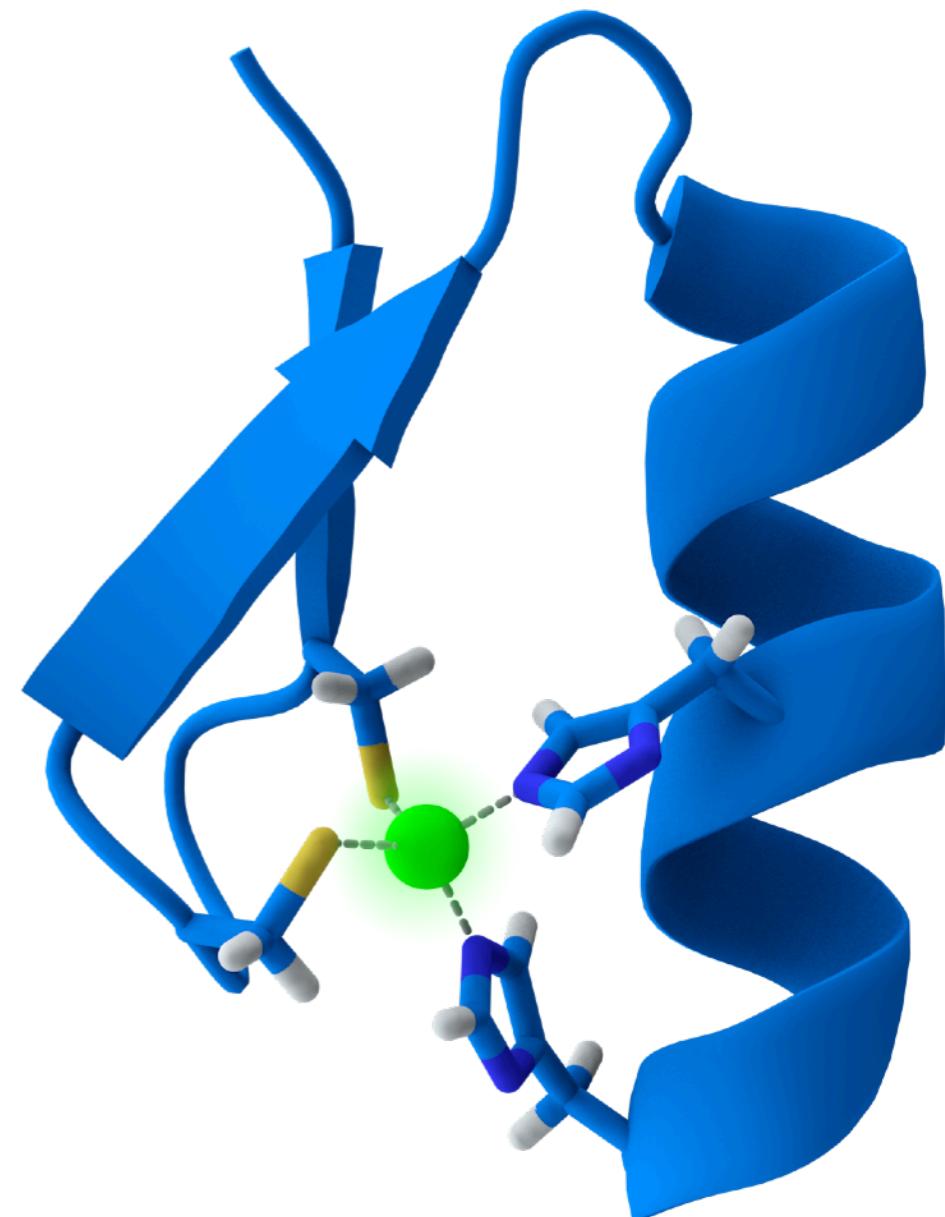


What are the features of protein domains that bind DNA?

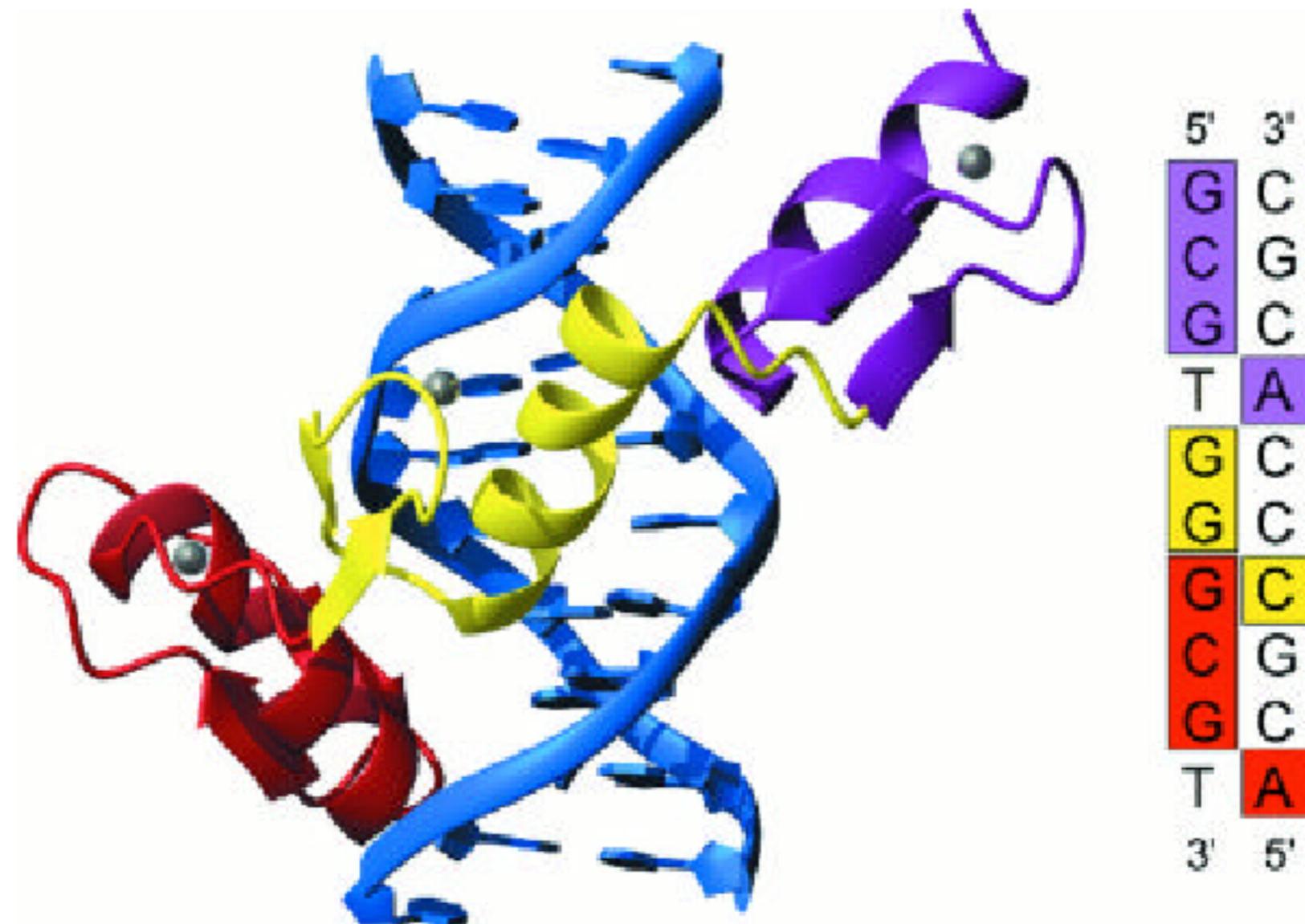
Zinc-containing DNA binding domains

(Zinc is coordinated with a combination of Cysteine and Histidine residues)

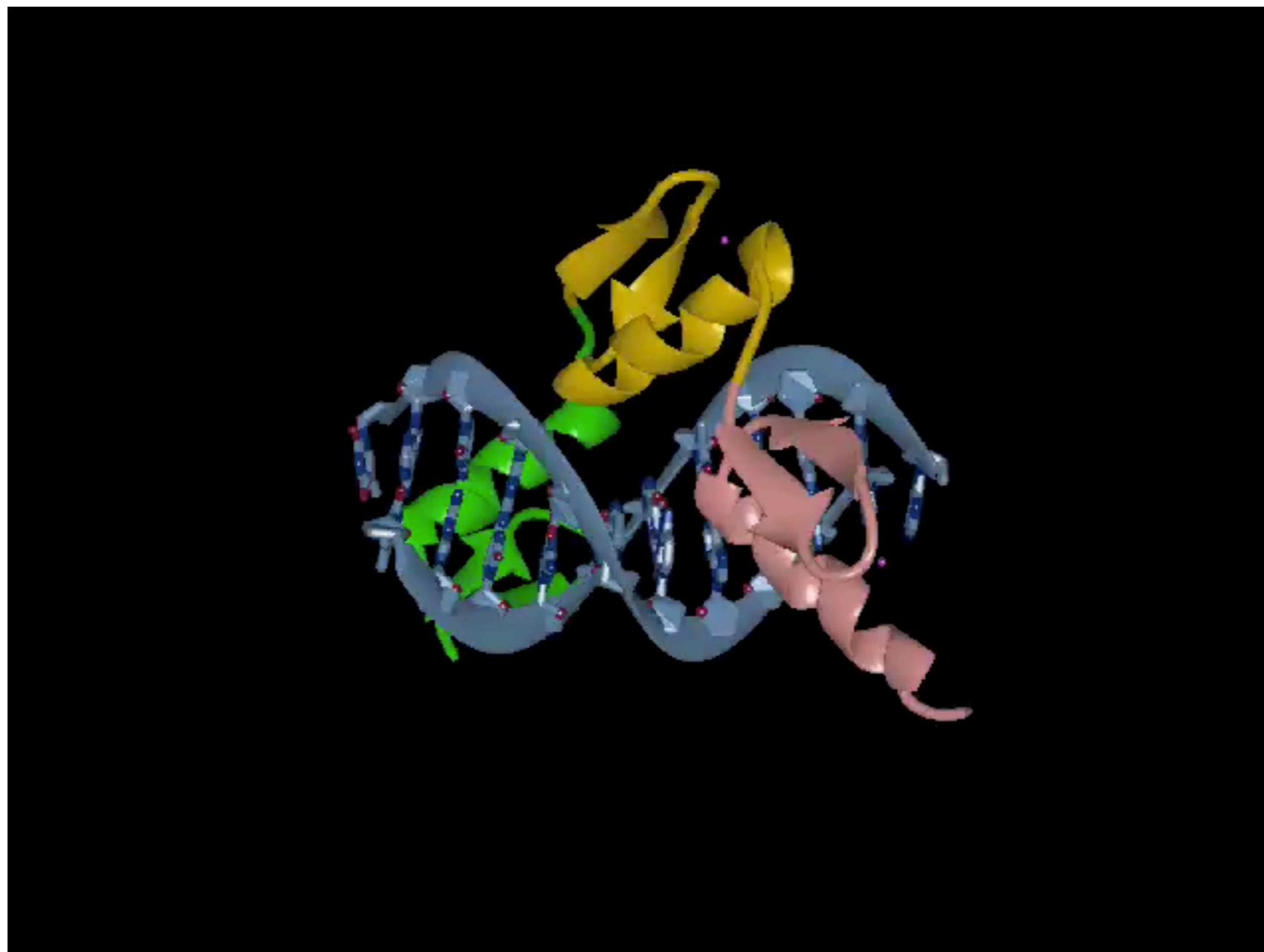
- Cys₂His₂ Class:
 - Beta-Beta-Alpha fold
 - Cys-X₂₋₄-Cys-X₁₂-His-X₃₋₅-His



Three zinc fingers of Zif268 follow the major groove with each fingers occupying ~3 bp.



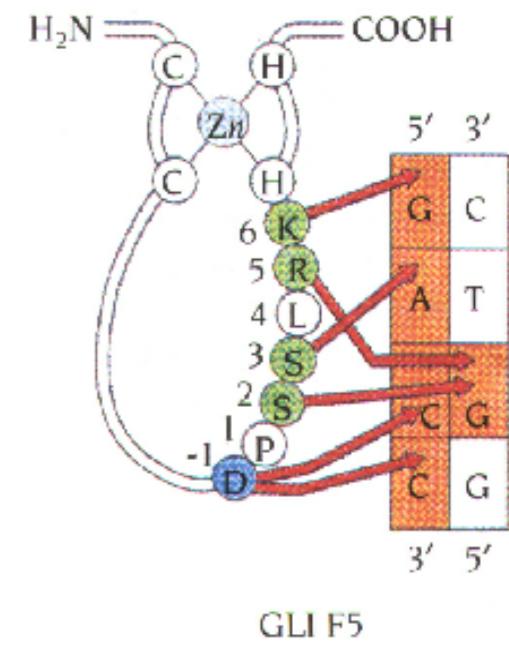
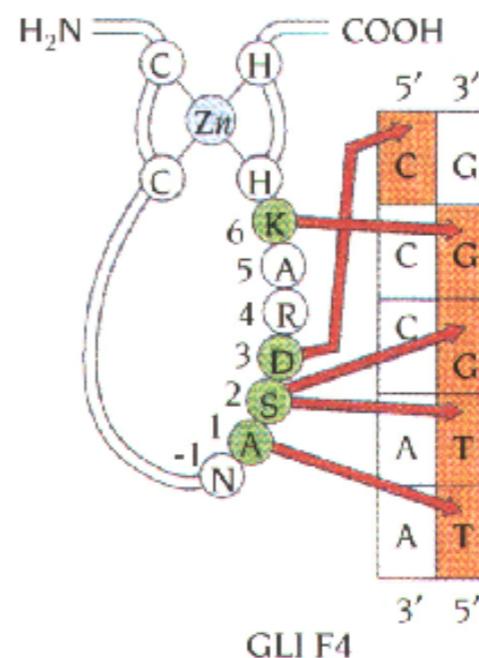
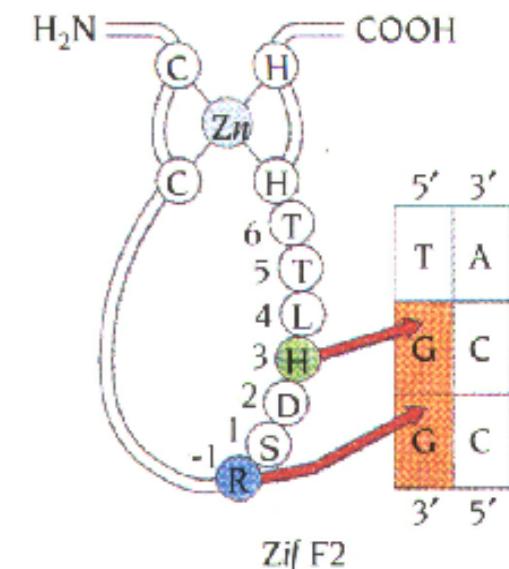
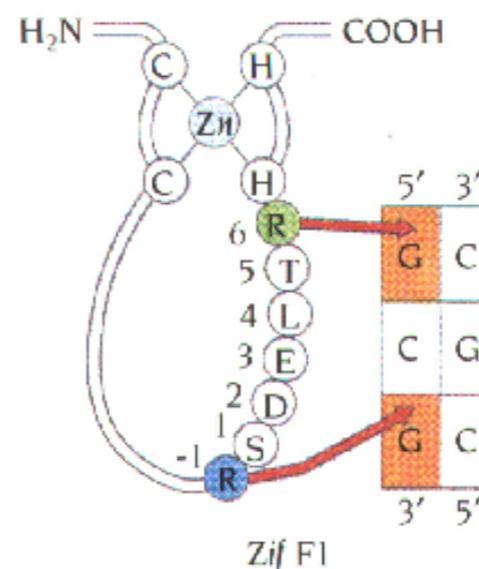
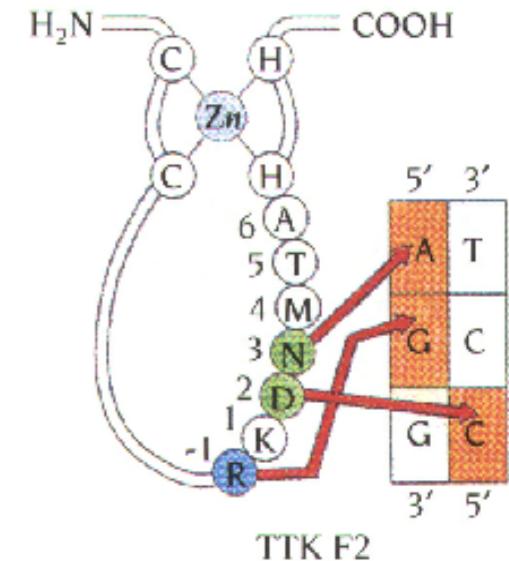
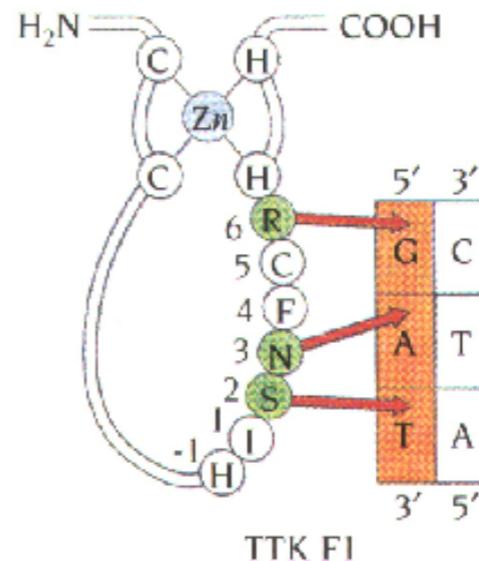
Song Tan lab: Zif268



http://www.personal.psu.edu/sxt30/movies/zif268dna_h264.mov

Comparing Zn Fingers

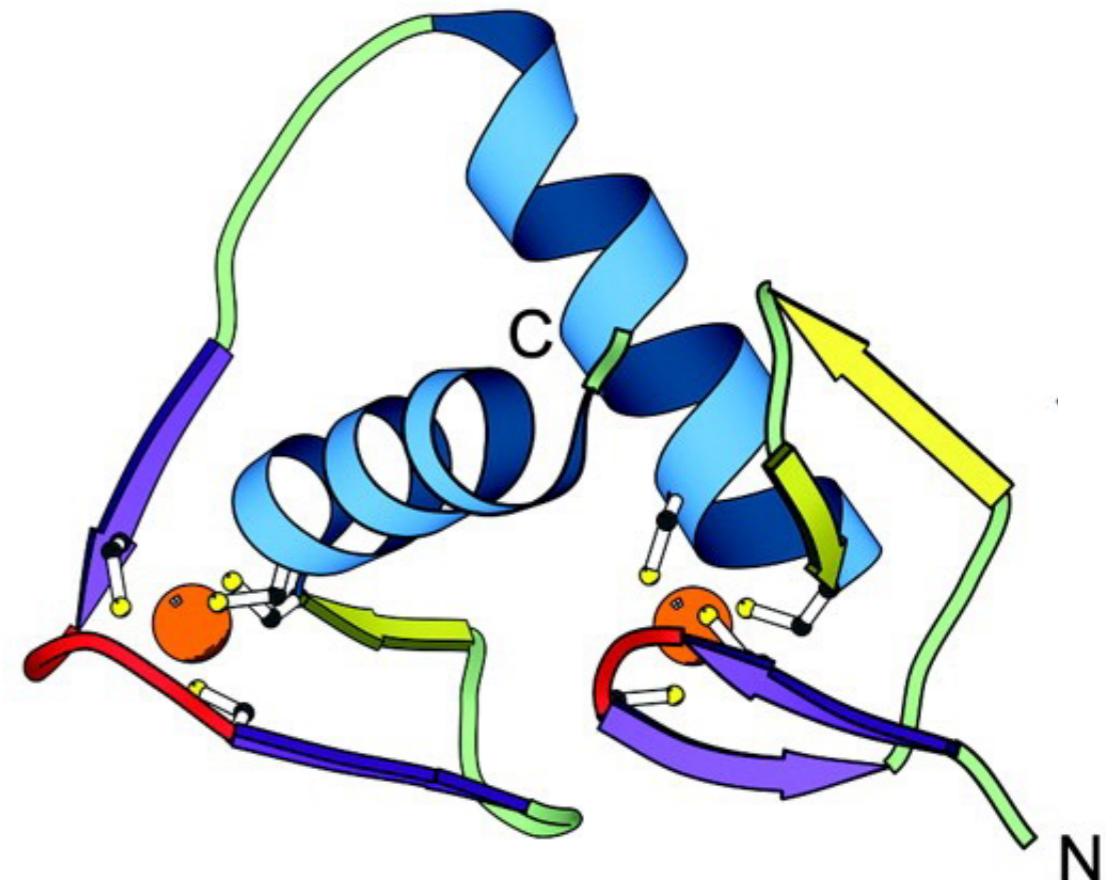
- Variations in a simple motif can provide a dramatic range of DNA sequence recognition.
- Zn Finger nucleases provide for directed mutagenesis. Geurts et al. Knockout rats via embryo microinjection of zinc-finger nucleases. *Science*. 2009;325:433.
- Have you ever heard of TALE & TALENs? They were the rage prior to CRISPR.



Zinc-containing DNA binding domains

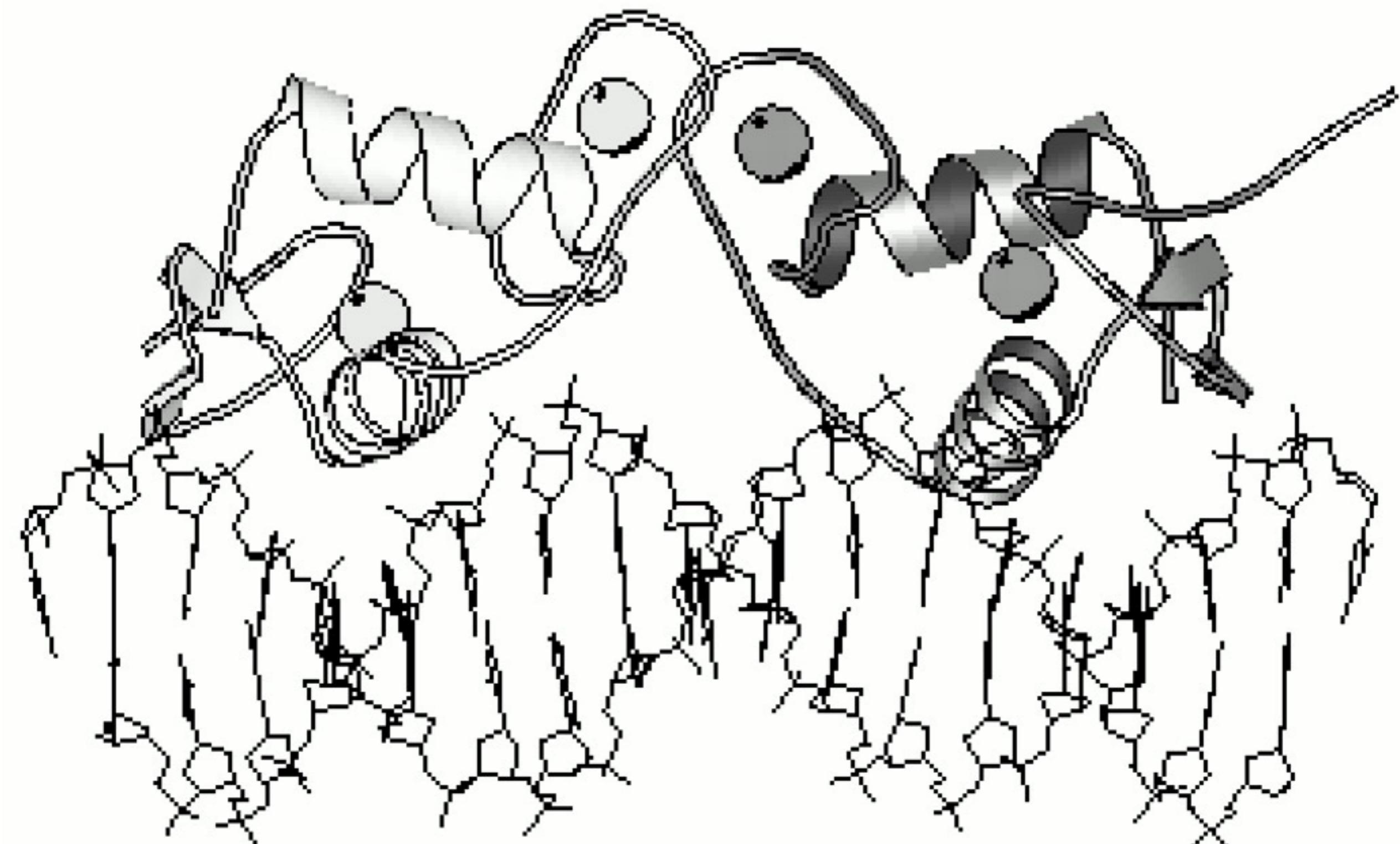
(Zinc is coordinated with a combination of four Cysteine residues)

- Treble-clef Class:
 - β -hairpin at the N-terminus and an α -helix at the C-terminus that each contribute two ligands for zinc binding (a loop and a second β -hairpin of varying length and conformation can be present between the N-terminal β -hairpin and the C-terminal α -helix)

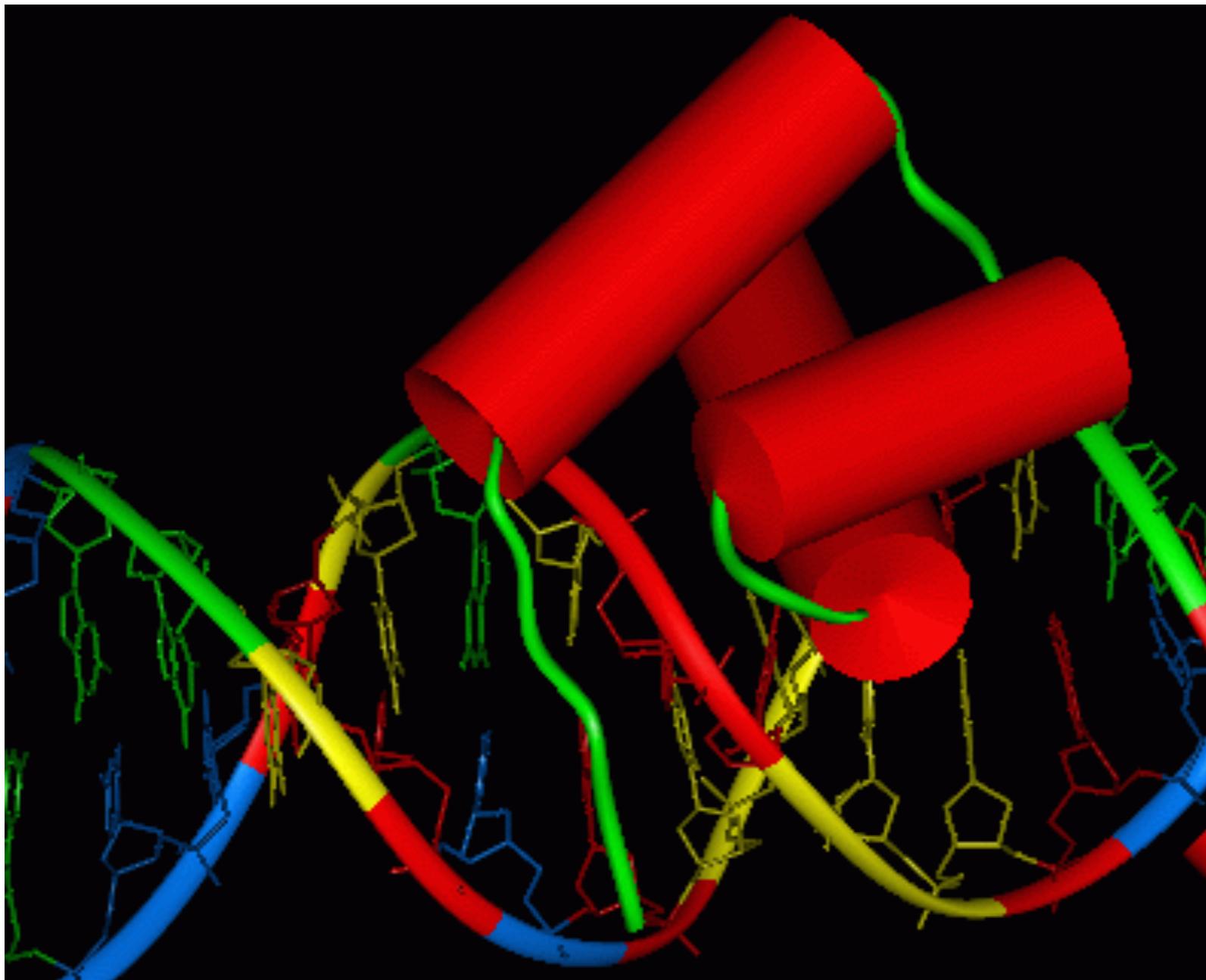


Estrogen Receptor/DNA Complex

(Zinc is coordinated with a combination of four Cysteine residues)



Engrailed Homeodomain/ DNA Complex



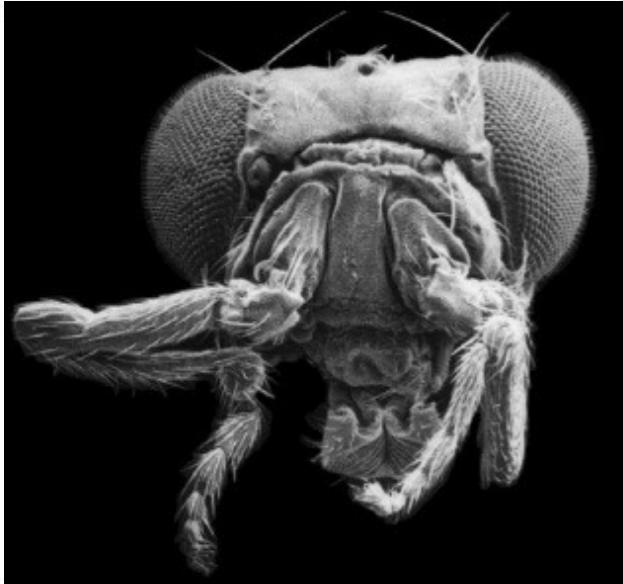
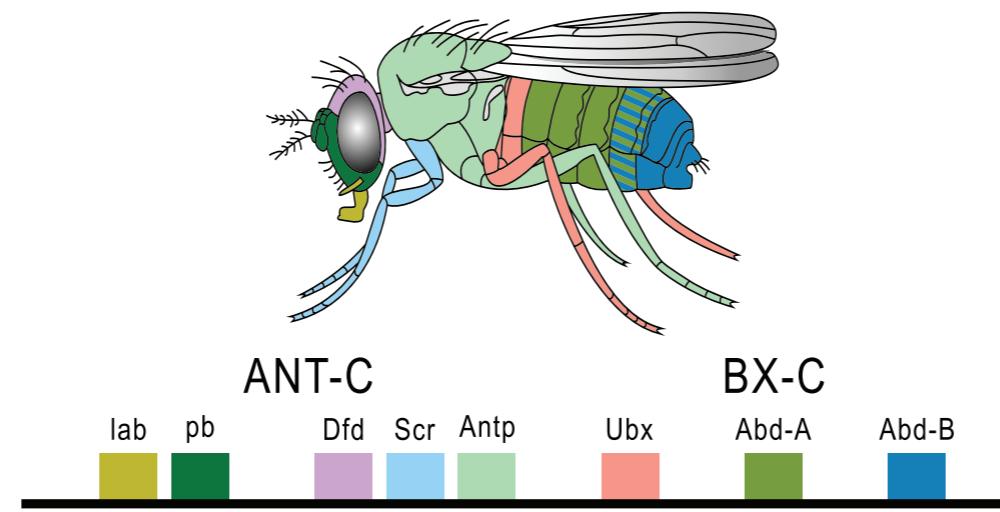
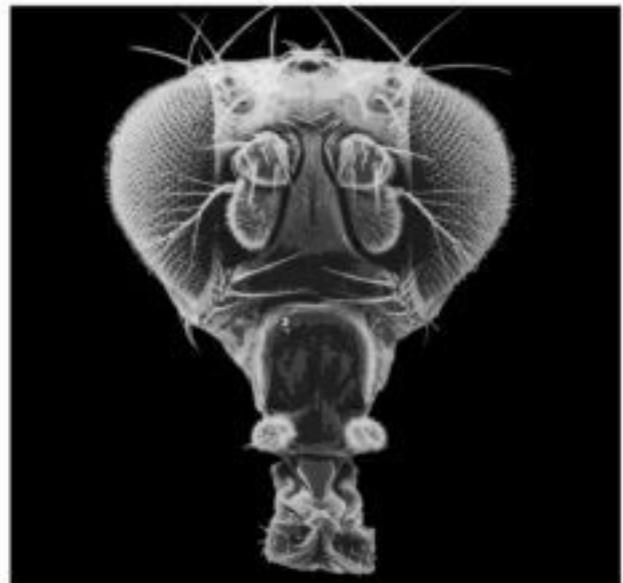
Both major and minor groove interactions by helix 3 and the N-terminal arm respectively

Song Tan lab: Engrailed

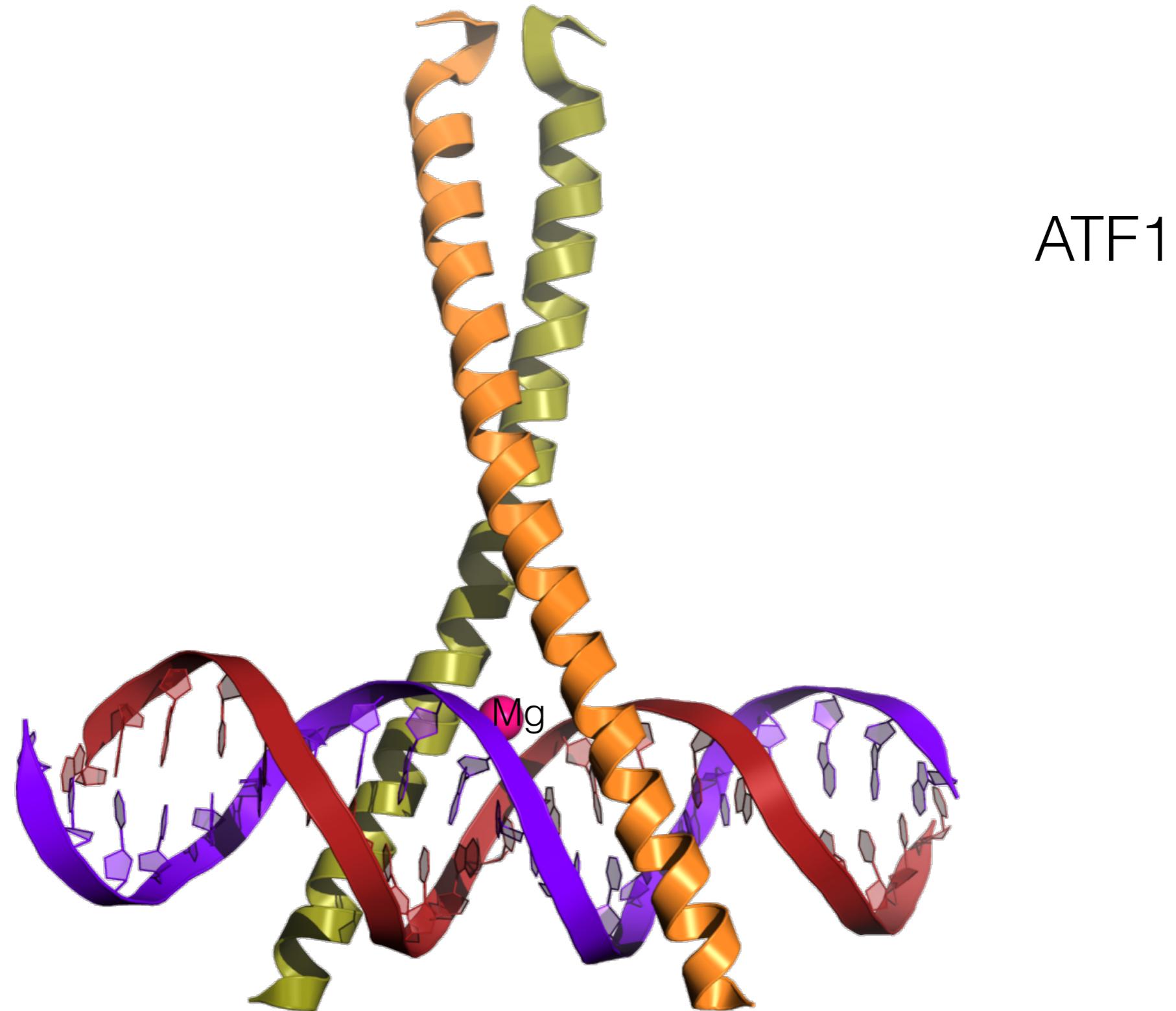


http://www.personal.psu.edu/sxt30/movies/engrdna_h264.mov

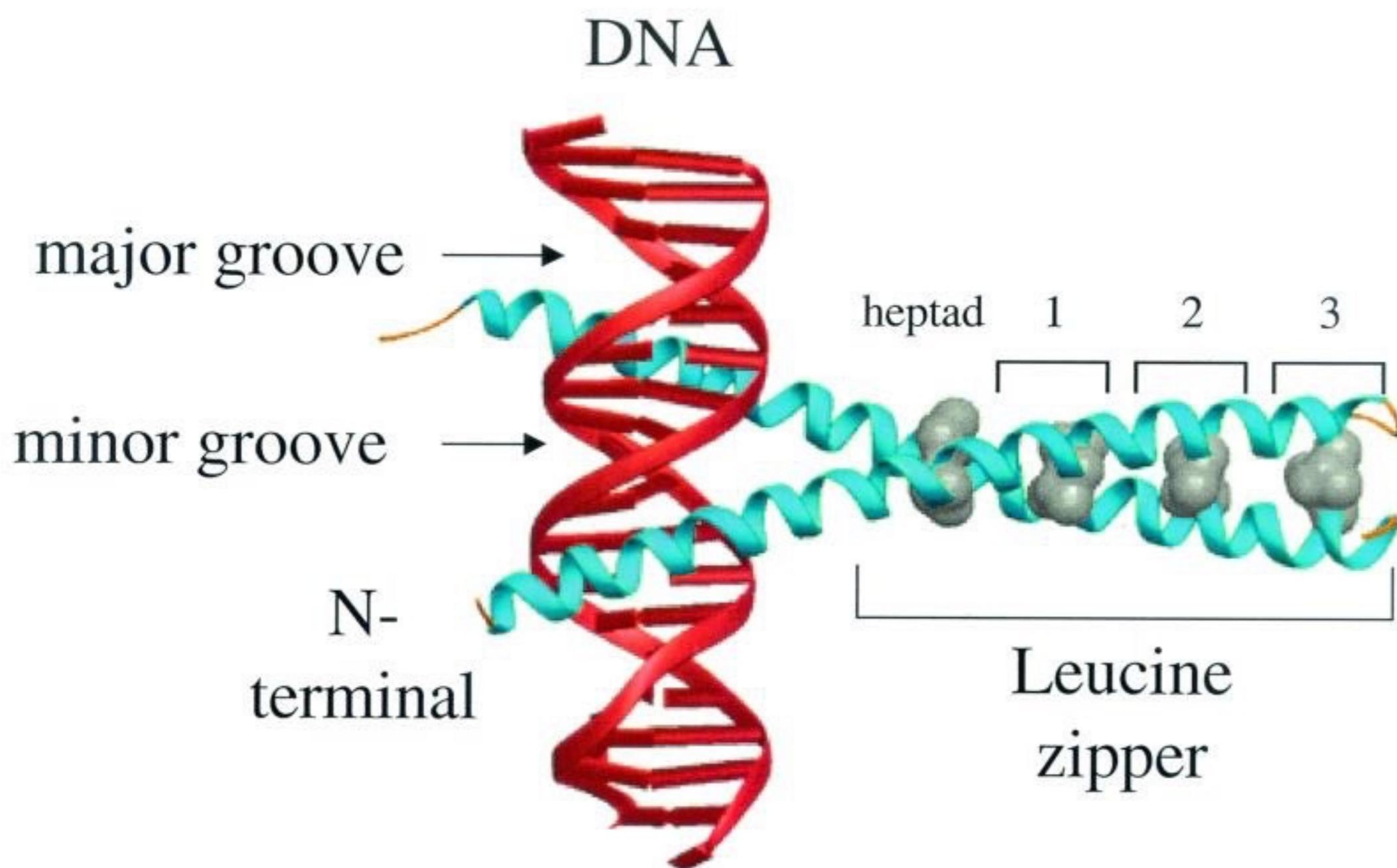
Engrailed Homeodomain (Hox genes also have homeodomains)



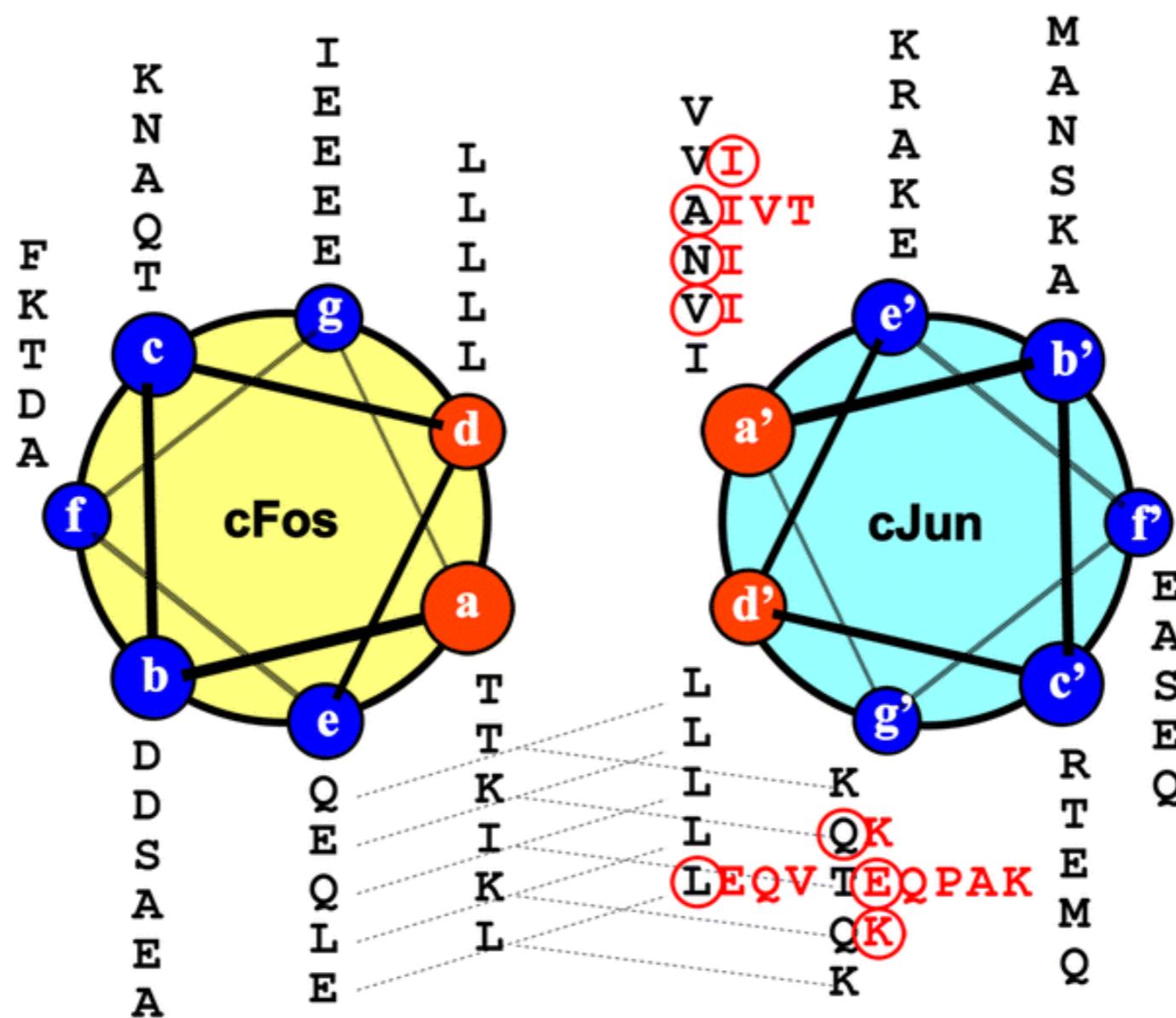
Leucine Zippers: dimerization of TFs



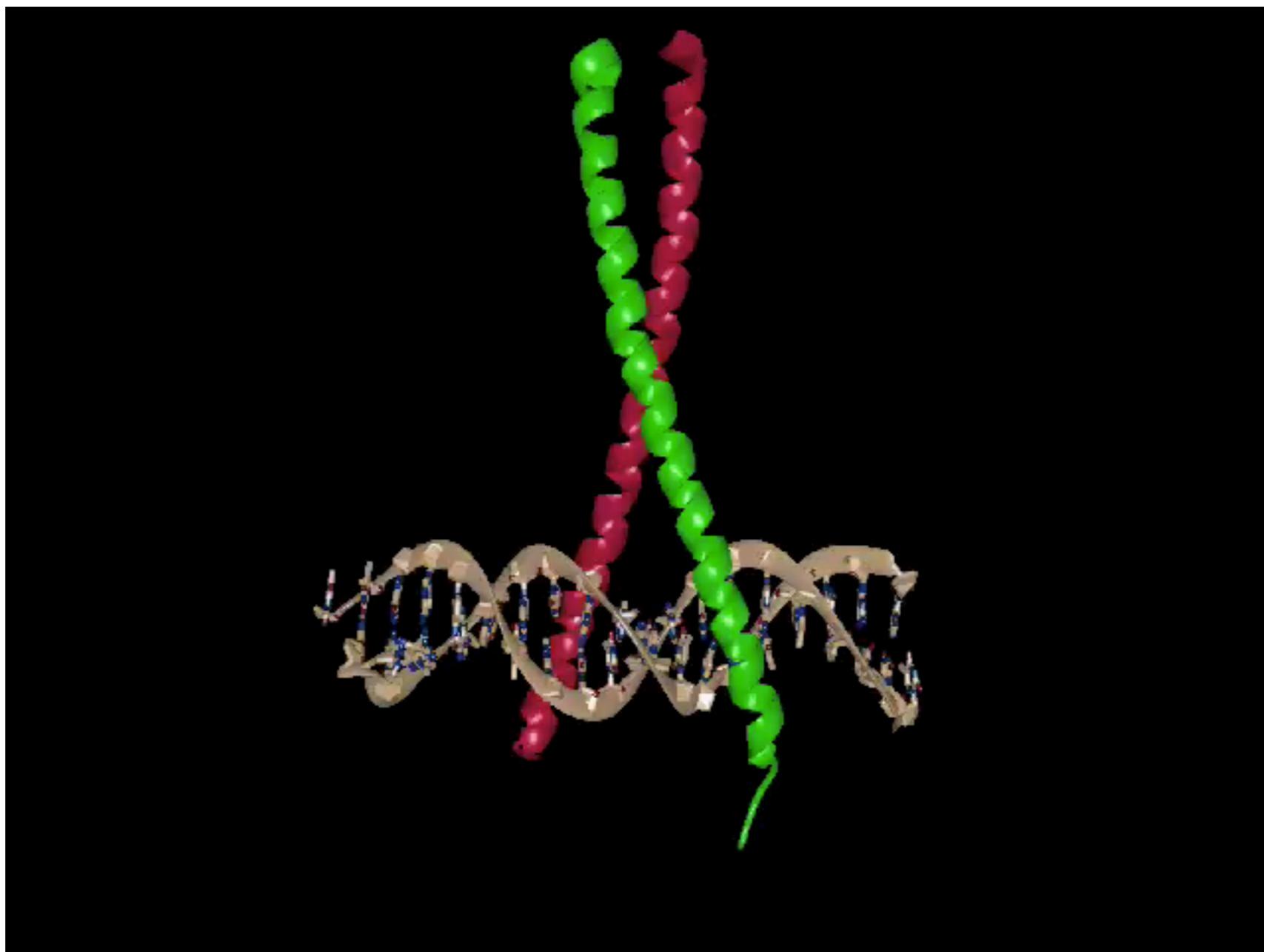
Leucine Zippers: dimerization of TFs



Leucine Zippers: dimerization of TFs

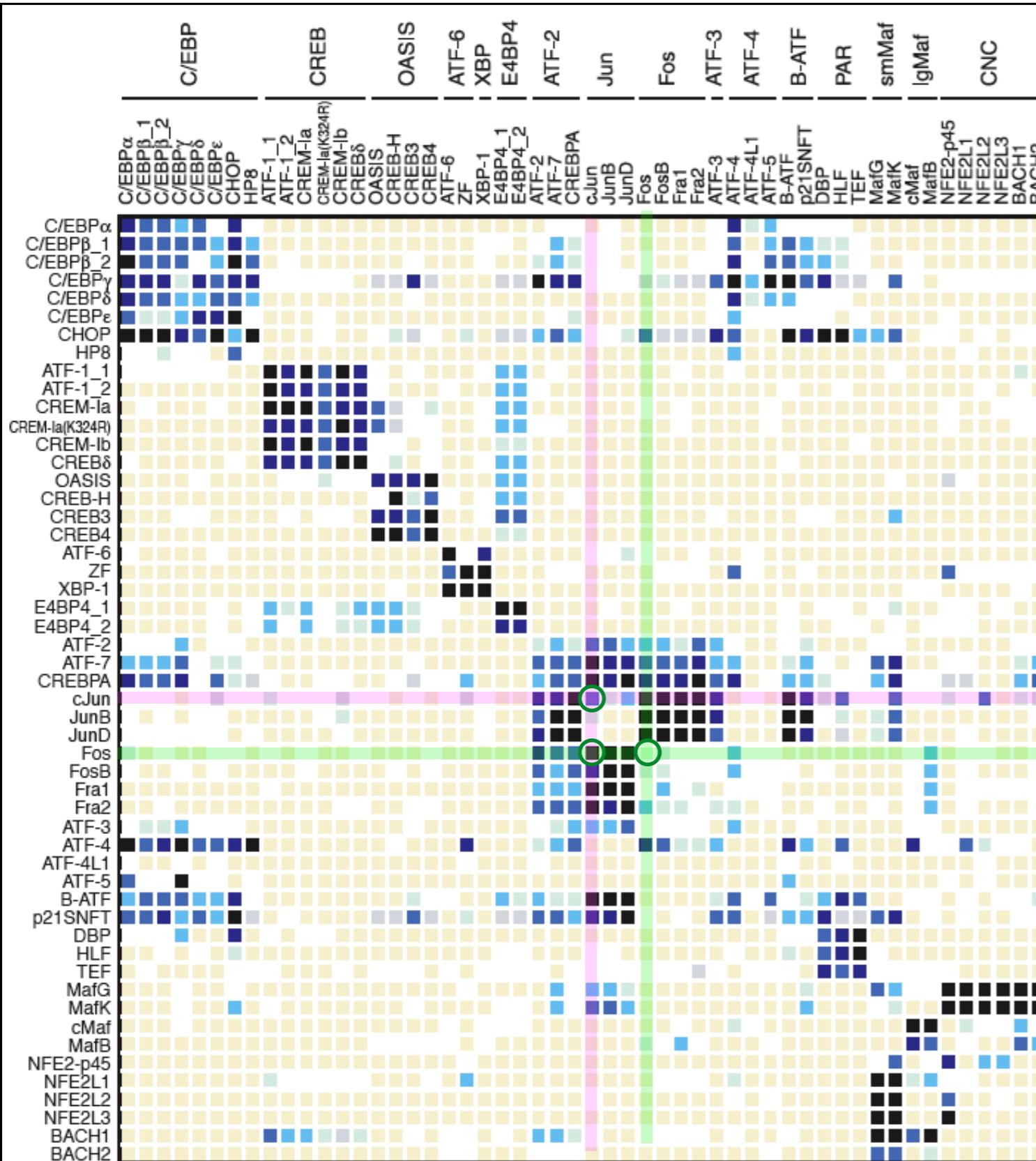


Song Tan Lab: bZIP



Have you heard of the transcription factor AP1?

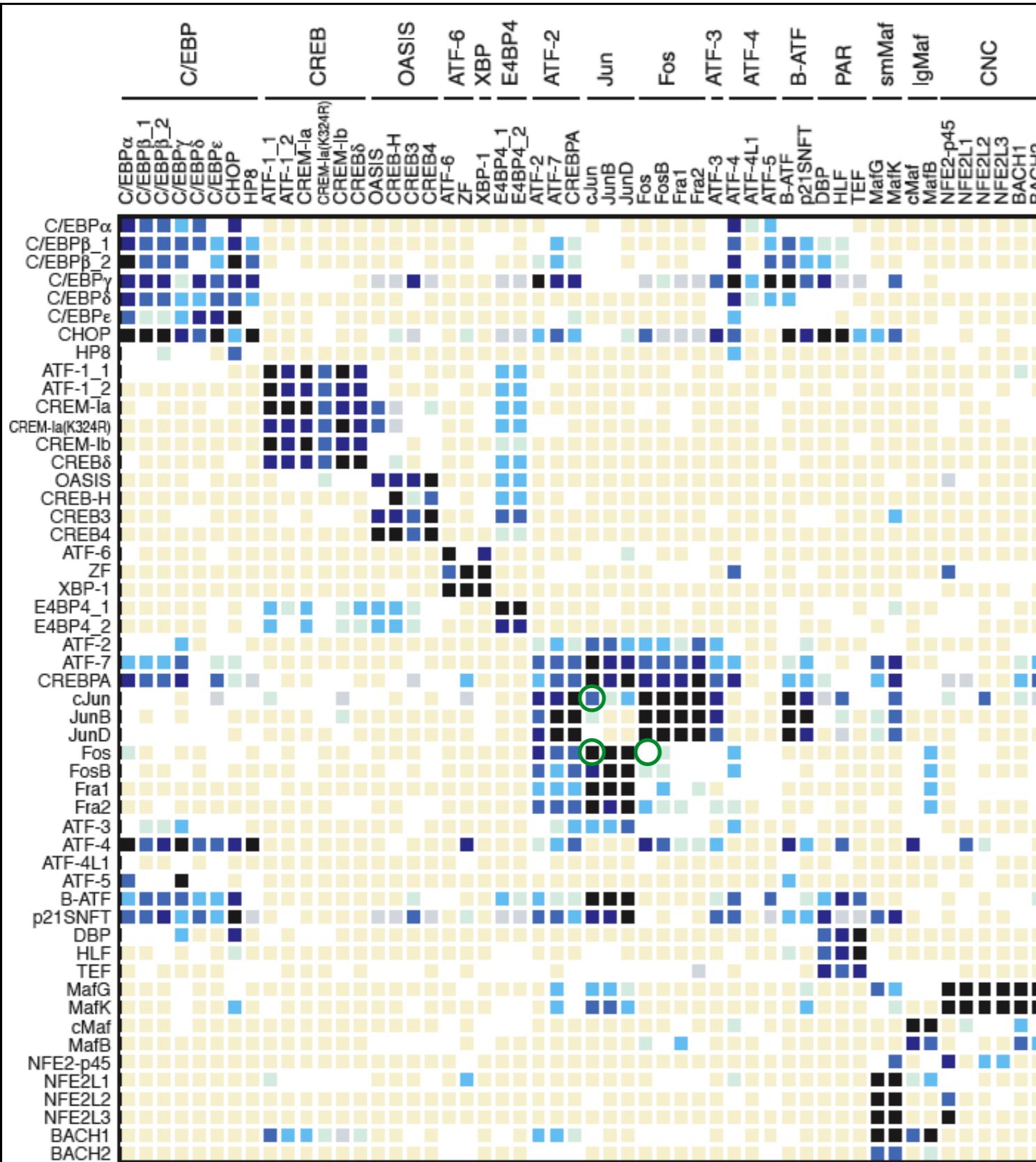
Interaction Matrix for 49 human bZIP Peptides



~14% of possible interactions detected.

Most between family members, but 136 between families.

Leucine Zippers Provide Specific Dimerization Interactions

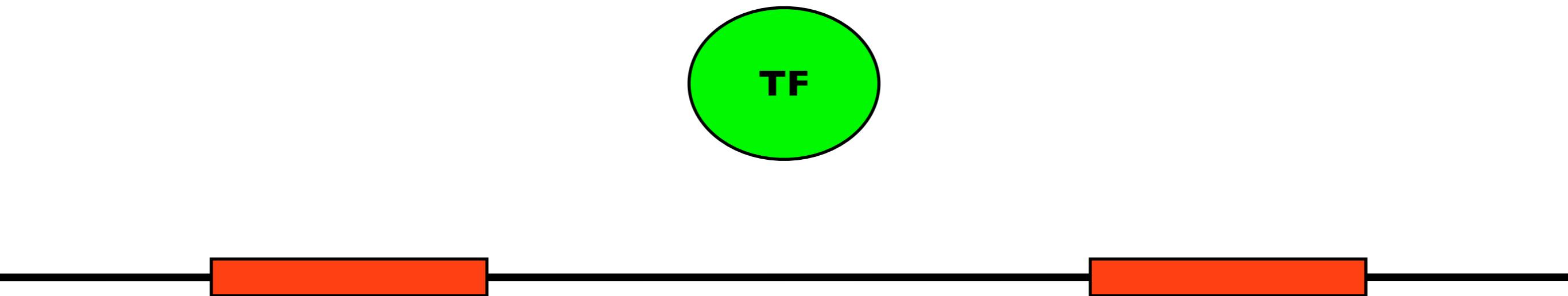


cJun/cFos heterodimer forms preferentially relative to homodimers

cJun/cJun forms but has two unfavorable charge interactions.

cFos/cFos does not form - four unfavorable charge interactions.

Chromatin affects TF binding in vivo



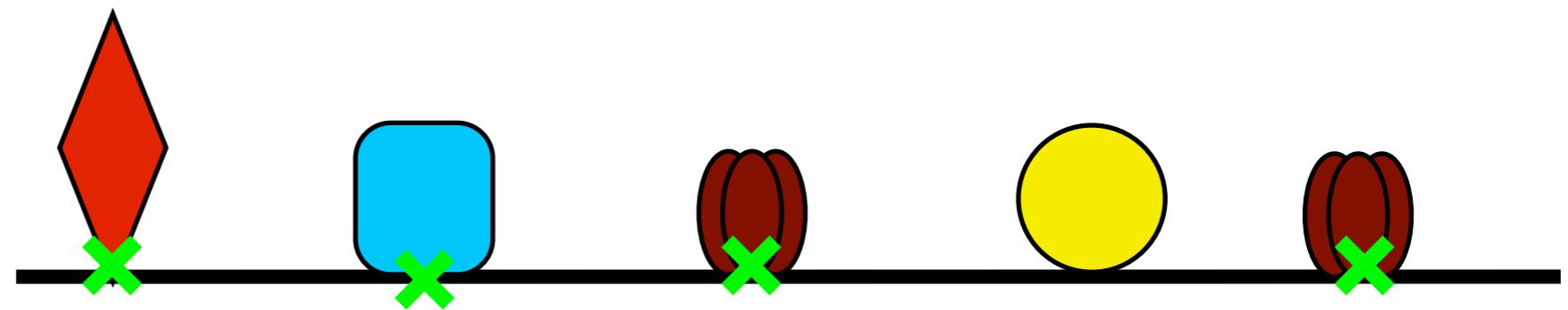
-Besides sequence, what influences TF binding?

HSF binds many sites after HS



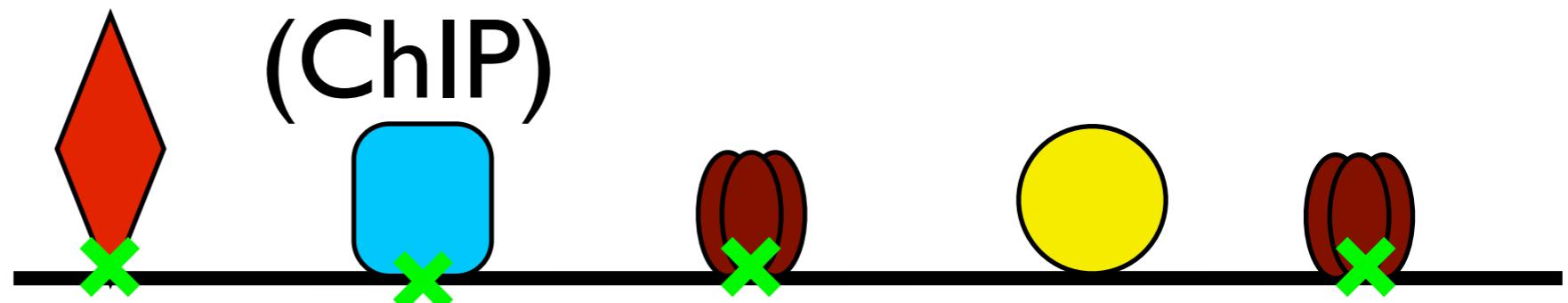
Chromatin Immunoprecipitation (ChIP)

Crosslink DNA and
Proteins

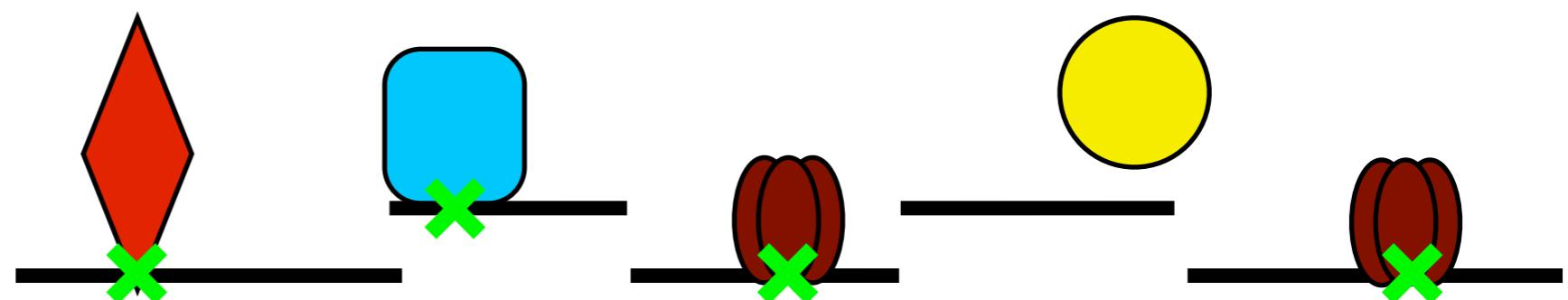


Chromatin Immunoprecipitation (ChIP)

Crosslink DNA and
Proteins

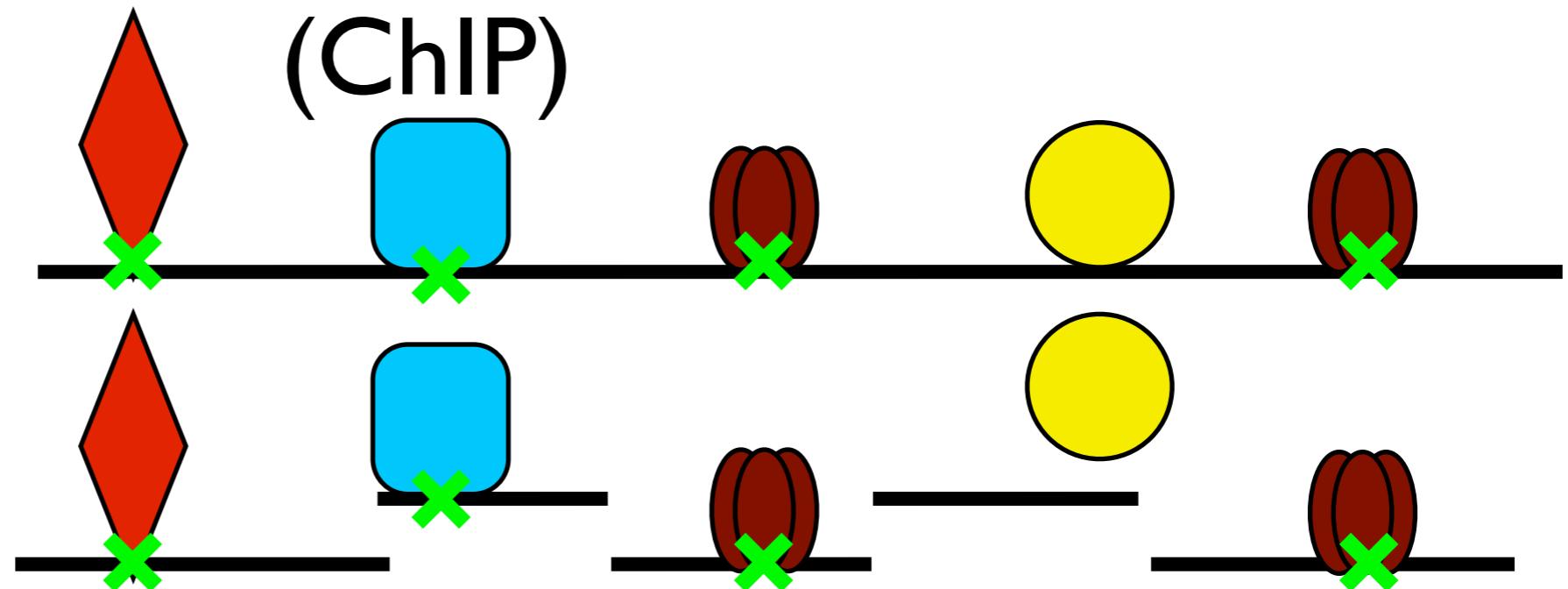


Shear DNA



Chromatin Immunoprecipitation (ChIP)

Crosslink DNA and
Proteins

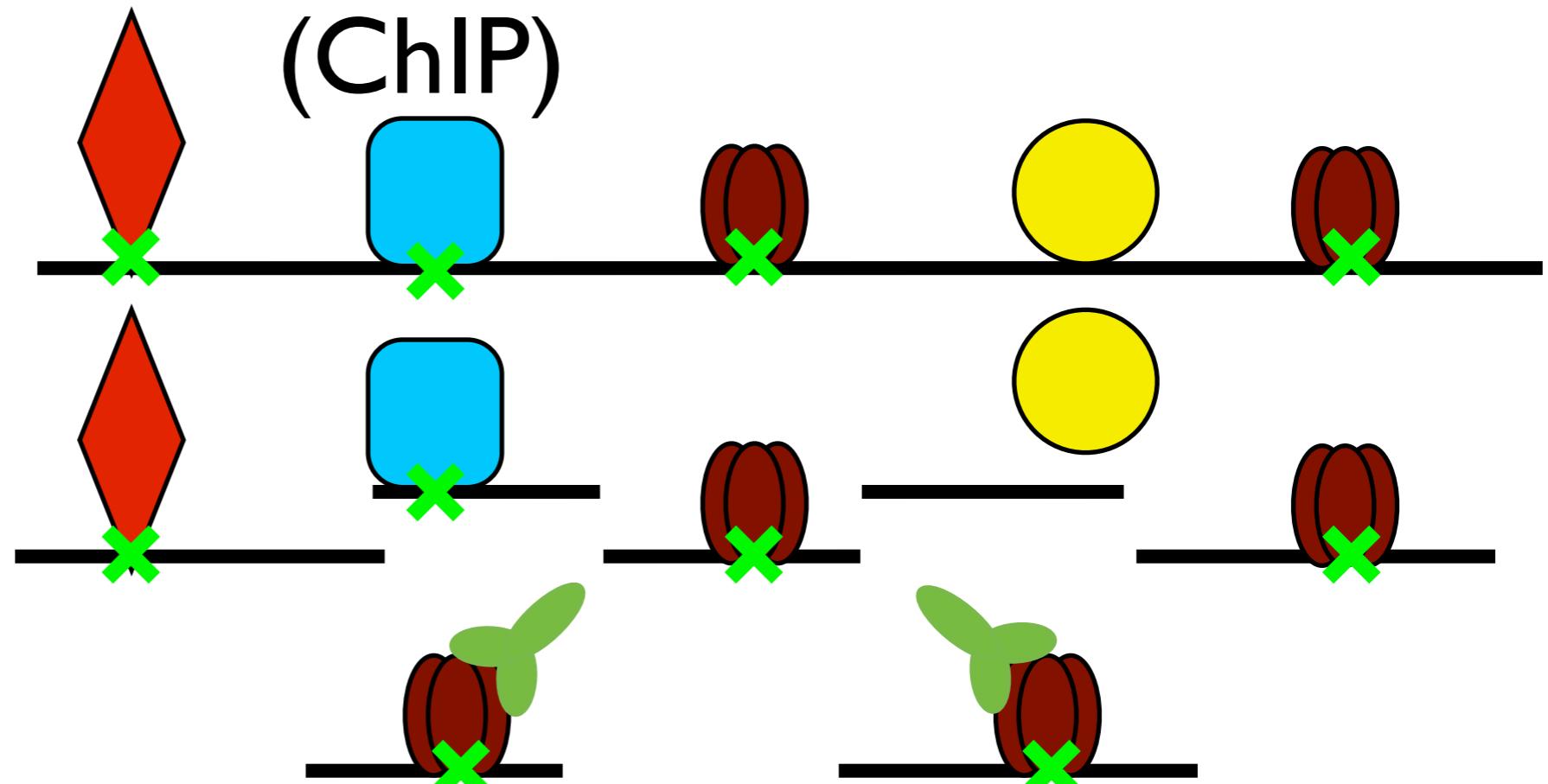


Shear DNA



Chromatin Immunoprecipitation (ChIP)

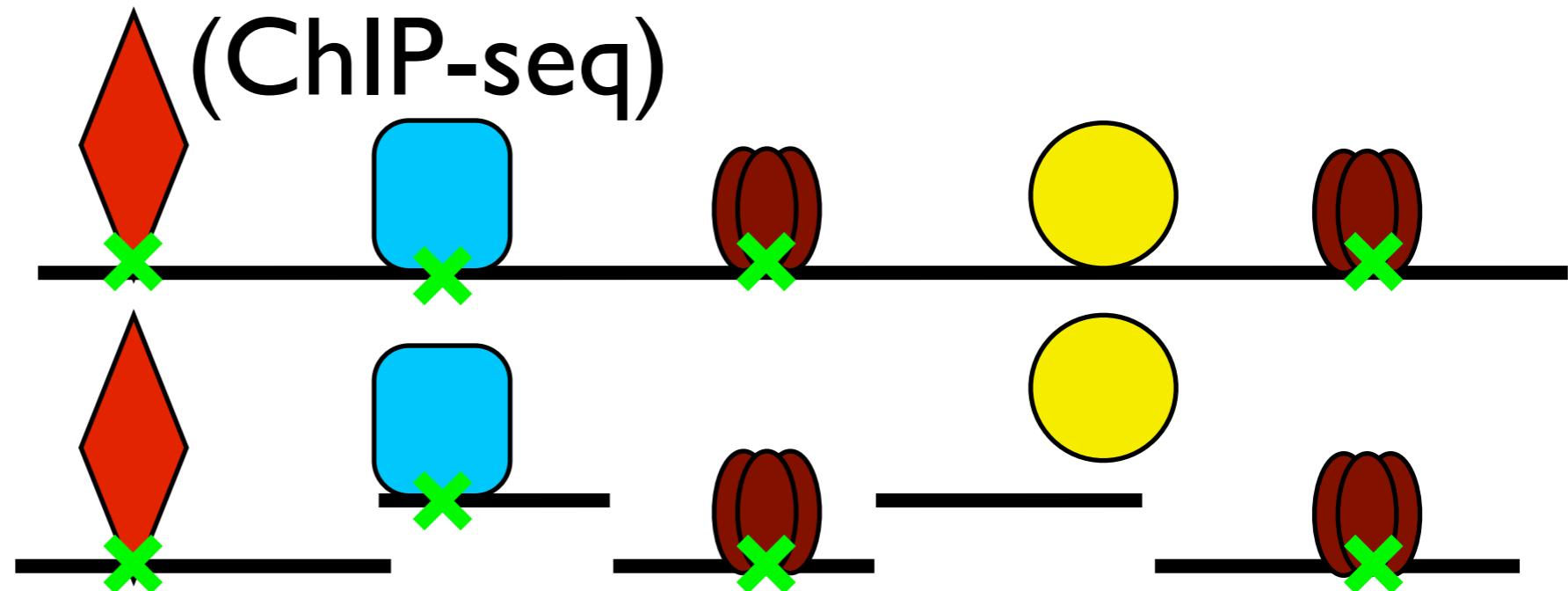
Crosslink DNA and
Proteins



Purify DNA

Chromatin Immunoprecipitation (ChIP-seq)

Crosslink DNA and
Proteins



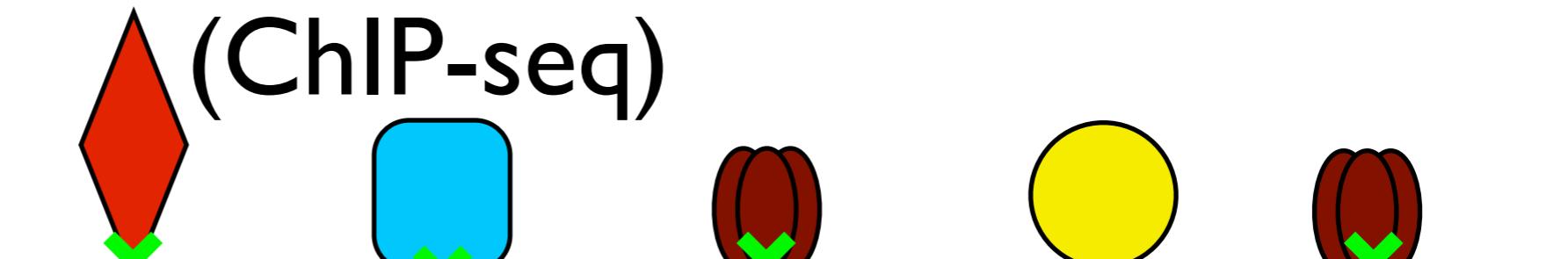
Immunoprecipitate

Purify DNA

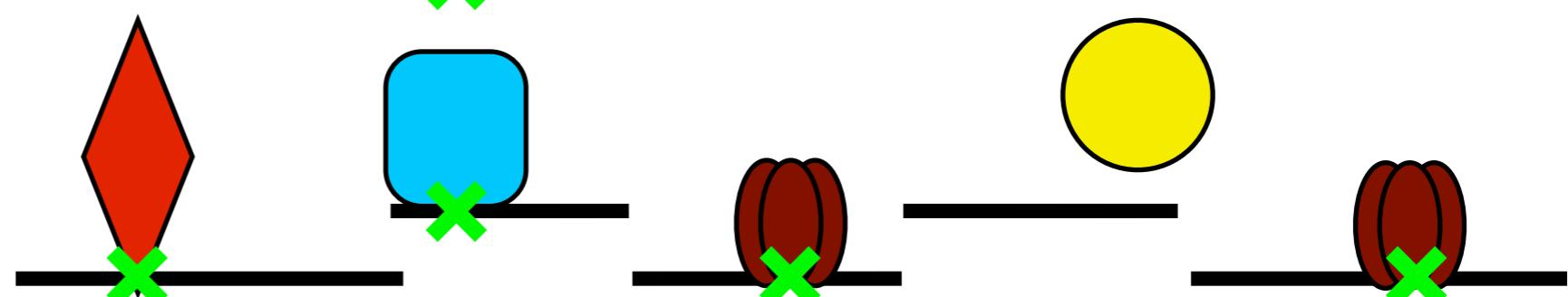
Ligate adapters

Chromatin Immunoprecipitation (ChIP-seq)

Crosslink DNA and
Proteins



Shear DNA



Immunoprecipitate



Purify DNA



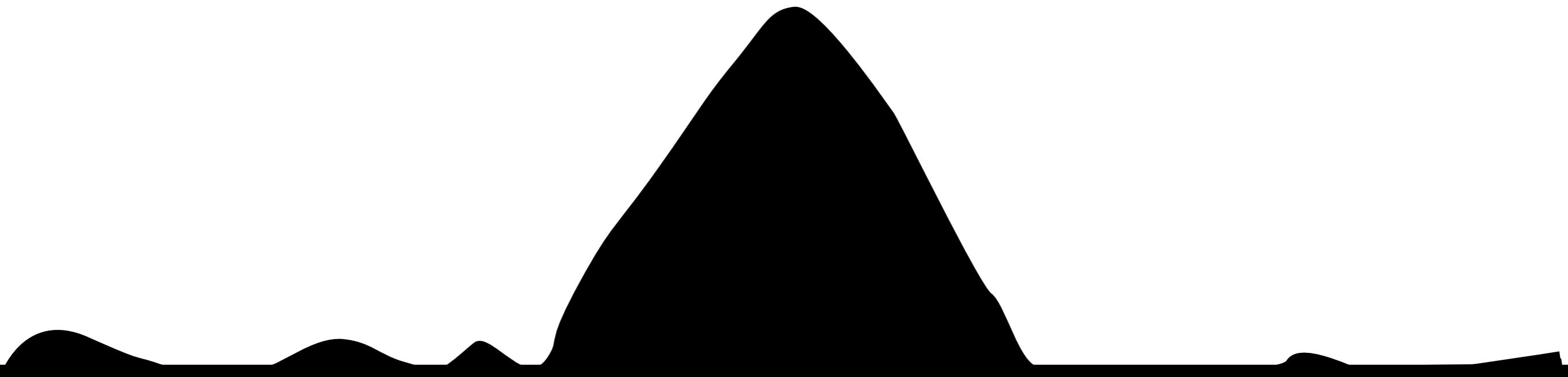
Ligate adapters



Sequence DNA ends

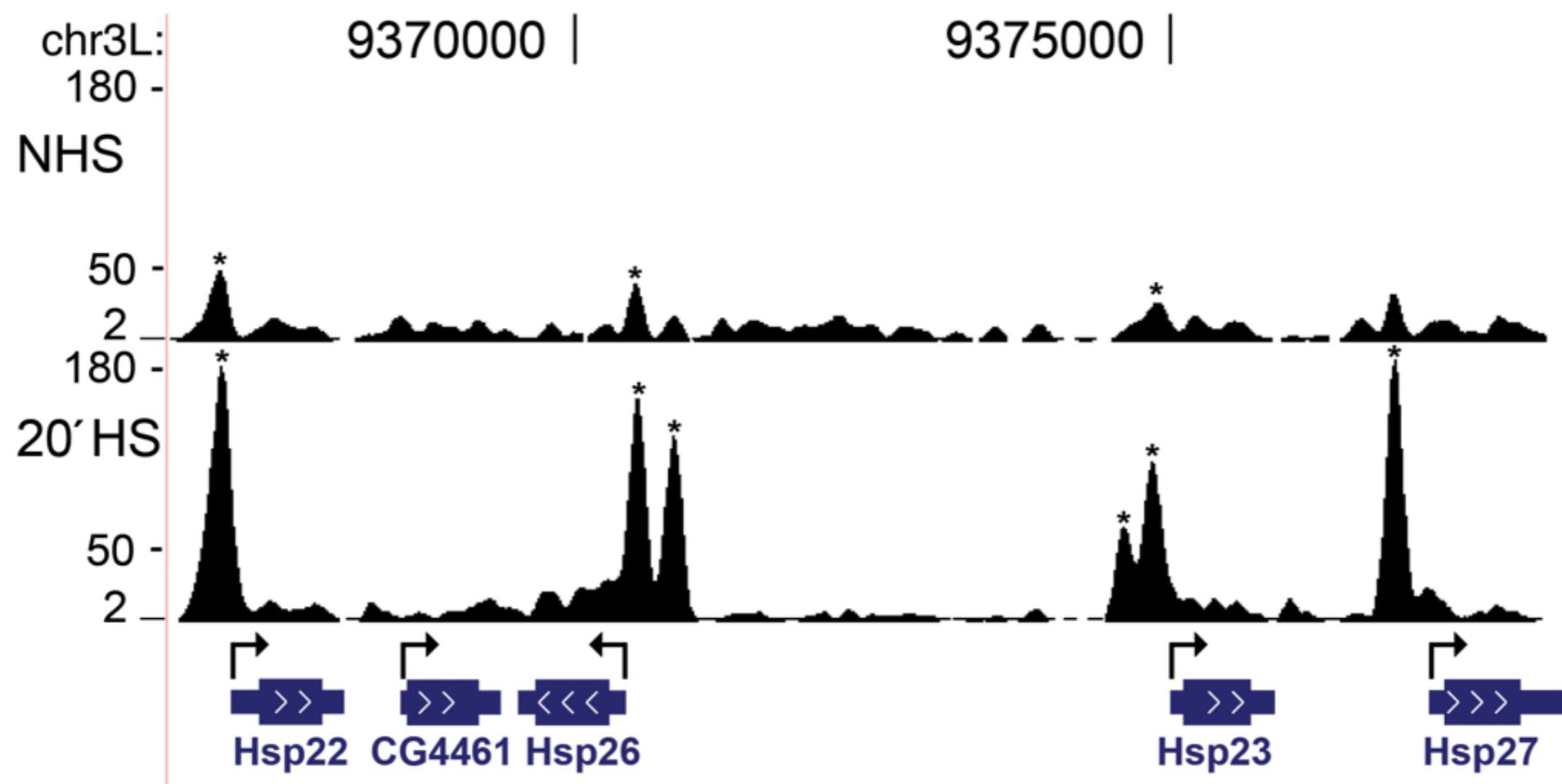


Chromatin Immunoprecipitation (ChIP-seq)

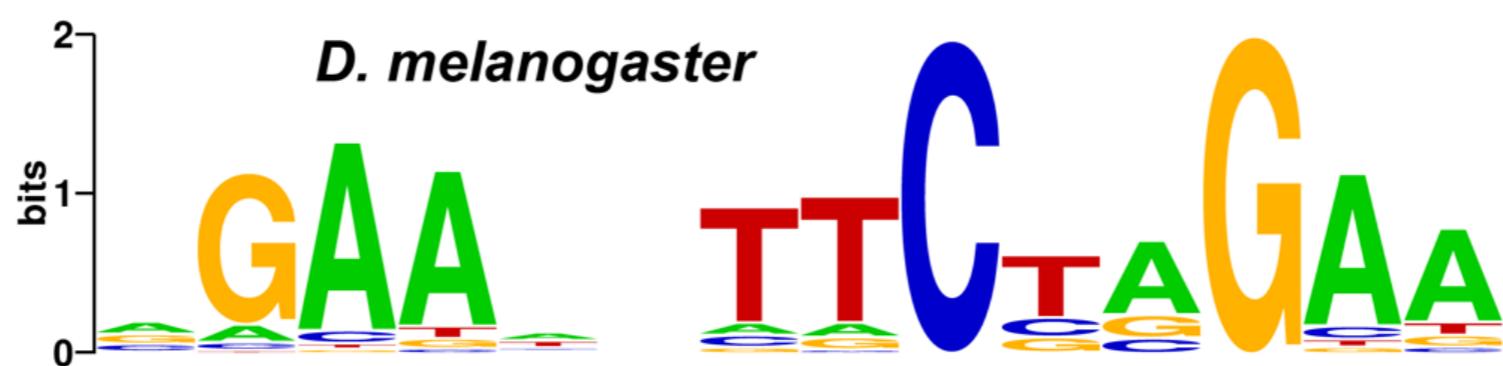


Peak calling: Zhang, et. al., Genome Biology 2008 (and many others)

HSF targets DNA inducibly

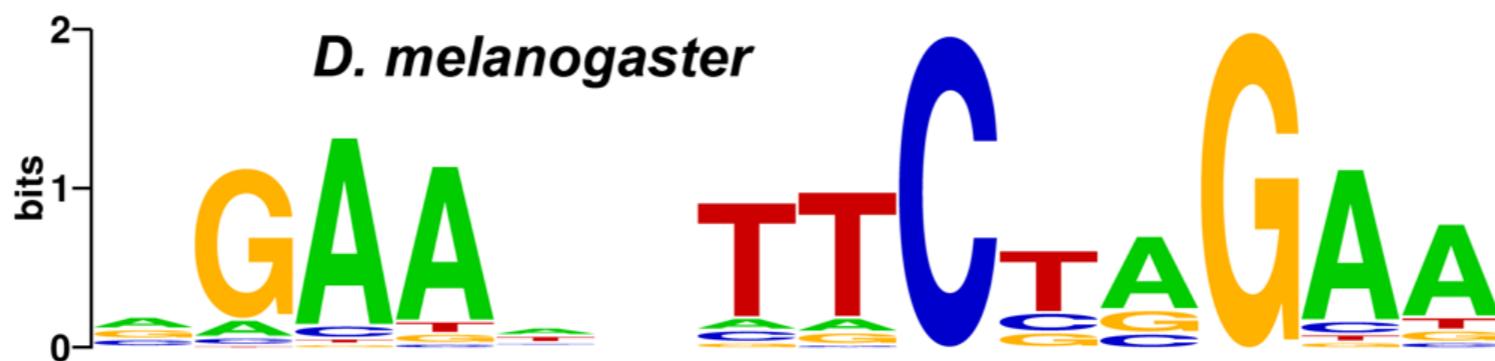


HSF targets a consensus motif



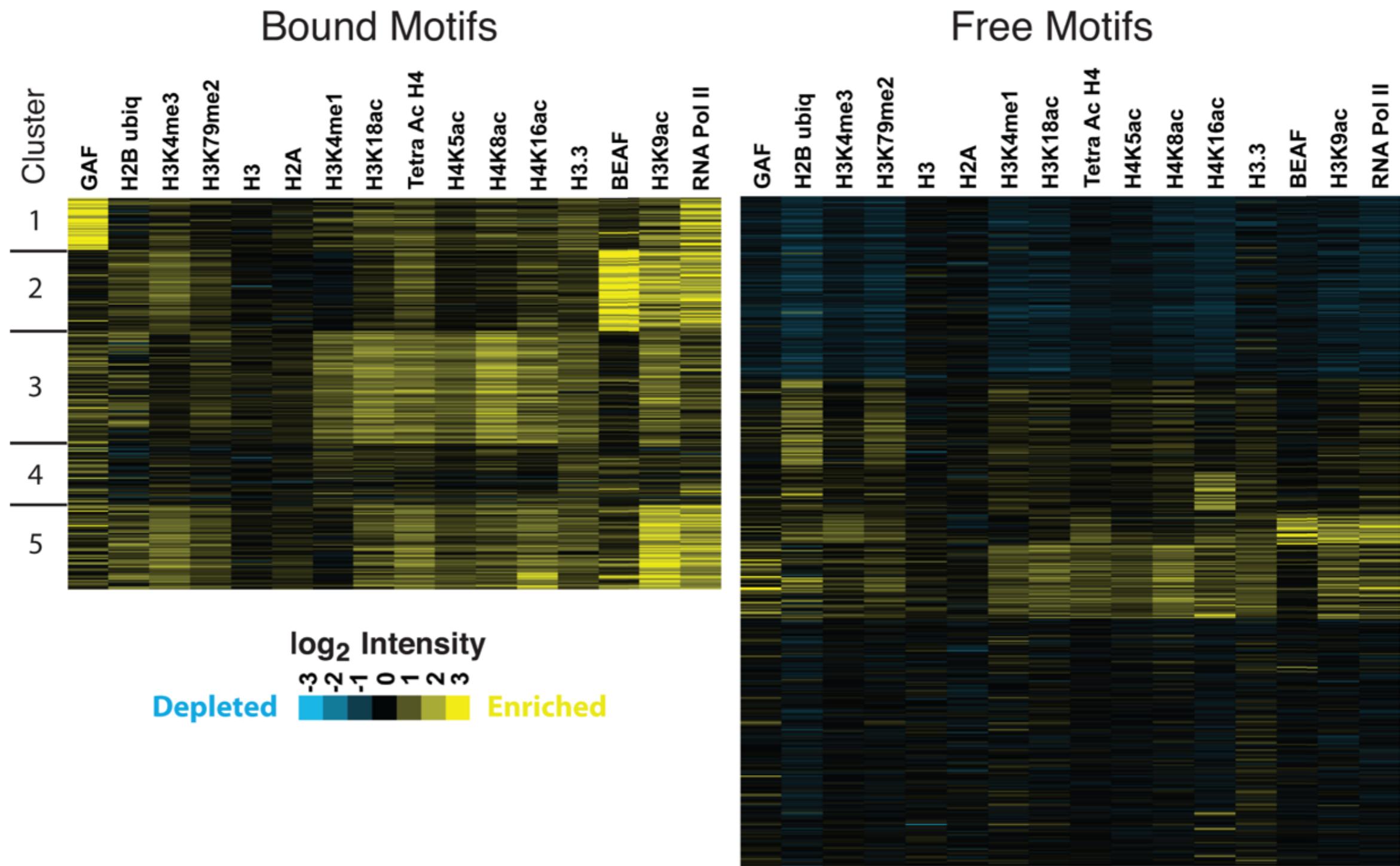
HSF binds a fraction of motifs in vivo

- Queried the *Drosophila* genome using this HSE matrix:

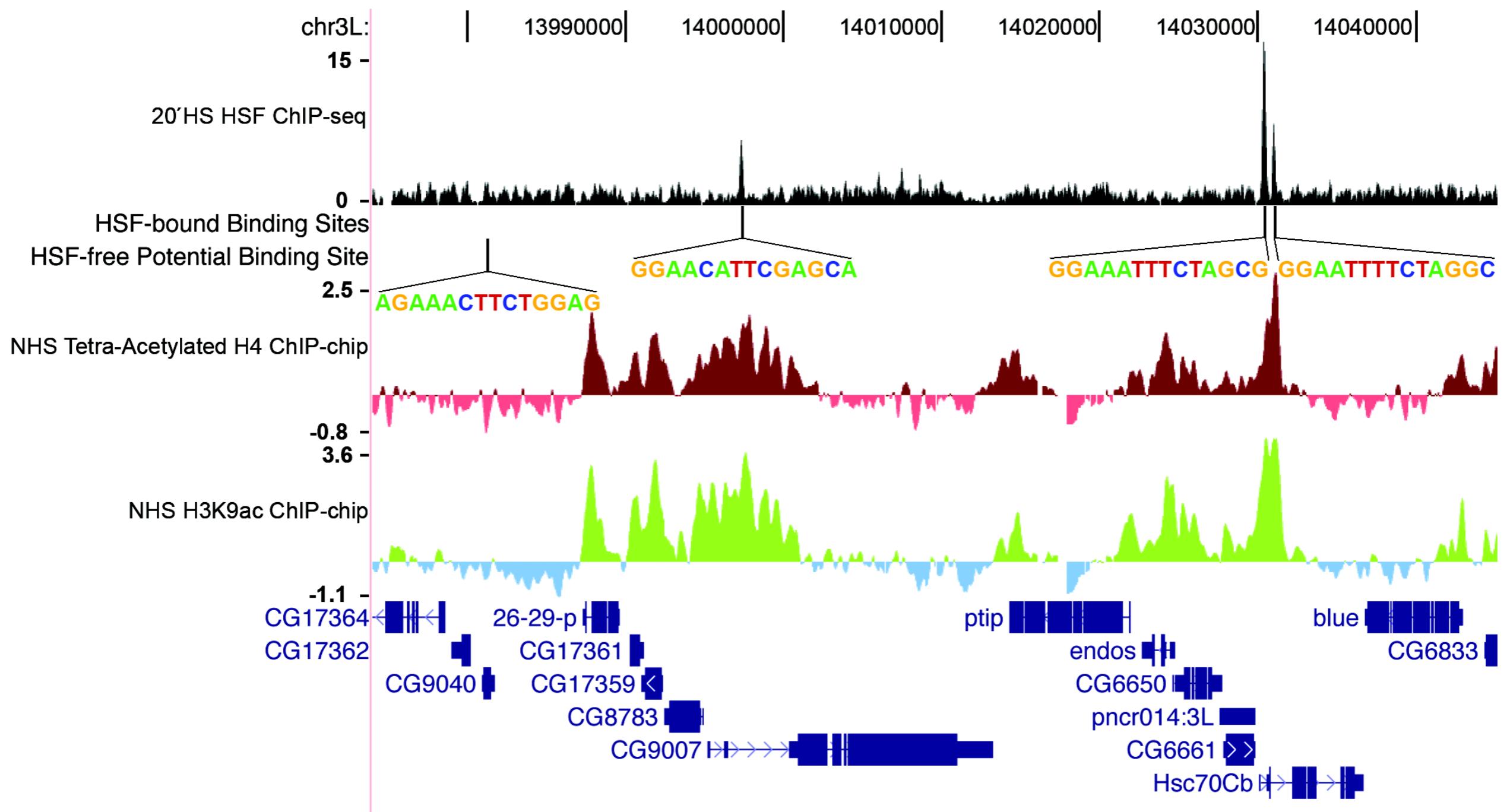


- Found 708 post-HS HSF-free motifs that conform stringently to this consensus HSE, compared to 442 HSF-bound motifs.
- Note that these are computational predictions of potential HSF binding sites

HSF targets motifs within active chromatin

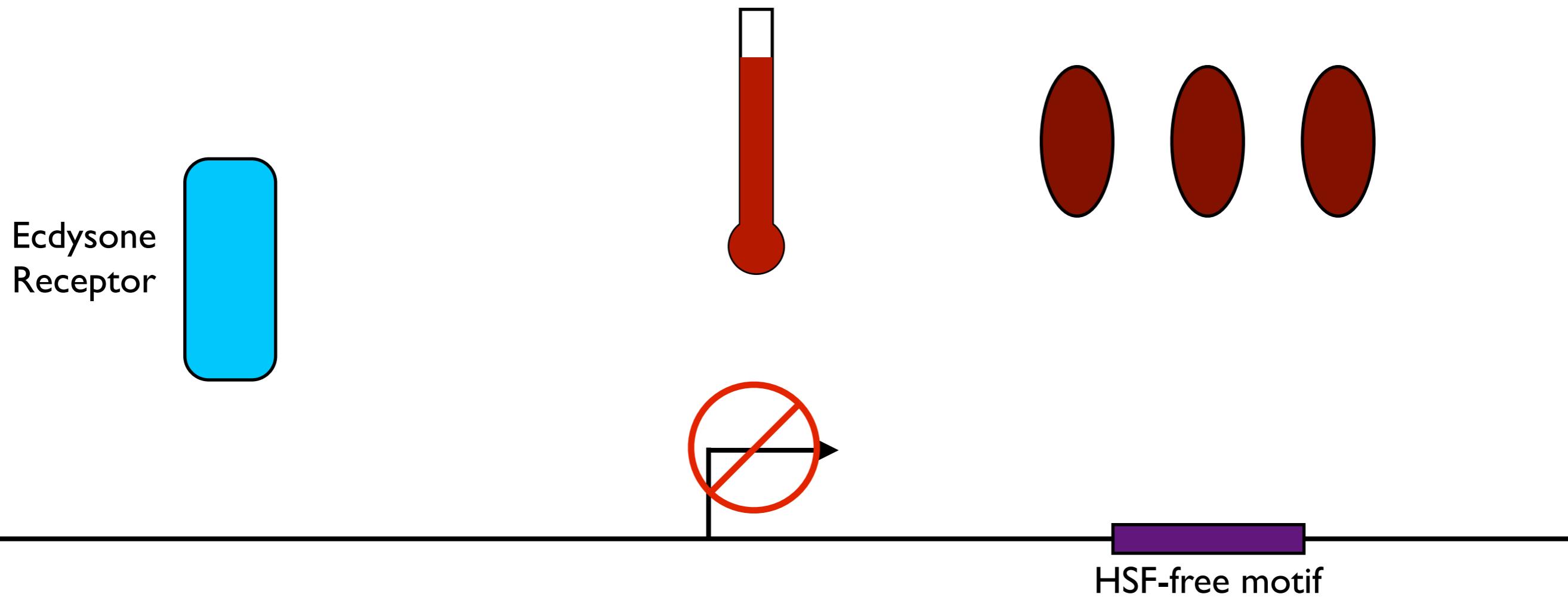


Correlative genomics allows one to develop a hypothesis;
we hypothesize that active chromatin permits TF binding

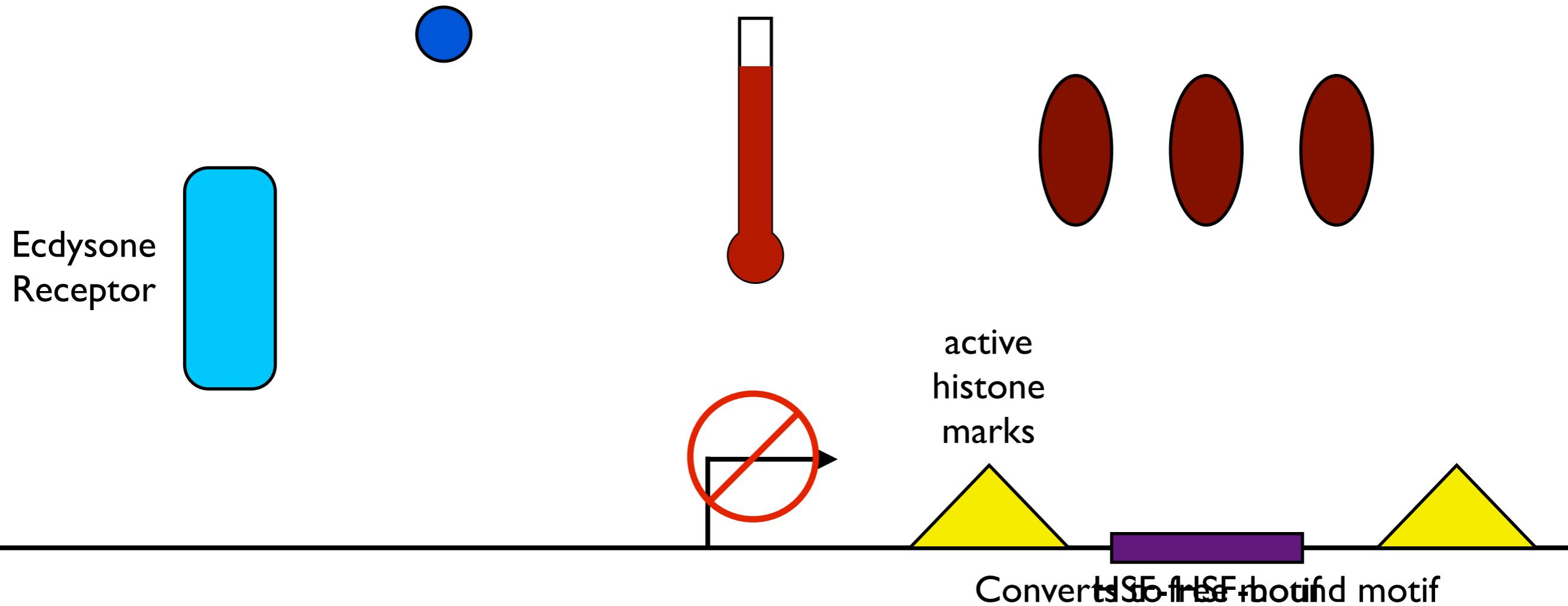


Changing the chromatin at an HSE

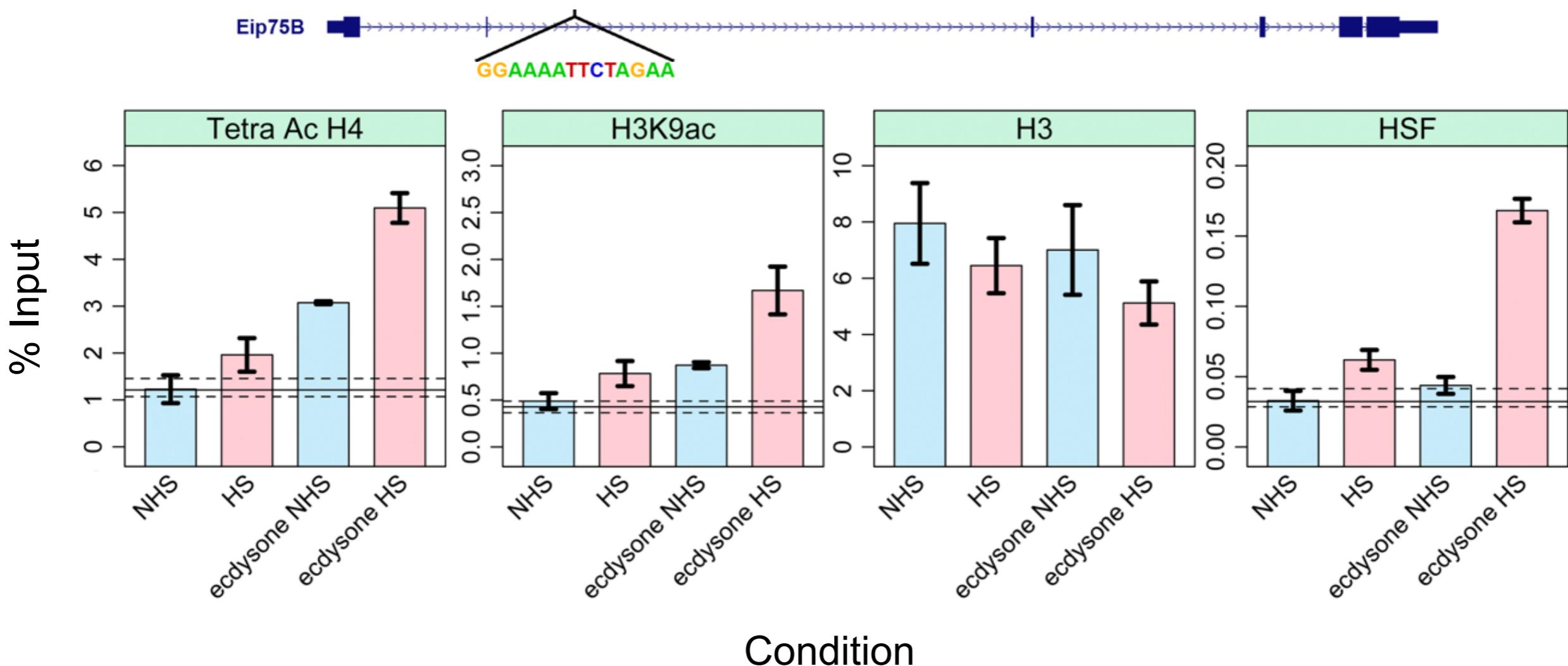
(now this is done trivially with CRISPR-dCas9 coupled to chromatin modifiers)



Changing the chromatin at an HSE



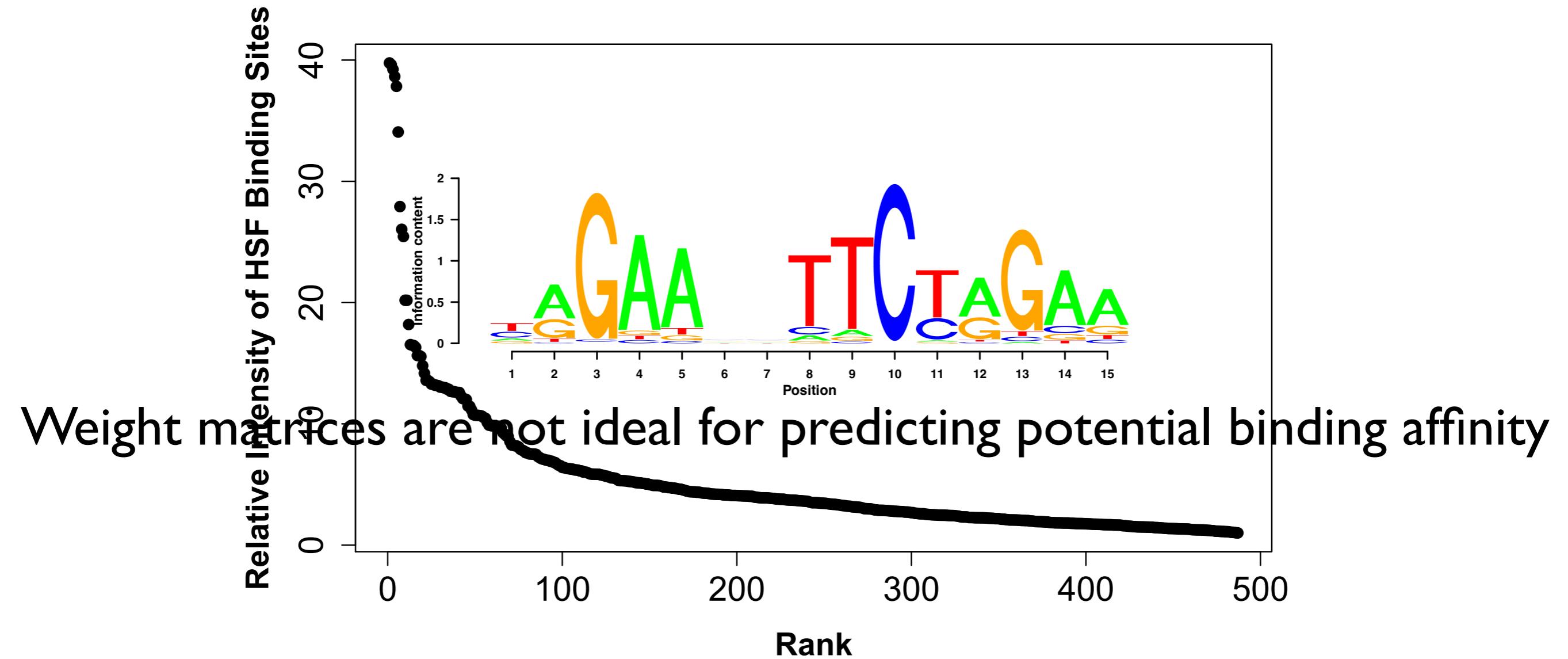
Converting and HSF-free to an HSF-bound motif



Biology is continuous, not discrete

- So far I have considered HSF-bound and HSF-free as two separate categories.
 - In reality, some low intensity binding sites look more like unbound regions and the intensity of binding is meaningful.
- Can we predict inducible TF binding intensity?

Predicting TF binding intensities



in vitro nucleic acid/protein binding (PB-seq)

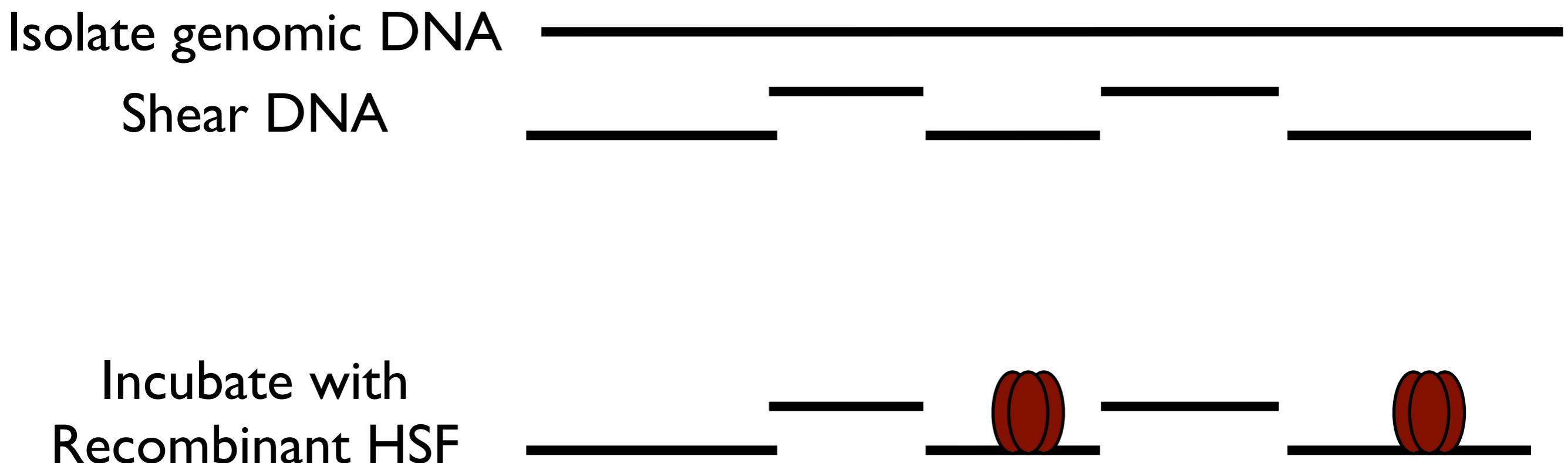
Isolate genomic DNA —————

in vitro nucleic acid/protein binding (PB-seq)

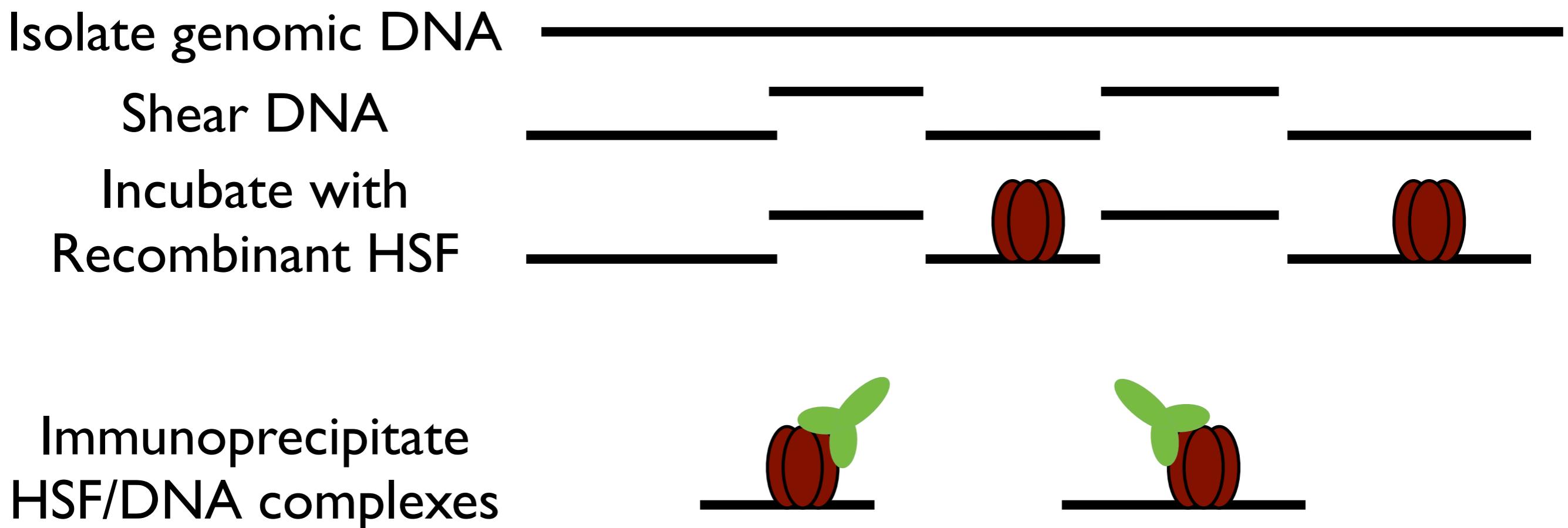
Isolate genomic DNA ——————

Shear DNA ——————

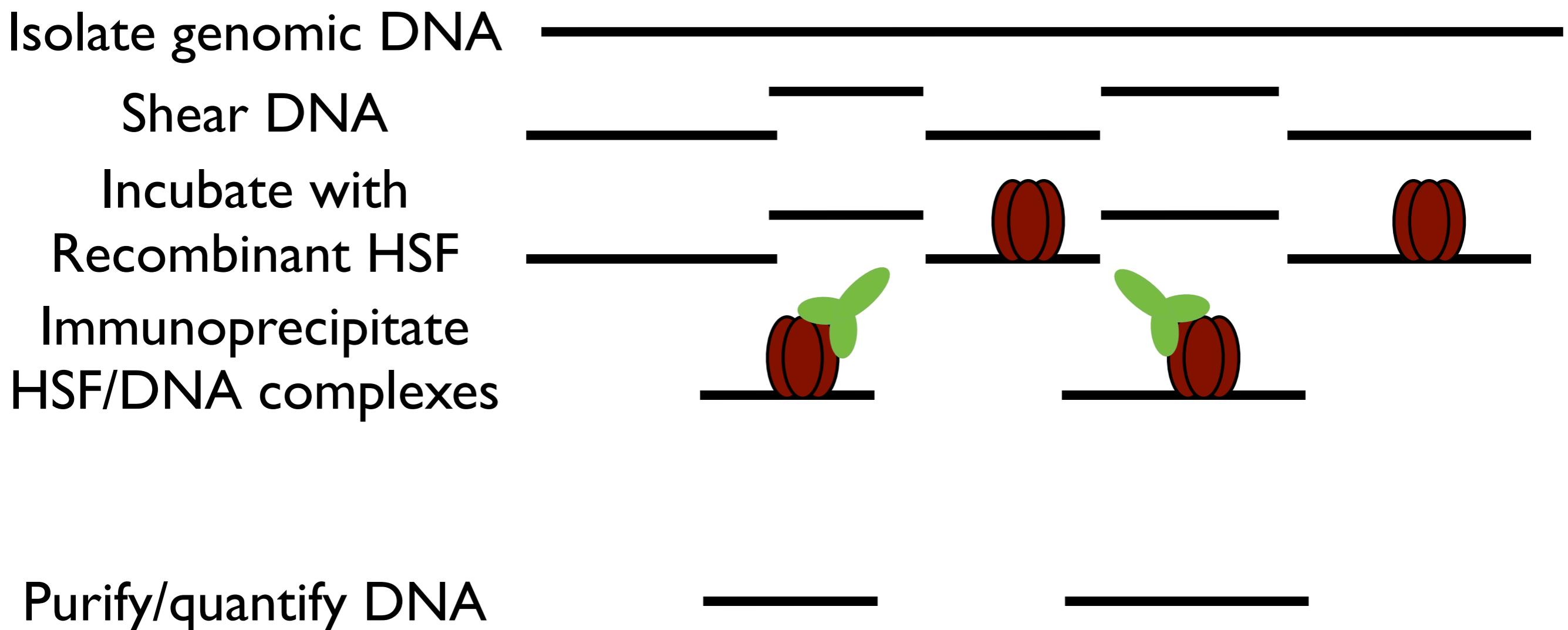
in vitro nucleic acid/protein binding (PB-seq)



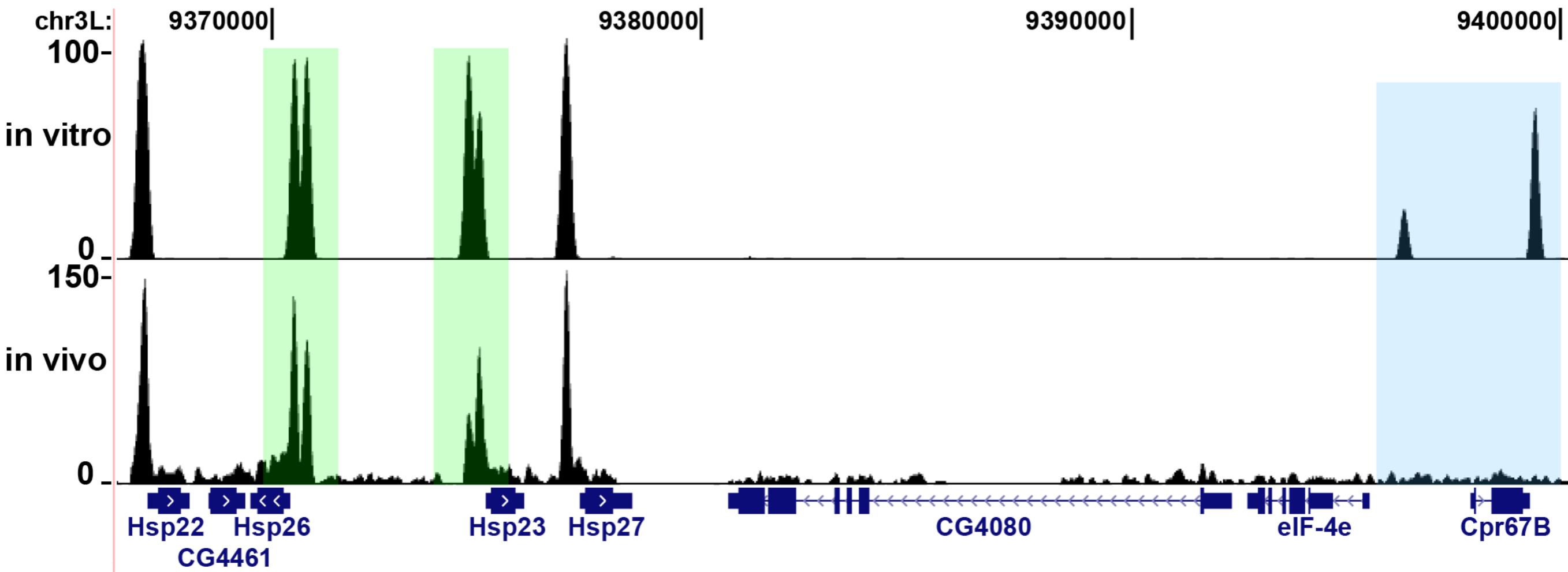
in vitro nucleic acid/protein binding (PB-seq)



in vitro nucleic acid/protein binding (PB-seq)

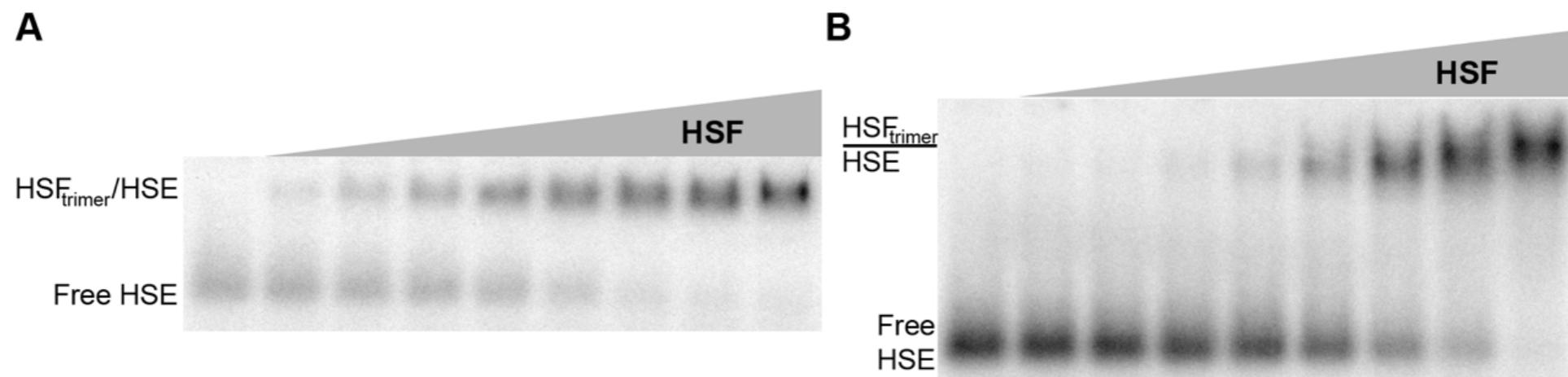


in vitro Binding Assay (PB-seq) reveals all potential binding sites and relative affinities



To transform these relative values into Kd measurements the absolute binding affinities for two genomic HSEs must be measured.

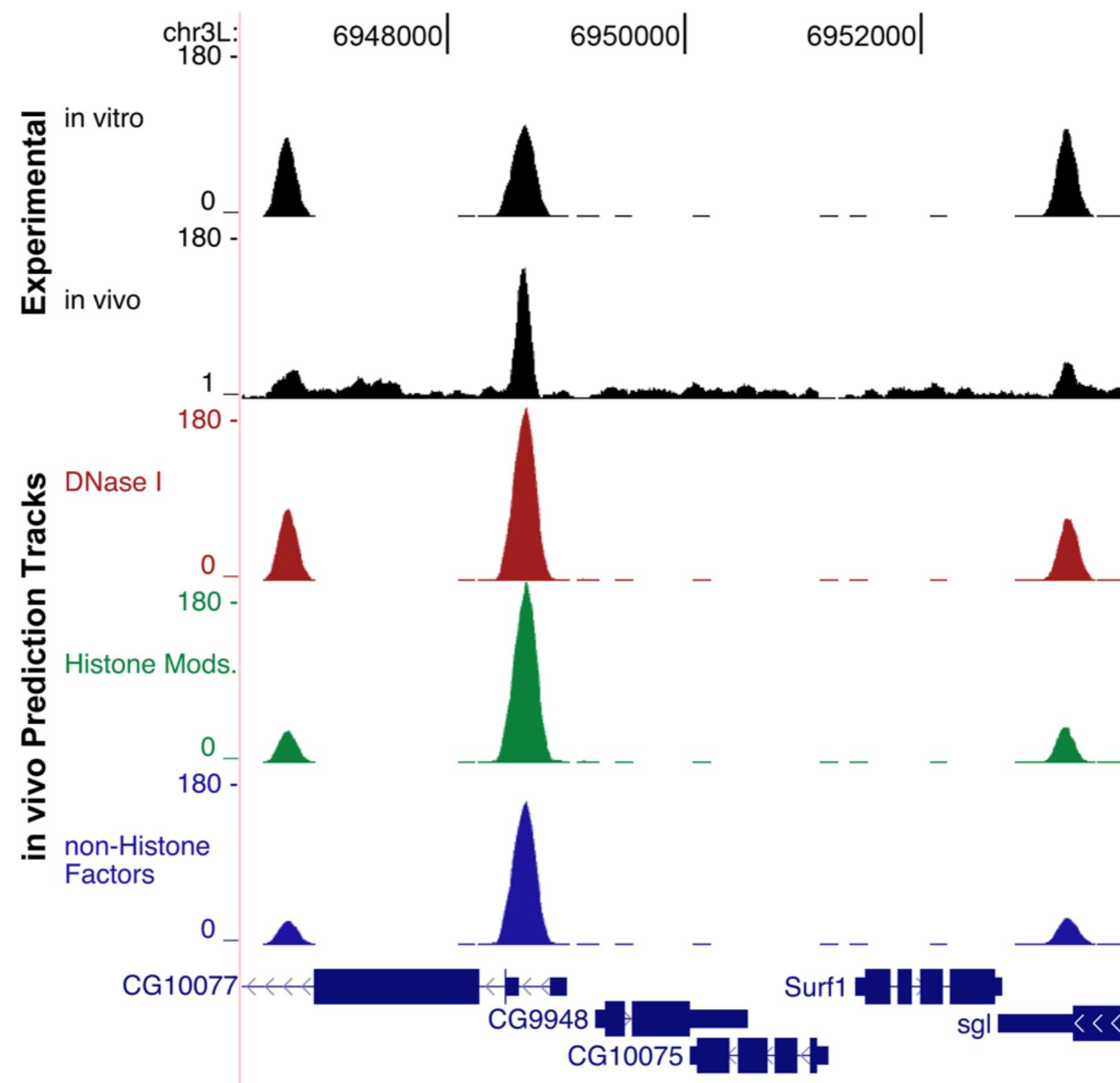
HSF binds to HSEs with picomolar to nanomolar affinity in vitro



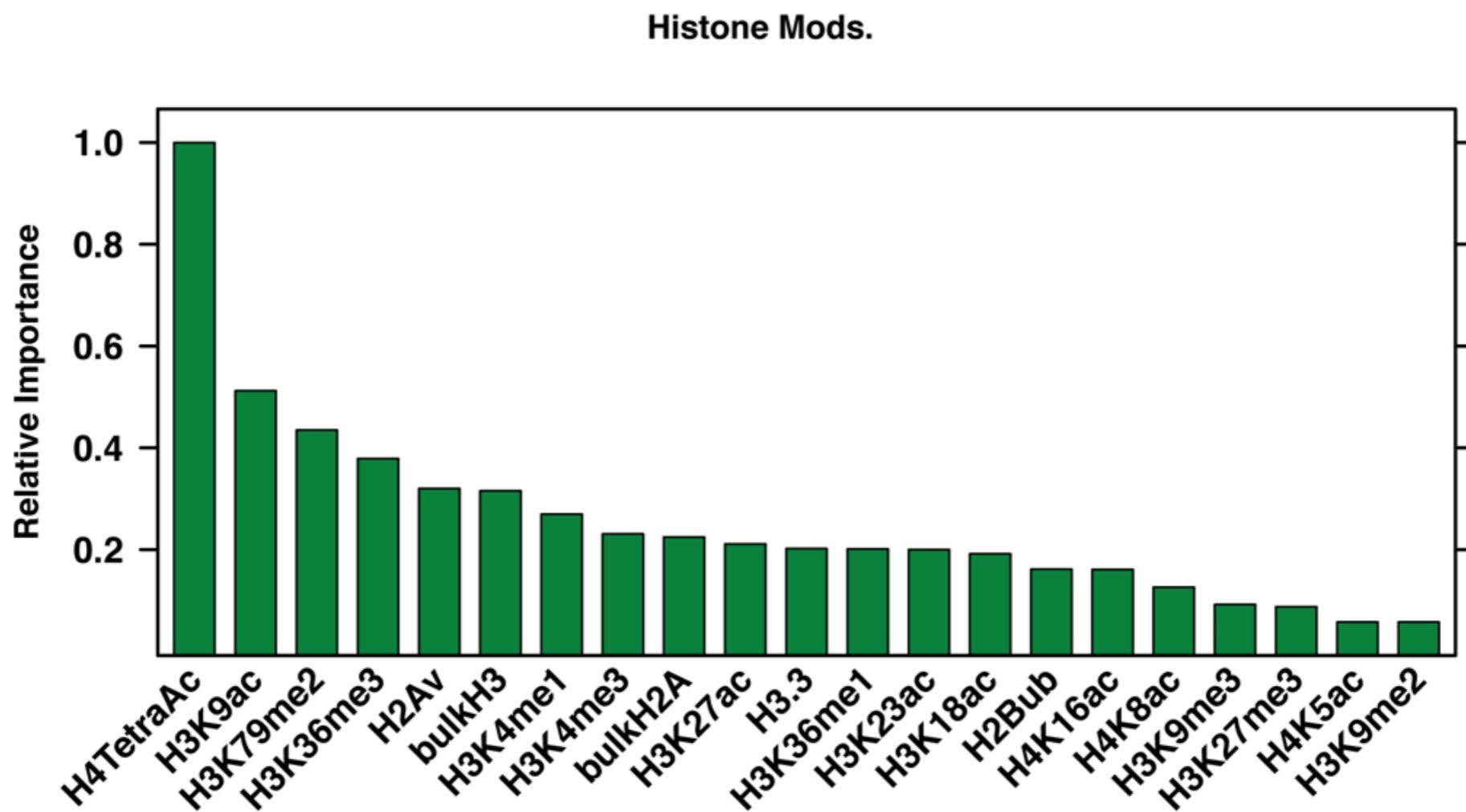
Prediction of Binding Profiles using available PB-seq Data and Genomic Chromatin Data from modENCODE

| | |
|-------------|----------|
| DNase I | H3K9me2 |
| GAF | H3K36me1 |
| H4K16ac | H3K36me3 |
| H4TetraAc | CP190 |
| MNase | SuHw |
| Chro(Chriz) | Ez |
| BEAF | H4K8ac |
| H3K4me3 | H3.3 |
| H3K27ac | H3K23ac |
| H3K9ac | H3K4me1 |
| H4K5ac | H3K9me3 |
| HP1 | CTCF |
| H3K27me3 | bulkH3 |
| H2Bub | bulkH2A |
| H3K79me2 | Pc |
| H2Av | H3K18ac |

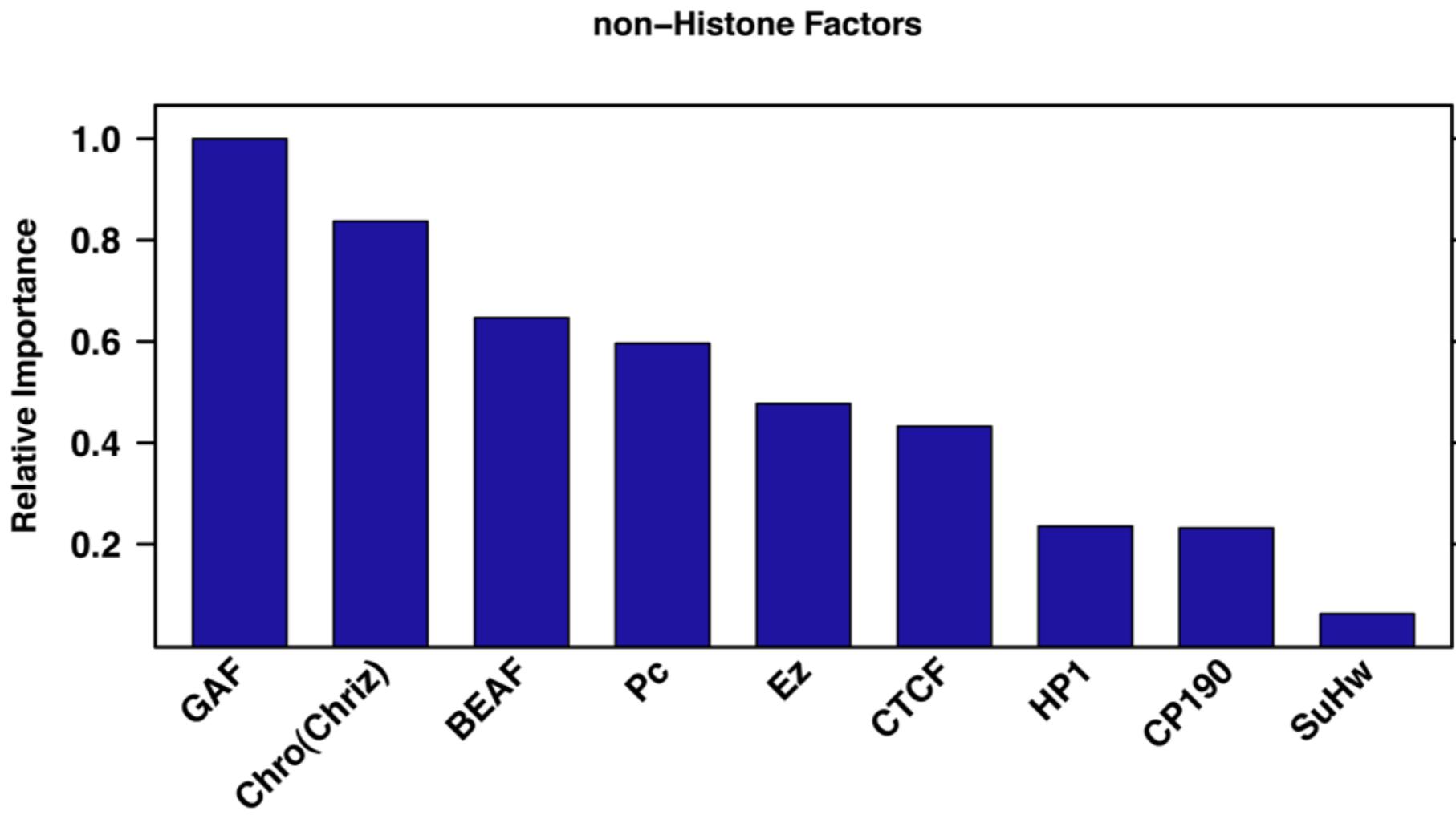
Regression models predict in vivo (ChIP) binding signal using the in vitro (PB-seq) data and NHS chromatin landscape



Histone Acetylation is the most influential modification for predicting HSF binding intensity

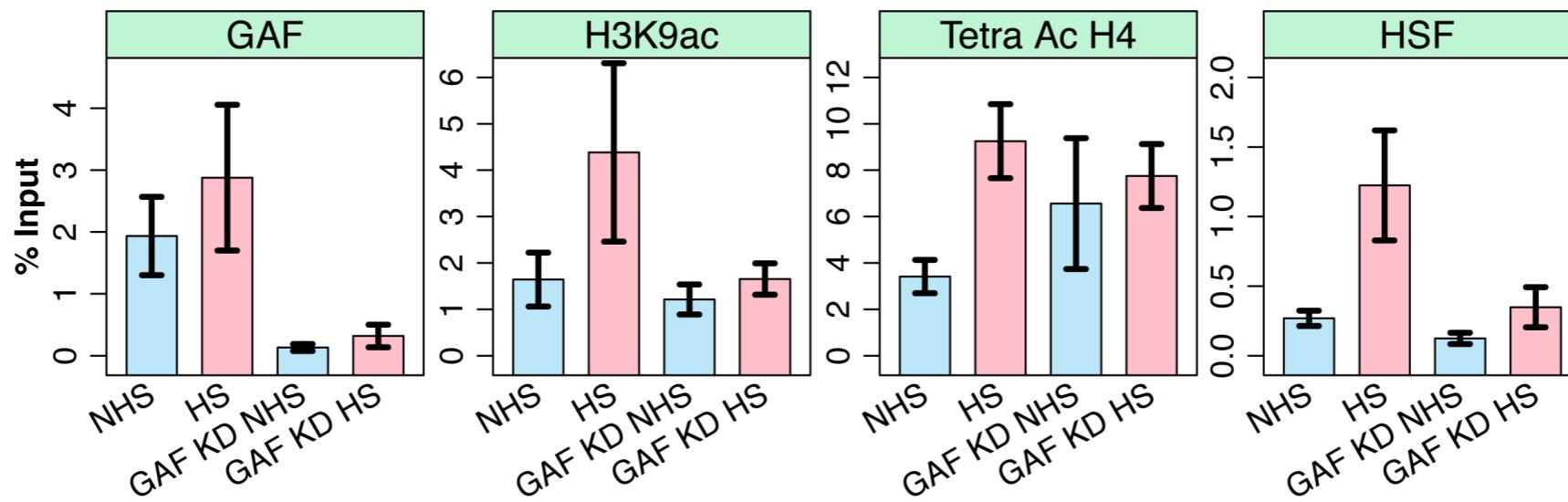


GAF is the most influential non-histone covariate in the predictive model

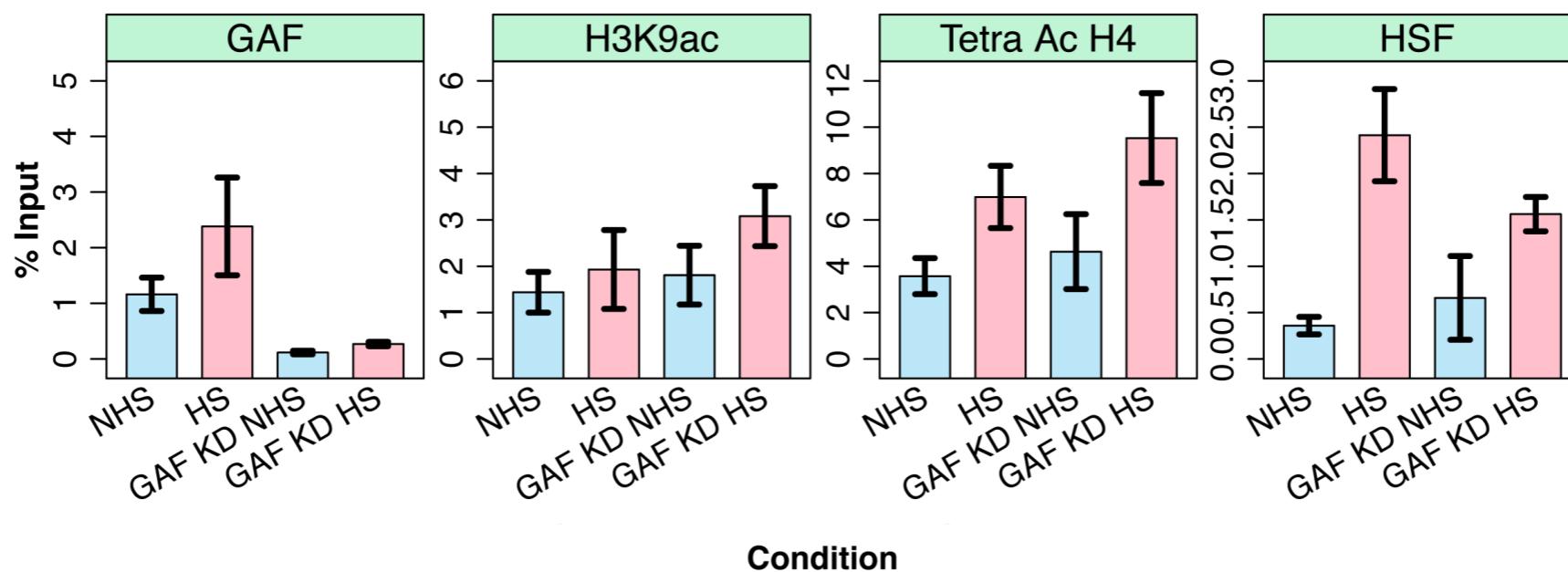


GAF depletion compromises HSF binding intensity

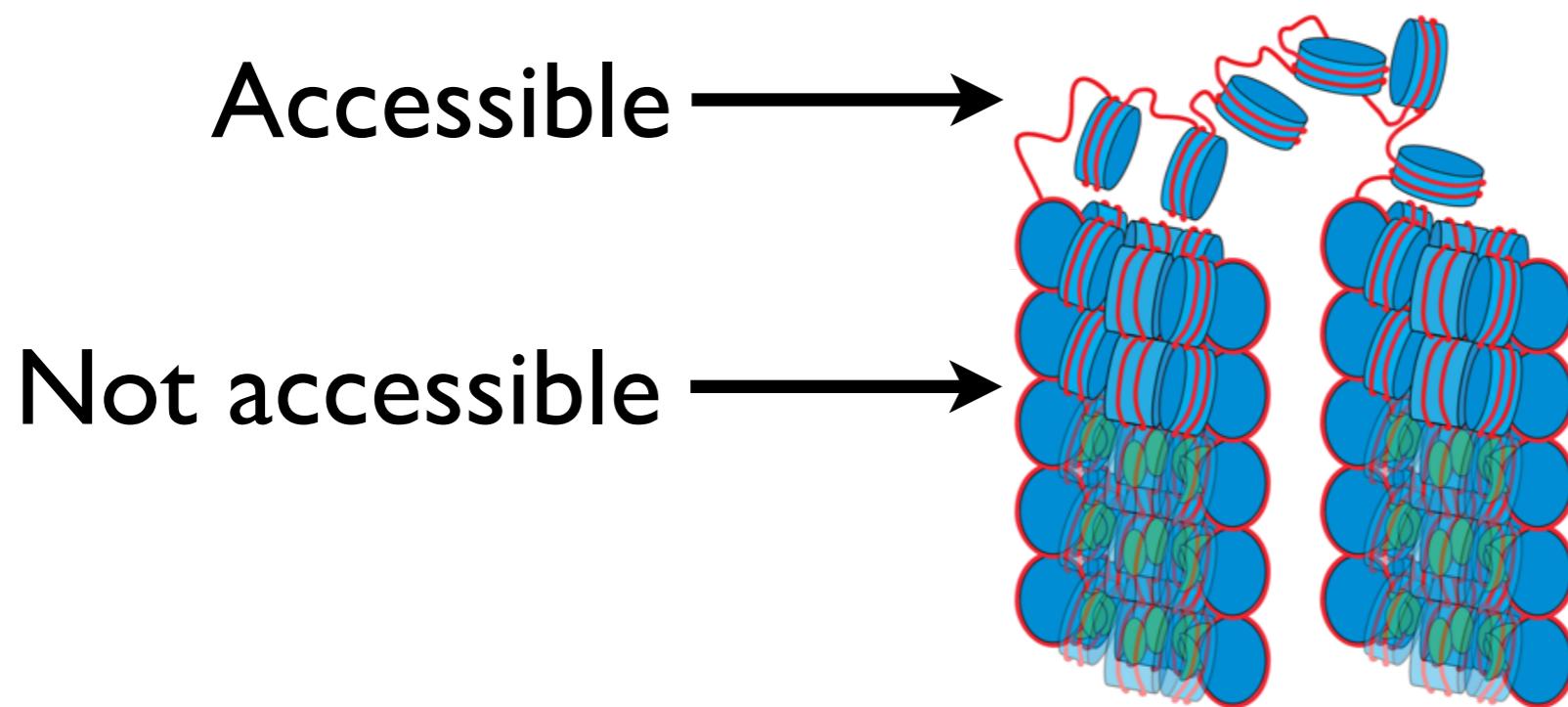
GAF KD ChIP chr2R: 1429426



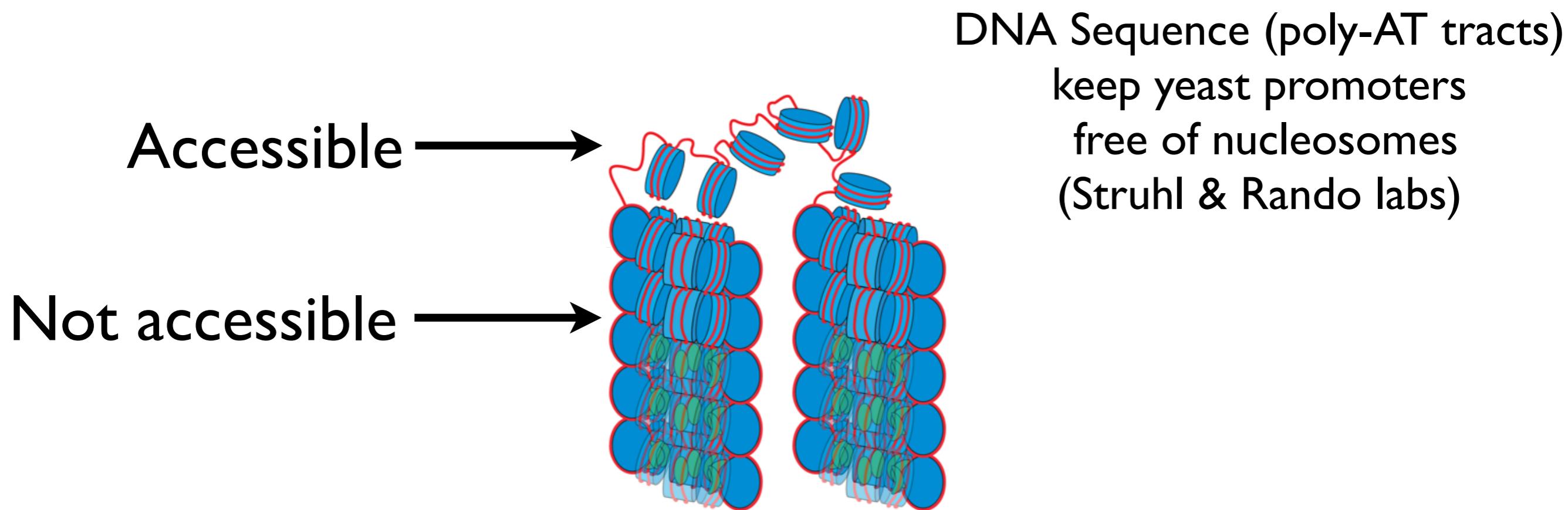
GAF KD ChIP chr2L: 5977789



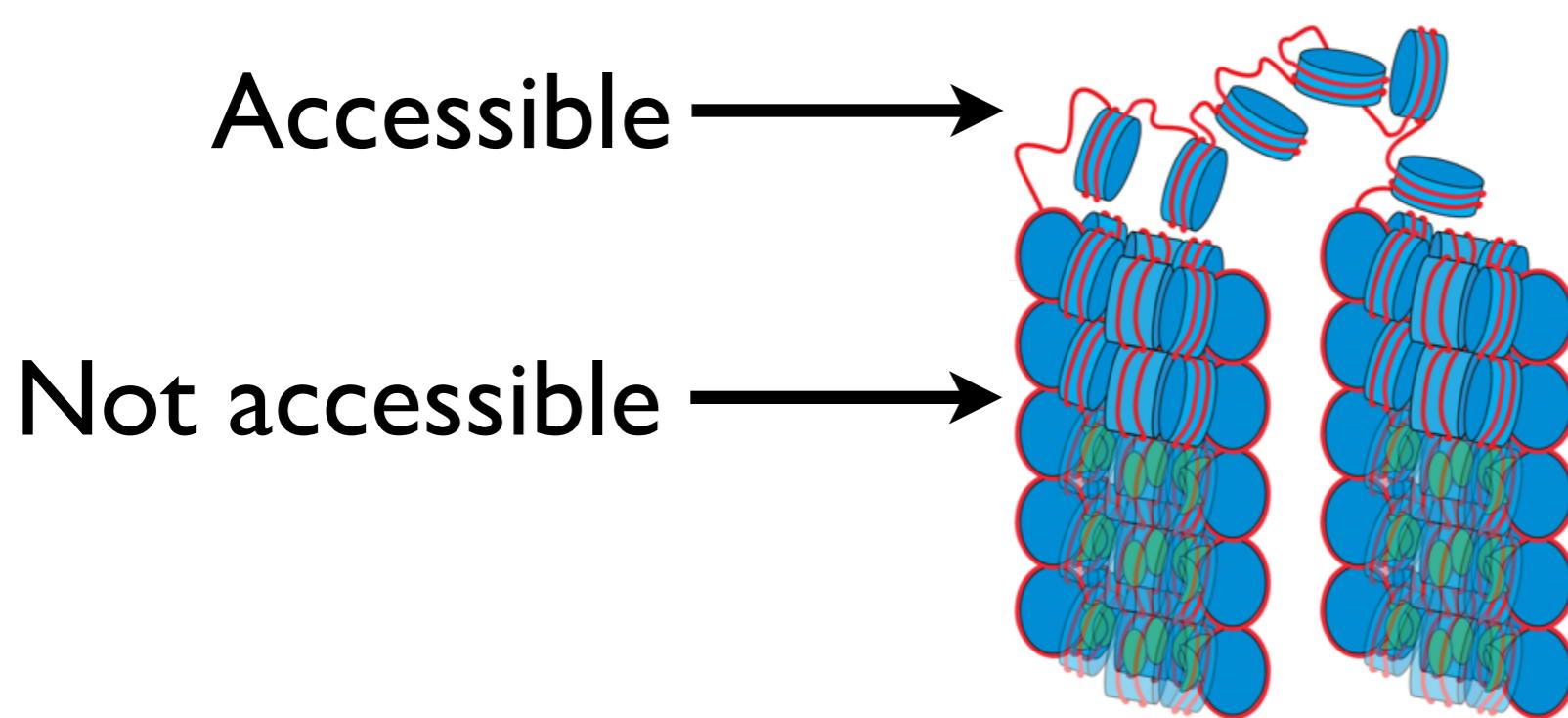
How do accessibility and active marks originate? working model:



DNA sequence directs accessibility



DNA sequence directs accessibility



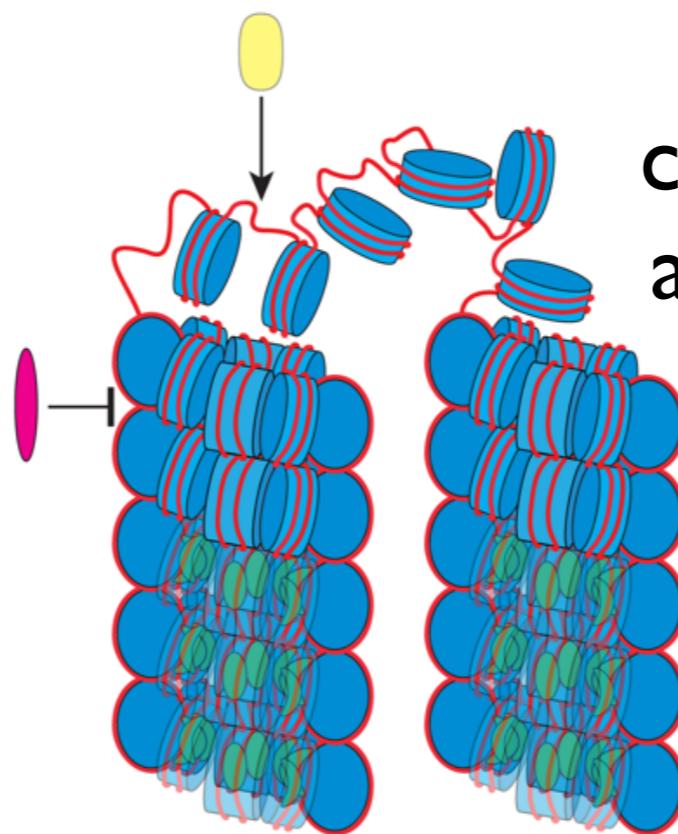
CpG islands favor nucleosome formation in vitro,
but are depleted in vivo.

HI preferentially binds
AT-rich linker DNA.

(Reviewed in Zlatanova and Yaneva,
DNA cell Biol. 1991)

Hypothesis: CpG islands are inherently refractory to higher order compaction by HI, which maintains the chromatin in a transiently accessible state.

Factors target linker DNA between nucleosomes



TFs target uncompacted chromatin and CpG islands are highly occupied *in vivo*.

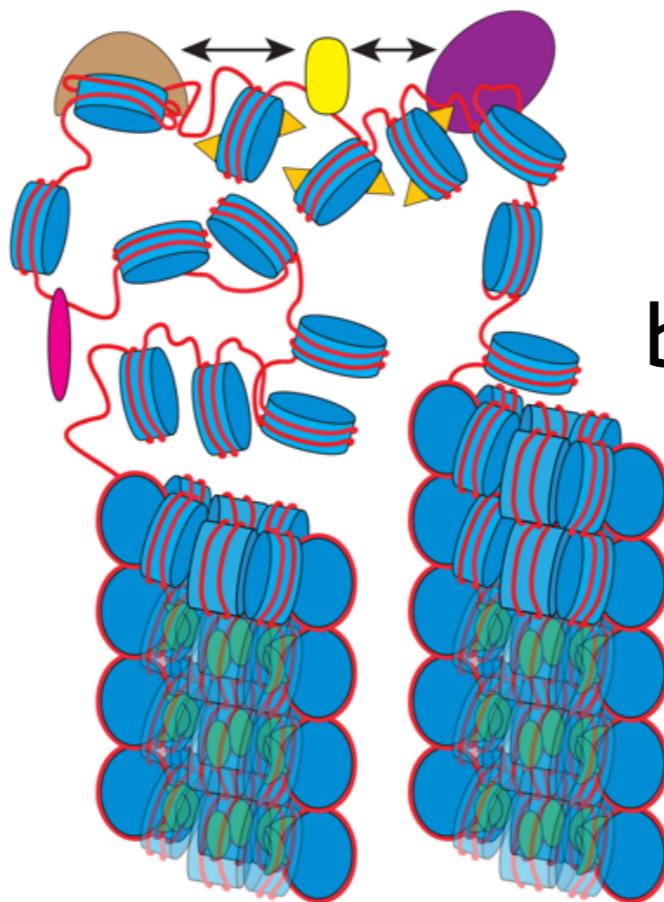
The Set I H3K4 methylation complex and a H3K36 demethylase has been shown to interact with unmethylated CpG-rich DNA *in vitro*.

(Ooi et al. Nature 2007, Zhou et al. Mol Cell Biol 2012)

Mammalian sequence-specific TFs, as a class, have a GC-bias in their cognate binding sites.

(Deaton et al. Genes Dev 2011)

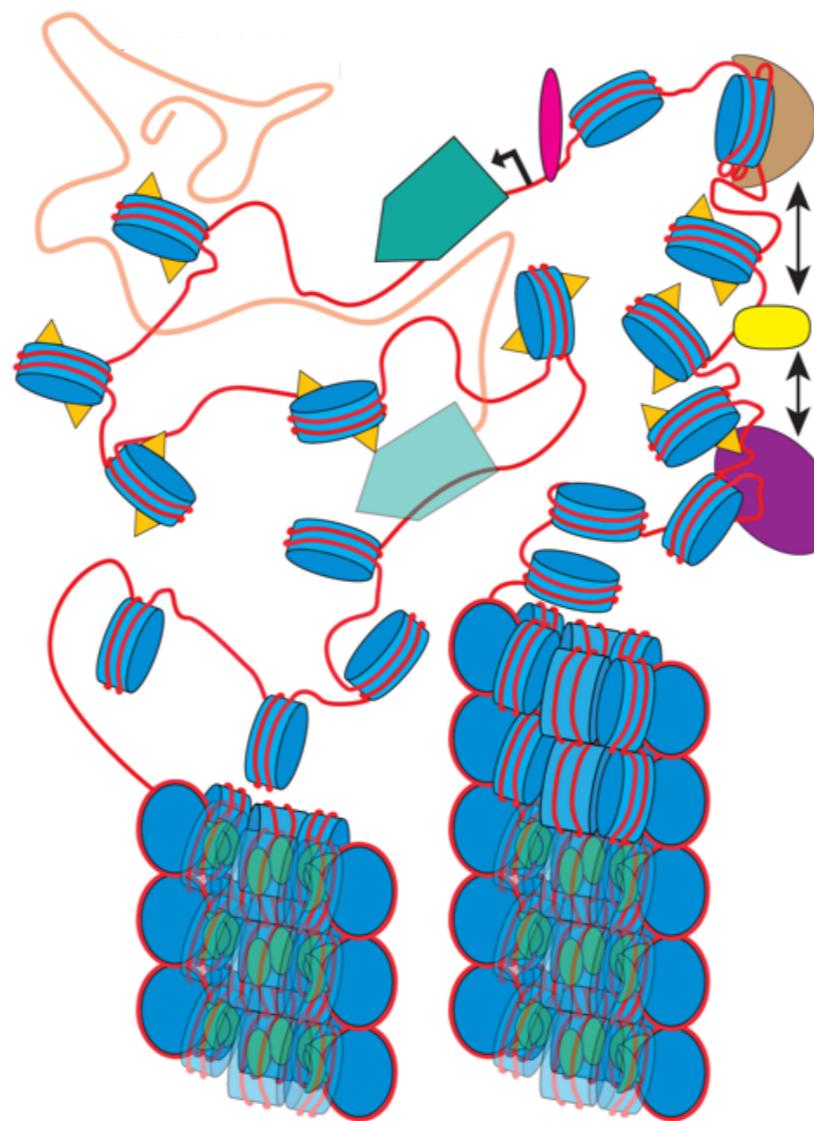
Binding expands accessible regions



TF binding increases
and expands the
boundaries of accessibility
and active marks.

Binding expands accessible regions

Then the pink TF can bind and active **transcription**, which ultimately controls cell fate.



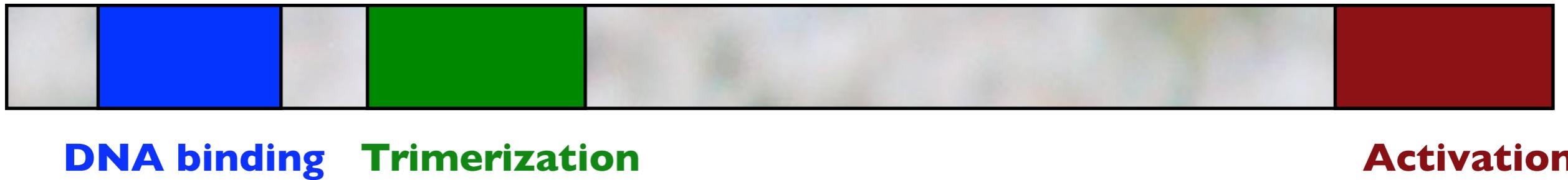
Is there a paradigm shift in biology, away from overly hypothesis-driven research?

- Starting today, here is how I would approach my PhD:
 - Identify a relevant question.
 - Design experiments that are as unbiased (and controlled!) as possible to address this question.
 - Analyze data and look for correlations.
 - Formulate hypotheses from the correlations.
 - Test the hypotheses.
- With an open mind and competently designed experiments/analyses, one can develop hypotheses that were inconceivable at the onset.

Summary: Part II

- Features of double-stranded DNA sequence can provide recognition features for proteins.
- DNA-binding transcription factors (TFs) represent a fairly large fraction of the proteome.
- TFs have domains that bind specific DNA elements and fall into distinct classes.
- These domains employ a variety of strategies to build a molecular protein complement to the DNA element.
- The repertoire of target DNA sequences that can be specifically recognize by these proteins is further enriched by heterodimerization and cooperative interactions.
- Chromatin state dictates TF binding, which can in turn influence chromatin structure

My favorite Transcription Factor: Drosophila HSF



DNA binding and activation domains of transcription factors are distinct and separable

GAL4 - a eukaryotic activator



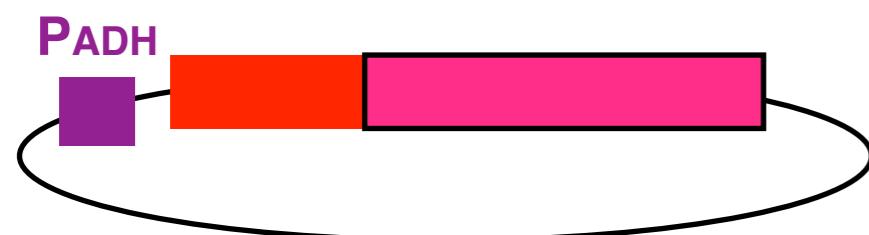
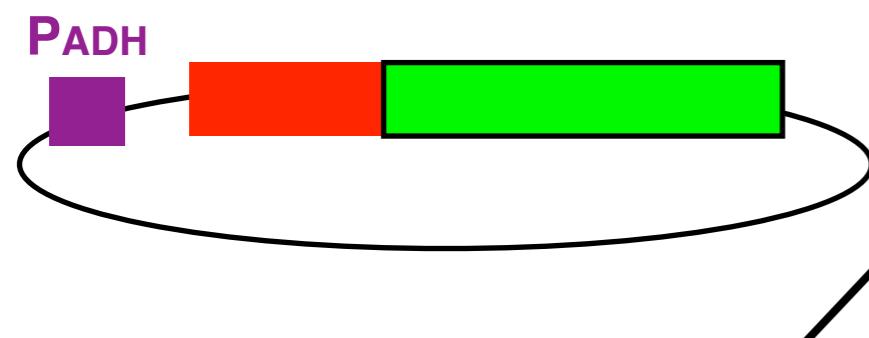
DNA
Binding
Domain to
UAS_G

LexA - a bacterial repressor



DNA
Binding
Domain to
LOP

Cut & Splice & Join
to Expression Vector



Transform Yeast
containing a
Gal1-lacZ
Reporter



Remove and
Substitute the
UAS_G

Δ UAS_G

**Vector Expression
Units of β -Gal**

LexA LexA- GAL4



Transcription Activation Activity can reside in one or more regions



| Construct | Units of β -Gal |
|---------------|-----------------------|
| GAL4 WT | 1900 |
| GAL4 1-147 | <1 |
| GAL4 1-238 | 110 |
| 1-147/768-881 | 110 |
| 1-238/768-881 | 1400 |

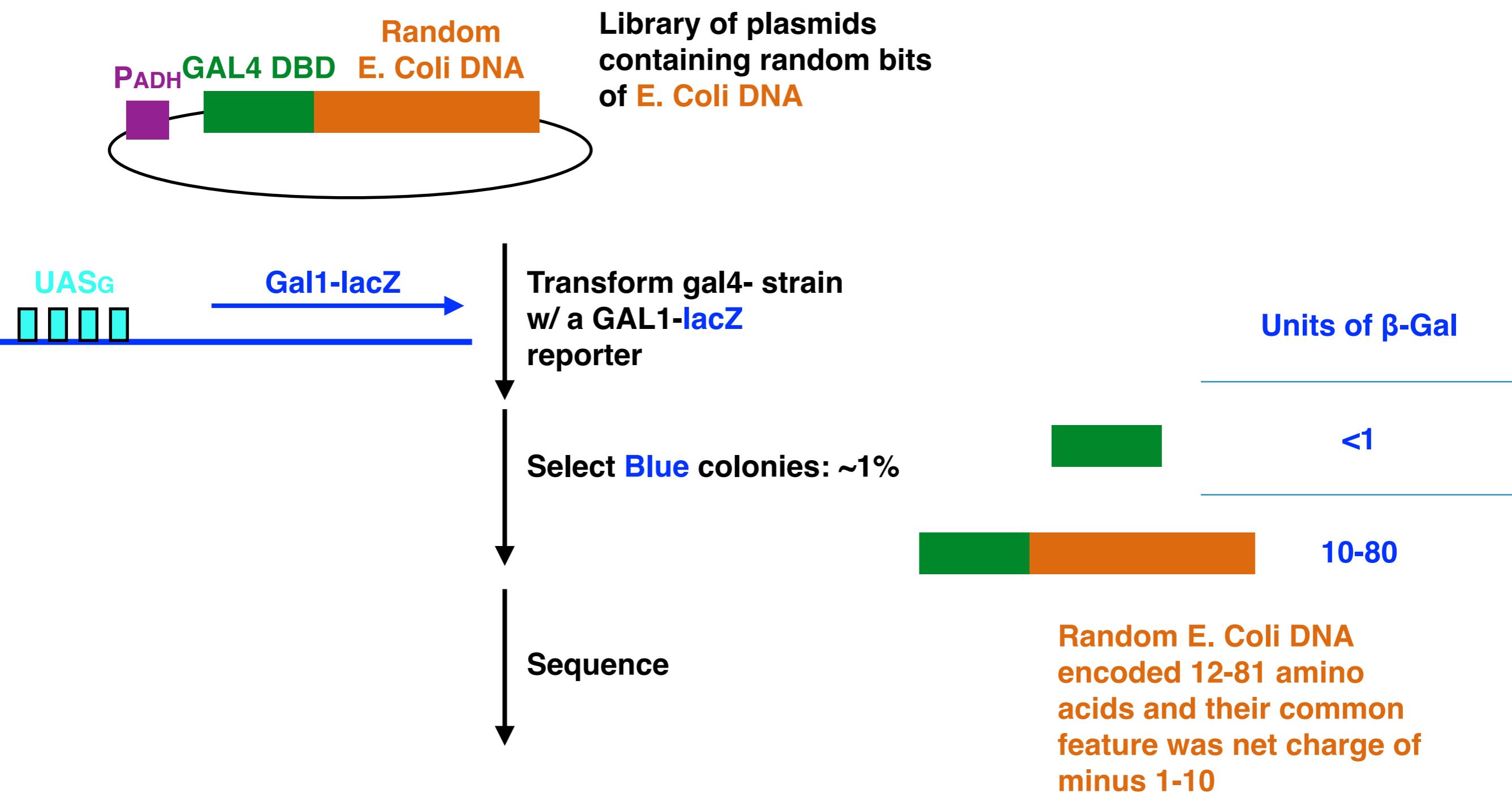
Generate deletions in vitro

Clone into Expression Vector

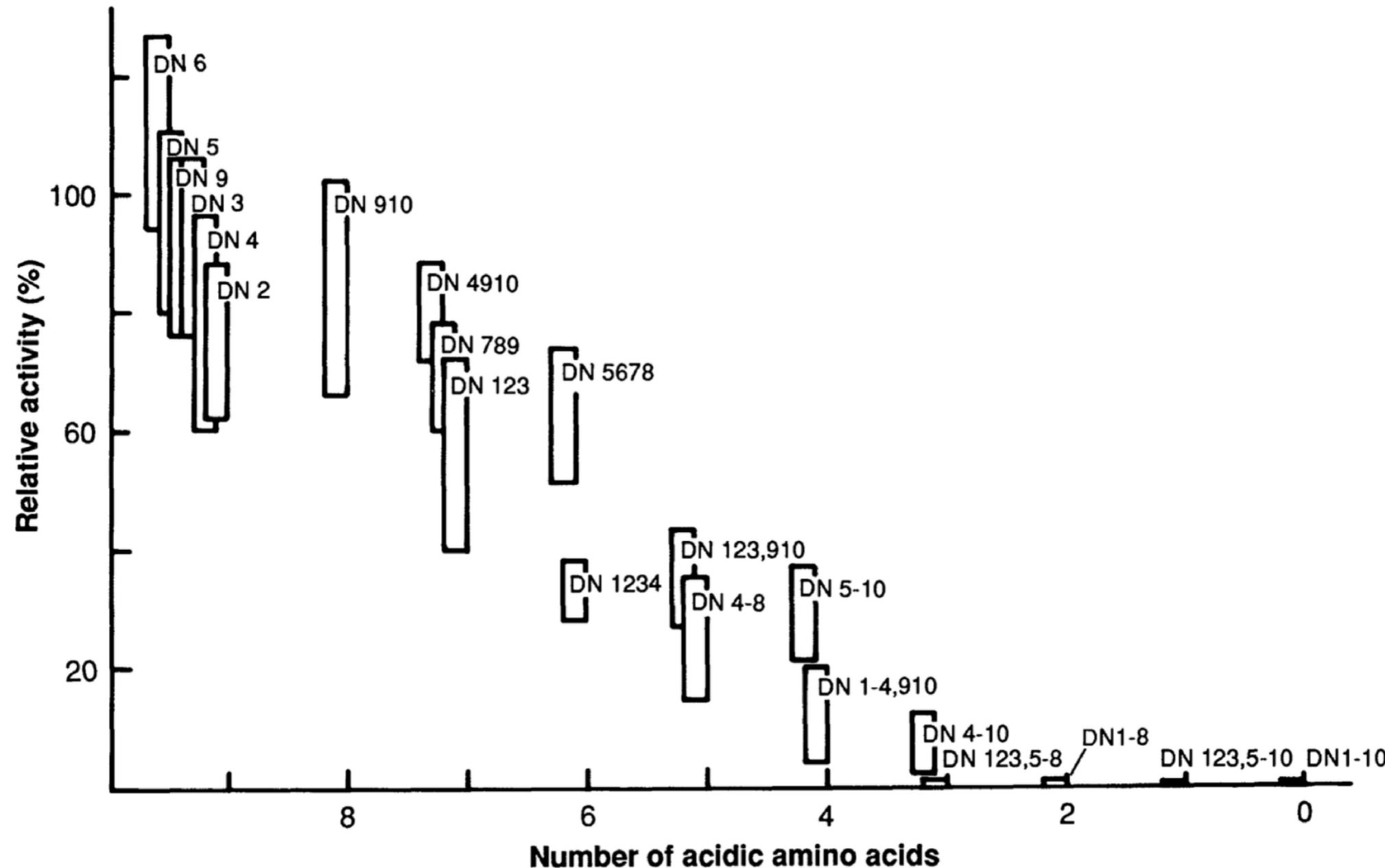
Transform gal4⁻ strain w/ GAL1-lacZ reporter

Downward arrows indicate the progression of the experimental steps from top to bottom.

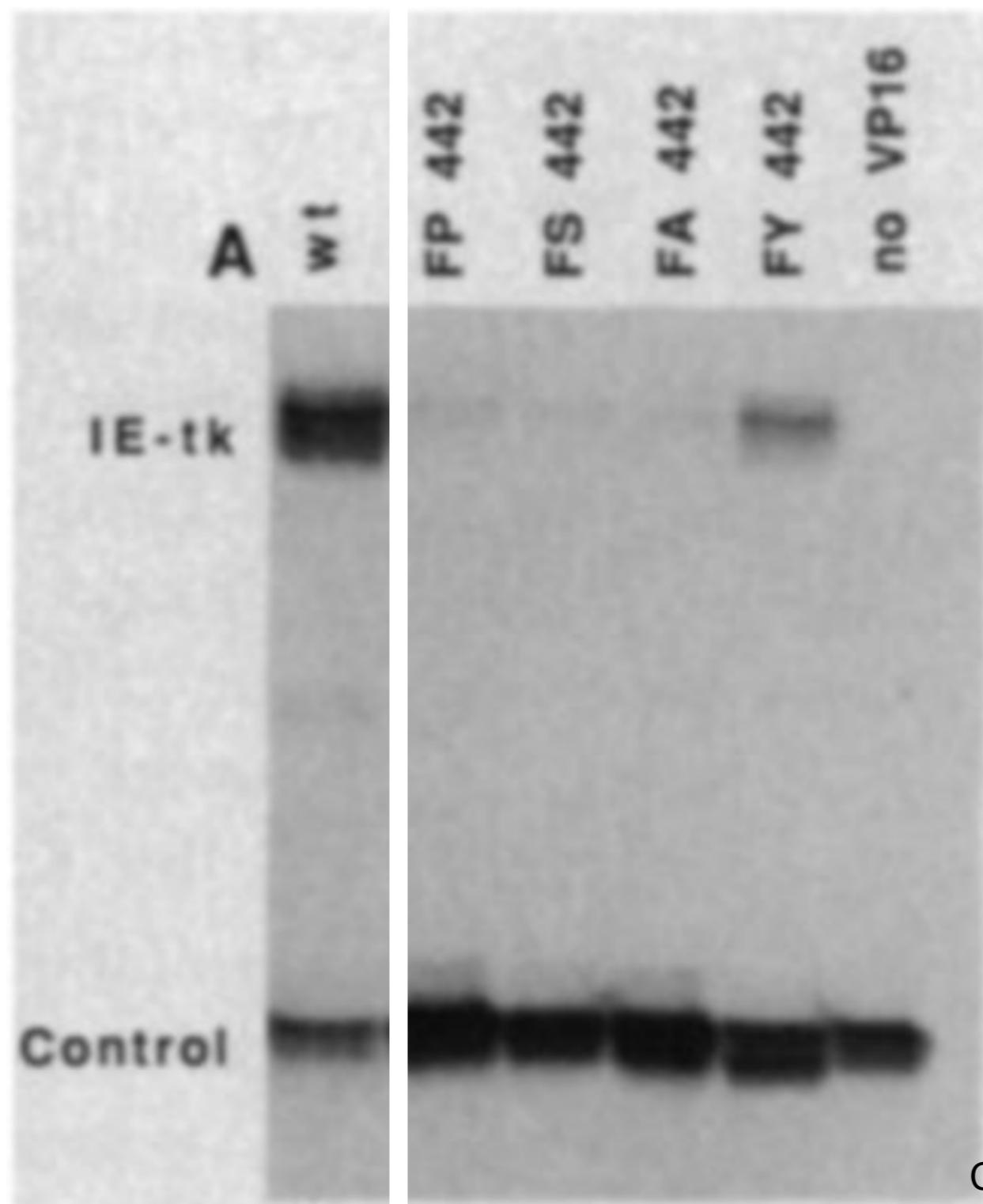
Transcription activation regions occur frequently and are often acidic



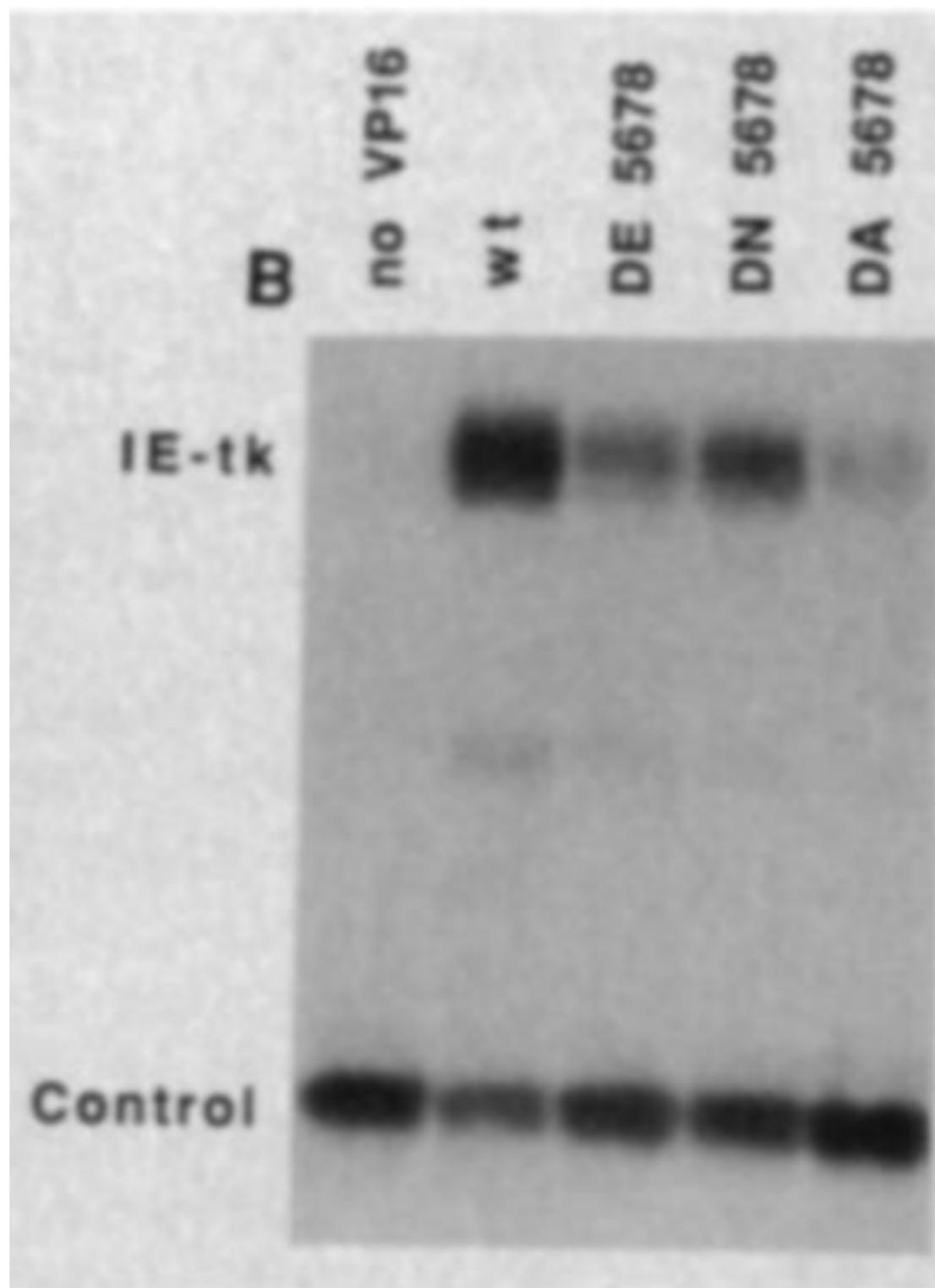
Negative charge correlates with transcriptional activity of VP16



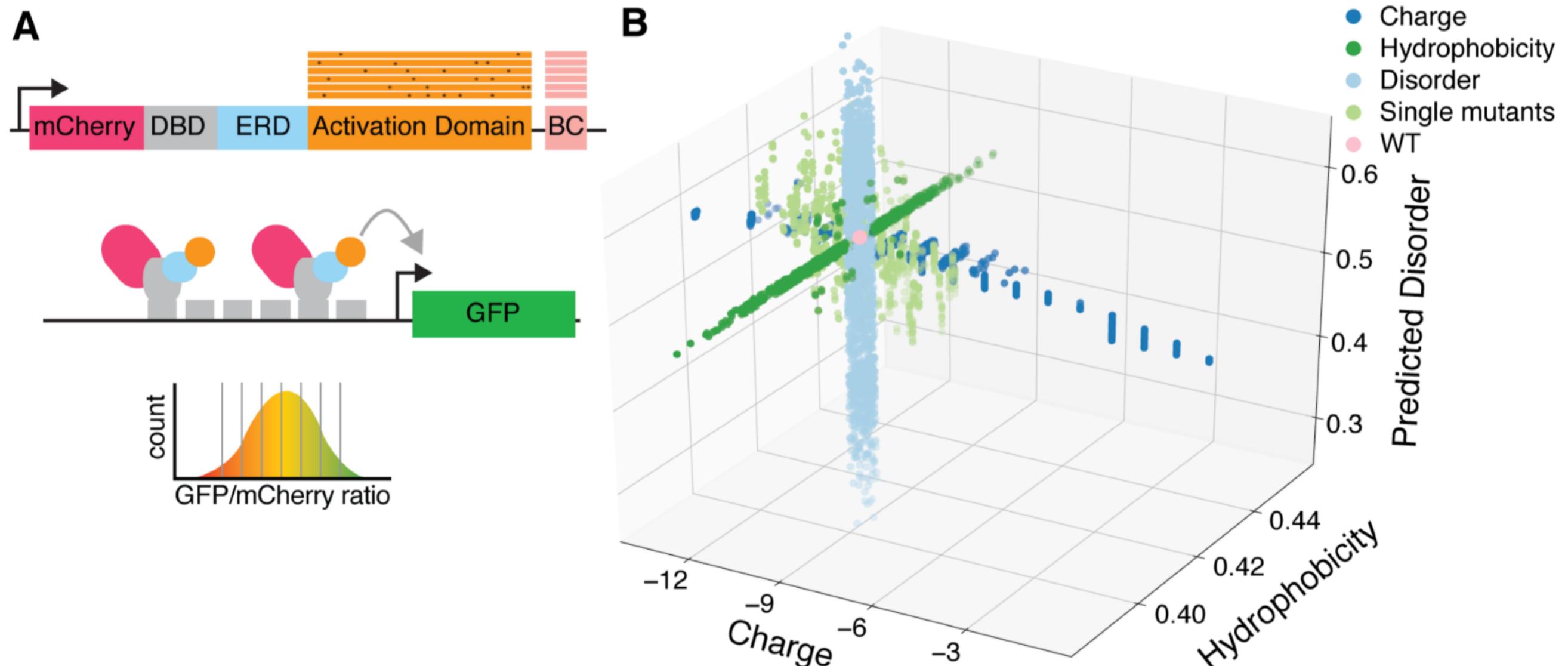
A single Phenylalanine is critical to VP16 function



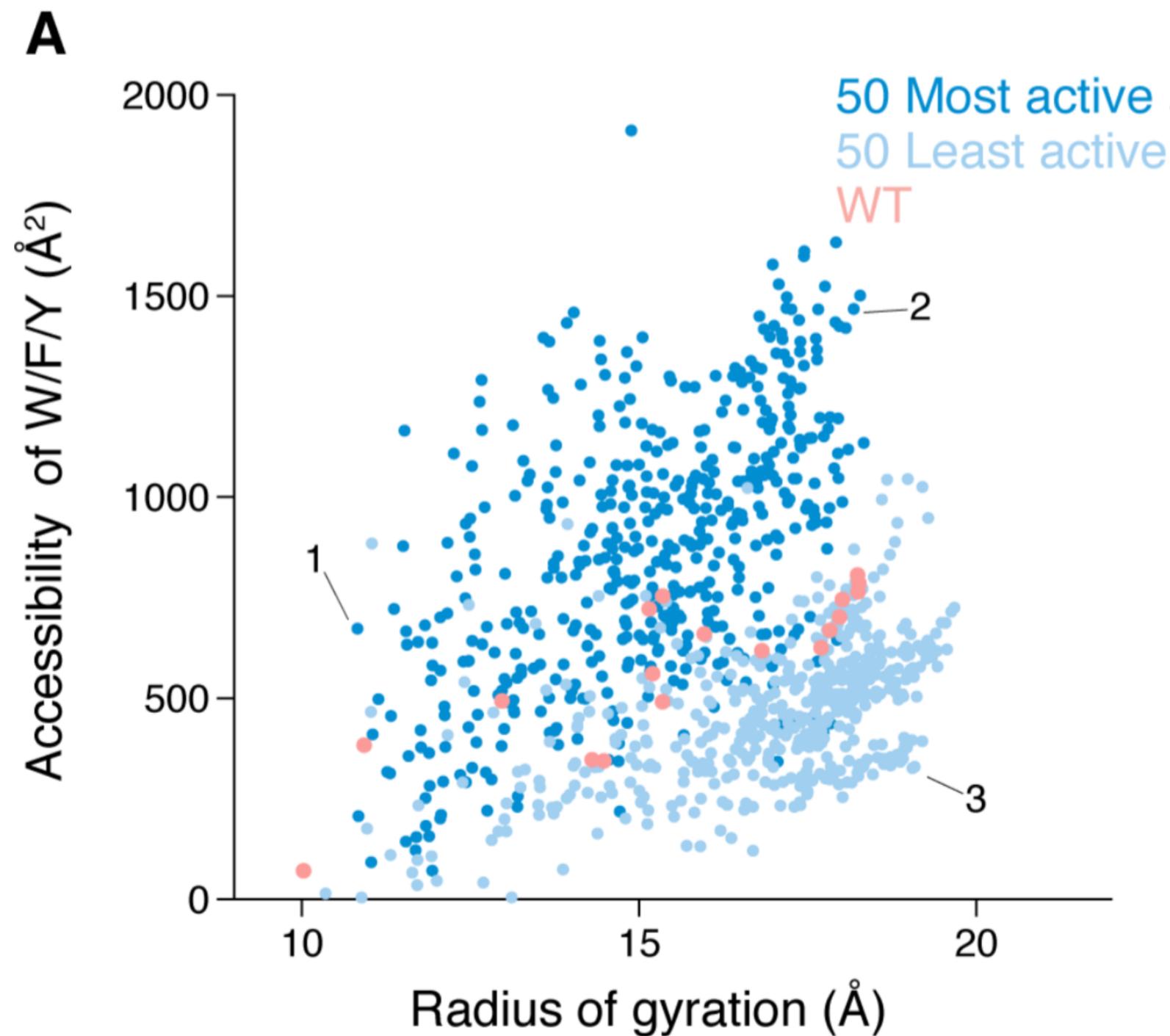
The environment surrounding Phe⁴⁴² affect VP16 function



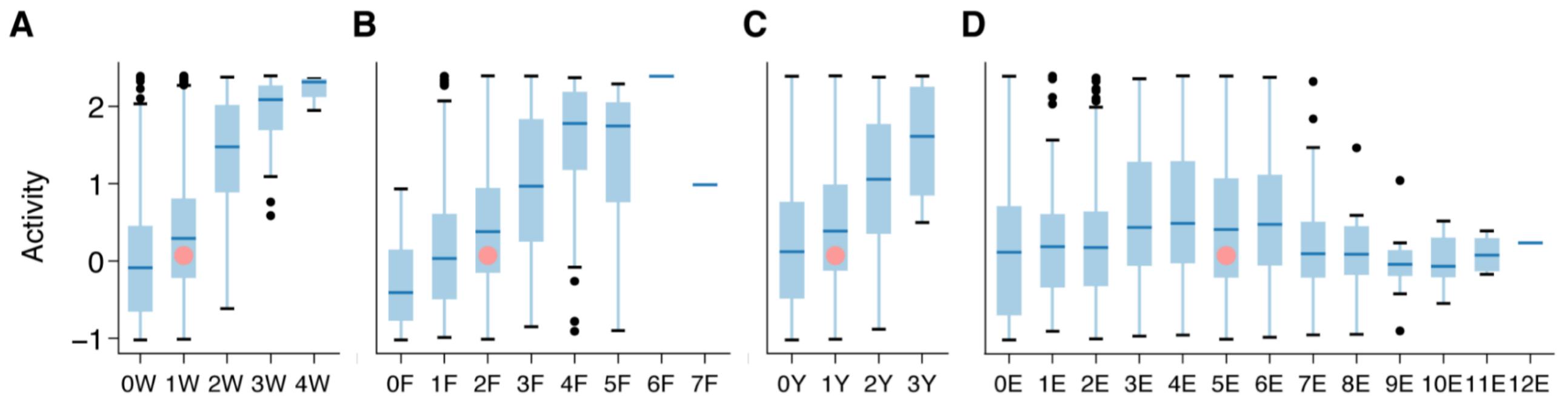
Measuring thousands of designed GCN4 activation domain mutants in parallel



Simulations reveal that highly active variants keep aromatic residues exposed to solvent



Aromatic residues control Gcn4 Activity



Conclusions: acidic residues regions keep two hydrophobic motifs exposed to solvent to mediate activity.

Activation domains come in several flavors

- SP1 - Q-rich domain (polar)
- CTF - P-rich activation domain (hydrophobic)
- NTF1 - I-rich activation domain (hydrophobic)
-

Summary: Part III

- DNA-binding transcription factors (TFs) often have distinct and separable domains (DBD and activation)
- Hydrophobic and Acidic residues are often critical for TF activation function, perhaps acidic residues keep hydrophobic solvent-exposed.
- There are many types of activation domains

Part IV

- Most TF binding events do not result in changes in gene expression.
- As a corollary, just because something exists does not mean it is functional.
- Too often people ask the question “what is the function of X”, when there is no evidence that X is functional.
- A contemporary example is lncRNAs.

How can we identify which cofactors interact with your favorite activation domain?

- Conventional chromatography of Mediator: Näär, et. al., *Science*, 1999; follow up work identified the KIX domain of Med15: Yang, et. al., 2006.
- immobilized template and label transfer: Fishburn, et. al., *Molecular Cell* 2005.
- Unbiased approach: BiolD or APEX tagging of TFs —Roux, et. al., *J Cell Biol.* 2012; Lam, et. al., *Nature Methods* 2014 (there are newer versions of each)