Correlation of Human Papillomavirus Type 16 and Human Papillomavirus Type 18 E7 Messenger RNA Levels with Degree of Cervical Dysplasia

Michael E. Scheurer,1 Guillermo Tortolero-Luna,2,4 Martial Guillaud,6 Michele Follen,2,3 Zhuo Chen,7 Laura M. Dillon,5 and Karen Adler-Storthz5

Departments of Epidemiology and Gynecologic Oncology, and Biomedical Engineering Center, The University of Texas M.D. Anderson Cancer Center; Division of Occupational and Environmental Health, The University of Texas School of Public Health; Department of Diagnostic Sciences, The University of Texas Dental Branch, Houston, Texas; 6Department of Cancer Imaging, British Columbia Cancer Research Centre, The British Columbia Cancer Agency, Vancouver, British Columbia, Canada; and 7Department of Hematology and Oncology, Emory University Winship Cancer Institute, Atlanta, Georgia

Abstract

Infection with certain types of human papillomavirus (HPV) is a necessary event in the development of cervical carcinoma; however, not all women who become infected with HPV will progress to cancer. Much is known about the molecular influence of HPV E6 and E7 proteins on the malignant transformation. Little is known about the additional factors needed to drive the process. Quantitative real-time PCR was used to quantify mRNA expression of the E7 gene in women exhibiting normal epithelium, low-grade squamous intraepithelial lesions (LSIL), and high-grade squamous intraepithelial lesions (HSIL). Prevalence of mRNA transcripts was lower among normal women (27%) than for women with LSIL (40%) and HSIL (37%). Mean levels ranged from 2.0 (in scale per 20 ng cDNA) among normal women to 4.2 among those with HSIL, with a significant trend (P = 0.008). This trend was only significant for HPV 18 transcripts if separately analyzed by HPV type. The transcriptional activity of HPV 18 is higher than that of HPV 16 and increases with increasing level of dysplasia. This is in concert with the findings of other studies, and reinforces the notion that HPV 18 is a more aggressive viral type. Real-time PCR of viral transcripts could provide a more efficient method to analyze the oncogenic potential within cells from a cervical swab, thus providing a way to better screen women who may progress to higher grade lesions or invasive carcinoma from those who will spontaneously regress. (Cancer Epidemiol Biomarkers Prev 2005;14(8):1948–52)

Introduction

Infection with certain types of human papillomavirus (HPV) is recognized as a necessary event in the development of cervical carcinoma, the third most common malignancy in women worldwide. However, infection with HPV does not guarantee an elevated risk of cancer. HPV is a ubiquitous infection, and an estimated 80% of all women will become infected with HPV at some point in their lives (1). However, very few of these infections will lead to dysplastic changes in cervical tissue. Of those that do develop mild dysplasia (LSIL), ~11% will progress to severe dysplasia (HSIL), and an equal fraction of those high-grade lesions will progress to invasive carcinoma (2). Epidemiologic and laboratory data have provided several insights into the molecular events that drive the malignant transformation. The HPV oncoproteins E6 and E7 are responsible for the ability of the virus to transform cells; however, the additional factors that contribute to that transformation still evade investigators. Uncontrolled expression of these two genes results from the integration of the viral genome into the host cell DNA.

Although integration is proposed as a mechanism for progression from cervical intraepithelial neoplasia to carcinoma (3), it is unknown at what point this event occurs. Integration is commonly seen in invasive cervical carcinoma; however, a consensus has not been reached for high-grade lesions. Estimates of integration vary from 5% to 100% in HSIL. Most studies have yielded no evidence of integration in low-grade lesions; however, a few have shown integration in LSIL and in histologically normal tissue (reviewed in ref. 4). More important may be the point at which transcriptional expression of these oncogenes pushes a cell toward malignancy.

In this study, quantitative real-time reverse transcription-PCR (RT-PCR) was used to isolate and quantitate mRNA of the E7 oncogene from HPV 16 and HPV 18 in cervical samples from women with abnormal Pap smears exhibiting normal histology and from women with intraepithelial lesions. Our hypothesis was that the amount of mRNA increases with degree of dysplasia.

Materials and Methods

Study Population. The cervical specimens used in this study were collected as part of a larger study to evaluate emerging optical technologies for cervical neoplasia. The study population consisted of women with abnormal Pap test results attending the colposcopy clinics at University of Texas M.D. Anderson Cancer Center, Lyndon B. Johnson General Hospital, and Memorial Hermann Hospital in Houston, Texas, and the British Columbia Cancer Agency in Vancouver, British Columbia, Canada, between October 2000 and July 2003. Institutional Review Board approval was received from each institution involved in the study, and participants gave written informed consent before being enrolled in the study.

For the present study, all study participants were identified as positive for HPV16 and/or HPV18 DNA by PCR. Cases...
were women classified by biopsy as having squamous intraepithelial lesions (SIL). Those with cervical intraepithelial neoplasia I or HPV-associated changes were classified as low-grade squamous intraepithelial lesions (LSIL); those with cervical intraepithelial neoplasia II or cervical intraepithelial neoplasia III were classified as high-grade squamous intraepithelial lesions (HSIL). Noncases were women classified with a normal biopsy or as having atypical squamous cells of undetermined significance.

**Data and Specimen Collection.** A demographic and epidemiologic risk factor questionnaire was given by the study nurse at each institution on each participant’s enrollment in the study. Clinical specimens were collected for various laboratory procedures, including histopathologic confirmation of disease, quantitative cytopathology and histopathology, HPV typing, and HPV DNA and mRNA analyses. Study physicians collected specimens for HPV DNA and RNA analyses using an endocervical cytobrush. DNA samples were placed in 250 µL PBS with 0.02% sodium azide immediately following cytologic sampling and stored at −80°C until extracted, ~1 month after collection. RNA samples were placed in 250 µL of lysis solution (Ambion, Austin, TX) immediately following cytologic sampling, and stored at −80°C until extracted; length of storage varied from 5 months to 3 years after collection.

**DNA Isolation and Detection.** Viral DNA was extracted from cervical cytobrush specimens using a commercially available kit (QiAamp DNA Mini Kit, Qiagen, Valencia, CA). Extracted DNA was stored at −80°C for no more than 2 weeks before PCR was done. Following the methods of Manos et al. (5), we then analyzed the samples for HPV DNA using MY9 and MY11, consensus HPV primers that amplify a 450 bp (5), we then analyzed the samples for HPV DNA using MY9 and MY11, consensus HPV primers that amplify a 450 bp fragment. PCR products were resolved by agarose gel electrophoresis, transferred to nylon membranes (Bio-Rad Laboratories, Hercules, CA), and hybridized to a 32P-labeled HPV consensus probe. Consensus probe-positive samples were then hybridized to 32P-labeled specific HPV 16 and HPV 18 probes on separate nylon membranes. Sample positivity was assessed by autoradiography following hybridization. DNAs extracted from HPV 18–positive HeLa cells, HPV 16–positive CaSkI cells, and a negative control without DNA were used as controls in the PCR and subsequent hybridization.

**RNA Isolation and Quantitation.** RNA was extracted as total mRNA from cervical cytobrush specimens using a commercially available kit (RNAqueous, Ambion), and the mRNA was then reverse transcribed into cDNA (RETROscript, Ambion) within 1 week of extraction. cDNA samples were purified using a NucAway Spin Column (Ambion) to remove unused deoxynucleotide triphosphates and any other DNA or RNA material <25 bases in length. The cDNA was quantified by spectrophotometry on the same day it was generated and was stored at −80°C for no more than 1 month before analysis by real-time PCR. An equal quantity of cDNA (20 ng) from each sample was analyzed by RT-PCR for quantitation of HPV 16 and HPV 18 E7 oncocene expression (cDNA levels reflect mRNA levels). Using the same amount of total cDNA (20 ng) from each sample allows for accurate comparison of the target mRNA between the samples independent of amount of sample collected. Samples were analyzed by RT-PCR according to the DNA results for each specimen. For example, if the sample was positive for HPV 16 DNA and negative for HPV 18 DNA, RT-PCR was done using only the HPV 16 primers. The RT-PCR technique used in this study was an absolute quantitative real-time PCR assay using the SYBR Green I (Molecular Probes, Eugene, OR) fluorescent intercalation dye. Each amplification experiment was done in a 96-well PCR plate covered with optical tape in the iCycler iQ real-time PCR instrument (Bio-Rad). A final volume of 25 µL was used containing 20 ng of cDNA template, 12.5 µL of iQ SYBR Green Supermix (Bio-Rad), 1.25 µL containing 0.2 µmol of either HPV 16 or HPV 18 forward and reverse primers, and water. A description of the construction of forward and reverse primers used can be found elsewhere (6). The reaction was subjected to denaturation at 95°C for 2 minutes followed by 40 cycles of denaturation at 95°C for 45 seconds, annealing at 62°C for 45 seconds, and elongation at 72°C for 45 seconds. Fluorescent data were specified for collection at the end of the elongation step in each cycle. SYBR Green I binds and intercalates into double-stranded DNA during the extension step of the amplification cycle.

To do an absolute quantitative RT-PCR, plasmids constructed with specific binding sites for HPV 16 and HPV 18 E7 primers were used to construct the standard curves for each set of patient samples that were assayed (6). A set of serial dilutions of the plasmids (1 × 10⁸, 1 × 10⁷, 1 × 10⁶, 1 × 10⁵, 1 × 10⁴, 1 × 10³, and 1 × 10² copies/µL) was used as PCR templates for the standard curve. Each standard dilution was tested in triplicate, and each patient sample was tested in duplicate (due to the low yield of the mRNA extraction process). Positive (SiHa for HPV 16 and HeLa for HPV 18) and negative controls were included in each experiment to ensure reproducible results. All mRNA assays were done by the investigator in a central laboratory.

For quality control purposes of the PCR assays, a small reproducibility study was done. Because we started the PCR with the same amount of total cDNA from each specimen, we were more concerned with the reproducibility of the PCR step rather than the RT step. As previously stated, the RT step was separately done from the PCR step to ensure that the quantity of each sample assayed was equivalent. Twelve samples were assayed on two separate days at least 1 week apart, and a one-factor random-effect ANOVA was done to estimate the coefficient of variation. A coefficient of variation of <20% is considered adequate for the reproducibility of the assay (7). The coefficient of variation calculated for this assay was 13.2%.

To assess the quality of the cDNA for RT-PCR, a random sample (n = 20) of the mRNA negative samples was analyzed by RT-PCR using SYBR Green I for the presence of the GAPDH gene. A random sample was done due to small amount of cDNA available for most specimens. All 20 samples were positive for GAPDH. Therefore, we are reasonably certain that, although HPV mRNA was not found, the cDNA was of sufficient quality to detect it if present.

**Control Cell Lines.** The cervical cancer cell lines HeLa and SiHa were purchased from American Type Culture Collection (Manassas, VA). HeLa cells express HPV 18 mRNA, and SiHa cells express HPV 16 mRNA. Both cell lines were grown in Eagle’s MEM with 10% fetal bovine serum and penicillin-streptomycin. The cells were incubated at 37°C in a 5% CO₂ atmosphere, and directly harvested in lysis buffer (Ambion). Total mRNA was isolated from the harvested cells using the same kit used for the patient samples (RNAqueous, Ambion), and cDNA was similarly generated (RETROscript, Ambion).

**Statistical Analyses.** Demographic characteristics were summarized, and differences by disease status were determined by the χ² or Fisher’s exact test for categorical variables or by ANOVA for continuous variables. Copy number of mRNA for each HPV type was transformed using the natural logarithm to minimize positive skewness. The Kruskal-Wallis nonparametric ANOVA method was used to discern differences in the mean mRNA levels among the categories of dysplasia (normal, LSIL, and HSIL). Cuzick’s nonparametric test for trend was used to assess the presence of a trend in
Table 1. Demographic characteristics of study population by histologic grade

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Total n = 340</th>
<th>Normal n = 223</th>
<th>LSIL n = 74</th>
<th>HSIL n = 43</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.760</td>
</tr>
<tr>
<td>White</td>
<td>208 (61%)</td>
<td>135 (61%)</td>
<td>43 (58%)</td>
<td>30 (70%)</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>46 (14%)</td>
<td>29 (13%)</td>
<td>11 (15%)</td>
<td>6 (14%)</td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>57 (17%)</td>
<td>38 (17%)</td>
<td>15 (20%)</td>
<td>4 (9%)</td>
<td></td>
</tr>
<tr>
<td>Other*</td>
<td>29 (9%)</td>
<td>21 (9%)</td>
<td>5 (7%)</td>
<td>3 (7%)</td>
<td></td>
</tr>
<tr>
<td>Age [mean (SD)]</td>
<td>41 (12.9)</td>
<td>45 (12.0)</td>
<td>35 (11.3)</td>
<td>32 (10.0)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.023</td>
</tr>
<tr>
<td>Not married</td>
<td>147 (44%)</td>
<td>83 (37%)</td>
<td>41 (56%)</td>
<td>23 (56%)</td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>190 (56%)</td>
<td>140 (63%)</td>
<td>32 (44%)</td>
<td>18 (44%)</td>
<td></td>
</tr>
<tr>
<td>Annual income</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.004</td>
</tr>
<tr>
<td>&lt;$20,000</td>
<td>47 (14%)</td>
<td>23 (10%)</td>
<td>14 (19%)</td>
<td>10 (24%)</td>
<td></td>
</tr>
<tr>
<td>$20,000-$49,999</td>
<td>112 (33%)</td>
<td>70 (31%)</td>
<td>30 (41%)</td>
<td>12 (29%)</td>
<td></td>
</tr>
<tr>
<td>$50,000+</td>
<td>143 (42%)</td>
<td>111 (50%)</td>
<td>22 (30%)</td>
<td>10 (24%)</td>
<td></td>
</tr>
<tr>
<td>Study site:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Houston</td>
<td>312 (92%)</td>
<td>222 (99%)</td>
<td>67 (91%)</td>
<td>23 (53%)</td>
<td></td>
</tr>
<tr>
<td>Vancouver</td>
<td>28 (8%)</td>
<td>1 (1%)</td>
<td>7 (9%)</td>
<td>20 (47%)</td>
<td></td>
</tr>
</tbody>
</table>

* Includes Native American, Asian, and other categories.

mRNA amounts across the levels of dysplasia. Dysplasia was scored for ranking as 1, 2, and 3 for normal, LSIL, and HSIL, respectively. All analyses were done using Intercooled STATA 8.2 (Stata Corp., College Station, TX).

Results

At the time the current study began, a total of 1,477 specimens had been collected for mRNA analysis. Of these, 940 (64%) were negative for dysplasia, 281 (19%) had LSIL, 228 (15%) had HSIL, and 28 (2%) were missing a definitive histologic diagnosis. DNA analysis by PCR had been completed for 870 participants, and, of those, 378 were positive for HPV 16 and/or HPV 18 DNA. Thirty samples were lost during the extraction process for various reasons; however, there was no significant difference in the histologic category of these samples (data not presented). Of the 348 samples remaining for quantitation, two were excluded because they were identified by histology to be from patients with squamous cell carcinoma. Therefore, 346 participants were identified for mRNA quantitation. Of those, 46% were positive for HPV 16 DNA alone, 27% were positive for HPV 18 DNA alone, and 29% were positive for both HPV 16 and HPV 18 DNA (data not presented). Of these, six were excluded from analysis due to missing histology data, leaving 340 for analysis of mRNA level by histologic grade.

Histologic analysis of biopsies from each patient revealed 223 (65%) subjects were negative for dysplasia, 74 (22%) had LSIL, and 43 (13%) had HSIL. This distribution of histologic grades is similar to that of the original 1,477 samples that were originally collected; therefore, any selection bias is unlikely. Demographic characteristics differed some by histologic group (see Table 1). Women in the normal category were older (mean age, 45 years), married (63%), had a higher income (>$50,000), and mostly from the Houston sites (99%) when compared with the LSIL and HSIL groups. Women identified with LSIL and HSIL were closer in age; mean ages were 35 and 32 years, respectively. Also of note was that women in the HSIL group were more likely to be from the Vancouver site and 32 years, respectively. Also of note was that women in the HSIL group were more likely to be from the Vancouver site and 32 years, respectively.

Of the 340 patients for whom RNA was quantitated, 105 (31%) yielded a positive result for either HPV 16 or HPV 18 mRNA (see Table 2). The prevalence by histologic grade varied slightly, 27% for normals, 41% for LSIL, and 37% for HSIL. However, these differences were not statistically significant (P = 0.074). Mean mRNA levels differed significantly between histologic categories; this significance seems to be due to higher levels among the HPV 18–positive samples.

Discussion

To our knowledge, this is one of the first studies to quantify the amount of viral mRNA present in normal and precancerous cervical epithelium. A few studies have looked for the presence of viral transcripts and have shown an increase in prevalence of HPV E6 and/or E7 mRNA transcripts (8-11). Whereas the current study found a significant difference in prevalence of mRNA transcripts between normal and dysplastic tissue, there was not as clear a difference between low-grade and high-grade lesions. However, mean levels of copy number of HPV 18 E7 mRNA did increase steadily from normal to low-grade to high-grade. This agrees with results from Wang-Johanning et al. (10) who reported that copy number of both RNA and DNA increased with increasing SIL grade.

The overall prevalence of E7 mRNA transcripts in the current study (31%) is lower than that of other more recent studies. Lamarcq et al. (12) reported an overall prevalence of 47% in HSIL, and Kraus et al. (11) found 46% of the biopsies tested contained mRNA. The inclusion of normal tissue in the current study likely accounts for the lower overall prevalence because detectible transcription from normal cells is less likely. The previous two studies did not include normal specimens and concentrated heavily on those with HSIL. In addition, the prevalence in HSIL (37%) is lower than those studies including only HSIL in the sample. This is possibly related to sample size differences between these studies and the current one. One study (12) included a total of 15 samples from HSIL patients, and the other (11)
Table 2. mRNA prevalence and amount by HPV type and histology

<table>
<thead>
<tr>
<th>HPV Type</th>
<th>Total</th>
<th>Normal</th>
<th>LSIL</th>
<th>HSIL</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive HPV 16 or 18 mRNA</td>
<td>410</td>
<td>223</td>
<td>74</td>
<td>41</td>
<td>0.074</td>
</tr>
<tr>
<td>Mean mRNA (SD)*</td>
<td>2.5 (2.6)</td>
<td>2.0 (2.1)</td>
<td>2.5 (2.6)</td>
<td>4.2 (3.7)</td>
<td>0.008</td>
</tr>
</tbody>
</table>

*Mean ln (mRNA) present in 20 ng of total cDNA from sample.

included 168 patients with HSIL. The current study included a much larger overall sample size (n = 340); however, the number of HSIL patients was much lower (n = 43). This difference in prevalence among HSIL specimens could also be due to the number of cervical scrapes that were taken at the study visit before collection for mRNA analysis, which was the last one taken. Therefore, it is possible that there were fewer viable squamous cells left for collection at that time. Methods to reuse collected material or combine tests would greatly enhance the yield from cervical samples for future studies. An example of this is the use of liquid-based cytology to perform Pap smears, after which the cells are available for HPV typing. These methods would be extremely helpful in the cervix because so few viable cells are collected during sampling. It is unlikely that storage time of specimens played an important role in prevalence estimates. There was no significant correlation (Pearson's r = 0.099) between length of storage and mRNA amount detected (data not shown). We also feel it is unlikely that storage temperature or collection media used played a significant role because both are standard protocol suggested by the manufacturer (Ambion), which is a leader in the collection and analysis of RNA for research purposes.

Results from the current study also indicate that the transcriptional activity of HPV 18 E7 is more indicative of the molecular activity in precancerous lesions than that of HPV 16. It has long been noted that the transforming ability of HPV 18 exceeds that of HPV 16, thus lending molecular evidence to the observation that HPV 18 is a more aggressive type. HPV 18–containing carcinomas have poorer prognosis than those with HPV 16 (13, 14), and HPV 18 is more efficient at immortalizing cervical cells (15, 16) and phosphorylation of the retinoblastoma protein (17, 18).

Most studies to date, including the current one, have been cross-sectional analyses of cervical material, either swabs or biopsies. However, to truly investigate the series of events leading up to malignancy, prospective studies are needed. Cuschieri et al. (19) recently reported that the detection of E6 or E7 transcripts in baseline samples helped predict those patients who were likely to carry a persistent infection. Further elucidation of these findings utilizing quantity of mRNA could be more revealing. There may be a threshold of production above which persistence is guaranteed. It would then be helpful to discern those factors that contribute to lasting mRNA transcription.

Misclassification of the main variable of interest (mRNA level) is a possibility in this study. However, quantitative real-time PCR was used for this purpose, and a small reproducibility study of the PCR assay done before the onset of the study revealed an acceptable coefficient of variation (13%). As this was a cross-sectional study, classification of the mRNA variable was based on one specimen taken at the time of the study visit. This does not take into account the potential temporal fluctuation in mRNA during the natural history of the lesions. To fully understand this phenomenon, a longitudinal repeated-measure cohort study would need to be done to increase the accuracy of the measurements and reduce any bias in the assessment of the cumulative exposure. On the other hand, misclassification of the outcome, level of dysplasia, is unlikely. Two independent pathologists evaluated the biopsies for all participants, and a third pathologist resolved any discrepancies.

One aspect of these results that should be considered in future studies is if samples that are determined HPV DNA negative by PCR show any significant increases in mRNA expression. Identification of HPV DNA by PCR was used simply as a screening process for the presence of HPV in the current study. Primers for the PCR detection of HPV DNA were based on the L1 viral gene, encoding the major structural protein, whereas the mRNA expression in this study was based on the E7 gene, of which expression is required for maintenance of malignancy. This is especially useful in screening for HPV DNA because the region of the L1 gene used is more highly conserved across viral types. It is possible that differences in detection exist based solely on the gene being used for the PCR assay.

Given all this information, the detection of high-risk HPV transcripts could serve as a marker for risk of the development of cervical cancer. It has been shown that E7 promotes the formation of benign lesions whereas E6 works to complete the malignant transformation (20). If the factors that contribute to the ability of E7 to advance benign lesions were identified, there might be a way to identify women at risk for more intensive screening versus those in whom the screening process could be minimized. There are several advantages to using RT-PCR for this purpose. First, typically with cervical sampling, very small amounts of cells or tissue are retrieved. RT-PCR can quantitate gene expression from very small amounts of starting material. In addition, the ease of the procedure lends itself to use in large-scale epidemiologic studies in which hundreds of subjects can be analyzed. Older studies utilizing fluorescent in situ hybridization and radiolabeled blotting techniques were too labor-intensive for these large studies (21) and lacked the sensitivity needed to detect the transcripts in low-grade and normal tissue (22). With continuing efforts to improve the reliability of multiplex RT-PCR reactions, multiple HPV types could be tested in one reaction, thus further decreasing the amount of cellular material needed for these studies.

In summary, the results presented here suggest that quantitative real-time PCR presents an effective means by which expression of the viral oncoproteins can be assessed in a high-throughput manner facilitating the discovery of markers that predict cervical progression. The main goal of this type of research is to hone the screening capabilities for HPV-related carcinoma. Several suggestions have been made on additional markers to be included in cervical cancer screening (HPV typing, HPV viral load, etc.); however, much research is still needed to meet the goal of a highly sensitive and specific screening program.

Acknowledgments
We thank all of the women who participated in the P01 study; without them our work could not be done. We also wish to thank the support of the P01 in both Houston and Vancouver, especially the study nurses who collected the samples and Nan Earle and Trey Kell for their help with data management.
References


