Consiglio Nazionale delle Ricerche
Istituto di Calcolo e Reti ad Alte Prestazioni

Identification and analysis of the intranuclear protein pattern in fluorescence microscopy images

L. Antonelli, F. Gregoretti, G. Oliva

RT-ICAR-NA-2021-02 **Novembre 2021**

Consiglio Nazionale delle Ricerche, Istituto di Calcolo e Reti ad Alte Prestazioni (ICAR) – Sezione di Napoli, Complesso Universitario Monte S. Angelo Via Cintia, 80126 Napoli, URL: *www.na.icar.cnr.it*

Consiglio Nazionale delle Ricerche
Istituto di Calcolo e Reti ad Alte Prestazioni

Identification and analysis of the intranuclear protein pattern in fluorescence microscopy images[1]

L. Antonelli, F. Gregoretti, G. Oliva[2]

RT-ICAR-NA-2021-02 **Novembre 2021**

---

[1] Questo rapporto include le slide della presentazione al Workshop "How can Scientific Computing help to study Life Sciences?" organizzato dall'Unità di Ricerca INdAM ICAR-CNR il 13 settembre 2021

[2] Istituto di Calcolo e Reti ad Alte Prestazioni, ICAR-CNR, Sezione di Napoli, Via P. Castellino 111, 80131 Napoli

Workshop "How can Scientific Computing help to study Life Sciences?"

Attività di disseminazione svolta presso l'Unità di Ricerca INdAM ICAR-CNR

http://www.na.icar.cnr.it/~maddalena.l/URINdAM.html

lunedì 13 settembre 2021

La giornata di lavori è stata mirata a promuovere e divulgare alcune delle attività svolte ed in corso presso l'Unità di Ricerca ICAR-CNR dell'INdAM (Istituto Nazionale di Alta Matematica), evidenziando il ruolo dei matematici e degli informatici nella risoluzione di problemi applicativi. Il focus di questa giornata è stato sulle applicazioni della biologia computazionale e della bioinformatica, che costituiscono sfide particolarmente avvincenti per diversi aspetti, data la mole di dati prodotti, la rapidità nella loro produzione, la complessità degli algoritmi atti alla loro elaborazione, le problematiche di sicurezza e protezione della privacy dei dati coinvolti, nonché per il loro ruolo nello studio della comprensione e la cura di importanti malattie.

Per sottolineare il contributo concreto apportato, le applicazioni di riferimento sono state scelte fra quelle che sono oggetto di collaborazioni scientifiche della UR con altri istituti ed enti di ricerca nel settore di riferimento. Esperti matematici, informatici, biologi, fisici e bioinformatici hanno illustrato gli elementi fondamentali della propria disciplina che entrano in gioco nelle ricerche oggetto della collaborazione.

Questo rapporto tecnico contiene le slide di una presentazione al Workshop tenuta da ricercatori dell'ICAR-CNR.

Identification and analysis of the intranuclear protein pattern in fluorescence microscopy images

Laura Antonelli

Institute for High Performance Computing and Networking (ICAR)

National Research Council of Italy

joint work with:

F. Gregoretti, G. Oliva (ICAR-CNR)

C. Lanzuolo's Laboratory (ITB-CNR and INGM)

workshop INDAM:

How can Scientific Computing help to study Life Science?

13th September 2021

Main Goals

Biological Goal

Analysis of the role of Polycomb Group Proteins (PcG) in the epigenetic signature of Laminopathies.

Scientific Computing Goal

Design of algorithms and software to automatically identify and analyze PcG proteins in fluorescence image sequences.

Acknowledgments

- *FIRB 2010 Project n.~RBFR106S1Z002*
- *EPIGEN Flagship Project*

Outline

1. **Fluorescence microscopy images**

   Introduction, features and issues

2. **Imaging Framework**

   Segmentation, 3D reconstruction and analysis

3. **Results**

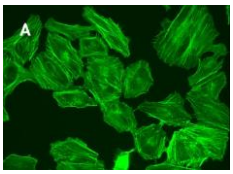
4. **Conclusions and Future Work**

Introduction

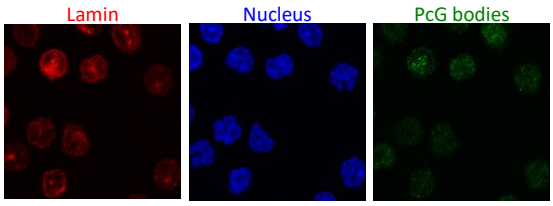
Principles of Microscopy Images

- Fluorescence microscopy has become an important **imaging technique** in cell biology. It is used in conjunction with **staining techniques** to visualize a whole range of intracellular structures.

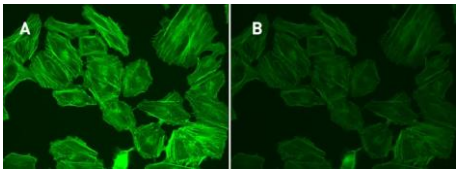
- The specimen is examined through a barrier filter that absorbs the short-wavelength light used for illumination and transmits the fluorescence, which is therefore seen as bright against a dark background.



Features and Issues



- When cells are excited by the illumination of a short wavelength, for example ultraviolet, the emergent rays are converted into longer wavelength light. Thus red, blue, or green light is emitted depending on the composition of the substance.
- The variety of fluorescent proteins and labeling techniques leads to considerable **differences in the appearance of cells**.



- Fluorophores lose their ability to fluoresce as they are illuminated in a process called **photobleaching**.
- Fluorescence microscopes produce images with very **low contrast**, since cells are sensitive to photodamage.

Why Use Automated Image Analysis?

High-Throughput image analysis

Manual analysis of a large volume of light microscopy images is **slow**, **time consuming** and **subject to observer variance**.

- number of image sequences: *tens of thousands*
- number of frames per sequence: *over 100*
- single frame sizes: *about* 10^3
- number of nuclei: *between 20 and 50*
- number of PcG bodies per nuclei: *up to around 100*

The large number of images generated in biological experiments that rely on advanced microscopy increases the demand of **automated image analysis tools**.

Sketch

framework

processing step

2D Segmentation
- image partition: nuclei regions and background
- PcG detection in nuclei regions

3D Reconstruction
- identification and reconstruction

analysis step

2D/3D Analysis
- nuclei features:
  area/volume, number, fluorescence, shape
- PcG features:
  area/volume, number, fluorescence distribution within nuclei, proximity to nuclear periphery

We have realized an **efficient** and **automatic imaging framework** in order to analyze the features of PcG in each image per sequence and in each cell per image

- The framework **integrates** algorithms written in C language for the **2D segmentation** and existing tools of the MatLab Image Toolbox for the **3D reconstruction**.
- The framework has **several functions** implemented in MatLab to analyze the PcG features which can be combined in order to create a customized analysis

Image segmentation

framework

processing step

2D Segmentation
- image partition: nuclei regions and background

The most challenging part of image analysis is usually determining which pixels in the image belong to each object (e.g., a protein, nucleus, or cell).
This task is known as **segmentation**.

The framework combines two segmentation methods:

[Chan, Esedoglu and Nikolova'06]

1. Region-based method based on the convex relaxation of the Chan-Vese model. This provides a two-region partition: nuclei regions and background using a combined image of lamin and nuclei images

$$min_{I,c_{in},c_{out}} F(I, c_{in}, c_{out}) = \underbrace{\int_\Omega |\nabla I| dx}_{\text{Regularization term}} + \lambda \underbrace{\int_\Omega \left((c_{in} - \bar{I})^2\right) I + \left((c_{out} - \bar{I})^2\right)(1 - I) dx}_{\text{Fidelity term}} \quad s.t.\ 0 \le I \le 1$$

c_{in} , c_{out} :mean values of the image $\bar{I}$ intensity of foreground and background

$\bar{I}$ *:Image to be segmented*
$\Omega \subset \mathbb{R}^2$ *:Image domain*

Numerical technique
- *first discretize then optimize*
  - Discretization step: all the quantities in the functional F are discretized
  - Optimization step: the alternating Split Bregman method

[Antonelli, De Simone CAIM 2016]

Image segmentation

framework

processing step

2D Segmentation
- image partition: nuclei regions and background

The most challenging part of image analysis is usually determining which pixels in the image belong to each object (e.g., a protein, nucleus, or cell).
This task is known as **segmentation**.

The framework combines two segmentation methods:

[Chan, Esedoglu and Nikolova'06]

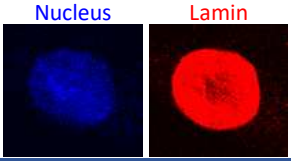
1. Region-based method based on the convex relaxation of the Chan-Vese model. This provides a two-region partition: nuclei regions and background using a combined image of lamin and nuclei images

$$min_{I,c_{in},c_{out}} F(I, c_{in}, c_{out}) = \underbrace{\int_\Omega |\nabla I| dx}_{\text{Regularization term}} + \lambda \underbrace{\int_\Omega \left((c_{in} - \bar{I})^2 \right) I + \left((c_{out} - \bar{I})^2 \right) (1 - I) dx}_{\text{Fidelity term}} \quad s.t.\ 0 \le I \le 1$$

c_{in} , c_{out} :mean values of the image $\bar{I}$ intensity of foreground and background

$\bar{I}$ *:Image to be segmented*
$\Omega \subset \mathbb{R}^2$ *:Image domain*

Nucleus Lamin



Segmentation

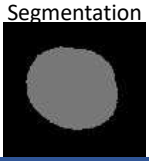


Image segmentation

framework

processing step

2D Segmentation
- image partition: nuclei regions and background
- PcG detection in nuclei regions:

- $thresh = p_{in}$
- **repeat**
  1. **Split image into** R_1 **and** R_2 **regions using** $thresh$
  2. **compute**
     μ_1 **= mean intensity of** R_1
     μ_2 **= mean intensity of** R_2
     $thresh = average(\mu_1, \mu_2)$
- **until** $(\mu_1 - \mu_2) < tol$

The most challenging part of image analysis is usually determining which pixels in the image belong to each object (e.g., a protein, nucleus, or cell).
This task is known as **segmentation**.

The framework combines two segmentation methods:

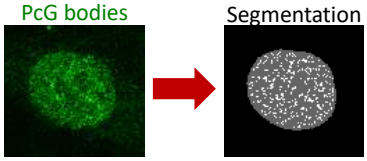
[Chan, Esedoglu and Nikolova SIAM J Appl Math 2006]

1. Region-based method based on the convex relaxation of Chan-Vese model. It provides a two-region partition: nuclei regions and background using a combined image of lamin and nuclei images

[Ball and Hall Tech Rep Stanf 1965]

2. Classification method based on the ISODATA applied only on the nuclei regions of the PcG image: the initial value of classifier threshold is p_{in} the mean intensity value of all nuclei regions in the PcG image.

PcG bodies Segmentation



[Gregoretti F, Cesarini E, Lanzuolo C, Oliva G, Antonelli L.. Methods Mol Biol. 2016]

Nuclei Identification and Reconstruction

framework

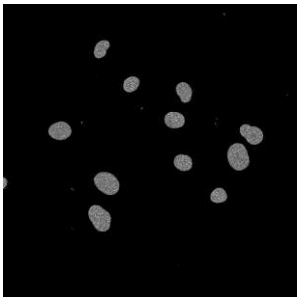
processing step

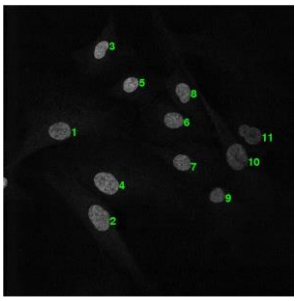
2D Segmentation
- image partition: nuclei regions and background
- PcG detection in nuclei regions

3D Reconstruction
- identification and reconstruction

Nuclei identification

Nuclei are numbered according to the number of connected components of the stack and are separated from each other.





Nuclei Identification and Reconstruction

framework

processing step

2D Segmentation
- image partition: nuclei regions and background
- PcG detection in nuclei regions

3D Reconstruction
- identification and reconstruction

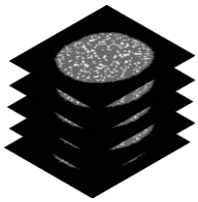
3D reconstruction of nuclei and PcG

The nuclei and the PcG bodies are reconstructed through a connected components algorithm using a 6-connectivity.

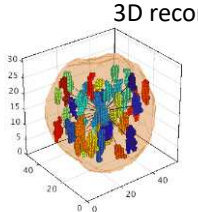
Random shuffling of detected PcG

PcG are scattered in the nucleus using a random distribution with the same mean and standard deviation as the actual PcG position distribution. Random location results are compared with the real location results in order to evidence a significance position.

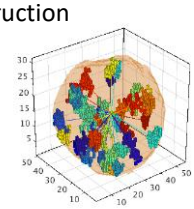
2D segmentation stack



3D reconstruction



Real PcG location



PcG random shuffling

Image analysis

framework

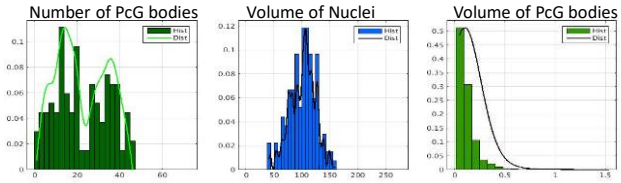
analysis step

2D/3D Analysis

- nuclei features:
  area/volume, number, fluorescence, shape
- PcG features:
  area/volume, number, fluorescence
  distribution within nuclei,
  proximity to nuclear periphery

PcG bodies: number and features

- Percentage/Number of Nuclei and PcG
- Volume of Nuclei and PcG



- The 'roundness' of each nucleus is evaluated by the eccentricity on the mean plane of the z-stack

$$\text{roundness} := 4\pi NCL_n.\text{area}/(NCL_n.\text{perimeter})^2 = \begin{cases} \approx 1 \rightarrow \text{nucleus is a circle} \\ \approx 0 \rightarrow \text{nucleus is not a circle} \end{cases}$$

NCL_n : n-th nucleus

Image analysis

framework

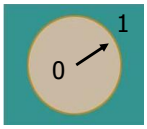
analysis step

2D/3D Analysis

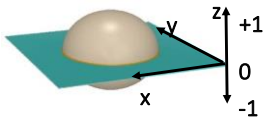
- nuclei features:
  area/volume, number, fluorescence, shape
- PcG features:
  area/volume, number, fluorescence
  distribution within nuclei,
  proximity to nuclear periphery

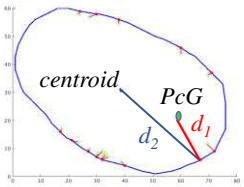
PcG distribution

- The distance from the nuclear centroid



- The distance from the mean z-axis





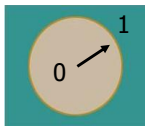
- The 'proximity' of the PcG from nuclear periphery is the ratio between:
  d_1 the minimum euclidean distance of PcG from the nuclear periphery
  d_2 the distance of the nuclear centroid from the point on nuclear periphery closest to the PcG

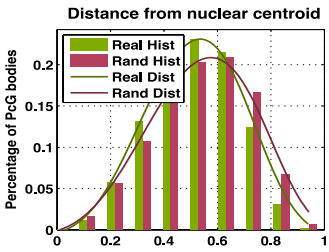
$$d_1/d_2 = \begin{cases} \approx 1 \rightarrow \text{PcG is near to the centroid} \\ \approx 0 \rightarrow \text{PcG is near to the periphery} \end{cases}$$

Image Analysis

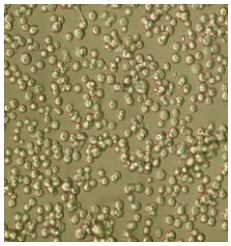
PcG bodies distribution

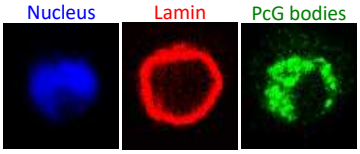
- The **distance from nuclear centroid** shows that the PcG bodies are excluded from periphery
- The **scattering around the mean z-axis** shows that PcG bodies are horizontally coplanar

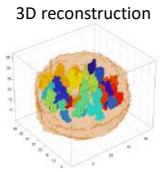


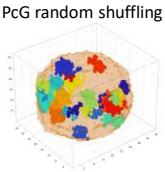


Drosophila S2









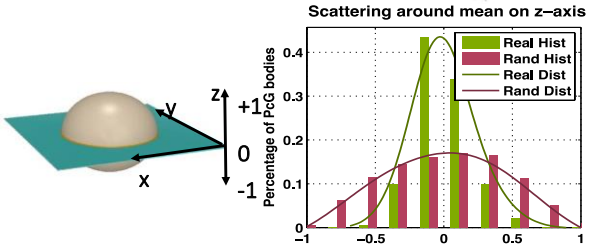
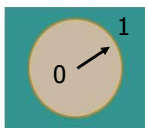
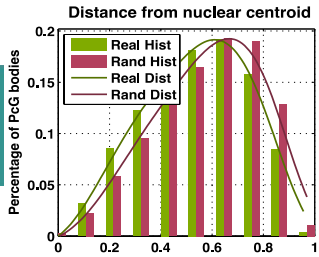


Image Analysis

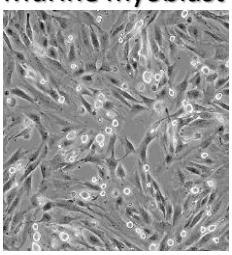
PcG bodies distribution

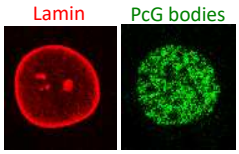
- The **distance from nuclear centroid** shows that the PcG bodies are excluded from periphery
- The **scattering around the mean z-axis** shows that PcG bodies are horizontally coplanar

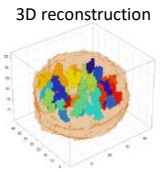


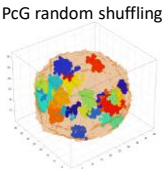


Murine myoblast









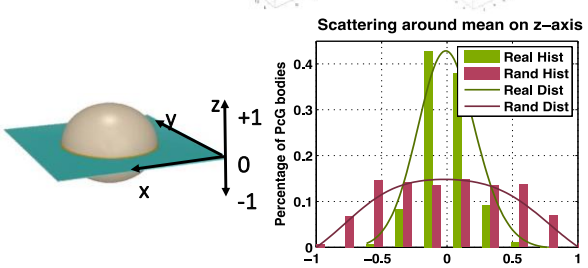
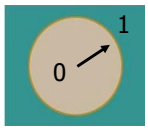
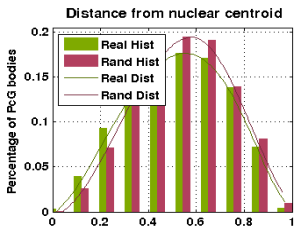


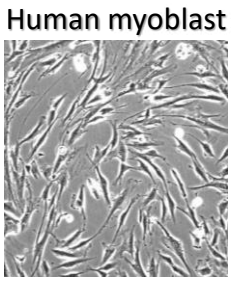
Image Analysis

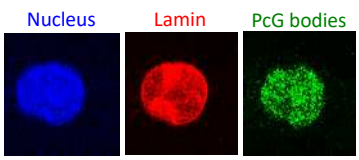
PcG bodies distribution

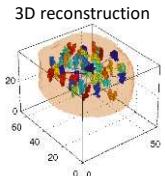
- The **distance from nuclear centroid** shows that the PcG bodies are excluded from periphery
- The **scattering around the mean z-axis** shows that PcG bodies are horizontally coplanar

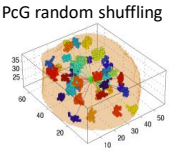


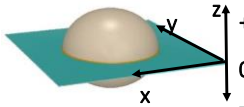


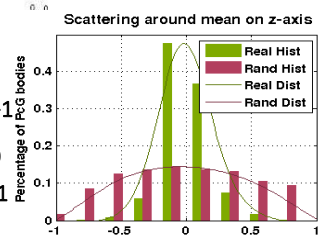












The positioning of PcG bodies is evolutionarily conserved, being horizontally coplanar and excluded from nuclear periphery.

References

- Sebestyén, Marullo, Lucini, Petrini, Bianchi, Valsoni, Olivieri, Antonelli, Gregoretti, Oliva, Ferrari, Lanzuolo *SAMMY-seq reveals early alteration of heterochromatin and deregulation of bivalent genes in Hutchinson-Gilford Progeria Syndrome*.
Nat Commun 11, (2020)
https://github.com/talisman4/2D-PcG-bodies-cell-image-analysis-HGP-Syndrome/tree/v1.0.0
github link of code

- Cesarini, Mozzetta, Marullo, Gregoretti, Gargiulo, Columbaro, Cortesi, Antonelli, Di Pelino, Squarzoni, Palacios, Zippo, Bodega, Oliva, Lanzuolo; Lamin A/C sustains PcG protein architecture, maintaining transcriptional repression at target genes.
***J Cell Biol* 9 (2015)**

- Fasciani, D'Annunzio, Poli, Antonelli, Gregoretti, Oliva et al.
MLL4-associated condensates counterbalance Polycomb-mediated nuclear mechanical stress in Kabuki syndrome
Nat Genet 52, (2020)

Future Work

- Hierarchical clustering of different cell nuclei populations

Thank you for your attention!

laura.antonelli@cnr.it