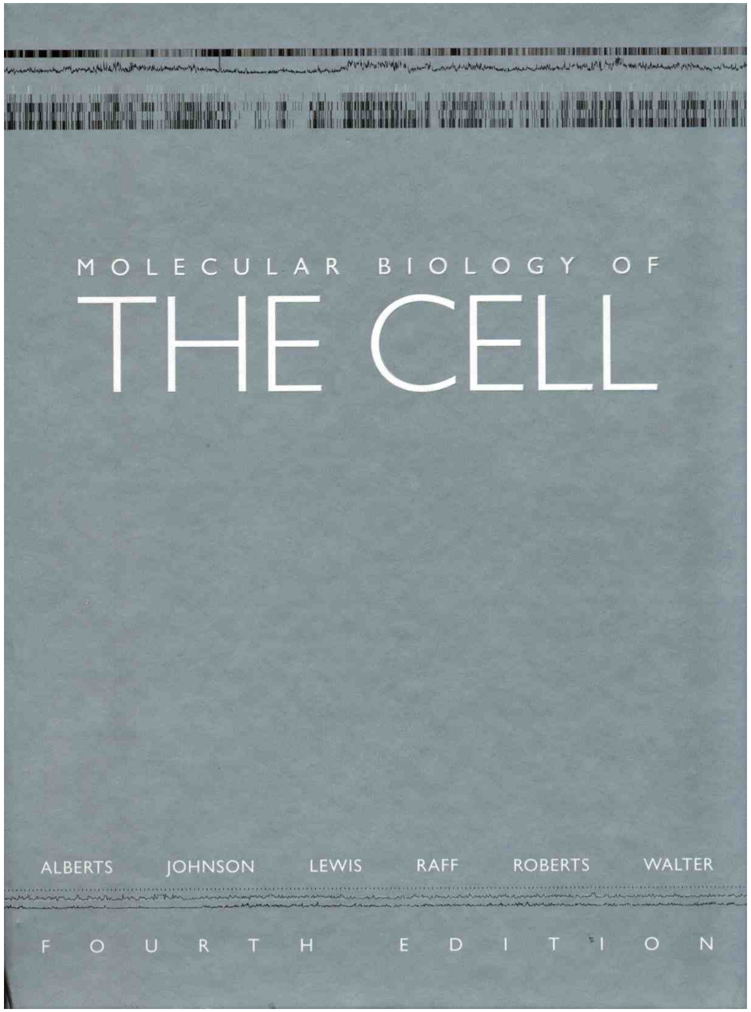
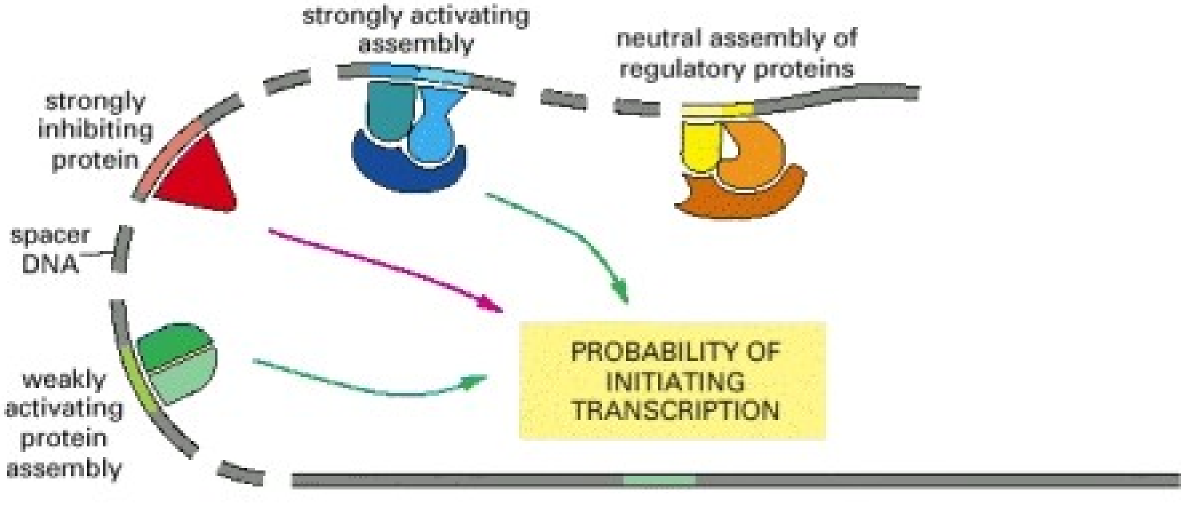
Widespread transcription at activity-dependent neuronal enhancers

Martin Hemberg

UC Berkeley
February 28, 2012

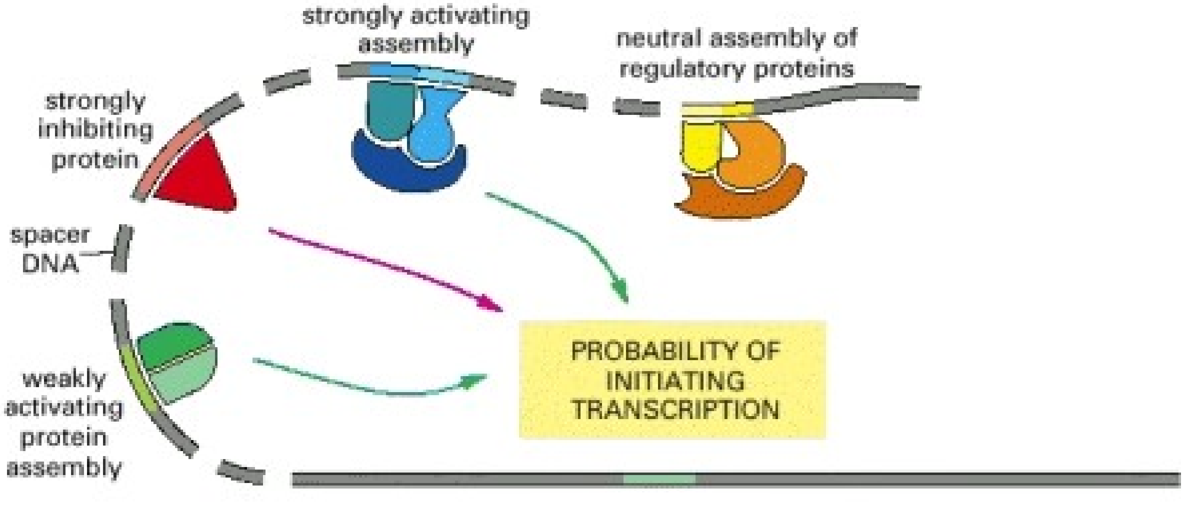
Textbook view of gene regulation





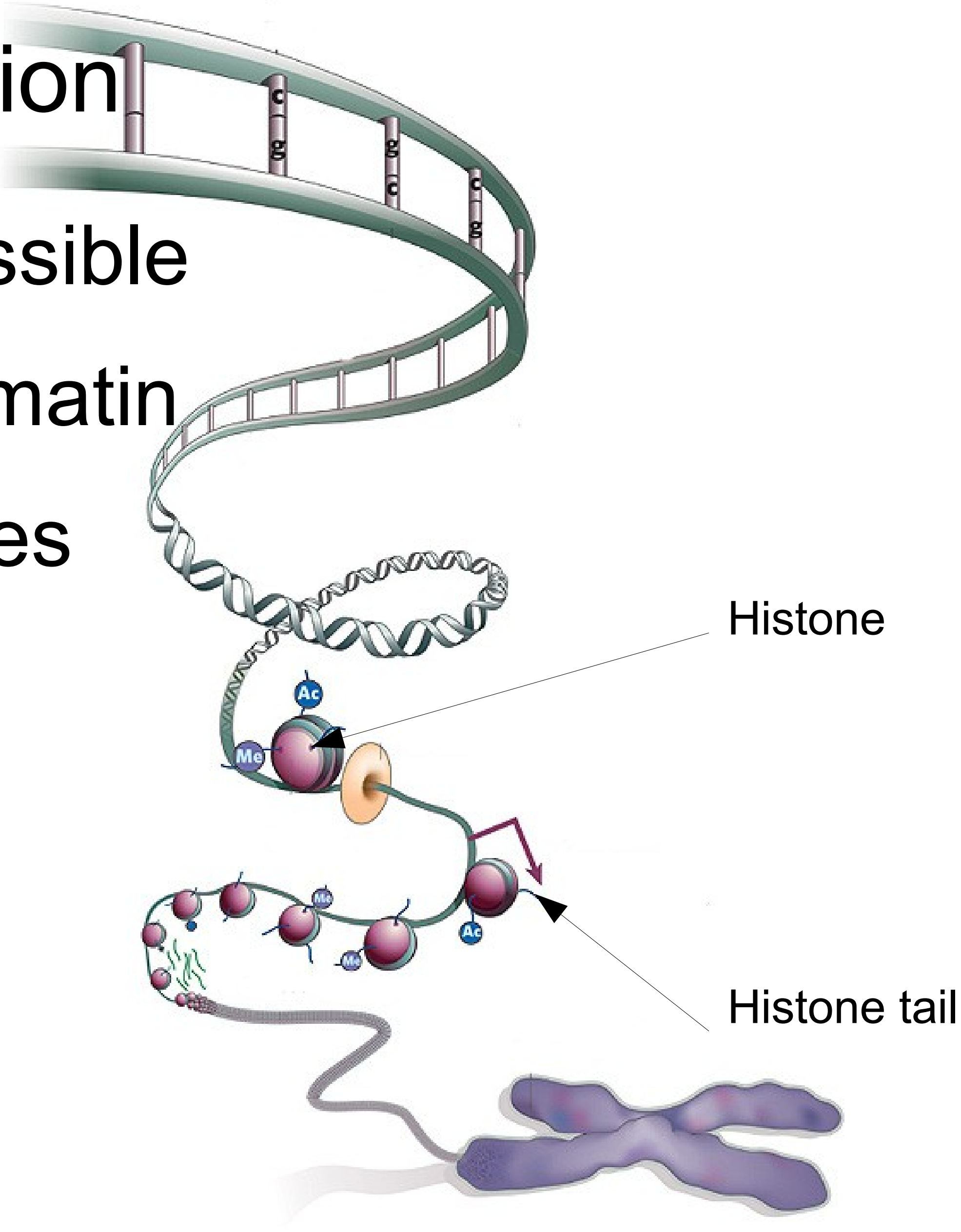
Transcription Factors (**TFs**) bind to DNA motifs

ACCTGACGTCAAACGTTTAA



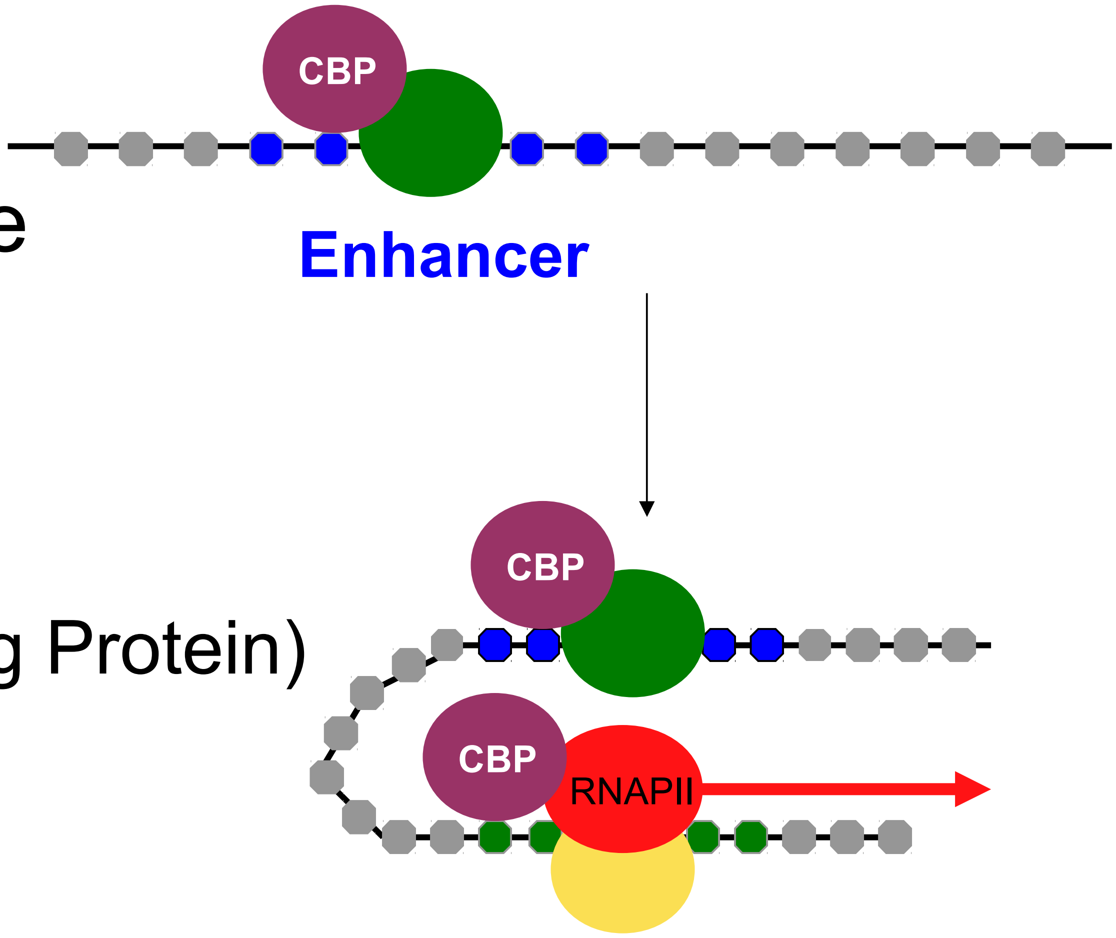
Post-translational modifications of histone tails correlate with function

- ~100 k loci or 1% accessible
- **H3K4Me1** – open chromatin
- **H3K4Me3** – active genes



(ENCODE, 2007)

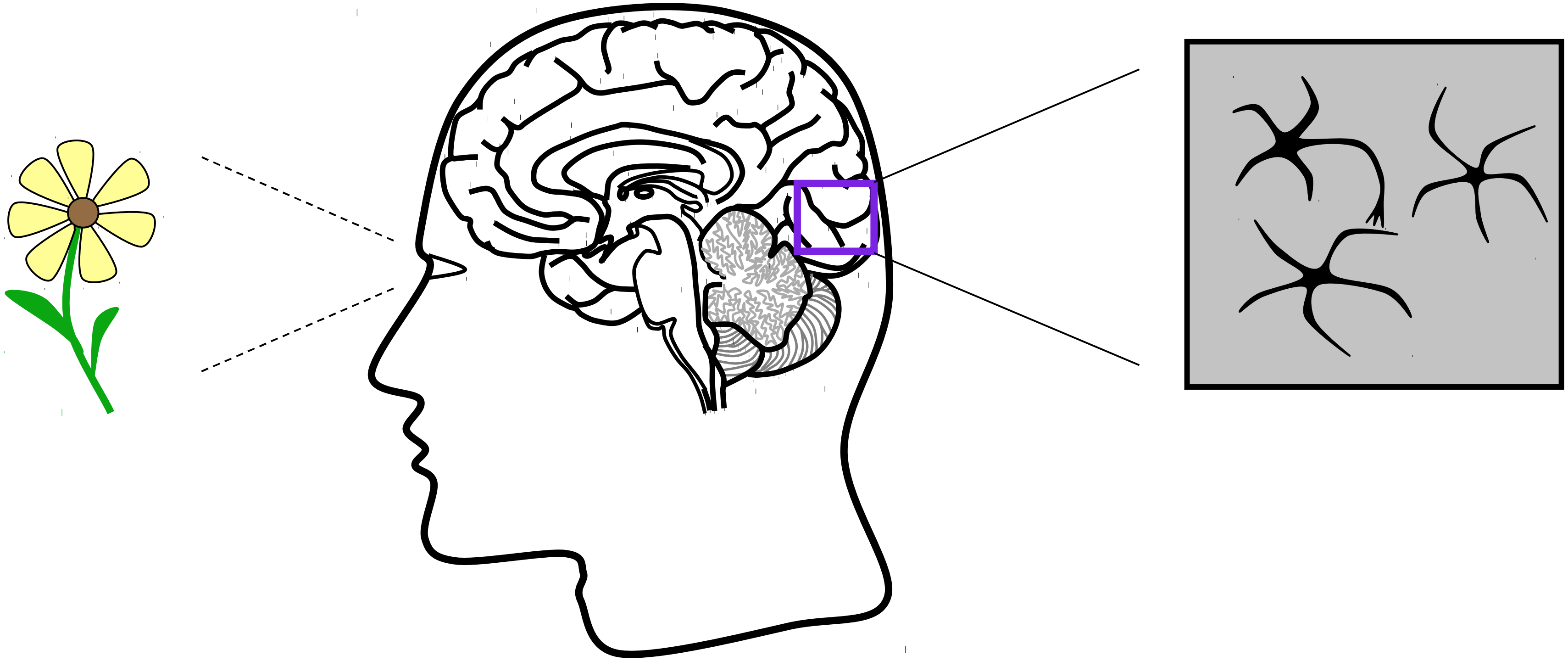
Enhancers are distal TF binding sites



- No universal sequence signature
- Up to 1Mb away
- Marked by
  - **CBP (**Creb Binding Protein)
  - H3K4me1 flanking
  - H3K4me3 absent
- Cell-type specific

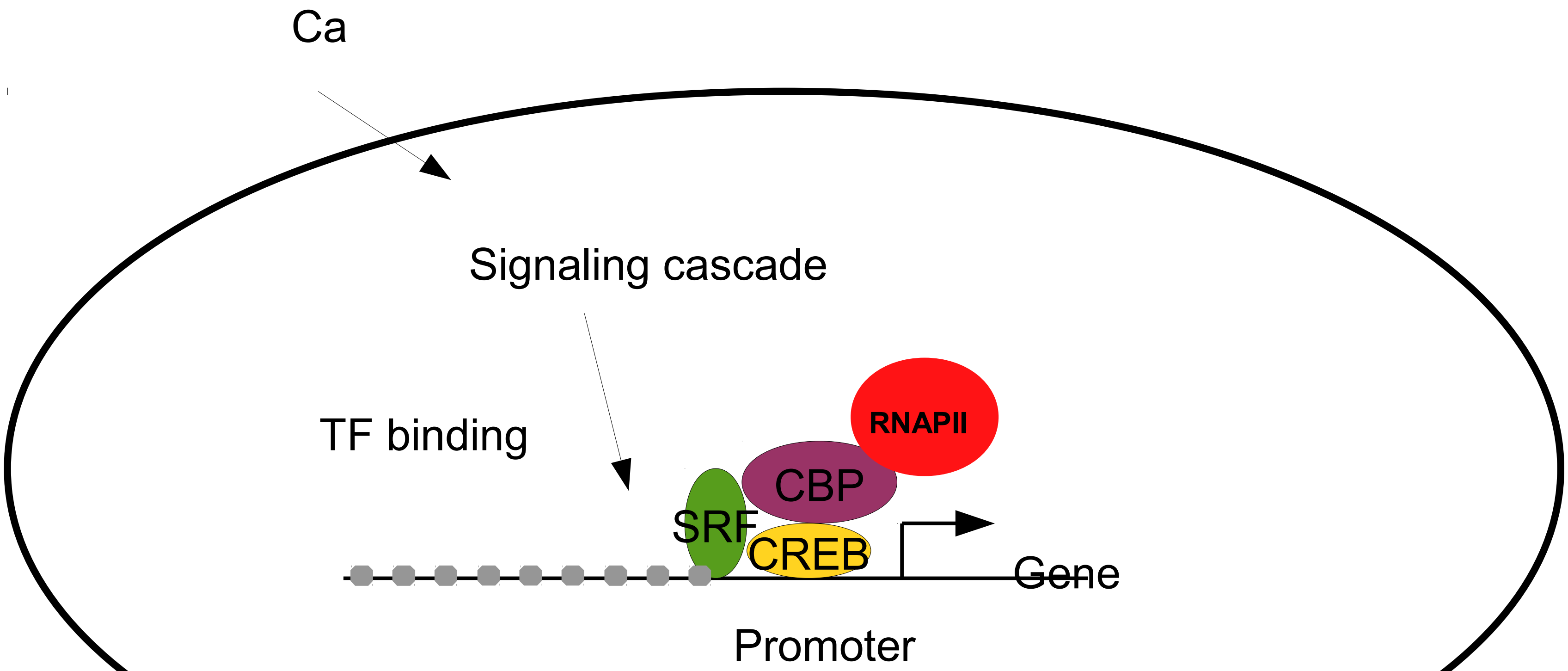
ENCODE, 2007
Heintzman et al, 2007
Roh et al, 2005
Visel et al, 2009

External stimuli change synapses

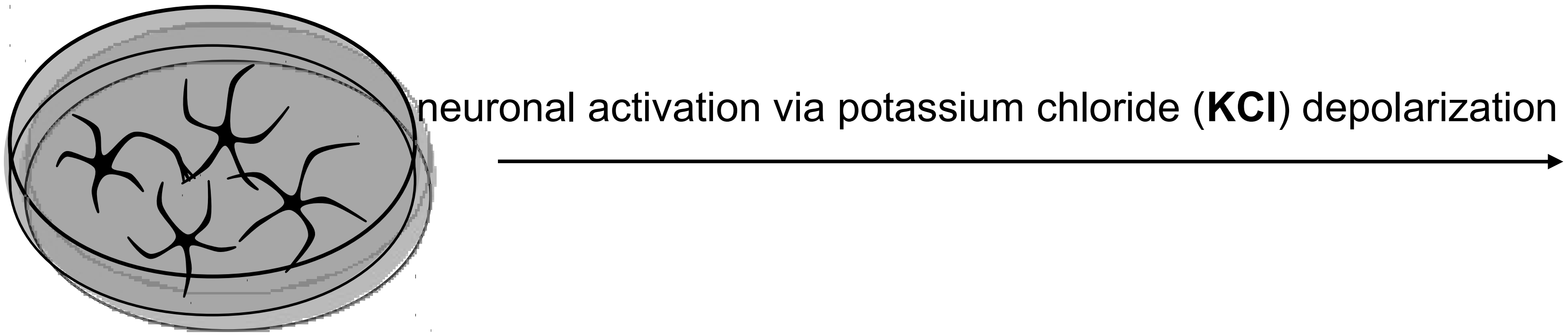


Hubel & Wiesel, 1970’s

Changes in synapses are driven by changes in gene expression



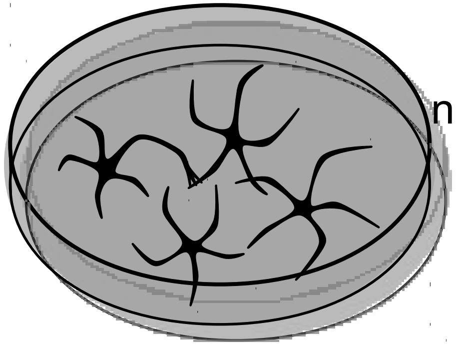
An experimental system for genome-wide study of activity dependent gene expression



neuronal activation via potassium chloride (**KCl**) depolarization

mouse cortical
neurons

An experimental system for genome-wide study of activity dependent gene expression



mouse cortical
neurons

neuronal activation via potassium chloride (**KCl**) depolarization

- KCl

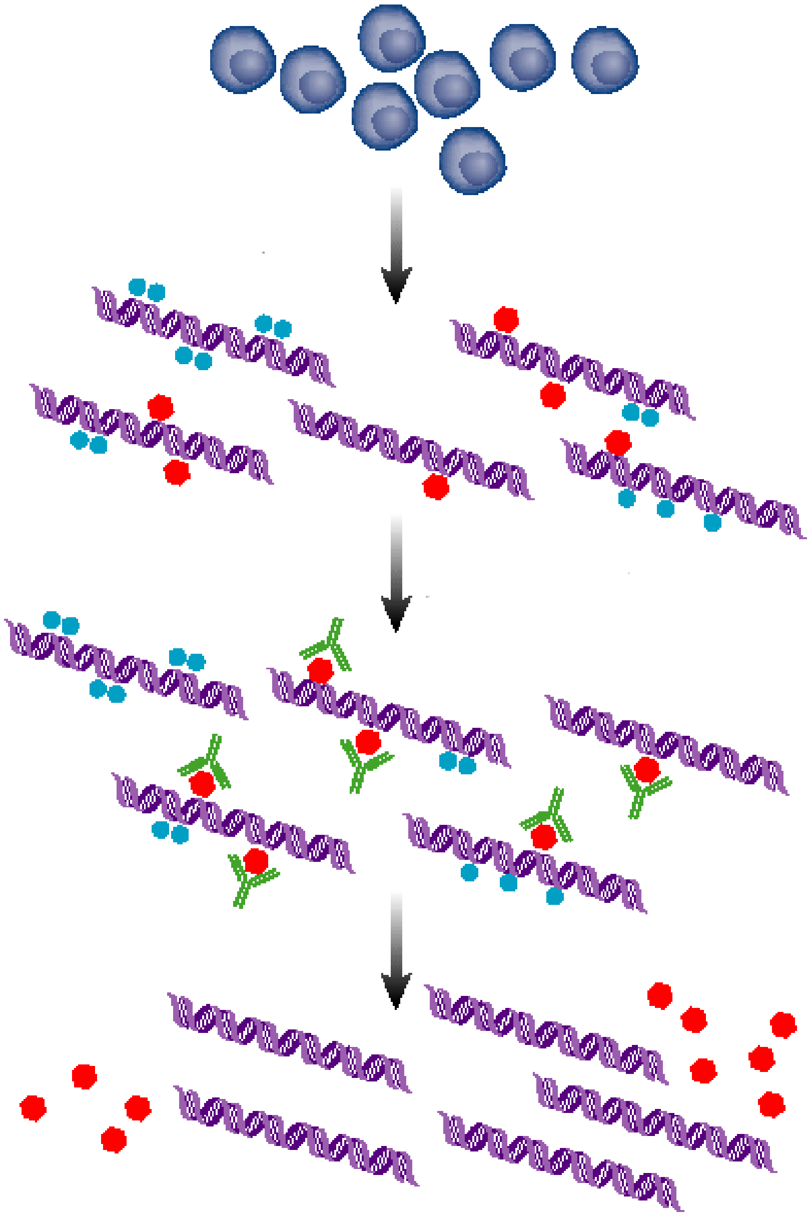
ChIP-Seq
RNA-Seq

+ KCl

ChIP-Seq
RNA-Seq

Jesse Gray
Tae-Kyung Kim
Greenberg Lab

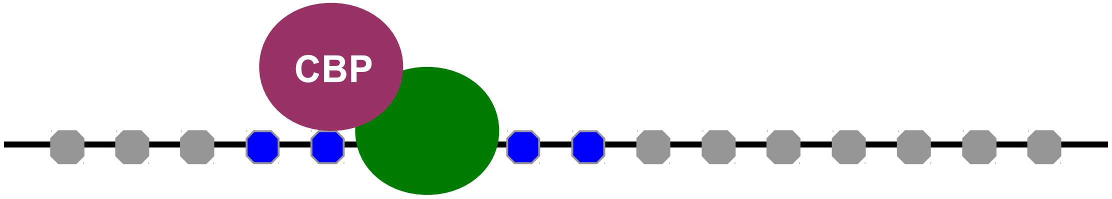
Chromatin immunoprecipitation and sequencing (**ChIP-Seq**) finds protein binding sites *in vivo*



(Mardis, 2007)

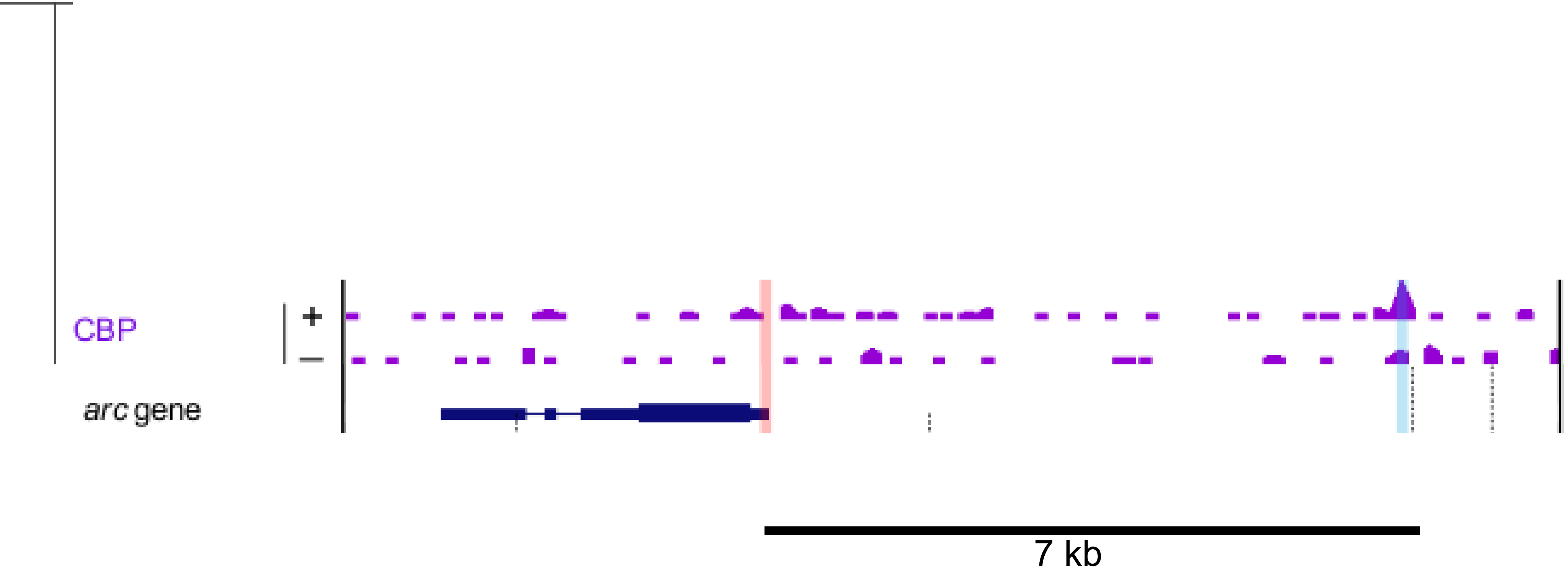
Where are activity-dependent enhancers located?

- **CBP** binding
- **H3K4me1** flanking
- **H3K4 me3** absent

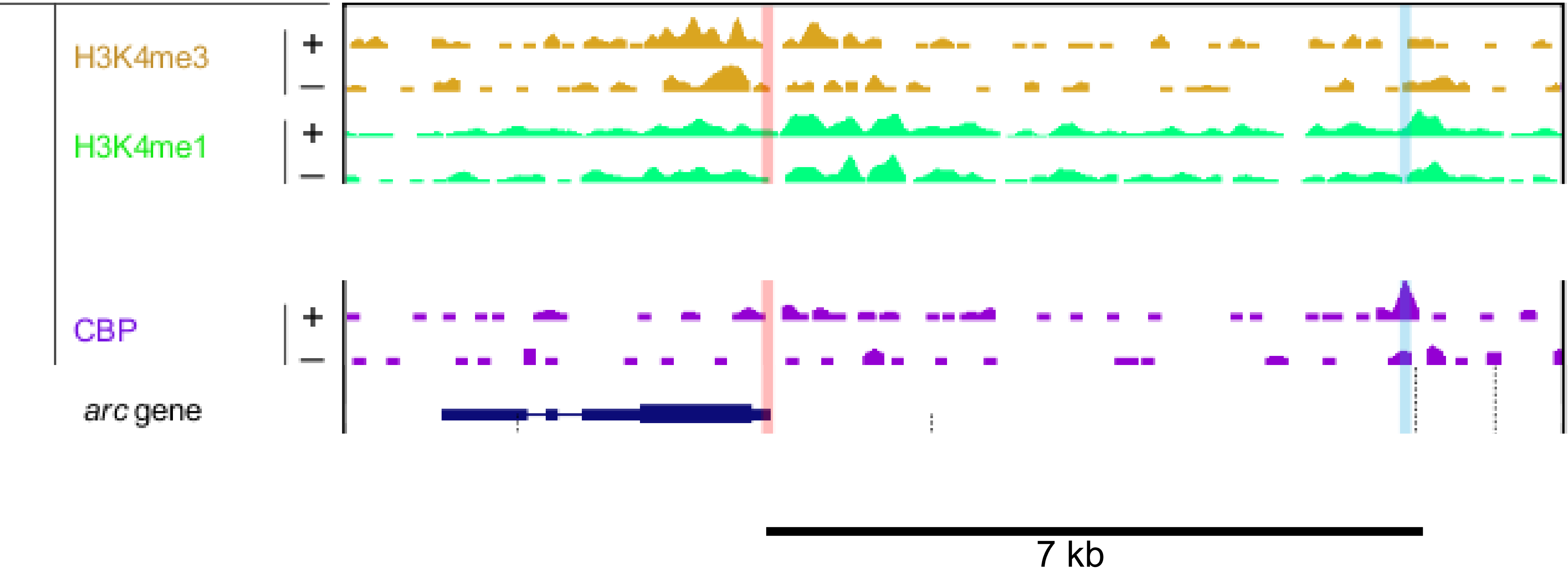


Enhancer
H3K4me1

Activity-regulated enhancer at the Arc-locus has inducible CBP binding

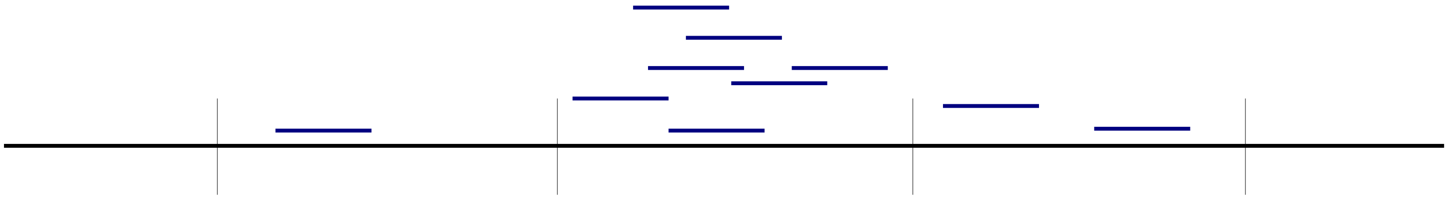


Distal CBP peaks have high levels of H3K4me1 and low levels of H3K4me3

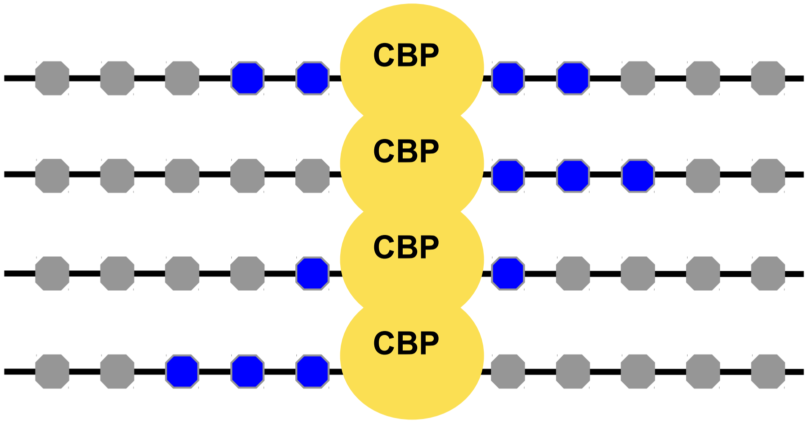


Identifying 28,000 CBP binding sites

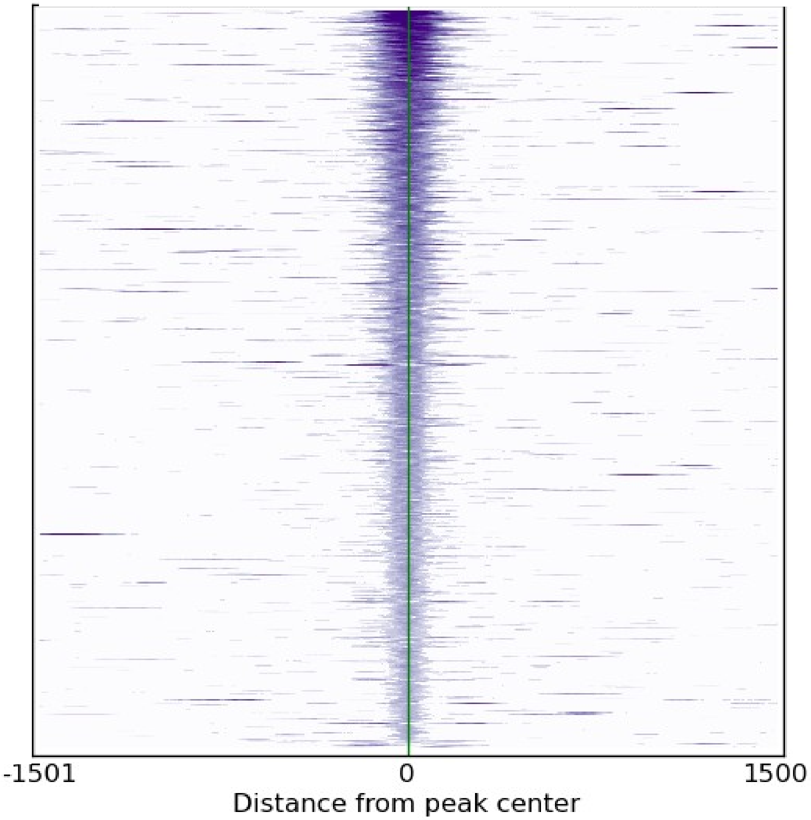
- Regions that have significantly more CBP than background

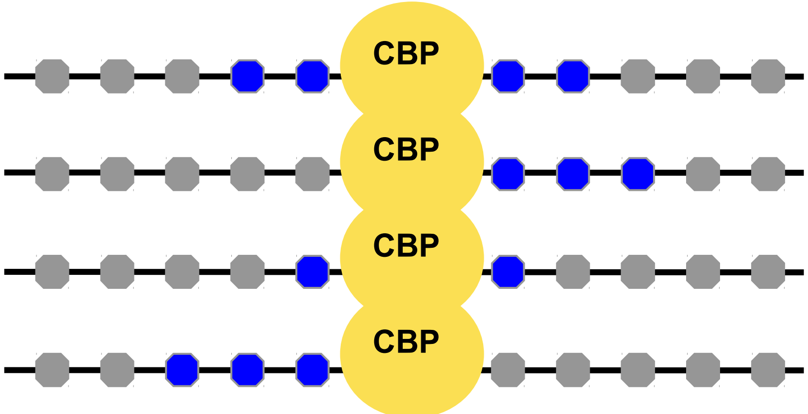


Aligning CBP peaks to calculate binding profiles

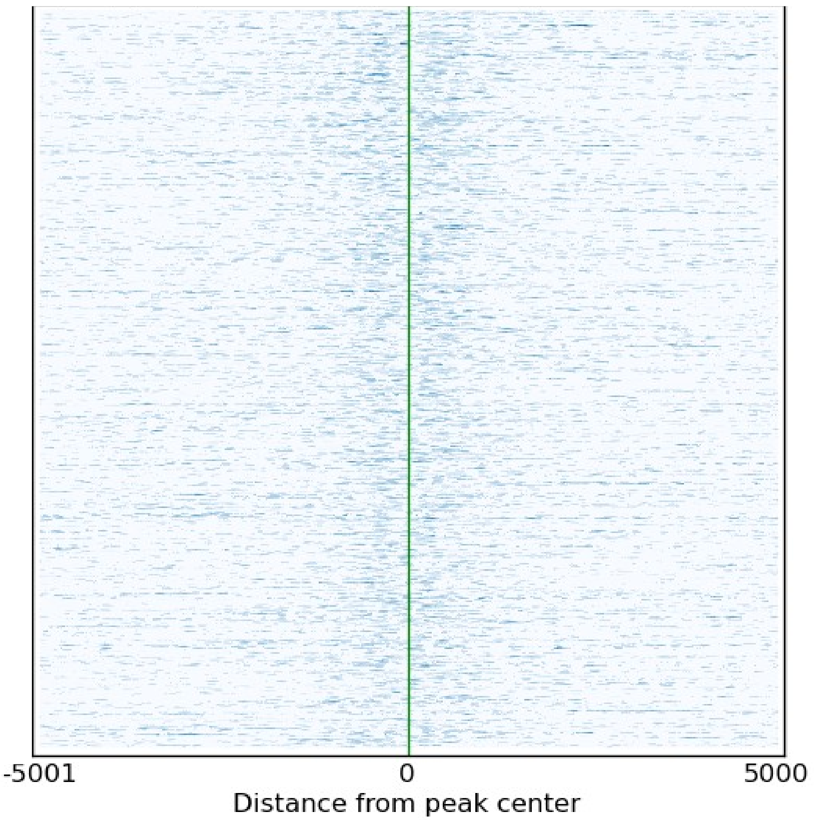


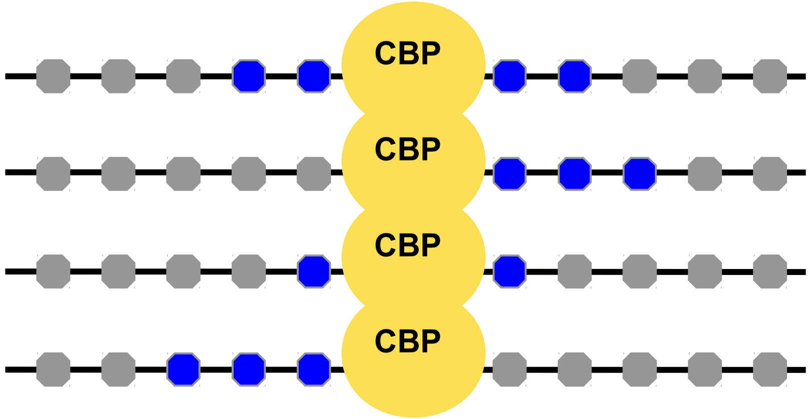
Aligning CBP peaks to calculate binding profiles



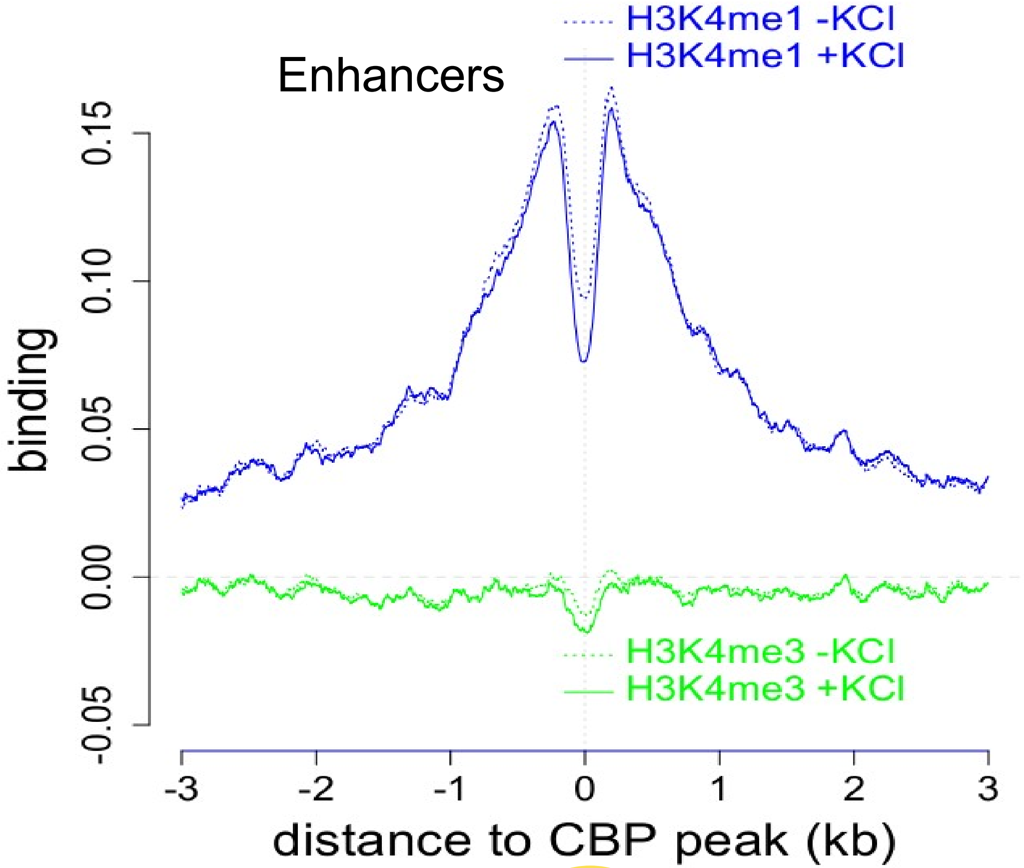


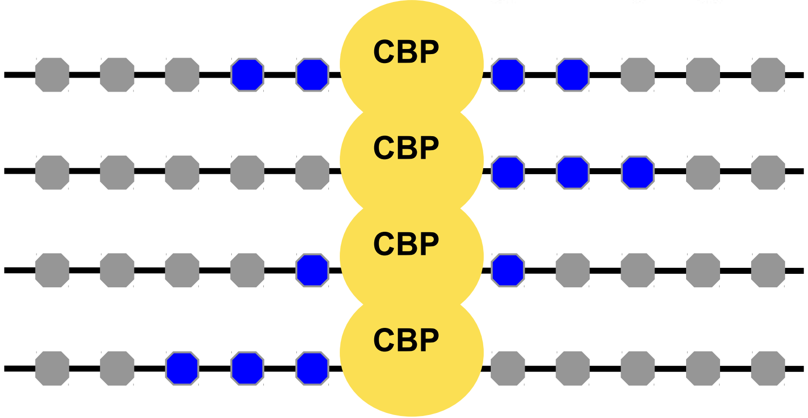
Aligning CBP peaks to calculate H3K4me1 binding profiles



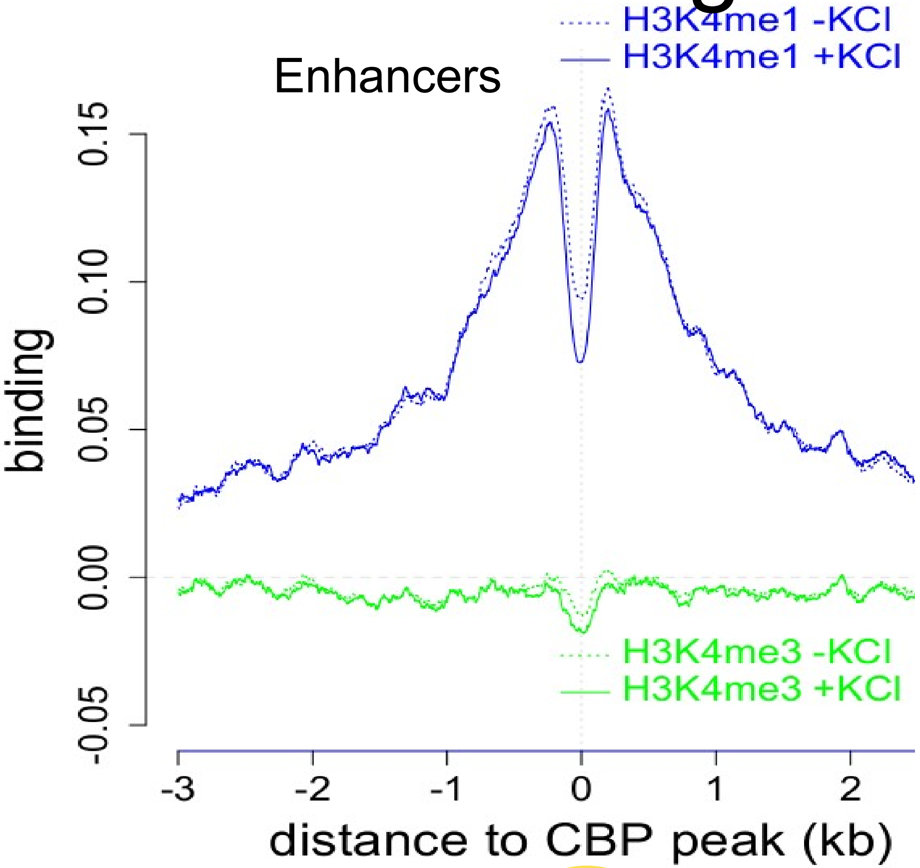


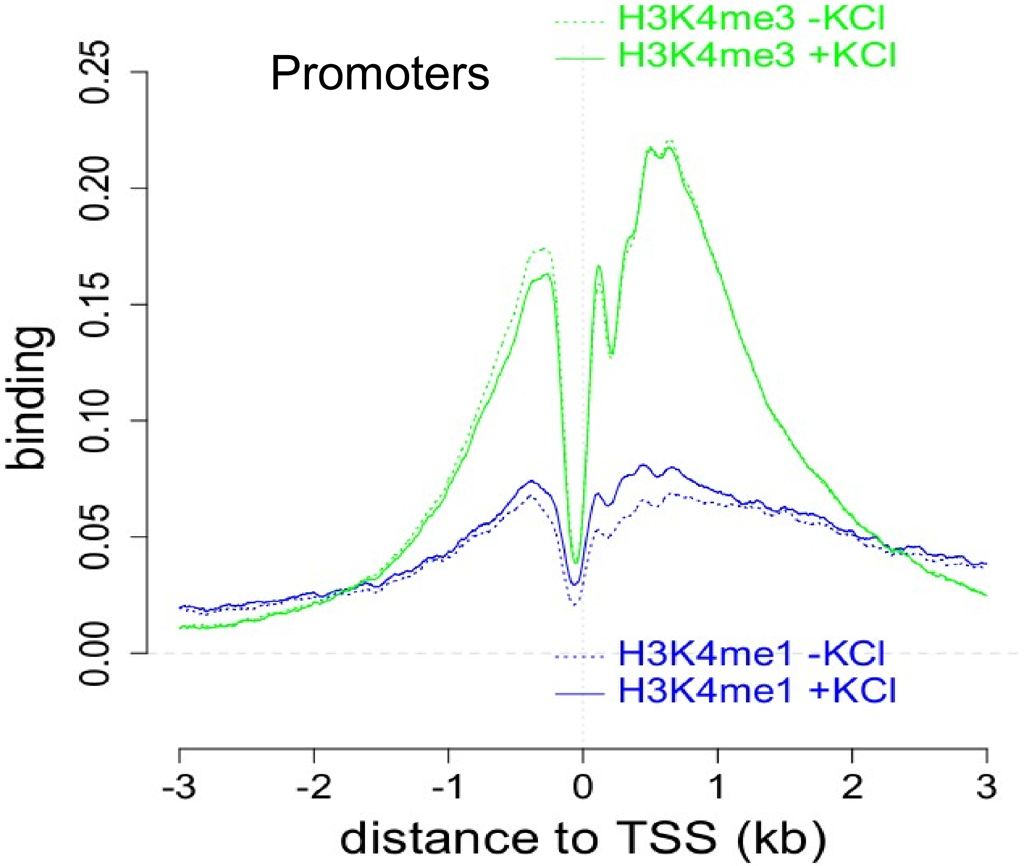
Enhancers have high levels of H3K4me1 and low levels of H3K4me3

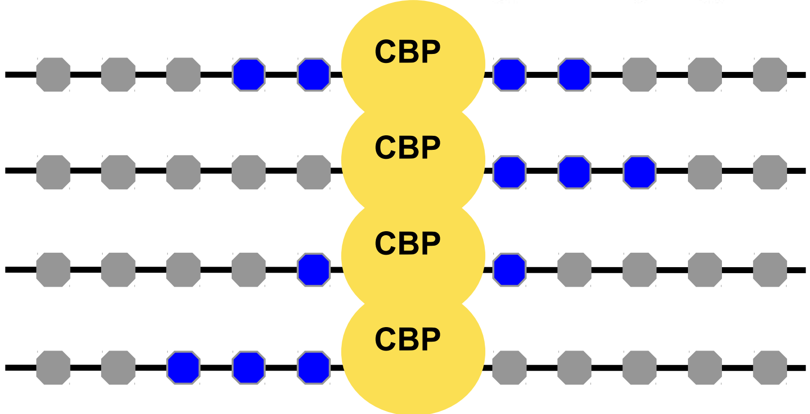




Transcription start sites have high levels of H3K4me1 and high levels of H3K4me3





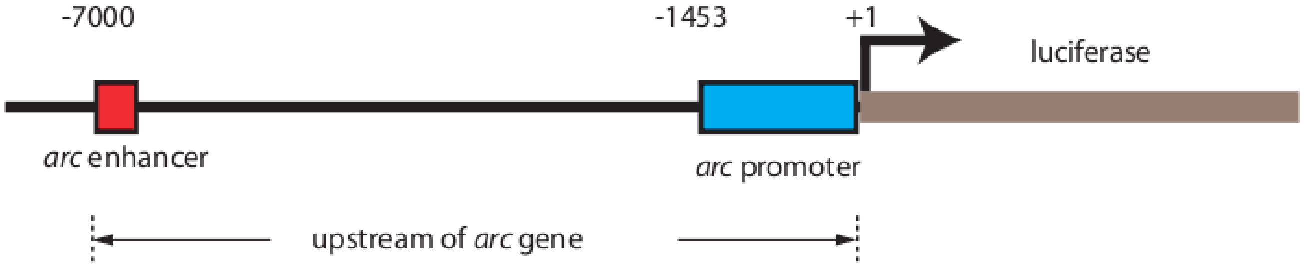


We identified 12k activity-dependent enhancers throughout the genome

- **CBP** binding
- **H3K4me1** flanking
- **H3K4me3** absent
    - ~**5000** extragenic enhancers
    - ~7000 intragenic enhancers

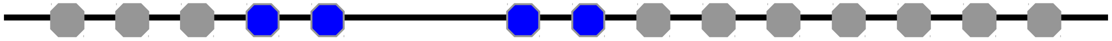
8/8 tested activity-dependent enhancers were validated using a luciferase assay

- **CBP** peak
- **High** levels of flanking **H3K4me1**
- **Low** levels of **H3K4me3**
  - **~5000** extragenic enhancers
  - ~7000 intragenic enhancers



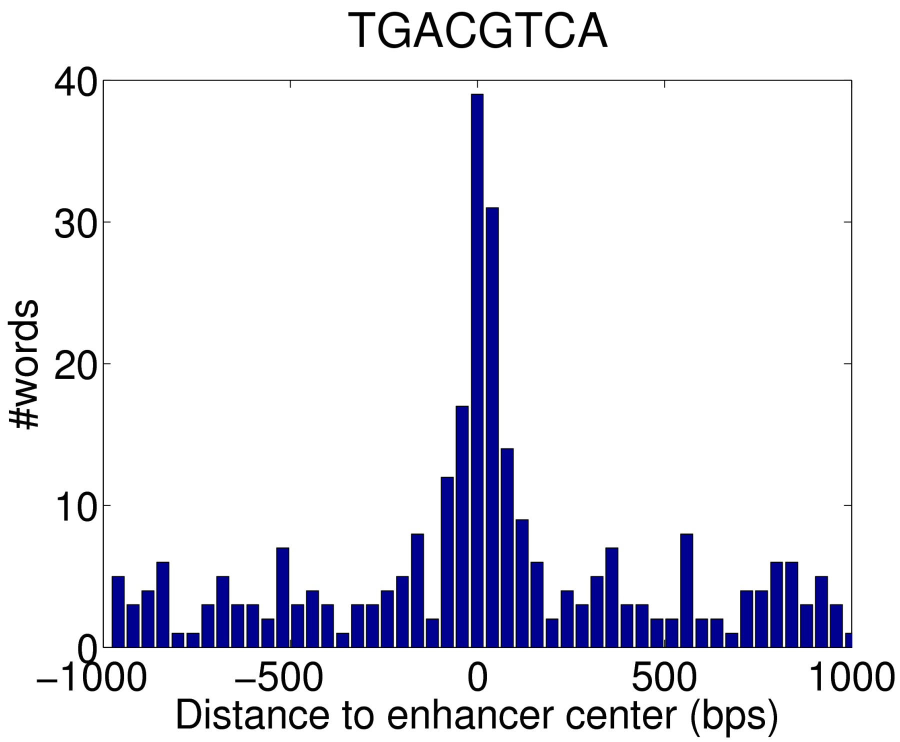
What mechanisms can we identify for enhancers in gene regulation?

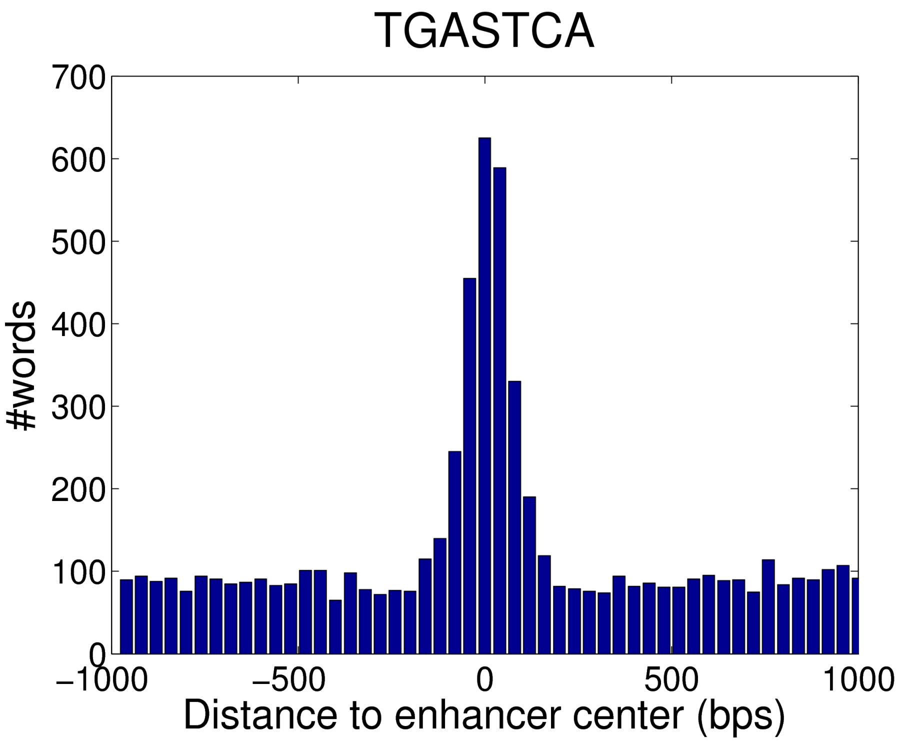
- Accessible chromatin



What motifs are enriched at enhancers?

- ~2000 words of known regulatory significance
- Calculate enrichment relative to flanks

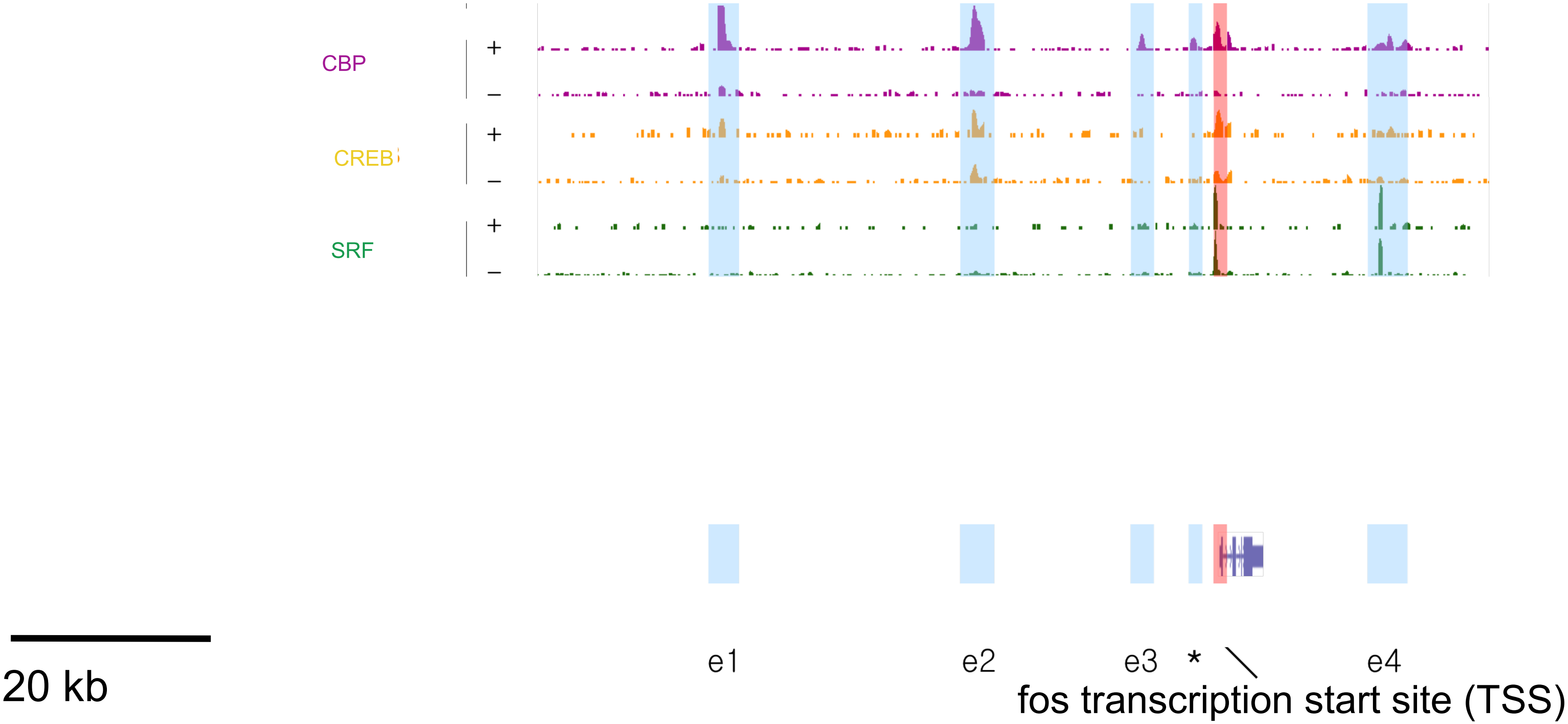

TGACGTCA


TGASTCA

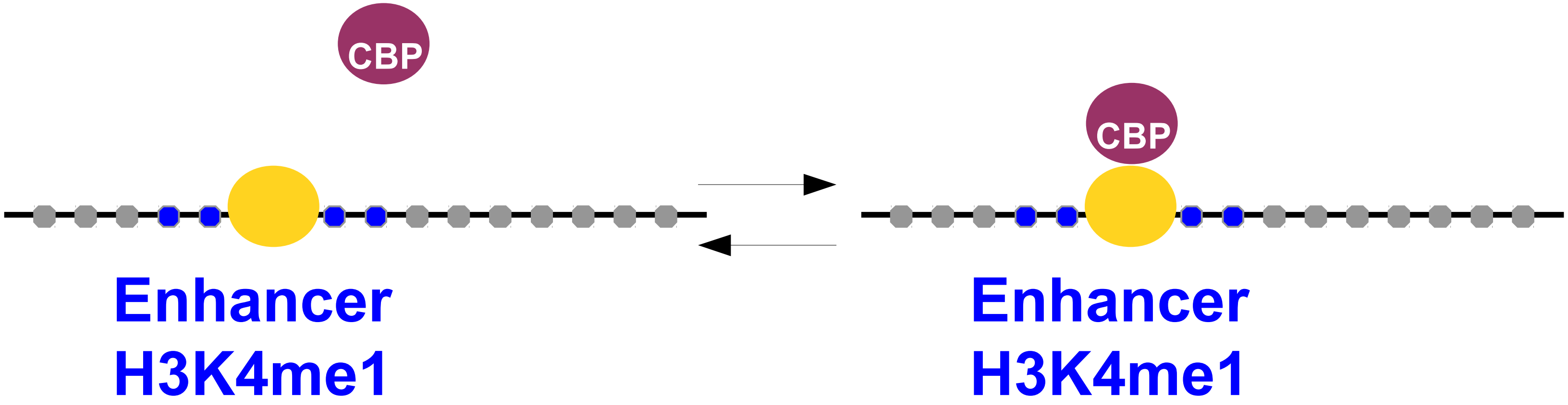
~100 enriched motifs are found

Word	**Enrichment**	**Known TF**
TGASTCA	4.74	Fos/Jun
TGACGTCA	6.41	Creb
CTAWWWATA	3.34	Srf
TCGTG	1.56	Npas4
CTGCCAAA	3.34	?

SRF and CREB binding at Fos enhancers

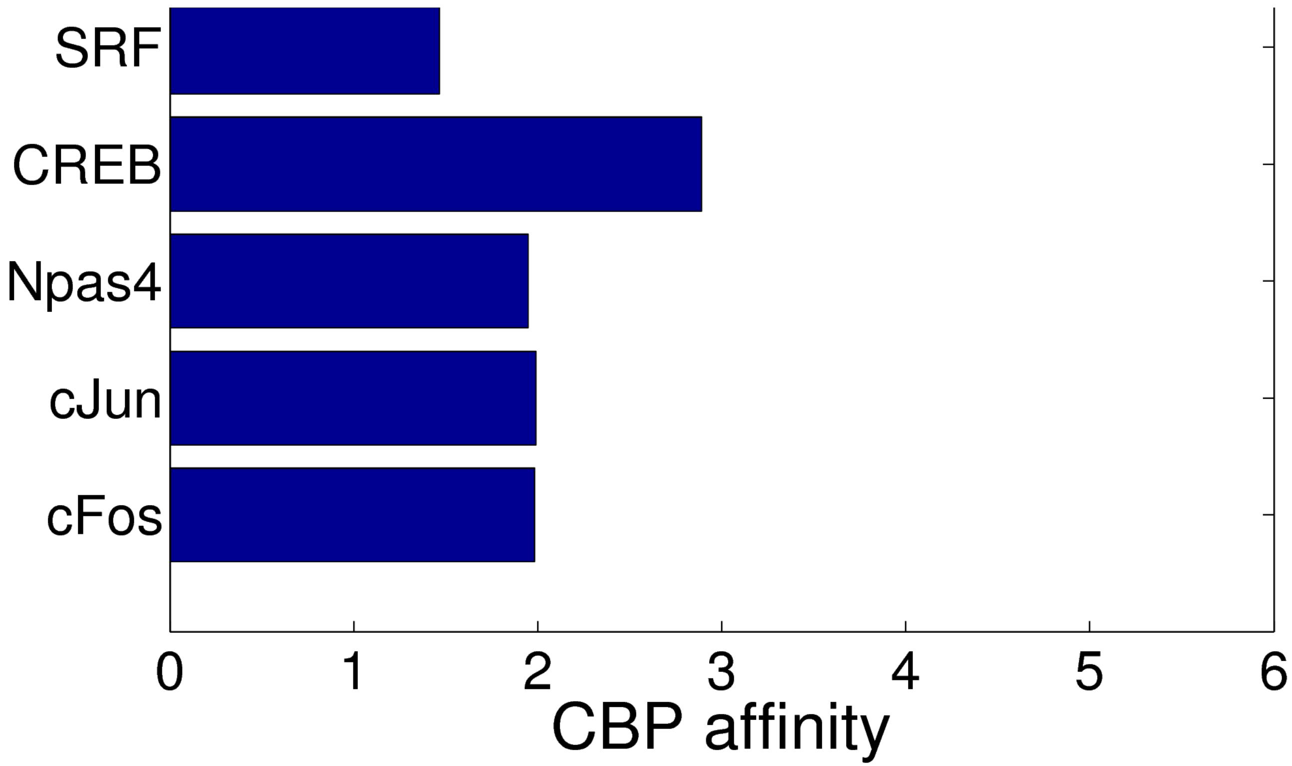


Is CBP binding determined by other TFs?

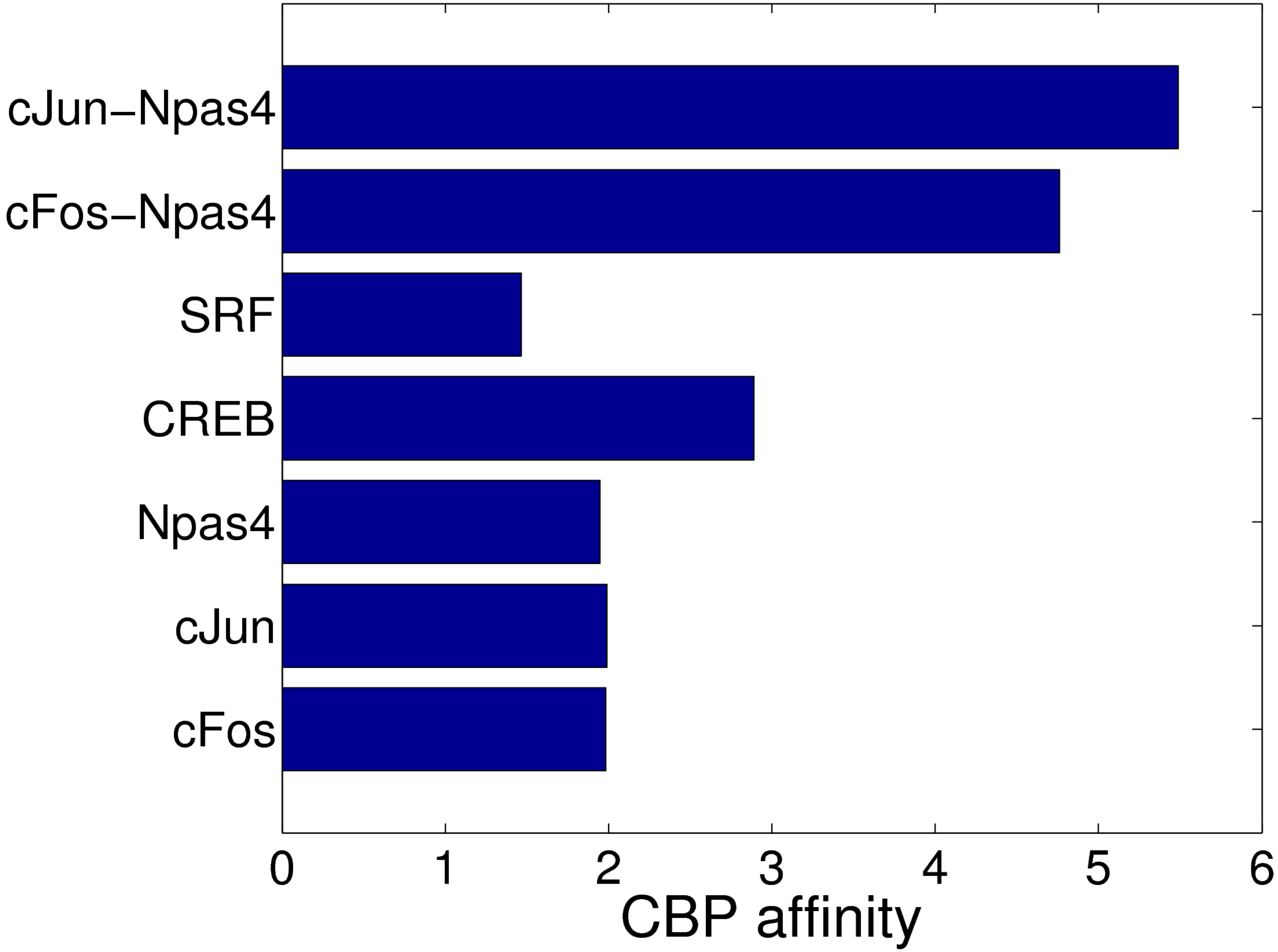


- Combinatorial regulation
  - Mechanisms unclear
  - CBP bottleneck?

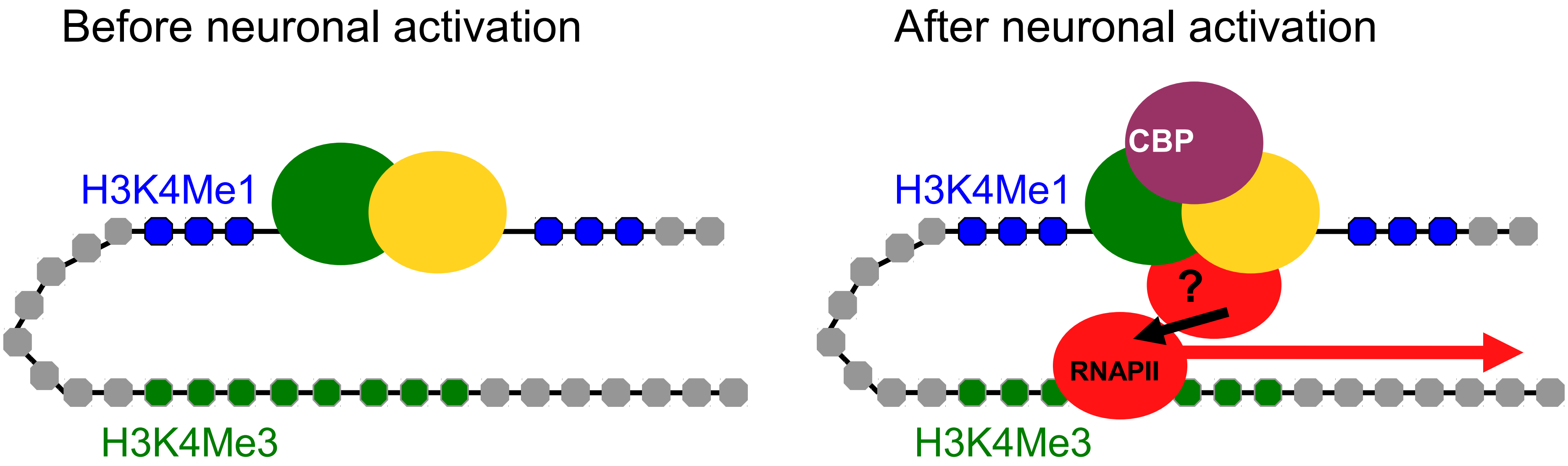
CBP levels determined by relative affinity of TF complexes



CBP levels determined by relative affinity of TF complexes

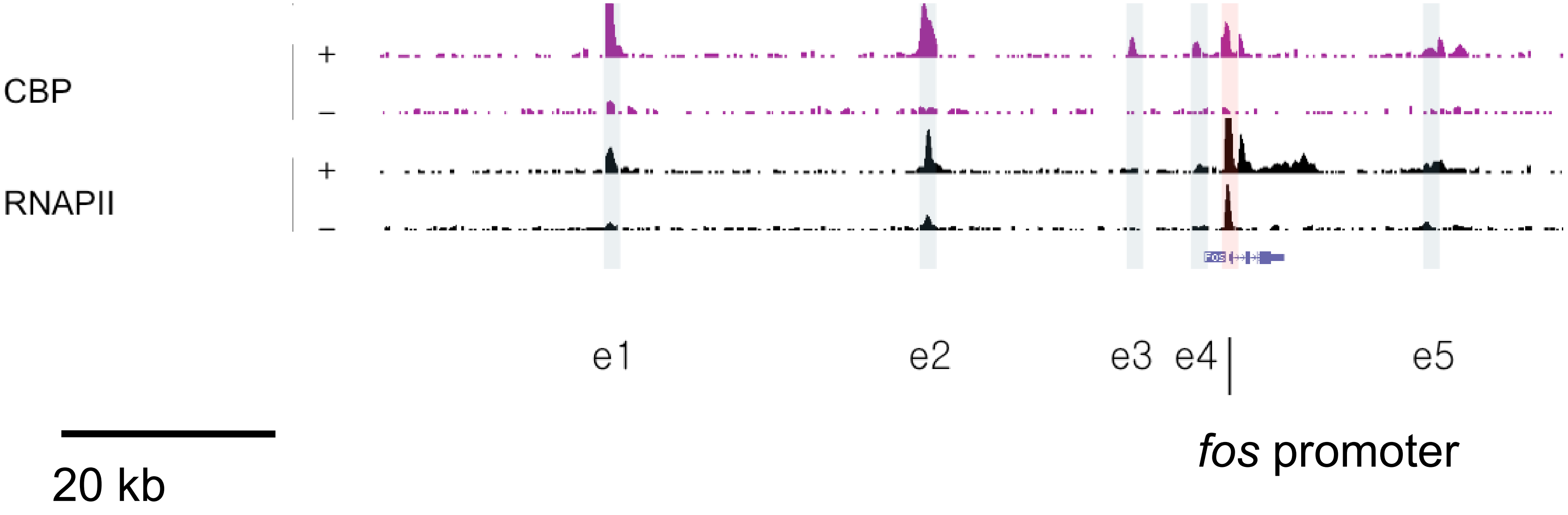


What is the role of CBP at enhancers?

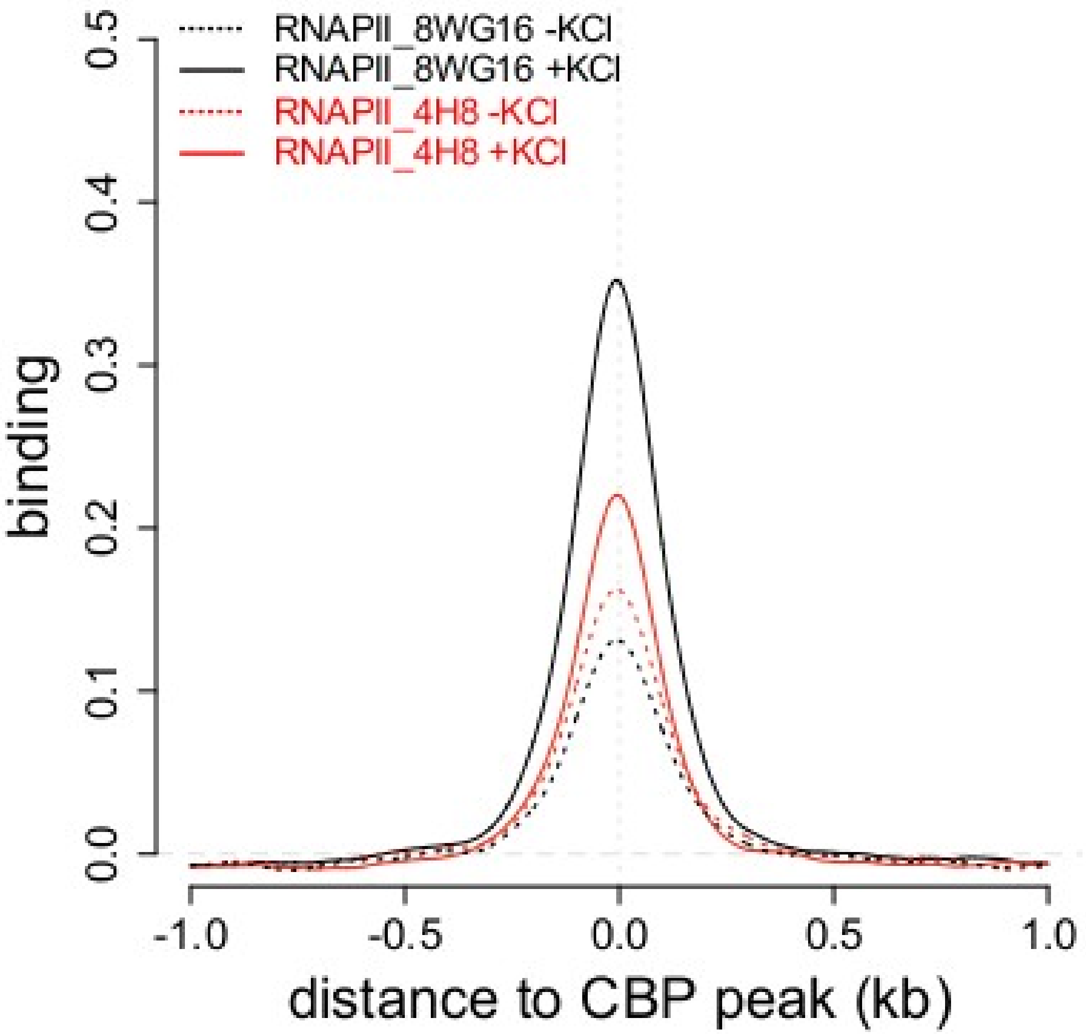


- Is CBP determined by TF combinations? YES
- Does RNAPII bind at enhancers?

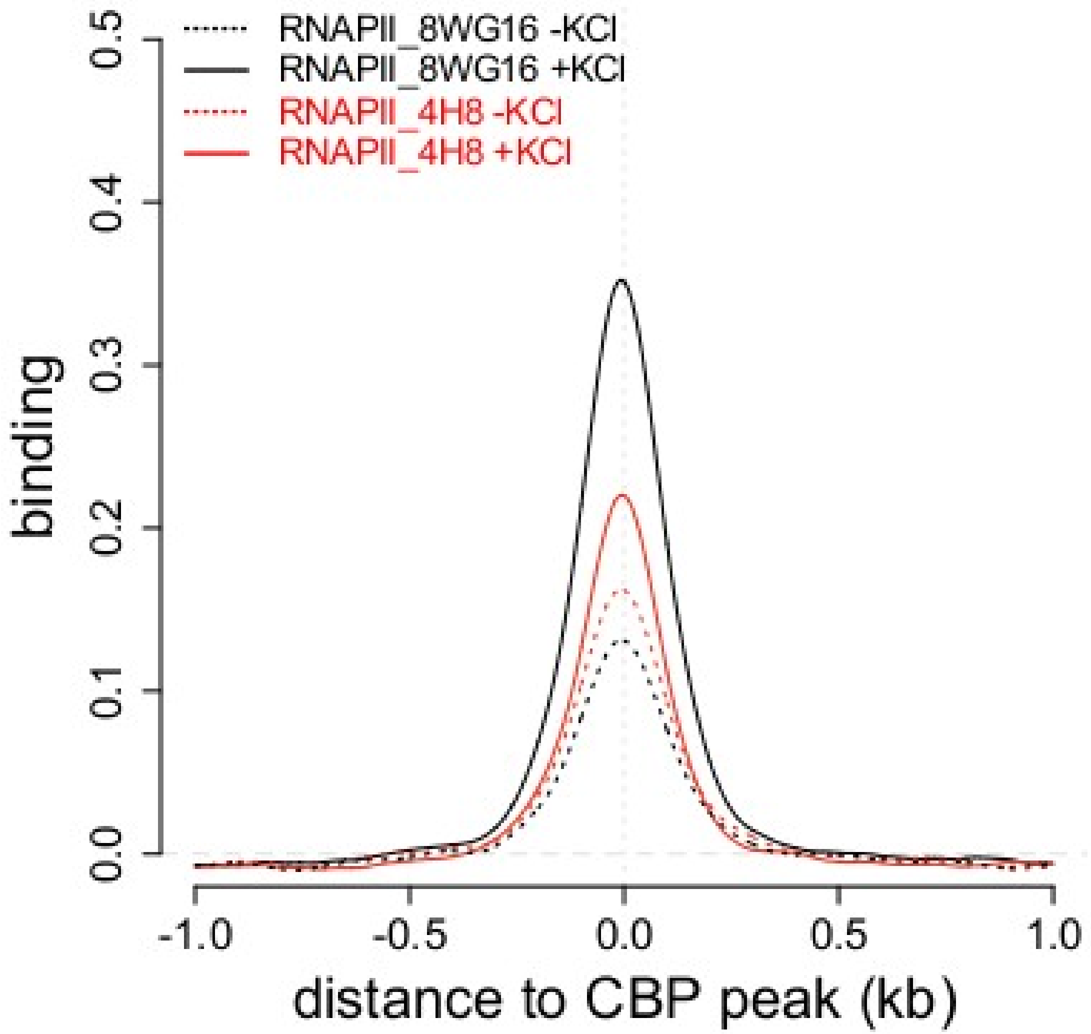
RNAPII is recruited to CBP binding sites at the *fos* locus

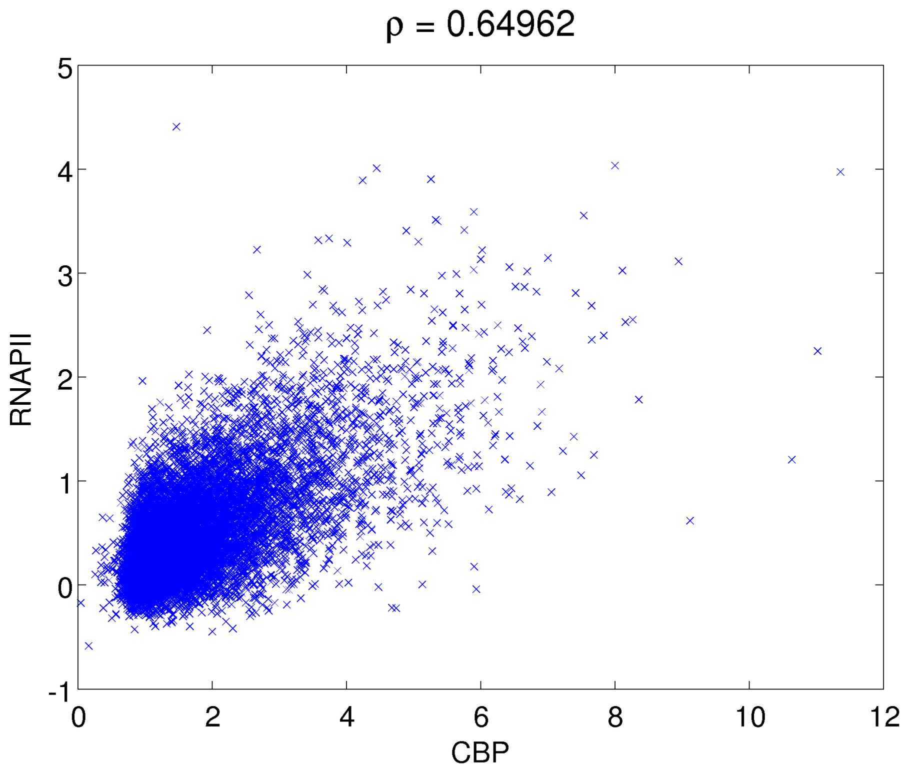


RNAPII is recruited at all enhancers

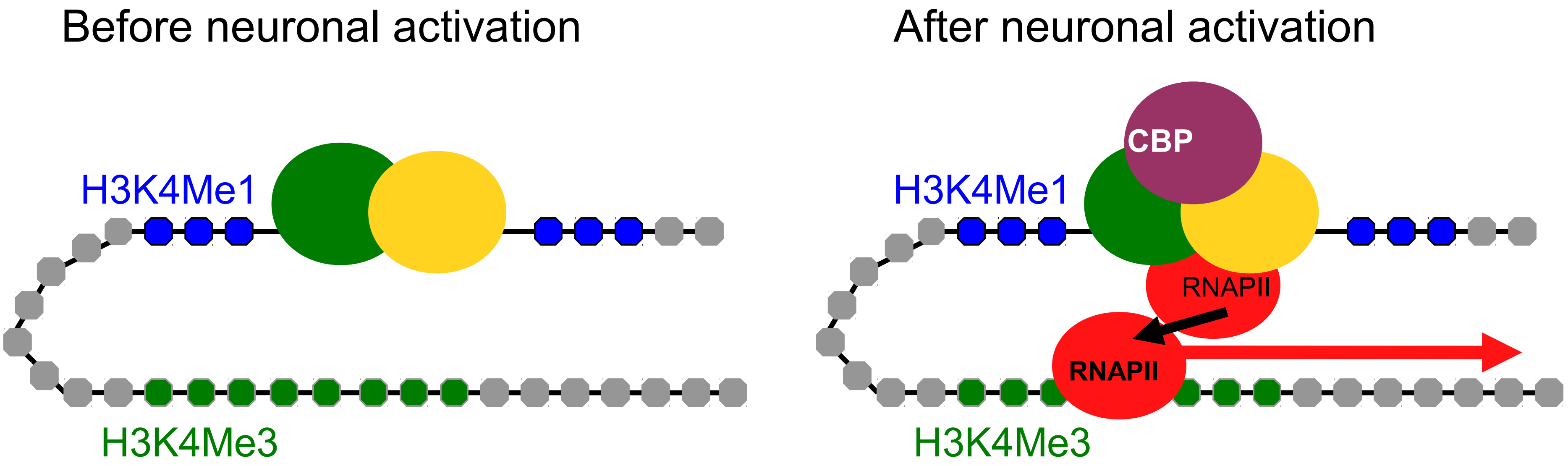


RNAPII is recruited at all enhancers



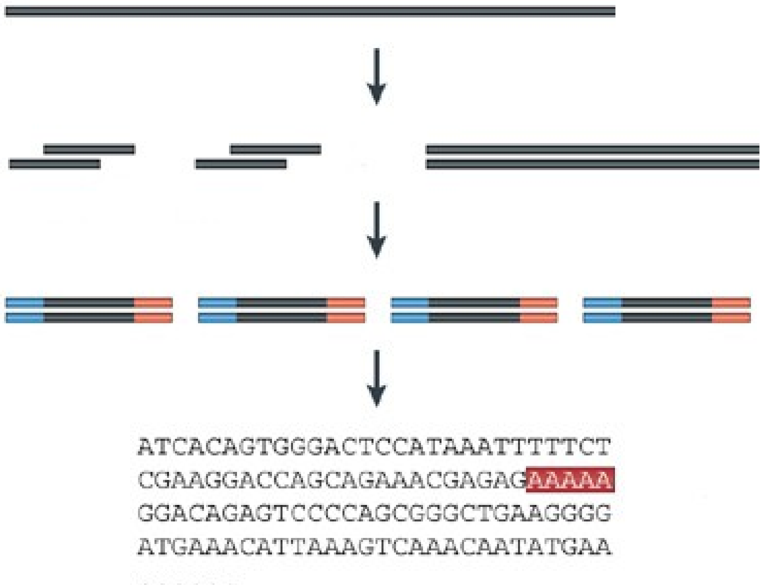


What is the function of RNAPII at enhancers?



- Is CBP determined by TF combinations? YES
- Does RNAPII bind at enhancers? YES
- Are transcripts produced at enhancers?

RNA-Seq reveals which parts of the genome are transcribed



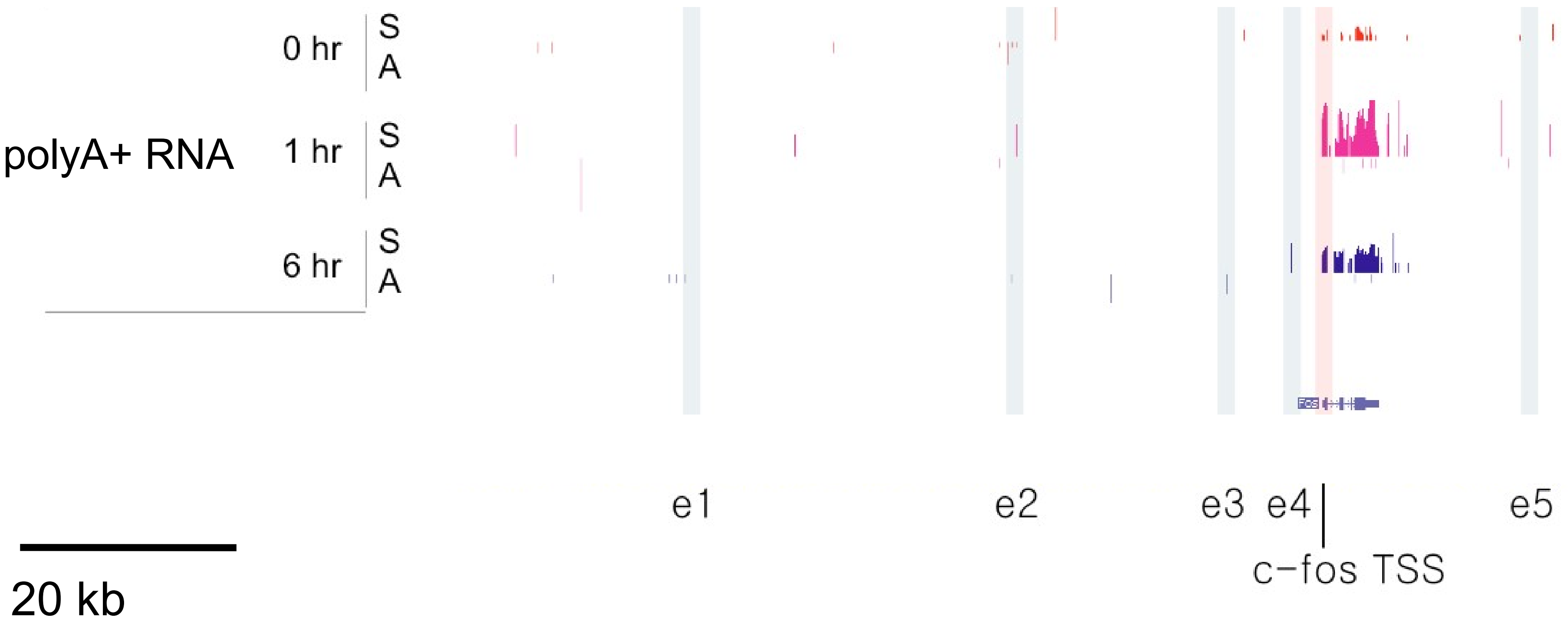
(Wang et al, 2009)

polyA tail is added to messenger RNAs (**mRNA**s)

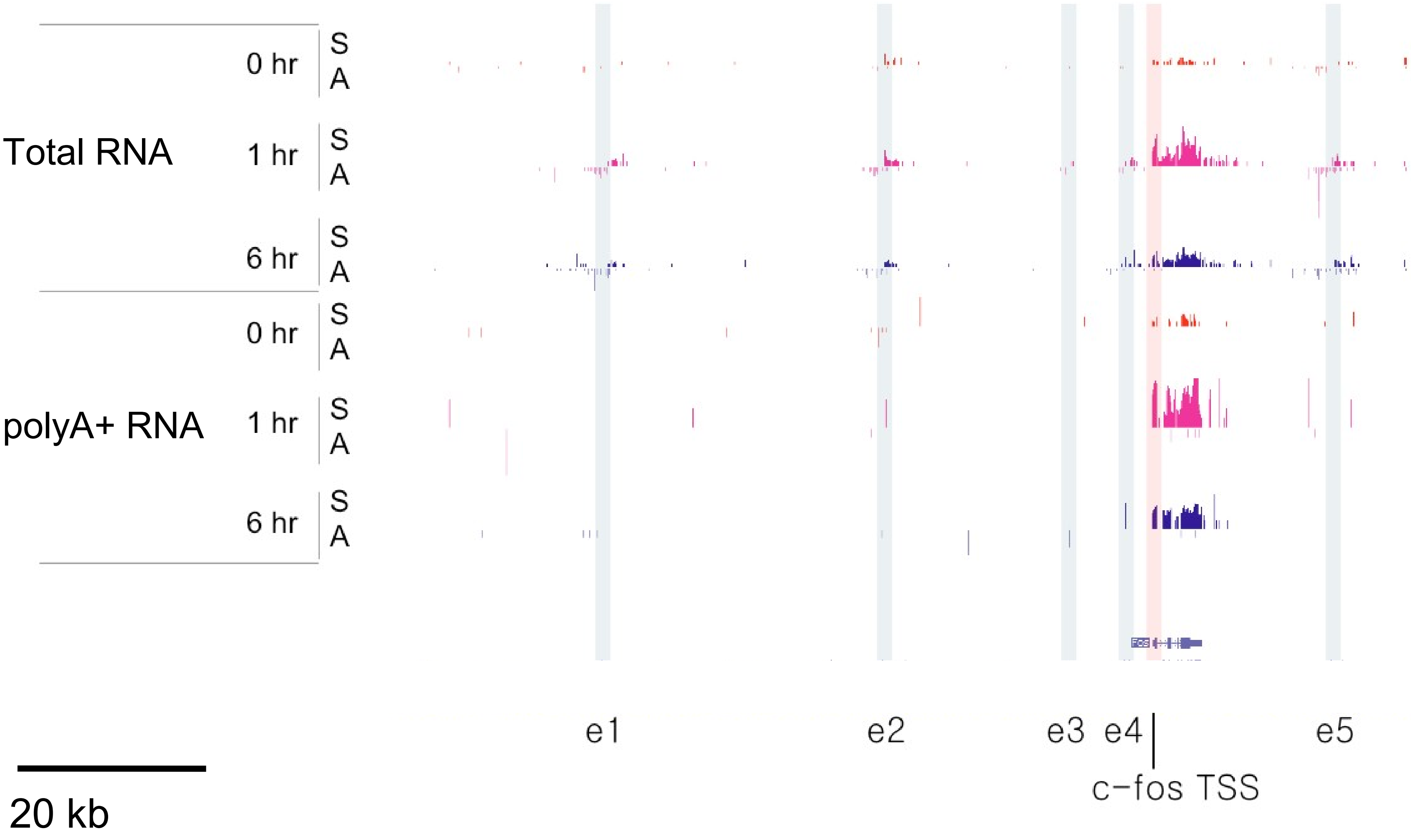
- Increases stability
- Allows transport out of nucleus

ACGUUUGUACCUAGCUAGCUUACGAG AAAAAAAAAAAAAAAAAAAA

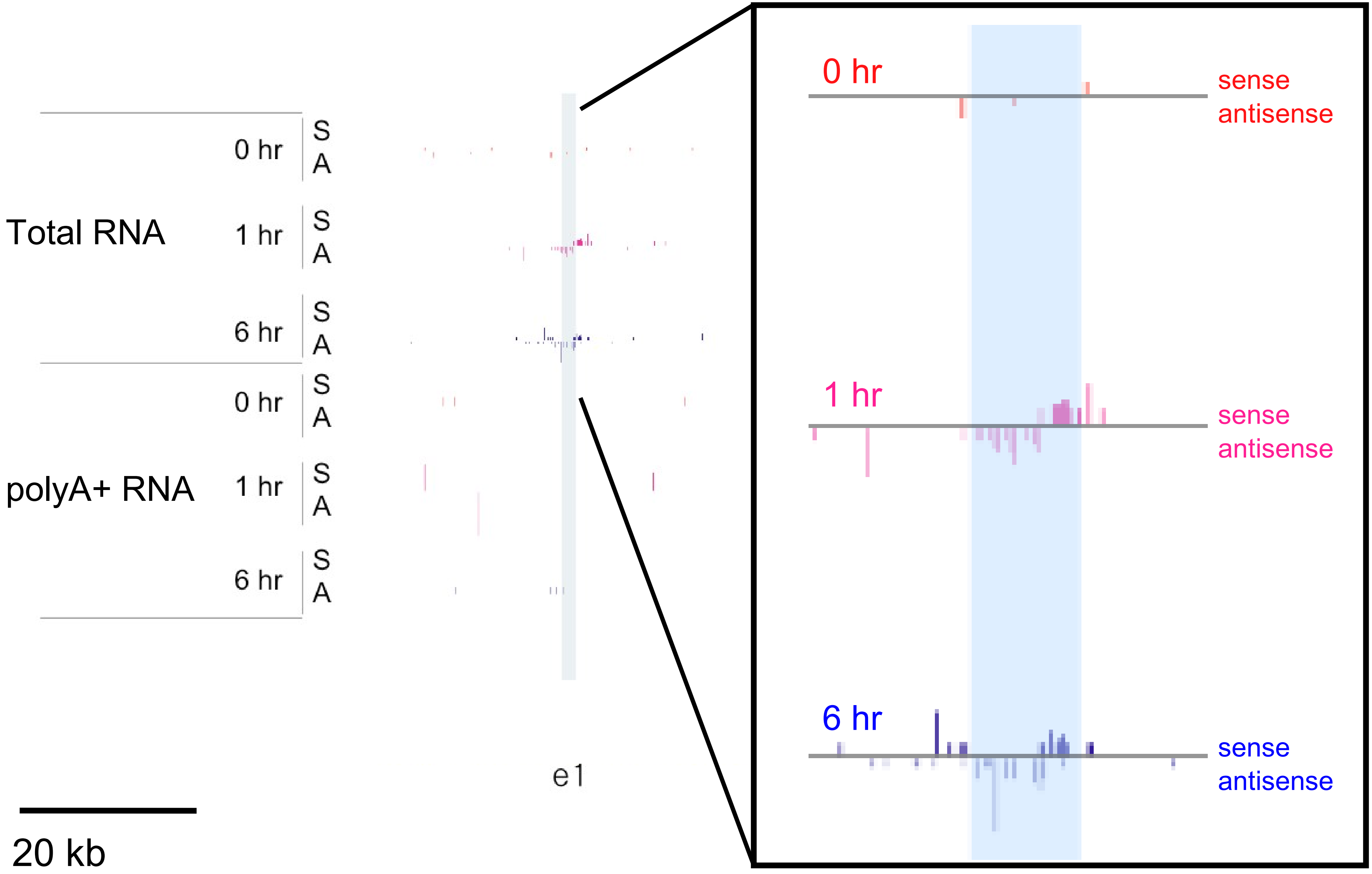
Transcription of mRNA at the *fos* locus



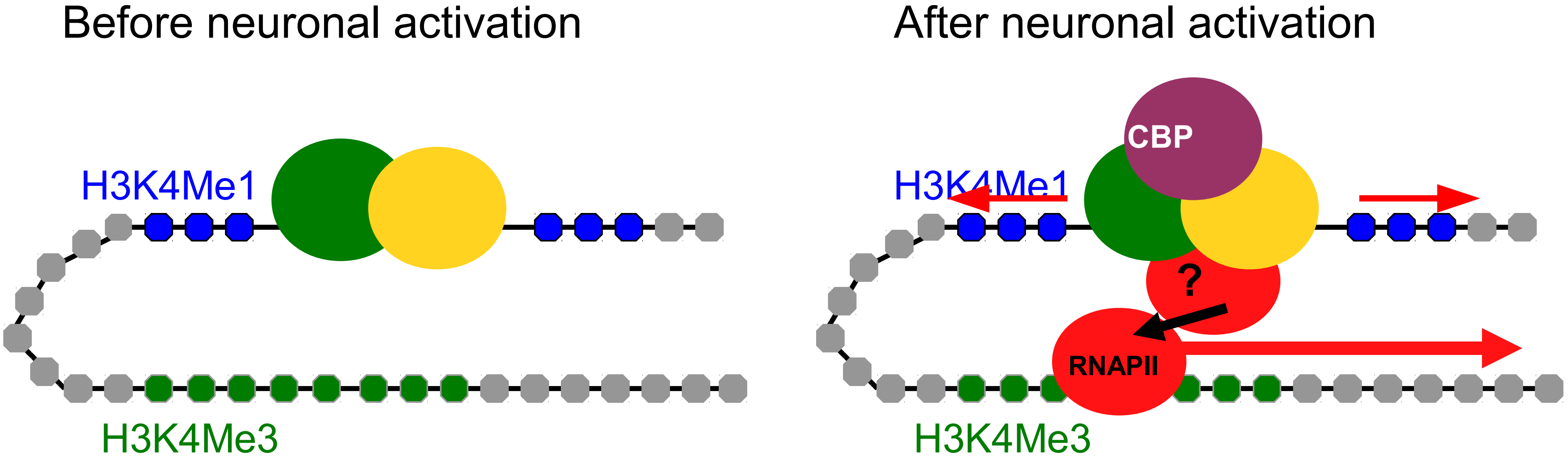
Transcription of total RNA at the *fos* locus



Transcription at enhancers is activity-dependent



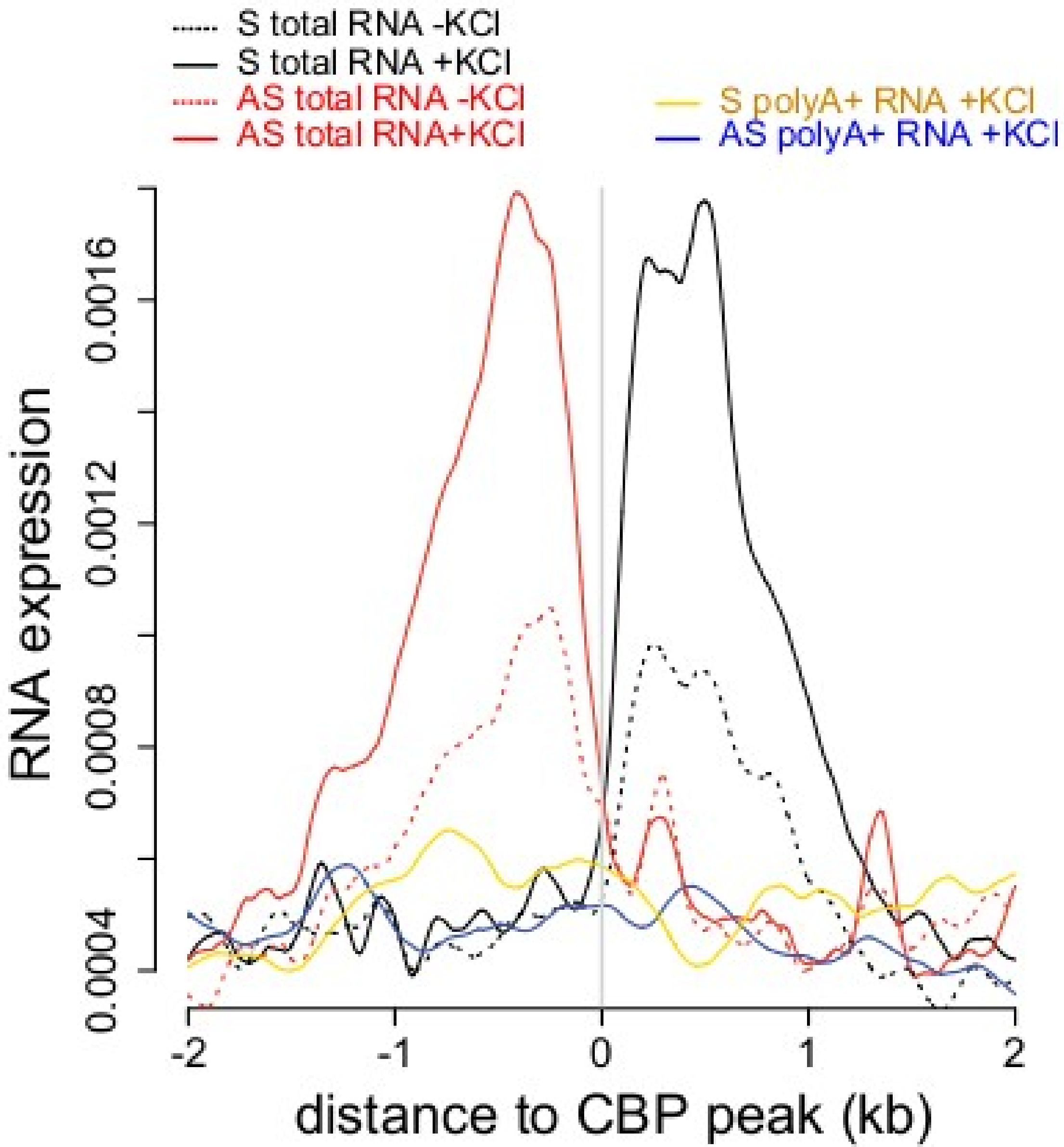
What are the properties of eRNAs?



eRNAs are induced by activity

eRNAs are not polyadenylated

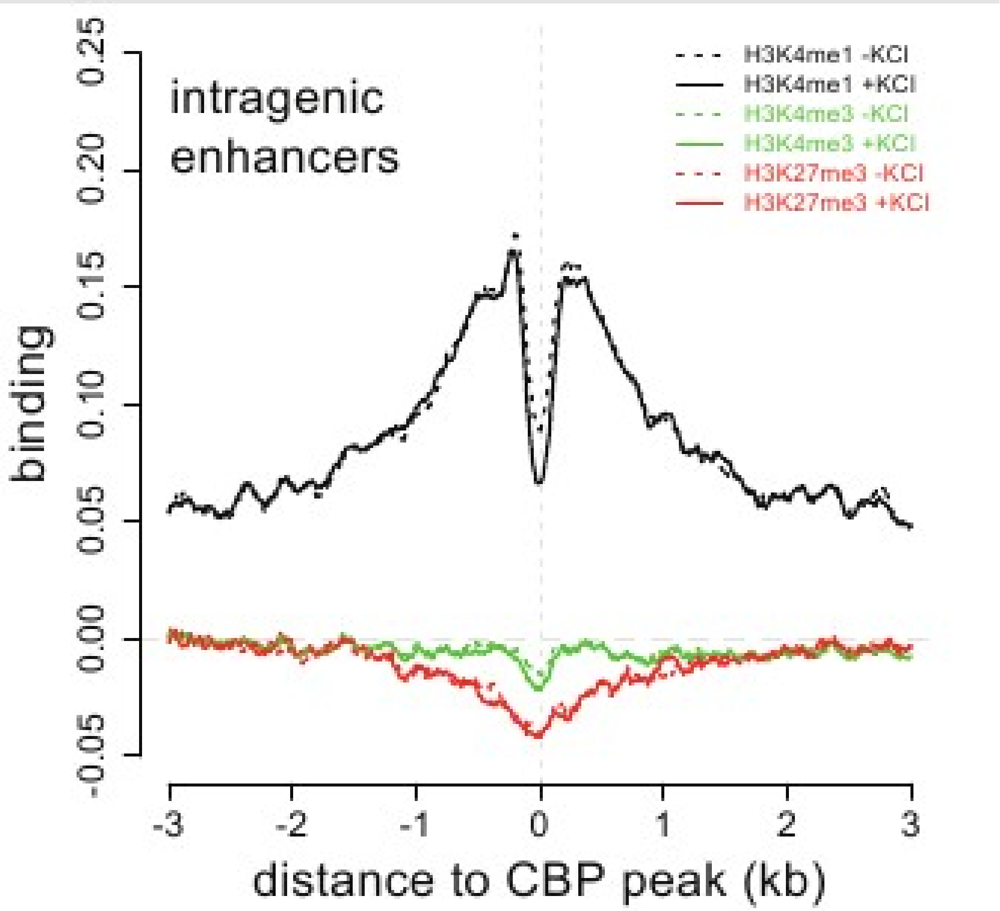
Properties of enhancer RNAs



- Inducible
- Low expression
- ~1.5 kb
- Bidirectional
- No polyA-tail
- Not protein-coding

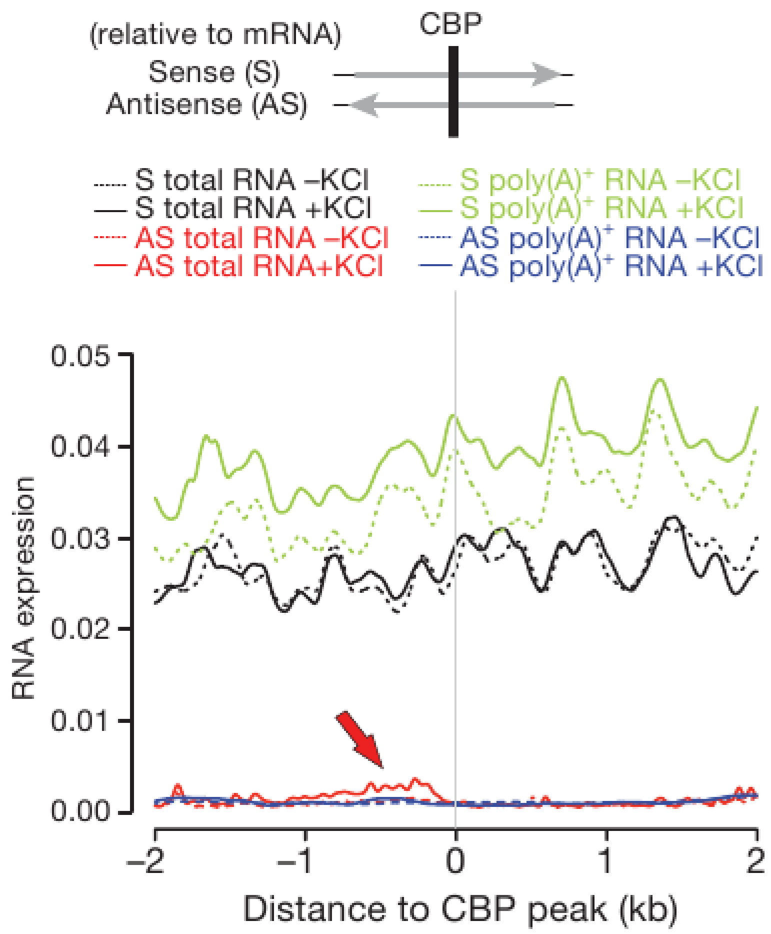
Intragenic enhancers

- ~7,000 enhancers overlapping introns
  - H3K4me1, but no H3K4me3



Intragenic enhancers are also transcribed

- ~7,000 enhancers overlapping introns
  - No signal detectable on sense strand
  - Significant anti-sense transcription



How abundant are eRNAs compared to mRNAs?

- Identify **all** transcripts in the genome
    - Wavelet-based algorithm for *de novo* detection of transcribed regions accounts for 99.8% of reads
        - mRNAs and annotated ncRNAs represent X%
        - eRNAs represent Y%
            - 1 in 10,000 reads is an eRNA read
            - MRNAs ~100 times more abundant

Why do eRNAs have such low abundance?

- eRNA production much slower than mRNA
- eRNA decay much faster than mRNA

A simple model of transcription

- eRNA production much slower than mRNA
- eRNA decay much faster than mRNA

$$
\begin{aligned}
\frac{dM}{dt} &= \frac{P_M k}{L_M} - \frac{M}{\tau_M} \\
\frac{dE}{dt} &= \frac{P_E k}{L_E} - \frac{E}{\tau_E}
\end{aligned}
$$

M – mRNA
E – eRNA
P – polymerase levels
k – elongation rate
L – length of transcript
tau – RNA half life

Half life of eRNAs relative to mRNAs

- eRNA production much slower than mRNA
- eRNA decay much faster than mRNA

$$\frac{dM}{dt} = \frac{P_M k}{L_M} - \frac{M}{\tau_M}$$

$$\frac{dE}{dt} = \frac{P_E k}{L_E} - \frac{E}{\tau_E}$$

M – mRNA
E – eRNA
P – polymerase levels
k – elongation rate
L – length of transcript
tau – RNA half life

$$\frac{\tau_E}{\tau_M} = \frac{E^*}{M^*}\frac{L_E}{L_M}\frac{P_M}{P_E}$$

eRNAs half life is approximately half an hour

- eRNA production much slower than mRNA
- eRNA decay much faster than mRNA

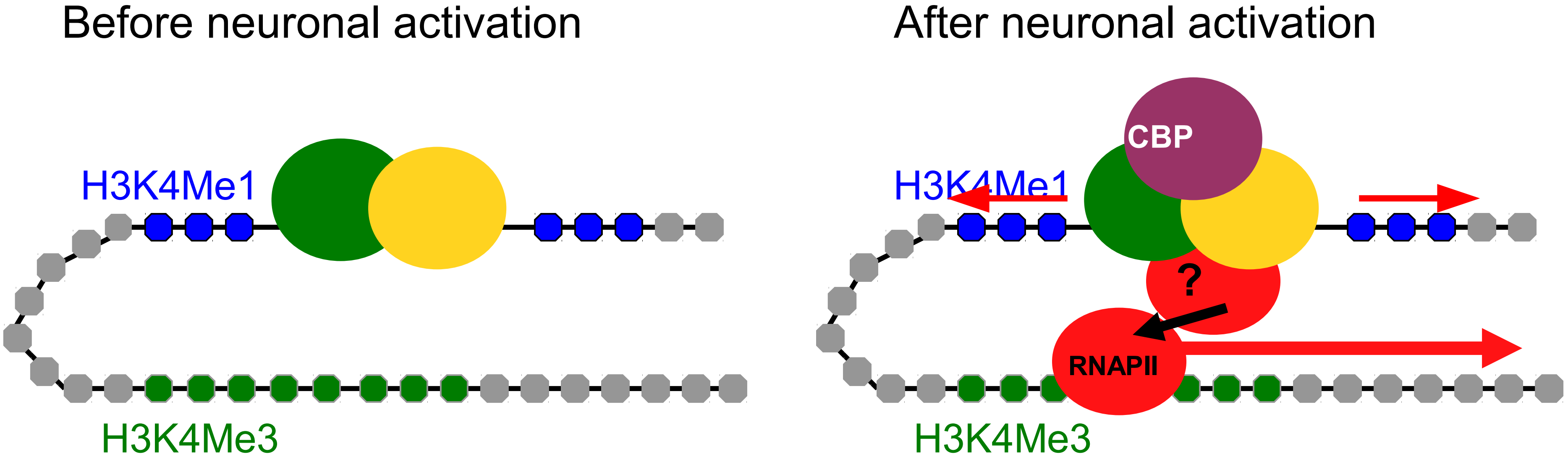
$$\frac{dM}{dt} = \frac{P_M k}{L_M} - \frac{M}{\tau_M}$$
$$\frac{dE}{dt} = \frac{P_E k}{L_E} - \frac{E}{\tau_E}$$

M – mRNA
E – eRNA
P – polymerase levels
k – elongation rate
L – length of transcript
tau – RNA half life

$$\frac{\tau_E}{\tau_M} = \frac{E^*}{M^*}\frac{L_E}{L_M}\frac{P_M}{P_E}$$

$$\tau_E \approx 10^{-2} \times \frac{1.5}{30} \times 5 \times \tau_M \approx 4 \times 10^{-2} \times 600\text{min} \boxed{= 24\text{min}}$$

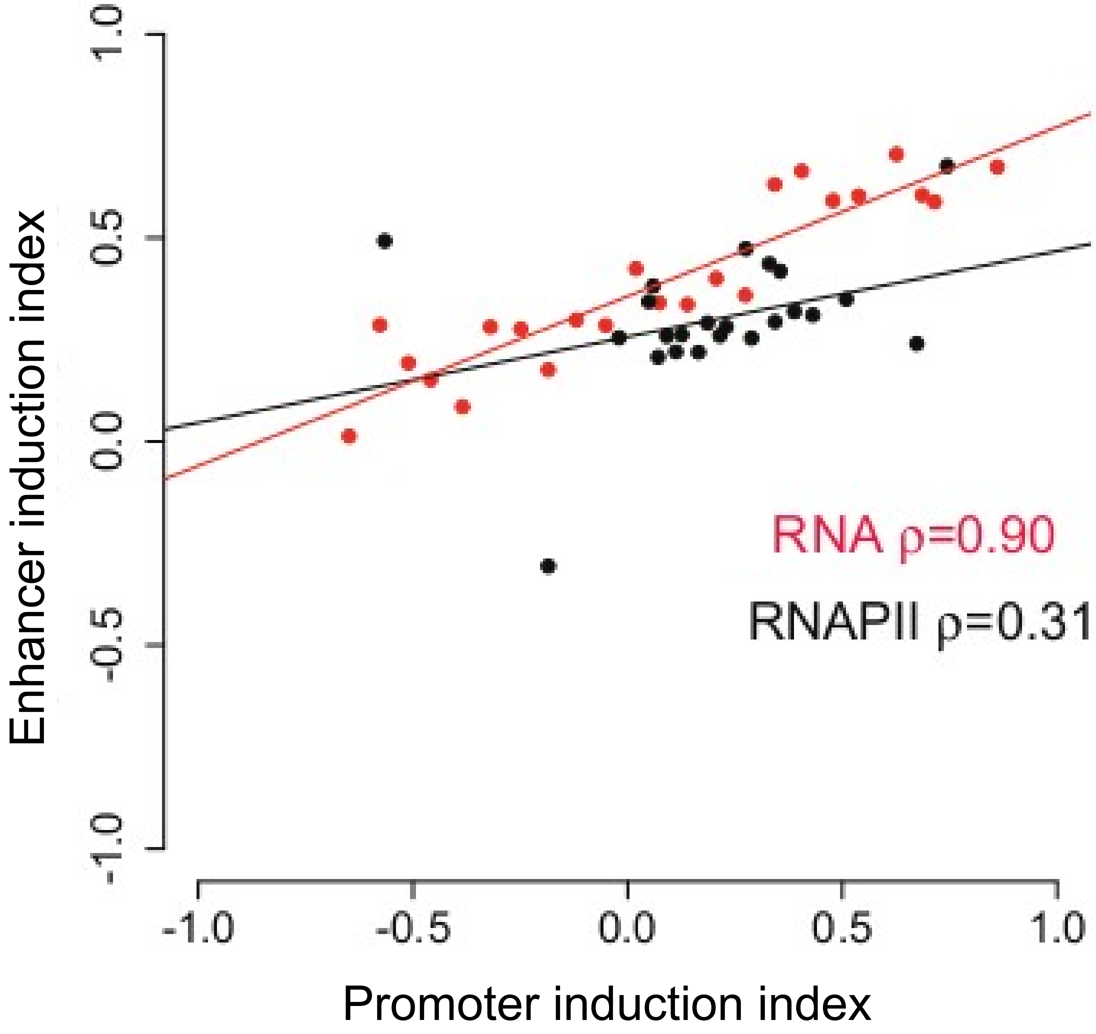
Enhancers recruit RNAPII and produce transcripts, but does it depend on promoter?



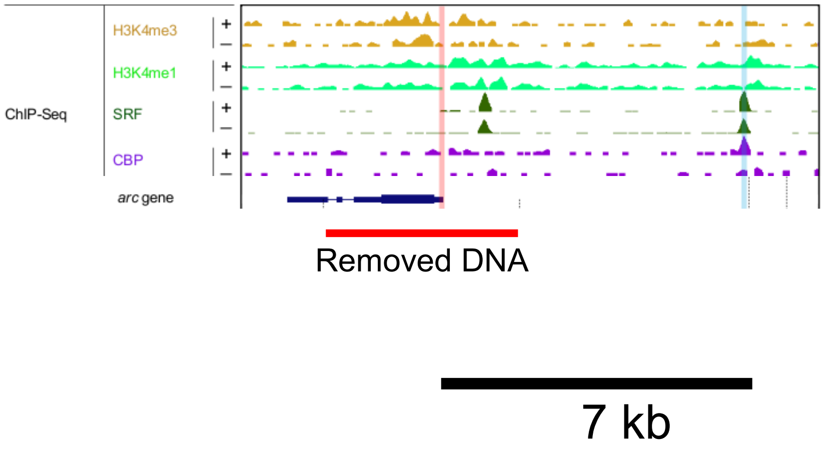
- Is CBP determined by TF combinations? YES
- Does RNAPII bind at enhancers? YES
- Are transcripts produced at enhancers? YES
- Is RNAPII recruitment independent?

eRNA induction is correlated with induction of nearby mRNAs

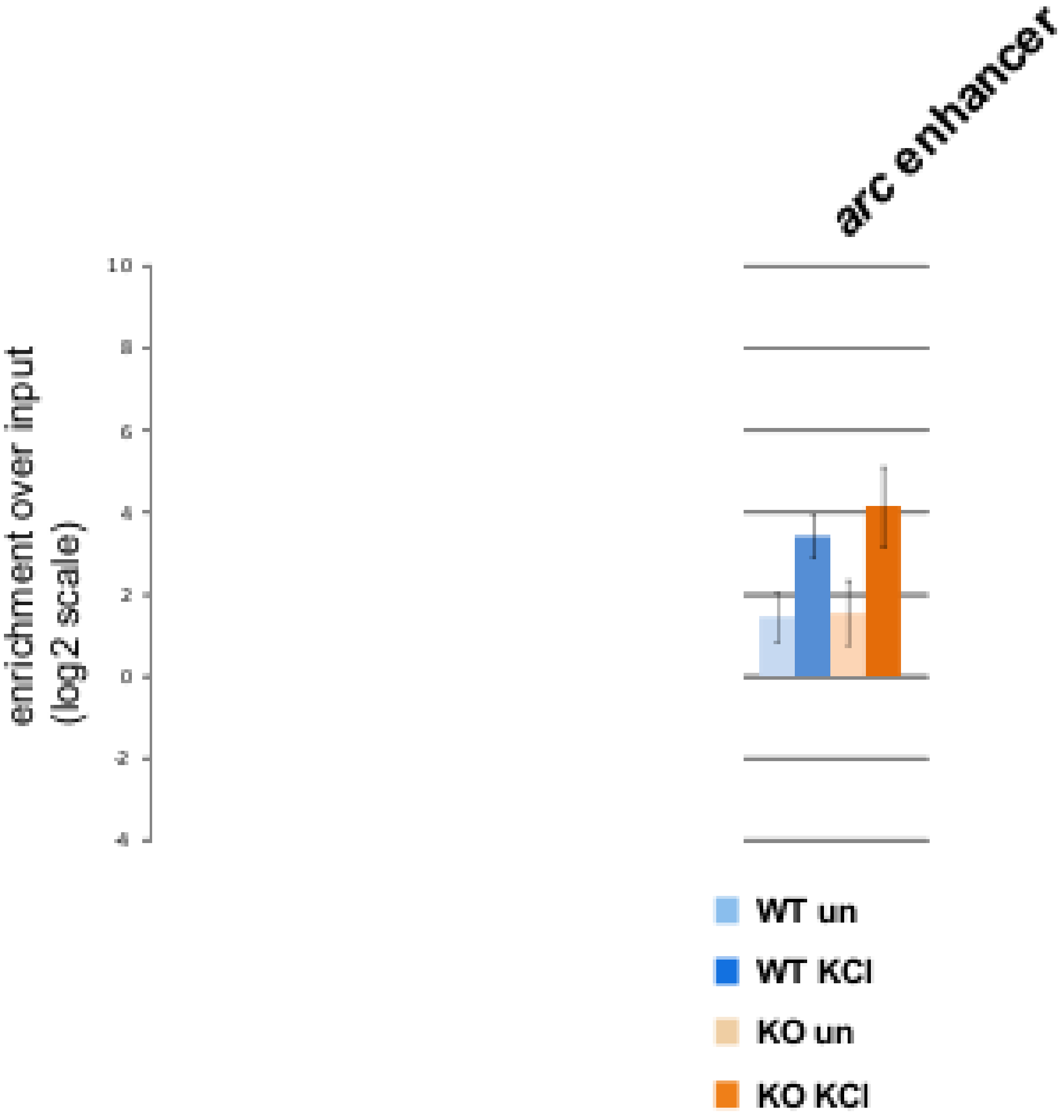
$$induction\ index = \frac{KCl - un}{KCl + un}$$

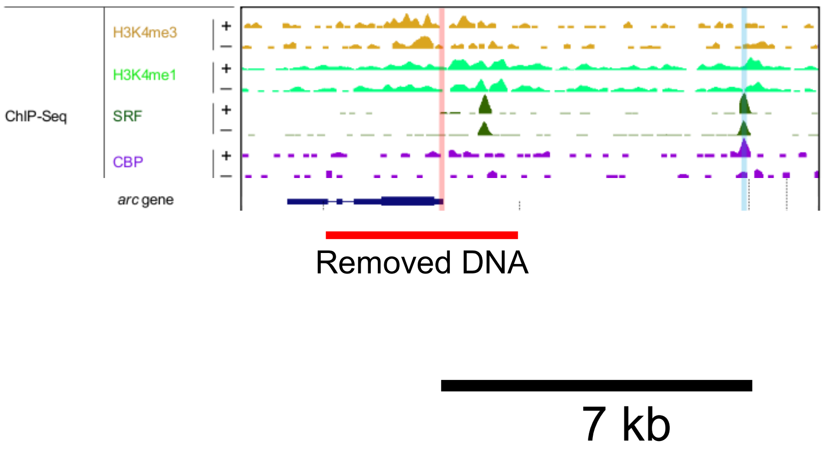


Deletion of the Arc-promoter

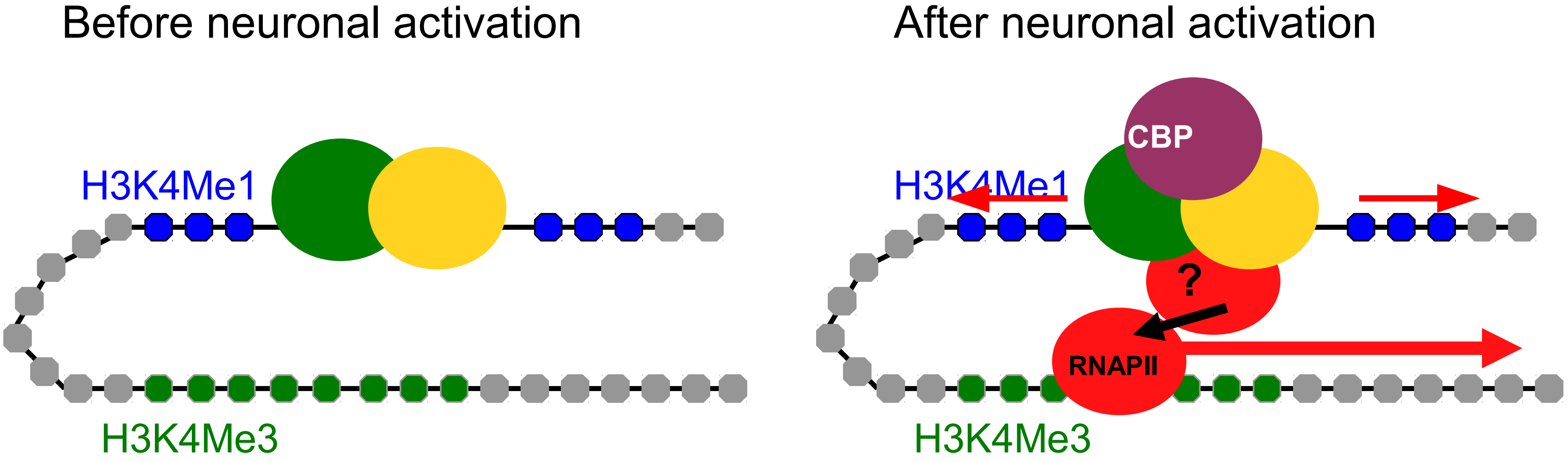


RNAPII levels are unchanged at the enhancer in the mutant before and after KCl



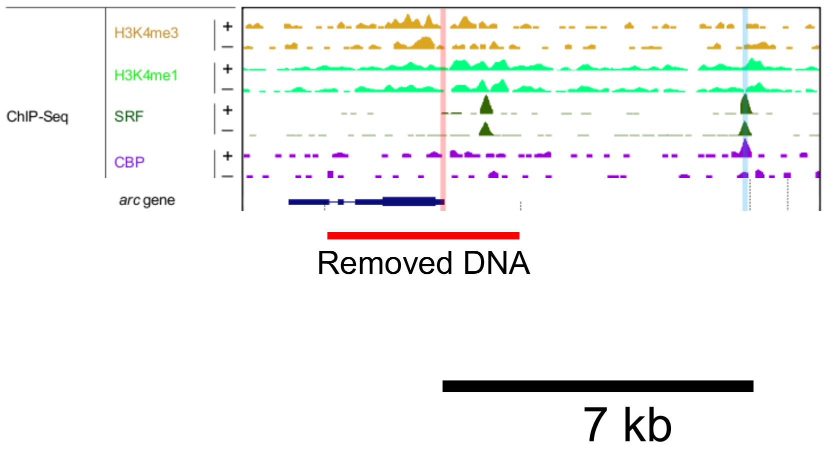


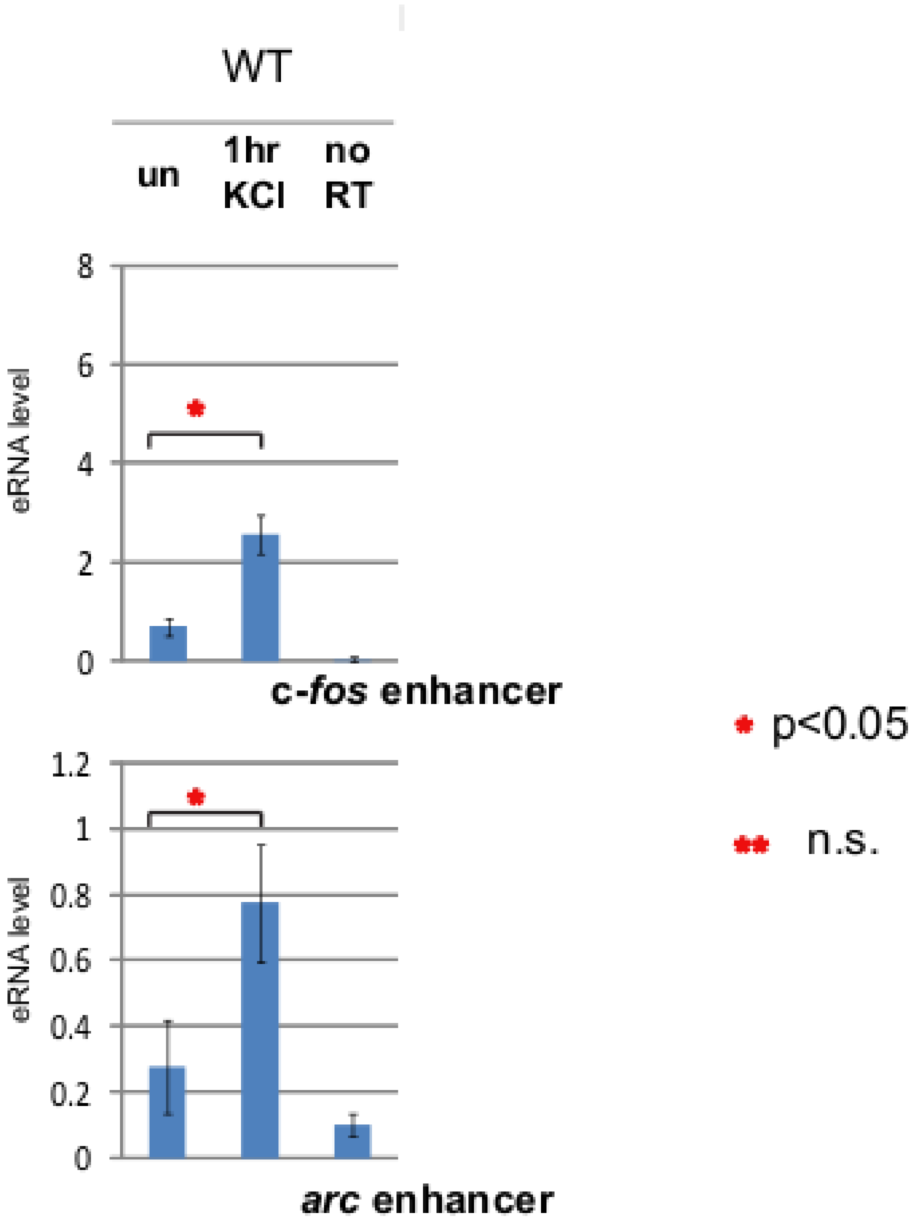
Enhancers bind RNAPII independently, does the transcription depend on the promoter?



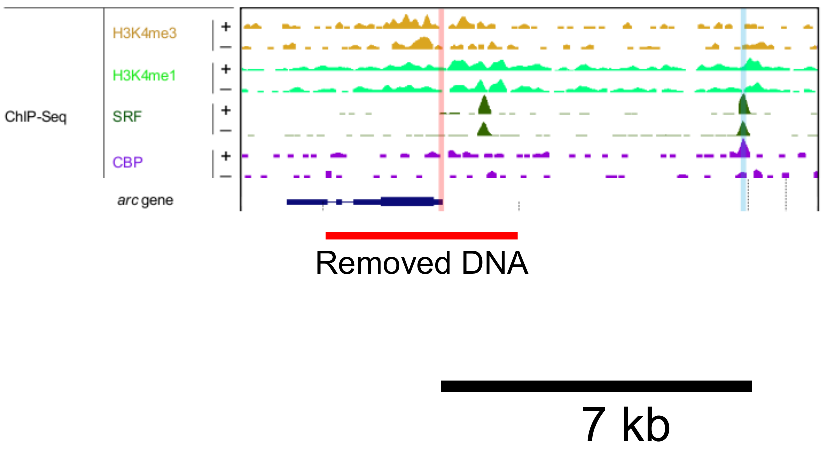
- Is CBP determined by TF combinations? YES
- Does RNAPII bind at enhancers? YES
- Are transcripts produced at enhancers? YES
- Is RNAPII recruitment independent? YES
- Is eRNA production independent?

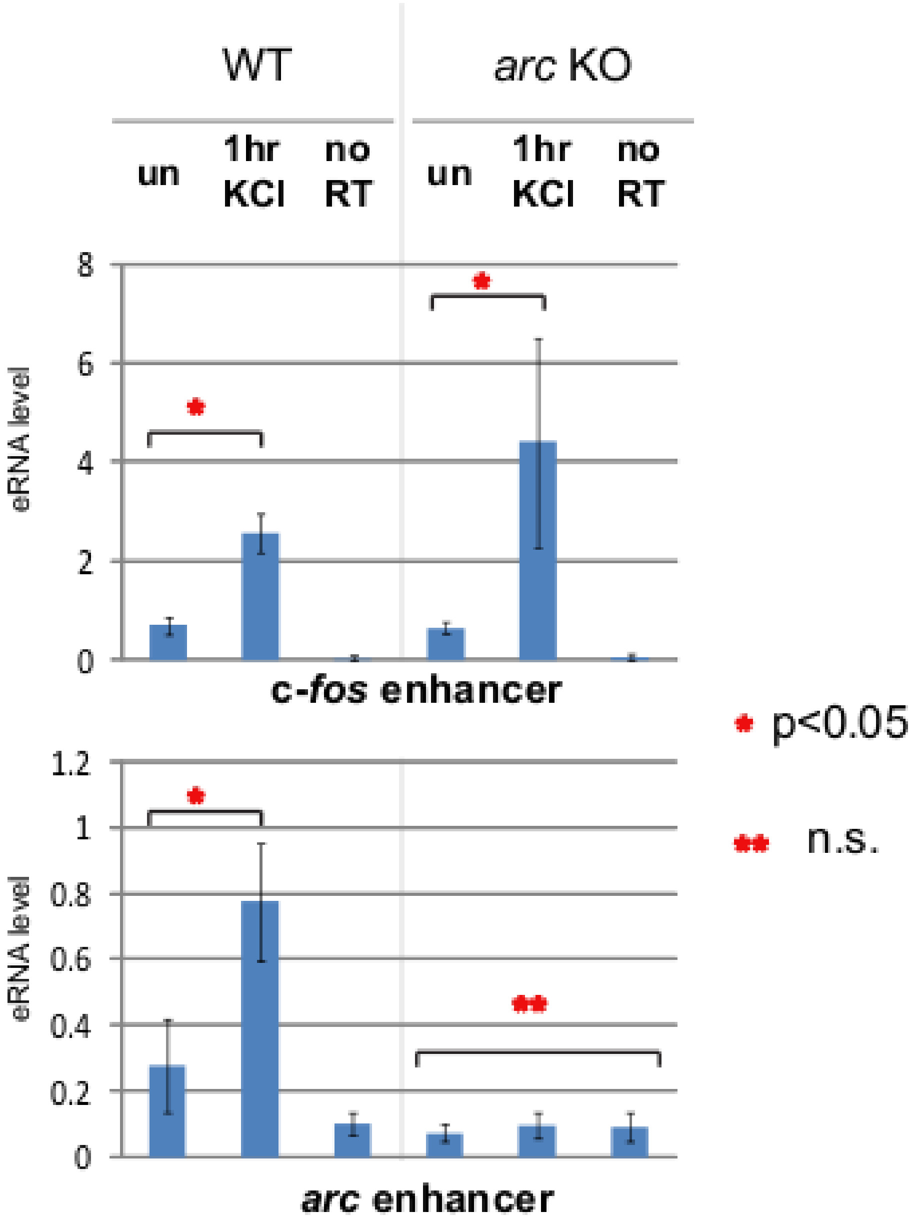
Transcription at the Fos and Arc enhancers



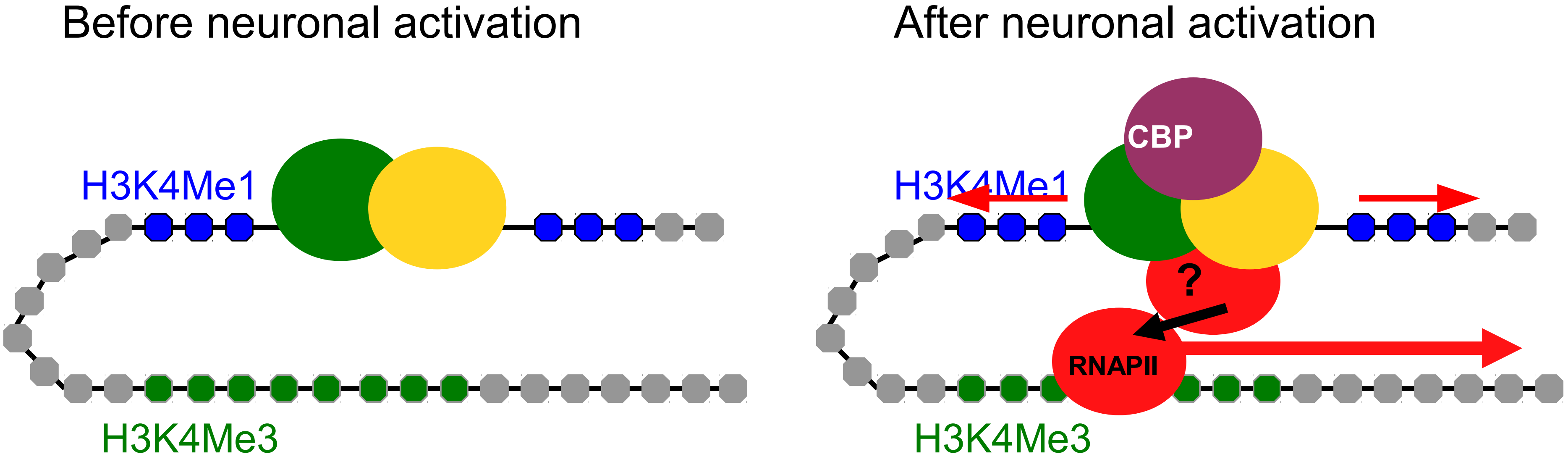


No transcription at Arc enhancer in mutant





Enhancers bind RNAPII independently, but the transcription is promoter-dependent



- Is CBP determined by TF combinations? YES
- Does RNAPII bind at enhancers? YES
- Are transcripts produced at enhancers? YES
- Is RNAPII recruitment independent? YES
- Is eRNA production independent? NO

What is the function of RNAPII at enhancers?

- Transcribe eRNAs
- Speed up recruitment of RNAPII at promoter

What is the function of RNAPII at enhancers?

- Transcribe eRNAs
- Speed up recruitment of RNAPII at promoter

$$\frac{dP_M}{dt} = k_p + N k_e c - \frac{P_M}{\tau}$$

P – polymerase levels
k_p – binding rate at promoter
k_e – binding rate at enhancer
N – number of enhancers
c – contact probability
tau – RNA half life

What is the function of RNAPII at enhancers?

- Transcribe eRNAs
- Speed up recruitment of RNAPII at promoter

$$\frac{dP_M}{dt} = k_p + Nk_ec - \frac{P_M}{\tau}$$

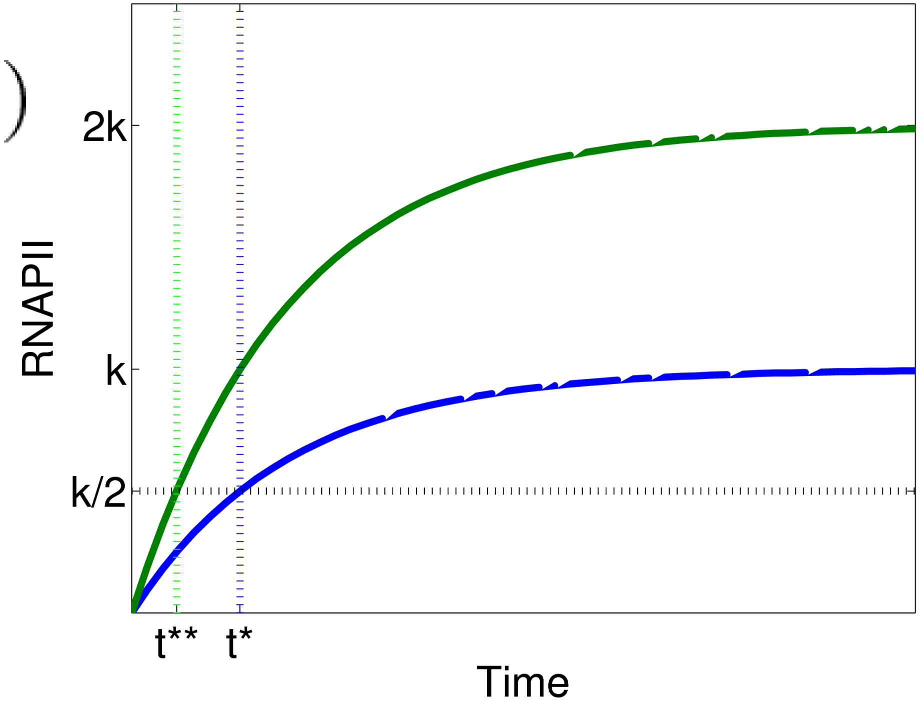
$$P_M(t) = (k_p + Nk_ec)(1 - e^{-t/\tau})$$

What is the function of RNAPII at enhancers?

- Transcribe eRNAs
- Speed up recruitment of RNAPII at promoter

$$\frac{dP_M}{dt} = k_p + Nk_e c - \frac{P_M}{\tau}$$

$$P_M(t) = (k_p + \boxed{Nk_e c})(1 - e^{-t/\tau})$$



What is the function of RNAPII at enhancers?

- Transcribe eRNAs
- Speed up recruitment of RNAPII at promoter

$$\frac{dP_M}{dt} = k_p + Nk_ec - \frac{P_M}{\tau} + \boxed{\sigma\sqrt{P_M(t)}\xi(t)}$$

→ Variance reduced by *(1 + Nc)*

What is the function of eRNAs?

> Science is always wrong. It never solves a problem without creating ten more.
>
> -George Bernard Shaw

- Noise
- Establish histone marks
- Transcript has function

eRNAs have been found in other cell types

doi:10.1038/nature09033 nature

ARTICLES

Widespread transcription at neuronal activity-regulated enhancers

Tae-Kyung Kim[1]*†, Martin Hemberg[2]*, Jesse M. Gray[1]*, Allen M. Costa[1], Daniel M. Bear[1], Jing Wu[3], David A. Harmin[1,4], Mike Laptewicz[1], Kellie Barbara-Haley[5], Scott Kuersten[6], Eirene Markenscoff-Papadimitriou[1]†, Dietmar Kuhl[7], Haruhiko Bito[8], Paul F. Worley[3], Gabriel Kreiman[2] & Michael E. Greenberg[1]

Histone H3K27ac separates active from poised enhancers and predicts developmental state

Menno P. Creyghton[a,1], Albert W. Cheng[a,b,1], G. Grant Welstead[a], Tristan Kooistra[c,d], Bryce W. Carey[a,e], Eveline J. Steine[a,e], Jacob Hanna[a], Michael A. Lodato[a,e], Garrett M. Frampton[a,e], Phillip A. Sharp[d,e], Laurie A. Boyer[e], Richard A. Young[a,e], and Rudolf Jaenisch[a,e,2]

OPEN ACCESS Freely available online PLoS BIOLOGY

A Large Fraction of Extragenic RNA Pol II Transcription Sites Overlap Enhancers

Francesca De Santa[1¶], Iros Barozzi[1¶], Flore Mietton[1¶], Serena Ghisletti[1], Sara Polletti[1], Betsabeh Khoramian Tusi[1], Heiko Muller[1], Jiannis Ragoussis[2], Chia-Lin Wei[3], Gioacchino Natoli[1]*

LETTER

doi:10.1038/nature09692

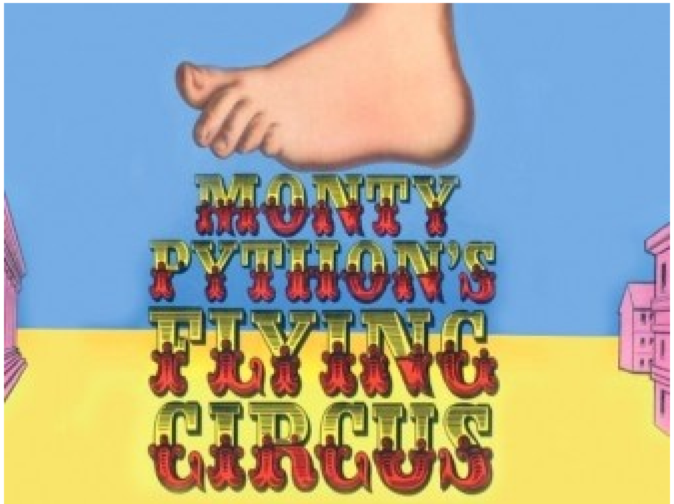
A unique chromatin signature uncovers early developmental enhancers in humans

Alvaro Rada-Iglesias[1], Ruchi Bajpai[1], Tomek Swigut[1], Samantha A. Brugmann[1], Ryan A. Flynn[1] & Joanna Wysocka[1,2]

Summary

- Identified ~12k activity-dependent enhancers
- Discovered and quantified novel mechanisms
  - Identified enriched motifs
  - Combinatorial affinity for CBP
  - Recruitment of RNAPII at enhancers
  - Transcription at enhancers
    - Properties of eRNA
    - Interaction with promoter necessary

And now something completely different....



Stochastic models of gene regulatory networks

- mRNAs often <10 copies per cell
- Describe using Master Equation (ME)
    - ME very difficult to solve, use Monte Carlo

Stochastic models of gene regulatory networks

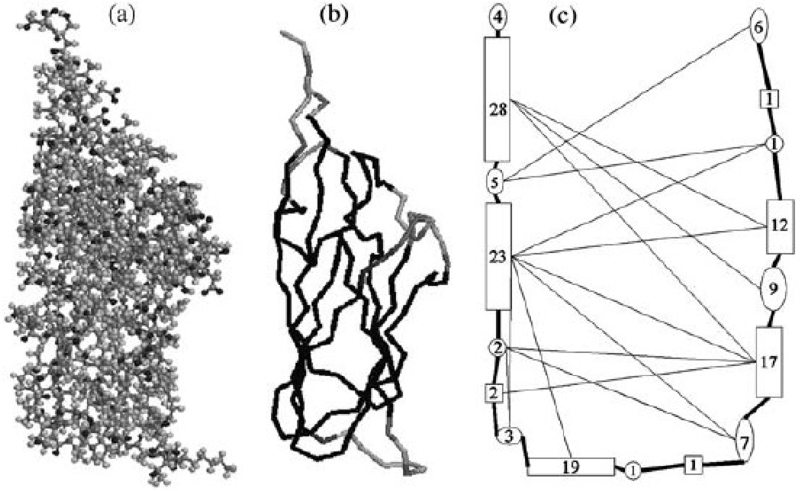
- mRNAs often <10 copies per cell
- Describe using Master Equation (ME)
    - ME very difficult to solve, use Monte Carlo
- How long do we need to run MC?
    - DCFTP-SSA **proven** to find stationary distribution
        - Provides certainty when using MC

Stochastic models of gene regulatory networks

- mRNAs often <10 copies per cell
- Describe using Master Equation (ME)
  - ME very difficult to solve, use Monte Carlo
- How long do we need to run MC?
  - DCFTP-SSA **proven** to find stationary distribution
    - Provides certainty when using MC
- Mixture model approach for analytical solutions and fits to experimental data

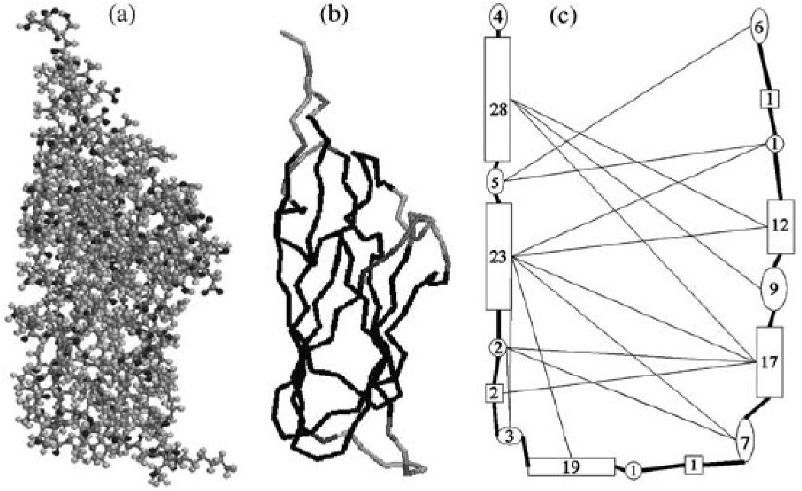
Assembly of viral capsids

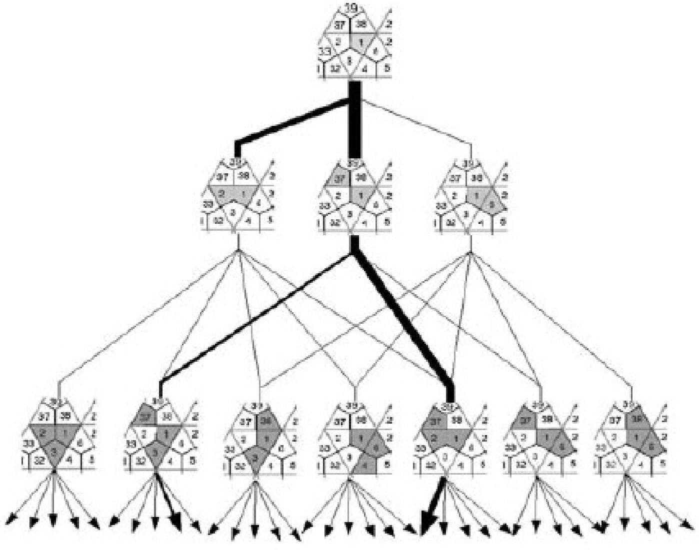
- Atomic-structure, coarse grain based on rigidity



Assembly of viral capsids

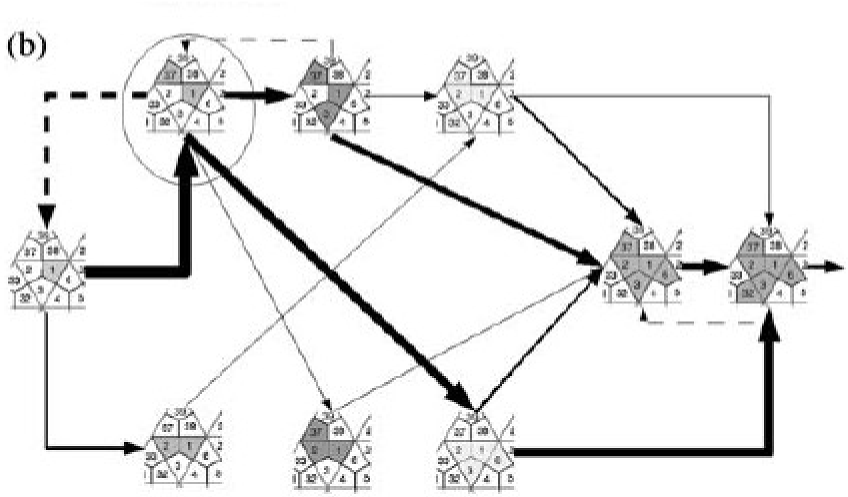
- Atomic-structure, coarse grain based on rigidity
- Oligomer association and dissociation rates
  - Association restricted by diffusion
  - Dissociation escape from multi-dimensional well





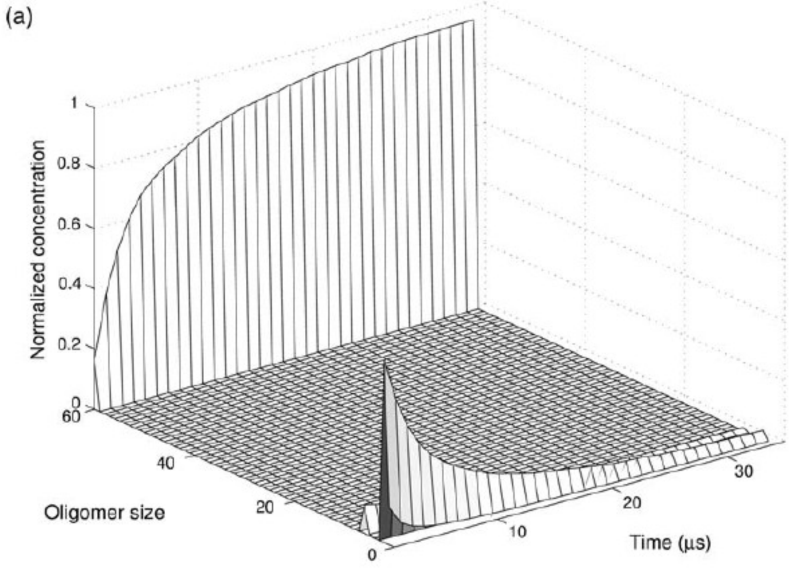
Assembly of viral capsids

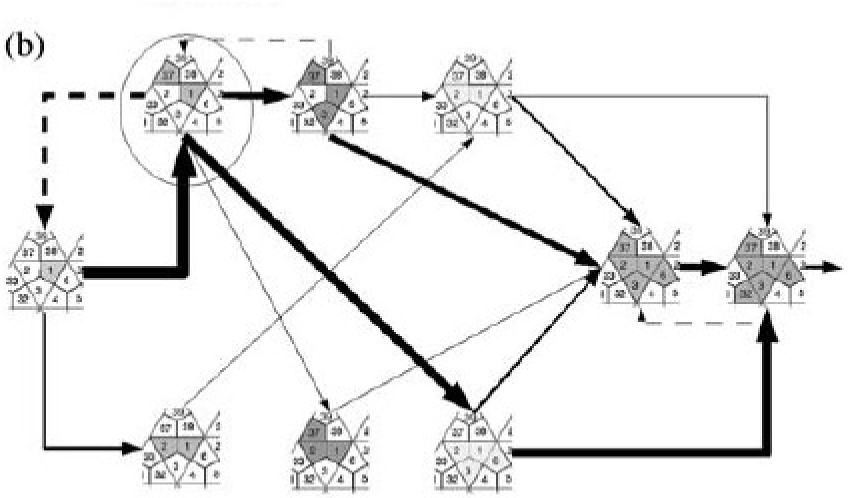
- Atomic-structure, coarse grain based on rigidity
- Oligomer association and dissociation rates
  - Association restricted by diffusion
  - Dissociation escape from multi-dimensional well
- Sample assembly paths



Assembly of viral capsids

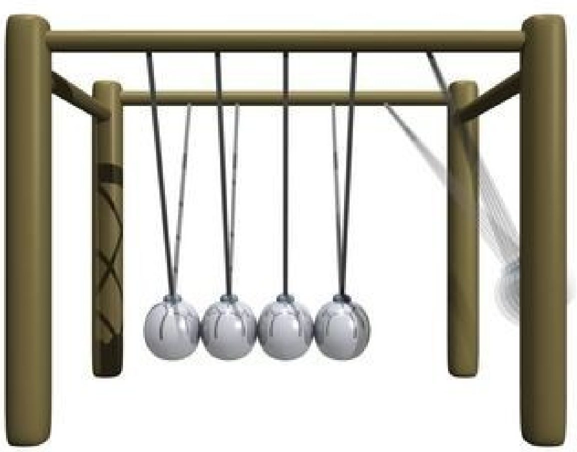
- Atomic-structure, coarse grain based on rigidity
- Oligomer association and dissociation rates
  - Association restricted by diffusion
  - Dissociation escape from multi-dimensional well
- Sample assembly paths

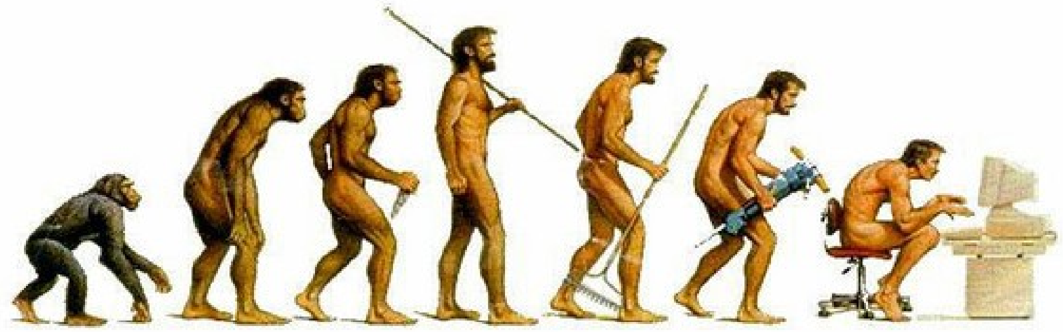




Future Work: Organizing principles of the genome

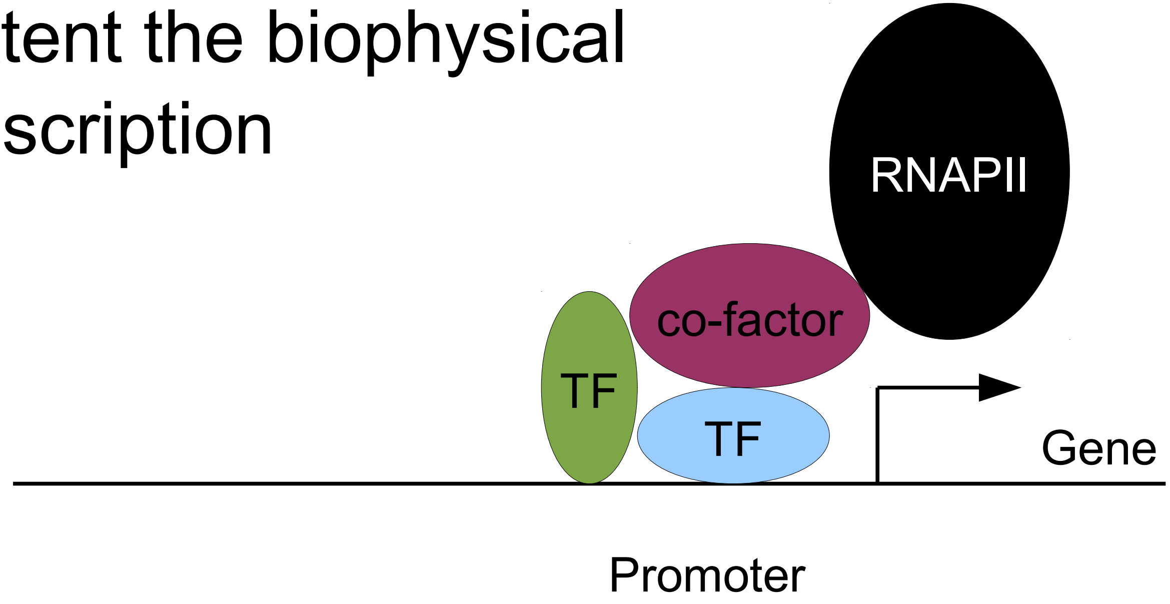
- Use genome-wide data to develop systems biology and biophysical type mathematical models of gene expression





Can biophysical models improve our understanding of TF binding and transcription?

- Use ChIP-Seq to test and compare biophysical models of TF binding
- Use RNA-Seq and synthetic biology approach to develop quantitative model of enhancer effect on expression
- Understand to what extent the biophysical properties affect transcription
  - DNA opening
  - DNA looping



Can we predict the structure of novel non-coding RNAs?

- Large number of non-coding RNAs discovered
- High-throughput experiments to probe structure
    - Sampling the folding and contact probability

Can models of stochastic gene expression be extended to entire transcriptome?

- Extend Poisson-Jacobi model combine with thermodynamic models
  - Develop MCMC methods
  - Develop robustness analyses
- Apply to single-cell RNA-Seq and FACS data
  - Global view of noise in gene expression

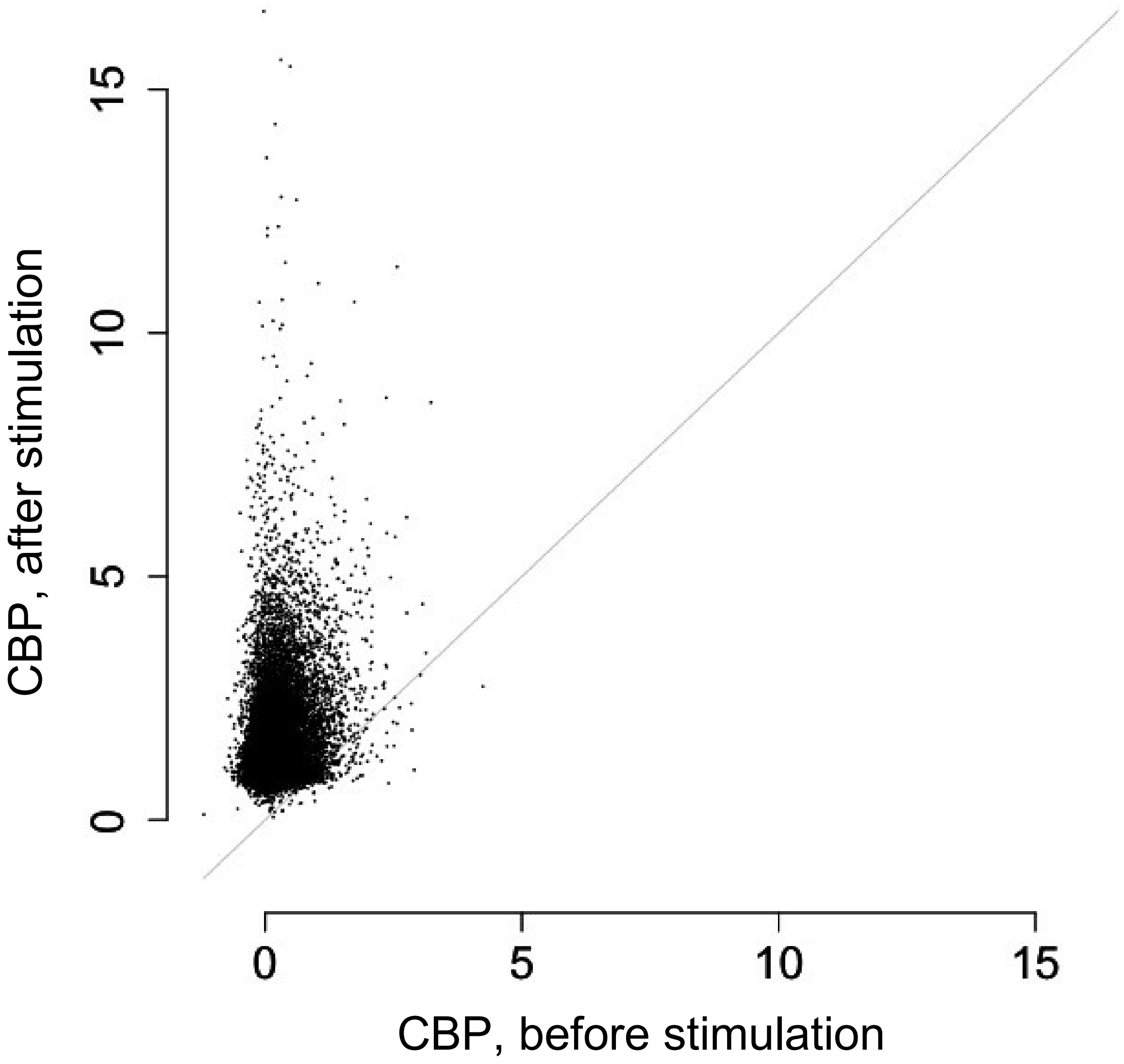
Acknowledgements

- Gabriel Kreiman
- Jesse Gray
- Tae-Kyung Kim
- Michael Greenberg
- Mauricio Barahona

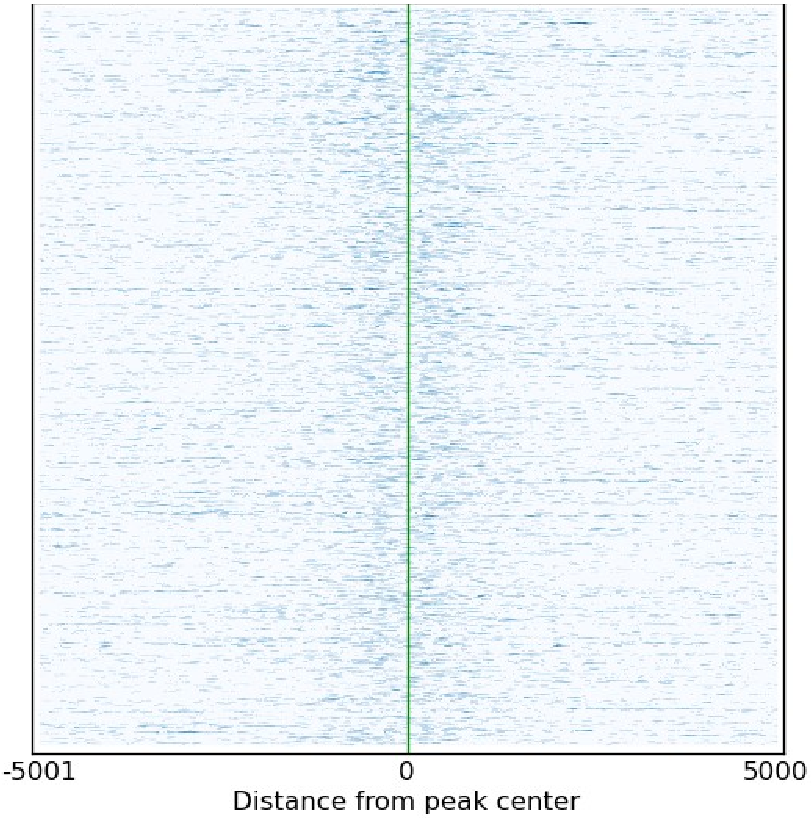
Thank You

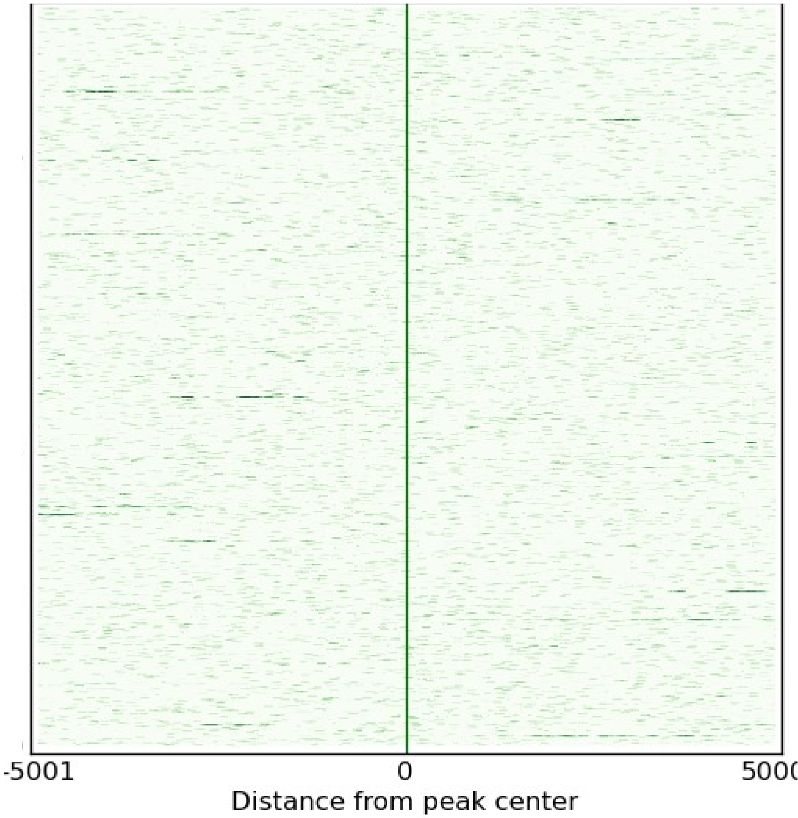
?

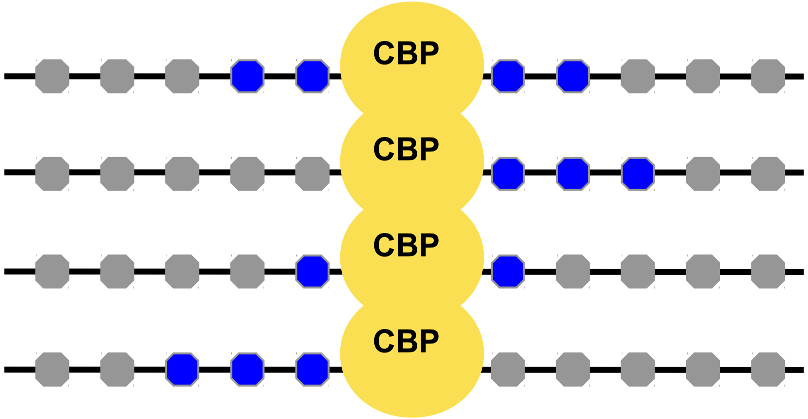
CBP binds in an activity regulated manner to ~28,000 sites throughout the genome



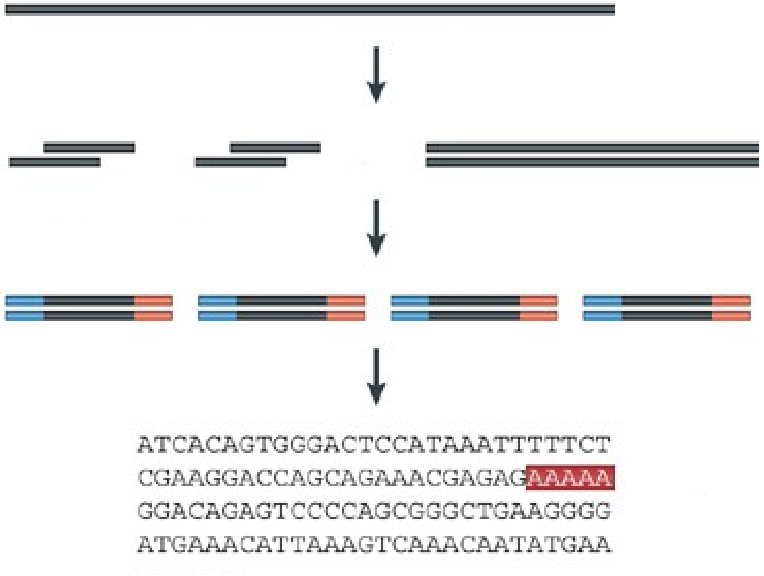
Aligning CBP peaks to calculate H3K4me1 and H3K4me3 binding profiles







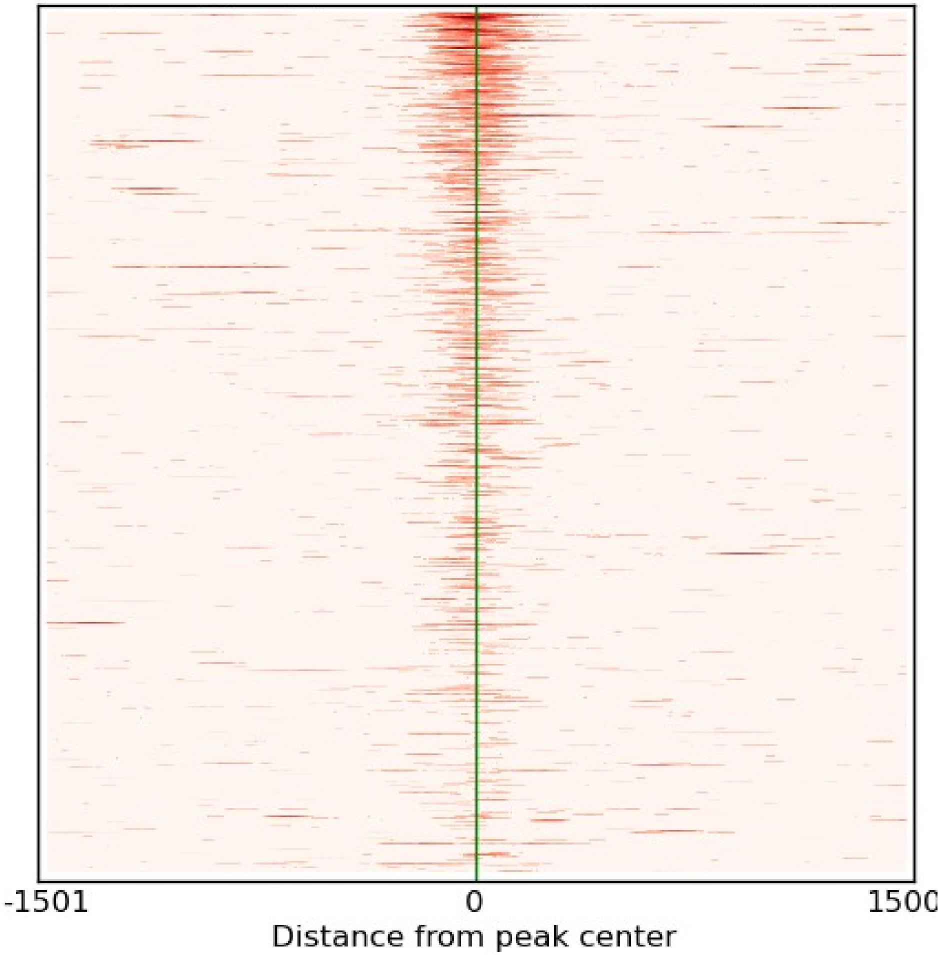
RNA-Seq reveals which parts of the genome are transcribed



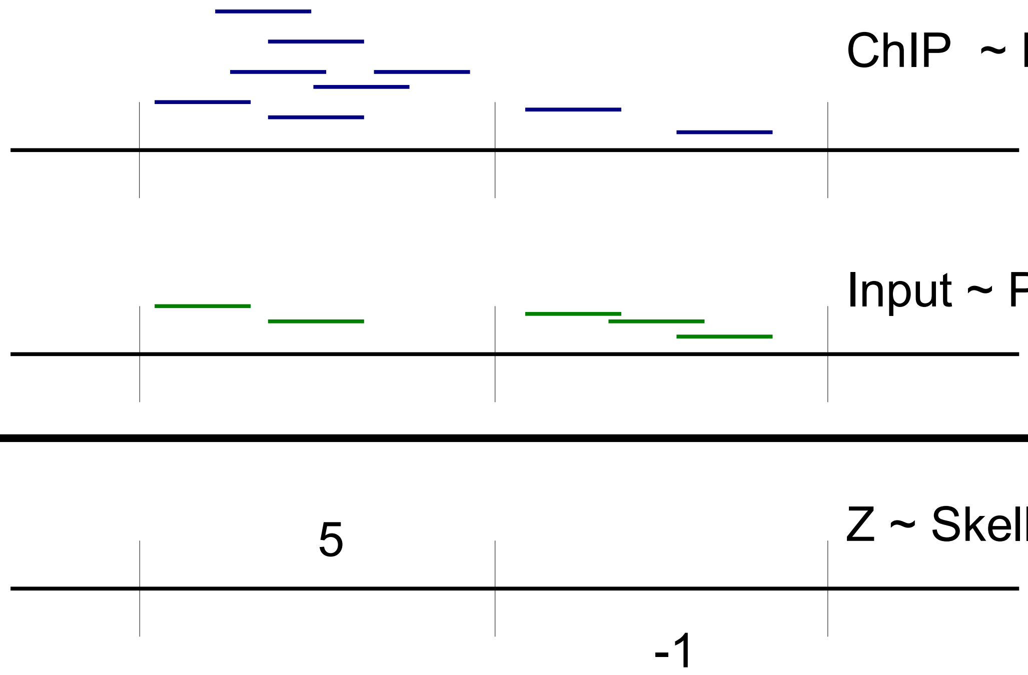
- Fragment
- RNA → cDNA
- 35 bp reads mapped to genome
  - Before and after KCl
  - Total RNA and polyA+

(Wang et al, 2009)

RNAPII binds at activity-dependent enhancers



Identifying regions with larger than expected number of ChIP-Seq reads

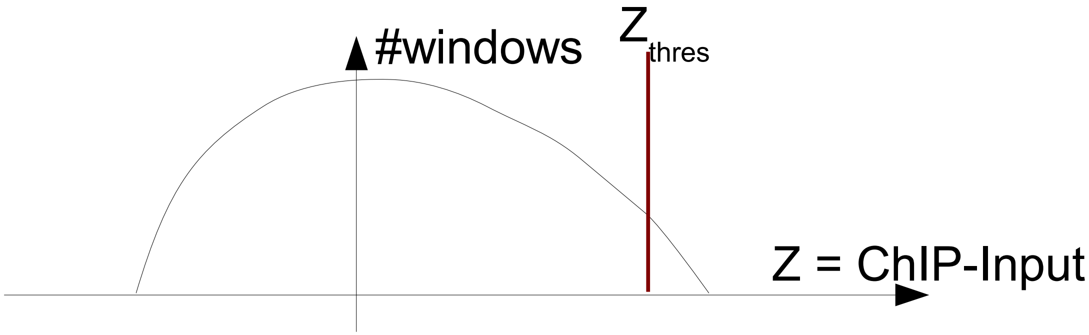


ChIP ~ Poisson(λ_c)

Input ~ Poisson(λ_i)

Z ~ Skellam(λ_c, λ_i)

- False Detection Rate (FDR) determine threshold



Use False Detection Ratio (FDR) to correct for multiple hypotheses

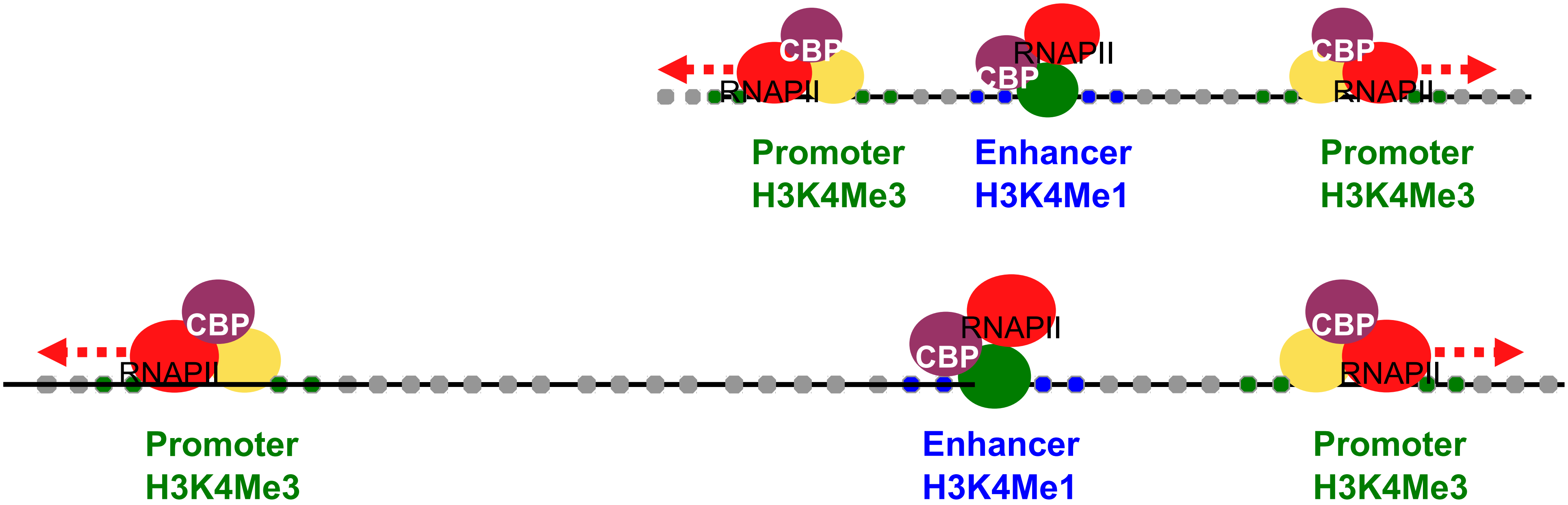
- Z_i = #ChIP reads - #input reads in window *i*
- ~1 read/100 bp
  - Assume #reads in window $P(k) = \lambda^k \exp(-\lambda)/k!$
    - Difference between two Poisson random variables
    - $Z_i \sim \text{Skellam}(z, \lambda_1, \lambda_2)$

$$p(x) = e^{-(\lambda_1+\lambda_2)} (\lambda_1/\lambda_2)^{x/2} I_x(2\sqrt{\lambda_1 \lambda_2})$$

- Millions of windows need to be tested
  - FDR - expected fraction of false positives

Can we learn more about enhancers by comparing their locations in multiple species?

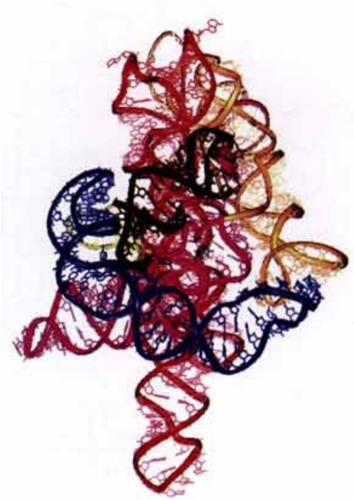
- Conservation of the genomic context of enhancers
- Evolutionary trajectories of enhancers and promoters



What is the structure of non-coding RNAs?

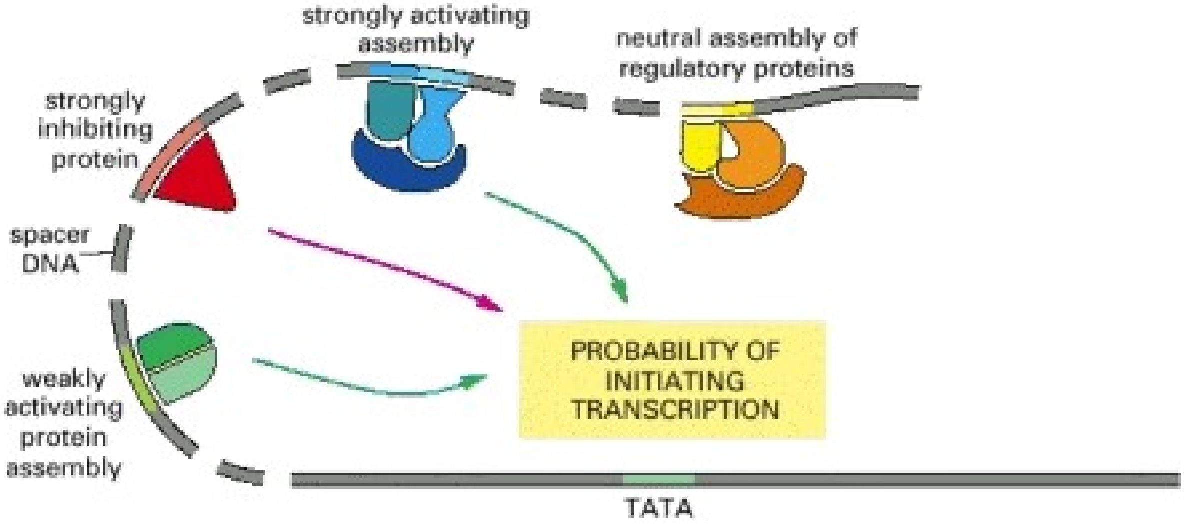
- Many classes of novel RNAs
- Structure → function
  - Structural motifs
  - Families of ncRNAs

......ACGUCCAAAUUCCCUAGGCUCAAGGCAUUCGAUCGGGAUUAUA..... →



Our understanding of gene expression is qualitative

Expression = f(TF1, TF2, ...; Motif1, Motif2, ...)



Conjectured order of events for eRNA

Open chromatin + Histone modifications + Motifs

Stimulus → TF binding

CBP binding

RNAPII binding

Interaction with promoter

Synthesis of eRNA and mRNA