

Analyse de données multi-omiques (Séance 2/7)

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

Pour rappel, 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 **continuer l'analyse des données RNAseq** avec la problématique de **l'identification des gènes différentiellement exprimés**. Ces gènes ont des niveaux d'expressions mesurés par RNAseq « significativement » différents entre deux conditions comparées. La notion de « significativité » est importante et repose sur **l'utilisation de tests statistiques**. Dans ce contexte, la méthode DESeq2 sera mise en application.

Organisation de la séance

Cette séance se décompose en plusieurs activités. Elles sont détaillées ci-dessous. Les temps de réalisation des activités sont indicatifs et peuvent être ajustés en fonction de vos besoins. Pour une bonne cohérence générale des apprentissages, chacune des tâches est à réaliser, pendant la séance, ou à l'issue de la séance. **À tout moment ne pas hésiter à solliciter l'équipe pédagogique pour d'éventuelles clarifications.**

❖ Activité 1 : TP Galaxy (partie 3)

Résumé

Pour chacun des fichiers BAM obtenus lors du précédent TP Galaxy (Partie 2), les niveaux d'expression des gènes sont quantifiés en utilisant le logiciel HTSeq-Count. Pour cela, il est nécessaire d'importer dans l'historique Galaxy un fichier GFF dans lequel les positions des régions codantes (CDS) sont indiquées. Les tables de « counts » obtenues sont contrôlées, elles servent de point de départ pour la recherche de gènes différentiellement exprimés avec DESeq2.

Travail à réaliser

☐ Suivre pas à pas les indications du fichier PDF « TP_Galaxy_Partie3.pdf ». Vérifier que les fichiers de comptage obtenus ne contiennent pas que des 0.
--

❖ **Activité 2 : Analyse différentielle, introduction**

Résumé

Ce cours a pour objectif de présenter les grands principes, piliers des méthodes d'analyse différentielle. Trois notions importantes seront présentées : la taille d'effet, la dispersion et la force de l'expression.

Travail à réaliser

☐ Relire les diapositives du cours. Les notions abordées sont-elles claires ? **Solliciter l'équipe pédagogique si des clarifications sont nécessaires.** Vous pouvez poser vos questions sur le forum de l'espace eCampus du cours.

Ressources complémentaires

- Vidéo de présentation du cours (Replay) : <https://youtu.be/eztUy86l4ds>

❖ **Activité 3 : TP Galaxy (Partie 4)**

Résumé

La méthode DESeq2 est mise en application, en utilisant Galaxy.

Travail à réaliser

☐ Suivre pas à pas les indications du fichier PDF « TP_Galaxy_Partie4.pdf ».

❖ **Activité 4 : La méthodologie DESeq2, les grands principes**

Résumé

Ce cours a pour objectif de présenter les grandes étapes de la méthodologie DESeq. Elle fait partie des procédures statistiques les plus classiquement utilisées, créée spécialement pour la problématique de l'analyse différentielle.

Travail à réaliser

☐ Relire les diapositives du cours. Les notions abordées sont-elles claires ? **Solliciter l'équipe pédagogique si des clarifications sont nécessaires.** Vous pouvez poser vos questions sur le forum de l'espace eCampus du cours.

☐ Un questionnaire d'évaluation des cours sur l'analyse différentielle sera réalisé au début de la prochaine séance (note incluse dans le CC).

Ressources complémentaires

- Vidéo de présentation du cours (Replay) : <https://youtu.be/XzbzaDNqXhQ>

❖ **Autonomie : Création d'un workflow Galaxy**

Résumé

Lors des TP précédents, nous avons réalisé les analyses de données RNAseq pas à pas. Les outils d'analyses (FASTQC, Bowtie2, htseq-count et DESeq2) ont été mis en application les uns à la suite des autres. Cette activité a pour objectif d'automatiser cette succession des analyses par la création d'un « workflow ». Les workflows sont utiles pour 1) assurer la reproductibilité des résultats et 2) permettre le partage d'une méthodologie d'analyse avec d'autres.

Ressources pédagogiques

- Tutoriel pour la création, l'édition et l'importation de Workflow dans Galaxy : <https://training.galaxyproject.org/training-material/topics/galaxy-interface/tutorials/workflow-editor/tutorial.html>

Travail à réaliser

- | | |
|--------------------------|---|
| ☐ | Suivre les indications du fichier PDF « TP_Galaxy_Partie5.pdf ». A cette étape de l'enseignement, les indications sont de moins en moins guidées, il est conseillé de prendre le temps de lire et de se questionner. Solliciter l'équipe pédagogique si des clarifications sont nécessaires. |
| ☐ | Soumettre une capture d'écran de votre Workflow sur eCampus (note incluse dans le CC). |

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