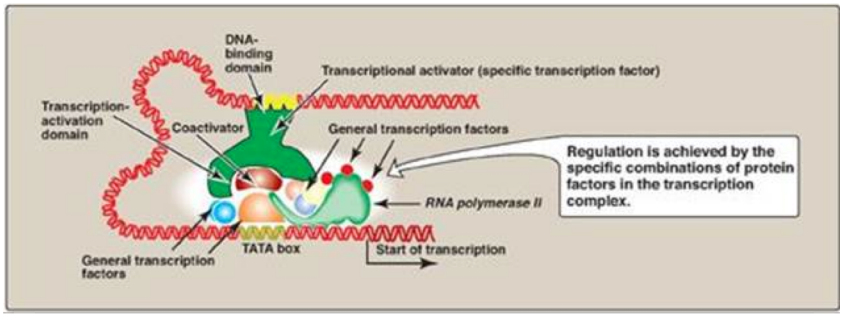
L4.4 – Transcriptional regulation (Transcription Factors)

Factors required to regulate transcription

- Polymerase
- GTFs



- Mediator complex
- DNA-binding regulatory transcription factors
- Corepressors and Coactivators (including different classes of chromatin-modifying enzymes)

AGENDA

1. Transcription factors (Definition, Families and Response Elements)
2. Mediator
3. Coregulators
4. Corepressors/Coactivator Exchange Model

Cell

Leading Edge
Review
Textbook

The Human Transcription Factors

Samuel A. Lambert,[1,9] Arttu Jolma,[2,9] Laura F. Campitelli,[1,9] Pratyush K. Das,[3] Yimeng Yin,[4] Mihai Albu,[2] Xiaoting Chen,[5] Jussi Taipale,[3,4,6,*] Timothy R. Hughes,[1,2,*] and Matthew T. Weirauch[5,7,8,*]

[1]Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada
[2]Donnelly Centre, University of Toronto, Toronto, ON, Canada
[3]Genome-Scale Biology Program, University of Helsinki, Helsinki, Finland
[4]Division of Functional Genomics and Systems Biology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Solna, Sweden
[5]Center for Autoimmune Genomics and Etiology (CAGE), Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA
[6]Department of Biochemistry, Cambridge University, Cambridge CB2 1GA, United Kingdom
[7]Divisions of Biomedical Informatics and Developmental Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA
[8]Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, Ohio, USA
[9]These authors contributed equally
*Correspondence: ajt208@cam.ac.uk (J.T.), t.hughes@utoronto.ca (T.R.H.), Matthew.Weirauch@cchmc.org (M.T.W.)
https://doi.org/10.1016/j.cell.2018.01.029

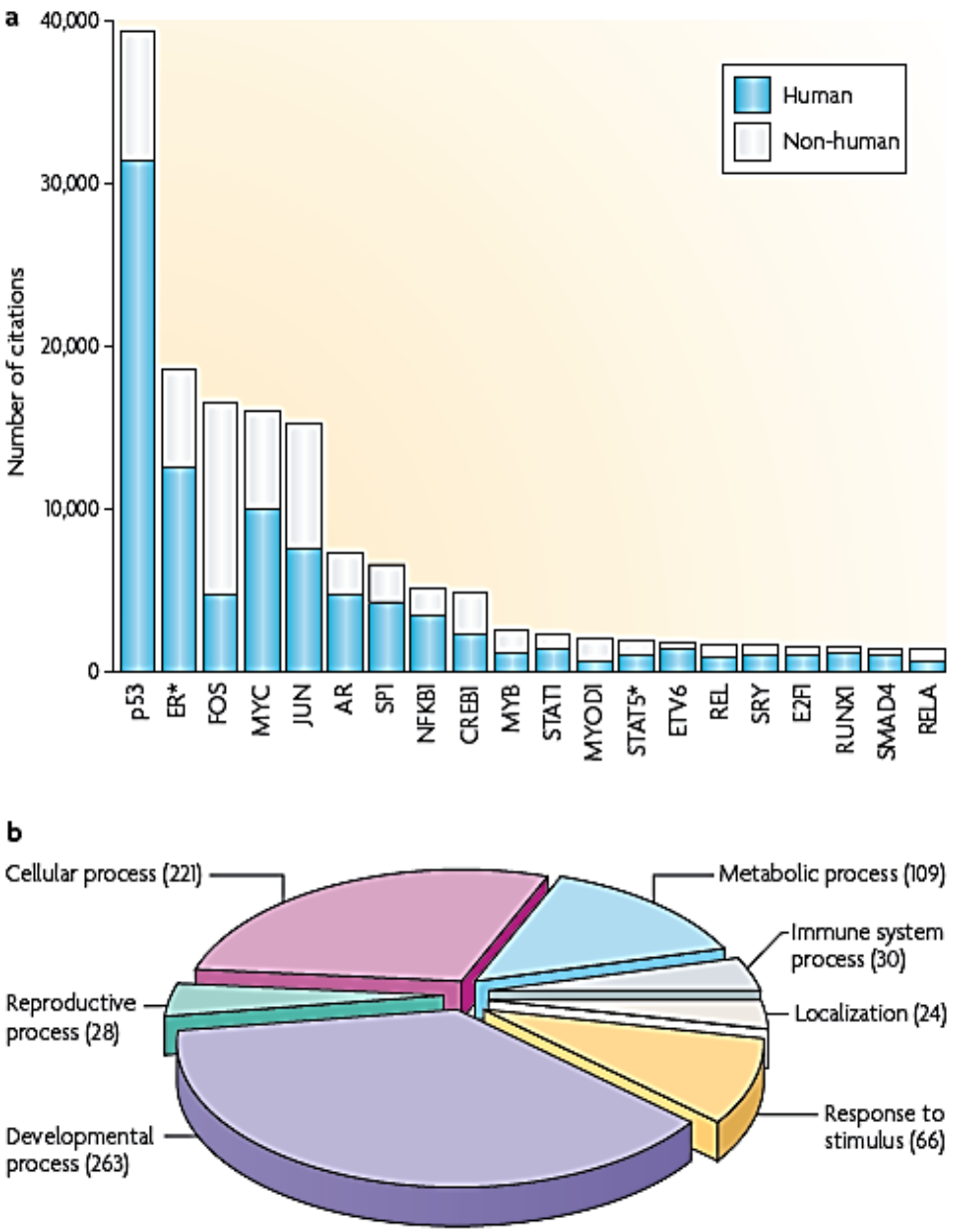
Transcription factors (TFs) recognize specific DNA sequences to control chromatin and transcription, forming a complex system that guides expression of the genome. Despite keen interest in understanding how TFs control gene expression, it remains challenging to determine how the precise genomic binding sites of TFs are specified and how TF binding ultimately relates to regulation of transcription. This review considers how TFs are identified and functionally characterized, principally through the lens of a catalog of over 1,600 likely human TFs and binding motifs for two-thirds of them. Major classes of human TFs differ markedly in their evolutionary trajectories and expression patterns, underscoring distinct functions. TFs likewise underlie many different aspects of human physiology, disease, and variation, highlighting the importance of continued effort to understand TF-mediated gene regulation.

Figure 1

Current state **of knowledge** about transcription factors in the human genome.

a | For the top 20 most cited transcription factors (TFs) in PubMed the number of studies performed in humans (blue bars) and in all other organisms (grey bars) is shown. ER* combines the citations for ERS1 and ERS2, which were indistinguishable in the literature search; similarly, STAT5* includes citations for both STAT5A and STAT5B.

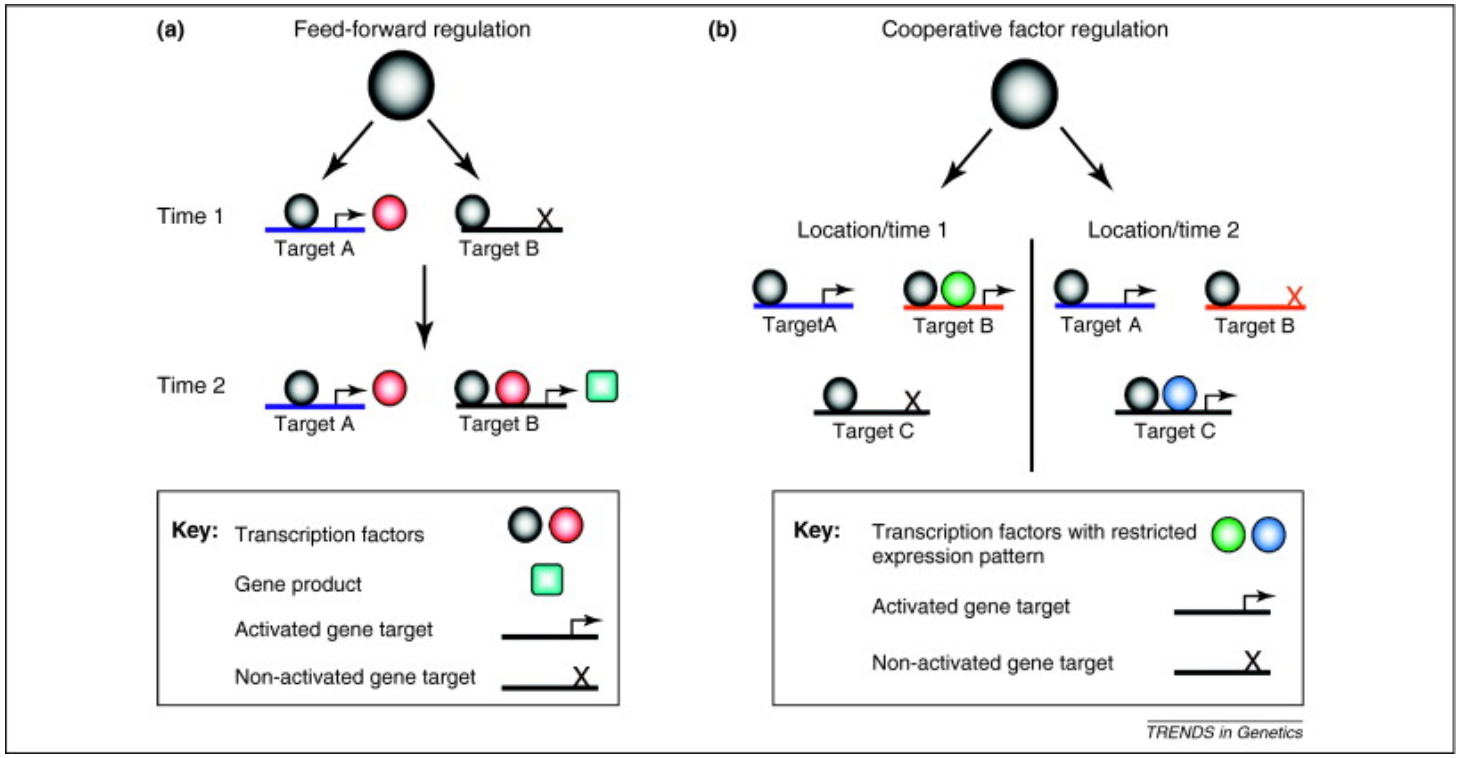
b | summary of biological processes regulated by TFs.
Annotations were obtained from the Gene Ontology database, excluding those based only in electronic annotation. Numbers of annotated TFs are given in parentheses; each gene can be annotated with more than one function.



Vaquerizas et al., Nature Reviews, 2009

Transcription Factors

>10% of the coding potential (2,000-3,000) of the Human Genome



Trends in Genetics 2011 27, 141-148DOI: (10.1016/j.tig.2011.01.001)

DNA-binding Transcription Factors (regulatory factors) (TF)

GO cathegory

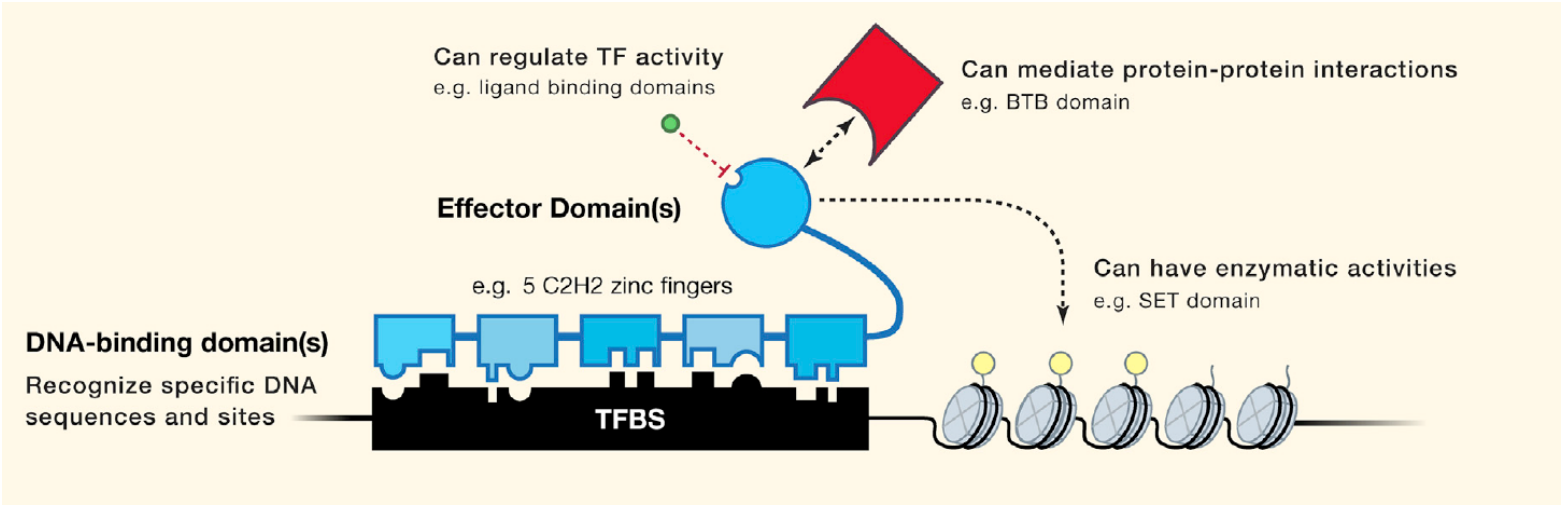
Do include:
- Putative proteins with similitude to known TFs
- Proteins that possess structural domains similar to DNA Binding domain (DBD) of known TFs

Do not include:
- Coactivators or corepressors
- Enzymes
- General Transcription Factors (basal PIC components)

TF database: http://www.transcriptionfactor.org/index.cgi?Home

Structures (DBD) http://www.rcsb.org/pdb/home/home.do

WHAT is a TRANSCRIPTION FACTOR?



DNA-binding domains (as well as dimerization domains, which are very often closely associated in transcription factors, display quite rigid 3D structures.

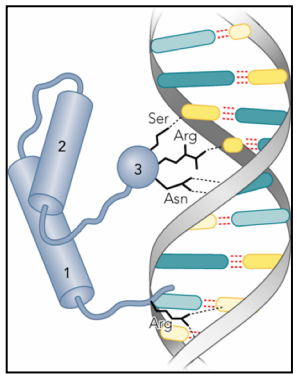
In sharp contrast, **transactivating domains** have never been resolved by cristallography, i.e. they are flexible and adaptable domains, which most likely assume different conformations, depending on interactions.

Trans-activating domain classification is rather based on aminoacid composition, i.e.:

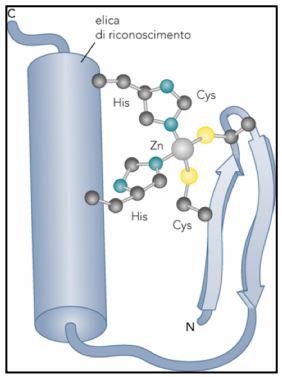
- acidic
- glutamine-rich
- glutamine/proline rich
- hydrophobic

Transcription Factor Families

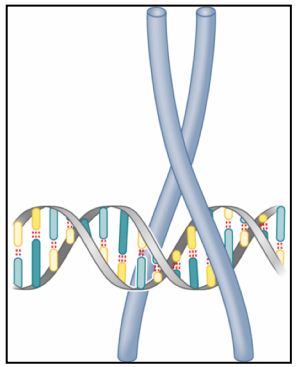
The specificity of gene expression is guaranteed by the existence of numerous transcription factor families with different DNA binding domains: helix-turn-helix, zinc finger, MADS, HMG box, leucine zipper…



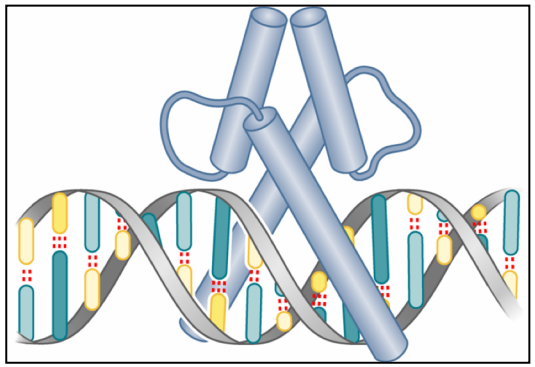
HOMEODOMAIN
(Hox, Matα)



ZINC FINGER
(Nuclear Receptors, Gal4)



LEUCINE ZIPPER
(Jun, Fos, CREB)



HELIX-LOOP-HELIX
(Myc, MyoD, SREBP)

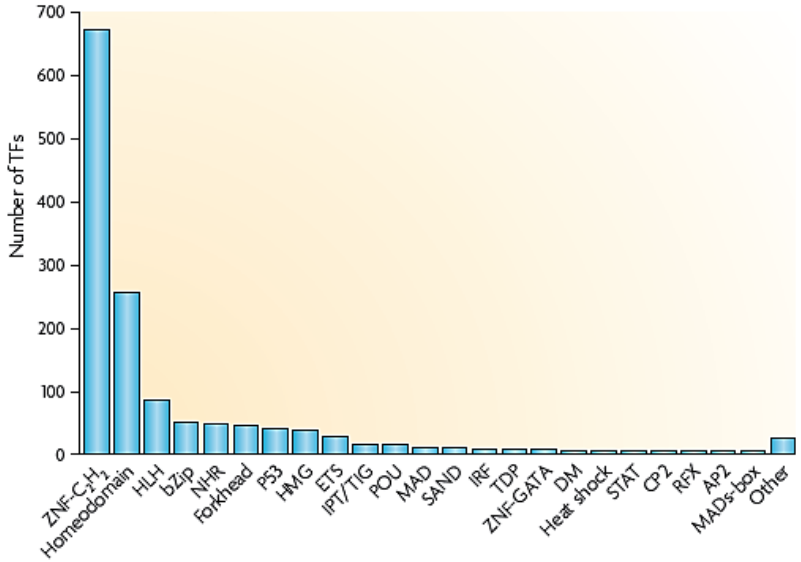


Figure 2 | **Transcription factors classified by DNA-binding domain.** Transcription factors (TFs) were classified into families according to their DNA-binding domain composition. InterPro parent–child relationships between DNA-binding domains were used as the basis for TF family definition (Supplementary information S1 (PDF)). TFs with multiple DNA-binding domains were classified in each of their respective families. Families with less than five members were classified as 'other'.

TFs Classes: constitutive/ubiquitous *versus* regulated/tissue-specific

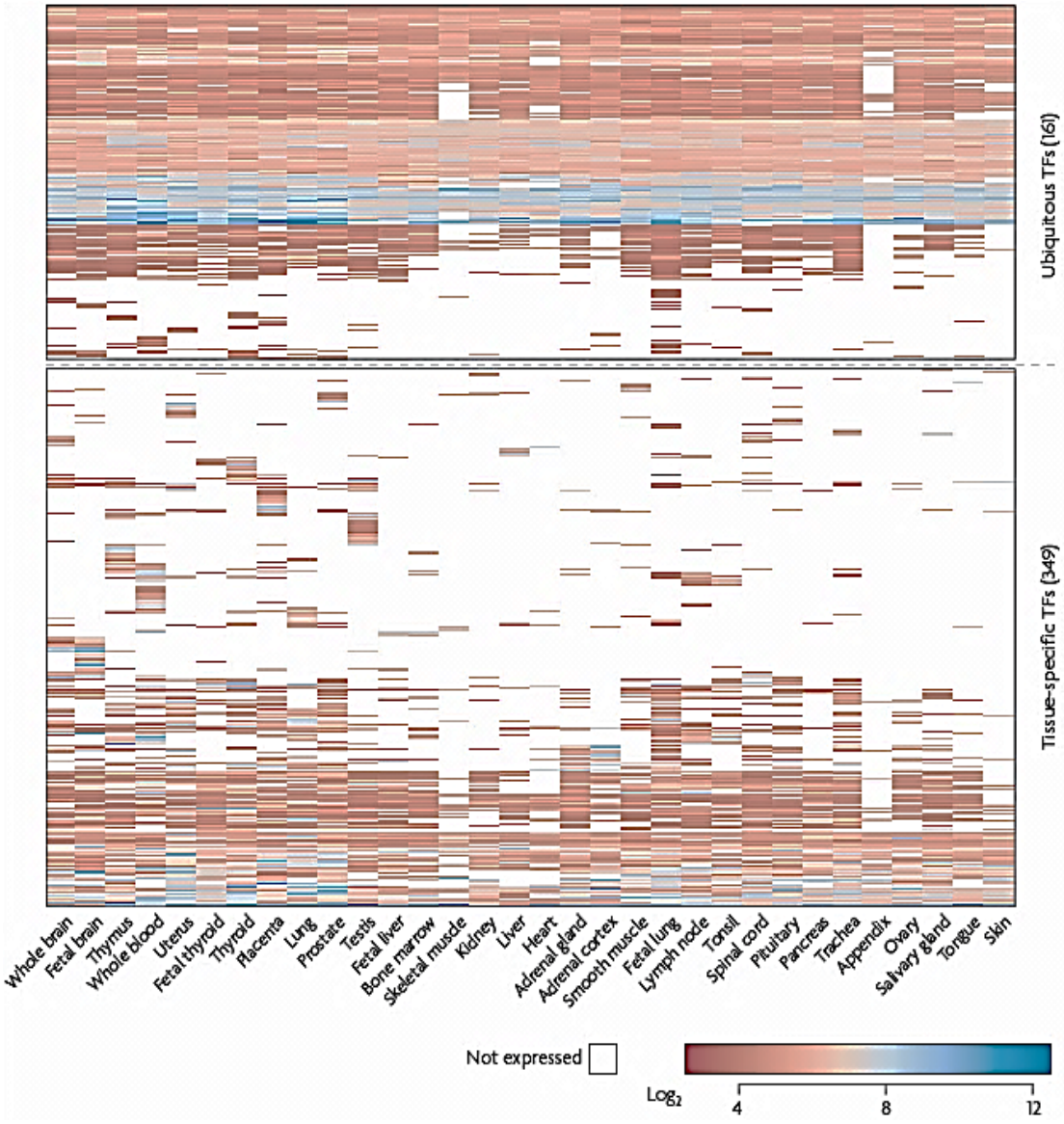


Figure 4 | **Heat map representation of transcription factor expression in 32 human organs and tissues.**
Heat map of transcription factor (TF) expression (rows) in 32 organs and tissues (columns). Intersecting cells are shaded according to expression level (dark red for low expression and blue for high expression). Ubiquitous and specific TFs are grouped according to their expression profiles using hierarchical clustering (before setting an expression level threshold). Ubiquitous regulators are expressed at similar levels across most tissues, whereas specific regulators are expressed at significantly different levels in certain tissues (supplementary information s1 (PDF)). Expression levels below the threshold of detection are depicted as white cells.

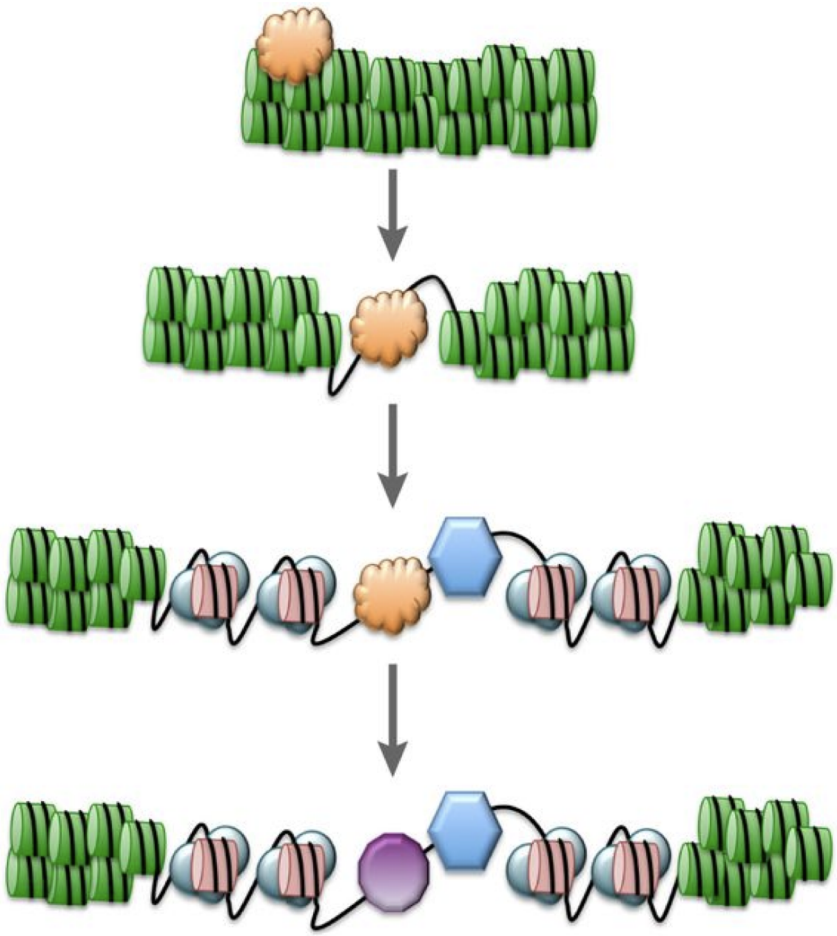
Functional Classification of TFs

Pioneer Factors:
transcription factors able to recognize their cognate DNA sequence even in compacted chromatin. Their binding is followed by histone modification at the nuclosomes flanking the enhancer

Tissue-specific Factors:
TFs that are expressed in precursors or in differetiated cells. Also called lineage-specific factors. They will bind pre-marked enhancers and activate them.

Signal-dependent factors:
TFs that are expressed or activated following a specific endogenous or exogenous stimulus. They will bind to and activate pre-marked enhancers (sometimes also to novel enhancers).

Pioneer TFs: FoxA example

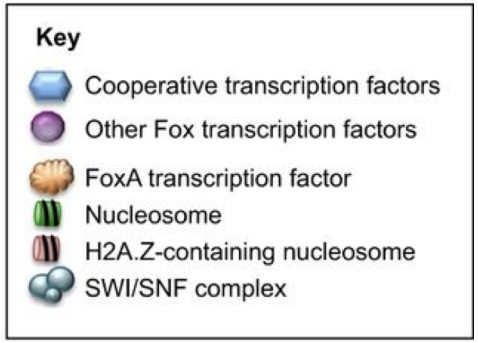


FoxA proteins scan nucleosomal DNA for their target sites

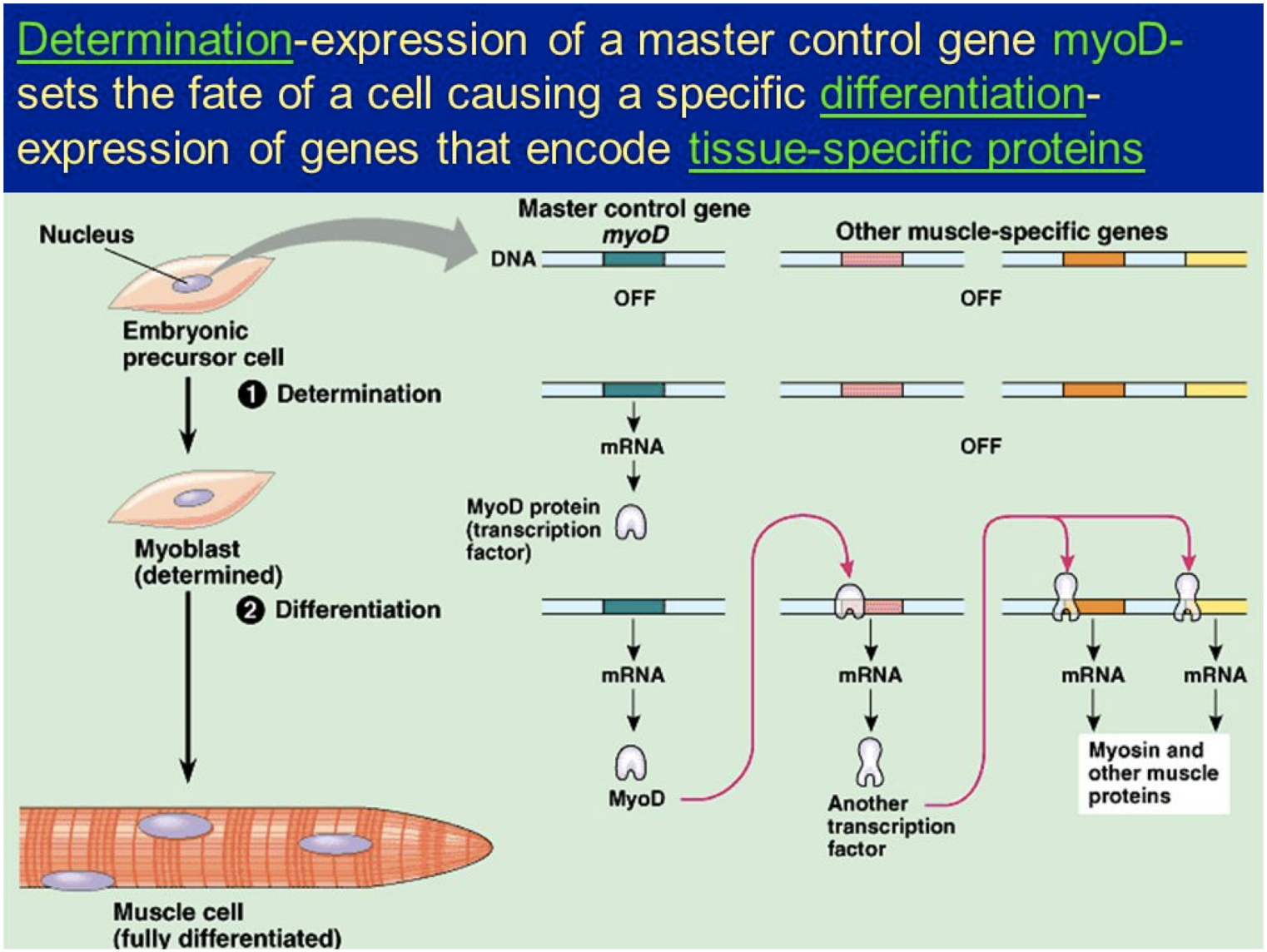
Once bound to their recognition site, FoxA transcription factors are able to displace histones and open chromatin

Once chromatin is open, FoxA recruits chromatin remodelers and histone variants, and other transcription factors are able to find and bind their target sites

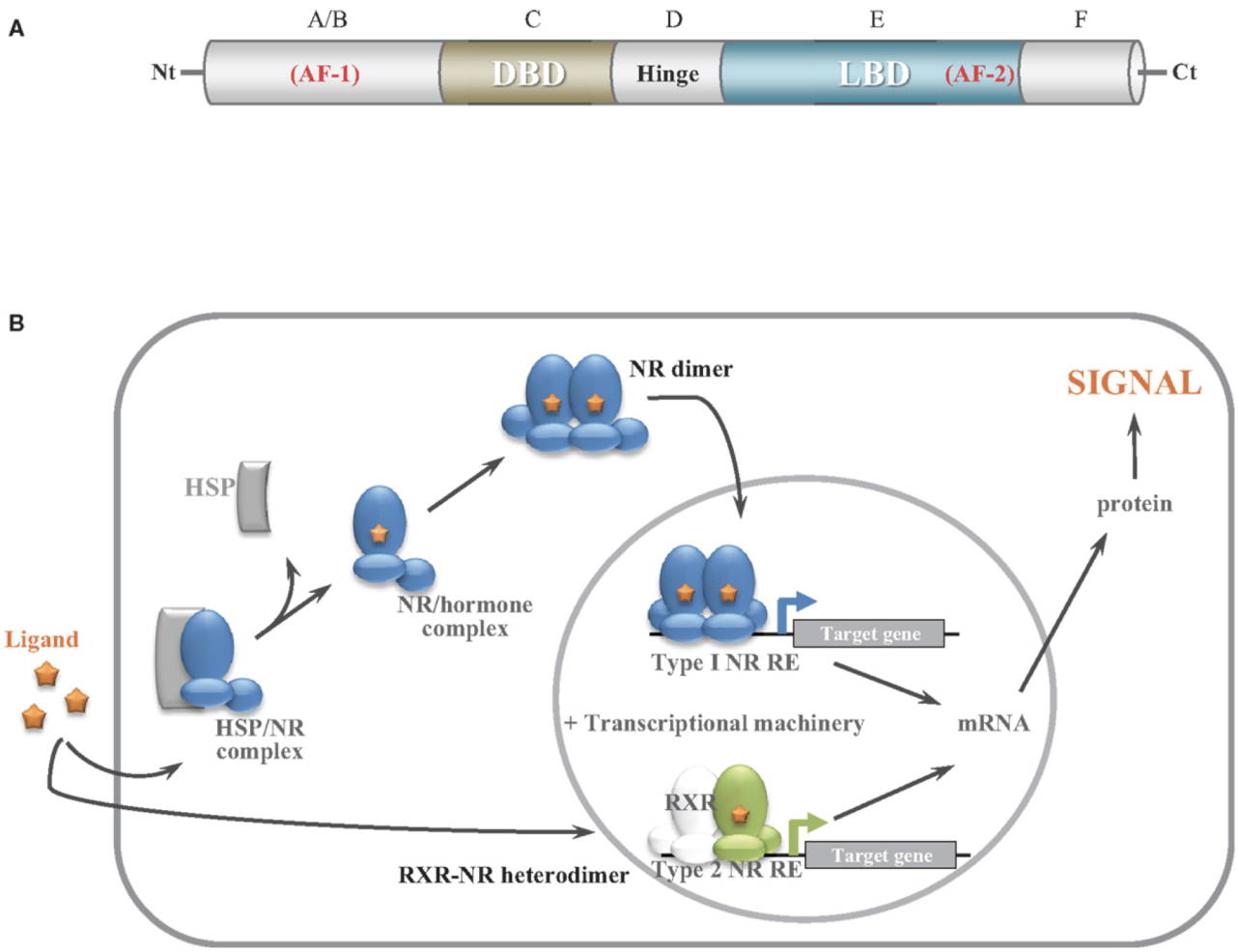
In some cases, the continued presence of FoxA transcription factors is not required for maintenance of tissue specification, especially when other Fox proteins compensate



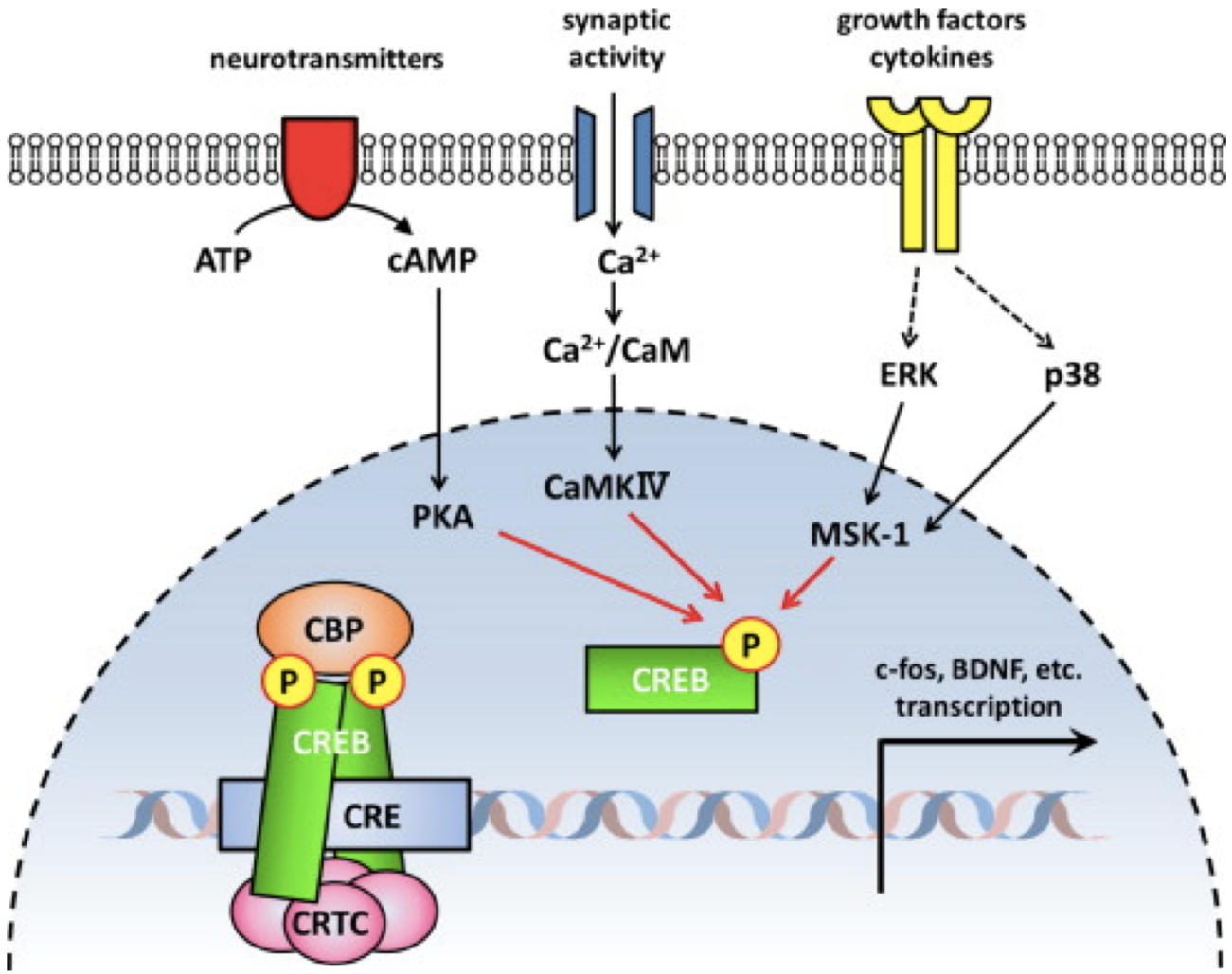
Tissue-specific TFs: MyoD example



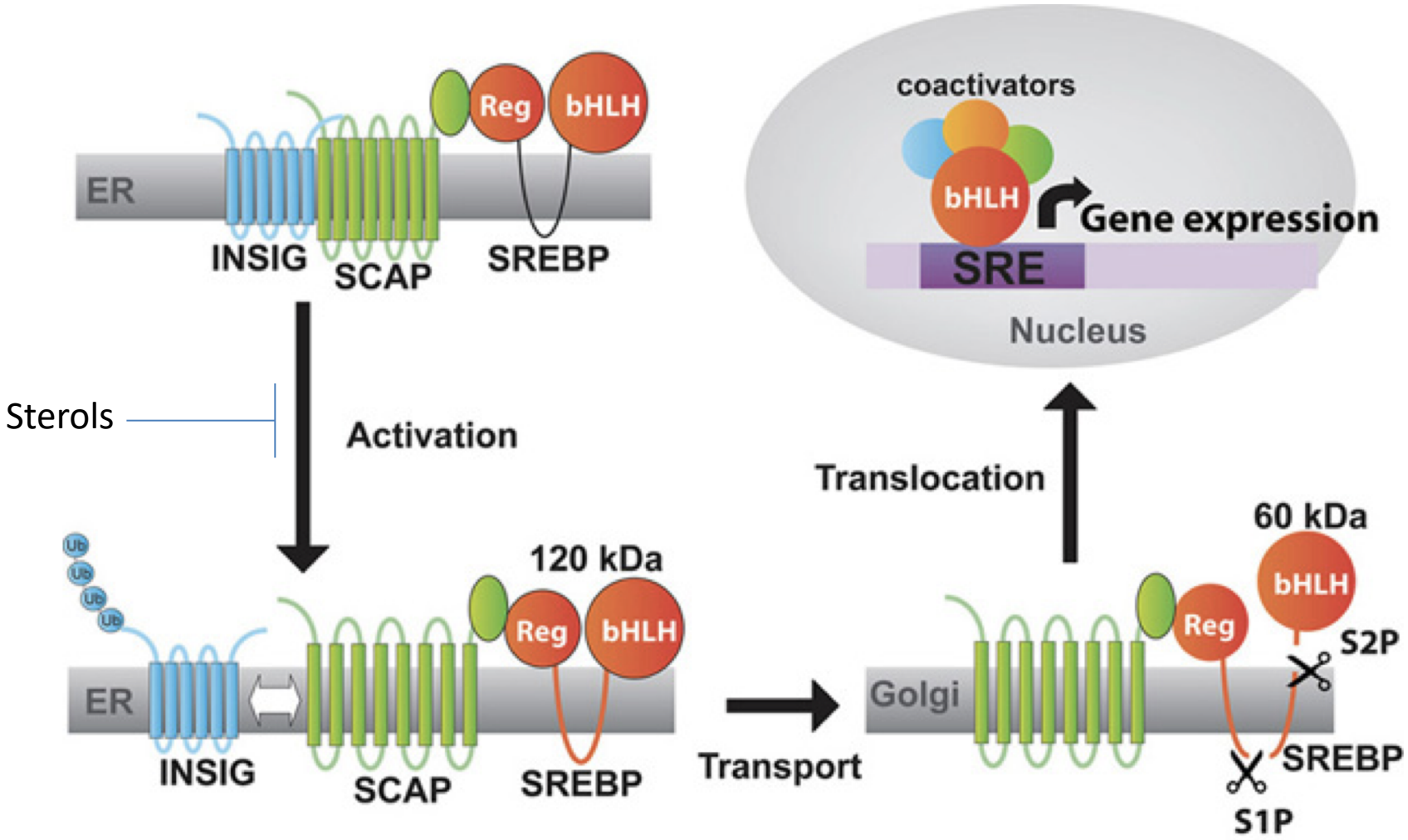
Signal-dependent TFs: Nuclear Receptors example (Ligand-based)

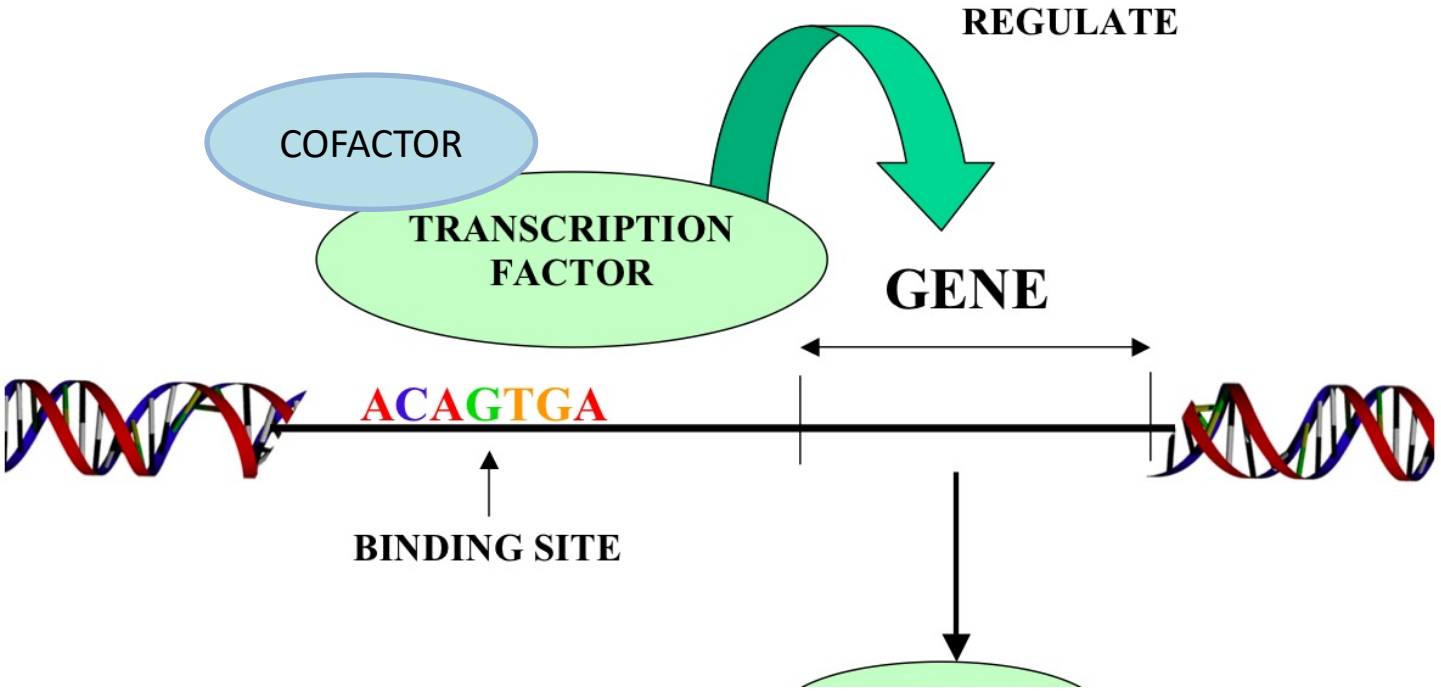


Signal-dependent TFs: CREB example (Phosphorylation-based)



Signal-dependent TFs: SREBP example (Cleavage-based)





DNA segments where short sequence motifs, 4 to 15 base long, called Response Elements and recognized by Transcription Factors, are juxtapposed.

Response Elements = **TFBS** (Transcription Factor Binding Sites)

How did we find which DNA sequences are bound by a given TF ?

Historical Approaches used to clone TFs and characterize their Response Elements

1° route:

isolating a promoter sequence, make deletional mutants and identify regulatory elements.

This is paralleled with Dnase I footprinting experiments using whole Nuclear Extract.

Once identified, the response elements are further analyzed by Band-shift (EMSA)

Proteins bound are then isolated by DNA affinity chromatography and identified.

This approach has led to the characterization of several tens of Transcription Factors.

2° route:

Several putative TFs were identified by homology cloning.

The binding site was then identified by **SELEX**

Finally, bioinformatic search for the binding site is performed on known genomic sequences.

FIRST ROUTE Example

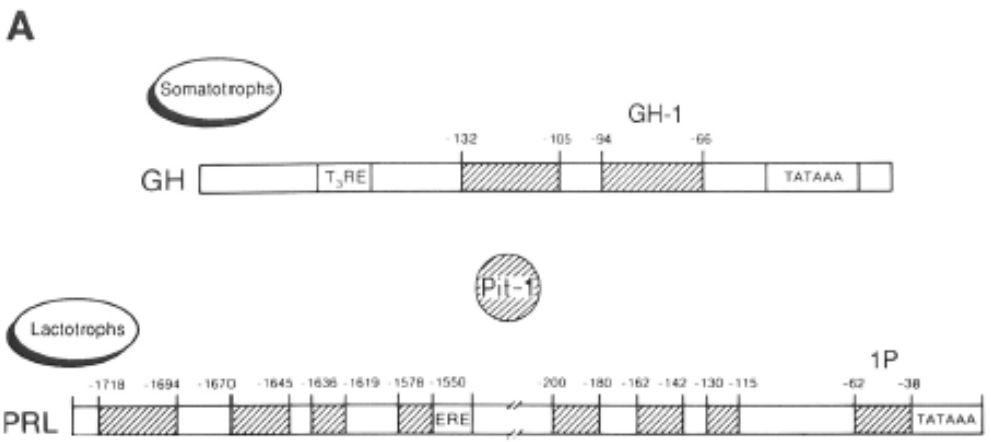
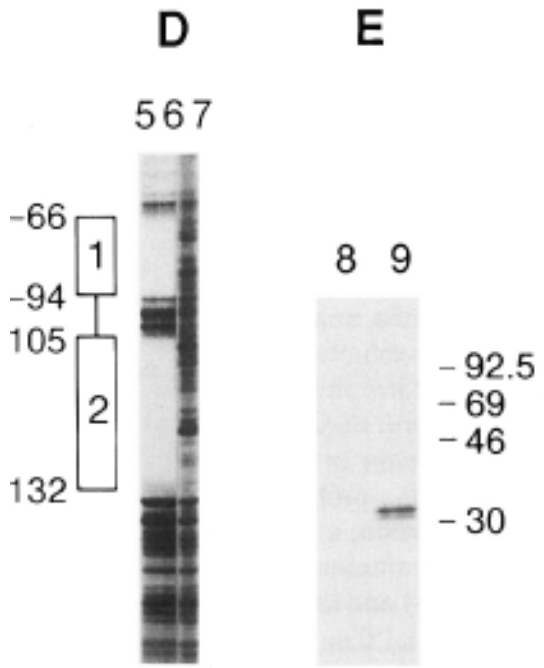


Figure 1. Characterization of a Pituitary Nuclear Protein That Binds to Prolactin Pit-I Cis-Active Elements
(A) Schematic diagram of the cell-specific elements in the rat prolactin and growth hormone genes that exhibit ceil-specific DNAase ! footprrnts and
are required for cell-specific gene expression. The factor(s) binding to these elements in the prolactin gene is referred to as Pit-I.



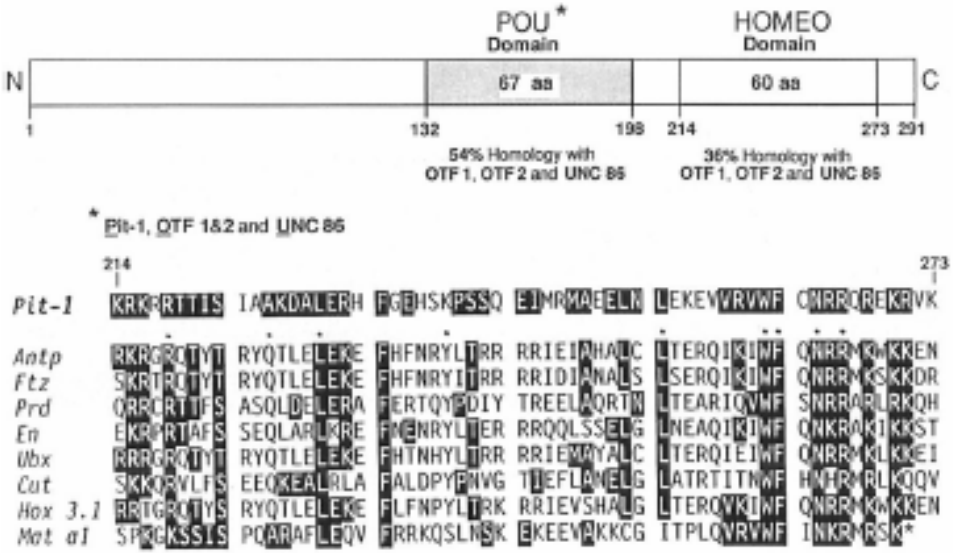


Figure 4. Putative Domains of Pit-1, and Comparison of the Carboxyl Terminus of Pit-1 with the Homeodomains of Several Regulatory Proteins

Ingraham et al., Cell 1988

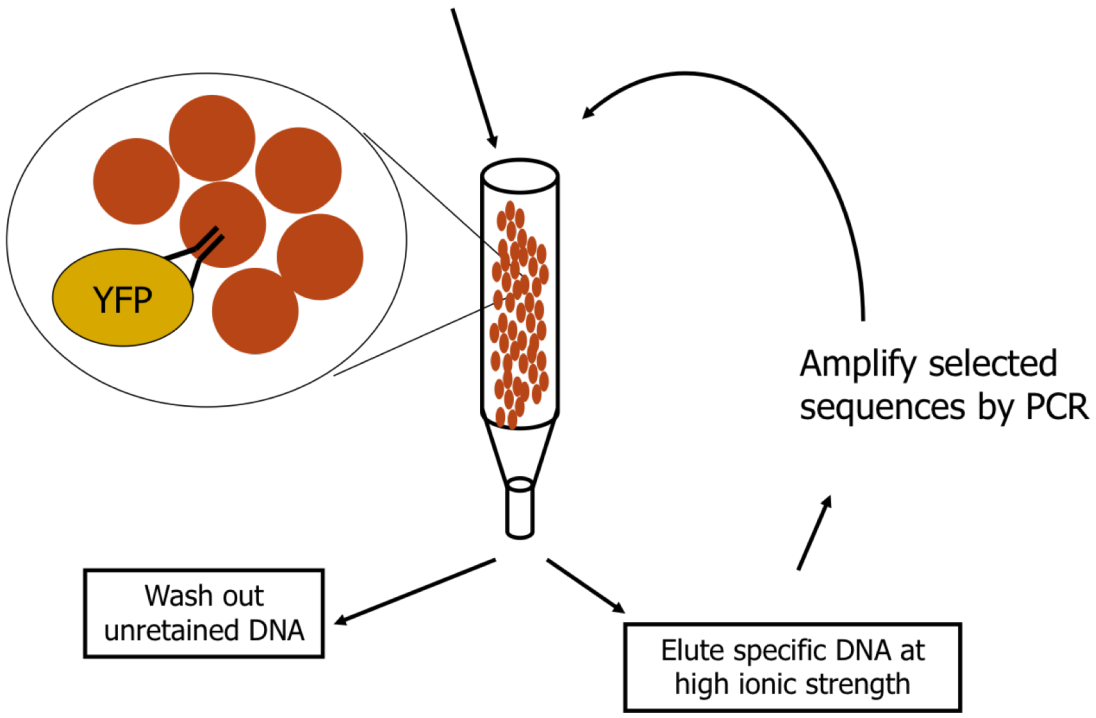
SECOND ROUTE Approach

Classic approach to identify a TF binding motif? SELEX

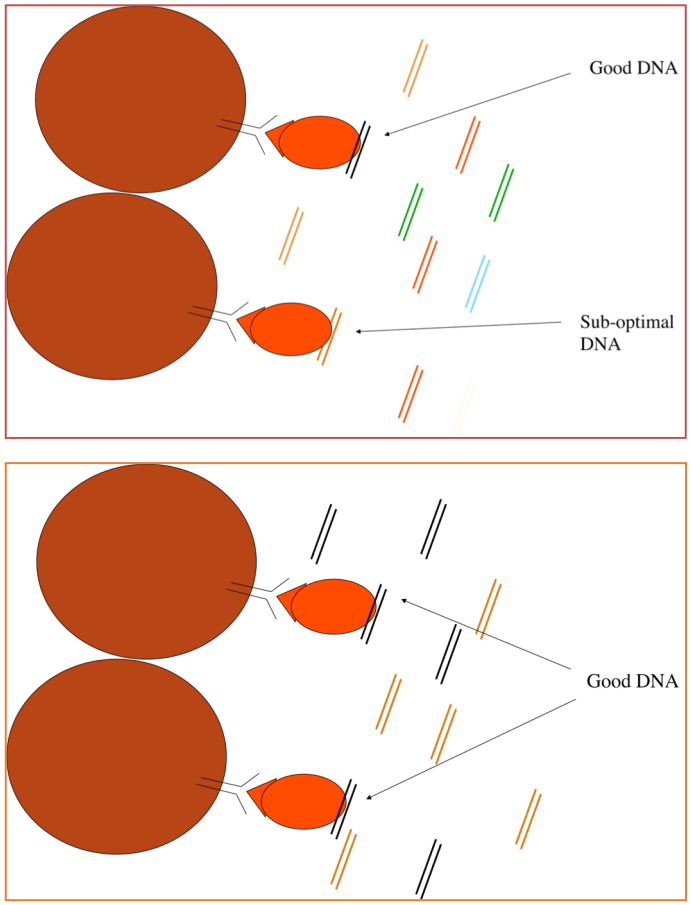
When a new DNA binding protein is under study, the sequence of DNA it interacts with can be selected using a process called Systematic Evolution of Ligands by Exponential Enrichment

```
5′sense primer- NNNNNNNNNNNN - reverse primer3′
3′remirp esnes- NNNNNNNNNNNN - remirp esrever5′
```

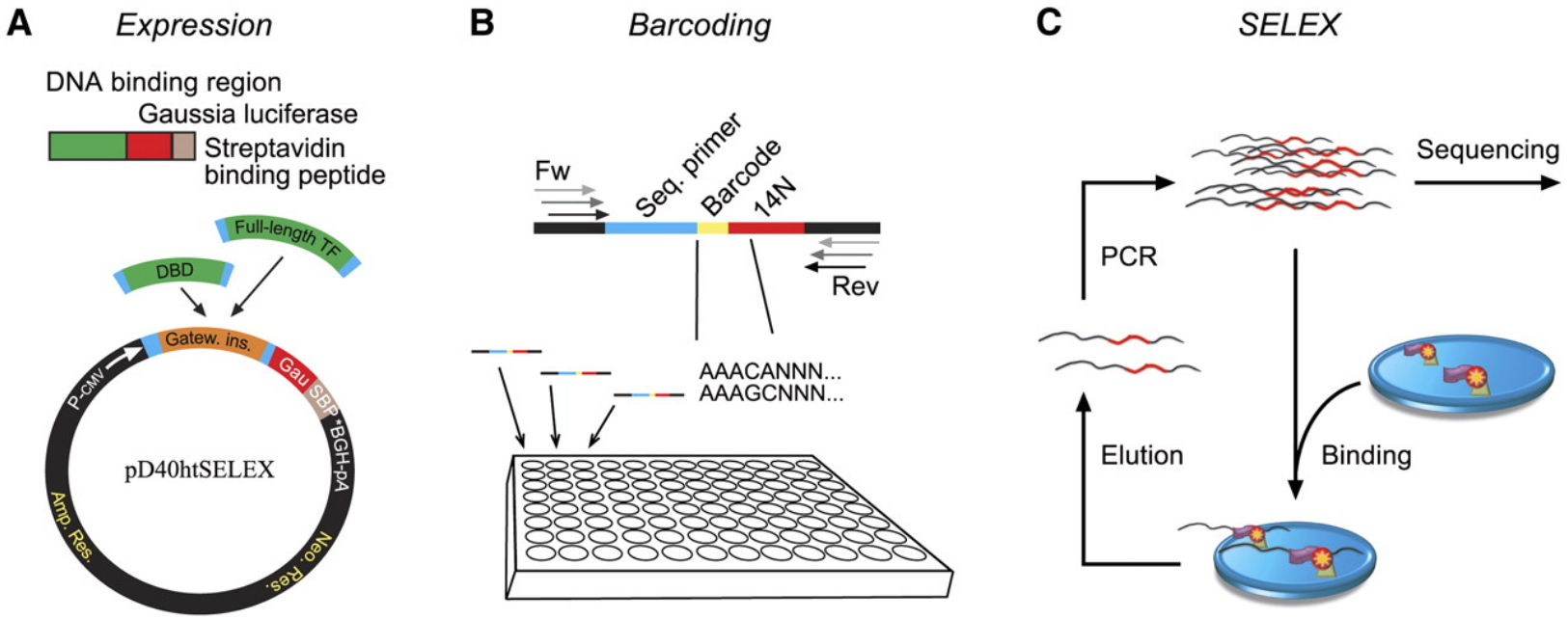
random (4^n sequences)



Exponential enrichment



Schematic description of the high-throughput SELEX process.



Arttu Jolma et al. Genome Res. 2010;20:861-873

Distance dendogram based on the minimum Kullback-Leibler divergences between TF position weight matrices from the Jaspar database and reference matrices of Figure5.

Arttu Jolma et al. Genome Res. 2010;20:861-873

Genome wide profiling of TFs binding to chromatin

- Using ChIP + microarrays or (best) ChIP-Seq it has been quite straightforward to obtain high-resolution maps of TF binding to chromatin using cell lines.

Insights from genomic profiling of transcription factors

Peggy J. Farnham

Abstract | A crucial question in the field of gene regulation is whether the location at which a transcription factor binds influences its effectiveness or the mechanism by which it regulates transcription. Comprehensive transcription factor binding maps are needed to address these issues, and genome-wide mapping is now possible thanks to the technological advances of ChIP–chip and ChIP–seq. This Review discusses how recent genomic profiling of transcription factors gives insight into how binding specificity is achieved and what features of chromatin influence the ability of transcription factors to interact with the genome. It also suggests future experiments that may further our understanding of the causes and consequences of transcription factor–genome interactions.

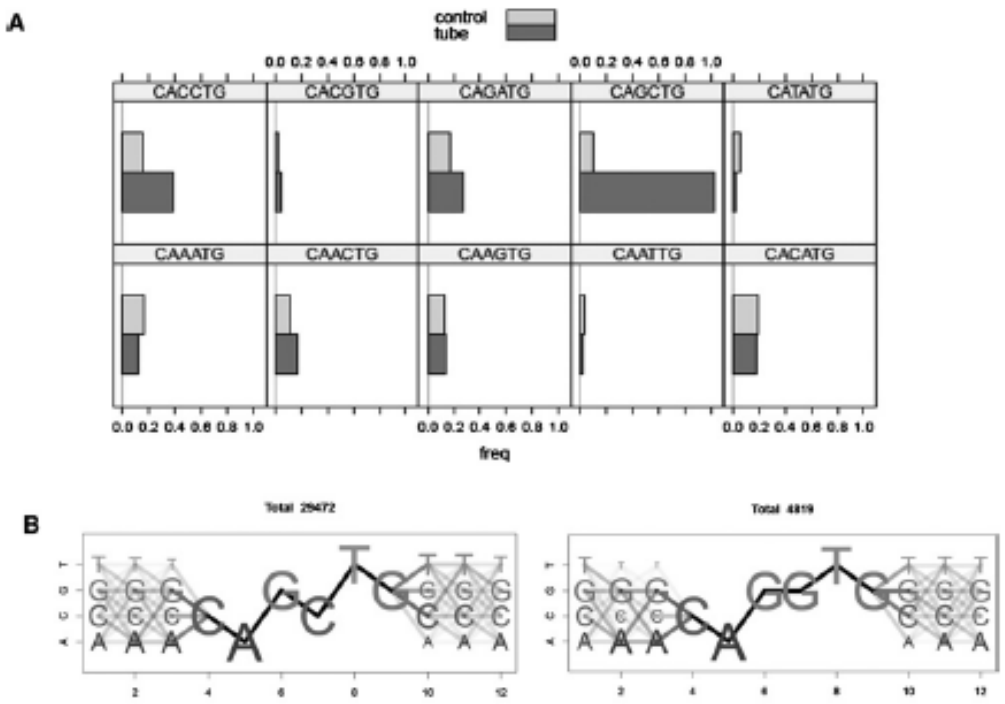
Remember : ChIp-Seq analysis identifies a region where the TF binds, not the TFBS

However, they can provide important insights into TFs – TFBSs interactions

1. Investigate TF preferences for distinct TFBS
2. Investigate how distinct factors of the same family select for TFBS
3. Identify cooperative networks

Example 1: ChIPseq for MyoD

(Investigate TF preferences for distinct TFBS)



Sequence Characteristics in Regions Bound by MyoD
in C2C12 Myotubes

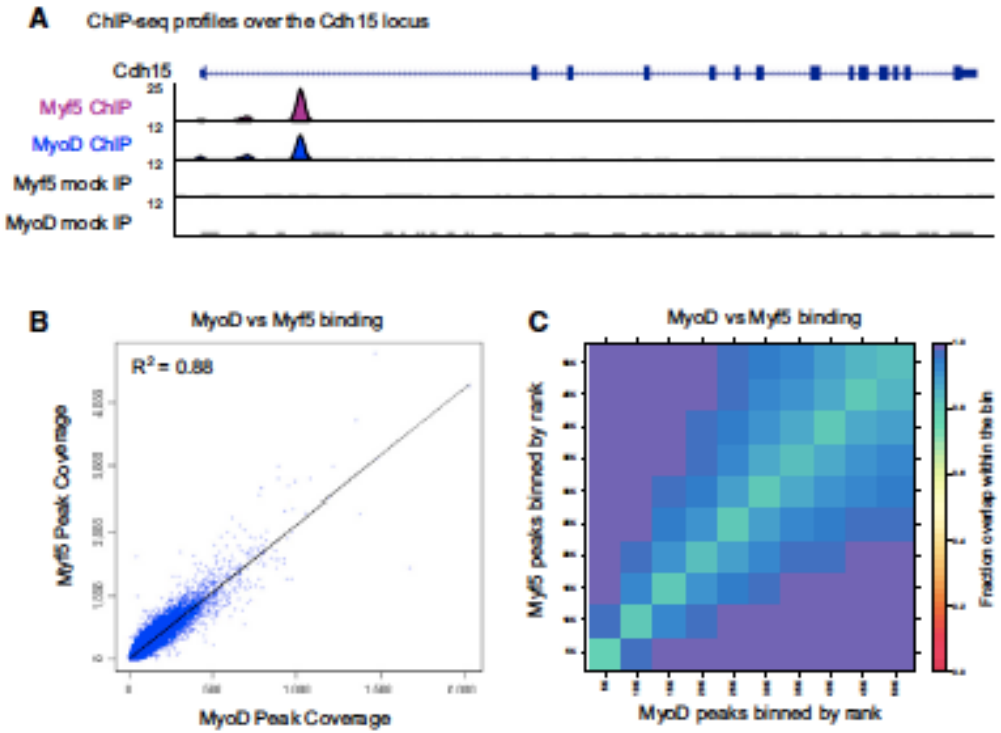
To understand the sequence determinants of MyoD binding in myotubes, we first examined E-box sequences in the 200 nt region centered at the MyoD peak summit.

A total of 96% of all peaks contain at least one canonical E-box (CANNTG), with an average of 2.4 E-boxes per peak, relative to 1.14 in comparable control regions : 72% of peaks have an E-box within 20 nt of the peak summit.

A strong sequence preference was observed for the E-box sequences in peak regions (Figure 1A): 74% of peaks have the CAGCTG E-box (enriched 11-fold); 32% have CACCTG (enriched 2.5-fold); and 89% have either or both of these twoE-box motifs.

Example 2: ChIP-seq for MyoD/Myf5

(Compare the binding of distinct factors of the same family to common TFBS)

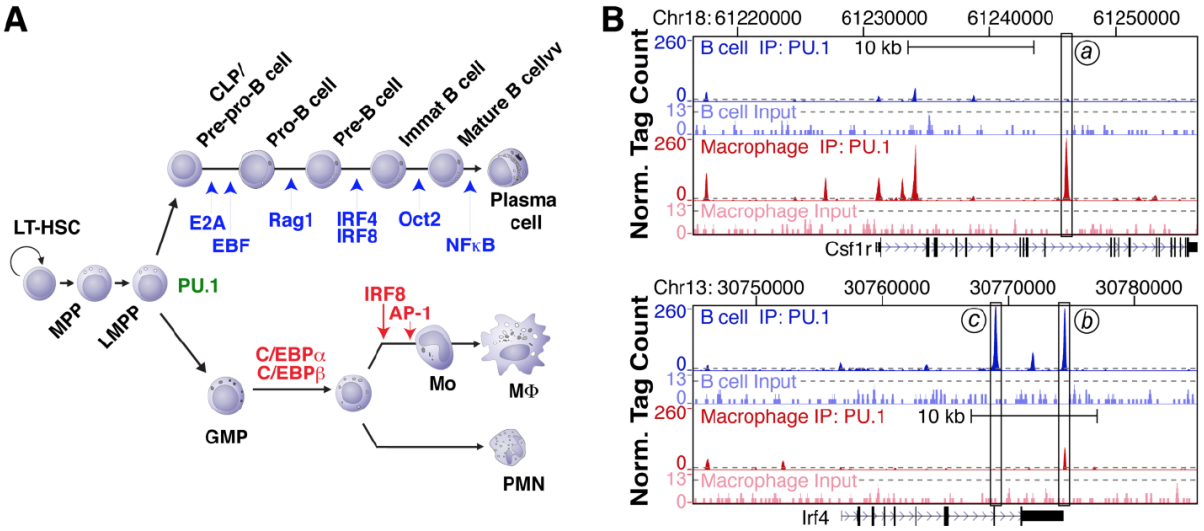


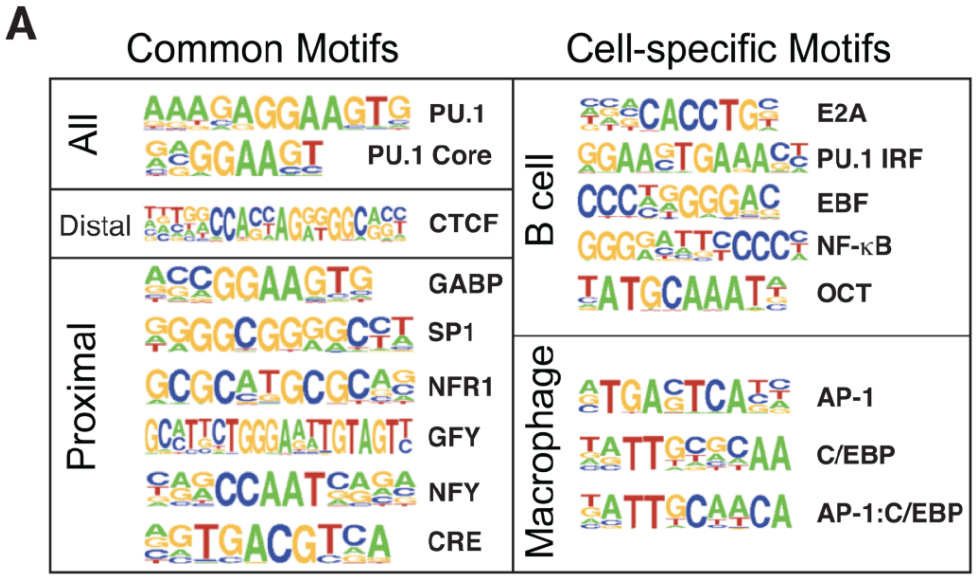
Conerly et al.
Dev Cell 2016

Figure 2. MyoD and Myf5 Bind the Same Subset of E-Boxes
(A) ChIP-seq reads over the Cdh15 promoter. (B) Overlap of MyoD and Myf5 ChIP-seq peaks plotted by peak coverage. (C) Rank comparison of the top 50,000 peaks.

MyoD and Myf5 bind the same sites genome-wide but have distinct functions: Myf5 induces histone acetylation without Pol II recruitment or robust gene activation, whereas MyoD induces histone acetylation, recruits Pol II, and robustly activates gene transcription

Example 3: ChIP-seq for PU.1
(Identify cooperative networks)





Heinz et al., Mol Cell, 2010

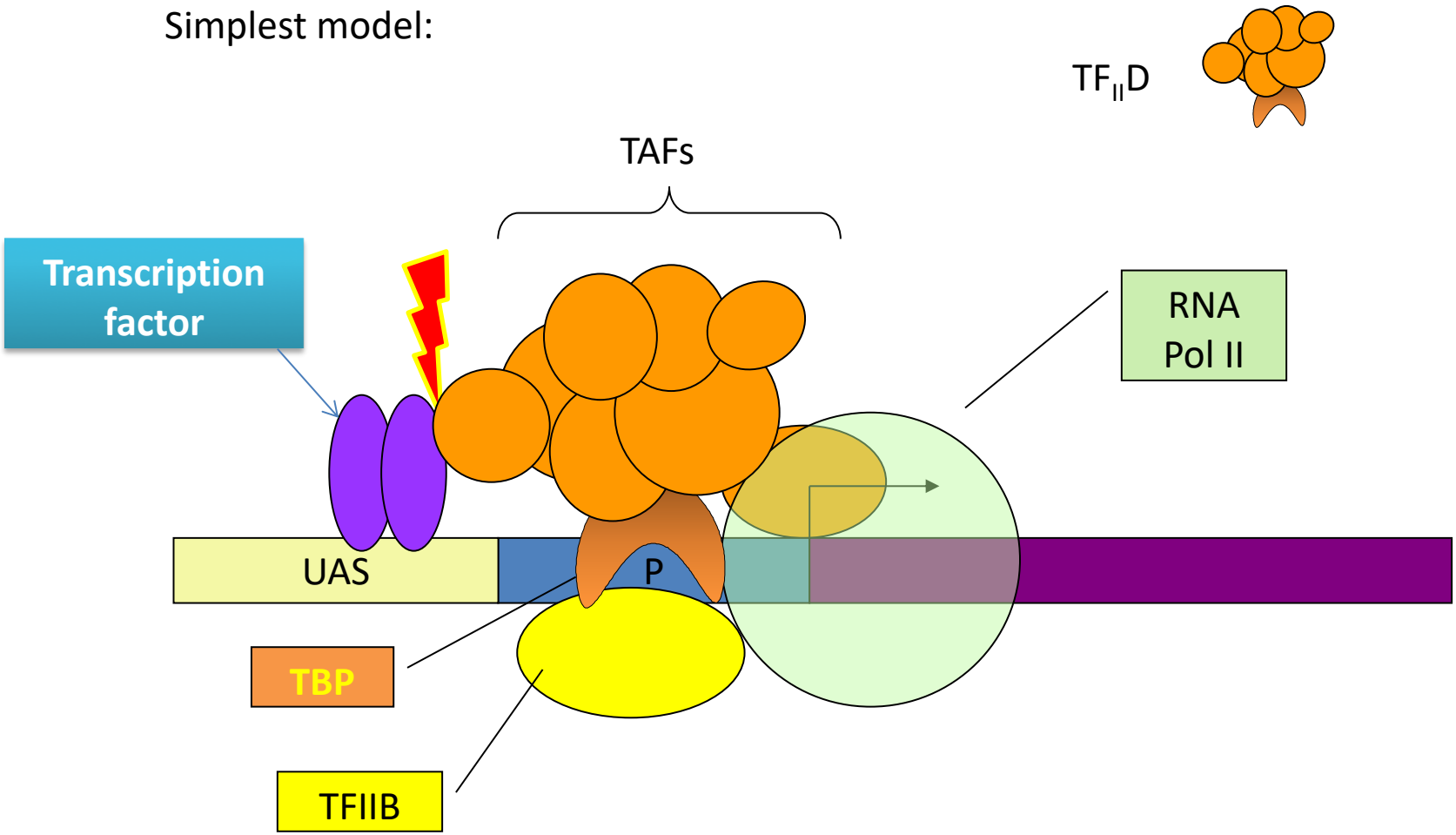
BREAK

AGENDA

1. Transcription factors (Definition, Families and Response Elements)
2. Mediator
3. Coregulators
4. Corepressors/Coactivator Exchange Model

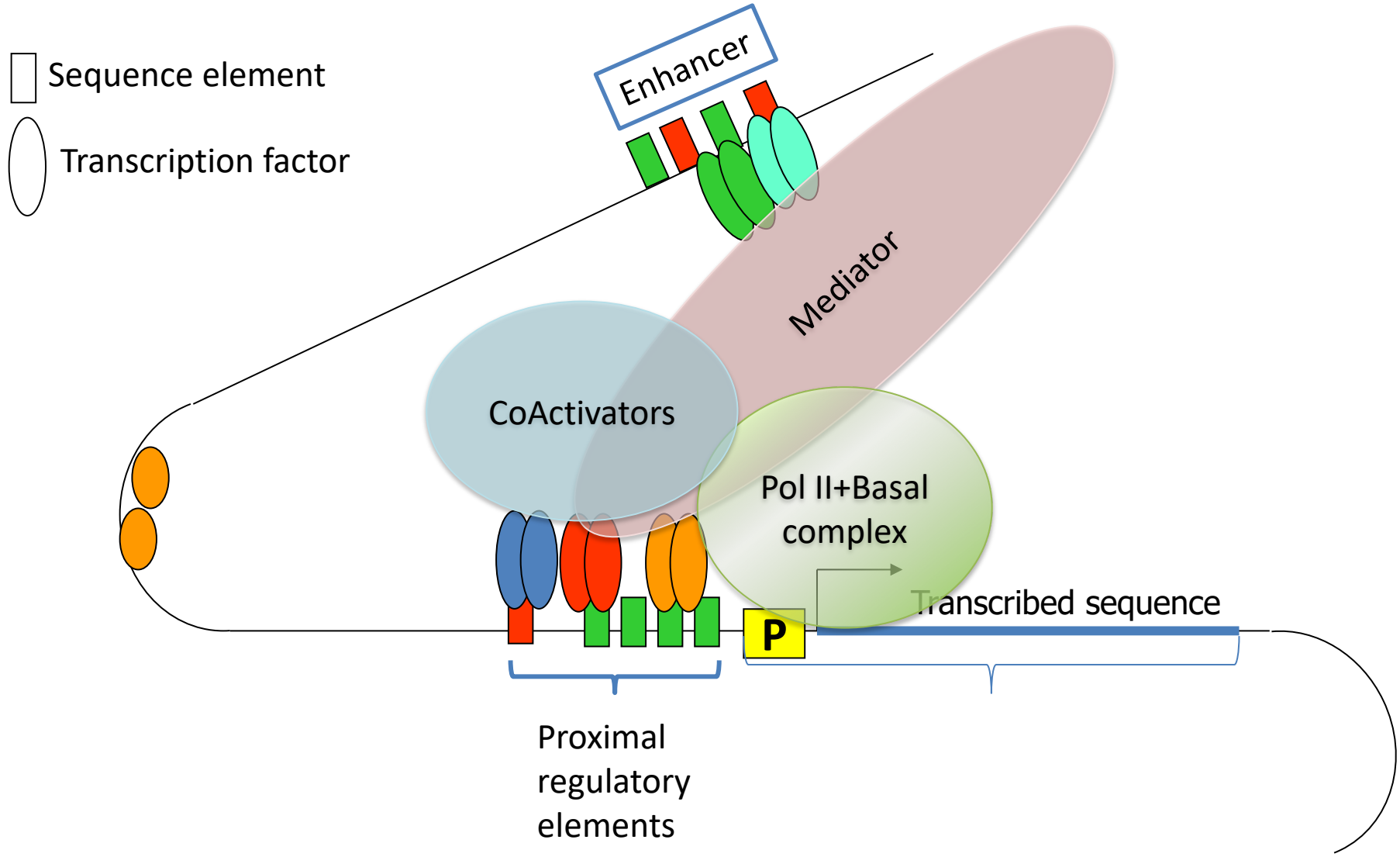
How do TFs regulate transcription ?

Simplest model:



Some Transcription Factors bound to proximal elements will contact components of the PIC, primarily TAF proteins. This was shown experimentally by reconstitution experiments.

From older papers by R. Tjian quoted in Levine, 2014


Sequence element
Transcription factor
Enhancer
Mediator
CoActivators
Pol II+Basal complex
Transcribed sequence
P
Proximal regulatory elements

HOW WERE THESE COMPONENTS IDENTIFIED?

Mediator

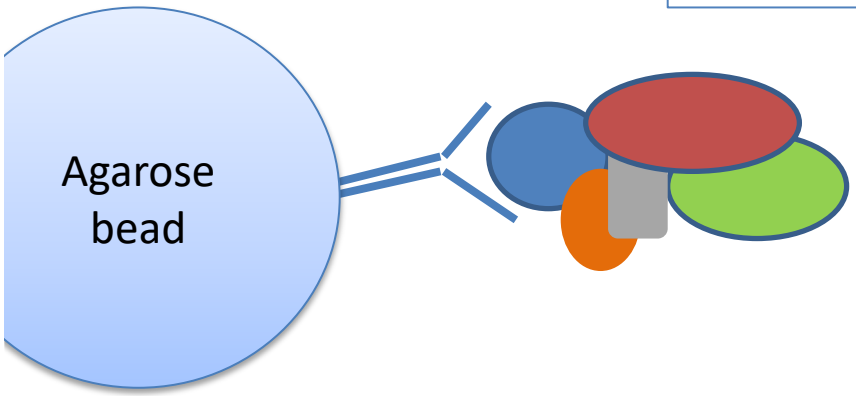
Cells or tissue

Nuclei

2000-2005

Nuclear extract (high salt)

Immunoprecipitation using anti-TF Ab (or other component)
or
using Tagged proteins and Ab against Tag (e.g. FLAG or HA epitope)



Mass spectrometry (identify)
Immunoblotting (validate)

Subunit compositions of mediator complexes.

Different forms of "Mediator" Complex

Unified subunit designation	DRIP Mediator-D	ARC Mediator-A	TRAP/ SMCC Mediator-T/S	PC2 Mediator-P	CRSP Mediator-C	NAT Mediator-N	hMediator Mediator-S	Murine Mediator Mediator-M	*S. cerevisiae*	*C. elegans*	*Drosophila*
		CBP/p300									
Med240	DRIP250	ARC250	TRAP240				ND				dTRAP240
Med230	DRIP240	ARC240	TRAP230			p230	ND	p160a	Nut1	Sop-1	dTRAP230
Med220	DRIP205	ARC205	TRAP220	(TRAP220)	CRSP200		ND	p160b	Gal11		dTRAP220
Med150	DRIP150	ARC150	TRAP170	TRAP170	CRSP150	p150	ND	Rgr1/p110	Rgr1		dTRAP170
Med130	DRIP130	ARC130	TRAP150	TRAP150b	CRSP130	p140/hSur2	hSur2			Sur-2	CG3695
Med105		ARC105/ TIG-1									
Med100	DRIP100	ARC100	TRAP100	TRAP100			ND		Sin4		dTRAP100
Med97	DRIP97		TRAP97			p95		Ring3/p96a	Srb4		
Med95	DRIP92	ARC92	TRAP95	TRAP95		p90	ND	p96b	Med1		dTRAP95
			TRAP93								
Med78	DRIP77	ARC77	TRAP80	TRAP80	CRSP77		ND	p78			dTRAP80
Med70	DRIP70-2	ARC70			CRSP70	p70	ND		Med2		
		ARC42		p37				p55	Pgd1/Hrs1		
Cdk8	(Cdk8)	(Cdk8)	hSrb10			p56/Cdk8	Cdk8		Srb10		dCdk8
Med36	DRIP36	ARC36		p36	CRSP34	p45	ND	p34	Med4		CG8609
Med34	DRIP34	ARC34	hMed7	hMed7	CRSP33	p37	Med7	Med7/p36	Med7	ceMed7	dMed7
						p36			Srb5		
Med33	DRIP33	ARC33	hMed6	(hMed6)		p33	ND	Med6/p32	Med6	ceMed6	CG9473
		ARC32	hTRF	hTRF			ND	TRF/p28a	Med8		
Cyclin C			hSrb11			p31/ Cyclin C	Cyclin C		Srb11		
						p30		p28b	Rox3		
						p23			Srb2		
			hSoh1	hSoh1		p22					
						p21			Med9/Cse2		
Med17	hSrb7		hSrb7	hSrb7		p17	ND	Srb7/p21	Srb7	ceSrb7	CG17397
Med10	hMed10		hNut2	hNut2		p14	ND		Med10/Nut2	ceMed10	dNut2
									Med11		
									Srb6		

Different forms of Mediator exist in different cell types/developmental stage and possibly gene context, depending on the kind of TFs bound.

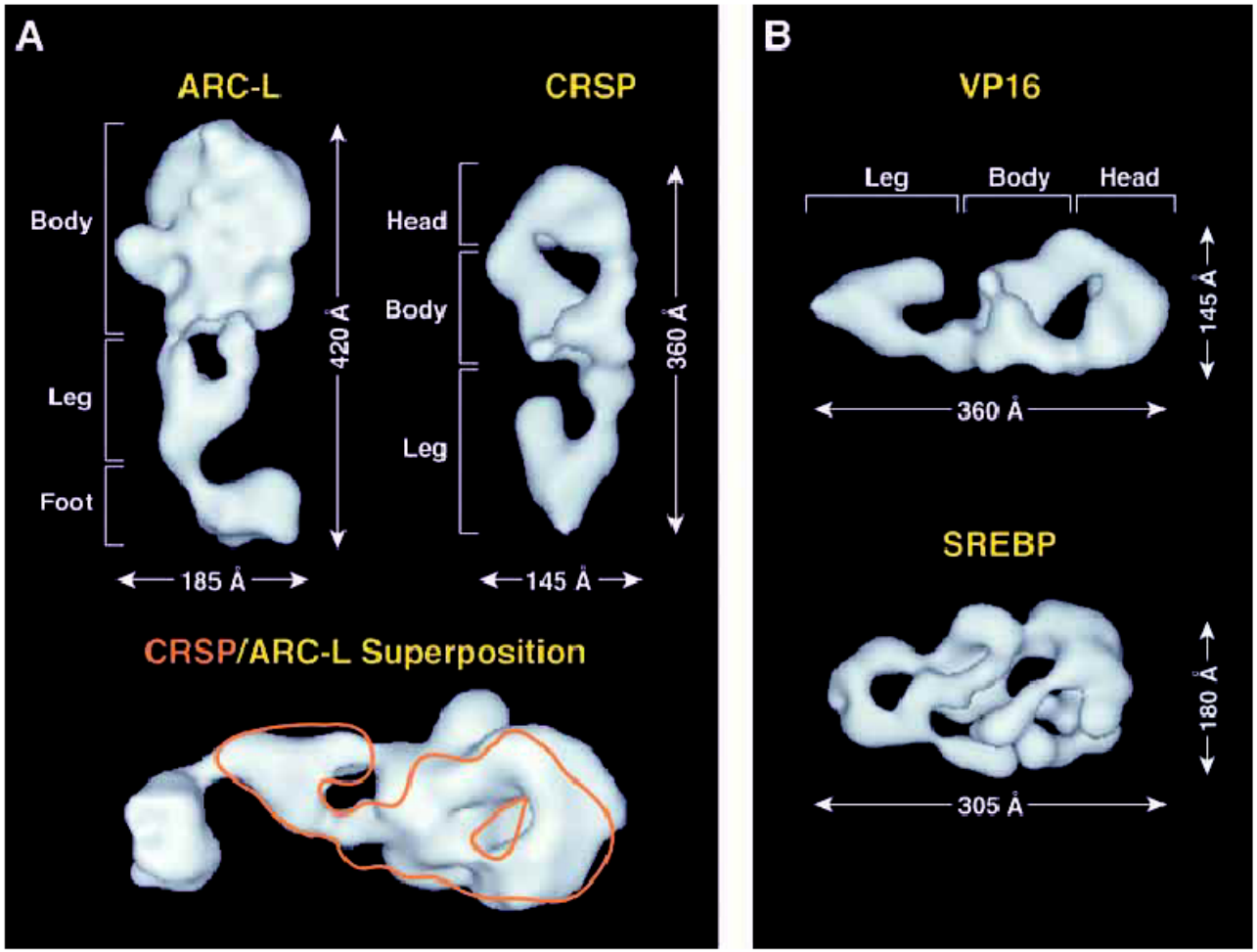
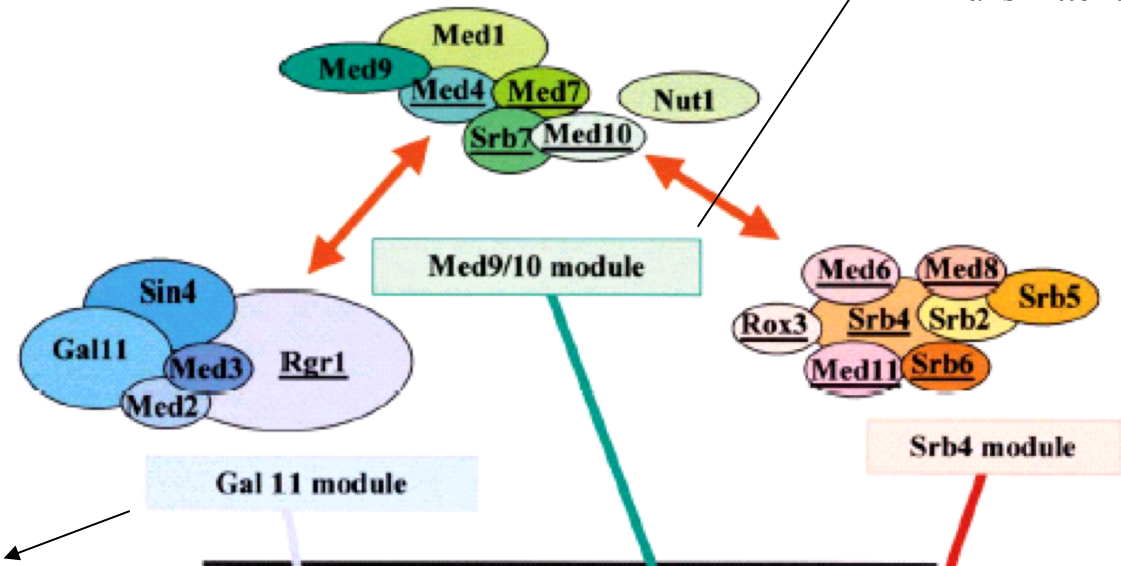


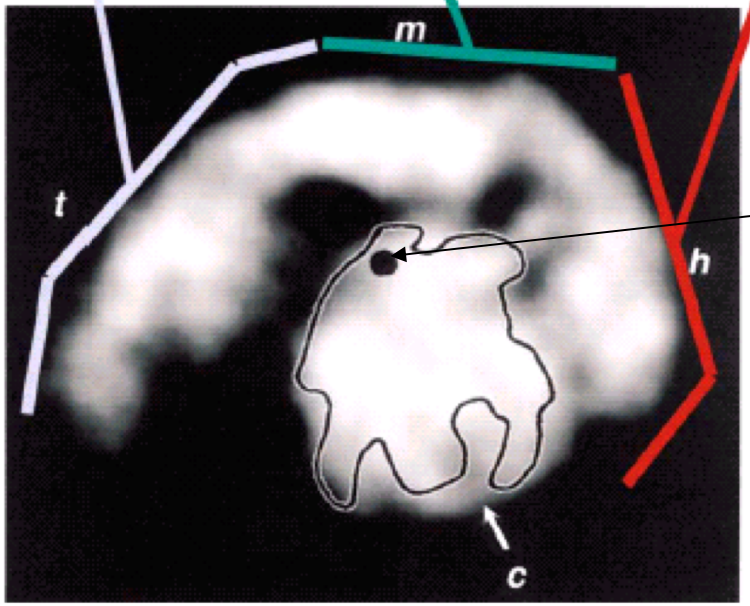
Fig. 2. Conformations of the mammalian mediator complexes. (A) EM composites of the ARC-L and CRSP complexes, which illustrate the size and structural differences between the two. (B) EM composites showing the distinctly different structural conformations adopted by CRSP when isolated via affinity interactions with either the VP16 or SREBP activator. EM composites were generously provided by Dylan Taatjes and Bob Tjian (Naar et al., 2002; Taatjes et al., 2002).

Mediator

Interactcs with CTD
Both activator and repressor signal
Transmitter to Pol

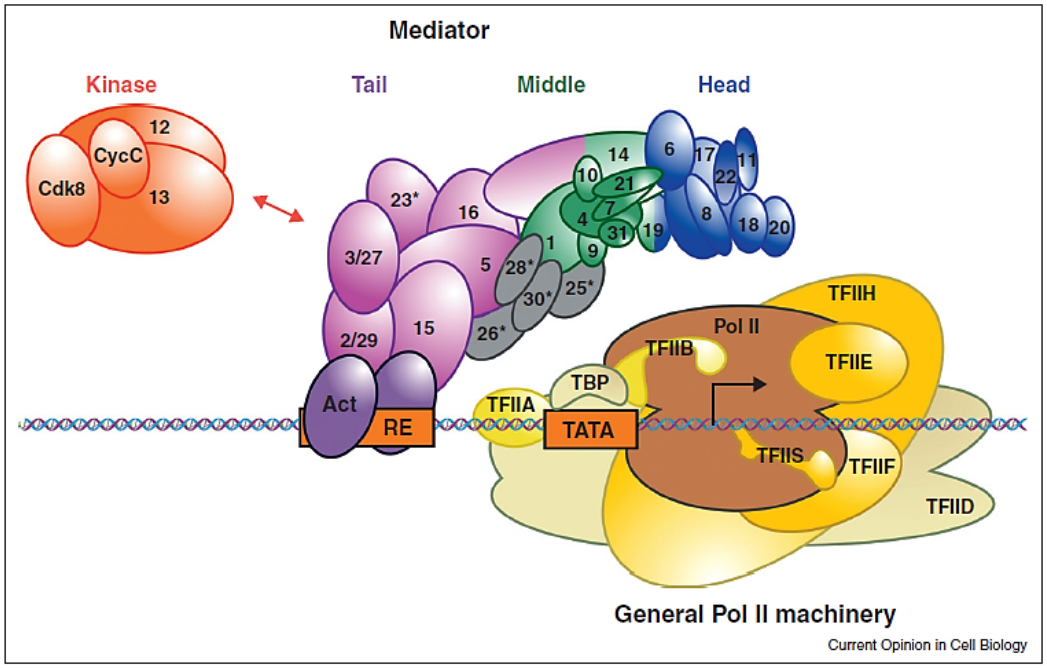


Role in contacting
trans-activators



CTD

Mediator



From:
Lariviere et al.
Curr Op Cell Biol.
2012, 24:305–313

RNA Polymerase II pre-initiation complex comprising Mediator and the general Pol II machinery at the promoter. Mediator bridges between gene specific activators (Act) bound to regulatory DNA elements (RE) and the general transcription machinery comprising Pol II and the general transcription factors TFIIA, TFIIB, TFIID/TBP, TFIIE, TFIIF, and TFIIH, and the factor TFIIS. The transcription start site is indicated with a black arrow. The Mediator modules head, middle, tail, and kinase are colored blue, green, purple, and orange, respectively. Mediator subunits Med14 and Med19 are probably bridging between modules and are therefore shown in two colors. Subunits that are not assigned to any module are colored gray. Yeast Mediator subunits Med2 and Med3 are identical to human Mediator subunits Med29 and Med27, respectively. Subunits present only in higher eukaryotes are marked with an asterisk.

Coregulators (Coactivators and Corepressors)

Large protein complexes exhibiting several functions:

- Interaction with Transcription Factors (sequence-specific TF)
- Histone PTM writers/erasers (i.eHAT histone acetyl transferase)
- Chromatin remodeling factors
    - (ATP-dependent chromatin remodelers are a class of proteins that «remodel» nucleosomes over DNA, usually in ATP-dependent fashion)

CoA and **CoR** may be recruited:

a) directly by Transcription Factors

b) through interaction with PIC or Mediator subunits

Prototype Coactivator: CBP

Cofactors

Histone acetylation

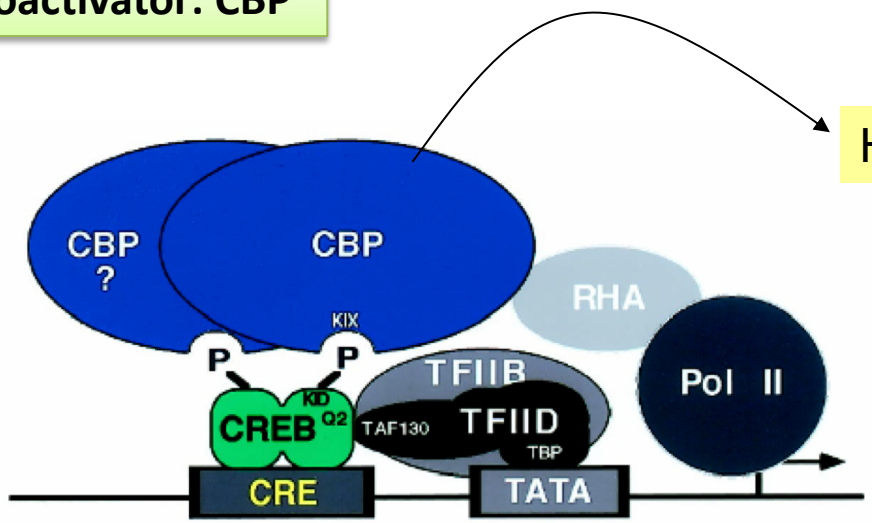
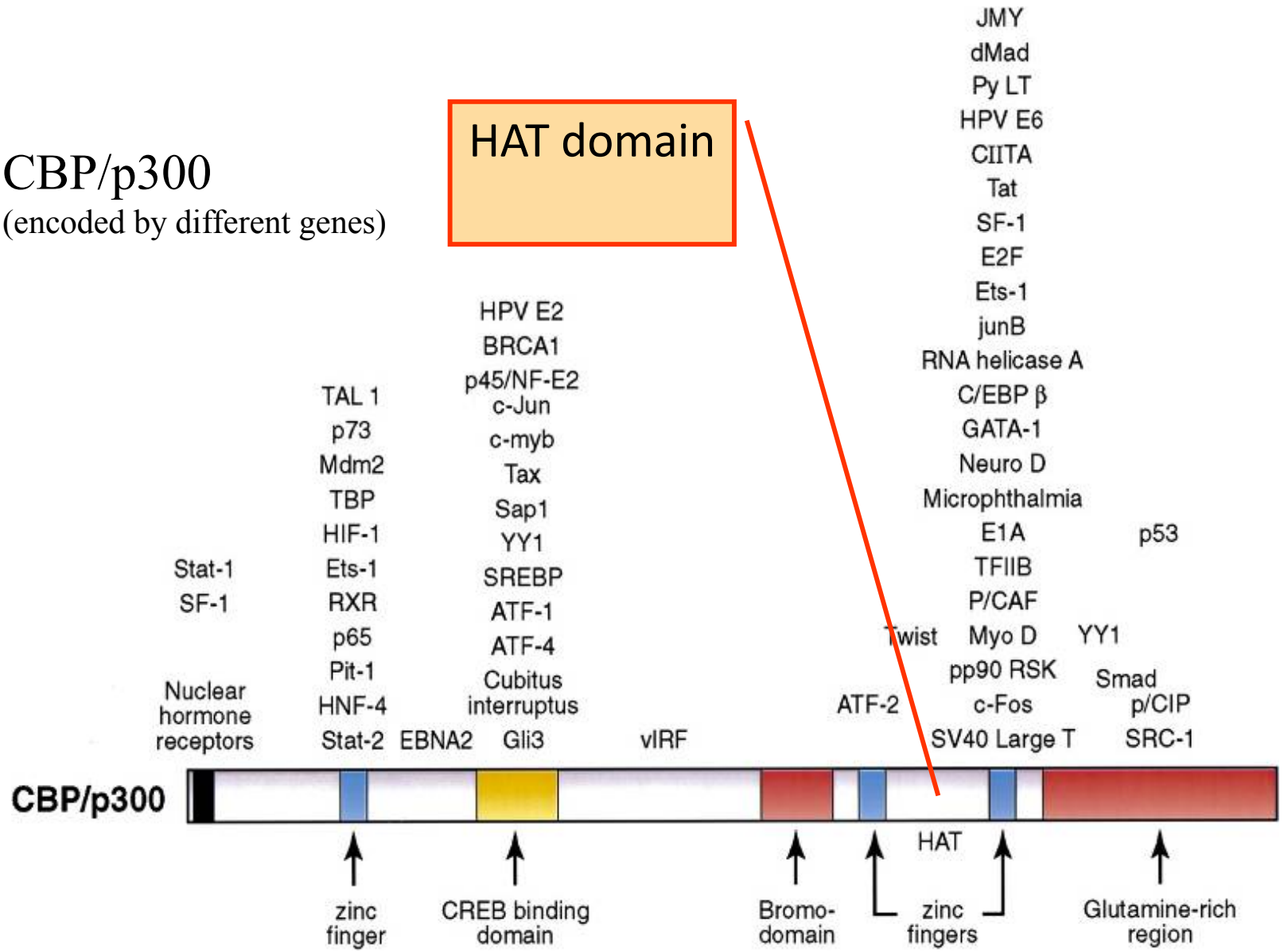


Figure 4 Multiple domains of CREB contribute to transcriptional activation. Different domains of CREB bind distinct coactivators and basal transcription factors to activate transcription. Shown is a CREB dimer bound to its cognate CaRE/CRE element on the promoter of a CREB target gene. Downstream of the CaRE/CRE is the TATA box, which binds the multiprotein TFIID basal transcription factor (via the TBP protein). Another factor within TFIID, TAF130, binds to the Q2 domain of CREB. The Q2 domain of CREB has also been shown to interact with TFIIB, which is a part of the basal transcription machinery as well. A distinct domain of CREB, the KID, contributes to signal-induced transcriptional activation. When phosphorylated at Ser133, the KID of CREB can bind to the KIX domain of the CBP. It is presently unclear whether CBP associates with Ser133–phosphorylated CREB as a dimer. CBP associates indirectly with Pol II via the RNA helicase A (RHA) protein. Therefore, recruitment of CBP to Ser133–phosphorylated CREB results in recruitment and stabilization of Pol II on the promoter of CREB target genes, whereas the Q2 domain interacts with other elements of the basal transcription machinery that are required for transcription, such as TFIID and TFIIB.

cAMP response pathway, CREB

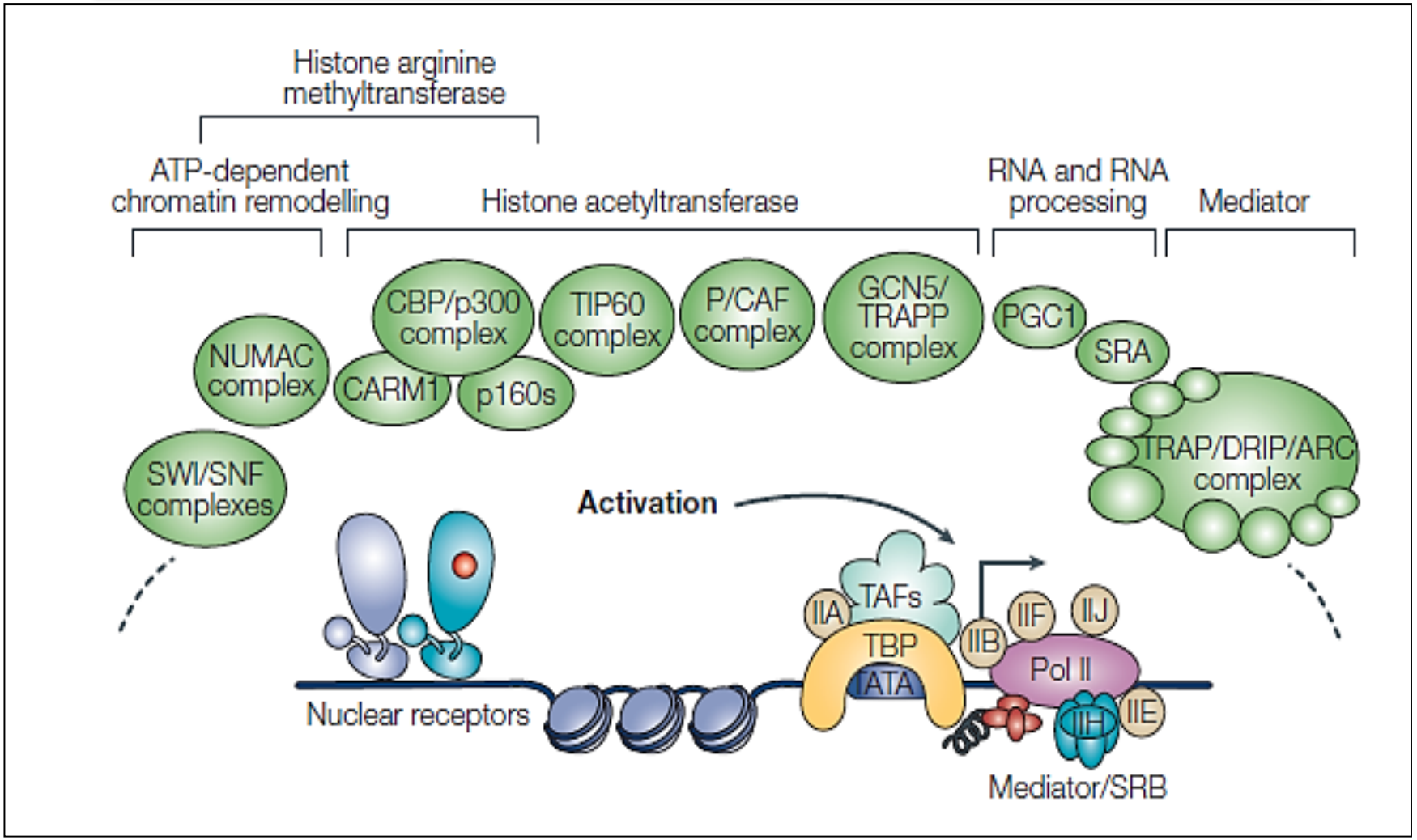
CBP = CREB binding protein (265KDa) and p300 are in fact **general coactivators**

CBP/p300
(encoded by different genes)



Cofactors

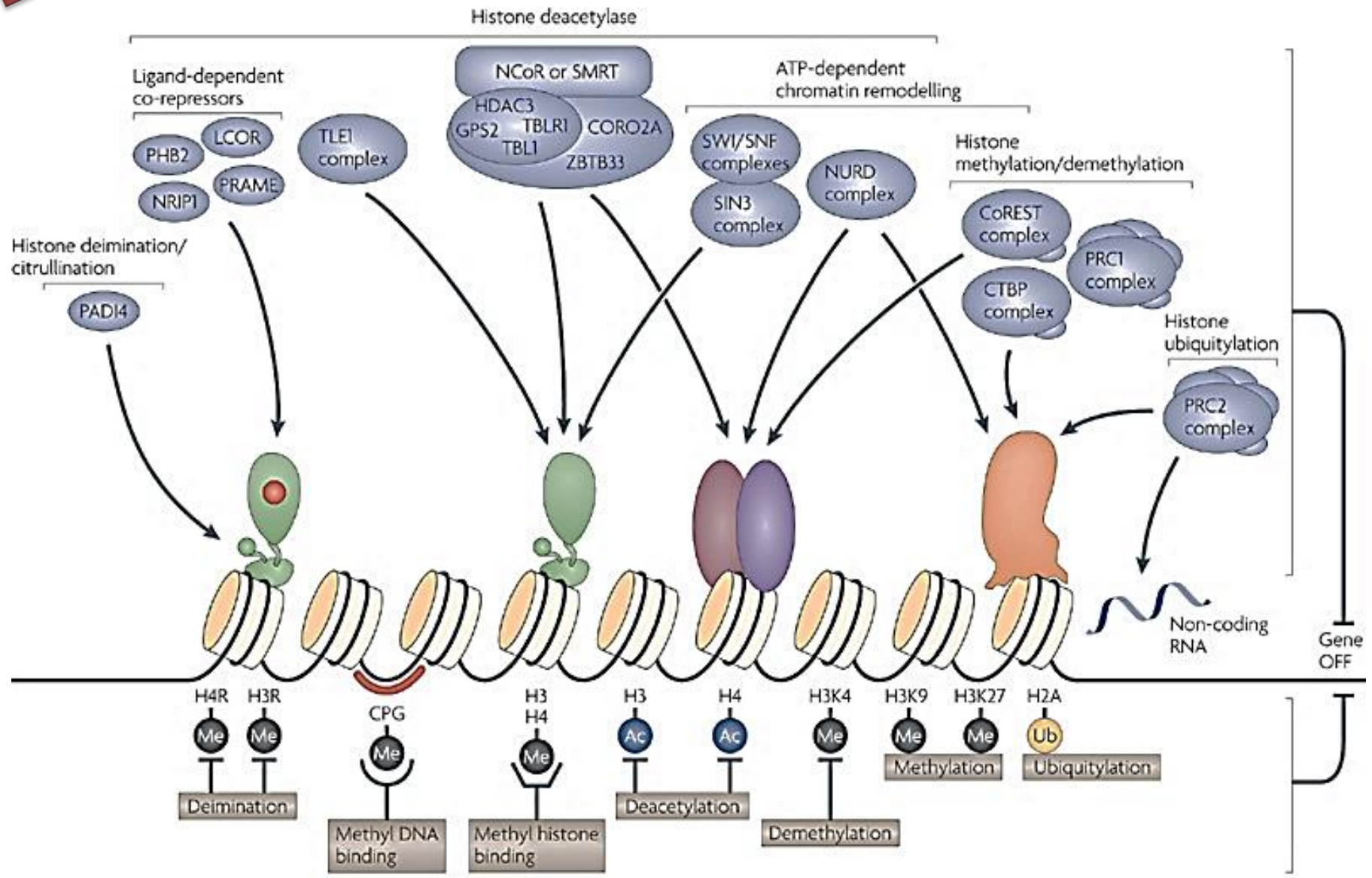
Co-activators participating in transcriptional activation by Nuclear Receptors



From: Perissi & Rosenfed, 2005, Nature Rev Mol Cell Biol, 6: 542-554

Cofactors

Co-repressors participating in transcriptional repression by Nuclear Receptors



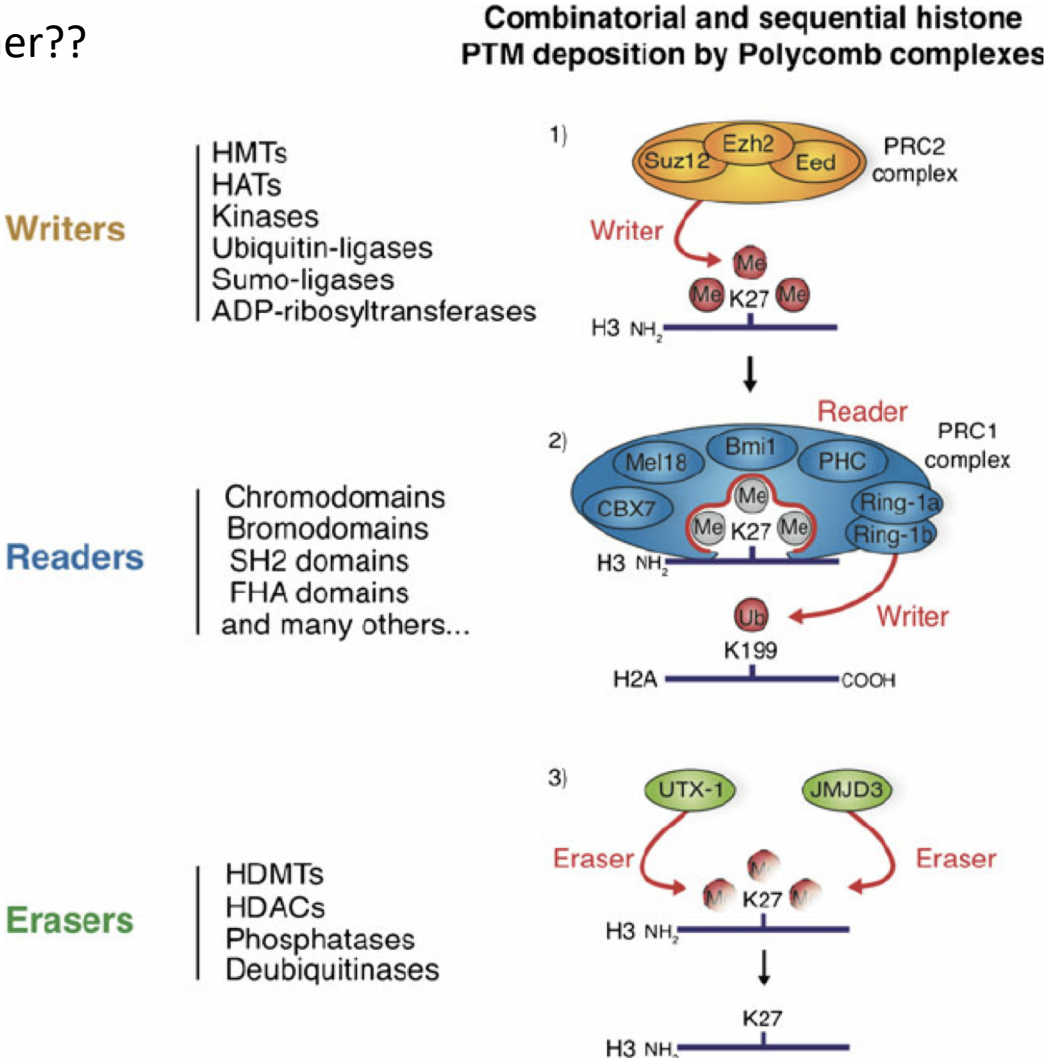
Perissi, et al. *Nature Reviews Genetics* **11**, 109-123 (February 2010)

Nature Reviews | Genetics

Discussion Topics

- Are these all binding together??

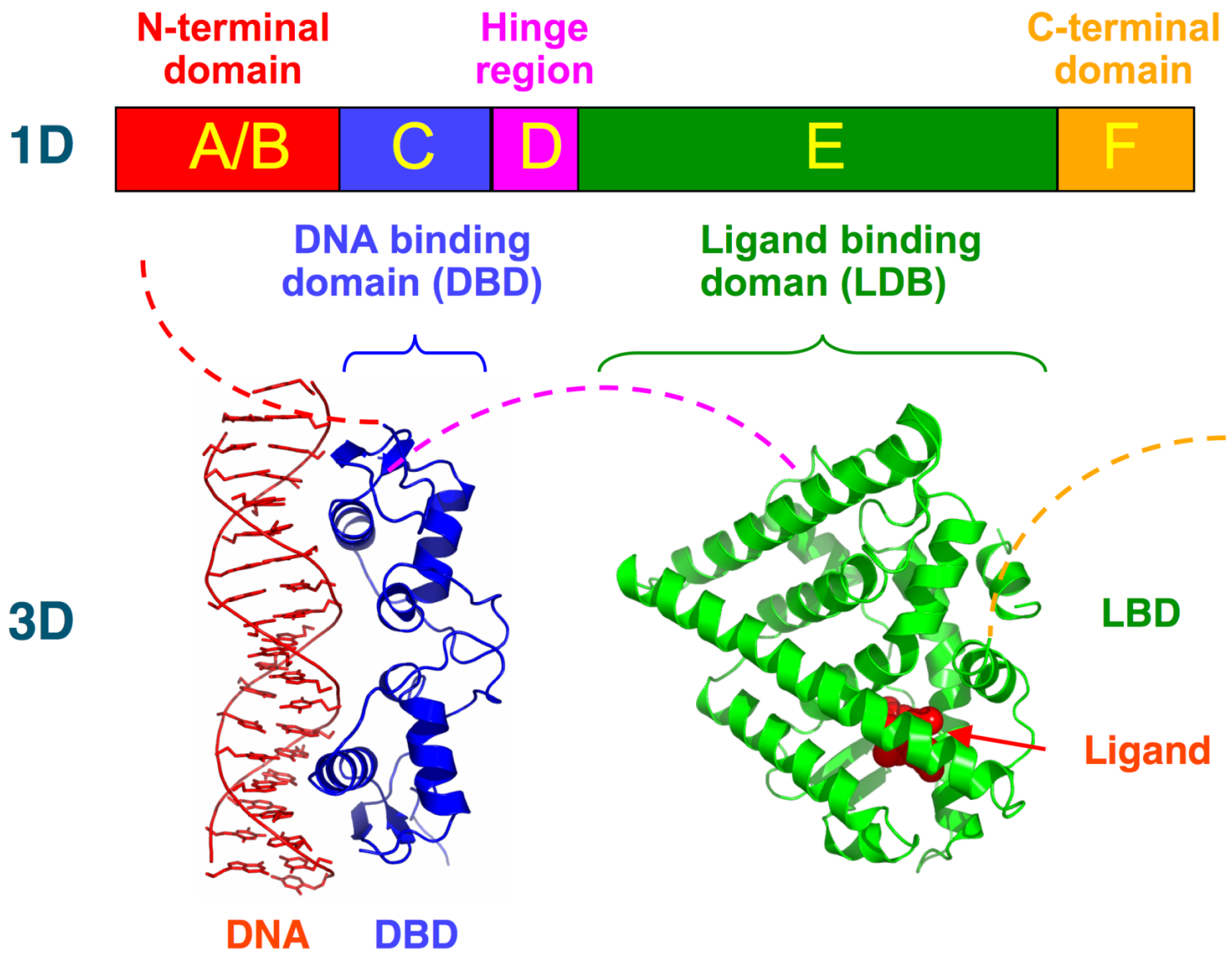
-> Writer/Eraser/Reader Code



- Are CoA and CoR binding to the same or different TFs?

-> Repressors/Activators/Inducible TFs

Structural Organization of Nuclear Receptors



NRs classification

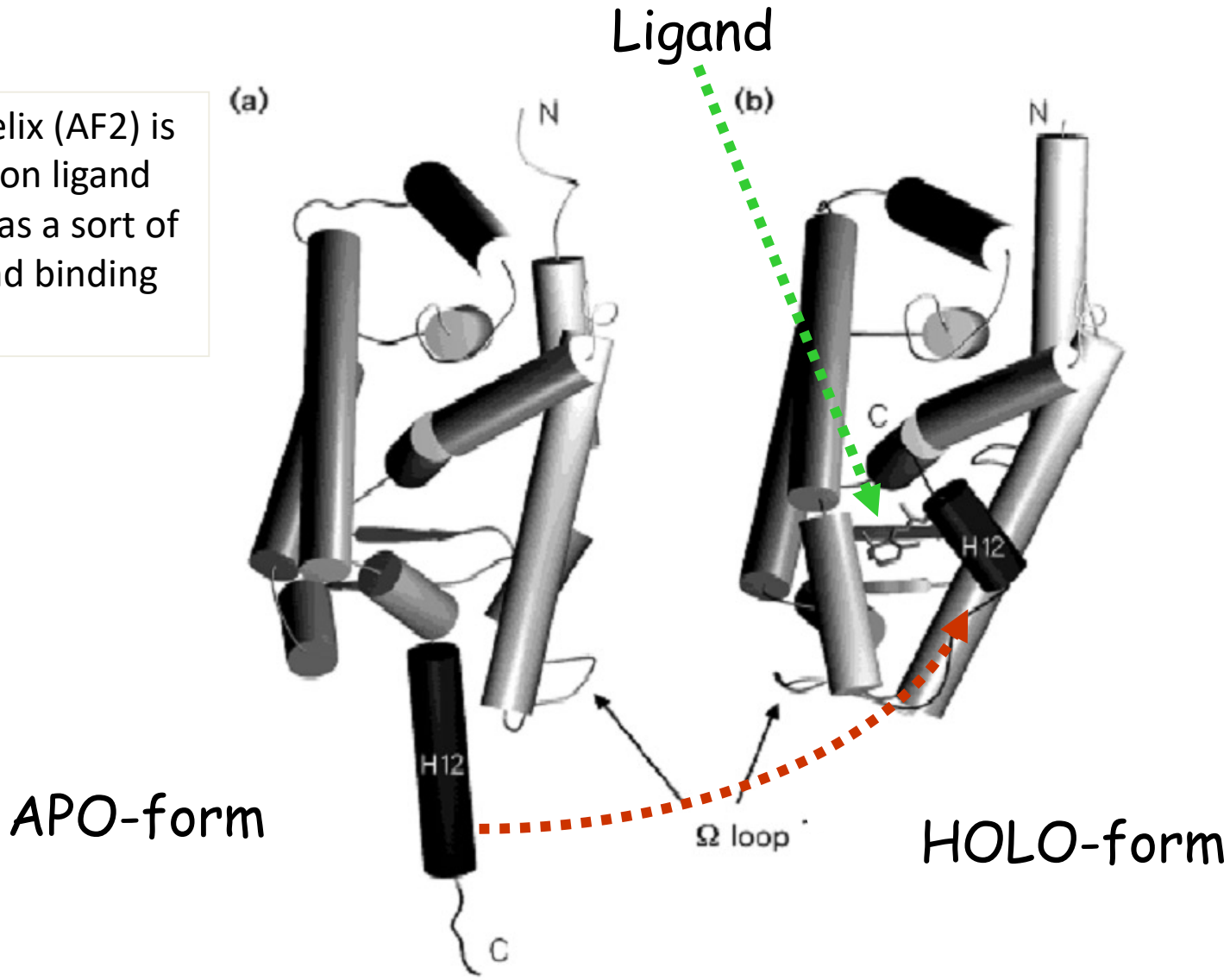
NUCLEAR RECEPTOR TYPE	NUCLEAR RECEPTOR MEMBERS
I (classical or steroid receptors)	Progestins receptor (PR) Estrogens receptor (ERα, ERβ) Androgens receptor (AR) Glucocorticoids receptor (GR) Mineralcorticoids receptor (MR)
II (RXR-heterodimeric receptors)	Thyroid hormone receptor (TRα, TRβ) All-*trans* retinoic acid receptor (RAR) 9-*cis* retinoic acid receptor (RXR) Vitamin D_3 receptor (VDR) Peroxisome proliferator receptor-γ (PPAR-γ)
III (Orphan nuclear receptors)	COUP-TFs X-linked orphan receptor (DAX-1) Rev-Erb

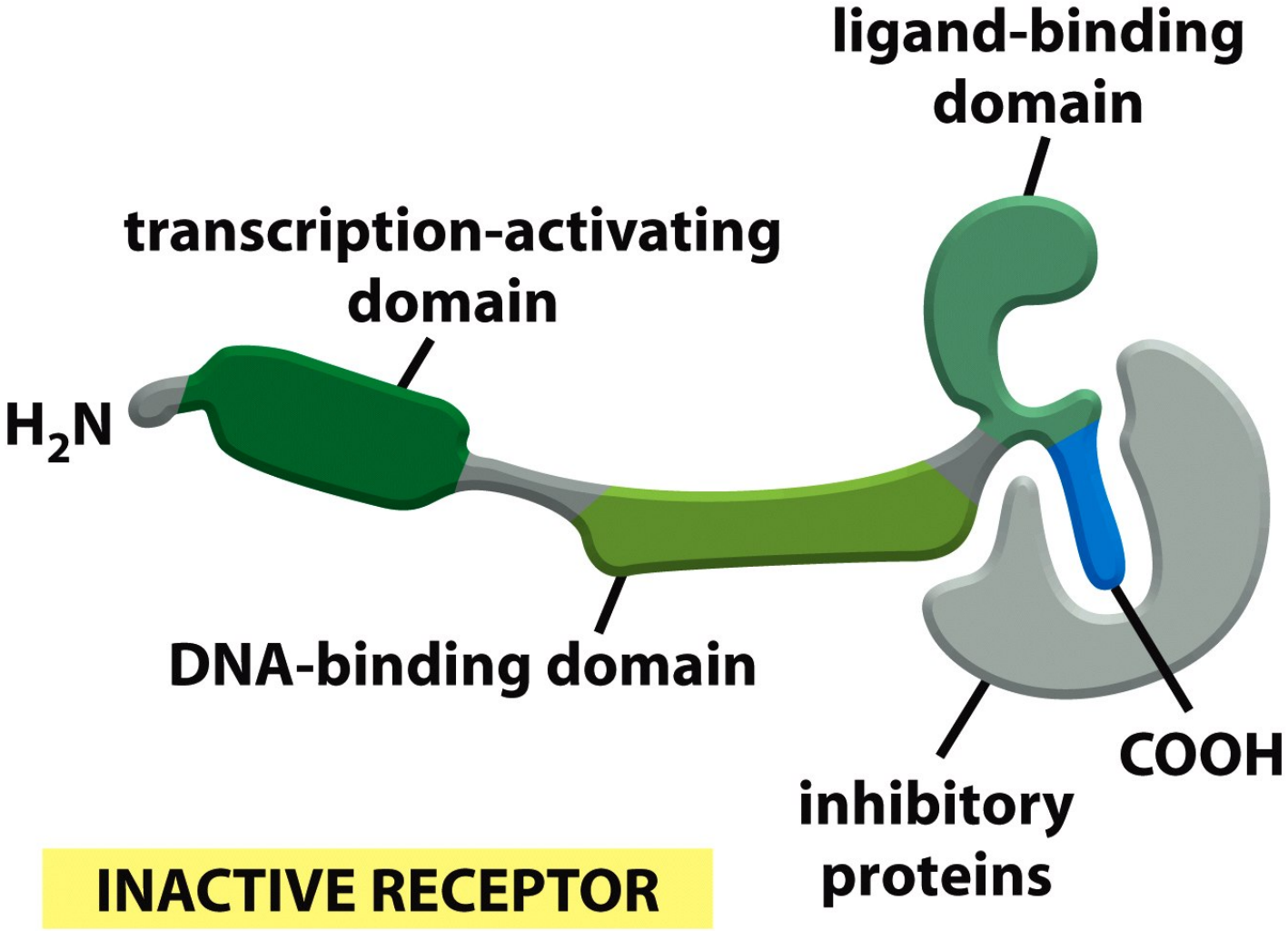
In H. sapiens there are 48 known nuclear receptor genes.

24 have known ligands 24 are orphan receptors

Ligand binding to the C-terminal domain of nuclear receptors induces a profound confo

NR C-terminal helix (AF2) is re-positioned upon ligand binding, serving as a sort of "lid" on the ligand binding pocket





hsp=heat shock proteins (chaperons)

Figure 15-14b *Molecular Biology of the Cell* (© Garland Science 2008)

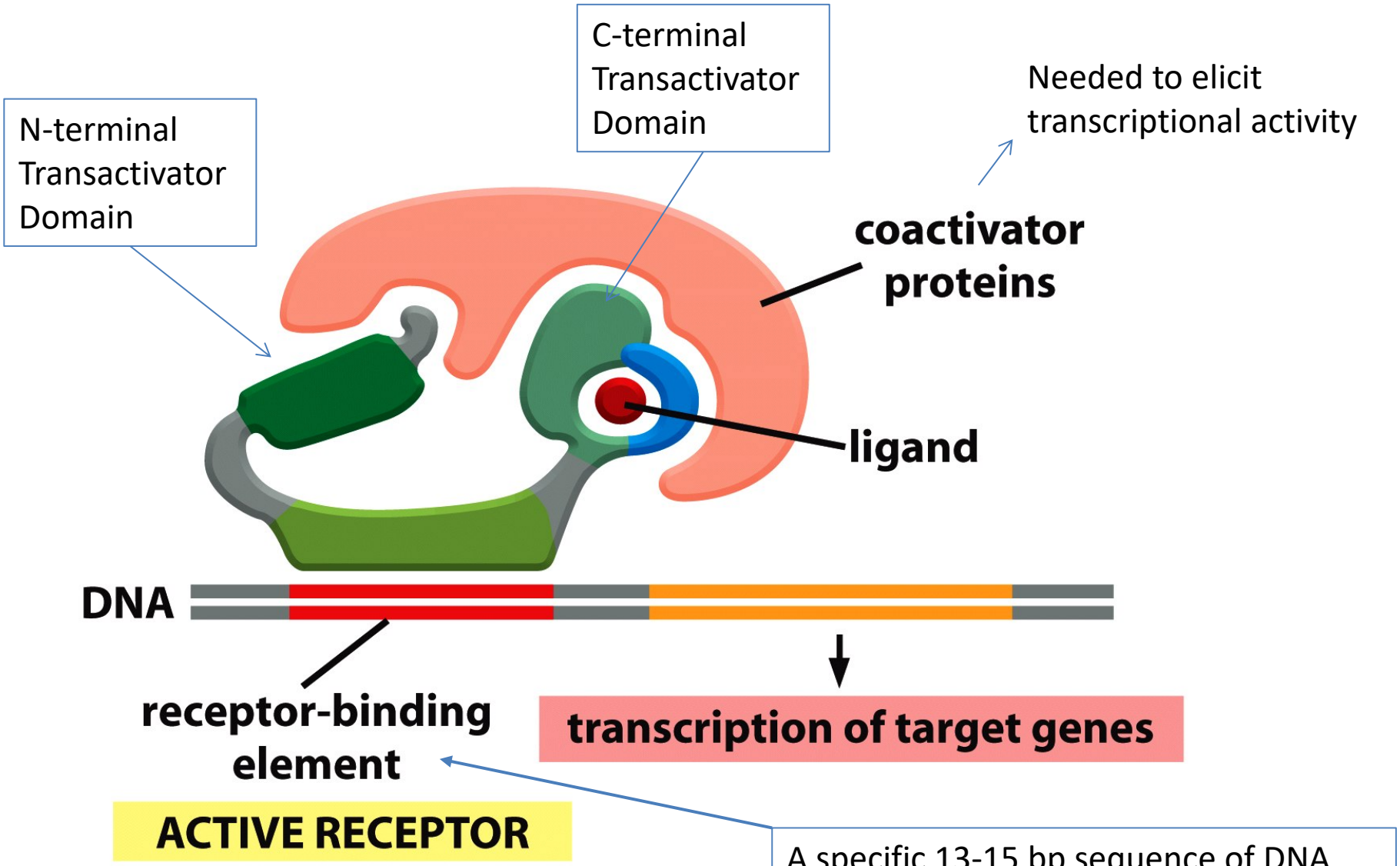
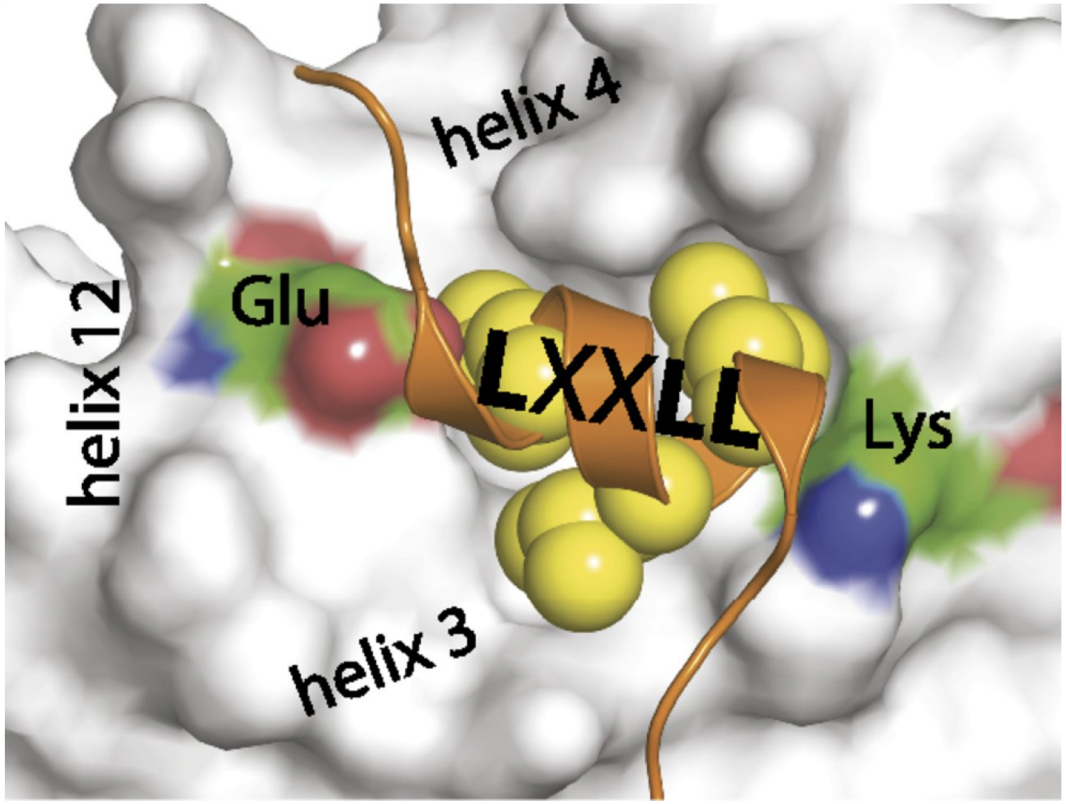


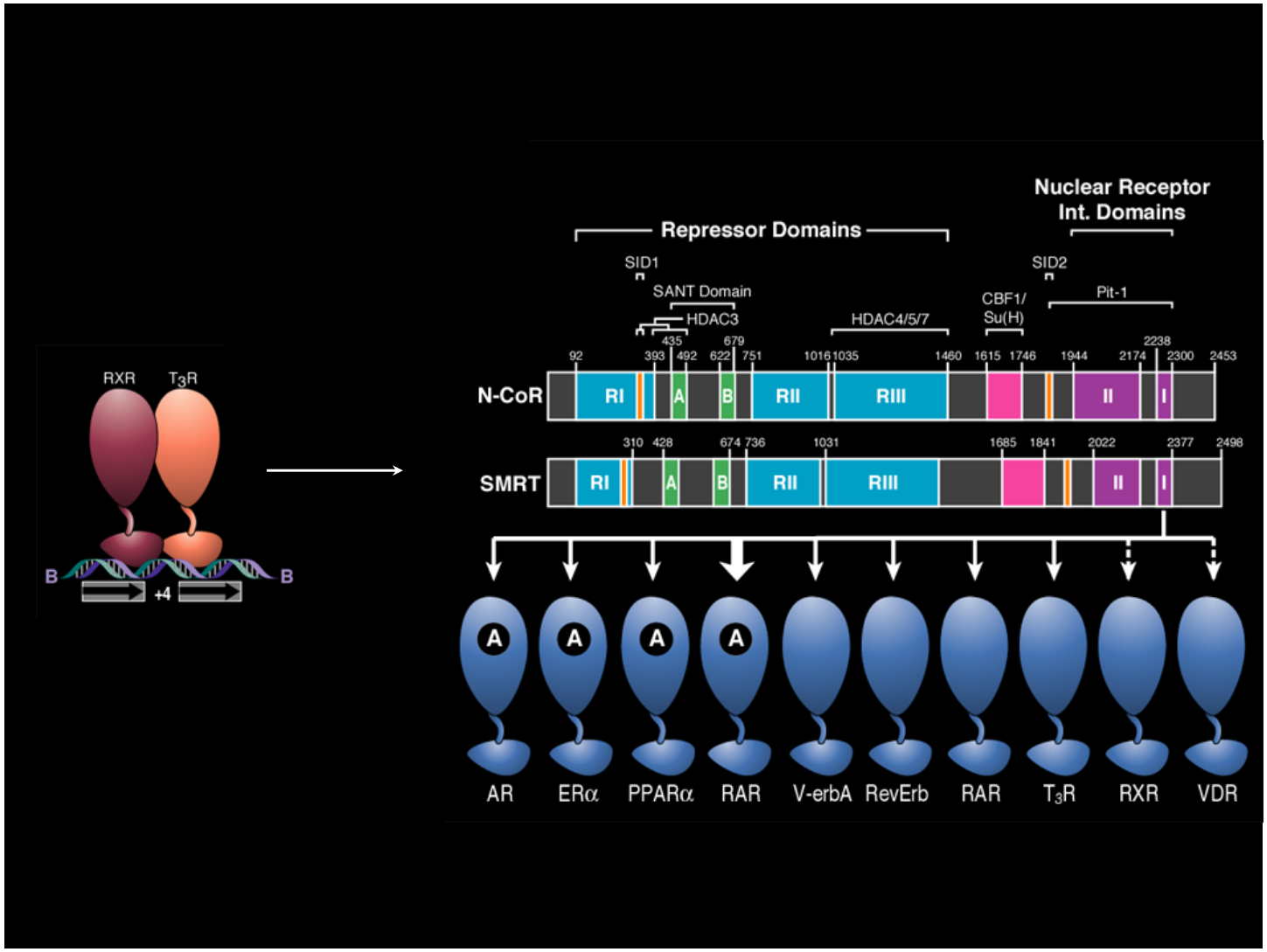
Figure 15-14c *Molecular Biology of the Cell* (© Garland Science 2008)

Interaction of co-activators (CoA) with the LBD of NRs is mediated by a common motif "**LXXLL**" present in all CoA

The LxxLL motif is flanked by charged residues that interact with opposite charges in the nuclear receptor LBD (in the active conformation) making a sort of **"charged clamp"**

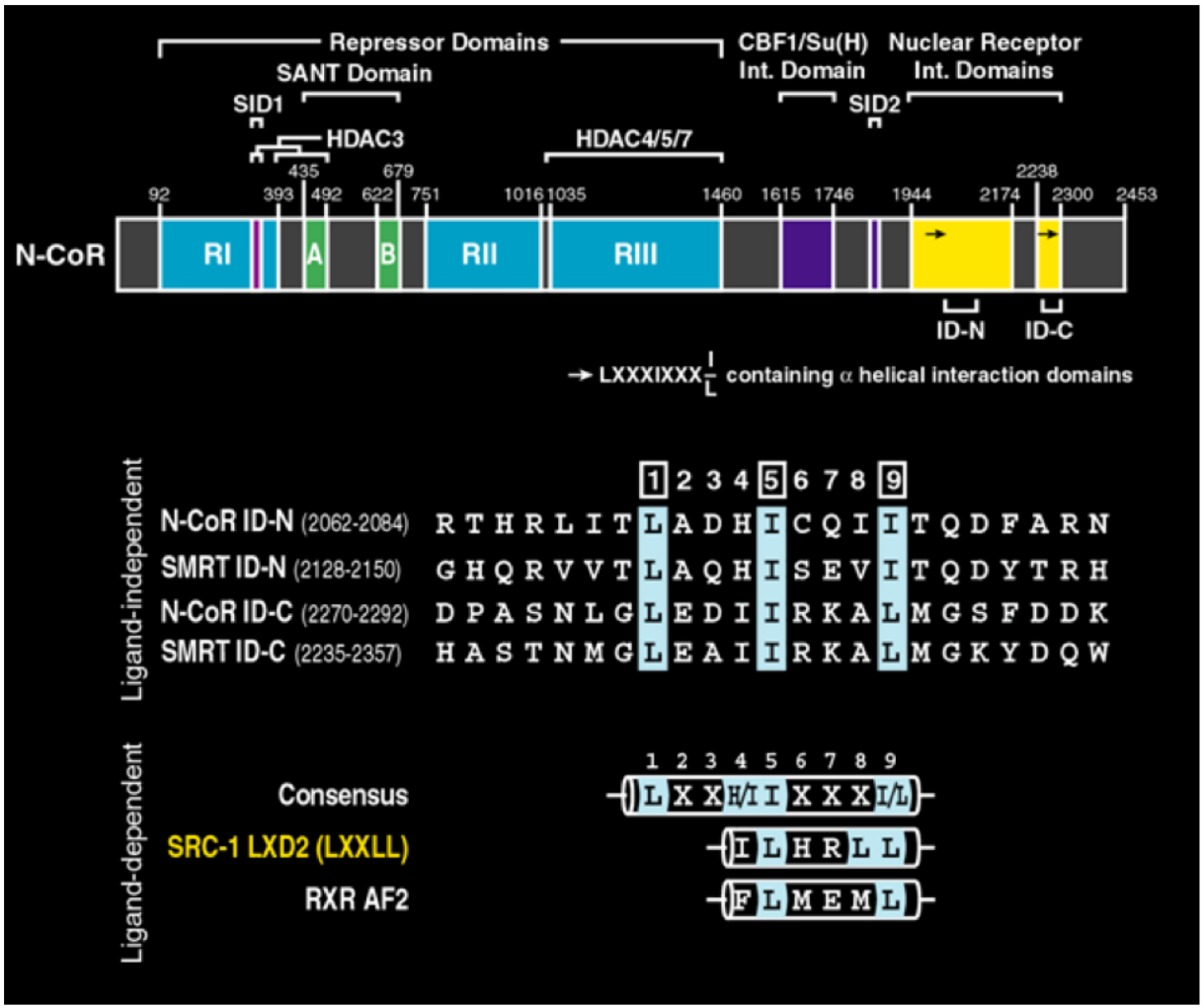


Corepressors binding to nuclear receptors



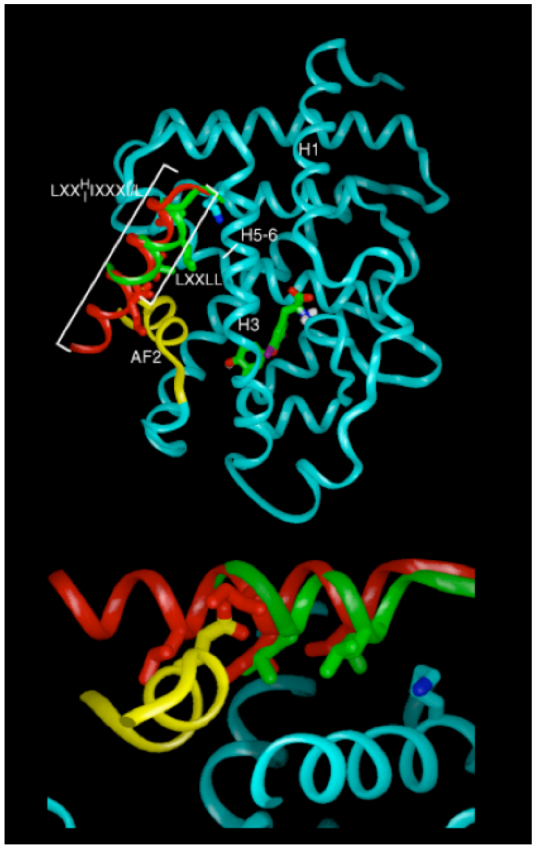
Slide adapted from MGRosenfeld

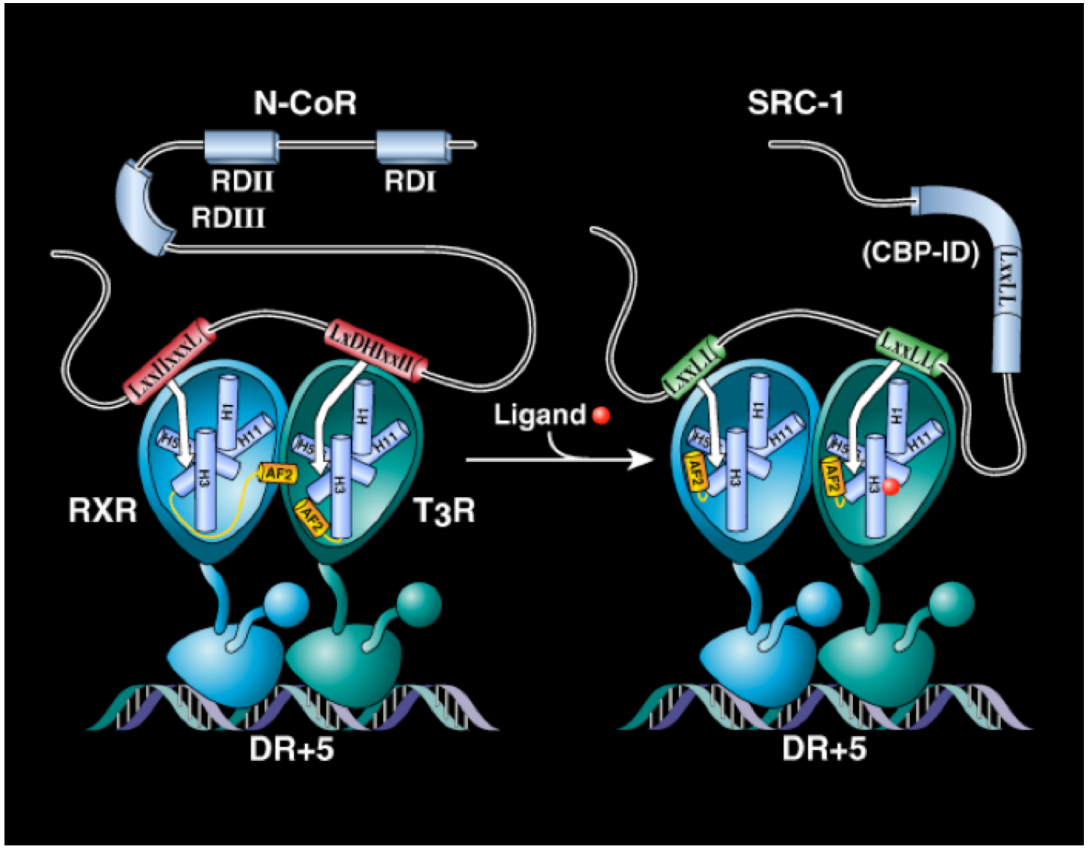
The Corepressors NCoR/SMRT contain an extended version of the coactivator LXXLL Motif



Perissi et al., G&D 1999

Schematic of Corepressor-Coactivator Helix Exchange





Perissi et al., G&D 1999

Ubiquitin-dependent exchange of corepressors for coactivators

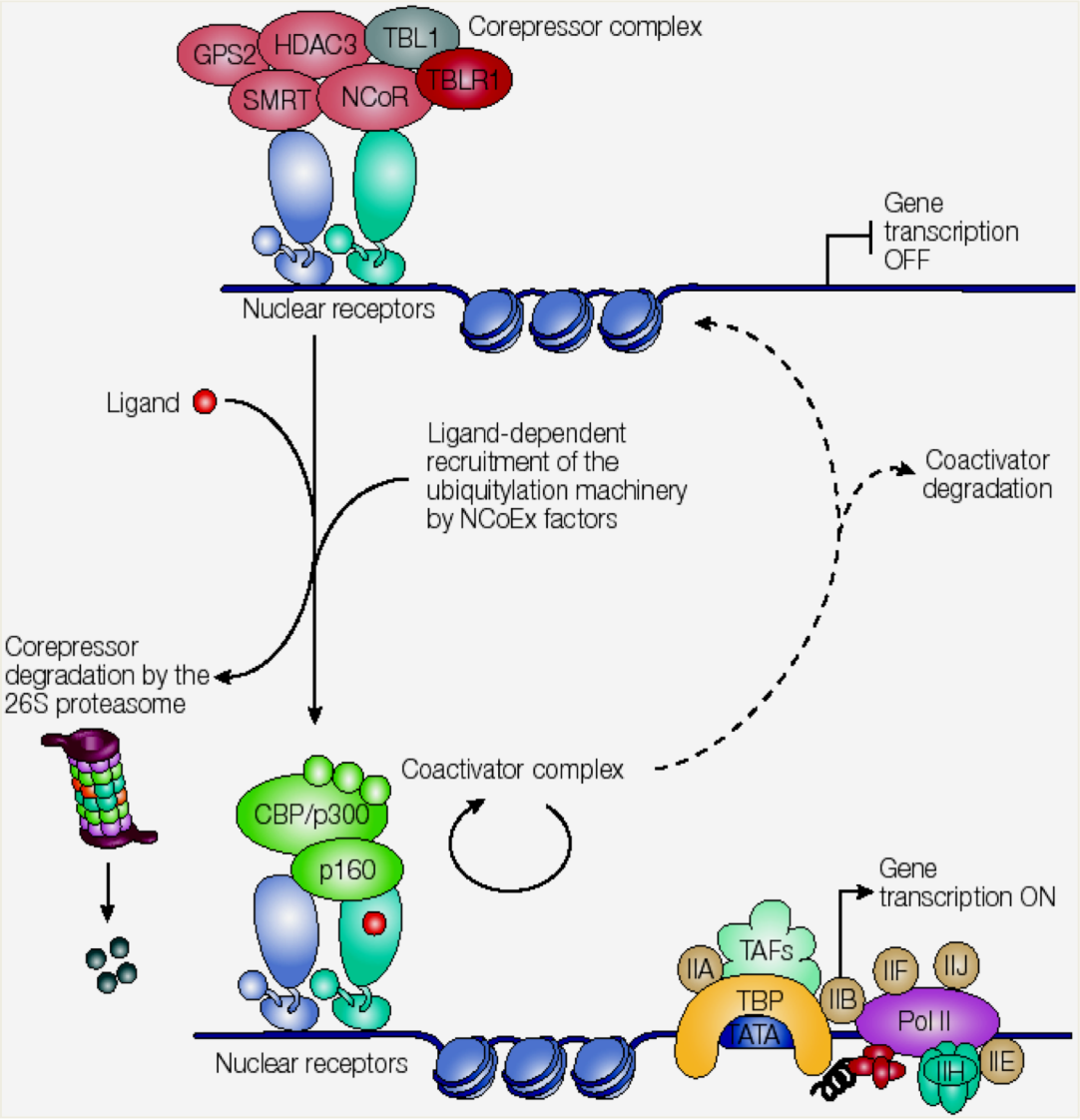
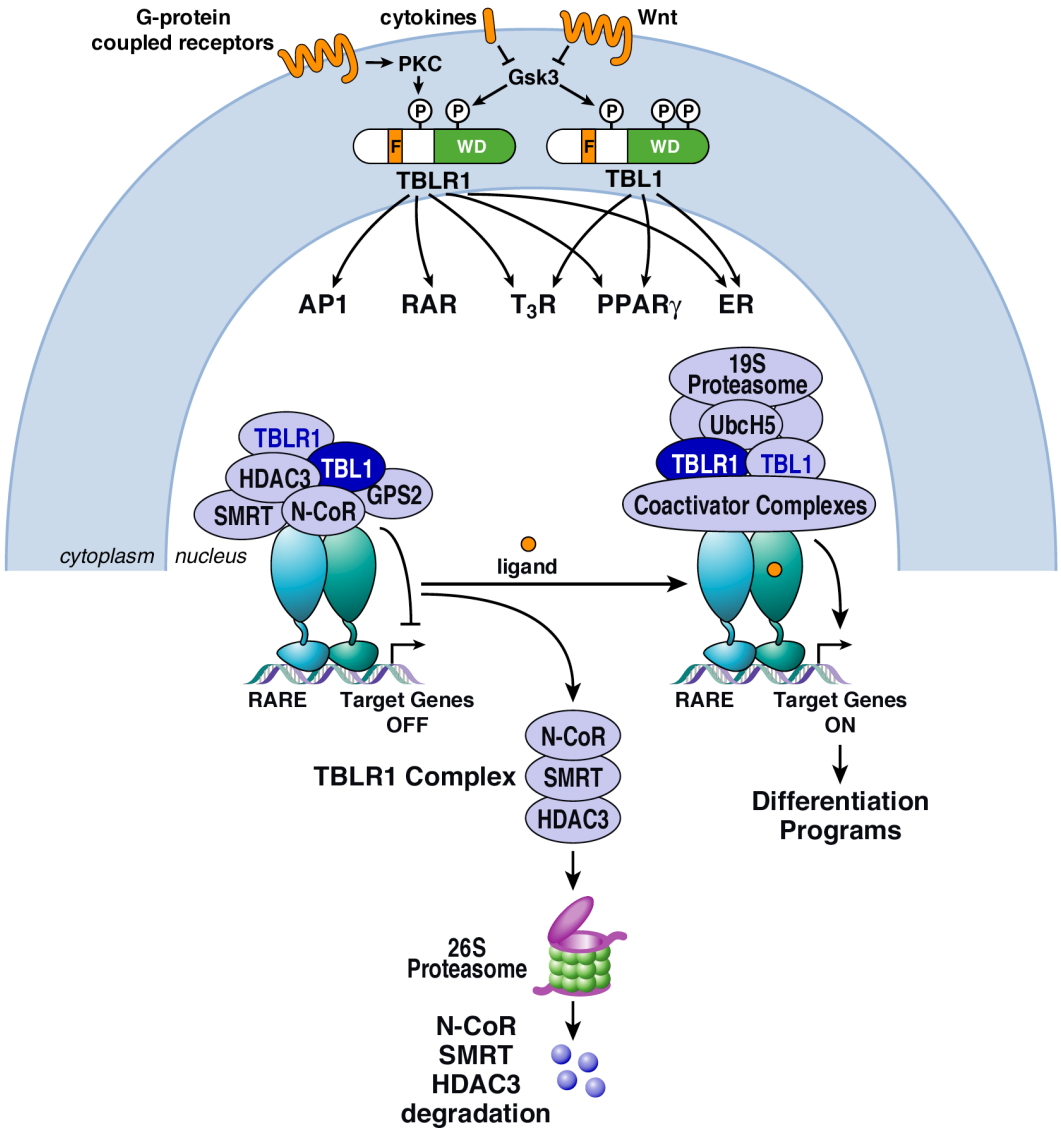


Figure 4 | Ubiquitin-dependent exchange of corepressors for coactivators. During the transition from gene repression to gene activation on ligand stimulation, there is a required exchange among cofactor complexes. This molecular switch is regulated by conformational changes in the nuclear receptor, which results in a different affinity for cofactor complexes, and by the recruitment of the ubiquitylation machinery, which has a fundamental role in the dismissal of the corepressor machinery. The diagram shows a generic, unliganded nuclear receptor heterodimer that recruits the NCoR and SMRT corepressor complexes to repress transcription. The nuclear corepressor exchange (NCoEx) factors TBL1 and TBLR1 are required to recruit the ubiquitylation machinery following ligand stimulation, thereby allowing the dismissal and degradation of the corepressor complex and the recruitment of coactivator complexes. Ubiquitin-dependent protein degradation events have also been associated with cycling of the receptor itself on the promoter and with coactivator turnover

Perissi and Rosenfeld, Nature Reviews, 2005

Regulation of the CoR/CoA exchange by signaling pathways



Perissi et al, Cell, 2004

The Corepressor-coactivator exchange model.

Many signalling pathways, which regulate «off-on» decision to developmentally important genes, similarly result in the described «**corepressor-coactivator exchange**».

Very popular examples are the **Notch** and **Wnt** pathways, but actually many other do the same.