Methylation of Human Papillomavirus Type 16 Genome and Risk of Cervical Precancer in a Costa Rican Population

Lisa Mirabello, Chang Sun, Arpita Ghosh, Ana C. Rodriguez, Mark Schiffman, Nicolas Wentzensen, Allan Hildesheim, Rolando Herrero, Sholom Wacholder, Attila Lorincz, Robert D. Burk

Background
Previous studies have suggested an association between human papillomavirus type 16 (HPV16) genome methylation and cervical intraepithelial neoplasia grade 3 (CIN3) (ie, cervical precancer) and cancer, but the results have been inconsistent.

Methods
We designed a case–control study within a large prospective cohort of women who underwent multiple screenings for cervical cancer in Guanacaste, Costa Rica. Diagnostic specimens were collected at the time of CIN3 diagnosis (n = 30 case subjects) and persistent HPV16 infection (persistence; n = 35 case subjects), prediagnostic specimens at the first HPV16-positive screening visit (n = 20 CIN3 case subjects; n = 35 persistence case subjects), and control specimens from women with infection clearance within 2 years (n = 34 control subjects). DNA extracted from specimens (cervical cells) was analyzed for methylation levels at 67 CpG sites throughout the HPV16 genome using pyrosequencing. Benjamini–Hochberg method was used to account for multiple testing. Associations between methylation levels and risk of CIN3 or persistence were assessed using logistic regression models to estimate odds ratios (ORs) and 95% confidence intervals (CIs).

Results
Increased methylation in diagnostic vs control specimens at nine CpG sites, three in each L1, L2, and E2/E4 genomic regions, was associated with an increased risk of CIN3 (third tertile [high] vs first and second tertiles combined [low], OR = 3.29 [95% CI = 1.16 to 9.34] to 11.12 [95% CI = 2.29 to 76.80]) and persistence. High methylation at three of these CpG sites was associated with a much higher risk when combined compared with low methylation at these sites (OR = 52, 95% CI = 4.0 to 670). In prediagnostic vs control specimens, increased methylation at a CpG site (nucleotide position 4261) in L2 was associated with an increased risk of CIN3.

Conclusion
In this HPV16-infected cohort, increased methylation of CpG sites within the HPV16 genome before diagnosis and at the time of diagnosis was associated with cervical precancer.


Persistent infection with different types of carcinogenic human papillomavirus (HPV) is the primary cause of cervical cancer—the third most common cancer in women worldwide (1). HPV infection is the most common sexually transmitted infection, but only a small percentage of women infected with HPV types of known carcinogenic potential progress to histologically confirmed cervical intraepithelial neoplasia grade 3 (CIN3) (ie, precancer), with a substantial fraction of those eventually developing invasive cancer (2,3). More than 90% of HPV infections clear within a few years of acquisition of the virus (4).

The major known determinant of infection outcome pathogenesis is the type of HPV. HPV type 16 (HPV16) is the most prevalent and persistent HPV and causes approximately 50–60% of all cervical cancer and CIN3 (5,6). A substantial proportion of persistent HPV16 infections lead to a diagnosis of CIN3, whereas few long-term persistent HPV16 infections neither progress nor regress (7–9). Approximately 30%–50% of CIN3 lesions progress to invasive cervical cancer (2,10). Despite multiple prospective analyses (11), the determinants of CIN3 development as a result of HPV16 infection are largely unknown, except for weak cofactors such as smoking, long-term oral contraceptive use, and multiparity.

Recently, methylation of the HPV16 genome has been suggested to be associated with CIN3 and cancer (12–19). Within the human genome, methylation of cytosines in the CpG dinucleotides (also known as CpG sites) clustered into islands associated with transcriptional promoters is an important cellular mechanism to regulate gene expression. Beyond regulating gene expression, cells use methylation as a defense mechanism against foreign agents (eg, viral DNA) (20,21). Methylation of HPV DNA by infected cells may alter the expression patterns of viral genes that are relevant for infection and transformation (22,23), and it may
provide important clues to help researchers understand why certain infections are cleared and certain infections persist with or without progression to precancer.

With approximately 8000 base pairs and eight genes (24), there are 113 CpG sites in the HPV16 genome that could be potentially methylated (17). Previous HPV16 methylation studies have been cross-sectional or have investigated limited regions of the HPV16 methylome (12–16,19,25). These studies have found differing frequencies of methylation at specific sites by infection outcome, and the results have been inconsistent.

The most comprehensive mapping of the HPV16 methylome to date found that methylation levels increased at specific regions in the early gene E5 open reading frame (ORF) (ie, a multifunctional protein with transforming activity), and the structural late genes L1 (ie, major capsid protein) and L2 (ie, minor capsid protein) with increased severity of the cervical lesions (17). In contrast, a recent prospective study investigated methylation at six CpGs sites in the E6 (ie, viral oncoprotein that can degrade tumor protein p53 [TP53]) gene promoter and enhancer in the upstream regulatory region (URR; regulates viral replication and gene expression) of HPV16 and found an association between higher methylation levels and a reduced likelihood of being diagnosed with CIN grade 2 (CIN2) or histologically confirmed higher grades (CIN2+) (18).

We evaluated HPV16 DNA methylation using specimens from both the time of diagnosis and before diagnosis among women selected from a prospective cohort. We examined whether HPV16 DNA methylation at multiple CpG sites, which include CpGs in all genomic regions, are associated with clearance of cervical HPV16 infection, development of CIN3, or persistent HPV16 infection without CIN2+.

**Methods**

**Study Population, Cervical Specimens, and Study Design**

Stored DNA isolated from cervical cells previously documented to contain HPV16 DNA were obtained from selected women with different outcomes from the population-based prospective cohort study in Guanacaste, Costa Rica (26). This is a longitudinal study of the natural history of HPV infection and cervical neoplasia amongst 10049 participants (27). Random census tracts representing approximately one-sixth of the Guanacaste province were selected. All women aged 18 years and older were recruited for screening and follow-up between June 1993 and December 1994. Follow-up intervals were determined by the severity of screening results. The participation rate was 93.6%, and loss to follow-up during the study period (June 1993 to August 2002) was less than 10% (28).

Women were screened by cervical visual images (Cervigram; National Testing Laboratories Worldwide, Fenton, MO), and cytological diagnosis was based on specimens collected with a Cervix Brush (Unimar, Wilton, CT), a broom-type collection device. Specimens were evaluated by three methods: conventional Papanicolaou smear, PapNet (Neuromedical Systems Inc, Suffern, NY), and a liquid-based ThinPrep (Cytyc Corp. [now part of Hologic], Marlborough, MA). A second cervical specimen was collected using a Dacron swab (Digene, Gaithersburg, MD; now part of Qiagen, Hilden, Germany) and placed into either ViraPap DNA transport medium or sample transport medium (Digene) for HPV testing. The methylation studies were conducted on DNA isolated from the swab-derived specimen. Details of the epidemiological study design and methods are described elsewhere (27).

The study protocol was reviewed and reapproved annually by the US National Cancer Institute and Costa Rican Institutional Review Boards.

Final diagnosis was determined by review of all histology slides as previously described (26). HPV16 infections were identified in 503 women participating in the study. Based on the complete study data (26), we selected 36 women representing each of three main infection outcomes using a case–control study design: 1) women who cleared the HPV16 infection in less than 2 years and did not have CIN2+ lesions (control subjects); 2) women with HPV16 infection and a histologically diagnosed CIN3 (CIN3 case subjects); and 3) women with more than 2 years of persistent HPV16 infection without observed CIN2+ lesions (persistence case subjects). We did not include women with a diagnosis of CIN2 in this study because it is an ambiguous endpoint, which combines acute infections and early precancerous lesions (29). Long-term HPV16 persistence without disease is an uncommon viral outcome. Because women with persistence made up the smallest group (n = 36 women), the number of women in the other two groups was matched to this group.

**CONTEXT AND CAVEATS**

**Prior knowledge**

Cervical cancer and histologically confirmed cervical intraepithelial neoplasia grade 3 (CIN3) (ie, precancer) is primarily caused by persistent infection with an oncogenic human papillomavirus (HPV); HPV type 16 (HPV16) being the most prevalent virus. Methylation of the HPV16 genome is suggested to be associated with CIN3 and cancer.

**Study design**

A case–control study was designed, and cervical specimens were collected at the time of CIN3 diagnosis or HPV16 persistence, before diagnosis or persistence, and from women who showed clearance of HPV16 within 2 years. HPV16 genome methylation at multiple CpG sites was analyzed, and association with development of CIN3 or persistence of infection was assessed using clearance specimens as the control group.

**Contribution**

Increased methylation at nine CpG sites in the L1, L2, and E2/E4 genomic region of HPV16 was associated with increased risk of CIN3 and persistence compared with control specimens. Three of these sites when combined showed a very high risk of CIN3. Increased methylation before diagnosis was also associated with increased risk of incident CIN3.

**Implications**

Cervical precancer was associated with increased HPV16 methylation, and this assay may serve as a biomarker for managing women with HPV16 infections of the cervix.

**Limitation**

Modest sample size and possible variability in methylation may have limited the ability to detect all associations.

From the Editors
To represent the common outcome of HPV acquisition (typically at young ages) followed by rapid clearance of infections, we randomly chose 36 women aged 25 years and younger as the control group. For case subjects with incident CIN3, we analyzed the HPV16 specimen collected at the screening visit closest to diagnosis of CIN3 (median time of 5 months before diagnosis) (CIN3 diagnostic specimen). The HPV16 specimen collected at the last screening visit was used for analysis from case subjects with persistence (persistence diagnostic specimen). We manually reviewed each woman’s visit history report to confirm that the viral disease outcome was classified correctly. We selected similar proportions of HPV16 variant lineages (European, non-European) among the three groups.

We also analyzed HPV16 specimens collected at the first HPV-positive screening visit (persistence diagnostic specimen) in case subjects with CIN3 and persistence. Twenty case subjects whose infections resulted in incident CIN3 had preceding HPV16-containing specimens available for analysis, and the median time between the first specimen and detection of CIN3 was 3 years (range = 1–7 years). For the case subjects with persistence but no CIN2+, the median time between the diagnostic and persistence diagnostic specimens collected was 6 years (range = 2–8 years).

DNA Isolation and Analysis of HPV16 DNA Methylation

Briefly, DNA was extracted from cervical cells and tested for the presence of HPV16 by polymerase chain reaction (PCR) using the MY09/MY11 primer set and type-specific dot blot hybridization as previously described (9). We focused on two regions within the HPV16 genome where CpG methylation was previously suggested to be associated with cervix precancer and cancer: the HPV16 L1 ORF and the URR region (13,15,17,30); in addition, we sampled each of the other viral genome ORFs for one or more CpG methylation sites. Primers for PCR were designed using MethPrimer (31) (http://www.urogene.org/methprimer/index1.html) or Epidesigner (http://www.epidesigner.com/). Oligonucleotide primers and target CpG sites are available from the authors (Dr R. D. Burk).

All primers were purchased from Integrated DNA Technologies, Inc (Coralville, IA).

The extent of CpG methylation was determined by EpiTYPER (Sequenom Inc, San Diego, CA) (32) or pyrosequencing (Qiagen, Valencia, CA) (33,34), according to the manufacturers’ recommendations. In brief, 1 µg of DNA (the amount was based on the manufacturer’s recommendation) from cervical specimens documented to contain HPV16 DNA were treated with freshly prepared bisulfitr using the EZ DNA methylation kit (Zymo Research, Orange, CA) according to the suggested protocol. Bisulfite-treated DNA preferentially deaminates unmethylated cytosine (C), which are converted into uracil (U), and further converted to thymine (T) by Taq polymerase during PCR amplification; methylated C’s remain unmodified. All segments were amplified by Platinum Taq high fidelity DNA polymerase (Invitrogen, Carlsbad, CA).

For EpiTYPER, the PCR was performed using a primer with a T7 promoter tag and the PCR product was in vitro transcribed into RNA, per the Sequenom protocol (Sequenom Inc). The transcribed RNA was then enzymatically cleaved following the Sequenom protocol using MassCLEAVE reagent at the C or U positions to generate short segments, and the mass for each segment was determined and distinguished by Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS; Sequenom Inc.). The ratio between the mass of the segments with guanine (G) and/or adenine (A) can be used to estimate the extent of methylation in the original DNA (32). This approach has the advantage of being able to survey CpG sites in a relatively large DNA segment but has a relatively low resolution for clustered CpG sites.

For pyrosequencing, the bisulfite-treated DNA was amplified with a primer containing biotin, captured with streptavidin-coated beads, and annealed to a sequencing primer (available from RDB). Sequence extension was performed with PyroMark Gold Q96 Reagents (Qiagen, Valencia, CA) that uses DNA polymerase with sequential addition of each dNTP; the released pyrophosphate was catalyzed by ATP sulfurylase, luciferase, and apyrase sequentially to generate light signals, which are detected by a charge-coupled device (CCD) camera (35) on a PSQ96 system (Qiagen, Valencia, CA). Because the bioluminometric response is linear for the sequential addition of dNTPs, the signal intensity reflects the original methylation ratio producing continuous values between 0 and 1 (33). This technology is rapid and economic for clustered CpG sites but has a relatively short read length. Combining pyrosequencing and EpiTYPER provides a means to efficiently interrogate the maximum number of methylation sites. In a previous study (19), the EpiTYPER and pyrosequencing assays were compared at two CpG sites within the L1 ORF, nucleotide positions 5611 and 5617, and the Kappa scores, a statistical quantity used to measure inter-rater agreement, were 1.0 and 0.85, respectively.

Twelve blinded quality control specimens were replicated from bisulfite treatment through the methylation pyrosequencing assays to assess variability at 23 randomly chosen CpG sites in L1, L2, and URR regions. Percent methylation was categorized using tertiles and then weighted kappa scores were calculated using squared differences between the categories for the original specimen and the replication as weights. The majority of the replicated CpG sites showed moderate to high agreement with weighted Kappa scores.

Two CpG sites in the L1 genomic region, nucleotide positions 6367 and 6457, had the same nucleotide composition, and thus the same mass and could not be distinguished by EpiTYPER, so data from these sites were not analyzed. DNA from nine women failed to amplify (two control subjects, six CIN3 case subjects, and one persistence case subject). The final sample size was 99 women (HPV16 DNA clearance [control] = 34 women, CIN3 case subjects = 30 women, and HPV16 DNA persistence case subjects = 35 women).

Statistical Analysis

The Kruskal–Wallis test was used to determine whether methylation at each individual CpG site from the diagnostic and prediagnostic specimens was associated with HPV infection outcome. We used the nonparametric Kruskal–Wallis test because methylation levels were not normally distributed in the three outcome groups and this test makes minimal assumptions about the distribution of methylation. We did not order the outcomes because persistence without progression is not necessarily a precursor of CIN3. We used the Benjamini–Hochberg method for reporting associations.
in the context of multiple hypotheses testing to ensure that the expected proportion of reported associations that are false positives, or the false discovery rate, was below a nominal level, which we set at .05 (36). We limited the subsequent analysis to estimate the magnitude of only those sites that were within the Benjamin–Hochberg .05 level. Logistic regression models were used to obtain the odds ratios (ORs) and 95% confidence intervals (CIs) for CIN3 and persistence case subjects, separately using the control subjects as the referent group for each individual CpG site. For obtaining odds ratios for a CpG site, methylation levels were dichotomized using the second tertile (ie, third tertile vs first and second tertiles combined), based on the distribution for that site in the control subjects. As an alternative approach, we also dichotomized the methylation levels at the median based on the distribution in the control subjects for each individual CpG site. To obtain the odds ratios for a combination of three CpG sites, we created a categorical variable representing the number of sites with percent methylation in the top tertile of methylation (0, 1, 2, or 3) and fitted a multivariable logistic regression model with the categorical variable as the predictor. We report odds ratios for CIN3 case subjects comparing women with any one, any two, or all three CpG sites methylated (ie, third tertile of methylation) vs women with low methylation (ie, first and second tertiles of methylation combined) at these three sites. Sensitivity was estimated as the proportion of women with CIN3 who have high methylation at all three of these CpG sites, and 1-specificity as the proportion of control subjects with high methylation at all three CpG sites.

To determine whether overall methylation of each gene separately was associated with CIN3 or persistence case subjects, incorporating information from all sites within each genomic region, a global test based on the adaptive rank truncated product method of combining the site-specific \( P \) values was performed (37). The \( P \) value corresponding to the global test for each region was obtained via the efficient permutation-based algorithm described in Yu et al. (37).

Longitudinal serial specimens (diagnostic and prediagnostic specimens) for case subjects with CIN3 or with persistence were evaluated with a Wilcoxon rank sum test to determine whether the methylation levels in the two specimens differed at each CpG site. Spearman rank correlations were used to investigate associations between viral methylation levels and age at enrollment within the case subjects with CIN3 and persistence separately, and in all subjects using a linear regression model adjusted for infection outcome. Case subjects with CIN3 and persistence were grouped into the following three age categories to group together adolescents and young adults, middle-aged adults, and older women (18–25, 26–50, and >50 years, respectively).

All statistical tests were two-sided, and an \( P \) value of less than .05 was considered statistically significant. Analyses were performed with SPSS version 15.0 (SPSS Inc, Chicago, IL) and R software (http://www.r-project.org/).

### Results

**Characteristics of the Study Population**

The study population included a total of 99 women for HPV16 genome methylation analysis: 34 women in the control group (cleared HPV infection in <2 years), 30 CIN3 case subjects, and 35 persistence case subjects. As expected (26), women with persistent infection were older than women with CIN3 (persistence vs CIN3, mean [SD] = 44 [17] vs 32 [12] years), and the control group was the youngest (mean [SD] = 22 [2] years). No statistically significant correlations were detected between age and viral genome methylation at a particular CpG site (Spearman rank correlation coefficient \( = -0.4 \) to 0.5), and methylation levels among the CIN3 or persistence case subjects were similar by age groups (18–25, 26–50, and >50 years) (data not shown). Thus, we did not detect associations between age and viral methylation levels.

**Diagnostic Specimens: Associations Between Individual CpG Methylation Sites and Infection Outcomes**

Before accounting for multiple tests (unadjusted), methylation of the diagnostic specimens at 28 (41.8%) of 67 CpG sites showed statistically significant associations with infection outcome (\( P < .05 \)): two in the URR region, one in E1, four in E2/E4, two in E5, one in E6, two in E7, seven in L1, and nine in L2 (Figure 1, A, and Supplementary Table 1, available online). After accounting for multiple tests (adjusted), methylation of the diagnostic specimens at nine CpG sites located in L1, L2, or E2/E4 genomic regions were associated with infection outcomes (Figure 1, B, and Table 1). The median methylation levels by outcome group and associations with infection outcome before and after adjustment for multiple testing, methylation levels were categorized into tertiles based on the distribution in the control subjects. The first and second tertiles were combined (low methylation) and compared with the third tertile (high methylation) in logistic regression models. Compared with the control subjects, all three CpG sites in the L1 genomic region were associated with an increased risk of persistence case subjects compared with control subjects and even higher levels for CIN3 case subjects in diagnostic specimens (Table 1). For the CpG sites that showed statistically significant associations between methylation levels and infection outcome after adjustment for multiple testing, methylation levels were categorized into tertiles based on the distribution in the control subjects. The first and second tertiles were combined (low methylation) and compared with the third tertile (high methylation) in logistic regression models. Compared with the control subjects, all three CpG sites in the L1 genomic region were associated with an increased risk of CIN3 (OR = 4.44, 95% CI = 1.49 to 13.26; OR = 6.71, 95% CI = 2.19 to 20.58; and OR = 3.61, 95% CI = 1.29 to 10.15) in diagnostic specimens (Table 1).

**L1 region.** Methylation levels at three of 17 different CpG sites assayed in the L1 region were statistically significantly associated with outcome status after adjustment for multiple tests (\( P = .022, P = .025, P = .049 \) (Figure 1, B, and Supplementary Table 1, available online). At each differentially methylated CpG site, we noted increased methylation for persistence case subjects compared with control subjects and even higher levels for CIN3 case subjects in diagnostic specimens (Table 1). For the CpG sites that showed statistically significant associations between methylation levels and infection outcome after adjustment for multiple testing, methylation levels were categorized into tertiles based on the distribution in the control subjects. The first and second tertiles were combined (low methylation) and compared with the third tertile (high methylation) in logistic regression models. Compared with the control subjects, all three CpG sites in the L1 genomic region were associated with an increased risk of CIN3 (OR = 4.44, 95% CI = 1.49 to 13.26; OR = 6.71, 95% CI = 2.19 to 20.58; and OR = 3.61, 95% CI = 1.29 to 10.15) in diagnostic specimens (Table 1).

**L2 region.** Methylation at three of 15 CpG sites assayed in the L2 region were statistically significantly associated with outcome
Figure 1. Association between human papillomavirus type 16 (HPV16) methylation at individual CpG sites and infection outcome in diagnostic specimens. Associations are shown for the diagnostic specimens collected at diagnosis of cervical intraepithelial neoplasia grade 3 or persistent infection. A) The dots represent the unadjusted P values (two-sided Kruskal–Wallis test) for 67 CpG sites in the eight HPV16 genes (inset legend). B) The dots represent the P values (two-sided Kruskal–Wallis test) for 67 CpG sites after adjustment for multiple tests using the Benjamini–Hochberg method. α = .05; the horizontal dashed line represents an extension of α, all dots (ie, individual CpG P values) above this dashed line represent statistically significant P values. The −log of the P values are shown on the y-axis (eg, a value of 2 on the y-axis is a P value of .01). E1 = early gene 1; E2/E4 = early genes 2 and 4; E5 = early gene 5; E6 = early gene 6; E7 = early gene 7; L1 = late gene 1; L2 = late gene 2; URR = upstream regulatory region.

Methylation levels were higher in case subjects with persistent infection and CIN3 compared with the control subjects in diagnostic specimens (Table 1). High methylation levels (third tertile vs first and second) were associated with increased risk of CIN3 at after adjustment for multiple tests (P = .049, P = .049, P = .017) (Figure 1, B, and Supplementary Table 1, available online).

Table 1. Association of human papillomavirus type 16 (HPV16) CpG site methylation with persistence and CIN3 in diagnostic specimens*

<table>
<thead>
<tr>
<th>Genomic region</th>
<th>Cpg site†</th>
<th>Control‡ (n = 34 subjects)</th>
<th>Persistence‡ (n = 35 subjects)</th>
<th>CIN3‡ (n = 30 subjects)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Median (IQR)</td>
<td>OR§ (95% CI)</td>
<td>Median (IQR)</td>
</tr>
<tr>
<td>E2/E4</td>
<td>3412</td>
<td>0 (0.16)</td>
<td>0.016 (0.98)</td>
<td>2.46 (0.89 to 6.74)</td>
</tr>
<tr>
<td>E2/E4</td>
<td>3415</td>
<td>0 (0.15)</td>
<td>0.010 (1.00)</td>
<td>1.49 (0.54 to 4.08)</td>
</tr>
<tr>
<td>E2/E4</td>
<td>3436</td>
<td>0 (0.07)</td>
<td>0.023 (0.97)</td>
<td>2.79 (1.01 to 7.69)</td>
</tr>
<tr>
<td>L2</td>
<td>4441</td>
<td>0.076 (0.65)</td>
<td>0.156 (0.97)</td>
<td>3.17 (1.11 to 9.03)</td>
</tr>
<tr>
<td>L2</td>
<td>5179</td>
<td>0.026 (0.27)</td>
<td>0.041 (0.85)</td>
<td>4.01 (1.47 to 10.91)</td>
</tr>
<tr>
<td>L2</td>
<td>5378</td>
<td>0.098 (0.28)</td>
<td>0.124 (1.00)</td>
<td>1.63 (0.47 to 5.88)</td>
</tr>
<tr>
<td>L1</td>
<td>5611</td>
<td>0.057 (0.37)</td>
<td>0.128 (0.82)</td>
<td>4.18 (1.47 to 11.89)</td>
</tr>
<tr>
<td>L1</td>
<td>6650</td>
<td>0.010 (0.11)</td>
<td>0.040 (0.74)</td>
<td>2.2 (0.75 to 6.48)</td>
</tr>
<tr>
<td>L1</td>
<td>7034</td>
<td>0.022 (0.41)</td>
<td>0.044 (0.28)</td>
<td>3.14 (1.17 to 8.41)</td>
</tr>
</tbody>
</table>

* The diagnostic specimen was collected from subjects at diagnosis of CIN3 (ie, precancer) or persistent infection. Median methylation levels, quantified using pyrosequencing assays, and the interquartile range, the difference between 75th and 25th percentiles, are shown for individual CpG sites. CI = confidence interval; CIN3 = cervical intraepithelial neoplasia grade 3; E1 = early gene 1; E2/E4 = early genes 2 and 4; E5 = early gene 5; E6 = early gene 6; E7 = early gene 7; IQR = interquartile range; L1 = late gene 1; L2 = late gene 2; OR = odds ratio.
† CpG sites that had statistically significant (P < .05) differences in methylation levels among outcome groups after adjustment for multiple tests in the diagnostic specimens.
‡ Women who cleared their HPV16 infection in less than 2 years were control subjects; persistence refers to the case subjects with more than 2 years of persistent HPV16 infection without development of CIN2 or higher grades (CIN2+); CIN3 refers to the case subjects with HPV16 infection and histologically confirmed CIN3.
§ Logistic regression models were used to estimate the associations between high methylation and persistence or CIN3 separately using the control subjects as the referent group. For these estimates, methylation levels were dichotomized using the second tertile, based on the distribution for that CpG site in the control subjects.
|| The association between methylation at each CpG site with infection outcome group (clearance, persistence, CIN3) after adjustment for multiple tests in the diagnostic specimens. P values were calculated using a two-sided Kruskal–Wallis test.
all three CpG sites compared with the control subjects, and the strongest association was noted at nucleotide position 5378 (OR = 11.12, 95% CI = 2.29 to 76.80) (Table 1). High levels of methylation were also associated with an increased risk of persistence at these CpG sites (Table 1). The associations with increased methylation were similar when methylation was dichotomized at the median (data not shown). Globally, methylation levels in the L2 genomic region were statistically significantly associated with HPV16 infection outcome (global P = .003).

**E2/E4 region.** Methylation levels at three of five CpG sites assayed in the E2 and E4 (overlapping ORF region; E2/E4) were statistically significantly associated with outcome status after adjustment for multiple testing (P = .022, P = .017, P = .049) (Figure 1, B, and Supplementary Table 1, available online). Methylation levels were highest in case subjects with CIN3 in diagnostic specimens (Table 1). We noted that all three CpG sites with high methylation were associated with a statistically significant increased risk of CIN3 (OR = 5.25, 95% CI = 1.77 to 15.61; OR = 3.82, 95% CI = 1.33 to 10.94; OR = 3.29, 95% CI = 1.16 to 9.34) and persistence compared with the control subjects (Table 1). Globally, methylation levels in the E2/E4 genomic region were statistically significantly associated with infection outcome (global P = .002).

Methylation analysis of diagnostic specimens showed that for the three CpG sites associated with the greatest risks of CIN3 in L1 (nucleotide position 6650; OR = 6.71, 95% CI = 2.19 to 20.58), L2 (nucleotide position 5378; OR = 11.12, 95% CI = 2.29 to 76.80), and E2/E4 (nucleotide position 3412; OR = 5.25, 95% CI = 1.77 to 15.61), the risk associated with high methylation (third tertile) at all three sites was much higher than that associated with low methylation at these three sites (OR = 52, 95% CI = 4.0 to 670). Among those with data on all three of these sites (a total of 23 control subjects and 22 CIN3 case subjects) and for having all three CpG sites methylated, the sensitivity was 60% (95% CI = 36% to 79%) and the specificity was 91% (95% CI = 72% to 99%) for CIN3. An increased risk of CIN3 was observed for women with high methylation at any one (OR = 3.0, 95% CI = 0.26 to 35.3) or two (OR = 8.0, 95% CI = 0.71 to 89.9) of the three CpG sites compared with women with low methylation at these sites.

After adjustment for multiple tests, no statistically significant association was noted between CpG methylation levels within the URR, E6, E7, or E1 regions and infection outcome.

**Prediagnostic Specimens: Analysis for Outcome Prediction**

Because we found statistically significant methylation differences among outcome groups using the HPV16 specimens collected at diagnosis, we prospectively examined the first prediagnostic HPV16 specimen collected to determine whether HPV16 methylation levels predicted outcome.

Based on the prediagnostic specimens, methylation levels at nine CpG sites were statistically significantly (P < .05) associated with HPV16 outcome status before accounting for multiple tests: two in the URR region, one in E7, four in L2, and two in L1 genomic regions (Figure 2, A and Supplementary Table 2, available online). Three of these CpG sites in L2 (nucleotide positions 856, 860, and 865).
4261, 4270, and 4277) and one in L1 (nucleotide position 5611) had methylation levels statistically significantly associated with outcome in both the diagnostic and prediagnostic specimens, with higher methylation levels in case subjects diagnosed with CIN3, and methylation levels at these CpG sites were not different between the diagnostic and prediagnostic specimens. Two CpG sites, one in L2 and one in L1, were statistically significantly associated with outcome after adjustment for multiple tests (Figure 2, B, and Table 2). A very high risk of CIN3 was associated with high methylation for the CpG site in L2 (nucleotide position 4261) compared with the control subjects (OR = 5.49, 95% CI = 1.49 to 20.16) (Table 2).

For case subjects with persistent infections, the diagnostic and prediagnostic specimens did not differ in methylation levels at most sites (Supplementary Table 3, available online); however, methylation was statistically significantly (P < .05) higher in the persistent diagnostic specimens at L1 (16.4%) of 67 CpG sites. In case subjects who developed CIN3, the CIN3 diagnostic specimens had statistically significantly (P < .05) higher methylation levels at 23 (34.3%) of 67 CpG sites compared with the CIN3 prediagnostic specimens. No sites showed statistically significantly lower methylation levels for the diagnostic compared with the prediagnostic specimens from either group.

**Discussion**

In this case–control study, we examined HPV16 DNA methylation at CpG sites in all viral genomic regions using specimens collected at the time of diagnosis and also before diagnosis from a set of HPV16-infected women in the prospective Guanacast, Costa Rica, cohort. In diagnostic specimens, we identified multiple CpG sites in L1, L2, and E2/E4 that have statistically significantly different methylation levels among HPV16 outcome groups. Methylation levels were elevated in case subjects with persistence without progression to CIN2+ and highest in case subjects with CIN3 compared with control subjects at all sites with statistically significantly different methylation levels. Combining high methylation at CpG sites in L1, L2, and E2/E4 showed a stronger association with CIN3 than any single site. In prediagnostic specimens, high methylation at a CpG site in L2 was strongly associated with risk of developing CIN3.

HPV16 is a relatively common sexually transmitted infection that causes greater than half of all cervical cancer. Nevertheless, only a small fraction of women who are infected with HPV16 develop histologically confirmed CIN3 or cancer. Our data show that distinct viral DNA methylation patterns are associated with viral clearance, viral persistence without progression to precancer or cancer, and CIN3. However, the differences in methylation levels between CIN3 case subjects and viral clearance control subjects were more pronounced at the time of CIN3 diagnosis than before diagnosis, suggesting that viral methylation patterns may be CIN3 diagnostic biomarkers.

Our findings using a large population-based cohort extend observations from smaller and less comprehensive studies (summarized in Table 3). Brandsma et al. (17) observed a pattern in 13 women of high methylation in HPV16 E5, L2, and L1 regions associated with high-grade lesions. Fernandez et al. (23) detected hypermethylation of L1 and L2 with tumorigenesis in specimens collected from 18 women. These results are consistent with the CpG methylation patterns we observed in the late genes. We could not confirm a statistically significant association with methylation levels in E5 among our outcome groups (8), although we did detect statistically significant associations with increased methylation in CIN3 before adjustment for multiple tests at the two CpG sites we examined in E5. Others have studied only the L1 and URR regions and observed hypermethylation of CpGs within L1 in high-grade and carcinoma lesions (15,25). L1 has also been shown to be hypermethylated in HPV18 carcinomas (38,39) and in HPV16- and HPV18-dependent penile carcinomas (40). We, in fact, have unpublished HPV18 data showing increased methylation at multiple CpG sites in L1 associated with CIN3+ suggesting methylation of the HPV genome with

### Table 2: Association of human papillomavirus type 16 (HPV16) CpG site methylation with persistence and CIN3 in prediagnostic specimens

<table>
<thead>
<tr>
<th>Genomic region</th>
<th>CpG site†</th>
<th>Control‡ (n = 34 subjects)</th>
<th>Persistence‡ (n = 35 subjects)</th>
<th>CIN3‡ (n = 20 subjects)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Median (IQR)</td>
<td>Median (IQR)</td>
<td>Median (IQR)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>OR (95% CI)</td>
<td>OR (95% CI)</td>
</tr>
<tr>
<td>L2</td>
<td>4261</td>
<td>0.036 (0.15)</td>
<td>0.082 (0.45)</td>
<td>2.87 (1.00 to 8.19)</td>
</tr>
<tr>
<td>L1</td>
<td>7136</td>
<td>0.132 (0.95)</td>
<td>0.0 (0.49)</td>
<td>0.24 (0.04 to 1.13)</td>
</tr>
</tbody>
</table>

* The prediagnostic specimen was collected from subjects at the first HPV16-positive screening visit. Median methylation levels, quantified using pyrosequencing assays, and the interquartile range, the difference between 75th and 25th percentiles, are shown for individual CpG sites. CI = confidence interval; CIN3 = cervical intraepithelial neoplasia grade 3; IQR = interquartile range; L1 = late gene 1; L2 = late gene 2; OR = odds ratio.
† CpG sites that had statistically significant (P < .05) differences in methylation levels among outcome groups after adjustment for multiple tests in the prediagnostic specimens.
‡ Women who cleared their HPV16 infection in less than 2 years were control subjects; persistence refers to the case subjects with more than 2 years of persistent HPV16 infection without development of CIN2 or higher grade (CIN2+); CIN3 refers to the case subjects with HPV16 infection and histologically confirmed CIN3.
§ Logistic regression models were used to estimate the associations between high methylation and persistence or CIN3 separately using the control subjects as the referent group. For these estimates, methylation levels were dichotomized using the second tertile, based on the distribution for that CpG site in the control subjects.
|| The association between methylation at each CpG site with infection outcome group (clearance, persistence, CIN3) after adjustment for multiple tests in the prediagnostic specimens. P values were calculated using a two-sided Kruskal–Wallis test.
Table 3. Literature review and comparison of the human papillomavirus type 16 (HPV16) methylation studies published to date*

<table>
<thead>
<tr>
<th>First author, year (reference)</th>
<th>Cross-sectional or case–control</th>
<th>Study design</th>
<th>Methylation assay</th>
<th>Sample size</th>
<th>Location of CpGs examined (No.)</th>
<th>Infection outcomes examined</th>
<th>Main finding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Badal, 2003 (12)</td>
<td>Yes</td>
<td>Cross-sectional</td>
<td>Restriction enzyme, bisulfite sequencing</td>
<td>81 (enzyme), 15 (sequencing)</td>
<td>URR (11)</td>
<td>Asymptomatic, CIN1, CIN3, invasive carcinoma</td>
<td>Methylation of URR and E6 decreases with progression</td>
</tr>
<tr>
<td>Kalantari, 2004 (15)</td>
<td>Yes</td>
<td>Cross-sectional</td>
<td>Bisulfite sequencing†</td>
<td>115</td>
<td>URR and part of L1 (19)</td>
<td>Asymptomatic, CIN1, CIN2, CIN3, squamous carcinoma</td>
<td>Methylation was highest in carcinomas</td>
</tr>
<tr>
<td>Bhattacharjee, 2006 (16)</td>
<td>Yes</td>
<td>Cross-sectional</td>
<td>Restriction enzyme, bisulfite sequencing</td>
<td>72 (enzyme), 13 (sequencing)</td>
<td>URR</td>
<td>Cytologically normal and cervical cancer</td>
<td>Methylation was significantly higher in cases compared with controls</td>
</tr>
<tr>
<td>Fernandez, 2009 (23)</td>
<td>Yes</td>
<td>Cross-sectional</td>
<td>Bisulfite sequencing</td>
<td>18</td>
<td>Genome (110)</td>
<td>Asymptomatic, CIN, cervical carcinoma</td>
<td>Methylation is progressively methylated with disease progression, particularly at L1 and L2</td>
</tr>
<tr>
<td>Ding, 2009 (13)</td>
<td>Yes</td>
<td>Cross-sectional</td>
<td>Bisulfite sequencing</td>
<td>53</td>
<td>URR (15)</td>
<td>LSIL, HSIL, SCC</td>
<td>Successively increased methylation of the URR with disease</td>
</tr>
<tr>
<td>Kalantari, 2009 (25)</td>
<td>Yes</td>
<td>Cross-sectional</td>
<td>Bisulfite sequencing</td>
<td>7</td>
<td>URR and part of L1 (19)</td>
<td>Normal, squamous and glandular cell carcinomas</td>
<td>Heterogeneous methylation patterns, and hypermethylation of L1 in advanced disease</td>
</tr>
<tr>
<td>Brandsma, 2009 (17)</td>
<td>Yes</td>
<td>Cross-sectional</td>
<td>Bisulfite sequencing</td>
<td>13</td>
<td>Genome (113)</td>
<td>Negative, ASC-US, LSIL, CIN1, and CIN2/3</td>
<td>Samples with the highest frequency of methylation, particularly in E5/L2/L1, were more severe disease</td>
</tr>
<tr>
<td>Hublarova, 2009 (14)</td>
<td>Yes</td>
<td>Cross-sectional</td>
<td>Restriction enzyme</td>
<td>121</td>
<td>URR (16)</td>
<td>Asymptomatic, CIN1, CIN2,3, invasive carcinoma</td>
<td>Methylome frequency decreases with progression of disease</td>
</tr>
<tr>
<td>Piyathilake, 2011 (18)</td>
<td>No</td>
<td>Cross-sectional</td>
<td>Pyrosequencing</td>
<td>75</td>
<td>URR/E6 (6)</td>
<td>CIN1 or less, CIN2+</td>
<td>Higher methylation was associated with a lower likelihood of CIN2+</td>
</tr>
<tr>
<td>Sun, 2011 (19)</td>
<td>Yes</td>
<td>Cross-sectional</td>
<td>EpiTYPER, pyrosequencing</td>
<td>85</td>
<td>URR, L1 (32)</td>
<td>less than CIN2, CIN2, CIN3+</td>
<td>Increased methylation of L1 was associated with CIN3+</td>
</tr>
<tr>
<td>This study</td>
<td>Yes</td>
<td>Cross-sectional</td>
<td>Pyrosequencing</td>
<td>108</td>
<td>Genome (67)</td>
<td>Cleared infections, persistent infection without CIN2+, CIN3</td>
<td>Increased methylation of L1/L2/E2/E4 with disease progression, and L1/L2 methylation predicts outcome</td>
</tr>
</tbody>
</table>

* We performed a literature search using PubMed and MEDLINE through February 30, 2011, using combinations of the search terms “HPV16” or “human papillomavirus type 16” and “methylation.” ASC-US = atypical squamous cells of undetermined significance; CIN = cervical intraepithelial neoplasia; E2 = early gene 2; E5 = early gene 5; E6 = early gene 6; HSIL = high-grade squamous intraepithelial lesion; L1 = late gene 1; L2 = late gene 2; LSIL = low-grade squamous intraepithelial lesion; SCC = squamous cell carcinoma; URR = upstream regulatory region.
† Bisulfite sequencing in this study included DNA cloning and sequencing.
increasing grades of cervical neoplasia may be a general phenomenon.

A prospective evaluation revealed that methylation changes at two CpGs in L2 and L1 were statistically significantly associated with infection outcome. The other study with a prospective component (see Table 3) only examined six CpGs in the URR region and found that higher methylation was associated with a reduced risk of being diagnosed with CIN2+ (18). We found that methylation at two CpG sites in the URR region were statistically significant but not after accounting for multiple tests. Other studies showed conflicting results: Some noted higher methylation in the URR region (often at E2-binding sites) in high-grade lesions and cancer cells (13,15,16,23), whereas others reported a lower level of methylation in the URR region in high-grade lesions and cancer (12,14). Discrepancies in these data may be due to differences in methylation assays, heterogeneous and small sample sizes, and/or disease misclassification.

Using both diagnostic and prediagnostic serial specimens, we identified higher methylation at CpG sites in L2 and L1 associated with an increased risk of CIN3. These proteins are expressed in the upper differentiated layers of the epithelium after genome replication and amplification (41,42). L1 and L2 are expressed during the early productive infection and are progressively less transcribed in the later phases of infection progression and transformation (42). The increased methylation we observed at CpG sites in L1 and L2 genomic regions in long-term infections (infections that persist with and without CIN3) may be contributing to the downregulation of these genes or might simply reflect neoplastic progression and/or loss of cell differentiation.

HPV16 DNA is found integrated into the cellular DNA in many cancers. HPV integration is considered a late event in the progression from CIN3 to cancer. Van Tine et al. (43) showed that silent integrated HPV16 copies can be reactivated by treatment with a demethylating agent in vitro, which suggests that HPV DNA methylation, at least in part, may be responsible for silencing transcription during neoplastic progression associated with integration.

We also report that higher methylation at CpG sites in E2/E4 at diagnosis were statistically significantly associated with CIN3. E2 is required for viral replication, and it is a transcription factor that can contribute to the repression of the viral oncoproteins E6 and E7 (44). E2 repression of E6 and E7 is reduced after integration and disruption of the E2 ORF. The function of E4 is not completely understood, but it is thought to be required for viral amplification and possibly viral assembly and release (42,45). Loss of E2 and E4 expression usually occurs with tumorigenesis. Increased methylation in these genomic regions in CIN3 may downregulate their expression, which may facilitate neoplastic progression (42).

Because increased viral DNA methylation has been observed in many cancer associated DNA viruses, including HPV16/18, the Epstein–Barr virus (46), Human Hepatitis B Virus (23), and human T-lymphotropic virus 1 (47), methylation may be a shared mechanism in response to exogenous viral DNA that also leads to cell survival and propagation. Others have suggested that HPV16 DNA methylation may be a biomarker of integration and disease progression, in particular hypermethylation of L1 (48). However, integration is not necessary for neoplastic transformation of HPV16-infected cells. A more interesting possibility is that methylation of the virus may predict which infections clear or persist with or without progression to CIN3. We observed that higher methylation is present before diagnosis of CIN3, possibly up to years earlier. Perhaps, methylation masks the virus from the immune system, as described for Epstein–Barr virus (49,50). It is possible that HPV16 actively regulates DNA methylation by recruiting DNA methyltransferases via the E7 oncoproteins (22) and uses methylation to block viral antigen presentation to escape immune surveillance.

A limitation of our study is the modest sample size and possible variability in methylation, which may limit our ability to detect an association. However, our study size is similar or larger than others who have detected an association with methylation of the E6 promoter.

In summary, we present diagnostic and prediagnostic HPV16 genome methylation data associated with infection outcome and suggest that methylation of CpG sites within L1, L2, and E2/E4 are biomarkers of or possibly mechanistically involved in viral progression to precancer and cancer. Viral CpG methylation of L1, L2, and E2/E4 might alter gene expression and thereby influence host response and pathogenesis. We have replication and follow-up studies underway to confirm our findings in a larger cohort of women and to test the generalizability of these observations to other carcinogenic HPV types.

References


Funding

This work was supported in part by the Intramural Research Program of the National Institutes of Health (NIH), National Cancer Institute (SU01CA078527 to RDB); Division of Cancer Epidemiology and Genetics Intramural Research Award (to LM) and through use of core facilities (CS, RDB) of the Einstein-Montefiore Center for AIDS funded by the NIH (AI-51519) and the Einstein Cancer Research Center (P30CA013330).

Notes

The authors are solely responsible for the study design, data collection, analysis and interpretation of the data, writing the article, and decision to submit the article for publication.

Affiliations of authors: Division of Cancer Epidemiology and Genetics, Department of Health and Human Services, National Cancer Institute, National Institutes of Health, Rockville, MD (LM, AG, MS, NW, AH, SW); Department of Pediatrics, Albert Einstein College of Medicine, Bronx, NY (CS, RDB); Proyecto Epidemiológico Guanacaste, Fundación INCIENSA, San José, Costa Rica (ACR, RH); Departments of Microbiology and Immunology, Epidemiology and Population Health, and, Obstetrics, Gynecology and Women’s Health, Albert Einstein College of Medicine, Bronx, NY (RDB); Queen Mary, University of London, Centre for Cancer Prevention, Wolfson Institute of Preventive Medicine, London, UK (AL).