Mapping the integration sites E1-E2 of HPV-16 and HPV-18 as a tool to evaluate different stages of cervical disease progression

Carestiato FN[1], Amaro-Filho SM[2], Moreira MA[2], Barbosa RH[1], Furtado YL[3], Passos MRL[1], Cavalcanti SMB[1].

[1]Department of Microbiology and Parasitology, Universidade Federal Fluminense, Niterói, Rio de Janeiro, Brazil;[2] Genetics Program, Instituto Nacional de Câncer, Rio de Janeiro, Brazil; [3]Institute of Gynecology, Universidade Federal do Rio de Janeiro (UFRJ), Brazil. (Contact e-mail: silviacavalcanti@vm.uff.br)

INTRODUCTION

Failure of the immune system to clear persistent HPV infections can lead to the development of cervical cancer (CC) after several decades. In precancerous lesions, most HPV genomes persist in an episomal state whereas, in many high-grade lesions, genomes are found integrated into the host chromosome. Although no apparent hotspots have been identified, HPV integration often occurs near common fragile sites, which are naturally occurring regions of genomic instability. The majority of CC contain one or many copies of HPV, integrated more or less randomly into the host chromosome, with the viral integration site frequently lying within the regulatory E1 or E2 genes. Over one-half of HPV 16-positive cancers and most HPV 18-positive malignancies contain integrated HPV genomes, suggesting that integration may, in some cases, contribute to malignant progression.

Recent studies have suggested that an important step in HPV carcinogenesis may be the coexistence of HPV episomes with integrated copies. Expression of the E1 and E2 viral replication proteins from episomes can initiate DNA replication from integrated viral origins, resulting in their amplification and the induction of chromosomal abnormalities. Replication of integrated origins also results in the activation of DNA repair and recombination systems, which increases the likelihood of acquiring cellular mutations, increased genomic instability and, eventually, malignant progression.

The E1 protein possesses DNA helicase and ATPase activities that catalyze the unwinding of DNA and recruits cellular replication machinery to viral origins. E2 is a DNA-binding protein that helps to load E1 onto origins and tethers the viral DNA to the host chromosome during segregation. Increased expression of E1 and E2 occurs upon differentiation and is necessary for genome amplification.

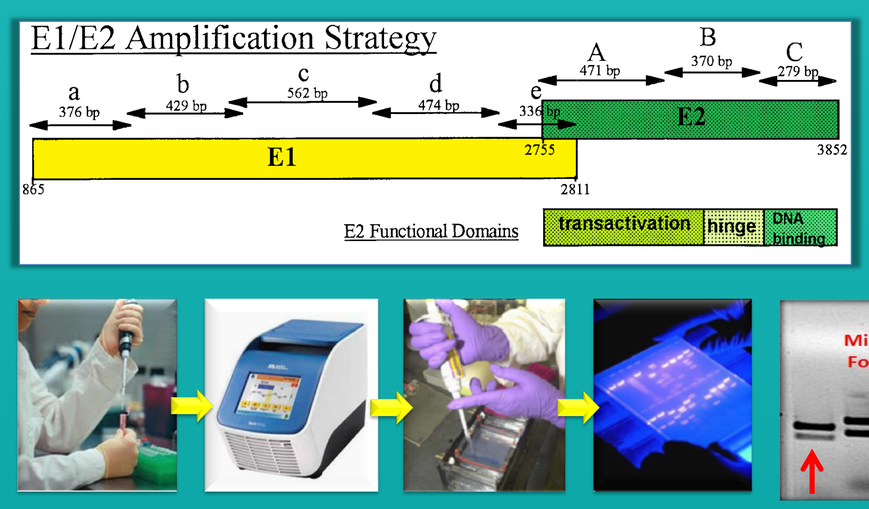
E2 proteins form complexes with E1 to initiate viral replication. E2 also regulates the expression of E6 and E7, and can exert suppressive or activating effects depending on the abundance of E2. Disruption of E2 ORF as a result of integration of viral genome into the host genome allows an uncontrolled overexpression of viral oncoproteins E6 and E7, which is a hallmark in CC. An elevation in the level of E6 and E7 is directly related to the increasing severity of neoplasia, and that the deregulated expression of these genes is directly responsible for the accumulation of genetic errors in the infected cell and the eventual integration of viral episomes into the host cell chromosome, which is seen in many CC.

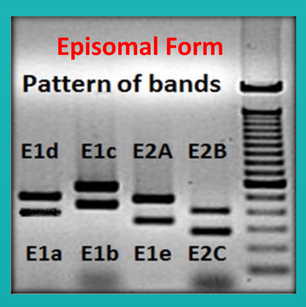
OBJECTIVE

To investigate the physical state (episomal or integrated) of HPV 16 and HPV 18 genome, using a PCR combining 10/11 primers pairs that covering the E1-E2 region, in samples from patients infected by HPV16 or 18, harboring cervical lesions in different stages of progression and cancer. In addition to establishing the mapping of rupture of the virus genome within this region investigated.

Primers			Sequências (5'-3')	Posição (nt)	Tamanho do Amplicon (pb)
HPV18	E1P1	FW	GGTGTGCATCCCAGCAGTAA	888-1403	515
		RW	GCCGCCACTACATACATTGC		
	E1P2	FW	GCGGCAATGTATGTAGTGGC	1400-1908	508
		RW	GCTGCAACACTACTTCGCAA		
	E1P3	FW	TCAACCACCAAAATTGCGAAGT	1877-2211	334
		RW	TCGTTTTTGGGCTCGCCTAT		
	E1P4	FW	GCAAACATTATAGGCGAGCCC	2181-2546	365
		RW	TGTCCAACACGTGGTCGTT		
	E1P5	FW	GGTGGCCATGTTAGATGATGC	2506-2895	389
		RW	GATTTTGTCCTGCAACGCACT		
	E1P6*	FW	GGTGTGCATCCCAGCAGTAA	863-2895	2032
		RW	GATTTTGTCCTGCAACGCACT		
	E2P1	FW	TCCAGATTAGATTTGCACGA	2786-3192	407
		RW	CAATTGTCTTTGTTGCCATC		
	E2P2	FW	ATACAAAACCGAGGATTGGA	3086-3388	303
		RW	ACTTCCCACGTACCTGTGTT		
	E2P3	FW	AACACAGGTACGTGGGAAGT	3369-3739	371
		RW	TTTCGCAATCTGTACCGTAA		
	E2P4	FW	GACCTGTCAACCCACTTCT	3598-3994	397
		RW	ACATGGCAGCACACATACAT		
	E2P5*	FW	TTAGATGATGCAACGACCAC	2495-4257	1763
		RW	CGGTGGGATACCATACTTTT		

Primers			Sequências (5'-3')	Posição (nt)	Tamanho do Amplicon (pb)
HPV16	E1a	FW	CCATGGCTGATCCTGCAG	863-1219	356
		RW	TCTCCTTTTTGCAGCTCT		
	E1b	FW	GACAGCGGGTATGGCAAT	1254-1663	409
		RW	CATTCCCCATGAACATGC		
	E1c	FW	AATAAATCAACGTGTTGCGATTGG	1548-2084	536
		RW	GTTTATAATGTCTACACATTGTTG		
	E1d	FW	GGATTGTGCAACAATGTG	2072-2527	455
		RW	TGGAGGGCATTTTAGTTG		
	E1e	FW	CAACTAAAATGCCCTCCA	2529-2845	316
		RW	CGCATGTGTTTCCAATAG		
	E1f*	FW	CCATGGCTGATCCTGCAG	863-2845	1982
		RW	CGCATGTGTTTCCAATAG		
	E2a	FW	CGAGGACAAGGAAAACGA	2738-3189	451
		RW	CTTGACCCTCTACCACAG		
	E2b	FW	GGTTTATATTATGTTCATGAAGG	3220-3599	379
		RW	TATGGGTGTAGTGTTACTATTACA		
	E2c	FW	GTAATAGTAACACTACACCCATA	3596-3853	257
		RW	GGATGCAGTATCAAGATTTG		
	E2d*	FW	TTAAGTTTGCACGAGGACGA	2721-3887	1167
		RW	CGCCAGTAATGTTGTGGATG		





Study design and participants - This cross-sectional study was performed with cervical samples (exfoliated cells and biopsies) representing the full spectrum of cervical pathology. Samples were collected from patients attending at Moncorvo Filho Hospital, and Instituto Nacional do Cancer (INCA), both in Rio de Janeiro, Brazil. This research was approved by both Ethical Committees. Our preliminary study involved 93 samples from patients infected by HPV16, harboring cervical lesions in different stages of progression, and 37 samples from HPV 18 CC.

DNA extraction and analysis - Samples were incubated in digestion buffer with proteinase K, then extracted with phenol: chloroform: isoamyl alcohol (25:24:1). DNA was precipitated with sodium acetate plus ice-cold ethanol, dry and suspended in Ultra-pure water.

MY09/11 consensual primers for HPV detection - Which amplify 450-bp DNA sequences in the L1 region, were used to detect generic HPV DNA via polymerase chain reaction (PCR). / **Specific genotyping** was performed by PCR amplification with primers from the E6 gene DNA sequences of HPV 16 and 18.

CASKI and HeLa lineage were used as positive control, and water as negative control in the PCR test. PCR products were analyzed on 1,5% agarose gel with ethidium bromide staining to visualize DNA under ultraviolet light, and their molecular weights 100-bp or 50-bp DNA ladder.

E1 *and* E2 *amplification* - For integration, were used a set of primers that overlap the regions, to cover the whole sequence. Each sample was tested for intact *E1/E2* DNA, and 9 separate amplification reactions to HPV16 and 10 to HPV18, with 3 microliters of the DNA extracted were used for each PCR. The MIX reaction and the thermal cycling parameters adopted were described by Vernon et al (1997) and Collins Constandinou-Williams et al (2009) . Controls positive and negative were used. Where a product did not amplify, this sample was repeated with all primers to verify absence of an *E1/E2* DNA fragment. If one or more *E1/E2* DNA fragments failed to amplify, the sample was recorded as having disrupted *E1/E2* DNA.

RESULTS

	HPV16 TOTAL (N=93)	HPV16 *-CC (N=43)	HPV16 CC (N=50)	HPV18 CC (N=37)
Episomal	N=26 (27,9%)	N=8 (18,6%)	N=18 (36%)	N=5 (13,5%)
Mixed	N=30 (32,2%)	N=16 (37,2%)	N=14 (28%)	N=8 (21,6%)
E1	22	12	10	-
E1/E2	8	4	4	-
E2	-	-	-	8
Integrated	N=37 (39,8%)	N=19 (44,2%)	N=18 (36%)	N=24 (64,9%)
E1	11	8	3	1
E1/E2	23	10	13	7
E2	3	1	2	16

Disruption of HPV 16	E1a			E1b			E1c			E1d			E1e			E2A			E2B			E2C		
	Pos	Mix	Neg	Pos	Mix	Neg	Pos	Mix	Neg	Pos	Mix	Neg	Pos	Mix	Neg	Pos	Mix	Neg	Pos	Mix	Neg	Pos	Mix	Neg
Normal-HSIL	8	17	18	33	9	1	36	7	-	33	8	2	33	9	1	35	6	2	26	10	7	25	8	10
Ca	26	9	15	44	1	5	44	1	5	44	1	5	35	3	12	38	0	12	41	2	7	38	1	11
Total	34	26	33	77	10	6	80	8	5	77	9	7	68	12	13	73	6	14	67	12	14	63	9	21

Disruption of HPV18	E1P1			E1P2			E1P3			E1P4			E1P5			E2P1			E2P2			E2P3			E2P4		
	Pos	Mix	Neg	Pos	Mix	Neg	Pos	Mix	Neg	Pos	Mix	Neg	Pos	Mix	Neg	Pos	Mix	Neg	Pos	Mix	Neg	Pos	Mix	Neg	Pos	Mix	Neg
Ca	35	-	2	35	-	2	32	1	2	28	1	6	28	-	7	25	-	10	24	-	11	5	-	30	25	2	8

DISCUSSION & CONCLUSION

Recent investigations suggest diverse and contradictory results related to integration pattern of viral in different stages of neoplastic progression. It has been that observed that viral integration mainly from specimens of high-grade lesions, whereas others found that viral integration takes place early during the course of infection detected in significant proportion of low-grade lesions. Additionally, has been postulated that viral integration is a consequence rather than a cause of chromosomal instability.

Frequently, in the preinvasive stages of CC, HPV is predominantly present in the episomal form, without changing the nucleotide sequence of DNA that regulates gene expression. Conversely, some degree of integration may be present in LSIL, suggesting that, LSIL may possibly represent a preinvasive lesion. Frequencies of HPV integrated genome in CC are variable (range from 30% to 100%) in different studies.

Our preliminary HPV16-results showed a reasonable number of patients with mixed forms (partially integrated): **32,2%(30/93).** It seems possible, that elimination of episomal forms may not be essential during tumorigenic transformation, and there could be some selective advantage of these forms in concomitant state for persistent and progressive HPV infection. Considering only integrated form, we have detected **39,8%(37/93),** this value is within the average of other studies. Finally, episomal forms were identified in **27,9%(26/93)** of the cases. However, in the cases of infection those of HSIL, this value drops to **18,6% (8/43)**, fact that should be further evaluated.

We identified predominantly integrated forms **64,9%(24/37)**, followed by mixed **21,6%(8/37)**, and episomal **13,5%(5/37) in** HPV18 infections, showing a different integration process.

Surprisingly, our data showed that E1 is more frequently broken **[91,9% (34/37)]** than E2 **[70,3%(26/37)]** in integrated forms HPV16+, and also partially broken in mixed forms, **E1 [100%(30/30)]** than E2 **[26,6%(8/30)]**, contrary to previous studies that described E2 as the gene region most commonly disrupted. Conversely, HPV 18 cases is consistent with the literature, with **95,8%(23/24)** E2 disrupted and **33,3%(8/24)** E1 at integrated forms, and in partially broken forms detected only E2 **[100% (8/8)].**

In the literature, has been demonstrated by sequence analysis that all sites of viral gene disruption occurred from E6 to L1 genes, more frequently in L1 gene (70%), followed by E1 gene (67%). Among possible explanations, there are: HPV integration into the host genomes does not appear to be an entirely random event but occurs preferentially at certain chromosomal locations, while HPV genomes could be disrupted at any gene, and cells with viral disruption at the L1 genes may be selected against during the clone selection process. And the use of cervical lesions at an early stage of cancer progression, possibly containing different cell clones.

Integration usually disrupts the E1 or E2 genes, potentially leading to a deregulation of viral gene expression. Among others, multiple-HPV infections, and different HPV variants, are also plausible examples of confounding variable.

Recent analyses have indicated that levels of E2 transcript and E2 protein expression in HPV-infected lesions do not correlate as closely as it has been previously thought, and that disruption of E2 protein expression is not always accompanied by disruption of the corresponding gene.

Aneuploidy can be detected in pre-malignant HPV-associated cervical lesions. These activities are limited to high-risk E6 and E7 proteins as none of them are seen in cells expressing their low-risk counterparts. Activation of the DNA damage response in cells containing both episomal and integrated forms of the viral genome could therefore result in chromosomal alterations and induction of genomic instability, which are likely to be important in the progression to malignancy.

Several molecular studies have suggested that the deregulation of E6/E7 expression, even in the absence of genome integration, is a critical event in determining neoplastic grade. Although it is not clear exactly how gene expression from the viral episome can become deregulated in early CIN. In these instances, deregulated gene expression may be driven by changes in cell signalling as can be brought about by hormonal changes, or epigenetic modifications such as viral DNA methylation, which may depend on the nature of the infected epithelial cell.

We herein describe a very specific methodology that can successfully map the HPV 16 and 18 genome fragile areas. Data are being analyzed in order to search for statistical correlation between integration and severity of the lesion but it has been observed that E1-E2 absences were suggestively frequent in HSIL and cancer. In a few cases, episomal forms were observed in cancer samples, suggesting additional biomarkers as involved in carcinogenesis.

It has been observed a higher frequency of intact forms for HPV16 than for HPV18 with predominance of disruption in E1/E2 together and E2 exclusive gene, respectively. The specific sites for disruption observed, suggest diversity in fragile areas between HPV types.

ACKNOWLEDGEMENTS