Automatic quantification of immunohistochemically stained cell nuclei based on standard reference cells

Petter Ranefall a,*, Kenneth Wester b, Ann-Catrin Andersson b, Christer Busch b and Ewert Bengtsson a

a Centre for Image Analysis, Uppsala, Sweden
b Department of Pathology, Uppsala University, Sweden

Received 8 January 1998
Revised 15 October 1998
Accepted 15 October 1998

Abstract. A fully automatic method for quantification of images of immunohistochemically stained cell nuclei by computing area proportions, is presented. Agarose embedded cultured fibroblasts were fixed, paraffin embedded and sectioned at 4 μm. They were then stained together with 4 μm sections of the test specimen obtained from bladder cancer material.

A colour based classifier is automatically computed from the control cells. The method was tested on formalin fixed paraffin embedded tissue section material, stained with monoclonal antibodies against the Ki67 antigen and cyclin A protein. Ki67 staining results in a detailed nuclear texture with pronounced nucleoli and cyclin A staining is obtained in a more homogeneously distributed pattern.

However, different staining patterns did not seem to influence labelling index quantification, and the sensitivity to variations in light conditions and choice of areas within the control population was low. Thus, the technique represents a robust and reproducible quantification method.

In tests measuring proportions of stained area an average standard deviation of about 1.5% for the same field was achieved when classified with classifiers created from different control samples.

Keywords: Image analysis, automatic quantification, immunohistochemically stained cell nuclei, control cells

1. Introduction

This is the third part of a series of studies intending to make quantification of immunohistochemistry (IHC) more objective, automatic and reproducible [23,24]. The incentive for these studies is that quantification of the proportions of specifically stained objects or regions in images is of significant interest in a growing number of biomedical applications. These applications includes histology and cytology where subjective semiquantification of various stainings performed on histological tissue sections, smears, imprints, etc. needs to be substituted by more objective techniques. By using specific stains, biological components of interest can be given a specific colour.

Crude and subjective quantifications with, sometimes, poor inter- and intraobserver reproducibility together with lack of proper controls and reference material has always been a limiting factor in IHC [3,
5,15,25,26,32]. Today, the access to powerful computerised image analysis methods [10,13,18,27] offers more objective quantification compared to subjective classifications into fictive categories regarding the extent, and approximations of the staining intensity of the IHC [1,4,8,19,20].

Most often the subjective quantifications are performed by a pathologist. When using computerised image analysis the evaluation can, in certain applications, be performed by technicians or engineers. Regarding reproducibility there are still standardisation problems to be solved, such as histoprocessing and staining, and reference point selection. In other words interactivity implies a subjective component also when using an objective quantification technique.

In our previous work [23,24], we dealt with a supervised method for classification as well as a fully automatic one. In this study we present a hybrid between the two techniques. It is based on a classifier, created unsupervised from an external cultured control sample.

The specimens are accompanied by control cells, sectioned and stained simultaneously. The image of the control cells serves as a standardised control regarding image qualities, such as illumination and colour properties. Since these cells have known characteristics regarding expression of protein to be examined, they serve as a means to control and standardise results.

The study shows that the proposed method is very stable and in that sense superior to the fully automatic method without controls, which in turn is associated with less variation than the intraoperator variation using supervised classification [24].

2. Materials and methods

2.1. Cell culture and media

Normal foreskin fibroblasts (CCD-1064Sk. ATCC CRL-2076) were split into passage 7 at a ratio 1:6 and grown as monolayers in 10 cm Petri-dishes. Cells were cultured in Eagles minimum essential medium (MEM) with Earl’s salt for three days and harvesting was performed by trypsinisation. The final concentration was $1.04 \times 10^6$ cells per Petri-dish, resulting in a total number of $33 \times 10^6$ cells. An aliquot was withdrawn for staining in propidium iodide and subsequent cell cycle analysis, using a FACSorter (Becton Dickinson Immunocytometry systems, CA, USA). The distribution in G0/G1, S-phase, and G2/M were of 68, 20, and 12%, respectively.

In a separate Petri-dish, cells were incubated with bromodeoxy uridine (BrdU, Amersham), according to instructions from the manufacturer, for two hours before harvesting. Cells were washed once in PBS and cytospin slides were prepared at a concentration of $7.0 \times 10^4$ cells per slide. After fixation in 2.5% paraformaldehyde (see below) the slides were stained by IHC, using a mouse monoclonal anti-BrdU antibody (DAKO, Glostrup, Denmark) diluted 1/40. The BrdU labelling index were calculated to 16%, by counting the number of BrdU labelled cells from a total number of 1000 cells.

2.2. Preparation of control cells, cell-agarose gel

The cells were washed twice in PBS, pH 7.4, at 4°C. A mild fixation in 2.5% paraformaldehyde in 0.1 M phosphate buffer, pH 7.2 for 15 min was followed by a further wash in PBS containing glycin. The cells were resuspended in 0.5 ml PBS and placed in a water bath at 40°C together with an equal volume of 6% Sea plaque® agarose (FMC BioProducts, Rockland, ME, USA) in 0.9% NaCl. When the solutions had reached a temperature of 40°C, the cells were added to the agarose solution and gently suspended with a pipette until a homogeneous cell-agarose mixture was achieved. The mixture was then
Fig. 1. IHC stained control cells.

Sections, 4 μm thick, from paraffin blocks of bladder carcinoma and cell-agarose gel were placed in pairs onto Super frost/plus® slides (Mentzel, Germany), baked at 40°C for 1 h and stored at 4°C until IHC was performed. Two mouse monoclonal antibodies were used in the study; MIB1 (Immunotech, Marseille, France) and anti cyclin A protein (Novocastra, Newcastle upon Tyne, UK). Six concentrations of each antibody were used, aiming to create a staining intensity spectrum, incubated for 16 h at 4°C. Prior to IHC heat mediated antigen retrieval (HIER) was obtained by boiling the slides in 0.01 M citrate buffer, pH 6.0, for 16 min at 750 W in a microwave oven (Whirlpool VIP34, Sweden). Blocking for endogenous peroxidase in 0.3% H2O2 and preincubation in 10% normal rabbit serum, both diluted in PBS and incubated for 20 min. As link antibody a biotinylated rabbit anti mouse (DAKO, Glostrup, Denmark) was applied, followed by a peroxidase labelled streptavidin biotin complex (DAKO), both diluted 1/200 and incubated for 30 min. The slides were developed in 0.03% DAB (Sigma, St. Louis, MO, USA) for 6 min and counterstained in Harris haematoxylin (Sigma). Finally, the slides were dehydrated through graded alcohols to xylene and mounted in organic mounting medium. Reagents were diluted in 0.5% BSA-C (Aurion, Wageningen, Netherlands) and incubations were performed at room temperature, unless otherwise stated. Washings, for 3 × 10 min, between incubation steps were done in 0.05 M Tris, pH 7.6, containing 0.3 M NaCl and 0.1% Tween 20®. Example of the IHC staining of a bladder carcinoma is shown in Fig. 2.
2.4. Image acquisition

The 756 × 572 pixels colour images with 3 × 256 grey levels were grabbed by a Sony DXC-151 colour video camera attached to a standard Olympus BH-10 optical microscope, using a 40× lens. This results in a pixel size of about 0.4 µm. The Rayleigh resolution criterion [9] gives a resolution limit of 0.24 µm for a wavelength of 550 nm and a numerical aperture of 0.7. We are thus not fully resolving the images but our application is not concerned with details of the nuclei texture. A larger field of view was considered more important than maximum resolution.

For all images Köhler illumination was maintained and the aperture iris diaphragm ring was fixed to 0.5.

It is important that the illumination be the same for the control sample and the fields of view to be classified.

2.5. Method

2.5.1. Training

First a principal component transform [14] of the RGB values of the control image is performed. The sign of the first principal component, PC1, is defined so that the sum of the components is positive. In this way high PC1 value corresponds to “light”.

Since the cultured cells do not form any histological structures, i.e., stroma, epithelia, muscle, etc., it is simple to segment the nuclei from the homogeneous background by thresholding of PC1. We used the approach of finding the threshold value \( d \), that maximises the split effect (SE) on the data set \( X \) as defined by Milvang [22] according to the formula:

\[
\text{SE}(X) = \max_d \left[ \sum_{x \in X} (x - \overline{X})^2 - \left( \sum_{x \in X_m} (x - \overline{X}_m)^2 + \sum_{x \in X_n} (x - \overline{X}_n)^2 \right) \right],
\]

(1)
where $X_m = \{ x \in X : x \leq d \}$, $X_n = \{ x \in X : x > d \}$, and $\overline{X}$ is the mean value of all $x \in X$. $X_m$ and $X_n$ are defined analogously.

Expressed in words it could be said that the split effect for a set is the maximum difference between the sum of all squared distances to the centre of gravity within the set and the sum of all squared distances to the centre of gravity within the two new sets after the split.

This step is not very crucial and other threshold criteria could be equally efficient. See result of this step in Fig. 3.

The next step is to define where to separate the different staining classes. To do this the colours from the segmented nuclei are transformed to Principal Component Hue (PCH), as defined in our previous work [24], with the PC1 as above. PCH is defined as the angle between the colour of interest and a reference vector, both projected to a plane orthogonal to PC1. The reference vector is defined by studying the PCH-histogram (see Fig. 4) and finding the direction with the least data content. After this a threshold level is chosen by using the optimal split effect as above (1). See result in Fig. 5.

The parameters for PC1, PC2, PC3, the reference angle and the threshold value are stored in a file. This file is used in the classification process.

2.5.2. Classification

To be able to use the stored classifier, the nuclei have to be segmented from the background in the specimen. We have done this by using $P^2/A$ histogram based global thresholding of the PC1 image, as defined in our previous work on automatic segmentation [24].

See example of segmented nuclei in Fig. 6.

When the nuclei are segmented from the background, the stored transformations and threshold values are used to transform the image into PCH and perform the thresholding.

The resulting PCH histogram is shown in Fig. 7 and the resulting segmentation is shown in Fig. 8.
Fig. 4. PCH histogram for control nuclei. This circular histogram is treated as a linear histogram by cutting the circle open at the ‘cut’ position, i.e., adding an offset so that the position of the cut gets the value 0.

Fig. 5. Result of training of control nuclei. The nuclei are segmented into two classes (here marked by grey and black), and the parameters needed to reproduce the transformation and segmentation are stored in a classifier file.
Fig. 6. Segmented nuclei in specimen. After this segmentation the stored classifier can be applied.

Fig. 7. PCH histogram for nuclei in specimen. The ‘cut’ and ‘threshold’ levels, respectively, are defined by the stored classifier.
3. Results

The method for automatic quantification of immunohistochemical nuclear staining was tested on an independent test set as follows. Samples from 7 patients were stained with either MIB1 or antibodies against the protein cyclin A. For each sample and staining one field was selected. Each of these 14 fields of view was grabbed with 7 different light settings. For each light setting 7 images with control cells were
Fig. 10. Each line shows the mean and the dashed lines shows the 95% confidence interval for each field of view for each light setting, i.e., for each group of 7 controls applied to the same field of view in Fig. 9. This shows how the result varies due to changes in light settings.

![Graphs showing data for Cyclin A and MIB1](image1)

Fig. 11. Result when using automatic method without control cells (each line represents a field of view). In the ideal case there should only be straight lines.

![Graphs showing data for Cyclin A and MIB1](image2)

also grabbed for each staining. Each control image was used to create a classifier as described above and the classifier was applied to all test images grabbed with the same light settings and from the same staining. This test scheme yielded in total 98 control images, 98 test images and 7 measurements for each test image, which gave a total of 686 values. These values are shown, grouped by field of view, in Fig. 9. Figure 10 shows the mean values for each field of view with the same light settings. For comparison our previously described automatic quantification method [24] was used on the same images. The results are shown in Fig. 11 and standard deviations for the two methods are shown in Fig. 12. The methods gave an average standard deviation of 1.5 and 3.3%, respectively. Thus, the method with control cells gives slightly better stability.
Fig. 12. The line shows standard deviations for each field of view using the method with controls and the dashed line shows standard deviations for each field of view using the automatic method without controls. The method with controls shows low variation (the average standard deviation was 1.5%) for all images, whereas the method without controls shows somewhat higher variation (the average standard deviation in this case was 3.3%), but the variations are still to be regarded as low for most images.

Fig. 13. This figure shows manual quantification plotted against the results for the proposed method. Each ‘*’ corresponds to one field of view for MIB1 and each ‘+’ corresponds to one field of view for Cyclin A. Since the manual quantification measured number of objects and the automatic method measured areas, we cannot expect to get exactly the same values. What is important is if the correlation is good. This test gave the correlation coefficient $r = 0.968$ for the Cyclin A images, and $r = 0.782$ for the MIB1 images. The overall correlation coefficient was $r = 0.626$.

To test the accuracy of the method, we compared it to manual quantification where all the negative and positive nuclei, respectively, were manually marked in each field of view. The manual counting was compared with the mean values for each field of view in Fig. 13. In this figure the manual results are plotted against the results for the proposed method and each ‘*’ corresponds to one field of view for MIB1 and each ‘+’ corresponds to one field of view for Cyclin A. Note that the automatic
method gave area proportions, whereas the manual counting gives us relations between number of objects. Therefore the results for the manual counting gives a lower fraction of positivity, but what is important here is that the two methods correspond well. This test gave the correlation coefficient $r = 0.968$ for the Cyclin A images, and $r = 0.782$ for the MIB1 images. The overall correlation coefficient was $r = 0.626$.

These tests gives us only indirect measures of the specificity and sensitivity of the proposed method. To be able to compute the percentage of false positive and false negative nuclei we need to identify each nuclei and classify it as either positive or negative. This could be done by using the watershed segmentation method presented in our previous article [23]. However, we have decided not to use this method in this study, since the watershed segmentation might make the objects too fragmented, and thus introduce another source of error when computing the specificity and sensitivity.

In Fig. 13 we can see a tendency that the method gives higher values for Cyclin A than for MIB1. It could be interesting to examine further if it is possible to compensate for this and thus get even better correlation.

4. Discussion

It could be argued that cells grown in culture, differ in many ways from cells in tissue and therefore are inappropriate for use as controls or references. This study shows that the use of embedded cultured control cells, sectioned and stained simultaneously provides a basis for standardisation of quantification of IHC by computerised image analysis.

The study shows that the present method is more stable than the fully automatic method without controls, which in previous work [24] showed less variation than the intraoperator variation using supervised classification. Without access to control cells, the automatic method could be used since it is objective and (at least) as stable as the supervised method for most kinds of cell images. But test situations including several classes of objects implying a need for subjective selection of the classes demands a supervised method.

In the tests described we tried to use exactly the same field of view under different lighting conditions. This was done manually using an earlier image as comparison. This was very thoroughly done, but to evaluate the accuracy we compared two images grabbed with the same lighting conditions, but deliberately displaced about one pixel in the $y$-direction. This, though statistically imperfect showed only minor differences, indicating that this type of error is of minor importance.

All values are measured per field of view, but it could be interesting to relate the measures to some kind of biological entity instead, e.g., clusters of nuclei. The selection of the fields of view, as well as the size and shape of the fields of view, is important when quantifying immunohistochemistry. But this issue is far beyond the scope of this article.

We have noticed that a problem might occur if the image is too light. Then the shape of the data could be distorted towards white. But, on the other hand, if the data is too dark then the colour range is not used properly. These problems can be detected through analysis of the distribution in colour space. In this study we have not used light settings producing images containing any completely white (255, 255, 255) pixels.

Control cells used as described here can correct for variations in image processing conditions such as light intensity and aperture settings. Hitherto, most results based on quantitative IHC evaluations have often encountered some scepticism regarding its validity due to poor inter and intra reproducibility, lack of
standardised IHC-methodology and histoprocessing regimes. Although, these factors have been focused on separately [2,7,11,12,17,21,29,33,34] an approach towards a concept taking all factors into consideration is still missing. A reliable external control cell system, correcting for these variations, would pave the way for extended use of IHC in future research applications. Increased knowledge on RNA and DNA levels, gained by the technical and methodological progress, demands reliable methods for complementary studies on the expression of proteins in their native cytological and histological environment in the tissue. Furthermore, the need for better inter and intra laboratory standardisation in the clinical application is of major importance [6,16,28–31]. Disagreement regarding the diagnostic and prognostic value of quantitatively analysing the expression of a number of proteins, involved in tumour biology, would probably come to a consensus if the IHC methodology was fully standardised and the results comparable. To fulfil these demands there is an absolute necessity to include some control and/or reference in IHC, at least when a quantitative evaluation is to be performed.

Acknowledgement

This project was supported by the Swedish National Board for Technological Development under grant no 5523-93-4991.

References


Submit your manuscripts at http://www.hindawi.com