Hierarchical Clustering of Human Papilloma Virus Genotype Patterns in the ASCUS-LSIL Triage Study

Nicolas Wentzensen1, Lauren E. Wilson1, Cosette M. Wheeler2, Joseph D. Carreon1, Patti E. Gravitt3, Mark Schiffman1, and Philip E. Castle1

Abstract
Anogenital cancers are associated with ∼13 carcinogenic human papilloma virus (HPV) types in a broader group that cause cervical intraepithelial neoplasia (CIN). Multiple concurrent cervical HPV infections are common, which complicates the attribution of HPV types to different grades of CIN. Here we report the analysis of HPV genotype patterns in the atypical squamous cells of undetermined significance–low-grade squamous intraepithelial lesion triage study with the use of unsupervised hierarchical clustering. Women who underwent colposcopy at baseline (n = 2,780) were grouped into 20 disease categories based on histology and cytology. Disease groups and HPV genotypes were clustered with the use of complete linkage. Risk of 2-year cumulative CIN3+, viral load, colposcopic impression, and age were compared between disease groups and major clusters. Hierarchical clustering yielded four major disease clusters: cluster 1 included all CIN3 histology with abnormal cytology; cluster 2 included CIN3 histology with normal cytology and combinations with either CIN2 or high-grade squamous intraepithelial lesion cytology; cluster 3 included older women with normal or low-grade histology/cytology and low viral load; and cluster 4 included younger women with low-grade histology/cytology, multiple infections, and the highest viral load. Three major groups of HPV genotypes were identified: group 1 included only HPV16; group 2 included nine carcinogenic types, plus noncarcinogenic HPV53 and HPV66; and group 3 included noncarcinogenic types, plus carcinogenic HPV33 and HPV45. Clustering results suggested that colposcopy missed a prevalent precancer in many women with no biopsy/normal histology and high-grade squamous intraepithelial lesion. This result was confirmed by an elevated 2-year risk of CIN3+ in these groups. Our novel approach to study multiple genotype infections in cervical disease with the use of unsupervised hierarchical clustering can address complex genotype distributions on a population level.

Cancer Res; 70(21); 8578–86. ©2010 AACR.

Introduction
More than 40 different types of human papilloma viruses (HPV) can infect the anogenital mucosa. Most of these types cause asymptomatic transient infections that may be associated with minor cytologic alterations, whereas approximately a dozen carcinogenic types can cause anogenital cancer (1). HPV16 is by far the most carcinogenic type. HPV18, 31, 33, and 45 follow, and together with HPV16 account for >90% of HPV-related cancers (2).

Traditionally, cervical cancer was thought to arise through increasingly severe grades of cervical intraepithelial neoplasia (CIN), defined by the extent and severity of cellular atypia. However, CIN1 is now known to represent acute HPV infection, whereas CIN3 is precancer (and includes carcinoma in situ). At the transition between acute infection and precancer is an equivocal and poorly reproducible diagnosis called CIN2 (3), which probably represents a mixture of precancer and HPV infection. The cytologic correlates used in screening are high-grade squamous intraepithelial lesion (HSIL) corresponding to CIN2/3, or low-grade squamous intraepithelial lesion (LSIL) corresponding to CIN1. The most common equivocal cytologic abnormalities are called atypical squamous cells of undetermined significance (ASCUS).

The multiple stages of cervical carcinogenesis are being redefined due to increased understanding of HPV natural history (1, 4). We distinguish acute HPV infection, a common and benign condition, from uncommon persistent infection that is the true risk factor for precancer and cancer. The prevalence of HPV types in the genital tract and the association of HPV types with different stages in CIN are related to multiple host and viral factors. HPV infection is easily transmitted by sexual contact, whereas poorly understood
immunologic factors are related to viral clearance/persistence. The risk of viral persistence and associated development of precancer vary by type. Some noncarcinogenic types from the $\alpha_3/\alpha_15$ species are preferentially detected in the vaginal epithelium, whereas $\alpha_7$ types, such as HPV18 and HPV45, are frequently found in endocervical lesions (5). There is no convincing evidence for interaction between multiple cervical HPV infections.

We wish to sort the different histologic and cytologic findings by their relationships with HPV natural history. However, HPV prevalence studies with the use of broad HPV genotyping have shown that multiple HPV infections are very common, especially in young women at the peak of their sexual activity (6, 7). Current genotyping assays allow detection of up to $\sim 40$ HPV genotypes from the same sample, generating complex HPV genotyping data that complicate the attribution of individual HPV genotypes to grades of CIN and corresponding cytology. Thus far, the complexity has been mainly addressed by restricting analyses to single genotype infections, by attributing genotypes to disease hierarchically based on the HPV prevalence in cervical cancers, or by combining genotypes within phylogenetic species. We previously showed the wide range of potential type attribution to all stages of cervical disease that can only be resolved by genotyping of individual lesions on the cervix (7).

The attribution of individual genotypes to cervical disease is further complicated by the imprecise ascertainment of cervical disease stages. Colposcopy and biopsy frequently miss prevalent precancer (8, 9), and even if the worst lesion is not missed, cytology and histology have both limited reproducibility, especially at the lower disease stages (10). For example, in a cross-sectional study, we showed that women with HSIL cytology and normal biopsy results had very similar HPV genotype patterns as women with biopsy-confirmed high-grade disease, suggesting that the worst lesion was frequently missed in colposcopy (11).

To further our understanding of the spectrum of histologic and cytologic abnormalities in relationship to HPV types, we analyzed HPV genotype distributions in 20 disease categories based on cytology and histology, and show their relation to subsequent 2-year risk of CIN3 in a large clinical trial called the ASCUS-LSIL triage study.

**Materials and Methods**

**Study population**

The ASCUS-LSIL triage study was a multicenter, randomized clinical trial conducted by the National Cancer Institute to compare three clinical management strategies for women referred with a community cytologic interpretation of ASCUS ($n = 3,488$) or LSIL ($n = 1,572$) cytology (12). At enrollment, cytology was repeated and HPV testing was done with the use of Hybrid Capture 2 (Digene Corporation, now Qiagen). Women in the immediate colposcopy arm were referred to colposcopy regardless of test results. In the HPV triage arm, women were referred to colposcopy if they were HPV-positive or received a cytology diagnosis of HSIL at their enrollment visit, and women in the conservative management arm underwent colposcopy only if they received an HSIL cytology diagnosis at enrollment. Our final analysis included 2,780 women.

---

1 ASCUS under the 1991 Bethesda system was slightly more inclusive, particularly of probable reactive changes and atypical squamous cells with possible HSIL, than the ASCUS category of the 2001 Bethesda system.
an HSIL cytology result at enrollment. Women were followed for 2 years, with cytology follow-ups every 6 months. Women with a HSIL cytology result at any of these follow-up visits were referred to colposcopy. Our analysis included all women referred for either ASCUS$^4$ or LSIL who underwent colposcopy at enrollment and who had a cytology diagnosis in the enrollment period ($n = 2,780$; Fig. 1).

Colposcopy was done by nurse practitioner colposcopists, general gynecologists, gynecology oncology fellows, or gynecologic oncologists. The type of medical training did not influence the sensitivity of colposcopy to detect CIN3+ in 2 years of follow-up (8). Before colposcopy, high-resolution photography of the cervix was done to evaluate visual screening and as additional colposcopy quality control.

Figure 2. Hierarchical clustering of disease groups by genotype prevalence. Columns, disease combinations based on histology (no biopsy, normal, CIN1, CIN2, CIN3) and cytology (normal, ASCUS, LSIL, HSIL); N, number of women in each disease combination; rows; HPV genotypes. The heatmap shows type-specific prevalence determined by line blot assay in gray scale for each type-disease stage combination. Darker gray, higher prevalence. Dendrograms are shown for both disease groups and HPV genotypes. 1 to 4, four disease clusters; 1 to 3, three genotype clusters; 2a to c, and 3a and b, subgroups.
At each study visit, a pelvic exam was done, and two cervical specimens were collected. One specimen was preserved in PreservCyt (Cytyc, now Hologic) for cytology and Hybrid Capture 2 testing, and the second was preserved in specimen transport medium (Qiagen). The National Cancer Institute and local Institutional Review Boards approved the study, and all participants provided written informed consent.

**HPV genotyping**

Line blot assay was done on enrollment specimen transport medium specimens as previously described for the detection of 27 individual (HPV6, 11, 16, 18, 26, 31, 33, 35, 39, 40, 42, 45, 51-59, 66, 68, 73, and 82-84) HPV types (13). A subset of specimens was retested by linear array, a commercialized version of line blot assay that tests for 37 HPV genotypes, including 26 detected by line blot assay as previously described (14).

Residual PreservCyt (Hologic) specimens, after being used for liquid-based cytology, were tested by Hybrid Capture 2 (Qiagen), a pooled-probe, signal amplification DNA test that targets a group of 13 carcinogenic HPV types.

**Pathology and treatment**

Treatment was based on cytologic and histologic diagnoses made by the clinical center pathologists as described previously. For quality control purposes, the Pathology Quality Control Group at Johns Hopkins Hospital reviewed referral smears, ThinPreps, and histology slides, and provided secondary diagnoses. Excisional treatment by the loop electrosurgical excision procedure was offered to any women receiving a clinical center histology diagnosis of CIN2 or worse, or a quality control diagnosis of CIN3 or worse. At the time of study exit, all women with persistent mild cervical abnormalities were offered treatment by loop electrosurgical excision procedure.

**Statistical methods**

Women were separated into 20 categories of cervical disease, formed by crossing the enrollment histology diagnosis with the cytology result. Disease combinations were consistently labeled with the use of the following format: ‘histology result/cytology result.’ Histology diagnoses were drawn from the quality control histology diagnosis when available. If a subject did not have a quality control histology diagnosis for the enrollment period, which was rare, her clinical center enrollment histology diagnosis was used. Histology diagnoses were classified as Normal, CIN1, CIN2, CIN3, or ‘No Biopsy’ (because no biopsy was taken at colposcopy, indicating a negative colposcopic impression). Enrollment cytology interpretation categories were Normal, ASCUS (including atypical squamous cells with possible HSIL), LSIL, or HSIL. With the use of the HPV infection status of each individual from the line blot assay results, we calculated type-specific HPV infection frequencies at enrollment for each diagnosis category. Also for each category, we calculated 2-year risk of CIN3+ diagnosis and described some other clinical and demographic data.

To examine patterns of HPV infection in these 20 diagnosis categories, we used hierarchical clustering to compare HPV genotype patterns in each disease group. We used complete linkage and a Euclidean distance metric. We simultaneously clustered both disease combinations and HPV genotypes, and created dendrograms to visualize the clustering with the use of the TreeView software. For sensitivity analyses, we also did the same hierarchical clustering: (a) with the use of the linear array results instead of the line blot assay results (linear array results were only available in women referred into the ASCUS-LSIL triage study with an ASCUS Pap), and (b) restricted to single-type HPV infections. We also recategorized the women based on worst 2-year histology and enrollment cytology, and did the same analyses to examine the effect of misclassification of disease at the time of enrollment on the HPV frequency patterns. We confirmed the clustering results by doing a k-means cluster analysis specifying three and four clusters.

Data analysis was done with the use of SAS version 9.1. Cluster 3.0 was used for clustering of HPV frequency arrays, and Java TreeView was used to construct frequency maps and dendrograms (15, 16).

**Results**

**HPV genotypes in disease groups**

Supplementary Table S1 displays the HPV genotype prevalence in all histology/cytology combinations. In 18 of the 20 groups, the most frequent genotype was HPV16, the exceptions being the groups normal/LSIL and CIN1/normal. Within each histology category, the women with HSIL cytology had the highest frequency of HPV16 infection; HPV16 frequency also increased with increasing histologic severity. Of 165 women with CIN3/HSIL, 117 (71%) were infected with HPV16. Multiple infections were very common in this population but varied significantly between disease groups from an average number of 0.74 types in women with normal/normal to an average number of 2.38 types in women with CIN2/LSIL. Across all histologic categories, most infections were found in women with LSIL cytology (Supplementary Table S1).

**Clustering of disease groups by genotype frequencies**

Unsupervised hierarchical clustering of disease groups by genotype frequencies yielded a tree with four major histology/cytology clusters (Fig. 2; Table 1). Cluster 1 included CIN3/ASCUS, CIN3/LSIL, and CIN3/HSIL. Cluster 2 included CIN3/normal, all less-severe histology with HSIL except for CIN1/HSIL, and CIN2/normal. Cluster 3 included normal histology or no biopsy with normal, ASCUS, or LSIL cytology. Cluster 4 included CIN2/ASCUS, CIN2/LSIL, normal/LSIL, and all CIN1 except for CIN1/normal.

We observed very similar clustering of disease combinations when restricting the analysis to women referred for ASCUS cytology only with the use of either line blot assay or linear array HPV genotyping data, although some histology/cytology combinations had <20 cases and were excluded. When restricting to cases with single-type HPV infections only, many combinations had only very small numbers and generated unstable clusters. K-means clustering specifying four disease clusters reproduced the same grouping.
pattern that was observed with unsupervised hierarchical clustering.

**Clustering of genotypes by genotype frequencies in disease groups**

HPV genotypes were clustered in three major groups (Fig. 2; Table 2): HPV16 clustered separately from all other HPV genotypes, driven by its high frequency across all disease stages (HPV cluster 1). Nine carcinogenic types, plus HPV66 and HPV53, were included in HPV cluster 2; the first subgroup included only α9 types, the second had mainly α7 types, and the third was dominated by α6 and α9 types. All remaining noncarcinogenic types, as well as HPV33 and HPV45, were included in HPV cluster 3; there was no specific distribution of HPV clades in the three subgroups of cluster 3. HPV genotypes in the first and second HPV clusters showed a differential distribution across disease categories, with their lowest prevalence in the disease cluster 3 and the highest prevalence in the disease cluster 4. Within the third HPV group, we observed 2 subgroups: (a) a subset of HPV genotypes with very low prevalence and no disease-specific distribution (HPV11, HPV26, HPV40, and HPV57 in HPV cluster 3a), and (b) 11 types with higher prevalence evenly distributed across all disease clusters (HPV clusters 3b+c). Exclusion of noncarcinogenic types had only a minor effect on disease clustering, whereas exclusion of carcinogenic types produced a completely different clustering of disease combinations, suggesting that the clustering was driven by carcinogenic types. Overall, the HPV clusters were less distinct than the disease clusters, as indicated by the distance metrics and visualized by the flat branching in the HPV genotype dendrogram.

**Clinical characteristics and risk of subsequent CIN3 in disease clusters**

We studied 2-year risk of CIN3 within the individual disease groups and within disease clusters (Fig. 3; Table 1). The 2-year risk of CIN3 differed significantly between the clusters ($P < 0.001$). Cluster 1 included only women with CIN3. Women in cluster 2 (including some women with CIN3 histology) had a 20% risk of CIN3 on average. The risk was much lower for women in cluster 4 (6.5%) and lowest for women in cluster 3 (3.5%). The clustering of women with normal or no histology and HSIL cytology in clusters with high-grade histology and cytology is also reflected by their higher risk of developing CIN3: among women with normal histology at enrollment, 16.5% of those with HSIL enrollment cytology (cluster 2) had the worst 2-year histology result of CIN3+ compared with 3.3% of women who had normal enrollment cytology (cluster 3; $P < 0.001$).

We also studied the clinical and demographic characteristics across the four disease clusters (Fig. 3). Women in cluster 3 were oldest with a median age of 27 years, whereas women in cluster 4 were youngest with a median age of 23 years, similar to clusters 1 and 2 (median age of

**Table 1. Two-year risk of high-grade disease and clinical characteristics of disease groups**

<table>
<thead>
<tr>
<th>Histology_cytology group</th>
<th>N</th>
<th>Dendrogram cluster</th>
<th>2-y risk CIN3+ (%)</th>
<th>2-y risk CIN2+ (%)</th>
<th>2-y risk confirmed CIN3 (%)</th>
<th>&gt;25 y at enrollment (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal_LSIL</td>
<td>352</td>
<td>4</td>
<td>6</td>
<td>10.2</td>
<td>3.7</td>
<td>33</td>
</tr>
<tr>
<td>CIN1_LSIL</td>
<td>280</td>
<td>4</td>
<td>5.7</td>
<td>12.1</td>
<td>3.2</td>
<td>31.1</td>
</tr>
<tr>
<td>CIN1_ASCUS</td>
<td>95</td>
<td>4</td>
<td>6.3</td>
<td>8.4</td>
<td>3.2</td>
<td>41.1</td>
</tr>
<tr>
<td>CIN1_HSIL</td>
<td>62</td>
<td>4</td>
<td>9.7</td>
<td>22.6</td>
<td>6.5</td>
<td>21</td>
</tr>
<tr>
<td>CIN2_LSIL</td>
<td>87</td>
<td>4</td>
<td>9.2</td>
<td>100</td>
<td>5.6</td>
<td>28.7</td>
</tr>
<tr>
<td>CIN2_ASCUS</td>
<td>52</td>
<td>4</td>
<td>1.9</td>
<td>100</td>
<td>0</td>
<td>21.2</td>
</tr>
<tr>
<td>Normal_normal</td>
<td>670</td>
<td>3</td>
<td>2.8</td>
<td>4.7</td>
<td>1.3</td>
<td>53.5</td>
</tr>
<tr>
<td>Normal_ASCUS</td>
<td>376</td>
<td>3</td>
<td>2.9</td>
<td>6.4</td>
<td>1.1</td>
<td>45.5</td>
</tr>
<tr>
<td>No biopsy_normal</td>
<td>34</td>
<td>3</td>
<td>8.8</td>
<td>14.7</td>
<td>2.9</td>
<td>29.4</td>
</tr>
<tr>
<td>CIN1_normal</td>
<td>73</td>
<td>3</td>
<td>1.4</td>
<td>4.1</td>
<td>0</td>
<td>49.3</td>
</tr>
<tr>
<td>No biopsy_LSIL</td>
<td>63</td>
<td>3</td>
<td>12.7</td>
<td>15.9</td>
<td>6.4</td>
<td>44.4</td>
</tr>
<tr>
<td>No biopsy_ASCUS</td>
<td>31</td>
<td>3</td>
<td>3.2</td>
<td>9.7</td>
<td>3.2</td>
<td>61.3</td>
</tr>
<tr>
<td>CIN2_normal</td>
<td>20</td>
<td>2</td>
<td>10</td>
<td>100</td>
<td>5</td>
<td>35</td>
</tr>
<tr>
<td>Normal_HSIL</td>
<td>109</td>
<td>2</td>
<td>11.9</td>
<td>19.2</td>
<td>8.3</td>
<td>43.1</td>
</tr>
<tr>
<td>CIN2_HSIL</td>
<td>81</td>
<td>2</td>
<td>9.9</td>
<td>100</td>
<td>8.6</td>
<td>33.3</td>
</tr>
<tr>
<td>No biopsy_HSIL</td>
<td>20</td>
<td>2</td>
<td>10</td>
<td>15</td>
<td>10</td>
<td>35</td>
</tr>
<tr>
<td>CIN3_normal</td>
<td>25</td>
<td>2</td>
<td>100</td>
<td>100</td>
<td>56</td>
<td>52</td>
</tr>
<tr>
<td>CIN3_ASCUS</td>
<td>29</td>
<td>1</td>
<td>100</td>
<td>100</td>
<td>44.8</td>
<td>20.7</td>
</tr>
<tr>
<td>CIN3_LSIL</td>
<td>53</td>
<td>1</td>
<td>100</td>
<td>100</td>
<td>45.3</td>
<td>35.9</td>
</tr>
<tr>
<td>CIN3_HSIL</td>
<td>165</td>
<td>1</td>
<td>100</td>
<td>100</td>
<td>78.2</td>
<td>37</td>
</tr>
</tbody>
</table>

(Continued on the following page)
24 y; \( P < 0.001 \). Semiquantitative viral load determined by Hybrid Capture 2 signal strength was highest in cluster 4 and lowest in cluster 3 (\( P < 0.001 \)). The frequency of high-grade colposcopy was highest in cluster 4, followed by clusters 2 and 1, a pattern confirmed for the review of cervigram images (\( P < 0.001 \)). The frequency of current smoking increased with increasing severity of disease in a cluster (1 > 2 > 4 > 3), further suggesting that smoking is indeed a cofactor for the development of precancerous lesions.

**Discussion**

In this analysis, we used a novel approach to study complex HPV genotype patterns in cervical disease and to address the misclassification of cervical disease stages defined by histology and cytology. Both histology and cytology are subjective methods with limited reproducibility. Colposcopic biopsy frequently misses the worst lesion on the cervix and augments the problem of disease misclassification (8, 9). Here, we did unsupervised hierarchical clustering of HPV genotyping data to agnostically define disease groups with similar HPV genotype patterns. We identified four disease clusters based on their unique HPV genotype distributions. The first cluster included women with CIN3 and HSIL, LSIL, or ASCUS cytology. The second cluster included mainly women with HSIL and/or CIN2. The third cluster included only women without high-grade histology or cytology and with low levels of viral infection, whereas the fourth cluster included women with mild-to-moderate dysplasia and cytologic signs of active viral infections (ASCUS and LSIL). We confirmed a previous finding that women with HSIL cytology but without histologically confirmed high-grade biopsy results clustered with those with histologically confirmed high-grade disease (11). We now show that these women also have a high risk of being detected with CIN3+ in the subsequent 2 years, corroborating that colposcopy-biopsy missed some prevalent precancer.

Women grouped in the four clusters had distinct clinical characteristics: women in the first cluster all had prevalent CIN3, and accordingly, they were the most likely to have visual and microscopic evidence of abnormalities and higher viral load. Women in the second cluster had the second highest risk of CIN3, had fewer abnormal cervical impressions.
younger women are likely to have multiple infections that reflect different stages of the natural history of HPV-related disease: younger women are likely to have multiple infections that are mainly productive rather than transforming, and associated with high viral load (cluster 4). Older women have fewer infections and lower viral load if there is no prevalent disease (cluster 3). Women with prevalent precancer have fewer infections with many carcinogenic types (most importantly HPV16) and may have high viral load (clusters 1 and 2). Admittedly, the age range in the ASCUS-LSIL triage study population is limited, and the difference in median age between clusters 3 and 4 is very small.

It is important to note that the risk prediction for women with CIN2 detected at baseline is limited by censoring because most women with CIN2 had a loop electrosurgical excision procedure, interrupting the natural history. The counterintuitive grouping of CIN2-ASCUS and CIN2-LSIL in disease cluster 4, whereas CIN2-normal is grouped in disease cluster 2, is driven by higher prevalence of genotypes from HPV cluster 3, which mainly includes noncarcinogenic types. Restricting the clustering to carcinogenic types led to a closer grouping of CIN2 disease groups.

The second dimension of cluster analysis identified three major clusters of HPV genotypes: HPV16 clustered separately from all other HPV genotypes and showed the closest association to risk of CIN3 among all types. The second HPV cluster included nine high-risk types, plus HPV53 and HPV66, and was found at higher frequencies in disease clusters 1, 2, and 4, whereas the third HPV cluster included mainly noncarcinogenic types, plus HPV33 and HPV45, and showed frequencies distributed more evenly across the disease clusters. Although HPV33 and HPV45 had the highest prevalence in CIN3/HSIL in cluster 3, their more uniform distribution in comparison with other carcinogenic types most likely caused their grouping with noncarcinogenic types. The carcinogenicity of HPV53, HPV66, and HPV68 has been widely debated. In the most recent IARC classification, HPV53 and HPV66 were considered possibly carcinogenic (cluster 2B), whereas HPV68 was classified as probably carcinogenic (cluster 2A) because of experimental and phylogenetic evidence but without strong supporting epidemiologic data (17, 18). The clustering of HPV53 and HPV66 with carcinogenic types in our analysis reflects their ability to cause a spectrum of mild and more severe precursor lesions, including CIN3, but little or no chance of invasion. The correlation of genotype clustering with the phylogenetic clades (19) and the WHO carcinogen classification (17) was quite good (Table 2), and we think that the remaining discrepancies are mainly related to the lack of cancers in the disease spectrum we analyzed. Still, whereas our approach can indicate which genotypes are important in the progression to precancer, it cannot identify the causal type in a multiple infection. Ultimately, lesion-specific genotyping is required to precisely attribute HPV genotypes to cervical precancer.

With our analytic approach, we were able to show the complex relation of HPV genotypes and cervical disease in 2,780 women from the ASCUS-LSIL triage study in a single figure. As exemplified in our study, it is possible to use this technique to address disease misclassification. For example, the same approach can be used to address the heterogeneity of CIN2, a very diverse category that, depending on the local histologic interpretation, may include a lot of low-grade diseases (more likely to be in cluster 4) or that is more similar to...
CIN3 (more likely to be in cluster 2). Our clustering approach allows studying type attribution to disease in different regions of the world by comparing type allocation with genotype clusters. Similarly, it can be used to analyze shifts in type attribution to disease in vaccinated populations. Due to the frequent misclassification of cervical disease, the disease groups used in the analysis are a combination of true results, cases in which the worst lesion was missed in colposcopy, and cases in which histology and/or cytology was undercalled or overcalled. Despite the common notion that cervical histology is the gold standard of disease ascertained, in our analysis, cytology was an important indicator of subsequent risk of CIN3+ within the group of normal/low-grade histology, reflected by similar HPV genotype patterns as found in histology-confirmed high-grade disease.

In summary, we present a novel solution to the handling of complex HPV genotype data in cervical disease and applied it in the prospective ASCUS-LSIL triage study. We show that HPV genotype patterns at various cervical disease stages are complex but distinctive when analyzed in aggregate. Our approach allows easily displaying HPV prevalence in disease groups and may be used to compare the distribution of HPV genotypes within diagnostic categories between different populations.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We thank Digene, Cytyc, National Testing Laboratories, DenVu, TriPath Imaging, and Roche Molecular Systems for donating or providing at reduced cost some of the equipment and supplies used in the ASCUS-LSIL triage study trial.

Grant Support

Intramural Research Program of the NIH and the National Cancer Institute. Roche Molecular Systems provided reagents and research support to the laboratory of P.E. Gravitt. C.M. Wheeler received support through her institution from Roche Molecular Systems for HPV genotyping studies.

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

Received 04/09/2010; revised 07/17/2010; accepted 08/06/2010; published OnlineFirst 10/19/2010.

References

papillomavirus genotype infections in cervical cancer progression in
the study to understand cervical cancer early endpoints and deter-
8. Gage JC, Hanson VW, Abbey K, et al. Number of cervical biopsies
biopsy, random cervical biopsy, and endocervical curettage in
the diagnosis of cervical intraepithelial neoplasia II or worse. Am J
10. Stoler MH, Schiffman M. Interobserver reproducibility of cervical
cytologic and histologic interpretations: realistic estimates from the
11. Wentzensen N, Schiffman M, Dunn ST, et al. Grading the severity of
cervical neoplasia based on combined histopathology, cytopathol-
yogy, and HPV genotype distribution among 1,700 women referred
12. Schiffman M, Adarianza ME, ASCUS-LSIL Triage Study. Design,
methods and characteristics of trial participants. Acta Cytol 2000;
44:726–42.
human papillomavirus types by using L1 consensus PCR products
by a single-hybridization, reverse line blot detection method. J Clin
14. Castle PE, Gravitt PE, Solomon D, Wheeler CM, Schiffman M.
Comparison of linear array and line blot assay for detection of human
papillomavirus and diagnosis of cervical precancer and cancer in the
atypical squamous cell of undetermined significance and low-grade
15. de Hoon MJ, Imoto S, Nolan J, Miyano S. Open source clustering
16. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and
display of genome-wide expression patterns. Proc Natl Acad Sci
18. Schiffman M, Clifford G, Buonaguro FM. Classification of
weakly carcinogenic human papillomavirus types: addressing the lim-
human papillomavirus types reflects viral evolution. Virology 2005;
337:76–84.
Hierarchical Clustering of Human Papilloma Virus Genotype Patterns in the ASCUS-LSIL Triage Study

Nicolas Wentzensen, Lauren E. Wilson, Cosette M. Wheeler, et al.

Cancer Res 2010;70:8578-8586. Published OnlineFirst October 19, 2010.

Updated version
Access the most recent version of this article at:
doi:10.1158/0008-5472.CAN-10-1188

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2010/10/08/0008-5472.CAN-10-1188.DC1

Cited articles
This article cites 19 articles, 5 of which you can access for free at:
http://cancerres.aacrjournals.org/content/70/21/8578.full.html#ref-list-1

Citing articles
This article has been cited by 6 HighWire-hosted articles. Access the articles at:
/content/70/21/8578.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.