

Transcription Factors

ChIP-seq measures TF binding to DNA.

ChIP-seq also measures histone modification, cofactor, and RNA Polymerase genomic locations—however, their occupancy are a consequence of TF binding.

Mike Guertin

Broad lecture goals:

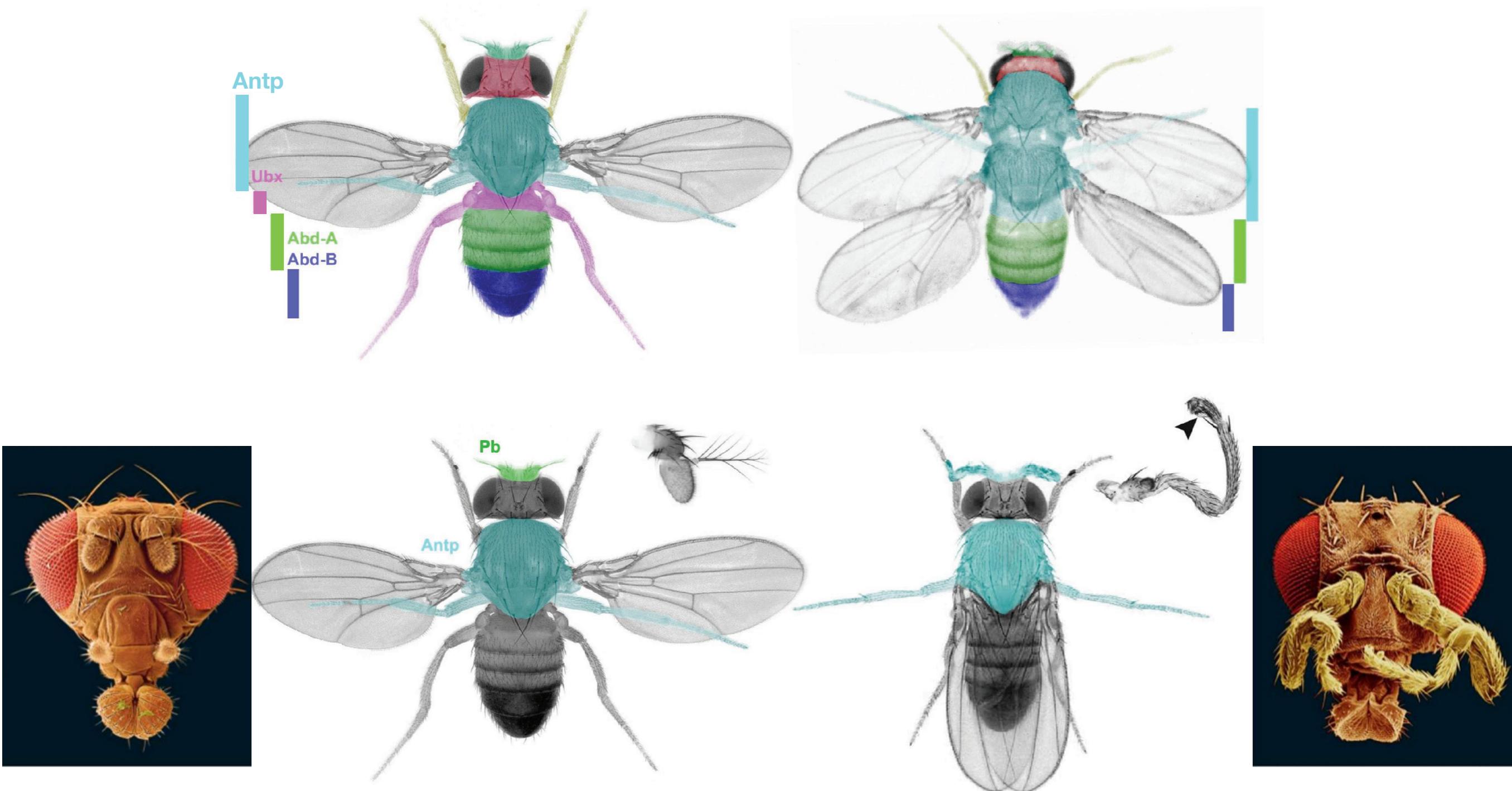
- Convince you of the importance of transcription factors in providing specificity in chromatin biology.
- Introduce classic experiments that defined principles of TF biology and provide references so one can follow up. Note that most molecular biology was interpreted through looking at bands on gels. **As a graduate student, you should aim to be able to take a well-written figure legend and figure and interpret the results.**
- Illustrate the point that biology is continuous, not discrete; relative quantification and controls are important.
- Emphasize the role of question-driven exploratory experiments (screens, molecular genomics, unbiased proximity label transfers, solving structures, etcetera) in defining principles of transcription factor biology.

Transcription dysregulation alters developmental patterning



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

Transcription dysregulation alters developmental patterning



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

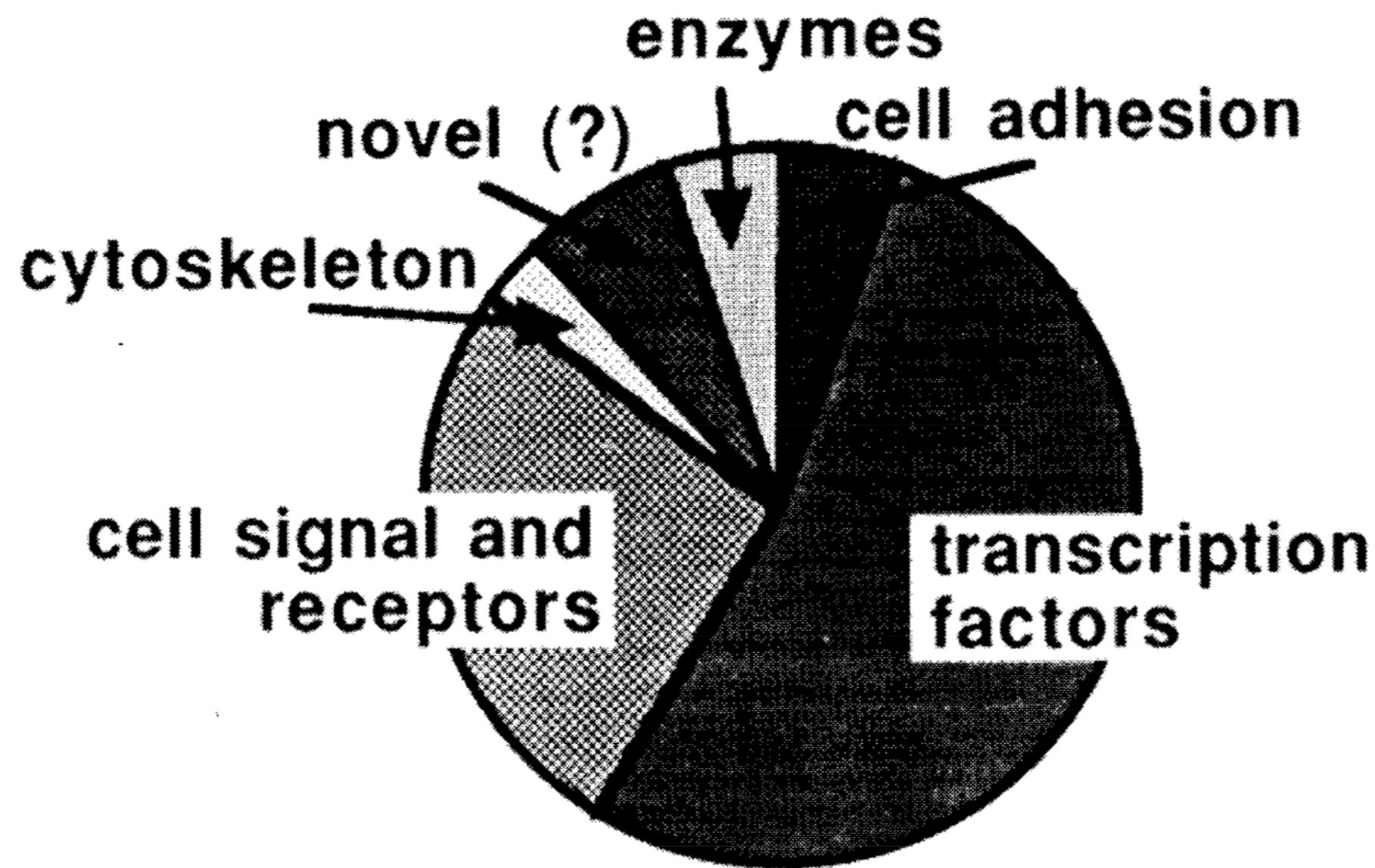
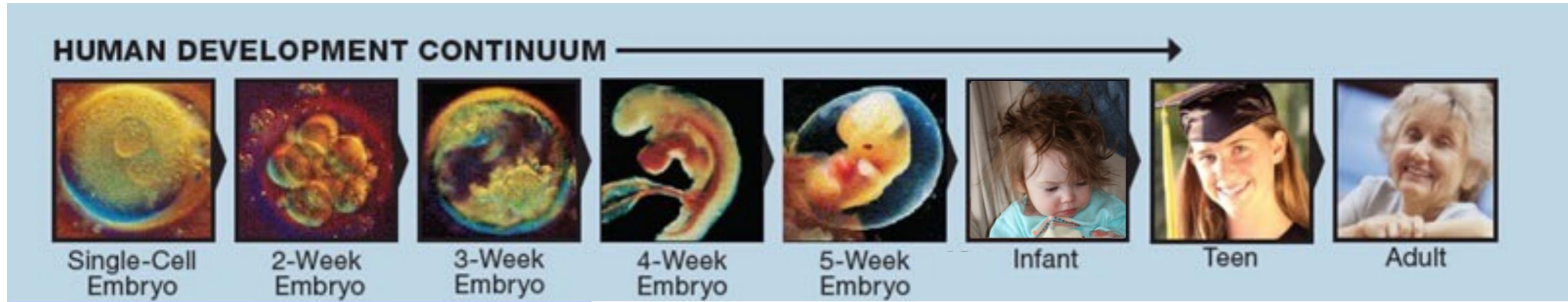
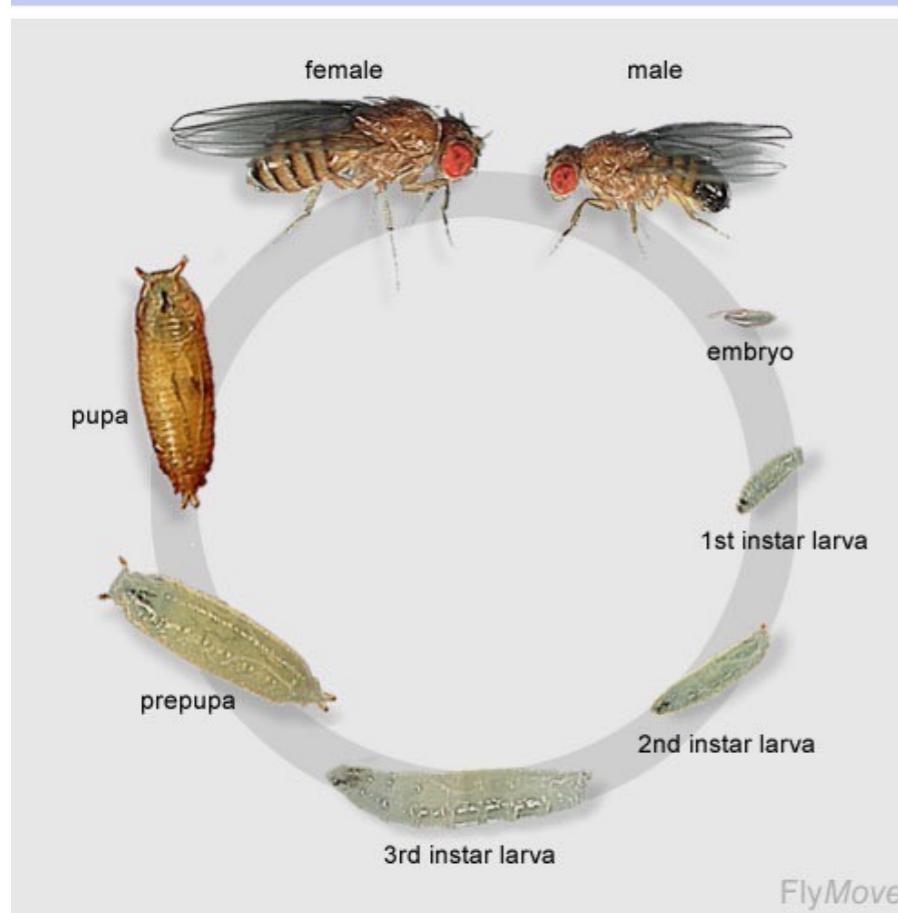


Figure 3. *Cellular Function of Heidelberg Mutations*. Based on the sequence of 75 cloned genes, most of the loci identified in Heidelberg encode transcription factors, or cell signals and receptors.

Transcription control is key in development and homeostasis



The life cycle of *Drosophila melanogaster*



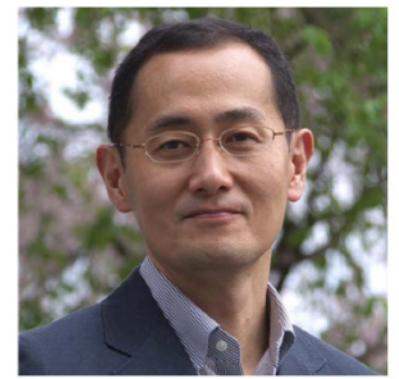
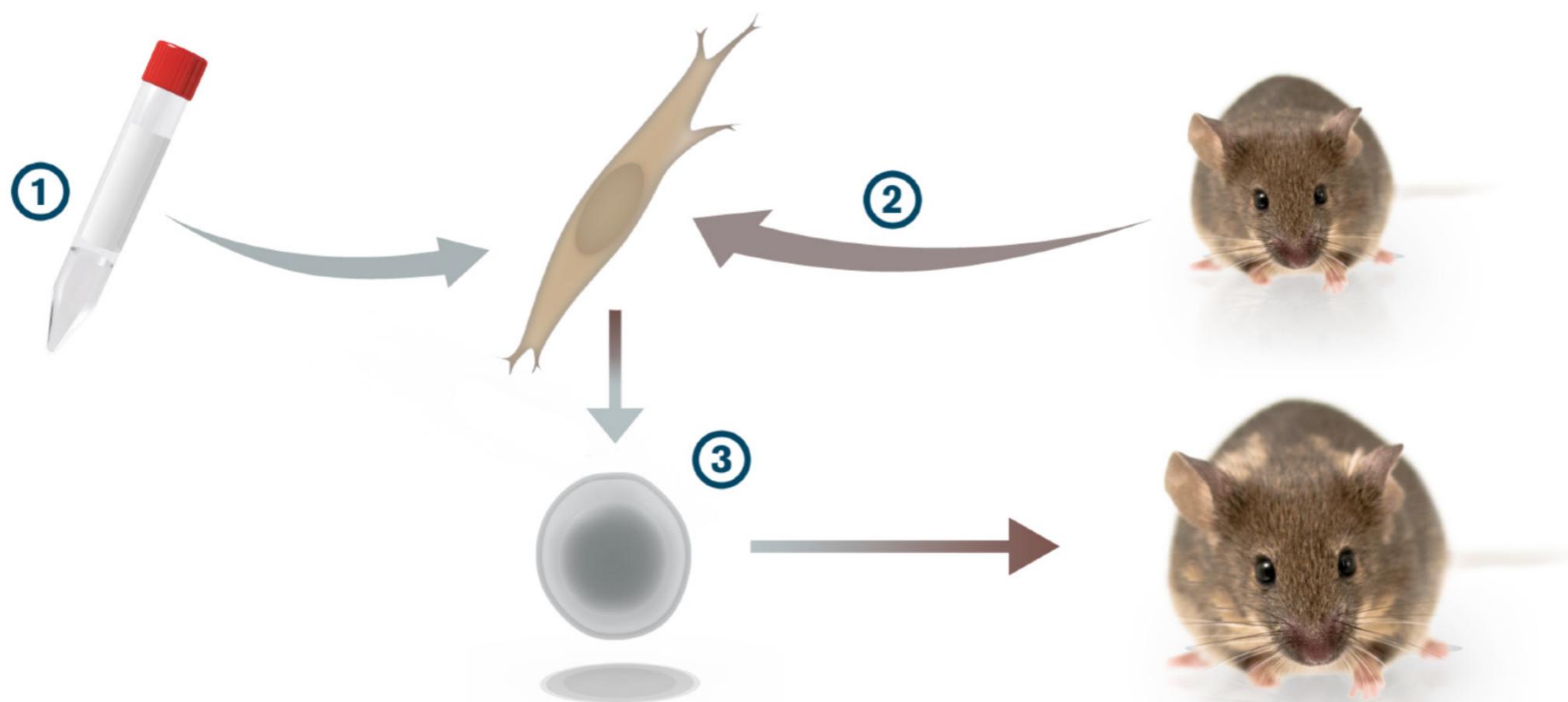
Embryonic cells progress from totipotent to a spectrum of more specialized states.

Much of this developmental regulation starts at transcription.

Cells need to respond to changing nutrients and environments.

Organisms have sophisticated programs of transcription regulation.

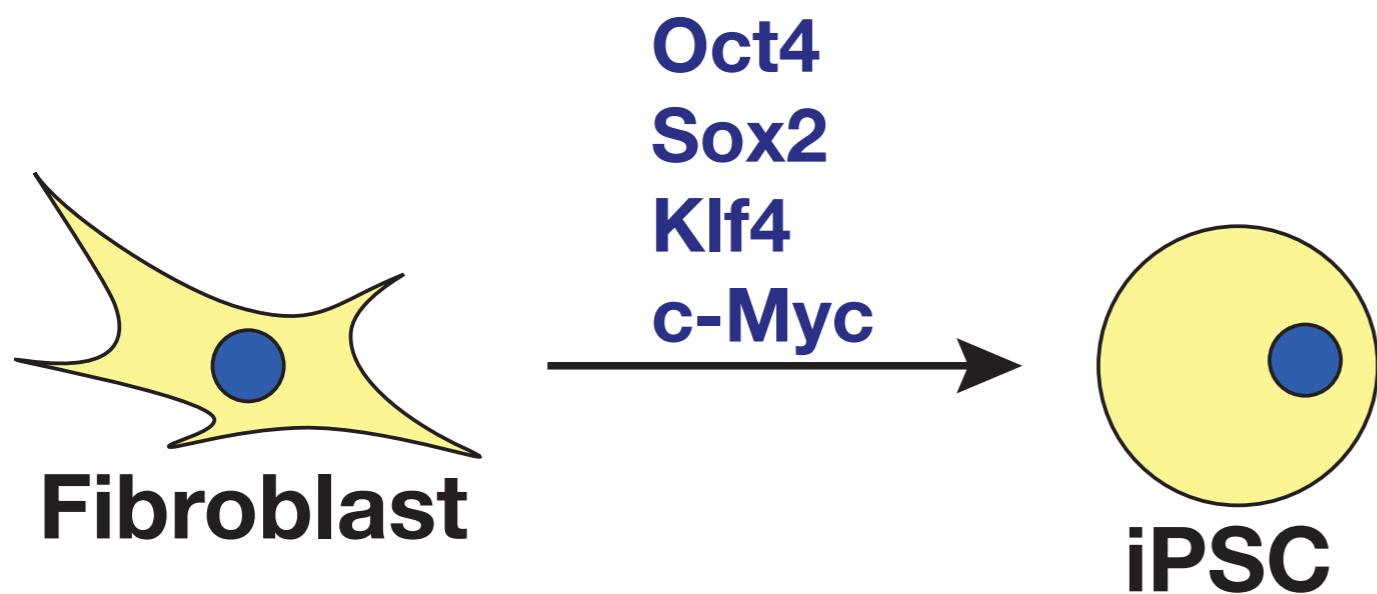
2012 Nobel in Physiology or Medicine: “for the discovery that mature cells can be reprogrammed to become pluripotent”

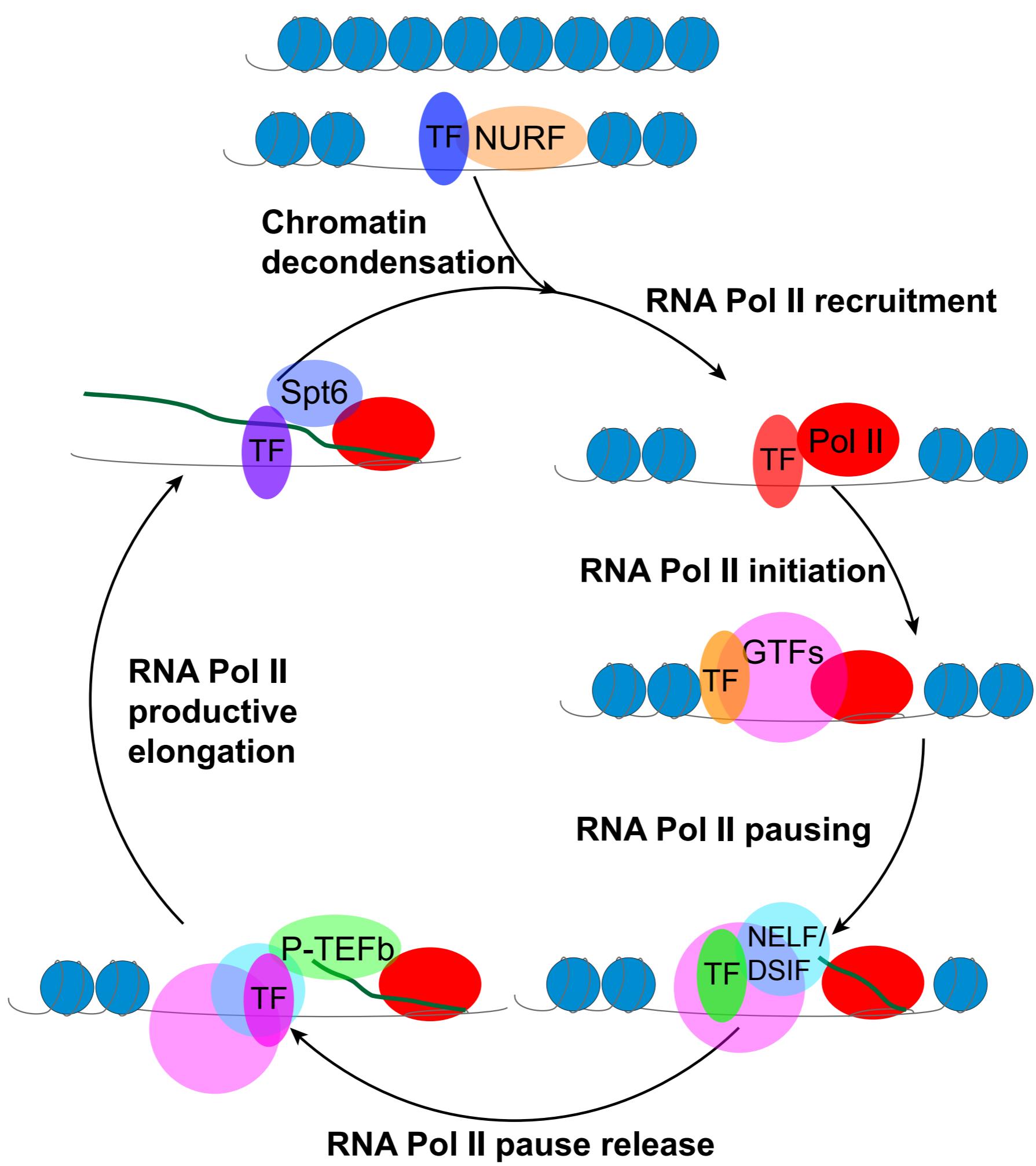


Shinya Yamanaka

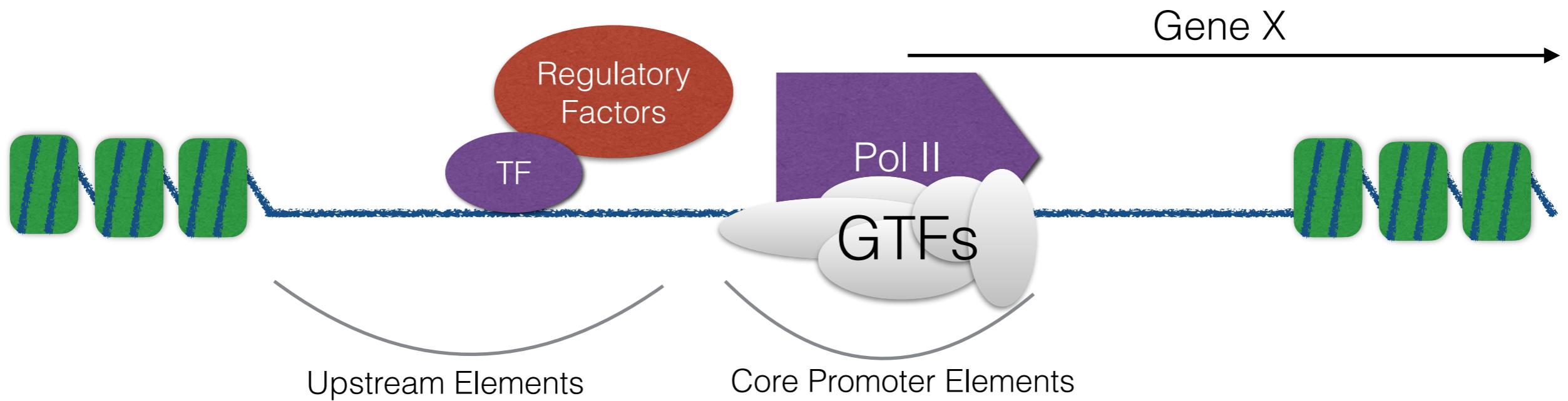
Shinya Yamanaka studied genes that are important for stem cell function. When he transferred four such genes (1) into cells taken from the skin (2), they were reprogrammed into pluripotent stem cells (3) that could develop into all cell types of an adult mouse. He named these cells induced pluripotent stem (iPS) cells.

Activating transcription factors changes cell identity





Transcription Regulation by Transcription Factors (TFs) is determined by DNA sequence



Linker Scanning Mutations of the thymidine kinase gene of HSV

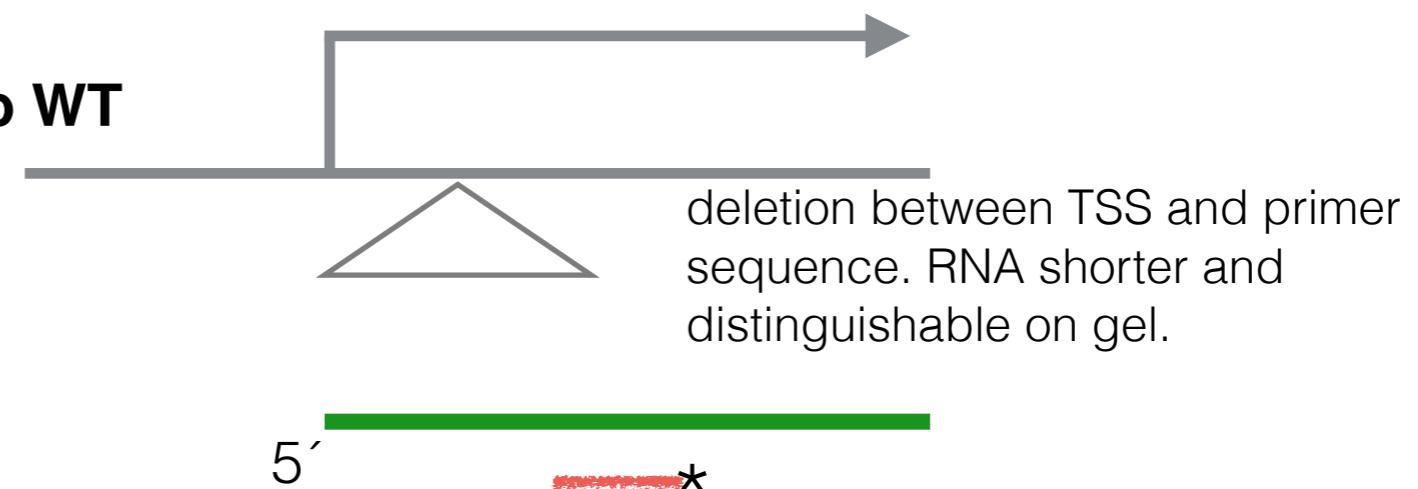
-120 -100 -80 -60 -40 -20 Cap 20
 CTATGATGAC ACAAAACCCG CCCAGCGTCT TGTCAATTGGC GAATTGAAAC ACGCAGATGC AGTCGGGGCG GCGCGGTCCG AGGTCCACTT CGCATATTAA GGTGACGCGT GTGGCCTCGA ACACCGAGCG ACCCTGCAGC WILD TYPE
CCCGATCCG ACAAAACCCG CCCAGCGTCT TGTCAATTGGC GAATTGAAAC ACGCAGATGC AGTCGGGGCG GCGCGGTCCG AGGTCCACTT CGCATATTAA GGTGACGCGT GTGGCCTCGA ACACCGAGCG ACCCTGCAGC LS -119 -109
 CTATGCCGGA **T**C~~CCG~~ACCCG CCCAGCGTCT TGTCAATTGGC GAATTGAAAC ACGCAGATGC AGTCGGGGCG GCGCGGTCCG AGGTCCACTT CGCATATTAA GGTGACGCGT GTGGCCTCGA ACACCGAGCG ACCCTGCAGC LS -115 -105
 CTATGATGAC **CCGATCCG** GCCCAGCGTCT TGTCAATTGGC GAATTGAAAC ACGCAGATGC AGTCGGGGCG GCGCGGTCCG AGGTCCACTT CGCATATTAA GGTGACGCGT GTGGCCTCGA ACACCGAGCG ACCCTGCAGC LS -111 -101
 CTATGATGAC ACAAAAC~~CGA~~**TCCG**AGCGTCT TGTCAATTGGC GAATTGAAAC ACGCAGATGC AGTCGGGGCG GCGCGGTCCG AGGTCCACTT CGCATATTAA GGTGACGCGT GTGGCCTCGA ACACCGAGCG ACCCTGCAGC LS -105 -95
 CTATGATGAC ACAAAACCCG CCCAGC~~CGA~~**TCCG**TTGGC GAATTGAAAC ACGCAGATGC AGTCGGGGCG GCGCGGTCCG AGGTCCACTT CGCATATTAA GGTGACGCGT GTGGCCTCGA ACACCGAGCG ACCCTGCAGC LS -95 -85
 CTATGATGAC ACAAAACCCG CCCAGCGTCT TGTCAATTGGC **CCCGATCCG** ACGCAGATGC AGTCGGGGCG GCGCGGTCCG AGGTCCACTT CGCATATTAA GGTGACGCGT GTGGCCTCGA ACACCGAGCG ACCCTGCAGC LS -84 -74
 CTATGATGAC ACAAAACCCG CCCAGCGTCT TGTCAATTGGC **CCCGATCCG** ACGCAGATGC AGTCGGGGCG GCGCGGTCCG AGGTCCACTT CGCATATTAA GGTGACGCGT GTGGCCTCGA ACACCGAGCG ACCCTGCAGC LS -80 -70
 CTATGATGAC ACAAAACCCG CCCAGCGTCT TGTCAATTGGC **GCCCGATCCG** CCGCAGATGC AGTCGGGGCG GCGCGGTCCG AGGTCCACTT CGCATATTAA GGTGACGCGT GTGGCCTCGA ACACCGAGCG ACCCTGCAGC LS -79 -69
 CTATGATGAC ACAAAACCCG CCCAGCGTCT TGTCAATTGGC GAATTGAAAC **CCGATCCG** AGTCGGGGCG GCGCGGTCCG AGGTCCACTT CGCATATTAA GGTGACGCGT GTGGCCTCGA ACACCGAGCG ACCCTGCAGC LS -70 -61
 CTATGATGAC ACAAAACCCG CCCAGCGTCT TGTCAATTGGC GAATTGAAAC ACGCAGATGC **ACCGATC** CG GCGCGGTCCG AGGTCCACTT CGCATATTAA GGTGACGCGT GTGGCCTCGA ACACCGAGCG ACCCTGCAGC LS -59 -49
 CTATGATGAC ACAAAACCCG CCCAGCGTCT TGTCAATTGGC GAATTGAAAC ACGCAGATGC AGTC~~CCG~~**ATCCG**GGTCCG AGGTCCACTT CGCATATTAA GGTGACGCGT GTGGCCTCGA ACACCGAGCG ACCCTGCAGC LS -56 -46
 CTATGATGAC ACAAAACCCG CCCAGCGTCT TGTCAATTGGC GAATTGAAAC ACGCAGATGC AGTCGGGGCG **CCGCGATC**CGGTCCTT CGCATATTAA GGTGACGCGT GTGGCCTCGA ACACCGAGCG ACCCTGCAGC LS -47 -37
 CTATGATGAC ACAAAACCCG CCCAGCGTCT TGTCAATTGGC GAATTGAAAC ACGCAGATGC AGTCGGGGCG **CCGATCCG**CGGTCCTT CGCATATTAA GGTGACGCGT GTGGCCTCGA ACACCGAGCG ACCCTGCAGC LS -42 -32
 CTATGATGAC ACAAAACCCG CCCAGCGTCT TGTCAATTGGC GAATTGAAAC ACGCAGATGC AGTCGGGGCG GCGCGGTCCG~~CCGATCCG~~TT CGCATATTAA GGTGACGCGT GTGGCCTCGA ACACCGAGCG ACCCTGCAGC LS -29 -18
 CTATGATGAC ACAAAACCCG CCCAGCGTCT TGTCAATTGGC GAATTGAAAC ACGCAGATGC AGTCGGGGCG GCGCGGTCCG AGGTCCACTT **CCGCGATCC** GGTGACGCGT GTGGCCTCGA ACACCGAGCG ACCCTGCAGC LS -21 -12
 CTATGATGAC ACAAAACCCG CCCAGCGTCT TGTCAATTGGC GAATTGAAAC ACGCAGATGC AGTCGGGGCG GCGCGGTCCG AGGTCCACTT CGCATATTAA GGTG~~CCGATCCG~~GGTGCCTCGA ACACCGAGCG ACCCTGCAGC LS -18 -6
 CTATGATGAC ACAAAACCCG CCCAGCGTCT TGTCAATTGGC GAATTGAAAC ACGCAGATGC AGTCGGGGCG GCGCGGTCCG AGGTCCACTT CGCATATTAA GGTGACGCGT GTG~~CCGATCCG~~GGTGCAGC ACCCTGCAGC LS -7 +3
 CTATGATGAC ACAAAACCCG CCCAGCGTCT TGTCAATTGGC GAATTGAAAC ACGCAGATGC AGTCGGGGCG GCGCGGTCCG AGGTCCACTT CGCATATTAA GGTGACGCGT GTGGCCTCGA ACAC~~CCG~~**ATCCG**GGTGCAGC LS +5 +15

Clusters of point mutants are generated at the point of joining 5' and 3' deletions, where linker sequence substitutes for tk sequence.

Assay expression of tk promoter mutants

Plasmid DNAs

Inject mutant DNA into frog oocyte nuclei, include **pseudo WT standard** as internal control



Isolate RNA

Primer Extension Assay

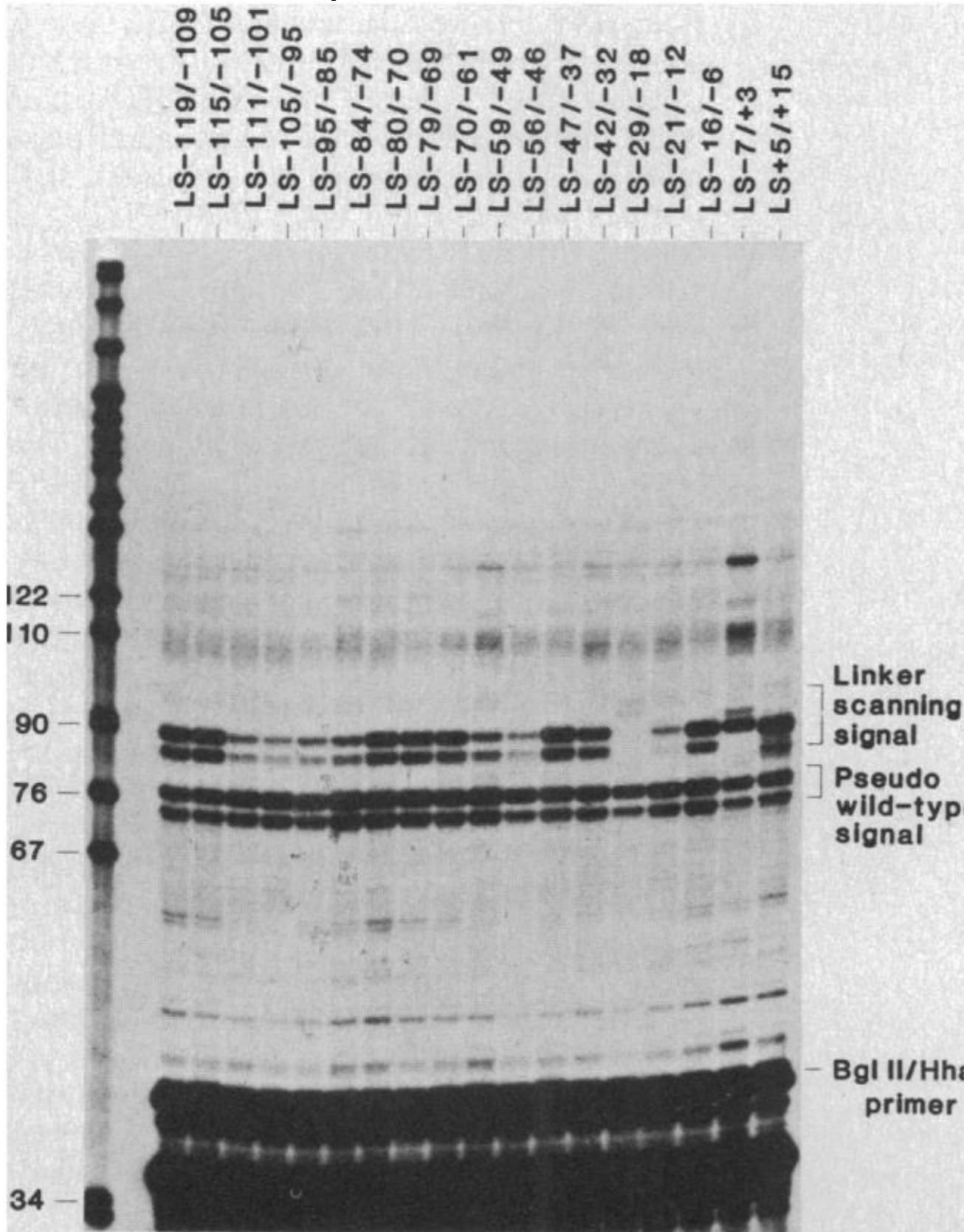
Gel Electrophoresis



tk promoter mutant

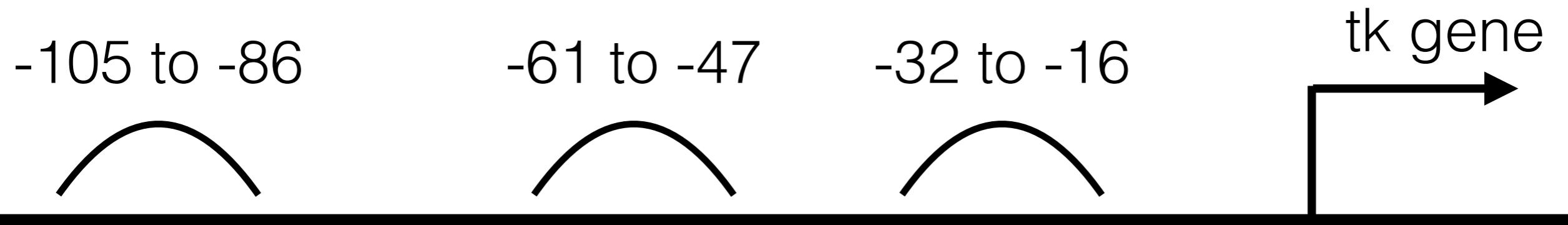
WT tk promoter; deletion of gene body

Expression data from Linker Scanning Mutants



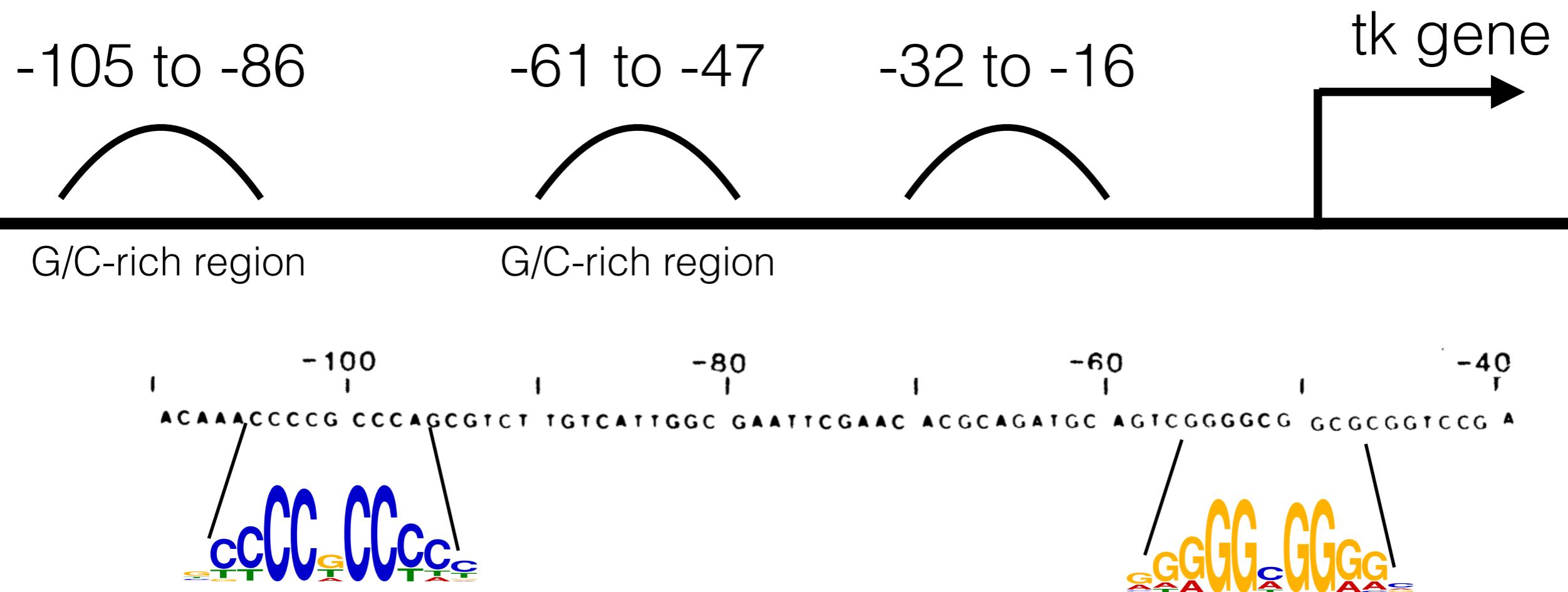
Short discontinuous regions of sequence are critical for basal expression.

Three promoter regions are critical for basal expression

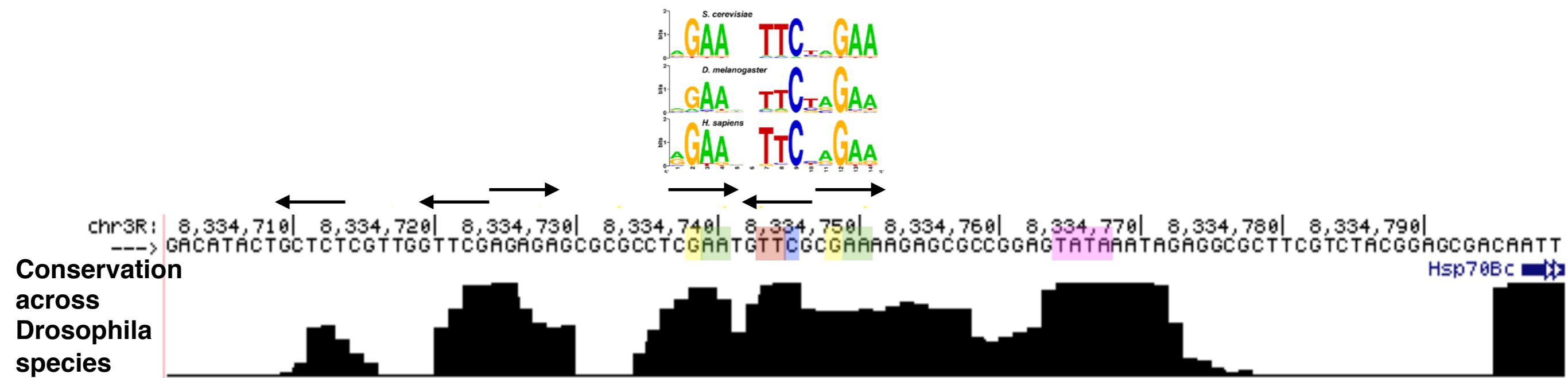


TATA element located
approximately 30bp
upstream of the TSS

Three promoter regions are critical for basal expression



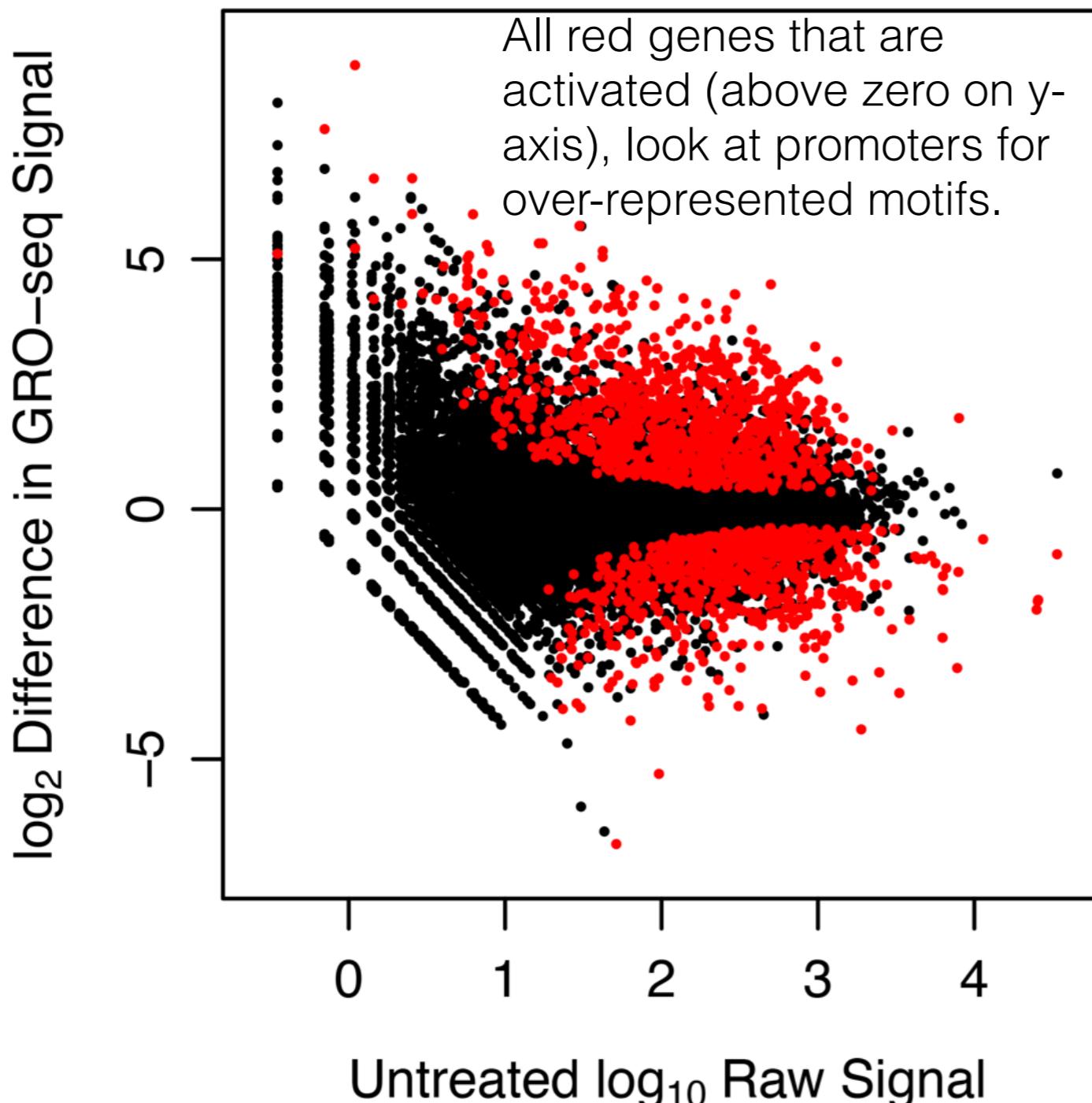
Evolutionary conservation and comparative genomics can identify crucial elements



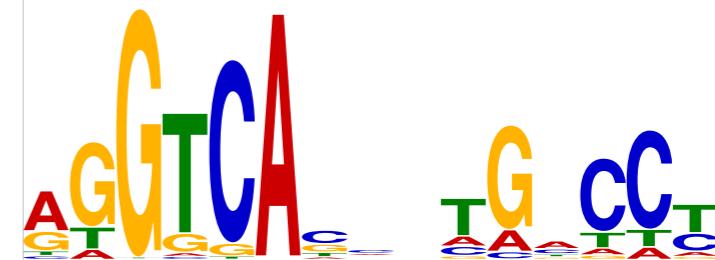
Significant changes in nascent transcription upon estrogen treatment in breast cancer cells

Collect the sequences of multiple (co-regulated) promoters within a species, search for common sequence motifs

Sustained Changes at both 10min and 40min



de novo motif analysis using MEME (or the alike) identifies the Estrogen Response Element, the known target of the Estrogen Receptor.

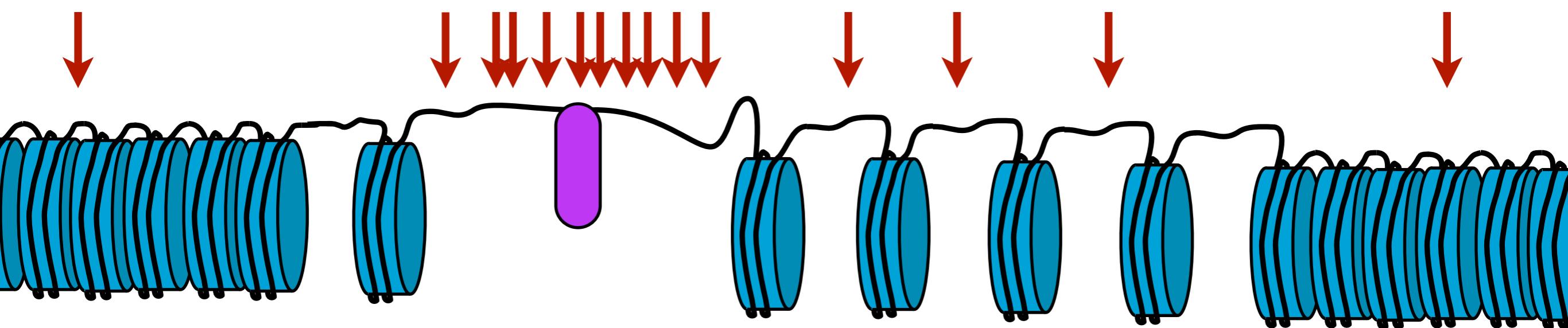


Note that not all regulatory elements bound by TFs are within promoters.

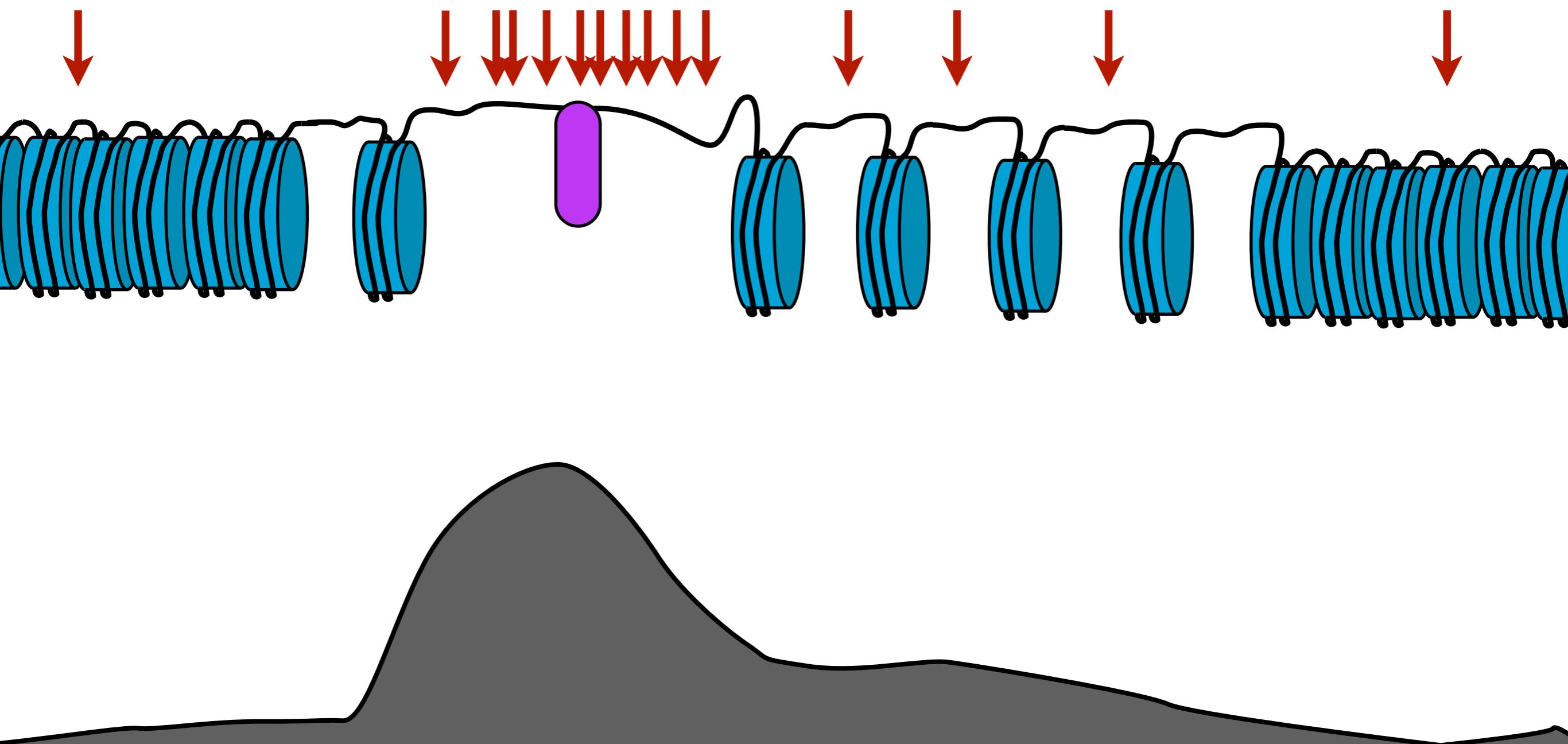
Identify All Active Regulatory Elements in a Cell Type: Enzyme Hypersensitivity (DNase-seq & ATAC-seq)

- A general measure of chromatin structure.
 - Factor/species-general
 - Changes in enzyme hypersensitivity landscape after drug treatment or throughout development can be used to identify novel regulatory elements and factors
 - Generally unbiased, but challenging to deconvolve
 - **TFs controlling chromatin landscape can be inferred from the data**

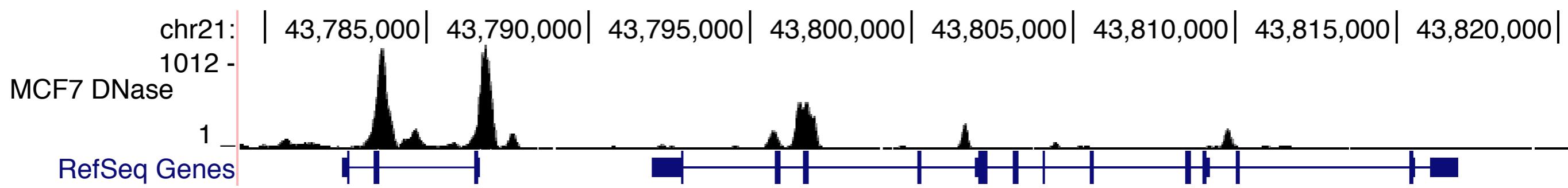
Enzyme Hypersensitivity



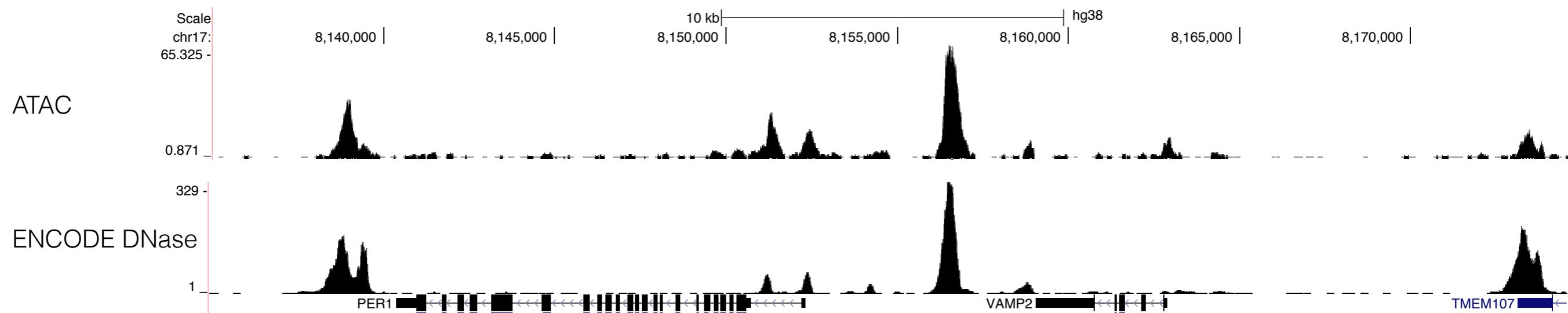
Enzyme Hypersensitivity



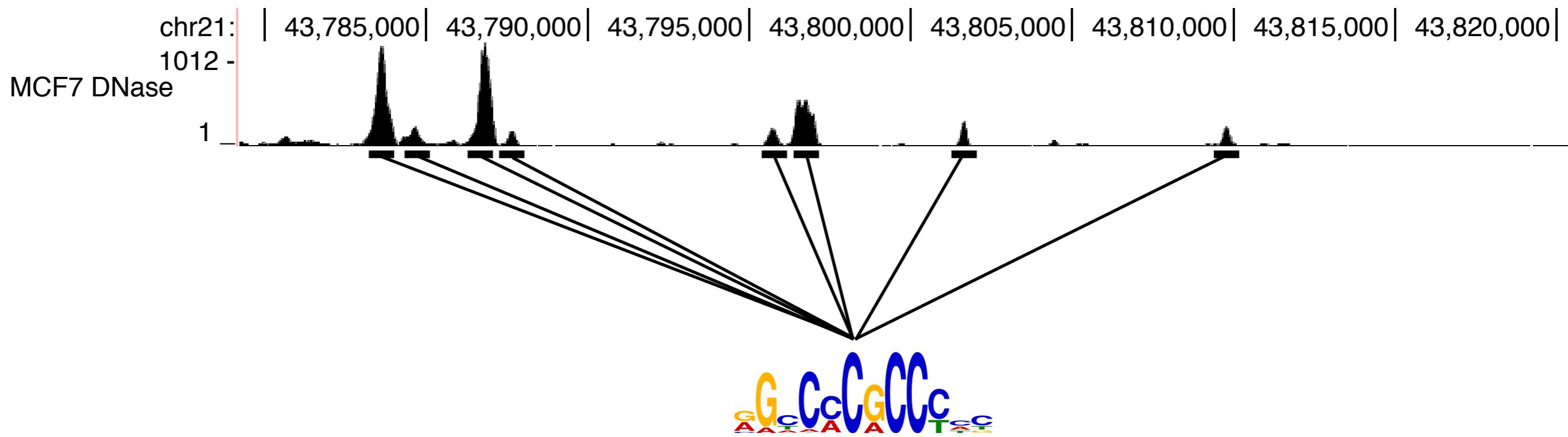
DNase-seq Data



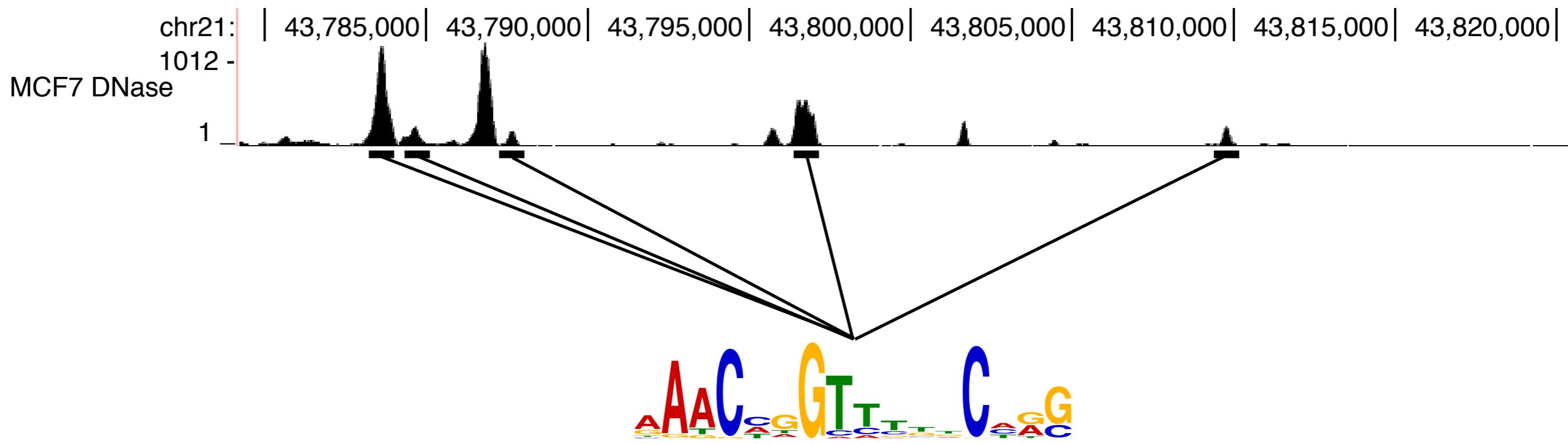
ATAC-seq vs. DNase-seq



DNase/ATAC identifies a repertoire of TF motifs

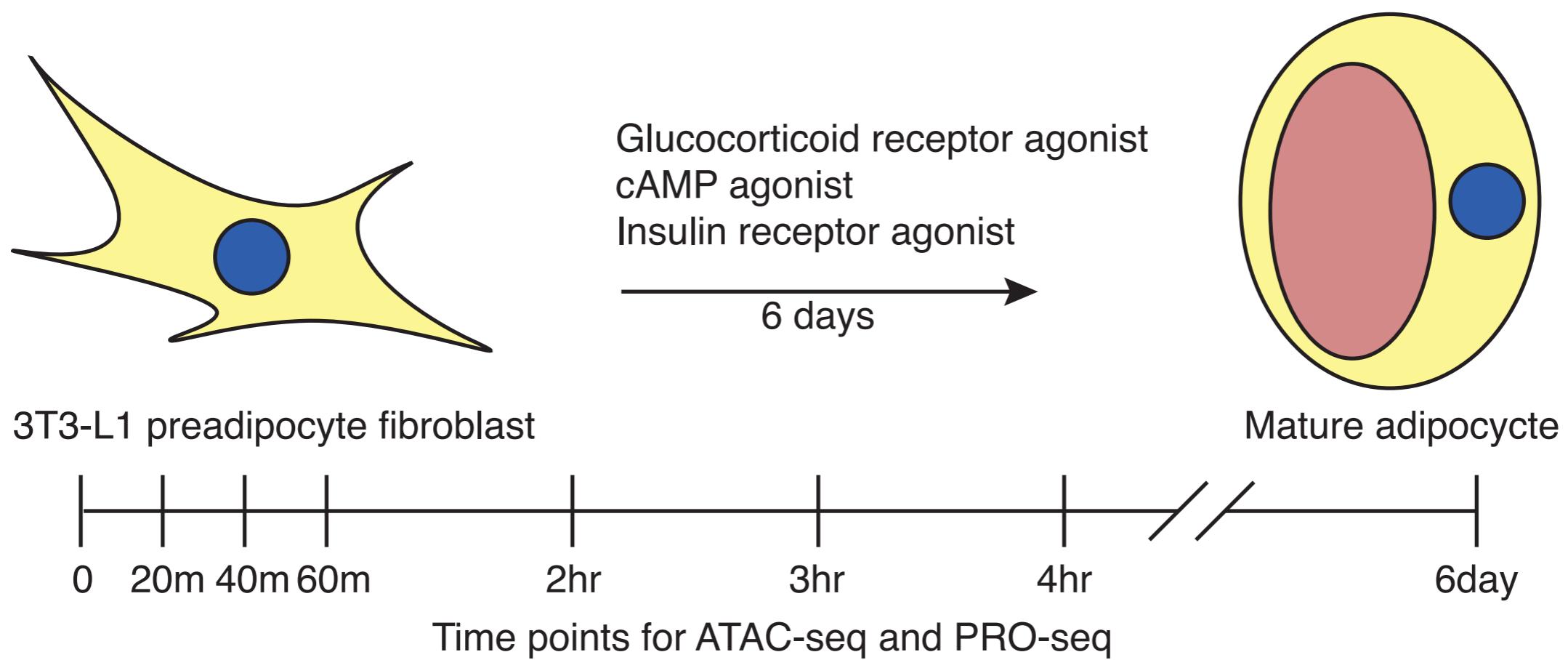


DNase/ATAC identifies a repertoire of TF motifs

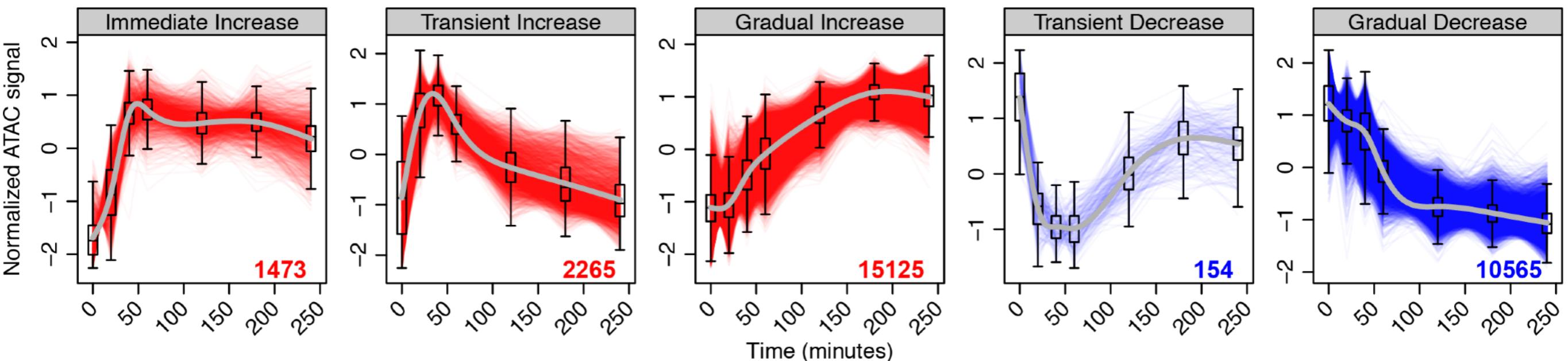


Identify sequence elements at hypersensitive site using iterative de novo motif analysis

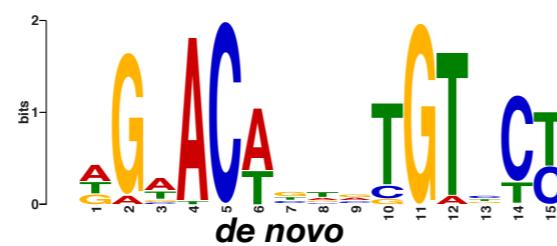
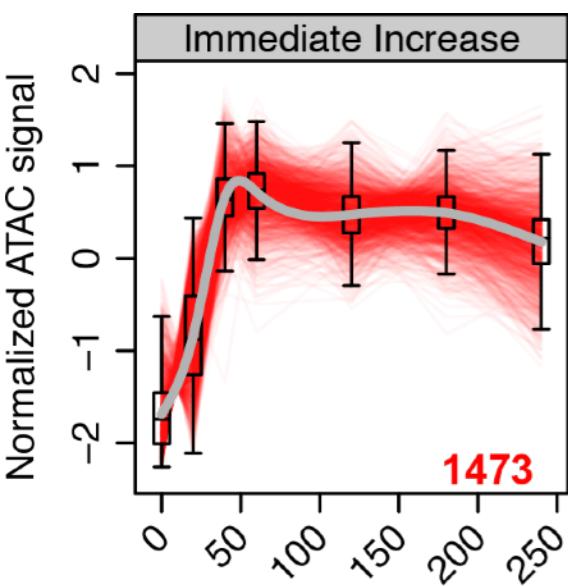
Experimental Design



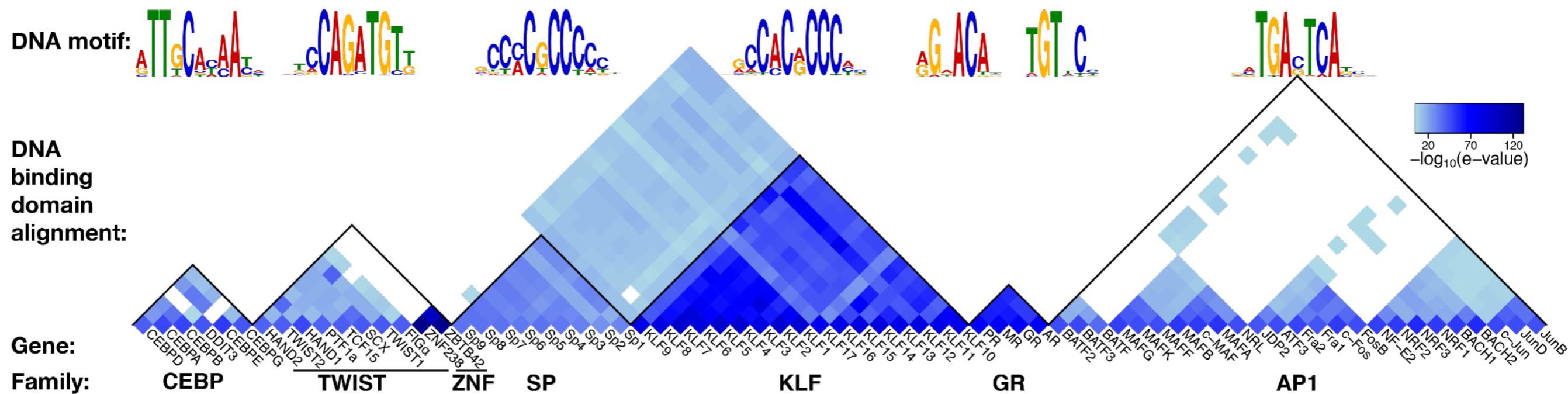
ATAC-seq peaks have distinct accessibility kinetics



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



14 TF-family motifs (top 6 shown) drive early adipogenesis changes in accessibility



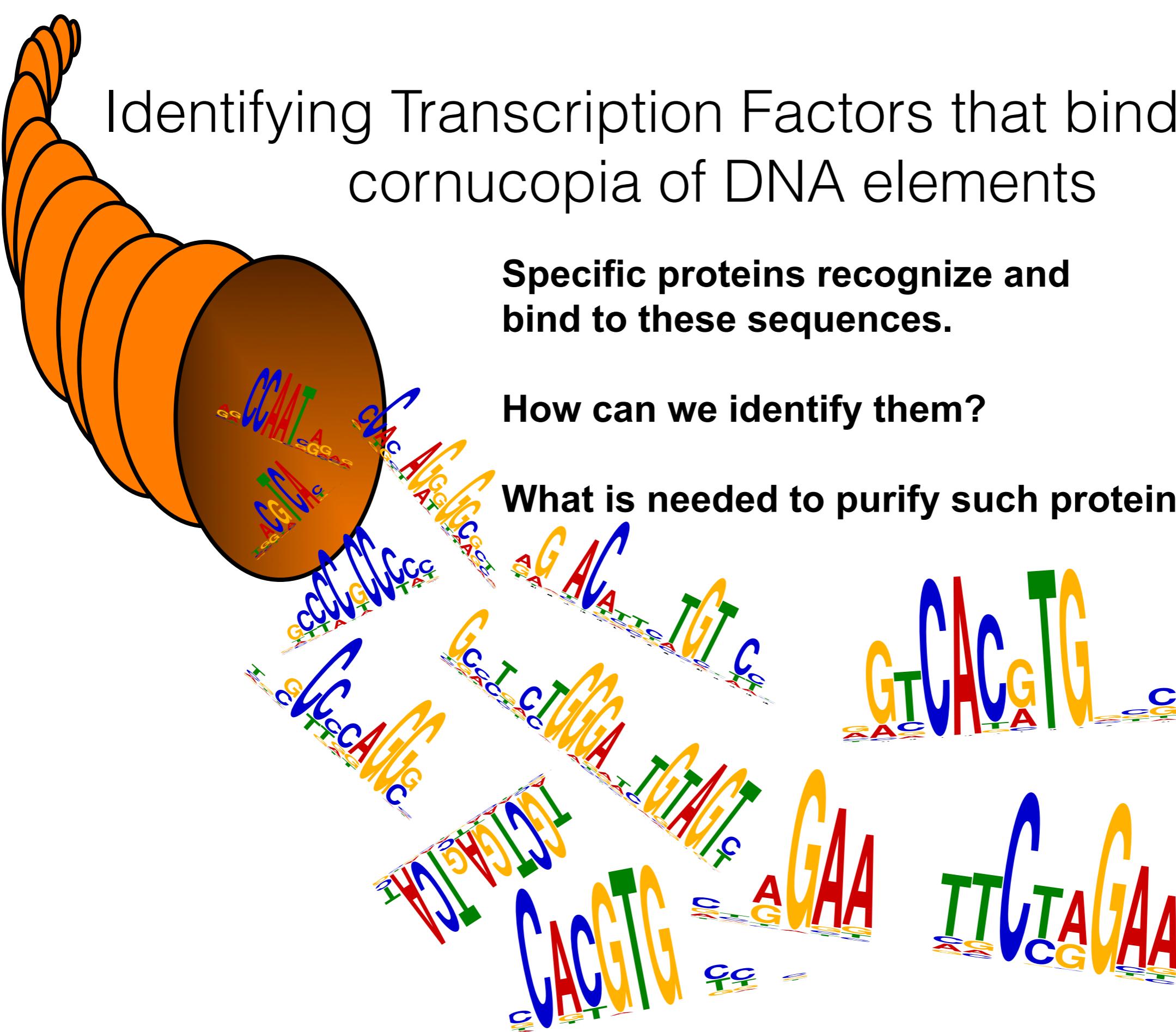
Paralogous TF DBD families that recognize each motif

Identifying Transcription Factors that bind to the cornucopia of DNA elements

Specific proteins recognize and bind to these sequences.

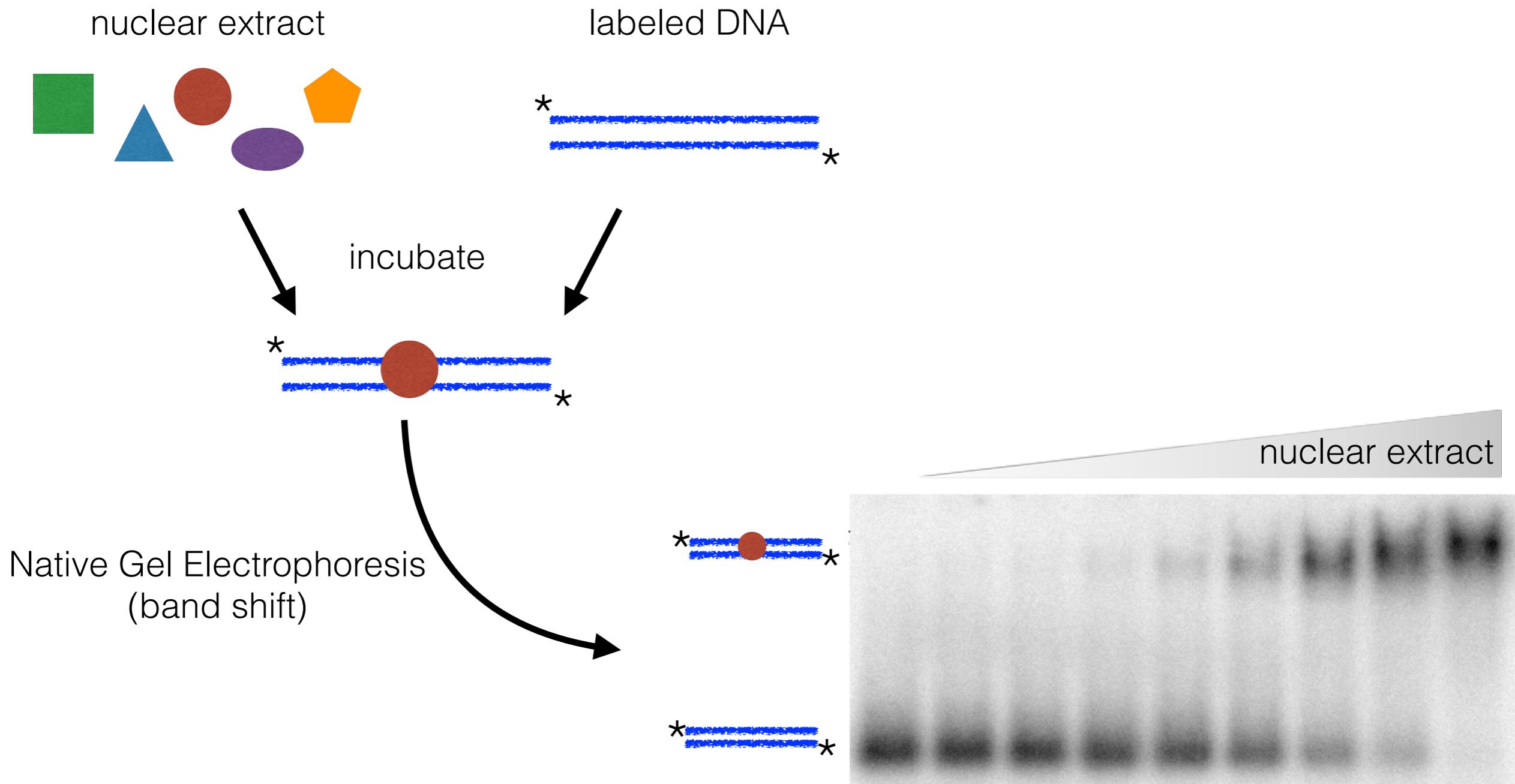
How can we identify them?

What is needed to purify such proteins?

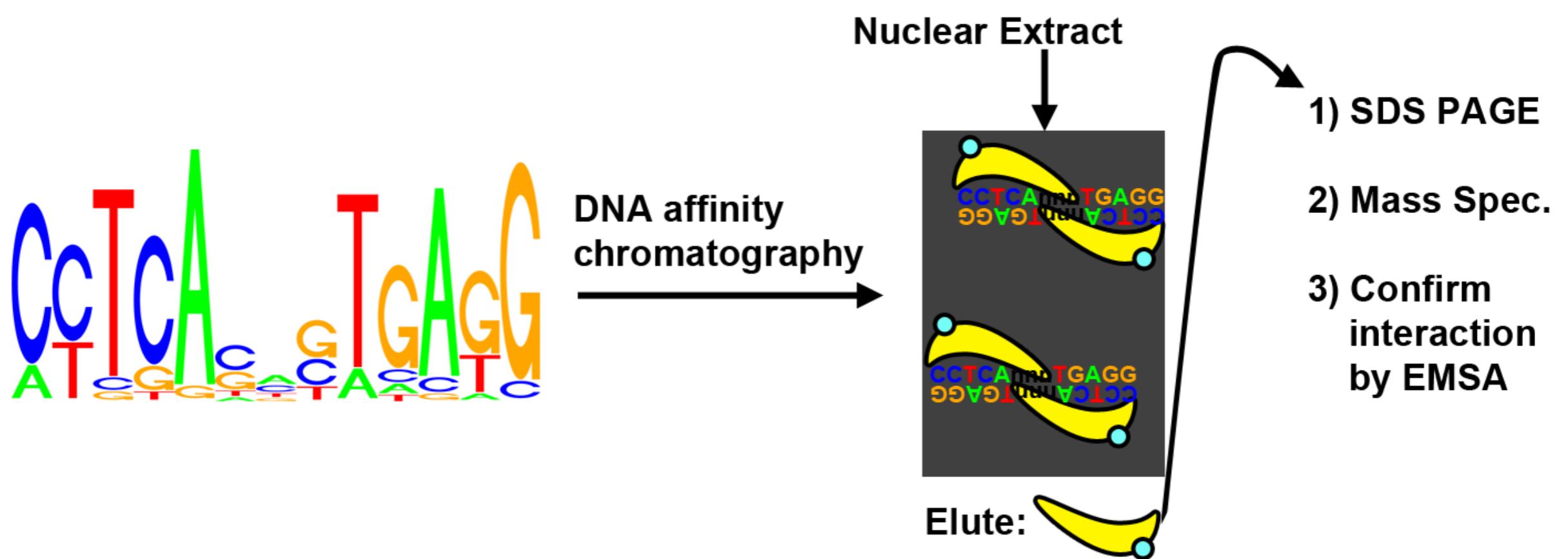


Knowing specific sequences helps create both an assay and tool for purifying.

Electrophoretic Mobility Shift Assay (EMSA) detect DNA binding factors



Purification of sequence-specific DNA-binding proteins



Order oligos with modest variants of your consensus sequence (include random flanking DNA). Biotinylate the ends of the duplexed DNA, bind to streptavidin beads/column, elute, compare eluate to nuclear extract by PAGE, and mass spec.

Summary: Part I

- Transcription and its regulation is specified by short DNA sequence elements.
- These elements interact with particular transcription factors.
- See Lambert et. al., The Human Transcription Factors, Cell 2018 for a review of TF/DNA binding