

Analyse de données multi-omiques (Séance 6/7)

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

Après avoir réalisé les analyses des données CHIP-seq pour localiser sur le génome ses sites de fixations, le but de cette séance est de rechercher dans les séquences des pics un ou plusieurs motifs nucléotidiques sur le(s)quel(s) ce régulateur pourrait se fixer. Pour mener à bien cette étude, on utilisera 2-3 autres logiciels disponibles sur <https://meme-suite.org/meme/>

Le premier logiciel intitulé « MEME » permet de rechercher des motifs plus ou moins conservés à partir d'un fichier fasta de séquences. Ces séquences sont celles que vous avez déterminé la semaine dernière issus des pics révélés par MCAS2. On utilisera également un autre fichier correspondant aux séquences des 136 pics obtenus avec le logiciel bPeak développé par Gaëlle et qu'on a utilisé pour le papier publié.

Pour le set de séquences donné, on analysera les séquences nucléotidiques des pics en faisant varier deux paramètres :

- Le fait que le motif soit présent ou non ou en plusieurs exemplaires
- Qu'il soit ou non palindromique

Ainsi pour un set de séquences 4 fichiers devront être produits afin de comparer les séquences des sites analysés.

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 : Organisation des enseignements**

Résumé

Les recherches réalisées par MEME sont longues surtout lorsque le nombre de séquences à analyser est important. De plus ce site web est très souvent sollicité et vous êtes nombreux à faire les mêmes requêtes. Enfin afin de saturer le site vous ne pouvez faire que 4 requêtes max par heure sauf si une ou plusieurs d'entre elles sont terminées avant.

Il est donc préférable de faire tourner MEME avec **un set de séquences plus restreint** afin de récupérer un premier logo représentant les séquences nucléotidiques conservées sur lesquelles DdrO pourrait se fixer puis de rechercher via un second logiciel appelé « FIMO » si cette séquence de fixation se retrouve dans la totalité des autres séquences présentes dans les fichiers issus de MACS et de bPeaks.

Vous allez donc créer en amont un fichier fasta avec une vingtaine de séquences de pics pour lesquelles vous êtes certains que DdrO se fixe bien sur ces séquences.

Durée

20 minutes.

❖ **Activité 2 : Analyse des données de séquences via MEME**

Résumé

Muni de votre nouveau fichier Fasta lancez MEME en suivant les indications données dans le fichier TP6. 4 requêtes pour 4 paramètres en tout. Attention de ne pas faire d'erreur dans votre requête sinon vous perdrez du temps ! N'oubliez pas d'indiquer pour chaque requête les paramètres utilisés.

Durée

20 minutes.

❖ **Activité 3 : Rechercher le motif dans les deux sets complets de séquences**

Résumé

Muni de vos logos il est très simple de rechercher directement si on retrouve le motif conservé 4 requêtes pour 4 paramètres en tout. Suivez les indications pour utiliser FIMO dans le fichier TP6.

Durée :20 minutes.

❖ Activité 4 : Recherche des logos dans l'ensemble des séquences promotrices

Résumé

Cette activité a pour objectif de s'interroger sur la pertinence de ce motif. Peut-on le retrouver ailleurs que dans les pics décelés par MACS2 ou bPeaks ? Si c'est le cas que peut-on en déduire ? Vous utiliserez FIMO et comme set de séquences le fichier seq_Prom_DRO.fasta trouvé sur ecampus

Durée

20 minutes.

❖ Activité 5 : comparaison des logos trouvés pour DdrO avec ceux déjà répertoriés dans les bases de données

Résumé

DdrO est un régulateur transcriptionnel appartenant à la famille HTH. De nombreux autres régulateurs bactériens sont également classés dans cette famille et leurs sites de fixation ont été répertoriés. Le but de cette activité est de comparer chacun des logos trouvés pour DdrO et de comparer les séquences de fixation de cette protéine avec celles d'autres régulateurs bactériens. Pour mener à bien cette étude vous utiliserez le logiciel TOMTOM. Suivez les indications dans le TP6.

Durée

20 minutes.