

Analyse bioinformatique des données à haut débit (D. Puthier / A. Gonzalez)

Code: SBBCU24L

Objectif(s): L'objectif de ce module est de familiariser l'étudiant avec la manipulation des données haut-débit issues, notamment, d'expériences de séquençage (RNA-Seq, ChiP-Seq,...). L'accent est mis ici sur les aspect informatiques (les aspect statistiques étant couvert dans le module 'Analyse des Données Génomiques') et la mise en place de procédures permettant une analyse reproductible de données biologiques (*cf* 'reproducible research').

Volume horaire: 15h de CM - 15h de TD - 0h de TP

Description: Les outils présentés et utilisés pour mettre en place les procédures de traitement seront notamment tophat2, bowtie2, samtools, bedtools et igvtools (...). On présentera des solutions informatiques pour articuler les procédures de traitement (shell, make, snakemake/Python/BioPython, R/BioC,...) et soumettre celles-ci à un cluster de calcul. Une attention particulière sera portée sur l'écriture de procédures lisibles, reproductibles intégrant un rapport (snakemake, R/Latex/Sweave, knitr, notebook,...). Les procédures mises en place seront utilisées pour analyser des données téléchargées depuis la base de données SRA et ENA.

Prérequis: 1ère année de master avec une formation initiale en génomique (type Introduction à la génomique BG1.1).