV16 DNA (data not shown). Every CaSki cell contains approximately 500 copies of integrated HPV16 DNA, and most of them are present as concatemers of the complete HPV16 genome (Yee *et al*., 1985; Baker *et al*., 1987). This feature is most likely the reason for the false priming as well as the failure to amplify and detect the much less-abundant viral-cellular junctions.





C.

Clone Number	3' or 5' Junction	RSO Primer	Sequence Length	HPV16 Part (start – end)	Length of HPV16 Part	Cellular Part (start – end)	Length of Cellular Part
1	3'	*SacI*	1639 bp	2387 – 3858	1472 bp	47967861 – 47968023	163 bp
2	3'	*SacI*	1639 bp	2387 – 3858	1472 bp	47967861 – 47968023	163 bp
3	3'	*SacI*	1639 bp	2387 – 3858	1472 bp	47967861 – 47968023	163 bp
4	5'	*XhoI*	1842 bp	6653 – 5321	1333 bp	46953669 – 46953161	509 bp

D.

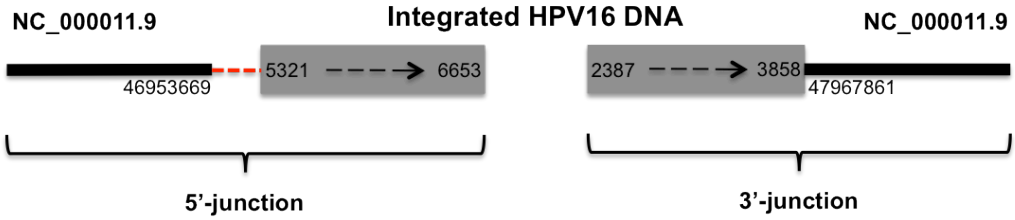


Figure 2-4: Identification of the 3' and 5' viral-cellular junction sequences in MRI-H196 by RS-PCR. The 2nd-PCR products and Southern hybridization with the HPV16-complete probe are shown for the 3'-junction (panel **A**) and 5'-junction RS-PCR (panel **B**). The SmartLadder was used as size marker. Four clones (panel **C**) obtained from the final products of two PCR reactions (red arrows in panel **A** and **B**) were sequenced. Panel **D** shows the structures of the 5' and 3' junctions based on the sequence data. The precise viral-cellular DNA boundary of the 5' junction is not known. The sequence gap (dashed red line) has a size of about 400 bp.

2.2 Genetic context of integrated HPV16 DNA in MRI-H186

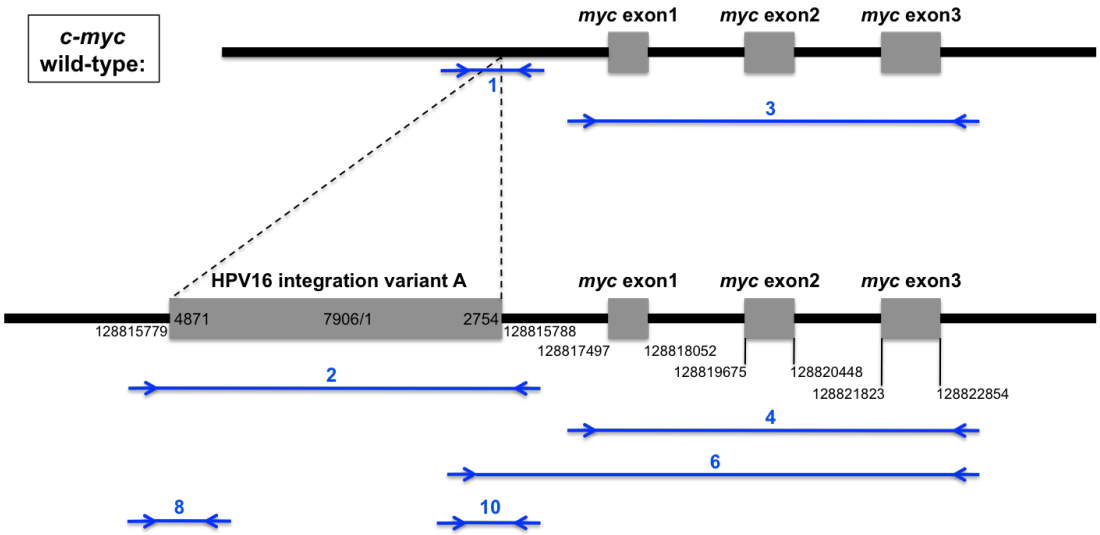
Because the integrated HPV16 DNA is located closely upstream of the *c-myc* gene in MRI-H186 cells, it was decided to clarify whether and to which extent this important cellular proto-oncogene is affected by HPV16 DNA integration.

2.2.1 The viral-cellular DNA organization

The human *c-myc* gene is mapped to chromosomal band *8q24.21*, and the three exons of *c-myc* expand from position 128817498 to 128822853 (NC_000008.9) of the forward DNA strand. As displayed in *Figure 2-5A*, the HPV16 DNA in MRI-H186 is integrated upstream and in sense orientation to the *c-myc* gene. The distance between the HPV16 DNA breakpoint at the 3' junction and the transcription initiation site of *c-myc* is only 1709 bp.

The integrated HPV16 DNA as determined by RS-PCR in MRI-H186 (see *Section 2.1.2*) is designated as HPV16 integration variant A (*Figure 2-5A*). To verify the viral-cellular DNA organization of HPV16 integration variant A, a series of long-PCR reactions was performed using different combinations of HPV16 and *c-myc* primers (table in *Figure 2-5B*). SiHa DNA was used as a control for the uninterrupted *c-myc* allele. As shown by the agarose gel in *Figure 2-5B*, all PCR products agreed with the expected sizes. Both the 5' and 3' junctions were confirmed by junction-specific PCR (reactions 8 and 10). Combining *c-myc* primers 128815697-F and 128816053-R, which are located directly upstream (5') and downstream (3') of integration variant A, a PCR product of the expected size (6.1 kb) was obtained (reaction 2). In addition, a small PCR product representing the *c-myc* wild-type allele was amplified not only from SiHa DNA as expected (reaction 1), but also in reaction 2 with MRI-H186 DNA as template. This result implied that either a normal chromosome 8 exists in MRI-H186 cells or that the mutated chromosome 8 with integrated HPV16 DNA also harbors an uninterrupted *c-myc* allele.

A.



B.

PCR	Primer Pair	Template	Expected Size	Observed Size
1	128815697-F 128816053-R	SiHa	356 bp	~400 bp (a.e.)
2		MRI-H186	6.1 kb	~6 kb (a.e.); ~400 bp
3	128817197-F 128822955-R	SiHa	5.8 kb	~6 kb (a.e.)
4		MRI-H186	5.8 kb	~6 kb (a.e.)
5	HPV16-2170-F 128822955-R	SiHa	negative	negative (a.e.)
6		MRI-H186	7.8 kb	~8 kb (a.e.)
7	128815697-F HPV16-5396-R	SiHa	negative	negative (a.e.)
8		MRI-H186	608 bp	~600 bp (a.e.)
9	HPV16-2573-F 128816092-R	SiHa	negative	negative (a.e.)
10		MRI-H186	486 bp	~500 bp (a.e.)

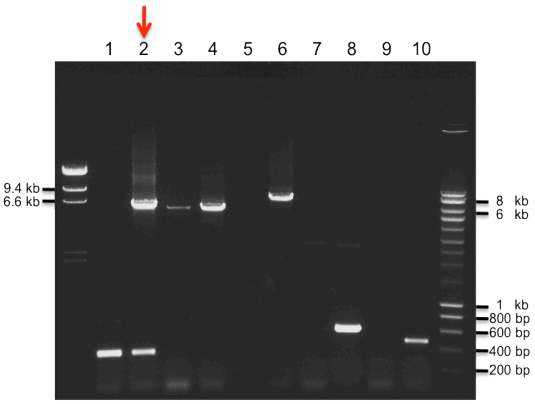
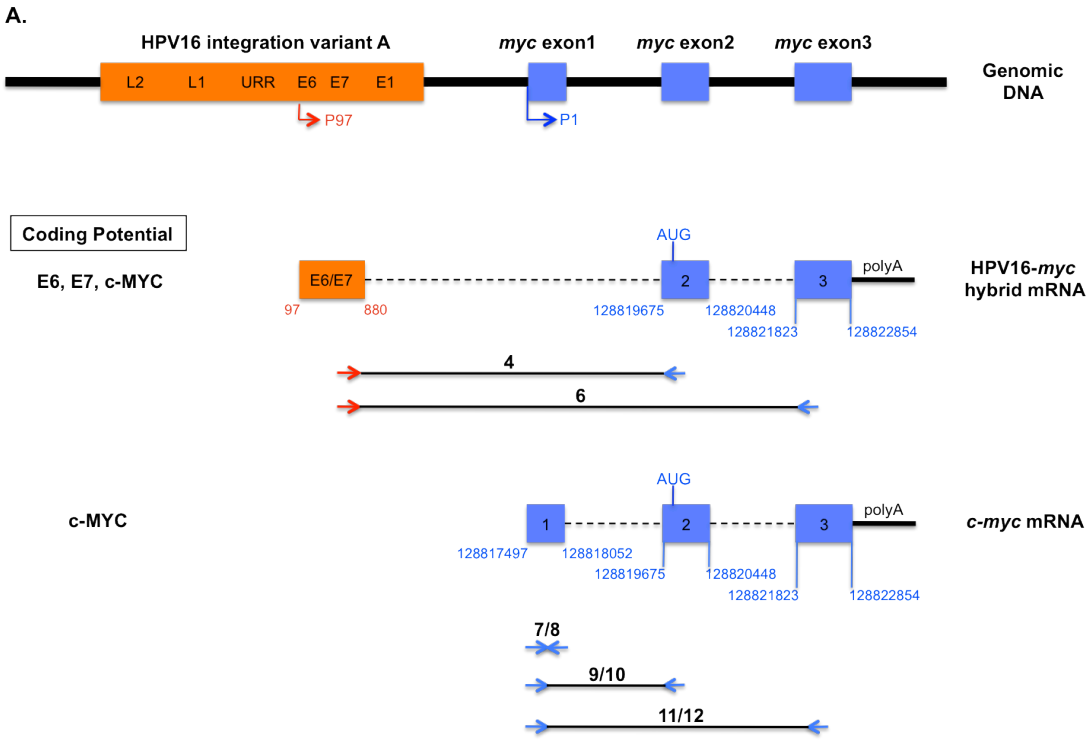


Figure 2-5: Long-PCR analysis to examine the genetic context of HPV16 integration variant A in MRI-H186. The junction sequences obtained from RS-PCR were combined to predict the genetic context (panel **A**) of HPV16 integration variant A and flanking cellular sequences. The genomic positions at the HPV16 integration junctions and of the three *c-myc* exons are given (NC_000008.9). The *c-myc* wild-type allele without integrated HPV16 DNA is shown above. The blue lines with the flanking arrows indicate the positions of the PCR products and their primers outlined in the table of panel **B**. The numbers correspond to the PCR numbers in the table. Five primer combinations (table in panel **B**) were used in long-PCR to test the predicted DNA structure in MRI-H186 (SiHa as control). PCR products are shown on the gel. The numbers above the lanes correspond to the PCR numbers in the table. The red arrow indicates the 6.1-kb band containing the HPV16 integration variant A. The λ/*Hin*d III (leftmost) and SmartLadder (rightmost) were used as size marker. (a.e. = as expected)

2.2.2 HPV16/*c-myc* transcripts

To investigate how HPV16 integration variant A might affect *c-myc* gene expression in MRI-H186, putative transcripts were predicted based on the architecture of integrated HPV16 DNA and *c-myc* gene (*Figure 2-6A*). The URR-E6-E7 region of HPV16 is entirely retained in MRI-H186 and therefore ensures expression of the viral oncogenes E6 and E7. The downstream-located *c-myc* transcription unit is fully intact. For that reason, transcription can be initiated at the HPV16 early promoter p97, and then continues through the 3' viral-cellular junction and the downstream *c-myc* gene, until it is terminated at the *c-myc* poly-A signal. RNA splicing will generate a viral-cellular hybrid transcript (HPV16-*myc* mRNA), which contains HPV16-E6 and E7 as well as exons 2 and 3 of *c-myc*. In addition, because the *c-myc* promoter P1 is maintained, the normal *c-myc* mRNA is also expected in MRI-H186. RT-PCR was applied to examine whether the two predicted transcripts are produced in MRI-H186 cells. In *Figure 2-6B*, the primer pairs and products are summarized in the table, and the experimental results are shown on the gel photo. With the exception of PCR numbers 7 and 8, all primers used are located in different exons, so that cDNA templates can be distinguished from DNA contamination. Not intended to amplify E6 splicing variants (see *Figure 1-2*), only primers in E7 (HPV16-811-F and HPV16-856-F) were selected for the HPV16 part. As a result, the lengths of cDNA amplicons checked on gel consisted with the expected sizes and thereby proved the existence of the two transcripts.

A.



B.

RT-PCR	Primer Pair	Template	Expected Size	Observed Size
1	HPV16-811-F 128817900-R	SiHa	negative	negative (a.e.)
2		MRI-H186	negative	negative (a.e.)
3	HPV16-811-F 128819716-R	SiHa	negative	negative (a.e.)
4		MRI-H186	112 bp	~100 bp (a.e.)
5	HPV16-856-F 128821873-R	SiHa	negative	negative (a.e.)
6		MRI-H186	850 bp	~850 bp (a.e.)
7	128817576-F 128817900-R	SiHa	325 bp	~300 bp (a.e.)
8		MRI-H186	325 bp	~300 bp (a.e.)
9	128817576-F 128819716-R	SiHa	519 bp	~500 bp (a.e.)
10		MRI-H186	519 bp	~500 bp (a.e.)
11	128817576-F 128821873-R	SiHa	1302 bp	~1300 bp (a.e.)
12		MRI-H186	1302 bp	~1300 bp (a.e.)

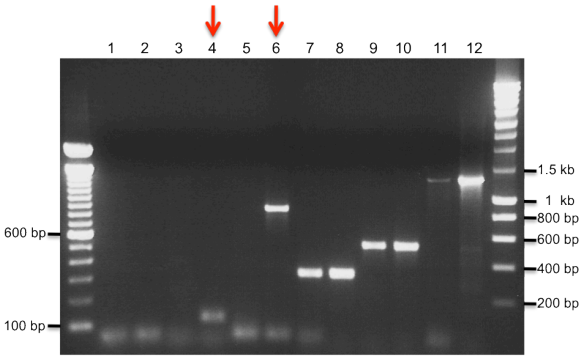


Figure 2-6: Prediction and RT-PCR analysis of HPV16/*c-myc* transcripts in MRI-H186. According to the viral-cellular DNA organization (top), two transcripts with different coding potentials were predicted (panel **A**). The dashed lines indicate the introns removed by splicing. In panel **A**, the numbered lines below the DNA and RNA structures indicate the primer pairs and RT-PCR products. The existence of the transcripts was checked by RT-PCR with six different primer pairs in altogether twelve reactions (table in panel **B**). Negative results were expected from reactions 1, 3 and 5 (because no product could be generated in SiHa using an HPV16 primer together with a *c-myc* primer) as well as from reaction 2 (because the 128817900-R in *c-myc* exon 1 does not bind to the HPV16-*myc* mRNA sequence). The RT-PCR products were checked on agarose gel. The reactions 4 and 6 (red arrows on gel photo) clearly showed the existence of the HPV16-*myc* hybrid mRNA. The 100 bp DNA Ladder (leftmost) and SmartLadder (rightmost) were used as size marker. (a.e. = as expected)

2.3 Expression level of *c-myc* in MRI-H186

2.3.1 *c-myc* Northern blot analysis

To detect *myc*-containing transcripts, Northern blot analysis was carried out using all three *c-myc* exons as DNA probes. Among them, only the *c-myc* exon 1 probe can distinguish between the HPV16-*myc* and the normal *c-myc* mRNA. For interpretation of *c-myc* expression level, poly-A+ mRNA from HeLa and CaSki cells was used for comparison with MRI-H186. It had been reported previously that *c-myc* gene is over-expressed in HeLa and CaSki cells (Dürst *et al*., 1987). Interestingly, the Northern blot analysis revealed that *c-myc* expression is strongly elevated in MRI-H186, even if compared with HeLa and CaSki (*Figure 2-7*). The normal *c-myc* mRNA (exons 1 to 3) has a length of 2362 nt. The HPV16-*myc* mRNA has a size of 2590 nt if E6 is present, or of 2408 nt if the spliced E6* is generated. Thus, the HPV16-*myc* and *c-myc* transcripts cannot be distinguished by size on Northern blot. In order to estimate the relative amount of the two transcripts in MRI-H186 cells, signal intensities produced with the three *c-myc* exon probes were compared to each other. The signal intensity produced by the exon 1 probe is very similar to the intensities from exon 2 and exon 3 probes. Since exon 1 is only present in the normal *c-myc* mRNA, but not in the HPV16-*myc* mRNA, it can be concluded that the normal *c-myc* mRNA is the predominant *myc*-containing transcript in MRI-H186.

Strikingly, an extra RNA band was observed in MRI-H186, which hybridized with *c-myc* exon 1 and exon 2, but not with exon 3 (red arrows in *Figure 2-7*). This RNA is shorter than the normal *c-myc* mRNA as well as the HPV16-*myc* hybrid mRNA, and is expressed in a notable amount. Identification of the structure and origin of this RNA will be explained in *Section 2.4*, together with the Northern blot analysis using HPV16-specific DNA probes.

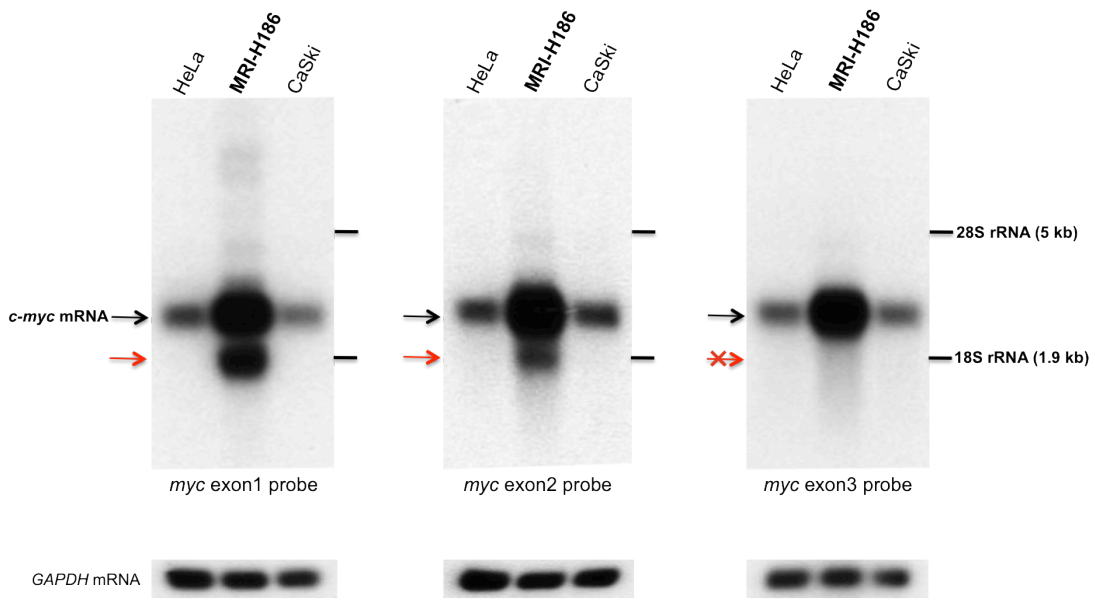


Figure 2-7: *c-myc* transcript level in MRI-H186. Poly-A+ mRNA was isolated from total RNA of HeLa, MRI-H186 and CaSki. Three Northern blots were prepared in parallel using equal amounts of the isolated mRNA, and hybridized with the *c-myc* exon 1, 2 and 3 DNA probes, respectively. The blots were post-hybridized with the *GAPDH* probe as sample-loading control. The red arrows show an unexpected band, which is only positive for *c-myc* exons 1 and 2, but not for exon 3.

2.3.2 c-MYC Western blot analysis

The Northern blot analysis showed that in MRI-H186 the normal *c-myc* mRNA is produced in large amount, compared to HeLa and CaSki. In the next step, the protein expression level was examined by Western blot analysis.

Translated from the AUG start-codon at the beginning of *c-myc* exon 2, the c-MYC oncoprotein, also named c-MYC-2 or p64, is composed of 439 amino acids. In addition, an alternative start codon CUG at the end of exon 1 is also active and gives rise to a c-MYC isoform called c-MYC-1 or p67, which is longer than c-MYC-2 with 15 additional residues at the N-terminus. This unusual CUG start-codon is not as efficient as the classic AUG, and therefore the c-MYC-1 protein is normally translated at a negligible scale.

With an anti-c-MYC antibody (clone 9E10) directed against an epitope at the C-terminus (see *Section 4.6*), c-MYC protein expression levels were analysed in MRI-H186 and other cell lines. Consistent with the up-regulation at RNA level, *c-myc* is also over-expressed at protein level in MRI-H186, compared to the non-cancerous cell line HaCaT as well as the cervical carcinoma cell lines C4-I, C4-II and MRI-H196 (*Figure 2-8*). Elevated c-MYC protein levels were also observed in HeLa cells as

expected, and in the HPV-negative cervical carcinoma cell line C33A. Unlike the great difference on Northern blots (see *Figure 2-7*), the c-MYC protein in MRI-H186 is expressed in similar amount as in HeLa.

In conclusion, the human *c-myc* gene is over-expressed at both RNA and protein levels in MRI-H186, associated with HPV16 DNA integration into this locus.

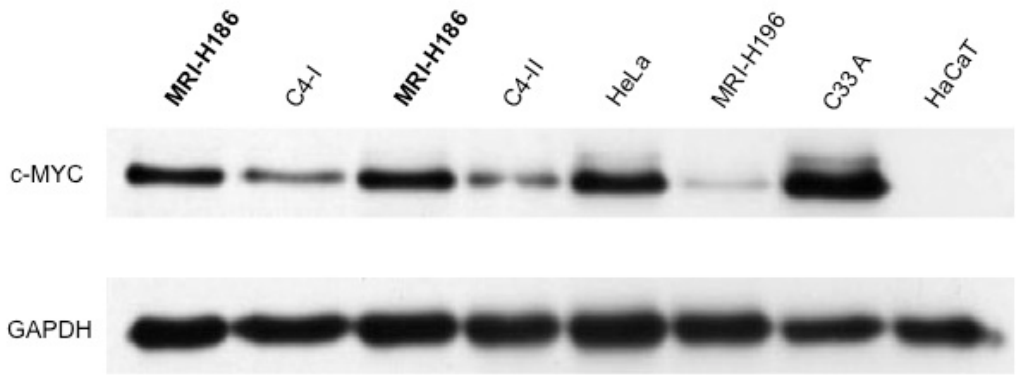


Figure 2-8: Expression level of c-MYC protein in MRI-H186. The c-MYC protein levels of various cancerous and non-cancerous cell lines were assessed by Western blot using antibody 9E10 directed against the C-terminus of human c-MYC protein. MRI-H186 cell lysates were prepared from two individual cell cultures to check reproducibility. GAPDH protein was used as an internal sample-loading control.

2.4 HPV16 transcript pattern in MRI-H186

Due to similar sizes the expected HPV16-*myc* hybrid mRNA (see *Figure 2-6*) seems to co-migrate with the normal *c-myc* mRNA on Northern blot, as deduced from the *c-myc* Northern blot shown in *Figure 2-7*. To directly visualize the HPV16-*myc* mRNA, Northern blot experiments were performed using HPV16-specific probes. Three DNA probes were used for detection, covering the complete HPV16 genome, the E6-E7 region and the E2 region, respectively. Besides MRI-H186, the HPV16-positive cell lines SiHa and CaSki were included in the analysis. In addition, it had to be clarified whether the extra-RNA hybridizing with *c-myc* exons 1 and 2 (see *Figure 2-7*) contains HPV16 sequences. The Northern results are shown in *Figure 2-9*.

The HPV16 transcript patterns are strongly different among the three cell lines. CaSki shows multiple robust bands on Northern blots, whereas the HPV16 transcripts of SiHa cells are hardly visible. For MRI-H186, hybridization with the HPV16-complete probe revealed three major mRNAs, which are indicated by the green, blue and red arrows in *Figure 2-9*. The mRNA indicated by the blue arrows seems to be the HPV16-*myc* hybrid mRNA, because it hybridized with the E6-E7

probe but not with E2. The mRNA with the strongest hybridization signal (red arrows) has a size very similar to the extra-RNA observed on the *c-myc* Northern blots. The mRNA hybridized with HPV16-complete and E2, but not with E6-E7. These results strongly suggested that a transcript containing sequences from the *c-myc* exon 1, exon 2 and HPV16-E2 could be existent in MRI-H186. The RNA band marked by the green arrows on the Northern blots seems to hybridize with HPV16-complete and E2, but not with E6-E7. The structure of this transcript has not been unravelled.

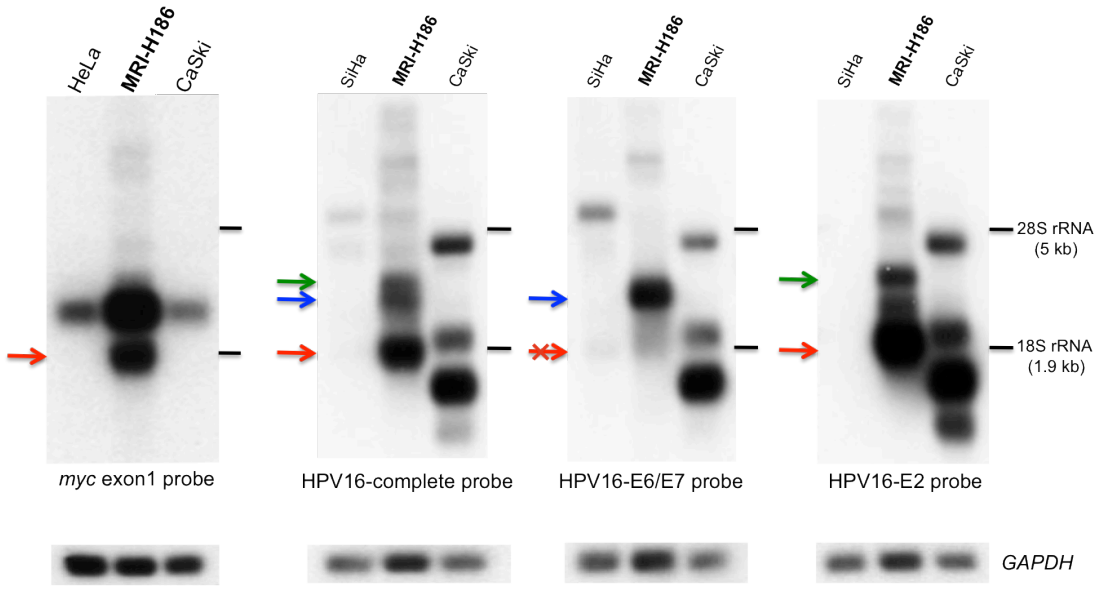


Figure 2-9: HPV16 transcript pattern in MRI-H186. Three Northern blots were prepared in parallel using poly-A+ mRNA. The blots were hybridized with three HPV16 DNA probes corresponding to the indicated different regions in the HPV16 genome, respectively. The Northern blot hybridized with the *c-myc* exon 1 probe was taken from *Figure 2-7* for comparison with the HPV16 Northern blots. The red arrows indicate the mRNA hybridizing with *c-myc* exons 1 and 2 as well as HPV16-E2. The blue arrows implicate the HPV16-*myc* hybrid mRNA. The green arrows indicate another mRNA hybridizing with HPV16-E2, but without further sequence information. The 28S and 18S rRNA bands were used as internal size marker. The blots were post-hybridized with the *GAPDH* probe as sample loading control.

2.5 Identification of *myc*-HPV16 hybrid mRNA in MRI-H186

In the *c-myc* and HPV16 Northern blot analysis, a novel transcript in MRI-H186 became apparent, which is composed of *c-myc* exons 1 and 2 as well as HPV16-E2 sequences (see *Sections 2.3.1* and *2.4*). Rapid amplification of cDNA ends (RACE) analysis was carried out in order to clarify the precise structure of this transcript. Details about the RACE procedure are explained in *Materials and Methods* (see *Section 4.27*). After the synthesis of 5' or 3'-RACE-ready cDNA, eight PCR reactions for the 5' and 3' RACE of MRI-H186 were performed using the universal adapter UPM in combination with eight different gene-specific primers, respectively. These primers have forward orientation for 3'-RACE or reverse orientation for 5'-RACE, and

are specific for HPV16-E7 or E2 and *c-myc* exon 1 or exon 2 (*Figure 2-10A*). Final products of all eight reactions (*Figure 2-10B*) were cloned and sequenced. Altogether, 32 sequences were obtained. After sequence comparison with the HPV16 genome as well as the *c-myc* cDNA reference (Accession number: NM_002467), 13 non-redundant sequences were obtained (*Figure 2-10C*).

With these experiments the novel transcript hybridizing with HPV16-E2 and *c-myc* probes could be identified. It is a *myc*-HPV16 hybrid transcript composed of *c-myc* exon 1, a truncated variant of exon 2, and the 3'-terminal part of HPV16-E2 (*Figure 2-10D*). The normal *c-myc* exon 2 is 774 nt long (128819675 – 128820448), whereas the truncated variant is only 195-nt long (128819675 – 128819869). The truncated *c-myc* exon 2 is joined to the 3' part of the HPV16 early region starting with the splice acceptor at position 3358 in E2/E4, and terminating with the poly-A signal at position 4215. Between the *c-myc* and HPV16 parts, the three nucleotides "AAG" are found in the hybrid mRNA. The origin of this tri-nucleotide sequence will be described later in *Section 2.6*. The RACE cDNA sequence of the *myc*-HPV16 mRNA was confirmed by convenient RT-PCR using primer 128817576-F (in *c-myc* exon 1) in combination with HPV16-3572-R (in HPV16-E2) followed by cloning and sequencing (data not shown). The cDNA sequence of *myc*-HPV16 mRNA is shown in *Appendix 2*.

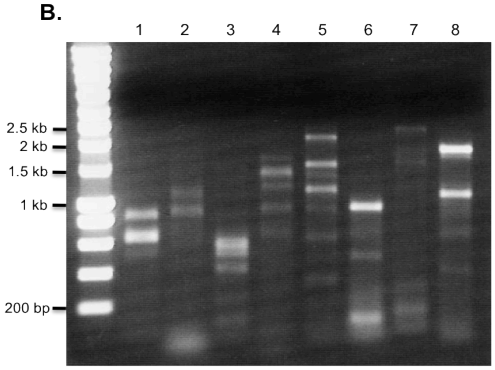
In addition to *myc*-HPV16, an HPV16-*myc*-HPV16 mRNA was also revealed by RACE (*Figure 2-10C and D*). It differs from the *myc*-HPV16 mRNA by replacing the *myc*-exon1 with HPV16-E6* (position 97–226). This mRNA is probably not expressed to a notable level in MRI-H186, because no corresponding band (below *myc*-HPV16 mRNA) was detected on Northern blots (see *Figures 2-7* and *2-9*). The existence of HPV16-*myc*-HPV16 mRNA was proven by RT-PCR using primer pair HPV16-103-F (in E6 and E6*) and HPV16-3572-R (in E2) followed by cloning and sequencing (data not shown).

Additional transcripts identified by RACE include the HPV16-*myc* and the normal *c-myc* mRNA as expected (see *Section 2.2.2*). Furthermore, HPV16 early transcripts were also found. These results strongly suggested that MRI-H186 cells may harbor integrated HPV16 copies containing an intact HPV16 early region. The HPV16 early transcripts were verified by RT-PCR using primers HPV16-103-F and HPV16-3572-R followed by cloning and sequencing (data not shown).

A.

RACE Reaction	5′ or 3′ RACE	Gene-Specific Primer	
		Gene (Exon)	Primer Position
1	5′	HPV16-E7	HPV16-855-R
2	5′	HPV16-E2/E4	HPV16-3565-R
3	5′	*c-myc* exon 1	128818021-R
4	5′	*c-myc* exon 2	128820371-R
5	3′	HPV16-E7	HPV16-683-F
6	3′	HPV16-E2/E4	HPV16-3386-F
7	3′	*c-myc* exon 1	128817671-F
8	3′	*c-myc* exon 2	128819726-F

B.



C.

Sequence	Sequence Frequency	RACE Reaction	Exon Composition	mRNA Designation
1	2	2; 7	*myc*-Ex1, *myc*-Ex2(tr),HPV16-E2	*myc*-HPV16
2	1	2	HPV16-E6*III, *myc*-Ex2(tr), HPV16-E2	HPV16-*myc*-HPV16
3	2	8	*myc*-Ex2(tr), HPV16-E2	*myc*-HPV16 **or** HPV16-*myc*-HPV16
4	1	2	E6*II, E7, E1^E4, E5	HPV16 early transcript
5	4	5	E7, E1^E4, E5	HPV16 early transcript
6	6	5	HPV16-E7, *myc*-Ex2, *myc*-Ex3	HPV16-*myc*
7	2	4	HPV16-E6, HPV16-E7, *myc*-Ex2	HPV16-*myc*
8	1	4	HPV16-E6*I, HPV16-E7, *myc*-Ex2	HPV16-*myc*
9	1	4	HPV16-E6*II, HPV16-E7, *myc*-Ex2	HPV16-*myc*
10	1	4	HPV16-E6*III, *myc*-Ex2	HPV16-*myc*
11	3	4	*myc*-Ex1, *myc*-Ex2	*c-myc*
12	4	7	*myc*-Ex1, *myc*-Ex2, *myc*-Ex3	*c-myc*
13	4	8	*myc*-Ex2, *myc*-Ex3	*c-myc* **or** HPV16-*myc*

D.

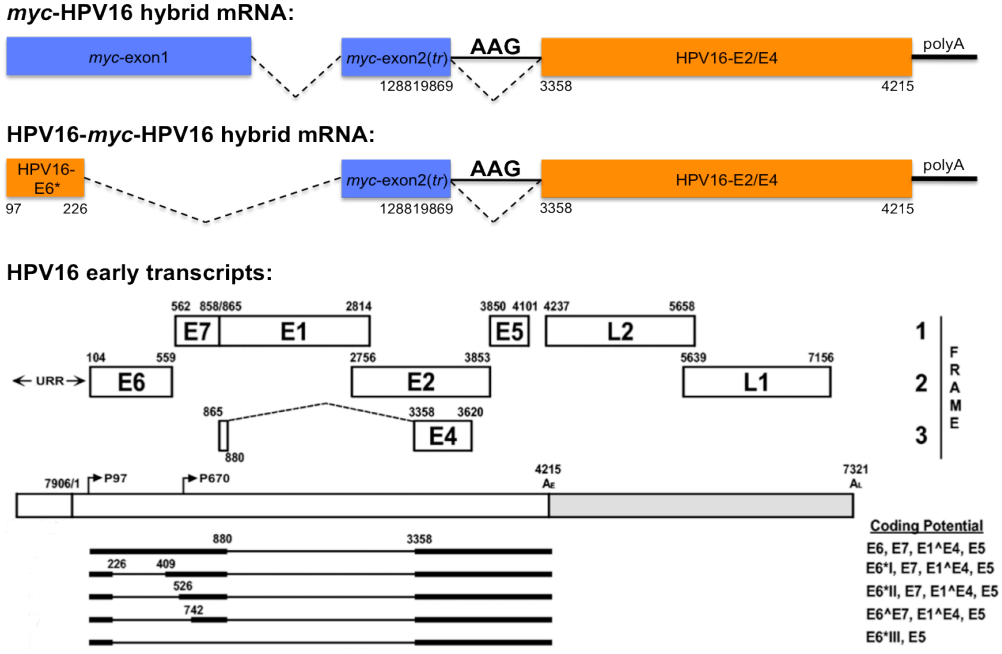
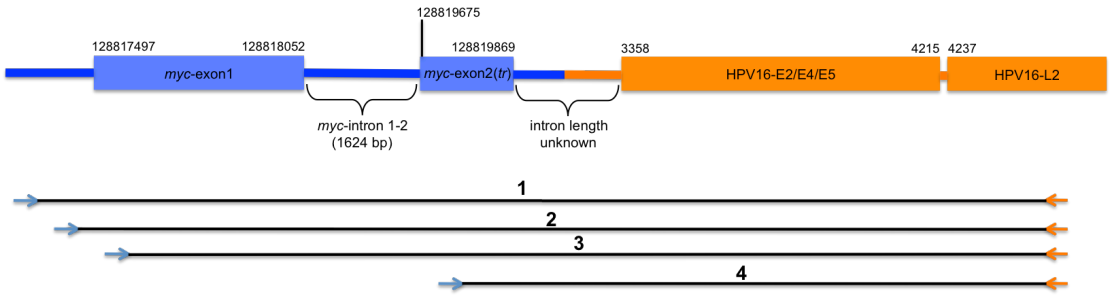


Figure 2-10: Identification of mRNA variants in MRI-H186 by RACE. Positions of gene-specific primers for RACE are shown in panel **A**. Final PCR products of the eight RACE reactions are shown on agarose gel (panel **B**). The SmartLadder was used as size marker. Thirteen non-redundant sequences each containing at least two exons are shown in panel **C** (Ex = exon; tr = truncated). The structures of two novel *myc*-HPV16 hybrid transcripts and HPV16 early transcripts identified by RACE are shown in panel **D**. The HPV16 transcript map was taken from Zheng and Baker (2006).

2.6 Genetic origin of the *myc*-HPV16 hybrid mRNA

The structures of the novel *myc*-HPV16 hybrid transcript could not be deduced from the DNA architecture of HPV16 integration variant A (see *Figure 2-5A*). Therefore, the HPV16 integration variant responsible for expression of the *myc*-HPV16 hybrid mRNA had to be identified. For this purpose, long-PCR reactions with combinations of five different *c-myc* forward primers and one HPV16 reverse primer were performed (*Figure 2-11*). The HPV16 reverse primer has been chosen with a location in the L2 ORF, assuming that the integrated HPV16 DNA extends beyond the early region poly-A site (position 4215). Selected PCR amplicons (red arrows in *Figure 2-11B*) were cloned and sequenced.

A.



B.

Reaction	Primer Forward	Primer Reverse	Product Observed
1	128816664-F	HPV16-5125-R	approx. 6 kb
2	128817197-F	HPV16-5125-R	approx. 5.5 kb
3	128817576-F	HPV16-5125-R	approx. 5 kb
4	128819700-F	HPV16-5125-R	approx. 3 kb
5	128821823-F	HPV16-5125-R	negative

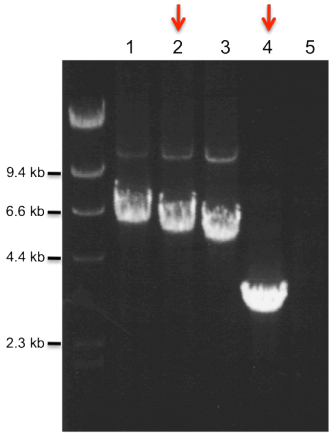


Figure 2-11: Long-PCR analysis to amplify the DNA architecture responsible for the *myc*-HPV16 mRNA. Based on the RACE results, the DNA structure for the *myc*-HPV16 hybrid mRNA was predicted (panel **A**). Five *c-myc* forward primers were used in combination with the same HPV16 reverse primer for long-PCR (panel **B**). The λ/*Hind*III was used as size marker. The PCR products of reactions 2 and 4 (red arrows on gel photo) were cloned and sequenced.

The sequences revealed a new genomic viral-cellular junction in MRI-H186 cells. The HPV16 DNA in this architecture is named HPV16 integration variant B. At the junction site, the *c-myc* gene is disrupted at position 128819869 inside of exon 2, and connected with HPV16 DNA at position 2201 within the E1 region (*Figure 2-12*). This DNA rearrangement activates the splice donor sequence “AAG’GTA” located at HPV16 position 2201–2206, which matches perfectly to the consensus sequence of splice donors (A/C-A-G’G-T-A/G), but is not used in the context of HPV16 early gene transcription (see *Figure 1-2*).

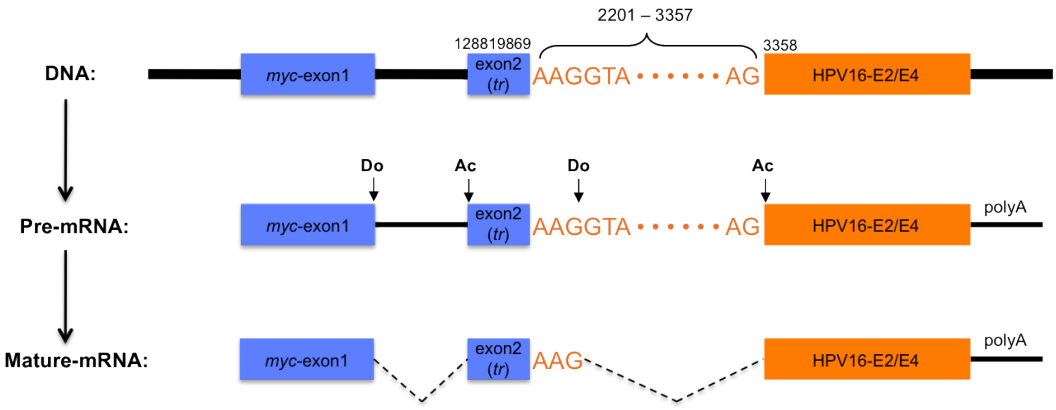


Figure 2-12: Genetic origin of the *myc*-HPV16 hybrid transcript. The structure of HPV16 DNA integration variant B is shown (top). The transcription procedure from DNA to *myc*-HPV16 mRNA is schematically demonstrated. (Do = splice donor; Ac = splice acceptor)

2.7 The novel MycHPV16E2 fusion protein in MRI-H186

2.7.1 Coding potential of *myc*-HPV16 hybrid mRNA

The amino acid sequence encoded by the *myc*-HPV16 hybrid mRNA was predicted from the cDNA sequence. Because the fused HPV16 part harbors both the E2 ORF and the overlapping E4 ORF, it was of interest to determine whether translation initiated at the *c-myc* start codon will continue in frame with HPV16-E2 or E4. As shown in *Figure 2-13*, the *myc*-HPV16 hybrid transcript has the potential to encode a MycHPV16E2 fusion protein.

The resulting fusion protein contains the first 60 amino acids of human c-MYC at the N-terminus, followed by the last 165 amino acids of HPV16-E2 at the C-terminus (*Figure 2-13B*). Both c-MYC and HPV16-E2 proteins are DNA-binding transcription factors with a similar domain organization, i.e., both start with a trans-activation

domain (TAD) and end with a DNA-binding domain (DBD), linked together by a flexibly structured hinge region. The MycHPV16E2 fusion protein contains the N-terminal 60 residues of the c-MYC TAD and the complete HPV16-E2 DBD, connected by the HPV16-E2 hinge region. The sequence of the MycHPV16E2 fusion protein is shown in *Appendix 2*.

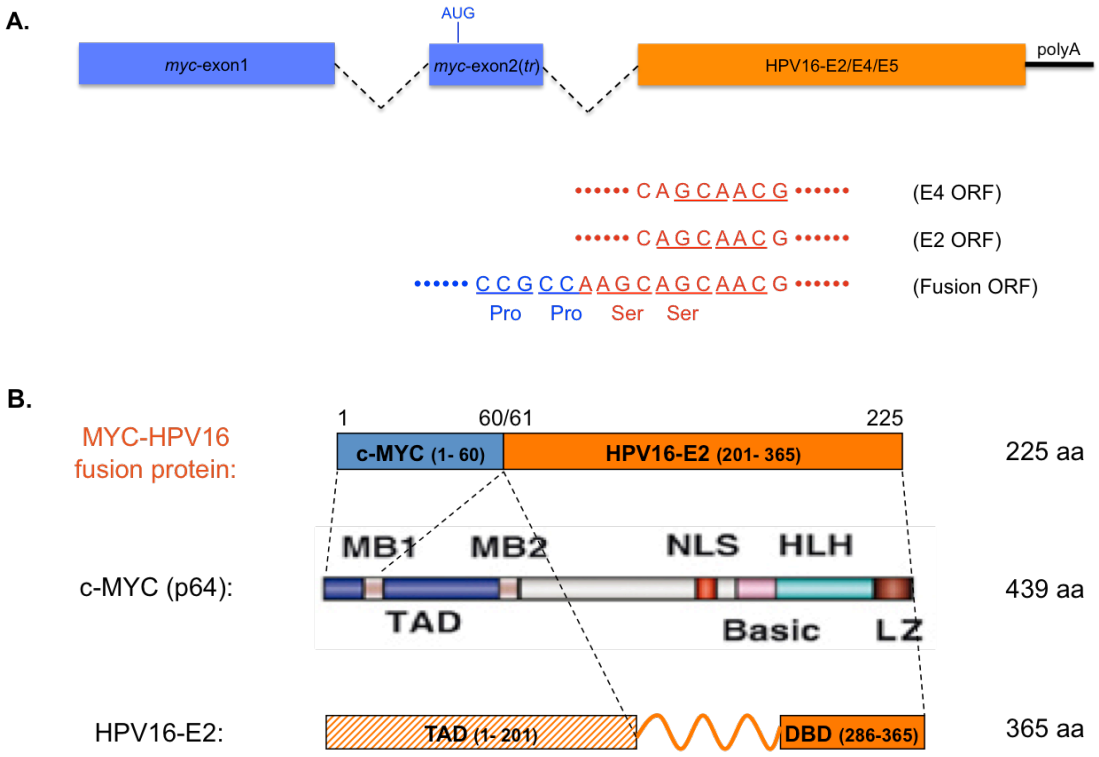


Figure 2-13: Coding potential of the *myc*-HPV16 hybrid mRNA. The HPV16 E2 and E4 ORF in the region affected by HPV16 DNA integration were compared with the fusion ORF (panel **A**). The comparison indicates that the truncated c-MYC ORF is fused to the HPV16-E2 ORF. The triplet codons involved are underlined in different colours (blue for *c-myc* and red for HPV16). The structure of the MycHPV16E2 fusion protein is shown in panel **B**. (TAD = trans-activation domain; DBD = DNA-binding domain)

2.7.2 Detection of MycHPV16E2 fusion protein

According to the predicted sequence, the novel MycHPV16E2 fusion protein is composed of 225 amino acids with a calculated molecular weight of 25.6 kDa, if the AUG start codon is used for translation. For detection of this protein on Western blot, the anti-c-MYC antibody specific for a C-terminal epitope (see *Section 2.3.2*) is not suitable. Likewise, the HPV16-E2 antibody commercially available recognizes the N-terminus of the E2 protein. Therefore, Western blot analysis was performed using a c-MYC antibody clone Y69 directed against an N-terminal epitope of c-MYC (see *Section 4.6*). This antibody should recognize both c-MYC and the MycHPV16E2

fusion protein. The results are shown in *Figure 2-14*. Indeed, MycHPV16E2 fusion protein was detected exclusively in MRI-H186 cell lysates, but not in other cell lines (*Figure 2-14*).

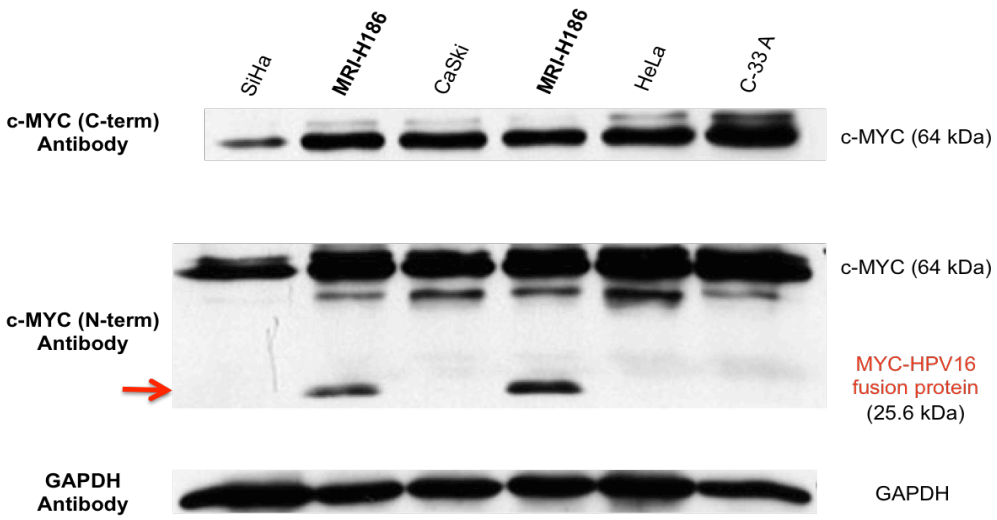


Figure 2-14: Detection of MycHPV16E2 fusion protein by Western blot. The c-MYC antibodies specific for the N-terminus (clone Y69) or the C-terminus (clone 9E10) were used for detecting the c-MYC-related signals. The MRI-H186 cell lysates from two individual preparations were analysed in parallel. GAPDH protein was used as an internal sample-loading control. The red arrow points to the MycHPV16E2 fusion protein.

2.8 Comprehensive analysis of HPV16 integration variants in MRI-H186

The DNA and RNA analysis mentioned before clearly demonstrated that at least two HPV16 DNA integration variants exist in MRI-H186, and both are located in the *c-myc* region. Integration variant A refers to the DNA structure identified by RS-PCR (see *Section 2.2.1*). Integration variant B is the DNA structure from which the MycHPV16E2 fusion protein derives (see *Section 2.6*). In addition, the *c-myc* wild-type allele unchanged by HPV16 DNA integration is also existent in MRI-H186 (see *Section 2.2.1*). This section is concerned with further investigation of HPV16 DNA integration variants existent in MRI-H186, through the combination of Southern blot and long-PCR. It involves the identification of novel HPV16 integration variants and characterization of their flanking cellular sequences.

2.8.1 HPV16 integration variants A and A+

Southern blot analysis was performed with genomic DNA digested with *Hind*III and *Xba*I, respectively. Both enzymes are HPV16 non-cutter, and therefore have recognition sites in the flanking cellular sequences. Deduced from the DNA sequence of integration variant A, cleavage should produce HPV16-cellular hybrid fragments of 16.8 kb (*Hind*III) and 12.8 kb (*Xba*I), respectively (*Figure 2-15A*). After hybridization with the HPV16-complete DNA probe, however, two bands from each digestion were observed on the blot (*Figure 2-15B*). These results indicated that there are two major HPV16 integration variants in MRI-H186. The two lower bands were compatible with the expected fragments (*Hind*III and *Xba*I) of integration variant A. The two larger bands with sizes greater than 23 kb for *Hind*III and about 20 kb for *Xba*I, should represent another integration variant.

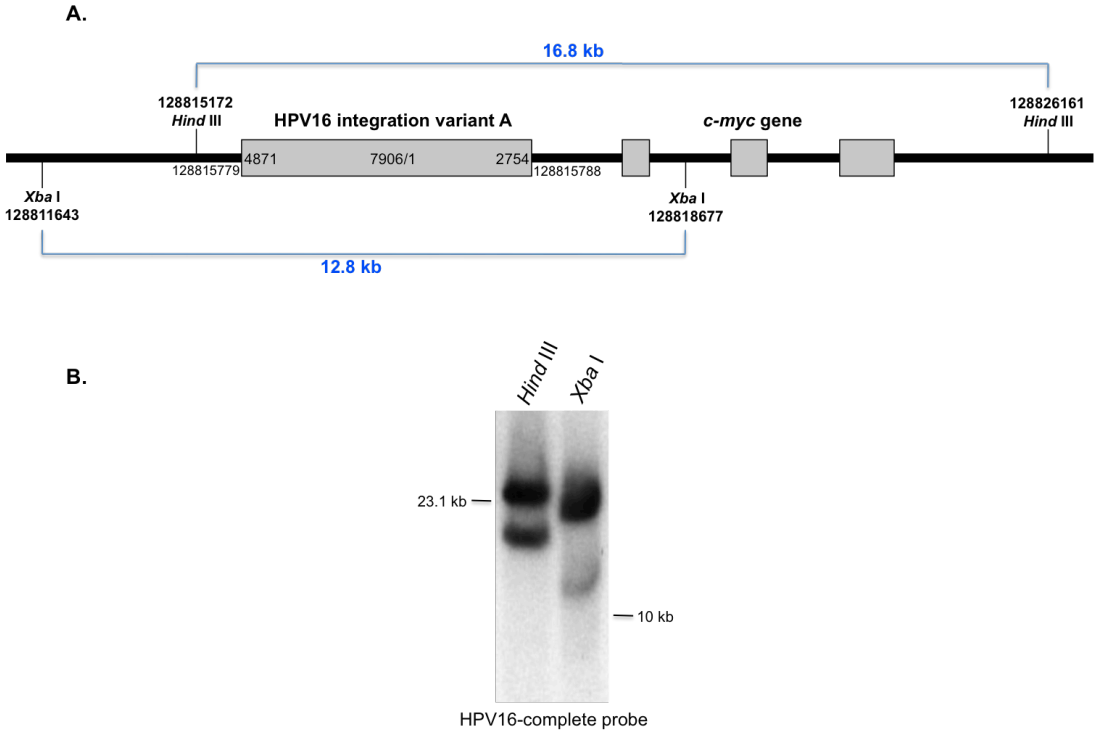


Figure 2-15: Southern blot analysis of HPV16 DNA in MRI-H186. Recognition sites for *Hind*III and *Xba*I flanking the HPV16 integration variant A are shown in panel **A**. MRI-H186 genomic DNA cleaved by *Hind*III or *Xba*I was blotted and hybridized with the HPV16-complete probe (panel **B**). Positions for the 23.1-kb band of λ/*Hind* III and the 10-kb band of SmartLadder marker are marked for size determination.

As mentioned in *Section 2.5*, RACE assay suggested that at least one copy of the complete HPV16 genome is available in MRI-H186. The size difference between the smaller and the larger fragment from each digestion on Southern blot is about 8 kb (see *Figure 2-15*). Therefore, it was assumed that the larger Southern fragments contain the same viral-cellular junctions as integration variant A, but include additionally a complete HPV16 genome. This hypothetical structure was named HPV16 integration variant A+ and is shown in *Figure 2-16*. The calculated sizes of A+ would be 24.7 kb (16.8 kb + 7.9 kb) for *Hind*III and 20.7 kb (12.8 kb + 7.9 kb) for *Xba*I, in agreement with the bands detected on Southern blot (see *Figure 2-15*).

In order to test the assumption, it was first attempted to amplify the complete integration variant A+ by long-PCR using cellular primers 128815697-F and 128816053-R (positions shown in *Figure 2-16*). However, the expected product of 14 kb was not obtained, probably because it is too long to be amplified. Furthermore, the same primers also bind to integration variant A as well as to the *c-myc* wild-type allele, and therefore favored the generation of these shorter amplicons of 6.1 kb and 357 bp, respectively (data not shown). To overcome this problem, long-PCR was performed with combinations of cellular and HPV16 primers located in E2 or L2 (*Figure 2-16*), since this region is deleted in integration variant A. As a result, the two PCR amplicons 1 and 2 with overlapping sequences in the HPV16-E2/L2 region were successfully amplified and cloned. One clone of each amplicon was fully sequenced and the sequence alignment was in complete agreement with the predicted structure of integration variant A+ (*Figure 2-16*).

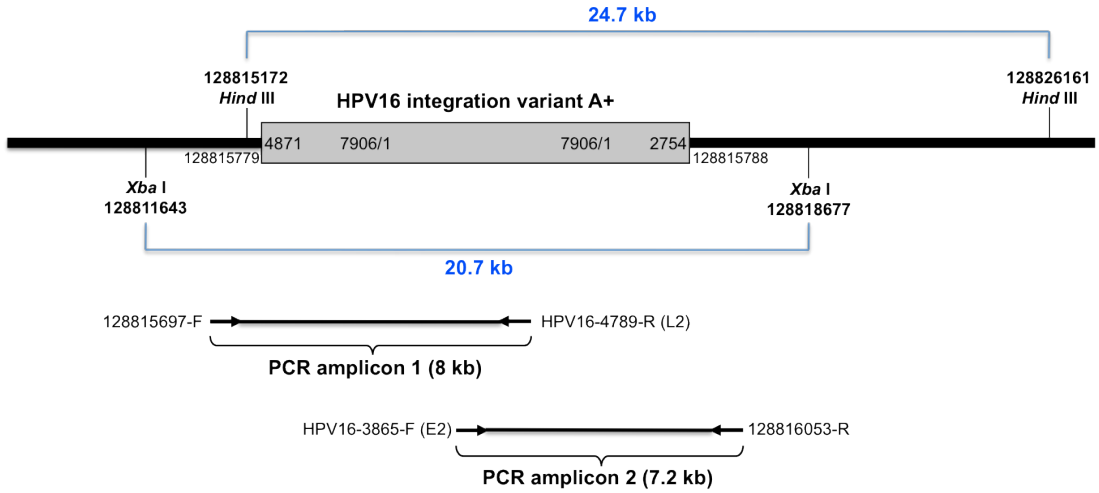
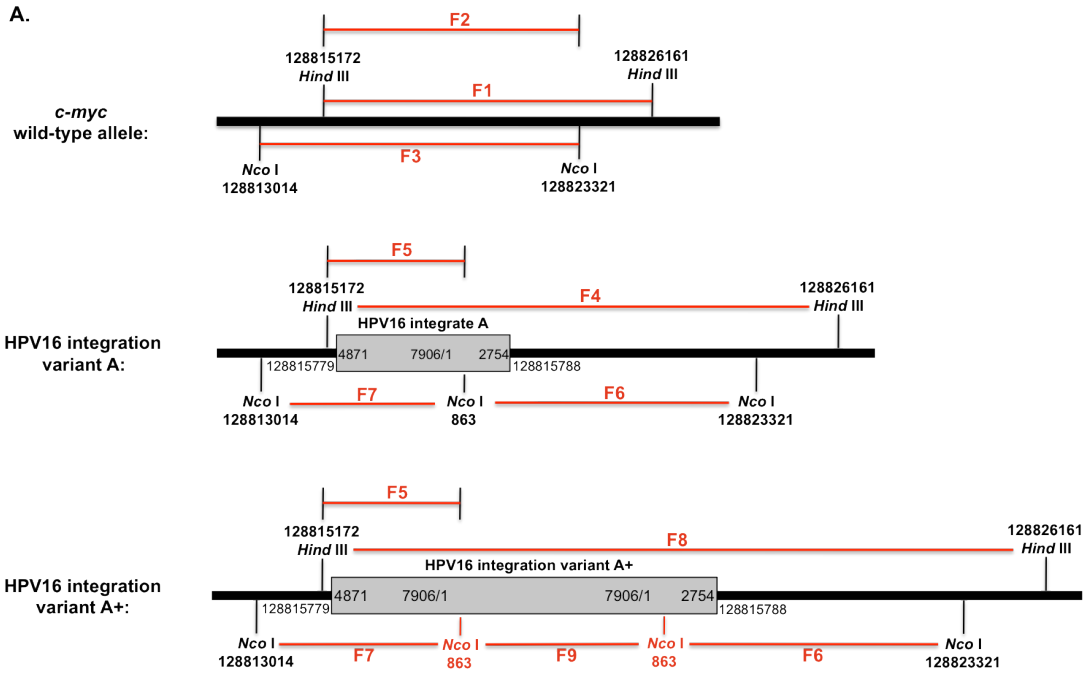


Figure 2-16: Long-PCR analysis of HPV16 integration variant A+ in MRI-H186. The predicted structure of integration variant A+ as well as the recognition sites for *Hind*III and *Xba*I are shown. The locations of the two overlapping amplicons 1 and 2 are indicated below.

Since integration variant A+ was not obtained in one piece by long-PCR, further Southern blot analysis was performed to confirm the predicted structure. Besides the HPV16 non-cutter (*Hind*III), an HPV16 single-cutter *Nco*I was used, in order to divide variant A+ into shorter fragments, which are suitable for size determination on Southern blot. Cleaving at position 863 of HPV16 DNA, *Nco*I cuts twice within variant A+ and provides three fragments: an entire HPV16 genome (7.9 kb), and two viral-cellular junctions identical to those found in variant A (*Figure 2-17A*). Southern blot analysis was performed using MRI-H186 DNA cleaved by *Hind*III and *Nco*I in single and double-digestions. The blots were hybridized with three probes: HPV16-complete, HPV16-E2 and *c-myc* exon 3. The expected and experimentally observed results are compiled as diagrams, table and autoradiographs in panels A, B and C of *Figure 2-17*, respectively. The bands observed on the blots agreed with the predicted fragment patterns of integration variants A and A+. In particular, the 7.9-kb *Nco*I fragment (F9) representing the complete HPV16 genome was detected with both the HPV16-complete and E2 probes. In addition, the *c-myc* wild-type allele previously identified by PCR (see *Section 2.2.1*) was detectable by the *c-myc* exon 3 probe (fragments F1, F2 and F3).

It needs to be noticed that unexpected HPV16-E2 hybridization signals were found at positions similar to those of integration variant A. These bands are labelled as “X” in *Figure 2-17C* with HPV16-E2 probe. Since integration variant A is supposed to be absolutely negative for E2, it should hybridize only with probes HPV16-complete and *c-myc* exon 3 (fragment F4). This paradox can be explained by two possible reasons: either there could be another unknown HPV16 integration variant in MRI-H186, or the HPV16-E2 probe cross-hybridizes with integration variant A. The latter possibility was ruled out, because the HPV16-E2 probe did not hybridize with the cloned integration variant A (data not shown).

A.



B.

DNA Region	Restriction Enzyme	Number of Fragments*	Expected Size	Specificity to DNA Probe	Observed Size on Southern Blot
c-myc wild-type allele	*Hind* III	one	11 kb (F1)	*c-myc* exon3	11 – 12 kb (as expected)
	Hind III + *Nco* I	one	8.1 kb (F2)	*c-myc* exon3	about 8 kb (as expected)
	Nco I	one	10.3 kb (F3)	*c-myc* exon3	not observable, hidden by F6
HPV16 integration variant A	*Hind* III	one	16.8 kb (F4)	HPV16-complete & *c-myc* exon3	15 – 20 kb (as expected)
	Hind III + *Nco* I	two	4.5 kb (F5)	HPV16-complete	about 4.5 kb (as expected)
			9.4 kb (F6)	HPV16-complete & *c-myc* exon3	about 9.5 kb (as expected)
	Nco I	two	6.7 kb (F7)	HPV16-complete	about 7 kb (as expected)
			9.4 kb (F6)	HPV16-complete & *c-myc* exon3	about 9.5 kb (as expected)
HPV16 integration variant A+	*Hind* III	one	24.7 kb (F8)	all three probes	about 24 kb (as expected)
	Hind III + *Nco* I	three	4.5 kb (F5)	HPV16-complete	about 4.5 kb (as expected)
			7.9 kb (F9)	**HPV16-complete & HPV16-E2**	**about 8 kb (as expected)**
			9.4 kb (F6)	HPV16-complete & *c-myc* exon3	about 9.5 kb (as expected)
	Nco I	three	6.7 kb (F7)	HPV16-complete	about 7 kb (as expected)
			7.9 kb (F9)	**HPV16-complete & HPV16-E2**	**about 8 kb (as expected)**
			9.4 kb (F6)	HPV16-complete & *c-myc* exon3	about 9.5 kb (as expected)

* Only fragments hybridizing with at least one of the three DNA probes are mentioned.

The Southern blots are shown in *Figure 2-17C* on next page.

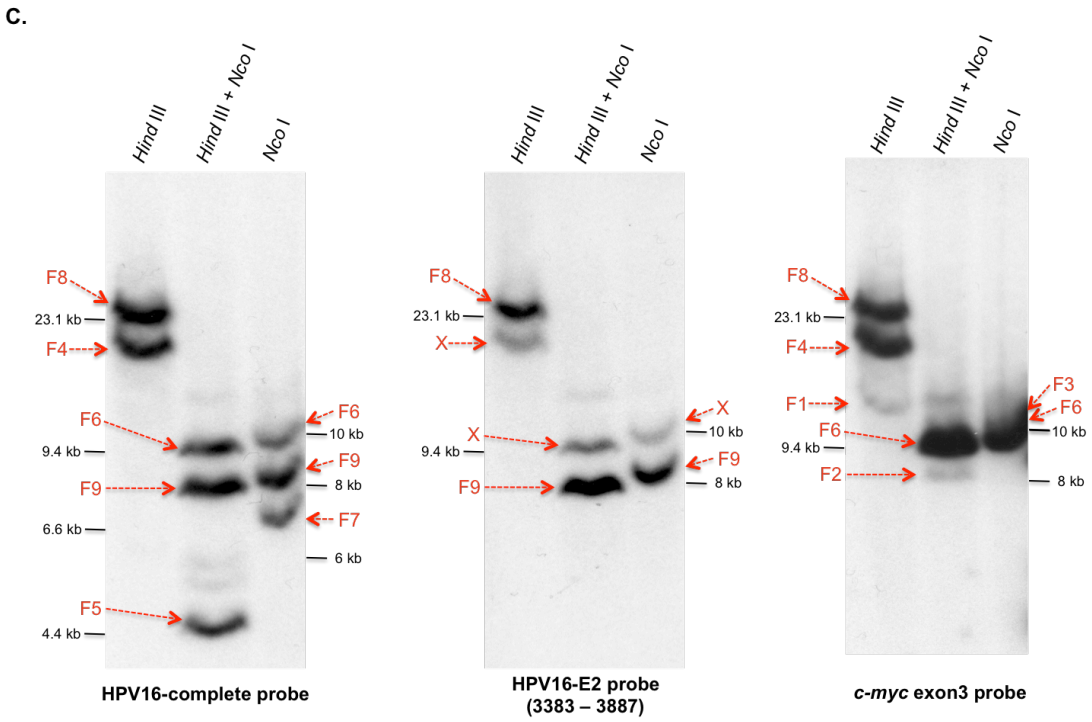


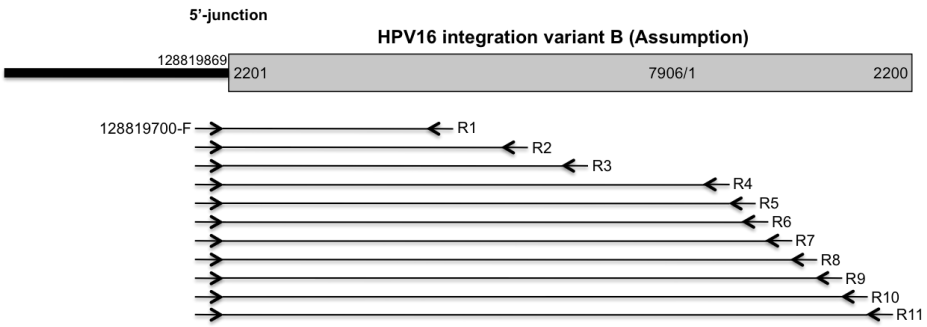
Figure 2-17: Southern blot analysis to verify the structure of HPV16 integration variants A and A+ as well as the *c-myc* wild-type allele in MRI-H186. The structures of *c-myc* wild-type allele and integration variants A and A+ as well as the recognition sites of *Hind*III and *Nco*I are demonstrated in panel **A**. For each DNA architecture, the total number, the expected sizes and the probe specificity of cleaved fragments (single or double-digested) were calculated and shown in panel **B**. Three Southern blots were prepared in parallel with MRI-H186 genomic DNA cut by *Hind* III, *Nco* I or both, and hybridized with the probes indicated (panel **C**). The expected fragments are designated by numbers from F1 to F9 (in red). Positions of λ/*Hind* III (leftmost) and SmartLadder (rightmost) markers are marked for size determination.

2.8.2 HPV16 integration variants B and A-B

The HPV16 integration variant B is the DNA template of the *myc*-HPV16 hybrid mRNA (see *Figure 2-12*). To detect this DNA architecture on genomic DNA Southern blot, the structure of the integrated HPV16 as well as its flanking cellular sequences to the next recognition sites of *Hind*III or *Nco*I needed to be further characterized. In variant B, a truncated *c-myc* exon 2 is fused to the downstream HPV16 DNA broken at position 2201 at the junction site. According to previous PCR data, the HPV16 sequence continues at least until position 5125 in the L2 ORF (see *Section 2.6*, *Figure 2-11*). In order to localize the 3' junction of variant B, a series of long-PCR reactions were performed using the *c-myc* forward primer 128819700-F (in exon 2) in combination with eleven HPV16 reverse primers covering almost the complete HPV16 genome (*Figure 2-18*). PCR products of expected sizes were observed in

reactions 1 to 7 until the reverse primer position 1098, but no amplification products were obtained from reactions 8 to 11 (*Figure 2-18C*). The results suggested that the 3' breakpoint of HPV16 integration variant B is located between positions 1098 and 1332.

A.



B.

PCR	Reverse Primer	Expected Size	Observed Size
1	HPV16-5712-R (**R1**)	3.7 kb	~4 kb (a. e.)
2	HPV16-6446-R (**R2**)	4.4 kb	~4.4 kb (a. e.)
3	HPV16-6951-R (**R3**)	4.9 kb	~5 kb (a. e.)
4	HPV16-209-R (**R4**)	6.1 kb	~6 kb (a. e.)
5	HPV16-773-R (**R5**)	6.6 kb	~6.6 kb (a. e.)
6	HPV16-817-R (**R6**)	6.7 kb	~6.6 kb (a. e.)
7	HPV16-1098-R (**R7**)	7 kb	~7 kb (a. e.)
8	HPV16-1332-R (**R8**)	7.2 kb	negative
9	HPV16-1692-R (**R9**)	7.6 kb	negative
10	HPV16-1805-R (**R10**)	7.7 kb	negative
11	HPV16-2186-R (**R11**)	8.1 kb	negative

C.

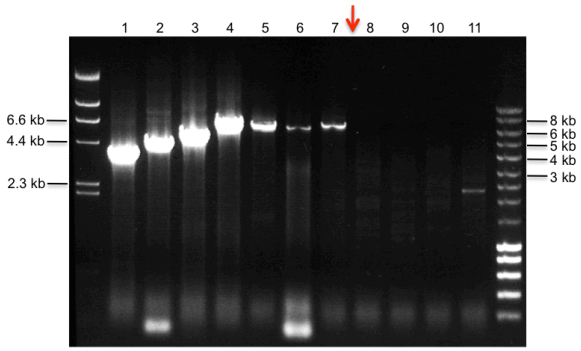


Figure 2-18: Long-PCR analysis to delimit the 3' breakpoint of HPV16 integration variant B. Positions of *c-myc* primer 128819700-F and the eleven reverse primers (R1-R11) throughout the HPV16 genome are shown in panel **A**. The expected sizes of PCR products are given in the table of panel **B**. The long-PCR results are shown in panel **C**. The λ/*Hind*III (leftmost) and SmartLadder (rightmost) were used as size marker. The observed amplicon sizes are included in the table of panel **B** (a.e. = as expected). The red line (in panel **B**) and arrow (in panel **C**) show the boundary between the positive and negative PCR results.

After delimiting the 3' HPV16 breakpoint, the next challenge was to identify the exact 3' viral-cellular junction sequence. The experiment was based on the assumption that insertion of integration variant B has not caused severe deletions in the *c-myc* gene, as observed in variants A and A+. Various long-PCR reactions were performed using the same forward primer 128819700-F as before in combination with reverse primers specific for *c-myc* exon 2 or exon 3. All primer combinations failed to amplify the 3' junction of integration variant B (data not shown). At that time, however, a novel 3' junction sequence was identified by Martine Peter (Institut Curie, Paris) using DIPS-PCR (see *Section 2.9*). In this sequence (*Appendix 1*), HPV16 DNA is broken at

position 1223 and linked to position 128744999 of chromosome 8. This position is located about 72.5 kb upstream of the *c-myc* gene. Since the 3' HPV16 breakpoint falls into the region between positions 1098 and 1332 determined by long-PCR, a tentative structure of variant B was assembled by combining the results from long-PCR and DIPS-PCR (*Figure 2-19*).

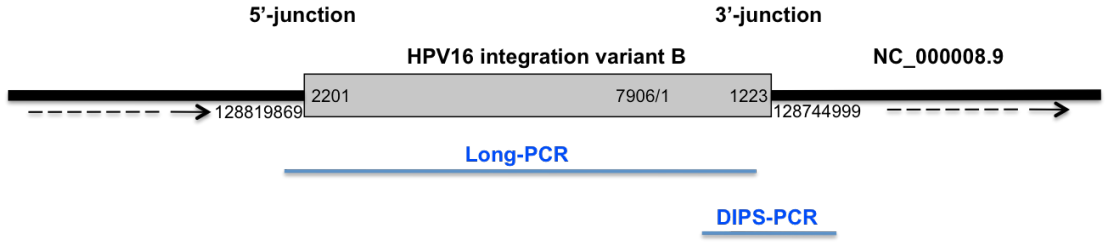
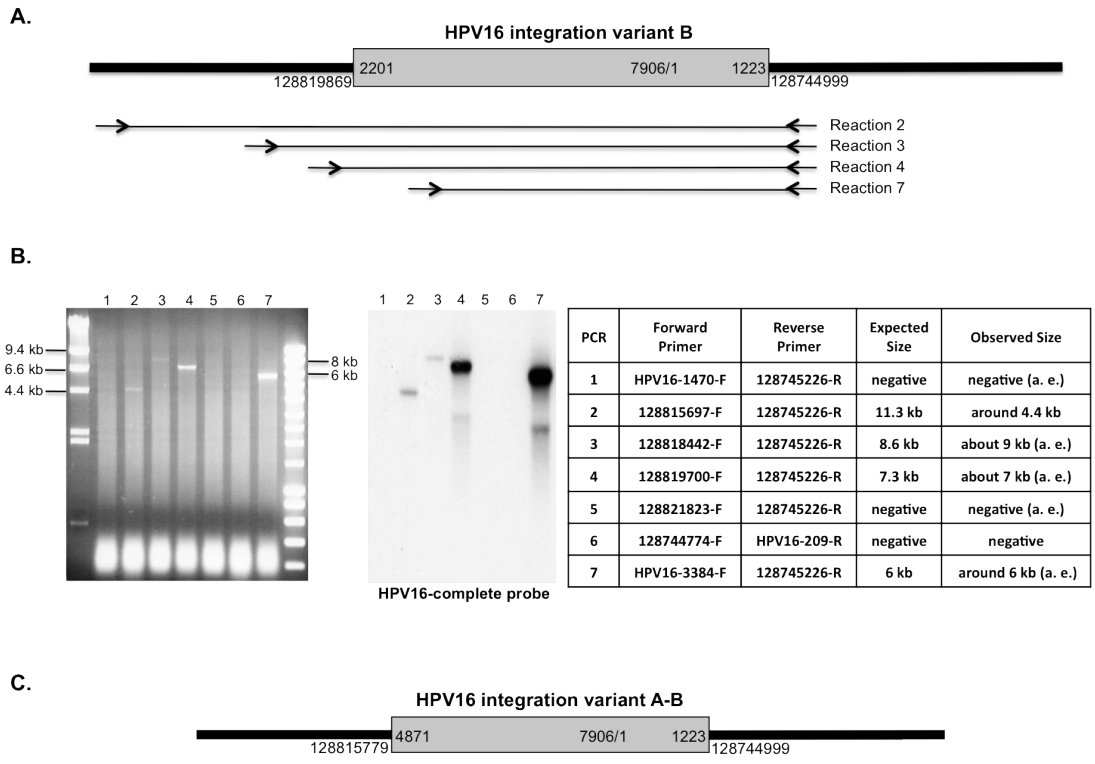


Figure 2-19: Predicted structure of HPV16 integration variant B. The results of long-PCR (see *Figure 2-18*) and DIPS-PCR (Martine Peter, Paris) were combined to predict the complete structure of integration variant **B**. The DNA regions determined by the two approaches are underlined.

To test the predicted structure, seven long-PCR reactions were designed for direct amplification of the complete variant B (*Figure 2-20*). Positive PCR products were expected only from four reactions (no. 2, 3, 4 and 7), while the other three served as negative controls. PCR products were obtained in all four reactions albeit at rather low yield, and hybridized with HPV16-complete DNA probe. Three of them showed the expected sizes on the gel. One of them (7.3 kb, lane 4) was cloned and sequenced. The sequence (data not shown) agreed with the predicted DNA structure of variant B.

The PCR product of reaction 2 showed a size of about 4.4 kb, in contrast to the expected size of 11.3 kb (*Figure 2-20B*). To clarify the structure of this fragment, the PCR product 2 was cloned and sequenced. According to the sequence (data not shown), this product turned out to be a novel HPV16 integration variant, which contains the same 5' junction as variant A (chr8-128815779/HPV16-4871) and the 3' junction of variant B (HPV16-1223/chr8-128744999). Therefore, this novel variant is named A-B (*Figure 2-20C*).



PCR	Forward Primer	Reverse Primer	Expected Size	Observed Size
1	HPV16-1470-F	128745226-R	negative	negative (a. e.)
2	128815697-F	128745226-R	11.3 kb	around 4.4 kb
3	128818442-F	128745226-R	8.6 kb	about 9 kb (a. e.)
4	128819700-F	128745226-R	7.3 kb	about 7 kb (a. e.)
5	128821823-F	128745226-R	negative	negative (a. e.)
6	128744774-F	HPV16-209-R	negative	negative
7	HPV16-3384-F	128745226-R	6 kb	around 6 kb (a. e.)

Figure 2-20: Long-PCR amplification of HPV16 integration variants B and A-B in MRI-H186. Primer positions of PCR reactions 2, 3, 4 and 7 for amplification of integration variant B are shown in panel **A**. Seven long-PCR reactions were performed (panel **B**, left), and the products were blotted and hybridized with HPV16-complete probe (panel **B**, middle). The primers as well as the expected and observed sizes of the PCR reactions are described in the table (panel **B**, right part). Negative PCR results were expected for PCR 1 (because the HPV16-1470-F sequence is not present in variant B), for PCR 5 (because the 128821823-F sequence is not present in variant B), and for PCR 6 (because the 128744774-F sequence is not present in variant B and HPV16-209-R has the "wrong" orientation). The λ/*Hind*III and SmartLadder were used as size marker. The PCR product of lane 4 (panel **B**) was cloned and sequenced. The sequence indicated a novel integration variant A-B (panel **C**). PCR reactions 1, 5 and 6 were expected to be negative, because the forward primers cannot hybridize. (a.e. = as expected)

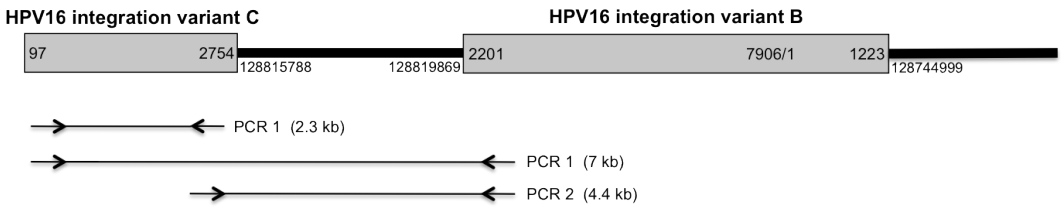
2.8.3 HPV16 integration variant C

An HPV16-*myc*-HPV16 hybrid mRNA was discovered by RACE assay (see *Section 2.5*). This transcript contains HPV16-E6* (position 97–226) instead of *c-myc* exon 1, followed by the truncated *c-myc* exon 2 and the 3' part of HPV16-E2. The transcript structure implied that at the DNA level another HPV16 integration variant (called C) should be located upstream of *c-myc* exon 1 and variant B. It was assumed that variant C has the same 3'-junction as variant A (*Figure 2-21A*). Long-PCR assays were designed and performed to test the assumption (*Figure 2-21B*). The expected 7-kb amplicon specific for the juxtaposed variants C_B was not obtained. Instead, a 2.3-kb amplicon was produced (*Figure 2-21B, reaction 1*). This product was also

expected because the primer pair used also binds to variants A, A+ and C. The amplicon from these templates is only 2.3 kb and preferentially amplified in the PCR reaction. In reaction 2, the 4.4-kb band of template C_B was observed with expected size, besides the 7.7-kb band from variant A+. In reaction 3, the forward primer HPV16-2881-F cannot bind to the predicted structure of integration variant C. As expected, no C_B specific product was found on the gel, but only the PCR product from A+. Comparison between PCR 2 and 3 clearly delimits the 3' HPV16 breakpoint of integration variant C between positions 2701 and 2881, compatible with position 2754. The 4.4-kb product of PCR 2 was cloned and sequenced. The sequence (data not shown) confirmed the predicted DNA arrangement of variants C_B, as shown in *Figure 2-21A*.

The 5' junction of variant C has not been determined. Nevertheless, variant C can be easily distinguished from variants A and A+ by the downstream cellular sequences. For variant C the downstream cellular sequence extends for 4082 bp until position 128819869 including *c-myc* exon 1 and truncated exon 2, while for variants A and A+ the cellular sequences extend further including all three *c-myc* exons (see *Section 2.8.1*).

A.



B.

PCR	Forward Primer	Reverse Primer	Expected Size (DNA template)	Observed Size
1	HPV16-103-F	HPV16-2450-R	7 kb (C_B); 2.3 kb (A, A+, or C)	~2.3 kb (a. e.)
2	HPV16-2701-F	HPV16-2450-R	4.4 kb (C_B); 7.7 kb (A+)	~4.4 kb (a. e.); ~8 kb (a. e.)
3	HPV16-2882-F	HPV16-2450-R	7.5 kb (A+)	~8 kb (a. e.)

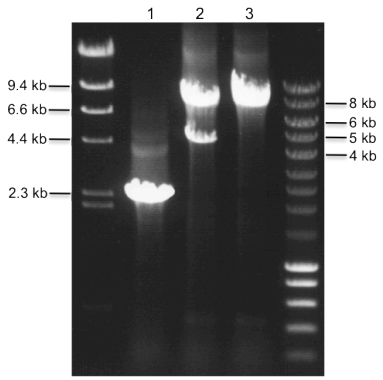
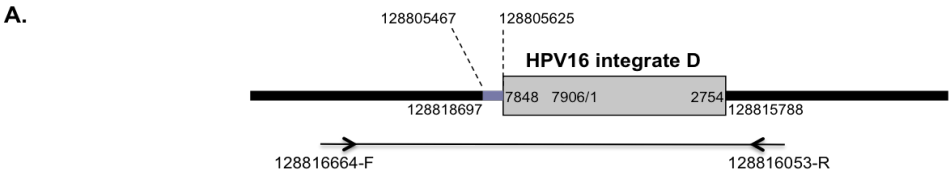


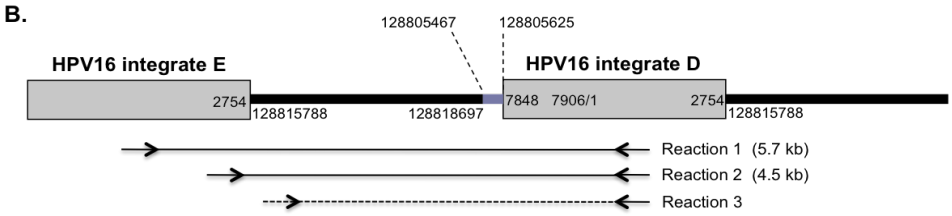
Figure 2-21: Identification of HPV16 integration variant C in MRI-H186 by long-PCR. The predicted structure of integration variant C based on RACE assay together with integration variant B is shown in panel **A**. Primer positions of three long-PCR reactions are marked by arrows. The expected long-PCR products and their DNA templates (integration variants) are listed in panel **B** (a. e. = as expected). Gel analysis of PCR products is shown. The λ/*Hind*III (leftmost) and SmartLadder (rightmost) were used as size marker.

2.8.4 HPV16 integration variants D and E

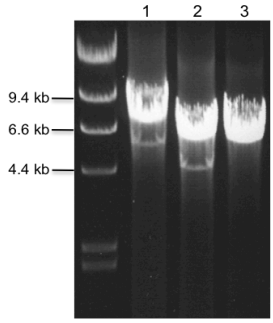
The identification of yet another integration variant was absolutely unexpected because it was obtained from a PCR reaction using primer pair 128816664-F and 128816053-R. This combination was originally considered as a negative control for amplifying integration variant A, because the forward primer is downstream from the reverse primer. Nevertheless, a robust band of around 5 kb was observed on the gel (data not shown). The fragment was cloned and sequenced completely. The sequence showed a novel HPV16 integration variant called D (*Figure 2-22A*). It contains the same 3' junction (HPV16-2754/chr8-128815788) as in variants A and A+, while its 5' HPV16 breakpoint is not at position 4871, but at position 7848. With this breakpoint, only the 3' segment of the URR containing the core promoter is retained upstream of the E6/E7 oncogenes, while the enhancer segment is deleted. Whether the E6/E7 genes are transcribed from this integration variant remains unknown. The 5' junction is featured by an additional DNA re-arrangement: a 159-bp fragment (position 128805467–128805625) originally located about 12 kb upstream from *c-myc* gene is inserted between position 128818697 (*c-myc*) and position 7848 (HPV16).

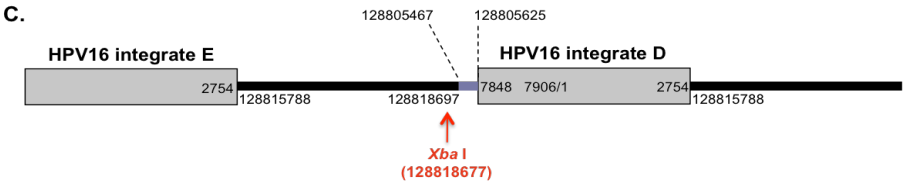
The structure shown in *Figure 2-22A* also revealed that the upstream and downstream cellular regions of HPV16 integration variant D share common sequences of 2910 bp (128815788–128818697). This raised the possibility that variant D is preceded by another integrated HPV16 copy. Long PCR was performed with the assumption that the suspected upstream HPV16 integration variant (called E) contains the same 3'-junction at position 2754 as D. The experimental design and the results are shown in *Figure 2-22B*. The primer pairs for PCR 1 and 2 are expected to produce amplicons of 5.7 kb and 4.5 kb from variants E_D as template, respectively. These sizes should be clearly distinguishable from the amplicons of 7.8 kb and 6.5 kb, expected from template A+. The primer pair for PCR 3, however, should only be able to amplify a product of 6.4 kb from template A+. The PCR results are in full agreement with the expectation (*Figure 2-22B*). However, PCR yield of the two E_D specific fragments was very low, and thus they were difficult to clone. To further verify the existence of integration variant E, diagnostic cleavage reactions with *Xba*I were performed (*Figure 2-22C*). *Xba*I has a recognition site in the cellular sequence between E and D, but not in the amplified HPV16 sequences of variant A+. After *Xba*I cleavage, the sub-fragments expected for variants E_D were obtained.





PCR	Forward Primer	Reverse Primer	Expected Size	Observed Size
1	HPV16-1470-F	HPV16-1332-R	7.8 kb (A+); 5.7 kb (D+E)	one band about 8 kb (a. e.); another band about 6 kb (a. e.)
2	HPV16-2701-F	HPV16-1332-R	6.5 kb (A+); 4.5 kb (D+E)	one band around 6.6 kb (a. e.); another band around 4.4 kb (a. e.)
3	HPV16-2881-F	HPV16-1332-R	6.4 kb (A+)	a band around 6.6 kb (a. e.)





Lane	Sample	Expected Size of *Xba*I subfragments	Observed Size
1	PCR 1	n. a.	n. a.
2	PCR 1 (*Xba*I)	4.2 kb; 1.6 kb	~4.5 kb (a. e.); ~1.6 kb (a. e.)
3	PCR 2	n. a.	n. a.
4	PCR 2 (*Xba*I)	3 kb; 1.6 kb	~3.2 kb (a. e.); ~1.6 kb (a. e.)
5	PCR 3	n. a.	n. a.
6	PCR 3 (*Xba*I)	negative	negative (a. e.)

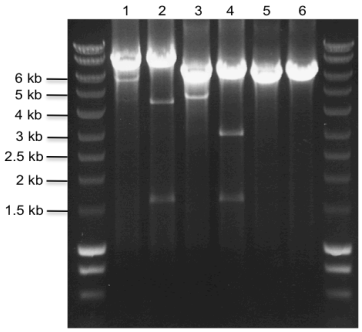


Figure 2-22: Determination of HPV16 integration variants D and E in MRI-H186. Panel **A** shows the structure of variant D and primer positions for the PCR reaction from which variant D was cloned. The blue line indicates the extra cellular sequence at the 5' junction. Three long-PCR reactions were performed to test the predicted DNA structure of E_D (panel **B**). Primer positions and expected size of PCR products are given in the table (a. e. = as expected; n. a. = not applicable). In panel **C**, the results of *Xba*I digestion of PCR products 1 to 3 are shown. The recognition site of *Xba*I for variants E_D is marked by the red arrow.

2.8.5 Summary of HPV16 integration variants identified in MRI-H186

A collection of seven HPV16 integration variants (A, A+, B, C, A-B, D and E) together with their flanking cellular sequences have been identified in MRI-H186, mainly by RS-PCR, long-PCR and Southern blot analysis. The structures are schematically outlined in *Figure 2-23* and *Table 2-24*. Variants A and A+ are regarded as the two major HPV16 DNA integrates in MRI-H186 mainly because they are visible as very strong bands on genomic Southern blots (see *Section 2.8.1*). Furthermore, either the 5' or 3' viral-cellular junction of variants A and A+ (the junction sequences are identical in A and A+) is also available in all other integration variants except in variant B. This suggests that all integration variants in MRI-H186 may originate from a primary integration event, which could be A or A+.

It needs to be noticed that not all HPV16 integration variants existent in MRI-H186 have been fully characterized. The 5' junction sequences of variants C and E remain unclear, and some weak bands on Southern blots (see *Figure 2-17C*) are still difficult to explain. Even for variants B and A-B, which have been amplified and fully sequenced, the expected restriction fragments have not been clearly identified on Southern blots. Furthermore, the order of all the integration variants on chromosome 8 is not known.

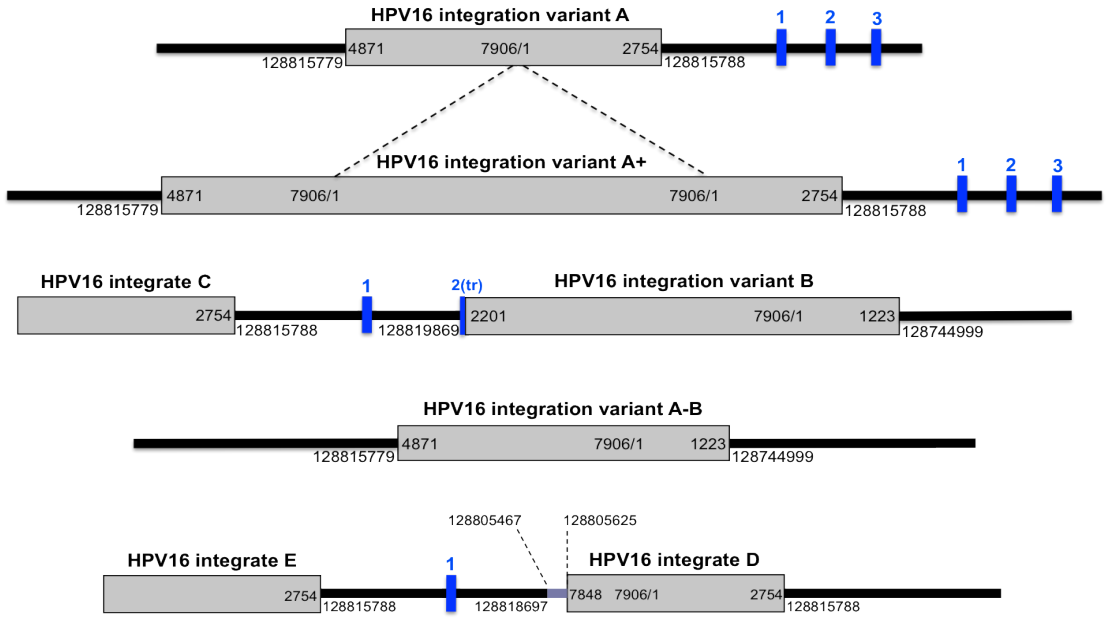


Figure 2-23: Summary of HPV16 DNA integration variants identified in MRI-H186. The schematic demonstration of all HPV16 integration variants characterized in MRI-H186 is shown. Positions of HPV16 and chromosome 8 (NC_000008.9) at the junction sites are given. The vertical blue bars and the numbers above indicate the *c-myc* exons (tr = truncated). The 5' junction of variants C and E were not determined. The downstream cellular sequences of variants A and A+ contain the intact *c-myc* gene (position 128817498 to 128822853).

Table 2-24: Summary of HPV16 DNA integration variants identified in MRI-H186.

HPV16 integration variant	5'-junction (cellular pos. HPV16 pos.)	3'-junction (HPV16 pos. cellular pos.)	Length of HPV16 part	Comments
A	chr8-128815779 HPV16-4871	HPV16-2754 chr8-128815788	5789 bp	major variant *c-myc* exons 1+2+3 HPV16-*myc* mRNA
A+	chr8-128815779 HPV16-4871	HPV16-2754 chr8-128815788	13695 bp	major variant *c-myc* exons 1+2+3 HPV16-*myc* mRNA HPV16 mRNAs
B	chr8-128819869 HPV16-2201	HPV16-1223 chr8-128744999	6929 bp	minor variant *c-myc* exons 1+2(tr) *myc*-HPV16 mRNA MycHPV16E2 protein
C	unknown	HPV16-2754 chr8-128815788	unknown	minor variant
A-B	chr8-128815779 HPV16-4871	HPV16-1223 chr8-128744999	4259 bp	minor variant
D	chr8-128818697 HPV16-7848	HPV16-2754 chr8-128815788	2813 bp	minor variant
E	unknown	HPV16-2754 chr8-128815788	unknown	minor variant
C_B tandem				HPV16-*myc*-HPV16 mRNA *c-myc* exons 1+2(tr) in cellular linker sequence (4081 bp; pos. 128815788–128819869)
E_D tandem				*c-myc* exon 1 in cellular linker sequence (2909 bp + 159 bp = 3068 bp; pos. 128815788-128818697 and pos. 128805467–128805625)

2.9 Supplementary data from Jérôme Couturier and Martine Peter

In order to obtain additional information about the molecular mechanisms of HPV16-induced insertional mutagenesis of *c-myc* proto-oncogene in MRI-H186, cooperation was arranged with Jérôme Couturier and Martine Peter from the Institut Curie (Paris, France). Using methods established in their laboratories and employed for their analysis of HPV16/18 integration in the *c-myc* region of cervical carcinoma samples (Couturier *et al*., 1991; Peter *et al*., 2006), they performed additional experiments with MRI-H186 cells, namely fluorescence in situ hybridization on MRI-H186 chromosomes, spectral karyotyping, DIPS-PCR and quantitative PCR for copy number determination in the *c-myc* region. The data obtained from Paris are summarized in this section.

2.9.1 Fluorescence *in situ* Hybridization analysis

Since multiple HPV16 DNA integration variants and the wild-type *c-myc* allele have been identified in MRI-H186 (see *Section 2.8.5*), it was interesting to find out whether MRI-H186 cells contain a normal chromosome 8 negative for integrated HPV16 DNA, or other HPV16 integration events outside of chromosome 8. To answer these questions, a dual-color fluorescence *in situ* hybridization (FISH) was performed by Jérôme Couturier. The chromosomes of MRI-H186 were hybridized with a blend of the *c-myc* and the HPV16 DNA probe labelled by different fluorescence dyes, respectively. The fluorescence signal of each dye was measured individually and merged together for evaluation. As shown in *Figure 2-25*, integrated HPV16 DNA is always co-localized with the *c-myc* gene in MRI-H186. There was definitely no HPV16 signal found on other chromosomes in this cell line, and no chromosome 8 without integrated HPV16 DNA was observed. Only two chromosomes 8 were detected in each MRI-H186 cell and both were positive for HPV16. Taken together, the FISH assay indicated that the HPV16 DNA integration variants co-localize with the *c-myc* sequences on chromosome 8.

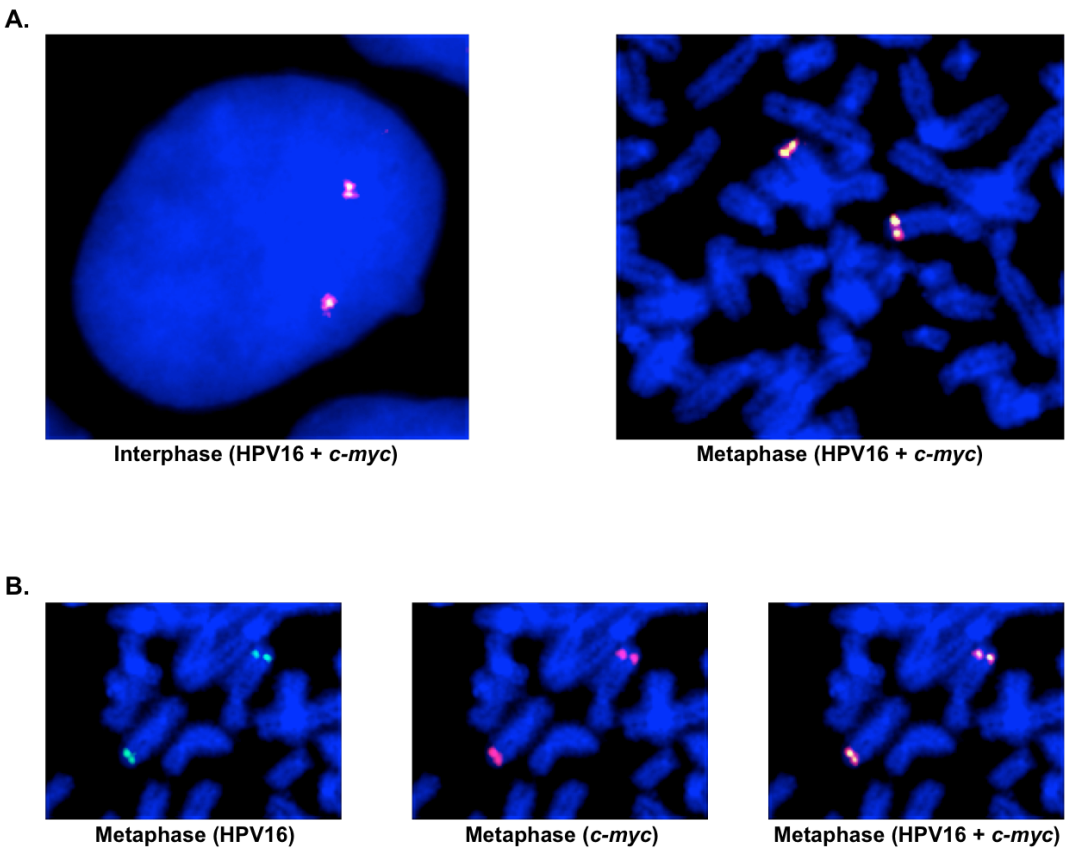


Figure 2-25: Dual-color FISH assay to demonstrate the co-localization of HPV16 DNA integrates and the *c-myc* gene. DAPI was used to visualize DNA/chromosomes (in blue). The HPV16 DNA probe was labelled with FITC (green light emission), while the *c-myc* probe was labelled with Spectrum Orange (red light emission). Cells were hybridized with a mixture of these three fluorescent dyes. The white light signal (red + green + blue = white) indicates the co-localization of HPV16 and *c-myc*. The experiments were performed and the figures were provided by Jérôme Couturier (Institut Curie, Paris).

2.9.2 Spectral karyotyping

In addition to FISH assay, the R-band karyotyping analysis of MRI-H186 cells was also performed by Jérôme Couturier. The cells appeared to be near-tetraploid and highly differed in the numbers of individual chromosomes. It was not possible to obtain a modal chromosome number for this cell line (Jérôme Couturier, personal communication). Therefore, a representative karyotype of MRI-H186 cells is shown in *Figure 2-26*. In agreement with the FISH results, only two chromosomes 8 were found in MRI-H186 cells, although other chromosomes showed great heterogeneity in their numbers.

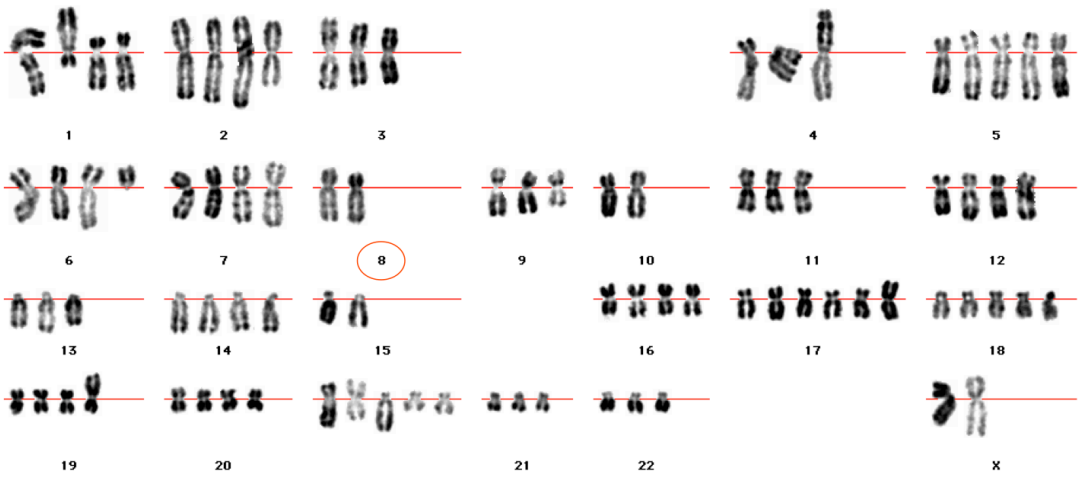


Figure 2-26: R-band karyotype of MRI-H186 cells. The MRI-H186 cells showed overall near-tetraploidy. Only two chromosomes 8 (indicated by the red circle) are present in these cells. The experiments were performed and the figure was provided by Jérôme Couturier (Institut Curie, Paris).

2.9.3 DIPS-PCR

In parallel with our RS-PCR analysis, the DIPS-PCR assay was performed by Martine Peter to determine the HPV16 DNA integration site in MRI-H186. Sequence data from DIPS-PCR revealed two viral-cellular junction sequences of MRI-H186 DNA. One 5'-junction is identical with the chr8-128815779/HPV16-4871 determined by RS-PCR (see *Section 2.1.2*). The other one is the 3'-junction of the integration variants B and A-B (see *Section 2.8.2*).

2.9.4 *c-myc* quantitative PCR

The genomic Southern analysis suggested that the HPV16 integration variants (especially A and A+) together with the flanking cellular sequences seem to be co-amplified in MRI-H186, because of the strong hybridization signals (see *Figure 2-17*). To determine the exact copy number of amplified DNA, quantitative PCR using ten amplicons throughout the *8q24.21* region was performed by Martine Peter (methods and primers as described in Peter *et al*., 2006). Two peaks of DNA copy number were detected from the *8q24.21* region in MRI-H186 (*Figure 2-27*). The higher peak with 26 copies represents the amplicon Q8 within the *c-myc* exon 1. This amplicon involves the *c-myc* wild-type allele as well as the flanking cellular sequences of almost all HPV16 integration variants characterized so far except the minor variant A-

B. The lower peak reflects the copy number of amplicons Q5 and Q6, which are located very close to each other. This cellular region is located about 14 kb downstream of the 3'-junction of integration variants B and A-B (see *Section 2.8.5*).

qPCR Amplicons	Position* on Chromosome 8 (NC_000008.9)	DNA Copy Number (per cell)	
		Normal human DNA	MRI-H186 DNA
Q1	128,243,427	2	2
Q2	128,364,902	2	2
Q3	128,491,860	2	2
Q4	128,655,929	2	2
Q5	128,758,926	2	**8**
Q6	128,759,250	2	**10**
Q7	128,770,803	2	2
Q8 (*c-myc*)	128,817,685	2	**26**
Q9	128,899,363	2	**4**
Q10	129,029,753	2	2

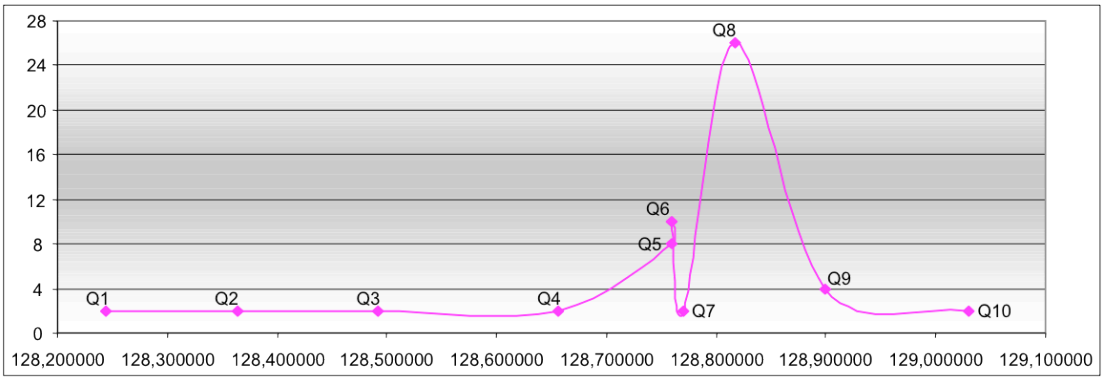


Figure 2-27: Determination of DNA copy numbers in chromosomal region *8q24.21* by quantitative PCR. Start positions* of the ten amplicons and the copy numbers determined by qPCR are shown in the table. The chart shows the copy numbers (Y-axis) and chromosomal locations (X-axis) of all the ten amplicons. The experiments were performed and the figures were provided by Martine Peter (Institut Curie, Paris).

2.10 Development of the novel amplification-selection-pyrosequencing strategy for HPV16 DNA integration analysis (ASP16)

The major goal of the thesis work was to determine HPV16 DNA integration sites in pre-cancerous and cancerous lesions obtained from cervical scrapes. In general, two fundamental challenges are recognized for clinical DNA samples. Firstly, the amount of DNA obtained from cervical scrapes is rather low, approximately 1 μg per sample. From each sample, we obtained about 100 ng of DNA for HPV16 integration analysis. In order to get enough template DNA for PCR-based determination of HPV16 integration sites, DNA samples have to be amplified accordingly. Secondly, clinical DNA samples are often degraded to some extent, during the preservation stage between cervical scraping and DNA isolation. These problems called for a reliable DNA whole genome amplification (WGA) method, which should tolerate DNA degradation. To combine the WGA with integration analysis, a novel strategy named amplification-selection-pyrosequencing for HPV16 (ASP16) was developed in this thesis, which mainly consists of three steps as follows:

Step 1 – Amplification (A): whole genome amplification of clinical DNA sample.

Step 2 – Selection (S): selection of HPV16-containing DNA from the WGA library.

Step 3 – Pyrosequencing (P): use of the next-generation DNA sequencing platform from Roche/454.

2.10.1 ASP16 Step 1 – Amplification

For the first step of ASP16 strategy, the GenomePlex whole genome amplification (GP-WGA) system from Sigma-Aldrich was applied according to the manufacturer's instruction. This is a PCR-based method tolerant to DNA degradation, because genomic DNA templates are fragmented at the very beginning (Langmore, 2002). This advantage is of particular interest for clinical DNA samples. The principal steps of GP-WGA are outlined in *Figure 2-28*. Genomic DNA is randomly fragmented into PCR-amplifiable sizes by chemical digestion and heating. Then, the denatured DNA fragments are exposed to a mixture of degenerate primers, which all contain the same GenomePlex universal adapter (GPUA) sequence at their 5' ends. Through a stepped isothermal amplification, DNA fragments are converted into the so-called OmniPlex library, which represents the whole genomic DNA input. In this library, all DNA fragments of 400 bp on average carry the uniform GPUA at both 5' and 3' ends.

Finally, the OmniPlex library is used as template for whole genome DNA amplification, which is achieved by PCR using GPUA as primer. To reduce amplification bias, PCR is performed with a limited number of thermal cycles (14 cycles). The input for GP-WGA can be as little as 10 ng of genomic DNA, which is easily obtainable from cervical scrapes. In the end, the DNA template will be amplified about 500-fold to a yield of around 5 μg.

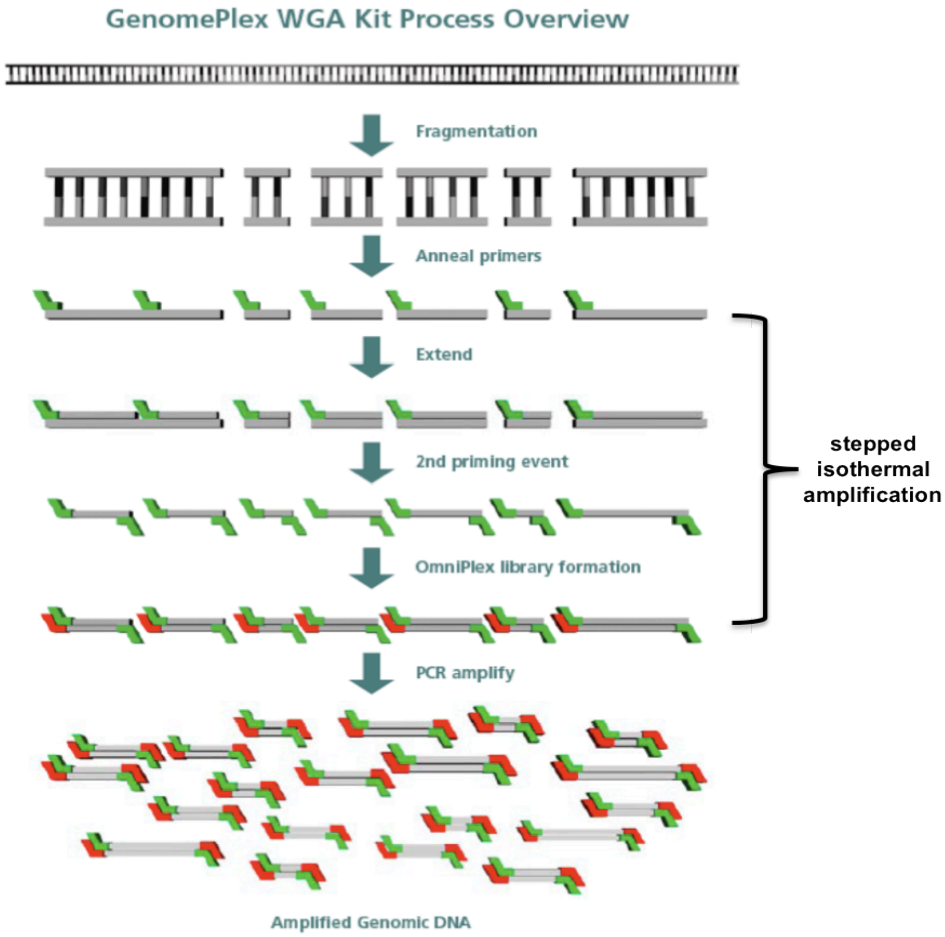


Figure 2-28: The GenomePlex whole genome amplification (GP-WGA) system. The procedure of GP-WGA is schematically demonstrated. Details about individual steps are explained in the text. The figure was taken from the Whole Genome Amplification Advisor provided by Sigma-Aldrich.

To test GP-WGA performance, four reactions were done with different HPV16-positive genomic DNA templates, respectively (*Figure 2-29B*). These included two cervical carcinoma cell lines (MRI-H186, MRI-H196) and two clinical pre-cancerous lesions (50073, 55246). The DNA quality of them was checked on agarose gel before WGA. Compared with the “intact” DNA of cell lines, the clinical samples exhibited a smear of moderately degraded DNA (*Figure 2-29A*). Robust products were obtained from all four GP-WGA reactions (*Figure 2-29B*). The observed product sizes ranged

between 200 bp and 2 kb, with the majority smaller than 1 kb. No difference in GP-WGA performance was found between intact DNA from cell lines and partially degraded DNA from clinical samples. With regard to allelic bias, a series of PCR using MRI-H186 and 50073 GP-WGA products as templates was performed to detect DNA regions in relation to HPV16 integration analysis (*Figure 2-29C*). PCR primers used in this assay are specific for HPV16-E6, HPV16-E2, *c-myc* and *GAPDH* genes. For MRI-H186, junction-specific PCR for both 5' and 3' viral-cellular junctions was also done (*Figure 2-29C*). All ten reactions showed final PCR products with the expected sizes. To conclude, the GP-WGA system is able to amplify the whole genome of moderately degraded clinical DNA samples.

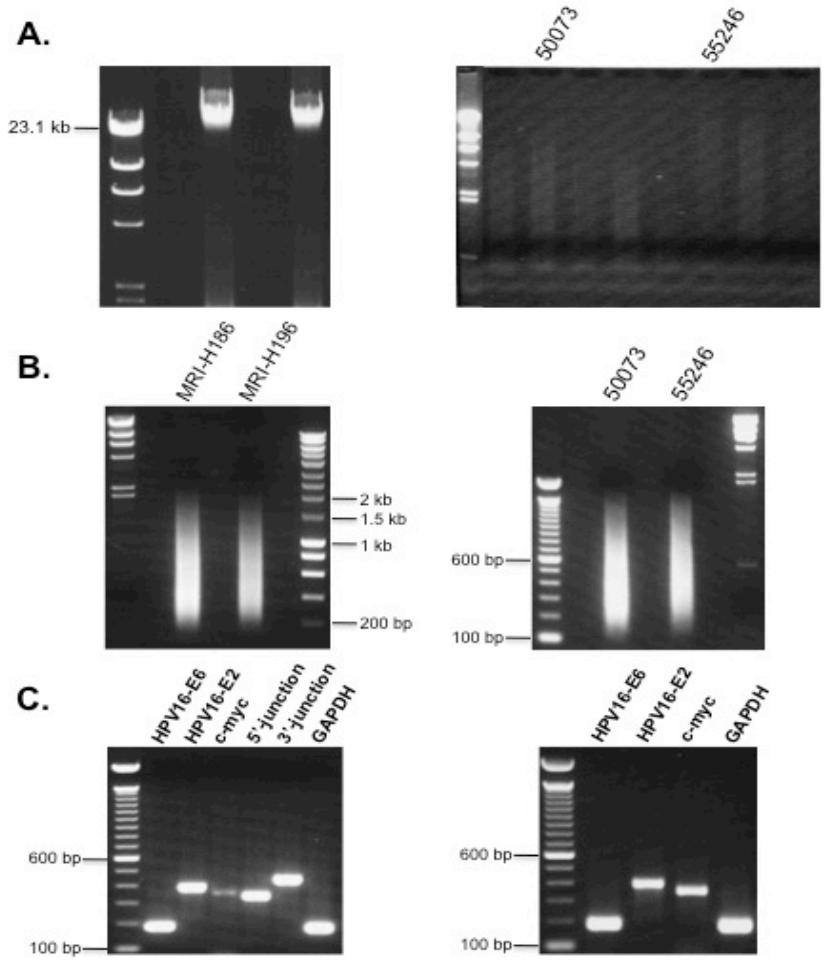


Figure 2-29: The performance of GenomePlex whole genome amplification (GP-WGA). Panel **A** shows DNA of cell lines MRI-H186 and MRI-H196 compared with DNA of altogether eight clinical samples including 50073 and 55246. "Intact" genomic DNA has a size of 50-100 kb, which runs on 0.8% agarose gels as a band above the 23.1-kb λ/*Hind*III fragment. Smears from top to bottom on gel indicate partially degraded genomic DNA of clinical samples. The λ/*Hind*III was used as size marker. Panel **B** shows the final products of four GP-WGA reactions. For each reaction, 4 μl from totally 75 μl was checked on gel. Ten PCR reactions, six for MRI-H186 (left) and four for 50073 (right), are shown in panel **C**. The PCR amplicons included the following: HPV16-E6 (184 bp), HPV16-E2 (395 bp), *c-myc* (356 bp), MRI-H186 5' viral-cellular junction (338 bp), MRI-H186 3' viral-cellular junction (448 bp) and *GAPDH* (173 bp). The 100 bp DNA Ladder was used as size marker.

2.10.2 ASP16 Step 2 – Selection

The GP-WGA system has shown promise to amplify partially degraded clinical DNA samples. However, it was expected that the combination of GP-WGA and RS-PCR is not suitable for determination of HPV16 DNA integration sites in cervical pre-cancerous lesions for two reasons. Firstly, the average length of GP-WGA products (about 400 bp) is too short for RS-PCR, because the latter requires cellular sequences to be relatively long for the specific binding of RSO-primers onto target sites. Secondly, RS-PCR had been shown to be inappropriate for DNA samples with high percentage of HPV16 episome/concatemer *versus* integrated DNA, such as the CaSki cell line (see *Section 2.1.2*). Therefore, an alternative procedure was developed in this thesis in order to amplify HPV16-containing sequences from GP-WGA products.

The GP-WGA library contains three species of DNA molecules: cellular genomic DNA, pure HPV16 DNA and HPV16-cellular junction DNA. For DNA sequencing, the rare HPV16-cellular junction DNA together with the pure HPV16 DNA need to be enriched against the huge background of cellular genomic DNA (*Figure 2-30*).

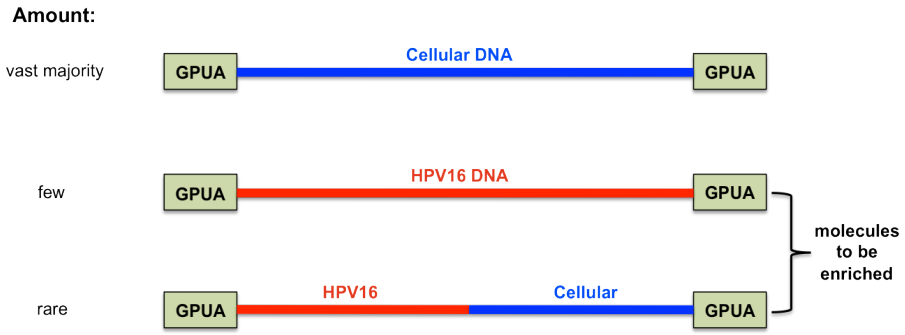


Figure 2-30: Three species of DNA molecules existent in GP-WGA libraries of HPV16-positive DNA samples. (GPUA= GenomePlex universal adapter)

As mentioned in *Section 2.10.1*, the GPUA plays an essential role for the GP-WGA procedure and defines the termini of all final WGA products. Therefore, the applicability of GPUA in combination with HPV16-specific primers for amplification and selection of HPV16-containing sequences from GP-WGA libraries was examined. Information about the exact sequence of GPUA is not given by the manufacturer. To unravel the GPUA sequence, an aliquot of MRI-H186 GP-WGA product was directly cloned into the pSC-A vector. Altogether 14 clones were randomly chosen and sequenced completely (data not shown). The 14 sequences

represent different regions of human genomic DNA, while they all carry the GPUA at both 5' and 3' ends as a common feature. The deduced GPUA sequence is shown in *Figure 2-31*.

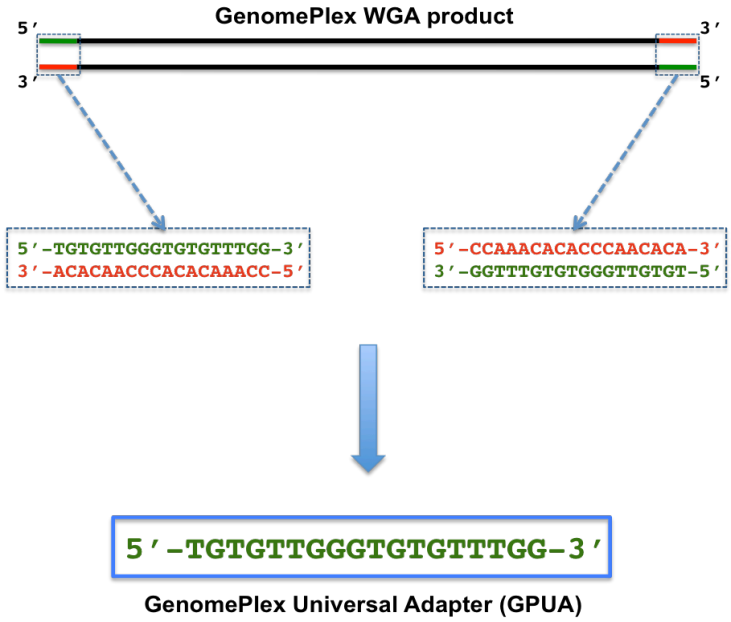


Figure 2-31: The GenomePlex universal adapter (GPUA) sequence. DNA sequences in green and red represent the GPUA sequence and its complementary strand, respectively.

Based on the sequence knowledge, it was then tested whether GPUA can be used as primer together with HPV16 primer for direct PCR amplification of HPV16-containing DNA. Three PCR reactions were performed with MRI-H186 GP-WGA product as template (*Figure 2-32*). GPUA was used as cellular primer (reverse or forward) in combination with an HPV16 E1 forward primer (HPV16-2573-F) and an L2 reverse primer (HPV16-5125-R), respectively. Both HPV16 primers can bind to the integrated HPV16 DNA in MRI-H186. It was therefore expected that HPV16-containing DNA would be specifically amplified with these primer combinations. As control, GPUA was used as both forward and reverse primer. The PCR results are shown in *Figure 2-32*. A DNA smear was observed in all three reactions, and obviously no HPV16 enrichment had occurred. Therefore, it was concluded that the combination of GPUA and HPV16 primer is not suitable for PCR amplification of HPV16-containing DNA directly from the GP-WGA library.

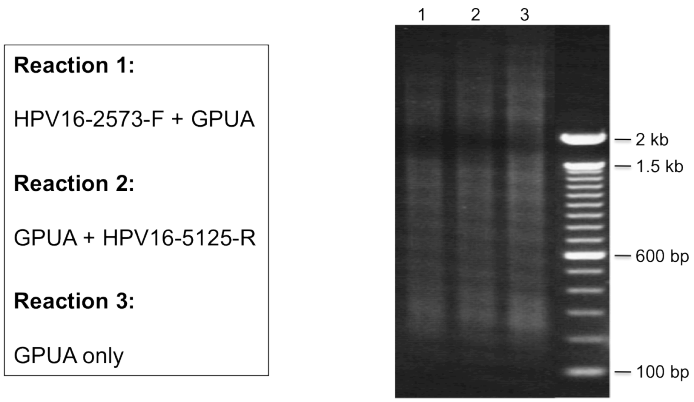


Figure 2-32: PCR for direct amplification of HPV16 DNA from the GP-WGA library. Primer combinations and gel analysis of three PCR reactions with MRI-H186 WGA product as template are shown. The 100 bp DNA Ladder was used as size marker.

Since the direct PCR amplification using GPUA together with HPV16 primer had failed, a more efficient approach for HPV16 DNA selection from GP-WGA libraries was required. As alternative, it was attempted to enrich HPV16 sequences by primer extension using 5'-biotinylated HPV16-specific primers (*Figure 2-33*). The resultant single-stranded DNA should be HPV16-positive and contain GPUA only at the 3' end. The single-stranded DNA products were purified from template DNA (GP-WGA library) after primer extension, using the biotin-streptavidin selection system with streptavidin-coated BioMag magnetic beads (see *Section 4.35.2*). In the next step, the single-stranded intermediates were converted into double-stranded DNA by PCR using GPUA and nested HPV16 primer in combination.

Taken together, the procedure designed to select HPV16-containing DNA from GP-WGA libraries is composed of linear amplification by primer extension, single-stranded DNA selection by biotin-streptavidin, and double-stranded DNA synthesis by PCR (*Figure 2-33*). This procedure should strongly reduce the genomic DNA background, and the final products are supposed to contain pure HPV16 DNA and HPV16-cellular junction DNA (see *Figure 2-30*).

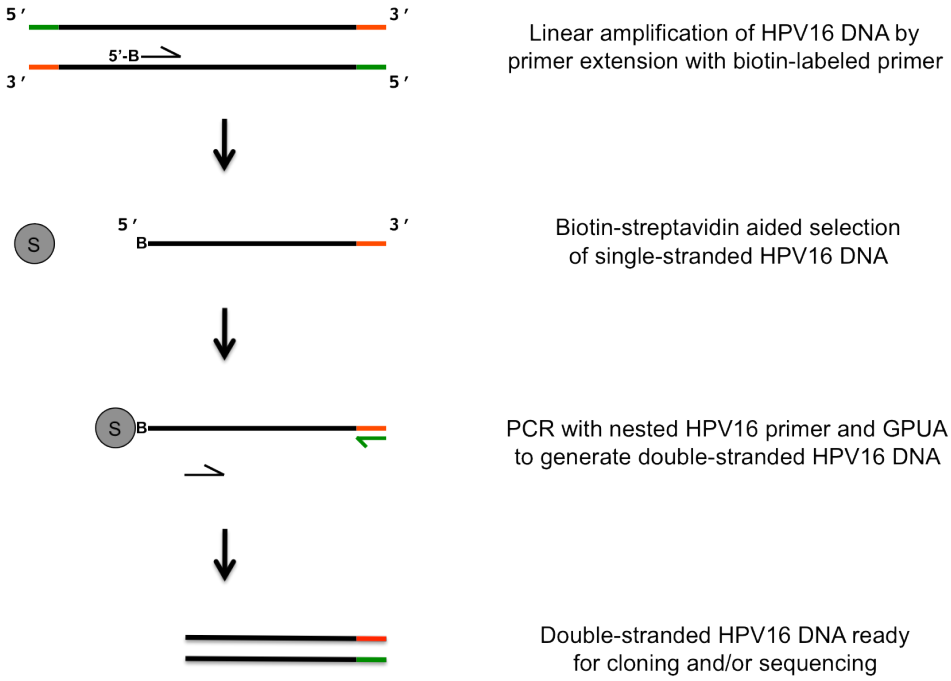


Figure 2-33: HPV16 DNA enrichment from the GP-WGA library. The GPUA and its complementary strand are shown in green and red, respectively. Positions of HPV16 primers are marked by black half-arrows. (S=streptavidin-coated magnetic bead; B=biotin; 5'-B=biotin labelled at the 5' end of HPV16 primer).

2.10.3 ASP16 Step 3 – Pyrosequencing

After HPV16 DNA enrichment, the next challenge was an effective way of DNA sequencing. Two aspects had to be taken into consideration. Firstly, HPV16 DNA breakpoints could be located anywhere within E1/E2 and L1/L2 regions, while the average size of GP-WGA products is only about 400 bp. Therefore, in order to cover all possible locations of viral-cellular junctions, multiple HPV16 primers with appropriate intervals have to be used. Secondly, clinical DNA samples contain not only integrated HPV16 DNA but also episomal HPV16 DNA to different extents. The two inherent problems are not well compatible with classical cloning and Sanger sequencing, which are time-consuming, laborious and expensive. In these aspects, the so-called "next-generation" sequencing technologies became more promising.

In 2006 and 2007, three next-generation sequencing systems were introduced, including *Genome Sequencer FLX* (Roche/454), *Genome Analyzer* (Illumina/Solexa) and *SOLiD* (Applied Biosystems) (Mardis, 2008; Shendure and Ji, 2008). A common strength of them is the massively parallel sequencing, which can provide huge amounts of sequence data in a single machine-run within hours. The Genome Sequencer FLX (GS-FLX) was chosen as a platform for the novel ASP16 strategy.

An essential aspect for this decision was the read length of about 250 bases, while the read lengths of the other systems are below 100 bases (Mardis, 2008; Shendure and Ji, 2008). The GS-FLX is mainly based on two principal techniques: emulsion PCR (emPCR) and pyrosequencing (Margulies *et al*., 2005). The emPCR precedes pyrosequencing, and it is used to amplify the template DNA for signal augmentation. Each DNA molecule bound to a magnetic bead is encapsulated into an emulsion sphere and clonally amplified to millions of identical copies (*Figure 2-34A*). The size of template DNA for emPCR is limited to about 400 bp, which fits the average length of GP-WGA products. The beads are deposited into a picotiter-plate for sequencing. Unlike Sanger sequencing, the pyrosequencing is a method featured on "*sequencing by synthesis"*. The pyrophosphates generated by DNA polymerization are quantitatively converted into light signals, without disrupting DNA elongation (*Figure 2-34B*). This technique makes it possible to sequence every single DNA molecule within a mixture of templates individually and simultaneously. The final effect of the whole procedure is: one DNA molecule = one bead = one sequence read.

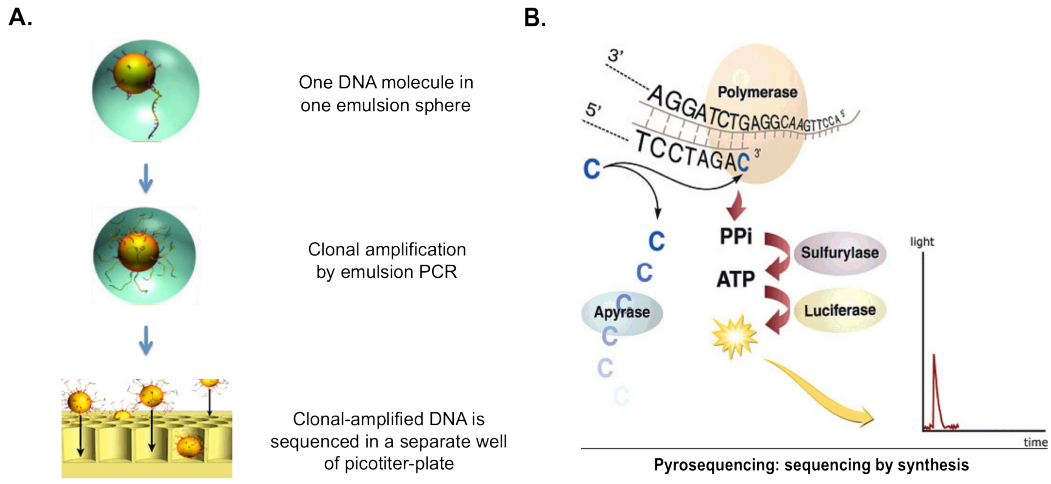


Figure 2-34: Emulsion PCR and pyrosequencing. Panel **A** describes the workflow of emulsion PCR. The figure was taken from http://www.454.com. Panel **B** shows the principle of pyrosequencing. The figure was taken from http://www.pyrosequencing.com.

In order to bind DNA onto magnetic beads in emPCR and to sequence PCR amplicons with GS-FLX, primers used in the nested-PCR step of HPV16 DNA enrichment (see *Figure 2-33*) have to be equipped with *Roche-A* and *Roche-B* adapter sequences. For ASP16 experiments, Roche-A was fused to the nested HPV16 primers, while Roche-B was coupled with GPUA (see *Section 4.12.5*). To utilize the GS-FLX system at its full capacity, a blend of multiple DNA samples can be sequenced simultaneously in one machine-run. To apply this possibility, each

DNA sample needs to be equipped with a molecular tag of unique character to differentiate between samples. Totally 24 distinct barcodes with a length of four nucleotides each were designed to represent different DNA samples in the ASP16 experiments (*see Section 4.12.5*). The 24 barcodes were individually joined with the GPUA primer, which is common to all nested PCR reactions.

In summary, the multiplex nested PCR in the HPV16 selection procedure was made compatible with sample pooling and GS-FLX pyrosequencing by a special primer design. All nested HPV16 primers have a bipartite structure carrying the *Roche-A* sequence at the 5' end. The GPUA primers have a tripartite structure composed of *Roche-B*, *4nt-barcode* and *GPUA* in 5' to 3' direction (see *Figure 2-35*). The whole experimental procedure of the ASP16 strategy is outlined in *Figure 2-35*.

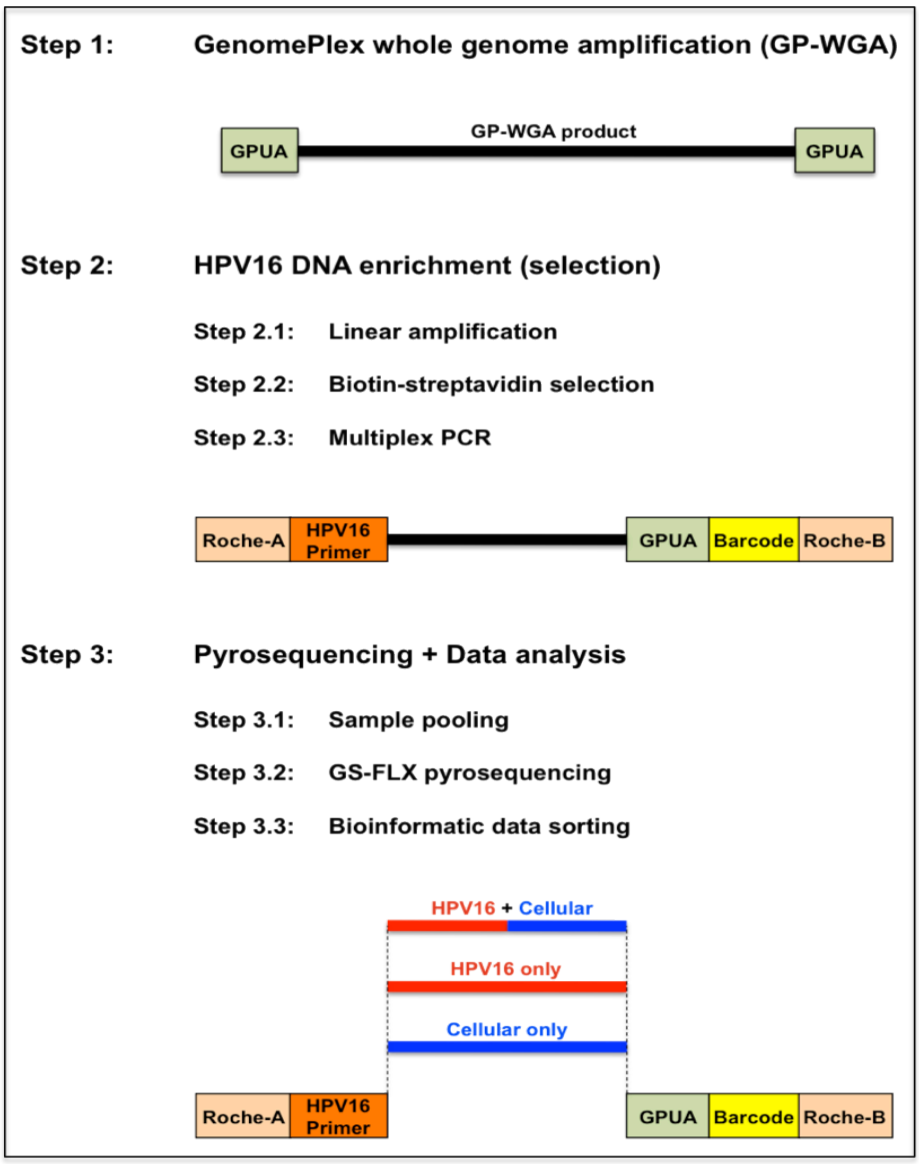


Figure 2-35: Summary of the amplification-selection-pyrosequencing strategy for HPV16 integration analysis (ASP16). The details are explained in the text.

2.11 The first ASP16 experiment (ASP16-1)

To test the whole concept of ASP16 for the first time, a set of 28 HPV16-positive DNA samples was prepared for pyrosequencing. This included four cervical carcinoma cell lines, three cytologically normal clinical samples and 21 cervical pre-cancerous lesions (*Table 2-36*).

Table 2-36: Information about DNA samples used in experiment ASP16-1.

Serial Number	Sample Name	Lesion Grade [a]	Sample ID (Barcode)
1	04-1921	HGSIL	*
2	00-1260	HGSIL	1B2
3	05-630	LGSIL	1B3
4	05-91	ASCUS	*
5	04-1898	normal	*
6	04-2019	normal	1B6
7	04-2277	normal	*
8	MRI-H186	cervical carcinoma cell line	1B8
9	MRI-H196	cervical carcinoma cell line	1B9
10	CaSki	cervical carcinoma cell line	1B10
11	SiHa	cervical carcinoma cell line	1B11
12	38953	CIN3	1B12
13	39061	CIN3	1B13
14	41962	CIN3	1B14
15	43317	CIN3	1B15
16	44495	CIN3	1B16
17	44969	CIN3	1B17
18	46116	CIN3	1B18
19	46168	CIN3	1B19
20	46645	CIN3	1B20
21	55246	CIN3	1B21
22	55602	CIN3	1B22
23	55619	CIN3	1B23
24	47630	CIN3	1B24
25	47664	CIN3	1B1
26	47898	CIN3	1B4
27	49136	CIN3	1B5
28	49231	CIN3	1B7

[a] HGSIL=high-grade squamous intraepithelial lesion; LGSIL=low-grade squamous intraepithelial lesion; ASCUS=atypical squamous cells of undetermined significance; CIN3=cervical intraepithelial neoplasia grade 3.

* Four samples were not included in the DNA mixture for pyrosequencing, due to their poor performance in the HPV16 DNA enrichment procedure, see *Figure 2-39*.

Primer design for HPV16 early and late regions

In the first ASP16 experiment (ASP16-1), it was planned to obtain both the 5' and 3' viral-cellular junction sequences of each DNA sample simultaneously. As described earlier, HPV16 DNA enrichment requires multiple HPV16 primers to cover the E1/E2 and L1/L2 regions completely. Therefore, 16 biotin-labelled HPV16 primers, eight (EP-1 to EP-8) for E1/E2 and eight (LP-1 to LP-8) for L1/L2, were designed for linear amplification (*Figure 2-37* and *Table 2-38*). Similarly, 16 nested HPV16 primers (EP-N1 to EP-N8; LP-N1 to LP-N8) were used for multiplex PCR (*Figure 2-37* and *Table 2-38*). Sequences of the primers are shown in *Section 4.12.5*. The primer interval was about 400 bp, which is in accordance with the average length of GP-WGA product and the size limit of template DNA for emPCR.

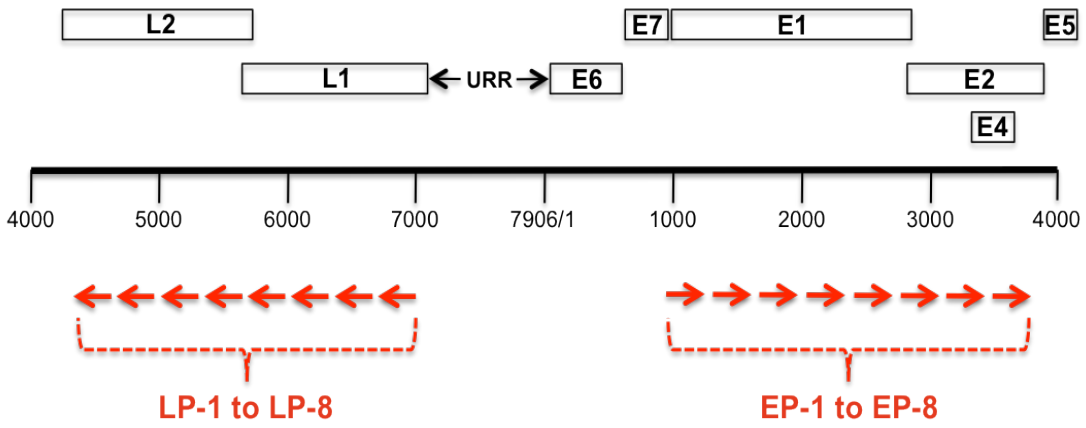


Figure 2-37: Primers used in experiment ASP16-1 to cover the early and late regions of integrated HPV16 DNA. Locations of 16 biotin-labelled HPV16 primers (LP-1 to LP-8; EP-1 to EP-8) for linear amplification are schematically demonstrated. The nested HPV16 primers for multiplex PCR are not shown. The exact positions of HPV16 primers are shown in *Table 2-38*. Primer sequences are given in *Section 4.12.5*. (LP=primer in the HPV16 late region; EP=primer in the HPV16 early region)

Table 2-38: Positions of HPV16 primers used in experiment ASP16-1.

ORF	Positions	Biotinylated Primer	Positions	Nested Primer	Positions
E1	865 to 2814	**EP-1**	810→831	**EP-N1**	857→877
		EP-2	1261→1282	**EP-N2**	1275→1296
		EP-3	1562→1582	**EP-N3**	1576→1597
		EP-4	1938→1959	**EP-N4**	1951→1972
		EP-5	2389→2411	**EP-N5**	2403→2424
		EP-6	2705→2726	**EP-N6**	2723→2742
E2	2756 to 3853	**EP-7**	3101→3122	**EP-N7**	3121→3143
		EP-8	3542→3563	**EP-N8**	3555→3576
L1	5561 to 7156	**LP-1**	6929←6951	**LP-N1**	6899←6918
		LP-2	6531←6552	**LP-N2**	6483←6505
		LP-3	6165←6186	**LP-N3**	6120←6139
		LP-4	5744←5766	**LP-N4**	5710←5729
L2	4237 to 5658	**LP-5**	5296←5318	**LP-N5**	5285←5304
		LP-6	4898←4917	**LP-N6**	4871←4890
		LP-7	4478←4501	**LP-N7**	4445←4466
		LP-8	4271←4290	**LP-N8**	4235←4254

The positions of totally 32 HPV16 primers (16 biotinylated primers and 16 nested primers) as well as the HPV16 ORFs involved are shown. Overall, the distance between two neighbouring primers is about 400 bp. (EP=early primer; EP-N=nested early primer; LP=late primer; LP-N=nested late primer)

Whole genome amplification and HPV16 DNA enrichment

At first, the GP-WGA was performed with 24 (serial number 1 to 24) out of the 28 DNA samples, respectively. For each amplified DNA sample, two linear amplifications (E and L) as well as two subsequent multiplex-PCR reactions (E and L) for the HPV16 early and late regions were carried out individually. In each single reaction, eight HPV16 biotinylated primers (for linear amplification) or nested primers (for PCR) were used in a multiplex way. Aliquots of all 48 multiplex-PCR products were checked on gel and hybridized with the HPV16-complete DNA probe (*Figure 2-39*). Staining intensities on gel and signal intensities on Southern blot showed large differences among samples. In particular for samples no. 1 to 7, only primer dimers were observed on the gel, and almost no HPV16 hybridization signals were detected on Southern blot (*Figure 2-39*, upper part). This was probably due to the poor DNA

quality of the seven samples, because the GP-WGA products of them were obviously shorter than other samples (data not shown). Overall, Southern hybridization correlated well with gel analysis in respect of signal intensity, which indicated the success of HPV16 DNA enrichment. Furthermore, signal intensities on Southern blot seem to correlate with HPV16 DNA copy number in the cell. For example, there is no difference in DNA quality among the three cell lines (no. 8 to 11), but the hybridization intensities were clearly different among them. The strongest signals came from CaSki (no. 10), which contains about 500 copies of HPV16 DNA per cell (Yee *et al*., 1985; Baker *et al*., 1987). By contrast, very weak signals were found in SiHa (no. 11), which harbors only 1 – 2 copy of integrated HPV16 DNA per cell (Yee *et al*., 1985; Baker *et al*., 1987).

Because of the poor product amounts and hybridization intensities, the samples no. 1, 4, 5 and 7 were eliminated from analysis and replaced by samples no. 25 to 28 (see *Table 2-36*). Thus, altogether 24 samples (sample-IDs 1B1 to 1B24) containing 48 multiplex-PCR reactions were purified separately and subjected to further analysis.

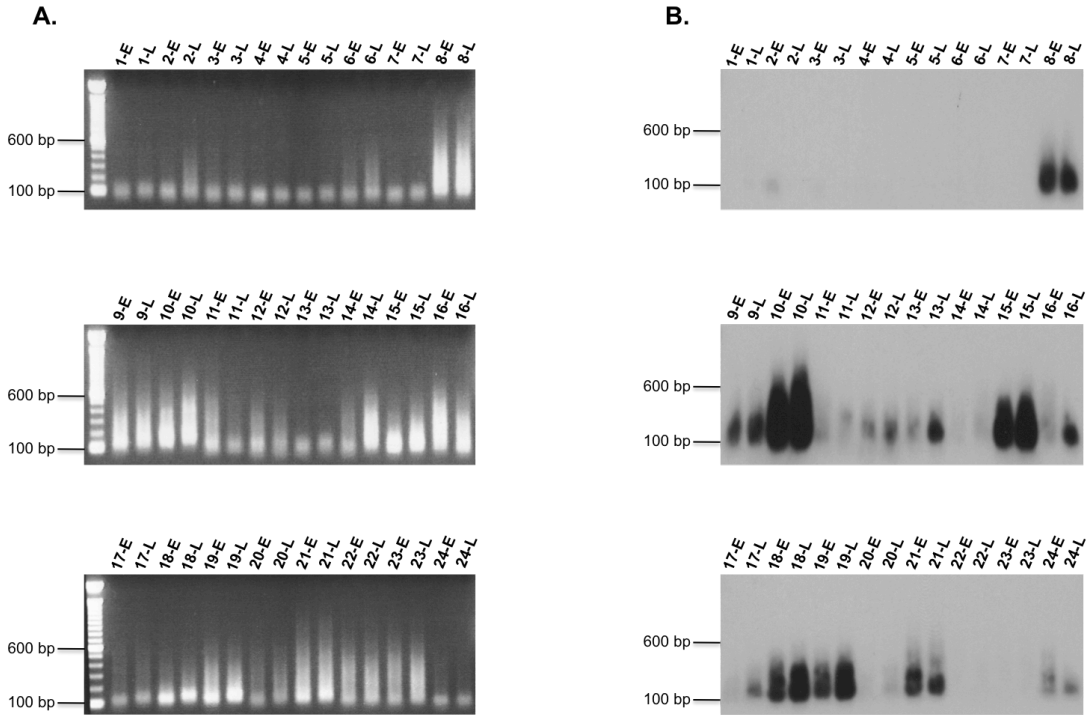


Figure 2-39: Gel analysis and Southern hybridization of final products after HPV16 DNA enrichment. Panel **A** shows the 48 final products of HPV16 DNA enrichment (for HPV16 **E**arly and **L**ate regions) from 24 DNA samples (serial number 1 to 24). The final products of samples no. 25 to 28 are not shown. For each reaction, 3 µl of PCR product was checked on agarose gel prior to DNA purification. The 100 bp DNA Ladder was used as size marker. Gel was blotted and hybridized with the HPV16-complete DNA probe (panel **B**).

The concentration of purified multiplex-PCR products was between 13 and 50 ng/μl. For pyrosequencing, 100 ng of each purified DNA from altogether 48 reactions were mixed together. DNA concentration of this mixture was determined as 21 ng/μl. The DNA mixture was checked on agarose gel, and appeared to be a smear from 100 bp to over 600 bp with a peak at the position about 150 bp (*Figure 2-40*).

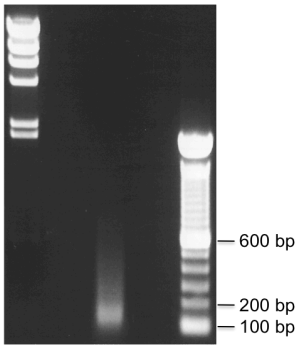


Figure 2-40: Gel analysis of the DNA mixture for pyrosequencing in ASP16-1. About 120 ng (6 μl) of the DNA mixture for pyrosequencing was checked on agarose gel. The λ/*Hind* III (leftmost) and 100 bp DNA Ladder (rightmost) were used as size marker.

Pyrosequencing

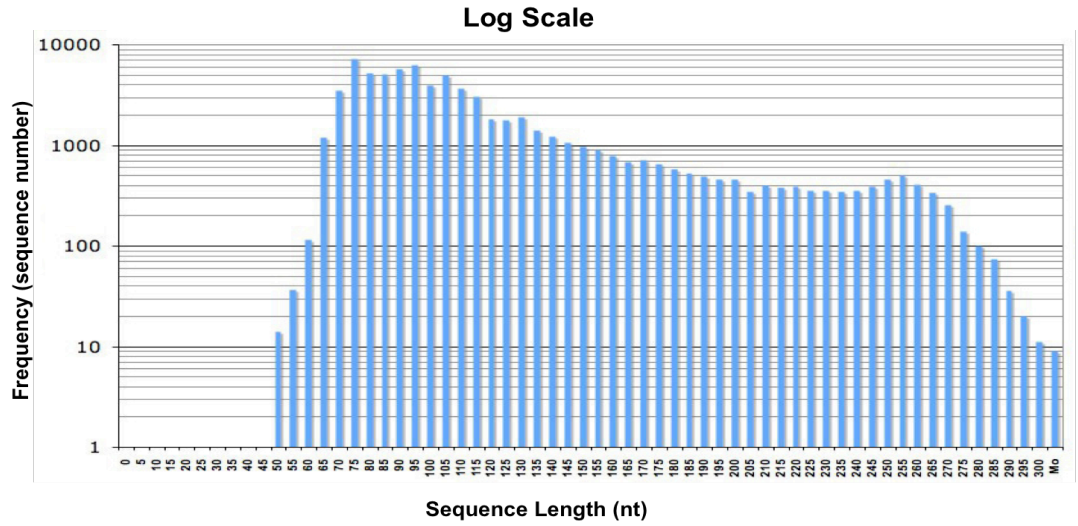
The emPCR amplification and GS-FLX pyrosequencing (see *Figure 2-34*) of the DNA sample mixture were performed by the DKFZ Genomics and Proteomics Core Facility. Sequencing can be primed with either *Roche-A* or *Roche-B*. To make sure that every sequence read has a barcode, Roche-B was chosen as sequencing primer for the ASP16 experiments, because the barcodes are located directly downstream of Roche-B. Dependent on experimental requirements, three formats of picotiter-plates are available as sequencing devices (*Figure 2-41*). Each plate has a unique capacity, and can be further divided into physically separated small regions. For the experiment ASP16-1, the device 25x75 with one medium region was applied.

Sequencing Kit	PicoTiterPlate Device	Number of Regions per Gasket	Read Length (bases)	Number of Reads per Region	Number of Bases per Region	Total Number of Reads per Gasket	Total Number of Bases (all regions)	
GS LR70	70x75	2 large regions	240	210,000	50.4 Mb	420,000	100.8 Mb	*
		4 medium regions	240	70,000	16.8 Mb	280,000	67.2 Mb	
		16 small regions	240	12,000	2.88 Mb	192,000	46.08 Mb	
GS LR25	25x75	1 medium region	240	70,000	16.8 Mb	70,000	16.8 Mb	
		4 small regions	240	12,000	2.88 Mb	48,000	11.52 Mb	
GS SR70	70x75	2 large regions	100	210,000	21 Mb	420,000	42 Mb	
		4 medium regions	100	70,000	7 Mb	280,000	28 Mb	
		16 small regions	100	12,000	1.2 Mb	192,000	19.2 Mb	

Figure 2-41: Sequencing devices available for different applications with GS-FLX. The picotiter-plate marked in red was used for the experiment ASP16-1 (*used for ASP16-2). The figure was taken from http://www.454.com.

Altogether 73283 sequence reads were obtained from ASP16-1. The entire data set amounted to 8.3 Mb, which did not reach the capacity of 16.8 Mb (see *Figure 2-41*). The average read length (Roche-B primer of 19 nt omitted) was 113 nt, which is a rather low value in comparison with the capacity of 250 nt. The average read length plus Roche-B (113+19=132 nt) agreed with the major DNA size observed on gel (see *Figure 2-40*). The overall distribution of sequence length is shown in *Figure 2-42*. The log-scale histogram shows that sequence reads from 50 to 300 nt were obtained by GS-FLX pyrosequencing in ASP16-1. Most sequences fell into the range between 65 and 145 nt, as demonstrated by the linear-scale histogram.

A.



B.

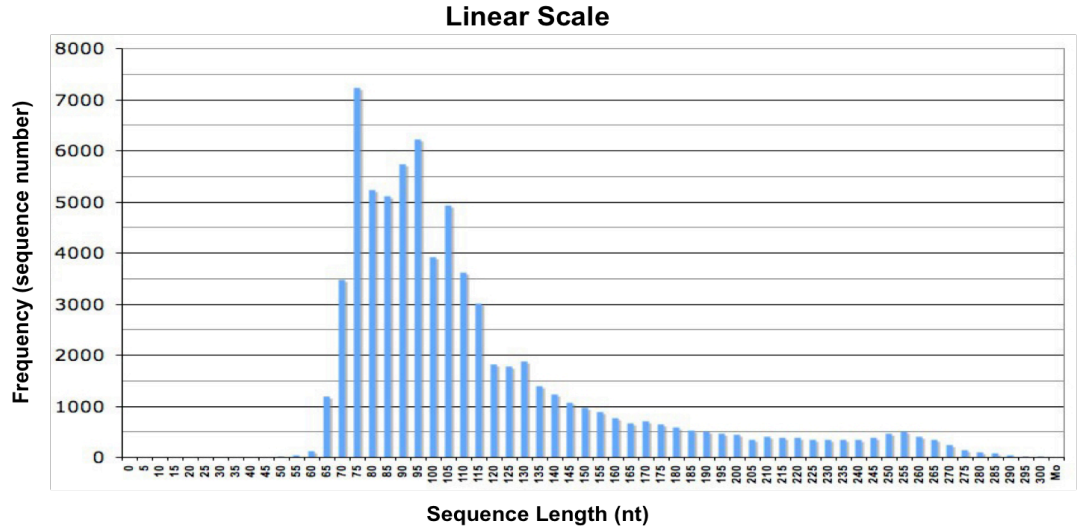


Figure 2-42: Overall length distribution of sequence reads from experiment ASP16-1. The X-axis represents increasing sequence lengths in 5-nt interval. The Y-axis shows the frequency (sequence number) of the corresponding read length. The length distribution is demonstrated in log scale (panel **A**) and linear scale (panel **B**). (Mo = more)

Bioinformatics data analysis

The pyrosequence reads were analysed with bioinformatics tools developed by Sasithorn Chotewutmontri in our laboratory (*Figure 2-43*). The whole data set of ASP16-1 was first sorted into 24 sample categories, according to the 4-nt barcodes. For each sample, sequence reads were individually compared with the HPV16 genome reference using the "BlastN" algorithm (≥11nt) from NCBI. A cut-off value of 28 nt was used to distinguish significant HPV16 sequences (≥28nt) from unspecific sequences caused by false priming. The significant HPV16 sequences were then classified into four groups by length of the non-HPV16 part. Sequences from group 1 have the longest non-HPV16 part, and are therefore the strongest candidates for viral-cellular junctions. Sequences from groups 1 to 3 were sequentially searched for matches in the human genome reference databases from NCBI using the "MegaBlast" algorithm (≥28nt). Sequences from group 4 were not analyzed, because the non-HPV16 parts are too short for significant matches.

Besides determining junction sequences, the programs were also applied to evaluate the specificity of individual nested HPV16 primers in ASP16 experiments. From each sample category, sequences were further sorted by the nested HPV16 primer sequences into 16 different groups. Within each primer group, number of significant HPV16 sequences *versus* total sequence number was compared.

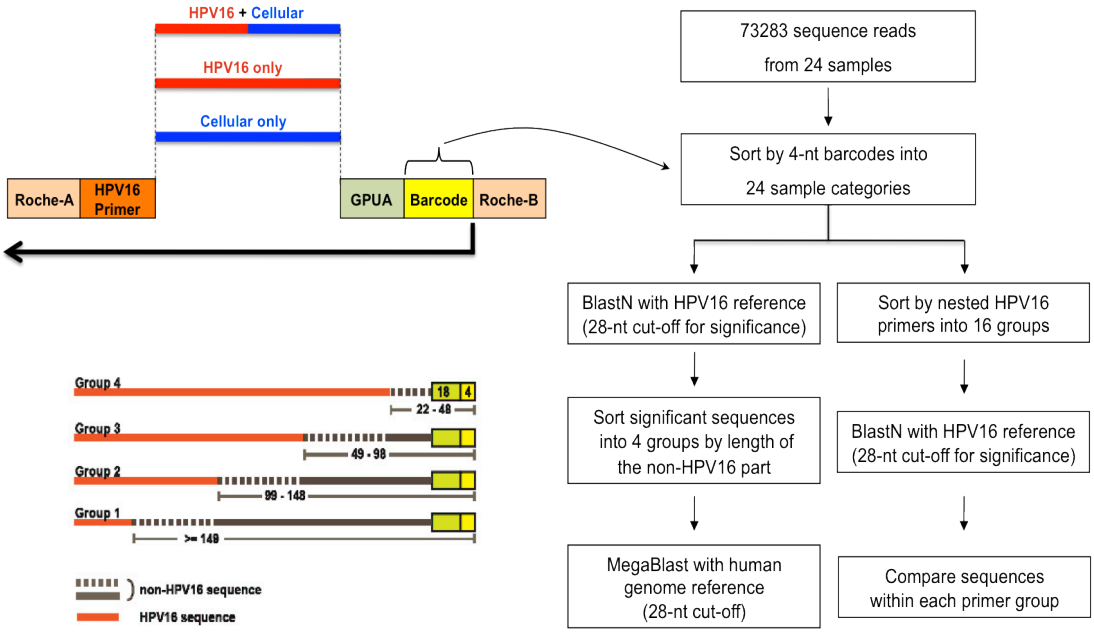
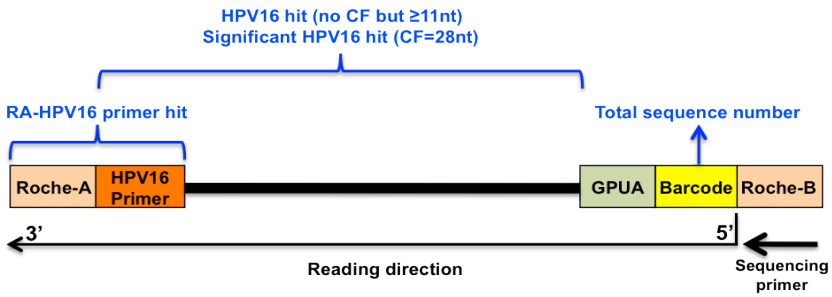


Figure 2-43: Bioinformatics data analysis of sequence reads obtained from ASP16-1. The individual steps are explained in the text.

To check the differences in sequence distribution among the 24 samples used for pyrosequencing, sequence numbers were calculated for each sample with four different filter criteria: *total sequence number*, *HPV16 hit without cut-off*, *RA-HPV16 primer hit*, and *HPV16 hit with 28-nt cut-off* (*Figure 2-44A*). The *total sequence number* refers to the number of sequence reads assigned to a given sample by the barcode, but without any data filtering. The *HPV16 hit without cut-off* represents those sequence reads, which have a match of at least 11 nt with the HPV16 genome reference. The *RA-HPV16 primer hit* specifies the sequence reads carrying Roche-A (RA) plus HPV16 primer at the 3' ends, which means they were sequenced completely. The *HPV16 hit with 28-nt cut-off* defines the significant HPV16 sequences, which contain at least 28 nt of HPV16 sequences.

As shown in *Figure 2-44B and C*, large differences in the *total sequence number* were observed among samples. For cell line CaSki (sample-ID 1B10), which harbors about 500 copies of HPV16 DNA per cell, 5895 sequence reads were obtained (the highest number in ASP16-1). Another cell line SiHa (sample-ID 1B11) with 1 – 2 copy of HPV16 DNA per cell, contained 1750 sequence reads (a rather low number). The lowest sequence number 599 came from clinical sample 1B3, which had shown low yield in GP-WGA (data not shown) and no HPV16 hybridization signal (see *Figure 2-39*), probably due to the poor quality of original DNA. For each sample, there was no clear difference between *total sequence number*, *HPV16 hits without cut-off* and *RA-HPV16 primer hits*. This indicated that most sequence reads from ASP16-1 had been fully sequenced, and also agreed with the short read length (113 nt on average). The performance of HPV16 DNA enrichment, reflected by the proportion of *HPV16 hits with 28-nt cut-off* to *total sequence number*, was different among samples. This might be caused by sample variations for example the poor DNA quality, but the exact reason still remains unclear. Nevertheless, HPV16-containing DNA has been specifically enriched in many of the 24 samples, which suggested that overall the whole concept for HPV16 DNA enrichment works.

A.



B.

Sample ID (Barcode)	Significant HPV16 hits (CF=28nt)	RA-HPV16 primer hits	HPV16 hits (no CF)	Total sequence number
1B3	78	568	585	599
1B6	2	1095	1145	1174
1B1	412	1173	1233	1253
1B13	871	1307	1350	1354
1B2	285	1287	1351	1373
1B17	689	1516	1611	1638
1B11	280	1606	1700	1750
1B4	570	1700	1770	1808
1B24	948	1785	1862	1877
1B5	1440	2065	2130	2145
1B12	575	2072	2149	2203
1B20	179	2052	2153	2204
1B14	101	2474	2587	2662
1B22	12	2866	3040	3133
1B23	20	2877	3054	3167
1B9	786	3443	3573	3656
1B8	661	3603	3736	3799
1B16	371	3666	3805	3901
1B21	1239	3999	4138	4190
1B7	4036	4928	5099	5106
1B18	4158	5214	5414	5439
1B15	3709	5455	5624	5663
1B19	3203	5568	5772	5844
1B10	4526	5721	5880	5895

C.

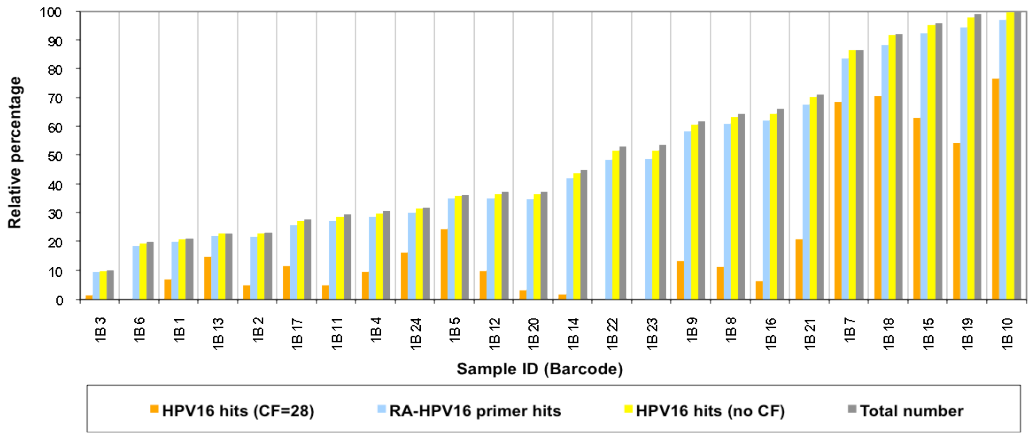


Figure 2-44: Sequence distribution of the 24 DNA samples in experiment ASP16-1. The general structure of the single sequence reads primed by Roche-B is shown in panel **A**. The absolute sequence numbers (panel **B**) and the relative numbers in percentage (panel **C**) compared to the total number of CaSki (1B10) are shown for each sample with four different filter criteria explained in the text. (RA=Roche-A; CF=cut-off; GPUA=GenomePlex universal adapter)

Identification of viral-cellular junctions

For cell line MRI-H186 (sample-ID 1B8), altogether nine pyrosequence reads were identified in experiment ASP16-1, which represent the already known 3' viral-cellular junction at HPV16-2754/chr8-128815788 (see *Section 2.8.5*). For the other three cell lines MRI-H196 (1B9), CaSki (1B10) and SiHa (1B11), however, no viral-cellular junction sequence was found in experiment ASP16-1. From the clinical samples, five potential junction sequences representing four unique viral-cellular junctions were obtained (*Table 2-45*). However, all the four junctions turned out to be false-positives, as determined by junction-specific PCR using an HPV16 primer together with a cellular primer (data not shown).

Table 2-45: Potential viral-cellular junction sequences of clinical DNA samples obtained in experiment ASP16-1.

Sequence ID	Sample ID	Sequence Read Length	Length of HPV16 part	HPV16 DNA Breakpoint	Chromosome
058833	1B7	102	33	3587	20
065134	1B7	102	33	3587	20
040959	1B8	109	36	3156	9
103364	1B8	112	35	3155	3
070687	1B17	113	36	1611	7

As the most important aspect, experiment ASP16-1 proved the concept of the ASP strategy, regarding the specificity of HPV16 sequence enrichment from whole-genome DNA samples and the capability to identify viral-cellular junction sequences. The key problem revealed by ASP16-1 was the short read length (113 nt on average) *versus* long intervals between HPV16 primers (about 400 bp). This means that only about one-fourth of the HPV16 early and late regions could be sequenced under this experimental condition. For a more comprehensive coverage of HPV16 DNA, it was required either to reduce primer interval by introducing new primers, or to increase the average read length by experimental optimizations.

2.12 The second ASP16 experiment (ASP16-2)

The most important modifications in the second experiment ASP16-2 were made in the HPV16 primer design, and in the decision to confine the integration analysis to the 3' junction in the HPV16 early region. This decision was made because the 3' junction downstream of the E6 and E7 oncogenes is more important for cervical carcinogenesis than the 5' junction (see *Figure 1-5*). It was intended to reach complete sequence coverage of the E1/E2 region in the pyrosequencing step, thereby enhancing the probability to determine viral-cellular junction sequences. To shorten the intervals between the primers, eight new biotin-labeled HPV16 primers (EP-9 to EP-16) for linear amplification as well as eight new nested HPV16 primers (EP-N9 to EP-N16) for multiplex PCR were added into experiment ASP16-2. By doing this, the primer interval was overall decreased from 400 bp to 200 bp in the HPV16 E1/E2 region (*Figure 2-46 and Table 2-47*).

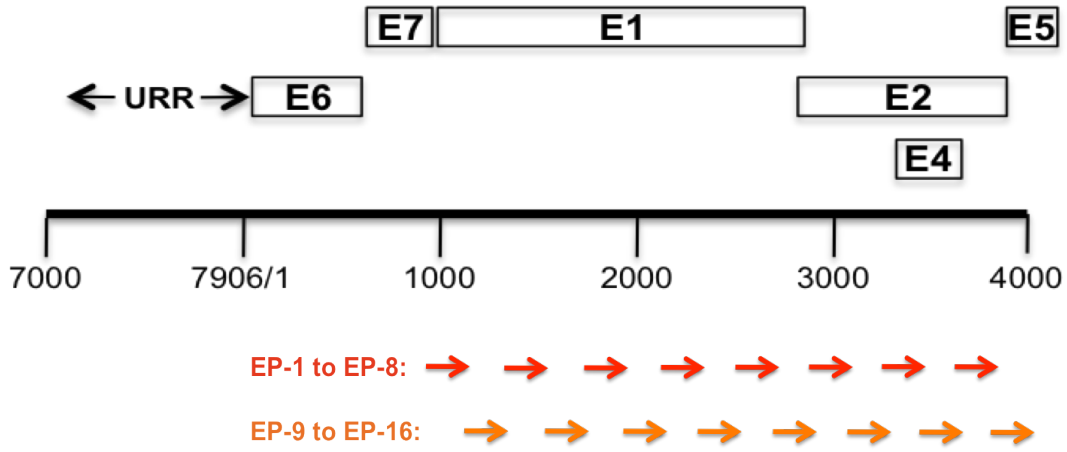


Figure 2-46: Primers used in experiment ASP16-2 to cover the E1/E2 region of HPV16 DNA. The locations of 16 biotin-labelled HPV16 primers (EP-1 to EP-16) for linear amplification are schematically demonstrated. The nested HPV16 primers for multiplex PCR are not shown. The exact positions of HPV16 primers are shown in *Table 2-47*. Primer sequences are given in *Section 4.12.5*. (EP=primer in the HPV16 early region)

Table 2-47: Positions of HPV16 primers used in experiment ASP16-2.

ORF	Positions	Biotin Primer	Positions	Nested Primer	Positions
E1	865 to 2814	EP-1	810→831	EP-N1	857→877
		EP-9	1046→1066	EP-N9	1064→1084
		EP-2	1261→1282	EP-N2	1275→1296
		EP-10	1394→1416	EP-N10	1411→1433
		EP-3	1562→1582	EP-N3	1576→1597
		EP-11	1760→1782	EP-N11	1785→1805
		EP-4	1938→1959	EP-N4	1951→1972
		EP-12	2151→2173	EP-N12	2174→2196
		EP-5	2389→2411	EP-N5	2403→2424
		EP-13	2539→2561	EP-N13	2569→2591
		EP-6	2705→2726	EP-N6	2723→2742
E2	2756 to 3853	EP-14	2912→2933	EP-N14	2933→2953
		EP-7	3101→3122	EP-N7	3121→3143
		EP-15	3318→3339	EP-N15	3339→3361
		EP-8	3542→3563	EP-N8	3555→3576
		EP-16	3762→3783	EP-N16	3778→3798

A collection of 21 HPV16-positive DNA samples was chosen for experiment ASP16-2 (*Table 2-48*). The three cervical carcinoma cell lines MRI-H186 (sample-ID 2B1), MRI-H196 (2B2) and SiHa (2B3) had already been included in ASP16-1. The 18 clinical DNA samples (sample-IDs 2B4 to 2B21) consisted of six cervical carcinomas and twelve high-grade pre-cancerous lesions (mostly classified as CIN3).

Table 2-48: Information about the DNA samples used in experiment ASP16-2.

Sample ID (Barcode)	Sample Name	Lesion Grade [a]	HPV16 Early Primer Mix (EPM) used for pyrosequencing
2B1	MRI-H186	cervical carcinoma cell line	EPM-1 + EPM-2
2B2	MRI-H196	cervical carcinoma cell line	EPM-1 + EPM-2
2B3	SiHa	cervical carcinoma cell line	EPM-1
2B4	07c134	cervical carcinoma	EPM-1 + EPM-2
2B5	07c368	cervical carcinoma	EPM-1 + EPM-2
2B6	07c381	cervical carcinoma	EPM-1
2B7	07c715	cervical carcinoma	EPM-1
2B8	07c716	cervical carcinoma	EPM-1
2B9	07c719	cervical carcinoma	*
2B10	61483	CIN3	EPM-1
2B11	62973	CIN3	EPM-1
2B12	63179	CIN3	EPM-1
2B13	63282	CIN3	EPM-1
2B14	64655	CIN3	EPM-1
2B15	64702	CIN3	EPM-1 + EPM-2
2B16	65468	CIN3	EPM-1 + EPM-2
2B17	65892	CIN3	EPM-1
2B18	66019	CIN3	*
2B19	67230	CIN3	EPM-1
2B20	d1	not determined	EPM-1
2B21	d2	not determined	EPM-1

[a] CIN3 = cervical intraepithelial neoplasia grade 3

* Due to weak or absent HPV16 hybridization signals (see *Figure 2-49B*), neither EPM-1 nor EPM-2 reactions were included in the pyrosequencing step of experiment ASP16-2.

In experiment ASP16-2, the GP-WGA and HPV16 DNA enrichment (including linear amplification and multiplex PCR) were prepared separately for all 21 DNA samples. For each sample, two linear amplifications and two multiplex PCR reactions were performed with different HPV16 primer mixtures. The early primer mix 1 (EPM-1) contained biotinylated primers EP-1 to EP-8 for linear amplification or nested primers EP-N1 to EP-N8 for multiplex PCR. The EPM-2 contained EP-9 to EP-16 for linear amplification or EP-N9 to EP-N16 for multiplex PCR. The primer positions in the HPV16 E1/E2 region are shown in *Figure 2-46 and Table 2-47*. The final products of

totally 42 reactions were purified individually. From each reaction, an aliquot of purified DNA was checked on gel and analyzed by Southern hybridization with the HPV16-complete DNA probe (*Figure 2-49*). The gel analysis showed clear differences between reactions with EPM-1 and EPM-2. EPM-1 gave rise to DNA smears from about 100 bp to 600 bp, whereas EPM-2 has mainly produced short products of about 100 bp, which most likely represent primer dimers. On Southern blot the majority of EPM-1 reactions showed strong HPV16 hybridization signals. By contrast, only 9 out of 21 EPM-2 reactions hybridized with the HPV16 probe. Finally, altogether 25 reactions including 19 EPM-1 and six EPM-2 reactions were pooled for pyrosequencing (*Table 2-48* and *Figure 2-49*).

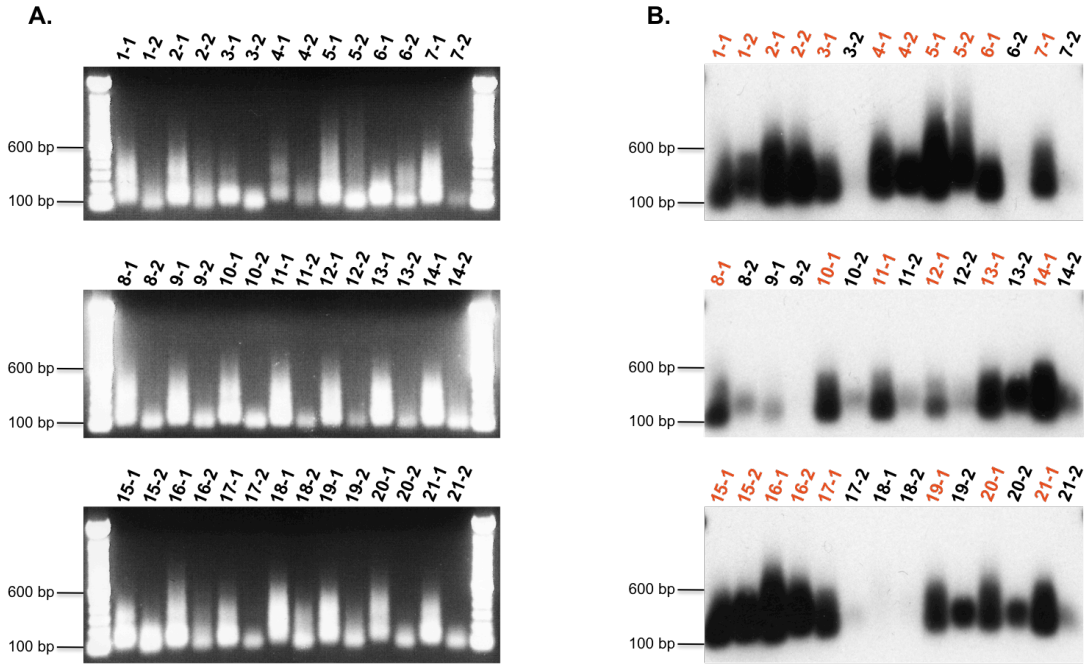


Figure 2-49: Gel analysis and Southern hybridization of DNA samples after HPV16 DNA enrichment in experiment ASP16-2. Panel **A** shows the 42 final products after HPV16 DNA enrichment from all 21 DNA samples. The first number refers to the sample ID (see *Table 2-48*), and the second number indicates EPM-1 and EPM-2, respectively. For each reaction, about 100 ng of purified PCR product was checked on agarose gel. The 100 bp DNA Ladder was used as size marker. The gel was blotted and hybridized with the HPV16-complete DNA probe (panel **B**). Reactions marked in red were mixed for pyrosequencing.

The final concentration of the pooled DNA was determined as 19 ng/µl. As shown in *Figure 2-50*, the majority of DNA was found in the size range between 100 and 200 bp, with an additional smear up to about 500 bp. In order to get more sequence reads, the pyrosequencing for ASP16-2 was performed using one large region in the 70x75 picotiter-plate (see *Figure 2-41*).

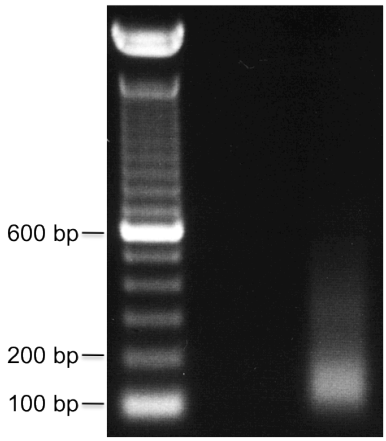


Figure 2-50: Gel analysis of the DNA mixture for pyrosequencing in experiment ASP16-2. 150 ng of the DNA mixture for pyrosequencing was checked on agarose gel. The 100 bp DNA Ladder was used as size marker.

The pyrosequencing outcome of experiment ASP16-2 is summarized in *Table 2-51*, in comparison with the capacity of the sequencing device applied. Altogether 223942 sequence reads were obtained in ASP16-2, and the average read length was 88 nt.

Table 2-51: Summary of the sequence data obtained from experiment ASP16-2.

	Picotiter-plate 70×75 (one large region)	ASP16-2 outcome
Number of reads	210,000	223,942
Read length (bases)	240	88 average (45 – 300 distribution)
Number of bases	50.4 Mb	19.7 Mb

The distribution of sequence length in ASP16-2 is shown in *Figure 2-52*. Most sequences were shorter than 105 nt. Striking was a unique peak at the position of 55 nt, which contained more than 37000 sequence reads (about 16.5% of all sequences). The DNA band of 74 nt (55 plus Roche-B of 19 nt) corresponding to this peak had not been observed before on agarose gel (see *Figure 2-50*). The length of 74 nt is even shorter than the minimal length defined by the fusion-primers used for ASP16, which is about 80 nt containing Roche-A (19 nt), HPV16 primer (about 20 nt), GPUA (18 nt), barcode (4 nt) and Roche-B (19 nt). This implied that the aberrant 74-nt product should be some kind of primer dimer. The nature of this peak (primer dimer) has been clarified and will be explained later in *Figure 2-56*.

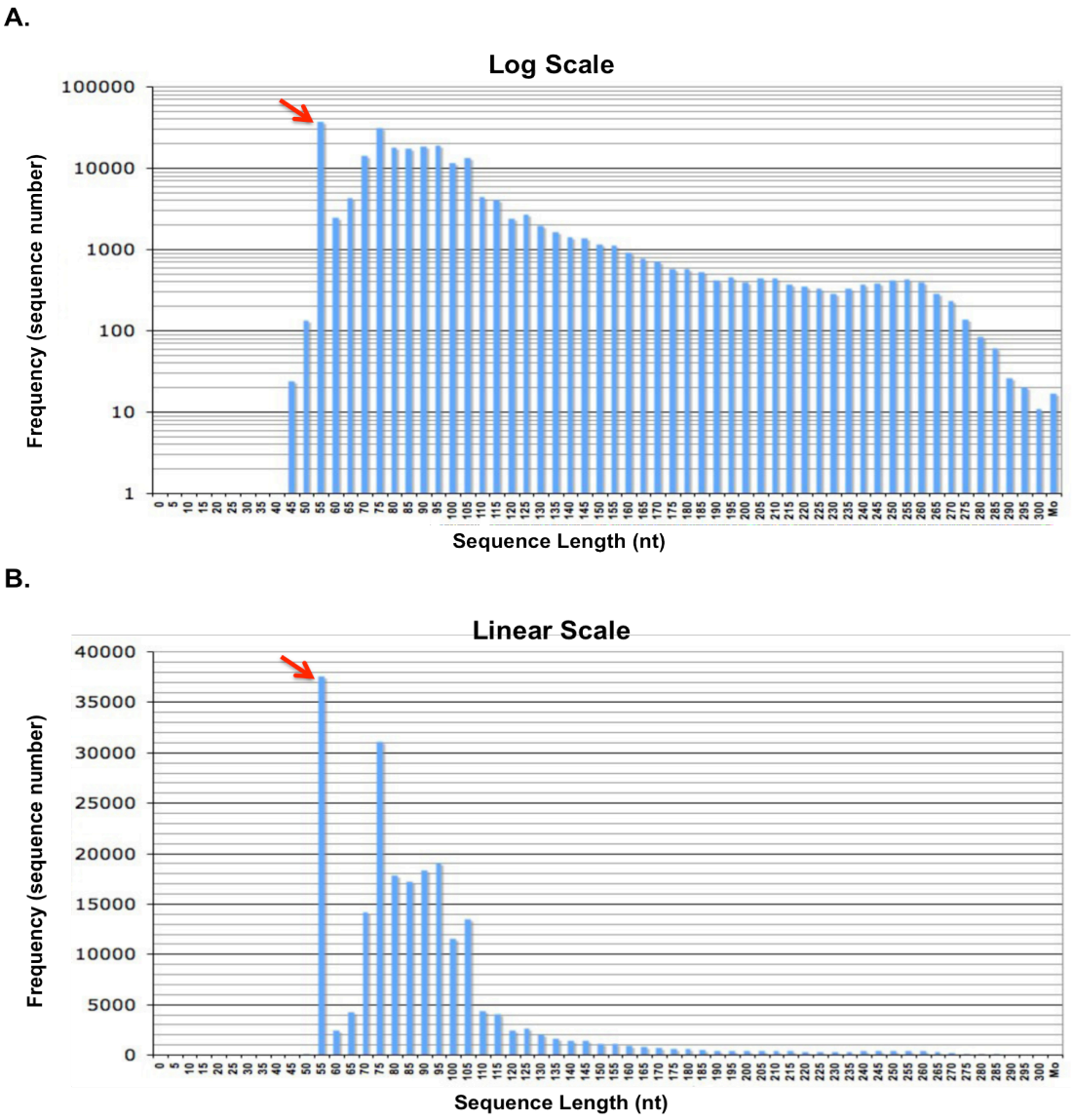


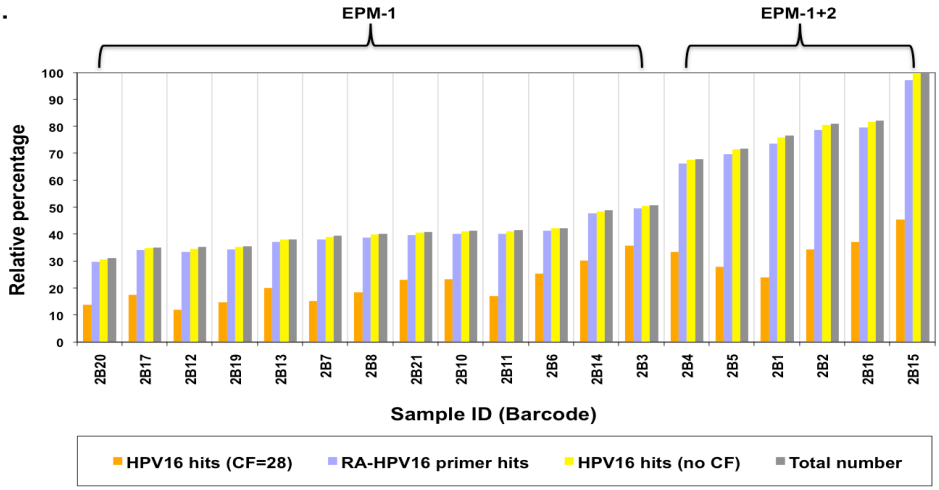
Figure 2-52: Overall length distribution of sequence reads from experiment ASP16-2. The X-axis represents increasing sequence lengths in 5-nt interval. The Y-axis shows the frequency (sequence number) of the corresponding read length. The length distribution is demonstrated in log scale (panel **A**) and linear scale (panel **B**). The red arrows point to the striking 55-nt peak. (Mo = more)

The 223942 sequences obtained from experiment ASP16-2 were analyzed using the programs described in *Figure 2-43*. Sequence distribution within each of the 19 samples used for pyrosequencing is shown in *Figure 2-53*, according to the four filter criteria: *total sequence number*, *HPV16 hit without cut-off*, *RA-HPV16 primer hit*, and *HPV16 hit with 28-nt cut-off*. Overall, the *total sequence number* of each sample was quite similar to each other, within the same category of either EPM-1 or EPM-1+2 (*Figure 2-53A and B*). Each sample contained 40–60 % of significant (≥28nt) HPV16 sequences (*Figure 2-53C*). These results indicated that the HPV16 DNA enrichment procedure is able to effectively select HPV16 sequences from whole-genome DNA.

A.

Sample ID (Barcode)	Significant HPV16 hits (CF=28nt)	RA-HPV16 primer hits	HPV16 hits (no CF)	Total sequence number	
2B20	3071	6563	6781	6867	EPM-1
2B17	3888	7499	7676	7744	
2B12	2663	7365	7628	7769	
2B19	3235	7567	7760	7834	
2B13	4439	8202	8357	8399	
2B7	3372	8366	8563	8672	
2B8	4072	8554	8765	8816	
2B21	5106	8721	8928	8989	
2B10	5155	8833	9041	9111	
2B11	3748	8824	9053	9140	
2B6	5600	9110	9279	9297	
2B14	6671	10493	10682	10744	
2B3	7853	10903	11135	11146	
2B4	7363	14569	14886	14940	EPM-1+2
2B5	6141	15317	15718	15766	
2B1	5297	16204	16707	16842	
2B2	7581	17295	17732	17802	
2B16	8203	17494	17961	18046	
2B15	9990	21343	21887	21977	

B.



C.

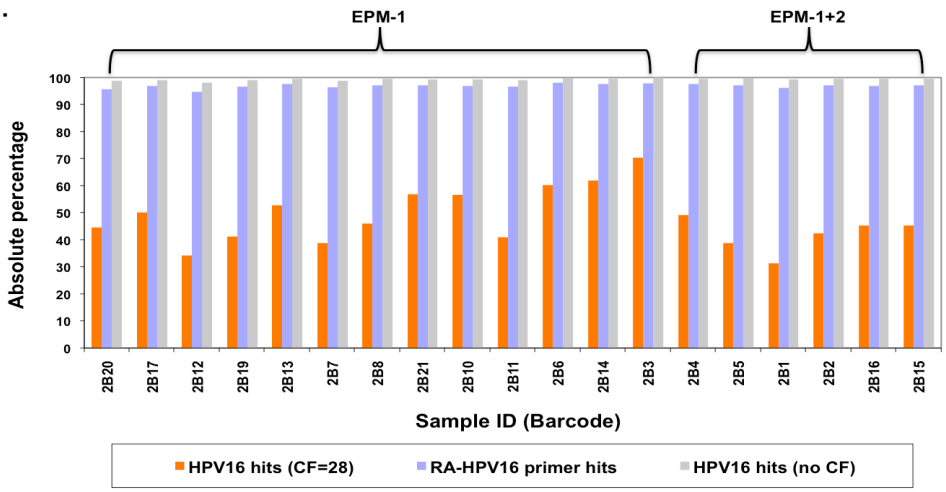


Figure 2-53: Sequence distribution of the 19 DNA samples analyzed in ASP16-2. The absolute sequence numbers (panel **A**) and the relative numbers in percentage (panel **B**) compared to the total number of sample 2B15 are shown for each sample with four different filter criteria. Panel **C** shows the absolute percentage of significant (≥28nt) HPV16 sequences for each sample (CF=cut-off; RA=Roche-A; EPM=HPV16 early primer mix)

To assess the specificity of the individual HPV16 primers, the percentage of significant (≥28nt) HPV16 sequences among the 19 samples was calculated for each nested HPV16 primer (EP-N1 to EP-N16) and illustrated as box plot (*Figure 2-54*). The median values showed large differences in specificity among the primers. Most of them worked specifically and gave rise to more than 60% of significant HPV16 sequences. Only two primers, EP-N1 and EP-N10, turned out to be very unspecific.

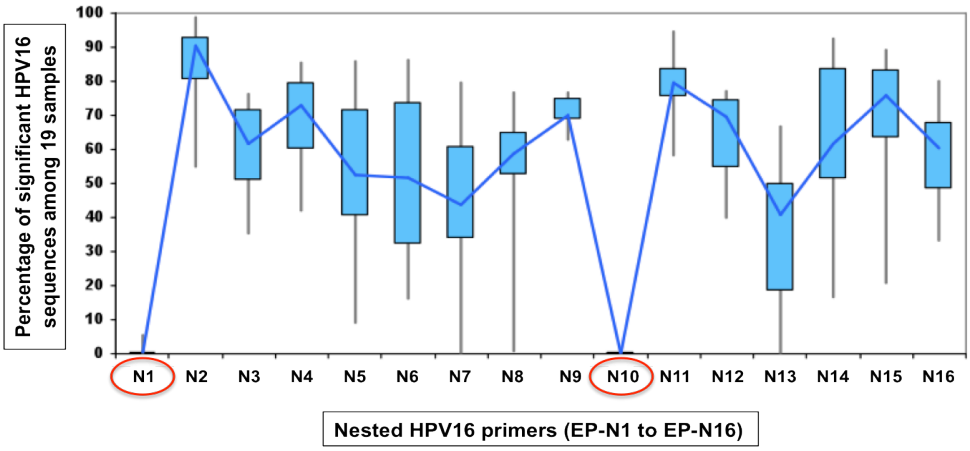


Figure 2-54: Percentage of significant HPV16 sequences among the nested HPV16 primers (EP-N1 to EP-N16) across the 19 DNA samples analyzed in ASP16-2. The five-number summary (Min, Q1, Median, Q3, Max) of sequence distribution is demonstrated by cyan boxes and grey lines outside. The blue line connects the median percentages of all primers. The red circles indicate the two unspecific primers EP-N1 and EP-N10.

To understand how EP-N1 and EP-N10 have impaired the experimental outcome, sequence numbers of *RA-HPV16 primer hits* and *significant (≥28nt) HPV16 hits* were listed for all 19 samples and 16 nested HPV16 primers (*Table 2-55*). For EP-N1, 15584 sequence reads in total were obtained from the 19 samples, while only 58 of them were significant HPV16 sequences. Inspection of the sequences revealed that most of them contained exactly the 21-nt EP-N1 sequence (HPV16 pos. 857 – 877), and then continued with non-HPV16 sequences from various cellular target sites (data not shown). These results imply that the HPV16 primer EP-N1 might have the tendency to target repetitive sequences of human genomic DNA, and therefore should be eliminated from further ASP16 experiments. By contrast, taking EP-N2 as example for a specific HPV16 primer, 14962 sequence reads were obtained from all 19 samples, and 13271 of them (about 89%) were significant HPV16 sequences.

For EP-N10, 38302 sequences from six samples were obtained, but only 27 of them passed the 28-nt cut-off criterion (*Table 2-55B*). Therefore, about 17% (38275 out of 223942) of total sequence reads in ASP16-2 were unspecific products induced by

primer EP-N10. Based on these data, it also became obvious that EP-N10 is most likely responsible for the 55-nt peak in *Figure 2-52*. Sequence comparison of EP-N10 and GPUA revealed a perfect complementarity of 8 nt at their 3' ends (*Figure 2-56*). Priming would create a primer dimer of 74 bp, which has a read length of 55-nt in experiment ASP16-2. These results clearly point to the fact that HPV16 primers with long stretches complementary to GPUA at 3' ends should be omitted.

Table 2-55: Comparison between *RA-HPV16 primer hits* and *significant HPV16 hits* of each nested HPV16 primer for all 19 samples in ASP16-2.

A.

	RA-HPV16 primer hits								Significant HPV16 hits (CF=28nt)							
	N1	N2	N3	N4	N5	N6	N7	N8	SN1	SN2	SN3	SN4	SN5	SN6	SN7	SN8
2B1	1308	790	2202	868	101	1096	757	1458	8	638	1528	634	53	671	225	703
2B2	429	104	2956	1101	124	511	1191	2646	2	57	2134	890	104	358	722	1760
2B3	169	961	5051	703	68	207	53	3668	3	914	3527	549	50	168	0	2527
2B4	115	1027	2553	708	82	320	975	1700	6	957	1819	605	60	276	774	1300
2B5	249	1371	3488	205	22	179	434	1818	6	1244	2666	156	2	140	304	1065
2B6	464	1069	5717	849	214	418	168	199	6	880	3592	642	173	238	2	2
2B7	1478	582	2667	581	87	443	589	1936	2	455	1188	367	42	144	118	1007
2B8	1258	527	2927	596	127	502	812	1800	6	396	1574	367	80	200	398	1002
2B10	675	598	3588	430	47	153	461	2881	0	557	2063	305	20	79	208	1859
2B11	1292	497	3045	383	49	286	777	2488	1	399	1454	179	16	47	294	1313
2B12	1240	358	2780	237	28	211	443	2063	2	247	979	100	13	44	88	1152
2B13	989	794	2621	437	36	259	772	2294	4	719	1492	260	17	81	329	1477
2B14	1044	1328	4115	900	130	428	1264	1284	3	1204	3128	702	91	237	772	459
2B15	538	1149	3700	902	171	606	1335	2923	0	1065	2701	741	147	470	892	1923
2B16	228	1277	3359	496	46	213	1004	2740	1	1262	2543	422	31	175	646	1923
2B17	800	427	2885	301	50	209	501	2323	4	371	1472	168	19	100	219	1493
2B19	1308	551	2259	491	43	270	819	1825	3	465	1094	263	15	80	341	915
2B20	1020	580	2012	501	50	277	648	1475	0	535	1019	304	18	71	247	836
2B21	980	972	2933	653	78	210	775	2119	1	906	1813	528	51	72	405	1273

B.

	RA-HPV16 primer hits								Significant HPV16 hits (CF=28nt)							
	N9	N10	N11	N12	N13	N14	N15	N16	SN9	SN10	SN11	SN12	SN13	SN14	SN15	SN16
2B1	931	6353	36	40	22	12	100	130	639	8	21	20	3	2	21	59
2B2	1344	5989	66	91	25	18	339	361	923	2	49	64	8	10	209	208
2B4	1563	5161	60	59	12	28	164	42	1196	0	47	45	6	19	146	14
2B5	660	6800	16	10	5	4	36	20	416	4	13	4	0	2	30	16
2B15	1672	7337	72	78	12	13	347	488	1187	13	61	60	6	12	239	340
2B16	924	6662	37	42	3	9	151	303	708	0	35	29	2	8	125	192

- CF=cut-off; N1-N16=EP-N1 to EP-N16; SN1-SN16=significant HPV16 hits of EP-N1 to EP-N16
- The red boxes indicate the unspecific primers EP-N1 and EP-N10.
- The green boxes indicate the specific primer EP-N2.
- ΣN1=15584 hits, ΣSN1=58 hits; ΣN10=38302 hits, ΣSN10=27 hits
- ΣN2=14962 hits, ΣSN2=13271 hits

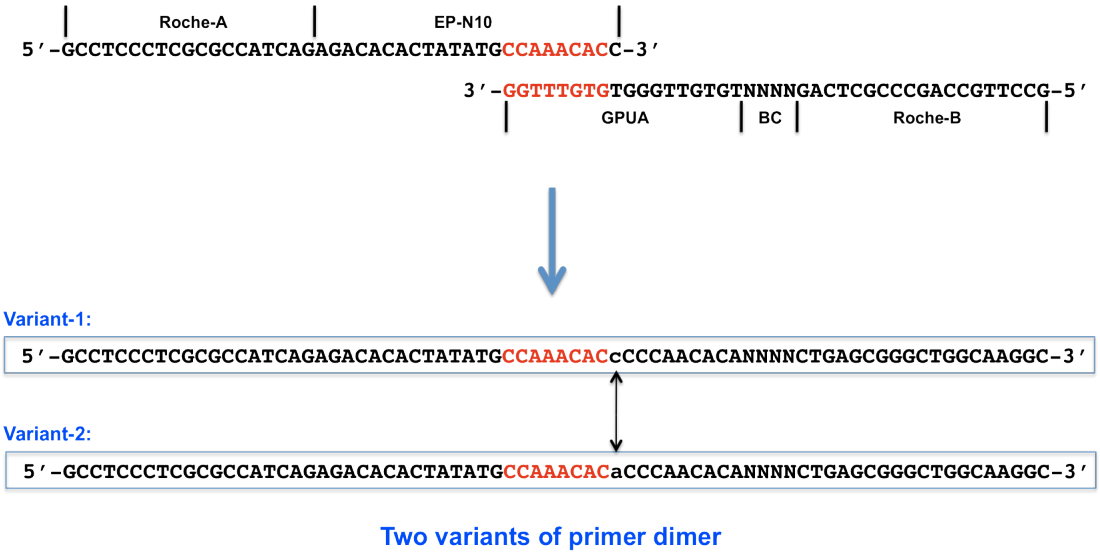


Figure 2-56: Origin and sequence of the 74-bp primer dimer. The 74-bp primer dimer responsible for the 55-nt peak in *Figure 2-52* was created by false priming between EP-N10 and GPUA. The DNA stretches marked in red show the 8-nt perfect match between EP-N10 and GPUA at their 3' ends. There exist two variants of this primer dimer, which differ to each other only in one nucleotide marked by black arrows. (BC=barcode)

Identification of HPV16-cellular junction sequences

The primary goal of the ASP16 strategy is to identify unknown HPV16-cellular junction sequences from clinical DNA samples. After barcode sorting of the pyrosequencing data, significant (≥28nt) HPV16 sequence reads were compared with the human genome reference sequence to search for viral-cellular junctions (see *Figure 2-43*). This was achieved by MegaBlast searching in NCBI databases with the standard criterion (28-nt cut-off of sequence match with human genome reference). In addition, for cell lines with already known viral-cellular junctions, pyrosequence reads could be compared directly with the known junction sequence. By doing this, pyrosequences with shorter cellular part (<28nt), which do not pass the standard searching criterion (28-nt cut-off), can be determined.

MRI-H186:

For cell line MRI-H186, two 3' viral-cellular junctions have been determined before by RS-PCR and long-PCR (see *Section 2.8.5*). The major 3' junction of MRI-H186 is located at HPV16-2754/chr8-128815788. In experiment ASP16-2, this junction should be found in sequence reads starting with the HPV16 primer EP-N6 (pos. 2723 – 2742). The distance between the 3' end of EP-N6 and the breakpoint is 12 nt. For

MRI-H186 (sample-ID 2B1), altogether 671 significant (≥28nt) HPV16 sequence reads were assigned to primer EP-N6 (see *Table 2-55A*). With MegaBlast searching (28-nt cut-off for the cellular part), 150 out of 671 sequence reads (22%) representing the junction HPV16-2754/chr8-128815788 were identified (class 1 in *Figure 2-57*). By comparing all the 671 reads to the known junction sequence, 320 reads (48%) turned out to contain the junction HPV16-2754/chr8-128815788. From them, 170 junction sequences contained cellular parts shorter than 28 nt (class 2 in *Figure 2-57*), and therefore were not detected by MegaBlast searching. The remaining 351 significant HPV16 reads (671-320=351) contained purely HPV16 sequences. These sequences are in agreement with the previous finding that complete HPV16 genomes are present in MRI-H186 (integration variant A+, see *Section 2.8.5*). Taken together, the results for MRI-H186 demonstrate convincingly the capacity of the ASP16 strategy for identification of viral-cellular junction sequences.

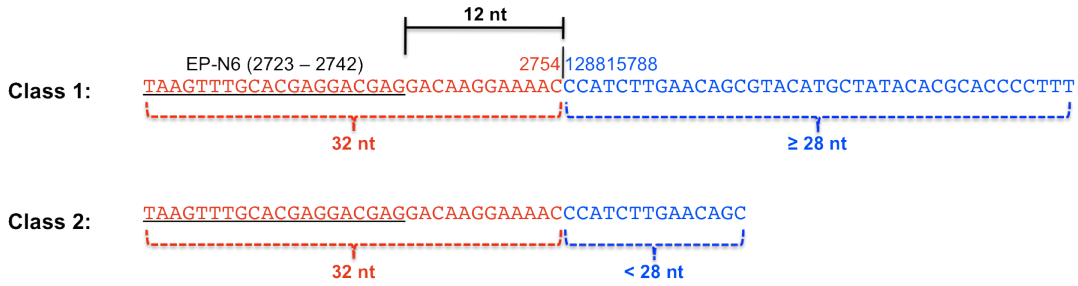


Figure 2-57: Viral-cellular junction sequences of MRI-H186 obtained from ASP16-2. For each class, one example of sequence reads is shown. The HPV16 sequence parts are shown in red. The cellular sequence parts (on chromosome 8) are shown in blue. Details about the two classes are explained in the text.

The minor 3' junction of MRI-H186 is located at HPV16-1223/chr8-128744999. In experiment ASP16-2, this junction is located between primers EP-N9 (pos. 1064 – 1084) and EP-N2 (pos. 1275 – 1296). The distance between EP-N9 and the junction is 139 nt. Primed by EP-N9, altogether 931 RA-HPV16 primer hits and 639 significant HPV16 hits were obtained (see *Table 2-55B*). The longest one of these sequences stops at the position 1219 of HPV16, thus not long enough to cross the 3' junction at position 1223.

MRI-H196:

For cell line MRI-H196, the 3' viral-cellular junction HPV16-3858/chr11-47967861 has been characterized by RS-PCR (see *Section 2.1.2*). In experiment ASP16-2, this junction should be expected from sequence reads starting with the HPV16 primer EP-N16 (pos. 3778 – 3798). The distance between the 3' end of EP-N16 and the breakpoint is 60 nt. For MRI-H196 (sample-ID 2B2), 208 significant HPV16 reads were assigned to EP-N16 (see *Table 2-55B*). However, no junction sequence was obtained by MegaBlast searching. By direct comparison between the known junction sequence and the 208 individual reads, one sequence containing the junction HPV16-3858/chr11-47967861 was identified. As shown in *Figure 2-58*, the cellular part of this sequence read was only 6-nt long. This sequence is also the longest one of the 208 reads. The rest of significant HPV16 sequences were too short to reach the junction site, and therefore contained purely HPV16 sequences.

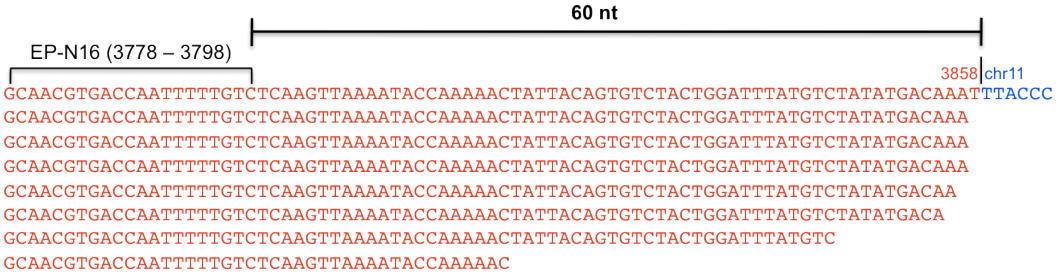


Figure 2-58: Sequence comparison between the viral-cellular junction sequence and pure HPV16 sequences of MRI-H196 obtained from ASP16-2. The 208 significant HPV16 sequences starting from EP-N16 were sorted according to their lengths (descending). The HPV16 sequences from the eight longest reads are shown in red. The cellular part (6 nt) of the longest sequence is shown in blue.

SiHa:

For cell line SiHa (sample-ID 2B3), the 3' viral-cellular junction known as HPV16-3133/chr13-74087558 (see *Section 2.1.1*) has been characterized before (Baker *et al.*, 1987; Meissner, 1999). The known junction sequence could not be identified in experiment ASP16-2 for reasons demonstrated in *Figure 2-59*. Firstly, the HPV16 primer EP-N7 (pos. 3121 – 3143) is located exactly at the junction site, and can bind onto DNA target only with the 5' part (pos. 3121 – 3133). In agreement with the priming site, no sequence was obtained for SiHa with primer EP-N7 (see *Table 2-55A,* row SN7). Secondly, the HPV16 primer EP-N14 (pos. 2933 – 2953), which is 180-nt upstream of the junction site, was not included into the pyrosequencing for SiHa (see *Table 2-48*). Thirdly, the HPV16 primer EP-N6 (pos. 2723 – 2742) is located 391-nt upstream of the junction site. For SiHa, 168 significant HPV16 reads

were assigned to EP-N6 (see *Table 2-55A*), but none of them extended to the junction site. The longest sequence ended at position 2926 of HPV16.

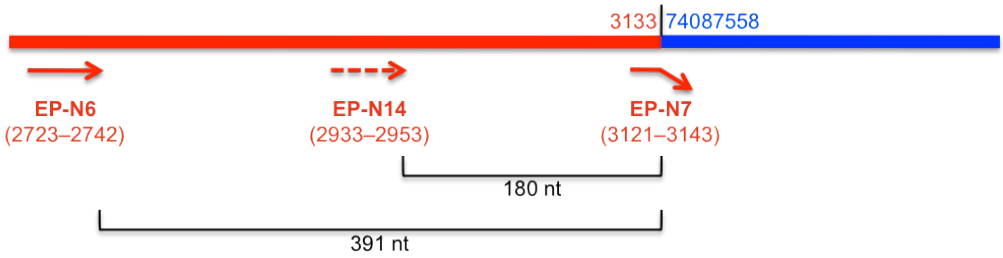


Figure 2-59: Limitations of HPV16 primers used in ASP16-2 preventing determination of the 3' viral-cellular junction in SiHa. Details are explained in the text.

In summary, the analysis of junction sites from cell lines clearly demonstrates that the ASP16 performance under the current conditions is extremely dependent on the distance between the 3' end of HPV16 primer and the junction site (*Table 2-60*). The main reason is the short average read length (only 88 nt) obtained in ASP16-2.

Table 2-60: Relationship between the distance of HPV16 primer to junction site and the yield of viral-cellular junction sequences.

Cell line (junction)	Distance between HPV16 primer and junction site	Significant HPV16 reads from primer	Sequence reads with viral-cellular junction	Identification by MegaBlast
MRI-H186 (major)	12 nt	671	320	150
MRI-H186 (minor)	139 nt	639	0	0
MRI-H196	60 nt	208	1	0
SiHa	391 nt	168	0	0

Clinical samples

From the 16 clinical DNA samples analyzed in experiment ASP16-2, a collection of 40 potential viral-cellular junction sequences representing 28 unique HPV16 DNA integration sites in 14 samples was obtained with MegaBlast searching (*Table 2-61*). To examine the potential junction sequences, junction-specific PCR assays were applied, for which an HPV16 forward primer was used in combination with a cellular reverse primer. Junction-specific PCRs were performed for all 28 potential junction sequences. Two integration sites could be confirmed by junction-specific PCR for cervical carcinomas 07c368 (sample-ID 2B5) and 07c381 (2B6), respectively. The experimental details will be explained later in *Sections 2.13 and 2.14*. All the other potential junctions were determined as false-positives.

Table 2-61: Potential viral-cellular junction sequences of clinical DNA samples obtained in experiment ASP16-2.

Unique Junction	Sequence ID	Sample ID (Barcode)	Sequence Length	Length of HPV16 part	HPV16 Breakpoint	Chromosome	Confirmed by Junction PCR
1	264864	2B5	180	87	1662	1	-/+
	369319	2B5	180	87	1662	1	
	321778	2B5	180	87	1662	1	
2	425227	2B6	221	61	2783	12	+
	340436	2B6	221	61	2783	12	
3	303044	2B7	106	29	3583	21	-
4	143460	2B8	114	28	884	7	-
	171376	2B8	114	28	884	7	
	297684	2B8	114	28	884	7	
	330467	2B8	114	28	884	7	
5	120501	2B8	108	29	3583	16	-
6	259802	2B8	183	28	3148	13	-
	399946	2B8	183	28	3148	13	
7	266576	2B10	177	28	1603	5	-
8	207966	2B10	152	29	3149	8	-
	475129	2B10	152	29	3149	8	
9	421622	2B10	120	46	1996	8	-
10	158429	2B11	161	29	3149	2	-
	296855	2B11	161	29	3149	2	
11	417998	2B12	133	29	3583	7	-
	233421	2B12	133	29	3583	7	
	392896	2B12	133	29	3583	7	
12	267802	2B13	148	30	886	1	-
13	374843	2B14	129	55	1329	5	-
14	101644	2B14	159	28	3148	1	-
	398469	2B14	159	28	3148	1	
15	249887	2B15	156	80	1655	X	-
16	201301	2B15	141	68	3188	18	-
17	304211	2B17	243	37	3157	6	-
18	259559	2B17	109	32	888	20	-
19	209907	2B17	196	116	1691	2	-
20	147955	2B17	123	43	3597	13	-
21	231983	2B19	116	28	1603	13	-
22	279325	2B19	183	31	1606	1	-
23	324509	2B19	250	30	3584	5	-
24	415536	2B19	110	28	3577	X	-
25	194623	2B20	119	34	3154	14	-
26	123560	2B20	119	28	3148	2	-
27	276102	2B20	119	28	3148	3	-
28	136223	2B21	99	28	3582	2	-

(The five sequences marked in blue represent two unique HPV16 DNA integration sites which were confirmed by junction-specific PCR. All other potential viral-cellular sequences could not be confirmed by junction-specific PCR.)

The original intention was to analyze only those clinical DNA samples by ASP16 strategy, for which the HPV16 DNA physical state had been estimated using E2-E6 quantitative PCR (see *Figure 1-6*) and which should contain a high percentage of integrated HPV16 DNA. For the samples included in experiment ASP16-2, however, the E2-E6 qPCR was actually performed after pyrosequencing. The results of E2-E6 qPCR are shown in *Table 2-62*. According to the qPCR data, sample 2B6 (E2/E6 ratio=0) should contain fully-integrated HPV16 DNA. Sample 2B5 (E2/E6 ratio=0.16) harbors mainly integrated HPV16 DNA (84%), although episomal DNA also co-exists. By contrast, all the other clinical samples contain high percentages of episomal HPV16 DNA. The qPCR data consisted with the outcome of experiment ASP16-2, which has only shown the junction sequences for 2B5 and 2B6.

Table 2-62: Comparison between HPV16 physical states and ASP16-2 outcome for clinical DNA samples.

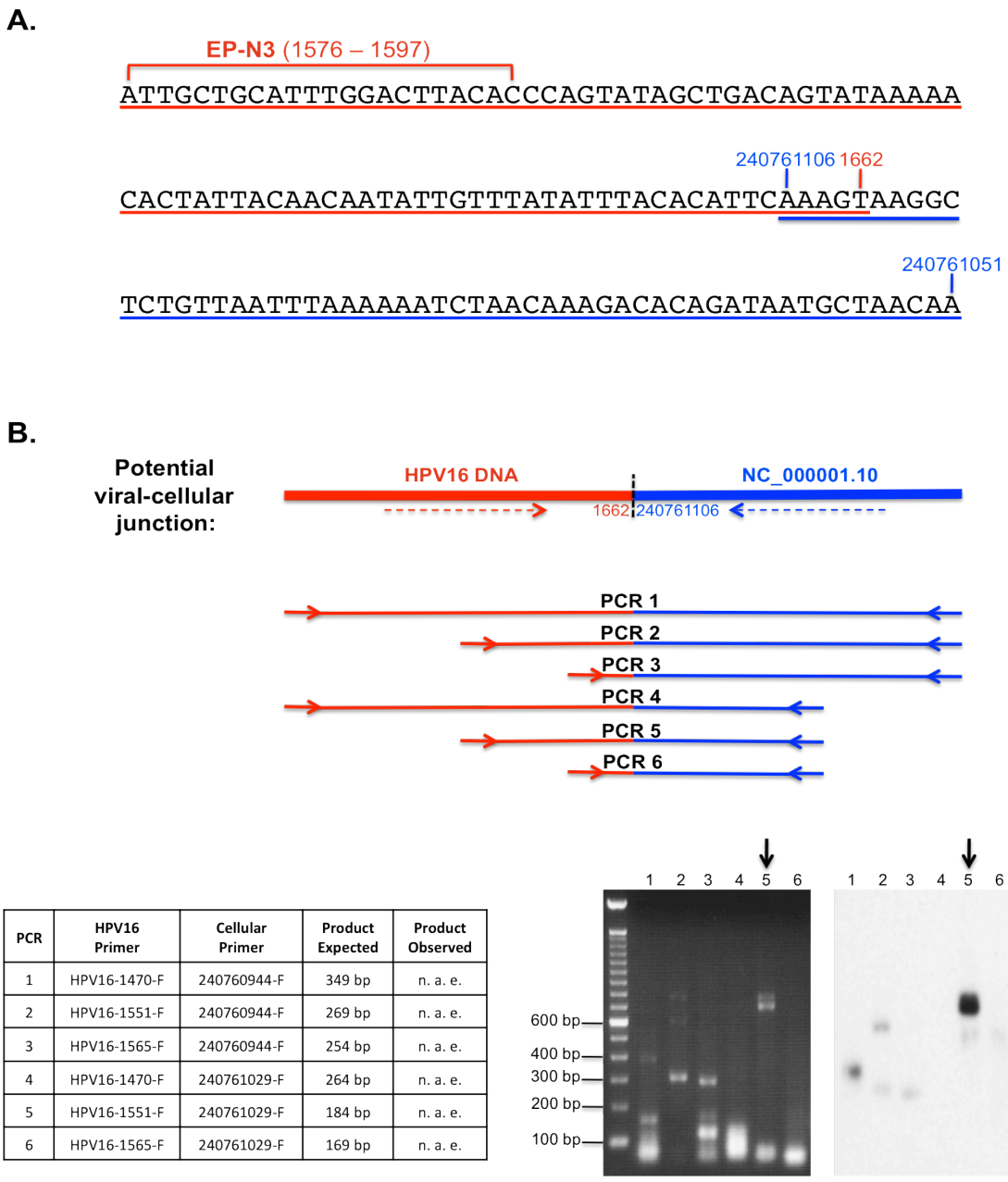
Sample ID (Barcode)	Sample Name	Lesion Grade	E2/E6 Ratio (%) [a]	Estimation of HPV16 Physical State [a]	HPV16 integration site identified in ASP16-2
2B4	07c134	cervical carcinoma	138	not possible	None
2B5	07c368	cervical carcinoma	16	mixed	1q43
2B6	07c381	cervical carcinoma	0	fully integrated	12q13.3
2B7	07c715	cervical carcinoma	48	mixed	None
2B8	07c716	cervical carcinoma	98	fully episomal	None
2B9	07c719	cervical carcinoma	76	mixed	Not included in pyrosequencing
2B10	61483	CIN3	63	mixed	None
2B11	62973	CIN3	70	mixed	None
2B12	63179	CIN3	n.d.		None
2B13	63282	CIN3	118	fully episomal	None
2B14	64655	CIN3	165	not possible	None
2B15	64702	CIN3	79	mixed	None
2B16	65468	CIN3	60	mixed	None
2B17	65892	CIN3	62	mixed	None
2B18	66019	CIN3	0	fully integrated	Not included in pyrosequencing
2B19	67230	CIN3	87	mixed	None
2B20	d1	n.d.	n.d.		None
2B21	d2	n.d.	n.d.		None

[a] E2/E6 ratio = 0%, fully integrated HPV16 DNA
E2/E6 ratio > 100%, experimental bias (estimation of HPV16 physical state not possible)
E2/E6 ratio between 0 and 100%, mixed (episomal + integrated)
E2/E6 ratio = 100%, fully episomal HPV16 DNA
(n.d. = not determined; E2/E6 qPCR was performed by Veronique Dalstein in Reims.)

2.13 HPV16 integration site of cervical carcinoma 07c368 (2B5)

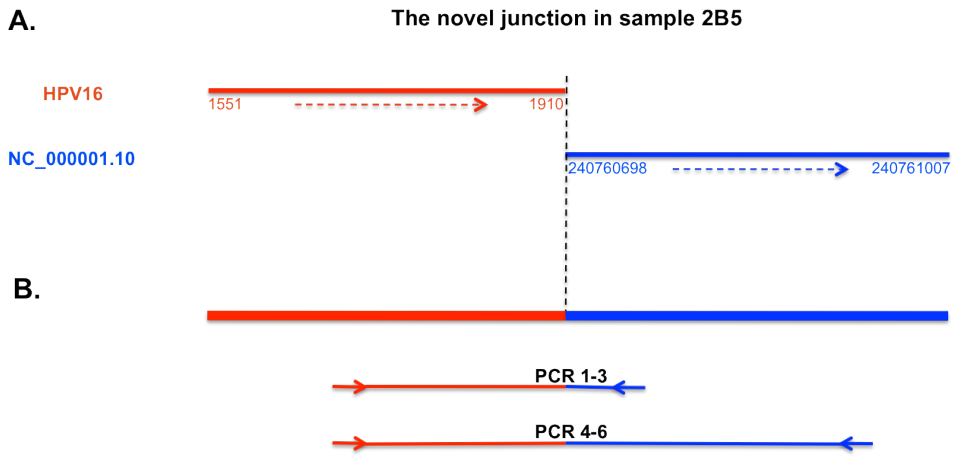
For clinical sample 2B5 (cervical carcinoma 07c368), three identical sequence reads derived from one potential viral-cellular junction were obtained in ASP16-2 (see *Table 2-61*). Starting with HPV16 primer EP-N3 (pos. 1576 – 1597), the HPV16 DNA breakpoint is located at position 1662 in the E1 gene, and the cellular DNA breakpoint on chromosome 1 at position 240761106 (NC_000001.10) (*Figure 2-63A*). To check the authenticity of this DNA organization, junction-specific PCR was performed using different combinations of HPV16 and cellular primers (*Figure 2-63B*). However, PCR products of expected sizes were not obtained. After Southern hybridization with HPV16-complete DNA probe, some positive signals were detectable on the blot. The PCR product of reaction 5 was cloned and sequenced. It turned out to be a novel viral-cellular junction sequence.

Distinct from the pyrosequence junction obtained in ASP16-2, the novel cloned viral-cellular junction has an HPV16 breakpoint at position 1910 in the E1 region and a cellular breakpoint at position 240760698 on chromosome 1 (NC_000001.10) (*Figure 2-64A*). To check the authenticity of the novel junction in 2B5, another series of junction-specific PCR reactions was performed using the viral forward primer HPV16-1793-F in combination with two cellular reverse primers (*Figure 2-64B*). PCR products with expected sizes were obtained from the GP-WGA library as template (reactions 1 and 4), and importantly also from the original 2B5 DNA as template (reactions 2 and 5), but not from MRI-H186 (reactions 3 and 6). The products were further confirmed by Southern hybridization with HPV16 DNA probe. Taken together, HPV16-1910/chr1-240760698 is the authentic viral-cellular junction of sample 2B5.



PCR	HPV16 Primer	Cellular Primer	Product Expected	Product Observed
1	HPV16-1470-F	240760944-F	349 bp	n. a. e.
2	HPV16-1551-F	240760944-F	269 bp	n. a. e.
3	HPV16-1565-F	240760944-F	254 bp	n. a. e.
4	HPV16-1470-F	240761029-F	264 bp	n. a. e.
5	HPV16-1551-F	240761029-F	184 bp	n. a. e.
6	HPV16-1565-F	240761029-F	169 bp	n. a. e.

Figure 2-63: Junction-specific PCR for the potential viral-cellular junction in sample 2B5. The potential junction sequence of sample 2B5 obtained from ASP16-2 is shown in panel **A** (HPV16 sequence underlined in red, cellular sequence underlined in blue). The junction-specific PCR and Southern hybridization (with HPV16 DNA probe) are shown in panel **B**. All six PCR reactions were performed using the 2B5 GP-WGA product as templates. The 100-bp DNA Ladder was used as size marker. (n.a.e.=not as expected)



PCR	DNA Template	Primer Forward	Primer Reverse	Product Expected	Product Observed
1	2B5 GP-WGA	HPV16-1793-F	240760735-R	154 bp	~150 bp (a.e.)
2	2B5 original DNA	HPV16-1793-F	240760735-R	154 bp	~150 bp (a.e.)
3	MRI-H186 GP-WGA	HPV16-1793-F	240760735-R	negative	negative (a.e.)
4	2B5 GP-WGA	HPV16-1793-F	240760912-R	331 bp	~350 bp (a.e.)
5	2B5 original DNA	HPV16-1793-F	240760912-R	331 bp	~350 bp (a.e.)
6	MRI-H186 GP-WGA	HPV16-1793-F	240760912-R	negative	negative (a.e.)

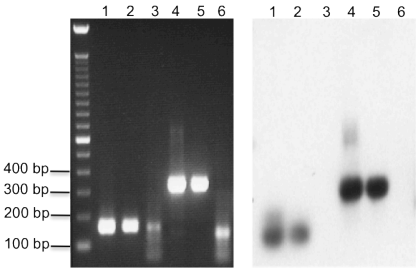


Figure 2-64: Junction-specific PCR for the novel viral-cellular junction of sample 2B5. The structure of the novel viral-cellular junction of 2B5 is shown in panel **A**. The junction-specific PCR and Southern hybridization (with HPV16 DNA probe) are shown in panel **B**. Junction-specific PCR was performed with both 2B5 GP-WGA and 2B5 original DNA as template, respectively. MRI-H186 GP-WGA was used as negative control. The 100-bp DNA Ladder was used as size marker. (a.e. = as expected)

In sample 2B5 (carcinoma 07c368), the integrated HPV16 DNA is located at chromosomal band *1q43* (*Figure 2-65*). A cellular gene called *Gremlin2* or *GREM2* (HGNC-ID: 17655) is located directly at the integration site. The *GREM2* gene contains two exons on the minus strand of chromosome 1, and encodes a BMP (bone morphogenic protein) antagonist family member. The integrated HPV16 DNA is inserted into the sole intron of *GREM2* (*Figure 2-65A*). Another gene named *RGS7* (*regulator of G-protein signaling 7*; HGNC-ID: 10003) is located downstream of the integrated HPV16 DNA (*Figure 2-65B*).

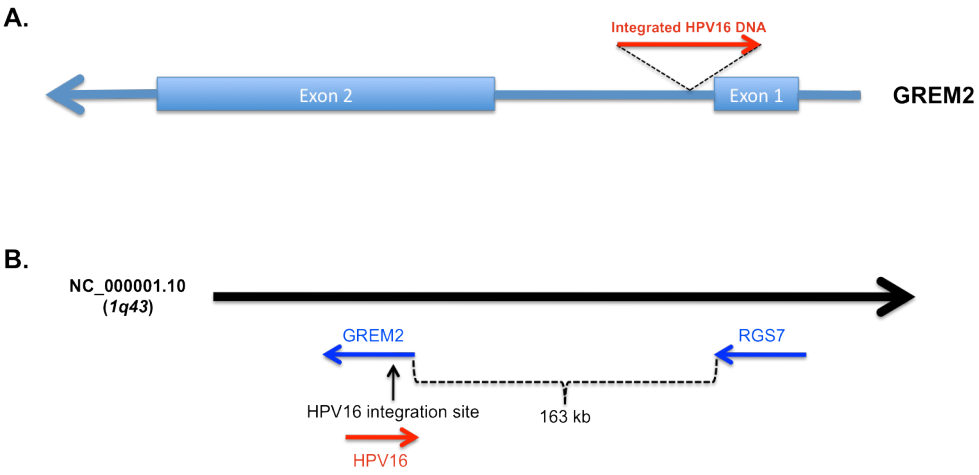


Figure 2-65: Schematic demonstration of the HPV16 DNA integration site (A) in sample 2B5 and the genetic context (B) near the junction site.

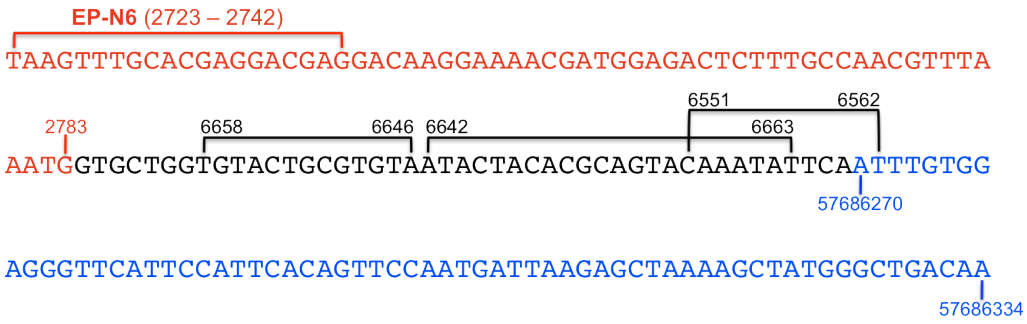
2.14 HPV16 integration site of cervical carcinoma 07c381 (2B6)

For clinical sample 2B6 (cervical carcinoma 07c381), two identical sequence reads representing one potential viral-cellular junction were obtained in ASP16-2 (see *Table 2-61*). The detailed sequence analysis revealed a complicated structure of this potential junction, which contained HPV16 early sequence, fused to 45 bp of strongly rearranged HPV16 L1 sequences and the downstream cellular sequence from human chromosome 12 (*Figure 2-66A*). Starting with primer EP-N6 (pos. 2723 – 2742), the HPV16 early sequence of this potential junction continues until position 2783 in the overlapping region of E1 and E2 ORFs. The cellular DNA breakpoint at the junction site is position 57686270 (NC_000012.11) of chromosome 12.

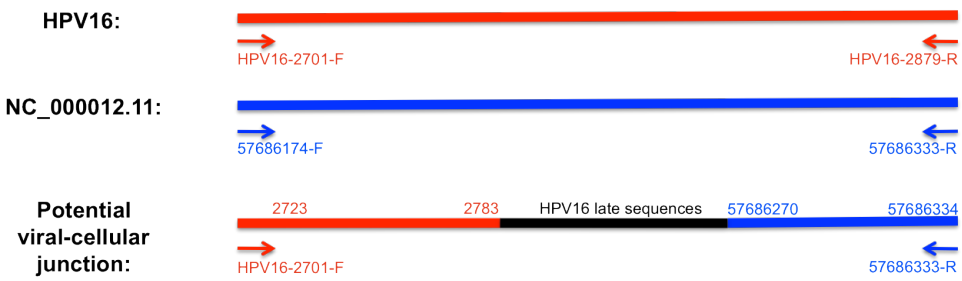
To examine the potential junction in sample 2B6, junction-specific PCR was performed using the viral forward primer HPV16-2701-F and the cellular reverse primer 57686333-R (*Figure 2-66B*; reactions 3 and 4). In addition, control reactions with either HPV16 (reaction 1) or cellular (reaction 2) primer pair alone were also done. PCR products with expected amplicon sizes of the potential junction were obtained from both the GP-WGA library as template as well as from the original 2B6 DNA as template (reactions 3 and 4). The PCR products of reactions 3 and 4 were cloned and sequenced. The sequences of the two clones are identical to each other, and also in full agreement with the pyrosequence. These results proved that the viral-cellular junction sequence of sample 2B6 obtained in ASP16-2 is authentic.

According to the E2-E6 qPCR data (see *Table 2-62*), the DNA sample 2B6 (E2/E6 ratio=0) was estimated to be HPV16 fully-integrated. As expected, in reaction 1 with primer pair HPV16-2701-F and HPV16-2879-R, no PCR product was observed, because the HPV16 reverse primer cannot bind to the junction sequence. In reaction 2 with cellular primers flanking the breakpoint 57686270, a band of expected size was detected, which suggests the existence of the related wild-type allele.

A.



B.



PCR	DNA Template	Primer Forward	Primer Reverse	Product Expected	Product Observed
1	2B6 GP-WGA	HPV16-2701-F	HPV16-2879-R	negative	negative (a.e.)
2	2B6 GP-WGA	57686174-F	57686333-R	159 bp	~160 bp (a.e.)
3	2B6 GP-WGA	HPV16-2701-F	57686333-R	193 bp	~200 bp (a.e.)
4	2B6 original DNA	HPV16-2701-F	57686333-R	193 bp	~200 bp (a.e.)

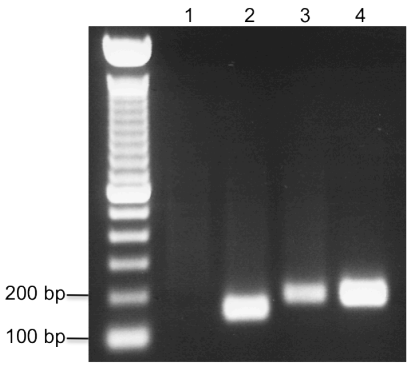


Figure 2-66: Junction-specific PCR for the viral-cellular junction sequence in sample 2B6. The potential viral-cellular junction sequence of sample 2B6 obtained in ASP16-2 is shown in panel **A** (HPV16 early sequence in red, cellular sequence in blue, the re-arranged HPV16 L1 sequences in black). Junction-specific as well as control PCRs are shown in panel **B**. Junction-specific PCR was performed with both 2B6 GP-WGA and 2B6 original DNA as template, respectively. The 100-bp DNA Ladder was used as size marker. (a.e. = as expected)

Deduced from the viral-cellular junction sequence, in sample 2B6 (carcinoma 07c381) the HPV16 DNA is integrated into chromosomal region *12q13.3*. The cellular gene *R3HDM2* (*R3H domain containing 2*; HGNC-ID: 29167) is located directly at the integration site. *R3HDM2* is composed of 22 exons and is located on the minus strand of chromosome 12. The integrated HPV16 DNA of sample 2B6 is located in the intron 8 of *R3HDM2* (*Figure 2-67A*). Three cellular genes with known functions including *INHBC* (inhibin, beta C; HGNC-ID: 6068), *INHBE* (inhibin, beta E; HGNC-ID: 24029) and *GLI1* (GLI family zinc finger 1; HGNC-ID: 4317), are situated downstream of the HPV16 integration site (*Figure 2-67B*).

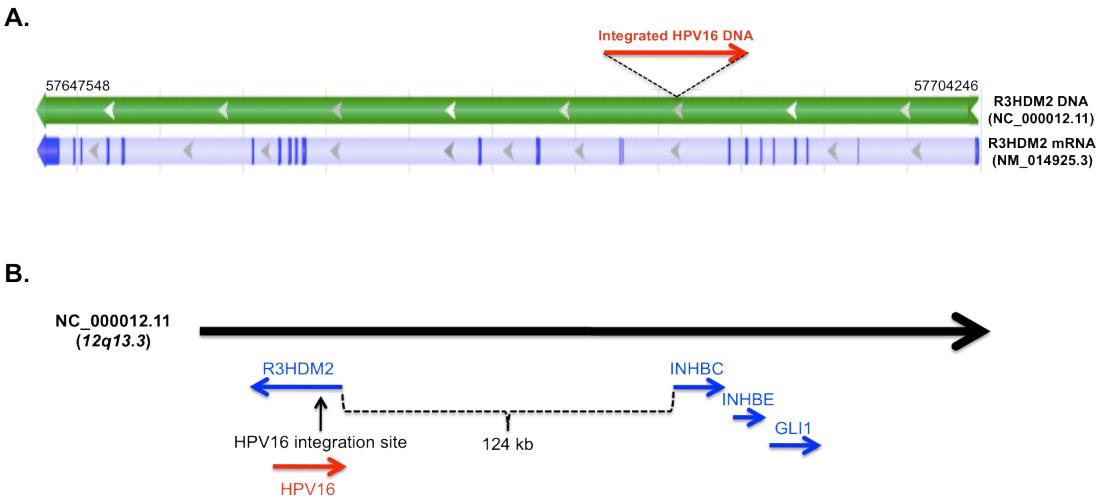


Figure 2-67: Schematic demonstration of the HPV16 DNA integration site (A) in sample 2B6 and the genetic context (B) near the junction site. The 22 exons of *R3HDM2* are shown as blue bars in panel **A**.

3. Discussion

3.1 HPV16 DNA integration analysis in MRI-H186

The present study has provided comprehensive insights into somatic DNA rearrangement and amplification in cervical carcinoma cell line MRI-H186, which are strongly coupled with HPV16 DNA integration. Altogether seven HPV16 integration variants have been identified in MRI-H186, and all variants are located on chromosomal band *8q24.21* in the *c-myc* region (see *Section 2.8.5*). Determined by signal intensities on Southern hybridization (see *Figure 2-17*), variants A and A+ are the two major products of integrated HPV16 DNA in MRI-H186. For the other five variants (B, C, A-B, D and E), no corresponding restriction fragments were detectable on Southern blots, although the flanking cellular sequences have been well characterized. This is most likely due to their low DNA copy number, compared to A and A+.

Sequence comparison between the viral-cellular junctions of all integration variants has suggested that in MRI-H186 there was a primary HPV16 DNA integration event, from which all the other variants have originated. This primary event should be A or A+, because they carry exactly the same junction sequences, and almost all the other variants have been generated from A or A+ by secondary events. However, it is impossible to judge whether the integration event A preceded A+ or *vice versa*. Variant A might be initially inserted into the human genome. Then, this architecture could have been converted to A+ by gain of a complete HPV16 genome from co-existing episomes (*Figure 3-1*). Otherwise, variant A+ might derive from a primary integration event of an HPV16 dimeric episome. The successive creation of A could be mediated by excision of a complete HPV16 genome from A+ (*Figure 3-1*).

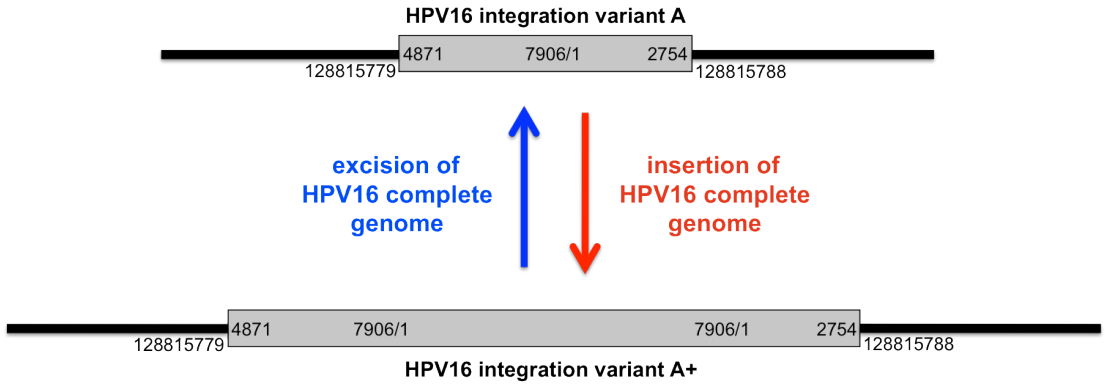


Figure 3-1: Possible conversions between HPV16 DNA integration variants A and A+ of MRI-H186. The structures of variants A and A+ with positions of viral-cellular junctions are shown. Details about the hypothesis are explained in the text.

Assessed by spectral karyotyping, MRI-H186 cells exhibit overall near-tetraploidy, with big differences in the numbers of individual chromosomes (see *Figure 2-26*). Despite this heterogeneity, only two chromosomes 8 have been found in each of the cells analysed (Jérôme Couturier, personal communication). This observation correlates with the FISH assay (see *Figure 2-25*), in which the co-localized HPV16 and *c-myc* signals were detected exclusively in the two chromosomes 8. As determined by quantitative PCR (see *Figure 2-27*), the copy number of *c-myc* gene is about 26 in each MRI-H186 cell (13 copies on each chromosome 8). The *c-myc* qPCR amplicon used for this assay is located in exon 1, which is present in the 3' flanking cellular sequences of almost all integration variants characterized (except A-B; see *Section 2.8.5*) as well as the *c-myc* wild-type allele, and therefore cannot distinguish between them. Most of the 26 *c-myc* copies, however, should be attributed to the two prominent integration variants A and A+ detected on Southern blot (see *Figure 2-17*). Taken together, these results have elucidated that HPV16 integration variants A and A+ together with their flanking cellular sequences, which contain the intact *c-myc* transcription unit, are amplified in MRI-H186 cells.

Co-amplification of integrated HPV16 DNA together with the flanking cellular sequences has also been reported for cervical carcinoma cell line HeLa (Lazo *et al*., 1989). Each HeLa cell harbors 10 – 50 copies of integrated HPV18 DNA (Schwarz *et al*., 1985), which are primarily mapped to chromosomal band *8q24* (Dürst *et al*., 1987; Lazo *et al*., 1989). The HPV18 integration site in HeLa is located about 500 kb upstream of the *c-myc* gene (Ferber *et al*., 2003). Southern analysis and chromosome *in situ* hybridization have shown that the integrated HPV18 DNA in HeLa is co-amplified with the flanking sequences from *8q24*, but the co-amplicons do not involve the *c-myc* coding region (Lazo *et al*., 1989). Recently, another group has reported that integrated HPV DNA is co-amplified with flanking cellular sequences in three out of five cervical carcinoma cell lines with integration sites at the *c-myc* locus on *8q24.21*, and this result was also verified in the original carcinomas (Peter *et al*., 2006). Estimated by qPCR with primer pairs distributed in *8q24.21*, the potential co-amplification of integrated HPV DNA together with *c-myc* coding region was speculated, but the co-amplicons had not been validated by cloning and sequencing (Peter *et al*., 2006). In comparison with the published examples, the present study on MRI-H186 has shown for the first time at sequence resolution the direct juxtaposition and co-amplification of integrated HPV16 DNA together with the complete transcription unit of *c-myc*. Since material of the original carcinoma from which cell line MRI-H186 was established was not available, it was impossible to directly

examine whether the HPV16 integration variants already existed in the primary tumor, or occurred later during the *in vitro* cell culture.

Regarding gene expression, HPV16 DNA integration in the *c-myc* region gives rise to two major hybrid transcripts in MRI-H186, the HPV16-*myc* and the *myc*-HPV16 mRNAs. Derived from integration variants A and A+, the HPV16-*myc* hybrid mRNA (see *Figure 2-6*) is a typical HPV-cellular co-transcript (see *Figure 1-5*). The HPV16-*myc* mRNA in MRI-H186 is a polycistronic mRNA with coding potentials for viral oncoproteins E6 and E7, as well as the human c-MYC protein. However, it is believed that only the upstream ORFs (in this case E6 and E7) could be efficiently translated from such a polycistronic mRNA in eukaryotes (Kozak, 2005). Therefore, this mRNA probably does not affect the *c-myc* expression level directly. To our knowledge, an HPV16-*myc* hybrid mRNA like that in MRI-H186 has not been described in other cell lines, in which HPV16 DNA is integrated in the *c-myc* region. A comparable mRNA structure has been described for the viral-cellular hybrid mRNA identified in cervical carcinoma cell line ME180 (Reuter *et al*., 1998). In ME180, the integrated HPV68 DNA splits the cellular tumor suppressor gene *APM-1* into two separate parts. The upstream part contains the untranslated exon 1, while the downstream part contains the protein-coding exons 2 and 3. A hybrid transcript composed of HPV68 E6, E7, and the complete *APM-1* ORF is produced. But, similar to the HPV16-*myc* mRNA in MRI-H186, the hybrid mRNA in ME180 does not support the translation of APM-1 protein either (Reuter *et al*., 1998).

The *myc*-HPV16 hybrid mRNA (see *Figure 2-10*), transcribed from integration variant B in association with the upstream cellular sequence, is a unique hallmark for MRI-H186 cells. It is expressed to a considerable scale in MRI-H186, as the second most abundant mRNA on *c-myc* Northern blot (see *Figure 2-7*), and the major mRNA on HPV16 Northern blot (see *Figure 2-9*). Structures analogous to this novel transcript have never been reported for any HPV type. In general, the typical HPV integration events give rise to viral-cellular hybrid transcripts, such as the HPV16-*myc* mRNA mentioned earlier. The unusual cellular-viral hybrid transcripts, in this case the *myc*-HPV16 mRNA, may only derive from secondary integration events by complicated DNA rearrangements. We speculated that the novel MycHPV16E2 in-frame fusion protein, as the end product of *myc*-HPV16 mRNA, might also be operative in cervical carcinogenesis, because it contains part of the trans-activation domain of c-MYC and the complete DNA-binding domain of HPV16 E2 (see *Figure 2-13*).

Functional analysis of the MycHPV16E2 fusion protein has been undertaken by Anna Tikidzhieva in the course of her Diplomarbeit (Tikidzhieva, 2009). It has been shown that the full-length E2 proteins of HPV16 and HPV18 can *trans*-activate a promoter through binding to the E2-binding sites (Phelps and Howley, 1987). However, for the early promoters of HPV16 (p97) and HPV18 (p105), the E2 proteins are actually repressors rather than activators of viral gene expression (Bernard *et al*., 1989; Romanczuk *et al*., 1990). This repression is thought to work in a dose-dependent manner, mediated by the occupancy status of E2-binding sites adjacent to viral promoter, and thereby sterically hinders the interaction between cellular transcription factors and their target sites for transcription initiation (Thierry and Howley, 1991; Tan *et al*., 1992; Tan *et al*., 1994; Demeret *et al*., 1994; Steger and Corbach, 1997). It has been suggested that the repression of viral oncogene expression caused by ectopic re-introduction of E2 protein leads to growth inhibition specifically in HPV-positive cervical carcinoma cell lines (Hwang *et al*., 1993; Dowhanick *et al*., 1995; Desaintes *et al*., 1997; Goodwin *et al*., 1998; Francis *et al*., 2000). Further investigation using HPV16 E2 mutants has proven that both the N-terminal *trans*-acting domain and C-terminal DNA binding domain of HPV16 E2 protein are required for the repression activity and cell death (Sakai *et al*., 1996; Nishimura *et al*., 2000). Transient transfection assays using luciferase reporter gene showed that in SiHa and HeLa cells exogenously expressed MycHPV16E2 protein *trans*-activates the co-transfected heterologous TK promoter, which is fused directly downstream to tandem repeated E2-binding sites (Tikidzhieva, 2009). Therefore, it seems that the partial c-MYC trans-activation domain at the N-terminus of MycHPV16E2 may be functional, because the DNA binding domain of HPV16 E2 has no *trans*-acting activity. Stable transfection assay have provided two stable clones from SiHa cells transfected with MycHPV16E2. One clone showed suppressed cell growth, but the other one did not show any obvious change compared to untransfected cells (Tikidzhieva, 2009). No stable clone could be isolated from HeLa cells transfected with MycHPV16E2, indicating that HeLa cell growth is not compatible with the presence of this protein (Tikidzhieva, 2009). From these results, it appears unlikely that the MycHPV16E2 fusion protein has oncogenic activity. More likely, the fusion protein is a transcription activator that probably causes cell death in HPV-infected cells, similar to HPV16 E2 protein. In MRI-H186 cells, the potential pro-apoptotic activity of endogenous MycHPV16E2 protein seems to be well tolerated.

Associated with HPV16 DNA integration into *c-myc* locus, the *c-myc* proto-oncogene is over-expressed at both mRNA and protein levels in MRI-H186 (see *Figure 2-7 and 2-8*). The human c-MYC protein is a multi-functional regulator involved in several important cellular processes including proliferation, differentiation, growth and apoptosis (Meyer and Penn, 2008). Aberrant *c-myc* expression has been frequently observed in a panel of human cancers (Dang, 1999; Grandori *et al*., 2000). The *c-myc* deregulation can be induced by gross genetic alterations such as viral insertional mutagenesis, chromosomal translocation and allelic amplification, or mediated by a variety of upstream cellular pathways (Meyer and Penn, 2008). Among them, *c-myc* over-expression caused by chromosomal translocation has been intensively characterized in human Burkitt's lymphoma. In these tumors, the *c-myc* locus on chromosome 8 is often juxtaposed to three different immunoglobin (Ig) loci on chromosomes 14, 2 and 22, respectively (Hecht and Aster, 2000). The translocation events lead to transcriptional regulation of *c-myc* promoters under the control of lymphocyte-specific Ig enhancers. In quiescent cells, the *c-myc* expression is negatively controlled by pausing of RNA polymerase II, which leads to premature termination of transcription (Eick and Bornkamm, 1986). By contrast, in Burkitt's lymphoma cells the blocked transcriptional elongation is released, thereby upregulating the *c-myc* expression compared to untransformed cells (Strobl and Eick, 1992). It has also been shown that Ig enhancers can modulate activities of promoters, which are located as far as about 500 kb apart (Chesi *et al*., 1998; Hecht and Aster, 2000). Therefore, it can easily be imagined that in MRI-H186 cells *c-myc* over-expression is caused by direct *cis*-activation of the *c-myc* promoter by the HPV16 enhancer (*Figure 3-2*), because the distance between HPV16 integration site and *c-myc* gene is only about 1.7 kb.

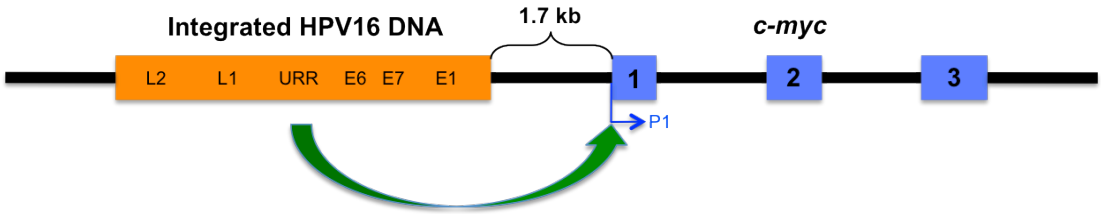


Figure 3-2: Hypothesized direct *cis*-activation of *c-myc* promoter by HPV16 enhancer. The structure of integrated HPV16 DNA (taking the variant A as example) and the flanking *c-myc* transcription unit is shown. The green arrow indicates the *cis*-activation of *c-myc* promoter (P1) by the HPV16 enhancer in upstream regulatory region (URR).

Concerning the fact that HPV DNA integration often occurs in *c-myc* locus (Dürst *et al*., 1987; Couturier *et al*., 1991; Sastre-Garau *et al*., 2000; Wentzensen *et al*., 2002; Ferber *et al*., 2003; Peter *et al*., 2006; Kraus *et al*., 2008; Dall *et al*., 2008), insertional mutagenesis of *c-myc* might not be a rare event but contribute to cervical carcinogenesis in a subset (about 10%) of cases. In an earlier study including 35 cervical carcinomas, DNA amplification and/or rearrangement of *c-myc* has been detected in about 90% of the samples (Ocadiz *et al*., 1987). Over-expression of *c-myc* has been reported in about 33% of altogether 93 cervical carcinomas (Bourhis *et al*., 1990). *c-myc* over-expression was shown to correlate with increased tumor size, overall relapse and distant metastasis, altogether suggesting a significant role of *c-myc* in tumor progression (Bourhis *et al*., 1990). Recent functional studies have shown that c-MYC protein can increase telomerase activity and thereby promote cellular immortalization, in cooperation with HPV16 E6 and E7 oncoproteins (Veldman *et al*., 2003; Liu *et al*., 2007; Liu *et al*., 2008). These results implicate that in cervical carcinogenesis the c-MYC oncoprotein may play an essential role for tumor progression.

3.2 The novel ASP16 strategy

In the present study, a novel strategy named ASP16 is introduced for the first time. The ASP16 strategy has been developed and examined during the thesis work, in order to determine HPV16 DNA integration sites at the sequence level in clinical pre-cancerous and cancerous lesions obtained from cervical scrapes. As proof of principle, two ASP16 experiments (ASP16-1 and 2) were performed in this thesis. Both experiments have clearly shown the successful whole genome amplification and the effective HPV16 DNA enrichment, starting with DNA samples isolated from cervical scrapes, which are generally of low quantity and often of poor quality. As far as we know, this is the first systematic attempt to identify HPV DNA integration sites in cervical scrapes. The PCR-based HPV DNA integration analyses (DIPS-PCR; RS-PCR) published so far have been focused on cervical carcinoma specimens rather than pre-cancerous lesions (Luft *et al*., 2001; Ziegert *et al*., 2003; Thorland *et al*., 2003; Ferber *et al*., 2003), probably because the DNA amounts obtained from the former are much more abundant than the latter. So far, HPV integration analysis of cervical scrapes has only been involved in the RT-PCR-based APOT assay at the cDNA level (Klaes *et al*., 1999; Wentzensen *et al*., 2002; Vinokurova *et al*., 2008; Kraus *et al*., 2008). However, due to RNA splicing, cDNA sequences do not usually

represent the viral-cellular DNA junctions directly. In comparison with the published techniques mentioned above, the novel ASP16 strategy seems to be the most plausible means for HPV16 DNA integration analysis in clinical samples from cervical scrapes, which requires only about 10 ng of genomic DNA input and tolerates the partial degradation of clinical DNA. For future applications, it is worthy of combining the ASP16 strategy at the DNA level and the APOT assay at the cDNA level for better experimental outcomes.

Concerning HPV16 DNA integration sites, the known 3' viral-cellular junction sequence of MRI-H186 cell line has been identified in the experiments ASP16-1 and 2. This clearly demonstrates that the whole concept of ASP16 strategy works. Based on the results of ASP16-1, some modifications on the experimental design were made for ASP16-2, including focus on the HPV16 early region, addition of new HPV16 primers to reduce primer interval, and expansion of pyrosequencing device to increase the total number of sequence reads. As a consequence, the performance of ASP16 has been improved in ASP16-2. This has helped to obtain much more sequence reads representing the already known 3' junction of MRI-H186, and to determine novel junction sequences from clinical DNA samples (carcinomas 2B5 and 2B6) as well. However, the problem of short average sequence length *versus* long primer interval remains to be solved. It was originally expected that the use of HPV16 early primer mix 2 (EPM-2) in ASP16-2 could reduce primer interval from about 400 to 200 bp, in association with primer mix 1 (EPM-1) also used in ASP16-1. But, the results from ASP16-2 have revealed that EPM-2 was not efficient enough, because the EP-N10 forms primer dimer together with GPUA. For optimization, unspecific primers EP-N10 (from EPM-2) and EP-N1 (from EPM-1) should be eliminated from future experiments or replaced by new ones. Even if the primer interval will be decreased to 200 bp in the future, the average length of pyrosequence reads (around 100 nt) obtained under the current experimental conditions is still too short. It is therefore necessary to remove primer dimers and PCR products shorter than 200 bp from the DNA mixture prepared for pyrosequencing.

In the experiment ASP16-2, HPV16 DNA integration sites have been identified and confirmed in two (cervical carcinomas 2B5 and 2B6) out of sixteen clinical DNA samples. Although the overall yield of authentic viral-cellular junctions was rather low, the ASP16 performance correlated well with HPV16 physical states of the clinical samples. Estimated by E2-E6 qPCR (see *Table 2-62*), carcinoma samples

2B5 and 2B6 should contain mainly (2B5; E2/E6 ratio=0.16) or exclusively (2B6; E2/E6 ratio=0) integrated HPV16 DNA. All the other fourteen clinical samples including three carcinomas and eleven pre-cancerous lesions, for which no authentic junction sequence was obtained in ASP16-2, showed high percentages of episomal HPV16 DNA. We learn from these results that for future ASP16 experiments the clinical DNA samples with low E2/E6 ratios should be preferentially conducted, in order to determine the HPV16 integration sites more efficiently. In the ongoing experiments ASP16-3 and ASP16-4 (performed in the PhD work of Sasithorn Chotewutmontri; unpublished data), this aspect has been taken into consideration, in addition to the experimental modifications mentioned before. In the experiment ASP16-3 for instance, HPV16 DNA integration sites have been determined in some cervical pre-cancerous lesions (CIN3) including the sample BV66019 with fully integrated HPV16 DNA, which was initially included in ASP16-2 (sample-ID 2B18), but not used for the pyrosequencing (see *Table 2-48* and *Figure 2-49*).

Taken together, the novel ASP16 strategy has offered the first opportunity to systematically explore HPV16 integration sites in clinical DNA samples obtained from cervical scrapes, with particular respect to the low quantity and poor quality of template DNA. This unique feature is of great interest for the incorporation of HPV16 DNA integration analysis into routine cervical screening programs, which continually provide pre-cancerous and cancerous lesions for investigation. Another potential DNA resource for ASP16 applications would be the large number of archival cervical biopsies collected previously and stored frozen over a long period of time, which therefore may not be compatible with the conventional methods applied for HPV integration analysis, such as DIPS-PCR, RS-PCR and APOT. In addition to viral-cellular junctions, the majority of pyrosequence reads from ASP16 experiments are purely HPV16 DNA sequences, altogether covering the complete HPV16 E1/E2 region. It has been suggested that mutations/variations of the HPV16 E2 gene associated with high-grade cervical pre-cancerous lesions and carcinomas may contribute to tumor progression by deregulating viral gene expression (Casas *et al*., 1999; Giannoudis *et al*., 2001). In this context, the huge amount of sequence data from ASP16 experiments can be used to detect sequence variations including point mutations, small deletions/insertions and gross DNA rearrangements of the HPV16 E1/E2 region. Based on the results of ASP16-2, one example of the mutation analysis has been illustrated by Nico Steinmeyer in his Diplomarbeit (Steinmeyer, 2009). In this case, mutation within E1 ORF has resulted in an additional recognition

site for a restriction enzyme, and the new site was confirmed by specific cleavage of PCR-amplicons containing the point mutation (Steinmeyer, 2009). Finally, after the establishment and optimization of ASP16, this strategy will be extended to other high-risk HPV types especially for HPV18 (then will be named ASP18), which has been detected as integrated form in almost all HPV18-positive cervical carcinomas analysed (Cullen *et al*., 1991; Hudelist *et al*., 2004; Vinokurova *et al*., 2008).

3.3 HPV16 DNA integration and insertional mutagenesis

It is apparent that HPV integration events may put intact cellular genes under the *cis*-control of viral regulatory elements. In some instances including *c-myc* discussed in *Section 3.1*, the aberrant regulation is strongly associated with activation of cellular proto-oncogene or inactivation of tumor suppressor gene (Reuter *et al*., 1998; Peter *et al*., 2006). Although a number of HPV integration sites have been identified in cervical carcinomas, less is known about the general role of insertional mutagenesis in cervical cancer progression. Based on the fact that integration sites are often mapped dozens to hundreds kb away from a protein-coding gene, some researchers argued that insertional mutagenesis might be a rare event and not important for cervical carcinogenesis (Wentzensen *et al*., 2002; Ziegert *et al*., 2003). Others, however, have argued that alterations of cellular gene expression caused by HPV integration may not be a rare event (Kraus *et al*., 2008). As seen from chromosomal rearrangements in Burkitt's lymphoma, gene expression can be modulated by enhancers located about 500 kb apart. Furthermore, it has also been reported that integrated HPV DNA is frequently found near microRNA-coding genes (Calin *et al*., 2004). These observations suggest that more efforts should be given in the future for investigating cellular candidate genes, whose expression may be affected by *cis*-elements of integrated HPV DNA.

In cervical carcinoma cell line MRI-H196, the integrated HPV16 DNA is mapped to chromosomal band *11p11.2* (see *Section 2.1.2*). The cellular gene *ptprj* (HGNC-ID: 9673) is located about 34 kb downstream of the 3' viral-cellular junction site (*Figure 3-3*). PTPRJ belongs to the protein tyrosine phosphatase (PTP) family composed of at least 37 members, including the well-known tumor suppressor protein PTEN (Tonks, 2006). Deletion of *ptprj* followed by loss of hetorozygosity (LOH) has been detected in a variety of lung, colon and breast cancers, indicating its tumor

suppressor activity (Ruivenkamp *et al*., 2002). However, gain of *ptprj* DNA copy number has also been reported recently for one glioma case, but without any information about the *ptprj* expression level (Preusser *et al*., 2007). The HPV integration sites published to date have not shown any connection with *ptprj* (Wentzensen *et al*., 2004; Kraus *et al*., 2008). In MRI-H196 cells, expression of *ptprj* might be upregulated by HPV16 enhancer elements, because the distance of about 34 kb is accessible for transcriptional *cis*-activation. On the other hand, there is also evidence from cell line ME180 implicating that tumor suppressor gene *APM-1* is silenced by promoter hypermethylation associated with HPV68 DNA integration into the *APM-1* gene (Elisabeth Schwarz, personal communication). The question here is whether this silencing effect may still be operative over a range of 34 kb in MRI-H196. Furthermore, it needs to be clarified whether the 1-Mb deletion of cellular DNA between the 5' and 3' viral-cellular junctions (see *Figure 2-4D*) is homozygous, because this region contains several apoptosis-related genes including *DDB2*, *MADD* and *PU.1*.

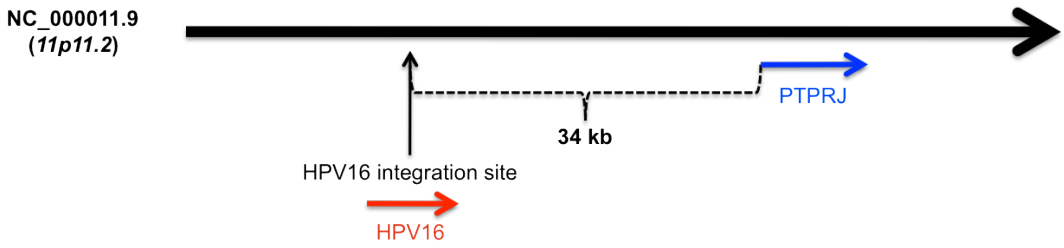


Figure 3-3: Schematic demonstration of the HPV16 DNA integration site in MRI-H196 and the genetic context near the junction site.

In cervical carcinoma 07c368 (2B5), the integrated HPV16 DNA is mapped to chromosomal band *1q43* (see *Figure 2-65*). The HPV16 integration site is located in the intron between exons 1 and 2 of the cellular gene *Grem2*. This gene encodes Gremlin2 protein, which is an antagonist of the bone morphogenetic protein (BMP) family essential for skeletal formation (Avsian-Kretchmer and Hsueh, 2004; Yanagita, 2005). BMPs belong to the tumor growth factor beta (TGFβ) superfamily, which either act as negative regulators of tumor growth or promote cell invasiveness and metastasis toward distant tissues (Massagué *et al*., 2000; Derynck *et al*., 2001; Tobin *et al*., 2002; Elliott and Blobe, 2005; Pardali and Moustakas, 2007). A growing number of evidence has shown the correlation between BMPs over-expression and osteoblastic bone metastasis in human prostate and breast cancers (Hamdy *et al*., 1997; Autzen *et al*., 1998; Masuda *et al*., 2003; Alarmo *et al*., 2008; Yuen *et al*.,

2008). In accordance with this correlation, several BMP antagonists including Noggin, Gremlin1 and SOSTDC1 have been reported to be downregulated in human malignancies (Suzuki *et al*., 2005; Schwaninger *et al*., 2007; Yuen *et al*., 2008; Blish *et al*., 2008; van Vlodrop *et al*., 2010). In some instances, the inactivation of *Gremlin1* gene has been linked with promoter hypermethylation (Suzuki *et al*., 2005; van Vlodrop *et al*., 2010). For sample 2B5, it needs to be clarified whether *Grem2* is inactivated and by which mechanism (LOH or methylation).

In cervical carcinoma 07c381 (2B6), the HPV16 DNA integration site is mapped to chromosomal band *12q13.3* (see *Figure 2-67*). The transcription unit of the cellular *R3HDM2* gene is disrupted by the integrated HPV16 DNA. It is impossible to speculate about the relevance to cervical carcinogenesis, because nothing is known about the function of R3HDM2 protein. By contrast, the cellular proto-oncogene *GLI1* located about 124 kb downstream of *R3HDM2* became more interesting (see *Figure 2-67B*). The GLI1 transcription factor is a member of the Kruppel zinc finger protein family, and belongs to the sonic hedgehog (SHH) signalling pathway (i Altaba *et al*., 2002; di Magliano and Hebrok, 2003). The *GLI1* gene was initially identified highly expressed in a malignant glioma, along with over 50-fold of *GLI1* DNA amplification (Kinzler *et al*., 1987). Later studies have frequently shown the upregulation of *GLI1* and activation of SHH pathway in basal cell carcinoma of the skin (Dahmane *et al*., 1997; Green *et al*., 1998; Nilsson *et al*., 2000). Recent data also suggest the role of *GLI1* over-expression in metastases from esophageal squamous cell carcinoma and pancreatic cancer (Mori *et al*., 2006; Feldmann *et al*., 2007). Thus, it would be interesting to investigate the expression level of *GLI1* and the metastasis status in sample 2B6.

Each of the HPV16 DNA integration sites identified in the present study has a unique feature of flanking cellular sequences. This is consistent with the results of other groups (Wentzensen *et al*., 2004; Kraus *et al*., 2008). The cellular proto-oncogenes or tumor suppressor genes located near or at the junction sites belong to different signalling pathways. Unlike the ubiquitous contributions of oncogenic E6 and E7 in cervical carcinogenesis, insertional mutagenesis of cellular genes may involve different molecular mechanisms. The potential insertional mutagenesis events implicated in this study, especially from samples 2B5 and 2B6, suggest a possible role of HPV-induced insertional mutagenesis in cervical cancer progression.

4. Materials and Methods

4.1 Chemical reagents

2'-deoxyadenosine-5'-triphosphate (dATP)	Roche Applied Science, Mannheim
2'-deoxycytidine-5'-triphosphate (dCTP)	Roche Applied Science, Mannheim
2'-deoxyguanosine-5'-triphosphate (dGTP)	Roche Applied Science, Mannheim
2'-deoxythymidine-5'-triphosphate (dTTP)	Roche Applied Science, Mannheim
Acetic acid glacial	Sigma-Aldrich, München
Agarose, for routine use	Sigma-Aldrich, München
Ammonium persulfate (APS), powder	Roche Applied Science, Mannheim
Ampicillin, sodium salt, powder	AppliChem, Darmstadt
Bacto agar	Becton Dickinson, Heidelberg
Bacto tryptone	Becton Dickinson, Heidelberg
Bacto yeast extract	Becton Dickinson, Heidelberg
Blue dextran	Sigma-Aldrich, München
β-mercaptoethanol	Sigma-Aldrich, München
Bovine Serum Albumin (BSA)	NEB, Frankfurt am Main
Bromophenol blue, sodium salt	Serva, Heidelberg
Chloroform	Roth, Karlsruhe
Coomassie Brilliant Blue G-250	Bio-Rad, München
DEPC	AppliChem, Darmstadt
Dimethylformamide (DMF)	Sigma-Aldrich, München
DMEM	Sigma-Aldrich, München
DMSO, cell culture grade	AppliChem, Darmstadt
EDTA, disodium salt dihydrate	Sigma-Aldrich, München
Ethanol absolute	Sigma-Aldrich, München
Ethidium bromide, powder	Sigma-Aldrich, München
Fetal Bovine Serum (FBS) Superior	Biochrom, Berlin
Ficoll 400	Serva, Heidelberg
Formaldehyde	Merck, Darmstadt
Formamide	Merck, Darmstadt
Glucose	Sigma-Aldrich, München
Glycerol	Sigma-Aldrich, München

Glycine	Sigma-Aldrich, München
HEPES, free acid	Sigma-Aldrich, München
Hydrochloric acid (HCl), 37% w/v	Sigma-Aldrich, München
Isoamyl alcohol	Sigma-Aldrich, München
Isopropanol	Sigma-Aldrich, München
Kanamycin sulfate, powder	AppliChem, Darmstadt
Lithium Chloride (LiCl)	Sigma-Aldrich, München
Magnesium chloride ($MgCl_2$)	Sigma-Aldrich, München
Magnesium sulfate ($MgSO_4$)	Sigma-Aldrich, München
Methanol	Sigma-Aldrich, München
Milk powdered, blotting grade	Roth, Karlsruhe
MOPS, free acid, powder	Sigma-Aldrich, München
NEN [α-^{32}P]dCTP	PerkinElmer, Rodgau-Jügesheim
NP-40	Sigma-Aldrich, München
Penicillin-streptomycin, liquid	Invitrogen, Karlsruhe
Phenol	Roth, Karlsruhe
Polyvinylpyrrolidone	Sigma-Aldrich, München
Potassium chloride (KCl), powder	Sigma-Aldrich, München
Potassium phosphate monobasic (KH_2PO_4)	Sigma-Aldrich, München
Proteinase K	Merck, Darmstadt
Redivue [α-^{32}P]dCTP	GE Healthcare Europe, München
RNase A	Qiagen, Hilden
RNase*Zap* spray	Sigma-Aldrich, München
Rotiphorese Gel 30 (37.5:1)	Roth, Karlsruhe
RPMI	Sigma-Aldrich, München
Sarkosyl, sodium salt	Sigma-Aldrich, München
Sodium acetate, powder	Sigma-Aldrich, München
Sodium chloride (NaCl), powder	Sigma-Aldrich, München
Sodium citrate dihydrate	Sigma-Aldrich, München
Sodium deoxycholate	Sigma-Aldrich, München
Sodium dodecyl sulfate (SDS), powder	Sigma-Aldrich, München
Sodium hydroxide (NaOH), powder	Sigma-Aldrich, München

Sodium phosphate dibasic (Na_2HPO_4)	Sigma-Aldrich, München
TEMED	Sigma-Aldrich, München
Trizma base, powder	Sigma-Aldrich, München
TRIzol Reagent	Invitrogen, Karlsruhe
t-RNA	Sigma-Aldrich, München
Trypsine/EDTA	Invitrogen, Karlsruhe
Tween 20	Sigma-Aldrich, München
UltraPure water, DNase/RNase-free	Invitrogen, Karlsruhe
X-Gal	Sigma-Aldrich, München
Xylene cyanol FF, sodium salt	Serva, Heidelberg

4.2 Buffer solutions (frequently used)

6× BPB DNA-loading buffer

0.25% (w/v) bromophenol blue
30% (v/v) glycerol
store at 4 °C
(BPB runs in the agarose gel
at the position of about 0.5 kb)

6× XC DNA-loading buffer

0.25% (w/v) xylene cyanol
30% (v/v) glycerol
store at 4 °C
(XC runs in the agarose gel
at the position of about 5 kb)

1 M Tris-HCl

1 M Trizma base
adjust pH with HCl
autoclave

0.5 M EDTA, pH 8.0

0.5 M EDTA
adjust pH with NaOH
autoclave

10× TE, pH 8.0

100 mM Tris-HCl, pH 8.0
10 mM EDTA, pH 8.0
autoclave

50× TAE buffer, per liter

242 g of Trizma base
57.1 ml of glacial acetic acid
100 ml of 0.5 M EDTA, pH 8.0
dissolve in deionized water

10× MOPS electrophoresis buffer

0.2 M MOPS
20 mM sodium acetate
10 mM EDTA, pH 8.0
dissolve in DEPC-treated water
filter-sterilized, store in dark

1× RNA-loading buffer

50% (v/v) formamide
2.2 M formaldehyde
1× MOPS-buffer
1% (v/v) Ficoll 400
0.02% (w/v) bromophenol blue

20× SSC

3 M NaCl
0.3 M sodium citrate
adjust pH to 7.0 with HCl
autoclave

3 M sodium acetate, pH 5.2

3 M sodium acetate
adjust pH with glacial acetic acid
autoclave

10× PBS, per liter

80 g NaCl
2 g KCl
14.4 g Na_2HPO_4
2.4 g KH_2PO_4
adjust pH to 7.4 with HCl
autoclave

10× TBS, per liter

80 g NaCl
2 g KCl
30 g Trizma base
adjust pH to 7.4 with HCl
autoclave

5× SDS-PAGE sample buffer

10% (w/v) SDS
10 mM β-mercaptoethanol
20% (v/v) glycerol
0.2 M Tris-HCl, pH 6.8
0.05% (w/v) bromophenol blue

5× SDS-PAGE running buffer

125 mM Trizma base
1 M glycine
0.5% (w/v) SDS

10× Western blotting buffer

250 mM Trizma base
2 M glycine

Western stripping buffer

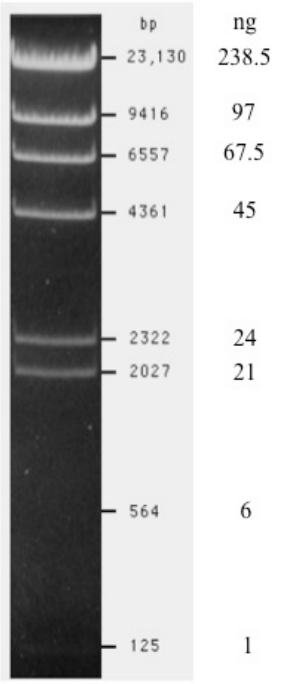
62.5 mM Tris-HCl, pH 6.8
2% (w/v) SDS
0.1 M β-mercaptoethanol

4.3 Molecular weight size markers

4.3.1 DNA markers

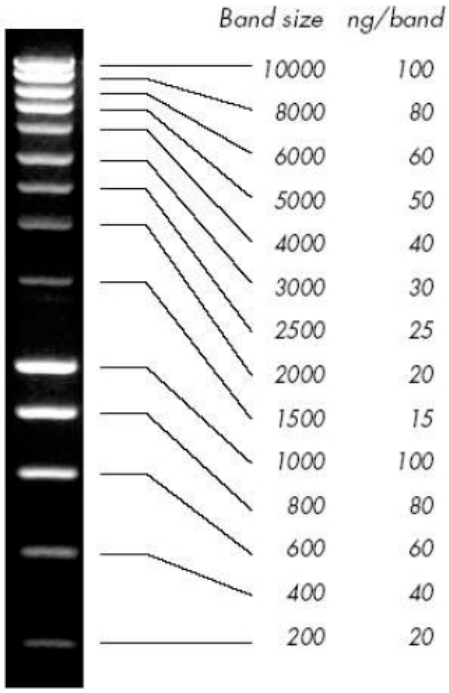
λ/*Hind* III (500 ng/lane)

(Invitrogen, Karlsruhe)



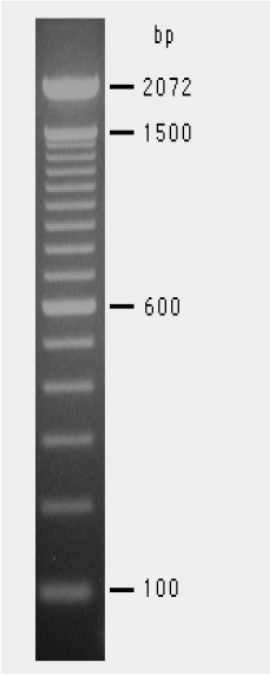
SmartLadder (5 µl/lane)

(Eurogentec, Köln)



100 bp DNA Ladder (500 ng/lane; not for DNA quantitation)

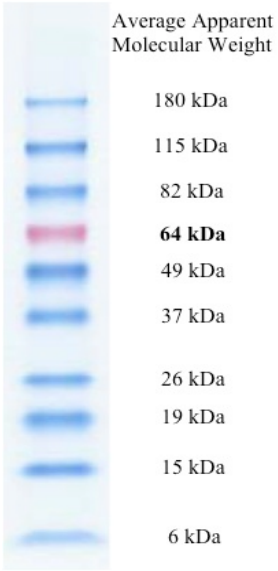
(Invitrogen, Karlsruhe)



4.3.2 Protein markers

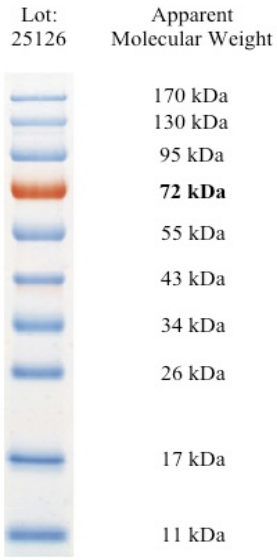
BenchMark prestained protein ladder

(Invitrogen, Karlsruhe)



PageRuler prestained protein ladder

(Fermentas, St. Leon-Rot)



4.4 Culture media

4.4.1 Bacterial culture media

LB medium

1% (w/v) Bacto tryptone
0.5% (w/v) Bacto yeast extract
1% (w/v) NaCl
dissolve in distilled water
pH = 7.0 (adjust with 5 N NaOH)
autoclave at 121 °C, store at RT

LB agar (with selective antibiotics)

LB contents + 1.5% (w/v) Bacto agar
dissolve in water, adjust pH, autoclave
cool to 50–60 °C, add antibiotics*
pour into petri dishes, 25 ml/plate
allow agar to solidify at RT
store plates inverted at 4 °C

(* final concentration: 100 µg/ml of ampicillin or 50 µg/ml of kanamycin)

SOB medium

2% (w/v) Bacto tryptone
0.5% (w/v) Bacto yeast extract
10 mM NaCl
2.5 mM KCl
adjust pH to 7.0, autoclave
add filter-sterilized $MgCl_2$/$MgSO_4$ (10 mM each)

SOC medium

SOB medium
20 mM glucose (filter-sterilized)
aliquot, store at –20 °C

4.4.2 Cell culture media

Complete DMEM culture medium

500 ml DMEM
50 ml FBS
5 ml Penicillin-streptomycin*

Complete RPMI culture medium

500 ml RPMI
50 ml FBS
5 ml Penicillin-streptomycin*

(* contains 50000 units of penicillin and 50000 µg of streptomycin in total)

Cell freezing medium

80% (v/v) complete cell culture medium
10% (v/v) FBS
10% (v/v) DMSO, cell culture grade

4.5 Restriction endonucleases (frequently used)

Enzyme	Recognition Sequence	Reaction Condition	Manufacturer
*Eco*R I (FastDigest)	5'-G▼AATTC-3' 3'-CTTAA▲G-5'	1× FastDigest buffer; at 37°C	Fermentas (St. Leon-Rot)
Hind III	5'-A▼AGCTT-3' 3'-TTCGA▲A-5'	1× React 2 buffer; at 37°C	Invitrogen (Karlsruhe)
Nco I	5'-C▼CATGG-3' 3'-GGTAC▲C-5'	1× NEBuffer 3; at 37°C	NEB (Frankfurt am Main)
Xba I	5'-T▼CTAGA-3' 3'-AGATC▲T-5'	1× React 2 buffer; at 37°C	Invitrogen (Karlsruhe)

4.6 Antibodies

4.6.1 Primary antibodies

Anti-c-MYC, clone 9E10 (Roche Applied Science, Mannheim)

Mouse monoclonal antibody (IgG); Working concentration: 1 µg/ml for Western blot;
Epitope: the C-terminus of human c-MYC protein (AA 410–419, EQKLISEEDL)

Anti-c-MYC, clone Y69 (Epitomics, Burlingame, U.S.A)

Rabbit monoclonal antibody (IgG); Working dilution: 1:10000 for Western blot;
Epitope: the N-terminus of human c-MYC protein (AA 1–20, MPLNVSFTNRNYDLDYDSVQ)

Anti-GAPDH (Abcam, Cambridge, UK)

Rabbit polyclonal antibody (IgG); Working dilution: 1:1000 for Western blot;
Immunogen: human GAPDH protein (native, full length)

4.6.2 Secondary antibodies

Peroxidase-conjugated AffiniPure Goat Anti-Mouse IgG (H+L)
(Jackson ImmunoResearch Europe, Suffolk, UK)
Working dilution: 1:10000 after rehydration in 1.5 ml of distilled water

Peroxidase-conjugated AffiniPure Goat Anti-Rabbit IgG (H+L)
(Jackson ImmunoResearch Europe, Suffolk, UK)
Working dilution: 1:10000 after rehydration in 1.5 ml of distilled water

4.7 Commercial kits

Advantage 2 PCR Kit	Becton Dickinson, Heidelberg
Chemiluminescence Detection Kit	AppliChem, Darmstadt
Dynabeads Oligo $(dT)_{25}$	Invitrogen, Karlsruhe
Expand HiFi^{PLUS} PCR Kit	Roche Applied Science, Mannheim
Expand Long Range PCR Kit	Roche Applied Science, Mannheim
FastStart Taq DNA Polymerase Kit	Roche Applied Science, Mannheim
GenomePlex WGA 2 Kit	Sigma-Aldrich, München
illustra MicroSpin G-25 Columns	GE Healthcare Europe, München
Paq5000 DNA Polymerase Kit	Stratagene, La Jolla, U.S.A
Qiagen Multiplex PCR Kit	Qiagen, Hilden
QIAquick Gel Extraction Kit	Qiagen, Hilden
QIAquick PCR Purification Kit	Qiagen, Hilden
Rediprime II DNA Labeling System	GE Healthcare Europe, München
SMART RACE cDNA Amplification Kit	Becton Dickinson, Heidelberg
StrataClone PCR Cloning Kit	Stratagene, La Jolla, U.S.A
SuperScript III Reverse Transcriptase Kit	Invitrogen, Karlsruhe
Taq DNA Polymerase Kit	Invitrogen, Karlsruhe
TOPO TA PCR Cloning Kit	Invitrogen, Karlsruhe

4.8 Consumables

BioMag streptavidin beads	Polysciences, Heidelberg
Biomax Light film	GE Healthcare Europe, München
Biomax MR film	GE Healthcare Europe, München
Cell culture flasks (25, 75, 175 cm^2)	Greiner, Frickenhausen
Cell scraper, plastic	Greiner, Frickenhausen
Cuvettes, plastic	Eppendorf, Hamburg
Falcon tube (15 and 50 ml)	Becton Dickinson, Heidelberg
Hybond-N+ nylone membrane	GE Healthcare Europe, München
Immobilon-P transfer membrane	Millipore, Schwalbach/Ts.
Microcentrifuge tube (1.5 & 2ml)	Steinbrenner, Wisenbach
Pasteur piptte, plastic	Neolab, Heidelberg
PCR-tube, 0.5 ml, thin-walled	Neolab, Heidelberg
Petri dish (for bacteria)	Greiner, Frickenhausen
Pipetting tips	Steinbrenner, Wisenbach
PP-tube (14 ml)	Greiner, Frickenhausen
Scalpel	Sterimed, Püttlingen
Sterile filter (0.22 & 0.45 µM)	Millipore, Schwalbach/Ts.
Whatman 3MM paper	Schleicher & Schuell, Dassel

4.9 Laboratory instruments

Biofuge Pico	Heraeus, Hanau
CO_2-incubator for cell culture	Heraeus, Hanau
Curix 60 film processor	AGFA Healthcare, Köln
Electrophoresis chamber	Renner, Dannstadt &
(for agarose gel)	Peqlab, Erlangen
Film cassettes	GE Healthcare Europe, München
Heating block	Eppendorf, Hamburg
Magnetic particle concentrator	Invitrogen, Karlsruhe
Megafuge 1.0R	Heraeus, Hanau

Mini-PROTEAN II Electrophoresis Cell	Bio-Rad, München
Mini Trans-Blot Transfer Cell	Bio-Rad, München
NanoDrop 1000 Spectrophotometer	Thermo Fisher, Waltham, U.S.A
Pipettes	Gilson, Düsseldorf
Thermocycler PTC-200	MJ Research, Waltham, U.S.A
Ultrospec 3000 pro	GE Healthcare Europe, München
Vortex	Bender & Hobein, Zürich, Switzerland

4.10 Software

HUSAR: Heidelberg Unix Sequence Analysis Resources (http://www.dkfz.de/menu/biounit)

NCBI: National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov)

Ensemble: European Bioinformatics Institute (EBI) databases (http://www.ensembl.org)

Computer programs for ASP16 strategy: written by Sasithorn Chotewutmontri (unpublished, ongoing PhD thesis, DKFZ)

4.11 Genome reference resources

4.11.1 Human genome reference sequences

Version: NCBI Build 36.3 (March 2008) and Build 37.1 (August 2009)

Organism: *Homo Sapiens*

NCBI taxonomy ID: 9606

4.11.2 HPV16 genome reference sequence

The complete HPV16 DNA sequence was originally published in 1985 (Seedorf *et al*., 1985; GenBank accession number: K02718). This sequence was revised according to the corrections made by Los Alamos National Laboratory in 1995 and 1997. The final version of the revised HPV16 reference sequence (version 1997) is shown in *Appendix 3*.

4.12 Oligonucleotides

4.12.1 HPV16 primers for routine PCR

All primers were synthesized by Thermo Fisher Scientific (Ulm). The number in the name indicates the start position of each primer according to the HPV16 reference sequence. (F = forward; R = reverse)

HPV16-103-F:	5′-AATGTTTCAGGACCCACAGG-3′
HPV16-209-R:	5′-GTTGCTTGCAGTACACACATTC-3′
HPV16-773-R:	5′-CTTTGTACGCACAACCGAAG-3′
HPV16-811-F:	5′-ATGGGCACACTAGGAATTGTG-3′
HPV16-817-R:	5′-TGCCCATTAACAGGTCTTCC-3′
HPV16-856-F:	5′-TAATCTACCATGGCTGATCCTG-3′
HPV16-1098-R:	5′-TTGTTTTGCTTCCTGTGCAG-3′
HPV16-1333-R:	5′-ATGGTGTTTCAGTCTCATGGC-3′
HPV16-1470-F:	5′-TGCAAAGGCAGCAATGTTAG-3′
HPV16-1551-F:	5′-TAAATCAACGTGTTGCGATTG-3′
HPV16-1565-F:	5′-GCGATTGGTGTATTGCTGC-3′
HPV16-1692-R:	5′-TAACACAACCATTCCCCATG-3′
HPV16-1793-F:	5′-CAAAATTGCGTAGTACAGCAGC-3′
HPV16-1805-R:	5′-CTACGCAATTTTGGAGGCTC-3′
HPV16-2171-F:	5′-GAGGTGATTGGAAGCAAATTG-3′
HPV16-2186-R:	5′-TGCTTCCAATCACCTCCATC-3′
HPV16-2450-R:	5′-CAACAGGGCACTGTAGCATC-3′
HPV16-2573-F:	5′-CTGGTACAGATTCTAGGTGGCC-3′
HPV16-2701-F:	5′-TTCTCAAGGACGTGGTCCAG-3′
HPV16-2879-R:	5′-TAGCACATTCTAGGCGCATG-3′
HPV16-2882-F:	5′-TATTACAAGGCCAGAGAAATGG-3′
HPV16-3384-F:	5′-TTATTAGGCAGCACTTGGCC-3′
HPV16-5125-R:	5′-TAGAGGTTAATGCTGGCCTATG-3′
HPV16-5398-R:	5′-TAGAGGGTACAGATGGTACCGG-3′
HPV16-5714-R:	5′-CATCCGTGCTTACAACCTTAG-3′
HPV16-6448-R:	5′-ATTTTCACCAACAGTACCAGCC-3′
HPV16-6955-R:	5′-CATCTTCTTTAGGTGCTGGAGG-3′

4.12.2 *myc* primers for routine PCR

All primers were synthesized by Thermo Fisher Scientific (Ulm). The number in the name indicates the start position of each primer according to the human genome reference sequence of chromosome 8 (Accession number: NC_000008.9). (F = forward; R = reverse)

128744774-F:	5′-AGGGACTCAATGAGTATGCATG-3′
128745226-R:	5′-ACCCATCAAATCCCTCTGC-3′
128815697-F:	5′-TTAAAGCTGAATTGTGCAGTGC-3′
128816053-R:	5′-ACGCTTCGACTTAGCTAGTTGC-3′
128816092-R:	5′-ACTGTATGTAACCCGCAAACG-3′
128816664-F:	5′-AACAGGCAGACACATCTCAGG-3′
128817197-F:	5′-TCAAACAGTACTGCTACGGAGG-3′
128817576-F:	5′-AGGGCTTCTCAGAGGCTTG-3′
128817900-R:	5′-TGTAAGTTCCAGTGCAAAGTGC-3′
128818442-F:	5′-TGGAGTAGGGACCGCATATC-3′
128819700-F:	5′-AACGTTAGCTTCACCAACAGG-3′
128819716-R:	5′-TTGGTGAAGCTAACGTTGAGG-3′
128821823-F:	5′-GAGGAGGAACAAGAAGATGAGG-3′
128821873-R:	5′-CTTTTCCACAGAAACAACATCG-3′
128822955-R:	5′-TCTTAAGACCACCCTCTGAAGG-3′

4.12.3 Primers for restriction-site PCR

All primers were synthesized by Thermo Fisher Scientific (Ulm). For HPV16 primers, the number in the name indicates the start position of each primer according to the HPV16 reference sequence. All primer sequences were taken from Thorland *et al* (2000). Names and compositions of the RSO primers are explained in *Section 2.1*. (D = downstream or forward; U = upstream or reverse)

```
HPV16-1546-26D:   5'-AGTAATAAATCAACGTGTTGCGATTG-3'
HPV16-1588-26D:   5'-GGACTTACACCCAGTATAGCTGACAG-3'
HPV16-2387-25D:   5'-TTTGTTTACAACCATTAGCAGATGC-3'
HPV16-6688-25U:   5'-AGTAGATATGGCAGCACATAATGAC-3'
HPV16-6653-25U:   5'-TGCGTGTAGTATCAACAACAGTAAC-3'
HPV16-5885-27U:   5'-ACTTATTGGGGTCAGGTAAATGTATTC-3'

RSO-BamHI:        5'-TAATACGACTCACTATAGGGAGANNNNNNNNNNGGATCC-3'
RSO-NheI:         5'-TAATACGACTCACTATAGGGAGANNNNNNNNNNGCTAGC-3'
RSO-SacI:         5'-TAATACGACTCACTATAGGGAGANNNNNNNNNNGAGCTC-3'
RSO-SalI:         5'-TAATACGACTCACTATAGGGAGANNNNNNNNNNGTCGAC-3'
RSO-NdeI:         5'-TAATACGACTCACTATAGGGAGANNNNNNNNNNCATATG-3'
RSO-PvuI:         5'-TAATACGACTCACTATAGGGAGANNNNNNNNNNCGATCG-3'
RSO-XbaI:         5'-TAATACGACTCACTATAGGGAGANNNNNNNNNNTCTAGA-3'
RSO-XhoI:         5'-TAATACGACTCACTATAGGGAGANNNNNNNNNNCTCGAG-3'
T7-primer:        5'-TAATACGACTCACTATAGGGAGA-3'
```

4.12.4 Primers for rapid amplification of cDNA ends

All primers were synthesized by Thermo Fisher Scientific (Ulm). For HPV16 primers, the number in the name indicates the start position of each primer according to the HPV16 reference sequence. For *myc* primers, the number indicates the start position of each primer according to the human genome reference sequence of chromosome 8 (Accession number: NC_000008.9). (F = forward; R = reverse)

```
HPV16-683-F:      5'-CAGCTGGACAAGCAGAACCGGACAG-3'
HPV16-855-R:      5'-TGGTTTCTGAGAACAGATGGGGCAC-3'
HPV16-3386-F:     5'-ATTAGGCAGCACTTGGCCAACCACC-3'
HPV16-3565-R:     5'-TGCAGTGAGGATTGGAGCACTGTCC-3'

myc-128817671-F:  5'-TAATTCCAGCGAGAGGCAGAGGGAG-3'
myc-128818021-R:  5'-CGTCTAAGCAGCTGCAAGGAGAGCC-3'
myc-128819726-F:  5'-TGACCTCGACTACGACTCGGTGCAG-3'
myc-128820371-R:  5'-CGTCGAGGAGAGCAGAGAATCCGAG-3'
```

Sequences of the following primers were taken from user manual of the SMART RACE cDNA amplification kit.

```
5'-CDS primer:    5'-(T)25VN-3'  (V = A, C or G)
3'-CDS primer A:  5'-AAGCAGTGGTATCAACGCAGAGTAC(T)30VN-3'
UPM-long :        5'-CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT-3'
UPM-short :       5'-CTAATACGACTCACTATAGGGC-3'
```

4.12.5 Primers for ASP16 strategy

All primers for ASP16 strategy were synthesized by Sigma-Aldrich (Steinheim am Albuch). Details about the primer combinations are explained in *Section 2.11* and *2.12*.

The following 24 primers were labeled with biotin at the 5' end (5B), and contain only HPV16 sequences. Start positions are given as numbers in primer names according to the HPV16 reference sequence. (EP = early primer; LP = late primer; F = forward; R = reverse)

```
EP-1(5B-810F):     5'-AATGGGCACACTAGGAATTGTG-3'
EP-2(5B-1261F):    5'-GGGTATGGCAATACTGAAGTGG-3'
EP-3(5B-1562F):    5'-GTTGCGATTGGTGTATTGCTG-3'
EP-4(5B-1938F):    5'-ACAGATGGTACAATGGGCCTAC-3'
EP-5(5B-2389F):    5'-TGGTTACAACCATTAGCAGATGC-3'
EP-6(5B-2705F):    5'-CAAGGACGTGGTCCAGATTAAG-3'
EP-7(5B-3101F):    5'-ACAGTGGAAGTGCAGTTTGATG-3'
EP-8(5B-3542F):    5'-GACAGTGCTCCAATCCTCACTG-3'

EP-9(5B-1046F):    5'-AGGCAGAAACAGAGACAGCAC-3'
EP-10(5B-1394F):   5'-GAGAGGGTGTTAGTGAAAGACAC-3'
EP-11(5B-1760F):   5'-GTGTGTCTCCAATGTGTATGATG-3'
EP-12(5B-2151F):   5'-ATGTGATAGGGTAGATGATGGAG-3'
EP-13(5B-2539F):   5'-TGCCCTCCATTATTAATTACATC-3'
EP-14(5B-2912F):   5'-CATATTAACCACCAAGTGGTGC-3'
EP-15(5B-3318F):   5'-CGGGTGGTCAGGTAATATTATG-3'
EP-16(5B-3762F):   5'-CATATGATAGTGAATGGCAACG-3'

LP-1(5B-6951R):    5'-TCTTTAGGTGCTGGAGGTGTATG-3'
LP-2(5B-6552R):    5'-TGGGCATCAGAGGTAACCATAG-3'
LP-3(5B-6186R):    5'-CCTGGATTTACTGCAACATTGG-3'
LP-4(5B-5766R):    5'-AGTAGTCTGGATGTTCCTGCATG-3'
LP-5(5B-5318R):    5'-ATAGAAGTAGGTGAGGCTGCATG-3'
LP-6(5B-4917R):    5'-TAATCCTAGGCGTGCCACTG-3'
LP-7(5B-4501R):    5'-CAGGAGCAAGTGTATCTGTAGCTG-3'
LP-8(5B-4290R):    5'-AAGTTGGGTAGCCGATGCAC-3'
```

The following 24 primers contain the 19-nt Roche-A (RA) sequence as 5' parts and HPV16 sequences as 3' parts. Start positions of the HPV16 parts are given as numbers in primer names according to the HPV16 reference sequence. (EP-N = nested early primer; LP-N = nested late primer; F = forward; R = reverse)

```
EP-N1(RA-857F):    5'-GCCTCCCTCGCGCCATCAG AATCTACCATGGCTGATCCTG-3'
EP-N2(RA-1275F):   5'-GCCTCCCTCGCGCCATCAG TGAAGTGGAAACTCAGCAGATG-3'
EP-N3(RA-1576F):   5'-GCCTCCCTCGCGCCATCAG ATTGCTGCATTTGGACTTACAC-3'
EP-N4(RA-1951F):   5'-GCCTCCCTCGCGCCATCAG TGGGCCTACGATAATGACATAG-3'
EP-N5(RA-2403F):   5'-GCCTCCCTCGCGCCATCAG AGCAGATGCCAAAATAGGTATG-3'
EP-N6(RA-2723F):   5'-GCCTCCCTCGCGCCATCAG TAAGTTTGCACGAGGACGAG-3'
EP-N7(RA-3121F):   5'-GCCTCCCTCGCGCCATCAG TGGAGACATATGCAATACAATGC-3'
EP-N8(RA-3555F):   5'-GCCTCCCTCGCGCCATCAG TCCTCACTGCATTTAACAGCTC-3'

EP-N9(RA-1064F):   5'-GCCTCCCTCGCGCCATCAG CACATGCGTTGTTTACTGCAC-3'
EP-N10(RA-1411F):  5'-GCCTCCCTCGCGCCATCAG AGACACACTATATGCCAAACACC-3'
EP-N11(RA-1785F):  5'-GCCTCCCTCGCGCCATCAG AGAGCCTCCAAAATTGCGTAG-3'
EP-N12(RA-2174F):  5'-GCCTCCCTCGCGCCATCAG GTGATTGGAAGCAAATTGTTATG-3'
EP-N13(RA-2569F):  5'-GCCTCCCTCGCGCCATCAG AATGCTGGTACAGATTCTAGGTG-3'
EP-N14(RA-2933F):  5'-GCCTCCCTCGCGCCATCAG CCAACACTGGCTGTATCAAAG-3'
```

```
EP-N15(RA-3339F): 5'-GCCTCCCTCGCGCCATCAG GTCCTACATCTGTGTTTAGCAGC-3'
EP-N16(RA-3778F): 5'-GCCTCCCTCGCGCCATCAG GCAACGTGACCAATTTTTGTC-3'

LP-N1(RA-6918R):  5'-GCCTCCCTCGCGCCATCAG GCAATTGCCTGGGATGTTAC-3'
LP-N2(RA-6505R):  5'-GCCTCCCTCGCGCCATCAG TGAACTGGCTAAATTTGCAGTAG-3'
LP-N3(RA-6139R):  5'-GCCTCCCTCGCGCCATCAG TTCCCCTATAGGTGGTTTGC-3'
LP-N4(RA-5729R):  5'-GCCTCCCTCGCGCCATCAG TGCGTGCAACATATTCATCC-3'
LP-N5(RA-5304R):  5'-GCCTCCCTCGCGCCATCAG GGCTGCATGTGAAGTGGTAG-3'
LP-N6(RA-4890R):  5'-GCCTCCCTCGCGCCATCAG CCCTGGTATGGGTGTGCTAC-3'
LP-N7(RA-4466R):  5'-GCCTCCCTCGCGCCATCAG GTTCCCAATGGAATATACCCAG-3'
LP-N8(RA-4254R):  5'-GCCTCCCTCGCGCCATCAG AGAACGTTTGTGTCGCATTG-3'
```

The following 24 primers contain the 19-nt Roche-B (RB) sequence as 5' parts and the 18-nt GenomePlex universal adapter (GPUA) sequence as 3' parts. Between RB and GPUA, there are 24 different 4-nt barcode sequences unique for each primer.

```
RB-GPUA-1:        5'-GCCTTGCCAGCCCGCTCAG TGAC TGTGTTGGGTGTGTTTGG-3'
RB-GPUA-2:        5'-GCCTTGCCAGCCCGCTCAG AGAC TGTGTTGGGTGTGTTTGG-3'
RB-GPUA-3:        5'-GCCTTGCCAGCCCGCTCAG TCAC TGTGTTGGGTGTGTTTGG-3'
RB-GPUA-4:        5'-GCCTTGCCAGCCCGCTCAG ACAC TGTGTTGGGTGTGTTTGG-3'
RB-GPUA-5:        5'-GCCTTGCCAGCCCGCTCAG TGTC TGTGTTGGGTGTGTTTGG-3'
RB-GPUA-6:        5'-GCCTTGCCAGCCCGCTCAG AGTC TGTGTTGGGTGTGTTTGG-3'
RB-GPUA-7:        5'-GCCTTGCCAGCCCGCTCAG TCTC TGTGTTGGGTGTGTTTGG-3'
RB-GPUA-8:        5'-GCCTTGCCAGCCCGCTCAG ACTC TGTGTTGGGTGTGTTTGG-3'
RB-GPUA-9:        5'-GCCTTGCCAGCCCGCTCAG CTGA TGTGTTGGGTGTGTTTGG-3'
RB-GPUA-10:       5'-GCCTTGCCAGCCCGCTCAG CAGA TGTGTTGGGTGTGTTTGG-3'
RB-GPUA-11:       5'-GCCTTGCCAGCCCGCTCAG CTCA TGTGTTGGGTGTGTTTGG-3'
RB-GPUA-12:       5'-GCCTTGCCAGCCCGCTCAG CACA TGTGTTGGGTGTGTTTGG-3'
RB-GPUA-13:       5'-GCCTTGCCAGCCCGCTCAG TAGC TGTGTTGGGTGTGTTTGG-3'
RB-GPUA-14:       5'-GCCTTGCCAGCCCGCTCAG ATGC TGTGTTGGGTGTGTTTGG-3'
RB-GPUA-15:       5'-GCCTTGCCAGCCCGCTCAG TACA TGTGTTGGGTGTGTTTGG-3'
RB-GPUA-16:       5'-GCCTTGCCAGCCCGCTCAG ATCA TGTGTTGGGTGTGTTTGG-3'
RB-GPUA-17:       5'-GCCTTGCCAGCCCGCTCAG TGCA TGTGTTGGGTGTGTTTGG-3'
RB-GPUA-18:       5'-GCCTTGCCAGCCCGCTCAG TCGA TGTGTTGGGTGTGTTTGG-3'
RB-GPUA-19:       5'-GCCTTGCCAGCCCGCTCAG AGCA TGTGTTGGGTGTGTTTGG-3'
RB-GPUA-20:       5'-GCCTTGCCAGCCCGCTCAG ACGA TGTGTTGGGTGTGTTTGG-3'
RB-GPUA-21:       5'-GCCTTGCCAGCCCGCTCAG CATC TGTGTTGGGTGTGTTTGG-3'
RB-GPUA-22:       5'-GCCTTGCCAGCCCGCTCAG CTAC TGTGTTGGGTGTGTTTGG-3'
RB-GPUA-23:       5'-GCCTTGCCAGCCCGCTCAG CTGC TGTGTTGGGTGTGTTTGG-3'
RB-GPUA-24:       5'-GCCTTGCCAGCCCGCTCAG CAGC TGTGTTGGGTGTGTTTGG-3'
```

4.12.6 Primers for junction-specific PCR of clinical samples

All primers were synthesized by Thermo Fisher Scientific (Ulm). The number in the name indicates the start position of each primer according to the human genome reference sequence of chromosome 1 (NC_000001.10) or chromosome 12 (NC_000012.11). (F = forward; R = reverse)

```
240760735-R:      5'-TTATTCTCAGCAGTTCCCAGC-3' (NC_000001.10)
240760912-R:      5'-AGCAGCTCCCAGGTCACTAG-3'  (NC_000001.10)
240760944-F:      5'-ATATTCCCAATGCTTGGCAC-3'  (NC_000001.10)
240761029-F:      5'-AAGTTCTAAGCGGATGGCTG-3'  (NC_000001.10)

57686174-F:       5'-TCTGCAGCTTCCCAATAAGC-3'  (NC_000012.11)
57686333-R:       5'-TGTCAGCCCATAGCTTTTAGC-3' (NC_000012.11)
```

4.13 Human continuous cell lines

All cell lines were provided by Elisabeth Schwarz and were obtained originally from the indicated sources.

CaSki:	cervical carcinoma,	HPV16-positive,	ATCC
MRI-H186:	cervical carcinoma,	HPV16-positive,	ATCC
MRI-H196:	cervical carcinoma,	HPV16-positive,	ATCC
SiHa:	cervical carcinoma,	HPV16-positive,	ATCC
C4-I:	cervical carcinoma,	HPV18-positive,	ATCC
C4-II:	cervical carcinoma,	HPV18-positive,	ATCC
HeLa:	cervical carcinoma,	HPV18-positive,	ATCC
C33A:	cervical carcinoma,	HPV-negative,	ATCC

HaCat: spontaneously immortalized keratinocytes, HPV-negative, Petra Boukamp (DKFZ)

4.14 Clinical DNA samples

The clinical DNA samples for HPV16 integration analysis were provided by our cooperation partners in Reims and Besançon. There, cervical scrapes were obtained from women participating in routine cervical screening programs. HPV16-positive high-grade cervical pre-cancerous lesions and carcinomas were sent to our group in Heidelberg for molecular determination of HPV16 DNA integration sites. Performed in Reims and Besançon, DNA was isolated from clinical samples using the EZ1 DNA tissue kit and BioRobot EZ1 from Qiagen (Hilden), according to the manufacturer's instructions.

4.15 *In vitro* cultivation of mammalian adherent cells

Taken from liquid-nitrogen tank, a vial of frozen cells (1 ml) was thawed quickly to 37°C. In a laminar flow hood, a 75-cm^2 tissue culture flask was filled with 14 ml complete cell culture medium pre-warmed to 37°C. Cells were suspended by repeated pipetting, transferred into the flask and mixed by gently shaking. Cell suspension was spread on flask bottom and then horizontally incubated at 37°C, 5% CO_2 and 95% relative humidity. Under this condition, adherent cells grow and expand to a monolayer culture. Cells were observed daily under a microscope, and medium was renewed every 2–3 days. For splitting, cells were rinsed briefly with 5 ml pre-warmed trypsine/EDTA solution, and incubated with another 3 ml solution at 37°C for minutes. After detachment from the flask bottom, cells were diluted with 7 ml of complete culture medium and mixed well by pipetting intensively. Dependent on the desired dilution rates, an appropriate volume of the cell suspension was combined with fresh medium to a total volume of 15 ml.

Tissue culture cells can be preserved in liquid nitrogen vapor phase over a long time. One 175-cm^2 flask of cells was trypsinized and centrifuged at 800 rpm for 10 min at 4°C. The

supernatant was discarded and the pellet was resuspended in 10 ml of complete medium. Centrifuged again, medium was removed and cells were dissolved with 5 ml of freezing medium. The cell suspension was mixed well and transferred into five sterile cryogenic vials (1 ml each). Cells were cooled slowly down to –80°C and stored in liquid-nitrogen tank.

4.16 Genomic DNA isolation from cultured mammalian cells

Medium was decanted and cells were washed twice with ice-cold PBS/EDTA. For a 175-cm^2 flask, 6 ml PBS/EDTA (1× PBS; 2 mM EDTA) and 3 ml of 3× lysis buffer (2× TE, pH 8.0; 3% sarkosyl, w/v) was applied. After shaking gently and briefly, the viscous lysate was collected into a 50 ml Falcon tube. Proteinase K was added at a final concentration of 0.2 mg/ml and incubated overnight at 55°C. Reaction was cooled down to RT, and 9 ml phenol* (50% phenol, 50% CIA) was added for deproteination. The reaction tube was placed into a mixer and rotated for 30 min. Centrifuged at 2000 rpm for 10 min at RT, the mixture was separated into a nucleic-acid containing upper aqueous phase and a lower organic phase. The aqueous phase (without interphase) was transferred into a new 50 ml Falcon tube using a plastic Pasteur pipette. To remove the residual phenol, CIA (chloroform : isoamyl alcohol = 24:1) was added into the aqueous phase with the same volume, rotated for 30 min and centrifuged for 10 min. The aqueous phase was moved into another tube, and 1/10 volume of 3 M sodium acetate (pH 5.2) plus 2 volumes of ice-cold ethanol absolute was added for DNA precipitation. Centrifuged at 5000 rpm for 30 min at 4°C, the DNA pellet was washed with 10 ml of 70% ethanol and centrifuged for 10 min. Ethanol was decanted and the pellet was air-dried for 30 min. Finally, the pellet was dissolved with 4 ml of 1× TE by overnight incubation at 4°C. Genomic DNA was stored at 4°C.

Concentration of genomic DNA cannot be accurately determined by spectrophotometer, due to the high molecular weight. Instead, it was estimated by agarose gel electrophoresis together with the λ/*Hind*III marker suitable for DNA quantitation. Samples were mixed with 6× BPB-buffer, loaded into a 0.8% agarose gel containing 0.5 µg/ml ethidium bromide, and separated with 1× TAE buffer.

4.17 Polymerase chain reaction

Polymerase chain reaction (PCR) is widely used for rapid amplification of specific target regions from trace amounts of DNA samples. All PCR reactions in this study were performed in a 50-µl format using the Thermocycler PTC-200 machinery with heated lid. Common materials for the reactions in one experiment were made as a homogenized master-mix to reduce pipetting error and to ensure comparable results among individual reactions. The final PCR products were examined by loading samples with 6× XC-buffer into a 1.5% agarose gel

for electrophoresis. The following commercial PCR systems were used for concrete experimental requirements, according to the manufacturer's instructions.

***Taq* DNA polymerase (Invitrogen, Karlsruhe)**

For routine PCR, up to 3 kb amplicon, non proof-reading, non hot-start

Reaction components	**Final conc.**
10× PCR buffer without Mg	1×
50 mM $MgCl_2$	1.5 mM
dNTP mix (10 mM each)	0.2 mM each
forward primer (10 μM)	0.2 μM
reverse primer (10 μM)	0.2 μM
Genomic DNA template	100 ng/reaction
Taq DNA polymerase (5 U/μl)	2 U/reaction

Cycling condition

Initial denaturation:	94 °C	for 3 min	
Denaturation:	94 °C	for 30 sec	} 35 cycles
Annealing:	60 °C	for 30 sec	
Extension:	72 °C	for 1 min/kb	
Final extension:	72 °C	for 10 min	
Cooling:	4 °C	hold	

Paq5000 DNA polymerase: (Stratagene, La Jolla, U.S.A)

Faster and cheaper than *Taq*, up to 6 kb, non proof-reading, non hot-start

Reaction components	**Final conc.**
10× Paq5000 Reaction Buffer	1×
dNTP-mix (10 mM each)	0.2 mM
forward-primer (10 μM)	0.2 μM
reverse-primer (10 μM)	0.2 μM
Genomic DNA template	100 ng/reaction
Paq5000 DNA polymerase (5 U/μl)	2.5 U/reaction

Cycling condition

Initial denaturation:	95 °C	for 3 min	
Denaturation:	95 °C	for 30 sec	} 35 cycles
Annealing:	60 °C	for 30 sec	
Extension:	72 °C	for 30 sec/kb	
Final extension:	72 °C	for 5 min	
Cooling:	4 °C	hold	

FastStart *Taq* DNA Polymerase (Roche Applied Science, Mannheim)

Hot-start, high specificity and sensitivity, up to 3 kb, non proof-reading

Reaction components	**Final conc.**
10× PCR buffer with 20 mM $MgCl_2$	1×
dNTP-mix (10 mM each)	0.2 mM each
forward-primer (10 μM)	0.2 μM
reverse-primer (10 μM)	0.2 μM
Genomic DNA template	100 ng/reaction
FastStart Taq DNA polymerase (5 U/μl)	2 U/reaction

Cycling condition

Initial denaturation:	95 °C	for	5 min	
Denaturation:	95 °C	for	30 sec	} 35 cycles
Annealing:	60 °C	for	30 sec	
Extension:	72 °C	for	1 min/kb	
Final extension:	72 °C	for	7 min	
Cooling:	4 °C		hold	

Expand Long Range PCR System (Roche Applied Science, Mannheim)

up to 25 kb amplicon, proof-reading, non hot-start

Reaction components	**Final conc.**
5× Buffer with 12.5 mM $MgCl_2$	1×
dNTP-mix (10 mM each)	0.5 mM each
forward-primer (10 μM)	0.3 μM
reverse-primer (10 μM)	0.3 μM
Genomic DNA template	100 ng/reaction
Expand Long Range Enzyme mix (5 U/μl)	3.5 U/reaction

Cycling condition

Initial denaturation:	94 °C	for	3 min	
Denaturation:	94 °C	for	30 sec	} 35 cycles
Annealing:	60 °C	for	30 sec	
Extension:	68 °C	for	1 min/kb	
Final extension:	68 °C	for	7 min	
Cooling:	8 °C		hold	

4.18 Restriction-site PCR

Restriction-site PCR (RS-PCR) is applied to determine the integration site of HPV16 DNA from the human genome at the sequence level (Sarkar *et al*., 1993; Thorland *et al*., 2000). The essence of this method is the use of restriction-site oligonucleotide (RSO) primer, which contains a T7-adapter at the 5' end, and the recognition site of an HPV16 non-cutting restriction enzyme at the 3' end (see *Section 2.1* and *4.12.3*). A random sequence of ten nucleotides was placed in the middle for efficient primer binding. The RS-PCR assay consists of two sequential rounds of PCR. In the first round, an HPV16-specific primer targeting the HPV16 early or late region was applied together with eight RSO-primers respectively, at a low stringent annealing temperature. The second round is a stringent semi-nested PCR, for which a nested HPV16 primer was used together with T7-adapter as primer pair. Both rounds of PCR were performed using the Expand High Fidelity PCR system (Roche, Mannheim), according to the user manual.

1st PCR

Reaction components	**Final conc.**
5× Buffer with 7.5 mM $MgCl_2$	1×
dNTP-mix (10 mM each)	0.2 mM each
HPV16-primer (5 μM)	0.1 μM
RSO-primer (50 μM)	2 μM
Genomic DNA template	100 ng/reaction
Expand HiFi Enzyme Blend (5 U/μl)	2.5 U/reaction

2nd PCR

Reaction components	**Final conc.**
5× Buffer with 7.5 mM $MgCl_2$	1×
dNTP-mix (10 mM each)	0.2 mM each
nested HPV16-primer (5 μM)	0.1 μM
T7-adapter (10 μM)	0.1 μM
1st PCR product	5 µl/reaction
Expand HiFi Enzyme Blend (5 U/μl)	2.5 U/reaction

Cycling condition

Initial denaturation:	94 °C	for	3 min	
Denaturation:	94 °C	for	30 sec	35 cycles
Annealing:	45 °C	for	30 sec (1st PCR)	
	60 °C	for	30 sec (2nd PCR)	
Extension:	68 °C	for	4 min	
Final extension:	68 °C	for	7 min	
Cooling:	8 °C		hold	

4.19 PCR clean-up

PCR clean-up was performed using the QIAquick PCR purification kit (Qiagen, Hilden) according to the user manual. PCR products were mixed with five volumes of buffer PB and transferred into the QIAquick spin column. To bind DNA, the tube was centrifuged at 13000 rpm for 1 min at RT. The flow-through was discarded and the column was placed back into the collection tube. DNA was washed once with 750 µl of buffer PE and centrifuged again. After the removal of flow-through, the column was placed back and centrifuged for another 1 min to discard the residual ethanol. The column was transferred into a new 1.5 ml tube, and 30 µl of buffer EB (10 mM Tris-HCl, pH 8.5) was carefully pipetted onto the center of the silica membrane. For a more efficient recovery, the elution was incubated for 1 min, followed by centrifugation for an additional minute. The eluted PCR products were stored at –20°C.

4.20 Gel extraction of DNA fragments

DNA fragments were extracted from agarose gel with the QIAquick gel extraction kit (Qiagen, Hilden) according to the user manual. Under UV light, the DNA band of interest was excised from the agarose gel with a scalpel. The gel slice up to 400 mg was exactly weighed and then covered with 3 volumes of buffer QG. The reaction was incubated at 50°C for 10 min, and mixed by vortexing vigorously every 2–3 min until the gel was completely dissolved. One gel volume of isopropanol was added and then mixed thoroughly. The sample with maximal volume of 800 µl was loaded into a QIAquick spin column and centrifuged at 13000 rpm for 1 min at RT to bind DNA onto the silica membrane. The flow-through was discarded, and the column was placed back to the same collection tube. The washing step and DNA elution procedure remains the same as described in *Section 4.19*. DNA eluate was stored at –20°C.

4.21 PCR cloning

Two commercial systems, TOPO TA cloning kit (Invitrogen, Karlsruhe) and StrataClone PCR cloning kit (Strategene, La Jolla, U.S.A), were used according to the user manuals. Both systems benefit from the DNA re-joining activity of DNA topoisomerase I, which is covalently bound to cloning vectors and assists in ligation between vector and insert DNA. Here, the StrataClone PCR cloning protocol is briefly described.

After pipetting the following components together: 3 µl of StrataClone cloning buffer, 2 µl of PCR product, and 1 µl of StrataClone vector mix, the reaction mixture was incubated at RT for 30 min. Following ligation, the reaction was placed on ice and up to 2 µl of the mixture was added to a vial of StrataClone SoloPack competent bacteria. The transformation reaction was mixed gently and incubated on ice for 30 min. The transformed bacteria were heat-shocked at 42°C for 45 sec. The reaction was immediately chilled on ice and incubated for 2 min. To recover the bacteria, 250 µl of SOC medium pre-warmed to 42°C was added and incubated

at 37°C for 1 h, with horizontal agitation at 180 rpm. During this time, two selective LB-agar plates were warmed to 37 °C and 40 µl of 2% (w/v) X-gal was spread evenly on each plate for blue-white screening. Finally, two different volumes of the recovered transformation mixture were plated onto the screening plates, and then incubated overnight at 37°C. Next day, white or light-blue but not dark-blue colonies were picked out for plasmid analysis.

4.22 Mini-preparation of bacterial plasmid DNA

The colony of choice was picked from the agar plate and inoculated into 4 ml LB-medium containing appropriate antibiotics. After shaking overnight at 180 rpm, 37°C, the bacterial culture was harvested. About 1.5 ml of culture was transferred into a 2 ml tube, centrifuged at 13000 rpm for 1 min to collect the bacterial pellet. The supernatant was discarded, only leaving a small volume at the bottom. The pellet together with the residual supernatant was resuspended by vortexing. To lyse bacteria and release plasmids, 300 µl of TENS buffer (1× TE, pH 8.0; 0.1 M NaOH; 0.5% SDS, w/v) was added and incubated for 5 min with horizontal agitations. To neutralize the alkali lysis buffer and favor the subsequent precipitation, 150 µl of 3 M sodium acetate (pH 5.2) was added, and mixed thoroughly by inverting the tube for several times. The reaction mixture was centrifuged at 13000 rpm for 2 min, and the supernatant was moved immediately into a 1.5 ml tube filled with 900 µl of ethanol absolute (–20°C) for DNA precipitation. Mixed well by inverting the tube, the precipitated plasmid DNA was spun down at 10000 rpm for 10 min at 4°C. The plasmid pellet was washed with 1 ml of 70% ethanol and centrifuging again for 5 min to remove the residual salt. The pellet was air-dried for 30 min to 1 h, and then resuspended with 50 µl of storage buffer (1× TE, pH 8.0; 20 µg/ml RNase A). The plasmid preparation was stored at 4°C.

4.23 Plasmid DNA sequencing

After mini-preparation and restriction analysis, plasmid DNA containing inserts with the expected sizes were sent to Andreas Hunziker at the DKFZ Genomics and Proteomics Core Facility for Sanger DNA sequencing.

4.24 Total RNA isolation from cultured mammalian cells

RNA isolation was performed using the TRIzol reagent (Invitrogen, Karlsruhe) according to the manufacturer's instruction. Cells were washed twice with ice-cold PBS. For the 75-cm^2 flask, 7.5 ml of TRIzol reagent containing phenol and guanidine isothiocyanate was applied. As a chaotropic agent, guanidine isothiocyanate lyses cells and protects RNA from RNase by denaturing all protein. The flask was agitated gently at RT for 5 min to completely dissociate the nucleoprotein complexes. The lysate was moved into a 15 ml Falcon tube, and 1.5 ml (1/5

of TRIzol volume) of chloroform was added for phase separation. The tube was closed firmly, shaken intensively by hand for 15 sec, and incubated at RT for 3 min. Centrifuged at 3000 rpm for 15 min at 4°C, the upper aqueous phase containing RNA was separated from the lower phenol-chloroform phase and the interphase both comprising protein and DNA. The aqueous phase was transferred into a new tube and mixed with 4 ml of ice-cold isopropanol for RNA precipitation. The reaction was incubated at RT for 10 min, and centrifuged at 4000 rpm for 10 min at 4°C. The RNA pellet was washed with 7.5 ml of 70% ethanol and centrifuged again for 5 min. Ethanol was decanted and the pellet was air-dried at RT for 30 min. Finally, RNA was resuspended in 100 µl of RNase-free water and incubated at 60°C for 10 min to dissolve completely. The RNA preparation was stored at –80°C.

The concentration was measured by pipetting 1.5 µl of RNA sample directly on NanoDrop spectrophotometer. For quality control, 2 µl of sample was mixed well with 18 µl of RNA-loading buffer, heated at 65 °C for 15 min to disrupt RNA secondary structure, and loaded onto a 1% agarose gel. The 1× MOPS buffer was used for RNA gel electrophoresis.

4.25 Poly-A$^+$ RNA enrichment

Poly-A$^+$ RNA was extracted from total RNA with Dynabeads Oligo $(dT)_{25}$ (Invitrogen, Karlsruhe) according to the user manual. Before loading samples, 200 µl of magnetic beads suspension was pipetted into a 1.5 ml RNase-free tube and placed into a magnet. Incubated at RT for 30 sec, the beads pellet was resuspended in 100 µl of 2× binding buffer (20 mM Tris-HCl, pH 7.5; 1 M LiCl; 2 mM EDTA). Incubated again in magnet for 30 sec, the supernatant was discarded, the tube was transferred into a normal rack and 100 µl of 2× binding buffer was added. Up to 75 µg of total-RNA was diluted with DEPC-treated water to 100 µl, and heated at 65°C for 3 min. The denatured RNA was transferred immediately into the beads solution. The reaction was incubated at RT for 5 min to allow the binding. The tube was placed back to the magnet and incubated for 30 sec. The pellet was washed twice with 200 µl of 1× wash buffer (10 mM Tris-HCl, pH 7.5; 150 mM LiCl; 1 mM EDTA). Following the washing, beads were dissolved with 10 µl of 1× elution buffer (2 mM EDTA), incubated at 65°C for 2 min and separated by magnet. The eluted mRNA was transferred into a new RNase-free tube and stored at –20°C.

4.26 Reverse transcription PCR

In reverse transcription PCR (RT-PCR), template RNA is first converted into cDNA by reverse transcription, followed by the amplification of cDNA using convenient PCR with gene-specific primer pairs. For the synthesis of full-length first-strand cDNA, Superscript III reverse transcriptase (Invitrogen, Karlsruhe) was used according to the user manual. The following components were pipetted into a 1.5 ml RNase-free tube, plus DEPC-treated water to 14 µl

total and mixed thoroughly: 5 μg of total RNA or 500 ng of poly-A^+ RNA, 1 μl of anchored oligo-$d(T)_{23}$ primer (50 μM) and 1 μl of dNTP-mix (10 mM each). The anchored oligo-dT primer (5'-TTTTTTTTTTTTTTTTTTTTTTTV-3'; V=A, C or G) binds to the poly-A tail of mRNA, whereas for those transcripts without this feature, random hexamer instead is usually applied for priming. The mixture was heated at 65°C for 5 min, chilled on ice and spun down briefly. To start the reverse transcription, 4 μl of 5× First-Strand buffer, 1 μl of 0.1 M DTT and 1 μl of Superscript III reverse transcriptase (200 U/μl) was added, mixed well and incubated at 50°C for one hour. Finally, the reaction was heated at 70°C for 15 min to inactivate the reverse transcriptase. The cDNA/RNA hybrids can be directly used as template for the subsequent PCR. To improve PCR sensitivity, it is possible to remove the RNA strand either with RNase H or alkali treatment as shown below. Following the reverse transcription, 2 μl of 2.5 M NaOH was added, mixed by vortexing and spun down briefly. The reaction was incubated at 37 °C for 15 min to digest RNA and ceased by neutralization with 10 μl of 2 M HEPES buffer, mixed well and spun down. The first-strand cDNA solution was stored at – 20°C. For a typical PCR, 3 μl of cDNA product was applied as template without any prior purification.

4.27 Rapid amplification of cDNA ends

The rapid amplification of cDNA ends (RACE) assay is applied to obtain the full-length cDNA sequences of specific mRNA molecules, whose partial sequences are previously known. The SMART RACE cDNA amplification kit (Becton Dickinson, Heidelberg) was used according to the user manual. The first step is to synthesize RACE-ready first-strand cDNA, which is split into two separate reactions for 5'-RACE and 3'-RACE. A 5 μl of mixture included with 1 μg of total-RNA, 1 μl of 5'-CDS primer and 1 μl of SMART II-A oligo was combined in a 0.5 ml PCR tube for the preparation of 5'-RACE-ready cDNA. Likewise, another 5-μl mixture containing 1 μg of total-RNA and 1 μl of 3'-CDS primer-A was prepared for 3'-RACE-ready cDNA. The tubes were incubated at 70°C for 2 min, and chilled immediately on ice. Centrifuged briefly, the following materials were added into both tubes: 2 μl of 5× First-Strand buffer, 1 μl of DTT (20 mM), 1 μl of dNTP-mix (10 mM each) and 1 μl of MMLV Reverse Transcriptase. The reactions (10 μl each) were incubated at 42°C for 1.5 hr in a thermal cycler. Finally, the cDNA was diluted with 100 μl of 1× TE buffer, followed by incubation at 72°C for 7 min to inactivate the residual reverse transcriptase. After reverse transcription, cDNA was subjected to RACE assay by using the Advantage 2 PCR Kit (Becton Dickinson, Heidelberg). For a 50-μl PCR reaction, the following components were mixed into a PCR tube for the 5'- or 3'-RACE analysis of a specific gene transcript: 34.5 μl of DNase-free water, 5 μl of 10× Advantage 2 PCR buffer, 1 μl of dNTP-mix (10 mM each), 2.5 μl of 5'- or 3'-RACE-ready cDNA, 5 μl of 10× UPM, 1 μl of GSP-forward (for 3'-RACE; 10 μM) or GSP-reverse (for 5'-RACE; 10 μM), and 1 μl of 50× Advantage 2 polymerase mix. The tube was placed into a thermal cycler, and the PCR condition is described as below:

Initial denaturation:	94 °C	for	3 min	
Denaturation:	94 °C	for	30 sec	} 35 cycles
Annealing:	68 °C	for	30 sec	
Extension:	72 °C	for	3 min	
Final extension:	72 °C	for	10 min	
Cooling:	4 °C		hold	

Subsequently, the RACE PCR products were cloned into pSC-A vector and sequenced with the Sanger method. The overlapping sequences resulting from the 5'- and 3'-RACE were assembled together to reach the full-length cDNA. Finally, the assumed full-length cDNA was verified by using both GSP-forward and reverse in one PCR reaction.

4.28 Nucleic acid transfer from agarose gel onto membrane

For the detection of specific nucleic acid molecules, DNA or RNA samples separated by gel electrophoresis are immobilized onto a solid support such as nylon membrane, quantitatively and also maintaining the same position as in the gel, through the vertically downward capillary transfer.

4.28.1 DNA transfer by alkaline capillary blotting (Southern blot)

Immediately after gel photographing, bands of DNA ladder were marked with a scalpel. Briefly rinsed in deionized water, the gel was immersed in 0.25 M HCl, and agitated for 10 min to partially depurinate DNA. BPB as pH indicator switches color from purple (base form) to yellow (acid form). The depurination step is essential for efficient transfer of longer fragments (>4 kb) like genomic DNA, and can be omitted for shorter DNA. After depurination, the gel was rinsed shortly in water, submerged in denaturation buffer (0.4 M NaOH), and shaken for 2× 15 min to denature DNA (BPB turns back to purple). Meanwhile, a piece of positively-charged Hybond-N+ nylon membrane larger than the gel was prepared, wet in water, and equilibrated together with two sheets of filter paper in denaturation buffer for 10 min. The apparatus for capillary transfer was assembled as follows (bottom-up):

- a stack of absorbent paper towels
- two sheets of filter paper (dry)
- two sheets of filter paper (saturated in denaturation buffer)
- Hybond-N+ nylon membrane
- agarose gel

The assembly was surrounded with plastic wrap and a small weight was placed on the top. The transfer is completed in 3-5 h. During the blotting, DNA was driven from the gel onto the

porous nylon membrane by capillary action, bound to the positively-charged groups and fixated by the alkaline solvent finally.

After blotting, the position of DNA ladder was marked for later size determination. The membrane was agitated gently in the neutralization buffer (3 M NaCl; 0.3 M sodium citrate; 0.5 M Tris-HCl, pH = 7.0) for 10 min, and dried thoroughly on filter paper.

4.28.2 RNA transfer by alkaline capillary blotting (Northern blot)

After electrophoresis and photographing, the position of the 28S and 18S rRNA were marked. The gel was rinsed briefly in DEPC-treated water, and soaked in five gel volumes of transfer buffer (10 mM NaOH; 3 M NaCl) for 20 min to partially hydrolyze RNA. A piece of Hybond-N+ nylon membrane was wet in DEPC-treated water, and incubated in transfer buffer together with two sheets of filter paper for 10 min. The apparatus assembly for Northern blot resembles the one used in Southern blot (see *Section 4.28.1*). The time for RNA transfer is limited to about one hour, to avoid the over-hydrolyzation caused by alkaline buffer. After blotting, the membrane was marked, incubated in 6× SSC for 5 min, and dried on filter paper.

4.29 Radioactive labeling of DNA probes

The labeling reaction was performed with the Rediprime II DNA labeling kit (GE Healthcare, München) according to the user manual. The dsDNA template (minimal length of 200–300 bp) to be labeled was diluted to 25 ng in 45 µl with DNase-free water, and incubated at 95°C for 5 min. The denatured DNA was chilled on ice for 5 min, centrifuged briefly, and transferred into a vial of Rediprime II reaction mixture. Five microliter (50 µCi) of [α-^{32}P]dCTP was added and mixed thoroughly. The labeling reaction was done at 37°C for 30 min, and terminated by 20 µl of stop solution (2× TNE; 0.1% BPB, w/v; 0.5% blue dextran, w/v).

During the labeling, a MicroSpin G-25 column (GE healthcare, München) was prepared for the clean-up by gel filtration chromatography according to the user manual. First, the resin materials in the spin-column were suspended by vortexing vigorously. The cap was loosened ¼-turn, and the bottom closure was removed. The column was placed into a collection tube, and centrifuged at 3000 rpm for 1 min. After the flow-through was discarded, 200 µl of 1× TNE buffer (1× TE, pH 8.0; 10 mM EDTA, pH 8.0; 100 mM NaCl; 0.2% SDS) was added and centrifuged again for equilibrium. The labeling reaction was mixed with 75 µl of 1× TNE buffer, transferred to the center of the column, and centrifuged at 3000 rpm for exactly 1 min. Larger DNA fragments pass through the column faster, whereas small molecules like impurities retard and in the gel matrix. The DNA eluate was either directly used for hybridization after a denaturation with 0.2 M NaOH (final conc.) for 10 min, or stored at 4°C for a short period.

4.30 Nucleic acid hybridization

To detect the specific nucleic acid sequence, denatured DNA probes were allowed to hybridize with DNA or RNA sample blotted on membrane. The binding specificity was ensured by a high-stringent hybridization temperature at 68°C. The blot was spread in a hybridization bottle smoothly and wet in water briefly. To equilibrate the membrane and reduce the hybridization background, the bottle was filled with appropriate volumes (5 ml for small and 10 ml for large bottles) of pre-hybridization buffer, and rotated at 68 °C for at least one hour. The buffer was replaced by pre-warmed hybridization buffer, and the amount of DNA probe required was calculated according to the volume of hybridization buffer. Generally, the radioactivity corresponding to 14 ng of DNA input was supplied for 10 ml of hybridization buffer. Hybridization was performed overnight at 68°C. Next day, the hybridization mixture was collected for reuse. The membrane was agitated gently in wash buffer (2× SSC; 0.1% SDS, w/v) at 68°C for 3× 30 min to remove the unspecific binding. Finally, the membrane was dried very briefly and packed with plastic wrap.

Because the DNA probe is present in an excessive amount compared to the complementary sequences fixed on the membrane, the radioactive probe can be re-used within the half-life of the ^{32}P-isotope (14 days). For this purpose, the hybridization mixture was denatured by heating in boiling water for 10 min, and immediately applied for the new hybridization.

Pre- and Hybridization buffer:
50 mM sodium phosphate buffer, pH 6.4
1% (w/v) SDS
1× Denhardt's solution
5× SSC
0.1 mg/ml tRNA

50× Denhardt's solution:
1% (w/v) Ficoll 400
1% (w/v) polyvinylpyrrolidone
1% (w/v) BSA

4.31 Hybridization signal capture by autoradiography

Due to the high-energy β-radiation emitted by ^{32}P-isotope, the hybridization signal can be captured by exposing an autoradiography film to the blot. For detecting low-abundant molecules, an intensifying screen was used to magnify the signal response linearly.

The blot was placed into the cassette in conjunction with intensifying screen, and a piece of MR-film (GE Healthcare, München) was inserted in between. The apparatus was closed firmly and kept in –80°C freezer. The duration of exposure depends on the abundance of DNA to be detected. In general, it takes hours for a robust band of PCR products or plasmid DNA, but days even for genomic Southern blots even with intensifying screen. Finally, the film was developed by the automatic Curix 60 film processor.

The Southern-blot can be used for re-hybridization, but the probe obtained from the last hybridization has to be stripped out. For this reason, a boiling solution of 0.1% (w/v) SDS was

poured onto the blot, agitated till the temperature down to RT. The blot was then rinsed briefly with 2× SSC and dried on filter paper.

4.32 Protein extraction from cultured mammalian cells

Cells were washed twice with ice-cold PBS. For a 10-cm dish, 1 ml of PBS was added and spread throughout the plate. Using a plastic scraper, cells were removed from the bottom and then collected into a 2 ml tube. Centrifuged at 2500 rpm for 5 min at 4°C, the cell pellet was resuspended in 1 ml of PBS and transferred into a new tube. Centrifuged again, the pellet was dissolved with 100 µl of lysis buffer. The lysis was accomplished by incubation on ice for 30 min. Finally, the cell lysate was centrifuged at 12000 rpm for 5 min at 4°C. The resulting supernatant was transferred into a 1.5 ml tube, mixed thoroughly and stored at –70°C.

RIPA buffer
50 mM Tris-HCl, pH 7.4
150 mM NaCl
1 mM EDTA
1% (v/v) NP40
0.5% (w/v) Na-deoxycholate
0.1% (w/v) SDS

Lysis buffer
RIPA buffer
add the following just before use
1 mM PMSF
Complete protease inhibitor, EDTA-free
(dissolve one tablet in 2 ml water, and
supply 40 µl per ml RIPA)

4.33 Protein concentration determination – Bradford assay

The Bradford assay is able to measure the protein concentration of complex sample solutions. It is based on the observation that Coomassie Brilliant Blue G-250 changes its absorbance maximum from 470 nm to 595 nm by binding to protein. The intensity of the absorbance shift is quantitatively related to the protein concentration. As relative standard, a dilution series with 2, 4, 6, 8, 10, 12, 14 µg of BSA was prepared with water to a volume of 800 µl each and mixed well. For samples to be tested, appropriate volumes (1–3 µl) of cell lysate were applied and diluted with water to 800 µl. After that, 200 µl of Coomassie dye was pipetted into disposable plastic cuvettes. The protein dilutions and water reference were transferred into the cuvettes and mixed thoroughly. The reaction was allowed to stay at RT for 10 min, and measured at 595 nm with the Ultrospec 3000 Pro spectrophotometer. The absorbance *versus* BSA amount was calculated to generate a linear regression standard curve and thereby the protein concentration of samples was determined.

4.34 Protein detection by Western blot

4.34.1 Denaturing SDS-polyacrylamide gel electrophoresis

SDS-polyacrylamide gel electrophoresis (SDS-PAGE) is a standard method to separate polypeptides based on the molecular weight. As an anionic detergent, SDS denatures polypeptide chains and provides them with a net negative charging proportional to their length. The separation system is a one-dimensional, vertical and discontinuous gel with stacking and resolving sections. For a small number of samples, two mini-gels (1-mm thick; 8×10-cm size) were prepared with the Bio-Rad Mini-PROTEAN II according to the manufacturer's instruction. First, the gel-casting apparatus was assembled, and the following reagents for the resolving gel of desired polyacrylamide concentration were added in a 50 ml Falcon tube:

	10%	**12.5%**
H_2O	4.05 ml	3.15 ml
Resolving buffer	2.5 ml	2.5 ml
(1.5 M Tris-HCl, pH 8.8)		
Rotiphorese Gel 30	3.3 ml	4.2 ml
10% (w/v) SDS	100 µl	100µl
TEMED	5 µl	5 µl
10% (w/v) APS, prepared daily	50 µl	50 µl

Immediately after addition of TEMED and ammonium persulfate (APS), contents were mixed well and poured quickly. The solution was overlaid with isopropanol and incubated at RT for 30 min to allow the polymerization of acrylamide monomers. The isopropanol was decanted and another mixture for the 5% stacking gel was prepared containing 3.45 ml of H_2O, 620 µl of stacking buffer (0.5 M Tris-HCl, pH 6.8), 830 µl of Rotiphorese Gel 30, 50 µl of 10% SDS, 6 µl of TEMED and 60 µl of 10% APS. The stacking gel solution was poured on top of the resolving gel and incubated for 20 min. The complete gel was placed in the electrophoresis chamber and filled with 1× running buffer. Before loading, cell lysate containing 30 µg of protein was combined with 4 µl of 5× sample buffer plus RIPA-buffer to 20 µl. By heating at 99°C for 10 min, disulfide bonds were broken by β-mercaptoethanol and protein was denatured by SDS. Samples were cooled on ice, spun down briefly and loaded onto the gel. Five microliters of pre-stained protein marker was applied directly without boiling. Proteins were separated by electrophoresis at 200 V for 45 min at RT.

4.34.2 Western blot

Protein can be transferred from the gel onto a supporting surface such as PVDF membrane. After SDS-PAGE, the stacking gel was cut off and the resolving gel was immersed in blotting buffer for 20–30 min with gentle agitation. A piece of Immobilon-P PVDF membrane was activated by methanol for 2–3 min, immediately rinsed with water and then equilibrated in blotting buffer. The gel was transferred by the wet-blotting procedure with the Mini Trans-Blot set. The blotting apparatus was assembled as below:

- light side of gel-holder cassette (near anode)
- a piece of fiber pad (pre-soaked in blotting buffer)
- a sheet of filter paper (pre-soaked)
- PVDF membrane
- protein gel
- a sheet of filter paper (pre-soaked)
- a piece of fiber pad
- dark side of gel-holder cassette (near cathode)

The cassette was closed firmly and placed in the buffer tank and then filled with ice-cold blotting buffer. A stir bar was put into the chamber to keep the buffer substances equally distributed during the whole procedure. The transfer was achieved by running at 400 mA, 100 Volt, for one hour at 4°C.

4.34.3 Immunodetection

The blot was immersed in ice-cold TBST (1× TBS; 0.1% Tween 20, v/v) with 5% (w/v) milk, and agitated at RT for one hour to block unspecific targets. The primary antibody from mouse or rabbit recognizing specific human protein was diluted in TBST with 5% milk to the optimal working concentration. After blocking, the blot was rinsed briefly in TBST and incubated with the antibody dilution either at RT for 1–2 h or overnight at 4°C. The blot was washed with TBST for 5 × 5 min at RT. The secondary antibody conjugated with horseradish peroxidase (HRP) and directed against the primary one was diluted and incubated with the blot for one hour at RT, followed by 5 × 5 min washing. Finally, the specific binding between protein and antibody was quantitatively converted to light signal mediated by the enhanced chemiluminescence (ECL) detection system. The same volume (1.5 µl) of solution A (luminol and enhancer) and solution B (hydrogen peroxide) was mixed and allowed to equilibrate at RT for 5 min. The mixture was applied to cover the blot completely and incubated for 2 min. The excess fluid was drained off and the blot was packed with transparent wrap. A piece of Biomax Light film (GE Healthcare, München) was exposed to the blot for detecting the protein bands.

After the data interpretation, the primary and secondary antibodies were stripped out of the blot by shaking at 50°C for 30 min in the stripping buffer. The membrane was washed with TBST at RT for 2× 10 min, and blocked again in TBST with 5% milk. Subsequently, primary antibody against human GAPDH protein as internal control was used for the normalization of protein amount among different samples.

4.35 ASP16 strategy for HPV16 DNA integration analysis

4.35.1 GenomePlex whole genome amplification

Whole genome amplification (WGA) was performed using the GenomePlex WGA 2 kit (Sigma-Aldrich, München) according to the manufacturer's instruction. To reach a representative and unbiased WGA, at least 10 ng of genomic DNA is required as starting material. For clinical samples the DNA template can be increased to 25–100 ng, dependent on the degree of DNA degradation. The template was diluted with DNase-free water to a volume of exactly 10 µl, and mixed well with 1 µl of 10× fragmentation buffer in a 0.5 ml PCR tube. To fragment genomic DNA efficiently but not excessively, the reaction was incubated at 95°C for exactly 4 min. The tube was immediately cooled on ice for minutes and spun down briefly. Two microliter of 1× library preparation buffer plus 1 µl of library stabilization solution was added and incubated at 95°C for 2 min. The reaction was chilled on ice and centrifuged briefly. After adding 1 µl of library preparation enzyme, the reaction contents totally 15 µl were mixed and placed in a thermal cycler. The OmniPlex library generation was accomplished by stepped isothermal reactions as follows:

Step 1:	16 °C	for	20 min
Step 2:	24 °C	for	20 min
Step 3:	37 °C	for	20 min
Step 4:	75 °C	for	5 min
Step 5:	4 °C		hold

To amplify the OmniPlex library, reagents including 7.5 µl of 10× amplification master mix, 47.5 µl of water and 5 µl of WGA DNA polymerase were added and mixed thoroughly. To reduce the experimental bias, the amplification reaction was performed with a limited cycle number in a thermal cycler according to the following condition:

Initial denaturation:	95 °C	for	3 min	
Denaturation:	94 °C	for	15 sec	} 14 cycles
Annealing/Extension:	65 °C	for	5 min	
Cooling:	4 °C		hold	

After cycling, 3–5 μl from the 75 μl final products were loaded into a 1.5% agarose gel to check the WGA performance. Normally, a successful amplification yields 3–7 μg of DNA. The amplified DNA fragments were stored at –20°C and were stable for months. The WGA products were directly used as DNA template for PCR without any prior clean-up.

4.35.2 HPV16 DNA enrichment from GenomePlex libraries

4.35.2.1 Linear amplification by primer extension

Sequences and combinations of HPV16 primers are shown in *Section 4.12.5*, *2.11* and *2.12*. All biotin-labeled HPV16 primers used for linear amplification were individually diluted with 1× TE buffer to a concentration of 10 μM. Dilutions of the eight primers used in one reaction were blended into one mixture, with a final concentration of 0.1 μM for each primer. For linear amplification of HPV16 DNA, the Paq5000 DNA polymerase (Stratagene, La Jolla, U.S.A) was applied according to the user manual. The following components were mixed in a 0.5 ml PCR tube: 39 μl of DNase-free water, 5 μl of 10× Paq5000 reaction buffer, 1 μl of dNTP-mix (10 mM each), 2 μl of mixed biotin-primer (0.1 μM each), 2.5 μl of GenomePlex WGA products (HPV16-positive) and 0.5 μl of Paq5000 DNA polymerase (5 units/μl). The tube was placed into a thermal cycler, and the targeted regions of HPV16 DNA were linearly amplified through primer extension under the cycling condition as shown below. The extent of amplification was about 45 fold, generating biotinylated single-stranded DNA molecules.

Initial denaturation:	95 °C	for	3 min	
Denaturation:	95 °C	for	30 sec	} 45 cycles
Annealing:	58 °C	for	75 sec	
Extension:	72 °C	for	30 sec (1 kb)	
Cooling:	4 °C		hold	

4.35.2.2 Biotin-streptavidin selection

Single-stranded DNA carrying a biotin at the 5' end can be extracted from the background of whole genomic DNA via the specific and strong association between biotin and streptavidin. The BioMag streptavidin beads (Polysciences, Heidelberg) were used for HPV16 DNA selection according to the user manual. A 75-μl suspension of BioMag streptavidin beads (1 mg/ml) was pipetted into a 1.5 ml tube. The tube was placed in a magnet and incubated for 1 min, to allow the separation of magnetic particles from the storage buffer. The supernatant was discarded, and the tube was removed from the magnet. The beads were washed with 100 μl of 2× binding buffer and incubated in the magnet for 1 min. After the removal of supernatant, particles were resuspended in 50 μl of 2× binding buffer and combined with 50 μl of linear-amplified ssDNA products. The mixture was incubated on a normal rack for 15–30 min at room temperature, and the contents were mixed carefully by tapping the tube

intermittently to keep beads in suspension. The tube was returned to the magnet and incubated for 1 min. The supernatant was removed, and particles were washed twice with 150 µl of wash buffer and three times with 150 µl of 1× TE buffer. In each wash step, the tube was removed from the magnetic field, mixed with buffer solutions, and incubated on the separator for 1 min. Finally, the reaction was dissolved in 30 µl of 1× TE buffer and stored at 4°C for a short period.

2× binding buffer
40 mM Tris-HCl, pH 7.8
2 M NaCl,
2 mM EDTA,
0.1% (v/v) Tween 20

Wash buffer:
20 mM Tris-HCl, pH 7.8
1 M NaCl
1 mM EDTA,
0.05% (v/v) Tween 20

4.35.2.3 Multiplex PCR

The selected HPV16-positive single-stranded DNA was amplified and converted to double-stranded DNA by multiplex PCR. In addition, the Roche-A (RA) and Roche-B (RB) sequencing primers needed for the later pyrosequencing were incorporated. Primer sequences and combinations are shown in *Section 4.12.5*, *2.11* and *2.12*. The RA-HPV16 nested primers were individually diluted with 1× TE buffer to a 10-µM stock solution. A 10× mixture was prepared by combining the eight RA-HPV16 primers used in one PCR reaction at a final concentration of 1 µM each. Multiplex PCR was performed using the Qiagen multiplex PCR kit (Qiagen, Hilden) according to the user manual. The following reagents were mixed in a 0.5 ml PCR tube and placed into a thermal cycler, including 25 µl of 2× Qiagen multiplex PCR master mix, 5 µl of 10× RA-HPV16 primer mixture (1 µM each), 2 µl of RB-GPUA primer (10 µM), 13 µl of DNase-free water and 5 µl of DNA template from the selection step. The reaction was accomplished with the condition as below:

Initial denaturation:	95 °C	for	15 min	
Denaturation:	94 °C	for	30 sec	} 35 cycles
Annealing:	58 °C	for	75 sec	
Extension:	72 °C	for	1 min (1 kb)	
Final extension:	72 °C	for	10 min	
Cooling:	4 °C	hold		

After PCR, the tube was placed in the magnet and incubated for 1 min. Finally, the supernatant containing multiplex PCR product was collected and purified with QIAquick PCR purification kit as described in *Section 4.19*. The concentration was determined by NanoDrop spectrophotometer. One hundred nanogram of purified DNA from each reaction was examined on a 1.5% agarose gel, and further analyzed by Southern hybridization with the HPV16-complete DNA probe.

4.35.3 Massively parallel pyrosequencing

The HPV16-positive PCR products from multiple DNA samples were mixed together in equal amounts (100 ng of each purified multiplex-PCR reaction). The DNA mixture was checked by NanoDrop spectrophotometer for quantitation and quality control. DNA was sent to the DKFZ Genomics and Proteomics Core Facility for massively parallel pyrosequencing with the Genome Sequencer FLX system (Roche, Mannheim), which gives rise to hundreds of thousand sequence reads (up to 250 nt) in one sequencing run. Each individual sequence read represents only one single-molecule DNA template. Because the unique 4-nt barcode designated to each sample had been incorporated into the RB-GPUA fusion primer, DNA mixture was sequenced using the Roche-B sequencing primer. By doing this, all sequence reads carry the barcode important for later sequence sorting, even if some of them are not fully sequenced.

4.35.4 Sequence analysis with bioinformatics tools

The pyrosequence reads were analyzed with a computer program written in Perl language by my colleague Sasithorn Chotewutmontri as part of her PhD thesis. Details about the procedure of sequence analysis are explained in *Section 2.11*.

References

Alarmo, E.-L., Korhonen, T., Kuukasjärvi, T., Huhtala, H., Holli, K. & Kallioniemi, A. *Bone morphogenetic protein 7 expression associates with bone metastasis in breast carcinomas.* Ann Oncol, 2008, Vol. 19(2), pp. 308-314

i Altaba, A.R., Sánchez, P. & Dahmane, N. *Gli and hedgehog in cancer: tumours, embryos and stem cells.* Nat Rev Cancer, 2002, Vol. 2(5), pp. 361-372

Arias-Pulido, H., Peyton, C.L., Joste, N.E., Vargas, H. & Wheeler, C.M. *Human papillomavirus type 16 integration in cervical carcinoma in situ and in invasive cervical cancer.* J Clin Microbiol, 2006, Vol. 44(5), pp. 1755-1762

Autzen, P., Robson, C.N., Bjartell, A., Malcolm, A.J., Johnson, M.I., Neal, D.E. & Hamdy, F.C. *Bone morphogenetic protein 6 in skeletal metastases from prostate cancer and other common human malignancies.* Br J Cancer, 1998, Vol. 78(9), pp. 1219-1223

Avsian-Kretchmer, O. & Hsueh, A.J.W. *Comparative genomic analysis of the eight-membered ring cystine knot-containing bone morphogenetic protein antagonists.* Mol Endocrinol, 2004, Vol. 18(1), pp. 1-12

Baker, C.C., Phelps, W.C., Lindgren, V., Braun, M.J., Gonda, M.A. & Howley, P.M. *Structural and transcriptional analysis of human papillomavirus type 16 sequences in cervical carcinoma cell lines.* J Virol, 1987, Vol. 61(4), pp. 962-971

Baseman, J.G. & Koutsky, L.A. *The epidemiology of human papillomavirus infections.* J Clin Virol, 2005, Vol. 32 Suppl 1, pp. S16-S24

Bedell, M.A., Hudson, J.B., Golub, T.R., Turyk, M.E., Hosken, M., Wilbanks, G.D. & Laimins, L.A. *Amplification of human papillomavirus genomes in vitro is dependent on epithelial differentiation.* J Virol, 1991, Vol. 65(5), pp. 2254-2260

Bernard, B.A., Bailly, C., Lenoir, M.C., Darmon, M., Thierry, F. & Yaniv, M. *The human papillomavirus type 18 (HPV18) E2 gene product is a repressor of the HPV18 regulatory region in human keratinocytes.* J Virol, 1989, Vol. 63(10), pp. 4317-4324

Blish, K.R., Wang, W., Willingham, M.C., Du, W., Birse, C.E., Krishnan, S.R., Brown, J.C., Hawkins, G.A., Garvin, A.J., D'Agostino, R.B., Torti, F.M. & Torti, S.V. *A human bone morphogenetic protein antagonist is down-regulated in renal cancer.* Mol Biol Cell, 2008, Vol. 19(2), pp. 457-464

Bory, J.-P., Cucherousset, J., Lorenzato, M., Gabriel, R., Quereux, C., Birembaut, P. & Clavel, C. *Recurrent human papillomavirus infection detected with the hybrid capture II assay selects women with normal cervical smears at risk for developing high grade cervical lesions: a longitudinal study of 3,091 women.* Int J Cancer, 2002, Vol. 102(5), pp. 519-525

Bosch, F.X., Lorincz, A., Muñoz, N., Meijer, C.J.L.M. & Shah, K.V. *The causal relation between human papillomavirus and cervical cancer.* J Clin Pathol, 2002, Vol. 55(4), pp. 244-265

Bosch, F.X., Manos, M.M., Muñoz, N., Sherman, M., Jansen, A.M., Peto, J., Schiffman, M.H., Moreno, V., Kurman, R. & Shah, K.V. *Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. International biological study on cervical cancer (IBSCC) Study Group.* J Natl Cancer Inst, 1995, Vol. 87(11), pp. 796-802

Boshart, M., Gissmann, L., Ikenberg, H., Kleinheinz, A., Scheurlen, W. & zur Hausen, H. *A new type of papillomavirus DNA, its presence in genital cancer biopsies and in cell lines derived from cervical cancer.* EMBO J, 1984, Vol. 3(5), pp. 1151-1157

Bourhis, J., Le, M.G., Barrois, M., Gerbaulet, A., Jeannel, D., Duvillard, P., Doussal, V.L., Chassagne, D. & Riou, G. *Prognostic value of c-myc proto-oncogene overexpression in early invasive carcinoma of the cervix.* J Clin Oncol, 1990, Vol. 8(11), pp. 1789-1796

Boyer, S.N., Wazer, D.E. & Band, V. *E7 protein of human papilloma virus-16 induces degradation of retinoblastoma protein through the ubiquitin-proteasome pathway.* Cancer Res, 1996, Vol. 56(20), pp. 4620-4624

Briolat, J., Dalstein, V., Saunier, M., Joseph, K., Caudroy, S., Prétet, J.-L., Birembaut, P. & Clavel, C. *HPV prevalence, viral load and physical state of HPV-16 in cervical smears of patients with different grades of CIN.* Int J Cancer, 2007, Vol. 121(10), pp. 2198-2204

Brown, D.R., Shew, M.L., Qadadri, B., Neptune, N., Vargas, M., Tu, W., Juliar, B.E., Breen, T.E. & Fortenberry, J.D. *A longitudinal study of genital human papillomavirus infection in a cohort of closely followed adolescent women.* J Infect Dis, 2005, Vol. 191(2), pp. 182-192

Calin, G.A., Sevignani, C., Dumitru, C.D., Hyslop, T., Noch, E., Yendamuri, S., Shimizu, M., Rattan, S., Bullrich, F., Negrini, M. & Croce, C.M. *Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers.* Proc Natl Acad Sci USA, 2004, Vol. 101(9), pp. 2999-3004

Casas, L., Galvan, S.C., Ordoñez, R.M., Lopez, N., Guido, M. & Berumen, J. *Asian-american variants of human papillomavirus type 16 have extensive mutations in the E2 gene and are highly amplified in cervical carcinomas.* Int J Cancer, 1999, Vol. 83(4), pp. 449-455

Castellsagué, X., Díaz, M., de Sanjosé, S., Muñoz, N., Herrero, R., Franceschi, S., Peeling, R.W., Ashley, R., Smith, J.S., Snijders, P.J.F., Meijer, C.J.L.M. & Bosch, F.X. *Worldwide human papillomavirus etiology of cervical adenocarcinoma and its cofactors: implications for screening and prevention.* J Natl Cancer Inst, 2006, Vol. 98(5), pp. 303-315

Castellsagué, X. & Muñoz, N. *Chapter 3: Cofactors in human papillomavirus carcinogenesis--role of parity, oral contraceptives, and tobacco smoking.* J Natl Cancer Inst Monogr, 2003(31), pp. 20-28

Castellsagué, X., de Sanjosé, S., Aguado, T., Louie, K.S., Bruni, L., Muñoz, J., Diaz, M., Irwin, K., Gacic, M., Beauvais, O., Albero, G., Ferrer, E., Byrne, S. & Bosch, F.X. *HPV and Cervical Cancer in the World. 2007 Report.* WHO/ICO Information Centre on HPV and Cervical Cancer (HPV Information Centre). Available at: www.who.int/hpvcentre

Chellappan, S., Kraus, V.B., Kroger, B., Munger, K., Howley, P.M., Phelps, W.C. & Nevins, J.R. *Adenovirus E1A, simian virus 40 tumor antigen, and human papillomavirus E7 protein share the capacity to disrupt the interaction between transcription factor E2F and the retinoblastoma gene product.* Proc Natl Acad Sci USA, 1992, Vol. 89(10), pp. 4549-4553

Cheng, S., Schmidt-Grimminger, D.C., Murant, T., Broker, T.R. & Chow, L.T. *Differentiation-dependent up-regulation of the human papillomavirus E7 gene reactivates cellular DNA replication in suprabasal differentiated keratinocytes.* Genes Dev, 1995, Vol. 9(19), pp. 2335-2349

Chesi, M., Bergsagel, P.L., Shonukan, O.O., Martelli, M.L., Brents, L.A., Chen, T., Schröck, E., Ried, T. & Kuehl, W.M. *Frequent dysregulation of the c-maf proto-oncogene at 16q23 by translocation to an Ig locus in multiple myeloma.* Blood, 1998, Vol. 91(12), pp. 4457-4463

Choo, K.B., Chen, C.M., Han, C.P., Cheng, W.T. & Au, L.C. *Molecular analysis of cellular loci disrupted by papillomavirus 16 integration in cervical cancer: frequent viral integration in topologically destabilized and transcriptionally active chromosomal regions.* J Med Virol, 1996, Vol. 49(1), pp. 15-22

Clavel, C., Masure, M., Bory, J.P., Putaud, I., Mangeonjean, C., Lorenzato, M., Nazeyrollas, P., Gabriel, R., Quereux, C. & Birembaut, P. *Human papillomavirus testing in primary screening for the detection of high-grade cervical lesions: a study of 7932 women.* Br J Cancer, 2001, Vol. 84(12), pp. 1616-1623

Clifford, G.M., Smith, J.S., Plummer, M., Muñoz, N. & Franceschi, S. *Human papillomavirus types in invasive cervical cancer worldwide: a meta-analysis.* Br J Cancer, 2003, Vol. 88(1), pp. 63-73

Couturier, J., Sastre-Garau, X., Schneider-Maunoury, S., Labib, A. & Orth, G. *Integration of papillomavirus DNA near myc genes in genital carcinomas and its consequences for proto-oncogene expression.* J Virol, 1991, Vol. 65(8), pp. 4534-4538

Cox, J.T., Schiffman, M., Solomon, D. & ASCUS-LSIL Triage Study (ALTS) Group. *Prospective follow-up suggests similar risk of subsequent cervical intraepithelial neoplasia grade 2 or 3 among women with cervical intraepithelial neoplasia grade 1 or negative colposcopy and directed biopsy.* Am J Obstet Gynecol, 2003, Vol. 188(6), pp. 1406-1412

Cullen, A.P., Reid, R., Campion, M. & Lörincz, A.T. *Analysis of the physical state of different human papillomavirus DNAs in intraepithelial and invasive cervical neoplasm.* J Virol, 1991, Vol. 65(2), pp. 606-612

Dahmane, N., Lee, J., Robins, P., Heller, P. & i Altaba, A.R. *Activation of the transcription factor Gli1 and the Sonic hedgehog signalling pathway in skin tumours.* Nature, 1997, Vol. 389(6653), pp. 876-881

Dall, K.L., Scarpini, C.G., Roberts, I., Winder, D.M., Stanley, M.A., Muralidhar, B., Herdman, M.T., Pett, M.R. & Coleman, N. *Characterization of naturally occurring HPV16 integration sites isolated from cervical keratinocytes under noncompetitive conditions.* Cancer Res, 2008, Vol. 68(20), pp. 8249-8259

Dang, C.V. *c-Myc target genes involved in cell growth, apoptosis, and metabolism.* Mol Cell Biol, 1999, Vol. 19(1), pp. 1-11

Daniel, B., Mukherjee, G., Seshadri, L., Vallikad, E. & Krishna, S. *Changes in the physical state and expression of human papillomavirus type 16 in the progression of cervical intraepithelial neoplasia lesions analysed by PCR.* J Gen Virol, 1995, Vol. 76 (Pt 10), pp. 2589-2593

Davy, C. & Doorbar, J. *G2/M cell cycle arrest in the life cycle of viruses.* Virology, 2007, Vol. 368(2), pp. 219-226

Dean, F.B., Hosono, S., Fang, L., Wu, X., Faruqi, A.F., Bray-Ward, P., Sun, Z., Zong, Q., Du, Y., Du, J., Driscoll, M., Song, W., Kingsmore, S.F., Egholm, M. & Lasken, R.S. *Comprehensive human genome amplification using multiple displacement amplification.* Proc Natl Acad Sci USA, 2002, Vol. 99(8), pp. 5261-5266

Demeret, C., Yaniv, M. & Thierry, F. *The E2 transcriptional repressor can compensate for Sp1 activation of the human papillomavirus type 18 early promoter.* J Virol, 1994, Vol. 68(11), pp. 7075-7082

Derynck, R., Akhurst, R.J. & Balmain, A. *TGF-beta signaling in tumor suppression and cancer progression.* Nat Genet, 2001, Vol. 29(2), pp. 117-129

Desaintes, C., Demeret, C., Goyat, S., Yaniv, M. & Thierry, F. *Expression of the papillomavirus E2 protein in HeLa cells leads to apoptosis.* EMBO J, 1997, Vol. 16(3), pp. 504-514

DiPaolo, J.A., Woodworth, C.D., Popescu, N.C., Notario, V. & Doniger, J. *Induction of human cervical squamous cell carcinoma by sequential transfection with human papillomavirus 16 DNA and viral Harvey ras.* Oncogene, 1989, Vol. 4(4), pp. 395-399

Dixon, E.P., Pahel, G.L., Rocque, W.J., Barnes, J.A., Lobe, D.C., Hanlon, M.H., Alexander, K.A., Chao, S.F., Lindley, K. & Phelps, W.C. *The E1 helicase of human papillomavirus type 11 binds to the origin of replication with low sequence specificity.* Virology, 2000, Vol. 270(2), pp. 345-357

Doorbar, J. *Molecular biology of human papillomavirus infection and cervical cancer.* Clin Sci (Lond), 2006, Vol. 110(5), pp. 525-541

Dowhanick, J.J., McBride, A.A. & Howley, P.M. *Suppression of cellular proliferation by the papillomavirus E2 protein.* J Virol, 1995, Vol. 69(12), pp. 7791-7799

Dyson, N., Howley, P.M., Münger, K. & Harlow, E. *The human papilloma virus-16 E7 oncoprotein is able to bind to the retinoblastoma gene product.* Science, 1989, Vol. 243(4893), pp. 934-937

Dürst, M., Croce, C.M., Gissmann, L., Schwarz, E. & Huebner, K. *Papillomavirus sequences integrate near cellular oncogenes in some cervical carcinomas.* Proc Natl Acad Sci USA, 1987, Vol. 84(4), pp. 1070-1074

Dürst, M., Gallahan, D., Jay, G. & Rhim, J.S. *Glucocorticoid-enhanced neoplastic transformation of human keratinocytes by human papillomavirus type 16 and an activated ras oncogene.* Virology, 1989, Vol. 173(2), pp. 767-771

Dürst, M., Gissmann, L., Ikenberg, H. & zur Hausen, H. *A papillomavirus DNA from a cervical carcinoma and its prevalence in cancer biopsy samples from different geographic regions.* Proc Natl Acad Sci USA, 1983, Vol. 80(12), pp. 3812-3815

Dürst, M., Kleinheinz, A., Hotz, M. & Gissmann, L. *The physical state of human papillomavirus type 16 DNA in benign and malignant genital tumours.* J Gen Virol, 1985, Vol. 66 (Pt 7), pp. 1515-1522

Eick, D. & Bornkamm, G.W. *Transcriptional arrest within the first exon is a fast control mechanism in c-myc gene expression.* Nucleic Acids Res, 1986, Vol. 14(21), pp. 8331-8346

Elliott, R.L. & Blobe, G.C. *Role of transforming growth factor Beta in human cancer.* J Clin Oncol, 2005, Vol. 23(9), pp. 2078-2093

Fehrmann, F. & Laimins, L.A. *Human papillomaviruses: targeting differentiating epithelial cells for malignant transformation.* Oncogene, 2003, Vol. 22(33), pp. 5201-5207

Feldmann, G., Dhara, S., Fendrich, V., Bedja, D., Beaty, R., Mullendore, M., Karikari, C., Alvarez, H., Iacobuzio-Donahue, C., Jimeno, A., Gabrielson, K.L., Matsui, W. & Maitra, A. *Blockade of hedgehog signaling inhibits pancreatic cancer invasion and metastases: a new paradigm for combination therapy in solid cancers.* Cancer Res, 2007, Vol. 67(5), pp. 2187-2196

Ferber, M.J., Thorland, E.C., Brink, A.A.T.P., Rapp, A.K., Phillips, L.A., McGovern, R., Gostout, B.S., Cheung, T.H., Chung, T.K.H., Fu, W.Y. & Smith, D.I. *Preferential integration of human papillomavirus type 18 near the c-myc locus in cervical carcinoma.* Oncogene, 2003, Vol. 22(46), pp. 7233-7242

Francis, D.A., Schmid, S.I. & Howley, P.M. *Repression of the integrated papillomavirus E6/E7 promoter is required for growth suppression of cervical cancer cells.* J Virol, 2000, Vol. 74(6), pp. 2679-2686

Franco, E.L., Villa, L.L., Sobrinho, J.P., Prado, J.M., Rousseau, M.C., Désy, M. & Rohan, T.E. *Epidemiology of acquisition and clearance of cervical human papillomavirus infection in women from a high-risk area for cervical cancer.* J Infect Dis, 1999, Vol. 180(5), pp. 1415-1423

FUTURE II Study Group. *Quadrivalent vaccine against human papillomavirus to prevent high-grade cervical lesions*. N Engl J Med, 2007, Vol. 356(19), pp. 1915-1927

Garland, S.M., Hernandez-Avila, M., Wheeler, C.M., Perez, G., Harper, D.M., Leodolter, S., Tang, G.W.K., Ferris, D.G., Steben, M., Bryan, J., Taddeo, F.J., Railkar, R., Esser, M.T., Sings, H.L., Nelson, M., Boslego, J., Sattler, C., Barr, E. & Koutsky, L.A. *Quadrivalent vaccine against human papillomavirus to prevent anogenital diseases.* N Engl J Med, 2007, Vol. 356(19), pp. 1928-1943

Giannoudis, A., Duin, M., Snijders, P.J. & Herrington, C.S. *Variation in the E2-binding domain of HPV 16 is associated with high-grade squamous intraepithelial lesions of the cervix.* Br J Cancer, 2001, Vol. 84(8), pp. 1058-1063

Goodwin, E.C., Naeger, L.K., Breiding, D.E., Androphy, E.J. & DiMaio, D. *Transactivation-competent bovine papillomavirus E2 protein is specifically required for efficient repression of human papillomavirus oncogene expression and for acute growth inhibition of cervical carcinoma cell lines.* J Virol, 1998, Vol. 72(5), pp. 3925-3934

Grandori, C., Cowley, S.M., James, L.P. & Eisenman, R.N. *The Myc/Max/Mad network and the transcriptional control of cell behavior.* Annu Rev Cell Dev Biol, 2000, Vol. 16, pp. 653-699

Green, J., Leigh, I.M., Poulsom, R. & Quinn, A.G. *Basal cell carcinoma development is associated with induction of the expression of the transcription factor Gli-1.* Br J Dermatol, 1998, Vol. 139(5), pp. 911-915

Halbert, C.L., Demers, G.W. & Galloway, D.A. *The E7 gene of human papillomavirus type 16 is sufficient for immortalization of human epithelial cells.* J Virol, 1991, Vol. 65(1), pp. 473-478

Hamdy, F.C., Autzen, P., Robinson, M.C., Horne, C.H., Neal, D.E. & Robson, C.N. *Immunolocalization and messenger RNA expression of bone morphogenetic protein-6 in human benign and malignant prostatic tissue.* Cancer Res, 1997, Vol. 57(19), pp. 4427-4431

Harper, D.M., Franco, E.L., Wheeler, C.M., Moscicki, A.-B., Romanowski, B., Roteli-Martins, C.M., Jenkins, D., Schuind, A., Clemens, S.A.C. & Dubin, G. *Sustained efficacy up to 4.5 years of a bivalent L1 virus-like particle vaccine against human papillomavirus types 16 and 18: follow-up from a randomised control trial.* Lancet, 2006, Vol. 367(9518), pp. 1247-1255

zur Hausen, H. *Papillomaviruses and cancer: from basic studies to clinical application.* Nat Rev Cancer, 2002, Vol. 2(5), pp. 342-350

zur Hausen, H. *Papillomaviruses causing cancer: evasion from host-cell control in early events in carcinogenesis.* J Natl Cancer Inst, 2000, Vol. 92(9), pp. 690-698

zur Hausen, H. *Human papillomaviruses and their possible role in squamous cell carcinomas.* Curr Top Microbiol Immunol, 1977, Vol. 78, pp. 1-30

zur Hausen, H. *Condylomata acuminata and human genital cancer.* Cancer Res, 1976, Vol. 36(2 pt 2), pp. 794

zur Hausen, H., Meinhof, W., Scheiber, W. & Bornkamm, G.W. *Attempts to detect virus-secific DNA in human tumors. I. Nucleic acid hybridizations with complementary RNA of human wart virus.* Int J Cancer, 1974, Vol. 13(5), pp. 650-656

Hawley-Nelson, P., Vousden, K.H., Hubbert, N.L., Lowy, D.R. & Schiller, J.T. *HPV16 E6 and E7 proteins cooperate to immortalize human foreskin keratinocytes.* EMBO J, 1989, Vol. 8(12), pp. 3905-3910

Hecht, J.L. & Aster, J.C. *Molecular biology of Burkitt's lymphoma.* J Clin Oncol, 2000, Vol. 18(21), pp. 3707-3721

Higgins, G.D., Uzelin, D.M., Phillips, G.E., McEvoy, P., Marin, R. & Burrell, C.J. *Transcription patterns of human papillomavirus type 16 in genital intraepithelial neoplasia: evidence for promoter usage within the E7 open reading frame during epithelial differentiation.* J Gen Virol, 1992, Vol. 73 (Pt 8), pp. 2047-2057

Ho, G.Y., Bierman, R., Beardsley, L., Chang, C.J. & Burk, R.D. *Natural history of cervicovaginal papillomavirus infection in young women.* N Engl J Med, 1998, Vol. 338(7), pp. 423-428

Holowaty, P., Miller, A.B., Rohan, T. & To, T. *Natural history of dysplasia of the uterine cervix.* J Natl Cancer Inst, 1999, Vol. 91(3), pp. 252-258

Hopman, A.H.N., Smedts, F., Dignef, W., Ummelen, M., Sonke, G., Mravunac, M., Vooijs, G.P., Speel, E.-J.M. & Ramaekers, F.C.S. *Transition of high-grade cervical intraepithelial neoplasia to micro-invasive carcinoma is characterized by integration of HPV 16/18 and numerical chromosome abnormalities.* J Pathol, 2004, Vol. 202(1), pp. 23-33

Hudelist, G., Manavi, M., Pischinger, K.I.D., Watkins-Riedel, T., Singer, C.F., Kubista, E. & Czerwenka, K.F. *Physical state and expression of HPV DNA in benign and dysplastic cervical tissue: different levels of viral integration are correlated with lesion grade.* Gynecol Oncol, 2004, Vol. 92(3), pp. 873-880

Hudson, J.B., Bedell, M.A., McCance, D.J. & Laiminis, L.A. *Immortalization and altered differentiation of human keratinocytes in vitro by the E6 and E7 open reading frames of human papillomavirus type 18.* J Virol, 1990, Vol. 64(2), pp. 519-526

Hughes, F.J. & Romanos, M.A. *E1 protein of human papillomavirus is a DNA helicase/ATPase.* Nucleic Acids Res, 1993, Vol. 21(25), pp. 5817-5823

Hutchinson, M.L., Cassin, C.M. & Ball, H.G. *The efficacy of an automated preparation device for cervical cytology.* Am J Clin Pathol, 1991, Vol. 96(3), pp. 300-305

Hwang, E.S., Riese, D.J., Settleman, J., Nilson, L.A., Honig, J., Flynn, S. & DiMaio, D. *Inhibition of cervical carcinoma cell line proliferation by the introduction of a bovine papillomavirus regulatory gene.* J Virol, 1993, Vol. 67(7), pp. 3720-3729

Jeon, S., Allen-Hoffmann, B.L. & Lambert, P.F. *Integration of human papillomavirus type 16 into the human genome correlates with a selective growth advantage of cells.* J Virol, 1995, Vol. 69(5), pp. 2989-2997

Jeon, S. & Lambert, P.F. *Integration of human papillomavirus type 16 DNA into the human genome leads to increased stability of E6 and E7 mRNAs: implications for cervical carcinogenesis.* Proc Natl Acad Sci USA, 1995, Vol. 92(5), pp. 1654-1658

Kalantari, M., Blennow, E., Hagmar, B. & Johansson, B. *Physical state of HPV16 and chromosomal mapping of the integrated form in cervical carcinomas.* Diagn Mol Pathol, 2001, Vol. 10(1), pp. 46-54

Kalantari, M., Karlsen, F., Kristensen, G., Holm, R., Hagmar, B. & Johansson, B. *Disruption of the E1 and E2 reading frames of HPV 16 in cervical carcinoma is associated with poor prognosis.* Int J Gynecol Pathol, 1998, Vol. 17(2), pp. 146-153

Kinzler, K.W., Bigner, S.H., Bigner, D.D., Trent, J.M., Law, M.L., O'Brien, S.J., Wong, A.J. & Vogelstein, B. *Identification of an amplified, highly expressed gene in a human glioma.* Science, 1987, Vol. 236(4797), pp. 70-73

Kjaer, S.K., van den Brule, A.J., Bock, J.E., Poll, P.A., Engholm, G., Sherman, M.E., Walboomers, J.M. & Meijer, C.J. *Human papillomavirus--the most significant risk determinant of cervical intraepithelial neoplasia.* Int J Cancer, 1996, Vol. 65(5), pp. 601-606

Kjaer, S.K., van den Brule, A.J.C., Paull, G., Svare, E.I., Sherman, M.E., Thomsen, B.L., Suntum, M., Bock, J.E., Poll, P.A. & Meijer, C.J.L.M. *Type specific persistence of high risk human papillomavirus (HPV) as indicator of high grade cervical squamous intraepithelial lesions in young women: population based prospective follow up study.* BMJ, 2002, Vol. 325(7364), pp. 572

Klaes, R., Woerner, S.M., Ridder, R., Wentzensen, N., Duerst, M., Schneider, A., Lotz, B., Melsheimer, P. & von Knebel Doeberitz, M. *Detection of high-risk cervical intraepithelial neoplasia and cervical cancer by amplification of transcripts derived from integrated papillomavirus oncogenes.* Cancer Res, 1999, Vol. 59(24), pp. 6132-6136

von Knebel Doeberitz, M., Rittmüller, C., zur Hausen, H. & Dürst, M. *Inhibition of tumorigenicity of cervical cancer cells in nude mice by HPV E6-E7 anti-sense RNA.* Int J Cancer, 1992, Vol. 51(5), pp. 831-834

Koutsky, L.A., Holmes, K.K., Critchlow, C.W., Stevens, C.E., Paavonen, J., Beckmann, A.M., DeRouen, T.A., Galloway, D.A., Vernon, D. & Kiviat, N.B. *A cohort study of the risk of cervical intraepithelial neoplasia grade 2 or 3 in relation to papillomavirus infection.* N Engl J Med, 1992, Vol. 327(18), pp. 1272-1278

Kozak, M. *Regulation of translation via mRNA structure in prokaryotes and eukaryotes.* Gene, 2005, Vol. 361, pp. 13-37

Kraus, I., Driesch, C., Vinokurova, S., Hovig, E., Schneider, A., von Knebel Doeberitz, M. & Dürst, M. *The majority of viral-cellular fusion transcripts in cervical carcinomas cotranscribe cellular sequences of known or predicted genes.* Cancer Res, 2008, Vol. 68(7), pp. 2514-2522

Langmore, J.P. *Rubicon Genomics, Inc.* Pharmacogenomics, 2002, Vol. 3(4), pp. 557-560

Lazo, P.A., DiPaolo, J.A. & Popescu, N.C. *Amplification of the integrated viral transforming genes of human papillomavirus 18 and its 5'-flanking cellular sequence located near the myc protooncogene in HeLa cells.* Cancer Res, 1989, Vol. 49(15), pp. 4305-4310

Liaw, K.L., Glass, A.G., Manos, M.M., Greer, C.E., Scott, D.R., Sherman, M., Burk, R.D., Kurman, R.J., Wacholder, S., Rush, B.B., Cadell, D.M., Lawler, P., Tabor, D. & Schiffman, M. *Detection of human papillomavirus DNA in cytologically normal women and subsequent cervical squamous intraepithelial lesions.* J Natl Cancer Inst, 1999, Vol. 91(11), pp. 954-960

Liu, X., Dakic, A., Chen, R., Disbrow, G.L., Zhang, Y., Dai, Y. & Schlegel, R. *Cell-restricted immortalization by human papillomavirus correlates with telomerase activation and engagement of the hTERT promoter by Myc.* J Virol, 2008, Vol. 82(23), pp. 11568-11576

Liu, X., Disbrow, G.L., Yuan, H., Tomaic, V. & Schlegel, R. *Myc and human papillomavirus type 16 E7 genes cooperate to immortalize human keratinocytes.* J Virol, 2007, Vol. 81(22), pp. 12689-12695

Lorenzato, M., Clavel, C., Masure, M., Nou, J.M., Bouttens, D., Evrard, G., Bory, J.P., Maugard, B., Quereux, C. & Birembaut, P. *DNA image cytometry and human papillomavirus (HPV) detection help to select smears at high risk of high-grade cervical lesions.* J Pathol, 2001, Vol. 194(2), pp. 171-176

Luft, F., Klaes, R., Nees, M., Dürst, M., Heilmann, V., Melsheimer, P. & von Knebel Doeberitz, M. *Detection of integrated papillomavirus sequences by ligation-mediated PCR (DIPS-PCR) and molecular characterization in cervical cancer cells.* Int J Cancer, 2001, Vol. 92(1), pp. 9-17

Lörincz, A.T. *Hybrid Capture method for detection of human papillomavirus DNA in clinical specimens: a tool for clinical management of equivocal Pap smears and for population screening.* J Obstet Gynaecol Res, 1996, Vol. 22(6), pp. 629-636

Madison, K.C. *Barrier function of the skin: "la raison d'être" of the epidermis.* J Invest Dermatol, 2003, Vol. 121(2), pp. 231-241

di Magliano, M.P. & Hebrok, M. *Hedgehog signalling in cancer formation and maintenance.* Nat Rev Cancer, 2003, Vol. 3(12), pp. 903-911

Mardis, E.R. *The impact of next-generation sequencing technology on genetics.* Trends Genet, 2008, Vol. 24(3), pp. 133-141

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.-J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.V., Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L.I., Jarvie, T.P., Jirage, K.B., Kim, J.-B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Begley, R.F. & Rothberg, J.M. *Genome sequencing in microfabricated high-density picolitre reactors.* Nature, 2005, Vol. 437(7057), pp. 376-380

Massagué, J., Blain, S.W. & Lo, R.S. *TGFbeta signaling in growth control, cancer, and heritable disorders.* Cell, 2000, Vol. 103(2), pp. 295-309

Masuda, H., Fukabori, Y., Nakano, K., Takezawa, Y., CSuzuki, T. & Yamanaka, H. *Increased expression of bone morphogenetic protein-7 in bone metastatic prostate cancer.* Prostate, 2003, Vol. 54(4), pp. 268-274

Meissner, J.D. *Nucleotide sequences and further characterization of human papillomavirus DNA present in the CaSki, SiHa and HeLa cervical carcinoma cell lines.* J Gen Virol, 1999, Vol. 80 (Pt 7), pp. 1725-1733

Meyer, N. & Penn, L.Z. *Reflecting on 25 years with MYC.* Nat Rev Cancer, 2008, Vol. 8(12), pp. 976-990

Middleton, K., Peh, W., Southern, S., Griffin, H., Sotlar, K., Nakahara, T., El-Sherif, A., Morris, L., Seth, R., Hibma, M., Jenkins, D., Lambert, P., Coleman, N. & Doorbar, J. *Organization of human papillomavirus productive cycle during neoplastic progression provides a basis for selection of diagnostic markers.* J Virol, 2003, Vol. 77(19), pp. 10186-10201

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.* Med Microbiol Immunol, 1987, Vol. 176(5), pp. 245-256

Mohr, I.J., Clark, R., Sun, S., Androphy, E.J., MacPherson, P. & Botchan, M.R. *Targeting the E1 replication protein to the papillomavirus origin of replication by complex formation with the E2 transactivator.* Science, 1990, Vol. 250(4988), pp. 1694-1699

Mori, Y., Okumura, T., Tsunoda, S., Sakai, Y. & Shimada, Y. *Gli-1 expression is associated with lymph node metastasis and tumor progression in esophageal squamous cell carcinoma.* Oncology, 2006, Vol. 70(5), pp. 378-389

Moscicki, A.B., Shiboski, S., Broering, J., Powell, K., Clayton, L., Jay, N., Darragh, T.M., Brescia, R., Kanowitz, S., Miller, S.B., Stone, J., Hanson, E. & Palefsky, J. *The natural history*

of human papillomavirus infection as measured by repeated DNA testing in adolescent and young women. J Pediatr, 1998, Vol. 132(2), pp. 277-284

Moscicki, A.-B., Shiboski, S., Hills, N.K., Powell, K.J., Jay, N., Hanson, E.N., Miller, S., Canjura-Clayton, K.L., Farhat, S., Broering, J.M. & Darragh, T.M. *Regression of low-grade squamous intra-epithelial lesions in young women.* Lancet, 2004, Vol. 364(9446), pp. 1678-1683

Muñoz, N., Bosch, F.X., de Sanjosé, S., Herrero, R., Castellsagué, X., Shah, K.V., Snijders, P.J.F. & Meijer, C.J.L.M. *Epidemiologic classification of human papillomavirus types associated with cervical cancer.* N Engl J Med, 2003, Vol. 348(6), pp. 518-527

Muñoz, N., Bosch, F.X., de Sanjosé, S., Tafur, L., Izarzugaza, I., Gili, M., Viladiu, P., Navarro, C., Martos, C. & Ascunce, N. *The causal link between human papillomavirus and invasive cervical cancer: a population-based case-control study in Colombia and Spain.* Int J Cancer, 1992, Vol. 52(5), pp. 743-749

Münger, K. & Howley, P.M. *Human papillomavirus immortalization and transformation functions.* Virus Res, 2002, Vol. 89(2), pp. 213-228

Münger, K., Phelps, W.C., Bubb, V., Howley, P.M. & Schlegel, R. *The E6 and E7 genes of the human papillomavirus type 16 together are necessary and sufficient for transformation of primary human keratinocytes.* J Virol, 1989b, Vol. 63(10), pp. 4417-4421

Münger, K., Werness, B.A., Dyson, N., Phelps, W.C., Harlow, E. & Howley, P.M. *Complex formation of human papillomavirus E7 proteins with the retinoblastoma tumor suppressor gene product.* EMBO J, 1989a, Vol. 8(13), pp. 4099-4105

Nagao, S., Yoshinouchi, M., Miyagi, Y., Hongo, A., Kodama, J., Itoh, S. & Kudo, T. *Rapid and sensitive detection of physical status of human papillomavirus type 16 DNA by quantitative real-time PCR.* J Clin Microbiol, 2002, Vol. 40(3), pp. 863-867

Narisawa-Saito, M., Yoshimatsu, Y., ichi Ohno, S., Yugawa, T., Egawa, N., Fujita, M., Hirohashi, S. & Kiyono, T. *An in vitro multistep carcinogenesis model for human cervical cancer.* Cancer Res, 2008, Vol. 68(14), pp. 5699-5705

Nilsson, M., Undèn, A.B., Krause, D., Malmqwist, U., Raza, K., Zaphiropoulos, P.G. & Toftgård, R. *Induction of basal cell carcinomas and trichoepitheliomas in mice overexpressing GLI-1.* Proc Natl Acad Sci USA, 2000, Vol. 97(7), pp. 3438-3443

Nishimura, A., Ono, T., Ishimoto, A., Dowhanick, J.J., Frizzell, M.A., Howley, P.M. & Sakai, H. *Mechanisms of human papillomavirus E2-mediated repression of viral oncogene expression and cervical cancer cell growth inhibition.* J Virol, 2000, Vol. 74(8), pp. 3752-3760

Nobbenhuis, M.A., Helmerhorst, T.J., van den Brule, A.J., Rozendaal, L., Voorhorst, F.J., Bezemer, P.D., Verheijen, R.H. & Meijer, C.J. *Cytological regression and clearance of high-risk human papillomavirus in women with an abnormal cervical smear.* Lancet, 2001, Vol. 358(9295), pp. 1782-1783

Nobbenhuis, M.A., Walboomers, J.M., Helmerhorst, T.J., Rozendaal, L., Remmink, A.J., Risse, E.K., van der Linden, H.C., Voorhorst, F.J., Kenemans, P. & Meijer, C.J. *Relation of human papillomavirus status to cervical lesions and consequences for cervical-cancer screening: a prospective study.* Lancet, 1999, Vol. 354(9172), pp. 20-25

Ocadiz, R., Sauceda, R., Cruz, M., Graef, A.M. & Gariglio, P. *High correlation between molecular alterations of the c-myc oncogene and carcinoma of the uterine cervix.* Cancer Res, 1987, Vol. 47(15), pp. 4173-4177

Ostör, A.G. *Natural history of cervical intraepithelial neoplasia: a critical review.* Int J Gynecol Pathol, 1993, Vol. 12(2), pp. 186-192

Papanicolaou, G.N. & Traut, H.F. *The diagnostic value of vaginal smears in carcinoma of the uterus. 1941.* Arch Pathol Lab Med, 1997, Vol. 121(3), pp. 211-224

Pardali, K. & Moustakas, A. *Actions of TGF-beta as tumor suppressor and pro-metastatic factor in human cancer.* Biochim Biophys Acta, 2007, Vol. 1775(1), pp. 21-62

Parkin, D.M., Bray, F., Ferlay, J. & Pisani, P. *Global cancer statistics, 2002.* CA Cancer J Clin, 2005, Vol. 55(2), pp. 74-108

Peh, W.L., Middleton, K., Christensen, N., Nicholls, P., Egawa, K., Sotlar, K., Brandsma, J., Percival, A., Lewis, J., Liu, W.J. & Doorbar, J. *Life cycle heterogeneity in animal models of human papillomavirus-associated disease.* J Virol, 2002, Vol. 76(20), pp. 10401-10416

Pei, X.F., Meck, J.M., Greenhalgh, D. & Schlegel, R. *Cotransfection of HPV-18 and v-fos DNA induces tumorigenicity of primary human keratinocytes.* Virology, 1993, Vol. 196(2), pp. 855-860

Peitsaro, P., Johansson, B. & Syrjänen, S. *Integrated human papillomavirus type 16 is frequently found in cervical cancer precursors as demonstrated by a novel quantitative real-time PCR technique.* J Clin Microbiol, 2002, Vol. 40(3), pp. 886-891

Peter, M., Rosty, C., Couturier, J., Radvanyi, F., Teshima, H. & Sastre-Garau, X. *MYC activation associated with the integration of HPV DNA at the MYC locus in genital tumors.* Oncogene, 2006, Vol. 25(44), pp. 5985-5993

Pett, M. & Coleman, N. *Integration of high-risk human papillomavirus: a key event in cervical carcinogenesis?* J Pathol, 2007, Vol. 212(4), pp. 356-367

Phelps, W.C. & Howley, P.M. *Transcriptional trans-activation by the human papillomavirus type 16 E2 gene product.* J Virol, 1987, Vol. 61(5), pp. 1630-1638

Popescu, N.C., DiPaolo, J.A. & Amsbaugh, S.C. *Integration sites of human papillomavirus 18 DNA sequences on HeLa cell chromosomes.* Cytogenet Cell Genet, 1987, Vol. 44(1), pp. 58-62

Preusser, M., Hoischen, A., Novak, K., Czech, T., Prayer, D., Hainfellner, J.A., Baumgartner, C., Woermann, F.G., Tuxhorn, I.E., Pannek, H.W., Bergmann, M., Radlwimmer, B., Villagrán, R., Weber, R.G. & Hans, V.H. *Angiocentric glioma: report of clinico-pathologic and genetic findings in 8 cases.* Am J Surg Pathol, 2007, Vol. 31(11), pp. 1709-1718

Remmink, A.J., Walboomers, J.M., Helmerhorst, T.J., Voorhorst, F.J., Rozendaal, L., Risse, E.K., Meijer, C.J. & Kenemans, P. *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, 1995, Vol. 61(3), pp. 306-311

Reuter, S., Bartelmann, M., Vogt, M., Geisen, C., Napierski, I., Kahn, T., Delius, H., Lichter, P., Weitz, S., Korn, B. & Schwarz, E. *APM-1, a novel human gene, identified by aberrant co-transcription with papillomavirus oncogenes in a cervical carcinoma cell line, encodes a BTB/POZ-zinc finger protein with growth inhibitory activity.* EMBO J, 1998, Vol. 17(1), pp. 215-222

Richart R.M. *Natural History of Cervical Intraepithelial Neoplasia*. Clinical Obstetrics and Gynecology, 1967, Vol. 10(4), pp. 748-784

Romanczuk, H. & Howley, P.M. *Disruption of either the E1 or the E2 regulatory gene of human papillomavirus type 16 increases viral immortalization capacity.* Proc Natl Acad Sci USA, 1992, Vol. 89(7), pp. 3159-3163

Romanczuk, H., Thierry, F. & Howley, P.M. *Mutational analysis of cis elements involved in E2 modulation of human papillomavirus type 16 P97 and type 18 P105 promoters.* J Virol, 1990, Vol. 64(6), pp. 2849-2859

Ruivenkamp, C.A.L., van Wezel, T., Zanon, C., Stassen, A.P.M., Vlcek, C., Csikós, T., Klous, A.M., Tripodis, N., Perrakis, A., Boerrigter, L., Groot, P.C., Lindeman, J., Mooi, W.J., Meijjer, G.A., Scholten, G., Dauwerse, H., Paces, V., van Zandwijk, N., van Ommen, G.J.B. & Demant, P. *Ptprj is a candidate for the mouse colon-cancer susceptibility locus Scc1 and is frequently deleted in human cancers.* Nat Genet, 2002, Vol. 31(3), pp. 295-300

Sakai, H., Yasugi, T., Benson, J.D., Dowhanick, J.J. & Howley, P.M. *Targeted mutagenesis of the human papillomavirus type 16 E2 transactivation domain reveals separable transcriptional activation and DNA replication functions.* J Virol, 1996, Vol. 70(3), pp. 1602-1611

Sarkar, G., Turner, R.T. & Bolander, M.E. *Restriction-site PCR: a direct method of unknown sequence retrieval adjacent to a known locus by using universal primers.* PCR Methods Appl, 1993, Vol. 2(4), pp. 318-322

Sastre-Garau, X., Favre, M., Couturier, J. & Orth, G. *Distinct patterns of alteration of myc genes associated with integration of human papillomavirus type 16 or type 45 DNA in two genital tumours.* J Gen Virol, 2000, Vol. 81(Pt 8), pp. 1983-1993

Saunier, M., Monnier-Benoit, S., Mauny, F., Dalstein, V., Briolat, J., Riethmuller, D., Kantelip, B., Schwarz, E., Mougin, C. & Prétet, J.-L. *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.* J Clin Microbiol, 2008, Vol. 46(11), pp. 3678-3685

Scheffner, M., Huibregtse, J.M., Vierstra, R.D. & Howley, P.M. *The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53.* Cell, 1993, Vol. 75(3), pp. 495-505

Scheffner, M., Werness, B.A., Huibregtse, J.M., Levine, A.J. & Howley, P.M. *The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53.* Cell, 1990, Vol. 63(6), pp. 1129-1136

Schiffman, M., Castle, P.E., Jeronimo, J., Rodriguez, A.C. & Wacholder, S. *Human papillomavirus and cervical cancer.* Lancet, 2007, Vol. 370(9590), pp. 890-907

Schiffman, M.H., Bauer, H.M., Hoover, R.N., Glass, A.G., Cadell, D.M., Rush, B.B., Scott, D.R., Sherman, M.E., Kurman, R.J. & Wacholder, S. *Epidemiologic evidence showing that human papillomavirus infection causes most cervical intraepithelial neoplasia.* J Natl Cancer Inst, 1993, Vol. 85(12), pp. 958-964

Schlecht, N.F., Kulaga, S., Robitaille, J., Ferreira, S., Santos, M., Miyamura, R.A., Duarte-Franco, E., Rohan, T.E., Ferenczy, A., Villa, L.L. & Franco, E.L. *Persistent human papillomavirus infection as a predictor of cervical intraepithelial neoplasia.* JAMA, 2001, Vol. 286(24), pp. 3106-3114

Schneider-Gädicke, A. & Schwarz, E. *Different human cervical carcinoma cell lines show similar transcription patterns of human papillomavirus type 18 early genes.* EMBO J, 1986, Vol. 5(9), pp. 2285-2292

Schwaninger, R., Rentsch, C.A., Wetterwald, A., van der Horst, G., van Bezooijen, R.L., van der Pluijm, G., Löwik, C.W.G.M., Ackermann, K., Pyerin, W., Hamdy, F.C., Thalmann, G.N. & Cecchini, M.G. *Lack of noggin expression by cancer cells is a determinant of the osteoblast response in bone metastases.* Am J Pathol, 2007, Vol. 170(1), pp. 160-175

Schwarz, E., Freese, U.K., Gissmann, L., Mayer, W., Roggenbuck, B., Stremlau, A. & zur Hausen, H. *Structure and transcription of human papillomavirus sequences in cervical carcinoma cells.* Nature, 1985, Vol. 314(6006), pp. 111-114

Seedorf, K., Krämmer, G., Dürst, M., Suhai, S. & Röwekamp, W.G. *Human papillomavirus type 16 DNA sequence.* Virology, 1985, Vol. 145(1), pp. 181-185

Shendure, J. & Ji, H. *Next-generation DNA sequencing.* Nat Biotechnol, 2008, Vol. 26(10), pp. 1135-1145

Sherman, L., Jackman, A., Itzhaki, H., Stöppler, M.C., Koval, D. & Schlegel, R. *Inhibition of serum- and calcium-induced differentiation of human keratinocytes by HPV16 E6 oncoprotein: role of p53 inactivation.* Virology, 1997, Vol. 237(2), pp. 296-306

Smith, J.S., Lindsay, L., Hoots, B., Keys, J., Franceschi, S., Winer, R. & Clifford, G.M. *Human papillomavirus type distribution in invasive cervical cancer and high-grade cervical lesions: a meta-analysis update.* Int J Cancer, 2007, Vol. 121(3), pp. 621-632

Snijders, P.J.F., Steenbergen, R.D.M., Heideman, D.A.M. & Meijer, C.J.L.M. *HPV-mediated cervical carcinogenesis: concepts and clinical implications.* J Pathol, 2006, Vol. 208(2), pp. 152-164

Solomon, D., Davey, D., Kurman, R., Moriarty, A., O'Connor, D., Prey, M., Raab, S., Sherman, M., Wilbur, D., Wright, T. & Young, N. *The 2001 Bethesda System: terminology for reporting results of cervical cytology.* JAMA, 2002, Vol. 287(16), pp. 2114-2119

Steger, G. & Corbach, S. *Dose-dependent regulation of the early promoter of human papillomavirus type 18 by the viral E2 protein.* J Virol, 1997, Vol. 71(1), pp. 50-58

Steinmeyer, N. *Mutationen und Integration von DNA humaner Hochrisiko-Papillomviren in anogenitalen Krebsvorstufen und Karzinomen*. Diplomarbeit, 2009, Hochschule Darmstadt

Strobl, L.J. & Eick, D. *Hold back of RNA polymerase II at the transcription start site mediates down-regulation of c-myc in vivo.* EMBO J, 1992, Vol. 11(9), pp. 3307-3314

Suzuki, M., Shigematsu, H., Shames, D.S., Sunaga, N., Takahashi, T., Shivapurkar, N., Iizasa, T., Frenkel, E.P., Minna, J.D., Fujisawa, T. & Gazdar, A.F. *DNA methylation-associated inactivation of TGFbeta-related genes DRM/Gremlin, RUNX3, and HPP1 in human cancers.* Br J Cancer, 2005, Vol. 93(9), pp. 1029-1037

Tan, S.H., Gloss, B. & Bernard, H.U. *During negative regulation of the human papillomavirus-16 E6 promoter, the viral E2 protein can displace Sp1 from a proximal promoter element.* Nucleic Acids Res, 1992, Vol. 20(2), pp. 251-256

Tan, S.H., Leong, L.E., Walker, P.A. & Bernard, H.U. *The human papillomavirus type 16 E2 transcription factor binds with low cooperativity to two flanking sites and represses the E6 promoter through displacement of Sp1 and TFIID.* J Virol, 1994, Vol. 68(10), pp. 6411-6420

Thierry, F. & Howley, P.M. *Functional analysis of E2-mediated repression of the HPV18 P105 promoter.* New Biol, 1991, Vol. 3(1), pp. 90-100

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, 2003, Vol. 22(8), pp. 1225-1237

Thorland, E.C., Myers, S.L., Persing, D.H., Sarkar, G., McGovern, R.M., Gostout, B.S. & Smith, D.I. *Human papillomavirus type 16 integrations in cervical tumors frequently occur in common fragile sites.* Cancer Res, 2000, Vol. 60(21), pp. 5916-5921

Tikidzhieva, A. *Ektopische Expression des in einer Zervixkarzinom-Zelllinie nachgewiesenen Fusionsproteins MycHPV16E2 zur Analyse der Wirkungen auf Tumorzellwachstum und Promotoraktivität*. Diplomarbeit, 2009, Universität Heidelberg

Tobin, S.W., Douville, K., Benbow, U., Brinckerhoff, C.E., Memoli, V.A. & Arrick, B.A. *Consequences of altered TGF-beta expression and responsiveness in breast cancer: evidence for autocrine and paracrine effects.* Oncogene, 2002, Vol. 21(1), pp. 108-118

Tonks, N.K. *Protein tyrosine phosphatases: from genes, to function, to disease.* Nat Rev Mol Cell Biol, 2006, Vol. 7(11), pp. 833-846

Tonon, S.A., Picconi, M.A., Bos, P.D., Zinovich, J.B., Galuppo, J., Alonio, L.V. & Teyssie, A.R. *Physical status of the E2 human papilloma virus 16 viral gene in cervical preneoplastic and neoplastic lesions.* J Clin Virol, 2001, Vol. 21(2), pp. 129-134

Unger, E.R., Vernon, S.D., Thoms, W.W., Nisenbaum, R., Spann, C.O., Horowitz, I.R., Icenogle, J.P. & Reeves, W.C. *Human papillomavirus and disease-free survival in FIGO stage Ib cervical cancer.* J Infect Dis, 1995, Vol. 172(5), pp. 1184-1190

Veldman, T., Liu, X., Yuan, H. & Schlegel, R. *Human papillomavirus E6 and Myc proteins associate in vivo and bind to and cooperatively activate the telomerase reverse transcriptase promoter.* Proc Natl Acad Sci USA, 2003, Vol. 100(14), pp. 8211-8216

Vernon, S.D., Unger, E.R., Miller, D.L., Lee, D.R. & Reeves, W.C. *Association of human papillomavirus type 16 integration in the E2 gene with poor disease-free survival from cervical cancer.* Int J Cancer, 1997, Vol. 74(1), pp. 50-56

Villa, L.L., Costa, R.L.R., Petta, C.A., Andrade, R.P., Ault, K.A., Giuliano, A.R., Wheeler, C.M., Koutsky, L.A., Malm, C., Lehtinen, M., Skjeldestad, F.E., Olsson, S.-E., Steinwall, M., Brown, D.R., Kurman, R.J., Ronnett, B.M., Stoler, M.H., Ferenczy, A., Harper, D.M., Tamms, G.M., Yu, J., Lupinacci, L., Railkar, R., Taddeo, F.J., Jansen, K.U., Esser, M.T., Sings, H.L., Saah, A.J. & Barr, E. *Prophylactic quadrivalent human papillomavirus (types 6, 11, 16, and 18) L1 virus-like particle vaccine in young women: a randomised double-blind placebo-controlled multicentre phase II efficacy trial.* Lancet Oncol, 2005, Vol. 6(5), pp. 271-278

de Villiers, E.-M., Fauquet, C., Broker, T.R., Bernard, H.-U. & zur Hausen, H. *Classification of papillomaviruses.* Virology, 2004, Vol. 324(1), pp. 17-27

Vinokurova, S., Wentzensen, N., Kraus, I., Klaes, R., Driesch, C., Melsheimer, P., Kisseljov, F., Dürst, M., Schneider, A. & von Knebel Doeberitz, M. *Type-dependent integration frequency of human papillomavirus genomes in cervical lesions.* Cancer Res, 2008, Vol. 68(1), pp. 307-313

Vizcaino, A.P., Moreno, V., Bosch, F.X., Muñoz, N., Barros-Dios, X.M. & Parkin, D.M. *International trends in the incidence of cervical cancer: I. Adenocarcinoma and adenosquamous cell carcinomas.* Int J Cancer, 1998, Vol. 75(4), pp. 536-545

van Vlodrop, I.J.H., Baldewijns, M.M.L., Smits, K.M., Schouten, L.J., van Neste, L., van Criekinge, W., van Poppel, H., Lerut, E., Schuebel, K.E., Ahuja, N., Herman, J.G., de Bruïne, A.P. & van Engeland, M. *Prognostic significance of Gremlin1 (GREM1) promoter CpG island hypermethylation in clear cell renal cell carcinoma.* Am J Pathol, 2010, Vol. 176(2), pp. 575-584

Walboomers, J.M., Jacobs, M.V., Manos, M.M., Bosch, F.X., Kummer, J.A., Shah, K.V., Snijders, P.J., Peto, J., Meijer, C.J. & Muñoz, N. *Human papillomavirus is a necessary cause of invasive cervical cancer worldwide.* J Pathol, 1999, Vol. 189(1), pp. 12-19

Wallin, K.L., Wiklund, F., Angström, T., Bergman, F., Stendahl, U., Wadell, G., Hallmans, G. & Dillner, J. *Type-specific persistence of human papillomavirus DNA before the development of invasive cervical cancer.* N Engl J Med, 1999, Vol. 341(22), pp. 1633-1638

Wanschura, S. *Untersuchungen zur Karyotypstabilität und HPV-Integration bei Cervicxcarcinomzellinien*. Diplomarbeit, 1992, Universität Bremen

Wentzensen, N., Ridder, R., Klaes, R., Vinokurova, S., Schaefer, U. & von Knebel Doeberitz, M. *Characterization of viral-cellular fusion transcripts in a large series of HPV16 and 18 positive anogenital lesions.* Oncogene, 2002, Vol. 21(3), pp. 419-426

Wentzensen, N., Vinokurova, S. & von Knebel Doeberitz, M. *Systematic review of genomic integration sites of human papillomavirus genomes in epithelial dysplasia and invasive cancer of the female lower genital tract.* Cancer Res, 2004, Vol. 64(11), pp. 3878-3884

Werness, B.A., Levine, A.J. & Howley, P.M. *Association of human papillomavirus types 16 and 18 E6 proteins with p53.* Science, 1990, Vol. 248(4951), pp. 76-79

Woodman, C.B., Collins, S., Winter, H., Bailey, A., Ellis, J., Prior, P., Yates, M., Rollason, T.P. & Young, L.S. *Natural history of cervical human papillomavirus infection in young women: a longitudinal cohort study.* Lancet, 2001, Vol. 357(9271), pp. 1831-1836

Yanagita, M. *BMP antagonists: their roles in development and involvement in pathophysiology.* Cytokine Growth Factor Rev, 2005, Vol. 16(3), pp. 309-317

Yee, C., Krishnan-Hewlett, I., Baker, C.C., Schlegel, R. & Howley, P.M. *Presence and expression of human papillomavirus sequences in human cervical carcinoma cell lines.* Am J Pathol, 1985, Vol. 119(3), pp. 361-366

Yuen, H.-F., Chan, Y.-P., Cheung, W.-L., Wong, Y.-C., Wang, X. & Chan, K.-W. *The prognostic significance of BMP-6 signaling in prostate cancer.* Mod Pathol, 2008, Vol. 21(12), pp. 1436-1443

Zheng, Z.-M. & Baker, C.C. *Papillomavirus genome structure, expression, and post-transcriptional regulation.* Front Biosci, 2006, Vol. 11, pp. 2286-2302

Ziegert, C., Wentzensen, N., Vinokurova, S., Kisseljov, F., Einenkel, J., Hoeckel, M. & von Knebel Doeberitz, M. *A comprehensive analysis of HPV integration loci in anogenital lesions combining transcript and genome-based amplification techniques.* Oncogene, 2003, Vol. 22(25), pp. 3977-3984

Appendix 1: viral-cellular junction sequences identified in cell lines

The 3' junction sequence of SiHa

This junction was determined by restriction-site PCR (see *Section 2.1.1*). The viral sequence part is shown in red, from HPV16 position 1588 to 3133. The blue sequence indicates the cellular part from position 74087558 to 74087156 of chromosome 13 (NC_000013.10).

GGACTTACACCCAGTATAGCTGACAGTATAAAAACACTATTACAACAATATTGTTTATATTTACACATT
CAAAGTTTAGCATGTTCATGGGGAATGGTTGTGTTACTATTAGTAAGATATAAATGTGGAAAAAATAGA
GAAACAATTGAAAAATTGCTGTCTAAACTATTATGTGTGTCTCCAATGTGTATGATGATAGAGCCTCCA
AAATTGCGTAGTACAGCAGCAGCATTATATTGGTATAAAACAGGTATATCAAATATTAGTGAAGTGTAT
GGAGACACGCCAGAATGGATACAAAGACAAACAGTATTACAACATAGTTTTAATGATTGTACATTTGAA
TTATCACAGATGGTACAATGGGCCTACGATAATGACATAGTAGACGATAGTGAAATTGCATATAAATAT
GCACAATTGGCAGACACTAATAGTAATGCAAGTGCCTTTCTAAAAAGTAATTCACAGGCAAAAATTGTA
AAGGATTGTGCAACAATGTGTAGACATTATAAACGAGCAGAAAAAAAACAAATGAGTATGAGTCAATGG
ATAAAATATAGATGTGATAGGGTAGATGATGGAGGTGATTGGAAGCAAATTGTTATGTTTTTAAGGTAT
CAAGGTGTAGAGTTTATGTCATTTTTAACTGCATTAAAAAGATTTTTGCAAGGCATACCTAAAAAAAAT
TGCATATTACTATATGGTGCAGCTAACACAGGTAAATCATTATTTGGTATGAGTTTAATGAAATTTCTG
CAAGGGTCTGTAATATGTTTTGTAAATTCTAAAAGCCATTTTTGGTTACAACCATTAGCAGATGCCAAA
ATAGGTATGTTAGATGATGCTACAGTGCCCTGTTGGAACTACATAGATGACAATTTAAGAAATGCATTG
GATGGAAATTTAGTTTCTATGGATGTAAAGCATAGACCATTGGTACAACTAAAATGCCCTCCATTATTA
ATTACATCTAACATTAATGCTGGTACAGATTCTAGGTGGCCTTATTTACATAATAGATTGGTGGTGTTT
ACATTTCCTAATGAGTTTCCATTTGACGAAAACGGAAATCCAGTGTATGAGCTTAATGATAAGAACTGG
AAATCCTTTTTCTCAAGGACGTGGTCCAGATTAAGTTTGCACGAGGACGAGGACAAGGAAAACGATGGA
GACTCTTTGCCAACGTTTAAATGTGTGTCAGGACAAAATACTAACACATTATGAAAATGATAGTACAGA
CCTACGTGACCATATAGACTATTGGAAACACATGCGCCTAGAATGTGCTATTTATTACAAGGCCAGAGA
AATGGGATTTAAACATATTAACCACCAGGTGGTGCCAACACTGGCTGTATCAAAGAATAAAGCATTACA
AGCAATTGAACTGCAACTAACGTTAGAAACAATATATAACTCACAATATAGTAATGAAAAGTGGACATT
ACAAGACGTTAGCCTTGAAGTGTATTTAACTGCACCAACAGGATGTATAAAAAAAACATGGATATACAGT
GGAAGTGCAGTTTGATGGAGACATATGCTGCTCAGCTTAGCTGATTAACCCTGGACAGATAAGACTGGA
GCCTATTGGGCATTGTCACAATGTCTAAAGATGACTGTTAGGAACCTGCTTCTGTGACTAGTGGTGGGG
ATAAGGGAATCCCAATGAAGGACACCTCAGATAACCCAGTATGTCTACTATACTTATATATCTAAATTC
TCATTGTAATTCCACCGGATTCACACTTATCACAGCTATGGAATATTTTATGAATTAGGTTGTTCTAGT
GAGTTCTTCGGAAGCCTGGAAATATAGTCCCCAGTTTGAGGATGCCACGTCTTGTGGCCTCAGTCATGG
CAGTAAGGCACAGTTGAGGTAGCTTGGGCTCTGCAACTGGAAAATAAGGGGTTGAAAGTTCTGGTAACT
TTTCATAGTGGGGGCTC

The 3' junction sequence of MRI-H186

This junction was determined by restriction-site PCR (see *Section 2.1.2*). The viral sequence part is shown in red, from HPV16 position 1588 to 2754. The blue sequence indicates the cellular part from position 128815788 to 128816380 of chromosome 8 (NC_000008.9).

GGACTTACACCCAGTATAGCTGACAGTATAAAAACACTATTACAACAATATTGTTTATATTTACACATT
CAAAGTTTAGCATGTTCATGGGGAATGGTTGTGTTACTATTAGTAAGATATAAATGTGGAAAAAATAGA
GAAACAATTGAAAAATTGCTGTCTAAACTATTATGTGTGTCTCCAATGTGTATGATGATAGAGCCTCCA
AAATTGCGTAGTACAGCAGCAGCATTATATTGGTATAAAACAGGTATATCAAATATTAGTGAAGTGTAT
GGAGACACGCCAGAATGGATACAAAGACAAACAGTATTACAACATAGTTTTAATGATTGTACATTTGAA
TTATCACAGATGGTACAATGGGCCTACGATAATGACATAGTAGACGATAGTGAAATTGCATATAAATAT
GCACAATTGGCAGACACTAATAGTAATGCAAGTGCCTTTCTAAAAAGTAATTCACAGGCAAAAATTGTA
AAGGATTGTGCAACAATGTGTAGACATTATAAACGAGCAGAAAAAAAACAAATGAGTATGAGTCAATGG
ATAAAATATAGATGTGATAGGGTAGATGATGGAGGTGATTGGAAGCAAATTGTTATGTTTTTAAGGTAT
CAAGGTGTAGAGTTTATGTCATTTTTAACTGCATTAAAAAGATTTTTGCAAGGCATACCTAAAAAAAAT
TGCATATTACTATATGGTGCAGCTAACACAGGTAAATCATTATTTGGTATGAGTTTAATGAAATTTCTG
CAAGGGTCTGTAATATGTTTTGTAAATTCTAAAAGCCATTTTTGGTTACAACCATTAGCAGATGCCAAA
ATAGGTATGTTAGATGATGCTACAGTGCCCTGTTGGAACTACATAGATGACAATTTAAGAAATGCATTG
GATGGAAATTTAGTTTCTATGGATGTAAAGCATAGACCATTGGTACAACTAAAATGCCCTCCATTATTA
ATTACATCTAACATTAATGCTGGTACAGATTCTAGGTGGCCTTATTTACATAATAGATTGGTGGTGTTT

ACATTTCCTAATGAGTTTCCATTTGACGAAAACGGAAATCCAGTGTATGAGCTTAATGATAAGAACTGG
AAATCCTTTTTCTCAAGGACGTGGTCCAGATTAAGTTTGCACGAGGACGAGGACAAGGAAAACCCATCT
TGAACAGCGTACATGCTATACACGCACCCCTTTCCCCCGAATTGTTTTCTCTTTTGGAGGTGGTGGAGG
GAGAGAAAAGTTTACTTAAAATGCCTTTGGGTGAGGGACCAAGGATGAGAAGAATGTTTTTTGTTTTTC
ATGCCGTGGAATAACACAAAATAAAAAATCCCGAGGGAATATACATTATATATTAAATATAGATCATTT
CAGGGAGCAAACAAATCATGTGTGGGGCTGGGCAACTAGCTAAGTCGAAGCGTAAATAAAATGTGAATA
CACGTTTGCGGGTTACATACAGTGCACTTTCACTAGTATTCAGAAAAAATTGTGAGTCAGTGAACTAGG
AAATTAATGCCTGGAAGGCAGCCAAATTTTAATTAGCTCAAGACTCCCCCCCCCCCAAAAAAAGGCACG
GAAGTAATACTCCTCTCCTCTTCTTTGATCAGAATCGATGCATTTTTTGTGCATGACCGCATTTCCAAT
AATAAAAGGGGAAAGAGGACCTGGAAAGGAATTAAACGTCCGGTTTGTCCGGGGAGGAAAGAGTTAACG
GTTTTTTTCACAAGGGTCTCTGCTGACTCCCCCGG

The 5' junction sequence of MRI-H186

This junction was determined by restriction-site PCR (see *Section 2.1.2*). The blue sequence indicates the cellular part from position 128814789 to 128815779 of chromosome 8 (NC_000008.9). The viral sequence part is shown in red, from HPV16 position 4871 to 5885.

GATCGCTTTAACCGAGGAGTTTGAGACCAGCCTGGGCAACATGACCAGACTGCCTCTCTACAAAAAGTT
TAAAAAATTAACCGGGTGTGGTGGTGCACTGCACTCCCAGCTACTGGGCTGGGGTATCAGGCTGAGGTA
GGAGGTTTGCTTTGAGCCCGGGGGGATCGAGGCTGCAGTGAGCTTTGATTGTGCCACTGCACTCCAGCC
TGGGTGACAGAAGGAGACCCTGTCTCAAAAATAATAAGAATAATAATTAATAATAATAGGCCAAACCAA
ATACCCATCACCTTCTGCTGTGCCTCCCCTTTCCCCAATAAATCCAGTGTCTTGCTTTCAAATTTTGTG
GTTAAAAAAGATGATGAGTTTCTAAGACGTGGGGGCTAAAGCTTGTTTGGCCGTTTTAGGGTTTGTTGG
AATTTTTTTTTCGTCTATGTACTTGTGAATTATTTCACGTTTGCCATTACCGGTTCTCCATAGGGTGAT
GTTCATTAGCAGTGGTGATAGGTTAATTTTCACCATCTCTTATGCGGTTGAATAGTCACCTCTGAACCA
CTTTTTCCTCCAGTAACTCCTCTTTCTTCGGACCTTCTGCAGCCAACCTGAAAGAATAACAAGGAGGTG
GCTGGAAACTTGTTTTAAGGAACCGCCTGTCCTTCCCCCGCTGGAAACCTTGCACCTCGGACGCTCCTG
CTCCTGCCCCCACCTGACCCCCGCCCTCGTTGACATCCAGGCGCGATGATCTCTGCTGCCAGTAGAGGG
CACACTTACTTTACTTTCGCAAACCTGAACGCGGGTGCTGCCCAGAGAGGGGGCGGAGGGAAAGACGCT
TTGCAGCAAAATCCAGCATAGCGATTGGTTGCTCCCCGCGTTTGCGGCAAAGGCCTGGAGGCAGGAGTA
ATTTGCAATCCTTAAAGCTGAATTGTGCAGTGCATCGGATTTGGAAGCTACTATATTCACTTAACACTT
GAACGCTGAGCTGCAAACTCAACGGGTAGCACACCCATACCAGGGTCTCGCCCAGTGGCACGCCTAGGA
TTATATAGTCGCACAACACAACAGGTTAAAGTTGTAGACCCTGCTTTTGTAACCACTCCCACTAAACTT
ATTACATATGATAATCCTGCATATGAAGGTATAGATGTGGATAATACATTATATTTTTCTAGTAATGAT
AATAGTATTAATATAGCTCCAGATCCTGACTTTTTGGATATAGTTGCTTTACATAGGCCAGCATTAACC
TCTAGGCGTACTGGCATTAGGTACAGTAGAATTGGTAATAAACAAACACTACGTACTCGTAGTGGAAAA
TCTATAGGTGCTAAGGTACATTATTATTATGATTTAAGTACTATTGATCCTGCAGAAGAAATAGAATTA
CAAACTATAACACCTTCTACATATACTACCACTTCACATGCAGCCTCACCTACTTCTATTAATAATGGA
TTATATGATATTTATGCAGATGACTTTATTACAGATACTTCTACAACCCCGGTACCATCTGTACCCTCT
ACATCTTTATCAGGTTATATTCCTGCAAATACAACAATTCCTTTTGGTGGTGCATACAATATTCCTTTA
GTATCAGGTCCTGATATACCCATTAATATAACTGACCAAGCTCCTTCATTAATTCCTATAGTTCCAGGG
TCTCCACAATATACAATTATTGCTGATGCAGGTGACTTTTATTTACATCCTAGTTATTACATGTTACGA
AAACGACGTAAACGTTTACCATATTTTTTTTCAGATGTCTCTTTGGCTGCCTAGTGAGGCCACTGTCTA
CTTGCCTCCTGTCCCAGTATCTAAGGTTGTAAGCACGGATGAATATGTTGCACGCACAAACATATATTA
TCATGCAGGAACATCCAGACTACTTGCAGTTGGACATCCCTATTTTCCTATTAAAAAACCTAACAATAA
CAAAATATTAGTTCCTAAAGTATCAGGATTACAATACAGGGTATTTAGAATACATTTACCTGACCCCAA
TAAGT

The 3' junction sequence of MRI-H196

This junction was determined by restriction-site PCR (see *Section 2.1.2*). The viral sequence part is shown in red, from HPV16 position 2387 to 3858. The blue sequence indicates the cellular part from position 47967861 to 47968023 of chromosome 11 (NC_000011.9). The sequence in black (TTAC) does not match with both the HPV16 and cellular sequences.

TTTGGTTACAACCATTAGCAGATGCCAAAATAGGTATGTTAGATGATGCTACAGTGCCCTGTTGGAACT
ACATAGATGACAATTTAAGAAATGCATTGGATGGAAATTTAGTTTCTATGGATGTAAAGCATAGACCAT
TGGTACAACTAAAATGCCCTCCATTATTAATTACATCTAACATTAATGCTGGTACAGATTCTAGGTGGC
CTTATTTACATAATAGATTGGTGGTGTTTACATTTCCTAATGAGTTTCCATTTGACGAAAACGGAAATC
CAGTGTATGAGCTTAATGATAAGAACTGGAAATCCTTTTTCTCAAGGACGTGGTCCAGATTAAGTTTGC
ACGAGGACGAGGACAAGGAAAACGATGGAGACTCTTTGCCAACGTTTAAATGTGTGTCAGGACAAAATA
CTAACACATTATGAAAATGATAGTACAGACCTACGTGACCATATAGACTATTGGAAACACATGCGCCTA
GAATGTGCTATTTATTACAAGGCCAGAGAAATGGGATTTAAACATATTAACCACCAGGTGGTGCCAACA
CTGGCTGTATCAAAGAATAAAGCATTACAAGCAATTGAACTGCAACTAACGTTAGAAACAATATATAAC
TCACAATATAGTAATGAAAAGTGGACATTACAAGACGTTAGCCTTGAAGTGTATTTAACTGCACCAACA
GGATGTATAAAAAAACATGGATATACAGTGGAAGTGCAGTTTGATGGAGACATATGCAATACAATGCAT
TATACAAACTGGACACATATATATATTTGTGAAGAAGCATCAGTAACTGTGGTAGAGGGTCAAGTTGAC
TATTATGGTTTATATTATGTTCATGAAGGAATACGAACATATTTTGTGCAGTTTAAAGATGATGCAGAA
AAATATAGTAAAAATAAAGTATGGGAAGTTCATGCGGGTGGTCAGGTAATATTATGTCCTACATCTGTG
TTTAGCAGCAACGAAGTATCCTCTCCTGAAATTATTAGGCAGCACTTGGCCAACCACCCCGCCGCGACC
CATACCAAAGCCGTCGCCTTGGGCACCGAAGAAACACAGACGACTATCCAGCGACCAAGATCAGAGCCA
GACACCGGAAACCCCTGCCACACCACTAAGTTGTTGCACAGAGACTCAGTGGACAGTGCTCCAATCCTC
ACTGCATTTAACAGCTCACACAAAGGACGGATTAACTGTAATAGTAACACTACACCCATAGTACATTTA
AAAGGTGATGCTAATACTTTAAAATGTTTAAGATATAGATTTAAAAAGCATTGTACATTGTATACTGCA
GTGTCGTCTACATGGCATTGGACAGGACATAATGTAAAACATAAAAGTGCAATTGTTACACTTACATAT
GATAGTGAATGGCAACGTGACCAATTTTTGTCTCAAGTTAAAATACCAAAAACTATTACAGTGTCTACT
GGATTTATGTCTATATGACAAATTTACCCATTGTAGGCTTCACTGGGAAGATTAATATTTAGTAAATAC
CTGAATTAGGTGCGAGAGCTGGCCATGTGGCTGTCTGAGGCAAAAGCATTCTAAGCATAGAAAACAACA
AATGTGGCGGCTAGGCGCGGTGGCTCACACCTGCAGCCCAGAACTTTGGGAG

The 5' junction sequence of MRI-H196

This junction was determined by restriction-site PCR but not completely sequenced (see *Section 2.1.2*). Therefore, it is shown in discontinuous sequence parts. The blue sequence indicates the cellular part from position 46953161 to 46953669 of chromosome 11 (NC_000011.9). The viral sequence part is shown in red, from HPV16 position 5321 to 6653.

GAGGGAAACAAATTCATTTAACAGACGTTTATTGTATAACTGATTTCTGCAGGCACTGCACATGGTCCC
AGAAATTTAAAGATGAATGAAGACTGTCTCTGCCCTCATGGAACTCACAATCTAGTAGGAAGAATTTTT
TAAAAAGACATAAATGGTTATAAAAGTGCATACTAGCTGGGCGCGGTGGCTCACGCCTGTAATCCCAGC
ACTTTAGGAGGCTGAGGCGGGTGGATCACCTGAGGTCAGGAGTTCGAGAGCAGCCTGACCAAAATGGAG
AAACCCCGTCTCTACTAAAAATACAAAATCAGCCAGGTGCAGTGCGCATGCCTGTAATCCCAGCTACTT
GGGAGGCTGAGGCAGGAAAATCTCTTGAACCCGGGAGGTGGAGGTTGCAGTGAGCCGAGATCACGCCAC
TGCAGTCCAGCCTGGGTGAAAGAGCGAGTCTCTGTCTCAAAAAAACAAAACAAAACAAAACAAAACAAA
ACAAAACAAAACTGCATAGTCTAGGA
ATAATGGATTATATGATATTTATGCAGATGACTTTATTACAGATACTTCTACAACCCCGGTACCATCTG
TACCCTCTACATCTTTATCAGGTTATATTCCTGCAAATACAACAATTCCTTTTGGTGGTGCATACAATA
TTCCTTTAGTATCAGGTCCTGATATACCCATTAATATAACTGACCAAGCTCCTTCATTAATTCCTATAG
TTCCAGGGTCTCCACAATATACAATTATTGCTGATGCAGGTGACTTTTATTTACATCCTAGTTATTACA
TGTTACGAAAACGACGTAAACGTTTACCATATTTTTTTTCAGATGTCTCTTTGGCTGCCTAGTGAGGCC
ACTGTCTACTTGCCTCCTGTCCCAGTATCTAAGGTTGTAAGCACGGATGAATATGTTGCACGCACAAAC
ATATATTATCATGCAGGAACATCCAGACTACTTGCAGTTGGACATCCCTATTTTCCTATTAAAAAACCT
AACAATAACAAAATATTAGTTCCTAAAGTATCAGGATTACAATACAGGGTATTTAGAATACATTTACCT

GACCCCAATAAGTTTGGTTTTCCTGACACCTCATTTTATAATCCAGATACACAGCGGCTGGTTTGGGCC
TGTGTAGGTGTTGAGGTAGGTCGTGGTCAGCCATTAGGTGTGGGCATTAGTGGCCATCCTTTATTAAAT
AAATTGGATGACACAGAAAATGCTAGTGCTTATGCAGCAAATGCAGGTGTGGATAATAGAGAATGTATA
TCTATGGATTACAAACAAACACAATTGTGTTTAATTGGTTGCAAACCACCTATAGGGGAACACTGGGGC
AAAGGATCCCCATGTACCAATGTTGCAGTAAATCCAGGTGATTGTCCACCATTAGAGTTAATAAACACA
GTTATTCAGGATGGTGATATGGTTGATACTGGCTTTGGTGCTATGGACTTTACTACATTACAGGCTAAC
AAAAGTGAAGTTCCACTGGATATTTGTACATCTATTTGCAAATATCCAGATTATATTAAAATGGTGTCA
GAACCATATGGCGACAGCTTATTTTTTTATTTACGAAGGGAACAAATGTTTGTTAGACATTTATTTAAT
AGGGCTGGTACTGTTGGTGAAAATGTACCAGACGATTTATACATTAAAGGCTCTGGGTCTACTGCAAAT
TTAGCCAGTTCAAATTATTTTCCTACACCTAGTGGTTCTATGGTTACCTCTGATGCCCAAATATTCAAT
AAACCTTATTGGTTACAACGAGCACAGGGCCACAATAATGGCATTTGTTGGGGTAACCAACTATTTGTT
ACTGTTGTTGATACTACACGCA

The 3' junction sequence of integration variant B in MRI-H186

This junction was determined by DIPS-PCR (see *Section 2.8.2*). The viral sequence part is shown in red, from HPV16 position 911 to 1223. The blue sequence indicates the cellular part from position 128744999 to 128745226 of chromosome 8 (NC_000008.9). The sequence in black (32 nt) does not match with both the HPV16 and cellular sequences.

ATGGATGGTTTTATGTAGAGGCTGTAGTGGAAAAAAAAACAGGGGATGCTATATCAGATGACGAGAACG
AAAATGACAGTGATACAGGTGAAGATTTGGTAGATTTTATAGTAAATGATAATGATTATTTAACACAGG
CAGAAACAGAGACAGCACATGCGTTGTTTACTGCACAGGAAGCAAAACAACATAGAGATGCAGTACAGG
TTCTAAAACGAAAGTATTTGGGTAGTCCACTTAGTGATATTAGTGGATGTGTAGACAATAATATTAGTC
CTAGATTAAAAGCTATATGTATAGAAAAACAAAGTAGAACCATCAGGGCTCTTACAGTTGTTAAAAGTG
ATGTGTCACTTTTAAACAATGGATATTTTGAATTCCAGAGAAAATGGTGTTAGGTAAAGAGTGACATTT
TATTATAAATTTCCTCCAATGTCATCTCATTTACTTTTAAATTGAACTTAGGTGTGTTTGAACTTGTCC
CTGAACTCTTAGTCACTCAAAACACAAACACACAGTTTGTAAAGAAAGTTTGCCTCCTGTGAGACCCAC
AAGGCAGAGGGATTTGATGGGT

Appendix 2: *myc*-HPV16 hybrid mRNA and MycHPV16E2 fusion protein

cDNA sequence of *myc*-HPV16 hybrid mRNA

The sequence with a length of 1607 nt is shown below, starting from the first nucleotide of *c-myc* exon1 to the early poly-A site of HPV16 (AATAAA, pos. 4215 – 4220). Details about the hybrid mRNA are explained in *Section 2.5*.

```
ACCCCCGAGCTGTGCTGCTCGCGGCCGCCACCGCCGGGCCCCGGCCGTCCCTGGCTCCCCTCCTGCCTC
GAGAAGGGCAGGGCTTCTCAGAGGCTTGGCGGGAAAAAGAACGGAGGGAGGGATCGCGCTGAGTATAAA
AGCCGGTTTTCGGGGCTTTATCTAACTCGCTGTAGTAATTCCAGCGAGAGGCAGAGGGAGCGAGCGGGC
GGCCGGCTAGGGTGGAAGAGCCGGGCGAGCAGAGCTGCGCTGCGGGCGTCCTGGGAAGGGAGATCCGGA
GCGAATAGGGGGCTTCGCCTCTGGCCCAGCCCTCCCGCTGATCCCCCAGCCAGCGGTCCGCAACCCTTG
CCGCATCCACGAAACTTTGCCCATAGCAGCGGGCGGGCACTTTGCACTGGAACTTACAACACCCGAGCA
AGGACGCGACTCTCCCGACGCGGGGAGGCTATTCTGCCCATTTGGGGACACTTCCCCGCCGCTGCCAGG
ACCCGCTTCTCTGAAAGGCTCTCCTTGCAGCTGCTTAGACGCTGGATTTTTTTCGGGTAGTGGAAAACC
AGCAGCCTCCCGCGACGATGCCCCTCAACGTTAGCTTCACCAACAGGAACTATGACCTCGACTACGACT
CGGTGCAGCCGTATTTCTACTGCGACGAGGAGGAGAACTTCTACCAGCAGCAGCAGCAGAGCGAGCTGC
AGCCCCCGGCGCCCAGCGAGGATATCTGGAAGAAATTCGAGCTGCTGCCCACCCCGCCAAGCAGCAACG
AAGTATCCTCTCCTGAAATTATTAGGCAGCACTTGGCCAACCACCCCGCCGCGACCCATACCAAAGCCG
TCGCCTTGGGCACCGAAGAAACACAGACGACTATCCAGCGACCAAGATCAGAGCCAGACACCGGAAACC
CCTGCCACACCACTAAGTTGTTGCACAGAGACTCAGTGGACAGTGCTCCAATCCTCACTGCATTTAACA
GCGCACACAAAGGACGGATTAACTGTAATAGTAACACTACACCCATAGTACATTTAAAAGGTGATGCTA
ATACTTTAAAATGTTTAAGATATAGATTTAAAAAGCATTGTACATTGTATACTGCAGTGTCGTCTACAT
GGCATTGGACAGGACATAATGTAAAACATAAAAGTGCAATTGTTACACTTACATATGATAGTGAATGGC
AACGTGACCAATTTTTGTCTCAAGTTAAAATACCAAAAACTATTACAGTGTCTACTGGATTTATGTCTA
TATGACAAATCTTGATACTGCATCCACAACATTACTGGCGTGCTTTTTGCTTTGCTTTTGTGTGCTTTT
GTGTGTCTGCCTATTAATACGTCCGCTGCTTTTGTCTGTGTCTACATACACATCATTAATAATATTGGT
ATTACTATTGTGGATAACAGCAGCCTCTGCGTTTAGGTGTTTTATTGTATATATTATATTTGTTTATAT
ACCATTATTTTTAATACATACACATGCACGCTTTTTAATTACATAATTACATAATGTAATTGTTACATA
TAATTGTTGTATACCATAACTTACTATTTTTTCTTTTTTATTTTCATATATAATTTTTTTTTGTGTTTG
TTTGTTTGTTTTTTAATAAA
```

MycHPV16E2 fusion protein sequence

The 225 amino acids sequence of MycHPV16E2 fusion protein is shown below. Details about this protein are explained in *Section 2.7.1*.

```
MPLNVSFTNRNYDLDYDSVQPYFYCDEEENFYQQQQQSELQPPAPSEDIWKKFELLPTPPSSNEVSSPE
IIRQHLANHPAATHTKAVALGTEETQTTIQRPRSEPDTGNPCHTTKLLHRDSVDSAPILTAFNSAHKGR
INCNSNTTPIVHLKGDANTLKCLRYRFKKHCTLYTAVSSTWHWTGHNVKHKSAIVTLTYDSEWQRDQFL
SQVKIPKTITVSTGFMSI
```

Appendix 3: HPV16 genome reference sequence (see *Section 4.11.2*)

ACTACAATAATTCATGTATAAAACTAAGGGCGTAACCGAAATCGGTTGAACCGAAACCGGTTAGTATAA
AAGCAGACATTTTATGCACCAAAAGAGAACTGCAATGTTTCAGGACCCACAGGAGCGACCCAGAAAGTT
ACCACAGTTATGCACAGAGCTGCAAACAACTATACATGATATAATATTAGAATGTGTGTACTGCAAGCA
ACAGTTACTGCGACGTGAGGTATATGACTTTGCTTTTCGGGATTTATGCATAGTATATAGAGATGGGAA
TCCATATGCTGTATGTGATAAATGTTTAAAGTTTTATTCTAAAATTAGTGAGTATAGACATTATTGTTA
TAGTTTGTATGGAACAACATTAGAACAGCAATACAACAAACCGTTGTGTGATTTGTTAATTAGGTGTAT
TAACTGTCAAAAGCCACTGTGTCCTGAAGAAAAGCAAAGACATCTGGACAAAAAGCAAAGATTCCATAA
TATAAGGGGTCGGTGGACCGGTCGATGTATGTCTTGTTGCAGATCATCAAGAACACGTAGAGAAACCCA
GCTGTAATCATGCATGGAGATACACCTACATTGCATGAATATATGTTAGATTTGCAACCAGAGACAACT
GATCTCTACTGTTATGAGCAATTAAATGACAGCTCAGAGGAGGAGGATGAAATAGATGGTCCAGCTGGA
CAAGCAGAACCGGACAGAGCCCATTACAATATTGTAACCTTTTGTTGCAAGTGTGACTCTACGCTTCGG
TTGTGCGTACAAAGCACACACGTAGACATTCGTACTTTGGAAGACCTGTTAATGGGCACACTAGGAATT
GTGTGCCCCATCTGTTCTCAGAAACCATAATCTACCATGGCTGATCCTGCAGGTACCAATGGGGAAGAG
GGTACGGGATGTAATGGATGGTTTTATGTAGAGGCTGTAGTGGAAAAAAAAACAGGGGATGCTATATCA
GATGACGAGAACGAAAATGACAGTGATACAGGTGAAGATTTGGTAGATTTTATAGTAAATGATAATGAT
TATTTAACACAGGCAGAAACAGAGACAGCACATGCGTTGTTTACTGCACAGGAAGCAAAACAACATAGA
GATGCAGTACAGGTTCTAAAACGAAAGTATTTGGGTAGTCCACTTAGTGATATTAGTGGATGTGTAGAC
AATAATATTAGTCCTAGATTAAAAGCTATATGTATAGAAAAACAAAGTAGAGCTGCAAAAAGGAGATTA
TTTGAAAGCGAAGACAGCGGGTATGGCAATACTGAAGTGGAAACTCAGCAGATGTTACAGGTAGAAGGG
CGCCATGAGACTGAAACACCATGTAGTCAGTATAGTGGTGGAAGTGGGGGTGGTTGCAGTCAGTACAGT
AGTGGAAGTGGGGGAGAGGGTGTTAGTGAAAGACACACTATATGCCAAACACCACTTACAAATATTTTA
AATGTACTAAAAACTAGTAATGCAAAGGCAGCAATGTTAGCAAAATTTAAAGAGTTATACGGGGTGAGT
TTTTCAGAATTAGTAAGACCATTTAAAAGTAATAAATCAACGTGTTGCGATTGGTGTATTGCTGCATTT
GGACTTACACCCAGTATAGCTGACAGTATAAAAACACTATTACAACAATATTGTTTATATTTACACATT
CAAAGTTTAGCATGTTCATGGGGAATGGTTGTGTTACTATTAGTAAGATATAAATGTGGAAAAAATAGA
GAAACAATTGAAAAATTGCTGTCTAAACTATTATGTGTGTCTCCAATGTGTATGATGATAGAGCCTCCA
AAATTGCGTAGTACAGCAGCAGCATTATATTGGTATAAAACAGGTATATCAAATATTAGTGAAGTGTAT
GGAGACACGCCAGAATGGATACAAAGACAAACAGTATTACAACATAGTTTTAATGATTGTACATTTGAA
TTATCACAGATGGTACAATGGGCCTACGATAATGACATAGTAGACGATAGTGAAATTGCATATAAATAT
GCACAATTGGCAGACACTAATAGTAATGCAAGTGCCTTTCTAAAAAGTAATTCACAGGCAAAAATTGTA
AAGGATTGTGCAACAATGTGTAGACATTATAAACGAGCAGAAAAAAAACAAATGAGTATGAGTCAATGG
ATAAAATATAGATGTGATAGGGTAGATGATGGAGGTGATTGGAAGCAAATTGTTATGTTTTTAAGGTAT
CAAGGTGTAGAGTTTATGTCATTTTTAACTGCATTAAAAAGATTTTTGCAAGGCATACCTAAAAAAAAT
TGCATATTACTATATGGTGCAGCTAACACAGGTAAATCATTATTTGGTATGAGTTTAATGAAATTTCTG
CAAGGGTCTGTAATATGTTTTGTAAATTCTAAAAGCCATTTTTGGTTACAACCATTAGCAGATGCCAAA
ATAGGTATGTTAGATGATGCTACAGTGCCCTGTTGGAACTACATAGATGACAATTTAAGAAATGCATTG
GATGGAAATTTAGTTTCTATGGATGTAAAGCATAGACCATTGGTACAACTAAAATGCCCTCCATTATTA
ATTACATCTAACATTAATGCTGGTACAGATTCTAGGTGGCCTTATTTACATAATAGATTGGTGGTGTTT
ACATTTCCTAATGAGTTTCCATTTGACGAAAACGGAAATCCAGTGTATGAGCTTAATGATAAGAACTGG
AAATCCTTTTTCTCAAGGACGTGGTCCAGATTAAGTTTGCACGAGGACGAGGACAAGGAAAACGATGGA
GACTCTTTGCCAACGTTTAAATGTGTGTCAGGACAAAATACTAACACATTATGAAAATGATAGTACAGA
CCTACGTGACCATATAGACTATTGGAAACACATGCGCCTAGAATGTGCTATTTATTACAAGGCCAGAGA
AATGGGATTTAAACATATTAACCACCAGGTGGTGCCAACACTGGCTGTATCAAAGAATAAAGCATTACA
AGCAATTGAACTGCAACTAACGTTAGAAACAATATATAACTCACAATATAGTAATGAAAAGTGGACATT
ACAAGACGTTAGCCTTGAAGTGTATTTAACTGCACCAACAGGATGTATAAAAAAACATGGATATACAGT
GGAAGTGCAGTTTGATGGAGACATATGCAATACAATGCATTATACAAACTGGACACATATATATATTTG
TGAAGAAGCATCAGTAACTGTGGTAGAGGGTCAAGTTGACTATTATGGTTTATATTATGTTCATGAAGG
AATACGAACATATTTTGTGCAGTTTAAAGATGATGCAGAAAAATATAGTAAAAATAAAGTATGGGAAGT
TCATGCGGGTGGTCAGGTAATATTATGTCCTACATCTGTGTTTAGCAGCAACGAAGTATCCTCTCCTGA
AATTATTAGGCAGCACTTGGCCAACCACCCCGCCGCGACCCATACCAAAGCCGTCGCCTTGGGCACCGA
AGAAACACAGACGACTATCCAGCGACCAAGATCAGAGCCAGACACCGGAAACCCCTGCCACACCACTAA
GTTGTTGCACAGAGACTCAGTGGACAGTGCTCCAATCCTCACTGCATTTAACAGCTCACACAAAGGACG
GATTAACTGTAATAGTAACACTACACCCATAGTACATTTAAAAGGTGATGCTAATACTTTAAAATGTTT
AAGATATAGATTTAAAAAGCATTGTACATTGTATACTGCAGTGTCGTCTACATGGCATTGGACAGGACA
TAATGTAAAACATAAAAGTGCAATTGTTACACTTACATATGATAGTGAATGGCAACGTGACCAATTTTT
GTCTCAAGTTAAAATACCAAAAACTATTACAGTGTCTACTGGATTTATGTCTATATGACAAATCTTGAT
ACTGCATCCACAACATTACTGGCGTGCTTTTTGCTTTGCTTTTGTGTGCTTTTGTGTGTCTGCCTATTA
ATACGTCCGCTGCTTTTGTCTGTGTCTACATACACATCATTAATAATATTGGTATTACTATTGTGGATA
ACAGCAGCCTCTGCGTTTAGGTGTTTTATTGTATATATTATATTTGTTTATATACCATTATTTTTAATA
CATACACATGCACGCTTTTTAATTACATAATGTATATGTACATAATGTAATTGTTACATATAATTGTTG
TATACCATAACTTACTATTTTTTCTTTTTTATTTTCATATATAATTTTTTTTTTTGTTTGTTTGTTTGT

TTTTTAATAAACTGTTATTACTTAACAATGCGACACAAACGTTCTGCAAAACGCACAAAACGTGCATCG
GCTACCCAACTTTATAAAACATGCAAACAGGCAGGTACATGTCCACCTGACATTATACCTAAGGTTGAA
GGCAAAACTATTGCTGATCAAATATTACAATATGGAAGTATGGGTGTATTTTTTGGTGGGTTAGGAATT
GGAACAGGGTCGGGTACAGGCGGACGCACTGGGTATATTCCATTGGGAACAAGGCCTCCCACAGCTACA
GATACACTTGCTCCTGTAAGACCCCCTTTAACAGTAGATCCTGTGGGCCCTTCTGATCCTTCTATAGTT
TCTTTAGTGGAAGAAACTAGTTTTATTGATGCTGGTGCACCAACATCTGTACCTTCCATTCCCCCAGAT
GTATCAGGATTTAGTATTACTACTTCAACTGATACCACACCTGCTATATTAGATATTAATAATACTGTT
ACTACTGTTACTACACATAATAATCCCACTTTCACTGACCCATCTGTATTGCAGCCTCCAACACCTGCA
GAAACTGGAGGGCATTTTACACTTTCATCATCCACTATTAGTACACATAATTATGAAGAAATTCCTATG
GATACATTTATTGTTAGCACAAACCCTAACACAGTAACTAGTAGCACACCCATACCAGGGTCTCGCCCA
GTGGCACGCCTAGGATTATATAGTCGCACAACACAACAGGTTAAAGTTGTAGACCCTGCTTTTGTAACC
ACTCCCACTAAACTTATTACATATGATAATCCTGCATATGAAGGTATAGATGTGGATAATACATTATAT
TTTTCTAGTAATGATAATAGTATTAATATAGCTCCAGATCCTGACTTTTTGGATATAGTTGCTTTACAT
AGGCCAGCATTAACCTCTAGGCGTACTGGCATTAGGTACAGTAGAATTGGTAATAAACAAACACTACGT
ACTCGTAGTGGAAAATCTATAGGTGCTAAGGTACATTATTATTATGATTTAAGTACTATTGATCCTGCA
GAAGAAATAGAATTACAAACTATAACACCTTCTACATATACTACCACTTCACATGCAGCCTCACCTACT
TCTATTAATAATGGATTATATGATATTTATGCAGATGACTTTATTACAGATACTTCTACAACCCCGGTA
CCATCTGTACCCTCTACATCTTTATCAGGTTATATTCCTGCAAATACAACAATTCCTTTTGGTGGTGCA
TACAATATTCCTTTAGTATCAGGTCCTGATATACCCATTAATATAACTGACCAAGCTCCTTCATTAATT
CCTATAGTTCCAGGGTCTCCACAATATACAATTATTGCTGATGCAGGTGACTTTTATTTACATCCTAGT
TATTACATGTTACGAAAACGACGTAAACGTTTACCATATTTTTTTTCAGATGTCTCTTTGGCTGCCTAG
TGAGGCCACTGTCTACTTGCCTCCTGTCCCAGTATCTAAGGTTGTAAGCACGGATGAATATGTTGCACG
CACAAACATATATTATCATGCAGGAACATCCAGACTACTTGCAGTTGGACATCCCTATTTTCCTATTAA
AAAACCTAACAATAACAAAATATTAGTTCCTAAAGTATCAGGATTACAATACAGGGTATTTAGAATACA
TTTACCTGACCCCAATAAGTTTGGTTTTCCTGACACCTCATTTTATAATCCAGATACACAGCGGCTGGT
TTGGGCCTGTGTAGGTGTTGAGGTAGGTCGTGGTCAGCCATTAGGTGTGGGCATTAGTGGCCATCCTTT
ATTAAATAAATTGGATGACACAGAAAATGCTAGTGCTTATGCAGCAAATGCAGGTGTGGATAATAGAGA
ATGTATATCTATGGATTACAAACAAACACAATTGTGTTTAATTGGTTGCAAACCACCTATAGGGGAACA
CTGGGGCAAAGGATCCCCATGTACCAATGTTGCAGTAAATCCAGGTGATTGTCCACCATTAGAGTTAAT
AAACACAGTTATTCAGGATGGTGATATGGTTGATACTGGCTTTGGTGCTATGGACTTTACTACATTACA
GGCTAACAAAAGTGAAGTTCCACTGGATATTTGTACATCTATTTGCAAATATCCAGATTATATTAAAAT
GGTGTCAGAACCATATGGCGACAGCTTATTTTTTTATTTACGAAGGGAACAAATGTTTGTTAGACATTT
ATTTAATAGGGCTGGTACTGTTGGTGAAAATGTACCAGACGATTTATACATTAAAGGCTCTGGGTCTAC
TGCAAATTTAGCCAGTTCAAATTATTTTCCTACACCTAGTGGTTCTATGGTTACCTCTGATGCCCAAAT
ATTCAATAAACCTTATTGGTTACAACGAGCACAGGGCCACAATAATGGCATTTGTTGGGGTAACCAACT
ATTTGTTACTGTTGTTGATACTACACGCAGTACAAATATGTCATTATGTGCTGCCATATCTACTTCAGA
AACTACATATAAAAATACTAACTTTAAGGAGTACCTACGACATGGGGAGGAATATGATTTACAGTTTAT
TTTTCAACTGTGCAAAATAACCTTAACTGCAGACGTTATGACATACATACATTCTATGAATTCCACTAT
TTTGGAGGACTGGAATTTTGGTCTACAACCTCCCCCAGGAGGCACACTAGAAGATACTTATAGGTTTGT
AACATCCCAGGCAATTGCTTGTCAAAAACATACACCTCCAGCACCTAAAGAAGATCCCCTTAAAAAATA
CACTTTTTGGGAAGTAAATTTAAAGGAAAAGTTTTCTGCAGACCTAGATCAGTTTCCTTTAGGACGCAA
ATTTTTACTACAAGCAGGATTGAAGGCCAAACCAAAATTTACATTAGGAAAACGAAAAGCTACACCCAC
CACCTCATCTACCTCTACAACTGCTAAACGCAAAAAACGTAAGCTGTAAGTATTGTATGTATGTTGAAT
TAGTGTTGTTTGTTGTGTATATGTTTGTATGTGCTTGTATGTGCTTGTAAATATTAAGTTGTATGTGTG
TTTGTATGTATGGTATAATAAACACGTGTGTATGTGTTTTTAAATGCTTGTGTAACTATTGTGTCATGC
AACATAAATAAACTTATTGTTTCAACACCTACTAATTGTGTTGTGGTTATTCATTGTATATAAACTATA
TTTGCTACATCCTGTTTTTGTTTTATATATACTATATTTTGTAGCGCCAGCGGCCATTTTGTAGCTTCA
ACCGAATTCGGTTGCATGCTTTTTGGCACAAAATGTGTTTTTTTAAATAGTTCTATGTCAGCAACTATG
GTTTAAACTTGTACGTTTCCTGCTTGCCATGCGTGCCAAATCCCTGTTTTCCTGACCTGCACTGCTTGC
CAACCATTCCATTGTTTTTTACACTGCACTATGTGCAACTACTGAATCACTATGTACATTGTGTCATAT
AAAATAAATCACTATGCGCCAACGCCTTACATACCGCTGTTAGGCACATATTTTTGGCTTGTTTTAACT
AACCTAATTGCATATTTGGCATAAGGTTTAAACTTCTAAGGCCAACTAAATGTCACCCTAGTTCATACA
TGAACTGTGTAAAGGTTAGTCATACATTGTTCATTTGTAAAACTGCACATGGGTGTGTGCAAACCGTTT
TGGGTTACACATTTACAAGCAACTTATATAATAATACTAA