Comparative analysis of 3D genomic data

Nathalie Vialaneix*†[1]

[1]INRAE – INRAE, Sigenae, Genotoul Bioinfo, BioInfoMics, MIAT UR875, Sigenae, INRAE, Castanet-Tolosan, France – France

Abstract

The three-dimensional conformation of the genome plays a central role in many biological processes, such as gene expression regulation. The Hi-C sequencing technique allows this 3D organization to be profiled by measuring interaction frequencies between genomic regions. The data obtained by this technique are summarized in the form of symmetric matrices, where each pixel (i,j) is a positive count that measures the number of interactions between genomic positions i and j. These counts are used as an estimate of the spatial proximity between i and j. The analysis of this type of data aims to extract functional regions, such as particularly dense genomic regions or genomic regions whose interactions vary significantly between two biological conditions. In this presentation, we will present the characteristics of these data and the analysis strategies that are implemented to answer these questions. In particular, we will show how constrained Hierarchical Agglomerative Clustering can be leveraged for the extraction of functional regions and the modeling of the hierarchical structure of the data. We will also present a review and statistical benchmark of comparative analysis strategies. During this presentation, I will discuss work carried out in collaboration with Marie Chavent, Sylvain Foissac, Pierre Neuvial, and Matthias Zytnicki, and within the context of the PhD theses of Nathanaël Randriamihamison and Élise Jorge.

Keywords: Next generation hypothesis testing

*Speaker

†Corresponding author: nathalie.vialaneix@inrae.fr