M2 internship – Coulon lab, Institut Curie, Paris – http://coulonlab.org/

Analysis of super-resolution microscopy data using signal theory and stochastic processes

Our ability to observe structures inside cells by fluorescent microscopy is fundamentally limited by the diffraction of light, imposing a ≈200 nm limit to the resolution that can be achieved by conventional imaging. In the recent years, *super-resolution* techniques have been developed – including single-molecule localization microscopy (SMLM) – now allowing to beat this limit and to observe objects at 20-40 nm resolution. However, many aspects of SMLM are inherently stochastic (e.g. partial labeling, fluorophore activation and blinking, localization errors…) so that each molecule in the sample does not equate to one-and-only-one detection in the reconstructed image, but rather to a distribution of detections – i.e. anything between no detection and a cluster of many detections. Image reconstruction approaches aim at minimizing these issues to obtain more accurate images. But this correction is only partial, so that quantitative interpretation of SMLM data remains a major challenge.

In the cell nucleus, these issues are particularly challenging. As opposed to many cytoplasmic objects (filaments, membrane…), most structure in the nucleus are less stereotyped: chromatin domains with local variations in density and possibly fractal properties, compartments/territories with partial overlap, clusters of factors with specific stoichiometries… Hence, the typical artifacts of SMLM (due to multiple detections of emitters) resemble the structures of interest. Methods for accurate and unambiguous analysis of nuclear structures are still needed.

Contrary to approaches aiming to minimize the 'multiple detection' artefacts during image reconstruction, others simply take them into account in the statistical analysis of the data without trying to correct them in the images:

i. Methods based on *spatial multi-point correlations* are very powerful in that they can, at the same time, characterize structures and make the effect of multiple detections very clear and identifiable [1].
ii. Another elegant approach to separate 'multiple detection' artefacts from real structural information relies on the fact that they have distinct *scaling properties* in across imaging conditions [2].
iii. Finally, *particle class averaging* methods, broadly used in cryo-electron microscopy, are very powerful for SMLM data in the case of stereotyped structures [3], as exemplified by the structure of the nuclear pore complex (NPC) [4] or benchmark studies on DNA origamis [5], reaching sub-nanometer precision.

The goal of this internship is to construct an analysis method, relying on spatio-temporal multi-point correlation functions/cumulant analysis and non-homogenous Poisson processes, to extract structural information in super-resolution data. Approaches based on multi-point correlation are common in material science to characterize the nano-structure of materials and will be here adapted to account for (i) the stochastic nature of SMLM and (ii) the heterogenous and non-stereotyped nature of molecular assemblies in the nucleus. Leveraging advantages of the 3 above-mentioned approaches, we expect this method to provide quantitative stoichiometric information, a multi-scale characterization of structures, with a possible increase in resolutive power.

Practically, the student will:

- Develop and implement this technique to obtain cumulant maps
- Simulate SMLM data and test the technique on simulated data
- Apply it to existing datasets with stereotypical structures (e.g. NPC [4], DNA origami [5]) as a proof of concept
- Apply it to multi-channel datasets of nuclear structures generated at the institute and with collaborators
- Propose and test solutions to the inverse problem of recovering full 3D structures (e.g. phase retrieval…)

Candidates should have **substantial programming skills** (Python preferred), knowledge in **signal processing theory** (auto/cross-correlation, higher order moments/cumulants) and **stochastic processes** (Poisson processes), a will to understand the physics behind SMLM methods, and a strong interest in quantitative analysis of biological data.
For application, please contact: Antoine Coulon at recruitment@coulonlab.org

Related literature:

[1] Sengupta, P., Jovanovic-Talisman, T., Skoko, D., Renz, M., Veatch, S. L., & Lippincott-Schwartz, J. (2011). Probing protein heterogeneity in the plasma membrane using PALM and pair correlation analysis. Nature Methods, 8(11), 969–975. http://doi.org/10.1038/nmeth.1704

[2] Baumgart, F., Arnold, A. M., Leskovar, K., Staszek, K., Fölser, M., Weghuber, J., et al. (2016). Varying label density allows artifact-free analysis of membrane-protein nanoclusters. Nature Methods. http://doi.org/10.1038/nmeth.3897

[3] Sieben, C., Douglass, K. M., Guichard, P., & Manley, S. (2018). Super-resolution microscopy to decipher multi-molecular assemblies. Current Opinion in Structural Biology, 49, 169–176. http://doi.org/10.1016/j.sbi.2018.03.017

[4] Szymborska, A., de Marco, A., Daigle, N., Cordes, V. C., Briggs, J. A. G., & Ellenberg, J. (2013). Nuclear pore scaffold structure analyzed by super-resolution microscopy and particle averaging. Science, 341(6146), 655–658. http://doi.org/10.1126/science.1240672

[5] Schnitzbauer, J., Wang, Y., Zhao, S., Bakalar, M., Nuwal, T., Chen, B., & Huang, B. (2018). Correlation analysis framework for localization-based superresolution microscopy. Proceedings of the National Academy of Sciences, 115(13), 3219–3224. http://doi.org/10.1073/pnas.1711314115