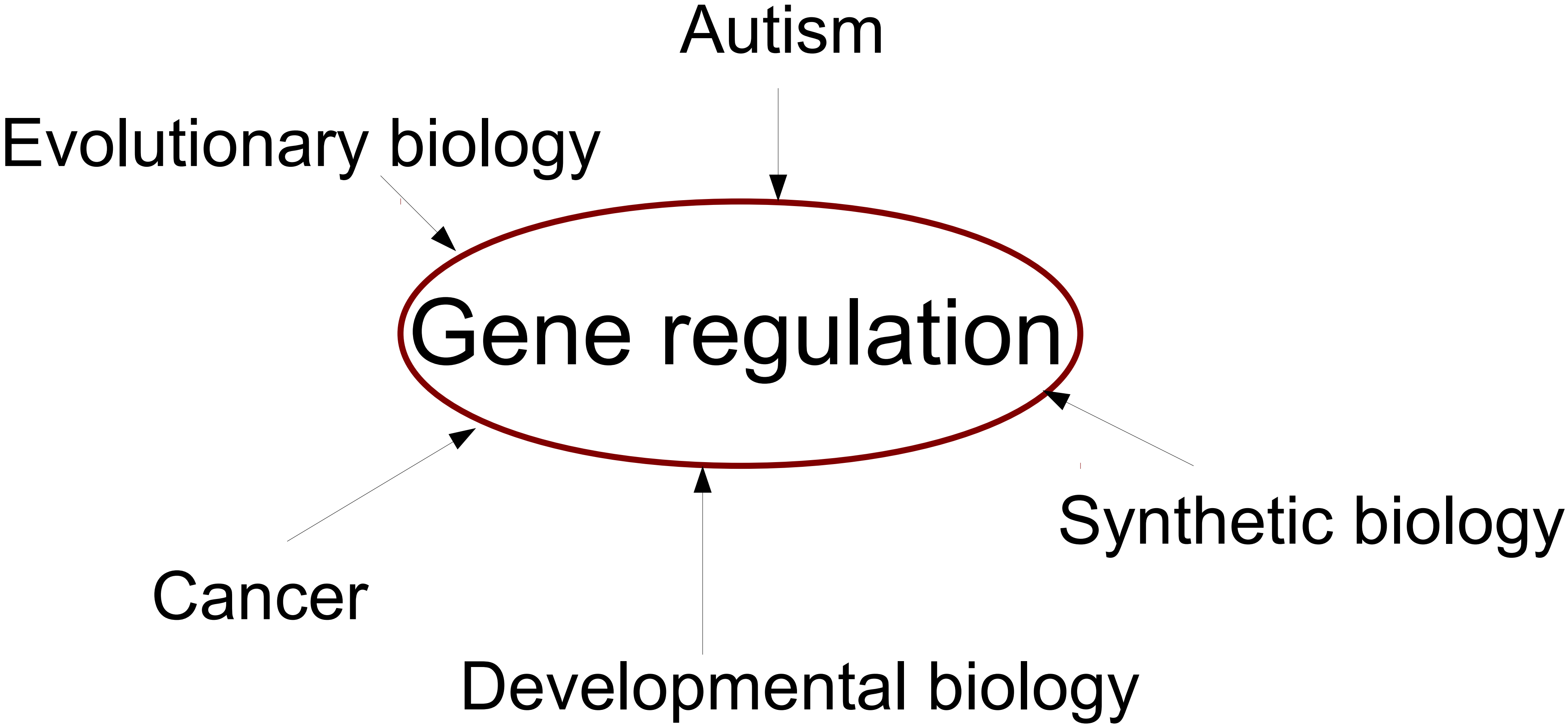
Models of distal enhancers of inducible gene expression

Martin Hemberg

Channing Division of Network Medicine
November 30, 2012

Why is gene regulation important?



Gene regulation

Autism

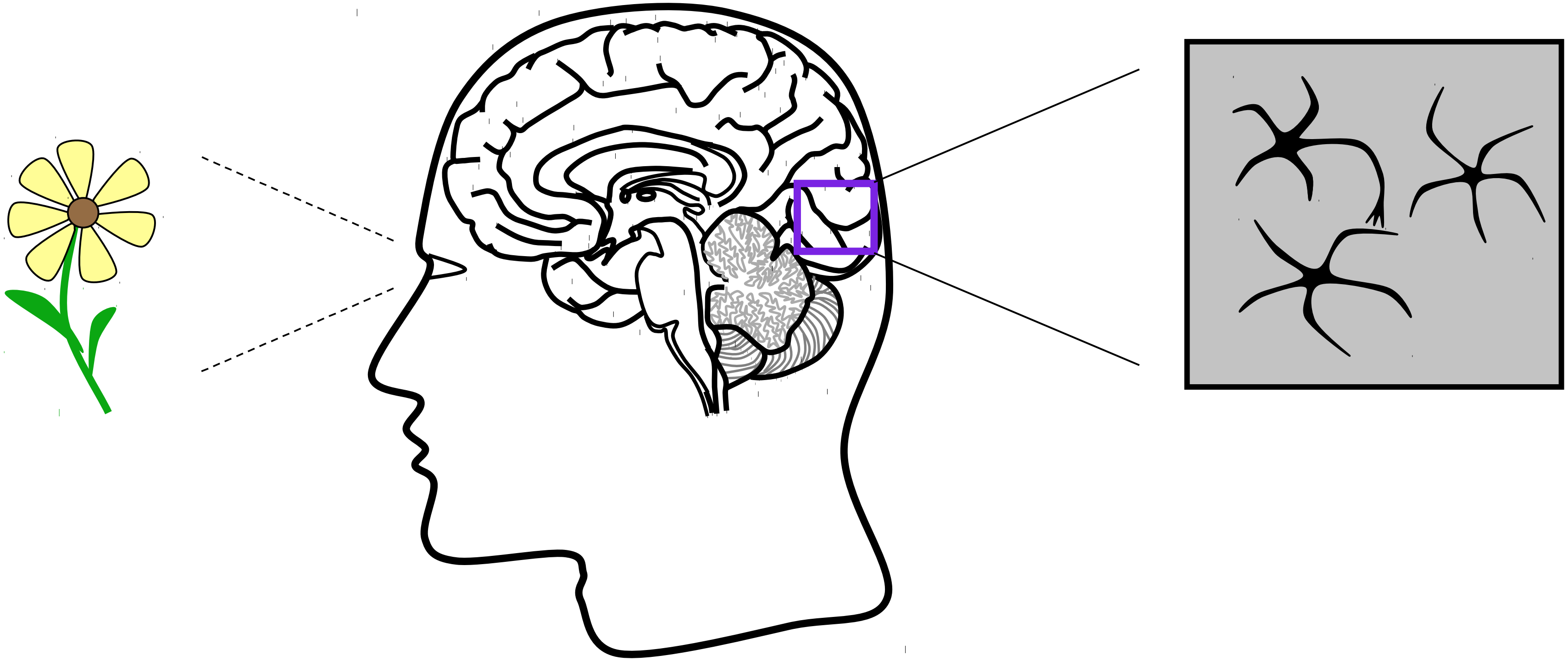
Evolutionary biology

Synthetic biology

Cancer

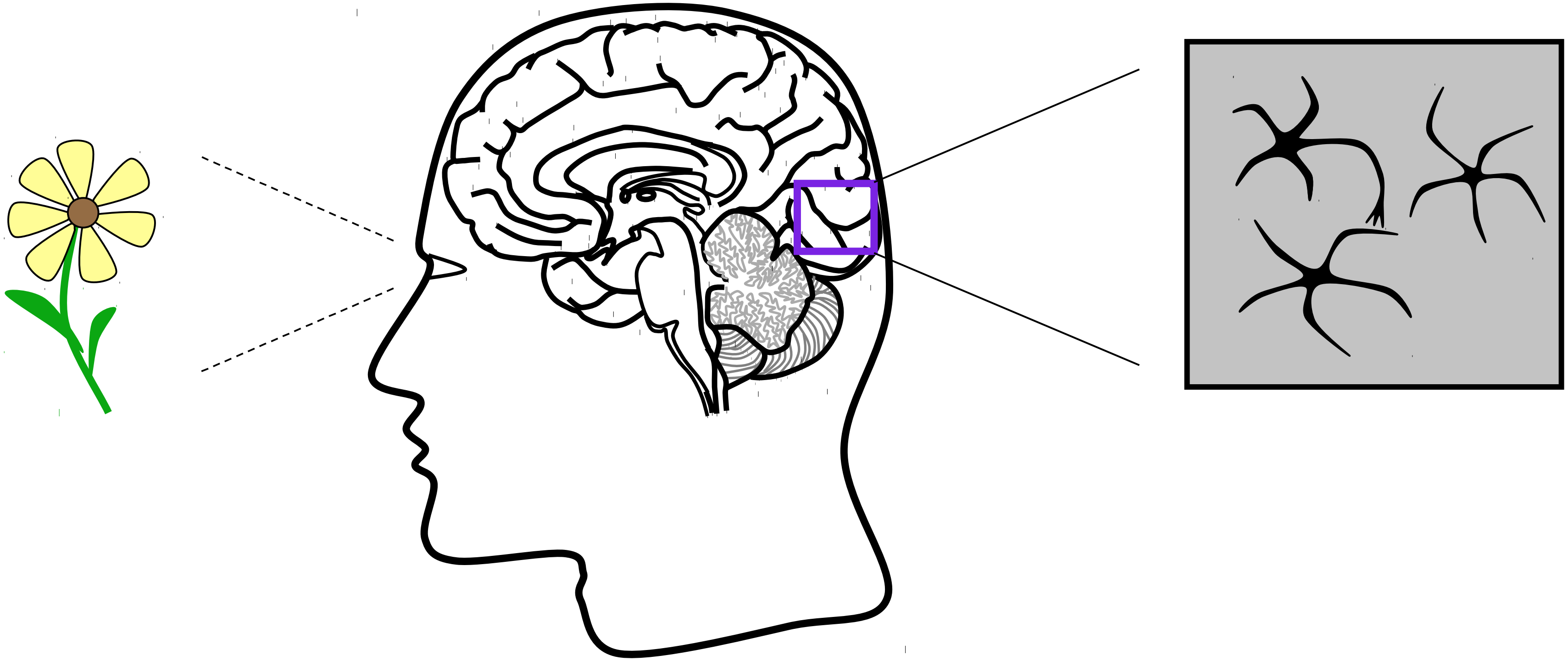
Developmental biology

Synapses change in response to external environmental stimuli



Hubel & Wiesel, 1970’s

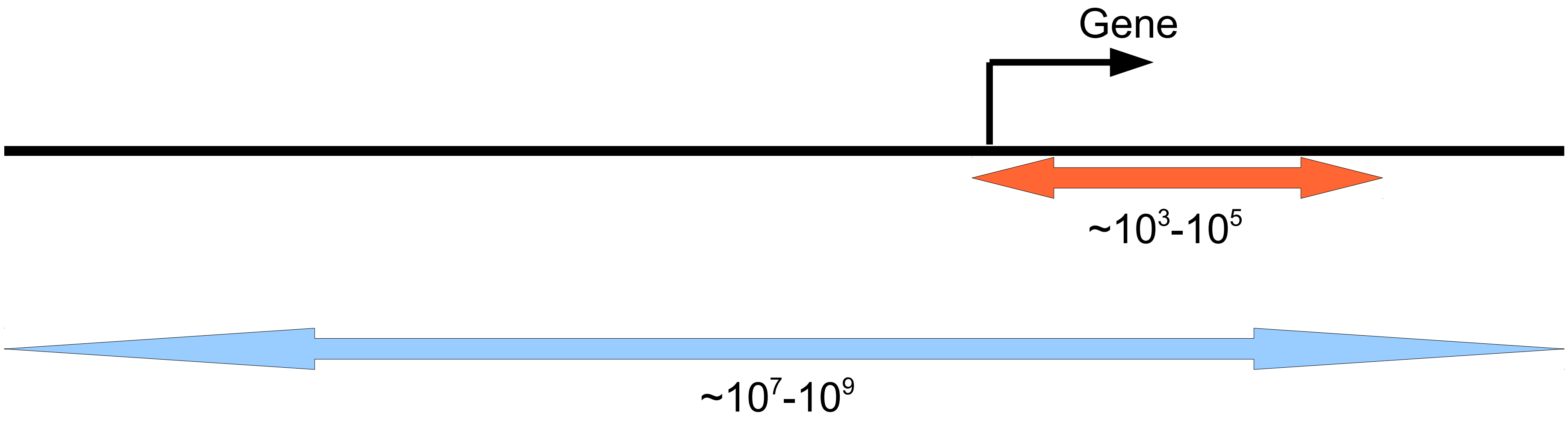
Activity dependent gene expression triggered by influx of Calcium ions



Hubel & Wiesel, 1970’s

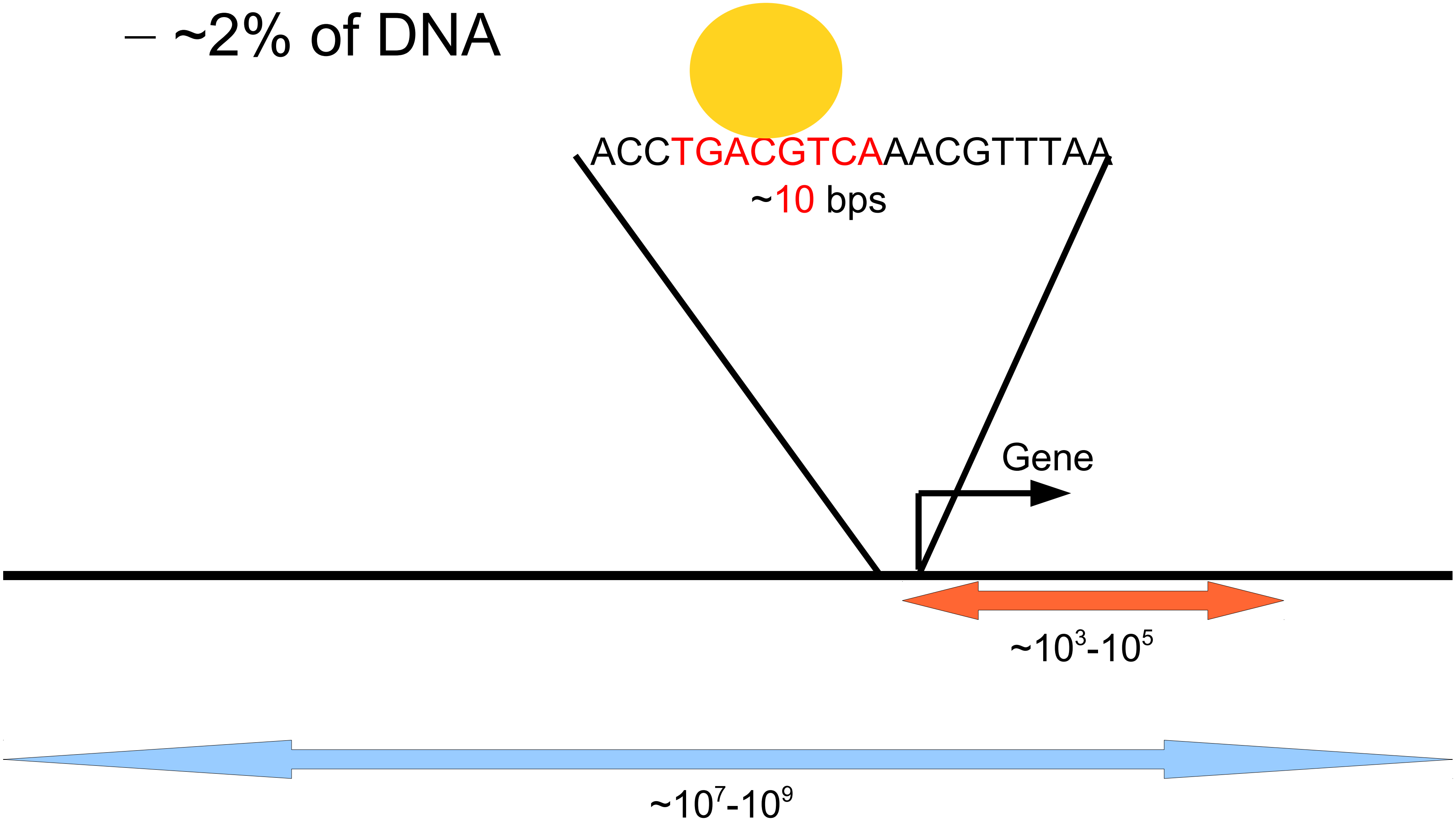
Mouse genome has ~3 billion bps

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



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

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



ACCTGACGTCAAACGTTTAA

~10 bps

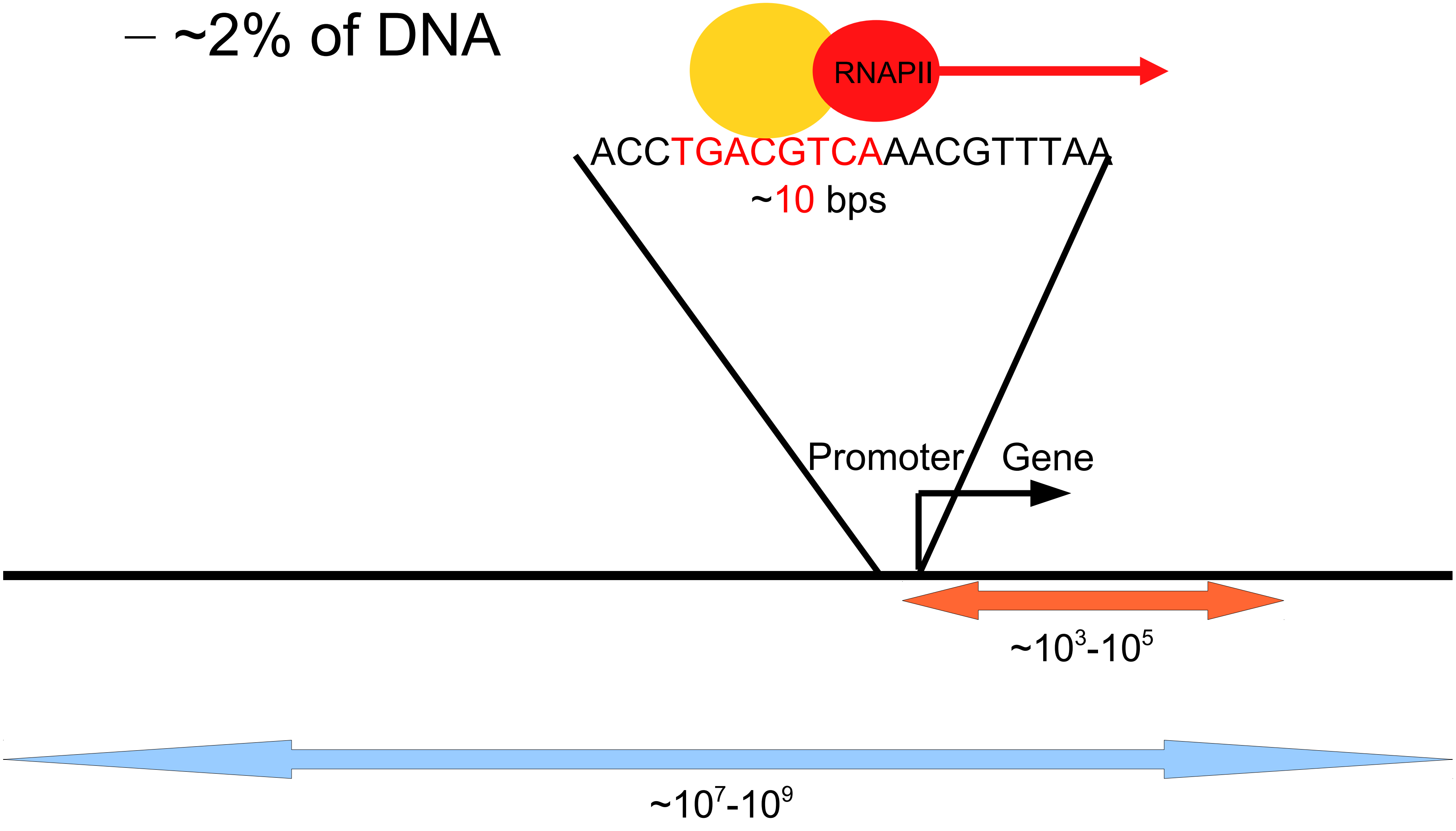
Gene

~10^3-10^5

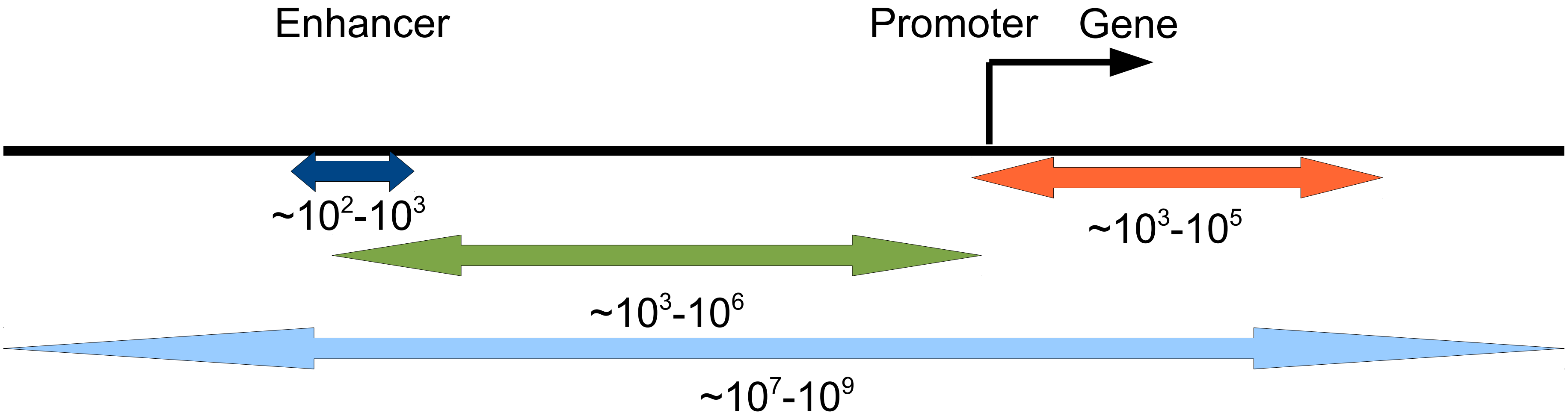
~10^7-10^9

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

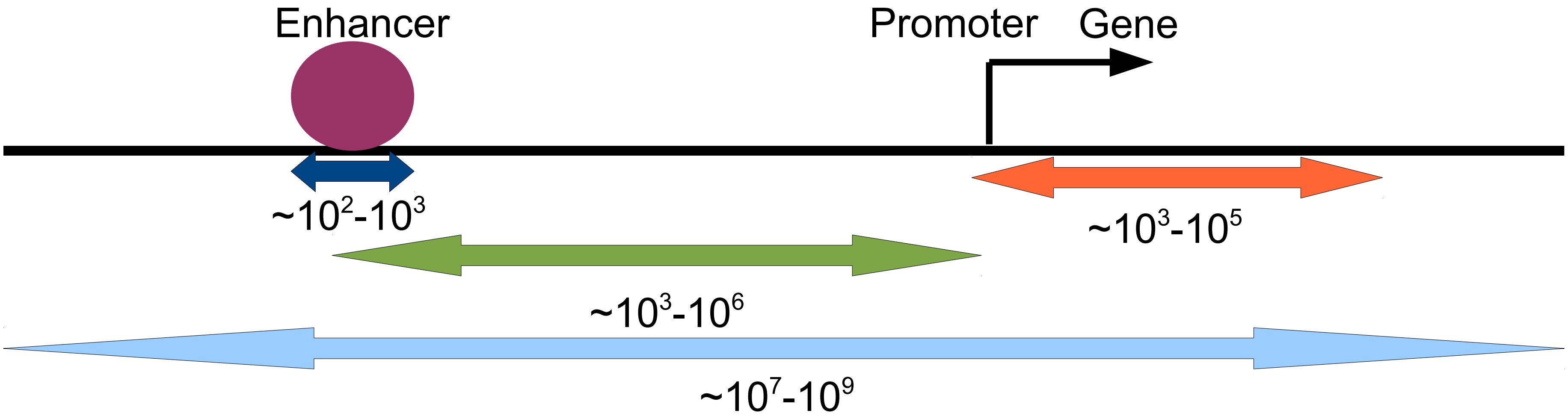
- ~25,000 genes
  - ~2% of DNA



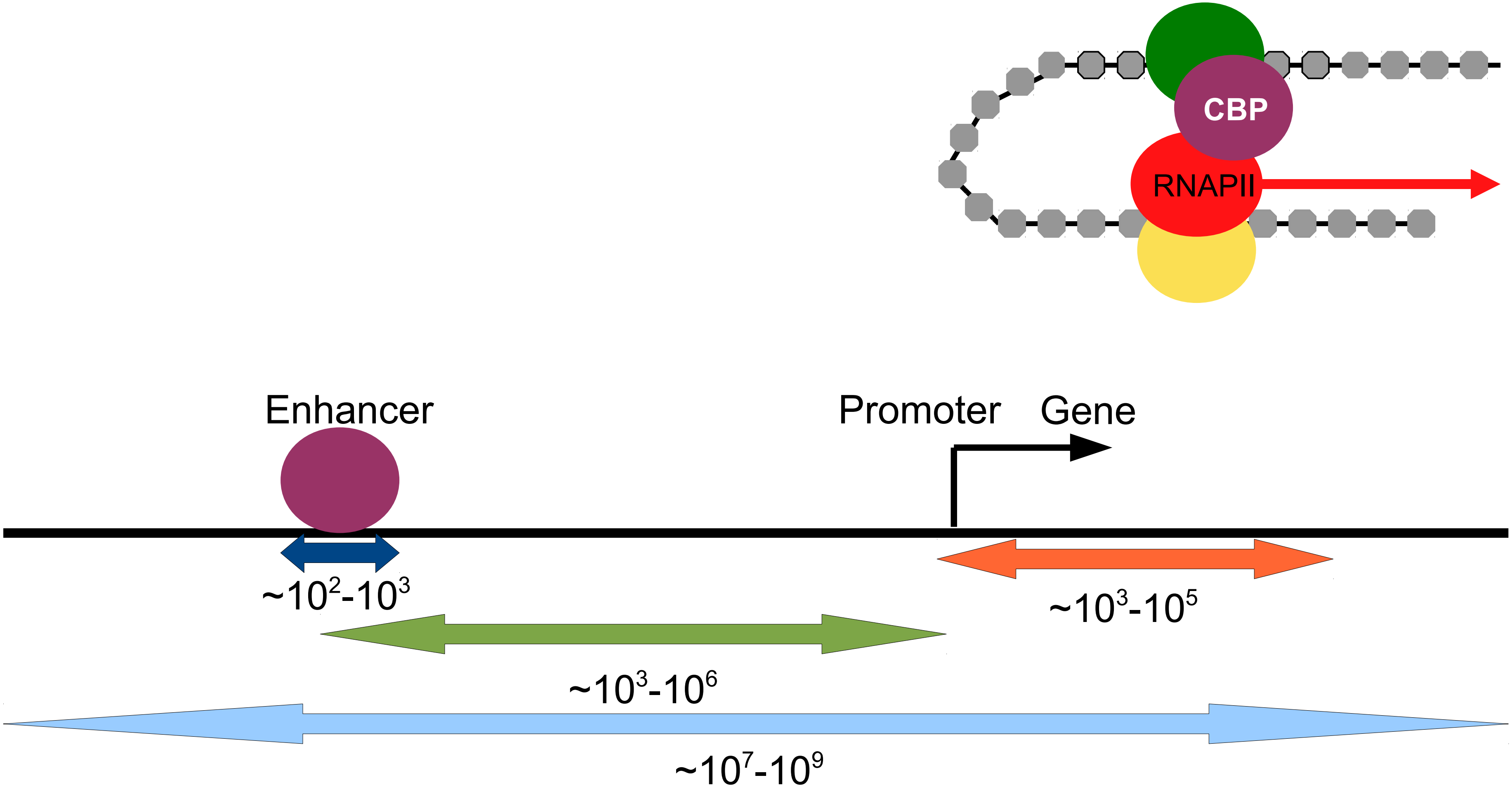
Enhancers are distal regulatory sequences



Enhancers characterized by **CBP** binding



The mechanism by which enhancers increase expression are poorly understood



CBP

RNAPII

Enhancer

Promoter

Gene

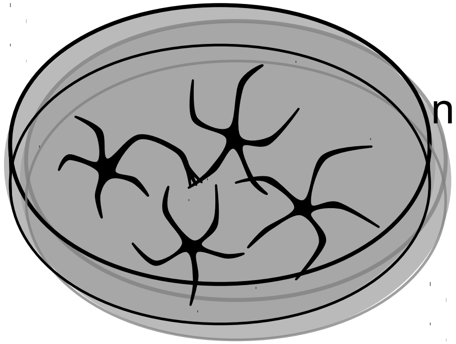
$\sim 10^2$-10^3

$\sim 10^3$-10^5

$\sim 10^3$-10^6

$\sim 10^7$-10^9

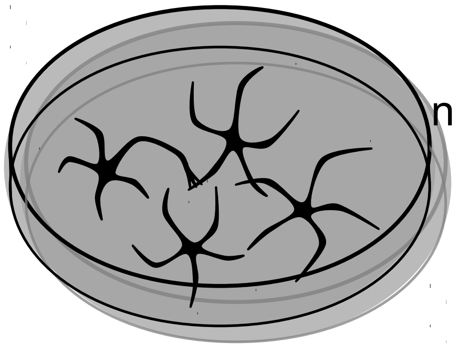
Cultured mouse cortical neurons for genome-wide study of activity dependent gene expression



mouse cortical neurons

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

Genome-wide data obtained using high-throughput sequencing



mouse cortical
neurons

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

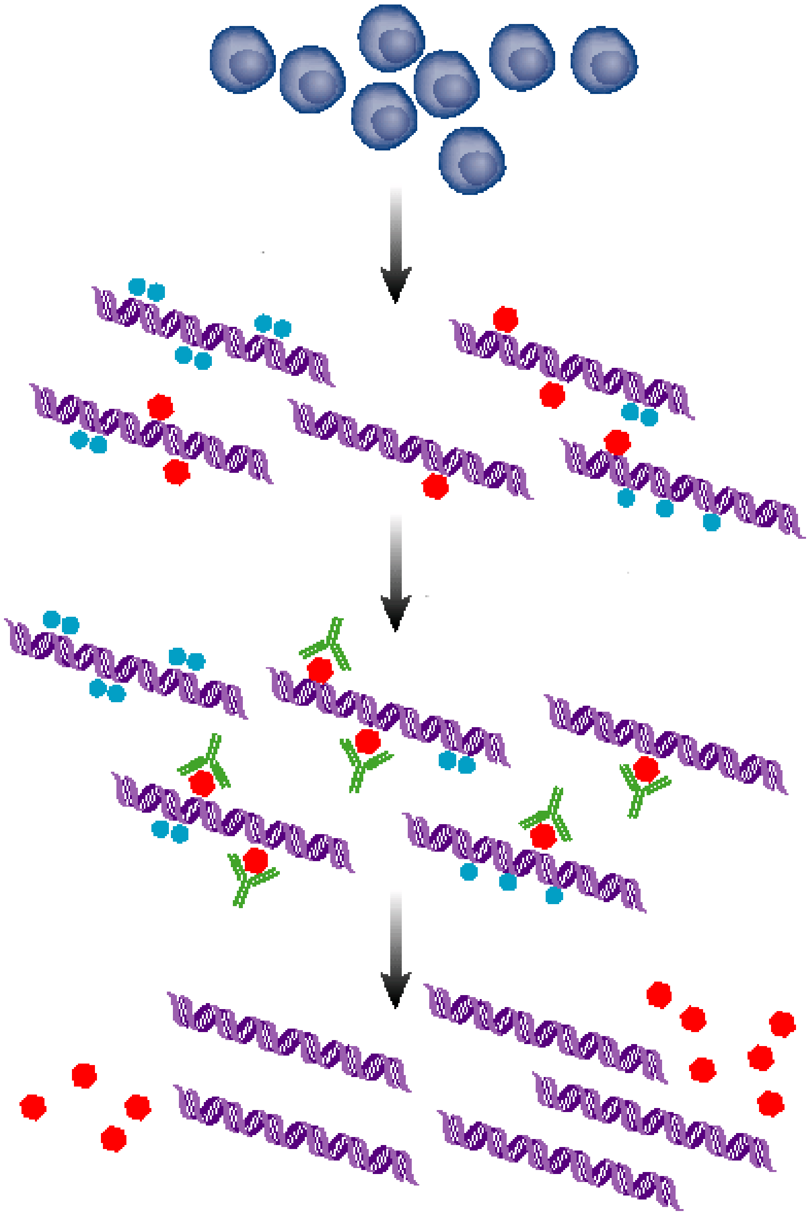
- KCl → + KCl

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

Jesse Gray
Tae-Kyung Kim
Greenberg Lab

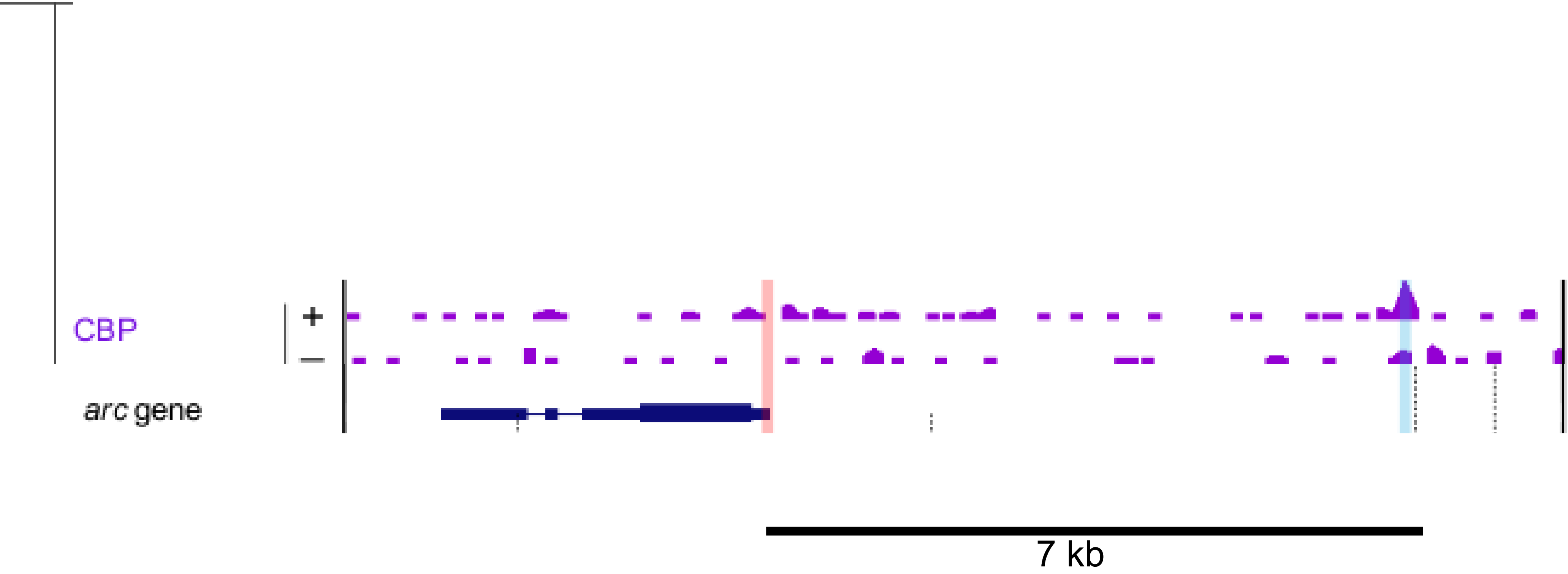
Chromatin immunoprecipitation and sequencing (**ChIP-Seq**) finds protein binding sites *in vivo*

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

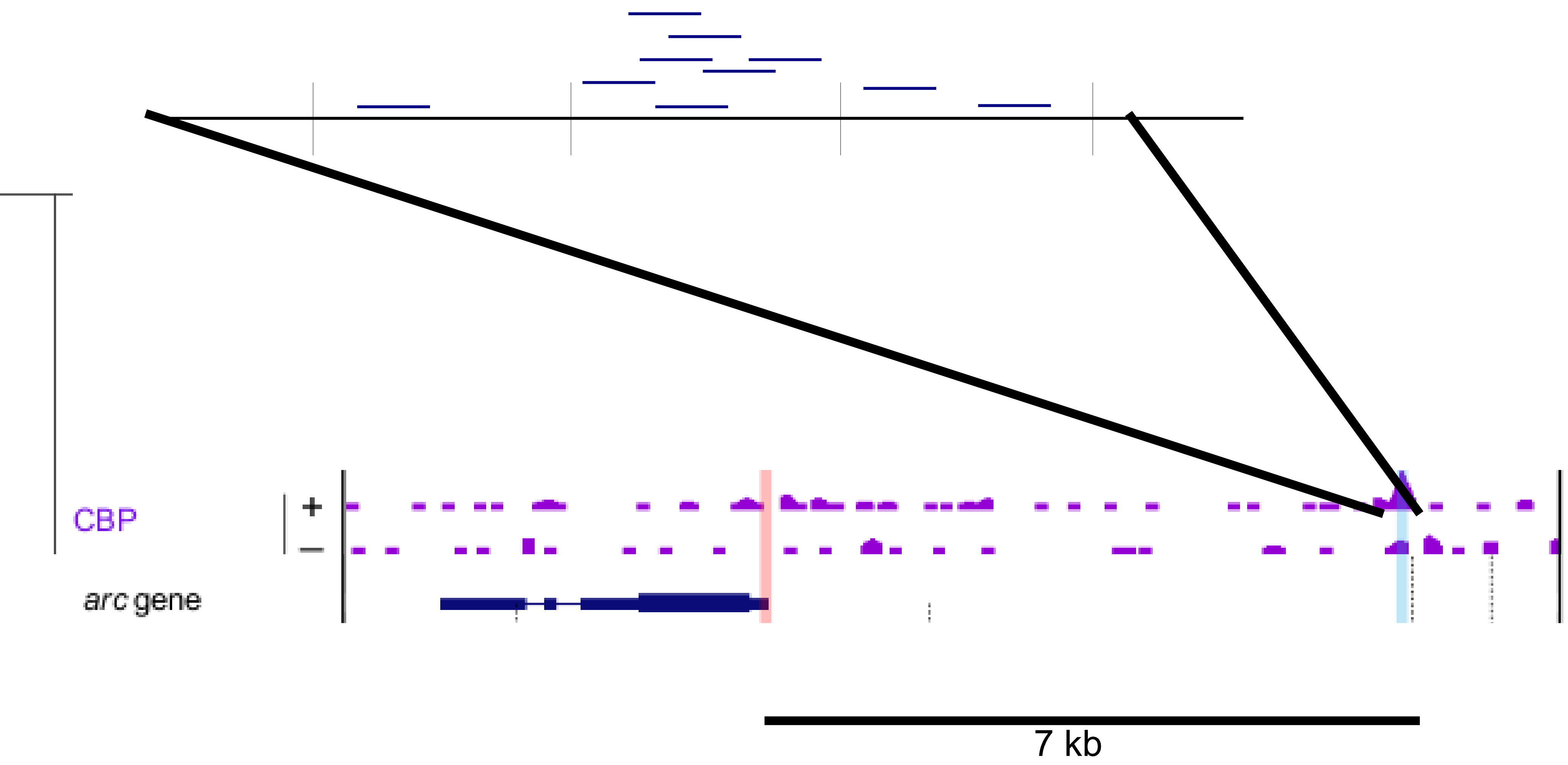


(Mardis, 2007)

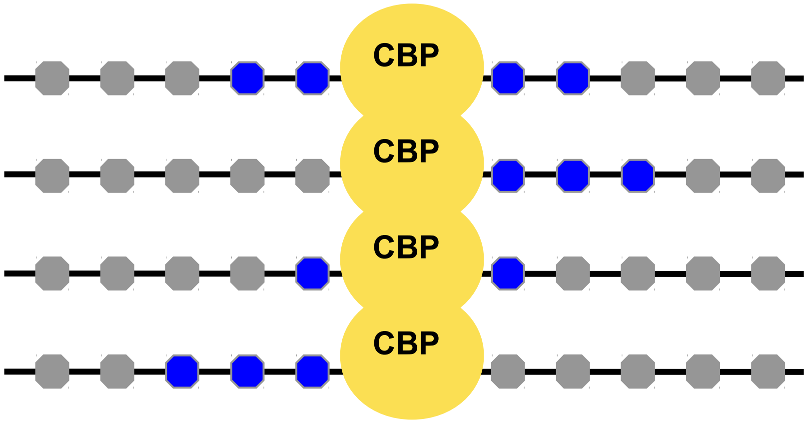
Inducible CBP binding at enhancers



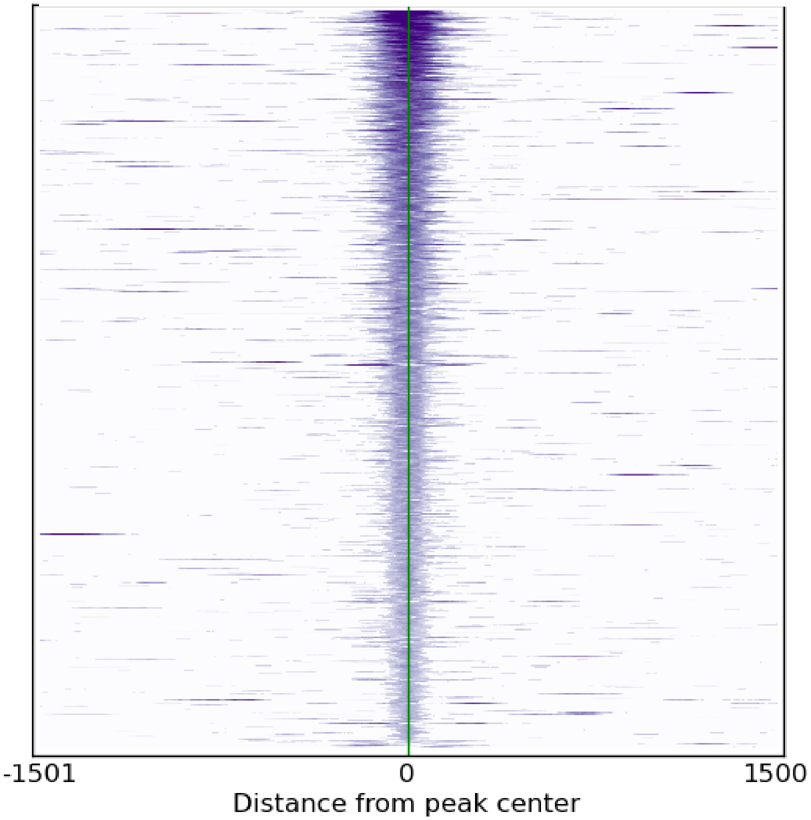
Peak-calling algorithm identifies ~28,000 CBP binding sites in two replicates

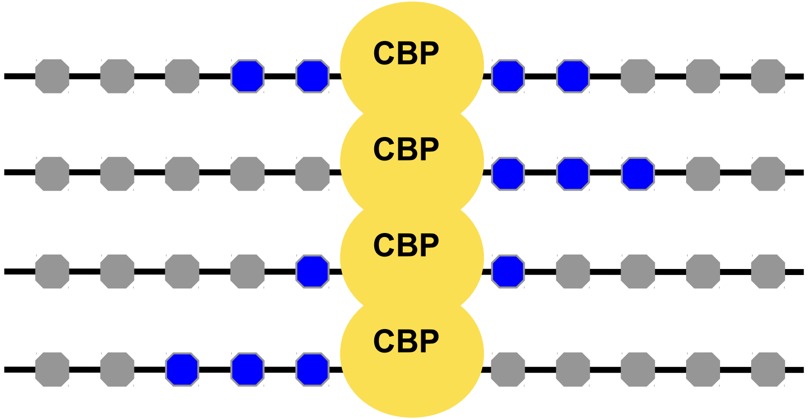


Align CBP peaks to obtain binding profiles

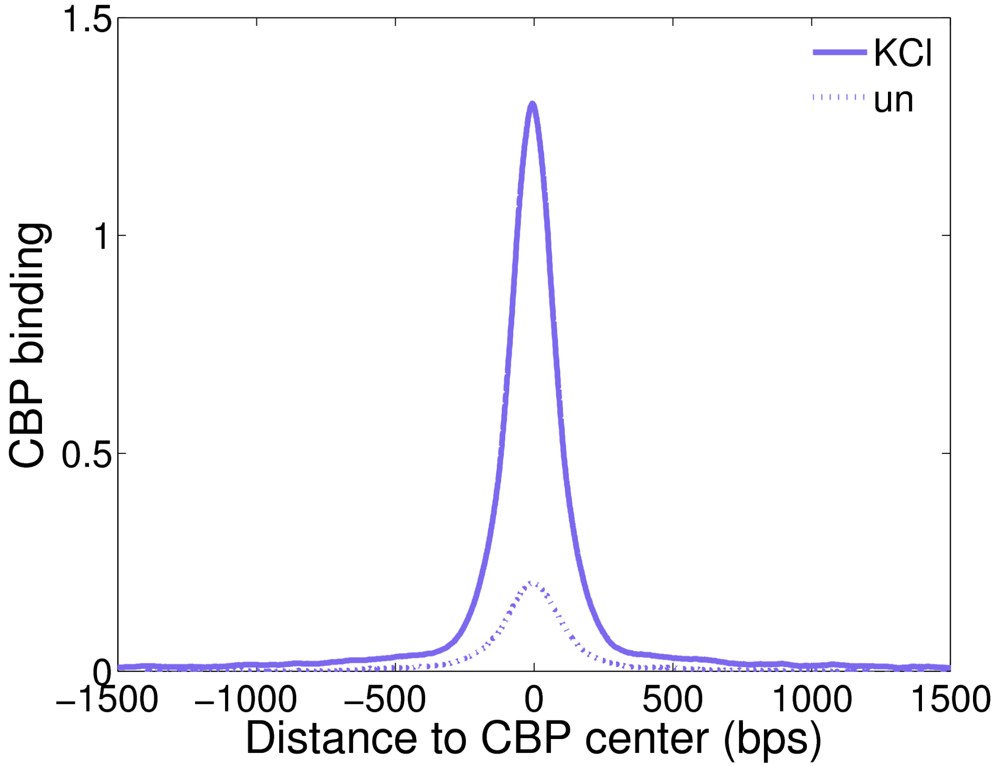


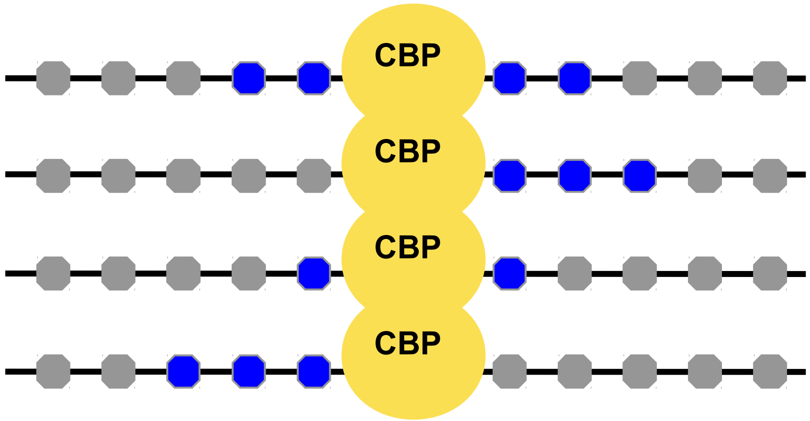
Align CBP peaks to obtain binding profiles



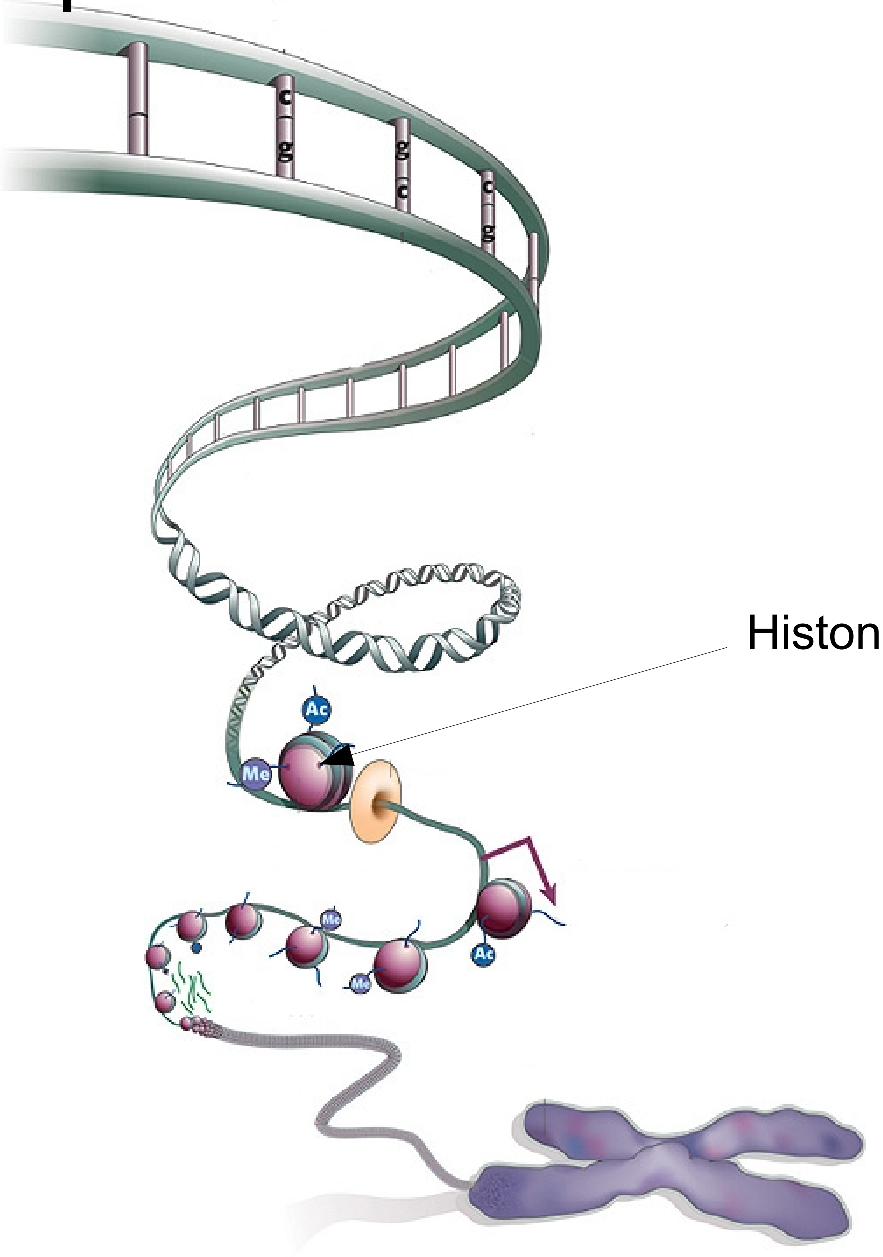


Average profile of CBP binding





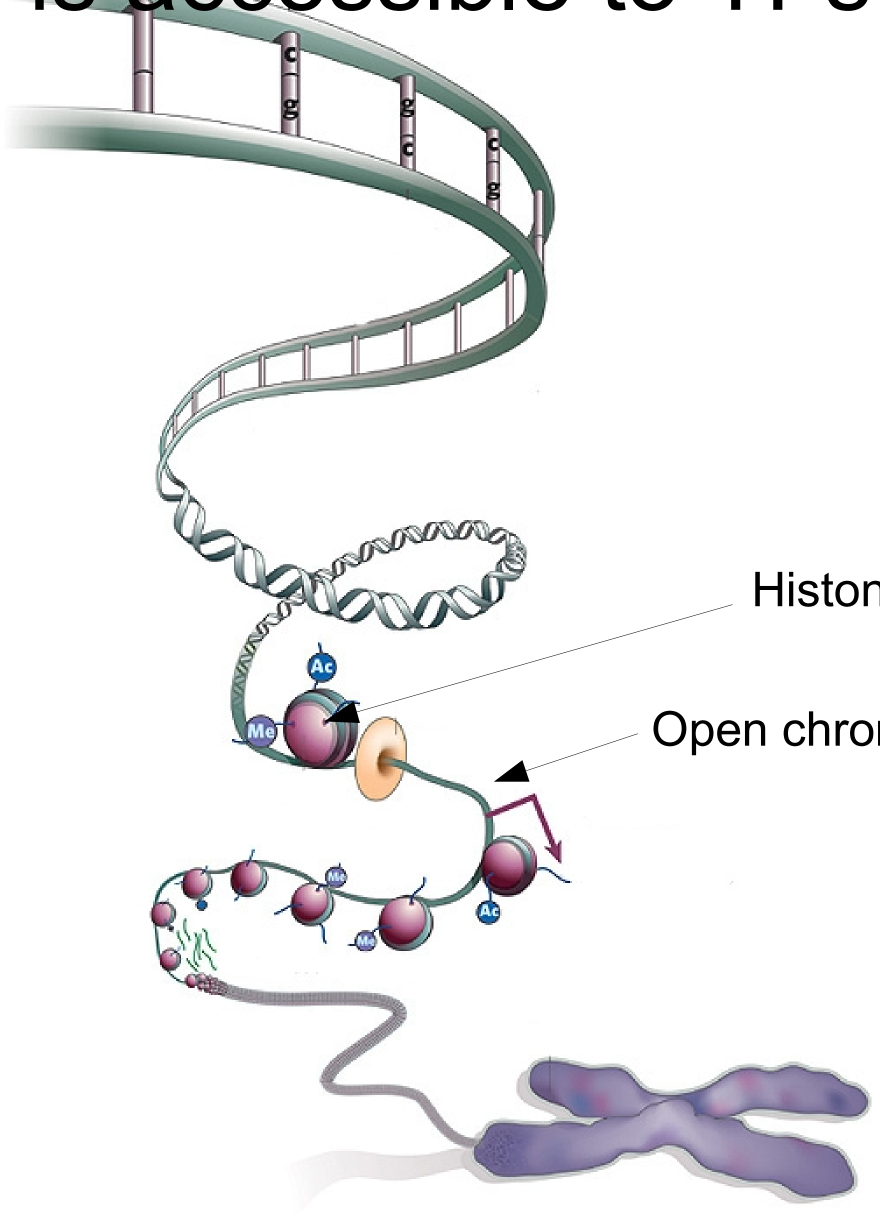
Histones prevent transcription factors from binding to DNA



Histone

(ENCODE, 2007)

Only 1% of the genome is accessible to TFs



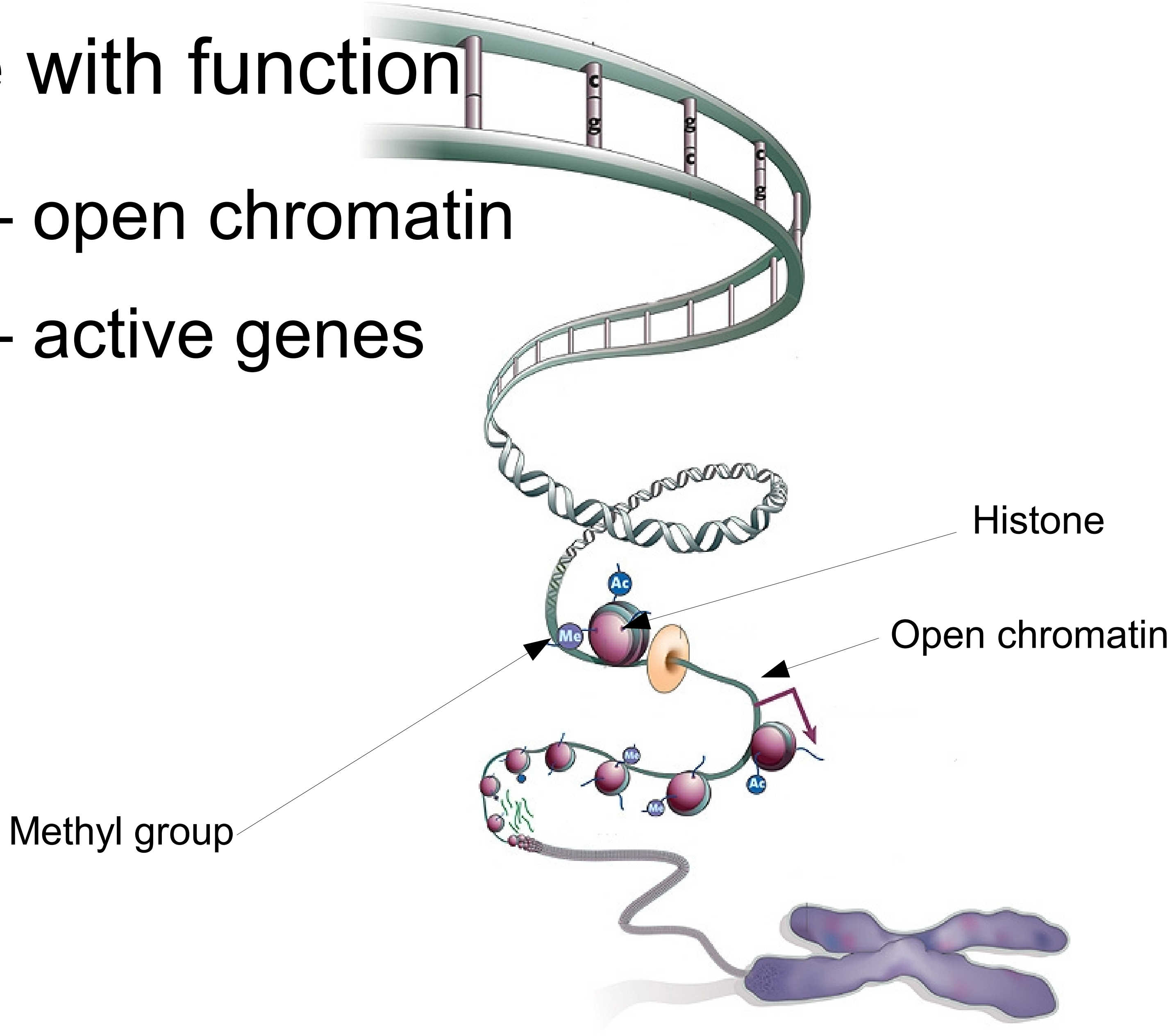
Histone

Open chromatin

(ENCODE, 2007)

Post-translational modifications of histone tails correlate with function

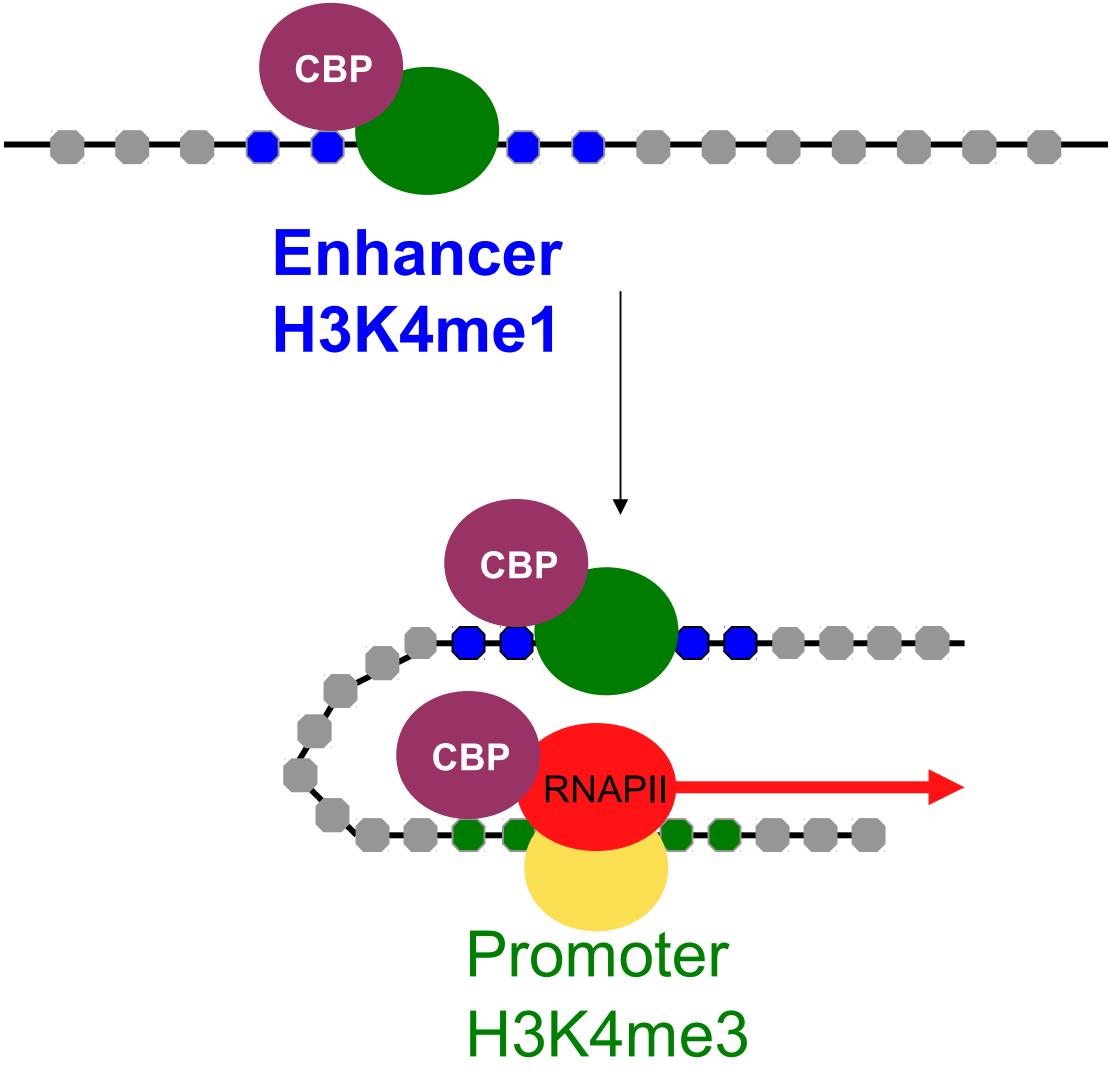
- **H3K4me1** – open chromatin
- **H3K4me3** – active genes



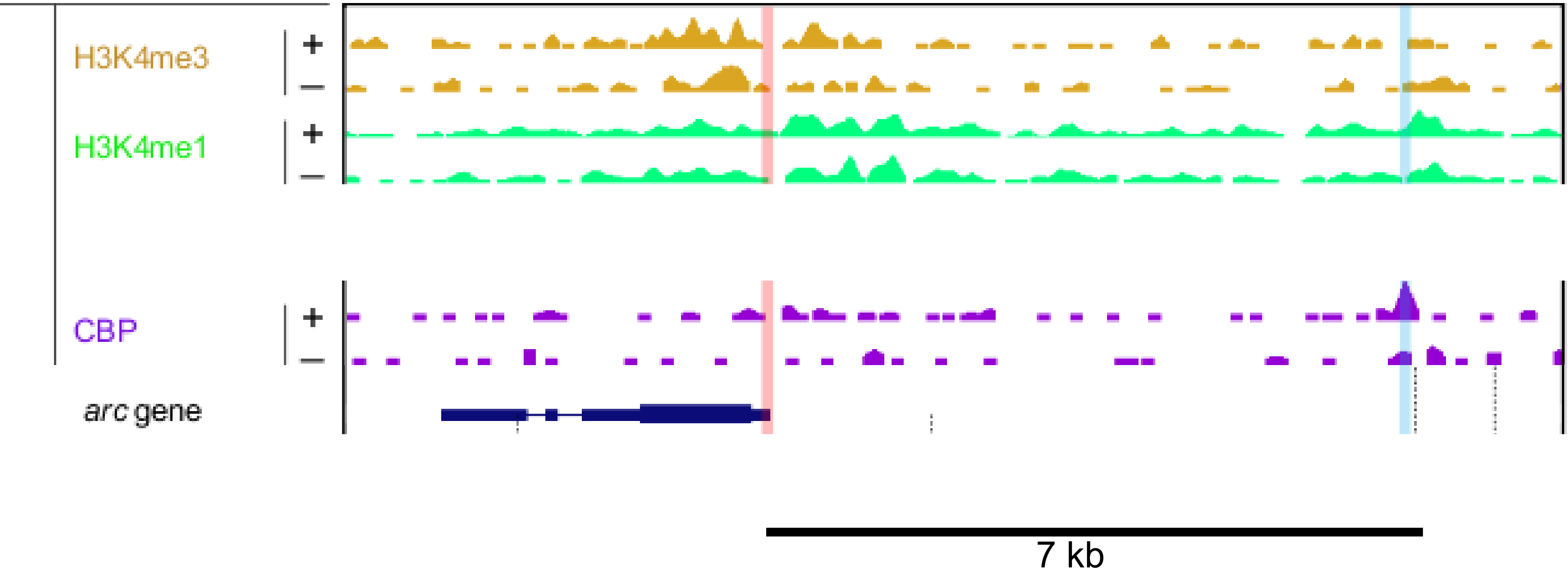
(ENCODE, 2007)

A combination of CBP and histone modifications identifies putative enhancers

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



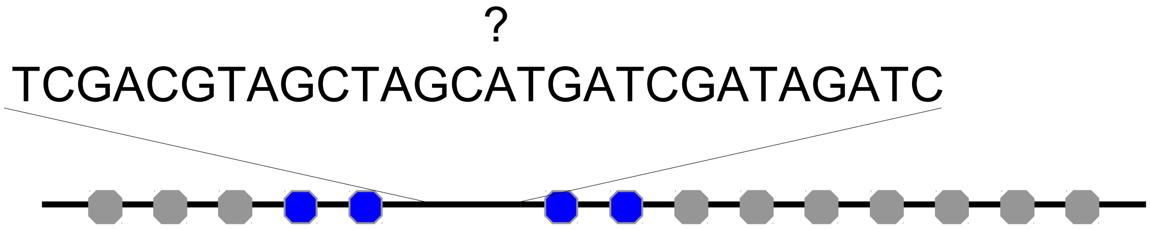
Distal CBP peaks have high levels of H3K4me1 and low levels of H3K4me3



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

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

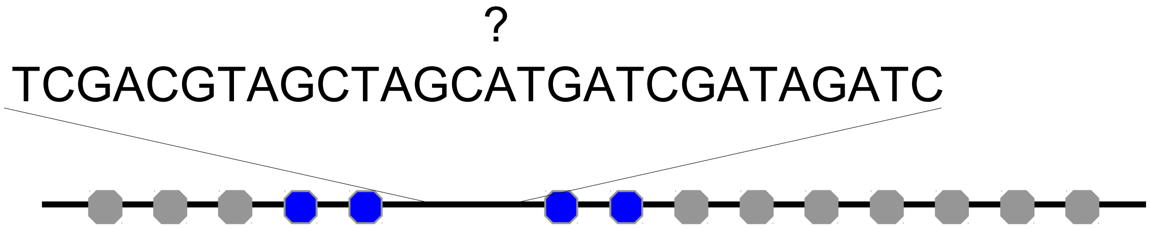
What TFs bind to enhancers?



Enhancer
H3K4me1

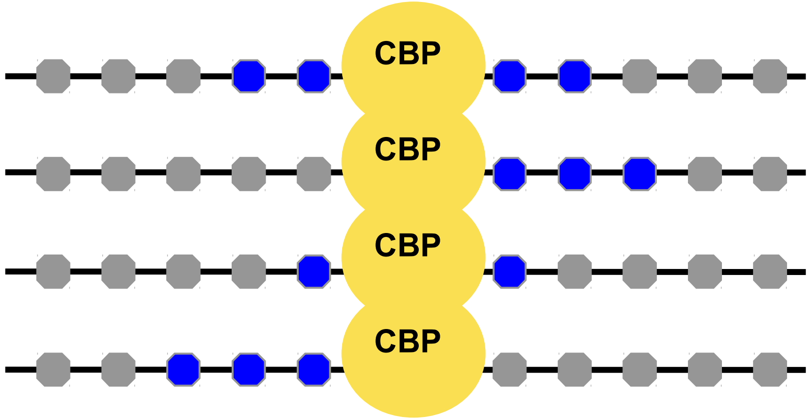
- CBP -
  CREB Binding Protein
  - >50 partners

~100 enriched motifs at enhancers


?
TCGACGTAGCTAGCATGATCGATAGATC

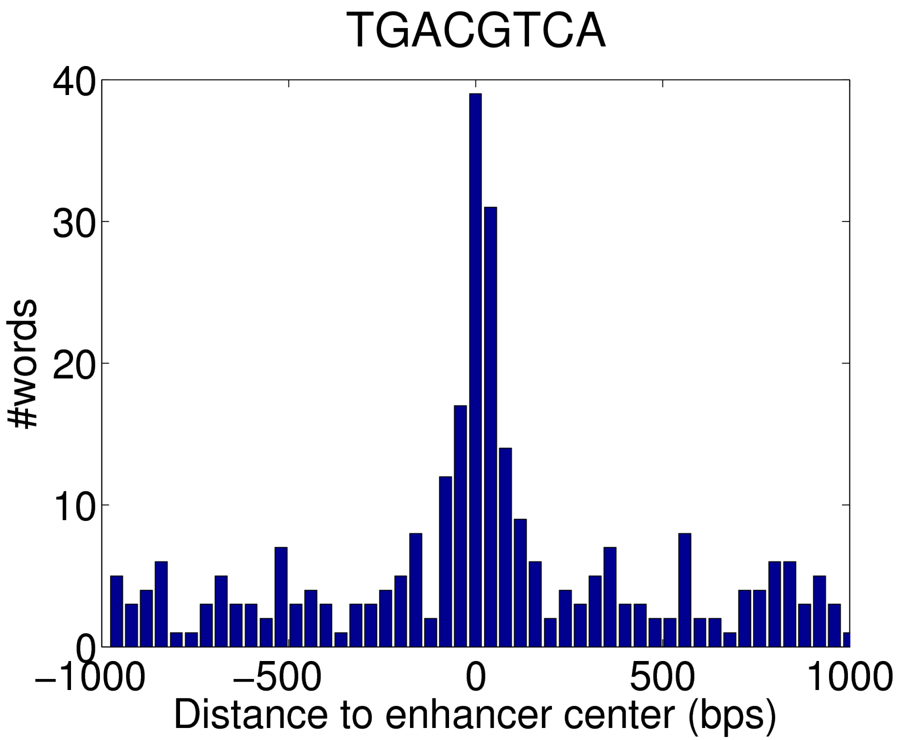
Enhancer
H3K4me1

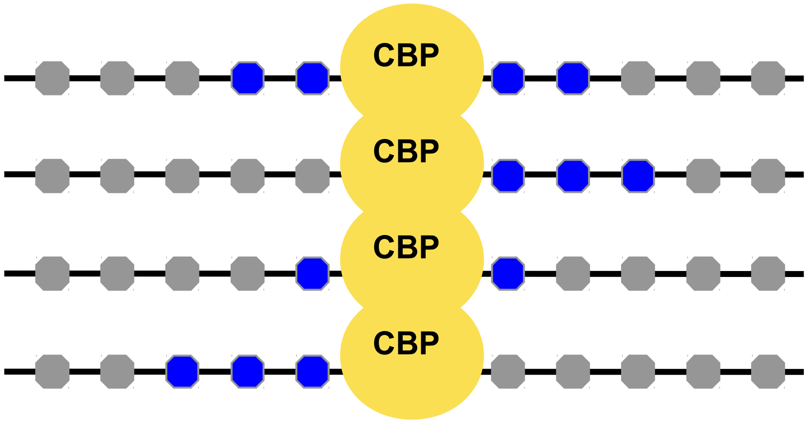
TCAGGCTGATGACGTCAAACCGTCGTTA
ACCTTTTGACGTCAAATTTACGCTAGTAT
TCGACGTAGCTAGCATGATCGATAGATC
CGTGACGTCAGTGCTCGTAAATCATAAG



- CBP - CREB Binding Protein
  - >50 partners

Enrichment of the CRE motif

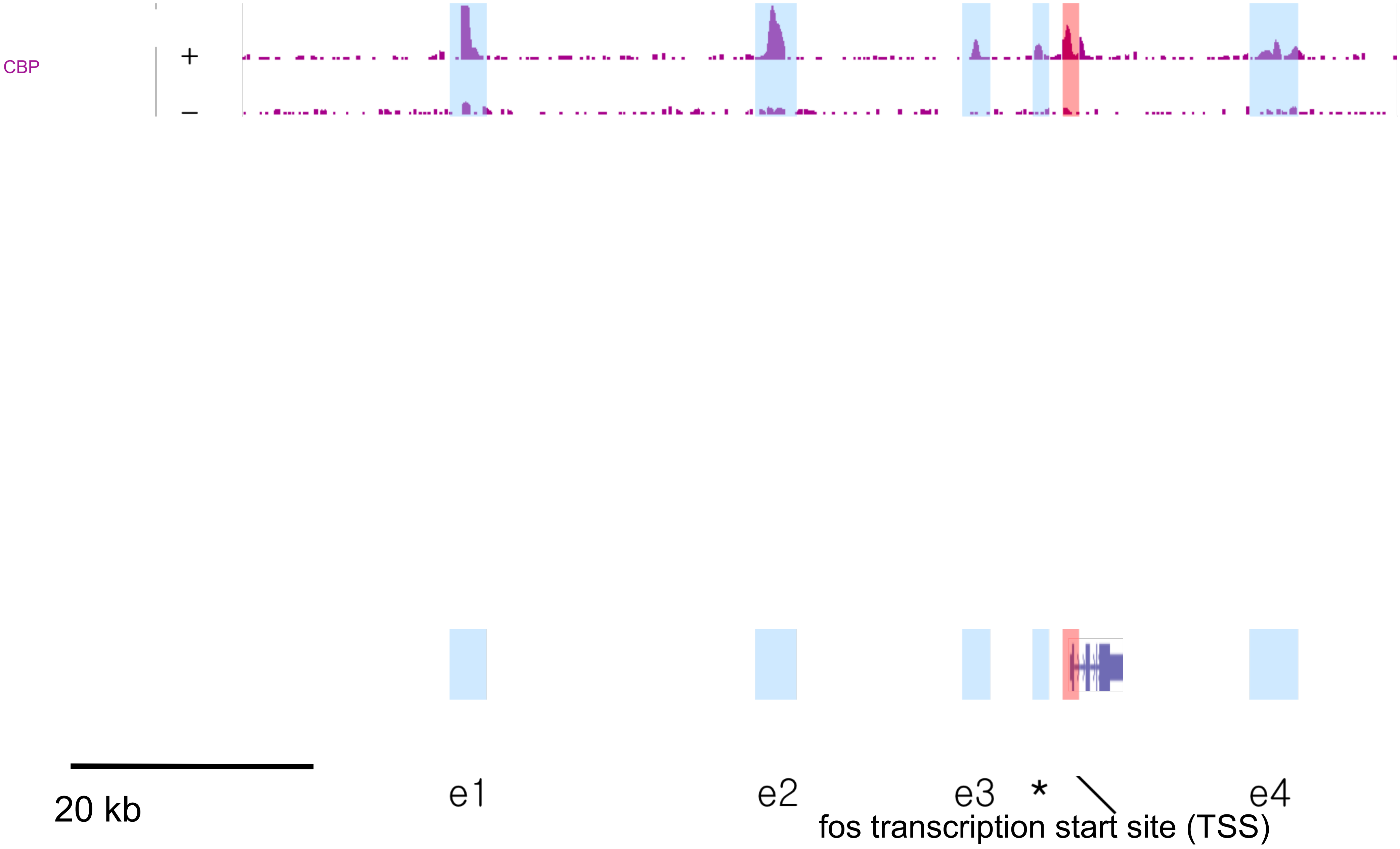




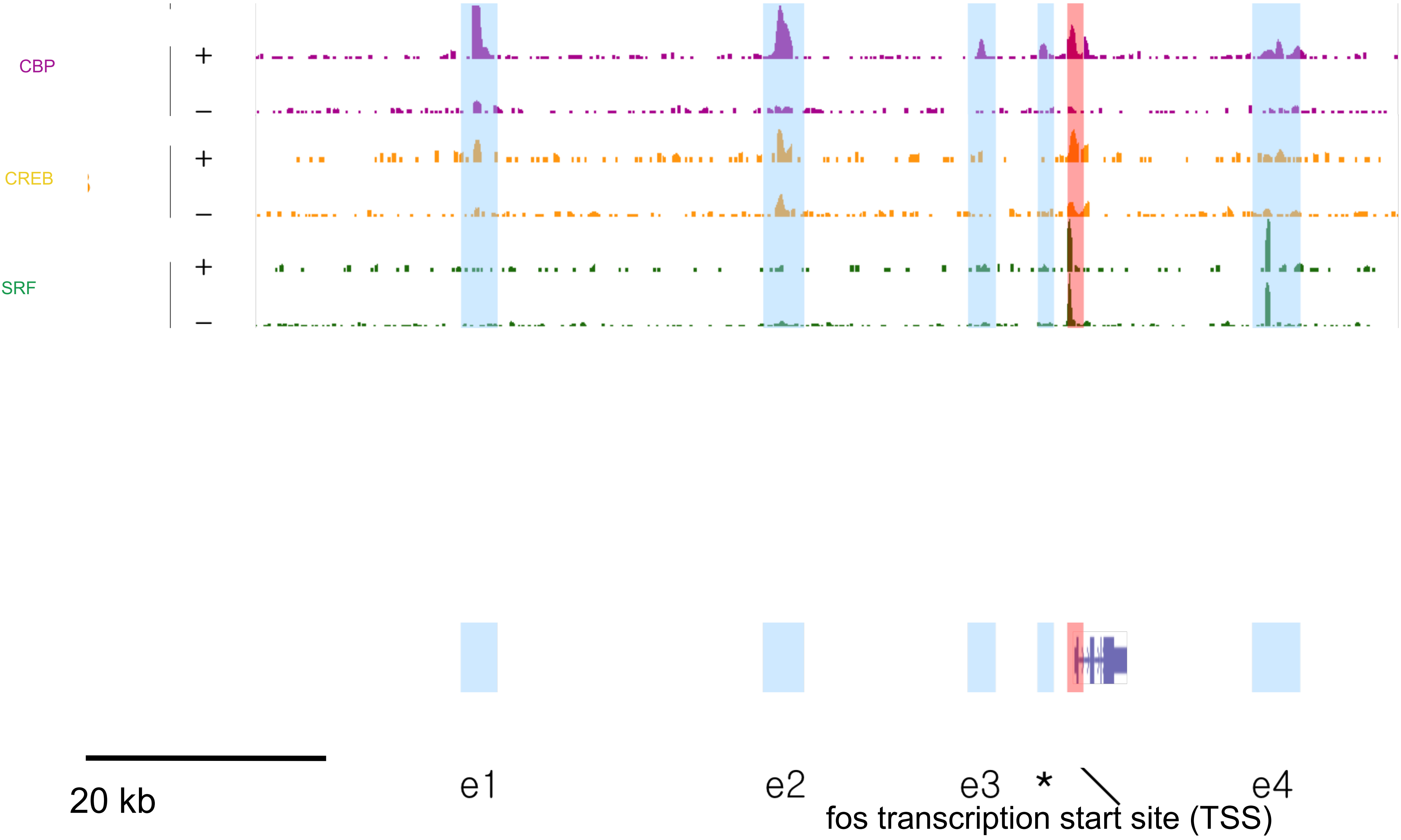
Motifs for several known TFs were identified

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

Several enhancers at the *fos* locus

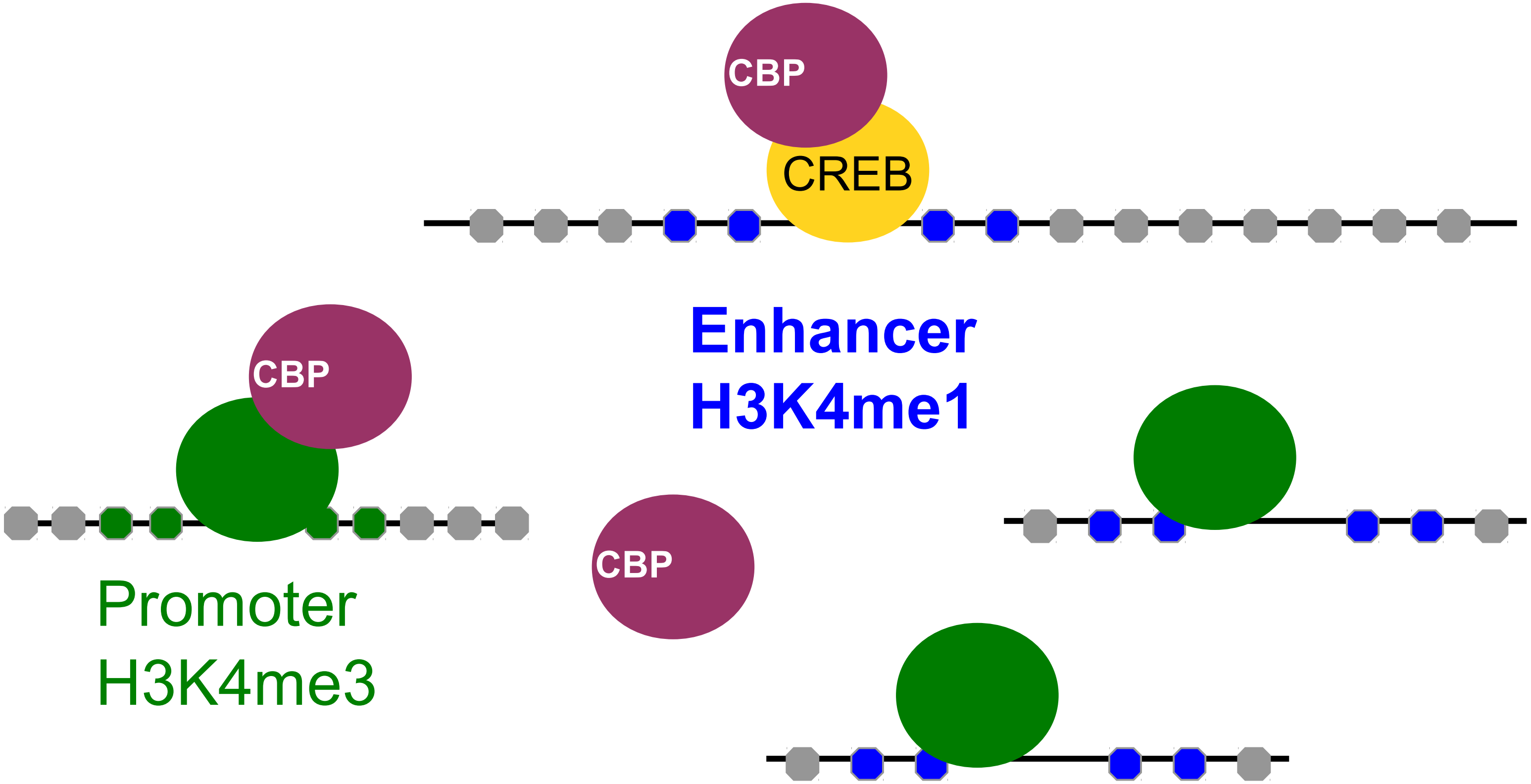


SRF and CREB binding at *fos* enhancers

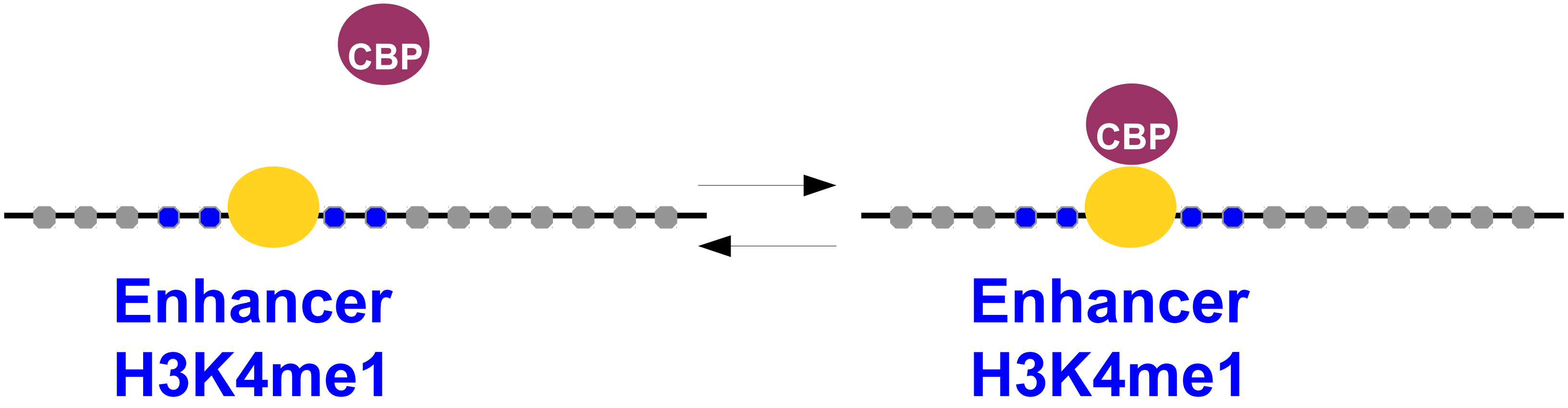


Is CBP binding determined by a combination of TFs?

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

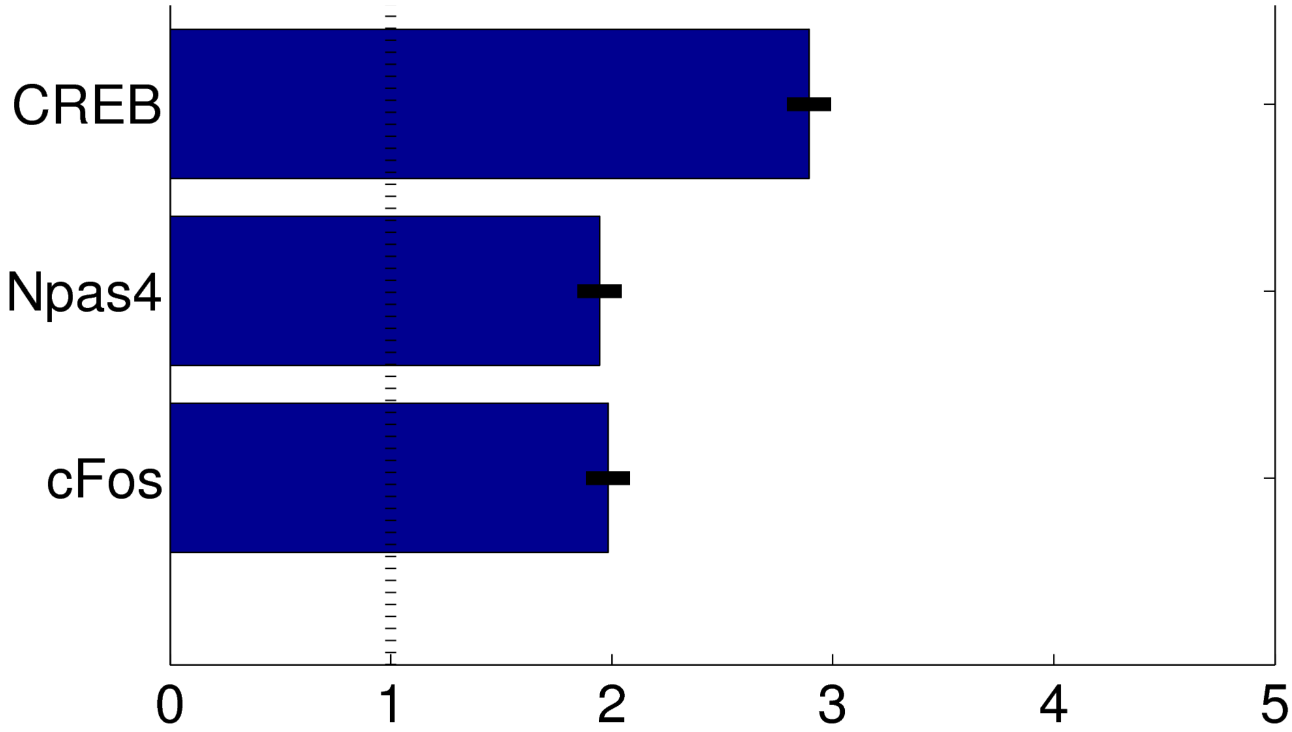


TFs compete for CBP



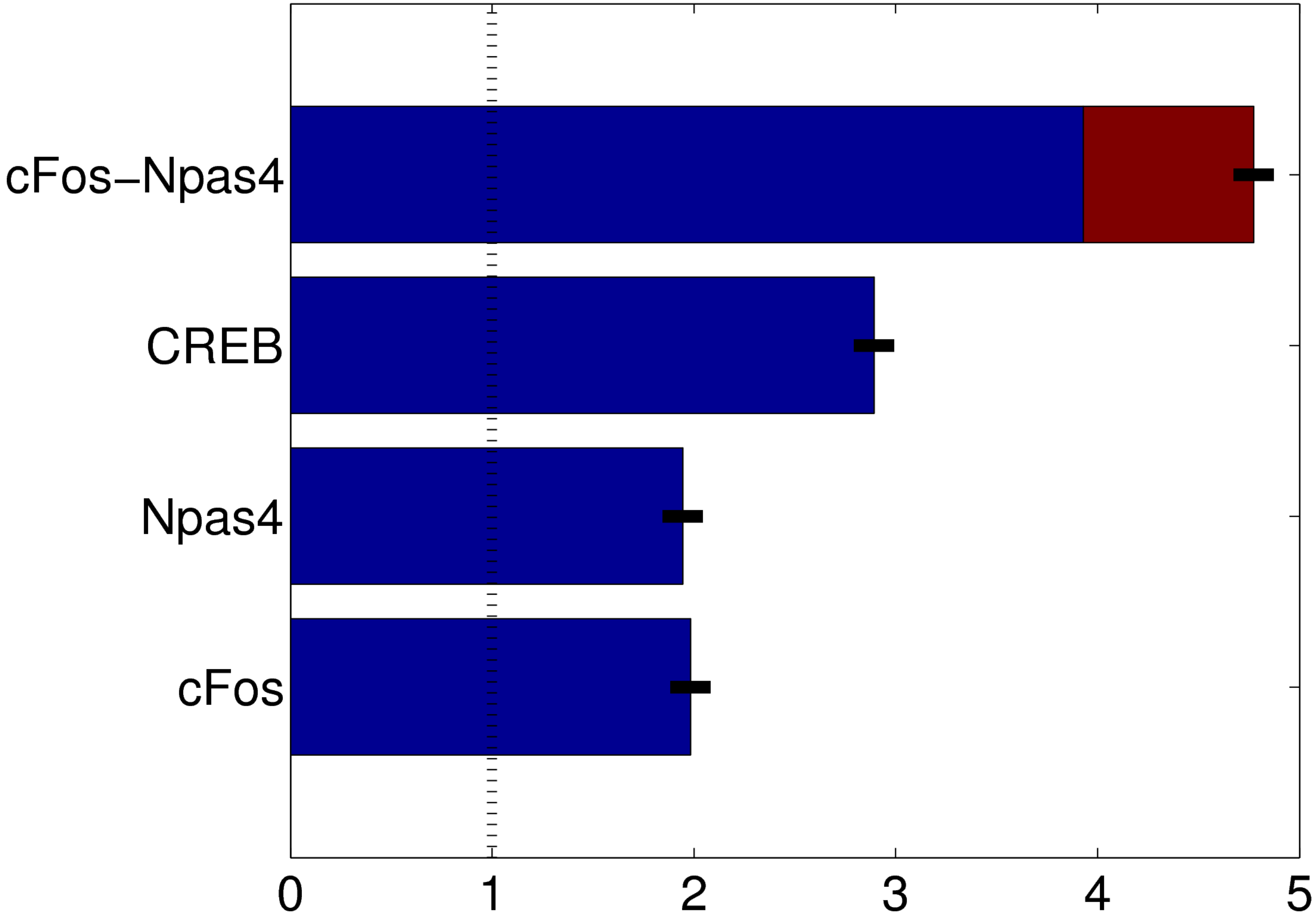
$$\text{CBP affinity} = \frac{\#\text{peaks with CBP}}{\#\text{peaks without CBP}}$$

Tfs have different affinities for CBP binding



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

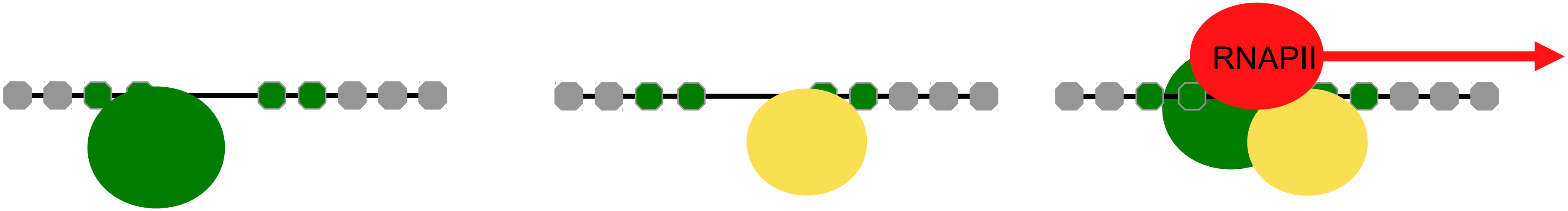
Synergistic effects for combinations of TFs



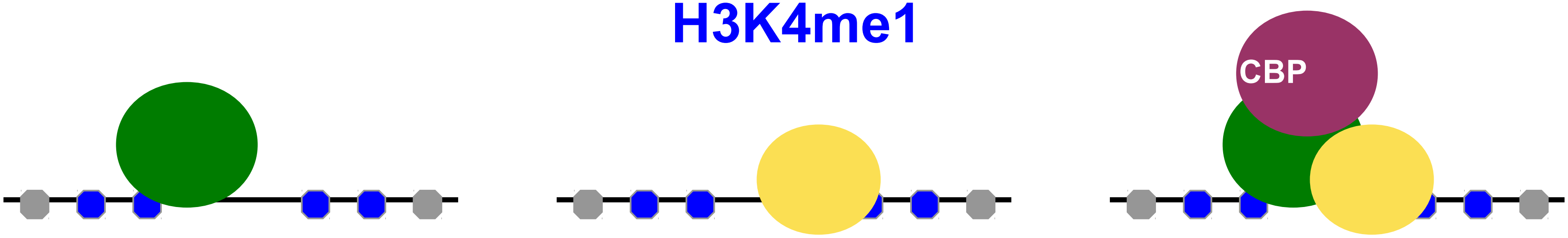
$$\text{CBP affinity} = \frac{\#\text{peaks with CBP}}{\#\text{peaks without CBP}}$$

Combinatorial code determines RNAPII levels at promoters and CBP at enhancers

Promoter
H3K4me3

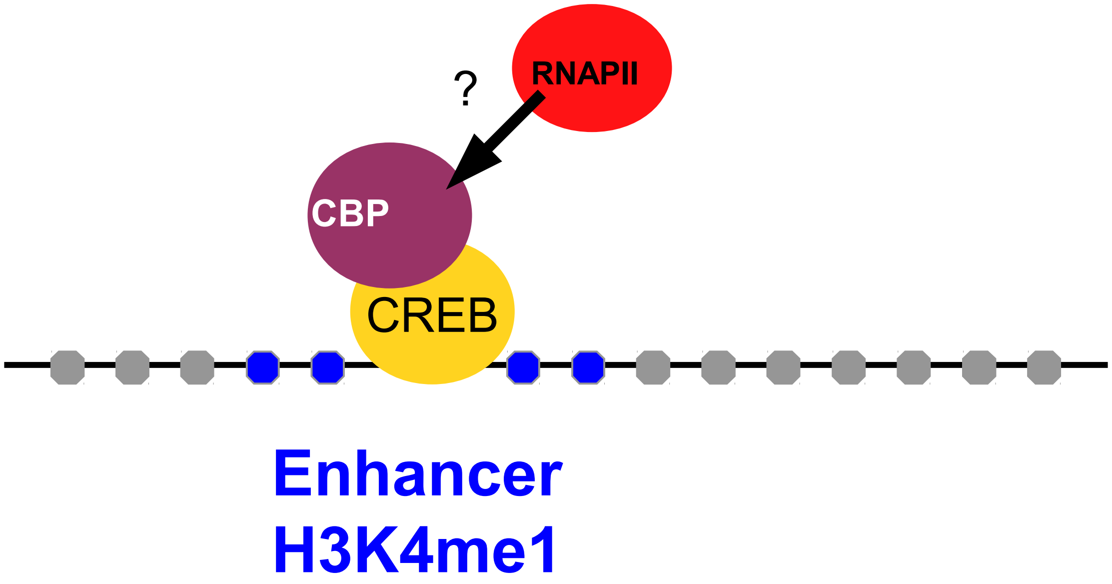


Enhancer
H3K4me1

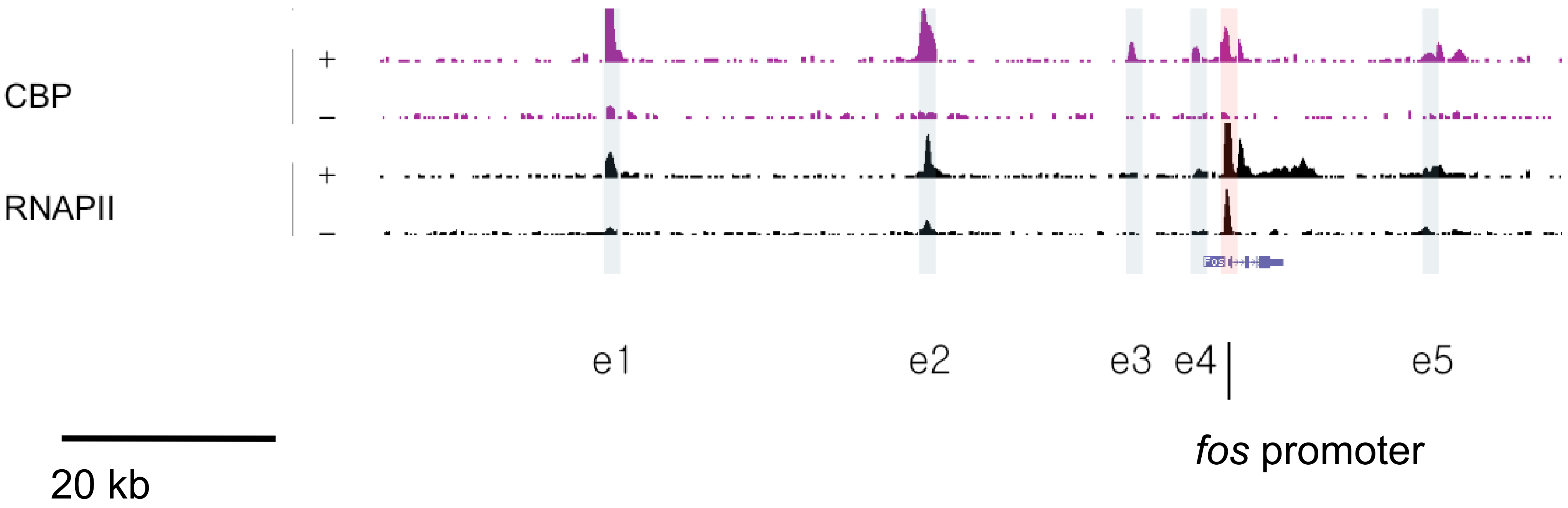


What is the function of CBP at enhancers?

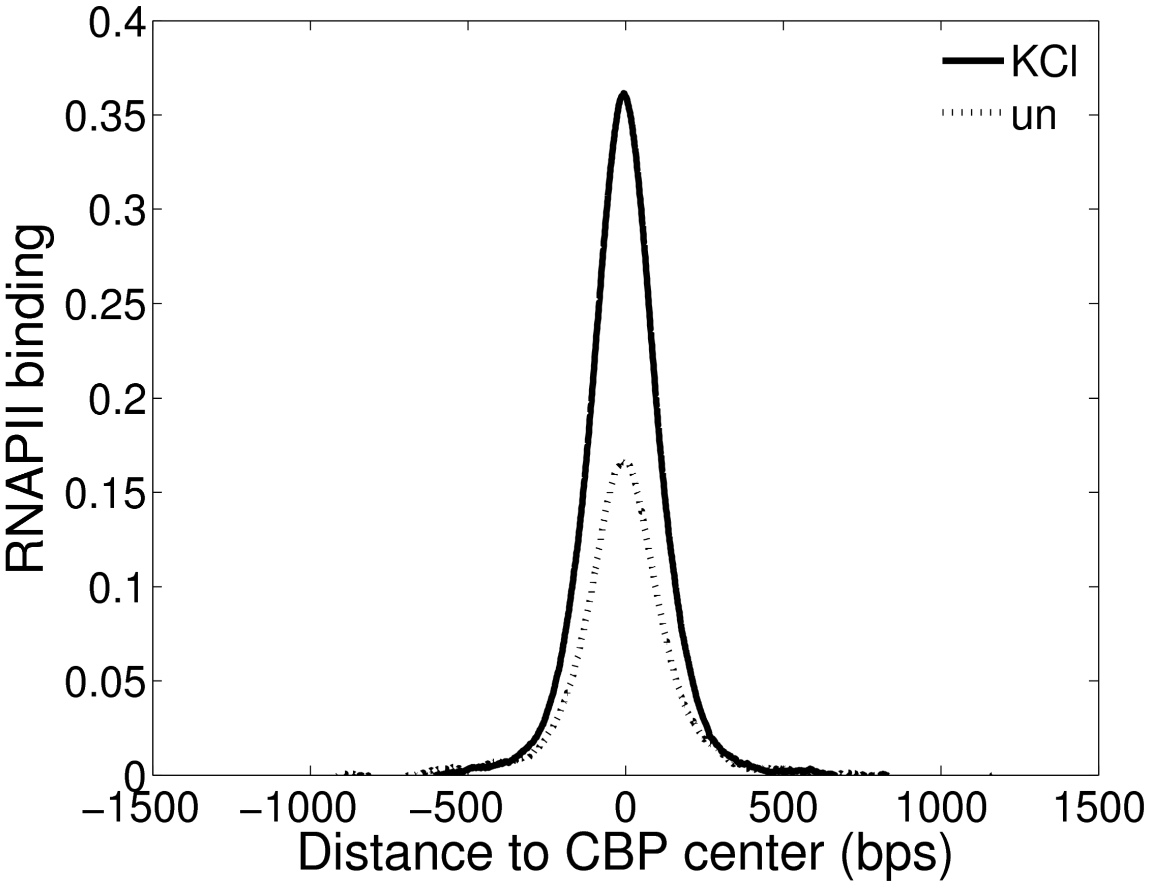
- Enriched for ~100 sequence motifs
- CBP binding determined by other TFs

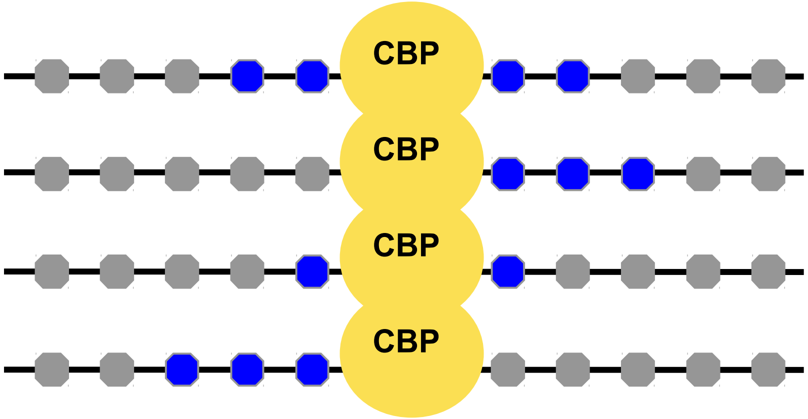


RNAPII is recruited to CBP binding sites at the *fos* locus

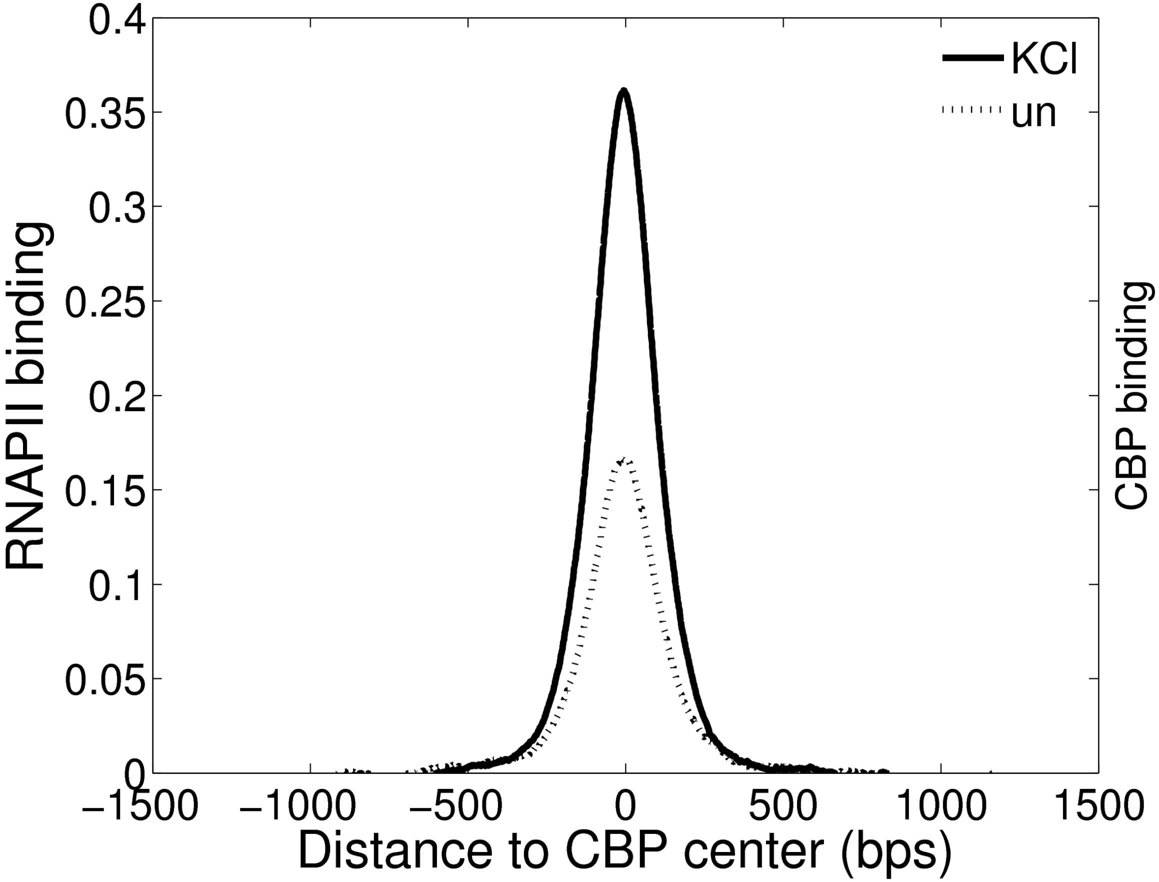


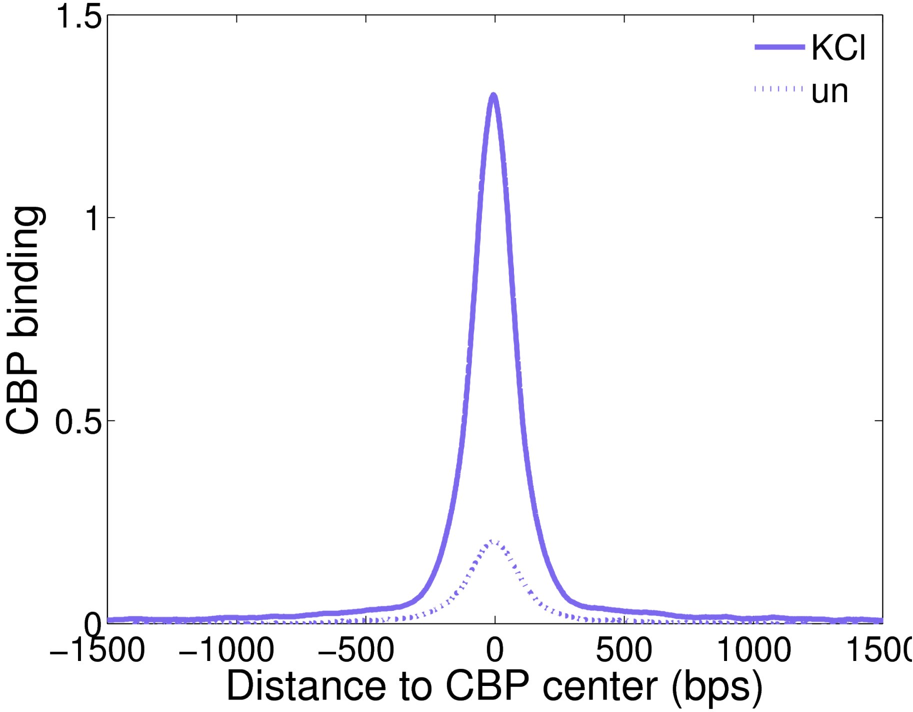
RNAPII is recruited at enhancers

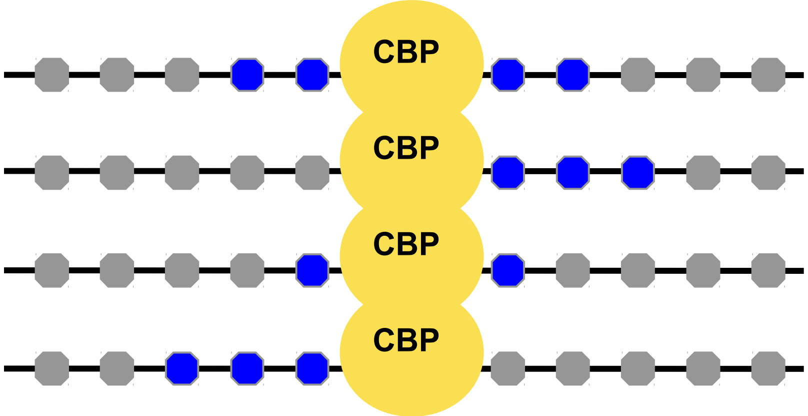




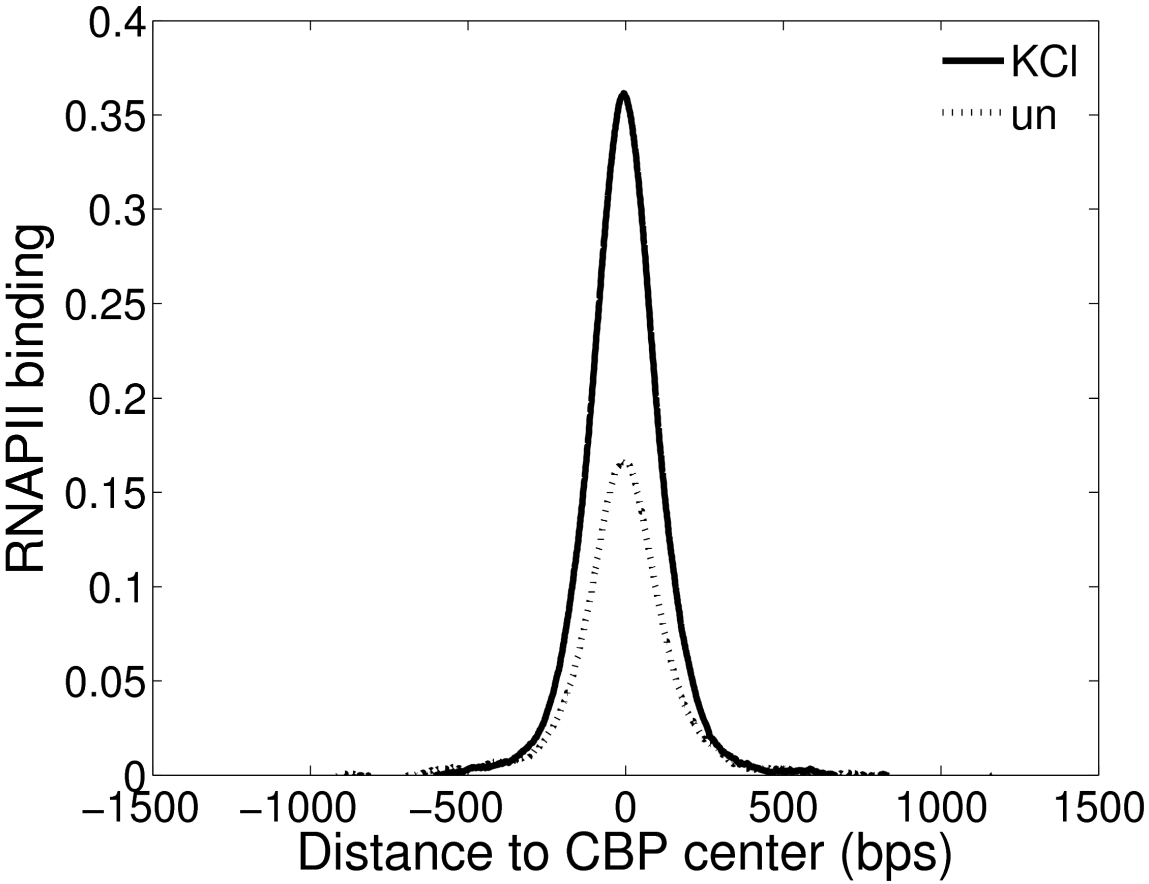
RNAPII is recruited at enhancers

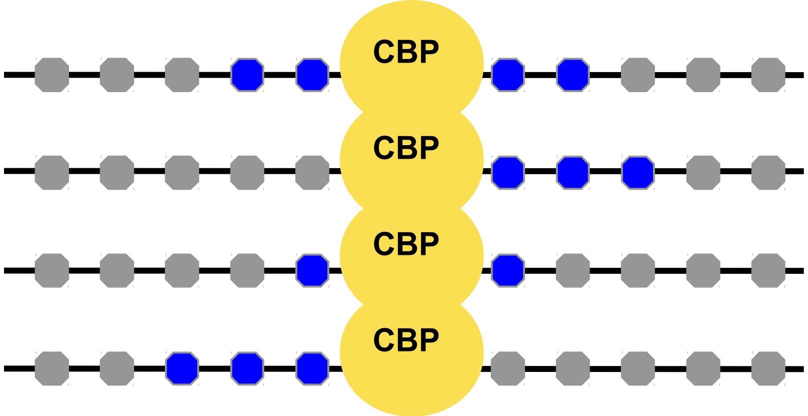


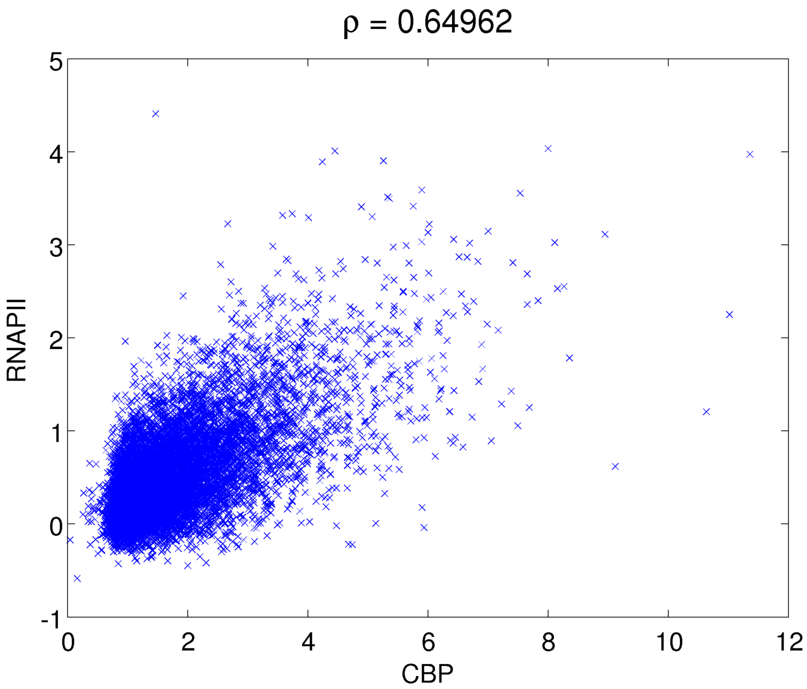




RNAPII is correlated with CBP

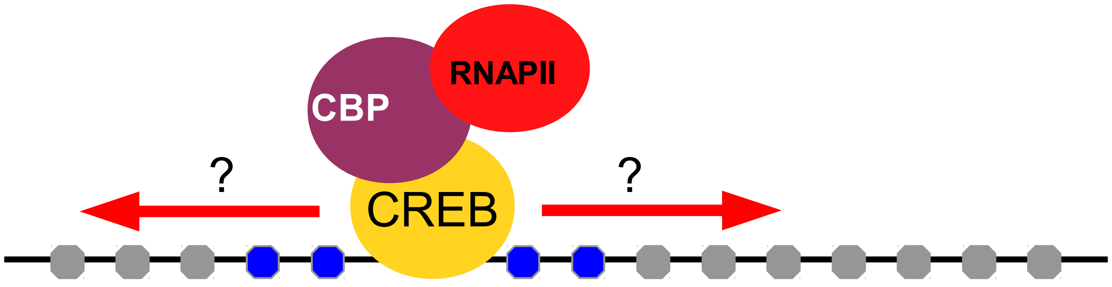






What is the function of RNAPII at enhancers?

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



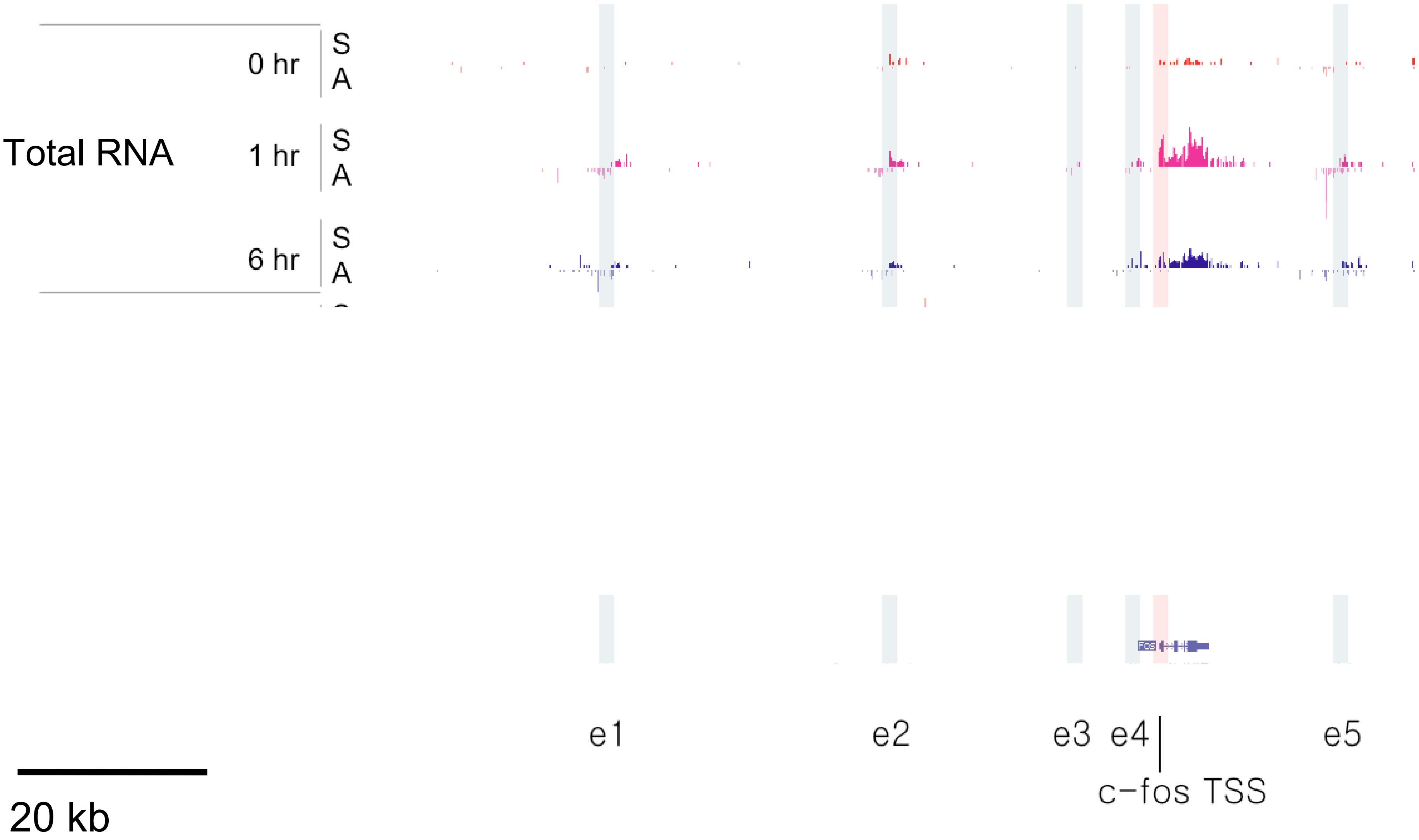
Enhancer
H3K4me1

RNA-Seq finds transcribed parts of the genome

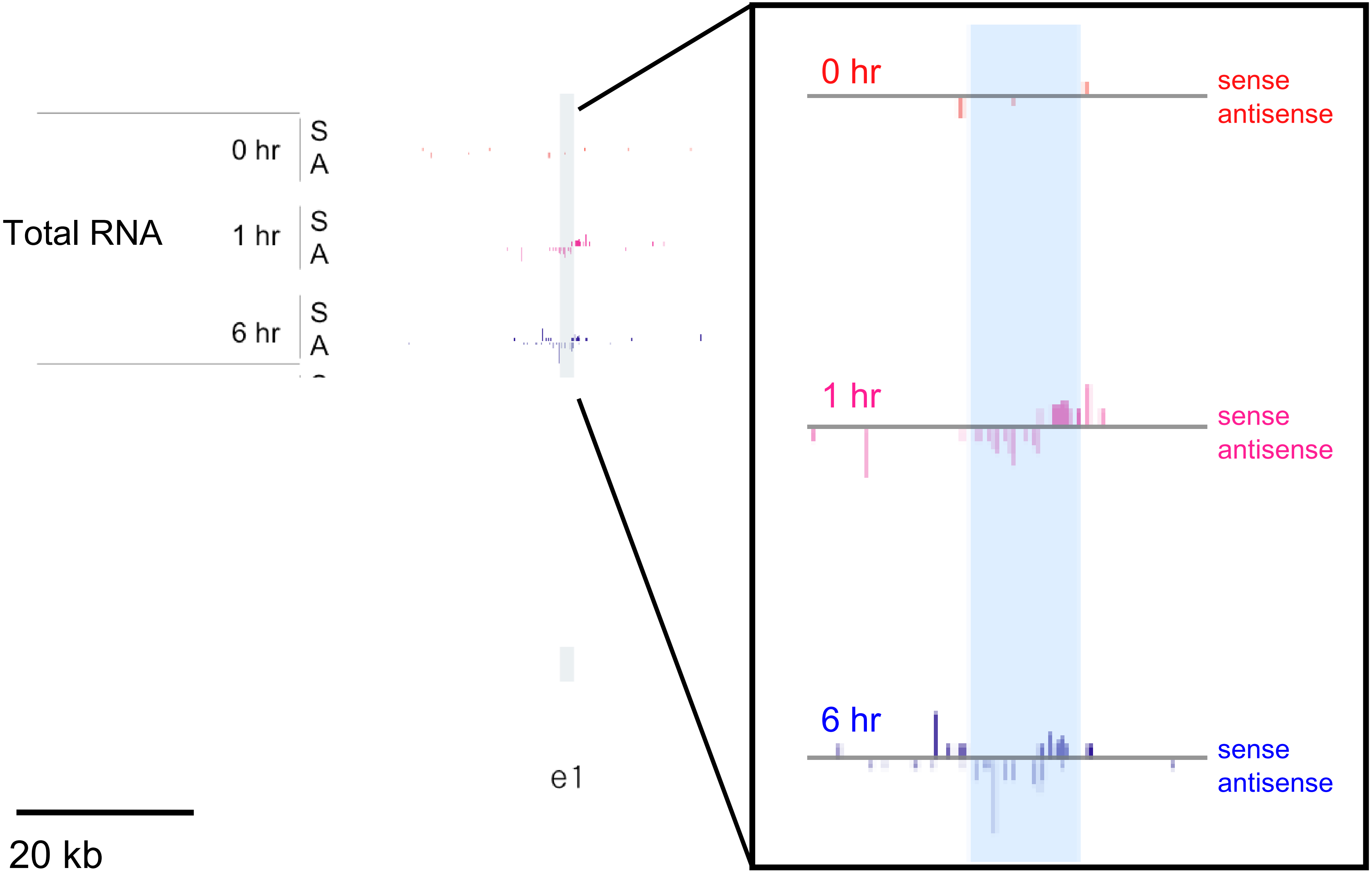
- Short **reads** mapped to reference genome
- ~5×10^6 reads
- #reads ~ RNA

(Wang et al, 2009)

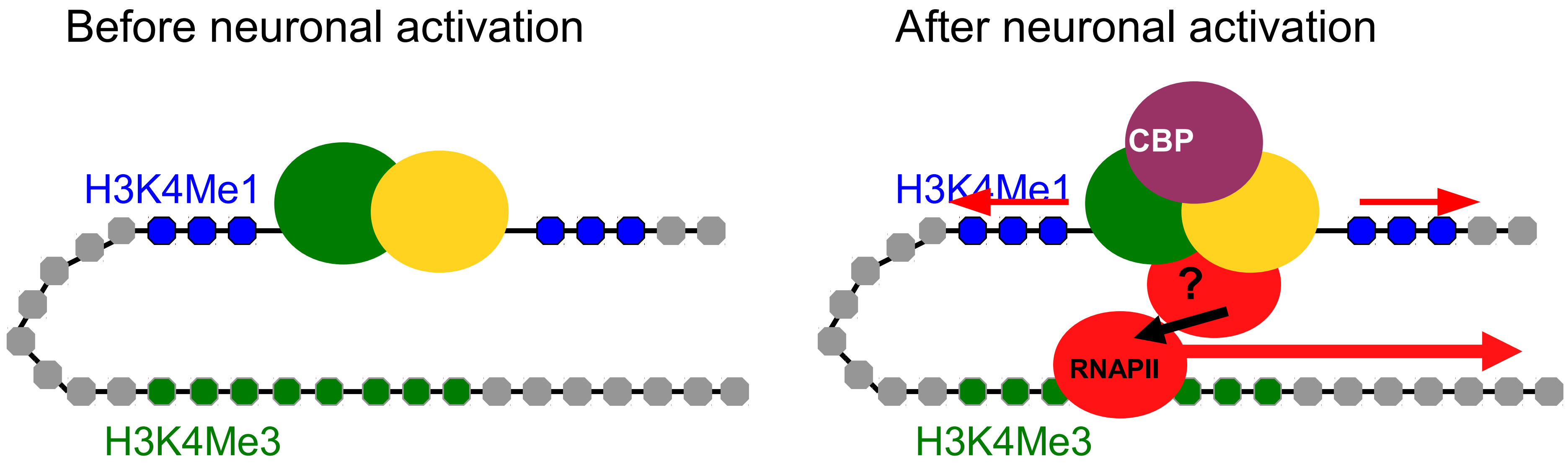
Transcription of total RNA at the *fos* locus



Transcription at enhancers is activity-dependent

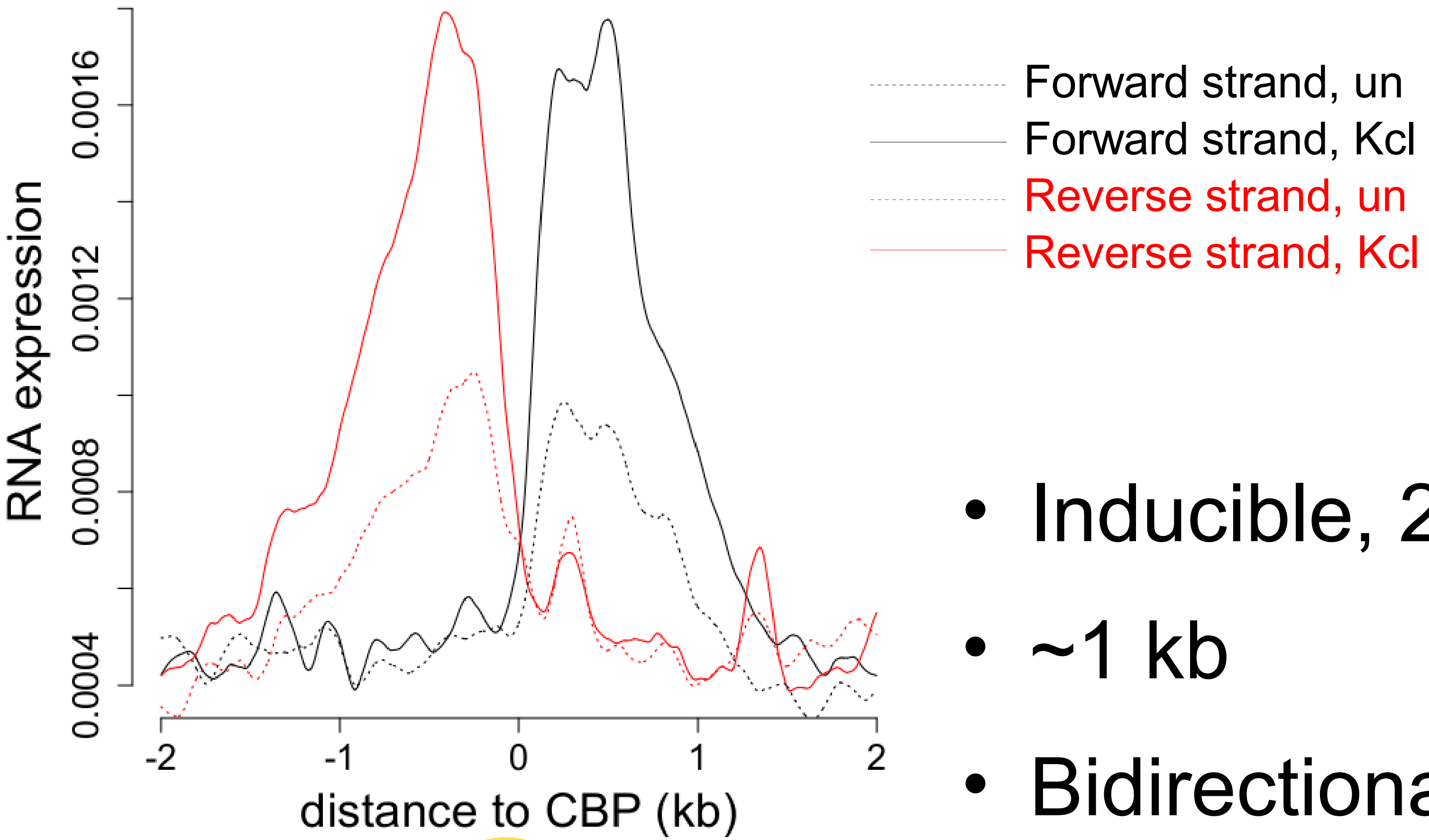


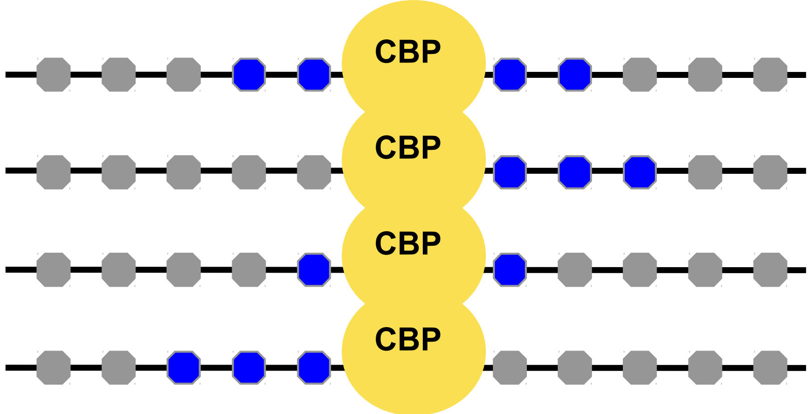
Enhancer RNAs (**eRNAs**) novel species



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

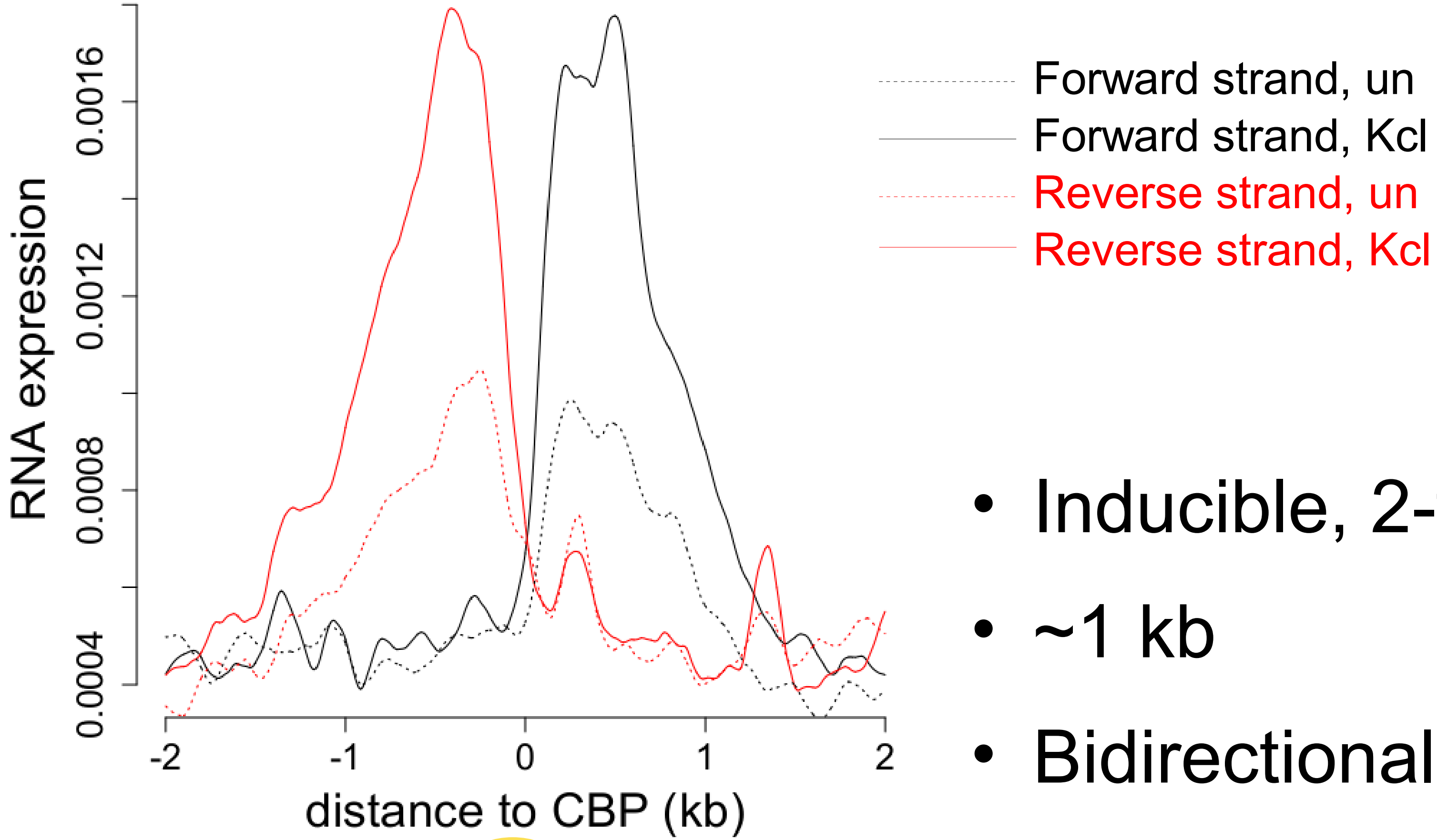
eRNAs are induced by activity

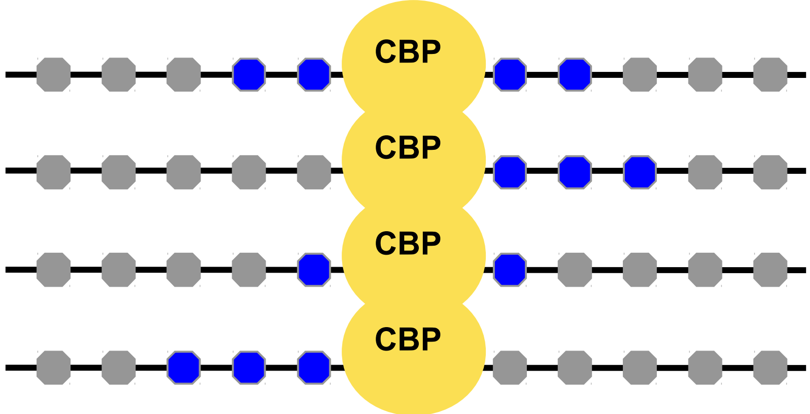




- Inducible, 2-fold
- ~1 kb
- Bidirectional

eRNAs are 100-fold lower than mRNAs

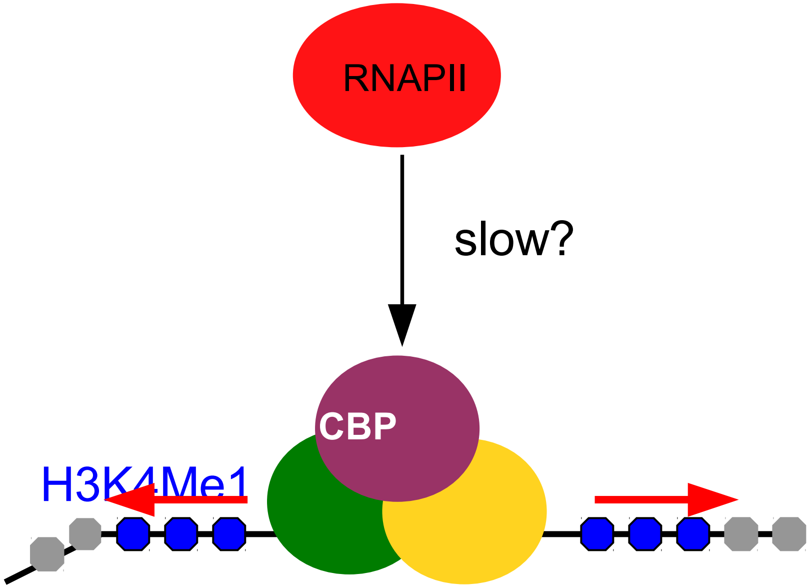


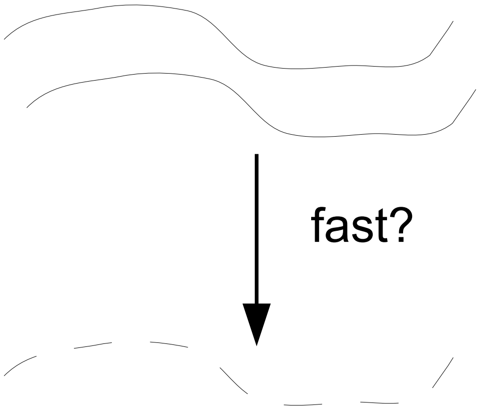


- Inducible, 2-fold
- ~1 kb
- Bidirectional
- 1 in 10k reads eRNA
  - Not at all CBP peaks
- Not protein-coding

Why do eRNAs have such low abundance?

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





Half life of eRNAs relative to mRNAs

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

Length of transcript

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

RNAPII level

Transcript level

half-life

eRNAs half life is less than half an hour

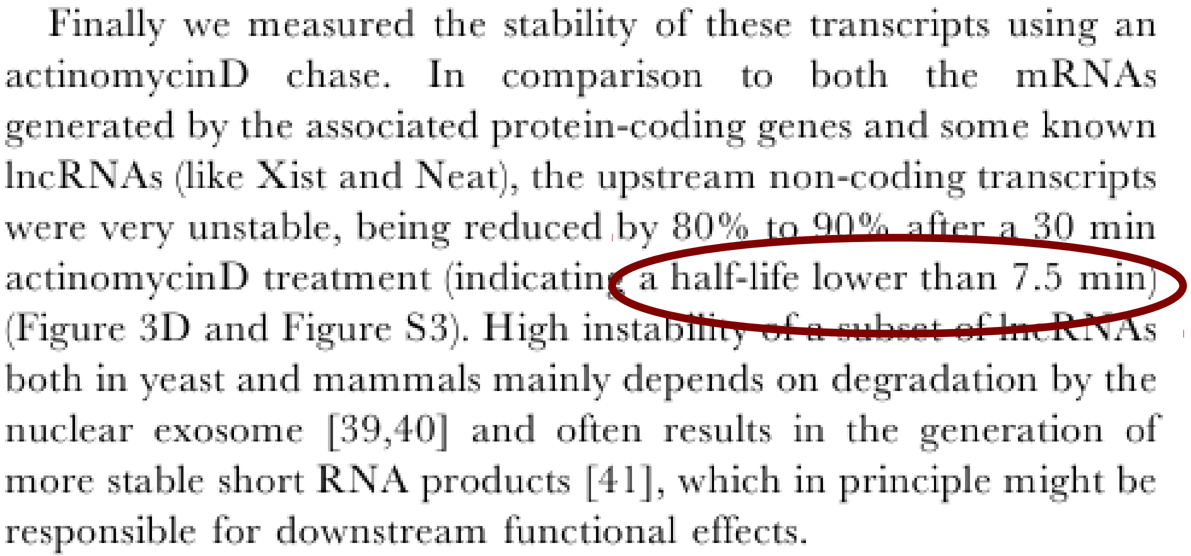
- eRNA production much slower than mRNA
- eRNA decay much faster than mRNA

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

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

Estimate consistent with experiments

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

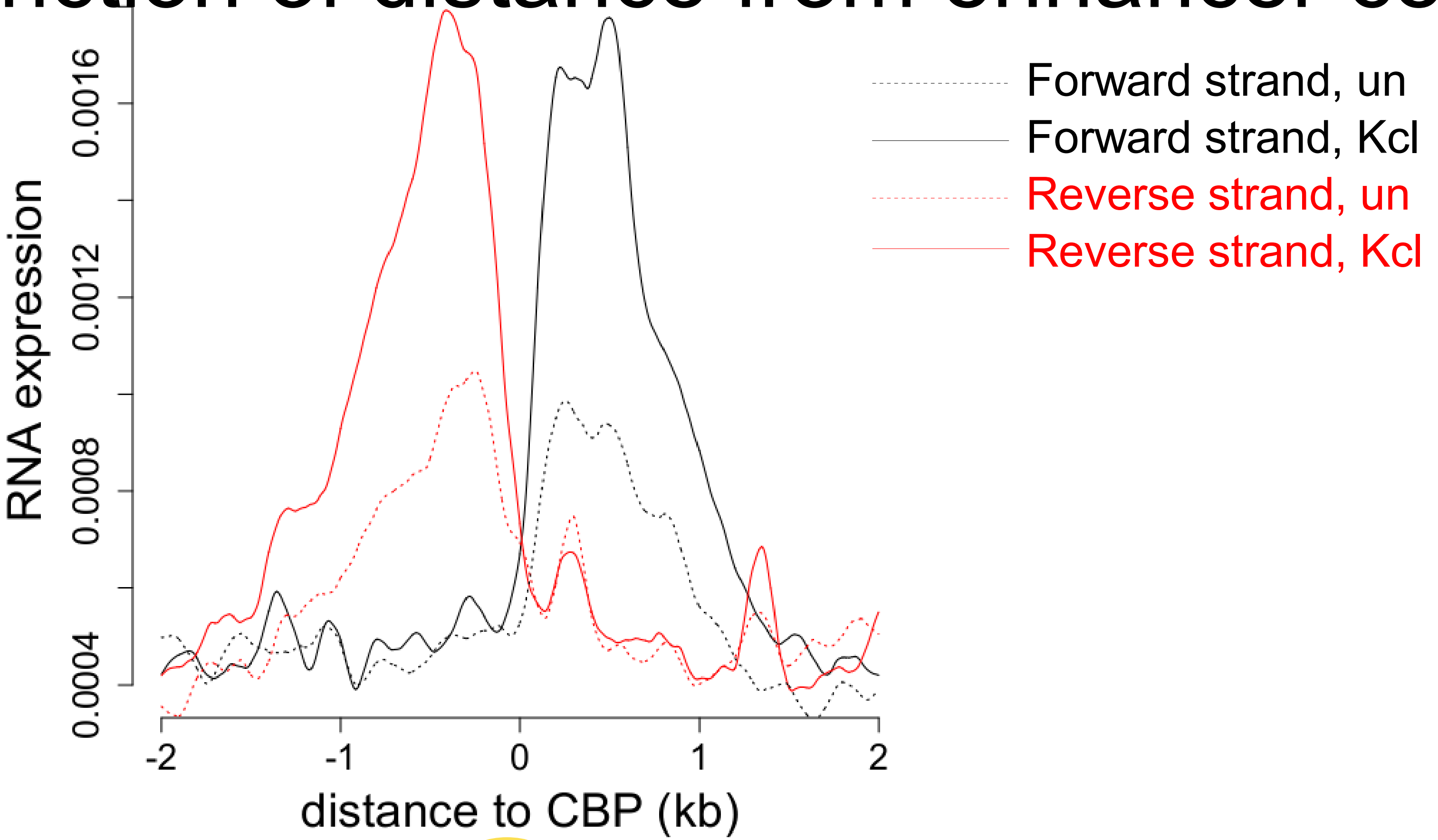


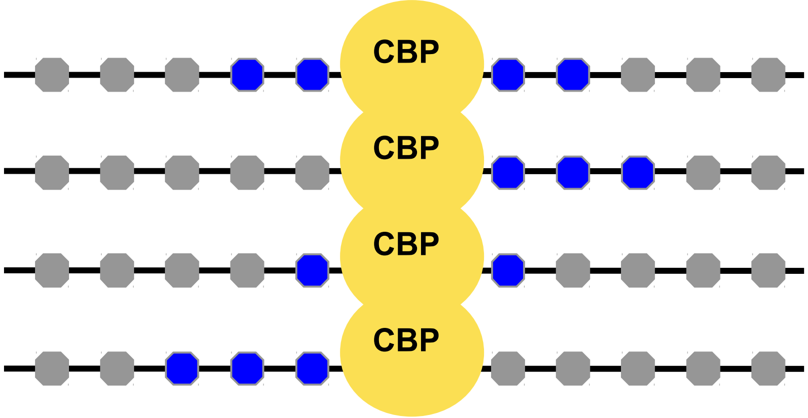
Finally we measured the stability of these transcripts using an actinomycinD chase. In comparison to both the mRNAs generated by the associated protein-coding genes and some known lncRNAs (like Xist and Neat), the upstream non-coding transcripts were very unstable, being reduced by 80% to 90% after a 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.

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

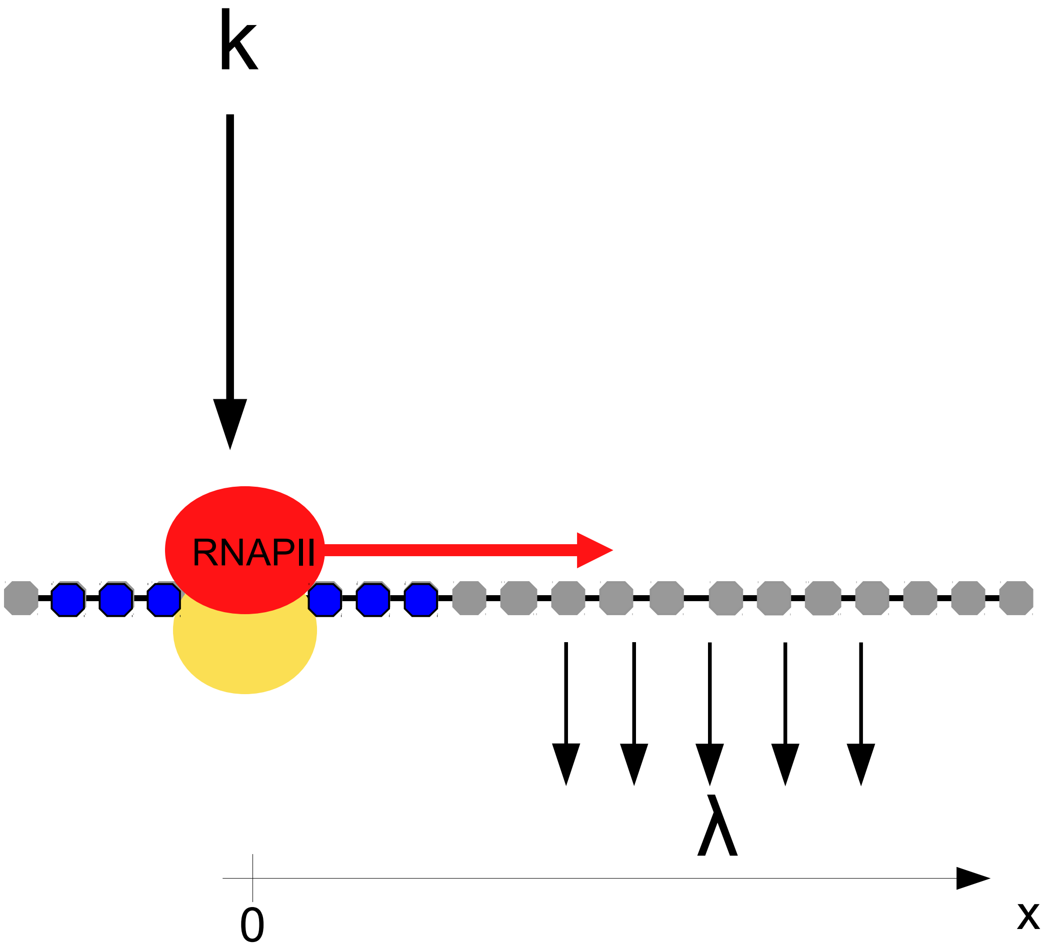
$$\tau_E \sim 10^{-2} \times 1 \times 2 \times \tau_M \sim 2 \times 10^{-2} \times 600\text{min} = \boxed{12\text{min}}$$

A quantitative model of eRNA levels as a function of distance from enhancer center

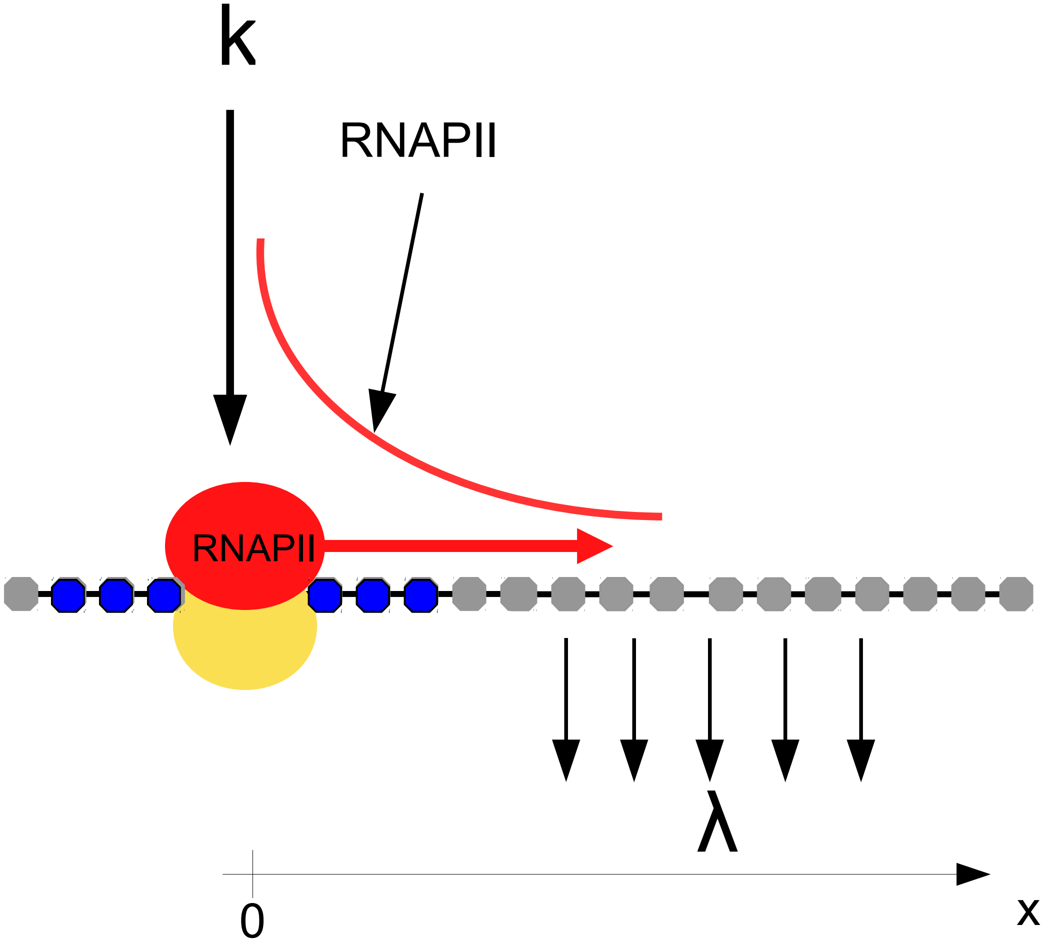




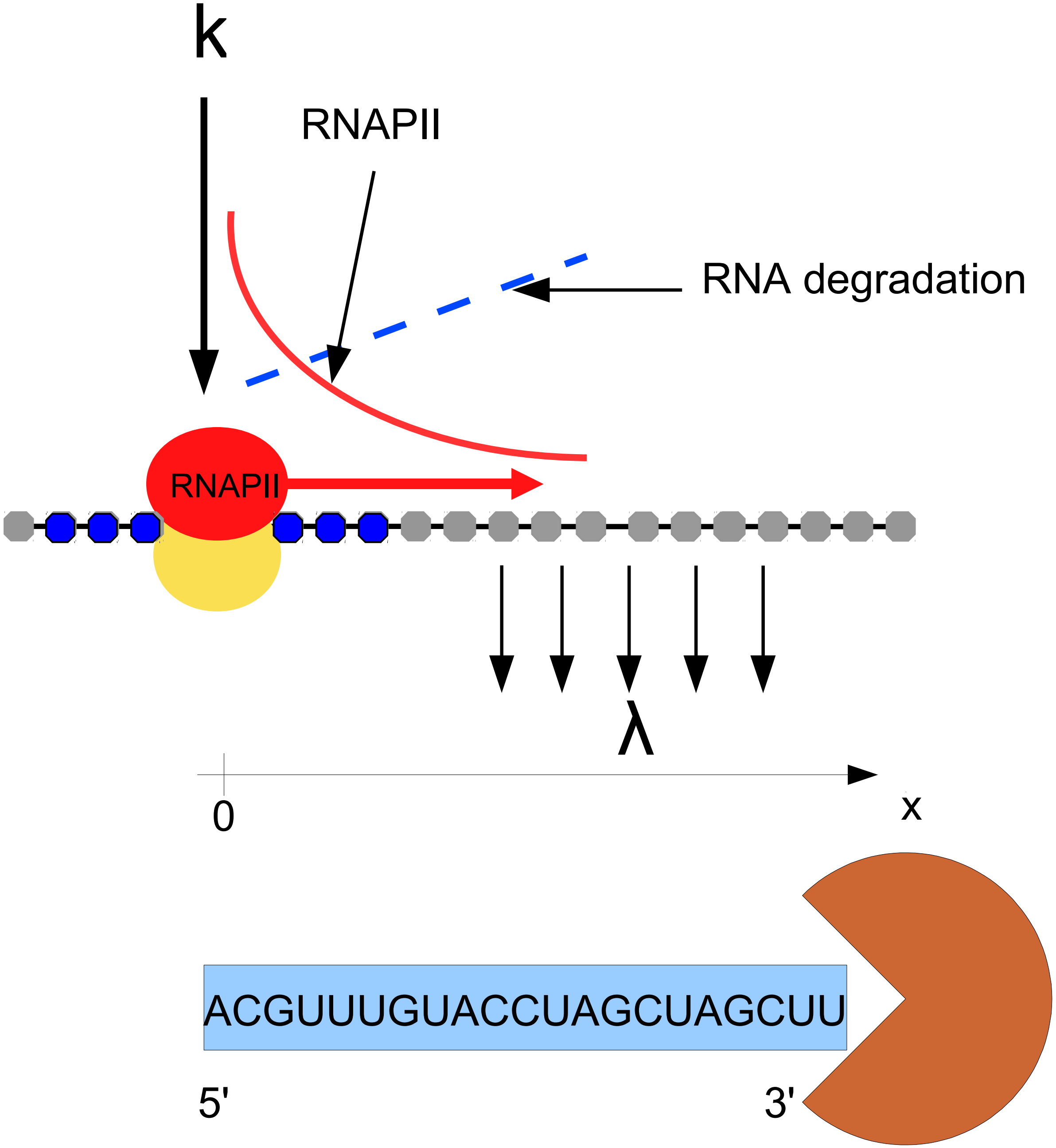
RNAPII binds and falls of at a constant rate



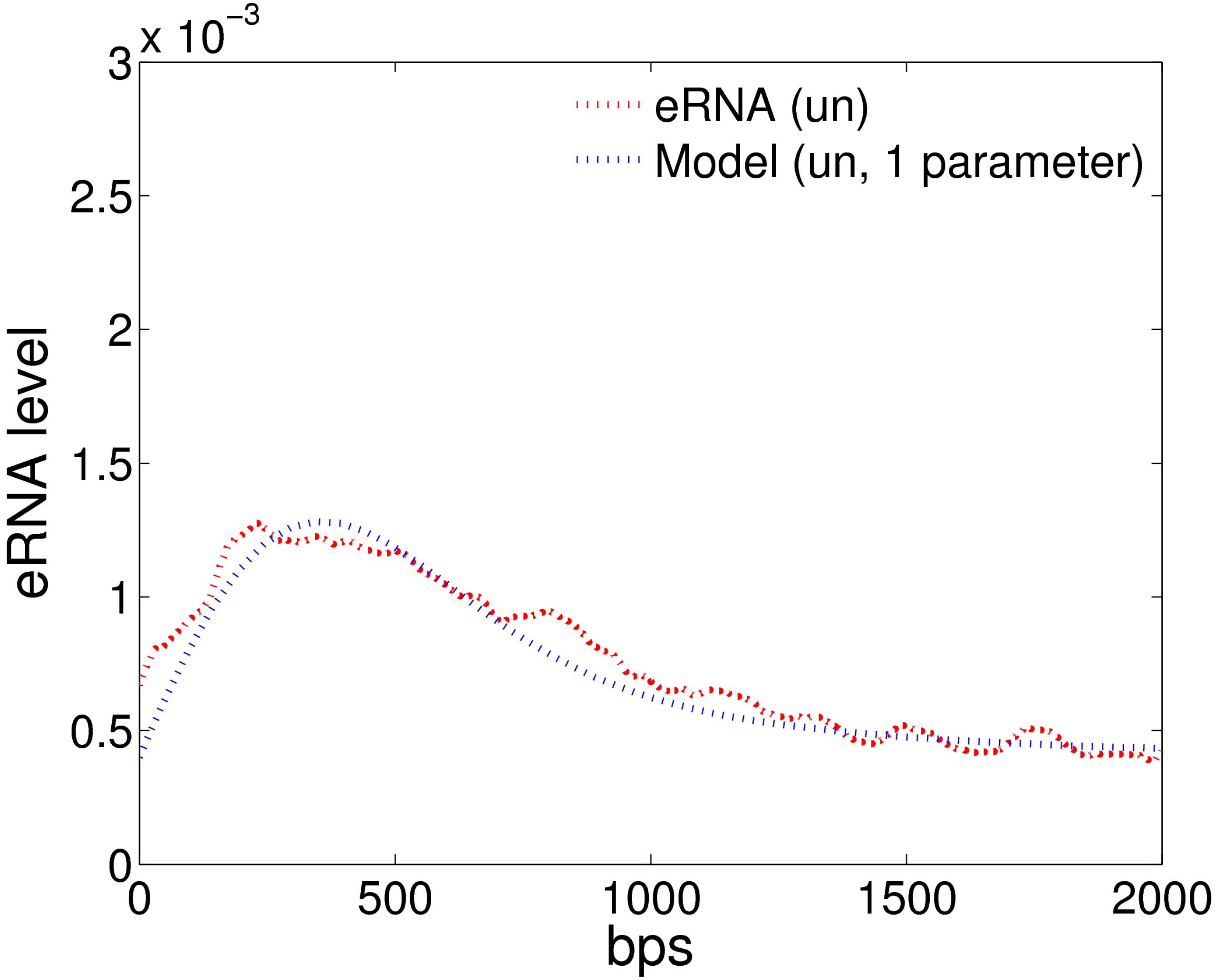
eRNA production proportional to RNAPII



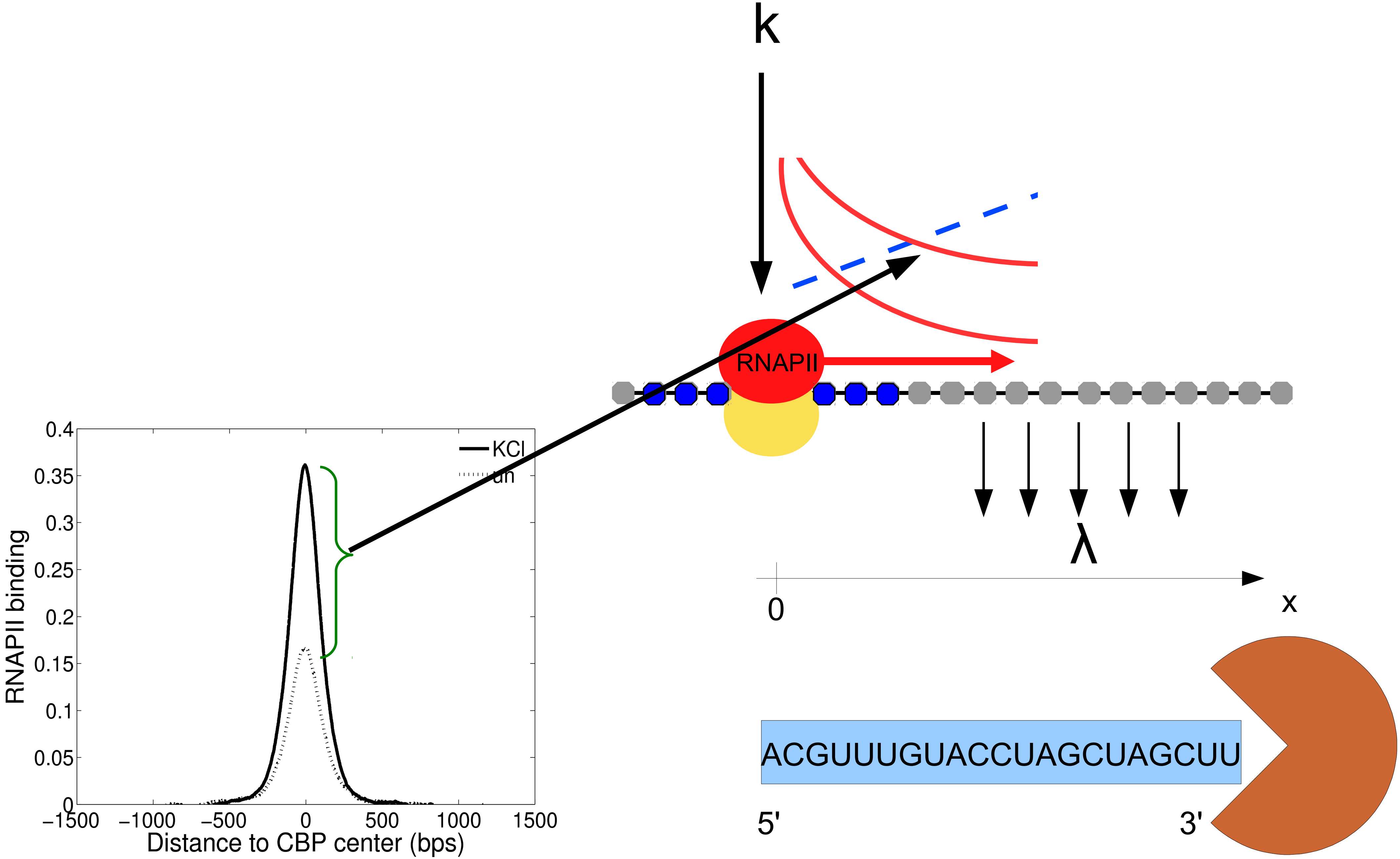
eRNA production proportional to RNAPII



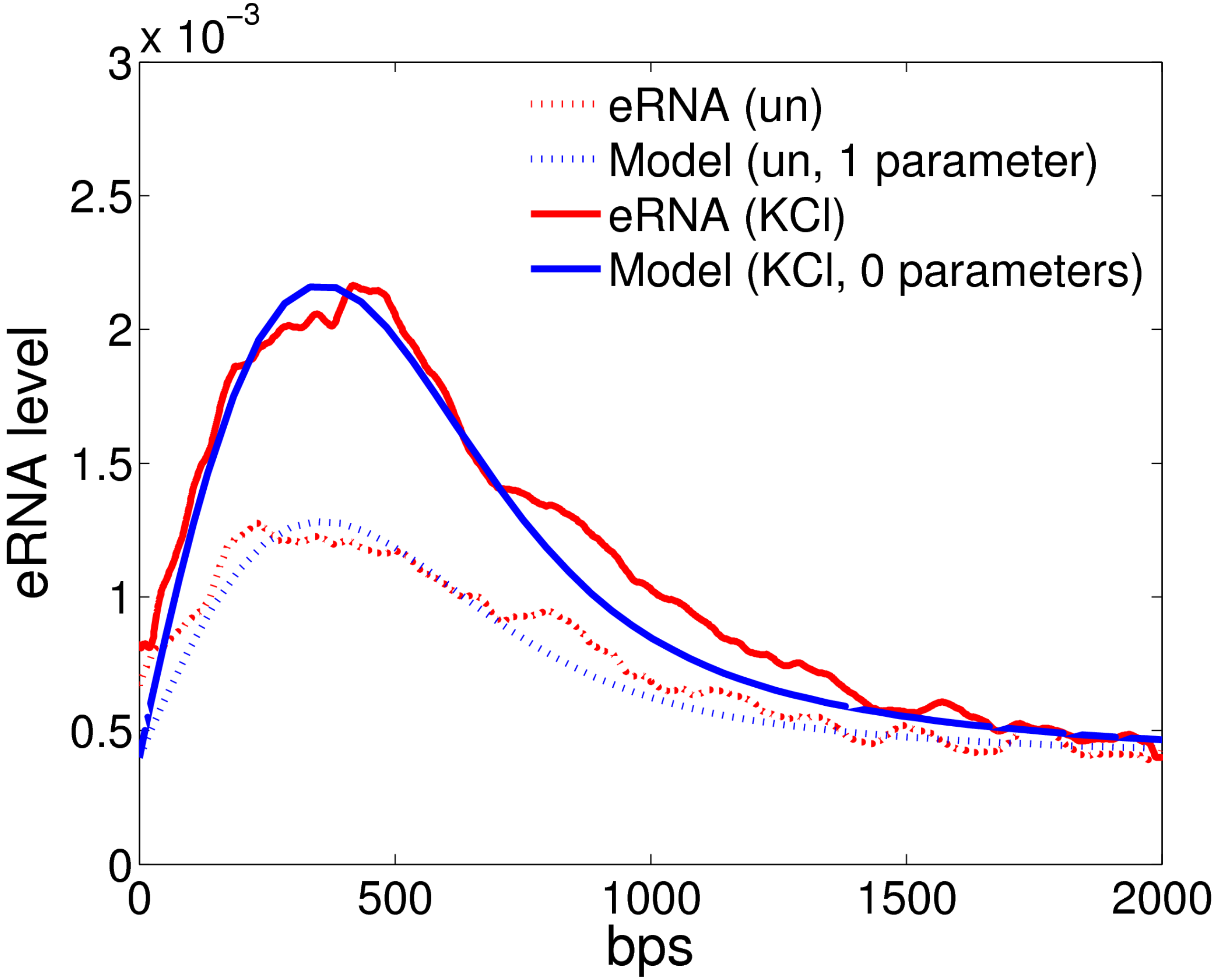
eRNA levels can be accurately predicted



Binding rate of RNAPII doubled after KCl

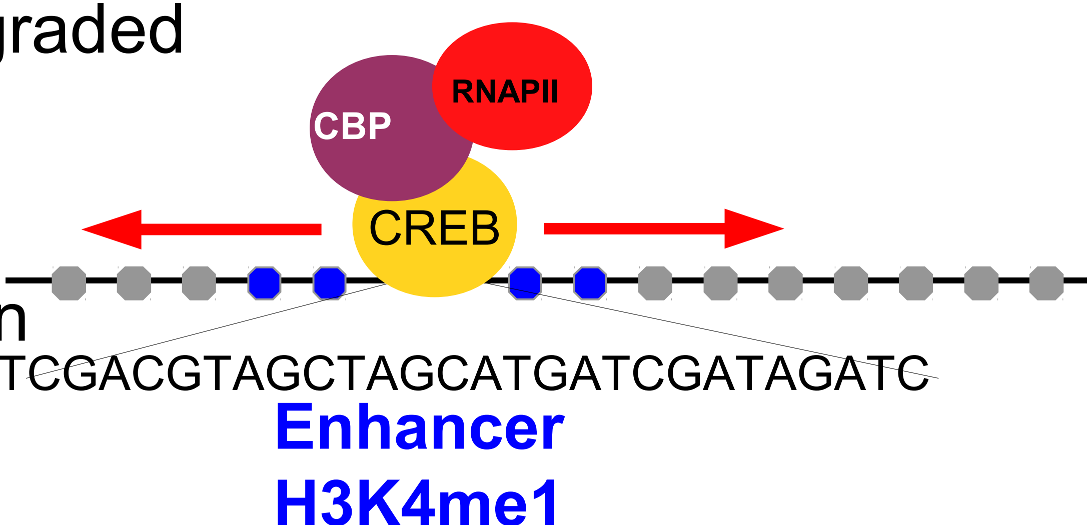


Excellent fit for eRNA after KCl without free parameters



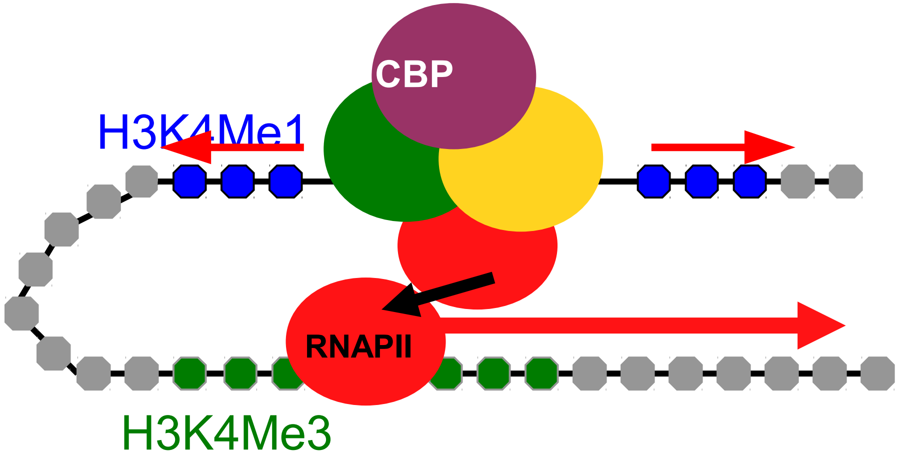
Properties of activity-dependent enhancers

- 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

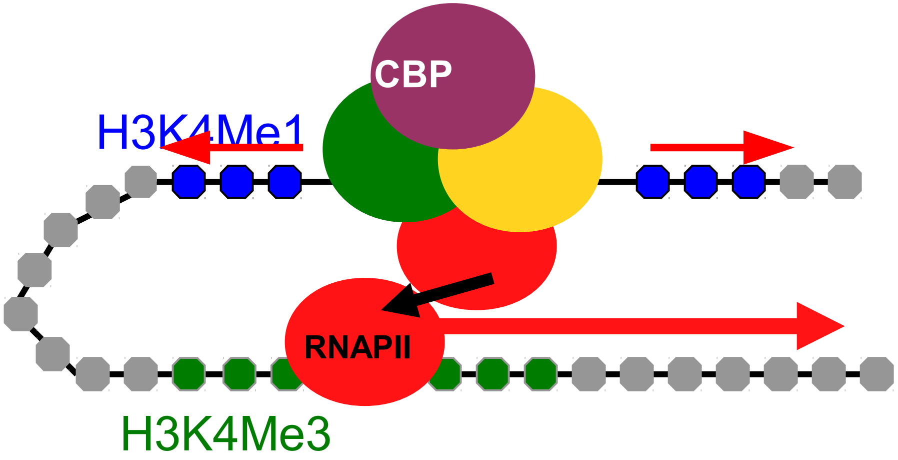


Science is always wrong. It never solves a problem without creating ten more.

-George Bernard Shaw

Recruitment of RNAPII at the promoter

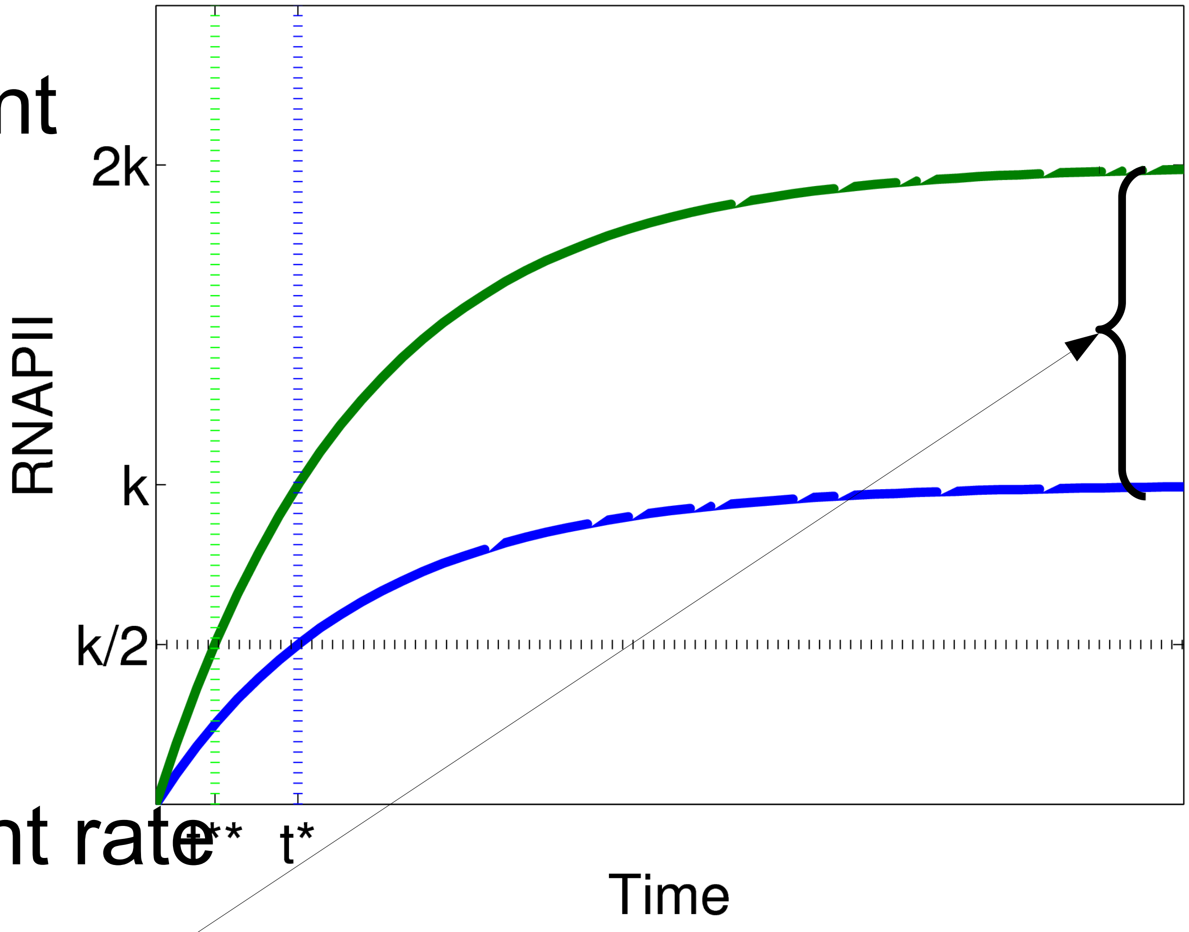
- Transcribe eRNAs
- Speed up recruitment of RNAPII at promoter



RNAPII recruitment rate
= [promoter rate]
+ N*[contact probability]*[enhancer rate]

Steady state level of RNAPII is increased

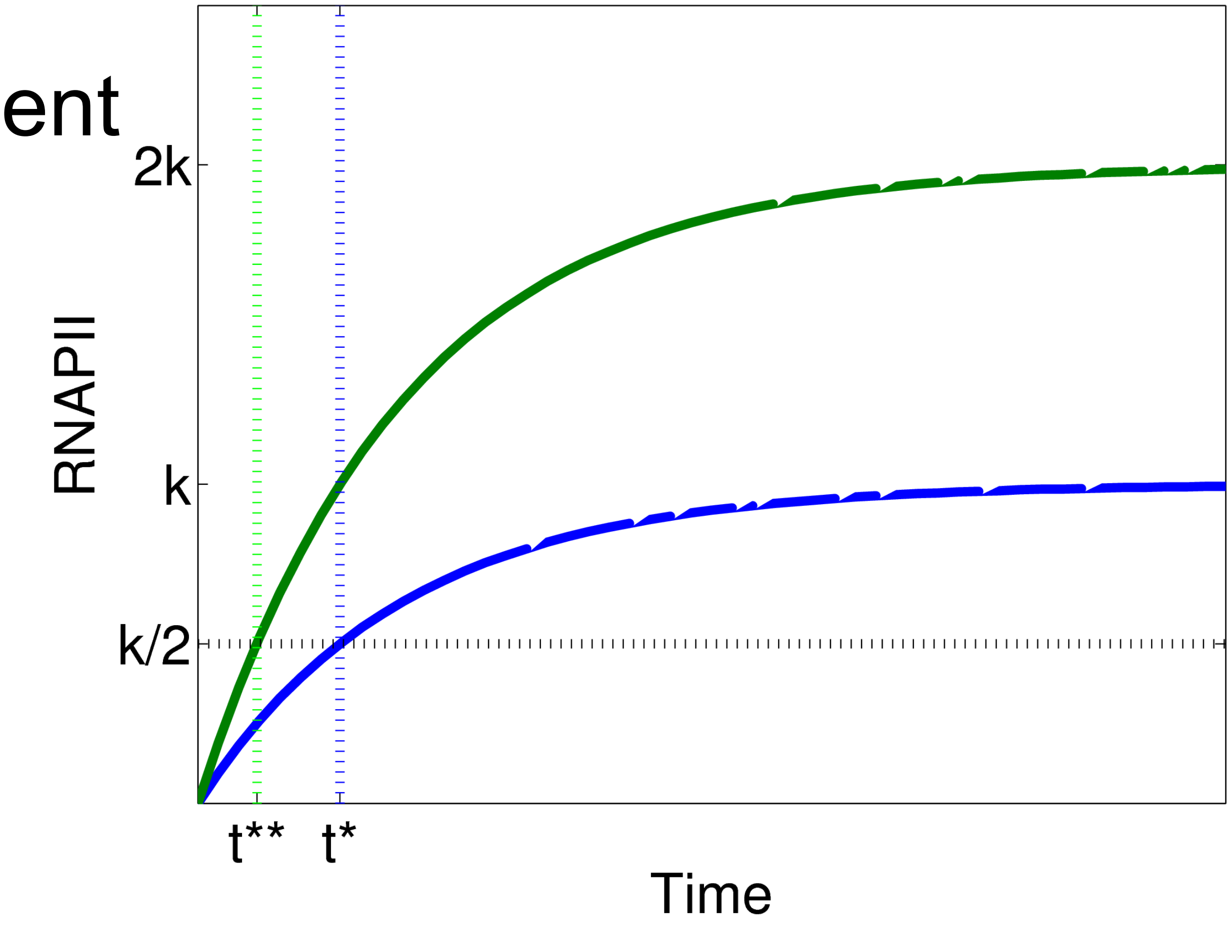
- Speed up recruitment of RNAPII at promoter



RNAPII recruitment rate
= [promoter rate]
+ N*[contact probability]*[enhancer rate]

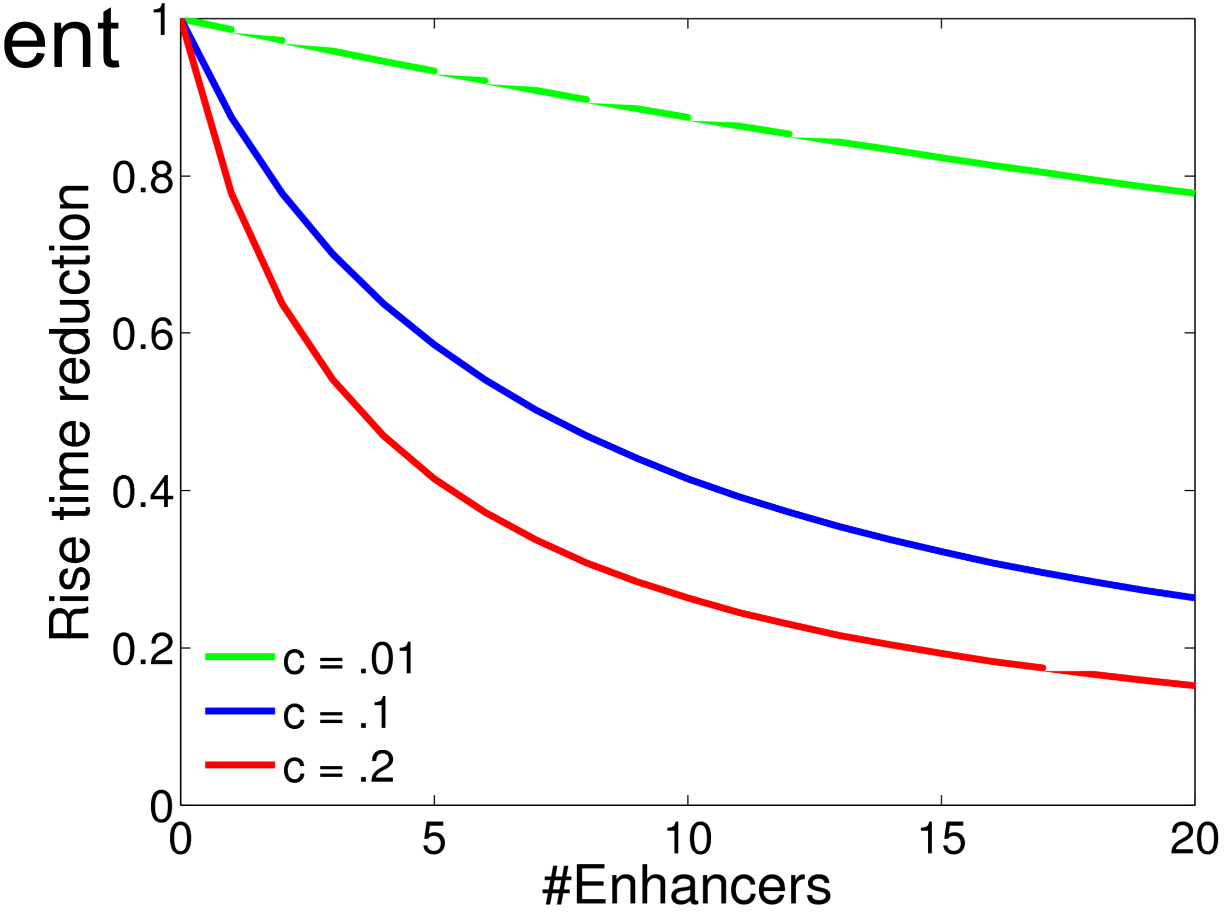
Rise time is reduced

- Speed up recruitment of RNAPII at promoter



Significant speed-up with ~5 enhancers

- Speed up recruitment of RNAPII at promoter



Enhancers may reduce the noise in RNAPII

- Speed up recruitment of RNAPII at promoter
- Reduce noise

RNAPII recruitment rate
= [promoter rate]
+ N*[contact probability]*[enhancer rate]

Reduction of noise proportional to the number of enhancers

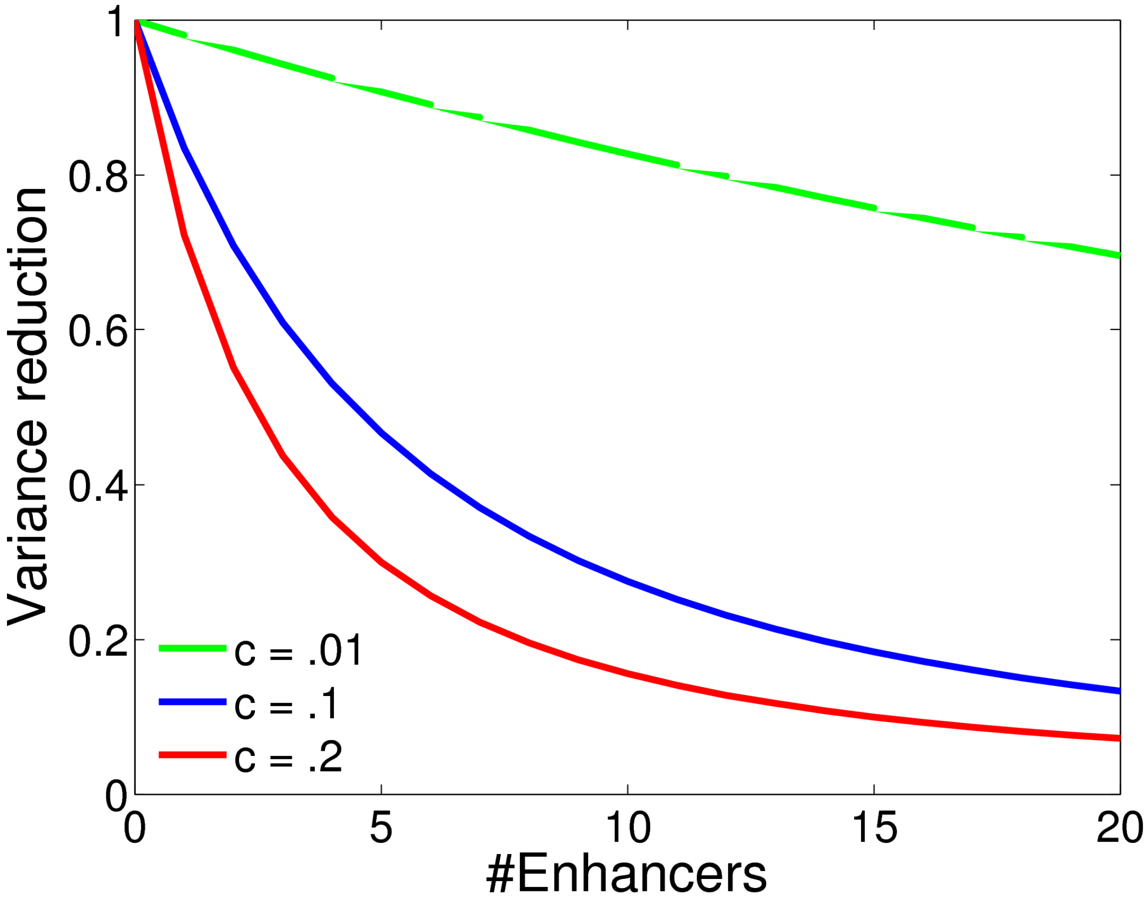
- Speed up recruitment of RNAPII at promoter
- Reduce noise

RNAPII recruitment rate
= [promoter rate]
+ N*[contact probability]*[enhancer rate]

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

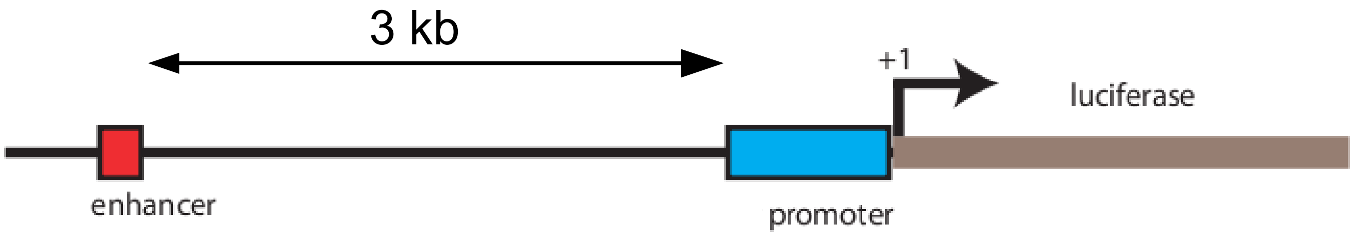
~50% reduction of noise with ~5 enhancers

- Speed up recruitment of RNAPII at promoter
- Reduce noise



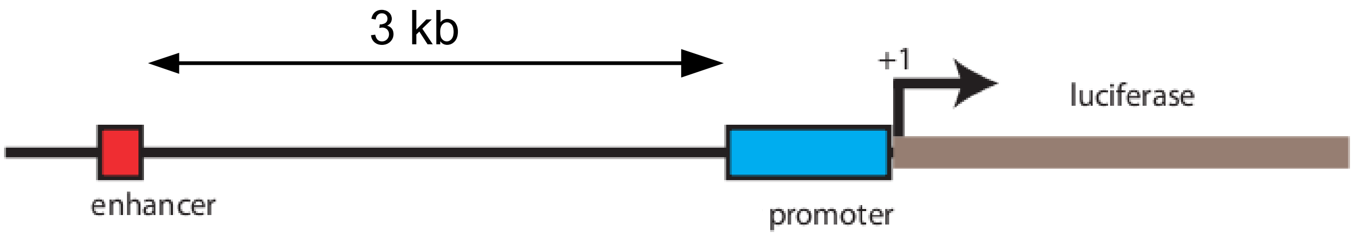
$$\frac{\text{Variance strong promoter}}{\text{Variance weak promoter with enhancers}} = \frac{\mathrm{Var}[(1+Nc)k]}{\mathrm{Var}[k]+N\mathrm{Var}[ck]} = \frac{(1+Nc)^2\mathrm{Var}[k]}{(1+Nc^2)\mathrm{Var}[k]} \sim N$$

Do eRNAs enhance gene expression?



- Luciferase construct to test enhancer function

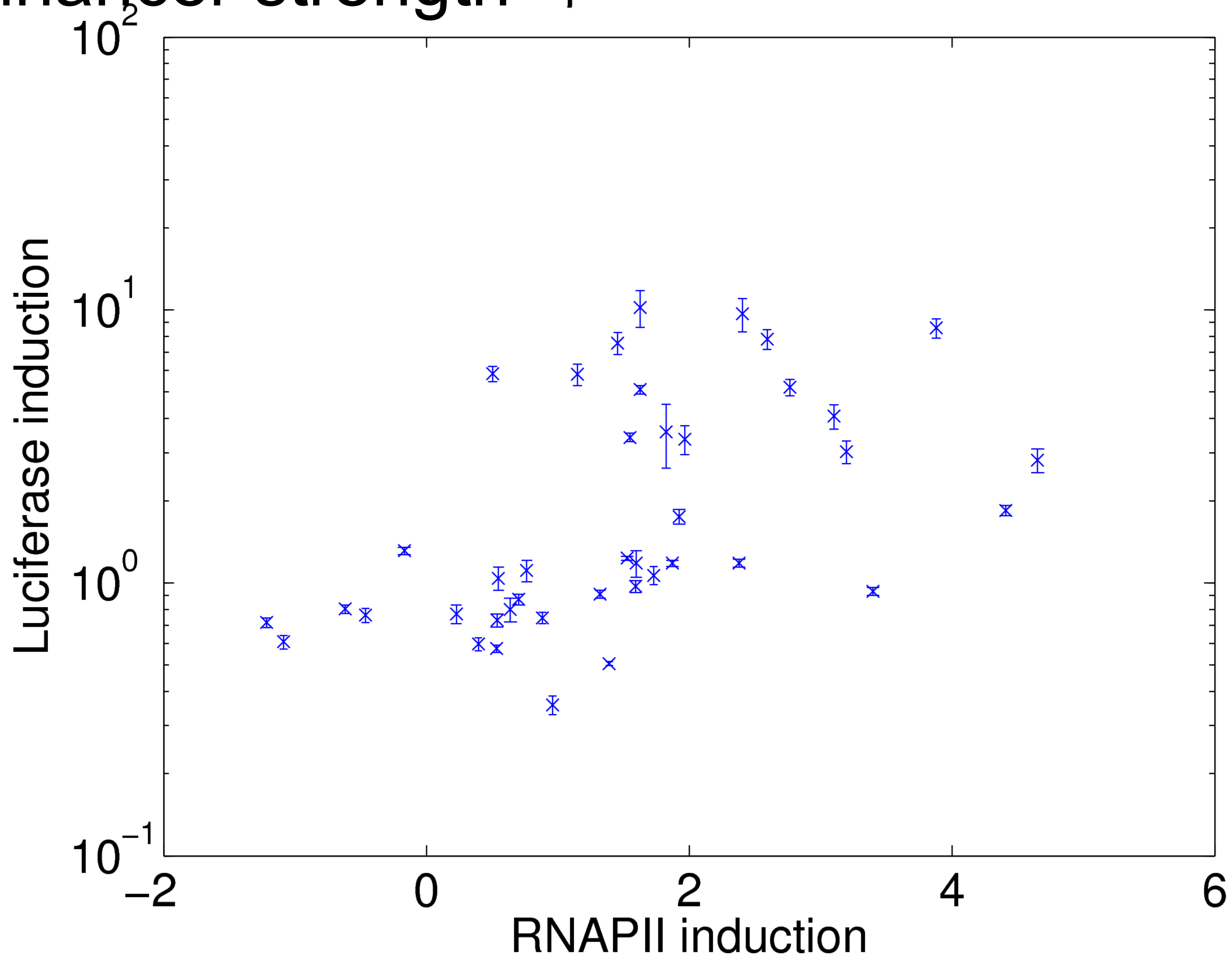
Do eRNAs enhance gene expression?



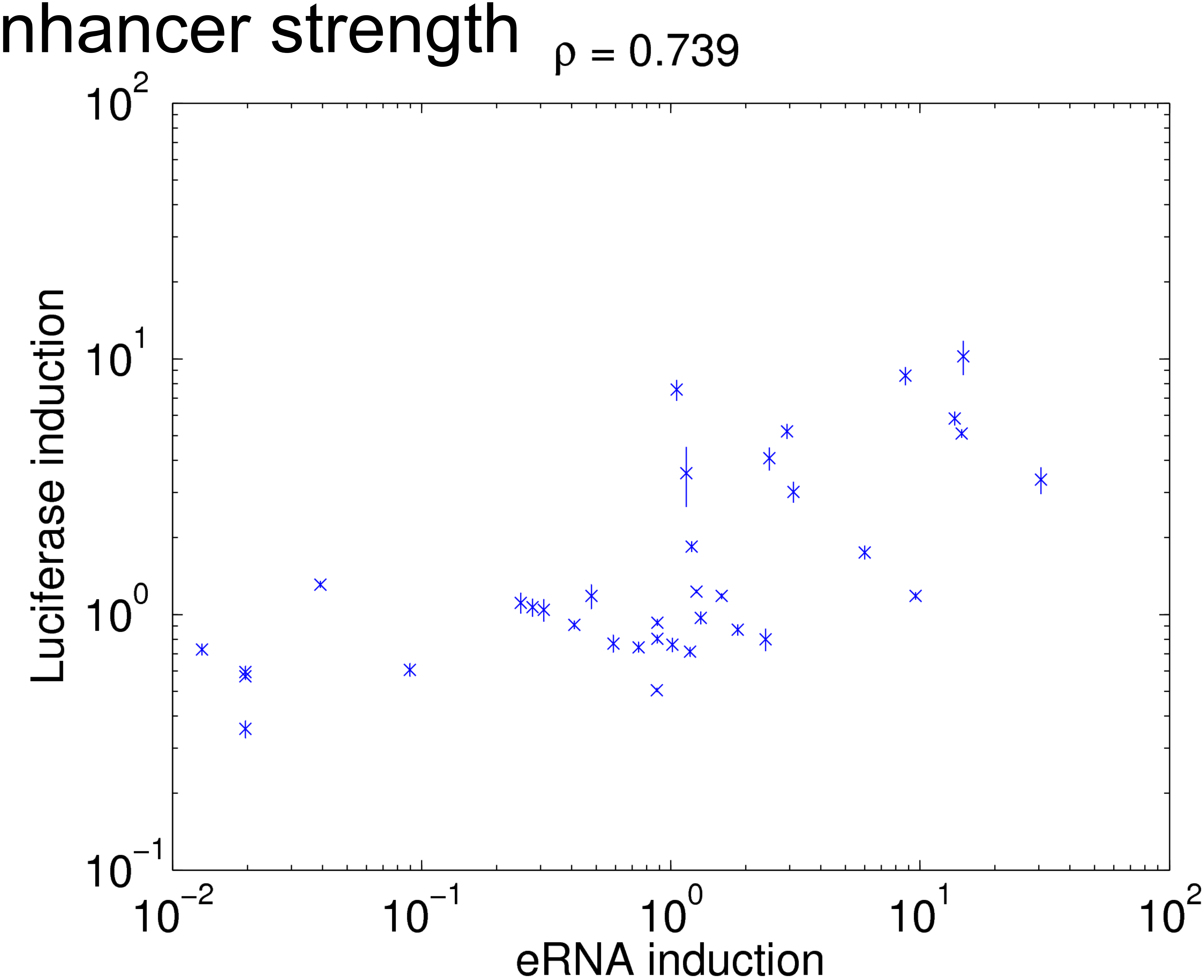
- Luciferase construct to test enhancer function
  - Narp gene

RNAPII induction is weakly correlated with enhancer strength

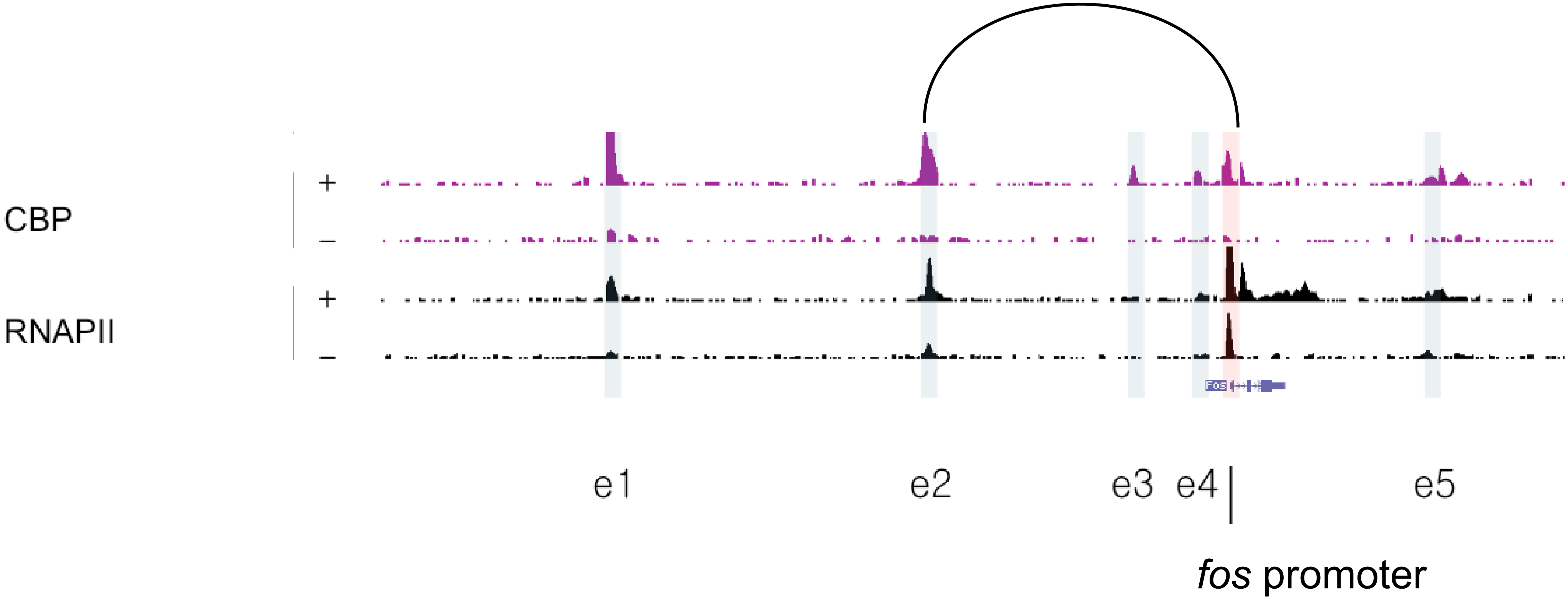
$\rho = 0.50$



eRNA induction is strongly correlated with enhancer strength

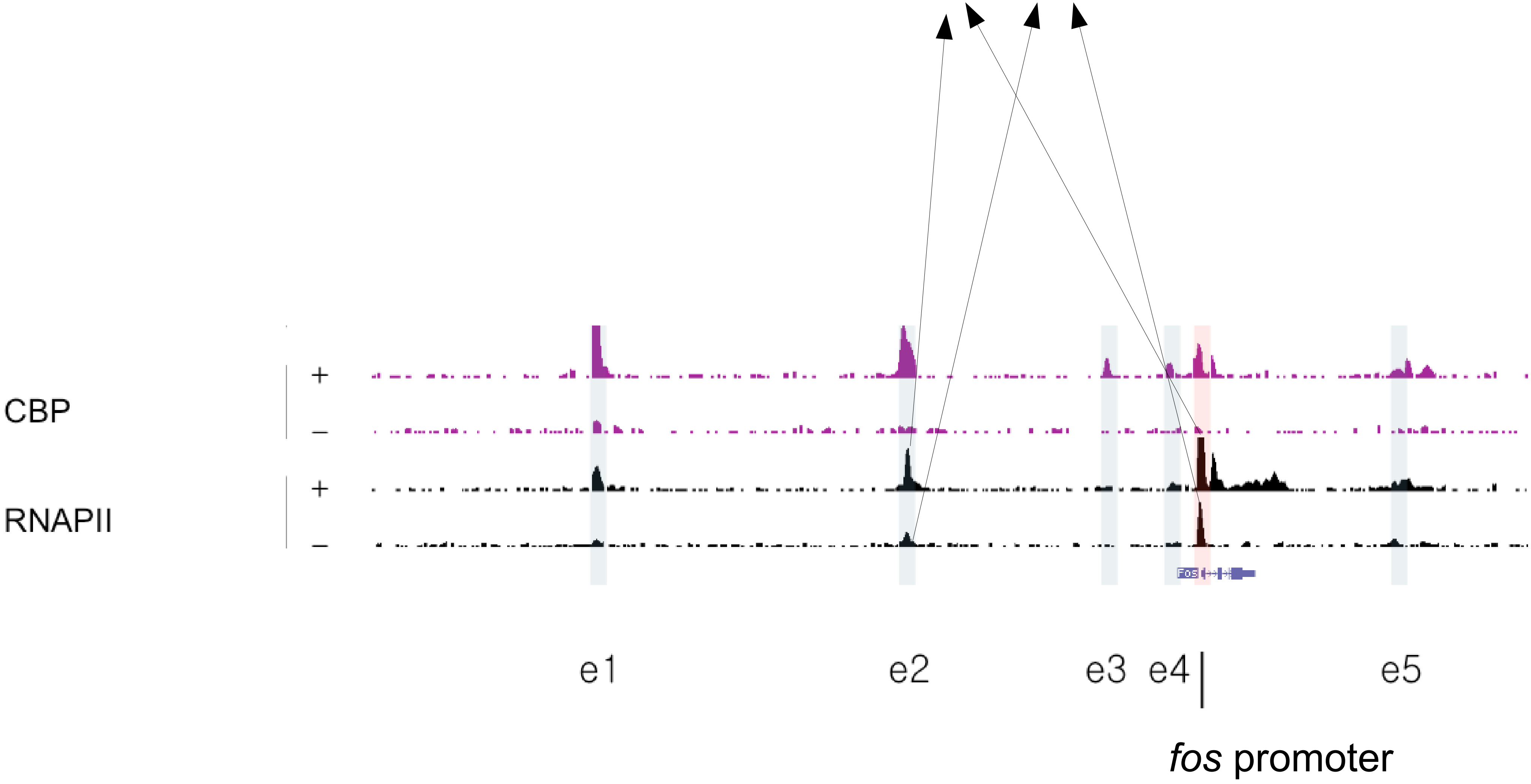


Is eRNA induction correlated with mRNA induction in vivo?



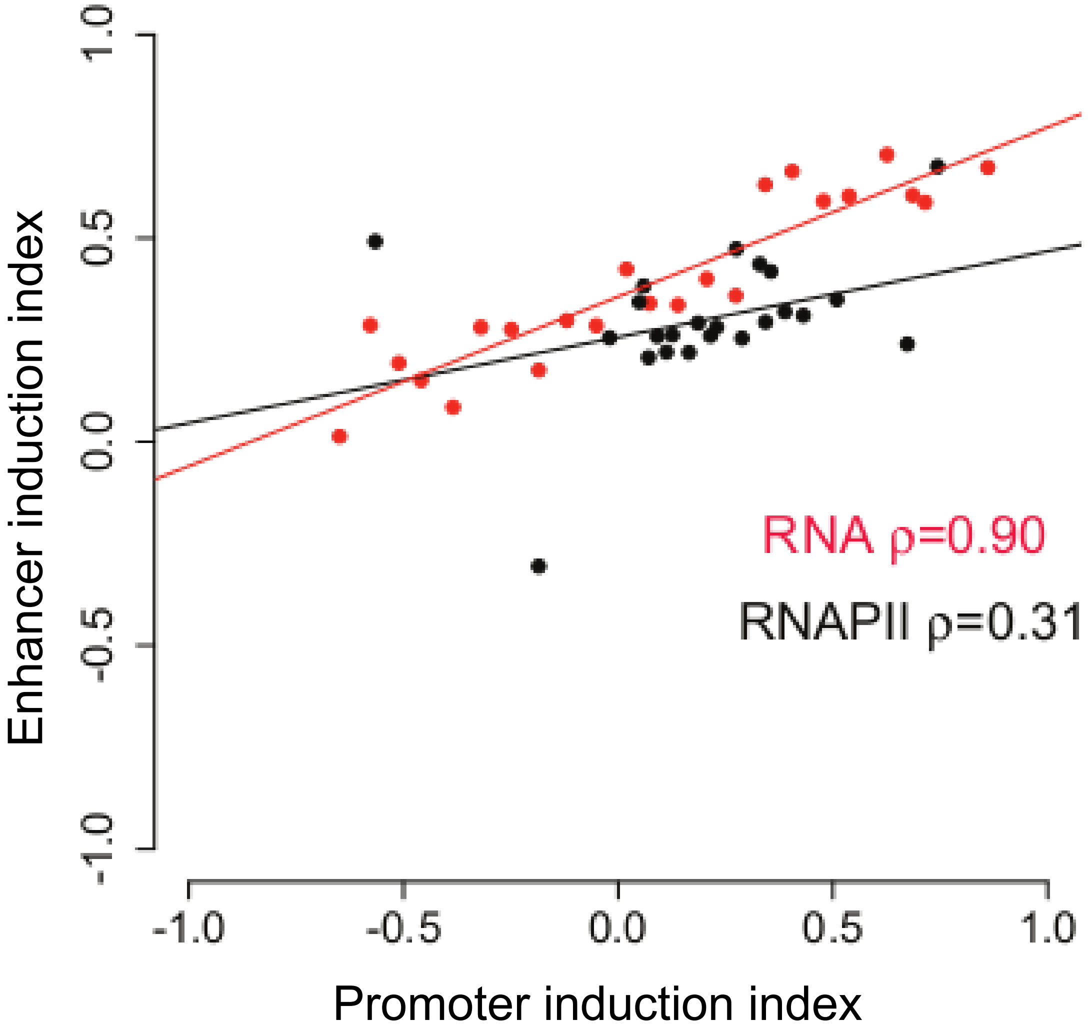
Normalized induction index to compare induction of RNAPII and transcription

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



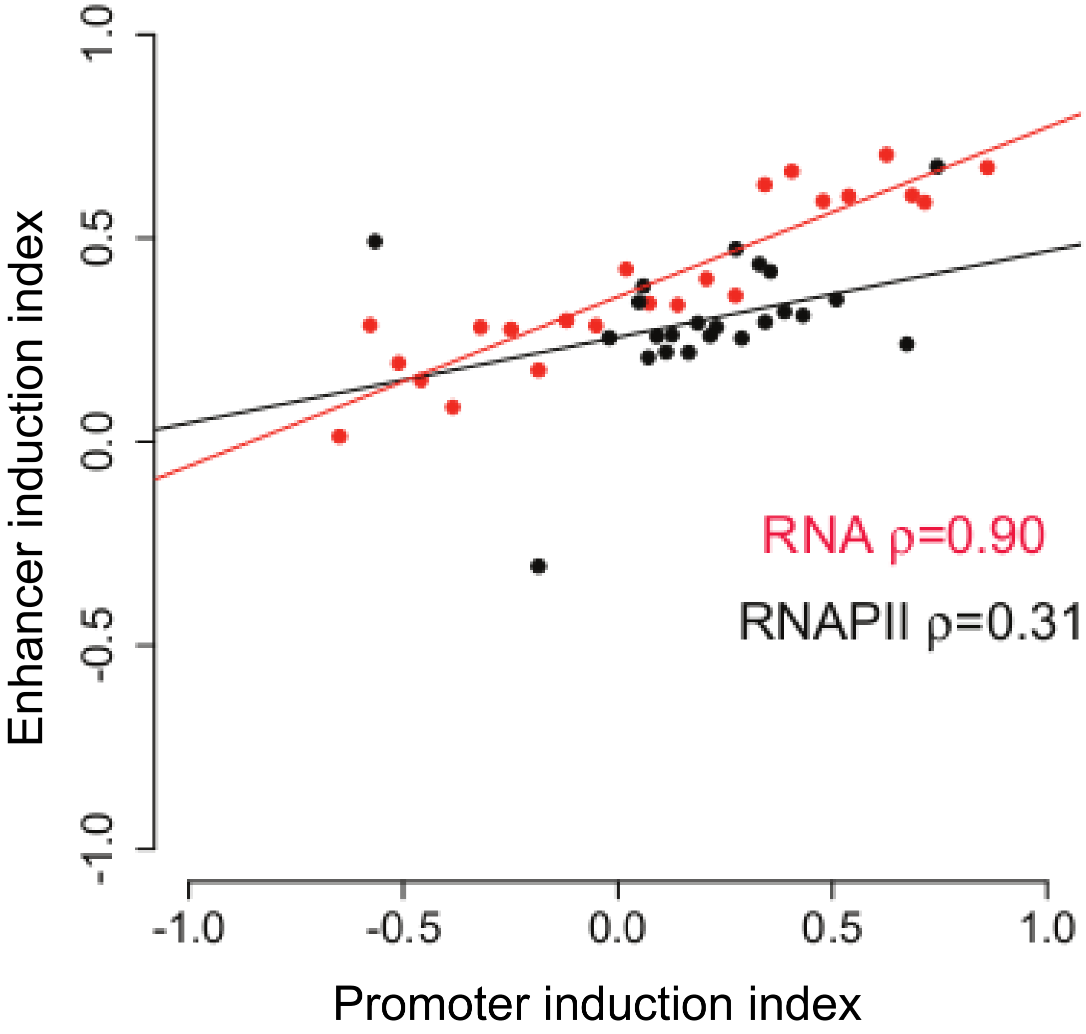
eRNA induction is correlated with induction of nearby mRNAs

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

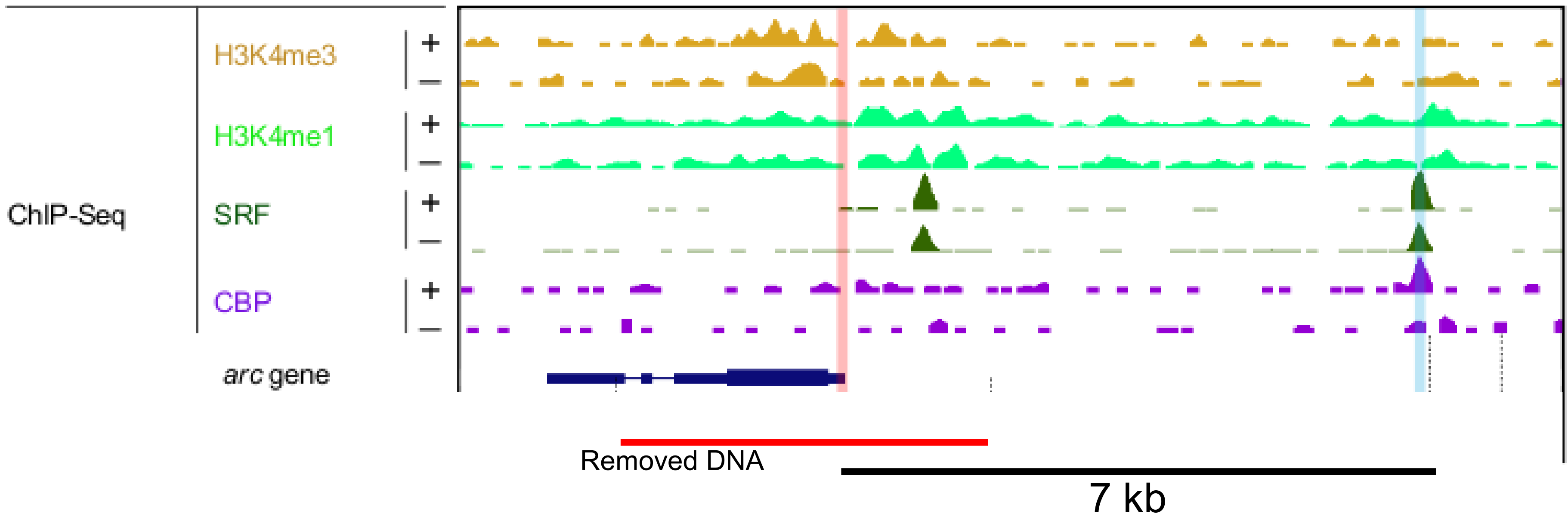


Do eRNAs depend on mRNAs?

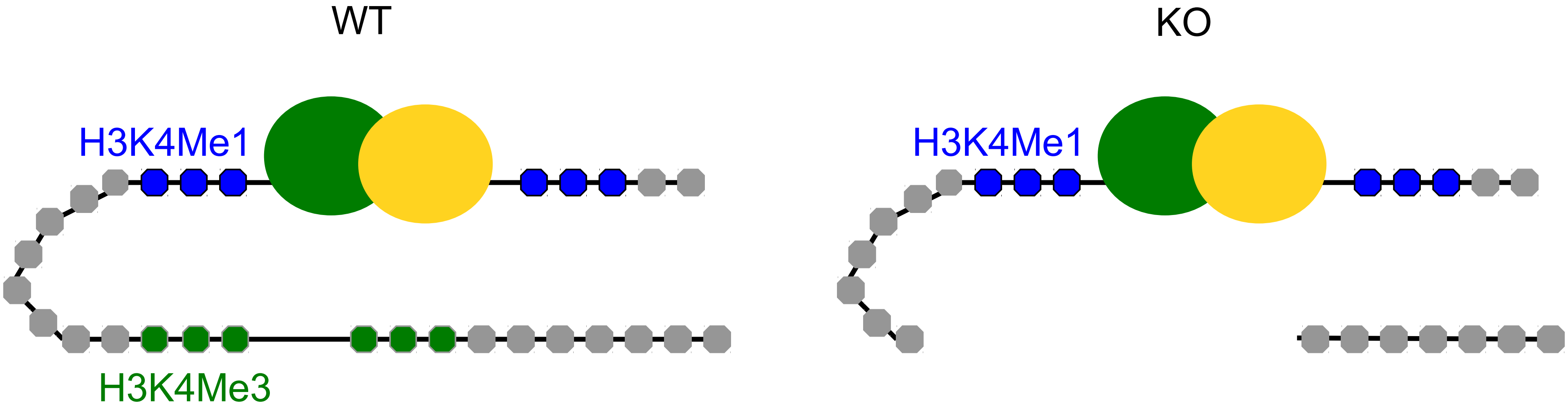
$$induction\ index = \frac{KCl - un}{KCl + un}$$



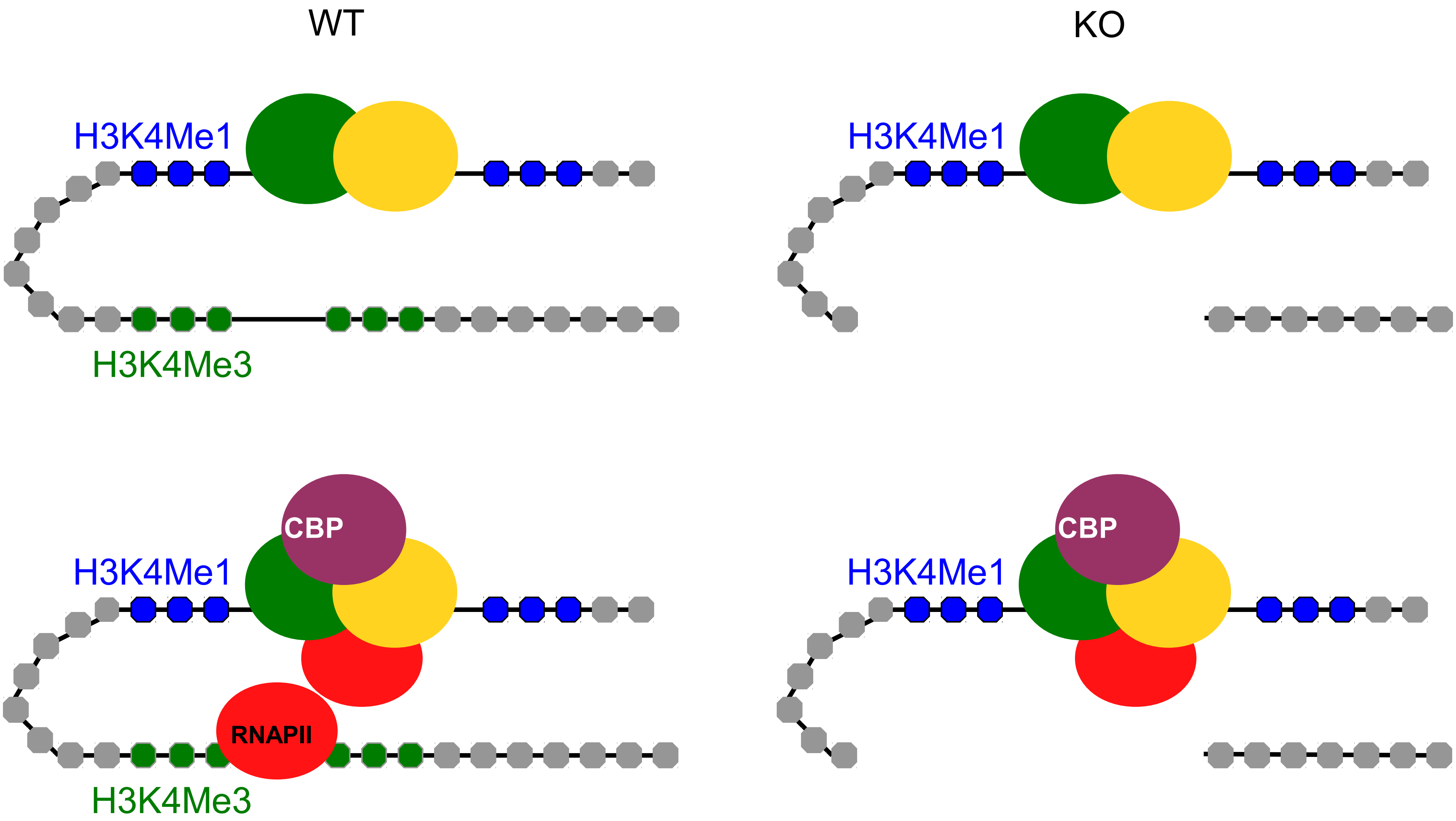
Knock-out experiment of the *arc*-promoter



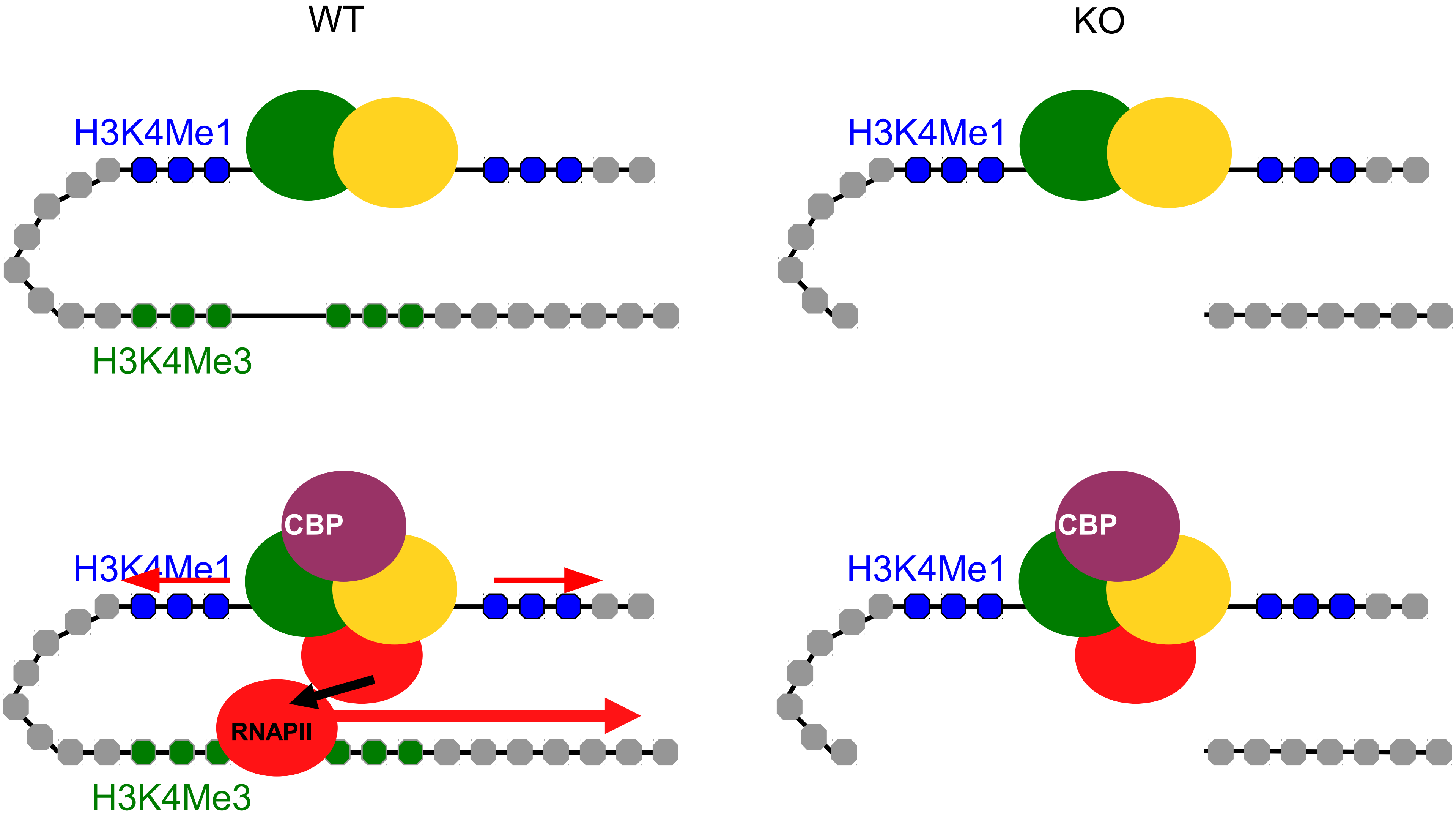
Are eRNAs independent of the promoter?



RNAPII increases 5-fold in both WT and KO



eRNAs are not present in KO



Summary

- Identified ~12k activity-dependent enhancers
- Discovered and quantified novel mechanisms
  - Identified enriched motifs and bound TFs

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

doi:10.1038/nature09033

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]

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]


AS

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]*

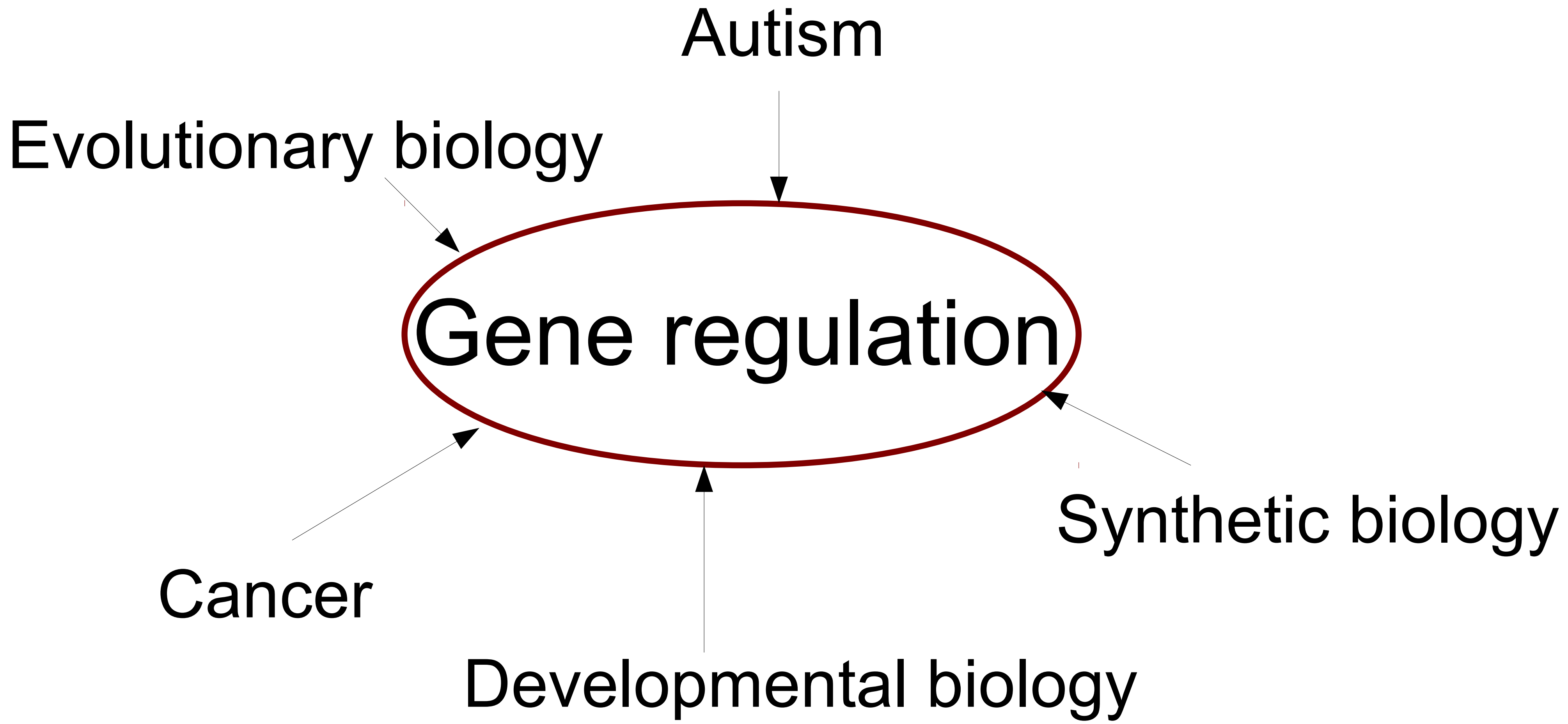

LETTER

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]

Why is gene regulation important?



Autism

Evolutionary biology

Gene regulation

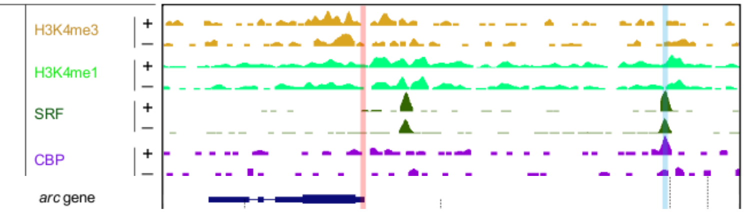
Synthetic biology

Cancer

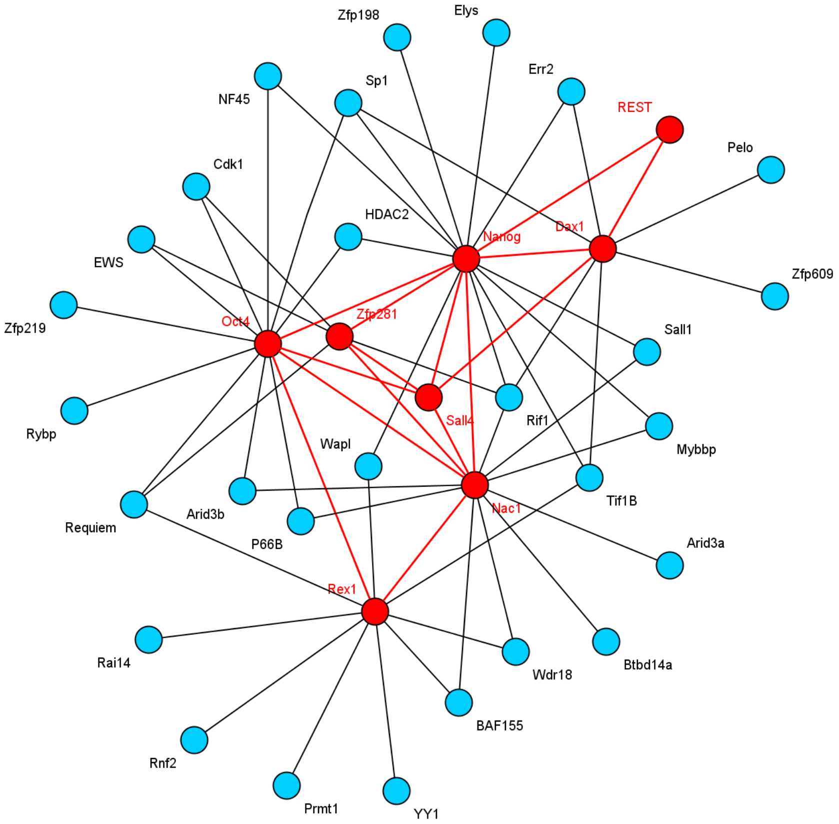
Developmental biology

Future Work: Organizing principles of the genome

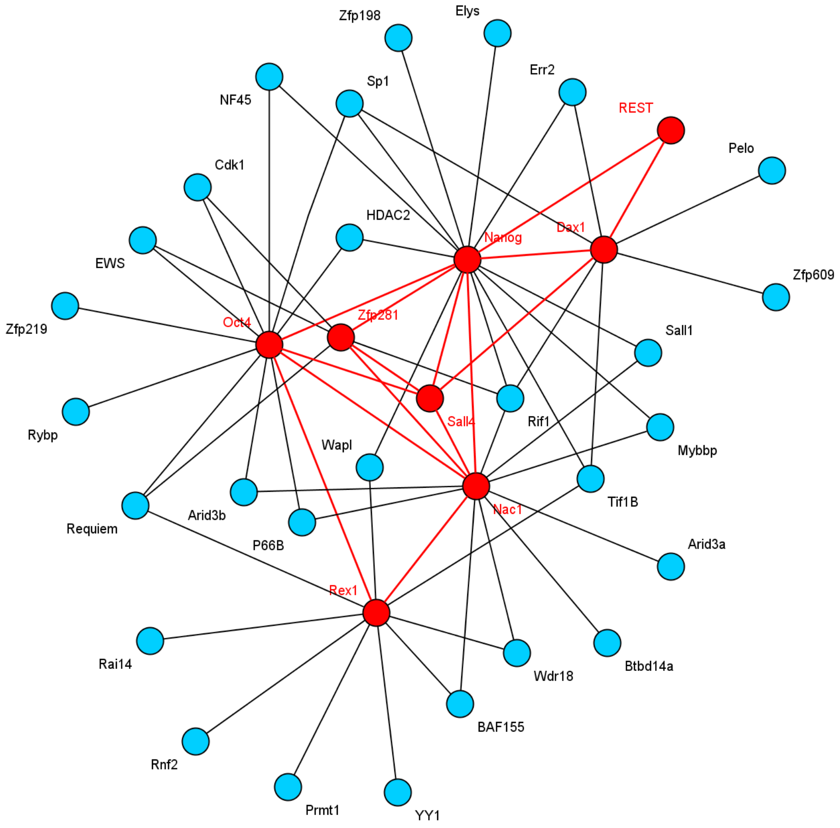
- Use genome-wide data to develop systems biology type models of gene regulation



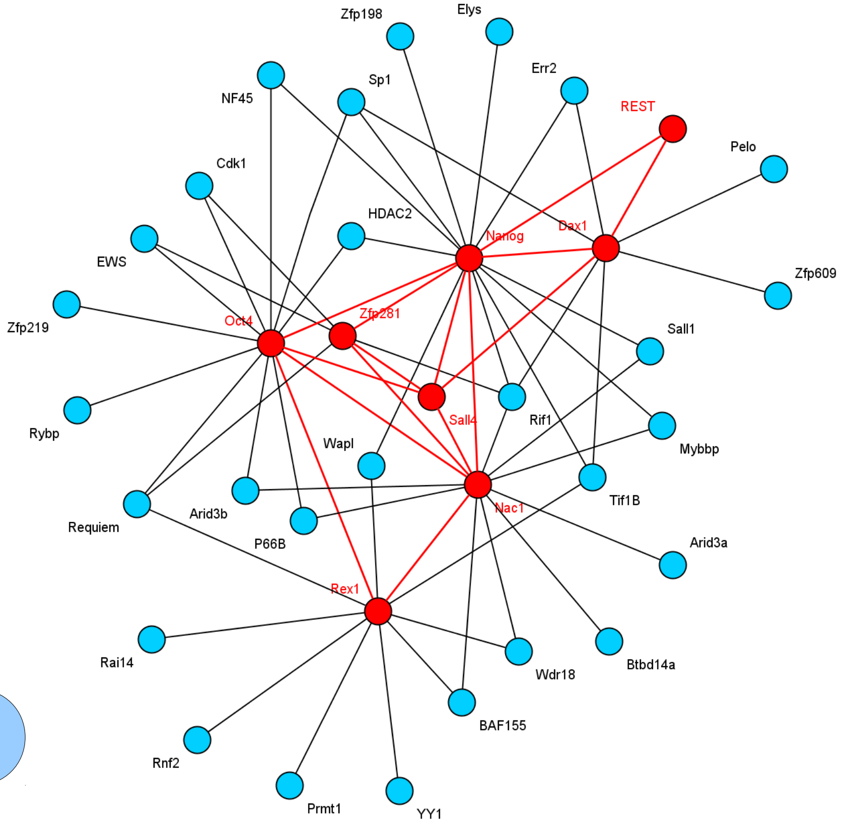
+

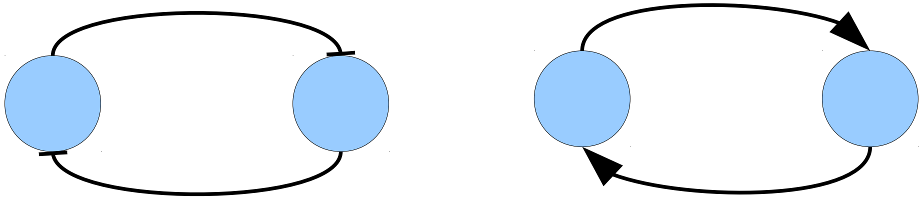


Topology is not sufficient for understanding function of gene regulatory networks

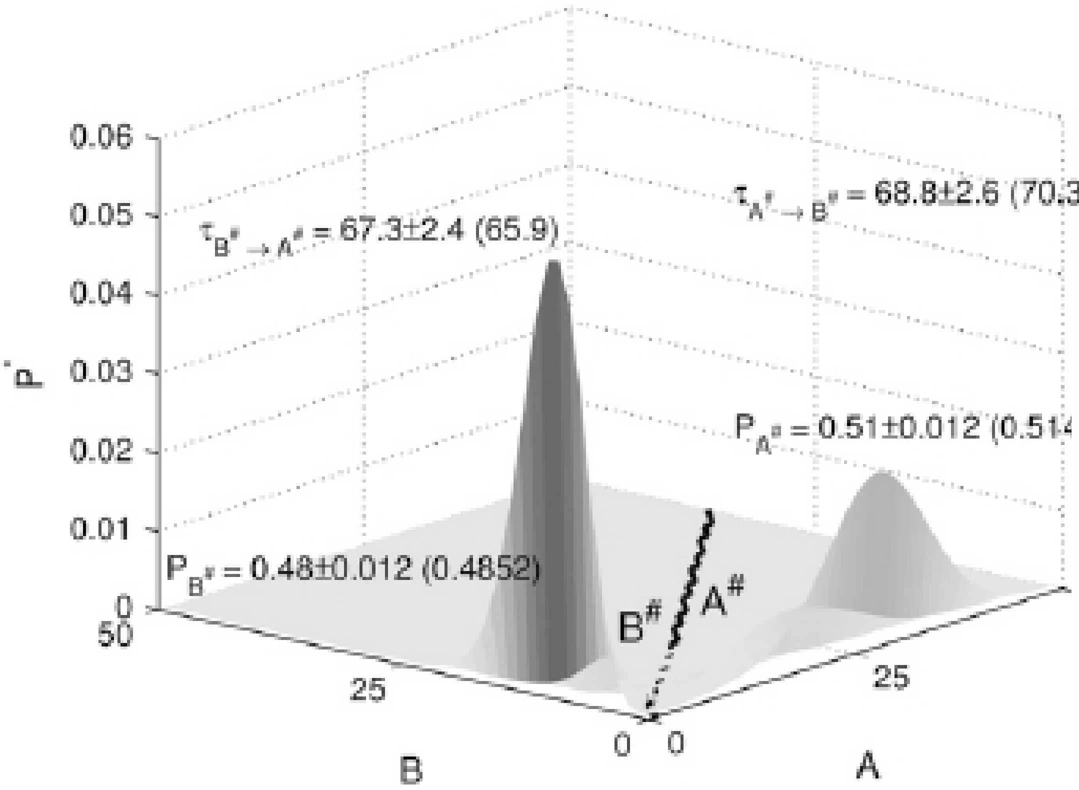


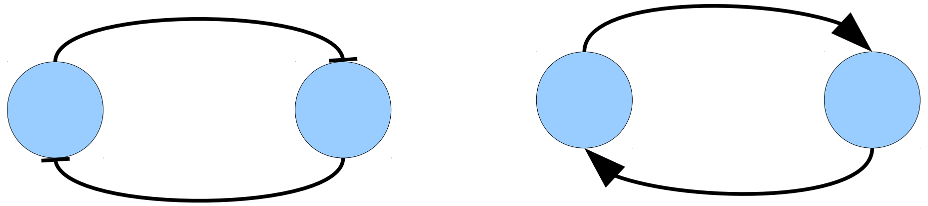
What are the dynamical properties of gene regulatory networks?

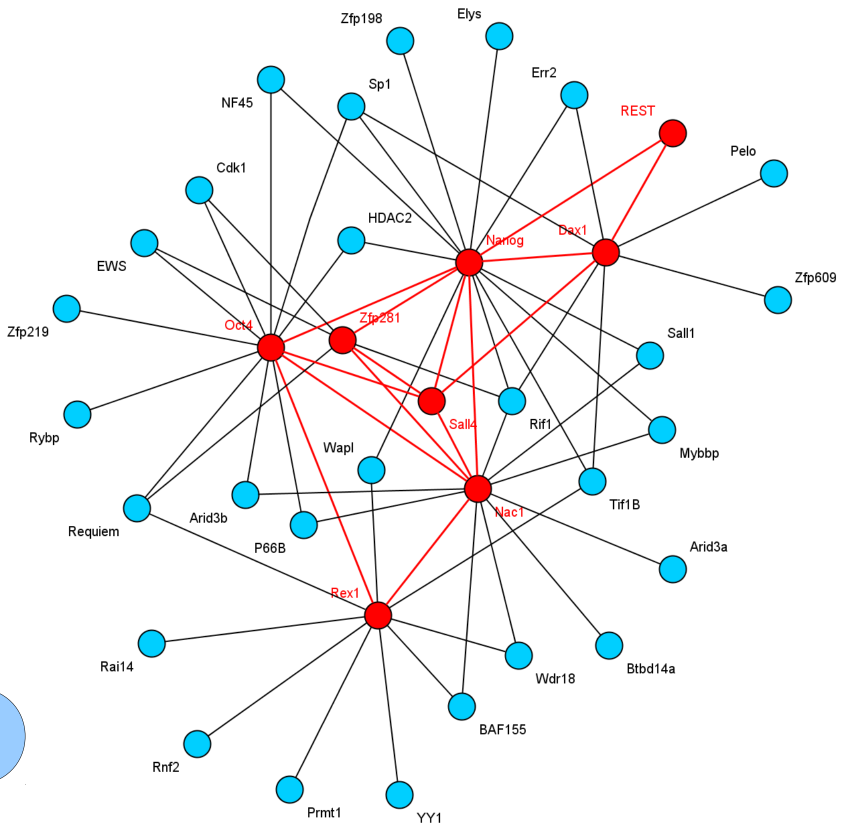




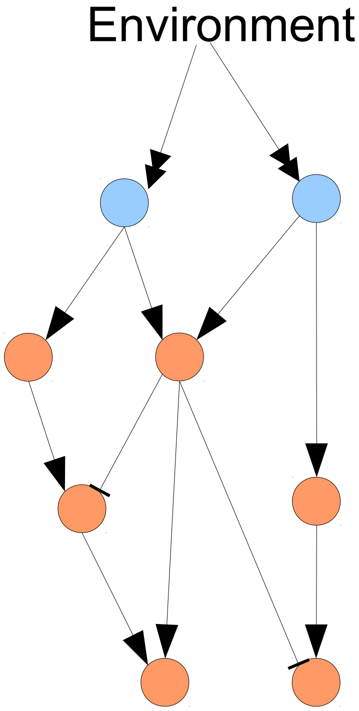
What is the role of noise in gene regulation?





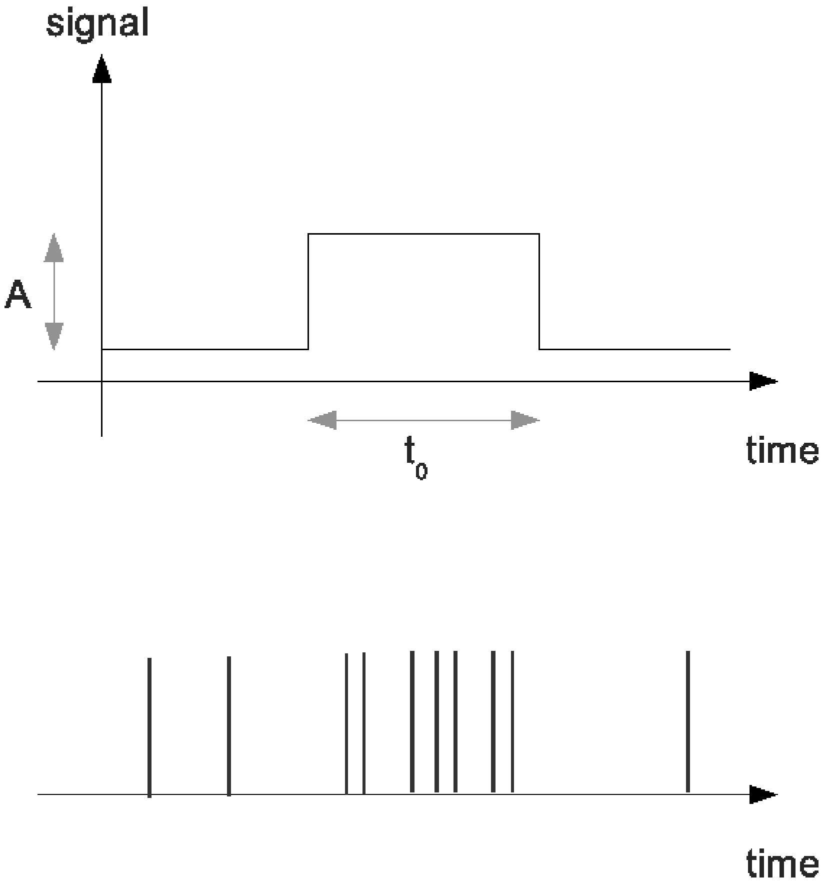


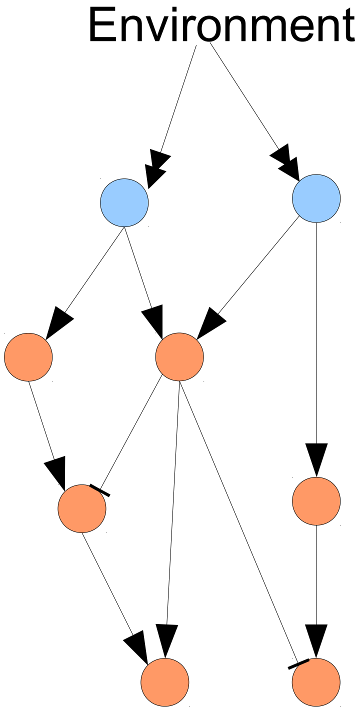
How is information propagated in gene regulatory networks?



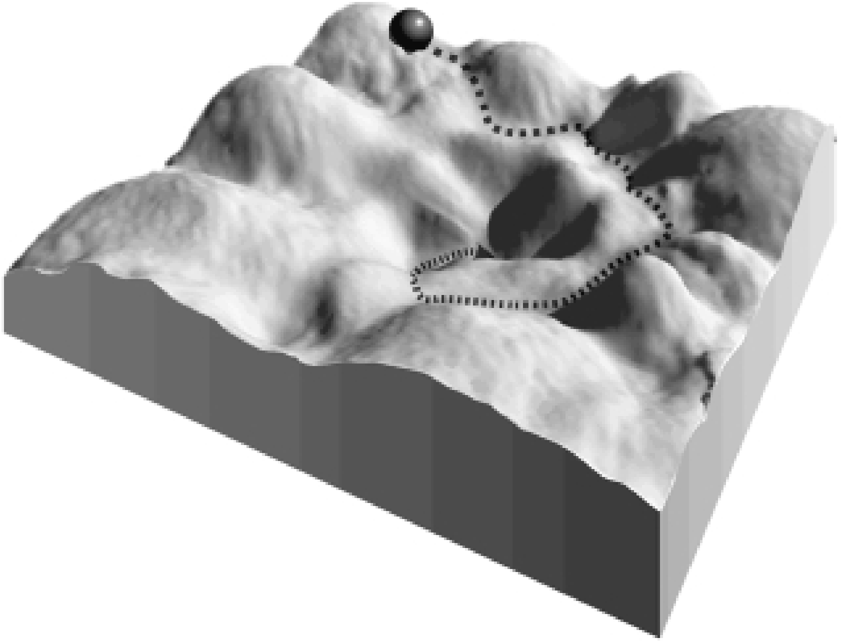
How can the input signals be inferred from observing the mRNA levels?

IMPORTANT

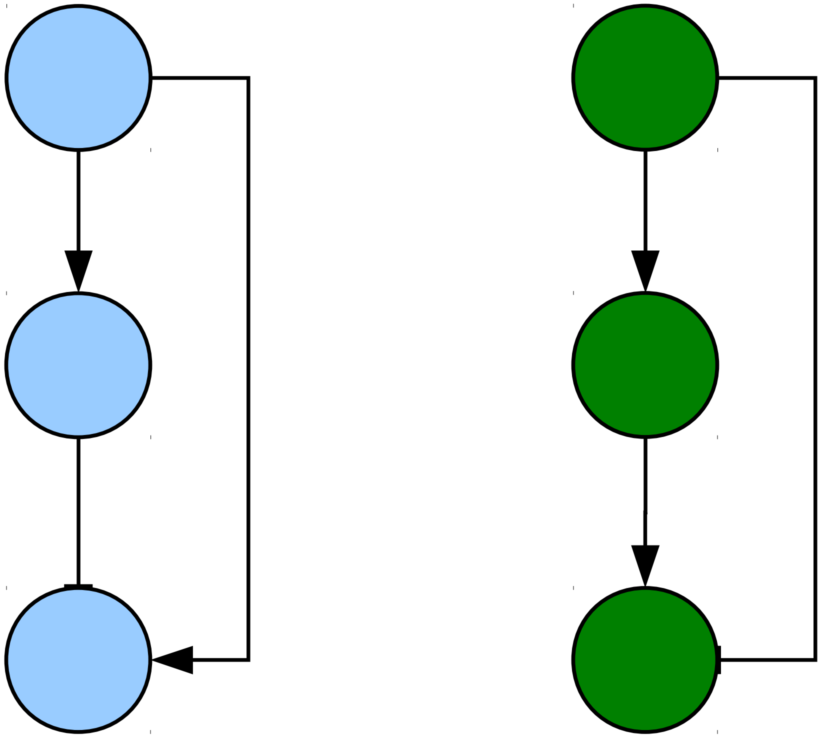




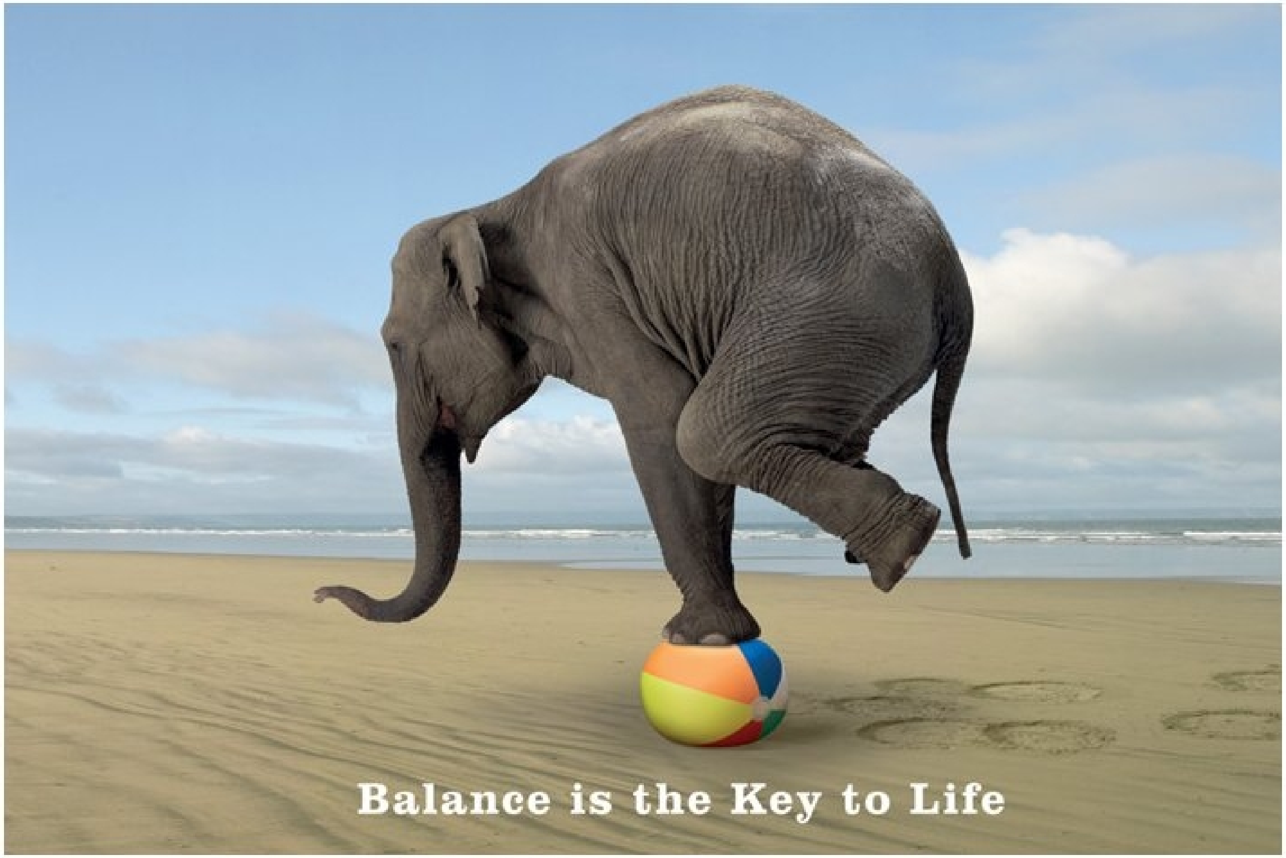
What are the limitations on control in gene regulatory networks?



What are different regulatory mechanisms optimized for?

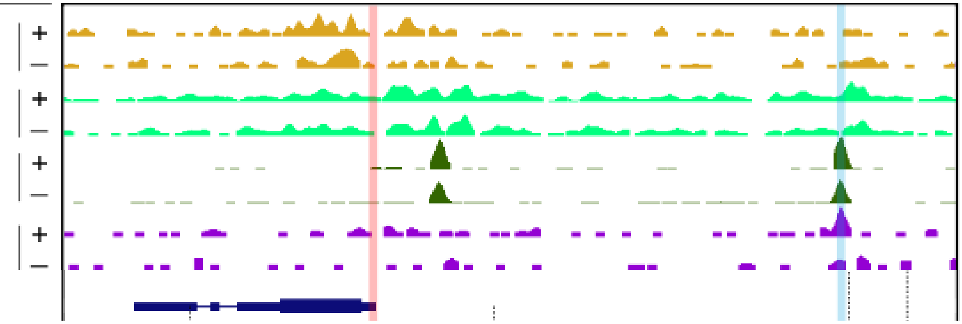


How robust is the system with respect to parametric perturbations?

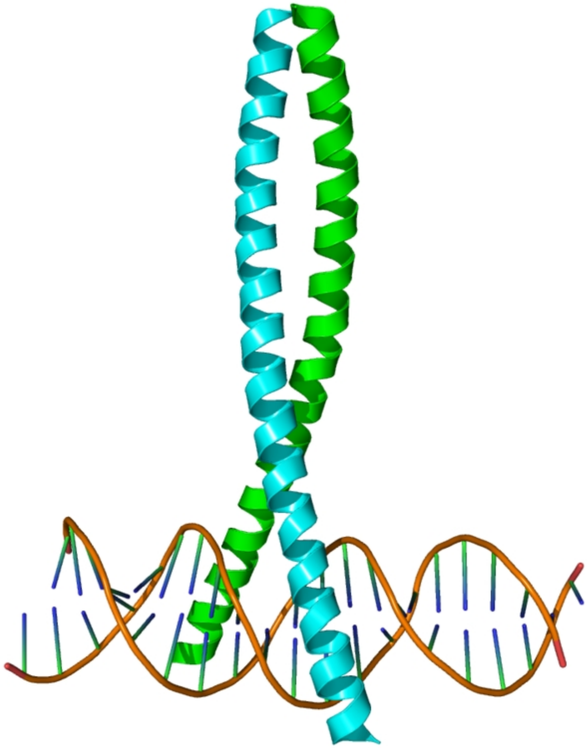


What is the biophysical basis of TF binding?

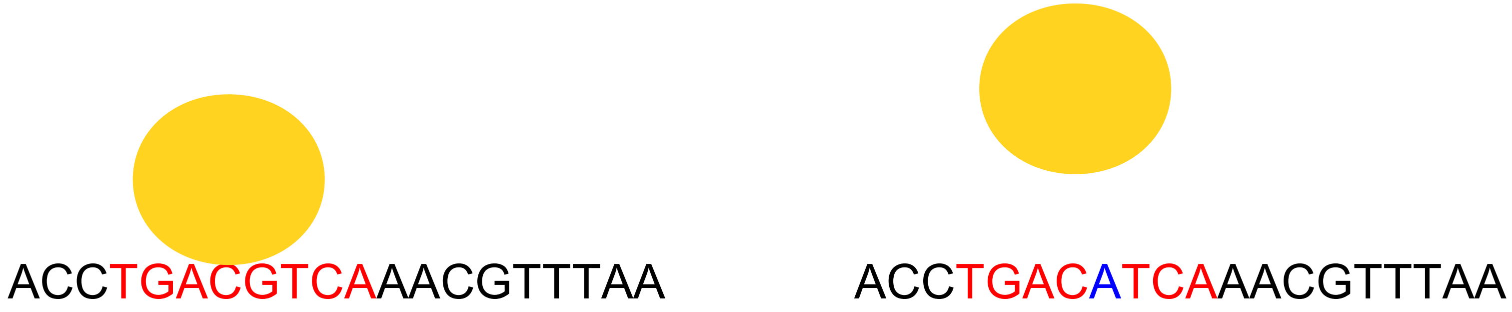
- X-ray structures
- ChIP-Seq binding



+



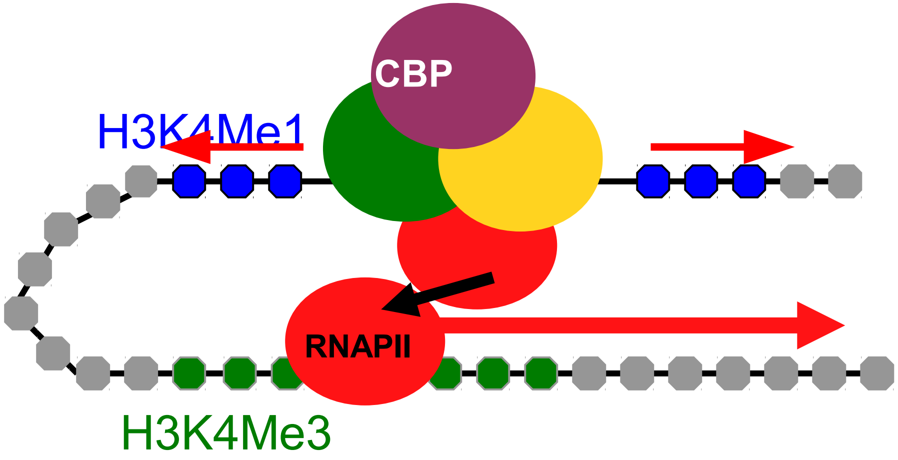
What is the impact of SNPs on TF binding?

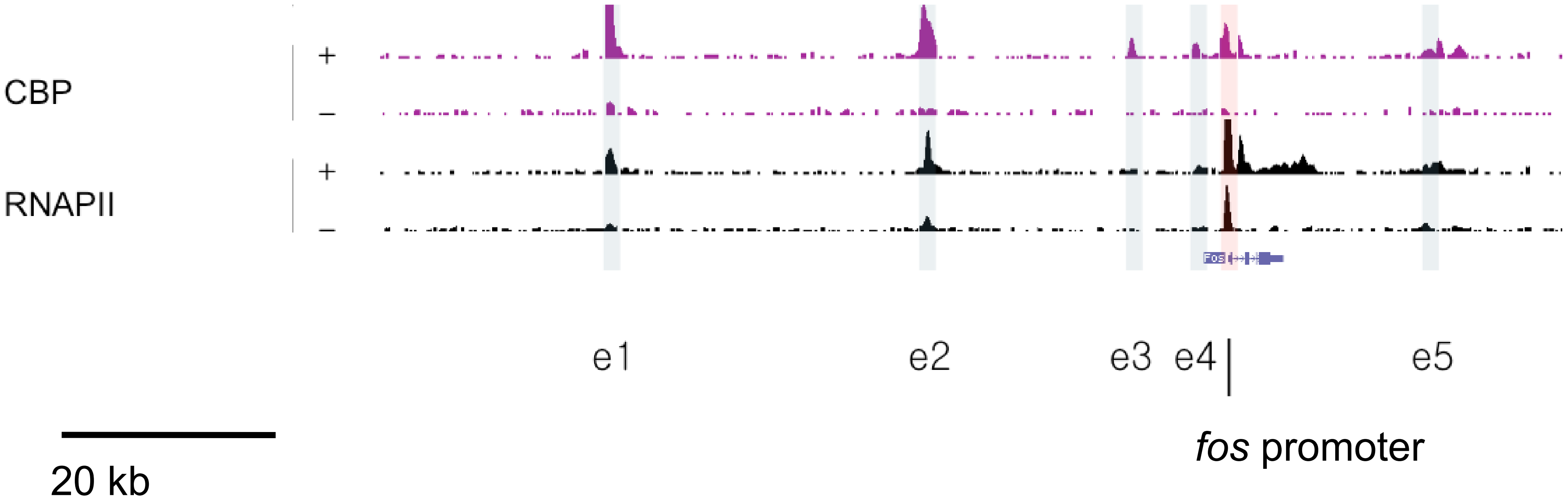


ACCTGACGTCAAACGTTTAA

ACCTGACATCAAACGTTTAA

What is the biophysical basis of DNA looping?





Acknowledgements

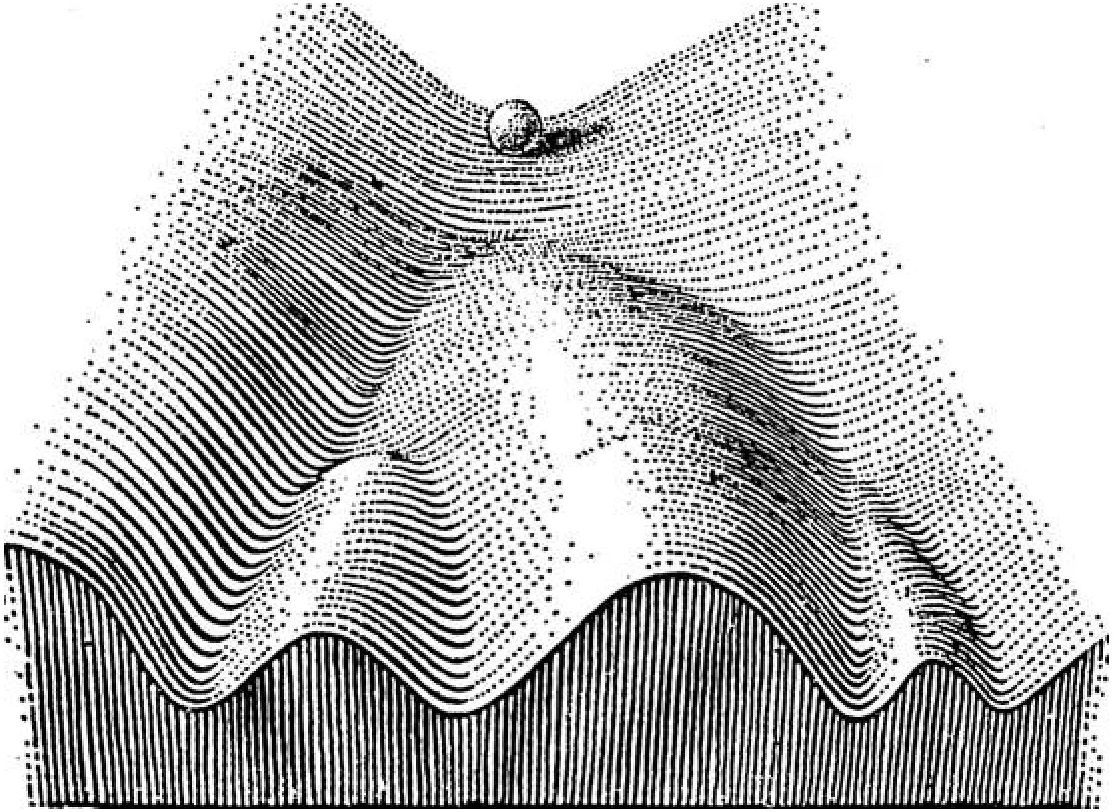
- Gabriel Kreiman
- Jesse Gray
- Tae-Kyung Kim
- Athar Malik
- Michael Greenberg

Thank You

?

Stochastic models of gene expression

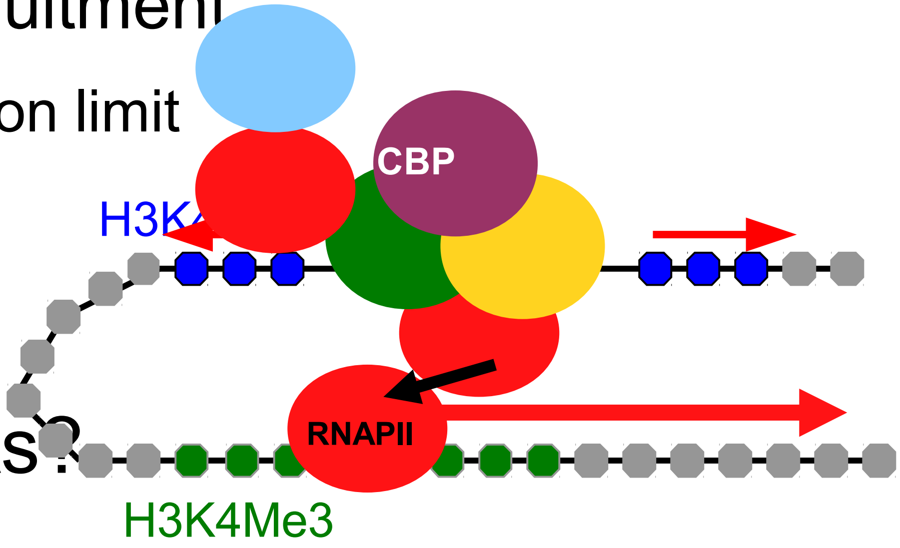
- Transitions between stable states



Waddington, 1953

Enzymes piggyback on the polymerase

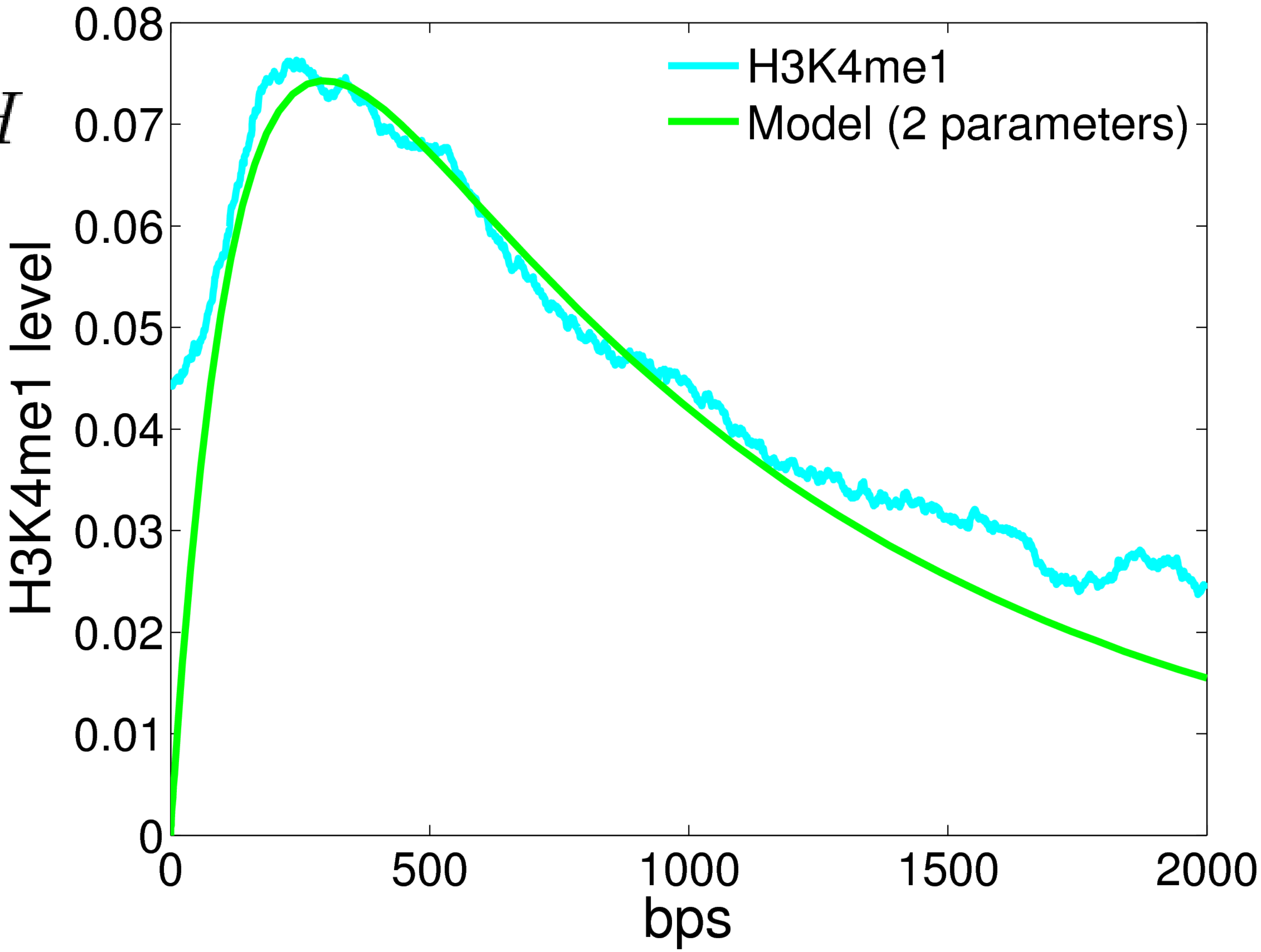
- What is the function of RNAPII at enhancers?
  - Increase rate of RNAPII recruitment
    - Possibly faster than diffusion limit
  - Faster rise-time
  - Reduced noise
- What is the function of eRNAs?
  - Noise
  - Transcription establishes histone modifications
  - Transcript has function



Establishing H3K4me1 levels at enhancers

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

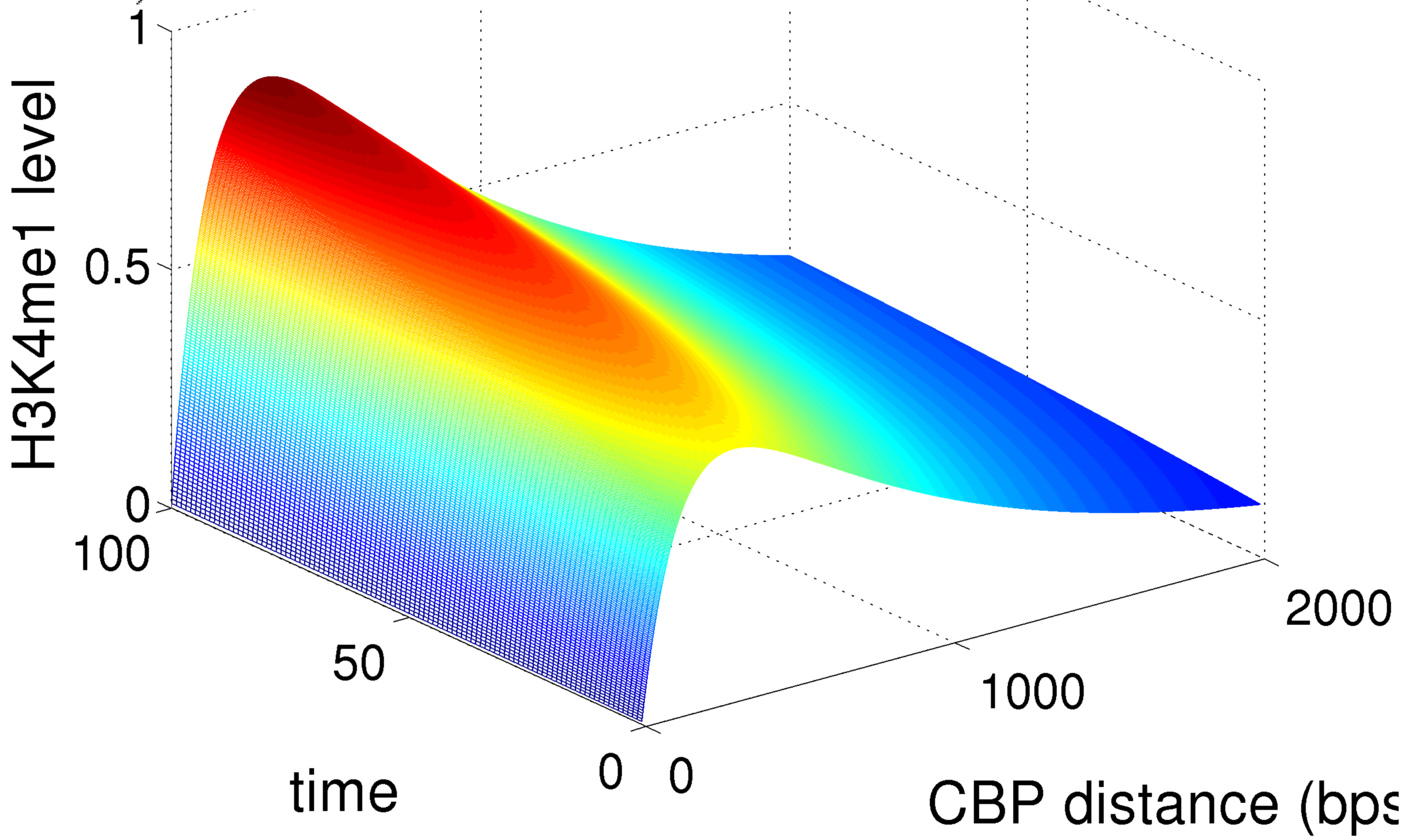
$$\frac{dH}{dx} = \kappa P(x) - \mu H$$



A PDE for histone levels

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

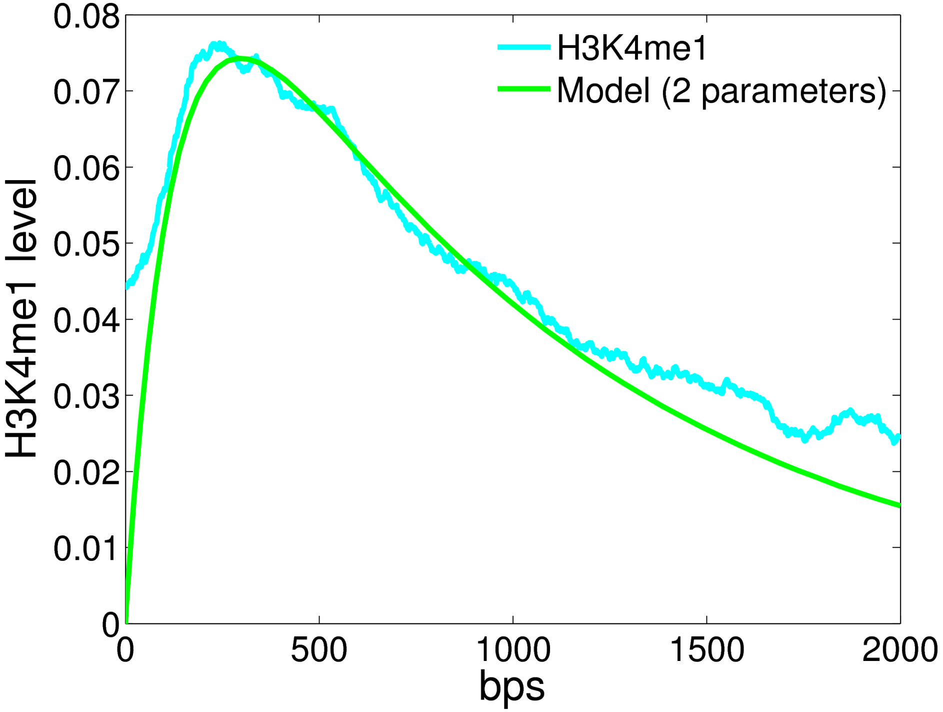
$$H(x,t) = \frac{k\kappa}{\mu_x(\mu_x - \lambda)}(e^{-\lambda x} - e^{-\mu_x x}) \times e^{-\mu_t t}$$

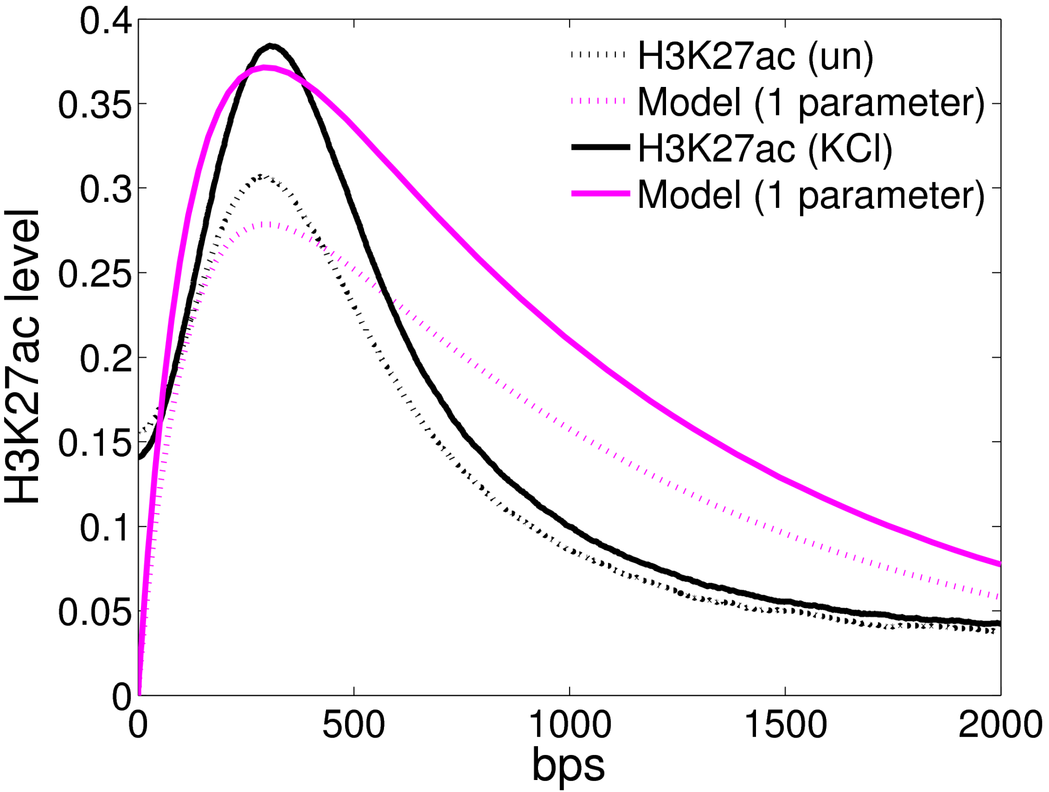


Histone methylation is not significantly changed, but histone acetylation is

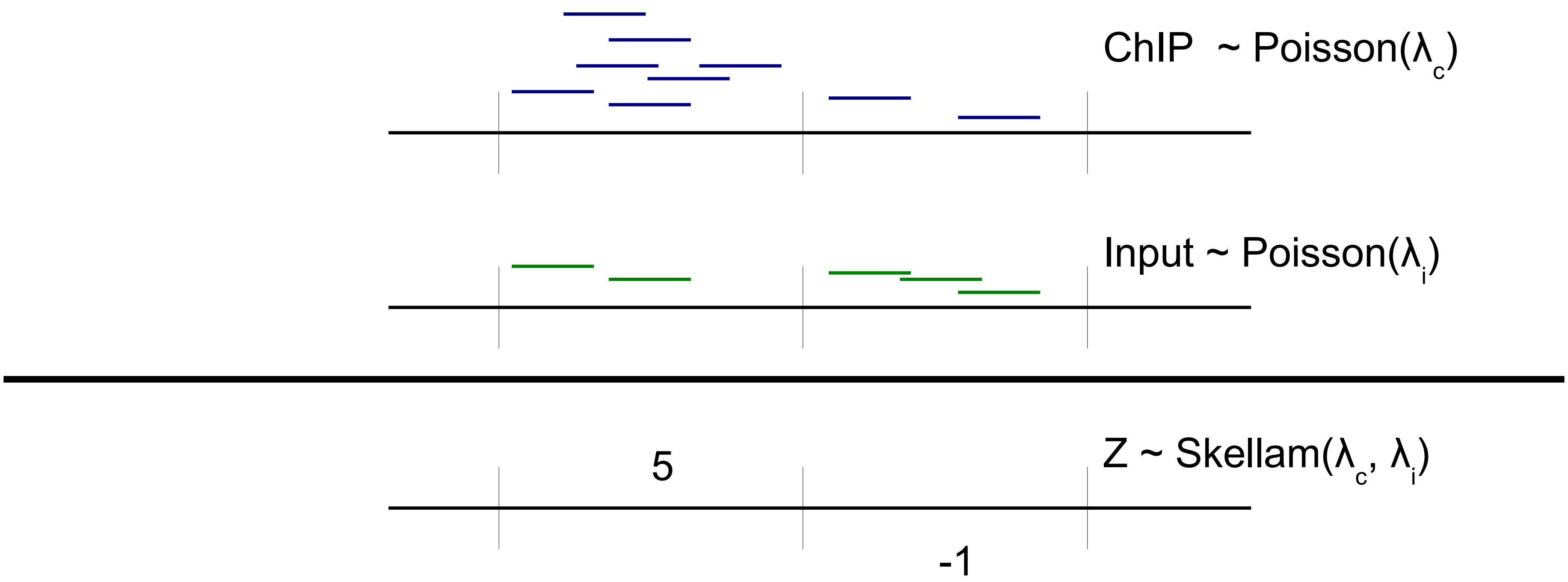
$$\frac{\partial H}{\partial x} + \frac{\partial H}{\partial t} = \kappa P(x,t) - \mu_x H - \mu_t H$$

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





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

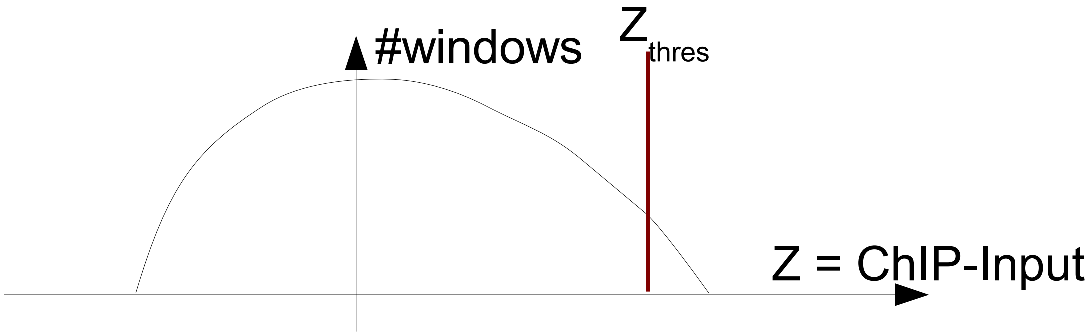


ChIP ~ Poisson(λ_c)

Input ~ Poisson(λ_i)

Z ~ Skellam(λ_c, λ_i)

- False Detection Rate (FDR) determine threshold



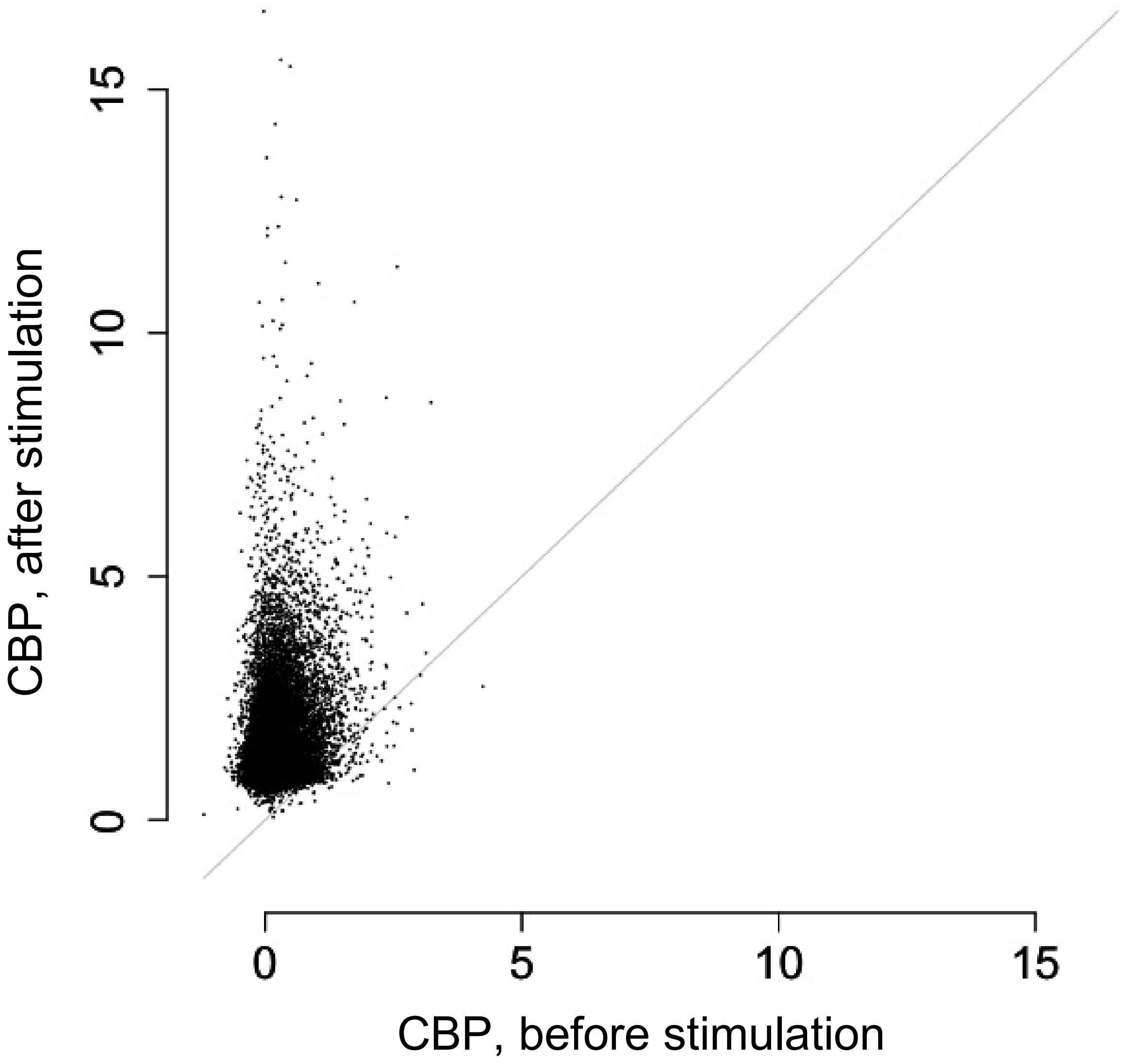
Use False Detection Ratio (FDR) to correct for multiple hypotheses

- 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
    - $Z_i \sim \text{Skellam}(z, \lambda_1, \lambda_2)$

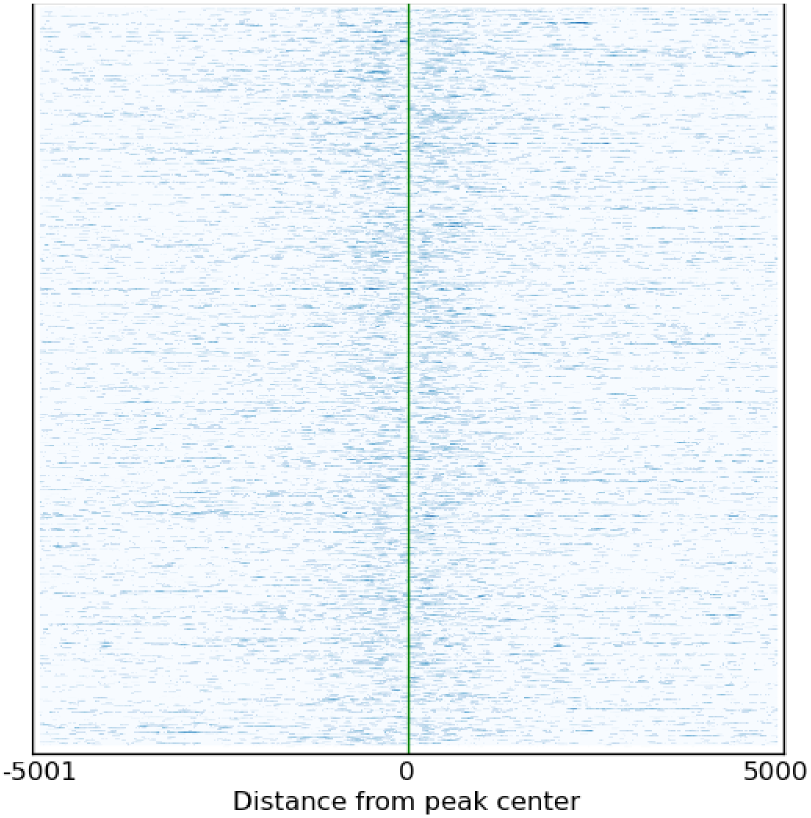
$$p(x) = e^{-(\lambda_1+\lambda_2)} (\lambda_1/\lambda_2)^{x/2} I_x(2\sqrt{\lambda_1\lambda_2})$$

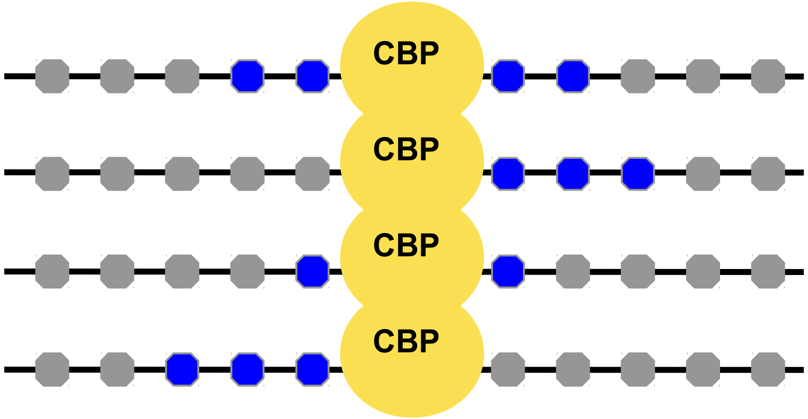
- Millions of windows need to be tested
  - FDR - expected fraction of false positives

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

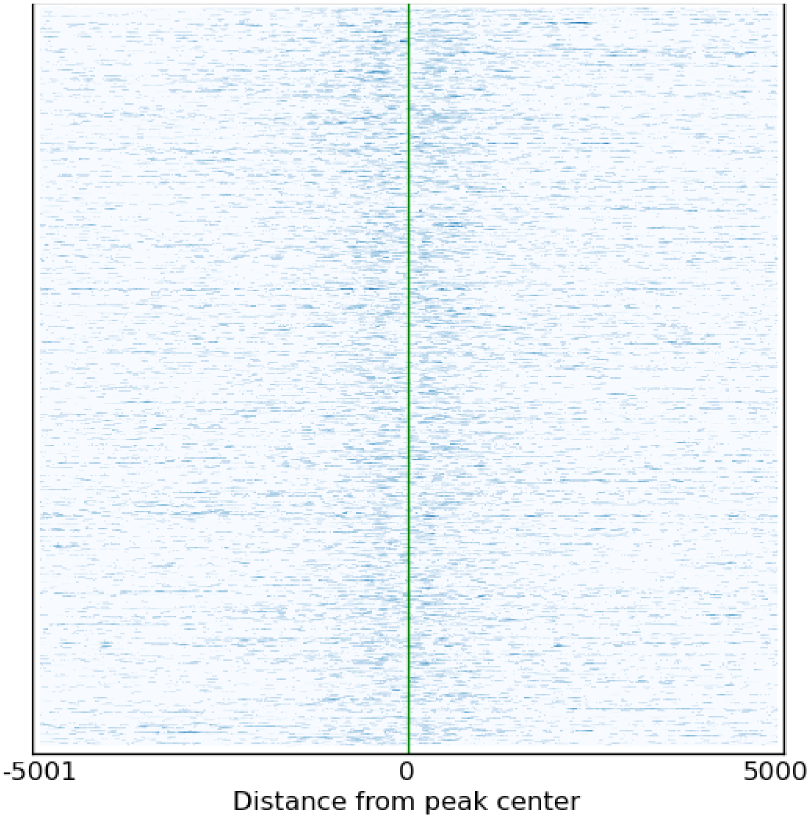


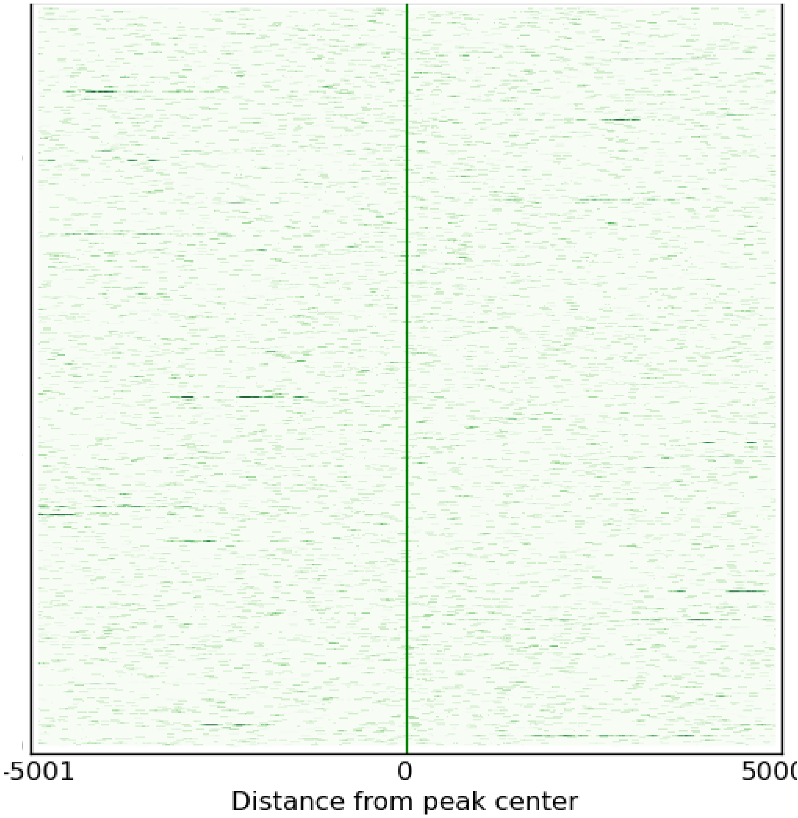
Aligning CBP peaks to calculate H3K4me1 binding profiles

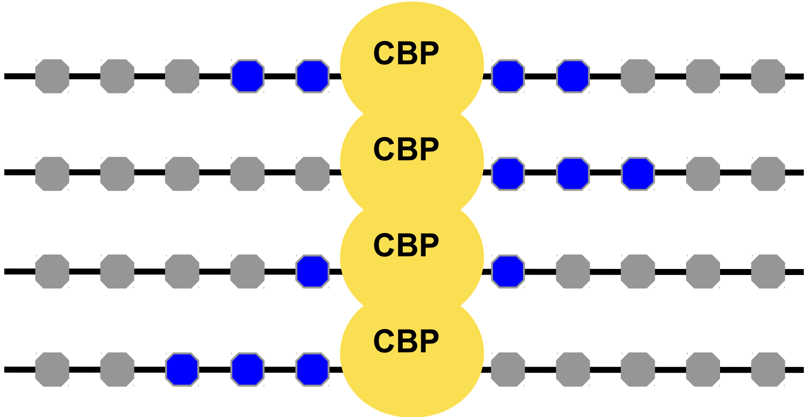




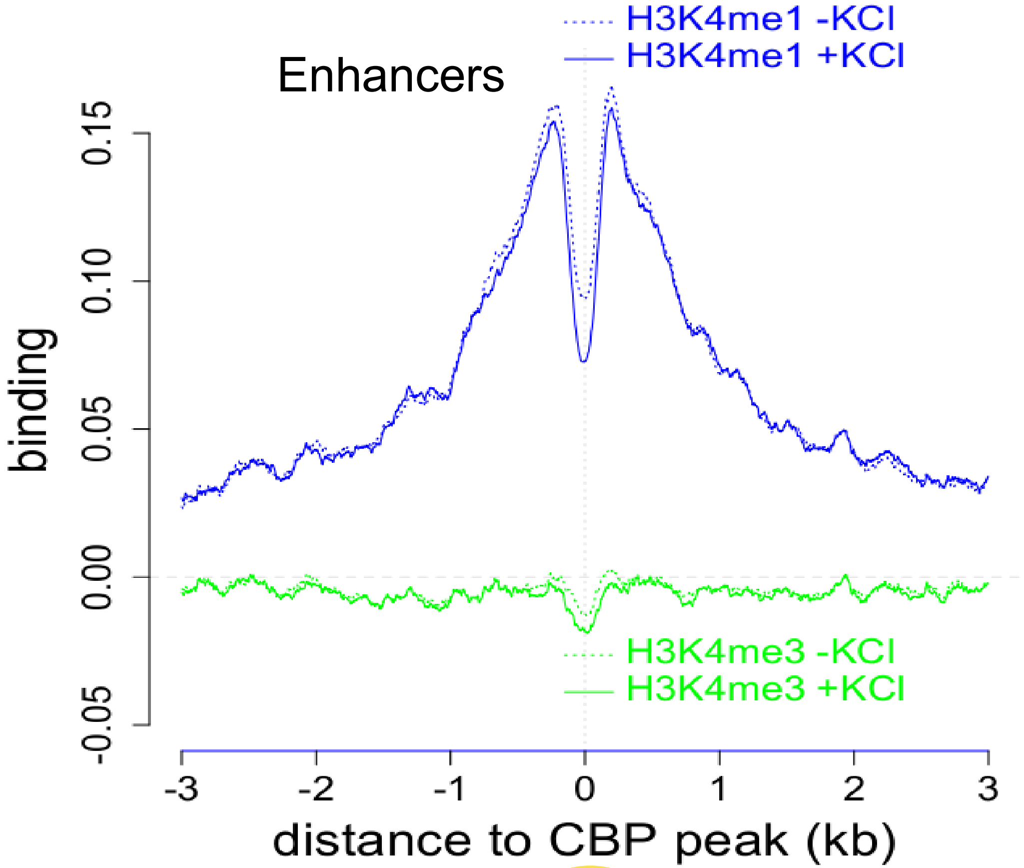
Aligning CBP peaks to calculate H3K4me1 and H3K4me3 binding profiles

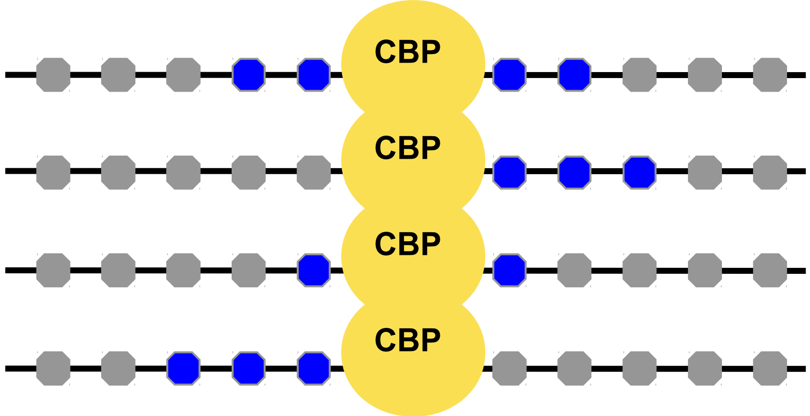




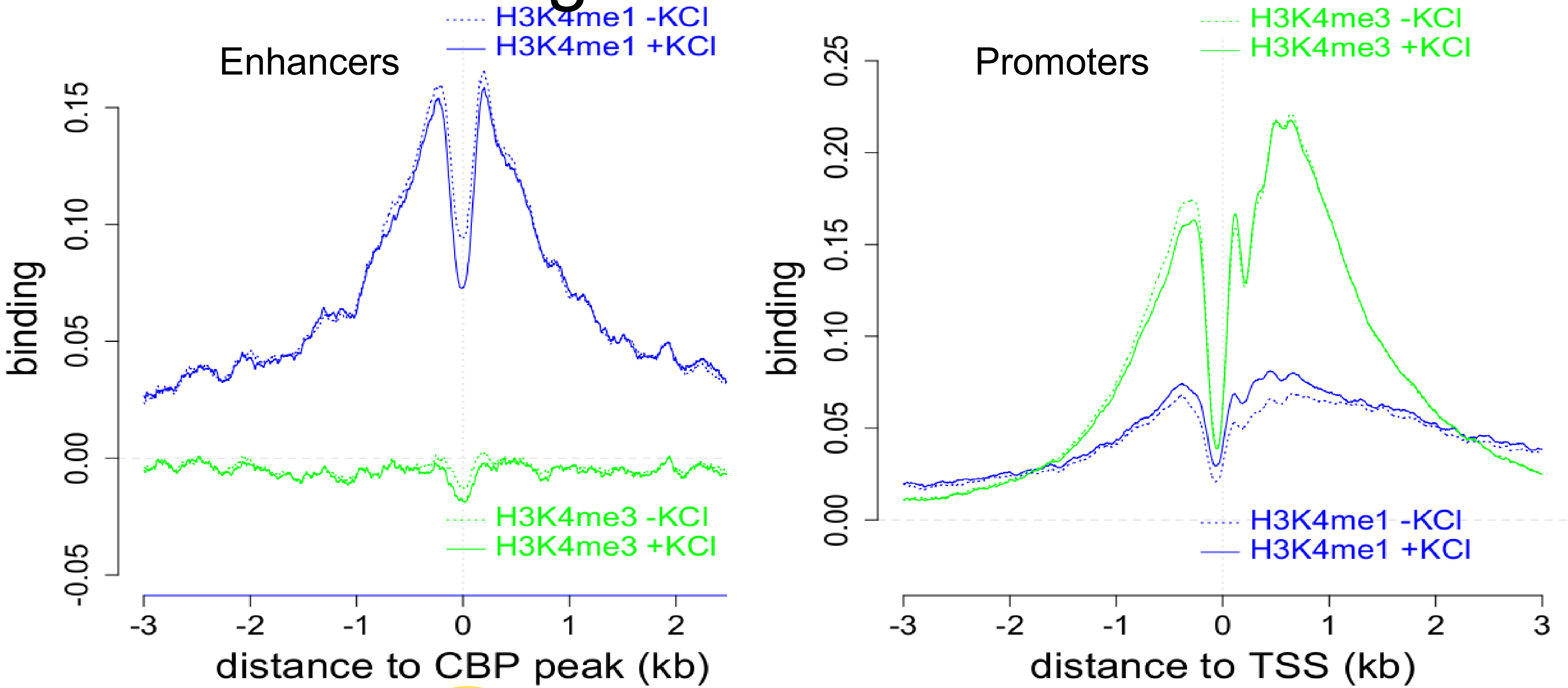


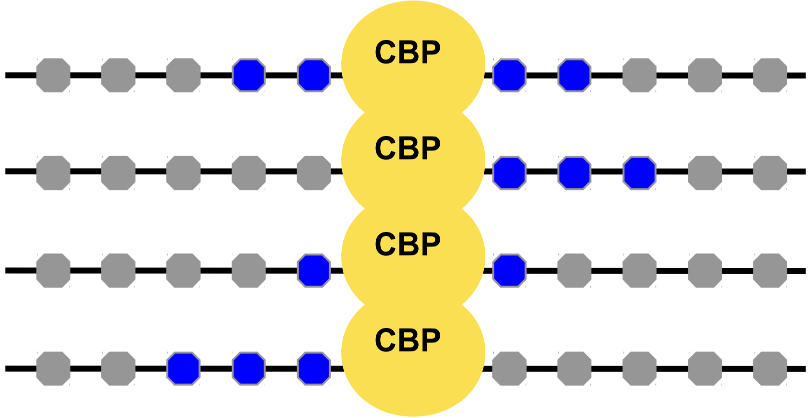
Enhancers have high levels of H3K4me1 and low levels of H3K4me3





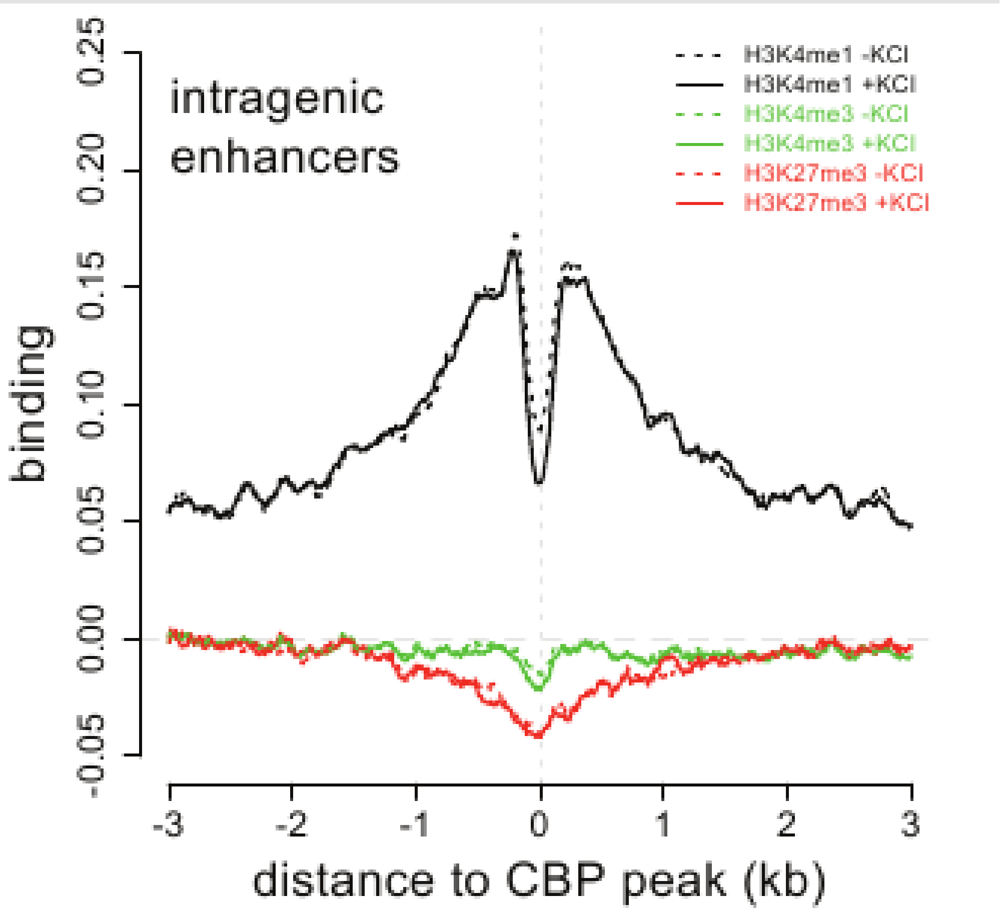
Transcription start sites have high levels of H3K4me1 and high levels of H3K4me3





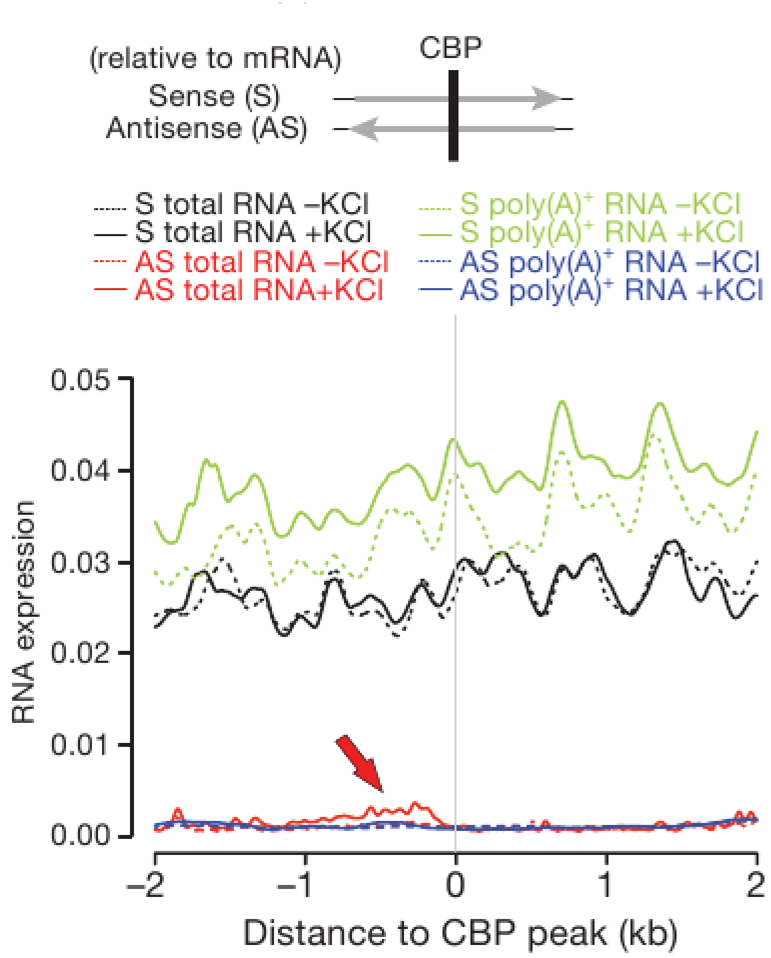
Intragenic enhancers

- ~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



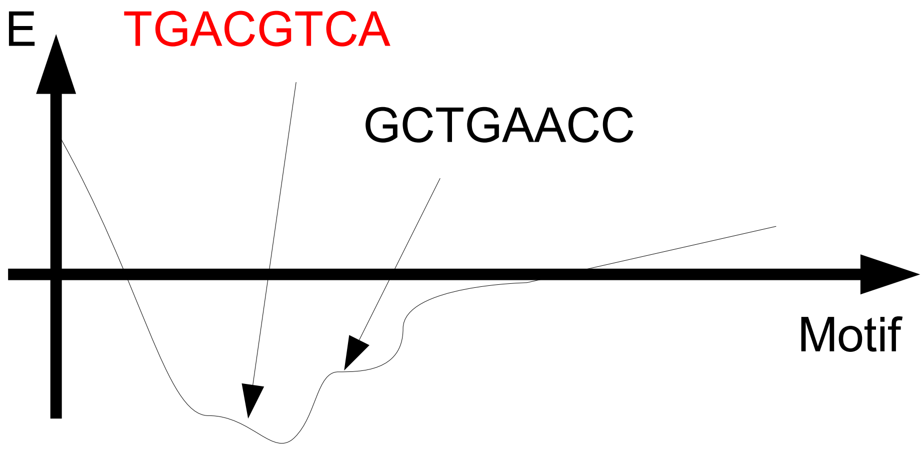
Can the read count be predicted from sequence motifs?

noise

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

#reads

Binding energy



Assume transfer function maximizes mutual information

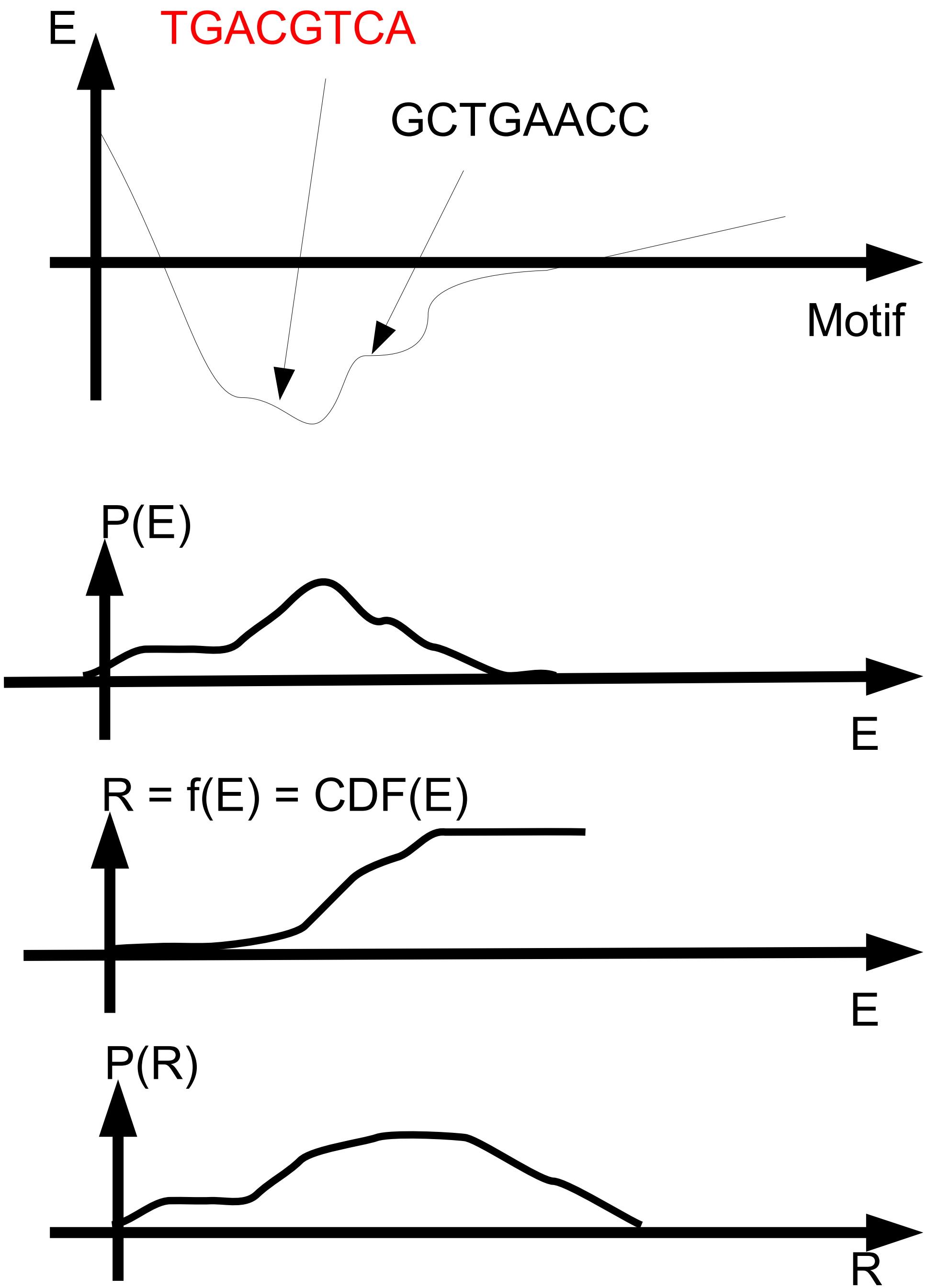
noise

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

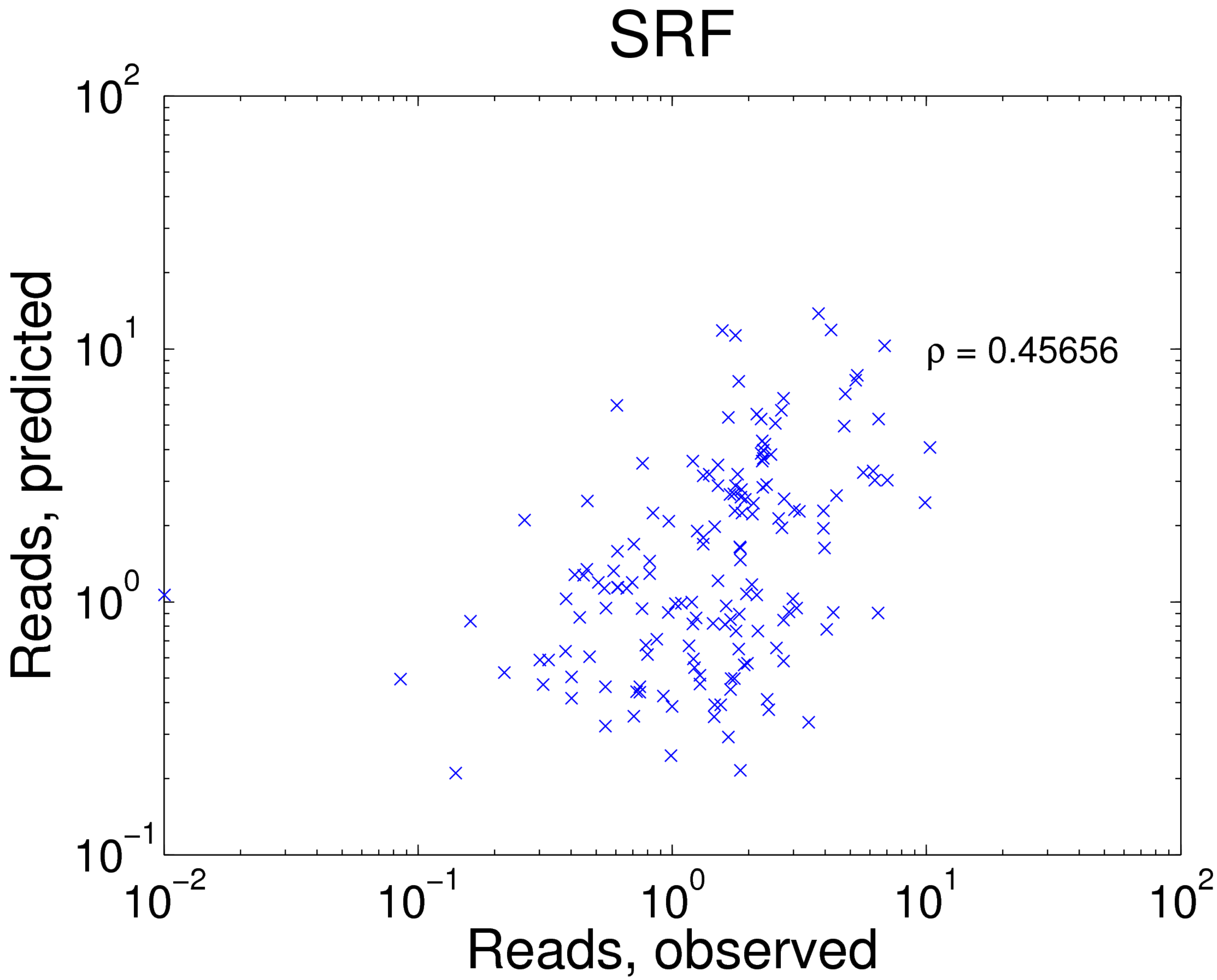
#reads

Binding energy (from JASPAR)

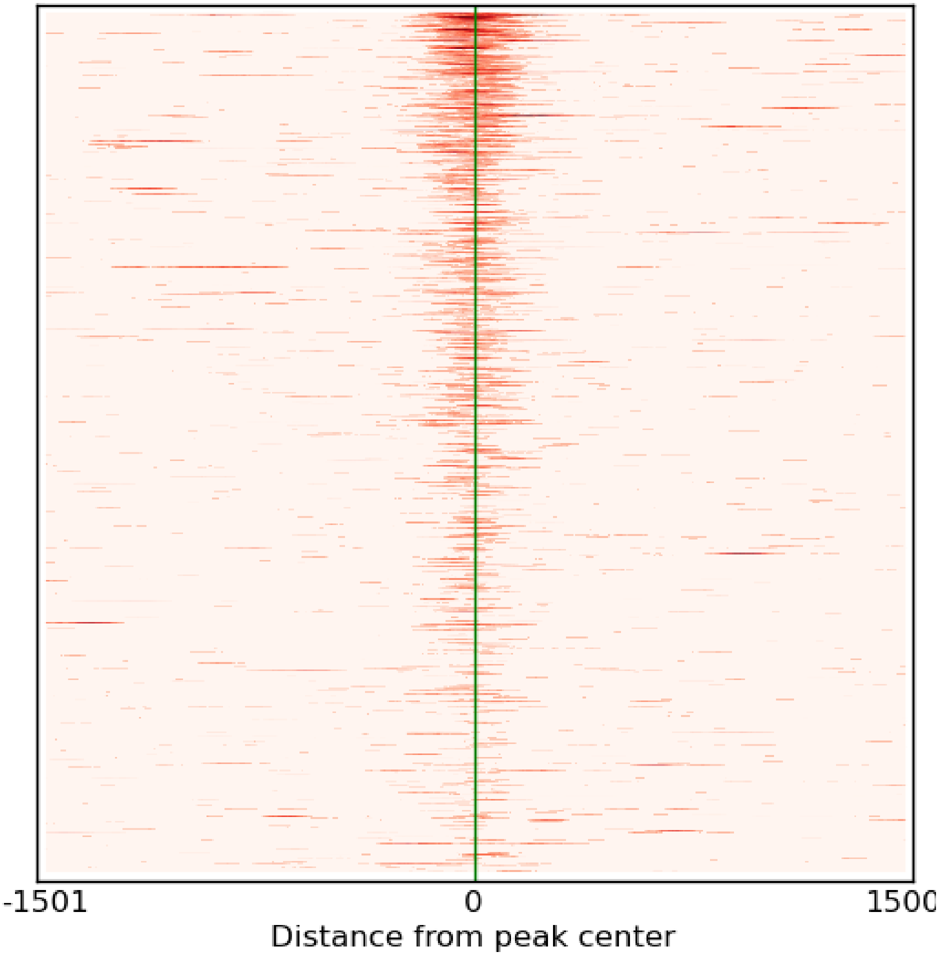
- *f* monotone
- max *I(R; E)*
- Noise small and gaussian



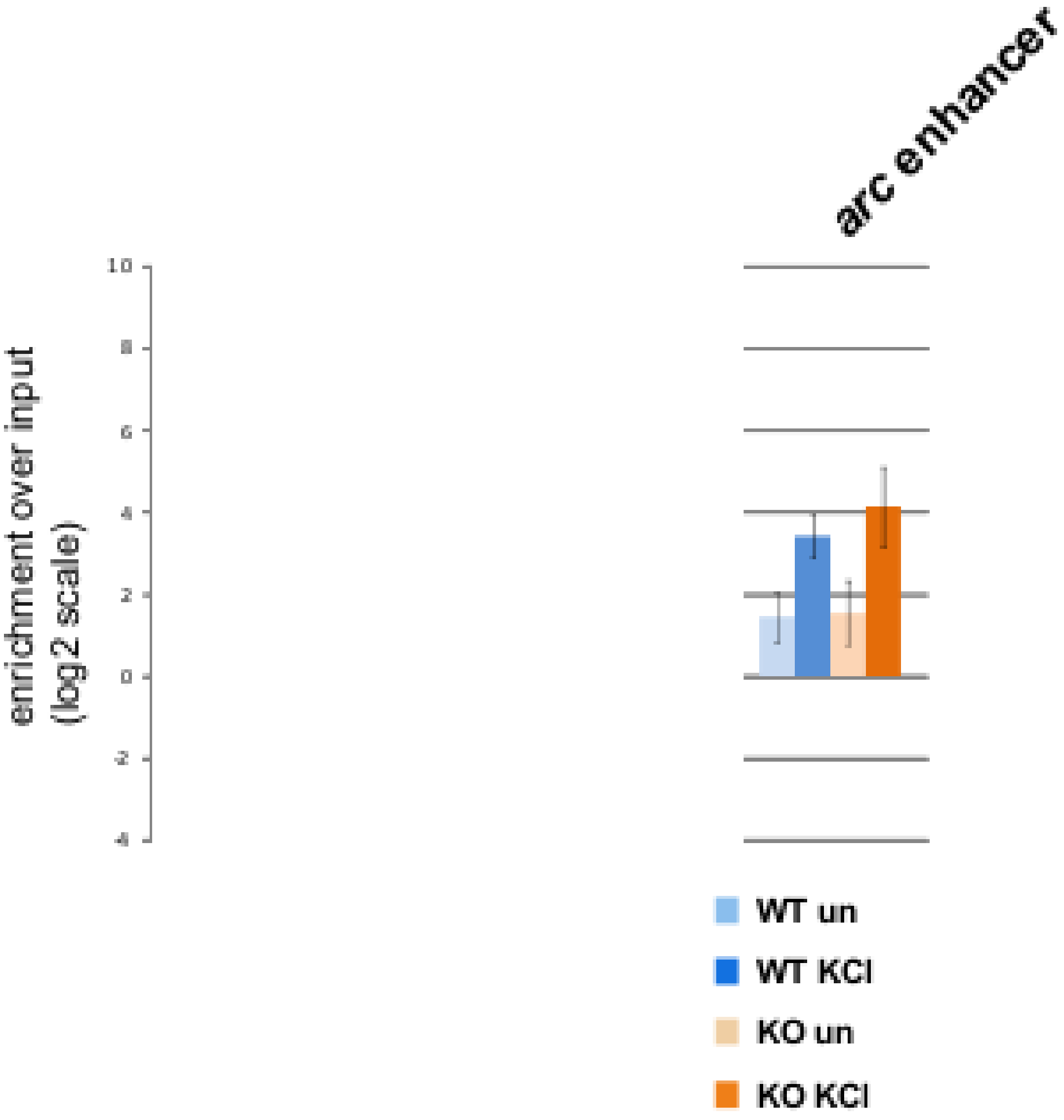
Number of reads can be predicted by binding energy

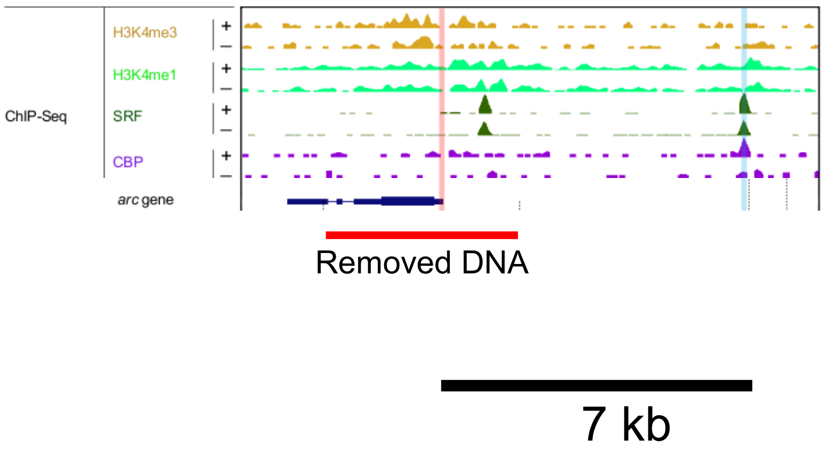


RNAPII binds at activity-dependent enhancers

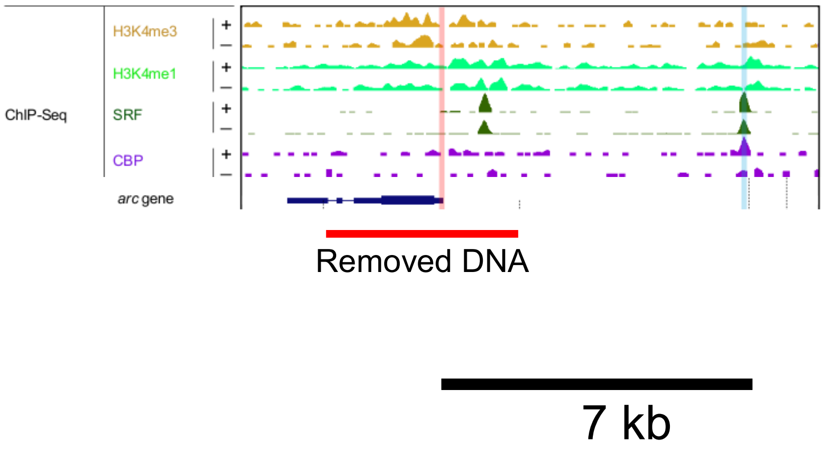


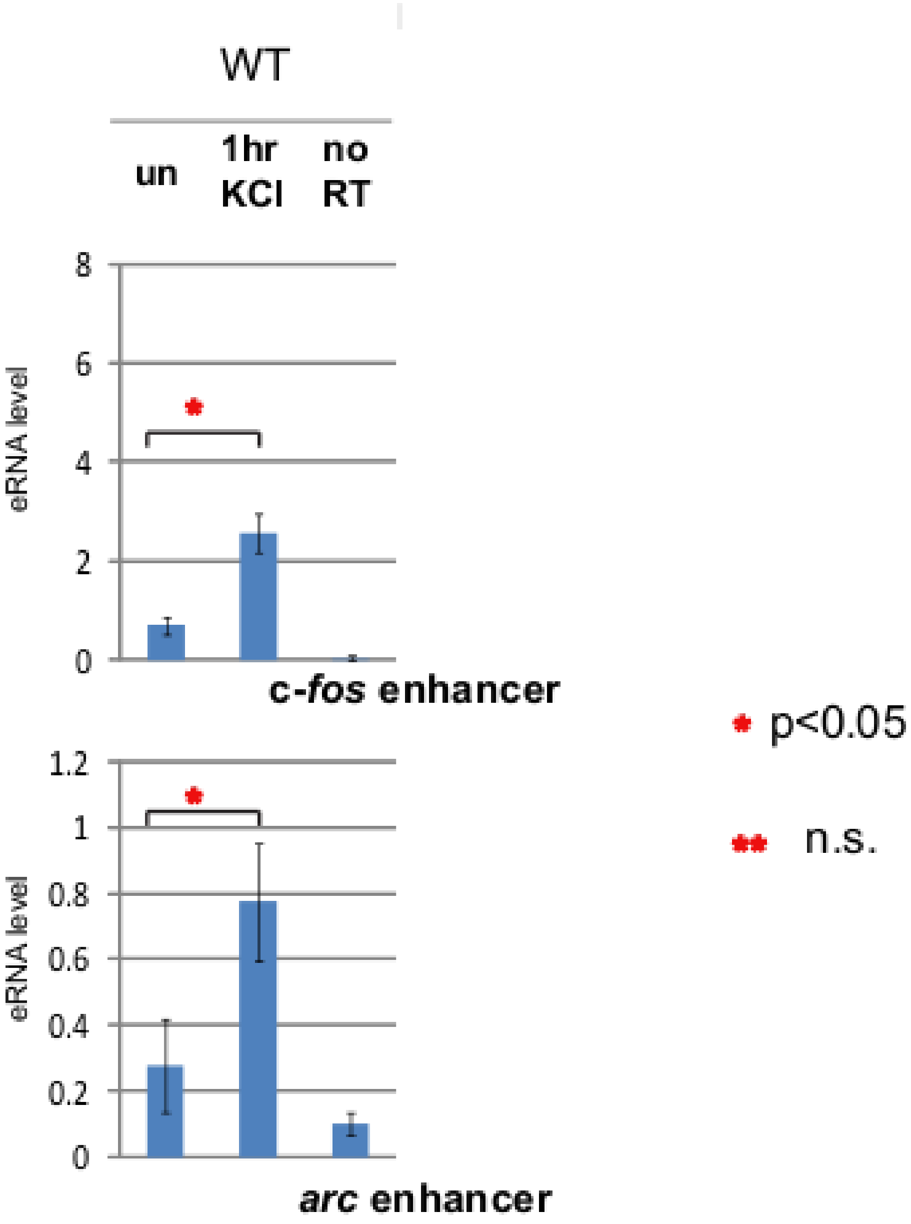
RNAPII levels are unchanged at the enhancer in the mutant before and after KCl



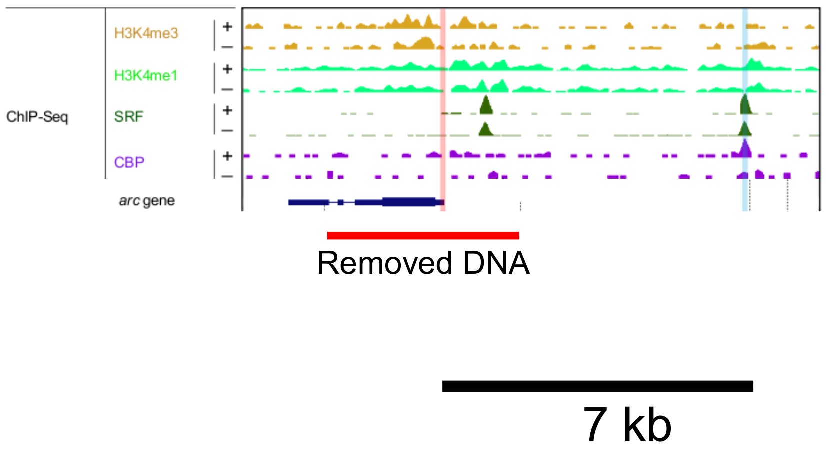


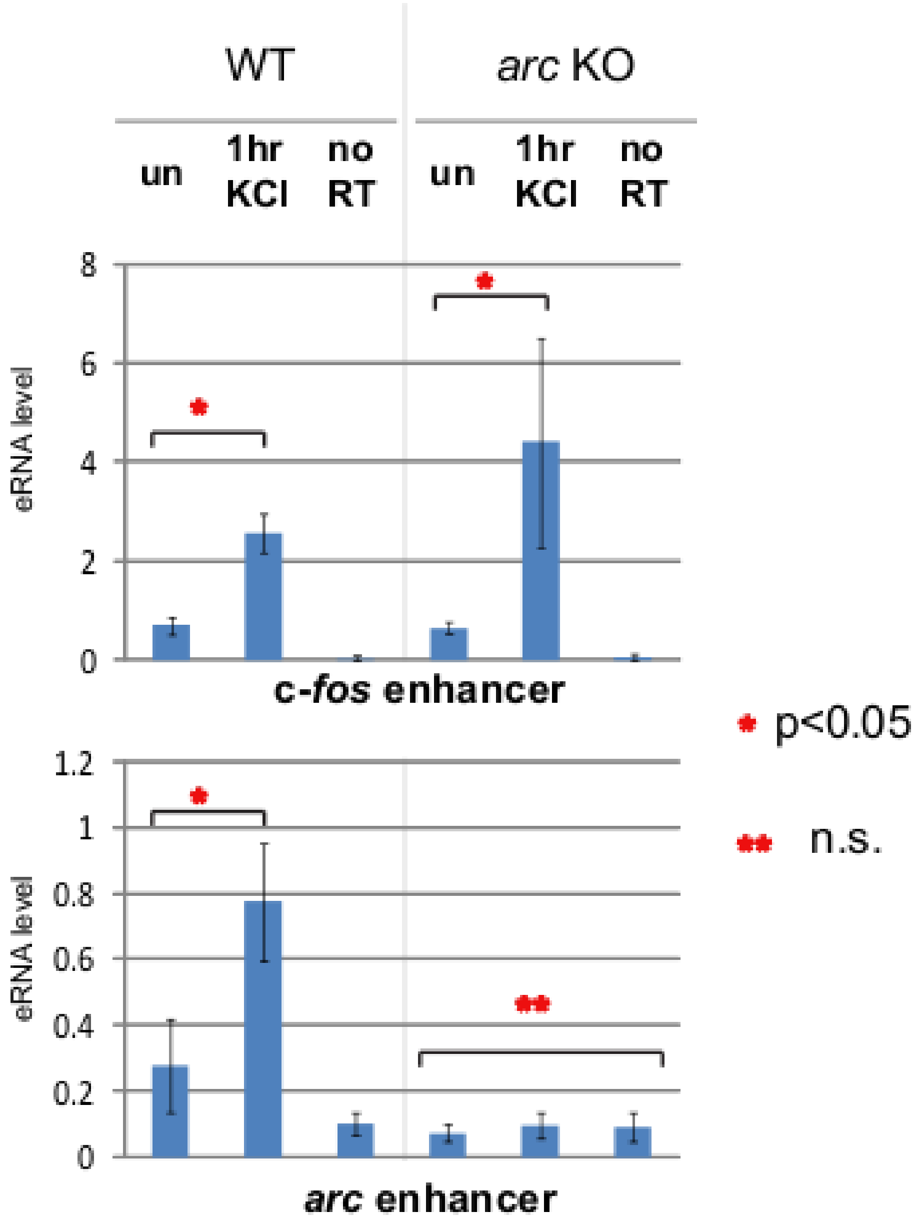
Transcription at the Fos and Arc enhancers





No transcription at Arc enhancer in mutant





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

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

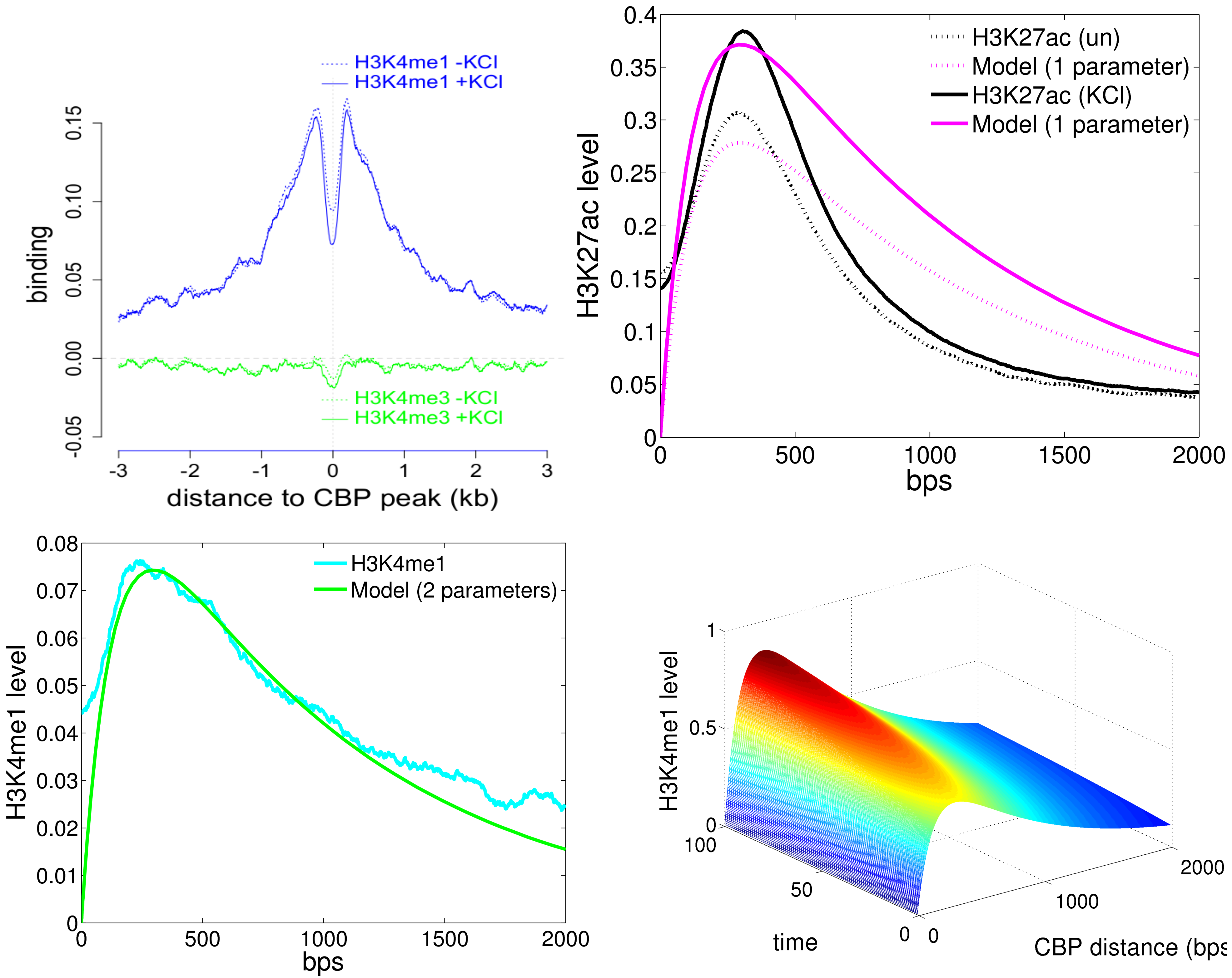
$$\tau_{decay} = \tau_{find} + \tau_{bp} L$$

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

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

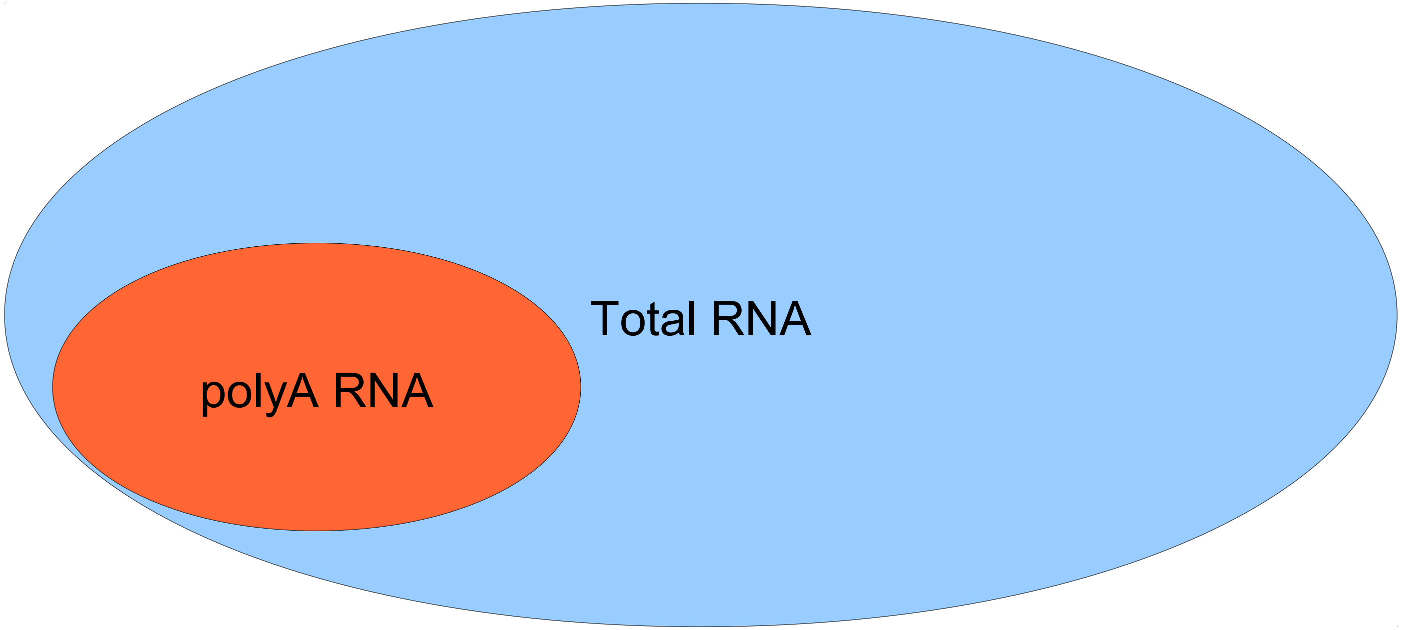
How abundant are eRNAs compared to mRNAs?

- 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


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)

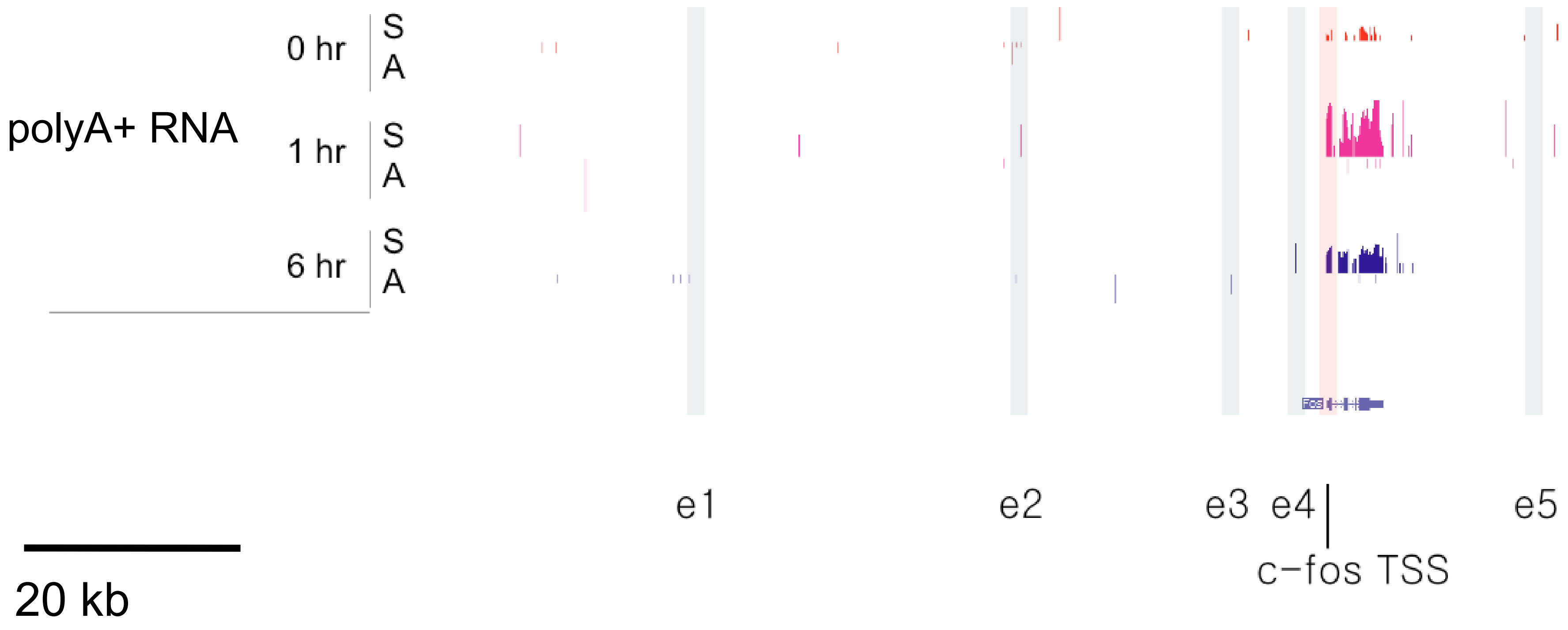

ACGUUUGUACCUAGCUAGCUUACGAG AAAAAAAAAAAAAAAAAAAA


Total RNA

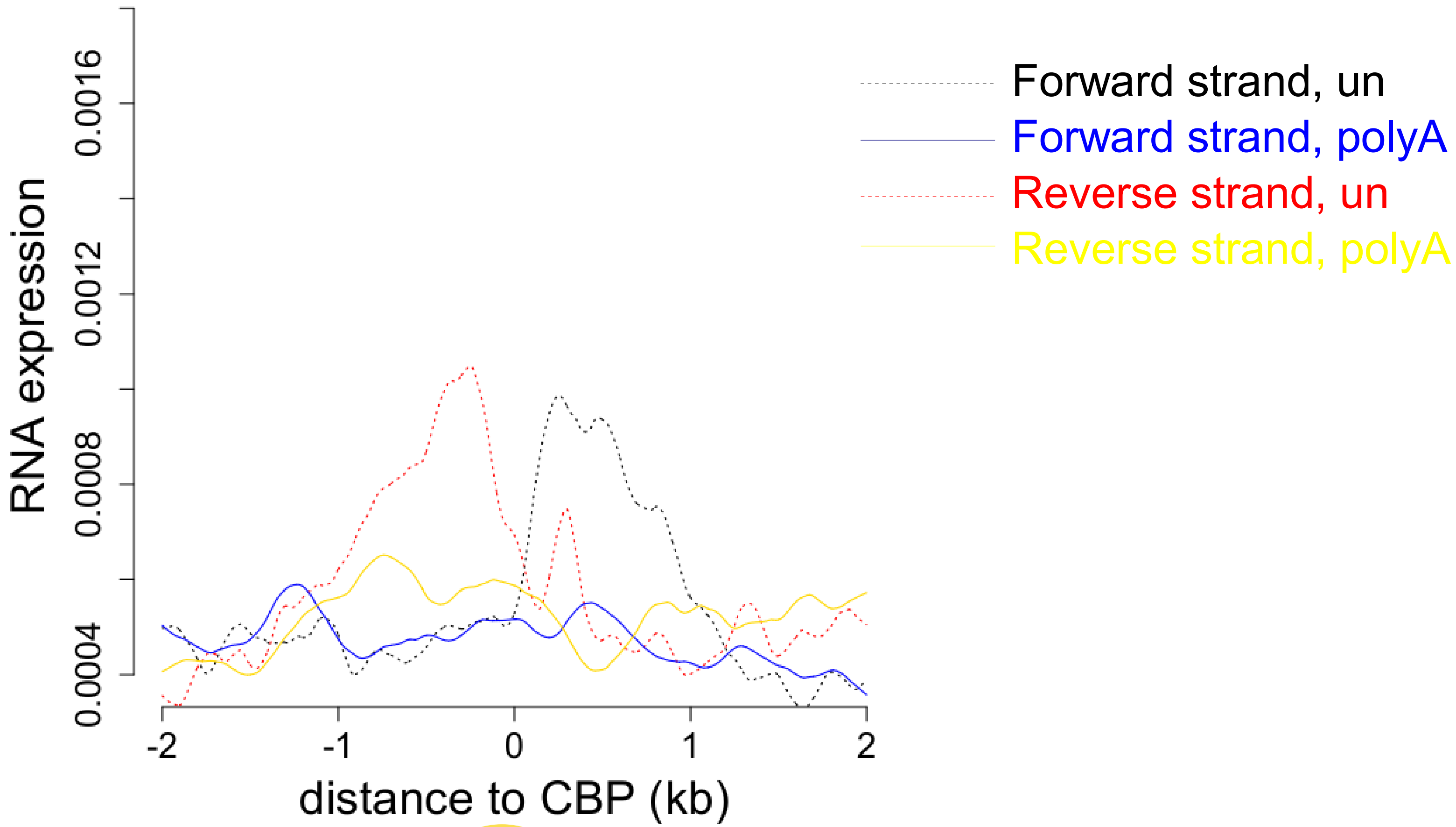
polyA RNA

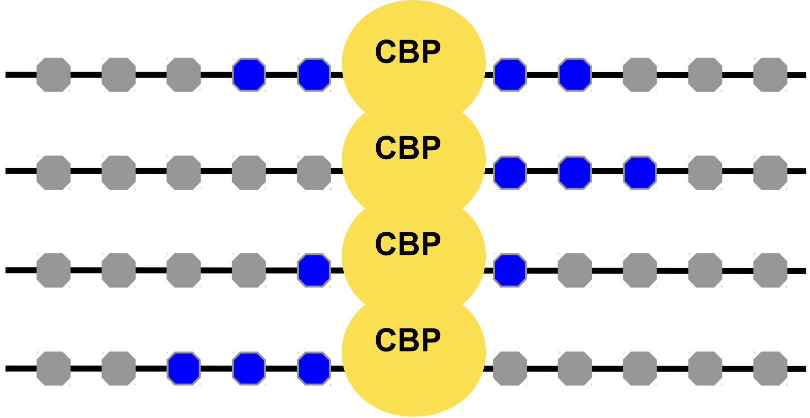
Transcription of mRNA at the *fos* locus

ACGUUUGUACCUAGCUAGCUUACGAG AAAAAAAAAAAAAAAAAAAAA



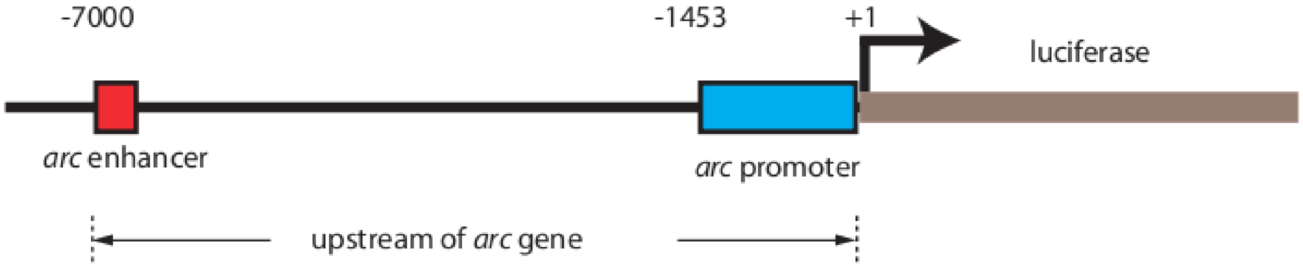
eRNAs are not polyadenylated





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

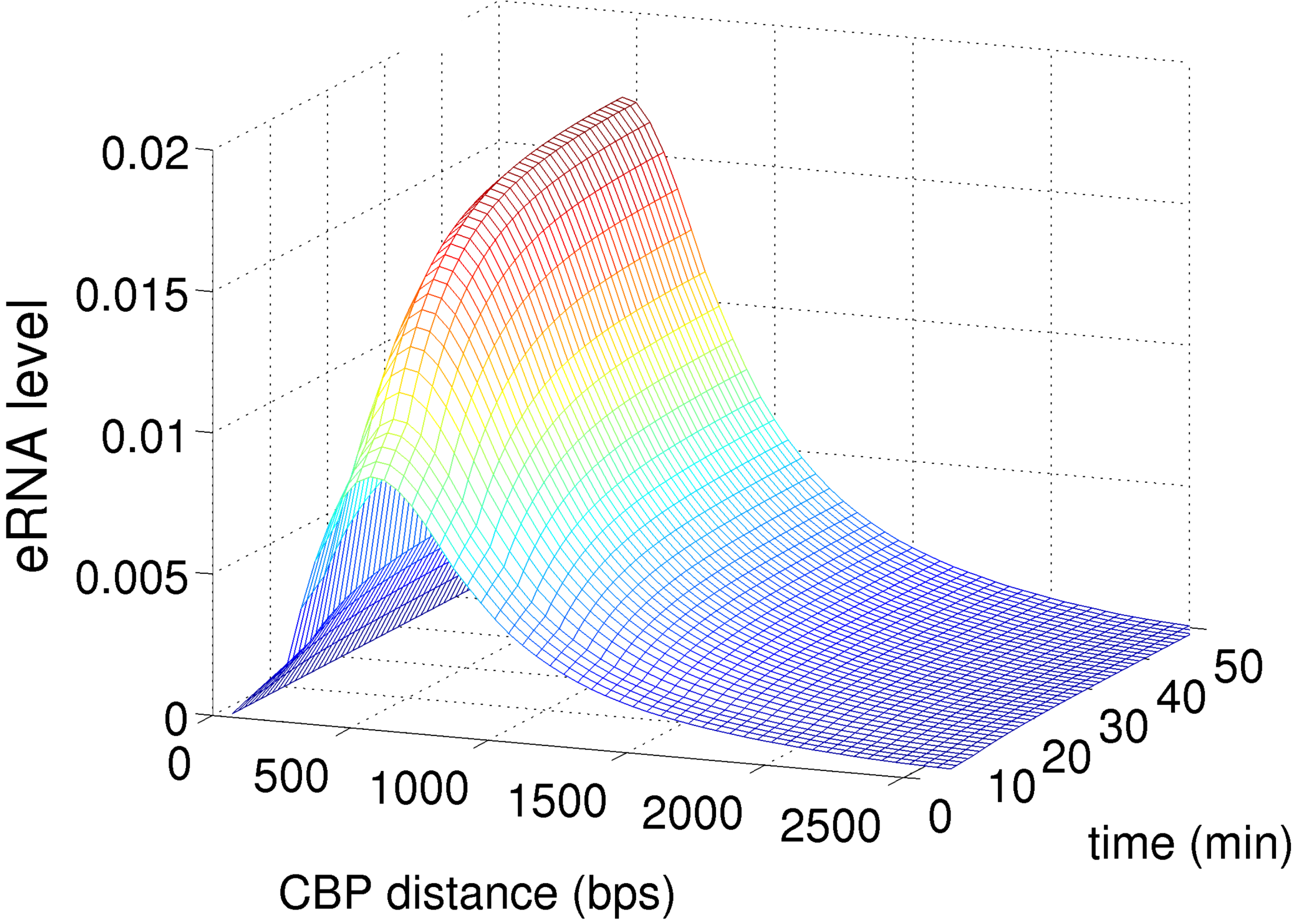
- **CBP** peak
- **High** levels of flanking **H3K4me1**
- **Low** levels of **H3K4me3**
    - 8/8 tested activity-dependent enhancers were validated using a luciferase assay



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