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,1 Christothea Constantinou-Williams,2 Kaisheng Wen,2 Lawrence S. Young,2 Sally Roberts,2 Paul G. Murray,2 and Ciaran B.J. Woodman2

1Cancer Research UK Clinical Trials Unit, Institute for Cancer Studies, and 2Cancer Research UK Institute for Cancer Studies, University of Birmingham, Edgbaston, Birmingham, United Kingdom

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; \(\chi^2\) = 0.11). 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; \(\chi^2\), 9.23; 1 degree of freedom (df); \(P = 0.002\)). 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 (2,011) 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 /C0

Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

S.I. Collins and C. Constantinou-Williams contributed equally to this work.

Requests for reprints: Stuart Collins, Cancer Research UK Clinical Trials Unit, Institute for Cancer Studies, University of Birmingham, Edgbaston, Birmingham, United Kingdom, B15 2TT. Phone: 44-0-121-414-8876; Fax: 44-0-121-414-3700; E-mail: s.i.collins@bham.ac.uk.

©2009 American Association for Cancer Research.

doi:10.1158/0008-5472.CAN-08-3099
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. A 158-bp fragment that was 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.

**Measurement of viral load.** HPV viral load was measured using a modified singleplex real-time PCR assay (24). In brief, sequence-specific primers and Fluorescein-labeled probes were designed for GAPDH, HPV16 E6, and HPV18 E7 (Supplementary Table S1). Genomic DNA (50 ng) and standards (10-fold plasmid dilutions between 10^6 and 10^10 copies of GAPDH, HPV16, and HPV18) were amplified using AmpliTaq Master Mix (Applied Biosystems) with 0.4 µmol/L of appropriate primer mix. Amplifications were performed using the ABI 7700 sequence detection system, and cycle conditions for GAPDH, HPV16 E6, and HPV18 E7 were as follows: 50 °C for 2 min, 95 °C for 12 min, followed by 50 cycles of 95 °C for 15 s and 55 °C for 30 s. The HPV16-positive cervical carcinoma cell line SiHa, and the HPV18-positive cell line HeLa, were used as positive controls. Standard curves 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 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). CaSkii, 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](image-url)

**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\(^2\) 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\(^2\) 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 primers. When the results of all E2 primers were positive, the E2 gene was considered intact. When at least one E2 primer gave a negative result and at least one primer gave a positive result, or when all E2 primers were negative but type-specific E6 could be detected, the E2 gene was considered to be disrupted (illustrative examples are shown in Fig. 2A and B). When all E2 primers and the E6 primer tested negative, the sample was categorized as type-specific HPV DNA negative, irrespective of the results of our earlier GP5+/GP6+ analysis; however, only two samples that tested positive using GP5+/GP6+-mediated PCR tested negative using E2/E6 PCR. After the first detection of a disrupted E2 gene, no further samples from that woman were tested for E2 or E6.

**Statistical analysis.** The baseline sample was defined as the first sample taken after study entry to test positive for one or both of HPV16 or HPV18 using the GP5+/GP6+ system. When the baseline sample was not available, as was the case for three women in the cross-sectional HPV16 analysis and one in the HPV18 analysis, the next available sample was used. Cross-sectional comparisons of the prevalence of E2 disruption in the baseline sample were made using contingency tables, with estimates of relative 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 χ\(^2\) 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 included twice in the cross-sectional comparison of the prevalence of E2 disruption, once for each type-specific analysis. Similarly, of these 10 women, 5 who had an intact HPV E2 gene in their baseline samples, were included in each type-specific analysis of time to first detection of E2 disruption.

**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 HPV18 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; χ\(^2\), 9.29; 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; χ\(^2\), 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 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-sectional 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 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 strongly associated with adenocarcinoma of the cervix, which is 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 strongly associated with adenocarcinoma of the cervix, which is increasing in incidence at a time when that of squamous cancer is strongly associated with adenocar...


Downloaded from cancerres.aacrjournals.org on August 18, 2017. © 2009 American Association for Cancer Research.
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.