Ruifrok AC, Johnston DA. Quantification of histochemical staining by color deconvolution. Anal Quant Cytol Histol 23: 291–299

Article *in* Analytical and quantitative cytology and histology / the International Academy of Cytology [and] American Society of Cytology · September 2001

Source: PubMed

CITATIONS
492

READS
3,766

2 authors:

Arnout Ruifrok
Netherlands Forensic Institute
66 PUBLICATIONS **3,818** CITATIONS

SEE PROFILE

Dennis A Johnston
Baylor University
298 PUBLICATIONS **15,405** CITATIONS

SEE PROFILE

Comparison of quantification of histochemical staining by hue-saturation-intensity (HSI) transformation and color-deconvolution.

Arnout C. Ruifrok, Ph.D., Ruth L. Katz, M.D., Dennis A. Johnston, Ph.D.

From the departments of Pathology (ACR, RLK) and Biomathematics (DAJ), University of Texas M.D. Anderson Cancer Center, Houston, TX, U.S.A.

Address reprint requests to: Arnout C. Ruifrok, Department of Pathology, Box 53, M.D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX, 77030
Phone: (713) 792 4088
Fax: (713) 792 3184
E-mail: arnout@mdanderson.org

Ruifrok AC, Katz RL, Johnston DA. Comparison of quantification of histochemical staining by hue-saturation-intensity (HSI) transformation and color-deconvolution.
Appl Immunohistochem Mol Morphol. 2003 Mar;11(1):85-91.

Abstract
We tested a recently developed flexible method of separation and quantification of immunohistochemical staining by means of color image analysis. An algorithm was recently developed to deconvolve the color information acquired with RGB cameras, to calculate the contribution of each of the applied stains, based on the stain-specific RGB absorption. The algorithm was tested using a set of lung-tumor samples labeled for the detection of Ki-67, an antigen expressed in proliferating cells, covering a wide range of staining levels. Quantification of the labeling was compared with HSI- based segmentation and manual analysis of the same samples. The recently developed deconvolution method performed significantly better than the HSI based system when compared to manual counting as gold standard. The deconvolution system showed significantly reduced variability in the LI determination, especially of highly labeled control samples. This resulted in significant increase in sensitivity of classification of samples with increased KI-67 labeling without changing the specificity, when compared to the HSI based method.

Keywords: image processing, computer-assisted; color separation, color deconvolution, immunohistochemistry.

Introduction:

Most imaging systems use transformation of the red-green-blue (RGB) image information to hue-saturation-intensity (HSI) or color specific representation for color segmentation (1-5). A problem with these systems is that they classify pixels as belonging to a certain color based on the relative contribution of each stain, ignoring that more than one stain may have contributed to the final color. During measurement of Ki-67 staining in lung tumor slides in our laboratory, the observation was made that dark-stained slides seemed to be under-counted. We developed a color-deconvolution system that can separate the contribution of up to three stains to the final color (6). Separation of the contribution of different stains is especially important in cases where the combination of stains results in a very dark staining, a situation causing classification problems using the HSI system. The color-deconvolution method is based on orthonormal transformation of the original RGB image, depending on user-determined color information about the stains used. The method provides the possibility to determine staining densities, even in areas where multiple stains are co-localized, making it possible not only to determine surface area and overall absorption in areas with a specific color, but also to determine densities and ratios of densities of stains in each area. After the color-deconvolution, images can be reconstructed for each stain separately, and be used for densitometry and texture analysis for each stain, using standard imaging methods.

To test the performance of the color-deconvolution system, we compared results of Ki-67 labeling with MIB-1 antibody in lung tumor tissue analyzed with the deconvolution method with results from analysis using HSI-based methods. We analyzed MIB-1 stained slides from lung-cancer specimens and the negative and positive control slides (breast tumor tissue) in multiple staining batches, and compared the two imaging based approaches with manual analysis. Parameters measured using the two systems are labeling index (LI), defined as the fraction of MIB-1 positive area divided by the fraction of nuclear (MIB-1+hematoxylin) area, the mean optical density (MOD) of the MIB-1 positive area, and the 'QuickScore', the product of the LI and the MOD. These measurements are compared to manual scoring of the slides, which was considered the 'gold standard'. Based on literature data (7), the LI is currently being used for estimation of the fraction of proliferating cells in the tissues under study in our laboratory. The differences in result between HSI based and color-deconvolution based results are illustrated with a comparison of foreground selection of diaminobenzidine (DAB) stained nuclei in positive control tissue.

Methods:

Specimens

Sections from breast tumor and lung tumor specimens were stained for KI-67 expression with MIB-1 antibody (DAKO, Carpinteria, California, USA), and DAB chromogen (Biogenics, Napa, California, U.S.A.). Counterstaining was performed with Mayer's Hematoxylin (Richard Allan, Kalamazoo, Michigan, USA) for 3-5 min. The samples were stained in 7 different batches, 6 batches with malignant lung tumors, and 1 batch with benign lung tumors, each containing multiple lung samples, one negative, and one positive control slide (breast tumor tissue with known Ki-67 labeling). Initial analysis was performed using a HSI based method. Of each of the 6 lung tumor batches two low

scoring samples, two intermediate scoring samples, and two high scoring samples, together with 4 samples that were 'flagged' during HSI analysis as unreliable, and 10 benign samples, were analyzed further with the deconvolution method, and compared to manual analysis. Manual counting of Ki-67 positive nuclei in these 50 samples was done blinded on number -coded slides.

Image acquisition

A Leica DMLB (Leica Microsystems Inc. Deerfield, Illinois, USA) microscope was equipped with a Hamamatsu C5810 chilled 3-chip color CCD camera (Hamamatsu, Bridgewater, New Jersey, USA), interfaced with an IBM computer (International Business Machines Corporation, Armonk, New York, USA) equipped with a Matrox Meteor digitizer board (Matrox Electronic Systems Ltd., Dorval, Quebec, Canada). Light and camera settings were standardized, resulting in average background values of 20±5 (mean ± standard deviation; scale 0-255 from white to black) for the red, green and blue channels. Linearity of the image-acquisition setup was tested using a stepped neutral density filter, and found to be linear with light intensity for all three colors within 2%, over the whole dynamic range of the camera, (correlation coefficients of the OD with grayscale values for red, green and blue: $R>0.996$). The images were captured with 20 X objective lenses.

Image processing

Images were analyzed based on HSI transformed images using SAMBA IPS software version 4.22 (Samba Technologies, Meylan, France). Areas of interest were selected manually using interactive cut-and-paste techniques.

For color-deconvolution, the 24-bit RGB images were transferred to a Macintosh G4 (Apple Computer, Cupertino, California, USA) and processed and analyzed using NIH image version 1.62, developed at the National Institutes of Health (NIH) and available on the internet from http://rsb.info.nih.gov/nih-image/. Custom macros were written for background correction and transformation from intensity to optical density (OD), to determine the color-vectors for the different stains, for calculation of the color-deconvolution matrix, and for the actual color-deconvolution of the images, according to the methods described in (6). The stored image of an empty field was used for determination of the light entering at each pixel ($I_{o,c}$), implicitly correcting for unequal illumination by subtraction of the background optical density.

Statistical analysis:

Data are presented as mean ± standard deviation. Normalization of results from each batch was performed assuming that the negative controls slide included in the same batch had actually no (0%) Ki-67 labeling, and that the positive controls had 100% Ki-67 positive labeling. The correlation of the results of automated analysis with manual counting was compared using the analysis of covariance. This was accomplished using SPSS (SPSS, Inc, Chicago, USA) and the custom program AOC (www.odin.mdacc.tmc.edu). This procedure follows the general plan described by Zar (8).

Results:
The first test we did was to compare the measurement results for the negative and positive control samples included in each of the 7 staining batches. The results are presented in table (1). As can be seen, considerable variability was observed in the measurements of LI, MOD and quickscore using the HSI based system, with a coefficient of variation (CV) of almost 50% for the LI of the positive control slides.

The data also clearly show that variability in LI and quickscore is significantly lower for the deconvolution than the HSI based method, with a CV of only 1% for the positive controls analyzed with the deconvolution system. Only the MOD for the negative controls shows a high variability with the deconvolution method. This high variability in MOD is easily explained by the very small positive areas in the negative controls, caused by small artifacts with highly variable OD. If areas smaller than the minimum area of a nucleus are excluded, no measurable positive areas are found in the negative controls. These results means that the HSI data suggest high variability between staining batches, and normalization of experimental data to the results of a negative and a positive control slides will result in considerable changes of values. However, the deconvolution method shows low variability, and limited changes as a result from normalization to negative and positive control slides.

To compare the performance of the two image analysis approaches with the overall pathologists impression, we analyzed the imaging results of 50 lung tumor samples covering the range from low Ki-67 positivity to high Ki-67 positivity. The same samples were counted in a blinded way by the pathologist, analyzed using the HSI based system, and analyzed using the deconvolution system. The manual counting was considered the 'gold standard' for the comparison. Figure (1) shows the correlation between manual count and the LI (not normalized; A and normalized B) using the HSI based and the deconvolution method. The normalization resulted in values less than 0% and more than 100% labeling compared to the controls for the HSI based data. Table (2) present the squared correlation coefficients (r^2) for the HSI method and the DECON method for all three parameters, with and without normalization of the data sets based on the negative and positive controls included in each staining batch.

It is clear that using the HSI data the highest correlation with manual counting is found for the normalized LI measurements, as might be expected based on the LI definition used for counting (relative area positive).
Using the deconvolution method, LI as well as quickscore give a high correlation with the manual analysis, while normalization did not improve these results. The correlation of LI and quickscore with manual counting was significantly higher for the deconvolution method than for the HSI based method ($p<0.05$) without normalization. After normalization of the experimental data to the included controls, the deconvolution method still performed better, although not significantly so for the LI ($p=0.19$).
The correlation between HSI based and deconvolution based analysis resulted in an r^2-0.58 for not normalized and 0.74 for normalized LI data somewhat but not significantly

higher after normalization, illustrating the improvement by normalization between batches, especially when the HSI based method is used.

The above data show that the deconvolution method performs significantly better than the HSI based analysis for Ki-67 labeling determination, especially when high levels of labeling can be expected (e.g. positive control slides).

Discussion:

The above data show that deconvolution method performs better than the HSI based method in determination of the Ki-67 labeling index, with manual counting as gold standard. This is especially clear for dark stained samples like positive controls, as apparent from the lower CV of the positive control values when the deconvolution method is used. What may be causing this difference in variability between the HSI method and the deconvolution method? An explanation for the large error rate in the positive controls may be the saturation of staining, causing mis-classification of the very dark areas. The HSI method excludes very dark nuclei as blue instead of brown; the HSI-classification is very sensitive to hue settings for the darkest areas, causing a large variability in the color determination. The deconvolution method does not suffer from this problem; very dark areas are contributing to the brown as well as the blue signal.The comparison between the two methods is illustrated in figure (2), showing the results of an image of a darkly stained area of a positive control breast tissue slide in the top panel (a). The panels below show the results of the segmentation using the HSI method (left panels, b and c), and the deconvolution method (right panels, d and e). The bottom panel shows the areas that were classified as DAB negative by the HSI method, but as DAB positive by the deconvolution method. It is clear from this example that a significant number of nuclei falsely classified as DAB negative using the HSI method are correctly classified as DAB positive using the deconvolution method.
The errors in LI determination of control slides resulted in normalized values for the HSI method that were negative or greater than 100 % for the lung tumor samples with extreme high and low Ki-67 LI. Normalized values below 0 or above 100% were not observed using the deconvolution method.
However, when comparing the two methods with respect to Ki-67 labeling analysis performance maybe a more important question is: What are the results if the HSI based and deconvolution methods are used in a setting where the most important decision would be based on a 'cut-off' value for determination of low and high Ki-67 labeling? How do the differences in degree of error between the two methods influence the determination of high Ki-67 labeling related to malignant cancer risk? To test that, we determined the sensitivity and specificity of classification of the lung cancer samples as 'increased Ki-67 labeling compared to benign' as a function of the critical value for cut-off applied. We determined sensitivity and specificity of classification as 'high Ki-67' using the mean value plus 1 to 4 times SD of the benign samples for each method as cut-off (figure 3). As expected, sensitivity tends to decrease and specificity tends to increase with increasing critical value for both methods. It is also clear that especially the

sensitivity for detection of increased levels of Ki-67 labeling is higher for the deconvolution method than for the HSI based method. Especially the critical value of mean plus 2 times SD, a commonly used cut-off value corresponding to the confidence level of 95% probability that an observed value is higher than the control value, results in significantly better classification using the deconvolution method than using the HSI based method.

Conclusions: The deconvolution method performed significantly better than the HSI based system when compared to manual counting as gold standard. Normalization of the data to positive and negative control slides included in each batch improved the performance of the HSI based system and increased the correlation between the HSI based and deconvolution system. The deconvolution system showed significantly reduced variability in the LI determination of highly labeled control samples. This resulted in significantly increase in sensitivity of classification of samples with increased KI-67 labeling without changing the specificity, when compared to the HSI based method.

Acknowledgements
The authors want to thank Sejal Patel and Abha Khana for expert technical assistance. In part supported by NCI contract RFP-N01-CN-85083-57

References

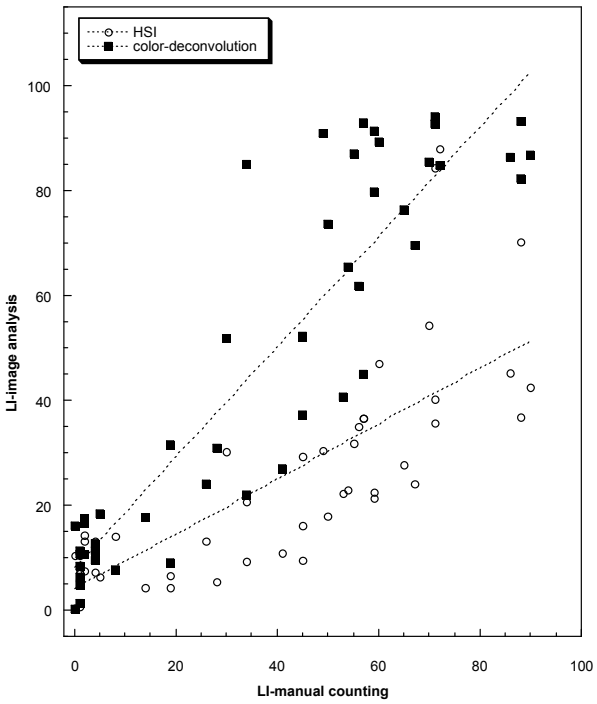
1. Lehr H-A, Mankoff DA, Corwin D, Santeusanio G, Gown AM: Application of Photoshop-based image analysis to quantification of hormone receptor expression in breast cancer. J Histochem Cytochem 1997;11:1559-1565.
2. Lehr H-A, van der Loos CM, Teeling P, Gown AM: Complete chromogen separation and analysis in double immunohistochemical stains using Photoshop-based image analysis. J Histochem Cytochem 1999;47:119-125.
3. Kai M, Nunobiki O, Taniguchi E, Sakamoto Y, Kakudo K: Quantitative and qualitative analysis of stain color using RGB computer color specification. Anal Quant Cytol Histol 1999;21:477-480.
4. Ruifrok AC: Quantification of immunohistochemical staining by color translation and automated thresholding. Anal Quant Cytol Histol 1997;19:107-113.
5. Montorini R,Diamanti L, Thompson D, Bartels HG, Bartels PH: Analysis of the capillary architecture in the precursors of prostate cancer: recent findings and new concepts. Eur Urol 1996;30:191-200.
6. Ruifrok AC, Johnston DA. Quantification if histochemical staining by color deconvolution. Analyt Quant Cytol Histol 2001;23:291-299.
7. Pinder SE, Wenyck P, Sibbering DM, et al. Assessment of the new proliferation marker MIB1 in breast carcinoma using image analysis: associations with other prognostic factors and survival. Br J Cancer 1995;71:146-149.
8. Zar JH: Biostatistical Analysis, 4th edition. Prentice-Hall, Upper Saddle River, New Jersey 1999.

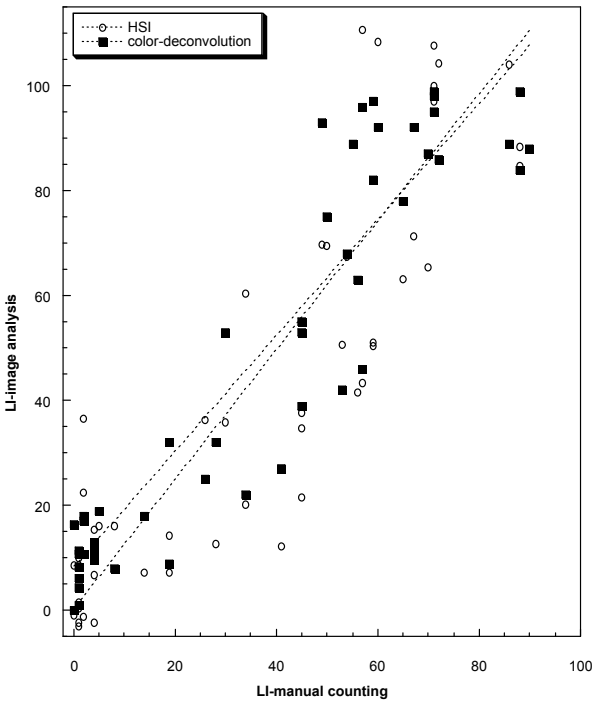
Legends:

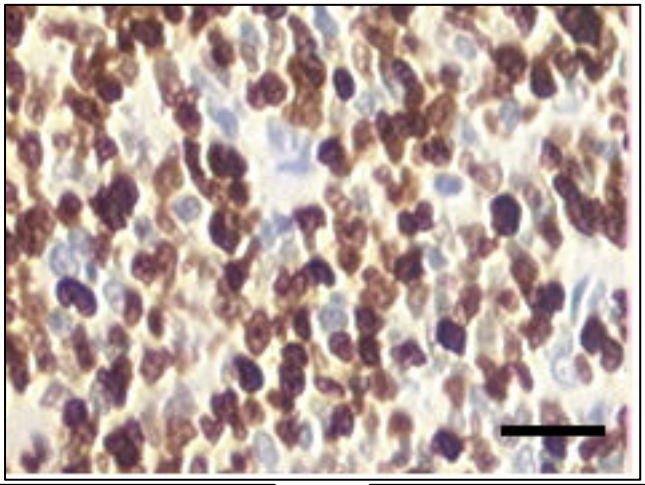
Figure 1: A) Correlation between manual counting of Ki-67 labeling index (LI) and automated LI determination using HSI-based and deconvolution-based methods, before normalization to positive and negative control slides included in the staining batches. HSI method: r=0.77, deconvolution method: r=0.91. B) Correlation between manual counting of Ki-67 labeling index (LI) and automated LI determination using HSI-based and deconvolution-based methods, after normalization to positive and negative control slides included in the staining batches. HSI method: r=0.85, deconvolution method: r=0.90.

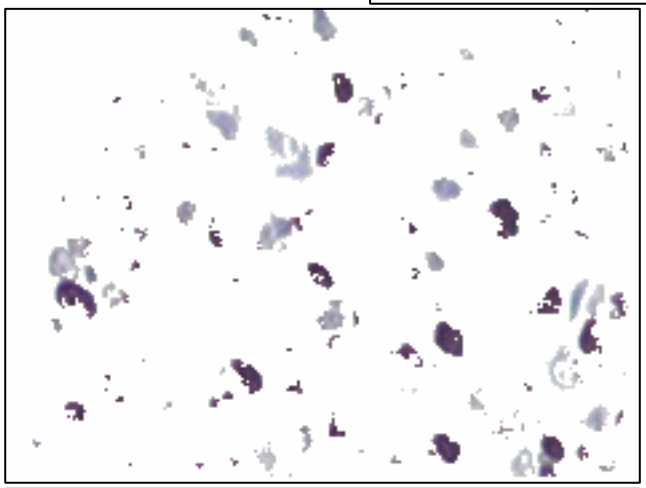
Figure 2: Comparison of image analysis results of a Ki-67 positive control slide with strong diaminobenzidine (DAB) staining (A). Left: blue segmented (B) and brown segmented(C) areas of sample A) using the HSI based segmentation method. Right: blue (D) and brown (E) components of sample A) based on the color-deconvolution method. The bottom panel shows nuclei falsely classified as negative for DAB by the HSI method, but as DAB positive using the color-deconvolution method. Original magnification 20X, Bar (panel A): 20 μm.

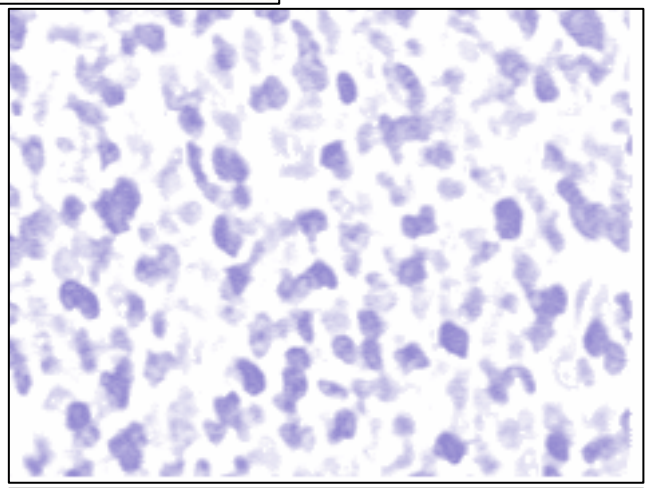
Figure 3: Sensitivity (A) and specificity (B) of classification of lung tumor samples as having an increased Ki-67 labeling index as a function of critical value used for cut-off. The value of mean plus 2 times standard deviation, corresponding to the confidence level of 95% probability that an observed value is higher than the control value, is commonly used as cut-off value. This cut-off value results in a sensitivity of 78% sensitivity and 78% specificity for the HSI based method, and 89% sensitivity and 100% specificity for the color-deconvolution based method.

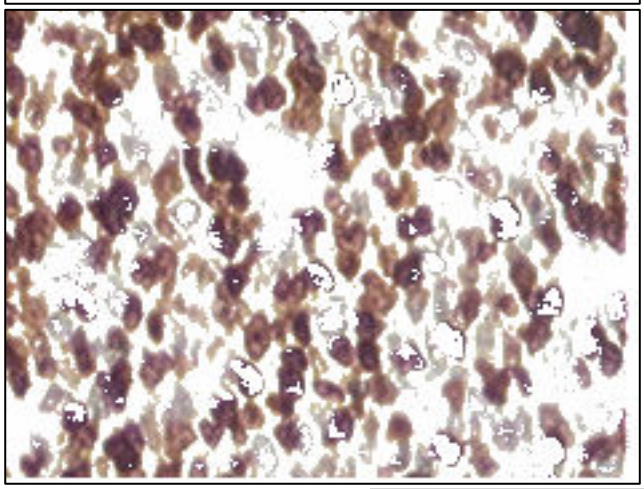

HSI
color-deconvolution
LI-image analysis
LI-manual counting
0
20
40
60
80
100

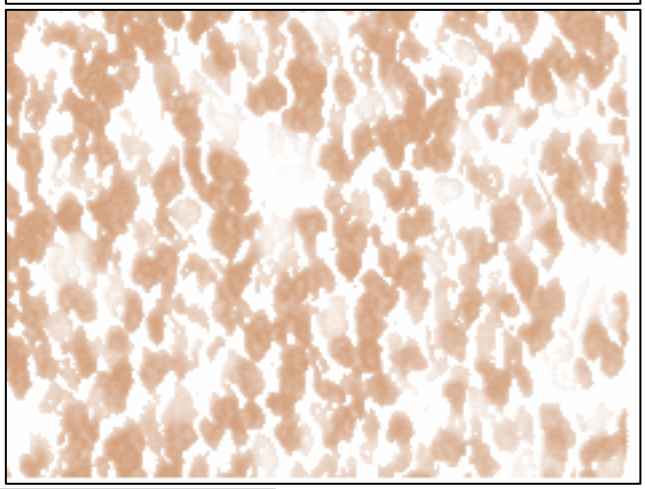

HSI
color-deconvolution
LI-image analysis
LI-manual counting
0
20
40
60
80
100

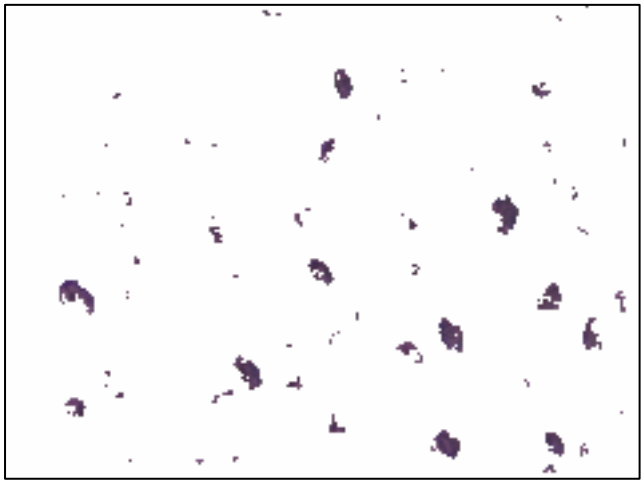


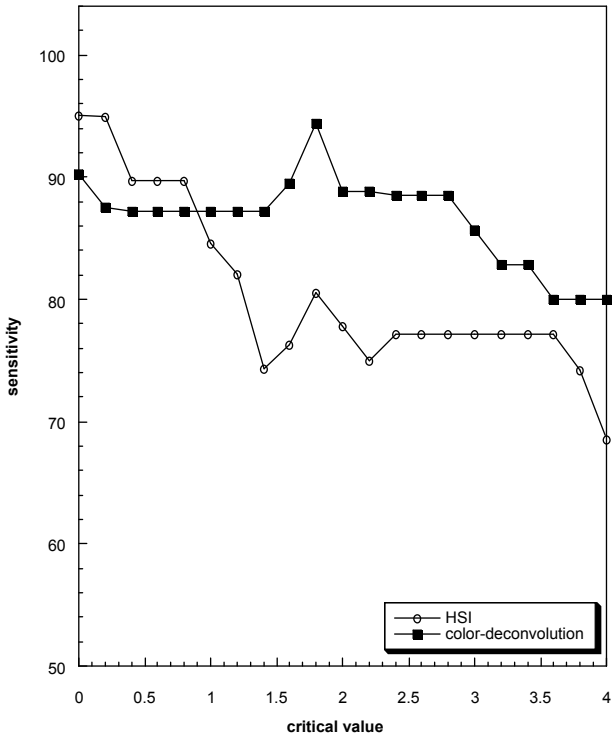











sensitivity
critical value
HSI
color-deconvolution

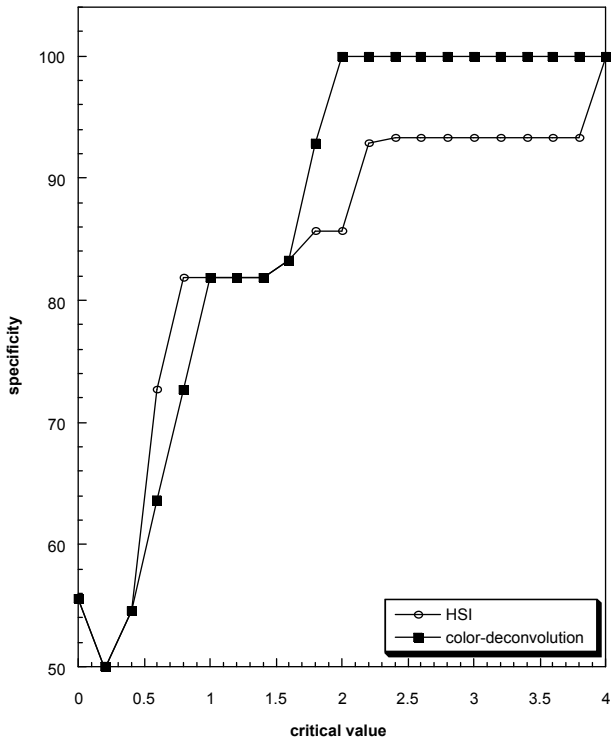

specificity
critical value
HSI
color-deconvolution

Table 1: Mean ± Standard Deviation of the automated measurements of the negative and positive control slides, using the HSI-based method and color-deconvolution (DECON):

	LI	MOD	QuickScore
Neg. contr. HSI	2.11±2.71	5.69±2.45	0.13±0.16
Pos. contr. HSI	49.48±23.97	43.86±17.54	24.86±21.35
Neg. contr. DECON	0.41±0.09	59.27±22.02	0.23±0.07
Pos. contr. DECON	96.95±1.38	96.07±25.13	93.37±25.23

LI: labeling index (area of Ki-67 positive nuclei/total nuclear area), MOD: mean optical density, QuickScore: product of LI and MOD.

Table 2). Correlation (r^2-values) of manual counting ("gold standard") with the HSI-based method and color-deconvolution (DECON) without and with normalization to positive and negative control slides included in each staining batch:

	HSI	HSI-normalized	DECON	DECON-normalized
LI	0.60	0.72	0.83	0.80
MOD	0.62	0.42	0.26	0.01
QuickScore	0.46	0.46	0.70	0.63

LI: labeling index (area of Ki-67 positive nuclei/total nuclear area), MOD: mean optical density, QuickScore: product of LI and MOD.