

MADT - Méthodes d'analyse de l'état de la chromatine

Anamaria Necsulea

anamaria.necsulea@univ-lyon1.fr

Février 2026

1 Introduction

Le consortium FAANG (<https://data.faang.org/home>) a pour objectif de construire une annotation fonctionnelle des génomes d'animaux. Il se concentre sur les espèces d'intérêt agronomique, dont le poulet. Ce consortium a généré des données de ChIP-seq pour plusieurs modifications d'histone, pour plusieurs organes chez le poulet adulte.

Dans le cadre de ce TP, nous allons analyser les données décrites dans le tableau suivant :

Numéro d'accension	Organe	Individu	Protéine ciblée
ERR10125084	reins	A	H3K27ac
ERR10125085	reins	B	H3K27ac
ERR10125069	reins	A	H3K27me3
ERR10125070	reins	B	H3K27me3
ERR10125039	reins	A	H3K4me3
ERR10125040	reins	B	H3K4me3
ERR10125054	reins	A	input DNA
ERR10125055	reins	B	input DNA

Les données de séquençage sont disponibles à l'adresse suivante : http://pbil.univ-lyon1.fr/members/necsulea/MADT_2026. Dans la même archive, vous trouverez le génome de référence et son annotation. Cela correspond au chromosome Z du poulet. Les données ont été récupérées à partir de la base de données Ensembl 113.

Au cours de ce TP, vous allez utiliser les outils suivants :

- FASTQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>)
- Bowtie 2 (<https://bowtie-bio.sourceforge.net/bowtie2/index.shtml>)
- samtools (<https://samtools.org/>)
- bedtools (<https://bedtools.readthedocs.io/en/latest/>)
- IGV (<https://igv.org/>)
- MACS (<https://github.com/macs3-project/MACS>)

Sur ces pages web, vous pouvez trouver une aide détaillée pour les différents outils. L'aide est également disponible en ligne de commande. Les outils vous sont fournis dans une image Singularity (chromatin_tools.sif). Ils sont donc accessibles par des commandes de ce type :

```
singularity exec chromatin_tools.sif samtools --help
```

L'utilisation des outils IA tels que ChatGPT, Claude, *etc* est strictement interdite au cours du TP. Utilisez un éditeur de texte tel que gedit pour écrire des scripts. Ces scripts devraient permettre de reproduire toutes vos analyses.

2 Contrôle de qualité de données

1. Combien a-t-on de lectures pour chaque échantillon ?
2. Réalisez un contrôle de qualité avec FASTQC (commande `fastqc`).

3 Alignement des lectures sur le génome

1. En utilisant Bowtie 2, plus précisément la commande `bowtie2-build`, construisez un index pour le génome de référence.
2. Alignez les lectures sur le génome de référence, toujours en utilisant Bowtie 2 (commande `bowtie2`).
3. En utilisant la commande `samtools sort`, triez par coordonnée l'alignement obtenu, en le convertissant au format BAM.
4. En utilisant `bedtools genomecov`, calculez la couverture génomique (c'est à dire, le nombre de lectures alignées à chaque position du génome) pour ces échantillons. Enregistrez les résultats au format `bedGraph`.
5. Visualisez les données en utilisant le navigateur de génome IGV. Commencez par charger le génome de référence (menu Genome, Load from file). Chargez ensuite les annotations des gènes, et les pistes au format `bedGraph` obtenues précédemment. Renommez les pistes pour plus de clarté.

4 Détection de pics

1. En utilisant la commande `macs3 callpeak`, détectez les régions enrichies en signal ChIPseq.
2. Visualisez les positions des pics grâce au navigateur IGV.
3. Est-ce que l'on peut observer une co-localisation des pics obtenus pour les différentes marques d'histones ? Vérifiez s'il existe des intersections entre les coordonnées des pics en utilisant l'outil `bedtools intersect`.

4. *Analysez la localisation des pics par rapport aux promoteurs des gènes.*