Cellular Oncology 29 (2007) 483–495
IOS Press


# Objective assessment of cancer biomarkers using semi-rare event detection

Jeroen A.W.M. van der Laak *, Albertus G. Siebers, Sabine A.A.P. Aalders, Johanna M.M. Grefte, Peter C.M. de Wilde and Johan Bulten

*Department of Pathology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands*

**Abstract.** Objective and reproducible assessment of cancer biomarkers may be performed using rare event detection systems. Because many biomarkers are not true 'rare events', in this study a semi-rare event detection system was developed. The system is capable of assigning a discriminant score to detected positive cells, expressing the extent and intensity of the immunocytochemical staining. A gallery image is constructed showing the diagnostically most interesting cells as well as quantitative data expressing the biomarker staining pattern. To increase scanning speed, an adaptive scanning strategy is studied in which scanning is aborted when a sufficient number of positive cells has been identified. System performance was evaluated using liquid based cervical smears, stained with an antibody directed against $p16^{INK4a}$ tumor suppressor protein. Overexpression of $p16^{INK4a}$ in cervix is related to high-risk HPV infection, which is associated with carcinogenesis. Reproducibility of the system was tested on specimens containing limited positivity. Quantitative analysis was evaluated using 10 cases within normal limits and 10 high grade lesions. The system was highly reproducible in detecting positive cells and in calculating discriminant scores (average CV 0.7%). Quantitative features were significantly increased in high grade lesions ($p < 0.001$). Adaptive scanning decreased scanning time with only minor impact on scanning results. The system is capable of automated, objective and reproducible assessment of biomarker expression and may be useful for a variety of applications.

Keywords: Rare event detection, automated image analysis, immunocytochemistry, cervical cytology, $p16^{INK4a}$ protein

## 1. Introduction

The main task in clinical pathology is to discriminate between benign and (pre)malignant tissues or cells. Visual microscopic identification of dysplastic cells is a tedious task, which is hampered by limited sensitivity and specificity. Recent advances in molecular techniques have identified a number of promising biomarkers, which may yield additional information or may even serve as a surrogate for traditional morphological analysis. These markers may reflect more subtle cellular processes, adding information which exceeds the conventional morphological evaluation of cells. Immunocytochemistry or in situ hybridization can be used to visualize possibly abnormal cells and may thus facilitate detection of abnormalities. For example, a promising biomarker for recognition of atypical or dysplastic cells in cervical cytology is cyclin dependent kinase inhibitor $p16^{INK4a}$ [22]. Overexpression of $p16^{INK4a}$ protein has been shown to be closely related to the presence of high risk human papilloma virus (HR-HPV), which is strongly associated with cervical carcinogenesis [3,27]. Overexpression of $p16^{INK4a}$ is hypothized to result from HR-HPV mediated inactivation of the pRB pathway [1, 27]. Immunocytochemical (ICC) detection of $p16^{INK4a}$ overexpression has been described as a possible aid in improving cervical cytology [10,19].

An advantage of using biomarkers is the possibility of automated (pre)screening by means of rare event detection systems. Such systems have been developed based on both image cytometry [11,17,20,24] and flow cytometry [5,6]. These systems typically scan an entire sample for the presence of positive events. When image cytometry is used, images of detected events may be presented to the pathologist for visual examination [17]. In this way, rare event detection may be applied for prescreening. Using rare event detection, one positive event in $10^6$ cells may be detected [11,17], underlining the superior sensitivity of such techniques over manual evaluation [16]. Rare event detection is highly

*Corresponding author: Jeroen A.W.M. van der Laak, PhD, Radboud University Nijmegen Medical Centre, Department of Pathology 824, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands. Tel.: +31 243614367; Fax: +31 243668750; E-mail: J.vanderlaak@pathol.umcn.nl

suitable for locating true rare events, such as disseminated tumor cells in lymph node specimens. However, in many applications positive events may also occur relatively frequently in normal specimens, due to technical imperfections in staining or to suboptimal immunocytochemical detection. For instance, $p16^{INK4a}$ has been described to stain squamous metaplastic cells, which are present in most normal cervical smears [1, 22,25]. Also, weak cytoplasmic $p16^{INK4a}$ staining may be observed in non-dysplastic epithelial cells [25].

Assessment of such biomarkers, which may occur relatively frequently in both normal and abnormal samples, requires more sophisticated rare event detection than is traditionally used. In samples with extensive positivity, only the diagnostically most interesting events should be presented to the pathologist. To be applicable in a routine diagnostic setting, scanning speed is an important parameter [11,16]. Next to the detection of positive events, the system may be used to quantitatively evaluate biomarker expression, thus yielding additional information which may support diagnosis making.

The aims of the present study were:

1. To develop a semi-rare event detection system which is capable of rating diagnostic importance of positive events. A gallery image is constructed showing only the most interesting cells/cell groups.

2. To explore the possibility of speed gain by applying an adaptive scanning technique, in which the sensitivity of rare event detection depends on the density of positive events.

3. To study the possibility of expressing biomarker appearance quantitatively. Evaluation of the semi-rare event detection procedure was performed using $p16^{INK4a}$ stained liquid based cytology slides.

## 2. Materials and methods

### 2.1. Patient selection

Cervical smears used in this study were taken from either women participating in the national Dutch cervical cancer screening program, or from women referred to a gynecologist. Cervical smears were cytologically diagnosed by liquid based Pap screening. Smears of 10 cases with cytological diagnosis ranging from 'within normal limits' (WNL) to 'high grade squamous intraepithelial lesion' (HSIL) were randomly selected without prior knowledge of $p16^{INK4a}$ status, to construct the segmentation algorithm and discriminant functions (training samples). To study reproducibility of detection of positive events on cell level, three new cervical smears were selected containing a limited number of visually identified $p16^{INK4a}$ positive cells. A separate set of cervical smears of 10 cases WNL and 10 HSIL were randomly selected for quantitative analysis.

### 2.2. Specimen preparation

Cervical specimens were collected using the Rovers® Cervex-Brush® (Rovers Medical Devices BV, Oss, The Netherlands) and subsequently rinsed in a ThinPrep® vial containing PreservCyt® transport medium (Cytyc Corporation, Boxborough, MA). The vials were processed using a ThinPrep® T2000 (Cytyc Corporation, Boxborough, MA). Using this system, blood, mucous and non-diagnostic debris are removed and a thin layer of cellular material is produced on a defined area of an object glass. The specimen was examined routinely (Pap test) in the normal laboratory setting. A second ThinPrep from the same vial was prepared for ICC. Cell density was assessed for three ThinPreps by visually counting all epithelial cells in 20 randomly selected microscopic fields on a computer screen (200× magnification), applying the forbidden line method [7]. The total number of cells for the entire ThinPrep was estimated from this.

### 2.3. Immunocytochemical staining

Specimens for ICC were collected in 96% ethanol after processing in the ThinPrep T2000 device. Specimens were air dried for 3 hours after remaining in ethanol for 24 hours. Staining of $p16^{INK4a}$ was performed using monoclonal antibody JC8 (Lab Vision Corporation, Fremont, CA) after post-fixation for 30 minutes in 10% neutral buffered formaldehyde. All reactions were performed at room temperature, unless stated otherwise.

After rinsing with phosphate-buffered saline (PBS), endogenous peroxidase activity was blocked by incubation for 30 minutes in PBS containing 3% $H_2O_2$. After rinsing with PBS, antigen retrieval for $p16^{INK4a}$ was performed using microwave boiling in a 10 mM sodium citrate buffer (pH 6.0) for 10 minutes. The slides were allowed to cool down for at least 30 minutes after boiling. After rinsing with PBS, slides were incubated with the primary antibody for 1 hour. Slides were rinsed with PBS and incubated for 30 minutes with Powervision (Immunologic, Duiven, The

Netherlands). After rinsing, the chromogen 3-amino-9-ethyl carbazole (AEC) was applied for 10 minutes at 37°C, resulting in red staining of $p16^{INK4a}$ positive cells. Slides were counterstained for 15 seconds with hematoxylin solution, rinsed in water, and mounted using Imsol-mount (aqueous-based mounting medium; Klinipath BV, Duiven, The Netherlands). Slides used in the present study were stained in different staining batches, to account for possible inter-batch staining variations.

### 2.4. Automated assessment

Immunostained ThinPreps were evaluated using a fully automatic measurement procedure. Images were acquired with an AxioCam MRc red green blue (RGB) CCD camera connected to an AxioPlan 2 Imaging microscope (Carl Zeiss, Germany). The CCD has square pixels of size 6.45 × 6.45 $um^2$ (Bayer arrangement) and shows a linear response. The 3 × 12 bit RGB camera signal was reduced to 3 × 8 bits by omitting the least significant bits for each camera channel. The microscope was equipped with an automatic scanning stage (Märzhäuser GmbH, Wetzlar, Germany) which may contain 8 object glasses. Microscope and stage were fully computer controlled. Scanning for possibly positive events ($p16^{INK4a}$ positive cells and cell groups) was initially performed at low magnification using a 10 × objective (Plan Apochromat, NA = 0.32, specimen level pixel size 1.06 × 1.06 $um^2$). Subsequently, a more accurate analysis of positive events detected at low magnification was performed using a 20× objective (Plan Apochromat, NA = 0.6, specimen level pixel size 0.525 × 0.525 $um^2$). Image acquisition and image processing were performed using custom macros in KS400 image analysis software (version 3.0, Carl Zeiss, Germany).

After loading specimens, the stage was moved through a predefined number of 7 fields of vision in each specimen in a stepwise manner. Autofocussing was attempted at low magnification in each of these fields. The focus z-position of the field of vision with highest focus score was taken as a starting focus position for the respective specimen. In the first specimen, the lamp voltage was automatically adjusted in the field of vision with highest focus score so that the background intensity (calculated as the 95th percentile of all intensity values in the image) was between 242 and 247. The resulting voltage value was used for scanning all specimens at low magnification. The same procedure was applied for high magnification. For both magnifications an empty microscopic image was acquired fully automatically.

#### 2.4.1. Scanning

The scanning procedure is depicted schematically in Fig. 1. Low magnification scanning was restricted to the predefined circular area in which the ThinPrep T2000 device deposits the cellular material. Fields of vision (size 1378 × 1060 $um^2$) were slightly overlapping and a 'forbidden line' method [7] was used to avoid measurement of objects on the edge of a field in two adjacent fields. Every individual field of vision was autofocussed. Grabbed images were corrected for unequal illumination by dividing each camera channel pixel-wise by the corresponding empty field camera channel after which the intensity was brought back to a byte value by multiplication by 240. A fixed factor of 240 was used to enable application of fixed thresholds for subsequent segmentation (Section 2.4.3). This procedure results in white balancing of the image. Subsequently segmentation of possibly $p16^{INK4a}$ positive cells was performed as described in Section 2.4.3. Positive objects belonging to the same group of cells were connected using a region growing procedure [4]. For each detected positive event xyz coordinates were stored. Automatic low magnification scanning was continued until 50 positive events had been detected, or until the entire cellular area had been scanned. Finally, the microscope was automatically switched to high magnification for more accurate analysis.

#### 2.4.2. High magnification analysis

Possibly $p16^{INK4a}$ positive events were relocated using the xyz-coordinates acquired during low magnification scanning. After autofocussing of the object of interest, an image was grabbed (size 134 × 134 $um^2$) which was corrected for unequal illumination using the stored empty field image applying the same procedure as described above. Segmentation of possibly $p16^{INK4a}$ positive cells was performed as described in Section 2.4.3. Next, geometric and densitometric features were calculated and a discriminant function (DF) value was calculated as described in Section 2.4.4. A gallery image was constructed containing subimages of the 25 events with highest DF values encountered during scanning.

#### 2.4.3. Object segmentation

The segmentation algorithm was devised on the basis of analysis of RGB pixel values of a large number of interactively identified objects, as well as on visual inspection of the absorption characteristics of the used dyes (Fig. 2). The training set of objects, which were selected from the 10 ThinPreps in the set of train-

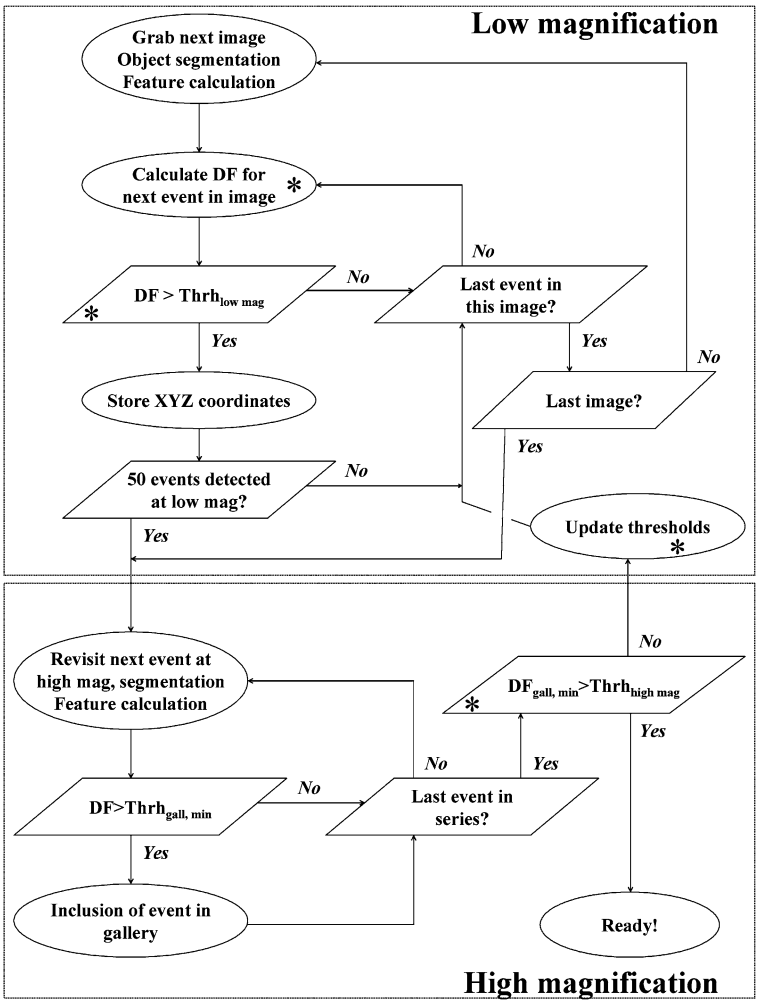


Fig. 1. Flow-chart for the automatic image analysis procedure developed in this study. The procedure may be subdivided into low magnification and high magnification parts. Asterisks indicate which steps are only performed in case of adaptive scanning.

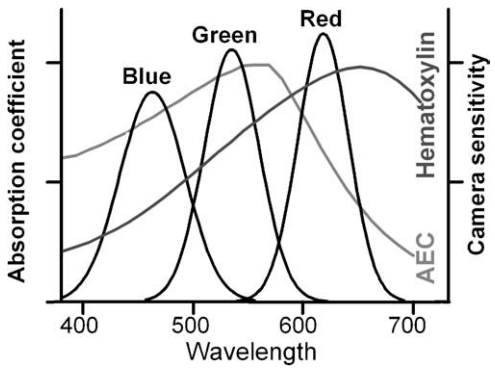


Fig. 2. Absorption curves of the dyes used to stain ThinPreps (AEC for immunocytochemical $p16^{INK4a}$ staining and hematoxylin as nuclear counterstain), together with the RGB camera sensitivity curves.

ing samples, encompassed $p16^{INK4a}$ positive staining, hematoxylin counterstain and various artifacts. By inspecting scatterplots of the RGB intensities and of ratios between pairs of RGB intensities, a set of threshold values was selected which defines the red AEC staining (data not shown). The steps of the resulting segmentation procedure for a small group of $p16^{INK4a}$ positive cells are shown in Fig. 3. From the original RGB image after shading correction (Fig. 3a), the ratio between the red intensity and green intensity values is calculated (Fig. 3b). The absorption curves in Fig. 2 disclosed that AEC is best characterized by using the combined red intensity (minimum AEC absorption) and green intensity (maximum AEC absorption). Values of the red-green intensity ratio exceeding 1.25 (Fig. 3c) were found to almost exclusively represent

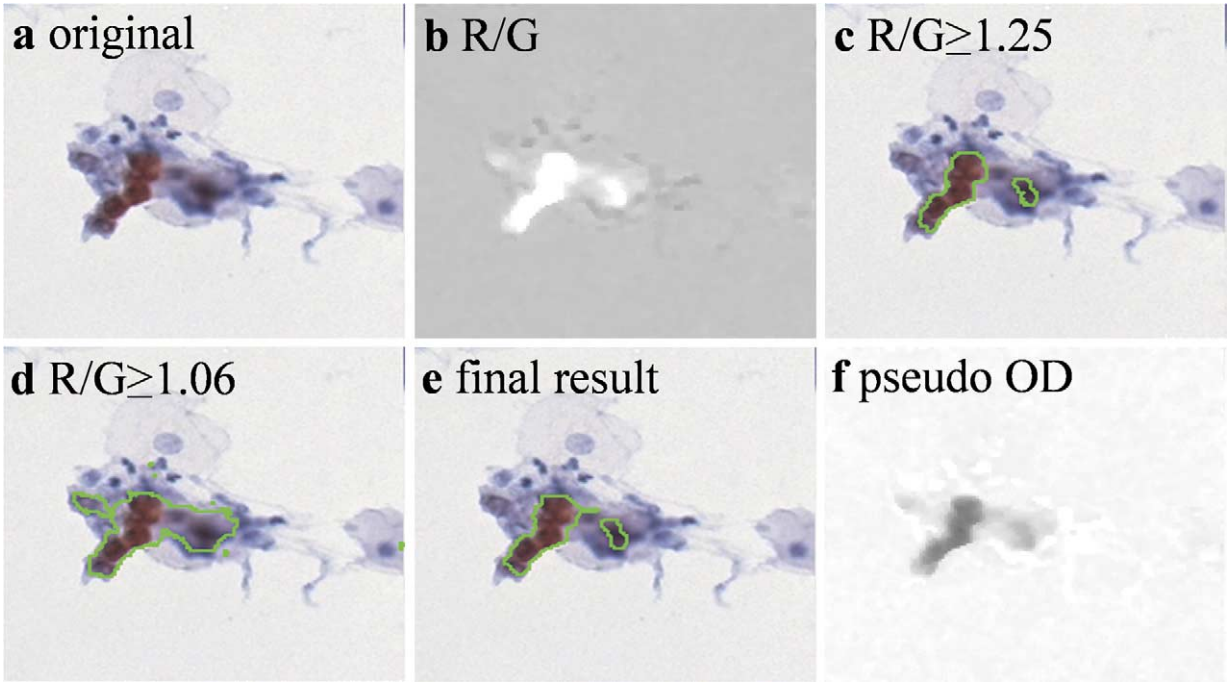


Fig. 3. Example of automated segmentation of $p16^{INK4a}$ positive staining in a small group of cells. (a) Original RGB image after shading correction; (b) R/G ratio, used to recognize the red stained positive nuclei; (c) Result of thresholding the RG ratio with threshold value of 1.25, superimposed over the original image; (d) Result of thresholding the RG ratio with the second, less strict, threshold value of 1.06, superimposed over the original image; (e) Final segmentation result; (f) Pseudo optical density image for assessment of IOD of AEC dye, calculated as G/R.

the AEC dye. Pixels with red-green intensity ratio between 1.06 and 1.25 (Fig. 3d) often represent the AEC dye, but may also include dense artifacts (e.g. dust particles), densely stained blue nuclei or white image background. Therefore, a second criterion was used to classify pixels with red-green intensity ratio between 1.06 and 1.25 in order to define $p16^{INK4a}$ positive staining. According to this second criterion, a pixel is classified as $p16^{INK4a}$ positive if it satisfies: red intensity $\geqslant$30 (grey value), blue-red intensity ratio $\leqslant$1.01 and green intensity $\leqslant$210 (grey value). A binary closing was applied to the resulting image after which small objects (area $\leqslant$20 pixels for low magnification or area $\leqslant$80 pixels for high magnification) were deleted. The final segmentation result is shown in Fig. 3e. Due to the presence of the hematoxylin counterstain, direct measurement of the optical density (OD) of AEC is not possible. Instead, a pseudo OD image for AEC was constructed by taking the negative log of the green intensity divided by the red intensity (Fig. 3f). It can be shown that this value linearly depends on the amount of AEC. The integrated OD calculated from these pseudo OD images will be called pseudo IOD.

#### 2.4.4. Discriminant function

A training set of $p16^{INK4a}$ positive events ($n = 614$) was obtained from the 10 ThinPreps in the set of training samples using the scanning procedure described above, and the images of all positive events encountered at high magnification were stored. For each detected positive event, the area, perimeter, pseudo IOD and mean, variance, skewness and kurtosis of the pseudo OD values of all object pixels were calculated and stored. Events were visually classified, blind for cytological outcome, for intensity and extensiveness of positive staining on an arbitrary scale ranging from 1 (hardly any positive staining visible) to 8 (large groups of strongly stained positive cells). To assign a score which describes the extend of positive staining on the basis of objective measurement features, a discriminant function was constructed using regression analysis.

Scatterplots of values of different features revealed that especially the pseudo IOD and the area of positive events show a strong, nonlinear relationship with visual classification. Fitting of different models showed that this relationship may be described by a logarithmic model (data not shown). Therefore, nonlinear re-

gression analysis was used to calculate a discriminant function according to the model:

$$\mathrm{DF} = b0 + b1 * \ln(\mathrm{IOD}) + b2 * \ln(\mathrm{area}).$$

Starting values for the regression parameters were obtained by first applying linear regression analysis to the log converted independents.

*2.4.5. Adaptive scanning*

For construction of a gallery image, exhaustive scanning of an entire ThinPrep may not be necessary. One might argue that once a sufficient number of $p16^{INK4a}$ positive events of sufficient diagnostic importance (as indicated e.g. by DF value) have been detected, scanning can be aborted. Especially in ThinPrep specimens containing a high density of positive events, this may improve scanning speed considerably. To study feasibility of this approach, an adaptive scanning algorithm was devised which scans a ThinPrep until a predefined number of $p16^{INK4a}$ positive events had been located with DF values exceeding a threshold value. Because more events with high DF values can be expected if the density of $p16^{INK4a}$ positive events is high, this threshold value was chosen to depend on the density of positive events. The relationship between threshold value and the density of positive events was studied using data obtained from 20 ThinPrep specimens which were also used for quantitative analysis. A scatterplot was produced of the number of positive events in a ThinPrep and the minimum DF value of events in the gallery image resulting from exhaustive scanning of the ThinPrep (Fig. 4). Non-linear regression analysis was used to study this relationship. A logarithmic function of the form:

$$Y = b0 + b1 * \ln(\text{number of positive events})$$

was found to fit the data best (data not shown). This expression may be regarded as yielding an upper limit to the threshold value required. A threshold value for adaptive scanning can therefore be derived by subtracting an offset from $Y$:

$$\begin{aligned}&\text{High Magnification Threshold}\\ &\quad = b0 + b1 * \ln(\text{number of positive events})\\ &\qquad - \text{offset}.\end{aligned}$$

Higher values for 'offset' result in faster scanning (25 objects fulfilling the threshold criterion are easier

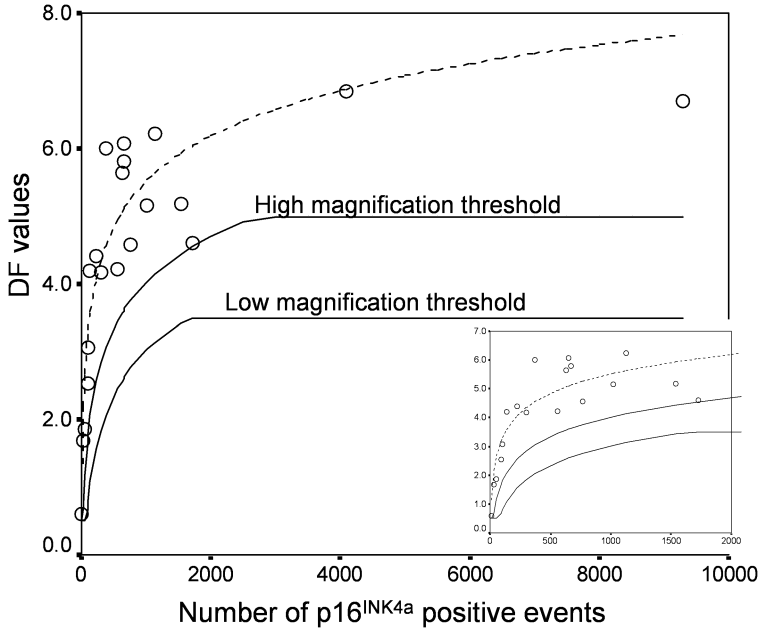


Fig. 4. Scatterplot of the number of $p16^{INK4a}$ positive events versus the minimum DF value (calculated at high magnification) of all events included in the gallery image for 20 ThinPrep specimens after exhaustive scanning. The dashed line shows the logarithmic relationship resulting from nonlinear regression analysis. Two negative offset values of $-2.5$ and $-1.5$ are used to derive thresholds for low and high magnification, respectively. The thresholds are maximized at values 3.5 and 5, respectively. The inset shows an enlargement of part of the same figure (number of pos. events $< 2000$).

found) but also result in inclusion in the gallery image of events which may be of less diagnostic importance. In the present study, we chose an offset value of 1.5. After 25 $p16^{INK4a}$ positive events were identified with DF values exceeding the high magnification threshold, images of these events were placed in a gallery image and the scanning was aborted. The minimum value for the high magnification threshold was set to 0.5; also the threshold was limited to a maximum value of 5.0. In this way, as long as the gallery image contains any positive events with DF $< 0.5$ scanning will never be aborted, whereas as soon as 25 events have been identified with DF $> 5.0$ scanning will always be aborted, independently of the density of positive events.

To further increase efficiency of the scanning, a low magnification threshold was defined analogous to the high magnification threshold. Events detected during low magnification scanning with DF value under the low magnification threshold were not analyzed at high magnification. In the present study, the offset used to calculate the low magnification threshold was 2.5. The minimum value for the low magnification threshold was set to 0.5; also this threshold was limited to a maximum value of 3.5. In this way, $p16^{INK4a}$ positive events with DF value smaller than 0.5 where never analyzed at high magnification, whereas positive events with DF value exceeding 3.5 were always analyzed at high magnification. To illustrate the effect of adaptive scanning, one HSIL was measured containing a high density of $p16^{INK4a}$ positive events. This ThinPrep was selected from the set used for quantitative analysis.

#### 2.4.6. Quantitative analysis

Cervical smears of 10 cases WNL and 10 HSIL were scanned exhaustively. All geometric and densitometric feature values calculated from $p16^{INK4a}$ positive events at high magnification, as well as the corresponding DF values were stored in a database for subsequent analysis. To describe the $p16^{INK4a}$ status of a specimen, a number of specimen level parameters may be of interest. For this purpose, the mean DF value over all $p16^{INK4a}$ positive events was calculated as well as the average pseudo IOD per positive event.

### 2.5. Statistics

Statistical analyses were performed using SPSS for Windows (version 11.0.1; SPSS Inc., Chicago, IL). Linear and nonlinear regression analysis was used to construct classifiers. Quantitative analysis of ThinPreps WNL and ThinPreps containing HSIL was evaluated using the non parametric Mann–Whitney $U$ test.

## 3. Results

### 3.1. Discriminant functions

Nonlinear regression analysis was used to construct a discriminant function for assigning a score which describes the extend of $p16^{INK4a}$ positive staining on the basis of area and pseudo IOD. All 614 objects in the training set were visually scored for extent and intensity of positive staining. This resulted in 42 objects in class 1, 97 in class 2, 221 in class 3, 9 in class 4, 122 in class 5, 61 in class 6, 41 in class 7 and 21 in class 8. The resulting DF is defined as:

$$\mathrm{DF} = -30.3 + 2.20 * \ln(\mathrm{IOD}) - 1.76 * \ln(\mathrm{area}).$$

### 3.2. Reproducibility on cell level

Three ThinPreps were selected which contained only a limited number of visually identified $p16^{INK4a}$ positive cells. Two ThinPreps which were cytologically classified as WNL contained two and three $p16^{INK4a}$ positive events, respectively. Total cell density of these ThinPreps was visually estimated to be 7360 and 3820 cells, respectively. A third ThinPrep, cytologically diagnosed as 'atypical squamous cells of undetermined significance' (ASC-US), contained six $p16^{INK4a}$ positive events in an estimated total of 3610 squamous cells. Each of these ThinPreps was measured 10 times. On average, measurement of a specimen took 45 minutes. Visual inspection of the resulting gallery images revealed that all visually identified positive events were present in galleries of all 10 measurement runs (i.e. sensitivity of detection for these events is 100%). Subimages of the $p16^{INK4a}$ positive events as well as the DF values resulting from these measurements are shown in Fig. 5 ((a,b) ThinPreps WNL, (c) ThinPrep containing ASC-US). As can be seen, reproducibility is high with average coefficient of variation (CV) of DF values of different measurement runs 0.7% (max 1.1%).

### 3.3. Adaptive scanning

The adaptive scanning algorithm scans a specimen until 25 events with DF value exceeding a threshold value have been identified at high magnification. The threshold was derived using nonlinear regression analysis of the relationship between the number of positive events in a ThinPrep and the minimum DF of

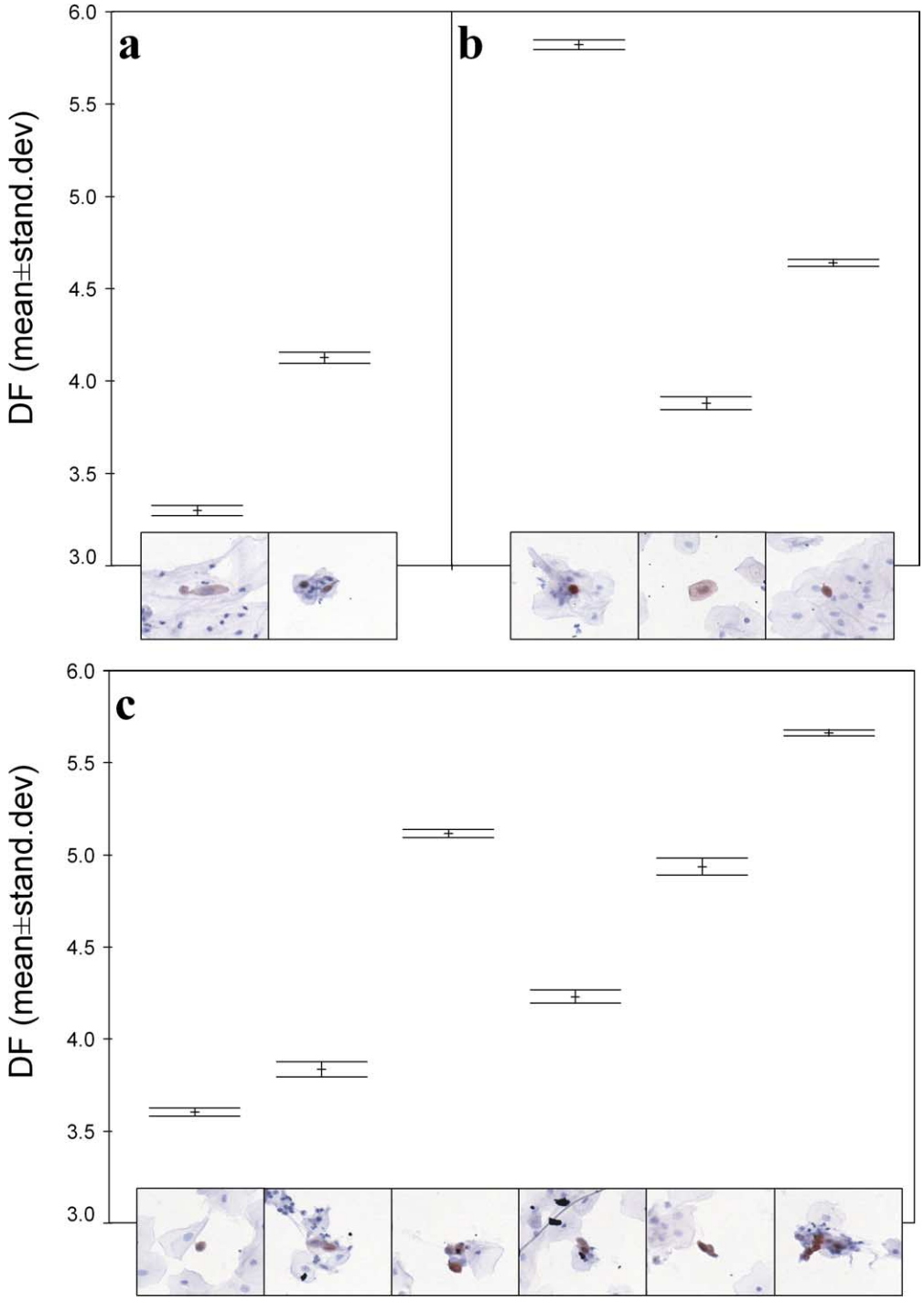


Fig. 5. Microscopic images of $p16^{INK4a}$ positive events in ThinPreps cytologically diagnosed as 'within normal limits' – (a) 2 events and (b) 3 events; and ASC-US – (c) 6 events. The corresponding discriminant function values (mean and standard deviation over 10 repeated measurements) are shown above the images. DF values are calculated over all positive cells/cell groups in an image.

events included in the gallery image (Fig. 4, dashed line). An offset value is subtracted from the resulting regression function to tune the adaptive scanning. In this study a value of 1.5 was used as offset. The threshold depended logarithmically on the total number of positive events in the ThinPrep:

$$\text{High Magnification Threshold} = -1.17 + 0.969 * \ln(\text{number of positive events}) - 1.5.$$

This threshold was not allowed to exceed a value of 5.0. The expected number of positive events in the entire ThinPrep is calculated during the scan by extrapolation:

$$\text{number of positive events} = 328 * \frac{\text{num. positive events found}}{\text{num. fields measured}}$$

(328 is the number of fields of vision in a ThinPrep for the instrumental setup used in this study). Figure 4 shows the high magnification threshold as function of the number of positive events.

Based on the result of nonlinear regression analysis, a second threshold may be defined for low magnification scanning, using a different offset value. In the present study, an offset of 2.5 was used for low magnification scanning:

$$\text{Low Magnification Threshold} = -1.17 + 0.969 * \ln(\text{number of positive events}) - 2.5.$$

The low magnification threshold was not allowed to exceed a value of 3.5 (Fig. 4). Positive events detected during low magnification scanning were only analyzed at high magnification if the DF value calculated at low magnification exceeded the low magnification threshold value. Because the calculation of DF at low magnification is less accurate compared to high magnification, the threshold was more conservative (i.e. the offset used is larger).

To illustrate the effect of adaptive scanning, one ThinPrep containing HSIL was scanned in which 305 $p16^{INK4a}$ positive cells were manually counted. Exhaustive scanning resulted in scanning of 328 fields of view. The gallery image resulting from exhaustive scanning is shown in Fig. 6c. Adaptive scanning stopped after analyzing 89 of all 328 possible fields of view (27%), reducing scanning time almost fourfold. Figure 6a shows all positive events analyzed during adaptive scanning, and Fig. 6b shows the gallery image resulting from this scan. In Fig. 6a, asterisks show which 25 events were selected for the gallery image in Fig. 6b. In Fig. 6c asterisks show which events selected in the exhaustive scan had also been detected in the adaptive scan. Exhaustive scanning resulted in detection of more intensely stained $p16^{INK4a}$ positive events.

### 3.4. Quantitative analysis

Results of quantitative analysis of scanning 10 ThinPreps WNL and 10 HSIL are shown in Fig. 7. As can be seen, a clear distinction can be made between WNL cases and HSIL cases based on mean DF value of $p16^{INK4a}$ positive events ($p < 0.001$), with only one HSIL in the range of the ThinPreps WNL. More overlap exists between the two groups of ThinPreps when considering the average pseudo IOD per positive object ($p < 0.01$).

## 4. Discussion

Large scale molecular analysis techniques have led to increased identification of promising biomarkers for a number of applications. For instance, $p16^{INK4a}$ has been described as a promising surrogate marker for dysplasia in cervical cytology [10]. If sufficiently sensitive and specific, the use of such a marker may decrease the subjectivity and inaccuracy of traditional cervical screening based on cytomorphology. However, contradicting data have been published concerning the diagnostic meaning of $p16^{INK4a}$ positivity in cervical smears. Most likely, this contradiction is the result of the lack of a uniform scoring method. Authors even go so far as to include traditional cytomorphologic criteria in the scoring system [25], undermining the possible gain in objectivity that a reliable marker offers. As was shown previously [9,14,16] visual assessment of (immunostained) specimens may be very inaccurate. It has been shown that even large, strongly stained cell clusters may be missed entirely by an experienced pathologist when screening a large series of specimens. Fully automated assessment of biomarker signals may facilitate uniform and objective scoring. Using automated image analysis, observer bias is re-

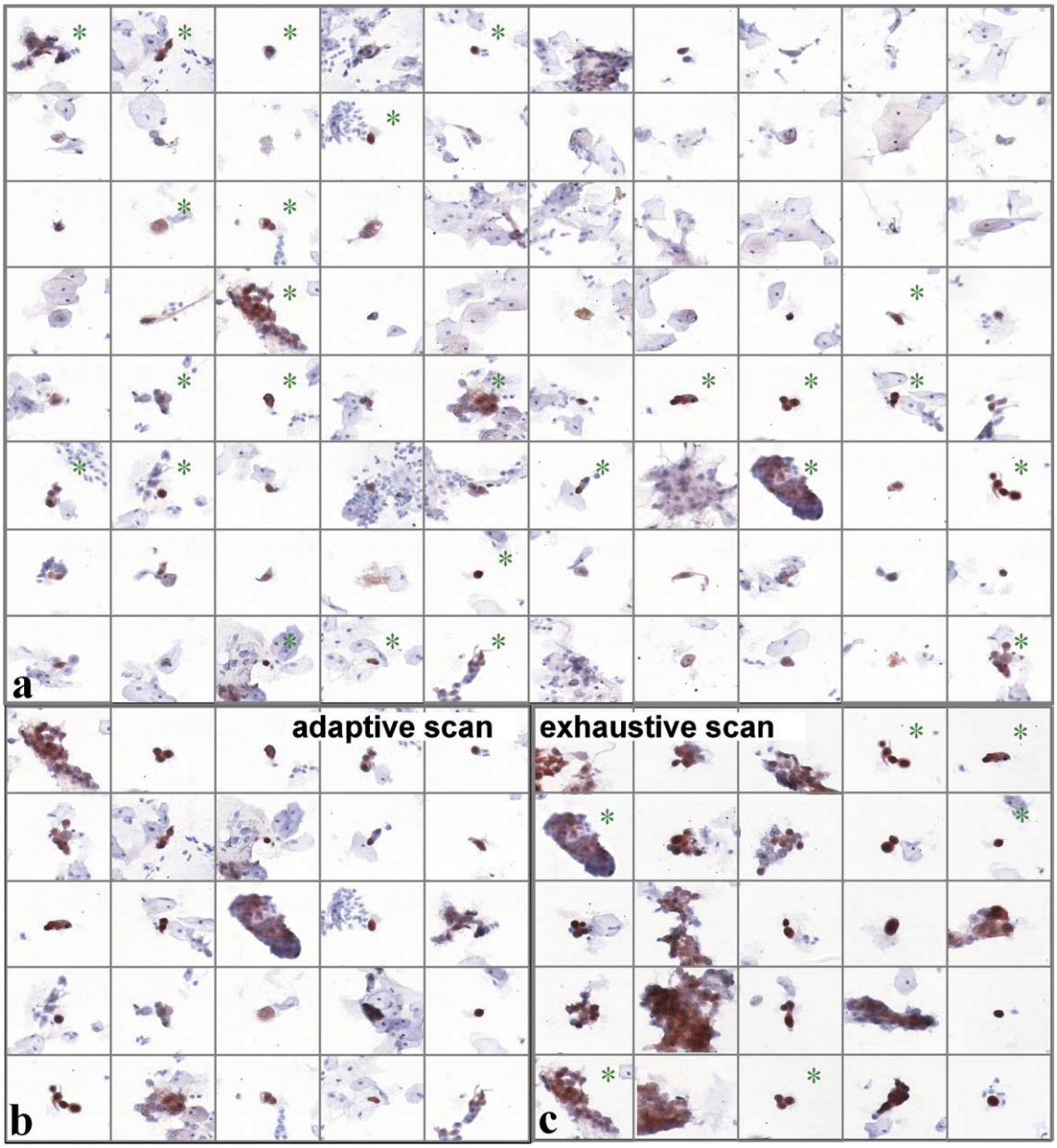


Fig. 6. Scanning results of an example ThinPrep containing HSIL, to illustrate the effect of the adaptive scanning procedure. (a) Overview of all $p16^{INK4a}$ positive events detected during adaptive scanning; asterisks indicate which events were selected for inclusion in the final gallery image; (b) Gallery image resulting from adaptive scanning; (c) Gallery image resulting from scanning the same ThinPrep exhaustively. Asterisks indicate which events were also present in the gallery image resulting from adaptive scanning (see (b)).

duced because of the use of objective measurements, and reproducibility is high [20].

In the present study, a semi-rare event detection system was described which uses statistical classifiers to rate the importance of detected positive cells and cell clusters. Similar to traditional rare event detection, the system was capable of locating 'true' rare events. However, traditional rare event detection systems are not devised to handle specimens with a relatively high density of positive events. If the percentage of false positive cells is high, classic rare event detection necessitates visual inspection of large numbers of cells [20]. The system presented here is capable of dealing with such specimens, rendering it useful

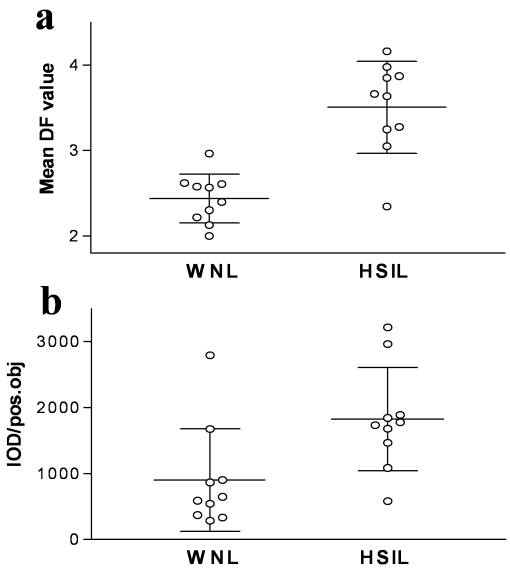


Fig. 7. Specimen level measurement results of quantitative $p16^{INK4a}$ analysis of 10 ThinPreps WNL and 10 ThinPreps containing HSIL. (a) Mean DF value; (b) Average IOD per positive event. Shown are individual values (open circles) as well as mean and standard deviation (horizontal bars).

for analysis of biomarkers which are either not fully specific for abnormal cells or biomarkers which show nonspecific background staining. Two previous studies [13,15] employing cell features for rare event detection aimed at differentiating between true and false positive events, rather than locating the most informative positive events. In cases with many positive events, the system described in the present paper focuses on positive cells and cell clusters exhibiting the most extensive and/or intense staining pattern. The underlying assumption is that positive events with more extensive/intense staining are in general of higher diagnostic importance, whereas most non-specific staining is limited in intensity and, to a lesser degree, in extent. For example, $p16^{INK4a}$ staining of normal cells is either limited in extent (metaplastic cells [22]) or is weak (non-dysplastic epithelial cells [25]). Thus, offering images of only the highest scoring events to a pathologist or cytotechnician will decrease problems with non-specific staining and at the same time increase efficiency. The ability to handle specimens with a high density of positive events is critical for such applications, as we aim at a uniform scoring method for all specimens, without prior knowledge of either cytological diagnosis or $p16^{INK4a}$ status.

The method described here will be a valuable tool for scientific research projects studying the diagnostic or prognostic significance of biomarkers. Speed of analysis is often not of major importance in scientific research, and may be sacrificed for the gain in objectivity and sensitivity. Whether the scanning method described here will be useful in routine diagnostics will mainly depend upon future availability of sensitive and reproducible biomarkers. Scanning speed is of higher importance in routine diagnostics [11,16,20], although objectivity and sensitivity are critical factors as well. Currently, scanning a specimen takes approximately 45 minutes, whereas human visual scanning takes only 5 minutes. However, the scanning method as described was not optimized for speed. In the present study we investigated an adaptive scanning strategy as a possible improvement to increase scanning speed. Compared to exhaustive specimen scanning, adaptive scanning resulted in detection of less extensively stained events. However, in the HSIL case presented here the gallery image reflected the $p16^{INK4a}$ status adequately. Many intensely stained $p16^{INK4a}$ positive events were detected and presented in the gallery image, with scanning time reduced almost fourfold. The adaptive scanning was conducted in such a way that for specimens with low density of positive events, a large part or even the entire cellular area will be scanned whereas for specimens with high density of positive events only a smaller part of the specimen is analyzed. In this way, adaptive scanning may improve scanning efficiency considerably without compromising on detection sensitivity.

In addition, scanning speed may further be increased by applying techniques originating from so called virtual microscopy [12], enabling high speed – high throughput analysis. This comprises digitizing an entire specimen after which image analysis may be performed on stored images 'off line'. At present, such systems are capable of digitizing an entire specimen in less than 10 minutes. As it can be expected that in the near future many routine pathology laboratories will witness the introduction of virtual microscopy, objective scoring methods for ICC such as the one described here will probably become available for a much larger audience than is presently the case.

The AxioCam MRc camera used in the present study is a single chip camera, producing 24 bit color images through interpolation of image data acquired using a so-called Bayer color filter pattern. A single chip camera allows high speed acquisition of high resolution images at an affordable price. Alternatively, images may be acquired using three chip cameras, or by application of different color filters in the light path. Three chip high resolution cameras are very expen-

sive, whereas use of separate color filters considerably slows down image acquisition. The drawbacks of using a single chip camera are loss of resolution [23] and the occurrence of false colors along object edges, caused by color interpolation ('demosaicing') [18]. The latter problem may be reduced by employing more complicated demosaicing algorithms, at the cost of operation speed [18]. In the present study, excellent results were obtained with the standard demosaicing algorithm of the AxioCam MRc single chip camera. Application of a three CCD camera may further improve object segmentation and classification.

To increase specificity of the detection of atypical cells, multiple markers may be used simultaneously [2]. For instance, $p16^{INK4a}$ immunocytochemistry may be combined with MIB-1 proliferation marker [22] or with ISH staining of oncogenic HPV types [8]. The semi-rare event detection presented in this study will be applicable for multiple markers, provided the segmentation procedure and discriminant functions have been adjusted to the specific staining patterns. Segmentation of such signals is challenging when transmitted light microscopy is used together with an RGB color camera. Segmentation may be facilitated by using appropriate color models which are better suited for transmitted light microscopy, such as the hue saturation density model [26] or color deconvolution [21]. In case of HPV ISH, texture of the positive signal is of importance to distinguish between episomal and integrated virus [8]. Texture features may be incorporated in the discriminant function to account for this.

In conclusion, the scanning procedure in the present study is highly reproducible in detecting positive cells and in calculating the discriminant score. Because only the diagnostically most interesting cells and cell groups are presented to the pathologist in a gallery image, screening may become less tedious [20]. The pathologist may review selected positive cells either on the computer gallery image or by automatic relocation under the microscope. In addition, quantitative features may be calculated describing the immunocytochemical staining patterns in an objective and reproducible manner. In the present study, the mean DF value over all $p16^{INK4a}$ positive events and the average pseudo IOD per positive event were analyzed for distinction between specimens WNL and specimens containing HSIL. Especially the mean DF value was found to discriminate between these two groups. Further studies may identify other features which are relevant for the distinction between different (pre)malignant stages. Assessment of such objective features, combined with suitable markers, will probably exceed the possibilities of human visual analysis.

## Acknowledgements

The authors wish to thank Monique Derikx for immunocytochemical stainings.

## References

[1] B.T. Andersen, C. Pfeifer, E.R. Clausen et al., Improvement of the CINtec(TM) p16(INK4a) assay for identifying cervical high-grade lesions in cytology, *Cytopathology* **16**(Supl. 2) (2006), 57.

[2] K. Astbury, C.M. Martin, M. Ring et al., Future molecular aspects of cervical cytology, *Curr. Diagn. Pathol.* **12** (2006), 104–113.

[3] J. Bulten, I. van der Avoort, W.J. Melchers et al., p14ARF and p16INK4A, two products of the same gene, are differently expressed in cervical intraepithelial neoplasia, *Gynecol. Oncol.* **101** (2006), 487–494.

[4] R.C. Gonzalez and R.E. Woods, *Digital Image Processing*, Addison-Wesley Publishing Company, Reading, Massachusetts, 1992.

[5] M.P. Grobusch, T. Hanscheid, B. Kramer et al., Sensitivity of hemozoin detection by automated flow cytometry in non- and semi-immune malaria patients, *Cytometry* **55B** (2003), 46–51.

[6] H.J. Gross, B. Verwer, D. Houck et al., Model study detecting breast cancer cells in peripheral blood mononuclear cells at frequencies as low as 10(-7), *Proc. Natl. Acad. Sci. USA* **92** (1995), 537–541.

[7] H.J. Gundersen, T.F. Bendtsen, L. Korbo et al., Some new, simple and efficient stereological methods and their use in pathological research and diagnosis, *APMIS* **96** (1988), 379–394.

[8] A.N. Kalof, M.F. Evans, L. Simmons-Arnold et al., p16INK4A immunoexpression and HPV in situ hybridization signal patterns: potential markers of high-grade cervical intraepithelial neoplasia, *Am. J. Surg. Pathol.* **29** (2005), 674–679.

[9] R. Klaes, A. Benner, T. Friedrich et al., p16INK4a immunohistochemistry improves interobserver agreement in the diagnosis of cervical intraepithelial neoplasia, *Am. J. Surg. Pathol.* **26** (2002), 1389–1399.

[10] R. Klaes, T. Friedrich, D. Spitkovsky et al., Overexpression of p16(INK4A) as a specific marker for dysplastic and neoplastic epithelial cells of the cervix uteri, *Int. J. Cancer* **92** (2001), 276–284.

[11] S.K. Kraeft, R. Sutherland, L. Gravelin et al., Detection and analysis of cancer cells in blood and bone marrow using a rare event imaging system, *Clin. Cancer Res.* **6** (2000), 434–442.

[12] R.K. Kumar, G.M. Velan, S.O. Korell et al., Virtual microscopy for learning and assessment in pathology, *J. Pathol.* **204** (2004), 613–618.

[13] A. Ladanyi, A.C. Sher, A. Herlitz et al., Automated detection of immunofluorescently labeled cytomegalovirus-infected cells in isolated peripheral blood leukocytes using decision tree analysis, *Cytometry* **58A** (2004), 147–156.

[14] F. McQueen and E. Duvall, Using a quality control approach to define an 'adequately cellular' liquid-based cervical cytology specimen, *Cytopathology* **17** (2006), 168–174.

[15] W.E. Mesker, F.S. Doekhie, J. Vrolijk et al., Automated analysis of multiple sections for the detection of occult cells in lymph nodes, *Clin. Cancer Res.* **9** (2003), 4826–4834.

[16] W.E. Mesker, H. Torrenga, W.C. Sloos et al., Supervised automated microscopy increases sensitivity and efficiency of detection of sentinel node micrometastases in patients with breast cancer, *J. Clin. Pathol.* **57** (2004), 960–964.

[17] W.E. Mesker, J.M. van de Burg, P.S. Oud et al., Detection of immunocytochemically stained rare events using image analysis, *Cytometry* **17** (1994), 209–215.

[18] J. Mukherjee, M.K. Lang and S.K. Mitra, Demosaicing of images obtained from single-chip imaging sensors in YUV color space, *Patt. Recogn. Let.* **26** (2005), 985–997.

[19] E.R. Nijhuis, N. Reesink-Peters, G.B.A. Wisman et al., An overview of innovative techniques to improve cervical cancer screening, *Cell. Oncol.* **28** (2006), 233–246.

[20] J.C. Oosterwijk, C.F. Knepfle, W.E. Mesker et al., Strategies for rare-event detection: an approach for automated fetal cell detection in maternal blood, *Am. J. Hum. Genet.* **63** (1998), 1783–1792.

[21] A.C. Ruifrok and D.A. Johnston, Quantification of histochemical staining by color deconvolution, *Anal. Quant. Cytol. Histol.* **23** (2001), 291–299.

[22] S. Sahebali, C.E. Depuydt, G.A. Boulet et al., Immunocytochemistry in liquid-based cervical cytology: analysis of clinical use following a cross-sectional study, *Int. J. Cancer* **118** (2006), 1254–1260.

[23] H.J. Tanke, R.J. Florijn, J. Wiegant et al., CCD microscopy and image analysis of cells and chromosomes stained by fluorescence in situ hybridization, *Histochem. J.* **27** (1995), 4–14.

[24] H.J. Tanke, J.C. Oosterwijk, W.E. Mesker et al., Detection of 'rare event' fetal erythroblasts in maternal blood using automated microscopy, *Early Hum. Dev.* **47**(Suppl.) (1996), S89–S93.

[25] M.J. Trunk, G. Dallenbach-Hellweg, R. Ridder et al., Morphologic characteristics of p16INK4a-positive cells in cervical cytology samples, *Acta Cytol.* **48** (2004), 771–782.

[26] J.A. van der Laak, M.M. Pahlplatz, A.G. Hanselaar et al., Hue-saturation-density (HSD) model for stain recognition in digital images from transmitted light microscopy, *Cytometry* **39** (2000), 275–284.

[27] M. von Knebel-Doeberitz, New markers for cervical dysplasia to visualise the genomic chaos created by aberrant oncogenic papillomavirus infections, *Eur. J. Cancer* **38** (2002), 2229–2242.