Disruption of the E2 Gene Is a Common and Early Event in the Natural History of Cervical Human Papillomavirus Infection: A Longitudinal Cohort Study

Stuart I. Collins, Christothea Constandinou-Williams, Kaisheng Wen, Lawrence S. Young, Sally Roberts, Paul G. Murray, and Ciaran B.J. Woodman

Abstract
Integration of high-risk human papillomavirus (HPV) types into the host-cell genome disrupts the HPV regulatory E2 protein, resulting in a loss of negative feedback control of viral oncogene expression; this disruption has been considered a critical event in the pathogenesis of cervical neoplasia, and a potential biomarker of progressive disease. However, using serial samples taken from a cohort of young women who were recruited soon after they first had sexual intercourse, we show that disruption of the E2 gene is a common and early event in the natural history of incident cervical HPV infections. The E2 gene was significantly more likely to be disrupted in women who tested positive for HPV18 in their baseline sample than in those who tested positive for HPV16 [26% versus 58%; relative risk, 2.26; 95% confidence interval (CI), 1.38–3.71; \( \chi^2, 9.23; 1 \) degree of freedom (df); \( P = 0.002 \)]. Among women with an intact E2 gene in their baseline sample, the median time to first detection of E2 disruption was also shorter for those who tested positive for HPV18 than HPV16 (5.7 versus 10.9 months; hazards ratio, 1.93; 95% CI, 0.84–4.44; \( \chi^2, 2.49; 1 \) df; \( P = 0.11 \)). This tendency for HPV18 to integrate early, coupled with the substantial reduction in viral load in HPV18-positive samples in which E2 is disrupted, may explain why HPV18-associated disease is often reported to be characterized by minor cytologic changes, which underestimate the severity of the underlying histologic abnormality. [Cancer Res 2009; 69(9):3828–32]

Introduction
The identification of high-risk human papillomavirus (HPV) types offers not only the prospect of effective primary prevention but also the possibility of improving the efficiency of cervical screening programs. Although most women will at some time have been infected with high-risk HPV types, very few will progress to invasive disease. Therefore, there is a continuing need for the identification of viral and host factors that modulate the risk of disease progression, and which might serve as predictive biomarkers. Integration into the host genome of high-risk HPV types, once considered a late event in cervical carcinogenesis but now reported to occur commonly in high-grade CIN, is one such factor (1–3).

HPV DNA can be found in cervical material in either epithelial or in integrated forms, or in both episomal and integrated forms. Viral integration occurs downstream of the E6 and E7 oncogenes in the E1 or E2 region, resulting in loss of negative feedback control of oncogene expression by the now disrupted viral regulatory E2 protein. Integrant-derived transcripts are more stable than those derived from episomal viral DNA, and HPV16 integration has been associated with a selective growth advantage for affected cells (4, 5).

When only exfoliated cervical cells are available for analysis, the identification of a small number of integrated forms in a background of mainly episomal forms is a substantial technical challenge, and one which has yet to be adequately addressed (6–8). A technically simpler proposition would be to test these samples for the disruption of the E2 gene. Although such an assay can only detect integrated forms of HPV when episomal forms are absent, this may not be a critical consideration because in vitro studies suggest that it is only the loss of regulatory episomal E2 in cells in which integration has already occurred, which confers a growth advantage on selected cells (9, 10). Were this sequence of events also to occur in vivo, then the testing of HPV-positive samples for the absence of E2 would merit consideration as a marker of disease progression.

However, before pressing yet another biomarker into clinical service, it is important to determine how often E2 is disrupted during the natural history of infection with high-risk HPV types. We have addressed this issue using longitudinal serial samples taken from a cohort of young women who were recruited soon after they first had sexual intercourse.

Materials and Methods

Study population. Women (2011) aged 15 to 19 y who visited one Birmingham Brook Advisory Centre (a family planning clinic) in Birmingham, United Kingdom, were recruited between 1988 and 1992, and were asked to reattend at intervals of 6 mo follow-up ended on 31st August 1997. At each visit, 2 cytologic samples were taken using the same Ayre’s spatula: the first was used to prepare a smear for cytologic evaluation; the second was placed into 10 mL of phosphate-buffered saline and stored at −80°C for subsequent virological examination; inevitably, cell numbers varied dependent on the cellular composition of the cytologic sample. We immediately referred all participants in whom a cytologic abnormality was identified to a research clinic, irrespective of the severity of that abnormality. In this clinic, a sample of the most severe colposcopic abnormality was removed for histologic examination. Colposcopic and
cytologic surveillance was maintained in these women and treatment was postponed until there was histologic evidence of high-grade CIN (CIN2 or CIN3), at which point, women left the study (11).

After all clinical follow-up had ended, cervical samples were tested for the presence of HPV DNA using a general primer-mediated (GP5+/GP6+) PCR, and further PCR tests were done with type-specific primers on samples that were HPV positive (11); we subsequently refer to this testing strategy as the "GP5+/GP6+ system." A 2-μL aliquot was taken from the stored sample and DNA was extracted using guanidinium thiocyanate acid; 100 ng of sample DNA were then used in a 50 μL PCR reaction according to a method previously described (12, 13). The study was approved by the appropriate ethical committee, and informed oral consent was obtained from all women. The study population for this analysis comprises the subset of women who were cytologically normal and HPV DNA negative at study entry and who first tested positive during follow-up for HPV16 or HPV18, or both.

HPV genotyping. For the present study, the remainder of the stored sample was pelleted. Proteinase K digestion and phenol/chloroform were used to extract DNA. The Nanodrop ND-1000 spectrophotometer was used to measure DNA concentration, according to the manufacturer's instructions. DNA was used to extract DNA. The Nanodrop ND-1000 spectrophotometer was used to measure DNA concentration, according to the manufacturer's instructions. The quality of which was assessed by amplification of a 158-bp fragment for GAPDH, HPV16 E6, and HPV18 E7 (Supplementary Table S1). Genomic DNA (50 ng) and standards (10-fold plasmid dilutions between 10^6 and 10^7 copies of GAPDH, HPV16, and HPV18) were used to generate measurements of viral load normalized for cellular DNA content.

Assessment of HPV E2 integrity. Adapting an approach previously described (19, 25), the integrity of the E2 gene was assessed using sets of sequence-specific primers that were designed to amplify overlapping fragments that spanned the full length of the HPV16 and HPV18 E2 genes (Fig. 1A and B). Details of these primer sequences and their melting temperatures are given in Supplementary Table S1. HPV16 primer pairs and HPV18 primer pairs were initially validated using DNA extracted from the cervical cell lines CaSki, SiHa, and W12 (kindly supplied by Margaret Stanley); the latter cell line is characterized by the loss of episomal forms of HPV16 and the emergence of integrated forms during long term in vitro cultivation (9). CaSki, which contains multiple copies of the HPV16 genome arranged in the host chromosomes as head-to-tail, tandemly repeated arrays, provided a positive E2 control (19). In SiHa, E2 is disrupted at nucleotides 3132 and 3384, and as would be predicted from this observation, primer set 2 (data not shown) did not amplify when this cell line was tested (26). We found that the E2 gene was intact in early passage W12 but disrupted in late passage with failure to amplify primer sets 4 and 5 (Fig. 2A), this is

Figure 1. A, location of primers used to amplify overlapping regions of the HPV16 E2 gene. Nucleotides are numbered according to the HPV16 whole genome sequence. B, location of primers used to amplify overlapping regions of the HPV18 E2 gene. Nucleotides are numbered according to the HPV18 whole genome sequence.
consistent with the reported virus-cellular junctions in late passage W12 at nucleotides 3732 and 4791 (4, 27). HPV18 primer pairs were validated using primary human foreskin keratinocytes transfected with episomal HPV18 (28). All four primer sets failed to amplify in the HeLa cell line (Fig. 2B); this is consistent with the reported virus-cellular junctions in this cell line at position 3100 and 5736 (29).

The sensitivity of the assay was assessed using 10-fold serial dilutions of HPV16 or HPV18 plasmid ranging from 10^9 to 1 copies. In brief, samples were amplified using GoTaq Green Master Mix with 0.4 μmol/L of appropriate primer mix. Cycling conditions for HPV16 and HPV18 E2 were as follows: 95°C for 5 min, followed by 60 cycles of 95°C for 30 s, melting temperatures for 1 min and 72°C for 2 min, then a final extension of 72°C for 10 min. PCR products were analyzed using electrophoresis on a 2% agarose gel. The sensitivity of each of the sets of primers used in the detection of full-length E2 was the same as that observed for HPV E6 (10^3 copies), thus minimizing the likelihood of differential sensitivity. When testing study samples, 50 ng of DNA were amplified using the conditions described above.

Classification of study samples. The following protocol was used when assigning study samples into one of three mutually exclusive and exhaustive categories: E2 intact, E2 disrupted, and type-specific HPV DNA negative.

First, a sample was tested using the full set of E2 categories:

1. HPV DNA was detected, and the date of the immediately preceding evaluable sample; censoring occurred on the earliest of the date of the sample in which the HPV type-specific negative sample was detected, or the date a woman’s last cytologic sample was taken. Estimates of cumulative risk obtained by unconditional maximum likelihood estimation and the associated 95% confidence intervals constructed using a normal approximation; tests of hypotheses were undertaken using the Pearson χ² test with continuity correction. Analyses of time to E2 disruption were undertaken using methods appropriate for interval-censored time-to-event data. Time to E2 disruption was measured from date of baseline sample until the interval between the date of the sample in which disruption of the E2 gene was first detected, and the date of the immediately preceding evaluable sample; censoring occurred on the earliest of the date of the sample in which the first HPV type-specific negative sample was detected, or the date a woman’s last cytologic sample was taken. Estimates of cumulative risk of E2 disruption were obtained using a nonparametric maximum likelihood estimator (30). Estimates of hazards ratios were obtained using a semiparametric method for modeling interval-censored time-to-event data as a generalized linear model (31): 95% confidence intervals were constructed from parameter estimates and their standard errors, and tests of hypotheses were undertaken using likelihood ratio tests. All tests of statistical significance were conducted at the 5% two-sided significance level. Women who tested positive for HPV16 and for HPV18 in their baseline samples (n = 10) were excluded from the cross-sectional cohort analysis because the first GP5+/GP6+ positive sample taken after study entry was not available. Disruption of the E2 gene was observed in a follow-up sample in 15 (42%) of 36 women with an incident HPV18 infection who tested positive for HPV E2/E6 at baseline; the E2 gene was intact in 21 (58%). Disruption of the E2 gene in the baseline sample was significantly more likely in women who tested positive for HPV16 than in those who tested positive for HPV16 [relative risk, 2.26; 95% confidence interval, 1.38–3.71; χ², 9.23; 1 degree of freedom (df); P = 0.002].

We next determined how often the E2 gene became disrupted during follow-up. Of the 49 women in the incident HPV16 cohort who had an intact E2 gene in their baseline sample, 2 were excluded from this analysis because the first GP5+/GP6+ positive sample taken after study entry was not available. Disruption of the E2 gene was observed in a follow-up sample in 19 of the remaining 47 women (median number of samples tested, 2; range, 2–9). Of 15 women in the incident HPV18 cohort who had an intact E2 gene in their first HPV18 GP5+/GP6+–positive sample, disruption of the E2 gene was observed in a follow-up sample in 7 (median number of samples tested, 2; range, 2–9). Median time to first detection of a disrupted E2 gene was 5.7 months for HPV18, compared with 10.9 months for HPV16; the difference was not statistically significant (hazards ratio, 1.93; 95% confidence interval, 0.84–4.44; χ², 2.49; 1 df; P = 0.111).

Results

We first determined how often the HPV E2 gene was disrupted in the baseline samples of women with incident infections. There were 66 women with an incident HPV16 infection who tested positive for HPV16 E2/E6 at baseline; the E2 gene was intact in 49 (74%) and disrupted in 17 (26%).

There were 36 women with an incident HPV18 infection who tested positive for HPV E2/E6 at baseline; the E2 gene was intact in 15 (42%) and disrupted in 21 (58%). Disruption of the E2 gene in the baseline sample was significantly more likely in women who tested positive for HPV18 than in those who tested positive for HPV16 [relative risk, 2.26; 95% confidence interval, 1.38–3.71; χ², 9.23; 1 degree of freedom (df); P = 0.002].

We next determined how often the E2 gene became disrupted during follow-up. Of the 49 women in the incident HPV16 cohort who had an intact E2 gene in their baseline sample, 2 were excluded from this analysis because the first GP5+/GP6+ positive sample taken after study entry was not available. Disruption of the E2 gene was observed in a follow-up sample in 19 of the remaining 47 women (median number of samples tested, 2; range, 2–9). Of 15 women in the incident HPV18 cohort who had an intact E2 gene in their first HPV18 GP5+/GP6+–positive sample, disruption of the E2 gene was observed in a follow-up sample in 7 (median number of samples tested, 2; range, 2–9). Median time to first detection of a disrupted E2 gene was 5.7 months for HPV18, compared with 10.9 months for HPV16; the difference was not statistically significant (hazards ratio, 1.93; 95% confidence interval, 0.84–4.44; χ², 2.49; 1 df; P = 0.111).
Finally, we investigated the relationship between disruption of E2 and viral load. We included in this analysis all samples taken on or after the baseline sample up to and including the sample in which E2 disruption was first detected or, when this did not occur, the last type-specific positive sample. A measure of viral load was available for 64 samples in which HPV16 E2 was present, and for 16 samples in which it was disrupted; this was also available for 21 samples in which HPV18 E2 was present, and 17 samples in which it was disrupted. The median viral load in HPV16-positive samples with an intact HPV16 E2 gene was 2,250 (interquartile range, 59–27,122) copies per 1,000 cells, compared with 1 (0.1–34) copy per 1,000 cells in samples in which E2 was disrupted. The median viral load in HPV18-positive samples with an intact HPV18 E2 gene was 705 (194–17,250) copies per 1,000 cells, compared with 28 (0.8–2,692) copies per 1,000 cells in samples in which E2 was disrupted.

**Discussion**

We have shown that disruption of the E2 gene is a common and early event after an incident HPV16 or HPV18 infection. This finding is consistent with recent cross-section studies describing the detection of integrated forms of HPV16 and HPV18 in women with low-grade CIN (32–34). The true incidence of integration events is almost certainly higher than we report because our assay was not designed to identify what are often described as “mixed infections,” i.e., those in which E2 continues to be detected because both episomal and integrated forms are present (32–34), nor could this assay reveal multicopy head-to-tail tandem repeat integrations in which E2 continues to be detected because only the flanking copies are disrupted (26). Unlike HPV18-associated cancers and high-grade CIN, in almost all of which an intact E2 gene can no longer be detected, head-to-tail tandem repeats have been reported in HPV16-associated tumors: this pattern of integration might offer a partial explanation for why we found that HPV18 E2 disruption was more common in our cohort (1, 8, 25, 35–38). In any event, it is difficult to escape the conclusion that integrated forms could be present even earlier than we report, and that their appearance, with or without the loss of episomal forms, is so common as to make it unlikely that they could be used as a marker of disease progression. It remains to be determined whether the detection of integrant-derived transcripts in cervical material will have a useful predictive value (39).

Differences between HPV16- and HPV18-associated disease are of particular interest because HPV18 is the second most commonly occurring infection in young women; and it is the HPV type most strongly associated with adenocarcinoma of the cervix, which is increasing in incidence at a time when that of squamous cancer is decreasing (40–42). On the other hand, HPV18 is rarely reported to be present at the same time as high-grade CIN is diagnosed (43–45). In an attempt to explain this paradox, it has been suggested that HPV18-associated disease rapidly progresses through the preinvasive stages of cervical neoplasia (42). However, we have previously shown an increased risk of high-grade CIN after HPV18 infection in our study population (11). Surprisingly, this analysis also revealed that whereas the risk of minor cytologic abnormalities was similar after an incident HPV16 or HPV18 infection, there was no excess risk of moderate or severe dyskaryosis after a HPV18 infection (46). We concluded, and others have since concurred (47), that the cytologic changes detected after HPV18 infection, unlike those detected after a HPV18 infection, underestimate the severity of the underlying histologic abnormality. Our finding that HPV18 integrates early offers a possible explanation for these observations. If the cytopathic effect observed in exfoliated cervical cells is a reflection of viral load, and if as we and others have reported, a low-viral load is associated with less severe cytologic abnormalities (reviewed in ref. 8), then the preponderance of minor cytlogic changes observed in women with HPV18-associated disease may simply reflect the decrease in copy number that follows the more rapid integration of HPV18.

These are important considerations because the benefits of cervical screening follow from the detection, investigation, and treatment of epithelial abnormality, with the decision to refer for colposcopic assessment usually based on the severity of the cytologic abnormality. If HPV16 infection is more likely to be followed by a severe cytologic abnormality than is HPV18 infection, then screening programs are more likely to interrupt the natural history of an HPV16 infection than that of an HPV18 infection. In many countries, screening programs have not prevented an increase in the incidence of adenocarcinoma of the cervix, despite their success in reducing that of squamous cell carcinoma (48–51). Although this failure has been in part attributed to the inaccessibility of these lesions to cytologic sampling, our observations on the relationship between changes in viral load and the loss of episomal forms offer another possible explanation for these trends.

As to how and why HPV18 episomes are so consistently and rapidly lost while integrated forms continue to be detected merits further investigation. For example, in vitro studies suggest that loss of episomal E2 with the emergence of integrated forms is associated with the endogenous activation of anti-viral genes, and therefore, the type-specific differences we have observed may be consistent with a recent report suggesting that HPV18 is more immunogenic than HPV16 (52, 53).

**Disclosure of Potential Conflicts of Interest**

None of the authors have any conflict of interest, and the grant giving body had no input whatsoever into the analysis or preparation of the manuscript.

**Acknowledgments**

Received 8/12/08; revised 1/14/09; accepted 1/19/09.

Grant support: Cancer Research UK (grant no. C965/A4634).

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

We thank the women who took part in this study, Francesca Lewis for technical support, and Shikha Bose for technical advice.

**References**


Disruption of the E2 Gene Is a Common and Early Event in the Natural History of Cervical Human Papillomavirus Infection: A Longitudinal Cohort Study

Stuart I. Collins, Christothea Constandinou-Williams, Kaisheng Wen, et al.