

1SC4210 - Biologie et statistique

Responsables : **Laurent LE BRUSQUET** , **Arthur TENENHAUS**

Type de cours : **Cours ST**

Nombre d'heures d'études élèves (HEE) : **60**

Nombre d'heures présentielles d'enseignement (HPE) : **30**

Année académique : **2024-2025**

Présentation, objectifs généraux du cours :

La cellule est un système complexe multi-échelle que les biologistes étudient depuis plusieurs siècles. L'essor de nouvelles technologies permet désormais de l'explorer et de mesurer son activité de plus en plus finement. L'objectif de la première partie du cours est de présenter les différentes facettes de la cellule pour mieux en appréhender le fonctionnement global. Les technologies de mesure seront également abordées.

La deuxième partie du cours présente les méthodes statistiques largement utilisées dans le domaine biomédical. À l'issue de cette partie, les élèves disposeront d'une boîte à outils utile pour répondre aux différentes questions soulevées par l'analyse de données biomédicales.

Ce cours sera très largement illustré sur des applications médicales afin de permettre aux élèves d'acquérir les connaissances et le vocabulaire nécessaire à la compréhension des applications proposées en EI (bases en biologie, en génomique, etc.), de comprendre l'origine de l'énorme volume de données liées aux applications médicales. C'est la raison pour laquelle, environ un tiers du volume global de ce cours sera réservée à nos partenaires médecins/biologistes.

Période(s) du cours (n° de séquence ou hors séquence) :

ST4

Prérequis :

Statistique et Apprentissage

Plan détaillé du cours (contenu) :

PARTIE I : INTRODUCTION A LA BIOLOGIE CELLULAIRE et MOLECULAIRE (12h)

La cellule est un système complexe multi-échelle que les biologistes étudient depuis plusieurs siècles. L'essor de nouvelles technologies permet désormais de l'explorer et de mesurer son activité de plus en plus finement. L'objectif de ce module est donc de présenter les différentes facettes de la cellule pour mieux en appréhender le fonctionnement global. Les technologies de mesure seront également abordées.

I.1. BIOLOGIE CELLULAIRE & GENETIQUE MOLECULAIRE

- *Organisation générale de la cellule eucaryote

- *Les acides nucléiques (ADN et ARN)

- *Les protéines

- *Le cycle cellulaire de la cellule eucaryote

- *La réplication de l'ADN chez les eucaryotes :

- *La régulation de l'expression des gènes chez les eucaryotes

- *Outils génie génétique

I.2. GENOMIQUE

I.2.1. Génomique structurale

- *Définition de la génomique – essor de la bioinformatique
- *Etat des lieux du séquençage des génomes
- *Structure et Organisation des génomes
- *Méthodes de séquençage des génomes

I.2.2. Génomique Fonctionnelle : l'expression génique

- * Le transcriptome:
- * Régulation transcriptionnelle & épigénétique:

I.2.3. Intégration des données

- * Omics : intérêts en biologie et difficultés en bioinformatique
- * Intégration post-analyse : intégration supervisée ou non-supervisée
- * Intégration pré-analyse : vers l'inférence de réseaux
- * Applications en recherche fondamentale et en médecine personnalisée

I.2.4. Conclusion & perspectives : évolution de la notion de gène & modifications des modèles classiques de la régulation de la transcription. Informations apportées par la single-cell genomic.

A chaque partie du cours seront associés des travaux pratiques sous R

PARTIE II : BIOSTATISTIQUE (12h)

À l'issue de la partie II, les élèves disposeront d'une boîte à outils utile pour répondre aux différentes questions soulevées par l'analyse de données biologiques.

- * Approche univariée et tests multiples (Bonferroni, False Discovery Rate...);
- * Modèle linéaire, ANOVA et modèles à effets mixtes.
- * GLM (régression, régression logistique, ...)
- * Analyse en Composantes Principales et Clustering Hiérarchique

PARTIE III : TRAVAUX DIRIGES/TP (9h)

Une grande partie des notions abordées dans la partie biostatistique sera mise en oeuvre devant machine au travers du logiciel R.

Déroulement, organisation du cours :

5*3h de cours pour la partie Biologie

6*3h de cours pour la partie Statistique. Cette partie sera ponctuée de mise en oeuvre des méthodes présentées sur données réelles.

Organisation de l'évaluation :

examen écrit durée 1H30

Moyens :

- Equipe enseignante : Arthur Tenenhaus et Marie-Anne Debily (MCF Evry)

Acquis d'apprentissage visés dans le cours :

Deux principaux objectifs :

1. Introduction à la biologie moléculaire et cellulaire
2. Biostatistique : apporter des compléments en statistique (utiles pour l'analyse de données biomédicales) par rapport au cours "Statistique et Apprentissage".

Bibliographie :

Tout le matériel de cours sera disponible sur Edunao (slides, données, scripts)

