

An automated tool to estimate chromatin compaction in stained nuclei

L. CUNEO^{(1)(2)(*)}, M. CASTELLO⁽²⁾⁽³⁾, F. BALDINI⁽²⁾ and A. DIASPRO⁽¹⁾⁽²⁾

⁽¹⁾ *DIFILAB, Department of Physics, University of Genoa - Genoa, Italy*

⁽²⁾ *Nanoscopy, Istituto Italiano di Tecnologia - Genoa, Italy*

⁽³⁾ *Genoa Instruments Srl - Genoa, Italy*

received 31 January 2022

Summary. — Chromatin organisation undergoes various structural changes during the entire cell life. The spatial arrangement of chromatin within the nucleus is not random, but how its architecture varies during cellular processes is still partially unknown. The aim of this work is to provide an automatic tool to quantify chromatin compaction from three-dimensional (3D) fluorescence microscopy images. We developed a tool to compute several morphological and statistical measures on stained nuclei. Furthermore, the nucleus is clustered, thus identifying different compactness regions. This workflow paves the way for biologists to study how chromatin structures form and behave in various cellular processes, by providing a parameter-free and automatic numerical quantification method of chromatin compaction.

1. – Introduction

Nuclear architecture and chromatin remodelling are the main regulators of cell identity and fate. In particular, chromatin spatial organisation is a well-orchestrated mechanism and its inappropriate regulation is the cause of many diseases, such as cancer [1]. Nevertheless, the molecular mechanisms involved in chromatin compaction are still partially unknown [2]. Fluorescence microscopy facilitates access to chromatin organisation at the nanoscale by optical means [3], allowing investigating how the 3D structure influences genome function. Indeed, 3D image stacks of stained nuclei, carrying information about chromatin, can be obtained by means of a confocal microscope. However, a method to quantify chromatin compaction from such images is required [4]. The aim of this work is to provide an original automatic tool to quantify chromatin compaction, in terms of numerical values related to fluorescence intensity, measurable from 3D images of stained nuclei. Since it is quite common to have multiple stained nuclei in the same Field of View (FoW), the first required task is nuclei segmentation [5]. This problem can be approached with several algorithms [6,7]. Among all possible techniques, training a neural network offers the highest degree of generalisation. Moreover, no expertise in data analysis is required, since it is a parameter-free method. For these reasons, we have applied the random forest algorithm [8] coupled with the watershed [9] method, in order to separate the nucleus of interest from both the background and other nuclei which

(*) E-mail: lisa.cuneo@iit.it

may be present in the FoW. Several morphological and statistical analyses are performed on each segmented nucleus. The underlying hypothesis is the following: the higher the number of fluorescent molecules, the higher the number of photons collected and the greater the compaction [10]. In general, this hypothesis is not always true, since high concentration of fluorescent molecules can lead to self-quenching, with a reduction of the recorded fluorescent intensity. Alternative and more sophisticated approaches based on FLIM (Fluorescence-Lifetime Imaging Microscopy), *e.g.*, FLIM-FRET [11], have been developed in the past by our group. However, these suffer from the need of very complex, high-cost and dedicated instrumentation, not available in any biology laboratory studying chromatin. A numerical value for the compaction can be thus calculated. Moreover, we cluster the nucleus, thus identifying different compactness regions. Finally, the tool is tested on confocal images of nuclei where the DNA is selectively marked with a fluorescent dye.

2. – Methods

The tool is implemented in Javascript, compatible with ImageJ and its 3D Suite [12].

2.1. Segmentation. – In this study, the segmentation purpose is twofold: we want to separate the nuclei from the background and from other nuclei eventually present in the same FoW. Since the first issue is related to difference in intensity while the second is related to shape, we decided to split the algorithm in two steps, as described below.

In order to classify the background and the nucleus, we apply a multi-threaded version of random forest, as set out by Fran Supek [13]. So far, 200 trees are generated, the depth of each tree is unlimited and the number of simultaneous threads used for computation is set equal to 8. The user can train the machine giving some examples, in other words, labelling some voxels as *background* or *nucleus*. A pre-trained model is also available. The features taken into account have been selected in such a way as to discriminate locally between plate-like, line-like, and blob-like image structures. The training features included were: Gaussian blur, derivatives, Hessian, Laplacian, structure (eigenvalues of second-moment matrix), edges (detected using the Canny algorithm [14]), difference of gaussians, mean and variance. By applying the trained classifier to any stack of images we obtain a mask, *i.e.*, a binary image. If holes are present in the output of the segmentation, they will be filled.

Once nuclei and background are distinguished, if there is more than one nucleus in the FoW, we will apply the watershed algorithm [13] on the segmented image in order to split them. The plugin works with two steps, namely: the former is aimed to select the seeds and can be obtained from local maxima, the latter is a thresholding step. Only seeds with values greater than a preset threshold will be used. A second threshold is used to cluster voxels with values greater than the threshold. If the centres of the two separated nuclei are too close, they will not be split. This minimal distance has been set up equal to 10 voxel (this value has to be chosen accordingly to the pixel size of the image). In this way, the Region Of Interest (ROI) is labelled by a mask.

2.2. Numerical analysis. – We calculate metrics both directly on the mask, obtaining morphological measurements, and on the raw voxel values belonging to the mask, obtaining statistical metrics on intensities. As concerns the morphological information, the following measurements on the selected nucleus are computed: volume V and surface S , expressed in μm^3 (calibrated unit) and in number of voxels (uncalibrated unit); the compactness, defined as $36\pi V^2/S^3$, which is dimensionless and maximized by a sphere with a value equal to 1; sphericity, that is the cube root of compactness. Moreover, a 3D

ellipsoid is fitted to the object. The values taken into account are: the three radii and their ratios, the calibrated volume of the ellipsoid and the ratio between the ellipsoid volume and the volume of the object. As concerns the statistical analysis, we computed the integrated density, *i.e.*, the sum of the intensities, the mean, the standard deviation, the minimum, the maximum and the median of the fluorescent intensity values. The chromatin compaction is calculated as density, *i.e.*, as the fraction between the integrated density and the calibrated volume. Finally, in order to have a comparable value of chromatin compaction between different nuclei, we normalise it with the sum of the intensities divided by the maximum intensity and the uncalibrated volume.

2.3. Cluster analysis. – In order to better study the organisation of chromatin in the nucleus, the ROI is clustered in a tunable number of sub-ROIs, thus identifying different compactness regions. We apply the k-means algorithm [15] with the Euclidean distance identifying four clusters: the background and three levels of compaction.

2.4. Sample preparation and imaging. – We tried our tool on SKNBE2 neuroblastoma cells. Cells were grown on RPMI 1640 medium (Sigma-Aldrich) with 10% fetal bovine serum (FBS) and maintained at 37 °C. After reaching the correct confluence, the cells were fixed with 4% paraformaldehyde in 0.1 M phosphate-buffered saline for 15 min. Then, the cells were incubated with Hoechst 33342 to detect the nuclear structure. Finally, the slides were sealed and 3D images were acquired with a Nikon A1R MP confocal microscope.

3. – Results

We tested the tool on images acquired as detailed in sect. 2.4. The results obtained through the segmentation and clustering elaboration are represented in fig. 1 for one particular nucleus (as an example), where a piece of the stack and the full 3D volume are plotted. While, as regards the numerical analysis, a subset of the metrics in output is presented in table I for four different nuclei. In particular, we reported (as an example) the volume, the surface, the compactness, the integrated intensity and the normalised chromatin compaction.

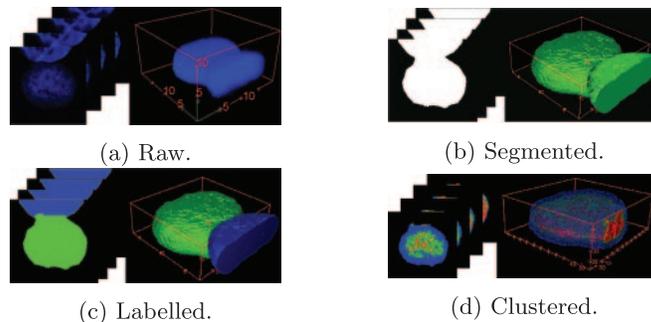


Fig. 1. – Example of 3D image stacks of stained nuclei, where DNA is selectively marked by fluorescent dye (Hoechst), analysed by the tool. Panel (a) shows the input raw data, panels (b) and (c) are, respectively, segmented and labelled images, while panel (d) is the clustered nucleus where colours represent low (blue), medium (green) and high (red) compaction.

TABLE I. – *Examples of morphological and statistical measurements.*

Vol. [μm^3]	Surface [μm^2]	Comp.	Int. Den. [A.U.]	Ch. Comp [%]
491.037	555.945	0.156	174.799	0.382
795.796	718.160	0.193	207.240	0.186
635.421	646.754	0.169	198.359	0.307
639.721	640.639	0.176	175.071	0.263

4. – Discussion

The major strength of this trainable tool is its generality. Indeed, it is possible to train the neural network to classify whatever type or part of a nucleus under study. Another important advantage is given by the absence of parameters to be set. This helps obtain objective quantification and measurements, not dependent on (heuristic) parameters decided by the user. Moreover, this makes the tool user-friendly, especially for people with no experience in data analysis and machine learning. Further improvements can be devoted to optimisation of segmentation and clustering analysis. For example, a better separation between nuclei and an automatic selection of the number of clusters can be implemented. Indeed, if two nuclei overlap too much, division could not be ideal, as in fig. 1(c). To conclude, the strategy in this work allows automatically performing different kinds of analyses, both morphologically and statistically, on stained nuclei. Thanks to the different output metrics of this method, a true comparison between two or more families of cells becomes easier than using other approaches. Moreover, by providing a numerical quantification of chromatin compaction, which impacts on gene expression, this workflow opens the way for biologists to study how chromatin structures form and behave in various cellular processes, from physiological to pathological ones, including cancer transformation, and finally can help understand the underlying genomic mechanisms.

REFERENCES

- [1] CREMER T. and CREMER C., *Nat. Rev. Genet.*, **2** (2001) 292.
- [2] DI STEFANO M. *et al.*, *Curr. Opin. Genet. Devel.*, **67** (2021) 25.
- [3] DIASPRO A. and BIANCHINI P., *Riv. Nuovo Cim.*, **43** (2020) 385.
- [4] DIASPRO A. *et al.*, *Image Vision Comput.*, **8** (1990) 130.
- [5] DIFATO F. *et al.*, *Microsc. Res. Tech.*, **64** (2004) 151.
- [6] SEO H. *et al.*, *Med. Phys.*, **47** (2020) 148.
- [7] SANNIHIT D. *et al.*, *Int. J. Eng. Appl. Phys.*, **1** (2021) 127.
- [8] BREIMAN L., *Mach. Learn.*, **45** (2001) 5.
- [9] SOILLE P. and VINCENT L. M., *Proc. SPIE.*, **1360** (1990) 240.
- [10] MASCETTI G. *et al.*, *Cytometry*, **23** (1996) 110.
- [11] PELICCI S. *et al.*, *J. Biophoton.*, **12** (2019) e201900164.
- [12] OLLION J. *et al.*, *Bioinformatics*, **29** (2013) 1840.
- [13] SUPEK F., <https://github.com/sdvillal/fast-random-forest>.
- [14] CANNY J., *IEEE Trans. Pattern Anal. Mach. Intell.*, **8** (1986) 679.
- [15] KANUNGO T. *et al.*, *IEEE Trans. Pattern Anal. Mach. Intell.*, **14** (2002) 881.