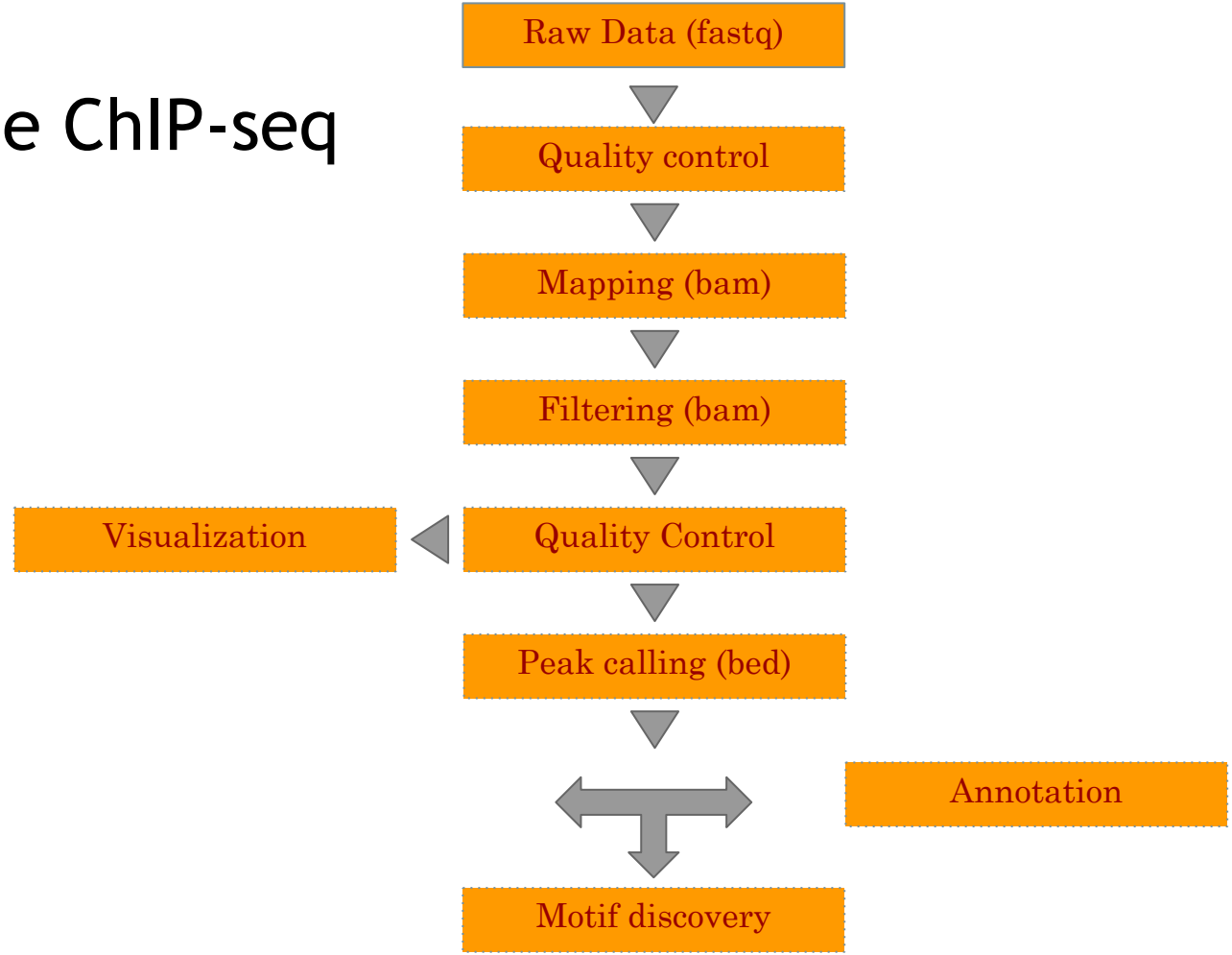
# Conclusions atelier Chip-Seq

École de bioinformatique AVIESAN-IFB 2017

# Bilan du pipeline ChIP-seq



Mettre photo

# Topics for discussion

*It's common practice to sequence the input deeper than the treatment. Why ?*

*Importance of the mapping tool ?*

*Single-end or paired-end sequencing ?*

*Which control to use ?*

*Why do we find peaks that do not have two opposite read densities ?*

*What to do with replicates ?*

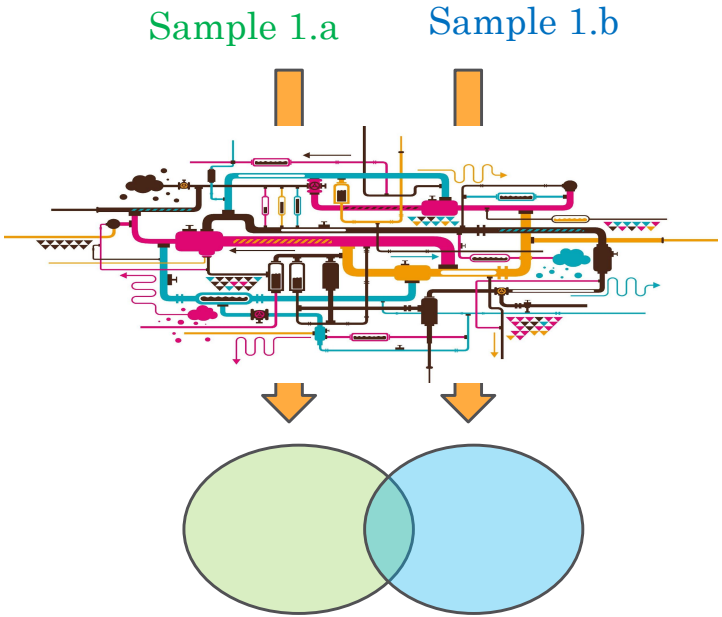
*Systematic biais ?*

*Low number of cells ?*

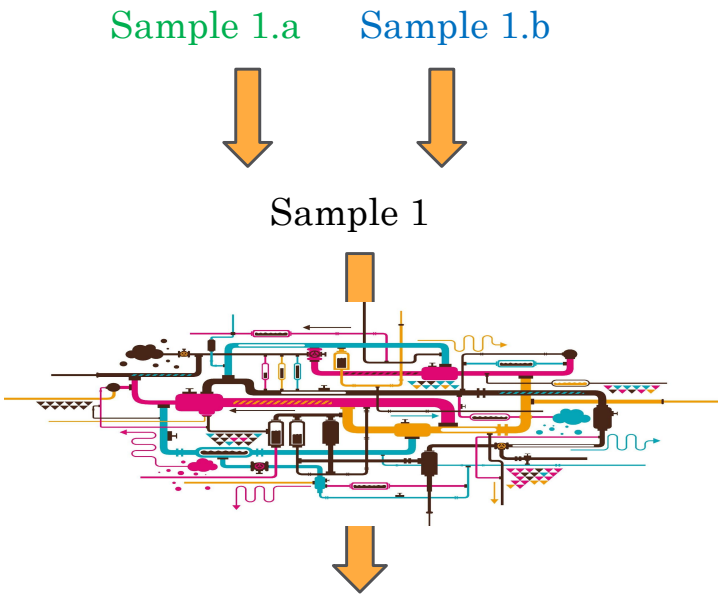
# How to deal with replicates

# How to deal with replicates

Analyze samples separately and takes union or intersection of resulting peaks




Merge samples prior to the peak calling (e.g recommended by MACS) => “pooling”

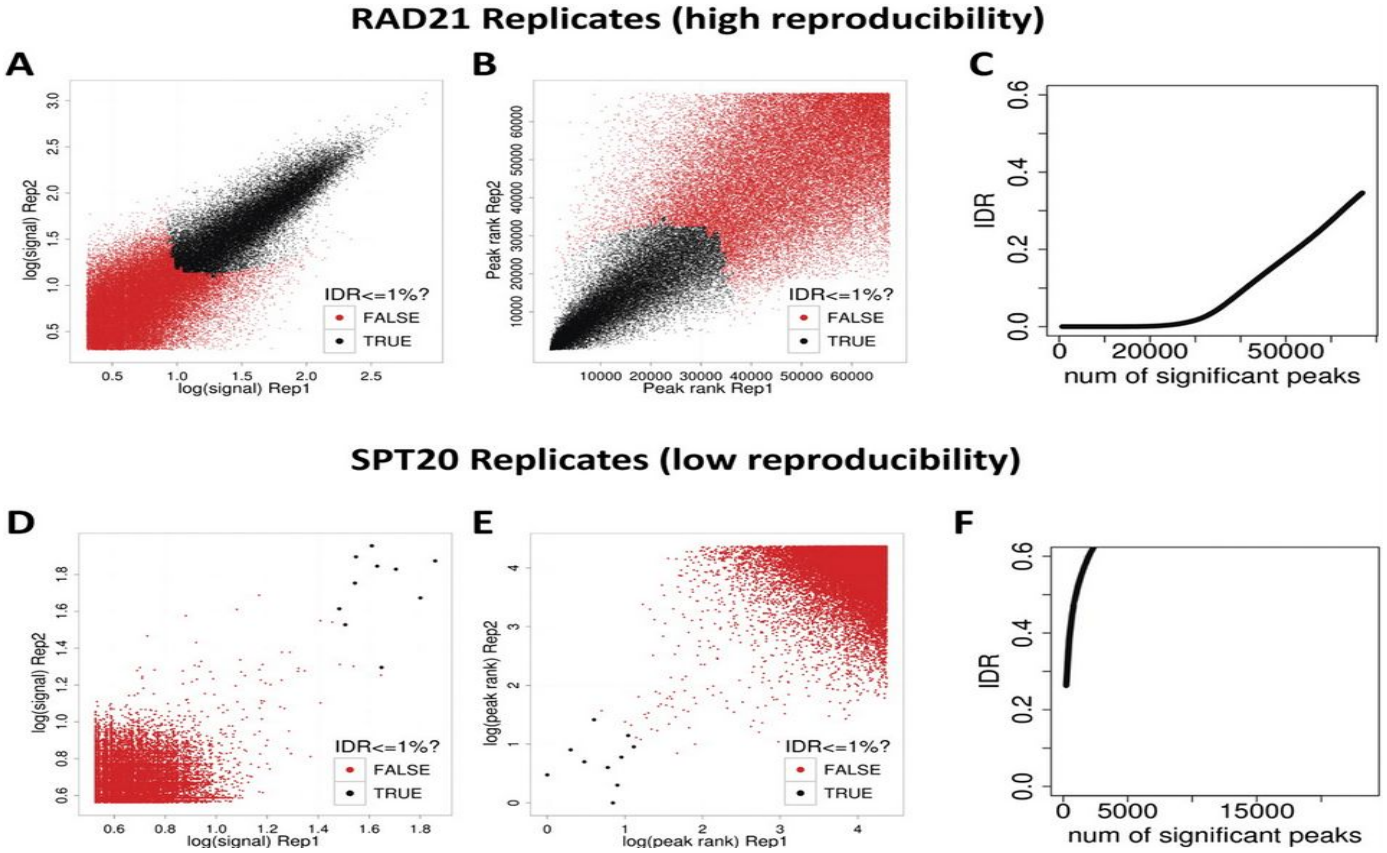


# IDR - Irreproducible Discovery Rate (ENCODE)

- Measures consistency between replicates
- Uses reproducibility in score rankings between peaks in each replicate to determine an optimal cutoff for significance.
- Idea:
  - The most significant peaks are expected to have high consistency between replicates
  - The peaks with low significance are expected to have low consistency

https://sites.google.com/site/anshulkundaje/projects/idr

# IDR




(!) IDR doesn't work on broad source data!

# Controls

# Different types of controls

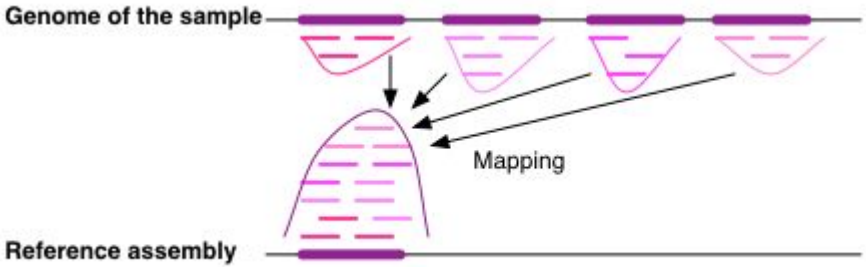
- **Input DNA** : controls biases due to chromatin fragmentation

=> higher background signal due to:

-more accessible chromatin

-more susceptible to shearing because of nucleotide composition

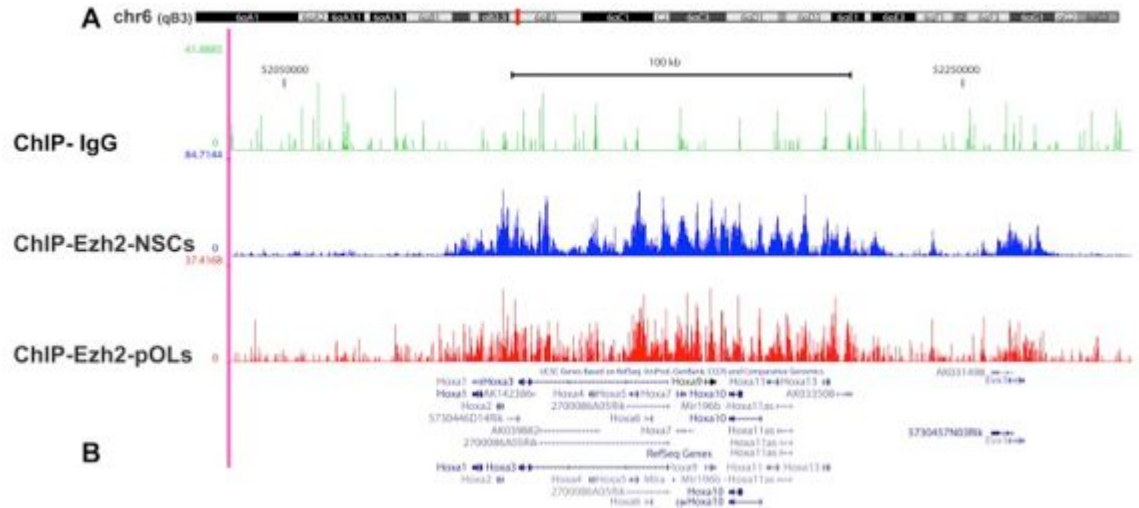
-Duplicated regions in the sample genome




- Commonly used, produces a complex library
- Most peak-calling programs were developped for using input DNA as control

# Different types of controls

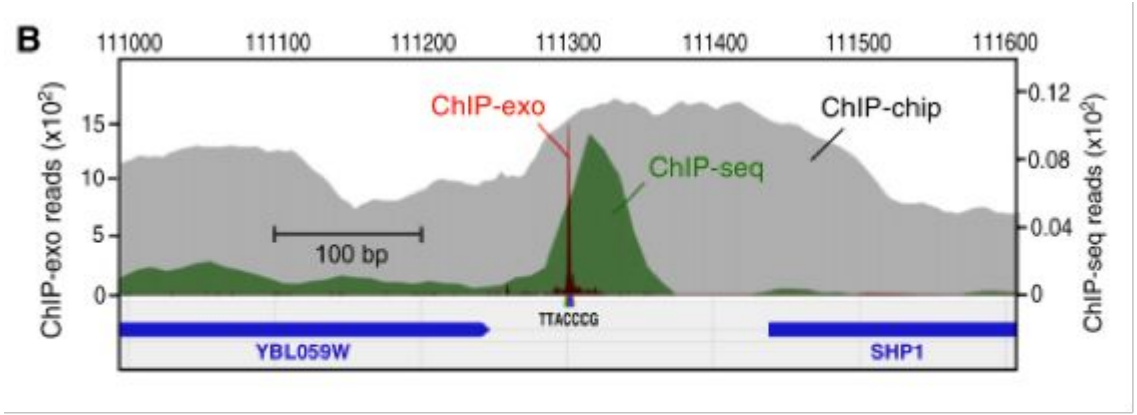
- **IgG** (or mock IP) : controls for non-specific IP enrichment
- Problem : low-complexity library (few reads)



# Different types of controls

- **Knockout of targeted TF**
- **Uninduced condition** (for inducible TFs)
  - Induced by Dexamethasone (Dex)
  - Control vehicule = Ethanol (EthOH)

# Beyond ChIP-seq : ChIP-exo


Rhee, H. S., & Pugh, B. F. (2011).

# Beyond ChIP-seq

**Experimental techniques**

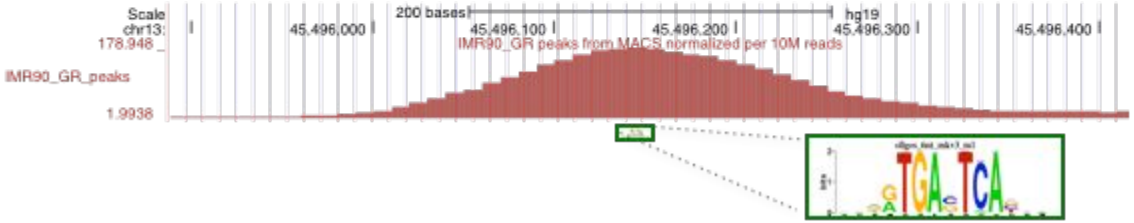
crosslink

sonication

antibody

**Improvement aimed**

higher resolution => 300bp to 1bp

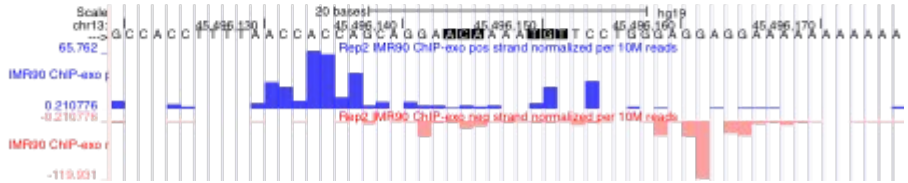


# Beyond ChIP-seq : ChIP-nexus

**Experimental techniques**


**Improvement aimed**

Get rid of PCR artifacts



# Beyond ChIP-seq : native ChIP

**Experimental techniques**

~~crosslink~~

~~sonication~~

antibody

endonuclease

**Improvement aimed**

Avoid formaldedyde crosslinking

- Formaldehyde crosslinking affects preferentially protein-protein interactions.
- Crosslinking could be the cause of hyper-signaling regions in highly transcribed sites.

# Atelier ChIP-seq: tour de table des données

Les questions qui pourraient moduler le pipeline d'analyse

Narrow peak ou broad peak ?

Paired-end ou single-end ?

Disponibilité du génome de référence (partie annotation) ?

Qualité de l'assemblage du génome ?