

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

Short title: p16 mRNA aiding diagnosis of CIN2+

Nataša Vasiljević¹, Paul D. Carter², Caroline Reuter¹, Rhian Warman¹, Adam R Brentnall¹, James R. Carton³, Jack Cuzick¹, Attila T. Lorincz^{1§}

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 EC1M 6BQ, UK

2. Centre for Molecular Pathology, Royal Marsden Hospital, Downs Road, Sutton, SM2 5PT, UK

3. Department of Histopathology, Charing Cross Hospital, Fulham Palace Road, London, W6 8RF, UK

§Corresponding author: Attila T Lorincz, Centre for Cancer Prevention, Wolfson Institute of Preventive Medicine, Barts and The London School of Medicine, Queen Mary University of London, Charterhouse Square, London EC1M 6BQ, UK; Tel.: 020-7882-3540; Fax: 020-7882-3890; E-mail: a.lorincz@qmul.ac.uk

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^{INK4A} protein

RT-qPCR Real time quantitative polymerase chain reaction

Appropriate category: Tumor Markers and Signatures

Novelty and Impact:

We present two novel approaches to aid diagnosis of CIN2/3, based on either quantitative PCR, measuring p16 mRNA levels, or digitalised reading of immunohistochemical p16 protein staining.

The main strength of these tests is their objectivity which may improve diagnosis of cervical intraepithelial neoplasia in locations with a lack of skilled pathologists. Furthermore, the application of these methods may improve accuracy of current diagnostics and therefore help avoid unnecessary treatments.

Abstract

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: 1) p16-mRNA assay and 2) digital analysis of sections stained for p16 protein.

The primary analysis considered 232 high-risk 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.

Introduction

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 [1, 2].

hrHPV infections and associated low grade lesions are quite common, but almost all spontaneously clear within a few years [3-6]. Depending on specific diagnostic criteria, expertise of the pathologist and level of surveillance an estimated 50% of CIN2 may regress without intervention [7]. The regression rate of CIN3 is incompletely assessed due to ethical considerations, but it is believed that a small percentage of CIN3 ever become invasive [8, 9]. Since the 1960s, in most countries of Europe and North America, cytological screening programmes have been implemented with 3 to 5 yearly screening intervals starting in a woman's mid-twenties. Women with moderate or severe dysplasia on cytology (also called HSIL) as well as hrHPV-positive women with mild or borderline abnormal findings (also called LSIL or ASCUS) are referred to colposcopy [10]. Some countries are also making the transition to use HPV testing as a primary screen [11]. The decision on who receives preventative excision treatment is predominantly based on a pathologist's CIN grading of biopsies collected during colposcopy, CIN2 being the most commonly set threshold for intervention [12]. Although the success of this approach is indisputable [13], it is difficult to implement effectively in low-resource settings. Morphological screening and grading methods are subjective and prone to variability with CIN3 and cancer shown to be the most, and CIN2 the least, reproducible diagnosis [14, 15]. This

might lead to unnecessary treatment, which is a concern due to the potential harm to patients and from a monetary perspective. After treatment, women commonly experience substantial pain and discomfort for weeks, however, the main morbidity of treatment is subsequent cervical insufficiency, leading to a higher risk of 2nd and 3rd trimester miscarriage and preterm delivery [16]. Biomarkers that improve triage of women to colposcopy and provide for a more accurate diagnosis of CIN3 destined to progress are therefore highly desirable.

Elevation of p16^{INK4A} protein, henceforth denoted as p16, in HPV-positive women may indicate a high-grade lesion with increased risk of development to cancer [17, 18]. p16 would normally be suppressed by retinoblastoma protein (pRb), however if a hrHPV infection persists, a continuous expression of the E6 and E7 oncoproteins inactivate p53 and pRb leading to p16 overexpression. p16 overexpression may therefore be considered a surrogate marker of persistent HPV infection and measurement of its expression allows triage of a higher risk group of hrHPV-positive women to colposcopy [19]. In addition, co-reading of p16 and hematoxylin and eosin (H&E) stained biopsies has been shown to improve observer agreement in the diagnosis of CIN [20, 21]. The reported sensitivity of p16 immunohistochemistry (IHC), however, varies due to difficulties in reproducible reading of the slides, lack of standardised reporting, and uncertainty about the best cut-off point. Detecting the presence of p16 mRNA by a molecular test rather than staining biopsy sections for the presence of the protein could diminish the subjectivity of p16 testing. In this study, we explore the potential of a novel real time (RT)-qPCR assay for measuring p16 mRNA expression to improve grading of CIN lesions in a set of formalin-fixed paraffin-embedded (FFPE) cervical biopsies from a UK colposcopy referral population [22]. As a comparison of performance, we also used p16 IHC staining graded by both an expert histopathologist (Visual IHC) and analysed as an H-Score by an automated digital scanner module (Digital IHC).

Materials and Methods

Patients

The Predictors 2 study comprised 1,099 women referred for colposcopy in London between September 2007 and October 2009 following an abnormal cytology result [22]. Women attended one or two colposcopies. For the current study, p16 testing was performed on 287 biopsy specimens that were available from the first colposcopy visit (Figure 1). Some women were recalled for a second colposcopy and additional biopsies taken where abnormalities were seen. None of the second biopsies were used for p16 testing. Women who met the standard UK criteria for treatment were referred to an expert clinic for large loop excision of the transformation zone (LLETZ). LLETZ tissues as well as second biopsies had histopathological review data available which was considered for the final reference diagnosis in the current study. However, the three p16 tests investigated here were done only on the first biopsies. All women in Predictors 2 received a patient information sheet explaining the study and provided written consent. Approvals were obtained from the relevant local research ethics committees.

Histopathological Diagnosis

In our main analyses, we dichotomized the specimens into a set with the diagnosis of CIN2 or CIN3 (CIN2/3) versus the set with a normal or CIN1 (<CIN2) reference diagnosis. However, additional diagnostic categories were provided by the pathologists and used in some of our secondary analyses. For the Predictors 2 study, H&E stained sections of all study specimens were read by the routine hospital pathologist(s) and reviewed by an expert consultant pathologist. In case of disagreement, diagnosis was arbitrated by a second expert pathologist [22]. The worst reviewed histology of the second biopsies and LLETZ were retrieved from Predictors 2 study. For the present study, the first biopsies were reviewed by a different histopathologist (JRC). To maximise the likelihood of accurate diagnosis, the worst reviewed histology of the first biopsy was based on JRC

review and Predictors 2 expert consultant pathologist, while hospital grading was used to arbitrate (Supplemental methods). The worst concordant diagnosis on any biopsy (first, second and if available the LEETZ 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 Hybrid Capture 2 HPV test (Qiagen, Hilden, Germany) and the BD HPV test (Becton Dickinson Diagnostics, Sparks, Maryland, USA) on exfoliated cervical cells from the concurrent sample [22]. hrHPV positive samples were defined as specimens positive by either of these two tests.

p16 immunohistochemistry 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 USA) according to the manufacturer's instructions. The histopathologist (JRC) performed the visual p16 readings, hence 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 [21]. 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 x percentage of strongly staining nuclei + 2 x percentage of moderately staining nuclei + percentage of weakly staining nuclei, giving a range of 0 to 300) was obtained, hence referred to as Digital IHC.

p16-mRNA Assay

Six 10µm 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 320µl of n-Hexadecane (Alfa Aesar, Heysham, UK) at 56°C for 3 minutes. 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, USA) 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 optimisation 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), *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, US) and melt curve analysis. Measurement of p16 expression, thereafter 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 s at 95°C and 10 s 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 1X 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* greater than 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 [24]. The relative expression of p16 was calculated using the $\Delta\Delta C_t$ method [25]. 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 look for significant differences in p16 expression stratified by diagnostic groups. Agreement between the

histopathologist reviews was explored with Cohen's Kappa statistics. Correlations were investigated by 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 to predict the reference diagnosis and DeLong confidence intervals for area under the curve (AUC) statistics were obtained. Statistical analyses used GraphPad Prism software (v5.03), Stata v11.2, and R [26]. 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 (Figure 1) and a trend of increasing p16-mRNA level was observed with increasing lesion severity $\chi^2_1= 45.5$ ($p<0.0001$, Figure 2). There was only one HPV negative patient diagnosed with CIN2+ and the p16-mRNA level in this woman was low, comparable to the women with a normal diagnosis (Figure 2). Generally, a significantly higher level of p16 expression was observed in HPV positive women compared to HPV negative women with a 7.2-fold difference observed between the HPV negative and HPV positive CIN1 groups ($p=0.007$, Figure 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 hospital pathologist(s) for <CIN2 versus CIN2/3, was overall fair to good with Kappas 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 (Supplemental tables 1-3). Agreement was, however, poor if all diagnostic categories were taken under consideration, with Kappas 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) (Supplemental table 1-3).

Comparisons of p16 measures in hrHPV samples

Performance of the three p16 assays was analysed in 232 hrHPV positive patients (Figure 1). The summary statistics of the two quantitative tests were overviewed in the 232 patients (Supplemental table 4). The association observed between p16-mRNA and Digital IHC was weak (Spearman rank correlation 0.22, $p=0.0009$, Figure 3a). There was a modest but significant association between the p16-mRNA assay and Visual IHC (Spearman rank correlation 0.41, $p<0.0001$, Figure 3b), as well as between Digital and Visual IHC (Spearman rank correlation 0.49, $p<0.0001$, Figure 3c).

Diffuse staining based on Visual IHC had a sensitivity at 85% (95%CI 77-91%) and specificity 70% (95%CI 61-77%) (Table 1).

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 assay and 0.67 (95%CI 0.60-0.74) for Digital IHC (Figure 4, Supplemental Figure 1). The Visual IHC combined with the p16-mRNA assay significantly improved the diagnostic accuracy from AUC=0.77 for Visual IHC alone to AUC=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 Figure 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, $\Delta\text{LR-}\chi^2= 13.8$, $p<0.0002$; c.f. Figure 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 show that p16-mRNA has the 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 (Figure 4).

We based our study on a colposcopy referral population to increase power for the endpoint of interest (CIN2/3). Only 287 of the first biopsies collected for the Predictor 2 cohort were available for p16 testing (Figure 1), limiting the study to be conducted without a power calculation. However, despite that these samples were a convenience set, they were similar to the percentage of the histological grades seen in 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 (Supplemental tables 1-3). 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-off [21] yielded the best results (Table 1 and Figure 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 (Figure 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 Figure 4). The Digital IHC had the weakest performance of all the tests (Figure 4 and Supplemental Figure 1). To the best of our knowledge, this utilisation 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 (Figure 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 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 [30],

while another reported a 6-fold increase in p16 mRNA expression in high-grade cytology specimens compared to normal cells and HPV-negative women [31].

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. We 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. This work was supported by Cancer Research UK programme grant (grant number C569/A10404).

The funders had no role in the study design, collection, analysis, interpretation of data, or the writing of this report.

Conflicts of Interest

None

References

1. de Villiers EM, Fauquet C, Broker TR et al. Classification of papillomaviruses. *Virology* 2004; **324**: 17-27.
2. zur Hausen H. Papillomaviruses and cancer: from basic studies to clinical application. *Nature reviews Cancer* 2002; **2**: 342-50.
3. Cuschieri KS, Cubie HA, Whitley MW et al. Persistent high risk HPV infection associated with development of cervical neoplasia in a prospective population study. *Journal of clinical pathology* 2005; **58**: 946-50.
4. Meijer CJLM, Helmerhorst TJM, Rozendaal L et al. HPV typing and testing in gynaecological pathology: has the time come? *Histopathology* 1998; **33**: 83-6.
5. Moscicki AB, Shiboski S, Broering J et al. The natural history of human papillomavirus infection as measured by repeated DNA testing in adolescent and young women. *J Pediatr-Ur* 1998; **132**: 277-84.
6. Onda T, Carter JJ, Koutsky LA et al. Characterization of IgA response among women with incident HPV 16 infection. *Virology* 2003; **312**: 213-21.
7. Nasiell K, Nasiell M, Vaclavinkova V. Behavior of Moderate Cervical Dysplasia during Long-Term Follow-Up. *Obstet Gynecol* 1983; **61**: 609-14.
8. McCredie MRE, Paul C, Sharples KJ et al. Consequences in women of participating in a study of the natural history of cervical intraepithelial neoplasia 3. *Aust Nz J Obstet Gyn* 2010; **50**: 363-70.
9. McCredie MRE, Sharples KJ, Paul C et al. Natural history of cervical neoplasia and risk of invasive cancer in women with cervical intraepithelial neoplasia 3: a retrospective cohort study. *Lancet Oncology* 2008; **9**: 425-34.
10. Arbyn M, Anttila A, Jordan J et al. European Guidelines for Quality Assurance in Cervical Cancer Screening. Second Edition-Summary Document. *Annals of Oncology* 2010; **21**: 448-58.
11. Saville AM. Cervical cancer prevention in Australia: Planning for the future. *Cancer Cytopathol* 2016; **124**: 235-40.
12. Saslow D, Solomon D, Lawson HW et al. American Cancer Society, American Society for Colposcopy and Cervical Pathology, and American Society for Clinical Pathology Screening Guidelines for the Prevention and Early Detection of Cervical Cancer. *American journal of clinical pathology* 2012; **137**: 516-42.
13. Peto J, Gilham C, Fletcher O, Matthews FE. The cervical cancer epidemic that screening has prevented in the UK. *Lancet* 2004; **364**: 249-56.
14. Robertson AJ, Anderson JM, Beck JS et al. Observer Variability in Histopathological Reporting of Cervical Biopsy Specimens. *Journal of clinical pathology* 1989; **42**: 231-8.
15. Stoler MH, Ronnett BM, Joste NE et al. The Interpretive Variability of Cervical Biopsies and Its Relationship to HPV Status. *Am J Surg Pathol* 2015; **39**: 729-36.
16. Kyrgiou M, Koliopoulos G, Martin-Hirsch P et al. Obstetric outcomes after conservative treatment for intraepithelial or early invasive cervical lesions: systematic review and meta-analysis. *Lancet* 2006; **367**: 489-98.
17. Doeberitz MV. New molecular tools for efficient screening of cervical cancer. *Dis Markers* 2001; **17**: 123-8.
18. Doeberitz MV. New markers for cervical dysplasia to visualise the genomic chaos created by aberrant oncogenic papillomavirus infections. *European journal of cancer* 2002; **38**: 2229-42.
19. Wentzensen N, Bergeron C, Cas F et al. Triage of women with ASCUS and LSIL cytology - Use of qualitative assessment of p16(INK4a) positive cells to identify patients with high-grade cervical intraepithelial neoplasia. *Cancer Cytopathol* 2007; **111**: 58-66.

20. Klaes R, Benner A, Friedrich T et al. p16(INK4a) immunohistochemistry improves interobserver agreement in the diagnosis of cervical intraepithelial neoplasia. *Am J Surg Pathol* 2002; **26**: 1389-99.
21. Reuschenbach M, Wentzensen N, Dijkstra MG et al. p16(INK4a) Immunohistochemistry in Cervical Biopsy Specimens A Systematic Review and Meta-Analysis of the Interobserver Agreement. *American journal of clinical pathology* 2014; **142**: 767-72.
22. Szarewski A, Mesher D, Cadman L et al. Comparison of seven tests for high-grade cervical intraepithelial neoplasia in women with abnormal smears: the Predictors 2 study. *Journal of clinical microbiology* 2012; **50**: 1867-73.
23. Klaes R, Friedrich T, Spitkovsky D et al. Overexpression of p16(ink4a) as a specific marker for dysplastic and neoplastic epithelial cells of the cervix uteri. *International Journal of Cancer* 2001; **92**: 276-84.
24. Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper - Excel-based tool using pair-wise correlations. *Biotechnol Lett* 2004; **26**: 509-15.
25. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. *Methods* 2001; **25**: 402-8.
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-J Am Med Assoc* 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 Epidem 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. *Modern Pathol* 2012; **25**: 465-70.

Table 1. The sensitivity and specificity of the Visual p16 IHC for CIN2/3 or CIN3 using combinations of the scoring categories, diffuse (D), focal (F), Sporadic (S) as the cut off.

Cut off	Positive tests (Number)	Sensitivity	Positive tests (Number)	Specificity
CIN2/3				
D	85	0.85 (95%CI 0.77-0.91)	40	0.70 (95%CI 0.61-0.77)
D+F	88	0.88 (95%CI 0.80-0.93)	69	0.48 (95%CI 0.39-0.56)
D+F+S	90	0.90 (95%CI 0.83-0.94)	84	0.36 (95%CI 0.29-0.45)
CIN3				
D	58	0.89 (95%CI 0.79-0.95)	40	0.70 (95%CI 0.61-0.77)
D+F	60	0.92 (95%CI 0.83-0.97)	69	0.48 (95%CI 0.39-0.56)
D+F+S	61	0.94 (95%CI 0.85-0.98)	84	0.36 (95%CI 0.29-0.45)

Table 2. Performance summary of the three tests. The combined AUC is for Visual with the quantitative measures fitted using a logistic regression.

p16 test	LR- χ^2 ¹⁾	Univariate (p-value)	Univariate AUC (95CI%)	Combined AUC	Δ LR- χ^2 (df=1)	Addition to Visual IHC (p-value)
CIN2/3 Visual IHC	74.9	<0.0001	0.77 (0.71-0.82)			
p16-mRNA	31.8	<0.0001	0.71 (0.64-0.77)	0.84	7.9	0.0049
Digital IHC	21.5	<0.0001	0.67 (0.60-0.74)	0.81	1.6	0.21
CIN3 Visual IHC ²⁾	66.8	<0.0001	0.77 (0.71-0.82)			
p16-mRNA	26.3	<0.0001	0.71 (0.64-0.79)	0.85	7.4	0.0064
Digital IHC	28.3	<0.0001	0.73 (0.65-0.80)	0.84	4.7	0.029

¹⁾ Likelihood ratio- chi square

²⁾ Univariate LR- χ^2 is on 4df for Visual, and 1df for RT-qPCR and IHC digital.

Figure.1 Flowchart showing the number of valid tests results for the women selected for the study.

Figure 2. p16-mRNA expression fold-difference in the first biopsies of 280 women stratified by worst reviewed histology and HPV status. Fold expression in CIN1, CIN2, CIN3 is shown in relation to the median expression of 48 normal biopsies. The Cuzick trend test confirmed a significant increasing trend ($p < 0.0001$) with lesion severity in HPV positive women. Higher p16 expression was observed in CIN1 women who were HPV positive compared to HPV negative ($p = 0.007$).

Figure 3. Comparison between three p16 measures. (a) p16 mRNA assay (RT-qPCR) vs Digital IHC (IHC digital), (CIN2+ +; < CIN1 circle); (b) p16-mRNA assay vs Visual IHC staining categories (Visual; N=Normal, S=Sporadic, F=Focal; D=Diffuse; grey=CIN2+, white = < CIN2); (c) Digital IHC vs Visual IHC staining categories (same format as (b)).

Figure 4. The performance of the different tests to detect CIN2/3 visualised with univariate ROC curves. The circles present the categorical visually read p16 IHC data from Table 1 where the cut offs are based on D (diffuse), D+F (D+focal) and D+F+S (D+F+sporadic) staining.