

# LE GÉNOME EN ACTION

## SÉQUENÇAGE HAUT DÉBIT ET ÉPIGÉNOMIQUE

# Régulation d'expression

la transcription d'une région de l'ADN nécessite

- ★ liaisons protéine-ADN (facteur de transcription et son site reconnu)
- ★ accessibilité de la chromatine

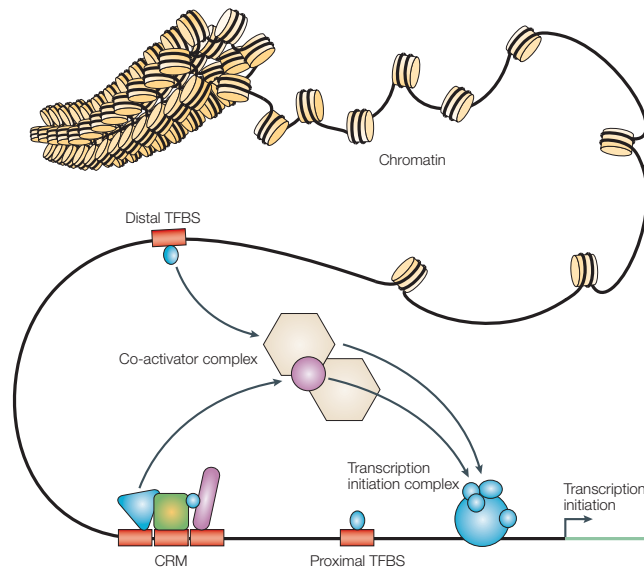
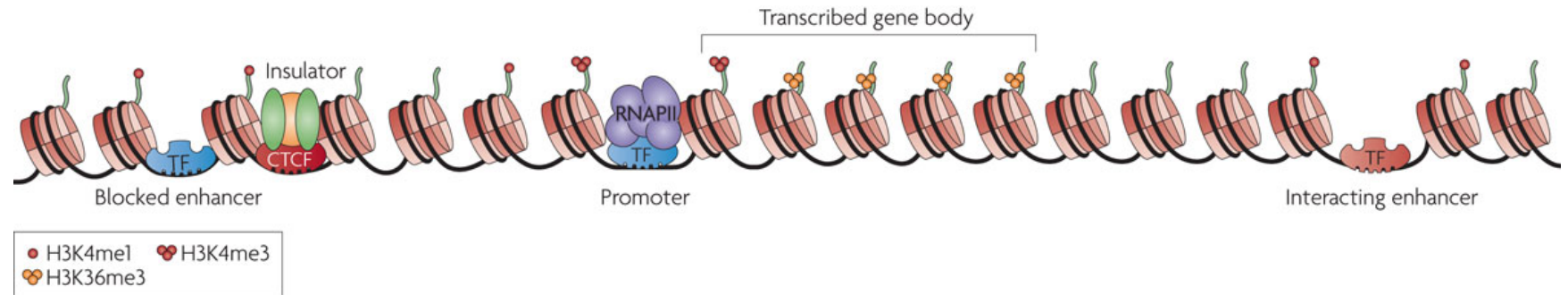


Figure 1 | **Components of transcriptional regulation.** Transcription factors (TFs) bind to specific sites (transcription-factor binding sites; TFBS) that are either proximal or distal to a transcription start site. Sets of TFs can operate in functional *cis*-regulatory modules (CRMs) to achieve specific regulatory properties. Interactions between bound TFs and cofactors stabilize the transcription-initiation machinery to enable gene expression. The regulation that is conferred by sequence-specific binding TFs is highly dependent on the three-dimensional structure of chromatin.

Wasserman & Sandelin *Nat Rev Genet* 5 :276 (2004)

# Annotation d'activité génomique



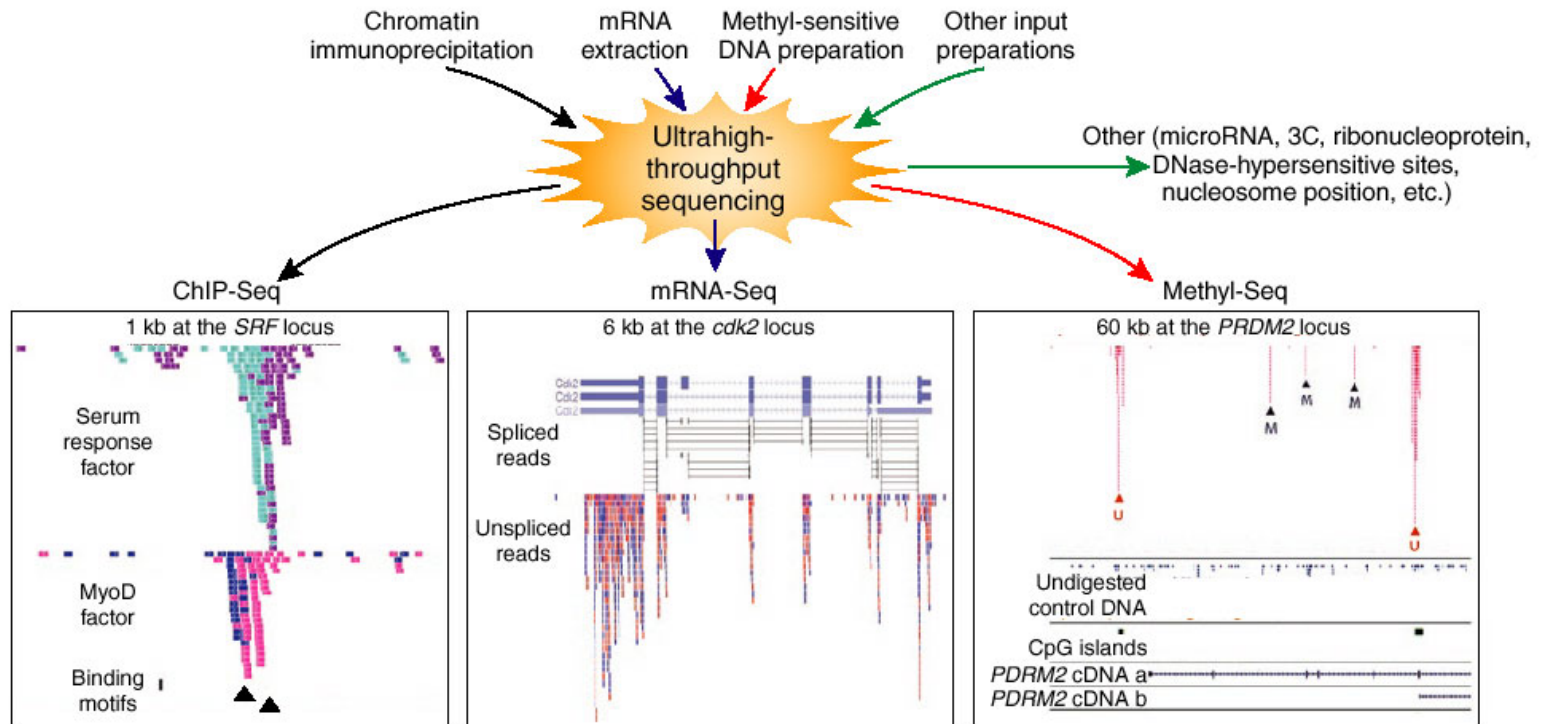
on veut annoter :

- ★ sites de liaison de facteurs de transcription
- ★ chromatine «ouverte»
- ★ méthylation de l'ADN
- ★ modification de histones
- ★ interaction ADN-ADN

Hawkins, Hon & Ren *Nat Rev Genet* 11 :476 (2010)

# On peut faire tout par séquençage...

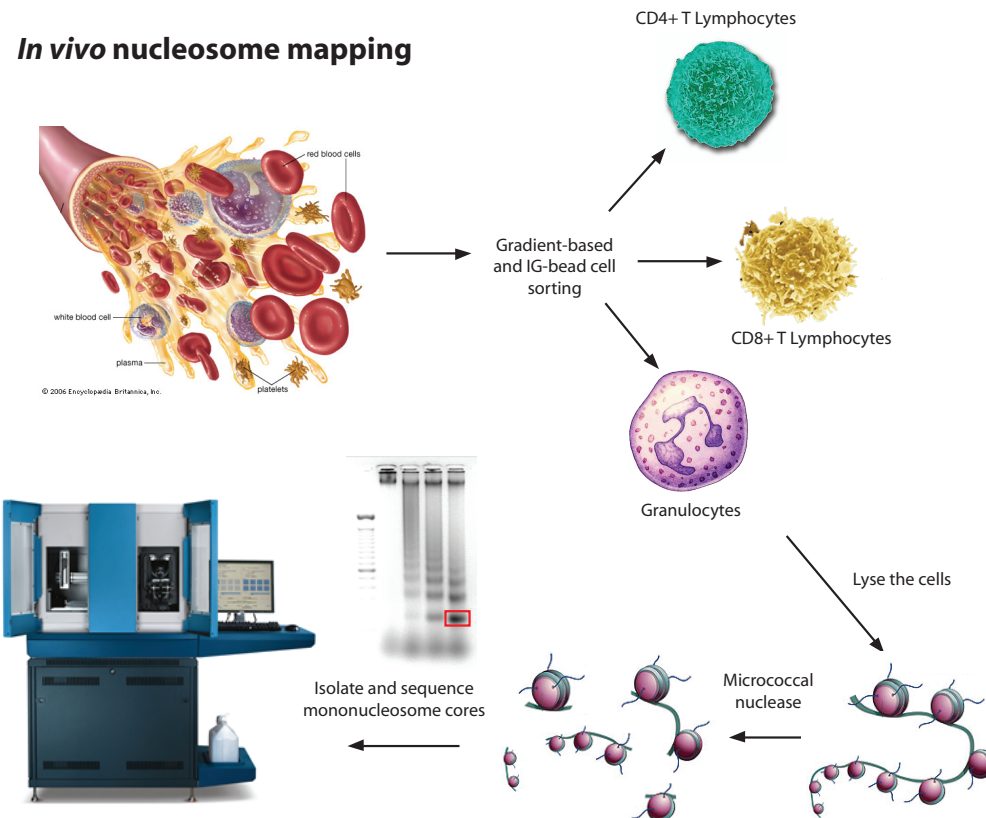
1. filtrage / enrichissement de régions d'intérêt
2. séquençage
3. alignement pour déterminer d'où viennent les morceaux



Wold & Myers *Nature Methods* 5 :19 (2008)

# MNase-seq : position des nucléosomes

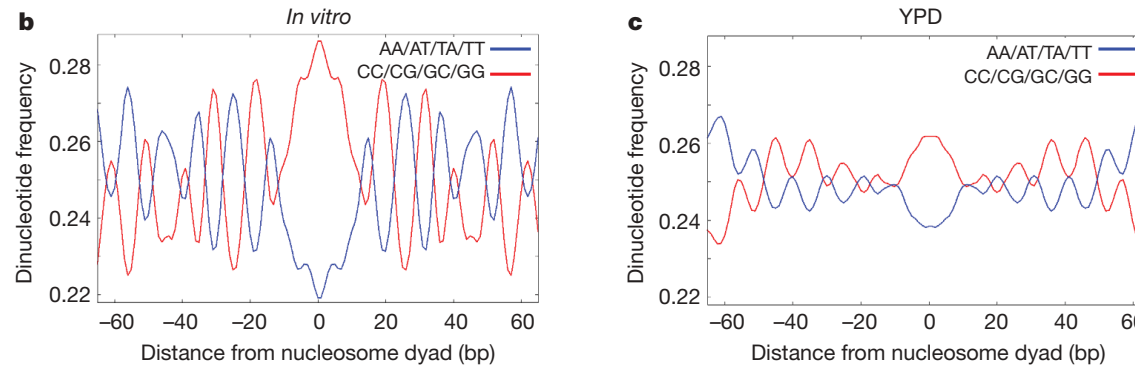
MNase :



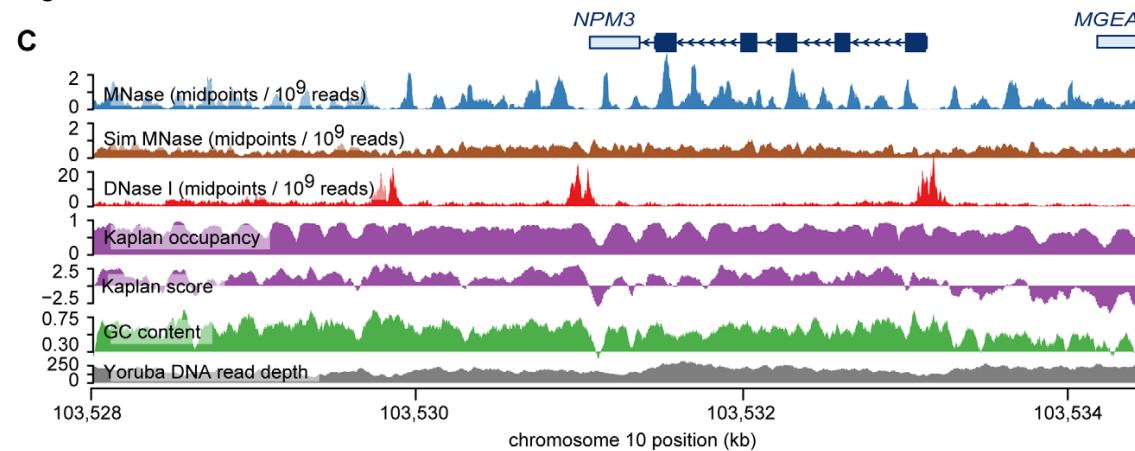
Valouev & al *Nature* 474 :516 (2011)

# Positionnement par séquence

placement déterminé par séquence génomique (dinucléotides)



... mais pas toujours exactement

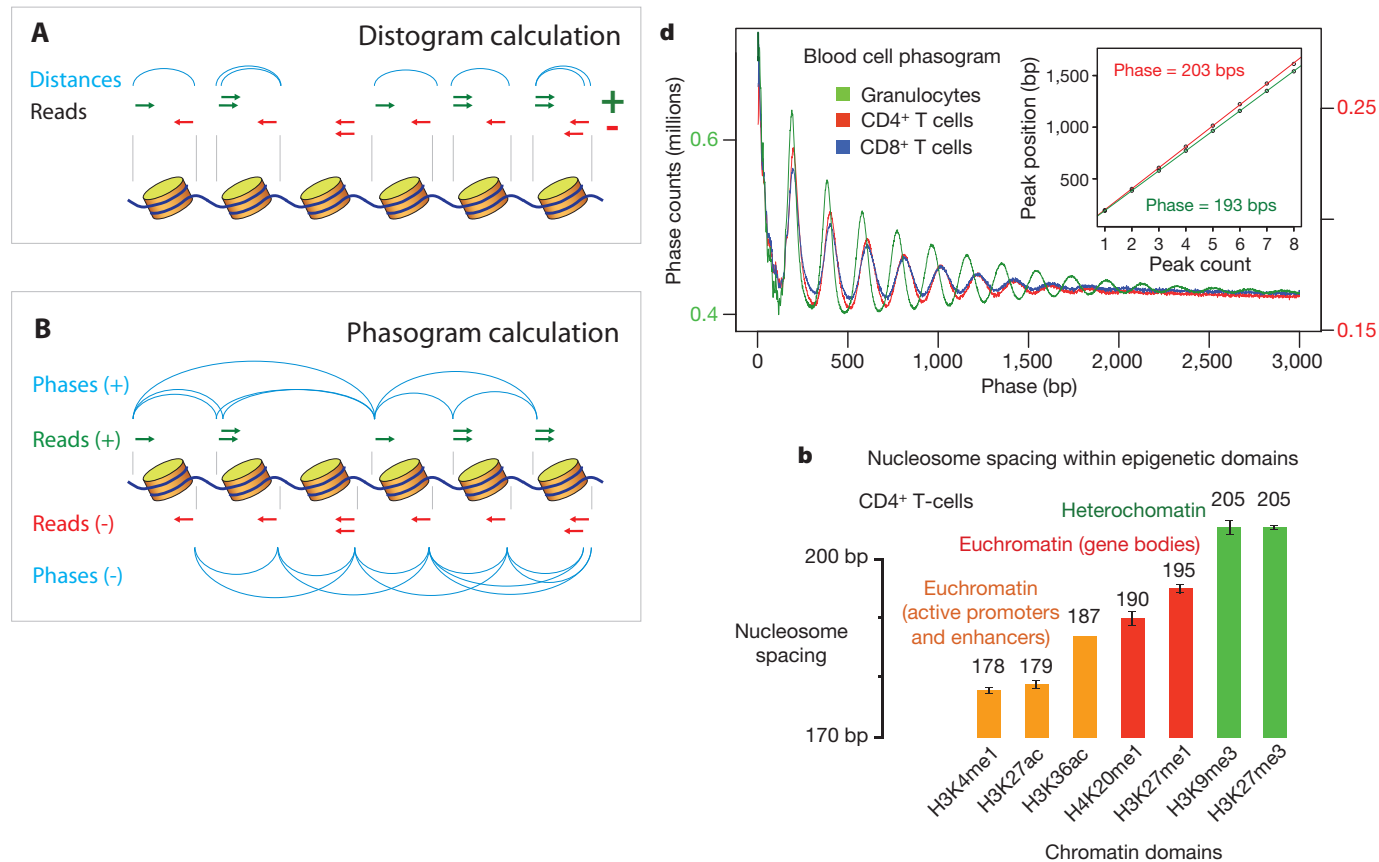


Kaplan & al *Nature* 458 :362 (2009) ; Gaffney & al. *PLoS Genetics* 8 :e1003036 (2012)

# Espace entre nucléosomes

plus proches dans des régions actives

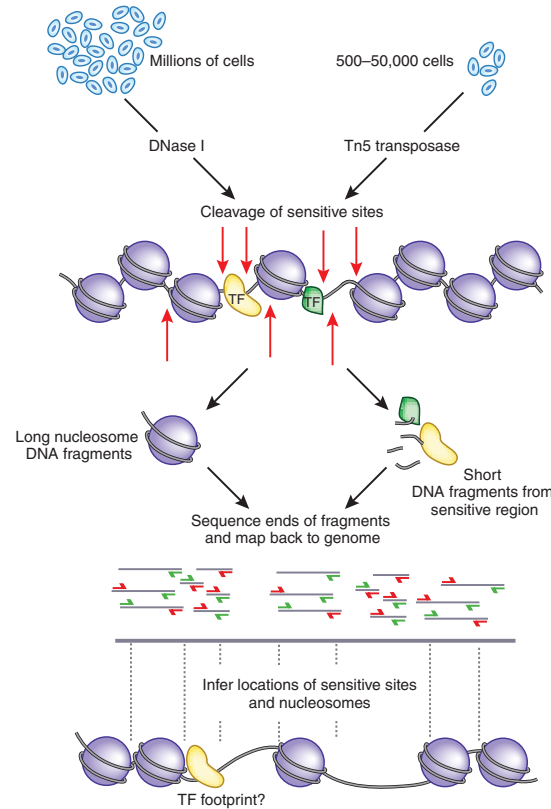
séquence + autres facteurs déterminent le placement



Valouev & al *Nature* 474 :516 (2011)

# DNase-seq : régions hypersensitives

on veut détecter la chromosome «ouverte» — site possible de liaison

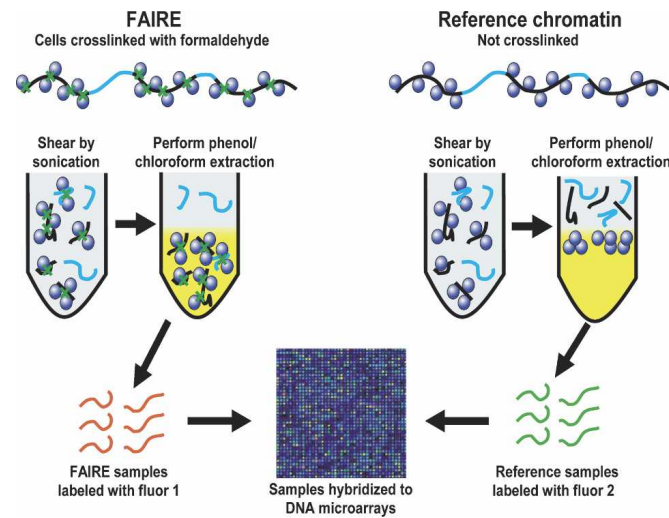


**Figure 1** | Mapping regulatory regions with paired-end DNase-seq or ATAC-seq. Short fragments come from nucleosome-depleted sequences, whereas long fragments originate from flanking nucleosomes.

Raj & McVicker *Nature Methods* 11 :39 (2014)



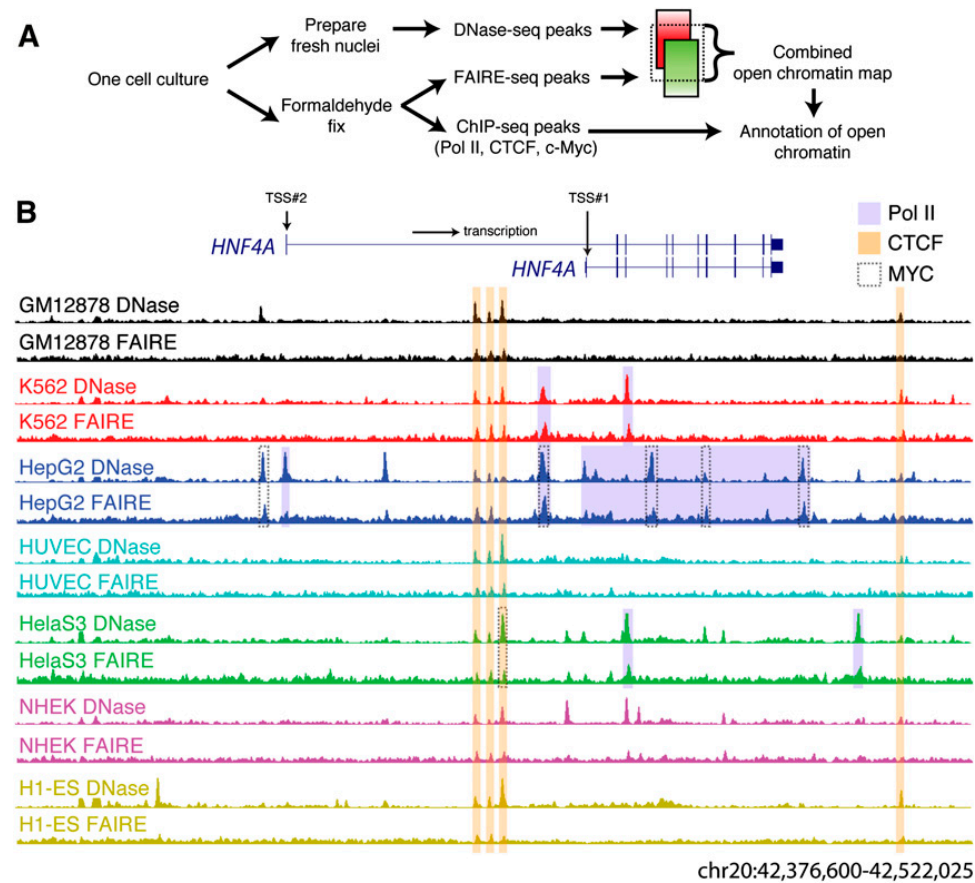
# FAIRE-seq : chromosome ouverte



**Figure 1.** FAIRE in human cells is illustrated on the *left*, while preparation of the reference is illustrated on the *right*. For FAIRE, formaldehyde is added directly to cultured cells. The crosslinked chromatin is then sheared by sonication and phenol-chloroform extracted. Crosslinking between histones and DNA (or between one histone and another) is likely to dominate the chromatin crosslinking profile (Brutlag et al. 1969; Solomon and Varshavsky 1985; Polach and Widom 1995). Covalently linked protein–DNA complexes are sequestered to the organic phase, leaving only protein-free DNA fragments in the aqueous phase. For the hybridization reference, the same procedure is performed on a portion of the cells that had not been fixed with formaldehyde, a procedure identical to a traditional phenol-chloroform extraction. DNA resulting from each procedure is then labeled with a fluorescent dye, mixed, and comparatively hybridized to DNA microarrays. In this case, we used high-density oligo-

# FAIRE et DNaseI

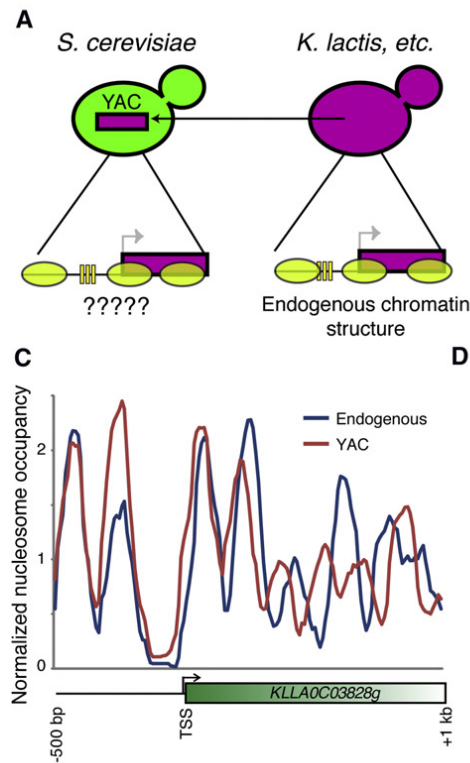
tous les deux trouvent la chromatine ouverte ( $\Leftarrow$  attachement de facteurs de transcription)



Song & al *Genome Res* 21 :1757 (2011)

# Où sont les nucléosomes ?

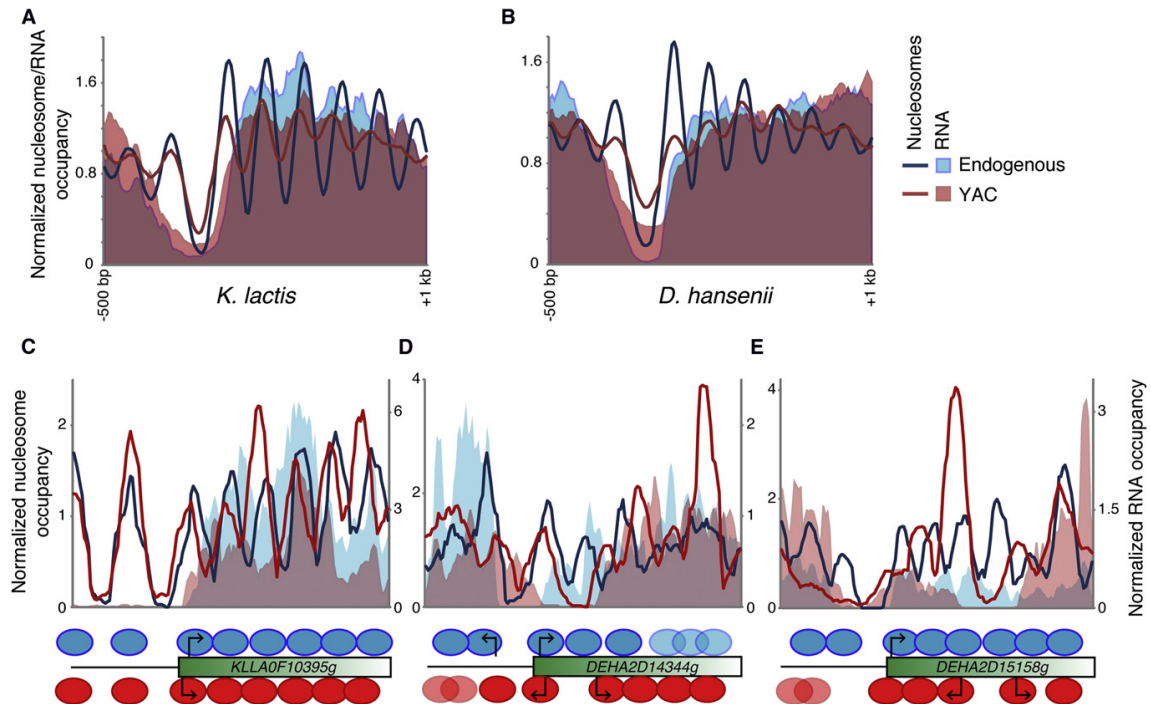
on transfère l'ADN (Yeast Artificial Chromosome) d'une espèce à un autre : qu'est-ce qui se passe aux nucléosomes ?



Hughes & al *Mol Cell* 48 :5 (2012)

# Ils bougent...

ce n'est pas simplement la séquence qui détermine le placement



**Figure 4. +1 Nucleosome Shifts Associated with Transcriptional Changes**

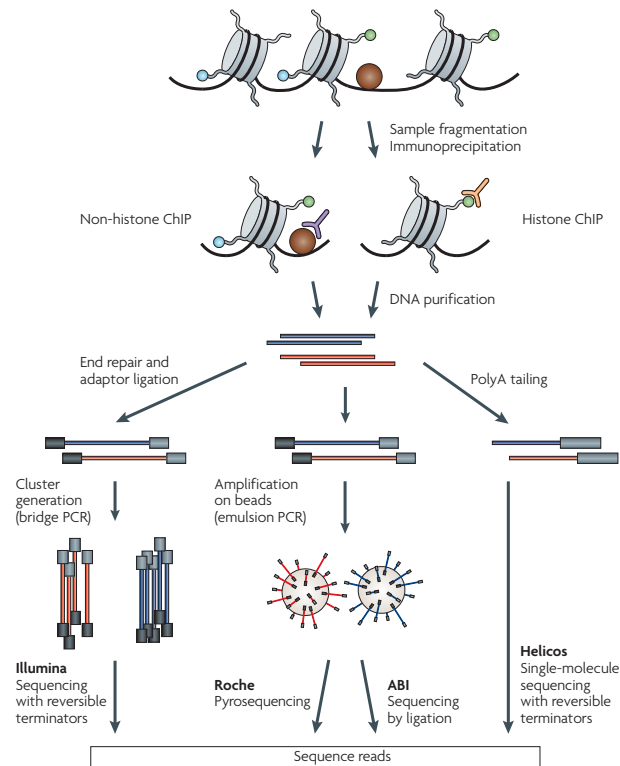
(A and B) Nucleosome data and RNA-Seq data are shown for *K. lactis* and *D. hansenii* genes in wild-type and YACs, as indicated. RNA-Seq data for YAC-derived transcripts are normalized independently from *S. cerevisiae* transcripts here—see Figures S4B and S4C for data normalized genome-wide.

(C–E) Examples of +1 nucleosome shifts associated with changes in transcription. (C) shows a moderate upstream shift in a +1 nucleosome with a similar change in transcript length, while (D) and (E) show large-scale NDR gain/loss with associated changes in transcription. Schematic interpretation of the nucleosome positioning for the endogenous gene is shown in blue above the rectangle, nucleosome positioning in the YAC is shown in red below the rectangle. Arrows



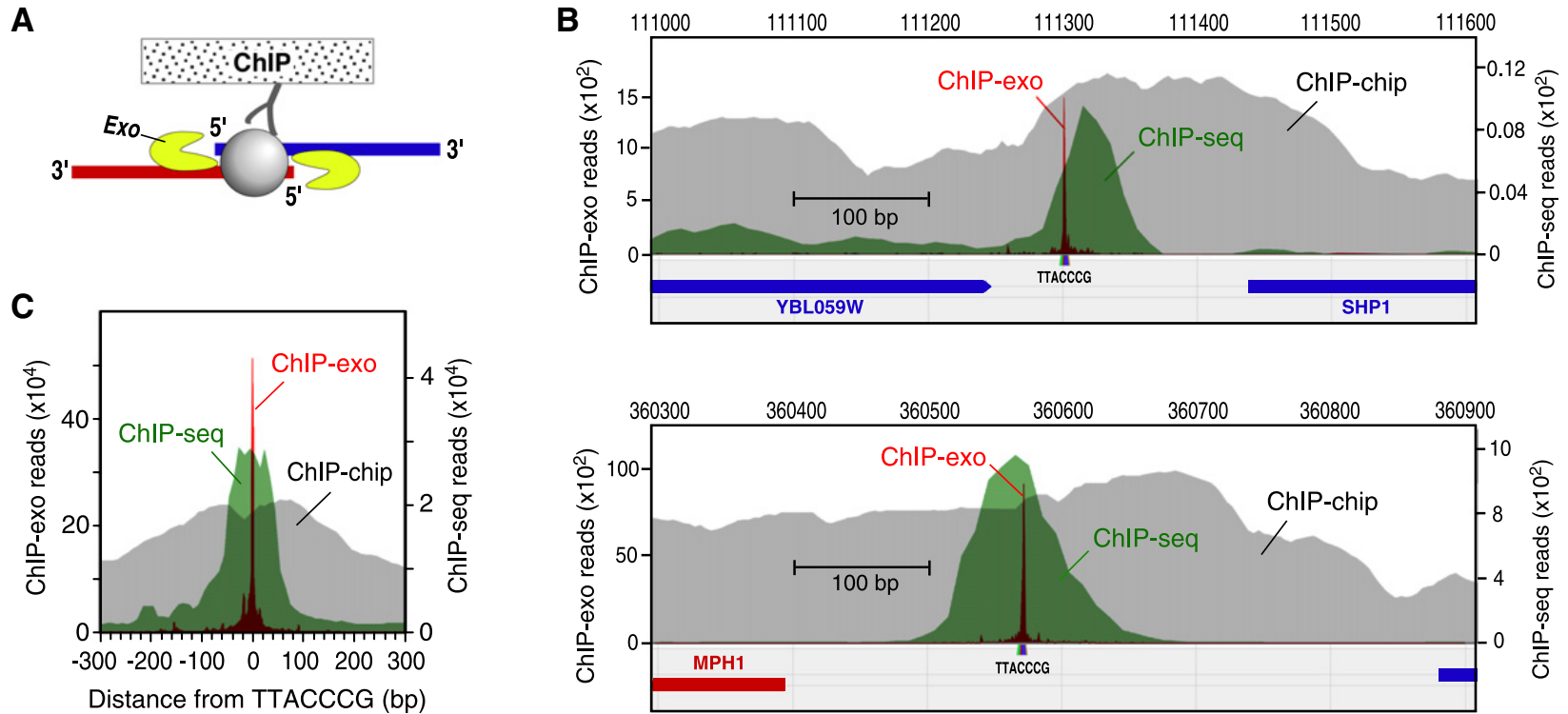
# CHiP-seq

1. CHiP = immunoprécipitation de chromatine : fixer les liaisons ADN-protéine (facteur de transcription ou histone)
2. fragmentation et filtrage (pour morceaux avec ADN+protéine)
3. séquençage de morceaux



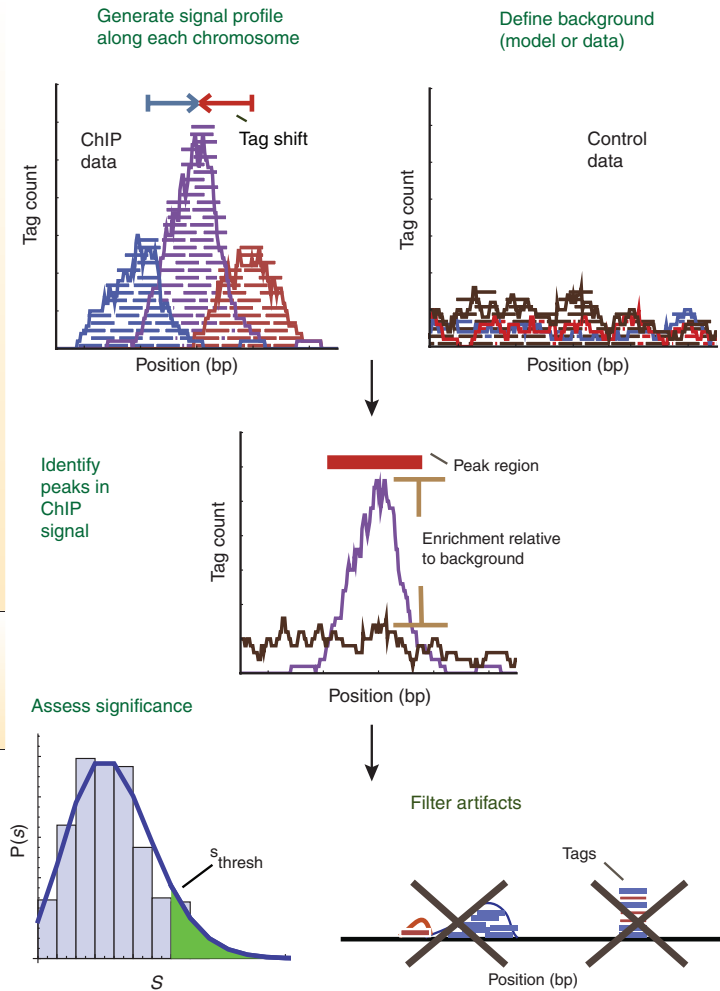
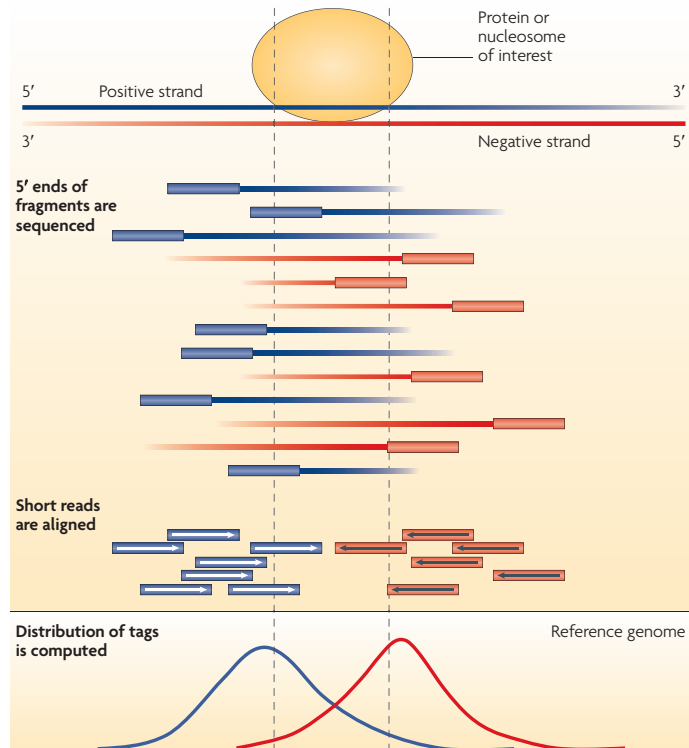
Park *Nat Rev Genet* 10 :669 (2009)

# ChIP-exo : résolution de 1pb



(exonucléase  $\lambda$  enlève le bout 5' précisément)

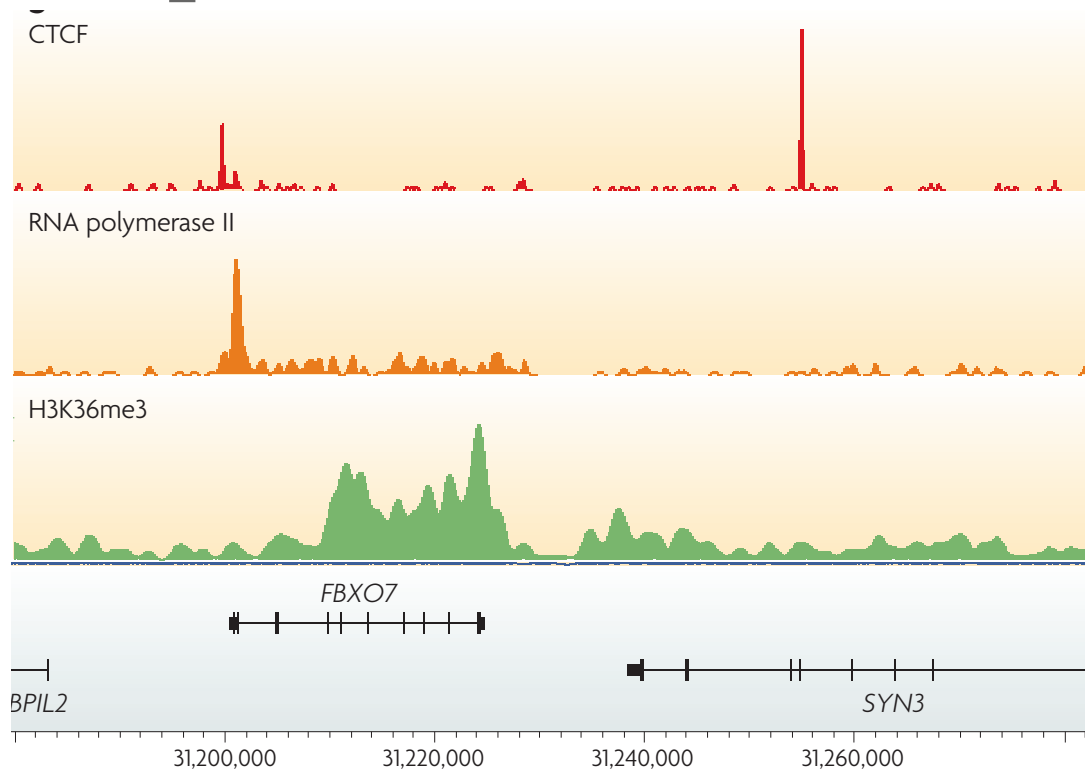
# Analyse de données



Park *Nat Rev Genet* 10 :669 (2009) ; Pepke & al *Nat Methods* 6 :S22 (2009)



# Pistes pour TFs et histones modifiées



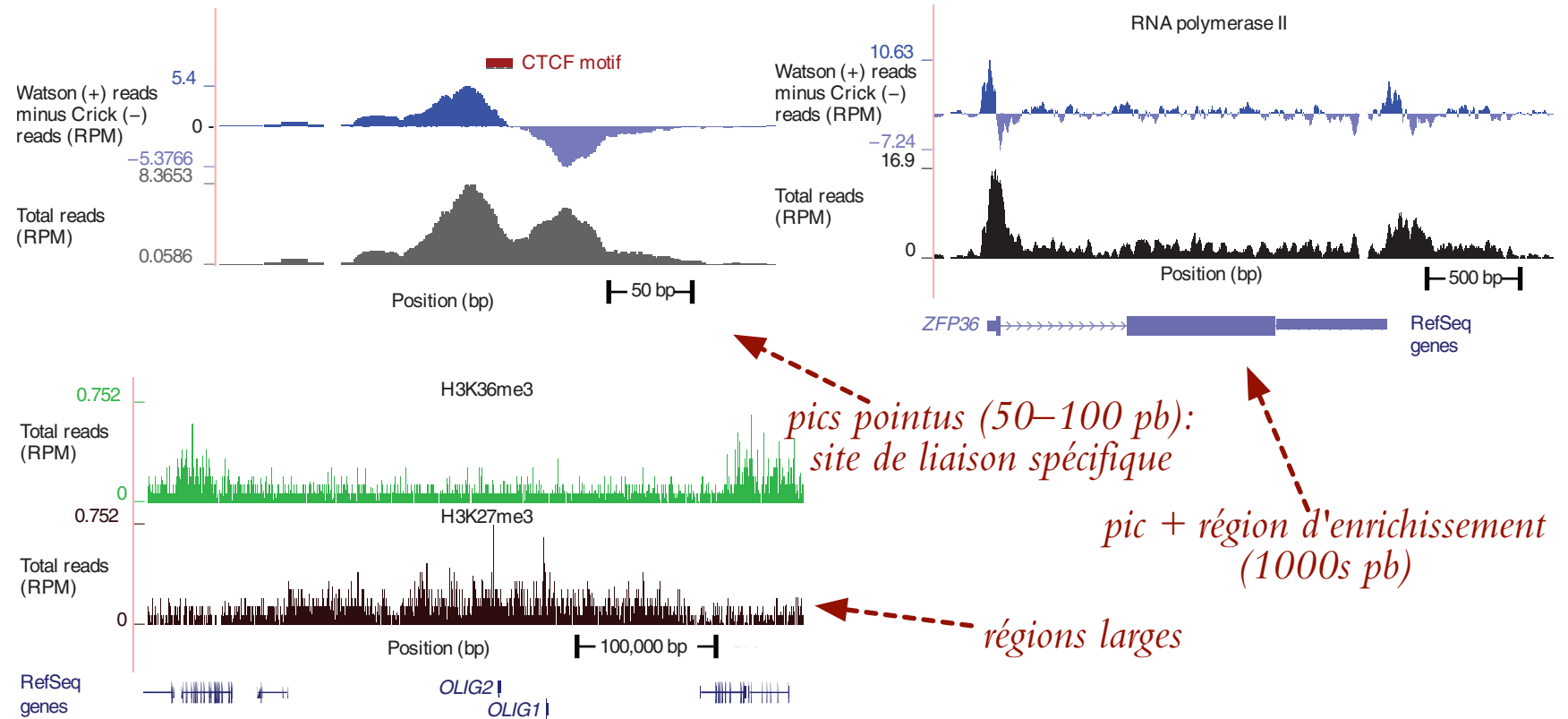
*CTCF (un élément isolant):  
pics pointus*

*polymérase d'ARN:  
pic pointu + région d'enrichissement*

*modification de histone H3  
(associé avec élongation de gène):  
pics élargis*

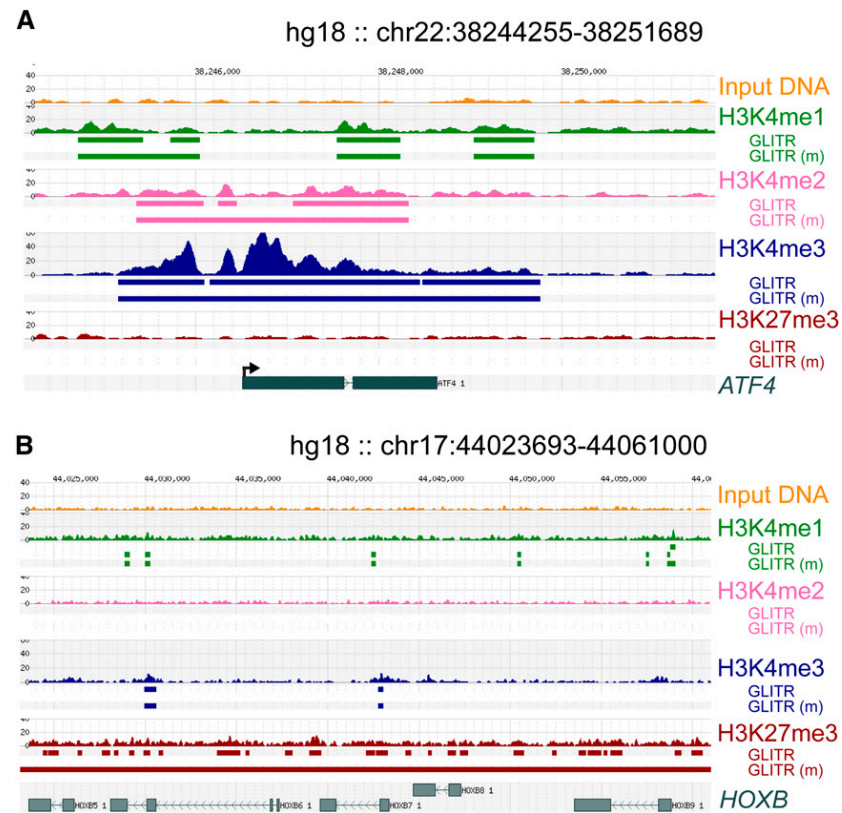
sites de liaison au début de la transcription ;  
modification de histone indique activité spécifique

# Pistes



Pepke & al Nat Methods 6 :S22 (2009)

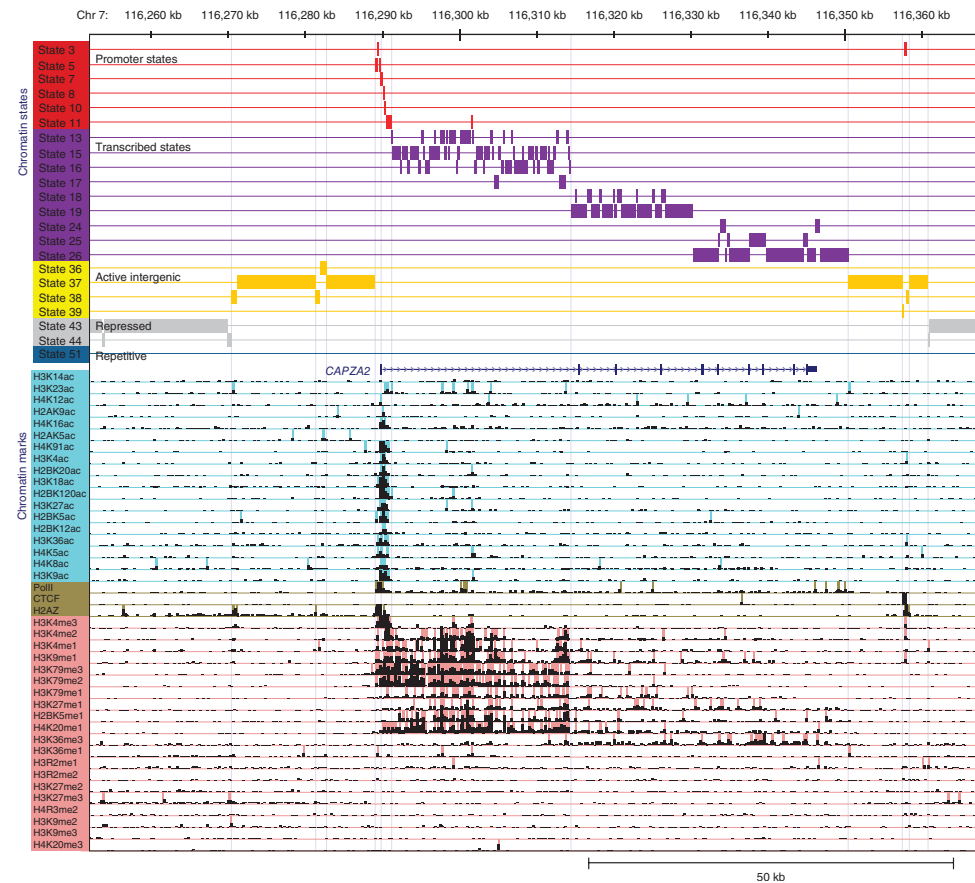
# Signaux d'activité et de silence



**Figure 3.** Marks of active and repressed genes in islets. (A) Human islet chromatin enriched for H3K4me1, H3K4me2, H3K4me3, H3K27me3, and control input DNA from four samples was processed and sequenced. Reads were pooled and aligned to the NCBI Genome Build 36.1-hg18, to determine regions that were enriched for binding by modified histones. Note the strong double peak surrounding the transcriptional start site (*left end*) of *ATF4*, and the weaker peaks for H3K4me1 and H3K4me2. In *B*, the *HOXB* cluster contains significantly enriched regions for H3K27me3 as well as H3K4me3 at several transcription start sites.

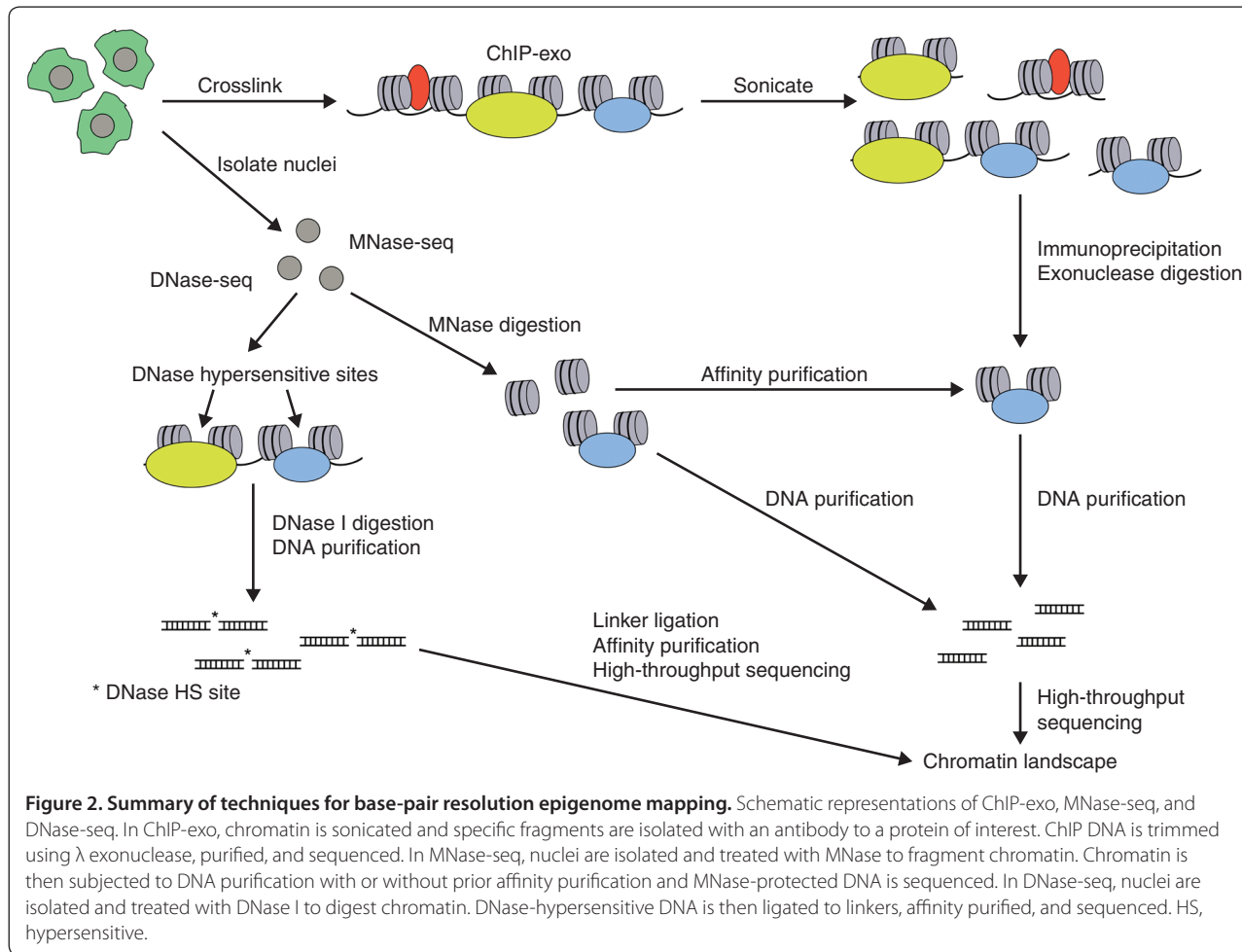
# HMM pour histones modifiées

émissions : modifications (vecteur binaire - Bernoulli indépendant) ;  
états : à créer comme nécessaire (évaluer complexité de modèle)



Ernst & Kellis *Nat Biotechnol* 28 :817 (2010)

# Méthodes d'épigénomique



**Figure 2. Summary of techniques for base-pair resolution epigenome mapping.** Schematic representations of ChIP-exo, MNase-seq, and DNase-seq. In ChIP-exo, chromatin is sonicated and specific fragments are isolated with an antibody to a protein of interest. ChIP DNA is trimmed using  $\lambda$  exonuclease, purified, and sequenced. In MNase-seq, nuclei are isolated and treated with MNase to fragment chromatin. Chromatin is then subjected to DNA purification with or without prior affinity purification and MNase-protected DNA is sequenced. In DNase-seq, nuclei are isolated and treated with DNase I to digest chromatin. DNase-hypersensitive DNA is then ligated to linkers, affinity purified, and sequenced. HS, hypersensitive.