

Analyses de données multi-omiques (Séance 1/7)

Gaëlle LELANDAIS et Fabrice CONFALONIERI

(gaelle.lelandais@universite-paris-saclay.fr – fabrice.confalonieri@universite-paris-saclay.fr)

Ce contenu est mis à disposition selon les termes de la licence [Creative Commons BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/)

Objectif(s) de la séance

Au cours de cet enseignement, nous allons discuter les problématiques d'analyses de données multi-omiques. Pour cela, nous avons décidé de vous placer dans une situation pratique, similaire à celle d'un(e) chercheuse / chercheur dans un laboratoire de génomique fonctionnelle. **Ainsi, votre objectif est de caractériser le régulon DdrO, chez la bactérie radio-résistante *Deinococcus radiodurans*.** Pour cela, nous répéterons les analyses de données multi-omiques présentées dans un article scientifique (Eugénie *et al.*, 2021). Cette séance a pour objectif de vous présenter le contexte scientifique du projet, et de **débuter l'analyse des données RNAseq** (utilisation de Galaxy, <https://usegalaxy.fr/>).

Organisation de la séance

Cette séance se décompose en plusieurs activités. Elles sont détaillées ci-dessous. Pour une bonne cohérence générale des apprentissages, chacune des tâches est à réaliser, pendant la séance, ou à l'issue de la séance.

❖ **Activité 1 : Organisation des enseignements**

Résumé

Au cours de cette activité, le planning et le programme des enseignements prévus sont présentés. L'accès aux ressources pédagogiques est vérifié, ainsi que la liste des participant(e)s. Les modalités d'évaluation des connaissances et des compétences sont rappelées.

Travail à réaliser

- | | |
|--------------------------|--|
| ☐ | Vérifier votre inscription au cours sur eCampus. |
| ☐ | Accéder aux documents de la séance. |

❖ **Activité 2 : Séminaire de recherche – Fabrice Confalonieri**

Résumé

L'article scientifique dont les analyses de données multi-omiques vont être répétées, a été publié en septembre 2021. Le chercheur qui a coordonné les analyses est Professeur à l'Université Paris Saclay (et co-responsable de cette UE ☺). Cette première activité a pour objectif de vous expliquer le contexte scientifique de l'étude, ainsi que les principaux résultats qui ont été obtenus (et sont présentés dans l'article). Les stratégies expérimentales employées sont citées (et détaillées dans l'activité suivante).

Travail à réaliser

- ☐ Relire les diapositives du cours. Les notions abordées sont-elles claires ? **Solliciter l'équipe pédagogique si des clarifications sont nécessaires.**

Ressources complémentaires

- Page Web du journal : <https://www.mdpi.com/2073-4409/10/10/2536>.
- Page Web de l'équipe de recherche : <https://www.i2bc.paris-saclay.fr/equipe-bacteria-and-archaea-radioresistance/>.
- Page Web de l'ANR (soutien financier) du projet : <https://anr.fr/Projet-ANR-19-CE12-0010>

❖ **Activité 3 : Débuter avec Galaxy**

Résumé

Cette activité a pour objectif de vous faire découvrir la plateforme Web Galaxy. Très complète en termes de programmes bioinformatiques disponibles, cette plateforme accessible *via* internet est également très simple d'utilisation. C'est la raison principale de sa popularité au sein de la communauté scientifique. Les ressources pédagogiques utilisées sont issues d'une communauté de formateurs : <https://training.galaxyproject.org/training-material/>.

Travail à réaliser

- ☐ Relire les diapositives du cours <https://training.galaxyproject.org/training-material/topics/introduction/slides/introduction.html#1>. Les notions abordées sont-elles claires ? **Solliciter l'équipe pédagogique si des clarifications sont nécessaires.**
- ☐ Se connecter à l'instance Galaxy France <https://usegalaxy.fr/>. Créer un compte. **Il est conseillé d'utiliser une adresse professionnelle. Un e-mail de confirmation pour activer votre compte doit être reçu.**
- ☐ Réaliser le tutoriel suivant : <https://training.galaxyproject.org/training-material/topics/introduction/tutorials/galaxy-intro-short/tutorial.html>. **Solliciter l'équipe pédagogique si des clarifications sont nécessaires.**

Ressources complémentaires

- Vidéo de présentation de Galaxy (en anglais) : <https://training.galaxyproject.org/training-material/videos/watch.html?v=introduction/tutorials/galaxy-intro-short/slides>.
- Comment importer des données de séquençage dans Galaxy : <https://training.galaxyproject.org/training-material/topics/introduction/tutorials/galaxy-intro-ngs-data-managment/tutorial.html#getting-ngs-data-to-galaxy>

❖ **Activité 4 : Pipeline d'analyse de données RNAseq**

Résumé

Au cours de cette activité, les grandes étapes d'une analyse de données RNAseq sont présentées : 1) contrôle de la qualité des séquences ; 2) alignement des séquences sur le génome de référence ; 3) visualisation des alignements et 4) recherche de gènes différentiellement exprimés.

Travail à réaliser

- ☐ Relire les diapositives du cours. Les notions abordées sont-elles claires ? **Solliciter l'équipe pédagogique si des clarifications sont nécessaires.**

❖ Activité 5 : TP Galaxy (partie 1)

Résumé

Une analyse d'un premier fichier FASTQ est réalisée avec Galaxy. Les étapes consistent à 1) Créer un historique de travail ; 2) Importer le fichier FASTQ depuis la base de données SRA ; 3) Réaliser le contrôle de la qualité des séquences ; 4) Télécharger le génome de référence et 5) Réaliser un alignement des séquences sur le génome de référence.

Travail à réaliser

☐

Suivre pas à pas les indications du fichier PDF « TP_Galaxy_Partie1.pdf ».

❖ Autonomie : TP Galaxy (partie 2)

Résumé

L'analyse de données RNAseq est étendue à 6 fichiers FASTQ. Pour cela, les identifiants SRA sont reportés dans un fichier texte qui est importé dans un nouvel historique Galaxy. Les fichiers FASTQ sont importés depuis la base de données SRA. Les outils FASTQC et BOWTIE2 sont mis en application sur la collection et les résultats des alignements sont visualisés avec IGV.

Travail à réaliser

☐

Suivre pas à pas les indications du fichier PDF « TP_Galaxy_Partie2.pdf ». Le temps nécessaire à l'importation des fichiers FASTQ depuis la base de données SRA peut être long. **Un historique de secours est disponible en cas de besoin.**

☐

Soumettre une capture d'écran de votre visualisation IGV sur eCampus **(note incluse dans le CC)**

/!\!\ La prochaine séance débutera par une évaluation « Quiz » (note incluse dans le CC) /!\!