

Modèles statistiques pour l'analyse des données transcriptome d'origine tiling-array

Les sondes des puces 'tiling array' couvrent l'ensemble d'un génome (régions codantes et non-codantes) et leur utilisation permet d'étudier les phénomènes de transcription à un niveau très précis. Il est ainsi possible de s'affranchir de l'annotation pour étudier l'expression du génome (la transcription des micro-ARNs, les gènes anti-sens, l'épissage alternatif, ...).

La première partie de cette thèse sera consacrée à une modélisation du signal d'hybridation pour détecter les régions transcrites. Les méthodes déjà développées pour des puces à ADN ou CGH (dizaine de milliers sondes) constituent des premières pistes : segmentation, modèle de mélange, HMM. Cependant, du fait de la très haute densité des nouvelles puces, ces approches devront être étendues, notamment pour proposer des algorithmes d'estimation efficaces.

La seconde partie portera sur une modélisation conjointe des données de transcription et d'autres données utilisées pour l'annotation structurale, notamment de type séquence (intégration de données de natures hétérogènes). L'objectif de cette partie est d'améliorer le processus de reconnaissance des régions transcrites.

L'unité d'accueil sera l'UMR AgroParisTech/INRA MIA 518 :

www.agroparistech.fr/mmip/math/index.html

L'encadrement sera assuré par Stéphane Robin (robin@agroparistech.fr) et Marie-Laure Martin-Magniette (mlmartin@agroparistech.fr). Pour plus d'information, nous contacter.

Statistical methods to transcriptome tiling-array data

Probes of tiling array cover the whole genome (genic and intergenic regions) and their use allows us to explore transcription events. It is thus possible to investigate the RNA degradation, transcription of micro-RNA or anti-sense genes.

The first part of the thesis will be devoted to a model of the signal to detect the transcriptional regions. Methods already developed for microarray or CGH array (tens of thousands probes) are initial approaches: segmentation, mixture model, HMM. However these approaches should be extended because of the very high density of the tiling array. Especially algorithms should be proposed to estimate the model parameters in an effective way.

The second part of the thesis will be devoted to a model of transcriptome data and others data used for structural annotation as well as DNA sequence (heterogeneous data integration). The objective will be to improve the detection of transcriptional regions.

The thesis will be held at UMR AgroParisTech/INRA MIA 518:

www.agroparistech.fr/mmip/math/index.html

Directors will be Stéphane Robin (robin@agroparistech.fr) and Marie-Laure Martin-Magniette (mlmartin@agroparistech.fr). Feel free to contact us for further informations.

