PhD project: Next generation tools for image based transcriptomics

Project Description

Gene expression is the process by which genetic information is transformed into functional products. RNA molecules are essential for this process; because they are either the functional gene products themselves or they serve as an intermediate molecule prior to the production of other functional products (proteins).

The study of gene expression has been one of the major fields in genome biology for many years and has triggered the development of both experimental techniques, such as microarrays or RNA sequencing, and computational methods in Bioinformatics. Traditionally, these studies focused on the number of RNA molecules, also called expression level.

More recently however, it has become apparent that it is not only the number of RNAs that matters, but also their localization inside the cell and inside the organism. Single molecule FISH (smFISH) is an imaging technique allowing the visualization of single RNA molecules (see Fig. 1). Recent advances in the experimental techniques now allow (1) to perform such experiments at a large scale, probing for hundreds of different RNAs individually in different cellular populations in a completely automated manner and (2) to jointly visualize many different RNAs in the same cell. These experimental breakthroughs are accompanied by a strong need for sophisticated and robust methods for the analysis of RNA localization patterns.

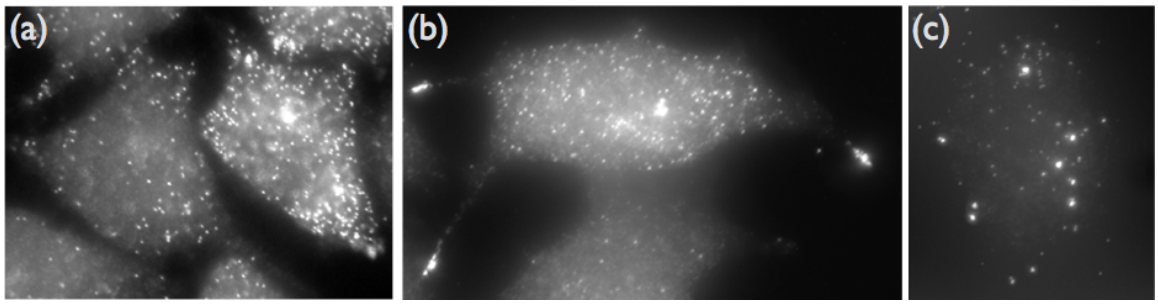

Figure 1 - Different localization patterns observed by smFISH.

Previously, we have developed tools for the automatic analysis of localization patterns in smFISH images. These methods rely on the detection of individual spots and the quantitative description of the resulting spatial distributions of points. With the new experimental developments and the new biological questions that can be addressed with these techniques, there is a need for new methodological developments, too.

The subject of this PhD thesis is the development of these next generation tools for image based transcriptomics. Concretely, there are 3 main objectives:

(1) To develop a statistical model to predict candidate RNAs that are likely to localize in specific regions in the cells. This is an essential step prior to performing screens, because it would allow experimental groups to keep the number of experiments in a reasonable range (typically ~1000), yet to allow unbiased choice of candidates and thus novelty detection. Second, such approaches will also allow to build a link between

sequence features of RNA and the localization pattern they produce. This will provide us with insights into the mechanism of the localization process.

(2) To build a tool for the visualization and analysis of multi-channel smFISH data, in which tens and hundreds of different RNAs are detected in the same cells. Today, it is still extremely challenging to visualize hundreds of channels together in one 3D volume. One idea is to visualize groups of RNAs, which are found by unsupervised learning techniques.

(3) To analyze challenging smFISH data sets generated during the project. These data sets will focus on particular biological processes, each coming with its own computational challenges. In particular, we will analyze RNA distributions for genes known to play a role in cell division. As the geometry of dividing cells is particular with respect to other cell cycle phases, there is a need for new methods for the identification of meaningful localization patterns. In addition, there are also recent efforts by other teams to measure and evaluate protein localization during cell division. It will be very interesting to use such complementary data sets in order to identify the correlations between RNA and protein localizations.

Research group and collaborators

Research group

The project will take place in the Centre for Computational Biology (CBIO — http://cbio.ensmp.fr), a joint laboratory between Mines ParisTech, one of the most prominent French engineering schools, and Institut Curie, a major hospital and research facility dedicated to cancer. CBIO benefits from an exceptional scientific environment with immediate access to experts and collaborators in biology and medicine, enabling a stimulating interdisciplinary exchange. The laboratory is located in the center of Paris, both in Mines ParisTech and in the nearby Institut Curie.

Collaborating groups

On this project, we are actively collaborating with the Zimmer group at the Institut Pasteur and the Bertrand group at the IGMM (Institut de Génétique Moléculaire de Montpellier). These two groups perform the experimental work, but are also involved in computational aspects.

Supervisor

The project will take place under the supervision of Thomas Walter.

Florian Müller (Zimmer group, Institut Pasteur) and Jean-Philippe Vert (director, CBIO) will co-supervise the project.

http://cbio.ensmp.fr/~twalter/index.html

Thomas.walter@mines-paristech.fr

Applications

The applicant should have a strong background in computer science and/or applied mathematics. Experience in machine learning, image analysis or computer vision, as well as good programming skills are required for this project. Basic knowledge in biology is an advantage, but not a requirement.

Applications are to be addressed to Thomas Walter (Thomas.Walter@mines-paristech.fr)