

Preferential integration of human papillomavirus type 18 near the *c-myc* locus in cervical carcinoma

Matthew J Ferber¹, Erik C Thorland¹, Antoinette ATP Brink², Anton K Rapp², Leslie A Phillips³, Renee McGovern⁴, Bobbie S Gostout⁴, Tak Hong Cheung⁵, Tong Kwok Hung Chung⁵, Wong Yick Fu⁵ and David I Smith^{*3}

¹Department of Biochemistry and Molecular Biology, Mayo Clinic, Rochester, MN 55905, USA; ²Department of Molecular Cell Biology, Leiden University Medical Center, Leiden 2333AL, The Netherlands; ³Division of Experimental Pathology, Mayo Clinic, Rochester, MN 55905, USA; ⁴Department of Obstetrics and Gynaecology, Mayo Clinic, Rochester, MN 55905, USA; ⁵Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong, Prince of Wales Hospital, Hong Kong, People's Republic of China

The development of cervical cancer is highly associated with human papillomavirus (HPV) infection. Greater than 99% of all cervical tumors contain HPV DNA. Integration of high-risk HPV has been temporally associated with the acquisition of a malignant phenotype. Recent work from our lab has shown that HPV16, the most common high-risk HPV associated with cervical carcinoma, preferentially integrates at loci containing human common fragile sites (CFSs). CFSs are regions of genomic instability that have also been associated with deletions, translocations, and gene amplification during cancer development. The current work shows that HPV18, the second most prevalent high-risk HPV type found in cervical tumors, preferentially targets the CFSs. We identified 27 unique HPV18 integrations in cervical tumors, of which 63% ($P < 0.001$) occur in CFSs. However, the distribution of HPV18 integrations found were profoundly different from those found for HPV16. Specifically, 30% of all HPV18 integrations occurred within the chromosomal band 8q24 near the *c-myc* proto-oncogene. None of the HPV16 integrations occurred in this region. Previous low-resolution mapping suggested that *c-myc* may be a target of HPV integration. Our data at nucleotide resolution confirm that in HPV18-positive cervical tumors, the region surrounding *c-myc* is indeed a hot spot of viral integration. These results demonstrate that CFSs are preferred sites of integration for HPV18 in cervical tumors. In addition, we have identified multiple cellular genes that have been disrupted by HPV18 integration in cervical tumors. Our results suggest that the sites of HPV18 integration are nonrandom and may play an important role in the development of cervical tumors.

Oncogene (2003) 22, 7233–7242. doi:10.1038/sj.onc.1207006

Keywords: common fragile sites; cervical cancer; human papillomavirus; viral integration; *c-myc*

Introduction

Cervical cancer is highly associated with the presence of human papillomavirus (HPV) infection. Greater than 99% of cervical cancers contain HPV sequences (Walboomers *et al.*, 1999; Bosch *et al.*, 2002). HPVs are a family of circular double-stranded DNA viruses that are subdivided into two classes, low and high risk, based on their association with benign lesions or invasive carcinomas, respectively. The two most commonly observed high-risk HPV subtypes found in cervical cancers are HPV16 and HPV18. These viruses infect basal epithelial cells and are frequently found in the episomal state in low- and high-grade intraepithelial lesions (LSIL and HSIL, respectively) (Cho, 1998). Integration of the HPV genome frequently results in the functional elimination of the viral E2 repressor and remodeling of the E6/E7 promoter, thus leading to the continued expression of the E6 and E7 oncoproteins (Ham *et al.*, 1991; McBride *et al.*, 1991; Bechtold *et al.*, 2003). These are potent antagonists to the p53 (Scheffner *et al.*, 1990) and pRB (Boyer *et al.*, 1996) tumor suppressor genes, and their overexpression contributes to the malignant phenotype. However, overexpression of E6 and E7 is only sufficient to immortalize primary epithelial cells; additional alterations such as activating ras mutations are required to transform them (Crook *et al.*, 1988). This suggests that additional human mutations need to take place in order to transform the cells.

We previously described work on the identification of multiple HPV16 integrations in cervical tumors using a rapid PCR-based technique to identify human sequences flanking the sites of viral integration (Thorland *et al.*, 2000, 2003). This work revealed that 48% ($P < 0.001$) of the integrations occurred within common fragile sites (CFSs). CFSs are described as specific loci that form gaps or breaks in metaphase chromosomes from cells that have been challenged with chemicals that induce replicative stress (Glover *et al.*, 1984). There are 89 CFSs present in the human genome, most of which are induced by the DNA polymerase inhibitor aphidicolin. Unlike the rare fragile sites (FRA11B, FRAXA, FRAXE, etc.), CFSs are present in all primates, lower

*Correspondence: DI Smith, Mayo Clinic and Foundation, 200 First Street SW, Rochester, MN 55905, USA;

E-mail: smith.david@mayo.edu

Received 7 April 2003; revised 11 July 2003; accepted 16 July 2003

mammals, and possibly yeast (Cha and Kleckner, 2002). The mechanism of CFS fragility is just beginning to be understood. Recent data suggest that the Ataxia-telangiectasia and Rad3-related (ATR) gene products are involved in maintaining DNA integrity at CFS loci (Casper *et al.*, 2002). Loss of ATR expression was shown to produce breakage without the addition of exogenous chemicals, indicating that without proper genome surveillance, CFSs are expressed in otherwise unaltered cells. This expression can promote genomic events such as translocations, sister chromatid exchange, chromosomal deletions and rearrangements, and most pertinent to the current study, the integration of tumor-associated viruses (Glover and Stein, 1987, 1988; Rassool *et al.*, 1991; Thorland *et al.*, 2000).

In this report, we describe our work characterizing HPV18 integrations by RS-PCR. HPV18 is the second most common high-risk HPV subtype, and is found in approximately 20% of cervical tumors (Walboomers *et al.*, 1999), has been associated with more aggressive adenocarcinomas of the uterine cervix (Kurman *et al.*, 1988). Here, we report on the identification of 27 unique HPV18 integrations in cervical tumors. These sites of HPV18 integrations were distinct from those we observed for HPV16 (Thorland *et al.*, 2000, 2003). Specifically, we have identified a cluster of HPV18 integrations surrounding the *c-myc* proto-oncogene similar to those already reported (Durst *et al.*, 1987; Popescu and DiPaolo, 1989; Couturier *et al.*, 1991).

Our previous analysis of HPV16 integration loci revealed a number of genes implicated in cancer development at or near the sites of viral integration, suggesting that alterations of these genes could play an important role in cervical tumorigenesis. In a similar fashion, we have identified multiple genes that are interrupted by HPV18 integration. These genes include tumor suppressors, proto-oncogenes, as well as genes involved in maintaining DNA integrity. Our results support a model of nonrandom HPV integrations in cervical tumors and indicate that genes at or near the sites of integration may play an important role in tumor development.

Materials and methods

Samples

HPV18-positive cervical carcinoma samples were obtained from 32 patients and three cell lines. The 32 cervical tumors were derived from patients from the Mayo Clinic, Leiden University, or the Chinese University of Hong Kong. The three cell lines used were HeLa, SW756, and C4-I (all available from ATCC). Cell culture was performed as recommended by the supplier. DNA was extracted using the IsoQuick kit (ORCA Research). Studies were limited to sensitive DNA-based methodologies.

PCR

RS-PCR is a technique that allows for the amplification of unknown sequence that lies adjacent to known sequence (Sarkar *et al.*, 1993). Specific conditions and reagents used, including RS primers and HPV18-specific primers, were

previously published (Ferber *et al.*, 2003). Briefly, by using a battery of HPV18-specific primers in conjunction with RS primers, which are semispecific for restriction endonuclease cleavage sites found throughout the human genome, in a two-step heminested PCR system, we are able to amplify HPV18-human junction fragments from cervical tumors. RS-PCR products were electrophoresed on 2.0% agarose gels. Positive bands were selected for sequencing. Additional PCR using HPV18- and human locus-specific primers was performed to confirm the RS-PCR results when a specific junction fragment was identified by sequence analysis (data not shown) (Ferber *et al.*, 2003).

Sequencing

PCR products were prepared for sequencing as previously reported (Ferber *et al.*, 2003). In brief, 5 μ l of each positive RS-PCR product was treated with exonuclease-1 and shrimp alkaline phosphatase. Treated aliquots were combined with 3.2 pmol of the proper sequencing primer and submitted to the Mayo Clinic Molecular Core Facility for sequencing.

Computer analysis, BAC clone selection, and FISH

BLASTN searches were performed as previously described (Thorland *et al.*, 2000). Results from these searches were used to identify bacterial artificial chromosome (BAC) clones and genes, if present, which represented the loci of viral integration. BACs were purified and used as probes for FISH analysis as described in Thorland *et al.* (2003) (Figure 1). These BAC clones were used as FISH probes on normal donor lymphocytes that had been stimulated to express CFSs by adding the DNA polymerase α/δ subunit inhibitor aphidicolin. Metaphase spreads were hybridized with fluorescently labeled BACs representing each integration loci to determine where each integration occurred relative to the CFS from that region (Figure 2a). Fluorescently labeled BACs hybridize to aphidicolin-induced breakage in one of the three possible signal-to-break relationships, proximal, crossing, or distal. Proximal defines a signal-to-break relationship where the labeled BAC hybridizes on the centromeric side of the breakage at the specific fragile site under examination. If breakage spans the region represented by the BAC probe, hybridization signal will be detected on both sides producing a crossing signal pattern. Distal defines a signal-to-break relationship where the labeled BAC hybridizes on the telomeric side of the breakage at the specific fragile site under examination (Figure 2b). A total of 20 clearly visible breaks are counted for each FS under examination. For each BAC, if 19 metaphases have proximal signals while the remaining has a single crossing or distal signal, the BAC is said to be on the fragile site, although at the proximal end of that CFS region. The same is true for a BAC that has one proximal and 19 distal hybridization signals, although that BAC would be localized at the distal end of the CFS region.

Results

HPV18 insertion mapping by RS-PCR and human genome sequence analysis

The RS-PCR technique (Sarkar *et al.*, 1993; Thorland *et al.*, 2000) was used to screen for HPV18 integration in a total of three HPV18-positive cervical carcinoma cell lines and 32 HPV18-positive primary cervical carcinomas. A total of 41 HPV18-human junctions were

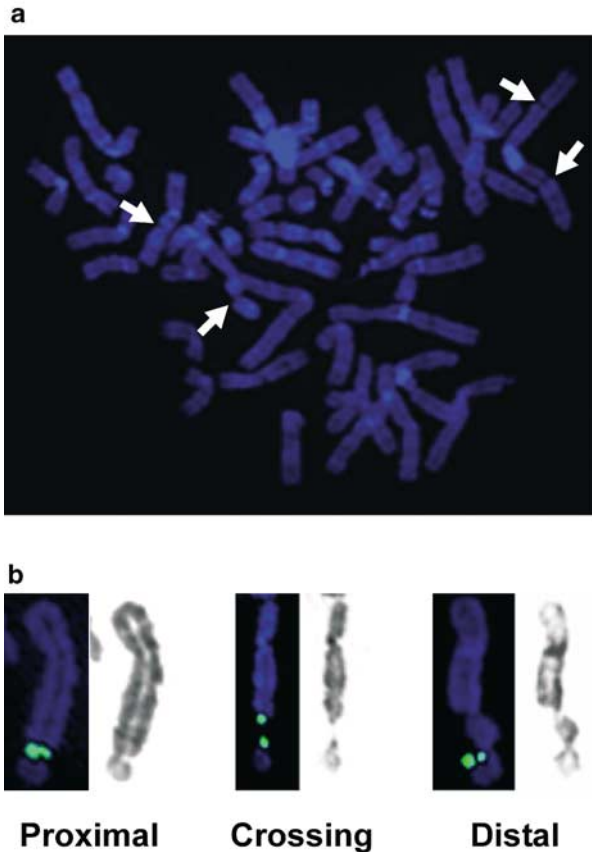


Figure 1 Schematic representation of the loci disrupted by HPV18 in CC cell lines and primary tumors. The HPV18 unit genome (X05015) is drawn to scale, and has been linearized in the E5 region. Arrows above the rectangles indicate HPV18 sequence (except HK15, where the second set of arrows above the rectangle indicate the portion of 723P16 that had been inverted). Arrows below the rectangles indicate host sequences (drawn to a scale different from the HPV18 genome) that correspond to the indicated BAC. BAC clone numbering of flanking sequences is based on the NCBI database, current as of January 2003. In cases where only a 5' junction is indicated, the 3' portion of the HPV18 genome contained in the tumor is unknown, and while the E2–E5 regions are shown, a part of these regions may be deleted in the tumor. The same is true for the L1 and L2 regions in cases where only a 3' junction fragment was identified. Diagonal lines represent human repetitive sequence as determined by the CENSOR database. Flanking regions that are completely repetitive could only be determined by PCR using a new set of unique primers to regions that lie outside the repeat element. Boldface print indicates tumors that contain at least one integration near the FRA8C\8q24.1\c-myc locus. Multiple integrations were found in tumors MC31, MC123, MC315, MC398, MC415, LU2, and LU8. Several tumors contain large deletions, and/or complex rearrangements. *Tumor MC415, and was isolated from a peripheral lymph node. •tumor MC360, where BAC 268N14 was replaced by BAC 620B22 for cytogenetic positioning of the integration relative to the CFS. These two BACs were found to overlap at the site of viral integration. The accession numbers and chromosomal locations for each of the host sequences illustrated are as follows: 414A9, AC027531, 8q24.1; 221n13, AC090023, 12q14.3; 216N221, AC023194, 8q21.2; hTERT, AY007685, 5p15; 255B23, AC020688, 8q24.1; 80K22, AF315312, 8q24.2; 488C13, AF111169, 14q24.3; 212E3, AF311103, 8p21.1; 27C2, AC099536, 3p14.2; 894D12, AL078605, 6q27; 2074B5, AC010386, 5p13.2; 3083H11, AC079453, 7q31.2; 268N14, AL391824, 1q31.1; 468O2, AC018992, 8q24.1; 55J15, AC103705, 8q24.2; CosF13825, AC005794, 19q13; 192G3, AC022516, 19q13; 170E16, AC026636, 15q23; 453E3, AC084393, 1q32.2; 382A18, AC018714, 8q24.1; 164E7, AC108036, 2q22.1; 55O6, AC022098, 19p13; 10K22, AC016667, 9q22; 723p16, AL132986, 14q24.1; 33A7, AC007833, 5q15

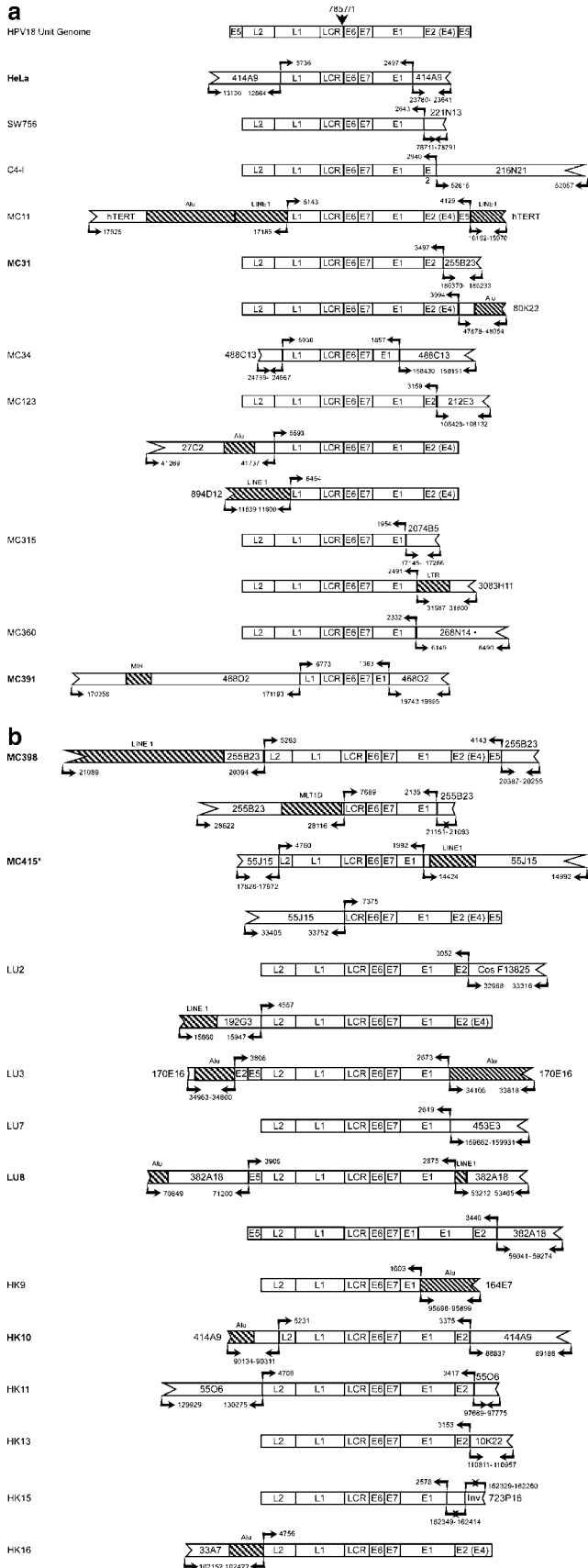
identified from 22 specimens, representing 30 integration events (Figure 1). Pathology data were available for 18 of the 22 specimens. MC11, MC123, LU3, LU7, HK10, HK13, and HK16 were all squamous cell carcinomas. MC31, MC34, MC360, MC391, MC398, MC415, LU2, LU8, HK9, HK11, and HK15 were all adenocarcinomas. Data for the remaining four samples were not available. Of the 30 events, both 5' and 3' junctions were identified in 11 cases, 5'-only junctions were found in five cases, and 3'-only junctions in 14 cases. Each integration event was confirmed by constructing new human loci-specific primers that were used for PCR in conjunction with a battery of HPV18 primers (DNS, Ferber *et al.*, 2003).

As was previously described for HPV16 (Thorland *et al.*, 2003, 2000), many of the HPV18 integration sites were found to occur in human repetitive sequences. Many of the integrations were also associated with deletions and/or complex rearrangements (Figure 1). Of the 41 junction fragments, 18 contain at least a portion of a repetitive element in the human-flanking sequence (nine Alu, seven line 1, and one each of LTR, MIR, and MLT1D). In the 11 cervical tumors and cell lines where both ends of the integration were isolated, we were able to determine the presence of HPV18 and human genomic deletions and/or rearrangements (Figure 1). In every case, the HPV18 genome had lost at least some coding sequences. Human deletions ranged from seven nucleotides (MC398) to over 130 kb (MC34 and MC391). Complex rearrangements were identified in seven of the 11 specimens.

Seven out of 22 cervical carcinoma specimens analysed contained more than a single 5' and/or a single 3' integration site (Figure 1). In these multi-integration specimens, five contain multiple hits near the same loci (MC31, MC398, MC415, LU2, and LU8), and may represent complex HPV18 insertion-duplication events. The remaining two specimens (MC123 and MC315) represent cases where single ends were found in different chromosomes, yet did not represent translocations involving the specific chromosomes, as suggested by long-range PCR (DNS).

HPV18 preferentially integrates at loci containing CFS

A total of 30 HPV18 integrations were identified. Several tumors contained three or more junction points within the same BAC, making the total number of unique integrations difficult to ascertain. For example, four junction loci were found in tumor MC398. Two of these were 5' junctions, two were 3' junctions, and all occurred within the region spanned by BAC 255B23 (Figure 1). We were unable to determine if cases like this represent two individual integration events or a duplication of a portion of the HPV18 genome at the time of integration. Thus, for the purpose of determining the frequency at which HPV18 integrates into CFS, this pattern of integration can only be counted as a single event, leaving the total number of informative integrations at 27. Using the FISH method outlined earlier, we were able to show that 17/27 (63%, $P < 0.001$) of the



HPV18 integrations rescued occurred in CFSs (Table 1), similar to that recently reported for HPV16 (Thorland *et al.*, 2003).

In addition, this work has positioned eight additional CFS with nucleotide resolution for the first time. These include FRA1K (1q31.1, BAC 453E3), FRA2F (2q22.1, BAC 164E7), FRA5D (5q15, BAC 33A7), FRA8C (8q24.1, BACs 468O2, 255B23, and 414A9), FRA9D (9q22.1, BAC 10K22), FRA14C (14q24.1, BAC 723P16), FRA15A (15q23, BAC 170E16), and FRA19A (19q13, COS F13825). This brings the total number of precisely localized CFSs to 17.

HPV18 integrations cluster around 8q24.1/c-myc/FRA8C

Previous studies have shown that several HPV integrations have occurred in the *c-myc* region (Durst *et al.*, 1987; Popescu *et al.*, 1990; Couturier *et al.*, 1991). In this study, we have identified 7/22 (32%), including the previously described HeLa cell line, specimens with HPV18 DNA integrated at this locus (Figure 1, all samples marked in boldface type, and Figure 3). Our greatly expanded data set, with nucleotide resolution shows that this 8q24.1 hot spot encompasses a 4Mb region, including a portion of the FRA8C CFS 5' of *c-myc* and extending to regions just 3' of *c-myc* (Figure 3). The HPV18 integrations in cervical tumors MC391, MC398, MC31, HeLa, HK10, and LU8 were all found to occur in FRA8C (top of Figure 3, box 1, and Table 1) represented by BACs 468O2, 255B23, 414A9, and 382A18, thus spanning a distance just over 4Mb (based on the UCSC genome browser November 2002 freeze). FISH data for MC391 and BAC 468O2 (Table 1) suggest that this integration is actually much closer to the other integrations in the cluster on FRA8C. In support of this theory, data from the human genome project indicate that there are several gaps that remain in this region and depending on how they are resolved, portions of this contig may be rearranged, bringing the distance across this region down to nearly 2 Mb.

The integrations in boxes 1 and 2 of Figure 3 define a much smaller region encompassing just less than 1 Mb. Box 1 itself represents a region spanning 250 kb, yet holds 5/8 (63%) of all *c-myc* region HPV18 integrations. Within that group, samples MC31, HeLa, HK10, and LU8 define a subcluster of approximately 88 kb in size, 530 kb 5' to the *c-myc* proto-oncogene. An integration cluster of this density and size has never been reported for any HPV in cervical carcinomas.

Box 2 illustrates the two primary cervical tumors, MC31 and MC415, with HPV18 integrations found

Figure 2 Representative metaphase spreads and chromosomes used to position sites of HPV18 integration relative to CFSs. (a) Typical metaphase morphology observed in this study. White arrows indicate clearly visible breaks. (b) Demonstration of the three unique hybridization patterns possible while positioning a CFS relative to a single BAC clone. Represented data is for BAC clone 468O2

Table 1 FISH data for BAC clones representing host-flanking sequence at the site of HPV18 integration

Tumor	Junction locus	BAC	Nearest CFS	Proximal	Crossing	Distal	On FS ^a
HeLa	8q24	414A9	8q24.1 (FRA8C)	1	0	19	Y
SW756	12q14.3	221N13	12q13.1 (FRA12A)	14	1	5	Y
C41	8q21.2	216N21	8q22.1 (FRA8B)	20	0	0	N
MC11	5p15	2191F10	No CFS	0	0	20	N
MC31	8q24	414A9	8q24.1 (FRA8C)	1	0	19	Y
	8q24	80k22	8q24.1 (FRA8C)	0	0	20	N
MC34	14q24.3	488C13	14q24.1 (FRA14C)	4	0	16	Y
MC123	3p14	27C2	3p14.2 (FRA3B)	N/D	N/D	N/D	Y
	6q26–27	894D12	6q26 (FRA6E)	N/D	N/D	N/D	N
	8p21.1	212E3	No CFS	N/A	N/A	N/A	
MC315	5p13.2	2074B5	5p14 (FRA5E)	0	0	20	N
	7q31	3083H11	7q31.2 (FRA7G)	18	0	2	Y
MC360	1q31.1	620B22	1q31 (FRA1K)	1	1	18	Y
MC391	8q24	468O2	8q24.1 (FRA8C)	3	1	16	Y
MC398	8q24	255B23	8q24.1 (FRA8C)	3	3	15	Y
MC415	8q24.2	55J15	FRA8C/FRA8D	N/D	N/D	N/D	N
LU2	19q13	192G3	19q13 (FRA19A)	0	0	4	N ^b
	19q13	F13825	19q13 (FRA19A)	0	2	4	Y ^b
LU3	15q23	107E16	15q22 (FRA15A)	3	0	17	Y
LU7	1q32.2	453E3	1q31 (FRA1K)	0	0	20	N
LU8	8q24.1	382A18	8q24.1 (FRA8C)	N/D	N/D	N/D	Y
HK9	2q22.1	164E7	2q21.3 (FRA2F)	0	1	19	Y
HK10	8q24	414A9	8q24.1 (FRA8C)	1	0	19	Y
HK11	19p13	55O6	No CFS	N/A	N/A	N/A	
HK13	9q22	10K22	9q22.1 (FRA9D)	0	1	19	Y
HK15	14q24.1	723P16	14q24.1 (FRA14C)	6	0	16	Y
HK16	5q15	33A7	5q15 (FRA5D)	10	2	10	Y

^aOn the FS is determined by identifying 20 visible breaks at the desired FS and counting the number of times the BAC hybridizes proximal, crossing, and/or distal.

^bFRA19A is an extremely low expressing 5-azacytidine-sensitive FS, and therefore does not contain a total of 20 counts.

N/A – not applicable.

N/D – not done due to existing map information.

Bold characters signifies integrations found in 8q24.1.

19 and 155 kb 3' to *c-myc*, respectively. FISH data corresponding to clones in this region show that these integrations do not take place in FRA8C or FRA8D (BAC 80K22 in Table 1 represents BACs 2056O22 and 2267H22 of Figure 3, box 2). The distal end of FRA8C is defined by BAC 382A18 (which overlaps to a large extent with BAC 414A9, but has been illustrated as a minimal overlapping contig in Figure 3) and is located 400 kb 5' to *c-myc*, while the proximal end of FRA8D begins roughly 1 Mb 3' from *c-myc* (DNS). This shows that the proto-oncogene *c-myc* is flanked on both sides by the highly unstable FRA8C and FRA8D CFSs.

HPV18 integration leads to genomic alteration of cellular genes

As was previously reported for HPV16 (Thorland *et al.*, 2003), many of the HPV18 integrations mapped in this study interrupted or deleted human genes. The interrupted genes are listed in Table 2. Tumor MC11 had HPV18 integrated in the 5' promoter region of hTERT, the catalytic subunit of human telomerase (Meyerson *et al.*, 1997). A full report on this integration can be found in Ferber *et al.* (2003).

The entire coding sequence of RING1B, a vital mRNA capping protein (Pillutla *et al.*, 1998), was deleted in tumor MC34. Tumor MC123 had three separate integrations. The first was in intron 4 of the fragile histidine triad (FHIT) tumor suppressor gene (Boldog *et al.*, 1993; Le Beau *et al.*, 1998). The second occurred in exon 7 of the KIAA1838 gene. The third integration was in exon 8 of the EPHX2 gene, which functions to metabolize xenobiotic epoxides (Fretland and Omiecinski, 2000). Tumor LU2 had HPV18 integrated in the 5' promoter region of CEACAM5, a transmembrane cell adhesion molecule (Eidelman *et al.*, 1993). Tumor LU7 was interrupted in intron 1 of CD34, the protein expressed on hematopoietic progenitor cells and involved in cell–cell adhesion (Simmons *et al.*, 1992). The candidate tumor suppressor LRP1B was disrupted at intron 41 in tumor HK9 (Liu *et al.*, 2000). Tumor HK11 contained HPV18 DNA in PRKACA, the catalytic subunit of cAMP-dependent protein kinase (Gonzalez and Montminy, 1989). Finally, in HK13, Fanconi's anemia complementation group (FANCC) was altered in intron 7. FANCC is one of the many genes involved in double-strand DNA breaks found in Fanconi's anemia patients (Ahmad *et al.*, 2002).

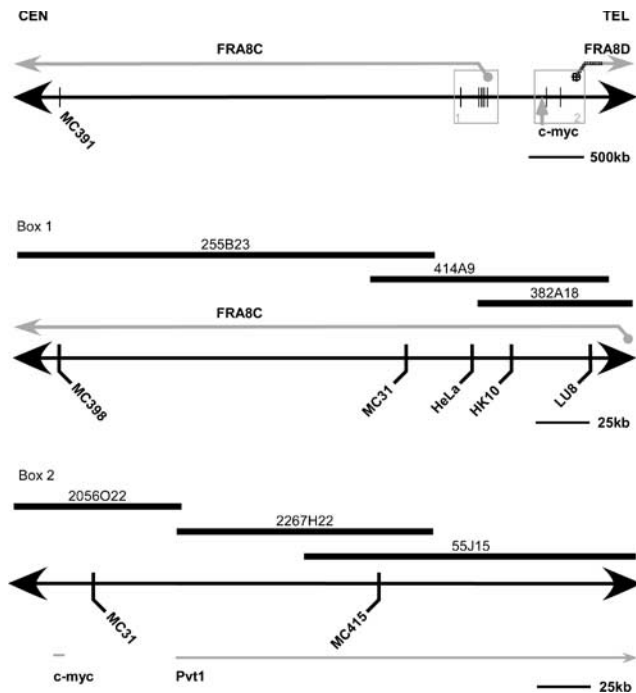


Figure 3 Detailed map of HPV18 integrations surrounding the FRA8C/*c-myc*/8q24.1 region. Each map has been drawn to the scale indicated. The top map shows the 4 mb region over which 8/27 HPV18 viral integrations were found in 7/22 cervical tumors. Horizontal black arrows indicate human genomic sequence. Vertical bars indicate the location of HPV18 integration. Gray or stippled horizontal bars indicate the relative locations of the CFSs FRA8C and FRA8D, respectively. Those in gray indicate defined regions, stippled indicates proposed regions. The vertical gray arrow indicates the position of the *c-myc* proto-oncogene. BAC 80K22 in Table 1 represents BACs 2056O22 and 2267H22 of box 2. CEN = centromeric. TEL = telomeric. Box 1, the enlarged region from above, showing the overlapping BAC contig and high density of HPV18 integration in cervical tumors MC31, HeLa, HK10, LU8, and the slightly further out MC398 in FRA8C. Box 2, the enlarged region surrounding the *c-myc* proto-oncogene illustrating the BAC contig around this locus. The *C-myc* and *Pvt-1* transcripts are illustrated by a gray bar or arrow, respectively. The box 1 cluster of HPV integrations is approximately 550 kb 5' to *c-myc*, and contains no known genes

The loci of HPV16 vs HPV18 integration in cervical carcinoma

Our lab had previously isolated 23 HPV16 integrations from cervical carcinoma specimens (Thorland *et al.*, 2003, 2000), and now report the isolation of 27 HPV18 integrations. When comparing the profile of integrations of the two HPV types, we see little overlap (Figure 4). Only the loci at 3p14, 5p15, and 8q21 have had both an HPV16 and an HPV18 viral integration in their respective cervical tumors. The 3p14 locus contains the FHIT gene and FRA3B (Zimonjic *et al.*, 1997); 5p15 contains the hTERT gene (Meyerson *et al.*, 1997). The 8q21 band does not contain a CFS or a gene that would suggest a reason for integration selection at this locus. HPV16 shows two clusters of integration, each consisting of three separate events. One cluster is located at 13q22, the other at

17q23. The cluster at 13q22 occurs over a tight, 200 kb region and contains several genes that may be important to the development of cervical carcinoma (Thorland *et al.*, 2003). RS-PCR of HPV18-positive tumors revealed a single large cluster at 8q24. Interestingly, RS-PCR has yet to identify a single HPV16 integration at 8q24. However, other groups using various techniques have identified HPV16 integration near 8q24 (Couturier *et al.*, 1991; Brink *et al.*, 2002; Wentzensen *et al.*, 2002).

Discussion

The development of cervical carcinoma is almost always associated with the presence of high-risk HPV infection (Walboomers *et al.*, 1999). The acquisition of a malignant phenotype is temporally associated with the integration of HPV into the human genome (Cullen *et al.*, 1991). Viral integration generally involves the interruption or deletion of the viral E2 repressor protein, which normally functions to regulate the expression of the E6 and E7 oncoproteins (Howley, 1996; Shah and Howley, 1996; Munger, 2002). This regulation has recently been suggested to also involve the remodeling of the E6/E7 promoter (Bechtold *et al.*, 2003). E6 and E7 interact with the tumor suppressor proteins p53 and pRB, respectively. In this manner, viral integration into the human genome provides a mechanism by which the cell accumulates genomic damage (zur Hausen, 1991). However, transient overexpression of E6 and E7 alone is not sufficient for transformation, suggesting that additional mutations are required (Hawley-Nelson *et al.*, 1989; Woodworth *et al.*, 1989). Given enough time, these mutations may occur without the need for viral integration. An alternative hypothesis is that the integration event itself may contribute to carcinogenesis. Indeed, several groups have reported the occurrence of large deletions or rearrangements at the site of HPV integration (Wilke *et al.*, 1996; Gallego *et al.*, 1997; Thorland *et al.*, 2003). In addition, HPV integration has been shown to occur near cellular proto-oncogenes (Durst *et al.*, 1987; Lazo *et al.*, 1989; Couturier *et al.*, 1991). Together these data suggest that HPV integration could directly influence gene expression by changing the normal human DNA composition. Therefore, we propose that in some cervical cancers, in addition to the resulting E6 and E7 expression, the integration event itself is a nonrandom, irreversible mutation that may contribute to the malignant phenotype. It is also possible that the CFS nature of the integration site is responsible for the acquisition of additional chromosomal aberrations. Indeed, chromosomal translocations are frequently found near HPV integration sites in cervical cancer (Popescu *et al.*, 1990; Koopman *et al.*, 1999; Macville *et al.*, 1999; Brink *et al.*, 2002).

Recent work from our lab has shown that for HPV16, the most common high-risk HPV subtype found in

Table 2 Host genes interrupted by HPV18 integration

Tumor	Integration locus	Gene	Position in gene	Function
MC11	5p15	HTERT	5' promotor region	Catalytic subunit of telomerase
MC34	14q24.3	RNGTT	Gene deletion	5'-terminal capping of mRNAs
MC123	3p14 6q26-27 8p21.1	FHIT KIAA1838 EPHX2	Intron 4 Exon 7 Exon 8 KO	Ap4A hydrolase; tumor suppressive activity Unknown Epoxide hydrolase; metabolize xenobiotic molecules
MC315	5p13.2	SLC1A3	Exons 4-7 KO	Human glutamate transporter
LU2	19q13	CEACAM5	5' promotor region	Transmembrane cell adhesion
LU7	1q32.2	CD34	Intron 1	Cell-cell adhesion; inhibition of hematopoietic differentiation?
HK9	2q22.1	LRP1B	Intron 41	LRP family; candidate tumor suppressor Lipoprotein catabolism Cell adhesion Embryonic development
HK11	19p13	PRKACA	Exon 10	Catalytic subunit C of cAMP-dependent protein kinase
HK13	9q22	FANCC	Intron 7	DNA double-strand breaks in Fanconi's anemia patients

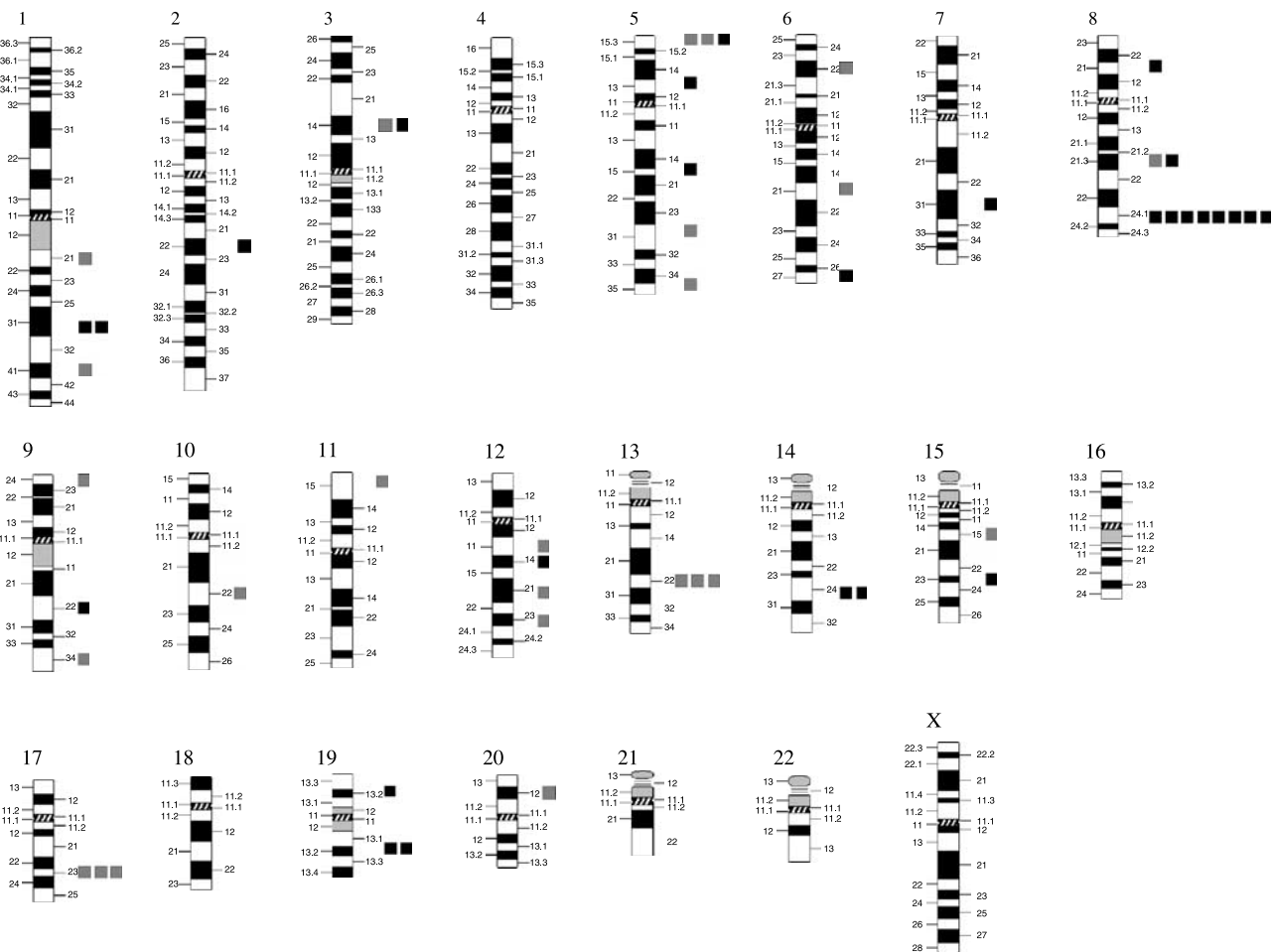


Figure 4 Comparison of HPV18V.S. HPV16 viral integrations as localized by RS-PCR. ■, HPV18 and ■, HPV16. Note the selection for the 8q24.1/*c-myc* /FRA8C loci in HPV18-positive cervical tumors

cervical tumors, integration occurs throughout the human genome. However, 48% ($P < 0.001$) of the integrations occurred in the 2% of the genome known as CFSs (Thorland *et al.*, 2003, 2000). In addition, several integrations were shown to disrupt cellular genes that may play a role in tumor development. HPV18, the second most common high-risk HPV type found in cervical tumors, is a distant relative of HPV16, and is more closely related to several low-risk HPVs than it is to HPV16 (Shah and Howley, 1996). However, HPV18 has been associated with more aggressive adenocarcinomas of the endocervix (Schwartz *et al.*, 2001). In the current study, we therefore sought to determine the relationship between HPV18 integration, CFSs, and cancer-related genes.

Using the established RS-PCR technique, we were able to identify 27 unique HPV18 viral integrations in 22 cervical carcinoma specimens. Positioning of these integrations relative to the CFSs revealed that 17/27 (63%, $P < 0.001$) occurred in CFS regions, demonstrating that these unstable genomic regions are also hot spots for HPV18 integrations. Given that non-CFS elements such as the *c-myc* proto-oncogene may be responsible for the large number of 8q24.1 integrations and thus skewing our data, we repeated the analysis without these data and found that 11/19 (58%, $P < 0.001$) integrations occur in CFSs. These data support the general statement that high-risk HPVs preferentially integrate at CFSs in cervical tumors. Interestingly, while both HPVs preferentially integrated at CFSs, the sites of HPV18 integration were for the most part distinct from those found in HPV16-positive cervical tumors (Figure 4). In addition, we identified a 4 Mb region surrounding *c-myc* and the FRA8C CFS that contained HPV18 viral integration in 7/22 (32%) of the cervical tumors analysed. While several groups have reported integrated HPV DNA in this region, the current study is of a much larger scale and provides integration data with nucleotide resolution.

The *c-myc* proto-oncogene has been shown to be deregulated in many tumor types including breast, colon, small-cell lung carcinomas, osteosarcomas, glioblastomas, melanoma, myeloid leukemias, and cervical tumors (reviewed in Pelengaris *et al.*, 2002). In Burkitt's lymphoma, translocations between the region surrounding *c-myc* and the immunoglobulin heavy chain, κ light, or λ light chains result in overexpression of c-Myc (Boxer and Dang, 2001; Ratsch *et al.*, 2002). These translocations occur over a region both 5' and 3' to *c-myc*, encompass 500 kb, and result in the placement of enhancer regions, from the Ig loci, more than several 100 kb from *c-myc*, yet result in upregulated c-Myc expression (Boxer and Dang, 2001; Ratsch *et al.*, 2002). In a similar fashion, HPV18 contains viral enhancers that may alter gene expression at this locus. While the cluster of integrations described in the current study represent a much broader region of over 4 Mb (Figure 3), the integrations found in tumors MC415 and the downstream end of MC31 (Figure 3, box 2) occur over a distance in agreement with described BL transloca-

tions. The integrations represented in box 1 of Figure 3, however, are over 500 kb 5' of *c-myc*, 300 kb farther 5' than any reported BL translocation. The tightest portion of this subcluster represents ~88 kb (between MC31 and LU8) and contains no known genes. While it is possible that these integrations affect the expression of *c-myc*, further studies are underway to determine if the presence of HPV DNA at this subcluster is sufficient to alter the expression of additional genes surrounding this locus.

During the course of this study, we found 10 examples where HPV18 integration disrupted known human genes (Table 2). The function of the KIAA1838 gene is unknown. Three others (EPHX2, SLC1A3, and RRGTT) perform functions that are difficult to tie to tumor development. The remaining six genes each have potential roles in the development or progression of cervical tumorigenesis.

Loss of tumor suppressor function is a hallmark of many cancers. In tumor MC123, HPV18 disrupted intron 4 of the tumor suppressor FHIT gene. The FHIT gene is located within FRA3B, the highest expressed human CFS, and has been shown to harbor HPV16 integrations (Glover and Stein, 1988; Wilke *et al.*, 1996). In addition, loss of FHIT expression has been found in a number of premalignant cervical lesions (Greenspan *et al.*, 1997; Huebner *et al.*, 1998). Together, this suggests that HPV18 integration at this site may promote cervical carcinogenesis.

Another hallmark of carcinogenesis is the acquisition of unrestrained replicative capacity. In many tumors, this is via the activation of telomerase (Kim *et al.*, 1994). hTERT expression is carefully regulated as it is the catalytic subunit of telomerase (Meyerson *et al.*, 1997). The placement of viral DNA containing powerful enhancers in the hTERT promoter region could result in gene activation, thus allowing the cell to divide indefinitely without the catastrophic effect of eroding telomeres. In this study, we identified a single HPV18 integration in tumor MC11, occurring 7 kb upstream from the start site of hTERT transcription. In addition, we recently described two additional HPV16 and two hepatitis B virus (HPV) integrations into the hTERT gene in cervical and hepatocellular carcinomas, respectively. Each of these samples overexpresses hTERT (Ferber *et al.*, 2003). Additional work is underway to determine if the overexpression is due to the presence of integrated viral DNA.

In tumor HK11, HPV18 DNA integrated into exon 10 of the cAMP-dependent protein kinase, catalytic subunit A (PRKACA) gene. PRKACA is one of the two catalytic subunits in the cAMP-dependent protein kinase (PKA) complex (Mayr and Montminy, 2001). Active PKA phosphorylates the transcription factor cyclic AMP response element-binding protein (CREB) (Gonzalez and Montminy, 1989). This phosphorylation leads to the activation of many genes, including BRCA1, retinoblastoma, and Bcl-2. Given the role of these genes in tumor suppression, disrupting PRKACA and thus inhibiting CREB function may have tumor-promoting effects within the cell.

Loss of genomic stability is another hallmark of cancer. In cervical cancer this loss is attributed to the E6 and E7 oncoproteins that inactivate p53 and pRB, respectively. In tumor HK13, HPV18 was integrated into the FANCC gene. Mutations in FANCC result in nonspecific double-stranded DNA breaks and are clinically manifested as Fanconi's anemia (Ahmad *et al.*, 2002). In cervical epithelial cells, integration into this gene could augment the effects of E6 and E7, thus creating an environment ripe for DNA damage and subsequent cellular transformation. Integration into FANCC is not restricted to HPV18. Wentzensen *et al.* (2002) previously detected a fusion transcript of HPV16 and part of FANCC.

Several additional genes that were disrupted by HPV18 integrations included three different cell-cell adhesion molecules. The first integration to disrupt a cell-cell adhesion molecule occurred in CEACAM5, a molecule that inhibits cellular differentiation (Eidelman *et al.*, 1993). Here, HPV18 DNA integrated into the 5' promoter region. An in-frame fusion transcript of HPV16 and CEACAM5 was described previously (Wentzensen *et al.*, 2002). The second cell-cell adhesion molecule, CD34 which is normally found on hematopoietic stem cells (Simmons *et al.*, 1992), was disrupted in intron 1 by HPV18 DNA. The third cell-cell adhesion molecule was LRP1B, a putative tumor suppressor that is deleted in non-small-cell lung cancer cell lines (Liu *et al.*, 2000). In tumor HK9, HPV18 was found to interrupt intron 41 of this gene. These three integrations

represent mutations that could promote tumor development by altering the way the epithelial cell interacts with neighboring cells or the extracellular matrix.

These data and that from Thorland *et al.* (2003, 2000), strongly suggest that HPV integration in cervical tumor cells is not random, and that it does target the CFSs. Furthermore, our data shows a large number of interesting genes at the sites of HPV18 integration, suggesting that in many tumors the integration event itself may trigger additional expression changes and thus provide growth advantage to the cell. Our work suggests that CFSs and important oncogenesis-related genes are frequent targets of HPV integration. Identifying the role that HPV insertion plays in gene dysregulation and subsequent cellular transformation will help to determine how viral integration contributes to malignant transformation of cervical epithelial cells. It is imperative that we understand the role of viral integration in the pathophysiology of cervical cancer, if we hope to formulate rational plans for molecular diagnostic and therapeutic interventions.

Acknowledgements

We thank Gert Jan Fleuren and Ed Schuurung for their contributions to this paper. This study was conducted as a part of MJF's doctoral thesis requirements in the Mayo Graduate School. DIS was supported by NCI Grant CA 48031. AATPB was financially supported by the Vanderes Foundation, Leiden, The Netherlands.

References

- Ahmad SI, Hanaoka F and Kirk SH. (2002). *BioEssays*, **24**, 439–448.
- Bechtold V, Beard P and Raj K. (2003). *J. Virol.*, **77**, 2021–2028.
- Boldog FL, Gemmill RM, Wilke CM, Glover TW, Nilsson AS, Chandrasekharappa SC, Brown RS, Li FP and Drabkin HA. (1993). *Proc. Natl. Acad. Sci. USA*, **90**, 8509–8513.
- Bosch FX, Lorincz A, Munoz N, Meijer CJ and Shah KV. (2002). *J. Clin. Pathol.*, **55**, 244–265.
- Boxer LM and Dang CV. (2001). *Oncogene*, **20**, 5595–5610.
- Boyer SN, Wazer DE and Band V. (1996). *Cancer Res.*, **56**, 4620–4624.
- Brink AA, Wiegant JC, Szuhai K, Tanke HJ, Kenter GG, Fleuren GJ, Schuurung E and Raap AK. (2002). *Cancer Genet. Cytogenet.*, **134**, 145–150.
- Casper AM, Nghiem P, Arlt MF and Glover TW. (2002). *Cell*, **111**, 779–789.
- Cha RS and Kleckner N. (2002). *Science*, **297**, 602–606.
- Cho KR. (1998). *The Genetic Basis of Human Cancer*. Vogelstein B and Kinzler K (eds). McGraw-Hill: New York, pp. 631–637.
- Couturier J, Sastre-Garau X, Schneider-Maunoury S, Labib A and Orth G. (1991). *J. Virol.*, **65**, 4534–4538.
- Crook T, Storey A, Almond N, Osborn K and Crawford L. (1988). *Proc. Natl. Acad. Sci. USA*, **85**, 8820–8824.
- Cullen AP, Reid R, Champion M and Lorincz AT. (1991). *J. Virol.*, **65**, 606–612.
- Durst M, Croce CM, Gissmann L, Schwarz E and Huebner K. (1987). *Proc. Natl. Acad. Sci. USA*, **84**, 1070–1074.
- Eidelman FJ, Fuks A, DeMarte L, Taheri M and Stanners CP. (1993). *J. Cell Biol.*, **123**, 467–475.
- Ferber MJ, Montoya DP, Yu C, Aderca I, McGee A, Thorland EC, Nagorney DM, Gostout BS, Burgart LJ, Boix L, Bruix J, McMahon BJ, Cheung TH, Chung TK, Wong YF, Smith DI and Roberts LR. (2003). *Oncogene*, **24**, 3813–3820.
- Fretland AJ and Omiecinski CJ. (2000). *Chem. Biol. Interact.*, **129**, 41–59.
- Gallego MI, Schoenmakers EF, Van de Ven WJ and Lazo PA. (1997). *Mol. Carcinogen.*, **19**, 114–121.
- Glover TW, Berger C, Coyle J and Echo B. (1984). *Hum. Genet.*, **67**, 136–142.
- Glover TW and Stein CK. (1987). *Am. J. Hum. Genet.*, **41**, 882–890.
- Glover TW and Stein CK. (1988). *Am. J. Hum. Genet.*, **43**, 265–273.
- Gonzalez GA and Montminy MR. (1989). *Cell*, **59**, 675–680.
- Greenspan DL, Connolly DC, Wu R, Lei RY, Vogelstein JT, Kim YT, Mok JE, Munoz N, Bosch FX, Shah K and Cho KR. (1997). *Cancer Res.*, **57**, 4692–4698.
- Ham J, Dostatni N, Gauthier JM and Yaniv M. (1991). *Trends Biochem. Sci.*, **16**, 440–444.
- Howley PM. (1996). *Fields Virology*, 3rd edn. Fields B, Howley PM, Griffin DE, Lamb RA, Martin MA, Roitman B, Straus SE and Knipe DM (eds). Lippincott-Raven: Philadelphia, pp. 2045–2076.
- Hawley-Nelson P, Vousden KH, Hubbert NL, Lowy DR and Schiller JT. (1989). *EMBO J.*, **8**, 3905–3910.
- Huebner K, Garrison PN, Barnes LD and Croce CM. (1998). *Annu. Rev. Genet.*, **32**, 7–31.

- Kim NW, Piatyszek MA, Prowse KR, Harley CB, West MD, Ho PL, Coviello GM, Wright WE, Weinrich SL and Shay JW. (1994). *Science*, **266**, 2011–2015.
- Koopman LA, Szuhai K, van Eendenburg JD, Bezrookove V, Kenter GG, Schuurin E, Tanke H and Fleuren GJ. (1999). *Cancer Res.*, **59**, 5615–5624.
- Kurman RJ, Schiffman MH, Lancaster WD, Reid R, Jenson AB, Temple GF and Lorincz AT. (1988). *Am. J. Obstet. Gynecol.*, **159**, 293–296.
- Lazo PA, DiPaolo JA and Popescu NC. (1989). *Cancer Res.*, **49**, 4305–4310.
- Le Beau MM, Drabkin H, Glover TW, Gemmill R, Rassool FV, McKeithan TW and Smith DI. (1998). *Genes Chromosomes Cancer*, **21**, 281–289.
- Liu CX, Musco S, Lisitsina NM, Yaklichkin SY and Lisitsyn NA. (2000). *Genomics*, **69**, 271–274.
- Macville M, Schrock E, Padilla-Nash H, Keck C, Ghadimi BM, Zimonjic D, Popescu N and Ried T. (1999). *Cancer Res.*, **59**, 141–150.
- Mayr B and Montminy M. (2001). *Nat. Rev. Mol. Cell. Biol.*, **2**, 599–609.
- McBride AA, Romanczuk H and Howley PM. (1991). *J. Biol. Chem.*, **266**, 18411–18414.
- Meyerson M, Counter CM, Eaton EN, Ellisen LW, Steiner P, Caddle SD, Ziaugra L, Beijersbergen RL, Davidoff MJ, Liu Q, Bacchetti S, Haber DA and Weinberg RA. (1997). *Cell*, **90**, 785–795.
- Munger K. (2002). *Front. Biosci.*, **7**, d641–d649.
- Pelengaris S, Khan M and Evan G. (2002). *Nat. Rev. Cancer*, **2**, 764–776.
- Pillutla RC, Shimamoto A, Furuichi Y and Shatkin AJ. (1998). *Genomics*, **54**, 351–353.
- Popescu NC and DiPaolo JA. (1989). *Cancer Genet. Cytogenet.*, **42**, 157–171.
- Popescu NC, Zimonjic D and DiPaolo JA. (1990). *Hum. Genet.*, **84**, 383–386.
- Rassool FV, McKeithan TW, Neilly ME, van Melle E, Espinosa Rd and Le Beau MM. (1991). *Proc. Natl. Acad. Sci. USA*, **88**, 6657–6661.
- Ratsch A, Joos S, Kioschis P and Lichter P. (2002). *Exp. Cell Res.*, **273**, 12–20.
- Sarkar G, Turner RT and Bolander ME. (1993). *PCR Methods Appl.*, **2**, 318–322.
- Scheffner M, Werness BA, Huibregtse JM, Levine AJ and Howley PM. (1990). *Cell*, **63**, 1129–1136.
- Schwartz SM, Daling JR, Shera KA, Madeleine MM, McKnight B, Galloway DA, Porter PL and McDougall JK. (2001). *J. Clin. Oncol.*, **19**, 1906–1915.
- Shah KV and Howley PM. (1996). *Fields Virology*, 3rd edn. Fields BN, Howley PM, Griffin DE, Lamb RA, Martin MA, Roitman B, Straus SE and Knipe DM. (eds). Lippincott-Raven: Philadelphia, pp. 2077–2109.
- Simmons DL, Satterthwaite AB, Tenen DG and Seed B. (1992). *J. Immunol.*, **148**, 267–271.
- Thorland EC, Myers SL, Gostout BS and Smith DI. (2003). *Oncogene*, **22**, 1225–1237.
- Thorland EC, Myers SL, Persing DH, Sarkar G, McGovern RM, Gostout BS and Smith DI. (2000). *Cancer Res.*, **60**, 5916–5921.
- Walboomers JM, Jacobs MV, Manos MM, Bosch FX, Kummer JA, Shah KV, Snijders PJ, Peto J, Meijer CJ and Munoz N. (1999). *J. Pathol.*, **189**, 12–19.
- Wentzensen N, Ridder R, Klaes R, Vinokurova S, Schaefer U and Doeberitz MK. (2002). *Oncogene*, **21**, 419–426.
- Wilke CM, Hall BK, Hoge A, Paradee W, Smith DI and Glover TW. (1996). *Hum. Mol. Genet.*, **5**, 187–195.
- Woodworth CD, Doniger J and DiPaolo JA. (1989). *J. Virol.*, **63**, 159–164.
- Zimonjic DB, Druck T, Ohta M, Kastury K, Croce CM, Popescu NC and Huebner K. (1997). *Cancer Res.*, **57**, 1166–1170.
- zur Hausen H. (1991). *Virology*, **184**, 9–13.