Published in final edited form as:

Gynecol Oncol. 2007 October ; 107(1 Suppl 1): S163–S169. doi:10.1016/j.ygyno.2007.07.050.

Epidemiologic Modeling of Cervical Dysplasia with Molecular and Cytopathological Markers

Michael E. Scheurer¹, Martial Guillaud², Guillermo Tortolero-Luna³, Michele Follen⁴, and Karen Adler-Storthz⁵

¹The University of Texas M. D. Anderson Cancer Center, Department of Epidemiology, Houston, Texas, USA

²The British Columbia Cancer Agency, BC Cancer Research Centre, Department of Cancer Imaging, Vancouver, British Columbia, Canada

³The University of Texas School of Public Health at Houston, Division of Occupational and Environmental Health and The University of Texas M. D. Anderson Cancer Center, Department of Gynecologic Oncology, Houston, Texas, USA

⁴The University of Texas M. D. Anderson Cancer Center, Department of Gynecologic Oncology and Biomedical Engineering Center, Houston, Texas, USA

⁵The University of Texas Dental Branch, Department of Diagnostic Sciences, Houston, Texas, USA

Abstract

Objectives—Conventional cervical screening is insufficient at identifying patients who are likely to progress from cervical dysplasia to carcinoma. Traditional epidemiologic studies have identified potential factors to aid in the discrimination between those lesions likely to progress from those likely to regress; however, there is still much to be learned. To examine the role of traditional epidemiologic factors in conjunction with molecular markers of human papillomavirus activity, we studied a group of women attending colposcopy clinics in Houston, TX and Vancouver, BC between October 2000 and July 2003.

Methods—Quantitative real-time PCR was used to measure mRNA expression of the human papillomavirus E7 gene, and quantitative cytology was used to gather information about the DNA index and chromatin features of the cells from these women. Logistic regression was used to establish predictor variables for histologic grade based on the epidemiologic risk factors and the molecular markers.

Results—The most predictive factors were mRNA level, DNA index, parity, and age. The ROC curve for the individual logits indicated excellent discrimination.

Conclusion—In accordance with other authors, these results suggest that molecular markers of the malignant process should be included in analyses looking to predict the progression potential of cervical lesions.

Reprints: Michael E. Scheurer, Ph.D., M.P.H., Department of Epidemiology, M. D. Anderson Cancer Center, 1155 Pressler St, Unit 1340, Houston, TX 77030, mescheur@mdanderson.org.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

human papillomavirus; mRNA; RT-PCR; cytometry; uterine cervix

Cervical cancer is the third most common malignancy in women worldwide. In 2007, 11,150 cases of invasive cervical cancer will be diagnosed among U.S. women, and approximately 3,700 U. S. women will die from this neoplasm.[1] The incidence and mortality rates for cervical cancer are much higher in many developing countries, where approximately 80 percent of all cases occur.[2] Recent worldwide estimates show that, in 2002, 493,243 women were diagnosed with cervical cancer worldwide, and 273,505 deaths from cervical cancer occurred. [2] In addition, over a five-year period over 1.4 million women around the world had cervical cancer.[2] Epidemiological and laboratory data have established the role of human papillomavirus (HPV) in the etiology of cervical cancer, and the molecular events that drive the malignant transformation are well documented.

Additionally, epidemiologic investigations have identified several risk factors for the progression of HPV-infected epithelium to high-grade lesions and cervical carcinoma.[3] Among the most consistent risk factors are oral contraceptive use, parity, cigarette smoking, and infection with certain STIs, especially *Chlamydia trachomatis* and herpes simplex virus 2 (HSV2).[4;5] It is not completely understood how these factors interact with the HPV-induced molecular events to drive the malignant transformation. In spite of the recently approved vaccines against HPV infection by two of the oncogenic types of HPV (16 and 18), the need still exists to identify risk factors and markers of progression for cervical cancer.

We quantitated HPV mRNA expression and cytometric features from cervical swab specimens. These molecular markers were then used along with the epidemiologic risk factors to build a polytomous regression model to predict level of dysplasia. This was done to ascertain which of the traditional epidemiologic risk factors continue to play an important role in carcinogenesis when considering the molecular events that occur.

Methods

Study design and population

This analytical cross-sectional study was performed using data from a multi-center Phase II clinical trial which employed fluorescence and reflectance point spectroscopy to diagnose cervical disease. Overall study designs, protocols, and preliminary results from the parent study are presented in detail elsewhere.[6] The current study assessed the potential role of HPV mRNA levels and nuclear morphometric characteristics (e.g., DNA index and chromatin condensation) as covariates in the classification of HPV-associated cervical dysplasia. The study population consisted of women with abnormal Papanicolaou test results attending the colposcopy clinics at M. D. Anderson Cancer Center, Lyndon B. Johnson General Hospital, and Memorial Hermann Hospital in Houston, Texas, USA, and the British Columbia Cancer Agency in Vancouver, British Columbia, Canada, between October, 2000, and July, 2003.

The study was approved by the IRB at each institution, and participants gave written informed consent prior to enrollment. All study participants were identified as positive for HPV16 and/or HPV18 DNA by PCR. Women with a normal biopsy or having a cytological diagnosis of atypical squamous cells of undetermined significance (ASCUS) were grouped into the Normal category. Women with cervical intraepithelial neoplasia (CIN) I or HPV-associated changes were grouped as low-grade squamous intraepithelial lesions (LSIL), those with CIN II or CIN III were grouped as high-grade SIL (HSIL).

Data and specimen collection

A demographic and epidemiologic risk factor questionnaire was completed upon each participant's enrollment in the study. Clinical specimens were collected for: histopathologic confirmation of disease, quantitative cyto- and histopathology, HPV typing, and HPV mRNA analyses. Study physicians collected specimens for HPV DNA and mRNA analyses using an endocervical cytobrush as described previously. [7]

Nucleic acid isolation and quantitation

Details of HPV DNA and HPV E7 mRNA quantitation have been reported elsewhere. [7] Briefly, viral DNA was extracted from cervical cytobrush specimens using a commercially available kit (Qiagen DNA Mini Kit, Qiagen, Valencia, CA) and analyzed for HPV L1 gene consensus sequences by PCR, followed by specific typing with HPV 16 and HPV 18 probes. Total mRNA was extracted from cervical cytobrush specimens using a commercially available kit (RNAqueous, Ambion, Austin, TX, USA), and reverse transcribed into cDNA (RETROscript, Ambion, Austin, TX, USA). An equal quantity of cDNA (20 ng) from each sample was analyzed by absolute real-time PCR for quantitation of HPV 16 and HPV 18 E7 mRNA expression according to the type of HPV DNA found in the sample. The cervical cancer cell lines, HeLa and SiHa were used as controls in the PCR reactions, as previously described. [7]

Quantitative cytology

Cytologic specimens were temporarily stored at 4°C in liquid-based fixative. ThinPrep slides were made from the cell suspension and stained for quantitative analysis using a Feulgen-based preparation, which is stoichiometric for DNA. All cytometric measures were performed in Vancouver, B.C., Canada, in a central laboratory.

Digital images of the Feulgen-stained nuclei were collected using the Cyto-Savant imaging system (Cancer Imaging, Vancouver, B.C., Canada). Slides were automatically loaded onto the microscope stage in batches of 50 and exhaustively scanned for abnormal (non-diploid) cells. A maximum of 2,000 diploid cells were measured per slide at a wavelength of 600 nm, using a 20× 0.75 aperture plan objective lens.

The digital images were then sorted using a multi-step classification system. First, cell objects were sorted by DNA content, shape, and size features using binary decision trees based on training datasets from previous studies into normal (diploid), abnormal (DNA index greater than 1.5), and “junk” (debris, overlapping cells, etc.) groups. DNA Index is the normalized DNA amount of each cell; a value of 1 indicates a diploid specimen. The normal cells were used as the diploid standard for normalization performed automatically during scanning. Second, after scanning, a cytotechnologist verified the normalization, removed any remaining “junk” from the diploid and non-diploid cell groups. Third, a cytopathologist reviewed the non-diploid cells microscopically and further classified them into sub-categories: “non-diploid” (“true abnormal”), “normal non-diploid cells”, and reactive cells.

A two-level quality control procedure for automatic cytometry was used. First, every slide batch included a Quality Assurance (QA) slide for the Feulgen staining process. Second, a reference slide containing multiple objects of the same known optical density and size was used to assess consistency of the scanning process.

After classification and quality control, nuclear feature measurements were performed on the digital cell images according to computations described elsewhere.[8] Three categories of features (Table 1) were measured for each cell image: Morphological features characterize increases in nuclear size and severe distortions in nuclear shape that are associated with

aneuploid progression of SIL. DNA content features estimate the absolute intensity, optical density levels of the nucleus (which are proportional to the DNA content of the cell), and the intensity distribution characteristics. Texture features describe the variations in optical intensity over the nuclear image and present an objective and quantitative method for characterization of changes in chromatin appearance.

For our multivariate analyses of the quantitative features, we were most interested in measuring DNA index and chromatin texture. The DNA index was calculated over the entire cell population on each slide. The chromatin texture feature was calculated as a multi-component score for only the normal diploid cell population as follows. We defined integrated HPV-associated changes (iHAC) as changes in chromatin features induced by expression of the E7 protein. Under the hypothesis that textural features measured by cytometry can discriminate those cells with E7 expression, we generated an individual cell score, which was a phenotypic measure of the degree and intensity of deviation of an epithelial cell with integrated HPV DNA. On the two extremes of this cellular scale, we first defined a group of normal diploid cells (DNA index between 0.9 – 1.1) selected from HPV DNA-positive and RNA-negative cervical specimens and then a group of cells from HPV DNA-positive and RNA-positive cervical specimens. By training the system on these two sets of phenotypically identical cells, we can be certain that the changes we detect are not due to some underlying process, in this case HPV infection. A stepwise linear discriminant analysis was performed between these two groups to select the most relevant cytometric features. The features selected by this process included: fractal_dimen, contrast, lowDNAcomp, fractal1_area, cl_prominence, avg_short_runs, lowVSmed_DNA, hiDNAamnt, den_dark_spot, and correlation. Correct classification by the discriminant analysis was 59 percent. This is similar to our findings for other cervical samples and lung tissue. A canonical score (iHAC score) based on the discriminant analysis was then calculated for each individual cell in each specimen. The threshold that maximized the discrimination of the training set was used as the cutoff value for an iHAC-positive cell in the test set. The chromatin changes were then reported as the percentage of cells for which the iHAC score was below the threshold. All statistical analyses for the generation of the iHAC score were performed with the STATISTICA package (StatSoft, Inc., Tulsa, OK, USA).

Statistical analyses and regression modeling

Differences in demographics by disease status were determined by Chi square test for categorical variables or analysis of variance (ANOVA) for continuous variables. Copy number of mRNA for each HPV type was transformed using the natural logarithm to minimize right-handed skewness. The Kruskal-Wallis nonparametric ANOVA method was used to discern differences in the mean mRNA levels, DNA index, and chromatin score among the categories of dysplasia (normal, LSIL, HSIL). Cuzick's nonparametric test for trend was used to assess the presence of a trend in mRNA levels, DNA index, and chromatin score across the levels of dysplasia. Dysplasia was scored for ranking as 1, 2, and 3 for normal, LSIL, and HSIL, respectively.

Polytomous and ordinal logistic regression modeling was performed utilizing the methods of Hosmer and Lemeshow.[9] The baseline polytomous model compares HSIL to Normal and LSIL to Normal. Three ordinal models were used: an adjacent category ordinal model comparing HSIL to LSIL and LSIL to Normal; a continuous ratio model comparing HSIL to LSIL/Normal and LSIL to Normal; and a proportional odds model comparing HSIL to LSIL/Normal and HSIL/LSIL to Normal. A list of potential covariates was established based on the study hypothesis and suggestions from the literature.[4;5] These included: RNA amount, DNA index, iHAC score, current smoking habits, excessive alcohol use (defined as reporting at least one occasion during the last month on which the participant drank more than 5 drinks, marital status (in married or married-like situation vs. not), parity, menopausal status, current OC use,

history of *C. trachomatis* and HSV-2, and number of sexual partners within the last year. Variables significant at the 0.20 alpha level in univariate analyses were considered for inclusion in the final model. Proper functional form (e.g., linear, dichotomous) was determined utilizing the Quartile Method.[9] Diagnostic graphs (residuals, leverage, and influence) were plotted to assess the fit of the final model for all covariate patterns. Any covariate patterns found to have unnecessary influence on the model were excluded from the analysis. An overall goodness-of-fit test was performed and an ROC curve plotted for each logit (Figure 1) of the most parsimonious model. Analyses were performed in Stata, v8.2 (Stata Corp, College Station, TX), and two-sided p-values are reported.

Results

When the current study began, 1,477 specimens had been collected for mRNA analysis (Figure 1). Analyses by PCR for HPV DNA was complete for 870 specimens; of which, 378 were positive for HPV 16 and/or HPV 18. From these, six samples were excluded due to missing histology, and two were excluded due to a diagnosis of squamous cell carcinoma. Thirty samples were subsequently lost during the mRNA extraction process; however, there was no significant difference in the histologic category of these samples (data not presented). Finally, 340 samples were available for mRNA analysis. Of those, 46 percent showed HPV 16 infection alone, 25 percent showed HPV 18 infection alone, and 29 percent showed co-infection with both types.

According to biopsy, 217 subjects were negative for dysplasia, 74 had LSIL, and 43 had HSIL. This distribution of histologic grades is similar to that of the 1,477 samples that were originally collected; therefore, any selection bias is unlikely (data not shown). Demographic characteristics differed slightly by histology (Table 2). Women in the normal category were older (mean age 45 years), married (59 percent), had a higher income (> \$50,000 [49 percent]), and mostly from the Houston site (77 percent) when compared to the LSIL and HSIL groups. Women identified with LSIL and HSIL were closer in age; mean ages were 36 and 32 years, respectively. Also of note, was that women in the HSIL group were more likely to be from the Vancouver site than any of the other groups. This is mostly due to the differences in standard of care for low-grade lesions between USA and Canada. There were no differences between the groups based on race and education level.

The initial logistic model included: total RNA (continuous as the natural log), DNA index (categorical), iHAC score (dichotomous), age (continuous), marital status (dichotomous), annual household income (dichotomous), current smoker (dichotomous), heavy drinker (dichotomous), parity (dichotomous), current OC use (dichotomous), number of sexual partners during the last year (dichotomous), history of *C. trachomatis* infection (dichotomous), history of genital warts (dichotomous), and history of HSV2 infection (dichotomous). All potential variables from the univariate analyses were consideration for inclusion in the final model (Table 3). Beginning with a model containing all potential covariates, the variable with the least significant p-value was removed and tested using the likelihood-ratio test until all variables left in the model significantly (at alpha=0.05) contributed to the model. The final baseline category polytomous model included the following variables: total RNA, DNA index, and number of pregnancies. Age, marital status, annual household income, and study site were assessed as potential confounders; however, only age significantly affected the model. The most significant variable in predicting histological grade was DNA index, and increasing age showed a significant decrease in prevalence of HSIL and LSIL when compared to normal (Table 4). Results from the three ordinal models (Table 5) support the findings from the polytomous model.

Discussion

In general, it is believed that higher grades of dysplasia have a higher probability of progression; however, epidemiologic and laboratory studies have been unable to predict the progression of precancerous lesions to carcinoma. The use of conventional cytology has also been unable to predict the progression of lesions in individual patients.[10] Poor intra- and inter-observer reproducibility of cytologic screening has been improved with the use of liquid-based cytologic methods, and investigators are developing semi-automated quantitative cytology based on digital image analysis of the specimens. However, quantitative pathology methods are not widely accepted by cytologists and pathologists, and screening programs are expensive to maintain due to the amount of resources that go to evaluating atypias and low-grade lesions. In addition, these analyses have not yet been able to identify a single marker of progression.

We report the use of conventional epidemiologic risk factors for HPV infection with molecular markers for carcinogenesis (mRNA expression and quantitative cytology) to model cervical dysplasia. To our knowledge, this is the first report to utilize both image cytometry and expression of HPV E7 mRNA from both HPV 16 and HPV 18. Some reports exist on the expression of viral oncogenes in cervical cancers and high-grade lesions, and others examined the relation of viral load to epidemiologic risk factors. However, evidence supporting viral load as a risk factor for progression is still unclear. Most studies associate a high viral load with high-grade cervical lesions [11-16]; however, newer studies suggest it is a predictor of low-grade lesions [16-18] and ASCUS [16] as well.

Giuliano et al. identified several epidemiologic factors (age, number of sexual partners, marital status, contraceptive use, and Chlamydial infection) as important predictors of infection with high-risk HPVs.[19] They also identified HPV viral load as a significant predictor of LSIL and HSIL; while parity was important for LSIL but not HSIL.[20] Our results also show age and parity as important predictors when considered with DNA index and mRNA level; however, parity appeared to convey a protective effect in our analysis. The effect with parity could be related to the older age of the women in the normal cytological category compared to the LSIL and HSIL groups.

Rajeevan et al. examined the effects of the E6/E7 transcripts in HPV 16-infected women. [21] They showed in a multivariate analysis that having high viral load, both E6 and E6*I transcripts, and age greater than 25 were all associated with CIN III; while only increased viral load was associated with CIN II. Smoking was the only other epidemiologic factor found significant, however this was in univariate analysis and only held for CIN III. Our univariate analysis also showed smoking to be highly related to both LSIL and HSIL. In addition, we showed that marital status, annual income, heavy alcohol use, number of pregnancies, current OC use, and number of sexual partners in the last year were also significant factors for LSIL and/or HSIL in univariate analysis. However, in the multivariate models these factors were excluded by the addition of the molecular factors. The only epidemiologic factor that remained in our model was number of pregnancies, with higher parity being protective for both LSIL and HSIL.

Castellsague and Muñoz suggest that cofactors of HPV infection can be classified into 3 groups: environmental or exogenous factors (OC use, tobacco use, diet, and STIs), viral factors (HPV type and variant, viral load, and viral integration), and host factors (endogenous hormones, genetic factors, and immune response). [4] In order to fully understand the risk factors that contribute to the progression of low- and high-grade cervical lesions, we must consider all three types of cofactors. The current study is the first step in this process; it includes the molecular markers that represent viral expression and consequences on cervical cells and the

traditional epidemiologic markers of environmental risk. The next logical step would be to include those factors related to host immune response. Much of the basic immunology concerning HPV infection is known.[22;23] However, the additional factors contributing to host immune response, or lack thereof, are illusive. The investigation of markers for immune response would provide that third level of evidence needed to fully investigate the factors contributing to progression.

The ultimate goal of epidemiologic modeling of progression is to improve the screening process for HPV-related disease and prevent cancer. The predictive power of the current study was excellent for HSIL (area under ROC curve = 0.9913) and was very good for LSIL (area under ROC curve = 0.8148) (Figure 2). The sample size for this analysis was sufficient to power the findings (ten events per variable in the model is sufficient).[9] With 43 subjects in the HSIL group, which was the least populated, there is enough power for the four predictor variables in the final model. Important factors identified for the cervix could also be important in other tissues, both other anogenital tissues and the oropharynx. As mentioned previously, there are several women each year whose Papanicolaou testing reveals some abnormality; however, the majority of these women would not develop cancer from these lesions. The results of the current study suggest that the molecular markers for progression are just as important, if not more so, than the traditional risk factors, and point to the importance of looking at host response markers as well as viral factors for the control of the malignant transformation in the cervix.

Acknowledgments

The laboratory work involved with this project was supported by the National Cancer Institute, grant 3P01CA82710, Michele Follen, MD, PhD, principal investigator. Michael E. Scheurer was supported by a cancer prevention fellowship, National Cancer Institute, grant R25CA57730, Robert R. Chamberlain, PhD, principal investigator.

References

1. American Cancer Society. Cancer Facts and Figures 2007. Atlanta (GA): American Cancer Society; 2007.
2. Ferlay, J.; Bray, F.; Pisani, P.; Parkin, DM. GLOBOCAN 2002: Cancer Incidence, Mortality and Prevalence Worldwide. IARC CancerBase No 5 version 2.0. IARC Press; Lyon: 2004. 9-13-2006
3. Scheurer ME, Tortolero-Luna G, Adler-Storthz K. Human papillomavirus infection: biology, epidemiology, and prevention. *Int J Gynecol Cancer* 2005;15(5):727–746. [PubMed: 16174218]
4. Castellsague X, Munoz N. Chapter 3: Cofactors in human papillomavirus carcinogenesis--role of parity, oral contraceptives, and tobacco smoking. *J Natl Cancer Inst Monogr* 2003;(31):20–28. [PubMed: 12807941]
5. Castle PE, Giuliano AR. Chapter 4: Genital tract infections, cervical inflammation, and antioxidant nutrients--assessing their roles as human papillomavirus cofactors. *J Natl Cancer Inst Monogr* 2003;(31):29–34. [PubMed: 12807942]
6. Nath A, Rivoire K, Chang S, West L, Cantor SB, Basen-Engquist K, et al. A pilot study for a screening trial of cervical fluorescence spectroscopy. *Int J Gynecol Cancer* 2004;14(6):1097–1107. [PubMed: 15571615]
7. Scheurer ME, Tortolero-Luna G, Guillaud M, Follen M, Chen Z, Dillon LM, et al. Correlation of human papillomavirus type 16 and human papillomavirus type 18 e7 messenger RNA levels with degree of cervical dysplasia. *Cancer Epidemiol Biomarkers Prev* 2005;14(8):1948–1952. [PubMed: 16103442]
8. Hanselaar AG, Poulin N, Pahlplatz MM, Garner D, MacAulay C, Matisic J, et al. DNA-cytometry of progressive and regressive cervical intraepithelial neoplasia. *Anal Cell Pathol* 1998;16(1):11–27. [PubMed: 9584897]
9. Hosmer, DW.; Lemeshow, S. *Applied Logistic Regression*. 2nd. New York, NY: Wiley-InterScience; 2000.

10. Bocking A, Nguyen VQ. Diagnostic and prognostic use of DNA image cytometry in cervical squamous intraepithelial lesions and invasive carcinoma. *Cancer* 2004;102(1):41–54. [PubMed: 14968417]
11. Bosch FX, Munoz N, de Sanjose S, Navarro C, Moreo P, Ascunce N, et al. Human papillomavirus and cervical intraepithelial neoplasia grade III/carcinoma in situ: a case-control study in Spain and Colombia. *Cancer Epidemiol Biomarkers Prev* 1993;2(5):415–422. [PubMed: 8220085]
12. Josefsson AM, Magnusson PK, Ylitalo N, Sorensen P, Qwarforth-Tubbin P, Andersen PK, et al. Viral load of human papilloma virus 16 as a determinant for development of cervical carcinoma in situ: a nested case-control study. *Lancet* 2000;355(9222):2189–2193. [PubMed: 10881891]
13. Morrison EA, Ho GY, Vermund SH, Goldberg GL, Kadish AS, Kelley KF, et al. Human papillomavirus infection and other risk factors for cervical neoplasia: a case-control study. *Int J Cancer* 1991;49(1):6–13. [PubMed: 1874571]
14. Swan DC, Tucker RA, Tortolero-Luna G, Mitchell MF, Wideroff L, Unger ER, et al. Human papillomavirus (HPV) DNA copy number is dependent on grade of cervical disease and HPV type. *J Clin Microbiol* 1999;37(4):1030–1034. [PubMed: 10074522]
15. Ylitalo N, Sorensen P, Josefsson AM, Magnusson PK, Andersen PK, Ponten J, et al. Consistent high viral load of human papillomavirus 16 and risk of cervical carcinoma in situ: a nested case-control study. *Lancet* 2000;355(9222):2194–2198. [PubMed: 10881892]
16. Flores R, Papenfuss M, Klimecki WT, Giuliano AR. Cross-sectional analysis of oncogenic HPV viral load and cervical intraepithelial neoplasia. *Int J Cancer* 2006;118(5):1187–1193. [PubMed: 16152619]
17. Lorincz AT, Castle PE, Sherman ME, Scott DR, Glass AG, Wacholder S, et al. Viral load of human papillomavirus and risk of CIN3 or cervical cancer. *Lancet* 2002;360(9328):228–229. [PubMed: 12133661]
18. Schlecht NF, Trevisan A, Duarte-Franco E, Rohan TE, Ferenczy A, Villa LL, et al. Viral load as a predictor of the risk of cervical intraepithelial neoplasia. *Int J Cancer* 2003;103(4):519–524. [PubMed: 12478669]
19. Giuliano AR, Papenfuss M, Abrahamsen M, Inserra P. Differences in factors associated with oncogenic and nononcogenic human papillomavirus infection at the United States-Mexico border. *Cancer Epidemiol Biomarkers Prev* 2002;11(9):930–934. [PubMed: 12223441]
20. Giuliano AR, Papenfuss M, De Galaz EM, Feng J, Abrahamsen M, Denman C, et al. Risk factors for squamous intraepithelial lesions (SIL) of the cervix among women residing at the US-Mexico border. *Int J Cancer* 2004;109(1):112–118. [PubMed: 14735476]
21. Rajeevan MS, Swan DC, Nisenbaum R, Lee DR, Vernon SD, Ruffin MT, et al. Epidemiologic and viral factors associated with cervical neoplasia in HPV-16-positive women. *Int J Cancer* 2005;115(1):114–120. [PubMed: 15688414]
22. Gravitt PE, Shah, KV. The Biology of Human Papillomavirus Infections. In: Rohan, TE, Shah, KV, editors. *Cervical Cancer: From Etiology to Prevention*. Boston: Kluwer Academic Publishers; 2004.
23. Woodworth CD. HPV innate immunity. *Front Biosci* 2002;7:d2058–d2071. [PubMed: 12165480]

Abbreviations

ASCUS	abnormal squamous cells of undetermined significance
cDNA	complementary deoxyribonucleic acid
CIN	cervical intraepithelial neoplasia
DNA	deoxyribonucleic acid
HPV	HPV, Human papillomavirus
HSIL	high-grade squamous intraepithelial lesion
iHAC	integrated HPV-associated changes
IOD	integrated optical density

IRB	institutional review board
LSIL	low-grade squamous intraepithelial lesion
mRNA	messenger ribonucleic acid
PCR	polymerase chain reaction
POR	prevalence odds ratio
RNA	ribonucleic acid
SIL	squamous intraepithelial lesion

Figure 1.
Flow chart of specimen inclusion in current analysis.

Figure 2.
ROC Curves for Individual Logits of Baseline Category Model. The curve for each logit in the polytomous model exhibited good (LSIL) to excellent (HSIL) fit for the data.

Table 1
Cytometric features measured for each cell

Category	Features
Morphological Features (43)	
Size (4)	area, mean_radius, variance_radius, maximum_radius
Shape (5)	eccentricity, spericity, elongation, compactness, inertia_shape
Boundaries (34)	low_freq_fft, freq_low_fft, harm01-32_fft
DNA Content (5)	DNA_index, OD_max, OD_var, OD_skew, OD_kurt
Texture Features (55)	
Discrete texture (20)	low, medium, high DNA amount
	low, medium, high DNA area
	low, medium, high, medium_high DNA compactness
	low, medium, high, medium_high DNA average distance
	low, medium, high density object
	low, medium, high center mass
	low_vs_medium, low_vs_high, low_vs_medium-high DNA
Markovian texture (7)	entropy, energy, contrast, correlation, homogeneity, cl_shade, cl_prommence
Non-Markovian texture (5)	density_light_spots, density_dark_spots, center_of_gravity, range_extreme, range_average
Fractal texture (3)	fractal_area1, fractal_area2, fractal_dimension
Run length texture (20)	short_runs_mean, short_run_stdv, short_run_min, sort_run_max, long_runs_mean, long_run_stdv, long_run_min, long_run_max, gray_level_mean, gray_level_stdv, gray_level_min, gray_level_max, run_length_mean, run_length_stdv, run_length_min, run_length_max, run_percent_mean, run_percent_stdv, run_percent_min, run_percent_max

Table 2

Demographic characteristics of women with abnormal Papanicolaou smear tested for quantitation of HPV mRNA by histological grade, Houston, TX, USA and Vancouver, British Columbia, Canada, October, 2000 - July, 2003

Characteristic	Total N=334	Normal N=218	LSIL N=74	HSIL N=42	p-value
Race					0.456
White	204 (62%)	131 (60%)	43 (59%)	30 (73%)	
Black	46 (14%)	29 (13%)	11 (15%)	5 (15%)	
Hispanic	57 (17%)	38 (18%)	15 (21%)	4 (10%)	
Other	25 (7%)	20 (4%)	4 (1%)	1 (2%)	
Age [Mean (SE)]	42 (0.72)	45 (0.84)	36 (1.4)	32 (1.3)	<0.001
Education					0.095
<HS	13 (4%)	5 (2%)	3 (4%)	5 (13%)	
HS/Some college	178 (53%)	117 (54%)	40 (55%)	21 (53%)	
College degree	87 (26%)	61 (28%)	18 (25%)	8 (20%)	
Advanced degree	38 (11%)	28 (13%)	7 (10%)	3 (7%)	
Refused/don't know	15 (5%)	7 (3%)	5 (7%)	3 (7%)	
Marital Status					0.007
Married	186 (56%)	136 (62%)	32 (44%)	18 (45%)	
Not married	145 (44%)	82 (38%)	41 (56%)	22 (55%)	
Annual Income					0.002
<\$20,000	47 (14%)	23 (11%)	14 (19%)	10 (25%)	
\$20,000-49,999	111 (34%)	69 (32%)	30 (41%)	12 (30%)	
\$50,000 +	139 (42%)	107 (49%)	22 (30%)	10 (25%)	
Refused/don't know	34 (10%)	19 (9%)	7 (10%)	8 (20%)	
Study Site					<0.001
Houston	306 (92%)	217 (99%)	67 (91%)	22 (53%)	
Vancouver	28 (8%)	1 (1%)	7 (9%)	20 (47%)	

Table 3

Results of univariate polytomous regression of histological grade, Houston, TX, USA and Vancouver, British Columbia, Canada, October, 2000 - July, 2003

Variable	HSIL vs. Normal		LSIL vs. Normal	
	POR (95% CI)	p-value	POR (95% CI)	p-value
In Total RNA	1.26 (1.08-1.47)	0.003	1.18 (1.02-1.37)	0.026
DNA Index				
Diploid	1.00		1.00	
Aneuploid	10.25 (4.87-21.56)	<0.001	1.72 (0.81-3.68)	0.160
iHAC Score				
<37	1.00		1.00	
=37	1.17 (0.61-2.27)	0.633	1.82 (1.05-3.16)	0.033
Age	0.90 (0.86-0.93)	<0.001	0.93 (0.91-0.96)	<0.001
Marital Status				
Not married	1.00		1.00	
Married	0.47 (0.24-0.93)	0.029	0.47 (0.27-0.81)	0.006
Annual Income				
<\$50,000	1.00		1.00	
\$50,000 +	0.39 (0.18-0.87)	0.021	0.43 (0.24-0.77)	0.005
Current Smoker				
No	1.00		1.00	
Yes	5.09 (2.35-11.03)	<0.001	2.89 (1.45-5.76)	0.003
Heavy Drinker				
No	1.00		1.00	
Yes	6.91 (2.75-17.35)	<0.001	2.61 (1.12-6.08)	0.026
No. Pregnancies				
= 2	1.00		1.00	
3 +	0.50 (0.26-0.96)	0.038	0.74 (0.43-1.26)	0.262
Current OC use				
No	1.00		1.00	
Yes	2.72 (1.18-6.26)	0.019	2.48 (1.24-4.95)	0.010
Num. sexual partners in last year				
0-1	1.00		1.00	
2+	6.01 (2.36-15.33)	<0.001	2.71 (1.07-6.83)	0.035
History of <i>C. trachomatis</i>				
No	1.00		1.00	
Yes	0.43 (0.05-3.38)	0.421	1.54 (0.56-4.26)	0.407
History of Genital Warts				
No	1.00		1.00	
Yes	2.60 (0.99-6.80)	0.052	1.32 (0.52-3.35)	0.560
History of Genital Herpes (HSV2)				
No	1.00		1.00	

Variable	HSIL vs. Normal		LSIL vs. Normal	
	POR (95% CI)	p-value	POR (95% CI)	p-value
Yes	2.13 (0.78-5.82)	0.140	1.13 (0.42-3.01)	0.806

Table 4

Results of baseline category polytomous regression of histological grade, Houston, TX, USA and Vancouver, British Columbia, Canada, October, 2000 - July, 2003

Variable	HSIL vs. Normal		LSIL vs. Normal	
	POR*	95% CI	POR*	95% CI
ln Total RNA	1.65	1.30-2.09	1.27	1.08-1.51
DNA Index				
Diploid	1.00		1.00	
Aneuploid	8.41	2.85-24.87	1.38	0.60-3.18
iHAC Score				
<37	1.00		1.00	
=37	0.31	0.11-0.86	1.29	0.70-2.37
No. Pregnancies				
= 2	1.00		1.00	
3 +	0.25	0.09-0.67	0.59	0.32-1.07

* adjusted for age and study site

Table 5

Results of ordinal regression of histological grade, Houston, TX, USA and Vancouver, British Columbia, Canada, October, 2000 - July, 2003

Variable	Adjacent Category Model [†]		Continuous Ratio Model [‡]		Proportional Odds Model [§]	
	POR*	95% CI	POR*	95% CI	POR*	95% CI
In Total RNA	1.27	1.14-1.42	1.30	1.15-1.48	1.33	1.16-1.52
DNA Index						
Diploid	1.00		1.00		1.00	
Aneuploid	2.51	1.49-4.24	2.88	1.58-5.22	2.68	1.36-5.28
iHAC Score						
<37	1.00		1.00		1.00	
=37	0.75	0.49-1.16	0.73	0.45-1.19	0.77	0.45-1.32
No. Pregnancies						
= 2	1.00		1.00		1.00	
3 +	0.53	0.34-0.81	0.48	0.30-0.78	0.44	0.26-0.74

* adjusted for age and study site

[†] Adjacent category model compares HSIL to LSIL and LSIL to Normal.[‡] Continuous ratio model compares HSIL to LSIL/Normal and LSIL to Normal.[§] Proportional odds model compares HSIL to LSIL/Normal and HSIL/LSIL to Normal