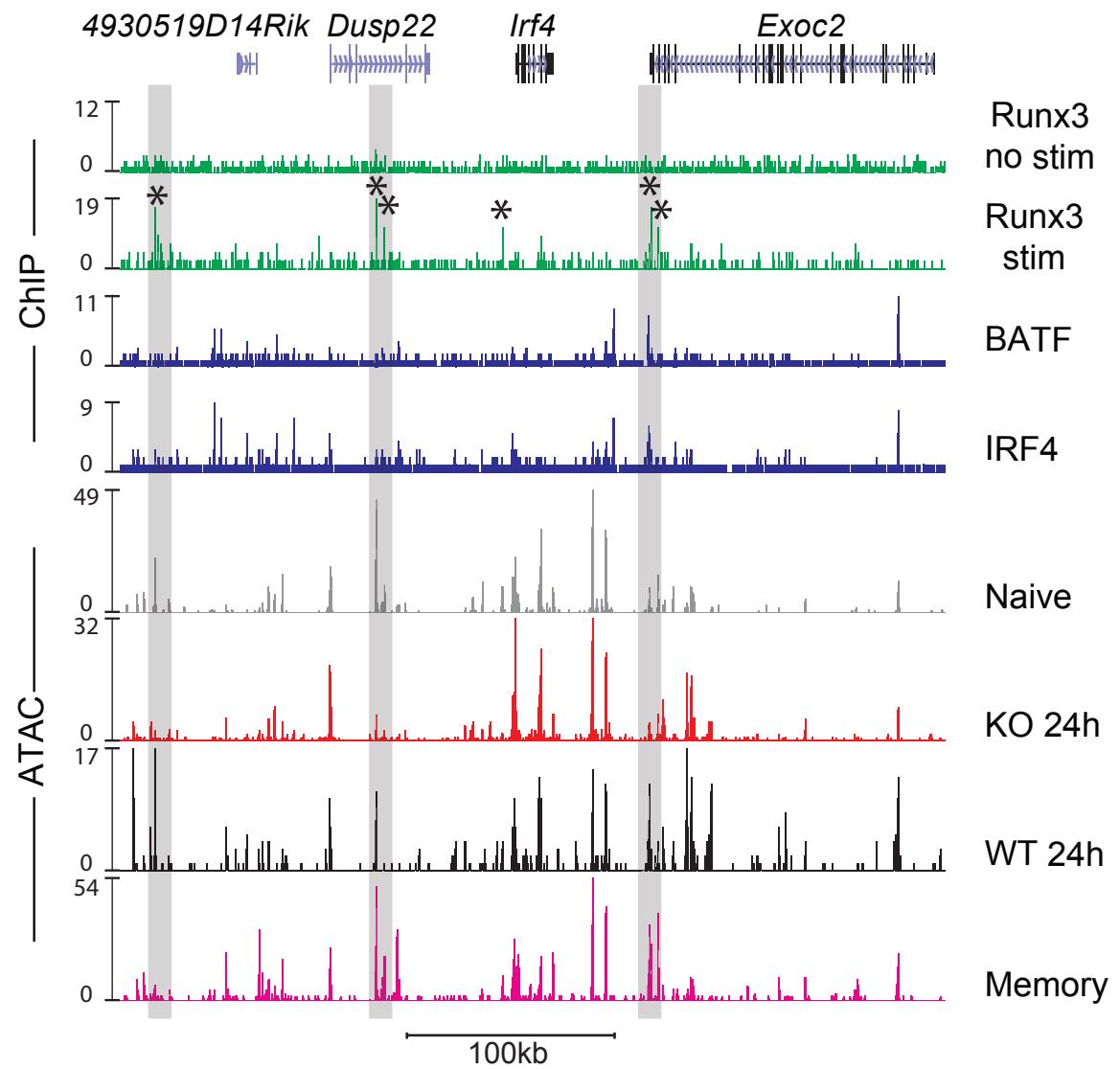
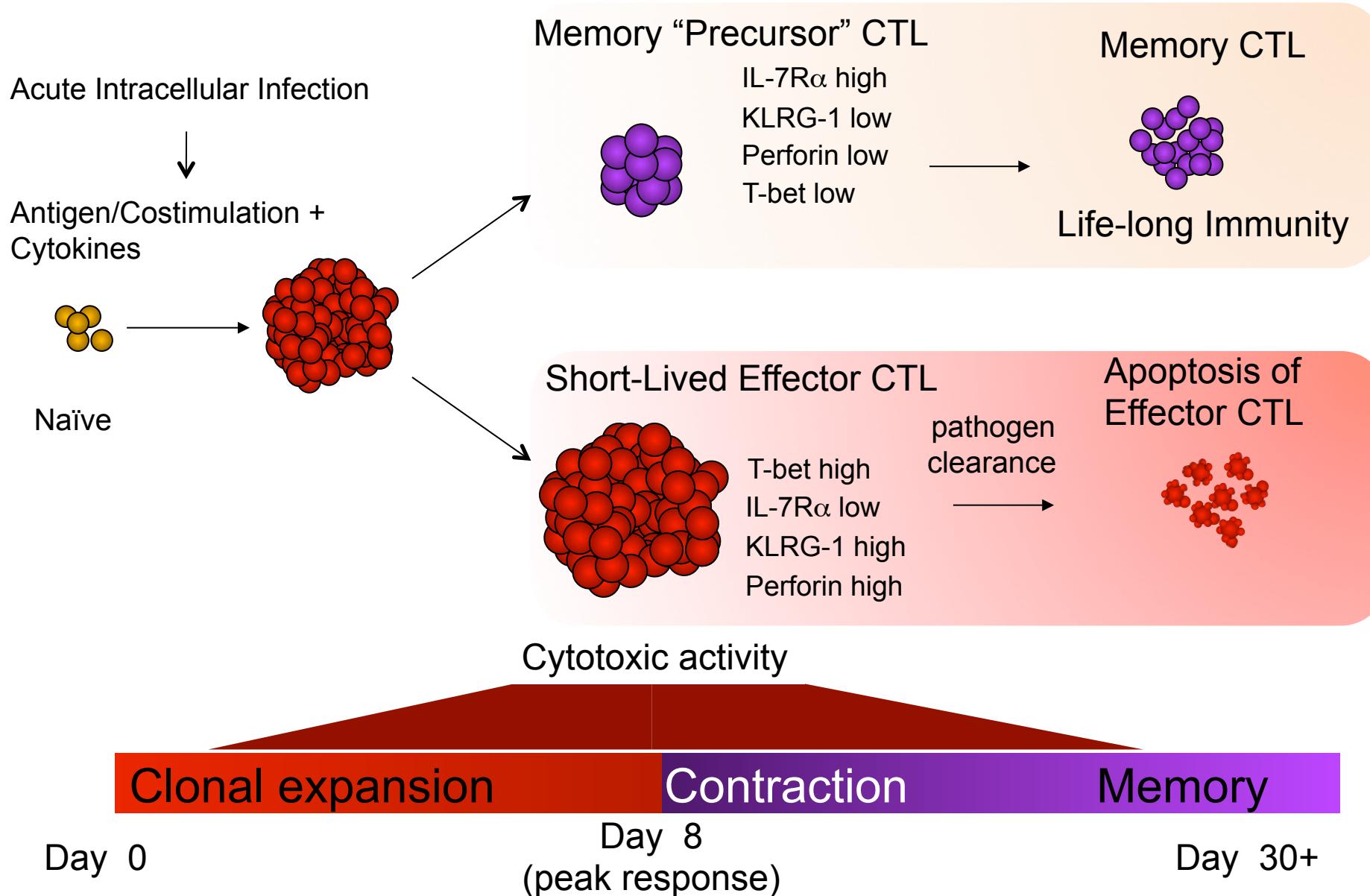


# Genome-scale Mapping of Protein-DNA Interactions in vivo

Matthew E. Pipkin, Ph.D.  
Department of Immunology and Microbiology  
The Scripps Research Institute-FL  
Phone: (561) 228-2182  
E-mail: [mpipkin@scripps.edu](mailto:mpipkin@scripps.edu)

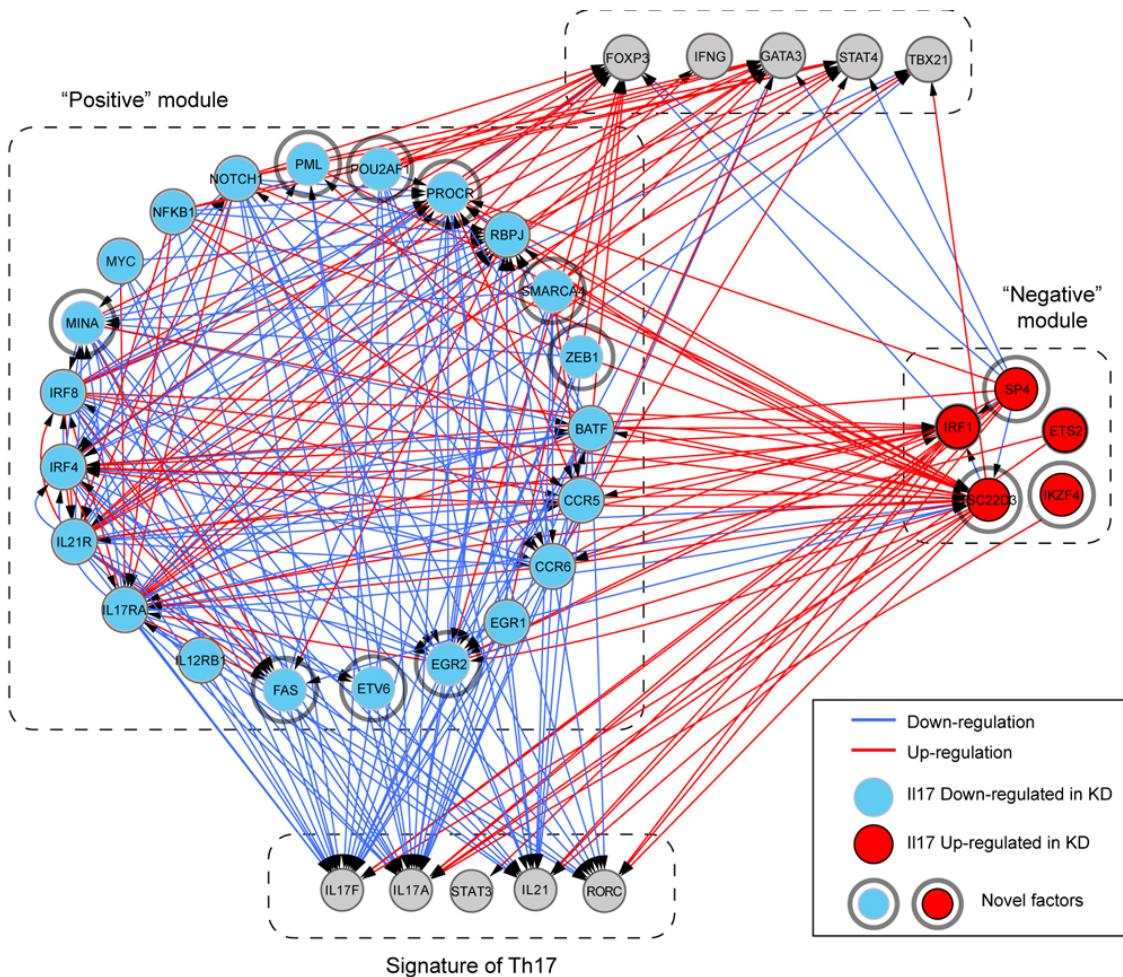


# Deciphering how gene expression controls cell differentiation

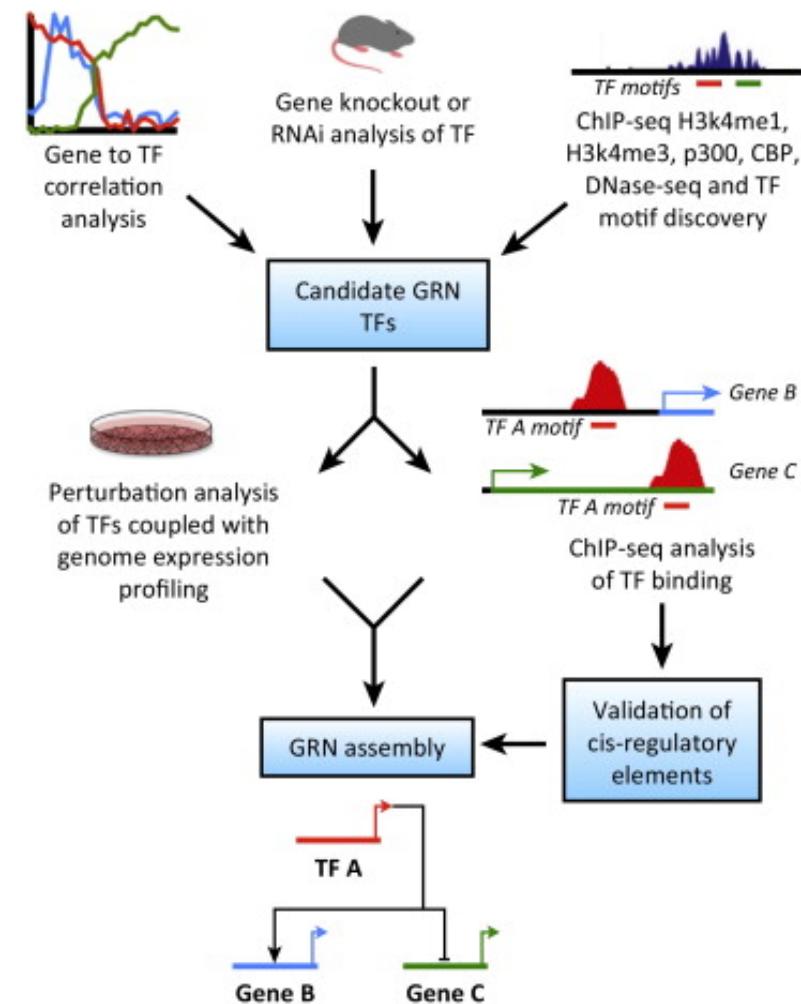


# Organizing Gene-Regulatory Networks into Circuits

## Gene Regulatory Networks Predicted from Gene Expression

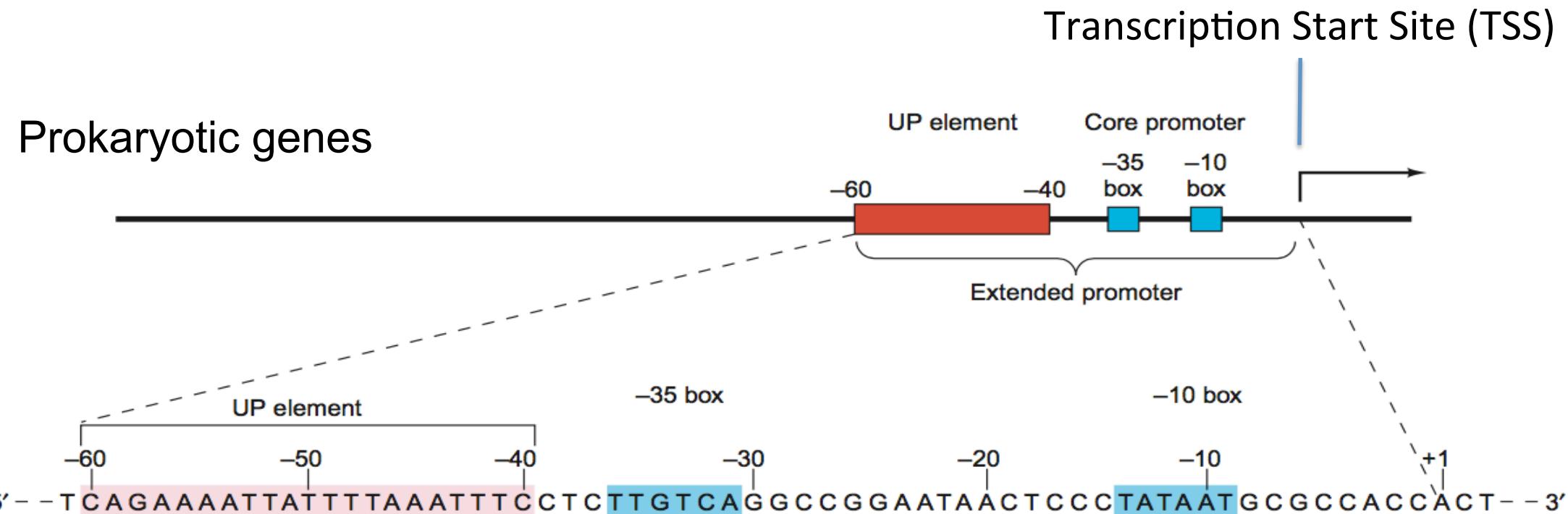


## Establishing Transcriptional Circuitry



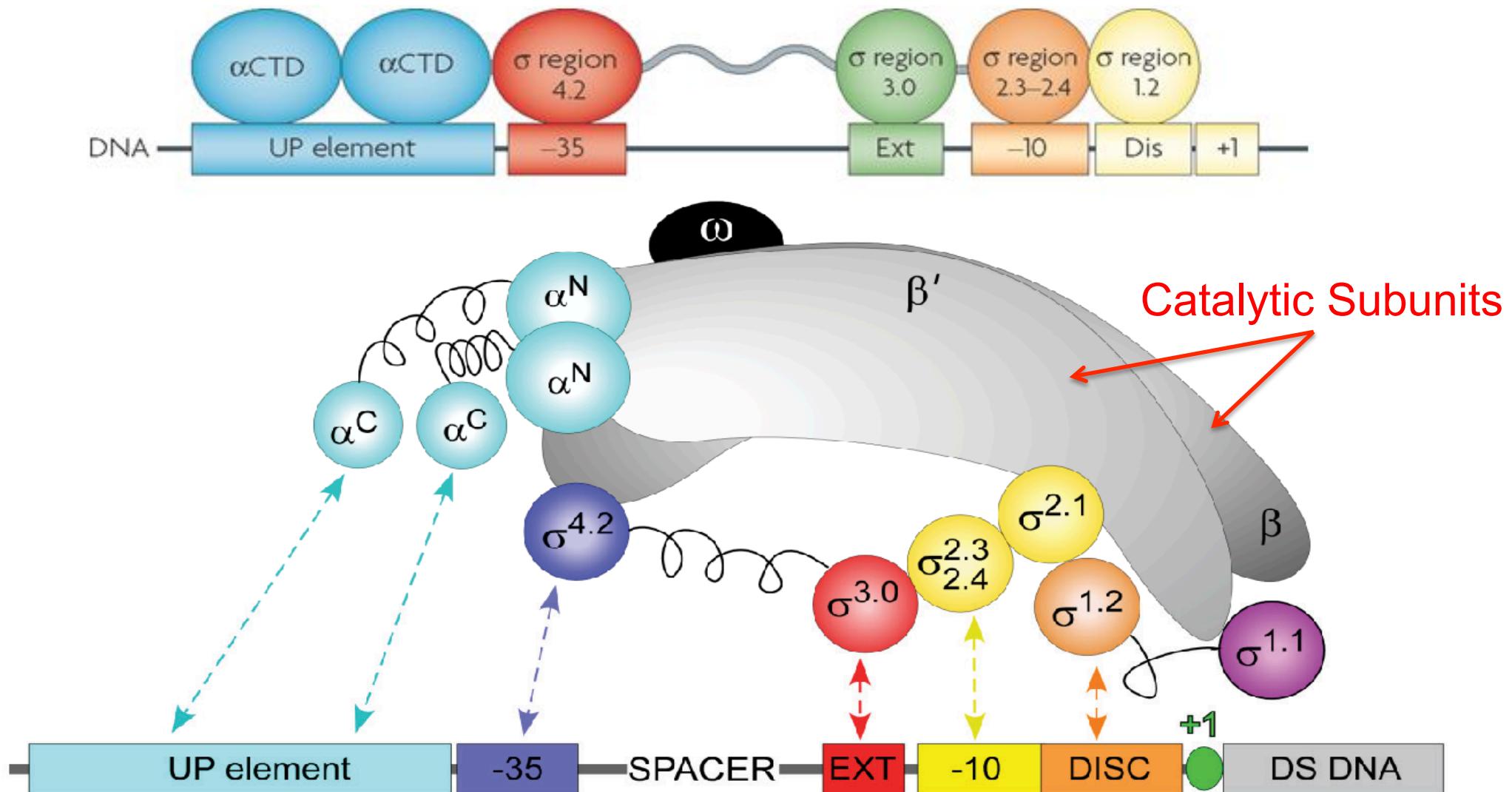
TRENDS in Immunol.

Transcription initiates from DNA-encoded cis-regulatory regions called promoters

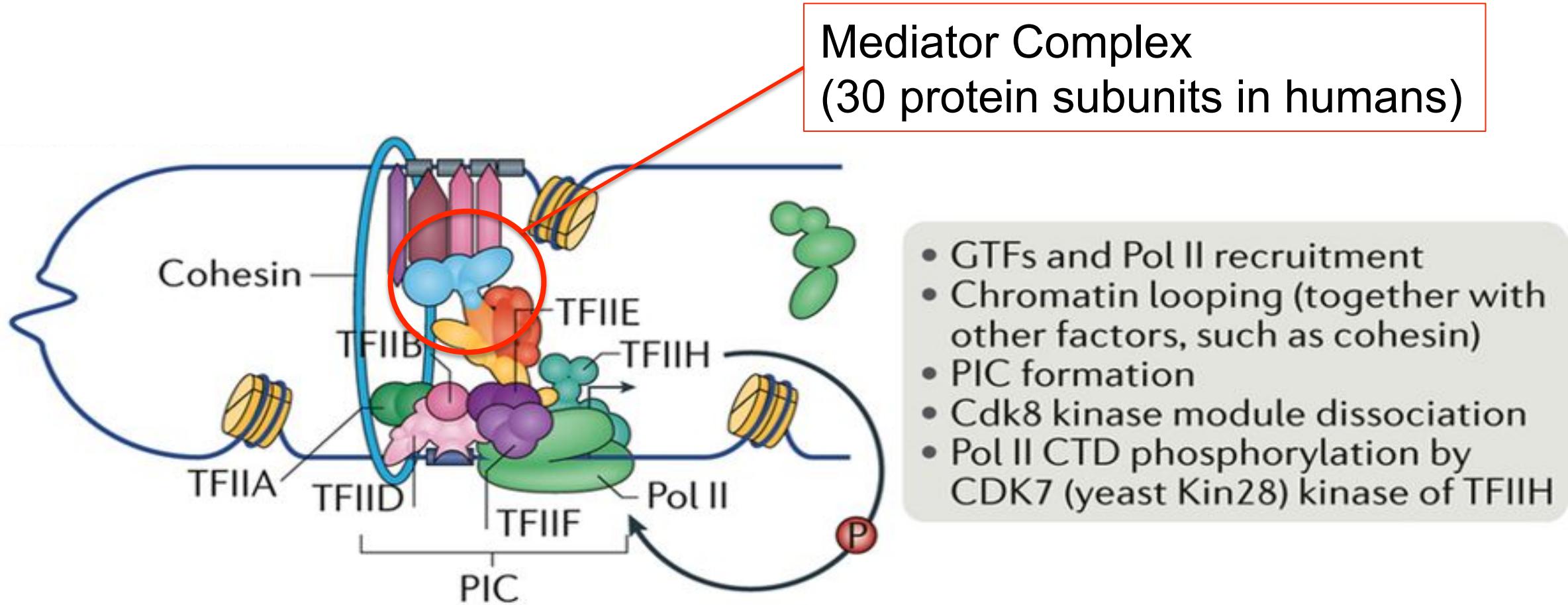


**Figure 6.6 The *rrnB* P1 promoter.** The core promoter elements ( $-10$  and  $-35$  boxes, blue) and the UP element (red) are shown schematically above, and with their complete base sequences (nontemplate strand) below, with the same color coding. (Source: Adapted from Ross et al., "A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase." *Science* 262:1407, 1993.)

# Recognition of promoter elements by multiple RNA Polymerase (RNA Pol) subunits position the holoenzyme at TSSs

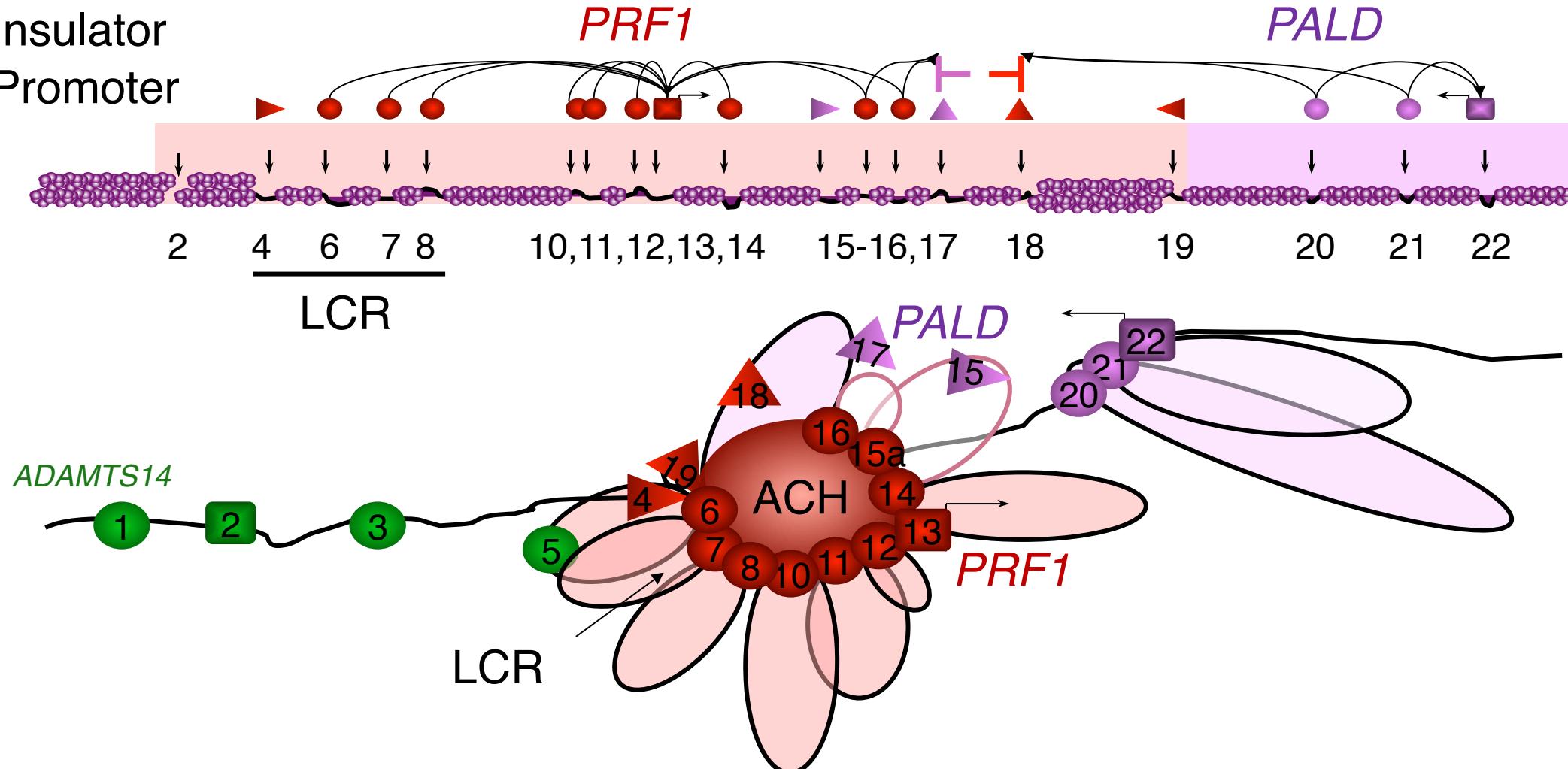


# Extensive Protein-DNA and Protein-Protein Interactions Govern the Recruitment and Activity of RNA Polymerase II

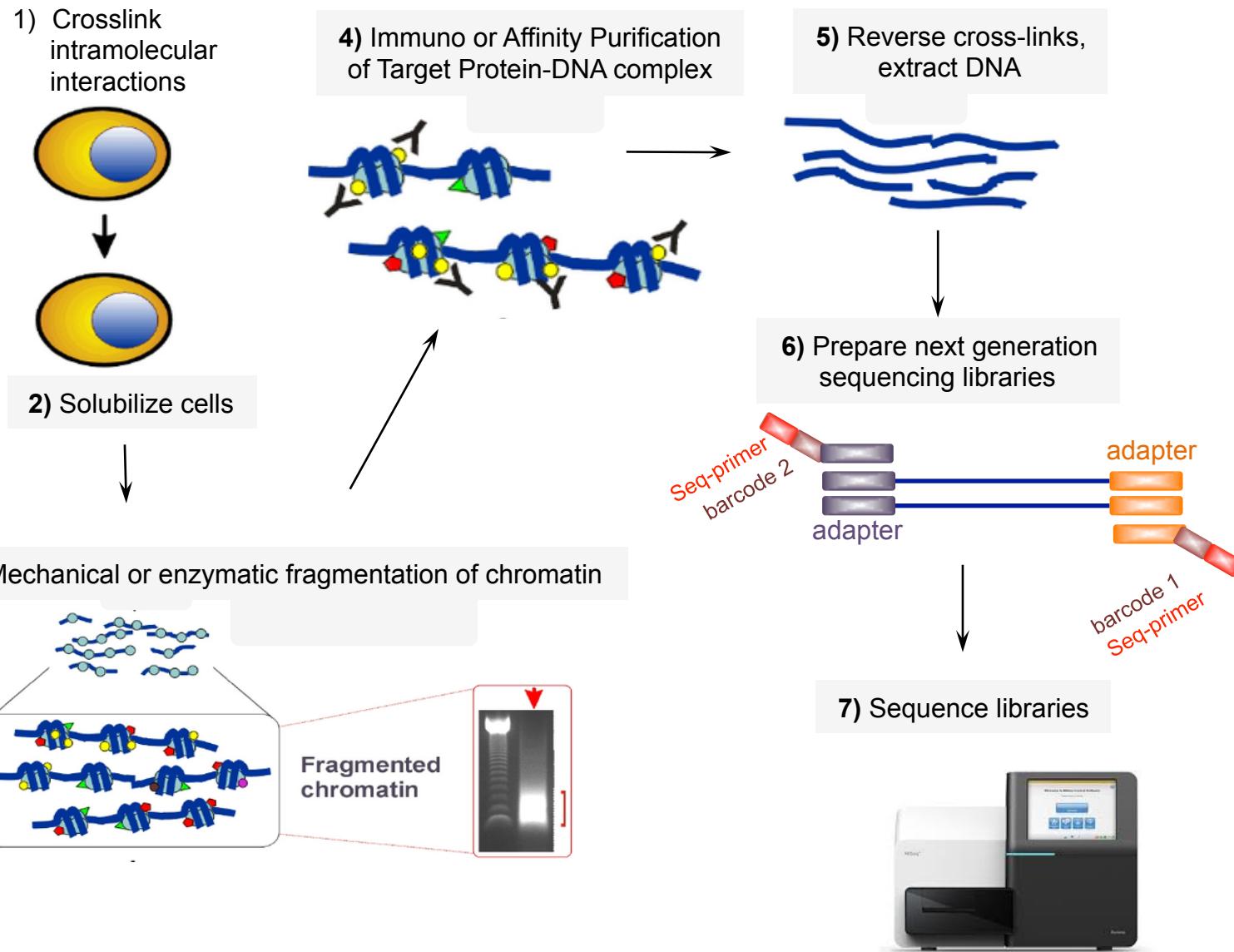


# Multiple Long-Range Regulatory Regions Converge Spatially to Regulate Transcription

- Enhancer
- Insulator
- Promoter

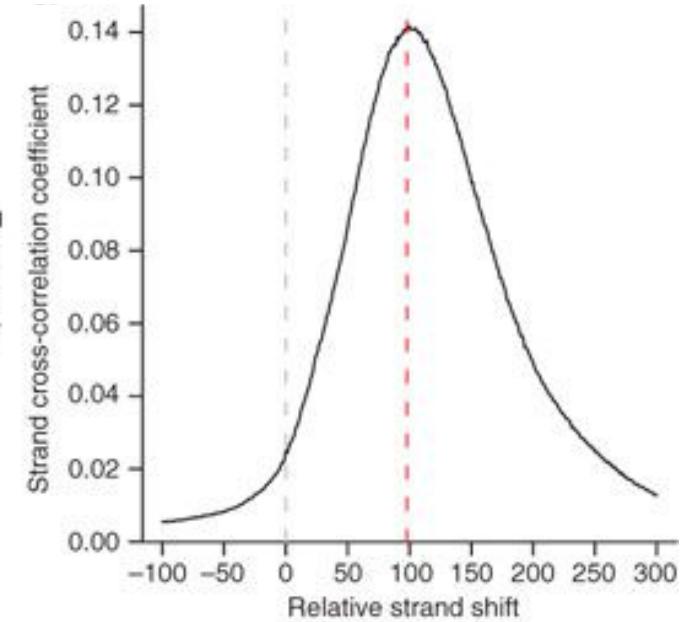
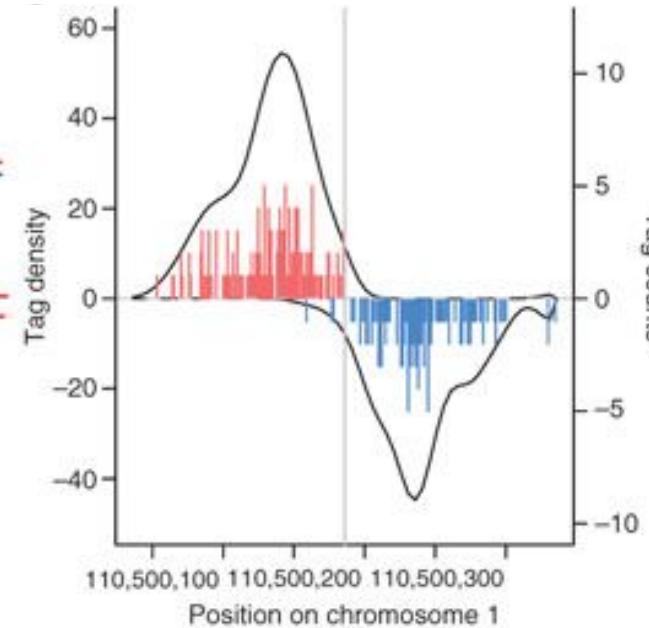
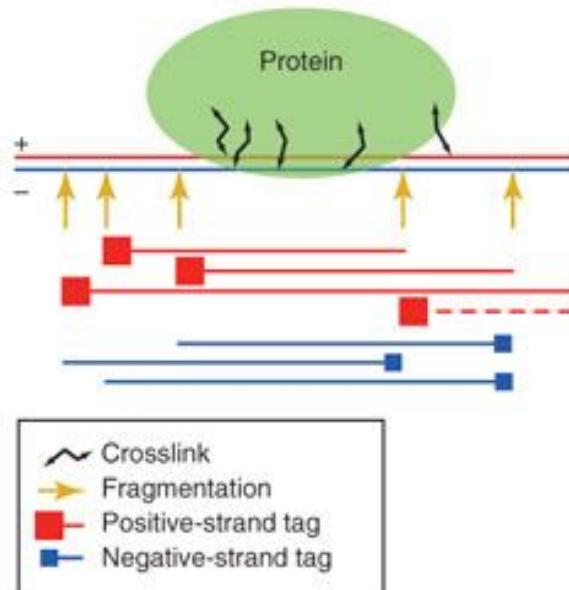
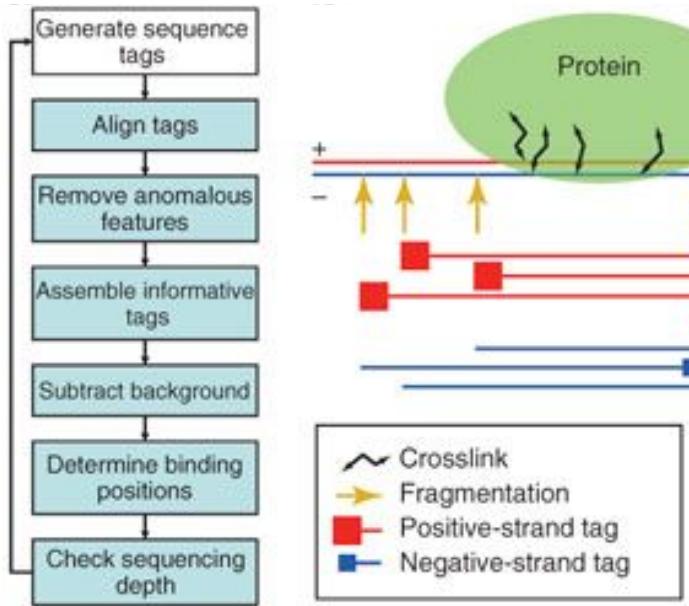


# Chromatin Immunoprecipitation Analysis Approach: Key Steps

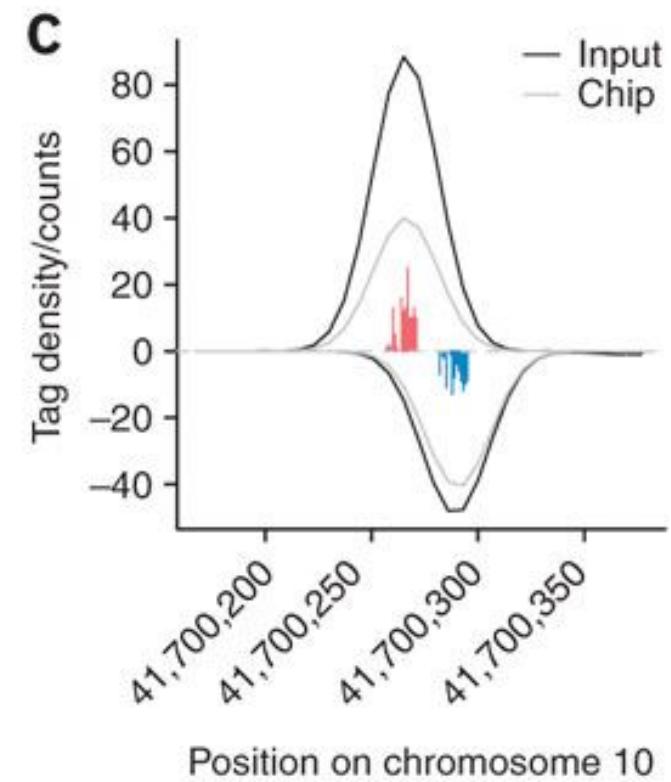
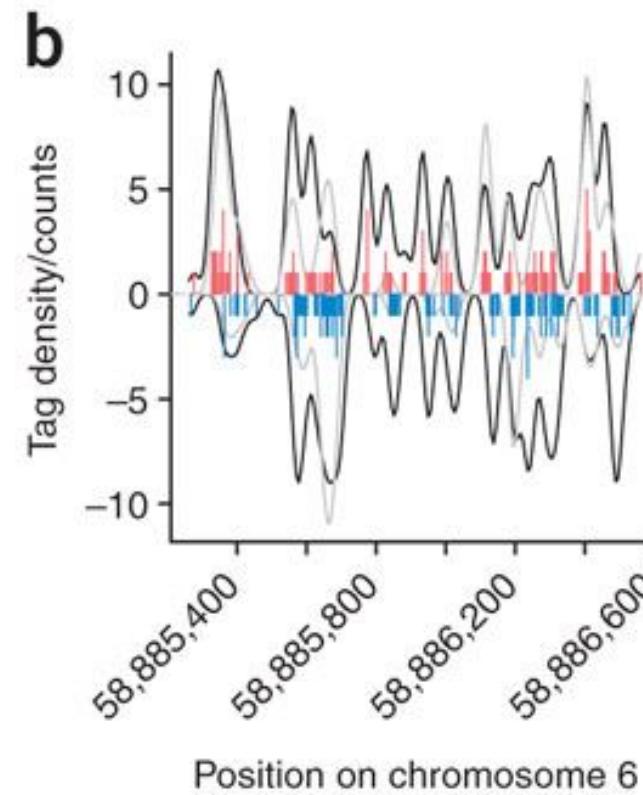
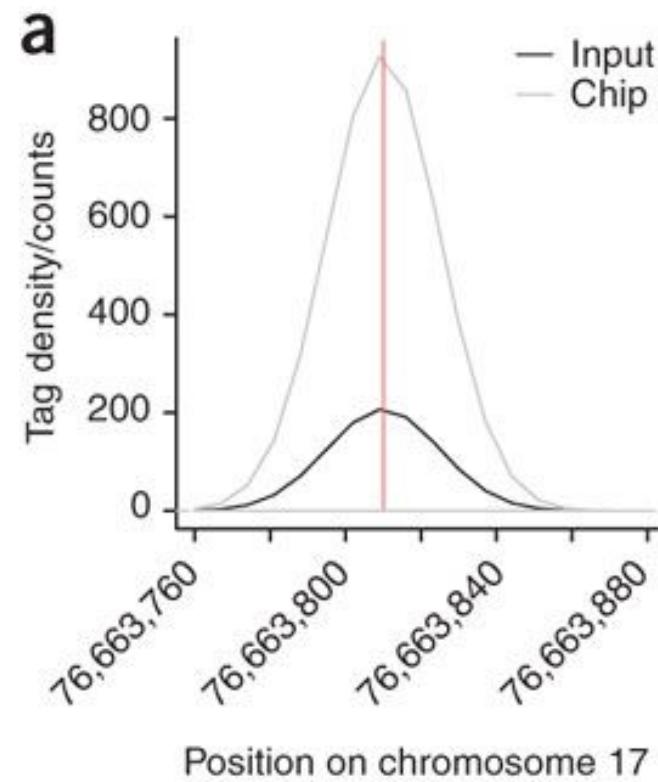


Rank/Motif	Motif	P-value	-log <sub>10</sub> (P-value)	% of Targets	STD(Bp)	Pos. M. Depth	Pos. M. (Std)	Motif File
1	TGTTCATACATA	1e-12661	2.915e+04	70.91%	15.19%	40.5bp	Pos. M. (Std)	motif file (n=65)
2	CTTGGCAG	1e-578	1.332e+03	27.14%	16.52%	54.0bp	Pos. M. (Std)	motif file (n=65)
3	TTTATTGGC	1e-384	8.860e+02	17.77%	10.53%	53.9bp	Pos. M. (Std)	motif file (n=62)
4	SCTCTGTAATAT	1e-164	3.783e+02	3.17%	1.28%	52.2bp	Pos. M. (Std)	motif file (n=62)
5	ATGACTCA	1e-151	3.485e+02	3.38%	1.47%	52.0bp	Pos. M. (Std)	motif file (n=65)
6	CCCACTCTGG	1e-107	2.485e+02	1.21%	0.35%	56.3bp	Pos. M. (Std)	motif file (n=69)

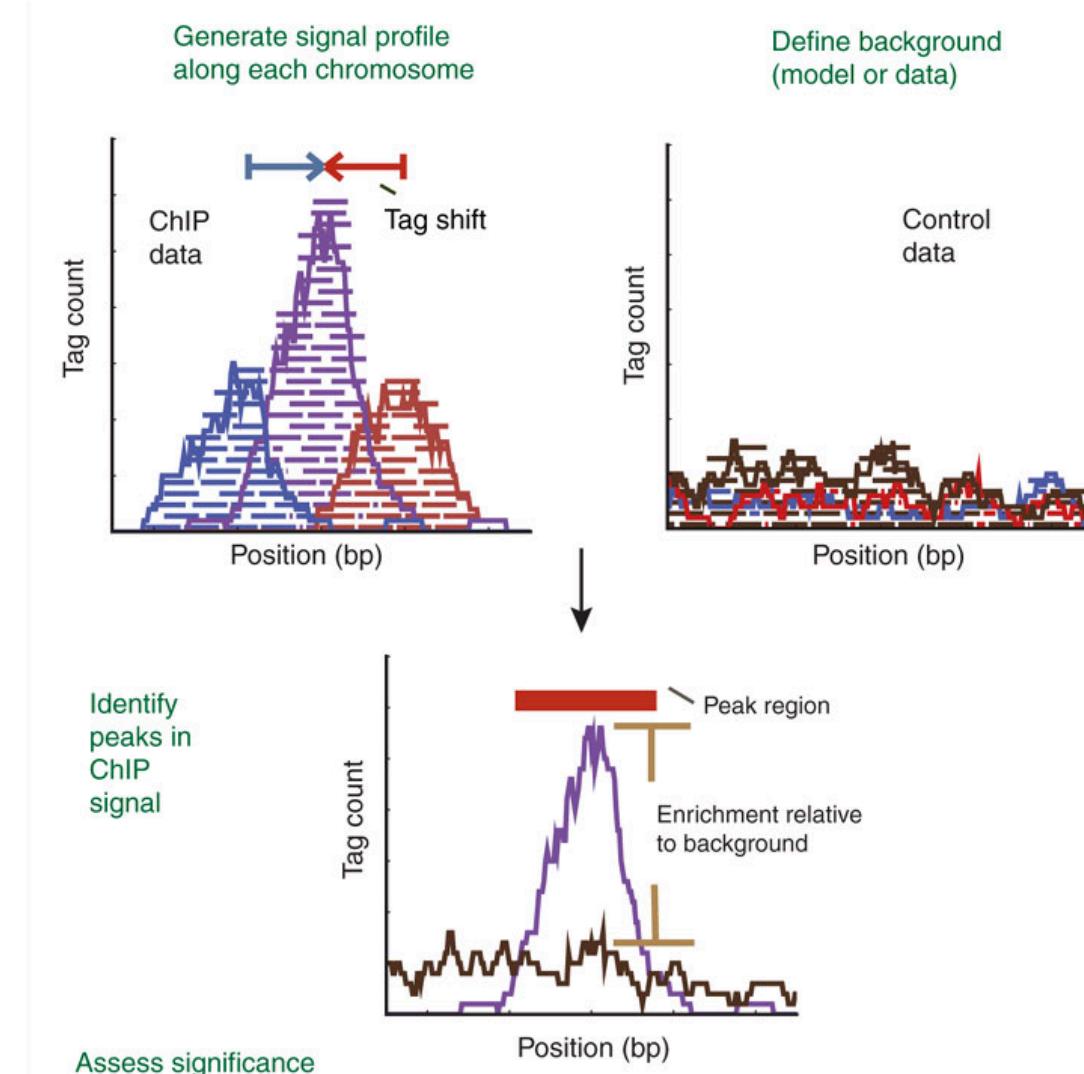
# Detecting protein binding events within ChIP-seq data



# Black lists: anomalies in background sequencing read distributions



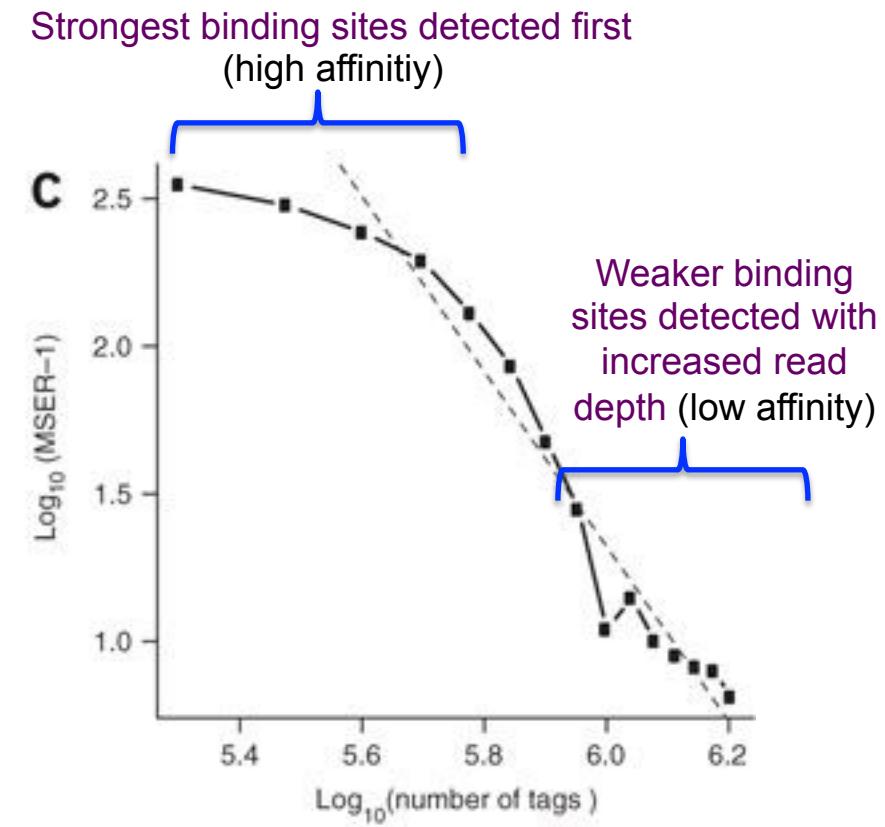
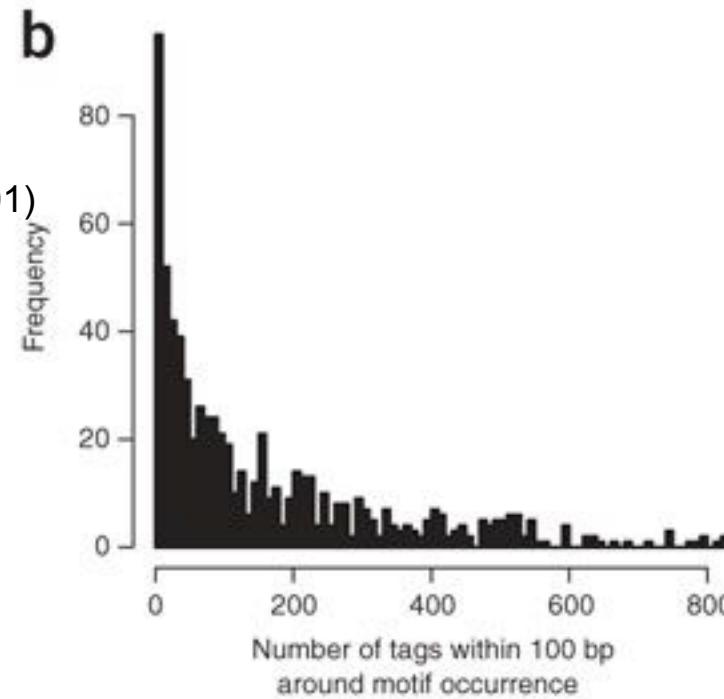
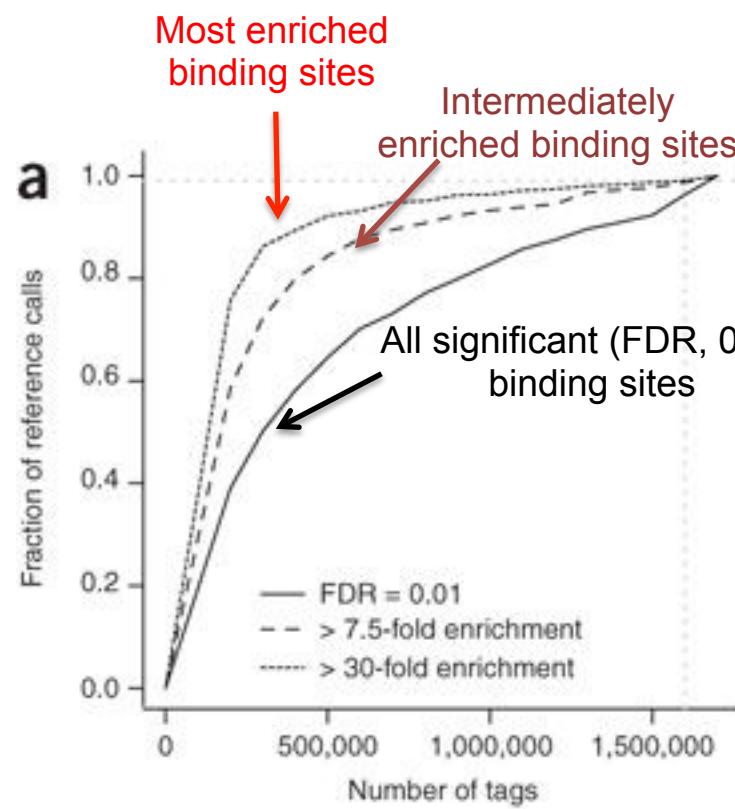
# “Calling” peaks that represent protein binding regions



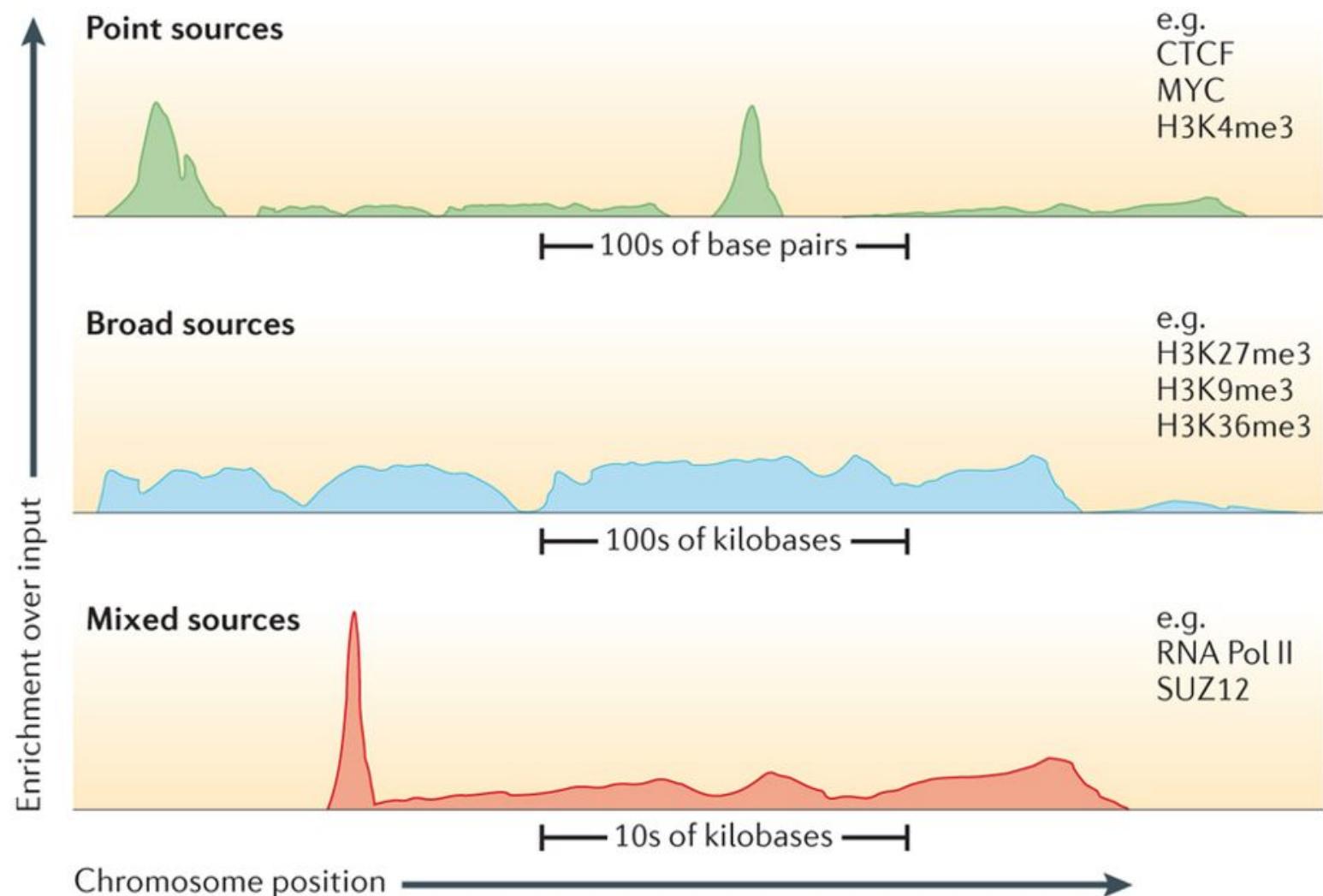
## Downstream Analyses

- 1) Annotate peaks to genomic features (e.g., TSS, Enhancers, Introns, Intergenic regions, etc.)
- 2) Differential binding analyses between different sample conditions (e.g., WT vs. KO)
- 3) Motif analyses to identify known and unknown TF motifs within peak regions

# Increased sequencing depth increases the number of true binding sites



# Considerations: Distinct features of protein-DNA interactions result in diverse patterns of genome coverage after ChIP-seq analyses



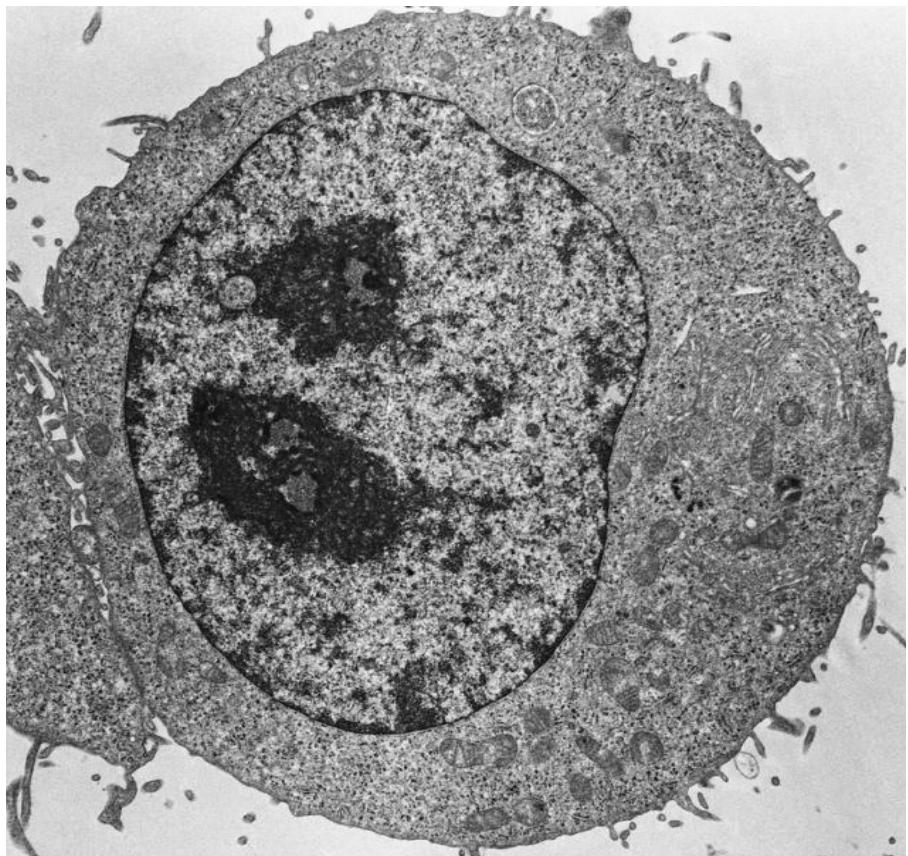
# Chromatin structure in nuclei is not homogeneous

## Rough dimensions of a typical lymphocyte

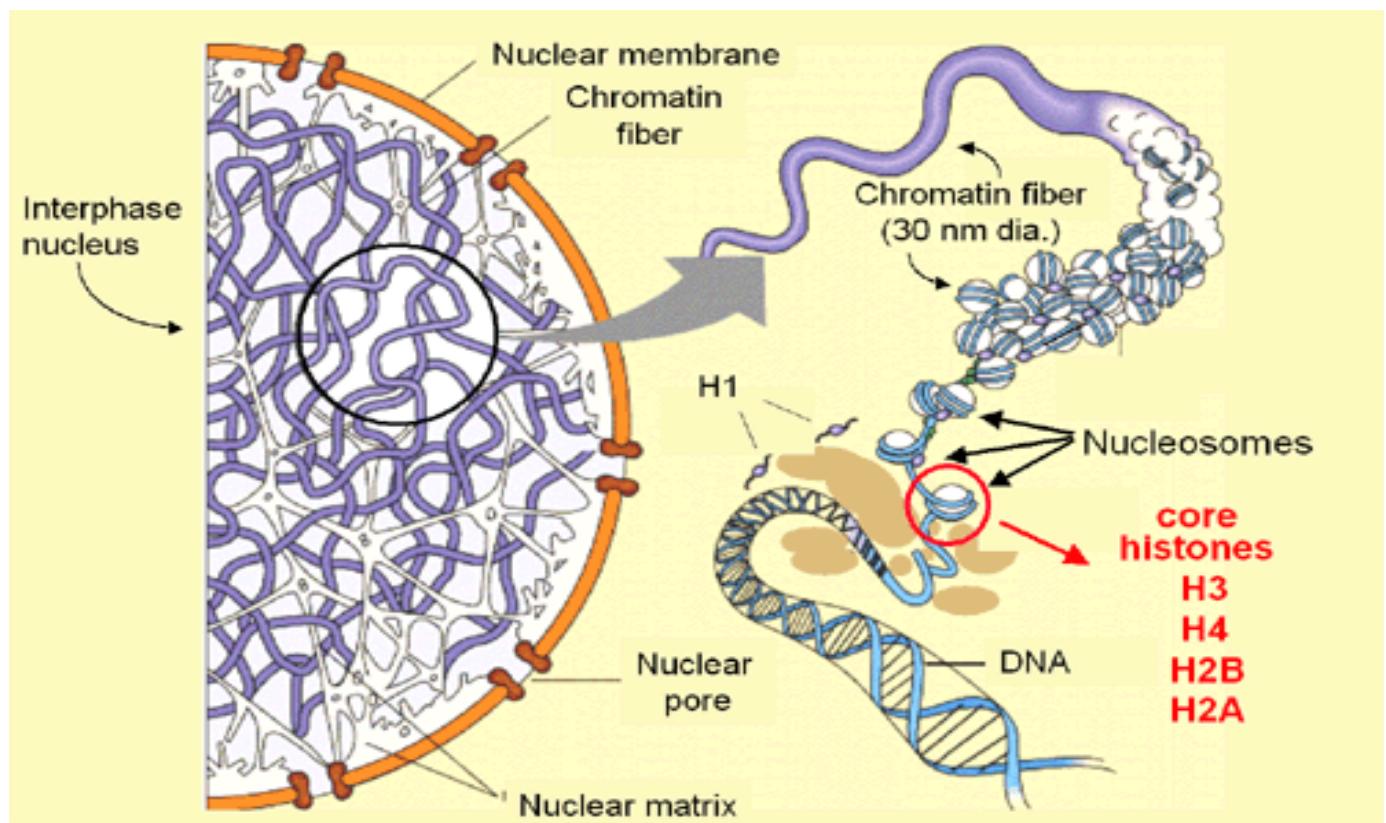
Cell diameter: ~10 microns

Nuclei diameter: ~ 2-5 microns

Nuclear volume: 100-500  $\mu\text{m}^3$



Nuclear DNA packaged in chromatin



Tem Photograph by David M. Phillips

Illustration by C.D. Allis

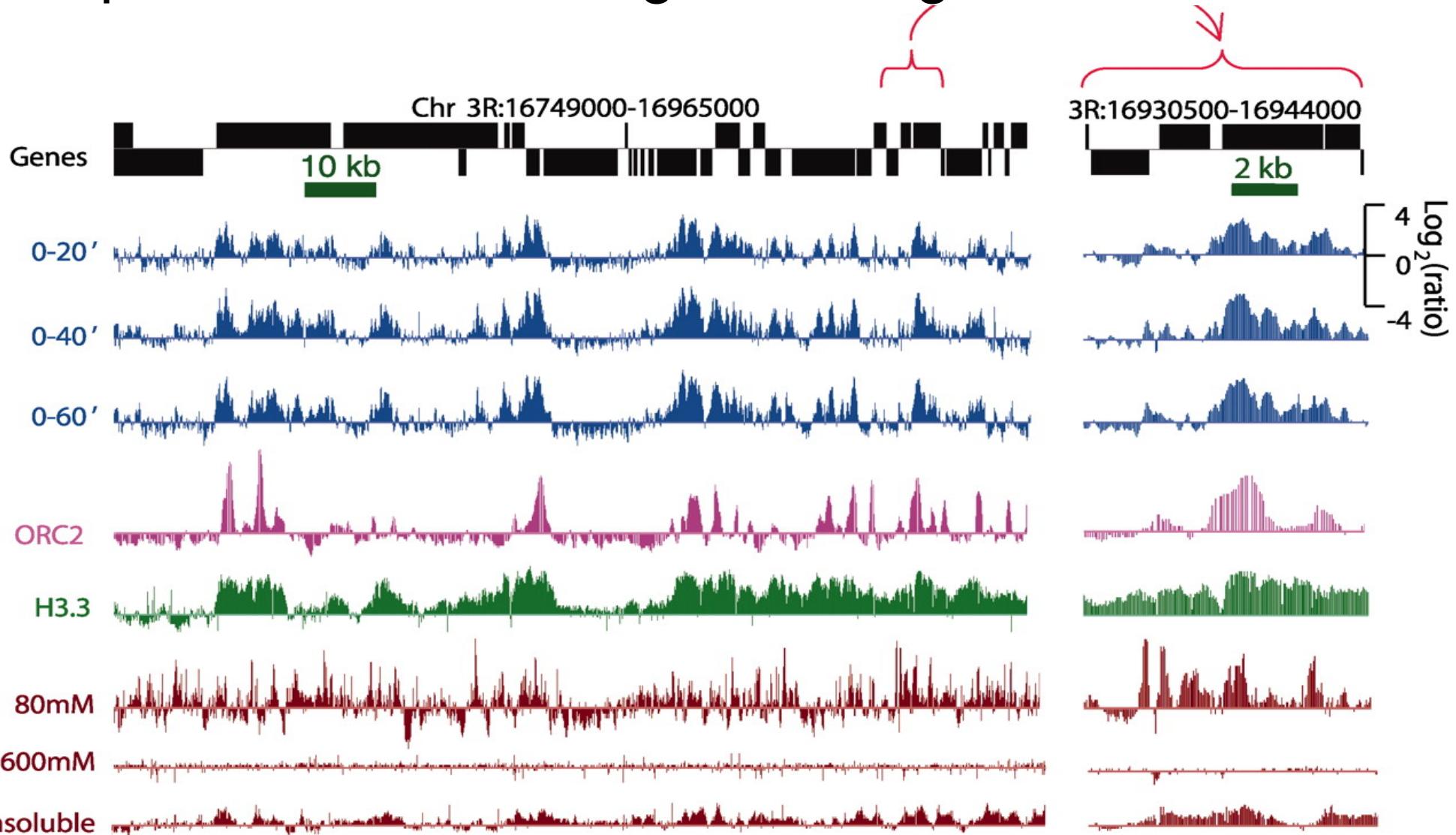
# Uneven nucleosome turnover across genomes *in vivo* correlates with nucleosome-protected DNA in regulated regions

MNase-protected regions enriched after SA-pulldown

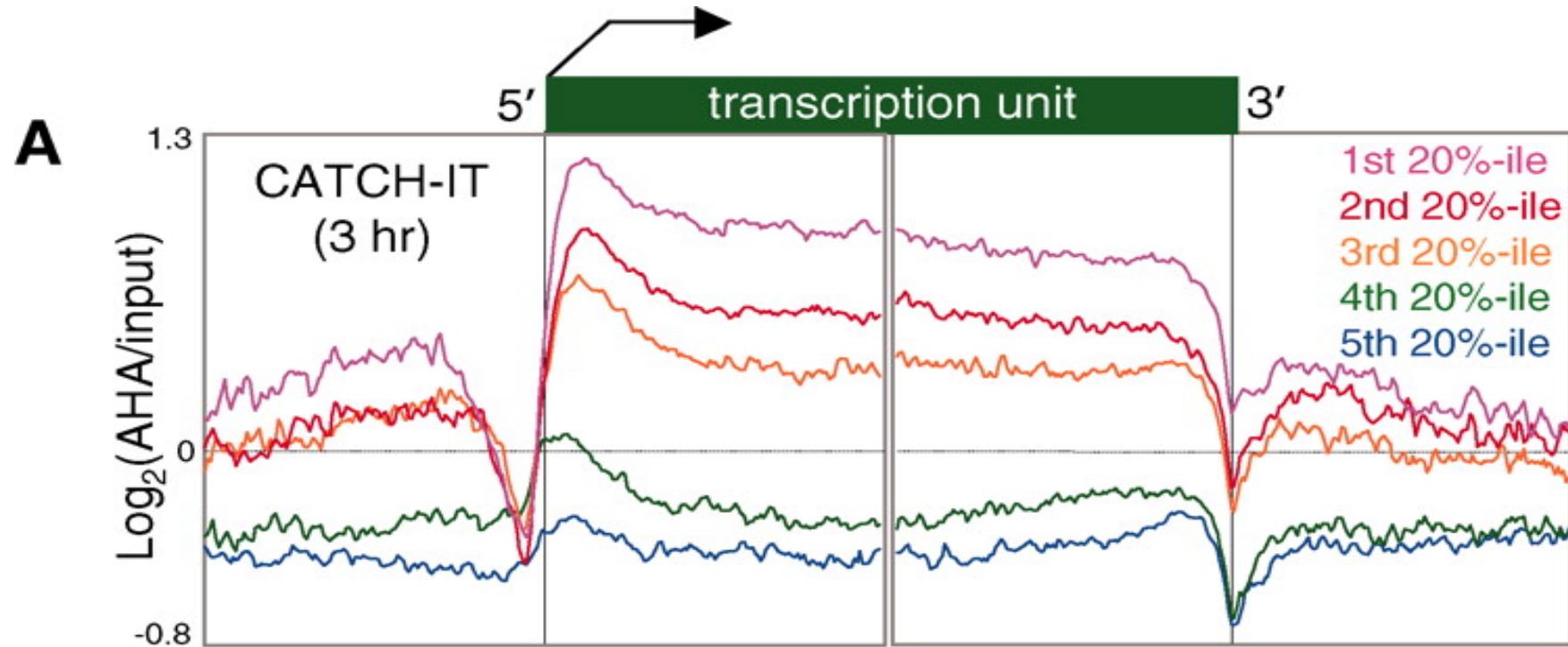
Origins of Rep

H3.3 ChIP

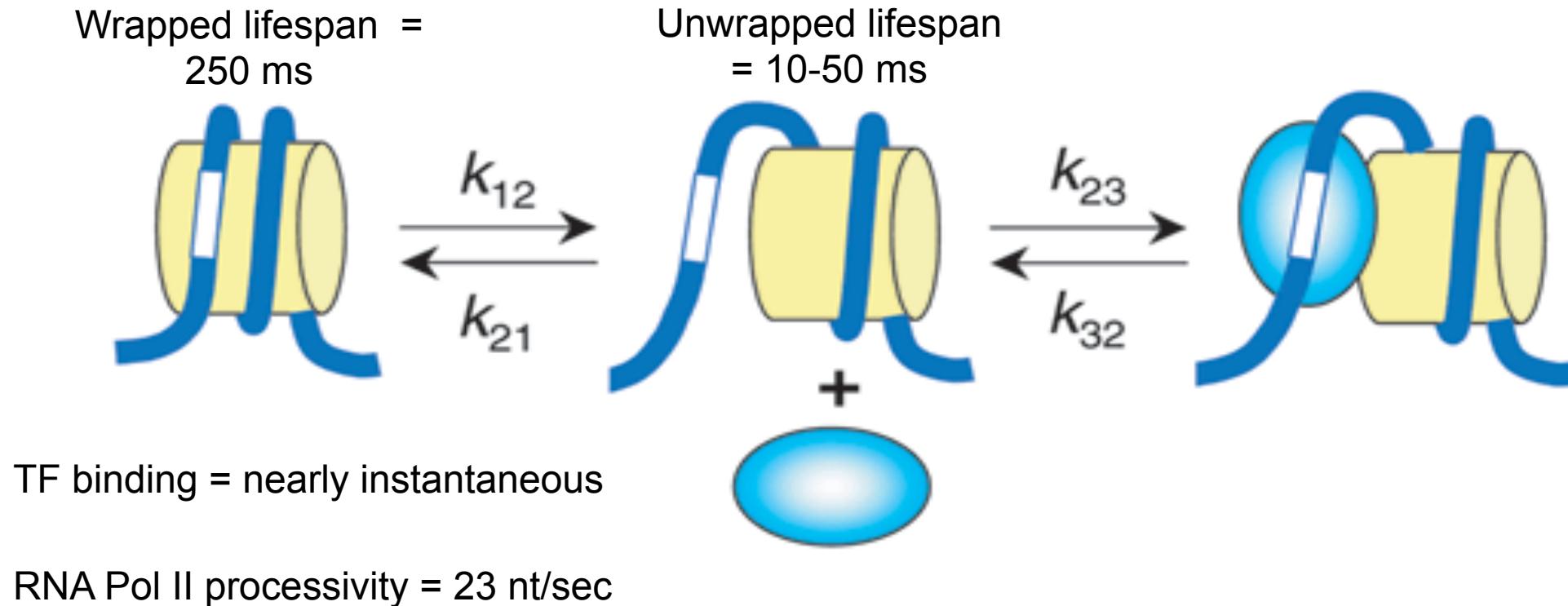
MNase-protected regions



# Rapid in vivo nucleosome turnover across transcribed genes



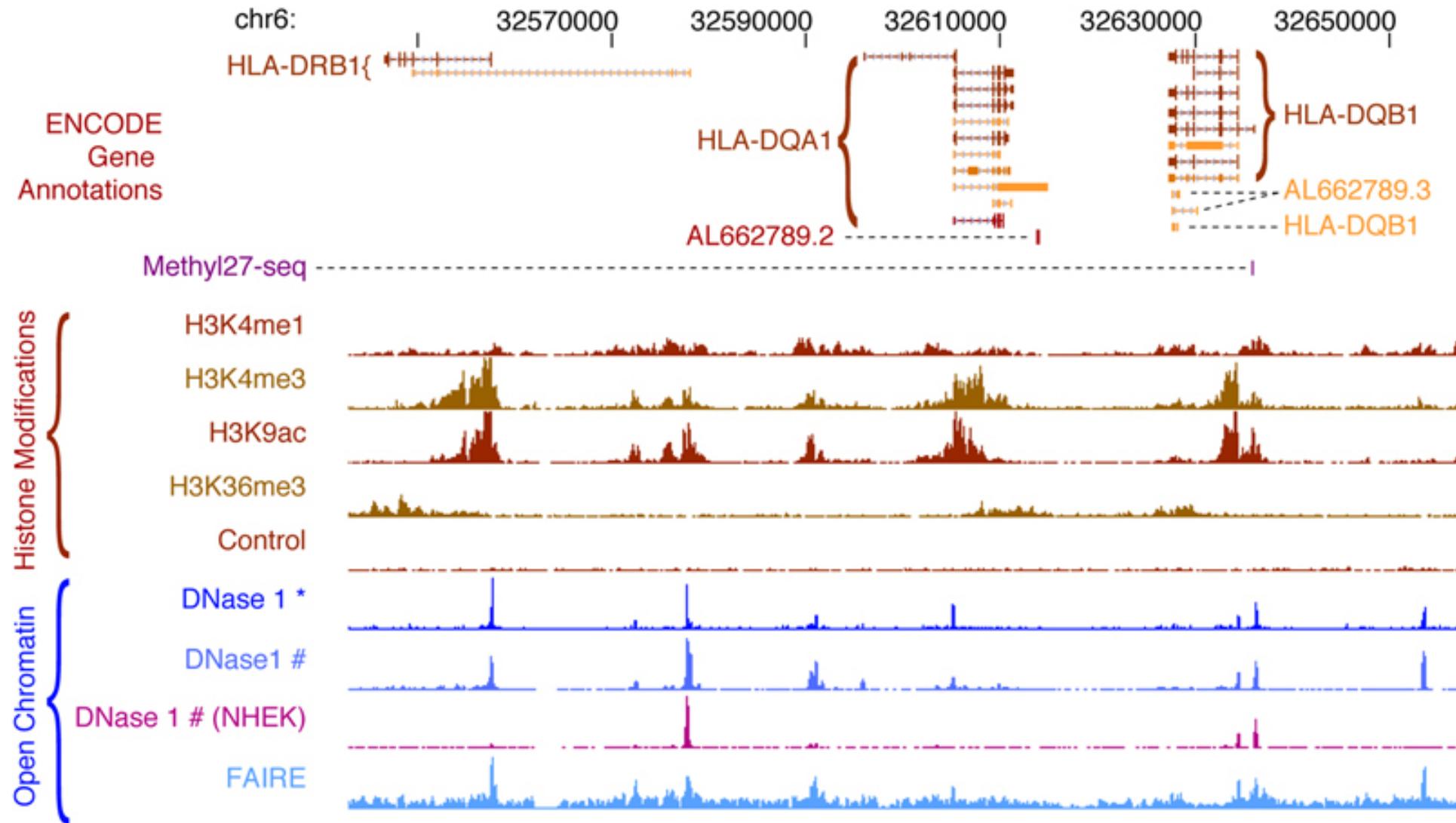
# Protein-DNA interactions are very dynamic and many transition states contaminate the ensemble ChIP preparations



## Implications:

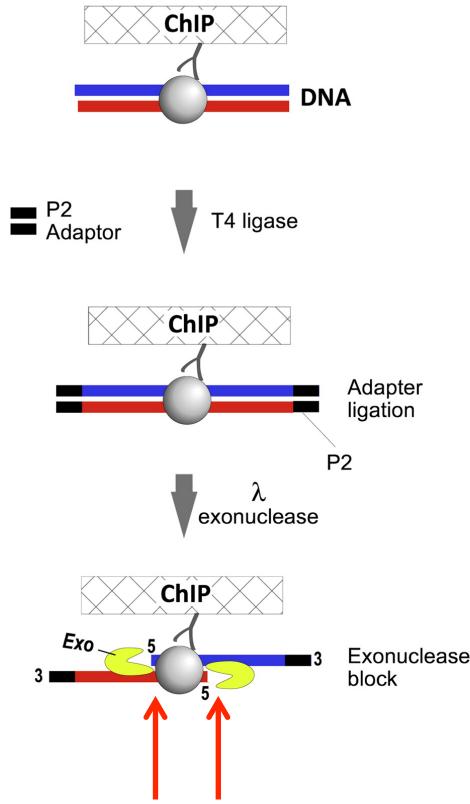
- TFs can passively enter nucleosomes without chromatin remodeling activity
- Major kinetic barrier to gene regulation is fast rewrapping of nucleosomes – RNA Pol II processivity on naked DNA is substantially slower than nucleosome dynamics
- Factors that delay rewrapping (rather than that accelerate unwrapping) would increase processivity

# Accessible Chromatin is Preferentially Solubilized Upon Sonication

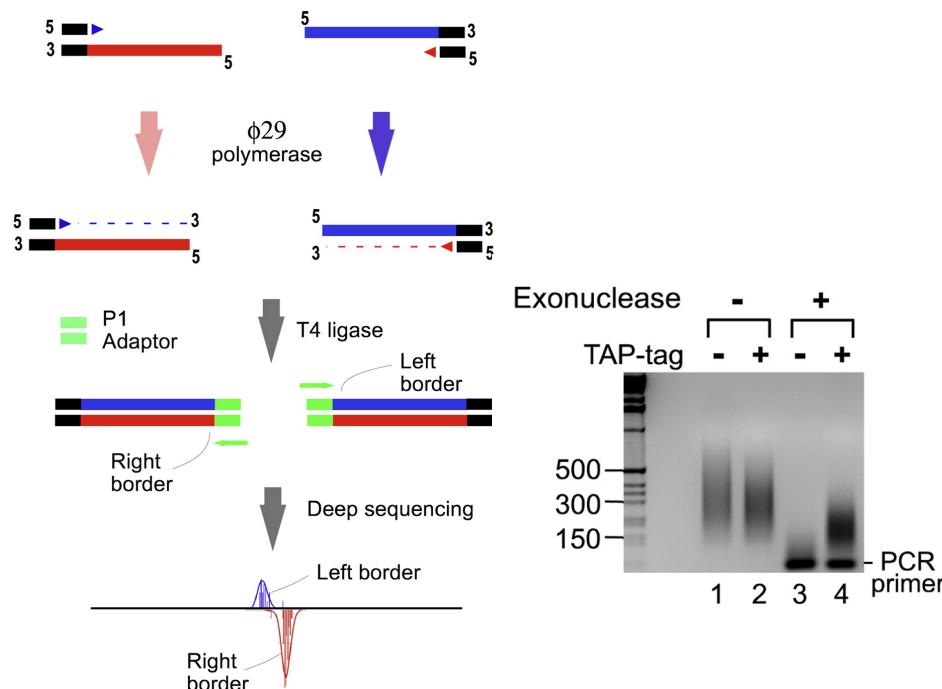


# Protection from exonuclease improves resolution of ChIP-seq

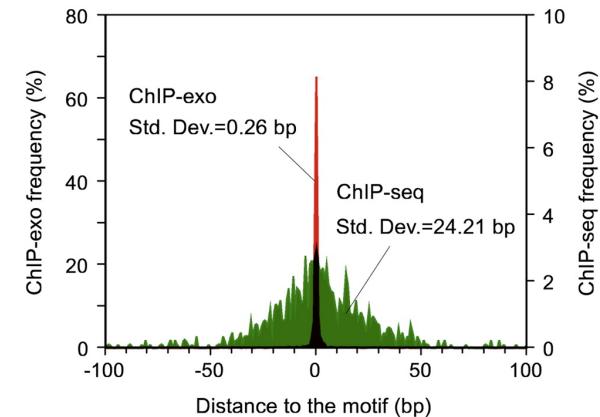
## Cross-link/Sonicate/IP



## Library Preparation

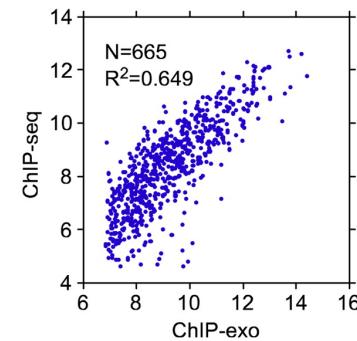


## Computational Analyses

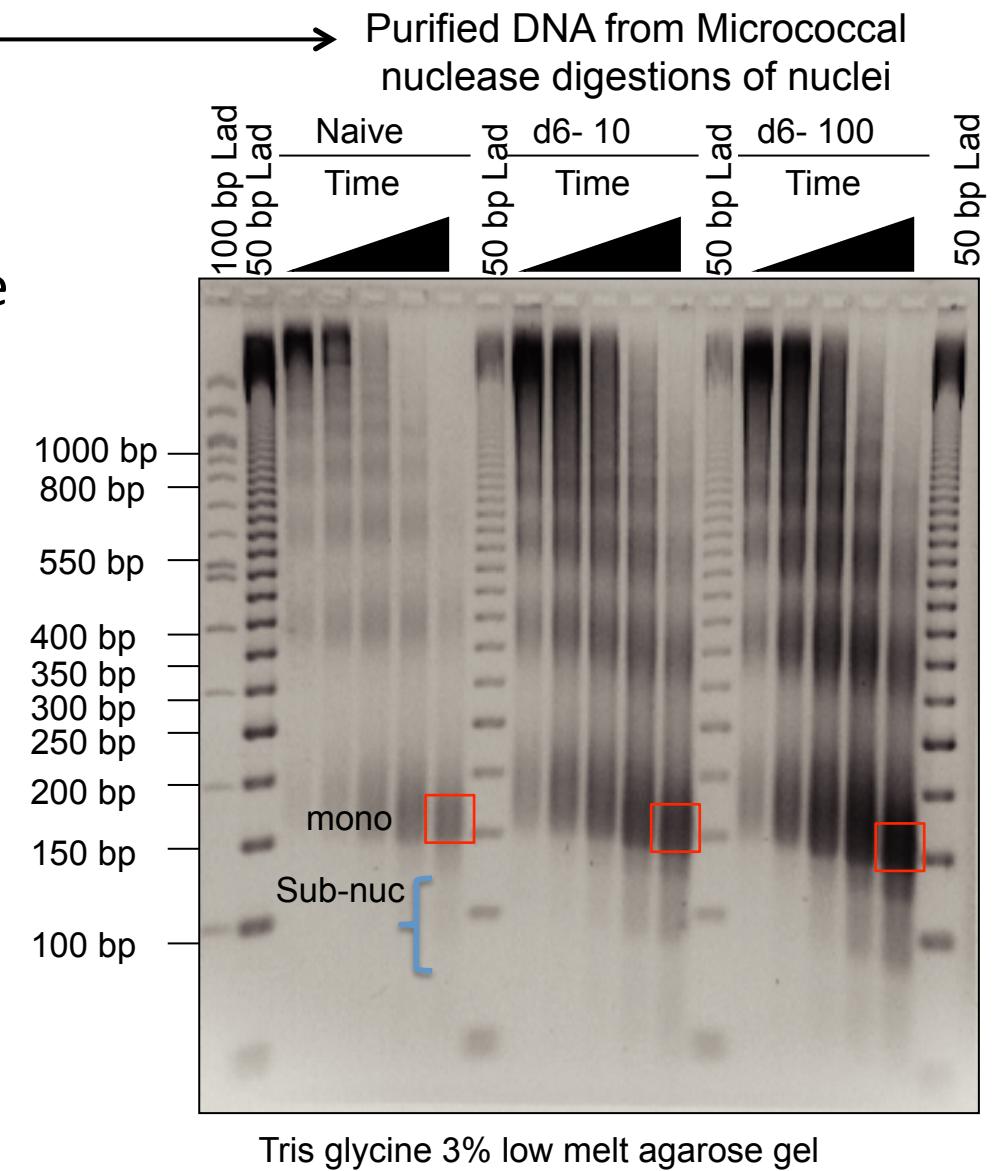
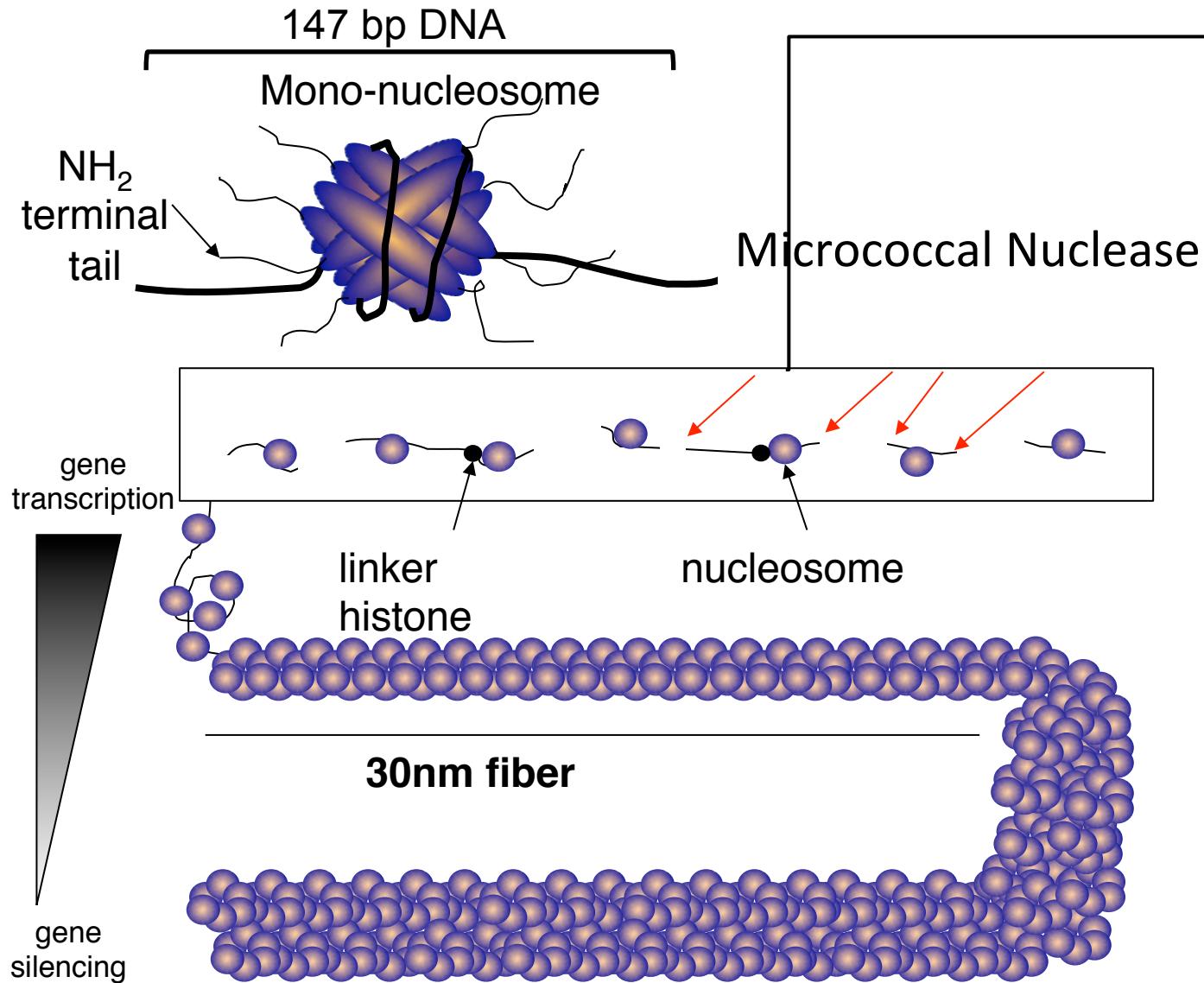


C

Tag counts per overlapped locations between ChIP-seq and ChIP-exo ( $\log_2$ )



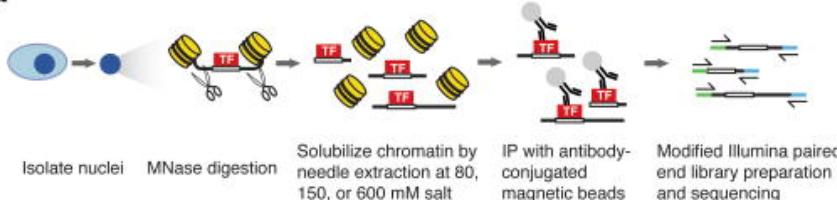
# Fragmentation of chromatin using Micrococcal nuclease (MNase)



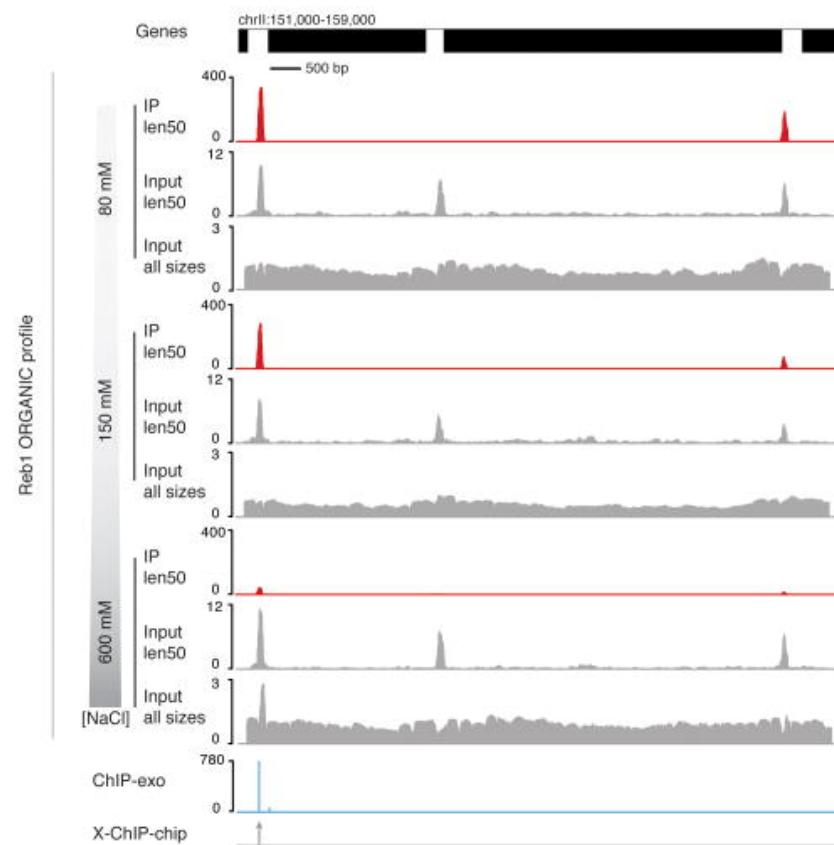
Tris glycine 3% low melt agarose gel

# ORGANIC ChIP enhances signal and spatial resolution of ChIP-seq

a



b

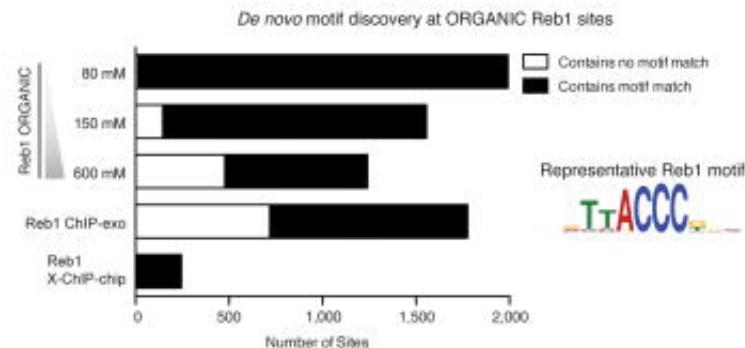


ORGANIC TF binding sites have characteristic binding site motifs

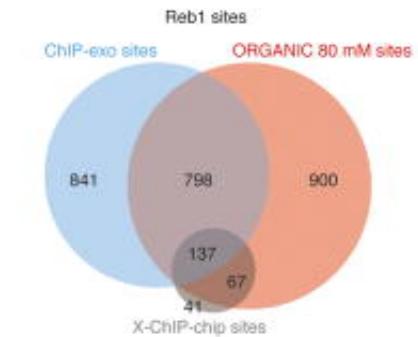
And

Massive increase in the number of binding sites over conventional ChIP

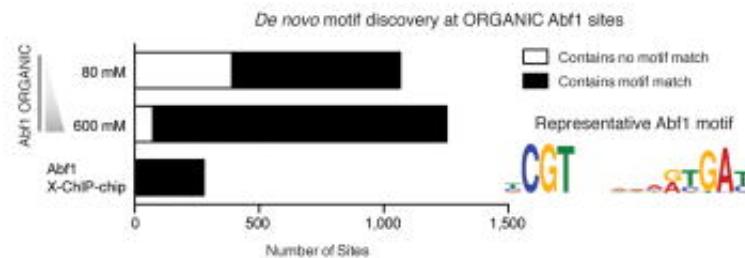
a



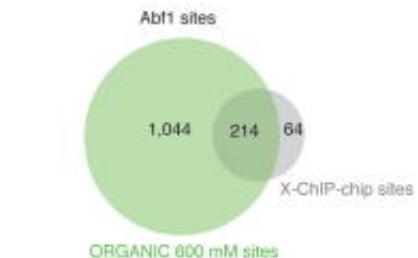
c



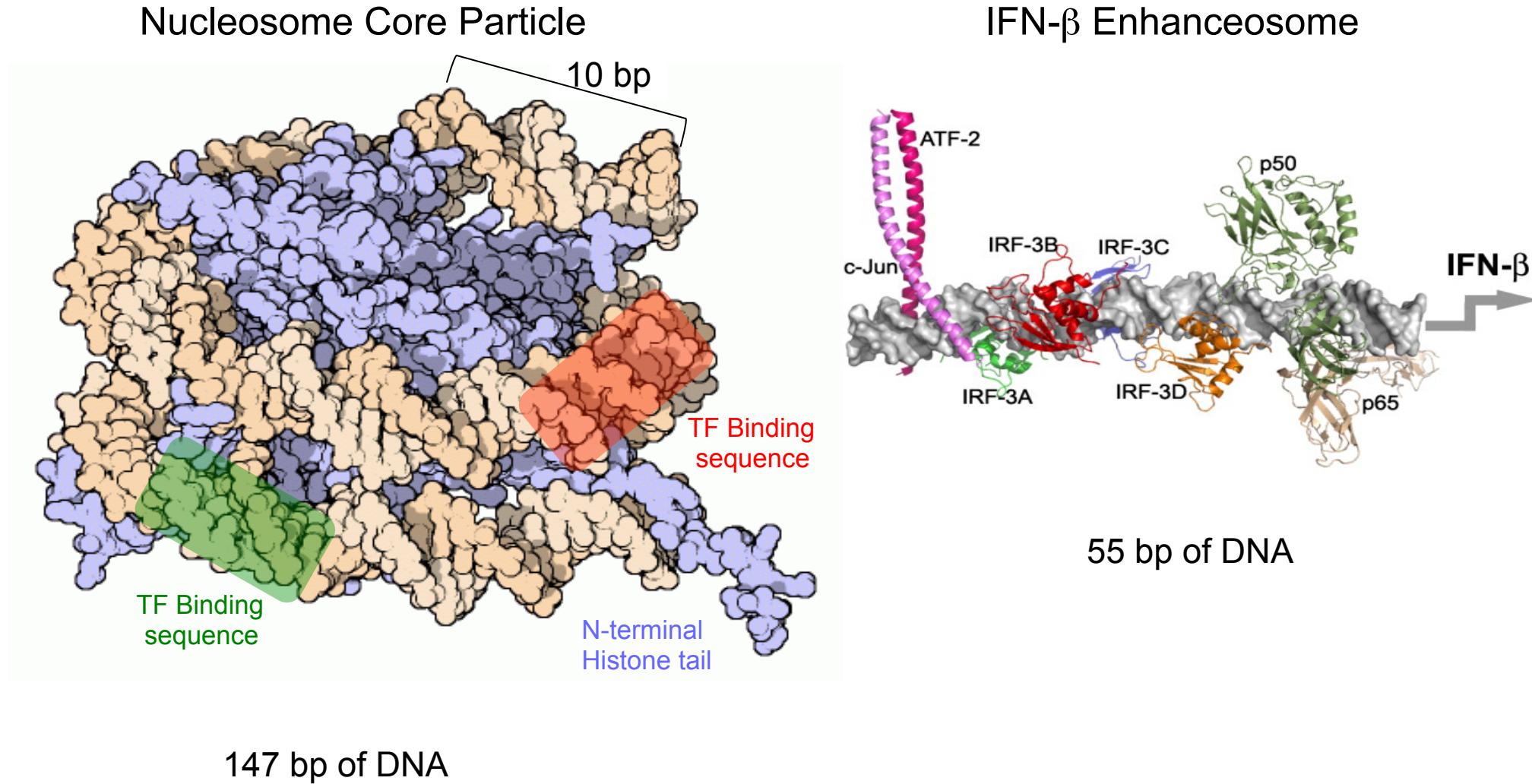
b



d



# Nucleosomes occlude the genome and their positioning at regulatory DNA is critical for gene transcription



The genome is organized in topologically associated domains (TADs) that coordinate developmentally regulated transcription

