

# Image Analysis of Breast Cancer Immunohistochemistry-stained Sections Using ImageJ: An RGB-based Model

T. VREKOUSSIS<sup>1</sup>, V. CHANIOTIS<sup>2</sup>, I. NAVROZOGLOU<sup>1</sup>, V. DOUSIAS<sup>1</sup>,  
K. PAVLAKIS<sup>3</sup>, E.N. STATHOPOULOS<sup>2\*</sup> and O. ZORAS<sup>4\*</sup>

<sup>1</sup>Department of Obstetrics and Gynecology, Medical School, University of Ioannina, Ioannina;  
Departments of <sup>2</sup>Pathology and <sup>4</sup>General Surgery, Medical School, University of Crete, Heraklion, Crete;  
<sup>3</sup>Department of Pathology, Medical School, University of Athens, Athens, Greece

**Abstract.** *Background: Image analysis of tissue sections using RGB image profiling is a modern accepted technique. Materials and Methods: A new method of RGB analysis, using the freeware ImageJ, is presented which can be applied to sections with either nuclear or cytoplasmic staining. The step-by-step process is presented and the method is tested using breast cancer specimens immunostained for CK-19 and estrogen receptors. Results: This image analysis easily discriminates CK-19 and estrogen receptor positivity in prepared breast cancer specimens. The method is easy to perform, without the need for previous image transformations. Conclusion: Compared to previous methods, this method proved more accurate in estimating the actual colours that an observer recognizes as positive after immunostaining. Further studies are needed to evaluate whether this method is efficient enough to be applied in clinical practice.*

Quantitative image analysis of immunostained formalin-fixed paraffin-embedded tissue sections has so far been performed using different methods and by combining or creating several software packages. The most commonly used method is the calculation of the mean value derived from the red channel of a red-green-blue (RGB) digital microphotograph of a stained section, as analyzed by commercially available software such as Adobe Photoshop (1). Using the Color Histogram module and the statistics of the red color distribution, the specimens under examination are compared. The higher the mean intensity on a scale of 0 to 255, the

more positive the specimen is. Variations of this method have been presented using spectral analysis (2) and composite mathematical calculations (3). However, it is difficult for these approaches to serve the everyday demands of a pathology laboratory due to their complexity, and they thus remain inapplicable. On the other hand, automated systems, supported by relevant software (for example ACIS), promise rapid and quantitative evaluation of large samples of immunostained sections. Nevertheless, this kind of systems is quite expensive for a low budget laboratory, not to mention the need for extremely large samples in order for the whole setup to be cost-effective.

Every colour in the RGB image mode has a unique identity consisting of the colour intensity of each of the three basic colours (red, green and blue). The intensity scale is defined from 0 (darkest) to 255 (brightest). Real monochromes are simply the three basic colours having RGB codes that are null for the other two basic colours; for example pure red is represented by 255 colours, from code [R0, G0, B0] to code [R255, G0, B0]. However humans perceive as red other colours as well, which have a high red intensity, for example code [R200, G50, B0] or [R200, G0, B50]. On the contrary, some colours with some red intensity are perceived as completely different colours; for example [R100, G150, B255] is recognized as cyan.

Based on the fact that when using 'red' substrates such as fast red, positive immunostaining is usually visualized as a red colour, we present a modification of the red channel analysis of RGB digital images that is easy to perform. The method is demonstrated in breast cancer tissue samples, showing its efficiency in evaluating both cytoplasmic and nuclear staining.

## Materials and Methods

*Software.* For image analysis the freeware ImageJ v1.33 as well as the Colour Histogram plug-in, both downloaded from the NIH website (<http://rsb.info.nih.gov/ij>) were used. Initial colour analysis regarding recognition of the range of the red colour conceived as positive in immunostaining was performed using Adobe® Photoshop® 7.0 (Adobe Systems Incorporated, U.S.A.).

\*These authors share senior authorship.

*Correspondence to:* O. Zoras, Associate Professor in Surgery, Department of General Surgery, University Hospital of Crete, 71 110 Heraklion, Crete, Greece. Tel: +30 2810392392, +30 2810392676, Fax: +30 810392392, e-mail: ozoras@med.uoc.gr

*Key Words:* Image analysis, immunohistochemistry, breast cancer.

*Method development.* By using the colour selection module of Adobe® Photoshop® 7.0, a diagram of the RGB distribution on a simplified colour spectrum (derived from a hue-saturation-intensity (HSI) model when S=100% and I=100%) was created, with the aim of presenting a schematic approximation of the way R, G and B intensity values alter. As shown in Figure 1A, the colour area usually seen in positive immunohistochemistry sections using a red chromogen was identified (by consensus of five observers) as a spectrum characterised by the absence of green. A two-dimensional spectrum of the red vs. blue channel values was then created (Figure 1B) in order to show that the red colour recognised by an individual observer is actually an area larger than that defined by a red monochrome. By consensus of five observers, this area was defined as the spectrum extending from R200 to R255 and from B0 to B50.

From all the above, taken together, it can be seen that false-positive or false-negative results may also be produced by using the red channel distribution histogram. This distribution gives information about the number of image pixels that present a certain red channel intensity, rather than a visible red colour. Thus the red channel distribution histogram may take into consideration colours that the observer would definitely recognise as negative. Additionally, even if unquestionable positivity was present, the mean value could be lower than that for the negative control, when the positively stained cells constitute a small percentage of the examined population. In order to exclude such a possibility, instead of considering the mean intensity, we attempted to count the number of image pixels containing visible red colours as presented by the spectrum in Figure 1B. We therefore defined the red intensity index ( $R_i$ ) as the normalized area under the curve of the red channel distribution between the limits R200 and R255. In mathematical terms,  $R_i$  is expressed, by the following formula

$$R_i = \frac{\int_{i=200}^{i=255} f_R(i) di}{A} = \frac{[AUC]_{i=200}^{i=255}}{A} \quad (\text{Equation 1}),$$

where  $i$  is the red channel intensity,  $A$  is the total image area in pixels given by the Colour Histogram module and  $f_R(i) di$  is the function of the red channel intensity distribution.

For the calculation of  $R_i$ , ImageJ v1.33 as well as the Colour Histogram plug-in were used. Every RGB digital image in jpeg format was loaded into the ImageJ platform. The Colour Histogram plug-in was used to obtain the red channel distribution and to extract it to either MSExcel or to commercially available mathematical applications. The line diagram of the intensity values from R200 to R255 was then plotted and the area under curve (AUC) was calculated by applying an integration module on the plotted function. Thus,  $R_i$  is practically expressed as a percentage of the area under analysis, transforming Equation 1 to the following

$$R_i = \frac{\int_{i=200}^{i=255} f_R(i) di}{A} \times 100\% \quad (\text{Equation 1a}).$$

As shown in Figure 1B, red colours are also contained in the spectrum ranging from B0 to B50. Thus, the blue intensity index is defined by:

$$B_i = \frac{\int_{i=0}^{i=50} f_B(i) di}{A} = \frac{[AUC]_{i=0}^{i=50}}{A} \quad (\text{Equation 2}),$$

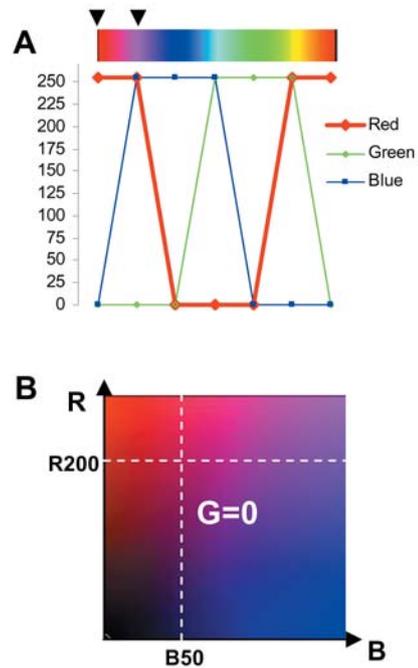


Figure 1. A, Schematic presentation of the RGB distributions. The arrows show the colour area seen in IHC-stained sections characterised by null green colour intensity. B, Two-dimensional diagram of the colour spectrum featuring null green colour contribution. The areas considered positive are designated as being higher than R200 and lower than B50.

where  $i$  is the blue channel intensity,  $A$  is the total image area in pixels given by the Colour Histogram module and  $f_B(i) di$  is the function of the blue channel intensity distribution. AUC calculation is performed by integrating the blue channel intensity distribution, as this is given by the colour histogram module, from B0 to B50.

*Immunohistochemistry.* Formalin-fixed paraffin-embedded breast cancer tissue sections from the archives of the University Laboratory of Pathology (Heraklion, Greece) stained for cytokeratin 19 (cytoplasmic reactivity) and estrogen receptors (nuclear reactivity) were used. Sections were photographed under the same magnification using a Nikon E80i microscope and a Nikon E80i digital camera and the obtained images were saved in jpeg format. Following the procedure described above (Table I), one positive sample as identified by two observers and its negative control (the same sample stained by substituting primary antibody with TBS) for each antibody were analysed and both their  $R_i$  and  $B_i$  were calculated.

*$R_i$  calculation using different integration limits.* In order to minimize background values, the  $R_i$  was calculated again, using higher than R200 values as lower integration limits, i.e. R210, R220, R230, R240 and R250. These calculations transformed Equation 1a to:

$$R_i = \frac{\int_{i=x}^{i=255} f_R(i) di}{A} \times 100\% \quad (\text{Equation 3}),$$

where  $x$  is the intensity defining the lower limit of the integration. Based on the above results presented in Figure 1B,  $x$  cannot take values less than 200.

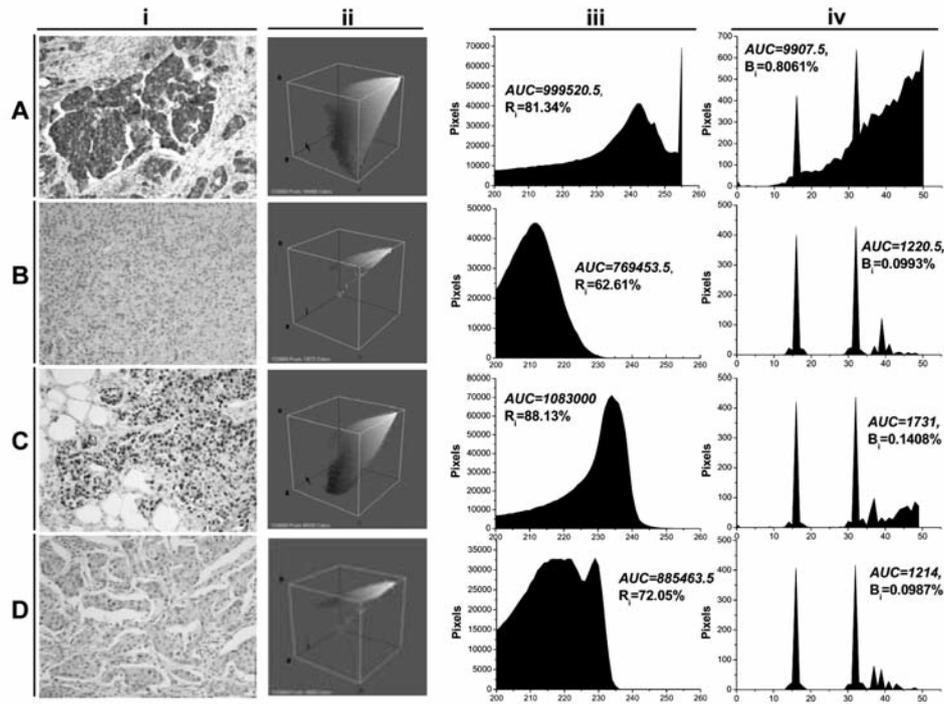


Figure 2. Calculation of  $R_i$  and  $B_i$ . Cytokeratin 19-positive (A) and negative (B), as well as estrogen receptor-positive (C) and -negative (D) sections (column i) were analysed and  $R_i$  (column iii) and  $B_i$  (column iv) were calculated. Colours of each section were also analysed by the 3D Colour Inspector module (column ii) giving a first qualitative approach.

## Results

As shown in Figure 1B, positive staining refers to a specific spectrum of colours ranging from red to purple. This is acceptable, since in several cases presenting weak positive reactivity, low red intensity is mixed with the existing blue of the haematoxylin counterstain, leading to colours near to purple. It is clear from Figure 2 that  $R_i$  and  $B_i$  for positively stained sections are higher than those of negatively stained sections. However,  $B_i$  gives extremely low values and thus evaluation using  $B_i$  is difficult and therefore not recommended. By altering the integration interval that defines the  $R_i$ , the background values of the negatively stained sections become minimal, whereas the difference between positive and negative sections seems to maximize (Figure 3). The minimum intensity, where background is minimal or null, is that recommended to be further used in large sample series.

## Discussion

Analysis of immunohistochemistry images is still a problem faced by the pathologist. Several models have been presented using different systems of image colour modes. Apart from RGB, there are reports using the cyan-magenta-yellow-black (CMYK) system (4), and the hue-saturation-intensity (HSI)

Table I. Brief description of the method.

- Acquire the micrograph in jpg format
- Load the micrograph file into the ImageJ platform
- Apply the Colour Histogram module to obtain the total pixel count of the image (designated as A in the  $R_i$  formula)
- Press the “List” command on the Colour Histogram module
- Copy the values of the Red Channel to mathematical software
- Integrate the distribution between values 200 and 255 (Designated as the AUC in the  $R_i$  formula)
- Calculate the  $R_i$  accordingly

system (5). All studies report a satisfactory result in discriminating the chromogen from the background. Most of them, however, use rather sophisticated software, a fact that makes it difficult for everyday practice.

The RGB-based methods presented so far use the mean red channel intensity to estimate staining reactivity (6). In order to capture the signal of the red colour, several techniques have been invented. The most common method is the application of software to isolate regions of interest and then to perform image analysis (7). This kind of approach is mainly applied in nuclear staining evaluation, since nuclei are easy to distinguish. However, in cases of cytoplasmic staining, by isolating several image regions, colours other than red may interfere with the result. On the other hand,

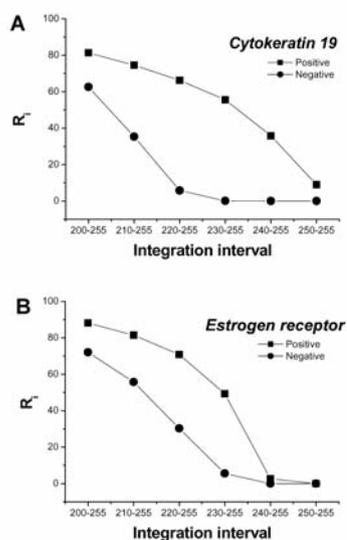


Figure 3. Background values are minimized by reducing the integration interval.

cropping objects from an original picture by isolating the background may lead to underestimation as well, since weak positivity visualized as purple may be considered as background. Moreover, normalization according to image size has been omitted so far, ignoring the fact that the simple red channel distribution is directly related to image size, since the AUC from 0 to 255 is image size dependent.

Our technique overcomes these drawbacks. Firstly, it is applicable to both cytoplasmic and nuclear staining without interfering with any part of the image. Normalization ensures that image size is not an interfering factor. Our method, especially when using  $B_i$ , may also overestimate some colours close to blue. Nevertheless, this does not alter the comparative result because the same method is applied to all the sections and the overestimation affects all the samples. This overestimation disappears when  $R_i$  calculation stringency is increased by changing the lower integration limit to a higher value than  $R_{200}$  (Figure 3).

A restriction of our method may be the need for immunostaining all the sections at the same time, in order to avoid staining and thus  $R_i$  variability. This problem may be overridden by automated immunostaining, using the same control in each batch and then normalizing the results between batches using the  $R_i$  of the control sections.

Additionally, in this way, several obstacles related to RGB image capture are also solved. It is already known that RGB image quality may be affected by the optical system used *i.e.* the light source, the microscope lenses and the camera as well (6). After normalization has been performed some of

Table II. Advantages of the presented method.

- Low cost
- Easy to perform
- Normalized AUC is image resolution independent
- Low variability after normalization of different batches

these parameters and especially the camera, do not interfere with the result. On the contrary, microscopy conditions *i.e.* light source, light intensity and magnification should remain unaltered to obtain comparable images.

In conclusion the suggested technique is simple, sensitive, easy to perform and inexpensive (Table II). A major contribution of this technique is the performance of comparative studies, since the method demands negative control sections to calibrate the integration interval used in the  $R_i$  calculation. Application to large sample series is imperative to evaluate correlation with other generally accepted techniques.

## References

- 1 Lehr HA, Mankoff DA, Corwin D, Santeusano G and Gown AM: Application of Photoshop-based image analysis to quantification of hormone receptor expression in breast cancer. *J Histochem Cytochem* 45: 1559-1565, 1997.
- 2 Ornborg RL, Woerner BM and Edwards DA: Analysis of stained objects in histological sections by spectral imaging and differential absorption. *J Histochem Cytochem* 47: 1307-1314, 1999.
- 3 Zhou R, Hammond EH and Parker DL: A multiple wavelength algorithm in color image analysis and its applications in stain decomposition in microscopy images. *Med Phys* 23: 1977-1986, 1996.
- 4 Pham NA, Morrison A, Schwock J, Aviel-Ronen S, Iakovlev V, Tsao MS, Ho J and Hedley DW: Quantitative image analysis of immunohistochemical stains using a CMYK color model. *Diagn Pathol* 2: 8, 2007.
- 5 Ruifrok AC, Katz RL and Johnston DA: Comparison of quantification of histochemical staining by hue-saturation-intensity (HSI) transformation and color-deconvolution. *Appl Immunohistochem Mol Morphol* 11: 85-91, 2003.
- 6 Taylor CR and Levenson RM: Quantification of immunohistochemistry issues concerning methods, utility and semiquantitative assessment II. *Histopathology* 49: 411-424, 2006.
- 7 Mofidi R, Walsh R, Ridgway PF, Crotty T, McDermott EW, Keaveny TV, Duffy MJ, Hill AD and O'Higgins N: Objective measurement of breast cancer oestrogen receptor status through digital image analysis. *Eur J Surg Oncol* 29: 20-24, 2003.

Received July 22, 2009

Revised November 5, 2009

Accepted November 9, 2009