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

Anamaria Necsulea

anamaria.necsulea@univ-lyon1.fr

Février 2026

1 Introduction

The FAANG consortium (`https://data.faang.org/home`) aims to obtain a functional annotation for animal genomes, focusing on animals of agricultural interest. This consortium has generated an extensive collection of ChIP-seq data for chicken, for several histone modification types and for several organs and developmental stages. pour objectif de construire une annotation fonctionnelle des génomes d'animaux.

During this practical session, we will use the data in the following table:

Accession number	Organ	Individual	ChIP-seq target
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

The sequencing data is available at the following address: `http://pbil.univ-lyon1.fr/members/necsulea/MADT_2026`. In the same archive, you will find the sequence of the reference genome and its annotation. This corresponds to the Z chromosome of chicken. The genome sequence and annotation were downloaded from Ensembl 113.
You will use the following tools:

- 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`)

On the corresponding websites, you can find detailed help pages for these tools. Help is also available in the command line. The tools are provided to you within a Singularity image(chromatin_tools.sif). They are thus available through this type of command:

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

The use of AI tools such as ChatGPT, Claude, *etc* is strictly forbidden during the practical session. Use a text editor such as `gedit` to write scripts, which should allow you to reproduce all your analyses.

2 Data quality

1. How many reads are there for each sample?
2. Perform a quality control analysis with FASTQC (`fastqc` in the command line).

3 Read mapping

1. Using the `bowtie2-build` command, construct an index for the reference genome.
2. Using the `bowtie2` command, align the reads on the reference genome.
3. Using `samtools sort`, sort the resulting alignments and save them in BAM format.
4. Using `bedtools genomecov`, compute the genomic coverage (that is, the number of reads aligned at each genome position) for these data. Save the results in bedGraph format.
5. View the results in IGV. Start by changing the reference genome (menu Genome, Load from file). Then load the genome annotations and the bedGraph tracks obtained previously. Rename the tracks for more clarity.

4 ChIP-seq peak detection

1. Using `macs3 callpeak`, detect genomic regions that are enriched in ChIPseq signal.
2. View these resuslts using the IGV browser.
3. Can we observe a co-localization between the different histone marks? Check for intersections between sets of peaks coordinates using `bedtools intersect`.
4. *Analyse the localisation of the ChIPseq peaks with respect to gene promoters.*