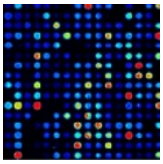
Séminaire – Math For Genomics

Séance du mercredi 4 avril 2018. 10h30.
Evry. IBGBI. LaMME.

Données Hi-C



Laboratoire de Mathématiques et Modélisation d'Évry
LaMME

université evry val-d'essonne

GENOPOLE

Nicolas SERVANT (Institut Curie, INSERM, Mines Paris Tech)

Title: Efficient processing of Hi-C data and application to cancer.

Over the past decade, major advances in high-throughput sequencing have allowed the development of new epigenetics approaches. Among them, the Hi-C technique was proposed as a genome-wide method to explore the chromatin organization in three-dimension (3D). Since then, the spatial organization of the genome and the physical interactions occurring within and between chromosomes has been described as a key factor of gene regulation and genome functions in general.

However, as any genome-wide sequencing data, Hi-C usually requires several millions to billions of paired-end sequencing reads, depending on genome size and on the desired resolution. Managing these data thus requires optimized bioinformatics workflows able to extract the contact frequencies in reasonable computational time and with reasonable resource and storage requirements. In this context, I will present HiC-Pro (https://github.com/nservant/HiC-Pro), an optimized and flexible pipeline for processing Hi-C data from raw reads to normalized contact maps. HiC-Pro maps reads, detects valid ligation products, performs quality controls and generates intra- and inter- chromosomal contact maps. It includes a fast implementation of the normalization and is based on a memory-efficient data format for Hi-C contact maps. Using HiC-Pro therefore allows to easily process high-throughput Hi-C data in a simple command line.

In addition, I will discuss the current computational challenges that emerge when Hi-C is applied on cancer cells. Given the important recent insights that chromosome conformation techniques have provided into 3D genome organization in a normal context, the application of such approach to a disease context offers the possibility to further explore the genome organization of cancer cells, and its impact on cell regulation. I will demonstrate why the Hi-C cancer data require dedicated normalization method, and how we can solve these issues.

Sarah OUADAH (AgroParisTech)

Title: Nonparametric multiple change-point estimation for analyzing large Hi-C data matrices

In this presentation, we propose a novel nonparametric approach for estimating the location of block boundaries (changepoints) of non-overlapping blocks in a random symmetric matrix which consists of random variables having their distribution changing from one block to the other. Our change-point location estimators are based on nonparametric homogeneity tests for matrices. We first provide some theoretical results for these tests. Then, we prove the consistency of our change-point location estimators. Some numerical experiments are also provided in order to support our claims. Finally, our approach is applied to Hi-C data which are used in molecular biology for better understanding the influence of the chromosomal conformation on the cells functioning.
This work is the fruit of a collaboration with Vincent Brault (Univ. Grenoble Alpes), Céline Lévy-Leduc (AgroParisTech) and Sarah Ouadah (AgroParisTech).