Analyse de données multi-omiques (Séance 3/7)

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 répétons les analyses de données multi-omiques présentées dans un article scientifique (Eugénie *et al.*, 2021). Cette séance a pour objectif de **terminer l'analyse des données RNAseq**. Vous savez en effet automatiser les analyses de données avec Galaxy, par la création d'un « workflow » et calculer les paramètres statistiques des analyses DESeq2 dans le but **d'identifier des gènes différentiellement exprimés**.

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 : Discussion des critères de sélection des gènes « candidats », cibles de DdrO**

Résumé

Les analyses de données RNAseq vous ont permis d'obtenir, pour chaque gène de la bactérie, un ensemble de valeurs numériques. Celles-ci doivent vous permettre de proposer une première liste de gènes, composée de ceux qui seraient de bons candidats pour être régulés par DdrO.

Travail à réaliser

☐	Réfléchir aux critères de sélection de votre première liste de gènes cibles de DdrO. La (re)lecture de l'article scientifique peut être utile à cette étape.
☐	Mettre en application ces critères sur les résultats de la comparaison des points de temps 4h00 et 1h00 de la souche D37 (résultat du TP4).

❖ **Activité 2 : Résultats de l'analyse différentielle, points de vigilance en statistiques**

Résumé

Ce cours a pour objectif d'expliquer pourquoi des risques d'erreurs sont associés à la mise en application des tests statistiques et comment interpréter avec prudence les résultats obtenus.

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 eCampus.

Ressource(s) complémentaire(s)

- Vidéo de présentation du cours (Replay) : <https://youtu.be/CiP7tO0jOsM>

❖ **Activité 3 : Visualisation des paysages transcriptomiques de gènes cibles de DdrO**

Résumé

Le logiciel IGV est utilisé pour comparer les fichiers BAM obtenus dans les conditions 4h00 et 1h00 de la souche D37. Cette fois, des régions génomiques dans lesquelles des gènes connus pour être des cibles de DdrO sont regardées.

Travail à réaliser

- Charger dans le logiciel IGV, les fichiers BAM obtenus à partir de l'analyse des échantillons 4h00 et 1h00 de la souche D37. Si besoin, un historique de secours est disponible ici : <https://usegalaxy.fr/u/gaellelelandais/h/backup---fichiers-bam---d37-1h00-et-4h00>
- Ajouter le fichier d'annotation du génome de *Deinococcus radiodurans*. Celui-ci permet de rechercher des gènes connus pour être des cibles.

❖ **Autonomie : Compilation des résultats des analyses DESeq2**

Résumé

Lors des TP précédents, vous avez créé un workflow Galaxy vous permettant d'automatiser la réalisation d'une recherche de gènes différentiellement exprimés avec la méthode statistique DESeq2. Son utilisation doit vous permettre de réaliser les comparaisons : 4H vs 1H, 6H vs 1H, 8h vs 1H, 16h vs 1h et 24h vs 1H pour les souches D37 et W37. Cela représente **10 analyses avec DESeq2**.

Travail à réaliser

- Rassembler dans un unique fichier les valeurs de LogFC et de p-values (ajustées), obtenues pour l'ensemble des gènes de la bactérie. Ce fichier est rendu sur eCampus (**note incluse dans le CC**).

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