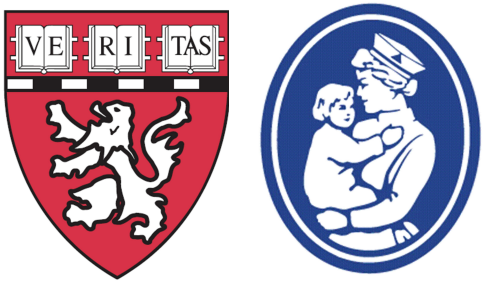
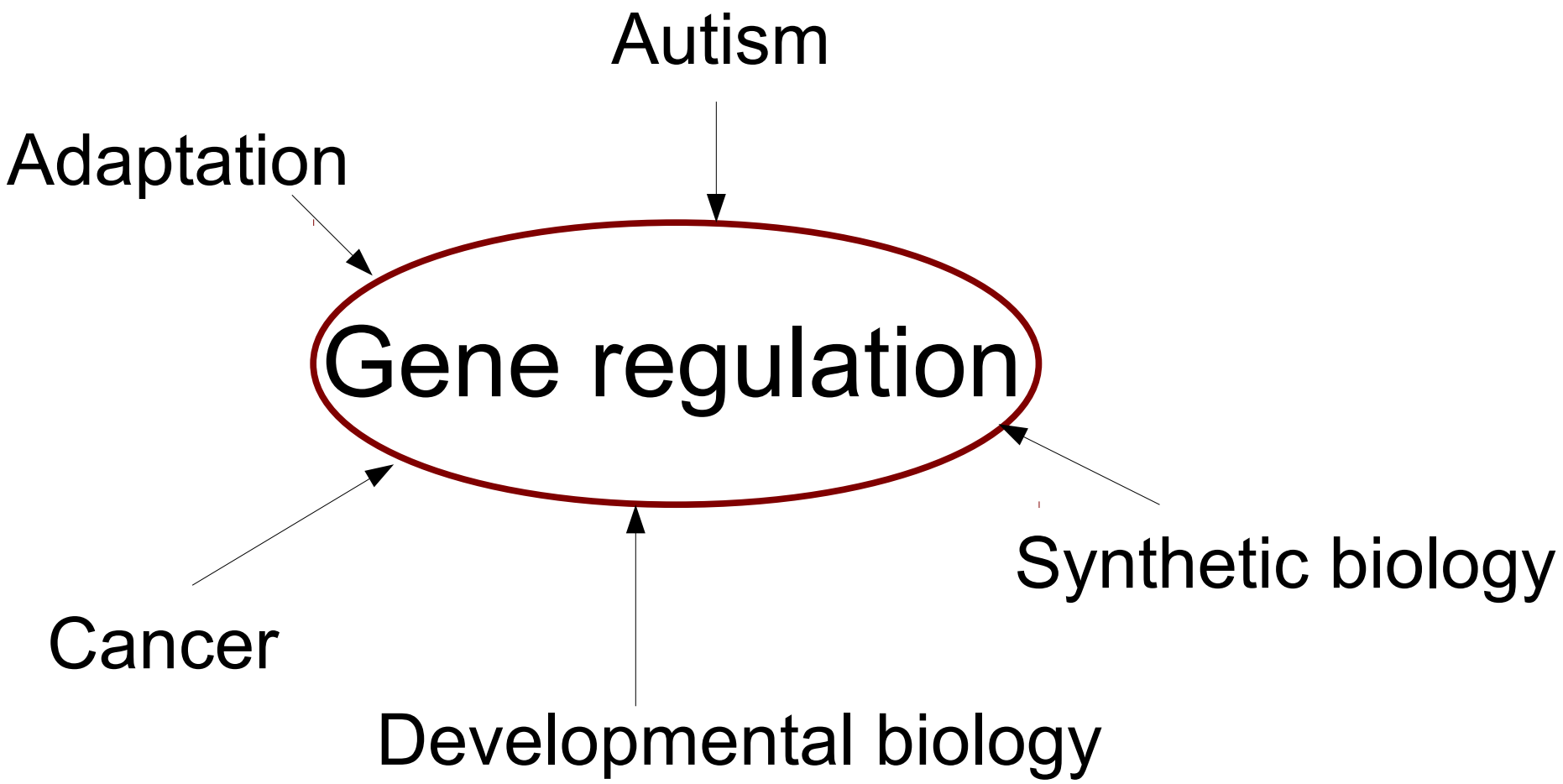
Mechanisms and models of distal enhancers of inducible gene expression

Martin Hemberg

UC Berkeley
February 28, 2012

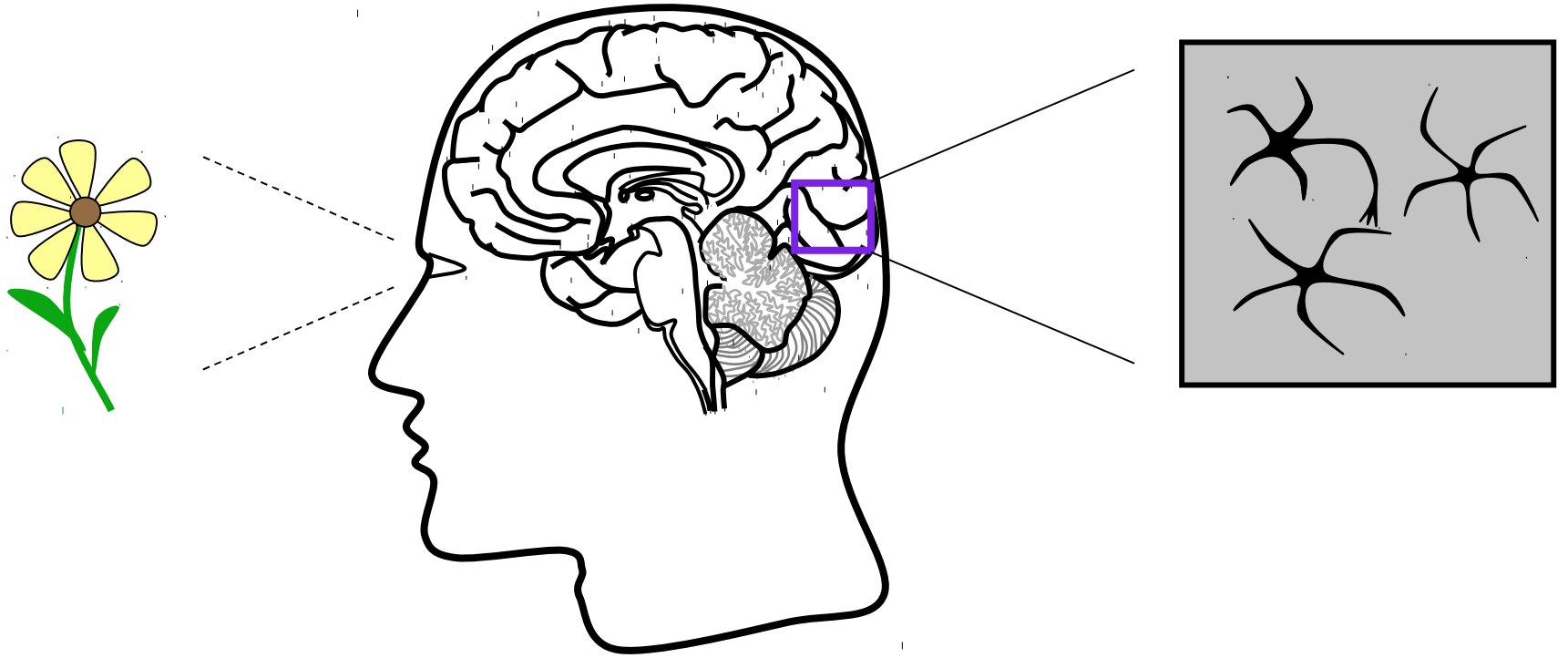


Why is gene regulation important?



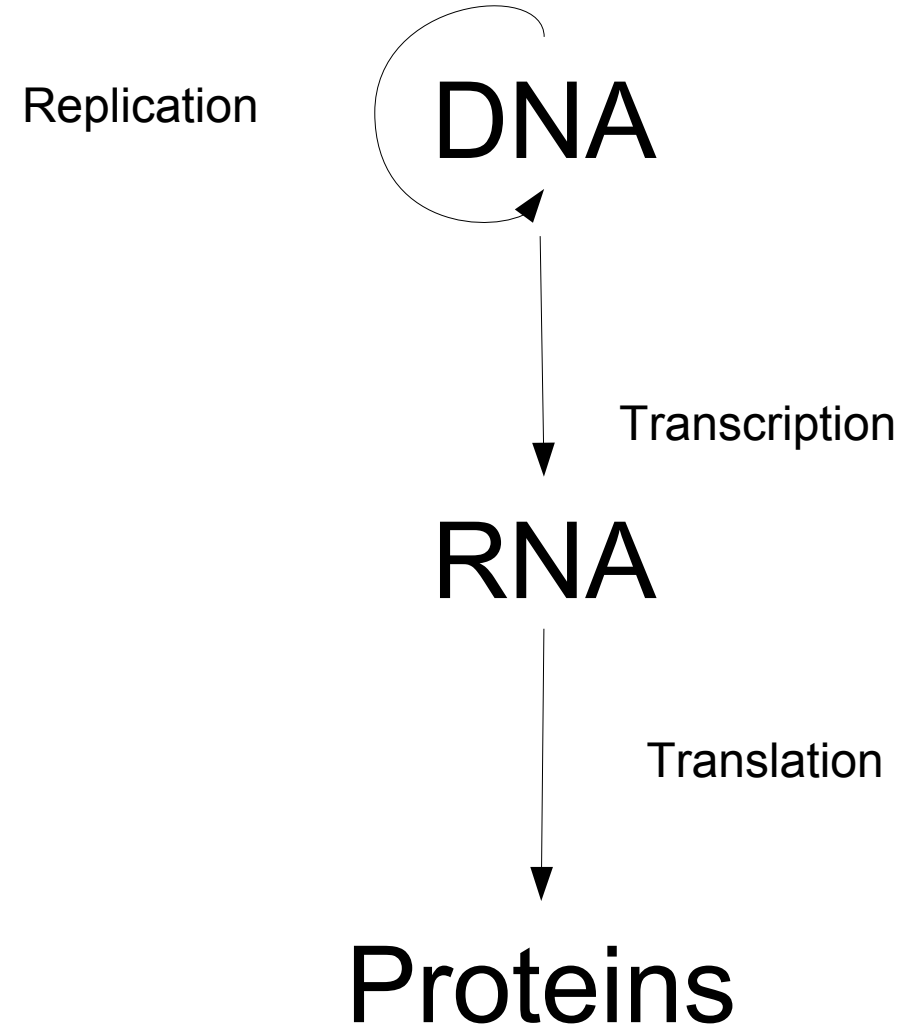
Synapses change in response to external environmental stimuli

- Caused by turning ~1000 genes on or off

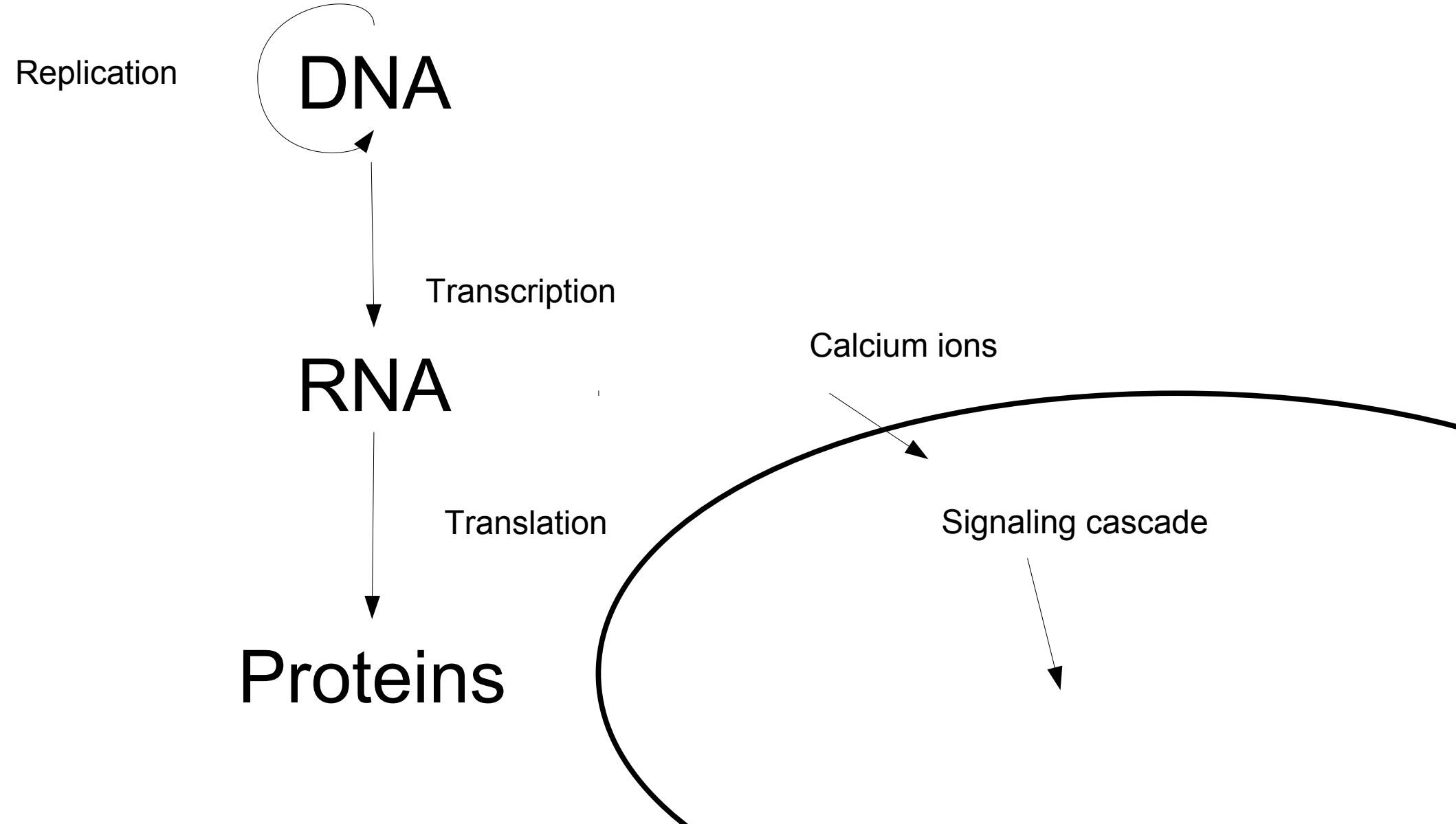


Hubel & Wiesel, 1970's

What is gene expression and gene regulation?

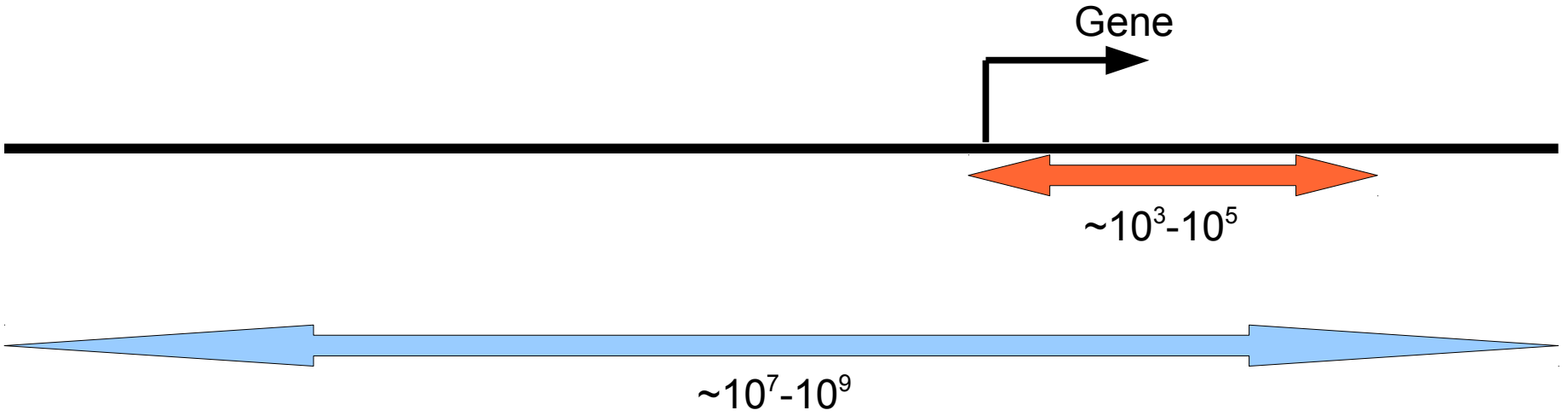


Activity dependent gene expression triggered by influx of Ca



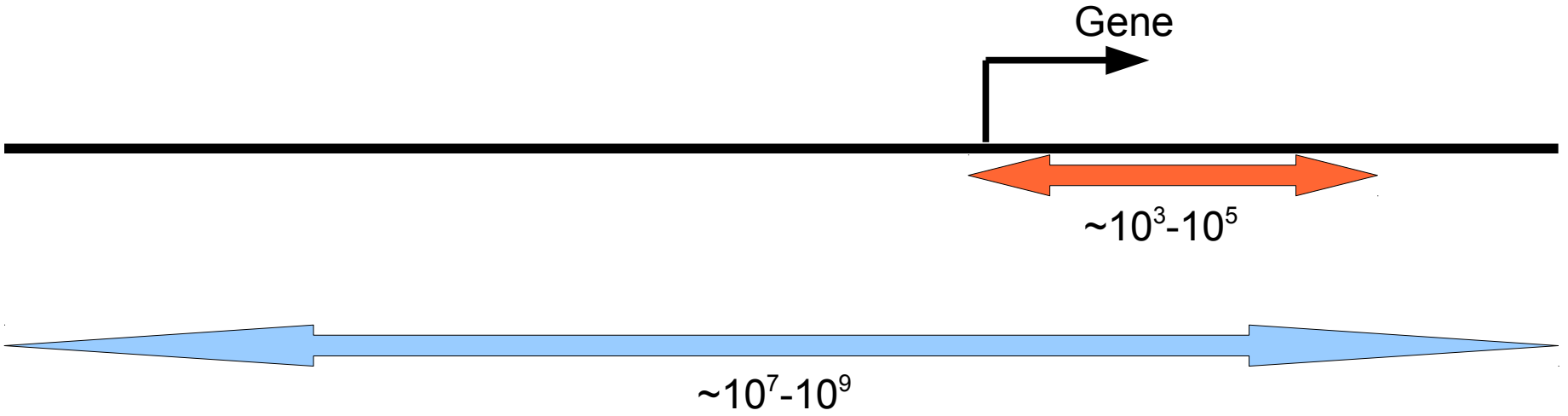
Mouse genome is large and has few genes

- ~25,000 genes
  - ~2% of DNA



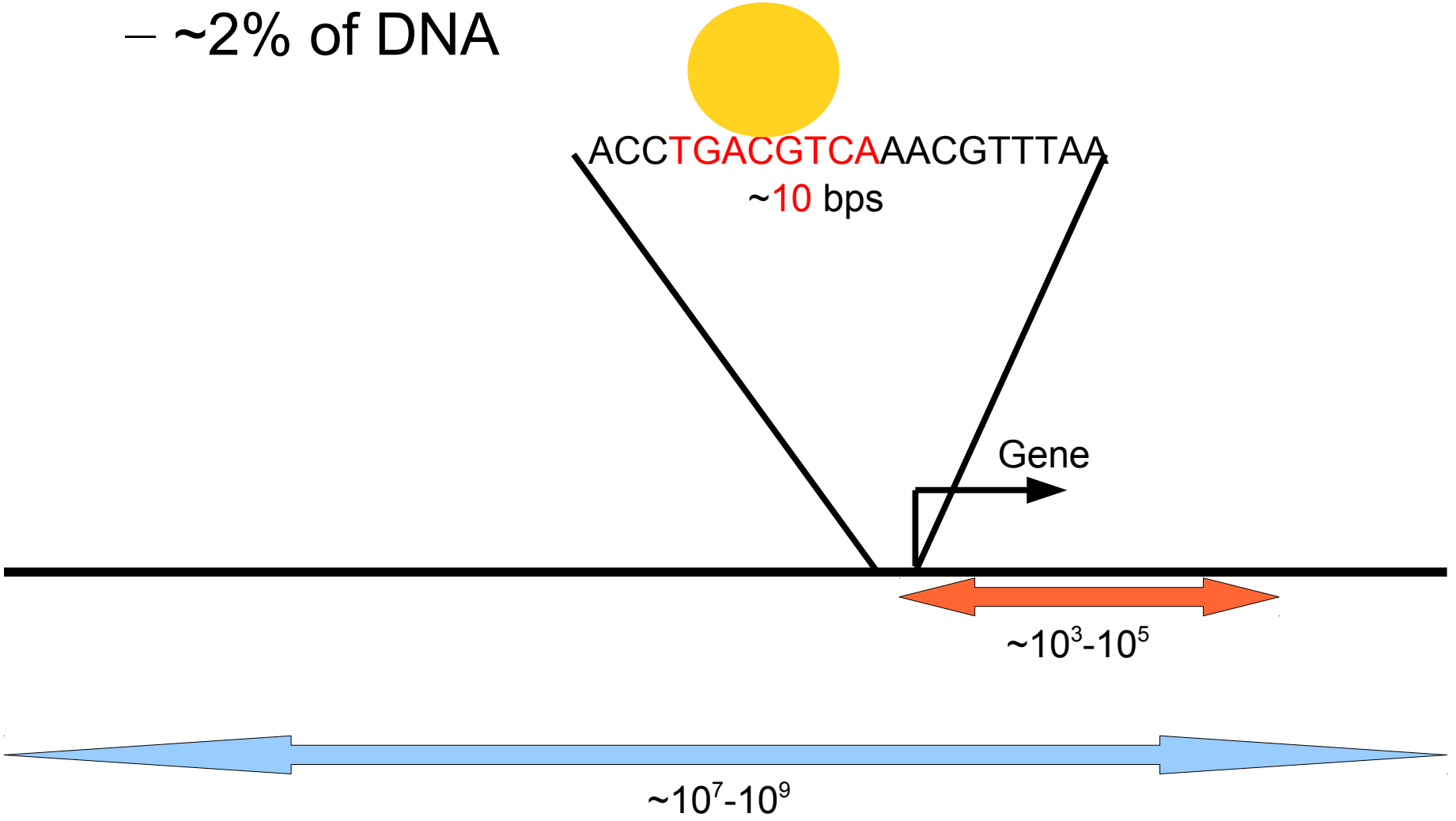
Bacterial genomes are compact

- ~25,000 genes
  - ~2% of DNA
  - Bacteria ~10^6 base pairs (**bps**)
    - 10^3-10^4 genes



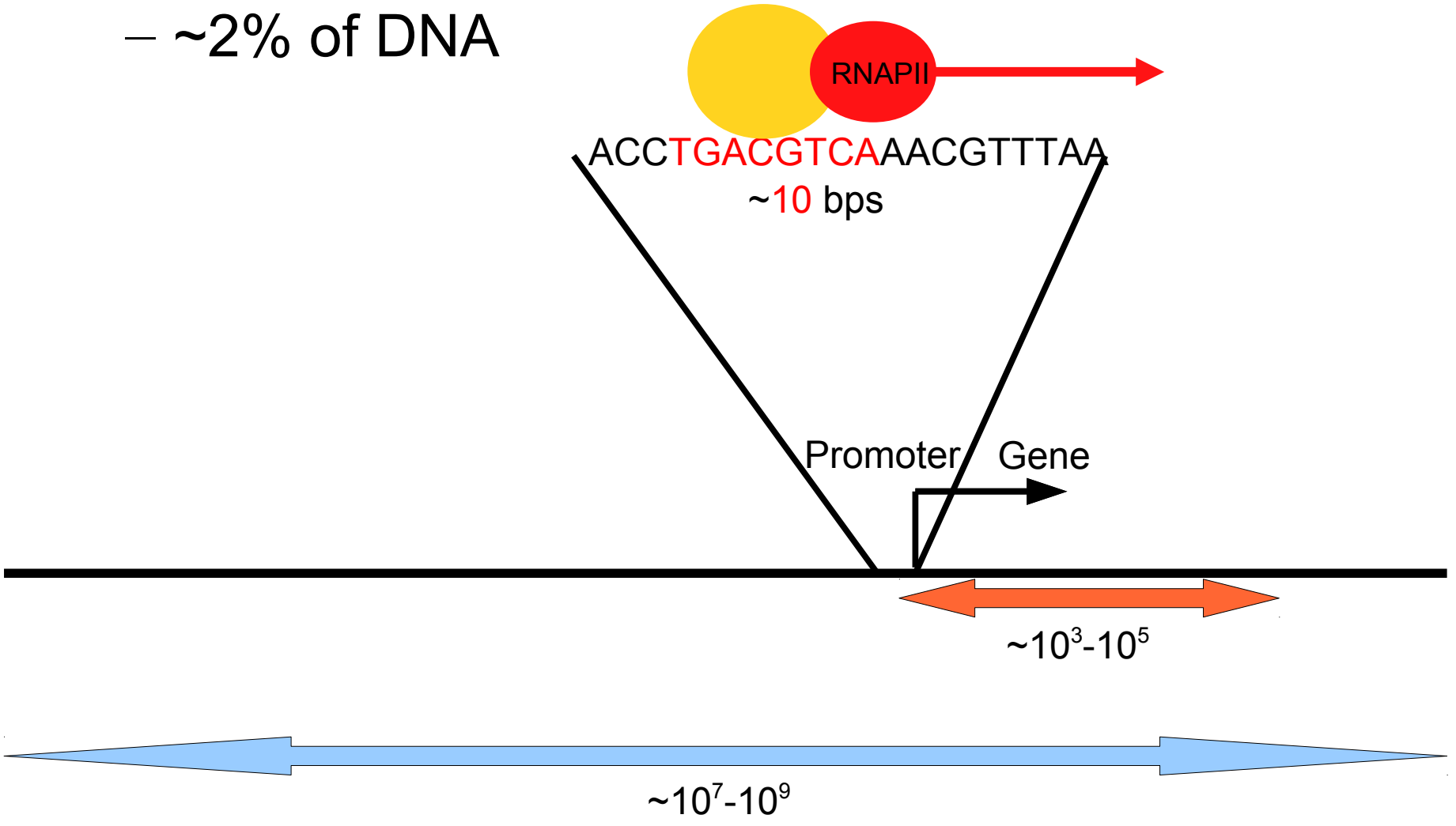
Transcription Factors (**TF**s) bind to DNA motifs

- ~25,000 genes
  - ~2% of DNA

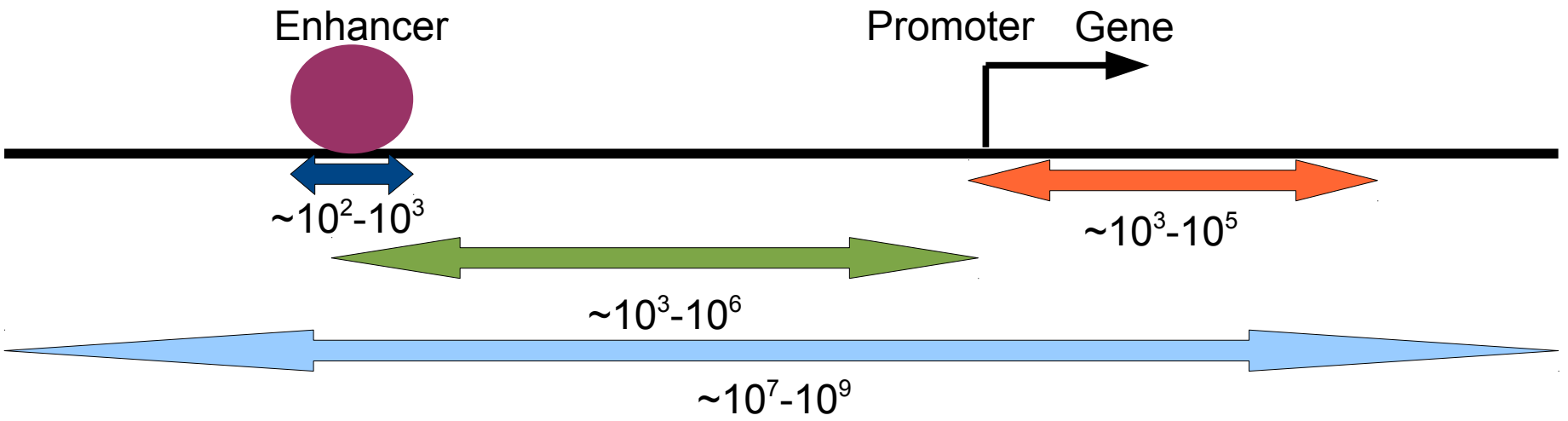


Transcription factors bind at promoter to recruit RNA Polymerase II (**RNAPII**)

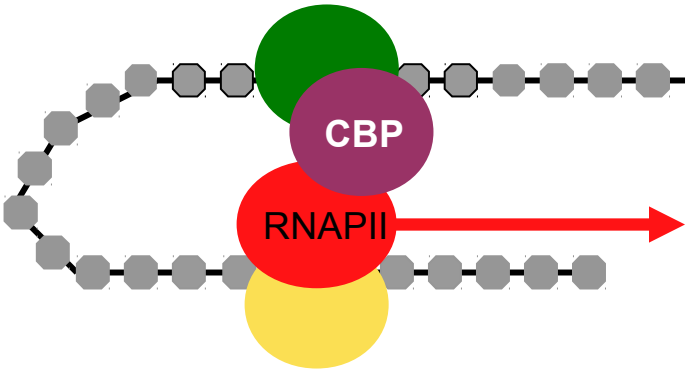
- ~25,000 genes
  - ~2% of DNA

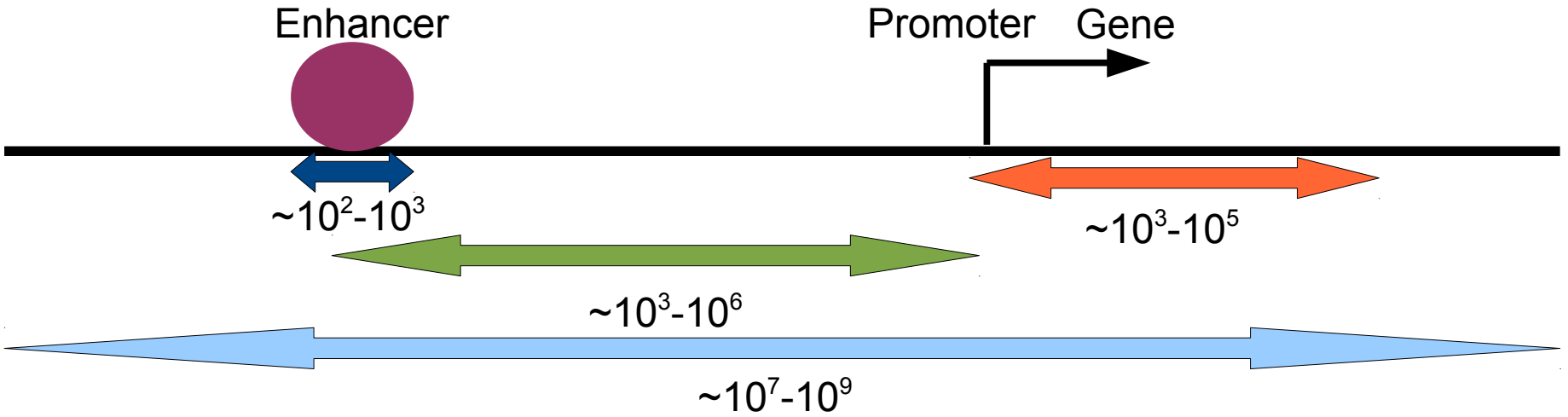


Enhancers are distal regulatory sequences

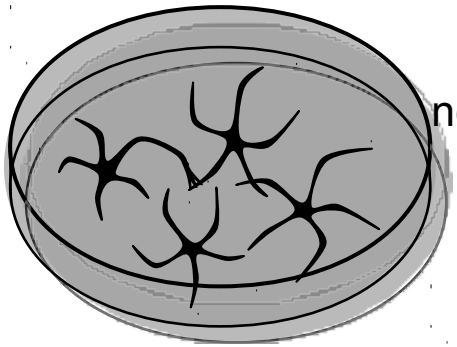


Enhancers characterized by **CBP** binding





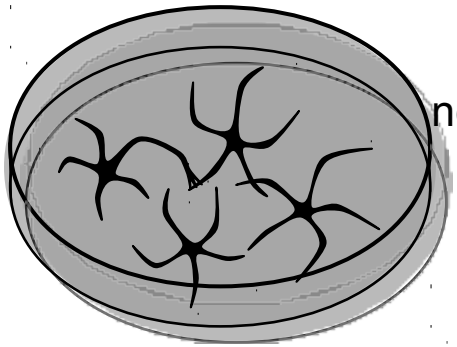
Cultured mouse cortical neurons for genome-wide study of activity dependent gene expression



mouse cortical neurons

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

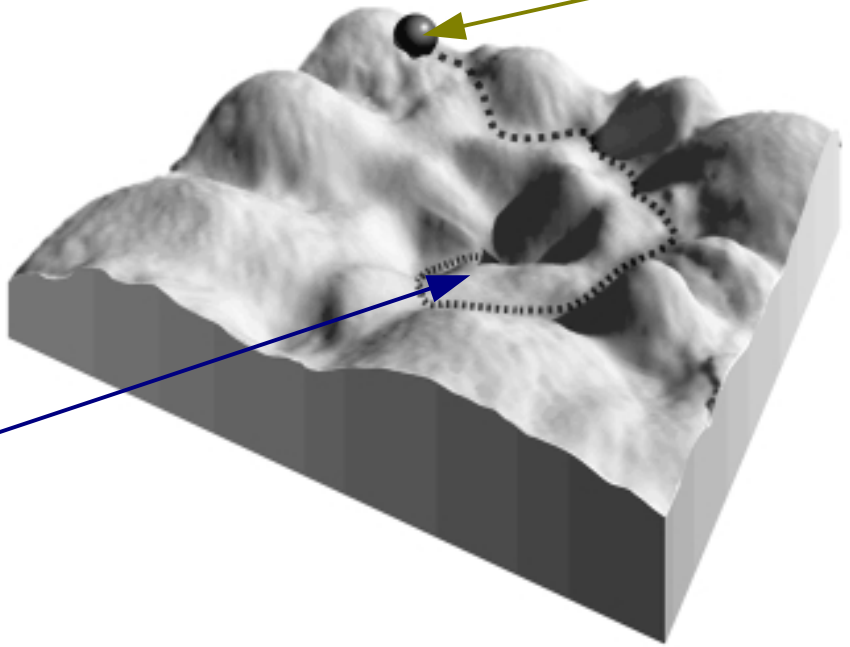
Potassium chloride (**KCl**) stimulation induces cells to change state



mouse cortical neurons

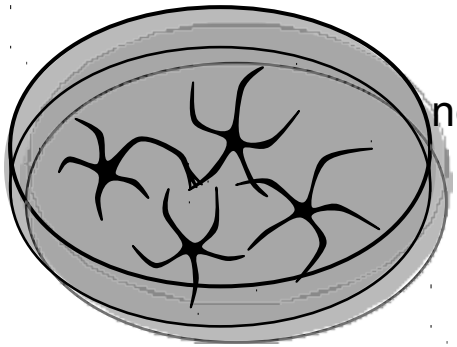
neuronal activation via potassium chloride (**KCl**) depolarization

Unstimulated (un)



Stimulated (KCl)

Genome-wide data obtained using high-throughput sequencing



mouse cortical neurons

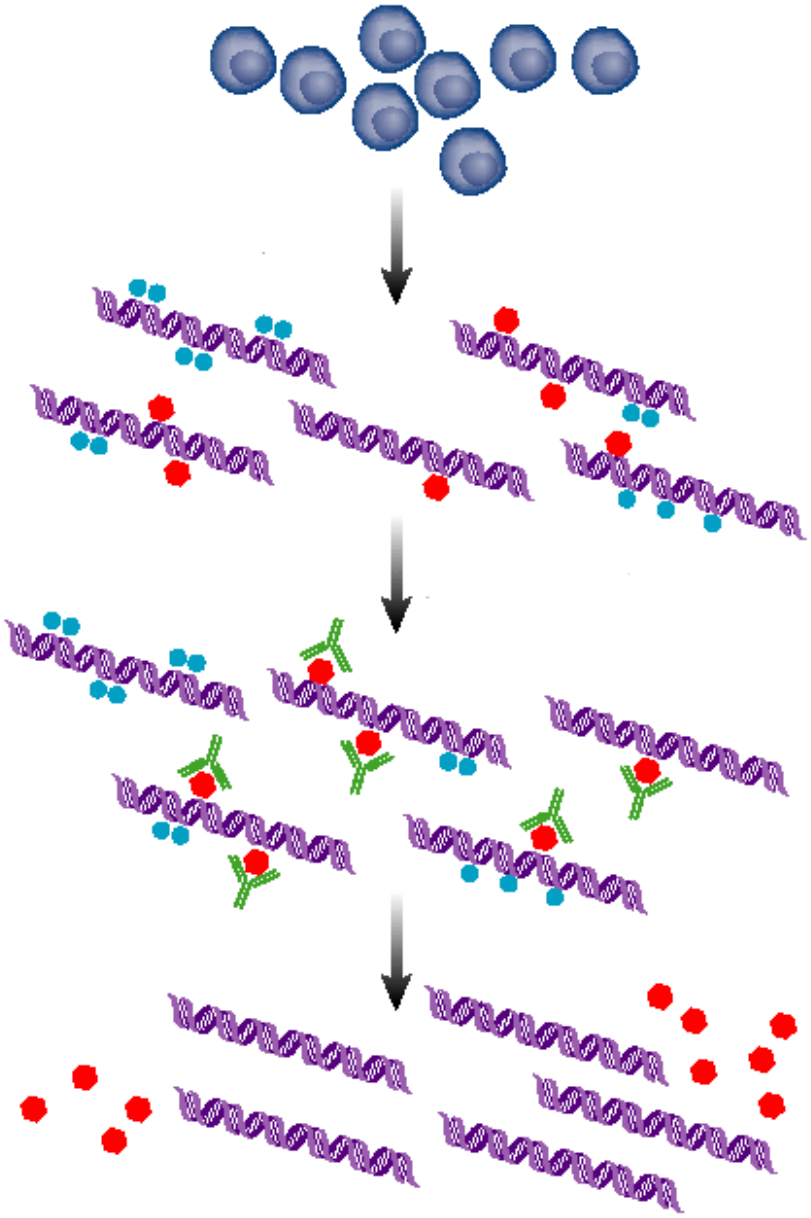
neuronal activation via potassium chloride (**KCl**) depolarization

- KCl	+ KCl
ChIP-Seq	ChIP-Seq
RNA-Seq	RNA-Seq

Jesse Gray
Tae-Kyung Kim
Greenberg Lab

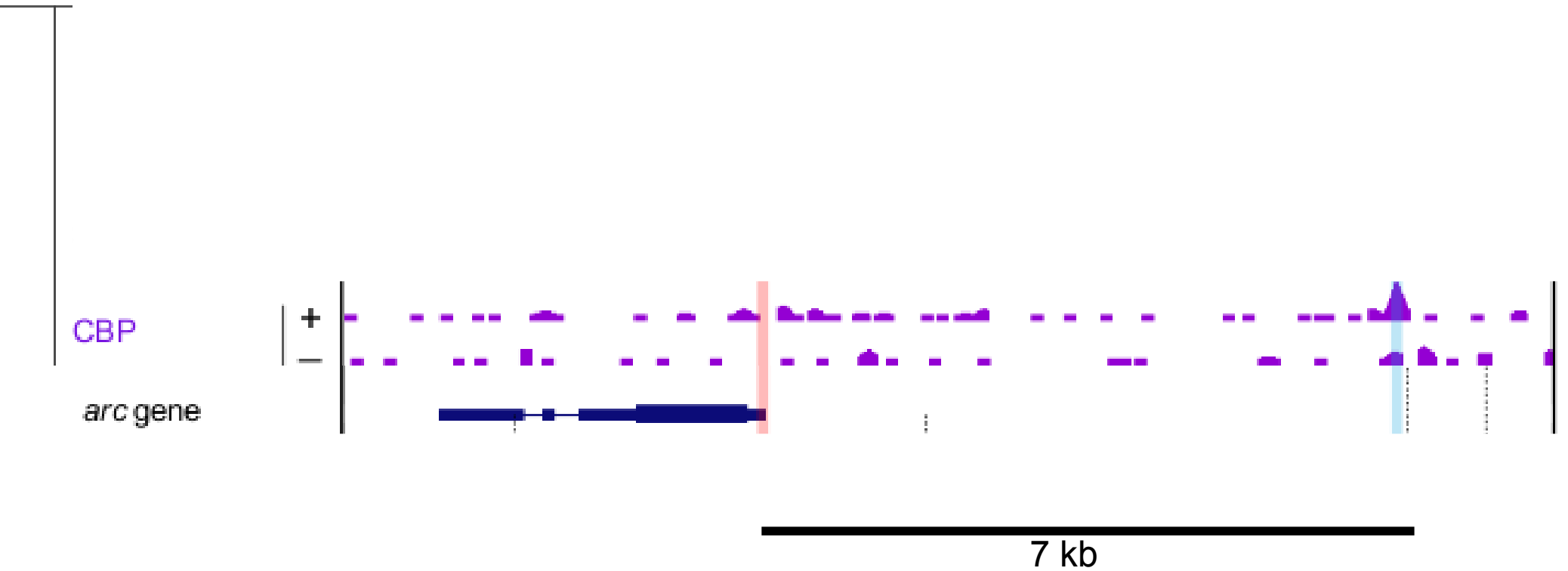
Chromatin immunoprecipitation and sequencing (**ChIP-Seq**) finds protein binding sites *in vivo*

- Short **reads** mapped to reference genome
- #reads ~ binding
- ~10^6 reads
- Unbiased



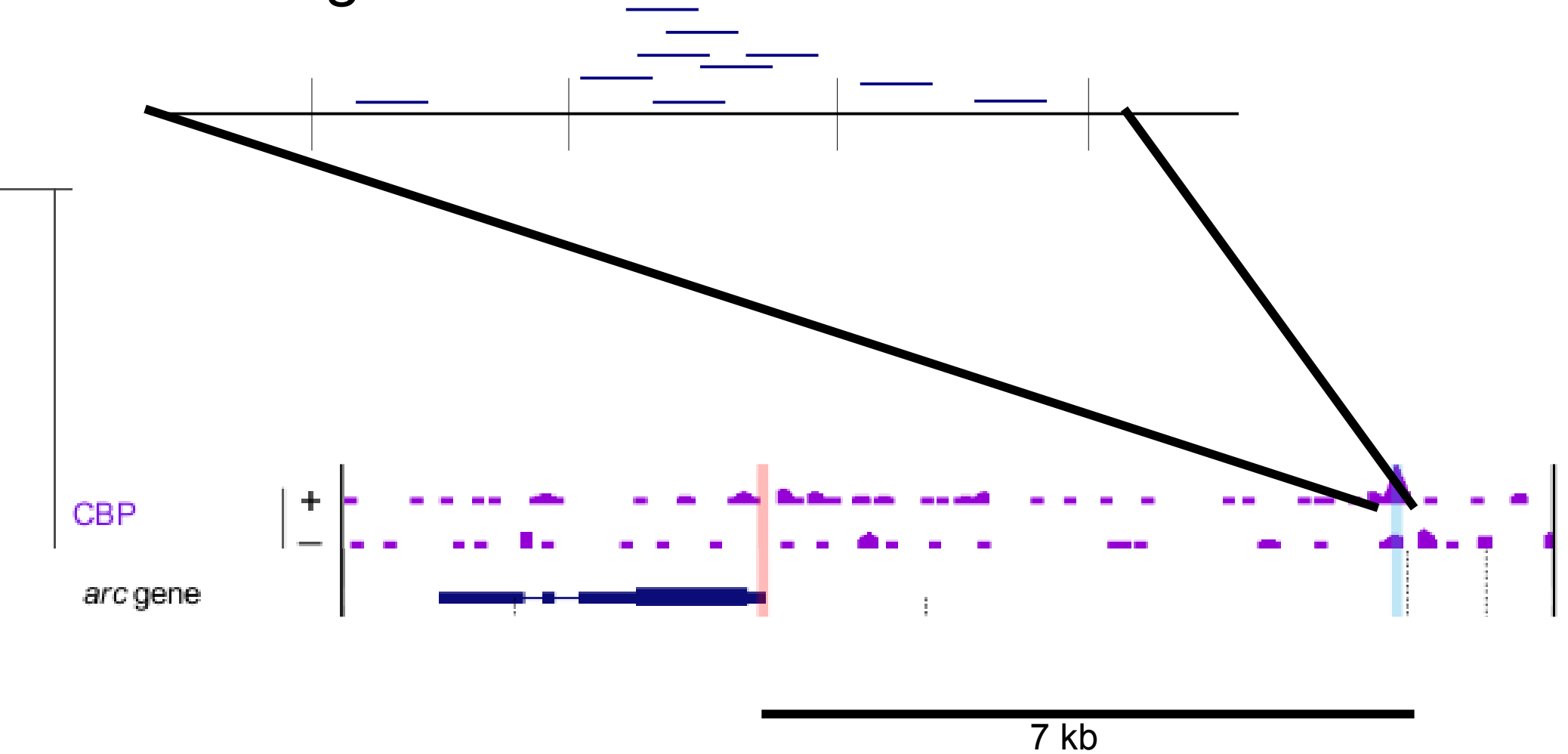
(Mardis, 2007)

Inducible CBP binding at enhancers

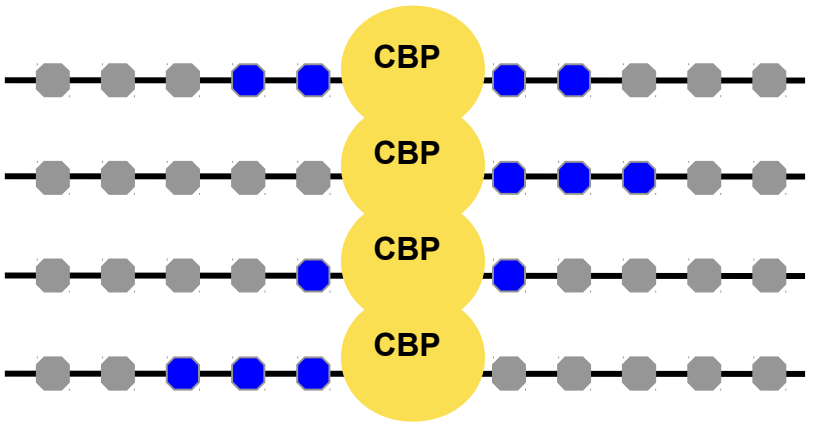


Identifying ~28,000 CBP binding sites in two replicate experiments

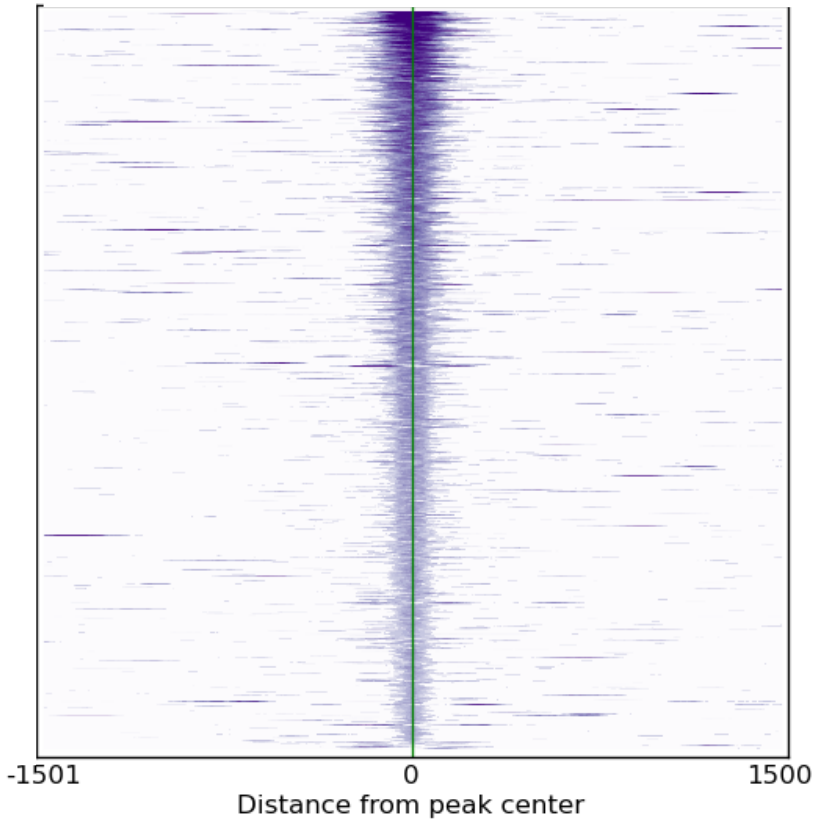
- Regions that have significantly more CBP than background

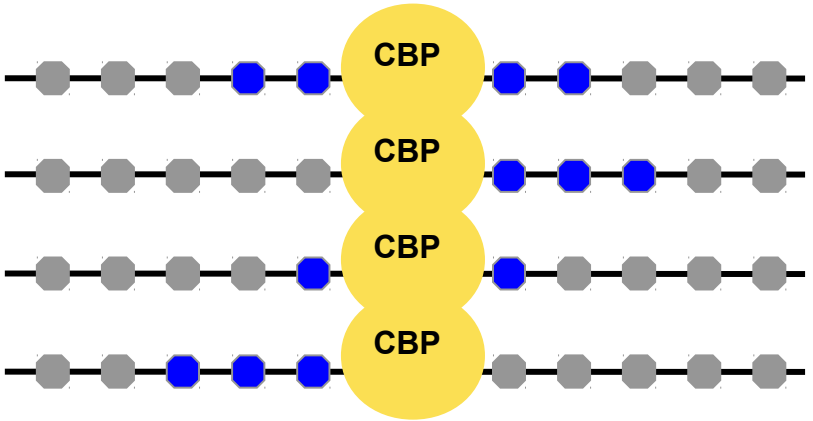


Aligning CBP peaks to calculate binding profiles

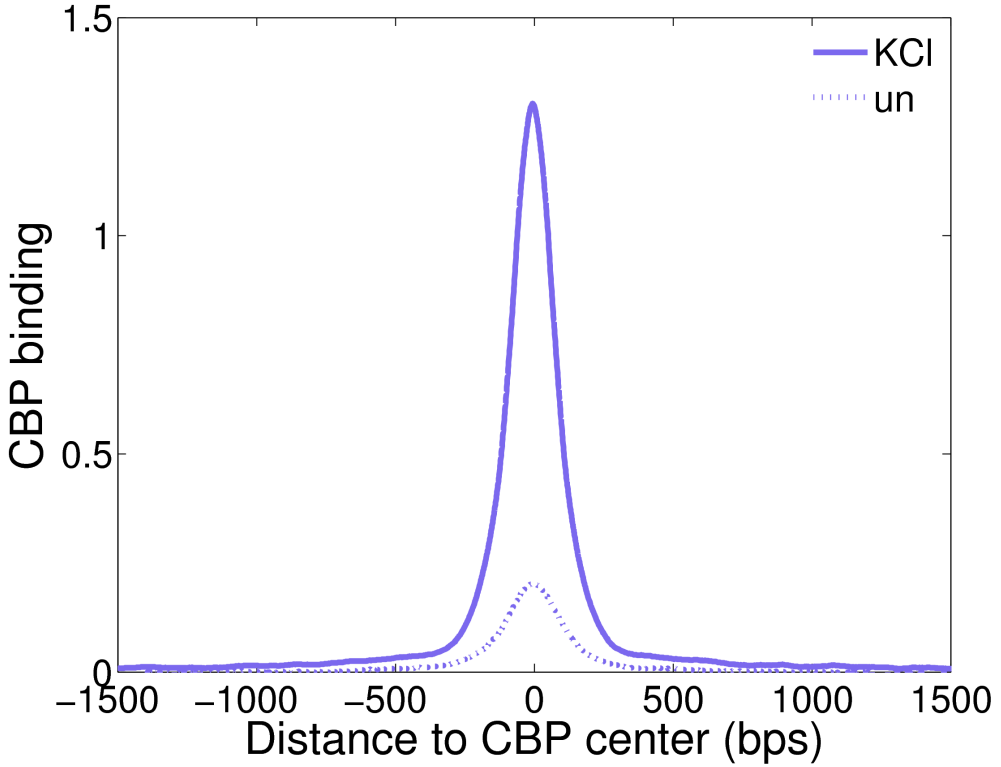


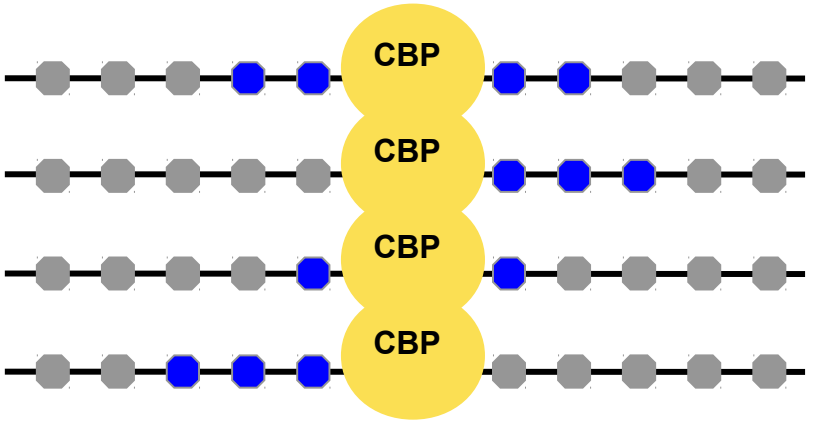
Aligning CBP peaks to calculate binding profiles





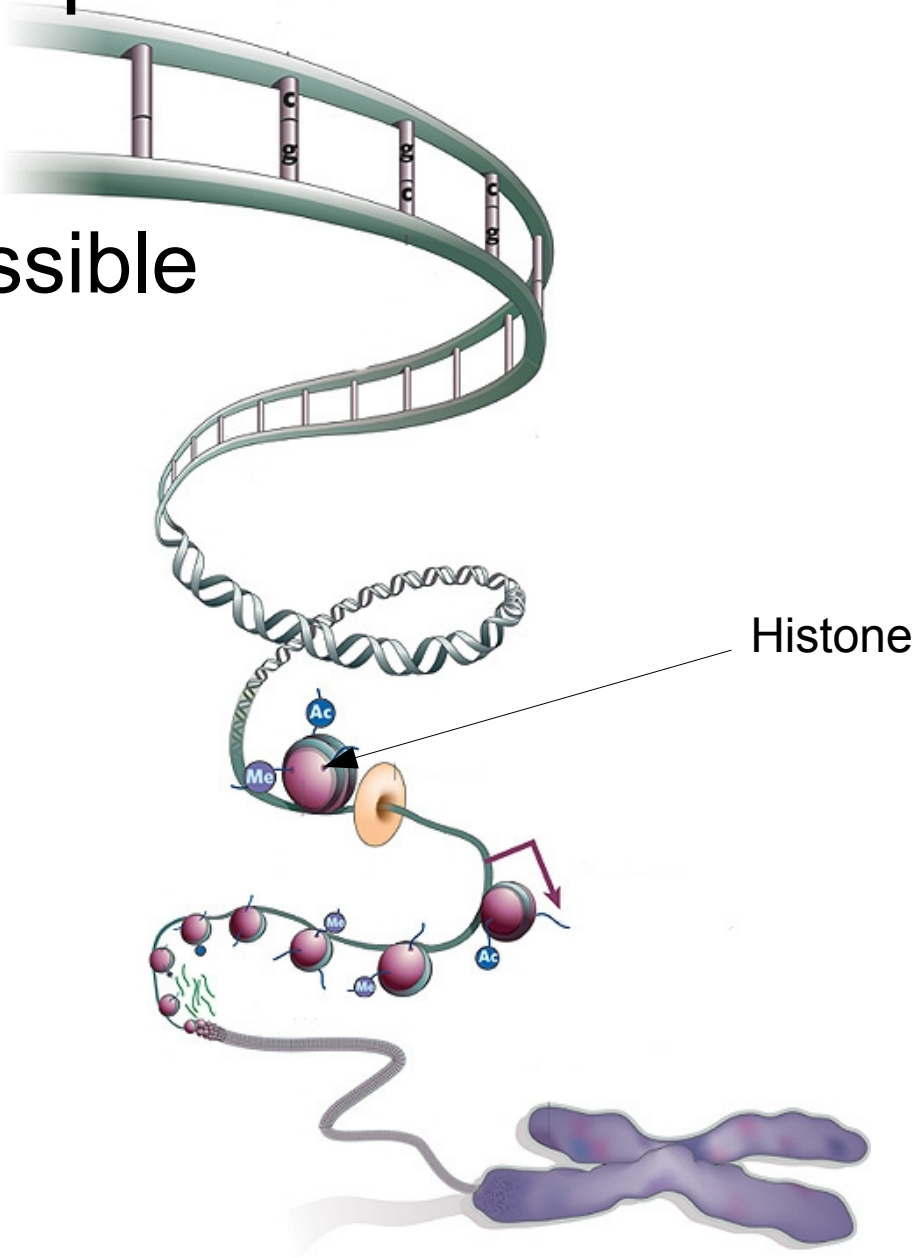
Average profile of CBP binding





Histones prevent transcription factors from binding to DNA

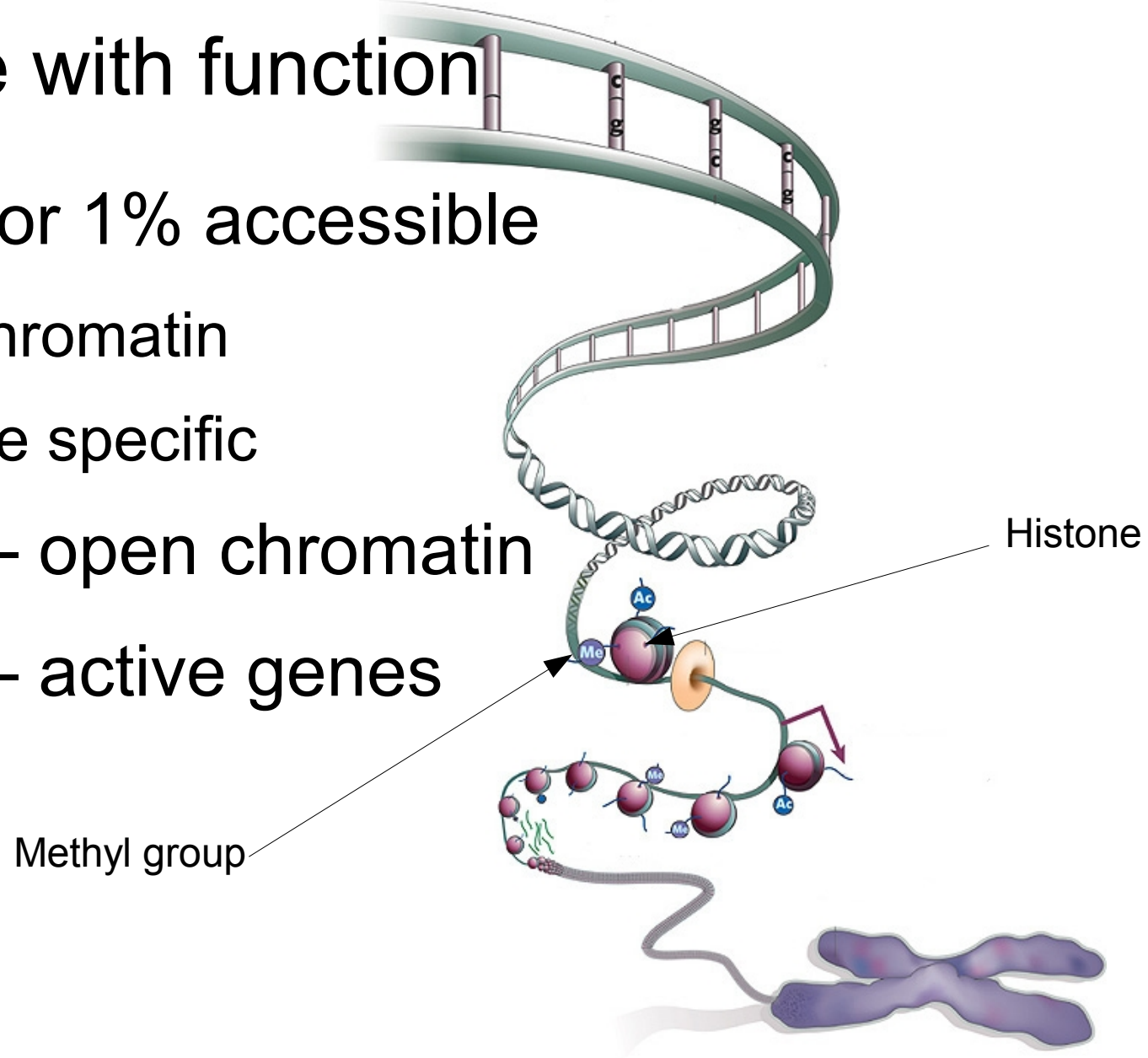
- ~100 k loci or 1% accessible
  - Open chromatin
  - Cell-type specific



(ENCODE, 2007)

Post-translational modifications of histone tails correlate with function

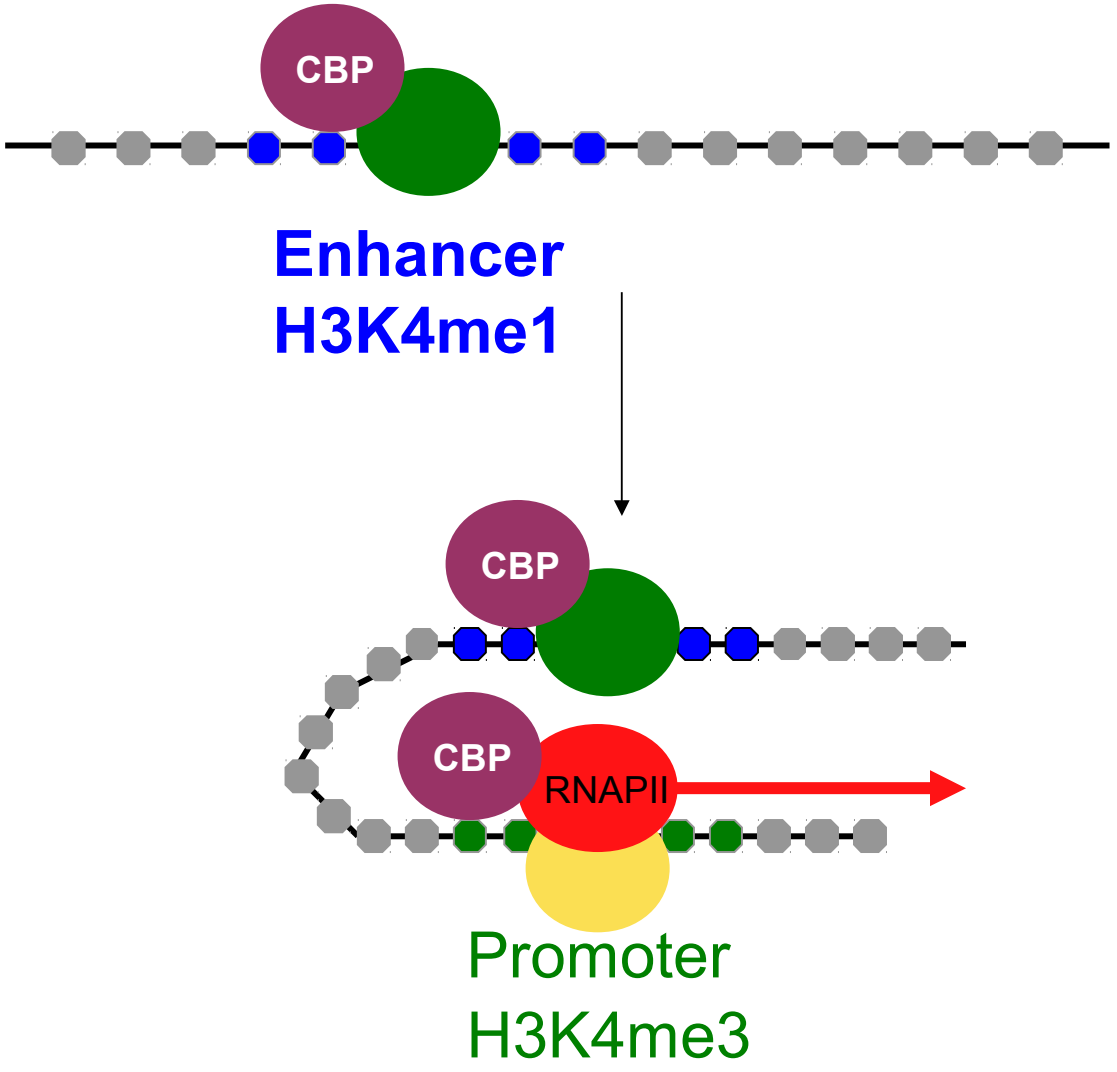
- ~100 k loci or 1% accessible
  - Open chromatin
  - Cell-type specific
- **H3K4me1** – open chromatin
- **H3K4me3** – active genes



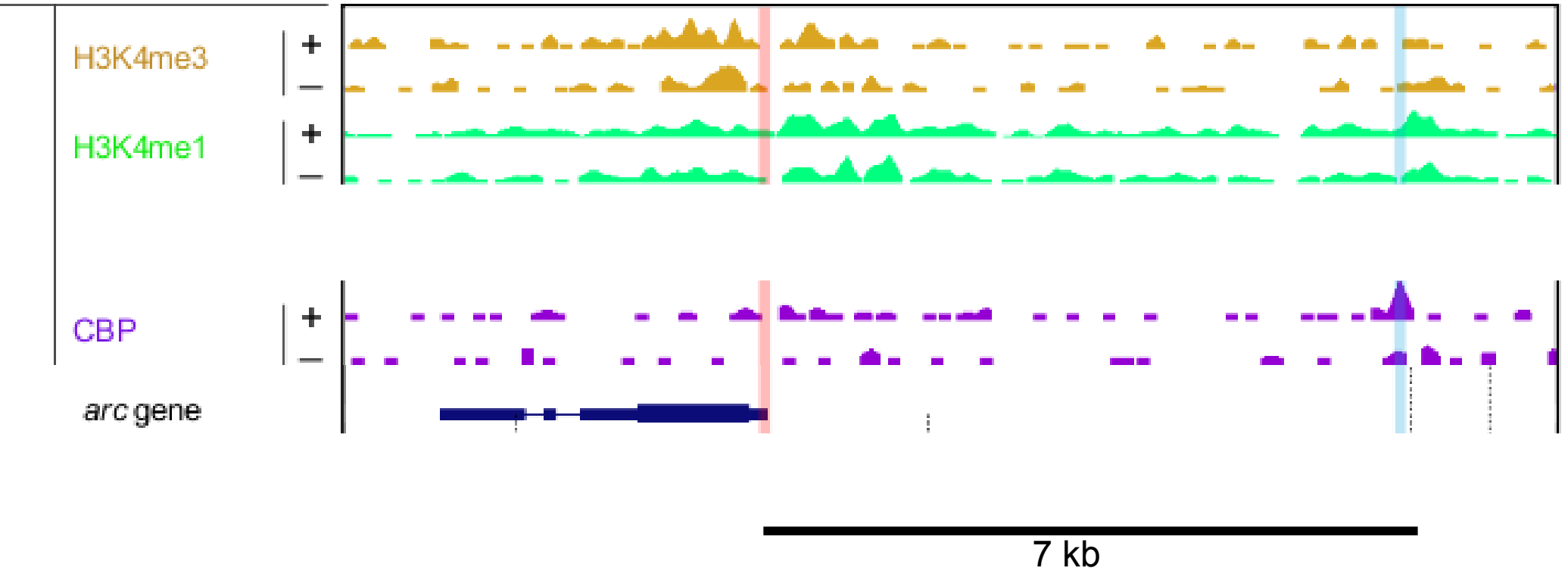
(ENCODE, 2007)

A combination of CBP and histone modifications identifies putative enhancers

- **CBP** binding
- **H3K4me1** flanking
- **H3K4me3** absent
  - Many unannotated promoters in the genome



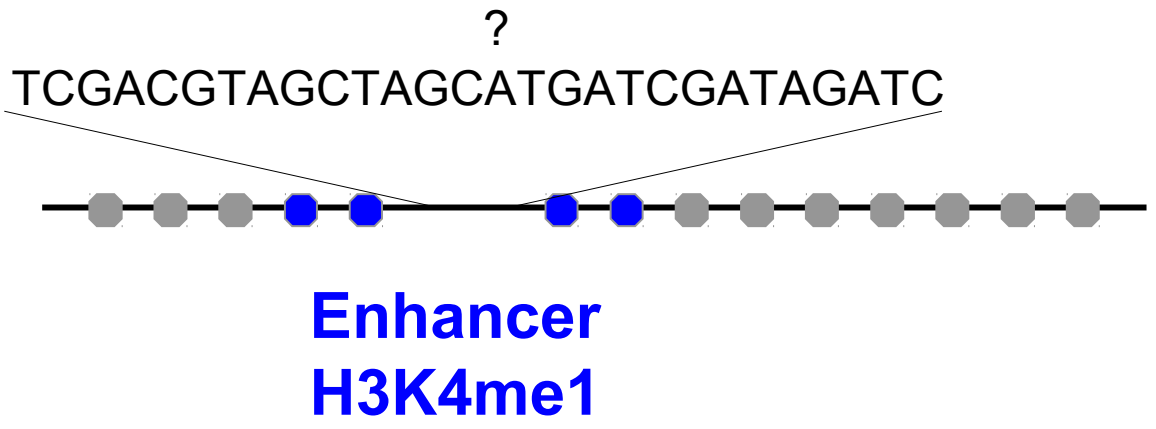
Distal CBP peaks have high levels of H3K4me1 and low levels of H3K4me3



We identified ~12,000 activity-dependent enhancers throughout the genome

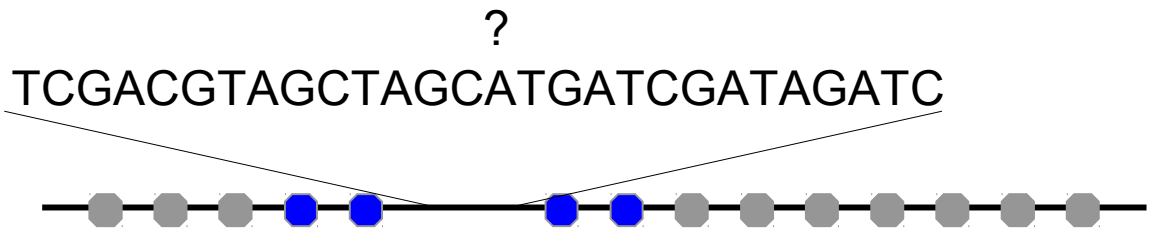
- **CBP** peak
- **High** levels of flanking **H3K4me1**
- **Low** levels of **H3K4me3**
  - Independently tested and validated 8 enhancers

What TFs bind to enhancers?



- CBP -
  CREB Binding Protein
  - >50 partners

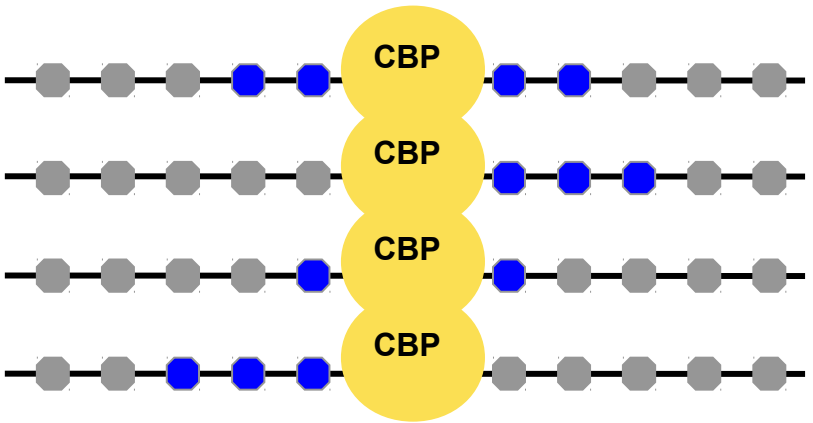
~100 enriched motifs at enhancers



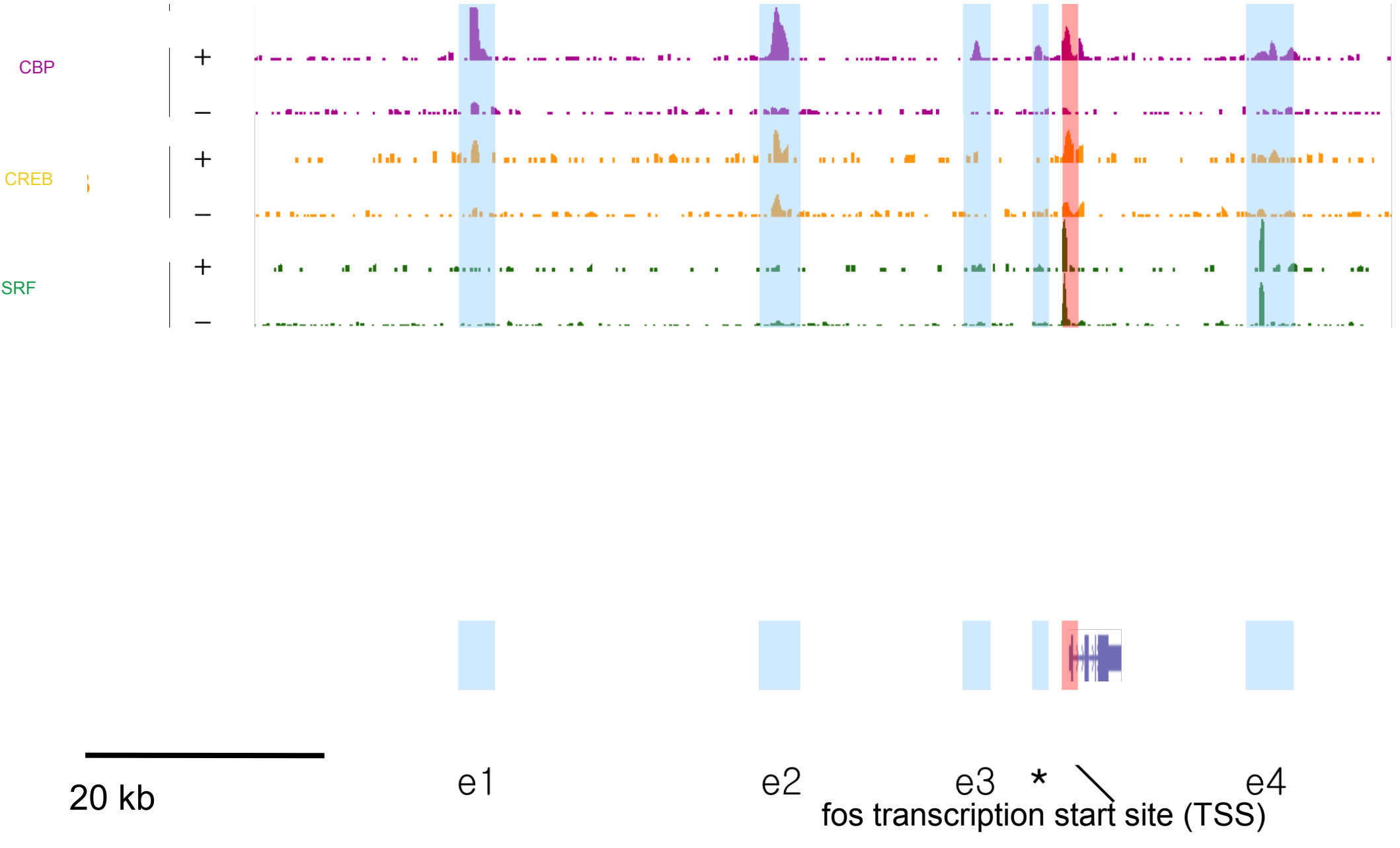
Enhancer
H3K4me1

TCAGGCTGATGACGTCAAACCGTCGTTA
ACCTTTTGACGTCAAATTTACGCTAGTAT
TCGACGTAGCTAGCATGATCGATAGATC
CGTGACGTCAGTGCTCGTAAATCATAAG

- CBP -
  - CREB Binding Protein
    - >50 partners

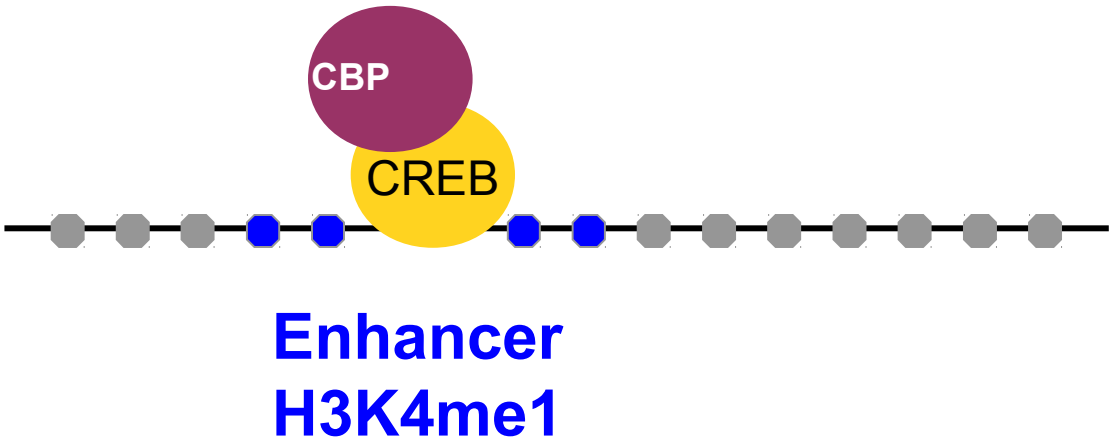


SRF and CREB binding at Fos enhancers

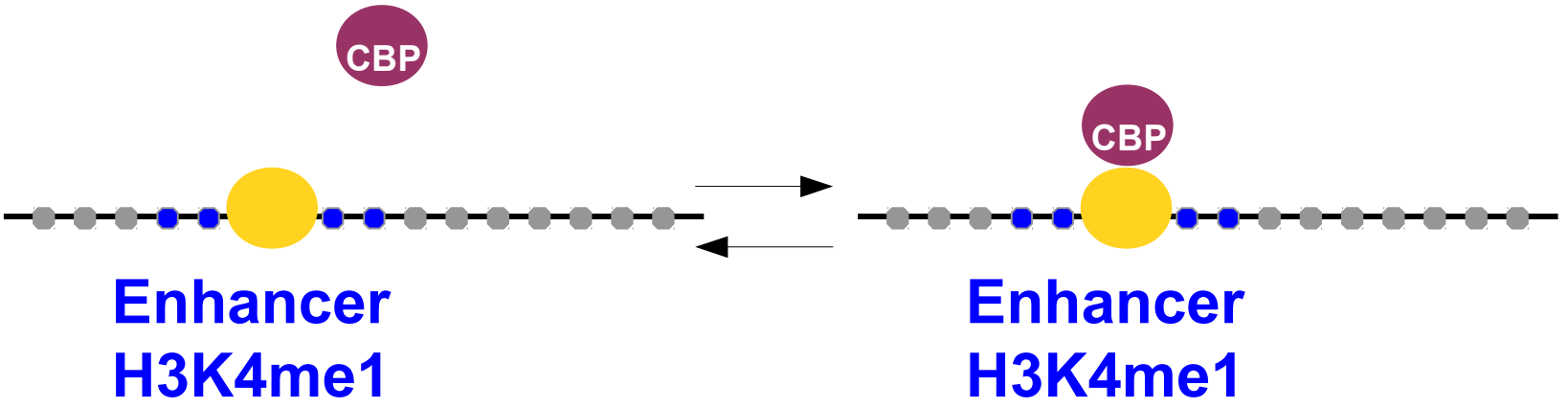


Is CBP binding determined by other TFs?

- Enriched for ~100 sequence motifs
- ChIP-seq reads predicted by sequence

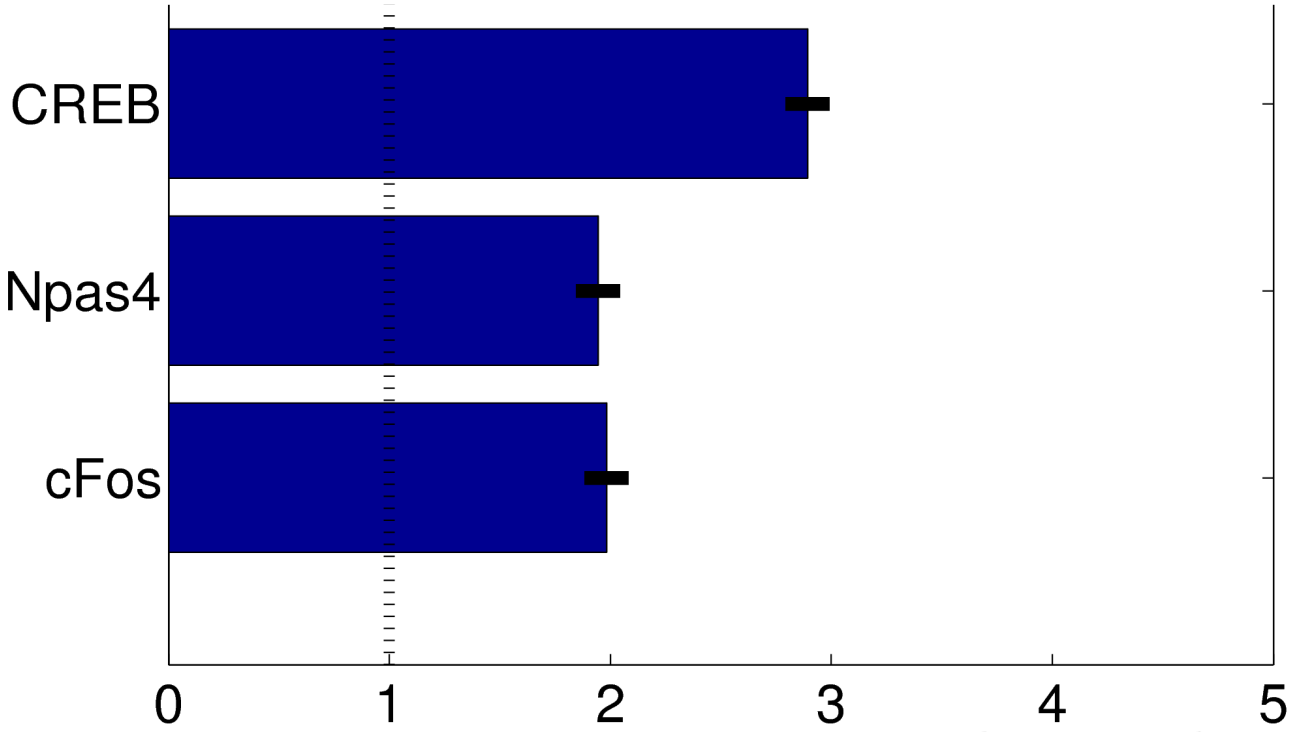


TFs compete for CBP



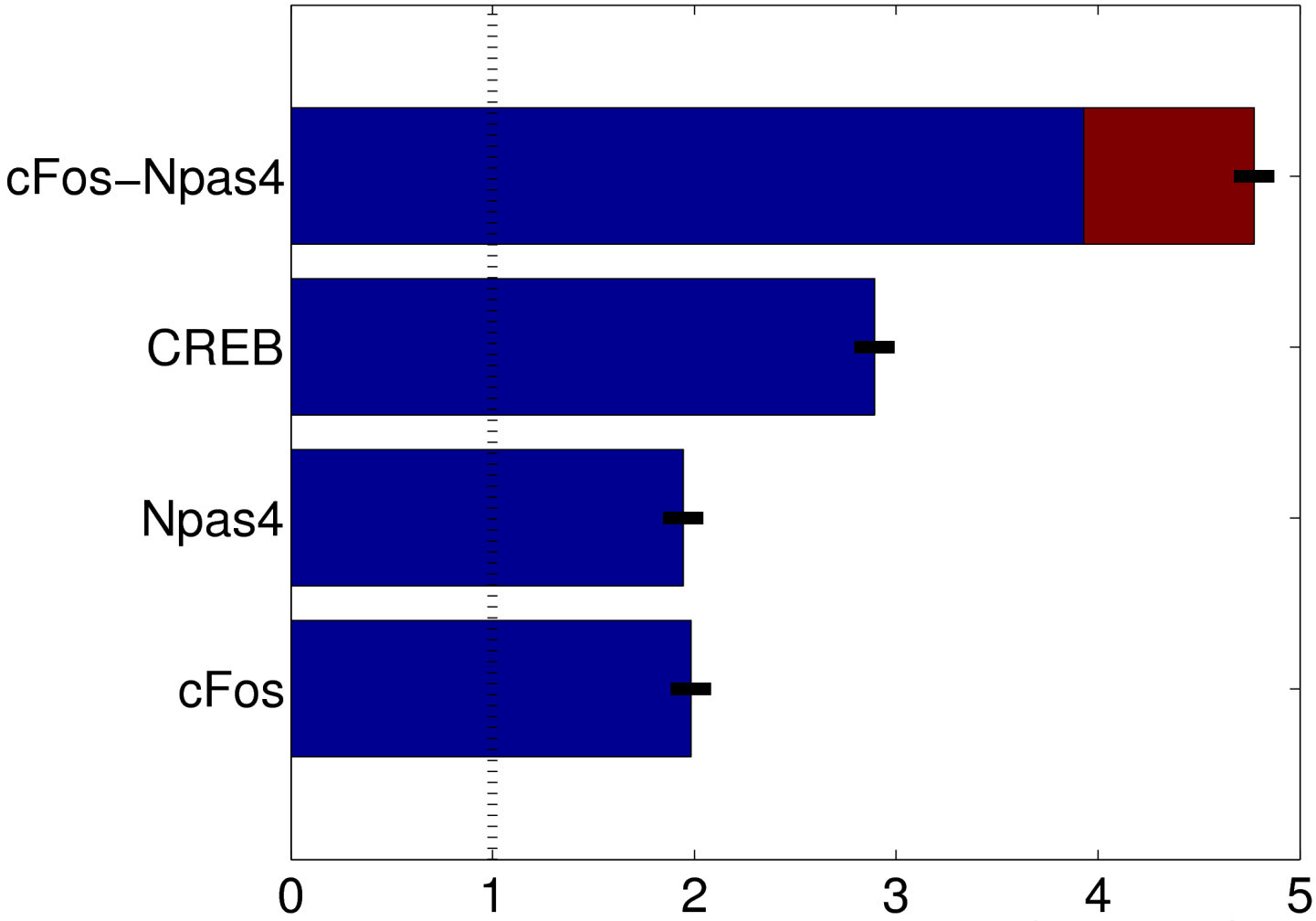
$$\text{CBP affinity} = \frac{\#\text{peaks with CBP}}{\#\text{peaks without CBP}}$$

CBP binding determined by affinity of TF



$$\text{CBP affinity} = \frac{\#\text{peaks with CBP}}{\#\text{peaks without CBP}}$$

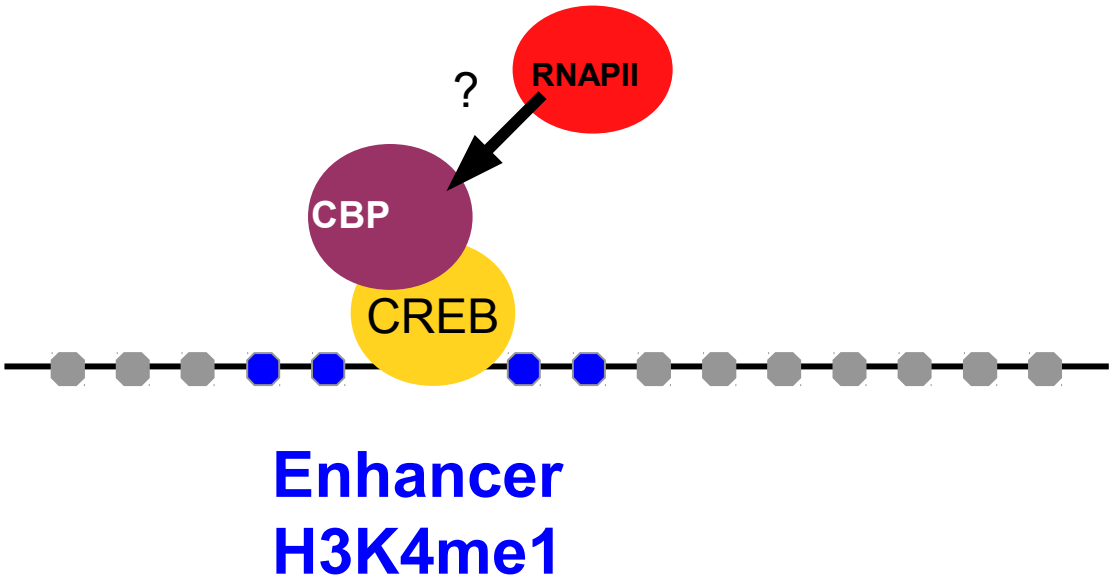
Synergistic effects for combinations of TFs



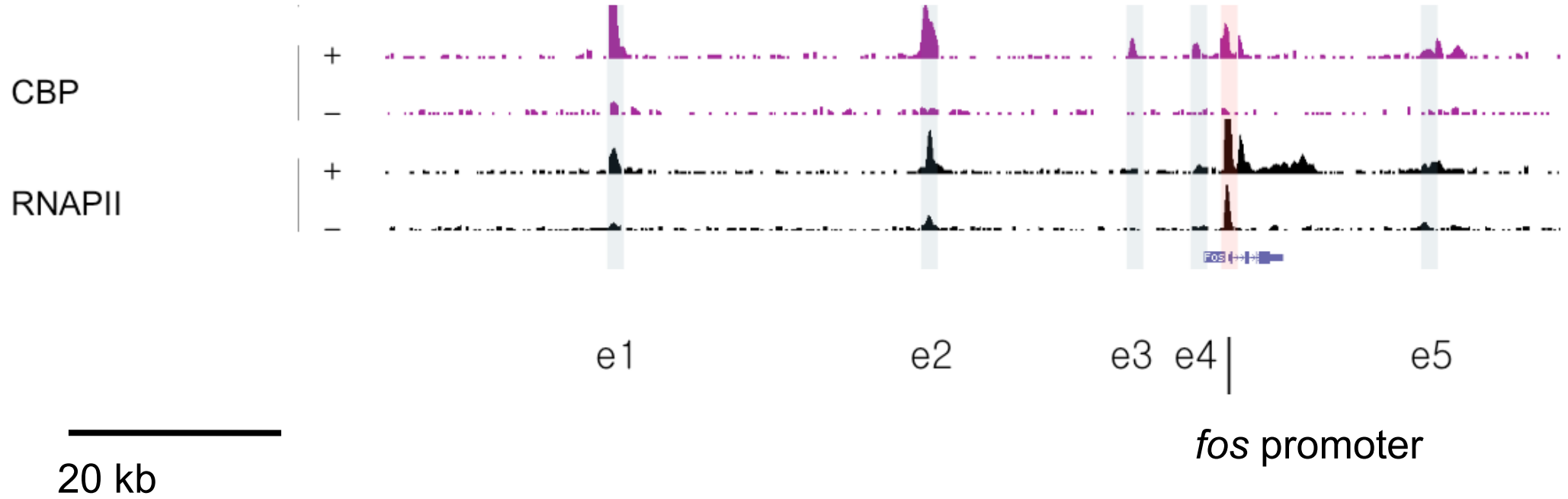
$$\text{CBP affinity} = \frac{\#\text{peaks with CBP}}{\#\text{peaks without CBP}}$$

What is the function of CBP at enhancers?

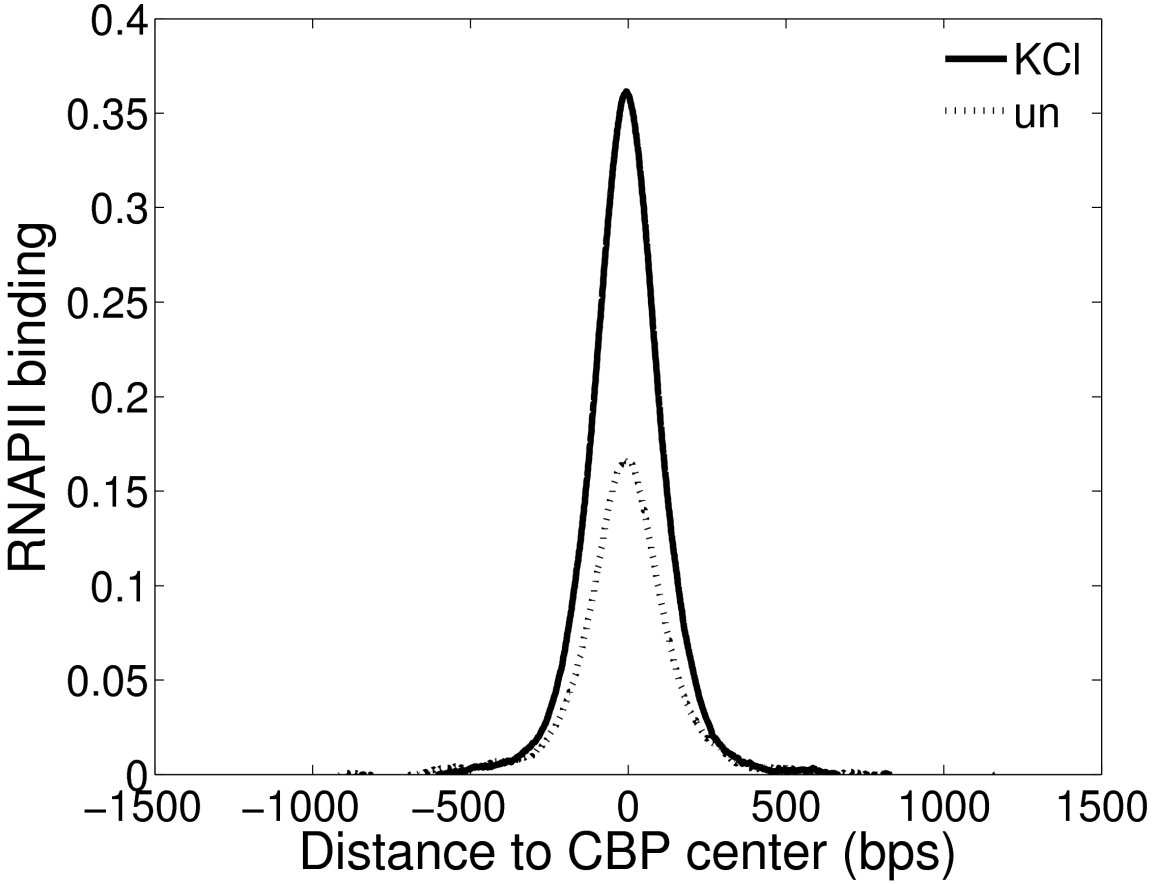
- Enriched for ~100 sequence motifs
- ChIP-seq reads predicted by sequence
- CBP binding determined by other TFs

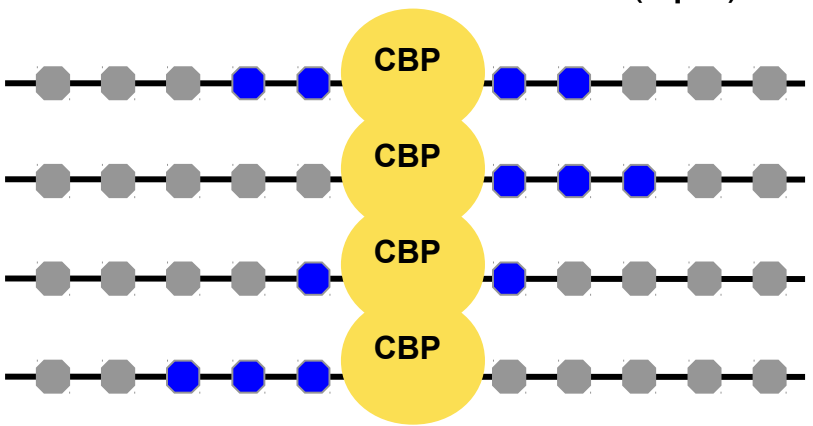


RNAPII is recruited to CBP binding sites at the *fos* locus

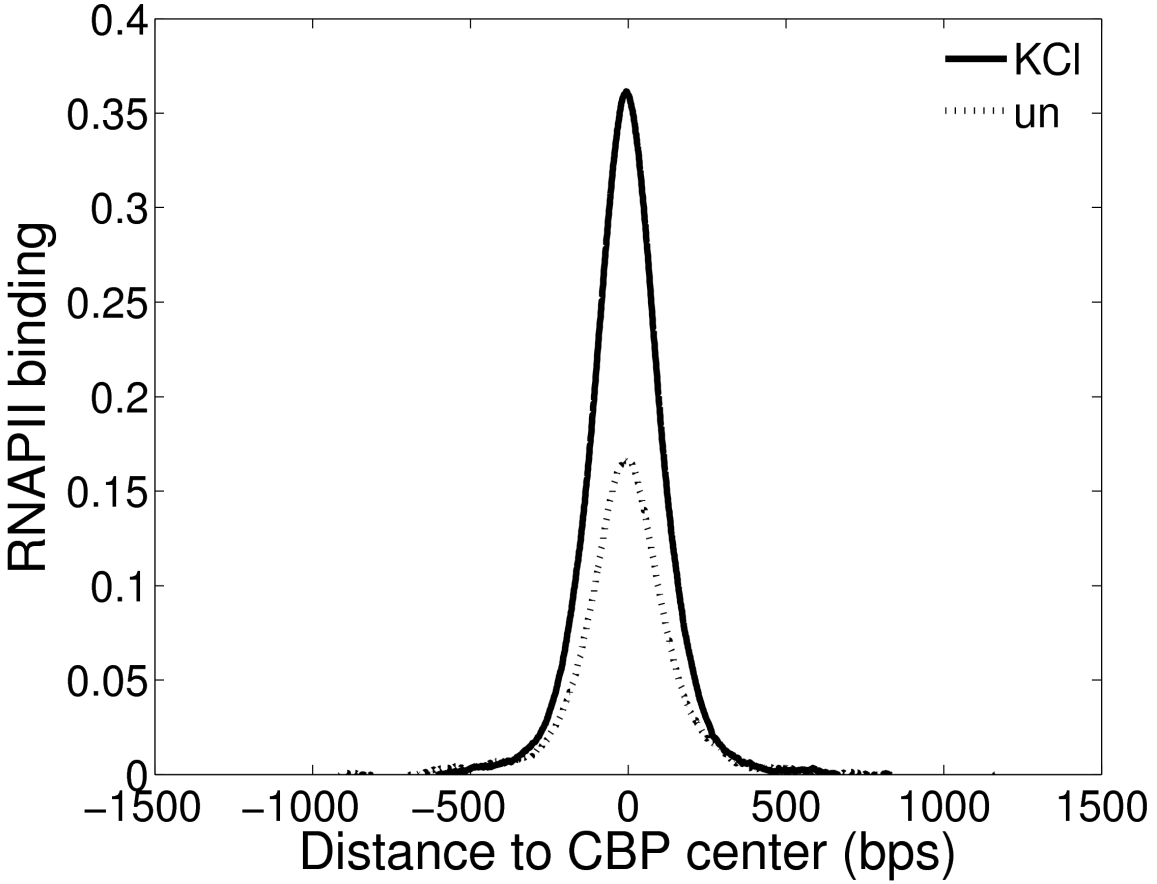


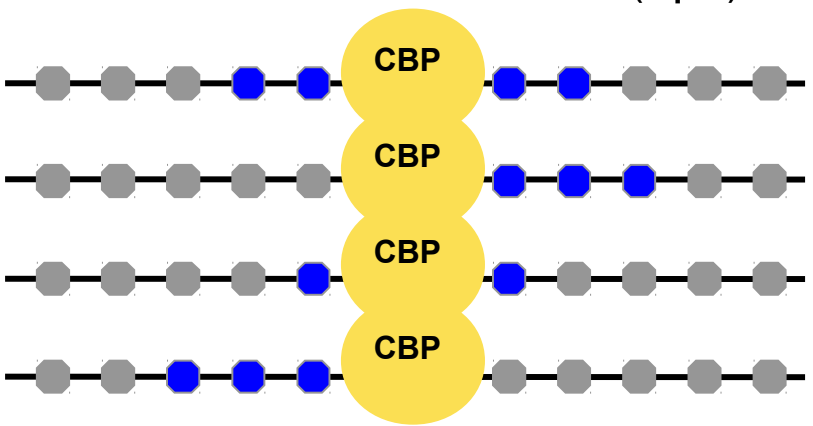
RNAPII is recruited at enhancers

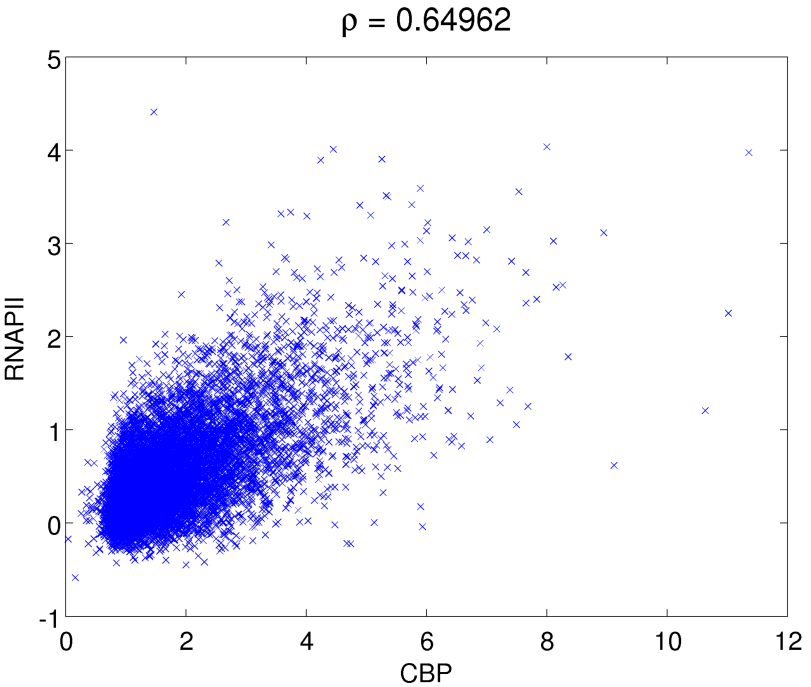




RNAPII is correlated with CBP

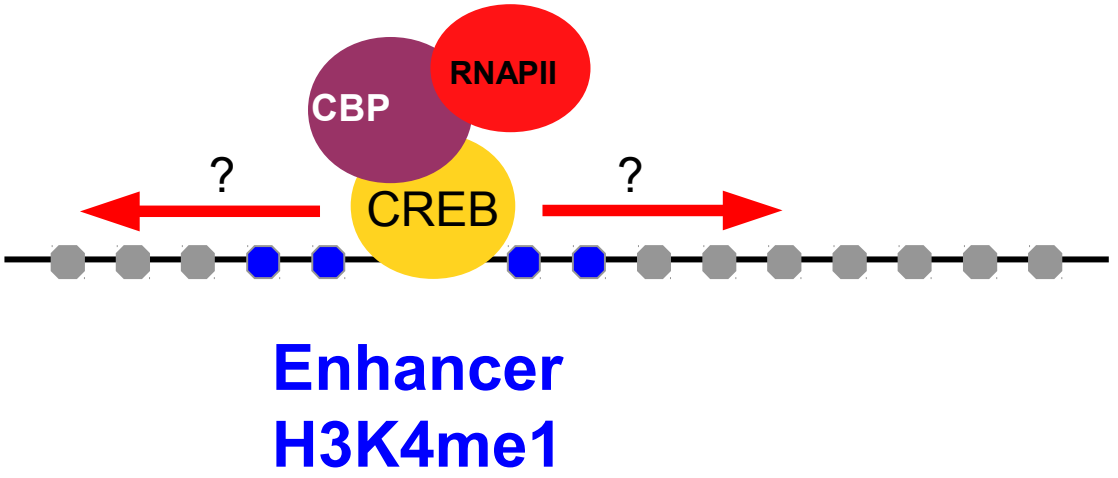






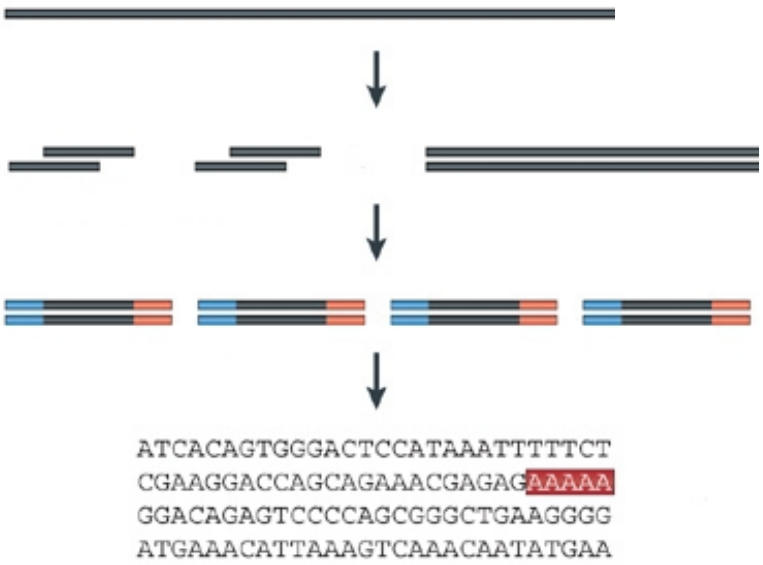
What is the function of RNAPII at enhancers?

- Enriched for ~100 sequence motifs
- ChIP-seq reads predicted by sequence
- CBP binding determined by other TFs
- CBP recruits RNAPII



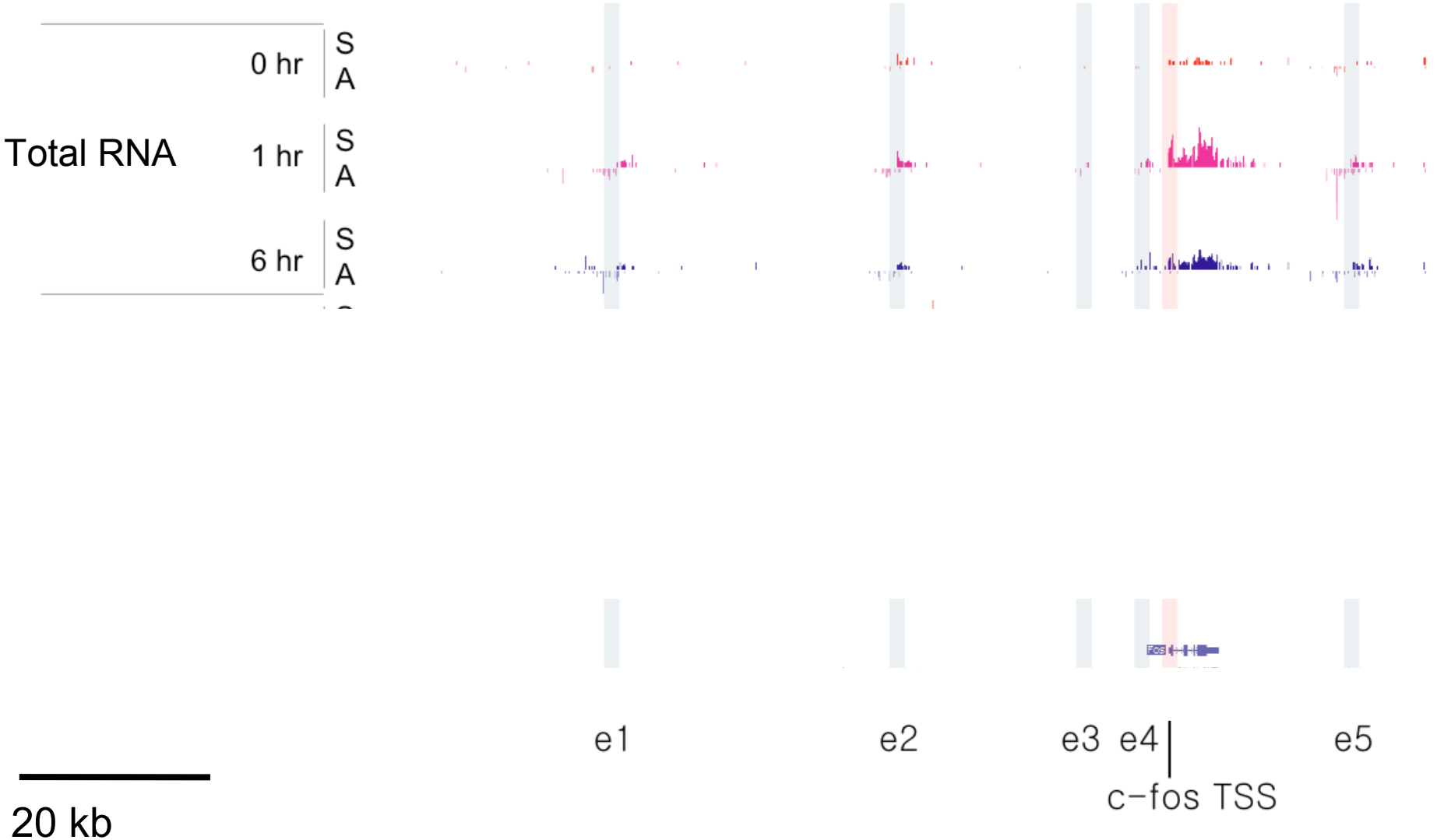
RNA-Seq finds transcribed parts of the genome

- Short **reads** mapped to reference genome
- ~5x10^6 reads
- #reads ~ RNA

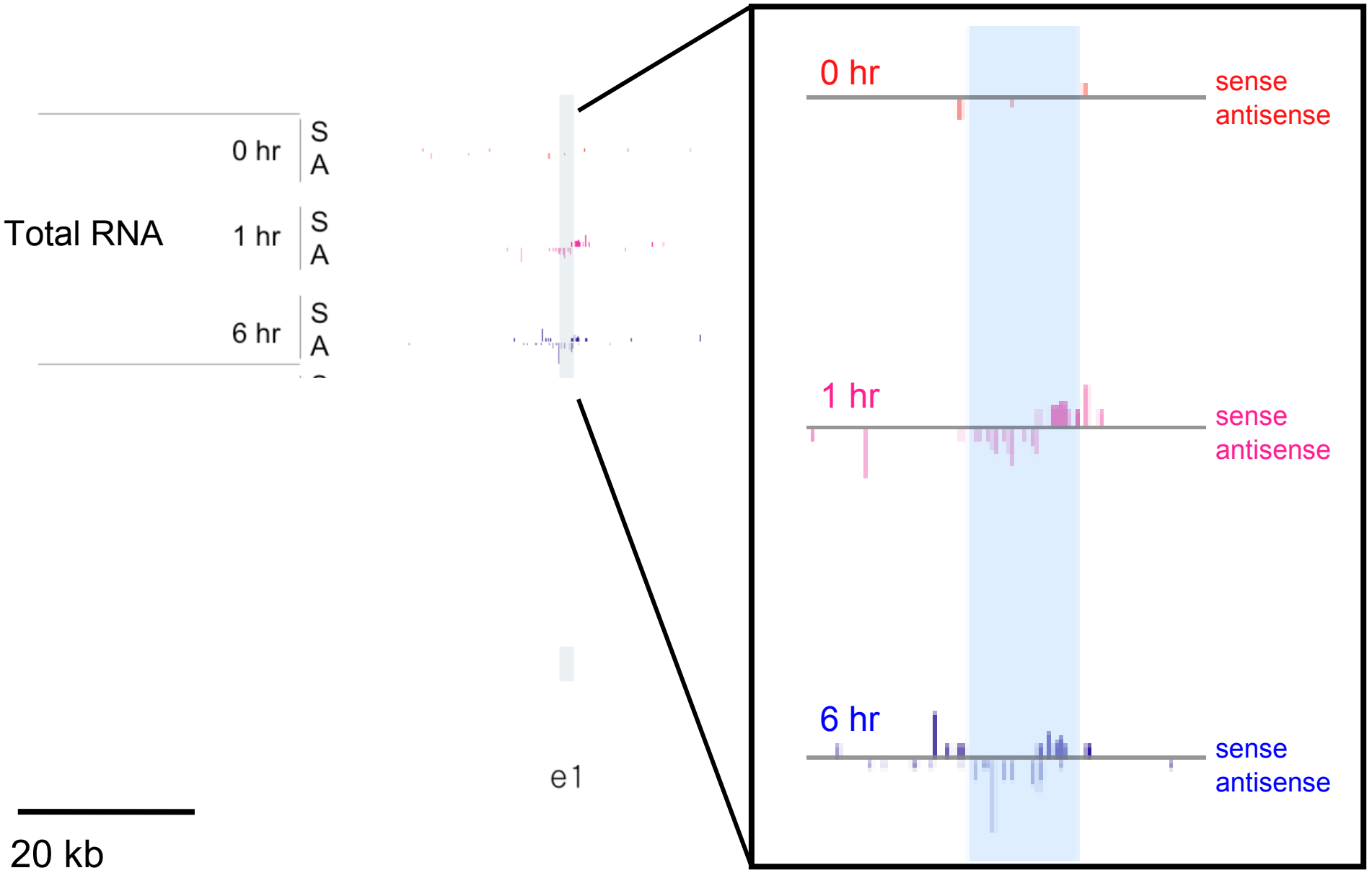


(Wang et al, 2009)

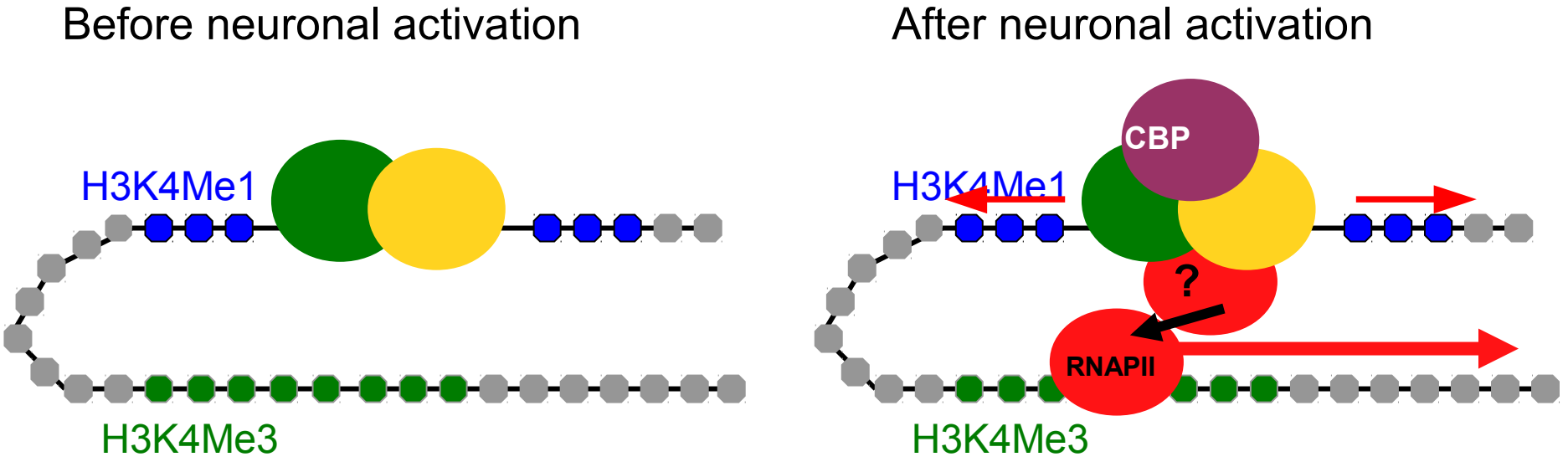
Transcription of total RNA at the *fos* locus



Transcription at enhancers is activity-dependent

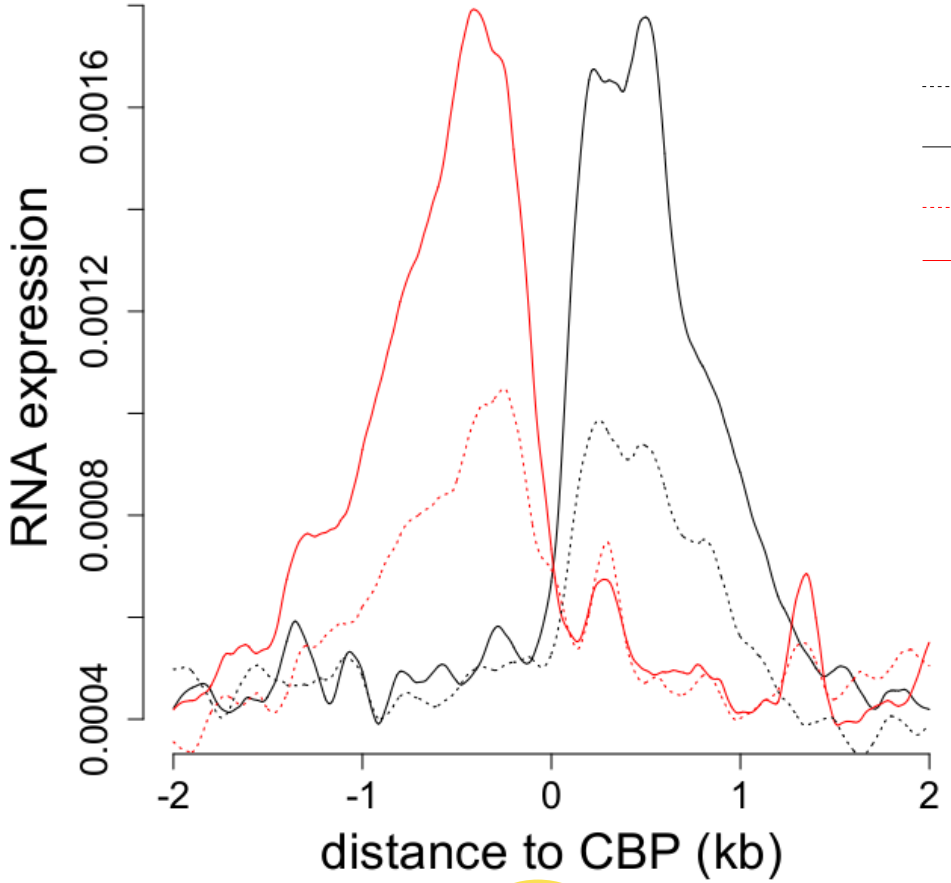


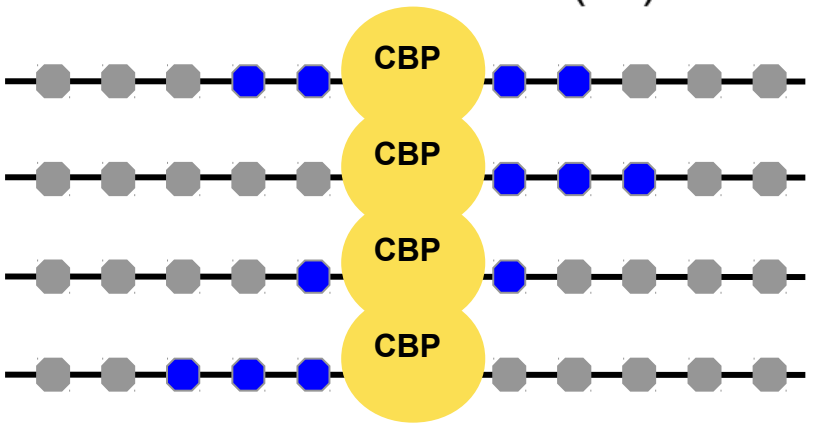
Enhancer RNAs (**eRNAs**) novel species



- mRNA, rRNA, tRNA, miRNA, snRNA, snoRNA, siRNA, piRNA, lncRNA, … ?

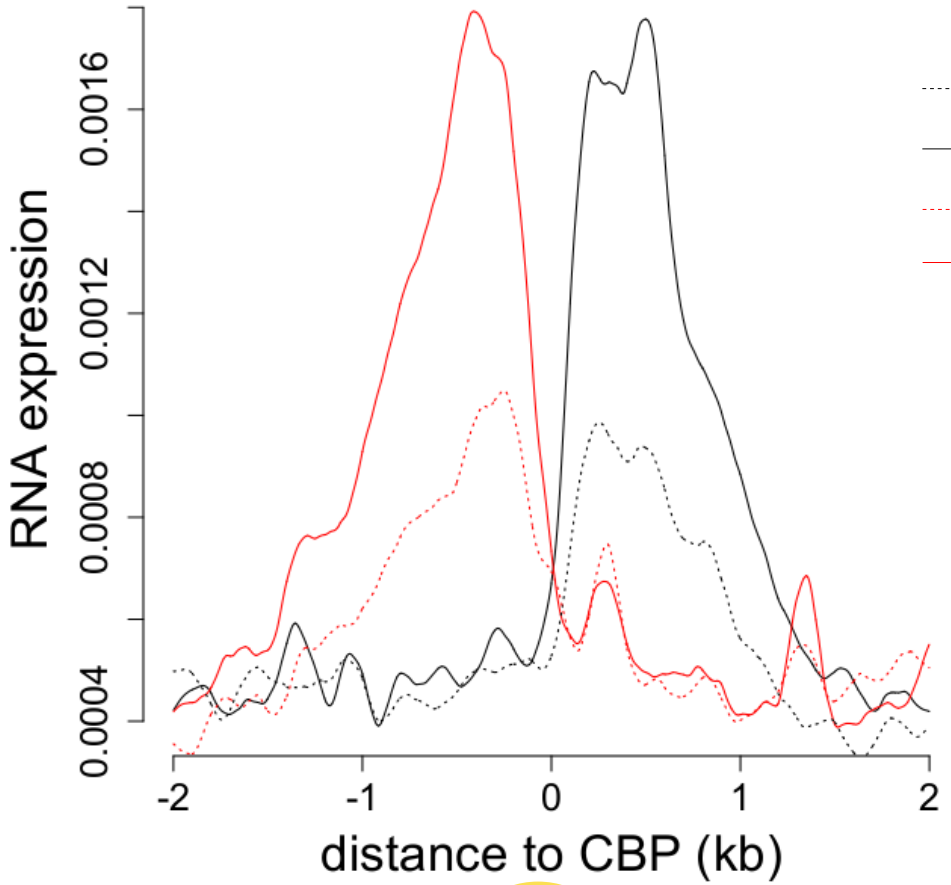
eRNAs are induced by activity

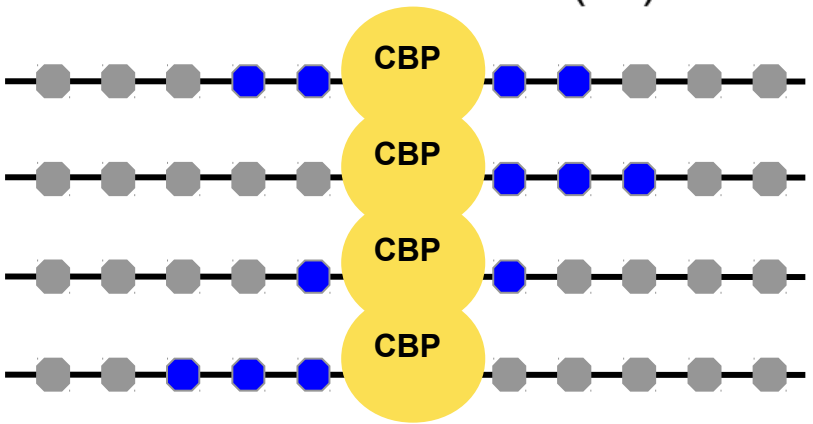




- Inducible, 2-fold
- ~1 kb
- Bidirectional

eRNAs are 100-fold lower than mRNAs

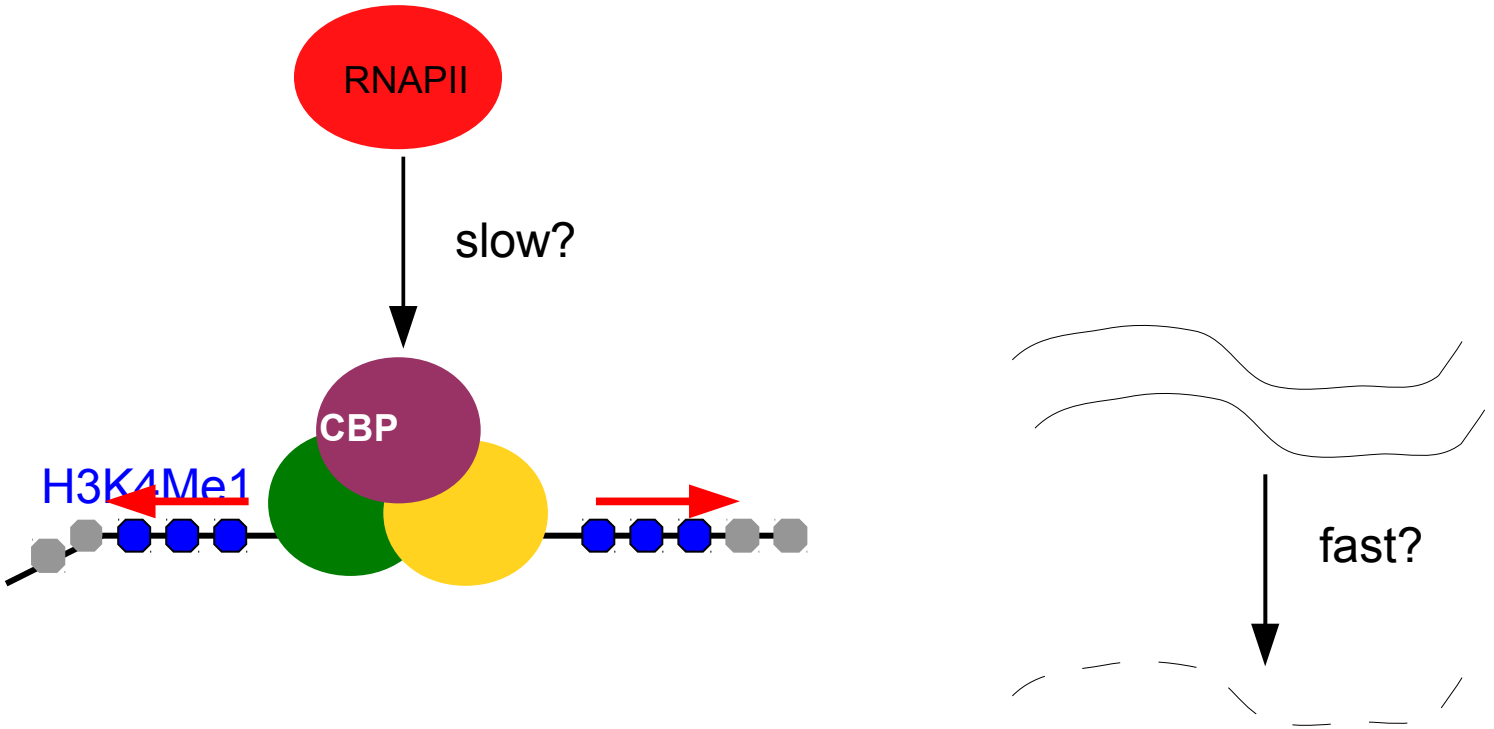




- Inducible, 2-fold
- ~1 kb
- Bidirectional
- 1 in 10k reads eRNA
- Not protein-coding

Why do eRNAs have such low abundance?

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



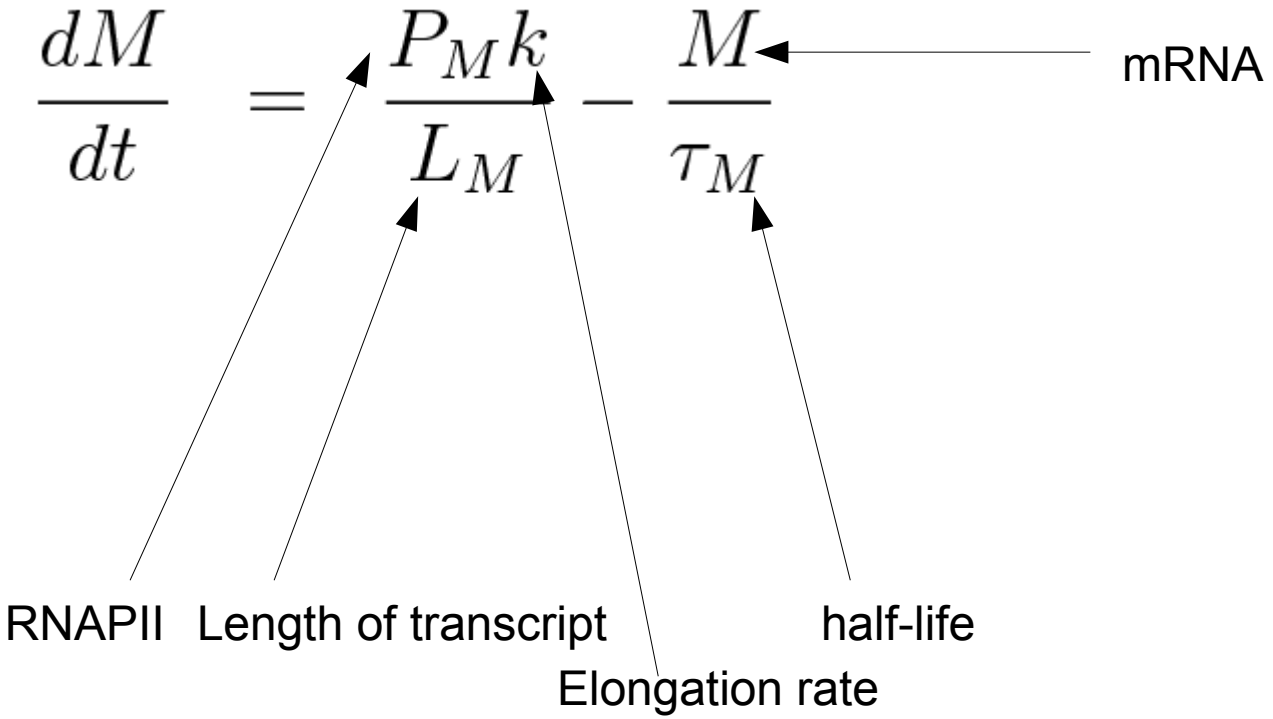
A model of mRNA production and decay

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

$$\frac{dM}{dt} = \underbrace{\frac{P_M k}{L_M}}_{\text{production}} - \underbrace{\frac{M}{\tau_M}}_{\text{decay}}$$

A model of mRNA production and decay

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



A model of eRNA production and decay

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

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

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

Half life of eRNAs relative to mRNAs

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

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

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

eRNAs half life is less than half an hour

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

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

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

$$\tau_E \sim 10^{-2} \times 1 \times 2 \times \tau_M \sim 2 \times 10^{-2} \times 600\text{min} = 12\text{min}$$

Estimate consistent with experiments

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

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

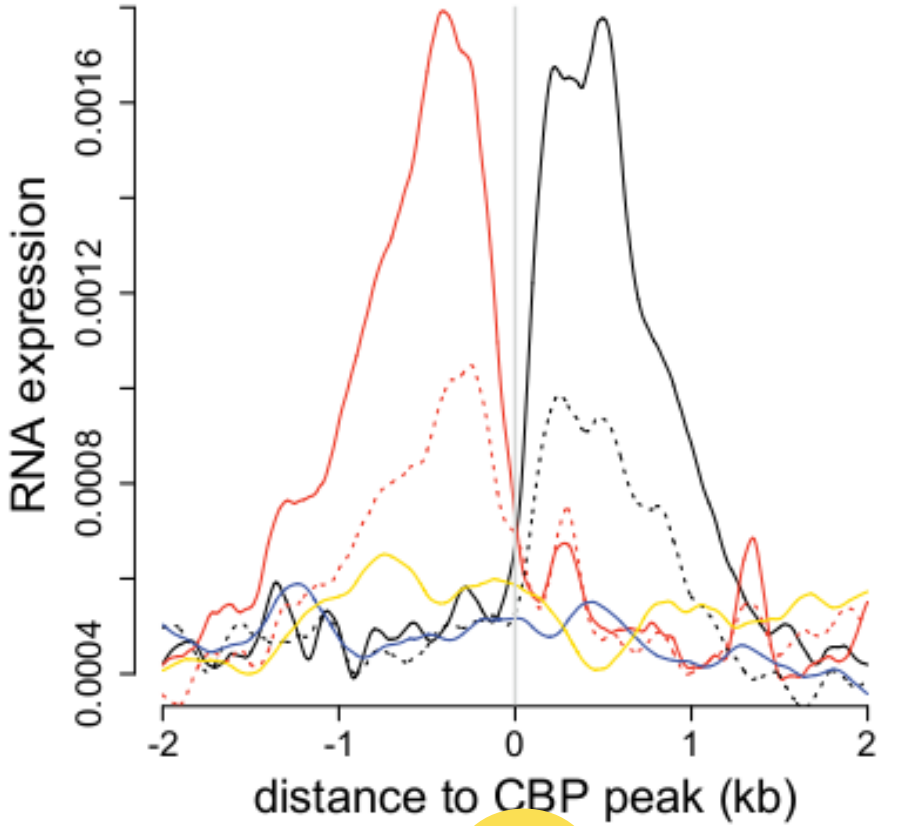
$$\frac{dE}{dt} = \frac{P_E k}{L_E} - \frac{E}{\tau_E}$$

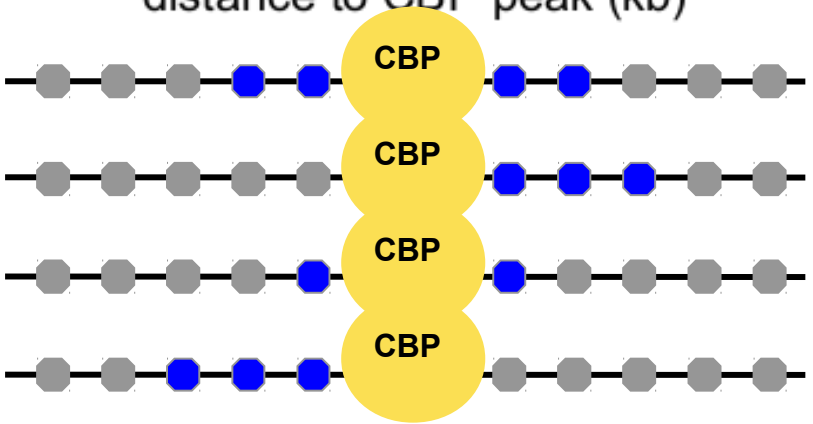
$$\frac{\tau_E}{\tau_M} = \frac{E^*}{M^*}\frac{L_E}{L_M}\frac{P_M}{P_E}$$

Finally we measured the stability of these transcripts using an actinomycinD chase. In comparison to both the mRNAs generated by the associated protein-coding genes and some known lncRNAs (like Xist and Neat), the upstream non-coding transcripts were very unstable, being reduced by 80% to 90% after a 30 min actinomycinD treatment (indicating a half-life lower than 7.5 min) (Figure 3D and Figure S3). High instability of a subset of lncRNAs both in yeast and mammals mainly depends on degradation by the nuclear exosome [39,40] and often results in the generation of more stable short RNA products [41], which in principle might be responsible for downstream functional effects.

$$\tau_E \sim 10^{-2} \times 1 \times 2 \times \tau_M \sim 2 \times 10^{-2} \times 600\text{min} = 12\text{min}$$

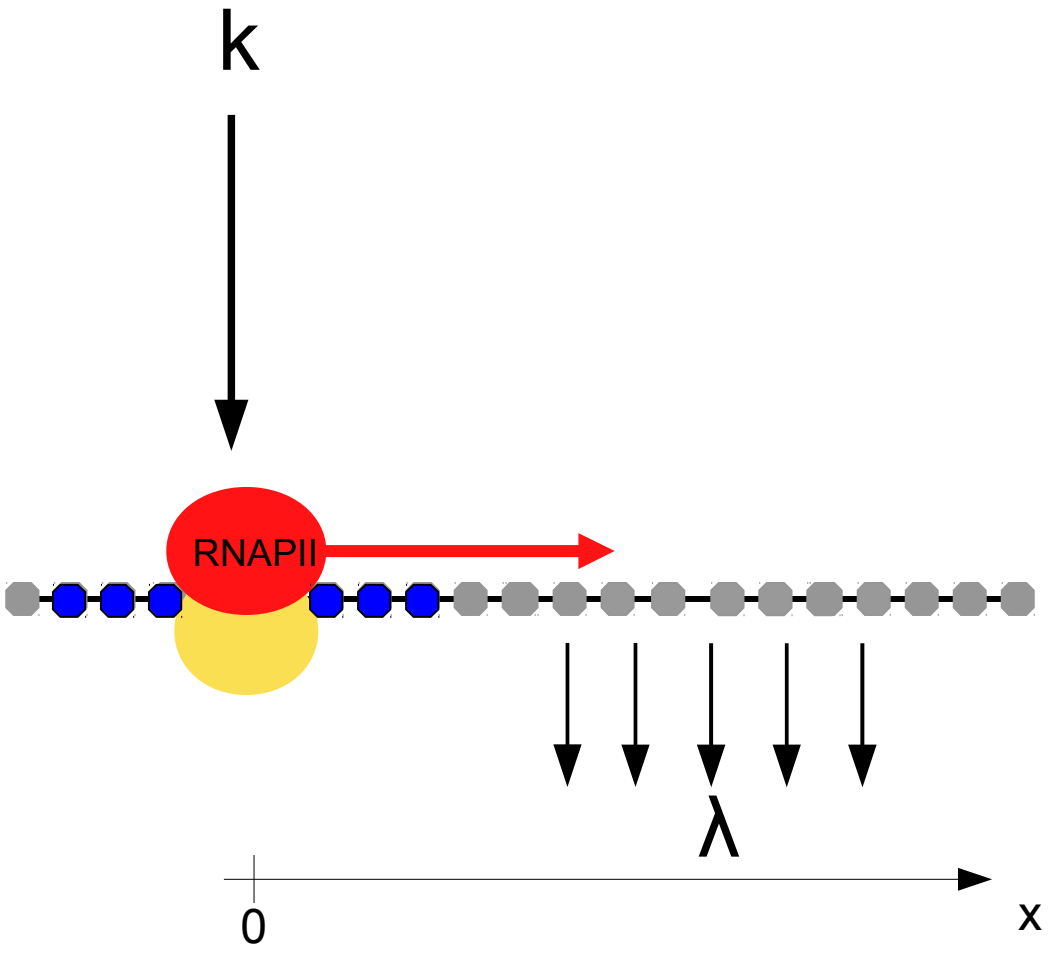
eRNA levels as a function of distance from center of enhancer





RNAPII binds and falls of at a constant rate

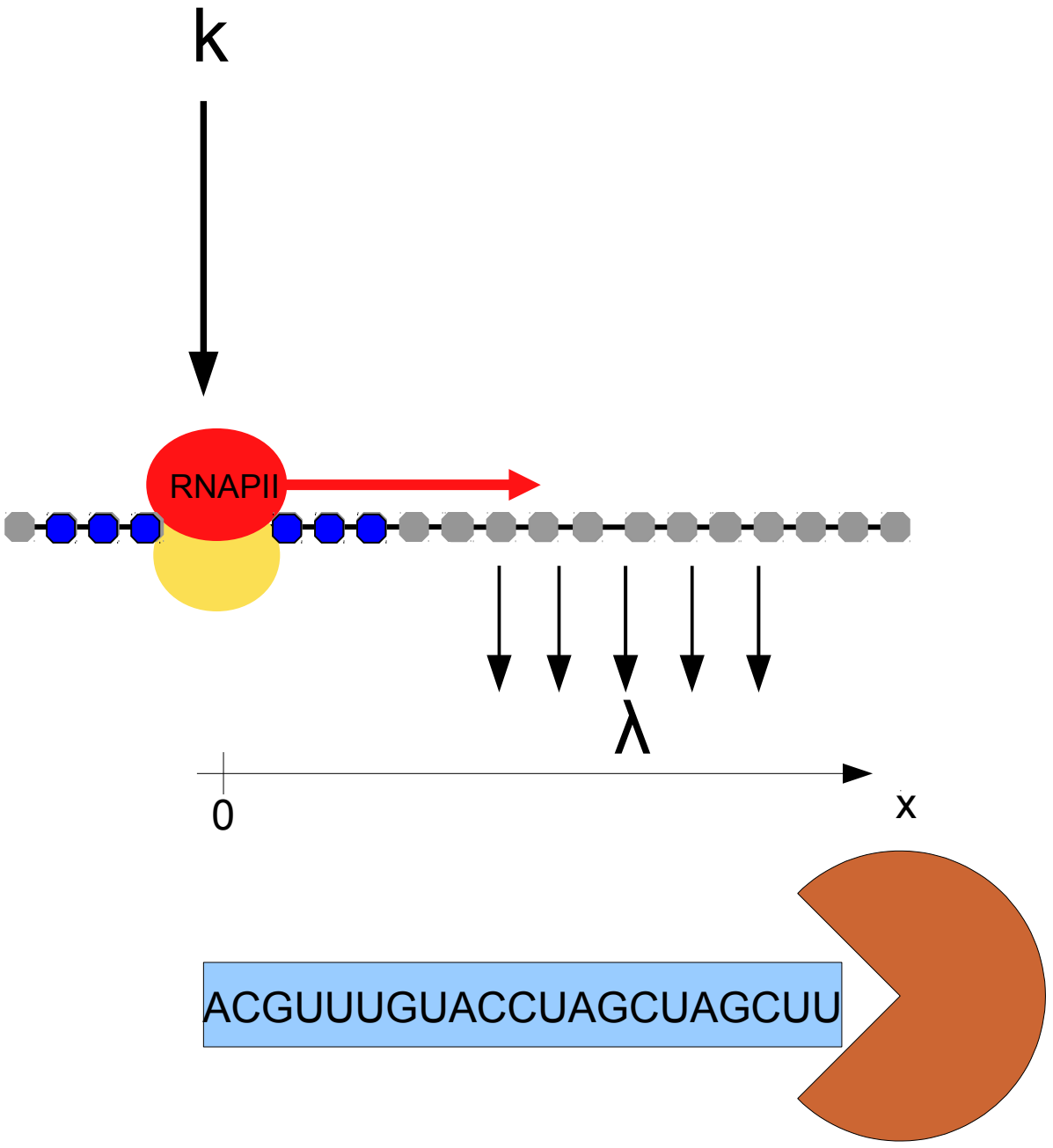
$$\frac{dP}{dx} = k - \lambda P$$



eRNA production proportional to RNAPII

$$\frac{dP}{dx} = k - \lambda P$$

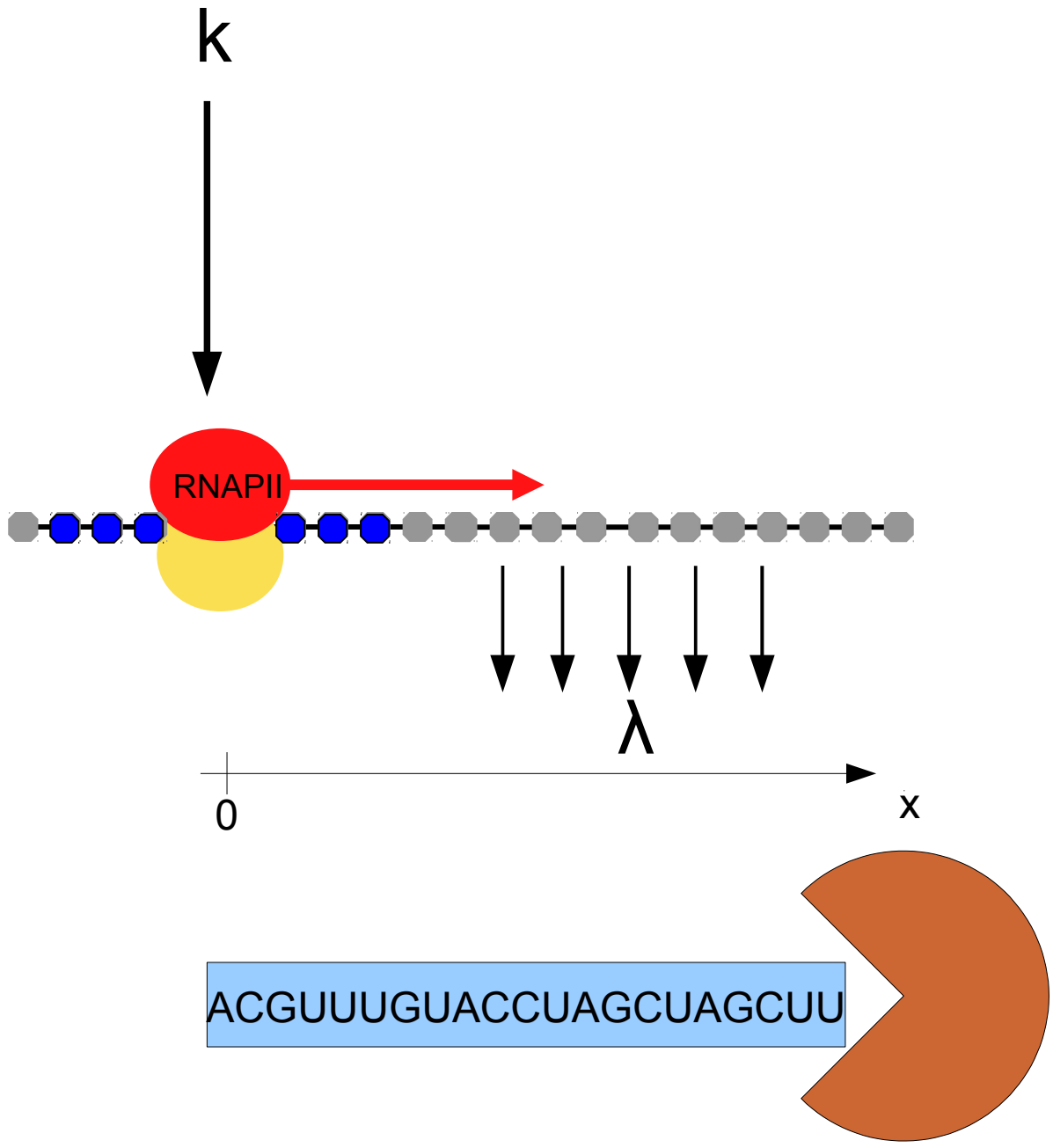
$$\frac{dE}{dx} = \gamma P(x) - \delta x E$$



Parameters can be estimated from literature

$$\frac{dP}{dx} = k - \lambda P$$

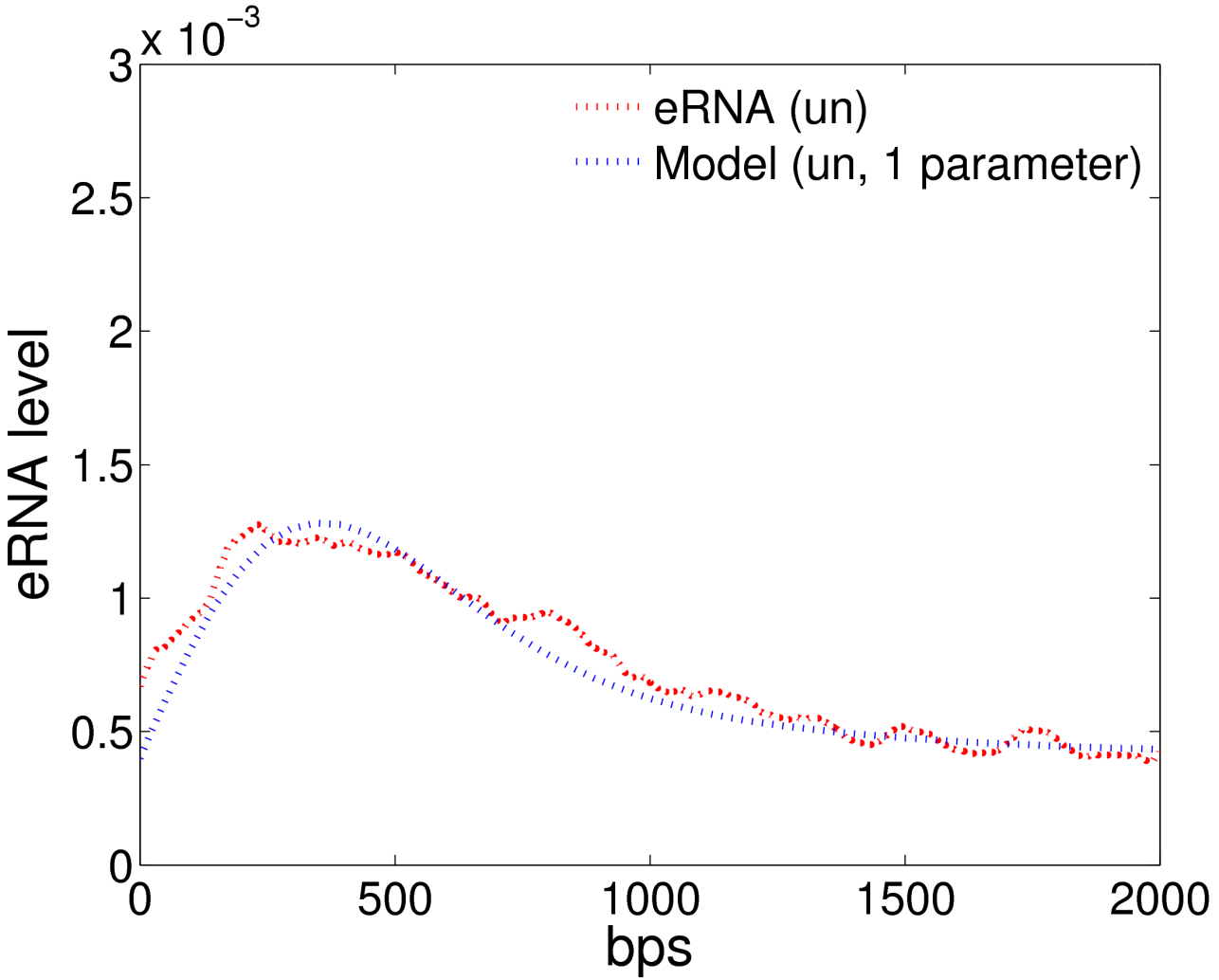
$$\frac{dE}{dx} = \gamma P(x) - \delta x E$$



eRNA levels can be accurately predicted

$$\frac{dP}{dx} = k - \lambda P$$

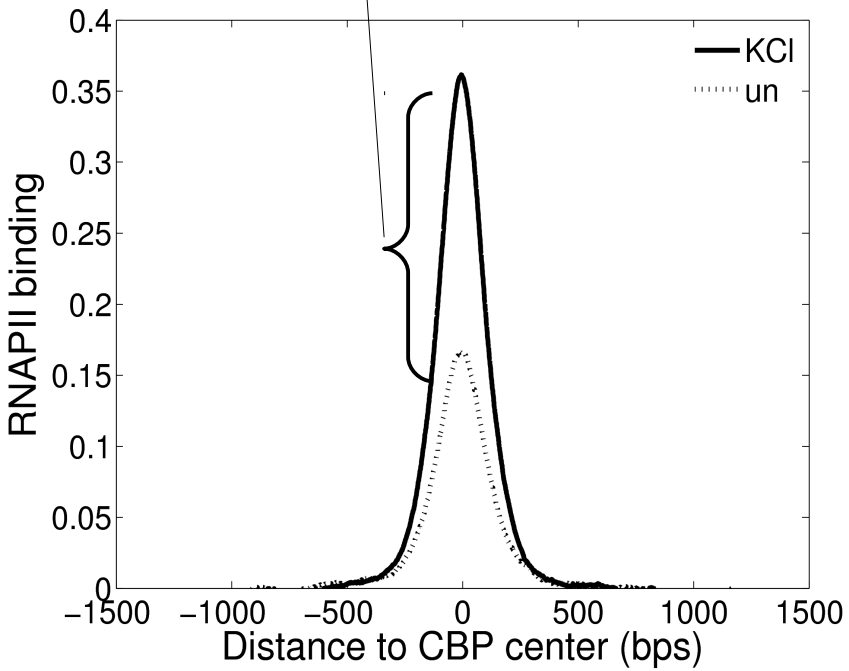
$$\frac{dE}{dx} = \gamma P(x) - \delta x E$$

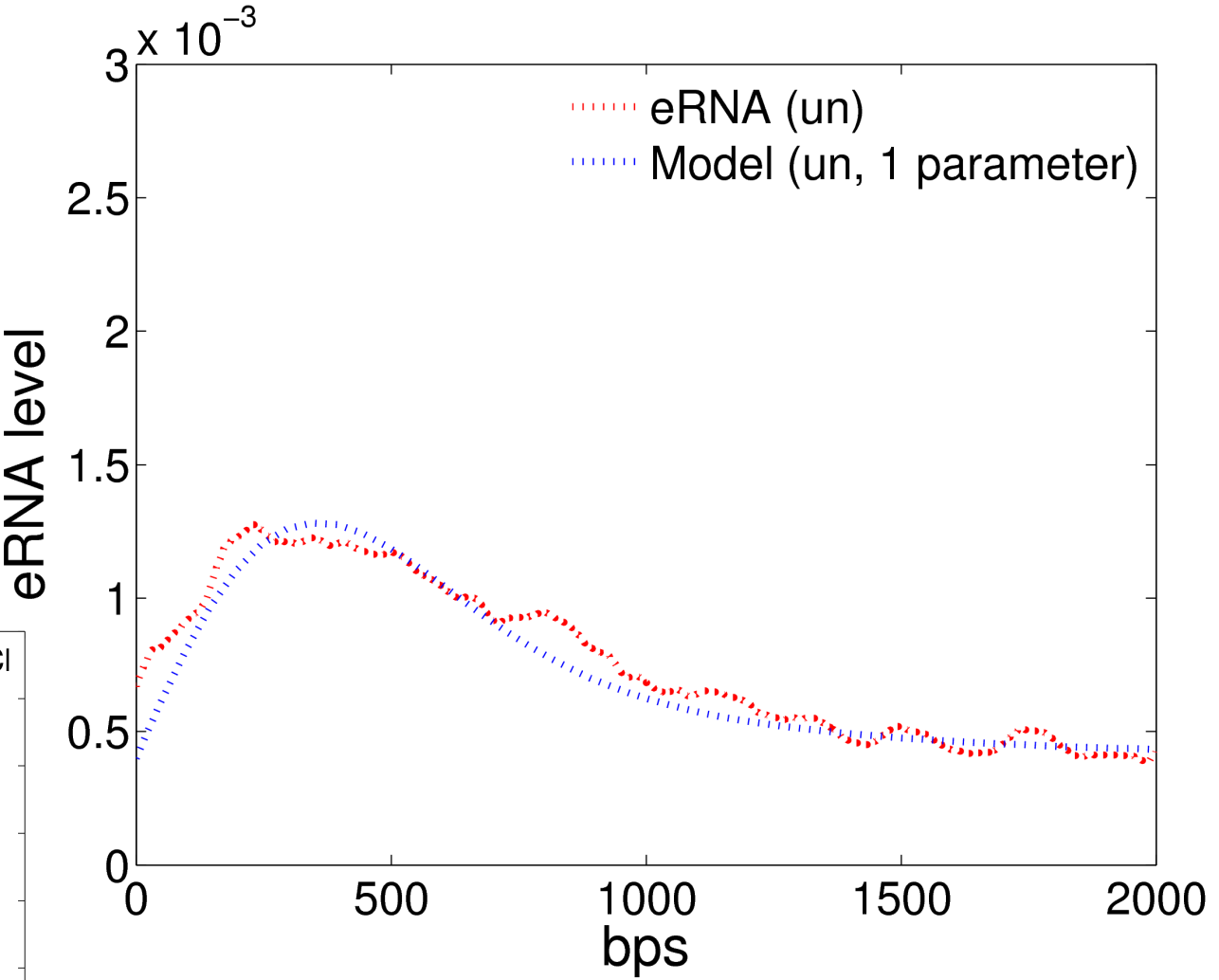


Binding rate of RNAPII doubled after KCl

$$\frac{dP}{dx} = k - \lambda P$$

$$\frac{dE}{dx} = \gamma P(x) - \delta x E$$

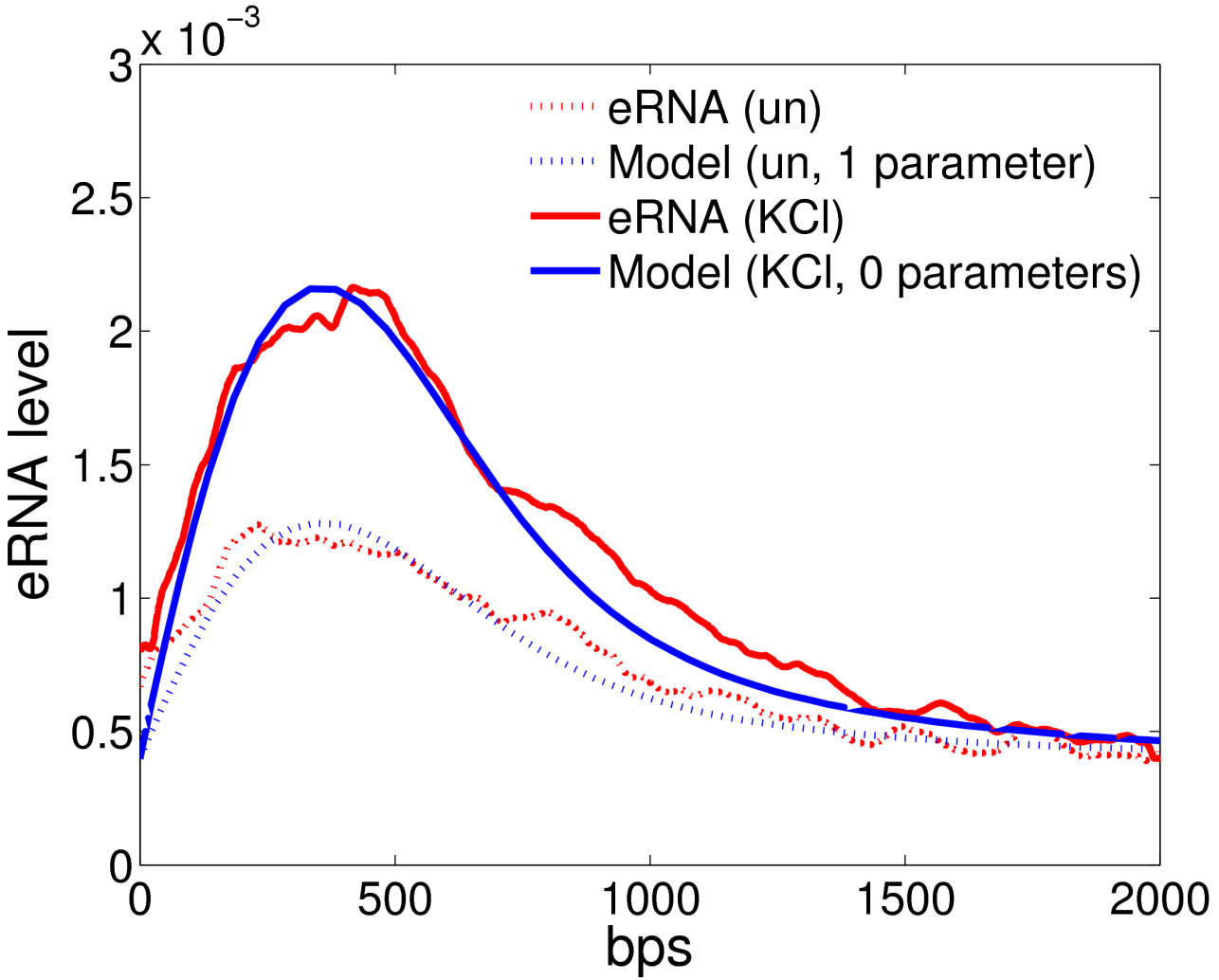




No free parameters for eRNA after KCl

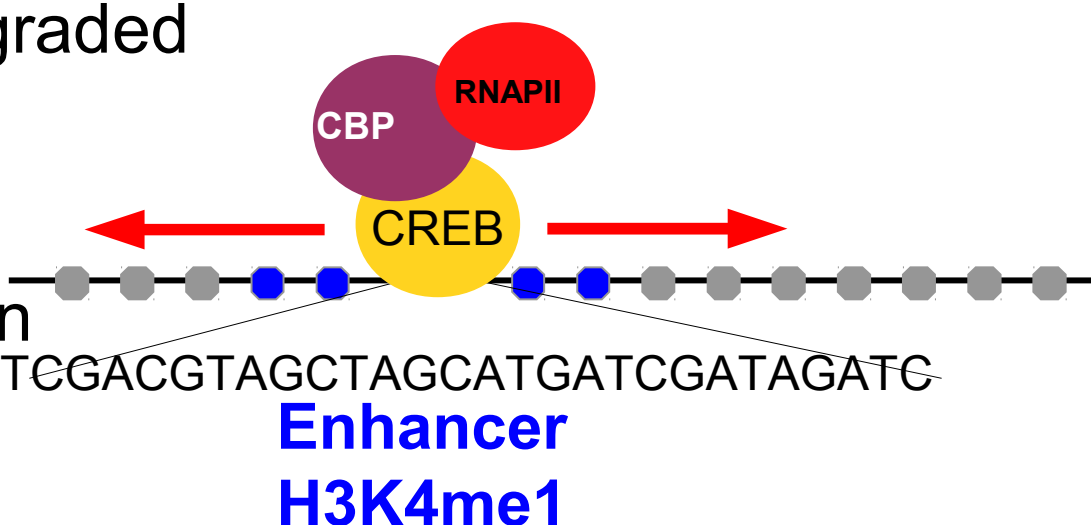
$$\frac{dP}{dx} = k - \lambda P$$

$$\frac{dE}{dx} = \gamma P(x) - \delta x E$$



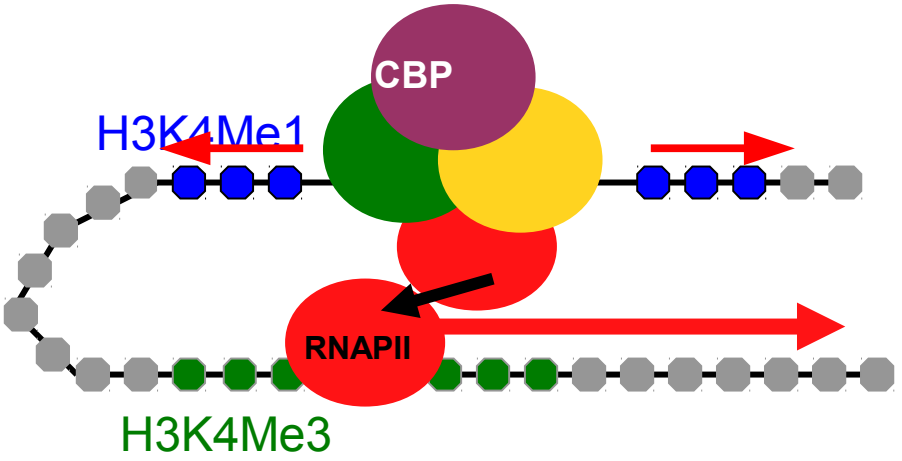
Properties of activity-dependent enhancers

- Enriched for ~100 sequence motifs
- ChIP-seq reads predicted by sequence
- CBP binding determined by other TFs
- CBP recruits RNAPII
- RNAPII synthesizes eRNAs
  - eRNAs are rapidly degraded
  - eRNA levels described by model of transcription



What is the function of RNAPII at enhancers?

- Transcribe eRNAs
- Speed up recruitment of RNAPII at promoter



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

Recruitment of RNAPII at the promoter

- Transcribe eRNAs
- Speed up recruitment of RNAPII at promoter

$$\frac{dP_M}{dt} = \underbrace{k_p + Nk_e c}_{\text{Binding rate}} - \underbrace{\frac{P_M}{\tau}}_{\text{decay}}$$

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

Difficult to estimate parameters

- Transcribe eRNAs
- Speed up recruitment of RNAPII at promoter

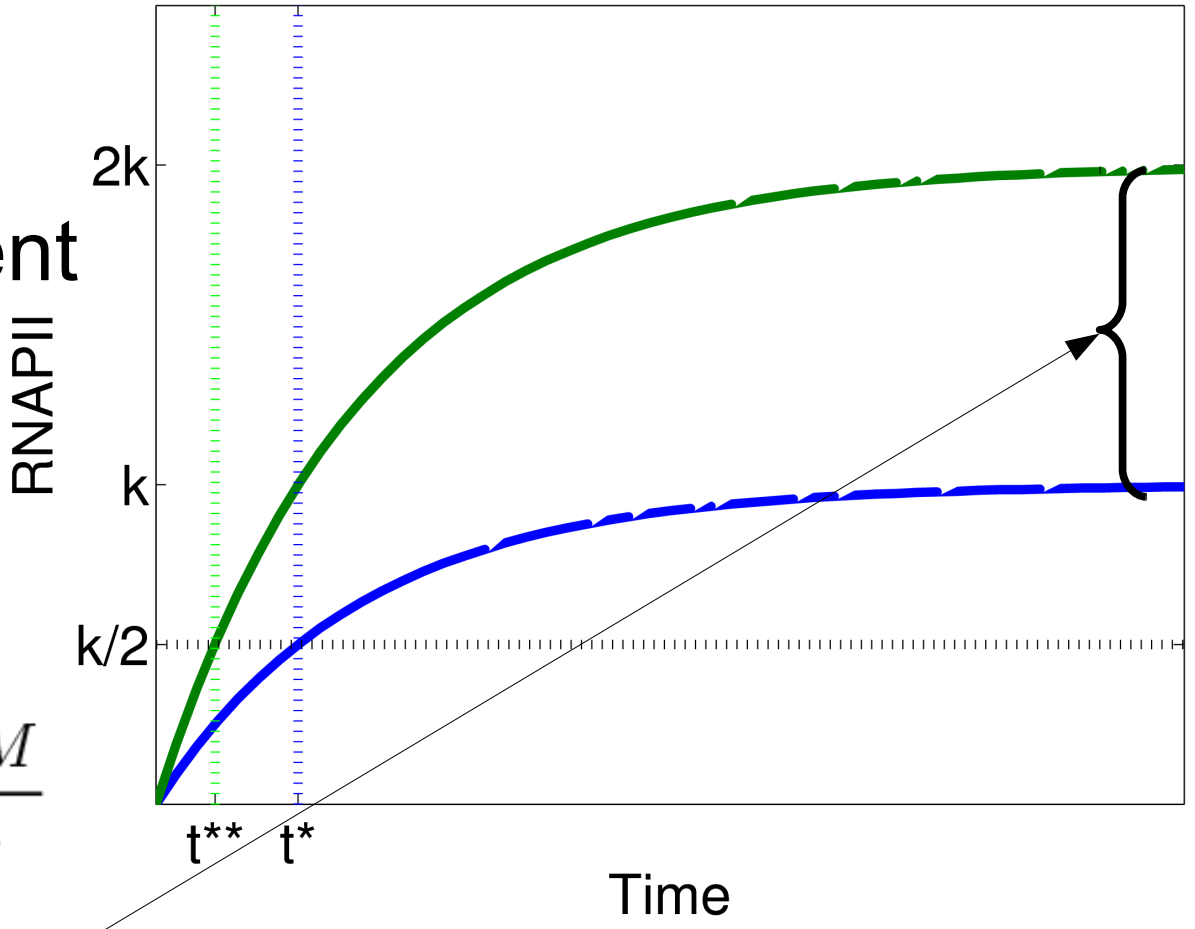
$$\frac{dP_M}{dt} = k_p + Nk_ec - \frac{P_M}{\tau}$$

$$P_M(t) = \underbrace{(k_p + Nk_ec)}_{\text{Steady state level}}(1 - e^{-t/\tau})$$

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

Steady state level of RNAPII is increased

- Transcribe eRNAs
- Speed up recruitment of RNAPII at promoter

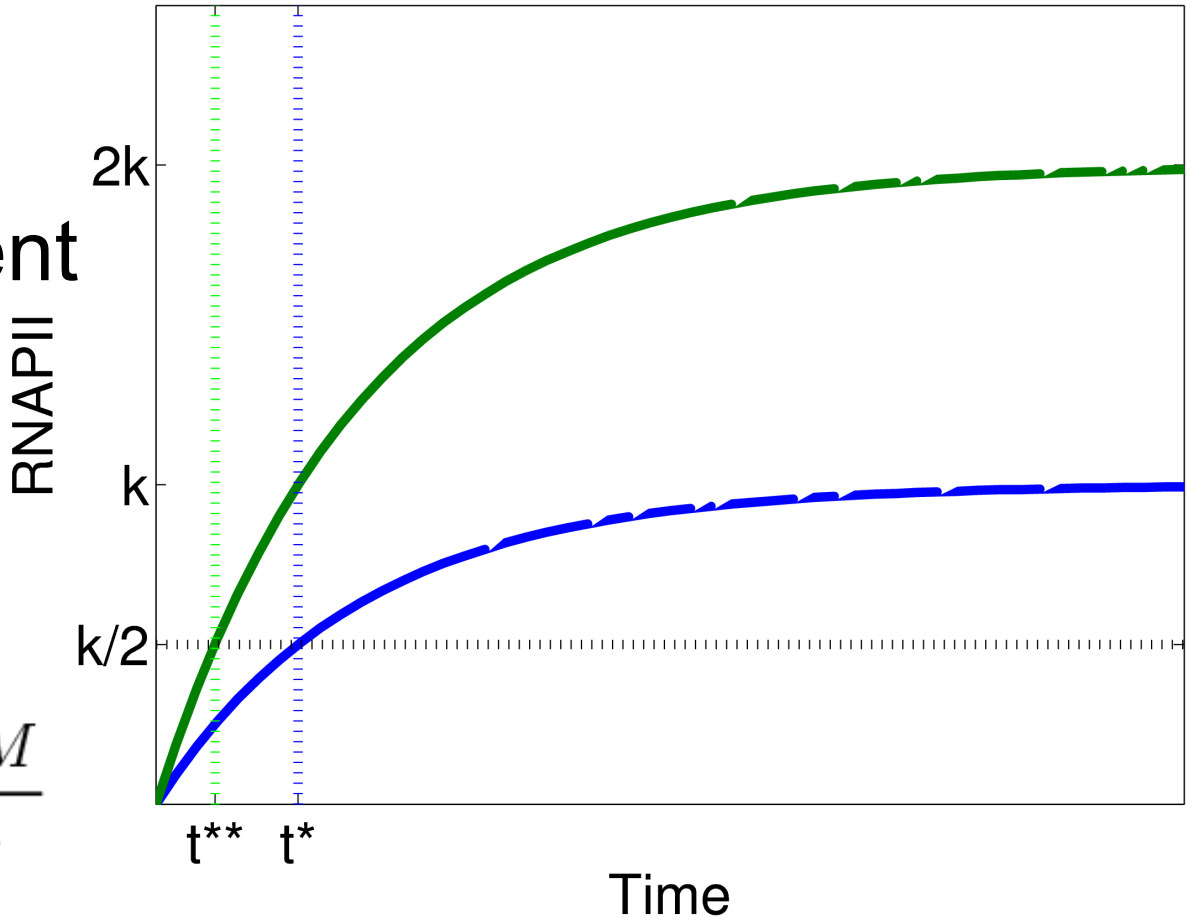


$$\frac{dP_M}{dt} = k_p + Nk_ec - \frac{P_M}{\tau}$$

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

Rise time is reduced

- Transcribe eRNAs
- Speed up recruitment of RNAPII at promoter

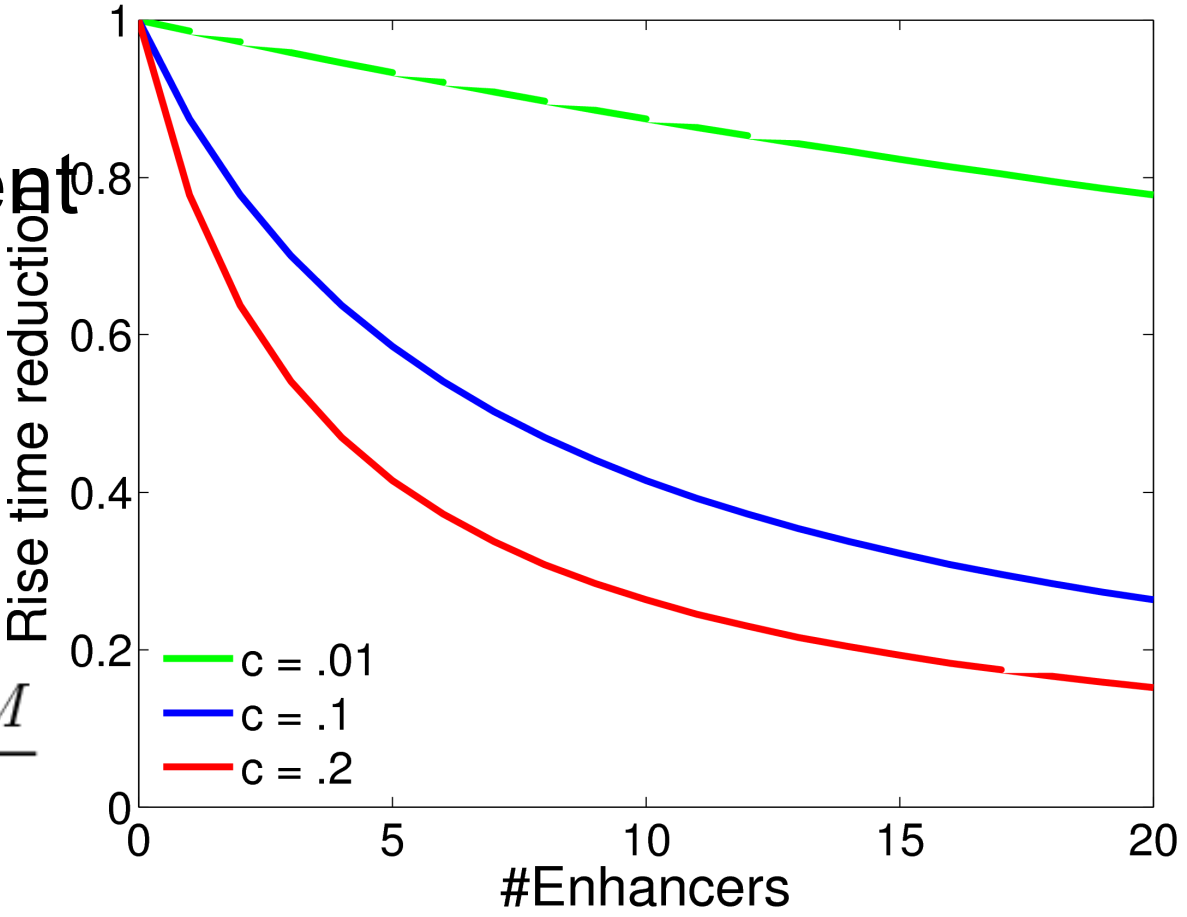


$$\frac{dP_M}{dt} = k_p + Nk_e c - \frac{P_M}{\tau}$$

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

Significant speed-up with ~5 enhancers

- Transcribe eRNAs
- Speed up recruitment of RNAPII at promoter

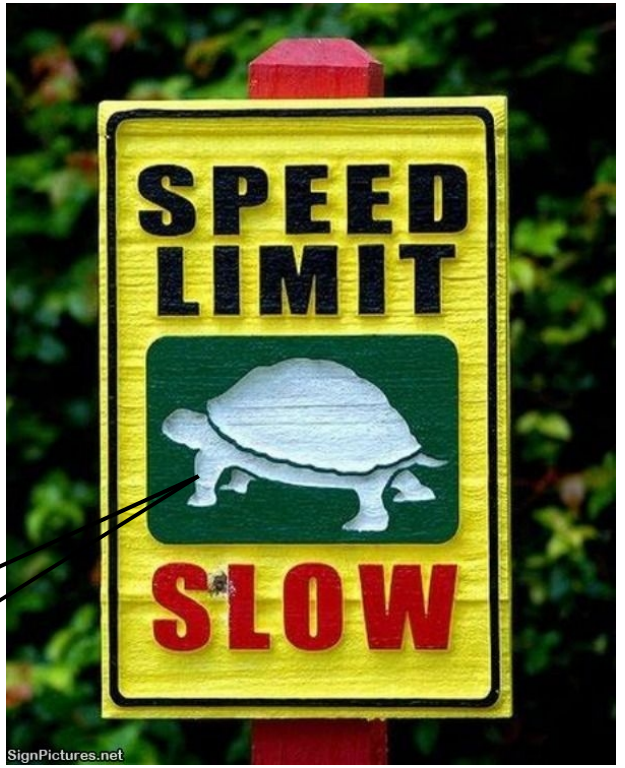


$$\frac{dP_M}{dt} = k_p + Nk_e c - \frac{P_M}{\tau}$$

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

Recruitment of RNAPII is diffusion limited

- Transcribe eRNAs
- Speed up recruitment of RNAPII at promoter



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

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

Enhancers may reduce the noise in RNAPII

- Transcribe eRNAs
- Speed up recruitment of RNAPII at promoter

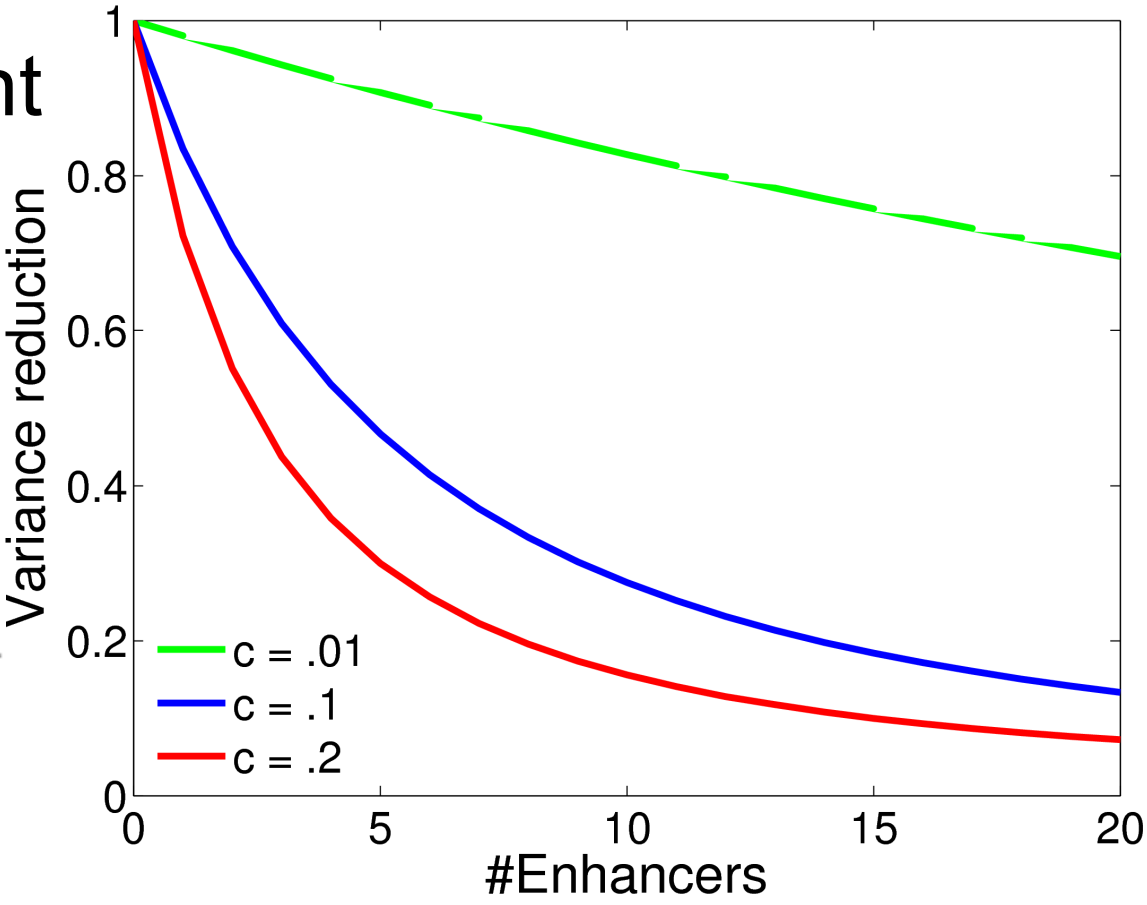
$$\frac{dP_M}{dt} = k_p + N k_e c - \frac{P_M}{\tau}$$

$\Gamma(\alpha, \beta)$

RNAPII noise reduction proportional to number of enhancers

- Transcribe eRNAs
- Speed up recruitment of RNAPII at promoter

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

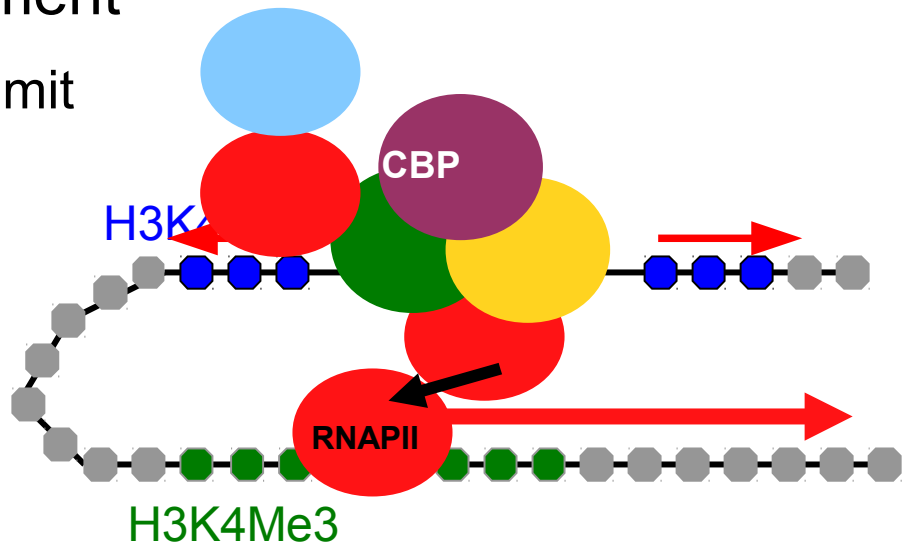


What is the function of eRNAs?

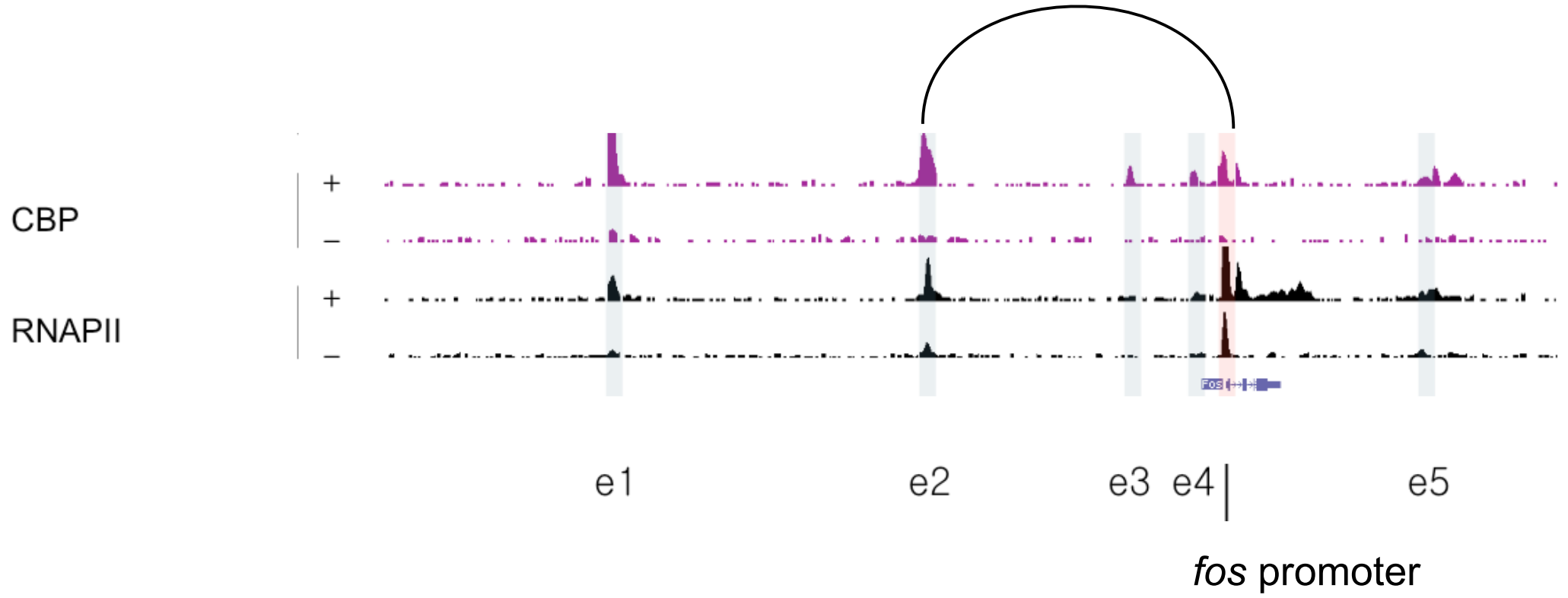
- What is the function of RNAPII at enhancers?
  - Increase rate of RNAPII recruitment
    - Possibly faster than diffusion limit
  - Faster rise-time
  - Reduced noise
- What is the function of eRNAs?
  - Noise
  - Transcription establishes histone modifications
  - Transcript has function

Enzymes piggyback on the polymerase

- What is the function of RNAPII at enhancers?
  - Increase rate of RNAPII recruitment
    - Possibly faster than diffusion limit
  - Faster rise-time
  - Reduced noise
- What is the function of eRNAs?
  - Noise
  - Transcription establishes histone modifications
  - Transcript has function

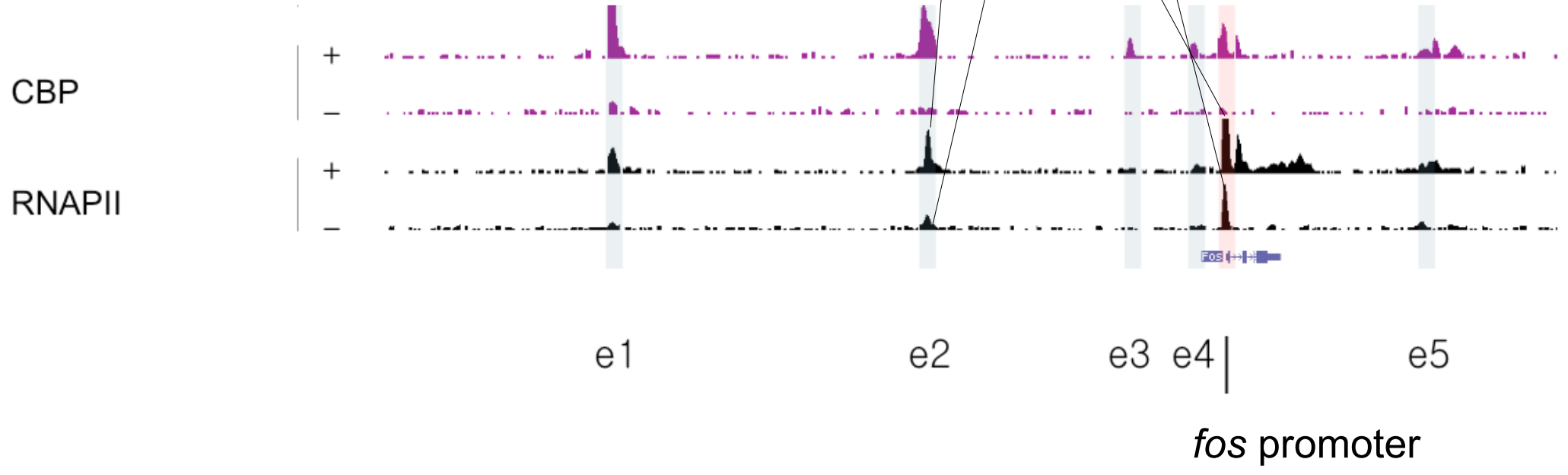


Pair each enhancer with nearest promoter and compare RNAPII and RNA



Calculate induction index for both RNAPII and transcription

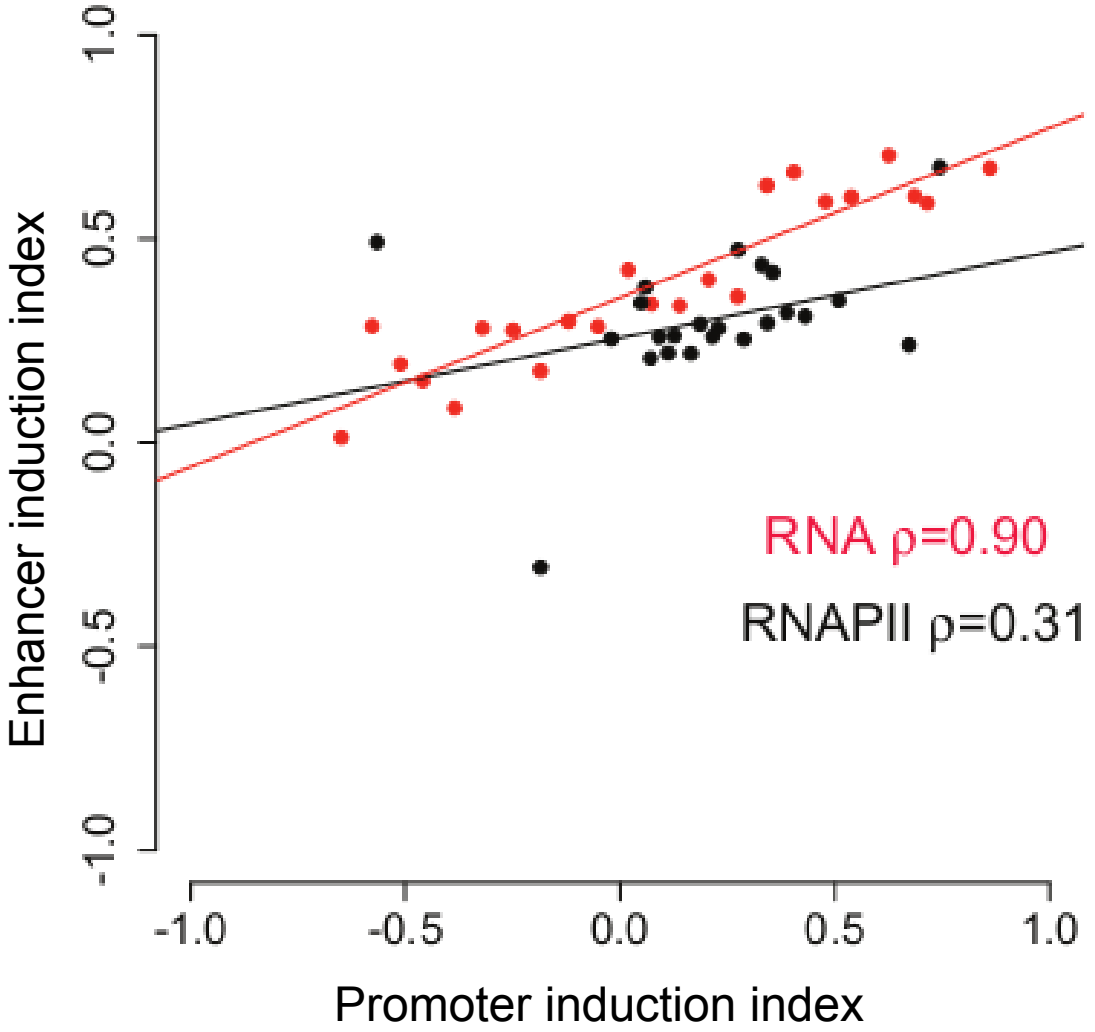
$$induction\ index = \frac{KCl - un}{KCl + un}$$



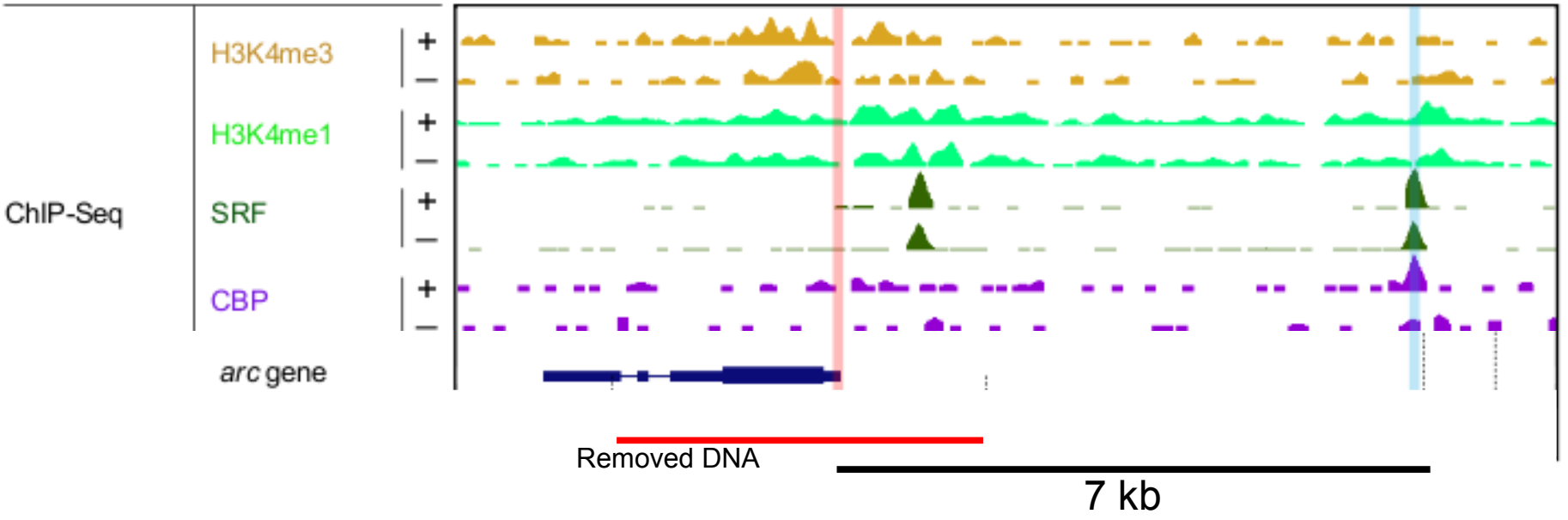
fos promoter

eRNA induction is correlated with induction of nearby mRNAs but not RNAPII

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



Deletion of the Arc-promoter confirms that RNAPII recruitment is independent but eRNA transcription is not.



- RNAPII same in knock-out +/- KCl
- eRNAs not present in knock-out

Summary

- Identified ~12k activity-dependent enhancers
- Discovered and quantified novel mechanisms
  - Identified enriched motifs and bound TFs
  - Combinatorial code for CBP affinity
  - Recruitment of RNAPII at enhancers
    - Faster recruitment to promoter
    - Reduce noise
  - Transcription at enhancers
    - Properties of eRNA
    - Model of RNAPII and eRNA levels
    - Interaction with promoter necessary

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

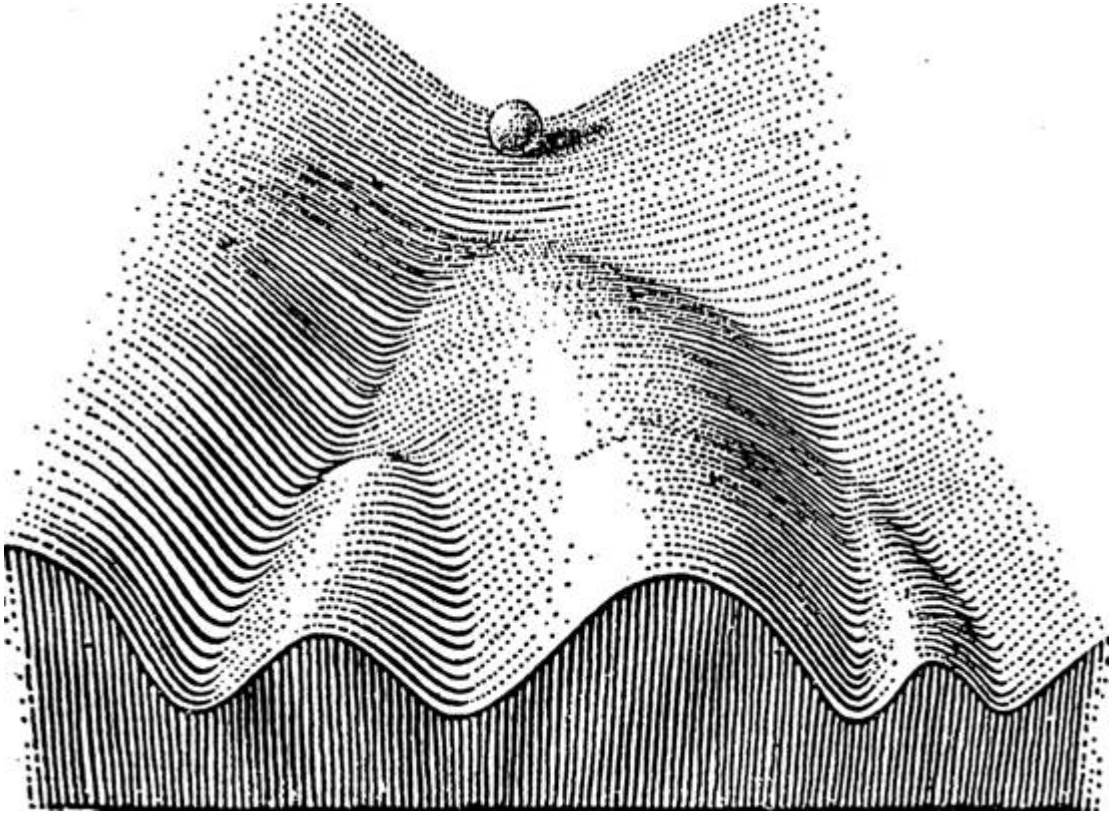
doi:10.1038/nature09692

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]

Stochastic models of gene expression

- Transitions between stable states



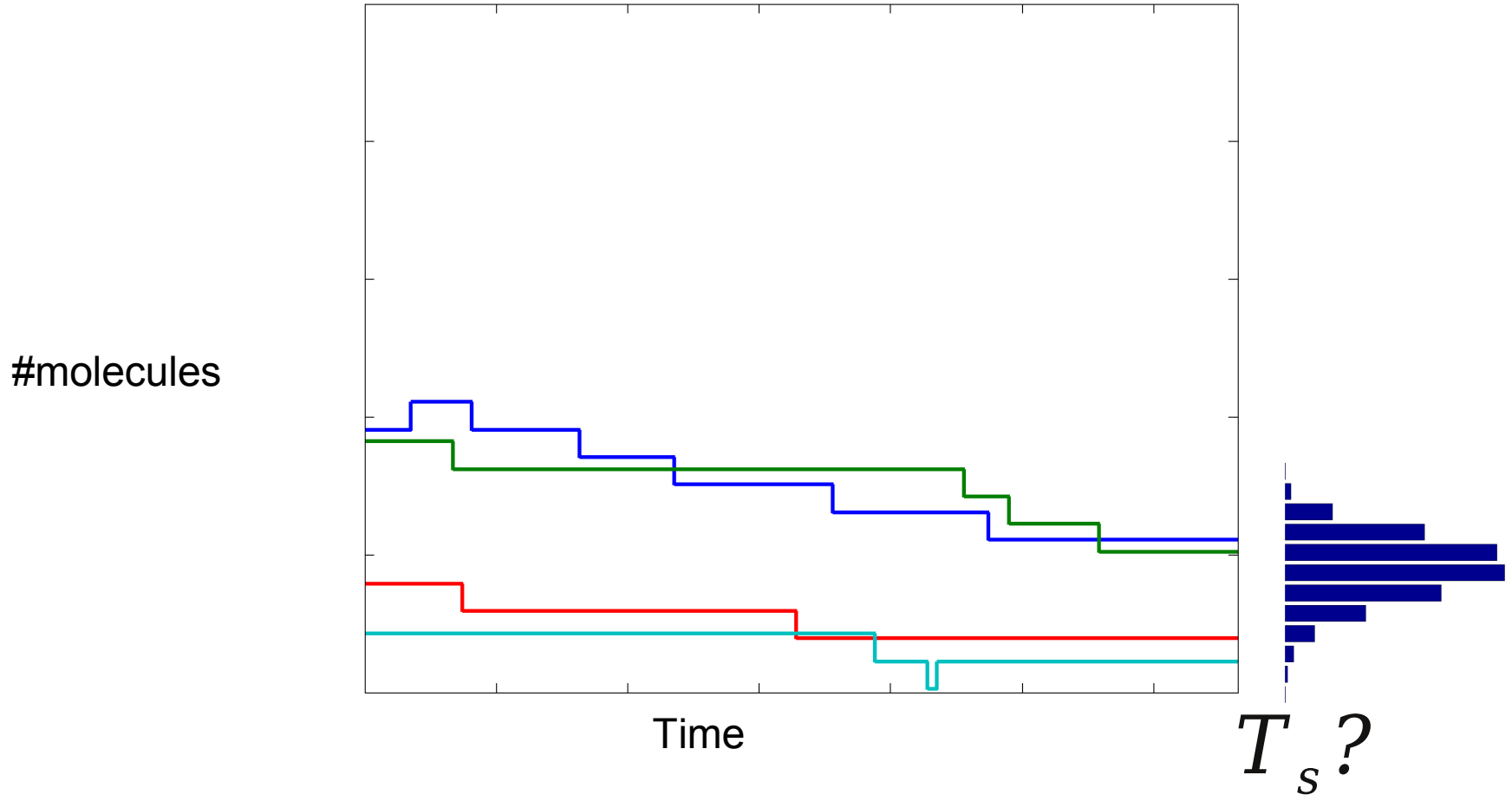
Waddington, 1953

Master Equation (**ME**) description

- Discreteness required since ~10 mRNAs/cell
- Use Markov Chain Monte Carlo (MCMC)
  - Gillespie's Stochastic Simulation Algorithm (**SSA**)

How long do we need to run MCMC?

- SSA simulates trajectories of system
  - Run repeatedly to estimate probability distribution



How long do we need to run MCMC?

- SSA simulates trajectories of system
  - Run repeatedly to estimate probability distribution
- Dominated Coupling From The Past SSA **proven** to reach stationary distribution

BMC Systems Biology

BioMed Central

Methodology article

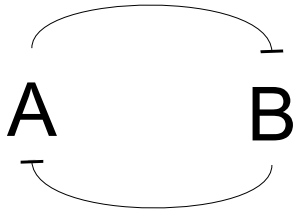
Open Access

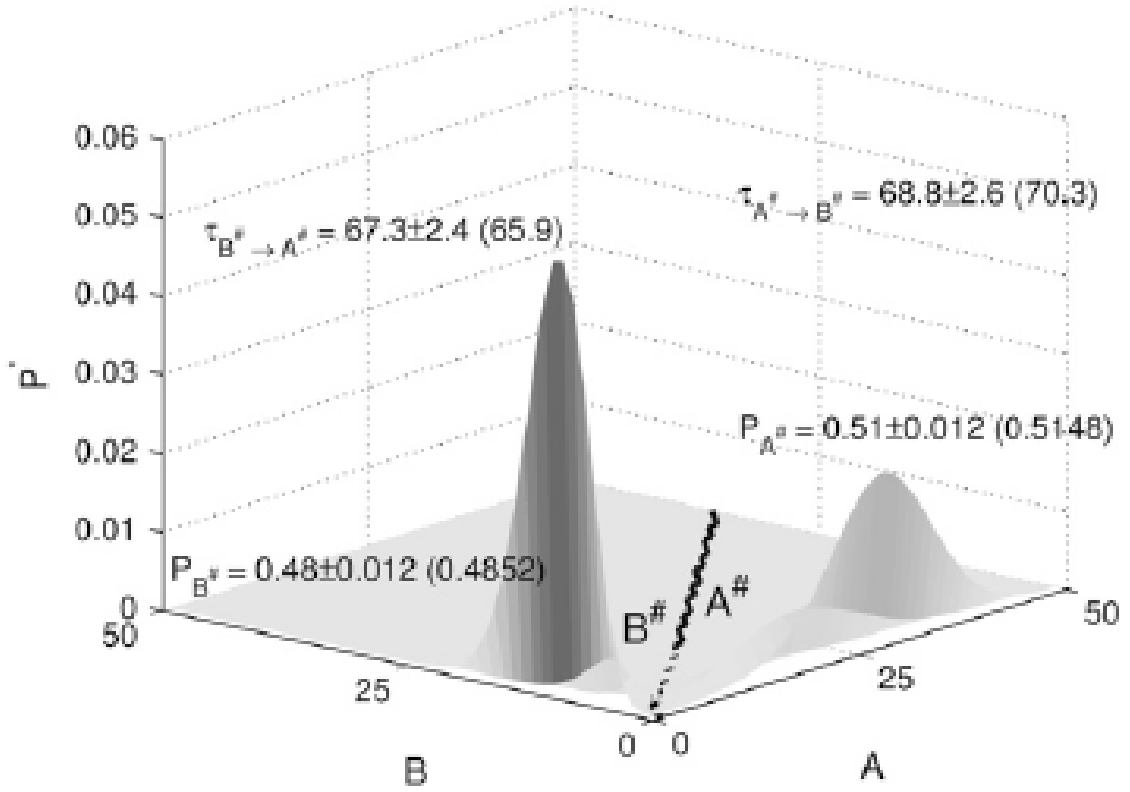
A Dominated Coupling From The Past algorithm for the stochastic simulation of networks of biochemical reactions

Martin Hemberg[1] and Mauricio Barahona*[1,2]

Address: [1]Department of Bioengineering, Imperial College London, South Kensington Campus, London SW7 2AZ, UK and [2]Institute for Mathematical Sciences, Imperial College London, South Kensington Campus, London SW7 2AZ, UK

Perfect sampling of transitions between steady states





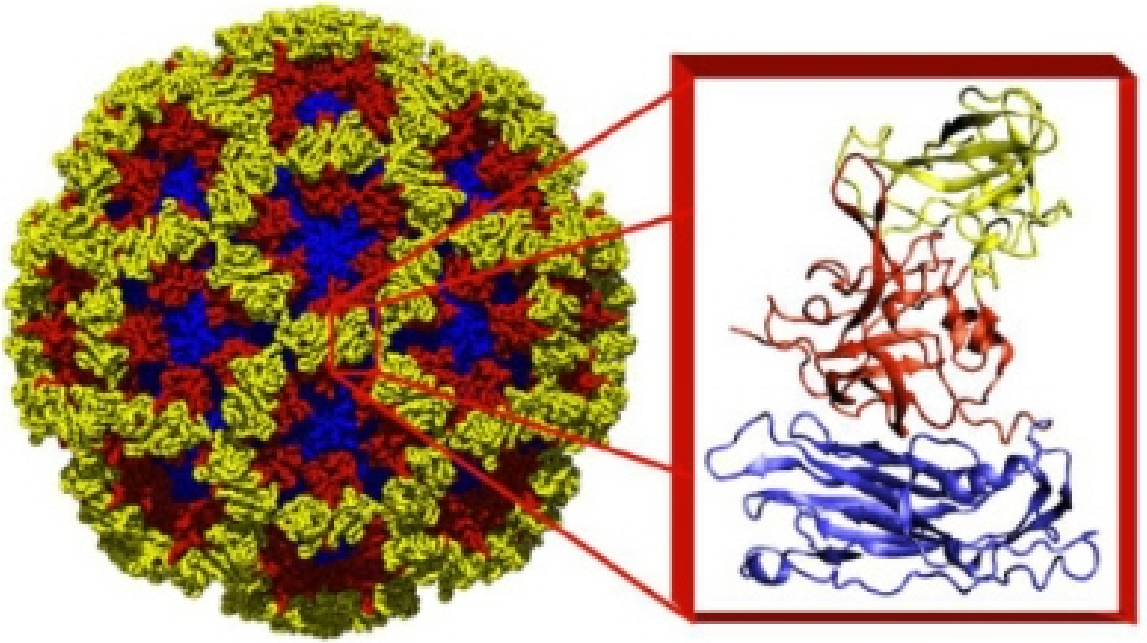
Biophysical Journal Volume 93 July 2007 401–410

Perfect Sampling of the Master Equation for Gene Regulatory Networks

Martin Hemberg and Mauricio Barahona
Department of Bioengineering and Institute for Mathematical Sciences, Imperial College London, London, United Kingdom

Assembly of viral capsids

- Protect viral genome
  - Self-assembly
  - Identical subunits
  - Icosahedral symmetry



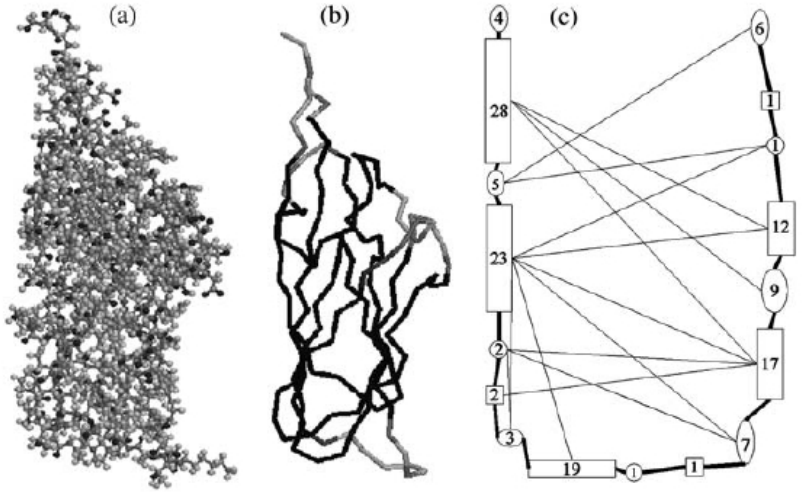
Biophysical Journal Volume 90 May 2006 3029–304:

Stochastic Kinetics of Viral Capsid Assembly Based on Detailed Protein Structures

Martin Hemberg,* Sophia N. Yaliraki,† and Mauricio Barahona*
*Department of Bioengineering and †Department of Chemistry, Imperial College London, London, United Kingdom

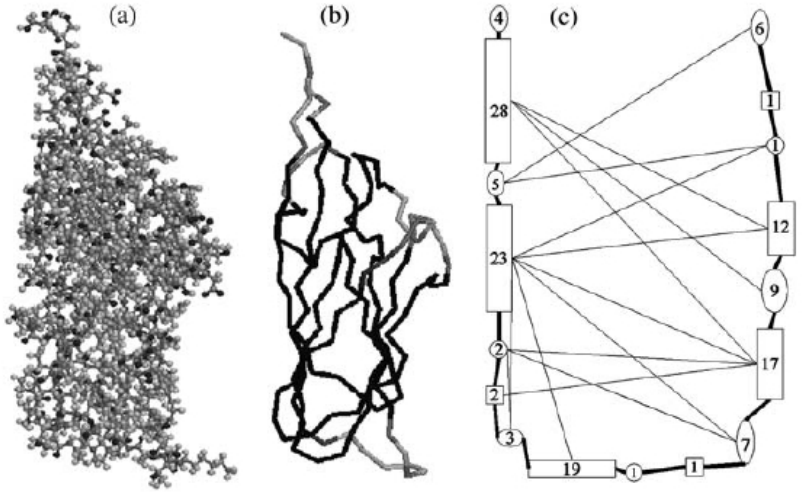
Coarse-grained protein model

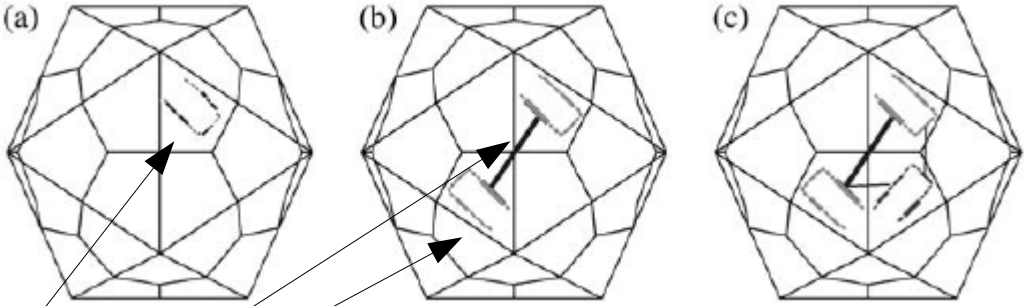
- Atomic-structure
- FIRST calculates rigidity of amino acids
- Identify ~20 rigid blocks



Use reduced representation for aggregates

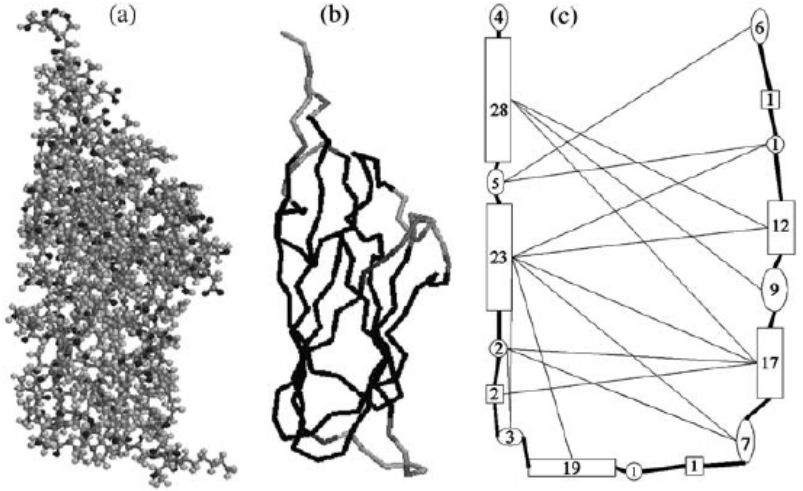
- Atomic-structure, identify rigid blocks
- Oligomer association and dissociation rates
  - Association restricted by diffusion

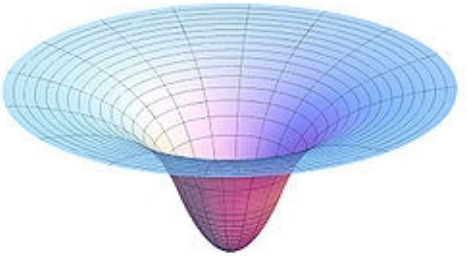


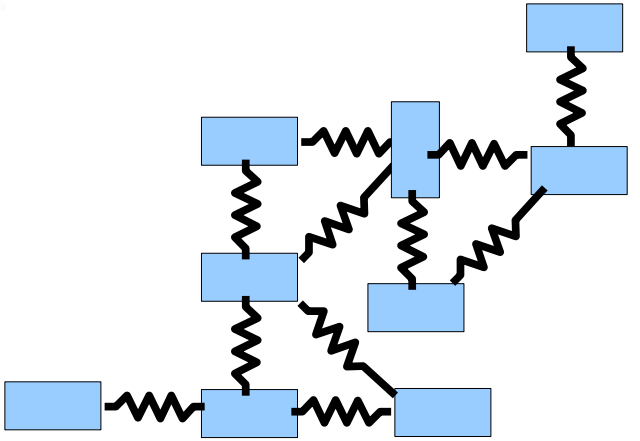


Aggregates modeled as mass-spring graph

- Atomic-structure, identify rigid blocks
- Oligomer association and dissociation rates
  - Association restricted by diffusion
  - Dissociation escape from multi-dimensional well

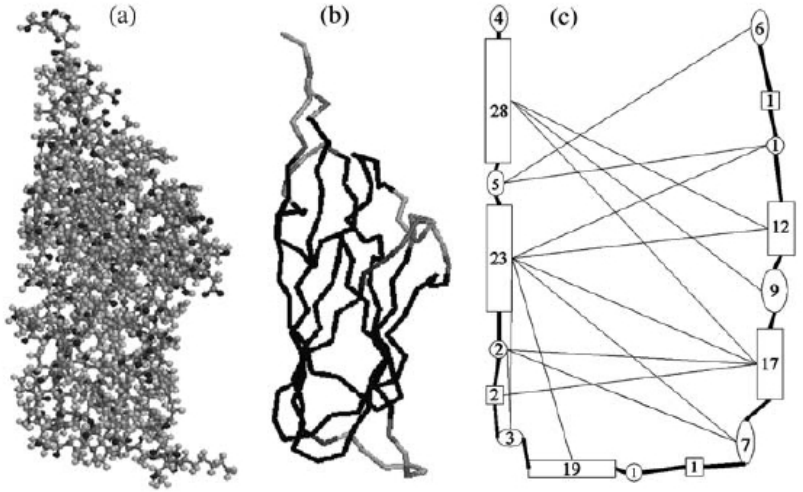


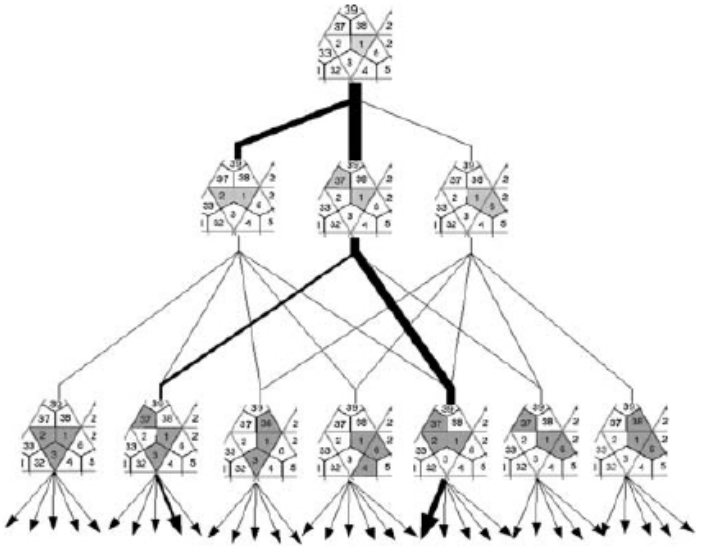




All reactions cannot be enumerated

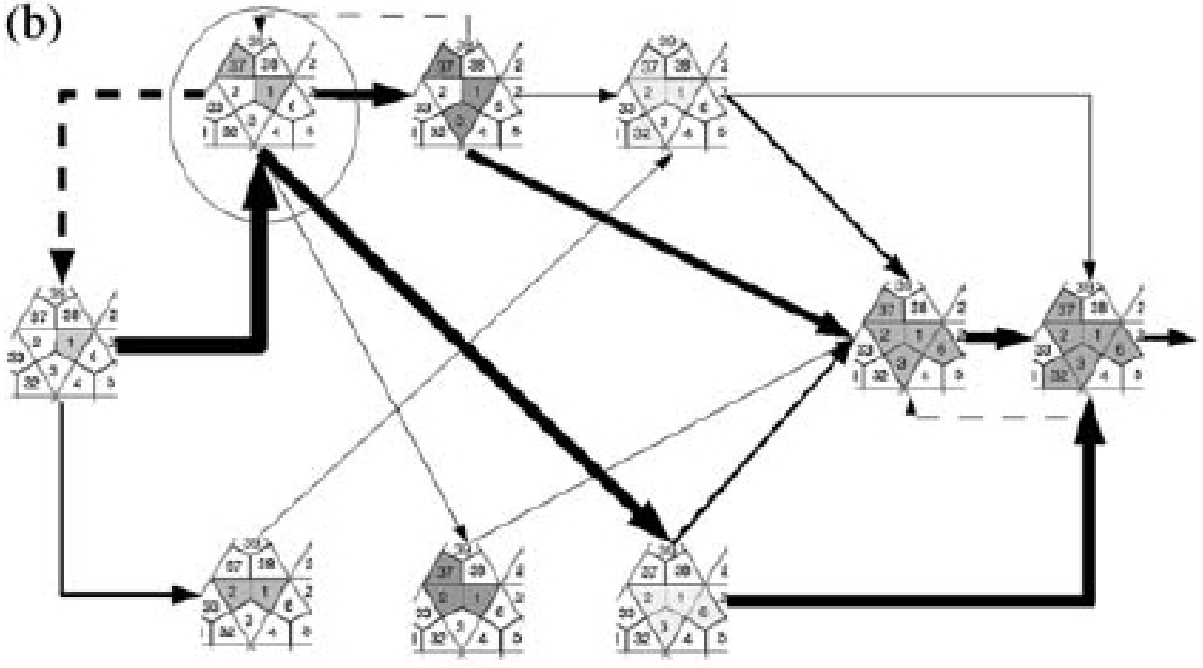
- Atomic-structure, identify rigid blocks
- Oligomer association and dissociation rates
  - Association restricted by diffusion
  - Dissociation escape from multi-dimensional well





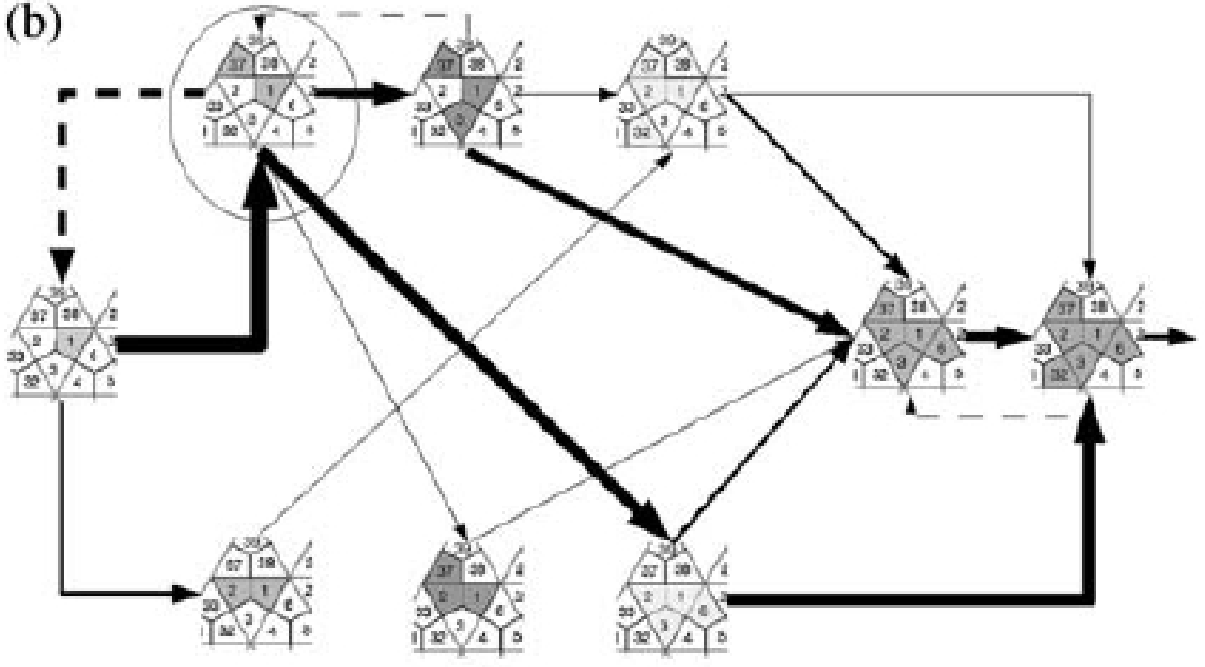
Probabilistic sampling of assembly paths

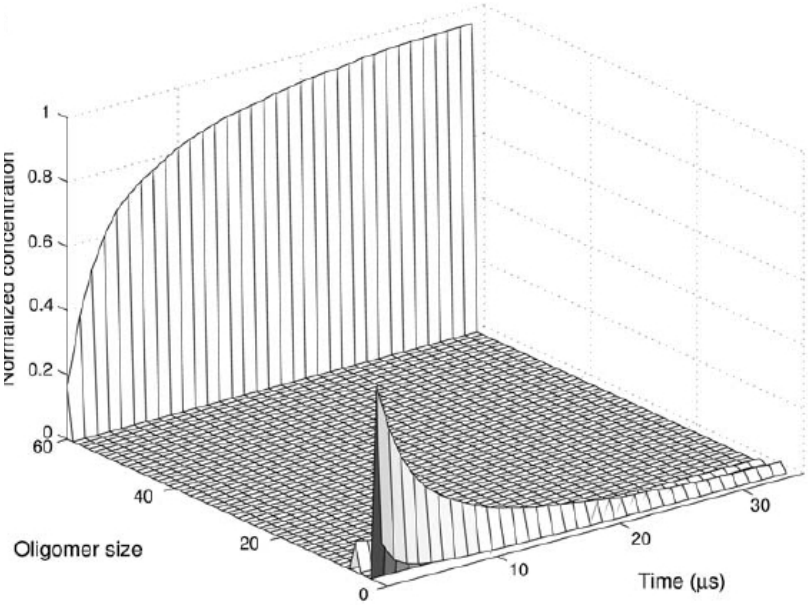
- Atomic-structure, identify rigid blocks
- Oligomer association and dissociation rates
  - Association restricted by diffusion
  - Dissociation escape from multi-dimensional well



Identify stable intermediaries

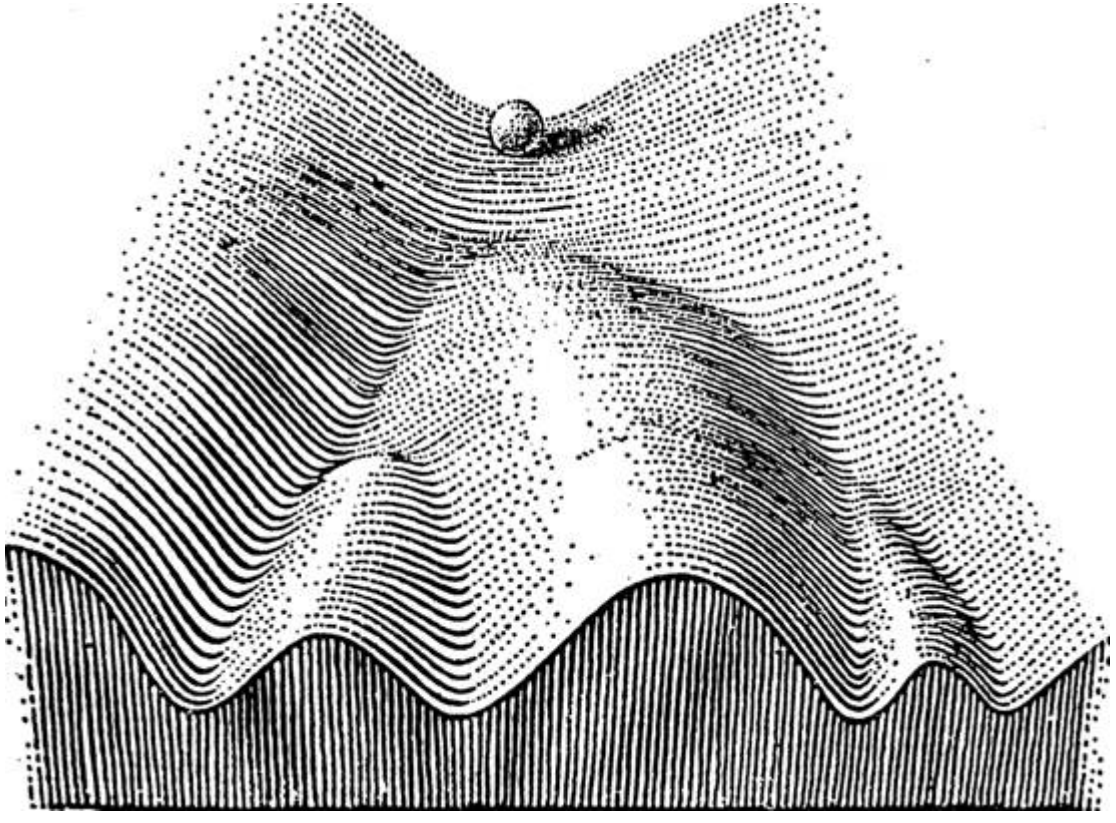
- Atomic-structure, identify rigid blocks
- Oligomer association and dissociation rates
  - Association restricted by diffusion
  - Dissociation escape from multi-dimensional well





Future Work: Organizing principles of the genome

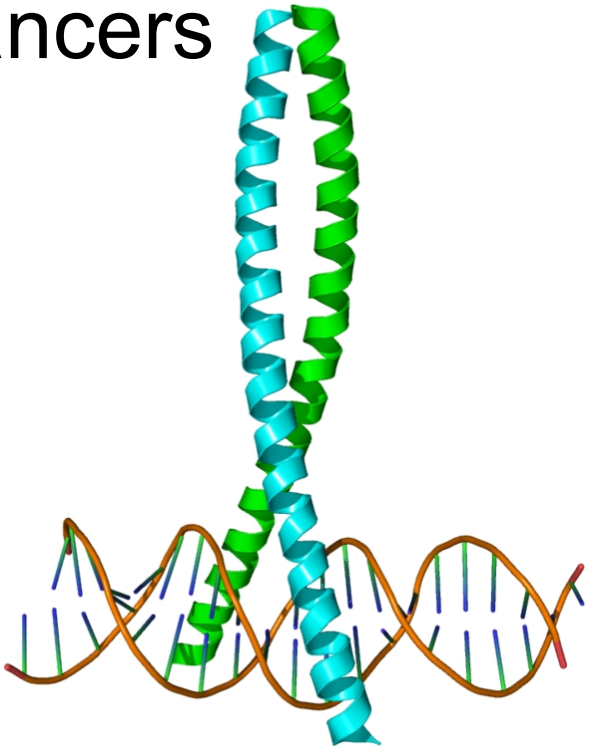
- Use genome-wide data to develop systems biology and biophysical models of gene regulation and gene expression

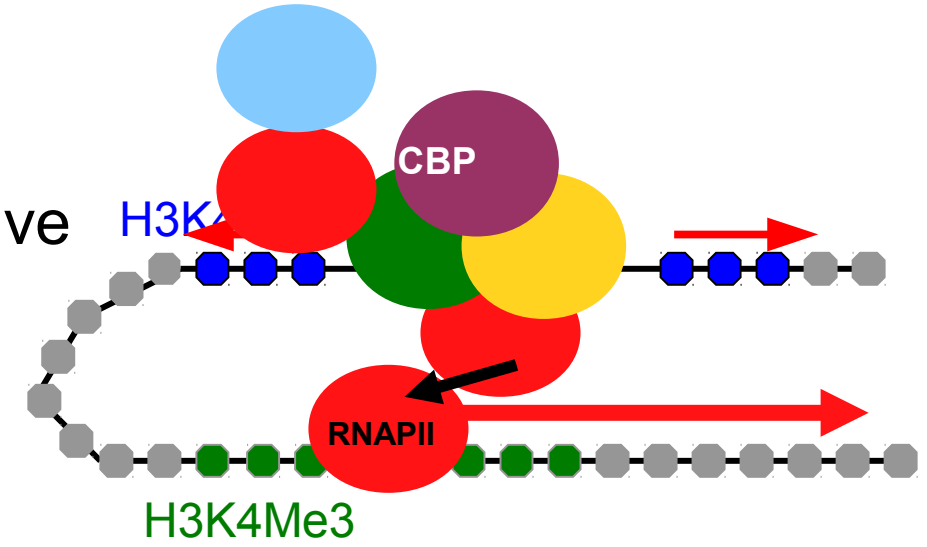


Waddington, 1953

Develop biophysical models of enhancers

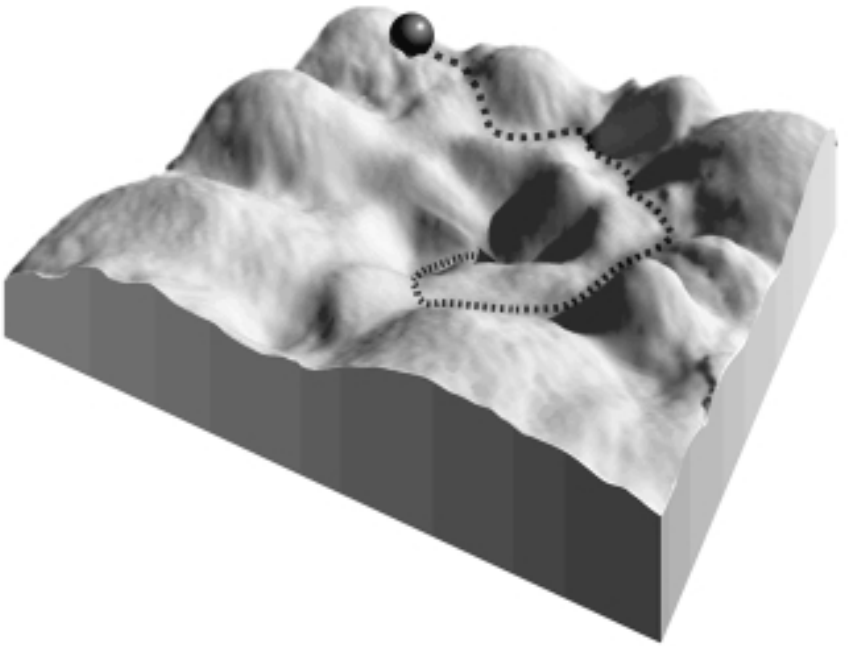
- TF-DNA binding
  - X-ray structures
  - ChIP-Seq binding
- DNA looping
  - Histones
    - H3K4me3 active
    - H3K27me3 repressive





Model stochastic gene expression for entire transcriptome

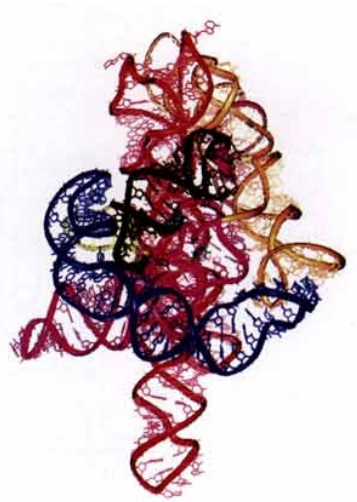
- Analytical models of gene expression noise
  - Parametric robustness
  - Time-scales
- Apply to genome-wide single-cell RNA-Seq
  - Propagation in pathways
  - Global factors
  - Dimensionality



Determine structure of RNAs

- Other species of novel non-coding RNAs
  - Identify structural motifs
- High-throughput sequencing of structure
  - PARS
  - SHAPE-Seq

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



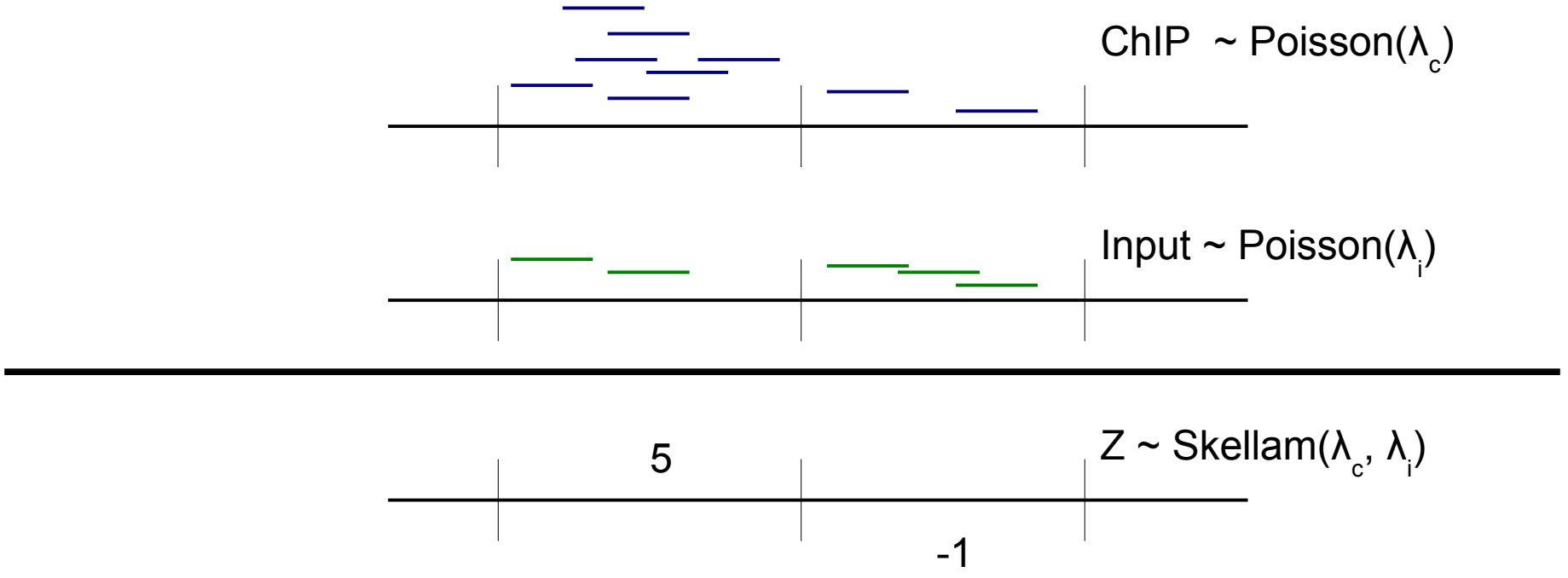
Acknowledgements

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

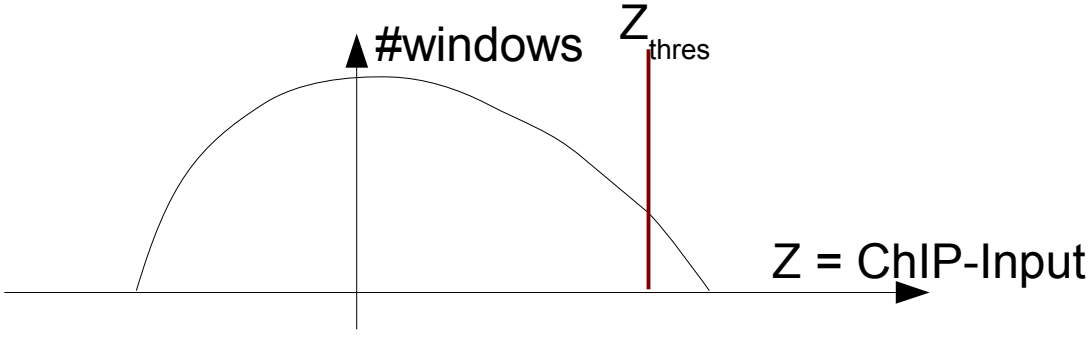
Thank You

?

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



- False Detection Rate (FDR) determine threshold



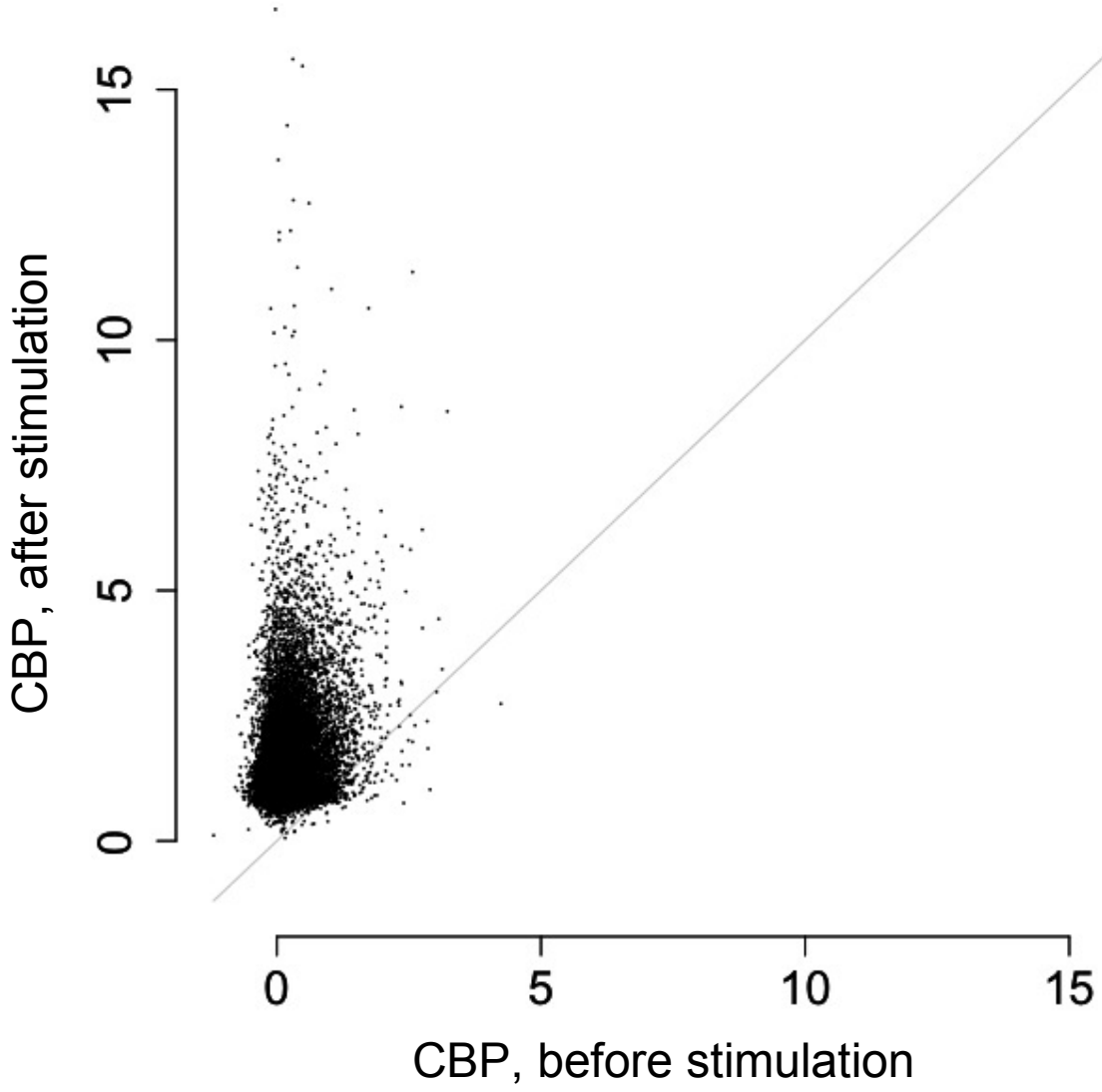
Use False Detection Ratio (FDR) to correct for multiple hypotheses

- 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 ~ Skellam(z, λ_1, λ_2)

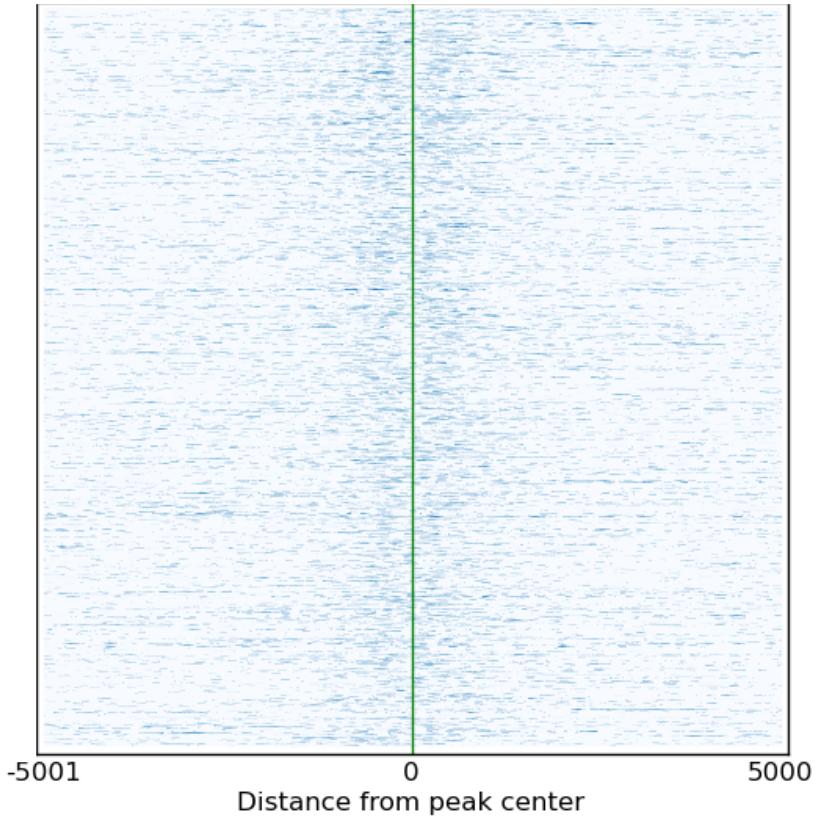
$$p(x) = e^{-(\lambda_1+\lambda_2)}(\lambda_1/\lambda_2)^{x/2} I_x(2\sqrt{\lambda_1\lambda_2})$$

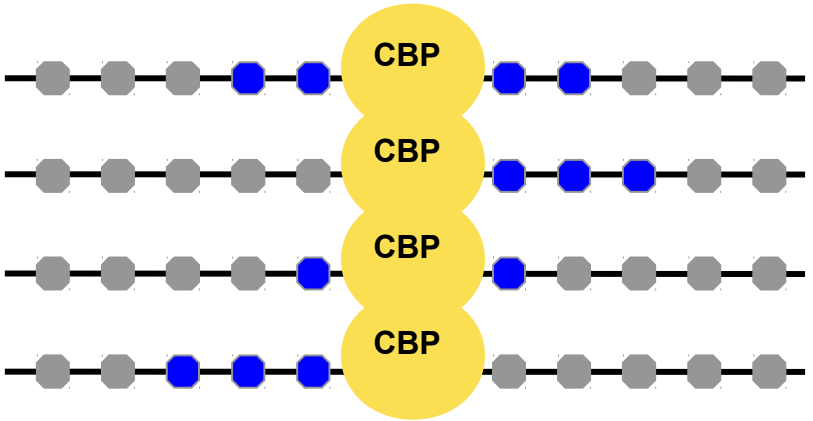
- Millions of windows need to be tested
  - FDR - expected fraction of false positives

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

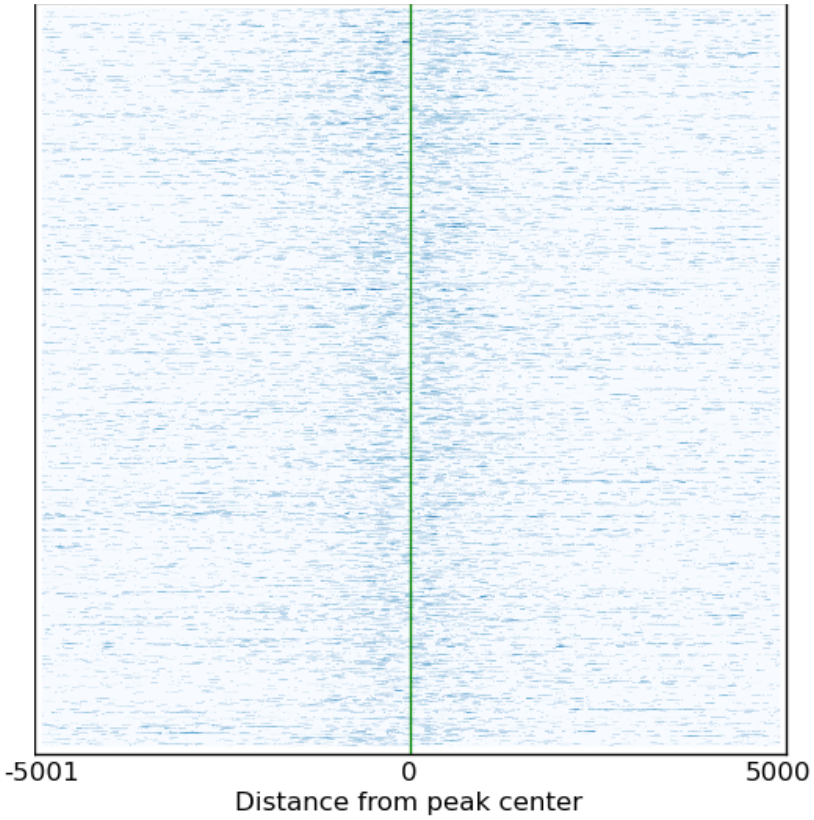


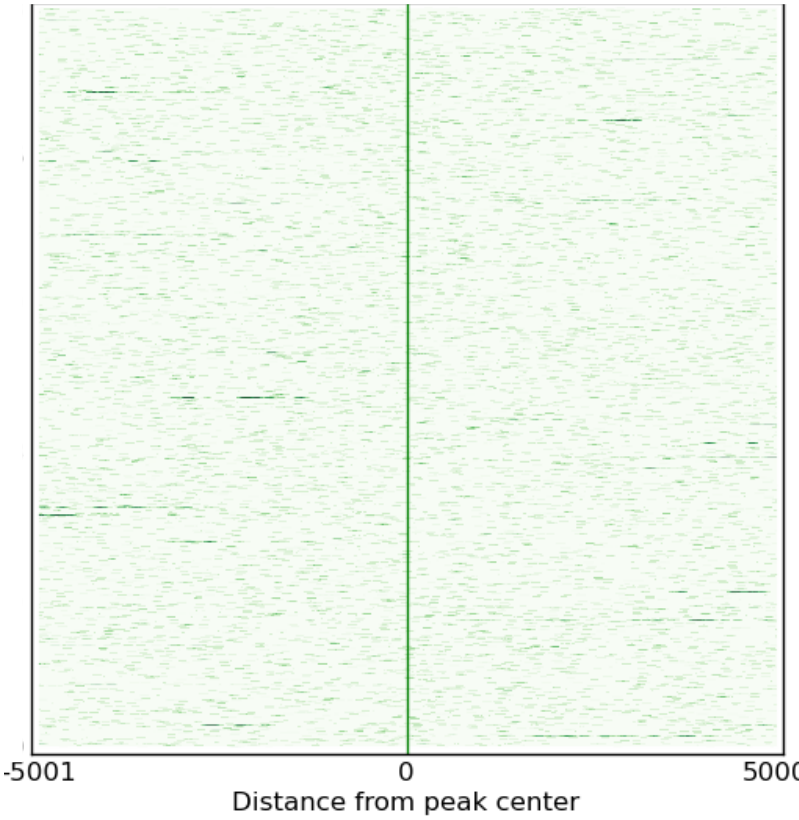
Aligning CBP peaks to calculate H3K4me1 binding profiles

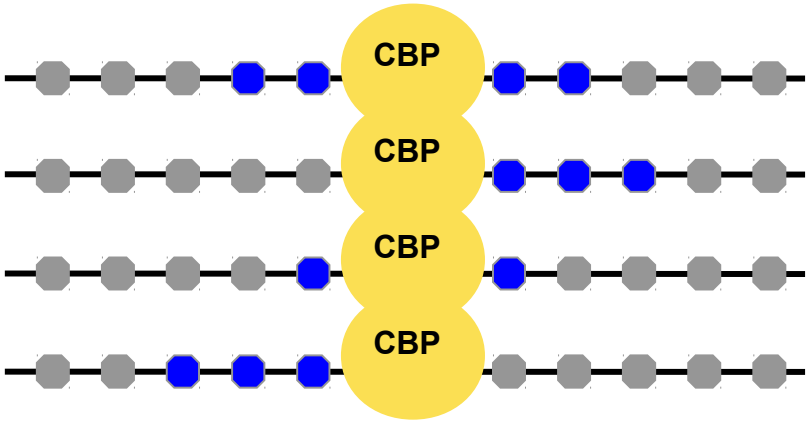




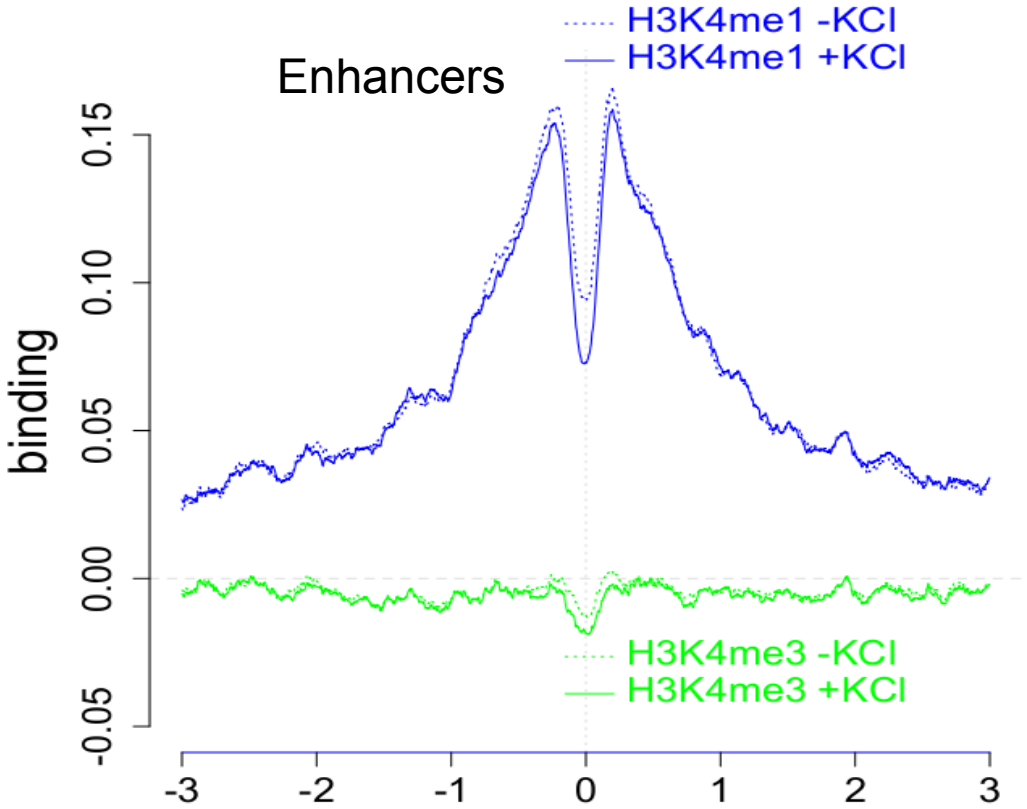
Aligning CBP peaks to calculate H3K4me1 and H3K4me3 binding profiles

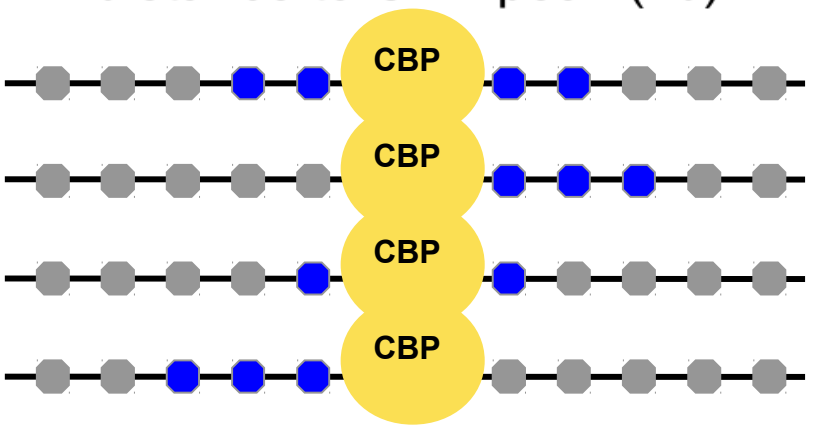




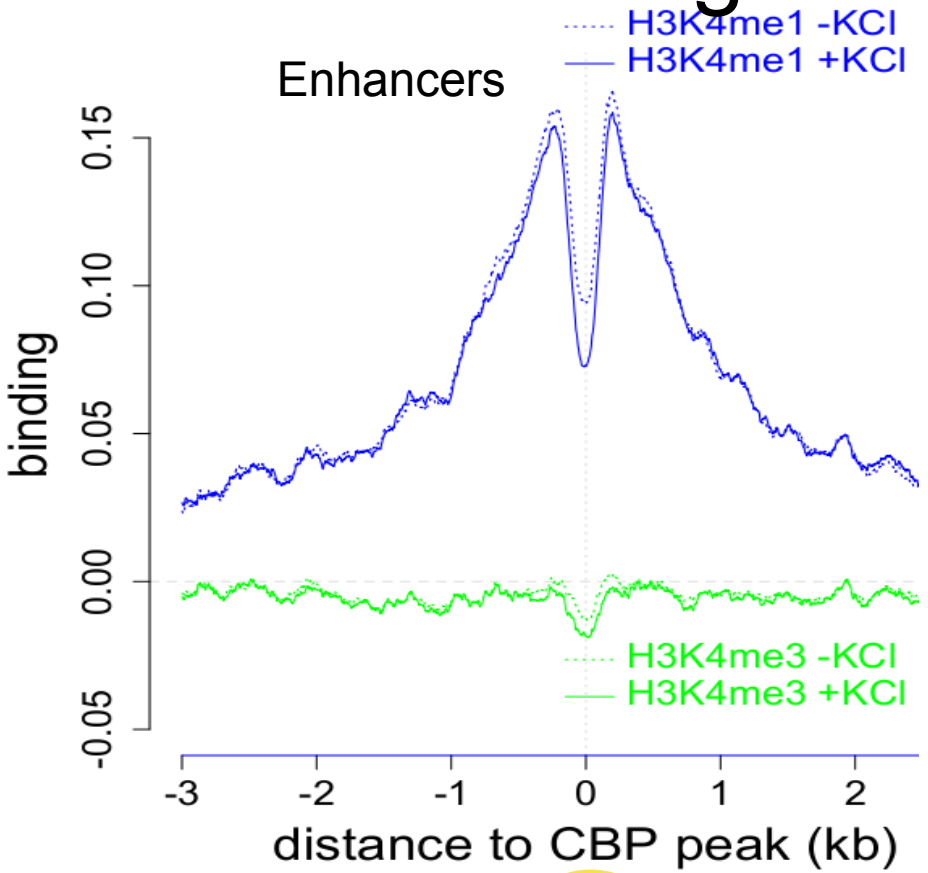


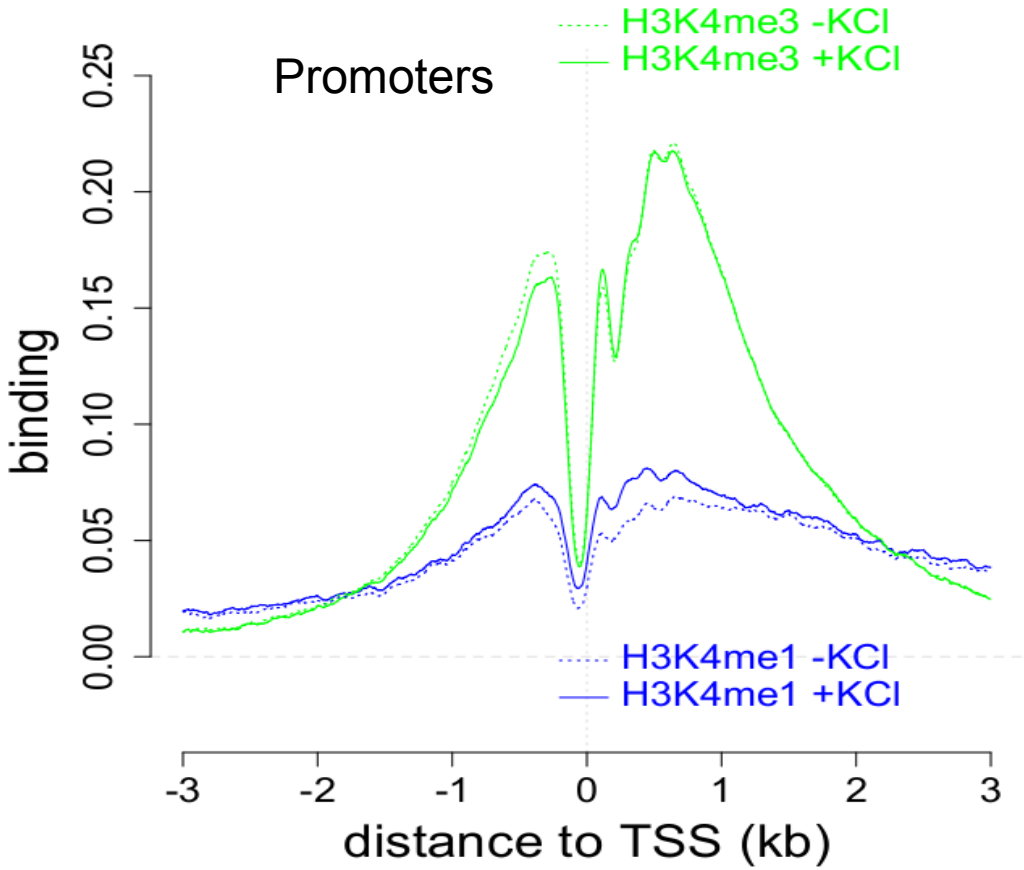
Enhancers have high levels of H3K4me1 and low levels of H3K4me3

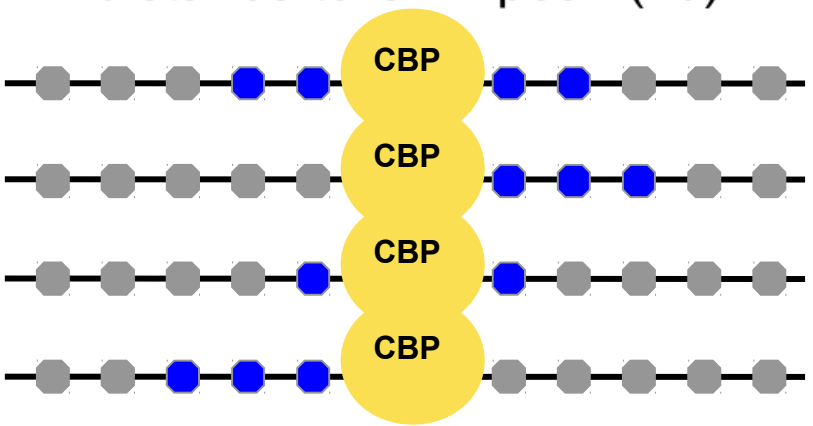




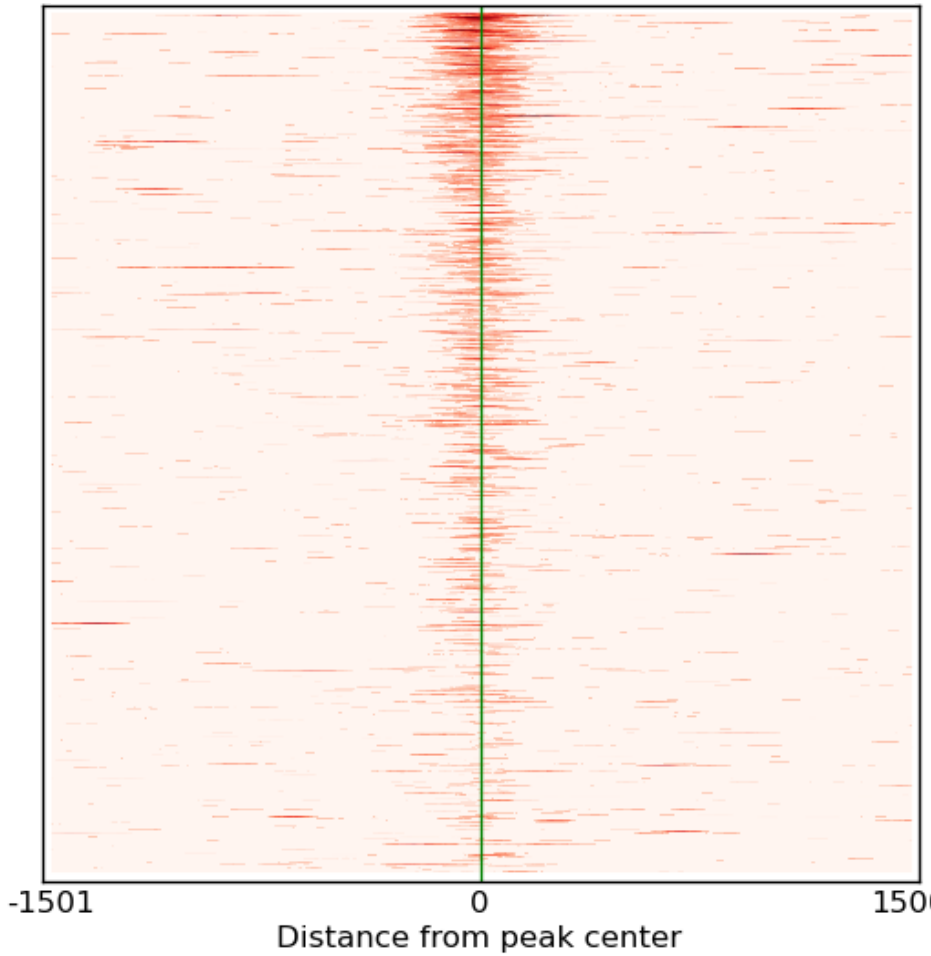
Transcription start sites have high levels of H3K4me1 and high levels of H3K4me3







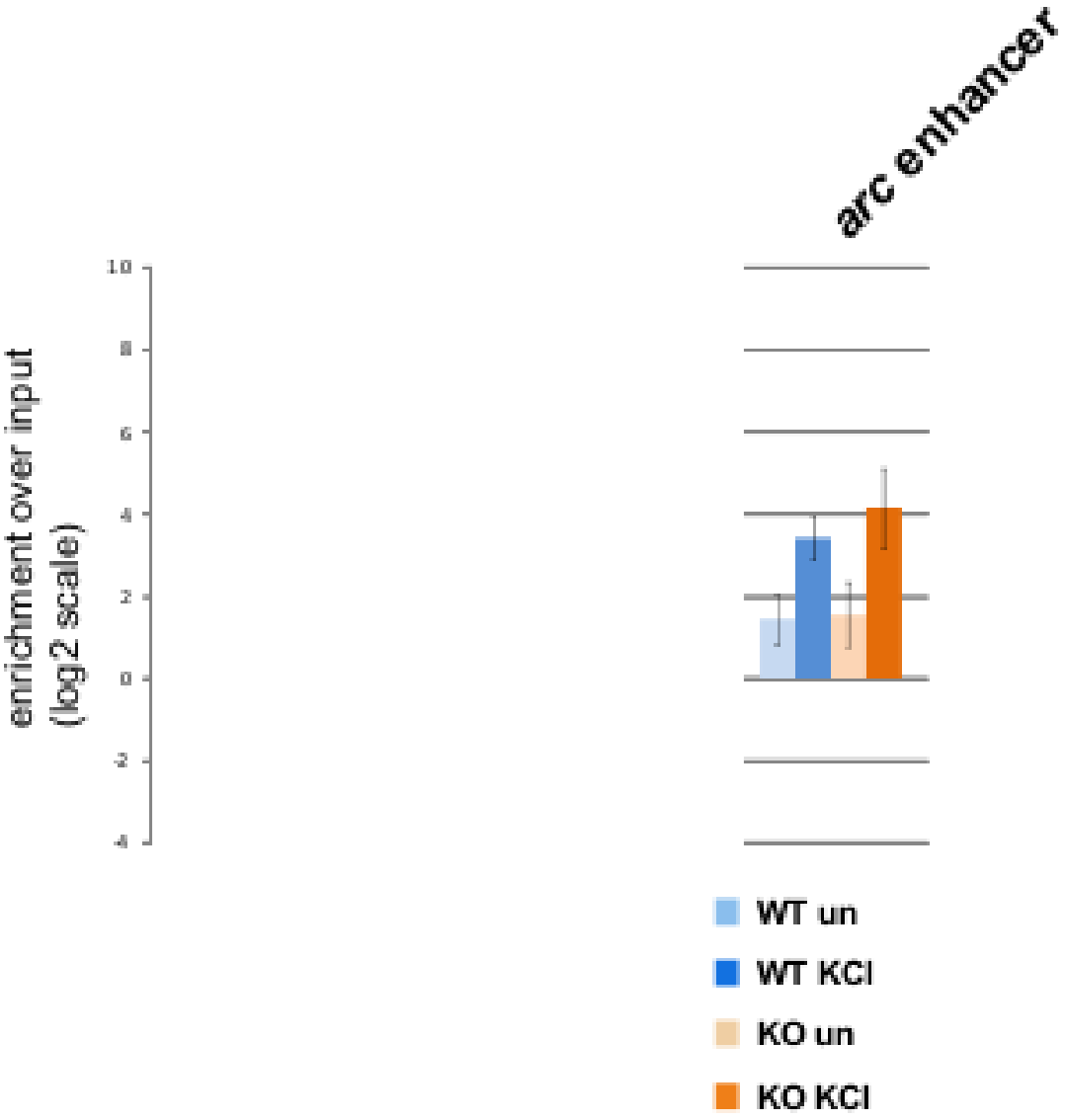
RNAPII binds at activity-dependent enhancers

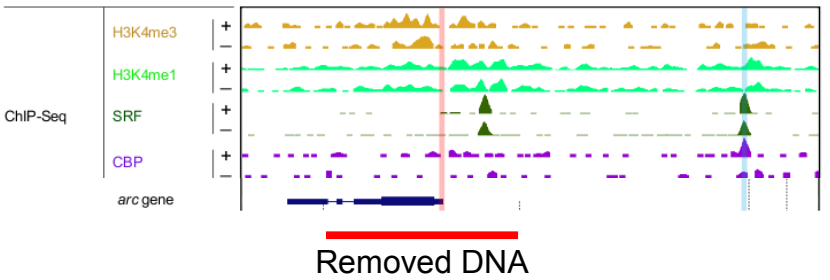


~100 enriched motifs are found

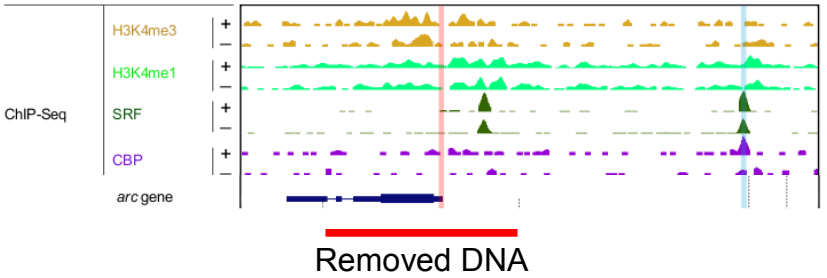
Word	**Enrichment**	**Known TF**
TGASTCA	4.74	Fos/Jun
TGACGTCA	6.41	Creb
CTAWWWATA	3.34	Srf
TCGTG	1.56	Npas4
CTGCCAAA	3.34	?

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

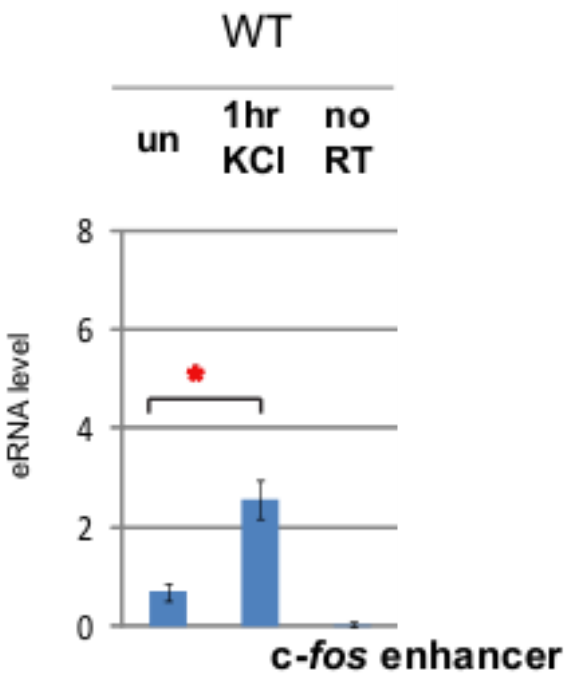


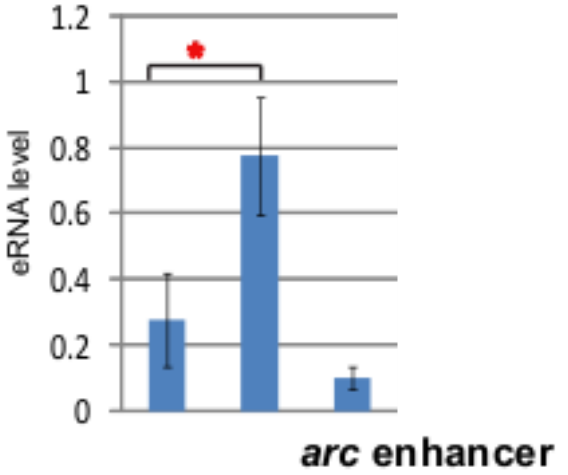


Transcription at the Fos and Arc enhancers



7 kb

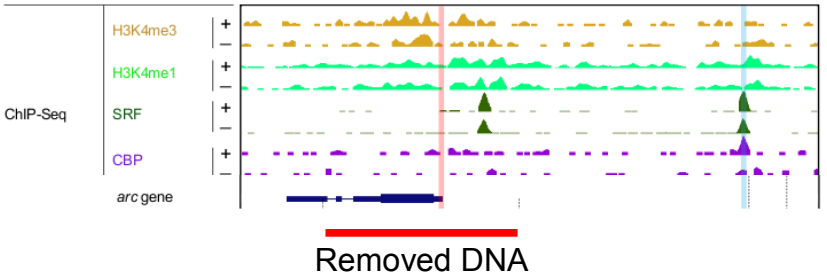




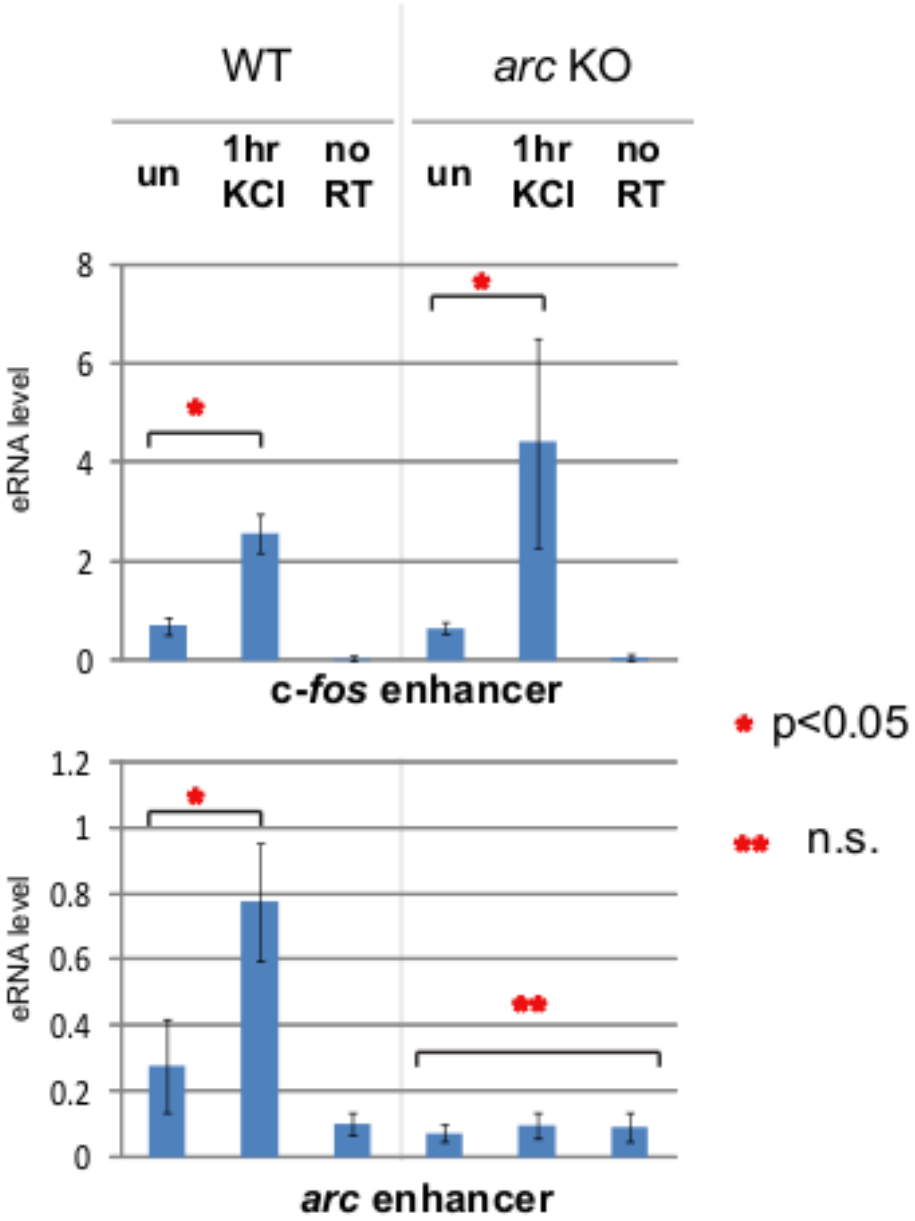
* p<0.05

** n.s.

No transcription at Arc enhancer in mutant



7 kb



Estimating the production rate of eRNAs

$$\frac{dE}{dt} = kN - \frac{E}{\tau_E}$$

$$k = \frac{E^*}{N\tau_E} \sim \frac{10^3}{10^4 \times 10^{-1}\mathrm{h}} = 1\mathrm{h}^{-1}$$

$$\frac{\text{Variance strong promoter}}{\text{Variance weak promoter with enhancers}} = \frac{\mathrm{Var}[(1+Nc)k]}{\mathrm{Var}[k] + N\mathrm{Var}[ck]} = \frac{(1+Nc)^2\mathrm{Var}[k]}{(1+Nc^2)\mathrm{Var}[k]} \sim N$$

Parameters for the eRNA fit

$$\lambda = \frac{k_{drop}\ \mathrm{s}^{-1}}{k_{elong}\ \mathrm{bp}^{-1}\mathrm{s}^{-1}} \sim \frac{2 \times 10^{-2}}{20}\ \mathrm{bp}^{-1} = 10^{-3}\ \mathrm{bp}^{-1}$$

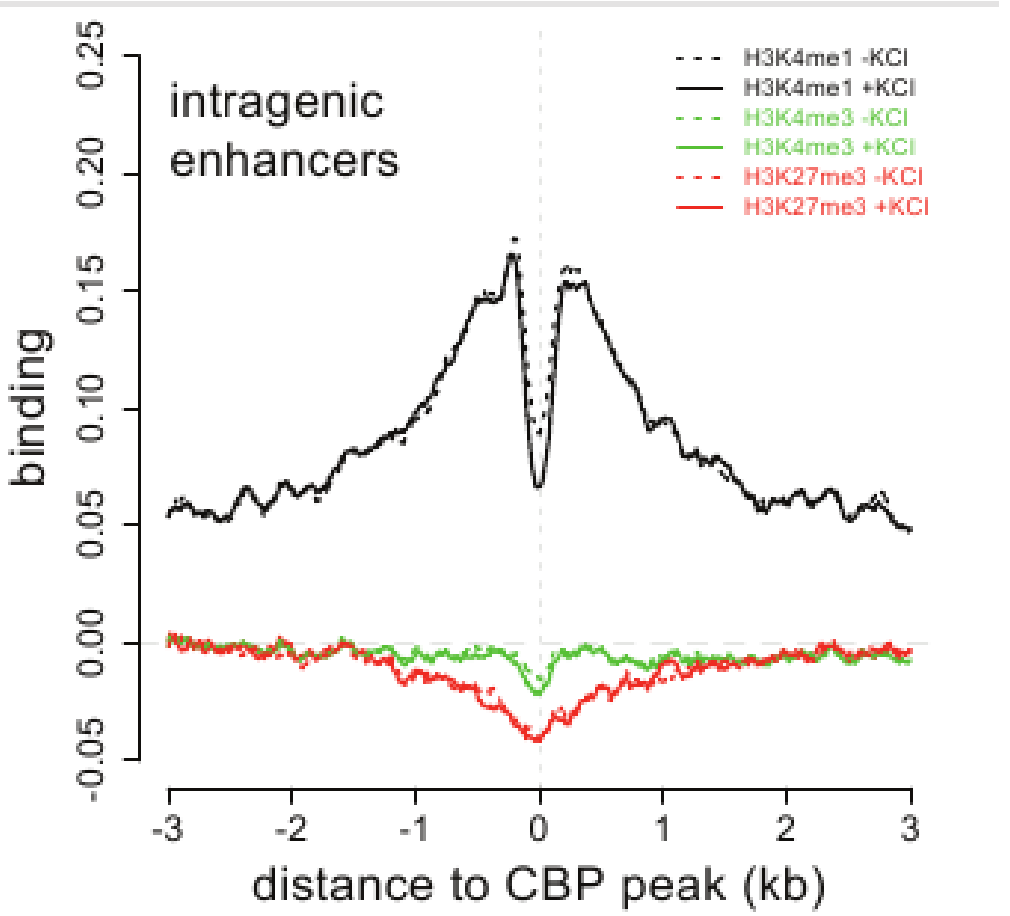
$$\tau_{decay} = \tau_{find} + \tau_{bp} L$$

$$H(x,t) = \frac{k\kappa}{\mu_x(\mu_x - \lambda)}(e^{-\lambda x} - e^{-\mu_x x}) \times e^{-\mu_t t}$$

$$E(x) = \sqrt{\frac{\pi}{2\lambda}}\frac{\gamma k}{\lambda} e^{-\delta^2/2\lambda - \lambda x^2/2} i\left[\mathrm{erf}\left(\frac{\delta i - \lambda i x}{\sqrt{2\pi}}\right) - \mathrm{erf}\left(\frac{\delta i}{\sqrt{2\lambda}}\right)\right]$$

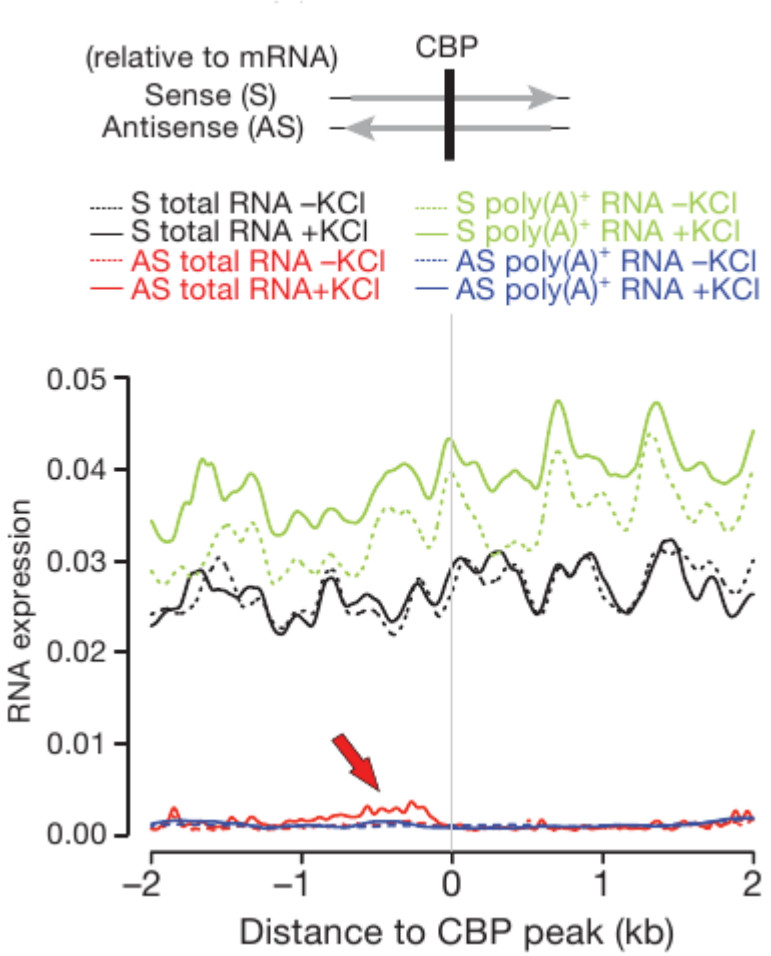
Intragenic enhancers

- ~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
    - Annotated RNAs ~ 98.3%
    - eRNAs ~ 0.02%
      - 1 in 10,000 reads is an eRNA read
      - mRNAs ~100 times more abundant

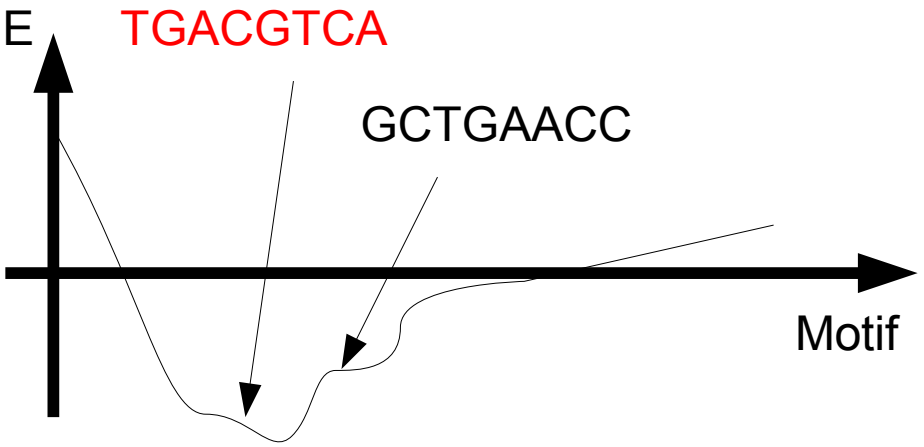
Can the read count be predicted from sequence motifs?

$$R = f\big(E(\theta)\big) + \xi$$

noise

#reads

Binding energy



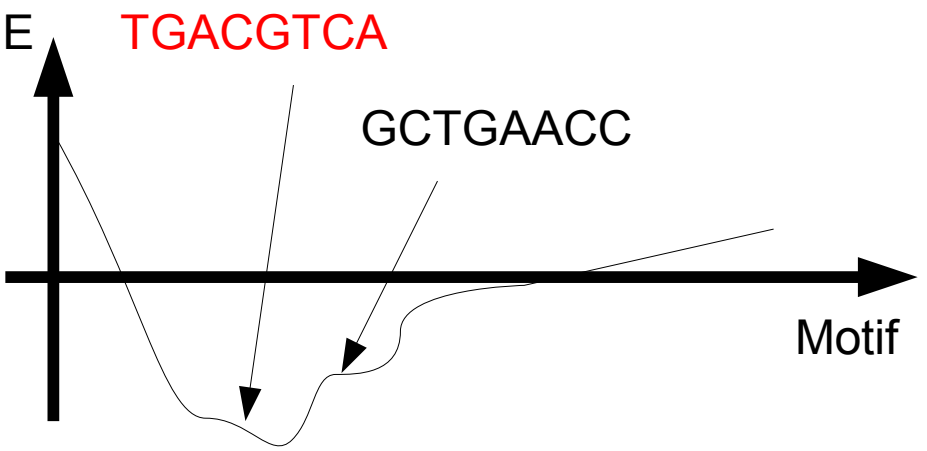
Assume transfer function maximizes mutual information

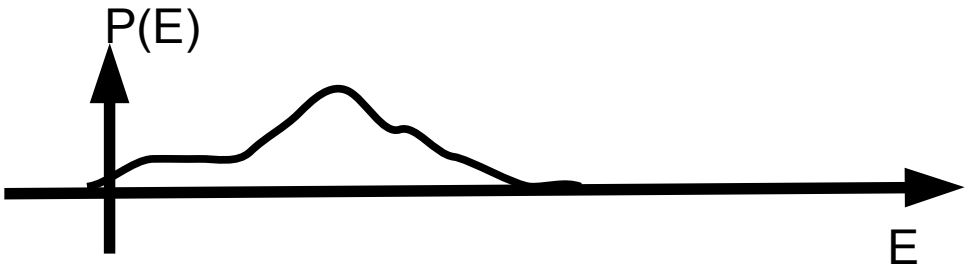
$$R = f\big(E(\theta)\big) + \xi$$

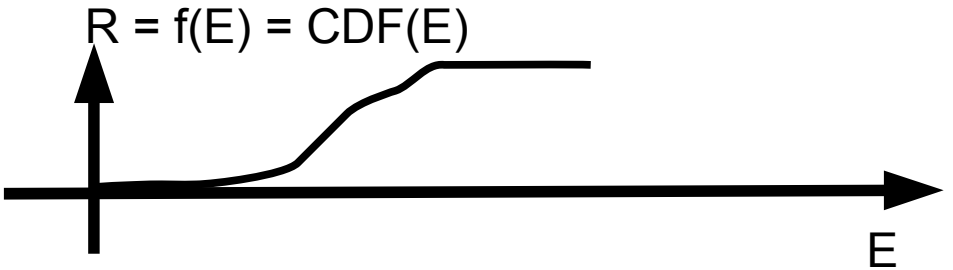
noise

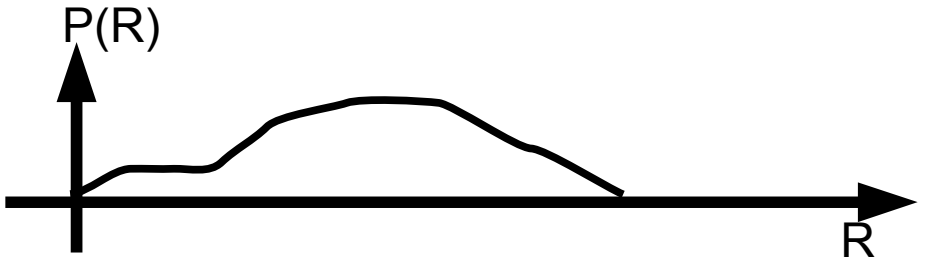
#reads

Binding energy (from JASPAR)



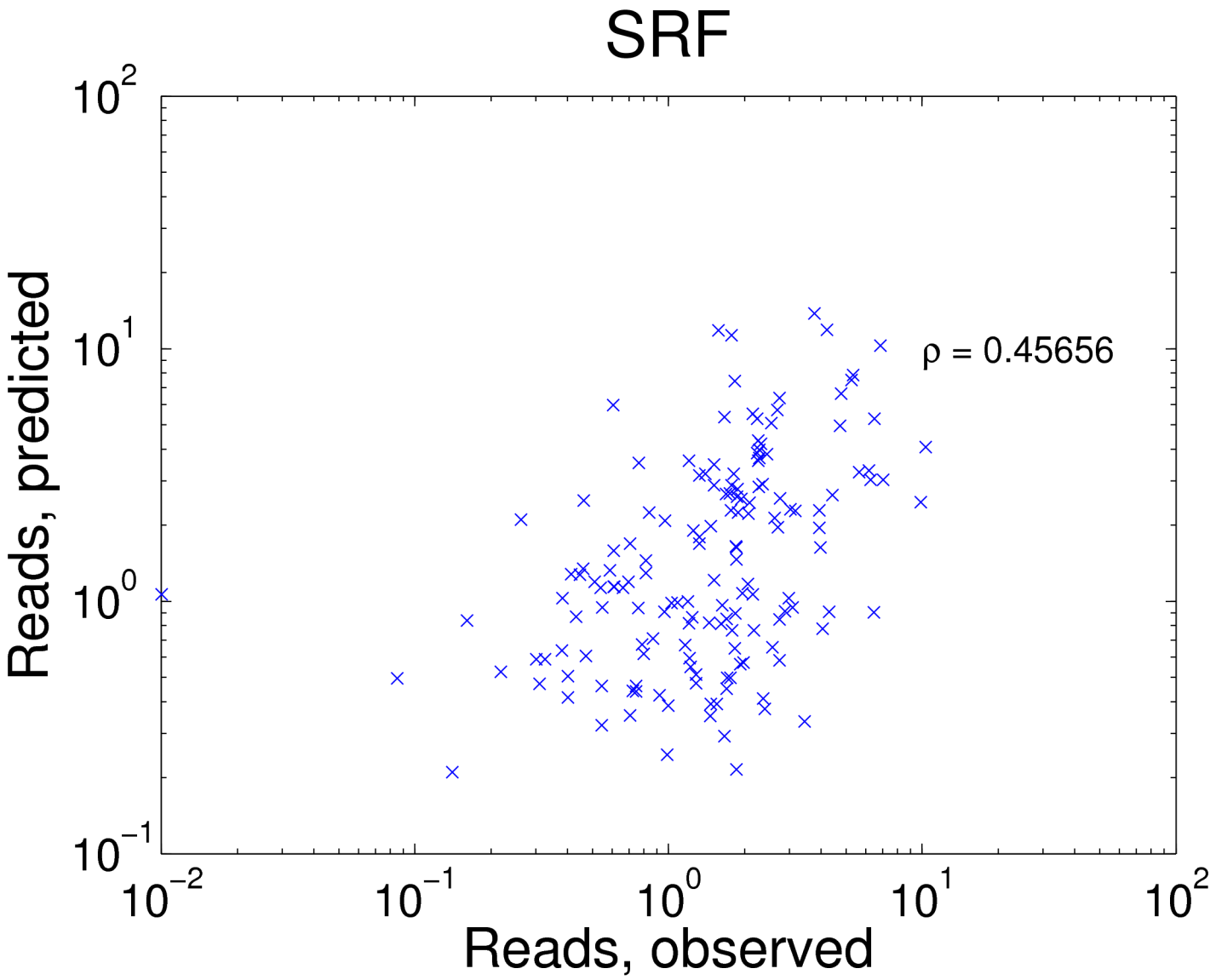






- *f* monotone
- max *I(R; E)*
- Noise small and gaussian

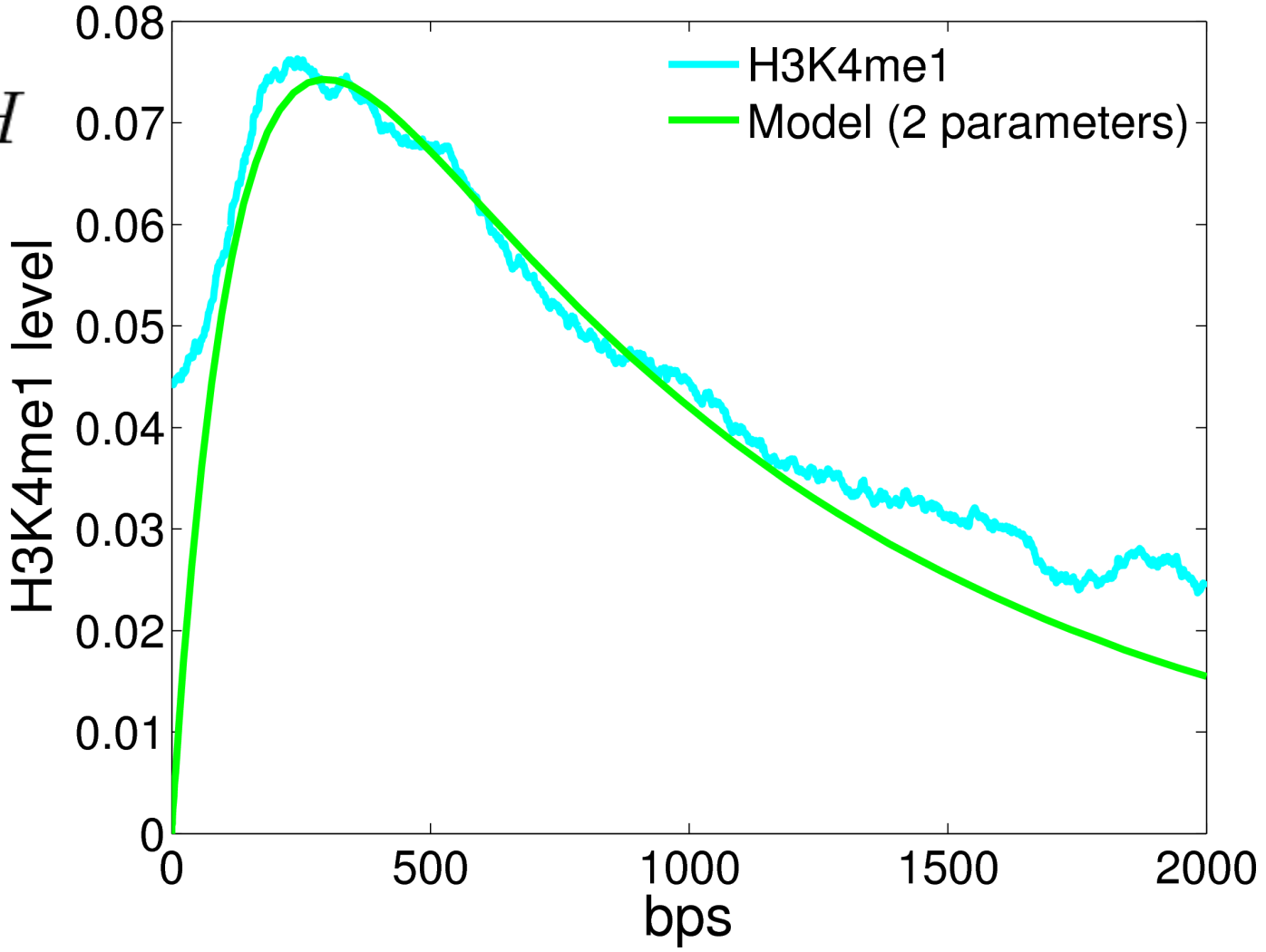
Number of reads can be predicted by binding energy



Establishing H3K4me1 levels at enhancers

$$\frac{dP}{dx} = k - \lambda P$$

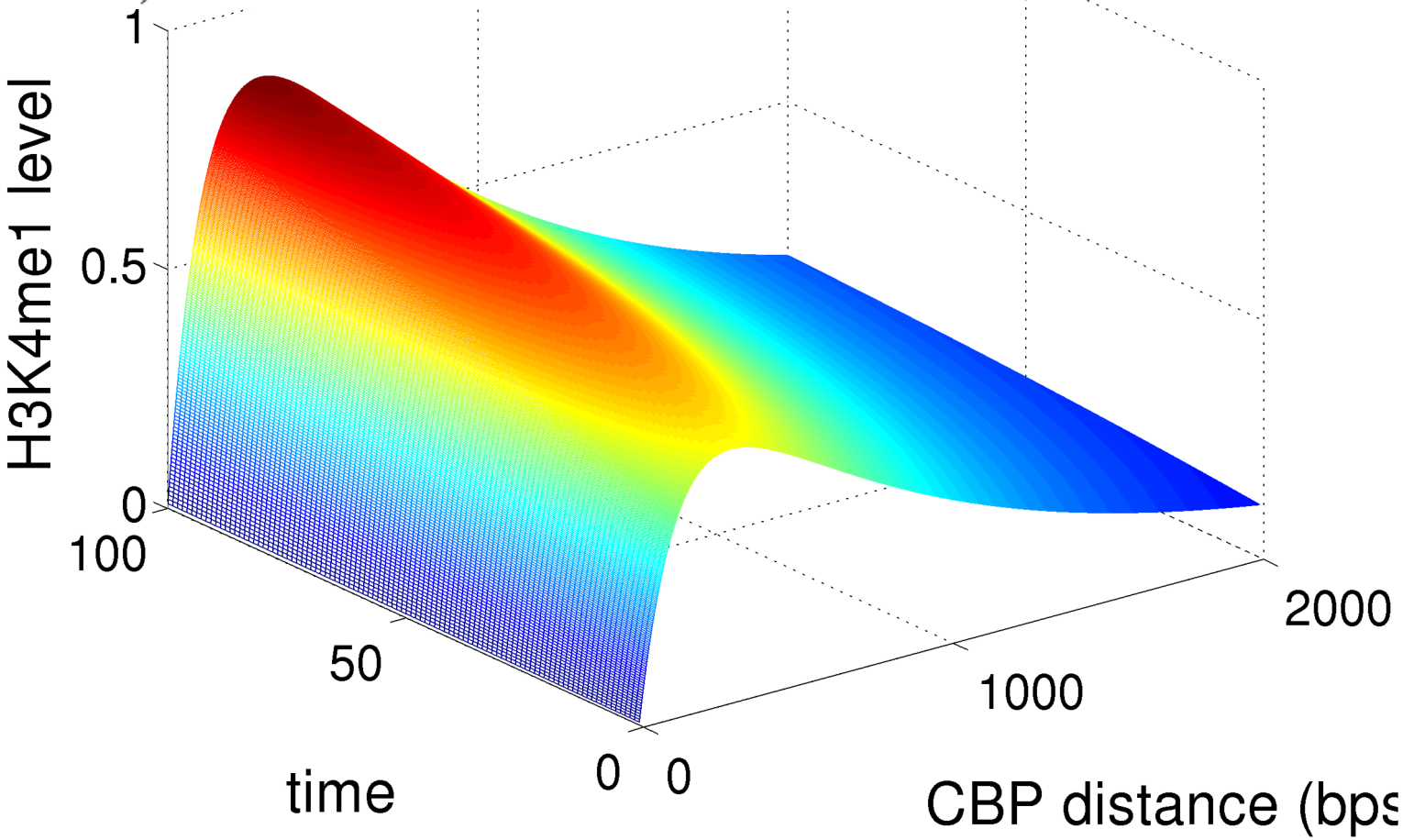
$$\frac{dH}{dx} = \kappa P(x) - \mu H$$

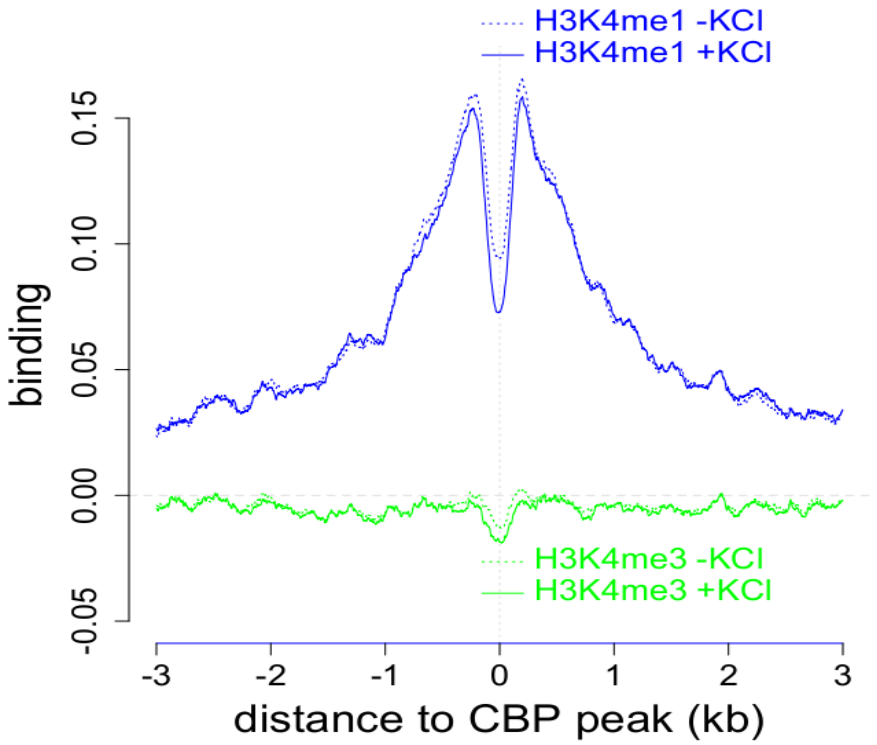


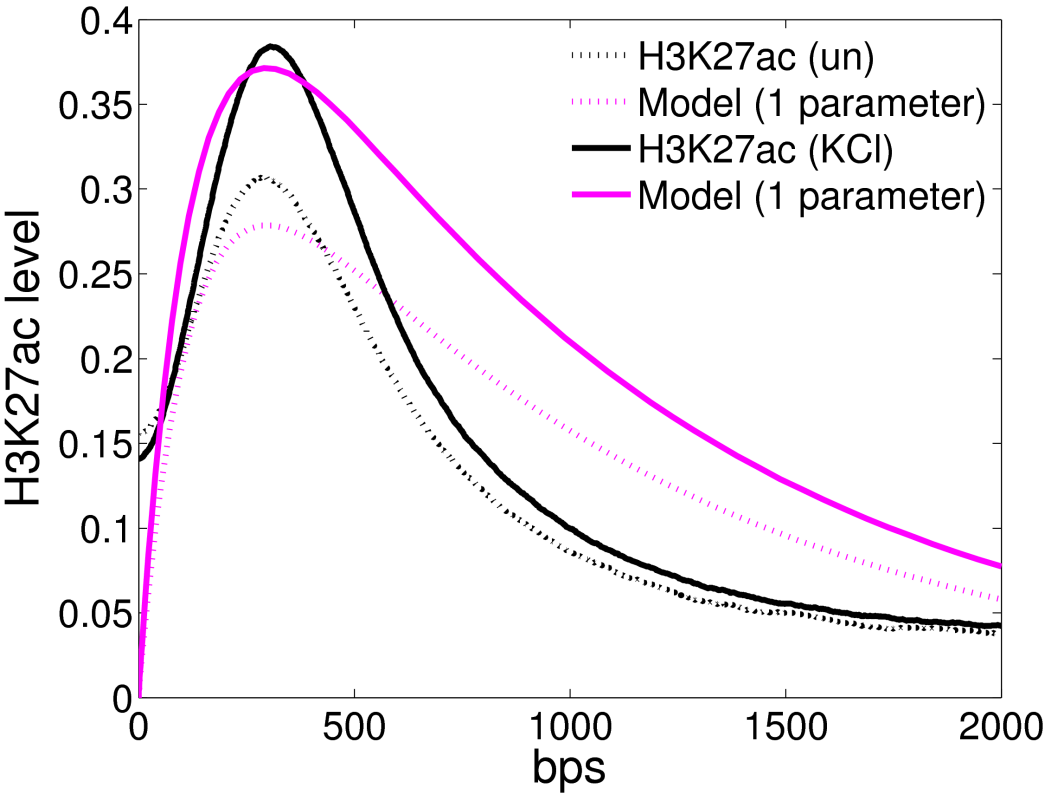
A PDE for histone levels

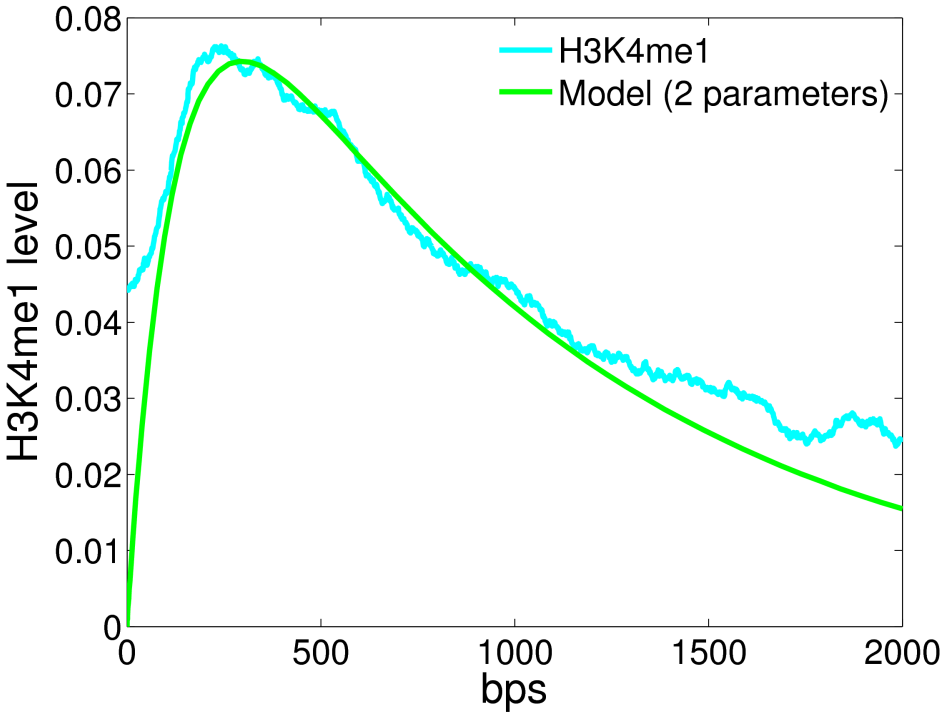
$$\frac{\partial H}{\partial x} + \frac{\partial H}{\partial t} = \kappa P(x,t) - \mu_x H - \mu_t H$$

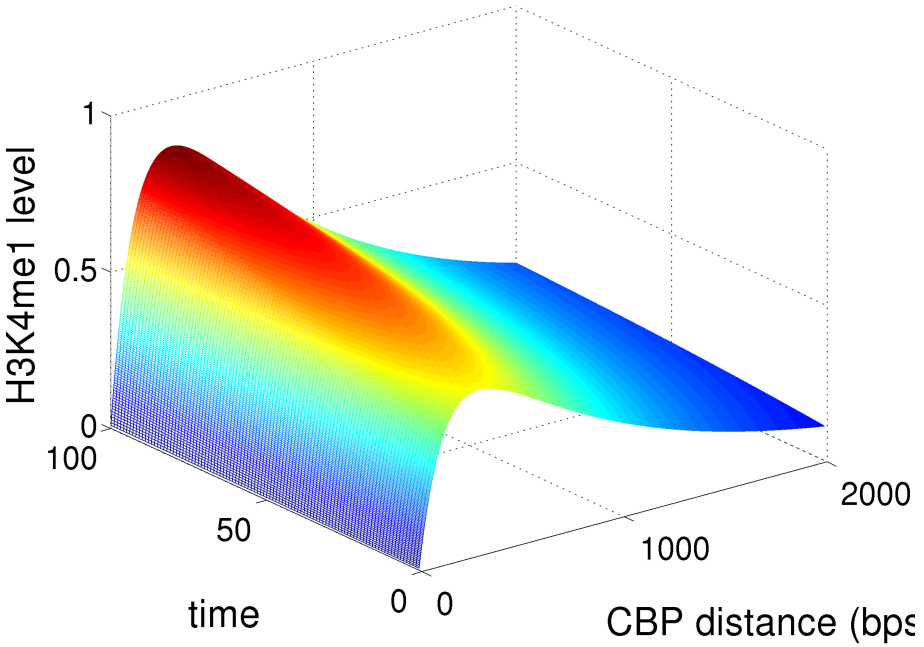
$$H(x,t) = \frac{k\kappa}{\mu_x(\mu_x - \lambda)}(e^{-\lambda x} - e^{-\mu_x x}) \times e^{-\mu_t t}$$



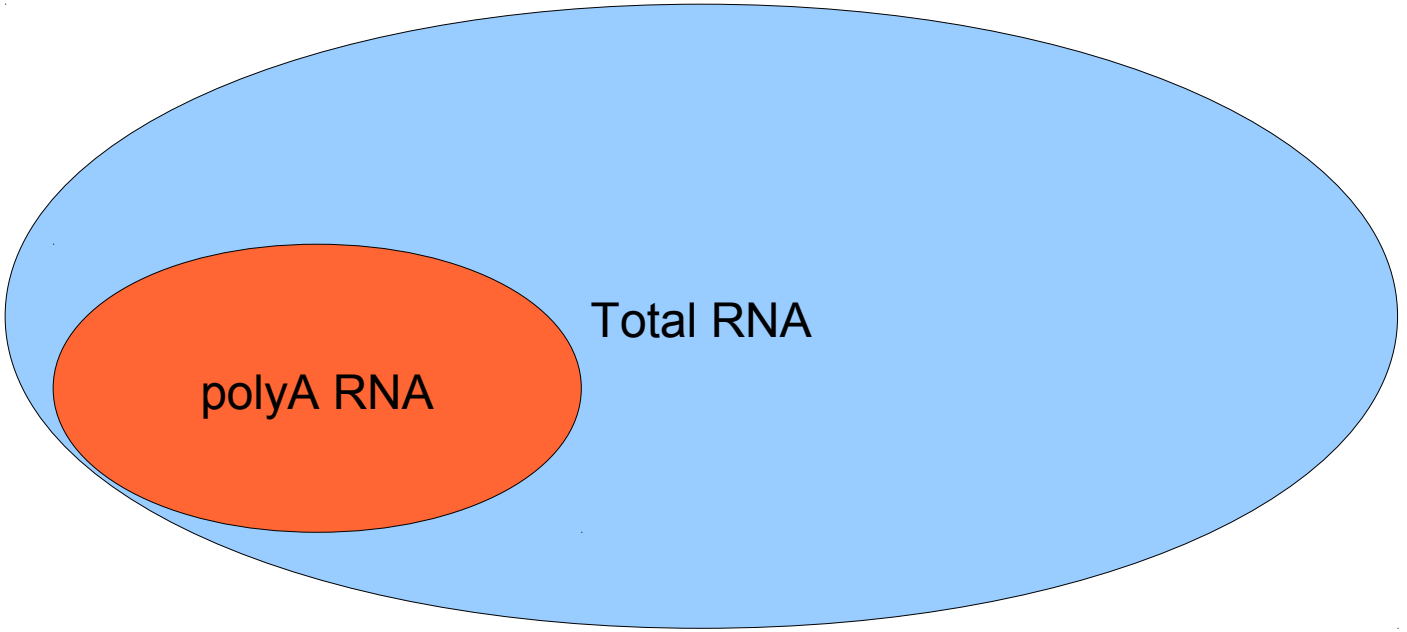

H3K4me1 -KCl
H3K4me1 +KCl
H3K4me3 -KCl
H3K4me3 +KCl
binding
0.15
0.10
0.05
0.00
-0.05
-3
-2
-1
0
1
2
3
distance to CBP peak (kb)


H3K27ac (un)
Model (1 parameter)
H3K27ac (KCl)
Model (1 parameter)
H3K27ac level
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
500
1000
1500
2000
bps


H3K4me1
Model (2 parameters)
H3K4me1 level
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
0
500
1000
1500
2000
bps

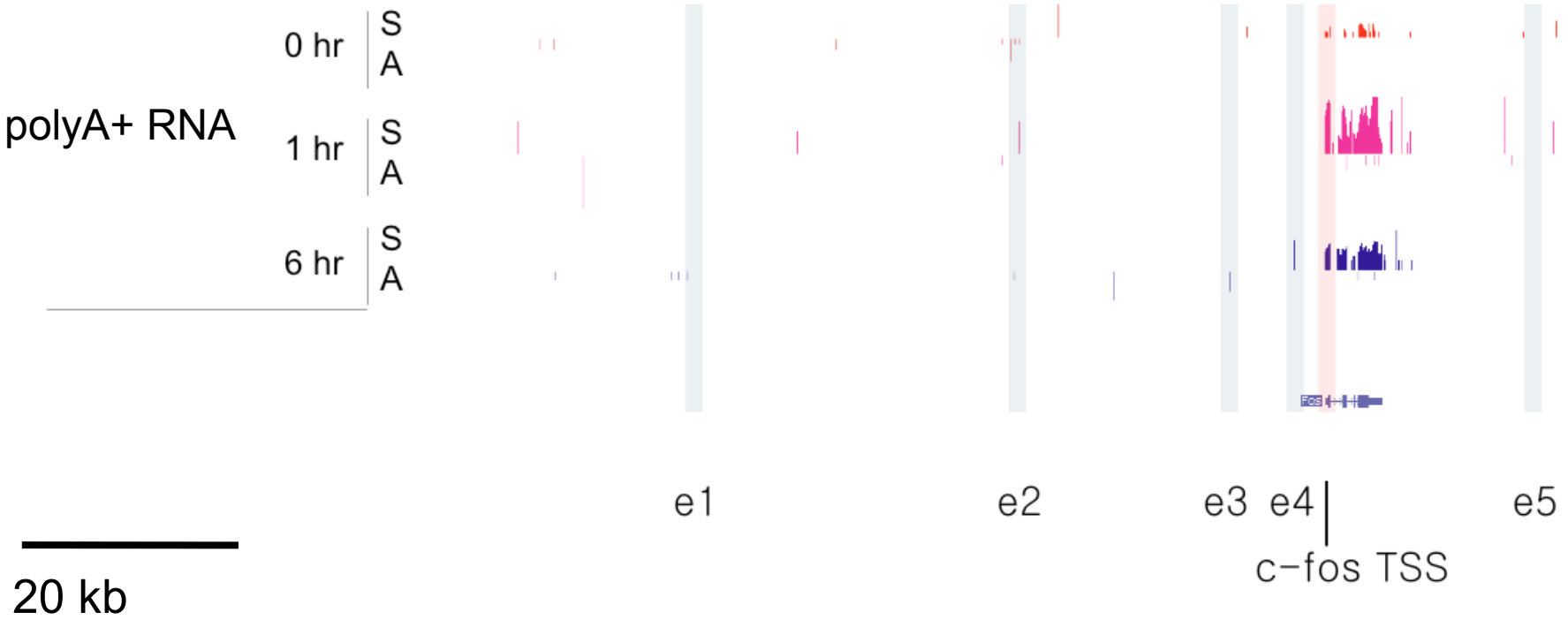

H3K4me1 level
1
0.5
0
100
50
0
time
0
1000
2000
CBP distance (bps

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

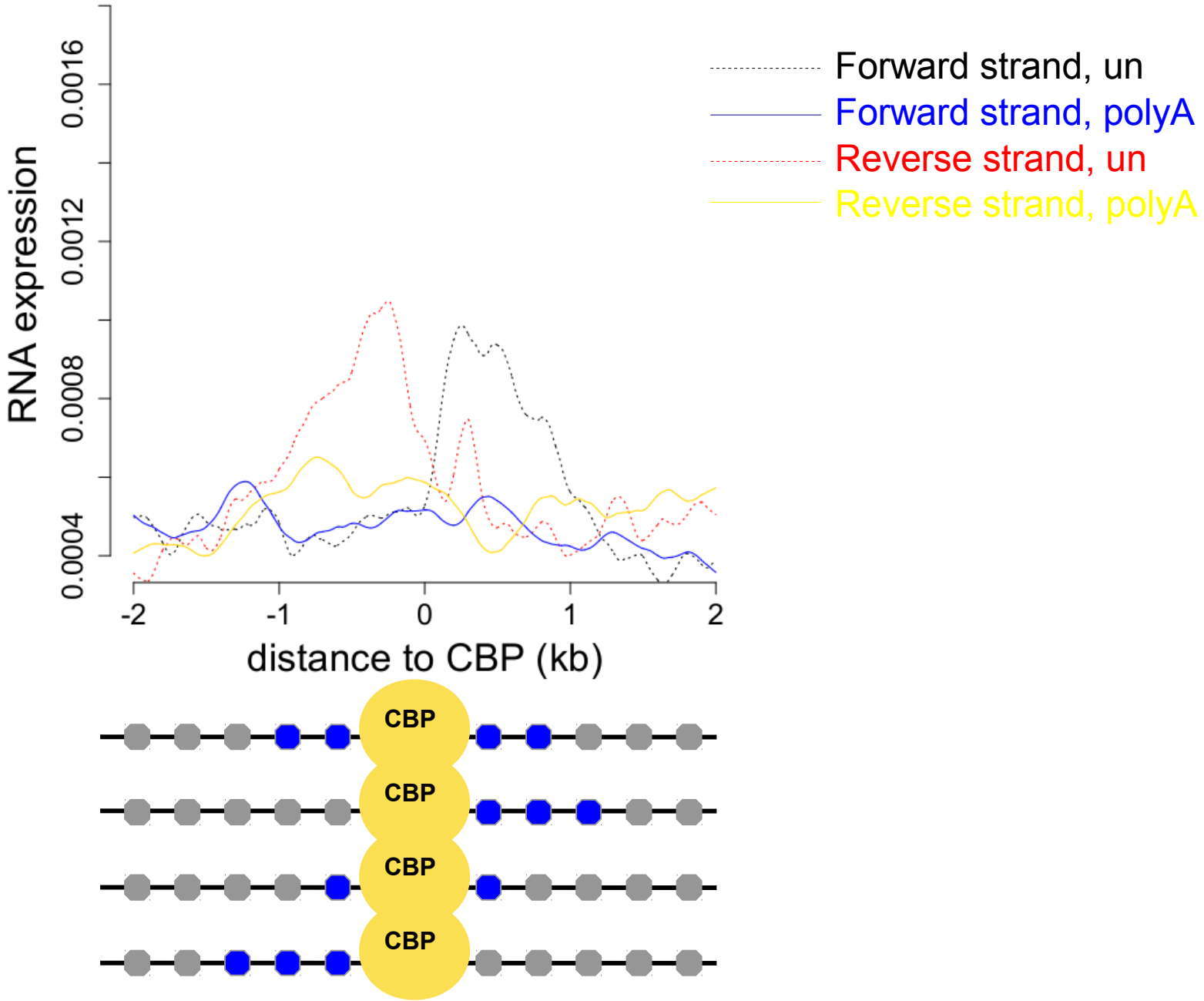


Transcription of mRNA at the *fos* locus

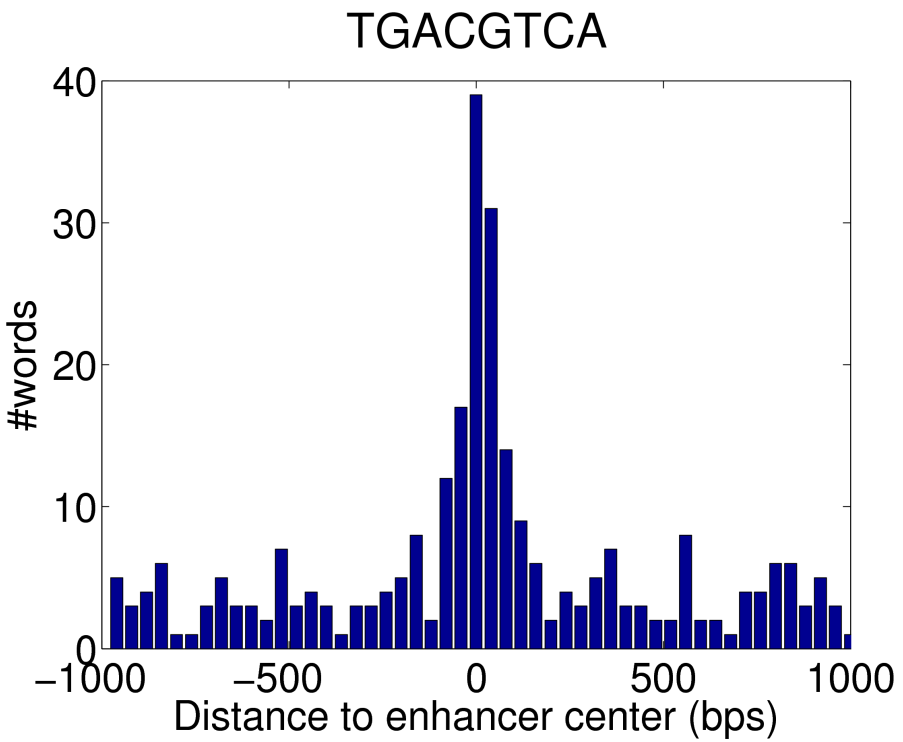
ACGUUUGUACCUAGCUAGCUUACGAG AAAAAAAAAAAAAAAAAAAA

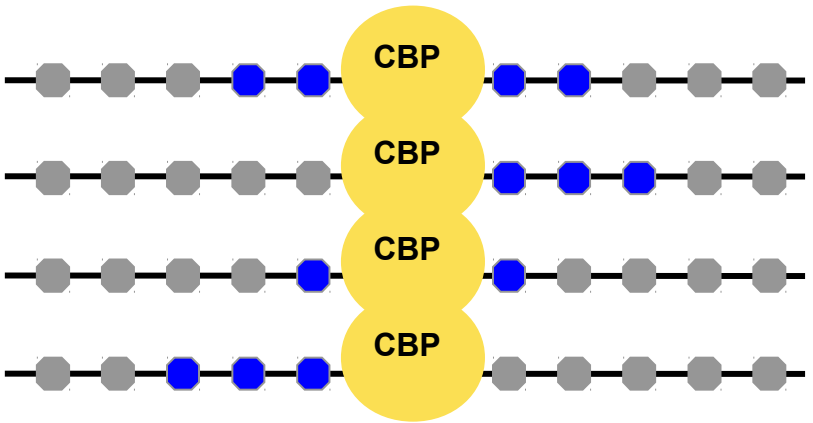


eRNAs are not polyadenylated



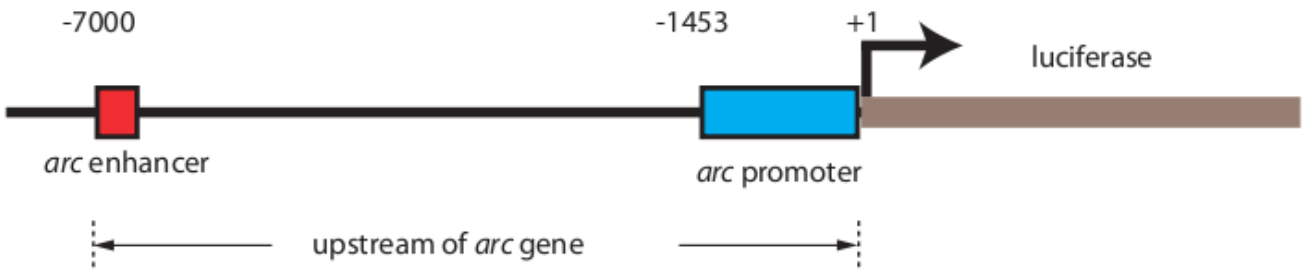
~100 enriched motifs found at enhancers





We identified ~12,000 activity-dependent enhancers throughout the genome

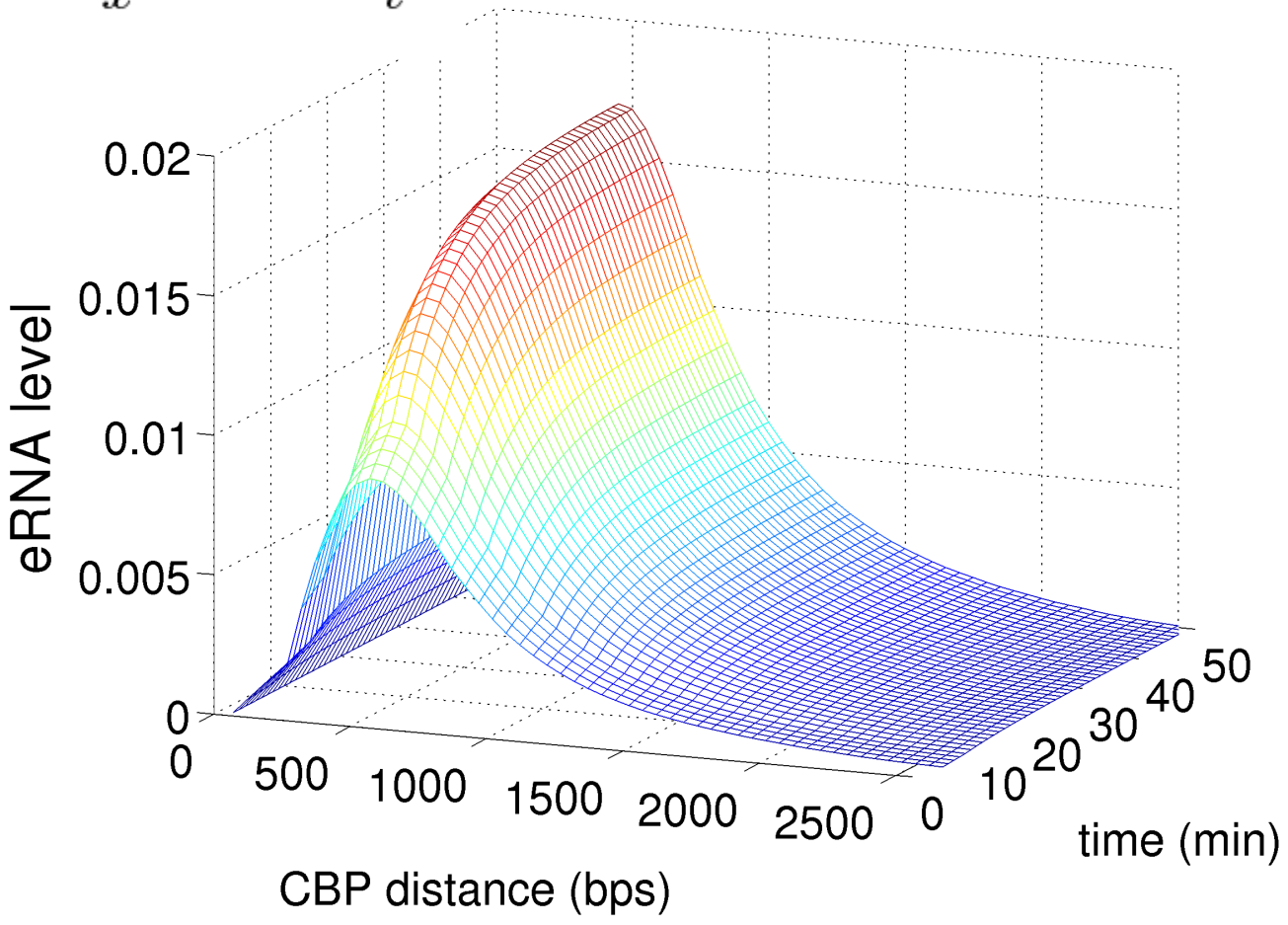
- **CBP** peak
- **High** levels of flanking **H3K4me1**
- **Low** levels of **H3K4me3**
  - 8/8 tested activity-dependent enhancers were validated using a luciferase assay



A PDE for eRNA levels

$$\frac{\partial P}{\partial x} + \frac{\partial P}{\partial t} = k(x,t) - \lambda_x P - \lambda_t P$$

$$\frac{\partial E}{\partial x} + \frac{\partial E}{\partial t} = \gamma P(x,t) - \delta_x x E - \delta_t t$$



Master Equation (**ME**) description

$$\frac{dP_j}{dt} = \sum_i W_{ij} P_i(t) - W_{ji} P_j(t)$$

P_j - **Probability** of having j molecules

W_{ij} - **Transition rate** from i to j