Modeling Spatial Genomic Interactions with the Hawkes model

Anna Bonnet*[1], Vincent Rivoirard[2], and Franck Picard[3]

[1]Laboratoire de Biométrie et Biologie Evolutive (LBBE) – Université Claude Bernard Lyon 1, Institut National de Recherche en Informatique et en Automatique, Centre National de la Recherche Scientifique : UMR5558 – 43 Bld du 11 Novembre 1918 69622 VILLEURBANNE CEDEX, France

[2]CEntre de REcherches en MAthématiques de la DEcision (CEREMADE) – Université Paris-Dauphine, Centre National de la Recherche Scientifique : UMR7534 – Place du Maréchal de Lattre de Tassigny 75775 - Paris Cedex 16, France

[3]Laboratoire de Biométrie et Biologie Evolutive (LBBE) – CNRS : UMR5558, Université Claude Bernard - Lyon I (UCBL), INRIA – 43 Bld du 11 Novembre 1918 69622 VILLEURBANNE CEDEX, France

Abstract

New sequencing techniques allow to study genomes with an unprecedented resolution; molecular processes can be captured through to the mapping of the whole genetic variants of a genome, the measure of the expression of
all genes of a cell, and now the structure of chromatin and its modifications. Epigenetics has been certainly one of the most active fields for the last few years, thanks to the development of ChIP-Seq protocoles that allow the spatial localization of DNA-proteins interactions.
ChIP-Seq data provide a map of chromatin modifications along chromosomes, and we focus on the statistical modeling of this spatial specificity of the data. This information can be used to characterize spatial interactions between genomic features, such as attraction/repulsion patterns, which may suggest relevant biological constraints. The modeling of such patterns is complex and calls for a statistical method that models interactions between spatial stochastic processes. Here we focus on the interplay between the starting points of DNA replication and chromatin marks in the human genome. DNA replication is a biological process that is intrisically spatial and the identification of replication origins at fine scales has revealed a complex interplay between genetic and epigenetic features in the early steps of DNA replication.
The methodology we propose relies on modeling spatial interactions thanks to the Hawkes model. We use this model and the associated methods to detect spatial interactions and estimate their intensities. The benefits of this method compared to standard approaches are plural, in particular it allows the reconstruction of sparse interaction functions, which is very convenient for interpretation. In addition, the multivariate aspect of our model is central compared to existing
pairwise strategies, since it allows to correct for spurious aliasing correlations. The application to the replication origins data allows us to determine which chromatin marks are associated to the initation of replication process and we also higlight attractive/repulsive effects between different chromatin marks.

*Speaker