Available online at www.sciencedirect.com

SCIENCE DIRECT

Gynecologic Oncology 96 (2005) 695–700

Gynecologic Oncology

www.elsevier.com/locate/ygyno

Human papillomavirus type 16 E2 and E6/E7 variants

David C. Swan[a,*], Mangalathu Rajeevan[a], Guillermo Tortolero-Luna[b], Michele Follen[b], Ruth Ann Tucker[a], Elizabeth R. Unger[a]

[a]*U.S. Department of Health and Human Services, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Public Health Service, Atlanta, GA 30333, USA*

[b]*Department of Gynecologic Oncology, University of Texas M.D. Anderson Cancer Center, Houston, TX 77030, USA*

Received 23 June 2004
Available online 19 December 2004

Abstract

Objectives. Polymorphisms in human papillomavirus (HPV) type 16 have been shown to be related to geographic areas and are broadly classified as European (E), African (Af), Asian (As), or Asian-American (AA). Certain variants have been reported as being more likely to cause cervical disease; our objectives were to identify new HPV16 polymorphisms, to determine the linkage of the E2 and E6/E7 regions and to determine the minimum sequence necessary to classify variants.

Methods. We sequenced the complete E2, E6, and E7 regions in all HPV16-positive cervical samples identified in a case-control study of pre-invasive cervical disease.

Results. In the 100 samples analyzed, only one new polymorphism was identified, a synonymous change, T3205A, in region E2. The frequency distribution of variants in the sample set was 37 European prototypes and 27 E-G350, 16 AA, 5 Af1, 2 Af2, 8 E-C109G, 3 E-G131G, and 2 As. As shown by others, region E7 varied much less than E6 and E2.

Conclusions. In each case, E2 changes were linked to the expected E6/E7 changes, and there was no evidence for recombination. The linkage between E2 and E6/E7 allows variant classification to be based on a short E6 sequence (nt 109–350).

Keywords: HPV16; Variants; E6; E7; E2; Cervical disease

Introduction

Human papillomaviruses (HPVs) are classified on the basis of their genomic DNA sequence. An HPV type has less than 90% similarity with other types at the nucleotide level [1]; HPVs with more than 98% similarity are considered type variants. Only HPV16 has been extensively sequenced to characterize variants: five major variant groups have been identified on the basis of sequences in the long control region (LCR) and E5 open reading frame [2–4]. These variant groups are referred to as European (E), Asian (As), Asian-American (AA), African-1 (Af-1), and African-2 (Af-2), loosely designating their geographic distribution. Sequencing of the E2, E6, L2, and L1 regions of the HPV16 genome [5–7] resulted in the identification of European prototype (Ep) variants E-G350 [T350G], E-G131 [A131G], and E-C109 [T109C] and North American variant NA-1 [8].

Nucleotide changes characteristic of a variant in one gene are generally linked to those in other genes, suggesting that variants are stable and are not subject to frequent recombination. In studies of variants, L1 polymorphisms have been linked to those in LCR, E5, and E6/E7 [6,9]. E2 nucleotide changes have been described and linked to those in E6 and L1 [10], E5 and E6 [7], and LCR [11]; however, a linkage between nucleotide changes in E6, E7, and E2 has not been shown previously. When attempting to evaluate the biologic significance of variants, it is important to understand and verify the linkage of polymorphisms throughout the genome.

Many of the nucleotide changes in the variants are in coding regions, and these non-synonymous changes may

* Corresponding author.
E-mail address: dswan@cdc.gov (D.C. Swan).

0090-8258/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygyno.2004.11.045

result in changes in the functional or antigenic domains of the virus that may be important in the host immune response or in functional aspects of the viral proteins. In vitro studies have shown, for example, that some E6 variant proteins have a much reduced ability to degrade p53 [12]. Other variations in E6 (E-G131T) result in amino acid changes in potential T-cell receptor or HLA binding regions [13].

Data correlating viral persistence and cervical disease progression with different variants are, however, inconclusive [14–17]. Some data suggest that women infected with certain variants have a significantly greater chance of developing a high-grade lesion [18]. In some studies, the variant with the T350G change in E6 (L83V) was associated with an increased risk of cervical disease progression compared with the prototype strain [5,19]. In a study in Costa Rica, a strong correlation was seen between non-European variants and disease [20], but other studies have shown no correlation between variant and disease [9].

In this study, we sequenced the complete E2, E6, and E7 regions of all HPV16-positive samples identified in a case-control study of pre-invasive cervical disease. The objectives were to identify new polymorphisms, to determine the linkage of the E2 and E6/E7 regions, and to determine the minimum sequence necessary to classify variants. In the 100 samples analyzed, only one new polymorphism was identified, a synonymous change, T3205A, in E2. The linkage between E2 and E6/E7 allows variant classification to be based on a short E6 sequence (nt. 109–350) rather than on long sequences of the HPV genome.

Materials and methods

Study design

The study design and population has been previously described [21]. In brief, this was a case-control study of pre-invasive cervical neoplasia, enrolling non-pregnant women who were residents of Harris County, Texas, and 18 years or older at the time of the study. Cervical disease status was determined on the basis of biopsy (cases) or cytology (controls) findings. We used L1 consensus polymerase chain reaction (PCR) on exfoliated cervical cells to detect HPV. End-point fluorescent PCR was used to type and quantitate HPV16, 18, 33, and 45 [22].

Sample preparation and sequencing

Cervical cytobrush samples had been stored at −70°C since the original DNA extraction [22]. Residual DNA was obtained from the stored cytobrushes for each HPV16-positive sample by using MasterAmp Buccal Swab DNA Extraction Solution (Epicentre Technologies, Madison, WI). Sequencing templates were generated by amplification of the E6/E7 (nt. 7740–943) and the E2 (nt. 2655–4008) regions from each extract. Reaction mixtures also contained 15 pmol of each primer, 0.125 U of *Taq* DNA polymerase (Platinum*Taq*, Gibco-BRL, Gaithersburg, MD), 250 μM dNTPs, 2.5 mM $MgCl_2$, and 1 × PCR buffer (Gibco-BRL). After an initial denaturation at 95°C for 2 min, reaction mixtures underwent 40 cycles at 95°C for 30 s and 62°C for 2 min. Ten microliters from each 50-μl PCR sample were run on 1% agarose gels to test for product; the remaining 40 μl from samples showing a visible band were purified by using Centricon YM-100 centrifugal filter devices. One microliter of this purified material was used as template in half-size (10 μl) cycle-sequencing reactions (BigDye Terminator Cycle Sequencing Kit, Applied Biosystems, Foster City, CA). Sequencing reaction mixtures were purified using Centricep Spin Columns (Princeton Separations, Adelphi, NJ), and run in the Applied Biosystems model 377 sequencing apparatus. Of the original 105 HPV16-positive samples, 100 yielded sufficiently intact DNA to give readable sequences.

Sequence data were compiled and analyzed using Sequencher sequence analysis software (Gene Codes Corp., Ann Arbor, MI). Sequence variations were detected in comparison to the reference HPV16R sequence (http://hpv-web.lanl.gov). Variants were classified on the basis of previously published E6 sequence changes: Ep (prototype sequence, no changes), E-G350 (T350G), E-C109G (T109C, T350G), E-G131G (A131G, T350G), As (T178G), Af1 (G132C, C143G, G145T, T286A, A289G, C335T), Af2 (T109C, G132T, C143G, G145T, T286A, A289G, C335T, G403G), AA (G145T, T286A, A289G, C335T, T350G).

Results

The nucleotide changes and variants identified in this study are shown in Table 1. Because 100 samples were sequenced, the sample number is the percent. Table 1 also shows the variant positions and coding changes in E6/E7 and E2 reported in the literature. The entire E2 gene was sequenced in all samples. Two high-grade lesion samples had small deletions (36 and 33 bp, respectively) in E2 that did not result in a reading frame change.

The prototype HPV16 sequence (HPV16R) has an A at position 2926. All samples in this study had a G at this position, as noted in a published correction to the reference sequence [23]. Only one previously unreported change was found, a synonymous change, T3205A, in E2. Three additional minor variants of the European prototype, Ep-a, Ep-b, and Ep-c, could be identified by using the E2 polymorphisms. No example of NA-1 was found in this population. No base changes were identified in the splice donor or two splice acceptor sites in E6. HPV16 variants predominated in this population. Even if the minor prototype variants (Ep-a, Ep-b, Ep-c) are included as prototype, only 36.6% of the samples were prototype.

Table 1
Variant positions in HPV16 E6/E7 and E2

Amino acid[a]			F-F	R-G	R-T/I	Q-E	Q-H	D-E	A-A	V-V	H-Y	L-V	L-L	S-S	M-V	F-S	F-S	L-R	L-P	P-Q	D-N	H-Q	T-T	D-D	T-K	H-Y	E-D	A-T	G-G	L-I	R-Q	N-D	P-A	I-T	I-T	P-S	T-P	A-T	E-K	T-N	T-T	L-L	S-S	F-V	S-A	Y-Y	T-K	T-T	S-S	W-C	R-R	D-E	V-V	I-T	
	31	83	109	131	132	143	145	178	286	289	335	350	403	532	647	732	789	795	846	864	2828	2860	2938	3043	3159	3161	3181	3182	3205*	3224	3249	3362	3377	3384	3387	3410	3425	3431	3449	3516	3517	3524	3538	3566	3575	3664	3684	3694	3706	3778	3784	3787	3805	3858	Samples
Prototype	C	A	T	A	G	C	G	T	T	A	C	T	A	A	A	T	T	T	T	C	G	C	A	C	C	C	A	G	T	T	G	A	C	T	T	C	A	G	G	C	T	T	A	T	T	T	C	T	T	G	T	C	T	T	No.
NA1			-	-	-	-	T	-	a	g	T	G	-	-																																											0
AA	-	-	-	-	-	-	T	-	a	g	T	G	-	g	-	C	C	G	-	-	-	A	-	-	A	-	C	A	-	A	A	G	G	-	C	T	-	-	A	A	c	-	c	G	-	c	A	a	c	T	-	A	g	C	16
Af1	-	c	-	-	C	G	T	-	a	g	T	-	-	-	-	-	C	G	-	-	-	A	-	-	A	T	C	A	-	-	A	G	G	-	-	T	C	-	A	A	-	-	-	G	-	c	A	a	-	T	g	A	-	C	5
Af2	t	-	c	-	T	G	T	-	a	g	T	-	g	-	G	-	C	G	-	-	-	A	-	t	A	T	-	A	-	-	A	G	G	-	-	T	-	A	A	A	c	-	c	G	-	-	A	a	c	T	-	A	g	C	2
As	-	-	-	-	-	-	-	G	-	-	-	-	-	-	G	-	-	-	C	-	a	-	-	-	A	-	-	-	-	-	A	-	-	C	-	T	-	-	A	-	-	c	-	-	-	-	A	-	-	-	A	-	-	2
E-C109G	-	-	c	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	8
E-G131G	-	-	-	G	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	A	-	-	g	-	-	-	-	-	-	-	-	-	-	C	-	T	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	3
E-G350	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	27
Ep	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	14
Ep-a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	13
Ep-b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	g	-	-	-	-	-	-	-	-	-	-	C	-	T	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	7
Ep-c	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3
			E6													E7						E2																																		100

[a] Amino acid (coding) changes shown in the first row, nucleotide numbering in the second row, and sequence in the third row are based on the HPV16 prototype sequence in HPV16R [23]. Synonymous base changes are in lower case letters, and non-synonymous changes are shown upper case letters.

* T3205A has not previously been reported.

By visual inspection of the tabulated sequences, we recognized a pattern in the association of sequence variation in the E6, E7, and E2 regions. We found that by assigning variant status based on variation in one region, for example, AA variant based on variation in E6, then variations for AA in E7 and E2 were predictable. This association or linkage of sequence variations was observed with all samples for the major variants.

We express viral load as the number of copies of viral DNA per microgram of cellular DNA. The range in viral load for all 100 women was more than 10 logs and showed a non-normal distribution (3.29×10^3 to 1.01×10^{15}; median, 7.02×10^8; average, 1.89×10^{13}). Viral load for each variant is shown in Fig. 1. While the median number of copies per microgram DNA differed between the variant subtypes, no significant associations were found for individual variants, although small sample numbers limit the precision of the estimation. A high viral load (relative to the median number of copies) was found more frequently in European prototype samples (68%) than in non-prototype samples (40%; $P = 0.0071$). However, high viral load was not associated with Ep-a,-b, or-c variants when compared with Ep, nor was it associated with the E6L83V variant when compared with all other types.

Table 2 shows the distribution of cervical disease by HPV16 variant. Cervical disease status was determined on the basis of biopsy findings for cases and on normal cytology for controls. There were no significant associations of individual variants with disease, and when all variants other than Ep-a,-b, and-c were grouped as non-prototype, the association with cervical intra-epithelial neoplasia II/III was not significant compared with the association with the prototype.

Table 2
Cervical Intra–Epithelial Neoplasia (CIN) disease status by HPV16 variant

Variant	Number (%) of cell samples in various disease stages			
	No disease ($n = 5$)	CIN I/HPV ($n = 26$)	CIN II ($n = 27$)	CIN III ($n = 42$)
AA	0 (0)	5 (19)	7 (26)	4 (10)
Af1	0 (0)	1 (4)	2 (7)	2 (5)
Af2	0 (0)	1 (4)	1 (4)	0 (0)
As	0 (0)	0 (0)	0 (0)	2 (5)
E-C109G	0 (0)	2 (8)	3 (11)	3 (7)
E-G131G	0 (0)	0 (0)	2 (7)	1 (2)
E-G350	3 (60)	5 (19)	4 (15)	15 (36)
Ep	0 (0)	7 (27)	3 (11)	4 (10)
Ep-a	1 (20)	3 (12)	3 (11)	6 (14)
Ep-b	1 (20)	1 (4)	0 (0)	5 (12)
Ep-c	0 (0)	1 (4)	2 (7)	0 (0)

Discussion

The main objectives of this study were to sequence the HPV16 E6, E7, and E2 regions of all HPV16-positive samples in a previously described case control study to identify new polymorphisms and to determine the linkage of the E2 and E6/E7 regions. We determined that these regions were linked. Because of the linkage, sequencing a short E6 sequence (nt. 109–350) would be sufficient to classify a sample into the major variant groups. To distinguish NA-1 from AA, it would be necessary to include a slightly longer sequence (nt. 109–532); NA-1 has wild-type A at nt. 532, whereas AA has a

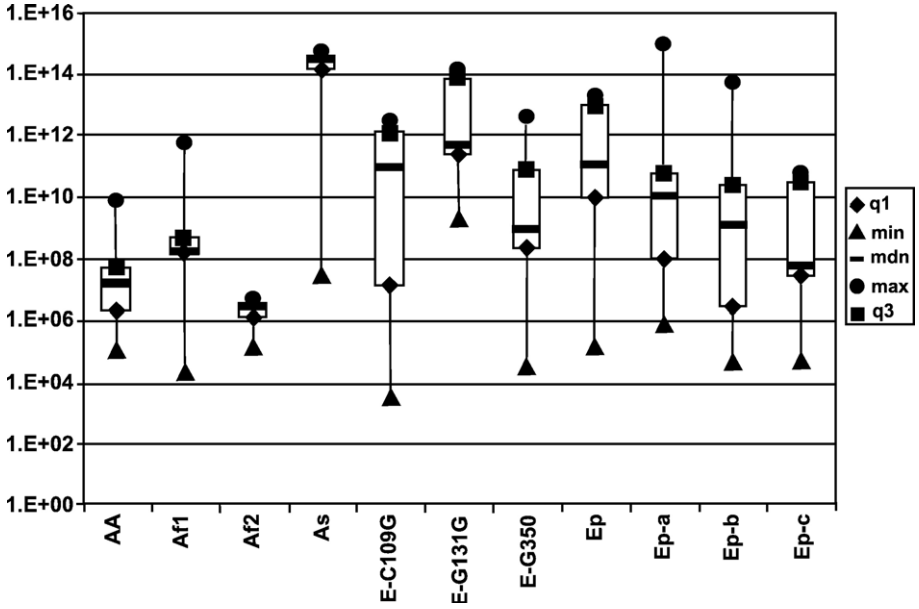

Fig. 1. Viral load by HPV16 Variant. Boxplot, q1 = first quartile, min = minimum, mdn = median, max = maximum and q3 = third quartile for HPV16 copies/μg for each of the variant groups. Median values by variant were as follows: AA, 1.87×10^7; Af1, 1.87×10^8; Af2, 2.61×10^6; As, 2.98×10^{14}; E-C109G, 9.75×10^{10}; E-G131G, 5.24×10^{11}; E-G350, 9.12×10^8; Ep, 1.12×10^{11}; Ep-a, 1.04×10^{10}; Ep-b, 1.15×10^9; Ep-c, 5.43×10^7.

synonymous G at this position. The uncommon North American variant (NA-1) [6], which is a derivative of the AA and Af2 variants, was not present in this patient population.

With the exception of two E-G350 samples that showed small E2 deletions, the E2 gene was sequenced in its entirety in all samples, suggesting that despite potential viral integration, intact E2 was also present. Three additional minor variants of the European prototype could be identified on the basis of nucleotide changes in E2 (Fig. 1): Ep-a, which is defined by changes T3205A (Gly silent) and T3575G (Ser to Ala); Ep-b, which is defined by changes A2938G (Thr silent), T3384C (Ile to Thr), C3410T (Pro to Ser), and C3684A (Thr to Lys); and Ep-c, which is defined by change C3410T (Pro to Ser). Of these changes, only the silent mutation T3205A has not previously been reported [7,11,24].

In this sample, most HPV16-positive samples were European variants (Table 1); of these, the European prototype predominated (37 samples) and was followed by the E-G350 European variant (27 samples). The prevalence of other variants was low. While this analysis found no association between variants and viral load or cervical disease, the interpretation is limited by the numbers of each variant in the study. Variant determination in other populations from different clinical settings and geographical locations may yield other results; this task will be easier because only a small region of E6 need be sequenced.

Sequence variations in LCR were not studied here. It is possible that viral gene expression through either sequence variation or methylation status in this region could impact clinical outcome. It would be interesting to see whether variations in the LCR were also linked to the regions studied here; this work may ultimately identify patterns which would serve as markers for disease progression.

Acknowledgment

We thank Ms. Deon Williams for sequencing the variant samples.

References

[1] Bernard HU, Chan SY, Manos MM, Ong CK, Villa LL, Delius H, et al. Identification and assessment of known and novel human papillomaviruses by polymerase chain reaction amplification, restriction fragment length polymorphisms, nucleotide sequence, and phylogenetic algorithms. J Infect Dis 1994;170:1077–85.

[2] Chan SY, Ho L, Ong CK, Chow V, Drescher B, Dürst M, et al. Molecular variants of human papillomavirus type 16 from four continents suggest ancient pandemic spread of the virus and its coevolution with humankind. J Virol 1992;66:2057–66.

[3] Ho L, Chan SY, Chow V, Chong T, Tay SK, Villa LL, et al. Sequence variants of human papillomavirus type 16 in clinical samples permit verification and extension of epidemiological studies and construction of a phylogenetic tree. J Clin Microbiol 1991;29:1765–72.

[4] Ho L, Chan SY, Burk RD, Das BC, Fujinaga K, Icenogle JP, et al. The genetic drift of human papillomavirus type 16 is a means of reconstructing prehistoric viral spread and the movement of ancient human populations. J Virol 1993;67:6413–23.

[5] Kämmer C, Tommasino M, Syrjänen S, Delius H, Hebling U, Warthorst U, et al. Variants of the long control region and the E6 oncogene in European human papillomavirus type 16 isolates: implications for cervical disease. Br J Cancer 2002;86:269–73.

[6] Yamada T, Manos MM, Peto J, Greer CE, Munoz N, Bosch FX, et al. Human papillomavirus type 16 sequence variation in cervical cancers: a worldwide perspective. J Virol 1997;71:2463–72.

[7] Eriksson A, Herron JR, Yamada T, Wheeler CM. Human papillomavirus type 16 variant lineages characterized by nucleotide sequence analysis of the E5 coding segment and the E2 hinge region. J Gen Virol 1999;80:595–600.

[8] Yamada T, Wheeler CM, Halpern AL, Stewart AC, Hildesheim A, Jenison SA. Human papillomavirus type 16 variant lineages in United States populations characterized by nucleotide sequence analysis of the E6, L2, and L1 coding segments. J Virol 1995;69:7743–53.

[9] Icenogle JP, Sathya P, Miller DL, Tucker RA, Rawls WE. Nucleotide and amino acid sequence variation in the L1 and E7 open reading frames of human papillomavirus type 6 and type 16. Virology 1991; 184:101–7.

[10] Casas L, Galvan SC, Ordoñez RM, 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;83:449–55.

[11] Veress G, Szarka K, Dong XP, Gergely L, Pfister H. Functional significance of sequence variation in the E2 gene and the long control region of human papillomavirus type 16. J Gen Virol 1999; 80:1035–43.

[12] Stoppler MC, Ching K, Stoppler H, Clancy K, Schlegel R, Icenogle J. Natural variants of the human papillomavirus type 16 E6 protein differ in their abilities to alter keratinocyte differentiation and to induce p53 degradation. J Virol 1996;70:6987–93.

[13] Ellis JR, Keating PJ, Baird J, Hounsell EF, Renouf DV, Rowe M, et al. The association of an HPV16 oncogene variant with HLA-B7 has implications for vaccine design in cervical cancer. Nat Med 1995; 1:464–70.

[14] Londesborough P, Ho L, Terry G, Cuzick J, Wheeler C, Singer A. Human papillomavirus genotype as a predictor of persistence and development of high-grade lesions in women with minor cervical abnormalities. Int J Cancer 1996;69:364–8.

[15] Xi LF, Koutsky LA, Galloway DA, Kuypers J, Hughes JP, Wheeler CM, et al. Genomic variation of human papillomavirus type 16 and risk for high grade cervical intraepithelial neoplasia. J Natl Cancer Inst 1997;89:796–802.

[16] Nindl I, Rindfleisch K, Lotz B, Schneider A, Durst M. Uniform distribution of HPV 16 E6 and E7 variants in patients with normal histology, cervical intra-epithelial neoplasia and cervical cancer. Int J Cancer 1999;82:203–7.

[17] Bontkes HJ, van Duin M, de Gruijl TD, Duggan-Keen MF, Walboomers JM, Stukart MJ, et al. HPV 16 infection and progression of cervical intra-epithelial neoplasia: analysis of HLA polymorphism and HPV 16 E6 sequence variants. Int J Cancer 1998; 78:166–71.

[18] Villa LL, Sichero L, Rahal P, Caballero O, Ferenczy A, Rohan T, et al. Molecular variants of human papillomavirus types 16 and 18 preferentially associated with cervical neoplasia. J Gen Virol 2000; 81:2959–68.

[19] Zehbe I, Wilander E, Delius H, Tommasino M. Human papillomavirus 16 E6 variants are more prevalent in invasive cervical carcinoma than the prototype. Cancer Res 1998;58:829–33.

[20] Hildesheim A, Schiffman M, Bromley C, Wacholder S, Herrero R, Rodriguez AC, et al. Human papillomavirus type 16 variants and risk of cervical cancer. J Natl Cancer Inst 2001;93:315–8.

[21] Tortolero-Luna G, Mitchell MF, Swan DC, Tucker RA, Wideroff L, Icenogle JP. A case-control study of human papillomavirus and cervical squamous intraepithelial lesions (SIL) in Harris County,

Texas: differences among racial/ethnic groups. Cad Saude Publica 1998;14(Suppl 3):149–59.
[22] Swan DC, Tucker RA, Tortolero-Luna G, Mitchell MF, Wideroff L, Unger ER, et al. Human papillomavirus (HPV) DNA copy number is dependent on grade of cervical disease and HPV type. J Clin Microbiol 1999;37:1030–4.
[23] Meissner J, Sequencing errors in reference HPV clones. In: Myers G, Baker C, Munger K, Sverdrup F, McBride A, Bernard HU, et al., editors. Human Papillomaviruses 1997: A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences. 3rd ed. Los Alamos: Theoretical Biology and Biophysics, Los Alamos National Laboratory; 1997. p. 110–23.
[24] Graham DA, Herrington CS. HPV-16 E2 gene disruption and sequence variation in CIN 3 lesions and invasive squamous cell carcinomas of the cervix: relation to numerical chromosome abnormalities. Mol Pathol 2000;53:201–6.