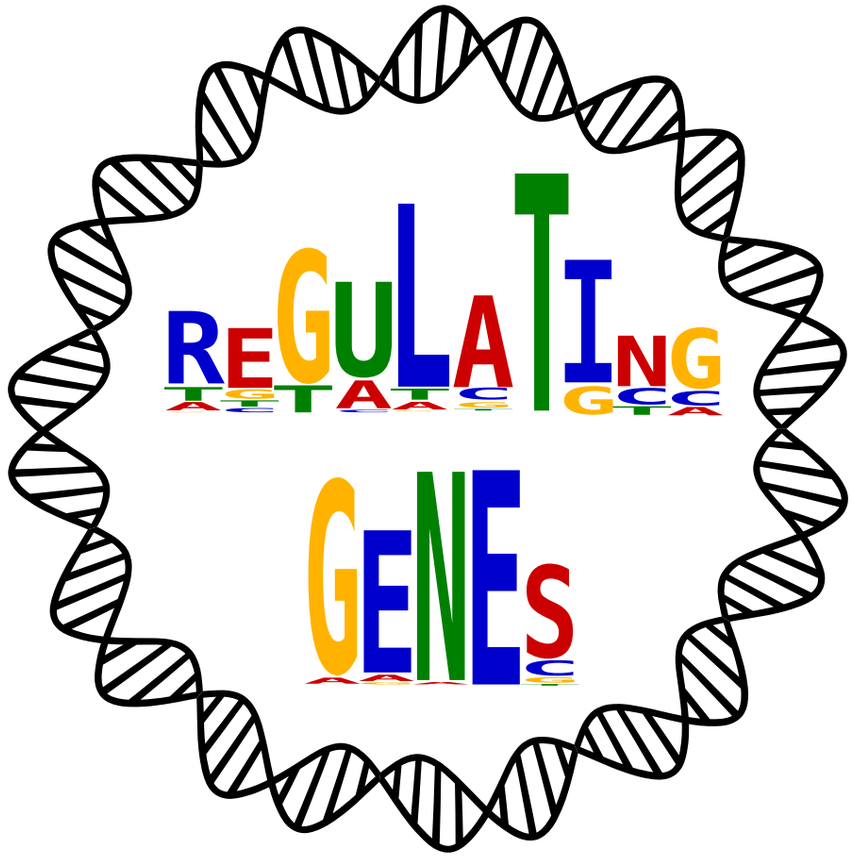
1) Molecular tracking of differentiation cascades;
2) Mechanisms of direct repression by a transcription activator




UCONN
HEALTH

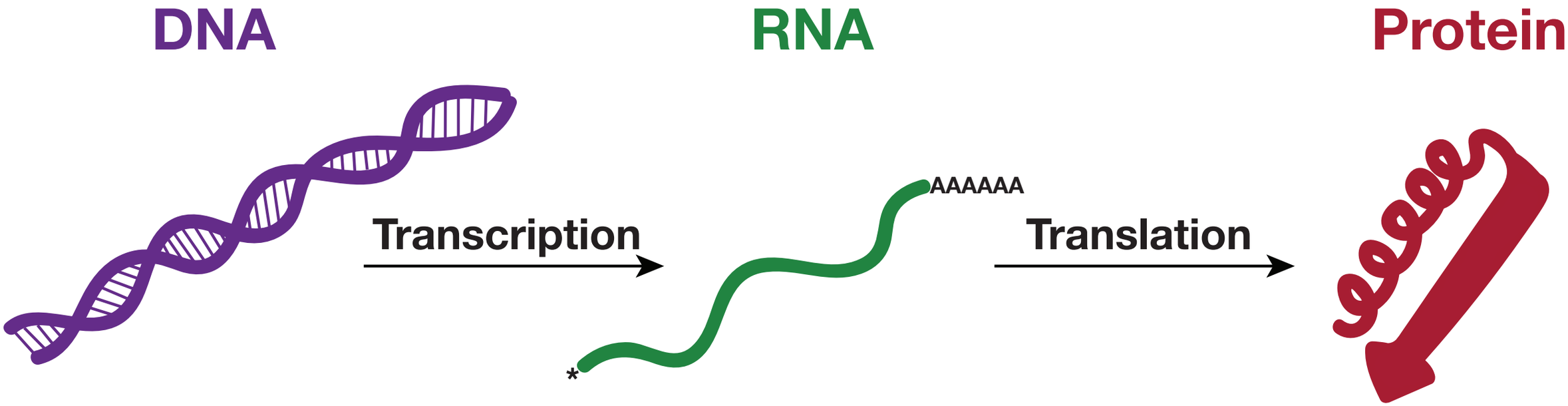
**Genetics and Genome Sciences Department**

Center for Cell Analysis & Modeling

GUERTIN LAB

Michael Guertin

# Gene Regulation

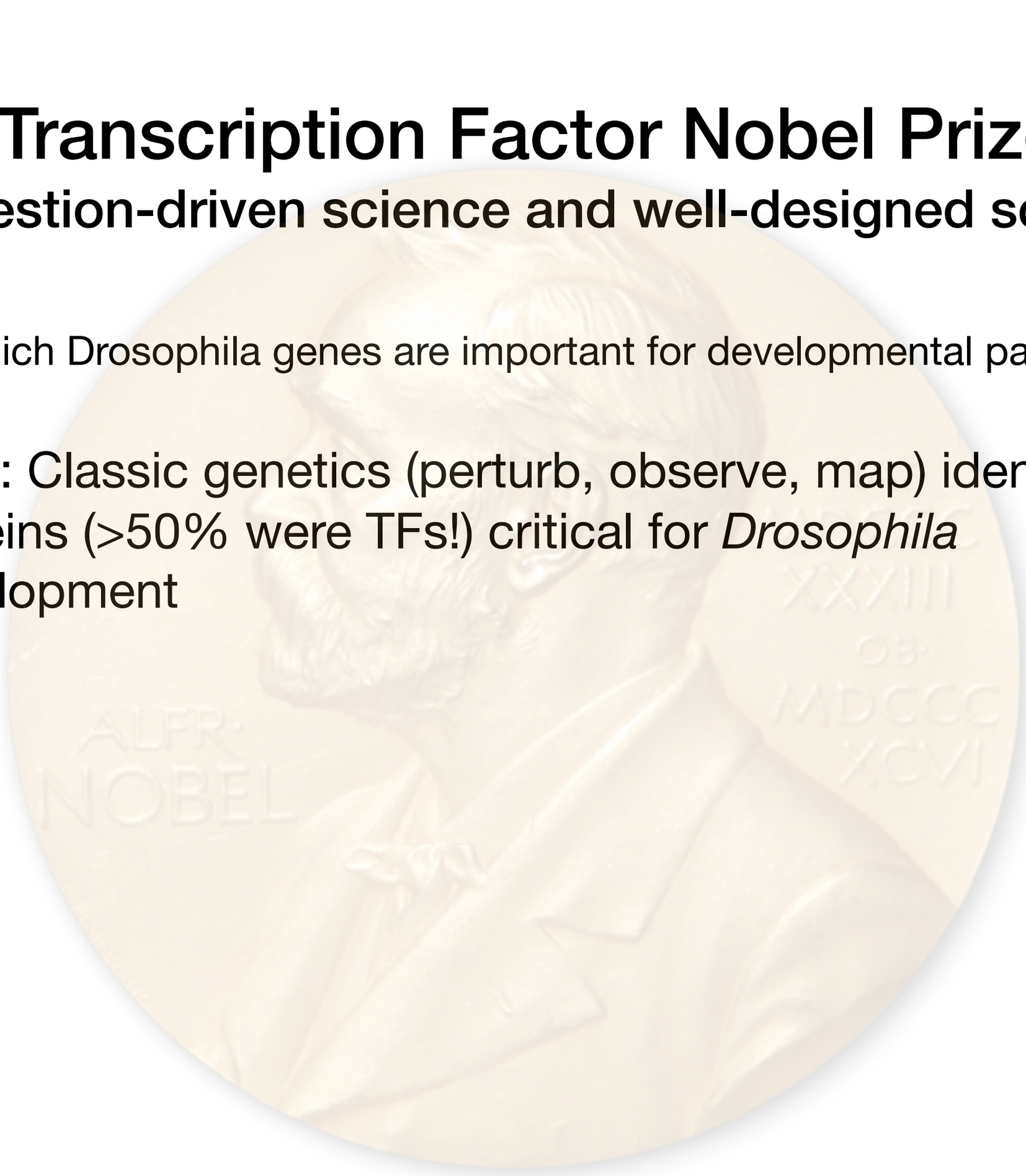


# Transcription Factor Nobel Prizes

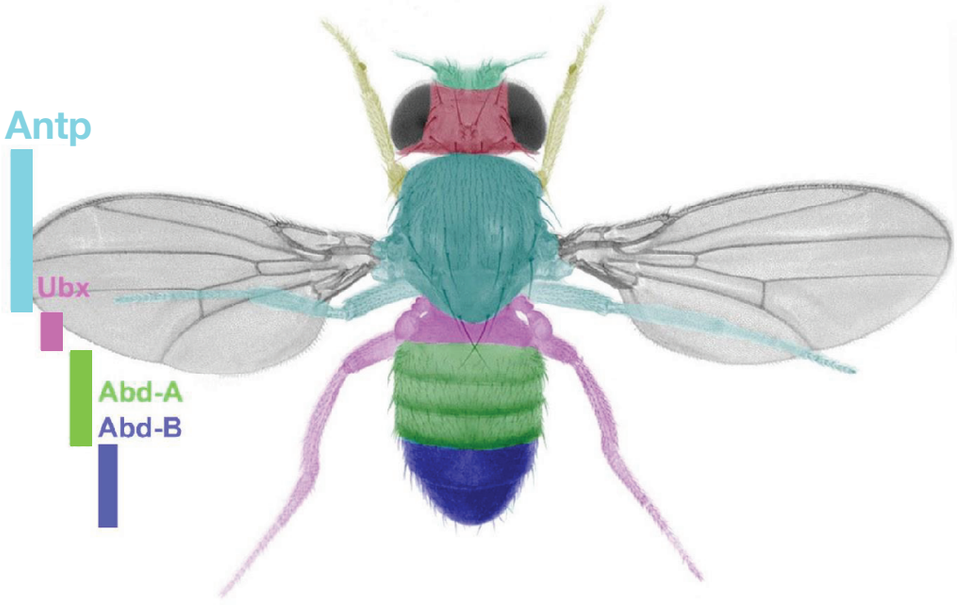
## (Question-driven science and well-designed screens)

Q: Which Drosophila genes are important for developmental patterning?

1995: Classic genetics (perturb, observe, map) identifies proteins (>50% were TFs!) critical for *Drosophila* development



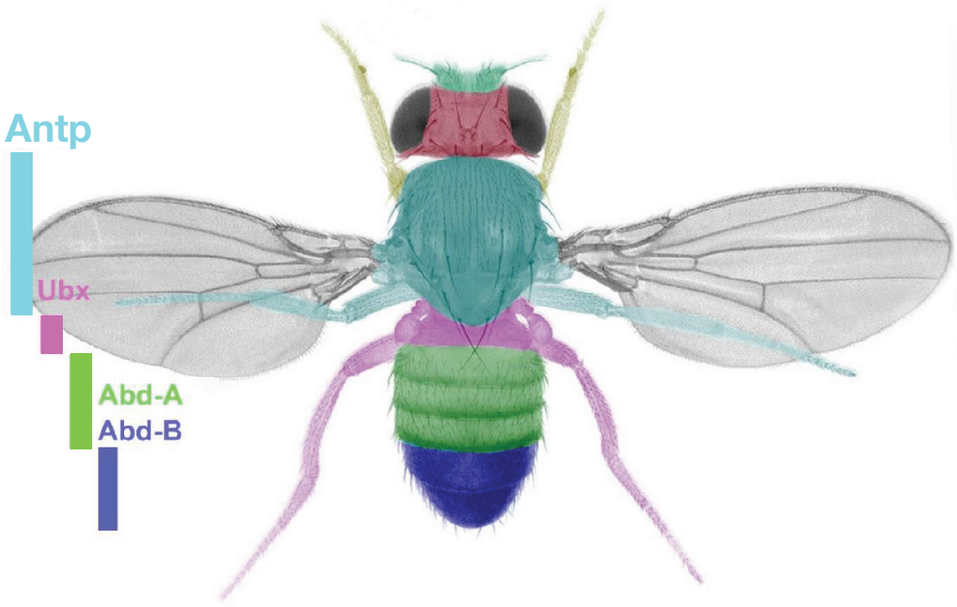
# Transcription dysregulation alters developmental patterning





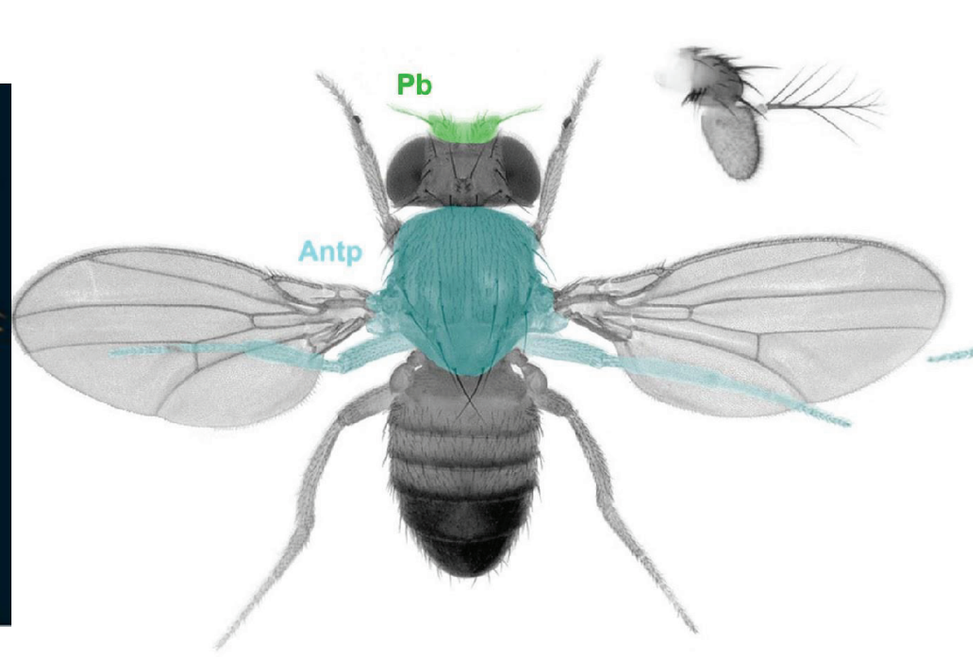

pseudocolored flies: Justin Crocker, Ed Lewis, Nicolas Gompel, and Welcome Bender

# Classic Genetics: Perturb and Map












pseudocolored flies: Justin Crocker, Ed Lewis, Nicolas Gompel, and Welcome Bender

pseudocolored SEM heads: Jürgen Berger

# Classic genetics (**perturb, observe, map**) found that Transcription Factors control developmental patterning

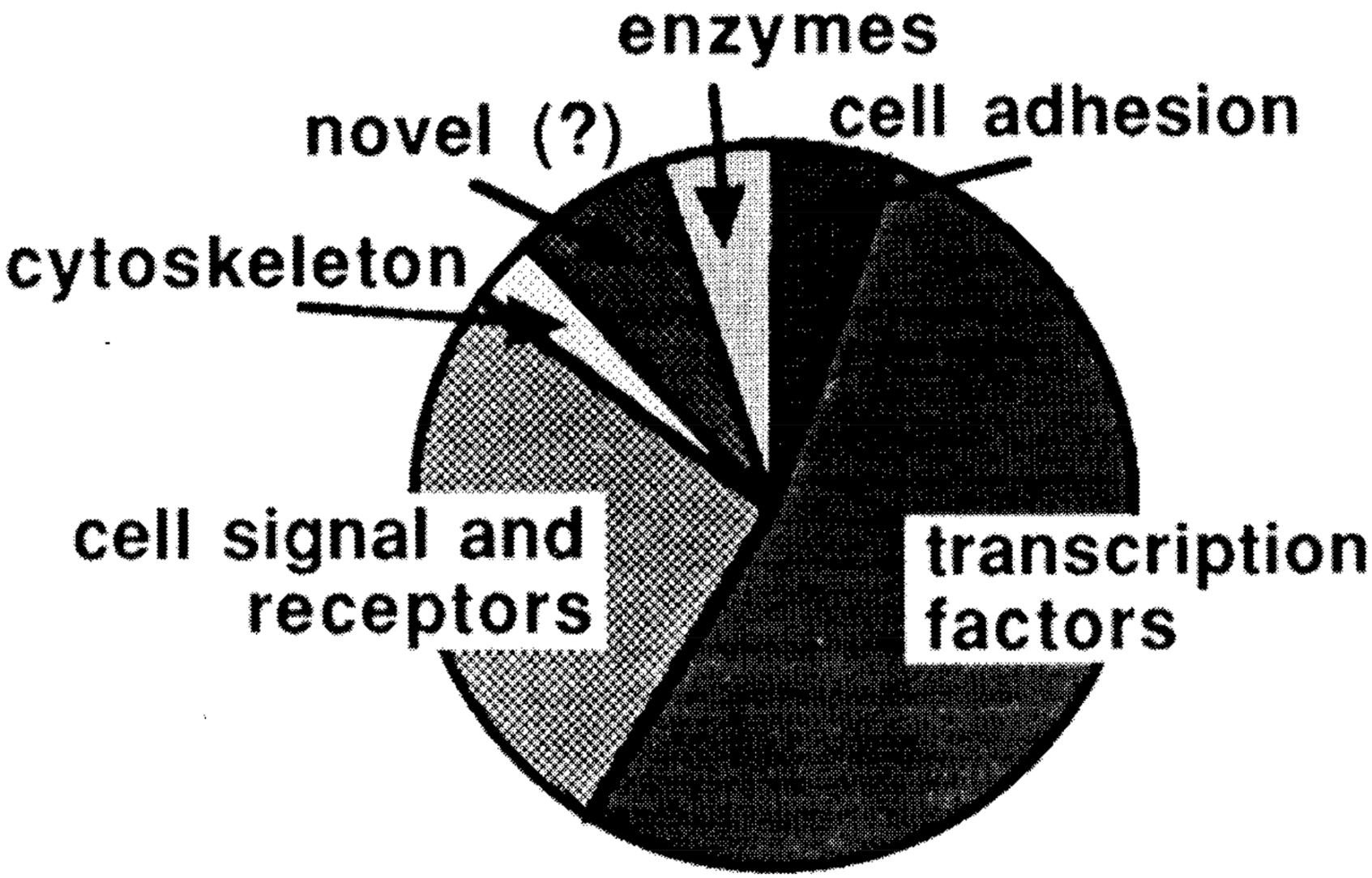



Figure 3. *Cellular Function of Heidelberg Mutations.* Based on the sequence of 75 cloned genes, most of the loci identified in Heidelberg encode transcription factors, or cell signals and receptors.

Eric Weischaus, Nobel Lecture 1995
Prize shared with Christiane Nüsslein-Volhard & Ed Lewis

# Transcription Factor Nobel Prizes
## (Question-driven science and well-designed screens)

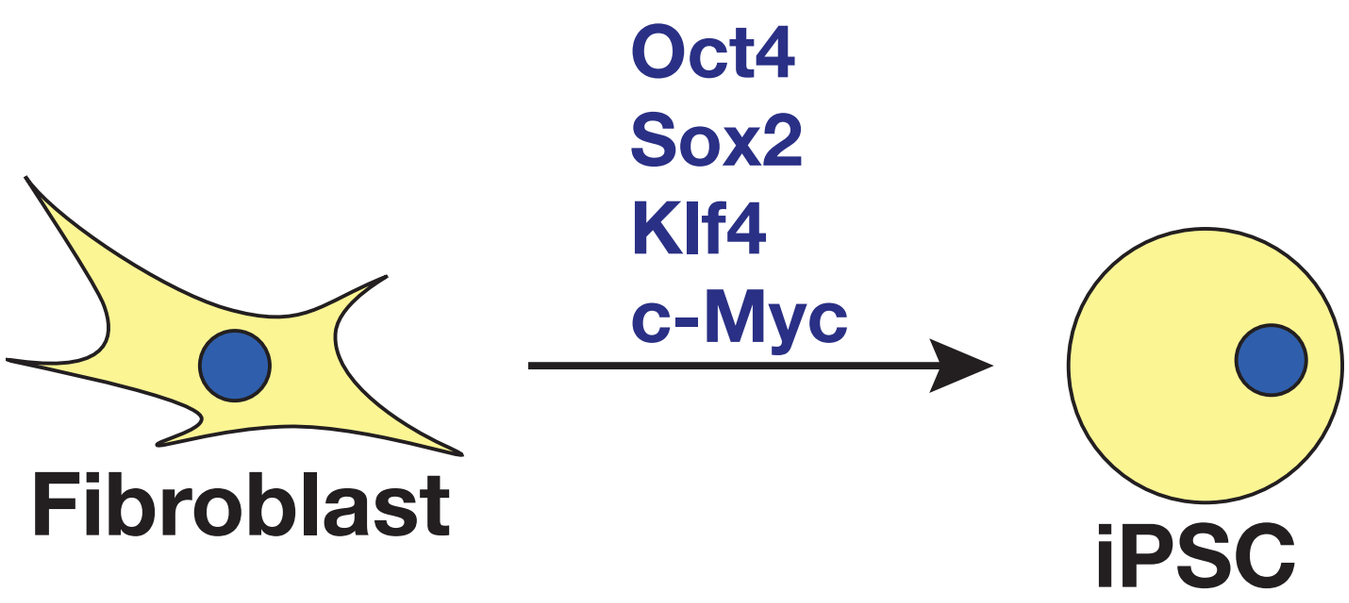
Q: Which Drosophila genes are important for developmental patterning?

1995: Classic genetics (perturb, observe, map) identifies proteins (>50% were TFs!) critical for *Drosophila* development

Q: Can we reprogram differentiated cells to a pluripotent state?

2012: Brute force screen with candidate genes identified from expression microarrays. Transduce 24 genes that are specifically expressed in embryonic stem cells into fibroblasts. Systematically narrow down the list: Oct3/4, Klf4, Sox2, and c-Myc (all TFs!)

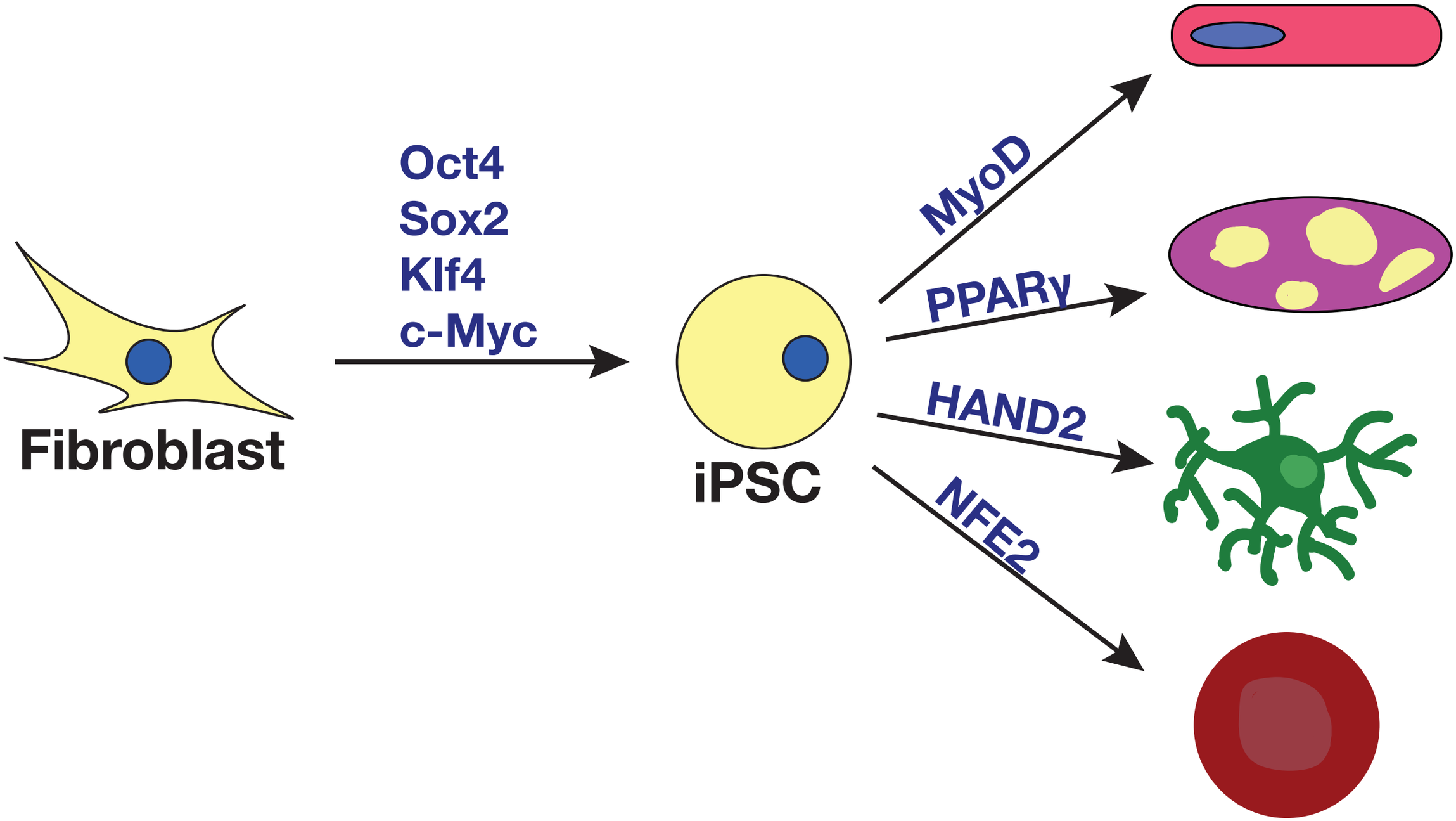
# Transcription factors drive changes in cell identity




Brute force screening approach:
express many genes in combination until we get iPSCs

Takahashi & Yamanaka, Cell 2006
2012 Nobel Prize shared with John Gurdon

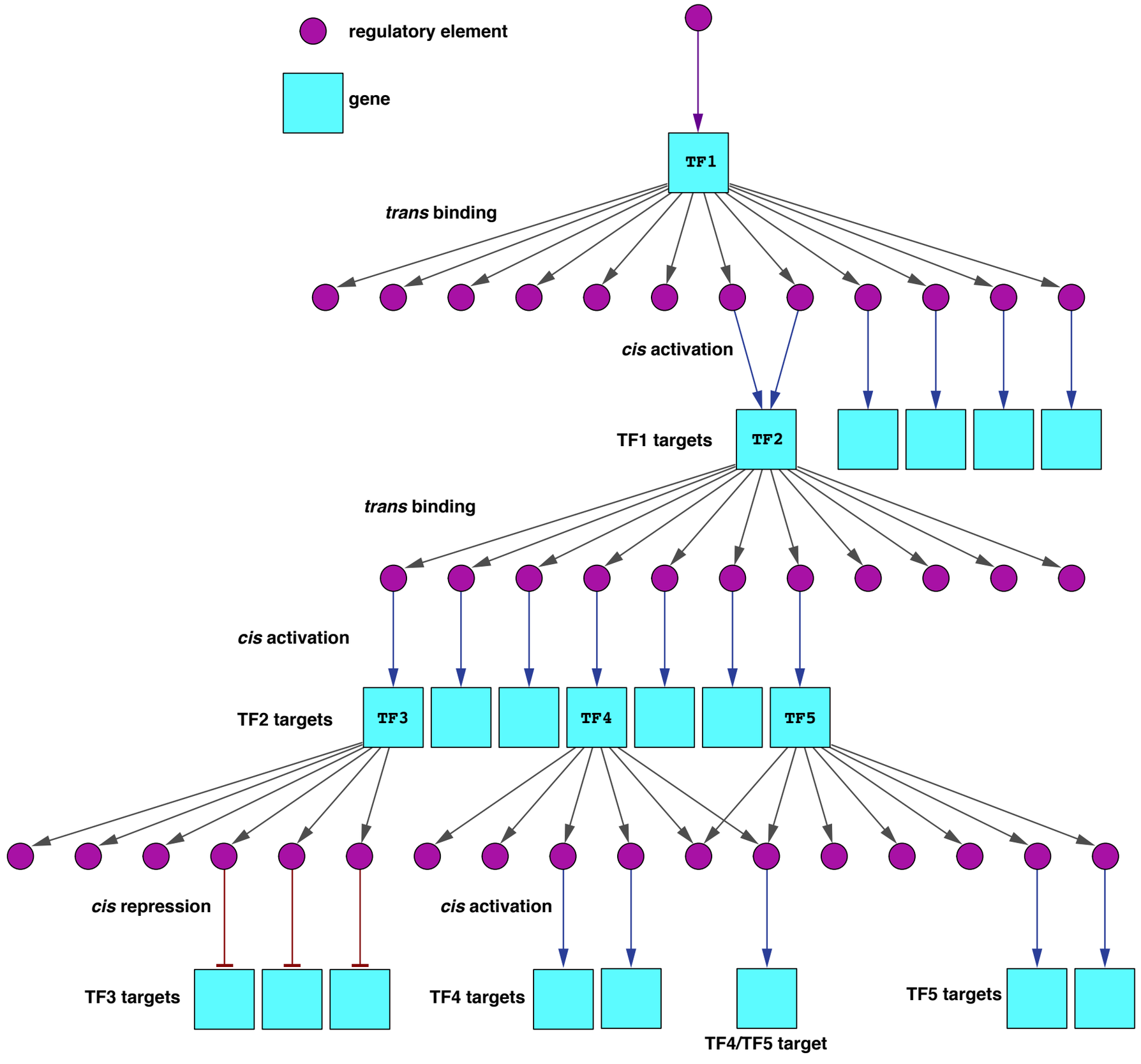
# Transcription factors drive changes in cell identity




Takahashi & Yamanaka, Cell 2006
2012 Nobel Prize shared with John Gurdon

Each research project in the lab starts out with a question—hypotheses come later.

# Q1: Which TFs and regulatory elements are important in regulatory cascades?



# Possible approaches and considerations

- ChIP-seq all the TFs throughout the cascade, but which TFs?

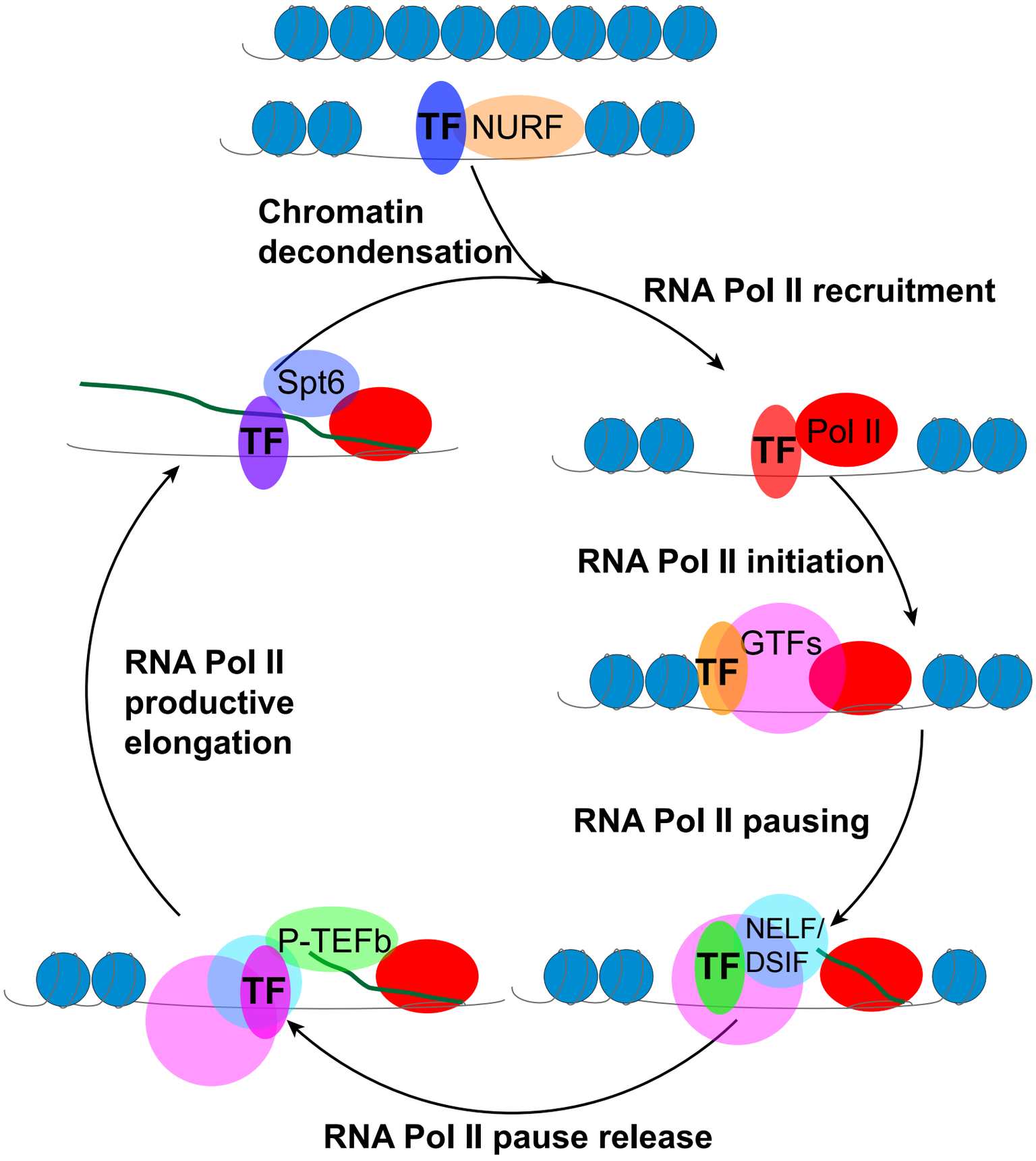
- ATAC-seq gives us putative regulatory elements, but which TFs are binding/dissociating during the cascade?

- RNA-seq quantifies changes in gene expression, but detection is delayed depending on RNA turnover rate and accumulation rate over basal RNA levels.

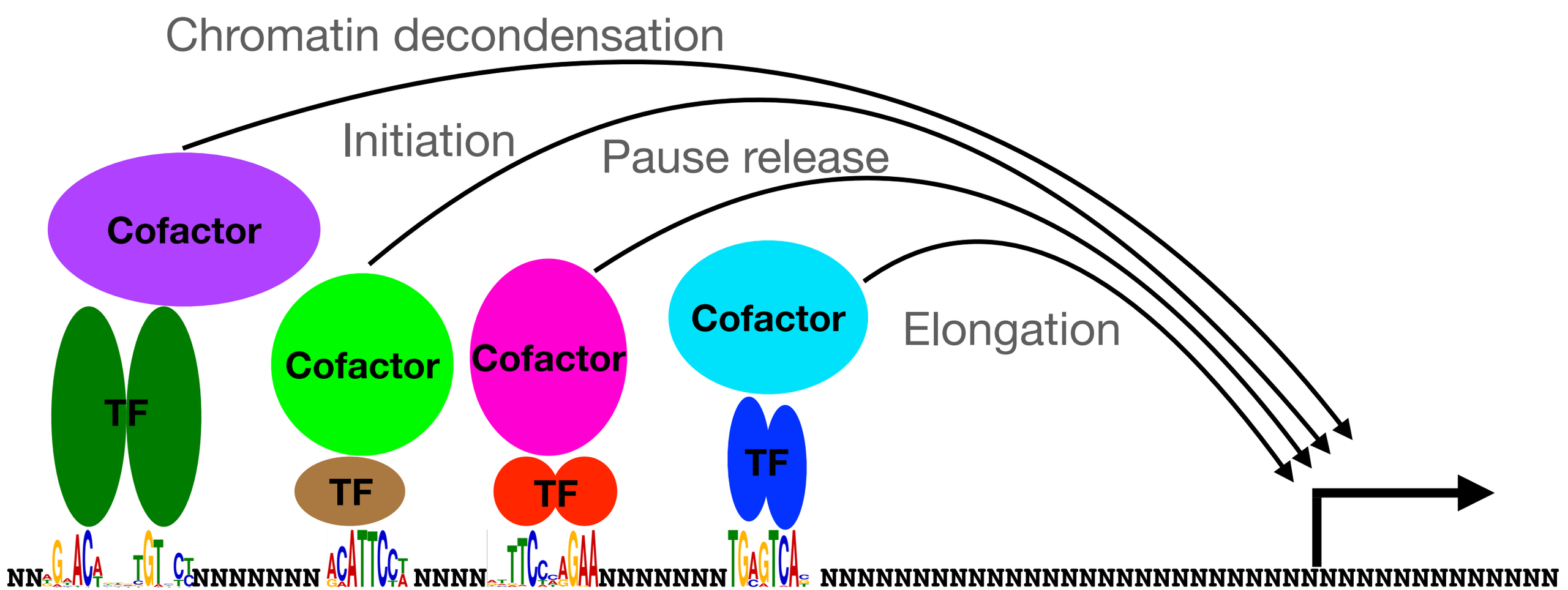
- Which time points do we choose?

- Which models can we genetically manipulate to validate candidate TFs and elements?

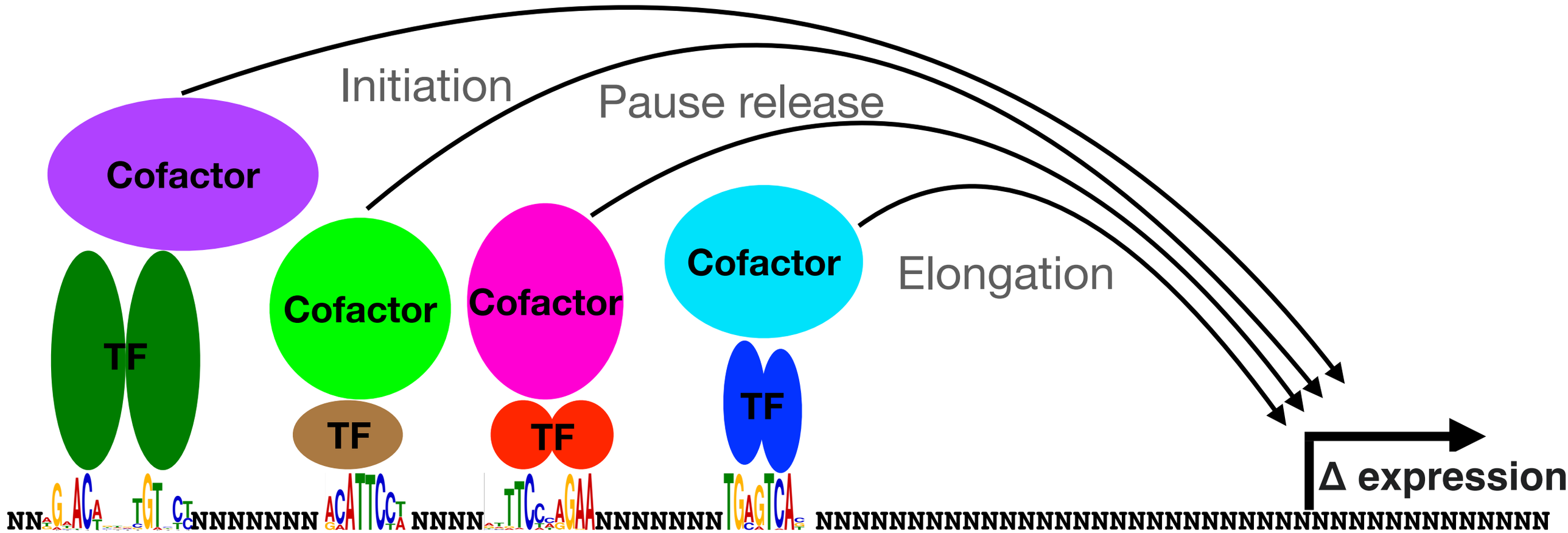
# Q2: What are the step(s) in transcription that each key TF regulates?



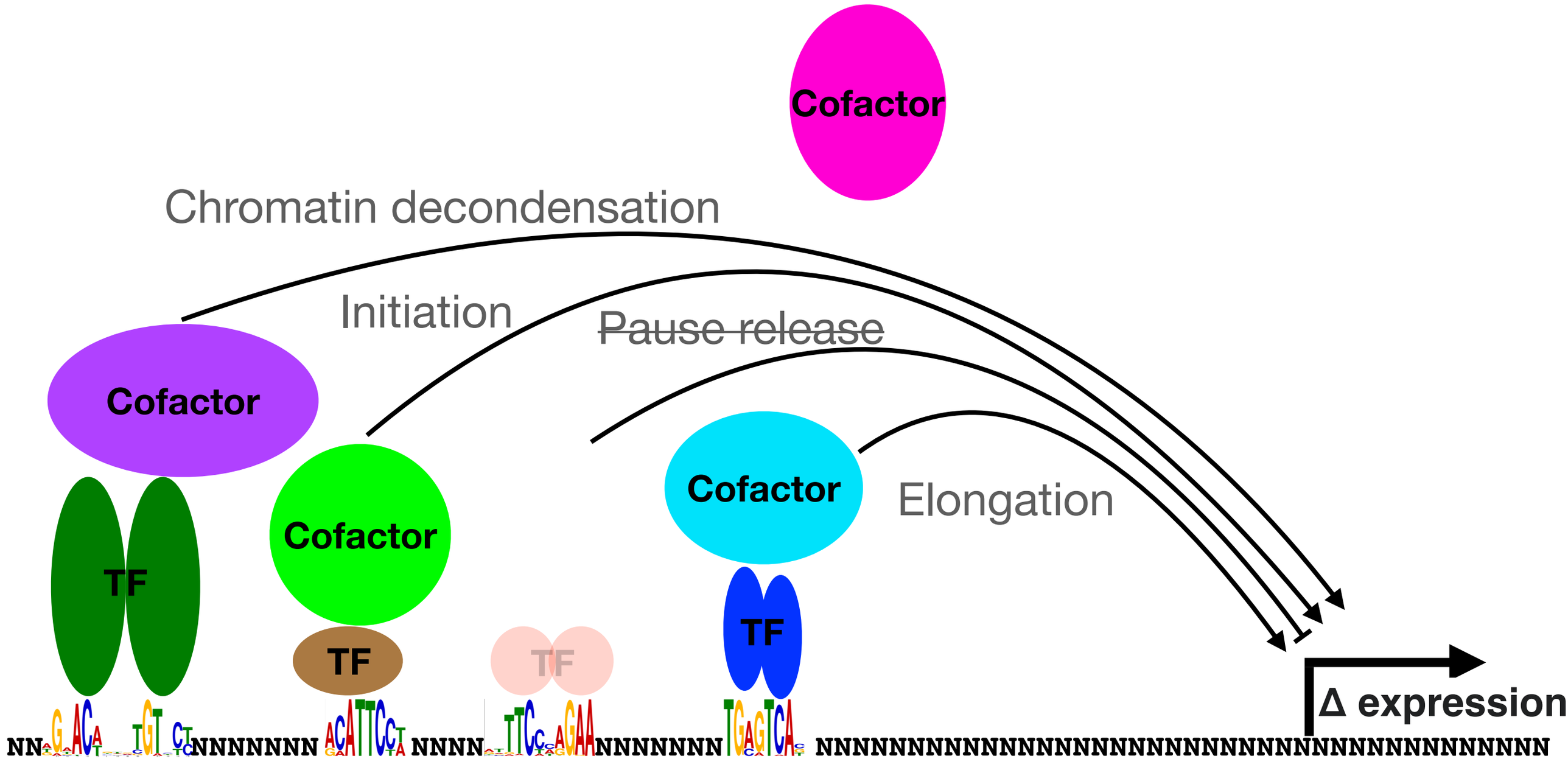
# Sequence-specific TFs specialize and work together


Chromatin decondensation
Initiation
Pause release
Elongation
Cofactor
Cofactor
Cofactor
Cofactor
TF
TF
TF
TF
NNNN

# Sequence-specific TFs specialize and work together


Chromatin decondensation
Initiation
Pause release
Elongation
Cofactor
Cofactor
Cofactor
Cofactor
TF
TF
TF
TF
Δ expression
NNNN

# Approach: rapidly degrade red TF and quantify changes in each step


Cofactor
Chromatin decondensation
Initiation
Pause release
Cofactor
Cofactor
Cofactor
Elongation
TF
TF
TF
TF
Δ expression
NNNN

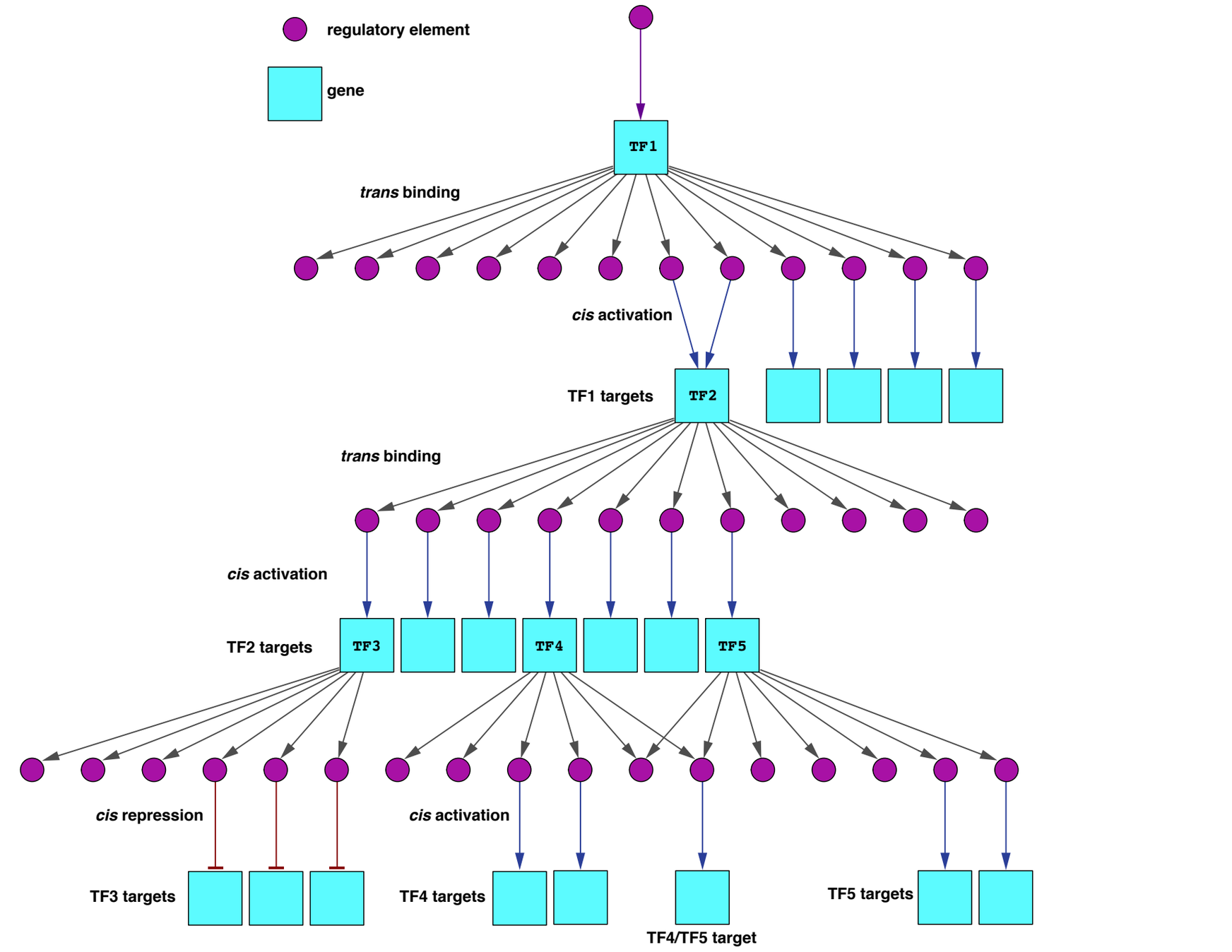
# Outline

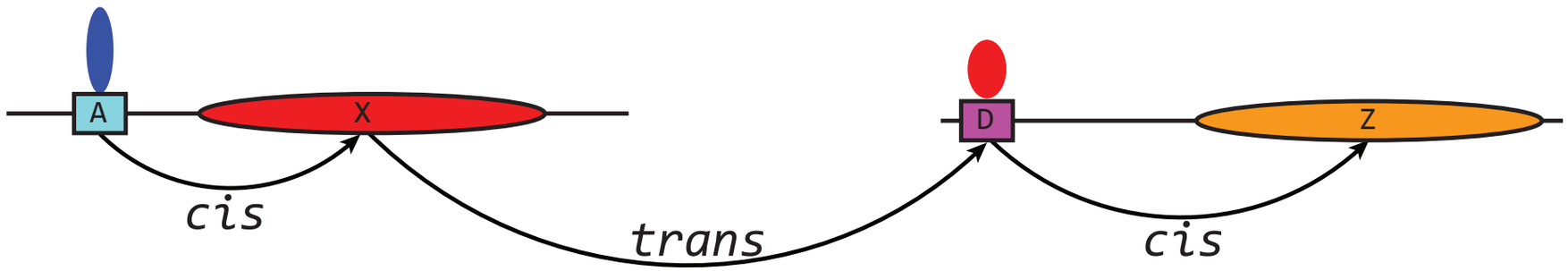
- Question-driven science and well-designed studies

- Mechanistic gene regulatory networks

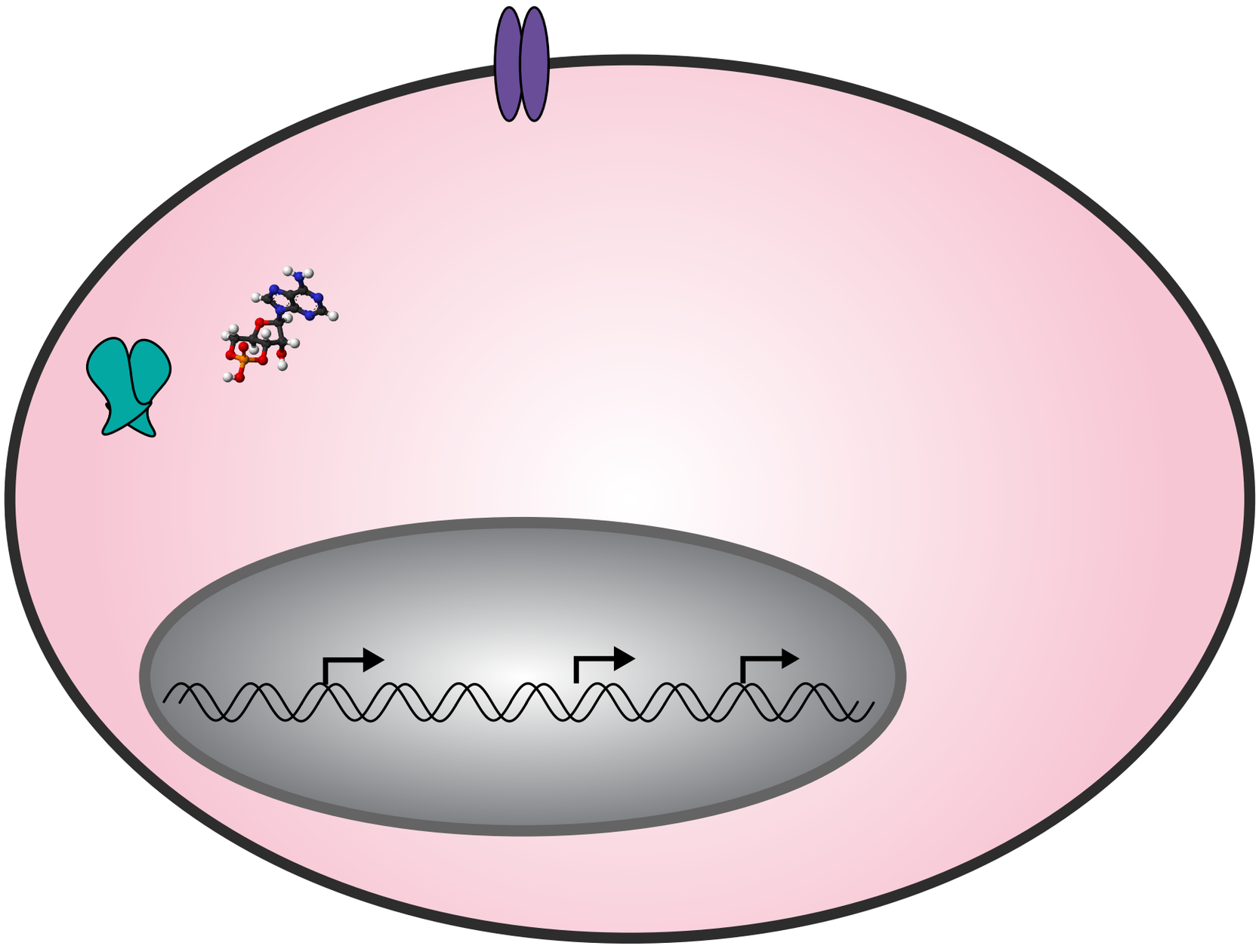
- Rapid depletion of the activator ZNF143 identifies unexpected direct repression of ZNF143 target genes

# Molecular tracking of differentiation cascades

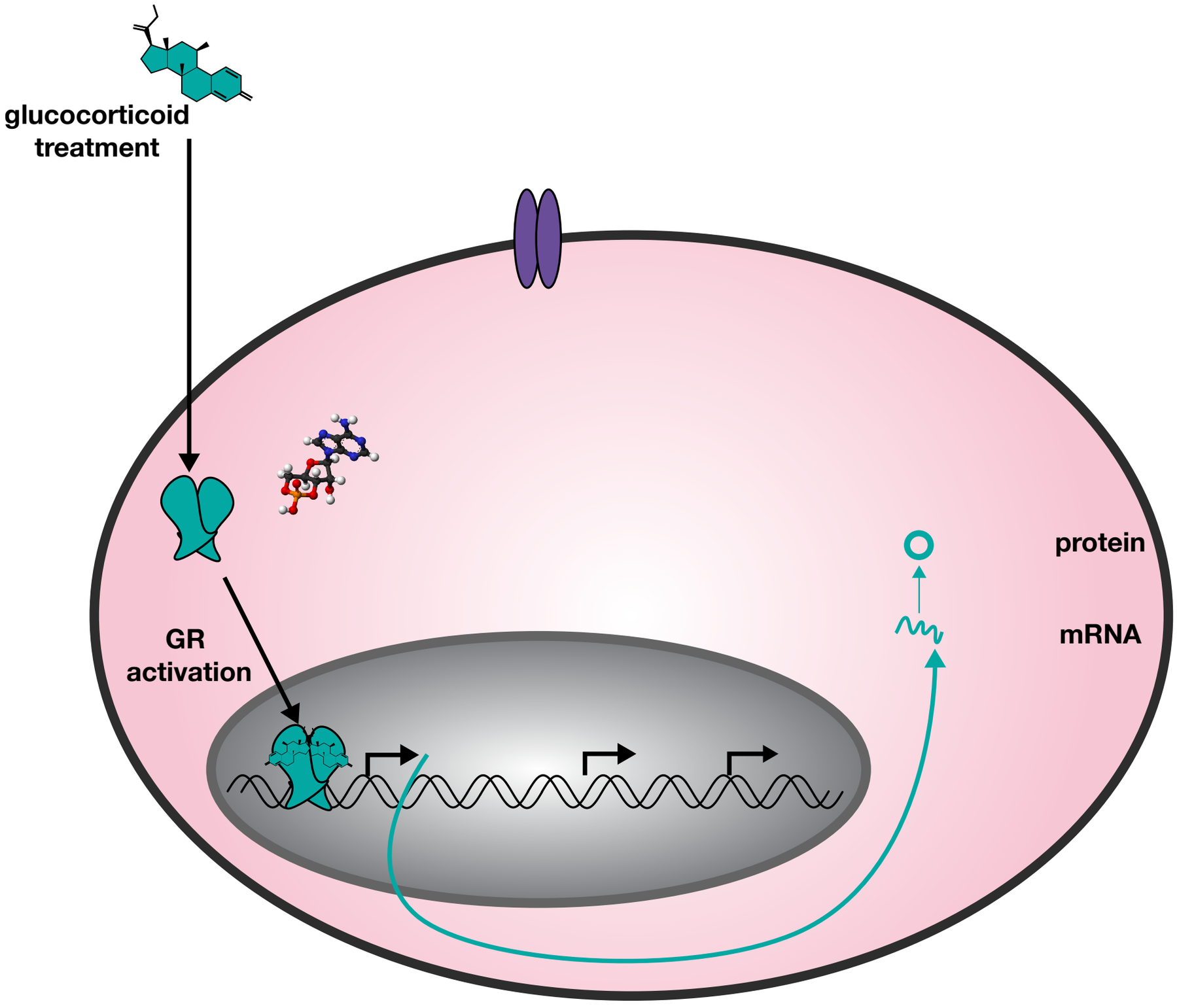


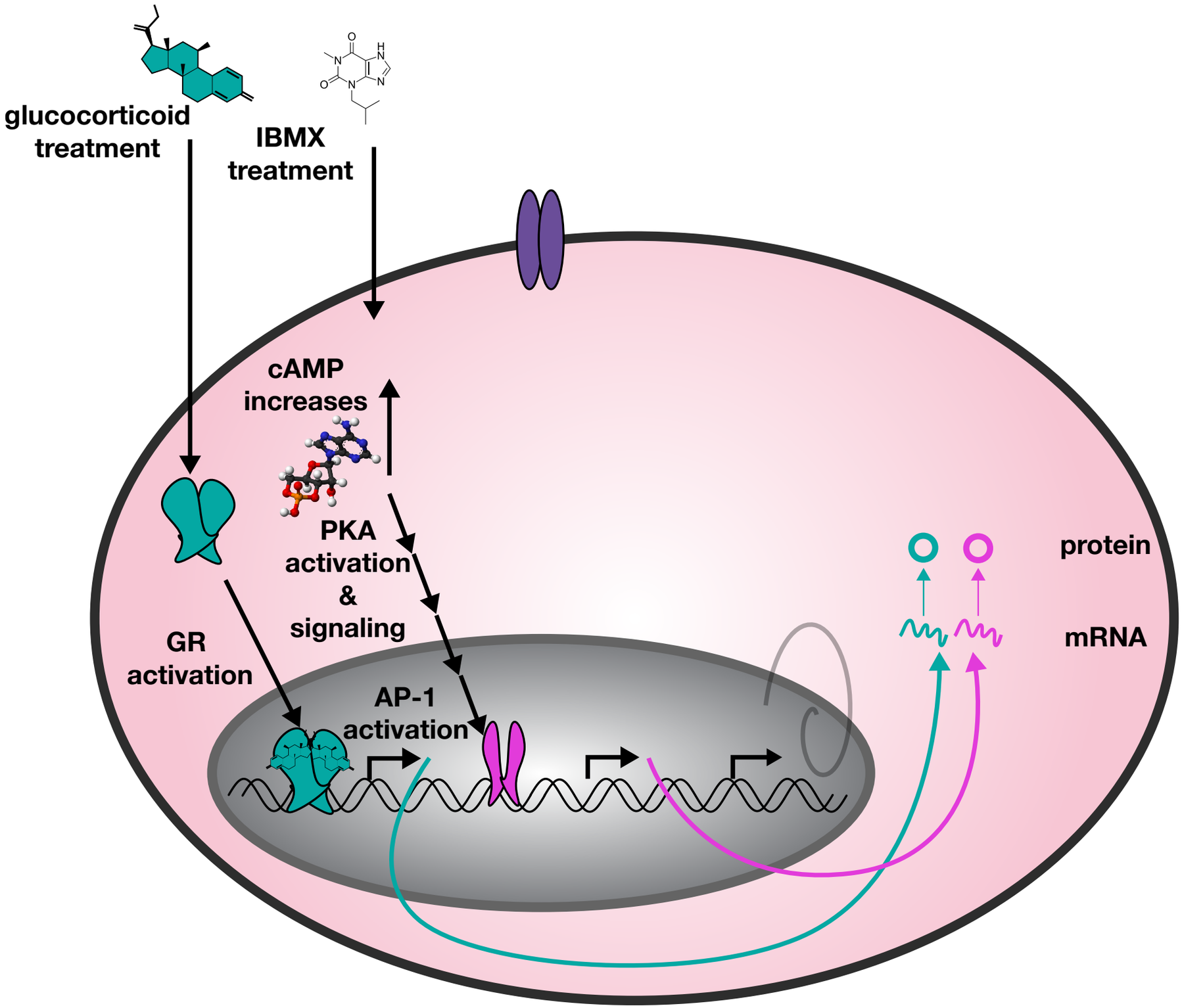
# Adipogenesis of 3T3-L1 cells



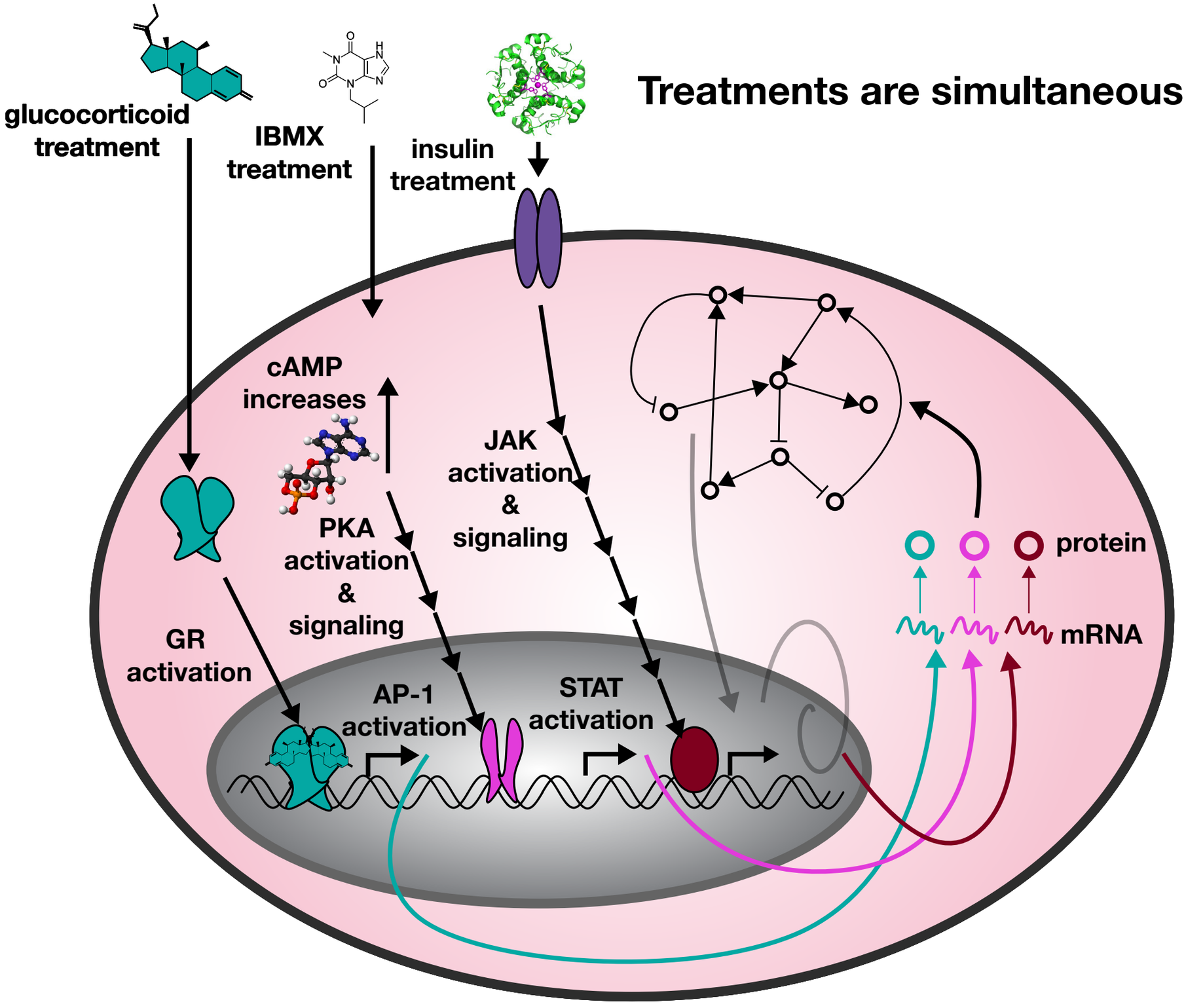
# Adipogenesis of 3T3-L1 cells



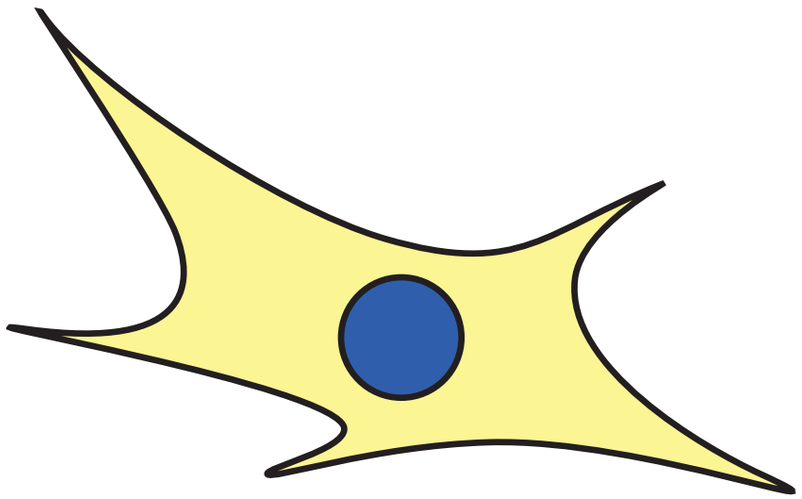
# Adipogenesis of 3T3-L1 cells



# Adipogenesis of 3T3-L1 cells

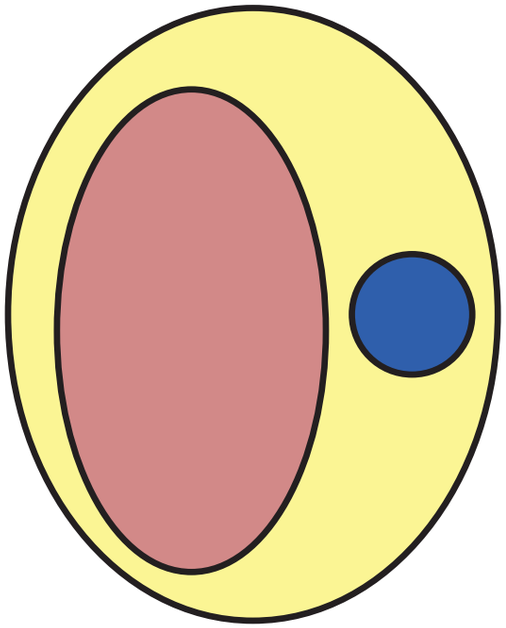


# Experimental Design



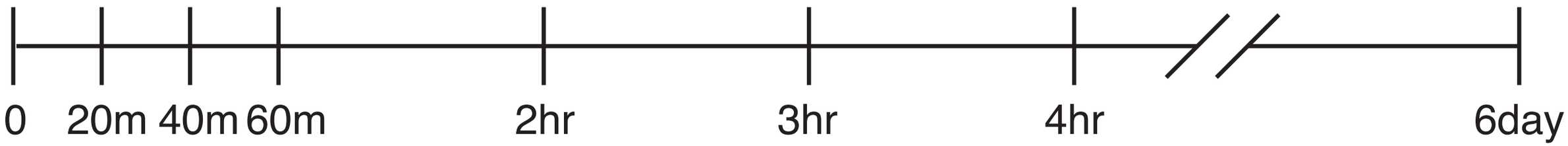
Glucocorticoid receptor agonist
cAMP agonist
Insulin receptor agonist

6 days



3T3-L1 preadipocyte fibroblast

Mature adipocycte




Time points for ATAC-seq and PRO-seq

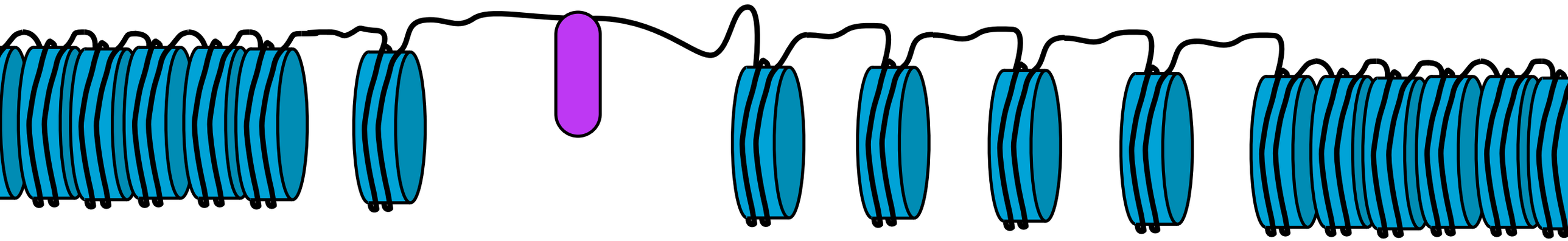
# Transposase Hypersensitivity (ATAC-seq)

- A general measure of chromatin structure.

# Transposase Hypersensitivity (ATAC-seq)

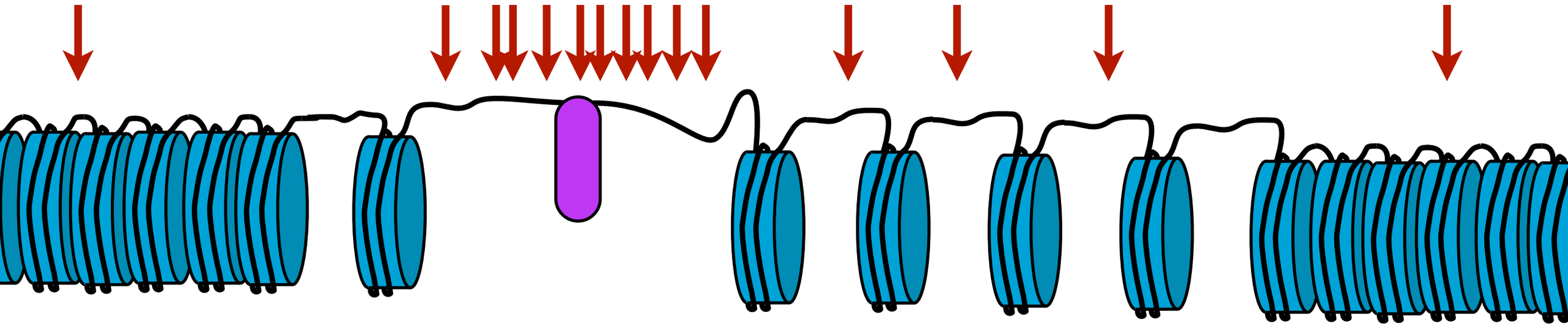
- A general measure of chromatin structure.
  - ATAC peaks and DNA motifs can be used to infer TF binding (i.e. if chromatin is accessible and contains a binding sequence, then a TF may be bound)

# Transposase Hypersensitivity (ATAC-seq)



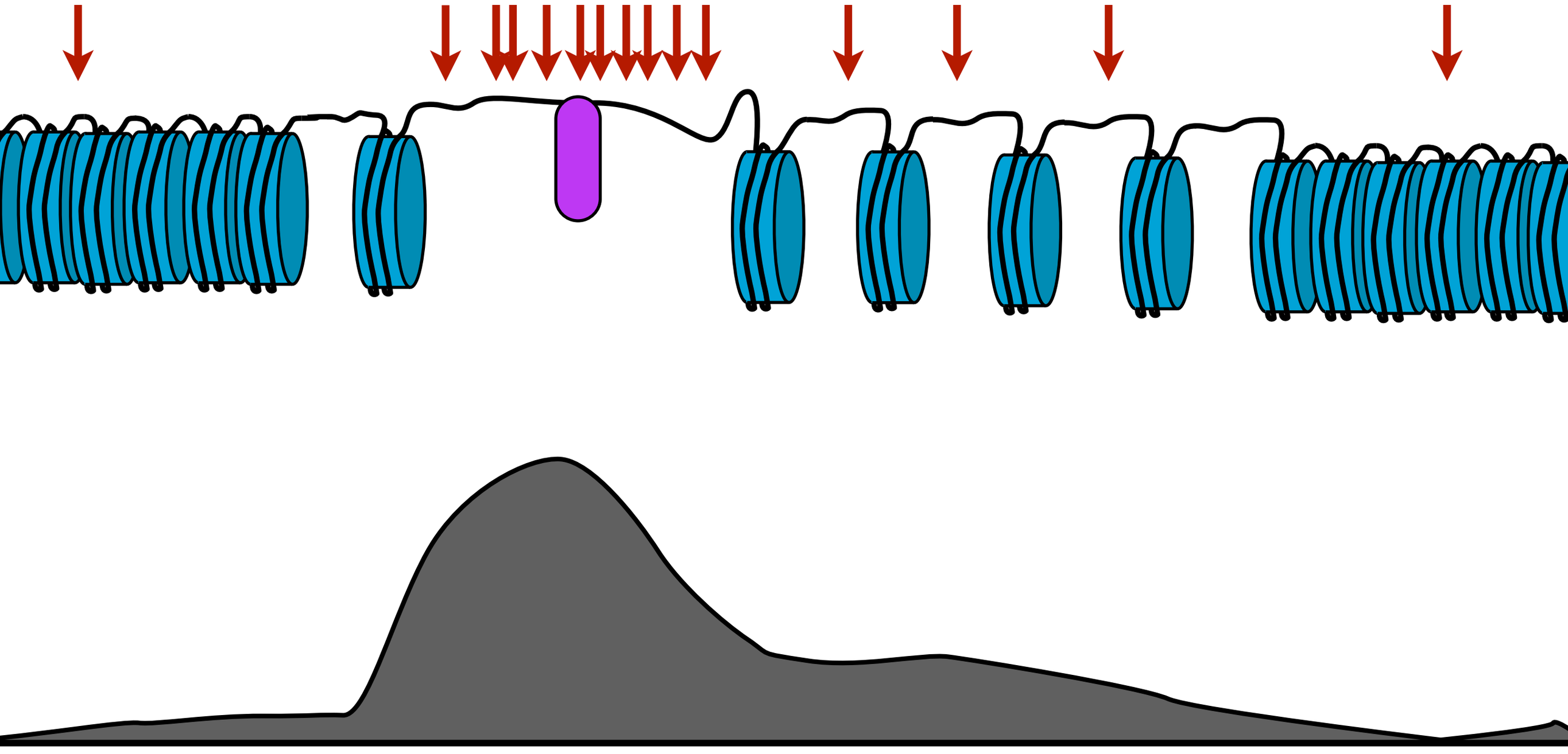
**Buenrostro, et. al. Nature Methods, 2013**

# Transposase Hypersensitivity (ATAC-seq)



**Buenrostro, et. al. Nature Methods, 2013**

# Transposase Hypersensitivity (ATAC-seq)



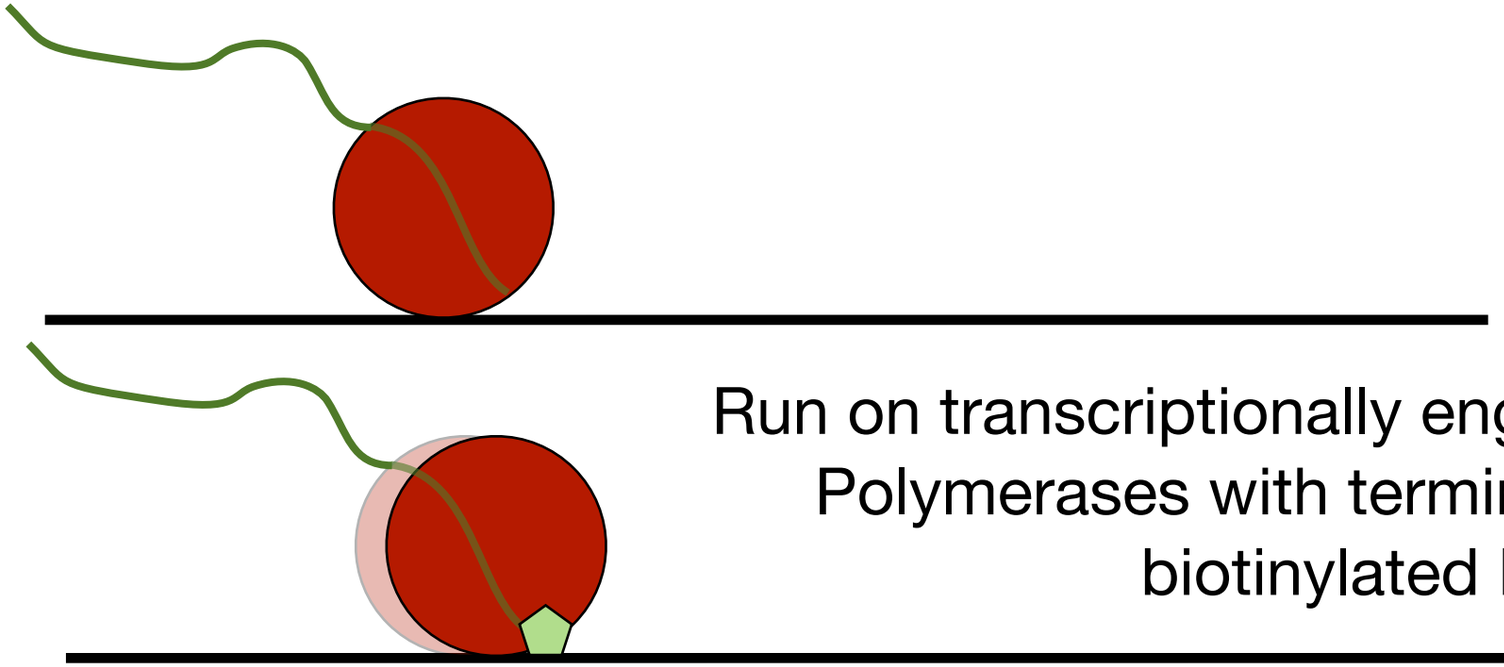
**Buenrostro, et. al. Nature Methods, 2013**

# PRO-seq detects nascent RNA



**Kwak, et. al. Science, 2013**

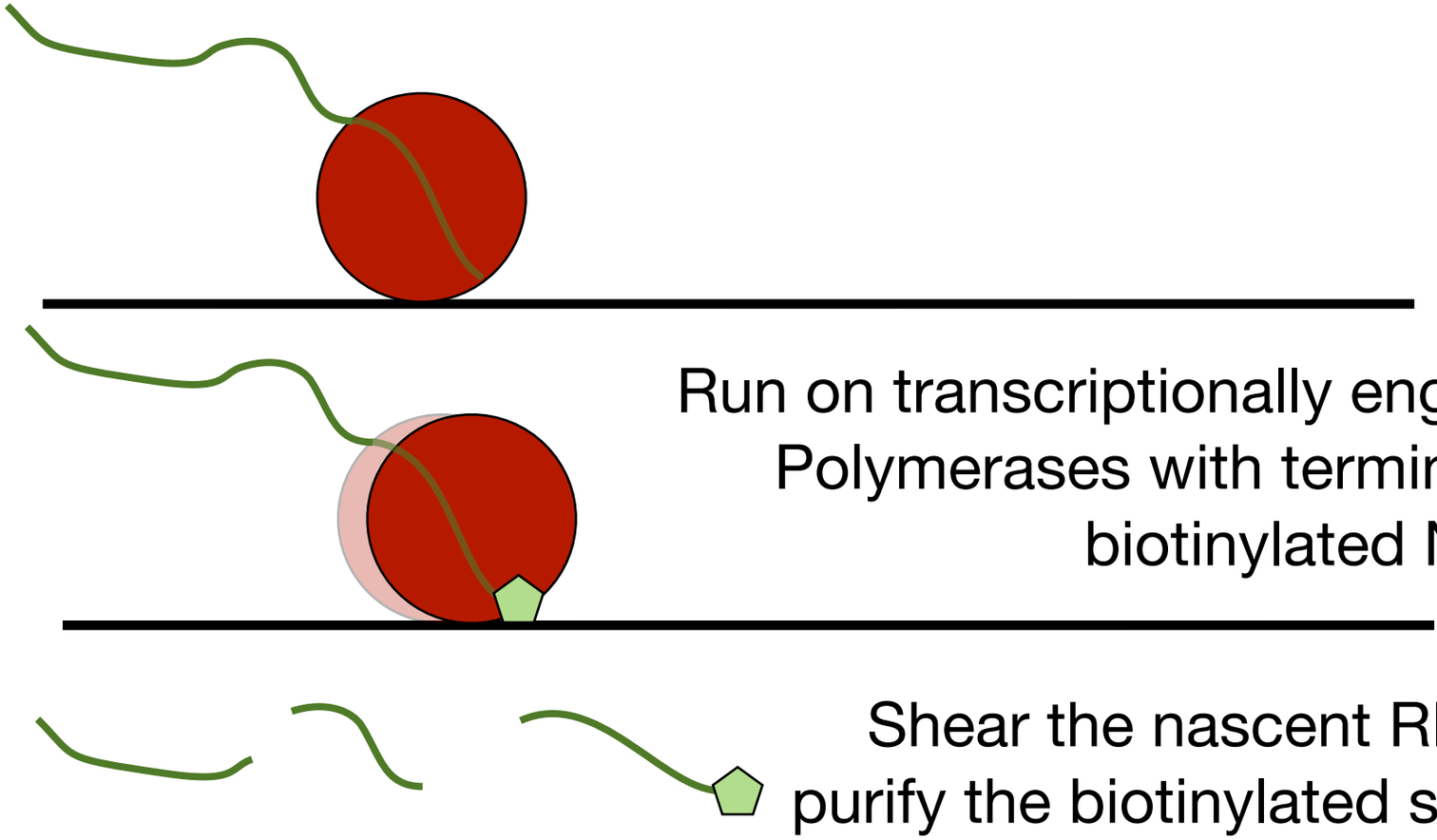
# Precision Genomic-Run On (PRO-seq)



Run on transcriptionally engaged Polymerases with terminating biotinylated NTPs.

**Kwak, et. al. Science, 2013**

# Precision Genomic-Run On (PRO-seq)

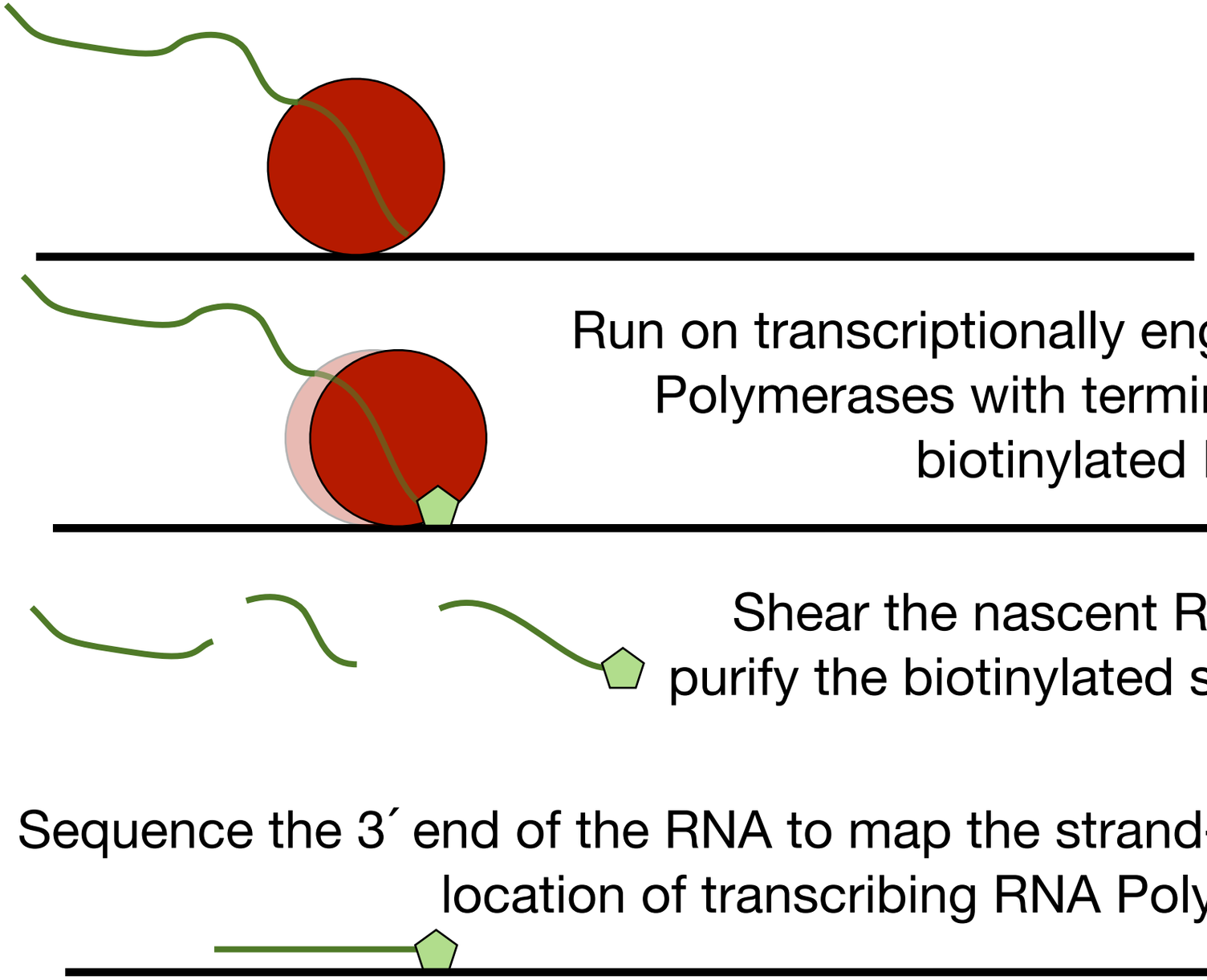


Run on transcriptionally engaged Polymerases with terminating biotinylated NTPs.

Shear the nascent RNA and purify the biotinylated species.

**Kwak, et. al. Science, 2013**

# Precision Genomic-Run On (PRO-seq)



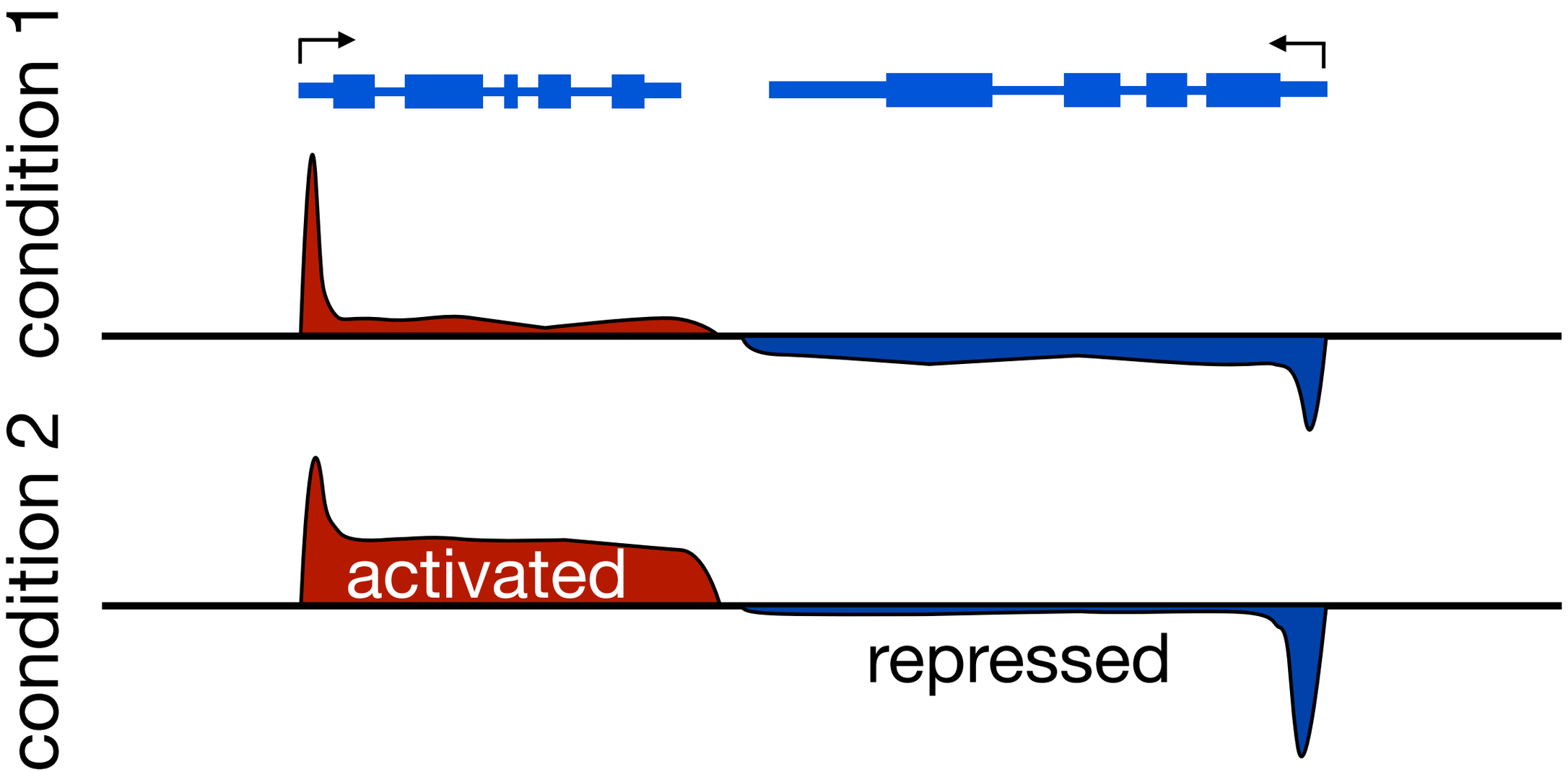
Run on transcriptionally engaged Polymerases with terminating biotinylated NTPs.

Shear the nascent RNA and purify the biotinylated species.

Sequence the 3´ end of the RNA to map the strand-specific location of transcribing RNA Polymerase.

**Kwak, et. al. Science, 2013**

# PRO-seq measures immediate responses



~200,000 peaks

ATAC signal

t0
20
40
60
120
180
240
6day

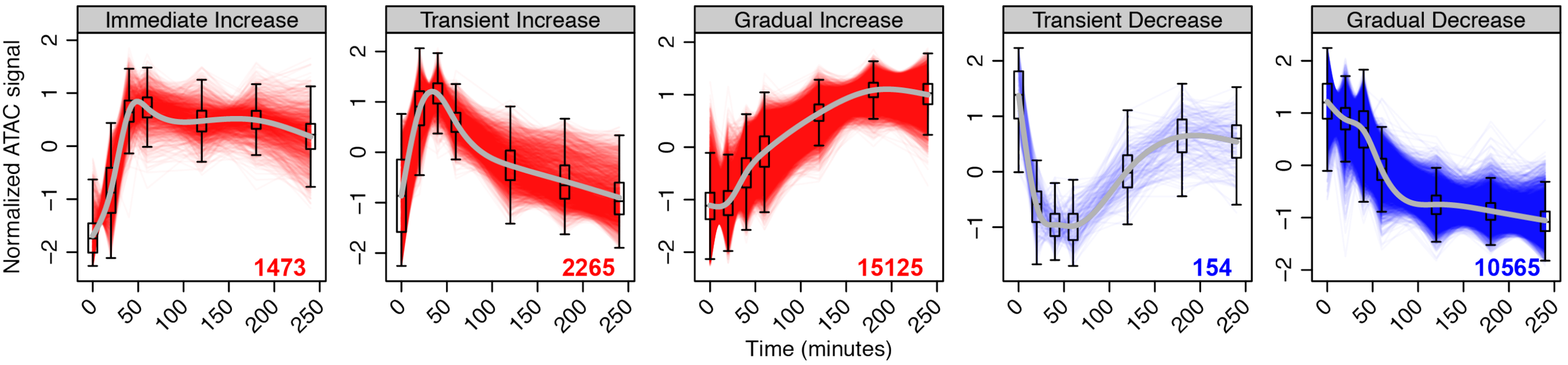
x2

t0_plus
t0_minus
20min_plus
20min_minus
40min_plus
40min_minus
60min_plus
60min_minus
2hr_plus
2hr_minus
3hr_plus
3hr_minus
4hr_plus
4hr_minus
6day_plus
6day_minus

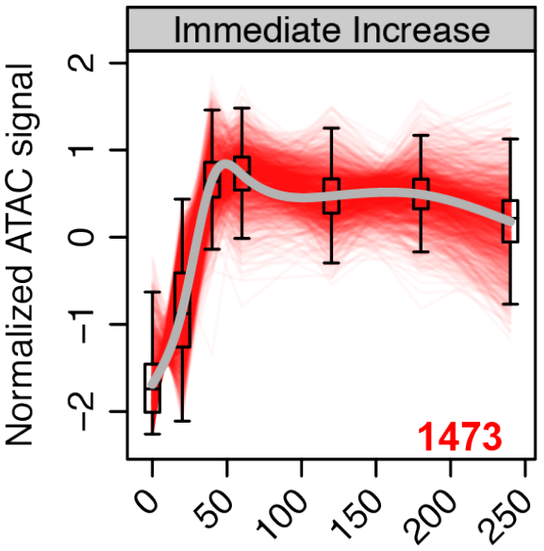
PRO signal

~30,000 genes

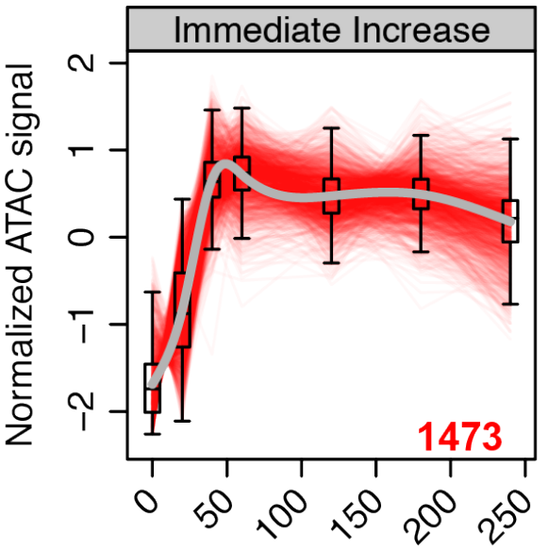
# Interesting ATAC-seq peaks are dynamic over the time course

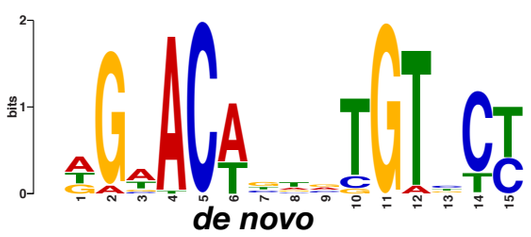


# *de novo* motif analysis identifies enriched sequence elements within dynamic ATAC peaks



# *de novo* motif analysis identifies enriched sequence elements within dynamic ATAC peaks






# Motif families that are enriched within dynamic ATAC peaks

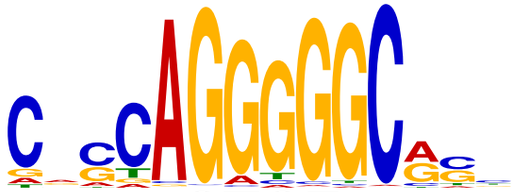

GR


TWIST


AP1


CEBP


bHLH


CTCF


SP


KLF


ELF


NFY

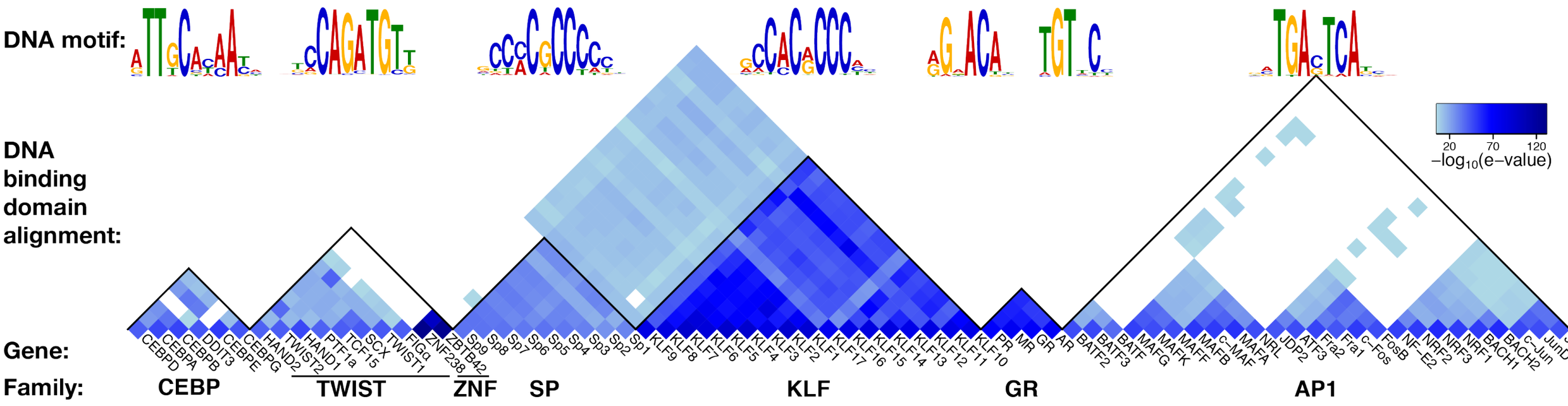

STAT


NRF


TFAP2


TEAD

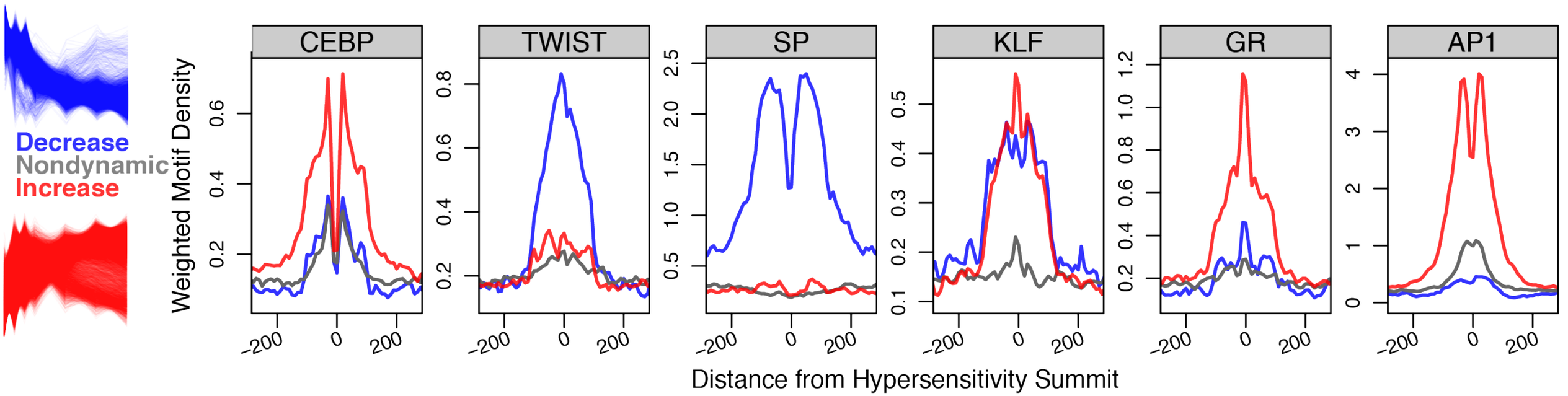
# 14 TF-family motifs (top 6 shown) drive early changes in chromatin and transcription



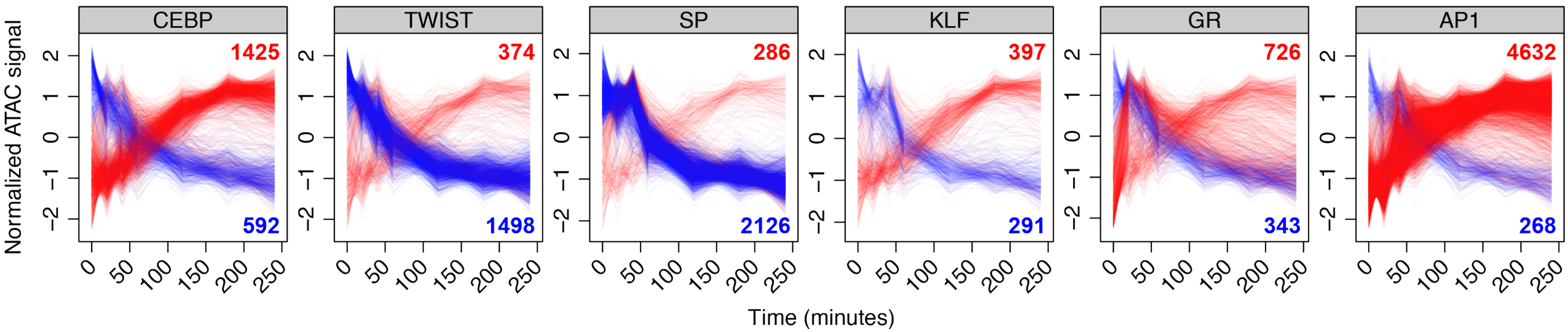

One motif ≠ One TF

Paralogous TF DBD families can recognize each motif

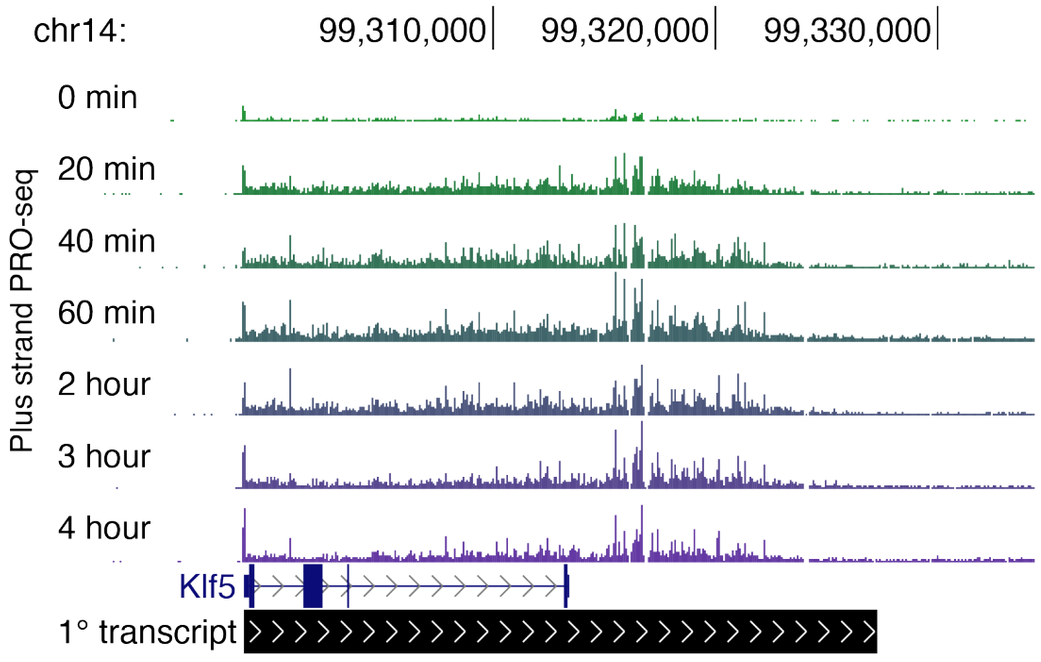
# CEBP, KLF, GR, and AP1 motifs associate w/ increasing chromatin accessibility
# TWIST and SP associate w/ decreasing chromatin accessibility

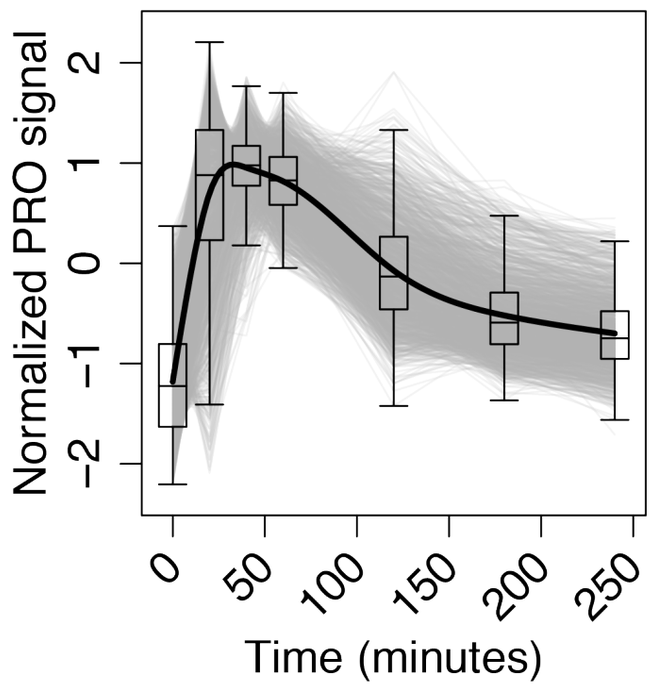


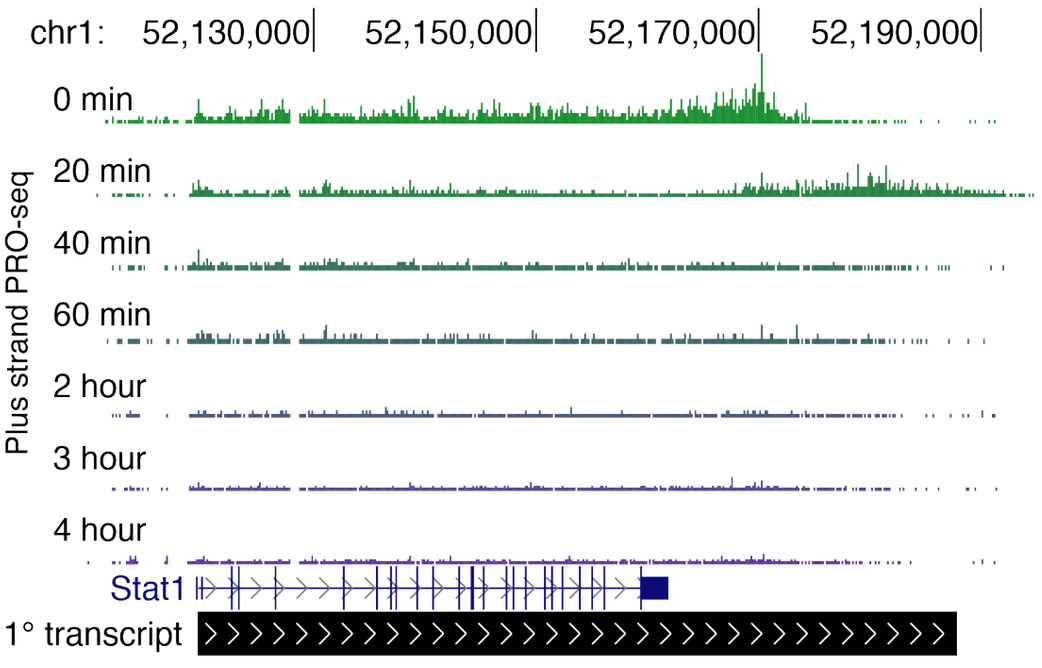
# CEBP, KLF, GR, and AP1 motifs associate w/ increasing chromatin accessibility
# TWIST and SP associate w/ decreasing chromatin accessibility

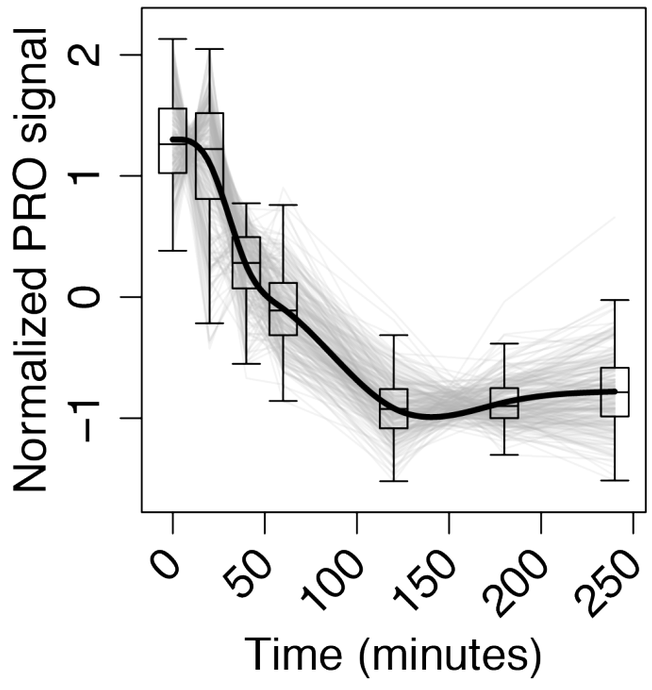


# Genes also have distinct activation and repression kinetics

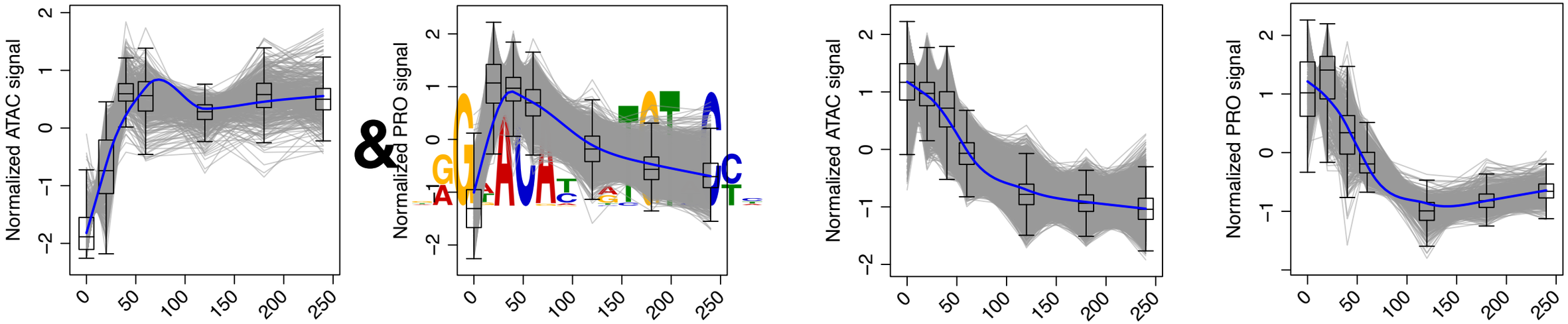


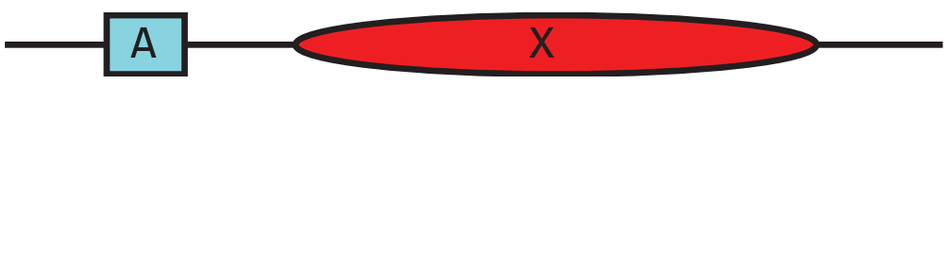



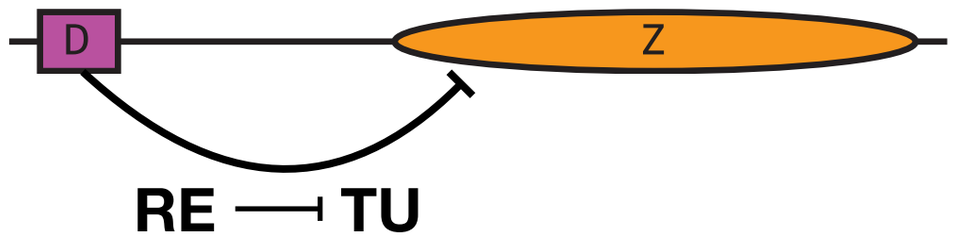



# Linking direct functional interactions between Regulatory Elements (RE) and Genes (TU)



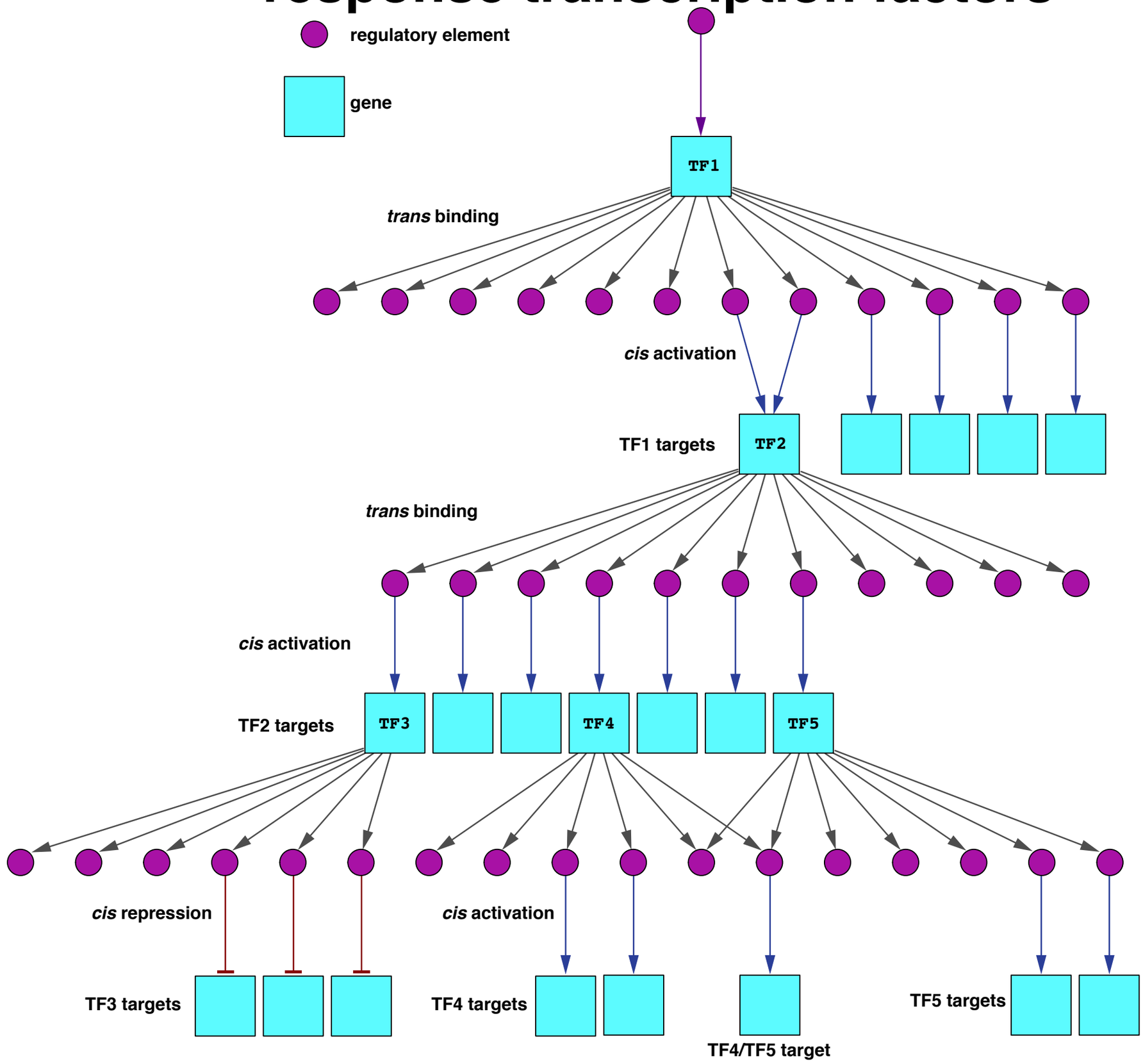




ATAC peak ↑ & gene PRO ↑ & time delay (PRO prior ATAC) & peak and gene are nearby

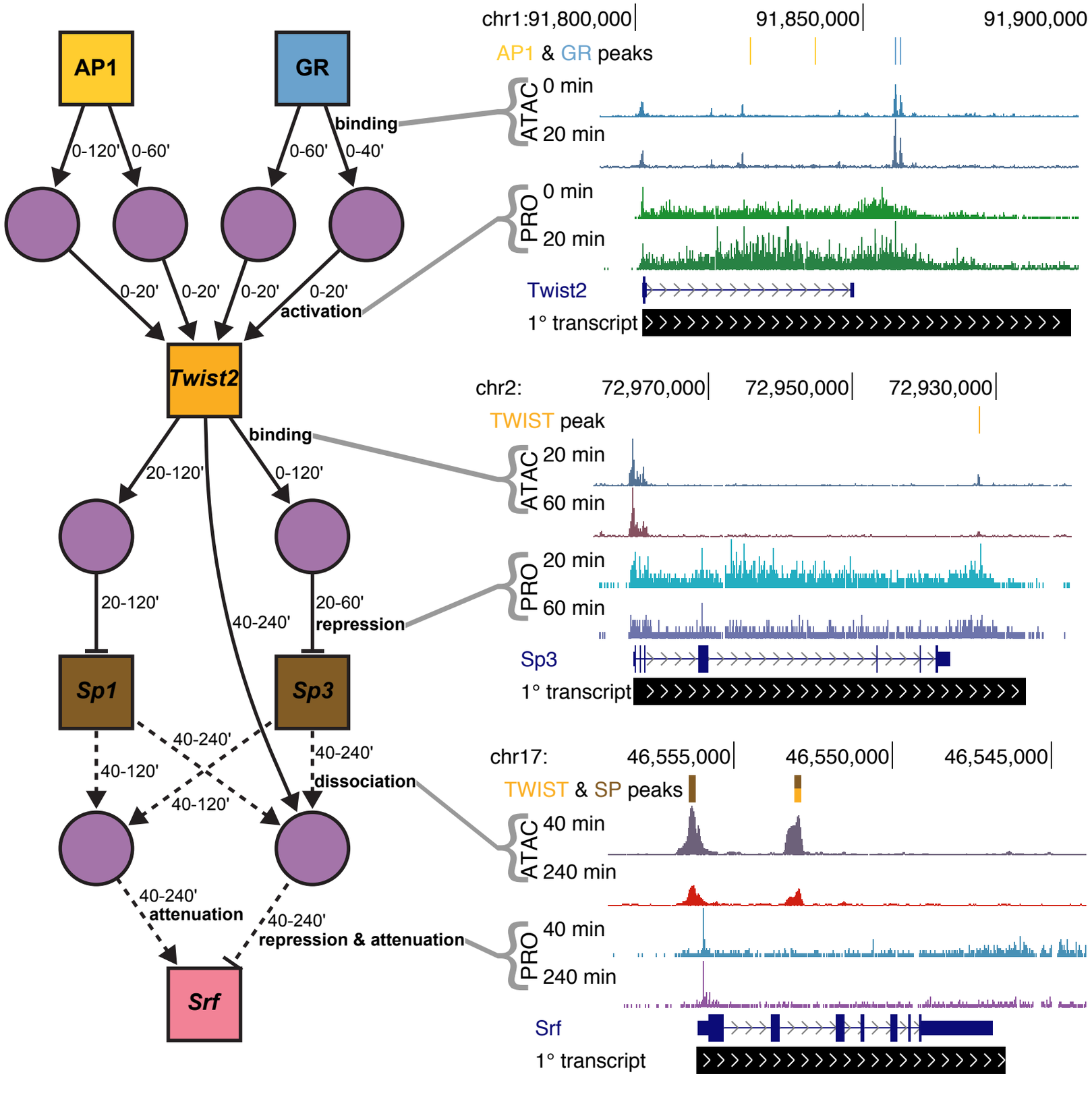
(repressor binding)

gene's motif within peak & ATAC peak ↓ & gene PRO ↓ & peak and gene are nearby

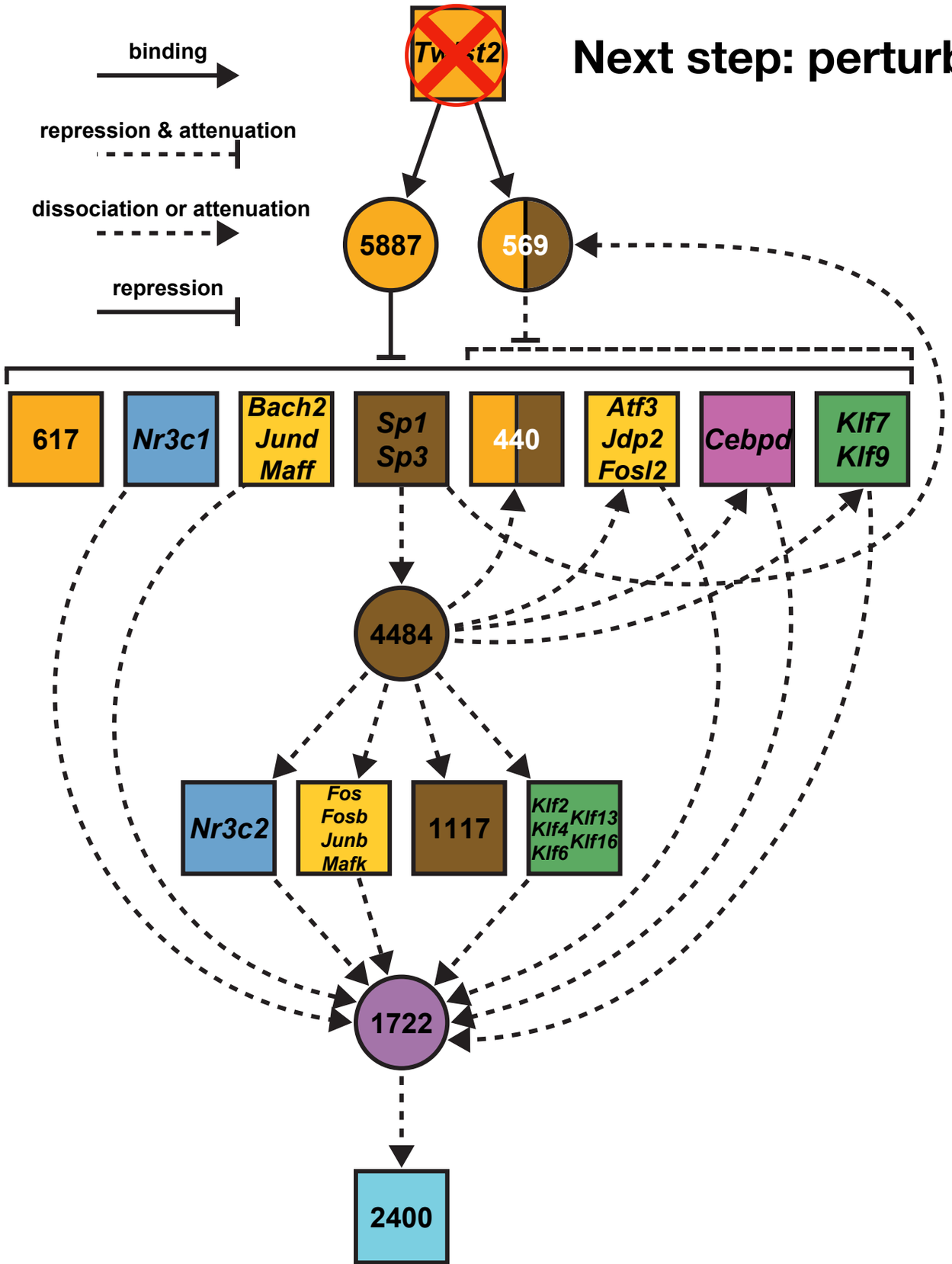
# We are interested in highly connected early response transcription factors



# TWIST2 is active early and transiently

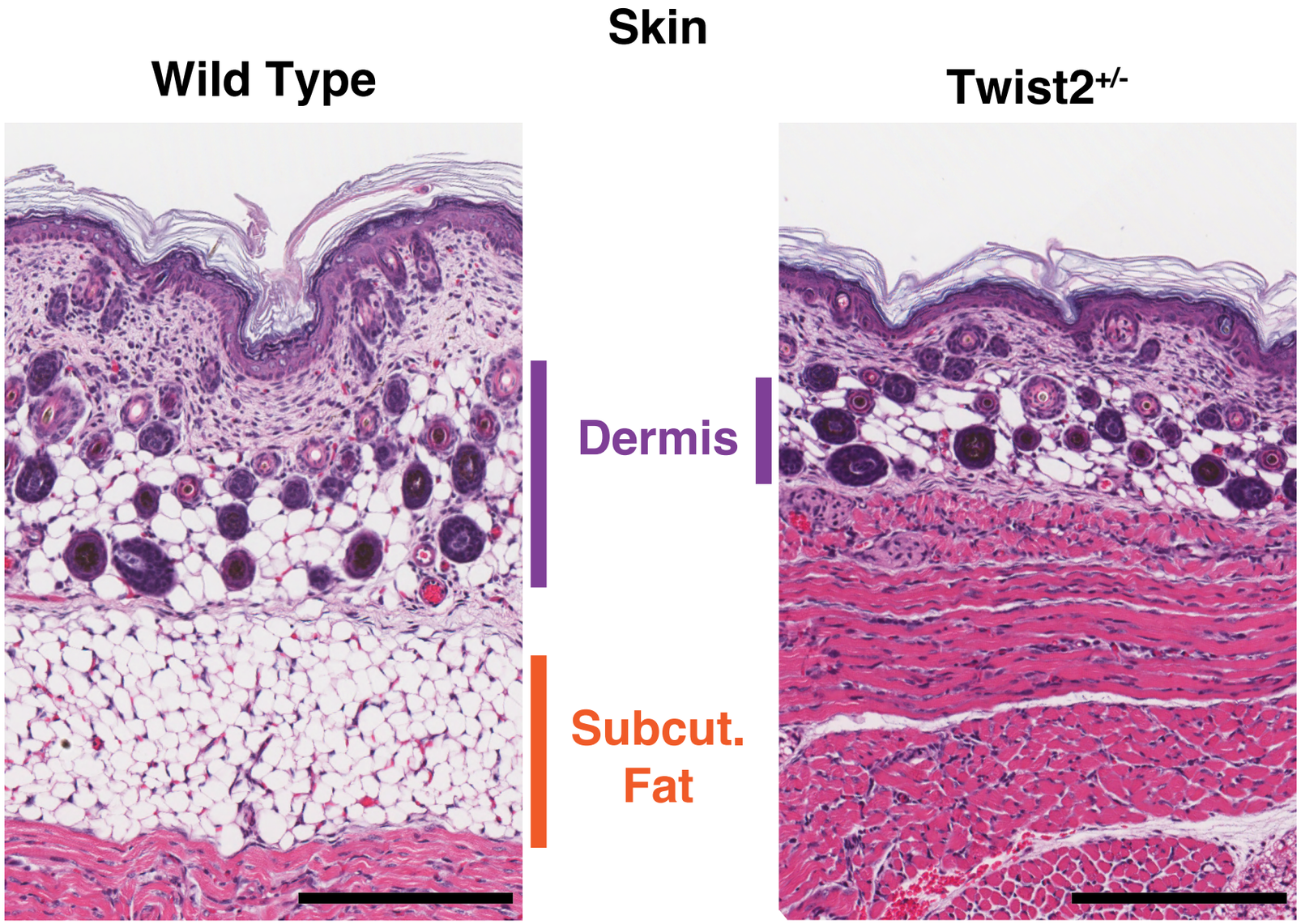


# TWIST2 is a highly connected early response gene

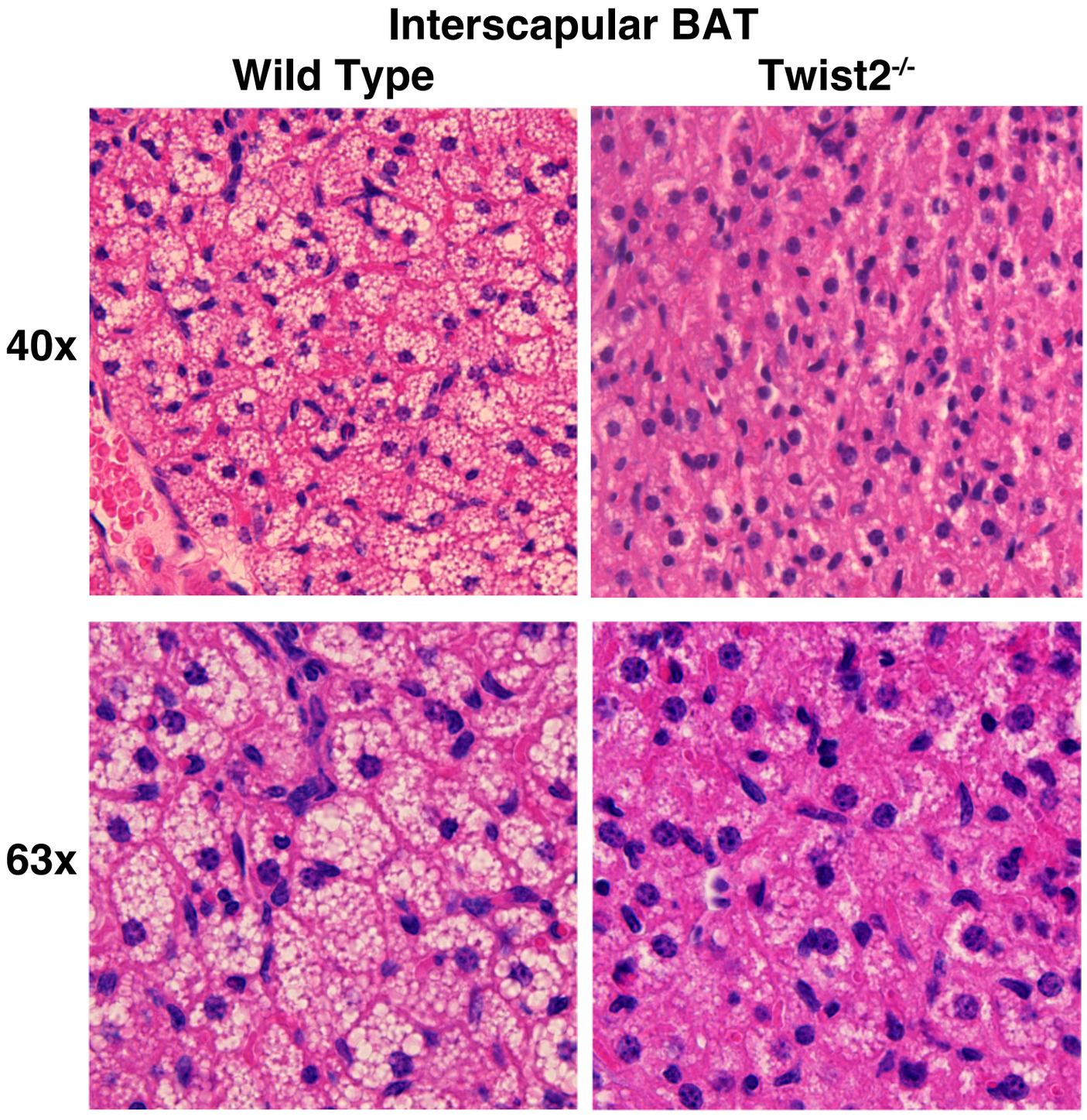



Next step: perturb and observe

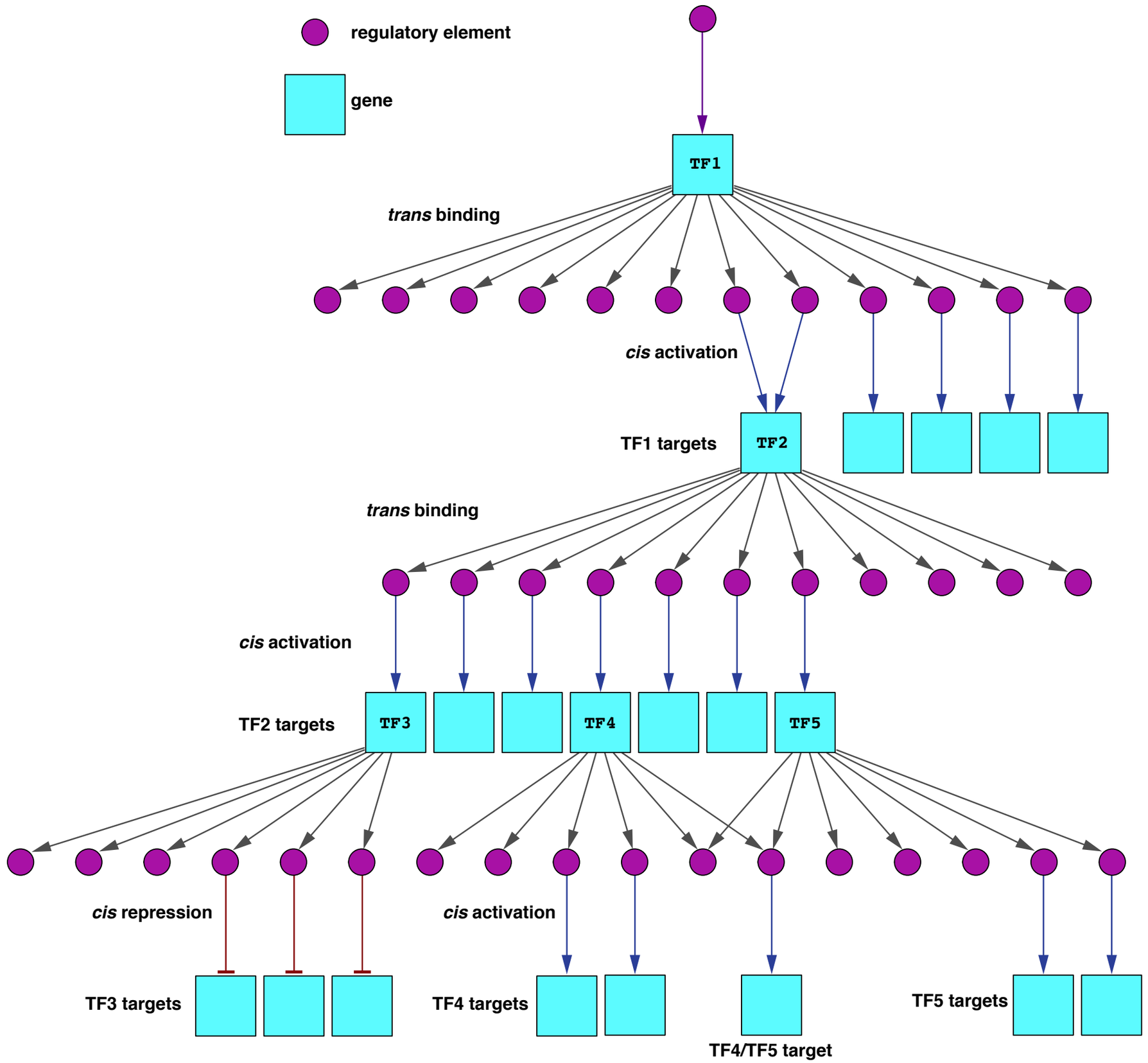
# Twist2$^{+/-}$ mice have a near absence of subcutaneous fat



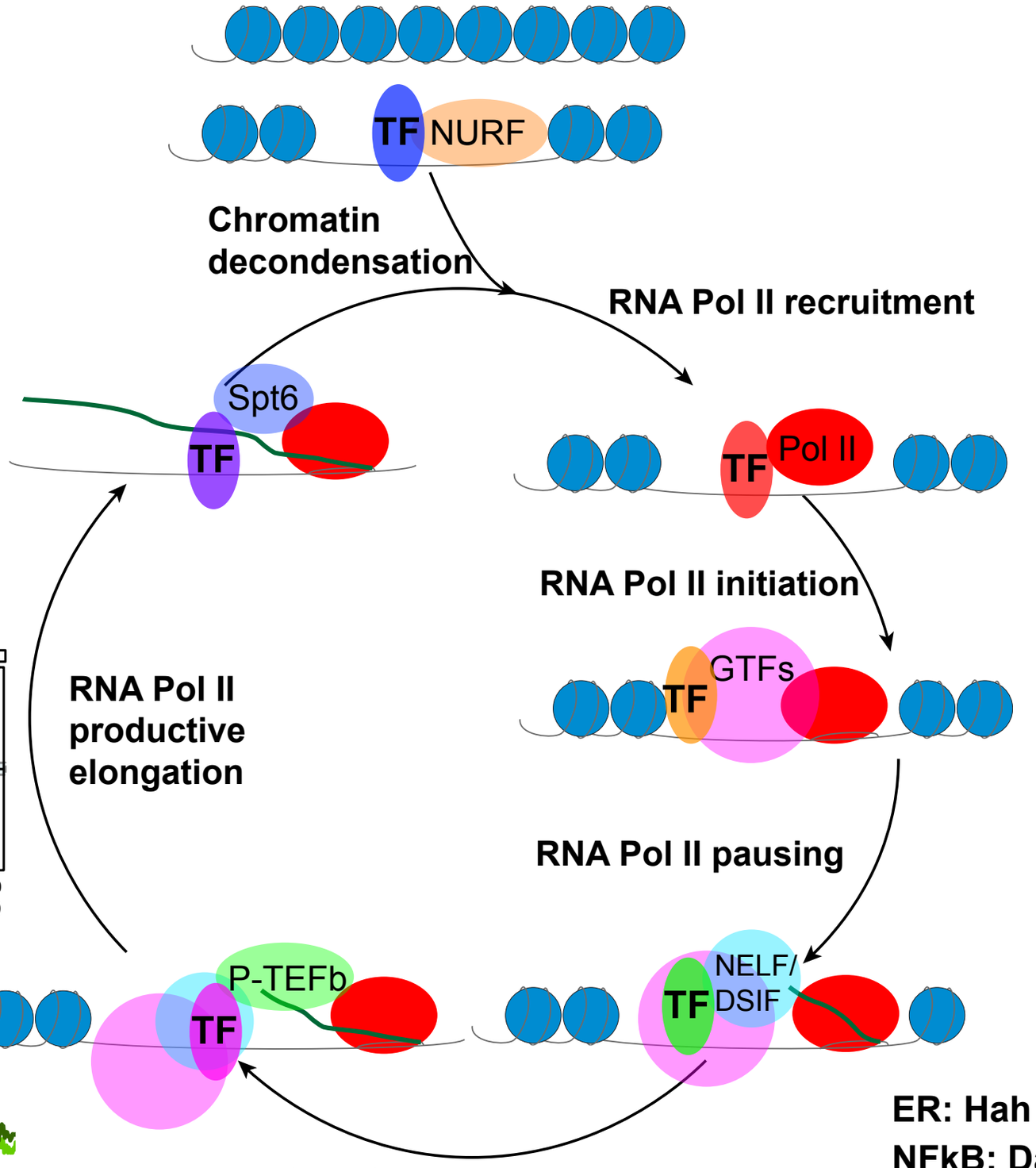
# Twist2$^{-/-}$ mice have reduced lipid droplets in brown fat



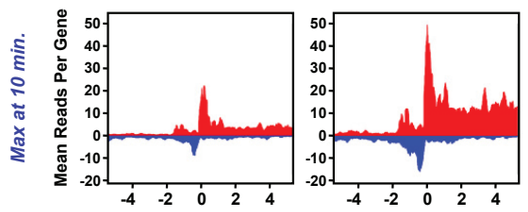
# What are the molecular functions of these key TFs?



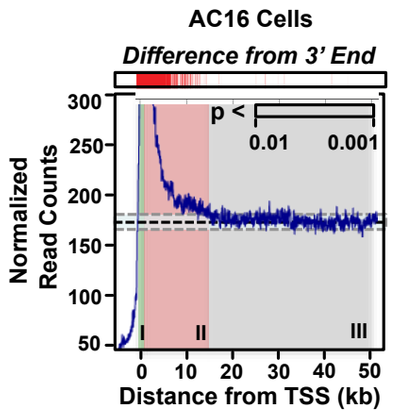
# Coupling rapidly inducible systems to nascent transcript profiling informs on TF function




10-40 min Estrogen

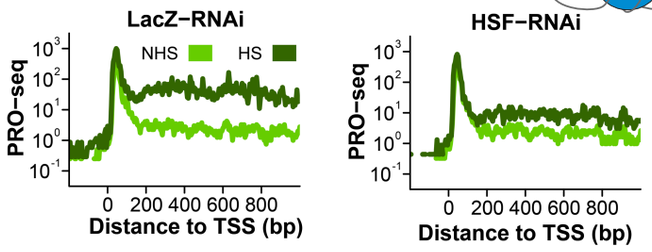



10-30 min TNFα



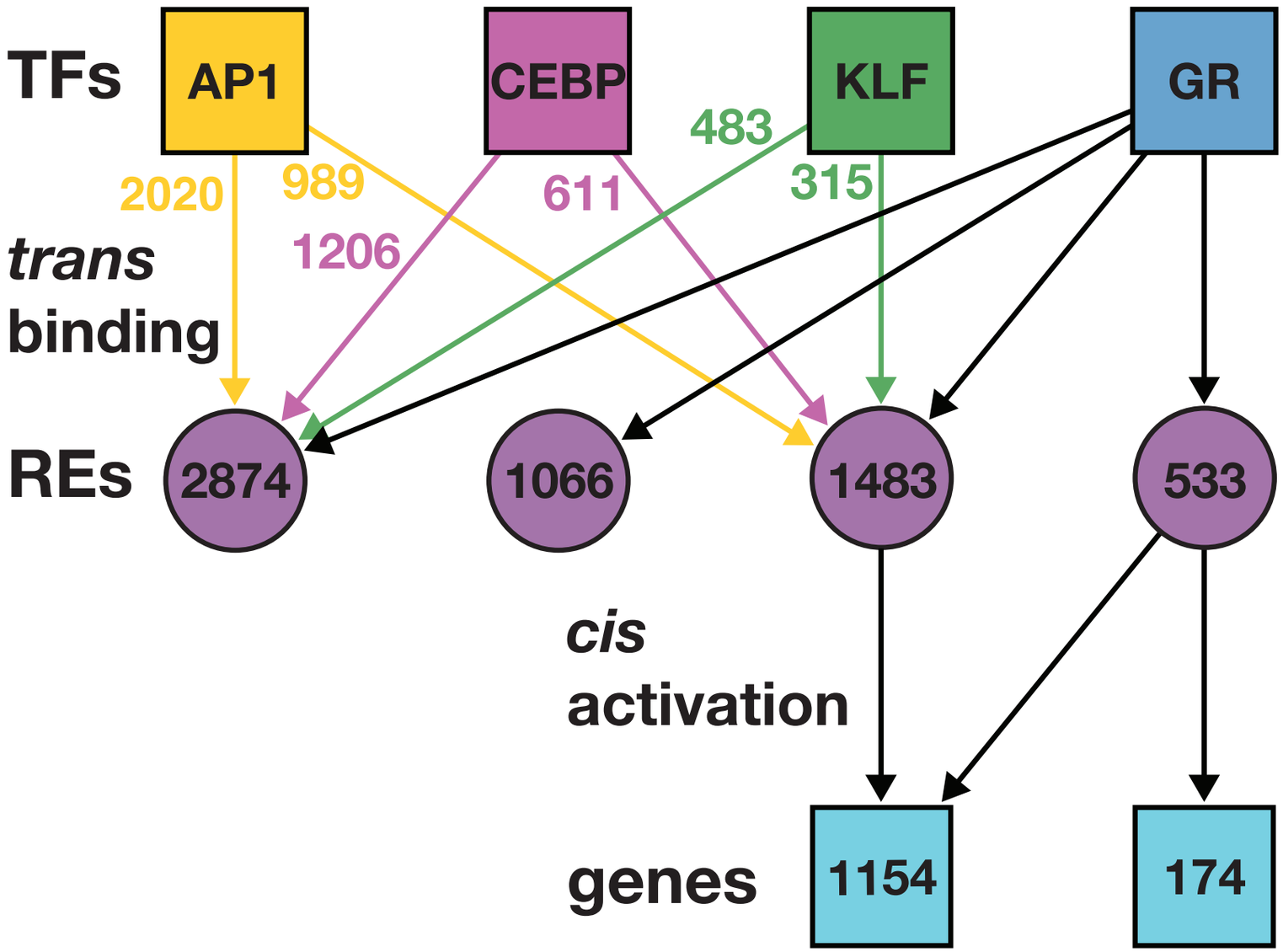

I. Paused
II. Accelerating
III. Fully Processive

10-20 min heat shock

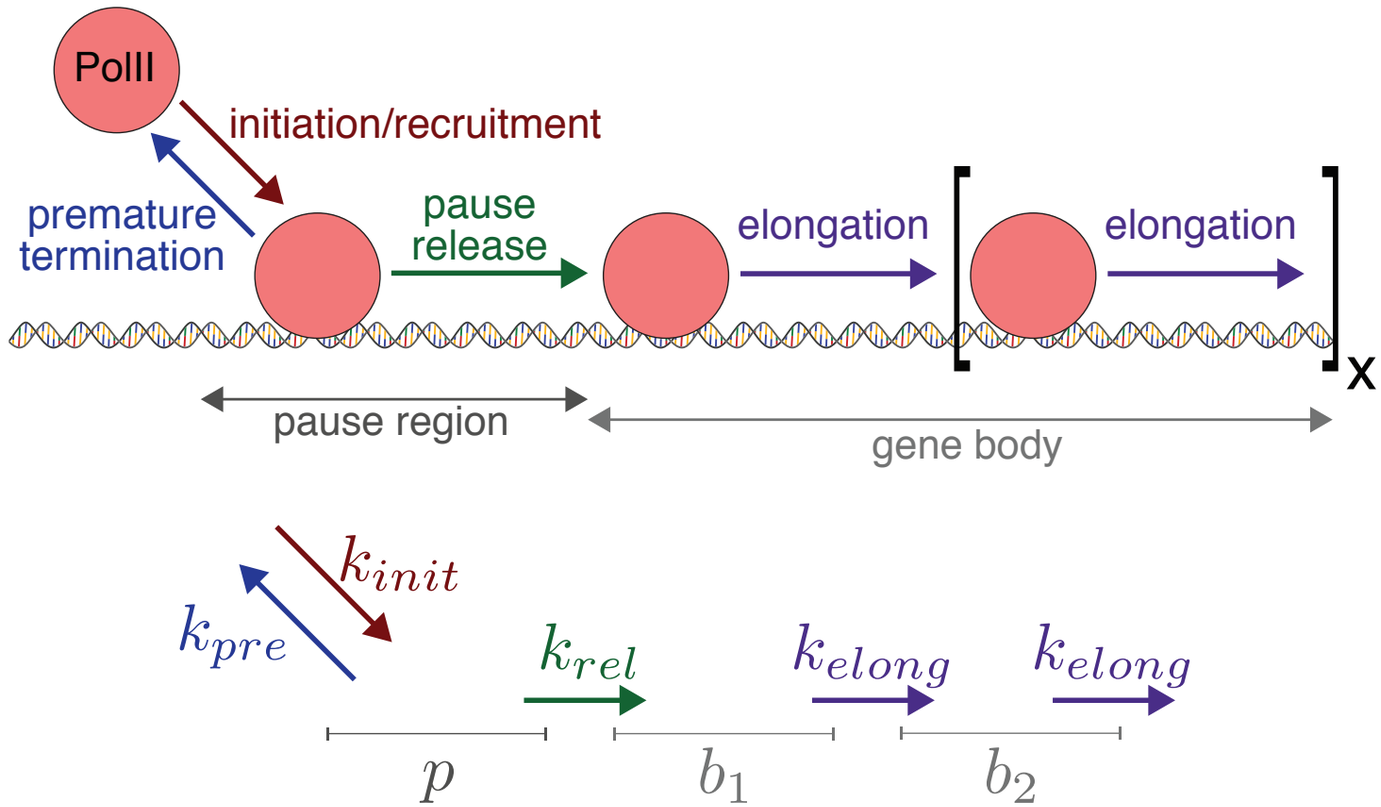


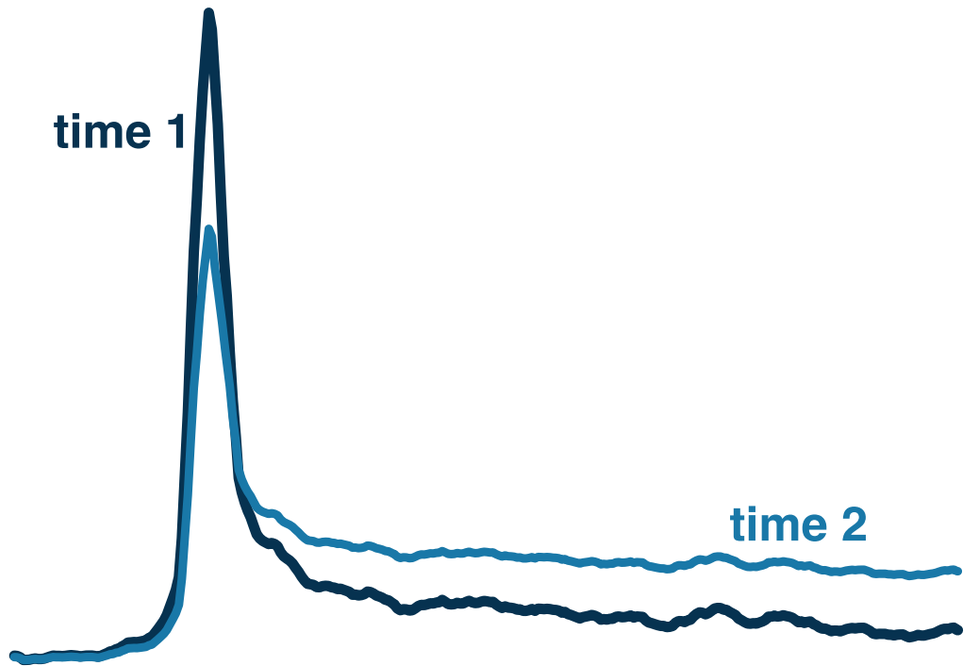

ER: Hah et. al., Cell, 2011
NFkB: Danko, et. al. Mol. Cell, 2014
HSF: Duarte et. al., G&D, 2016

# Simplified networks identify genes that are primarily regulated by a single factor



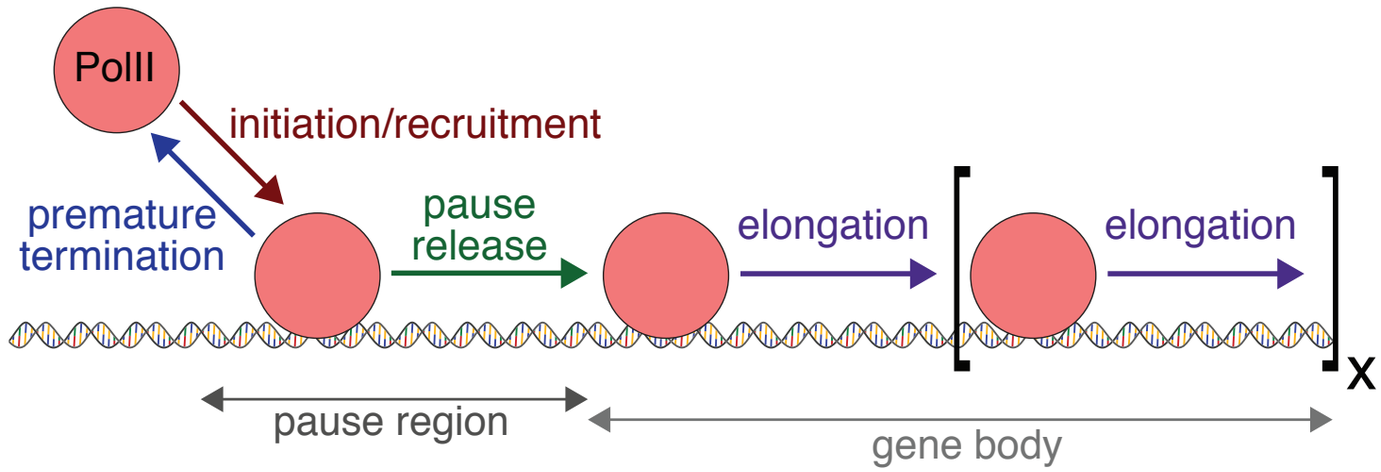
# Estimating rates with a model that considers RNA Polymerase density within the pause region and gene body

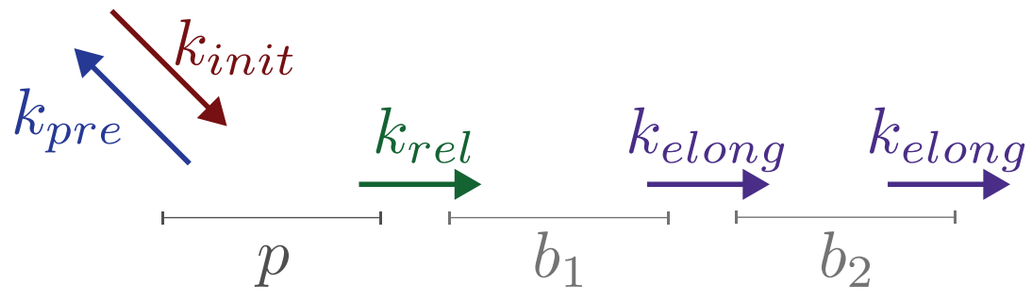



**RNA polymerase composite profile**

# We can determine the step(s) that a TF regulates by quantifying RNA polymerase density changes in genic regions



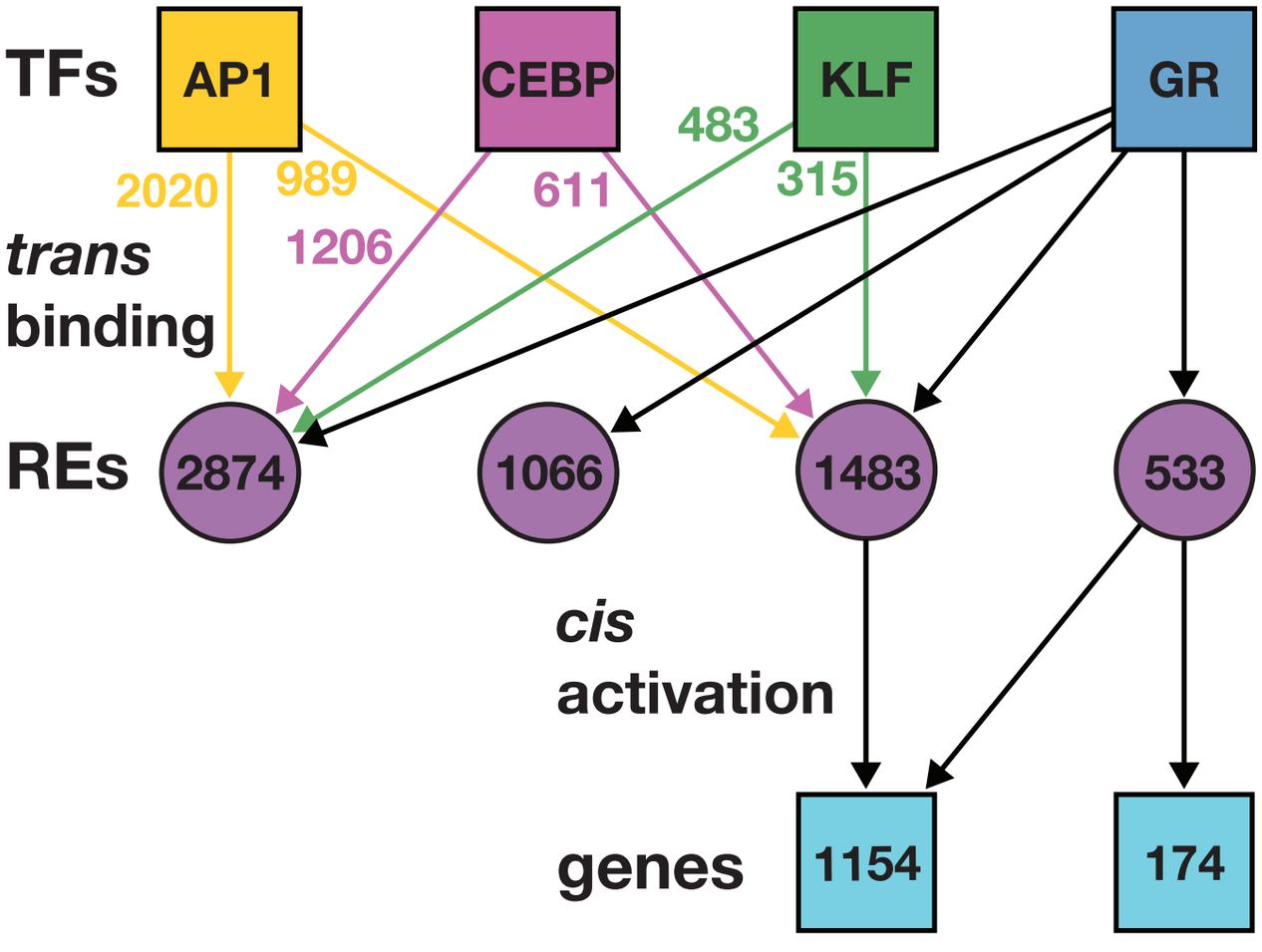

$$\frac{dp}{dt} = k_{init} - (k_{pre} + k_{rel})\, p$$

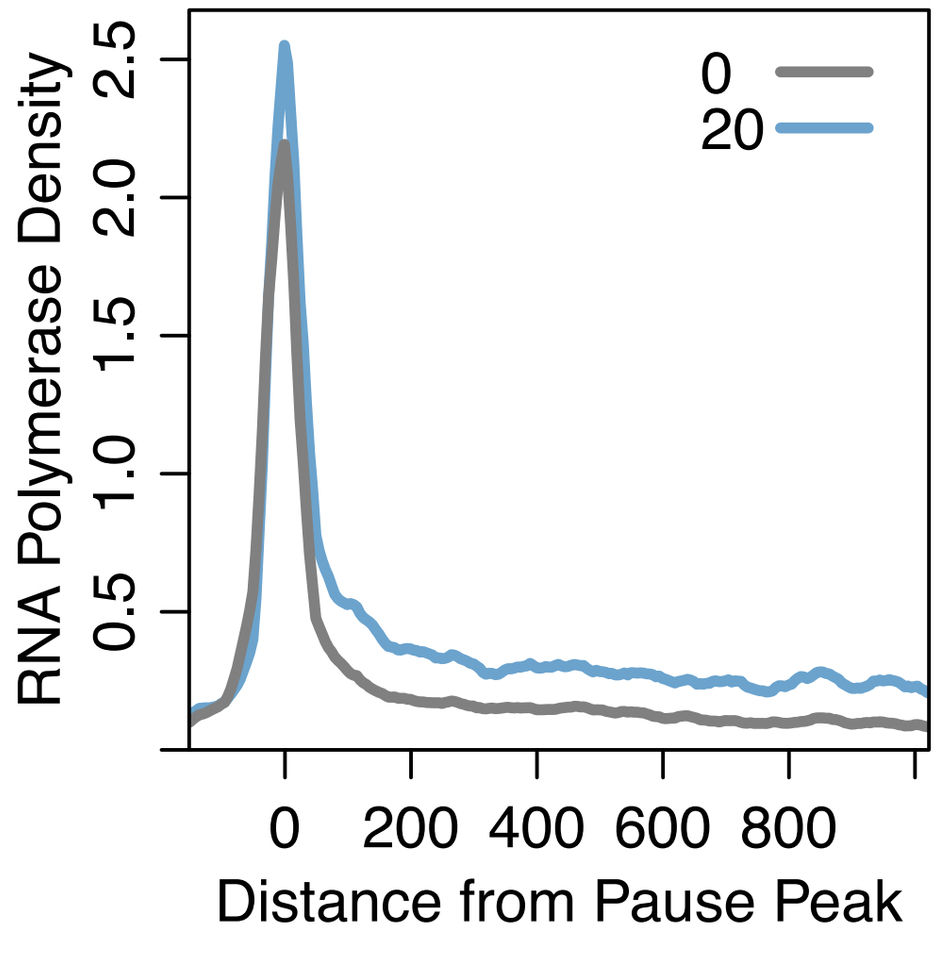
$$\frac{db_1}{dt} = k_{rel}\, p - k_{elong}\, b_1$$

$$\frac{db_2}{dt} = k_{elong}\, b_1 - k_{elong}\, b_2$$

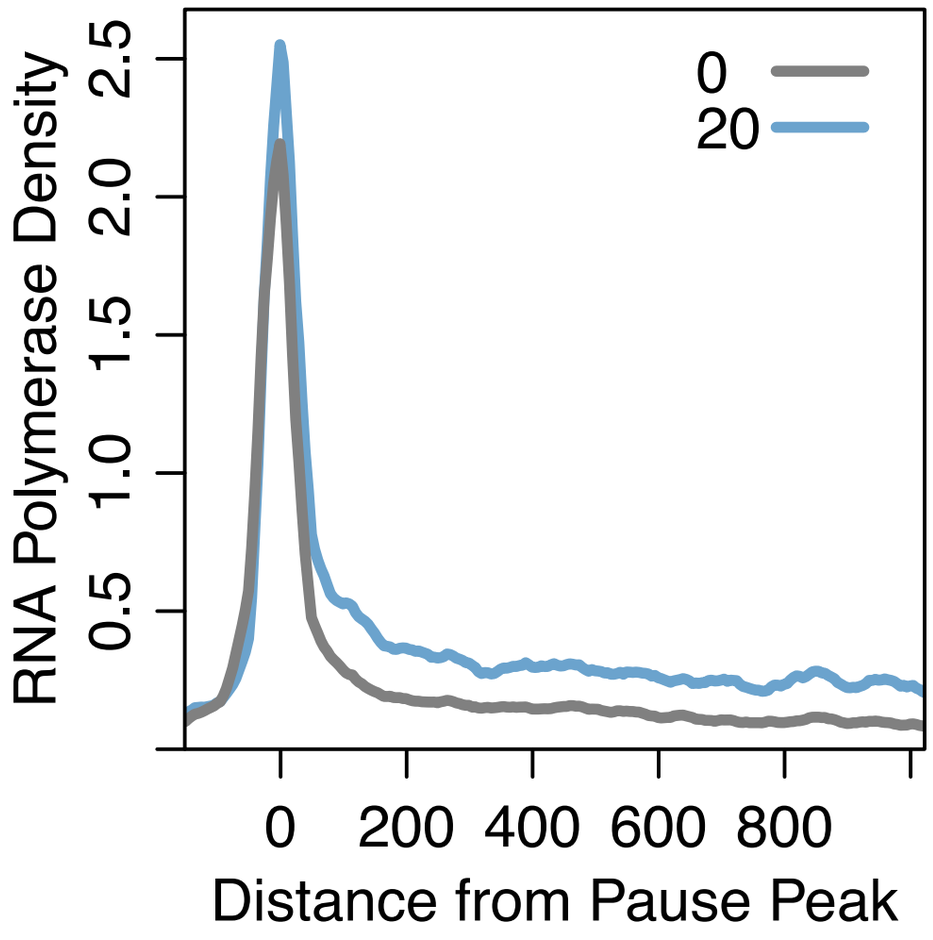
$$\frac{db_x}{dt} = k_{elong}\, b_{x-1} - k_{elong}\, b_x$$

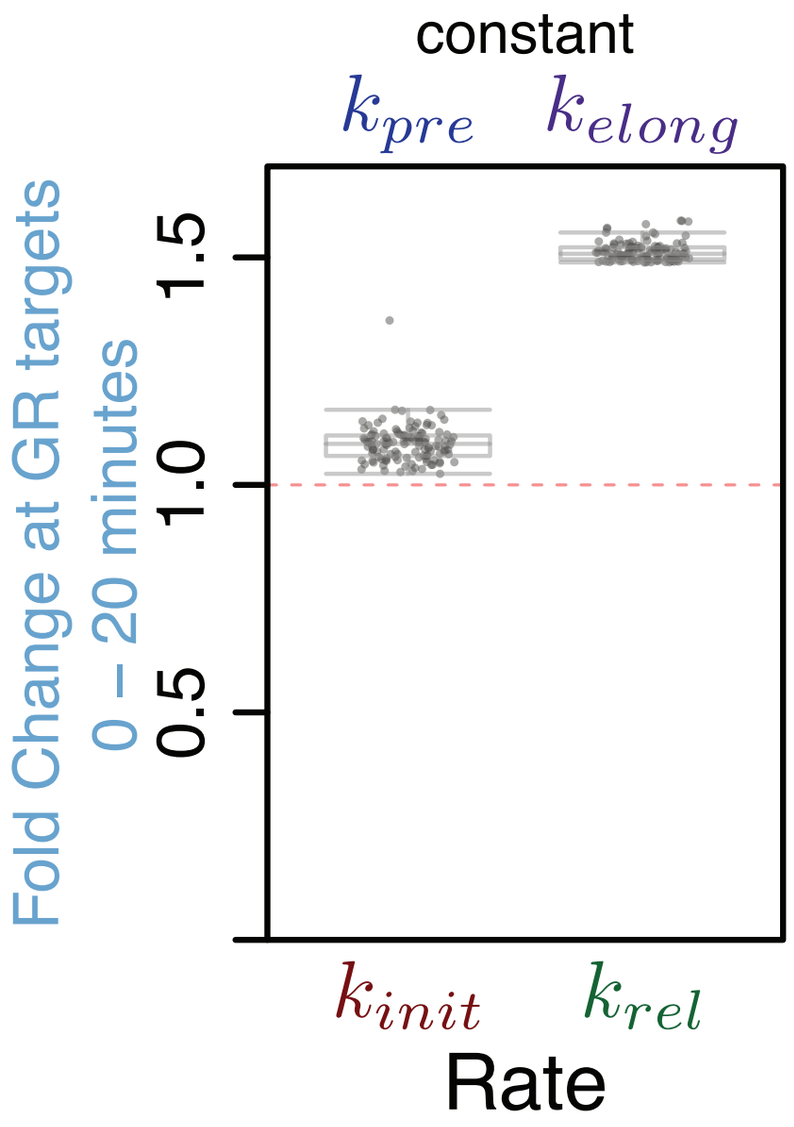
# Simplified networks identify genes that are primarily regulated by a single factor



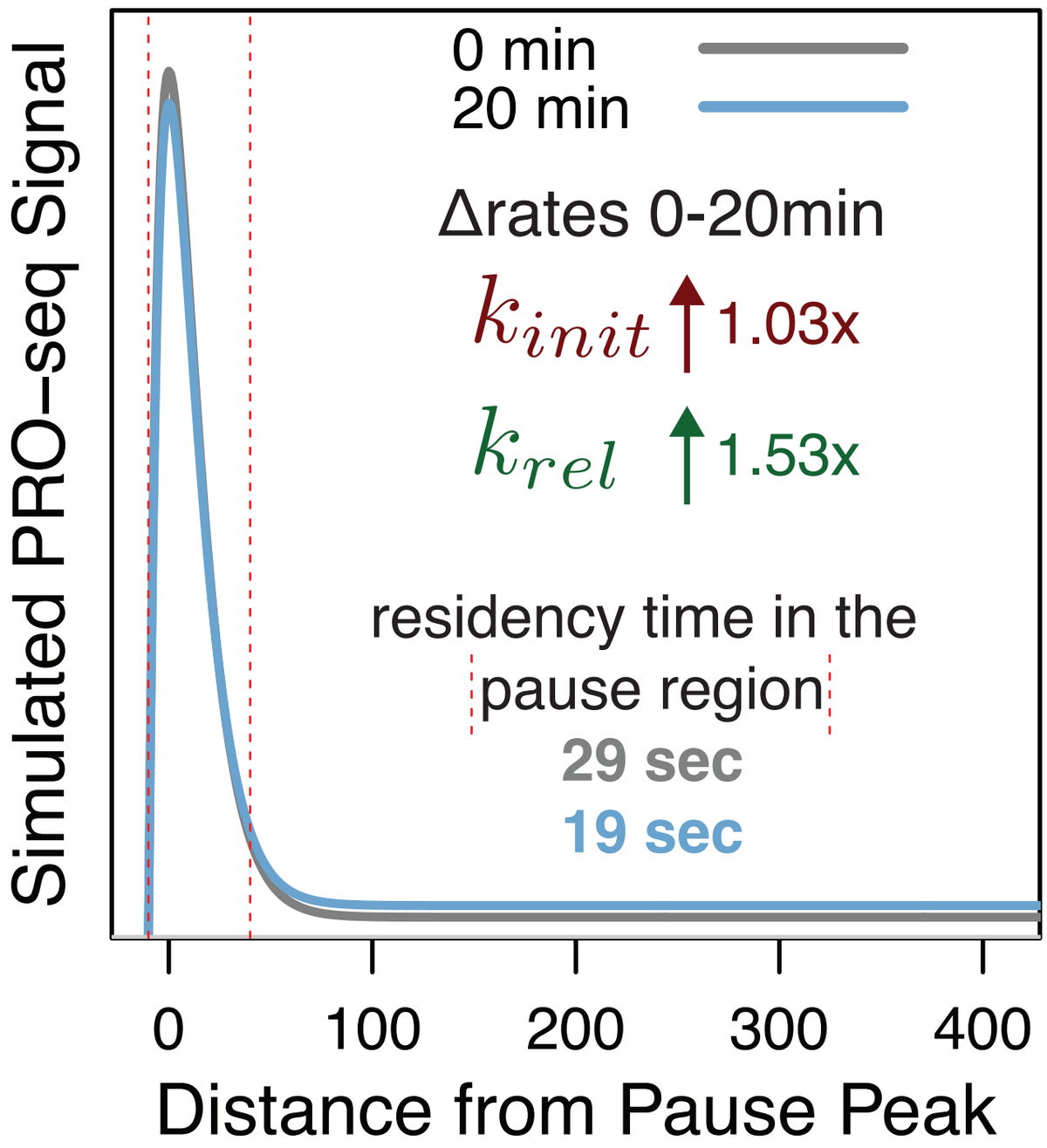



# GR preferentially regulates pause release

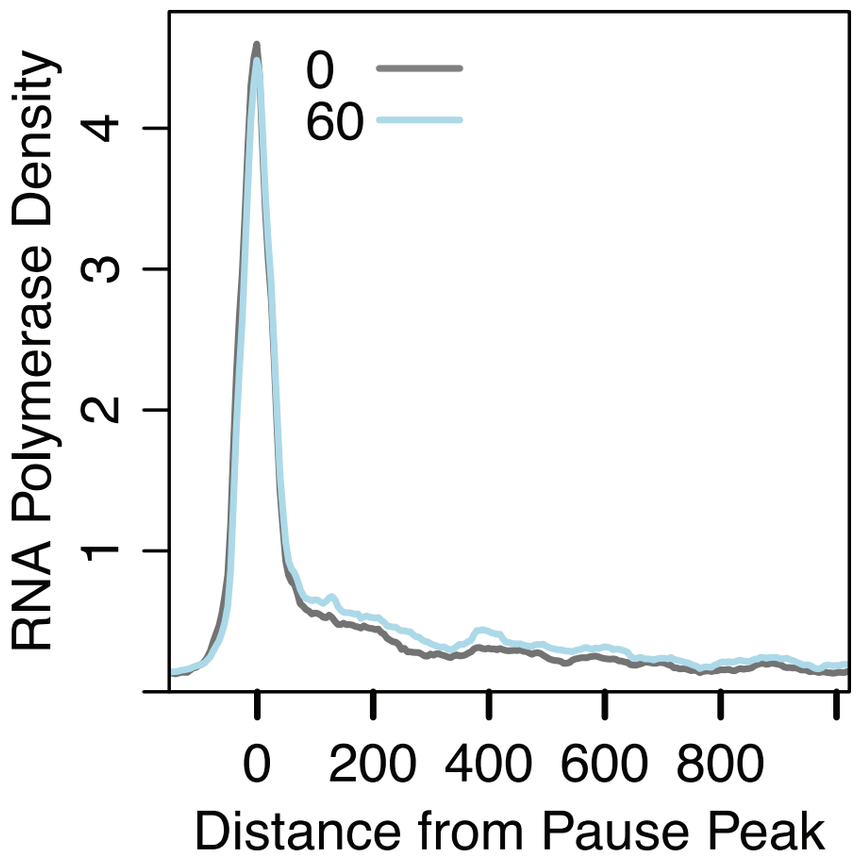


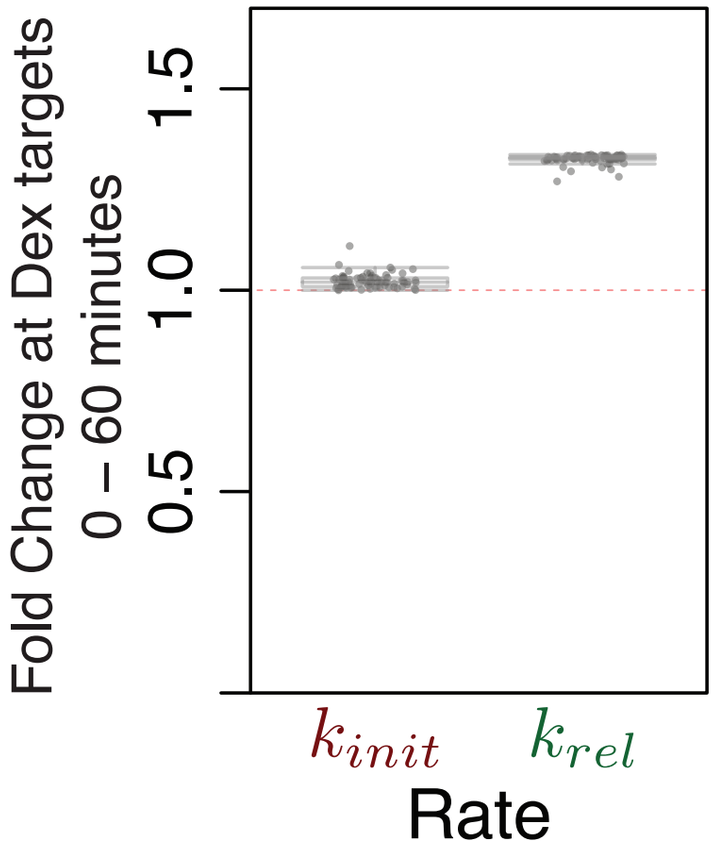



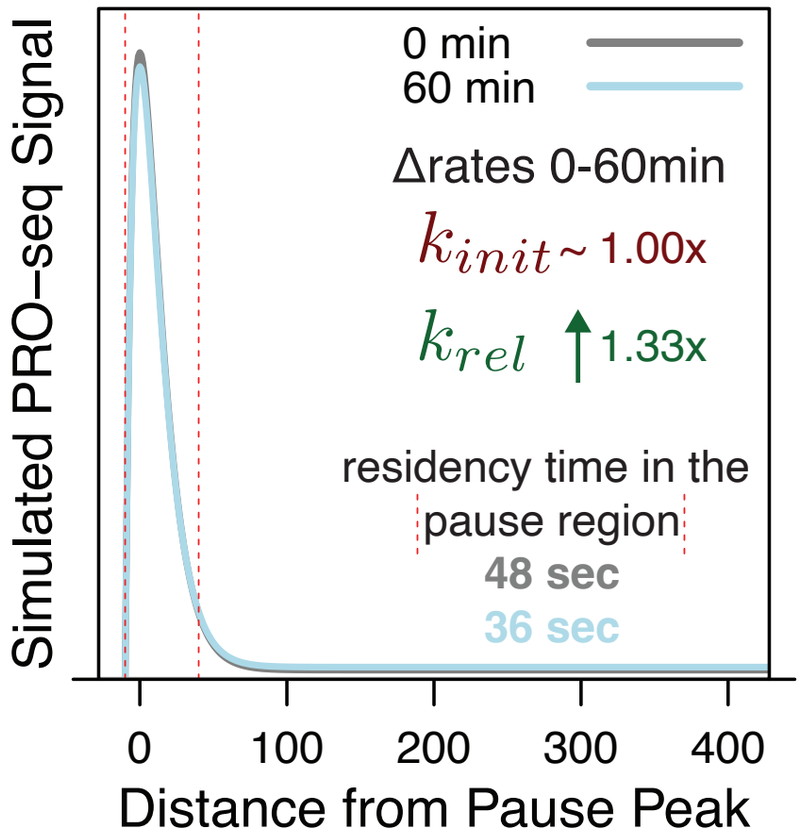
# GR preferentially regulates pause release



# GR preferentially regulates pause release in another system: Leukemia cells treated with dexamethasone for 1 hour



# Adipogenesis Network Conclusions

- Kinetic ATAC and PRO data can be used to infer key functional interactions between regulatory elements and genes.

- TWIST2 is a highly connected node in the adipogenesis network.

- TWIST2 deletion cause an absence of subcutaneous fat deposits.

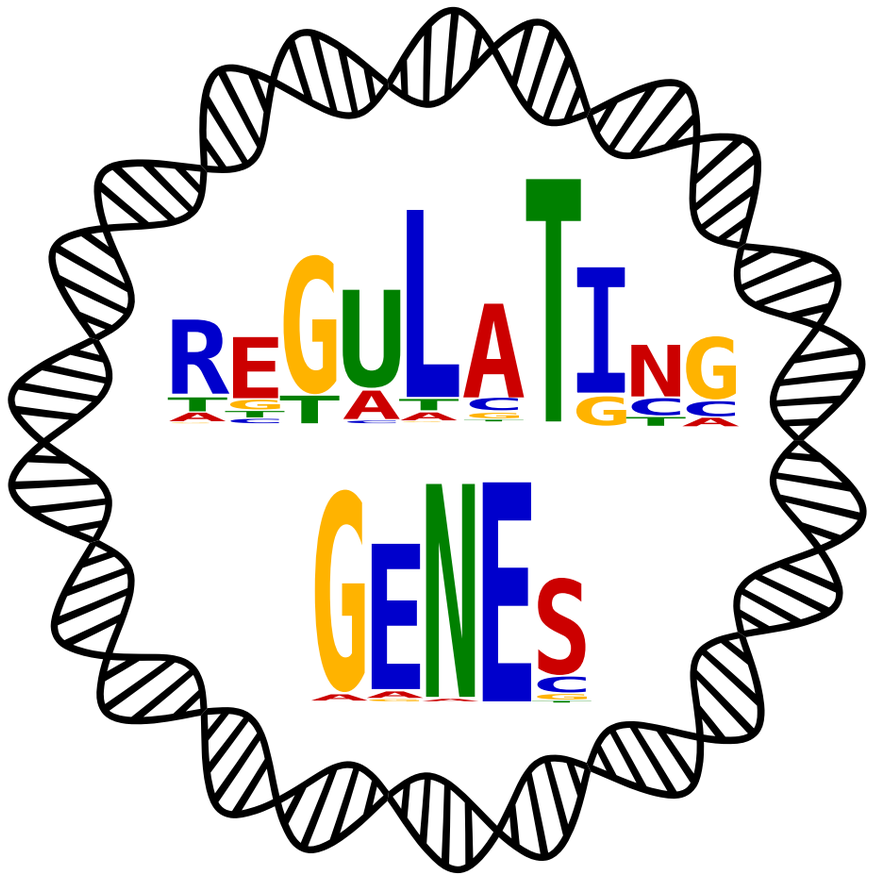
- GR regulates pause release

- Rapidly inducible systems are necessary to provide these mechanistic insight.

# TWIST2 Future

- We generated a conditional Twist2 KO mouse
- Isolate the effect of Twist2 in proadipocytes and perform organismal phenotyping (glucose tolerance/insulin sensitivity)
- Molecular mechanisms of Twist1 and Twist2 coopertivity and antagonism

# 2) Mechanisms of direct *cis* Repression by a Transcription Activator



# A history of rapidly inducible TFs we study.
# A new era and democratizing the study of TFs.

Inherent rapid inducibility

- 2010-2021: Heat Shock Factor in Drosophila, yeast, and human cells
- 2012-2023: Estrogen Receptor in human cells
- 2017-2023: Glucocorticoid Receptor in rats and mouse & human cells
- 2023: Androgen Receptor in human cells
- 2023: TRPS1
- 2023: ZNF143
- 2024+: SPDEF, GATA3…

**cannot rapidly induce or inhibit; often essential**

# A history of rapidly inducible TFs we study.
# A new era and democratizing the study of TFs.

Inherent rapid inducibility

- 2010-2021: Heat Shock Factor in Drosophila, yeast, and human cells
- 2012-2023: Estrogen Receptor in human cells
- 2017-2023: Glucocorticoid Receptor in rats and mouse & human cells
- 2023: Androgen Receptor in human cells
- 2023: TRPS1
- 2023: **ZNF143**
- 2024+: SPDEF, GATA3, *

**cannot rapidly induce or inhibit;**
**essential;**
**implicated in chromatin looping;**
**regulates spliceosomal U RNAs;**
**Our preliminary analyses suggested a role in Estrogen signaling**

# Functional perturbation experiments:
## Limitations of using RNAi depletion and knock outs to study molecular mechanisms

- Chronic depletion can result in compensatory feedback.

- Secondary (and post secondary) effects can dominate after extended depletion.

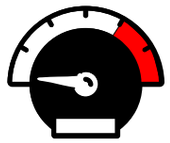
- Extended depletion can be lethal.

- Degron tagging is an attractive alternative

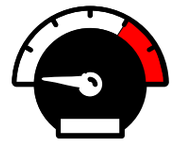
# What does a car's radiator do?



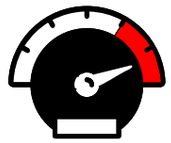
Speed



Tire Pressure

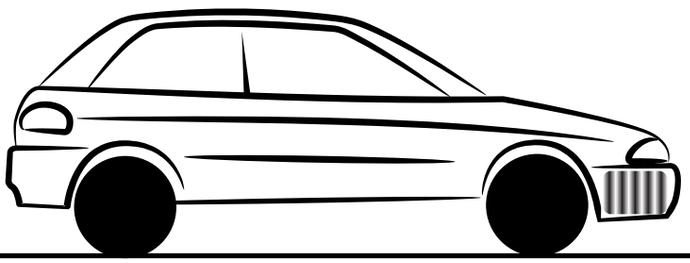


Temperature

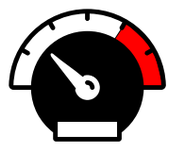


Prevents engine overheating (3 days to fully deplete fluid)

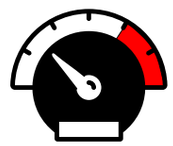
Prevents engine overheating (3 days to fully deplete fluid)



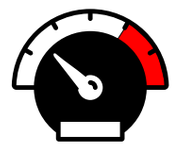
Speed



Tire Pressure



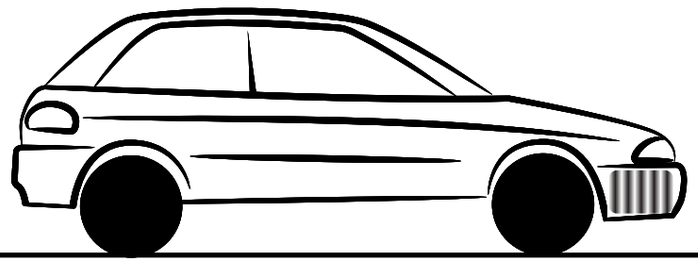
Temperature



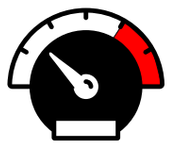
Control

The radiator keeps the car healthy;
The radiator affects the starter, speed, tire pressure, and temperature

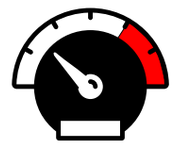
# What does a car’s radiator do?



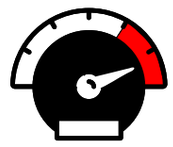
**Speed**



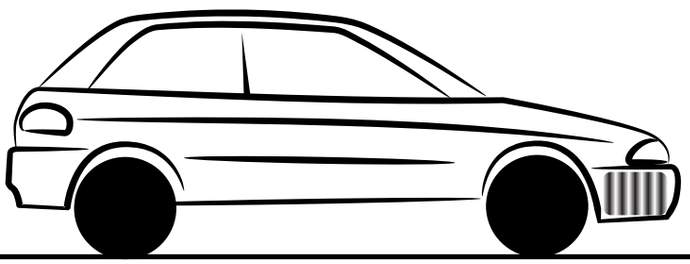
**Tire Pressure**



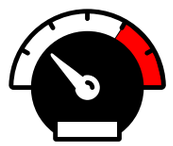
**Temperature**



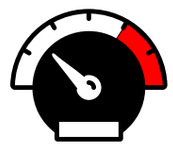
Monitor the engine and vehicle systems continually



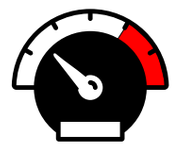
**Speed**



**Tire Pressure**



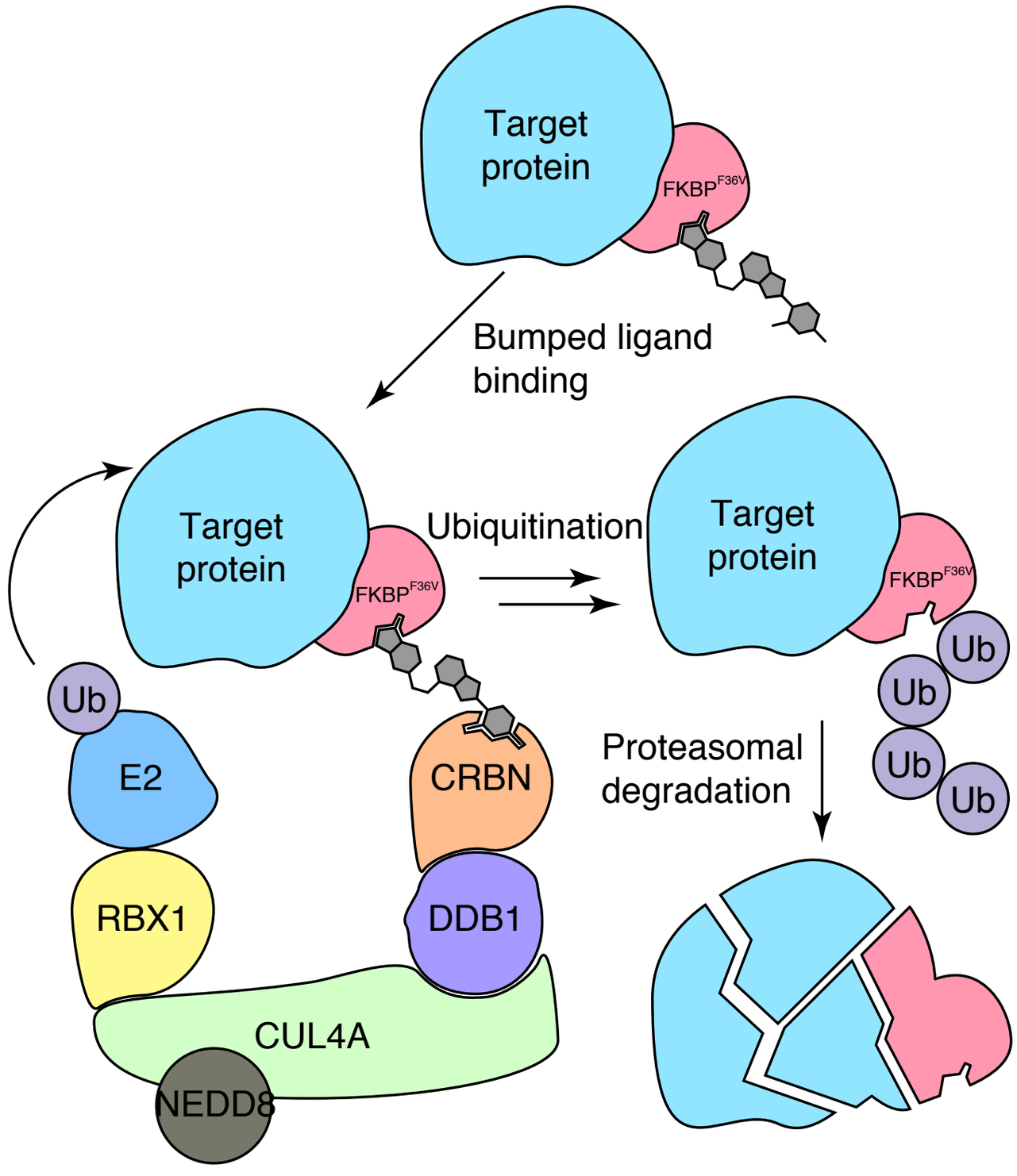
**Temperature**



Control

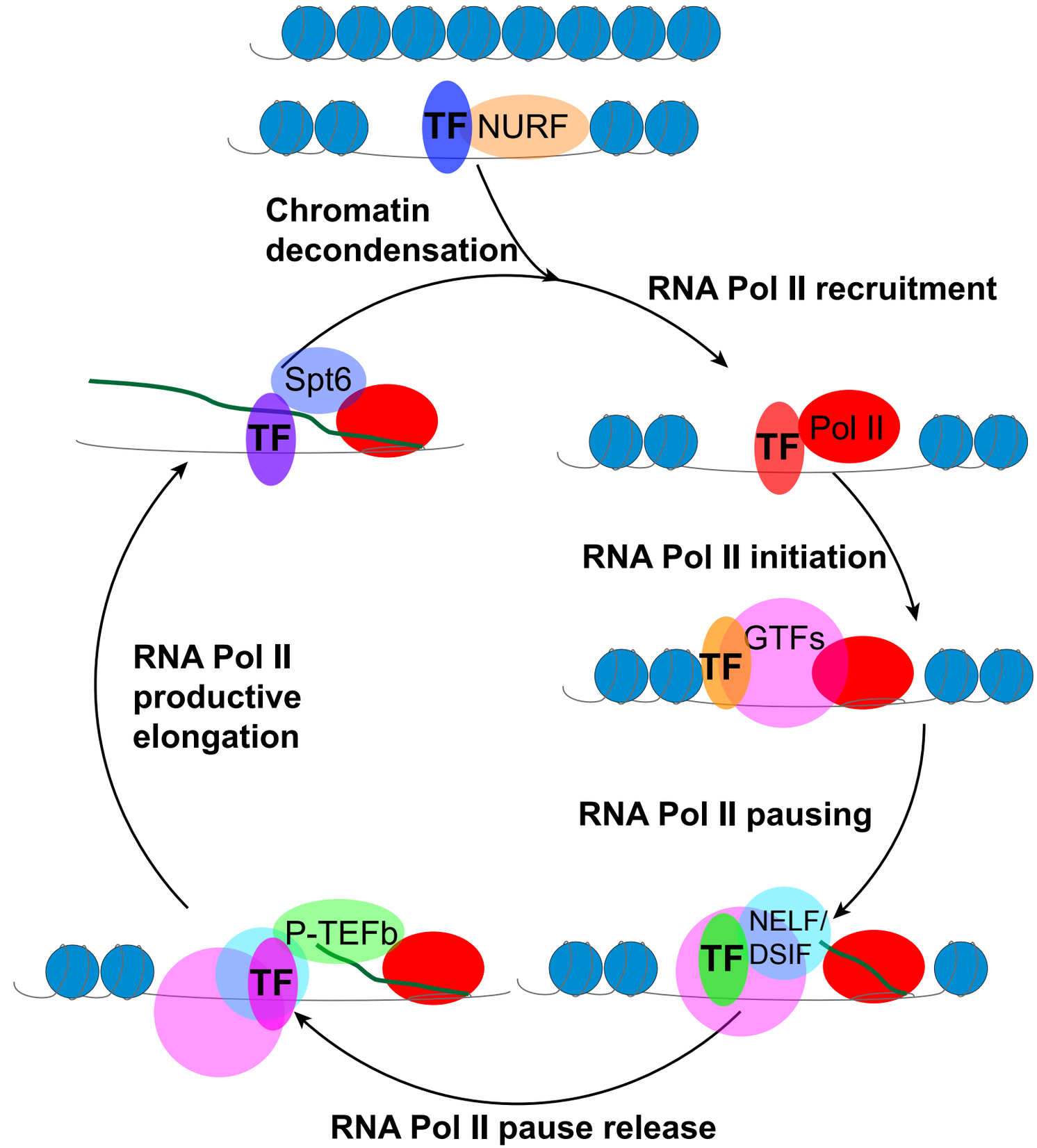
The radiator directly regulates temperature;
all other effects are indirect effects of the car catching fire.

# dTAG: rapid and specific protein degradation

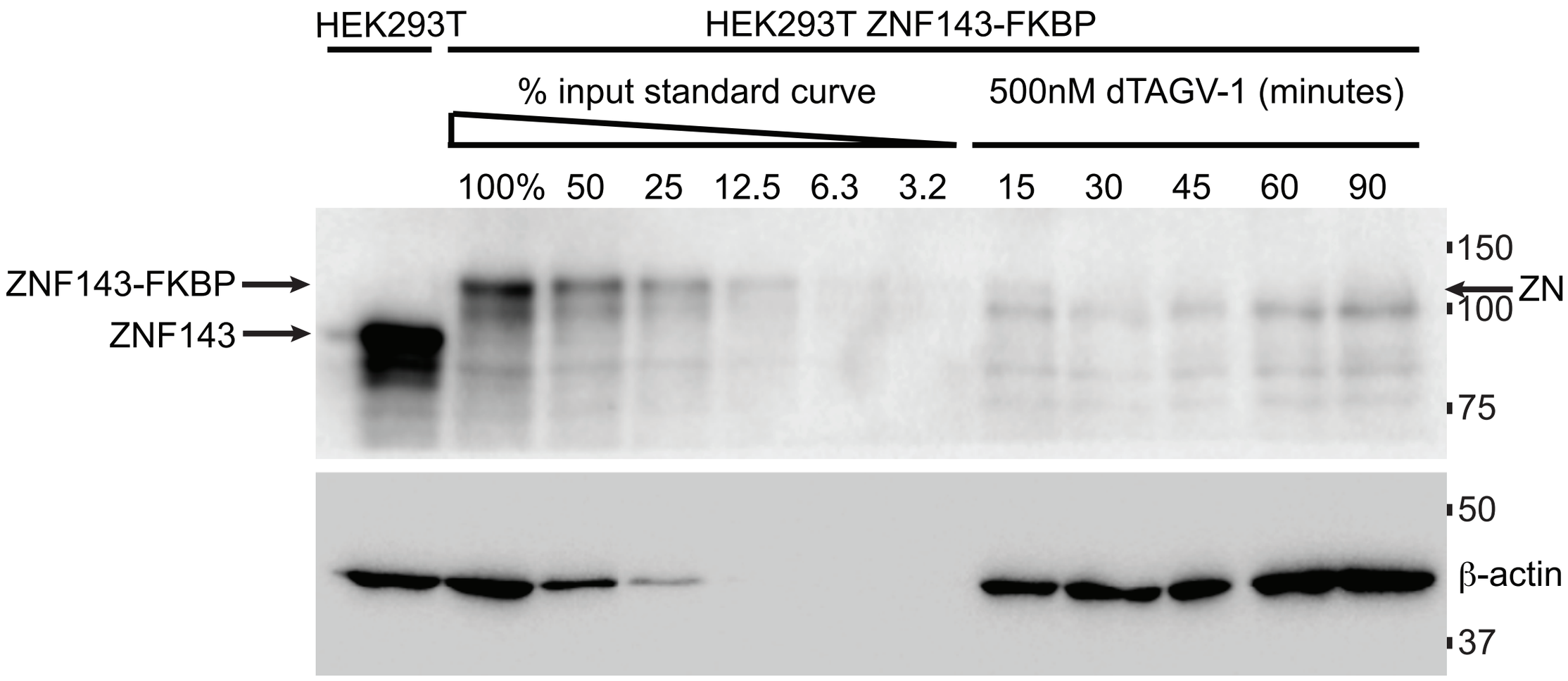



Nabet, et. al., Nat Chem Biol. 2018

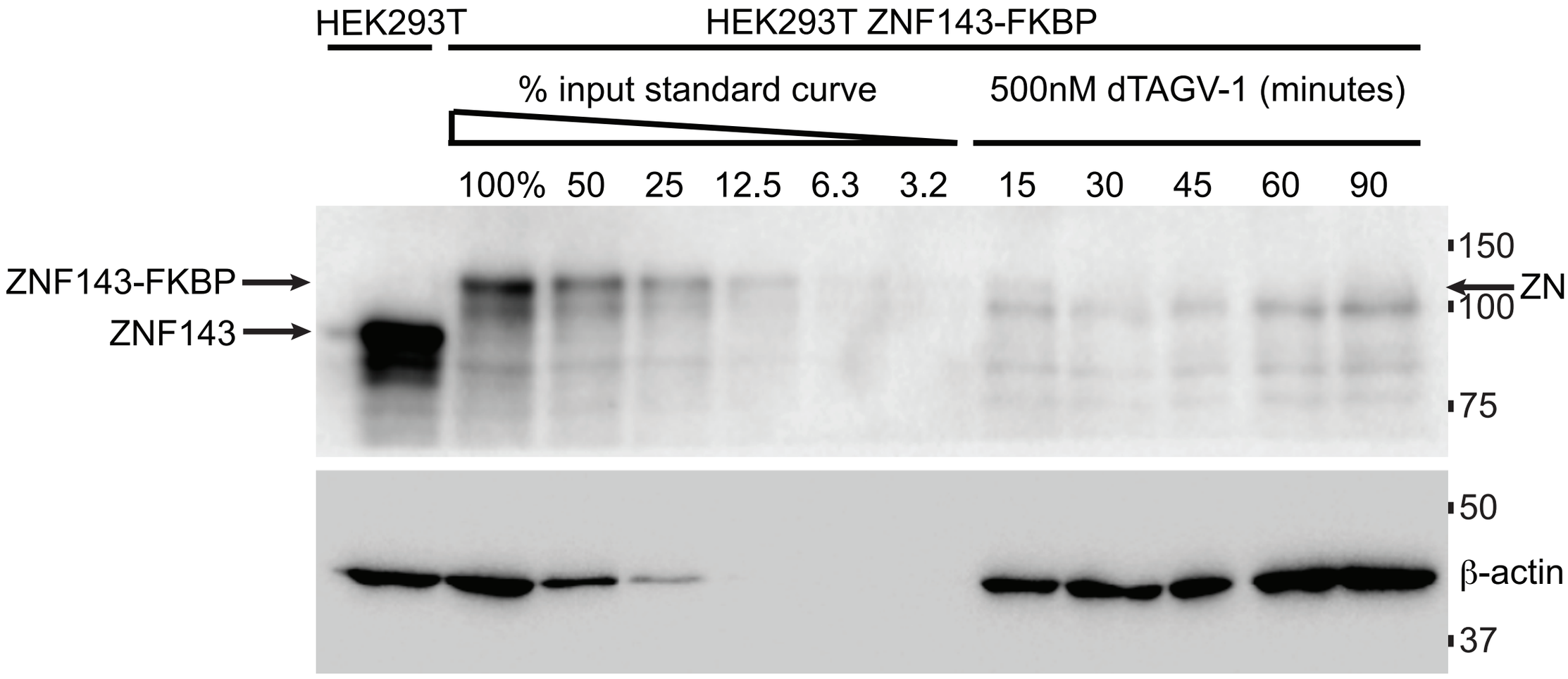
# What is/are the molecular function(s) of ZNF143?



# Degron tagged ZNF143 is degraded at 30 minutes

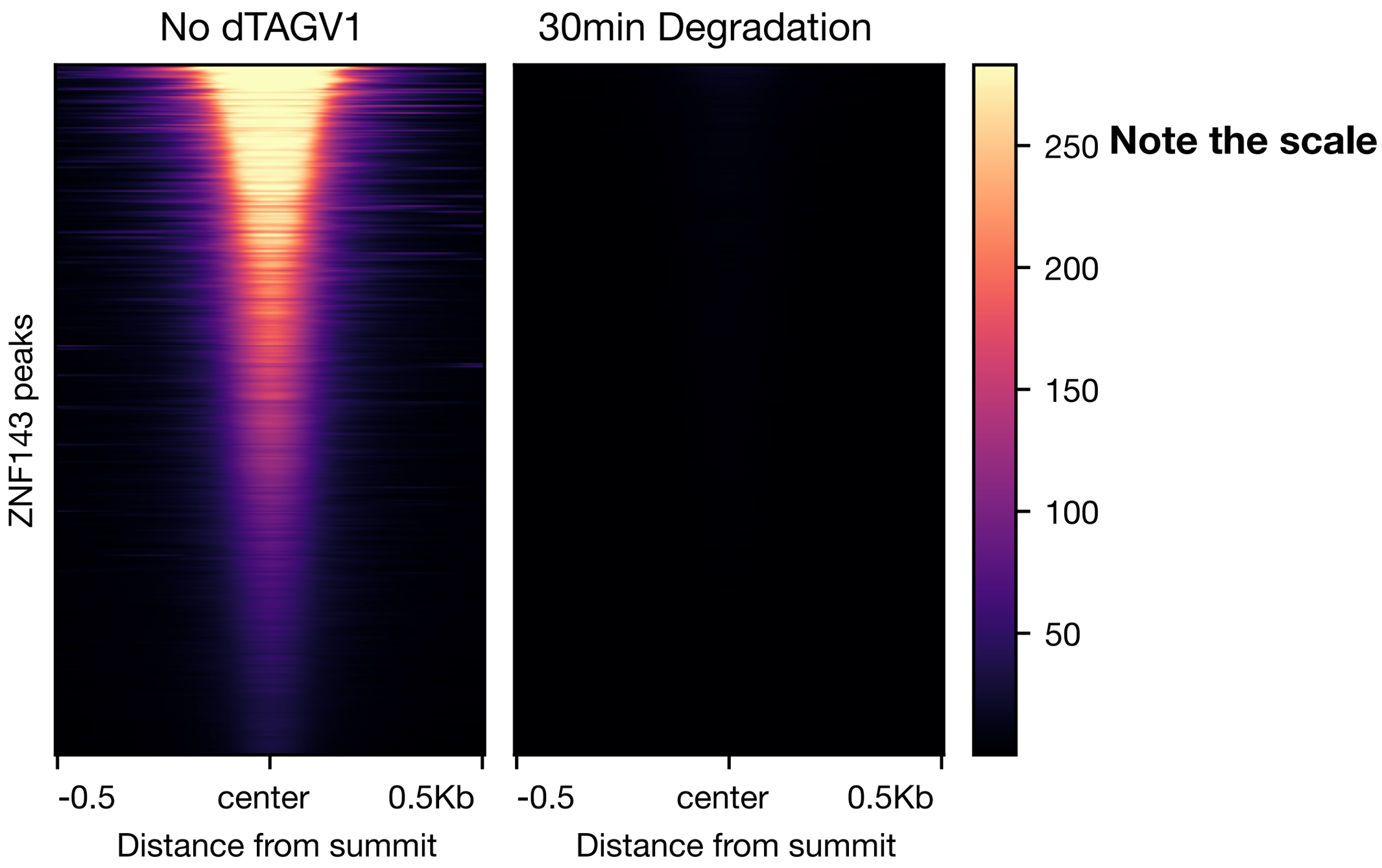


# Degron tagged ZNF143 is degraded at 30 minutes

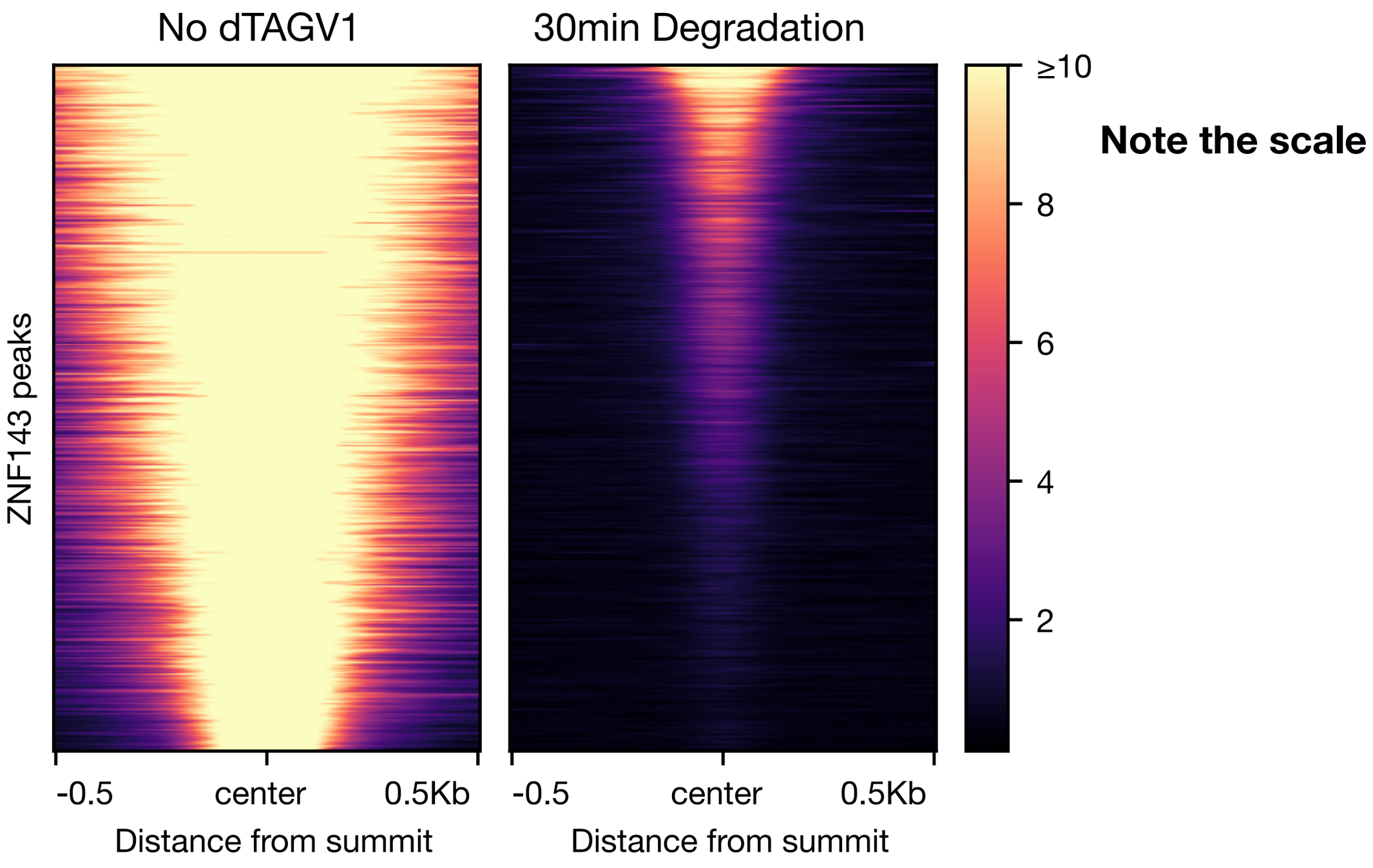



****All degradation data in this presentation are after 30min****

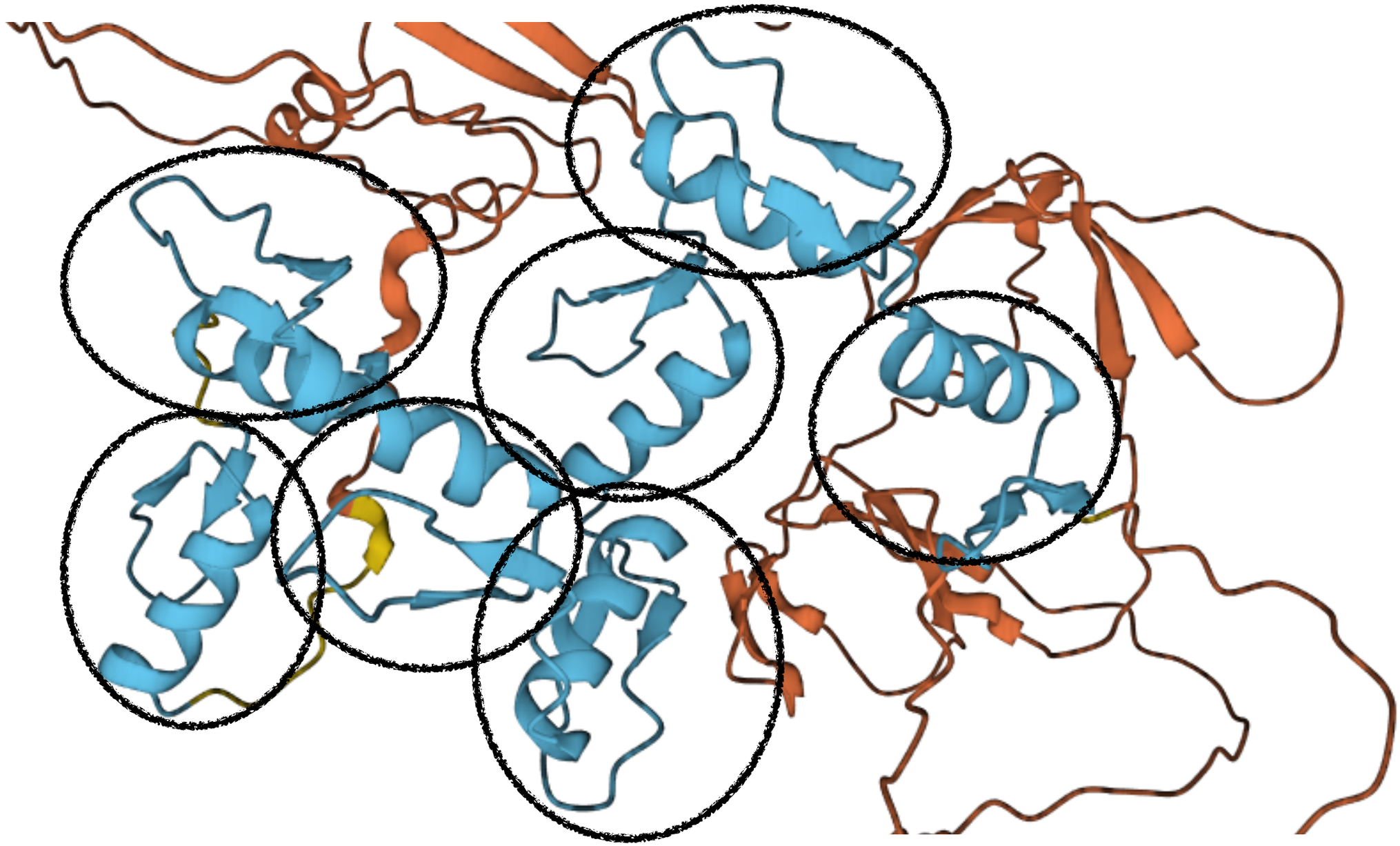
# ZNF143 is off chromatin at 30 minutes



# ZNF143 is off chromatin at 30 minutes

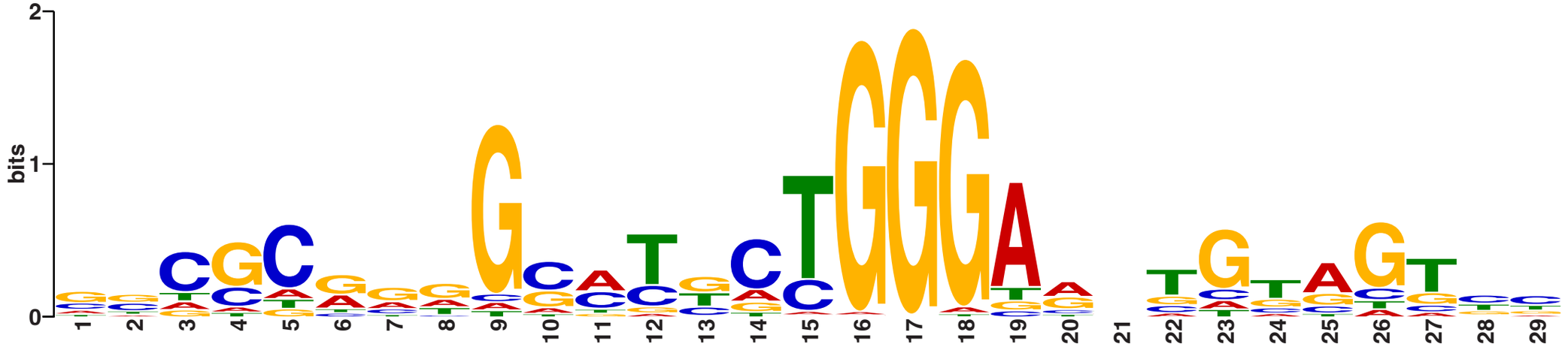


# ZNF143 has 7 blue Zn Fingers



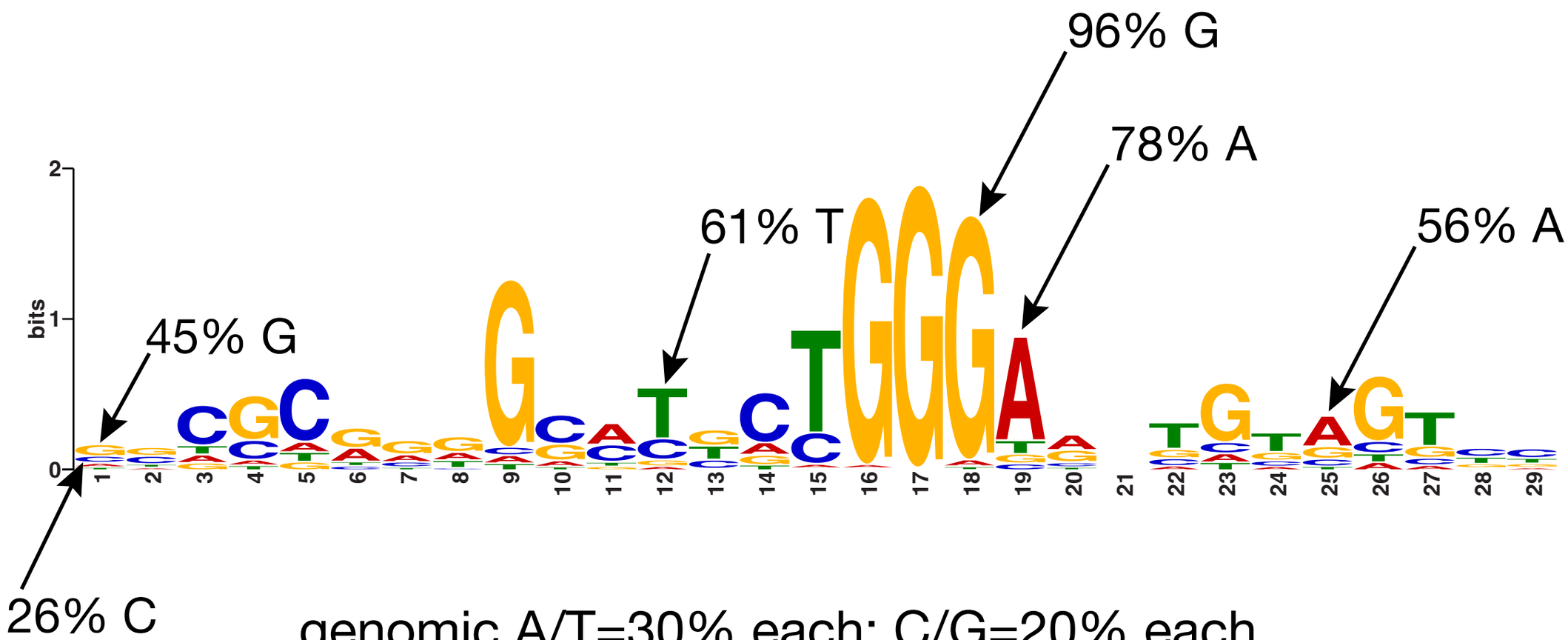
AlphaFold

# ZNF143’s 7 Zn fingers span a 29 base degenerate motif




# All ZNF143 ChIP peaks have a ZNF143 motif

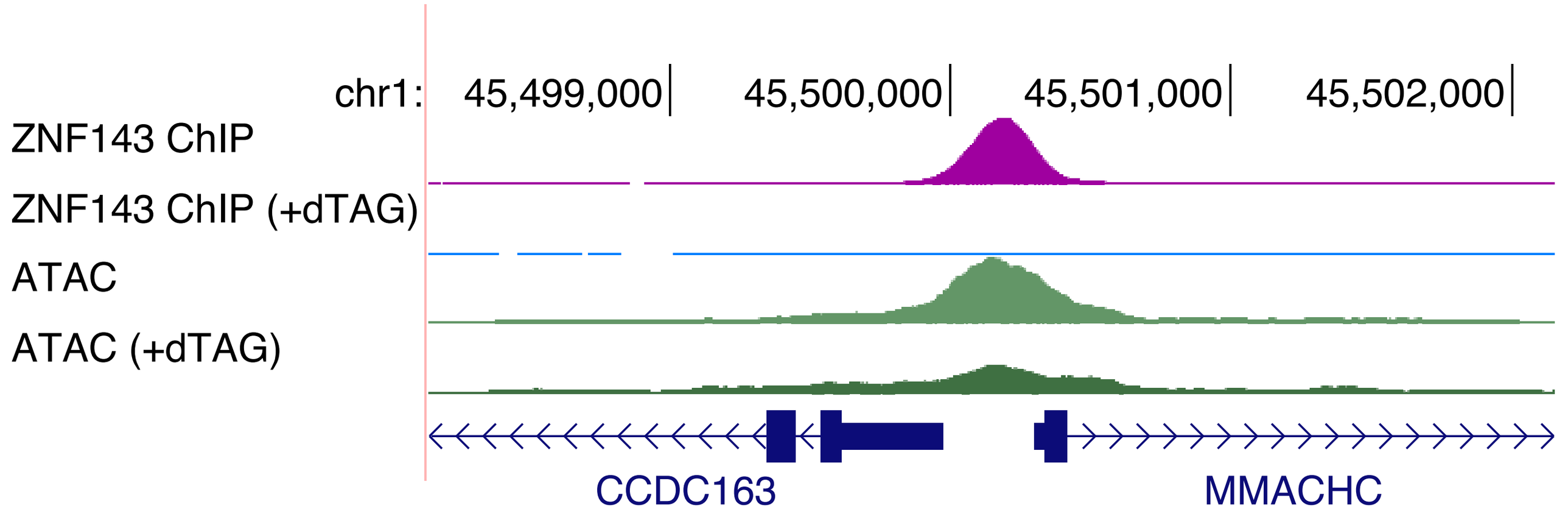
# ZNF143's 7 Zn fingers span a 29 base degenerate motif



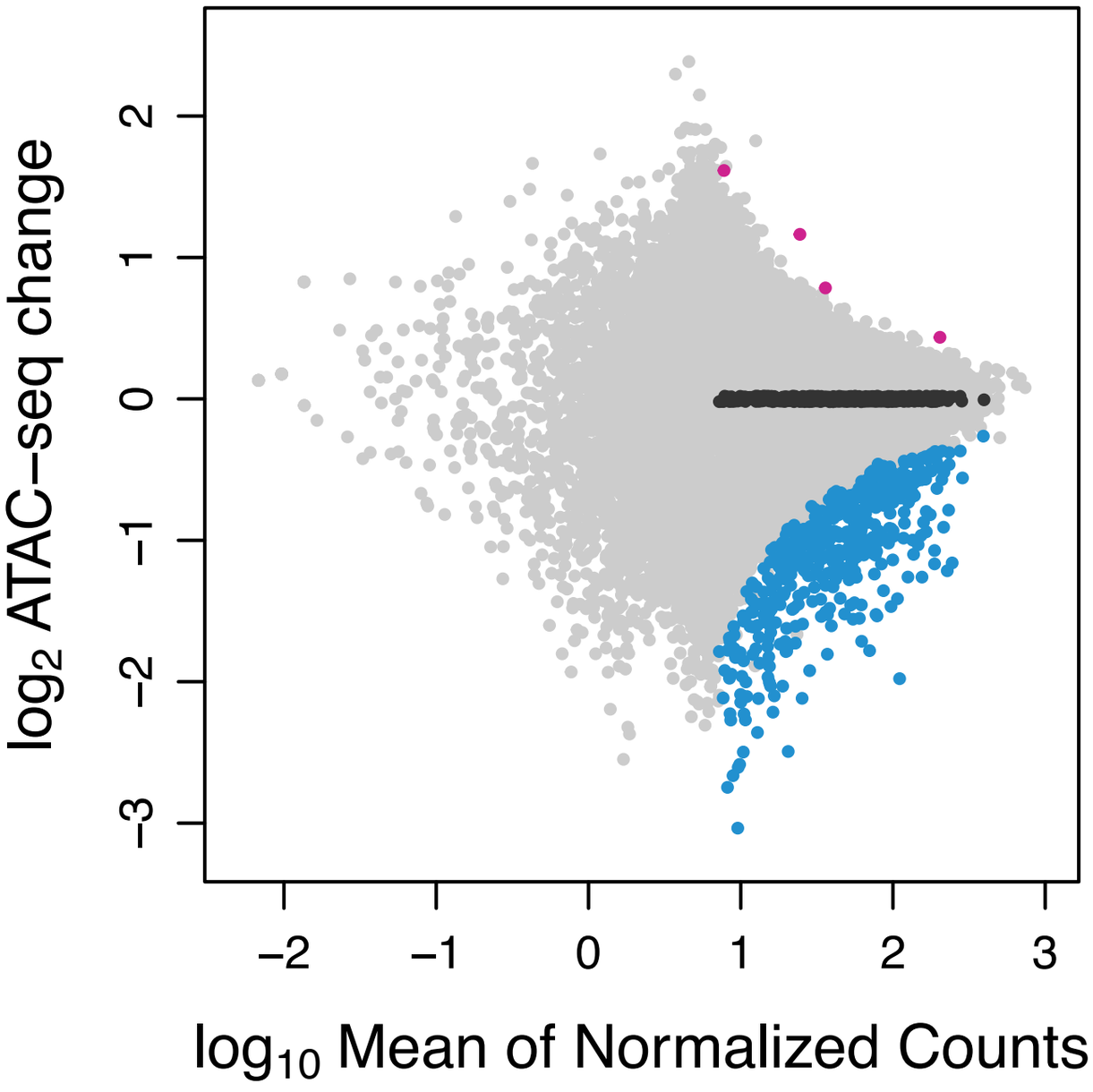

genomic A/T=30% each; C/G=20% each

# All ZNF143 ChIP peaks have a ZNF143 motif

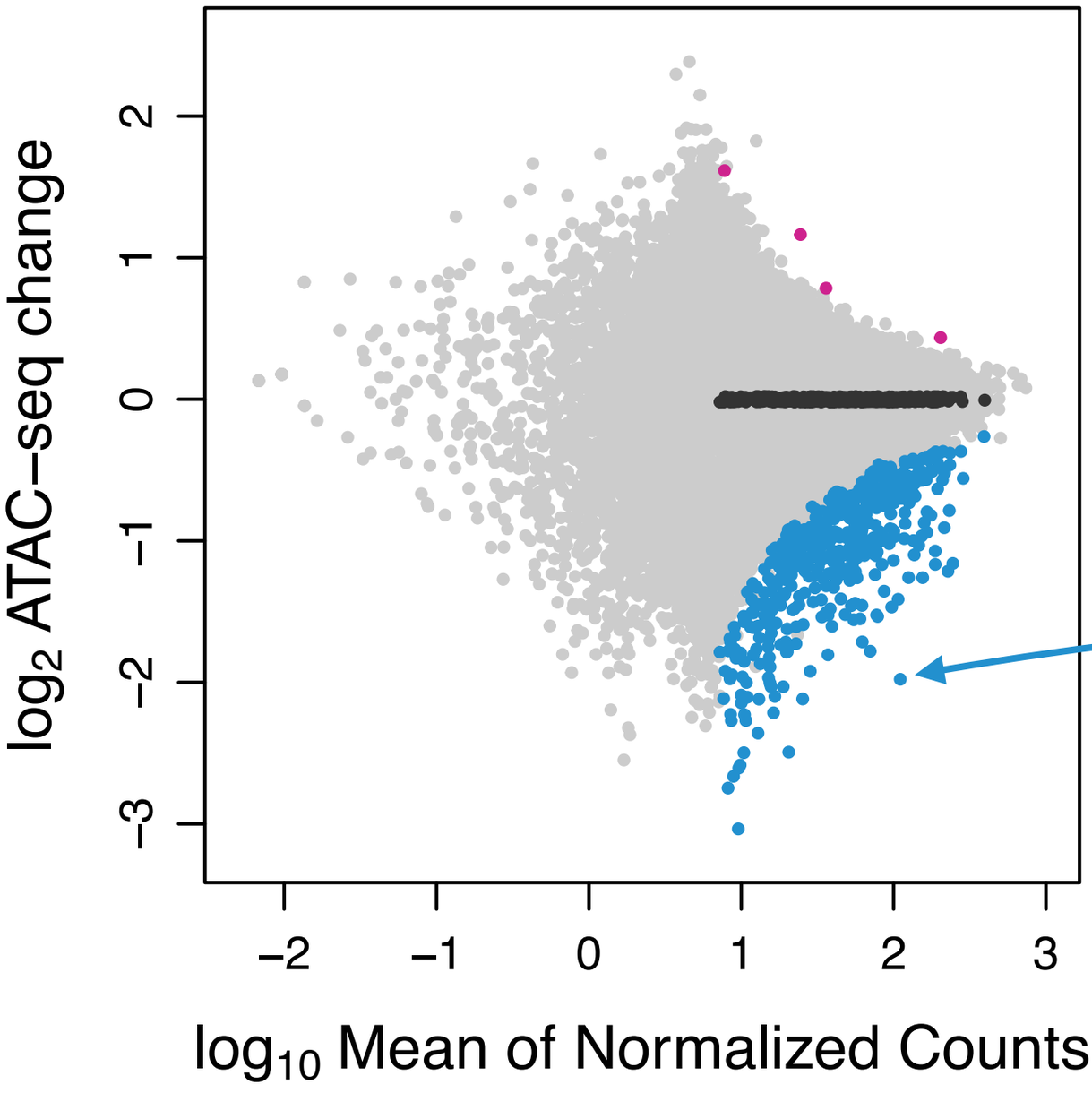
# Use ATAC-seq to quantify changes in chromatin

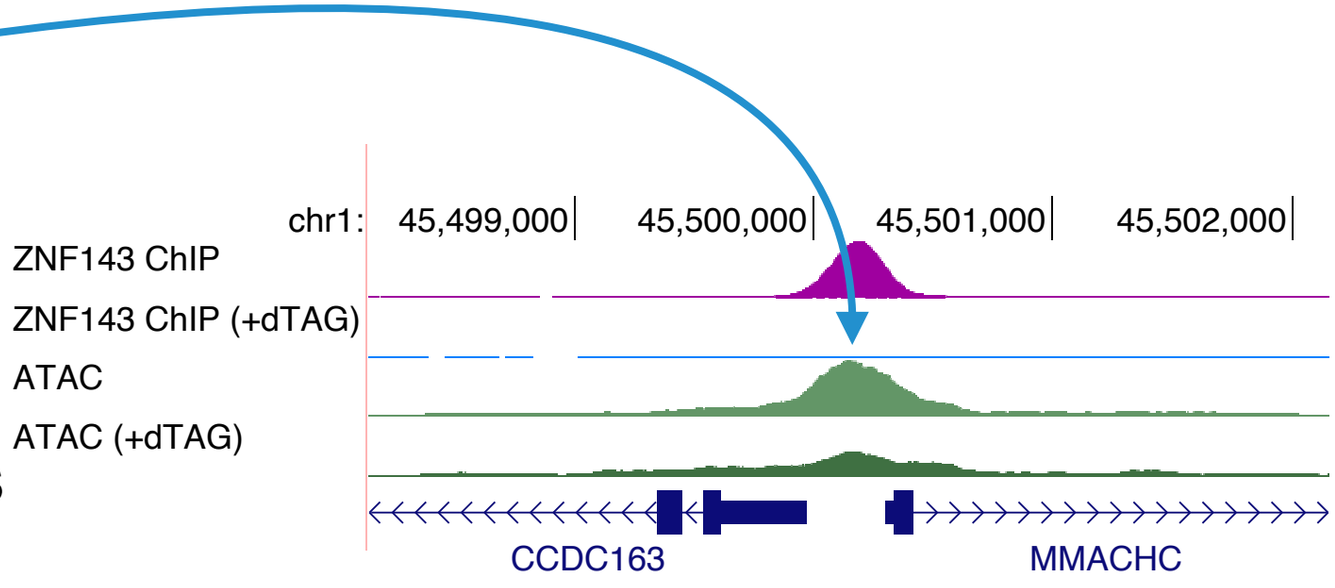


# 30 min of ZNF143 depletion causes decreased chromatin accessibility

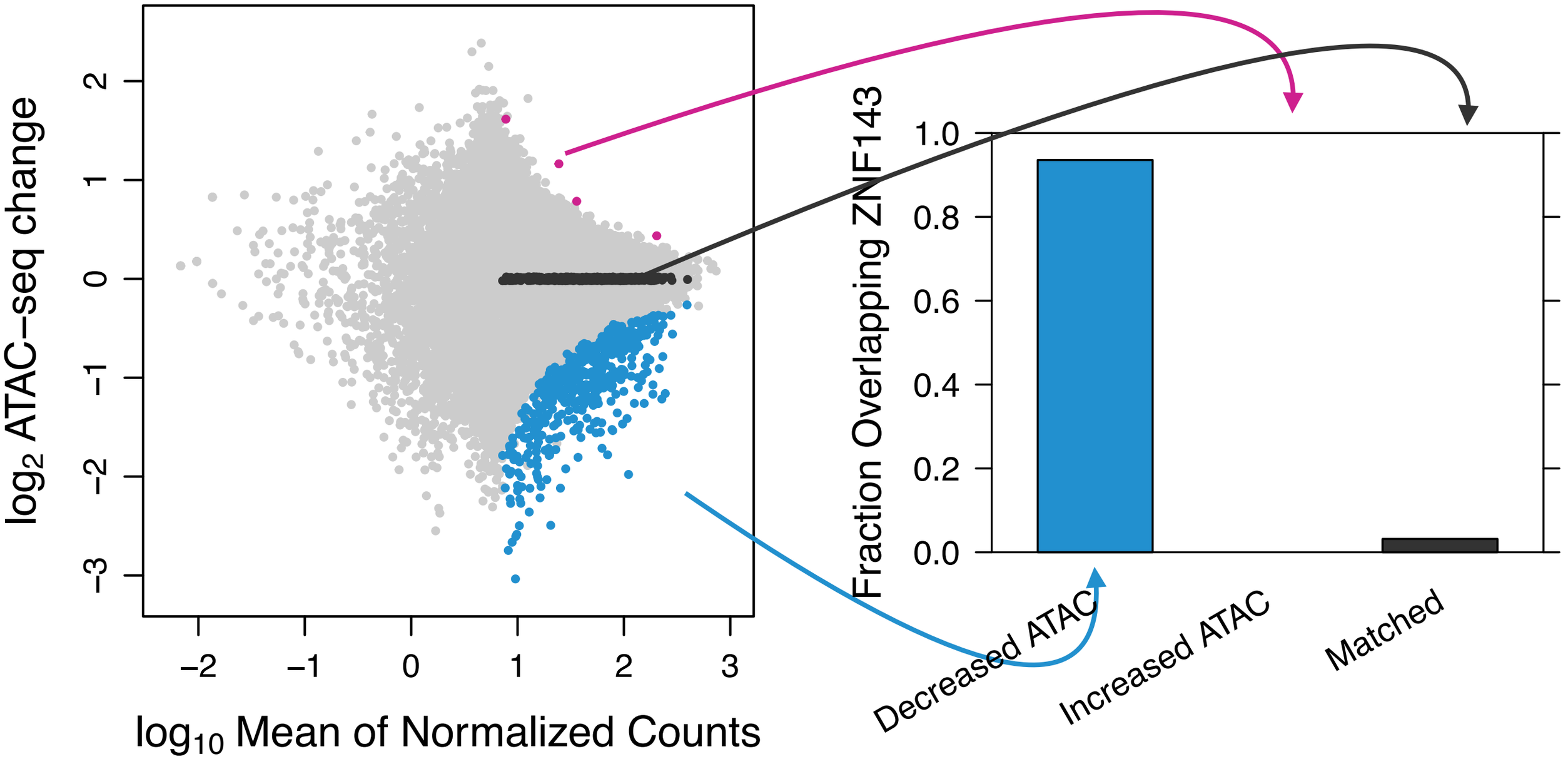


# ZNF143 maintains open chromatin structure

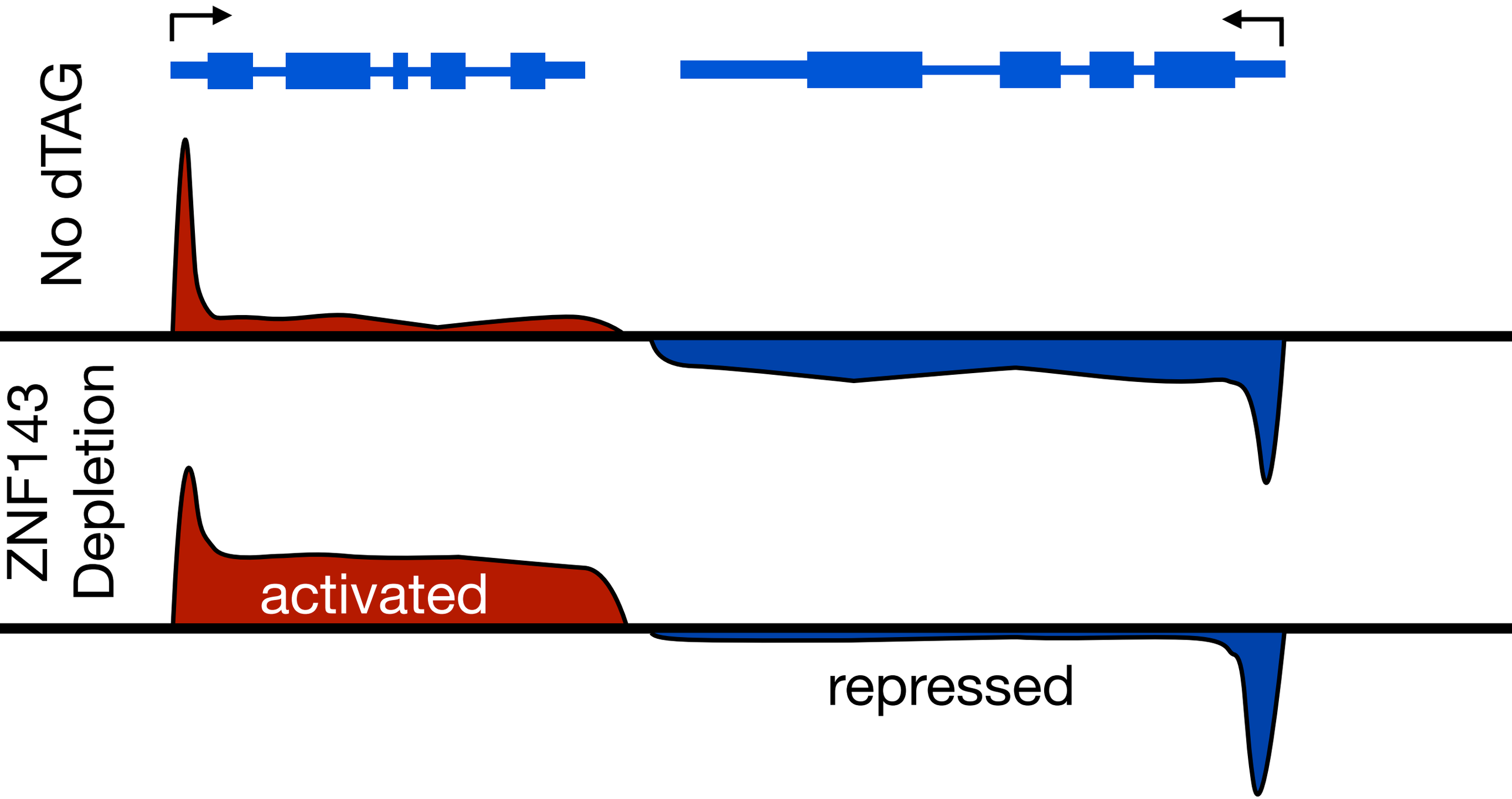


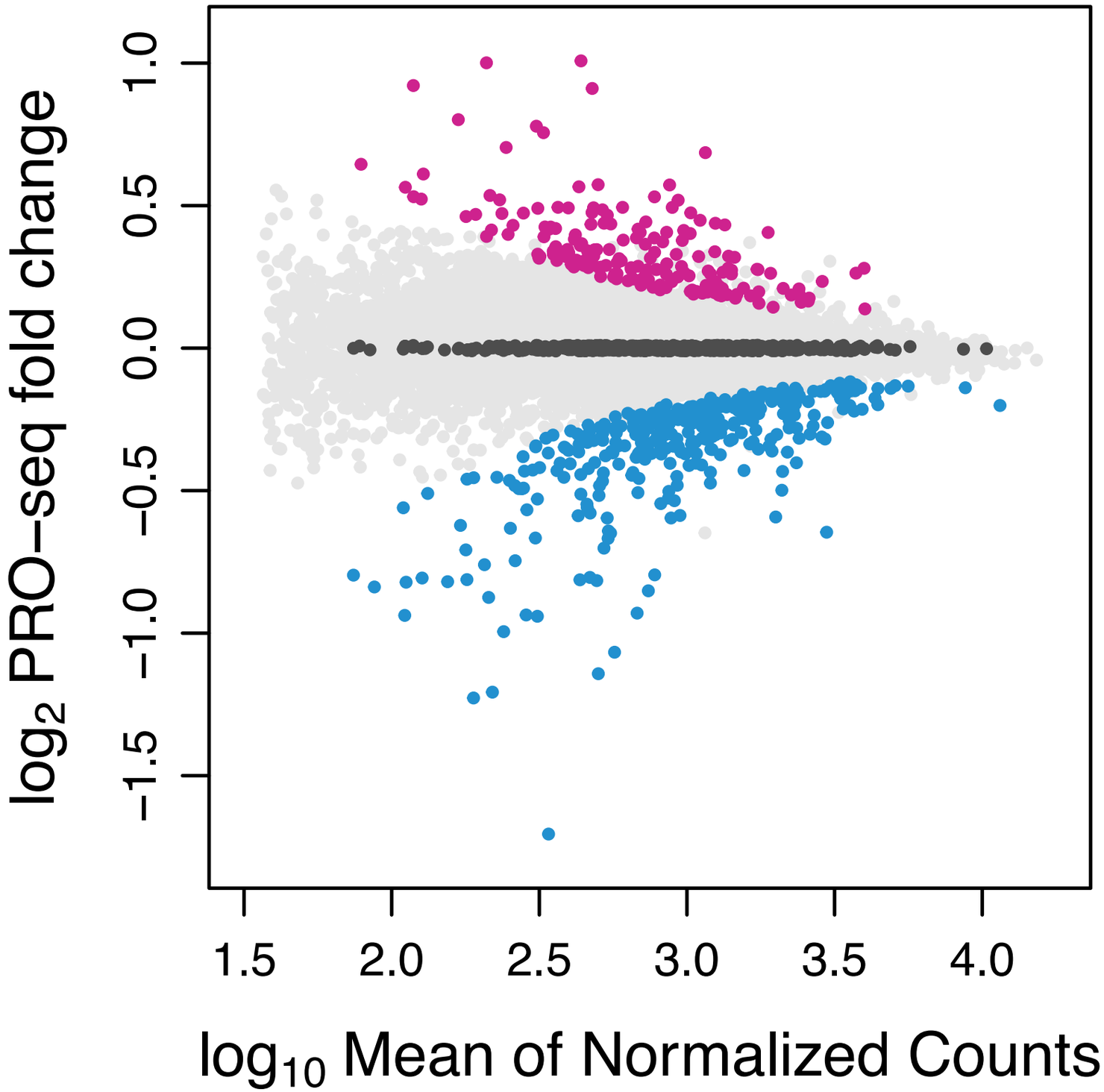
# 94% of decreased ATAC peaks overlap ZNF143; ZNF143 degradation leads to decreased accessibility



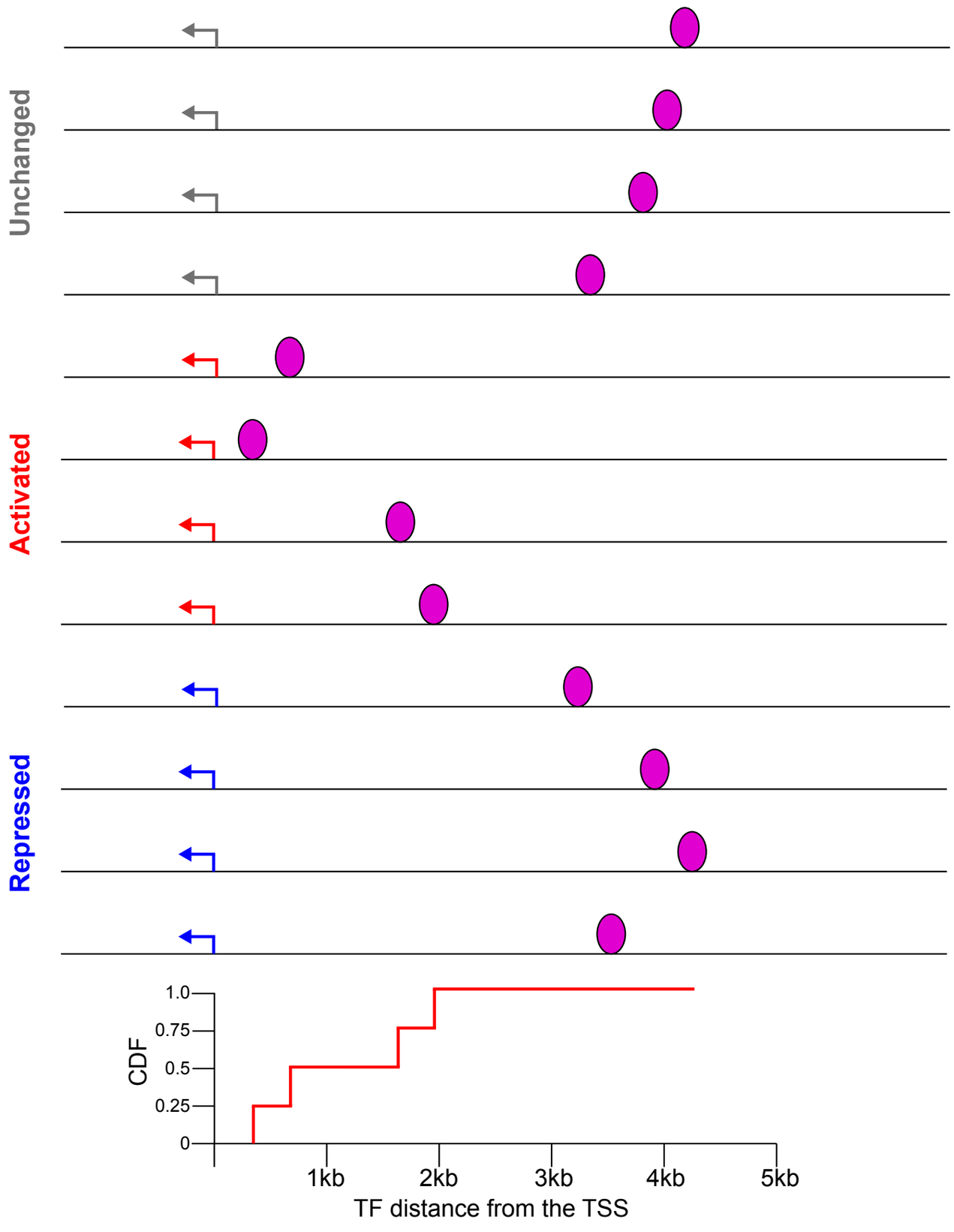
# Use PRO-seq to quantify changes in gene expression



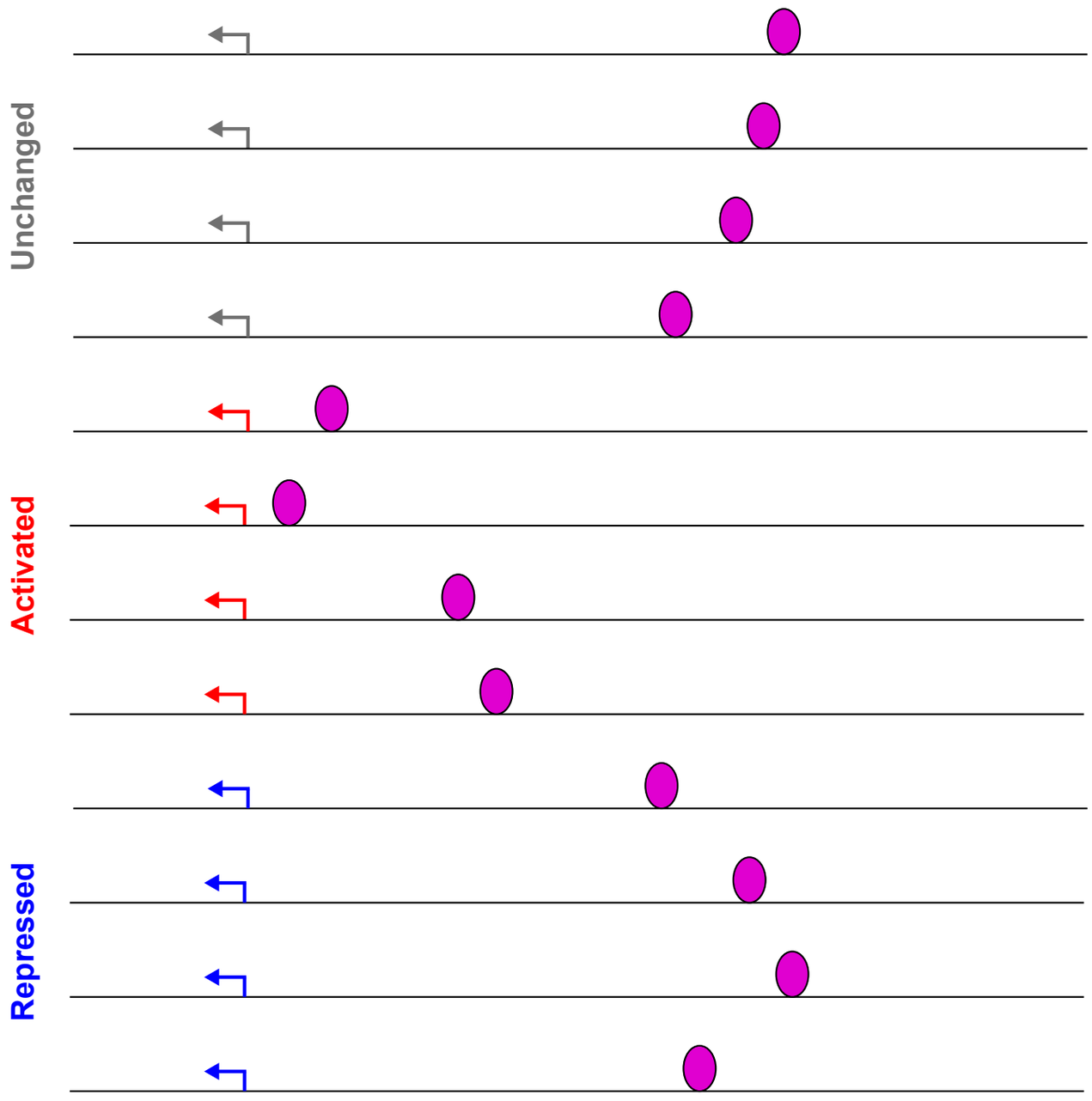
# Gene expression goes up and down after ZNF143 degradation



# Interpreting TF binding site/TSS cumulative distributions

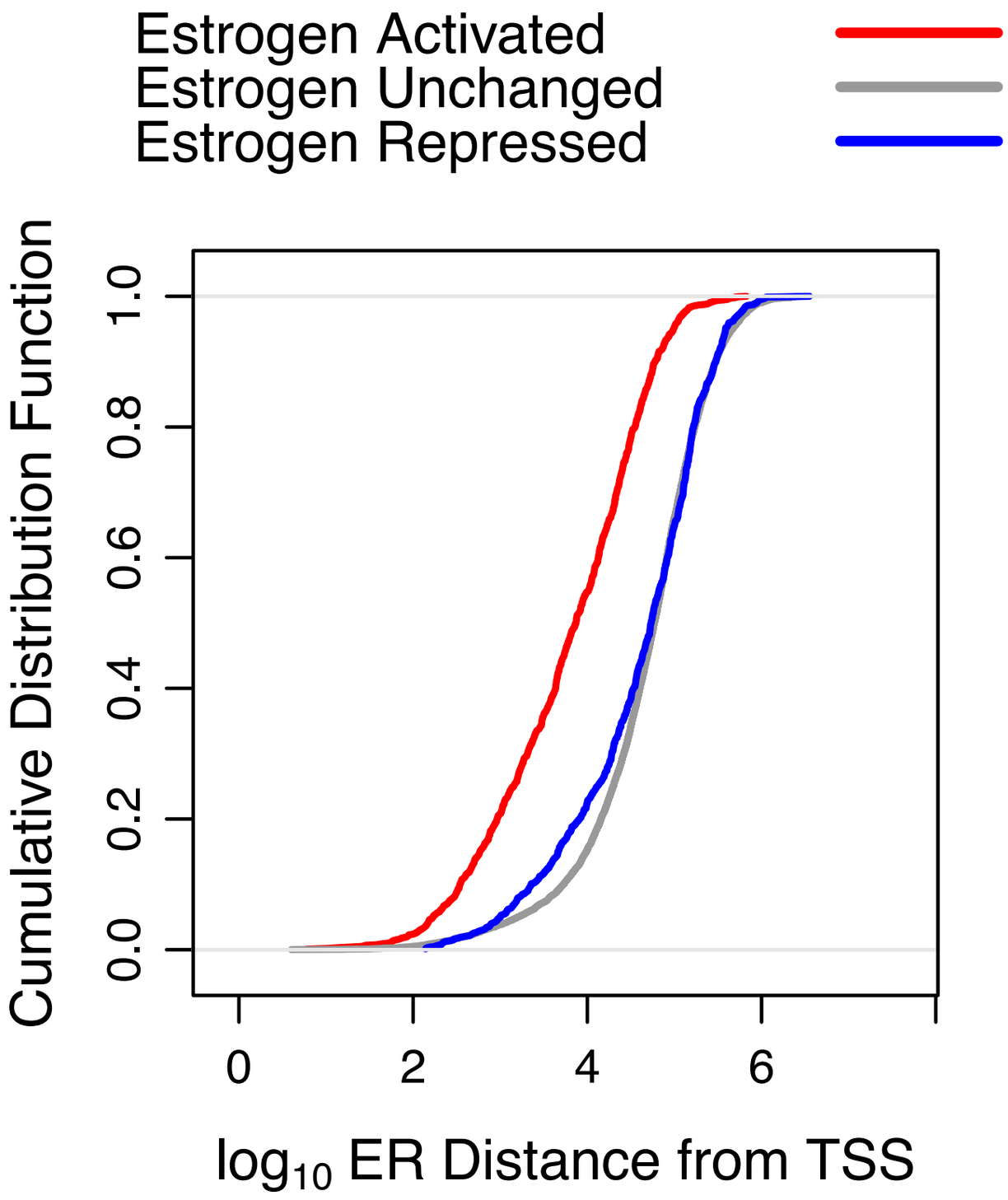


# Left shifted CDF indicates that the TF regulates that gene set

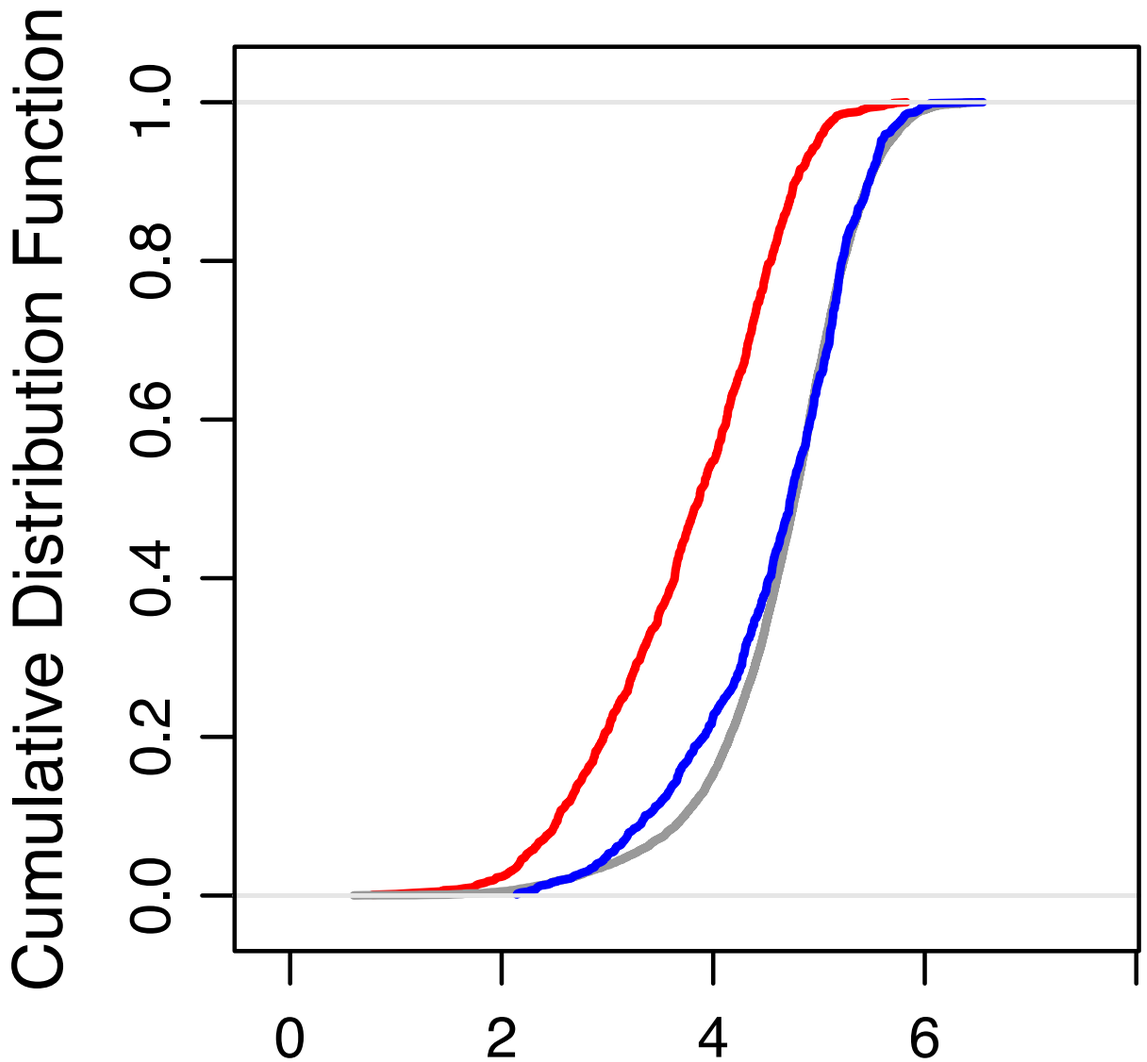



Left-shift indicates that the activated genes tend to be closer to pink TF binding sites.

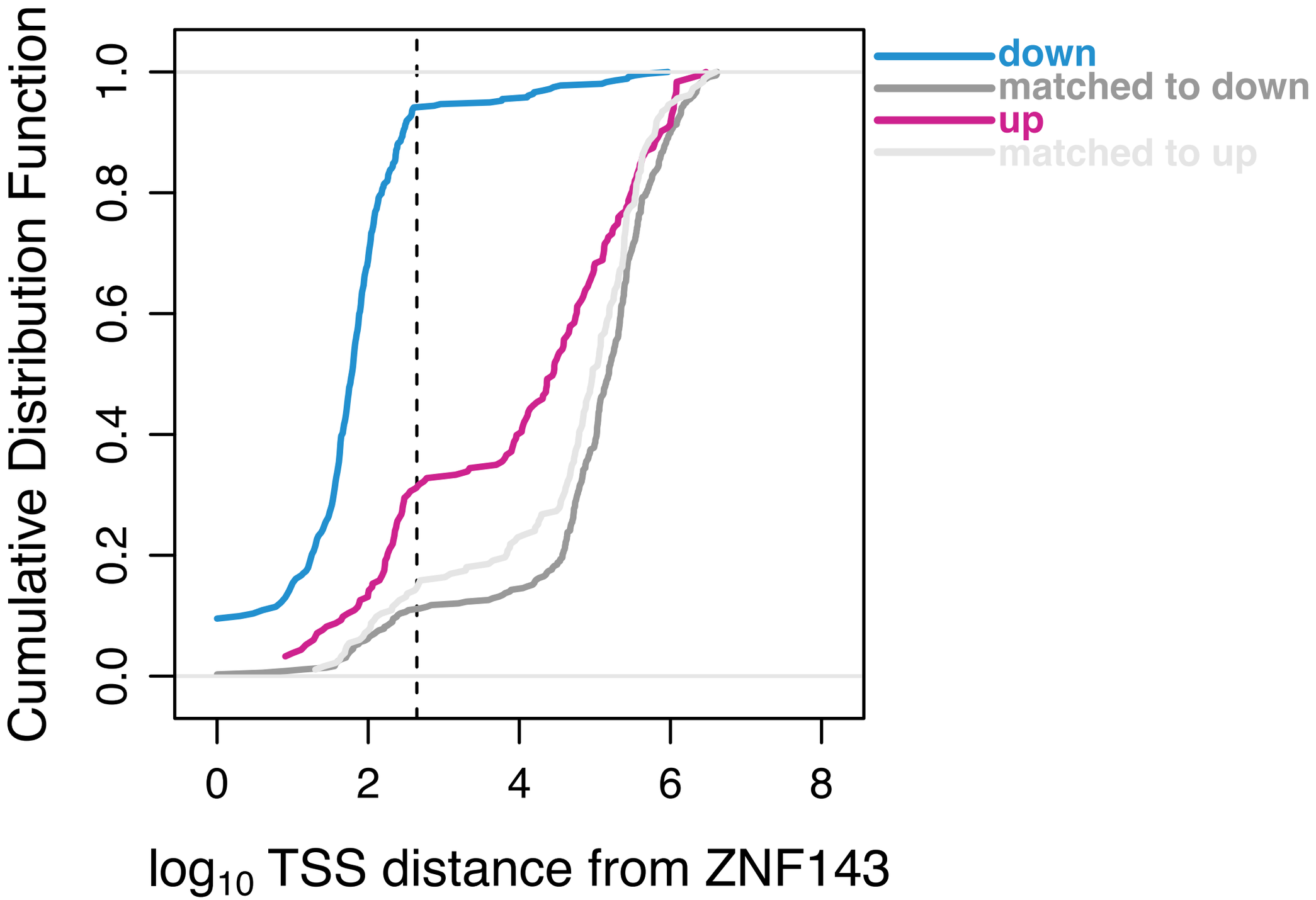
# Immediate estrogen activated genes are close to inducible Estrogen Receptor binding; ER is an activator



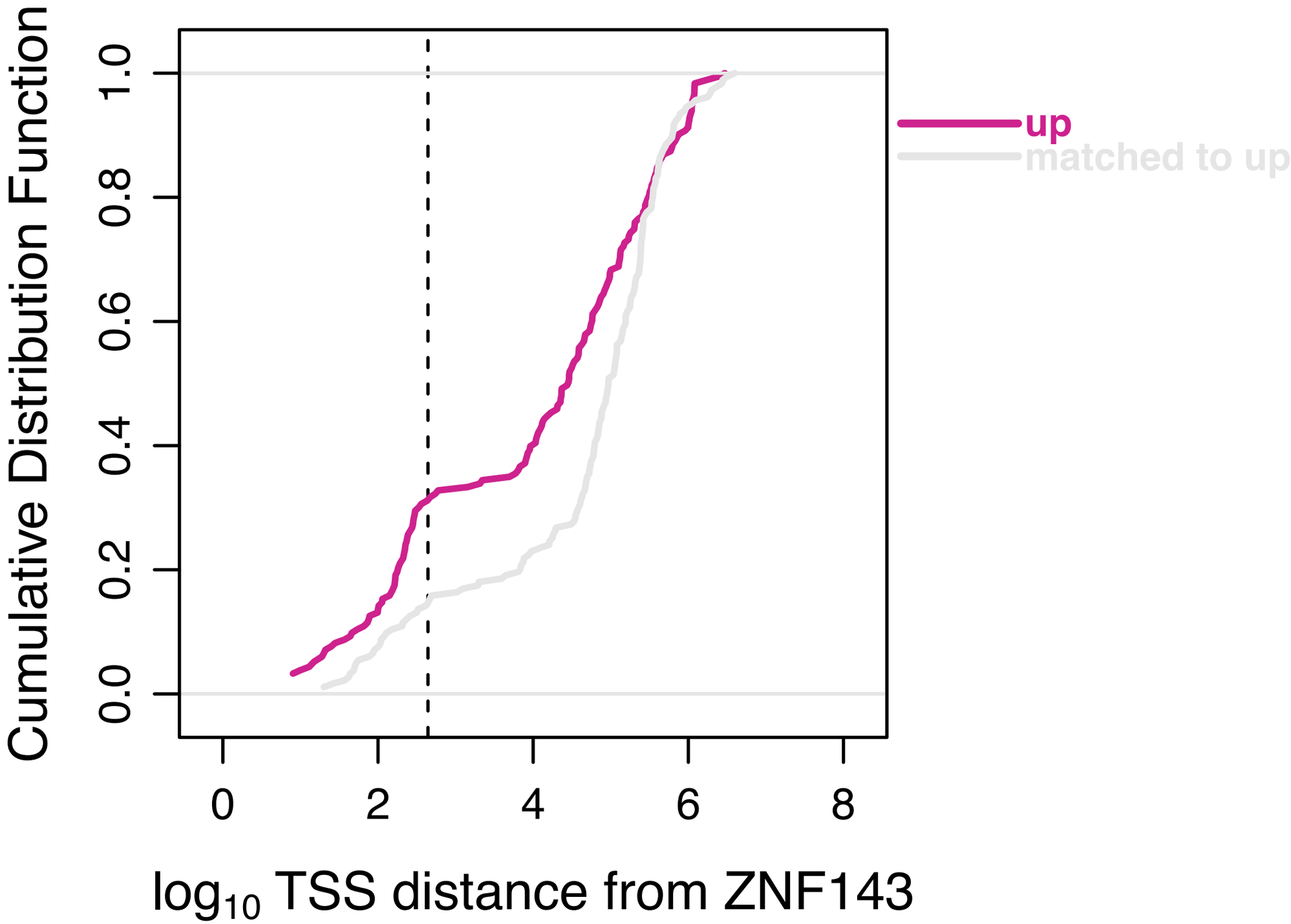
**The shift left of either activated or repressed genes**
**is what I observed/confirmed previously for HSF, ER, GR, and AR**
**(not both!)**



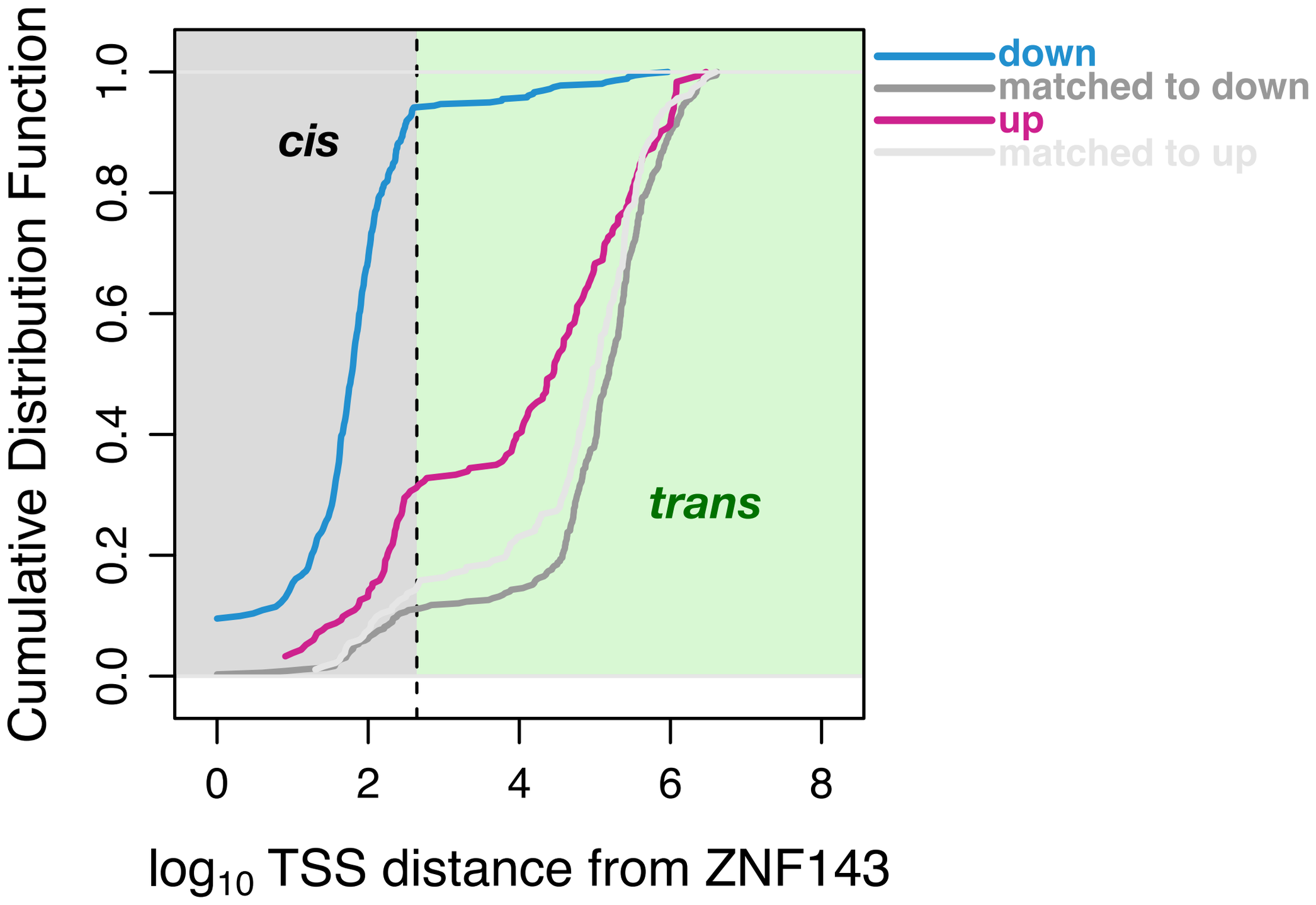
# 94% of down and 31% of up are within 440 bases of ZNF143



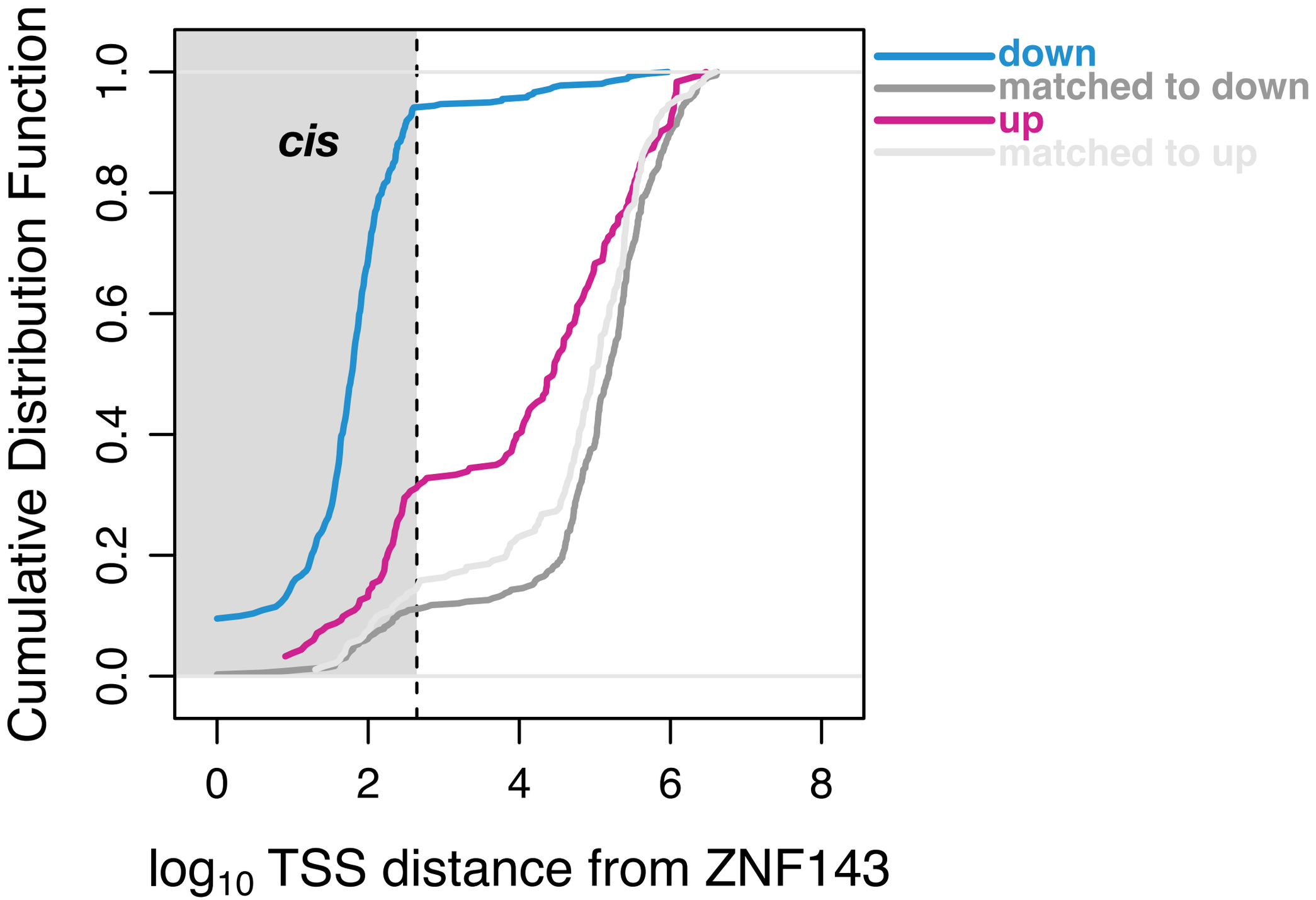
# Up genes significantly closer to ZNF143 challenged my TF world view



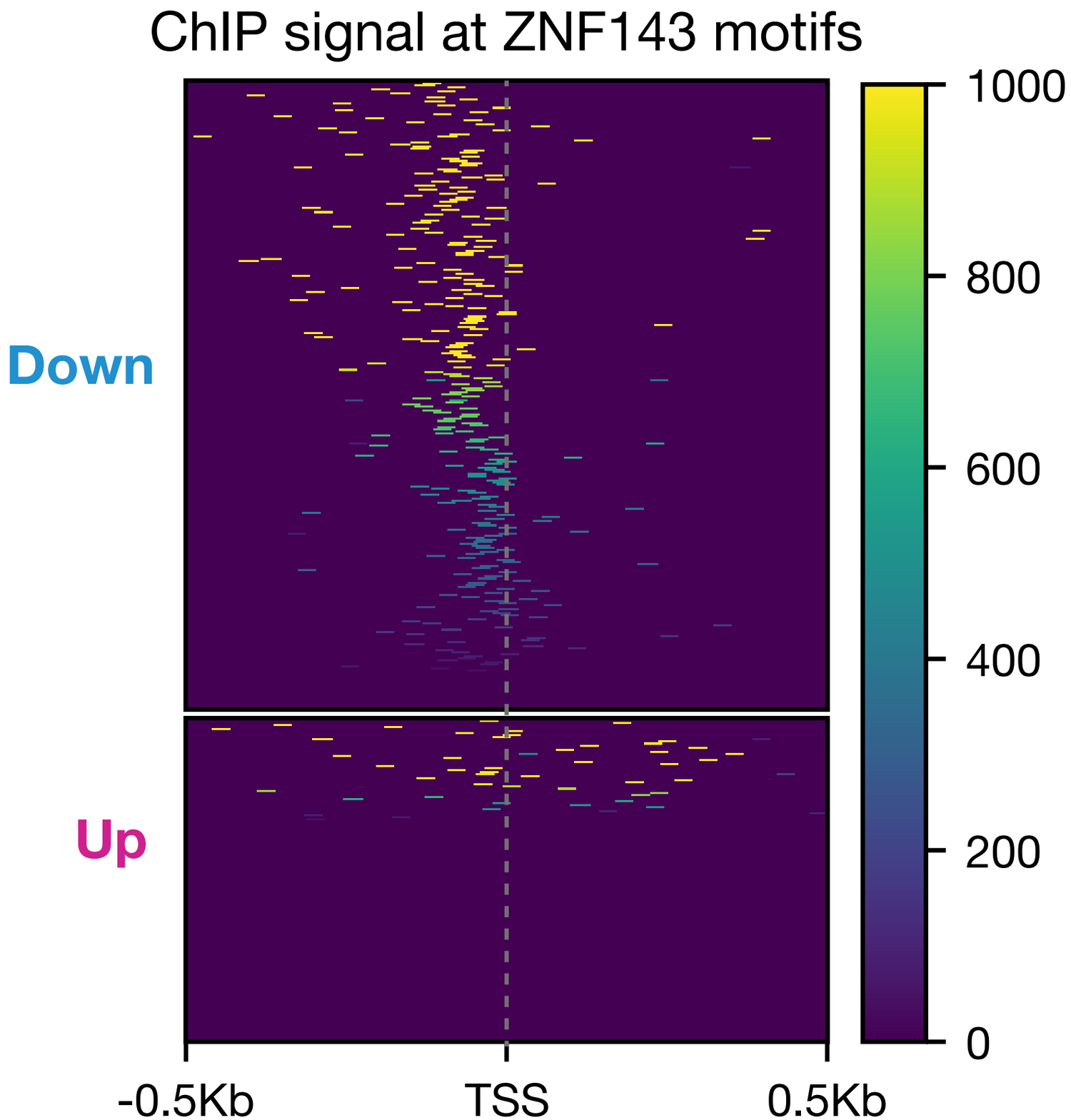
# Within 440 bases is considered ZNF143 *cis* targets



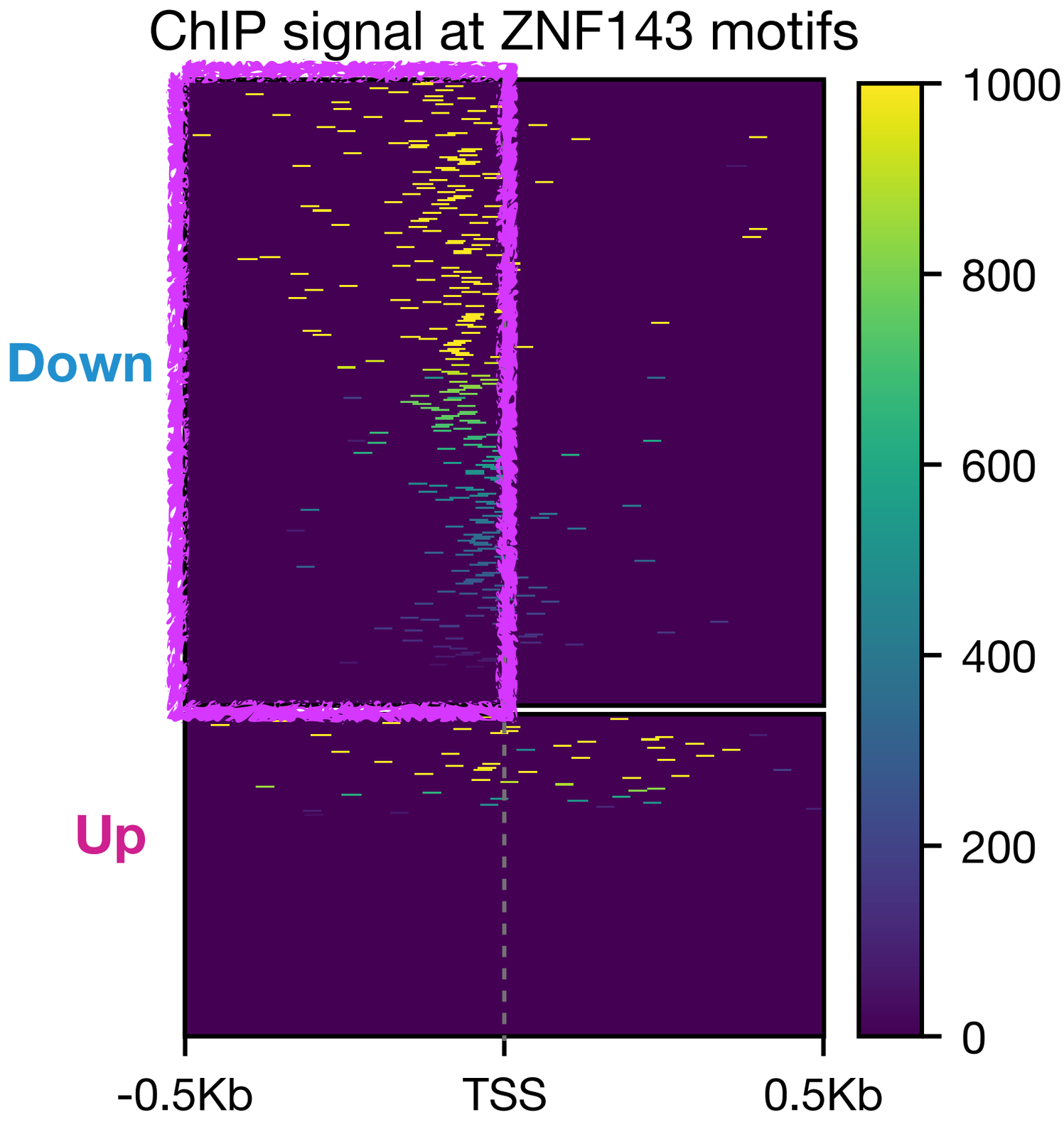
# How does ZNF143 activate gene targets?



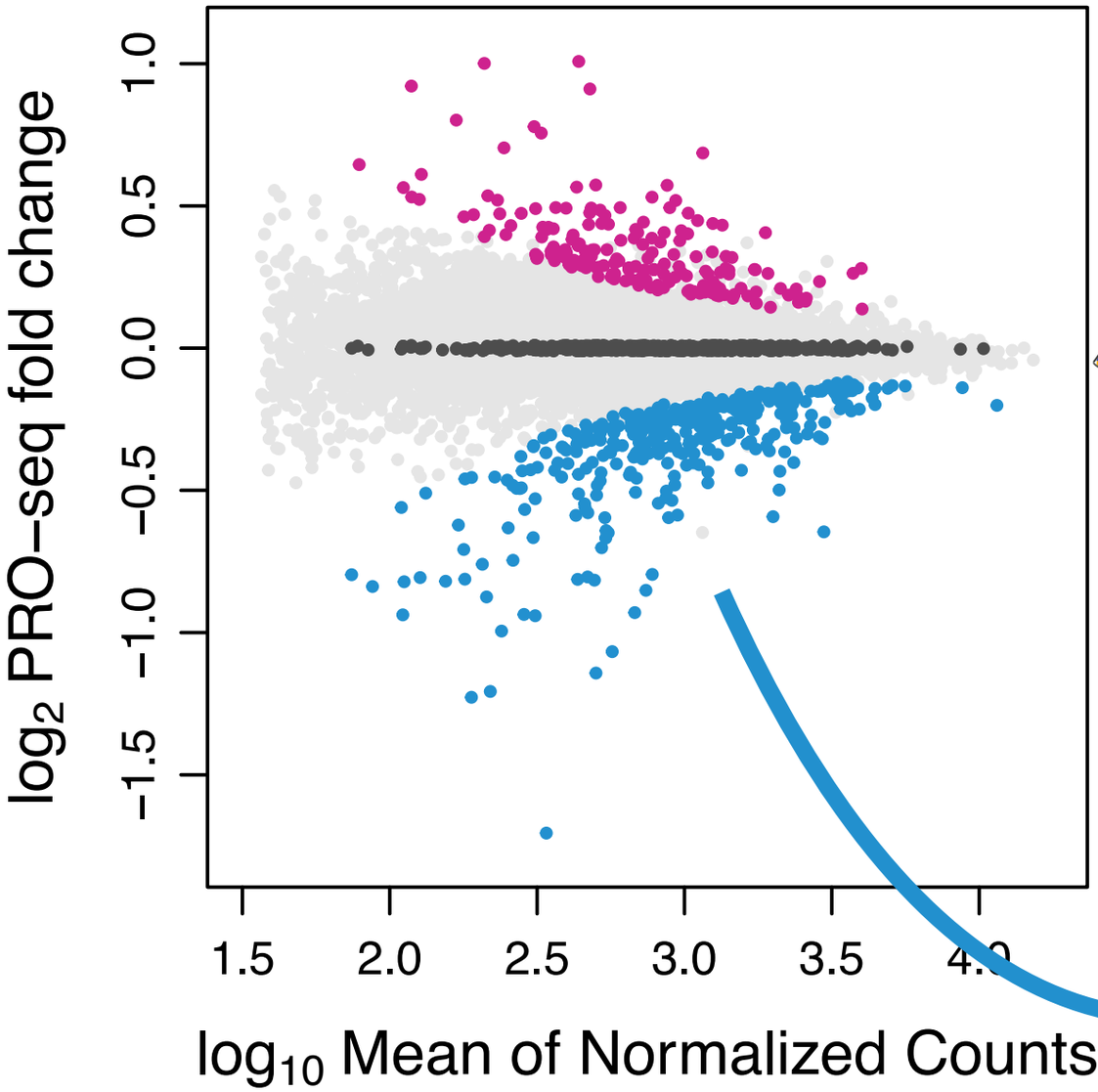
# ZNF143 binding sites distribution relative to TSSs

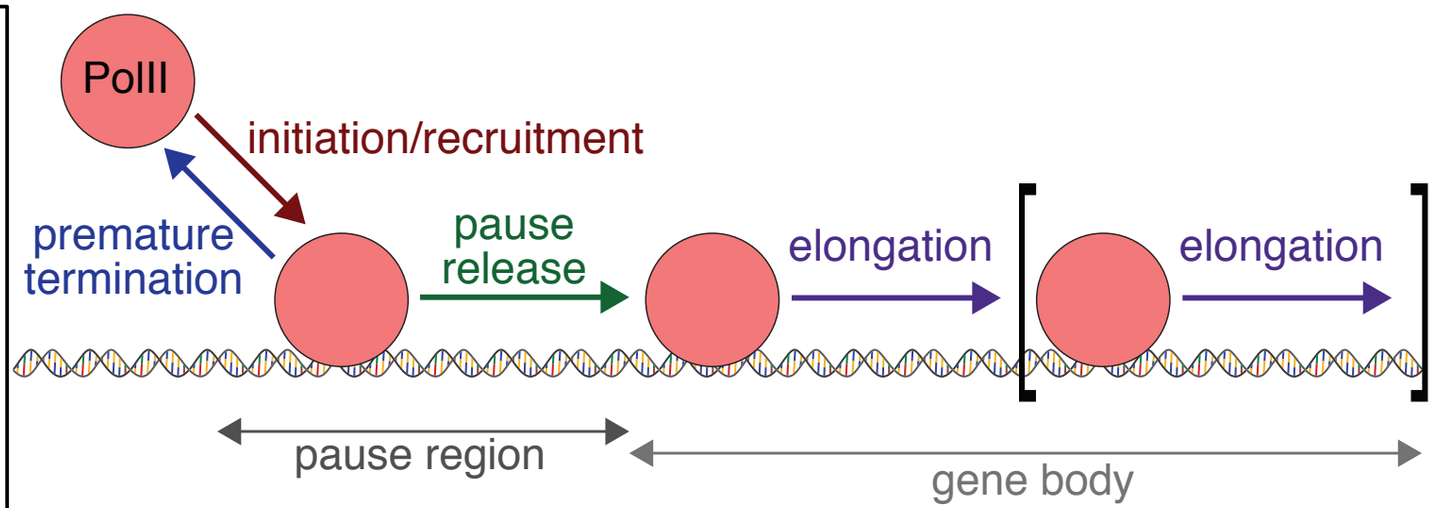


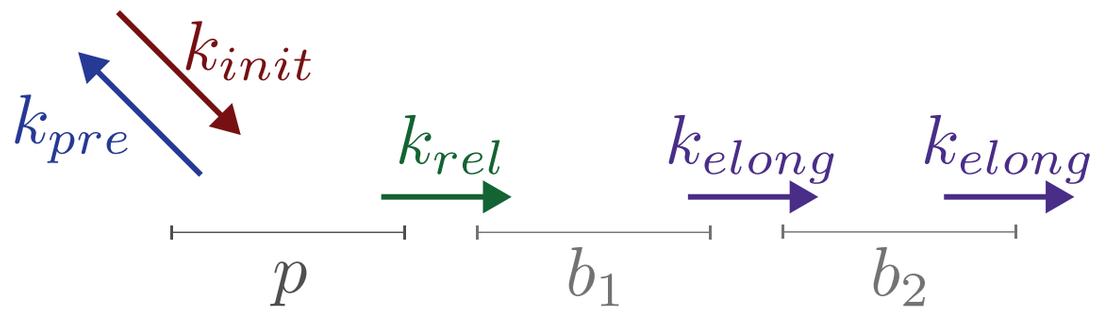
# ZNF143 stimulates transcription from the promoter



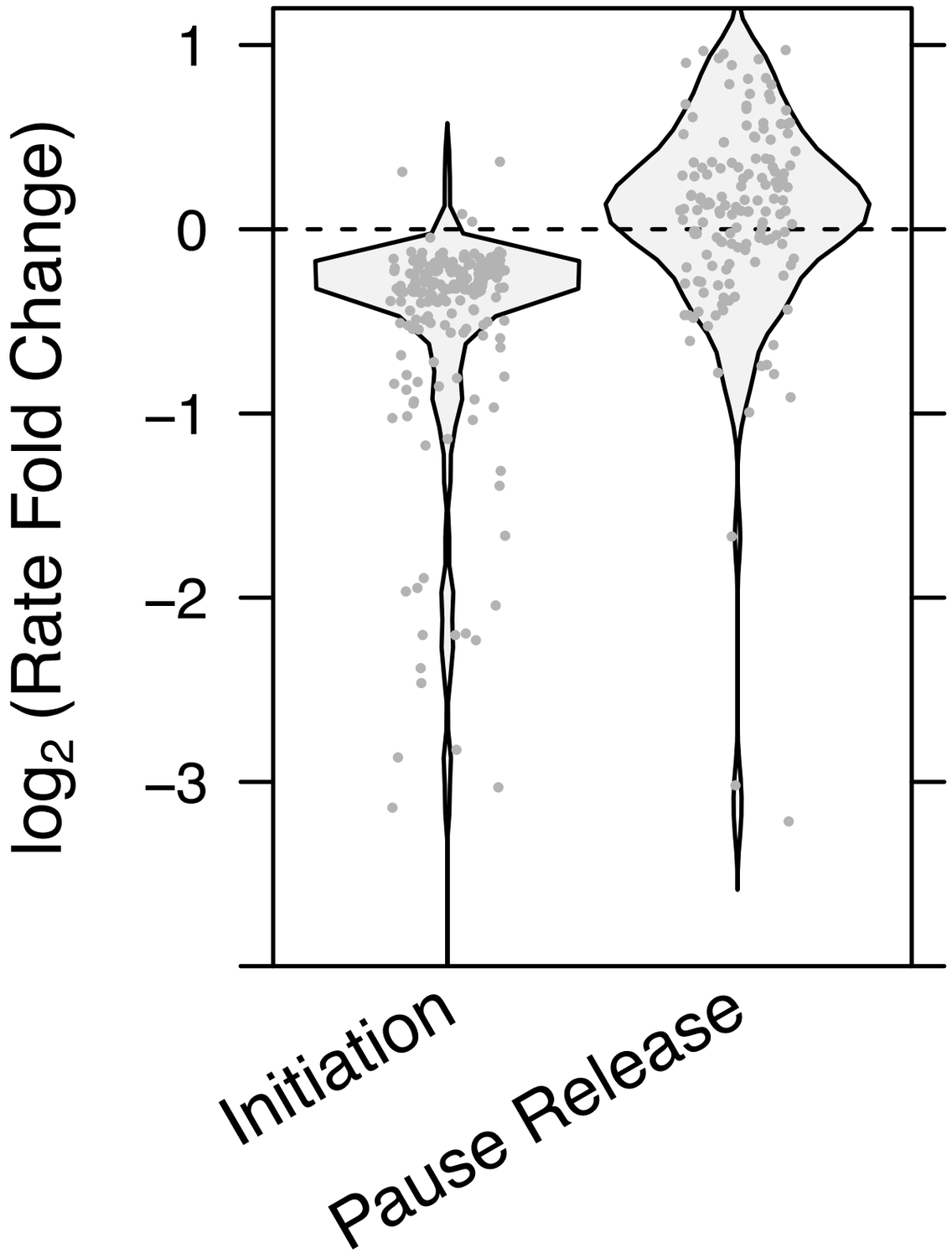
# Compartment modeling on each down gene



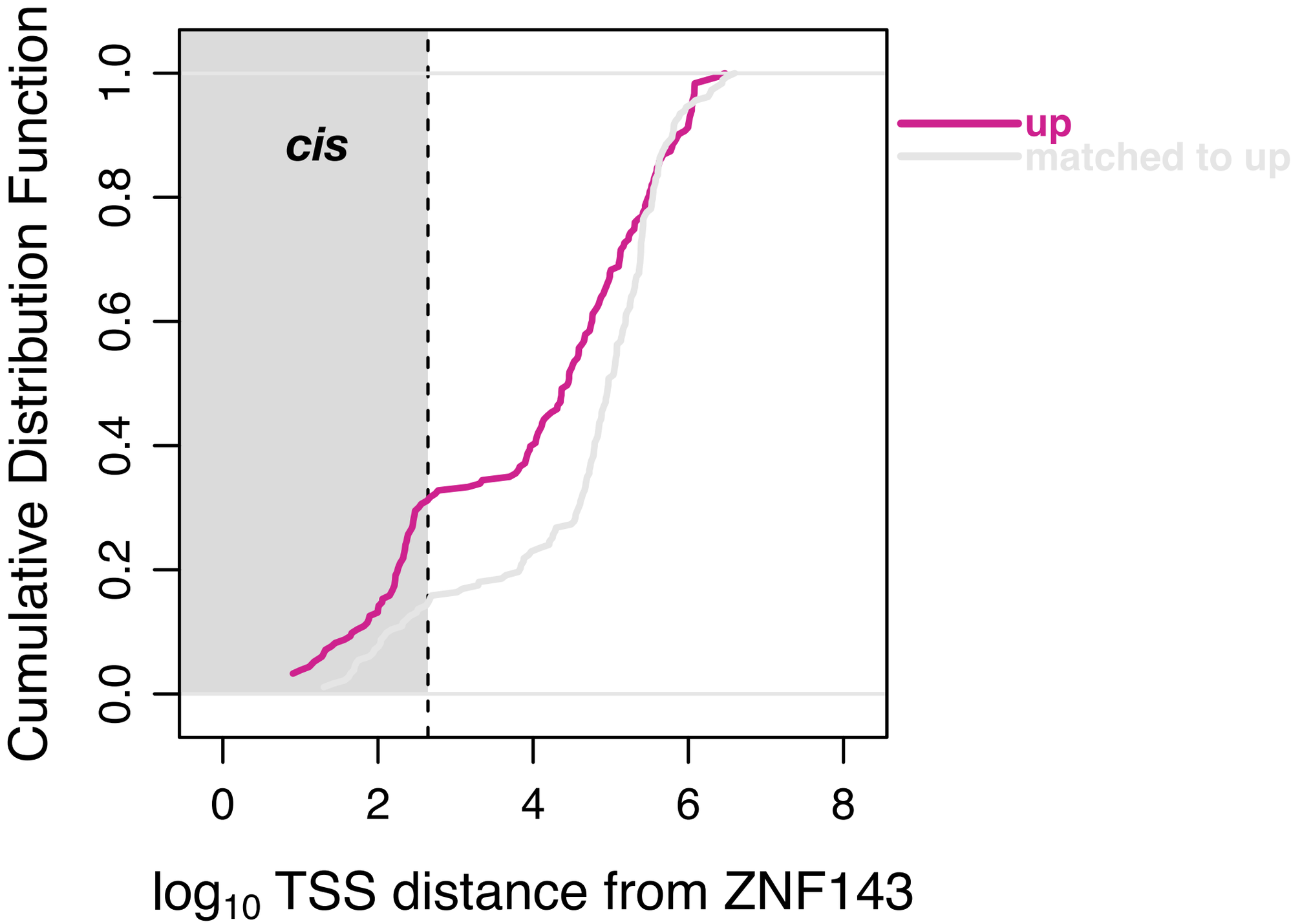



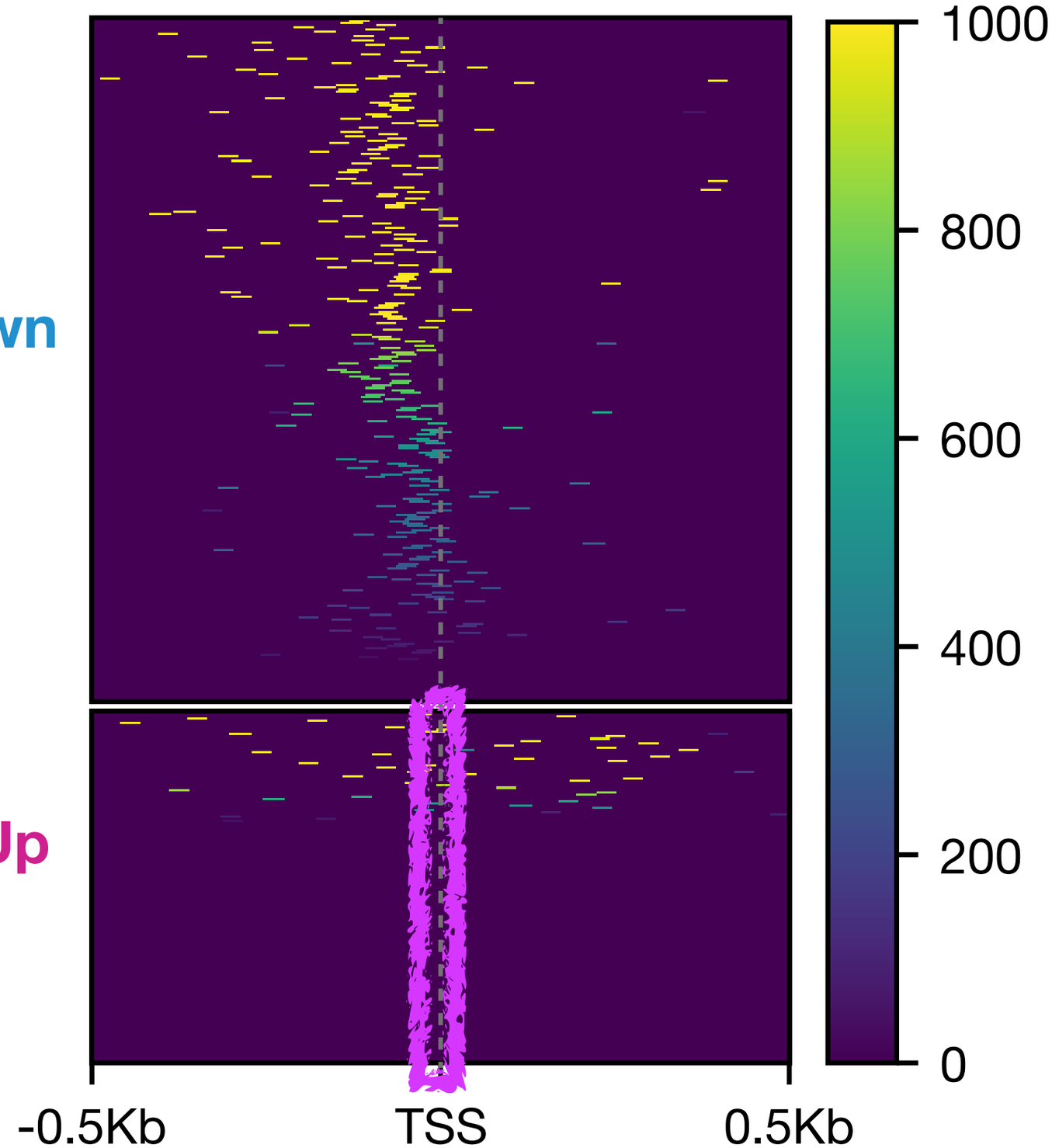
# ZNF143 predominantly regulates initiation



# How does ZNF143 repress gene targets?

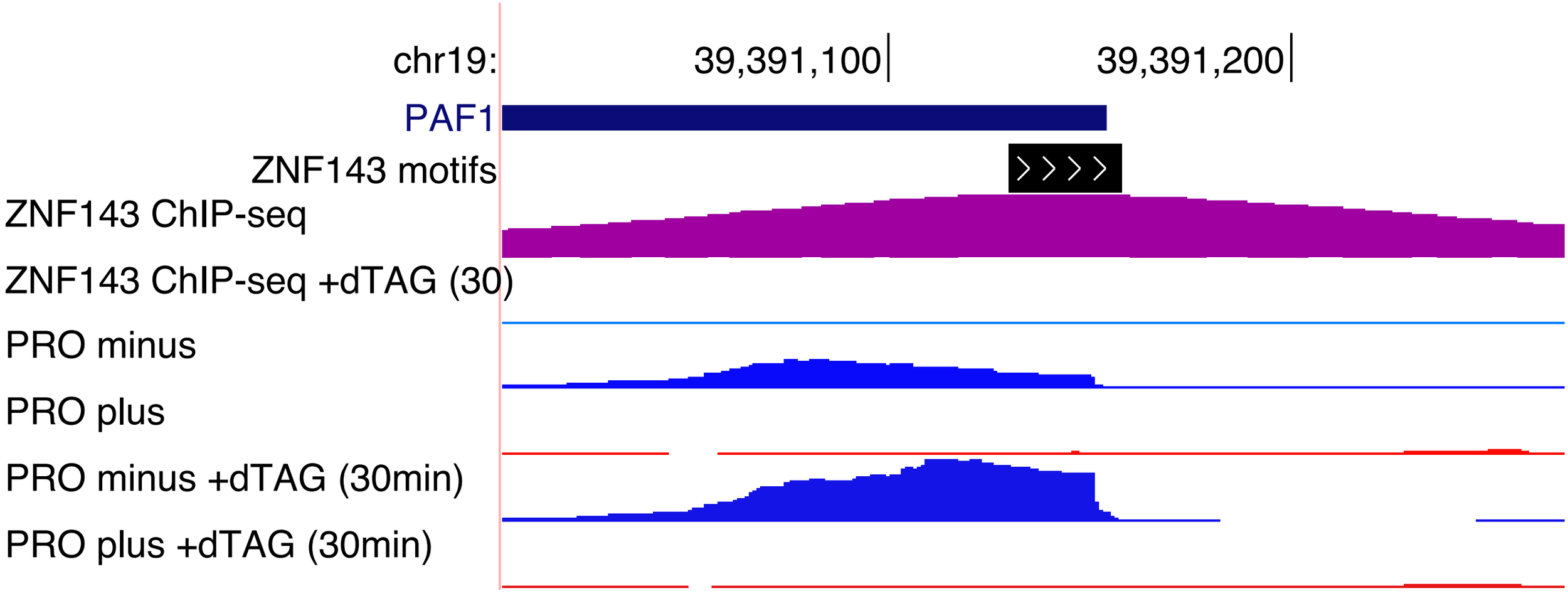


# What explains TSS-bound ZNF143 depletion increasing gene expression?

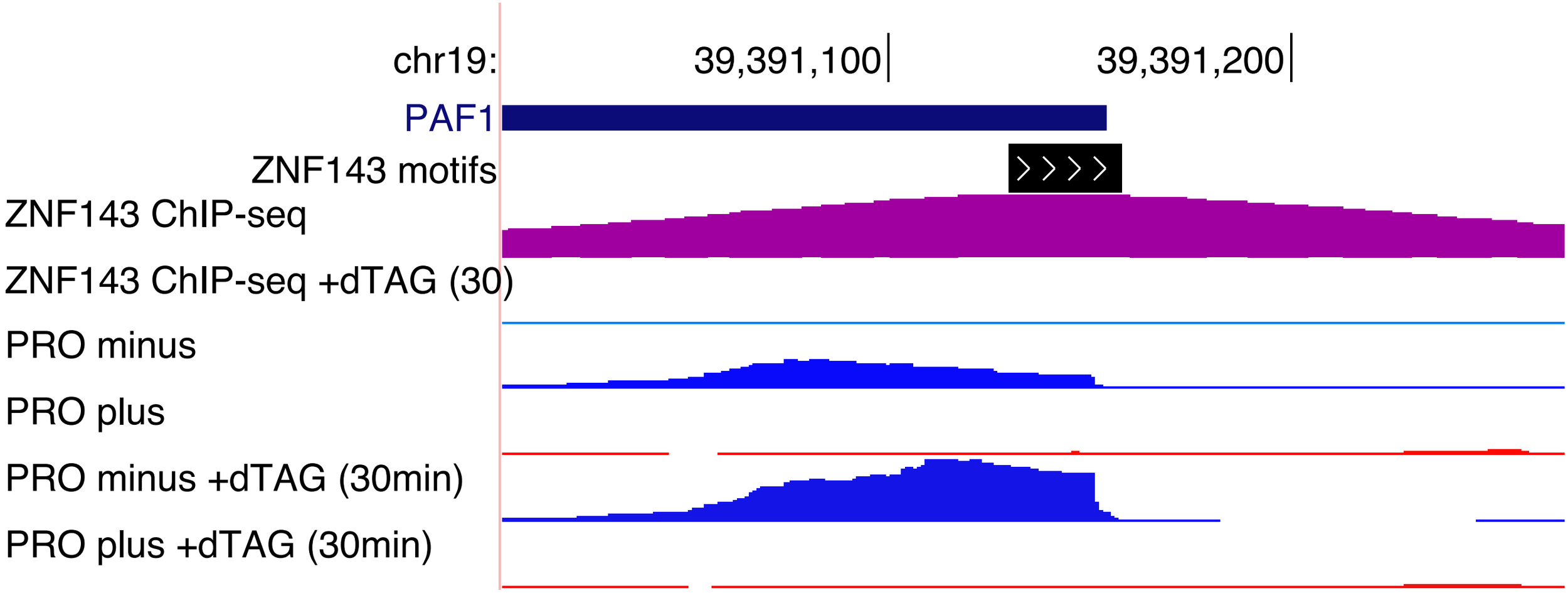



**183 up**
**57 in *cis***
**5 over the TSS**

# ZNF143 binding at the TSS may be an impediment at 5 up genes

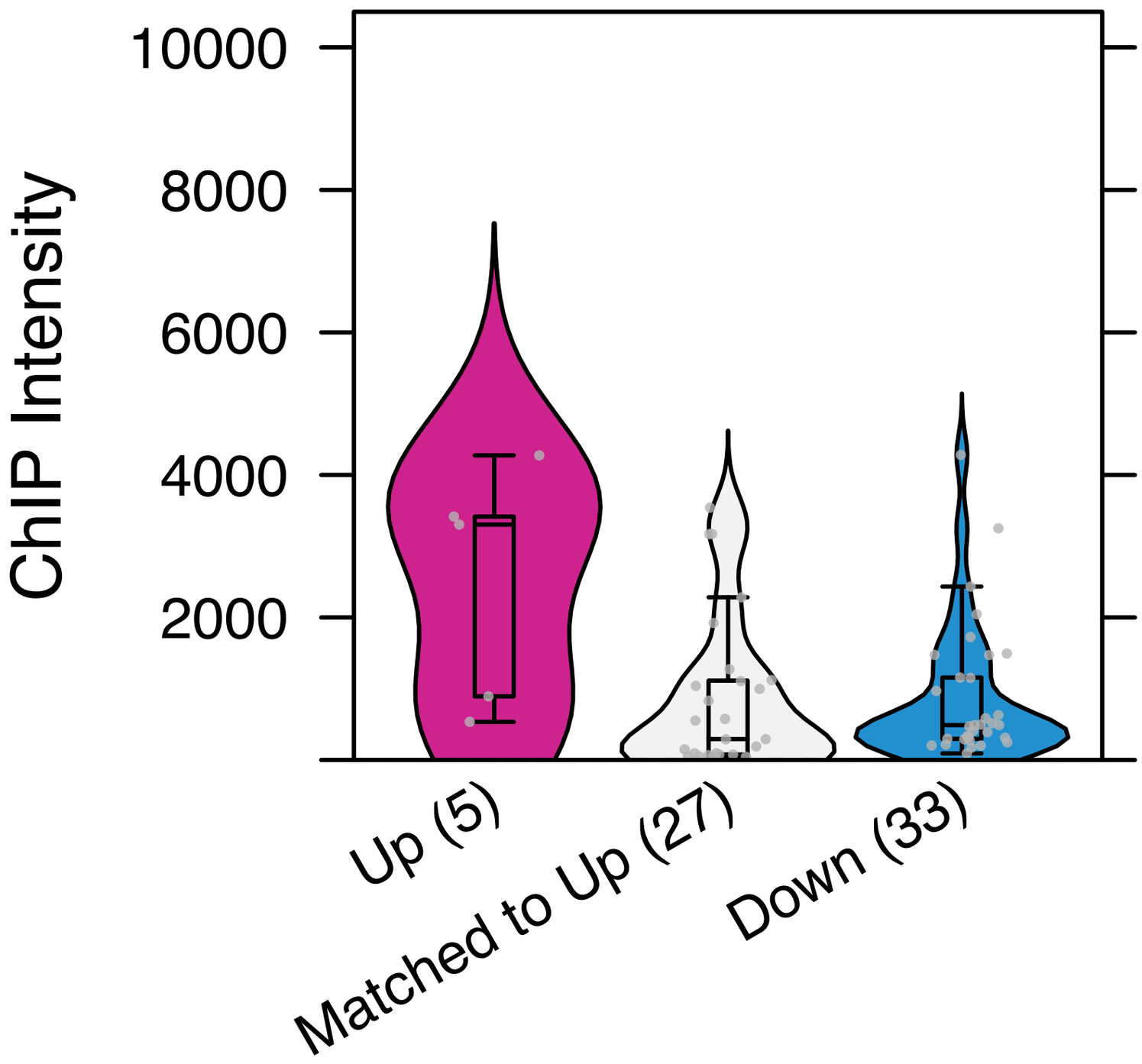


# ZNF143 binding at the TSS may be an impediment at 5 up genes

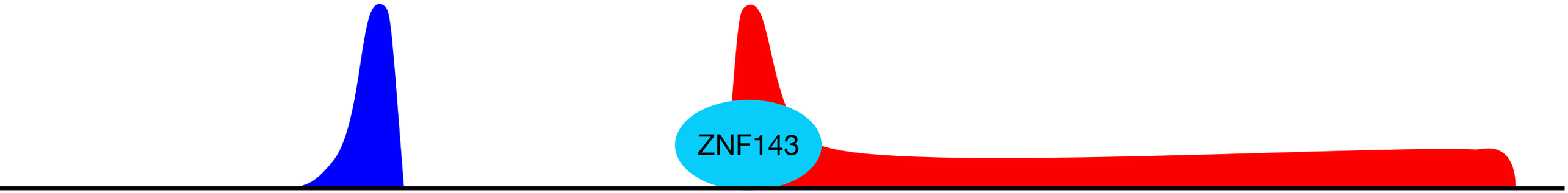



**The typical convincing functional experiment:**
**degrade ZNF143 rapidly and measure expression by PRO at these genes**

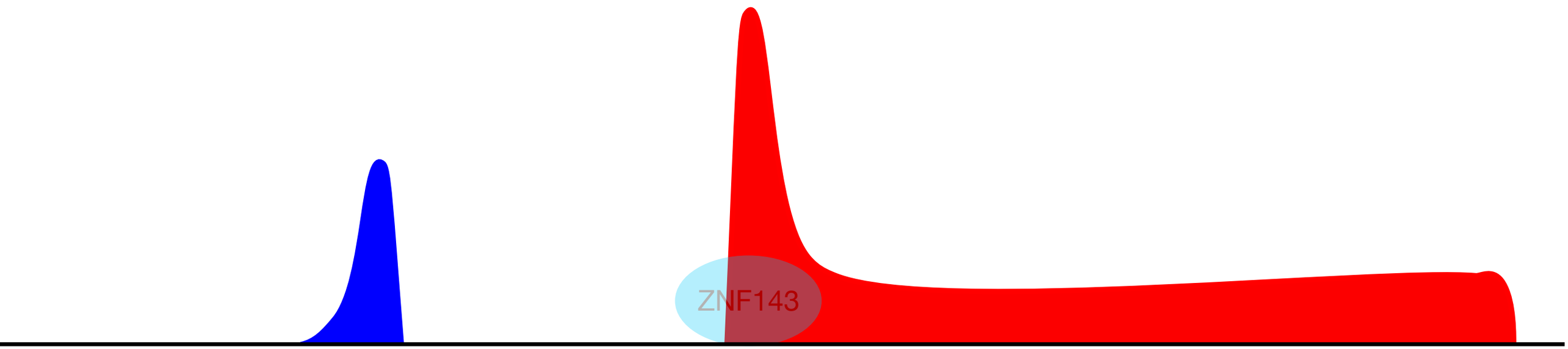
# ZNF143 binding is relatively strong at the 5 TSS-bound up genes



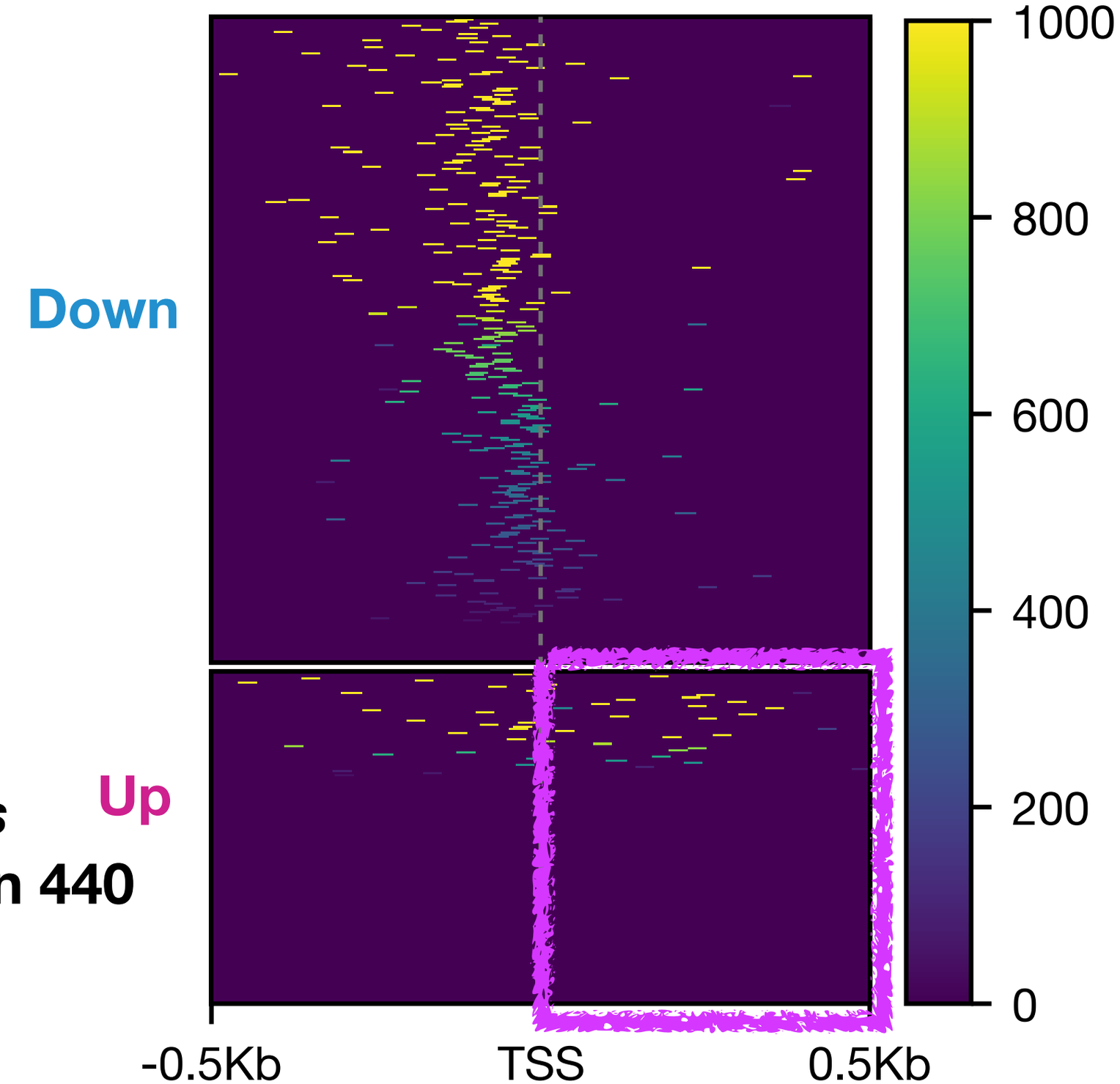
# ZNF143 blocks the TSS to attenuate transcription



# ZNF143 blocks the TSS to attenuate transcription

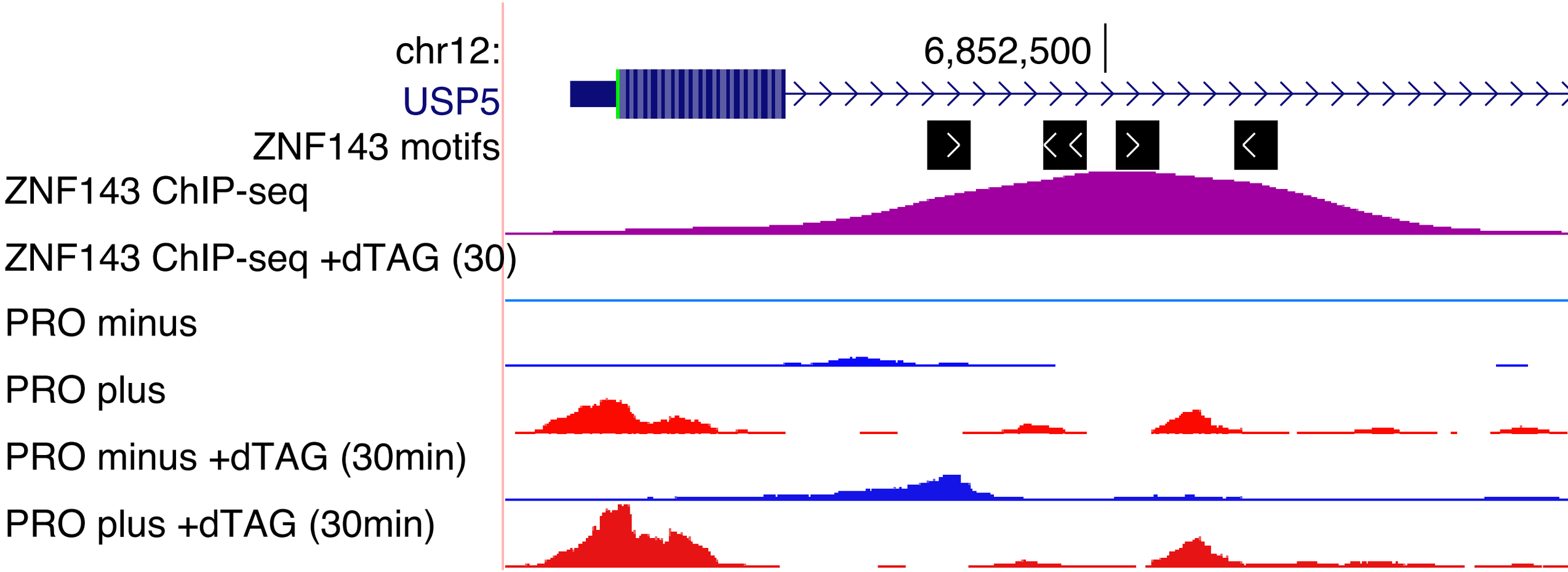


# What explains downstream-bound ZNF143 depletion increasing gene expression?

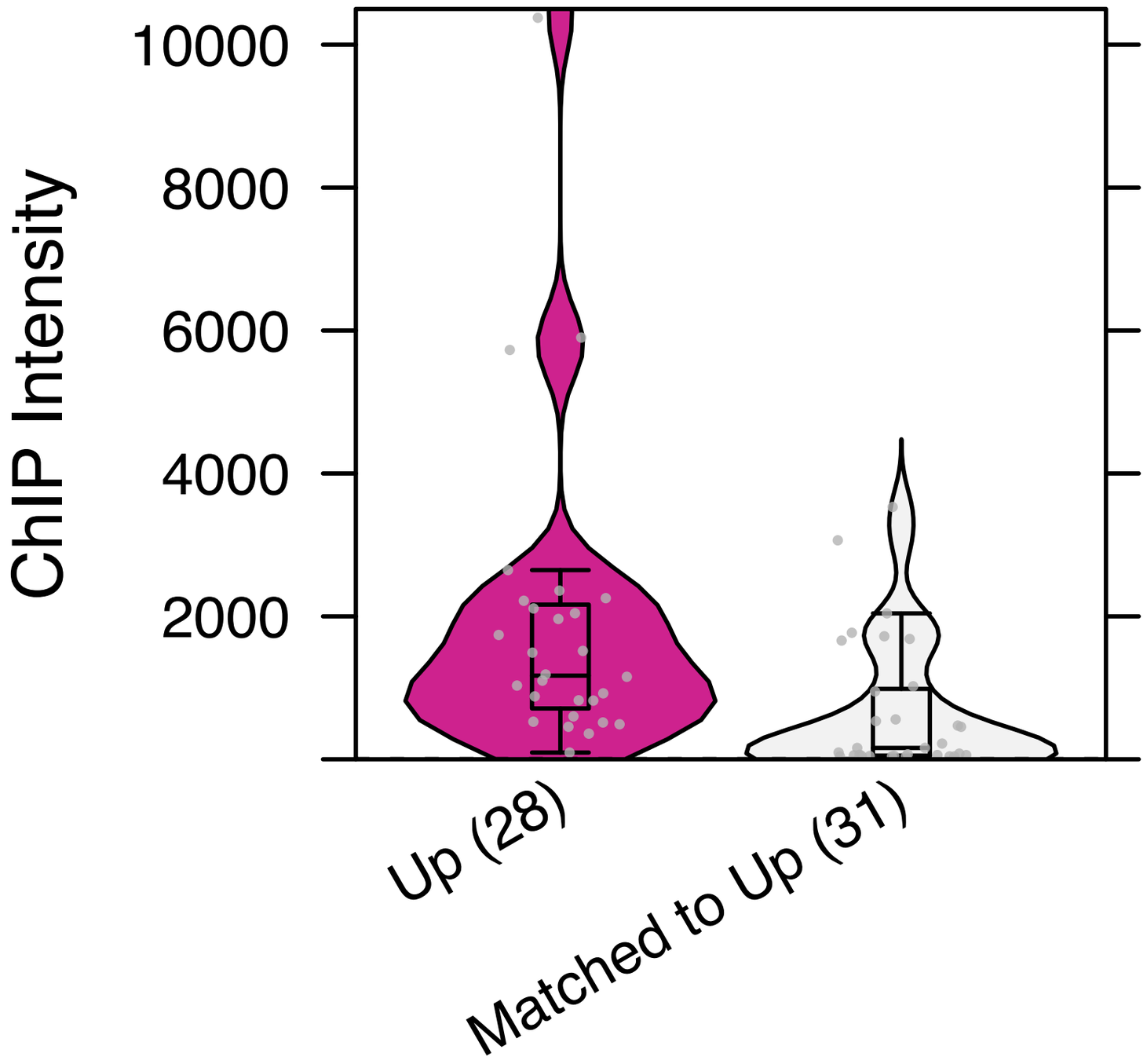



**183 up**
**57 in *cis***
**28 genic w/in 440**

# ZNF143 binding at the strongest ZNF143 peak may act as a roadblock



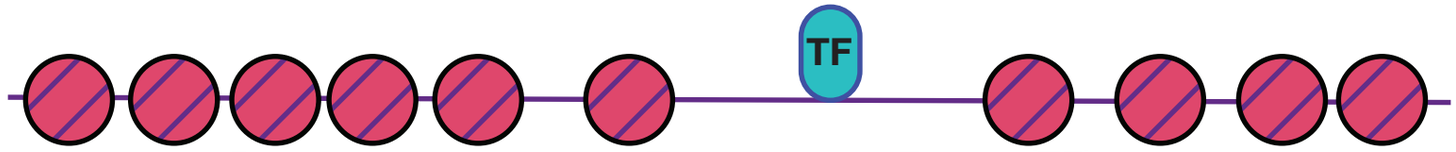
# Binding within up genes is relatively strong and may act as a road block



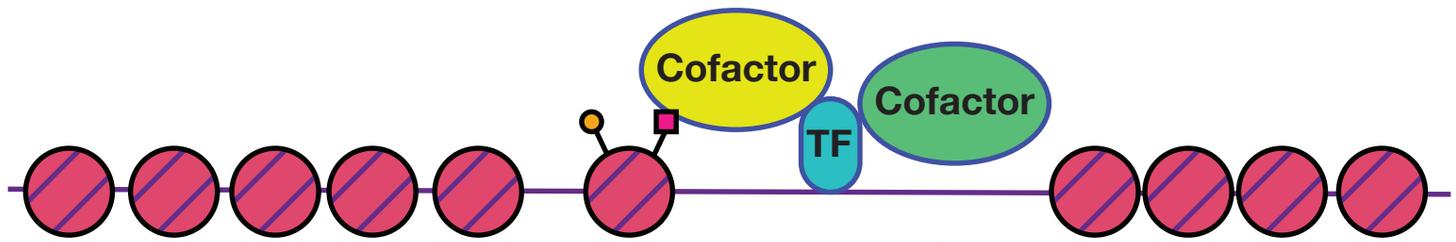
# Bidirectional Transcription is a hallmark of regulatory regions



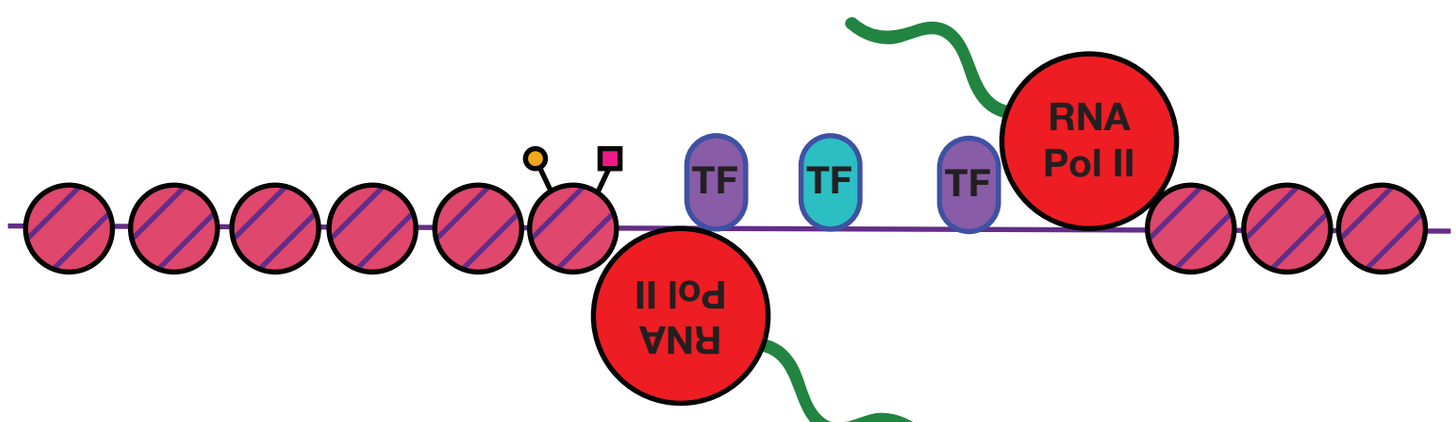
DNA sequence influences nucleosome positions/dynamics and chromatin compaction




DNA sequence directs Transcription Factor (TF) binding

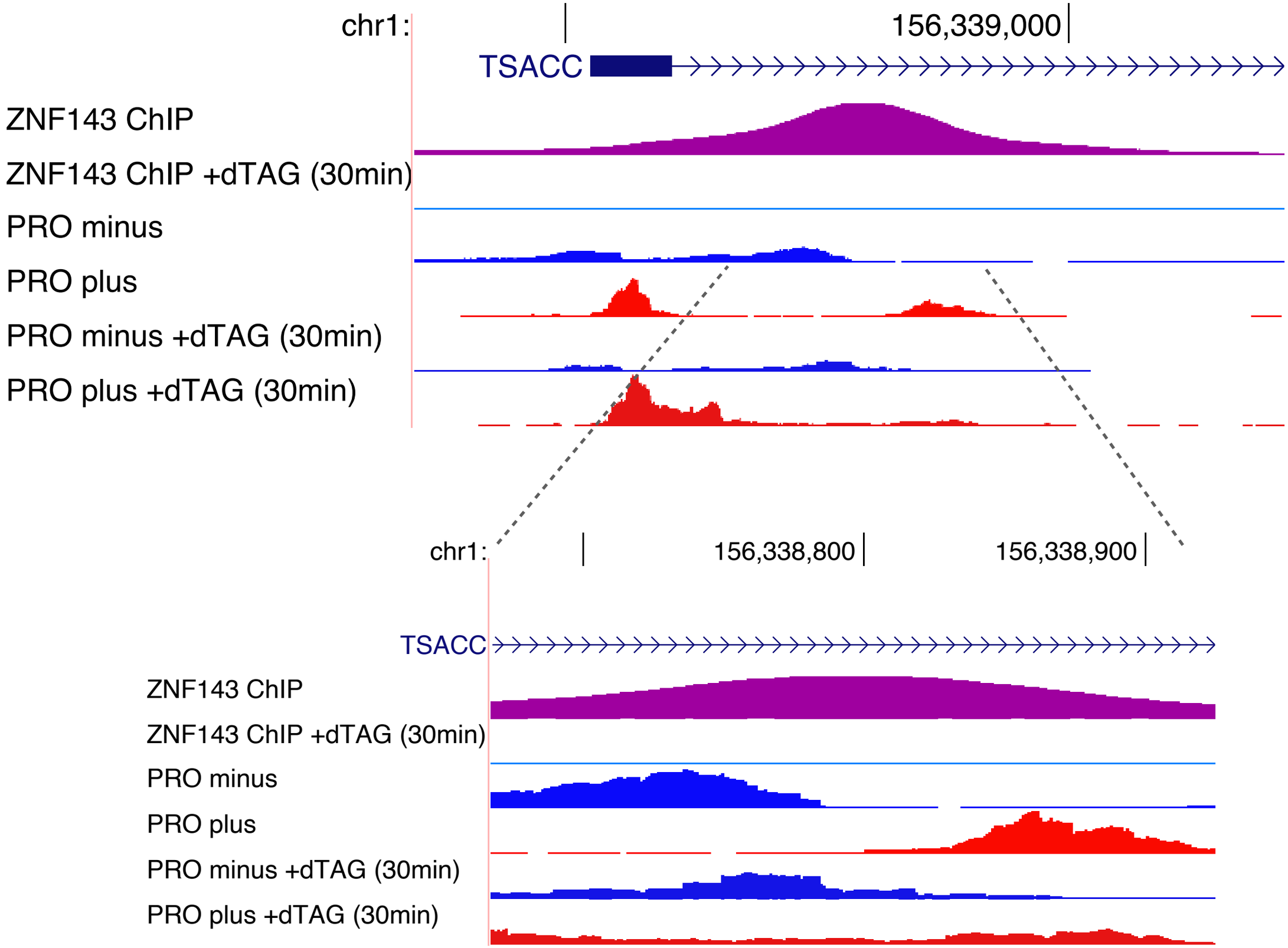



sequence-specific TFs recruit chromatin modifying/remodeling factors

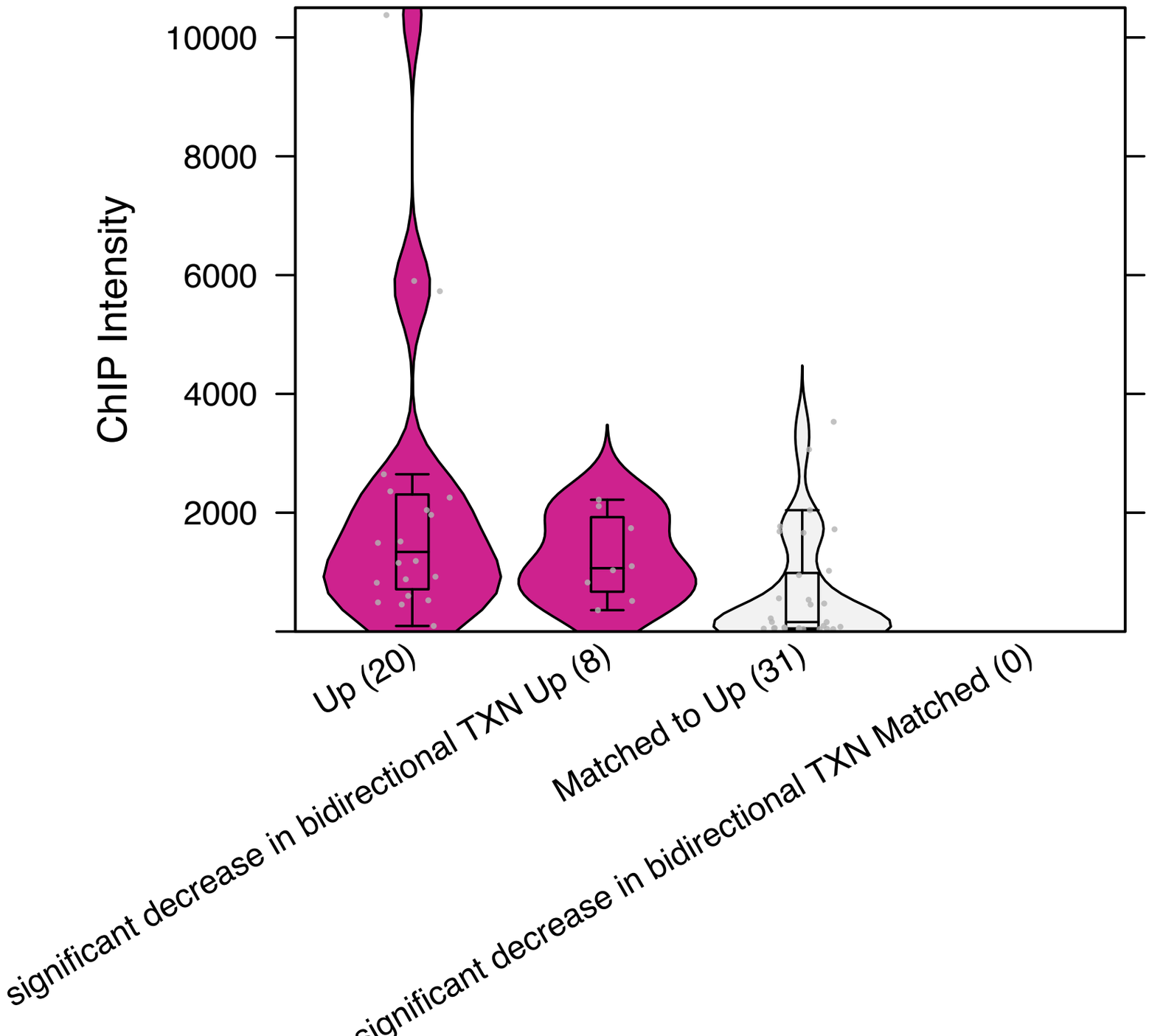



sequence-specific TFs recruit initiation factors and RNA Polymerase

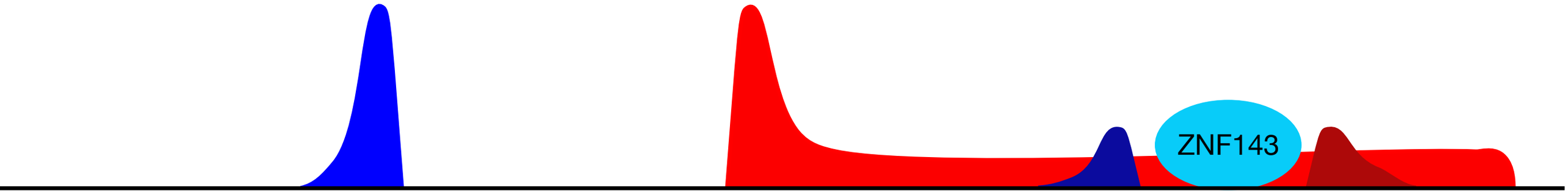
# Binding within genes stimulates bidirectional transcription that may act as a road block



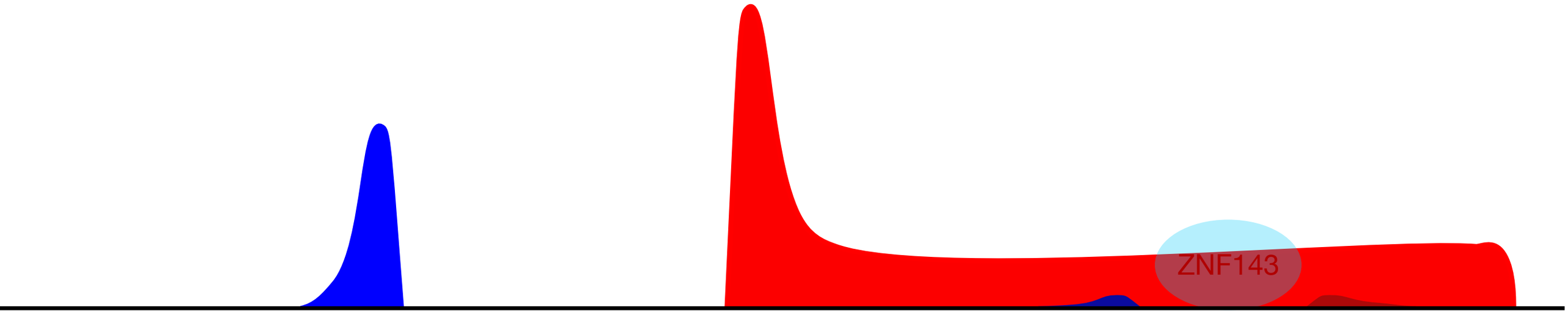
# Bidirectional transcription within genes may act as a road block



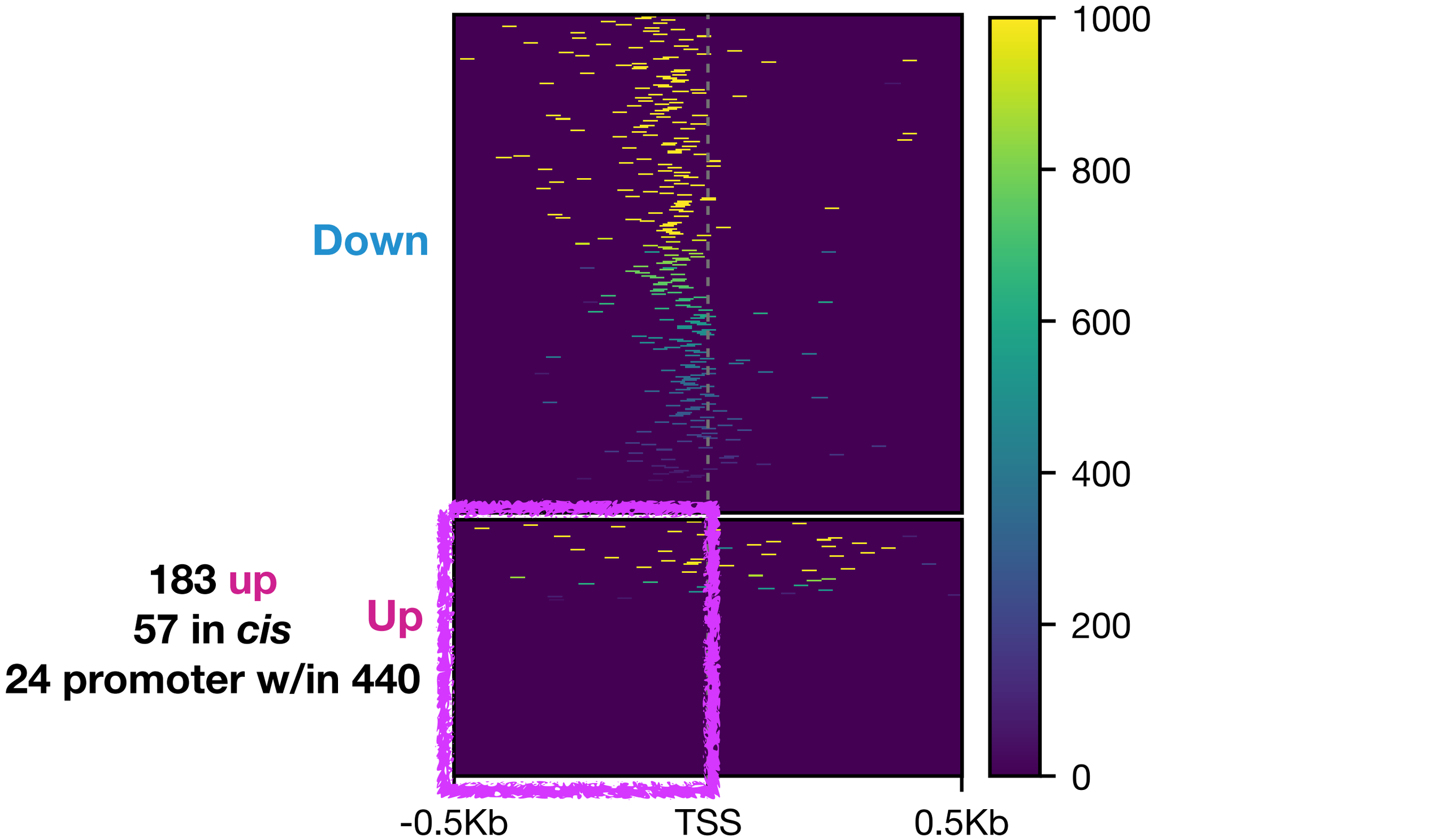
# ZNF143 and/or ZNF143-stimulated bidirectional TXN may act as physical barriers



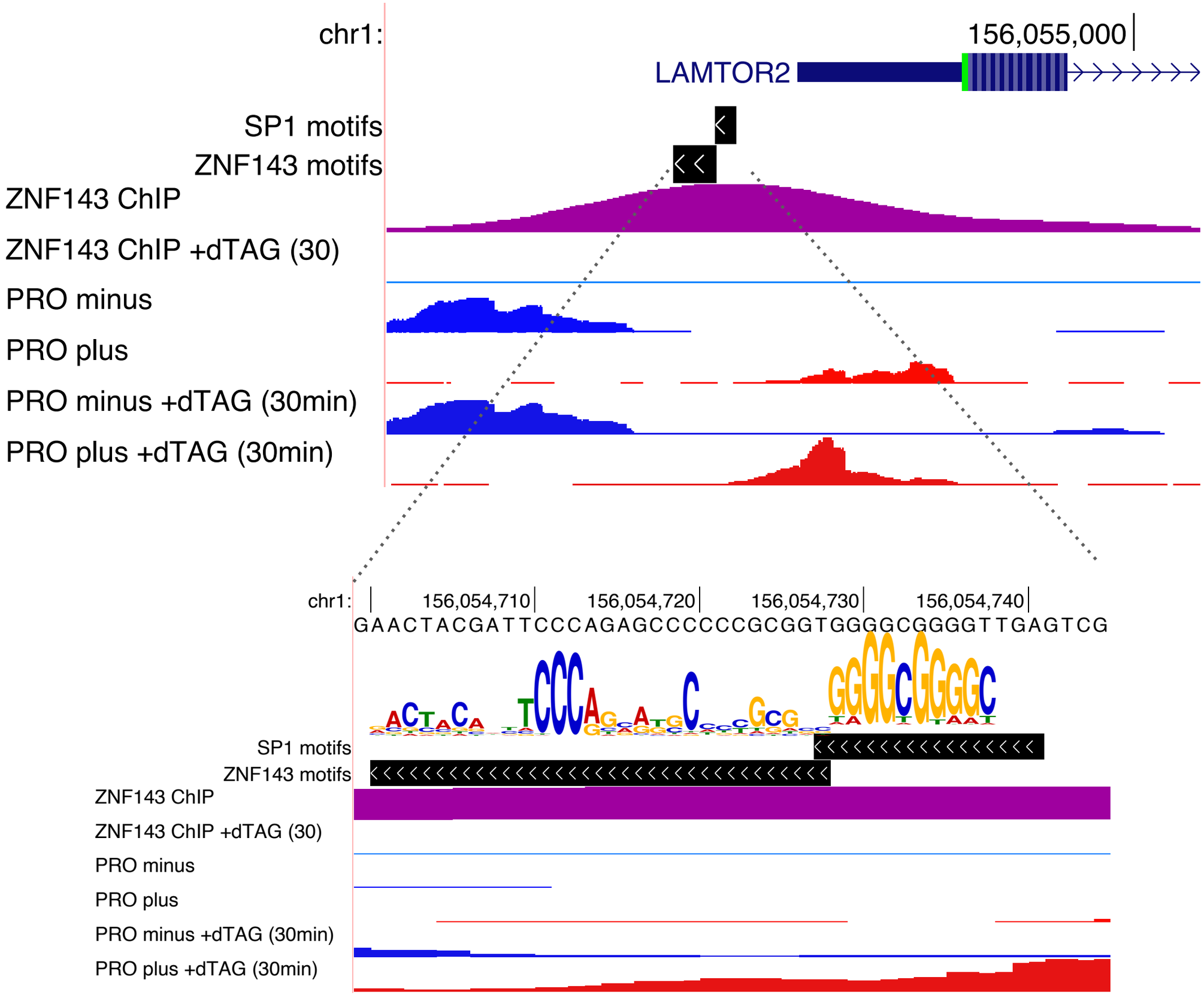
# ZNF143 and/or ZNF143-stimulated bidirectional TXN may act as physical barriers



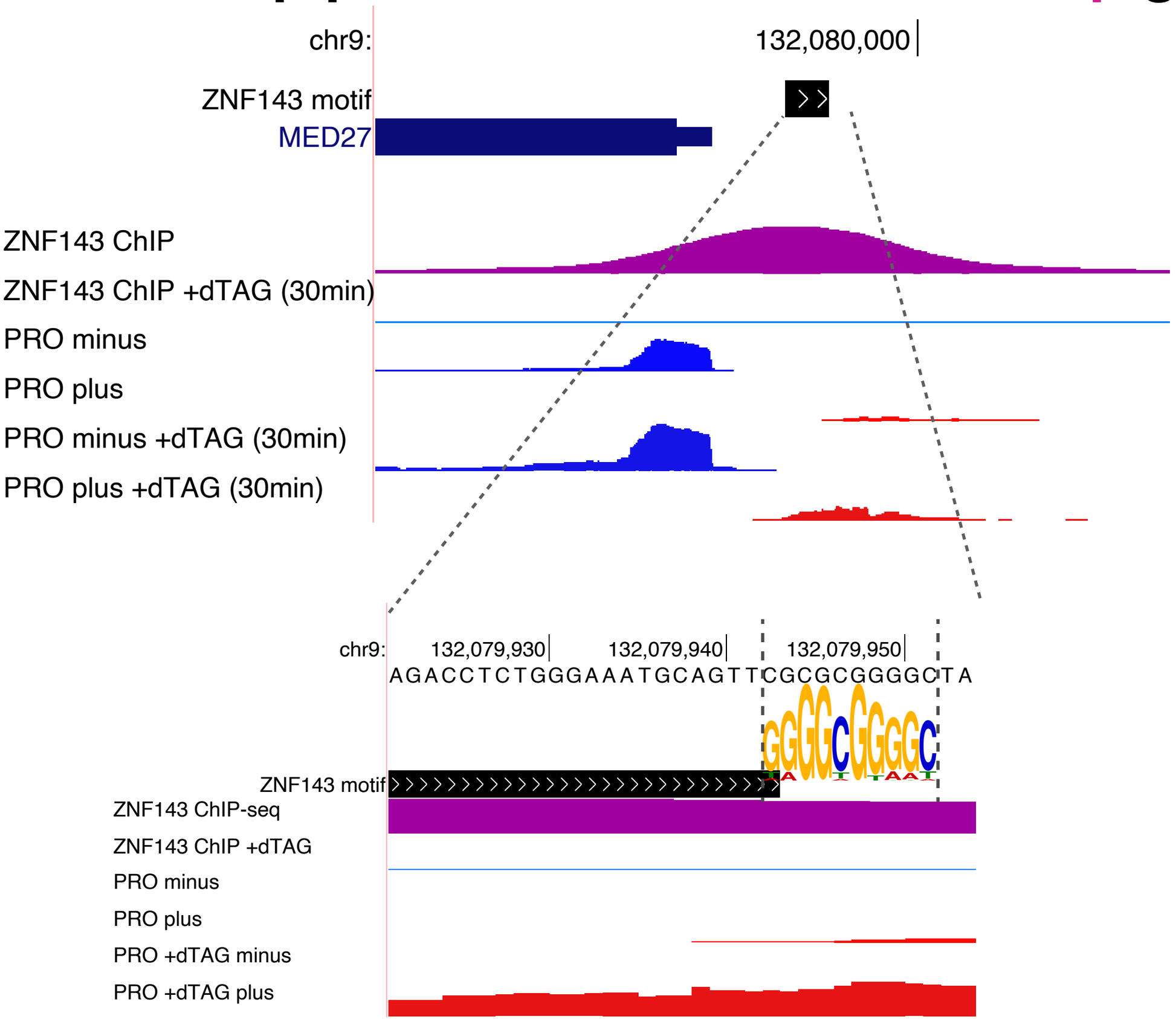
# What explains upstream-bound ZNF143 depletion increasing gene expression?



# SP motifs overlap promoter ZNF143 sites at up genes

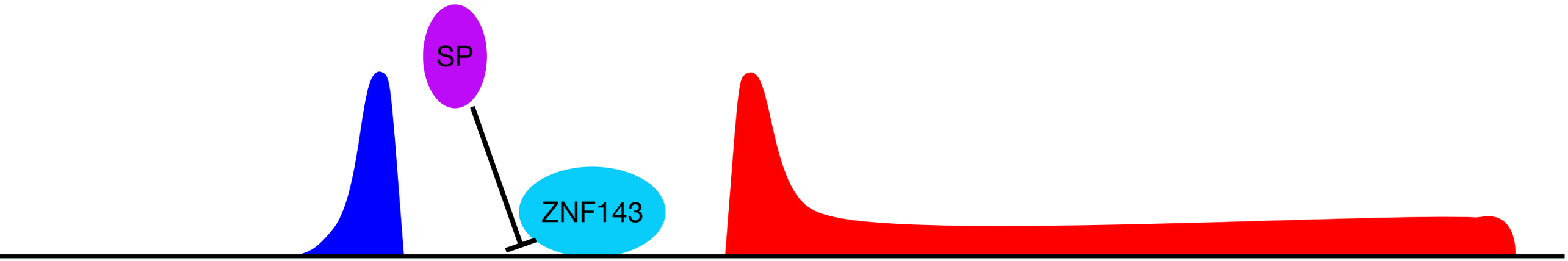


# SP motifs overlap promoter ZNF143 sites at up genes

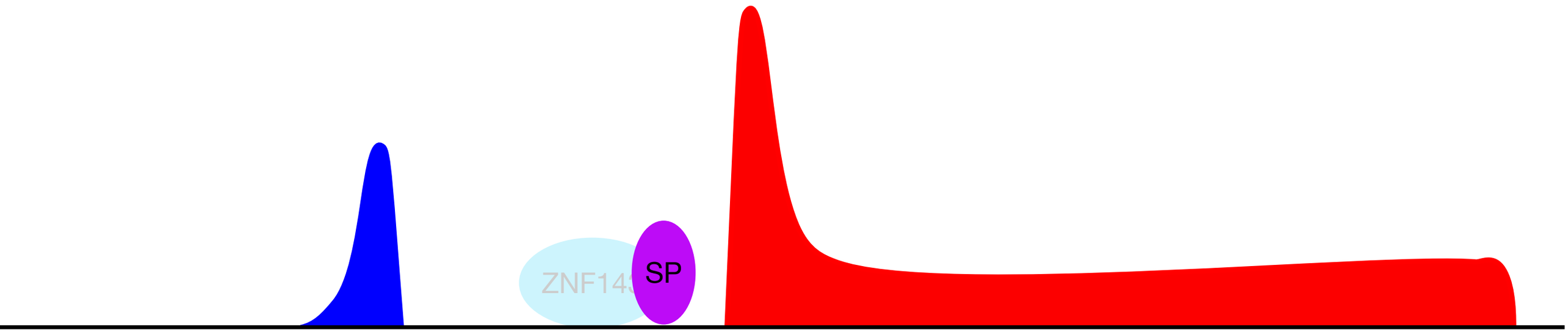


**SP motifs overlap promoter ZNF143 sites at 46% (11/24) of up genes**

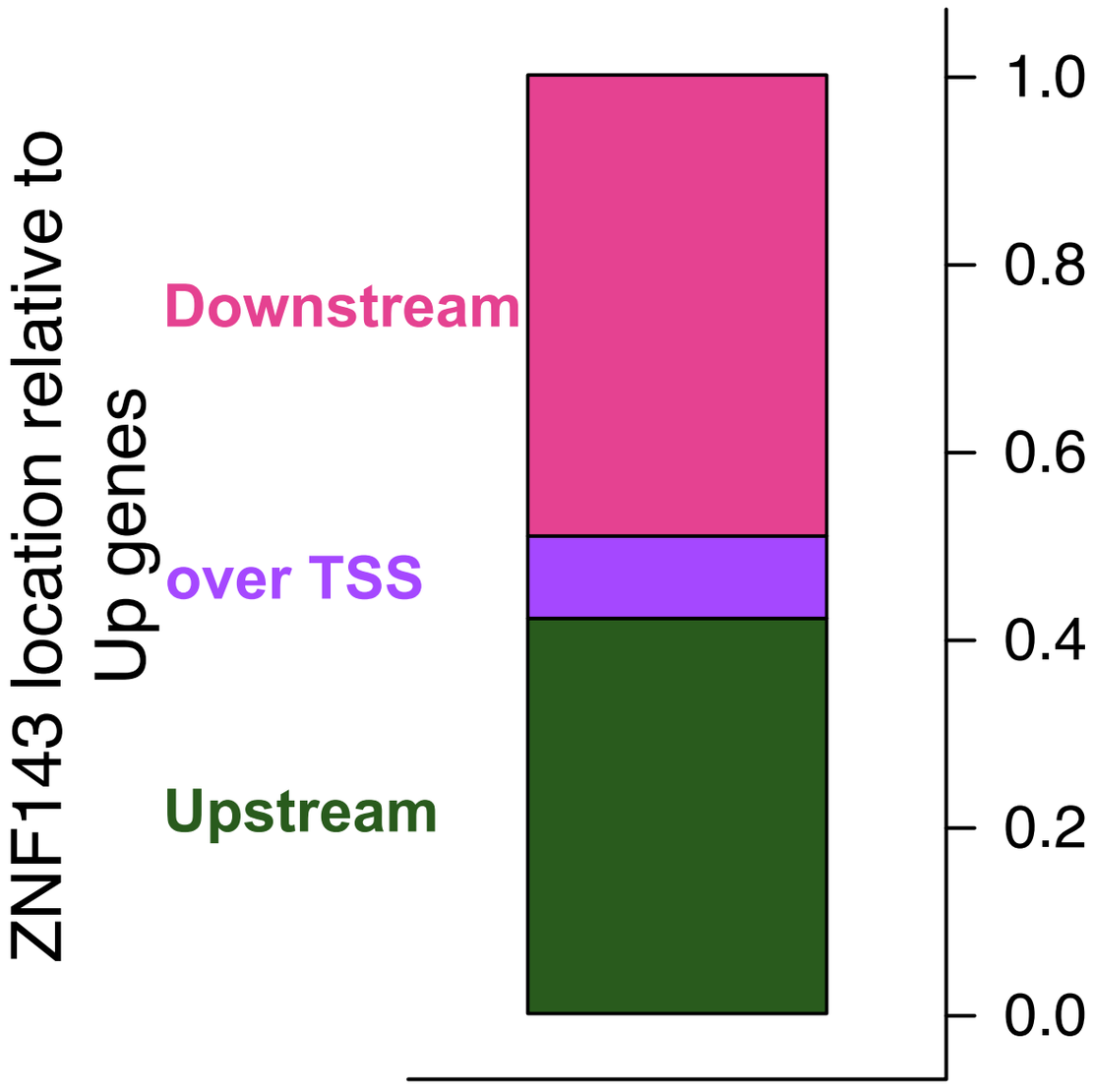
# ZNF143 may outcompete better activators such as SP/KLF factors for promoter access



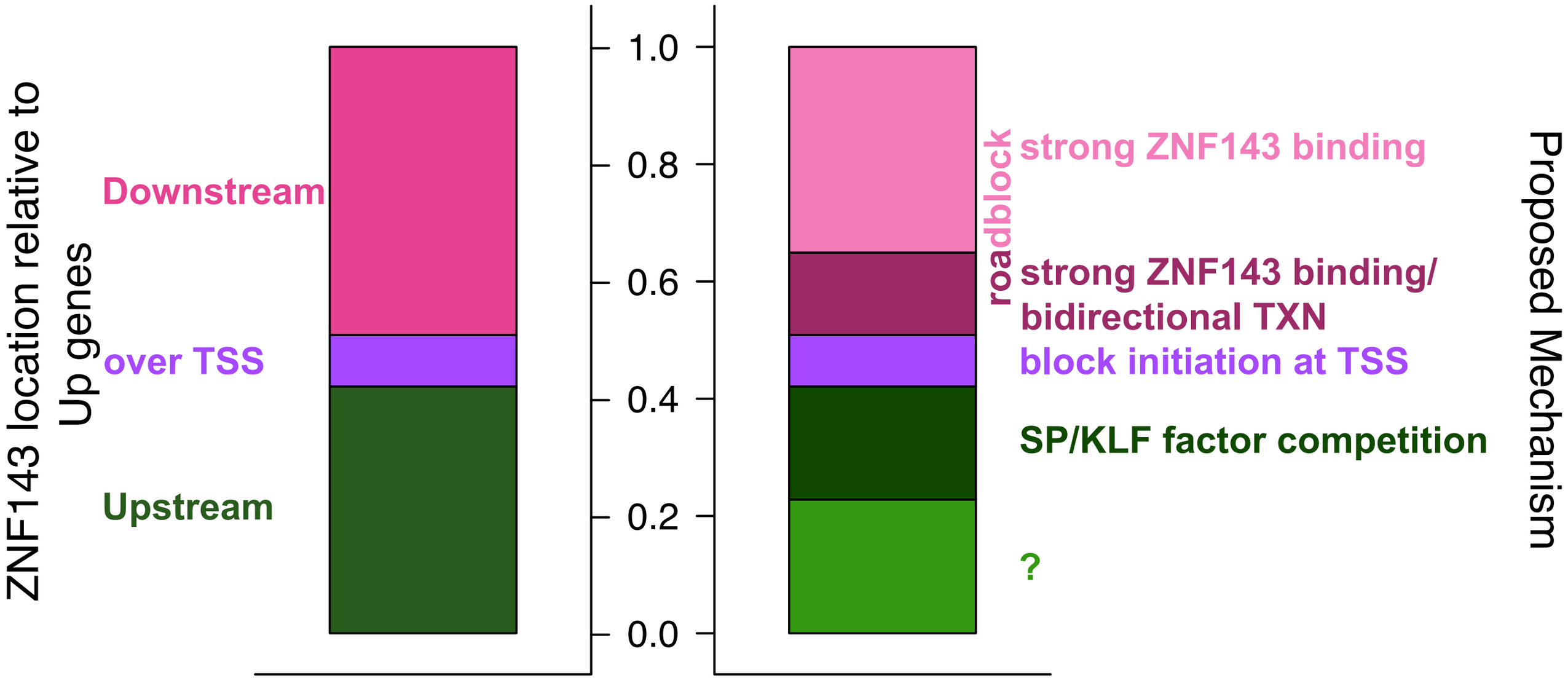
# ZNF143 may outcompete better activators such as SP/KLF factors for promoter access



# Proposed repressive mechanisms of an Activator



# Proposed repressive mechanisms of an Activator



# ZNF143 Conclusions/Future

- The molecular functions of ZNF143:

    1. Bind DNA

    2. Stimulate RNA polymerase initiation

- The molecular functions do not change when ZNF143 represses target genes.

- To do:

    1. ChIP-seq SP/KLF factors to test redistribution



**www.guertinlab.org**
**https://github.com/guertinlab**
**@guertin.bsky.social**
**guertin@uchc.edu**

**Arun Dutta** (adipogenesis)

**Jinhong Dong** (ZNF143 degron)

Rudradeep Mukherjee (compartment modeling & Twist1/2)

Tommy Scott (Estrogen response & TRPS1 degradation)

Siyu Sun (GATA3 degron & Estrogen response)



UCONN
SCHOOL OF MEDICINE



National Institute
of General Medical
Sciences

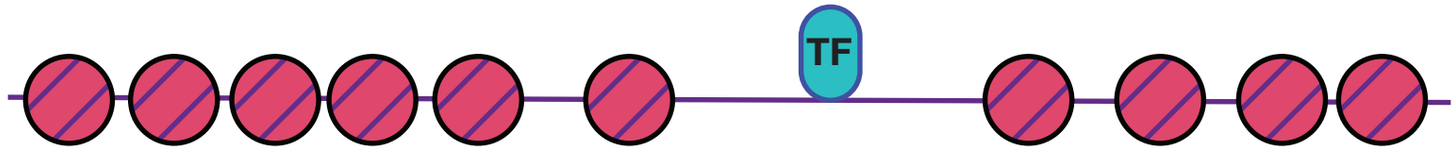


National **Heart**
**Lung and Blood** Institute

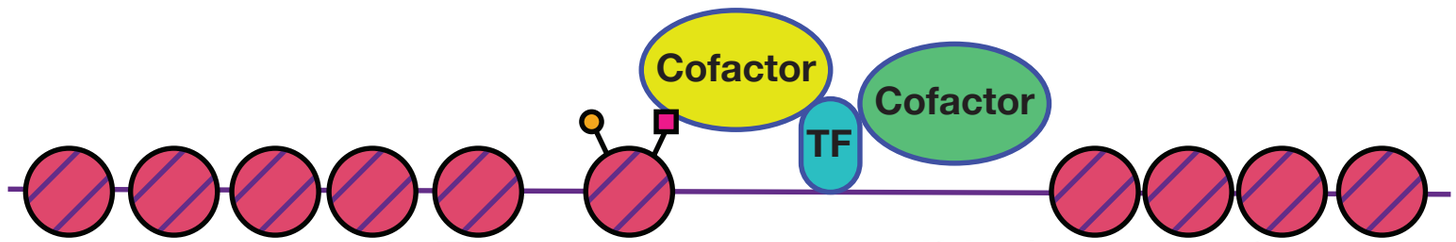
# Bidirectional Transcription is a hallmark of regulatory regions



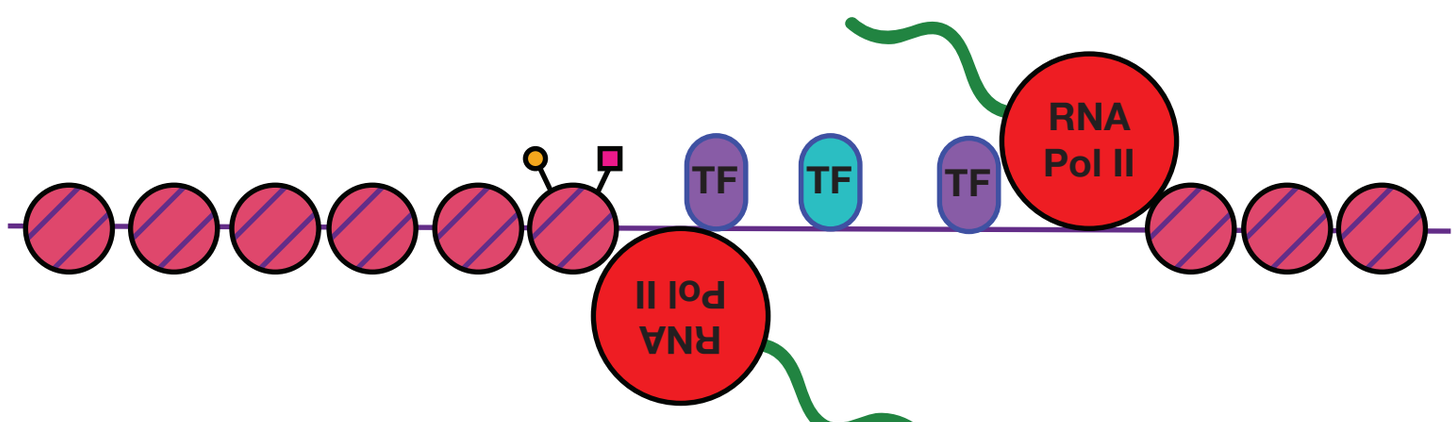
DNA sequence influences nucleosome positions/dynamics and chromatin compaction




DNA sequence directs Transcription Factor (TF) binding

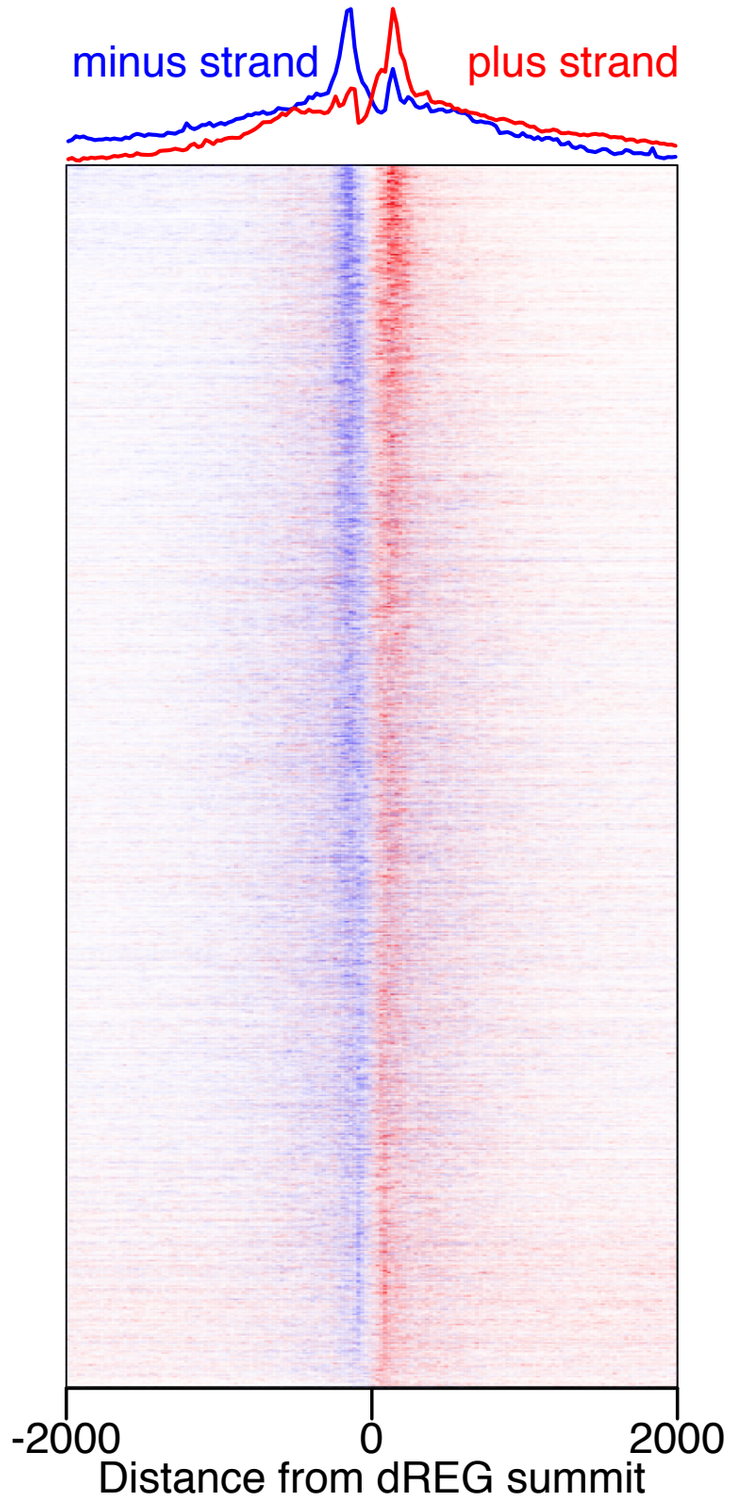



sequence-specific TFs recruit chromatin modifying/remodeling factors

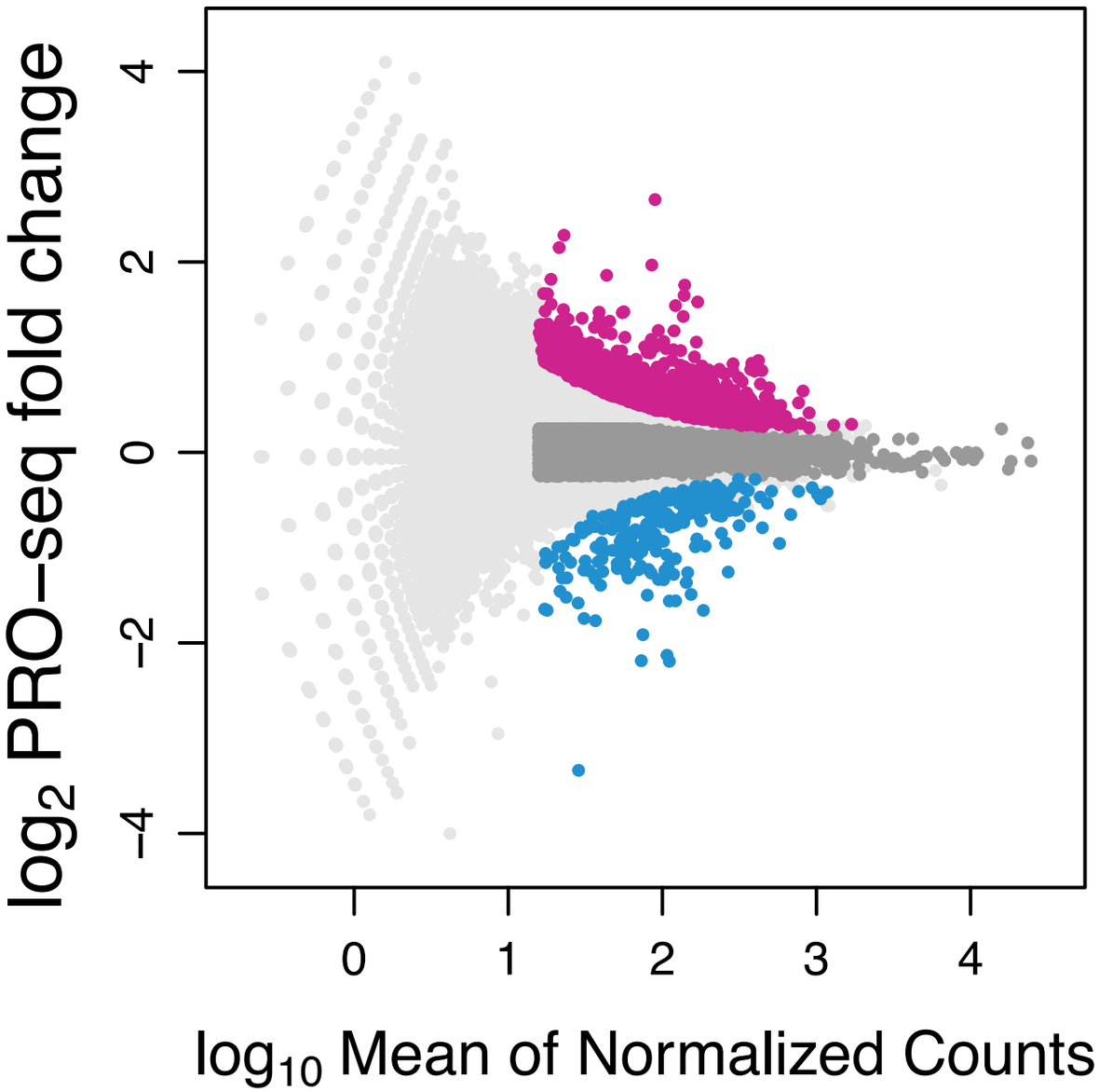



sequence-specific TFs recruit initiation factors and RNA Polymerase

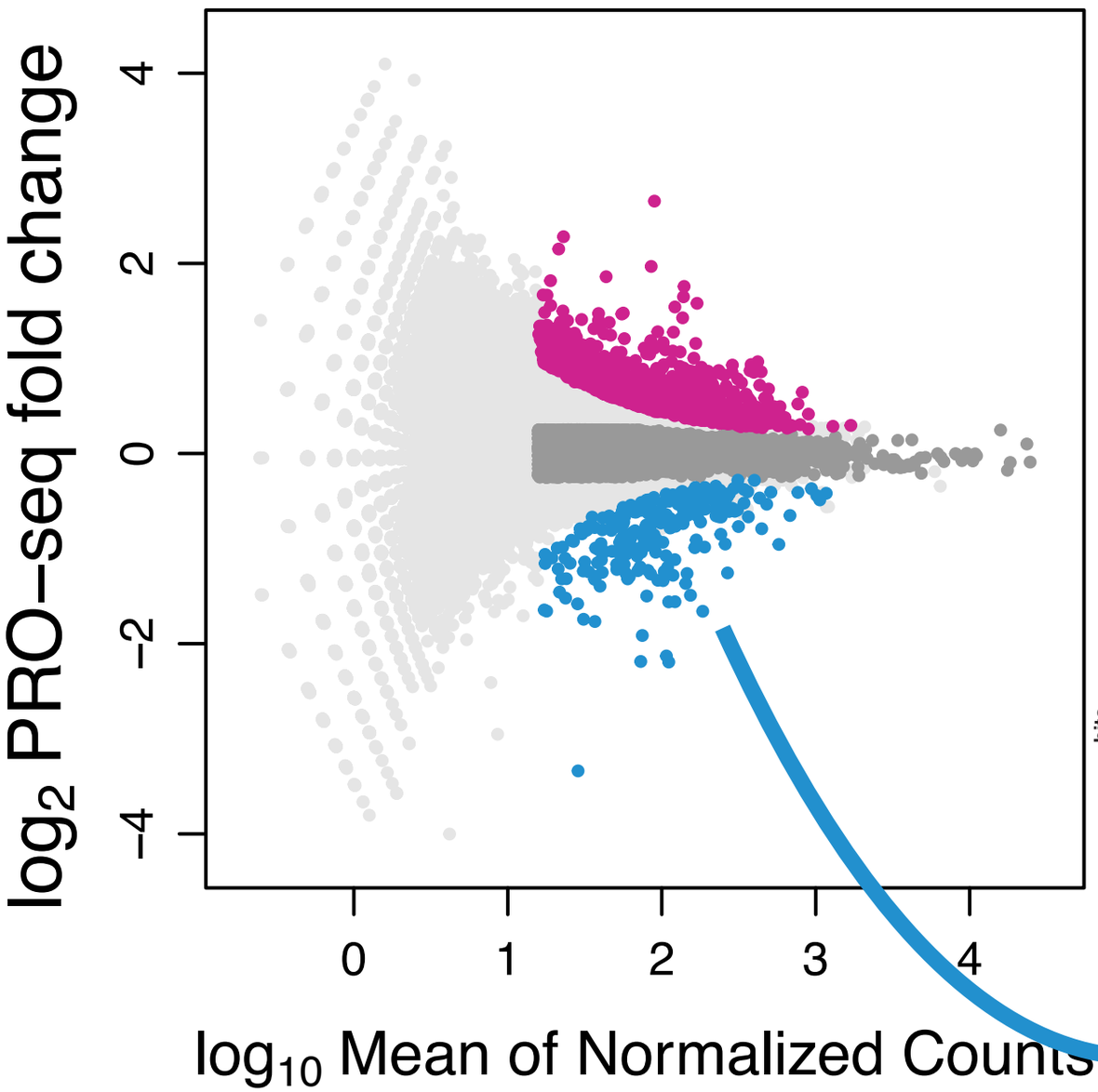
# Bidirectional transcription signatures from PRO-seq independently identifies putative regulatory regions

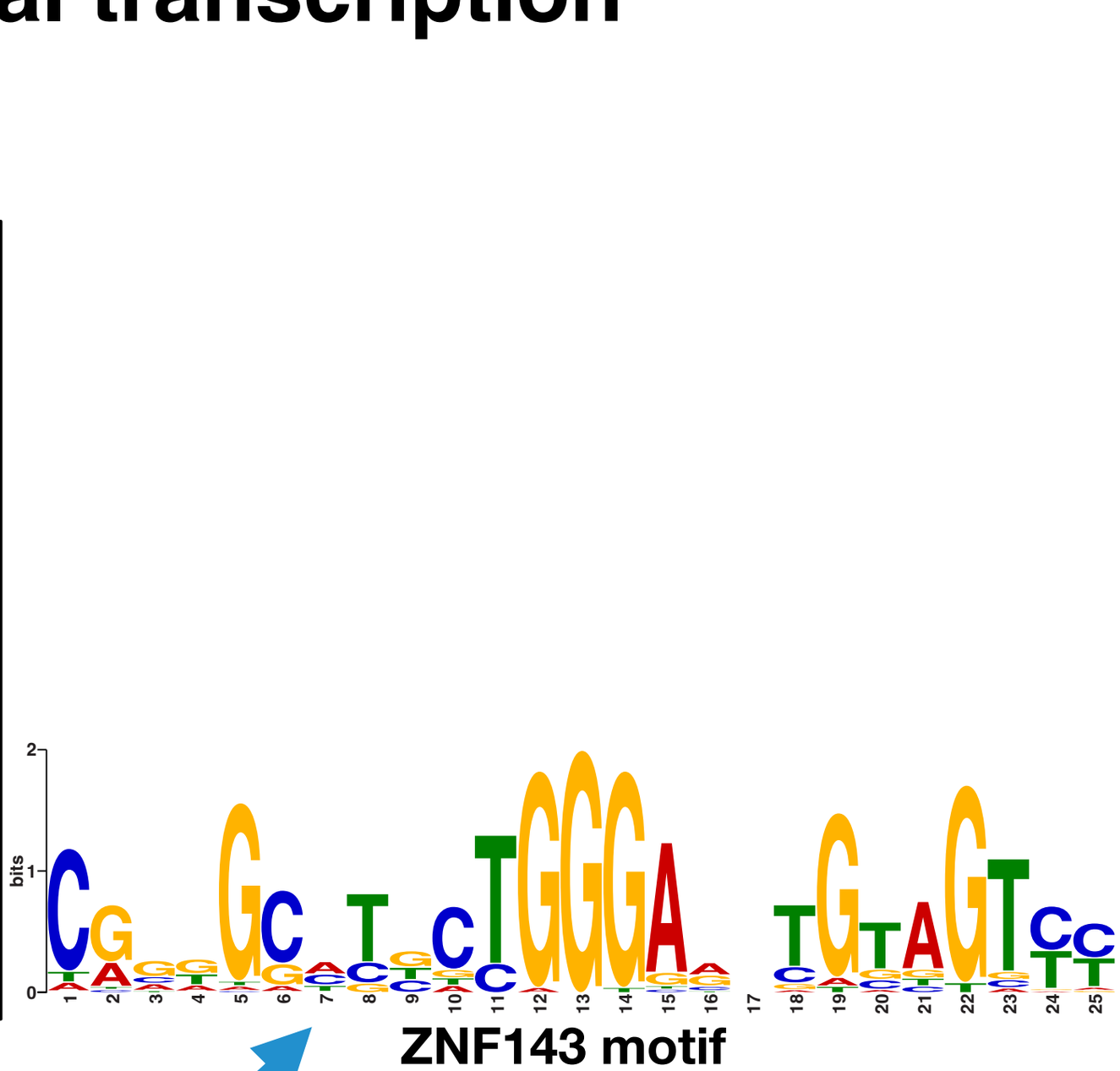


# 30 min of ZNF143 depletion causes both increased and decreased bidirectional transcription

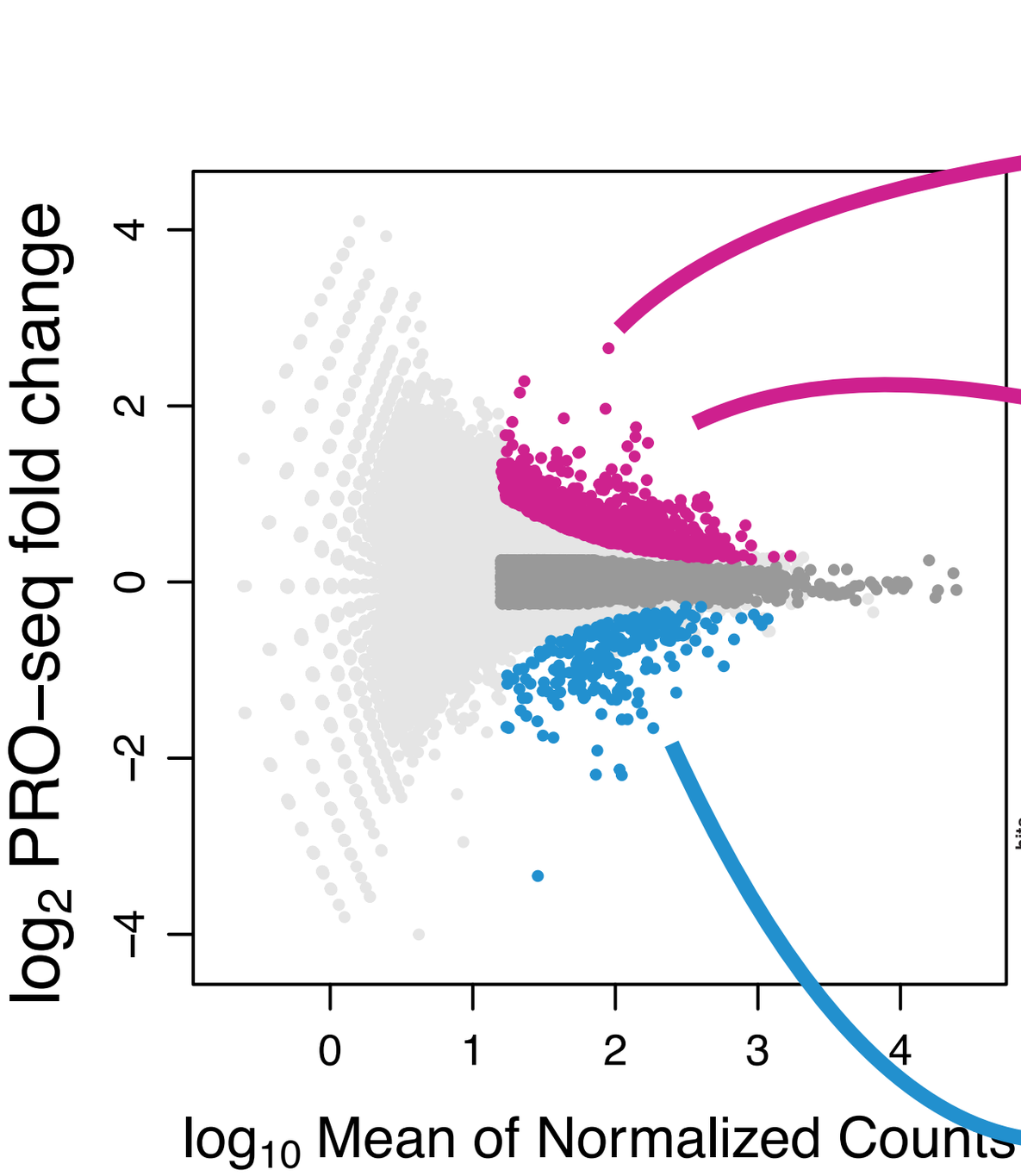


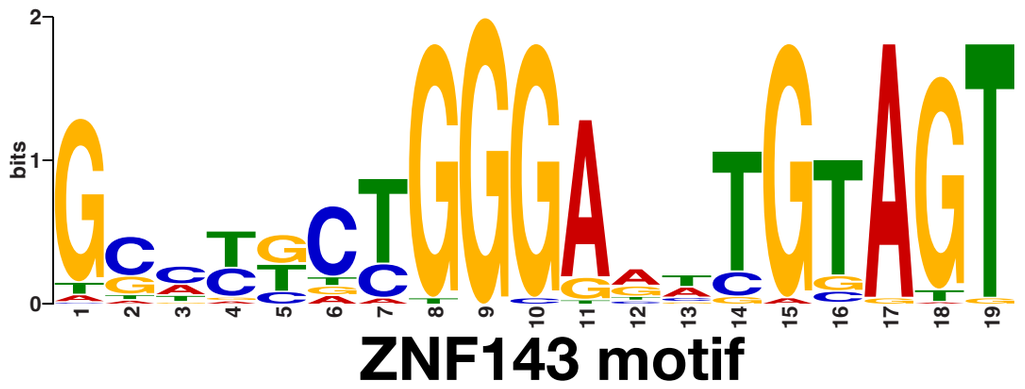
# The ZNF143 motif is found *de novo* at decreased bidirectional transcription

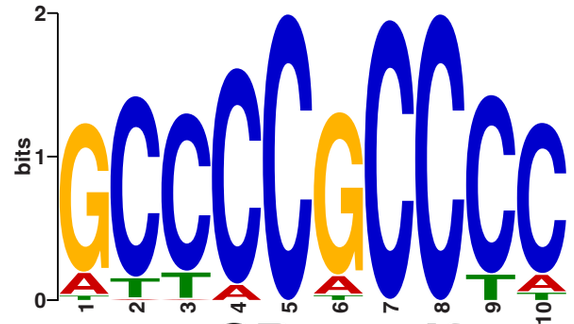


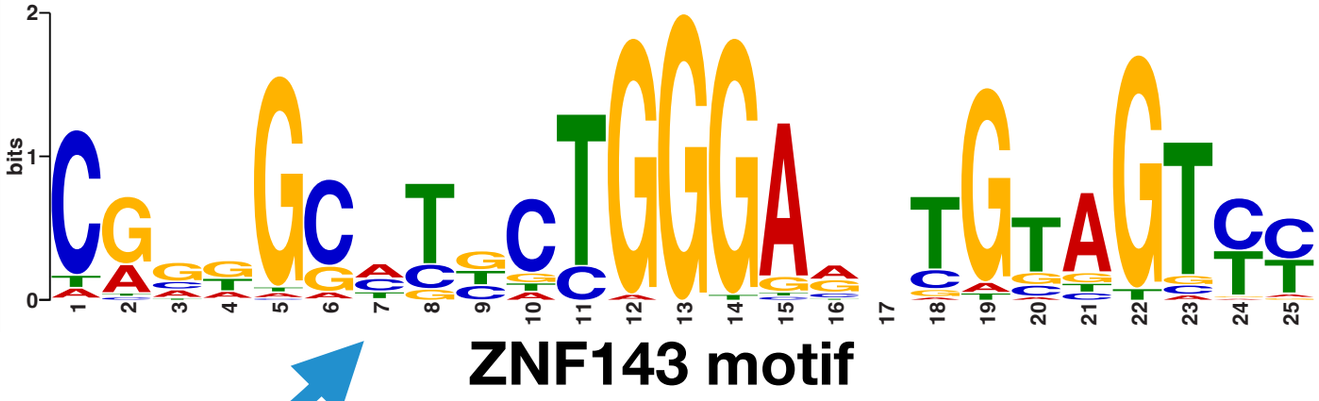



# The ZNF143 motif and others are found at increased bidirectional transcription



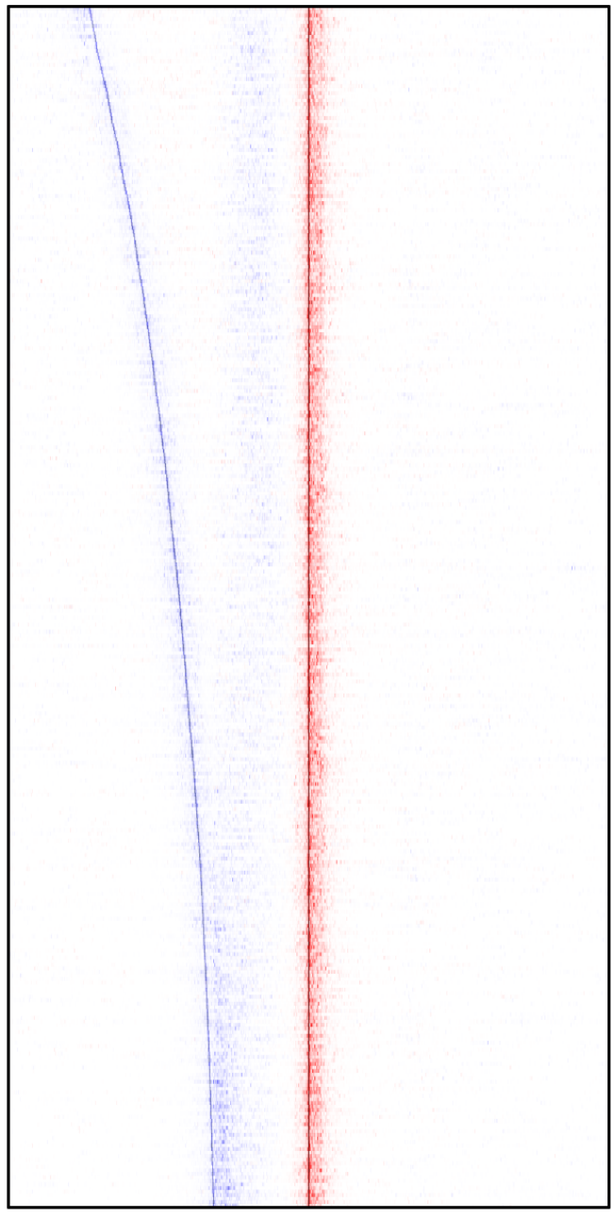
# SP motifs are enriched within promoters

5´ end of PRO-seq reads

Antisense

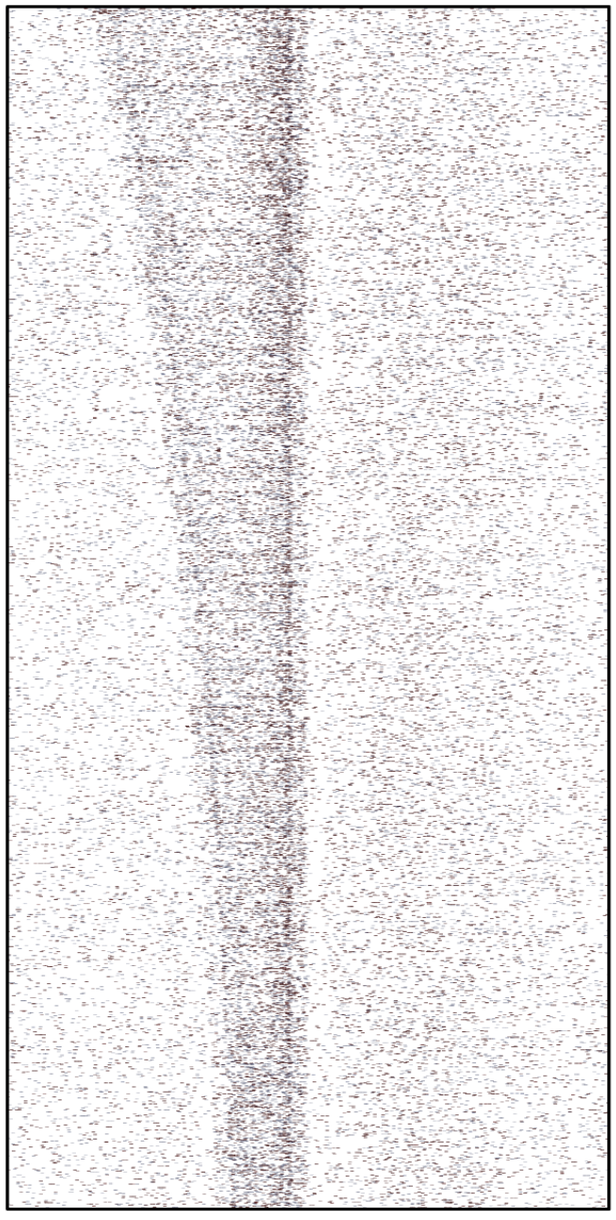
Sense

Bidirectional Promoters



–700 0 700

SP1 motif



–700 0 700