# FISH IMAGE ANALYSIS SYSTEM FOR BREAST CANCER STUDIES

Z.Theodosiou AUTH,Greece | F.Raimondo AUTH,Greece | M.E.Garefalaki AUTH,Greece | G.Karayannopoulou AUTH,Greece | K.Lyroudia AUTH,Greece | I.Pitas AUTH,Greece | I.Kostopoulos AUTH,Greece

## ABSTRACT

HER2/neu gene amplification is being evaluated by fluorescent in situ hybridization (FISH). In order to avoid interobserver variations in the assessment of HER2/neu status, an integrated FISH image analysis system is developed to automate the classification of FISH images from breast carcinomas. Using a two-stage algorithm, for nuclei and dot detection, and combining results from multiple images taken from a slice for overall case classification, FISH signals ratio per cell nucleus were measured and cases were classified as positive or negative. The system consists of functions for red spot detection, green spot detection, nuclei segmentation and FISH signal ratio. Therefore, it provides the capability to manually correct the resulted images after the analysis.

## 1. INTRODUCTION

Oncogenes refer to genes whose activation can contribute to the development of cancer. Activation can occur through gene amplification such that more of the protein encoded by the gene is present, Osborne et al (1). HER-2/neu (C-erbB-2) oncogene is amplified and over-expressed in approximately 20-40% of breast cancers. It is also associated with disease outcome in gastrointestinal, pulmonary, genitourinary and other neoplasms. Over expression of HER-2/neu is usually a consequence of gene amplification, in which multiple copies of the gene appear through the gerome.

HER-2/neu status can be determined by analyzing the numbers of gene copies centrally or the amount of protein peripherically. Fluorescence in situ Hybridization (FISH) and immunohistochemistry (IHC) are currently regarded as the standard screening techniques. When a standardized IHC assay is performed on specimens that are carefully fixed, processed and embedded there is an excellent correlation between gene copy status and protein expression levels. Advantages of IHC testing include its wide availability, relatively low cost, easy preservation of stained slides, and the use of a familiar routine microscope. Disadvantages of IHC include the impact of preanalytic issues including storage, duration and type of fixation, intensity of antigen retrieval, type of antibody (polyclonal versus monoclonal), nature of system control samples, and most importantly, the difficulties in applying a subjective slide scoring system Ross et al (2). FISH is a technique that can directly identify a specific region of DNA or RNA in a cell and therefore enables several new areas of cytogenetic investigation by allowing visual determination of the presence and normality of specific genetic sequences in single metaphase or interphase cells. FISH testing for c-erbB-2 should meet the following criteria: (a) the inclusion of a chromosome 17 control to allow for correction of the HER2 signal number for chromosome 17 aneusomy (seen in over of 50% cases); (b) comprehensive standarization of methodology; and (c) validated controls, Ellis et al (3). The FISH technique, which is morphological driven and can be automated, has the advantages of a more objective scoring system and the presence of a built-in internal control consisting of the two c-erbB-2 gene signals present in all non-neoplastic cells in the specimen. Fluorescence *in situ* hybridization is a direct *in situ* technique that is relatively rapid and sensitive. No cell culture is needed in order to apply this method and results are easier to interpret than karyotype. FISH method can be combined with immunostaining. Disadvantages of FISH testing include the high cost of each test, the long time needed for slide scoring, the requirement for a fluorescence microscope, the inability to preserve the acquired image for storage and review, and, occasional a difficulty in identifying the invasive tumour cells (2). The main disadvantage of the FISH is the cost which is ten times higher than the cost of IHC , Sauer et al(4).

Several methods have been proposed for the automated evaluation of FISH signals, even though they were not applied directly for measuring Her-2/neu gene amplification of breast samples. Most methods focused on automatic spot counting whereas only very few focused on case-based classification of FISH images.

Netten et al (5) focused on automatic counting of dots per cell nucleus in slides of lymphocytes from cultured blood. Solorzano et al (6) developed a method to study leukocytes in blood samples. Kozubek et al (7) developed a system that acquired 2-D and 3-D FISH images and performed image analysis on both. Lerner et al (8, 9) proposed a FISH image classification system based on the properties of in- and out-of-focus images captured at different focal planes. The signals were classified as real or artifacts and the images that contained no artifacts were considered to be the in-focus image. This methodology is in contrast with the

methods described above that rely on auto-focusing mechanisms. Recently, Chawla et al (10) developed an automated system for analyzing FISH signals from brain hippocampal and cortical sections. Their objective was to examine temporal gene transcription activity for which manual counting was time-consuming considering that a stack of images had to be examined Based on the above, there seems to be a potential for further development of systems for the automated case-based reading of FISH images, particularly for the application of HER-2/neu evaluation in breast carcinomas samples. Such a system should take into account multiple images of a specific case and quantify the HER-2/neu status in a collective manner. In this paper a toolbox for the automated classification of FISH images from breast carcinomas samples is presented.

## 2. FISH images

We employed four patient cases, two of which were previously classified by an expert as positive and two that were classified as negative, in order to evaluate the precision of the system. The breast tissue slides were prepared using the following procedure.

Paraffin sections of 4μm thickness were incubated overnight at $60^{o}$ C. Deparaffinization, pretreatment; enzyme digestion and fixation of slides were performed using the Vysis Paraffin Pretreatment kit according to the manufacturer's recommended protocol. Slides were deparaffinized in xylene, dehydrated in 100% ethanol and immersed in pretreatment solution. Proteolysis of neoplastic cells was performed by immersing the sections in protease solution at 37 $\hat{u}$ C for 12 minutes. Tissue sections were denaturated at $85^{o}$ C for 2 minutes, then the PathVysion HER-2 DNA Probe (LSI HER-2/CEP17 probe, Abbott GmbH and Company, KG, Wiesbaden-Delkenheim, Germany) was added and hybridization took place at 37 $\hat{u}$ C in a moist chamber for 14-18h (overnight incubation). The following day the slides were washed with post-hybridization buffer (2X SSC and 0,3% NP-40) at $72^{o}$ C for 2 minutes, followed by counterstaining of the nuclei with 4, 6-diamino-2phenylindole dihydrochloride (DAPI).

For each case, at least 60 non overlapping nuclei were scored for both Her-2/neu (red spot) and chromosome 17 (green spot) signals by image analysis. Hybridization signals were enumerated utilizing a Zeiss, Axioskop 2 microscope equipped with a 100 Watt mercury lamp (HBO 100) and an automatic filter wheel system with the following filters: BP360/51 DAPI filter, BP485/17 FITC filter-spectrum green, BP560/18 Rhodamine filter-spectrum orange. Plan-Neofluar lens with magnification of x100, NA=1.3 and a pixel size of 0.24 μm were used when reading the images, along with manual focusing. Images were grabbed using a CV-M300 2/3 CCD camera (JAI, Copenhagen, Denmark). The camera had a high S/N ratio of >58 db and an effective pixel resolution of 752 (horizontal) x 582 (vertical). During clinical reading, the images were processed using the Meta Systems software (Altlussheim, Germany) in order to adjust contrast in the different color channels. This software contains a shading correction algorithm to account for non-uniform illumination. Her-2/neu gene amplification was determined by a ratio of Her-2/neu gene copies to chromosome 17 centromeres. According to the manufacturer's recommendations the cases with a ratio $\geq 2$ were determined as amplified, while those having a ratio $< 2$ as not amplified.

## 3. IMAGE ANALYSIS SYSTEM

The FISH image analysis system, hereafter simply called the System, is a module for to the Volumetric Image Processing, Analysis and Visualization software package, EIKONA3D for Windows. Accordingly it is supported with a user interface, making it very practical and simple to use.

The System (Fig.1) contains functions for red spot detection, green spot detection, nuclei segmentation and automated FISH signal ratio, by employing multistage algorithms for spot detection and nuclei segmentation. Initially, a sequence of images can be loaded into the input volume and then according to the function of choice, the system implements the corresponding algorithm and displays the results on the screen thus giving the advantage of correcting the results manually in a step by step fashion.

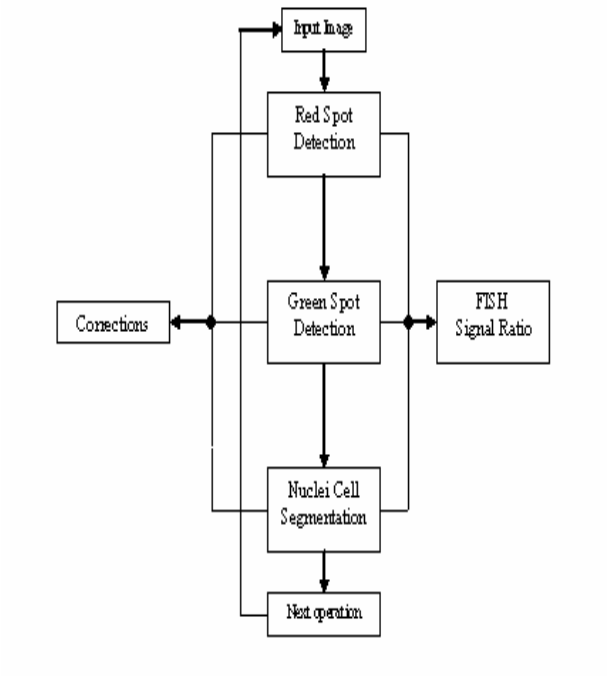


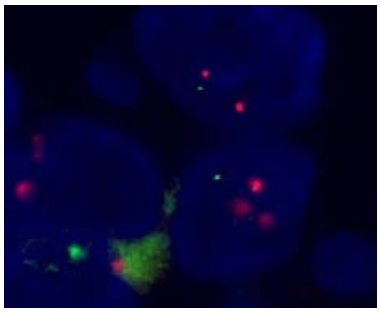

**Figure 1.** The structure of the system

Using the result from each step, the System can calculate the FISH signal ratio.

### 3.1 Red/Green Spot Detection

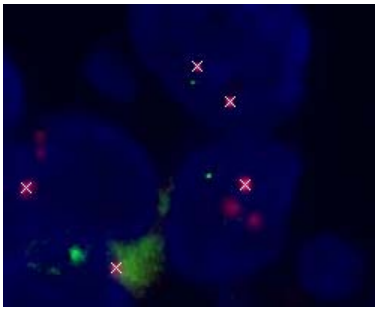
In case of calling the Red Spot Detection and Green Spot Detection function, the system implements the spot detection algorithm for red and green spots, respectively. The algorithm keeps the pixel's coordinates that contain a spot, and then draws a cross in the position of each spot that is detected. The first processed image illustrates only the crosses on the corresponding spot detected positions. Combining the first processed image with the input image, the system creates the final processed image of red/green spot detection function. This image illustrates the crosses on the corresponding detected spot. During this procedure, the user can check the precision of the spot detection and correct the errors manually.

The user has the opportunity to correct manually the processed image, using the peripheral components of the computer. If the user wants to delete or draw a cross, the specific location can be defined using the mouse cursor and then by pushing the "D" button of the keyboard and left button of the mouse, in the same time, a cross will appear or disappear on this position.



(a)

(b)

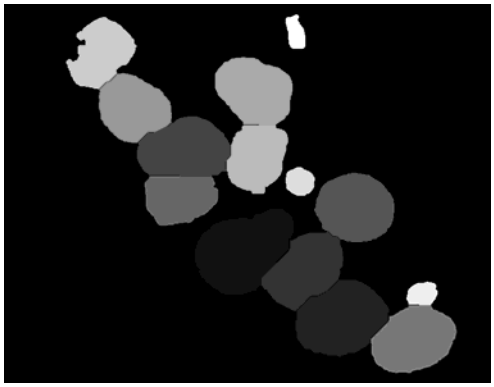


(c)

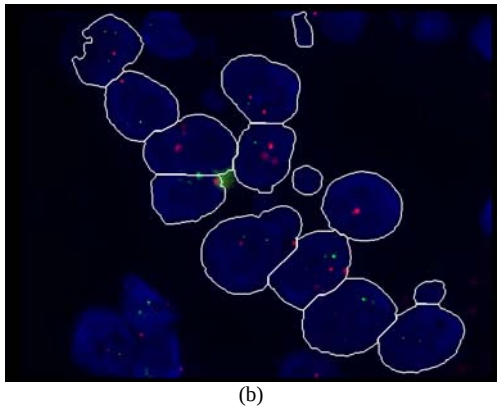
**Figure 2:** Red spot detection **(a)** the input image; **(b)** the first processed image; **(c)** the final processed image.

### 3.2 Nuclei Segmentation

In the same way, if the user chooses the nuclei segmentation function, the system implements the nuclei segmentation algorithm. The first processed image represents the segmented nucleus with different shades of grey colour corresponding to each nuclei. Combining the input image with the perimeters of the segmented nucleus, the final processed image of nuclei segmentation function can be created. The final image provides the precision of the nuclei segmentation. In case of indiscretion, the user can split or merge the erroneous segmented nucleus. If two separate nuclei are represented as one in the processed image, they can be split by following the procedure described next. We press the "S" button (inform the system for splitting procedure) and using the mouse we choose two points. Then, the system calculates the mathematical equation of the straight line that passes through the two clicking points. The nucleus will be split into two different nuclei based on this straight line.



(a)



(b)

**Figure 3:** Nuclei segmentation **(a)** first processed image ;**(b)** the final processed image.

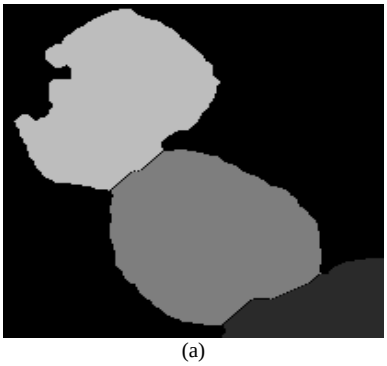
In the same way, if one independent nucleus is represented as two separate nuclei in the processed image, we press the "M" button (inform the toolbox for merging procedure) and then we choose the two nuclei by clicking the left and right button of the mouse. The second nucleus is painted with the color of the first nucleus, making the two appear as one nucleus.

### 3.3 FISH signal ratio

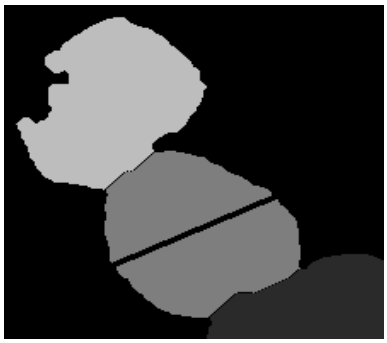
The system can calculate the FISH signal ratio, by choosing the ratio function. In this case the system implements the spot detection algorithm for red and green spots, respectively. Furthermore, it implements the nuclei segmentation to detect the segmented nucleus. For every segmented nucleus where at least one red spot is present, a ratio is calculated, defined as:

$$ratio = \frac{N_R}{N_G}, \quad (1)$$

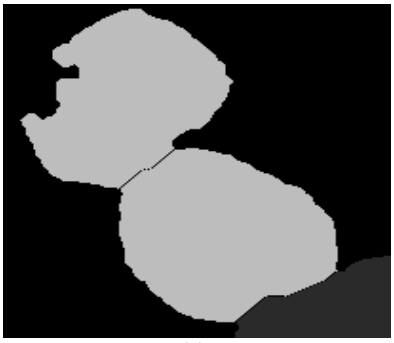
Where $N_R$ and $N_G$ are, the number of red and green spots that are presented in the segmented nucleus, respectively. In the instances where the number of green spots is zero, and the number of red spots is nonzero, it is assumed that at least one red spot is present. This procedure is implemented for every sequence of image and calculates the average ratio for the valid nucleus (they have at least one red spot) by associating all the images.



(a)



(b)



(c)

**Figure 4 :**Nuclei correction **(a)** The image before the correction; **(b)** The nuclei splitting **;(c)** The nuclei merging.

## 4. MULTISTAGE ALGORITHM

The classification is based on the accurate measurement of red/green spot ratio per cell nucleus, so the system employs a multistage algorithm, combining two stages for spot detection and cell nuclei segmentation respectively.

### 4.1 FISH Spot Detection

The spot detection stage is applied on red and green channels of FISH image and includes mainly stages of top-hat filtering, binary thresholding, grey level template matching and contrast evaluation. It begins with top-hat filtering, using a disk of 4-pixel radius as structuring element, for noise removal. A modification of the algorithm proposed in (10) was used to estimate two thresholds for the top-hat, red and green channel output respectively.

Although red and green spots usually have the greatest channel intensity, it is likely that many valid spots have red/green level value smaller than that of false signals. For better spot detection, the characteristic grey level trend of every spot is employed, to compare the grey level "shape" of every candidate spot with the spot shape template obtained from the average shape of a set of valid labeled spots. The normalized cross correlation (11) is used for the measurement of the similarity between every candidate spot and the spot template. For the estimation of the spot shape template, a 7x7 window positioned on every spot center is saved as a template for every red and green channel respectively. Two spot template windows $T_G$ and $T_R$ are estimated by averaging the respective spot channel intensities:

$$T_R(x,y) = \frac{1}{N_R}\sum_{i=1}^{N_R} f_{R_i}(x,y), \quad (2)$$

$$T_G(x,y) = \frac{1}{N_G}\sum_{i=1}^{N_G} f_{G_i}(x,y), \quad (3)$$

Where $N_R$ and $N_G$ are the number of used red and green spots, $x = 1,...,7$ and $y = 1,...,7$ are coordinates in a 7x7 window and $f_{R_i}$ and $f_{G_I}$ are red and green intensities of the $i-th$ spot image. For each new test spot image, the normalized cross correlation $C_R$ between $T_R$ the respective channel intensity $I_R$, is calculated as follows:

$$C_R(u,v) = \frac{\sum_{u,v}[T_R(x,y)-\bar{T}][I_R(x-u,y-v)-\bar{I}_R(u,v)]}{\{\sum_{u,v}[T_R(x,y)-\bar{T}]^2\sum[I_R(x-u,y-v)-\bar{I}_R(u,v)]^2\}^{0.5}}, \quad (4)$$

Where $\bar{T}_{R_{u,v}}$ is the mean value of $T_R$, while $\bar{I}_R(u,v)$ is mean value of red channel $I_R$ around the 7-pixel neighborhood of pixel $(u,v)$. The normalized cross correlation $C_G(u,v)$ for the green channel is computed in a similar fashion. The selection of red/green spot positions is accomplished, using two thresholds $Th_R$, $Th_G$. Spots with values of $C_R$ and $C_G$ lower than $Th_R$, $Th_G$ respectively, are discarded.

Finally, a further processing step, a contrast measure, is performed to discard spots that have similar shape to the template and low channel intensity contrast with respect to their surrounding pixels, making them invisible to the human eye. The contrast measure is performed with the follow method. Two vectors $v_{for}$ and $v_{back}$, are created for each spot, that consist of three values, using the information of the red, blue and green channel. For $v_{for}$, each value is estimated calculating the average channel intensity of the pixels of a 5x5 window positioned on every spot center. The corresponding three values of $v_{back}$ is estimated calculating the average channel intensity of the background pixels around binary object perimeter. The contrast measure $C_M$ is calculated as follows:

$$C_M = \frac{\left\| v_{for} - v_{back} \right\|}{\left\| v_{back} \right\|} > T_{C_M} \quad (5)$$

The, $Th_R$, $Th_G$ and $T_{C_M}$ are empirically estimated.

### 4.2 Cell Nuclei Segmentation

The second algorithm is used for the cell nuclei segmentation, consists of a non-linear blue channel correction step, a global thresholding by Otsu algorithm (12), a grey level hole classification by a geometric rule and of the marked watershed transform using local h-dome maxima as markers.

In order to reduce the gray level difference between dark regions and more illuminated ones on the blue channel, a nonlinearity correction step was performed applying the square root function to the blue channel, normalized by its maximum grey level $p_{MAX}$, as shown in Eq.6.

$$p_{out} = \sqrt{\frac{p_{in}}{p_{MAX}}} \quad (6)$$

Where the $p_{out}$ and $p_{in}$ are the pixels grey levels. By applying the opening morphological operator, to the image blue channel, using a disk of 4-pixel radius as structuring element, the intensity of the gray level peaks due to the presence of spots is reduced. A top-hat filtering using an 80-pixel radius disk as structuring element is also applied to reduce the blue channel intensity of regions where color diffusion was caused from non-ideal staining.

The Otsu algorithm is employed to determine the threshold for the initial nuclei segmentation. The resulted image contains holes in a single nucleus body region which they have to be filled and holes in inter nuclei zones of overlapping nuclei which they should not be filled. In order to fill the holes of first type, a geometric approach is used.

The last step of the nuclei segmentation algorithm is consisted of the marked watershed algorithm (13) using local h-dome maxima as markers (14), which is employed to detect borders in overlapping nuclei clusters.

## 5. RESULTS AND DISCUSSION

The classification of FISH images demands at least 60 non overlapping nuclei for every case, which are scored for both Her-2/neu (red spot) and chromosome 17 (green spot) signals. Cases that have FISH signal ratio lower than two are classified as negative. Otherwise they are classified as positive.

In order to evaluate the precision of the system, we employed four patient cases, two of which were previously classified by an expert as positive and two that were classified as negative.

TABLE 1. The Four cases which were used for the evaluation

| | # images | Classification |
|---|---|---|
| Case #1 | 15 | negative |
| Case #2 | 15 | negative |
| Case #3 | 15 | positive |
| Case #4 | 15 | positive |

The number of images available, the values of ratio and classification by the system for each of the testing cases are listed in table 2.

TABLE 2. Number of available images for each case, the FISH signal ratio and the classification for each of the testing cases.

| | Ratio | classification |
|---|---|---|
| Case #1 | 1.1912 | negative |
| Case #2 | 1.1774 | negative |
| Case #3 | 2.1488 | positive |
| Case #4 | 2.1551 | positive |

It can be seen from Table 2 that all cases can be correctly classified as either positive or negative. Despite the small number of cases, these preliminary results are encouraging for the further testing of the system in clinical trials. A larger database of FISH images of breast tissue is being prepared in order to

examine how well these results can generalize in a broader population.

## 6. CONCLUSION

In this paper a system for image analysis and automated evaluation of HER-2/neu status with FISH images was presented. The user can employ this system for analyzing the FISH images and measuring FISH signal ratio per cell nuclei, having the opportunity of correcting manually the results from spot detection and nuclei segmentation before the ratio measurement. The performance for the case-based classification on the 4 testing FISH cases showed the ability of the system to distinguish between all positive and negative cases. Despite the small number of cases, the classification results were encouraging for the further testing of the method in clinical trials. Its usage can simplify the classification of FISH images, relieving the scientists from the difficult identification of the invasive tumour cells and the time-consuming scoring of the slide.

## ACKNOWLEDGMENT

This work was supported by the EU project Biopattern: Computational Intelligence for biopattern analysis in Support of eHealthcare, Network of Excellence Project No. 508803.

## REFERENCES

1. Osborne C., Wilson P., Tripathy D., 2004, Oncogenes and Tumor Suppressor Genes in Breast Cancer: Potential Diagnostic and Therapeutic Applications. The Oncologist, 9, 361-377

2. Ross J.S., Fletcher J.A., Linette G.P., Stec J., Clark ED, Ayers M., Symmans W.F., Pusztal L., Bloom K.J., 2003, The HER-2/neu Gene and Protein in Breast Cancer 2003: Biomarker and Target of Therapy, The Oncologist, 8, 307-325

3. Ellis I.O., Dowsett M., Bartlett J., Walker R., Cooke T., Gullick W., Gusterson B., Mallon P., Barrett L. Recommendations for HER@ testing in the UK.

4. Sauer T., Wiedswang G., Boudjema G., Christensen H., Kåresen R., 2003, Assessment of HER-2/neu overexpression and /or gene amplification in breast carcinomas: should in situ hybridization be the method of choice? APMIS, 111,444-50

5. Netten H., Young I. T., J.van Vliet L., Tanke H. J., Vroljik H, Sloos W. C.R., 1997, FISH and Chips: Automation of fluorescent dot counting in interphase cell nuclei, Cytometry, 28(1), 1-10

6. Solorzano C. O., Santos A., Vallcorba I., Garcia-Sagredo J-M, Pozo F., 1998, Automated FISH spot counting in interphase nuclei: Statistical validation and data correction, Cytometry, 31(2), 93-99

7. Kozubek M., Kozubek S., Lukasova E., Mareckova A., et. Al, 1999, High-resolution cytometry of FISH dots in interphase nucleus nuclei, Cytometry, 36(4), 279-293

8. Lerner B., 2004, Bayesian fluorescence in situ hybridisation signal classification, Artificial Intelligence in Medicine, 30(3), 301-316

9. Lerner B., Clocksin W.F., Dhanjal S., Hulten M.A., Bishop C.M., 2001, Automatic signal classification in fluorescence in situ hybridization images, Cytometry, 43(2), 87-93

9. Chawla M. K., Lin G., Olson K., Vazdarjanova A., Burke S. N., et. Al., 2004, "3D-catFISH: a system for automated quantitative three-dimensional compartmental analysis of temporal gene transcription activity imaged by fluorescence in situ hybridization, Journal of Neuroscience Methods, 139, 13-24

10. Rosin P.L., Unimodal thresholding, 2001, Pattern Recognition, 34(11), 2083-2096

11. Gonzales R.C., Woods R.E., 2003, "Digital Image Processing", Addison Wesley, 2nd edition

12. Otsu N., 1979, A thresholding selection method from graylevel histogram, IEEE Trans. on Systems, Man and Cybernetics, 9 (1), 62-66

13. Beucher S., Meyer F., 1992, The morphological approach to segmentation: the watershed transformation, in E.R. Doughterty (Ed.), Mathematical morphology in image processing, New-York: Dekker

14. Vincent L., 1993, Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms, IEEE Trans. Image Processing, 2 (2), 176-201