

Analyse statistique de données génomiques (3 crédits) (Ex-ASDG 1) [van Helden J., Puthier D.]

Code: SBBCU23L

Objectif(s): Acquisition des connaissances théoriques et pratiques sur les différentes méthodes d'analyses statistiques mises en place dans l'étude des données génomiques.

Volume horaire: 18h de CM - 18h de TD - 0h de TP

Description

Avec l'avènement des méthodes haut-débit (séquençage massivement parallèle, microarrays,...) les statistiques sont devenues incontournables pour le biologiste s'intéressant à la génomique. Ce module propose de découvrir et de mettre en pratique un certain nombre de méthodes statistiques couramment utilisées dans ce domaine. Les exemples sont notamment tirés de l'analyse du transcriptome (RNA-Seq, microarrays) ou de l'épigénome (ChIP-Seq). La majorité des exemples d'analyse sont mis en pratique à l'aide du logiciel R et des bibliothèques du projet Bioconductor. Le programme général dont le contenu peut être adapté en fonction de l'actualité est le suivant:

- Méthodes d'analyse du transcriptome et de l'épigénome.
- Introduction à R
- Méthodes de normalisation (Quantile, lowess, TMM, RLE,...)
- Distributions (loi normale, binomiale, poisson, hypergéométrique, binomiale négative,...)
- Probabilités
- Tests d'hypothèse (test de Welsh's, test hypergéométrique,...)
- Correction des tests multiples (E-value, Bonferroni, FDR, Benjamini-Hochberg, ...)
- Read mapping probabilities (binomial)
- Méthodes de classification non-supervisées: distances et méthodes associées (classification hiérarchique, k-means,...)
- Méthodes de classification supervisées (LDA, SVM, KNN)

Prérequis

1ère année de master avec une formation initiale en génomique, probabilités et statistiques.