p16 Immunohistochemistry Interpretation by Nonpathologists as an Accurate Method for Diagnosing Cervical Precancer and Cancer

Guang-Dong Liao, MD[1,2], **Le-Ni Kang, BS**[2], **Wen Chen, PhD**[2], **Xun Zhang, MD**[3], **Xiao-Yang Liu, BS**[3], **Fang-Hui Zhao, MD, PhD**[2], **Mark H. Stoler, MD**[4], **Anne Mills, MD**[4], **Ming-Rong Xi, MD**[1], **You-Lin Qiao, MD, PhD**[2], and **Philip E. Castle, MPH, PhD**[5,6]

[1]Department of Gynecology and Obstetrics, the West China Second University Hospital, Sichuan University, Key Laboratory of Birth Defects and Related Diseases of Women and Children (Sichuan University), Ministry of Education, Chengdu, China

[2]Department of Epidemiology, Cancer Institute and Hospital, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing, China

[3]Department of Pathology, Cancer Institute and Hospital, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing, China

[4]Department of Pathology, University of Virginia, Charlottesville, VA

[5]Global Cancer Initiative, Chestertown, MD

[6]Global Coalition Against Cervical Cancer, Arlington, VA

Abstract

Objective—We conducted a pilot study of whether nonpathologists could accurately diagnose cervical precancer in biopsies using only a basic light microscope, evaluating $p16^{INK4a}$ immunohistochemistry (p16 IHC) of biopsies, and video-based training for both.

Materials and Methods—Using biopsies collected as part of a screening study conducted in rural China, we randomly selected 50 biopsies with a precancerous diagnosis of cervical intraepithelial neoplasia grade 2 (CIN2) or more severe (CIN2+) and 50 biopsies with diagnosis of CIN less severe than CIN2, and stained them for p16 using a commercial IHC kit. Twelve nonpathologists of varying educational backgrounds living in Beijing, China received video training and were assigned one of 4 sets of 25 CIN2+ and 25 CIN less severe than CIN2 for evaluation. A pathologist reviewed all 100 cases.

Results—The mean sensitivity and specificity of the p16 IHC staining scored by the nonpathologists were 91.7% and 94.1%, respectively, compared to scoring by the pathologist. The readers and the pathologist agreed on p16 IHC scoring for 42 (84%) of the 50 slides of CIN less severe than CIN2 and 37 (74%) of the 50 CIN2+ slides. The mean sensitivity and specificity for

consensus CIN2+ of p16 IHC as scored by the readers were 88% and 87%, respectively, versus an overall sensitivity and specificity by the pathologist of 96% and 92%, respectively.

Conclusions—We demonstrated that nonpathologists can accurately diagnose CIN2+ using p16 IHC alone.

Whereas the recent World Health Organization recommendations[1] to use alternative screening (high-risk HPV testing or VIA vs Papanicolaou testing) and management strategies (e.g., screen and treat vs diagnostic verification) will hopefully expand screening to the women who need them the most, many countries may not adopt screen-and-treat strategies because of concerns of overtreatment, especially in the context of multiple screens over a lifetime. Thus, there is a need to expand diagnostic services to manage the women who screen positive and identify those with cervical precancer, cervical intraepithelial neoplasia grade 2 (CIN2), grade 3 (CIN3), or adenocarcinoma in situ (AIS), and early-stage invasive cervical cancer (CIN2+) for treatment.

The capacity gap for pathology in low- and middle-income countries remains a big challenge, especially in Sub-Saharan Africa, where there is one tenth the number of pathologists per person to provide diagnostic services.[2] Strategies to address this gap must be developed to prevent cervical pathology from being a bottleneck to the expansion of secondary cervical cancer prevention programs.

We hypothesized that p16[INK4a] immunohistochemistry (p16 IHC) is a sufficiently robust test for CIN2+ that nonpathologists with basic, video-based training on the use of a light microscope and reading of p16 IHC could accurately diagnose CIN2+. p16 IHC is currently recommended as an adjunctive molecular stain to clarify the diagnosis of CIN2, the threshold for treatment, i.e., to differentiate between low-risk (p16 IHC–negative) and high-risk (p16 IHC–positive) CIN2.[3] To test our hypothesis, we evaluated the diagnostic performance of educated nonpathologists living in Beijing, China and attending Peking University reading p16 IHC staining of biopsies collected from women living in rural China.

MATERIALS AND METHODS

This study was nested within a Screening Technologies to Advance Rapid Testing—Utility and Program (START-UP) project.[4,5] In 2010–2011, 7,541 women aged 25 to 65 years living in Yangcheng, Xinmi, and Tonggu counties were enrolled. The institutional review boards of Cancer Institute/Hospital, Chinese Academy of Medical Sciences in China and the Program for Appropriate Technology in Health of the United States approved the START-UP project. The written consent for START-UP project included the use of the specimens for additional testing. Details of inclusion/exclusion criteria, screening, management, and diagnosis of participants are described in detail elsewhere.[4,5]

Evaluation of p16 IHC

We enrolled nonpathologists attending Peking Union Medical College for secondary education. Eligibility criteria included people who were willing and able to provide informed consent. Participants were asked to provide some basic information on age, education, and training. The p16 IHC substudy underwent review by the institutional review board of Cancer Institute/Hospital, Chinese Academy of Medical Sciences. All subjects signed a written informed consent before entry into this study.

A random sample of 50 biopsies diagnosed as CIN2+ and 50 biopsies diagnosed as CIN less severe than CIN2 (negative or CIN1) from the START-UP project were selected for evaluation. Of the 100 diagnoses, 43 diagnoses (37 of which were CIN2+) were based on a joint review/adjudication between the 2 study pathologists (X.Z. and M.S.). The remaining 57 diagnoses (13 of which were CIN2+) were separate reviews by the 2 study pathologists, who agreed on all but 2 diagnoses (both were CIN2 by M. S. and CIN3 by X.Z.) that were conservatively assigned the less severe diagnosis. Four-micron-thick cuts of the formalin-fixed paraffin-embedded (FFPE) biopsies were made, placed on slides, and labeled with a new identifier. Slides were stained for p16^{INK4a} using CINtec (Ventana, Tucson, AZ) according to the manufacturer's instructions.

All readers (n = 12) were provided a short video on how to set up and use a light microscope (http://www.screencast.com/t/BIuLDwHwb) (n.b., Olympus BH2 was used in this study) and how to evaluate p16 IHC (http://www.screencast.com/users/PathologyAM/folders/p16%20Training%20/media/9f60765b-201c-4792-91ae-23423a9dbb33), the latter of which instruct readers to consider only diffuse p16 IHC staining as positive and focal p16 IHC staining as negative (as defined in detail in Darragh et al.[3]). Videos were generated using PowerPoint with voice-over recorded on Camtasia Recording Software (TechSmith, Okemos, MI).

Each reader was randomly assigned to one of 4 sets of 25 p16 IHC-stained CIN2+ and 25 p16 IHC-stained slides of CIN less severe than CIN2 so that all slides were reviewed by 6 readers. Each reader was instructed to score the slide as diffuse staining (positive), focal staining (negative), or no staining (negative). All slides were also scored by the study pathologist.

Analysis

For each reader and the study pathologist for this analysis (X. Z.) rating the same set of 50 slides, sensitivity and specificity for CIN2+ were calculated. A mean sensitivity and specificity for p16 IHC staining and CIN2+ diagnoses for all scores by all readers on the 100 slides was calculated. An overall sensitivity and specificity for CIN2+ for the study pathologist (X.Z.) rating the 100 slides was calculated. A mean total agreement and Kappa value between all scores of the review by the participants and the score by the pathologist was calculated for all slides.

RESULTS

Table 1 shows the characteristic of the readers. Nine women and 3 men participated. The mean, median, and range of ages were 25, 25, and 22 to 29 years, respectively. Three readers are undergraduate students, 6 readers are master's degree candidates, and 3 readers are doctorate degree candidates. Only one reader had professional laboratory training, none had pathology training, and only one self-reported being skilled in using a microscope. The median, mean, and range of times to evaluate the 50 slides were 60.5, 59.7, and 41 to 75 minutes, respectively.

The percent p16 IHC positive ratings as scored by the study pathologist (X.Z.) by grade of histologic diagnosis were 2.3% for negative (n = 44), 50.0% for CIN1 (n = 6), 95.8% for CIN2 (n = 24), 95.7% for CIN3 (n = 23), and 100% for cervical cancer (n = 3). Using the rating of the p16 IHC by the pathologist as an analytical endpoint, the mean sensitivity and specificity of the readers' rating of the p16 IHC were 91.7% and 94.1%, respectively. There was complete agreement by all readers and the pathologist on the p16 IHC results for 42 (84%) of the 50 slides of CIN less severe than CIN2 and 37 (74%) of the 50 CIN2+ slides. The mean total agreement between the participants and the study pathologist in reviewing the p16 IHC was 92.8% and kappa was 0.857.

Table 2 shows the sensitivity and specificity of both the participants and the study pathologist reviewing the p16 IHC for clinical endpoint of histologically confirmed CIN2+. For the individual sets of 50 reviews, the sensitivity for CIN2+ ranged from 72% to 96% for the readers scoring the p16 IHC and 96% for all sets for the pathologist scoring the p16 IHC. The specificity for CIN2+ ranged from 76% to 96% for the readers scoring the p16 IHC and 88% to 96% for the pathologist scoring the p16 IHC. The mean sensitivity and specificity for CIN2+ of the readers were 88% and 87%, respectively. Using the Lower Anogenital Squamous Terminology (LAST) definition of high-grade squamous intraepithelial lesion (HSIL), CIN3+ and p16 IHC-positive CIN2 (as read by X.Z.), and the mean sensitivity and specificity for HSIL were 90% and 87%, respectively. The mean sensitivity and specificity for CIN2+ of the study pathologist (X.Z.) scoring the p16 IHC were 96% and 92%, respectively. Thus, the Youden Index and the positive likelihood ratio, as metrics for overall accuracy, for CIN2+ were higher for the study pathologist (88% and 12, respectively) than the readers (75% and 6.8, respectively) scoring the p16 IHC.

Figure 1 shows examples of the p16 IHC staining: a consensus p16 IHC-positive (true positive) CIN3 (Figures 1A, E), a consensus p16-IHC-negative CIN2 (Figures 1B, F), a consensus p16 IHC-positive CIN1 (false positive) (Figures 1C, G), and a consensus p16 IHC-negative (true negative) CIN1 (Figures 1D, H).

DISCUSSION

In a proof-of-principle study, we demonstrated that with simple didactic training via videos, we can train educated nonpathologists to evaluate p16 IHC with excellent agreement with a senior pathologist and achieve very good clinical performance versus the criterion standard, a morphologic assessment of hematoxylin-and-eosin–stained tissue by pathologists. Once

the microscope was set up for reading p16 IHC, the readers reviewed the slides at a rate of approximately 1 to 1.5 minutes per slide. We did not provide any ongoing feedback and training to the readers, which we expect would have improved the overall performance further still.

This approach focuses on the need for low- and middle-income countries, like Sub-Saharan Africa,[2] where there may be insufficient numbers of pathologists, which could limit expansion of screening programs where diagnostic verification for treatment decisions is required. In higher-resource settings, there are sufficient pathologists to rely on for morphologic assessment, which is the criterion standard, and there is no need to use this approach in these settings. However, it is worth noting that recent US pathology (LAST) guidelines for diagnosis of cervical neoplasia recommend the use of p16 IHC to clarify equivocal diagnoses, specifically CIN2. [2] Since virtually all CIN3 biopsies would be expected to test p16 positive, [2] accurate review of p16 IHC would achieve similar sensitivity for CIN2+ as the LAST guidelines.

An alternative strategy to basing treatment solely on the interpretation of p16 by a nonpathologist, as considered here, would be to incorporate laboratory personnel, such as histotechnologists, who would be already processing the biopsies to create FFPE tissue and create slides, to do the p16 IHC staining, and screen the slides to rule out CIN2+ then send only the subset of slides of interest to the local pathologist to "rule in" CIN2+ by reviewing the p16 IHC slide, hematoxylin and eosin slide, or both. That is, have laboratory personnel read more sensitively but less specifically to reduce the number of slides that need to be expertly reviewed by a pathologist, of which low- and middle-income countries are in short supply. This is akin to how cytologic examination is done in many countries: cytotechnologists screen the cytology slide and primarily send only those suspicious of abnormalities to the pathologist for review.

To simulate this alternate strategy, we redefined p16 IHC positivity to any p16 staining, either diffuse or focal staining, as scored by the readers, which would have achieved a 94% sensitivity and 50% specificity for CIN2+. Then, restricting the pathologist rating only to those positive for any p16 as judged by the readers, the overall sensitivity and specificity both would have been 92%, which would have been a little more sensitive and specific than basing management on the scoring by the nonpathologist alone. For a typical CIN2+ prevalence of 20% in a colposcopy referral population, such a strategy would have reduced the number of slides to be reviewed by the pathologist by approximately 40%, with only a small decrement in the overall sensitivity for CIN2+ based on morphologic assessment. However, such a strategy was not evaluated in this study, and its performance would need to be confirmed.

We noted that several CIN1s were consensus p16 IHC positive. As previously reported,[6–10] a significant fraction (20%–40% or even higher) CIN1 tissues will have diffuse p16 staining. Although these cases did not develop CIN2+ after a 1-year follow-up and rescreening, 2 of the 3 cases had evidence of persistent high-risk HPV infection, which is an important risk factor for the development of cervical precancer and cancer.[11–13] Furthermore, women with p16 IHC-positive CIN1 are at a significantly higher subsequent risk of CIN2+ than women

with p16 IHC-negative CIN1.[9] Thus, in some instances, these cases are not false positives but will develop into a clinically important disease and if possible should be monitored closely.

The concern that treating all p16-positive cases might result in significant overtreatment also needs to be weighed against the incredible high frequency of overcalling of CIN1 and CIN2 in many routine practices, many of which would then test p16 IHC negative. An objective biomarker like p16 helps address such potential for overcall that will likely lead to unnecessary treatment in both low- and high-resource settings.[2] In low- and middle-income countries, where the prevalence of cervical precancer and cancer is high because of a lack of screening, we believe that the benefits of the proposed method of diagnosis far outweigh the risks of overtreating a few p16-positive CIN1s. It offers an important alternative for those countries that will not accept screen-and-treat strategies (that would lead to much greater overtreatment) now endorsed by the WHO[1] but do not have sufficient numbers of pathologists to provide diagnostic services for scaled-up screening programs.

We conclude that p16 IHC has tremendous potential to address the gap in pathologic services that may inhibit the expansion of cervical cancer screening services globally. Basic light microscopes and well-educated laypeople (nonpathologists) are readily available in all low- and middle-income countries. With basic capabilities for creating and processing FFPE tissues, any low- or middle-income country could rapidly develop good diagnostics for cervical precancer and cancer that support the expansion of cervical cancer screening programs to reduce the burden of this preventable cancer. A large-scale evaluation of this alternative approach to cervical diagnosis is warranted.

Acknowledgments

References

1. New guidelines on screening and treatment for cervical cancer. South Africa: World Health Organization; 2013. Reference type: pamphlet

2. Adesina A, Chumba D, Nelson AM, Orem J, Roberts DJ, Wabinga H, et al. Improvement of pathology in sub-Saharan Africa. Lancet Oncol. 2013; 14:e152–7. [PubMed: 23561746]

3. Darragh TM, Colgan TJ, Cox JT, Heller DS, Henry MR, Luff RD, et al. The Lower Anogenital Squamous Terminology Standardization Project for HPV-Associated Lesions: background and consensus recommendations from the College of American Pathologists and the American Society for Colposcopy and Cervical Pathology. J Low Genit Tract Dis. 2012; 16:205–42. [PubMed: 22820980]

4. Zhao FH, Jeronimo J, Qiao YL, Schweizer J, Chen W, Valdez M, et al. An evaluation of novel, lower-cost molecular screening tests for human papillomavirus in rural China. Cancer Prev Res (Phila). 2013; 6:938–48. [PubMed: 23878179]

5. Qiao YL, Jeronimo J, Zhao FH, Schweizer J, Chen W, Valdez M, et al. Lower cost strategies for triage of human papillomavirus DNA-positive women. Int J Cancer. 2014; 134:2891–901. [PubMed: 24248915]

6. Galgano MT, Castle PE, Atkins KA, Brix WK, Nassau SR, Stoler MH. Using biomarkers as objective standards in the diagnosis of cervical biopsies. Am J Surg Pathol. 2010; 34:1077–87. [PubMed: 20661011]

7. Redman R, Rufforny I, Liu C, Wilkinson EJ, Massoll NA. The utility of p16(Ink4a) in discriminating between cervical intraepithelial neoplasia 1 and nonneoplastic equivocal lesions of the cervix. Arch Pathol Lab Med. 2008; 132:795–9. [PubMed: 18466028]

8. Keating JT, Cviko A, Riethdorf S, Riethdorf L, Quade BJ, Sun D, et al. Ki-67, cyclin E, and p16INK4 are complimentary surrogate biomarkers for human papilloma virus-related cervical neoplasia. Am J Surg Pathol. 2001; 25:884–91. [PubMed: 11420459]

9. Liao GD, Sellors JW, Sun HK, Zhang X, Bao YP, Jeronimo J, et al. p16INK4A immunohistochemical staining and predictive value for progression of cervical intraepithelial neoplasia grade 1: a prospective study in China. Int J Cancer. 2014; 134:1715–24. [PubMed: 24105727]

10. Crum CP. Our wages of CIN. Obstet Gynecol. 2012; 120:1261–2. [PubMed: 23168747]

11. Castle PE, Rodriguez AC, Burk RD, Herrero R, Wacholder S, Alfaro M, et al. Short term persistence of human papillomavirus and risk of cervical precancer and cancer: population based cohort study. BMJ. 2009; 339:b2569.doi: 10.1136/bmj.b2569.:b2569 [PubMed: 19638649]

12. Rodriguez AC, Schiffman M, Herrero R, Wacholder S, Hildesheim A, Castle PE, et al. Rapid clearance of human papillomavirus and implications for clinical focus on persistent infections. J Natl Cancer Inst. 2008; 100:513–7. [PubMed: 18364507]

13. Kjaer SK, Frederiksen K, Munk C, Iftner T. Long-term absolute risk of cervical intraepithelial neoplasia grade 3 or worse following human papillomavirus infection: role of persistence. J Natl Cancer Inst. 2010; 102:1478–88. [PubMed: 20841605]
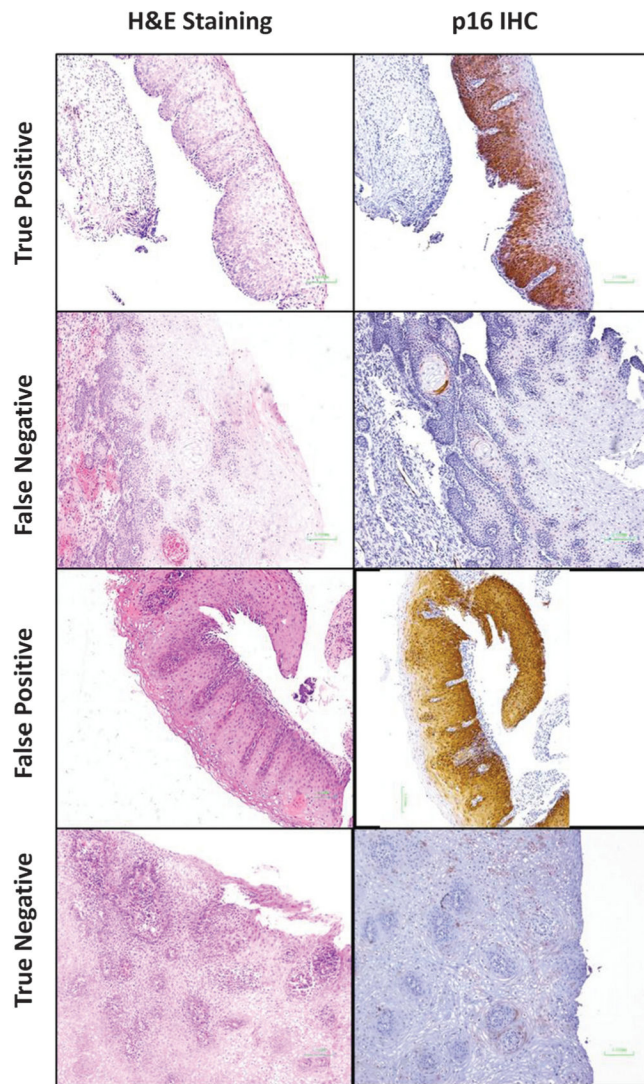
FIGURE 1.
Examples of hematoxylin and eosin (H&E) staining and p16INK4a immunohistochemical (p16 IHC) staining for diagnoses of cervical intraepithelial neoplasia grade 2 (CIN2) or more severe diagnosis (CIN2+).

TABLE 1

Characteristics of the Nonpathologist Readers of the p16INH4a Immunohistochemistry

Reader ID	Gender	Age	Education level	Professional background	Training experience in pathology	Experience in microscope operation
1	Female	25	Doctoral candidate	Preventive medicine	None	General
2	Female	25	Master's degree candidate	Preventive medicine	None	General
3	Male	22	Undergraduate	Preventive medicine	None	General
4	Female	24	Master's degree candidate	Preventive medicine	None	General
5	Female	27	Undergraduate	Laboratory	None	General
6	Female	24	Master's degree candidate	Preventive medicine	None	Skilled
7	Female	29	Doctoral candidate	Preventive medicine	None	General
8	Female	23	Undergraduate	Preventive medicine	None	General
9	Female	25	Master's degree candidate	Preventive medicine	None	General
10	Female	23	Master's degree candidate	Preventive medicine	None	General
11	Male	25	Master's degree candidate	Preventive medicine	None	General
12	Male	28	Doctoral candidate	Preventive medicine	None	General

TABLE 2

Clinical Performance of p16INK4a Immunochemistry as Evaluated by Nonpathologist Readers and a Pathologist to Identify Women With Histologic Diagnosis of CIN2 or More Severe Diagnoses (CIN2+)

	Reader result				Pathologist result			
	Sensitivity (%)	Specificity (%)	YI	PLR	Sensitivity (%)	Specificity (%)	YI	PLR
Reader 1	96	96	92	24	96	96	92	24
Reader 2	96	80	76	4.8	96	88	84	8.0
Reader 3	88	88	76	7.3	96	88	84	8.0
Reader 4	72	96	68	18	96	96	92	24
Reader 5	92	76	68	3.8	96	88	84	8.0
Reader 6	96	84	80	6.0	96	96	92	24
Reader 7	96	84	80	6.0	96	88	84	8.0
Reader 8	96	84	80	6.0	96	88	84	8.0
Reader 9	76	88	64	6.3	96	96	92	24
Reader 10	92	88	80	7.7	96	88	84	8.0
Reader 11	76	92	68	9.5	96	96	92	24
Reader 12	84	88	72	7.0	96	96	92	24
Mean/Overall[a]	88	87	75	6.8	96	92	88	12

YI, Youden Index (= sensitivity + specificity − 1); PLR, positive likelihood ratio.

[a] The average performance of all reads for all tissues was calculated. The pathologist read all slides only once and the overall performance is presented.