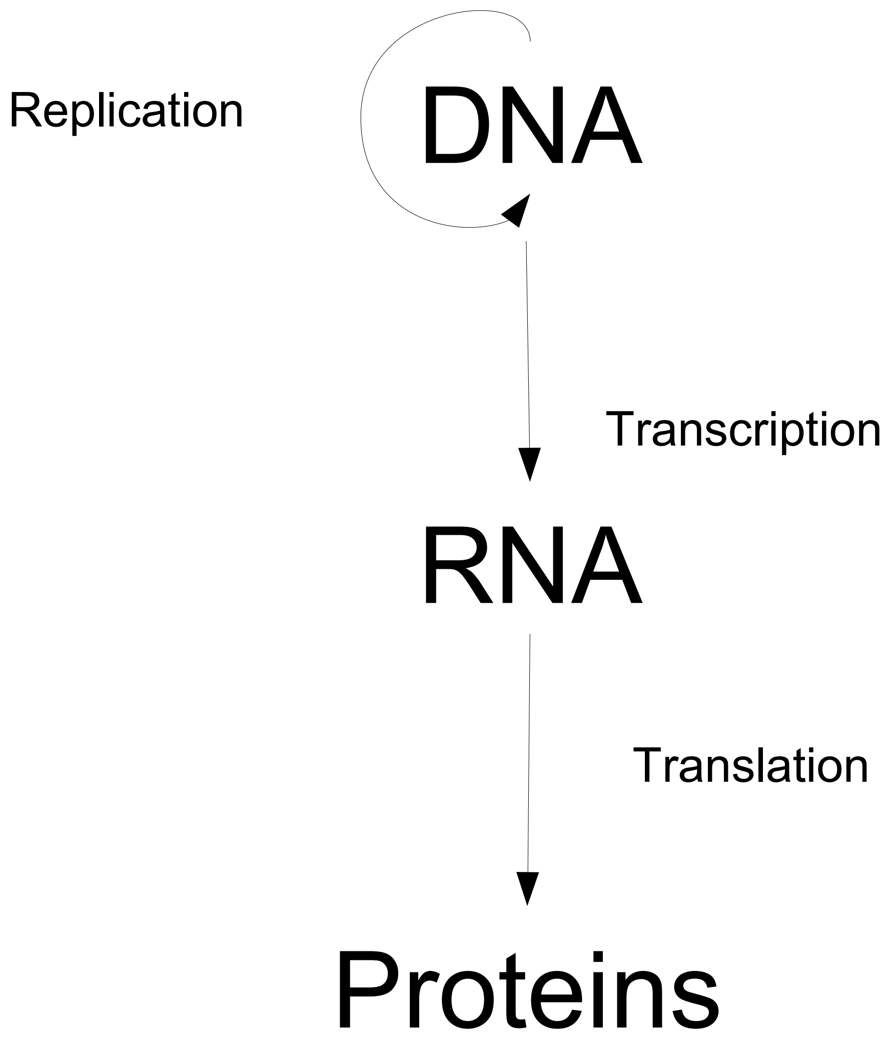
Widespread transcription at activity-dependent neuronal enhancers

Martin Hemberg

Imperial College London
December 12, 2011

Central dogma of molecular biology



Replication

DNA

Transcription

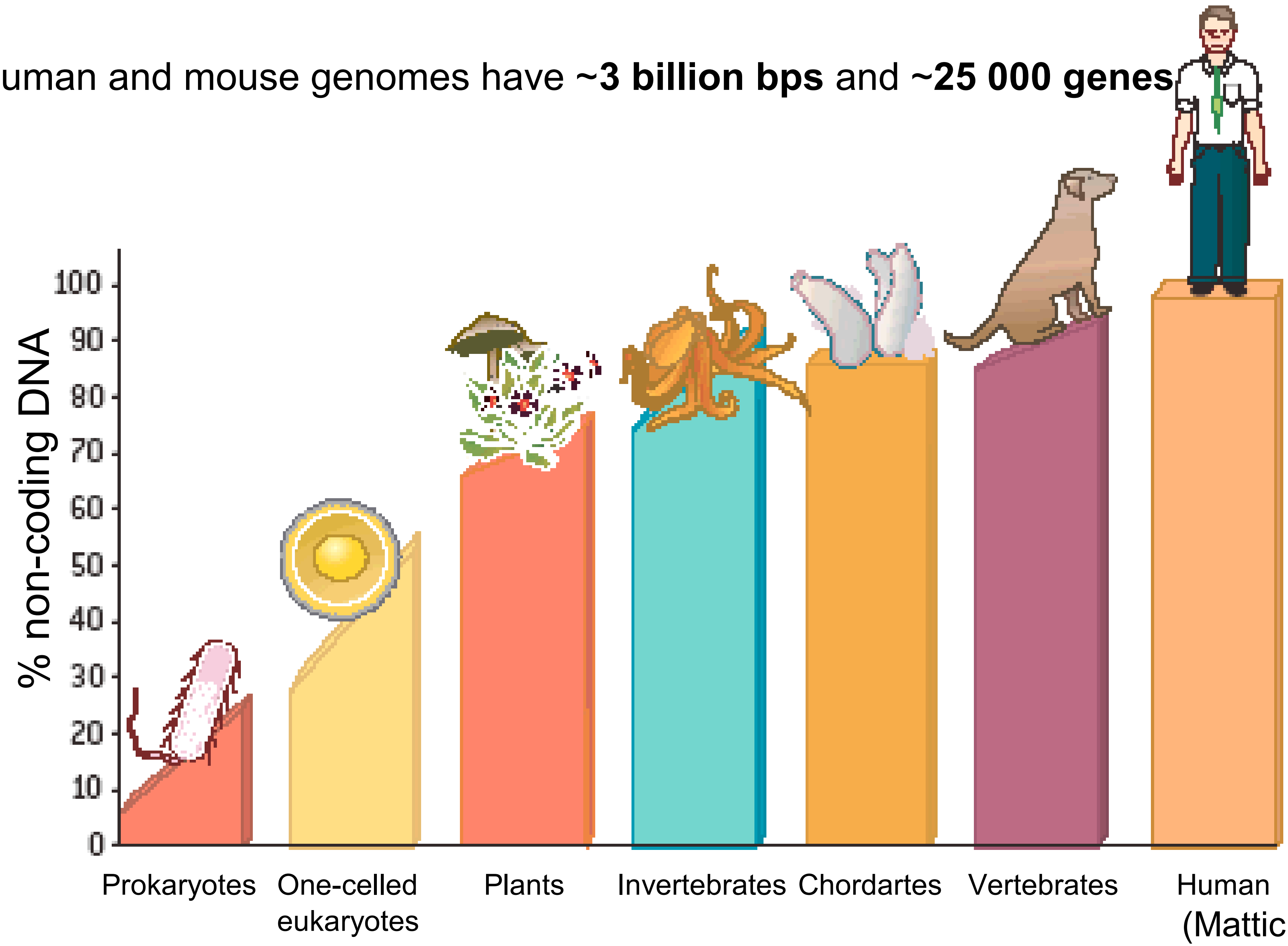
RNA

Translation

Proteins

Most of the genome is **not** protein-coding

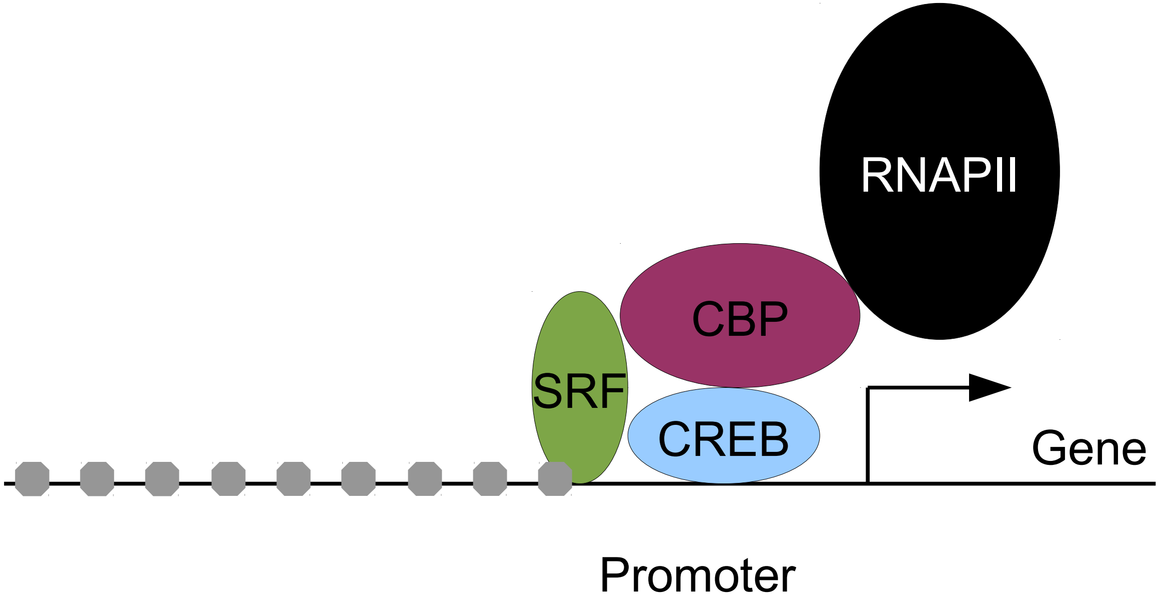
Human and mouse genomes have ~**3 billion bps** and ~**25 000 genes**



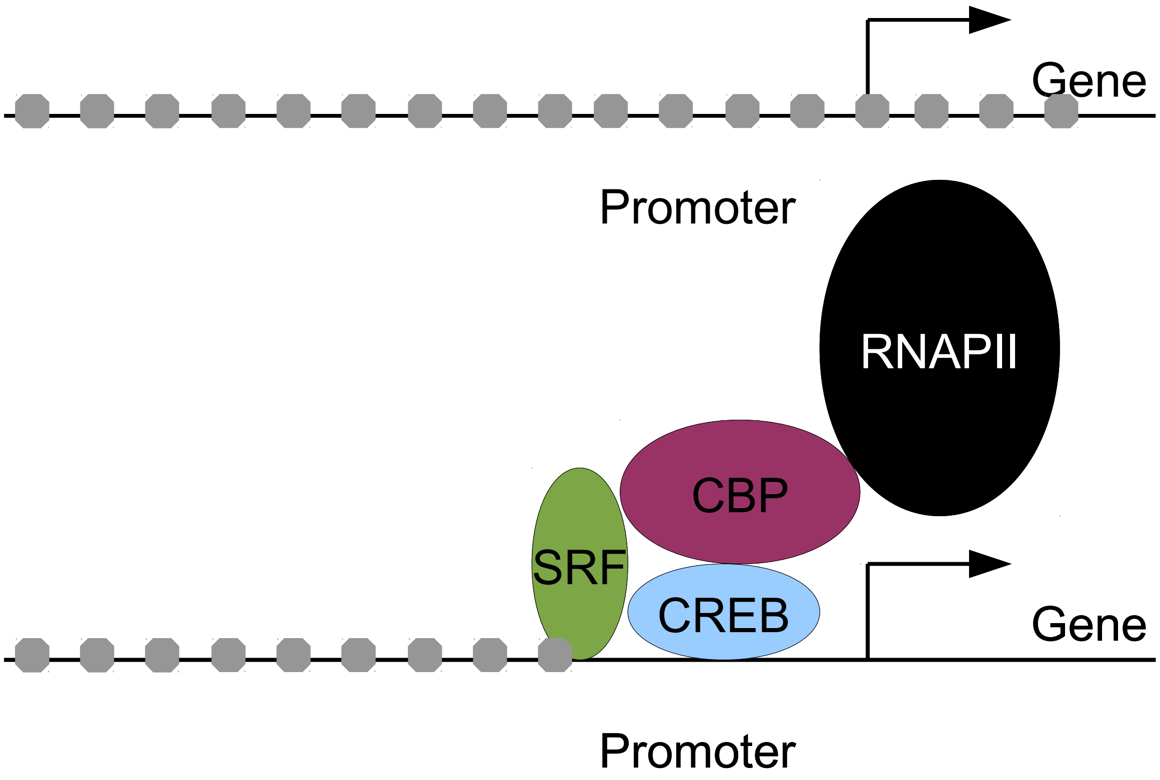
(Mattick, 2004)

Transcription factors (**TFs**) regulate gene expression by binding to the DNA

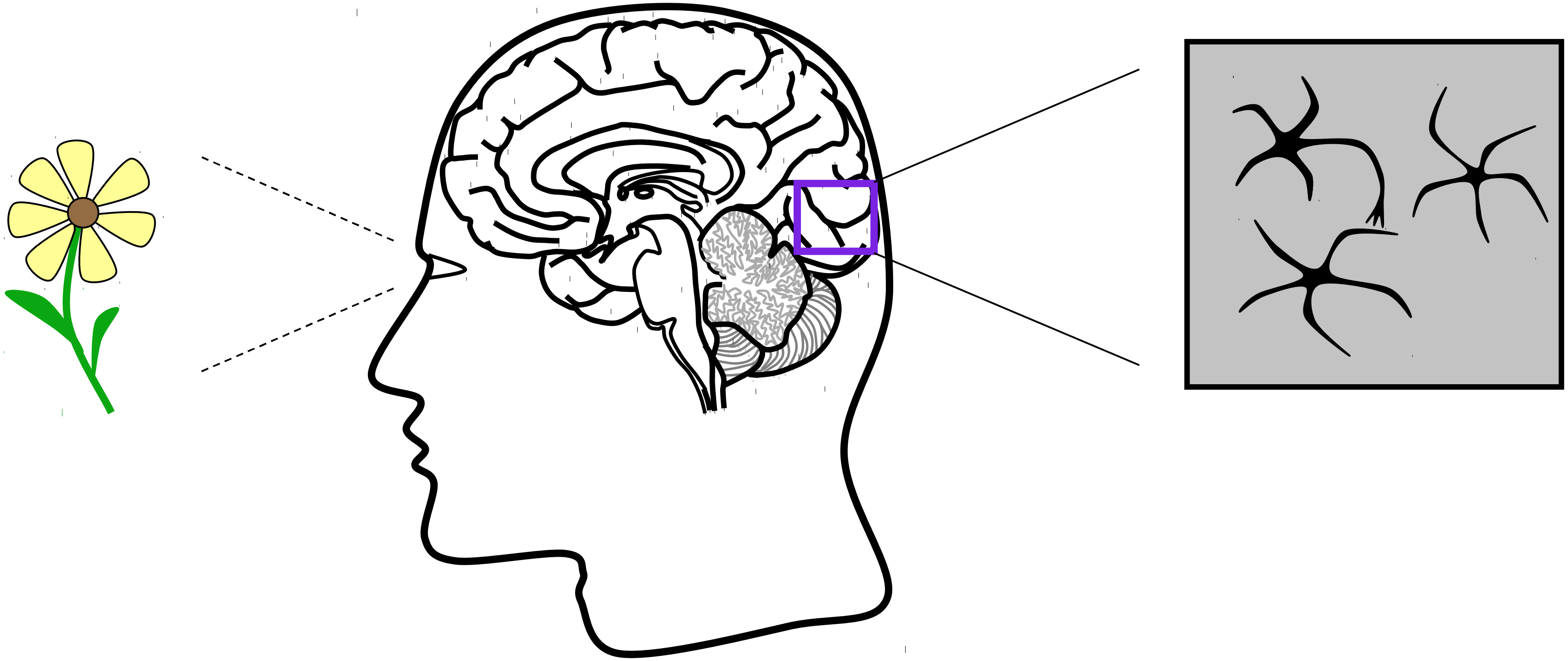
- TFs recruit **RNAPII** for transcription



TF binding is cell-type specific

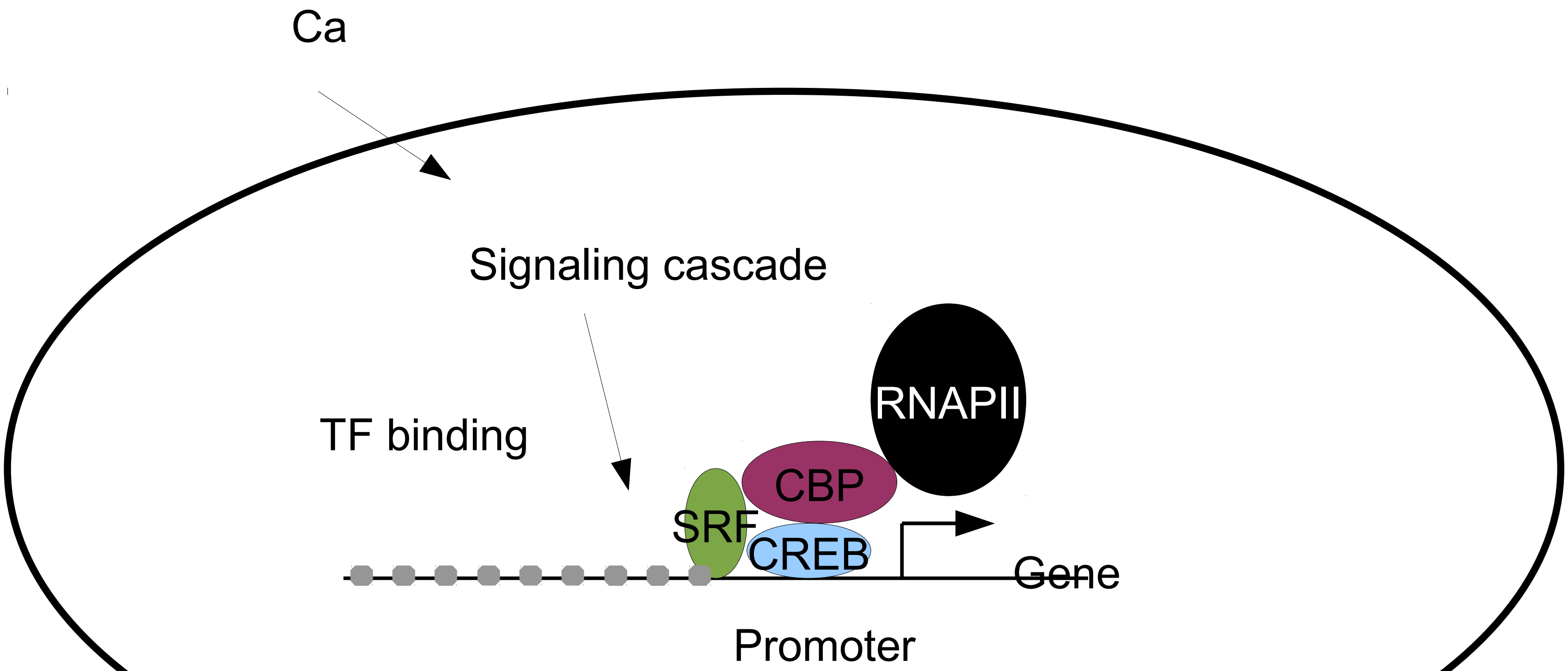


External stimuli change synapses

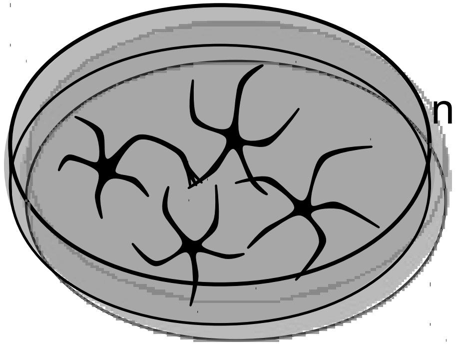


Hubel & Wiesel, 1970’s

Changes in synapses are driven by changes in gene expression



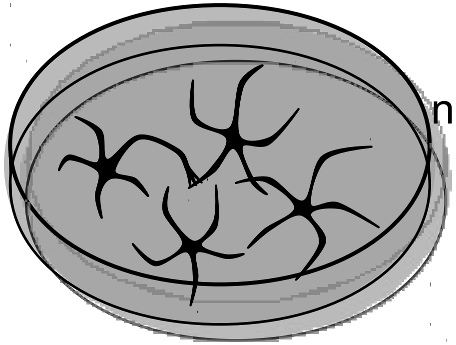
An experimental system for genome-wide study of activity dependent gene expression



mouse cortical
neurons

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

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



mouse cortical
neurons

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

- KCl

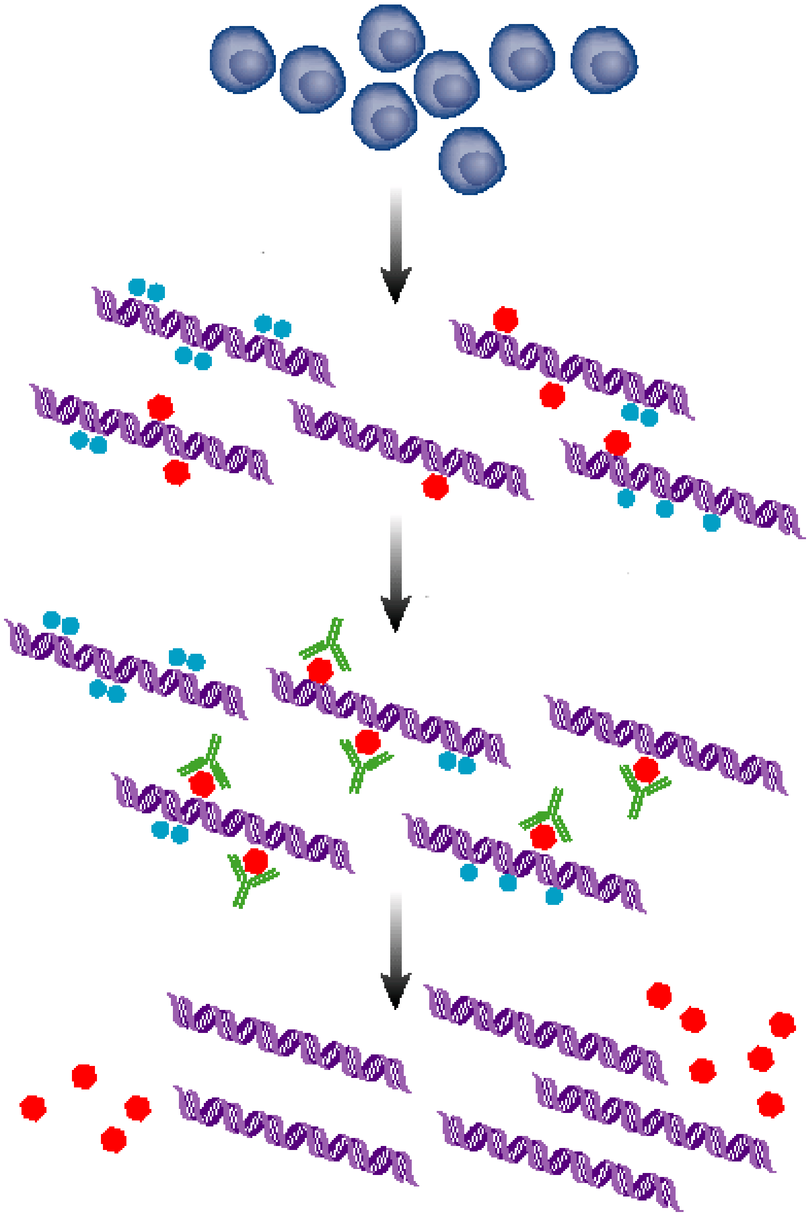
ChIP-Seq
RNA-Seq

+ KCl

ChIP-Seq
RNA-Seq

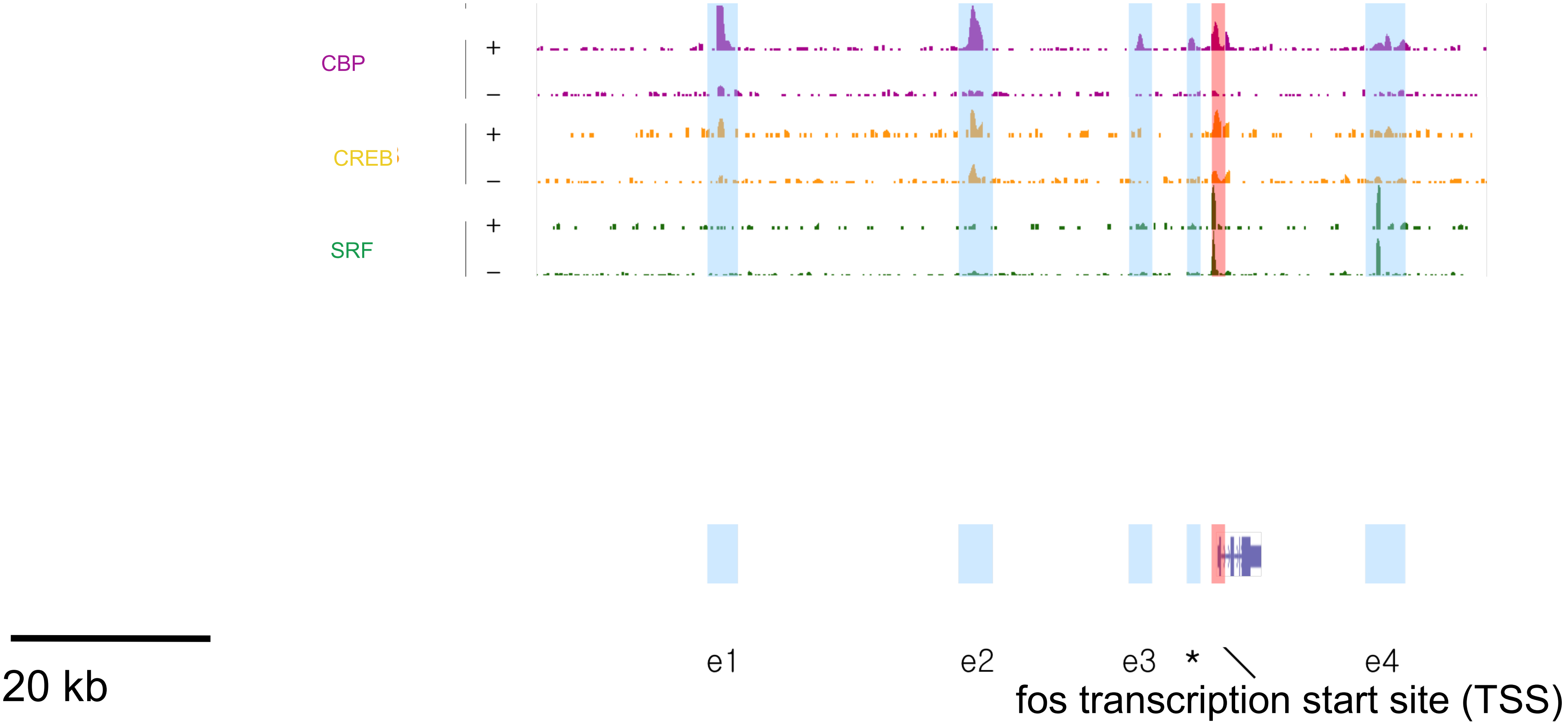
Jesse Gray
Tae-Kyung Kim
Greenberg Lab

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



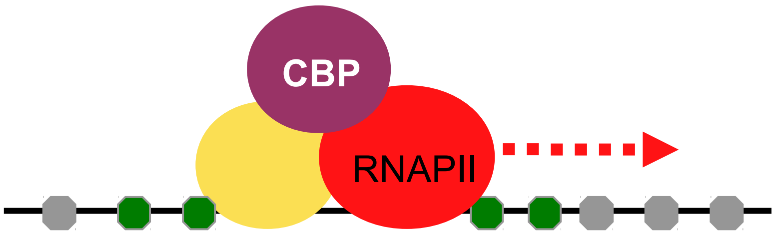
(Mardis, 2007)

Binding of **CBP** depends on activity at the *fos* promoter and flanking loci



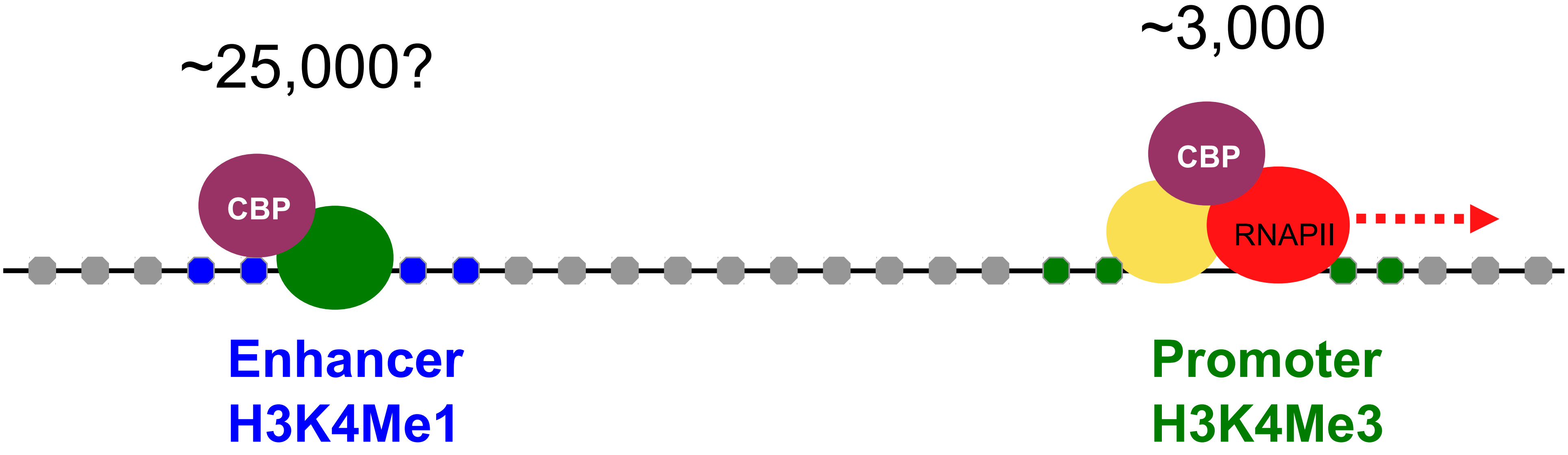
Only ~3000 CBP peaks at promoters

~3,000



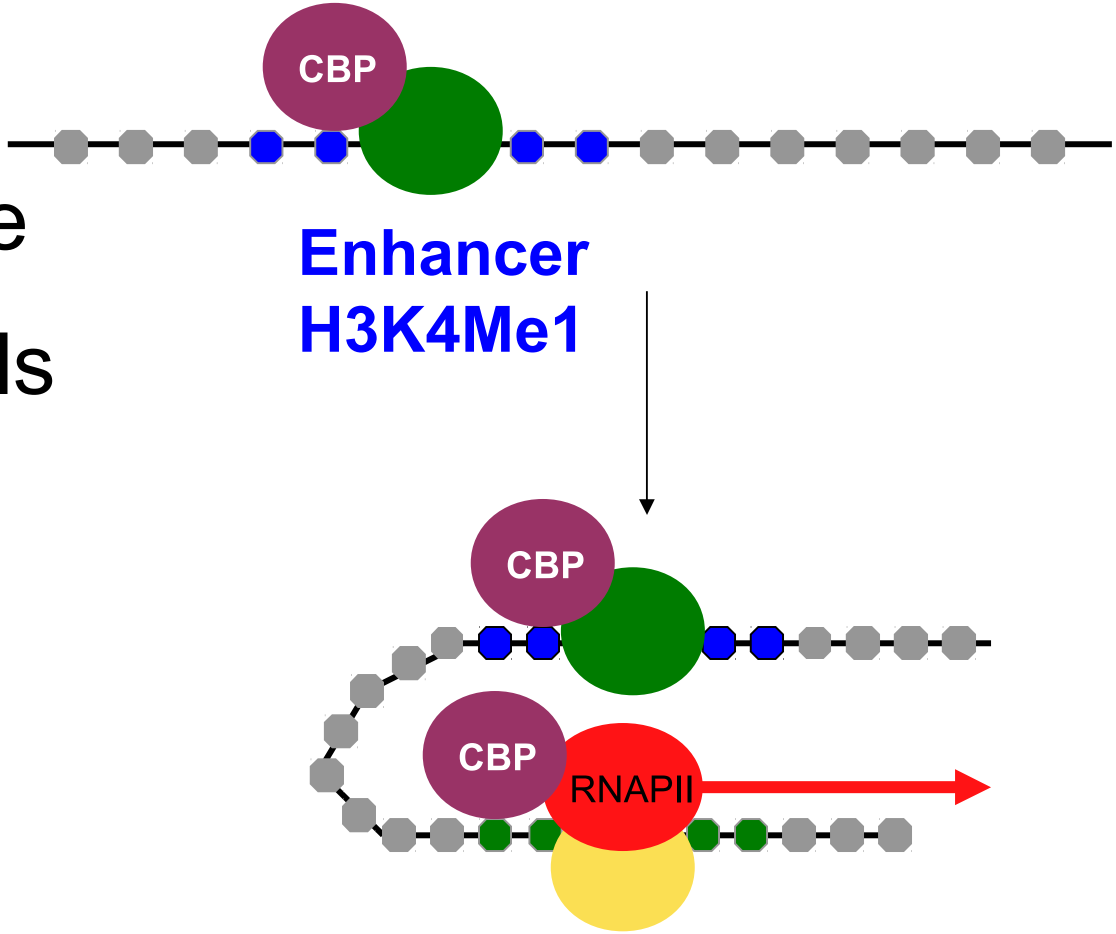
Promoter
H3K4Me3

CBP hypothesized to bind at enhancers



Enhancers are distal TF binding sites

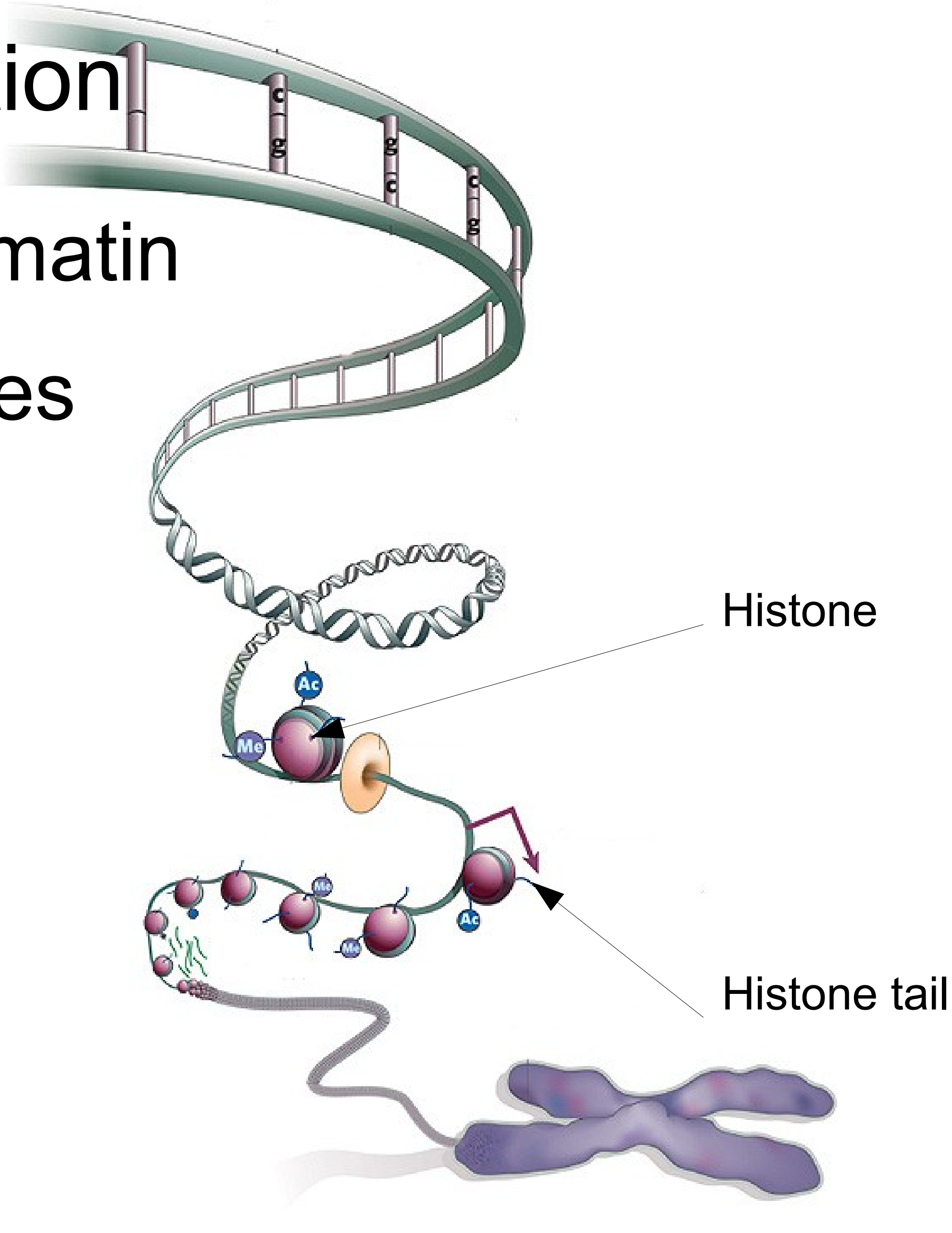
- No universal sequence signature
- Marked by **high** levels of **H3K4me1**



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

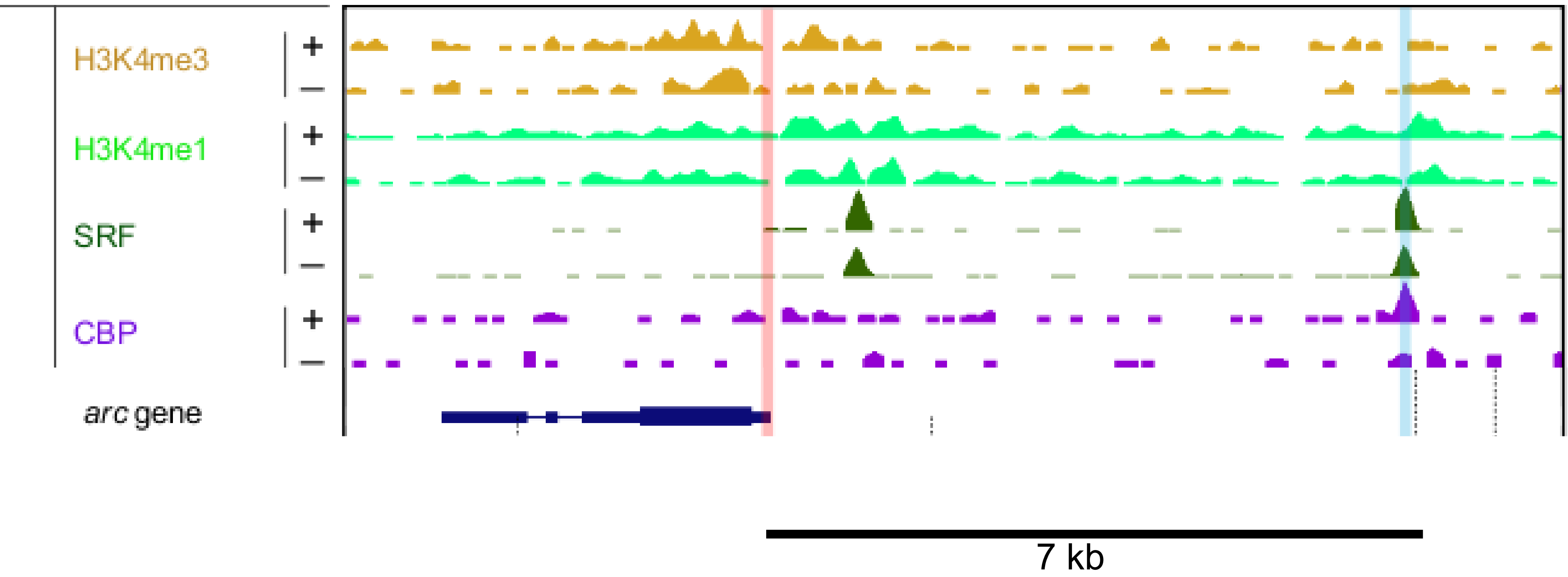
Post-translational modifications of histone tails correlate with function

- **H3K4Me1** – open chromatin
- **H3K4Me3** – active genes

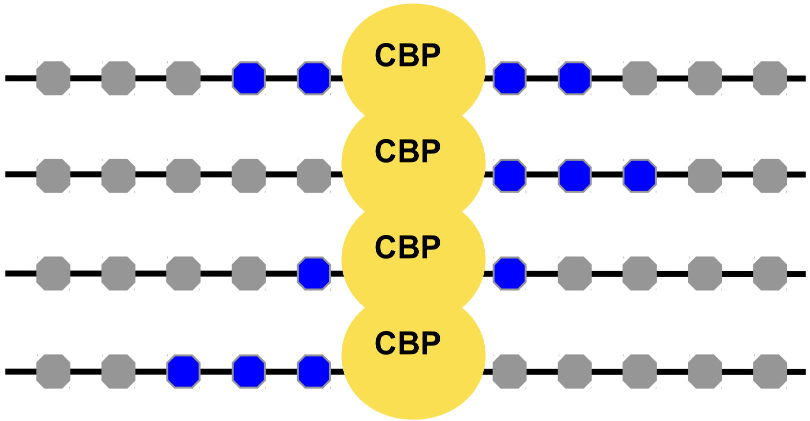


(ENCODE, 2007)

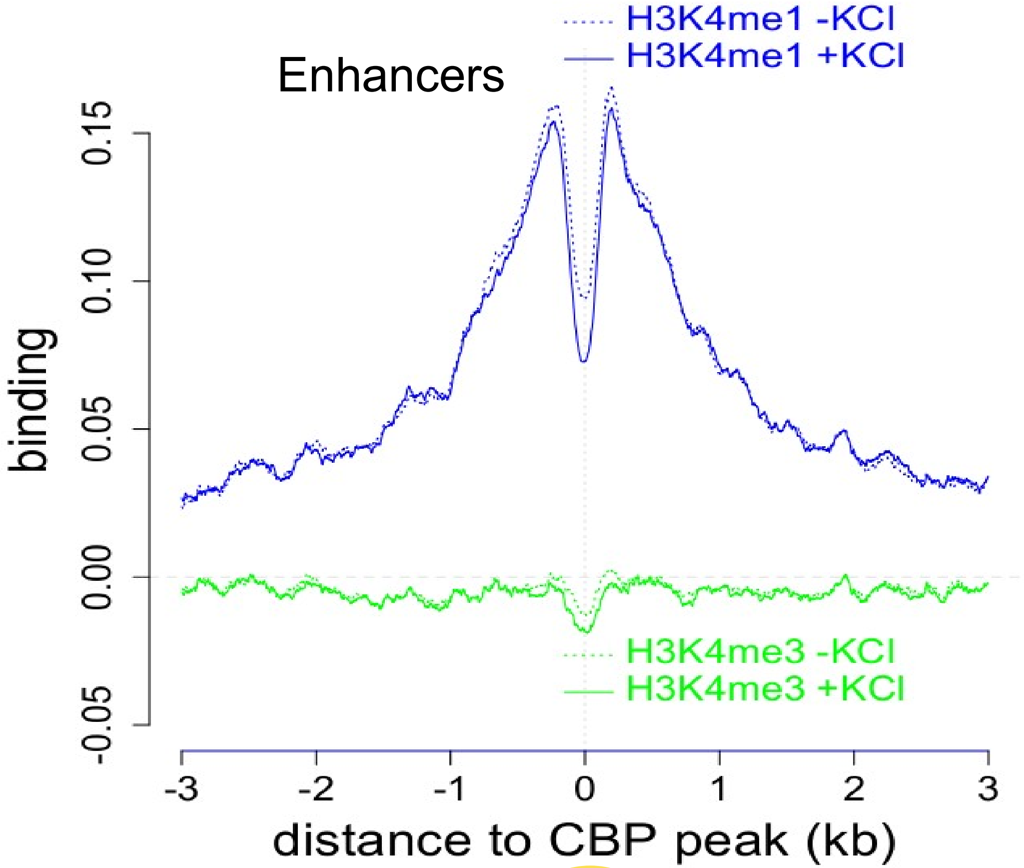
Distal CBP peaks have high levels of H3K4me1 and low levels of H3K4me3

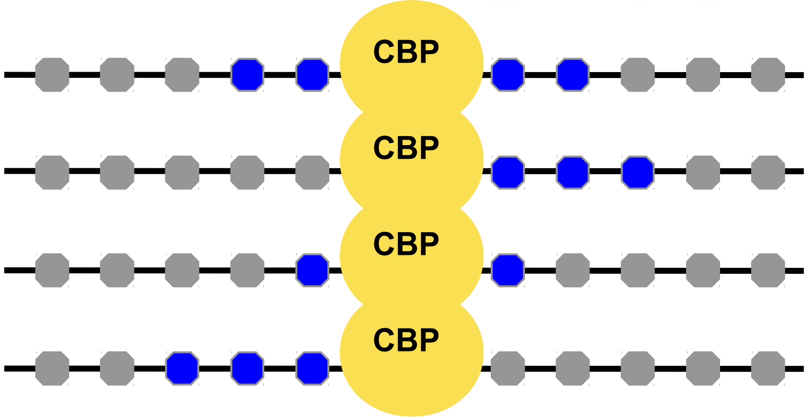


Aligning CBP peaks to calculate binding profiles

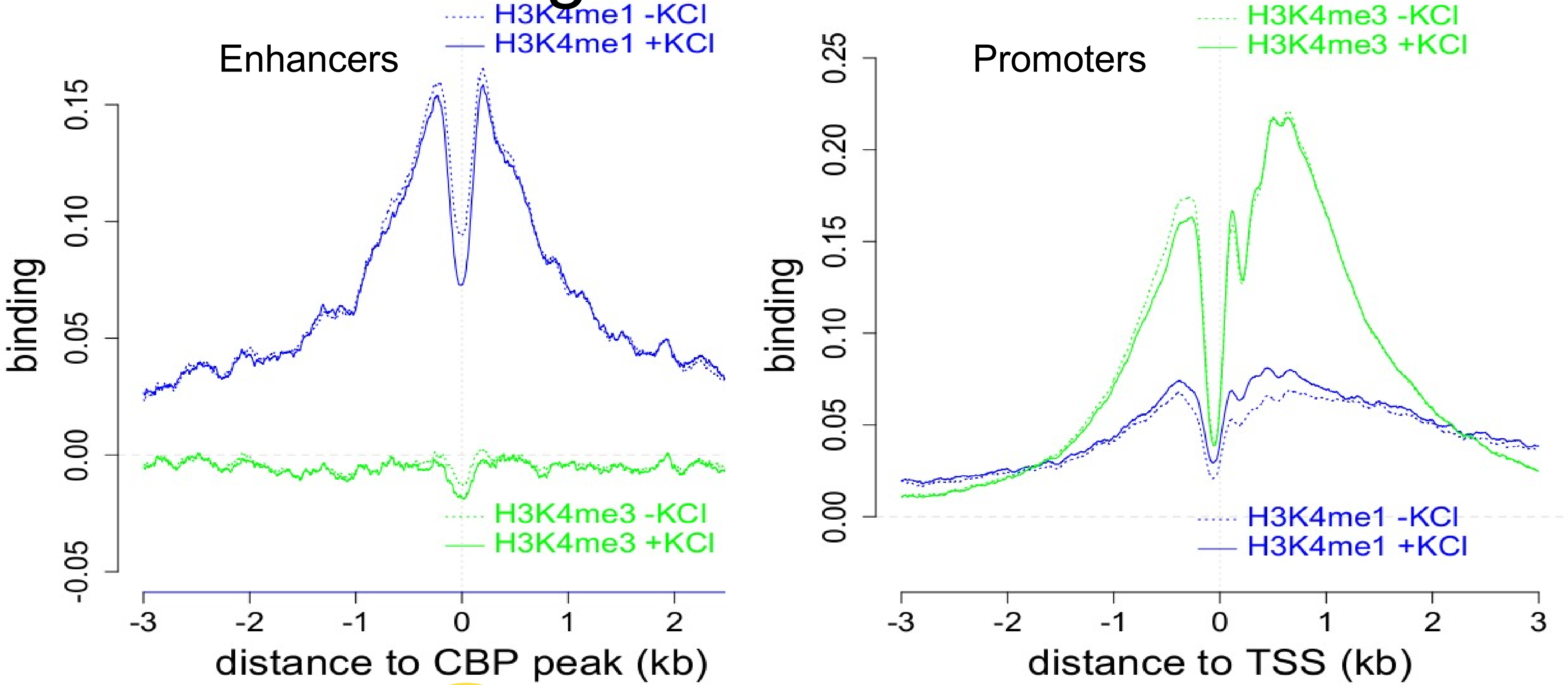


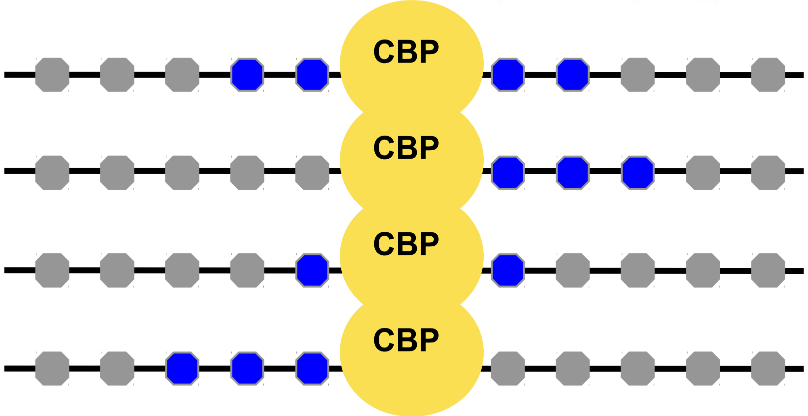
Enhancers have high levels of H3K4me1 and low levels of H3K4me3





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





Criteria for identifying activity-dependent enhancers

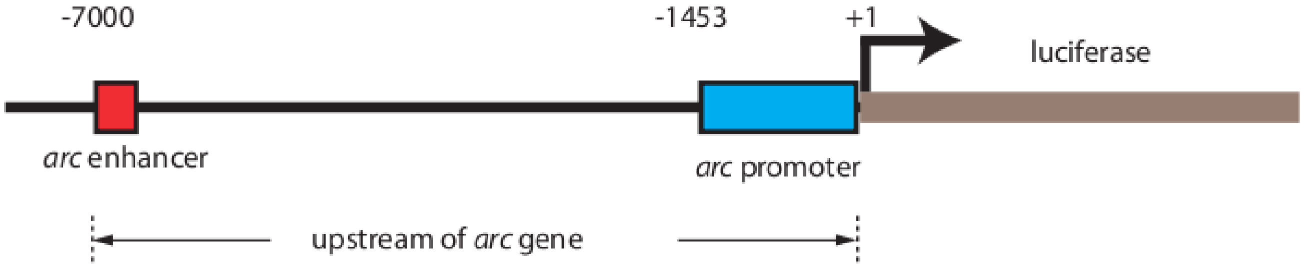
- **CBP** peak
- **High** levels of flanking **H3K4me1**
- **Low** levels of **H3K4me3**

We identified 12k activity-dependent enhancers throughout the genome

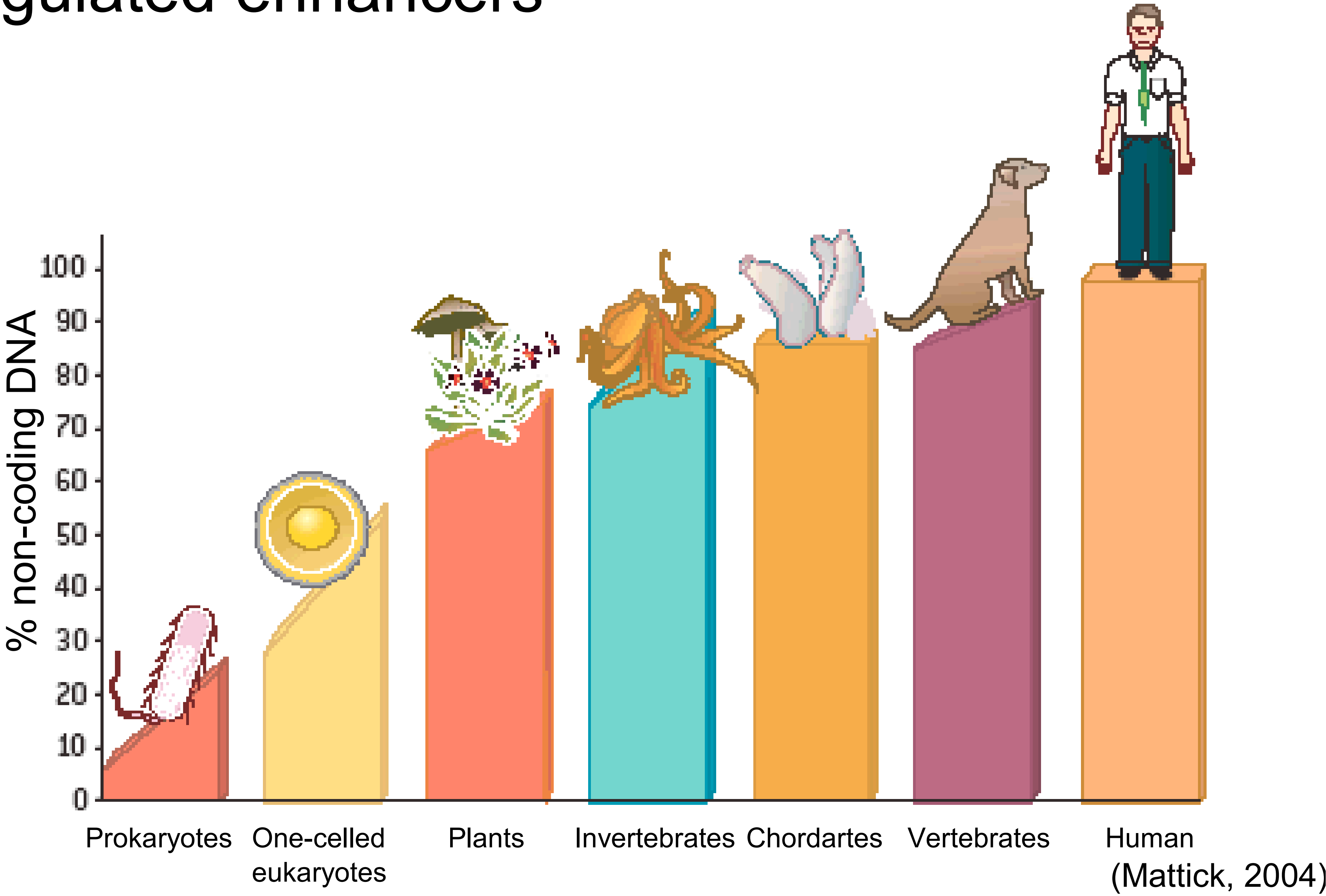
- **CBP** peak
- **High** levels of flanking **H3K4me1**
- **Low** levels of **H3K4me3**
    - ~**5000** extragenic enhancers
    - ~7000 intragenic enhancers

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

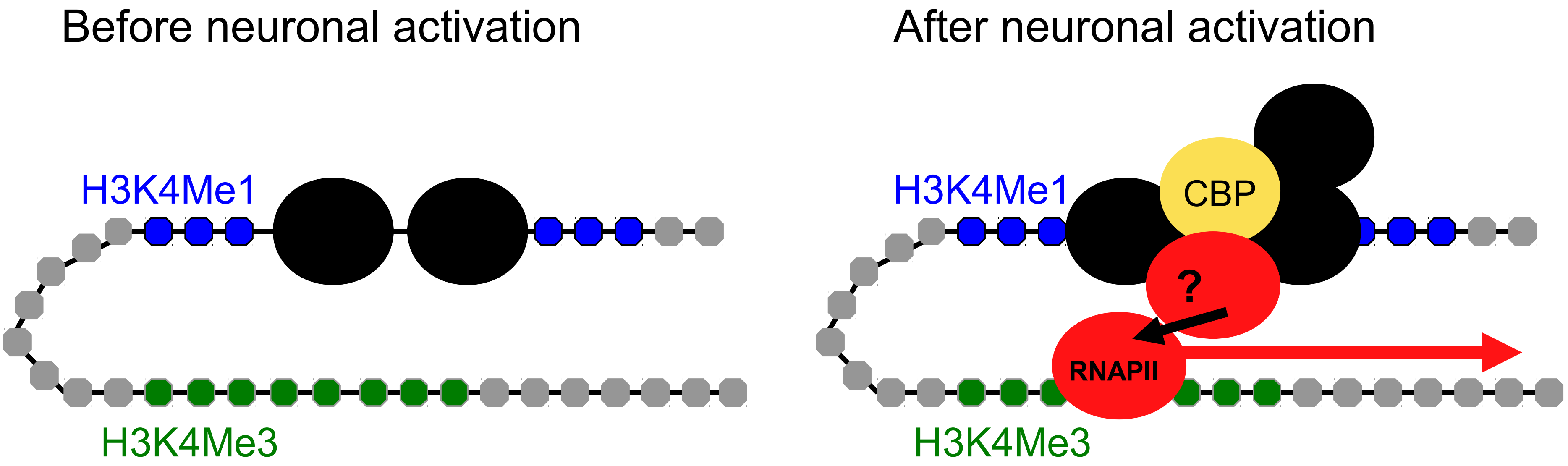


Non-coding DNA serves as activity-regulated enhancers



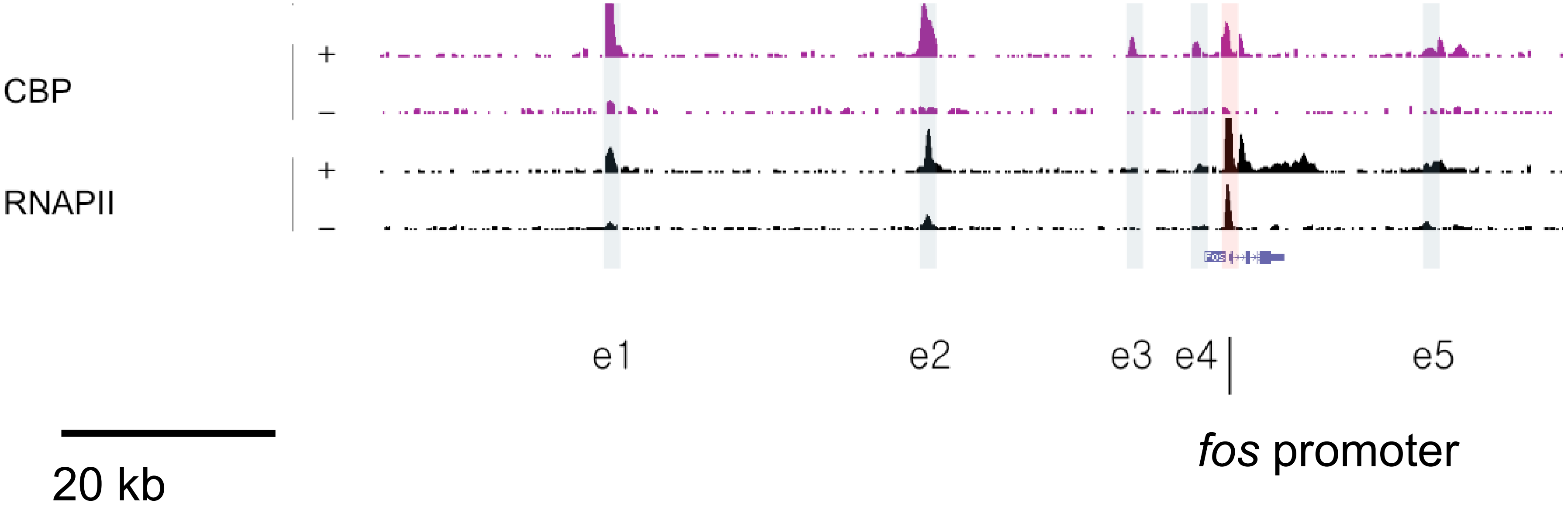
(Mattick, 2004)

Is RNAPII recruited at enhancers?

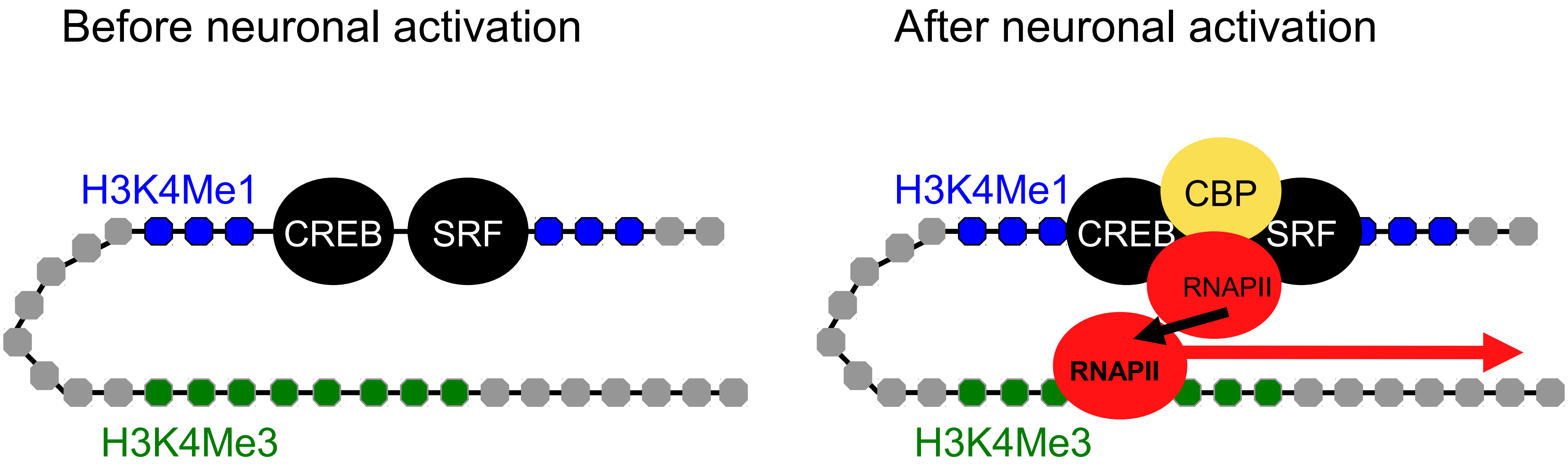


- Does RNAPII bind at enhancers?

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



What is the function of RNAPII at enhancers?

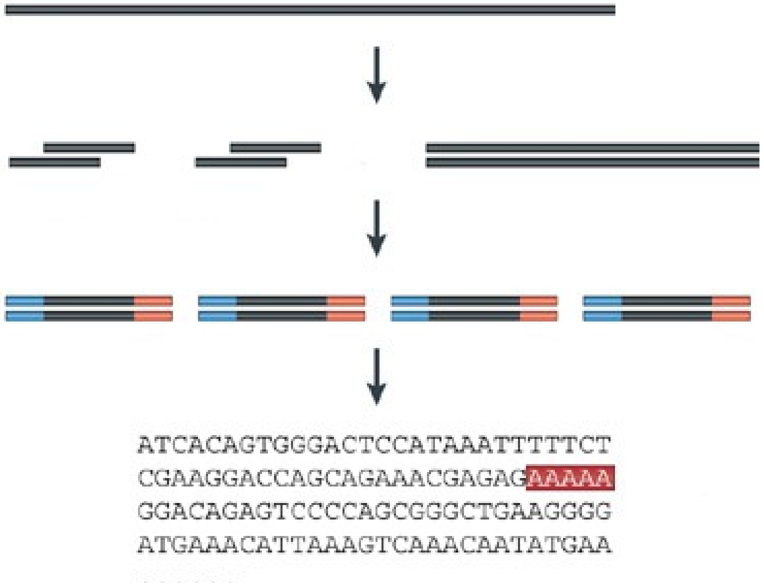


Before neuronal activation

After neuronal activation

- Does RNAPII bind at enhancers? YES
- Are transcripts produced at enhancers?

RNA-Seq reveals which parts of the genome are transcribed



```
ATCACAGTGGGACTCCATAAATTTTTCT
CGAAGGACCAGCAGAAACGAGAGAAAAA
GGACAGAGTCCCCAGCGGGCTGAAGGGG
ATGAAACATTAAAGTCAAACAATATGAA
```

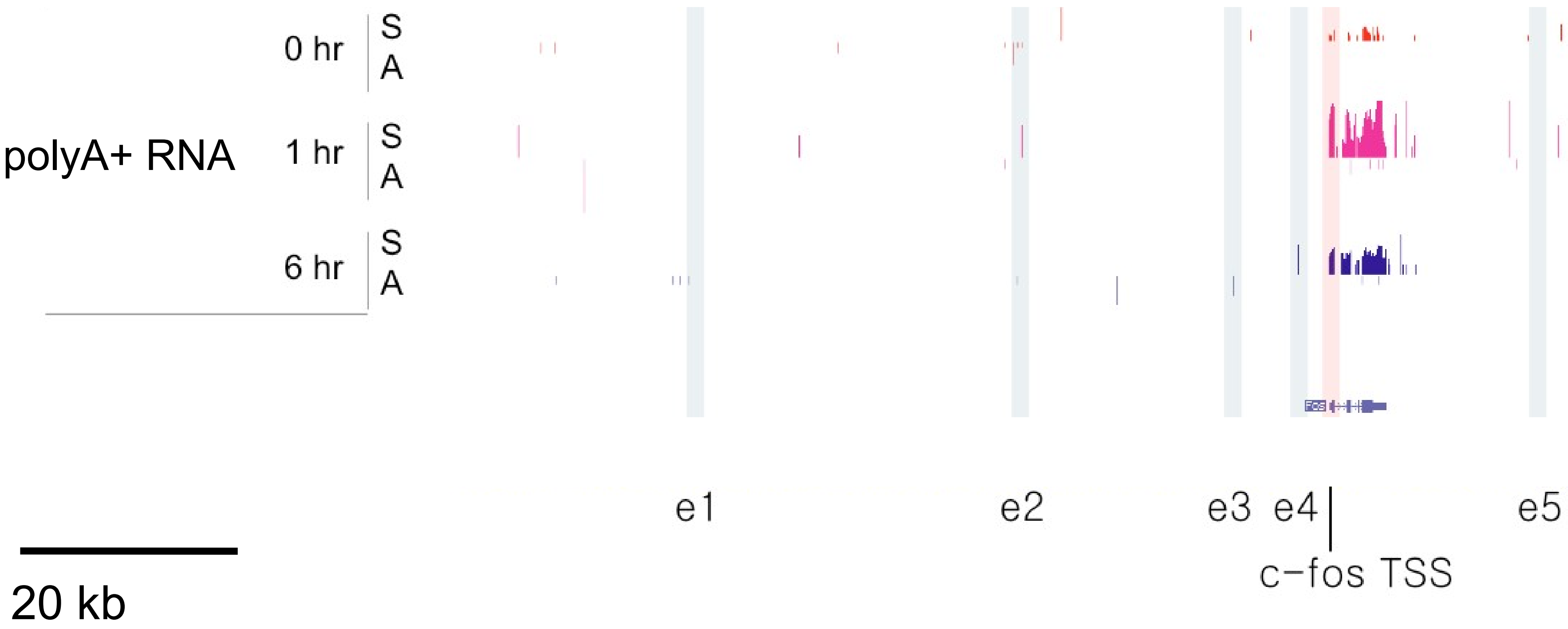
(Wang et al, 2009)

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

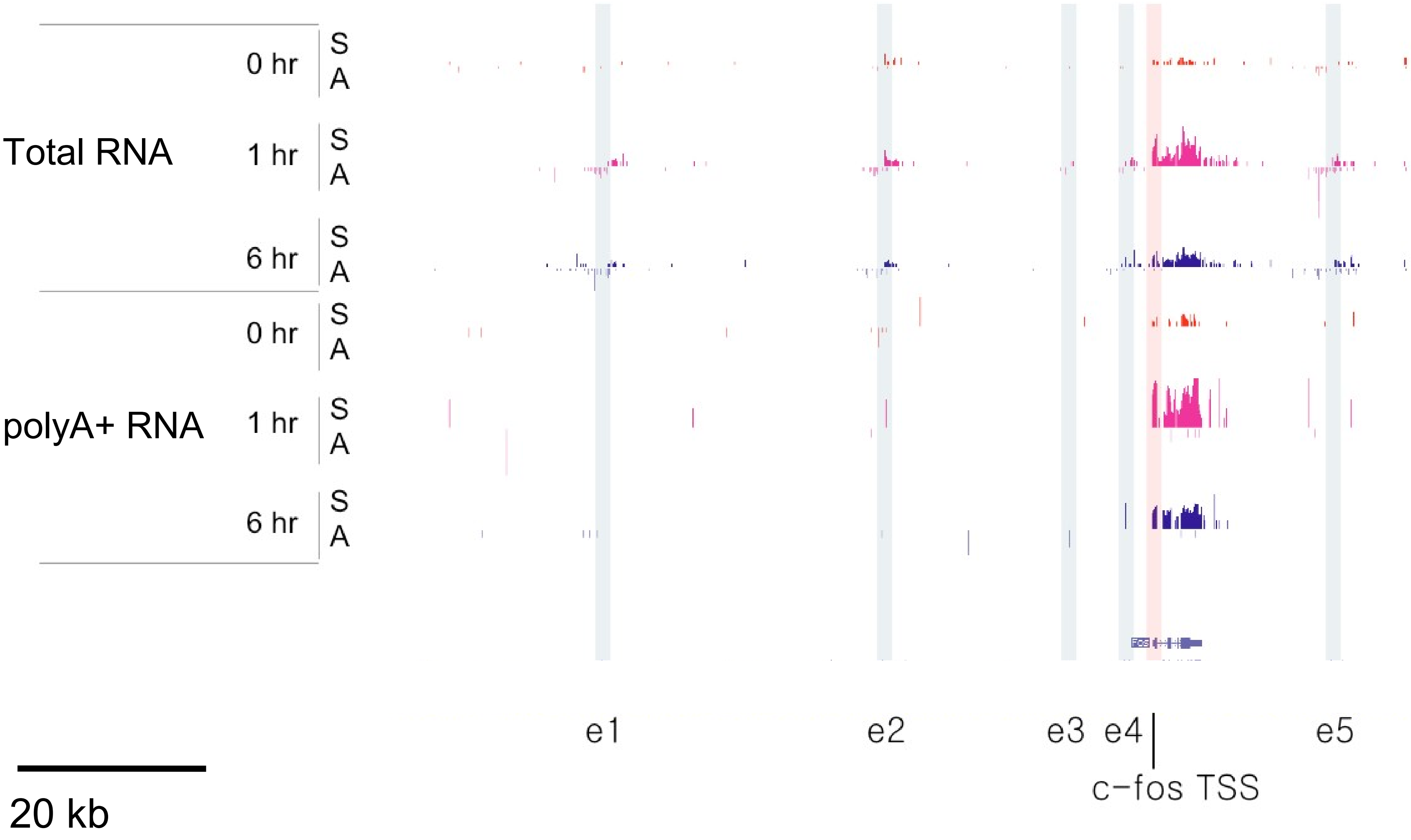
- Increases stability
- Allows transport out of nucleus

ACGUUUGUACCUAGCUAGCUUACGAG AAAAAAAAAAAAAAAAAAAA

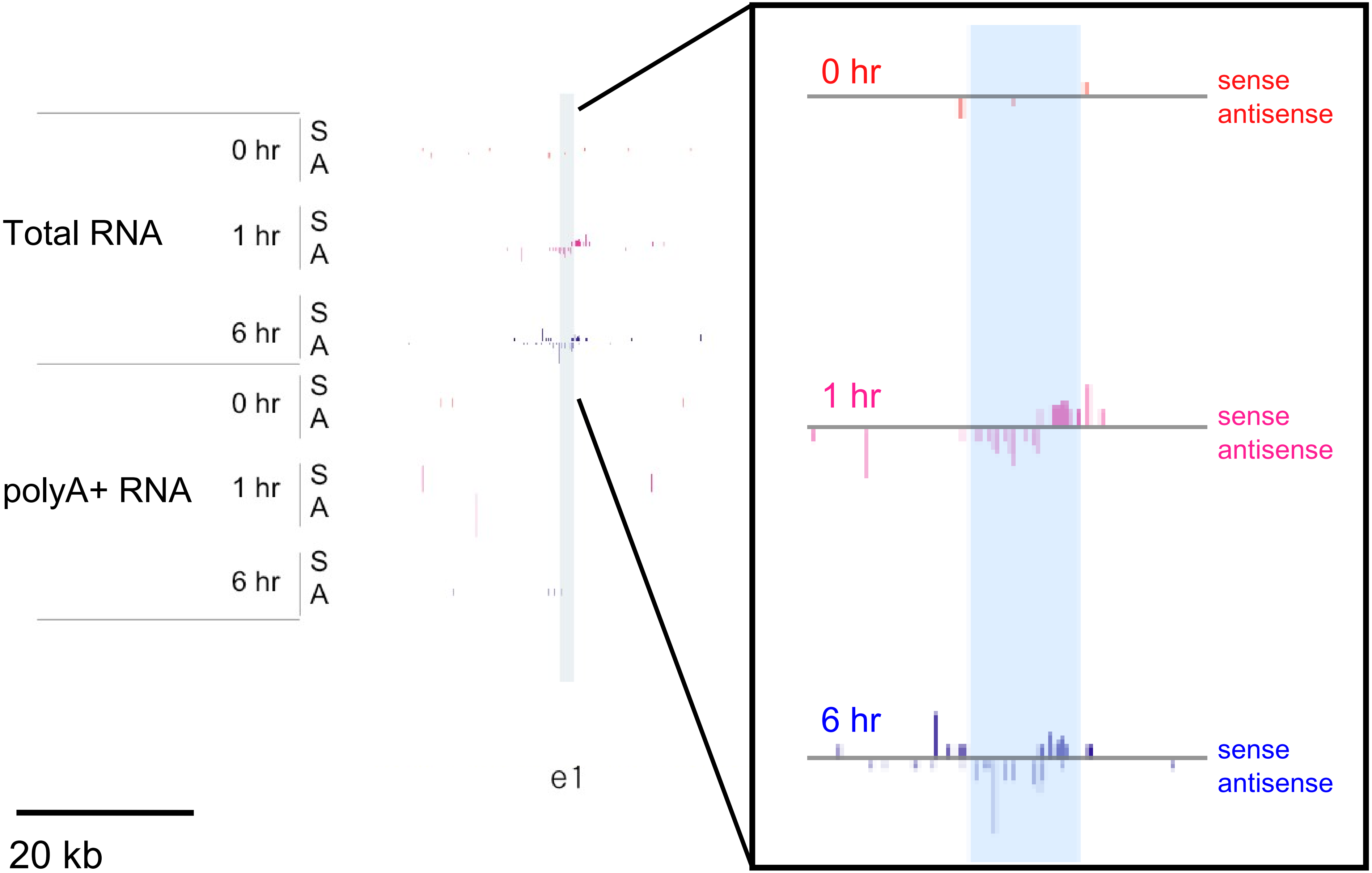
Transcription of mRNA at the *fos* locus



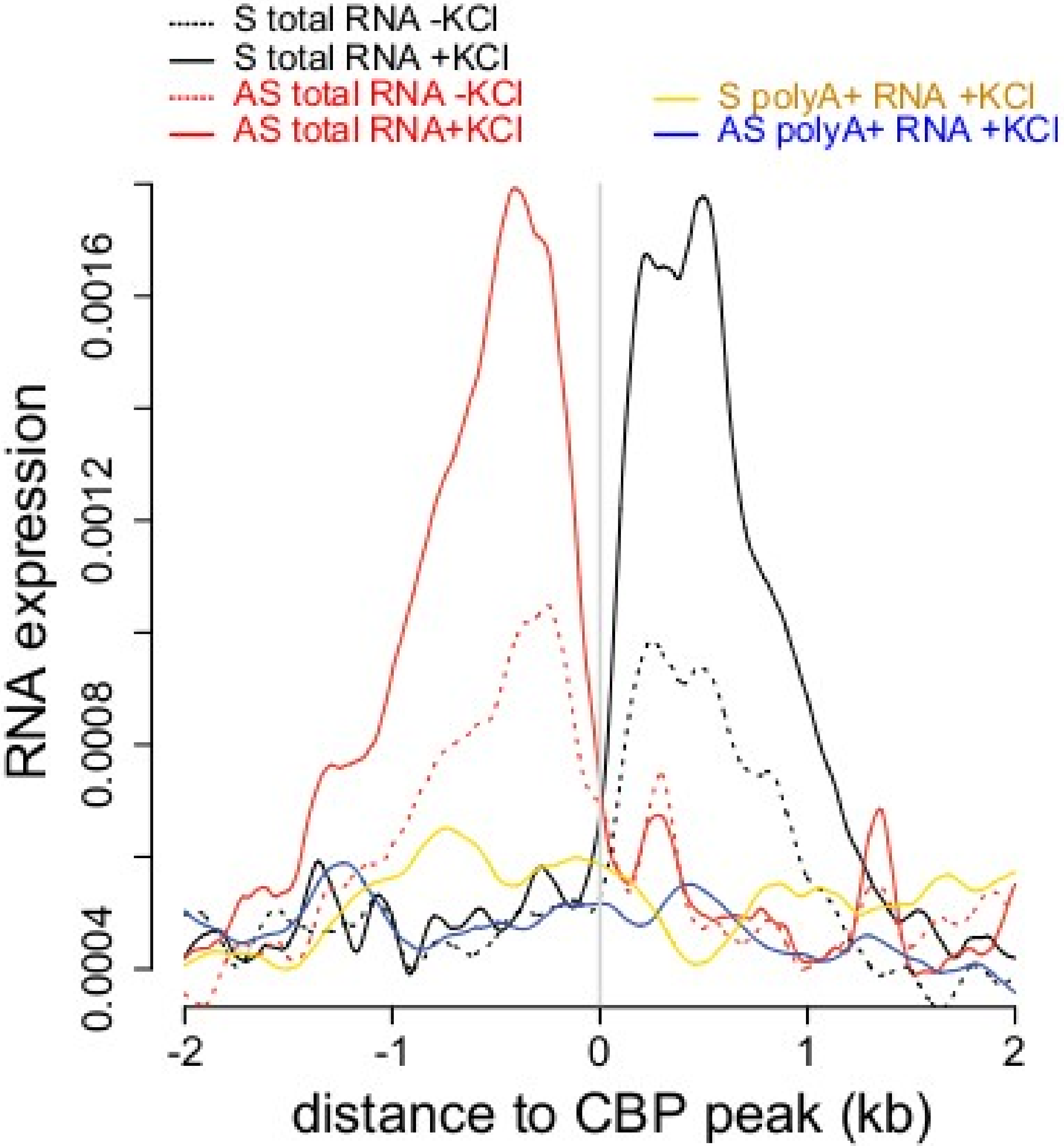
Transcription of total RNA at the *fos* locus



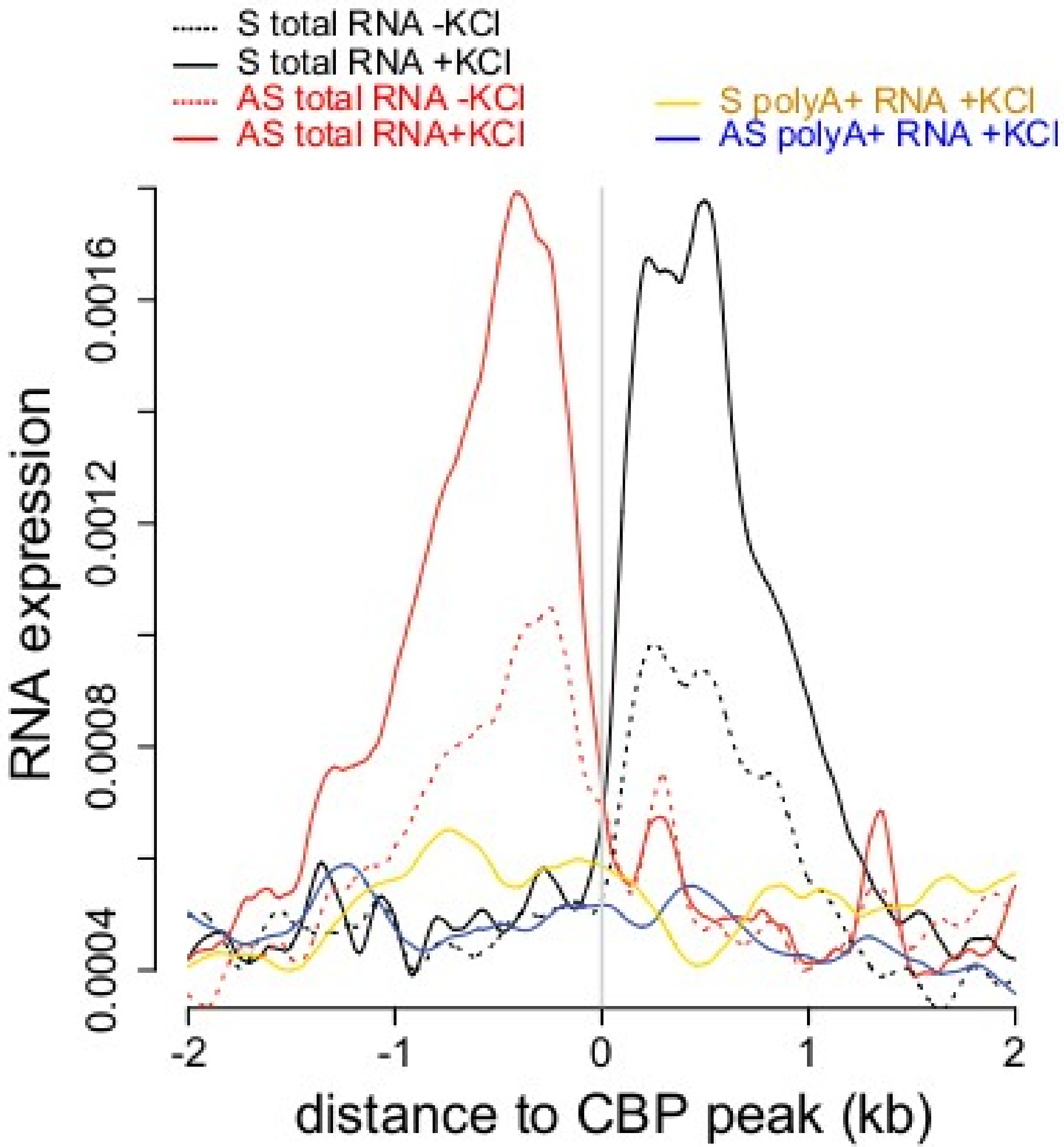
Transcription at enhancers is activity-dependent



Average profile of enhancer RNAs (eRNAs)



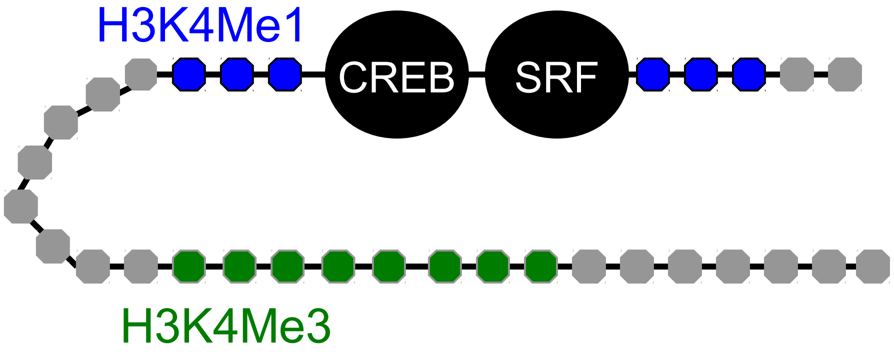
Properties of enhancer RNAs



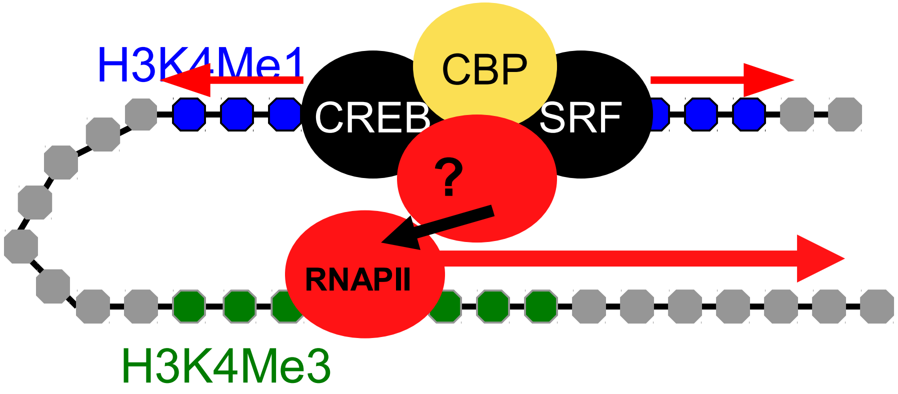
- Inducible
- Low expression
- ~1.5 kb
- Bidirectional
- No polyA-tail
- Not protein-coding

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

Before neuronal activation



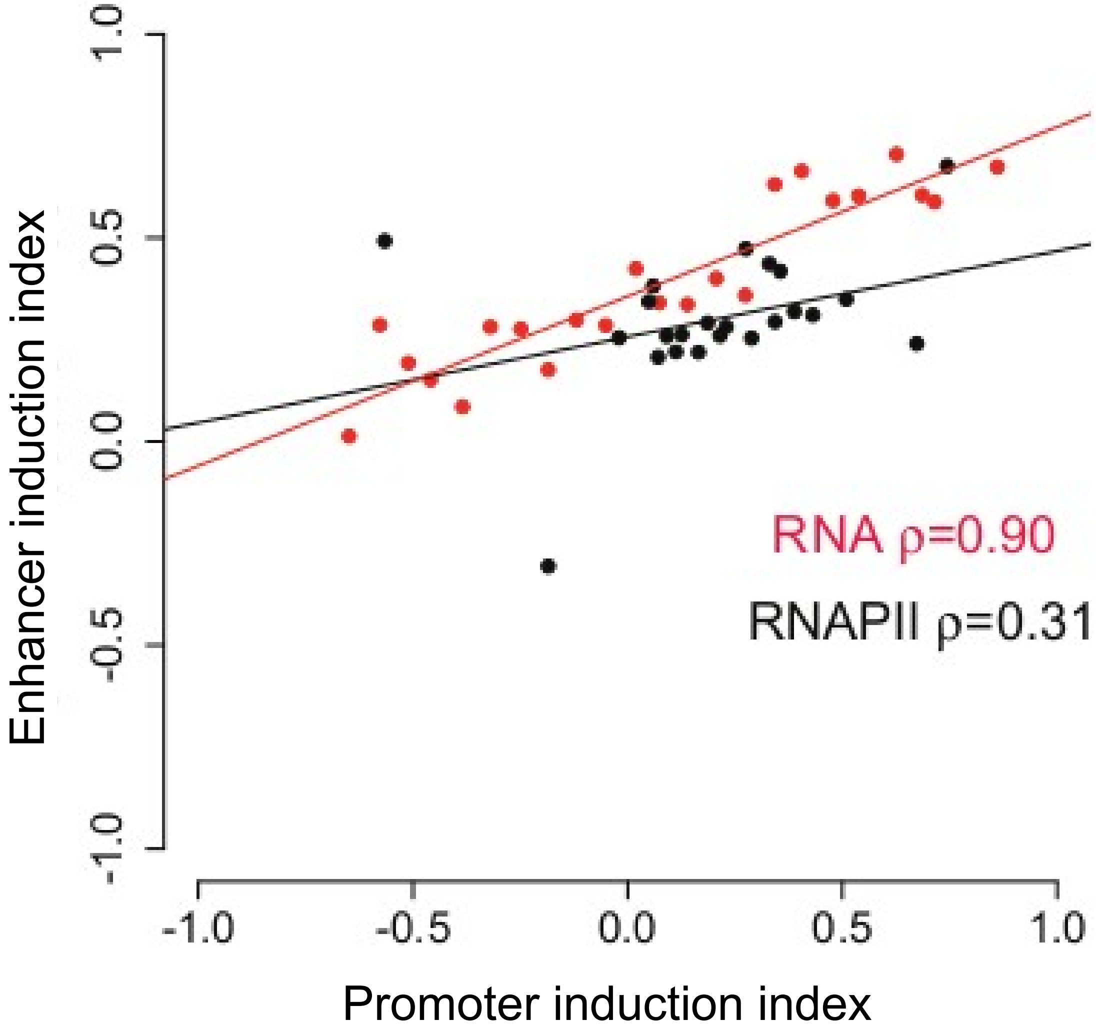
After neuronal activation



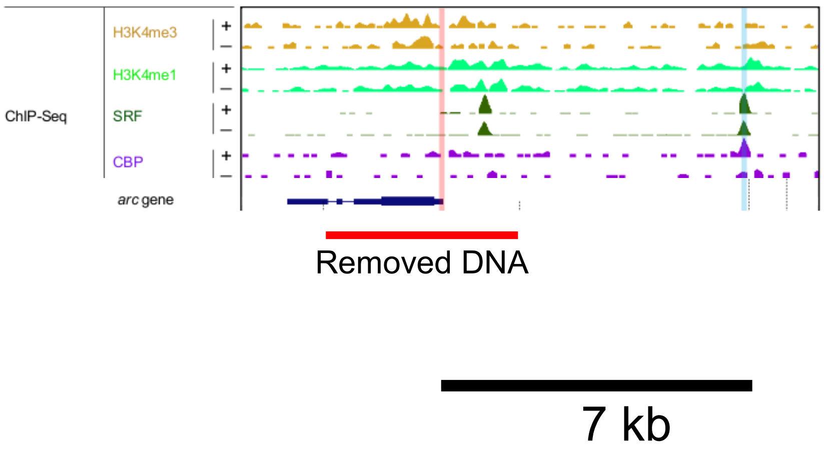
- Does RNAPII bind at enhancers? YES
- Are transcripts produced at enhancers? YES
- Is RNAPII recruitment independent?

eRNA induction is correlated with induction of nearby mRNAs

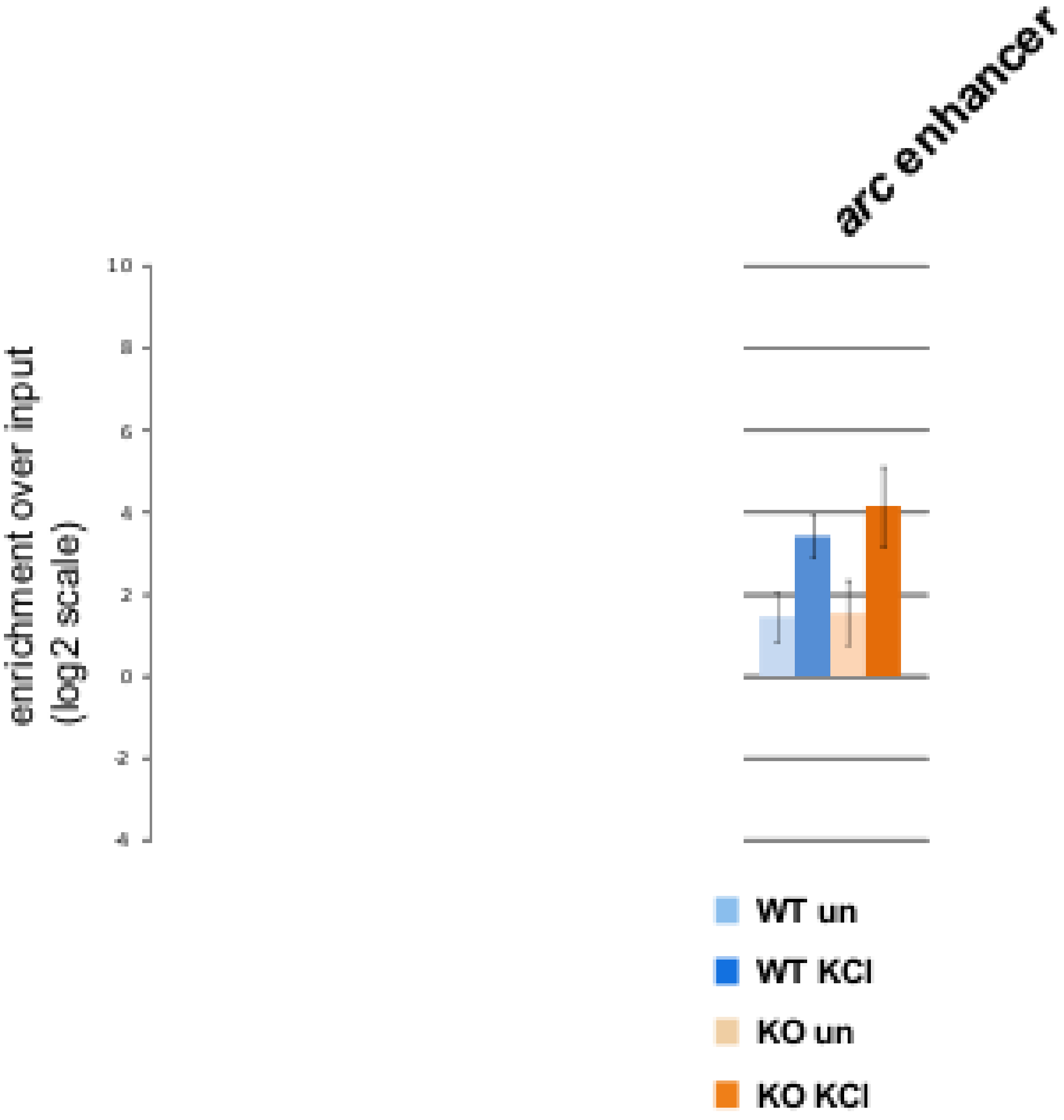
$$induction\ index = \frac{KCl - un}{KCl + un}$$

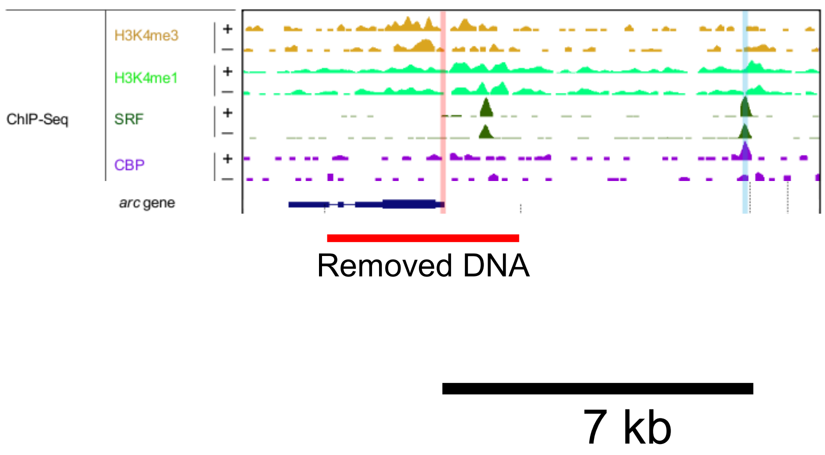


Knock-out experiment confirms that RNAPII recruitment is independent of the promoter

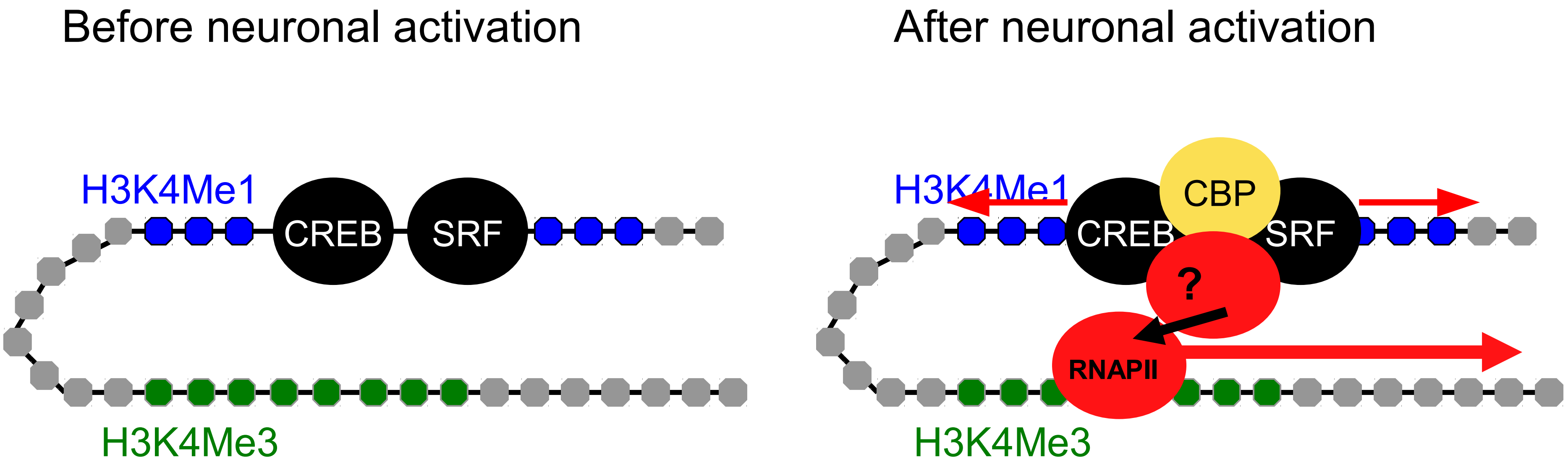


Knock-out experiment confirms that RNAPII recruitment is independent of the promoter



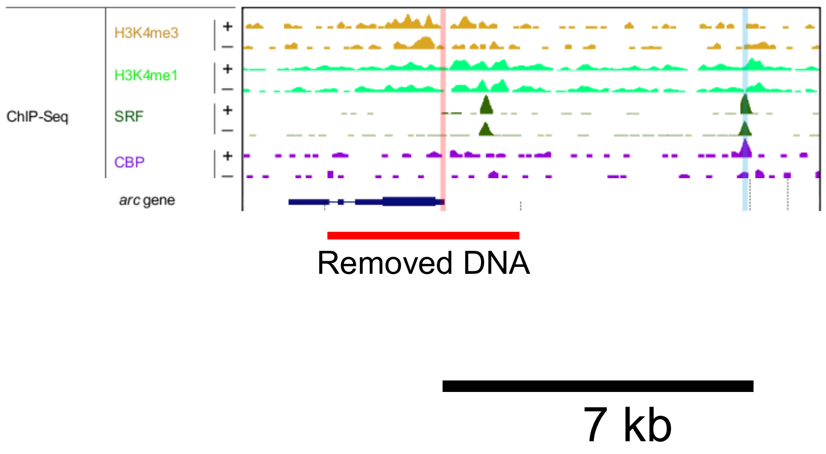


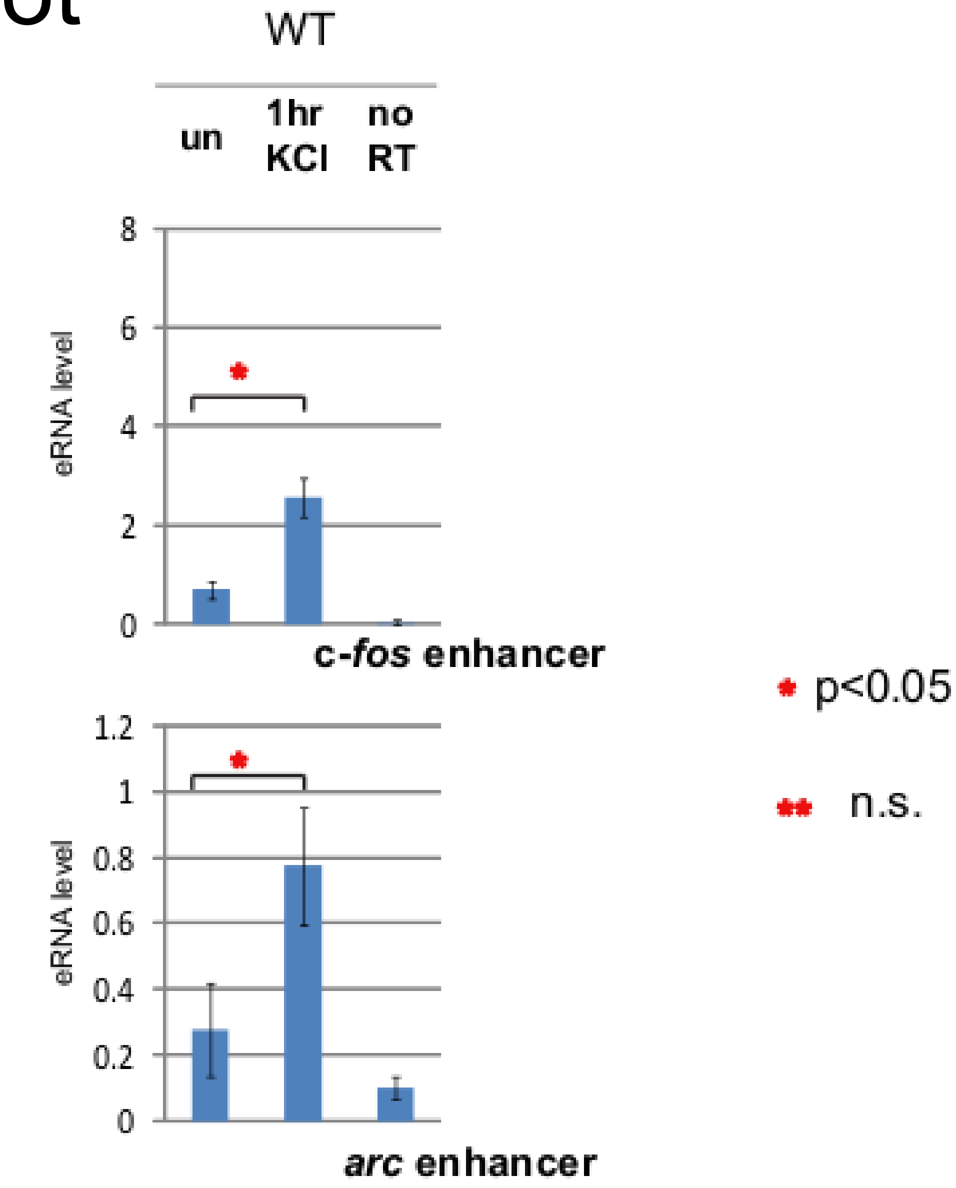
Enhancers bind RNAPII independently, does the transcription depend on the promoter?



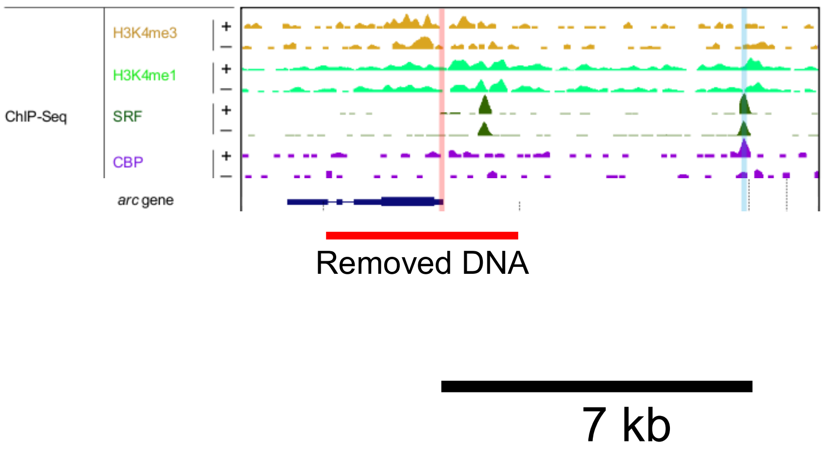
- Does RNAPII bind at enhancers? YES
- Are transcripts produced at enhancers? YES
- Is RNAPII recruitment independent? YES
- Is eRNA production independent?

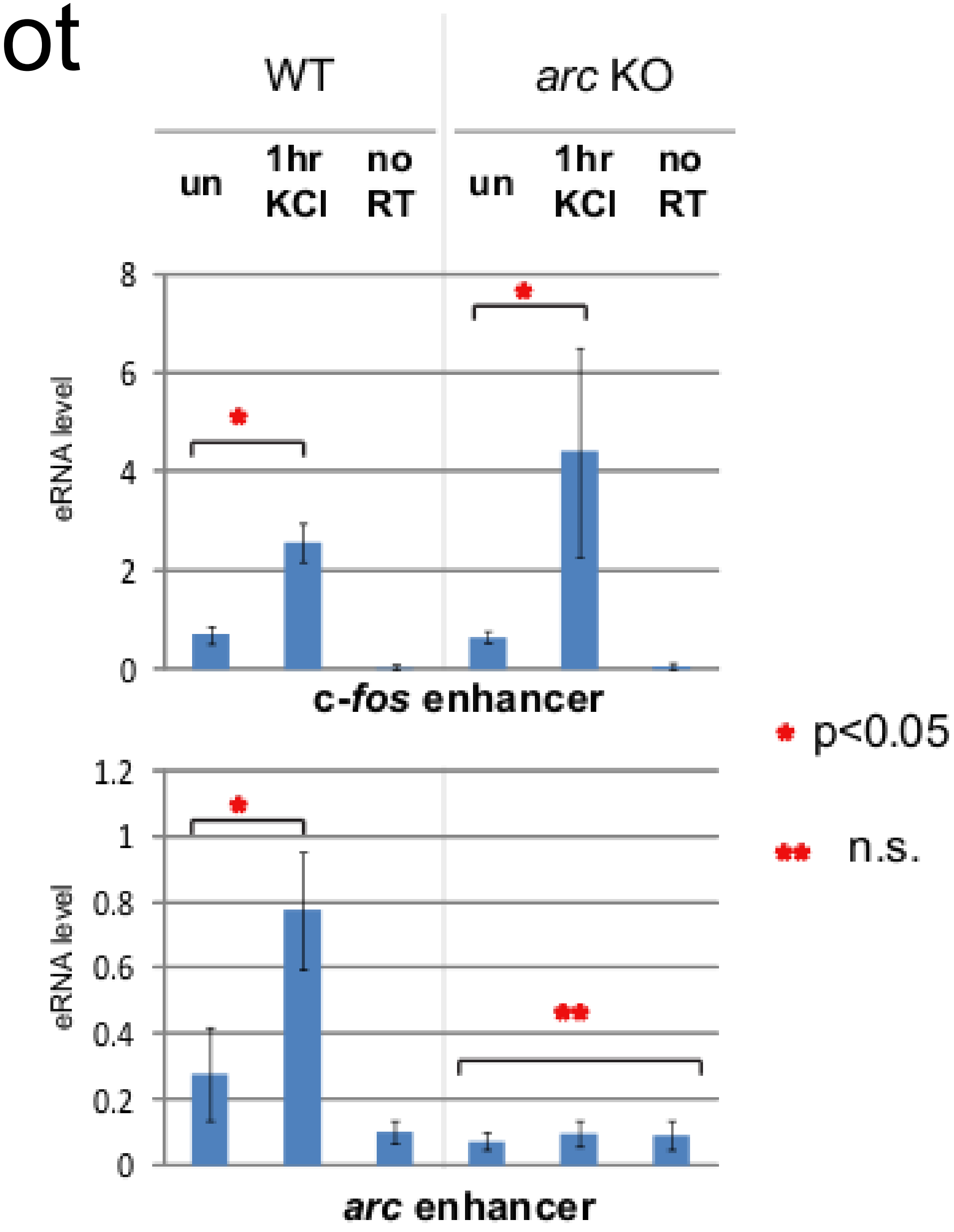
Knock-out experiment confirms that RNAPII recruitment is independent of the promoter but eRNA synthesis is not



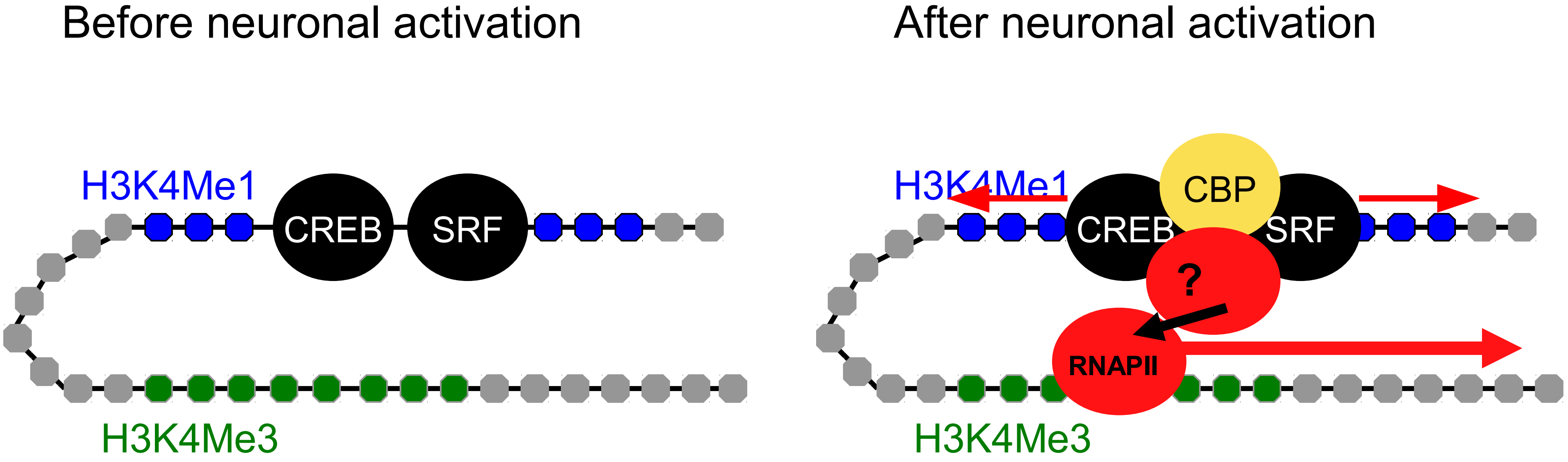


Knock-out experiment confirms that RNAPII recruitment is independent of the promoter but eRNA synthesis is not





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



- Does RNAPII bind at enhancers? YES
- Are transcripts produced at enhancers? YES
- Is RNAPII recruitment independent? YES
- Is eRNA production independent? NO

We have not yet been able to determine 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

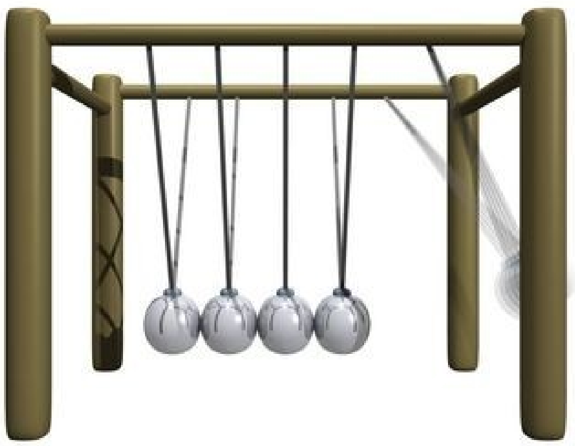
doi:10.1038/nature09692

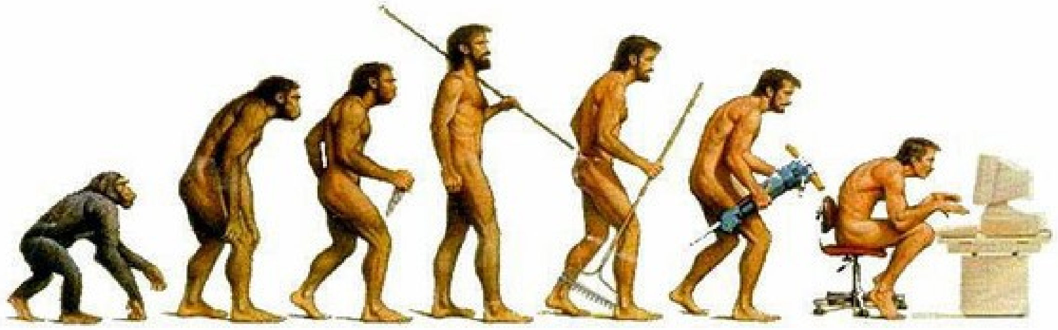
A unique chromatin signature uncovers early developmental enhancers in humans

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

Future Work: Organizing principles of the genome

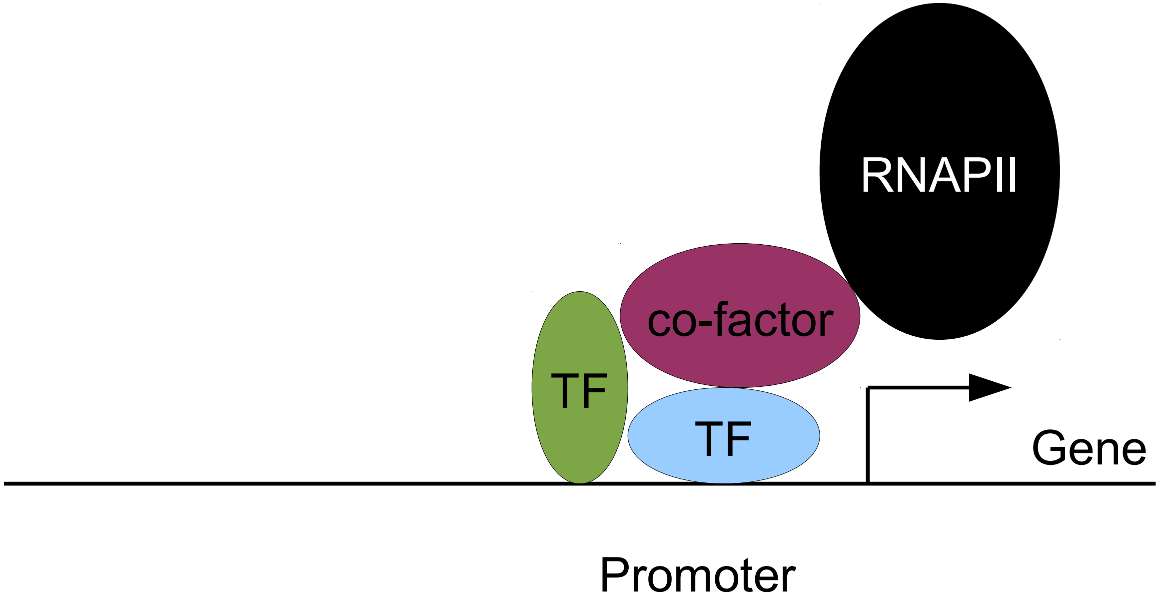
- Use genome-wide data to develop systems biology and biophysical type mathematical models of gene expression





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

- Use ChIP-Seq to test and compare biophysical models of TF binding

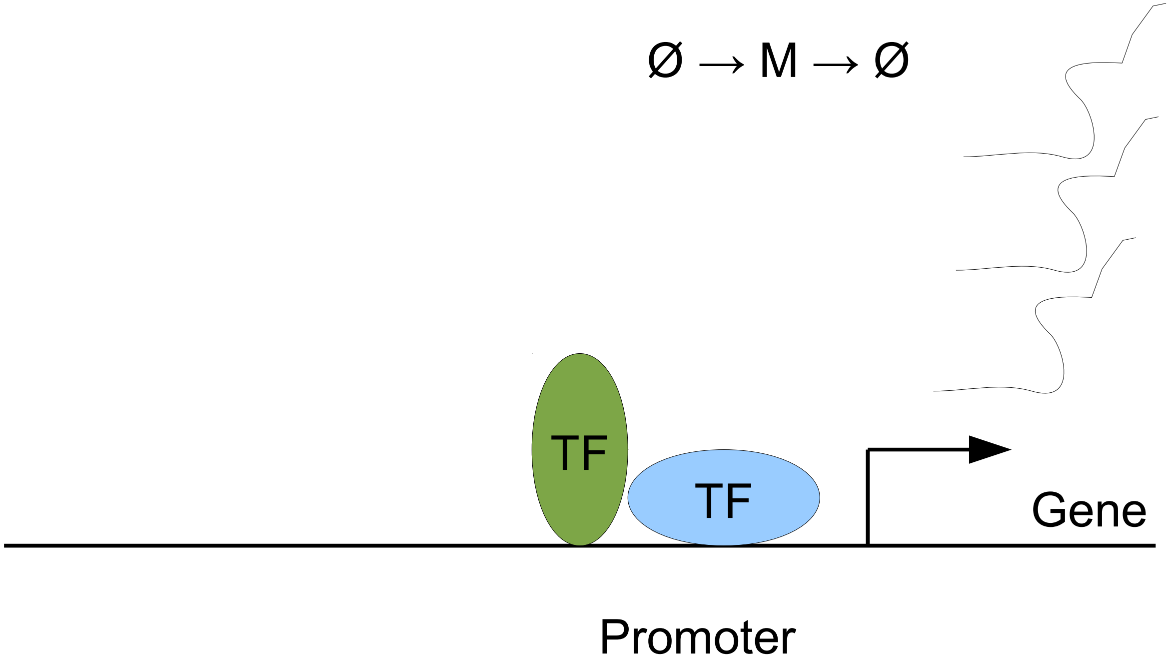


Can we develop a mathematical framework for parametric noise and robustness?

- Molecular noise does not fit observations

$$dP(m)/dt = kP(m-1) + P(m+1) - (m+k)P(m)$$

$$\varnothing \rightarrow M \rightarrow \varnothing$$

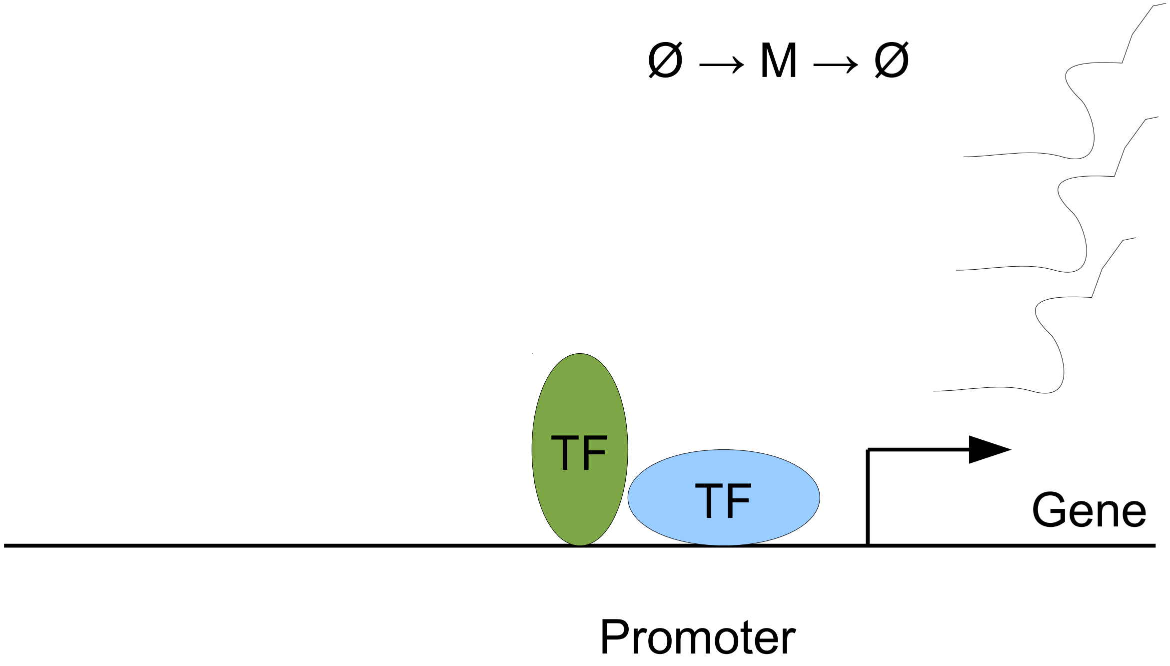


Can we develop a mathematical framework for parametric noise and robustness?

- Molecular noise does not fit observations
- Incorporate parametric noise

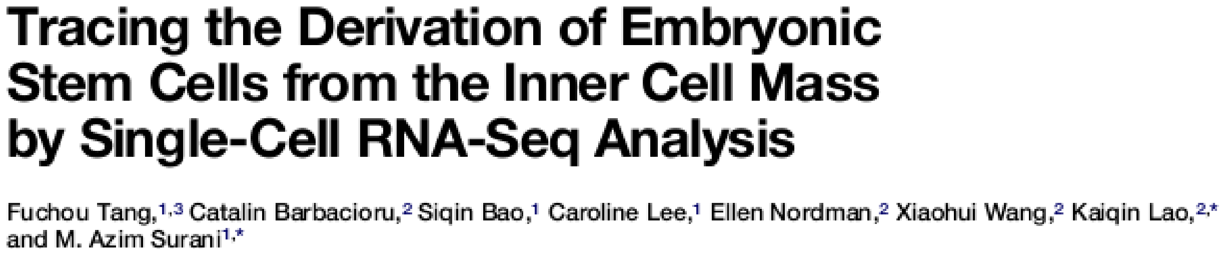
$$dP(m)/dt = kP(m-1) + P(m+1) - (m+k)P(m)$$

$$\varnothing \rightarrow M \rightarrow \varnothing$$



What is the impact on the phenotype from gene expression noise?

- RNA-Seq for single cells
- Global view of noise in gene expression
  - Pathways
  - Proximity
  - Cell-types
  - Propagation



Tracing the Derivation of Embryonic Stem Cells from the Inner Cell Mass by Single-Cell RNA-Seq Analysis

Fuchou Tang,[1,3] Catalin Barbacioru,[2] Siqin Bao,[1] Caroline Lee,[1] Ellen Nordman,[2] Xiaohui Wang,[2] Kaiqin Lao,[2,*] and M. Azim Surani[1,*]

Acknowledgements

Gabriel Kreiman, **Children's Hospital Boston**

Wui Ip

Enrique Tobis

Michael Greenberg, **Harvard Medical School**

Tae-Kyung Kim

Jesse Gray

Allen Costa

Daniel Bear

David Harmin

Mike Laptewicz

Eirene Markenscoff-Papadimitriou

Molecular Genetics Core
Children's Hospital Boston
Kellie Haley
Josh Davis
Hal Schneider

Life Technologies
Rob David
Jingwei Ni
Scott Kuersten
Gina Costa
Kevin McKernan

Harvard Medical School
Biopolymer facility
Kristin Waraska
Robert Steen

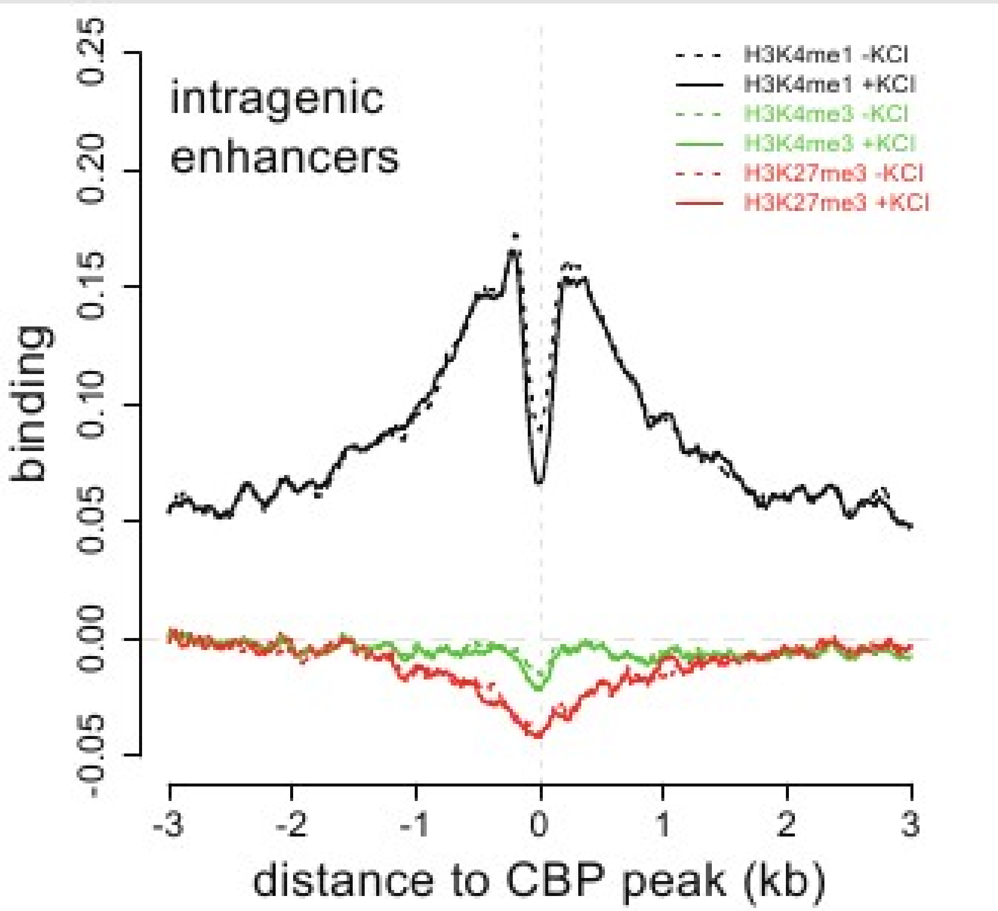
Johns Hopkins
Jing Wu, Paul Worley Lab

Thank You

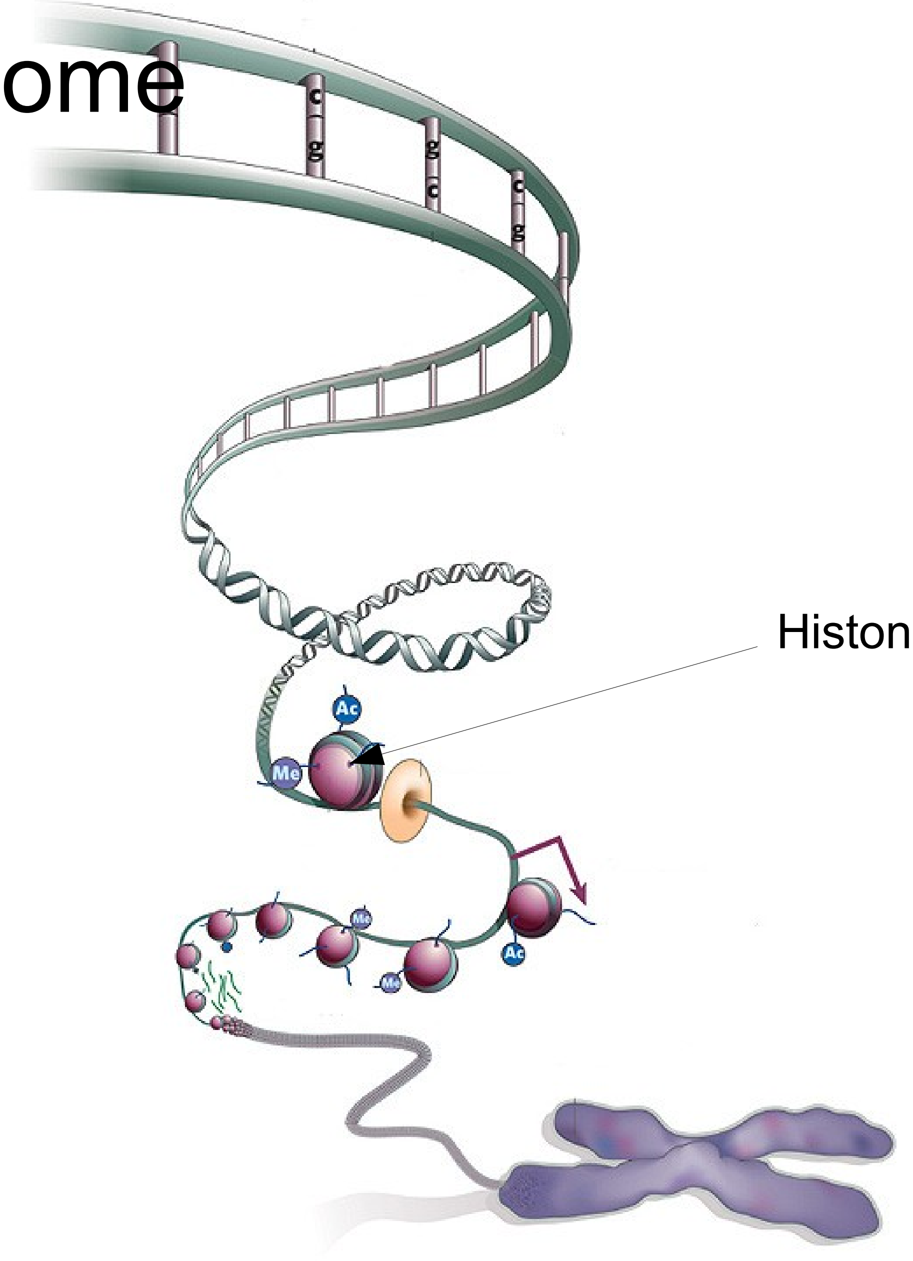
?

Intragenic enhancers

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



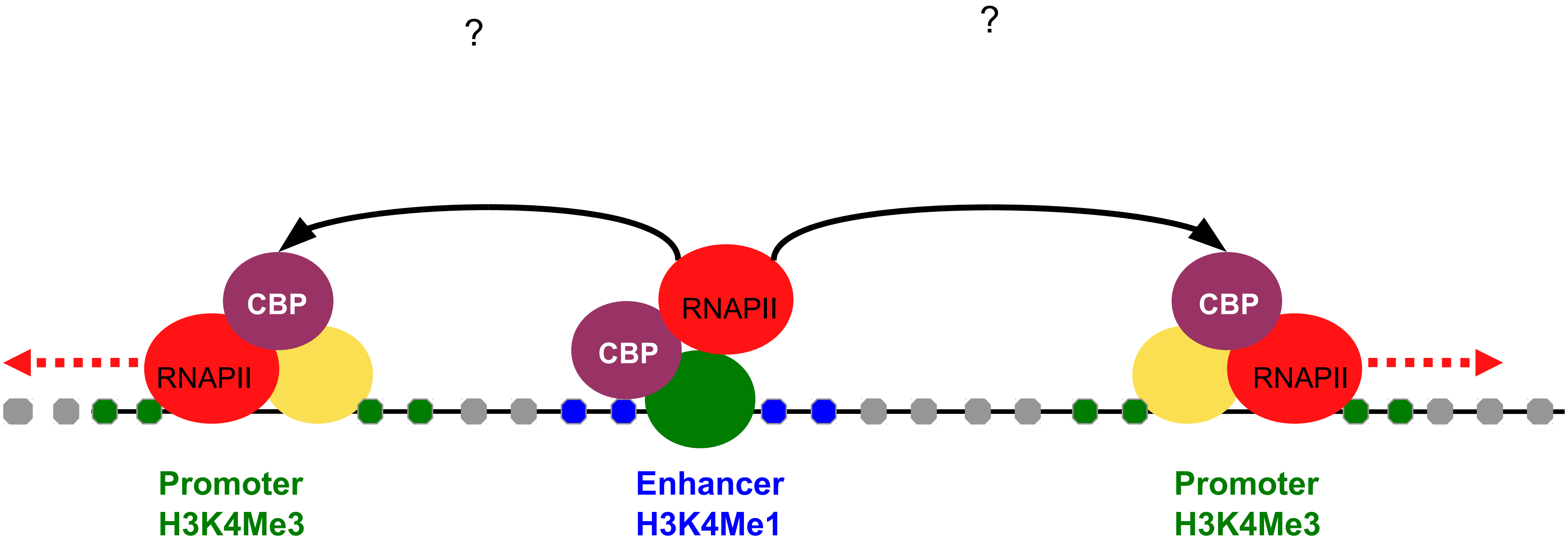
Additional layers of modifications determine the function of the genome



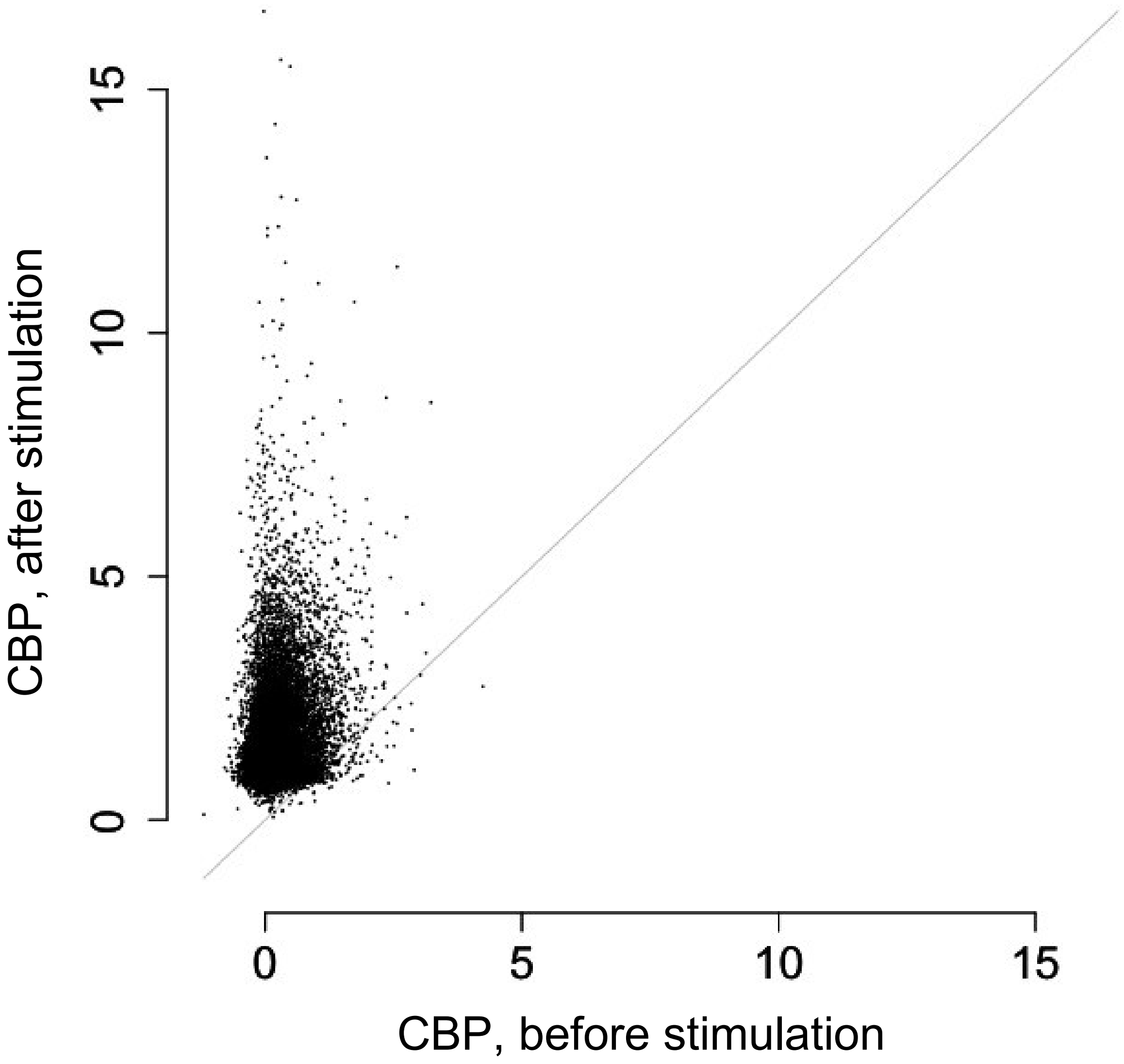
Histone

(ENCODE, 2007)

How do eRNA levels relate to mRNA levels?

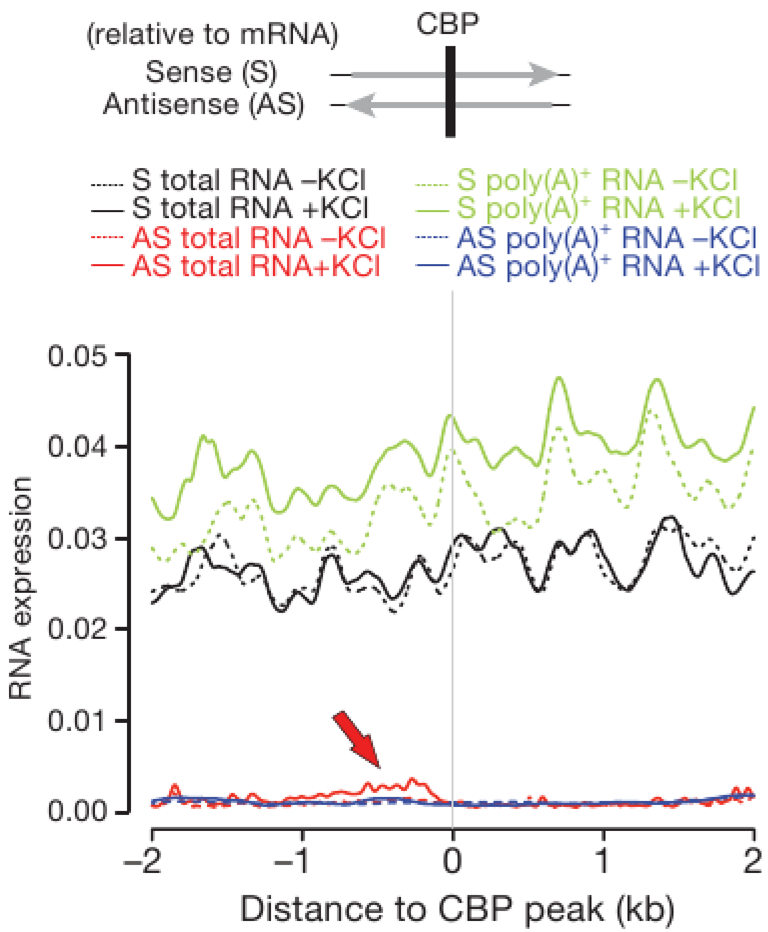


CBP binds in an activity regulated manner to ~28,000 sites throughout the genome

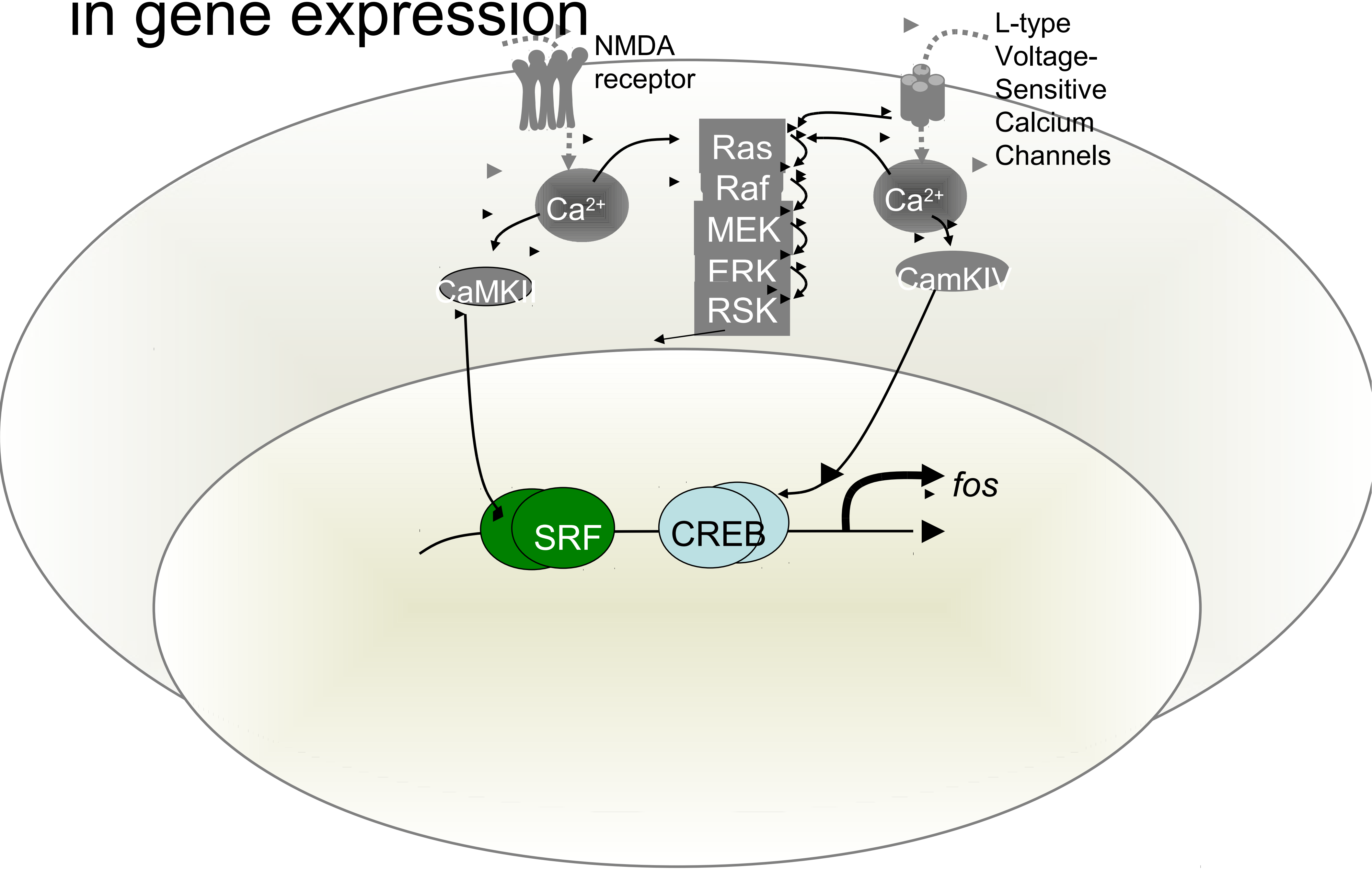


Intragenic enhancers are also transcribed

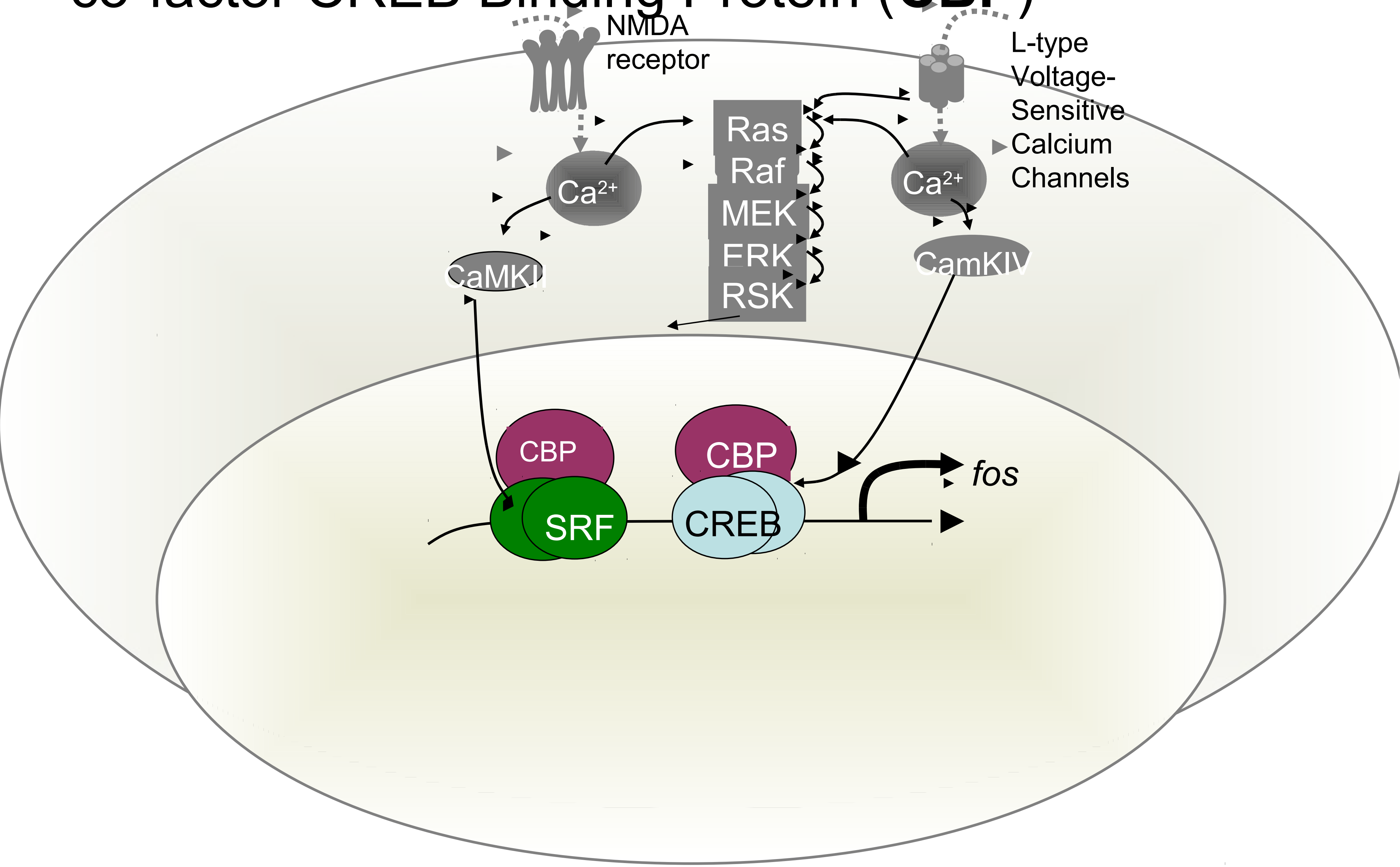
- ~7,000 enhancers overlapping introns
  - No signal detectable on sense strand
  - Significant anti-sense transcription



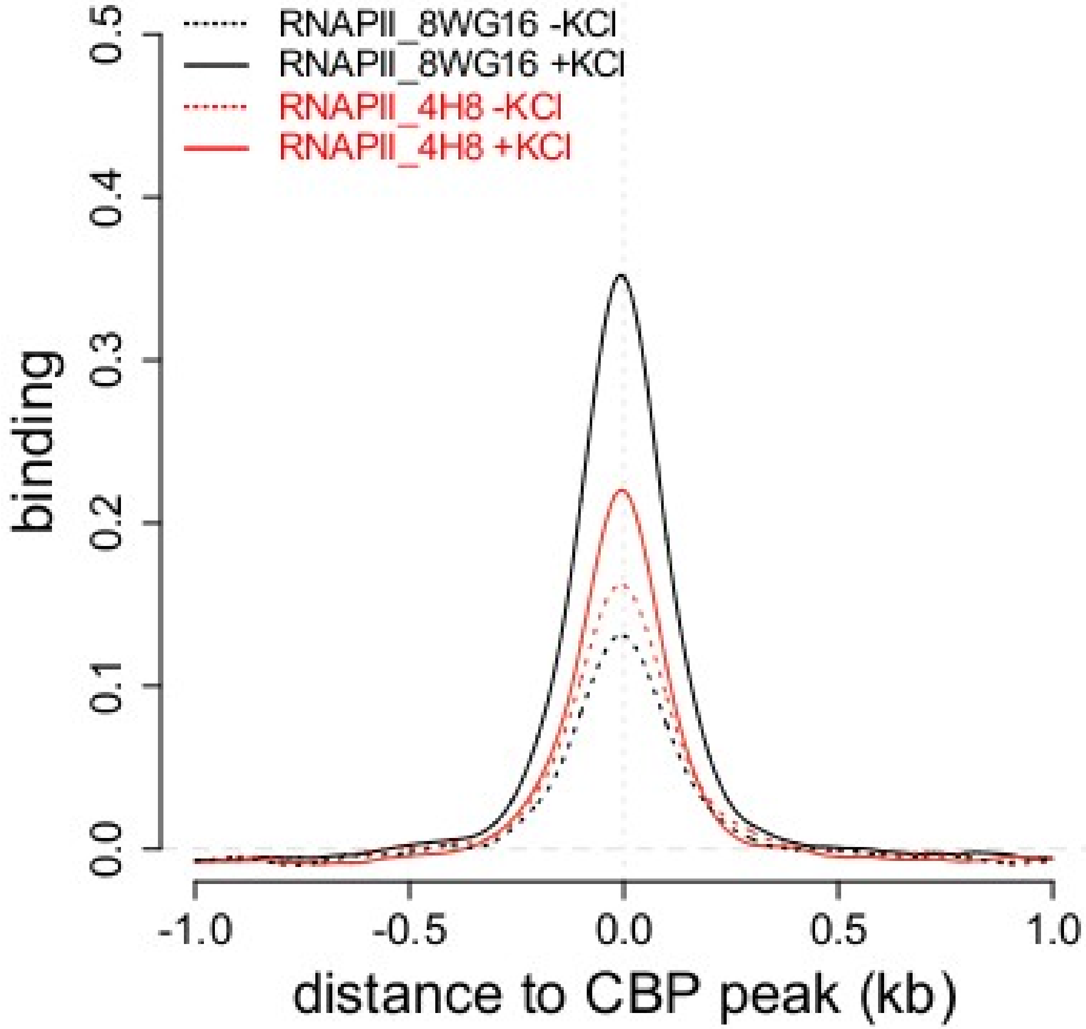
Changes in synapses are driven by changes in gene expression

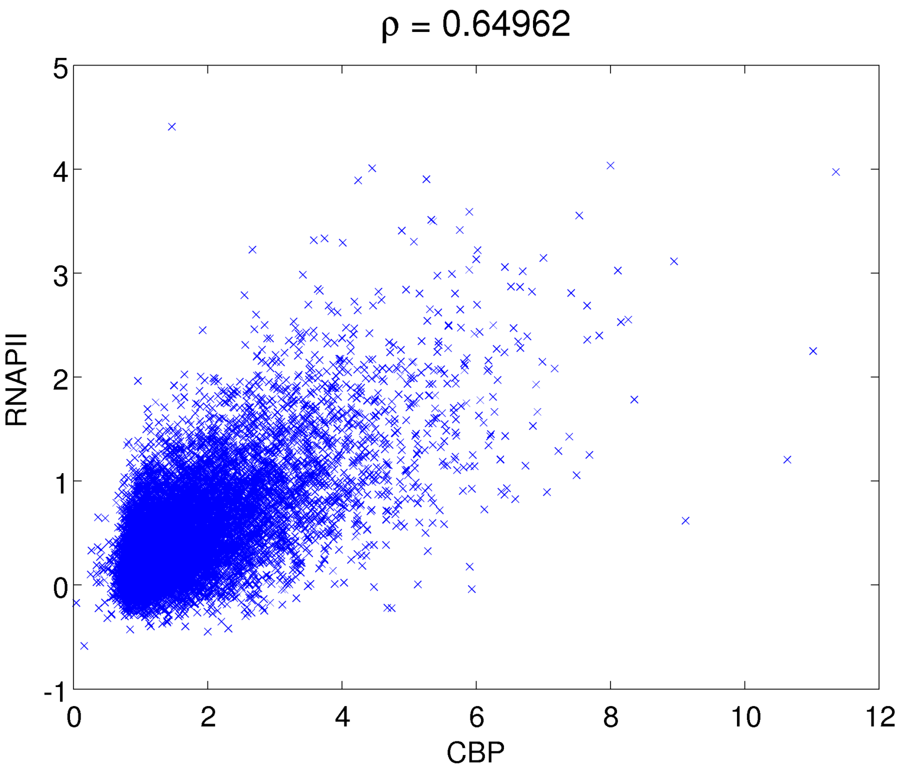


fos is activated by phosphorylation and binding of co-factor CREB Binding Protein (**CBP**)



RNAPII is recruited at all enhancers





Conjectured order of events for eRNA

Open chromatin + Histone modifications

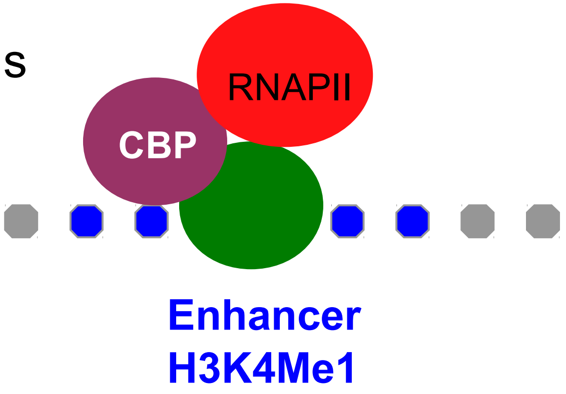
Stimulus → TF binding

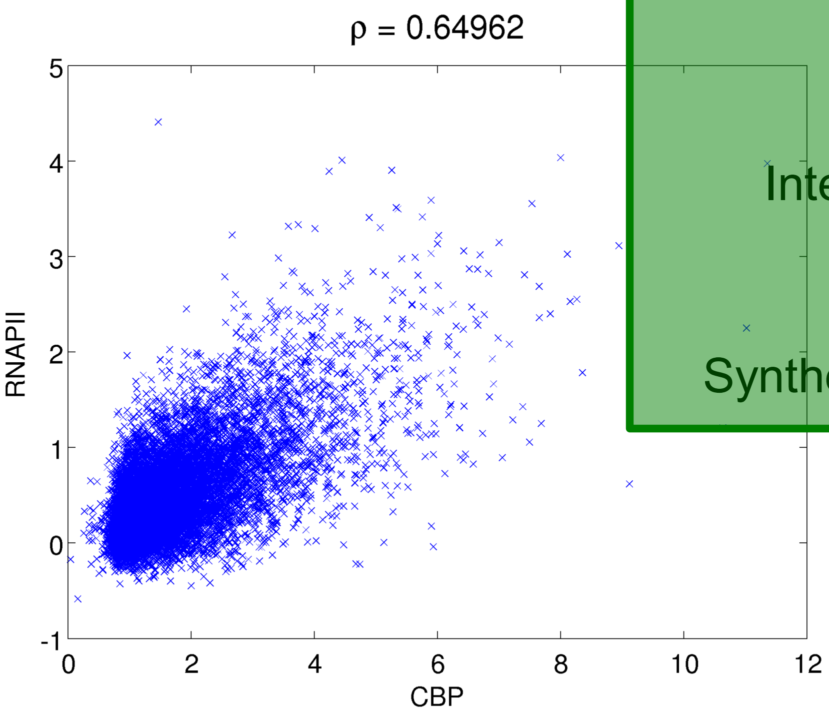
CBP binding

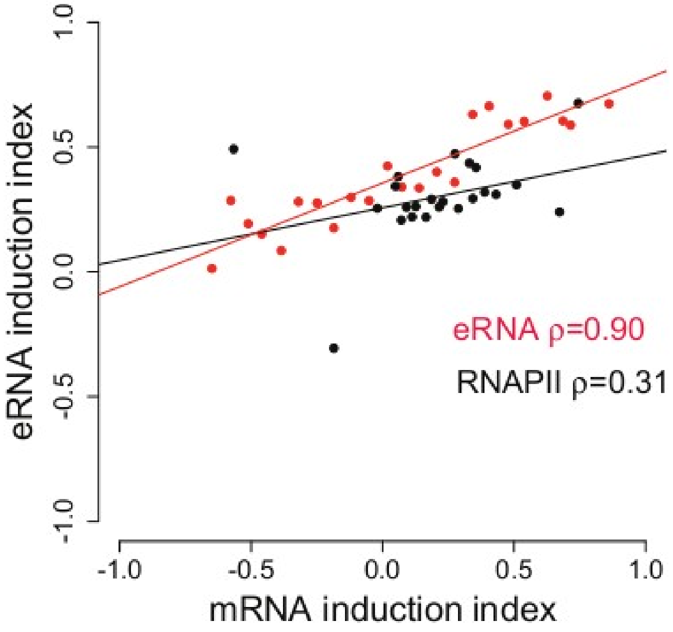
RNAPII binding

Interaction with promoter

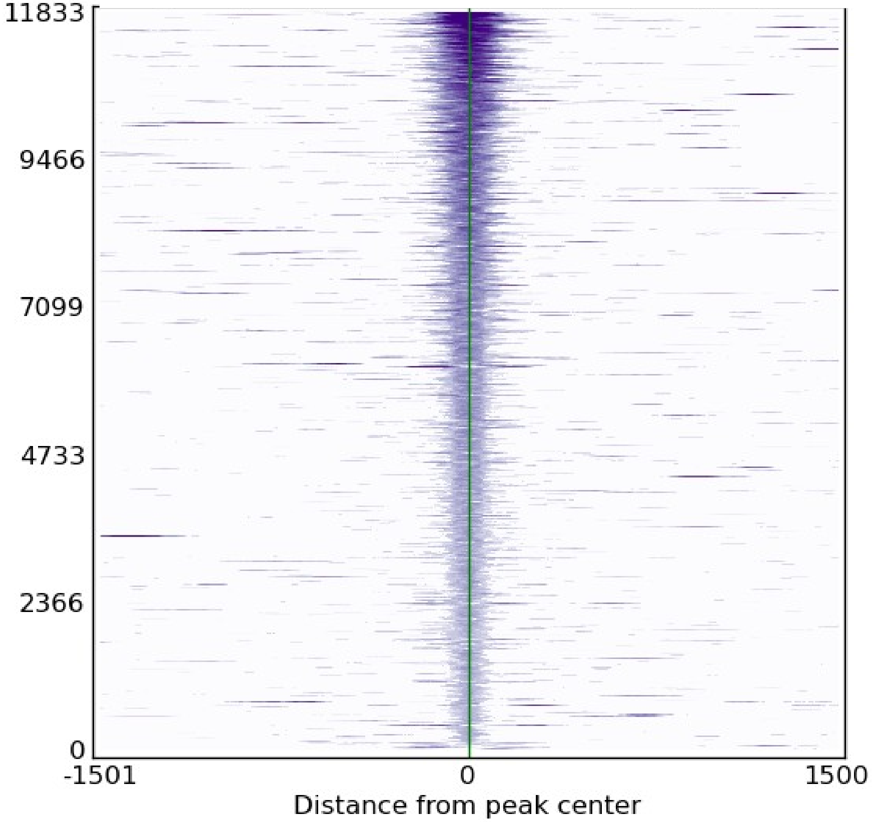
Synthesis of eRNA and mRNA

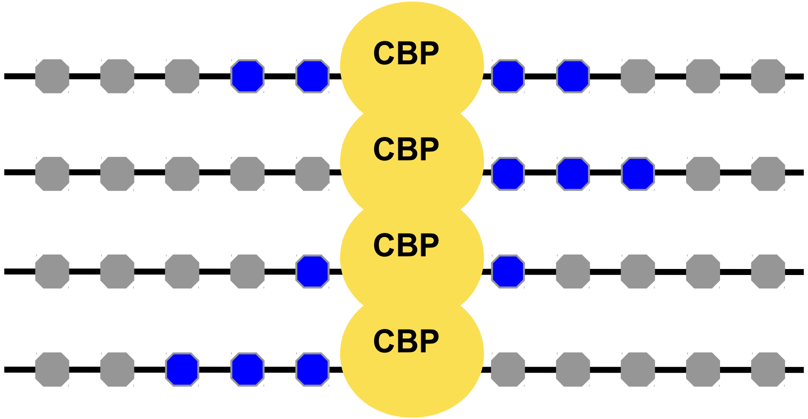




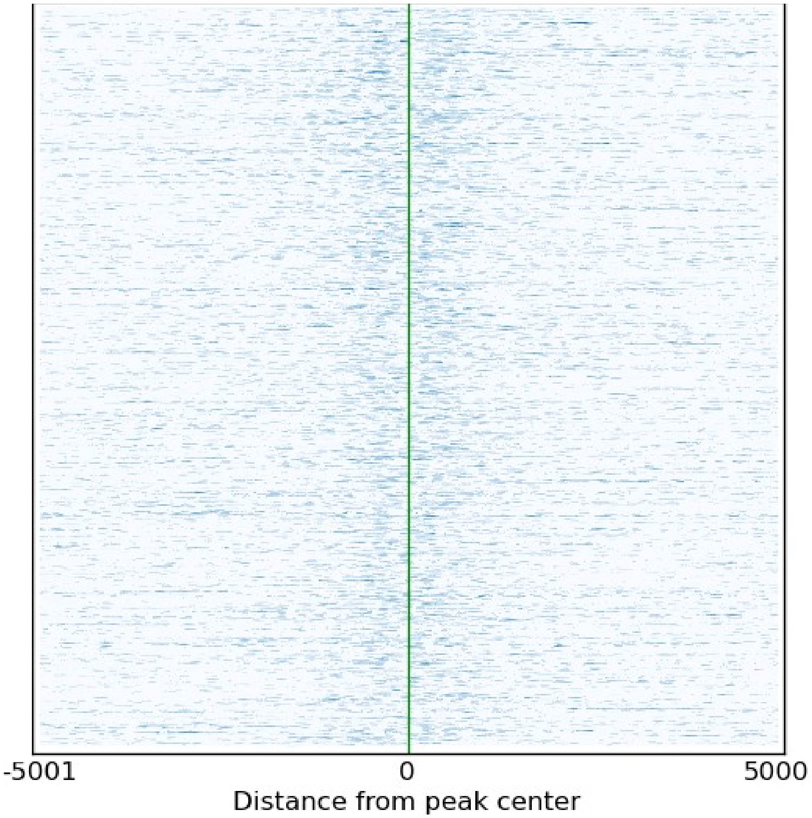


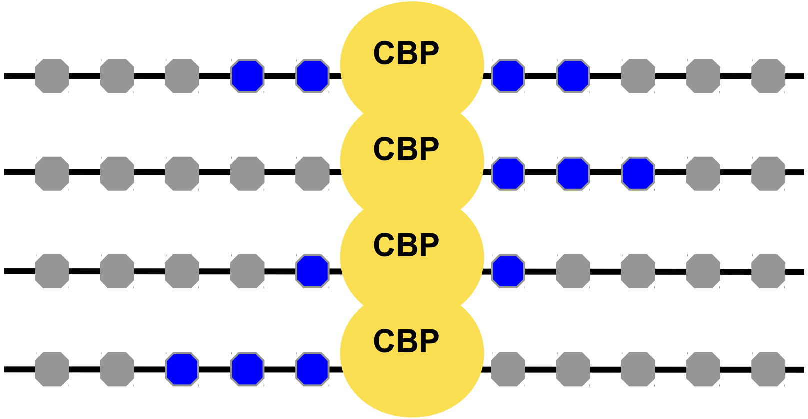
Aligning CBP peaks to calculate binding profiles



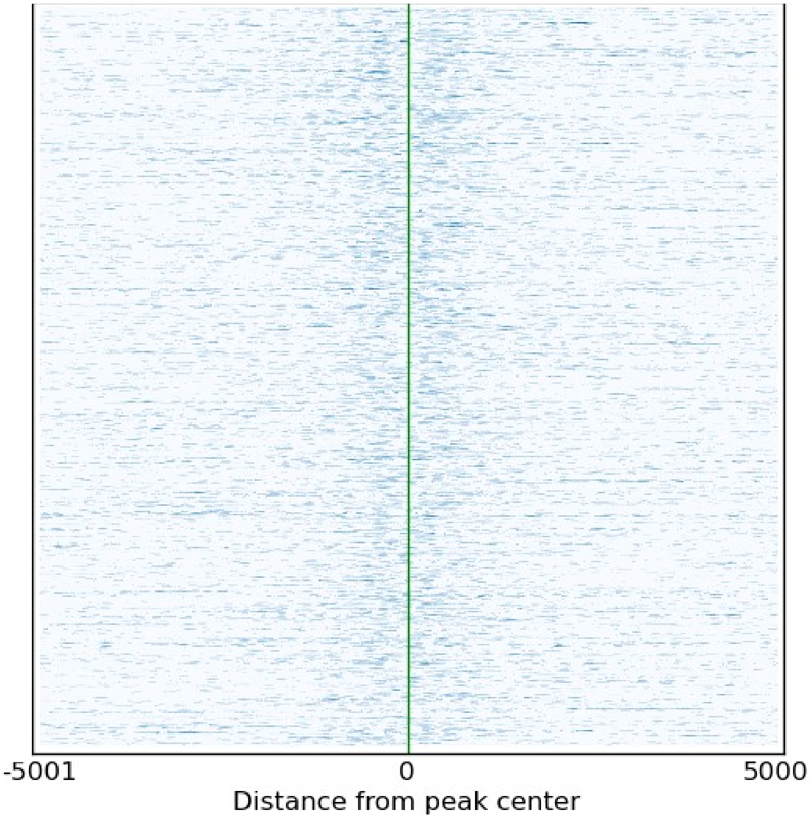


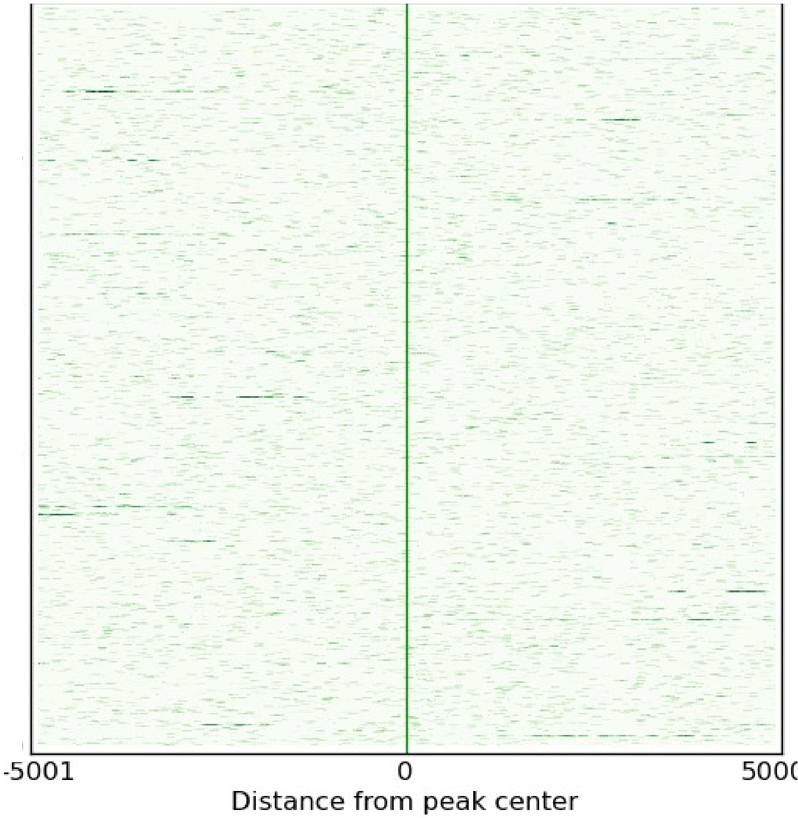
Aligning CBP peaks to calculate H3K4me1 binding profiles

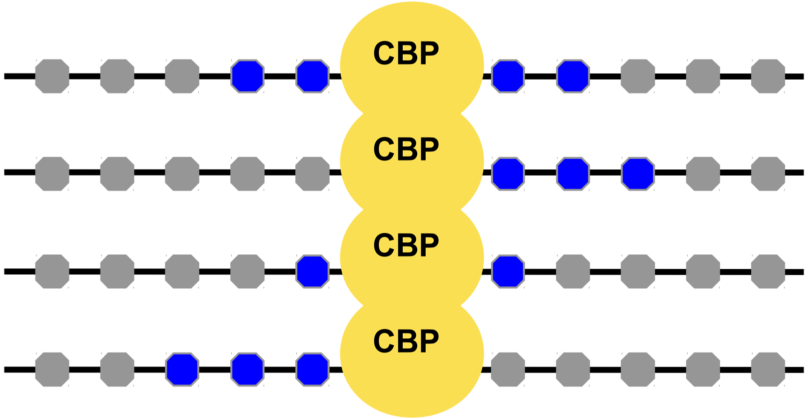




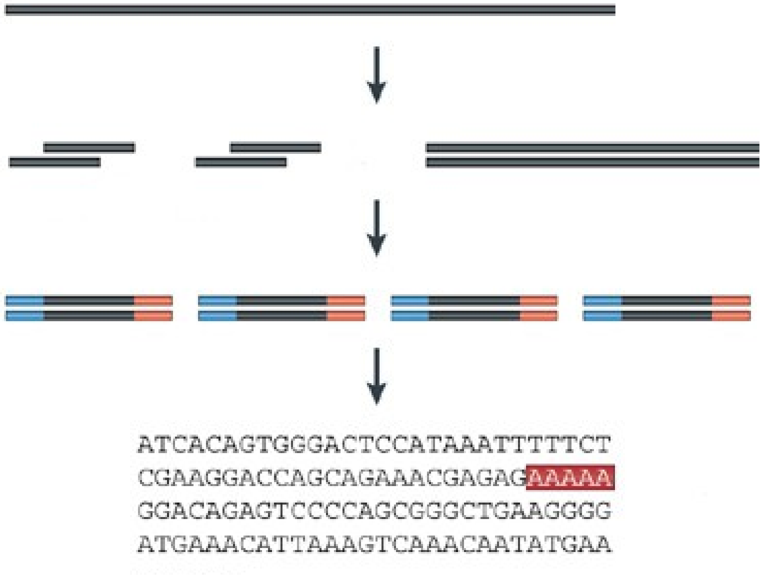
Aligning CBP peaks to calculate H3K4me1 and H3K4me3 binding profiles







RNA-Seq reveals which parts of the genome are transcribed

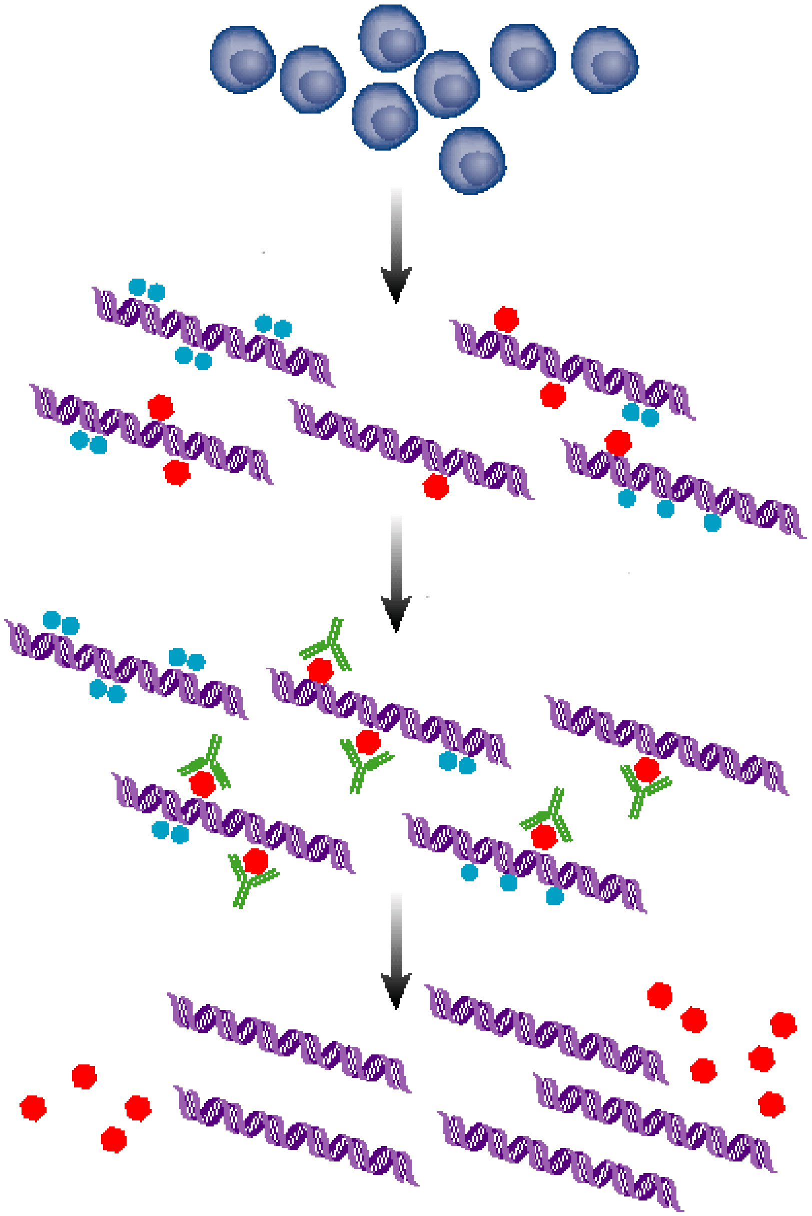


- Fragment
- RNA → cDNA
- 35 bp reads mapped to genome
  - Before and after KCl
  - Total RNA and polyA+

(Wang et al, 2009)

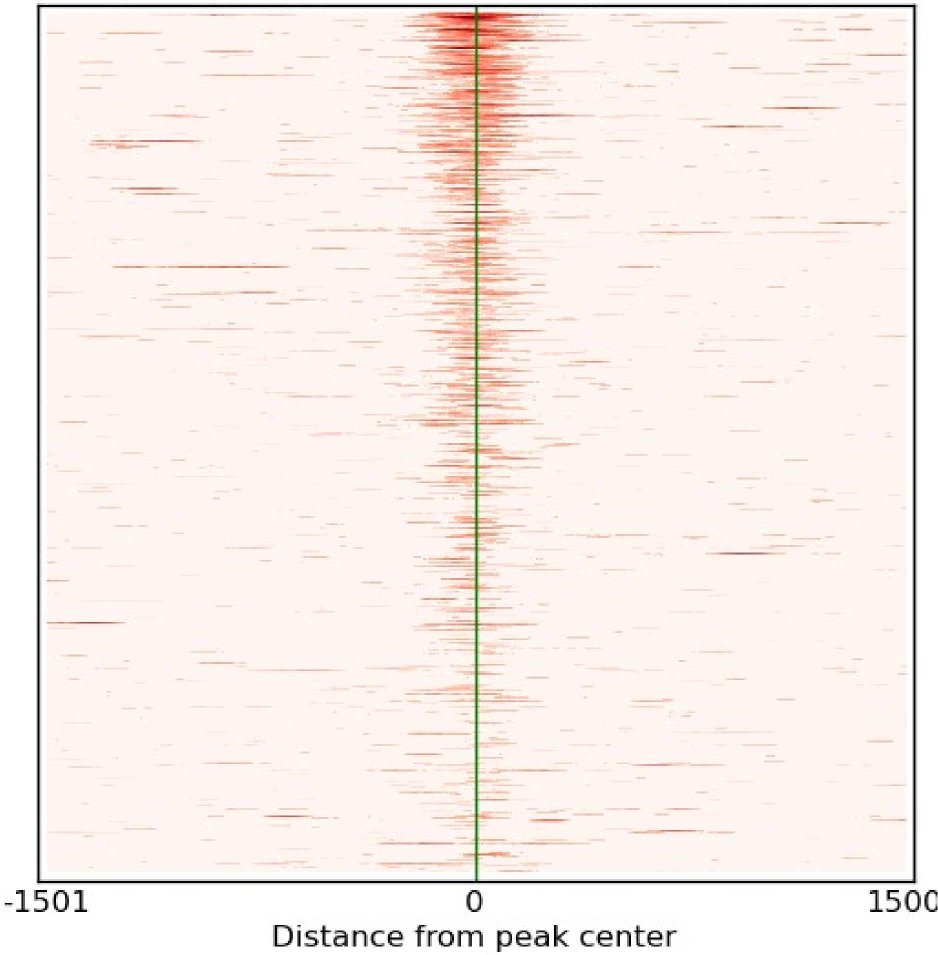
Chromatin immunoprecipitation and sequencing (**ChIP-Seq**) finds protein binding sites *in vivo*

- Fragment DNA
- Extract with antibody
- Sequence fragments
- Map to reference genome

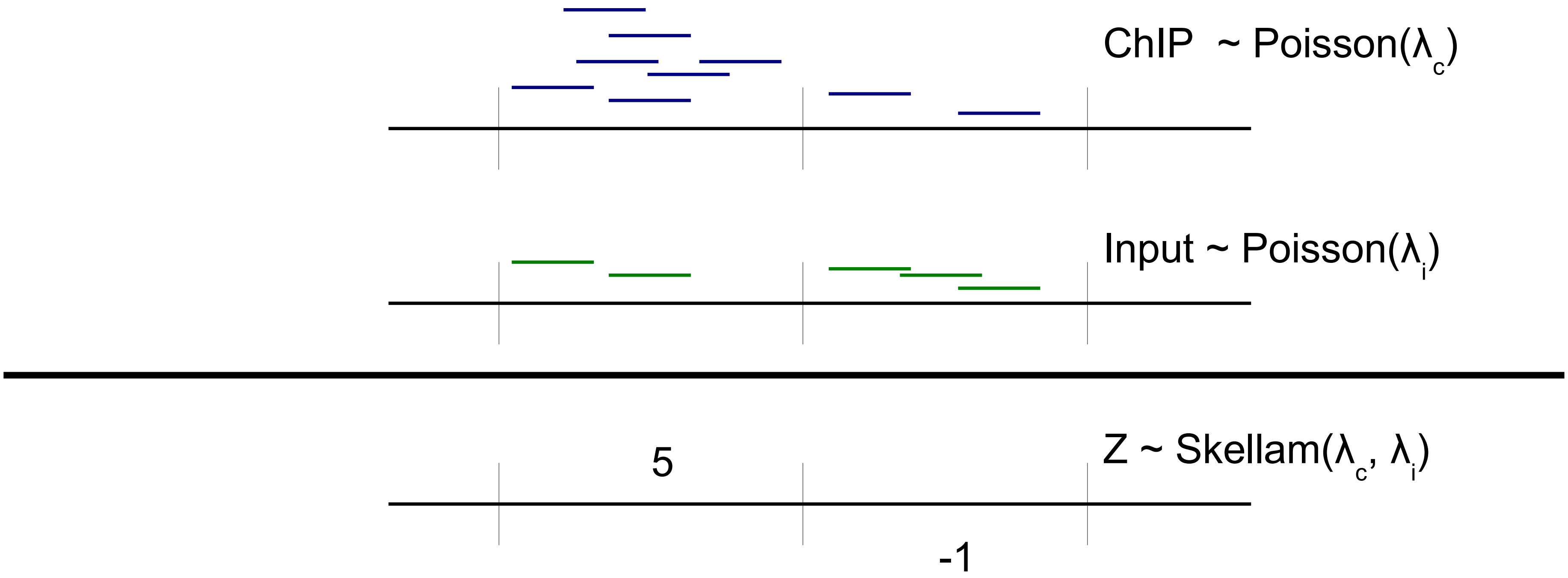


(Mardis, 2007)

RNAPII binds at activity-dependent enhancers



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

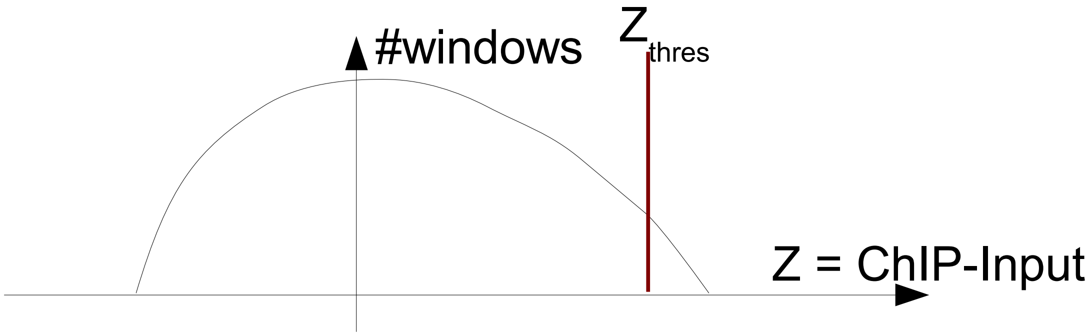


ChIP ~ Poisson(λ_c)

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

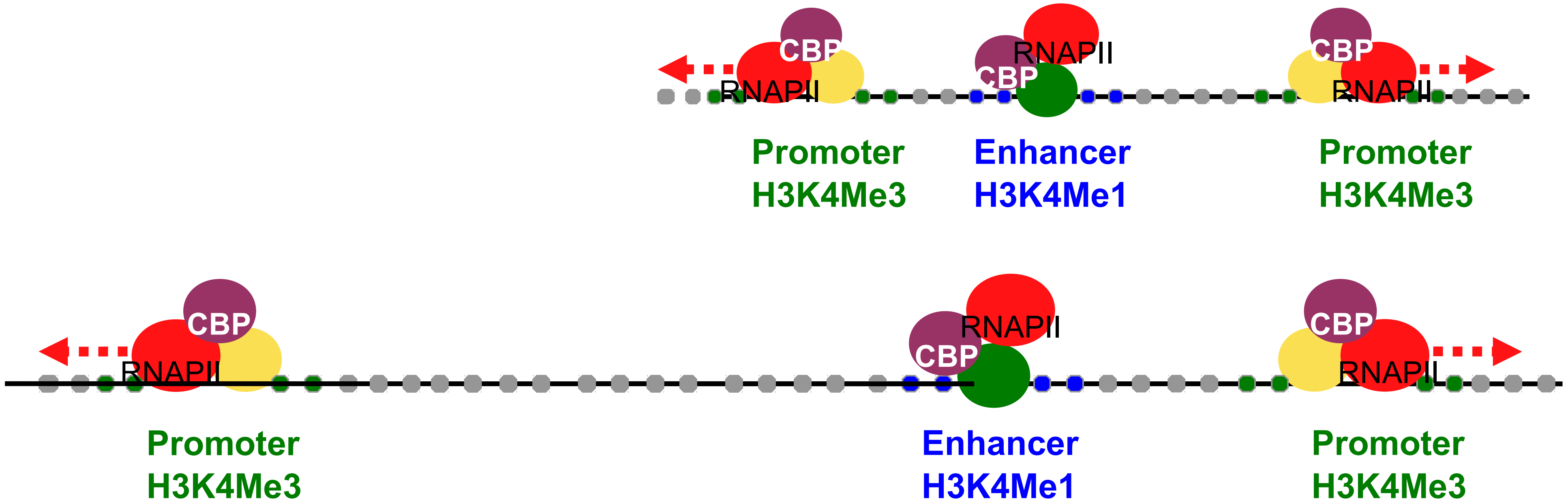
- Z_i = #ChIP reads - #input reads in window *i*
- ~1 read/100 bp
  - Assume #reads in window $P(k) = \lambda^k \exp(-\lambda)/k!$
    - Difference between two Poisson random variables
    - $Z_i \sim \text{Skellam}(z, \lambda_1, \lambda_2)$

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

- 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..... →