Bioinformatic approaches to regulatory genomics and epigenomics

376-1347-00L - 2022 | week 05

Pierre-Luc Germain

ETH Zürich

Plan

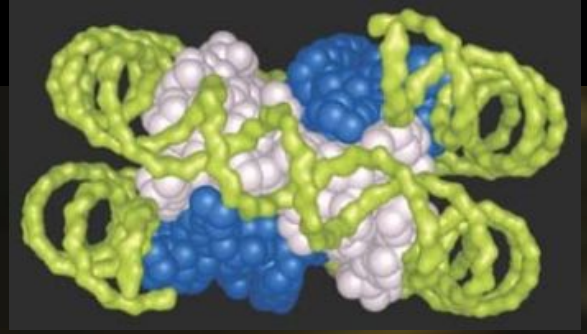
- The ‘histone code’ & functional elements

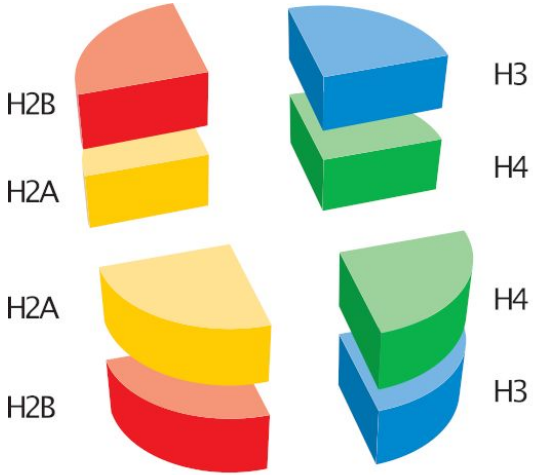
- Debriefing on the assignment
- Exploring overlaps

- p300 & enhancers

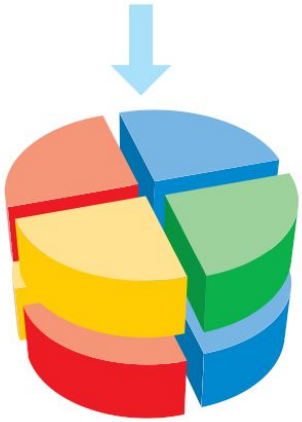
Nucleosome

(Krebs et al., 2018)



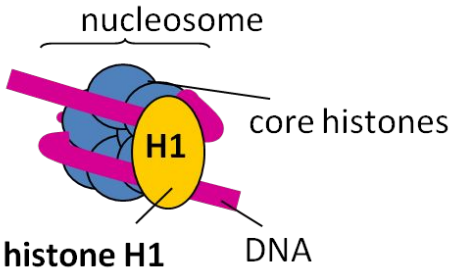


core histones



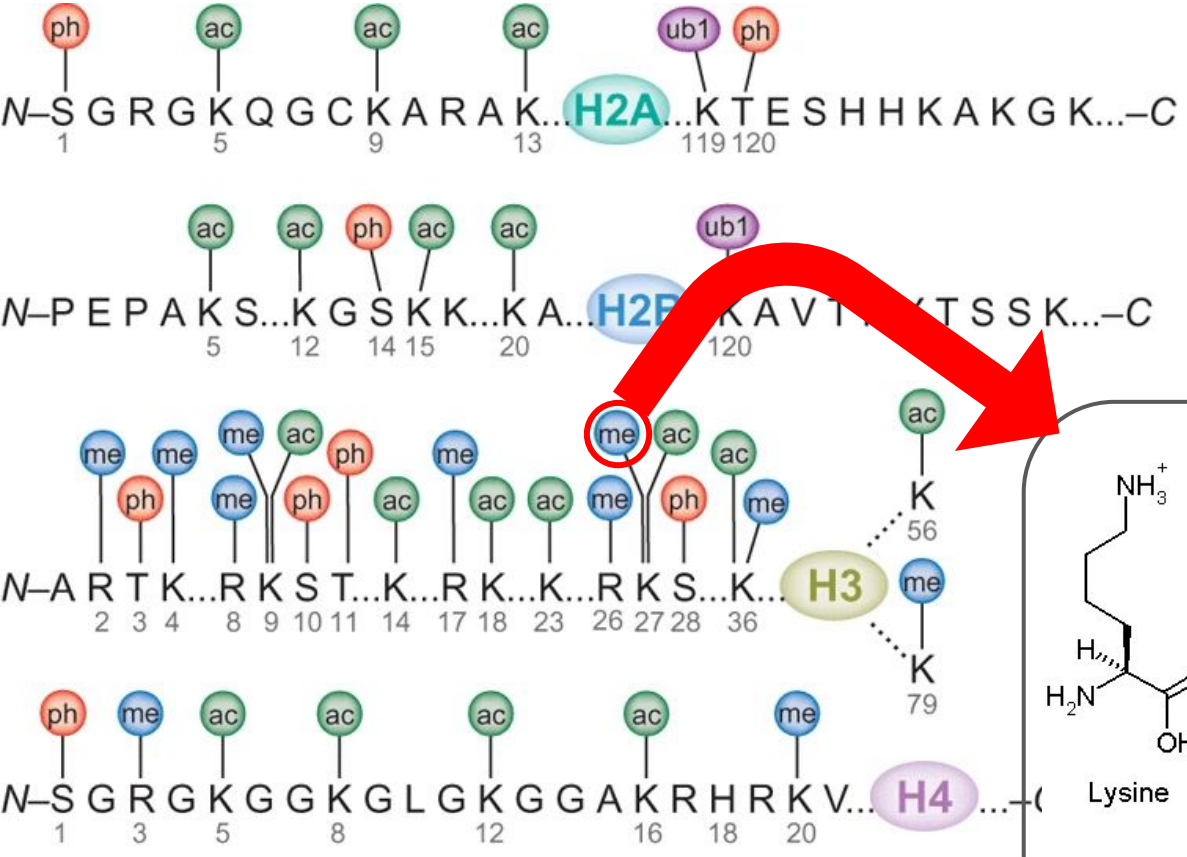
histone octamer

beads–on–a–string nucleosome array



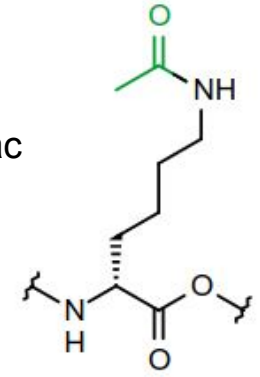
(Adapted from David O Morgan - The Cell Cycle. Principles of Control. Wikimedia Commons)

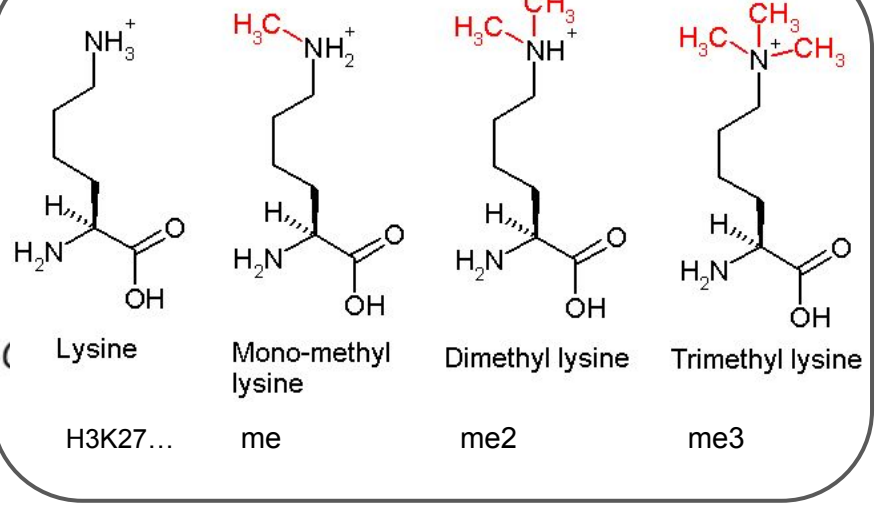
Many residues on the histone tails can be post-translationally modified



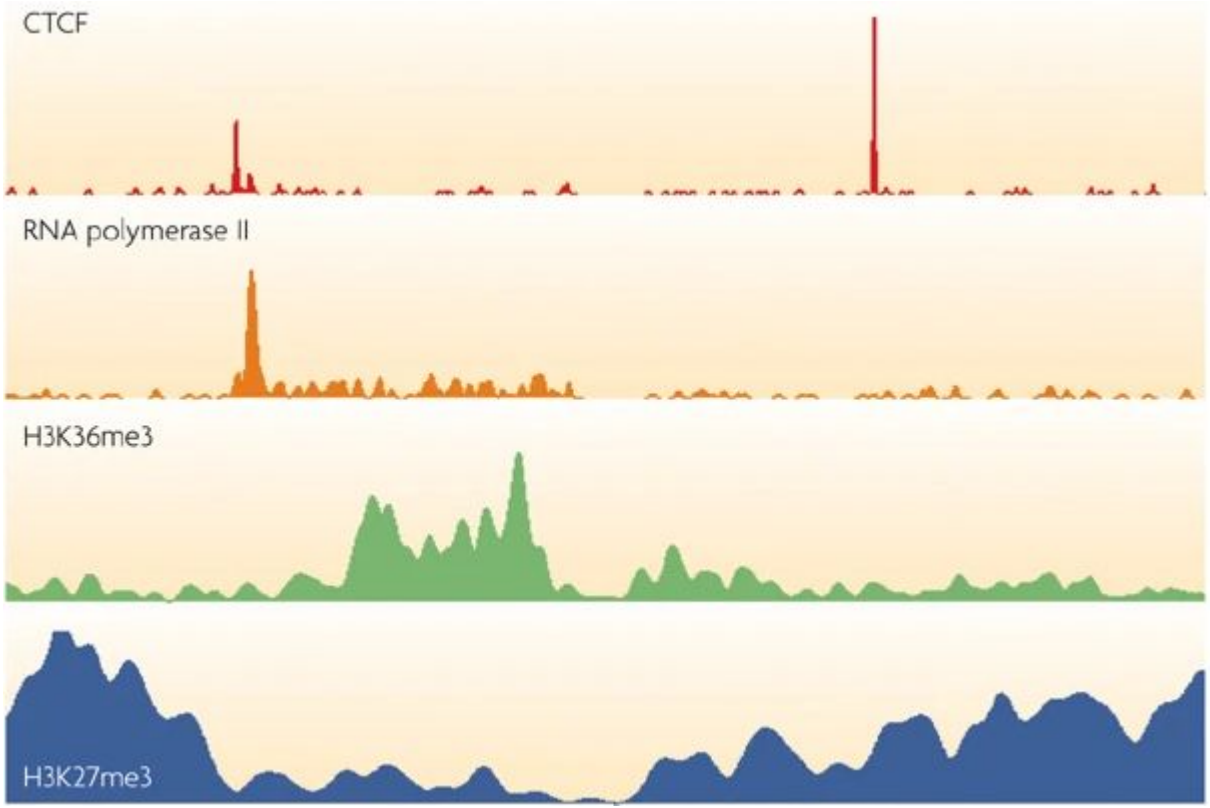
(Bhaumik, Smith and Shilatifard 2007)

H3K27ac





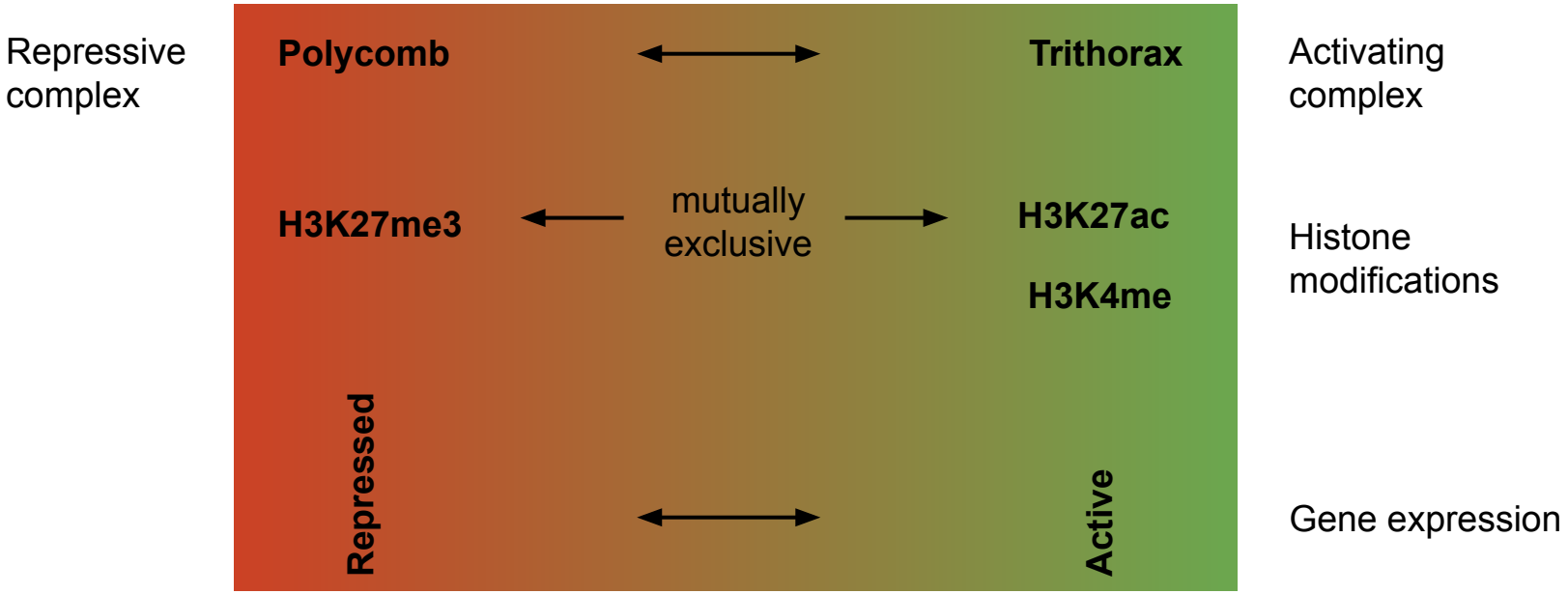
Some histone modifications appear to be very localized, e.g. happening on a specific nucleosome, while most are much more broadly distributed



The strategy of calling ‘peaks’ must therefore be adapted
(e.g. “broad” option of most peak-callers)

(Park, Nat Rev Gen 2009)

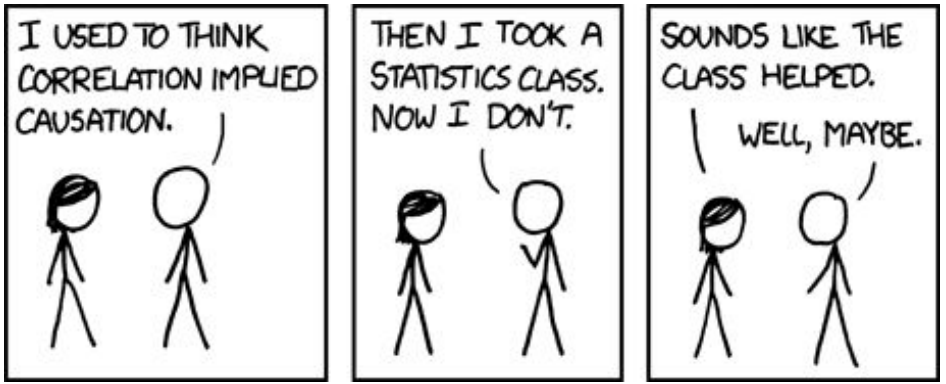
There is a very strong association of certain histone marks and activation or repression



But which comes first?

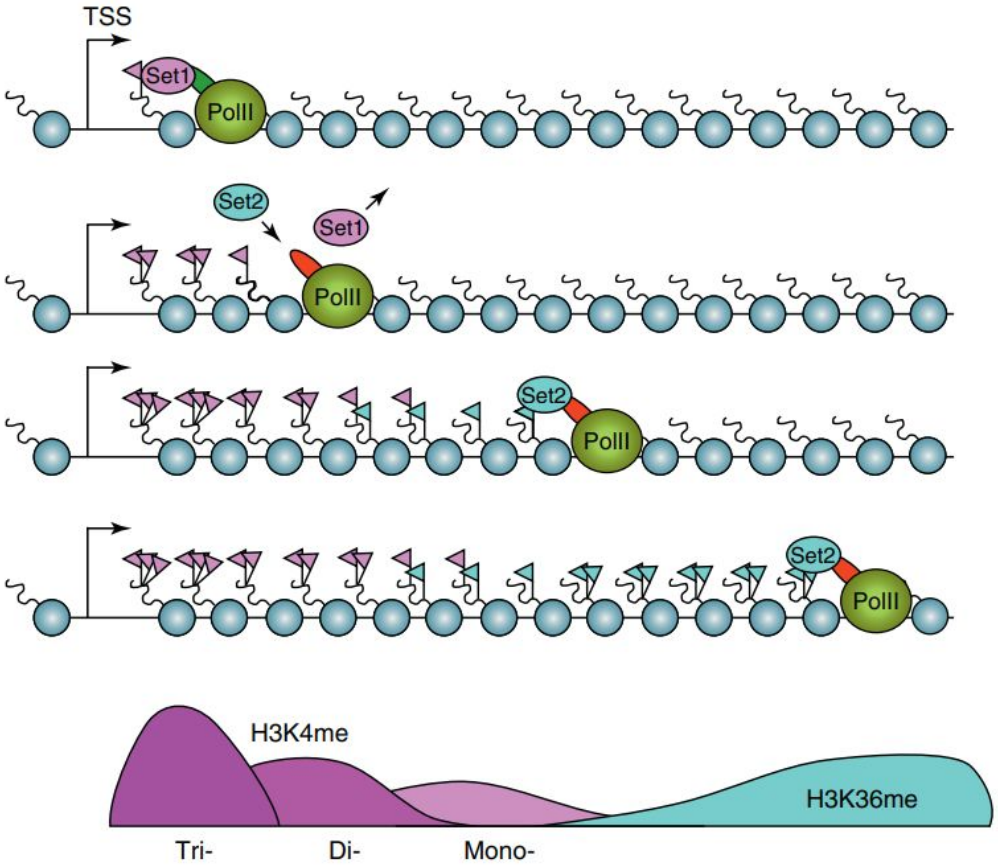
Causality or correlation?

Are histone modifications **responsible** for activation/repression, or are they merely associated **side-effects**?



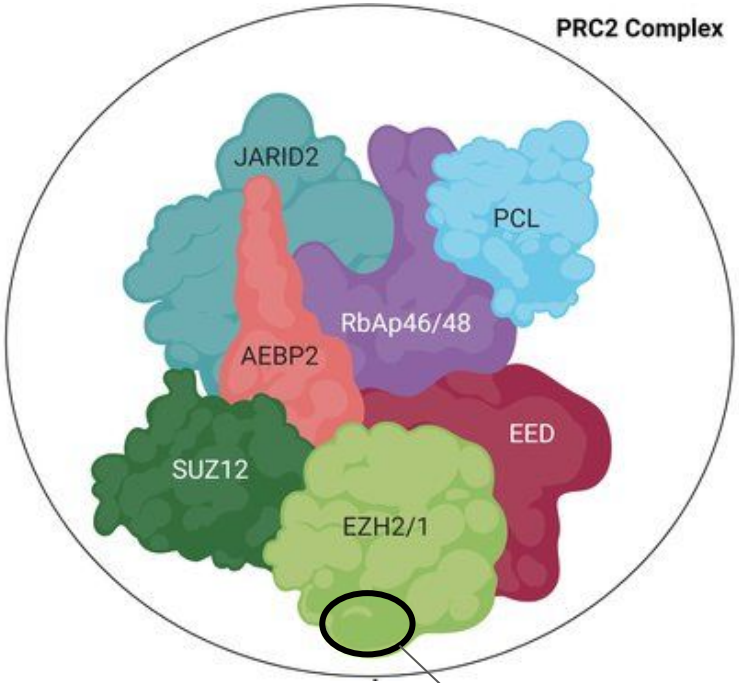
(https://xkcd.com/552)

Transcription-mediated histone modification



(Henikoff and Shilatifard 2011)

The example of H3K27me3, chiefly deposited by the polycomb repressive complex (PRC2)



EHZ2's SET domain catalyzes the addition of a 3rd methyl group to H3K27, i.e. H3K27me2 → H3K27me3

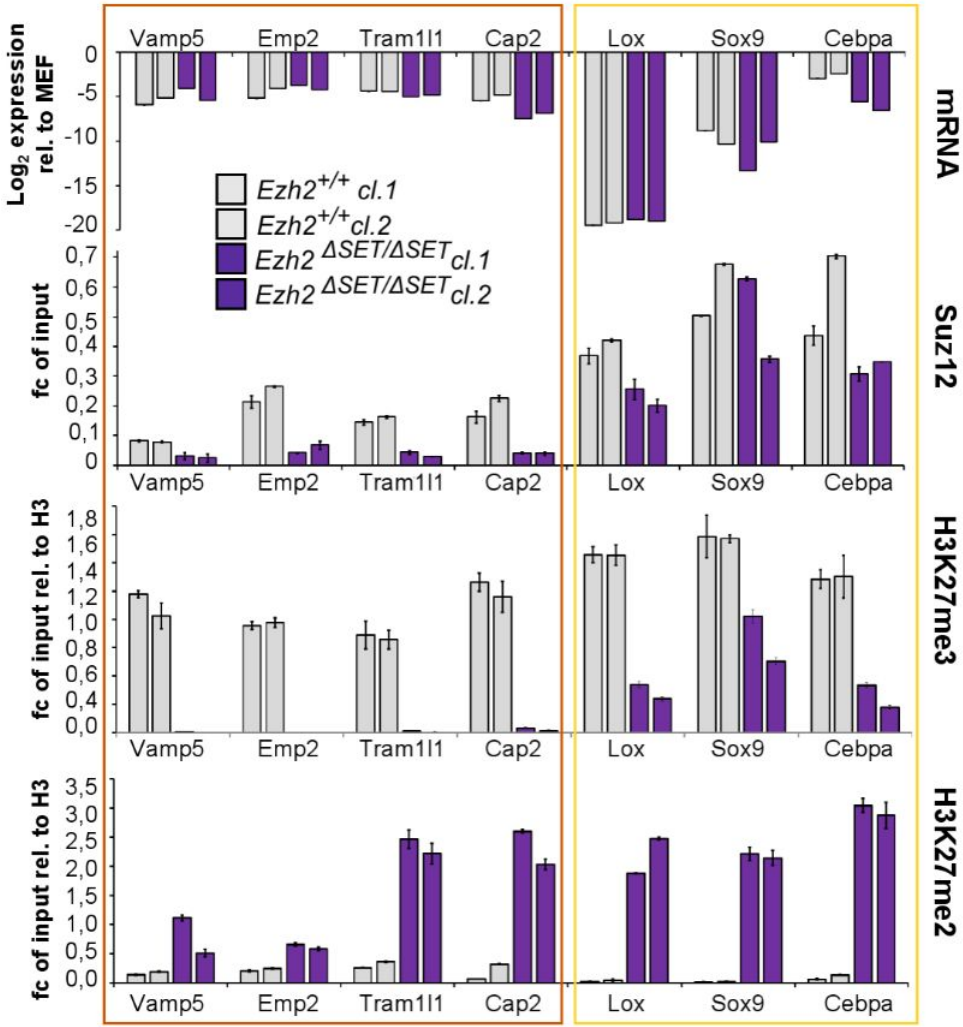
(Adapted from Zoroddu et al., 2021)

Abolishing the enzymatic activity of *Ezh2*, the gene responsible for depositing H3K27me3, abolishes (most of) the mark but does not prevent the repression of the target genes, nor cellular reprogramming

(Fragola et al., PLoS Genetics 2013)

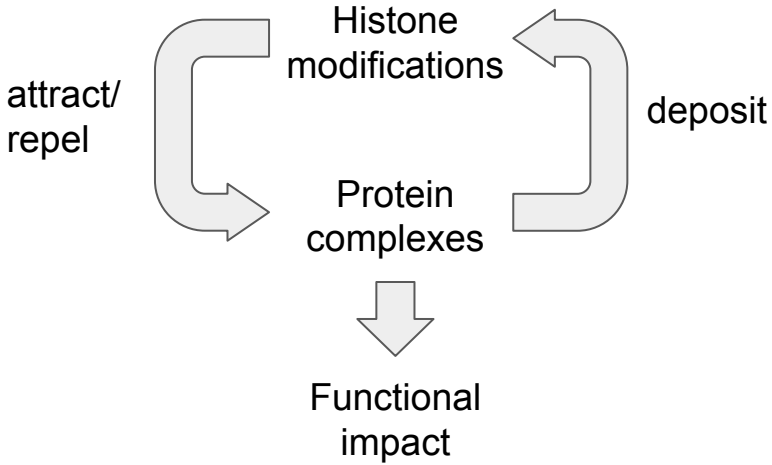
Similarly, the loss of H3K4me3 appears to have no effect on nascent transcription

(Murray et al., bioRxiv 2019)



Causality or correlation?

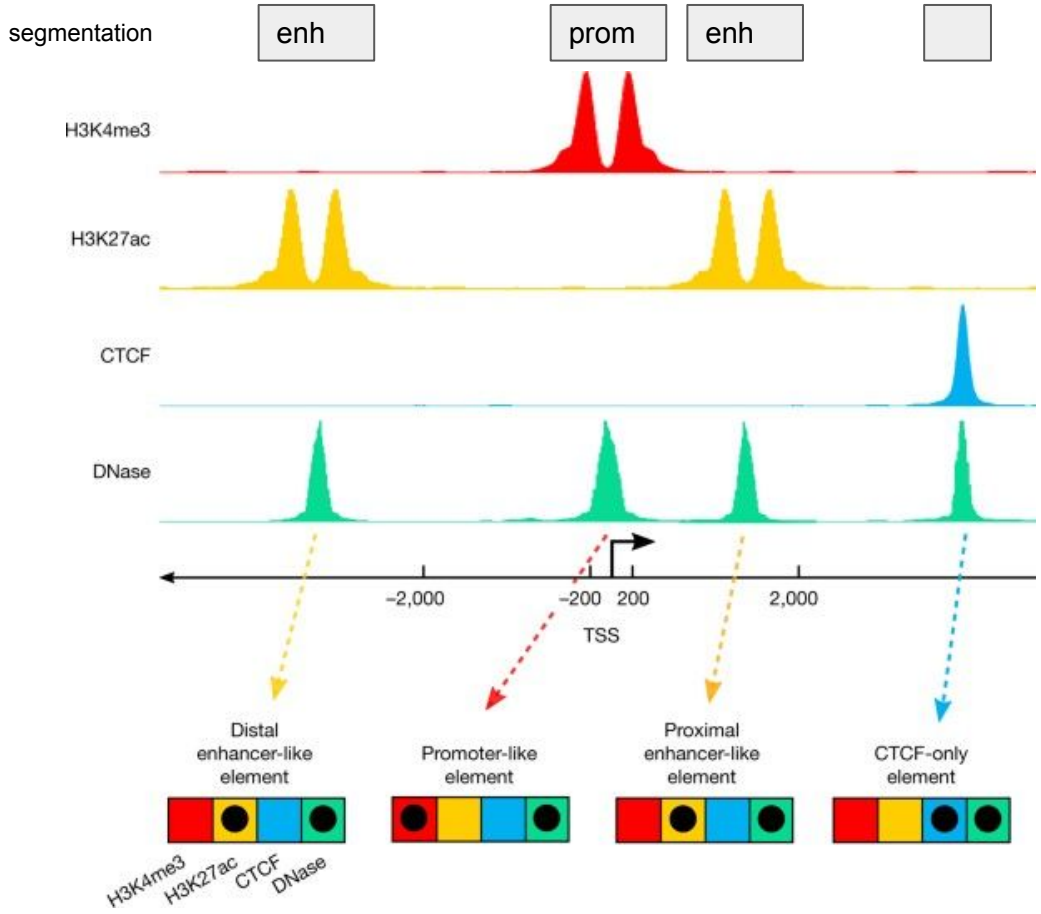
The jury is still out, but the current most likely view is somewhere in the middle



Whether they're causative or not, they can serve as **proxies** for function.

This means that profiling a few histone modifications gives an overview of the epigenomic landscape of a cellular state which would otherwise require profiling all the potentially-relevant factors/complexes

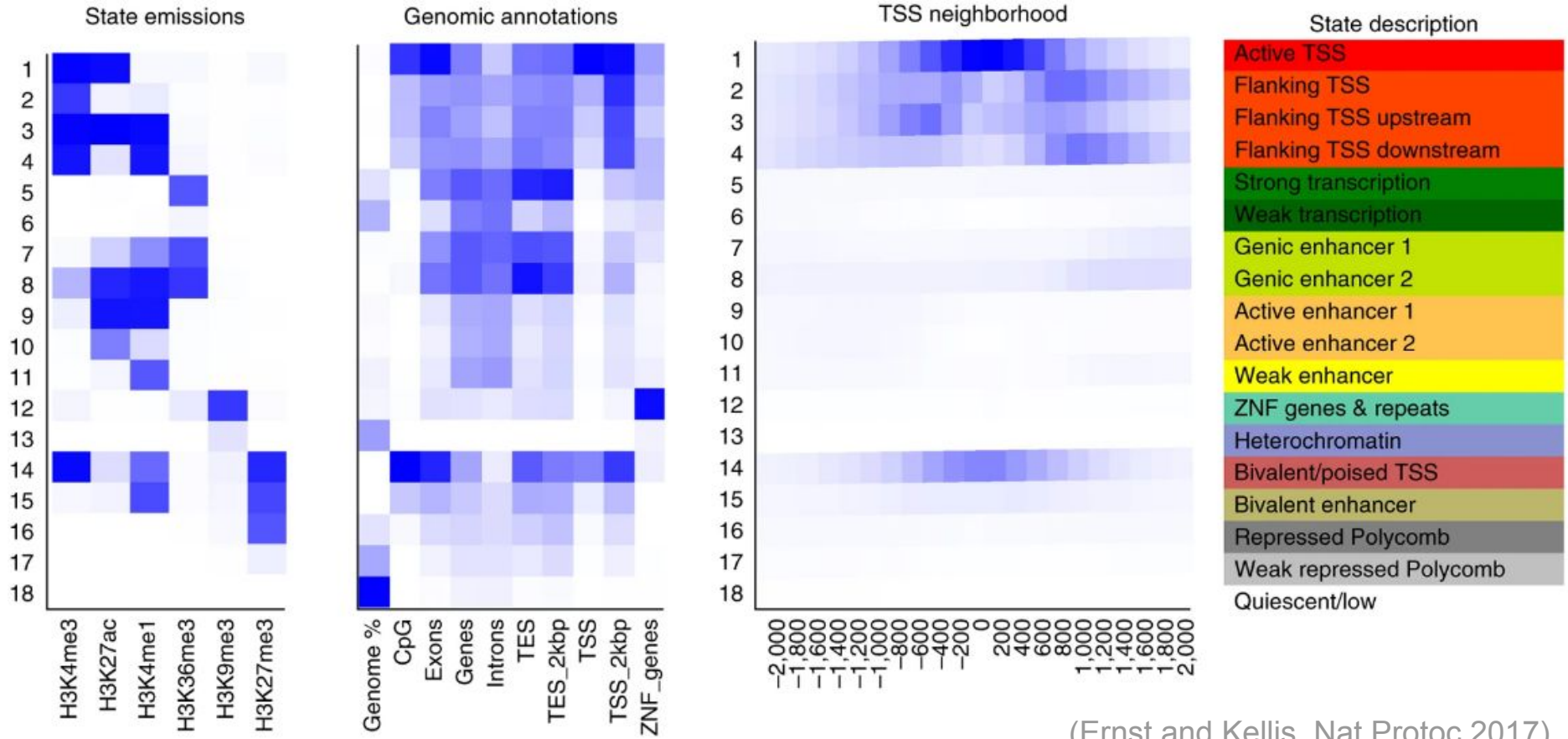
A signature-based encyclopedia of DNA elements



ENCODE's "signature strategy":

- Different types of functional genetic elements are associated with different chemical signatures
- We can identify functional elements by identifying these signatures genome-wide

So how many kinds of functional elements/states are there?



(Ernst and Kellis, Nat Protoc 2017)

Some stuff is pretty clear:

- **Transcription start site (TSS):**
  - **H3K4me3** is almost always associated with a
  - Active TSS are marked by **H3K27ac**
  - So-called “poised” (or bivalent) TSS are instead marked by both **H3K4me3** and **H3K27me3**

- **Enhancers:**
  - Most enhancers have **H3K4me1**
  - Active enhancers are marked by **H3K27ac**
  - So-called “poised” (or bivalent) enhancers are marked by **H3K4me1** and **H3K27me3**

- Repressed elements are marked by **H3K27me3**

- Heterochromatin is marked by **H3K9me3**
- Insulators: CTCF+cohesin

Assignment debriefing & practical

Inconsistent seqlevels

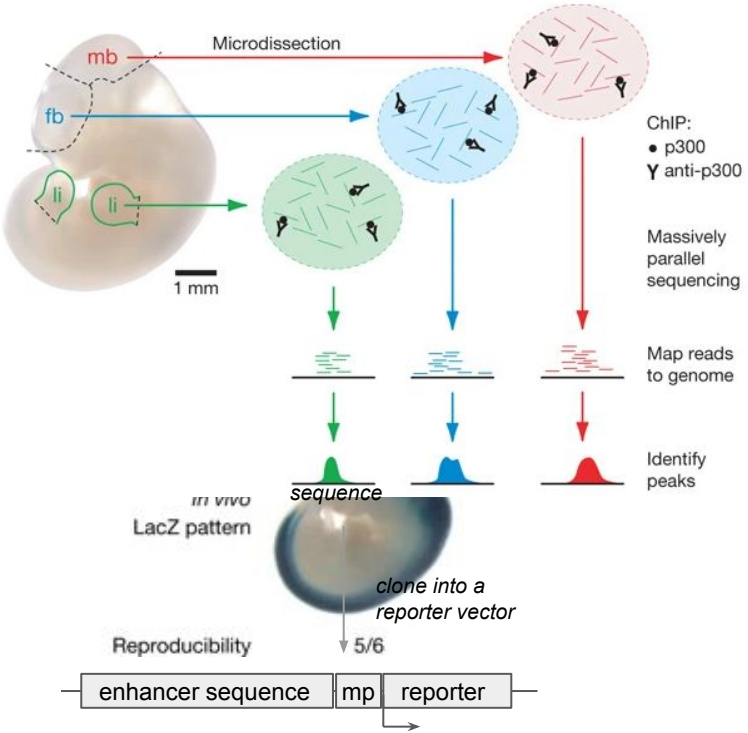
```
overlap_H3K27ac <- overlapsAny(peaks_p300, peaks_H3K27ac)
```

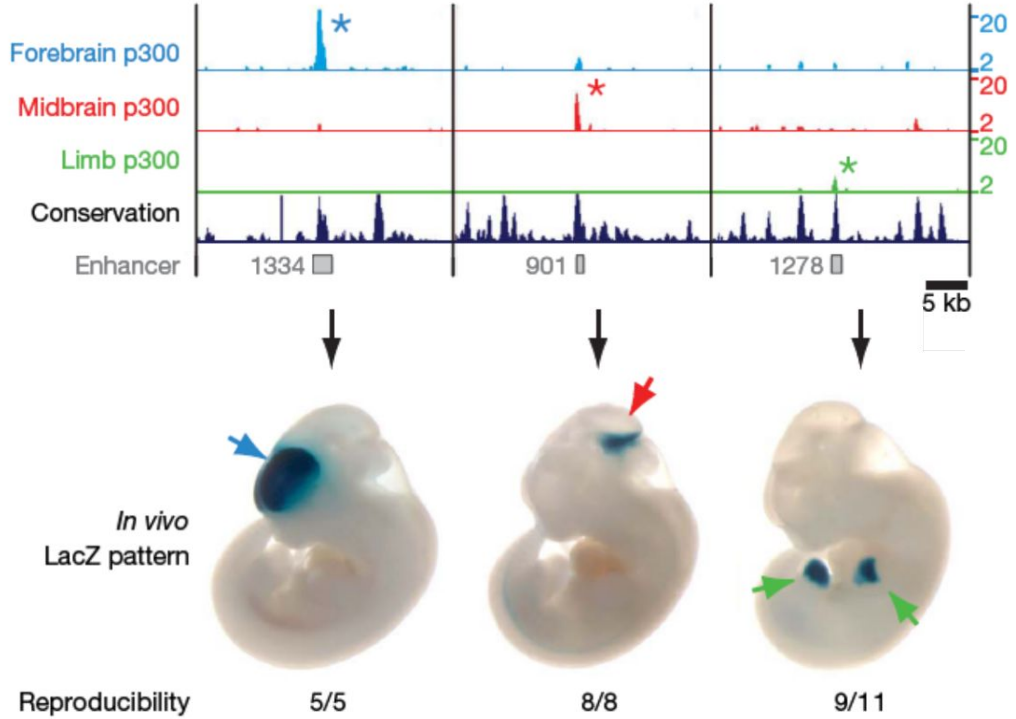
```
## Warning in .Seqinfo.mergexy(x, y): The 2 combined objects have no sequence levels in common. (Use
##   suppressWarnings() to suppress this warning.)
```

```
table(overlap_H3K27ac)
```

```
## overlap_H3K27ac
## FALSE
##  6394
```

p300 and validation of enhancer activity





(Adapted from Visel et al., 2009)

vista enhancers: http://enhancer.lbl.gov

Assignment

- Using the peaks you downloaded last week, identify bivalent domains (H3K27me3 + H3K4me3) in mouse embryonic stem cells (mESC)

- What happens to those regions upon differentiation?
  - Choose a differentiated cell type (e.g. hepatocytes, neural progenitor, or smooth muscle cells)
  - Download the H3K27me3 and H3K4me3 peaks from this cell type
  - How many of the mESC bivalent domains are, in this differentiated cell type, overlapping either mark or their combination?