

ChIP-seq & BigData workshops

Analytical part

Raw data available (2024-2025 session)

```
bmoindro3@clust-slurm-client2:~/shared/projects/big_data_psaclay/students_M2/chip-seq/2024-2025$ tree
.
├── fastq_files
│   ├── Input_RA_PC.fastq.gz
│   ├── Input_RA_PD.fastq.gz
│   ├── Input_RB_PF.fastq.gz
│   ├── Input_RB_PI.fastq.gz
│   ├── K27_RA_PA.fastq.gz
│   ├── K27_RB_PH.fastq.gz
│   ├── K27_RB_PI.fastq.gz
│   ├── K4_RA_PA.fastq.gz
│   ├── K4_RA_PC.fastq.gz
│   ├── K4_RB_PG.fastq.gz
│   ├── K4_RB_PJ.fastq.gz
│   ├── K9_RA_PB.fastq.gz
│   ├── K9_RA_PD.fastq.gz
│   ├── K9_RB_PG.fastq.gz
│   ├── K9_RB_PH.fastq.gz
│   ├── Mock_RA_PC.fastq.gz
│   ├── Mock_RA_PD.fastq.gz
│   ├── Mock_RB_PF.fastq.gz
│   ├── Mock_RB_PI.fastq.gz
│   ├── PolIII_RA_PB.fastq.gz
│   ├── PolIII_RA_PE.fastq.gz
│   ├── PolIII_RB_PF.fastq.gz
│   └── PolIII_RB_PJ.fastq.gz
```

ChIP-seq & BigData workshops

Analytical part

Raw data available (2024-2025 session)

```

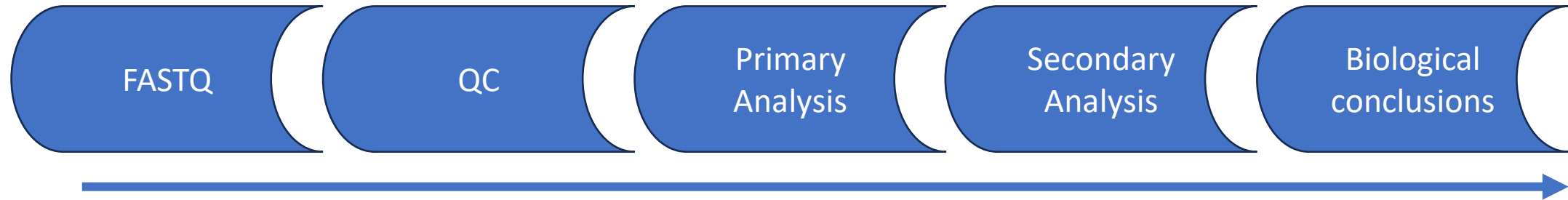
bmoindro3@clust-slurm-client2:~/shared/projects/big_data_psaclay/students_M2/chip-seq/2024-2025$ tree
.
├── fastq_files
│   ├── Input_RA_PC.fastq.gz
│   ├── Input_RA_PD.fastq.gz
│   ├── Input_RB_PF.fastq.gz
│   ├── Input_RB_PI.fastq.gz
│   ├── K27_RA_PA.fastq.gz
│   ├── K27_RB_PH.fastq.gz
│   ├── K27_RB_PI.fastq.gz
│   ├── K4_RA_PA.fastq.gz
│   ├── K4_RA_PC.fastq.gz
│   ├── K4_RB_PG.fastq.gz
│   ├── K4_RB_PJ.fastq.gz
│   ├── K9_RA_PE.fastq.gz
│   ├── K9_RA_PD.fastq.gz
│   ├── K9_RB_PG.fastq.gz
│   ├── K9_RB_PH.fastq.gz
│   ├── Mock_RA_PC.fastq.gz
│   ├── Mock_RA_PD.fastq.gz
│   ├── Mock_RB_PF.fastq.gz
│   ├── Mock_RB_PI.fastq.gz
│   ├── PolIII_RA_PE.fastq.gz
│   ├── PolIII_RA_PF.fastq.gz
│   ├── PolIII_RB_PF.fastq.gz
│   └── PolIII_RB_PJ.fastq.gz
└──

```

fastq_files	Ab condition	Replicate	Pair
Input_RA_PC.fastq.gz	Input	ReplicateA	PA
Input_RA_PD.fastq.gz			PB
Input_RB_PF.fastq.gz			PC
Input_RB_PI.fastq.gz			PD
K27_RA_PA.fastq.gz	H3K27me3	ReplicateB	PE
K27_RB_PH.fastq.gz			PF
K27_RB_PI.fastq.gz			PG
K4_RA_PA.fastq.gz	H3K4me3		PH
K4_RA_PC.fastq.gz			PI
K4_RB_PG.fastq.gz			PJ
K4_RB_PJ.fastq.gz			
K9_RA_PE.fastq.gz	H3K9me3		
K9_RA_PD.fastq.gz			
K9_RB_PG.fastq.gz			
K9_RB_PH.fastq.gz			
Mock_RA_PC.fastq.gz	IgG		
Mock_RA_PD.fastq.gz			
Mock_RB_PF.fastq.gz			
Mock_RB_PI.fastq.gz			
PolIII_RA_PE.fastq.gz	Pol2.S5		
PolIII_RA_PF.fastq.gz			
PolIII_RB_PF.fastq.gz			
PolIII_RB_PJ.fastq.gz			

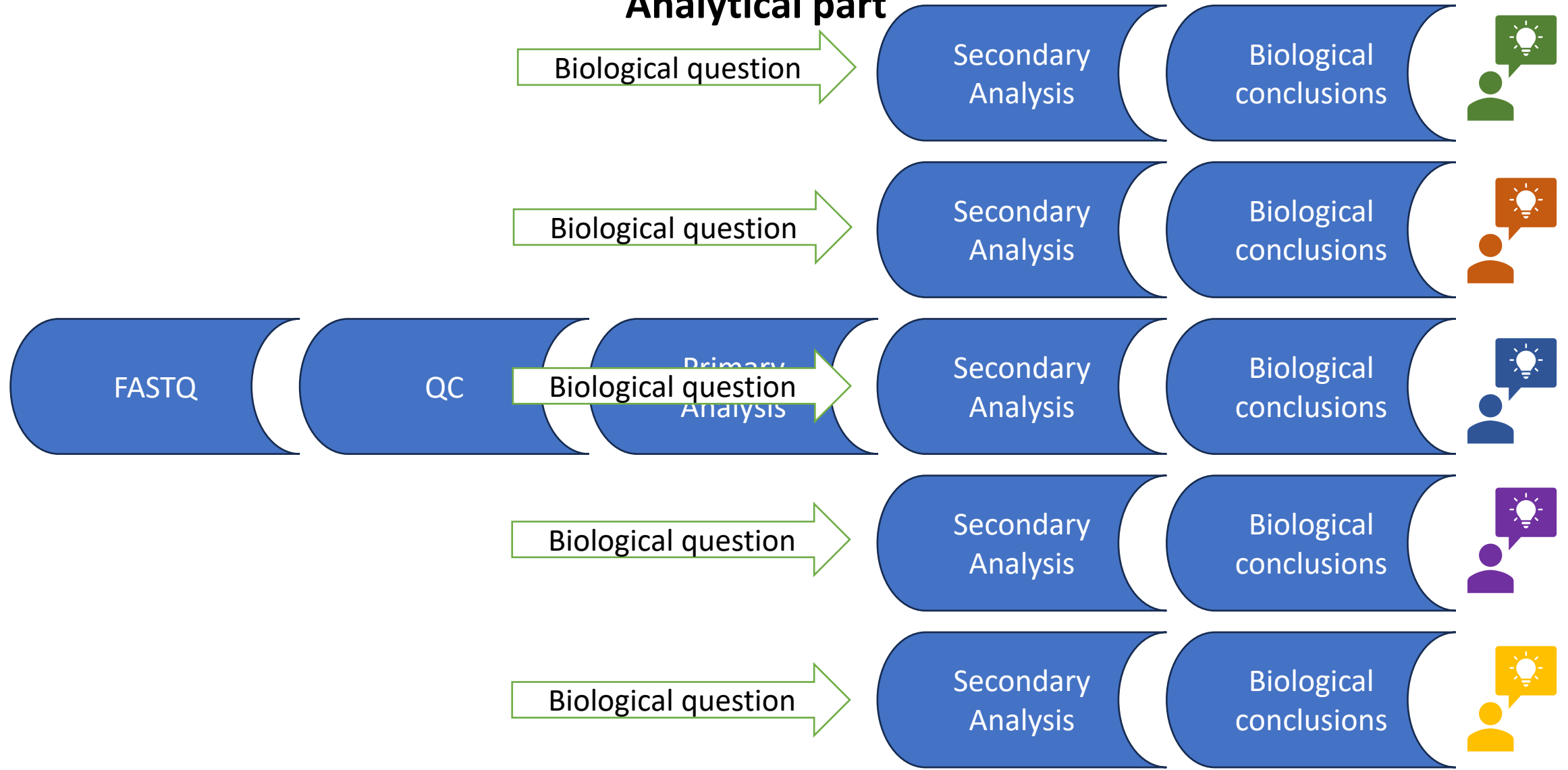
ChIP-seq & BigData workshops

Analytical part



ChIP-seq & BigData workshops

Analytical part



ChIP-seq & BigData workshops

Bioinformatic ressources



JupyterLab: Core
System Installed App



RStudio Server: Core
System Installed App



Desktop core

<https://ondemand.cluster.france-bioinformatique.fr/>

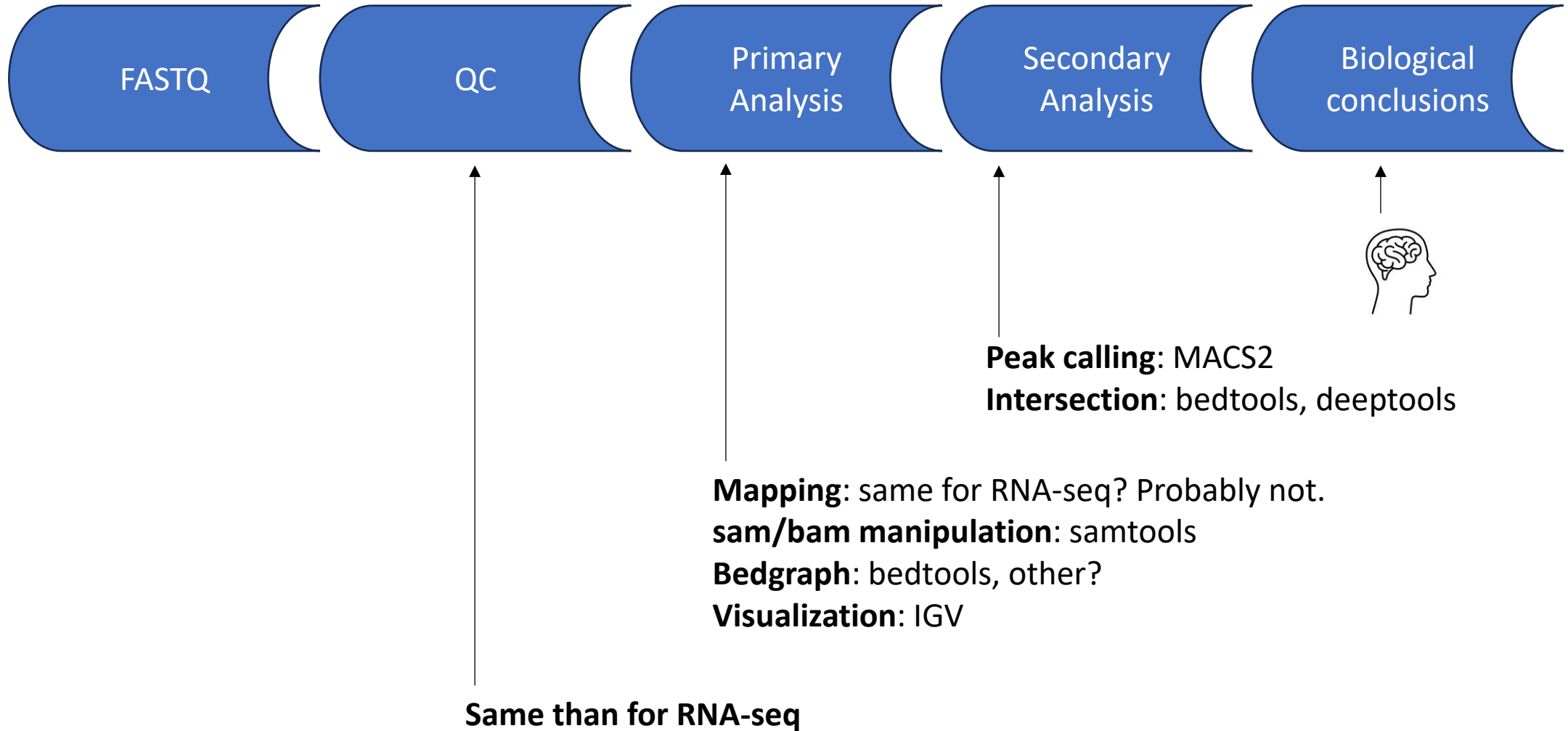
+ FTP



<https://usegalaxy.fr/> ←
<https://usegalaxy.eu/>

ChIP-seq & BigData workshops

Bioinformatic tools



ChIP-seq & BigData workshops Evaluation

2nd report:

40% of the grade

Due: 22 Nov 2024

Importantly: the analyses done and shown must be used to **convey a biological/technical message**. They aim at **answering the biological/technical question raised in the introduction**. Every **pair of students** can tackle its **own biological/technical question**.

Type of *in-silico* analyses expected in this report

- Some **informative numbers** regarding the sequencing run

- One or several **bedgraphs**

- **Intersections** between the ChIP-data generated during the practical and something else (other ChIP data; genomic feature; ...)

universitè
Magna Graecia
UNIVERSITÀ DELLA SALERNO

ChIPseq data analysis

Master 2 GenE2
ChIPseq workshop
2022/2023

[Students' first and last names]

[Title]

Cover page

Fig1

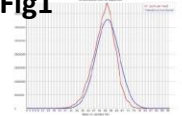


Figure 1: Complete legend of the figure

Fig2

Run	Library	Size	Reads	Quality
Run1	Lib1	150 bp	100M	30
Run2	Lib2	150 bp	100M	30
Run3	Lib3	150 bp	100M	30
Run4	Lib4	150 bp	100M	30

Table 1: Complete legend of the table

Intro

Very brief general introduction

Explicitly write down the biological/technical question you aim to tackle in the report

Pres. of data

How is the output of the sequencing run?
What about the quality of the sequencing?

Methods

Methods

If the method has been already used during the "big data" workshop, just cite the program and the key parameters. Otherwise provide a brief description of what you have done and how.

Fig3

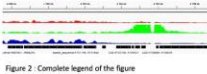


Figure 2: Complete legend of the figure

Fig4

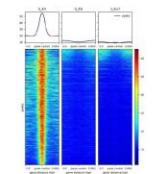


Figure 3: Complete legend of the figure

Fig5

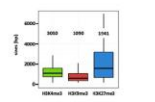


Figure 4: Complete legend of the figure

Figure 5: Complete legend of the figure

Results of the analysis

Results of the analysis = Interpretation

Here, your aim is to answer (to the best as the data can) a biological/technical question raised in the introduction. Select the piece of data you need to provide to support your reasoning.

- Describe general observations of the ChIP data. If a replicate has been excluded, justify why.
- The data presented in this section should demonstrate how the experiment's objectives (= answer to biological/technical question) have been met.
- In the text, convey the main findings of the experiment. Text and findings must be supported by the Figures (and/or tables).
- Interpret your findings in light of the field's knowledge.

Conclusion

Conclusion

Summarize the main findings. Comment on the quality of the ChIPseq. Conclude on the biological relevance of your results.

No new information should be given in the conclusion

Propose further work or potential improvements identified during the experiments/analyses.

Writing instructions

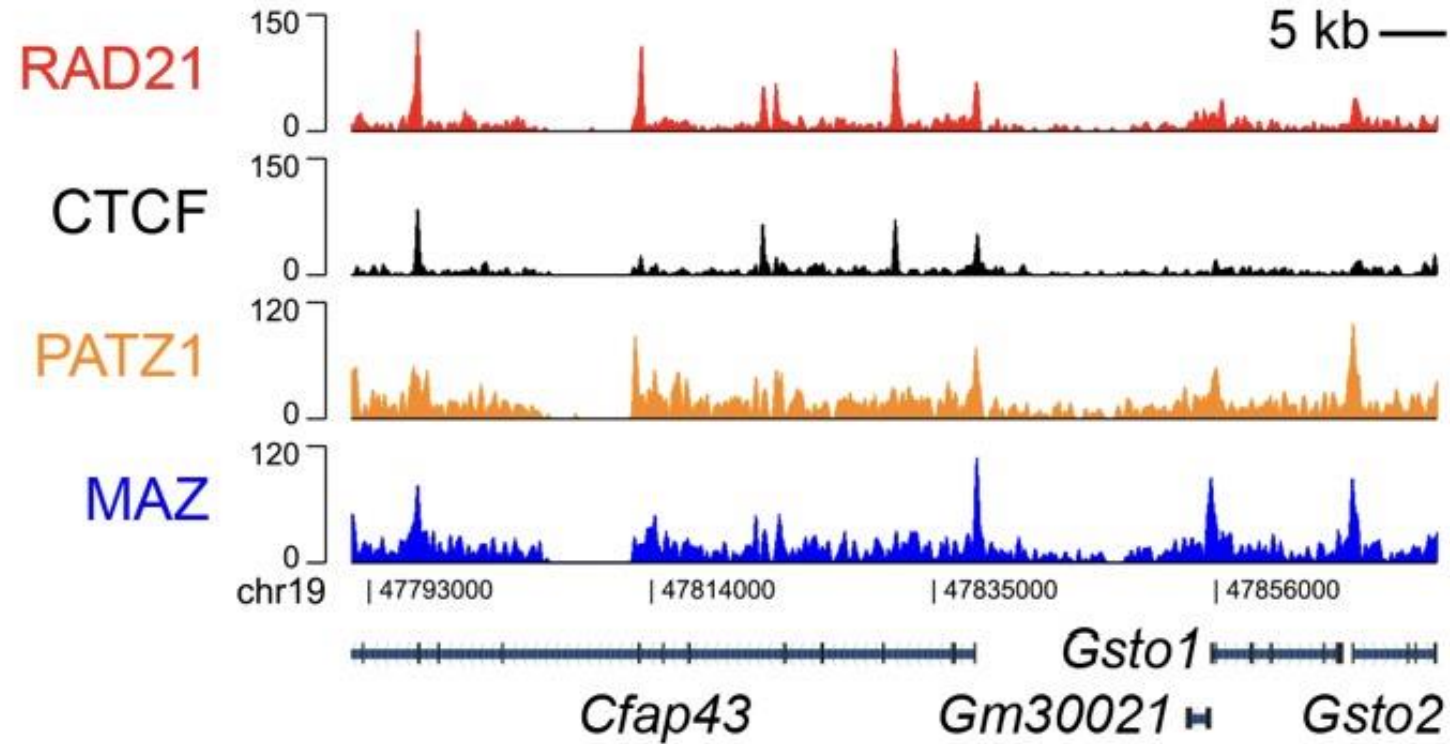
- Write with your own words.
- Copy-paste (from websites, existing publications, ...) is prohibited (plagiarism).
- Cite your sources when needed
- Avoid sprawling sentences. Instead, adopt concise / efficient / precise / scientific writing (as done in publications).

Additional instructions

- one report per two-person team
- English.
- A4 page (21 x 29.7 cm)
- Font for the main text: Arial, 11pt, single-space
- Margin: 2.5cm (1 inch) on left/right/top/bottom
- Figures in the main text. With Legend, axis, units, etc...
- Electronic version (not printed). PDF. Deposited on [eCampus \(?\)](#)

ChIP-seq & BigData workshops

Bedgraph Visualization, an example

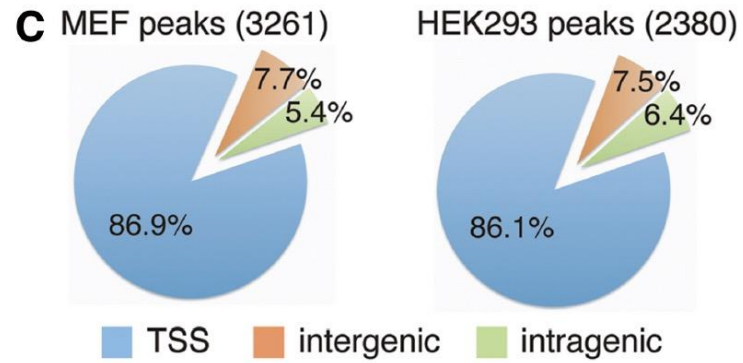


Can be done at different scales
(from gene to whole chromosome),
depending on your needs

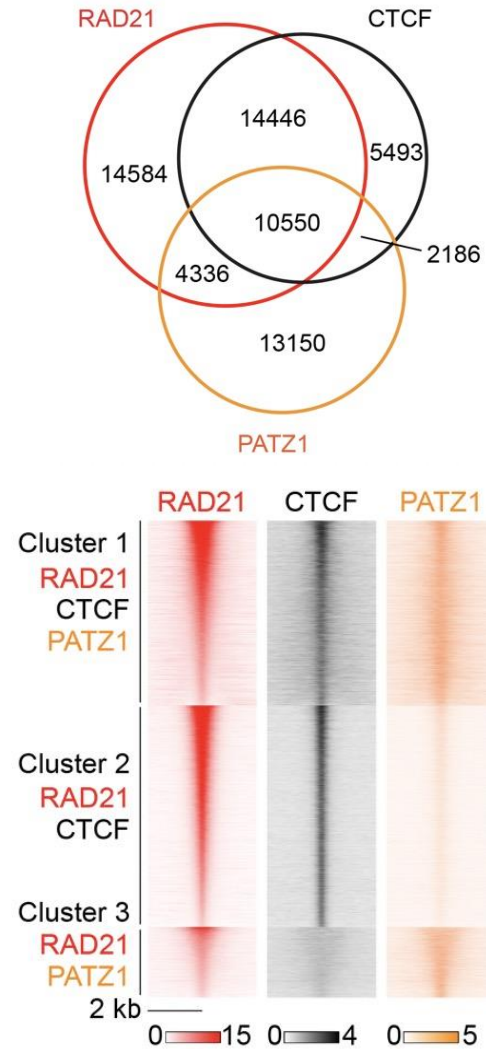
ChIP-seq & BigData workshops

Intersection: a few examples, among many...

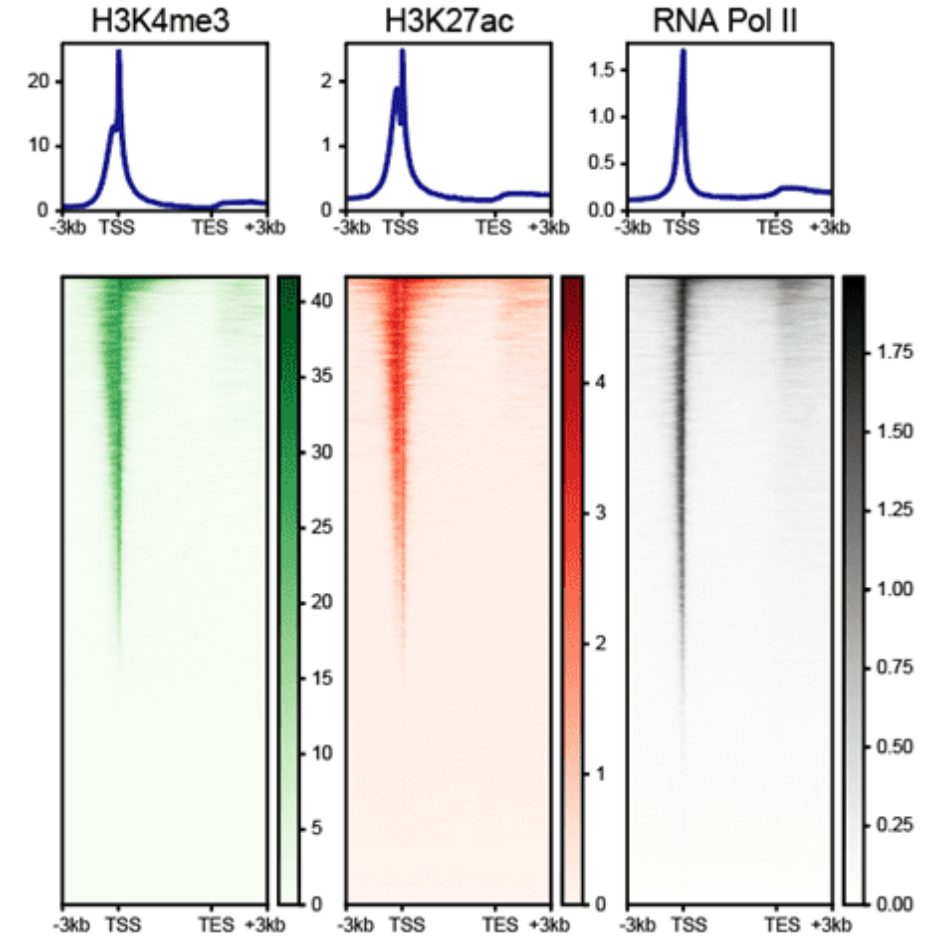
ChIP <-> Genome Annotation



ChIP peak <-> ChIP peak



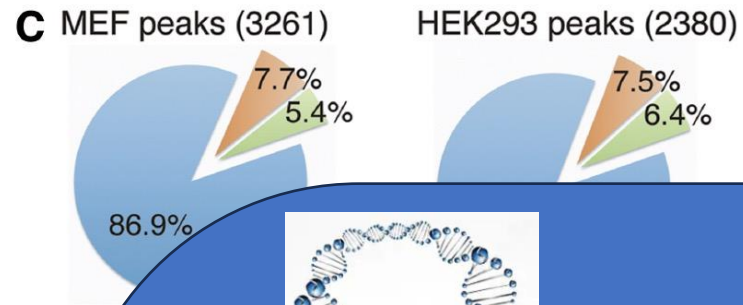
ChIP <-> ChIP <-> Genomic feature



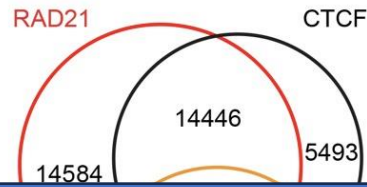
ChIP-seq & BigData workshops

Intersection: a few examples, among many...

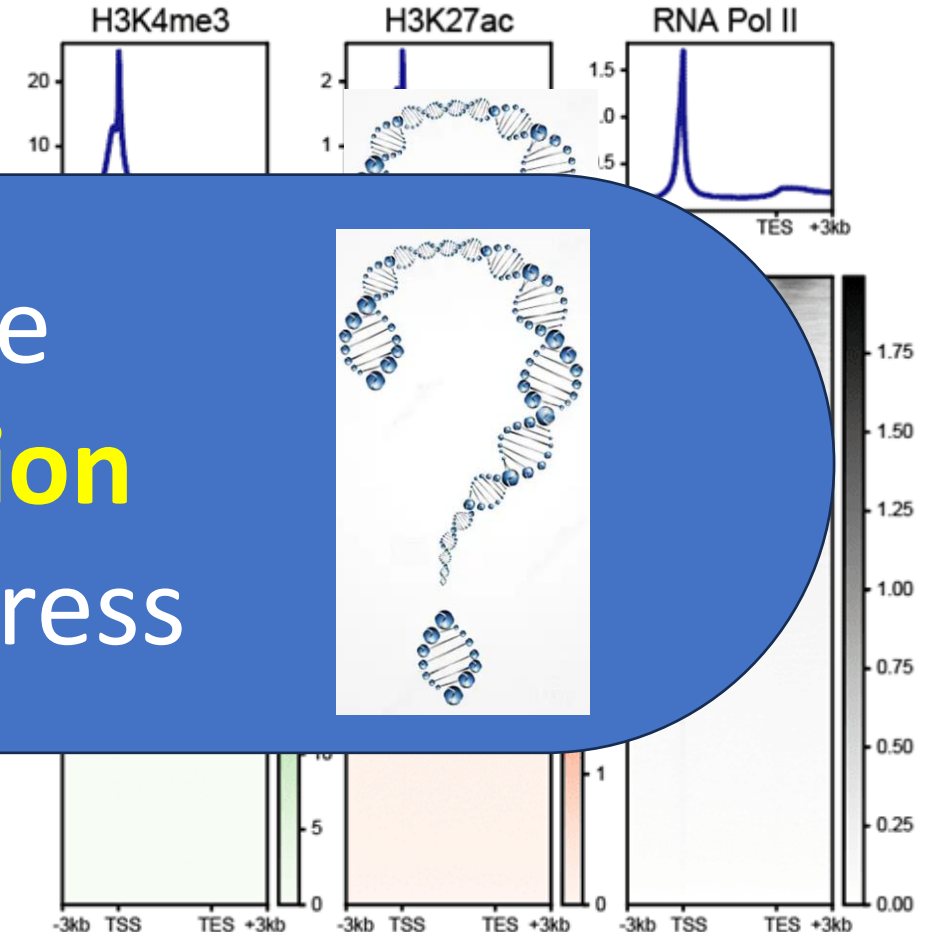
ChIP <-> Genome Annotation



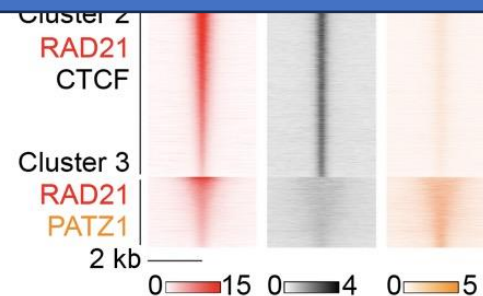
ChIP peak <-> ChIP peak



ChIP <-> ChIP <-> Genomic feature



Depends on the **biological question** you WANT to address



ChIP-seq & BigData workshops Program

Monday	18/11	09h00 - 11h00	N0-001 (I2BC)	Benoit Moindrot	ChIPseq data analyses
		13h30 - 16h00	N0-001 (I2BC)	Pierre Grognet	ChIPseq data analyses
Tuesday	19/11	09h30 - 12h00	N0-009 (I2BC)	Gaëlle Lelandais	ChIPseq data analyses
		14h00 - 17h30	—	Autonomous work	ChIPseq data analyses
Wednesday	20/11	09h30 - 12h00	N0-001 (I2BC)	Pierre Grognet	ChIPseq data analyses
		14h00 - 17h30	N0-001 (I2BC)	Benoit Moindrot	ChIPseq data analyses
Thursday	21/11	09h30 - 12h00	—	Autonomous work	Report writing
		14h00 - 17h00	—	Autonomous work	Report writing
Friday	22/11	09h30 - 12h00	N0-001 (I2BC)	Gaëlle Lelandais	Basics of using R and RStudio
		14h00 - 17h30	N0-001 (I2BC)	Autonomous work	ChIPseq report writing and RNAseq report feedback (G. Lelandais)