Gynecologic Oncology

Gynecologic Oncology 100 (2006) 372 – 378

# HPV16 E2 gene disruption and polymorphisms of E2 and LCR: Some significant associations with cervical cancer in Indian women

Bornali Bhattacharjee, Sharmila Sengupta *

*Human Genetics Unit, Indian Statistical Institute, 203, B.T. Road, Kolkata 700 108, India*

Received 15 April 2005
Available online 24 October 2005

**Abstract**

*Objectives.* We evaluated the status of the HPV16 E2 gene (disrupted or intact), nucleotide sequence alterations within intact E2 genes and LCR of HPV16 isolates in a group of CaCx cases (invasive squamous cell carcinomas, $n = 81$) and population controls (normal cervical scrapes, $n = 27$) from Indian women.

*Methods.* E2 disruption was detected by amplifying the entire E2 gene with single set of primers, while overlapping primers were used to determine if any particular region got selectively disrupted. Nucleotide variations in E2 and LCR were analyzed by PCR amplification followed by bi-directional sequencing. The associations between the viral factors and CaCx were analyzed using Fisher's Exact or Chi-squared test and interpreted as OR (95% CI) and $P$ values.

*Results.* E2 disruption was significantly higher among the cases [3.38 (1.07–10.72); $P = 0.02$], which was maximum in the region between nucleotides 3650 and 3872 (DNA-binding region). The European (E) variant was found to be the prevalent subgroup (87.76% among cases and 96.30% among the controls), and the remaining samples were Asian-American variants. Among the E subgroup, variation at position 7450 (T > C) within the E2-binding site-IV was found to be significantly higher among the E2 undisrupted cases (21/37; 56.76%), compared to controls (5/18; 27.78%) [3.41 (1.01–11.55); $P = 0.03$].

*Conclusions.* Besides HPV16 E2 disruption, LCR 7450T > C variation within undisrupted E2 of E subgroup appears to be a major factor contributing to the risk of CaCx development in Indian women. Furthermore, polymorphisms in the E2 gene of HPV16 may not be significant for disease risk.

*Keywords:* Cervical cancer; Human papillomavirus type 16; E2 disruption; LCR variation; Indian women

## Introduction

Human papillomavirus (HPV) infection has been considered to be the major etiological agent for the development of cervical carcinoma (CaCx). HPV DNA is found in 99.7% of CaCx cases, of which 50% are HPV16 DNA-positive [1–3]. The viral genes that contribute to transformation of cells are E6 and E7. The E2 viral protein regulates transcription of these genes through its binding to sites adjacent to the promoter responsible for this transcription. Thus, an increase in expression of E2 results in repression of E6/E7 expression [4], while loss of E2 function most commonly by integration into the viral genome confirms an upregulation of E6/E7 proteins [5,6].

The viral genome contains a region spanning 1024 base pair termed as the long control region (LCR) that regulates replication and transcription. The entire LCR is divided into three functionally distinct segments, viz., the 5′, the central and the 3′ segment. The 5′ region of the LCR has a high A + T content (up to 85%) and contains a negative regulatory element acting at the level of late mRNA stability [7] and a nuclear matrix attachment region that represses viral oncoprotein expression [8,9]. The central segment (400 base pair) functions as an epithelial-specific transcriptional enhancer [10–12]. Multiple transcription factors have been found to bind in vitro to more than 20 sites within the HPV16 enhancer sequence which include nuclear factor-1 (NF1, NF-IL6), Jun–Fos

heterodimers (AP-1, AP-2), transcriptional enhancer factor-1 (TEF-1), Octamer-1 (Oct-1), glucocorticoid responsive element (GRE), Yin Yang 1 (YY1) and papillomavirus silencing motifs [13–18]. The 3′ segment contains the origin of replication and the E6 and E7 promoter, P97 [19,20]. Thus, it is a complex network of events that guide either to cause positive or negative effect.

A number of factors could be involved in modifying processes that disrupt the natural progression of HPV infection towards CaCx development. One such is the disruption of the E2 gene, often observed in CaCx but not in premalignant lesions [21–23]. As a result of disruption, the E2 protein in its truncated form is unable to displace the cellular transcription factors, Sp1 or TFIID, from their cognate binding sites next to the P97 promoter, resulting in the increased expression of the E6 and E7 genes [5,22]. This phenomenon is often attributed to disease progression from CIN to invasive CaCx [10,11,24] and to poor disease-free survival.

HPV16, harboring intact E2 in the episomal form, are often found to coexist with disrupted forms in CaCx [25], but in relatively low frequencies. In fact, a few studies have also reported an increased frequency of intact episomal HPV16 DNA in CaCx [26], reflecting a lack of repression of E6 and E7 genes by intact E2. Taken together, such findings indicate that E2 gene disruption may not be a prerequisite for CaCx development. Among alternative mechanisms of enhanced viral oncogene expression, altered E2 functions, as a result of variations in the E2 ORF, have been reported by a few studies [27–30], including one from India [24]. Some studies have also observed elevated levels of E6 and E7 in primary or metastasizing tumors carrying intact E2, due to deletions or point mutations affecting one or more binding sites of the transcription factor YY1 [31,32] in the viral LCR.

HPV16 has been found to be the most prevalent high risk HPV type in CaCx in India [33] as well as in HPV-infected individuals of the normal population [33,34]. In this study, we evaluated the status of the E2 gene (disrupted or intact) in a group of CaCx cases (invasive squamous cell carcinomas) and population controls (cytologically normal) harboring HPV16 DNA. Furthermore, we also determined the nucleotide sequence alterations within intact E2 genes and the corresponding LCR of the HPV16 isolates. Our objective was to determine the viral factors that were significantly associated with HPV16-related CaCx in Indian women.

## Materials and methods

### *Sample collection and DNA isolation*

The malignant cervical biopsy tissues used for this study were all histopathologically confirmed as invasive squamous cell carcinomas. These were derived from married subjects aged 27–80 years (median age: 50 years), attending a cancer referral hospital (Cancer Centre Welfare Home and Research Institute, South 24 Parganas, West Bengal, India). All of these malignant subjects were clinically diagnosed to have tumors of stage III and above, as per FIGO classification. The control samples were normal cervical scrapes confirmed by Pap smear test. These were derived from married women aged 16–80 years (median age: 32 years), with no previous history of cervical dysplasia/malignancy. These women were attending a Reproductive and Child Health Clinic (Child in Need Institute, South 24 Parganas, West Bengal, India) for routine contraception and reproductive healthcare counseling. All samples (biopsy tissues and cervical scrapes) were collected from the participants with informed consent approved by the institutional ethical committee for human experimentation.

DNA was isolated following the method of Miller et al. [35] with modifications. Briefly, cell pellets were collected and suspended in lysis buffer (10 mM Tris, 400 mM NaCl and 2 mM EDTA, pH 8.0 and 10% SDS). Lysates were digested overnight in the presence of 5 mg/ml Proteinase K containing 1% SDS and 2 mM EDTA, at 37°C. After digestion, 6 M NaCl was added, and the DNA was ethanol-precipitated.

### *HPV16 screening*

Isolated DNA was quantitated, and 200 ng of DNA was used for screening HPV16 by PCR amplification using E6-specific primers and RFLP [36], with modification [34]. CaSki and SiHa DNA were used as positive controls and Hela DNA as negative, along with water blank. DNA from a total of 81 HPV16-positive malignant samples and 27 HPV16-positive normal samples were used subsequently for the study.

### *E2 amplification by PCR and confirmation of nucleotide variations by RFLP and sequencing*

HPV16-positive samples were subjected to PCR amplification of the E2 ORF using a single set of primers (W1/W2) as indicated in Table 1. A 1026 bp fragment was amplified by 35 cycles of PCR [37]. The reaction mixture (20 μl) contained 200 ng target DNA, 100 ng primers, 50 mM KCl, 1.5 mM $MgCl_2$, 50 mM Tris–HCl (pH 8.3), 50 mM $(NH_4)_2SO_4$, 25 μM of each dNTP and 0.5 U of thermostable DNA Polymerase (Roche Diagnostics). The presence of E2 amplicon in agarose gel confirmed an intact E2, while lack of specific band indicated disruption. CaSki cell DNA, which harbors an intact E2, was used as a positive control.

Those samples that gave amplification with the primer mentioned above were further amplified with three sets of overlapping primer pairs [A1/A2, B1/B2 and C1/C2] [27] in separate reactions. The reaction mixture (20 μl) contained 100 ng target DNA, 20 ng primers, 50 mM KCl, 1 mM $MgCl_2$, 50 mM Tris–HCl (pH 8.3), 50 mM $(NH_4)_2SO_4$, 25 μM of each dNTP and 1 U of thermostable DNA Polymerase (Roche Diagnostics). The primer sequences, PCR conditions and amplicon sizes are summarized in Table 1. The amplicons were confirmed by electrophoresis in 1.5–2% agarose gels.

Bi-directional sequencing was carried out for the E2 1026 bp fragment using the first set of primers, i.e. W1/W2. Twenty nanograms of each PCR product was treated with 0.6 U Exonuclease I and Shrimp Alkaline Phosphatase at 37°C for an hour to remove the excess dNTPs and unused primers with subsequent enzyme inactivation at 80°C for 20 min and subjected to sequencing. Sequencing was done in an ABI Prism™3100 automated sequencer using dye terminator chemistry.

### *LCR amplification by PCR and sequencing*

HPV16-positive samples were further subjected to amplification of LCR. The reaction mixture (20 μl) contained 100 ng target DNA, 10 ng primers, 50 mM KCl, 1.5 mM $MgCl_2$, 50 mM Tris–HCl (pH 8.3), 50 mM $(NH_4)_2SO_4$, 25 μM of each dNTP and 1 U of thermostable DNA Polymerase (Roche Diagnostics). The primer sequences [27], PCR conditions and amplicon sizes are summarized in Table 1. The amplicons were confirmed by electrophoresis in 2% agarose gels. Bi-directional sequencing was carried out using the same set of primers.

### *Sequence analysis of E2 and LCR*

ABI trace files thus generated were analyzed using the PHRED software (http://www.mbt.washington.edu/phrap.docs/phred.html), which assigns quality scores to each base. The PHRED outputs for the given amplicons were then aligned using PHRAP software. The resulting assemblies were viewed using CONSED, which allows identification of sequence differences as well

Table 1
PCR conditions with different primers and their position on the E2 and LCR region

| E2 primers (5′–3′) | Positions | Product | Conditions |
|---|---|---|---|
| W1: ATG AAA ATG ATA GTA CAG AC | 2811 | 1026 bp | Denaturation: 95°C/1 m |
| W2: CCA GTA GAC ACT GTA ATA G | 3837 | | Annealing: 50°C/2 m |
| | | | Extension: 72°C/1 m 30 s (35 cycles) |
| A1: ATG AAA ATG ATA GTA CAG AC | 2810 | 661 bp | Denaturation: 95°C/1 m |
| A2: TGG ATA GTC TGT GTT TCT TCG | 3471 | | Annealing: 55°C/45 s |
| | | | Extension: 72°C/30 s (35 cycles) |
| B1: CGA AGA AAC ACA CTA CAC CCA TA | 3448 | 203 bp | Denaturation: 94°C/1 m |
| B2: TAA AGT ATT AGC ATC ACC TT (32) | 3649 | | Annealing: 50°C/1 m |
| | | | Extension: 72°C/1 m (35 cycles) |
| C1: GTA ATA GTA ACA CTA CAC CCA TA | 3596 | 277 bp | Denaturation: 94°C/1 m |
| C2: GGA TGC AGT ATC AAG ATT TGT (32) | 3872 | | Annealing: 50°C/1 m |
| | | | Extension: 72°C/1 m 30 s (35 cycles) |
| LCR primers (5′–3′) | | | |
| L1: GCT TGT GTA ACT ATT GTG TCA | 7289 | 731 bp | Denaturation: 94°C/15 s |
| L2: GTC CAG AAA CAT TGC AGT TCT (33) | 114 | | Annealing: 55°C/45 s |
| | | | Extension: 72°C/1 m (35 cycles) |

as access to the individual chromatograms to scrutinize each putative variant.

### *Statistical analysis*

The associations of the various viral factors with CaCx were determined using Fisher's Exact Test or Chi-squared test as appropriate, and the risk for associations was determined using odds ratio (OR) and 95% confidence interval (CI). A $P$ value of less than 0.05 was considered as statistically significant.

## Results

### *HPV16 E2 gene disruption*

Thirty of the eighty-one cases (37.04%) and four of the twenty-seven controls (14.81%) showed E2 disruption. Each sample was amplified twice, using the W1/W2 primers, to confirm the results. Thus, E2 disruption was significantly higher among the cases [$P$ = 0.018; OR (95% CI) = 3.38 (1.07–10.72)]. Subsequently, each portion of E2 gene was amplified individually using the three sets of overlapping primers indicated in Table 1 to map the exact region that got most frequently disrupted. The DNA-binding region showed maximum disruption of 86.67% followed by 56.67% of region encoding the hinge and 44% of transactivation domain, respectively. The primers C1 and C2 (Table 1) covered the region encoding the DNA-binding domain, exclusively, showing lack of amplification in 86.67% (26 out of 30) of the malignant cases against 50% (2 of the 4) controls. All of these isolates showed disruption while using the primers W1/W2. The prevalence of intact episomal E2 DNA was 63% (51 out of 81) among the malignant cases.

### *HPV16 E2 sequence variations*

The numbering of nucleotide positions considered in this study is based on the revised sequence of HPV16 (HPV16R) available in the HPV16 Sequence Database (Los Alamos National Laboratory), and the regions described are in accordance with that of Veress et al. [28]. The results of E2 sequence analysis (spanning nt 2811–3873 using primers W1/W2) of 49 cases and 23 controls are summarized in Table 2. Two samples could not be sequenced from the malignant group. Polymorphisms were identified at a few new positions such as 2988, 3007, 3025 3058, 3161, 3187, 3236, 3253 and 3605. Of these, variations at positions 3007, 3058, 3236 and 3605 were absent among the cases or prominent among the controls. A total of 12 variations (8 novel and 4 previously reported) were found in the transactivation domain. In this region, nucleotide changes at 3058, 3187 and 3253 (all novel variants) failed to show amino acid changes, while the remaining 9 out of 12 variations resulted in amino acid changes as indicated in Table 2. There were 2 nucleotide variations in the hinge and 5 in the DNA-binding regions, respectively, most of which were reported earlier. In the DNA-binding region, the novel variation at 3605 resulted in amino acid change, while those at 3538 and 3964 (also reported previously by others) failed to do so.

E2 sequences of 3684C > A, together with 3694T (prototype), are characteristic of the European variant of HPV16 found in SiHa and CaSki cells, while 3684C > A and 3694T > A together denote Asian-American variants (AA) [27,29]. Based on such classification, 6 samples from among the cases (6/49; 12.24%) and 1 of the controls (1/27; 3.7%) were identified as Asian-American (AA) variants compared to the sequence data of HPV16R. Therefore, remaining 43 cases (87.76%) and 26 controls (96.30%) were all European (E) variants of HPV16. Furthermore, the AA variants showed nucleotide changes involving all three domains of E2, such as the transactivation (3161C > T, 3181A > C, 3182G > A, 3249G > A, and 3224T > A/C), hinge (3516, 3517CT > AA/AC) and DNA binding (3538A > C, 3566T > G, 3684C > A and 3694T > A).

### *HPV16 sequence variations in the LCR*

LCR sequencing of 42 out of the 49 malignant cases (E2 undisrupted and sequenced) and 19 out of 23 cytologically

Table 2
Variations in DNA and amino acid sequences of HPV16 E2 variants compared to the reference sequence

| Polymorphism | | | | | Malignant ($n$ = 49) | Non-malignant ($n$ = 23) |
|---|---|---|---|---|---|---|
| Domain | Nucleotide | | Amino acid residue | | | |
| | Number | Change | Number | Change | | |
| Transactivation | 2988 | C > T | 78 | Thr > Met | 2 | 1 |
| | 3007 | C > T | 84 | Asp > Asp | 0 | 1 |
| | 3025 | A > C | 90 | Glu > Asp | 1 | 0 |
| | 3058 | G > A | 101 | Val > Val | 1 | 4 |
| | 3161 | C > T | 136 | His > Tyr | 4 | 1 |
| | 3181 | A > C | 142 | Glu > Asp | 2 | 0 |
| | 3182 | G > A | 143 | Ala > Thr | 6 | 1 |
| | 3187 | A > C | 144 | * | 2 | 0 |
| | 3224 | T > A | 157 | Leu > Ile | 3 | 0 |
| | | T > C | 157 | Leu > Leu | 1 | 1 |
| | 3236 | C > G | 161 | His > Asp | 0 | 2 |
| | 3249 | G > A | 165 | Arg > Gln | 6 | 1 |
| | 3253 | A > G | 166 | * | 2 | 0 |
| | 3516,3517 | CT > AA/AC | 254 | Thr > Lys/Asn | 4/2 | 1/0 |
| Hinge | 3538 | A > C | 261 | * | 6 | 1 |
| DNA binding | 3566 | T > G | 271 | Phe > Val | 3 | 1 |
| | 3605 | A > G | 286 | Gln > Asp | 0 | 2 |
| | 3684 | C > A | 310 | Thr > Lys | 7 | 4 |
| | 3694 | T > A | 313 | * | 7 | 1 |

* No amino acid change.

normal controls could be carried out, while the remaining samples failed to amplify. Of the malignant cases, 37 were E and 5 AA variants, while among the controls, 18 belonged to the E and 1 to the AA lineage, respectively. All the samples were amplified twice to confirm the results, and it was noted that those which failed to amplify for LCR did amplify with the E6 primers (used for HPV screening), confirming the integrity of the samples. Further experiments need to be done to confirm LCR status in case of those samples that failed to amplify with LCR primer.

The sequence variations recorded among the various cases and controls are depicted in Table 3. Among the E lineage, variation at position 7450T > C within the E2-binding site-IV was found to be significantly higher among the cases (21/37; 56.76%), compared to controls (5/18; 27.78%) [$P$ = 0.03; OR (95% CI) = 3.41 (1.01–11.55)]. Among the samples that harbored this variation, 71.42% (15/21) malignant cases in contrast to 40% (2/5) of controls had no other variation involving either LCR or E2. It was also noted that only among the samples that harbored 7450T (like the HPV16R) a variation of 7521G > A was found to be higher among the cases (50%, 8/16) compared to the controls (38.46%; 5/13). However, this variation was not statistically significant. Besides such variations, a few other changes were also noted. Such changes were either present exclusively among the cases (7394 C > T within the GRE/1 binding site, 7568T > G, 7703G > T and 7713T > G) or were proportionately higher among the controls (7521G > A, 7714T > G and 7775T > A), though not statistically significant.

The AA variants showed nucleotide variations at a few identified transcription factor binding sites such as 7394C > T and 7395C > T within the GRE/1, 7485A > C and 7489G > A within GRE/2, 7689C > A and 7743T > G within TEF-1, 7729A > C, 7764C > T, 7786C > T within YY14 and 7868G > A within E2-binding site II. There were also differences in the sets of polymorphisms within these 6 AA variants, based on which, they could further be divided into sub-groups. Among the 5 AA case samples, 2 were found to belong to subgroup, ‘a’ or ‘b’ based on polymorphism at 7743 T > G. The remaining 3 cases along with the single control were found to belong to subgroup ‘c’.

## Discussion

In this study, we analyzed E2 disruption as well as E2 and LCR sequence variations in HPV16 isolates from CaCx cases and cytologically normal subjects to identify key viral factors associated with disease development. Our data indicate that 37.04% of the HPV16-positive CaCx against 14.81% of controls harbor disrupted E2, which showed significant association with disease development. Such disruption was most prominent in the region encoding the DNA-binding domain of E2 protein (nucleotides 3596–3872). It was also noted that 63% of cases harbored undisrupted or intact E2 genes. Among such cases, several variations were noted in the E2 sequences, chiefly within the region encoding the transactivating domain, but none showed statistically significant association with CaCx. However, a variation of 7450T > C in the LCR of HPV16 within the E2-binding site-IV was noted, which showed significant association with CaCx. This variation was restricted to the HPV16 isolates having intact E2 and belonging to the E lineage, which was most predominant in this region (88% among cases) compared to the AA lineage found in 12.24% of cases.

Our findings reinforced earlier reports [21–23] on association of E2 disruption with CaCx, wherein such cases harbored

Table 3
Variations in HPV16-LCR sequences compared to the reference sequence

| Nucleotide | | Transcription factor binding site | Malignant ($n$ = 42) | | Non-malignant ($n$ = 19) | |
|---|---|---|---|---|---|---|
| Number | Change | | European variant ($n$ = 37) | Asian-American variant ($n$ = 5) | European variant ($n$ = 18) | Asian-American variant ($n$ = 1) |
| 7394 | C > T | GRE/1 | 2 | 2 | 0 | 0 |
| 7395 | C > T | GRE/1 | 0 | 2 | 0 | 0 |
| 7450 | T > C | – | 21 | 0 | 5 | 0 |
| 7482 | A > G | – | 1 | 0 | 0 | 0 |
| 7485 | A > C | GRE/2 | 0 | 5 | 0 | 1 |
| 7489 | G > A | GRE/2 | 0 | 5 | 0 | 1 |
| 7497 | Int-del T | – | 1 | 0 | 1 | 0 |
| 7521 | G > A | – | 8 | 0 | 5 | 0 |
| 7550 | A > G | – | 0 | 0 | 2 | 0 |
| 7568 | T > G | – | 3 | 0 | 0 | 0 |
| 7669 | C > T | – | 0 | 5 | 0 | 1 |
| 7689 | C > A | TEF-1 | 0 | 5 | 0 | 1 |
| 7703 | G > T | – | 1 | 0 | 0 | 0 |
| 7713 | T > G | – | 1 | 0 | 0 | 0 |
| 7714 | T > G | – | 5 | 0 | 5 | 0 |
| 7729 | A > C | – | 0 | 5 | 0 | 1 |
| 7743 | T > G | TEF-1 | 0 | 2 | 0 | 0 |
| 7764 | C > T | – | 0 | 5 | 0 | 1 |
| 7775 | T > A | – | 0 | 0 | 1 | 0 |
| 7786 | C > T | YY14 | 0 | 5 | 0 | 1 |
| 7826 | G > A | – | 0 | 1 | 0 | 0 |
| 7868 | G > A | E2-II | 0 | 3 | 0 | 0 |

37.04% disruption in contrast to 14.81% in controls. The region between the nucleotides 3596–3872, which encodes a part of the DNA-binding region of E2 protein was found to be more prone to disruption, while the region encoding the transactivation domain of E2 was more intact. It has been identified that Th cell responses against the HPV16 E2 C-terminal domain occur at the time of virus clearance [39,40]. Therefore, frequent disruption of the C-terminal domain of E2, noted for the first time in this study, could possibly be related to the process of attenuation of viral clearance or promotion of viral persistence.

There were variations in E2 sequences among the cases as well as controls, but no significant difference was recorded in view of the small number of samples showing such variations. However, variations were most prominent in the regions encoding the transactivation domain, compared to the hinge and DNA-binding regions. It is known that E2 protein harbors linear epitopes of which one is present in the transactivation domain between amino acids 121–140 [41]. The variation at nucleotide position 3161C > T noted in this study, which results in an amino acid change (histidine to tyrosine), falls within this region. Furthermore, the region (amino acids 18–41) utilized by E2 for association with E1 to promote viral replication [42] was found to be devoid of any variations. In other words, the replication efficiency of HPV16 could possibly remain unaffected in such cases harboring intact E2. Similar observation was also made in another study from the Southern part of India [24]. Several novel polymorphisms with or without amino-acid changes were also noted in this region in our study. However, the functional significance of such polymorphisms needs to be tested.

A lack of prominence of 3684C > A (T310K) change within the DNA-binding domain of E2 was noted in our study (14.28% among the cases compared to 17.39% among the controls). This was contradictory to that reported by Giannoudis et al. [27] and Casas et al. [29] but was in line with that of Veress et al. [33], who also reported lack of association of various E2 polymorphisms, including 3684C > A, based on functional assays. In fact, the variation of 3694T > A in the DNA-binding region of E2 does not result in an amino-acid change. This was also noted in our study, and we found this polymorphism to be proportionately higher among the cases (14.28%) compared to the controls (4.34%), though not statistically significant. The implication of such finding remains to be elucidated, although some studies have pointed out that this variation along with 3684C > A could be related to conformational alterations of DNA structure. Overall, our study, like that of Veress et al. [28], further emphasizes the fact that sequence variations in the HPV16 E2 region may not be a major mechanism related to enhanced expression of E6 and E7 oncoproteins.

The LCR region harbors binding sites for many cellular transcription factors along with E2-binding sites, and functional studies have also reasoned out that variations in the LCR do contribute to transformation capacity of high risk HPV [42]. A myriad of cellular transcription factors along with HPV16 E2 modulates the expression of viral oncoproteins. The expression level of each of these proteins as well as polymorphisms in their binding sites sets the threshold for virulence. One such probable switch that has been identified in our study is the variation at nucleotide 7450T > C, two nucleotides upstream of the core E2-binding site, at the 5′ end

of the LCR spanning nucleotides 7453–7464 [43]. Significant association of this variation was noted with CaCx in our study. Furthermore, a majority of the cases (71.42%) were found to harbor this change solely in absence of other LCR or E2 variations. Therefore, this points to the fact that this variation in the E2-binding site could probably result in loss of transcriptional repression of E2 enhancing the virulence of HPV16. However, our finding was contradictory to an earlier study [44], which identified increased occurrence of this variation in asymptomatic controls. This former study related the observation to reduced virulence attributable to altered binding of E2 at the site proximal to 5′ end of the LCR and reduced transcriptional activation. But, in support of our findings, another study has also identified that this E2-binding site can mediate E2's activation of transcription, especially under conditions where the other E2-binding sites are disrupted [45].

It has been reported that the variation at nucleotide 7521G > A is a commonly observed change noted in 80% of the HPV16 isolates sampled worldwide [38]. The study by Burk et al. [46] also identified this variation in 61% of squamous cell carcinomas sampled worldwide. In our study, the occurrence of the variation at 7521G > A was only 21.62% among the cases (SCC), which was again restricted to the E variants lacking 7540T > C variation in those isolates harboring intact E2. Therefore, taken together, our findings of the LCR variations could possibly be lineage-specific or related to host genetic factors. Such findings thus provide a framework for further studies of the functional effects of these polymorphisms to identify their biological relevance.

In this study, we considered E2 and LCR sequence variations to confirm the lineages of HPV16 isolates. Therefore, the data on those samples that had undisrupted E2 were considered. It was noted that prevalence of AA variants was 12.24% among the malignant cases compared to 3.7% among the controls. Probably, this is the first study to report the prevalence of such HPV16 variants in malignant tumors as well as population controls from India. Earlier, an international study of invasive cervical cancer by Yamada et al. [47] identified that AA variants clustered in Central and South America and in Europe, limited only in Spain.

There is a possibility that the product of the E2 gene from an episomal form of HPV could suppress the transcription activation potential of HPV with disrupted E2 in cases harboring mixed infections. This study therefore throws some light, like a few other studies [24,27–30], on the alternative mechanism of loss of E2 repression that could lead to sustained E6/E7 expression under conditions of mixed infections or in cases harboring episomal HPV16. We therefore refrained from resolving the mixed status of the viral genome, which involves more advanced techniques such as real time PCR.

In summary, our study is the first of its kind addressing both the E2 and LCR sequence variations in HPV16 isolates prevalent among Indian women. We identified that, besides E2 disruption, LCR 7450T > C variation within the E2-binding site towards the 5′ end of LCR in samples harboring undisrupted E2 was significantly associated with HPV16-positive CaCx, where the virus belonged to the E lineage. This polymorphism thus appears to have strong biological plausibility towards facilitating cell transformation, probably through enhanced viral oncogene expression concomitant with loss of repressor activity of E2, which needs to be functionally tested. Furthermore, polymorphisms in the E2 gene of HPV16, some of which were functionally relevant particularly for immune evasion, did not appear to be significant for disease risk. A study involving larger sample sizes of HPV16-positive cases and controls would probably help in confirming this. Such studies are ongoing as well as identification of the role of other factors, such as methylation of HPV16 LCR in modulating the oncogenic potential of episomal HPV16. Nevertheless, the finding of an overrepresentation of 7450C > T LCR variation of HPV16 E-isolates harboring intact E2 in malignant samples could be utilized for singling out HPV16-positive women of this population for closer monitoring.

## Acknowledgments

We thank Dr. Chandralekha Duttagupta (formerly of Biochemistry Unit, Indian Statistical Institute, Kolkata, India) for her constant inspiration and untiring support through this work; Professor Partha P Majumder and Dr. Sourabh Ghosh of Human Genetics Unit, Indian Statistical Institute, Kolkata, India, for their comments on the manuscript and for providing guidelines for statistical analysis of the data; Cancer Centre Welfare Home and Research Institute (Thakurpukur, South 24 Parganas, West Bengal, India), Child in Need Institute (Pailan, South 24 Parganas, West Bengal, India), for their support in sample collection; and all members of Human Genetics Unit, Indian Statistical Institute, Kolkata, India for technical support during the work.

## References

[1] Bosch FX, Manos MM, Munoz N, et al. Prevalence of human papillomavirus in cervical cancer: a world wide perspective. International biological study on cervical cancer (IBSCC) Study Group. J Natl Cancer Inst 1995;87:796–802.

[2] Das BC, Sharma JK, Gopalkrishna V, Luthra UK. Analysis by polymerase chain reaction of the physical state of human papillomavirus type 16 in cervical preneoplastic and neoplastic lesions. J Gen Virol 1992;73:2327–36.

[3] Londesborough P, Ho L, Terry G, Cuzick J, Wheeler CM, 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.

[4] Tan S-H, Leong LE-C, Walker PA, Bernard H-U. The human papillomavirus type 16 E2 transcription factor binds with low cooperativity to two flanking sites and represses the E6 promoter through displacement of Sp1 and TFIID. J Virol 1994;68:6411–20.

[5] Jeon S, Allen-Hoffmann BL, Lambert PF. Integration of human papillomavirus type 16 into the human genome correlates with selective growth advantage of cells. J Virol 1995;69:2989–97.

[6] Jeon S, Lambert PF. Integration of human papillomavirus type 16 DNA into the human genome leads to increased stability of E6 and E7 mRNAs: implications for cervical carcinogenesis. Proc Natl Acad Sci U S A 1995;92:1654–8.

[7] Mittal R, Pater A, Pater MM. Multiple human papillomavirus type 16 glucocorticoid response elements functional for transformation, transient expression, and DNA protein interactions. J Virol 1993;67:5656–9.
[8] Stunkel W, Bernard HU. The chromatin structure of the long control region of human papillomavirus type 16 represses viral oncoprotein expression. J Virol 1999;73:1918–30.
[9] Tan SH, Bartsch D, Schwarz E, Bernard HU. Nuclear matrix attachment regions of human papillomavirus type 16 point toward conservation of these genomic elements in all genital papillomaviruses. J Virol 1998;72:3610–22.
[10] Cripe TP, Haugen TH, Turk JP, et al. Transcriptional regulation of the human papillomavirus-16 E6/E7 promoter by a keratinocyte-dependent enhancer, and by viral E2 trans-activator and repressor gene products: implications for cervical carcinogenesis. EMBO J 1987;6:3745–53.
[11] Gloss B, Chong T, Bernard H-U. Numerous nuclear proteins bind the long control region of human papillomavirus type 16: a subset of 6 of 23 Dnase I-protected segments coincides with the location of the cell-type-specific enhancer. J Virol 1989;63:1142–52.
[12] Sibbet GJ, Cuthill S, Campo MS. The enhancer in the long control region of human papillomavirus type 16 is up regulated by PEF-1 and down regulated by Oct-1. J Virol 1995;69:4006–11.
[13] Chan WK, Chong T, Bernard H-U, Klock G. Transcription of the transforming genes of the oncogenic human papillomavirus type 16 is stimulated by tumor promoters through AP1 binding sites. Nucleic Acids Res 1990;18:763–9.
[14] Ishiji T, Lace M, Parkkinnen S, et al. Transcriptional enhancer factor (TEF)-1 and its cell-specific co-activator activate human papillomavirus-16 E6 and E7 oncogene transcription in keratinocytes and cervical carcinoma cells. EMBO J 1992;11:2271–81.
[15] Khare S, Pater MM, Tang SC, Pater A. Effect of glucocorticoid hormones on viral gene expression, growth, and dysplastic differentiation in HPV16-immortalized ectocervical cells. Exp Cell Res 1997;232:353–60.
[16] O'Connor MJ, Bernard H-U. Oct-1 activates the epithelial-specific enhancer of human papillomavirus type 16 via a synergistic interaction with NF1 at a conserved composite regulatory element. Virology 1995;207:77–88.
[17] O'Connor MJ, Tan SH, Tan CH, Bernard H-U. YY1 represses human papillomavirus type 16 transcription by quenching AP-1 activity. J Virol 1996;70:6529–39.
[18] O'Connor MJ, Stunkel W, Zimmerman H, Koh CH, Bernard H-U. A novel YY1-independent silencer represses the activity of human papillomavirus type 16 enhancer. J Virol 1998;72:10083–92.
[19] Dürst M, Glitz D, Schneider A, zur Hausen H. Human papillomavirus type 16 (HPV 16) gene expression and DNA replication in cervical neoplasia: analysis by in situ hybridization. Virology 1992;189:132–40.
[20] Smotkin D, Wettstein FO. Transcription of human papillomavirus type 16 early genes in cervical cancer and a cervical cancer derived cell line and identification of the E7 protein. Proc Natl Acad Sci U S A 1986;83:4684–9.
[21] Matsukura T, Kanda T, Furuno A, Yoshikawa H, Kawana T, Howley PM. Cloning of monomeric human papillomavirus type 16 DNA integrated within cell DNA from cervical carcinoma. J Virol 1986;58:979–82.
[22] 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.
[23] Schwarz E, Freese UK, Gissmann L, Mayer W, Roggenbuck B, Stremlau A, zur Hausen H. Structure and transcription of human papillomavirus sequences in cervical carcinoma cells. Nature 1985;314:111–4.
[24] Narayanan S, Abraham P, Peedicayil A, et al. E2 sequence variations of HPV 16 among patients with cervical neoplasia seen in the Indian subcontinent. Gynecol Oncol 2004;95:363–9.
[25] Matsukura T, Koi S, Sugase M. Both episomal and integrated forms of human papillomavirus type 16 are involved in invasive cervical cancers. Virology 1989;172:63–72.
[26] Cullen AP, Reid R, Lorincz AT. Analysis of the physical state of different human papillomavirus DNAs in intraepithelial and invasive neoplasm. J Virol 1991;65:606–12.
[27] Giannoudis A, Van Duin M, Snijders PJF, Herrington CS. Variation in the E2-binding domain of HPV 16 is associated with high-grade squamous intraepithelial lesions of the cervix. Br J Cancer 2001;84:1058–63.
[28] Veress G, Szarka K, Dong X-P, 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.
[29] Casas L, Galvan SC, Ordonez RM, et al. 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.
[30] Graham DA, Herrington CS. HPV16 E2 gene disruption and sequence variation in CIN 3 lesions and invasive squamous cell carcinomas of the cervix: relation to numerical chromosome abnormalities. J Clin Pathol Mol Pathol 2000;53:201–6.
[31] Dong XP, Stubenraunch F, Beyer-Finler E, Pfister H. Prevalence of deletions of YYI-binding sites in episomal HPV 16 DNA from cervical cancers. Int J Cancer 1994;58:803–8.
[32] May M, Dong XP, Beyer-Finkler F, Stubenraunch F, Fuchs PG, Pfister H. The E6/E7 promoter of extrachromosomal HPV16 DNA in cervical cancer escapes from cellular repression by mutation of target sequences for YY1. EMBO J 1994;13:1460–6.
[33] Das BC, Gopalkrishna V, Hedau S, Katiyar S. Cancer of the uterine cervix and human papillomavirus infection. Curr Sci 2000;78:52–63.
[34] Duttagupta C, Sengupta S, Roy M. Oncogenic Human papillomavirus (HPV) infection and uterine cervical cancer: a screening strategy in the perspective of rural India. Eur J Cancer Prev 2002;11:447–56.
[35] Miller SA, Dykes D, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 1988;16:1215.
[36] Haraf DJ, Nodzenski E, Brachman D, et al. Human papillomavirus and p53 in head and neck cancer: clinical correlates and survival. Clin Cancer Res 1996;2:755–62.
[37] Park JS, Hwang ES, Nakoong SE, et al. Physical status and expression of HPV genes in cervical cancers. Gynecol Oncol 1997;65:121–9.
[38] Tornesello ML, Buonaguro FM, Buonaguro L, Salatiello I, Giraldo EB, Giraldo G. Identification and functional analysis of sequence rearrangements in the long control region of human papillomavirus type 16 Af-1 variants isolated from Ugandan penile carcinomas. J Gen Virol 2000;81:2969–82.
[39] Bontkes HJ, De-Gruijl TD, Walboomers JMM. Human papillomavirus type 16 E2-specific T-helper lymphocyte responses in patients with cervical intraepithelial neoplasia. J Gen Virol 1999;80:2453–9.
[40] De Jong A, Van der burg SH, Kwappenberg KMC. Frequent detection of human papillomavirus 16 E2-specific T-helper immunity in healthy subjects. Cancer Res 2002;62:472–9.
[41] Storey A, Piccini A, Bouvard V, Banks L. Mutations in the human papillomavirus type 16 E2 protein identify a region of the protein involved in binding to E2 protein. J Gen Virol 1995;76:819–26.
[42] Watts KJ, Thompson CH, Cossart YE, Rose BR. Variable oncogene promoter activity of human papillomavirus type 16 cervical cancer isolates from Australia. J Clin Microbiol 2001;39:2009–14.
[43] Kämmer C, Warthorst U, Martinez NT, Wheeler CM, Pfister H. Sequence analysis of the long control region of human papillomavirus type 16 variants and functional consequences for P97 promoter activity. J Gen Virol 2000;81:1975–81.
[44] Schmidt M, Kedzia W, Jozefiak AG. Intratype HPV 16 sequence variation within LCR of isolates from asymptomatic carriers and cervical cancers. J Clin Virol 2001;23:65–77.
[45] Demeret C, Desaintes C, Yaniv M, Thierry F. Different mechanisms contribute to the E2-mediated transcriptional repression of human papillomavirus type 18 viral oncogenes. J Virol 1997;71:9343–9.
[46] Yamada T, Manos MM, Wheeler CM, et al. Human papillomavirus type 16 sequence variation in cervical cancers: a worldwide perspective. J Virol 1997;71:2463–72.
[47] Burk RD, Terai M, Gravitt PE, et al. Distribution of human papillomavirus types 16 and 18 variants in squamous cell carcinomas and adenocarcinomas of the cervix. Cancer Res 2003;63:7215–20.