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

D A Graham, C S Herrington

Abstract

**Aim**—To test the hypothesis that, because the human papillomavirus (HPV) E2 protein represses viral early gene transcription, E2 gene sequence variation or disruption could play a part in the induction of the numerical chromosome abnormalities that have been described in squamous cervical lesions.

**Methods**—The integrity and sequence of the E2 gene from 11 cervical intraepithelial neoplasia (CIN) grade 3 lesions and 14 invasive squamous cell carcinomas, all of which contained HPV-16, were analysed by the polymerase chain reaction (PCR). The E2 gene was amplified in three overlapping fragments and PCR products sequenced directly. Chromosome abnormalities were identified by interphase cytogenetics using chromosome specific probes for chromosomes 1, 3, 11, 17, 18, and X.

**Results**—E2 gene disruption was present in significantly more invasive carcinomas (eight of 14) than CIN 3 lesions (one of 11) (p = 0.03). No association was found between E2 disruption and the presence of a numerical chromosome abnormality. The E2 gene from the non-disrupted isolates was sequenced and wild-type (n = 5) and variant (n = 11) sequences identified. Variant sequences belonged to European and African classes and contained from one to 15 amino acid substitutions. Although numerical chromosome abnormalities were significantly more frequent in invasive squamous cell carcinoma than CIN 3 (p = 0.04), there was no significant relation between the presence of sequence variation and either histological diagnosis or chromosome abnormality.

**Conclusions**—These data do not support the hypothesis that E2 gene disruption or variation is important in the induction of chromosome imbalance in these lesions. However, there is a relation between E2 gene disruption and the presence of invasive disease.

Keywords: papillomavirus; E2 gene; cervix; chromosome

Human papillomaviruses (HPV) are DNA tumour viruses that are associated with the formation of epithelial tumours. Over 80 types of HPV have now been described and are categorised as low risk (for example, HPV-6 and HPV-11), intermediate risk (for example, HPV-31 and HPV-33), and high risk (for example, HPV-16 and HPV-18), based on their association with clinical disease. The HPV genome encodes two classes of genes that have been divided into two functional groups; the early (E) genes (E1, E2, E4, E5, E6, and E7) and the late (L) genes (L1 and L2). The E6 and E7 genes encode oncoproteins that, among other properties, bind to the p53 and retinoblastoma (pRb) proteins, respectively. The E1 protein is involved in viral DNA replication, the E4 protein in productive viral infection, and the E5 protein has some transforming properties. The L1 and L2 genes encode the major and minor capsid proteins, respectively.

The papillomavirus E2 protein is a modulator of papillomavirus transcription and replication, having major roles in viral DNA replication and as a repressor of viral early gene transcription. It has three functional domains: the N-terminal transactivation domain, the hinge region, and the C-terminal DNA binding domain. In vitro studies have shown that the HPV-16 and HPV-18 early promoters, which regulate the transcription of the oncogenic E6 and E7 genes, are regulated by the viral E2 protein. The expression of both the E6 and E7 genes of the high risk HPV-16 is necessary for the efficient immortalisation of primary human keratinocytes, and the E2 protein is able to repress E6 and E7 expression by binding to sites adjacent to the major early promoter. The integration of HPV DNA into host cellular DNA is associated with neoplastic progression and, when HPV integrates, the viral DNA frequently breaks in the E1–E2 region. Consequently, disruption of the E2 open reading frame results in loss of E2 protein function, leading to uncontrolled E6 and E7 gene expression. This is also in keeping with the fact that E2 gene mutation or disruption can increase the immortalisation capacity of HPV-16.

In a previous study, we identified distinct patterns of numerical chromosome abnormality in high grade intraepithelial and invasive squamous lesions of the cervix. In view of the in vitro evidence that the E2 protein can abrogate a mitotic control checkpoint, we hypothesised that the presence of numerical chromosome abnormalities in these lesions might be associated with the retention of intact E2 genes. Similarly, because mutation of the E2 gene can alter viral immortalisation capacity, we analysed the DNA sequence of the E2 gene in those lesions in which it was not disrupted to...
test the hypothesis that, in the absence of disruption, sequence variation might be important in the induction of chromosome abnormalities.

Materials and methods

CLINICAL MATERIAL

Eleven HPV-16 positive cervical intraepithelial neoplasia (CIN) grade 3 lesions and 14 HPV-16 positive invasive squamous cell carcinomas from our previous study of chromosome abnormalities\(^1\) were included in our present study. HPV-16 was identified using the GP5+/GP6+ generic PCR system and dot blot hybridisation, as described previously.\(^2\) An aliquot of 20 µl of DNA was used for each PCR reaction, which contained, in a total volume of 50 µl, 1× PCR buffer (Gibco BRL, Paisley, Scotland, UK), 1.5 mM MgCl\(_2\), 0.2 mM of each dNTP, and 1.25 units of AmpliTaq DNA polymerase. Hot start was achieved using Ampliwax PCR gems (Perkin-Elmer, Warrington, Cheshire, UK) and the following parameters were used: 95°C for four minutes then 40 cycles of 95°C for one minute, 58°C for two minutes, and 72°C for 1.5 minutes, followed by 72°C for seven minutes and a final resting temperature of 4°C. The reference HPV-16 clone (obtained from Dr E-M de Villiers, Heidelberg, Germany) and DNA extracted from paraflin wax embedded CaSki cells were amplified as positive controls. The omission of template DNA served as a negative control. Detection was performed by running 10 µl of amplified DNA on a 1.5% agarose gel, staining with ethidium bromide, and photographing over an ultraviolet transilluminator.

ANALYSIS OF E2 GENE INTEGRITY

DNA quality was demonstrated by amplification of a 536 bp β-globin fragment.\(^3\) The E2 gene was amplified in three overlapping fragments (amplifiers A, B, and C), each of which is smaller in size than the internal control amplifier, using the following primer pairs. Nucleotide positions (nt) are according to the HPV-16R sequence.\(^2\)

Amplifier A (475 bp product): A1, 5'-AGGACGAGGACAGAA-3' (nt 3735–3753); A2, 5'-ACTTGGACCTCTACACAGTTA-3' (nt 3187–3210). Amplifier B (477 bp product): B1, 5'-TTGTGAAGACATGACTAACT-3' (nt 3172–3193); B2, 5'-TAAGTATGTAGCATCACCTTT-3' (nt 3630–3649). Amplifier C (276 bp product): C1, 5'-GTAATAGATAGCATCACCCATT-3' (nt 3597–3618); C2, 5'-GGATGCGATATCGAAGATTGT-3' (nt 3853–3873).

These primers were validated using DNA extracted from paraflin wax embedded CaSki and SiHa cells, which contain intact and disrupted E2 genes, respectively.\(^2\) DNA was extracted from three 6 µm thick, formalin fixed, paraflin wax embedded sections using proteinase K (Boehringer-Mannheim, Lewes, Sussex, UK) and the supernatant treated with Instagene (Bio-Rad, Hemel Hempstead, Hertfordshire, UK) to remove cellular debris, as described previously.\(^2\) An aliquot of 20 µl of DNA was used for each PCR reaction, which contained, in a total volume of 50 µl, 1× PCR buffer (Gibco BRL, Paisley, Scotland, UK), 1.5 mM MgCl\(_2\), 0.2 mM of each dNTP, and 1.25 units of AmpliTaq DNA polymerase. Hot start was achieved using Ampliwax PCR gems (Perkin-Elmer, Warrington, Cheshire, UK) and the following parameters were used: 95°C for four minutes then 40 cycles of 95°C for one minute, 58°C for two minutes, and 72°C for 1.5 minutes, followed by 72°C for seven minutes and a final resting temperature of 4°C. The reference HPV-16 clone (obtained from Dr E-M de Villiers, Heidelberg, Germany) and DNA extracted from paraflin wax embedded CaSki cells were amplified as positive controls. The omission of template DNA served as a negative control. Detection was performed by running 10 µl of amplified DNA on a 1.5% agarose gel, staining with ethidium bromide, and photographing over an ultraviolet transilluminator.

E2 SEQUENCE ANALYSIS

PCR products from the isolates, the HPV-16 clone, and CaSki cells were purified from agarose gels using the QIAquick\(^\text{TM}\) gel extraction kit (Qiagen Ltd, Crawley, Sussex, UK) and the supernatant treated with proteinase K (Boehringer-Mannheim, Lewes, Sussex, UK) and the supernatant treated with proteinase K (Boehringer-Mannheim, Lewes, Sussex, UK). DNA was used for each PCR reaction, which contained, in a total volume of 50 µl, 1× PCR buffer (Gibco BRL, Paisley, Scotland, UK), 1.5 mM MgCl\(_2\), 0.2 mM of each dNTP, and 1.25 units of AmpliTaq DNA polymerase. Hot start was achieved using Ampliwax PCR gems (Perkin-Elmer, Warrington, Cheshire, UK) and the following parameters were used: 95°C for four minutes then 40 cycles of 95°C for one minute, 58°C for two minutes, and 72°C for 1.5 minutes, followed by 72°C for seven minutes and a final resting temperature of 4°C. The reference HPV-16 clone (obtained from Dr E-M de Villiers, Heidelberg, Germany) and DNA extracted from paraflin wax embedded CaSki cells were amplified as positive controls. The omission of template DNA served as a negative control. Detection was performed by running 10 µl of amplified DNA on a 1.5% agarose gel, staining with ethidium bromide, and photographing over an ultraviolet transilluminator.

Table 1  Summary of the relation between E2 gene disruption, histological diagnosis, and interphase cytogenetic data

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Karyotype</th>
<th>1</th>
<th>3</th>
<th>17</th>
<th>18</th>
<th>X</th>
<th>Non-amplifying Segment(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIN 3</td>
<td>Aneusemic</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>None</td>
</tr>
<tr>
<td>CIN 3</td>
<td>Aneusemic</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>None</td>
</tr>
<tr>
<td>CIN 3</td>
<td>Tetrasomic</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>None</td>
</tr>
<tr>
<td>CIN 3</td>
<td>Disomic</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>None</td>
</tr>
<tr>
<td>CIN 3</td>
<td>Disomic</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>None</td>
</tr>
<tr>
<td>CIN 3</td>
<td>Disomic</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>None</td>
</tr>
<tr>
<td>ISCC</td>
<td>Aneusemic</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>None</td>
</tr>
<tr>
<td>ISCC</td>
<td>Aneusemic</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>None</td>
</tr>
<tr>
<td>ISCC</td>
<td>Tetrasomic</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>None</td>
</tr>
<tr>
<td>ISCC</td>
<td>Aneusemic</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>None</td>
</tr>
<tr>
<td>ISCC</td>
<td>Aneusemic</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>None</td>
</tr>
<tr>
<td>ISCC</td>
<td>Aneusemic</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>None</td>
</tr>
<tr>
<td>ISCC</td>
<td>Aneusemic</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>None</td>
</tr>
<tr>
<td>ISCC</td>
<td>Disomic</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>None</td>
</tr>
<tr>
<td>ISCC</td>
<td>Disomic</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>None</td>
</tr>
<tr>
<td>ISCC</td>
<td>Disomic</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>None</td>
</tr>
<tr>
<td>ISCC</td>
<td>Disomic</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>None</td>
</tr>
<tr>
<td>ISCC</td>
<td>Disomic</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>None</td>
</tr>
<tr>
<td>ISCC</td>
<td>Disomic</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>None</td>
</tr>
<tr>
<td>ISCC</td>
<td>Disomic</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>None</td>
</tr>
<tr>
<td>ISCC</td>
<td>Disomic</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>None</td>
</tr>
<tr>
<td>ISCC</td>
<td>Disomic</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>None</td>
</tr>
</tbody>
</table>

The data for each chromosome refer to the number of chromosomes identified, as reported previously.\(^1\) The definitions of aneusemic, tetrasomic, and disomic are given in the text.

CIN 3, cervical intraepithelial neoplasia grade 3; ISCC, invasive squamous cell carcinoma.

CHROMOSOME ANALYSIS

Interphase cytogenetics was performed in each case for chromosomes 1, 3, 11, 17, 18, and X using pericentromeric probes, as described previously.\(^2\)\(5\)

STATISTICS

Proportions were compared using Fisher’s exact test and a significance level of p < 0.05 was used.

Results

INTERPHASE CYTOGENETICS

The lesions that were used in our study had been investigated previously for numerical chromosome abnormalities using pericentromeric chromosome probes specific for chromosomes 1, 3, 11, 17, 18, and X.\(^1\)\(8\) Lesions were classified as disomic (no numerical chromosome abnormality), tetrasomic (duplication of chromosome number with no imbalance), or aneusemic (numerical chromosome imbalance).
ance) (table 1). Aneusomy was present in significantly more invasive carcinomas (nine of 14) than CIN 3 lesions (two of 11) (p = 0.04).

DISRUPTION OF THE E2 GENE

The overlapping primer pairs were validated initially using DNA extracted from CaSki and SiHa cells: all three fragments were amplifiable from CaSki cells but only one (amplimer C) could be amplified from SiHa cells, as predicted from the published sequences.24 The 536 bp \(\beta\)-globin fragment was amplified in all 25 cases. Amplification of the HPV-16 E2 gene from these lesions demonstrated a failure of amplification of one or more amplimer, indicating disruption, in significantly more invasive squamous cell carcinomas (eight of 14) than CIN 3 lesions (one of 11) (p = 0.03; fig 1). There was, however, no significant relation between the presence of aneusomy and E2 gene disruption (p = 1; table 1).

E2 GENE SEQUENCE ANALYSIS

The HPV-16 isolates that did not show evidence of disruption were sequenced, as were amplimers from the HPV-16 clone and from CaSki cells, and compared with the wild-type HPV-16R E2 reference sequence22 to investigate variation within the E2 gene. Mutations were identified in a total of 24 positions, 18 of which resulted in amino acid changes (fig 2). A guanosine residue was also identified in position 2926 in all isolates, in addition to CaSki cells and the HPV-16 clone: this has been identified recently as a sequencing error in the HPV-16R sequence.22 Several consistent variants were identified. The most common individual base change was C to T at position 2410. This was present in isolates in which it was the only variant residue, or in which it was associated with other single base changes (group 1). It was also present in variants that showed more widespread variation (groups 2 and 3). Group 2 variants had the same sequence as the E2 gene in CaSki cells.24 Group 3 variants showed a much greater variation, with a total of 20 variant positions. Overall, there was no relation between the presence of E2 mutation and either histological diagnosis (p = 1.0) or chromosome imbalance (p = 1.0).

Variation in the transactivation domain was only identified in groups 2 and 3. In group 2, the single mutation in this domain is silent. In group 3, there are five mutations that lead to amino acid substitutions. Variation was also observed in the hinge and DNA binding regions, with a greater number of mutations again being present in the group 3 variants. Two different mutations were identified at position 3377: C to G was present in group 3, and produces a proline to arginine amino acid substitution; C to T, which produces a proline to serine substitution, was identified in a single isolate in group 1. Some mutations within the hinge region also affect the sequence of the E4 gene. These were identified only in group 3 variants and are as follows: L62T (3516, C to A and 3517, T to C); Q69P (3538, A to C); H78Q (3566, T to G).

Discussion

Disruption of the E2 gene was significantly more frequent in invasive squamous cell carcinomas than in CIN 3 lesions. Although this is in keeping with viral integration being important in neoplastic progression, disruption of the E2 gene does not necessarily equate to viral integration. E2 genes can be intact in the presence of E2 mutation and either histological diagnosis (p = 1.0) or chromosome imbalance (p = 1.0). Variation in the transactivation domain was only identified in groups 2 and 3. In group 2, the single mutation in this domain is silent. In group 3, there are five mutations that lead to amino acid substitutions. Variation was also observed in the hinge and DNA binding regions, with a greater number of mutations again being present in the group 3 variants. Two different mutations were identified at position 3377: C to G was present in group 3, and produces a proline to arginine amino acid substitution; C to T, which produces a proline to serine substitution, was identified in a single isolate in group 1. Some mutations within the hinge region also affect the sequence of the E4 gene. These were identified only in group 3 variants and are as follows: L62T (3516, C to A and 3517, T to C); Q69P (3538, A to C); H78Q (3566, T to G).
Within HPV-16, five phylogenetic branches have been defined on the basis of variation within the E6 and L1 genes, each predominating within specific geographical regions: these branches have been designated E (European), Aa (Asian), AA (Asian-American), and Af (African).

A recent report that details variation in the E2 hinge region and its relation to the more extensively studied E6 and L1 variants allows the approximate identification of the groups of isolates identified in our study. Group 1 variants show the 3362 A to G mutation in AA variants, whereas Casas et al. did not find this mutation.28 This does not affect the identification of our variants as AA.

Mutations in the transactivation domain were present both in group 2 and 3 variants. The single mutation in group 2 variants is silent and hence no amino acid substitutions in this domain are present. By contrast, the group 3 variants contain several mutations that lead to amino acid substitutions in this domain. The H35Q mutation is within the E1 interaction site, suggesting that this mutation identifies them as Af group 2a (Af2a) variants.27 A recent report detailing a series of AA variants supports this interpretation. Specifically, our group 3 isolates possess a 3410 G to T mutation, which does not occur in AA variants.28 It is of note that Eriksson and colleagues did not find the 3362 A to G mutation in AA variants, whereas Casas et al. did.28 This does not affect the identification of our variants as Af2a.

Figure 2  Summary of the position of the E2 mutations identified in the human papillomavirus 16 (HPV-16) isolates. The isolates are grouped according to the combination of mutations identified. The resulting amino acid substitution (if any) is given below each variant position. GenBank accession numbers have been assigned as follows: AF193425 (isolate 6); AF193426 (isolate 7); AF193427 (isolates 8–10); AF193428 (isolates 11 and 12); AF193429 (isolates 13–15); AF193430 (isolate 16). CIN, cervical intraepithelial neoplasia grade 3; ISCC, invasive squamous cell carcinoma.
variants containing this mutation were shown to be capable of supporting viral DNA replication in carcinomas in which the E1–E2 region was retained, indicating that this mutation does not abrogate this function. Mutations within the hinge region are of less clear functional importance, although they could potentially alter the three dimensional relation between transactivation and DNA binding domains. This possibility is of particular relevance to the P219S (3410 C to T) mutation, which occurred in all but one of the variant isolates, because the replacement of proline with serine at this position could significantly alter the secondary and tertiary protein structures. The frequency of this mutation raises the possibility that variants with this alteration could be involved in the development of high grade intraepithelial and invasive disease. However, an analysis of low grade intraepithelial lesions is required to assess this hypothesis.

The ability to bind to DNA, particularly that in the long control region of the HPV genome, is central to the function of the E2 protein. When binding to DNA, the E2 protein forms a homodimer, which interacts with the DNA molecule through its DNA binding helices formed from residues 292–309. Therefore, the T310K mutation, present both in group 2 and group 3 variants, might affect the three dimensional structure of this helix, and hence its ability to bind to DNA. However, a recent study showed very similar transcriptional activities when E and AA variants were compared. Although in vivo function cannot necessarily be inferred from in vitro functional studies, these data suggest that the amino acid sequence variation identified might not be relevant to this function of the protein.

Another possible effect of amino acid substitution in the E2 gene is an alteration of the host immune response to this protein. Linear epitopes have been described within the E2 protein: one in the transactivation domain (residues 121–140), three in the hinge region (residues 181–200, 241–260, and 271–290), and one in the DNA binding domain (residues 328–346). It is interesting that only the amino acid substitutions present in the group 3 variants, and that in the single group 4 isolate, affect these regions. Whether this is of biological relevance remains to be determined.

The lack of a relation between the presence of sequence variation and either histological diagnosis or chromosome imbalance suggests that any functional difference between the E2 variants is unlikely to be involved in determining either the development of aneuomy or progression to invasive carcinoma. However, these data do not exclude the possibility that the variants might differ in their ability to induce progression of a productive HPV-16 infection to CIN 3.

In conclusion, E2 gene disruption appears to be related to stromal invasion, but not to the presence of numerical abnormalities of chromosomes 1, 3, 11, 17, 18, or X. A variety of E2 gene variants were identified in the lesions analysed, but there was no association between the presence of these variations and either stromal invasion or chromosome abnormality. However, this does not exclude the possibility that E2 gene variation is important in determining the natural history of HPV-16 infection or the likelihood of progression to CIN 3. Further data are required regarding the prevalence of the E2 variants described across the spectrum of cervical squamous neoplasia.

We thank Wellbeing and the Royal College of Obstetricians and Gynaecologists, UK for funding the initial phase of this study.

24 Meissner JD. Nucleotide sequences and further characterization of human papillomavirus DNA present in CaSki, SiHa and HeLa cervical carcinoma cell lines. J Gen Virol 1999;80:1725–33.