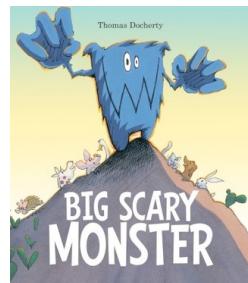
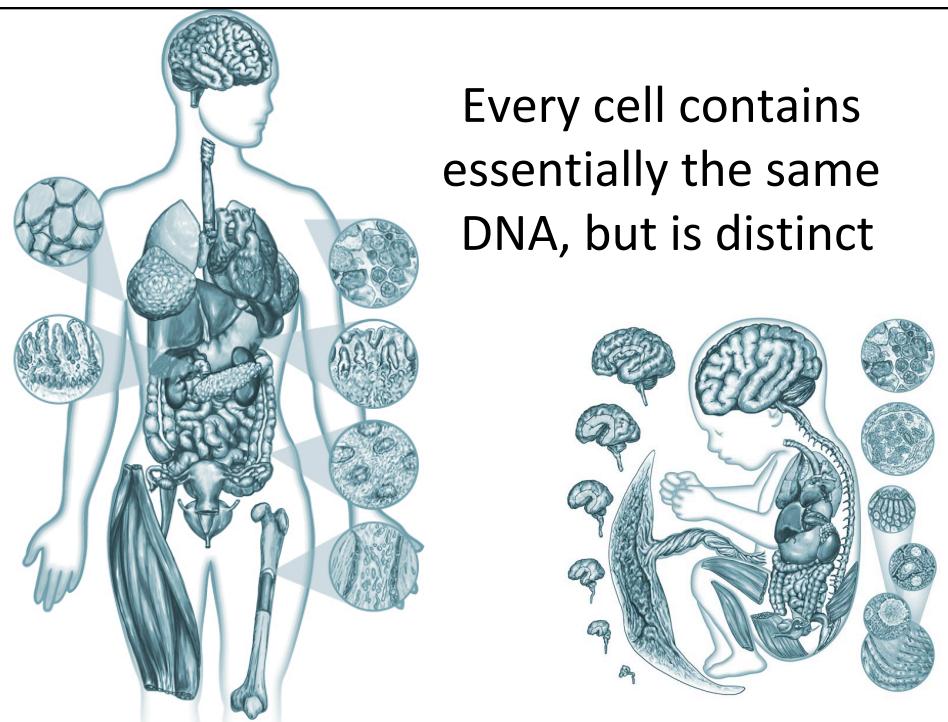


“One sees great things from the valley; only small things from the peak.”

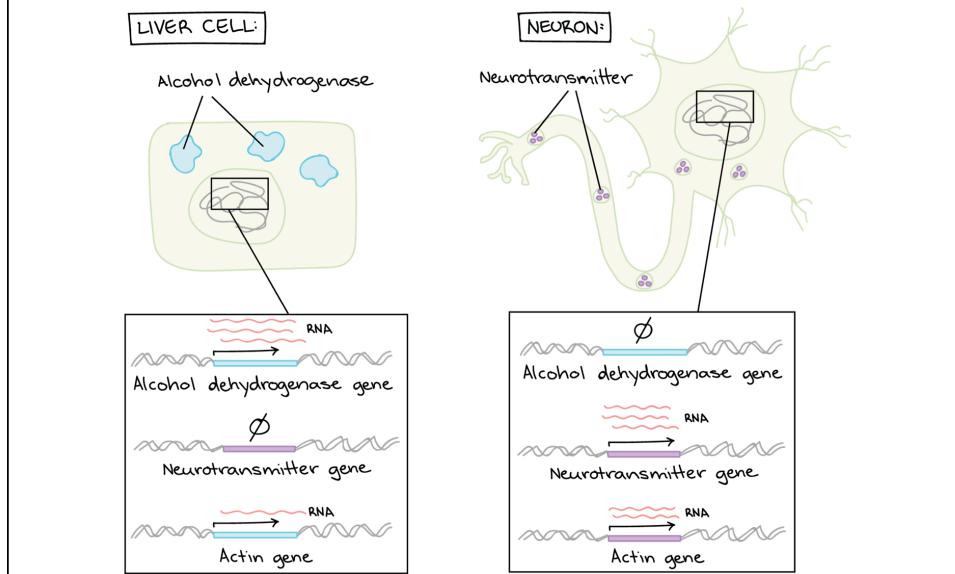
---Gilbert K. Chesterton



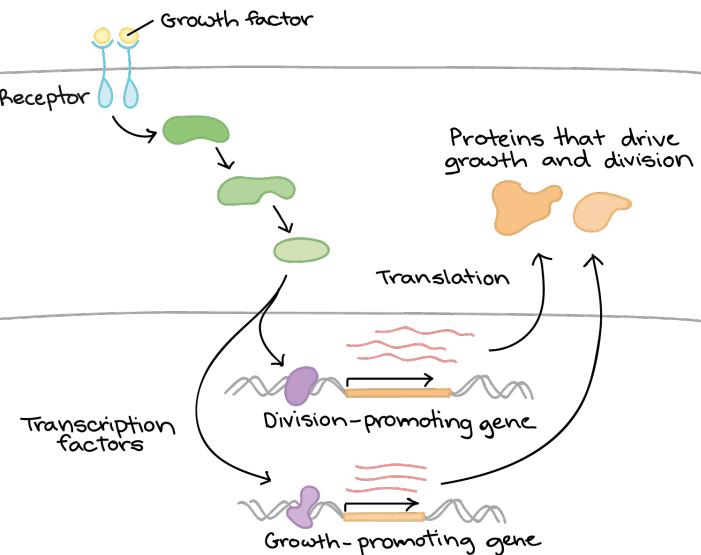
Every cell contains
essentially the same
DNA, but is distinct



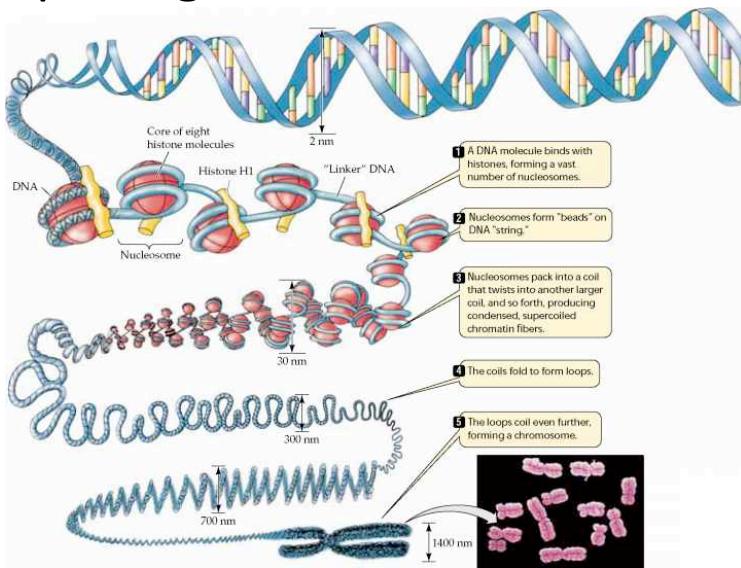
The difference between cell types is distinct patterns of expression.



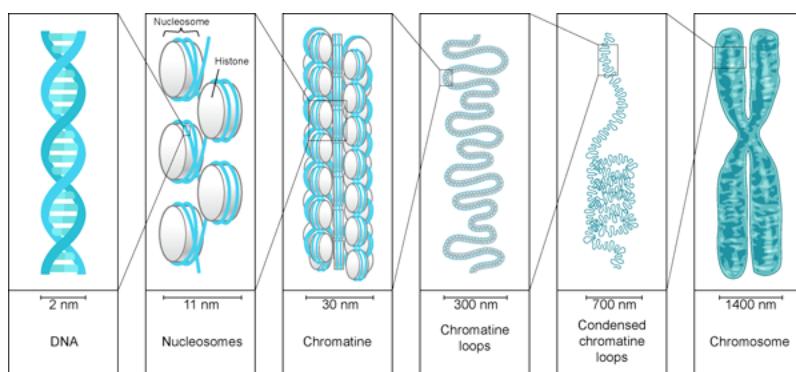
Transcription is also at the heart of how cells respond to environment changes.



Inside the cell, DNA is densely packaged into chromatin.

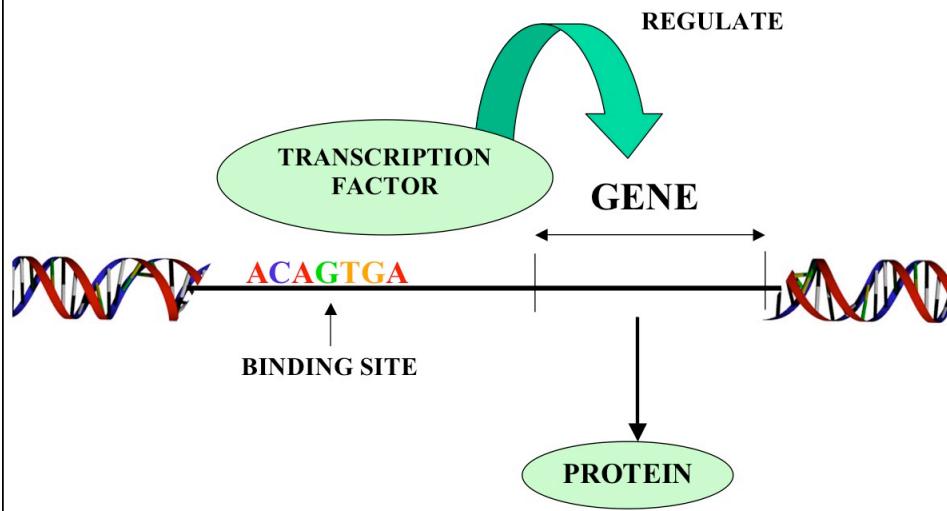


Nucleosomes allow all that DNA to be stuffed into the nucleus.

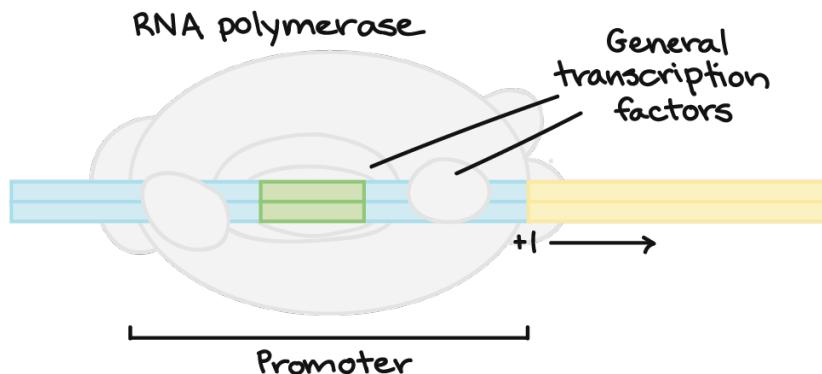


We will talk about studies into nucleosomes next week.

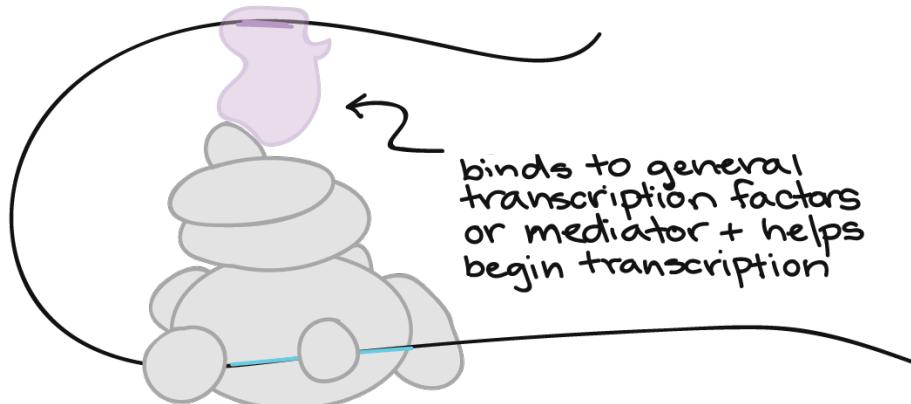
Transcription factors are DNA binding proteins that alter transcription.



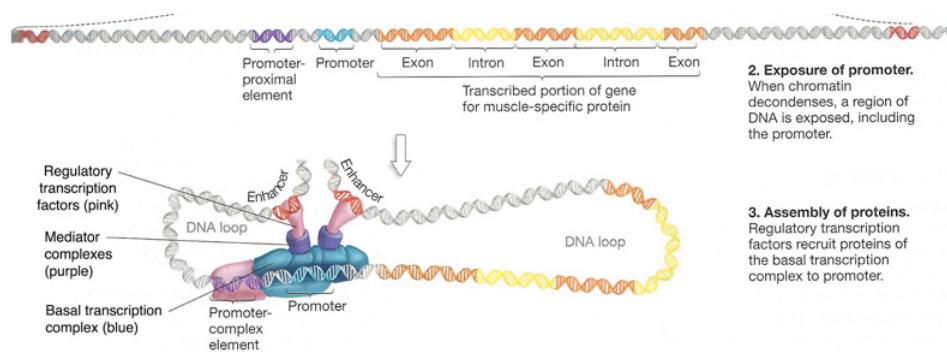
RNA polymerase can attach to the promoter only with the help of proteins called **basal (general) transcription factors**.



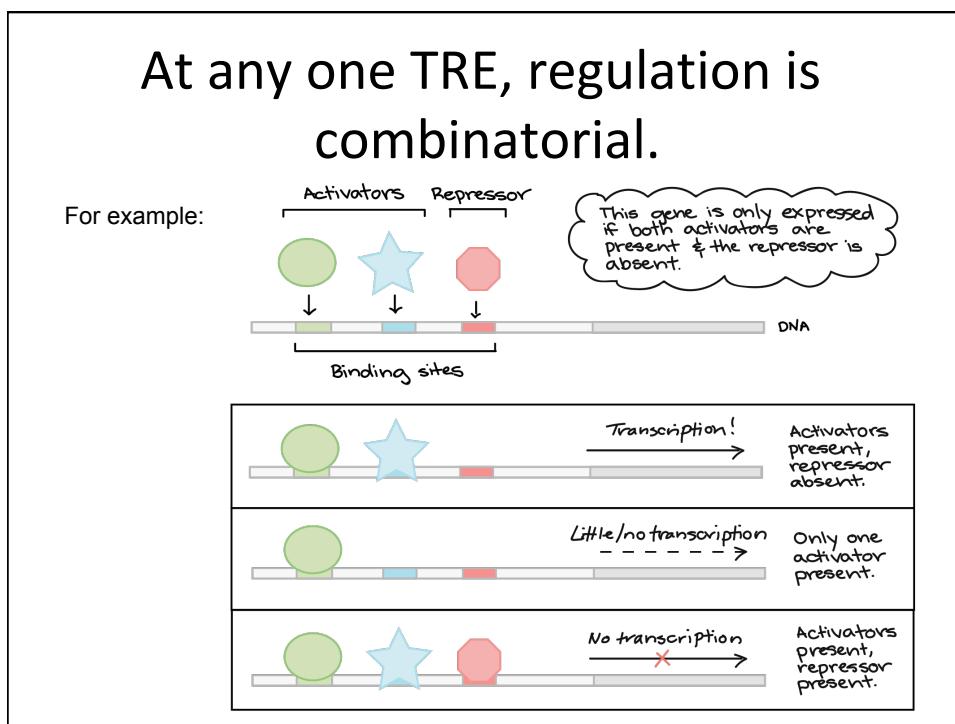
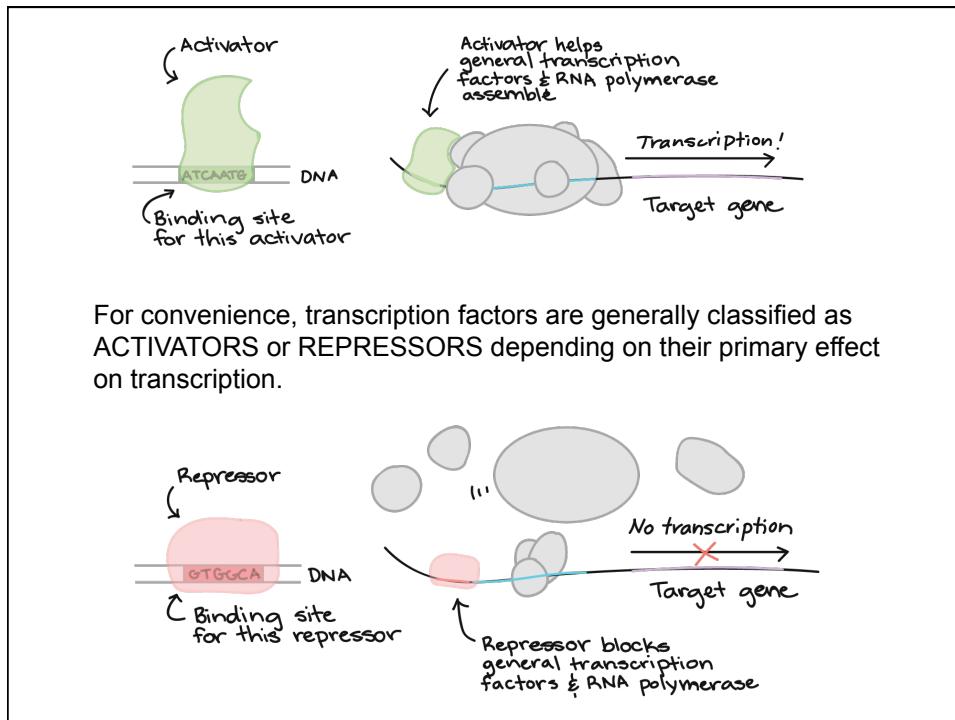
Specific (regulatory) transcription factors control the expression of specific, individual genes.



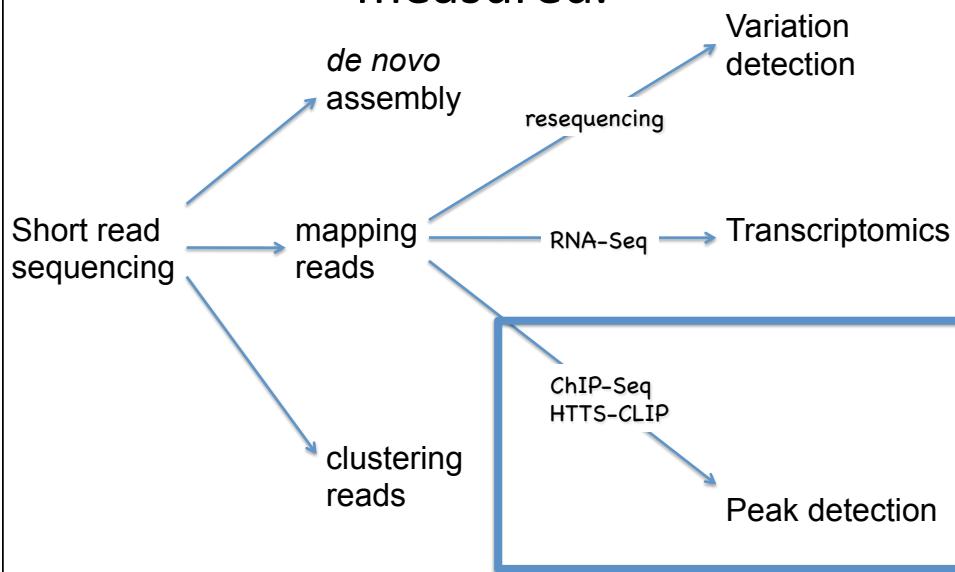
These DNA binding proteins bind to transcriptional regulatory elements (TREs)



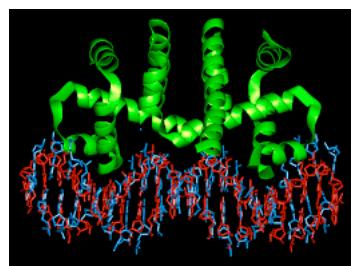
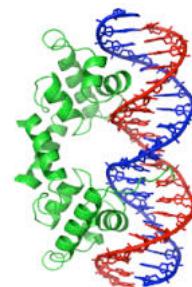
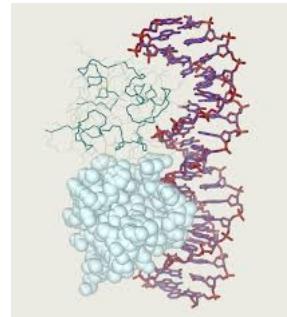
- Promoters are immediately upstream of gene (proximal)
- Enhancers are at some distance from the start of the gene (distal)



Protein-DNA interactions can be measured.



Many distinct protein to DNA binding confirmations.



Importantly: TFs have two functions

1. They bind to DNA. i.e., a protein binds to a particular fragment (sequence) of the DNA.

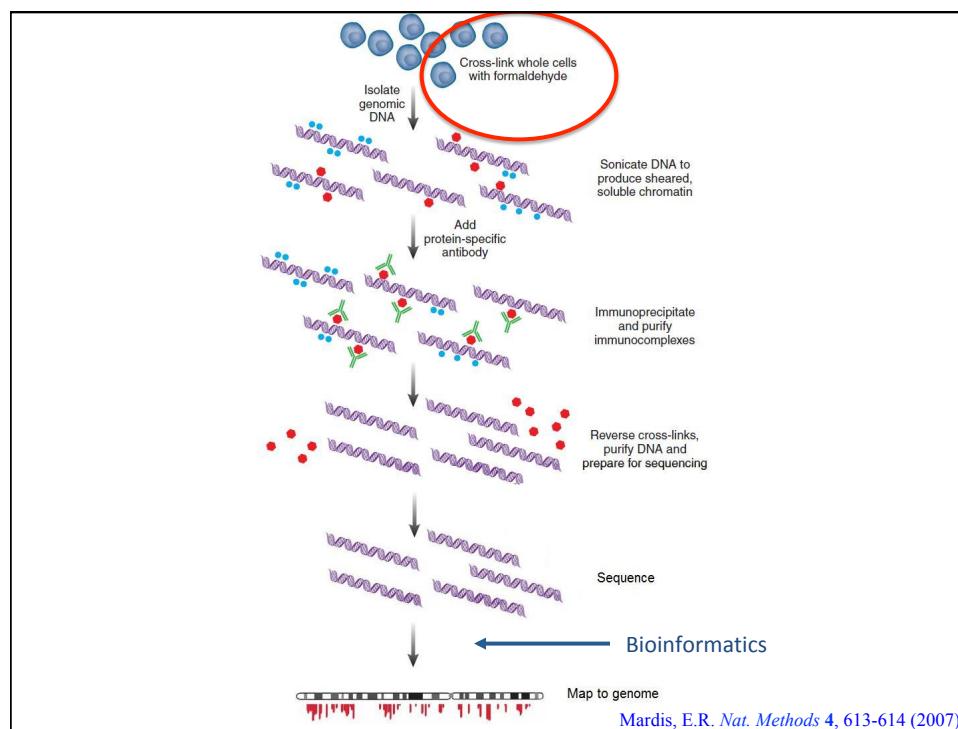
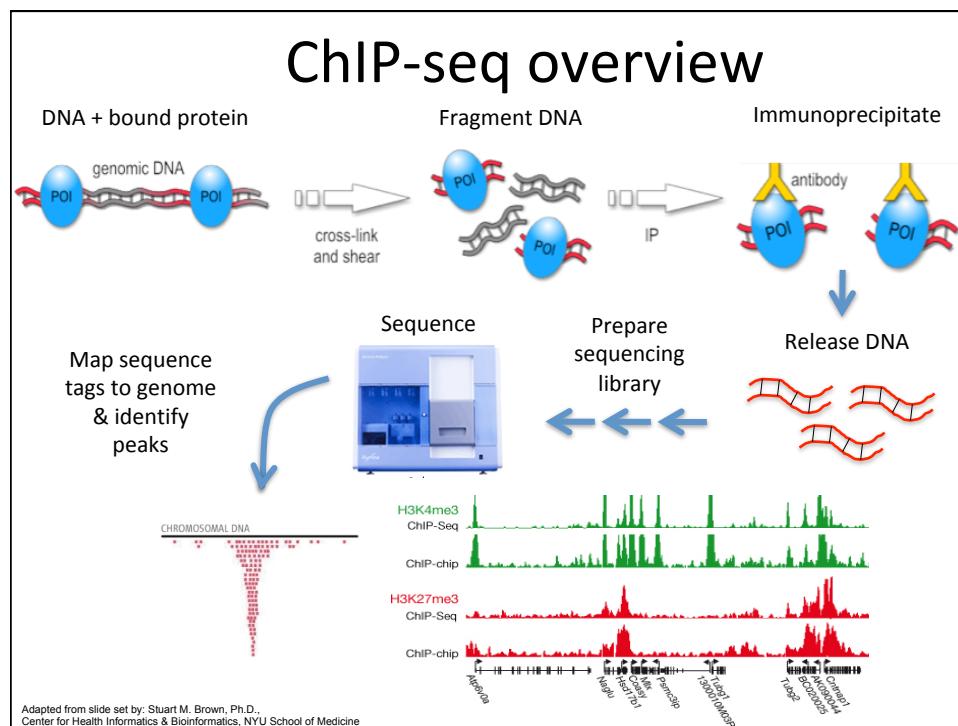
TODAY we'll talk about BINDING.
FRIDAY we'll talk about sequence motifs.

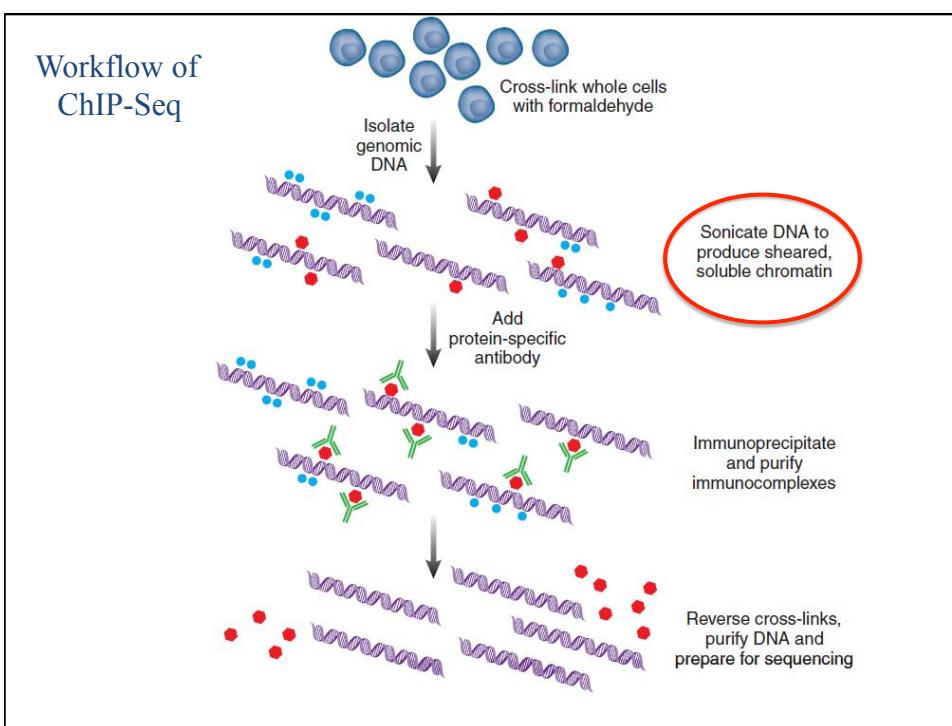
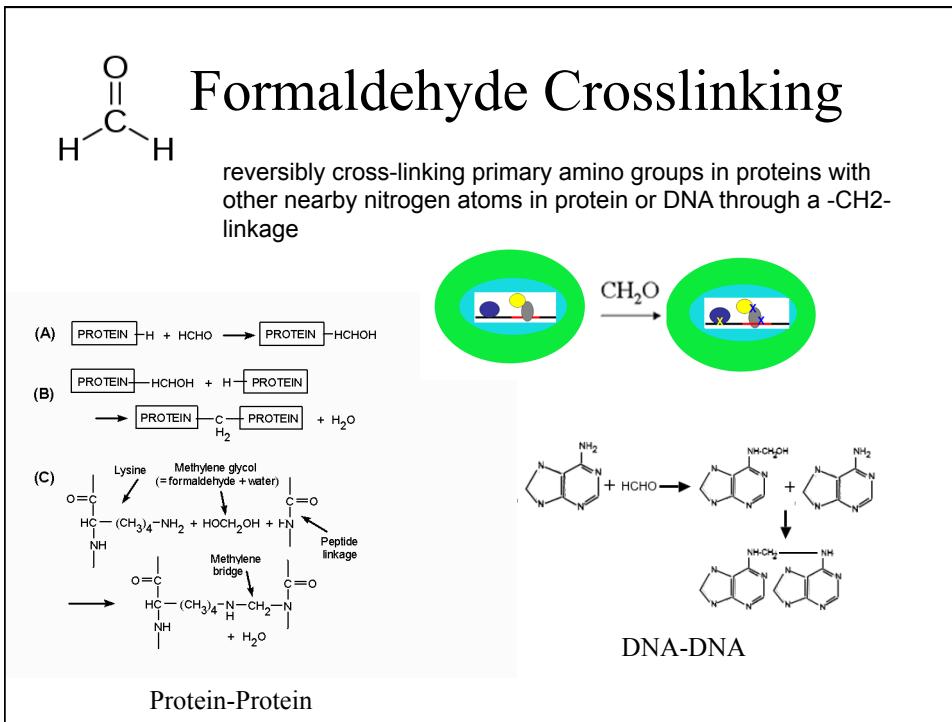
2. They modify transcription nearby.

Monday we'll talk about TF interactions with polymerase.

Chromatin Immunoprecipitation (ChIP)

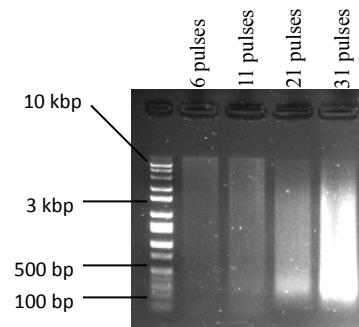
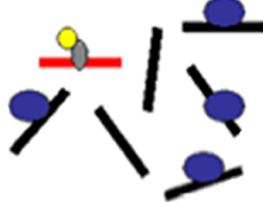
- ChIP is a method to investigate protein-DNA interaction (binding) *in vivo*.
- The output of ChIP is enriched fragments of DNA that were bound by a particular protein.
- The identity of DNA sequence motif (what the protein recognizes) need to be *inferred* from binding data.



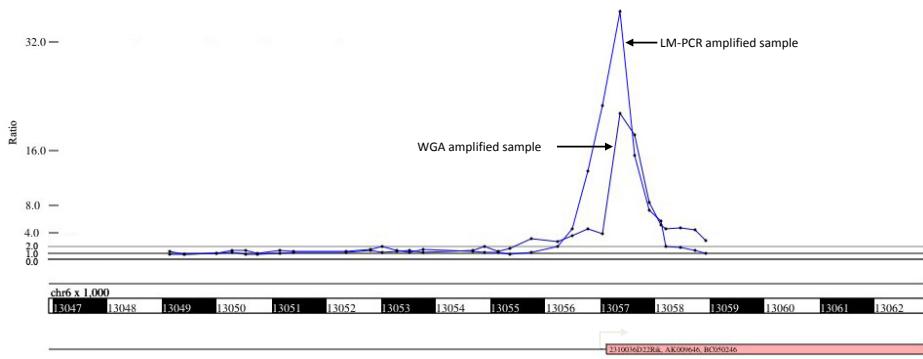


Sonication Check

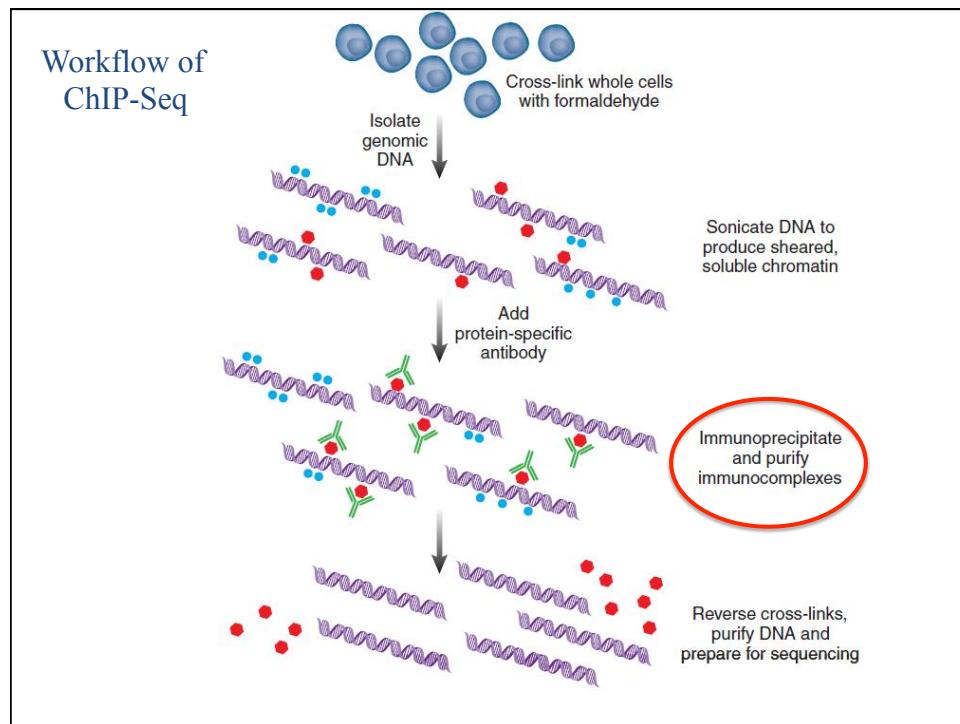
Sonication results in many different sized fragments and should be verified for your experiment



Test Experiment:
H3K4me3 ChIP in Olig2-stage cells shows peaks directly over promoters



We find peaks over promoters as expected, but peaks resulting from each amplification technique differ in height and location



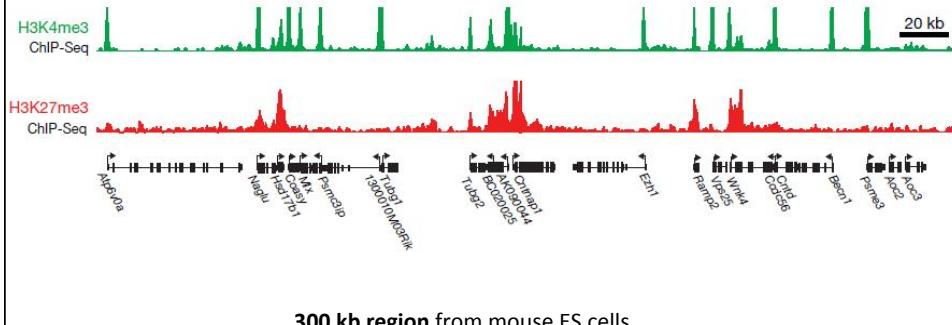
Antibody Validation

Step 1: Show reactivity on a Western (denatured epitope).

Step 2: Show IP ability. Perform IP, run Western and re-probe with same Ab (IP-Western)

Step 3: Rough quantification, true band should be >50% of combined signal from other bands (ENCODE guidelines)

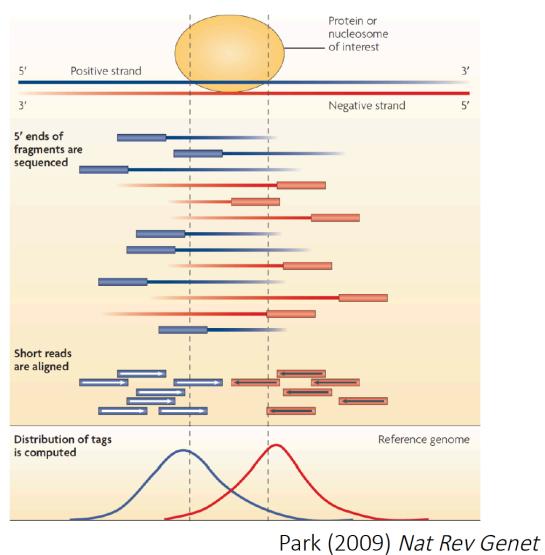
Reads are then mapped to genome.



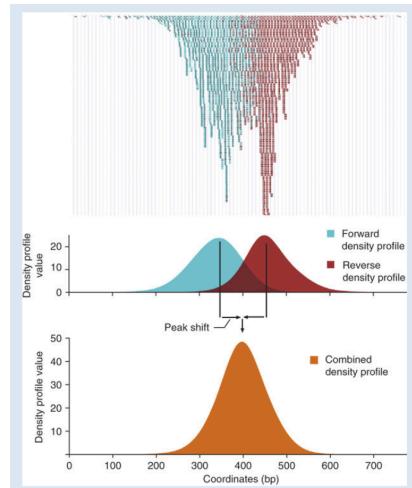
Mikkelsen,T.S. et al. *Nature* **448**, 553-562 (2007)

What do we expect at bound sites?

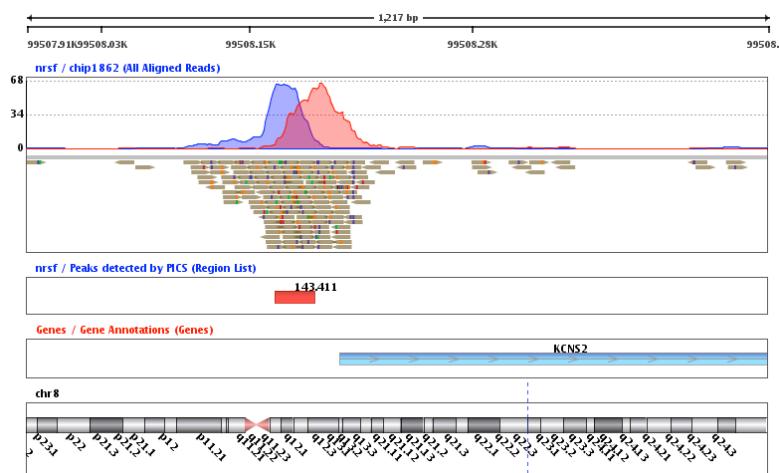
- Fragments contain the TF binding site at a (mostly) random position within them.
- Reads are (randomly) from left or right edges (sense or antisense) of fragments.
- Therefore, detection requires:
 - Shift reads appropriately
 - Background estimation
 - Peak identification

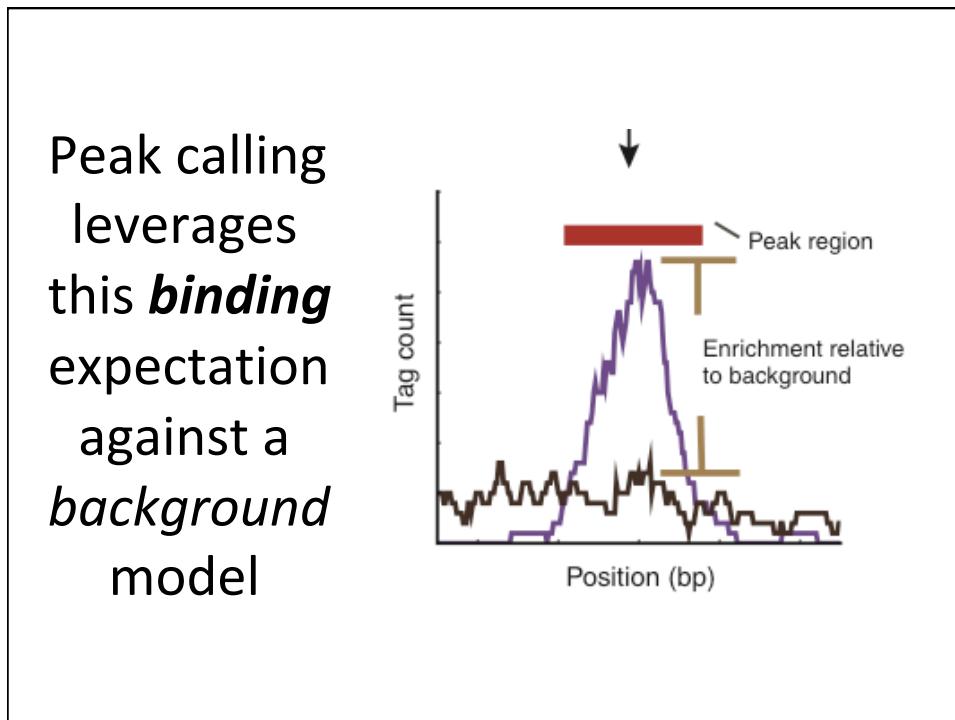
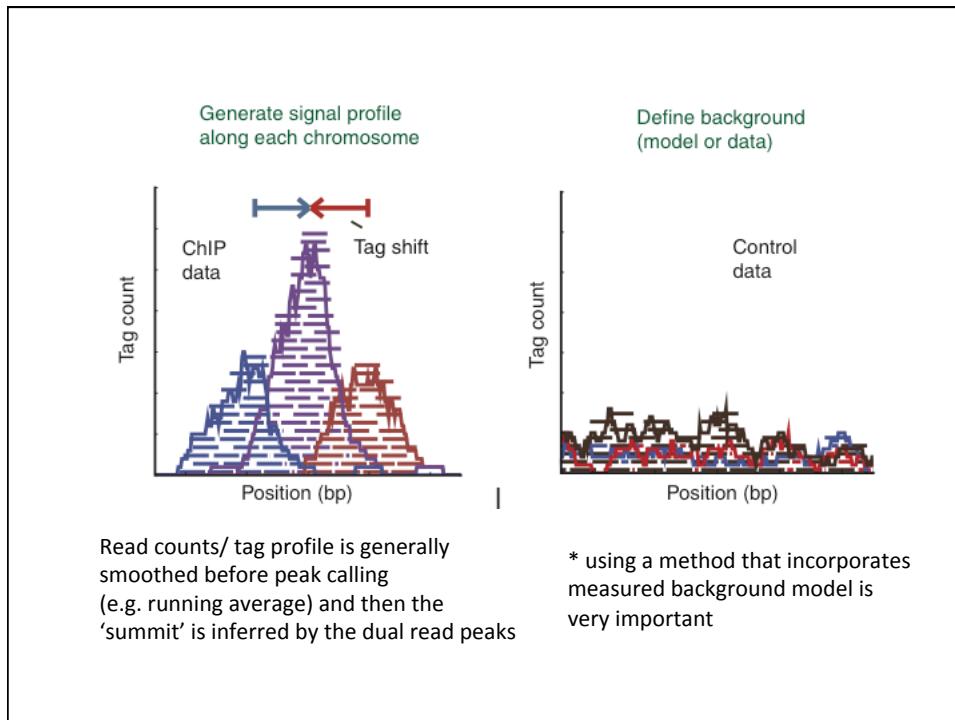


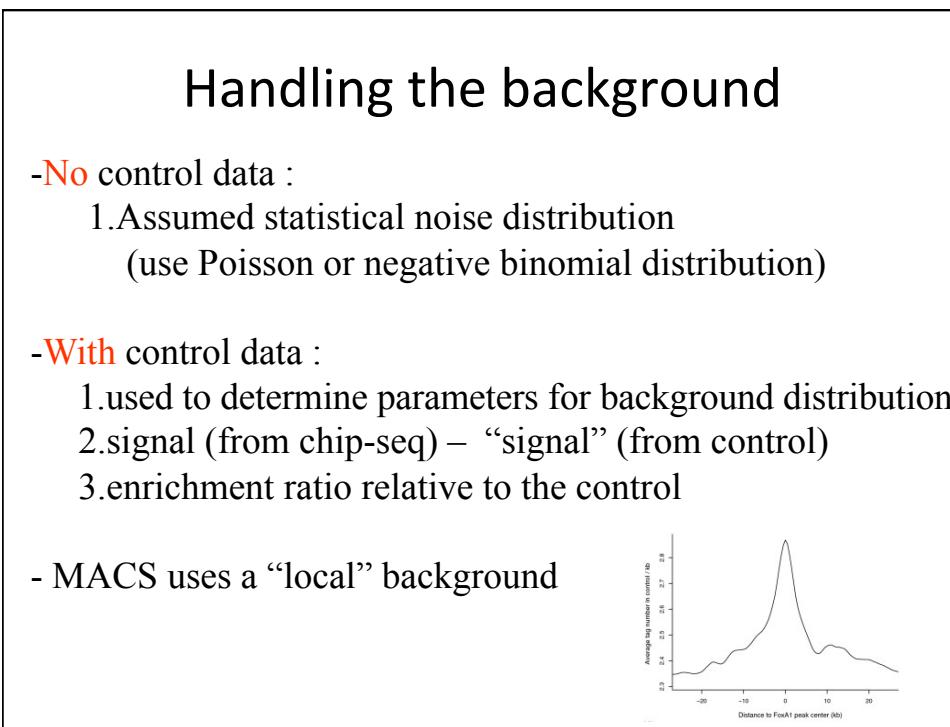
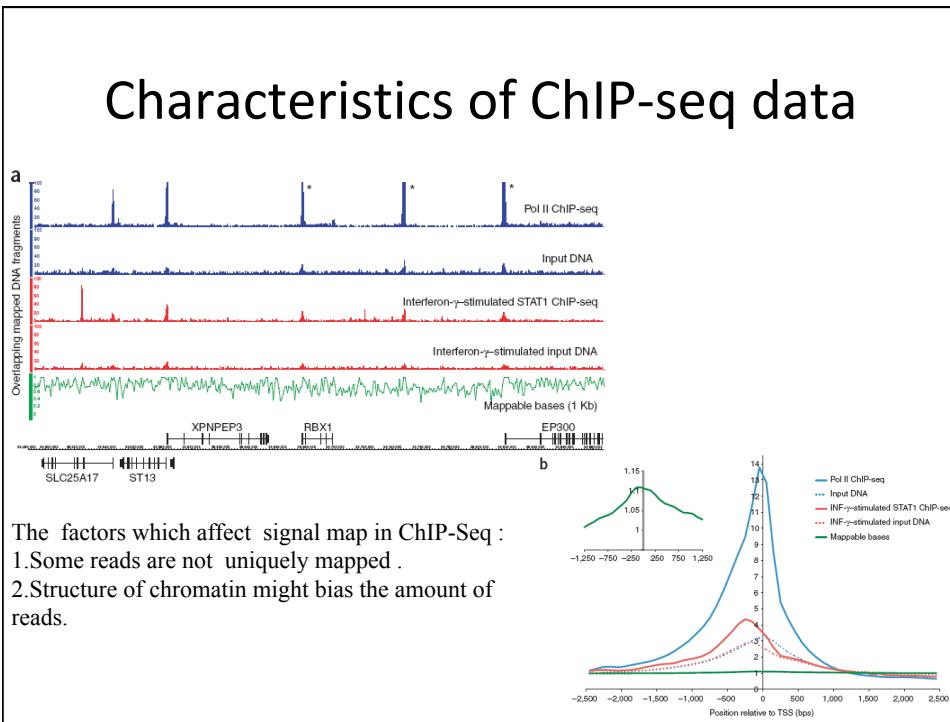
The true peak comes from a “shift”
of the obtained reads.



For example ...







Peak call criteria

Peak call criteria are for selection of candidate peaks.

- 1.absolute signal threshold
- 2.minimum enrichment to the background

Sliding window scan the genome (window size = peak-shift + std dev)

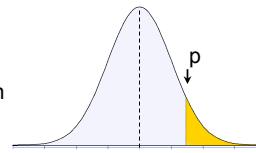
- a) Count # + and - chIP and input reads
- b) Calculate binomial p-value for chIP enrichment and +/- skew

$$p\text{-value} = 1 - \sum_{k=0}^{t-1} \binom{n}{k} p^k (1-p)^{n-k}$$

n = # chIP + # input window reads
 t = observed # reads in chIP window
 p = expected probability of chIP read in window

Estimate FDRs

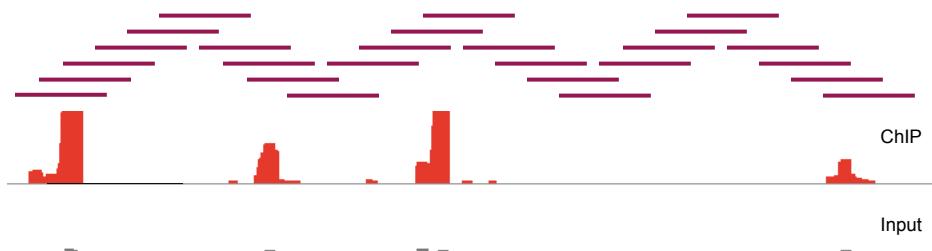
- a) Use Storey's q-value method to convert binomial p-values
- b) Calculate empirically using an input control null distribution



Merge overlapping windows with an FDR <1% into binding peaks

Windows → Binding Peaks

Overlapping windows



Input

Binomial

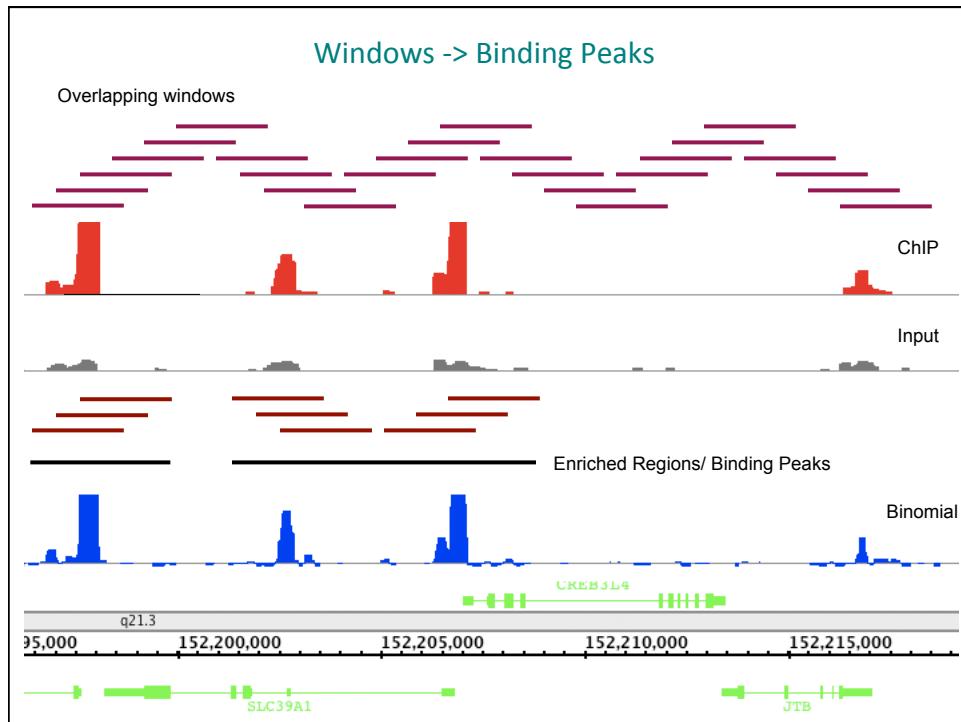
q21.3

95,000 152,200,000 152,205,000 152,210,000 152,215,000

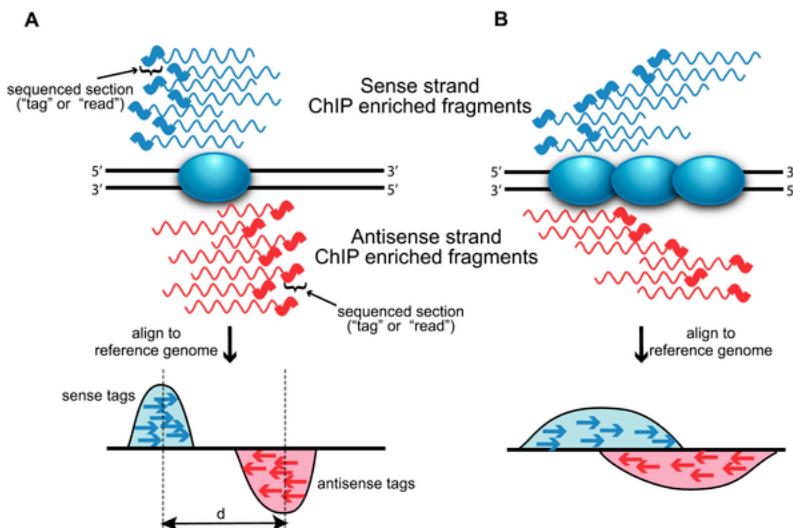
SLC39A1

CREB3L4

JTB

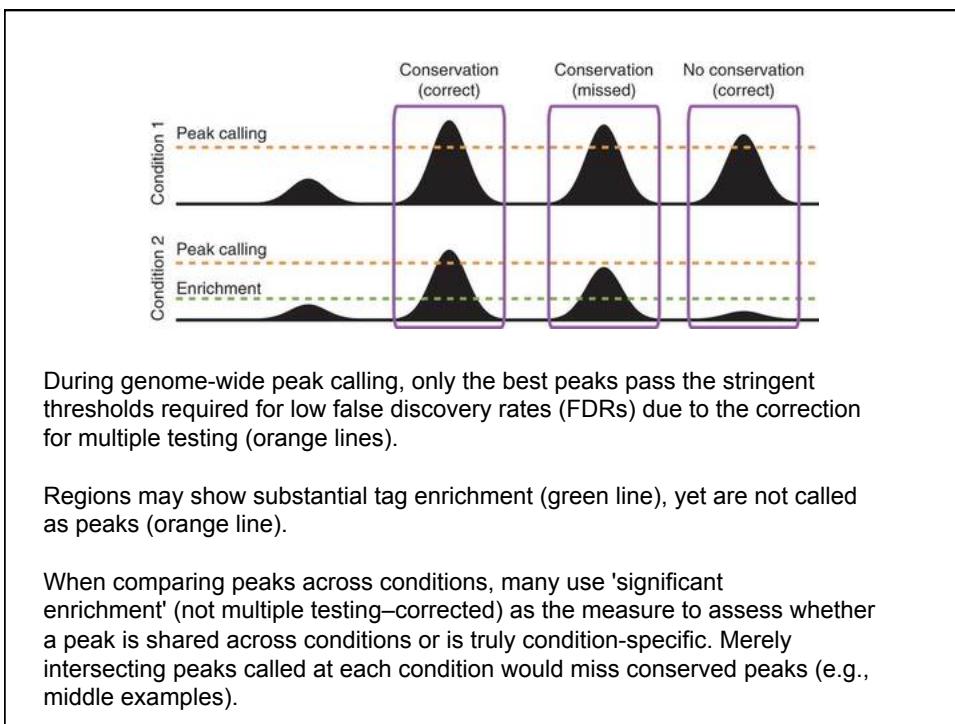
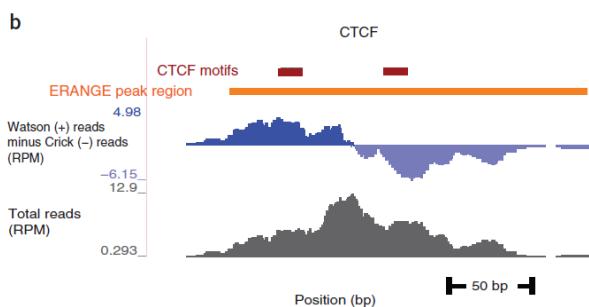


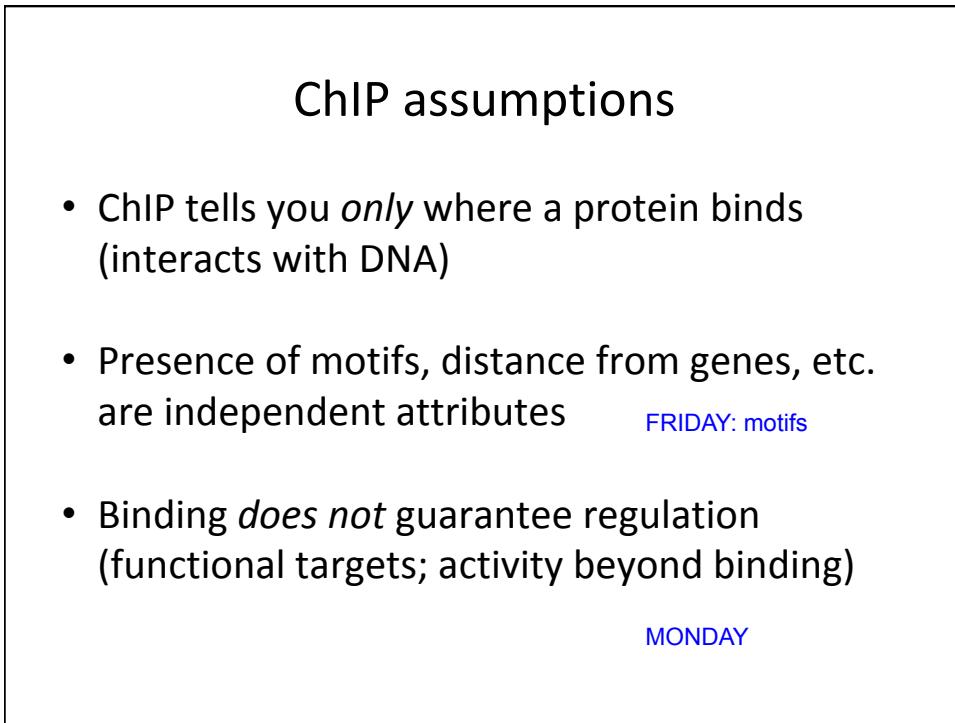
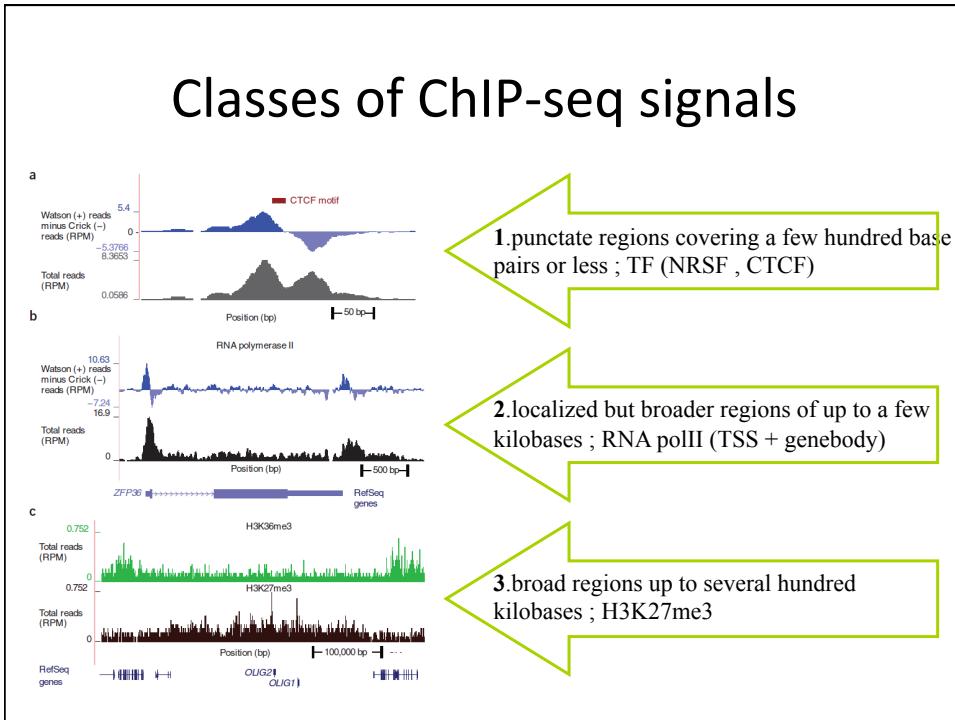
If we know the shear distribution,
we can even infer adjacent binding.



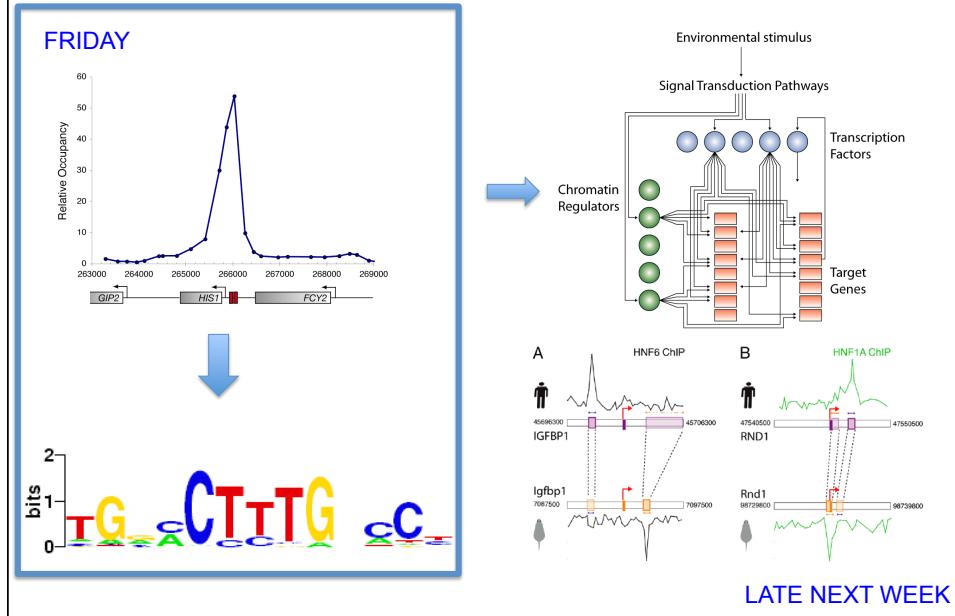
But peak deconvolution is tricky.

1. Solving spatial or temporal convolutions of multiple biologically true events.
2. Weak signal site can be confirmed by ChIP-quantitative PCR, bidding motif, protection assay, etc.



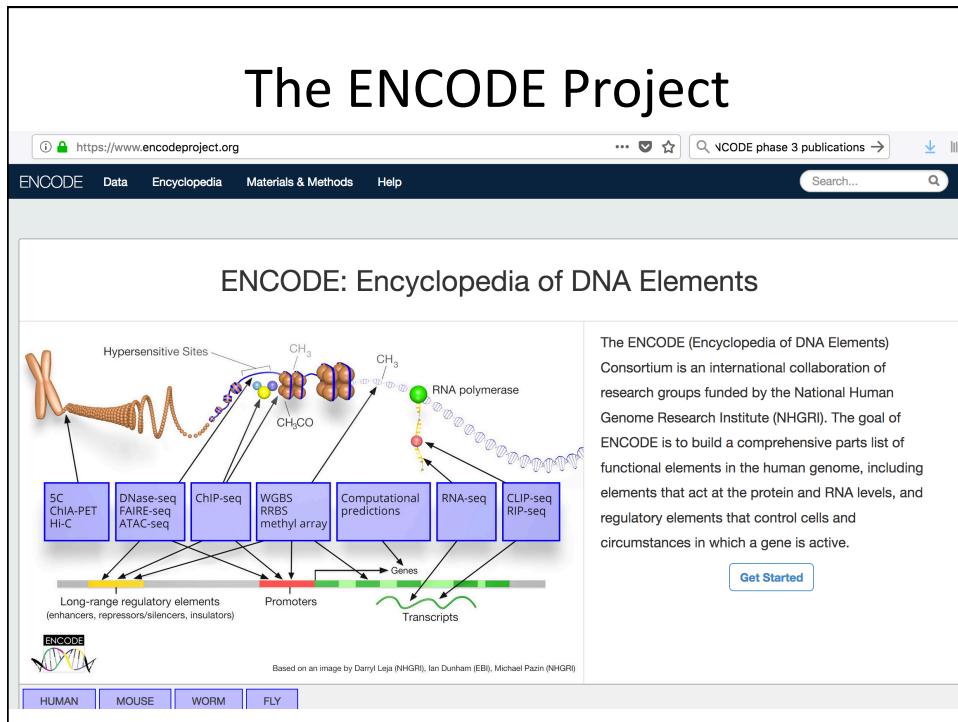


So now what do we do with the peaks?



ENCODE

- The **Encyclopedia of DNA Elements (ENCODE)** is a public research consortium initiated by the US National Human Genome Research Institute (NHGRI) in September 2003.
- The goal is to find all functional elements in the human genome.
- All data generated in the course of the project will be released “rapidly” into public databases.



- Pilot phase – 2003-2007 – method evaluation
 - 1% of genome
- Production phase 2007-2012
 - September 2012 – 30 papers published
 - 442 scientists
 - 31 labs
 - 147 different types of cells with 24 types of experiments
 - 1,642 experiments
 - Data released
- Subsequently have had multiple rounds of additional funding, data releases, and publications.

