Manuscript Information

Journal name:	Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc
Manuscript #:	62500
Manuscript Title:	Stratification of HPV-Induced Cervical Pathology using the Virally-Encoded Molecular Marker E4 in Combination with p16 and MCM
Principal Investigator:	J Doorbar (jd121@cam.ac.uk)
Submitter:	Nature Publishing Group (repositorynotifs@nature.com)

Manuscript Files

Type	Fig/Table #	Filename	Size	Uploaded
manuscript	1	article_1.docx	216873	2015-03-06 08:02:27
figure	1	figure_1.tif	4697364	2015-03-06 08:02:28
figure	2	figure_2.tif	71062752	2015-03-06 08:02:47
figure	3	figure_3.tif	51728456	2015-03-06 08:03:01
figure	4	figure_4.tif	51134704	2015-03-06 08:03:15
figure	5	figure_5.tif	2615728	2015-03-06 08:03:15
figure	6	figure_6.tif	2154524	2015-03-06 08:03:16
figure	7	figure_7.tif	1457204	2015-03-06 08:03:17

Stratification of HPV-Induced Cervical Pathology using the Virally-Encoded Molecular Marker E4 in Combination with p16 or MCM

Heather Griffin[1,2], Yasmina Soneji[2], Romy Van Baars[3], Rupali Arora[4], David Jenkins[3], Miekel van de Sandt[3], Zhonglin Wu[2], Wim Quint[3], Robert Jach[5], Krzysztof Okon[5], Hubert Huras[5], Albert Singer[4] and John Doorbar[1,2*]

1. Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QP, UK
2. National Institute for Medical Research, Mill Hill, London NW7 1AA UK.
3. DDL Diagnostic Laboratory, Visseringlaan 25, 2288 ER, Rijswijk, The Netherlands.
4. University College Hospital, Euston Road, London, NW1 2BU, UK.
5. Department of Gynecology and Oncology, Jagiellonian University College, 31-501 Krakow, Poland

*Corresponding author: John Doorbar, Department of Pathology, Tennis Court Road, Cambridge, CB2 1QP, United Kingdom; Phone +44 (0)1223 336633; email: jd121@cam.ac.uk

Running title: Cervical Neoplasia Molecular Stratification

Keywords; Human Papillomavirus, Cervical Neoplasia, Molecular Biomarkers, E4, Cervical Screening, Disease Stratification

Abstract

High-risk HPV types cause cervical lesions of varying severity, ranging from transient productive infections to high-grade neoplasia. Disease stratification requires the examination of lesional pathology, and possibly also the detection of biomarkers. $P16^{INK4a}$ and MCM are established surrogates of high-risk HPV E6/E7 activity, and can be extensively expressed in high-grade lesions. Here we have combined these two cellular biomarkers with detection of the abundant HPV-encoded E4 protein in order to identify both productive and transforming lesions. This approach has allowed us to distinguish true papillomavirus infections from similar pathologies, and has allowed us to divide the heterogeneous CIN2 category into those that are CIN1-like and express E4, and those that more closely resemble non-productive CIN3. To achieve this, 530 lesional areas were evaluated according to standard pathology criteria and by using a multiple staining approach that allows us to superimpose biomarker patterns either singly or in combination onto an annotated haematoxylin & eosin image. Conventional grading of neoplasia was established by review panel, and compared directly to the composite molecular pathology visualised on the same tissue section. The detection of E4 coincided with the onset of vacuolation, becoming abundant in koilocytes as the MCM marker declined and cells lost their defined nuclear margins as visualised by standard H&E staining. Of the dual marker approaches, $p16^{INK4a}$ and E4 appeared most promising, with E4 generally identifying areas of low-grade disease even when $p16^{INK4a}$ was present. Extensive $p16^{INK4a}$ expression usually coincided with an absence of E4 expression or its focal retention in sporadic cells within the lesion. Our results suggest that a straightforward molecular evaluation of HPV life-cycle deregulation in cervical neoplasia may help improve disease stratification, and that this can be achieved using complementary molecular biomarker pairs such as MCM/E4 or more promisingly, $p16^{INK4a}$/E4 as an adjunct to conventional pathology.

Introduction

The International Agency for Research on Cancer states that 4.8% (610,000) of new cancers occurring in 2008 worldwide were attributable to human papillomavirus (HPV) infection, with substantially higher incidence and mortality rates in developing versus developed countries (1, 2). HPV has a causal aetiology for cancers of the cervix uteri, penis, vulva, vagina, anus and oropharynx, including the base of the tongue and tonsils (3, 4). To date, more than 200 different types of HPV have been identified on the basis of sequence analysis (http://pave.niaid.nih.gov/#home, (5)), with each type showing a defined epithelial tropism and disease association (3, 6). High-risk Alpha HPV (high-risk HPV) types are the primary cause of cervical neoplasia and cervical cancer, a disease that affects over half a million women per year worldwide (3, 7, 8).

In squamous cervical epithelium, the mitotically active basal and parabasal cells occupy the cell layers immediately above the basal lamina (9). As cells enter the mid zone, they begin to mature, and show a general increase in the volume of clear vacuolated cytoplasm containing glycogen. The differentiated superficial cells are characterised by larger areas of cytoplasm and smaller pyknotic nuclei (10), with the HPV life cycle being linked to the differentiation status of this epithelium. HPV infects actively dividing basal cells during wound healing (11), with the viral genome maintaining itself episomally at a low copy number. In the lower epithelial layers, high-risk-HPV express E6 and E7 proteins in order to stimulate cell proliferation and to ensure that differentiating cells are retained in the cell cycle (12, 13). These viral proteins stimulate the synthesis of cellular proteins necessary for S-phase entry, allowing the replication of viral episomes and initiation of the productive viral phase. As virus-infected cells approach the epithelial surface, an increase in late promoter activity leads to an elevation in replication (E1, E2), and accessory proteins (e.g. E4), followed by the assembly of infectious virions in the superficial cell layers (3). Current thinking suggest that the levels and/or activity of E6 and E7 are lower in CIN1 (low-grade squamous-intraepithelial lesion (LSIL) or mild dysplasia) where basaloid cells occupy the lower third of epithelium, than in CIN2/3 (high-grade squamous-intraepithelial lesion (HSIL) or moderate to severe dysplasia), where such cells occupy from two-thirds to the full thickness of the epithelium (14). Such deregulated viral gene expression is thought to underlie pathologic phenotype, and predisposes the cell to the accumulation of genetic errors and the eventual progression to cancer (14). High-grade CIN (CIN2/3; HG-CIN) lesions often do not support a productive virus life-cycle, but represent a transforming viral phase or abortive virus life cycle, even though the majority of cells within the lesion still maintain intact viral episomes (15, 16). Most cervical HPV infections are cleared within 1-2 years. However, women with persistent high-risk HPV infections can develop HG-CIN (17-19). Current treatments for HG-CIN focus on eliminating abnormal HPV-infected precancerous cells by surgical excision. The use of molecular criteria that distinguish lesions supporting an abortive or transforming, rather than a productive phase could thus be of value in clinical practice (20).

Our understanding of the molecular biology of HPV infection and the organisation of the HPV life-cycle during cancer progression provides a rational basis for marker selection (21). A major class of biomarkers useful in cervical screening represents proteins activated as a consequence of expression of the viral oncogenes E6 and E7. E7 associates with the Rb cell cycle regulator and releases the E2F transcription factor, which subsequently activates genes necessary for DNA replication (22). The most extensively evaluated cell cycle markers include $p16^{INK4a}$, MCM (minichromosome maintenance protein), and Ki-67 (23-25). When present in the upper layers of cervical lesions, these proteins can be regarded as surrogate markers of viral oncogene expression (26, 27). In low-grade cervical lesions, such proteins are typically confined to the lower epithelial layers, but extend into the higher epithelial layers in HSIL. Marked inter-and intra-observer variability even on histology indicates the need for biomarkers that are specific, sensitive and reproducible (23, 24).

The purpose of this study was to correlate the two most widely used molecular surrogates of E6/E7 expression and life-cycle deregulation ($p16^{INK4a}$ and MCM) with detection of the abundant HPV E4 protein, which is highly expressed in productive HPV infections and which marks the initiation of the late stage of the virus life cycle. Previous studies have suggested that these two distinct classes of biomarker are complementary and that they could be used together to establish grade of neoplasia and to distinguish virus infections from other similar pathologies that may have a different prognosis (12, 28). In this study we have correlated the expression of all three biomarkers with the precise pathology characteristics that are currently used to establish lesion grade, in order to establish whether a combined molecular and standard pathology approach might offer advantages over conventional methods in the future. To do this we have carried out multiple biomarker stains on tissue sections containing different grades of neoplasia, followed by the digital overlay of each biomarker pattern either singly or in combination onto the haematoxylin and eosin (H&E) pathology. Our results indicate that a dual biomarker approach using E4 and $p16^{INK4a}$ can distinguish HPV-associated CIN1 from other, probably non-viral pathologies, and may be used to divide the CIN2 group according to the extent of life-cycle deregulation.

Materials and methods

Clinical samples

Loop electrosurgical excision procedure (sometimes referred to as LEEP) and hysterectomy specimens from women treated at the Jagiellonian University Medical College, Krakow, Poland (103 biopsy specimens), as well as punch biopsies taken from women during colposcopy at the Gynaecological outpatient clinic of Hospital Clínic, Barcelona, Spain and Reinier de Graaf Groep, Voorburg, the Netherlands (34 biopsy specimens), were used in this study (29). Patient data for all the samples was anonymised. All biopsies were fixed in buffered formalin and embedded in paraffin.

Selection of cervical biopsy sections and whole tissue PCR

Twenty five serial sections of 5 μm were cut from each formalin-fixed paraffin-embedded histology specimen. Sections 8 and 15 were taken for whole tissue section-PCR analysis and analysed separately, and sections 12, 13, and 14 were mounted on Laser Capture Microscopy membrane slides (Zeiss, Cambridge, UK). Slide 11 was used for E4 Mab and MCM2 immunohistochemistry followed by $p16^{INK4a}$ and H&E staining. Slide 10 was used for pan E4 monoclonal and MCM2 immunohistochemistry followed by H&E staining. Extensive measures were implemented during sectioning to prevent cross-contamination.

Immunohistochemical staining

E4 detection was carried out using both TVG 405 (a HPV16,18, 31, 35 and 45-specific E4 Fab (28)) and a newly developed pan-specific E4 monoclonal antibody (FH1.1) reactive against the high-risk HPV types 16, 18, 31, 33, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 67 and 70 (manuscript in preparation), depending on the HPV types present. Although TVG 405 was used for the majority of the stains, both antibodies produce equivalent staining patterns. MCM2 and $p16^{INK4a}$ immunohistochemistry was performed on all cervical biopsy sections according to standard procedures. For epitope retrieval, slides were incubated in solution D pH 9.0 (Dako, Glostrup, Denmark) for 10 min at room temperature prior to autoclaving for 2 min at 121°C. The HPV anti-E4 Fab TVG 405 was directly conjugated to Alexa 488 and was diluted 150-fold prior to use. The concentrated monoclonal E4 antibody (FH1.1) was diluted 100-fold and detected using an anti-mouse Alexa 488 conjugated antibody (dilution 1:150, Vector, Peterborough, UK). E4 staining was combined with MCM2 using a primary rabbit polyclonal antibody to MCM2 (Abcam, Cambridge, UK). MCM antibody detection was carried out using an anti-rabbit Alexa 594 conjugated antibody (dilution 1:150, Vector, Peterborough, UK). Nuclear counterstain was performed with 4'-6-diamidino-2-phenylindole (DAPI, 1 mg/ml 200- to 500-fold diluted, Sigma, St-Louis, MO, USA) before mounting in Citifluor medium (Agar Scientific, Essex, UK) for fluorescence scanning and digital imaging on a Pannoramic Slide Scanner (3D Histotech, Hungary).

Sections which had been stained with the Fab fragment to HPV E4 and a rabbit polyclonal antibody to MCM2 were then stained using a mouse primary antibody to $p16^{INK4a}$ clone JC8 diluted 1:20 (Santa Cruz Biotechnology Inc., Santa Crua, CA, USA) which was detected using an anti-mouse biotinylated antibody (dilution 1:150, Vector, Peterborough, UK) followed by development using an ABC kit (Vector) and red AEC-reagent (Sigma, St Louis, USA). After scanning and digital imaging of the $p16^{INK4a}$ stain (using the Pannoramic Slide Scanner (3D Histotech, Hungary) as described above), slides were H&E stained with Carazzi Haematoxylin (x2) prior to rescanning under bright field illumination.

For $p16^{INK4a}$, strong nuclear and cytoplasmic staining was considered as positivity. For MCM2, strong nuclear staining was considered as positivity. The $p16^{INK4a}$ and MCM2 results were reported in a semi-quantitative fashion as negative or focal staining, staining up to 1/3 of the epithelium, staining up to 2/3 of the epithelium or staining above 2/3 of the epithelium. Scoring was based on the highest

category within an annotated area. The immunohistochemistry staining and reading was carried out by three individuals (HG, YS and RvB).

HPV DNA detection and laser capture micro-dissection

HPV typing was carried out on all cases using a PCR/Line Probe Assay (SPF_{10}-PCR-DEIA-$LiPA_{25}$, Version 1, Labo Biomedical Products BV, Rijswijk, The Netherlands) as described previously (30, 31) in order to assign HPV type(s) to a lesion. This is referred to in the text as whole tissue section whole tissue section-PCR. In CIN1 areas that were E4-negative, an adjacent section was analysed by laser capture microdissection-PCR in order to establish whether HPV DNA was detectable in the lesional area (32). Fluorescence *in situ* hybridisation (FISH) was carried out on a subset of these E4-negative lesions (including all the total-agreement CIN1 cases)to confirm that E4-negativity correlated with an absence of genome amplification (12, 33). Sequence analysis of the SPF_{10} regions was done for three of the laser capture microdissection regions that remained untypeable by PCR/Line Probe Assay ($LiPA_{25}$) as described previously (34). We excluded cases with HPV types that were undetectable by the antibodies used in this study.

Pathological diagnosis and grading

Initial diagnosis and the grading of discrete areas with different pathologies was carried out on the H&E stained section according to standard criteria by an expert pathologist at UCL (London, UK). Annotated regions were classified as 'non-CIN', when HPV-associated pathology was absent, but where other histological changes, including immature metaplasia and squamous hyperplasia were seen. When putative HPV-associated changes were noticed, they were classified either as CIN1, CIN2, CIN3 or a combination thereof (e.g. CIN1/2, CIN2/3). Grading was then independently re-assessed by two expert pathologists using the same tissue section. Discrete areas within the tissue section were examined in this study, because we wanted to correlate the precise relationship between markers of infection and associated pathology, although we appreciate that in routine practice, the highest grade of disease on the tissue section must be used to determine treatment options. If all pathologists agreed on the diagnosis of disease severity, it was considered as total agreement. If 2/3 pathologists agreed, it was considered consensus agreement. Furthermore, all areas were scored for the presence of koilocytosis (i.e. superficial cells with perinuclear atypia and cytoplasmic cavitation (see p209 of reference (10)), or the presence of vacuolation, when not all the koilocyte characteristics were apparent. The pathologists were blinded to the HPV status and immunohistochemistry results.

Results

Development of an Overlay Approach for Pathology/Biomarker Correlation

In order to correlate the expression patterns of multiple biomarkers with disease pathology, we first developed an 'image-overlay' approach. Using this methodology, the distribution of key molecular markers were imaged and captured separately following immunofluorescence staining or immunohistochemistry (Fig.1A). Our marker panel included $p16^{INK4a}$, which is an established marker of deregulated high-risk HPV gene expression, and MCM, which identifies cells 'in cycle'. Both are considered as surrogates of E6/E7 expression when present in HPV-associated cervical neoplasia. An important aspect of this study is the use of such markers in combination with the E4 biomarker, which represents a separate category of marker that marks the onset of productive infection. The staining and image-capture regime is shown in Fig.1A. At the end of this procedure, the tissue section was cleared of visible biomarker staining as part of the H&E staining regime in order to allow precise visualisation of the pathology features used in conventional diagnosis. To accurately correlate molecular biomarker patterns with disease pathology, all of the H&E-stained slides were subsequently subjected to pathology review, and discrete lesional areas exhibiting distinct pathologies were annotated as shown in Fig.1A. Individual biomarkers were then digitally superimposed onto the annotated H&E image in order to establish the molecular pathology of disease. The images shown in Fig.1B shows the distribution of E4 (coloured green) in regions of low-grade disease, the distribution of MCM (coloured red), which marks cells in cycle, and the distribution of $p16^{INK4a}$ (coloured brown) which is most prominent in the regions of HG-CIN. As the molecular biomarker stains and H&E are all carried out on the same tissue section, individual markers can be overlayed in different combinations to fully understand how HPV activity and cellular gene expression relates to conventional pathology. The complementarity of the HPV E4 biomarker (green), and surrogate markers of E6/E7 expression ($p16^{INK4a}$ (brown) and MCM (red)) is shown in the low-power images in Fig.1B, as are the distinct biomarker distributions seen in low and high-grade disease areas. During the course of the study, 530 areas with discrete cervical pathologies (as shown in Fig.1A) were identified and analysed using the biomarker-overlay approach.

Pathology Agreement

Of the 530 annotated regions examined in this study, all pathologists agreed on the diagnosis in 146 instances, while in 282 areas there was a consensus by 2/3 pathologists. 102 lesional areas were the subject of total disagreement, with pathologists disagreeing as to the precise CIN grade (e.g. CIN1 or CIN2, or CIN2 or CIN3) or whether diagnosis should be CIN1 or non-CIN (i.e. metaplasia, inflammation etc). In total, 276 lesional areas (52%) were classified as CIN1 by at least one pathologist. Only in 25/276 (5%) areas was there total agreement on the CIN1 diagnosis however, with an additional 8 areas receiving a range of diagnosis comprising CIN1/2. 284 lesional areas were classified as CIN2 by at least one pathologist, but in only 12/284 (4%) did all three pathologists agree on the diagnosis. In 6 of these, one of the pathologists used either a CIN2/CIN3 (4 areas) or a CIN1/CIN2 (2 areas) classification. A total of 311 individual areas were classified as CIN3 by at least one pathologist, but in

only 68 areas was there total agreement on the CIN3 diagnosis amongst all three pathologists. Of these, 23 received a diagnosis of CIN2/CIN3 by one of the pathologists. The areas with total agreement were subsequently used to establish the specific criteria that might allow the use of molecular markers as surrogates of CIN grading. HPV-typing analysis revealed that the majority of cases (90%) contained only a single HPV type by whole tissue section-PCR, and encompassed single-infections by HPV types 16, 18, 31, 33, 35, 39, 51, 52, 53, 56, 58, 59 and 66. No additional types were detected in cases with multiple infections, which contained between two and five HPV types, all of which could be detectable using the combination of E4 antibodies outlined in this study.

Total Agreement CIN1 Generally has a Well-Defined Molecular Pathology Pattern

HPV-induced changes are considered pathognomonic of low-grade squamous-intraepithelial lesions, and of these the most significant is nuclear atypia characterized by variation in nuclear size, hyperchromasia, and irregularity or wrinkling of the nuclear membrane. Additional characteristics of HPV infection include acanthosis and koilocytosis. The **33** CIN1 regions examined here had recognisable characteristics, with koilocytosis and clear evidence of cytoplasmic vacuolation in **32** of the areas (see Fig.2A(i) and 2A(ii)). The majority of these 'total-agreement' CIN1 lesions (i.e. **27** out of **33** (82%)), showed expression of the E4 biomarker (Fig.2A(i) and (ii)), although in **3** areas there was tissue damage at the epithelial surface that occurred during the biopsy procedure or during histologic processing). In contrast to the unambiguous CIN1 shown in Fig.2A(i) and 2A(ii), the lesion shown in Fig.2A(iii) was diagnosed as CIN1/basal cell hyperplasia by one of the three pathologist reviewers. HPV52 was detected within the tissue by laser capture microdissection, but this was at the limits of detection compared to the productive CIN1 cases described above, requiring DNA sequencing to confirm the infecting type as HPV52. Only five other CIN1 lesions were found to be devoid of E4, with one of these being reclassified as higher grade (CIN3) following visualisation of mitotic figures in the upper epithelial third on pathology review, as well as nuclear atypia in the basal layer despite a low-grade pathology in the mid-epithelial layers. The presence of $p16^{INK4a}$ and MCM2 expression in the upper two-thirds of the epithelium, and the absence of E4 supported this revised grading. As reported previously, the presence of E4 correlated closely with the onset of viral genome amplification as visualised by fluorescent *in situ* hybridisation (FISH) (6, 12, 33). Viral genome amplification was not seen in any of the 6 E4-negative CIN1 described here. Two of the CIN1 E4-negative areas were tangentially sectioned and thus difficult to interpret, while one showed clear inflammation. Interestingly, the remaining E4-negative CIN1 was amongst those classified as CIN1/2 by one of the pathologists (shown in Fig.2B (iii)) and differed markedly from the above examples, in containing clearly identifiable koilocyte-like cells that are suggestive of productive HPV infection (Fig.2B (iii)), but only limited MCM2 and $p16^{INK4a}$ expression which was restricted to the basal and suprabasal layers (MCM2 staining shown in Fig.2B (iii)). In the koilocyte-like cells (arrowed in Fig.2B(iii)) there was no evidence of cell-cycle entry, viral genome amplification (as determined by fluorescence *in situ*

hybridisation; data not shown) or the presence of E4 (see Fig.2B (iii)), distinguishing this lesion from the majority of productive CIN1, which were E4-positive (see Fig.2B(i) and 2B(ii)), and which could be confirmed as productive HPV infections. Although HPV type 33 could be detected in this lesion by laser capture microdissection, it is worth noting that HPV can sometimes also be detected in apparently normal cervical tissue as an asymptomatic or latent infection (32, 35), and that HPV positivity in CIN1 can be between 50-70% (36). In the absence of a definitive pathognomic biomarker pattern or other evidence of viral gene expression, viral causality remains uncertain.

HPV_E4 Accumulation in CIN1 Occurs during Vacuolation and Koilocyte Development

To correlate virus-associated pathology with biomarker presence more precisely, the E4/MCM2 biomarker pair was overlayed onto H&E stained images in all of the low-grade lesional areas described above. This analysis revealed a reproducible relationship between marker presence and viral pathology (Fig.2B). Cells with well defined nuclear margins, before the first appearance of E4 stained strongly with MCM2 (cells labelled '1' in Fig.2B (i) and (ii)). The appearance of E4 coincided with the first signs of cellular vacuolisation in HPV associated CIN, with vacuolisation becoming more apparent in these E4-positive cells upon further differentiation (cells labelled '2' in Fig.2B (i) and (ii)). E4 was typically more apparent in koilocytic cells however (labelled '3' in Fig.2B), which we suspect arise from these first E4-positive vacuolated cells. The koilocytotes and the cells immediately above them accumulate higher levels of the E4 biomarker and lose their well-defined nuclear margins in the H&E stain, with lower and more diffuse MCM2 expression (cells labelled '4' in Fig.2B (i) and (ii)). The larger nuclear size often seen in these cells is compatible with what might be expected following cell cycle arrest in G2, which is thought to be required for genome amplification and E4 accumulation (37). Interestingly, a subset of CIN1 showed a noticeable elevation, rather than decline in MCM2 expression as the E4 biomarker appeared, with little basal cell expression (data not shown), similar to what is occasionally seen in HPV16 organotypic raft cultures (14). This is reminiscent of benign productive lesions (i.e. warts) caused by low-risk papillomaviruses. When taken alongside the variation in $p16^{INK4a}$ staining patterns seen in CIN1, the data suggests some variation in viral gene expression in CIN1.

These distinctive biomarker correlations were not apparent in the atypical CIN1/CIN2 (see koilocyte-like cells marked by 'K' in Fig.2B (iii)). Interestingly, the non-productive infection shown in Fig.2A(iii) shows (upon close inspection) only the very first signs of koilocyte formation, in agreement with its distinct molecular pathology. While MCM2 and $p16^{INK4a}$ staining have very similar distributions here, in productive HPV infections such as that shown in Fig.2A(ii), robust MCM2 expression persists into the upper epithelial layers beyond those where $p16^{INK4a}$ is found. Given that MCM is a marker of E6/E7-mediated cell cycle entry in such lesions, and that cell cycle entry is required for genome amplification, this pattern fits well with our understanding of the HPV life cycle

organisation. By contrast, p16^{INK4a} is considered to be a marker of deregulated E6/E7 gene expression, which may account for its more limited distribution.

Diagnostic Relevance of the HPV_E4 biomarker in p16^{INK4a}/MCM-positive CIN3

In **63** cases where there was total agreement on the CIN3 diagnosis (out of **68**), no E4 could be detected by any of the E4 antibodies used in this study. P16^{INK4a} and MCM expression extended from 2/3 of the epithelium up to the surface with nuclear crowding, pleomorphism and loss of polarity clearly visible (Fig. 3A(i) and (ii)). Although absence of the E4 biomarker was by far the majority pattern in the CIN3 group, two of the E4-negative cases were also p16^{INK4a}-negative, even though MCM2 levels extended to above 2/3 of the epithelium. Amongst the five CIN3 areas where E4 was detected, the majority (i.e. **4** out of **5**) contained only single E4-positive cells, or small clusters of such cells in isolated regions of differentiation close to the epithelial surface. These isolated cells or cell-groups generally showed some evidence of koilocytosis and viral cytopathic effect, similar to what is found more extensively in low-grade disease, (cells marked by arrows 1 and 2 in Fig.3C (panel B)), although it was notable that not all vacuolated cells were E4-positive (cells marked by arrow 'v' in Fig.3C (panel A)). Only in one CIN3 was E4 expression extensive, and in this case the lesional area exhibited some degree of tangential sectioning with mixed or heterogeneous pathology. MCM expression extended throughout the epithelium and facilitated the visualisation of nuclear atypia, nuclear crowding (as seen in Fig.3B(ii)), pleomorphism and loss of normal cell polarity. P16^{INK4a} expression was typically found to extend from between two-thirds of the epithelium up to the surface and was broadly complementary to the pattern of staining seen with E4 (Fig.3A and B).

Division of the CIN2 Group according to HPV_E4 and p16^{INK4a}/MCM2 Biomarker Patterns

Studies of the pathological diagnosis of CIN2 have shown that it is only poorly reproducible (38). The E4 biomarker was apparent in **7/12** (58%) CIN2 areas with total agreement, using both E4 antibodies described in this study. Staining for E4 was however often patchy, being interspersed by regions where p16^{INK4a} and MCM clustered close to the epithelial surface (see Fig.4A). This was particularly evident when the areas designated as CIN2 were large. **4** of the E4-positive areas showed prominent koilocytosis, with E4 being expressed in a subset of koilocytes upon epithelial differentiation (Fig.4A and 5A). P16^{INK4a} levels extended up to and above 2/3 of the epithelium. The MCM marker typically extended into the upper two thirds of the epithelium.

Interestingly, **5** of the agreed CIN2 areas were E4-negative and showed only limited signs of vacuolation and an absence of koilocytosis. One area was found to have a mitotic figure higher than 2/3 of the epithelium. Two of the E4-negative areas were in the endocervical glands, (i.e. a site that we suspect does not support the 'normal' productive life-cycle). All 4 E4-negative areas showed high p16^{INK4a} and MCM levels up to or above 2/3 of the epithelium, suggesting that at the molecular level they may in fact be CIN3-like.

Re-classification of Ambiguous Pathology using the Dual Marker Approach

To clearly assess how molecular patterns relate to pathology, we examined all 146 total-agreement pathologies according to both E4 status and distribution of $p16^{INK4a}$ (Fig 6A). As each pathology diagnosis was the consensus from three pathologists, the total number of individual diagnostic opinions in this group was 438 (i.e. 3x146). Lesions classified as CIN1 and which express E4 (Fig. 6A; 'CIN1' column (green) left-most graph)) are supporting the productive phase of the virus life cycle, and typically exhibit limited $p16^{INK4a}$ expression. Almost all E4-positive CIN1 were however positive for $p16^{INK4a}$ to some extent. By contrast, the majority of CIN3 were E4-negative or very limited expression (Fig. 6A 'CIN3' column (red) right-most graph) and showed a much more extensive $p16^{INK4a}$ distribution (i.e. extending through 2/3 of the epithelium), in agreement with their status as transforming rather than productive infections. Although numbers were limited, the total-agreement E4-negative CIN2 generally showed a more extensive $p16^{INK4a}$ distribution than was seen in the E4-positive group. Although clinical practice suggests that CIN2 and CIN3 could be considered together as HSIL, it appears that E4-positive 'CIN1-like' lesions are more common in the CIN2 group than CIN3, which reflects the heterogeneity of CIN2. The E4 and $p16^{INK4a}$ biomarkers were absent in nearly all cases where there was total agreement that the pathology was not HPV-associated and had another cause (e.g. metaplasia, inflammation etc.). It would therefore appear that biomarker patterns reflect the underlying pathology

As part of our aim was to establish how molecular biomarkers might eventually improve diagnostic accuracy, we next went on to prepare a second chart, but this time included the 384 cases where there was partial or total pathology-disagreement amongst the panel of pathologists (Fig.6B). In this case, the total number of individual pathologies examined was 530 (384 + 146) and the total number of individual opinions plotted in Fig.6B was 1590 (3 x 530). In areas considered to be low grade by at least one pathologist, E4 expression was generally accompanied by low levels of $p16^{INK4a}$, similar to the total-agreement group. Interestingly, a significant fraction of the E4-negative areas were found to be $p16^{INK4a}$-negative, and may not have a causal association with HPV. Non-viral mimics of CIN1 could include metaplasia or inflammation. The remaining E4-negative 'CIN1' generally showed more extensive $p16^{INK4a}$ than the E4-positive group.

The majority of areas that received a CIN2-classification were E4-negative, with levels of $p16^{INK4a}$ expression higher in this group than in the E4-positive group. E4-positive 'CIN2' areas were equally divided according to whether E4 expression was extensive, or restricted to focal regions similar to those seen in the total-agreement CIN3 category (Fig.6A). The E4-negative CIN2 that express high levels of $p16^{INK4a}$ are a potentially important group that may warrant follow-up . The E4-positive CIN3 typically expressed only very limited/focal levels of E4, and generally showed extensive $p16^{INK4a}$ expression supporting the high-grade diagnosis.

A fraction of the areas classified as CIN2 or CIN3 by at least one pathologist lacked expression of both the $p16^{INK4a}$ and E4 biomarker, and it may be that these pathologies are not HPV-associated. These areas only became evident when the consensus and total disagreements were combined, hence they can be considered as cases that are difficult to judge. Further analysis of these biomarker-negatives where 2 of the pathologists agreed, identified 2 CIN2 areas judged to be metaplasia or CIN3 by the third pathologist, as well as 2 CIN3 areas considered to be metaplasia or CIN2.

Amongst the areas where there was total disagreement between the panel of pathologists, 43 areas were $p16^{INK4a}$-negative, with two of these being E4-positive, although the extent of E4 expression was quite restricted. The majority (41) were both E4 and $p16^{INK4a}$ negative. One of our pathologists graded 32 of these as non-CIN, in agreement with the E4/$p16^{INK4a}$ patterns, whereas another pathologist identified 24 CIN1 amongst this group of 43 areas. In 15 out of these 24 pathology areas, a different grading (CIN1, CIN3 and non-CIN (metaplasia/inflamation)) was given by each of the three pathologists. In another 7 E4/ $p16^{INK4a}$-negative areas, all three pathologists reported the presence of CIN, but each gave a different diagnosis (CIN1, CIN2, CIN3). Such inter-observer variability highlights the problem of reliably discriminating true HPV-associated changes from other similar pathologies such as metaplasia or inflammation, and supports our hypothesis that the staining of molecular markers in addition to conventional pathology should help to improve overall diagnostic accuracy. More specifically, our data suggest that the molecular patterns seen in the unequivocal cases (Fig. 6A), and which are likely to be true indicators of CIN pathology, could be used to refine the grading of disease in case where pathology-diagnosis was equivocal (Fig. 6B).

Discussion

The viral aetiology of cervical neoplasia has indicated an important role for HPV testing during cervical screening (1, 39). HPV-negativity means that screening intervals can be extended and that there is a low risk of cancer progression (40-43), whereas the detection of HPV DNA highlights a need for triage to establish the presence of active disease and to confirm its severity. The situation is complicated however by the fact that transient HPV infections that are not likely to progress to cancer are common in young women, and that the presence of HPV in a cervical cytology or whole tissue biopsy sample does not necessarily mean the presence of a precancerous lesion, and may in some instances represent only sexual deposition and not even infection (25). As a result, there is interest in the use of biomarker approaches as an adjunct to pathology in order to confirm infection and facilitate disease stratification. Of these, $p16^{INK4a}$ has been most well characterised, and its presence in cervical neoplasia is generally taken as indicating deregulated HPV 'oncogene expression' (25, 27, 44). Although the discrimination between high and low-grade cervical disease eventually determines treatment options, robust markers of low-grade cervical pathology that can be used alongside $p16^{INK4a}$ are not well developed. In this study therefore, we considered the contribution of the abundant HPV-

encoded E4 protein in discriminating between high and low grade disease and in identifying true viral infections from regions with similar pathology. An additional class of HPV E6/E7-induced cellular protein (MCM), similar to Ki-67, was also used (45-47). We conclude from this that the detection of E4 with $p16^{INK4a}$ provides additional molecular detail regarding the extent of HPV gene and life-cycle deregulation, and that such a dual marker approach can help confirm viral aetiology and disease status in both high and low grade disease.

To achieve the above goals, we used a research/evaluation-methodology that allowed us to superimpose the distribution of each marker onto a standard H & E stained pathology image, allowing us to correlate the features that are currently used for pathology diagnosis with the presence of molecular markers. During this process, particular attention was paid to individual discrete pathologies found in each lesion, in order to establish a reference of how each biomarker relates to the underlying pathology of the tissue. From these focused observations, a common pattern of $p16^{INK4a}$ loss and E4 appearance was apparent as the HPV infected cell underwent the process of epithelial differentiation. In general, this correlated with the cytopathic effects that mark the onset of productive infection during the process of koilocytosis and initiation of cell vacuolisation (8). Interestingly, the appearance of koilocytes was frequently characterised by the presence of abundant E4 in the upper epithelial layers, and a restriction of $p16^{INK4a}$ to the lower epithelial layers within a lesion. This pattern of gene expression was more reliably followed in low-grade than high-grade disease. In general, difficulties in distinguishing normal squamous epithelium from low-grade squamous intraepithelial lesions (as reported by McCluggage & Bharucha (48)) reflect the problem of discriminating between superficial vacuolated cells and true koilocytes. The observation that the onset of E4 expression coincides with the appearance of vacuolated cells that subsequently go on to form koilocytes as they differentiate may thus have diagnostic significance. In addition to this correlation between E4 expression and pathology, we noticed a clear and consistent difference in the distribution of MCM, which is considered to be a surrogate marker of E6/E7 expression, and $p16^{INK4a}$ which is thought to be an indicator of E6/E7 deregulation (Fig. 7) (25, 44, 49). Again, this was more obvious in low-grade pathologies than in CIN3, where the distribution of these two E6/E7 surrogates was in most cases very similar.

The underlying basis for the molecular biomarker patterns observed above can to a large extent be drawn from our understanding of HPV-driven neoplastic progression and normal HPV life-cycle deregulation. Thus high levels of E6 and E7 can stimulate the appearance of $p16^{INK4a}$ throughout the epithelium by interfering with the normal functions of the retinoblastoma protein (pRb) and the KDM6 histone demethylase (16, 26). Interestingly, recent studies have suggested that changes in viral gene expression may underlie such phenotypic variation in CIN1 and CIN2 (14), but that host genetic changes may be additionally required for the development of CIN3 (15). The MCM marker, as shown previously for Ki-67, has a distinct pattern from $p16^{INK4a}$ in low-grade disease, presumably because its induction marks the cell cycle entry that is ordinarily stimulated by HPV to allow genome

amplification (see Fig. 7) (12, 24, 49). In this case, cell cycle entry associated with productive infection can be distinguished from that seen in proliferating cells by the detection of E4 and MCM together. When used singly however, the presence of markers such as MCM and Ki-67 can sometimes be difficult to interpret, as these proteins may also be up-regulated during inflammation and metaplasia. The E4 biomarker has comparable limitations when used alone, but facilitates the visualisation of most, if not all HPV-associated disease areas when used in combination. The molecular regulation that leads to E4 accumulation has previously been investigated, and involves deposition of E4 in the infected cell as amyloid fibres, at around the time that genome amplification begins during productive infection (50-52). Precisely why productive infection and E4 expression are retarded in high-grade disease is only poorly understood however, although our recent work suggests a link to cell cycle duration, and the accumulation of E4 only as the G2 phase of cell cycle lengthens (37, 53, 54). These preliminary results offer some initial insight as to why E4 and $p16^{INK4a}$ produce largely complementary patterns of staining as neoplastic severity increases.

Although the pathology overlay analysis clearly shows a relationship between E4 abundance and regions of low-grade pathology, the main purposes of using such markers is to improve diagnostic accuracy and to determine the most appropriate regimen for the treatment of disease. For cervical neoplasia, treatment generally follows the diagnosis of CIN2 or higher, and typically involves surgical removal of the infected tissue and disease margins (e.g. by cone biopsy). While this treatment is generally effective in preventing neoplastic progression, such intervention has been linked to an increased risk of pre-term delivery at childbirth (55, 56). Current thinking also suggests that there is considerable heterogeneity amongst CIN2, and that there is a significant problem of overtreatment within this group, partly accentuated by the difficulty of reliable diagnosis (57). From the work described here, it appears that lesions designated as CIN2 by one or more pathologists can be divided into two groups on the basis of whether significant E4 expression is retained (i.e. they are CIN1-like), or whether E4 expression is lost or is restricted to focal areas close to the epithelial surface (i.e. they are CIN3-like). In general, the absence or focal expression of the E4 biomarker correlated with established pathology features of high-grade disease, including high cell density that extended towards the epithelial surface, the presence of cells with a high nuclear to cytoplasmic ratio and a more extensive $p16^{INK4a}$ expression throughout the epithelium. When our molecular understanding of the HPV life-cycle is considered alongside these pathology correlations, it would seem that there is a persuasive rationale for the use of a dual molecular marker approach alongside conventional pathology to establish disease severity and the associated risk of cancer.

For routine diagnostics, an additional advantage of the molecular biomarkers described here lies in their ease of use. Both $p16^{INK4a}$ and MCM are already available commercially for diagnostic purposes (49, 58). Interestingly, several studies have advocated the use of HPV DNA *in situ* hybridisation alongside $p16^{INK4a}$ to confirm diagnosis (48, 59), a rationale which given the correlation between E4 expression and viral genome amplification, has some similarities to the E4/$p16^{INK4a}$ dual

biomarker approach described here. Although E4 detection reagents remain to be commercially developed, it is the most abundantly expressed HPV protein during the viral life-cycle, and can be detected in lesions without the need for signal amplification systems (28, 33, 53). The recent development of a new pan-specific HPV E4 antibody that is broadly cross-reactive against high-risk E4 proteins should facilitate their further evaluation. It is generally realised that distinguishing high-grade HPV-induced disease from other events such as immature metaplasia, atypical immature metaplasia (AIM), reactive/reparative changes or atrophy from precancer is sometimes impossible on the basis of pathology alone. The use of $p16^{INK4a}$ immunohistochemistry and detection of high-risk HPV can assist in determining viral aetiology and has shown that a proportion of these lesions are morphologically difficult CIN3 (10). The situation with low-grade cervical disease is also often very difficult. In the ASCUS-LSIL Triage Study (ALTS) for instance (60), only 43% of biopsies initially classified as low-grade HPV-associated lesions were classified as low-grade disease on review, with most discrepancies being explained by an inability of the pathologist panel to reliably discriminate between HPV-associated disease, reactive squamous proliferations, and other situations where a HPV-like low grade phenotype may be apparent (e.g. candidiasis, trichomoniasis). The vacuolated cells shown in Fig.2B (iii) that lack markers of productive HPV infection may in fact be pseudokoilocytes. Such issues contribute to the generally low inter-observer agreement in the detection of low-grade HPV-associated disease, which contrasts sharply with the excellent agreement for invasive cervical lesions and even high-grade disease, which are excellent and moderately good respectively. As shown here, E4 expression is typically associated with discrete pathology features of low-grade HPV-associated disease, and it has been suggested that careful attention to cytological and histological changes can be used to discriminate between viral and non-viral pathologies (10). In reality however, such relatively subtle changes are not easy to establish under routine screening conditions, and as has been shown for all of the markers described here, their use not only marks, but also draws attention to the region of abnormality. In our study, this was apparent in one region of CIN3 that was missed by all three pathologists, but was identified as being both $p16^{INK4a}$- and MCM2-positive throughout the epithelium following biomarker analysis with confirmation of CIN3 status upon pathology review. In this instance the CIN3 region comprised a small glandular lesion within a large cone biopsy.

When taken together, it appears that conventional pathology combined with the use of the E4 and $p16^{INK4a}$ biomarkers has advantages over pathology alone, or the use of pathology in combination with $p16^{INK4a}$ staining for the stratification of HPV-associated cervical neoplasia, which already provides information regarding clinical outcome (61). In particular, the use of E4 facilitates the identification of low-grade viral disease where the protein is typically abundant, and distinguishes such cases from non-viral pathologies that may need different management strategies. Our data also suggests that the various and distinct expression patterns of both $p16^{INK4a}$ and E4 in CIN2, may allow categorisation of this heterogeneous group into a CIN1-like productive infection or a CIN3-like transforming infection group according to the extent of HPV deregulation and life-cycle completion.

We cannot be sure at present whether this would help in determining patient management strategies, but suspect from our data that this might be the case. The abundance of E4 and its ease of detection, should facilitate a potential role during routine diagnosis and triage following cervical screening.

Acknowledgements

JD, HG, YS and ZW were funded by the UK Medical Research Council. RvB was funded by the Stichting Pathologie Ontwikkeling en Onderzoek (SPOO) Foundation, The Netherlands.

List of abbreviations

CIN: cervical intra-epithelial neoplasia; DAPI: 4'-6-diamidino-2-phenylindole; H&E: hematoxylin and eosin; MCM: minichromosome maintenance protein; PCR: polymerase chain reaction.

Figure Legends

Figure 1. **Overlay of biomarker patterns onto annotated Hemotoxylin and Eosin Pathology.**

A. Individual tissue sections were subject to immunofluorescence staining to detect the biomarkers E4 using a TVG405-Alexa 488 conjugate (green), MCM using an anti-mouse Alexa 594 secondary antibody (red) and cellular DNA using 4',6-Diamidino-2-Phenylindole (DAPI; blue). Individual stains were recorded digitally at high resolution (lower panels) before the section was processed for $p16^{INK4a}$ staining and development in 3_Amino_9_ethylcarbazole (AEC; brown). The AEC image was digitally captured, before the tissue section was cleared of the AEC substrate and stained with Carazzis (X2) Hemotoxylin and Eosin. All H&E stained images were then examined independently by three pathologists, and regions with discrete pathology phenotypes recorded. The primary categories of HPV-associated pathology comprised CIN1, CIN2 and CIN3, with the general term 'non-CIN' being used to encompass a variety of non HPV-associated pathologies such as inflammation and metaplasia. A 'normal' classification was given when there was no histological abnormality.

B. The 3-colour immunofluorescence stain is shown on the left. Single channel coloured images (collected as described in A) were extracted from the immunofluorescence image or from the $p16^{INK4a}$ AEC stain and were superimposed onto the H&E images (under the heading 'Pathology overlay'). The simple dual marker molecular pathologies (i.e. not overlayed onto the H&E image) are shown on the right to reveal the relative distributions of E4/MCM and E4/ $p16^{INK4a}$ (under the heading 'Dual marker molecular pathology). The image shown is typical of those used to prepare the more detailed images of neoplasias used in subsequent figures.

Figure 2. **Distribution of the HPV_E4, $p16^{INK4a}$ and MCM2 biomarkers in lesions unambiguously classified as CIN1.**

A. Biomarker patterns typically associated with CIN1. The E4/MCM (green/red) biomarker patterns are shown in the immunofluorescence image to the left of the figure. To facilitate comparison with lesional pathology, the E4/MCM (green/red) and $p16^{INK4a}$ (brown) biomarkers are overlayed onto (and shown alongside) the basic H & E stain in the central part of the figure. The E4/MCM (green/red) and E4/ $p16^{INK4a}$ (green/brown) biomarker overlays are shown separately to the right of the figure so that their relative distributions can be observed. The Biomarker patterns seen in HPV associated CIN1 are described below.

i) Most CIN1 are productive infections, with MCM (red) reaching and extending into the E4 expressing layers (green). In this lesion, MCM and $p16^{INK4a}$ have broadly similar distributions.

ii) Although $p16^{INK4a}$ (brown) and MCM (red) expression can have different distributions, MCM again extends into the E4-positive layers (green). In this lesion, MCM extends higher into the epithelial layers than $p16^{INK4a}$.

iii) A small number of consensus CIN1 (6/33) failed to show E4 (green) biomarker expression. $p16^{INK4a}$ (brown) expression extends into the upper epithelial layers and is more extensive than MCM (red).

B. Correlation of Biomarker patterns with Pathology. The E4/MCM (green/red) biomarker distribution in consensus CIN1 are shown as immunofluorescence images to the left of the figure, and overlayed onto H & E pathology towards the centre. The higher magnification shown on the right allows correlation of specific pathology characteristics with the accumulation of E4 and the decline of MCM.

i) In the majority of CIN1, the E4 (green) biomarker accumulates during the process of koilocyte formation. Arrows labelled 1 to 4 show the progressive vacuolation that accompanies E4 accumulation (beginning at arrow 2 and highest at arrow 4), and the loss of the MCM biomarker (lowest at arrow 4).

ii) Despite some differences in cell morphology and tissue architecture amongst CIN1, the distribution of biomarkers described above (i) is conserved.

iii) In one (out of 33) consensus CIN1, koilocyte-like cells we present but were devoid of markers of productive infection, including the E4 (green) biomarker and associated MCM (red) staining.

Figure 3. **Relevance of the HPV_E4, $p16^{INK4a}$ and MCM biomarkers in consensus CIN3 lesions**

A. Biomarker Patterns Typically Associated with CIN3. The E4/MCM (green/red) and $p16^{INK4a}$ (brown) biomarker images are formatted as outlined in Fig.2A. The biomarker patterns seen in HPV-associated CIN3 are described below.

i) In the majority of CIN3, the E4 (green) biomarker is absent, and both the MCM (red) and $p16^{INK4a}$ (brown) biomarkers extend uniformly through the full thickness of the epithelium.

ii) Despite differences in lesion size and morphology, the biomarker patterns described above (i) were broadly similar in other CIN3.

B. Expression of the E4 biomarker is an occasional occurrence in CIN3.

i) In a small number of CIN3 (4/68), focal areas of E4 (green) were apparent. Both $p16^{INK4a}$ (brown) and MCM (red) extended through the full thickness of the lesion with evidence of nuclear crowding as seen in Fig 3A. Panel A is shown enlarged in Fig. 3C.

ii) In one case (out of 68), the E4 (green) biomarker was extensive, but was confined to regions showing low-grade pathology that lacked strong $p16^{INK4a}$ (brown) biomarker staining. Panel B is shown enlarged in Fig. 3B.

C. Pathology associated with E4 expression in CIN3. The three images to the left show an enlargement of Fig. 3B(i) panel A. Although vacuolated cells are sometimes apparent in CIN3 (arrows marked 'v'), they are not necessarily associated with expression of the E4 biomarker (Fig. 2B(i)). The pathology associated with E4 expression in CIN3 is more clearly shown in the images to the right, which is an enlargement of panel B from Fig.3B(ii). This region of low-grade pathology was contained within the CIN3 area. The pattern of vacuolation and koilocyte formation is comparable to that seen in CIN1 (Fig. 2B), with E4 expression beginning at the first sign of vacuolation (arrows labelled 1) and becoming prominent as the MCM biomarker is lost (arrows labelled 2).

Figure 4. **Expression of HPV_E4 along with $p16^{INK4a}$ and MCM suggests two categories within the consensus CIN2 group.**

A. HPV_E4-positivity amongst consensus CIN2. The E4/MCM (green/red) and $p16^{INK4a}$ (brown) biomarker images are formatted as outlined in Fig.2A. The biomarker patterns seen in E4-positive HPV-associated CIN2 are described below.

i) High cell density and the expression of MCM precedes the accumulation of E4 in a subset of cells showing evidence of vacuolation. $P16^{INK4a}$ and MCM expression extend throughout the full thickness of the epithelium. Panel A is enlarged in Fig.5.

ii) Despite differences in epithelial thickness, the biomarker pattern is preserved in other consensus CIN2.

iii) As in CIN1, some CIN2 show a more extensive expression of MCM into the upper epithelial layers when compared to $p16^{INK4a}$. E4 expression is limited to cells close to the epithelial surface. Panel B is enlarged in Fig.5 to illustrate the correlation between pathology and biomarker patterns.

B. HPV_E4 negativity amongst consensus CIN2.

i) Loss of the E4 biomarker in a proportion of CIN2, can be associated with extensive expression of the $p16^{INK4a}$ /MCM biomarkers throughout the thickness of the epithelium, as well as an absence of obvious differentiation at the level of pathology. Panel C is enlarged in Fig. 5.

ii) MCM can extend closer throughout the epithelium more robustly than $p16^{INK4a}$ without the expression of the E4 biomarker. The panel shown in D is enlarged in Fig. 5 to show the absence of key pathology features apparent in the E4-positive CIN2.

Figure 5. **Expression of the HPV_E4 biomarker in CIN2 is associated with discrete regions of CIN1-like pathology.**

A.

i) Pathology associated with E4 expression in CIN2. The six images show enlargements of the regions that are boxed in Fig. 4A(i) and (iii) (H&E images). The pattern of vacuolation and koilocyte formation is similar to that seen in CIN1 (Fig. 2B), with E4 expression, vacuolation and MCM decline coinciding closely (arrows 1&2).

B.

ii) The six images show an enlargement of the regions that are boxed in Fig 4B(i) and (ii). In these lesions, vacuolated cells are sometimes apparent (arrows marked 'v'), but are not necessarily associated with expression of the E4 biomarker (Fig. 4B(i) and (ii)).

Figure 6. **Division of Cervical Pathologies according to the Presence and Distribution of the Molecular Markers E4 and $p16^{INK4a}$.**

A. Lesional areas where there was total agreement amongst the panel of pathologists. The columns in each graph show the individual diagnostic opinions provided by the pathologist panel after review of the H&E stained slides, as either non-CIN, CIN1, CIN2 or CIN3. Graphs shown in A include only lesional areas where there was total agreement amongst the pathologist panel. This standard pathology-grading is stratified according to whether the diagnosed areas were subsequently found to be HPV-E4-positive (green edged columns/left-most graph) or E4-negative (red edged columns/right-most graph), and to what extent the $p16^{INK4a}$ expression extended through the epithelium. Lesional areas showing full thickness p16 staining are indicated as dark brown columns, with lower levels of staining being shown in lighter shades of brown. Lesional areas that lacked $p16^{INK4a}$ staining are indicated by white columns. In general, the patterns fit with our current model of life-cycle deregulation in high-grade neoplasia, with an absence of both markers in the 'total-agreement' non-CIN group. As described in the text, E4 expression in CIN3 (marked by an asterisk) was typically sporadic and in a small number of cells close to the epithelial surface.

B. All lesional areas including those where there was disagreement amongst the panel of pathologists. The columns show the individual diagnostic opinions provided by the pathologist panel on all lesional areas, irrespective of whether there was agreement or disagreement between individual pathologists. Labelling is as described in 'A' above. Because

the H&E-based diagnostic opinion often differed between the individual pathologists, the pattern of the $p16^{INK4a}$ and E4 staining typical of non-CIN, CIN1, CIN2 and CIN3 (shown in Fig.6A) is less apparent, with some evidence of virus infection in the 'proposed' non-CIN group, and an absence of biomarker staining in some 'proposed' CIN1.

Figure 7. **Molecular Principles Underlying the use of p16, MCM and E4 as HPV-Associated Disease Biomarkers**

A. In uninfected epithelium, the cellular MCM protein (red) is usually detectable at low levels only in the basal and parabasal cell layers as a result of cell-cycle stimulation by growth factors. This facilitates the phosphorylation of pRb by cyclin-dependent kinases, the release of the E2F transcription factor, and the regulated expression of MCM. During normal metaplasia or wound healing, MCM may also be detected in the upper epithelial layers. The cellular $p16^{INK4a}$ protein is also stimulated by E2F, but does not usually accumulate to detectable levels in uninfected epithelium. It provides feedback-regulation on the activity of cyclin-dependent kinases. $p16^{INK4a}$ is sometimes visualised as a weak cytoplasmic stain in cells undergoing senescence (pale brown). The HPV-encoded E4 protein is never expressed in uninfected epithelium and E4 antibodies show no reactivity with cellular proteins.

B. (i) In HPV-infected epithelial tissue, the high-risk E6 & E7 genes (red) are expressed together from the viral early promoter (PE), and function to drive cell-cycle entry in order to allow cell proliferation and genome amplification. The high-risk E4 gene (green) is expressed from a spliced mRNA, and becomes abundant following the activation of the viral late promoter (PL) as the infected cell exits the cell cycle and commits to true differentiation.

(ii) E6 and E7 are expressed at low level in the cell, but the consequences of their presence can be visualised by alterations in the presence of $p16^{INK4a}$ and MCM. The association of E7 with pRb leads to E2F release irrespective of growth factor stimulation. This allows MCM and also $p16^{INK4a}$ to accumulate to higher levels than are typically seen in uninfected epithelium where expression is dependent on cyclin-dependent kinase activation. The E7 protein also acts to increase the transcription of $p16^{INK4a}$ as a result of epigenetic modification of the $p16^{INK4a}$ promoter. In this context, $p16^{INK4a}$ and MCM can be used with caution as surrogate markers of E6/E7 deregulation.

The viral E4 protein becomes abundant in the upper layers of HPV-infected epithelium as a result of viral late promoter activation and the cleavage of the full-length E4 protein by calpain. Calpain-cleavage exposes a C-terminal multimerization motif in E4, which allows its assembly into amyloid-like fibres. The high-level accumulation of E4-amyloid is thought to coincide with progression of the infected cell through the G2 phase of the cell cycle and eventually to cell-cycle exit, explaining its appearance as MCM levels decline.

References

1. Bosch FX, Broker TR, Forman D*, et al.* Comprehensive control of human papillomavirus infections and related diseases. Vaccine 2013;31 Suppl 8:I1-31.

2. Ferlay J, Shin HR, Bray F*, et al.* Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer 2010;127:2893-917.

3. Doorbar J, Quint W, Banks L*, et al.* The biology and life-cycle of human papillomaviruses. Vaccine 2012;30 Suppl 5:F55-70.

4. Westra WH. The morphologic profile of HPV-related head and neck squamous carcinoma: implications for diagnosis, prognosis, and clinical management. Head Neck Pathol 2012;6 Suppl 1:S48-54.

5. de Villiers EM, Fauquet C, Broker TR, Bernard HU, zur Hausen H. Classification of papillomaviruses. Virology 2004;324:17-27.

6. Peh WL, Middleton K, Christensen N*, et al.* Life cycle heterogeneity in animal models of human papillomavirus-associated disease. Journal of Virology 2002;76:10401-16.

7. Cubie HA. Diseases associated with human papillomavirus infection. Virology 2013;445:21-34.

8. Jenkins D. Histopathology and cytopathology of cervical cancer. Dis Markers 2007;23:199-212.

9. Arwert EN, Hoste E, Watt FM. Epithelial stem cells, wound healing and cancer. Nat Rev Cancer 2012;12:170-80.

10. Kurman RJ, Ellenson, L.H., Ronnett, B.M. Blaustein's Pathology of the Female Genital Tract. Springer US: 2010.

11. Pyeon D, Pearce SM, Lank SM, Ahlquist P, Lambert PF. Establishment of human papillomavirus infection requires cell cycle progression. PLoS Pathog 2009;5:e1000318.

12. Middleton K, Peh W, Southern S*, et al.* Organization of human papillomavirus productive cycle during neoplastic progression provides a basis for selection of diagnostic markers. J Virol 2003;77:10186-201.

13. Bodily J, Laimins LA. Persistence of human papillomavirus infection: keys to malignant progression. Trends in microbiology 2011;19:33-9.

14. Isaacson Wechsler E, Wang Q, Roberts I*, et al.* Reconstruction of human papillomavirus type 16-mediated early-stage neoplasia implicates E6/E7 deregulation and the loss of contact inhibition in neoplastic progression. J Virol 2012;86:6358-64.

15. Hafner N, Driesch C, Gajda M*, et al.* Integration of the HPV16 genome does not invariably result in high levels of viral oncogene transcripts. Oncogene 2008;27:1610-7.

16. Khleif SN, DeGregori J, Yee CL*, et al.* Inhibition of cyclin D-CDK4/CDK6 activity is associated with an E2F-mediated induction of cyclin kinase inhibitor activity. Proc Natl Acad Sci U S A 1996;93:4350-4.

17. Chen HC, Schiffman M, Lin CY*, et al.* Persistence of type-specific human papillomavirus infection and increased long-term risk of cervical cancer. J Natl Cancer Inst 2011;103:1387-96.

18. Geraets DT, van Doorn LJ, Kleter B*, et al.* Long-term follow-up of HPV16-positive women: persistence of the same genetic variant and low prevalence of variant co-infections. PLoS One 2013;8:e80382.

19. Quint KD, de Koning MN, Quint WG, Pirog EC. Progression of cervical low grade squamous intraepithelial lesions: in search of prognostic biomarkers. Eur J Obstet Gynecol Reprod Biol 2013;170:501-6.

20. Cantor SB, Atkinson EN, Cardenas-Turanzas M*, et al.* Natural history of cervical intraepithelial neoplasia: a meta-analysis. Acta Cytol 2005;49:405-15.

21. Doorbar J, Cubie H. Molecular basis for advances in cervical screening. Mol Diagn 2005;9:129-42.

22. Roman A, Munger K. The papillomavirus E7 proteins. Virology 2013;445:138-68.

23. Tornesello ML, Buonaguro L, Giorgi-Rossi P, Buonaguro FM. Viral and cellular biomarkers in the diagnosis of cervical intraepithelial neoplasia and cancer. BioMed research international 2013;2013:519619.

24. von Knebel Doeberitz M, Reuschenbach M, Schmidt D, Bergeron C. Biomarkers for cervical cancer screening: the role of p16(INK4a) to highlight transforming HPV infections. Expert review of proteomics 2012;9:149-63.

25. Ikenberg H, Bergeron C, Schmidt D*, et al.* Screening for cervical cancer precursors with p16/Ki-67 dual-stained cytology: results of the PALMS study. J Natl Cancer Inst 2013;105:1550-7.

26. McLaughlin-Drubin ME, Crum CP, Munger K. Human papillomavirus E7 oncoprotein induces KDM6A and KDM6B histone demethylase expression and causes epigenetic reprogramming. Proceedings of the National Academy of Sciences of the United States of America 2011;108:2130-5.

27. Doorbar J. Papillomavirus life cycle organization and biomarker selection. Disease Markers 2007;23:297-313.

28. Griffin H, Wu Z, Marnane R*, et al.* E4 antibodies facilitate detection and type-assignment of active HPV infection in cervical disease. PLoS One 2012;7:e49974.

29. van der Marel J, van Baars R, Quint W, *et al.* The impact of human papillomavirus genotype on colposcopic appearance: a cross-sectional analysis. BJOG 2014.

30. Kleter B, van Doorn LJ, ter Schegget J, *et al.* Novel short-fragment PCR assay for highly sensitive broad-spectrum detection of anogenital human papillomaviruses. The American journal of pathology 1998;153:1731-9.

31. Kleter B, van Doorn LJ, Schrauwen L, *et al.* Development and clinical evaluation of a highly sensitive PCR-reverse hybridization line probe assay for detection and identification of anogenital human papillomavirus. J Clin Microbiol 1999;37:2508-17.

32. Quint W, Jenkins D, Molijn A, *et al.* One virus, one lesion--individual components of CIN lesions contain a specific HPV type. The Journal of pathology 2012;227:62-71.

33. Doorbar J, Foo C, Coleman N, *et al.* Characterization of events during the late stages of HPV16 infection in vivo using high-affinity synthetic Fabs to E4. Virology 1997;238:40-52.

34. de Sanjose S, Quint WG, Alemany L, *et al.* Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study. Lancet Oncol 2010;11:1048-56.

35. Kalantari M, Garcia-Carranca A, Morales-Vazquez CD, *et al.* Laser capture microdissection of cervical human papillomavirus infections: copy number of the virus in cancerous and normal tissue and heterogeneous DNA methylation. Virology 2009;390:261-7.

36. Guan P, Howell-Jones R, Li N, *et al.* Human papillomavirus types in 115,789 HPV-positive women: a meta-analysis from cervical infection to cancer. International journal of cancer Journal international du cancer 2012;131:2349-59.

37. Davy CE, Jackson DJ, Raj K, *et al.* Human Papillomavirus Type 16 E1^E4-Induced G2 Arrest Is Associated with Cytoplasmic Retention of Active Cdk1/Cyclin B1 Complexes. J Virol 2005;79:3998-4011.

38. Carreon JD, Sherman ME, Guillen D, *et al.* CIN2 is a much less reproducible and less valid diagnosis than CIN3: results from a histological review of population-based cervical samples. Int J Gynecol Pathol 2007;26:441-6.

39. Cuzick J, Bergeron C, von Knebel Doeberitz M, *et al.* New technologies and procedures for cervical cancer screening. Vaccine 2012;30 Suppl 5:F107-16.

40. Bulkmans NW, Rozendaal L, Voorhorst FJ, Snijders PJ, Meijer CJ. Long-term protective effect of high-risk human papillomavirus testing in population-based cervical screening. Br J Cancer 2005;92:1800-2.

41. Elfstrom KM, Smelov V, Johansson AL, *et al.* Long-term HPV type-specific risks for ASCUS and LSIL: A 14-year follow-up of a randomized primary HPV screening trial. Int J Cancer 2014.

42. Ronco G, Dillner J, Elfstrom KM*, et al.* Efficacy of HPV-based screening for prevention of invasive cervical cancer: follow-up of four European randomised controlled trials. Lancet 2014;383:524-32.

43. Kocken M, Uijterwaal MH, de Vries AL*, et al.* High-risk human papillomavirus testing versus cytology in predicting post-treatment disease in women treated for high-grade cervical disease: a systematic review and meta-analysis. Gynecol Oncol 2012;125:500-7.

44. Litjens RJ, Hopman AH, van de Vijver KK*, et al.* Molecular biomarkers in cervical cancer diagnosis: a critical appraisal. Expert opinion on medical diagnostics 2013;7:365-77.

45. Maiorano D, Lutzmann M, Mechali M. MCM proteins and DNA replication. Curr Opin Cell Biol 2006;18:130-6.

46. Murphy N, Ring M, Heffron CC*, et al.* Quantitation of CDC6 and MCM5 mRNA in cervical intraepithelial neoplasia and invasive squamous cell carcinoma of the cervix. Mod Pathol 2005;18:844-9.

47. Murphy N, Ring M, Heffron CC*, et al.* p16INK4A, CDC6, and MCM5: predictive biomarkers in cervical preinvasive neoplasia and cervical cancer. J Clin Pathol 2005;58:525-34.

48. McCluggage WG, Walsh MY, Thornton CM*, et al.* Inter- and intra-observer variation in the histopathological reporting of cervical squamous intraepithelial lesions using a modified Bethesda grading system. British journal of obstetrics and gynaecology 1998;105:206-10.

49. Piaton E, Advenier AS, Carre C*, et al.* p16(INK4a) /Ki-67 dual labelling as a marker for the presence of high-grade cancer cells or disease progression in urinary cytopathology. Cytopathology 2013;24:327-34.

50. McIntosh PB, Martin SR, Jackson DJ*, et al.* Structural analysis reveals an amyloid form of the human papillomavirus type 16 E1^E4 protein and provides a molecular basis for its accumulation. Journal of Virology 2008;82:8196-203.

51. Khan J, Davy CE, McIntosh PB*, et al.* Role of calpain in the formation of human papillomavirus type 16 E1^E4 amyloid fibers and reorganization of the keratin network. Journal of Virology 2011;85:9984-97.

52. Wang Q, Kennedy A, Das P*, et al.* Phosphorylation of the human papillomavirus type 16 E1^E4 protein at T57 by ERK triggers a structural change that enhances keratin binding and protein stability. J Virol 2009;83:3668-83.

53. Doorbar J. The E4 protein; structure, function and patterns of expression. Virology 2013;445:80-98.

54. Davy C, Doorbar J. G2/M cell cycle arrest in the life cycle of viruses. Virology 2007;368:219-26.

55. Ortoft G, Henriksen T, Hansen E, Petersen L. After conisation of the cervix, the perinatal mortality as a result of preterm delivery increases in subsequent pregnancy. BJOG 2011;117:258-67.

56. Armarnik S, Sheiner E, Piura B*, et al.* Obstetric outcome following cervical conization. Arch Gynecol Obstet 2011;283:765-9.

57. Kinney W, Hunt WC, Dinkelspiel H*, et al.* Cervical excisional treatment of young women: A population-based study. Gynecol Oncol 2014;132:628-35.

58. Henderson D, Hall L, Prpic N*, et al.* The selection and characterization of antibodies to minichromosome maintenance proteins that highlight cervical dysplasia. J Immunol Methods 2011;370:1-13.

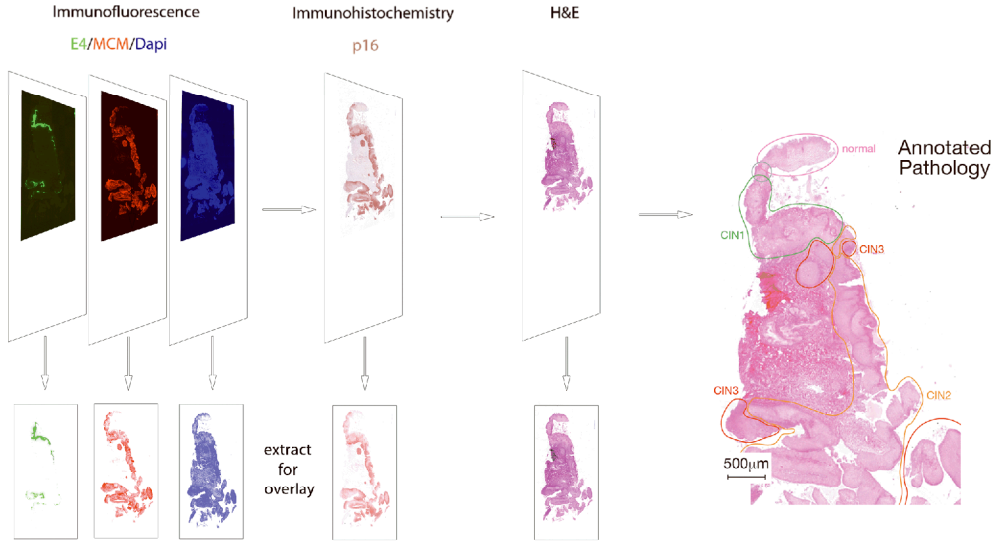
59. Galgano MT, Castle PE, Atkins KA*, et al.* Using biomarkers as objective standards in the diagnosis of cervical biopsies. Am J Surg Pathol 2010;34:1077-87.

60. Stoler MH, Schiffman M, Atypical Squamous Cells of Undetermined Significance-Low-grade Squamous Intraepithelial Lesion Triage Study G. Interobserver reproducibility of cervical cytologic and histologic interpretations: realistic estimates from the ASCUS-LSIL Triage Study. JAMA 2001;285:1500-5.

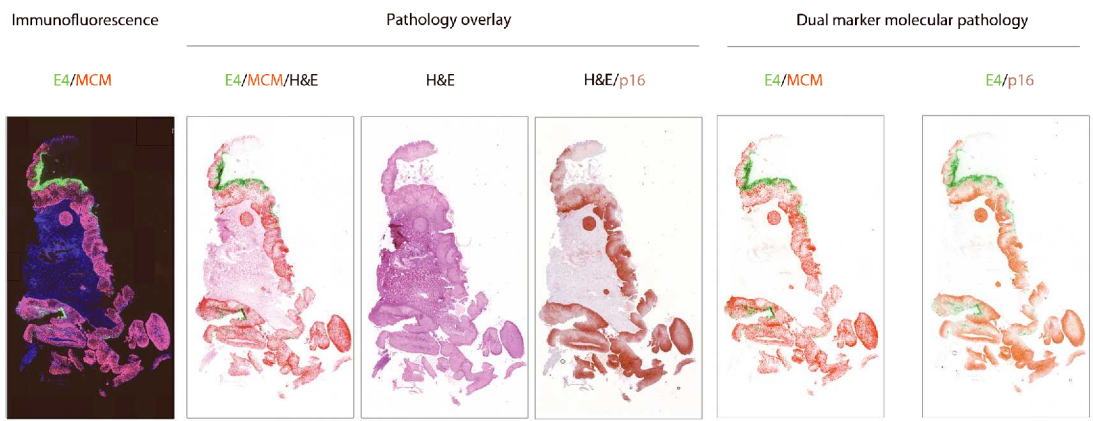
61. Vrdoljak-Mozetic D, Krasevic M, Versa Ostojic D*, et al.* HPV16 genotype, p16/Ki-67 dual staining and koilocytic morphology as potential predictors of the clinical outcome for cervical low-grade squamous intraepithelial lesions. Cytopathology 2013.

Type of file: figure
Label: 1
Filename: figure_1.tif

A


Immunofluorescence
E4/MCM/Dapi
Immunohistochemistry
p16
H&E
extract
for
overlay
normal
Annotated
Pathology
CIN1
CIN3
CIN3
CIN2
500µm

B


Immunofluorescence
Pathology overlay
Dual marker molecular pathology
E4/MCM
E4/MCM/H&E
H&E
H&E/p16
E4/MCM
E4/p16

Type of file: figure
Label: 2
Filename: figure_2.tif

A

Immunofluorescence

Pathology Overlay

Dual Marker Molecular Pathology

E4/MCM

E4/MCM/H&E

H&E

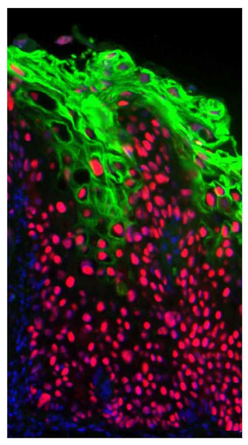
H&E/p16

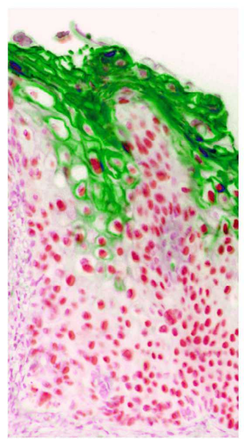
E4/MCM

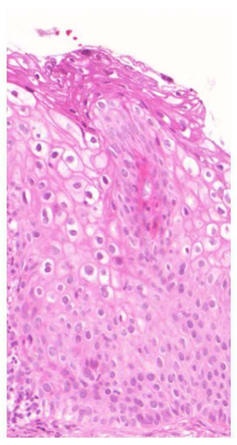
E4/p16

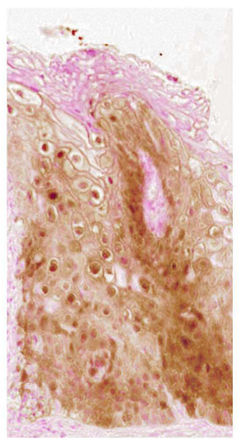
i

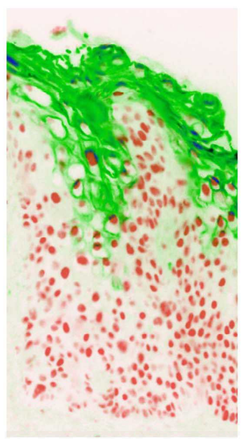
HPV51

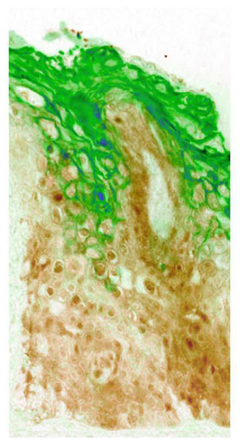






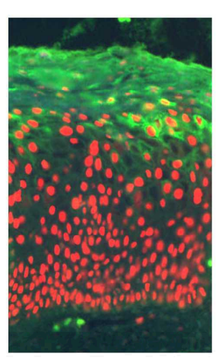


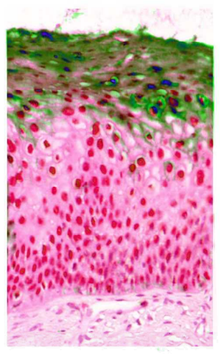


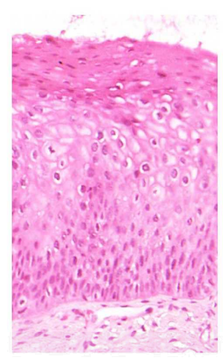


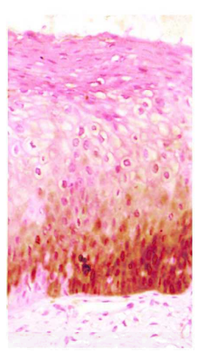
ii

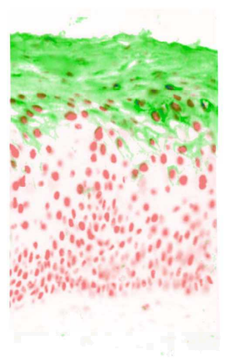
HPV16
HPV31
HPV58

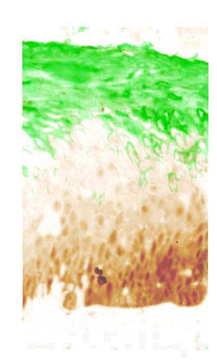






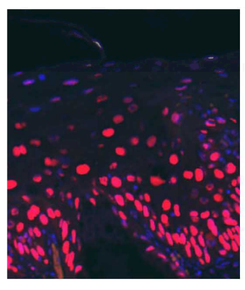


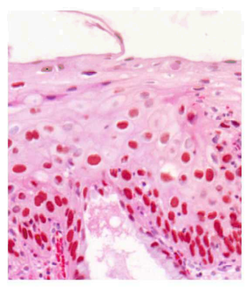


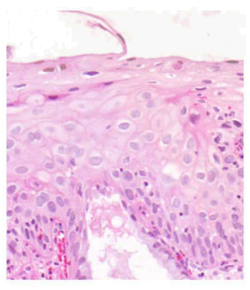


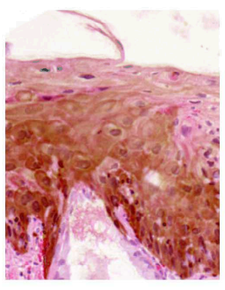
iii

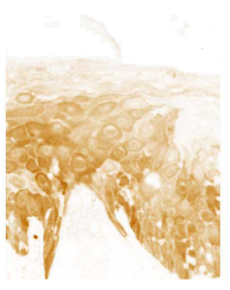
HPV52











B

Immunofluorescence

Pathology Overlay

Pathology/Biomarker Correlation

E4/MCM

E4/MCM/H&E

H&E

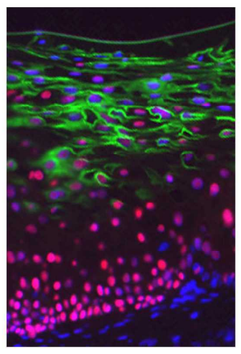
H&E

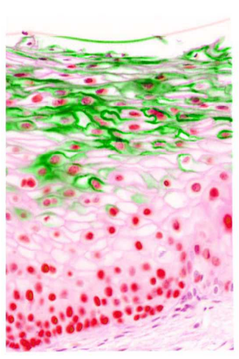
E4/MCM/H&E

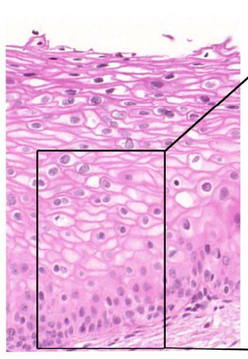
E4/MCM

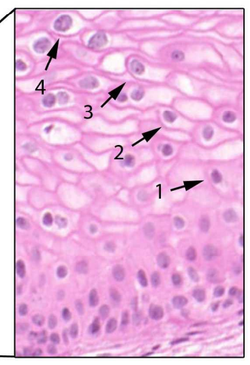
i

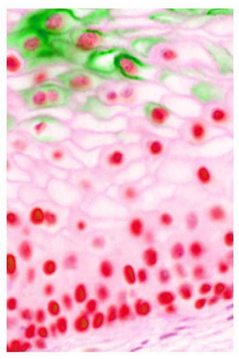
HPV52

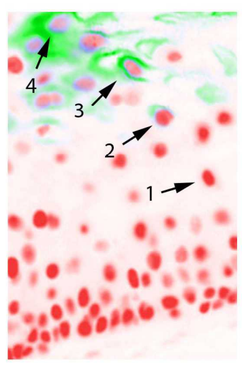






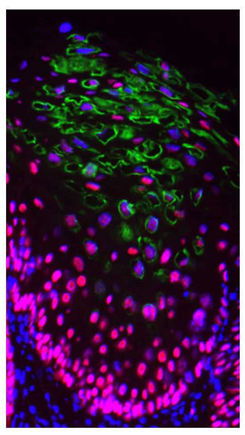


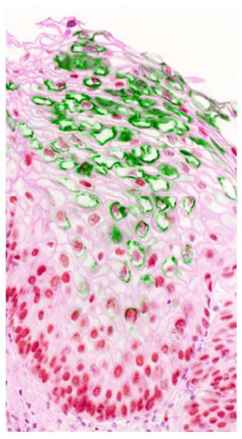


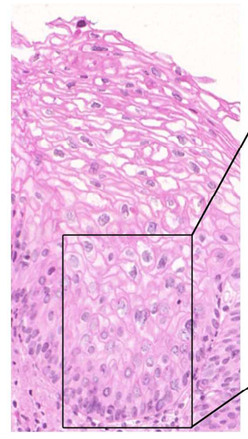


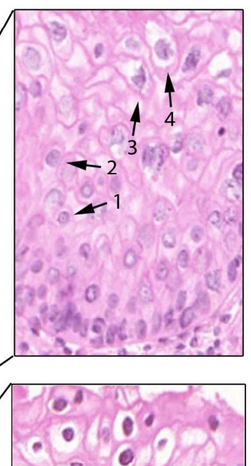
ii

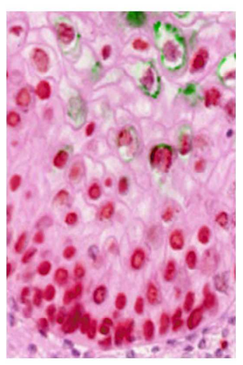
HPV56

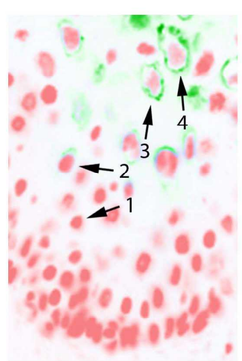






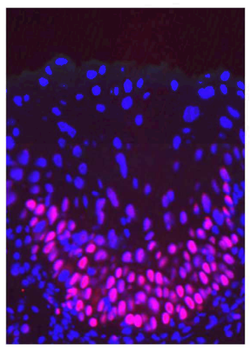


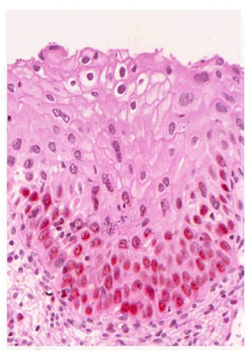


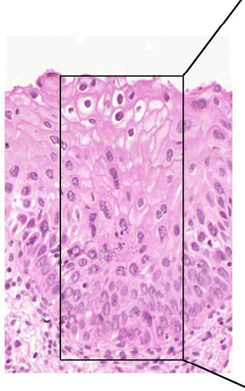


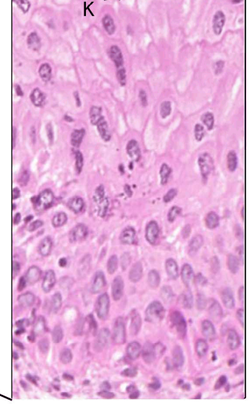
iii

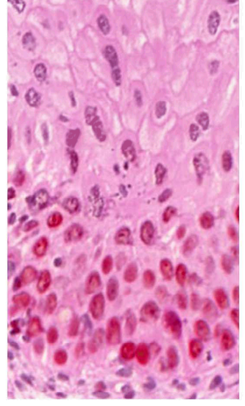
HPV33

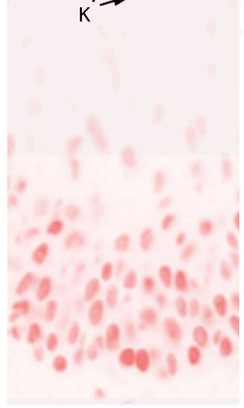












Type of file: figure
Label: 3
Filename: figure_3.tif

A

Immunofluorescence

Pathology overlay

Dual Marker Molecular Pathology

E4/MCM

E4/MCM/H&E

H&E

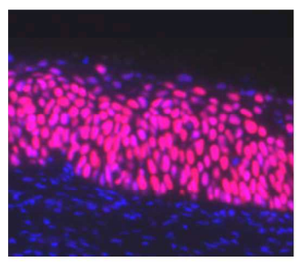
H&E/p16

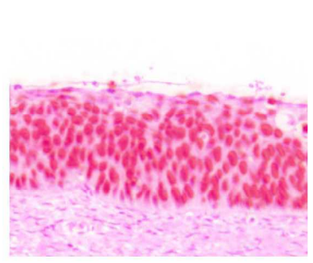
E4/MCM

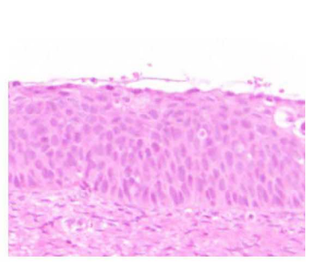
E4/p16

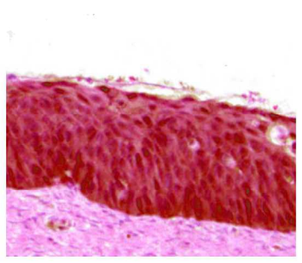
i

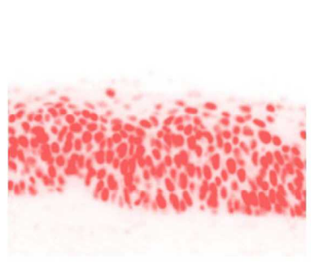
HPV11
HPV16
HPV53

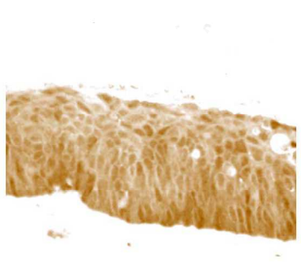






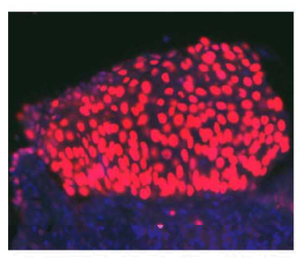


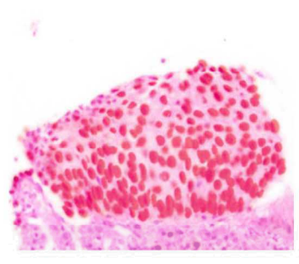


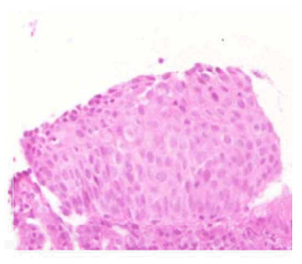


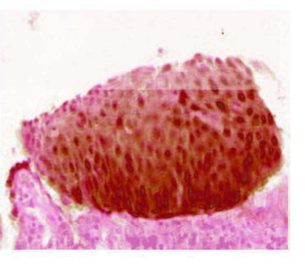
ii

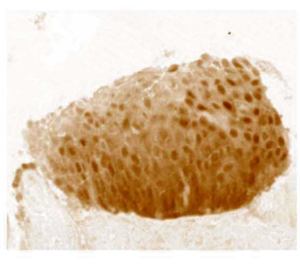
HPV16











B

Immunofluorescence

Pathology overlay

Dual Marker Molecular Pathology

E4/MCM

E4/MCM/H&E

H&E

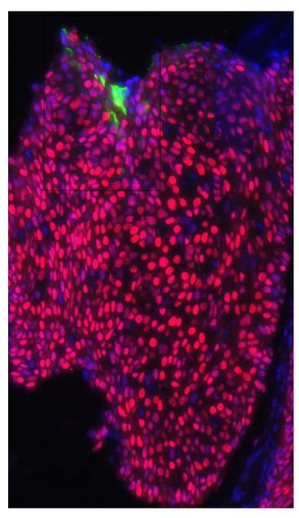
H&E/p16

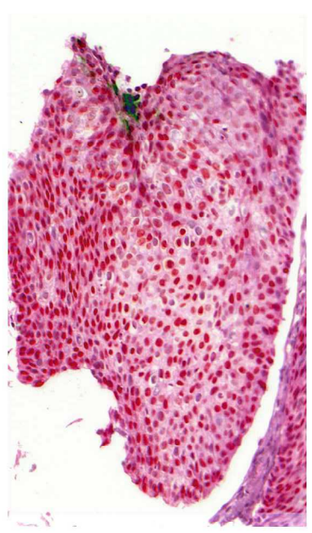
E4/MCM

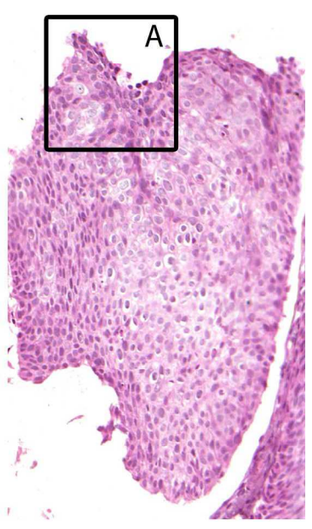
E4/p16

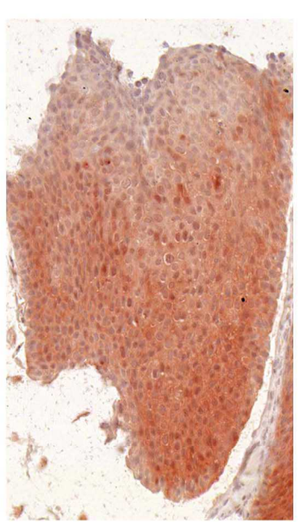
i

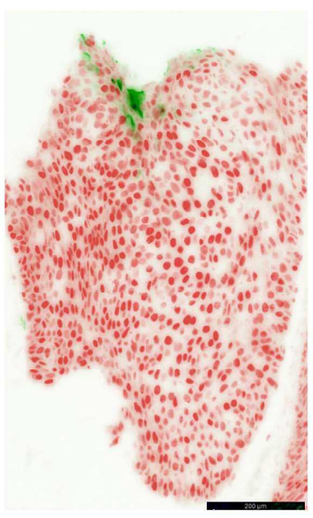
HPV16

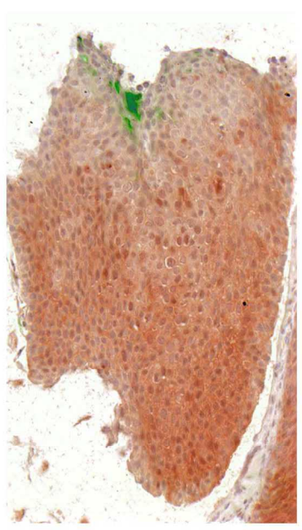






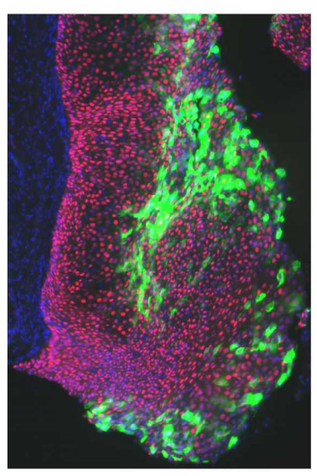


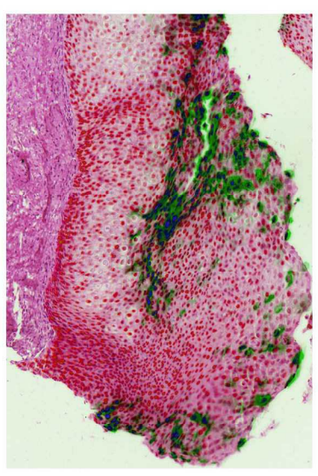


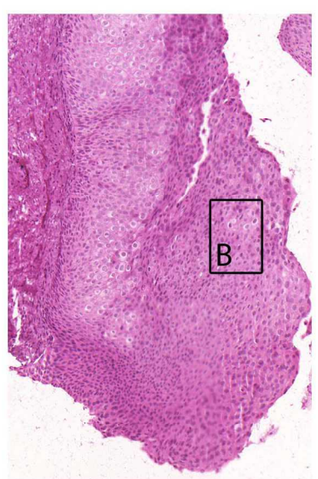


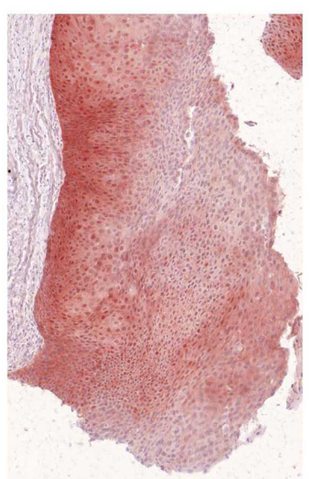
ii

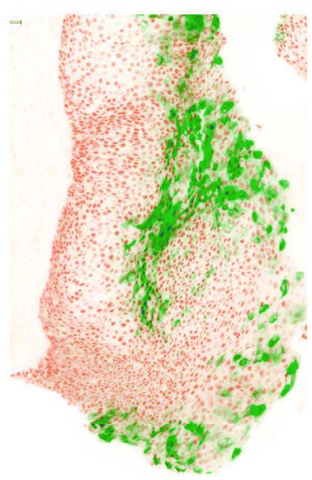
HPV16

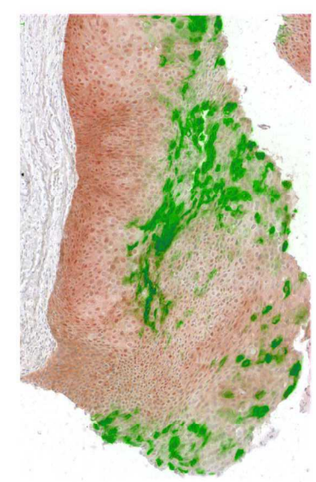












C

Pathology/Biomarker Correlation

Pathology/Biomarker Correlation

H&E

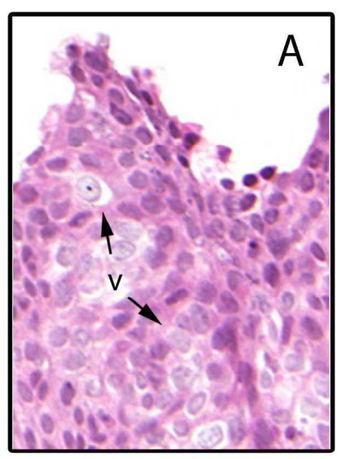
E4/MCM/H&E

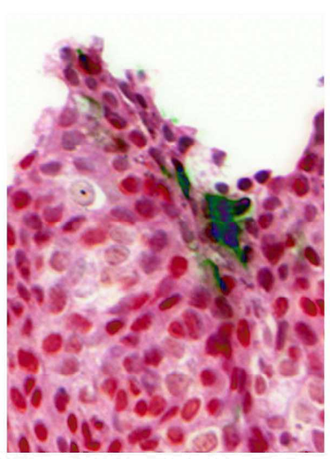
E4/MCM

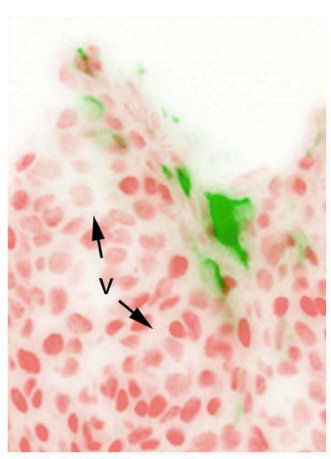
H&E

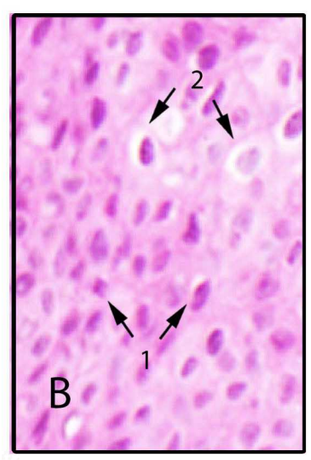
E4/MCM/H&E

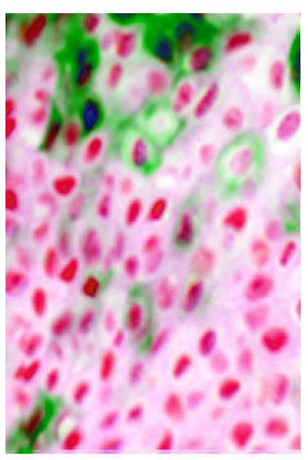
E4/MCM

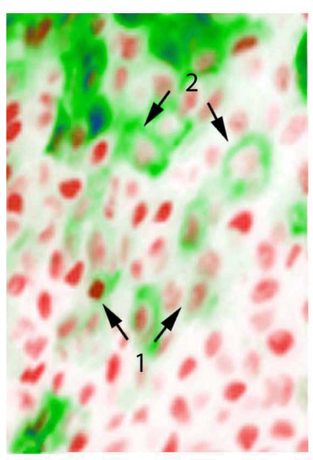












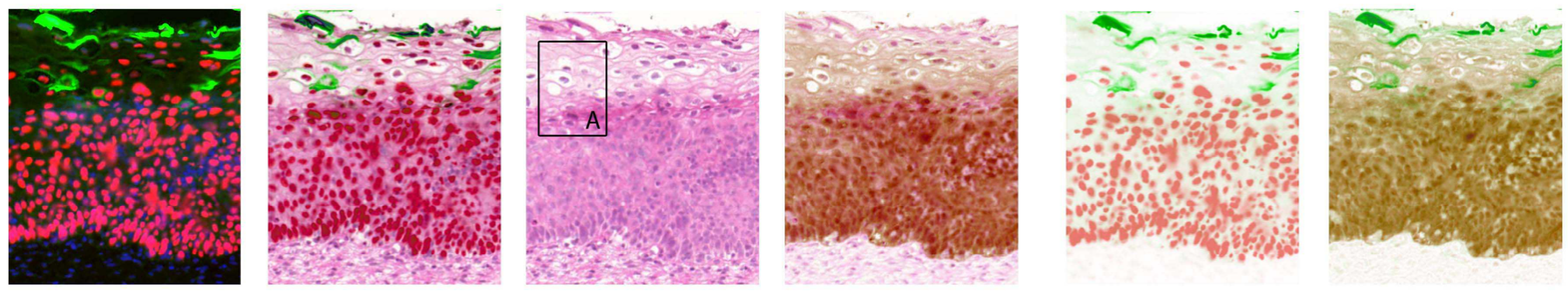
Type of file: figure
Label: 4
Filename: figure_4.tif

A

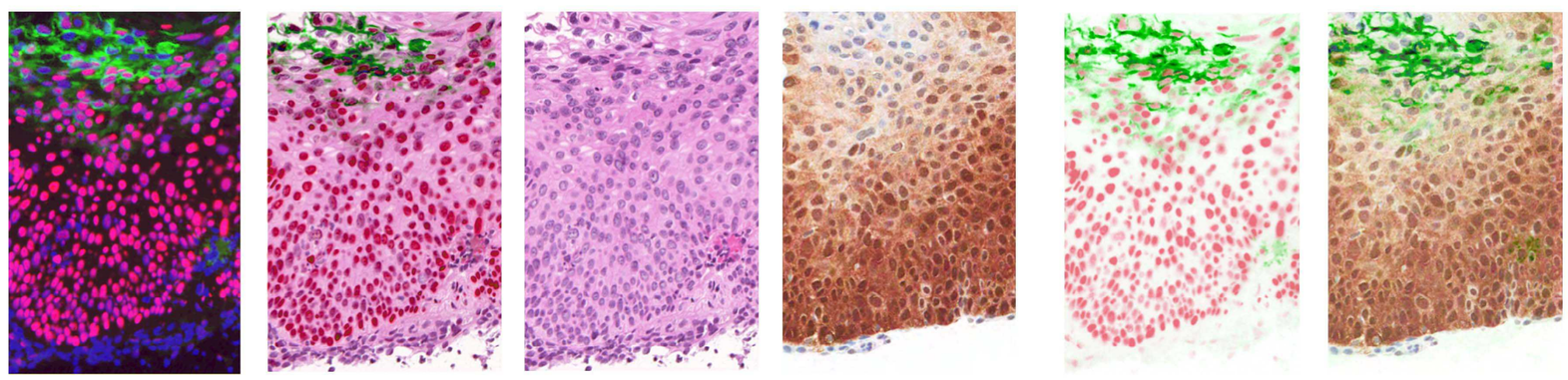
Immunofluorescence | Pathology overlay | Dual Marker Molecular Pathology

E4/MCM | E4/MCM/H&E | H&E | H&E/p16 | E4/MCM | E4/p16

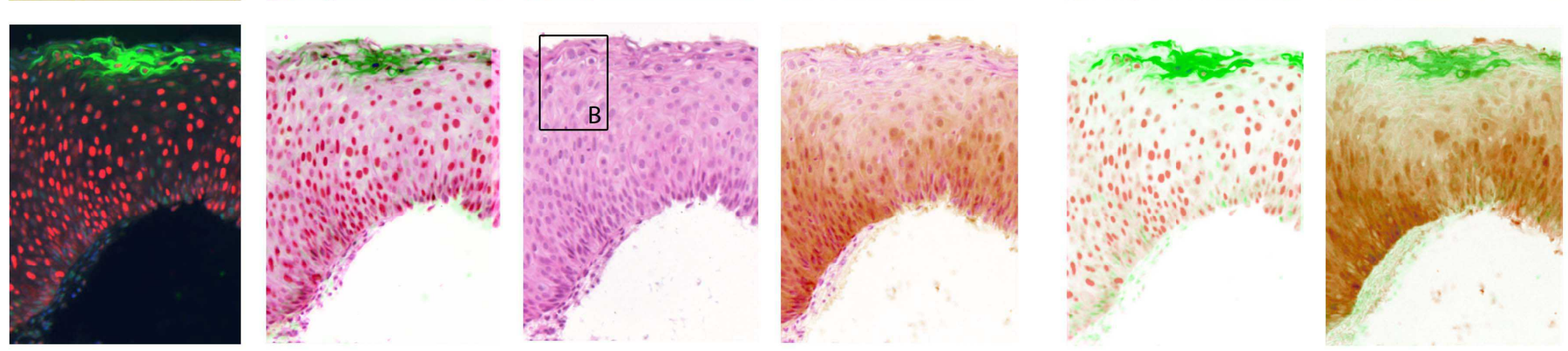
i HPV16



ii HPV52



iii HPV18 HPV31 HPV33

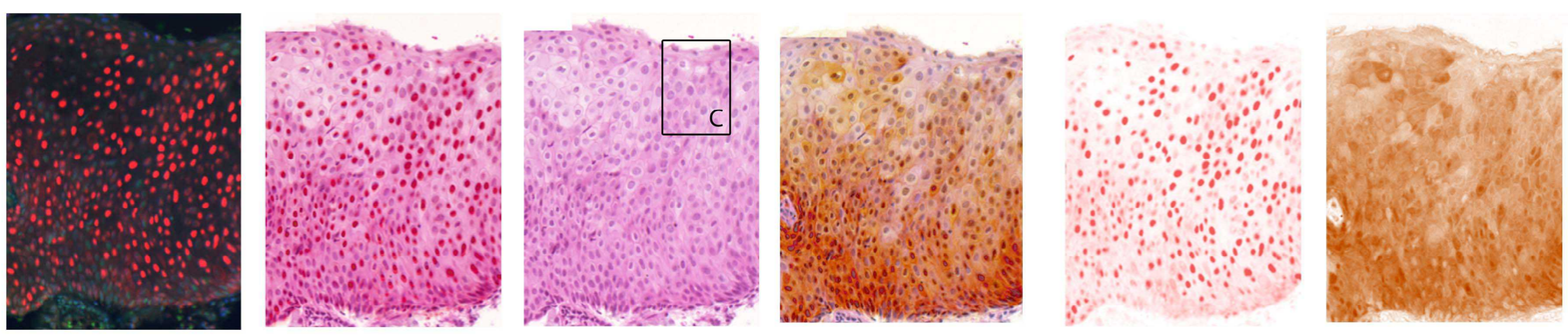


B

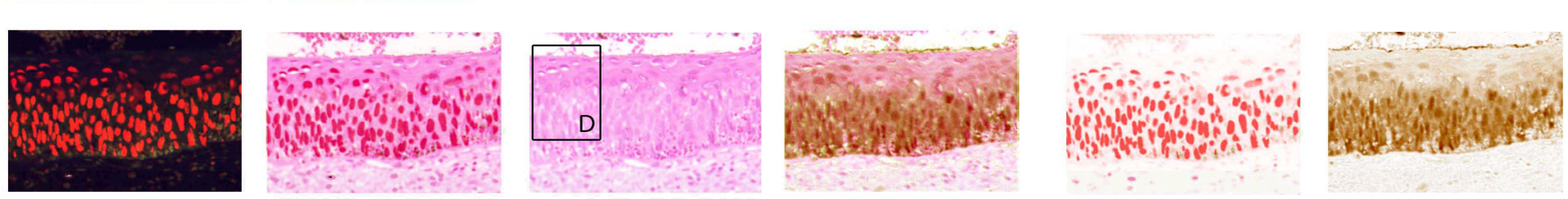
Immunofluorescence | Pathology overlay | Dual Marker Molecular Pathology

E4/MCM | E4/MCM/H&E | H&E | H&E/p16 | E4/MCM | E4/p16

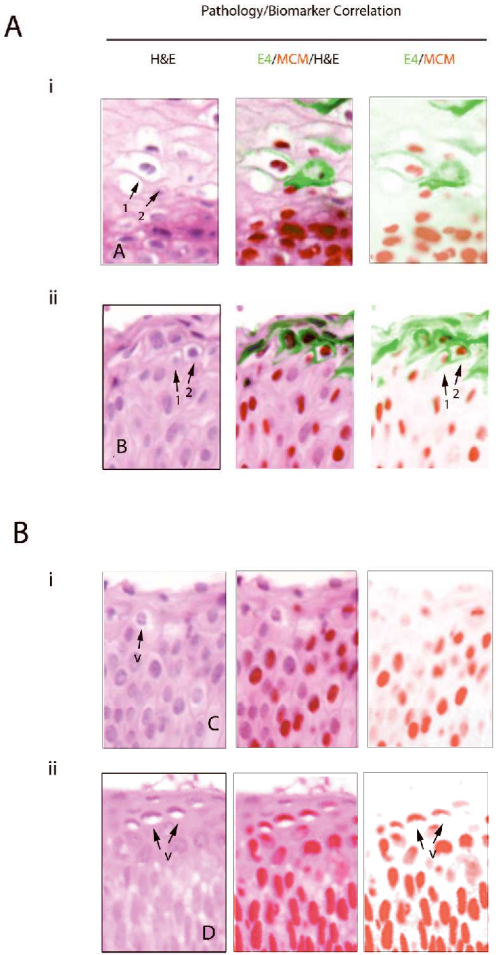
i HPV33



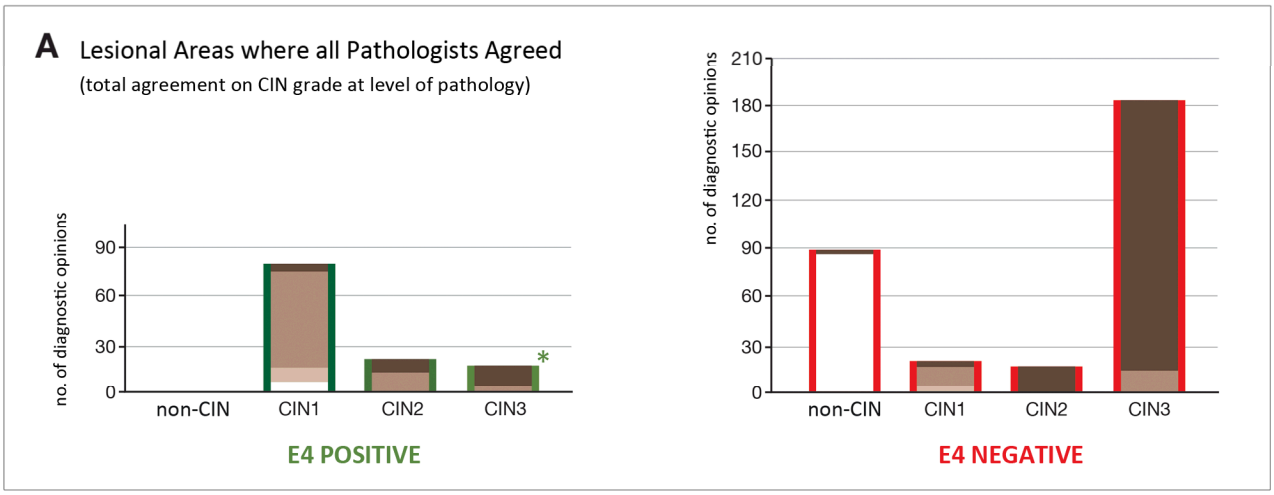
ii HPV58

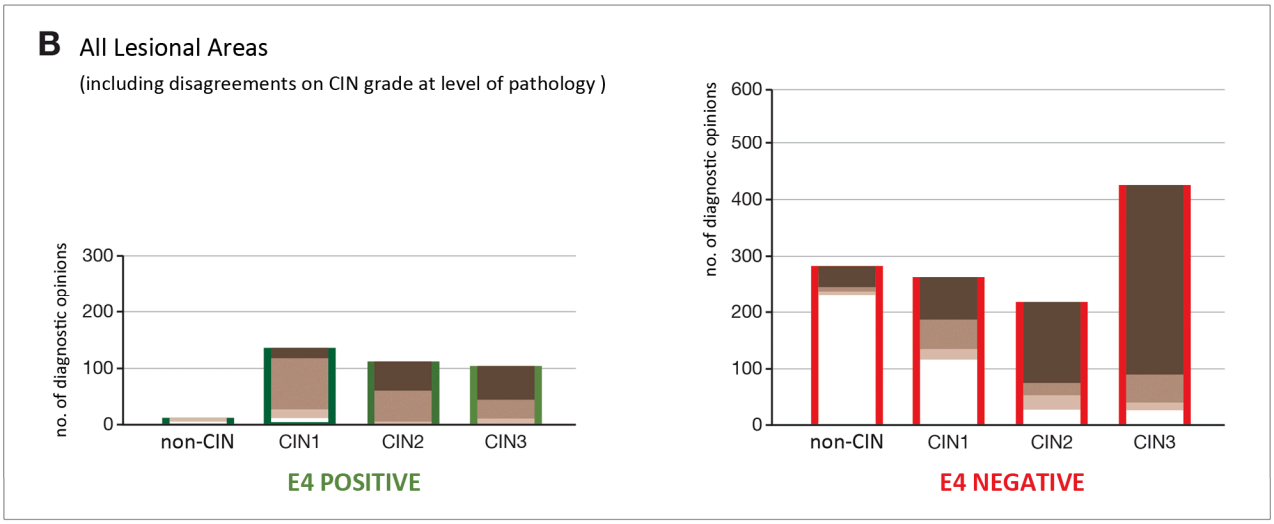


Type of file: figure
Label: 5
Filename: figure_5.tif


A
Pathology/Biomarker Correlation
H&E
E4/MCM/H&E
E4/MCM
i
1
2
A
ii
1
2
B
B
i
v
C
ii
v
D
v

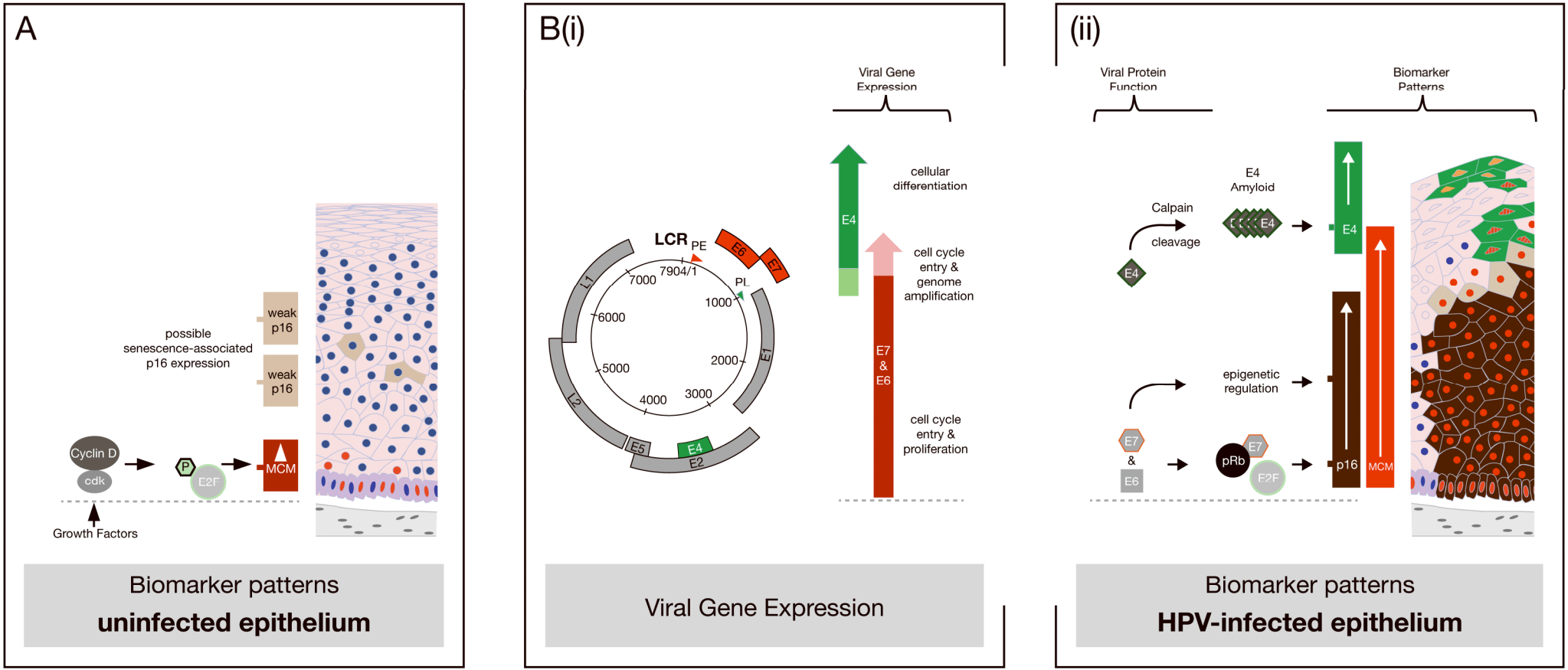
Type of file: figure
Label: 6
Filename: figure_6.tif





E4 positive	E4 negative	KEY
		p16 above 2/3
		p16 up to 2/3
		p16 up to 1/3
		p16 negative

Type of file: figure
Label: 7
Filename: figure_7.tif


A
possible
senescence-associated
p16 expression
weak
p16
weak
p16
Cyclin D
cdk
P
E2F
MCM
Growth Factors
Biomarker patterns
uninfected epithelium
B(i)
LCR
PE
E6
E7
7904/1
7000
PL
1000
6000
2000
5000
3000
4000
L1
L2
E1
E5
E4
E2
Viral Gene
Expression
E4
cellular
differentiation
cell cycle
entry &
genome
amplification
E7
&
E6
cell cycle
entry &
proliferation
Viral Gene Expression
(ii)
Viral Protein
Function
Biomarker
Patterns
Calpain
cleavage
E4
E4
Amyloid
E4
E4
epigenetic
regulation
E7
&
E6
pRb
E7
E2F
p16
MCM
Biomarker patterns
HPV-infected epithelium