

The Majority of Viral-Cellular Fusion Transcripts in Cervical Carcinomas Cotranscribe Cellular Sequences of Known or Predicted Genes

Irene Kraus,^{1,3,4} Corina Driesch,⁴ Svetlana Vinokurova,⁵ Eivind Hovig,² Achim Schneider,⁶ Magnus von Knebel Doeberitz,⁵ and Matthias Dürst⁴

¹Institute of Pathology, Rikshospitalet University Hospital; ²Department of Tumor Biology, Institute for Cancer Research, The Norwegian Radium Hospital, Oslo, Norway; ³NorChip AS, Klokkestrua, Norway; ⁴Gynäkologische Molekularbiologie, Frauenklinik der Friedrich-Schiller-Universität, Jena, Germany; ⁵Department of Applied Tumor Biology, Institute of Pathology, University of Heidelberg, Heidelberg, Germany; and ⁶Frauenklinik mit Hochschulambulanz der Charité, Campus Benjamin Franklin und Campus Mitte, Berlin, Germany

Abstract

Integration of human papillomavirus (HPV) DNA into the host genome is a frequent event in cervical carcinogenesis and is reported to occur at randomly selected chromosomal sites. However, as the databases are being up-dated continuously, the knowledge based on sequenced viral integration sites also expands. In this study, viral-cellular fusion transcripts of a preselected group of 74 cervical carcinoma or cervical intraepithelial neoplasia grade 3 (CIN3) biopsies harboring integrated HPV16, HPV18, HPV31, HPV33, or HPV45 DNA were amplified by 3'-rapid amplification of cDNA ends PCR and sequenced. Consistent with previous reports, integration sites were found to be distributed throughout the genome. However, 23% (17 of 74) of the integration sites were located within the cytogenetic bands 4q13.3, 8q24.21, 13q22.1, and 17q21, in clusters ranging from 86 to 900 kb. Of note is that clusters 8q24.21 and 13q22.1 are within 1.5 Mbp of an adjacent fragile site whereas clusters 4q13.3 and 17q21 are >15 Mbp distant to any known fragile sites. It is tempting to speculate that as yet unknown fragile sites may be identified on the basis of HPV integration hotspots. No correlation between HPV type and specific integration loci was found. Of 74 fusion transcripts, 28 contained cellular sequences, which were homologous to known genes, and 40 samples contained sequences of predicted genes. In 33 fusion transcripts, both viral and cellular sequences were in sense orientation, indicating that the gene itself or upstream sequences were affected by integration. These data suggest that the influence of HPV integration on host gene expression may not be a rare effect and should encourage more detailed analyses. [Cancer Res 2008;68(7):2514–22]

Introduction

Persistent infection with high-risk human papillomavirus (HPV) is the main risk factor for the development of high-grade precancerous lesions [cervical intraepithelial neoplasia grade 2 (CIN2) and CIN3] and cervical carcinoma (1, 2), with HPV16 being the most prevalent type, followed by HPV18, HPV31, HPV33, and HPV45 (3). The immortalizing and transforming potential of the

virus lies in its oncogenes E6 and E7, underlined by their constitutive expression in HPV-induced cervical carcinomas (4, 5). The E6 and E7 oncoproteins mediate mitogenic and antiapoptotic stimuli by interacting with numerous regulatory proteins of the host cell that control the cell cycle (6, 7). Moreover, the viral oncoproteins induce mitotic defects and genomic instability by uncoupling centrosome duplication from the cell division cycle (8, 9). During malignant progression, viral DNA is frequently integrated into the host genome (10, 11). A number of experimental studies show that integration results in altered expression of E6 and E7 thereby providing the cells with a selective growth advantage (12). Moreover, viral sequences expressed as fusion transcripts together with human sequences have been reported to have a longer half-life as compared with episome-derived transcripts, which is explained by the absence of instability elements normally present in the 3' untranslated region of the episome-derived transcripts (13). In addition, due to the frequent disruption of the E2 region during the integration process, the viral transcriptional control by E2 is lost and, as a consequence, the expression of E6 and E7 increases (14). However, it needs to be considered that in many precancers and cancers, episomal viral genomes are present together with the integrated viral DNA, suggesting that viral E2 is available in *trans* to regulate E6 and E7 expression (15). E6 and E7 expression levels may also be influenced by human *cis*-acting regulatory elements present in the vicinity of the viral integration site (16). Indeed, differences in the oncogene expression pattern have been observed by RNA-RNA *in situ* hybridization between low-grade and high-grade CIN but not between high-grade CIN and invasive cancer (17, 18). E6 and E7 transcripts were not detectable or were expressed at very low levels in the basal and parabasal layers of low-grade CIN. In contrast, high-grade CIN invariably expressed E6 and E7 mRNA in basal cells as is the case for cancer cells, but the signal intensity varied considerably irrespective of histologic grading. These differences in steady-state levels of oncogene transcripts were also evident in analyses using quantitative reverse transcription-PCR (RT-PCR) technology (19, 20). In a recent analysis, no correlation was found between the expression levels of viral oncogene transcripts and the physical state of the viral genome, which is in clear contrast to the observations made in cell culture models (21).

To date, more than 200 virus-host integration sites have been mapped (11, 22). All integration loci examined were unique, involving all chromosomes, and fragile sites, translocation break points, and transcriptionally active regions were found to be preferred sites for integration (23–25). In addition to the effects on viral oncogene expression, integration may also influence the

Note: I. Kraus and C. Driesch contributed equally to this work.

Requests for reprints: Matthias Dürst, Gynäkologische Molekularbiologie, Frauenklinik der Friedrich-Schiller-Universität, Bachstrasse 18, 07743 Jena, Germany. Phone: 49-3641-933720; Fax: 49-3641-934272; E-mail: matthias.duerst@med.uni-jena.de.
©2008 American Association for Cancer Research.
doi:10.1158/0008-5472.CAN-07-2776

expression of cellular genes. Reuter et al. (26) report on the functional inactivation of APM-1, a putative tumor suppressor gene, which had resulted from insertional mutagenesis in combination with the deletion of the second allele of the gene. Moreover, in a number of tumors, integration sites have been mapped near the *MYC* locus. Of note are the high levels of *MYC* expression in these tumors, suggesting that integration may influence transcription of neighboring genes (27–29). However, a systematic analysis of the disruption or deregulation of cellular genes as a consequence of viral integration has not been done.

The number of characterized integration sites is still too small to fully understand all aspects of HPV integration. The aim of this study was to analyze in further detail viral integration sites at the RNA level to underscore current knowledge and to provide new insights in this process. In addition to HPV16 and HPV18, we have also analyzed integration sites for HPV31, HPV33, and HPV45.

Materials and Methods

Clinical samples. In this study, biopsies from a preselected group of CIN3 and cervical carcinomas known to harbor integrated HPV DNA were analyzed. Sixty-five cervical carcinoma samples (International Federation of Gynecology and Obstetrics stages IA–IVB) were obtained from patients treated at the Department of Gynecology of Friedrich-Schiller-Universität, Jena, Germany, between 1995 and 2003 ($n = 29$) and at the Norwegian Radium Hospital, Oslo, Norway, between 1995 and 2003 ($n = 36$). Nine CIN3 samples were obtained from women admitted to Østfold County Hospital, Norway, for treatment in the period October 1999 to October 2001 ($n = 5$) and to the Norwegian Radium Hospital in October 2002 through October 2003 ($n = 4$). All biopsies were histologically verified. The samples were positive for HPV16 ($n = 15$), HPV18 ($n = 27$), HPV31 ($n = 2$), HPV33 ($n = 7$), or HPV45 ($n = 23$). For HPV typing, we used the GP5+/GP6+ enzyme immunoassay approach as described by Jacobs et al. (30) and nucleic acid sequence-based amplification as previously described (4, 31).

Nucleic acid isolations. Total RNA was isolated with the RNeasy Mini Kit following the protocol for total RNA extraction from animal tissues (Qiagen). Samples were homogenized by using the QIAshredder (Qiagen). An additional DNase treatment step was included for all samples. For some of the HPV18-, HPV45-, HPV31-, and HPV33-positive samples collected in Norway, nucleic acids were extracted as described earlier (4, 31) and DNA was removed by DNase treatment for 20 min at 70°C (Sigma-Aldrich). Total RNA was eluted in 40 µL of RNase-free water and stored at –70°C.

Reverse transcription. Total RNA (~500 ng) was reverse transcribed using 20 units of Superscript II reverse transcriptase (Invitrogen) and an oligo(dT)₁₇ primer coupled to a linker sequence (32), referred to as Frohman primer. The reaction was done for 1 h at 42°C in a final volume of 20 µL. To allow for proper primer annealing, the primer and template were incubated at 70°C for 10 min before the reverse transcription. To control for RNA integrity and first-strand cDNA quality, PCR using glyceraldehyde-3-phosphate dehydrogenase primers was done as previously described (33).

Amplification of integrate-derived fusion transcripts. HPV oncogene transcripts derived from integrated viral genomes (viral-cellular fusion transcripts) were amplified using the amplification of papillomavirus oncogene transcript (APOT) assay (34). APOT is based on a 3'-rapid amplification of cDNA ends PCR. HPV E7 primers are used as forward primers, an adapter primer complementary to the linker sequence in the Frohman primer as first reverse primer, and the Frohman primer as the second nested reverse primer. The reactions were done as previously described (34)⁷ with minor modifications for HPV16 and HPV18: the reaction mixture was subjected to an initial denaturation step for 3 min at 94°C, followed by 30 cycles of denaturation at 94°C for 0.5 min, primer annealing, and elongation at

72°C for 2 min. For HPV16, annealing temperatures of 61°C and 67°C for first and second PCR, respectively, were used; for HPV18, 61°C and 70°C. The reaction was ended by a final elongation step at 72°C for 6 min. Five microliters of the first PCR were used as template for the nested PCR step.

APOT amplification products were visualized by 1.2% agarose gel electrophoresis. Viral-cellular fusion transcripts (often referred to as off-size bands) deviate from the characteristic size of the major viral transcript (E6*1-E7-E1_vE4-E5) derived from episomal viral DNA. HPV-positive cell lines like HeLa and CaSki, which are known to harbor multiple copies of integrated viral DNA, served as controls in the APOT assay and invariably show off-size bands only.

PCR. For a number of randomly selected samples ($n = 15$), the existence of the fusion transcript was confirmed by semi-nested RT-PCR using viral and cellular integration site-specific primers. The PCR amplification was carried out in a total volume of 50 µL containing as final concentrations 1× PCR buffer [20 mmol/L Tris-HCl (pH 8.4), 50 mmol/L KCl], 1.5 mmol/L MgCl₂, 0.2 mmol/L deoxynucleotide triphosphates, 0.25 µmol/L (250 pmol) of each primer, and 1 unit of Taq DNA polymerase (Invitrogen). The first DNA denaturation step was done for 3 min at 94°C, followed by 30 cycles of denaturation for 45 s at 94°C, annealing for 45 s at 55°C, and extension for 1 min at 72°C, before a final extension for 6 min at 72°C. Five microliters of the first PCR were used as template for the nested PCR step. The temperature profile was as for the first PCR step except for the annealing temperature, which was 57°C.

Expression of predicted genes was explored by semi-nested RT-PCR using gene-specific primers only. The reaction conditions were the same as above.

Cloning and sequence analysis. Viral-cellular fusion transcripts were excised from the agarose gel and extracted using the QIAquick Gel Extraction Kit from Qiagen. The isolated amplicons were cloned into the TOPOPCR4 vector using the TA Cloning Kit (Invitrogen), then subjected to DNA sequence analysis; for some samples, the PCR products were sequenced directly. Samples were sequenced using LI-COR 4200S (MWG-Biotech) or the CEQ 2000 XL DNA Capillary Electrophoresis System from Beckman Coulter. The genomic localization of the human sequences, as well as chromosomal fragile sites located adjacent to the evident viral integration locus, were determined by database query using National Centre for Biotechnology Information (NCBI) Build 36 (35) and the University of California, Santa Cruz (UCSC) hg18 (March 2006) human genome assemblies (36).

Results

Viral-cellular fusion transcripts of a preselected group of biopsies comprising 65 cervical carcinomas and 9 CIN3 known to harbor integrated viral DNA were amplified, cloned, and sequenced. The cellular sequences of the fusion transcripts were first characterized by the NCBI human megaBlast database alignment tool. Additionally, all sequences were analyzed using the UCSC Blat database to identify expressed sequence tags and genes of probable identity.

Mapping viral-cellular fusion transcripts to chromosomal bands. The chromosomal reference of all viral-cellular fusion transcripts with respect to Giemsa-stained bands was taken from the UCSC database (37). All bands targeted by integration were aligned and are presented in Table 1. Viral-cellular fusion transcripts and thus the corresponding integration sites were found to be distributed throughout the genome, targeting all chromosomes except for chromosomes 21 and 22 (Fig. 1). Twenty-one integration loci, of which most are assigned to sub-bands, have thus far not been reported in the literature. These are marked with a suffix in Table 1. Moreover, several cytogenetic bands showed three or more HPV integration events. Five sites on 8q24.21 were located within a distance of ~1 to 860 kb upstream of the *MYC* gene. For

⁷ S. Vinokurova, personal communication.

Table 1. Summary of all HPV fusion transcripts analyzed

Sample	Pathology	HPV*	Integration locus	Fragile site involved [†]	Gene [†]			Orientation
					Accession code	Name	Spliced or fused to	
T183	SCC	16	1p36.11	FRA1A (1p36)	<i>C1orf201</i>	<i>chromosome 1 open reading frame 201</i>	n.a.	AS
T627	SCC	45	1p21.3 [§]	<i>FRA1E (1p21.2)</i> <i>FRA1M (1p21.3)</i>	<i>AK024564</i>		n.a.	AS
D3571	SCC	45	1p21.3 [§]	<i>FRA1E (1p21.2)</i> <i>FRA1M (1p21.3)</i>	—			
T1981	SCC	18	2p25.1 [§]	—	CYS1	cystin 1	Intron 2	S
T174	SCC	16	2p24.3	<i>FRA2C (2p24.2)</i>	<i>hmm88622</i>		n.a.	AS
T644	SCC	16	2p13.1 [§]	<i>FRA2E (2p13)</i>	<i>hmm230426</i>		Intron 4	S
D3918	SCC	18	2q22.1	<i>FRA2F (2q21.3)</i>	LRP1B	low density lipoprotein-related protein 1B	n.a.	AS
T654	SCC	45	2q22.1	—	LRP1B	low density lipoprotein-related protein 1B	n.a.	AS
F251	CIN3	45	2q24.1	—	GPD2	glycerol-3-phosphate dehydrogenase 2 (mitochondrial)	Exon 20	S
T4177	SCC	18	2q24.2	—	ITGB6	integrin, β ₆	n.a.	AS
D3931	SCC	31	2q32.2	—	<i>snerklar</i>		n.a.	AS
D3994	SCC	45	2q33.3	<i>FRA2I (2q33)</i>	KLF7	kruppel-like factor 7	n.a.	AS
D4034	SCC	18	2q34	<i>FRA2I (2q33)</i>	<i>NT_005403.1081</i>		Intron 1	S
D3622	SCC	45	3p26.2	—	<i>hmm12935</i>		n.a.	AS
D3829	SCC	45	3q28	<i>FRA3C (3q27)</i>	LEPREL1	leprecan-like 1	Exon 9	S
D4023	SCC	45	3q28	<i>FRA3C (3q27)</i>	—			
T2341	SCC	18	4p15.31	<i>FRA4D (4p15)</i>	KCNIP4	kv channel interacting protein 4	n.a.	AS
D3899	SCC	18	4q12 [§]	<i>FRA4B (4q12)</i>	<i>hmm35522</i>		Exon 4	S
D3421	SCC	18	4q13.3	—	<i>chr4.75.005a</i> <i>chr4_622.1</i>		n.a.	AS
D3458	SCC	45	4q13.3	—	<i>NT_006216.31</i>		n.a.	AS
D4049	SCC	45	4q13.3	—	—			
T1051	SCC	16	4q23 [§]	—	<i>chr4_782.1</i>		Intron 2	S
D3545	SCC	18	4q28.3 [§]	—	<i>NT_016354.412</i> SLC7A11	solute carrier family 7, member 11	Intron 1 Exon 8	S S
D3826	SCC	45	4q32.3 [§]	—	<i>NT_022792.8</i>		Intron 2	S
D4026	SCC	18	5q31.3	—	ARHGAP26	rho GTPase activating protein 26	n.a.	AS
D4060	SCC	18	5q33.1 [§]	—	SH3TC2	SH3 domain tetratricopeptide repeats 2 family member, possibly N-myristoylated	n.a.	AS
D3601	SCC	18	6p25	<i>FRA6B (6p25.1)</i>	RREB1	ras responsive element binding protein 1	Exon 4	S
V204	CIN3	18	6p22.1	<i>FRA6C (6p22.2)</i>	<i>chr6_368.1</i> <i>NT_007592.447</i>		Intron 3 Intron 4	S S
T3359	SCC	33	6q15 [§]	<i>FRA6G (6q15)</i>	<i>chr6.92.003.a</i>		n.a.	AS
V828/1	CIN3	45	7p13 [§]	<i>FRA7D (7p13)</i>	—			
D3835	SCC	18	7q31.1	<i>FRA7G (7q31.2)</i>	IMMP2L	IMP2 inner mitochondrial membrane peptidase-like	Exon 6	S
D4039	SCC	45	7q31.2	<i>FRA7G (7q31.2)</i>	CFTR	cystic fibrosis transmembrane conductance regulator, ATP-binding cassette (sub-family C, member 7)	n.a.	AS

(Continued on the following page)

Table 1. Summary of all HPV fusion transcripts analyzed (Cont'd)

Sample	Pathology	HPV*	Integration locus	Fragile site involved [†]	Gene [‡]			Orientation
					Accession code	Name	Spliced or fused to	
T2594	SCC	16	8q12.1 [§]	—	<i>C8orf72</i>	<i>chromosome 8 open reading frame 72</i>	Intron 4	S
D4045	SCC	18	8q21.2	—	<i>NT_008046.7</i>		Intron 2	S
D4056	SCC	45	8q21.2	—	<i>NT_008046.7</i>		Intron 2	S
D3385	SCC	33	8q22.3 [§]	—	<i>BX119528</i>		n.a.	AS
					<i>chr8_1098.1</i>		n.a.	AS
					<i>NT_008046.331</i>		n.a.	AS
					<i>g18734.t1</i>		n.a.	AS
T315	SCC	16	8q24.21	<i>in all cases flanked by FRA8C (8q24.1)</i>	<i>hmm75097</i>		n.a.	AS
T4694	SCC	16	8q24.21	<i>FRA8E (8q24.1)</i>	<i>blevyby</i>		n.a.	AS
F284	CIN3	18	8q24.21		<i>chr8_1126.1</i>		Intron 2	S
					<i>chr8_1235.1</i>		Intron 3	S
D3464	SCC	18	8q24.21		<i>sweeker</i>		Exon 3	S
D3631	SCC	45	8q24.21		<i>sweeker</i>		n.a.	AS
T3817	SCC	16	9p24.1	—	KIAA1432	connexin 43 interacting protein 150 kDa	n.a.	AS
F343	CIN3	45	9p11.2	—	<i>NT_078045.4</i>		n.a.	AS
			9p12	—	<i>HTR007521.9.538.0</i>	<i>contactin associated protein-like 3</i>	Intron 11	S
			9p12	—	<i>chr9_414.1</i>		n.a.	AS
			9p12	—	<i>chr9_406.1</i>		n.a.	AS
			9p12	—	<i>swerskabu</i>		Exon 2	S
			9p12	—	<i>paskobu</i>		Exon 2	S
			9p13.1	—	<i>neytoyby</i>		Exon 2	S
T209	SCC	33	9q21.13 [§]	—	<i>chr9.76.001a</i>		n.a.	AS
D3708	SCC	18	10q24.31	—	NPM3	nucleophosmin/nucleoplasmin 3	Exon 5	S
T954	SCC	33	11p15.4	FRA11C (11p15.1) FRA11I (11p15.1)	MRV11	murine retrovirus integration site 1 homolog	Exon 2	S
D3647	SCC	45	11q22.2 [§]	—	<i>hmm26110</i>		n.a.	AS
V144	SCC	18	12p11.2 [§]	—	SSPN	sarcospan (Kras oncogene-associated gene)	Exon 2	S
F380	CIN3	45	12q21.33	FRA12B (12q21.3)	—			
D3819	SCC	45	13q21.32	FRA13B (13q21)	PCDH9	protocadherin 9	n.a.	AS
T3283	SCC	16	13q22.1	<i>in all cases flanked by FRA13B (13q21)</i>	<i>NT_024524.1096</i>		n.a.	AS
T3706	SCC	16	13q22.1		<i>BG182794</i>		Intron 1	S
D3996	SCC	18	13q22.1		KLF5	Kruppel-like factor 5	Exon 4	S
D3772	SCC	18	13q22.1		<i>BG182794</i>		Intron 1	S
					<i>chr13_463.1</i>		n.a.	AS
F250	CIN3	45	13q22.1		<i>NT_024524.1096</i>		n.a.	AS
D3632	SCC	45	13q22.1		<i>BG182794</i>		Exon 2	S
					<i>chr13_463.1</i>		n.a.	AS
T4290	SCC	18	14q11.2 [§]	—	<i>C14orf94</i>	<i>chromosome 14 open reading frame 94</i>	n.a.	AS
D3484	SCC	45	14q24.1	FRA14C (14q24.1)	RAD51L1	RAD51-like 1	n.a.	AS
T221	SCC	18	15q21.1 [§]	FRA15A (15q22)	<i>sehe</i>		n.a.	AS
T1729	SCC	18	15q25.3 [§]	—	AGBL1	ATP/GTP binding protein-like 1 family member	Intron 22	S
D3962	SCC	33	16q22.1	FRA16B (16q22.1)	COG8	component of oligomeric Golgi complex 8	Exon 2	S

(Continued on the following page)

Table 1. Summary of all HPV fusion transcripts analyzed (Cont'd)

Sample	Pathology	HPV*	Integration locus	Fragile site involved [†]	Gene [‡]			Orientation
					Accession code	Name	Spliced or fused to	
T2147	SCC	16	17q21.2	—	EIF1	eukaryotic translation initiation factor 1	Exon 2	S
T1543	SCC	45	17q21.2	—	<i>chr17_706.1</i>		Intron 10	S
D3773	SCC	31	17q21.2	—	<i>KRT42P</i>	<i>keratin 42 pseudogene</i>	Exon 11	S
T2322	SCC	33	18q11.2 [§]	—	<i>C18orf17</i>	<i>chromosome 18 open reading frame 17</i>	n.a.	AS
T4653	SCC	16	19p13.11	FRA19B (19p13)	ZNF506	zinc finger protein 506	n.a.	AS
T4916	SCC	18	19q13.11	FRA19A (19q13)	<i>LOC126235</i>	<i>similar to 40S ribosomal protein S4, X isoform</i>	Exon 1	S
T4743	SCC	16	19q13.2	FRA19A (19q13)	CEACAM6	carcinoembryonic antigen-related cell adhesion molecule 6 (nonspecific cross-reacting antigen)	n.a.	AS
T4251	SCC	18	20p12.1	<i>FRA20B (20p12.2)</i>	FLRT3	fibronectin-like domain-containing leucine-rich transmembrane protein 3	Exon 2	S
T1013	SCC	16	20p11.21	FRA20A (20p11.23)	<i>hmm65227</i>		n.a.	AS
V815	CIN3	45	20q13.2	—	<i>NT_011362.413</i>		n.a.	AS
D3335	SCC	33	Xp22.2	<i>FRAXB (Xp22.31)</i>	MID1	midline 1	Intron 1	S
V863	CIN3	18	Xp22.31	<i>FRAXB (Xp22.31)</i>	—			
D3923	SCC	18	Xq25 [§]	—	<i>SMARCA1</i>	<i>SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 1</i>	n.a.	AS

NOTE: —, no entry of fragile sites ("Fragile site involved" column) genes ("Accession code" column) in public databases. n.a., not applicable because fusion transcript is in antisense orientation.

Abbreviations: SCC, squamous cervical carcinoma. S/AS, sense/antisense orientation of the cellular gene relative to the viral coding strand.

*HPV accession numbers used as reference for the nucleotide sequences of the respective viral genomes are NC_001526 (HPV16), NC_001357 (HPV18), NC_001527 (HPV31), NC_001528 (HPV33), and NC_001590 (HPV45).

[†] Fragile sites located at a distance of up to 5 Mbp adjacent to the integration locus are indicated in italics.

[‡] Known genes ($n = 28$) and predicted genes ($n = 40$) found in the fusion transcripts. Six sequences did not show any similarity other than human genomic sequence. Predicted genes are shown in italics (AceView Gene Models with Alt-Splicing, Human mRNAs from Genbank, Human ESTs That Have Been Spliced, Genscan Gene Predictions, N-SCAN Gene Predictions, Geneid Gene Predictions, Augustus Gene Predictions Using Hints, Swiss Institute of Bioinformatics Gene Predictions from mRNA and ESTs); hmm-sequences were displayed by NCBI Blast, taken from the Gnomon gene prediction program.

[§]Integrations not previously reported with regard to the respective sub-band containing one digit to the band number ($n = 21$).

^{||}Rare fragile site.

13q22.1, six sites were located within a region of 380 kb between the two genes *KLF5* and *KLF12*. For 4q13.3, three sites were spanning a region of 86 kb between the *IL8* gene and *CXCL6*; for 17q21, a region of 900 kb was targeted by three integration sites. No correlation was seen between HPV type and integration loci.

HPV integration within fragile sites. All integration loci were checked for the presence of fragile sites. By use of the NCBI fragile site map viewer, we found that 42 of the integrations were located in or close to a fragile site (Table 1): 21 within common or rare fragile sites and 21 up to 5 Mbp adjacent to a common or rare fragile site. Thirty-two samples were not associated with any fragile site.

Splicing within viral sequences. All viral-cellular fusion transcripts were spliced (Fig. 2). In all cases, the splice donor site was within the 5' end of the viral E1 open reading frame (ORF). The nucleotide positions for the splice donor sites were 880 for HPV16, 929 for HPV18 and HPV45, and 877 and 894 for HPV31 and HPV33, respectively. Almost all transcripts (71 of 74) were spliced directly to cellular sequences. Only in two cases splicing involved the E1 and E4 ORFs before continuing into the human sequence: In the HPV18-positive sample D3772, an ag-ta splice acceptor signal at position 3,432 was used, and in the HPV45-positive sample D3458, an ag-at splice acceptor signal was used at position 3,421. The viral sequences were fused without further splicing to human sequences from nucleotides 3,759 and 3,638, respectively. Exceptional also was

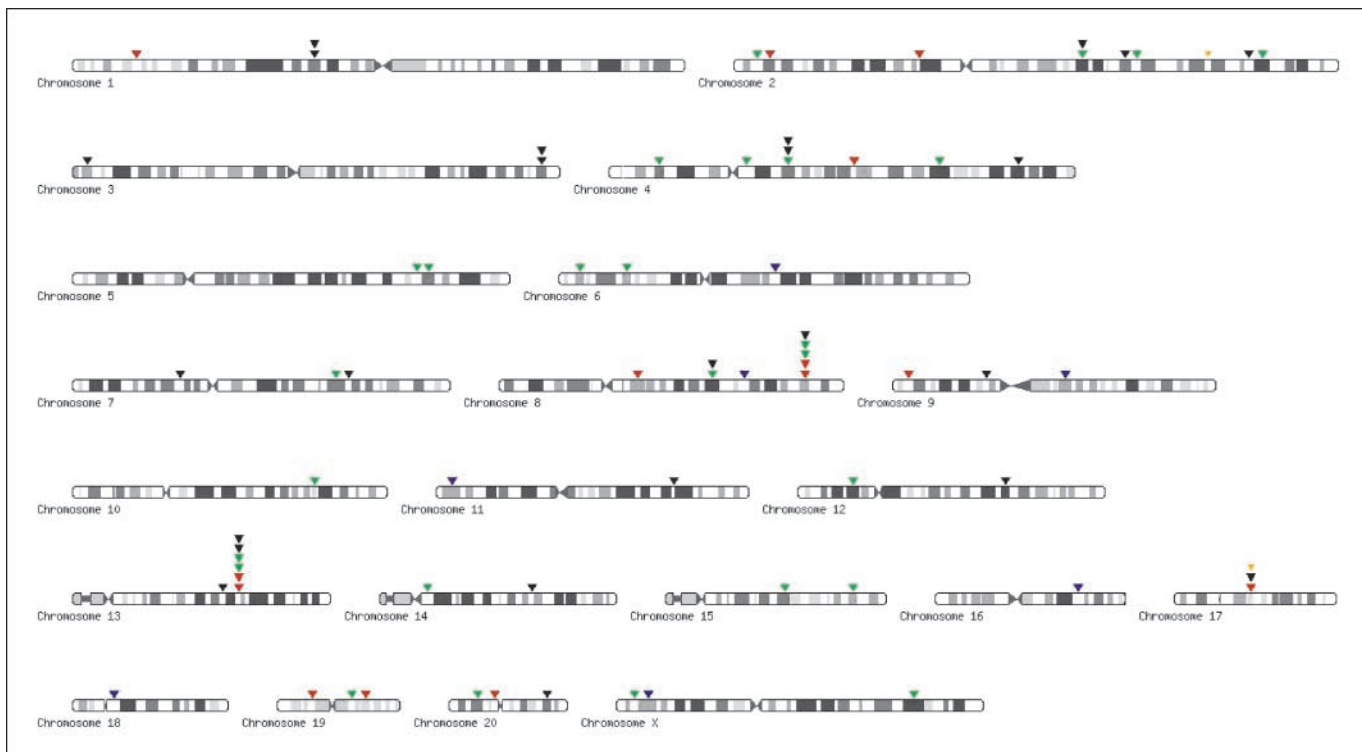


Figure 1. Integration sites were found randomly distributed throughout the genome, located over all chromosomes except for chromosomes 21 and 22. Red, HPV16; green, HPV18; black, HPV45; yellow, HPV 31; blue, HPV 33.

a fusion transcript that used a splice acceptor within E1 at nucleotide 2,779 before the transcript continued into the human sequence at nucleotide 3,033 within E2 (Fig. 2).

Cotranscribed cellular sequences show homology to known and predicted genes. Of 74 fusion transcripts, 28 cellular sequences corresponded to known genes. Forty transcripts contained sequences of predicted genes, as revealed by searching with the UCSC Blat tool (Table 1). Six samples did not show any similarity other than genomic sequences; however, for sample F380, the use of characteristic splice sites and the presence of a polyadenylation [poly(A)] signal are suggestive of an unidentified gene in this area. In some cases, cDNA synthesis was initiated at

A-rich stretches rather than at the 3' poly(A) tail. This resulted in shorter cDNAs but still permitted the identification of authentic viral-cellular fusion transcripts.

Splicing to cellular sequences and orientation of ORFs. Among the 28 fusion transcripts containing sequences of known cellular genes, 15 transcripts had both the viral and the cellular sequences in sense orientation (i.e., in a coding orientation). In most of these cases (12 of 15), the viral sequence was spliced to a cellular exon sequence (Table 1). For all but one (sample D3708, NPM3) of these transcripts, the regular cellular splice acceptor site was used. For fusion transcripts containing sequences of predicted genes, the results are more complex: For several transcripts, two or

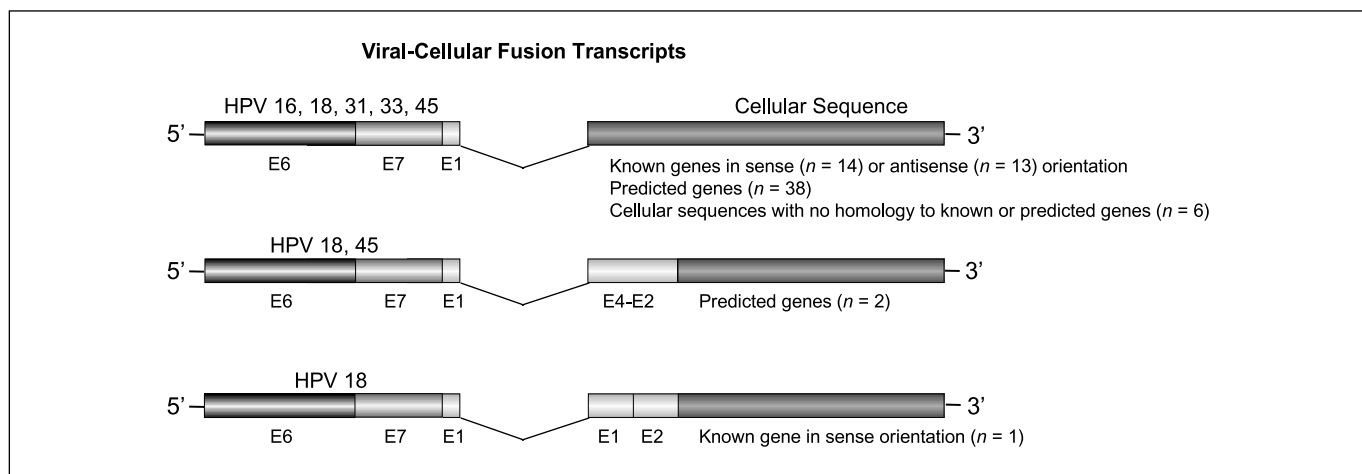


Figure 2. Three different species of fusion transcripts were observed. The predominant species showed splicing from the viral E1 ORF directly to cellular sequences.

more predicted genes were found on the same chromosomal location and often on both strands of the DNA. Among the 40 samples containing fusion transcripts with sequences of predicted genes, 6 transcripts showed fusion in sense to an exon of a predicted gene.

Discussion

DNA-based PCR methods allow the determination of the precise chromosomal location of viral integration sites. Examples for these techniques are human interspersed repetitive sequence-specific PCR (38), restriction-site PCR (39), detection of integrated papillomavirus sequences by ligation-mediated PCR (40), and restriction enzyme cleavage, self-ligation, inverse PCR (41). The integration assay of the present study is based on the amplification of cDNA by the APOT approach (34). This strategy was chosen because we wanted to limit our analyses to integration sites with a transcriptionally active viral genome. In tumors with multiple integration sites, only one integration site seems to be active (42). This integration site is not only relevant for the dysregulated expression of the viral oncogenes but may also provide further insights into the effect integration may have on the host genome. By the APOT assay, viral-cellular fusion transcripts are readily detected because the fragment size differs from episome-derived transcripts.

The DNA sequence of the cellular part of all cloned fusion transcripts was assigned to Giemsa-stained chromosomal bands taken from UCSC database (37). This enabled an easier comparison with data from the literature. Consistent with previous reports, integration sites were found located over all chromosomes, except for chromosomes 21 and 22. Besides, it is evident that several integration sites are clustered in some chromosomal regions. Twenty-three percent (17 of 74) of the integration sites analyzed were located within the cytogenetic bands 4q13.3, 8q24.21, 13q22.1, and 17q21.2. The overall physical distance between the integration sites within each cluster ranged from 86 to 900 kb, which represents exceedingly short regions in genomic terms. The most prominent integration sites are the ones in the location of 8q24.21. In a review, Wentzensen et al. (11) report on 12 clinical samples with HPV integration within band 8q24. In addition, the cell line HeLa shows HPV18 integration in this region (27, 28, 43–45). However, all previous reports refer to the broader bands 8q24, 8q24.1, or 8q24.2. Several studies also report that integration has taken place within or close to the *MYC* proto-oncogene, also known as *c-MYC*, previously mapped to the region 8q24.12–13 (28, 45). According to the UCSC hg18 database, the *MYC* gene is located within band 8q24.21. In our study, the sequences identified on 8q24.21 matched neither the *MYC* gene nor any other gene of known function. However, all integrations were located within a distance of 1 to 860 kb upstream *MYC*. Integrations of HPV close to the *MYC* gene may lead to deregulation and activation of *MYC*, which was previously shown for five cell lines harboring HPV integration in this region (46). On the other hand, the *MYC* gene is commonly deregulated in cancers and it is therefore difficult to substantiate that the deregulation is actually a result of HPV integration. Still, Peter and colleagues showed that in cell lines with integration at other chromosomal sites, no *MYC* mRNA or protein overexpression was observed, indicating a role of HPV integration in the up-regulation observed in the other cell-lines. Moreover, Ferber et al. (28) have found only HPV18 integrations and no HPV16 integrations to be associated with the *MYC* locus. Hence,

it was speculated that this could be an explanation for the generally more aggressive tumors associated with HPV18. Our study, however, does not confirm this hypothesis because we, in addition to two HPV18 integrations, found two HPV16 as well as one HPV 45 integration on 8q24.21. In analogy to the clustering of integration sites on chromosome 8, all of six integration sites on 13q22.1 were located within a region of 380 kb adjacent to the gene *KLF5*. The influence the viral sequences exert on *KLF5* has not been examined, but it is of interest to note that, by ontology, *KLF5* is termed as a tumor suppressor gene. Moreover, clusters with three viral integration sites each are located on bands 4q13.3 and 17q21.2. Further evidence for HPV integration sites within these two sub-bands is provided in the literature (22, 47). Of interest also are the two integration sites within the gene *LRP1B* (low-density lipoprotein-related protein 1B) on 2q22.1. HPV integration within this gene has also been reported in two previous independent studies (25, 28). *LRP1B* belongs to the low-density lipoprotein receptor gene family, and through interactions with multiple ligands, these receptors play a wide variety of roles in normal cell function and development (48). Further, based on the finding about its inactivation in different tumor types, *LRP1B* has been reported to be a candidate tumor suppressor gene (49–51).

Viral integration within fragile sites has previously been reported in several studies (23–25). Depending on their mode of induction and frequency within the population, fragile sites are grouped as rare or common (52). Fragile sites are regions that form gaps, breaks, or rearrangements on chromosomes exposed to replication stress, and genes in or near these sites are therefore disposed to foreign DNA integration. It is also argued that the frequently observed integration within fragile sites can be related to their greater susceptibility to integration-induced chromosomal alterations (22, 24, 28, 29, 39, 53). Of the 74 integrations presented in our study, 21 were located within a common or rare fragile site. It is also of note that the clustered integration sites on 8q24.21 and 13q22.1 were located within 1.52 and 0.83 Mbp, respectively, of an adjacent fragile site and that the clusters on 4q13.3 and 17q21 were located >15 Mbp distant to any known fragile site. Thus, it is tempting to speculate that as yet unknown fragile sites may be identified on the basis of HPV integration hotspots (Table 1).

In total, 28 of 74 (37.8%) samples showed integration within known genes and 40 of 74 (54.1%) samples within predicted genes. Considering the fact that coding sequences comprise only 2% of the genome (54), these data strongly indicate that integration as such is not an entirely random event but may frequently occur in transcriptionally active regions, which may simply be more accessible for integration. This was confirmed in additional experiments in which the transcriptional activity for a subset of the predicted genes ($n = 10$) was analyzed in normal cervical tissue and CIN. In all cases, transcriptional activity was shown in at least one of the tissues examined by semi-nested RT-PCR. Moreover, direct evidence for the involvement of cellular genes is provided by those fusion transcripts ($n = 18$) in which viral 5' coding sequences are fused to cellular exons. For these cases, it remains to be shown whether viral integration has resulted in disruption of the gene or is located upstream of the coding sequences.

The results reported here expand on previous studies about HPV16 and HPV18 integration sites. We also provide the first detailed sequence data on HPV31, HPV33, and HPV45 fusion transcripts identified by the APOT assay. Our findings strongly suggest that there is no correlation between HPV type and specific integration loci. Furthermore, the observed clustering of integration

sites in the cytogenetic bands 4q13.3, 8q24.21, 13q22.1, and 17q21 may reflect chromosomal regions, which are genetically unstable and which may be relevant to the development of cervical cancer. In addition, the relatively high number of integration sites in either defined genes or predicted genes suggests that the disruption or deregulation of specific genes through HPV integration could contribute to the process of transformation and carcinogenesis. This needs to be addressed systematically in future analyses.

Acknowledgments

Received 7/20/2007; revised 11/27/2007; accepted 1/30/2008.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank Gunnar Kristensen (Department of Gynaecologic Oncology) and Kathrine Lie and Ruth Holm (Department of Pathology, Norwegian Radium Hospital, Oslo, Norway); Björn Hagmar (Institute of Pathology, Rikshospitalet University Hospital, Oslo, Norway); and Lars Espen Erno (Department of Gynaecology and Obstetrics, Østfold County Hospital, Fredrikstad, Norway) for providing clinical samples and for histologic verification of the biopsies.

References

- Nobbenhuis MA, Walboomers JM, Helmerhorst TJ, et al. Relation of human papillomavirus status to cervical lesions and consequences for cervical-cancer screening: a prospective study. *Lancet* 1999;354:20–5.
- Remmink AJ, Walboomers JM, Helmerhorst TJ, et al. The presence of persistent high-risk HPV genotypes in dysplastic cervical lesions is associated with progressive disease: natural history up to 36 months. *Int J Cancer* 1995;61:306–11.
- Munoz N, Bosch FX, de Sanjose S, et al. Epidemiologic classification of human papillomavirus types associated with cervical cancer. *N Engl J Med* 2003;348:518–27.
- Kraus I, Molden T, Holm R, et al. Presence of E6 and E7 mRNA from human papillomavirus types 16, 18, 31, 33, and 45 in the majority of cervical carcinomas. *J Clin Microbiol* 2006;44:1310–7.
- zur Hausen H. Papillomaviruses causing cancer: evasion from host-cell control in early events in carcinogenesis. *J Natl Cancer Inst* 2000;92:690–8.
- Mantovani F, Banks L. The human papillomavirus E6 protein and its contribution to malignant progression. *Oncogene* 2001;20:7874–87.
- Munger K, Howley PM. Human papillomavirus immortalization and transformation functions. *Virus Res* 2002;89:213–28.
- Duensing S, Lee LY, Duensing A, et al. The human papillomavirus type 16 E6 and E7 oncoproteins cooperate to induce mitotic defects and genomic instability by uncoupling centrosome duplication from the cell division cycle. *Proc Natl Acad Sci U S A* 2000;97:10002–7.
- Munger K, Hayakawa H, Nguyen CL, Melquiot NV, Duensing A, Duensing S. Viral carcinogenesis and genomic instability. *EXS* 2006;(96):179–99.
- 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;66:1515–22.
- 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;64:3878–84.
- Pett MR, Alazawi WO, Roberts I, et al. Acquisition of high-level chromosomal instability is associated with integration of human papillomavirus type 16 in cervical keratinocytes. *Cancer Res* 2004;64:1359–68.
- Jeon S, Allen-Hoffmann BL, Lambert PF. Integration of human papillomavirus type 16 into the human genome correlates with a selective growth advantage of cells. *J Virol* 1995;69:2989–97.
- Romanczuk H, Howley PM. Disruption of either the E1 or the E2 regulatory gene of human papillomavirus type 16 increases viral immortalization capacity. *Proc Natl Acad Sci U S A* 1992;89:3159–63.
- Arias-Pulido H, Peyton CL, Joste NE, Vargas H, Wheeler CM. Human papillomavirus type 16 integration in cervical carcinoma *in situ* and in invasive cervical cancer. *J Clin Microbiol* 2006;44:1755–62.
- von Knebel Doeberitz M, Bauknecht T, Bartsch D, zur Hausen H. Influence of chromosomal integration on glucocorticoid-regulated transcription of growth-stimulating papillomavirus genes E6 and E7 in cervical carcinoma cells. *Proc Natl Acad Sci U S A* 1991;88:1411–5.
- Dürst M, Bosch FX, Glitz D, Schneider A, zur HH. Inverse relationship between human papillomavirus (HPV) type 16 early gene expression and cell differentiation in nude mouse epithelial cysts and tumors induced by HPV-positive human cell lines. *J Virol* 1991;65:796–804.
- Stoler MH, Rhodes CR, Whitbeck A, Wolinsky SM, Chow LT, Broker TR. Human papillomavirus type 16 and 18 gene expression in cervical neoplasias. *Hum Pathol* 1992;23:117–28.
- de Boer MA, Jordanova ES, Kenter GG, et al. High human papillomavirus oncogene mRNA expression and not viral DNA load is associated with poor prognosis in cervical cancer patients. *Clin Cancer Res* 2007;13:132–8.
- Rosty C, Sheffer M, Tsafir D, et al. Identification of a proliferation gene cluster associated with HPV E6/E7 expression level and viral DNA load in invasive cervical carcinoma. *Oncogene* 2005;24:7094–104.
- Häfner N, Driesch C, Gajda M, et al. Integration of the HPV16 genome does not invariably result in high levels of viral oncogene transcripts. *Oncogene*. Epub 2007 Sep 10.
- Yu T, Ferber MJ, Cheung TH, Chung TK, Wong YF, Smith DI. The role of viral integration in the development of cervical cancer. *Cancer Genet Cytogenet* 2005;158:27–34.
- Koopman LA, Szuhai K, van Eendenburg JD, et al. Recurrent integration of human papillomaviruses 16, 45, and 67 near translocation breakpoints in new cervical cancer cell lines. *Cancer Res* 1999;59:5615–24.
- Thorland EC, Myers SL, Gostout BS, Smith DI. Common fragile sites are preferential targets for HPV16 integrations in cervical tumors. *Oncogene* 2003;22:1225–37.
- Ziegert C, Wentzensen N, Vinokurova S, et al. A comprehensive analysis of HPV integration loci in anogenital lesions combining transcript and genome-based amplification techniques. *Oncogene* 2003;22:3977–84.
- Reuter S, Bartelmann M, Vogt M, et al. 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;17:215–22.
- Dürst M, Croce CM, Gissmann L, Schwarz E, Huebner K. Papillomavirus sequences integrate near cellular oncogenes in some cervical carcinomas. *Proc Natl Acad Sci U S A* 1987;84:1070–4.
- Ferber MJ, Thorland EC, Brink AA, et al. Preferential integration of human papillomavirus type 18 near the c-myc locus in cervical carcinoma. *Oncogene* 2003;22:7233–42.
- Ferber MJ, Montoya DP, Yu C, et al. Integrations of the hepatitis B virus (HBV) and human papillomavirus (HPV) into the human telomerase reverse transcriptase (hTERT) gene in liver and cervical cancers. *Oncogene* 2003;22:3813–20.
- Jacobs MV, Snijders PJ, van den Brule AJ, Helmerhorst TJ, Meijer CJ, Walboomers JM. A general primer GP5+/GP6(+)-mediated PCR-enzyme immunoassay method for rapid detection of 14 high-risk and 6 low-risk human papillomavirus genotypes in cervical scrapings. *J Clin Microbiol* 1997;35:791–5.
- Kraus I, Molden T, Erno LE, Skomedal H, Karlsen F, Hagmar B. Human papillomavirus oncogenic expression in the dysplastic portio; an investigation of biopsies from 190 cervical cones. *Br J Cancer* 2004;90:1407–13.
- Frohman MA. Rapid amplification of complementary DNA ends for generation of full-length complementary DNAs: thermal RACE. *Methods Enzymol* 1993;218:340–56.
- Häfner N, Gajda M, Altgassen C, et al. HPV16–6 mRNA is superior to cytokeratin 19 mRNA as a molecular marker for the detection of disseminated tumour cells in sentinel lymph nodes of patients with cervical cancer by quantitative reverse-transcription PCR. *Int J Cancer* 2007;120:1842–6.
- Klaes R, Woerner SM, Ridder R, et al. Detection of high-risk cervical intraepithelial neoplasia and cervical cancer by amplification of transcripts derived from integrated papillomavirus oncogenes. *Cancer Res* 1999;59:6132–6.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–10.
- Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. *Genome Res* 2002;12:996–1006.
- Furey TS, Haussler D. Integration of the cytogenetic map with the draft human genome sequence. *Hum Mol Genet* 2003;12:1037–44.
- Carmody MW, Jones M, Tarraza H, Vary CP. Use of the polymerase chain reaction to specifically amplify integrated HPV-16 DNA by virtue of its linkage to interspersed repetitive DNA. *Mol Cell Probes* 1996;10:107–16.
- Thorland EC, Myers SL, Persing DH, et al. Human papillomavirus type 16 integrations in cervical tumors frequently occur in common fragile sites. *Cancer Res* 2000;60:5916–21.
- Luft F, Klaes R, Nees M, et al. Detection of integrated papillomavirus sequences by ligation-mediated PCR (DIPS-PCR) and molecular characterization in cervical cancer cells. *Int J Cancer* 2001;92:9–17.
- Kalantari M, Blennow E, Hagmar B, Johansen B. Physical state of HPV16 and chromosomal mapping of the integrated form in cervical carcinomas. *Diagn Mol Pathol* 2001;10:46–54.
- Van Tine BA, Knops J, Broker TR, Chow LT, Moen PT, Jr. *In situ* analysis of the transcriptional activity of integrated viral DNA using tyramide-FISH. *Dev Biol (Basel)* 2001;106:381–5.
- Cannizzaro LA, Dürst M, Mendez MJ, Hecht BK, Hecht F. Regional chromosome localization of human papillomavirus integration sites near fragile sites, oncogenes, and cancer chromosome breakpoints. *Cancer Genet Cytogenet* 1988;33:93–8.
- Popescu NC, DiPaolo JA. Integration of human papillomavirus 16 DNA and genomic rearrangements in immortalized human keratinocyte lines. *Cancer Res* 1990;50:1316–23.
- 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;21:419–26.
- 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;25:5985–93.
47. Shera KA, Shera CA, McDougall JK. Small tumor virus genomes are integrated near nuclear matrix attachment regions in transformed cells. *J Virol* 2001;75:12339–46.
48. Liu CX, Li Y, Obermoeller-McCormick LM, Schwartz AL, Bu G. The putative tumor suppressor LRP1B, a novel member of the low density lipoprotein (LDL) receptor family, exhibits both overlapping and distinct properties with the LDL receptor-related protein. *J Biol Chem* 2001; 276:28889–96.
49. Langbein S, Szakacs O, Wilhelm M, et al. Alteration of the LRP1B gene region is associated with high grade of urothelial cancer. *Lab Invest* 2002;82:639–43.
50. Liu CX, Musco S, Lisitsina NM, Yaklichkin SY, Lisitsyn NA. Genomic organization of a new candidate tumor suppressor gene, LRP1B. *Genomics* 2000; 69:271–4.
51. Sonoda I, Imoto I, Inoue J, et al. Frequent silencing of low density lipoprotein receptor-related protein 1B (LRP1B) expression by genetic and epigenetic mechanisms in esophageal squamous cell carcinoma. *Cancer Res* 2004;64:3741–7.
52. Schwartz M, Zlotorynski E, Kerem B. The molecular basis of common and rare fragile sites. *Cancer Lett* 2006; 232:13–26.
53. Wilke CM, Hall BK, Hoge A, Paradee W, Smith DI, Glover TW. FRA3B extends over a broad region and contains a spontaneous HPV16 integration site: direct evidence for the coincidence of viral integration sites and fragile sites. *Hum Mol Genet* 1996;5: 187–95.
54. Shabalina SA, Spiridonov NA. The mammalian transcriptome and the function of non-coding DNA sequences. *Genome Biol* 2004;5:105.