Role of quantitative p16\textsuperscript{INK4A} mRNA assay and digital reading of p16\textsuperscript{INK4A} immunostained sections in diagnosis of cervical intraepithelial neoplasia

Nataša Vasiljević\textsuperscript{1}, Paul D. Carter\textsuperscript{2}, Caroline Reuter\textsuperscript{1}, Rhian Warman\textsuperscript{1}, Adam R. Brentnall\textsuperscript{1}, James R. Carton\textsuperscript{3}, Jack Cuzick\textsuperscript{1} and Attila T. Lorincz\textsuperscript{1}

\textsuperscript{1}Centre for Cancer Prevention, Wolfson Institute of Preventive Medicine, Barts and The London School of Medicine, Queen Mary University of London, Charterhouse Square, London, United Kingdom
\textsuperscript{2}Centre for Molecular Pathology, Royal Marsden Hospital, Sutton, United Kingdom
\textsuperscript{3}Department of Histopathology, Charing Cross Hospital, London, United Kingdom

Visual interpretation of cervical biopsies is subjective and variable, generally showing fair to moderate inter-reader agreement in distinguishing high from low grade cervical intraepithelial neoplasia (CIN). We investigated the performance of two objective p16 quantitative tests in comparison with visual assessment: (i) p16-mRNA assay and (ii) digital analysis of sections stained for p16 protein.

The primary analysis considered 232 high-risk human papilloma virus positive (HPV+) samples from diagnostic cervical specimens. A p16 RT-qPCR (p16-mRNA assay) was run on mRNA extracted from formalin-fixed paraffin-embedded sections. Two p16 immunohistochemistry (IHC) readings, a visual read by a histopathologist (Visual IHC) and a digital read of a high-resolution scan (Digital IHC), were done on adjacent sections. The worst reviewed CIN grade (agreed by at least two histopathologists) from up to two biopsies and a loop excision was taken, with CIN2/3 as the primary endpoint.

Visual IHC attained a specificity of 70\% (95\%CI 61–77) for 85\% (95\%CI 77–91) sensitivity. The four-point Visual IHC staining area under the curve (AUC) was 0.77 (95\%CI 0.71–0.82), compared with 0.71 (95\%CI 0.64–0.77) for p16-mRNA and 0.67 (95\%CI 0.60–0.74) for Digital IHC. Spearman rank-order correlations were: visual to p16-mRNA 0.41, visual to digital 0.49 and p16-mRNA to digital: 0.22. The addition of p16-mRNA assay to visual reading of p16 IHC improved the AUC from 0.77 to 0.84 (p = 0.0049). p16-mRNA testing may be complementary to visual IHC p16 staining for a more accurate diagnosis of CIN, or perhaps a substitute in locations with a lack of skilled pathologists.

Key words: p16 RT-qPCR, p16 Immunohistochemistry, digital H-score, human papilloma virus, cervical intraepithelial neoplasia

Abbreviations: AUC: area under curve; CIN: cervical intraepithelial neoplasia; IHC: immunohistochemistry; HPV: human papilloma virus; LLETZ: large loop excision of the transformation zone; p16: p16\textsuperscript{INK4A} protein; RT-qPCR: real time quantitative polymerase chain reaction

Additional Supporting Information may be found in the online version of this article.
Grant sponsor: Cancer Research UK programme grant; Grant number: CS69/A10404
DOI: 10.1002/ijc.30783

Infection by high risk human papilloma virus (hrHPV) and subsequent failure to attend for cervical screening are the principal factors responsible for high cervical cancer burden. After host exposure to HPV, the virus may become established in tissues as one or more areas of normal-appearing epithelium which over time remain visually flat or perhaps assume warty features. In early stages of HPV infection, the histopathological examination of biopsies usually appears morphologically normal. They might, however, have various minor changes or sometimes show architecture consistent with cervical intraepithelial neoplasia grade 1 (CIN1). Over a period of years persistent hrHPV infection can transform the affected tissue into high grade intraepithelial disease, either CIN2 or the more worrisome CIN3, which are regarded as precancerous lesions.\textsuperscript{1,2}

hrHPV infections and associated low grade lesions are quite common, but almost all spontaneously clear within a few years.\textsuperscript{3–6} Depending on specific diagnostic criteria, expertise of the pathologist and level of surveillance an estimated 50\% of CIN2 may regress without intervention.\textsuperscript{7} The regression rate of CIN3 is incompletely assessed due to ethical considerations, but it is believed that a small percentage of CIN3
What’s new?

Visual examination of cervical tissue obtained by biopsy provides information on morphological changes associated with human papillomavirus (HPV) infection and transformation to precancerous cervical intraepithelial neoplasia 2/3 (CIN2/3). The accuracy of CIN2/3 diagnosis by visual interpretation, however, could be improved through the use of complementary objective tests. This study describes two such tests: quantitative p16 mRNA analysis and digital analysis of immunohistochemical p16-protein staining. CIN2/3 diagnostic accuracy benefitted especially from the addition of p16-mRNA testing. Both new approaches could prove useful in locations lacking trained pathologists and could help avoid overtreatment in HPV-positive women, many of whom recover spontaneously.

Int. J. Cancer: 141, 829–836 (2017) © 2017 The Authors International Journal of Cancer published by John Wiley & Sons Ltd on behalf of UICC
biopsies and LLETZ were retrieved from the Predictors 2 study. For the present study, the first biopsies were also reviewed by a different histopathologist (JRC). To maximise the likelihood of an accurate diagnosis, the worst reviewed histology of the first biopsy was based on JRC’s review and the Predictors 2 expert consultant pathologist, while hospital grading was used to arbitrate (Supporting Information methods). The worst concordant diagnosis on any biopsy (first, second and if available the LLETZ surgical specimen) was used to determine the final reference diagnosis for comparisons of the p16 tests.

**HPV testing**

HPV status of the 287 patients was determined in the Predictors 2 study by the Hybrid Capture 2 HPV test (Qiagen, Hilden, Germany) and the BD HPV test (Becton Dickinson Diagnostics, Sparks, MD) on exfoliated cervical cells from the concurrent sample.22 hrHPV positive samples were defined as specimens positive by either of these two tests.

**p16 IHC for visual and digital reading**

Two 5 µm sections were cut from the FFPE first biopsy blocks for H&E and p16 IHC staining. p16 IHC was done using the CINtec p16 Histology assay (Ventana, Tucson, AZ) according to the manufacturer’s instructions. The histopathologist (JRC) performed the visual p16 readings, henceforth referred to as Visual IHC, and was blinded to all patient data. Visual IHC was scored using a semi-quantitative scale with four categories: normal, sporadic, focal or diffuse.23 Diffuse staining was the cut-off point for calling specimens positive.25 The p16 stained sections were also scanned using a Panoramic 250 digital slide scanner (3D Histech Kft, Budapest, Hungary). The images were analysed using the Nuclear Quant module (version 1.15.1.9) from which the H-score (3 × percentage of strongly staining nuclei + 2 × percentage of moderately staining nuclei + percentage of weakly staining nuclei, giving a range of 0–300) was obtained, henceforth referred to as Digital IHC.

**p16-mRNA assay**

Six 10 µm sections from the first biopsies were cut directly into 1.5 ml microcentrifuge tubes and stored at −70°C. The tissue sections were deparaffinised in 320 µl of n-Hexadecane (Alfa Aesar, Heysham, UK) at 56°C for 3 min. Total RNA was extracted using the RNeasy FFPE Kit (Qiagen) according to the manufacturer’s instructions. A Nanodrop 1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA) was used to quantify RNA and up to 300 ng of DNase-treated total RNA was reverse transcribed using the QuantiTect...
Reverse Transcription Kit (Qiagen) according to the manufacturer’s instructions. A region of the gene RPLP0 was selected after optimization experiments as the reference gene. For this test, equal amounts of mRNA from 20 samples were tested on six housekeeping genes obtained from Qiagen: RPLP0 (Cat. no. QT00075012), ACTB (Cat. no. QT00095431), B2M (Cat. no. QT00088935), GAPDH (Cat. no. QT01192646), HPRT1 (Cat. no. QT00059066) and PGK1 (Cat. no. QT00013776). RPLP0 showed the most stable expression (data not shown). Further, p16 (Cat. no. QT00089964) and RPLP0 primers were tested for amplification efficiency and specificity using a 10-fold dilution of Stratagene QPCR Human Reference Total RNA (Agilent Technologies, Santa Clara, CA) and melt curve analysis. Measurement of p16 expression, hereafter called p16-mRNA assay, was performed on a Rotor-Gene 6000 (Qiagen) with the following cycling conditions: 95°C for 5 min, then 40 cycles of 5 sec at 95°C and 10 sec at 60°C with a final melting curve measuring step from 60 to 99°C. Two microliters (~25 ng) cDNA was amplified in a volume of 25 μl containing Rotor-Gene SYBR Green PCR master mix (Qiagen) and 1× dilution of p16 or RPLP0 primer as per manufacturer’s instructions. Each sample was run in duplicate and the average Ct value was calculated. qPCR runs included a four points standard curve, a No-RT control and a No Template Control in duplicate for both primers. Any sample with a Ct value for RPLP0 >24 was excluded from the analysis due to insufficient amount of cDNA.

**Statistical analyses**

The expression stability of the reference genes was determined using BestKeeper (version 1) software. The relative expression of p16 was calculated using the ΔΔCt method. Fold expression of p16 in CIN1, CIN2 and CIN3 was analysed in relation to the median expression in normal biopsies. A Cuzick test for trend and a Mann-Whitney U test were used to assess differences in p16 expression between diagnostic groups. Agreement between the histopathologist reviews was measured with Cohen’s Kappa statistic. Correlations were investigated by the Spearman rank correlation coefficient. Logistic regression was applied to determine if p16 methods in combination performed better than Visual IHC alone in the first biopsy, and DeLong confidence intervals for area under the receiver operating characteristic curve (AUC) statistics were obtained. Statistical analyses used GraphPad Prism software (v5.03), Stata v11.2, and R. A statistical test was accepted as significant if the two-sided p value was <0.05.

**Results**

**p16 Expression**

Expression of p16 was successfully measured in 280 of 287 biopsies (Fig. 1) and a trend of increasing p16-mRNA level was observed with increasing lesion severity $\chi^2 = 45.5$ ($p < 0.0001$, Fig. 2). There was only one HPV negative patient diagnosed with CIN2+ and the p16-mRNA level in this woman was low, comparable with women with a normal diagnosis (Fig. 2). A higher level of p16 expression was observed in HPV positive women compared with HPV negative women, with a 7.2-fold difference observed between the HPV negative and HPV positive CIN1 groups ($p = 0.007$, Fig. 2).

**Agreement between histopathological reviews**

The agreement between the pathologists on the first biopsy, where all specimens were read by the two study experts and the hospital pathologist(s) for <CIN2 versus CIN2/3, was overall fair to good, with Kappa coefficients of 0.62 (95% CI 0.51–0.73), 0.78 (95% CI 0.69–0.87) and 0.64 (95% CI 0.54–0.75) for the pairwise comparisons (Supporting Information Tables S1–S3). Agreement was, however, poor if all diagnostic categories were taken under consideration, with Kappa coefficients of 0.27 (95% CI 0.17–0.37), 0.37 (95% CI 0.28–0.46) and 0.43 (95% CI 0.34–0.53) (Supporting Information Table S1–S3).

**Comparisons of p16 measures in hrHPV samples**

Performance of the three p16 assays was analysed in 232 hrHPV positive patients (Fig. 1, Supporting Information Table S4). The association between p16-mRNA and Digital IHC was weak (Spearman rank correlation 0.22, $p = 0.0009$, Fig. 3a). There was a modest but significant association between the p16-mRNA assay and Visual IHC (Spearman rank correlation 0.41, $p < 0.0001$, Fig. 3b), as well as between Digital and Visual IHC (Spearman rank correlation 0.49, $p < 0.0001$, Fig. 3c).

Diffuse staining based on Visual IHC had a CIN2+ sensitivity of 85% (95% CI 77–91%) and specificity of 70% (95% CI 61–77%; Table 1).
The four-point Visual IHC staining AUC was 0.77 (95% CI 0.71–0.82), compared with 0.71 (95% CI 0.64–0.77) for the p16-mRNA assay and 0.67 (95% CI 0.60–0.74) for Digital IHC (Fig. 4, Supporting Information Figure S1).

The Visual IHC combined with the p16-mRNA assay significantly improved the AUC from 0.77 for Visual IHC alone to 0.84 \( (p < 0.0049) \) for the combination test. In contrast, Digital IHC did not significantly improve the AUC of the Visual IHC test \( (p = 0.21) \); Table 2 and Fig. 4). A combination of p16-mRNA and Digital IHC also improved the ability to detect CIN2/3 compared with either test alone \( (AUC = 0.76, \Delta LR-\chi^2 = 13.8, p < 0.0002; \text{c.f. Fig. 3}) \).

### Discussion

This is the first study to measure p16-mRNA levels and utilise an automated computer-generated H-score from digitally scanned images of p16 IHC stained cervical biopsy sections. The performance of these two tests for identifying high grade intraepithelial disease was assessed in comparison with expert visual reading of p16 IHC staining patterns. The aim of the study was to determine the performance of the two quantitative p16 tests univariately and in combination with the Visual IHC results. While p16 staining of biopsies has been shown to improve inter-reader reliability,\(^{20,21}\) this standard method requires a subjective visual interpretation by a trained...
morphologist. Therefore, objective quantitative methods, such as those investigated here, could complement the expert pathologist H&E reading in routine practice. An objective p16 test would be particularly useful in low to middle-income regions where there may be a lack of sufficiently trained pathologists. We showed that p16-mRNA potential to improve the visual diagnosis based on H&E interpretation plus p16 visual scoring and may lead to a more accurate approach to diagnosing CIN2/3 (Fig. 4).

We based our study on a colposcopy referral population to increase power for the endpoint of interest (CIN2/3). Although only 287 of the first biopsies collected in the Predictors 2 cohort were available for p16 testing (Fig. 1), this provided ample power. The histological grade distribution was comparable with the complete colposcopy population. The original biopsies were read by the hospital histopathology team and the sections were subsequently reviewed by two expert histopathologists. Dichotomization of the data at a cut-off relevant to the decision to treat or not (<CIN2 and CIN2/3), showed good agreement on H&E readings between the histopathologists. However, taking into consideration all diagnostic categories, the agreement was poor (Supporting Information Tables S1–S3). In our study a trend of the expert reviewers downgrading low grade lesion diagnoses was observed, an occurrence that has previously been reported.27 Comparing the relative performance of the three p16 approaches tested here, the Visual IHC interpretation with diffuse staining as the cut-off21 yielded the best results (Table 1 and Fig. 4). Other studies based on visual interpretation of p16 staining have shown similar sensitivity and specificity to our study.28,29 The novel p16-mRNA assay performed similarly to the Visual IHC and, at 90% sensitivity, the specificity observed for the mRNA assay was not different than for Visual IHC (Fig. 4). In addition, logistic regression suggested that complementing p16 Visual IHC reading with a p16-mRNA assay could significantly improve accuracy of diagnosis from an initial AUC of 0.77 to an AUC of 0.84 (p = 0.0049) (Table 2 and Fig. 4). The Digital IHC had the weakest performance of all the tests (Fig. 4 and Supporting Information Figure S1). To the best of our knowledge, this utilization of H-score for p16 expression has not been investigated before. Although such a digital test would avoid the drawback of working with mRNA, a requirement of the p16-mRNA assay, the scanning equipment is quite expensive and would therefore also be less likely to be applicable in low income regions than use of an inexpensive PCR test.

There was a significant increase in expression of p16 (p<0.0001) in CIN2/3 compared to <CIN2 biopsies (Fig. 2), an observation supported by the previous reports that expression of p16 is elevated in higher grade lesions.17,18 As there was only one HPV negative CIN2+, no meaningful comparison could be done in this group. The HPV negative CIN1 lesions had significantly lower p16 expression than the HPV

![Figure 4. The performance of the different tests to detect CIN2/3 visualized with ROC curves. The circles present the categorical visually read p16 IHC data from Table 1 where the cut off s are based on D (diffuse), D + F (D + focal) and D + F+S (D + F+sporadic) staining. [Color figure can be viewed at wileyonlinelibrary.com](image)](image)

Table 2. Performance summary of the three tests

| p16 test   | LR-$\chi^2$ | Univariate (p-value) | Univariate AUC (95 CI%) | Combined AUC$^3$ | ALR-$\chi^2$ (df = 1) | Addition to Visual IHC (p-value) |
|------------|-------------|----------------------|-------------------------|-----------------|----------------------|----------------------------------|
| CIN2/3 Visual IHC | 74.9 | <0.00001 | 0.77 (0.71–0.82) |                  |                      |                                  |
| p16-mRNA  | 31.8 | <0.00001 | 0.71 (0.64–0.77) | 0.84 | 7.9 | 0.0049 |
| Digital IHC | 21.5 | <0.00001 | 0.67 (0.60–0.74) | 0.81 | 1.6 | 0.21 |
| CIN3 Visual IHC | 66.8 | <0.00001 | 0.77 (0.71–0.82) |                  |                      |                                  |
| p16-mRNA  | 26.3 | <0.00001 | 0.71 (0.64–0.79) | 0.85 | 7.4 | 0.0064 |
| Digital IHC | 28.3 | <0.00001 | 0.73 (0.65–0.80) | 0.84 | 4.7 | 0.029 |

1Likelihood ratio $\chi^2$.
2Univariate LR-$\chi^2$ is on 4 df for visual and 1 df for RT-qPCR and IHC digital.
3The combined AUC is for Visual with the quantitative measures fitted using a logistic regression.
positive CIN1 lesions ($p = 0.007$) as would be expected due to lack of HPV interference with the p16 pathway. Our study has several novel features including the comparisons of the p16-mRNA assay and the digitally read H-score to each other and to the visual readings. Although there have been some earlier studies of p16 expression using RT-qPCR in cervical specimens, all have been on small underpowered sample sets. Upregulation of p16 at both the mRNA and protein level in a small number of carcinomas and cell lines compared to normal tissue was reported in one study, while another reported a 6-fold increase in p16 mRNA expression in high-grade cytology specimens compared to normal cells and HPV-negative women.

An important limitation of our study was that it was performed on biopsy specimens and there were no comparisons to corresponding exfoliated cervical cells. Unfortunately, a pilot study found that the quality of RNA was too poor in this archived material to obtain valid RT-qPCR measurements (data not shown). The measurement of p16-mRNA levels in exfoliated cervical specimens deserves further attention in studies designed to address the mRNA instability problem. It is possible that a combination of p16-mRNA with other biomarkers such as DNA methylation in exfoliated cell specimens may provide a substantially improved triage test for reducing the number of women requiring referral to colposcopy.

In conclusion, we present two novel approaches, a p16-mRNA assay and digital reading of p16 IHC, to aid the objective diagnosis of CIN2/3. Application of these measures in a triage test might help to avoid unnecessary treatment in HPV-positive women. Visual interpretation of p16 IHC is likely to be the best available method, but our data show that use of a p16-mRNA assay may be complementary and improve the accuracy of the diagnosis. We also envisage the possibility of a future biomarker panel including p16-mRNA in settings where the diagnostic emphasis is on nucleic acid testing. Analysis of p16 levels alone or in combination with other biomarkers, shows promise to improve diagnosis of cervical intraepithelial neoplasia in locations with a lack of skilled pathologists.

**Acknowledgements**

The authors thank all the participating patients. They thank Louise Cadman and Janet Austin for providing us with the samples and collecting, updating and maintaining the Predictors 2 cohort. The assistance of Dr Mangesh Thorat was instrumental in designing the process to produce the H-score. The funders had no role in the study design, collection, analysis, interpretation of data or the writing of this report.

**Conflicts of Interest**

None
26. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing, retrieved from https://www.R-project.org 2016.

27. Stoler MH, Schiffman M, Undetermin ASC. Interobserver reproducibility of cervical cytologic and histologic interpretations - Realistic estimates from the ASCUS-LSIL triage study. JAMA 2001; 285:1500–5.

28. Galgano MT, Castle PE, Atkins KA, et al. Using Biomarkers as Objective Standards in the Diagnosis of Cervical Biopsies. Am J Surg Pathol 2010; 34:1077–87.

29. Wang SS, Trunk M, Schiffman M, et al. Validation of p16(INK4a) as a marker of oncogenic human papillomavirus infection in cervical biopsies from a population-based cohort in Costa Rica. Cancer Epipem Biomar 2004;13:1355–60.

30. Ivanova TA, Golovina DA, Zavalishina LE, et al. Up-regulation of expression and lack of 5' CpG island hypermethylation of p16 INK4a in HPV-positive cervical carcinomas. BMC Cancer 2007;7.

31. von Keyserling H, Kuhn W, Schneider A, et al. p16(INK4a) and p14(ARF) mRNA expression in Pap smears is age-related. Mod Pathol 2012;25:465–70.