



Analyse statistique de données d'expression de gènes

Objectifs et contenu

Les technologies telles que le séquençage à haut débit (RNA-seq) et les puces à ADN permettent de générer des quantités importantes de données sur le transcriptome. Plusieurs problèmes statistiques se posent pour la planification d'expérience et pour l'analyse des données transcriptomiques : analyses exploratoires, normalisation des données, modélisation des variances pour la construction des statistiques de tests afin de détecter les gènes différentiellement exprimés, corrections pour les tests multiples, classification et modèles de mélange afin d'obtenir des groupes de gènes à fonction biologique similaire, inférence de réseaux de gènes.

Le module sera accompagné de travaux pratiques (sur R/Bioconductor) centrés sur une analyse de données publiques réelles.

Les notions abordées sont les suivantes : plans d'expérience, tests d'hypothèse, correction pour tests multiples, packages R/Bioconductor pour l'analyse de données transcriptomiques (DESeq2, limma), analyses d'enrichissement, Glasso.

Intervenant(s)

Andrea Rau (CR INRA)

Page personnelle : <http://www6.jouy.inra.fr/gabi/Presentation/Annuaire-des-pages-professionnelles/RAU-Andrea>

Page du laboratoire : <http://www6.jouy.inra.fr/gabi/>

Grégory Nuel (DR CNRS, LPMA, UPMC)

Page personnelle : <http://nuel.perso.math.cnrs.fr/>

Page du laboratoire : <http://www.proba.jussieu.fr/>

Prérequis

Des bases en statistique sont requises.

Durée et emploi du temps

Non précisés.

Infrastructure requise

Non précisée.

