BI 8 LECTURE 15

COMBINATORIALITY IN *CIS* AND *TRANS*: TRANSCRIPTIONAL CONTROL IN MULTICELLULAR EUKARYOTES

Ellen Rothenberg

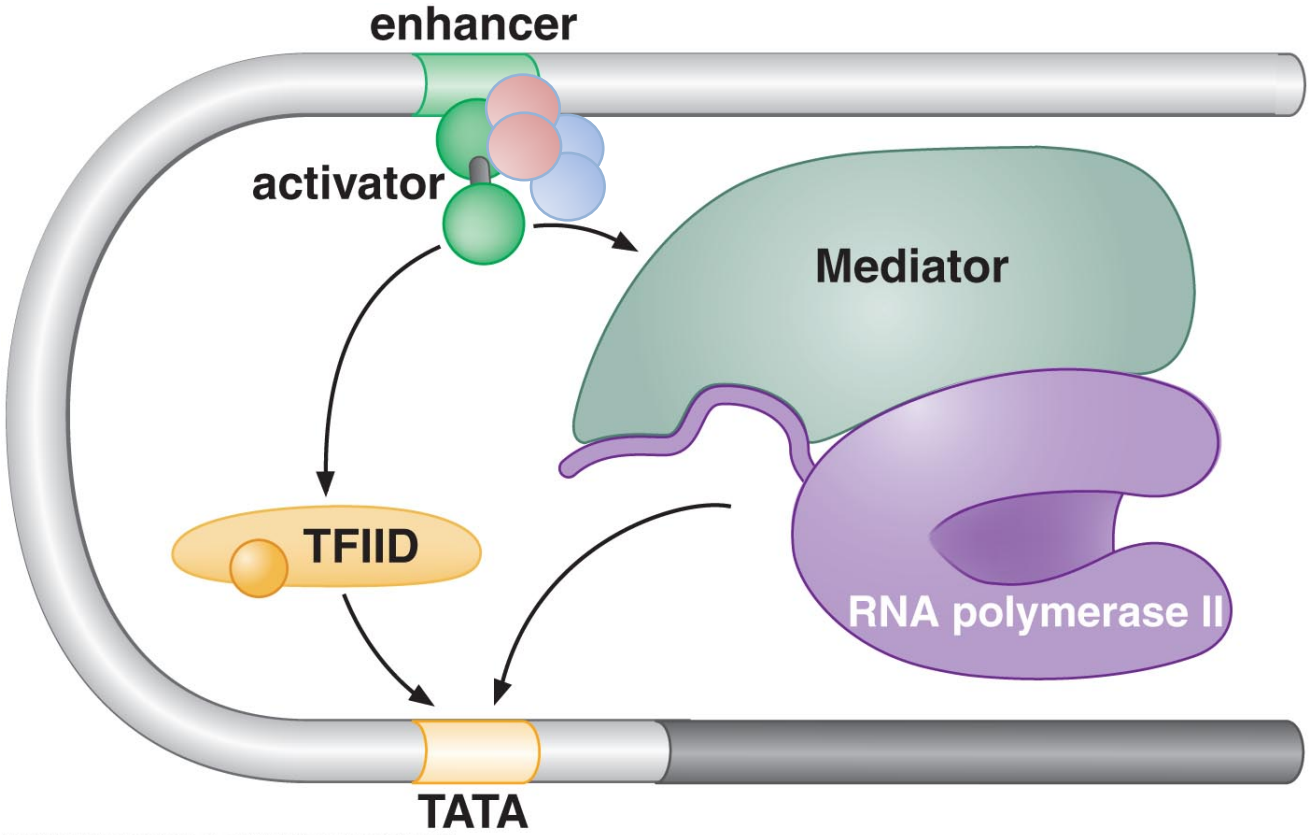
23 February 2016

Chapter 7 and suggested reading: Spitz & Furlong 2012

In eukaryotes…

Real TFs often work by looping enhancers to promoters: help to recruit RNA polymerase II via Mediator Complex

can be direct, or via coactivator binding to activation domain



Different factors work *together* to define cell identity & activate correct genes

- Factors turned on via cell history
- Factors turned on by signals from outside cell

Key to eukaryotic, multicellular life: transcriptional control that responds correctly to multiple conditions for expression

- Default for most genes is not just “on”
- Many genes are used in various tissues, but under different circumstances: “**OR**” logic is crucial
- Transcriptional control has to distinguish not only presence/ absence of one nutritional factor, but ***COMBINATIONS*** of circumstances
- Transcriptional control has to be a logic processor

Changing regulatory state: Multiple ways that transcription factors can respond to signals received by a cell

- Transcription factor binding or activity is modified (+ or -) by phosphorylation (enzyme travels to nucleus)
- Transcription factor can travel from cytoplasm to nucleus
- Cofactor for transcription factor can travel from cytoplasm to nucleus – can toggle transcription factor between activating & repressive function, etc.
- New transcription factor is synthesized in response to signal (as a result of all of the above!)

How do we know???

Functional analysis of TFs and cis-reg target elements requires a ***transfection system*** **…**to put defined genes and defined factors into a "clean" environment

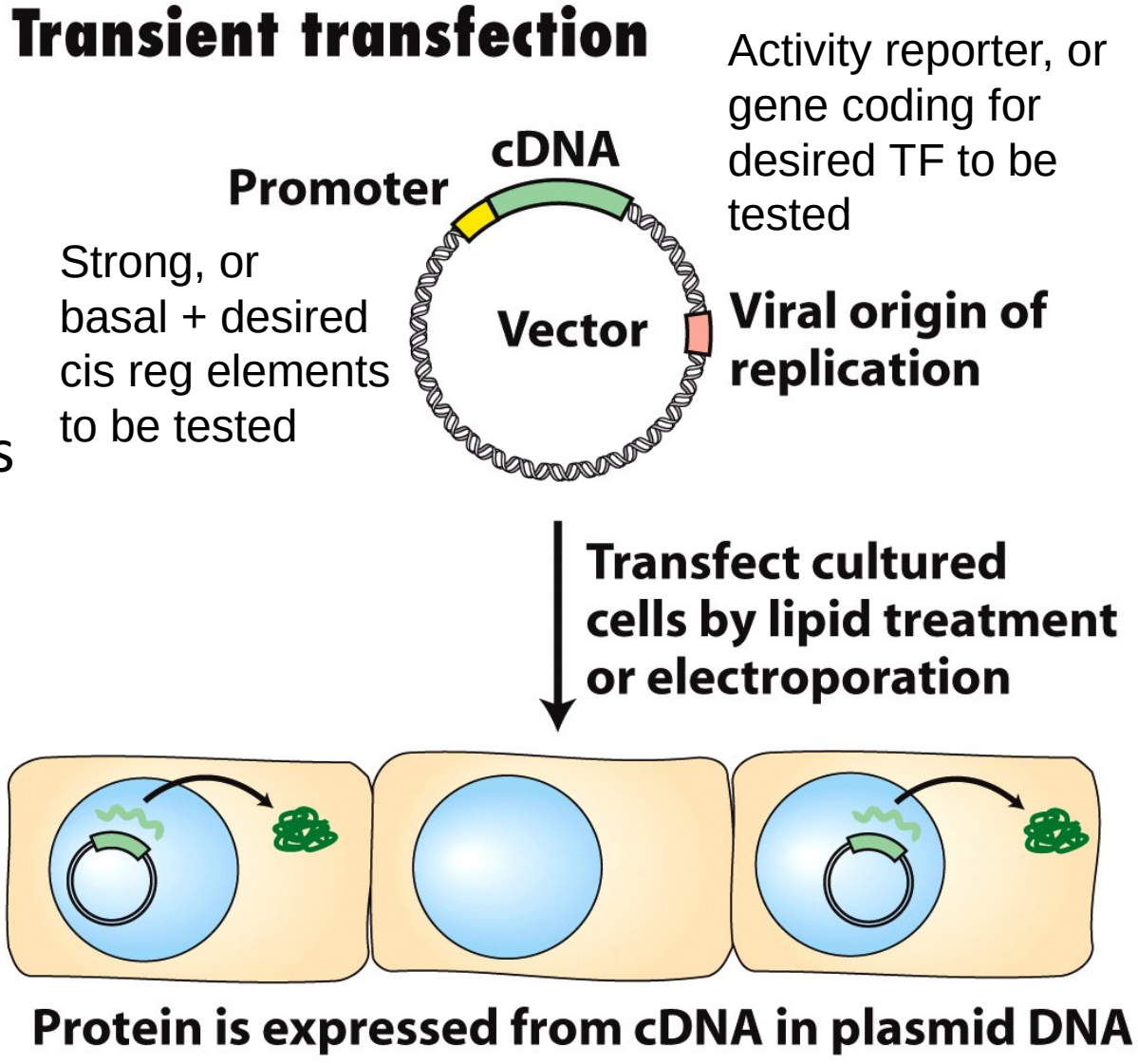


Figure 5-32a
Molecular Cell Biology, Sixth Edition

A plasmid coding for a ***TF*** can be transfected at the same time as the one testing its possible ***target site***, to test if expression is now enhanced

To study regulatory interactions on a gene integrated into chromosome (with histone marks, etc.), you can select for stably transfected cells

only integrated genes will persist through many rounds of replication

Combine genes of interest with gene coding for drug resistance enzyme and make cells grow a long time in presence of drug…

Surviving clones will contain transgene stably incorporated into host DNA

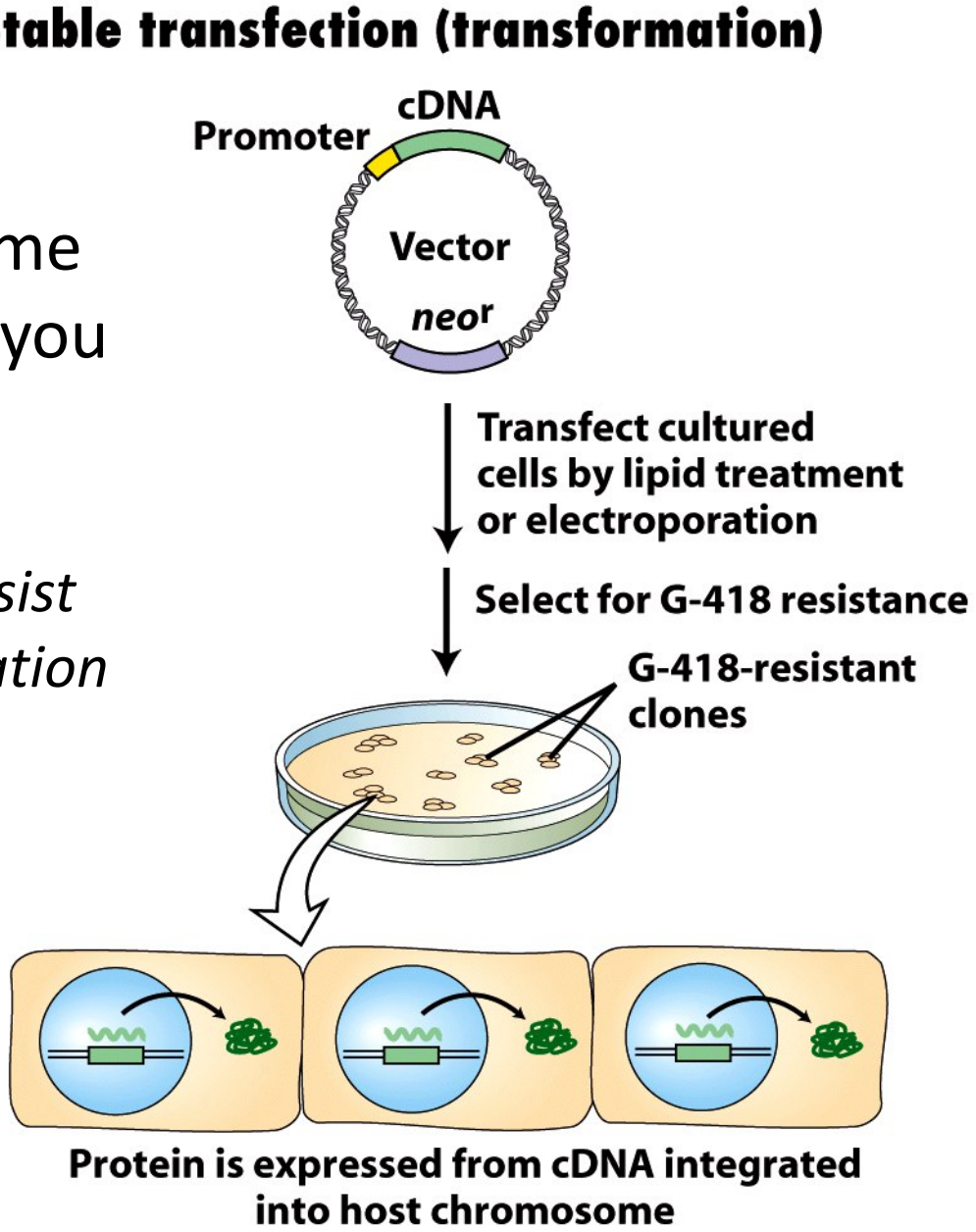
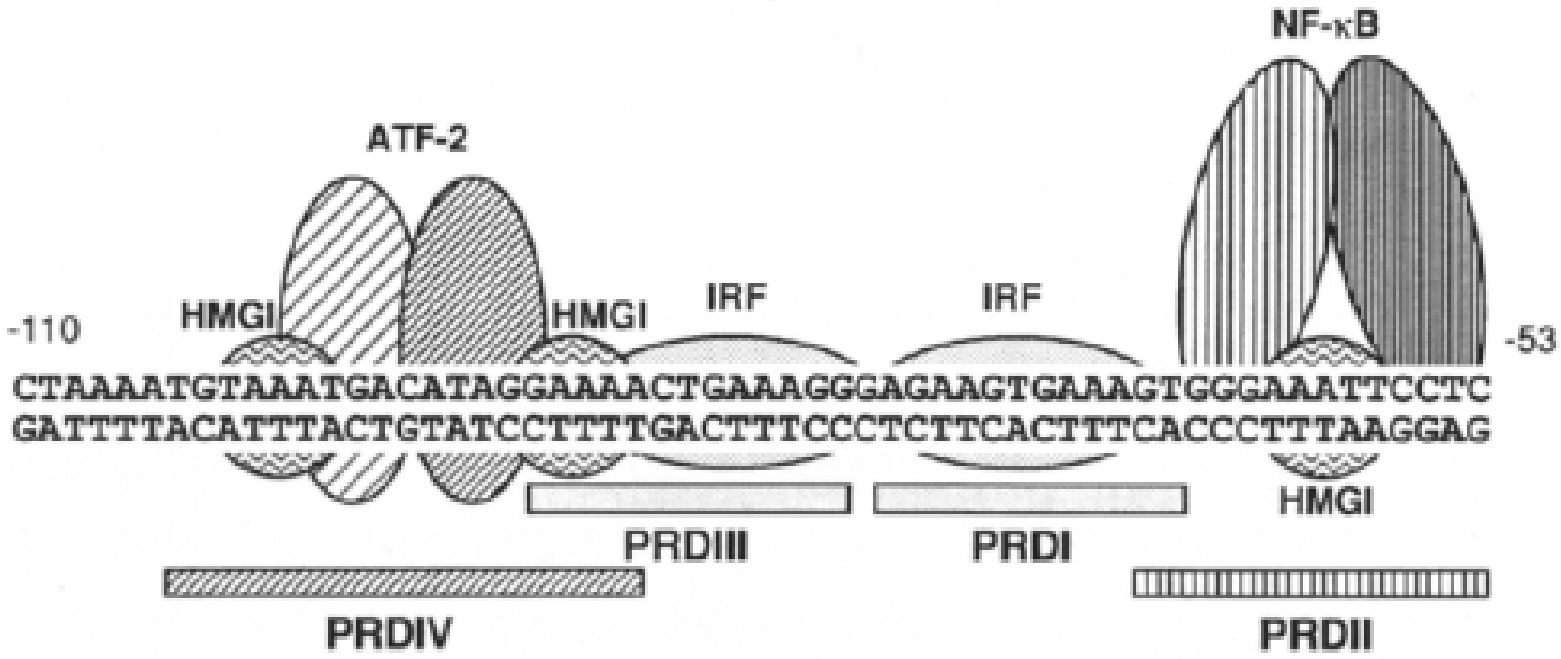


Figure 5-32b
Molecular Cell Biology, Sixth Edition

The IFNβ enhancer: "computation" of physiological triggering conditions by combinatorial transcription factor binding to a single *cis*-reg element

- Interferon-β is crucial antiviral protein produced by cells in innate immune system in response to infection
- Even in the right cell type, this gene should be off most of the time
- Gene needs to "know" exactly WHEN to turn on, and how strongly

The IFNβ enhanceosome complex components: contributing transcription factors and their sites



Logic processing by an ultra-compact mammalian cis-reg element

Crude initial deletion studies previously broke up the IFNβ promoter region into four domains (“PRD”), using footprinting

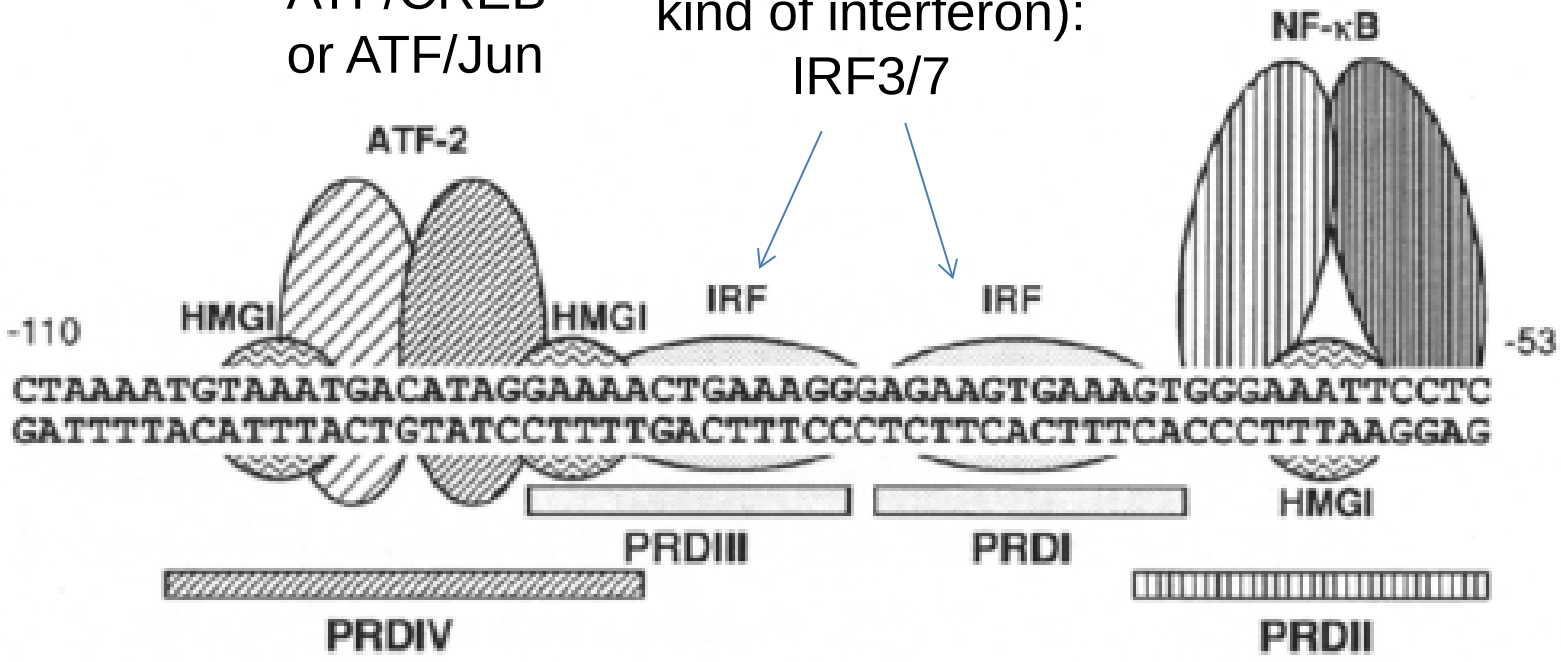
HMG I(Y)= Hmga1 in new nomenclature

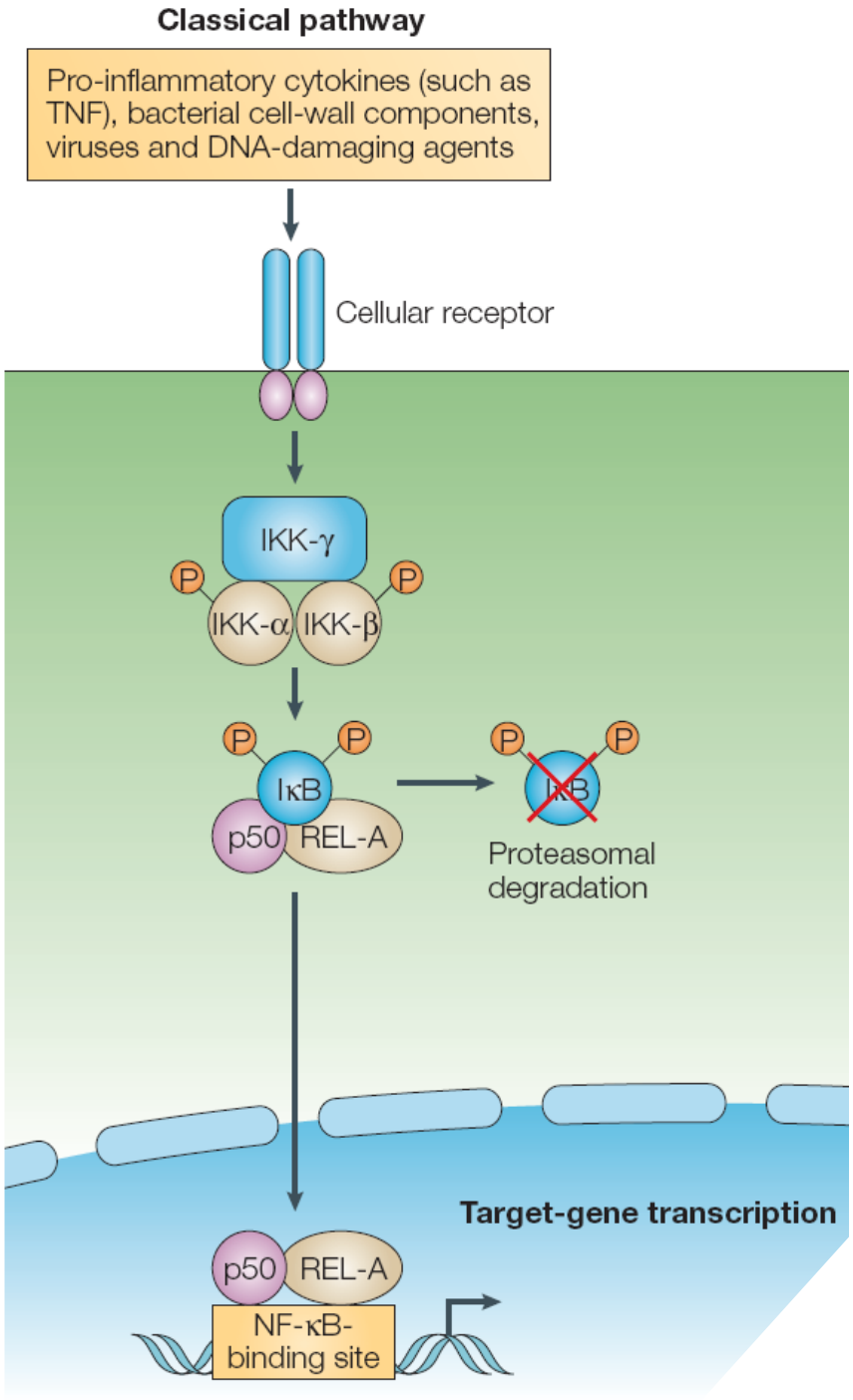
The IFNβ enhanceosome complex components: induction-dependent transcription factors

Inducible (by cAMP): ATF/CREB or ATF/Jun

Inducible (by a different kind of interferon): IRF3/7

Inducible (by TNFα): NFκB



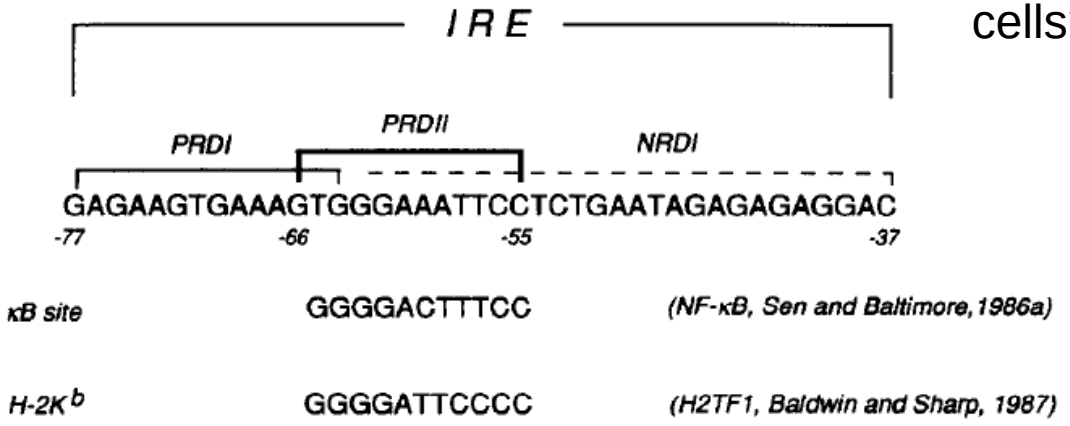


NF-κB is made long in advance of use, but held in cytoplasm by a tethering complex (IκB) until released to go to the nucleus by signaling

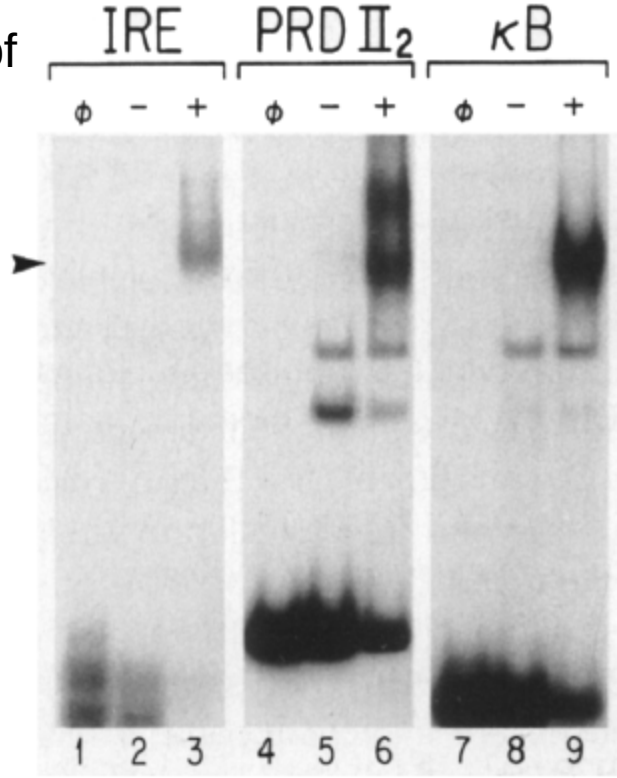
Cell exposure to cytokines, immune signals, bacteria, etc. cause breakdown of tethering complex

NF-κB transcription factor, diagnosed by DNA binding, is activated in immune cells by stimulation (+)

In vitro binding assays, using IFNβ enhancer element PRDII in parallel with another well-studied NF-κB target site:
electrophoretic mobility shift shows stimulation needed for appearance of site-binding activity in nucleus (nuclear extracts)



Stim of cells?

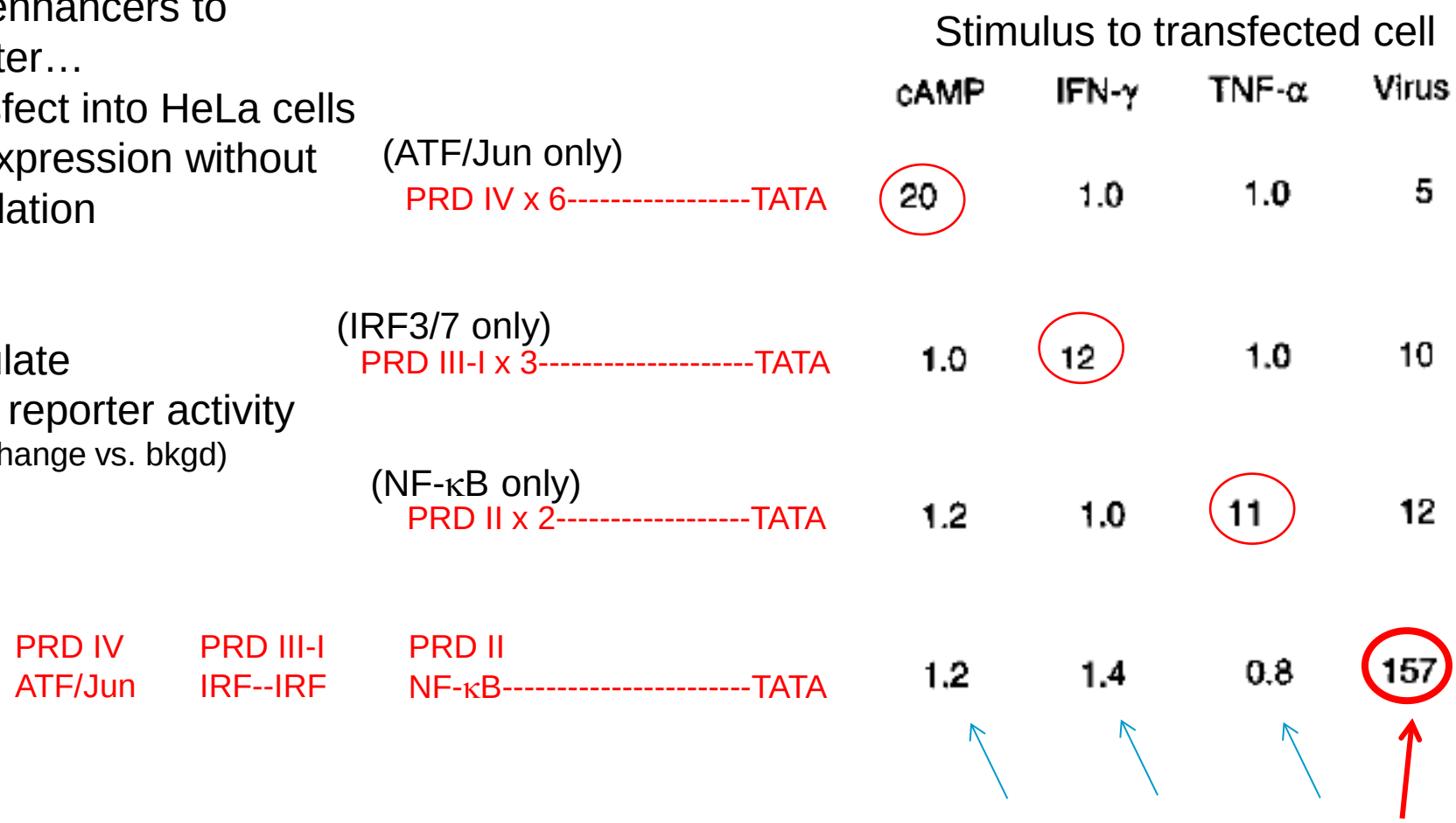


Lenardo et al., Cell, 1989

Response properties of the intact IFNβ enhancer are ***different*** from those of the individual transcription factor binding site elements

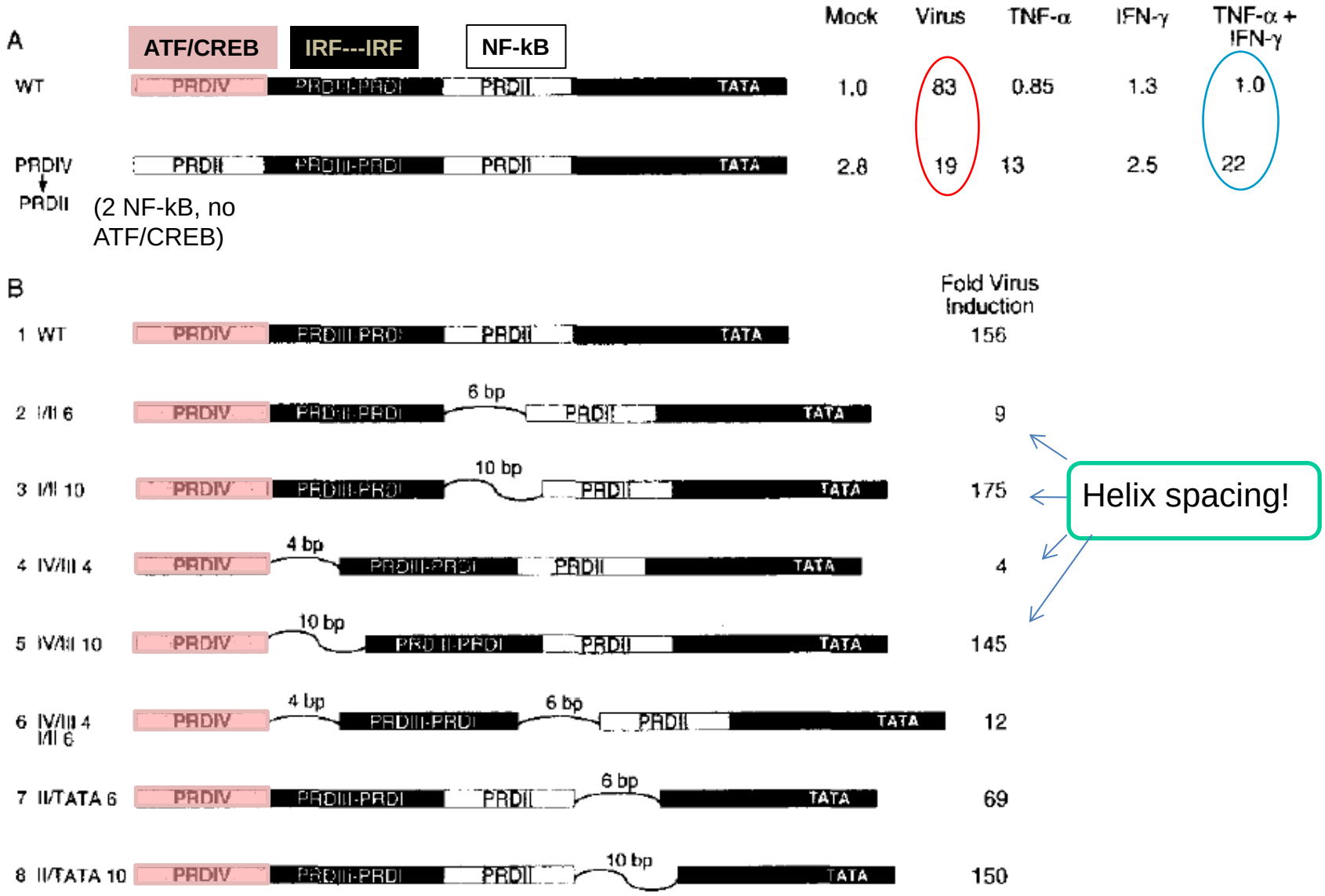
Link enhancers to reporter…
Transfect into HeLa cells
NO expression without stimulation

But:
Stimulate
Read reporter activity
(Fold change vs. bkgd)



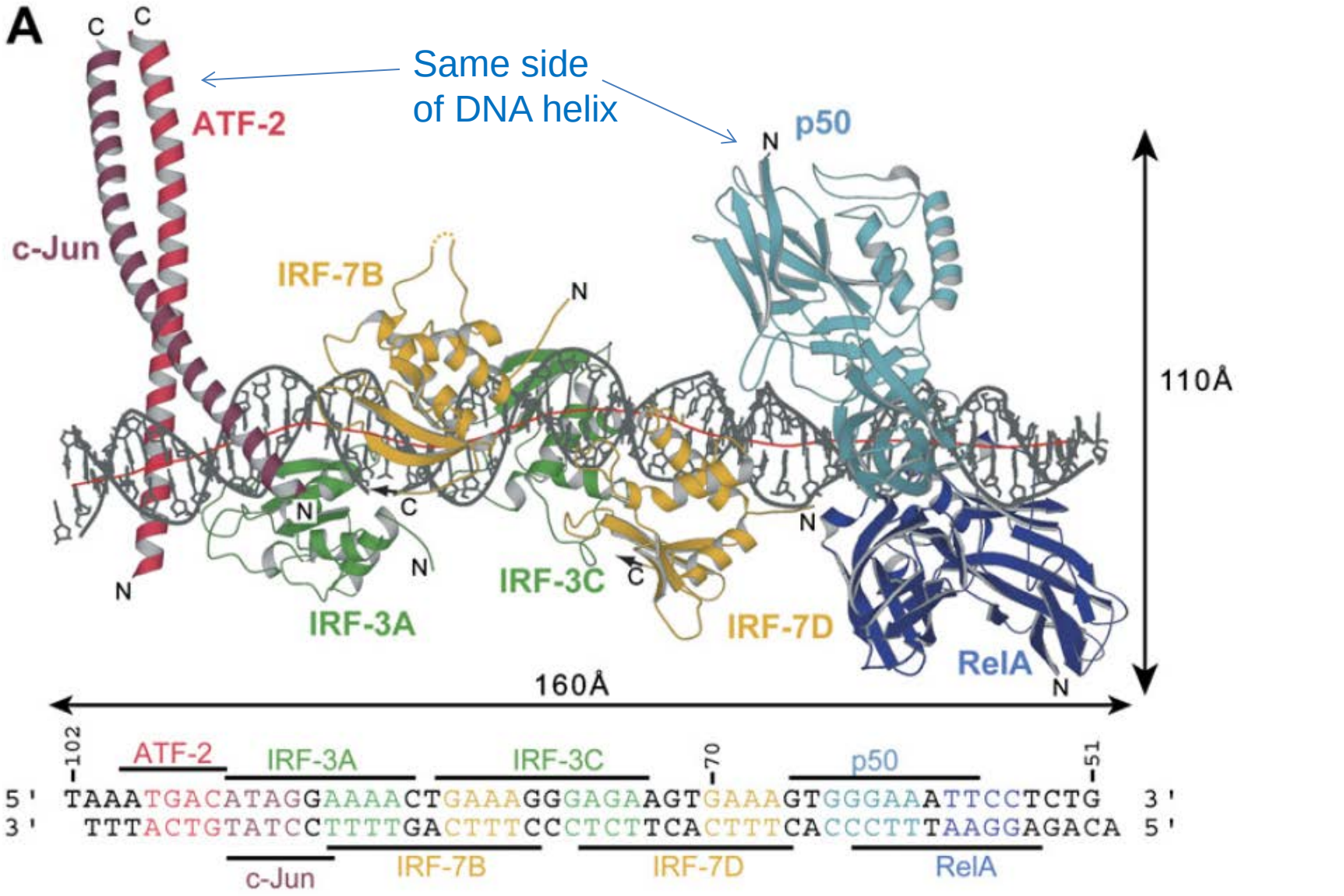
(Thanos & Maniatis, Cell 1995)

Details of enhancer architecture dramatically affect response to physiological stimulus, virus!



3-D structure of the IFNβ enhanceosome complex

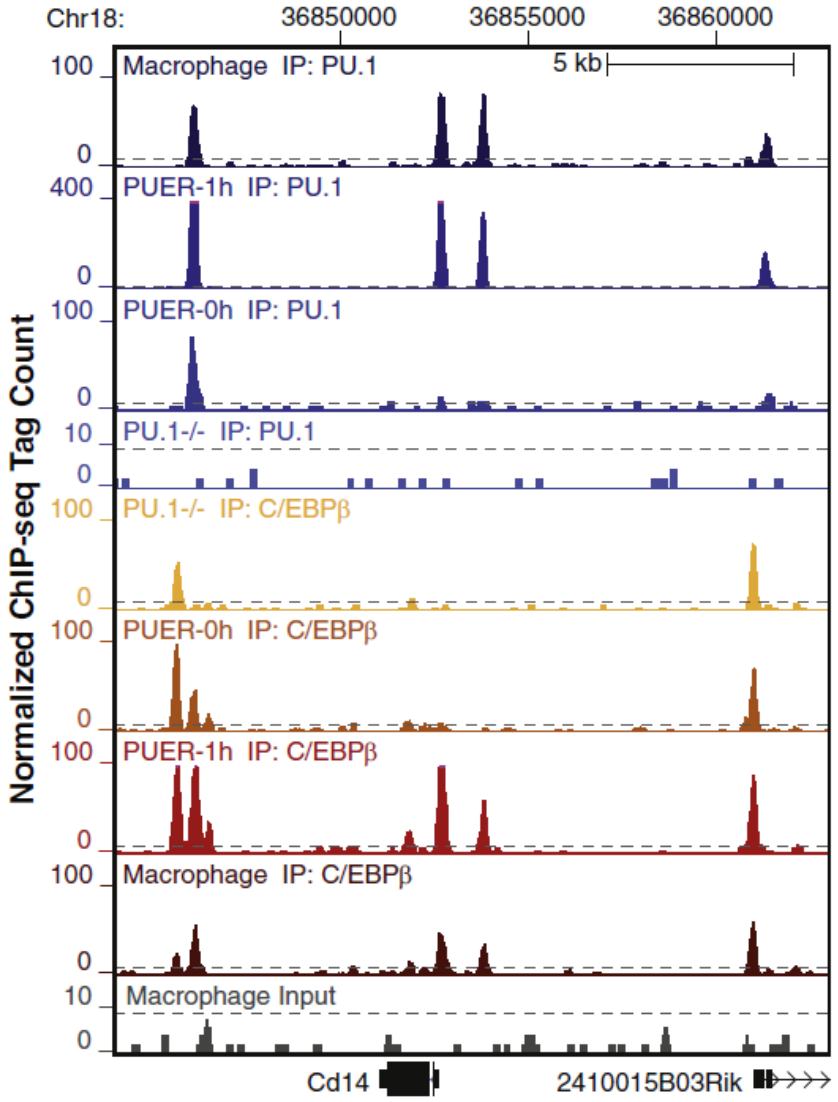
Daniel Panne,[1] Tom Maniatis,[2] and Stephen C. Harrison[1,*] Cell *129*, 1111–1123, June 15, 2007



Induction-dependent transcription factors can make an entire enhanceosome induction-dependent: linking molecular & cell biology

- For some induction dependent genes, some of the required binding factors are always present… but these are *necessary but not sufficient*
- Several are strictly induction-dependent… though also *necessary but not sufficient*
- Combinatorial requirement for these factors creates combinatorial requirement for particular kinds of signaling
- May combine with cell type specific transcription factors and timing factors

A TF already present in a mammalian cell may not be able to bind its site without "a little help from its friends"



- PU.1 and C/EBP factors usually bind together at many sites in macrophages
- In a mutant PU.1-negative cell line, C/EBPβ binds far fewer sites despite being expressed at normal levels
- Restoring PU.1 to the cell (using a cute "ER fusion" trick) rapidly recruits C/EBPβ to sites that it could not bind to, alone

(Heinz, … Glass 2010 Mol Cell)

Transcription factor binding specificity: how do we know it?

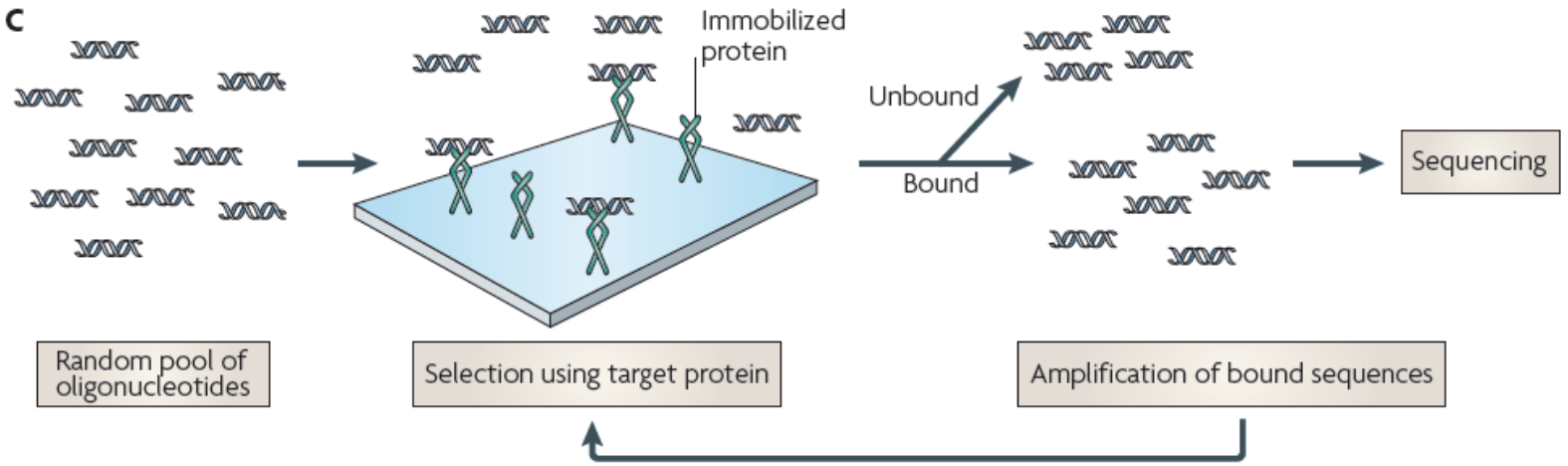
- Genetic evidence is not enough to identify sites where transcription factors act in animals with large genomes
  - Cannot assume compact, predictably localized “operator” sites
- Need biochemistry
- Starting material: use purified protein to identify sites
- Or – use purified sites to identify proteins

- Build on prior knowledge: by 2014, many previous tests have been made (though many unknowns remain)
  - Transcription factors of a given structural family often have similar DNA binding specificity
  - Complex statistical analysis of experimental results: “PWM” site predictors

Methods of testing protein:DNA binding specificity

Depend on purified protein & artificial DNA…

Start by cloning the gene for the transcription factor of interest, express it from a highly-transcribed plasmid in bacteria or other host, purify it, and use it to distinguish preferred from non-preferred DNA from a random library…



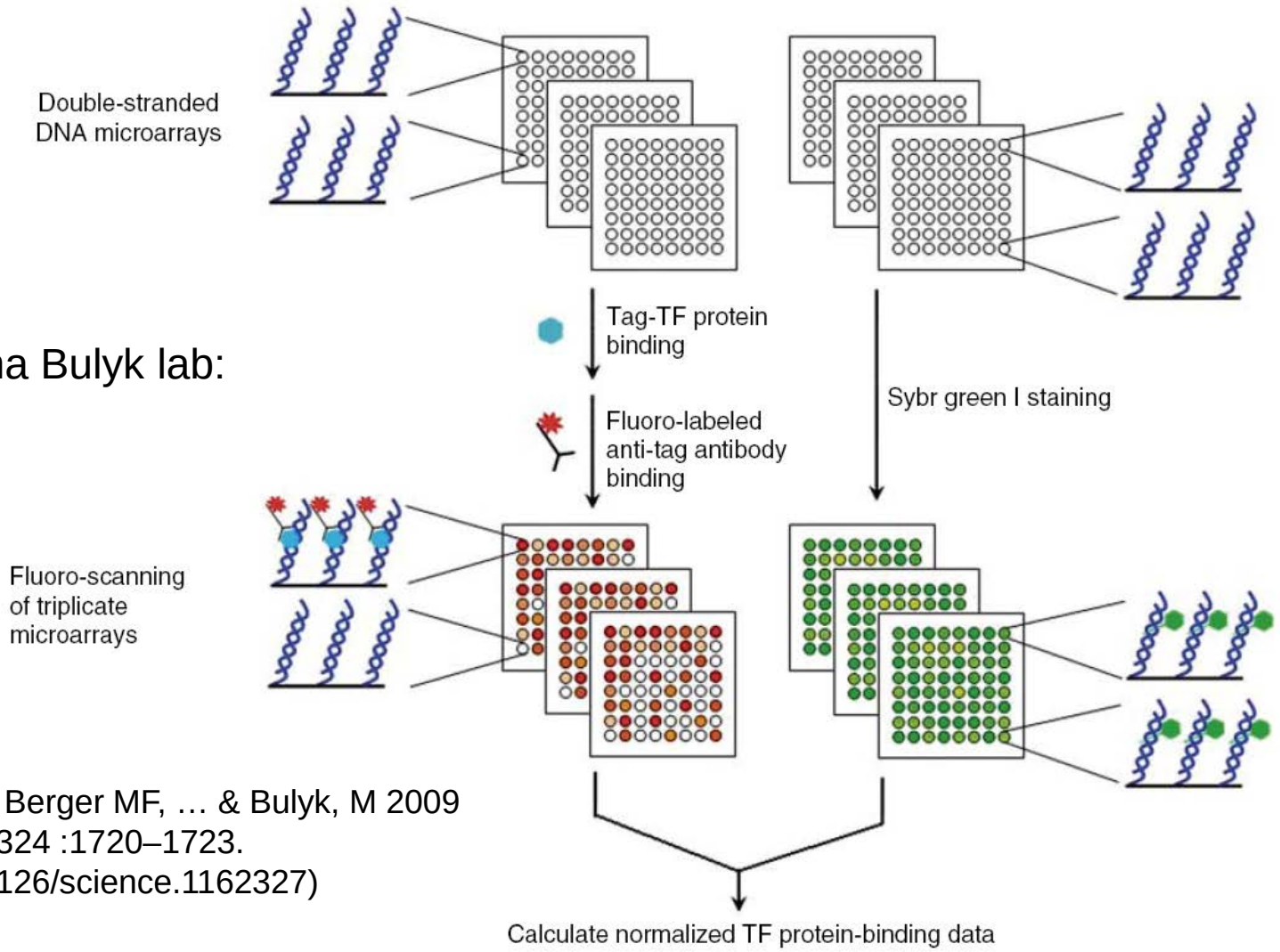
"SELEX"

If your "random library" DNA sequences are designed to be flanked with known PCR priming sites, you can amplify the "winners" in vitro and repeat… and repeat…

- *Reviewed by Stormo & Zhao, Nat Rev Genet, 2010*

Protein-binding microarrays: lay out huge # of DNA sequences in predefined matrix, have an antibody ready to detect factor, then follow where purified factor goes

Martha Bulyk lab:



Badis G, Berger MF, … & Bulyk, M 2009
Science 324 :1720–1723.
(doi:10.1126/science.1162327)

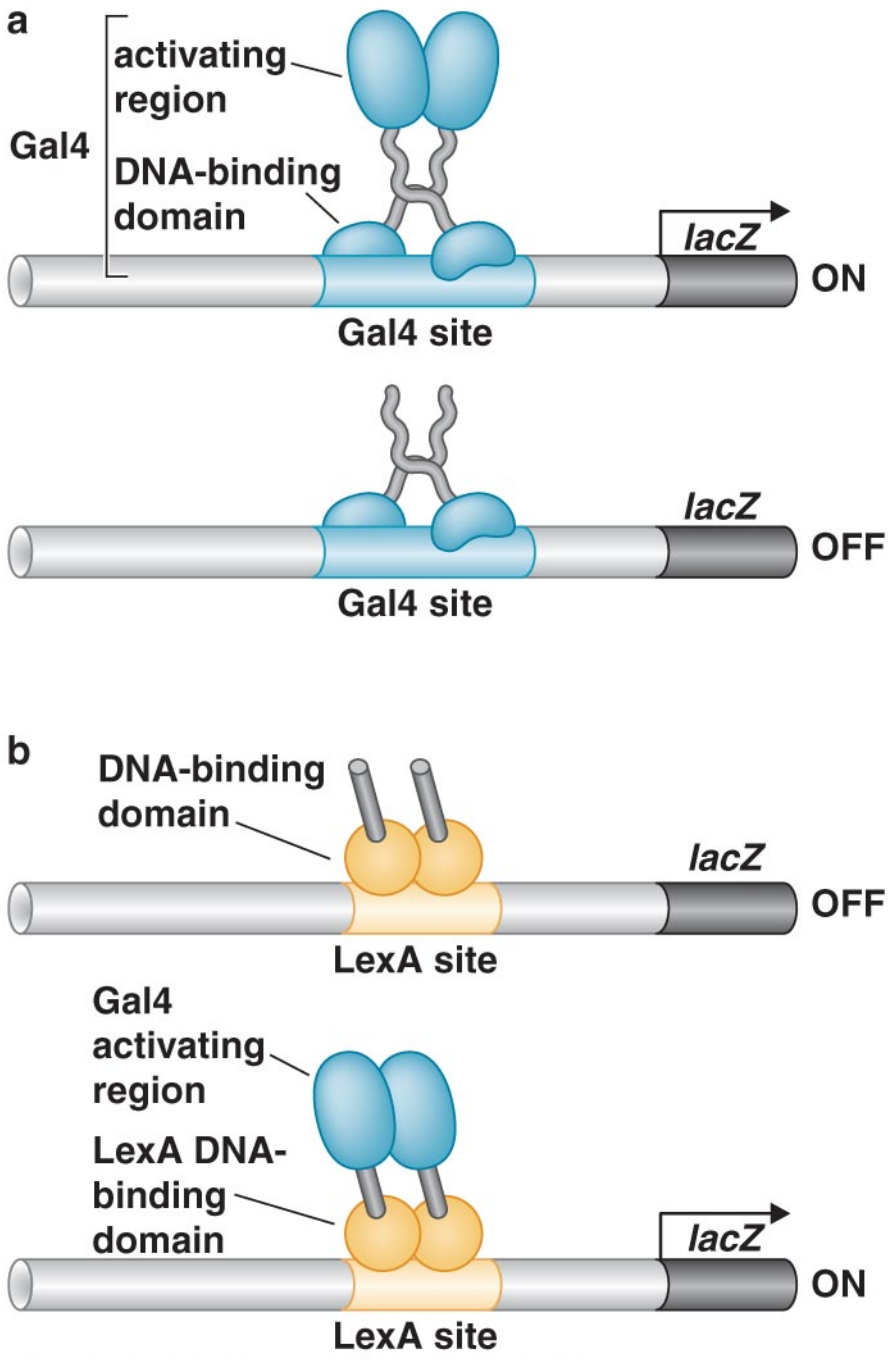
Reviewed in Wang & Liu, Journal of Endocrinology (2011) 210, 15–27

How do transcription factors work?

- DNA binding domains are often separate from “effector” domains
  - Different exons
- “Transactivation” domains
  - Q-rich
  - Acidic
  - Other…
- Regulatory domains for other factor interactions

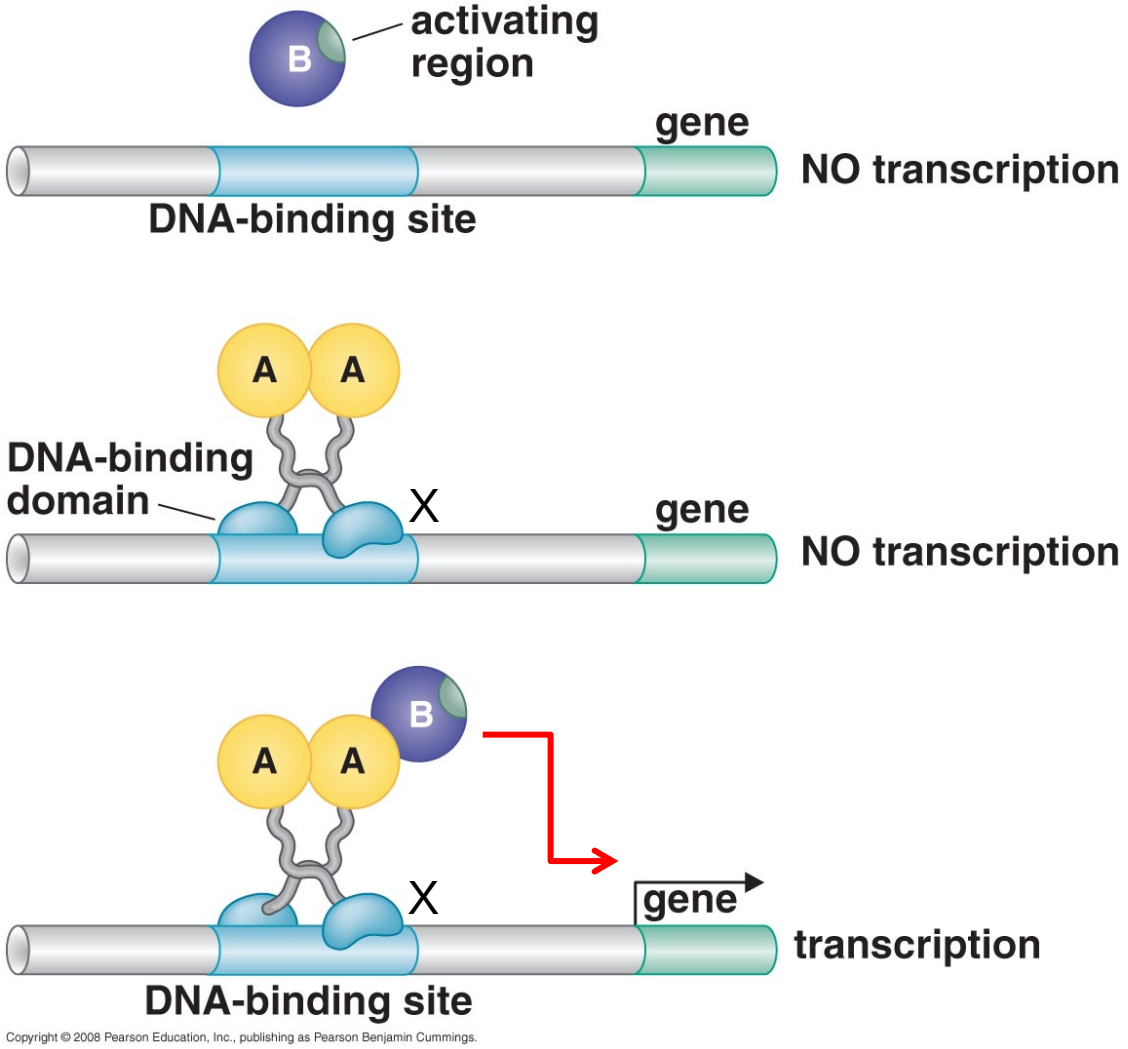
These functional domains are especially important for eukaryotic transcription factors that often work FAR from promoters

Modular transcription factor structure makes chimeras easy to make experimentally (and in evolution)

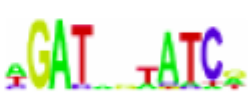


The “yeast 2-hybrid assay”: widely useful application of a basic principle of TF structure

- Engineer fusion of “Bait” protein “A” to an isolated DNA binding domain of transcription factor “X”
- Make a reporter construct that requires activity of transcription factor X
- Transfect, express in yeast
- Also express a cDNA library of multiple proteins (B_1, B_2, …) engineered to be fused with a transcription activation domain…
- Only for a B that interacts with A will reporter go on



DNA binding specificity of whole transcription factor protein usually matches specificity of isolated DNA binding domain when tested as purified protein on short test sites

Protein	Primary Motif DNA Binding Domain	Primary Motif Full Length	Secondary Motif DNA Binding Domain	Secondary Motif Full Length
Max				
Bhlhb2				
Gata3				
Rfx3				
Sox7				

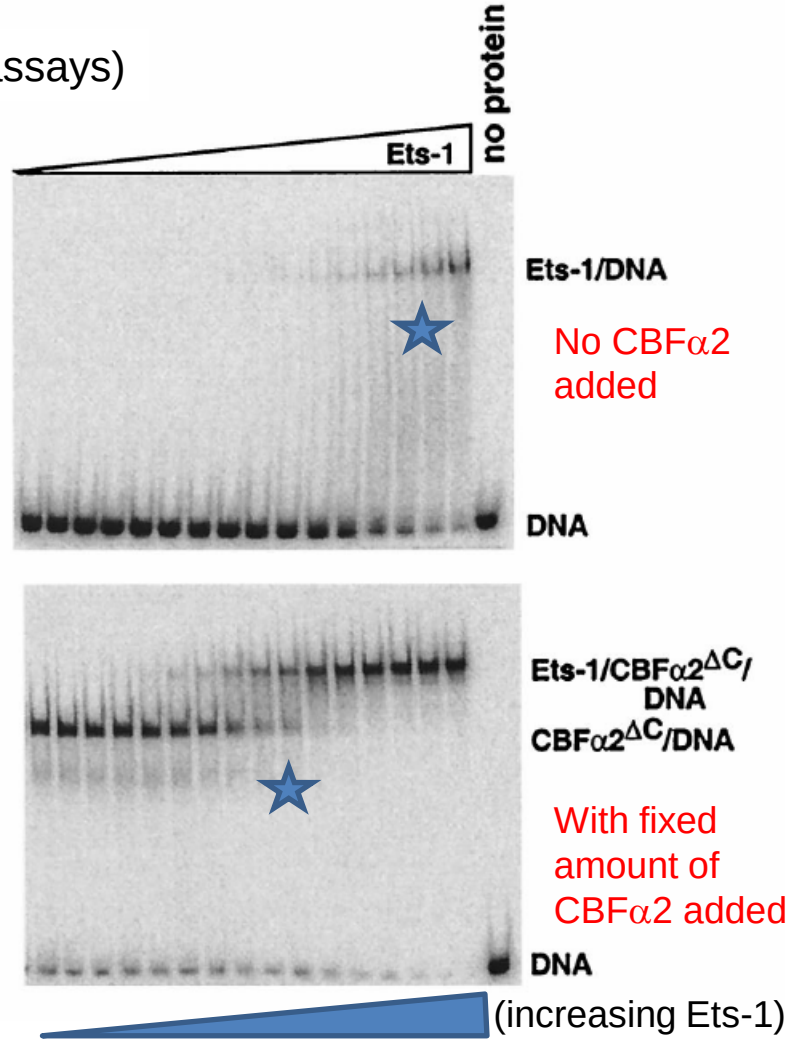
Badis, …. & Bulyk, Science 2009

Crucial importance of cooperativity in transcription factor action

- Transcription factors recognize DNA with too little specificity
  - Often degenerate recognition sites (e.g. CANNTG (E protein); WGATAR (GATA factor); GGAA (Ets family))
  - Single sites can occur >10^6 times in a ~3X10^9 bp genome (10^3 x too many?!)
- Some have weak occupancy without partners, others weak function
  - "Too-easy" binding factors are often weak transcriptional activators on their own
- Partner interaction sensing mechanisms
  - Direct (e.g. Ets1 and CBFα2)
  - Via bridging factors/ coactivators
  - Activity as transcriptional activator (& usefulness in txf "team") can depend on ability to recruit coactivators

What cooperativity can look like: mammalian blood transcription factors, Ets-1 and CBFα2

(Gel shift assays)



Probe has sites for Ets-1 *and* CBFα2…

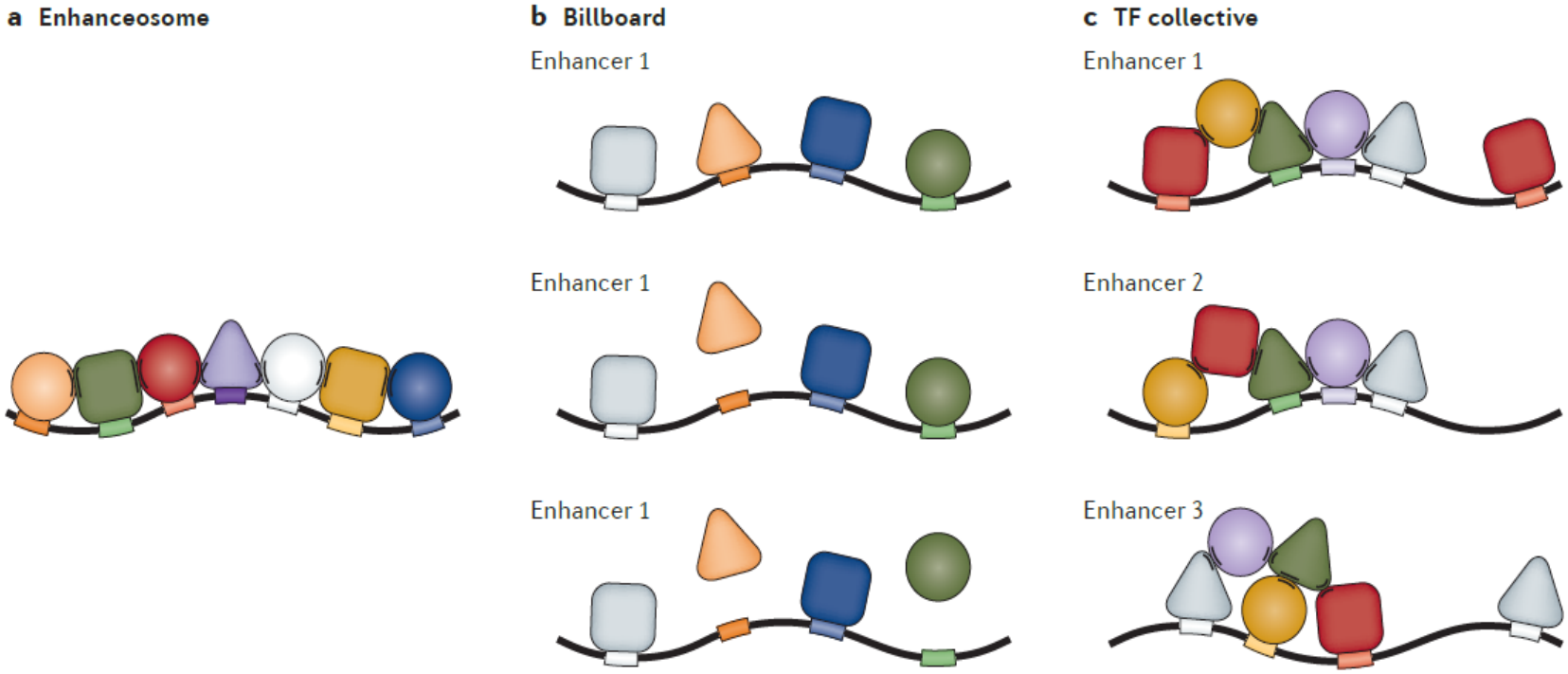
We are titrating how much Ets-1 it takes to shift 50% of the probe to a slower migration by binding

Complex with CBFα2 alone is smaller (faster running) than complex with CBFα2 + Ets-1

Much lower concentration needed!

Goetz TL, Gu TL, Speck NA, Graves BJ. Mol Cell Biol. 2000 Jan;20(1):81-90

Exactly ***how*** does enhancer binding site combination set requirement of TFs needed for gene expression? 3 models



	a Enhanceosome	b Billboard	c TF collective
Protein DNA	• Highly cooperative DNA binding • DNA sequence acts as scaffold	• Cooperative and additive binding	• Cooperative DNA binding • Both DNA and protein may act as scaffold
Protein interface	• Fixed (formed from a higher-order TF–DNA complex)	• Variable	• Variable
Motif	• Fixed motif composition (sites for all factors must be there) • Fixed motif positioning (grammar)	• Fixed motif composition • Flexible motif grammar	• Flexible motif composition (as different TFs directly bind to DNA) • Flexible motif grammar
Output	• Unitary (requires the integrated activity of all TFs)	• Only requires a subset of factors to be active	• Collective (requires most TFs but not clear if it requires all TFs)

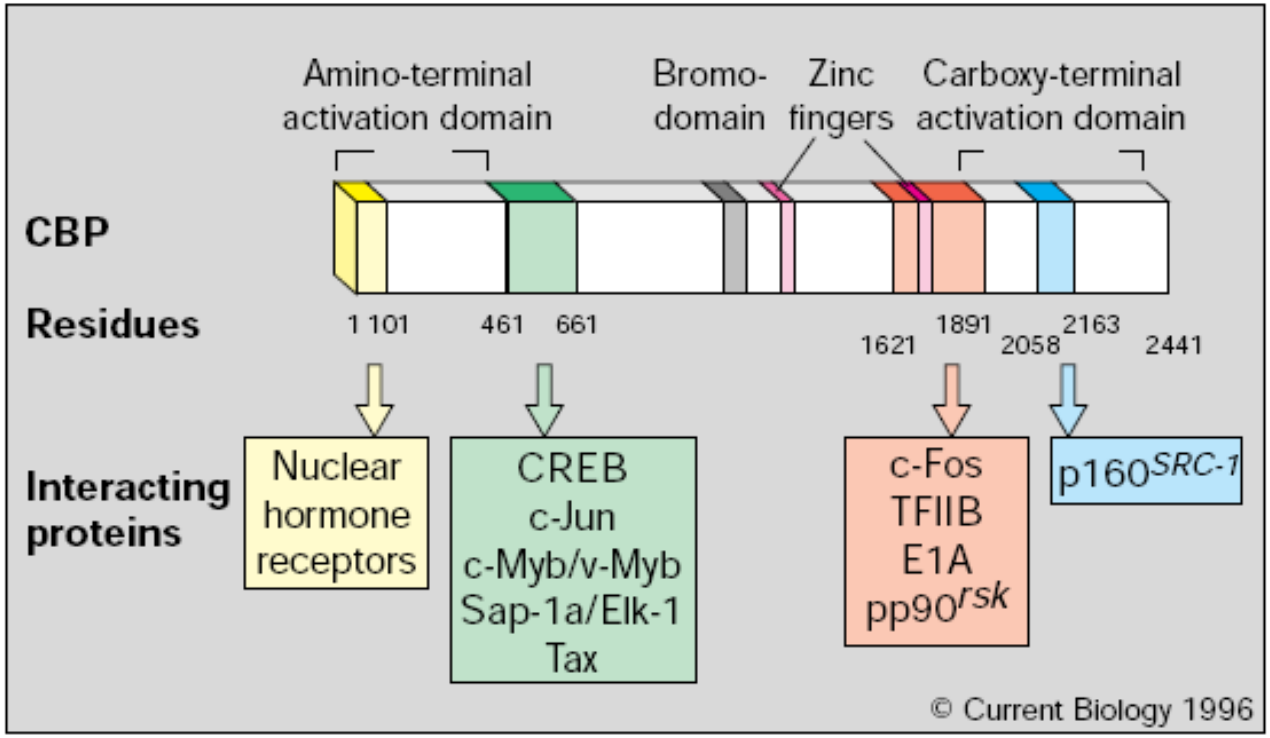
(from Spitz & Furlong, 2012; after Arnosti et al 2005)

Using coactivators to create combinatorial bridges

- Partner interaction sensing mechanisms
  - Direct (e.g. Ets1 and CBFα2)… but this is not the only way
  - Via alterations in histone packing architecture (for next time)
  - **Via bridging factors/ coactivators**
  - Activity as transcriptional activator (& usefulness in txf "team") can depend on ability to recruit coactivators

- General purpose coactivators
  - p300 and CBP
  - Recruitment of histone modification machinery (to be continued!!)
- Dedicated cell type-defining coactivator
  - CIITA
  - Creates "teamwork"

CBP and p300 are huge multidomain proteins with interaction sites for many different transcription factors



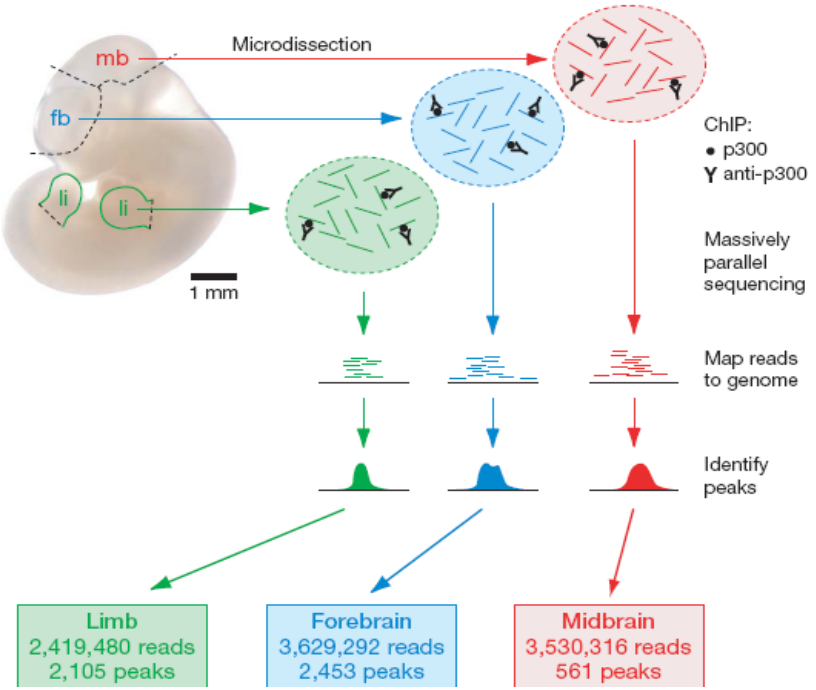
Mouse CBP. Regions 1–101, 461–661, 1621–1891 and 2058–2163 contain binding sites for the proteins indicated below. The bromodomain comprises residues 1107–1171; the two zinc fingers comprise residues 1284–1312 and 1708–1733. The domain structure of p300 appears to be basically similar.

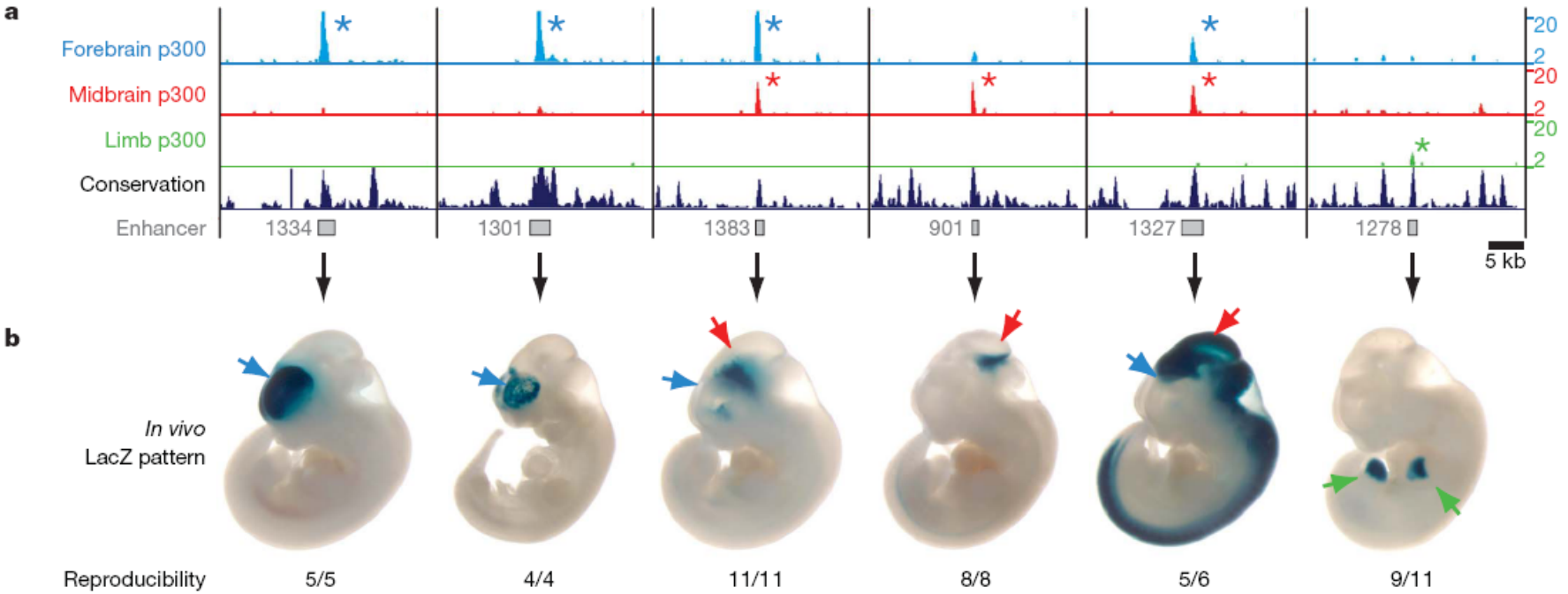
Janknecht, Hunter, Curr Biol 1996

Profile of p300 binding genome-wide can locate sites of active enhancers

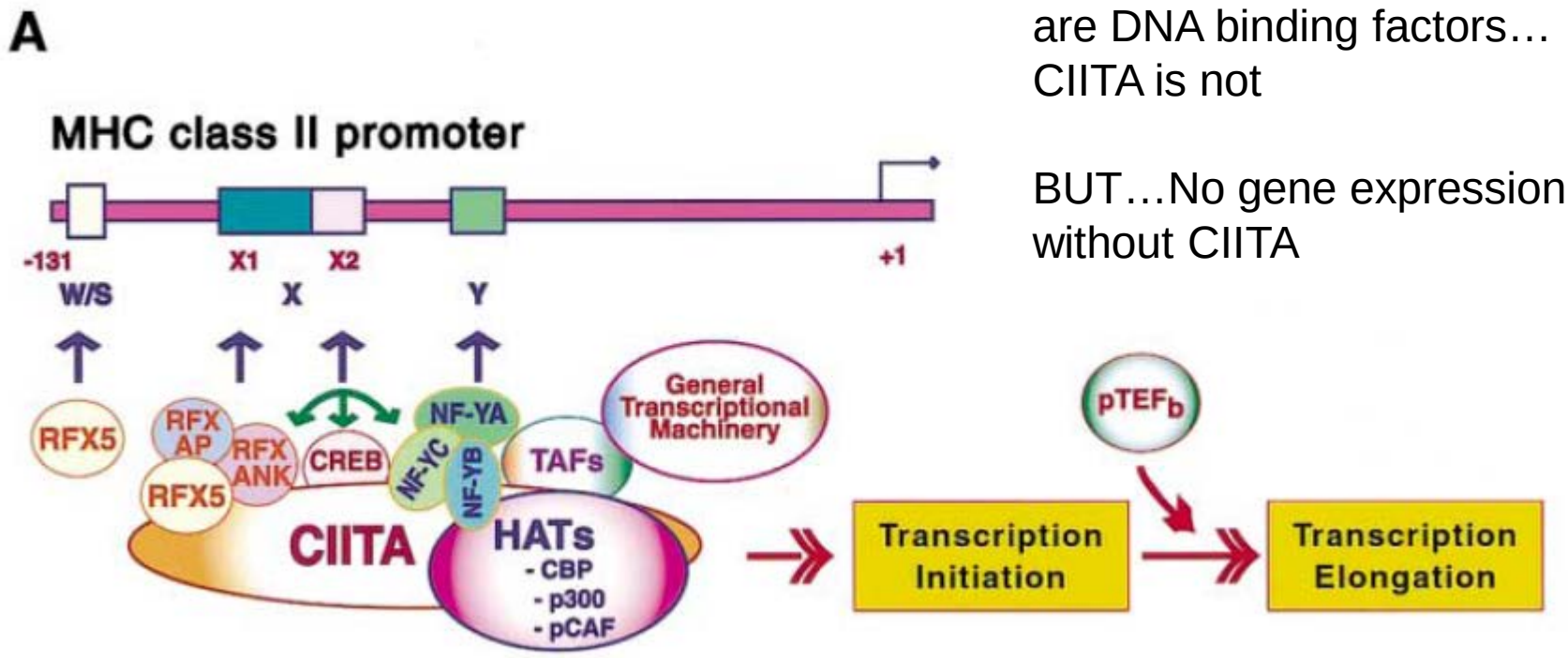
Crosslink chromatin, use antibodies vs. p300 to precipitate bound DNA, and sequence

(Visel, … Pennacchio 2009, *Nature)*





The “master cis-regulatory motif” for a large group of antigen presentation genes is a cluster of sites for “friends of CIITA”



Rfx5, RfxAP, CREB & NF-Y are DNA binding factors… CIITA is not

BUT…No gene expression without CIITA

Ting, Trowsdale, Cell, Vol. 109, S21–S33, April, 2002

Combinatoriality functions

- Dose detection by cooperative site occupancy
- Boolean “AND” logic via structurally required factor interactions

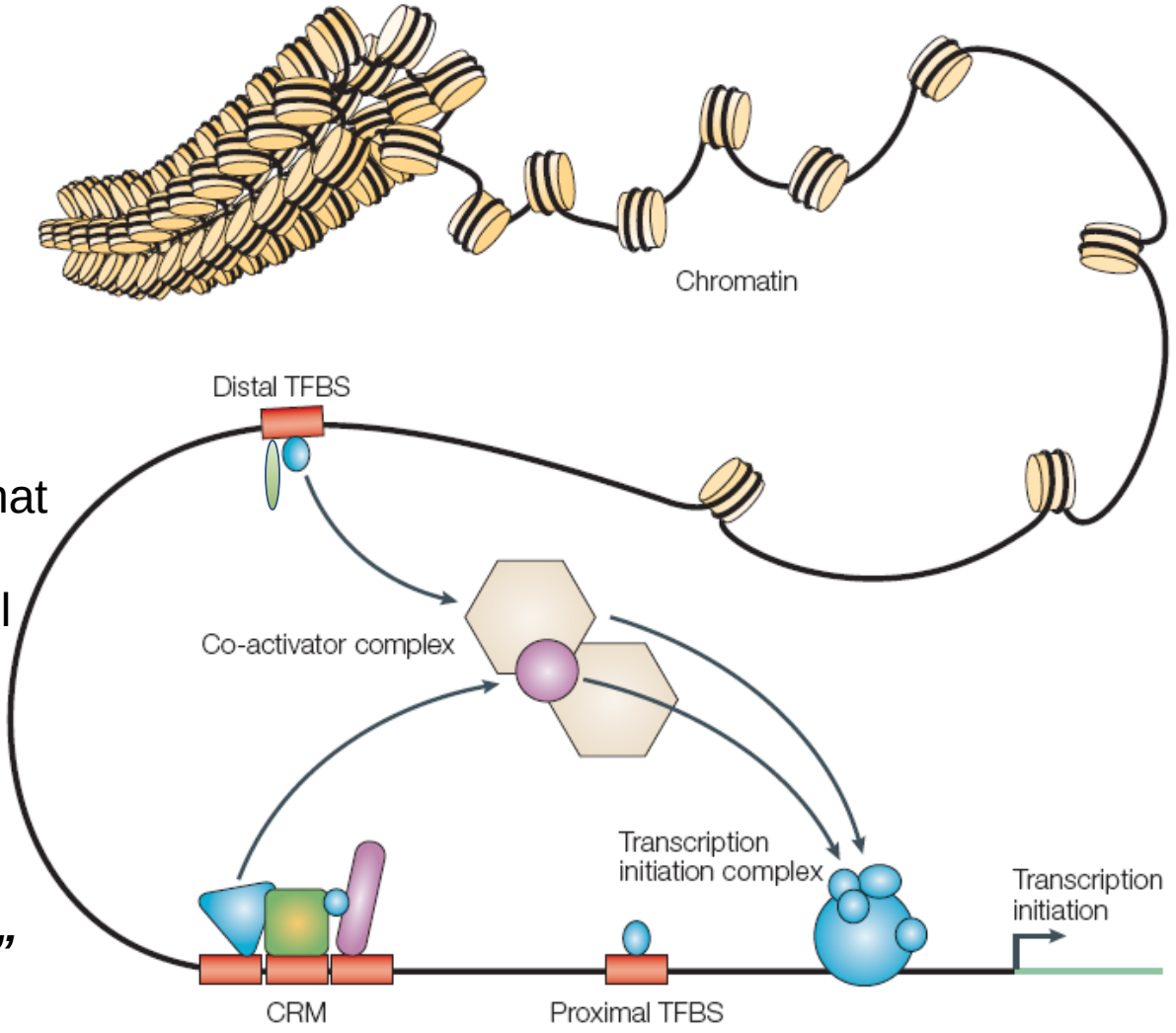
- Quorum mechanisms to make “fuzzy logic” decisions

- Some transcription factors ***change function*** via coactivator vs. corepressor recruitment
  - Major theme in signaling-dependent transcription

Overview of transcriptional regulation in metazoans

Transcription factors bind in clusters at cis-regulatory elements that can be far from the promoters they control

Each cis-regulatory element has its own rule for how many bound transcription factors are "enough"



Key elements of transcription control logic in multicellular eukaryotes

- Enhancers represent a series of DNA sequences recognizable by TFs: determine which TFs are needed for control

- Multiple transcription factors collaborate at each cis-regulatory element
  - "AND" logic (or "additive OR" logic)
  - Both distal enhancers and proximal promoters have this feature

- Multiple cis-regulatory elements can control expression of the same gene
  - Can have alternative, sometimes even antagonistic roles
  - Can be associated with different promoters
  - Use of distinct cis-regulatory elements is CRUCIAL for developmental gene regulation

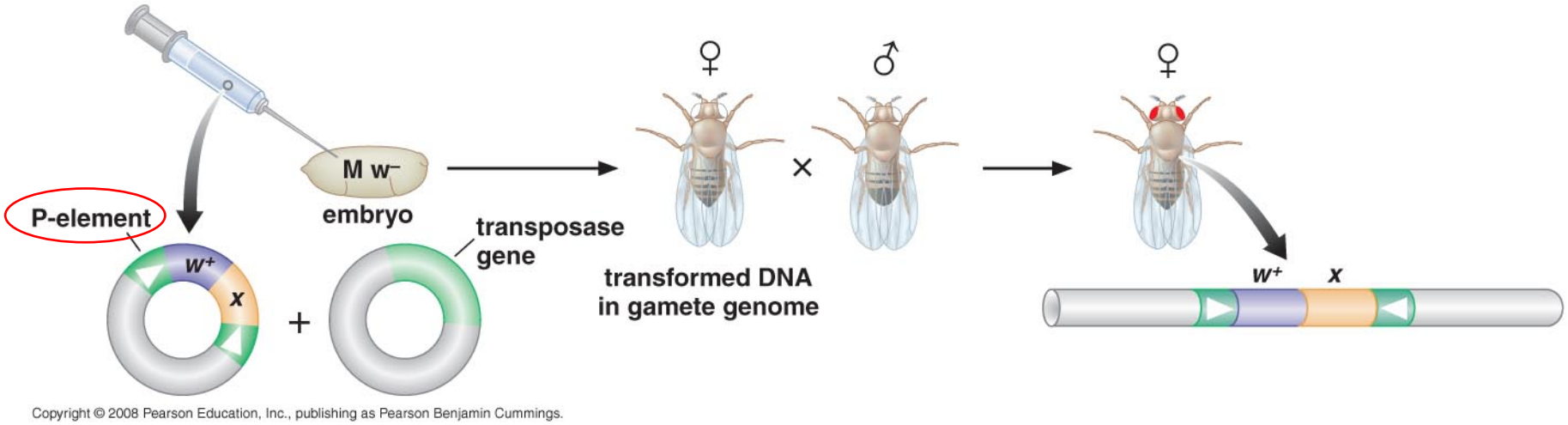
To dissect mechanisms involved in developmental gene regulation, need experimental systems to compare trans-acting factor impacts on cis-reg element targets in multiple cell types from a single organism

- Transfection is always the strategy
- Cis-effects and trans-effects are always compared

  …but somehow these interactions need to be viewed in different types of cells, with different developmental histories, in parallel

- ➢ Go beyond “regular” transfection techniques – transient or stable transfection of cell lines – to make transgenic animals

Specialized transgenic production in Drosophila takes advantage of a transposase that is activated during mating: rapidly get whole, developed animals carrying gene of interest



Drosophila naturally have a transposable element – a repeated DNA sequence that can encode its own recombination enzyme – and integrates randomly into the chromosomal DNA. Can be made into a vector to insert your gene of interest, X….

Detect the flies with X in the genome by screening for linked gene “W+”.

Techniques exist to *replace* normal DNA sequences in mouse genome precisely with desired mutants

Key: having a cell type that has the developmental potential of a fertilized egg, but grows in culture:

ES cells

ES cells re-introduced into mouse embryos enter germline, make whole normal mice

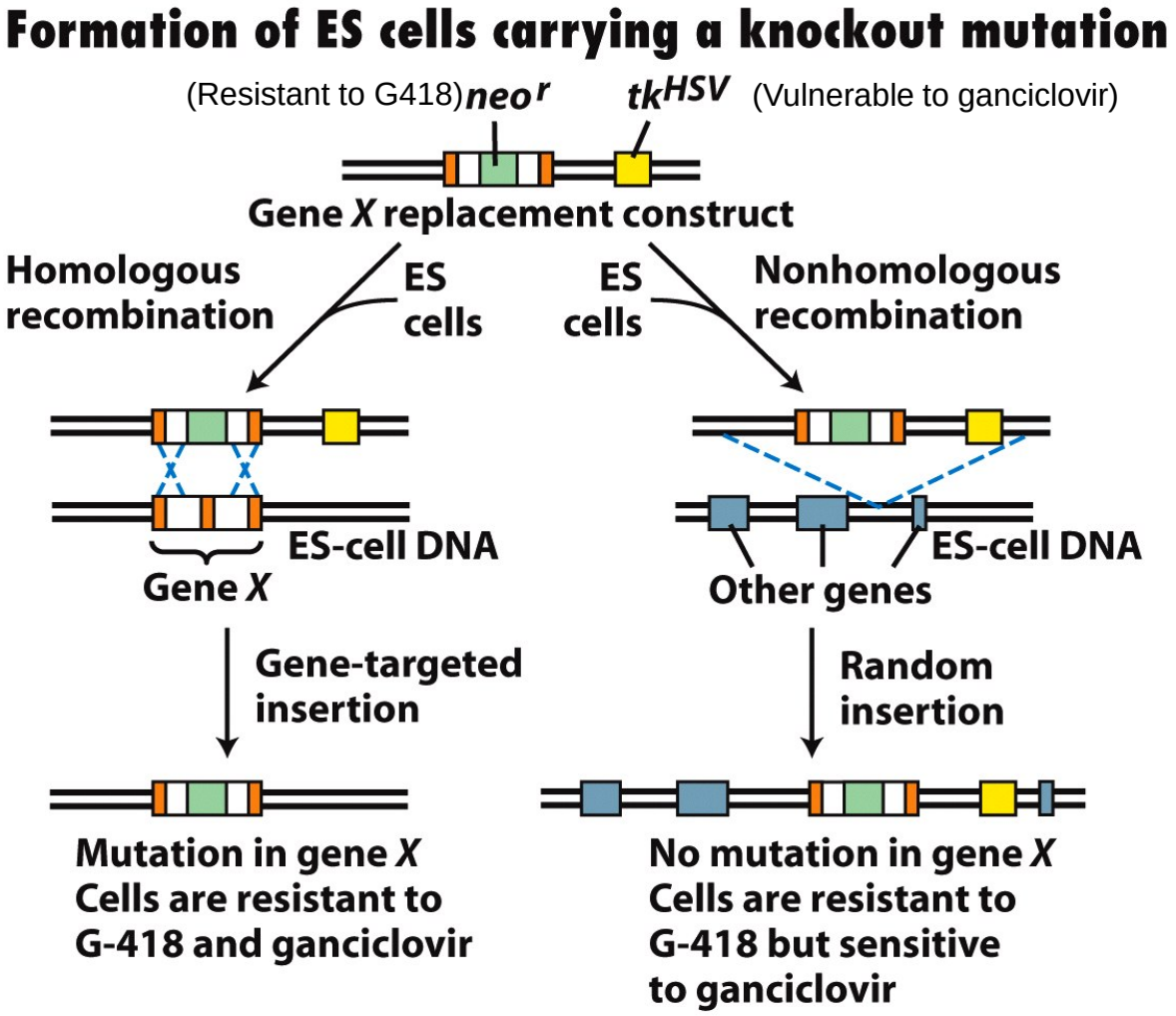


Figure 5-40a
Molecular Cell Biology, Sixth Edition

WANTED

UNWANTED, DISCARDED