



02
DÉC.
2010

🕒 de 11h à 12h

SÉMINAIRE

Statistical and algorithmic developments for the analysis of SNP/CGH arrays, with application to breast cancer data.

Guillem Rigail

AgroParisTech, Paris

Basal-like breast cancers are among the breast cancers with the poorest prognoses and patients do not benefit from any targeted therapy yet. We aim to identify the deregulated signaling pathways using genomic, transcriptomic and proteomic (RPPA) data in order to identify therapeutic targets. In this talk, I will focus on the analysis of SNP and CGH data. More specifically, I will discuss several statistical and algorithmic challenges directly related to their statistical analysis.

- 1) Normalisation One important issue when analyzing SNP profiles is their normalisation. Indeed, especially with tumour profiles, it cannot be assumed that most of the genome is normal and it has been shown that not taking these genomic alterations into account while normalising leads to over-correction. We propose a method to estimate the signal (or copy number) and correct technical artefacts simultaneously.
- 2) Exact and Fast segmentation A CGH profile can be viewed as a succession of segments representing regions in the genome that share the same DNA copy number. Multiple-change-point detection methods constitute a natural framework for their analysis and the detection of breakpoints. However, recovering the optimal position of the breakpoints is not an easy task, especially for large SNP profiles such as Affymetrix SNP 6.0. We propose an algorithm to recover quickly the best segmentation (the maximum likelihood estimate).
- 3) Assessing the quality of a segmentation Assessing the quality of a segmentation and in particular the confidence we have in a particular breakpoint is a difficult problem. We propose algorithms and statistical methods to assess and take into account the quality of possible segmentations.