

Automatic 3D analysis of genes localization in multi-FISH images

Michael Gué^a and Thomas Boudier^a

^aInstitut Curie Section Recherche, U759 INSERM - Laboratoire Imagerie Integrative, Centre Universitaire, Bâtiment Raymond Latarjet, 91405 Orsay CEDEX, France.

ABSTRACT

3D-Fish is a technique that permits to visualize 3D positions of specific genes in cell nuclei. However the manual analysis of such images is complex, it requires the 3D localization of small spots in noisy images. Furthermore in order to obtain valid statistical results, thousands of images must be analyzed. We developed a software called *Smart3D-Fish* in order to automate the detection of spots (corresponding to genes labeled with FISH probes) and to calculate distances between corresponding genes. The segmentation of spots is based on various parameters. *Smart3D-Fish* can handle any number of fluorescent spots and channels, it can also incorporate the images of counterstained nuclei which can be used, for example, to normalize all measured distances according to the nuclear volume. It allowed us to analyze the relative position of genes involved in leukemia cells. The software was extended to allow the analysis of multi-labeled genes through a co-localization procedure, enabling the simultaneous analysis of seven genes in cell nuclei from three fluorescent dyes. Last it allows the 3D registration of segmented spots positions between images allowing an extended genomic analysis.

Keywords: fluorescence, optical microscopy, 3D-FISH, image processing, segmentation, ImageJ.

1. INTRODUCTION

Cell imaging techniques are integral to successful life science studies. Fluorescence In Situ Hybridization (FISH) is used to study the organization and positioning of chromosomes, subchromosomal regions, and/or specific sequences in cellular preparations by hybridization of probes with complementary DNA sequences in cells. A protocol of cell fixation has been developed to perform FISH in the three-dimensionally (3D) preserved structure of cell nuclei. This technique, named 3D-FISH, has opened the way to study the organization of the chromatin in interphase nuclei.¹⁻³

Relative positions of genes inside nuclei and their dynamics in interphase cells could play a major role in chromosomal rearrangements. For instance, some studies about chromatin organization in interphase nuclei have indicated that the spatial proximity of genes could be a key factor in the induction of translocations.⁴⁻⁸ Furthermore, the differences in relative position of chromosome territories in different cell-lineages could contribute to the large variety of translocations, especially in hematological cancer.⁹

To further address this relationship, the relative spatial distribution of genes potentially involved in chromosomal translocations must be accurately analyzed. In a 3D-FISH experiment, the images analysis can be a heavy workload and complex task for the biologist. In order to obtain reliable results multiple steps are required : manual cropping for selecting regions of interest, manual segmentation of spots corresponding to the localization of genes and finally, computation of the 3D distances between the spots. Further, thousands of image stacks need to be analyzed for reaching statistical significance, the analysis can be hence very time consuming.

The goal of the present work is to provide an user-friendly software in order to improve the quality of rather high through put analysis of 3D-FISH images, and the subsequent positioning of genes and measurement of intergenic distances. With the help of this software, named *Smart3D-FISH*,¹⁰ image stacks corresponding to thousands of cells can be automatically analyzed overnight on a standard desktop computer. It can analyze

Send correspondence to T.B.

T.B.: E-mail: thomas.boudier@snv.jussieu.fr, Telephone: +33(1) 69 86 31 72

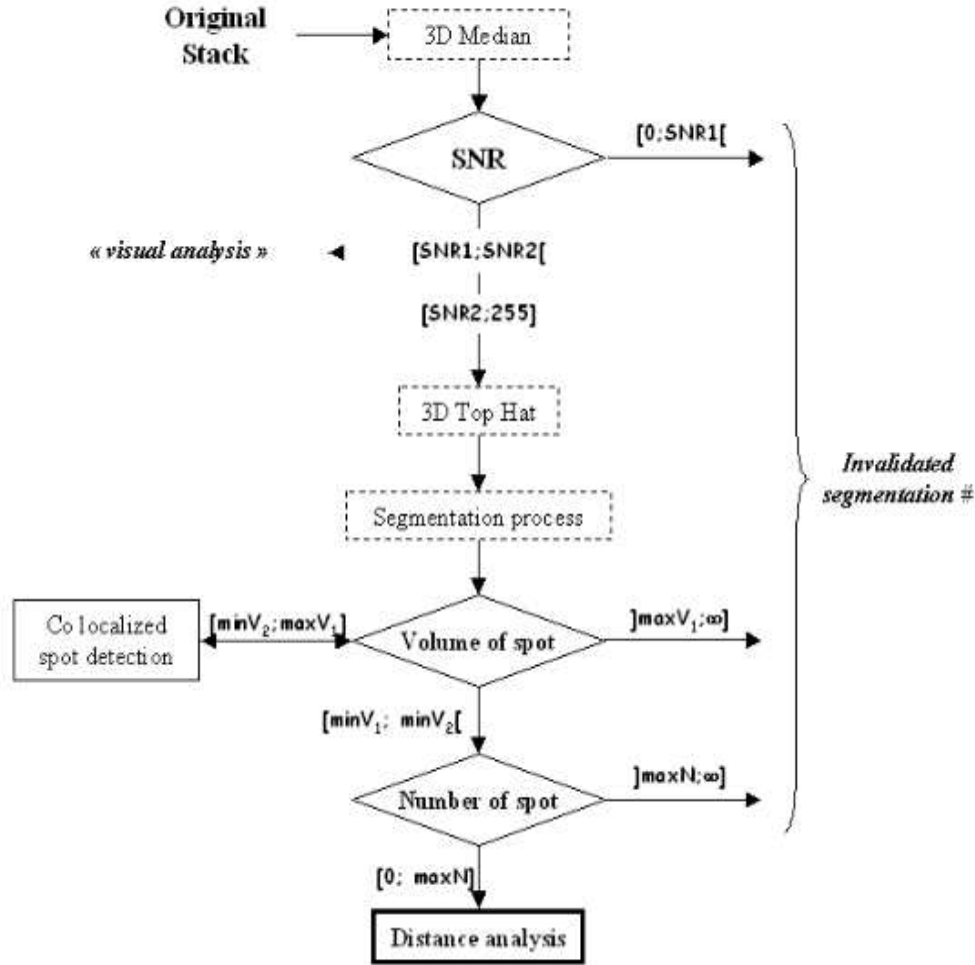


Figure 1. The flow chart of the main part of our software that will segment and analyze spots in a 3D-FISH image. The images are first filtered by a 3D median filter and a 3D TopHat filter. Spots are then segmented and separated if needed. Whether the segmentation leads to results that are in accordance with user parameters or not, the results are saved into different files.

images with a virtually unlimited number of color channels (probes) and spots. The results are saved into text files that can be directly incorporated into standard spread-sheets software.

This new software has been successfully applied in a 3D-FISH experiment aimed at measuring the 3D radius distribution and intergenic distances of genes involved in acute leukemia, namely, MLL, AF4 and ENL, in two lymphoblastic cell lines.¹¹ *Smart3D-FISH* is available as a set of plugins for the image analysis ImageJ software.¹² It can be freely downloaded at the following URL: <http://www.snv.jussieu.fr/~wboudier/softs.html>.

2. MATERIAL AND METHODS

Various plugins were developed for a complete and automatic analysis of 3D FISH experiment images, starting from the automatic cropping of cells in wide-field images coming out a fluorescent microscope, and the analysis of spots distributions to the 3D registration of spots in order to obtain an average spots distribution.

2.1. Cropping of cells

The first task, after acquisition on a 3D fluorescent microscope is to localize and extract the various Regions Of Interest (ROIs) corresponding to nuclei. The numerous channels, each one corresponding to a specific probe labeling a specific gene, may be present in one or multiple files. In the case of a single file containing all the channels, the various channels must be split before cropping. The ROIs corresponding to nuclei are segmented by thresholding the projection image of the counterstained nucleus image, the segmented area is then extended by dilatation, a mask is created and will be used for the cropping in other channels. However only ROIs presenting a circularity greater than a threshold are retained (0.65 in practice). A optional watershed can be run to separate two close nuclei. Properties of optical microscopy can induce a shift between the different color channels. The ROIs are then shifted in X-Y and Z in order to align the different channels. Once the ROIs in the different channels have been extracted, they can be processed and analyzed.

2.2. Spots segmentation

Possible procedure for 3D processing is to apply 2D processing in a slice-by-slice manner. However, slice by slice processing tends to ignore weak signals which could be considered as pixels of noise in a 2D slice but which are part of a signal in 3D. Consequently three new classes *Image3D*, *IntImage3D* and *ColorImage3D* were created, and standard and useful procedures were implemented to work with 3D images. A *ColorImage3D* is simply a set of unlimited number of *IntImage3D*.

Figure 1 shows the flow chart of the overall algorithm. The first step consists in removing the background noise corresponding mostly to randomly distributed unspecific labeling. The median filter, well suited for salt and pepper noise was adapted in 3D and was used, in our case, with a neighborhood of radius equal to two pixels. Then a 3D TopHat filter is applied to enhance spots against the background. The second step is the segmentation of spots. A spot is considered as a compact set of pixels with a volume inside a specific range and including at least one bright pixel. A bright pixel is determined to correspond to a pixel whose value is greater than 99.95% of the histogram, *i.e.* with a value greater than v (v defined by : 99.95% of all pixel values are in the range $[0 - v]$). This value of 99.95% of histogram was computed by estimating the average volume of one spot and considering that a stack could comprise up to four spots.

More precisely, the pixel displaying the maximum intensity in the 3D image, but greater than 99.95% of the histogram, is detected. This pixel will act as a seed from where a spot will be segmented. A local threshold is then computed and the seed is extended, by 3D connectivity, to adjacent pixels whose value is greater than this local threshold, in order to form an object. If the final segmented object is touching one border of the image in the X-Y plane, it is removed. These procedures (detection of maximum pixel and 3D connectivity) are applied until no more unsegmented pixels have a value greater than 99.95% of the histogram. The segmentation is performed on all the different stacks except for the counterstained nucleus image.

For the counterstained nucleus image a special segmentation is automatically performed. The radius of the first median filter applied to reduce the noise is increased to a radius of four pixels in order to get a more homogeneous signal. The image is then thresholded using the classical “Isodata” algorithm.¹³

For both spots and nucleus segmentation, a 3D mathematical morphology closing procedure is applied followed by a 3D mathematical morphology opening procedure (radius 2 for gene images, radius 4 for nucleus images) in order to remove very small objects and make objects more compact. A 3D fill holes procedure may also be added to the preceding procedures, especially in the case where the nucleus may display a large inner nucleolus.

2.3. Analysis and validation of the segmentation

A first analysis of the quality of the subsequent segmentation is done after the application of the median filter. If the computed SNR (maximum - minimum value of gray level in the 3D image) is lower than the SNR1 threshold fixed by the user, the measurements will be saved into an invalidated spots file. If the SNR is in the interval $[SNR1, SNR2]$ a “visual analysis” message will be displayed in the results file. A second analysis of the validity of the segmentation is based on the volume and number of objects. The list of segmented objects is analyzed to detect whether the segmented objects correspond to gene spots or not. Segmented objects with volumes lower than the minimal volume $minV_1$, fixed by the user, are removed. In the same way, objects with volumes greater

than the maximum volume $\max V_1$ are also removed. Then, if the number of spots exceeds the value fixed by the user (2 in the present work), the segmentation is considered as invalid and further measurements will be performed but the results will be redirected into the invalidated spots file.

In all cases, results of measurements are saved into text files. Different text files are created, the first text file will store the results for correct segmentation, the others will store the results for incorrect segmentation, that is to say, where one parameter (SNR, volume or number of spot) does not correspond to the values fixed by the user. In this last case, annotations will indicate the nature of the problem.

2.4. Spot separation procedure

The mathematical morphology procedures may not be sufficient to separate two closely located spots. Close spots may be merged into one big spot with a volume roughly twice the average volume of one spot. The user can fix the minimum volume that may correspond to a merged spot ($\min V_2$). The slices of the stack containing the object are projected along the Z axis onto a 2D plane in order to find local maxima that may correspond to the center of the two merged spots. The obtained projection is then smoothed so that local maxima are more easily detected. If two or more local maxima are encountered, they are separated into two clusters using a k-means algorithm. If the center of the two clusters are distant more than a fixed distance, corresponding approximately to the diameter of one spot, the object is separated into two smaller objects. The centers of these two clusters are the new x and y center coordinates of the two new objects. The z value of the two new objects is determined as the centers of the primary segmented object along the Z axis for the x and y coordinates of the two new spots. Pixels of the primary big object are then re-labeled according to their distance to the centers of these two new objects. If the primary object cannot be split into two objects, it is kept in the list of segmented objects but will be annotated as “large spot”.

2.5. Automation for a serie of nuclei

Once the parameters for spots have been determined (volumes and number), one can use them on a serie of cropped nuclei. The directory containing all the nuclei images with the different channels along with the number of the first cell, the total number of cells and the spots parameters are entered in a configuration file. This file can then be processed and the analysis is automatically performed on all the cell nuclei images.

2.6. Spots registration

The results files will display the 3D localization of the centers of segmented spots. It is then possible to analyze the distribution of the spots inside the studied nuclei by registering the positions of spots found. However there are some constraints based on the underlying nuclear architecture, the positions of genes must be registered on a position of the same gene in another image. Since one gene has two loci in a normal cell, there are two possibilities of registration. Therefore when registering N genes we have 2^N possibilities of registration. For each possibility the optimal transformation between the two sets of positions is computed based on a SVD decomposition.¹⁴ The mean square distance is computed between the transformed set of positions in the first cell and the second set of positions in the second cell. The transformation giving the minimal error between the 2^N possibilities is kept as the transformation between the two sets of positions. From these sets of transformation an average distribution of the gene spots can be computed.

2.7. Co-localization procedure and analysis of multi-fish images

In order to analyze a large number of genes labeled with fluorescent probes, the number of fluorescent dyes can be increased, but this may be expensive and the acquisition procedure may be more difficult. An alternative approach is to label a gene with more than one fluorescent dye. With this approach, using three fluorescent dyes (called A, B and C), three genes can be labeled with only one dye, three more genes can be labeled with a combination of two dyes (A-B, A-C and B-C), finally a last gene can be labeled with the three probes (A-B-C). Consequently seven genes can be labeled with three standard fluorescent dyes using this combinatorial approach.

During the spot segmentation of the three images corresponding to the hybridization of the three probes A, B and C, eight spots are detected (in the case there is no strict co-localization of genes); for instance for probe A, we find two spots for the gene labeled with A only, two for the gene labeled with A-B, two for A-C and two

for A-B-C, the same for color B and C. After the segmentation procedure since we have all the positions of the spots in the three images, it is then quite simple to detect if a spot is present in one, two or three images. The co-localization is performed by comparing the bounding boxes of the spots and the mutual intersection volume that should be more than a certain percentage of each volume (in practice 60%). The spots present in more than one image are then removed from their original image and copied into a “virtual” image corresponding to a “virtual” dye corresponding to the combination of two or three dyes. We hence end up with seven images containing each one two spots and a normal analysis can then be run.

3. RESULTS AND DISCUSSION

The various developed algorithms are available as plugins and macros for ImageJ. The first task, concerning the cropping of cell nuclei, was implemented as a macro that crops the nuclei based on the projection of the stained nucleus images. The main plugin called *FISH3D* will process and analyze the images, ask for the various parameters needed, analyze the images and write the results into text files. Actually three different text files are created : the first one where all the parameters are respected, one (with a *BAD* extension) will comprise the segmentation presenting a default, either the SNR is under the threshold, the number or volumes or spots are not correct, and a third file (with a *UNDER* extension) with all the segmentation where only the number of spots is under the correct value and that may correspond to the co-localization of two spots. For the automation of the analysis there are two other plugins called *FISH3DConfig* that will help the user create the configuration file and *MFISH3D* that will run the analyzes on all the images mentioned in the configuration file.

Our software, run on more than 1000 images, gives good results since about 98% of images are correctly segmented in the correct results file. In the *BAD* results file most of images are correctly segmented but are not in accordance with the user requirements, these images usually reflects a problem in the experimental or acquisition procedure. Some of the data saved in the *BAD* results file can be however retrieved and added to the final analysis. These data correspond to images presenting a low SNR (under the SNR1 value) but with spots correctly segmented and having required volumes. The results in the *UNDER* file concerns mainly the segmentation that lead to the segmentation of only one spot (assuming two spots for correct nuclei) and usually the two spots were really co-localized and in a small fraction of cases the two spots were close but the spot separation procedure was not able to separate them.

We have applied our program to the analysis of distances in leukemia cells nuclei to analyze the distance between three genes *MLL*, *AF4* and *ENL* that are involved in 11q23 translocations. The results obtained are contradictory to the “static contact first”¹⁵ model which states that there is a correlation between the proximity of genes inside the nucleus and the translocation frequency found in cancer. Our results on 11q23 translocation favour the “dynamic breakage first” model proposed by Aten *et al.*¹⁶ The further analysis of multi-probes labeled genes and registration of set of positions have been validated on phantom images, and experiments are currently being made in order to test the plugins on real data.

These results probe, if needed, the interest of this software for the generalized and automated study of nuclear structure.

REFERENCES

1. M. Manuelidis, “Individual interphase chromosome domains revealed by in situ hybridization,” *Hum Genet* **71**(4), pp. 288–93, 1985. 0340-6717 Journal Article.
2. R. M. Zirbel, U. R. Mathieu, A. Kurz, T. Cremer, and P. Lichter, “Evidence for a nuclear compartment of transcription and splicing located at chromosome domain boundaries,” *Chromosome Res* **1**(2), pp. 93–106, 1993. 0967-3849 Journal Article.
3. P. Lichter, T. Cremer, J. Borden, L. Manuelidis, and D. C. Ward, “Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant dna libraries,” *Hum Genet* **80**(3), pp. 224–34, 1988. 0340-6717 Journal Article.
4. E. Lukasova, S. Kozubek, M. Kozubek, J. Kjeronska, L. Ryznar, J. Horakova, E. Krahulcova, and G. Horneck, “Localisation and distance between abl and bcr genes in interphase nuclei of bone marrow cells of control donors and patients with chronic myeloid leukaemia,” *Hum Genet* **100**(5-6), pp. 525–35, 1997. 0340-6717 Journal Article.

5. S. Kozubek, E. Lukasova, A. Mareckova, M. Skalnikova, M. Kozubek, E. Bartova, V. Kroha, E. Krahulcova, and J. Slotova, "The topological organization of chromosomes 9 and 22 in cell nuclei has a determinative role in the induction of t(9,22) translocations and in the pathogenesis of t(9,22) leukemias," *Chromosoma* **108**(7), pp. 426–35, 1999. 0009-5915 Journal Article.
6. H. Neves, C. Ramos, M. G. da Silva, A. Parreira, and L. Parreira, "The nuclear topography of abl, bcr, pml, and raralpha genes: evidence for gene proximity in specific phases of the cell cycle and stages of hematopoietic differentiation," *Blood* **93**(4), pp. 1197–207, 1999. 0006-4971 Journal Article.
7. M. N. Nikiforova, J. R. Stringer, R. Blough, M. Medvedovic, J. A. Fagin, and Y. E. Nikiforov, "Proximity of chromosomal loci that participate in radiation-induced rearrangements in human cells," *Science* **290**(5489), pp. 138–41, 2000. 0036-8075 Journal Article.
8. J. J. Roix, P. G. McQueen, P. J. Munson, L. A. Parada, and T. Misteli, "Spatial proximity of translocation-prone gene loci in human lymphomas," *Nat Genet* **34**(3), pp. 287–91, 2003. 1061-4036 Journal Article.
9. I. Alcobia, A. S. Quina, H. Neves, N. Clode, and L. Parreira, "The spatial organization of centromeric heterochromatin during normal human lymphopoiesis: evidence for ontogenically determined spatial patterns," *Exp Cell Res* **290**(2), pp. 358–69, 2003. 0014-4827 Journal Article.
10. M. Gué, C. Messaoudi, J. S. Sun, and T. Boudier, "Smart 3D-FISH: automation of distance analysis in nuclei of interphase cells by image processing.," *Cytometry A* **67**, pp. 18–26, Sep 2005.
11. M. Gué, J.-S. Sun, and T. Boudier, "Simultaneous localization of MLL, AF4 and ENL genes in interphase nuclei by 3D-FISH: MLL translocation revisited.," *BMC Cancer* **6**, p. 20, Jan 2006.
12. W. Rasband, "Imagej,"
13. T. Ridler and S. Calvard, "Picture thresholding using an iterative selection method.," *System, Man and Cybernetics* , pp. 630–632, 1978.
14. K. S. Arun, T. S. Huang, and S. D. Blostein., "Least square fitting of two 3-d point sets," *Transactions on Pattern Analysis and Machine Intelligence* **9**(5), 1987.
15. L. Parada and T. Misteli, "Chromosome positioning in the interphase nucleus," *Trends Cell Biol* **12**(9), pp. 425–32, 2002. 0962-8924 Journal Article Review Review, Tutorial.
16. J. A. Aten, J. Stap, P. M. Krawczyk, C. H. van Oven, R. A. Hoebe, J. Essers, and R. Kanaar, "Dynamics of DNA double-strand breaks revealed by clustering of damaged chromosome domains.," *Science* **303**, pp. 92–95, Jan 2004.