Modified Automated Scoring system for Immunohistochemical staining using commercially available low cost software for image analysis

Alla Ghani, * (M.B.Ch.B), (F.I.C.M.S. Path),
Mahmood Yaseen** M.B.Ch.B, M.Sc. (Microbiology),
Ibraheem Yaseen, ***. M.B.Ch.B, M.Sc. (Path),

Summary::

J Fac Med Baghdad 2006; Vol. 48, No.3
Received Jan. 2006
Accepted May 2006

Background: *During the past several years, there has been a rapidly escalating clinical need to perform IHC stains that require quantitative interpretation. Automated Cellular Imaging System is used to analyze immunohistochemically stained slides, primarily for cancer-related diagnostics. Studies have shown that the device offers accuracy, precision, and reproducibility of immunostained slide analysis exceeding that possible with manual evaluation, which was the prevailing method.*

Aim of the study *In this article we will demonstrate that meaningful image analysis of immunohistochemical staining studies can be performed using inexpensive, widely distributed graphics software (Adobe Photoshop) on a personal computer. Also we will try to use a modified digital scoring system depending on the percentage of pixels that are showing a given stain with regard to the total area of the slide. We select three sets for each antigen (one is optimally stained, one is insufficiently stained and third one is not stained one)*

materials and methods *Thirty digital pictures of immunohistochemically stained slides with monoclonal antibodies against different antigens, from standard quality control lab (NORDIQC) and then were analyzed by Adobe Photoshop software, then we distribute the percent of pixels showing a giving band of the brown color into five groups, then we compare those results in the three groups.*

Results: *Results showed that there was significant difference between color bands of the same tissue among optimal and insufficient staining ($P<0.05$), also there was significant difference between the group of slides that were optimally stained with those insufficiently stained ($p<0.05$), that's to say the procedure of scoring that was done was accurate in discriminating between optimal staining and insufficient staining*

Conclusion: *Each slide was converted into a matrix of data that describe every pixel in the slide and by that we can compare between all slides that's to say we convert the visual manual evaluation into an automated objective analysis, which is the first step in establishing quantitative immunohistochemistry.*

Key words: Quantitiative immunohistochemistry, Adobe photoshop, Automated Cellular Imaging System

Introduction:

WHEN immunohistochemistry (IHC) was first introduced as an adjunct in surgical pathology diagnosis, the interpretation of tissue staining was largely qualitative. The fact that IHC interpretation was qualitative, rather than semi-quantitative, minimized the impact of many known inconsistencies among laboratories with regard to reagents and methods. During the past several years, there has been a rapidly escalating clinical need to perform IHC stains that require quantitative interpretation. The level of cellular expression for certain analytes, notably HER-2 and estrogen and progesterone receptor proteins, are linked to particular therapies. In this emerging clinical paradigm of individualized medicine accurate quantitative data are important to reach the correct treatment decision. [1]

Automated Cellular Imaging System (ACIS)

This system is used to analyze immunohistochemically (IHC) stained slides, primarily for cancer-related diagnostics. Studies have shown that the device offers accuracy, precision, and reproducibility of immunostained slide analysis exceeding that possible with manual evaluation, which was the prevailing method. ACIS is comprised of an automated microscope; a digital camera; and computerized image-processing technology for detecting, counting, and classifying cells based on color, size, and shape.[2,3] However, the high cost and the complexity of these image analysis systems, requiring major hardware and software investments, severely limit their practicability in the routine diagnostic laboratory. [4]

In this article we demonstrate that meaningful image analysis of immunohistochemical

Assistant Professor, Chairman of Pathology Department, Al Nahrian University, College of Medicine
** Assistant Lecturer, Medical Microbiology Department, Al Nahrian University, College of Medicine.
*** Pathology Department, Al Nahrian University, College of Medicine

staining studies can be performed using inexpensive, widely distributed graphics software (Adobe Photoshop) on a personal computer. We demonstrate that this technique, which requires nothing but a microscope, a camera, and a personal computer, yields data that correlate with the score that was given to the slides by a known quality control lab. Also we will try to use a modified digital scoring system depending on the percentage of pixels that are showing a given stain with regard to the total area of the slide.

materials and methods:

Digital Slides pictures

All photomicrographs were obtained from standard quality control lab ***NORDIQC*** (which is a professional and scientific organization independent of economical or political interests. The NordiQC work is primarily based on routine immunostaining of slides from standard processed human histological specimens with varying expression of antigens. The stains presented at the web-site originates from Nordic laboratories participating in schemes, Institute of Pathology , Aalborg Hospital , P.O.Box 561, DK-9100 Aalborg, Denmark). Files were saved in uncompressed TIFF format so that their sizes ranged between 20 and 25 MB. When a portion of the original image file was selected for further evaluation, the size of the modified file ranged from 1 to 20 MB, depending on the amount of image being evaluated by Qualitative IHC.

Selected Antigens

Thirty slides were taken after they were stain with monoclonal antibodies against ASMA (6 slides), BCL-2 (4 slides), CD34 (6 slides), CD30 (5 slides), and PSA (9 slides), for each antigen we chose two sets of the same section one is of optimal staining (OPT), and the second one is insufficient staining (INS). Two Slides without any staining were taken as No stain group (NIL). Enumeration and criteria for selection is shown in details in Table(7).

Image analysis

Each picture was analyzed by commercially available software called Adobe Photoshop © according to the following scheme which was organized by us: The software used was Photoshop, version 7.0 (Adobe Systems; Mountain View, CA). Three x20 fields were chosen so as to best reflect the overall immunostaining contained on the entire slide. The procedure for determination of immunostaining intensity was done by using the Magic Wand tool in the Select menu of Photoshop, the cursor was placed on an DAB positive cells. The tolerance level of the Magic Wand tool was adjusted so that the entire positive cells were selected. Using the Similar command in the Select menu, all immunostained cells were automatically selected. Once the different chromogens are selected, quantification is accomplished using the Histogram command in the **Image** menu. This display is rarely if ever used by graphic designers but rather serves as an internal measurement of tonal distribution as the basis for automated **image** manipulation (map commands). When Histogram is selected, a display appears on the screen depicting the the luminosity (color) of all pixels within the selected area, including median and standard deviation. Furthermore, this display shows the number of pixels that are covered by the selected area. Because the number of pixels reflects a surface area on the **image**, important spatial information can be obtained for the specific chromogen (and hence the cells expressing a certain antigen) and can be expressed as percentage of the entire **image** or in µm2. [5] The brown colour of positive DAB was classified by the software into 256 band (0-255), where 0 was the darkest Brown color while 255 was the lightest brown, by making a graph where the X axes shows the number of pixles that show a given brown color band while the Y axes was the 256 bands of the brown colour, after that we convert those graphs and data into an excel file where we divide the 256 bands into 5 groups that's is to say 5 groups of the brown color as follow (table.1) and (fig.2):

Table (1) The Brown color bands division into 5 groups

Brown color bands	Grade	Description
0-50	*****	Very Dark Brown
51-100	****	Dark Brown
101-150	***	Brown
151-200	**	Light brown
201-256	*	Very light brown

Examples of image analysis are shown in figure (3 and 4)

Statistical Analysis

We use ANOVA , t-test for measuring the significant of difference between the results among the group of the same antigen and between the two main groups of optimal and insufficient staining using the Microsoft XP Excell software, P value was =0.05.

Results:

The average percentage of pixels that showed very dark brown stain (grade *****) was higher in optimal stained group ($9.83\% \pm 3.34$) than those of insufficiently stained group ($1.41\% \pm 0.69$) and both groups were higher than non stained groups ($0\% \pm 0$), the optimal stain was (6.97X) higher that insufficient stain for this color grade.

While the average percentage of pixels that showed dark brown stain (grade ****) was higher in optimal stained group ($6.82\% \pm 1.86$) than those of insufficiently stained group ($1.81\% \pm 0.81$) and both groups were higher than non stained groups ($0\% \pm 0$), the optimal stain was (3.76X) higher that insufficient stain for this color grade.

The average percentage of pixels that showed brown stain (grade ***) was higher in optimal stained group

(8.65% ± 2.01) than those of insufficiently stained group (3.54% ± 1.79) and both groups were higher than non stained groups (0% ± 0), the optimal stain was (2.44X) higher that insufficient stain for this color grade.

The average percentage of pixels that showed light brown stain (grade **) was higher in optimal stained group (15.40% ± 3.39) than those of insufficiently stained group (10.36% ± 3.59) and both groups were higher than non stained groups (0.40% ± 0.15), the optimal stain was (1.49X) higher that insufficient stain for this color grade.

In contrast to the above, the average percentage of pixels that showed very light brown stain (grade *) was lower in optimal stained group (52.1% ± 7.35) than those of insufficiently stained group (82.88% ± 6) and both groups were higher than non stained groups (99.74% ± 0.15).

Results showed that there was significant difference between color bands of the same tissue among optimal and insufficient staining (P<0.05), also there was significant difference between the group of slides that were optimally stained with those insufficiently stained (p<0.05), that's to say the procedure of scoring that was done was accurate in discriminating between optimal staining and insufficient staining. Results were summarized in tables (2,3,4, and 5) and figure (1)

Table (2): The average percentage of pixels showing a given color band for all groups

Color grade	opt	ins	NIL	Opt/Ins	Significant
*****	1.41	9.83415	0.00	6.97	The difference was significant (P<0.05)
****	1.8126	6.83039	0.00	3.76	
***	3.5438	8.72768	0.00	2.44	
**	10.356	15.7469	0.40	1.49	
*	82.876	58.8609	99.58	0.63	

Table (3): Descriptive statistics for the Optimal staining group according to staining grades

	*****	****	***	**	*
Mean	9.83	6.83	8.72	15.74	58.86
St.Error	3.34	1.86	2.01	3.39	7.35
Median	2.99	4.35	5.71	11.63	45.18
St.Deviation	12.97	7.22	7.78	13.16	28.46
Range	0-45.32	0.13-23.51	0.24-25.59	1.048-44.14	11.88-98.50

Table (4): Descriptive statistics for the Insufficient staining group according to staining grades

	*****	****	***	**	*
Mean	1.41	1.81	3.54	10.35	82.87
St.Error	0.69	0.81	1.79	3.59	6.00
Median	0.00	0.06	0.41	4.16	94.18
St.Deviation	2.69	3.14	6.96	13.90	23.26
Range	0-7.77	0-11.43	0-26.46	0.25-39.08	39.28-99.74

Table (5): Descriptive statistics for the No staining group according to staining grades

	*****	****	***	**	*
Mean	0.00	0.00	0.00	0.40	99.58
St. Error	0.00	0.00	0.00	0.15	0.15
Median	0.00	0.00	0.00	0.40	99.58
St. Deviation	0.00	0.00	0.00	0.21	0.22
Range	0-0.00	0-0.00	0-0.01	0.25-0.55	99.42-99.74

Table (6) : Median percentage of pixels showing a giving color grade in all groups

	*****	****	***	**	*
OPT	2.99	4.35	5.71	11.63	45.18
INS	0.00	0.06	0.41	4.16	94.18
NIL	0.00	0.00	0.00	0.40	99.58

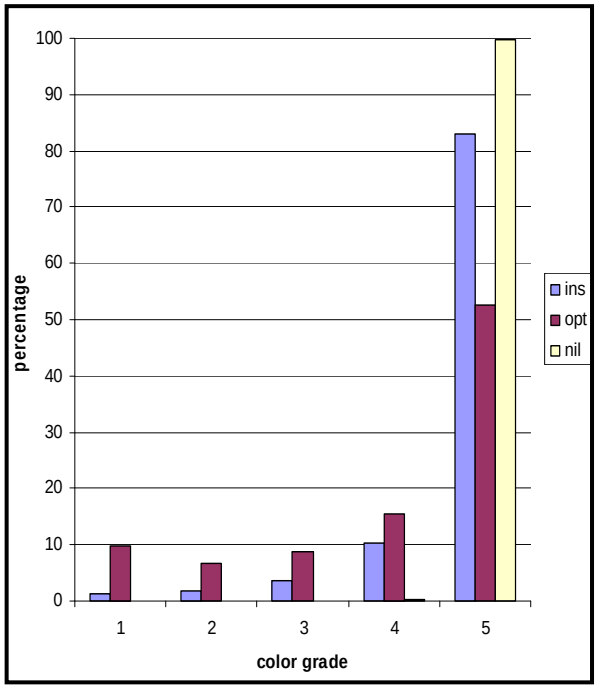

Fig.(1) Showing the percentage of pixels showing a given color grade in all groups

Discussion

Biopharmaceutical companies and research organizations conducting drug discovery using image analysis have found that ACIS gives them the ability to quantitate IHC analysis precisely, something not possible with manual evaluation. Because the instruments are likely to be employed in clinical trials testing, image analysis may emerge as the preferred method to qualify patients for certain therapies when the drugs reach the market. Therefore, medical laboratories using IHC to select patients for therapy, or to monitor that therapy, could use the ACIS image analysis technology to perform such testing accurately.[6,7] Depending on the experience level of the pathologists and the volume of slide reading they undertake, different individuals interpreting the same slides manually could arrive at results that diverge considerably, most notably at the treatment/no-treatment decision point. Current manual IHC reporting analysis, is only semiquantitative; it relies on subjective scoring, using a scale of 0 to 3+, and produces less than

standardized results. As mentioned, studies indicate that the level of accuracy and the degree of reproducibility from pathologist to pathologist increase significantly with the assistance of ACIS.[8]

Our method which was based on the same principle of ACIS but the difference that we use a commercially available image analysis software (Adobe Photoshop) that can be acquired by any researcher cause it is cheap and available to everyone. We tried to simulate the program that was used in ACIS which scores by counting individual pixels of chromogen color and converting the count to one of 256 distinguishable levels of color intensity. Measurement of staining intensity is thus objective. With the aid of Microsoft Excel program that we prepare to measure the average intensity of the color in each group of pixels and convert them to graphs that is easily distinguishable.

Concluding the modified ACIS-assistance analysis, the pathologist generates a written report that includes graphical and numerical data on each marker analyzed. Results are either reported as the exact percentage of positively stained cells. Universal adoption of these standardized scoring and reporting methods would enable detailed comparisons for clinical research and outcomes studies that are not possible with manual assessment techniques. [8]

From the results we can see that the optimal stained groups were having higher percentage of their cells showing the higher grades of the Brown color of the DAB stain, these results were higher than insufficiently stained group and both of them are higher than the no stain group. We can see from the results that in optimal group the percentages for pixels that show high grade stain (*****, ****, and ***) were higher in optimal group than others and the percentage of pixels showing the weakest stain (grade *) was lower in optimal group than insufficient stained group and both are lower than that of non stained group. That's to say whenever the stain is optimal, we will have lower percentage of weak positive. For each color grade the optimal stained group was higher than other group except the grade *, but the highest ratio of optimal to insufficient stain was highest in grade ***** where it was about 7 times higher, the lowest ratio was in grade ** and * (1.49 and 0.6 respectively), so we can use that ratio as discriminating variable between optimal and insufficient staining, that's to say whenever the percentage of pixels that showed grade ***** was 10% or higher, then the stain is optimal.

Other thing to notice that the grade * which was the weakest brown color which can be considered as negative stain, its percentage was highest in no stain group, followed by insufficient stained group while the lowest percentage was in optimal group, the percentage in both no stain and insufficient stain were (99.6 and 82.8 respectively), so we can speculate that whenever grade * percentage was between (80-100%) we will have negative staining or insufficient stain.

When we take the median for percentage of pixels showing a giving color grade, the results were more accurate cause median is the number in the middle of a set of numbers; that is, half the numbers have values that are greater than the median, and half have values that are less, by this the odd value which are too how or too low in set of numbers will not affect the results. From that median percentage of pixels showing (*****) grade was (2.99%, 0%,and 0%) in optimal staining group, insufficient staining group and no staining group respectively, while median percentage of pixels showing (****) grade was (4.35%, 0.06%,and 0%) in optimal staining group, insufficient staining group and no staining group respectively. The median percentage of pixels showing (***) grade was (5.71%, 0.41%, and 0%) in optimal staining group, insufficient staining group and no staining group respectively. The median percentage of pixels showing (**) grade was (11.63%, 4.16%, and 0.40%) in optimal staining group, insufficient staining group and no staining group respectively. Finally median percentage of pixels showing (*) grade was (45.18%, 94.18%, and 99.58%) in optimal staining group, insufficient staining group and no staining group respectively. From that we can say that the staining to be optimal , it must be with following criteria (total percentage of strong staining (grade ***** and ****) must be equal or more than 7.34% of the total stained area.

Each slide was converted into a matrix of data that describe every pixel in the slide and by that we can compare between all slides that's to say we convert the visual manual evaluation into an automated objective analysis, which is the first step in establishing quantitative immunohistochemistry.

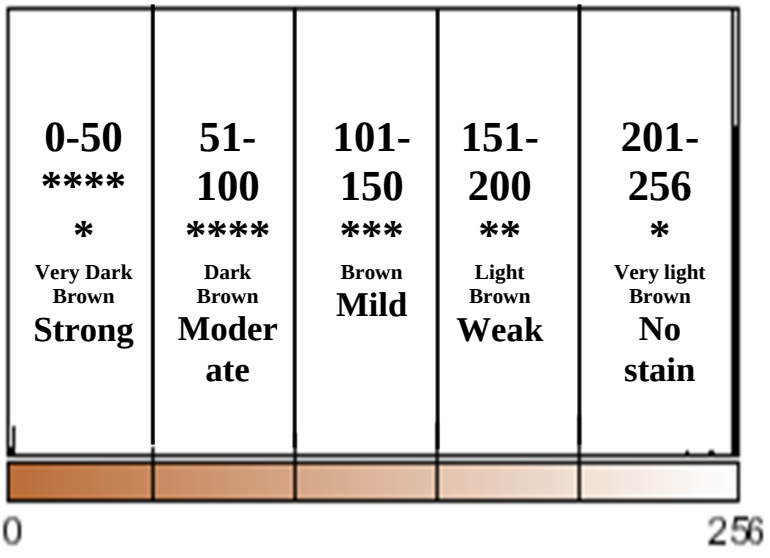

Figure (2): The distribution of brown color bands into 5 grades

Table (7): Details of the slides were chosen, the monoclonal antibody used, the type of tissue which was examined, number of optimal and insufficient stained slides, criteria for choosing optimal or insufficient stains)

No.	Antigen	Tissue	OPT	INS	Description
1.	ASMA	normal appendix	1	1	Criteria for assessing an ASMA staining as optimal included: A strong and distinct cytoplasmic reaction of the appendiceal smooth muscle cells (vessels and muscular layers) and myofibroblasts, the myoepithelial cells of the glands and ducts of the breast fibrocytic disease and the uterine leiomyosarcoma, and a focal cytoplasmic reaction of the two GIST The most frequent causes of insufficient stainings were -Inappropriate choice of primary Ab -Inappropriate epitope retrieval - Insufficient HIER, i.e. too short heating time - Too low concentration of the primary antibody.
		leiomyosarcoma	1	1	
		small intestinal GIST	1	1	
		ASMA Appendix	**OPT**	**INS**	
		*****	7.777186	2.409951	
		****	4.638461	2.797357	
		***	5.713291	3.012438	
		**	39.08905	6.758228	
		*	42.78202	85.02203	
		ASMA Leiomyosarcoma	**OPT**	**INS**	
		*****	2.998186	0.028505	
		****	13.57606	1.391552	
		***	19.70977	9.407878	
		**	23.67194	25.84478	
		*	40.04405	63.32729	
		ASMA GIST	**OPT**	**INS**	
		*****	0.988598	0	
		****	1.474475	0	
		***	2.86214	0.002591	
		**	6.912413	0.392589	
		*	87.76237	99.60482	
2.	Bcl-2	normal tonsil	1	1	Criteria for assessing a Bcl-2 staining as optimal included: In both tonsils a moderate to strong distinct cytoplasmic staining of the peripheral B-lymphocytes in the mantle-zone and T-lymphocytes in the interfollicular areas and germinal centres (whereas the germinal B-lymphocytes should be negative) and a strong staining of all the follicular lymphomas. The most frequent cause of insufficient staining was: too diluted primary antibody.
		Hodgkin's lymphoma	1	1	
		Bcl-2 Tonsils	**OPT**	**INS**	
		*****	18.912	0	
		****	12.764	0	
		***	11.500	0	
		**	11.639	0.348536	
		*	45.182	99.65146	
		Bcl-2 lymphoma	**OPT**	**INS**	
		*****	17.911	0	
		****	23.517	0	
		***	25.593	0.018139	
		**	21.090	6.208862	
		*	11.886	93.773	
3.	CD34	Normal liver	1	1	Criteria for assessing a CD34 staining as optimal were: a strong and distinct cytoplasmic reaction with membrane accentuation in the endothelial cells of small vessels in the appendix and the liver (portal tracts and zone 1 sinusoids), tumour cells of the GISTs, and a high proportion of tumour cells in the myeloid sarcoma, without any staining of epithelial and smooth muscle cells. The most frequent causes of insufficient stainings (often in combination) were: -A too low primary Ab concentration -An insufficient HIER (too low pH and/or too short heating time) - A less sensitive visualization system
		GIST	1	1	
		Breast	1	1	
		CD34-liver	**OPT**	**INS**	
		*****	1.261822	7.777186	
		****	1.608997	4.638461	
		***	3.146482	5.713291	
		**	7.762399	39.08905	
		*	86.2203	42.78202	
		CD34-breast	**OPT**	**INS**	
		*****	27.6782	0.016148	
		****	17.62399	0.126874	
		***	14.28258	0.410611	
		**	12.63668	1.630911	
		*	27.77855	97.81546	
		Cd34-gist	**OPT**	**INS**	

		*****	9.490196	0.004614	
		****	11.40138	0.06113	
		***	18.19377	0.484429	
		**	27.74625	4.161476	
		*	33.1684	95.28835	
4.	CD30	Hodgkin lymphoma	1	1	Criteria for assessing a CD30 staining as optimal included: A distinct membranous staining of activated B- and T-cells in the tonsil, the anaplastic large cell lymphoma, the embryonal carcinoma and the choriocarcinoma as well as a strong and a distinct membranous and dot-like (Golgi) staining of the Reed-Sternberg and Hodgkin's cells in the Hodgkin's lymphoma. The most frequent causes of insufficient stainings were: - Inappropriate choice of epitope retrieval (i.e., proteolytic pre-treatment) - Too low concentration of the primary antibody
		Embryonal carcinoma	1	1	
		Cd30-Hodgk	**OPT**	**INS**	
		*****	0.242291	0.001296	
		****	0.303187	0.010365	
		***	0.66468	0.062192	
		**	2.41902	0.542887	
		*	96.37082	99.38326	
		cd30-embercancer	**OPT**	**INS**	
		*****	0.513086	0.001296	
		****	1.99404	0.001296	
		***	11.83338	0.012957	
		**	44.14226	0.558435	
		*	41.51723	99.42602	
5.	PSA	Prostatic hyperplasia	1	1	Criteria for assessing a PSA staining as optimal included: A moderate to strong distinct cytoplasmic staining of the hyperplastic prostate glands and the three prostate adenocarcinomas. A weak to moderate reaction of the prostate stroma was accepted. No reaction should be seen in the kidney. The most frequent causes of insufficient staining were (often in combination): -Too low concentration of the primary Ab -Inappropriate choice of primary Ab -Inappropriate epitope retrieval (proteolytic pre-treatment) or no retrieval
		Kidney	1	1	
		Adenocarcinoma (8)	1	1	
		Adenocarcinoma (6)	1	1	
		PSA kidney	**OPT**	**INS**	
		*****	0	7.777186	
		****	0.145115	4.638461	
		***	1.058564	5.713291	
		**	5.178803	39.08905	
		*	93.61752	42.78202	
		psa-gleason8	**OPT**	**INS**	
		*****	1.237367	0	
		****	2.538222	0	
		***	6.484841	0	
		**	18.22104	0.255247	
		*	71.51853	99.74475	
		PSA hyper	**OPT**	**INS**	
		*****	11.77637	1.383778	
		****	4.357346	2.076963	
		***	5.046644	1.78155	
		**	10.51049	3.354496	
		*	68.30915	91.40321	
		psa-gleason6	**OPT**	**INS**	
		*****	45.3291	1.767297	
		****	5.516973	11.4382	
		***	4.419539	26.46022	
		**	6.147966	21.04561	
		*	38.58642	39.28868	
6.	PSA	Kidney	1		No Stain was observed in the slides
7.	CD30	Tonsil	1		No Stain was observed in the slides
Total					Total : 30 slides

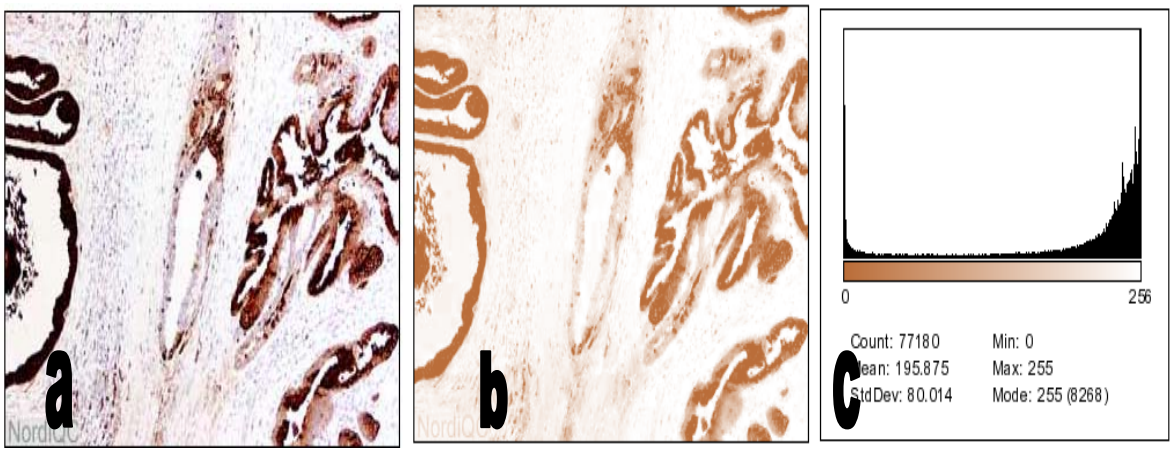

Fig (3) (a) Optimal PSA staining of the prostate hyperplasia. A moderate to strong staining is seen in almost all epithelial cells. A weak stromal reaction is unavoidable. (b) Extraction of the DAB stain only from the whole section with removal of all other colours (c) the histogram of the DAB color according the level of the brown color on X axis and number of pixels that show that level of color intensity on Y axis

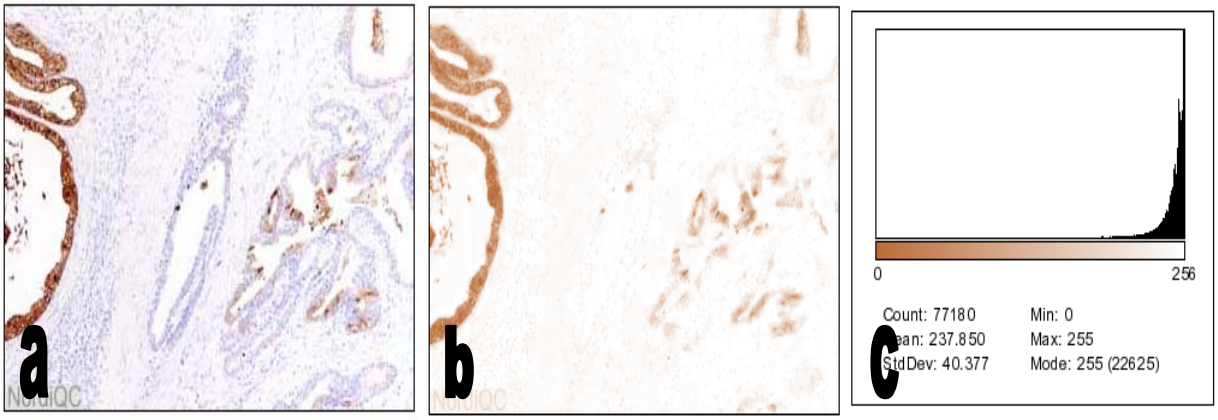

Fig (4) (a) Insufficient PSA staining of the prostate hyperplasia. A large proportion of epithelial cells are unstained. (b) Extraction of the DAB stain only from the whole section with removal of all other colours (c) the histogram of the DAB color according the level of the brown color on X axis and number of pixels that show that level of color intensity on Y axis

References

1. Seshi R. Sompuram, Vani Kodela, Keming Zhang, Halasya Ramanathan, Gail Radcliffe, Peter Falb, and Steven A. Bogen, A Novel Quality Control Slide for Quantitative Immunohistochemistry, Testing". Journal of Histochemistry and Cytochemistry, Vol. 50, 1425-1434, November 2002.

2. S Wang et al., "Automated Cellular Imaging System (ACIS)–Assisted Quantitation of Immunohistochemical Assay Achieves High Accuracy in Comparison with Fluorescence In Situ Hybridization Assay as the Standard," American Journal of Clinical Pathology 116 (2001): 495–503

3. K Bloom et al., "Comparison of HER-2/neu Analysis Using FISH and IHC When Hercep Test Is Scored Using Conventional Micros-copy and Image Analysis," Breast Cancer Research and Treatment 64 (2000): 99.

4. Hans-Anton Lehr, David A. Mankoff, David Corwin, Guiseppe Santeusanio, and Allen M. Gown, Application of Photoshop-based Image Analysis to Quantification of Hormone Receptor Expression in Breast , Journal of Histochemistry and Cytochemistry, Vol. 45, 1559-1566,

5. *Hans-Anton Lehr, Chris M. van der Loos, Peter Teeling, and Allen M. Gown,* ***Complete Chromogen Separation and*** *Analysis* ***in Double Immunohistochemical Stains Using Photoshop-based*** *Image Analysis* ***Journal of Histochemistry and Cytochemistry, Vol. 47, 119-126, January 1999***

6. C Taylor, "The Total Test Approach to Standardization of Immunohistochemistry," Archives of Pathology and Laboratory Medicine 124 (2000): 945–951.

7. T Seidal, AJ Balaton, and H Battifora, "Interpretation and Quantification of Immunostains," American Journal of Surgical Pathology 25, no. 9 (2001): 1204–1207.

8. Apfelbach W. and Jose de la Torre-Bueno, Instrumentation Development, Making image-analysis technology accessible to community pathologists By developing a virtual network for distributing the benefits of its technology among users, an instrument manufacturer expands its market. Carl IVD Technology January/February 2003