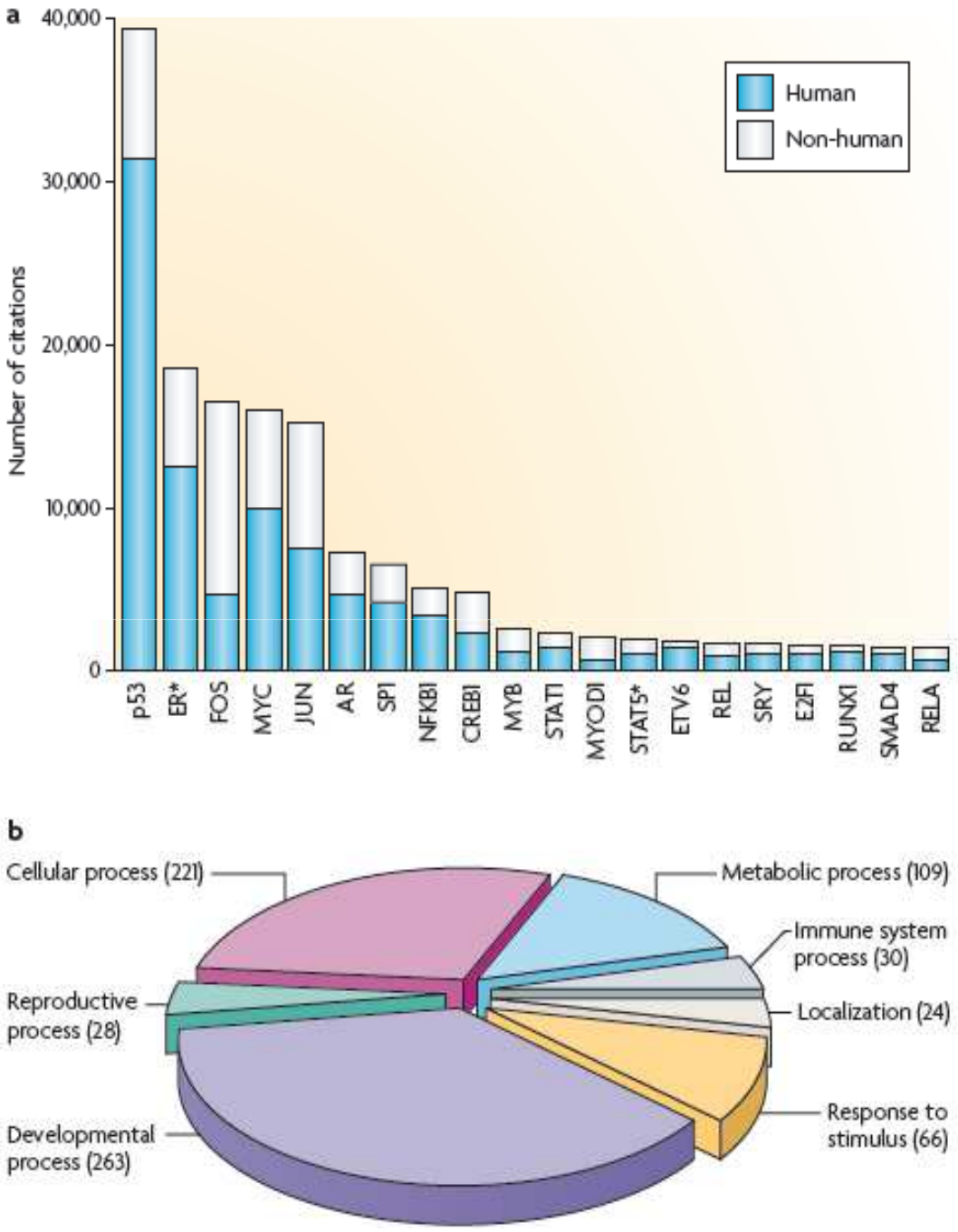
A census of human transcription factors: function, expression and evolution

Juan M. Vaquerizas[*], *Sarah K. Kummerfeld*[‡§], *Sarah A. Teichmann*[‡] *and Nicholas M. Luscombe*[*||]

Abstract | Transcription factors are key cellular components that control gene expression: their activities determine how cells function and respond to the environment. Currently, there is great interest in research into human transcriptional regulation. However, surprisingly little is known about these regulators themselves. For example, how many transcription factors does the human genome contain? How are they expressed in different tissues? Are they evolutionarily conserved? Here, we present an analysis of 1,391 manually curated sequence-specific DNA-binding transcription factors, their functions, genomic organization and evolutionary conservation. Much remains to be explored, but this study provides a solid foundation for future investigations to elucidate regulatory mechanisms underlying diverse mammalian biological processes.

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.



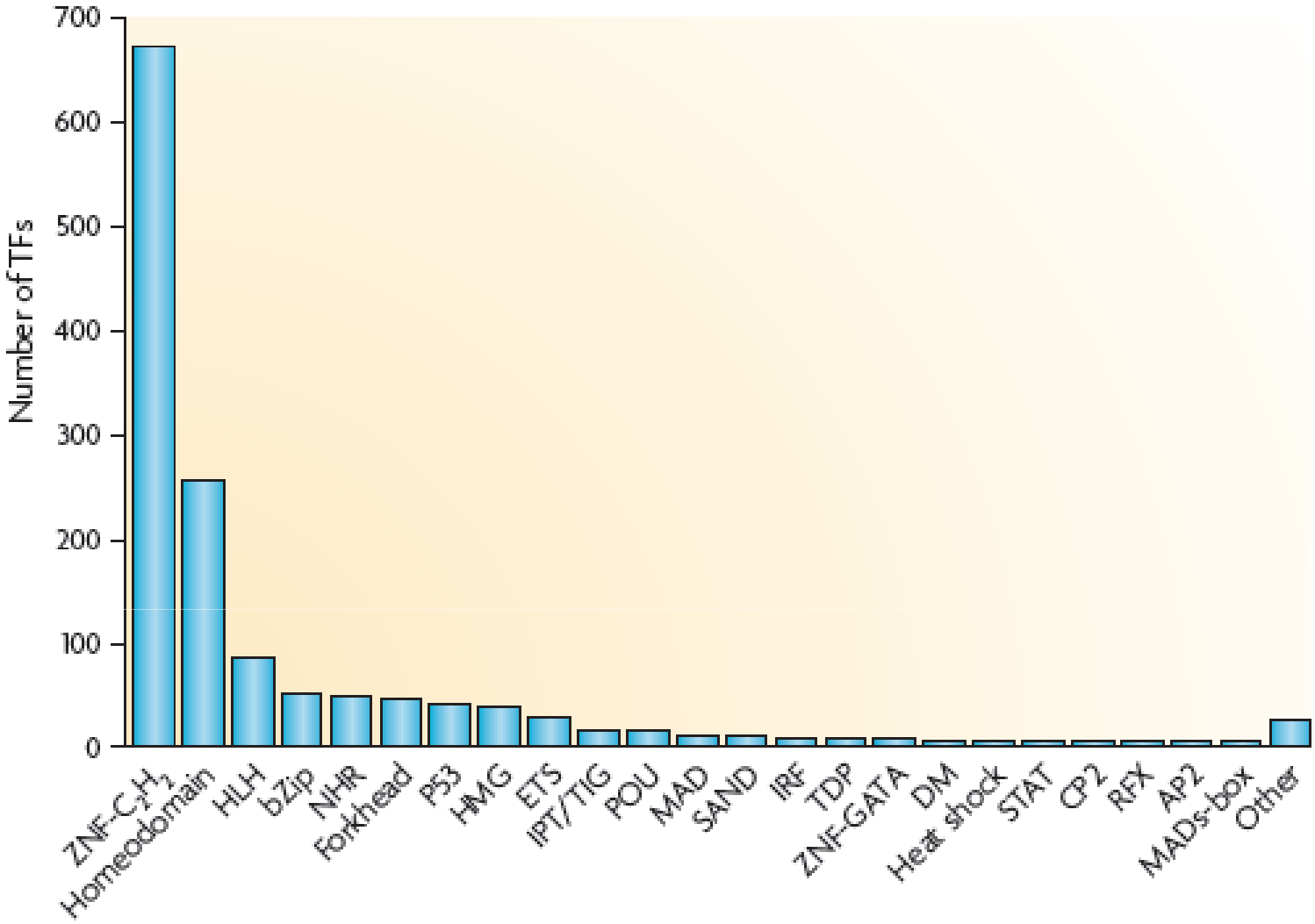


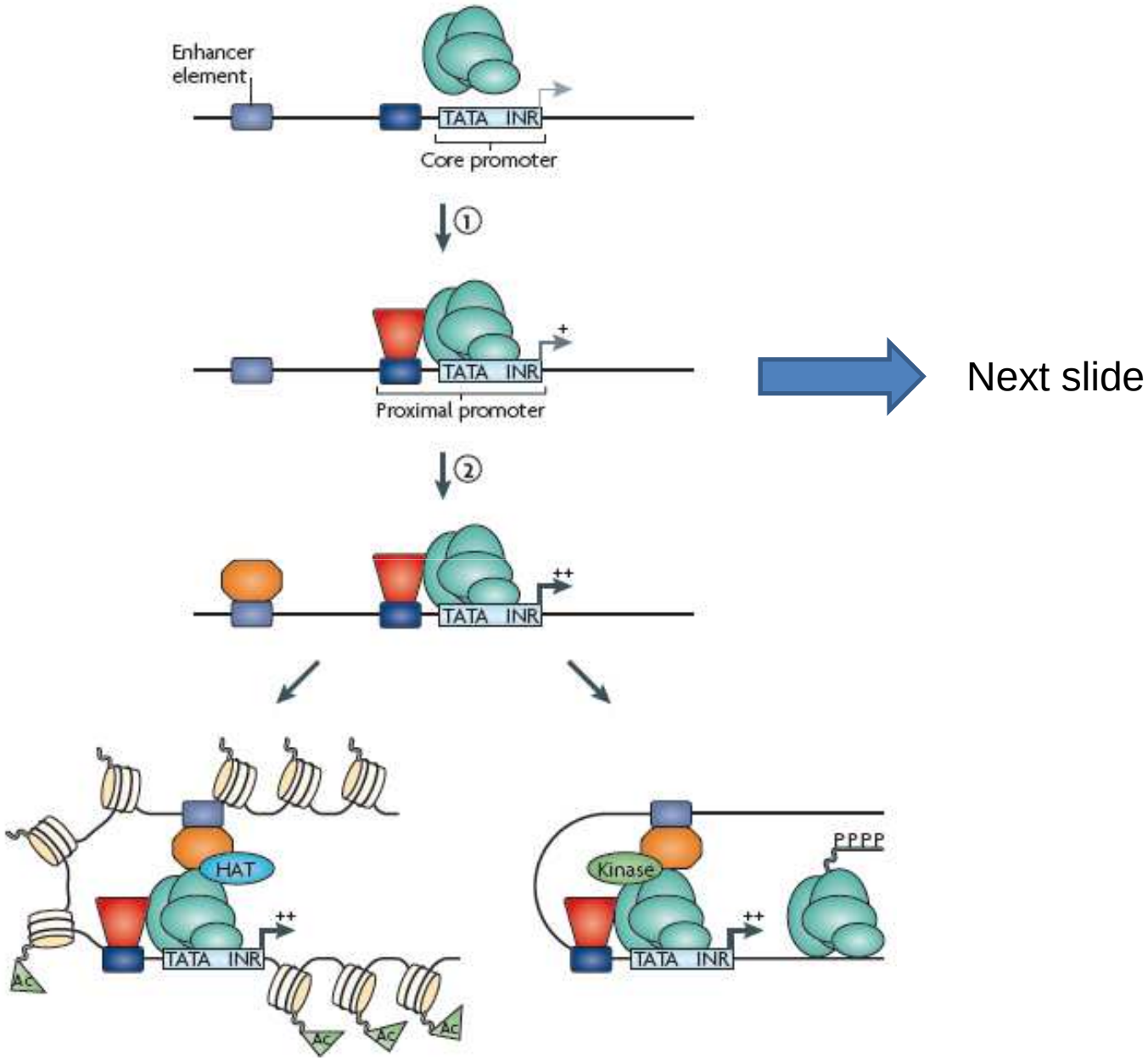
Figure 2 | **Transcription factors classified by DNA-binding domain.** Transcription factors (TFs) were classified into families according to their DNA-binding domain composition. InterPro parent–child relationships between DNA-binding domains were used as the basis for TF family definition (Supplementary 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'.

Insights from genomic profiling of transcription factors

Peggy J. Farnham

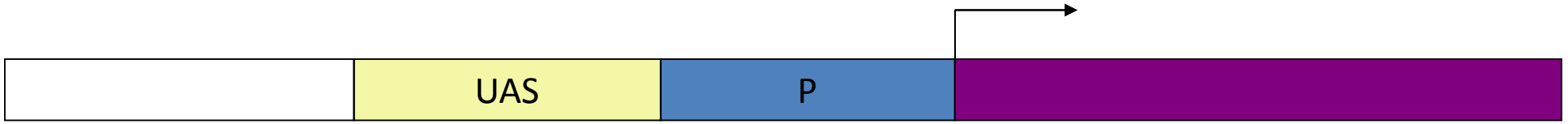
Abstract | A crucial question in the field of gene regulation is whether the location at which a transcription factor binds influences its effectiveness or the mechanism by which it regulates transcription. Comprehensive transcription factor binding maps are needed to address these issues, and genome-wide 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.

REVIEW


Enhancer element
TATA INR
Core promoter
1
+
TATA INR
Proximal promoter
2
++
TATA INR
HAT
++
TATA INR
Ac
Ac
Ac
Ac
Kinase
PPPP
++
TATA INR

Next slide

la situazione più semplice è:



sebbene vi siano diversi tipi di promotore (es. TATA-dep, Inr-dep, CpG, etc)
il promotore minimo è:

1) estremamente inefficiente *in vivo*
2) non regolato

ed è prevalentemente un **elemento di posizione**

Is the interaction of a DNA-bound regulatory TF with the basal PIC (general TFII + PolII)

DIRECT ?

Older experiments were made by reconstructing “in vitro” a minimal transcription system.
This was made by constructing a reporter gene driven by a basal promoter combined with variable UAS elements, binding specific transcription factors
The performance of these regulatory elements were then tested in experiments of “in vitro transcription”, using purified, recombinant proteins.

To understand these approaches, first think about what you should know about basal transcription.

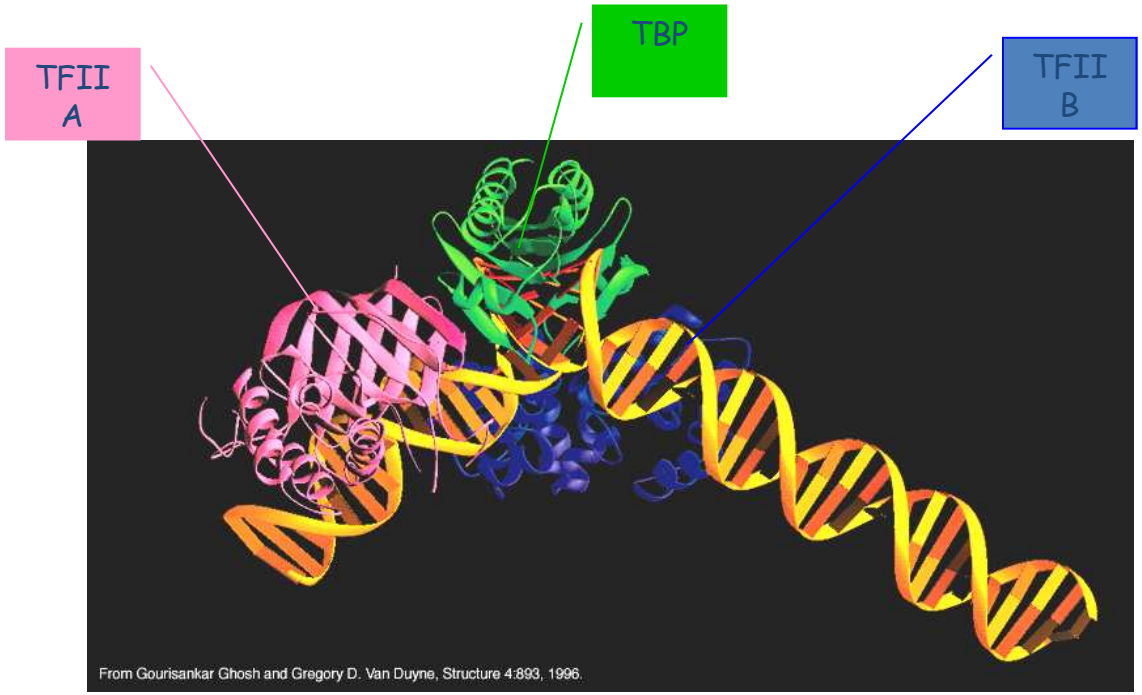
The general factor TFIID has the primary role of **promoter recognition**: it is composed of **TBP** (TATA-binding protein) and 9-14 TAFs.

TAFs (TBP-associated proteins) have different roles in PIC assembly:

-possess histone acetylase activity (HAT)

-make contacts with DNA and recognize Inr, DPE

-mediate contacts with transcription factors ?=



Human	Drosophila	Yeast	Essential in Yeast	Features	New nomen clature
TAF250	TAF250/230	TAF145/130	Yes	HAT, G1/S arrest	TAF1
TAF150	TAF150	TAF150	Yes	G2/M arrest	TAF2
TAF140		TAF47	Yes		TAF3
TAF130/135	TAF110	TAF48	Yes		TAF4
TAF100	TAF80/85	TAF90	Yes	G2/M arrest	TAF5
TAF80	TAF60/62	TAF60	Yes	H4 like	TAF6
TAF55		TAF67	Yes		TAF7
(bab71460)		TAF65	Yes		TAF8
TAF32/31	TAF40/42	TAF17	Yes	H3 like	TAF9
TAF30		TAF25	Yes	G1/S arrest	TAF10
TAF28	TAF30beta	TAF40	Yes	H3 like	TAF11
TAF20/15	TAF22, TAF28/30 alfa	TAF61/68	Yes	H2B like	TAF12
TAF18		TAF19	Yes	H4 like	TAF13
		TAF30	no		
TAF105				B-cell specific	

TFIID = TBP + 9 to 14 TAF

TBP= TATA binding protein
TAF = TBP associated factors

TFIID factor has since long time been appreciated as an essential component to “recognize” the transcriptional initiation site.

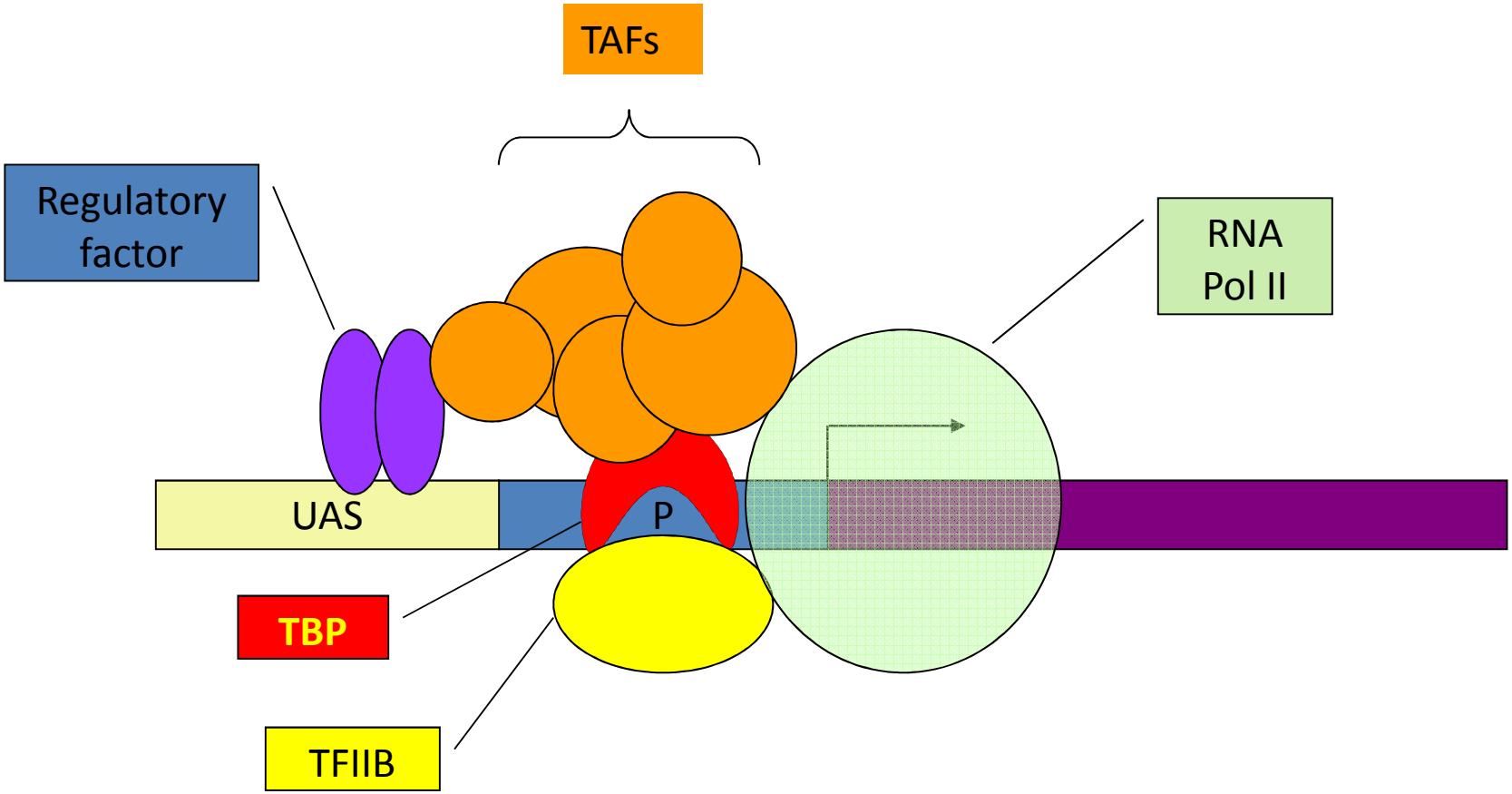
It is composed of TBP + variable number of subunits called TAFs (TBP associated factors)

← Comparison of yeast and higher eukaryotic TAFIIs.

Homologues of TAFIIs are arranged in rows, human TAFIIs (column 1), DrosophilaTAFIIs (column 2), yeast TAFIIs (column 3).

TAFIIs required for viability in yeast are indicated as essential.

(HAT, histone acetyltransferase).



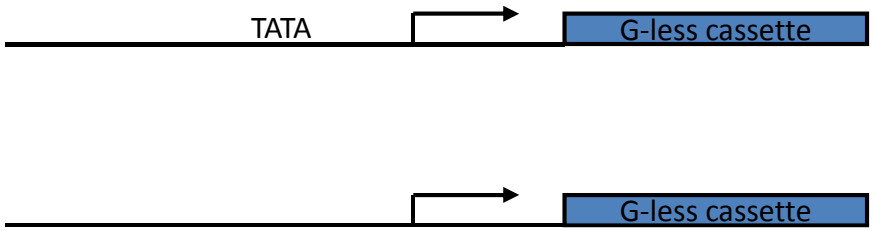
Regulatory Transcription Factors bound to proximal elements will contact components of the PIC, primarily TAF proteins. This was shown experimentally by reconstitution experiments.

Although TFIID purifies as a multimeric protein, starting from recombinant TAFs and recombinant TBP expressed in E. coli, R. Tjian's group has shown that **partial complexes** can be assembled and tested in "*in vitro* transcription" experiments.

These experiments showed that transcription from TATA-box containing promoters can be correctly initiated using only TBP, TFIIB and RNA Pol II.

On the contrary, TATA-less promoters require TFIID (i.e. the TAFs) for initiation. In particular, a sub-complex containing TAFII-250 and TAFII-150 (drosophila) was enough to initiate transcription from a INR-containing promoter and TAFII40+TAFII60 was enough for a DPA-containing promoter.

in vitro transcription experiment:



recombinant proteins added to the "in vitro" transcription mixture

TBP TFIIB PolII	**TFIID(1)** TFIIB PolII	TBP TFIIB PolII **TAFs** (partial)
+	+	+
-	+	+/- (*)

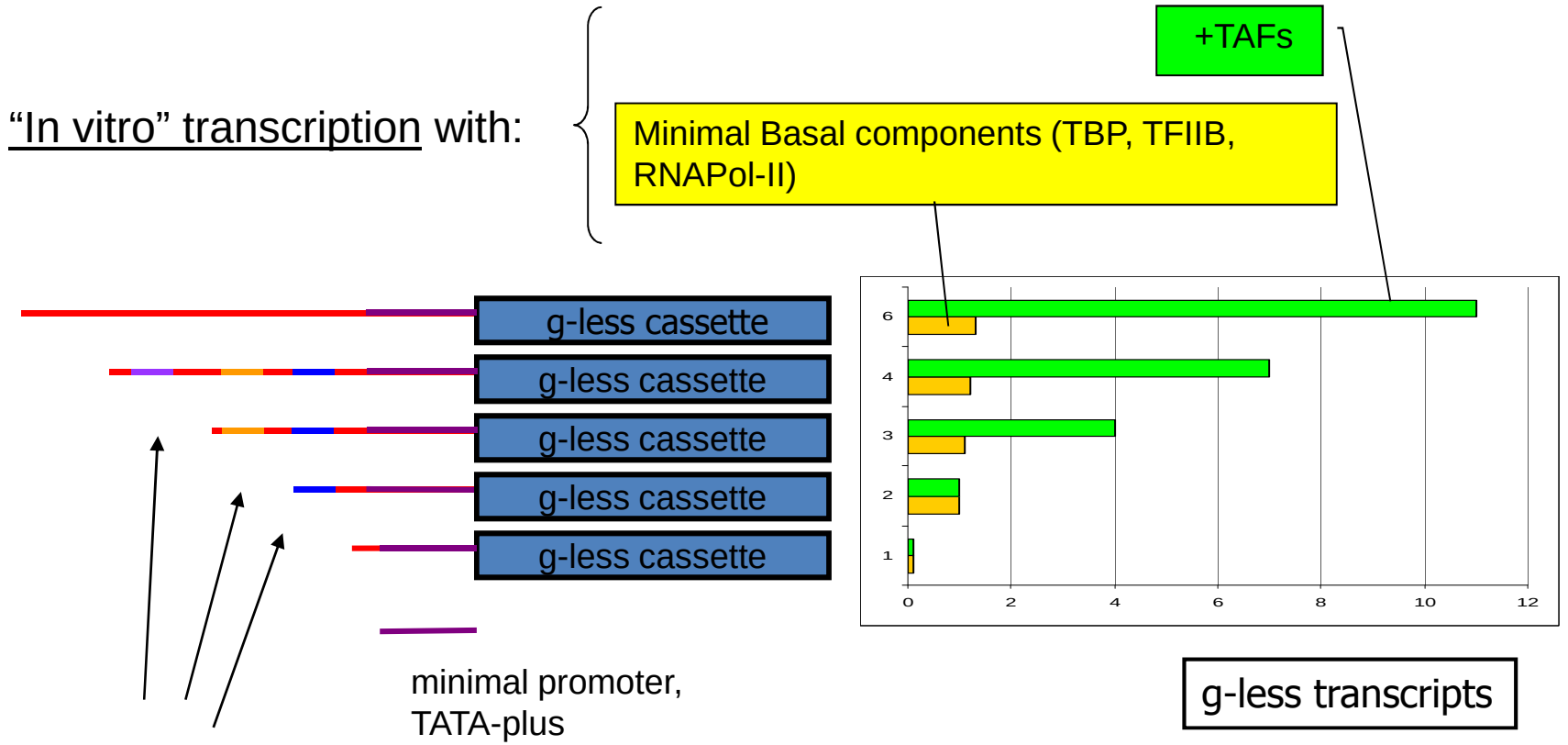
(*) depending on TAFs added

(1) the entire protein, composed of TBP-TAFs

These experiments demonstrated that reconstituted transcription “in vitro” is possible, starting from few recombinant proteins.
Note that TATA-plus promoters do not require TAFs for **basal** activity

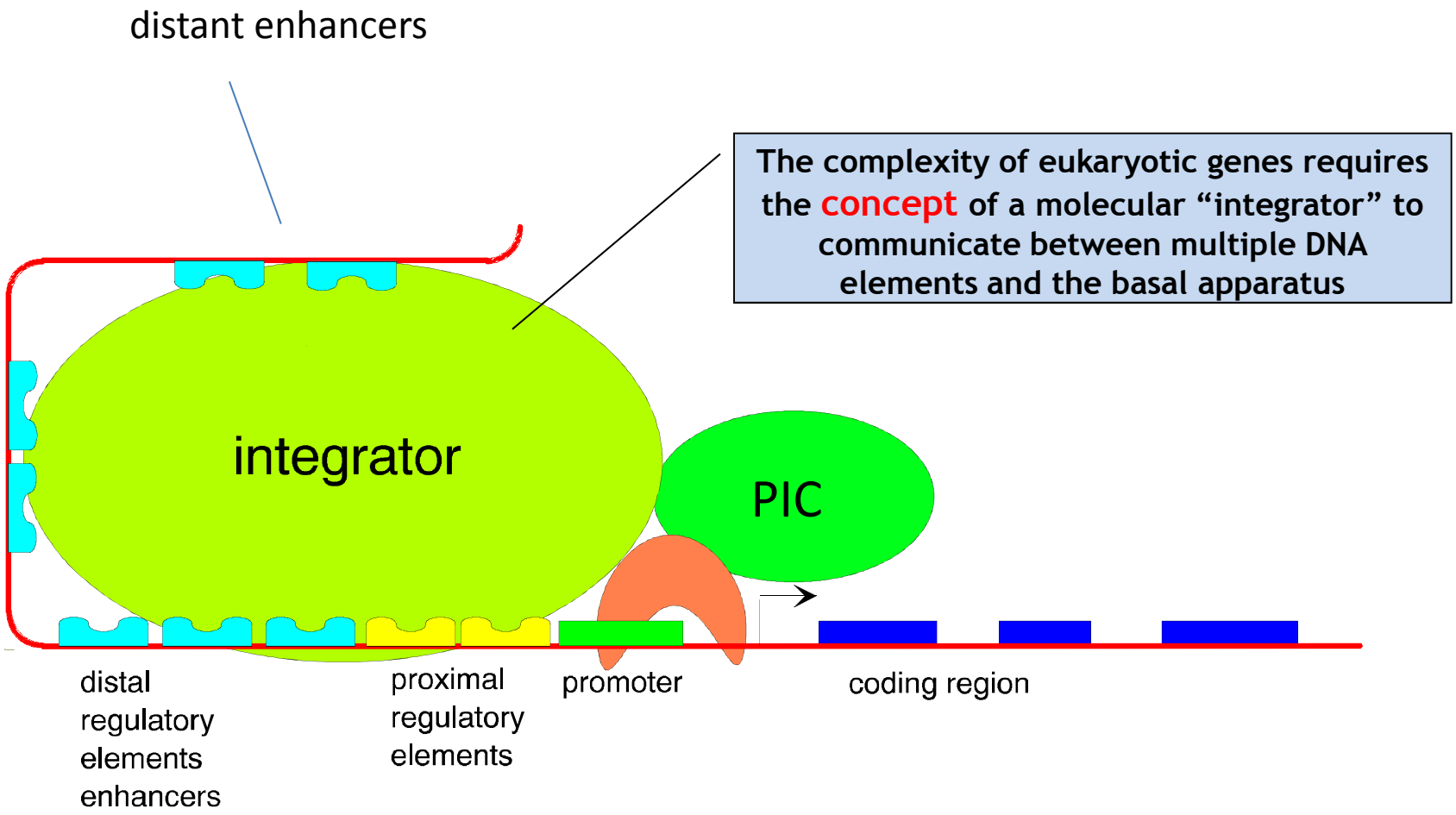
Now, the same experiment was done by adding 5’ UAS (sequence elements recognizing specific regulatory Transcription Factors) and adding to the “in vitro” transcription mixture the appropriate recombinant TF.

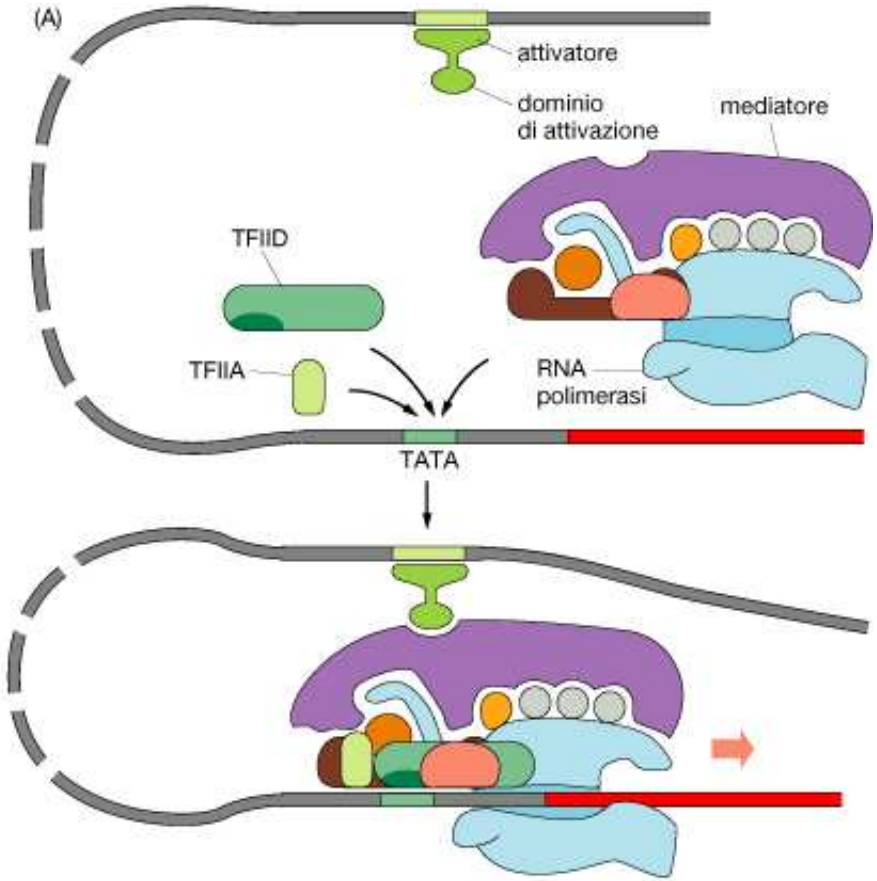
See the next scheme:



“Trans-activation” is seen only when specific TAFs are added to the mixture, allowing conclusion the direct contact between regulatory (activating) TF and basal PIC requires TAFs, i.e. transcription factors interact with TAFs.

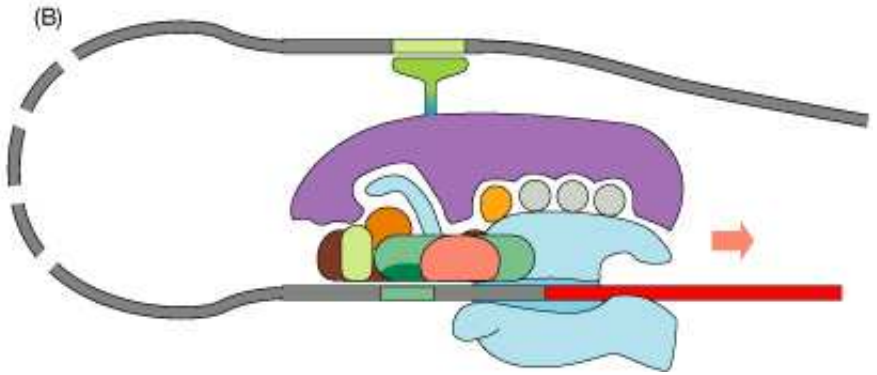
Problem: the effect of multiple elements and multiple TFs is “compositional”


distant enhancers
The complexity of eukaryotic genes requires the concept of a molecular "integrator" to communicate between multiple DNA elements and the basal apparatus
integrator
PIC
distal regulatory elements enhancers
proximal regulatory elements
promoter
coding region



The Mediator, a megadalton complex interacting with TF and PIC components, was isolated ancd characterized in 2004-05.

There are several version of Mediator in the cell nucleus

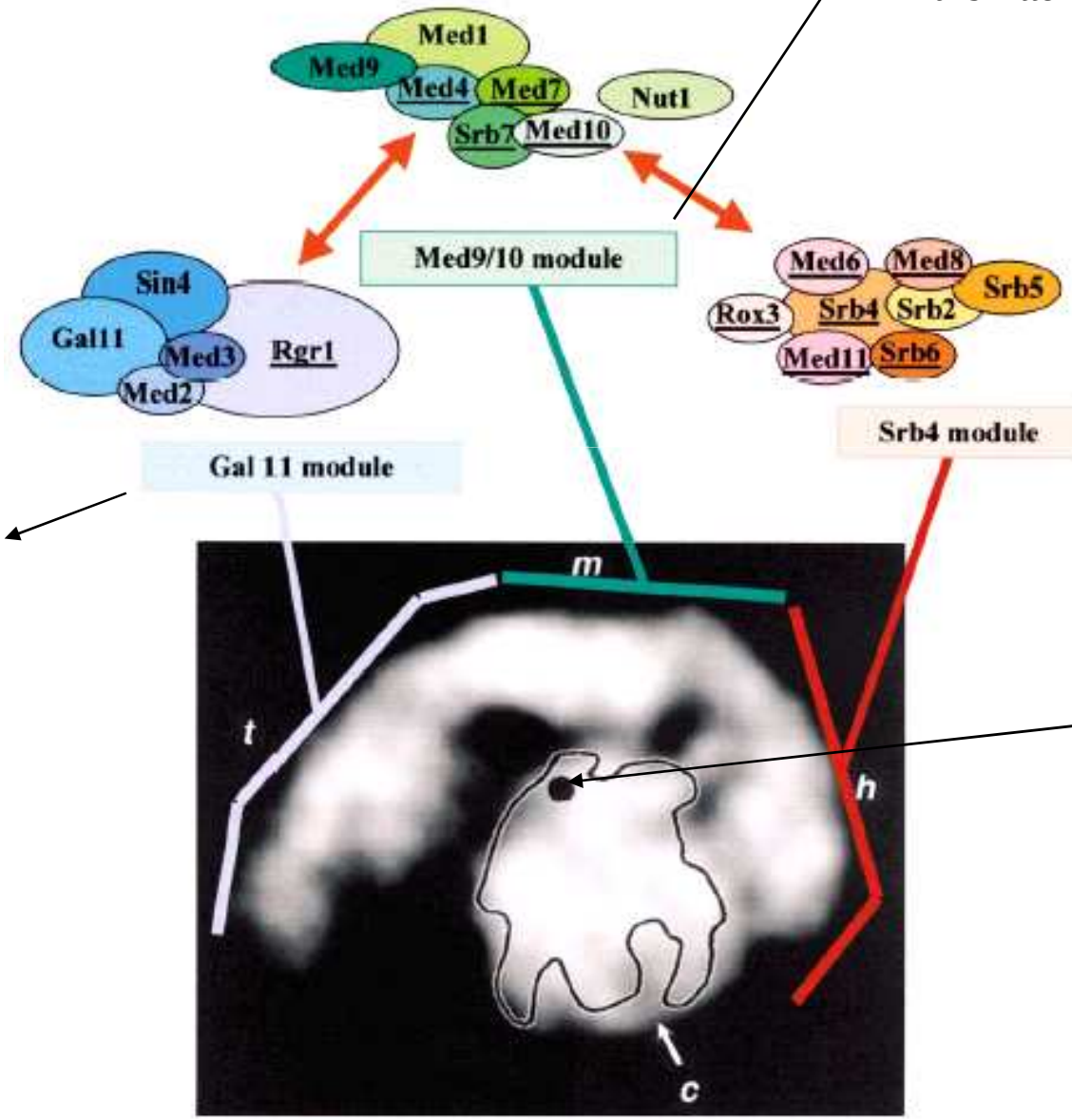


Il complesso del "Mediatore"

Subunit compositions of mediator complexes.

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		

Interactcs with CTD
Both activator and repressor signal
Transmitter to Pol



Role in contacting
trans-activators

CTD

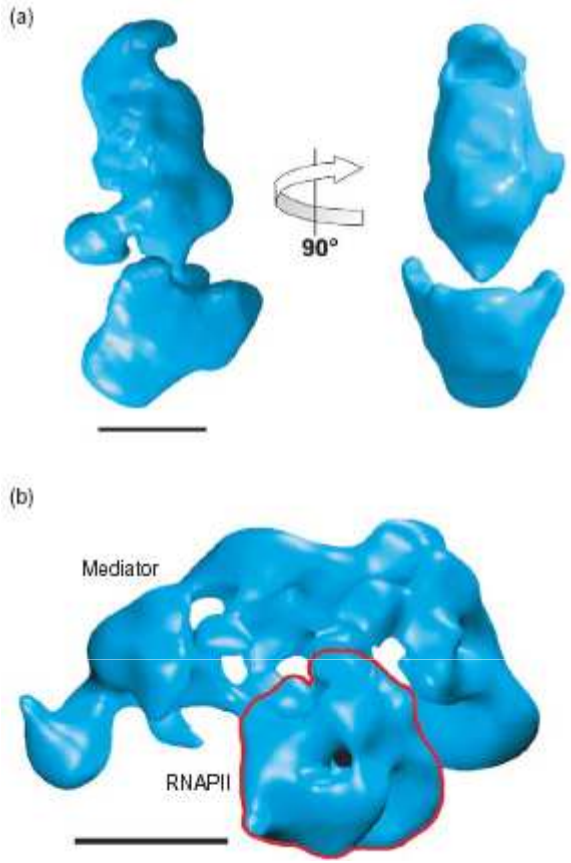


Figure 1. Structure of the yeast Mediator and holoenzyme complexes. (a) A 3D reconstruction of the yeast Mediator structure was calculated from images of individual particles imaged in an electron microscope after preservation in stain. Mediator has a compact, roughly triangular shape. A large domain at the bottom is linked by a thin connection to the top portion of the structure. The resolution of the reconstruction is 35- A° , and the scale bar represents 100 A° . (b) Structure of the Mediator–RNA polymerase II holoenzyme complex calculated from electron microscope images of individual particles preserved in stain. Previous characterization of the polymerase and Mediator structures led to identification of the Mediator and RNA polymerase II (red outline) portions of the holoenzyme structure. In the holoenzyme, Mediator adopts an extended conformation, embracing the central polymerase density. The resolution of the reconstruction is w35 A° , and the scale bar represents 100 A° . Part (b) reproduced, with ermission, from Ref. [19].

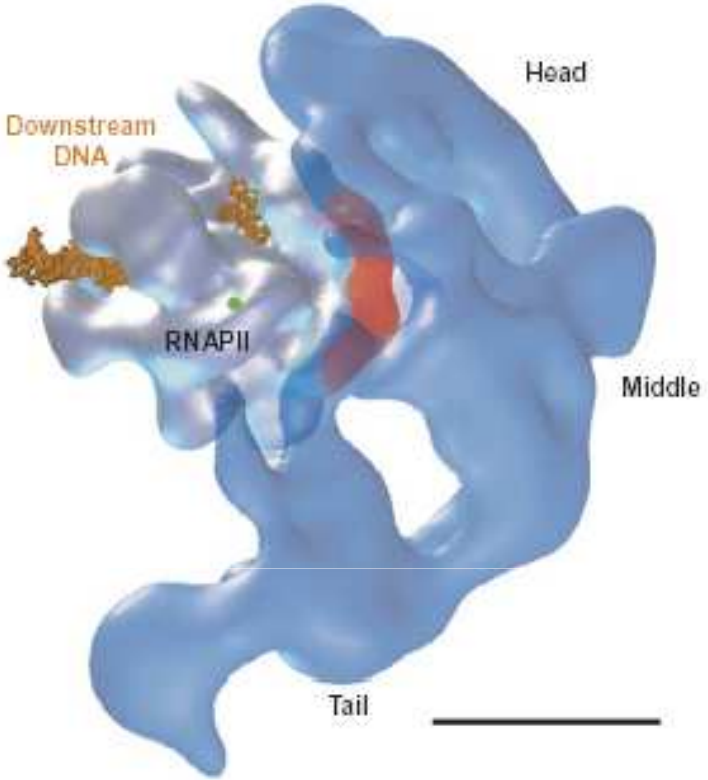


Figure 3. Interaction of Mediator and RNA polymerase II (RNAPII) in the holoenzyme complex. The precise orientation of RNAPII in the holoenzyme complex was
established by 2D cross-correlation analysis between holoenzyme and RNAPII projections. The figure shows a cryoelectron microscopy reconstruction of polymerase fitted into the extended Mediator structure in the orientation
determined by cross-correlation analysis. Multiple contacts between Mediator and RNAPII are established in the holoenzyme complex, involving mostly the head and middle domains, and distributed around the Rpb3–Rpb11 polymerase subunits (highlighted in red). The small green circle indicates the point where the carboxyterminal
domain of Rpb1 (the largest polymerase subunit), crucial for Mediator polymerase interaction, emanates from the surface of the enzyme. The bacterial homolog of the Rpb3–Rpb11 complex, the a2 homodimer, is involved in transcription regulation in bacteria, suggesting a conservation between prokaryotes and eukaryotes of the RNA polymerase surface involved in regulation. The scale bar represents 100 A° .

Different forms of Mediator exist and Mediator conformation is dependent on the kind of **activator** is bound

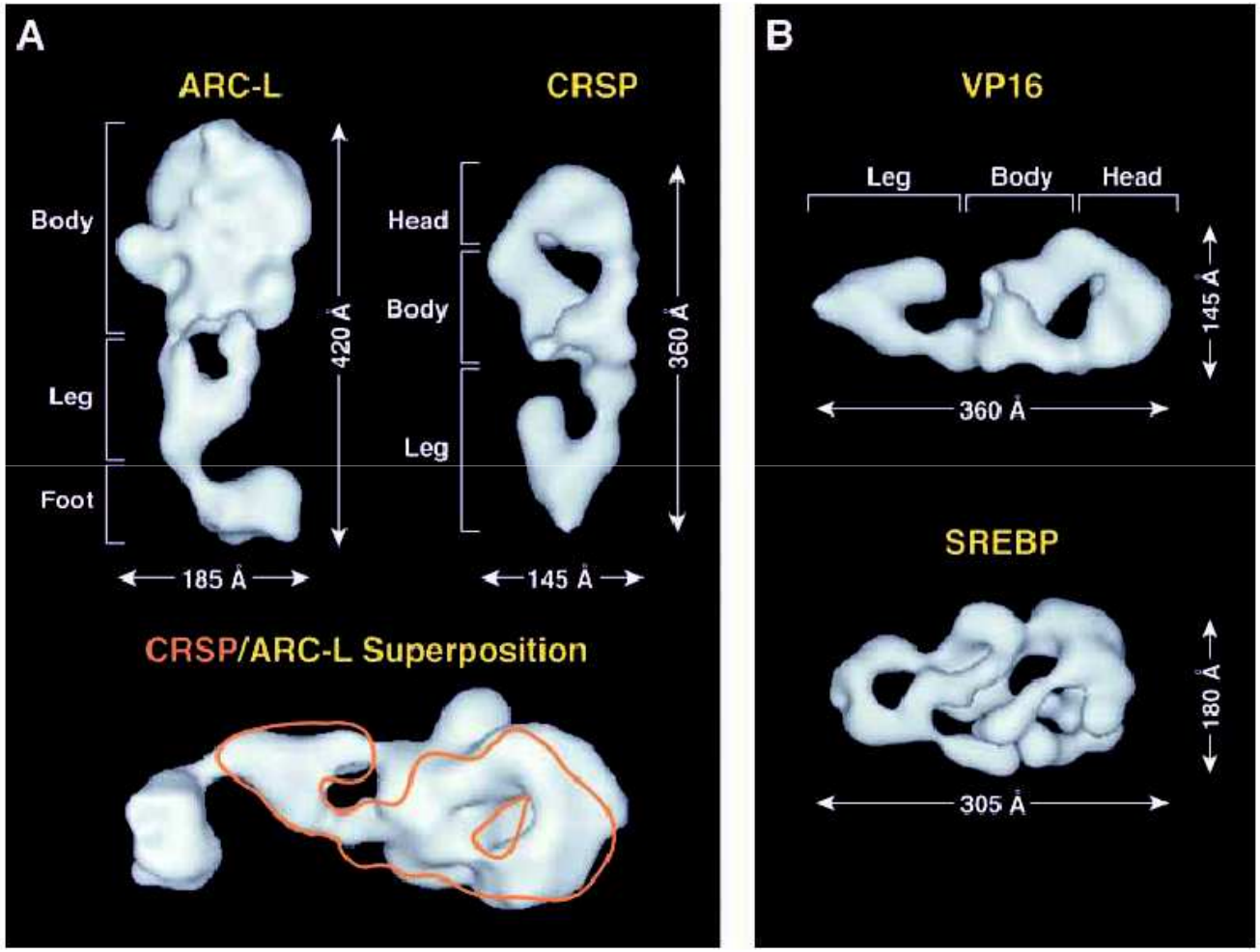


Fig. 2. Conformations of the mammalian mediator complexes. (A) EM composites of the ARC-L and CRSP complexes, which illustrate the size and structural differences between the two. (B) EM composites showing the distinctly different structural conformations adopted by CRSP when isolated via affinity interactions 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).

CONCLUSIONS

Direct interaction of regulatory TFs with the basal machinery is possible

but

“in vivo” is greatly enhanced through interaction with megadalton multifunctional protein complexes called “Mediator”

Second question:

-where are enhancers (regulatory modules) placed in genomes?

-how can we define “response elements” for a specific TF ?

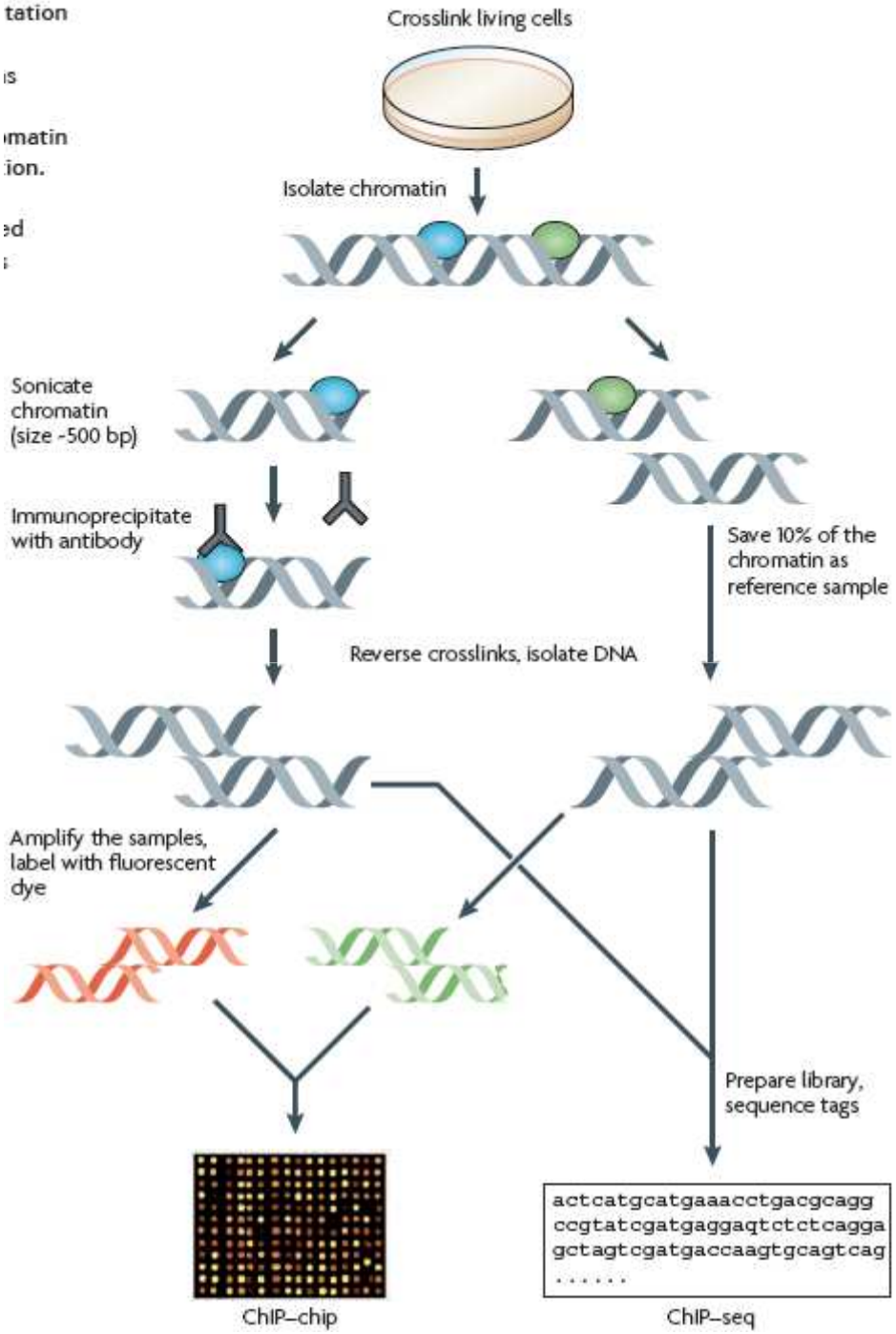
-is TF binding to a cognate sequence wherever sufficient for transactivation?

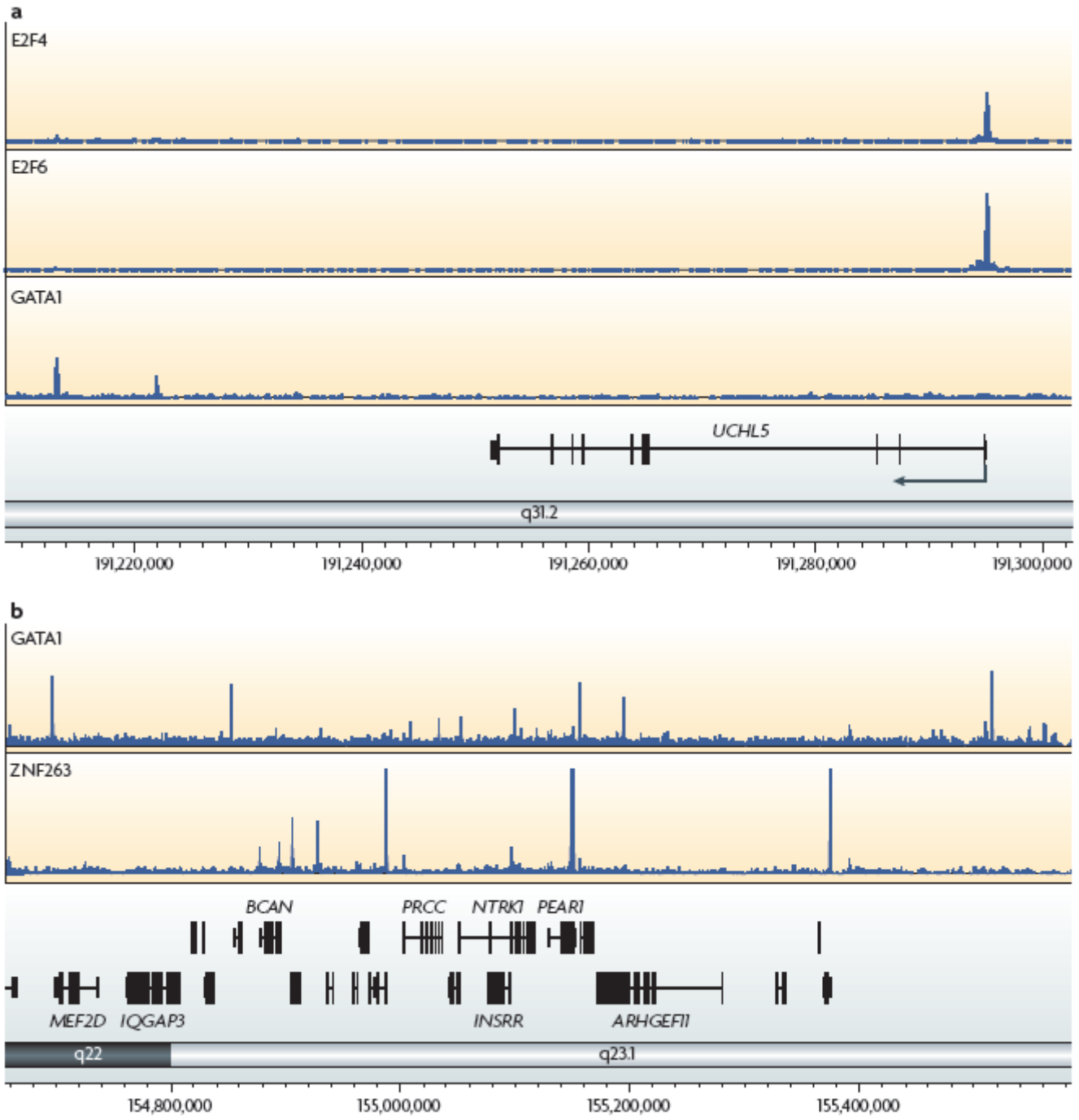
tation

is

matin
ion.

ed




a
E2F4
E2F6
GATA1
UCHL5
q31.2
191,220,000
191,240,000
191,260,000
191,280,000
191,300,000
b
GATA1
ZNF263
BCAN
PRCC
NTRK1
PEAR1
MEF2D
IQGAP3
INSRR
ARHGEF11
q22
q23.1
154,800,000
155,000,000
155,200,000
155,400,000

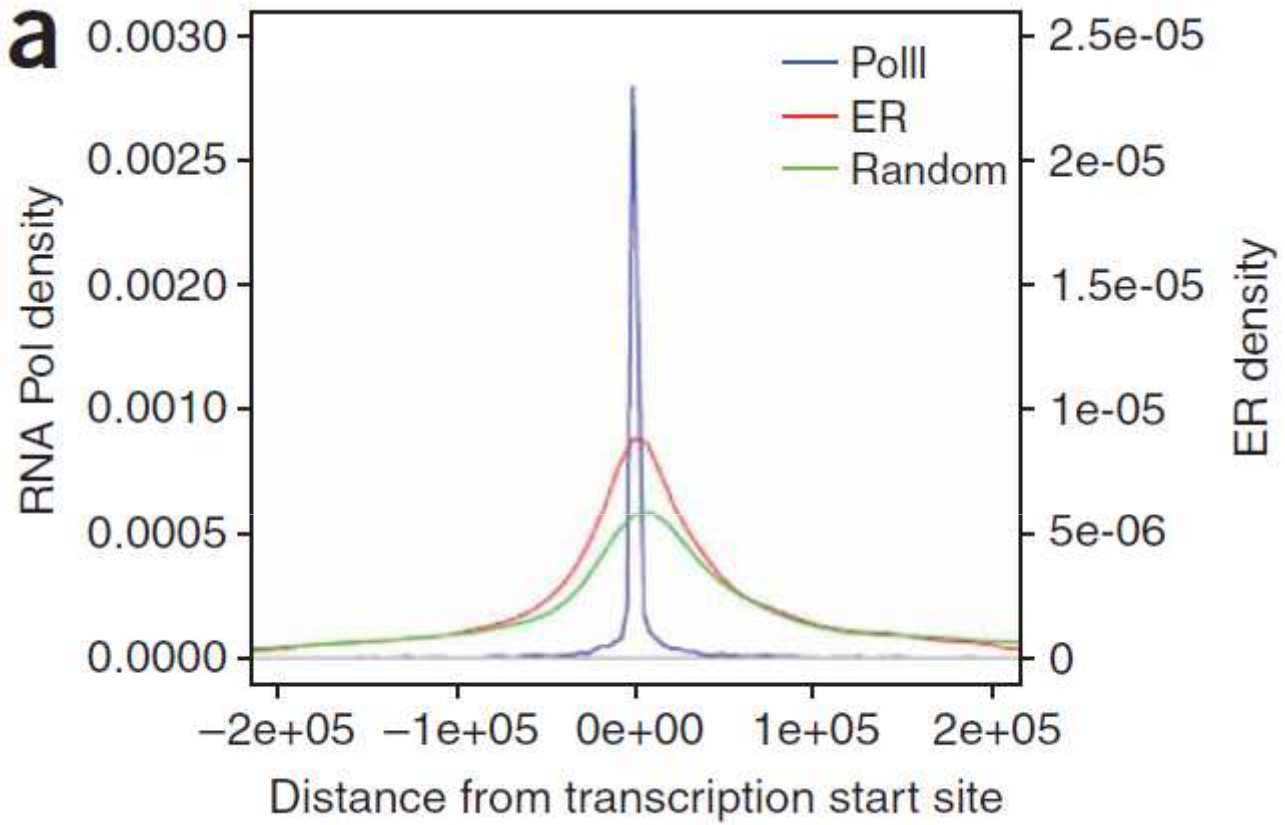
Genome-wide analysis of estrogen receptor binding sites

Jason S Carroll[1], Clifford A Meyer[2,3], Jun Song[2,3], Wei Li[2,3], Timothy R Geistlinger[1], Jérôme Eeckhoute[1], Alexander S Brodsky[4], Erika Krasnickas Keeton[1], Kirsten C Fertuck[1], Giles F Hall[5], Qianben Wang[1], Stefan Bekiranov[6,8], Victor Sementchenko[6], Edward A Fox[5], Pamela A Silver[5,7], Thomas R Gingeras[6], X Shirley Liu[2,3] & Myles Brown[1]

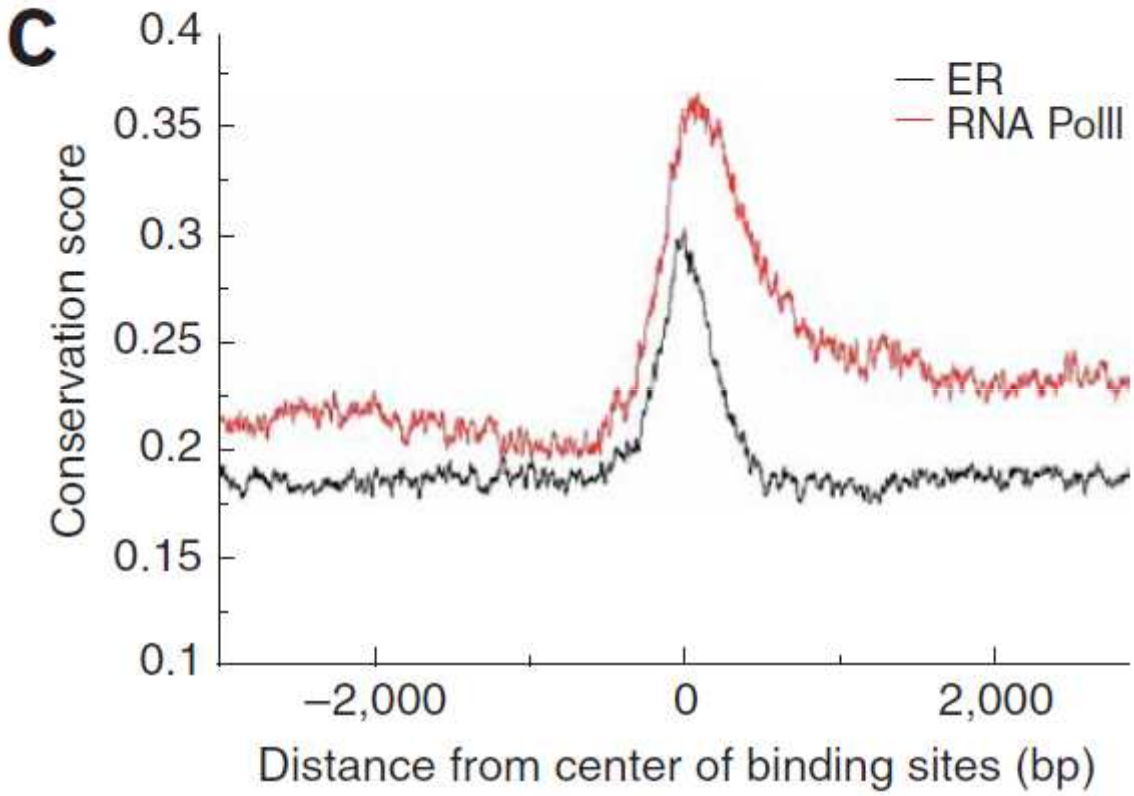
The estrogen receptor is the master transcriptional regulator of breast cancer phenotype and the archetype of a molecular therapeutic target. We mapped all estrogen receptor and RNA polymerase II binding sites on a genome-wide scale, identifying the authentic *cis* binding sites and target genes, in breast cancer cells. Combining this unique resource with gene expression data demonstrates distinct temporal mechanisms of estrogen-mediated gene regulation, particularly in the case of estrogen-suppressed genes. Furthermore, this resource has allowed the identification of *cis*-regulatory sites in previously unexplored regions of the genome and the cooperating transcription factors underlying estrogen signaling in breast cancer.

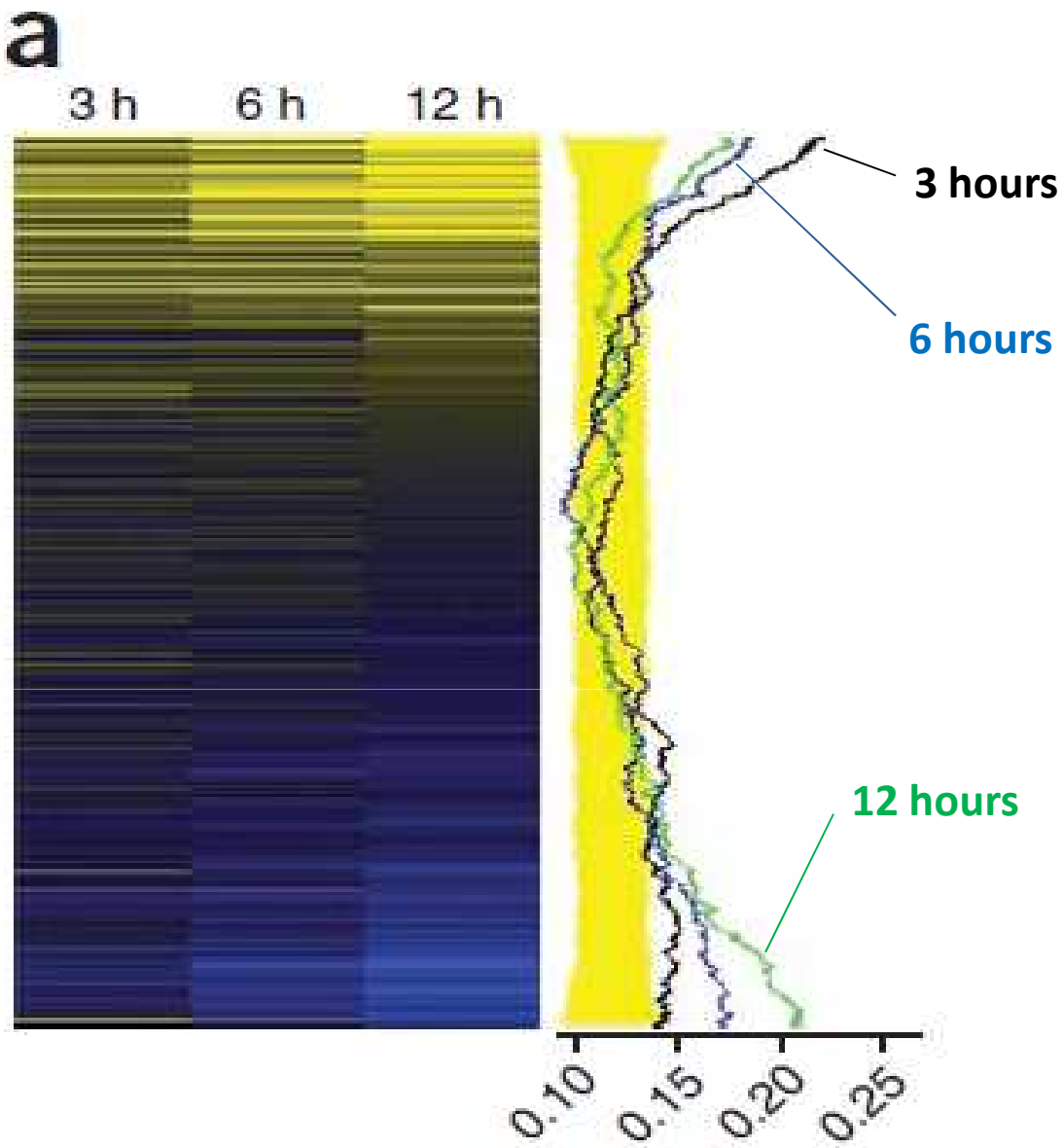
We deprived MCF-7 cells of hormones for 3 d and then synchronously induced transcription by the addition of estrogen for a brief period of time (45 min) known to result in maximal estrogen receptor–chromatin binding.

We used estrogen receptor–specific and RNA PolII–specific antibodies for ChIP and prepared precipitated chromatin.
We hybridized ChIP chromatin and input DNA to the Affymetrix Human tiling 1.0 microarrays representing the entire nonrepetitive human genome sequence (NCBI build 35) tiled at 35-bp resolution.

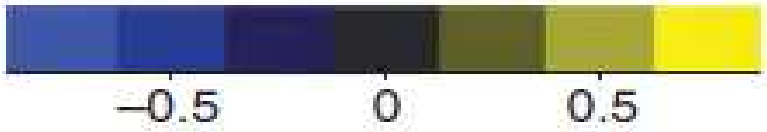


(a) Location of estrogen receptor (ER) and RNA PolII sites relative to transcription start sites (TSS) of RefSeq genes. The scale on the left represents RNA PolII distribution, and the scale on the right represents estrogen receptor and random distribution.

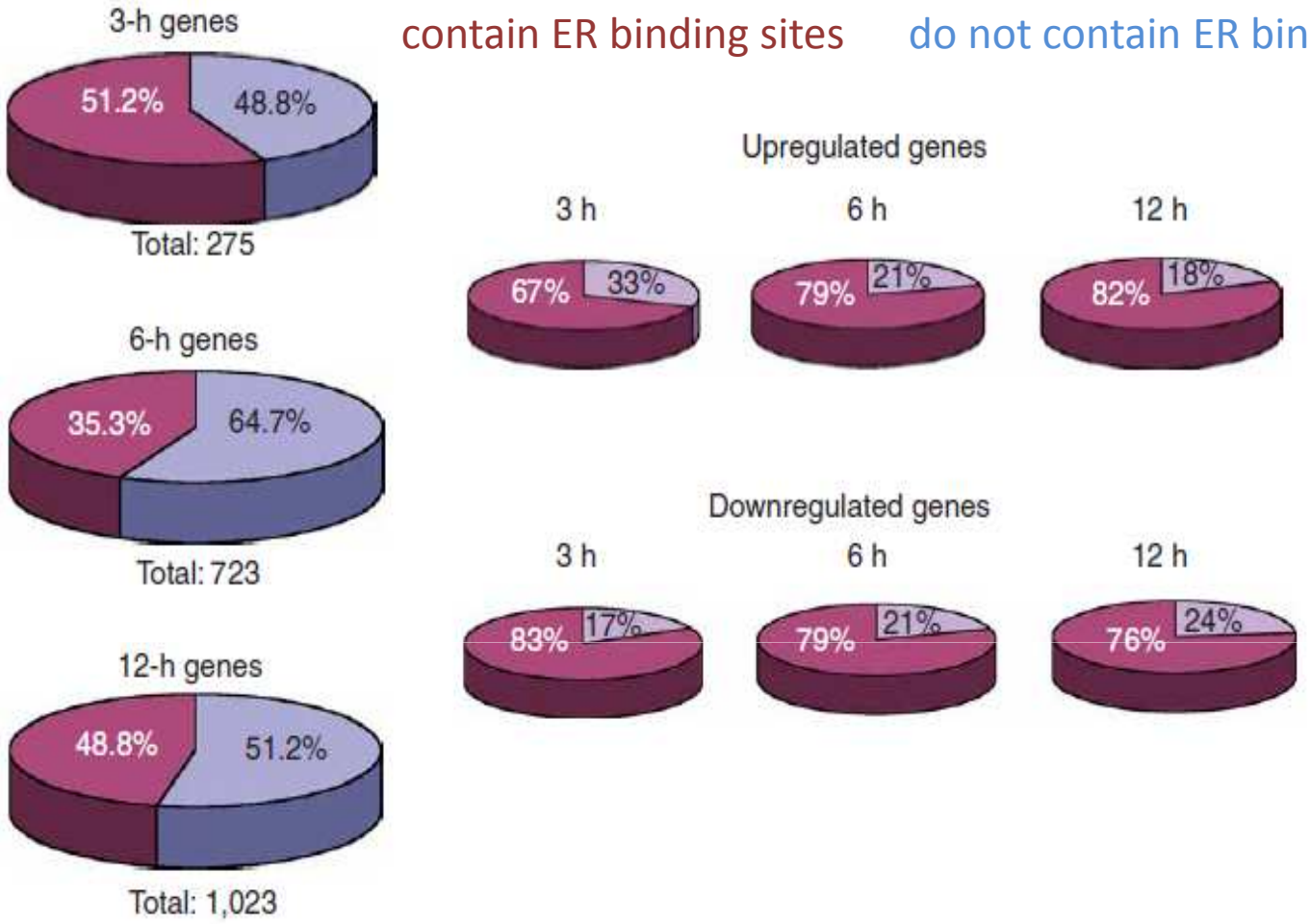

C
0.4
0.35
0.3
0.25
0.2
0.15
0.1
Conservation score
ER
RNA PolII
–2,000
0
2,000
Distance from center of binding sites (bp)



Expression

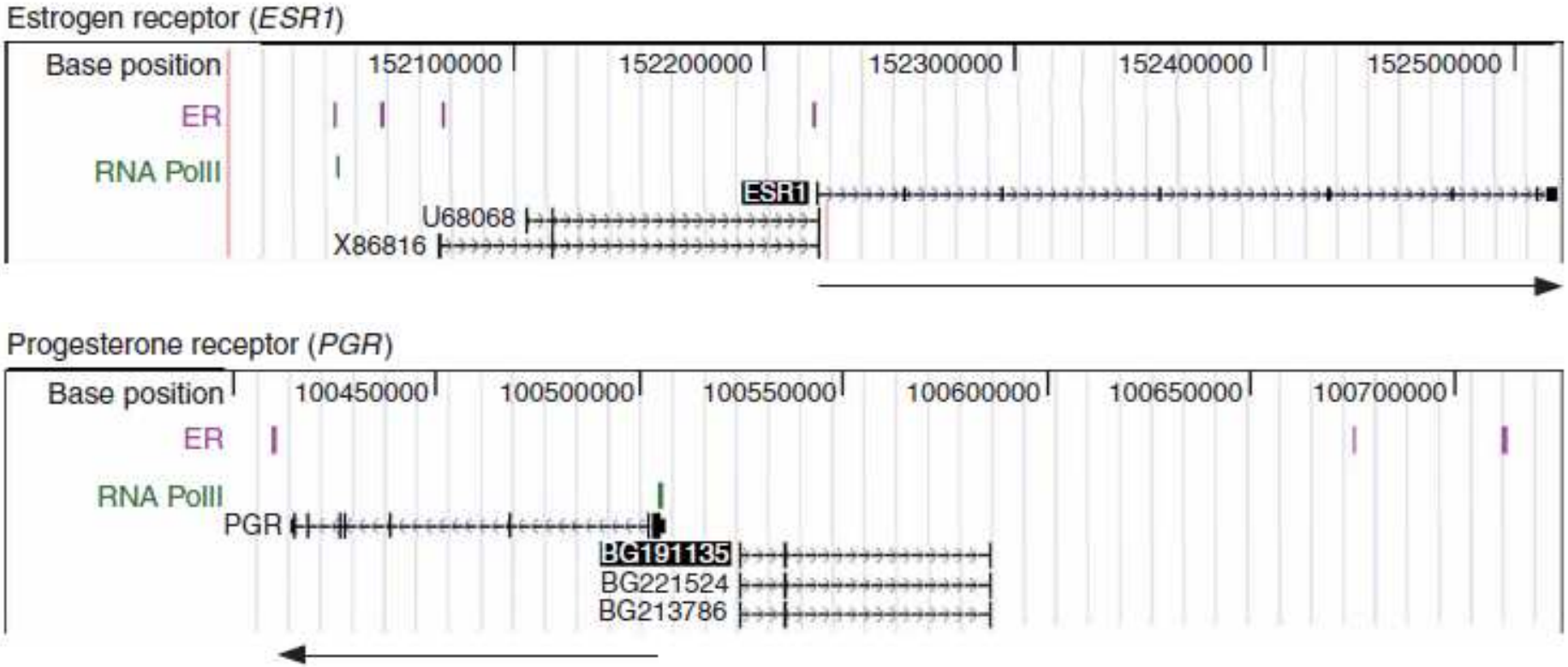


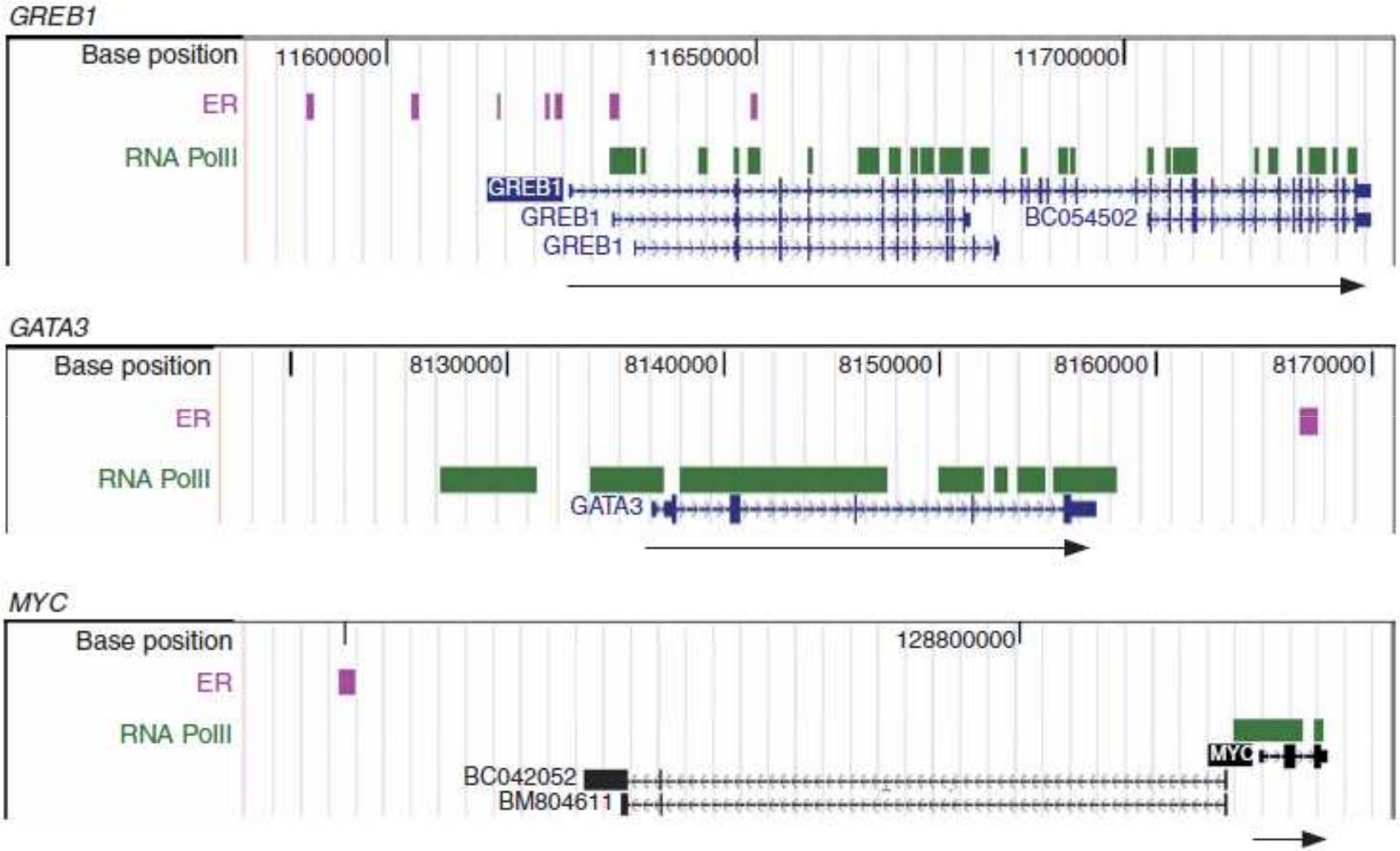
Estrogen-mediated transcript changes and correlation with estrogen receptor binding.
(a) Expression changes of all genes as ranked by Welch t statistic at 3, 6 and 12 h relative to 0 h. Induction of gene expression relative to 0 h is represented as yellow and repression as blue. The graph represents the fraction of genes with an estrogen receptor binding site within 50 kb of the transcription start site. Genes were ranked by Welch t statistic between 3, 6 and 12 h and 0 h (control). The black (3 h), blue (6 h) and green (12 h) lines represent 2,000 gene moving averages of the fraction of genes that have one or more estrogen receptor binding sites within 50 kb of the transcription start site. The yellow band is a 99% confidence interval for the binding site moving average of genes in the 25%–50% 12-h t statistic range.



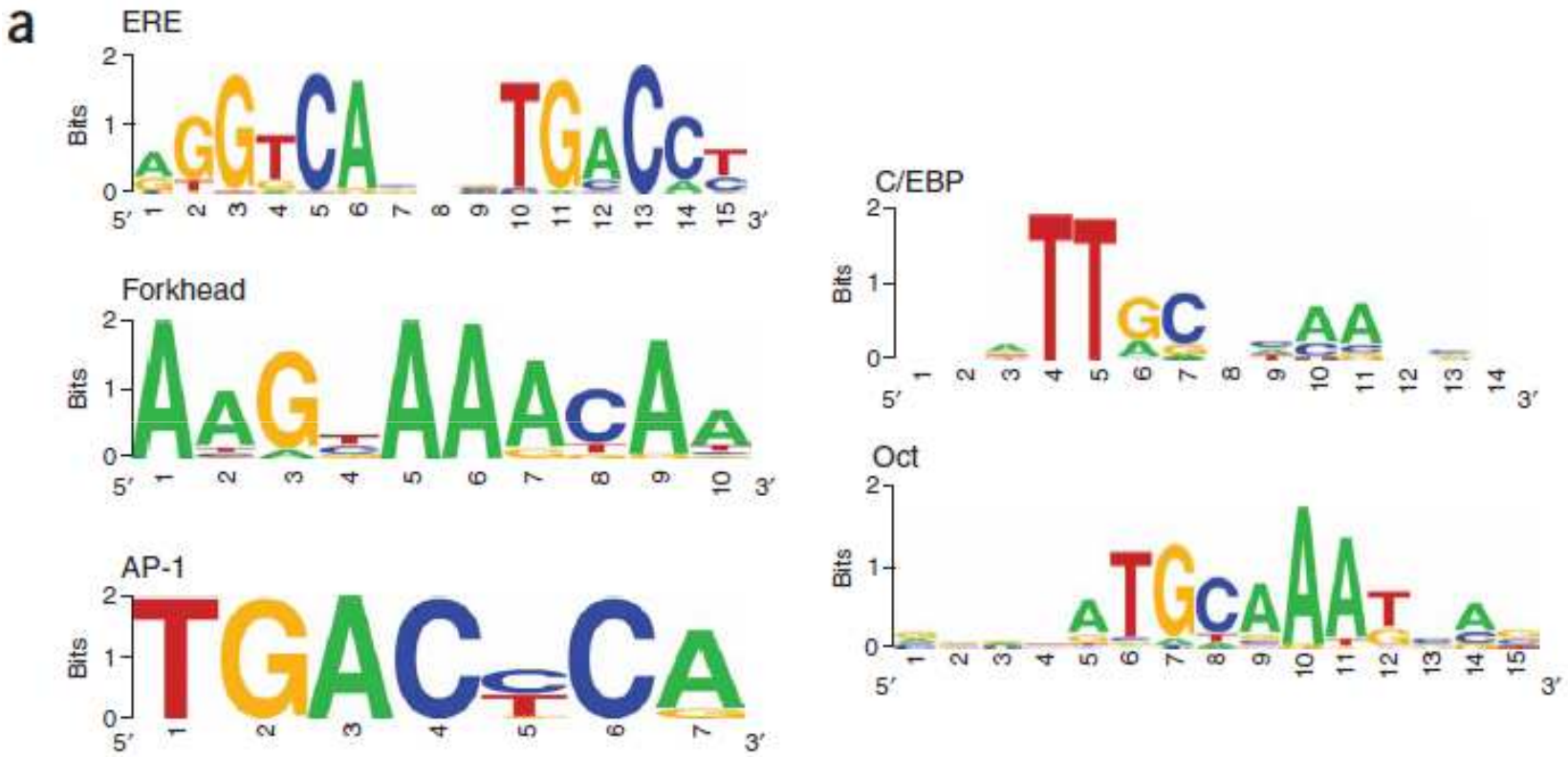
(b) Summary of estrogen mediated expression changes over a time course (0, 3, 6 and 12 h). Shown are the number of differentially expressed genes after estrogen treatment, relative to the vehicle-treated control (0 h). Blue segments represent upregulated genes, and red segments represent downregulated genes.
(c) Percentage of genes upregulated or downregulated at each time point (relative to time 0 h) that contain estrogen receptor binding sites within 50 kb (purple sector).

Examples of ER binding sites proximal to the regulated gene

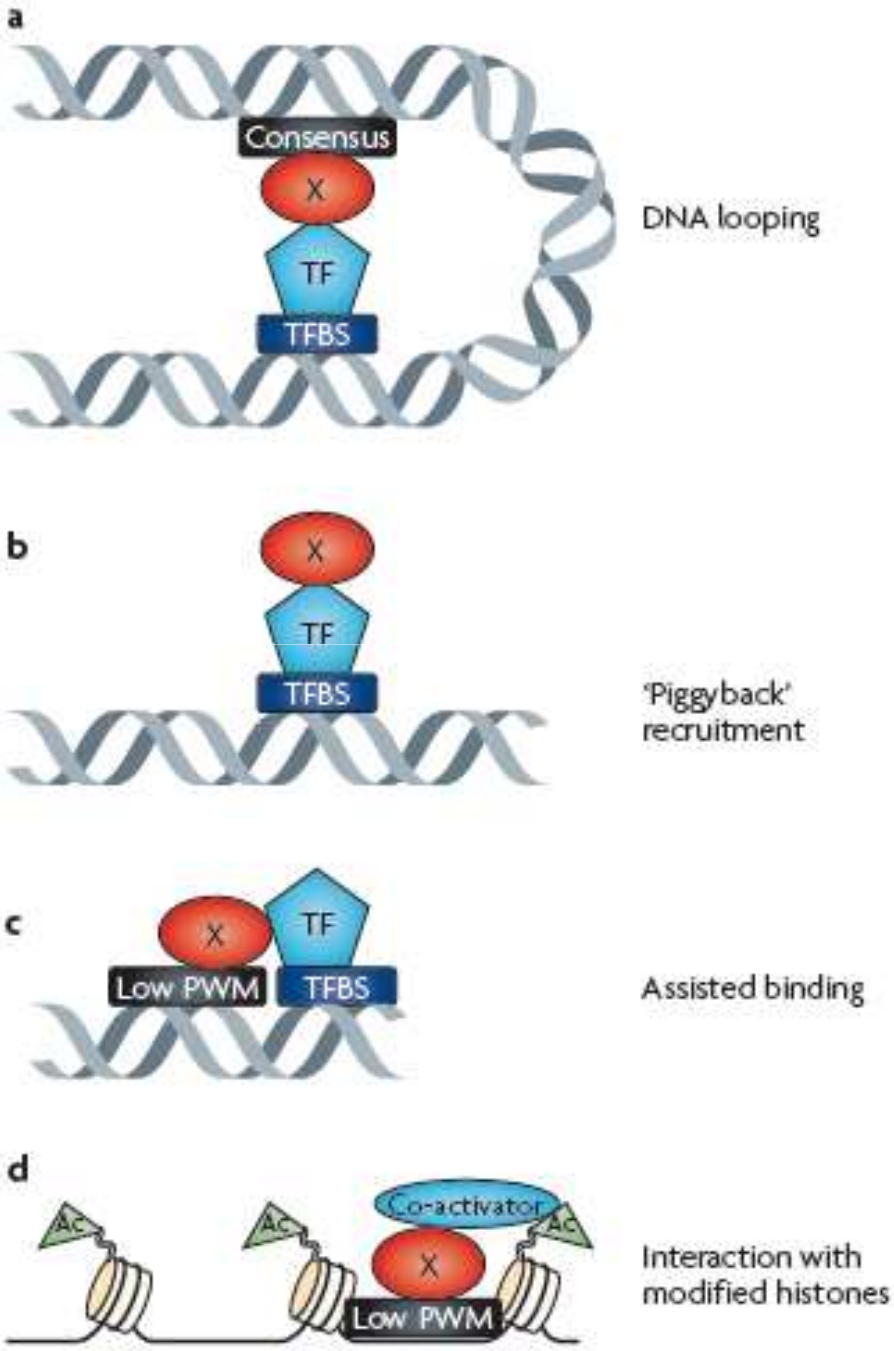



GREB1
Base position
11600000
11650000
11700000
ER
RNA PolII
GREB1
GREB1
GREB1
BC054502
GATA3
Base position
8130000
8140000
8150000
8160000
8170000
ER
RNA PolII
GATA3
MYC
Base position
128800000
ER
RNA PolII
MYC
BC042052
BM804611

In the -500, +500 iterval around ER binding peaks, the following matrices of TF binding sites were found



Note that more than half of ER-binding locations does not possess any ERE. So how can the ER protein be recruited to these sites?

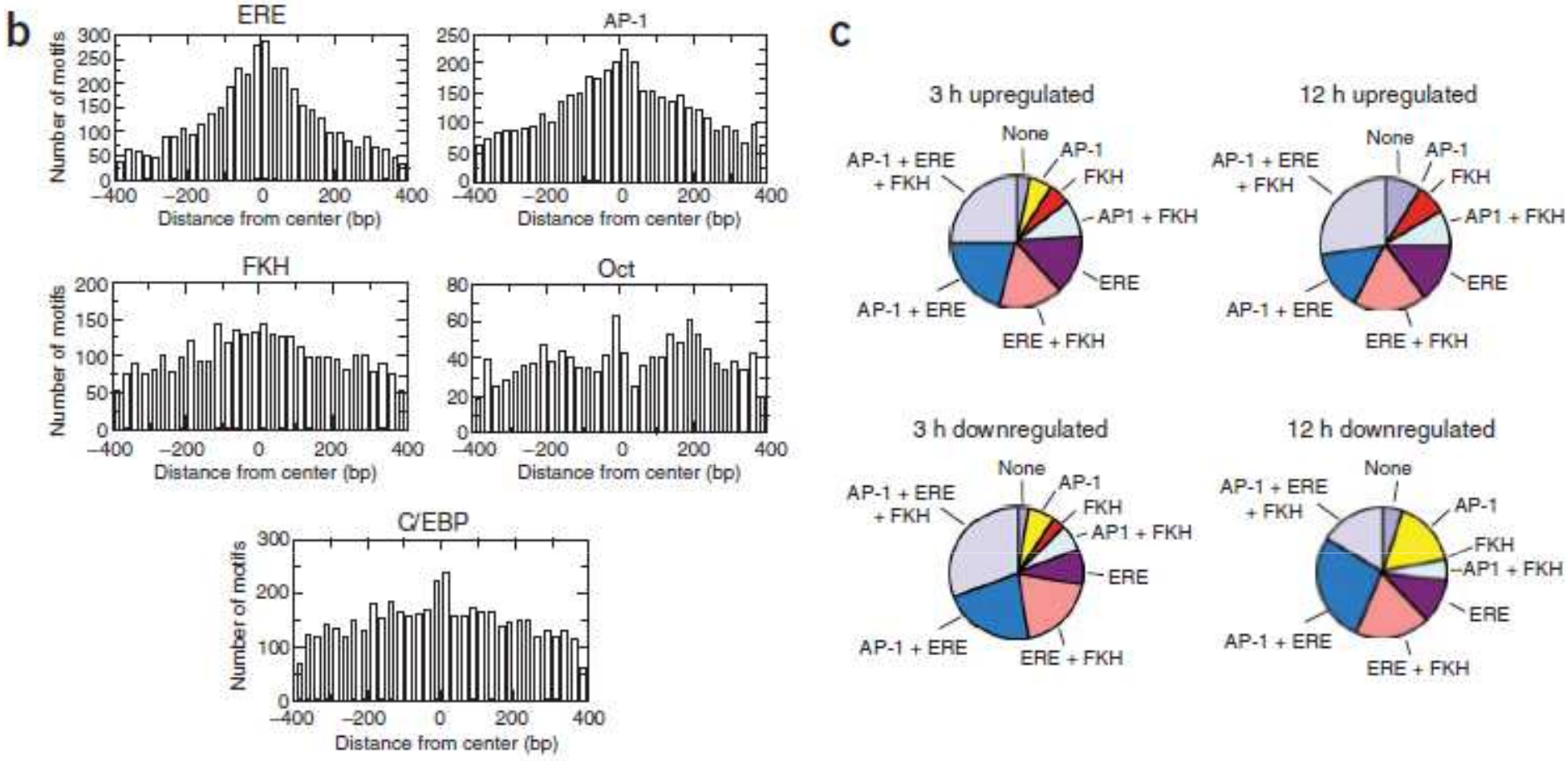


Artifact due to looping

Recruitmetn through protein-protein interaction with another TF

Binding to “poor” nonconsensus element aided by another TF

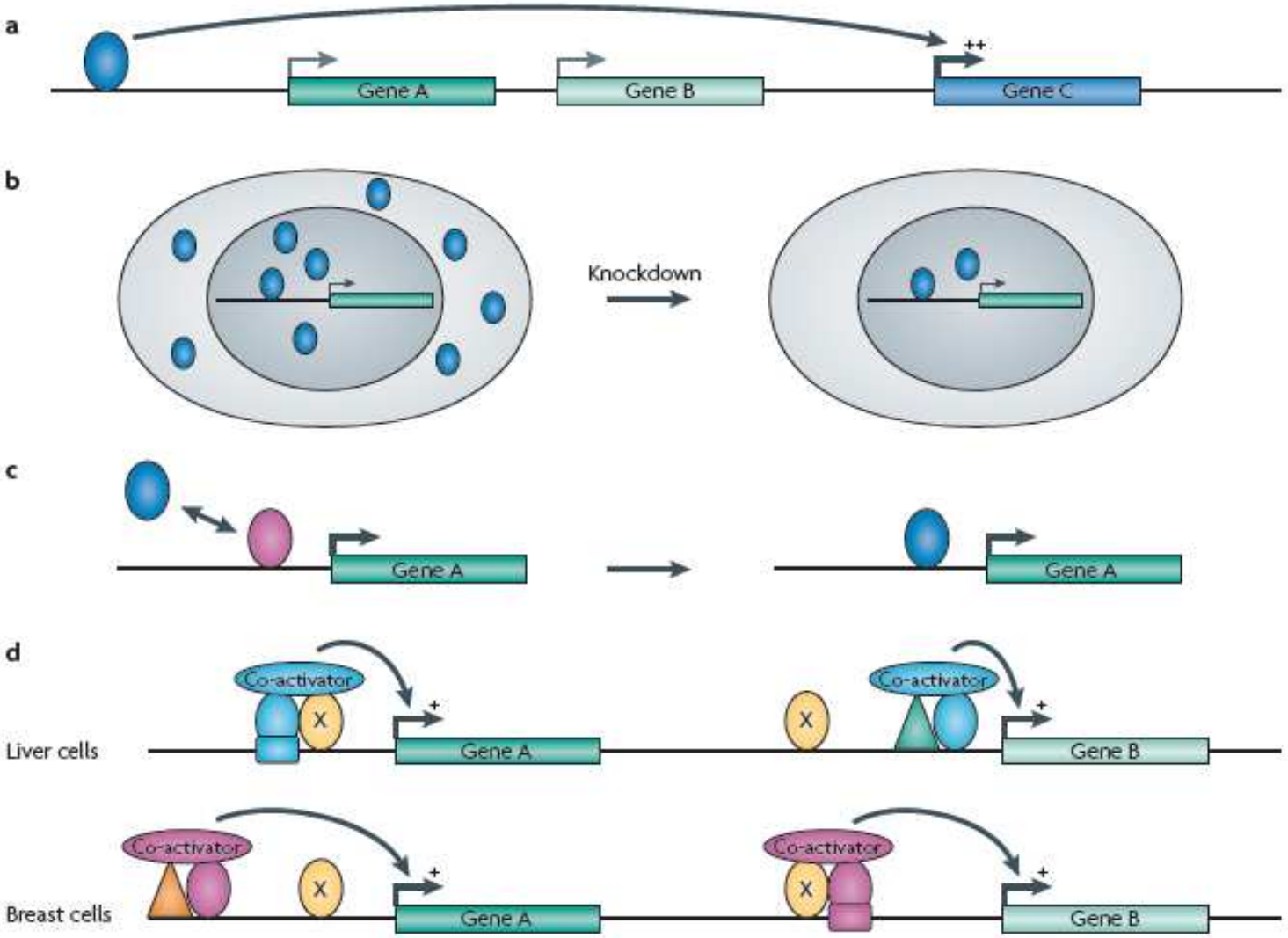
Binding to “poor” nonconsensus element aided by modified chromatin



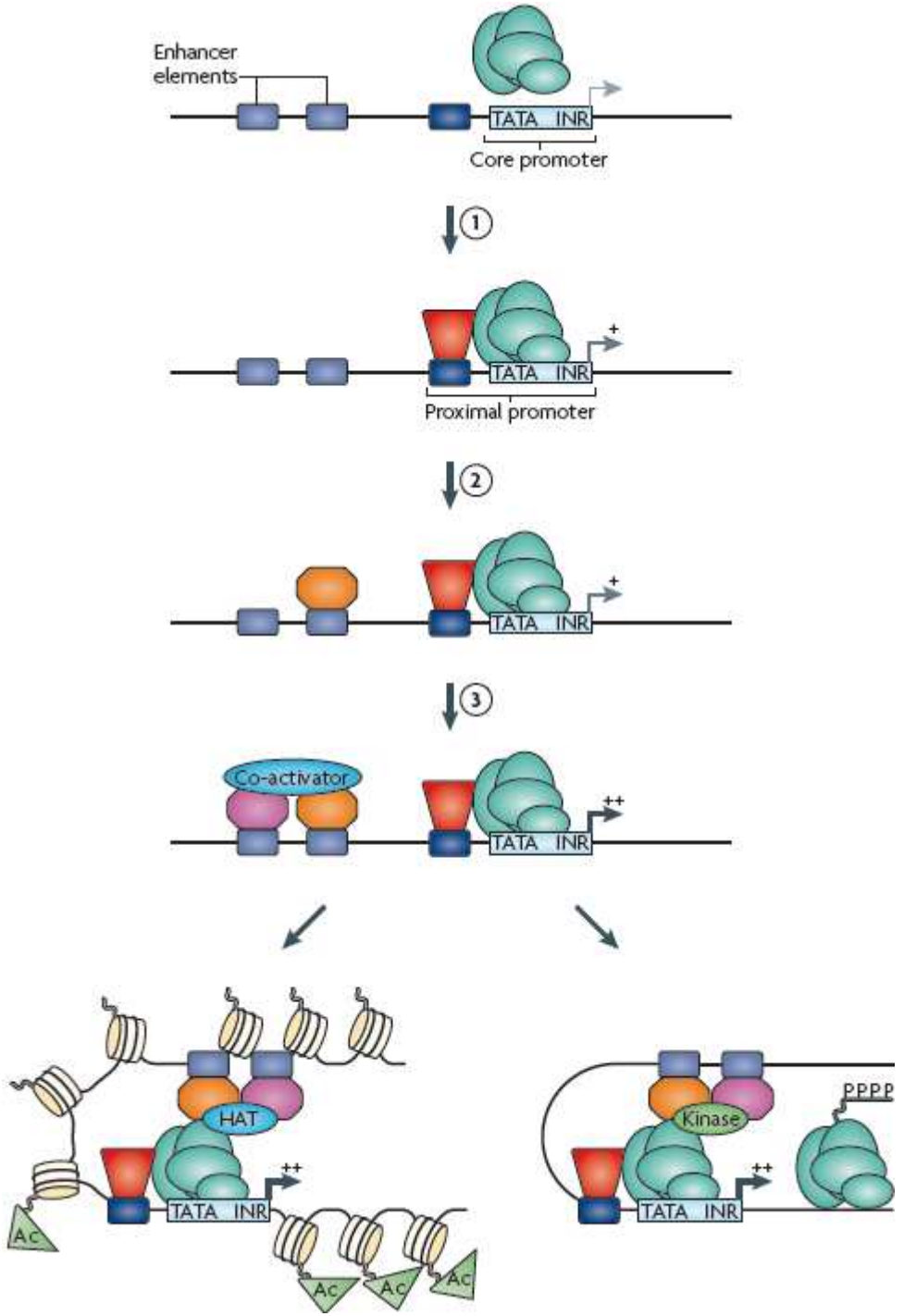
Other TF binding sistes found enriched in ER-regulated genes

Serious problem:

how to attribute an enhacer-promoter couple with no experimental data?


a
Gene A
Gene B
++
Gene C
b
Knockdown
c
Gene A
Gene A
d
Liver cells
Co-activator
X
+
Gene A
X
Co-activator
+
Gene B
Breast cells
Co-activator
X
+
Gene A
Co-activator
X
+
Gene B

Evolved model



DATABASES

Entrez Gene: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene
IFNB1 | UCHL5

UniProtKB: http://www.uniprot.org
FOXA2 | GATA1 | NANOG | NRSF | OCT4 | oestrogen receptor | p53 | p63 | SOX2 | SRF | STAT1 | TAF1 | TCF4 | ZNF263

FURTHER INFORMATION

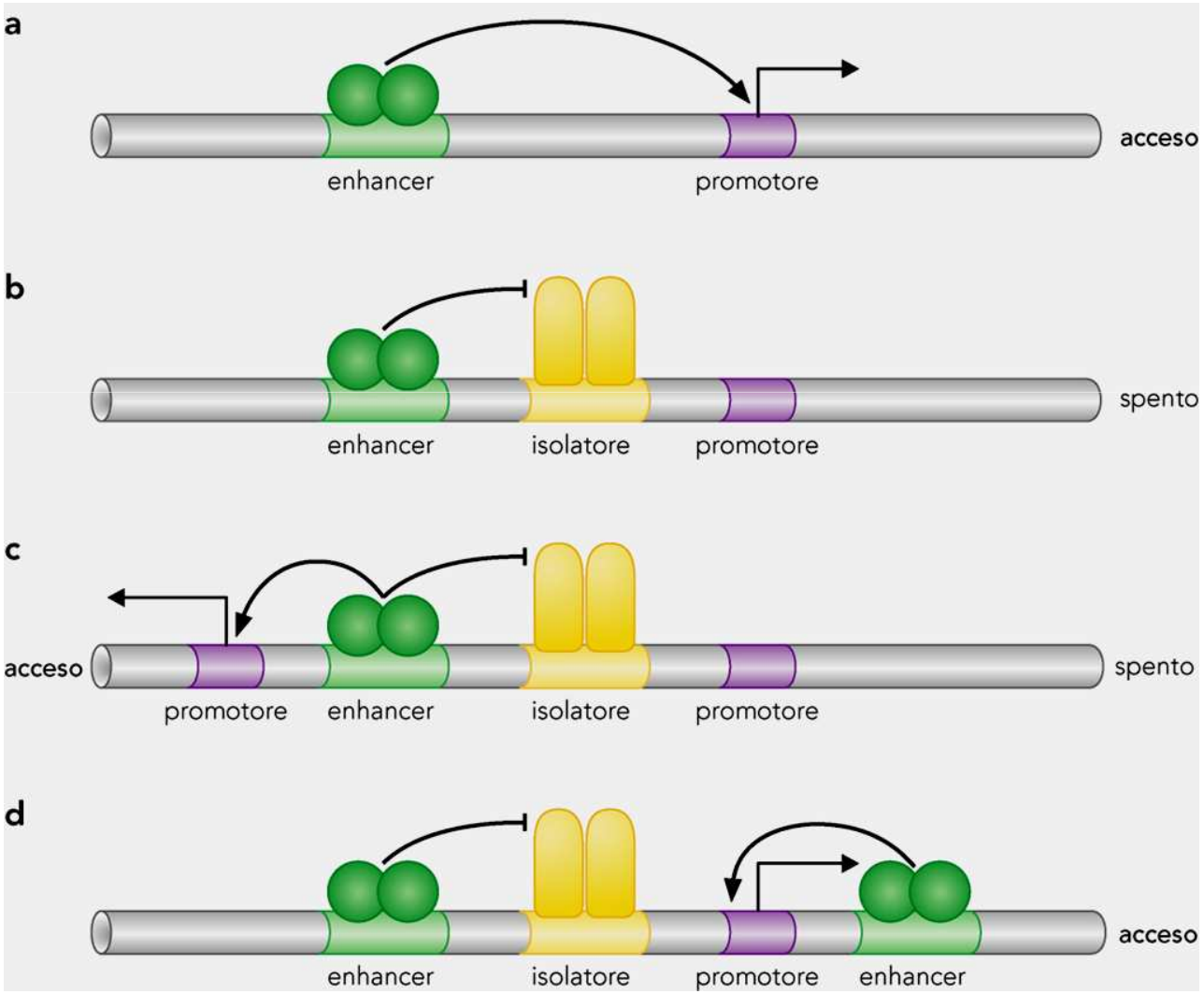
Peggy J. Farnham's homepage:
http://www.genomecenter.ucdavis.edu/farnham
ENCODE Data Coordination Center at UCSC:
http://www.genome.ucsc.edu/ENCODE
The ENCODE Project: http://www.genome.gov/10005107
NIH Roadmap Epigenomics Program:
http://nihroadmap.nih.gov/epigenomics/

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

Enhancer-promoter interaction

Gli enhancers sono sequenze regolatrici composte di molteplici siti di legame per fattori trascrizionali, localizzati in punti molto variabili del gene, anche a distanze considerevoli (50-100 Kb).

Perchè un enhancer non regola qualsiasi gene nelle vicinanze?



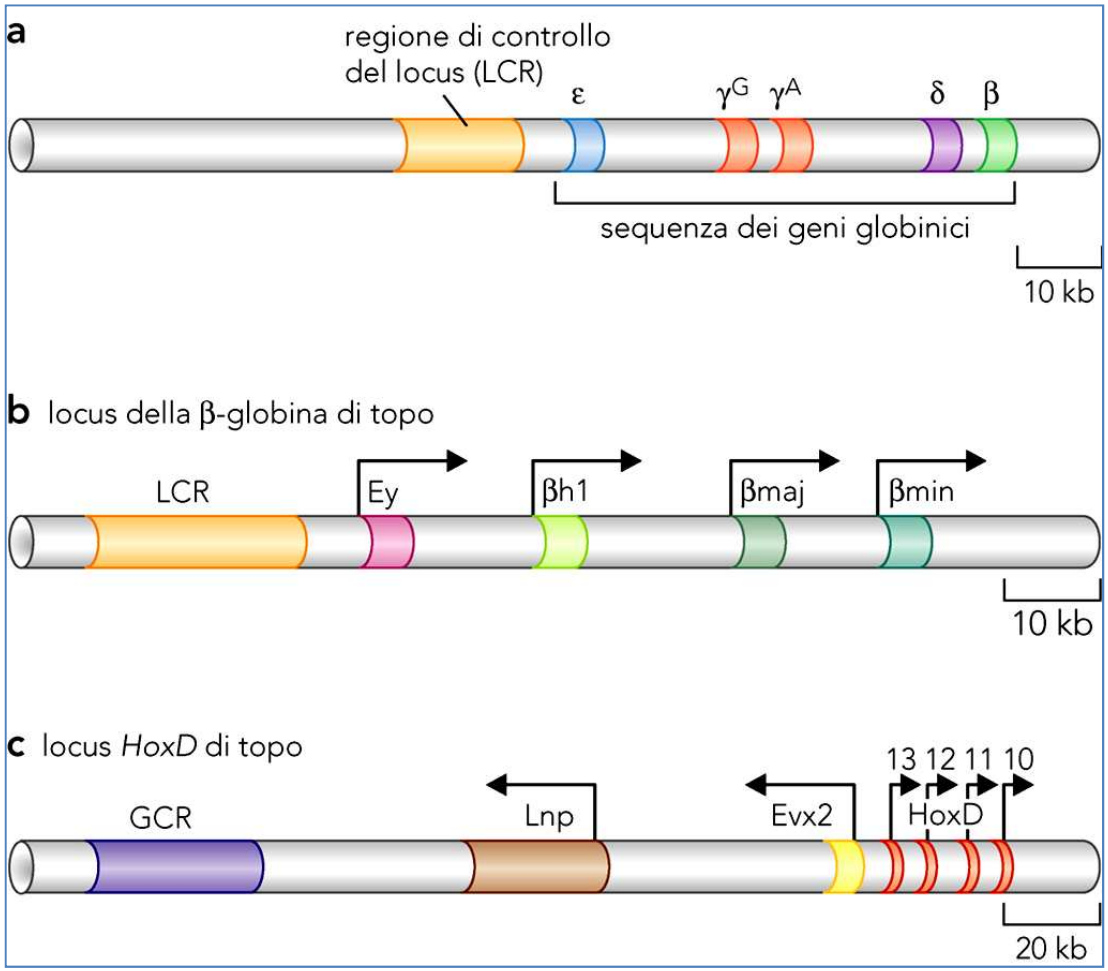
Particolari sequenze, chiamate isolatori (*insulators*), limitano l'effetto di un enhancer

L’esempio di eve di Drosophila ci ha illustrato
The example of Drosophila eve transcriptional regulatory enhancers has illustrated the fact that many genes possess more than one enhancer control “module” that may respond to different transcription factors in different cells/ tissues/ developmental stages etc.

On the opposite,

Gene clusters that derive from gene duplication events are often controlled by one or more common enhancer module, which interacts in a exclusive fashion with one or the other promoter. These common control regions are called LCR (*locus control region*).

LCRs control the sequential and exclusive use of one promoter at the time. An example is given by the globin gene clusters, containing embryonic, phoetal and adult versions of the globin proteins. Another example is given by the gene clusters encoding homeoproteins, that are expressed following a precise spatial order in the body.



Transcription factors mediate long-range enhancer–promoter interactions

Ilias K. Nolis[a,1], Daniel J. McKay[b,1], Eva Mantouvalou[a], Stavros Lomvardas[c,2], Menie Merika[a], and Dimitris Thanos[a,3]

[a]Institute of Molecular Biology, Genetics and Biotechnology, Biomedical Research Foundation, Academy of Athens, 4 Soranou Efesiou Street, Athens 11527 Greece; and [b]Integrated Program in Cellular, Molecular and Biophysical Studies and [c]Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032

Edited by Tom Maniatis, Harvard University, Cambridge, MA, and approved October 9, 2009 (received for review March 5, 2009)

We examined how remote enhancers establish physical communication with target promoters to activate gene transcription in response to environmental signals. Although the natural IFN-β enhancer is located immediately upstream of the core promoter, it also can function as a classical enhancer element conferring virus infection-dependent activation of heterologous promoters, even when it is placed several kilobases away from these promoters. We demonstrated that the remote IFN-β enhancer "loops out" the intervening DNA to reach the target promoter. These chromatin loops depend on sequence-specific transcription factors bound to the enhancer and the promoter and thus can explain the specificity observed in enhancer–promoter interactions, especially in complex genetic loci. Transcription factor binding sites scattered between an enhancer and a promoter can work as decoys trapping the enhancer in nonproductive loops, thus resembling insulator elements. Finally, replacement of the transcription factor binding sites involved in DNA looping with those of a heterologous prokaryotic protein, the λ repressor, which is capable of loop formation, rescues enhancer function from a distance by re-establishing enhancer–promoter loop formation.

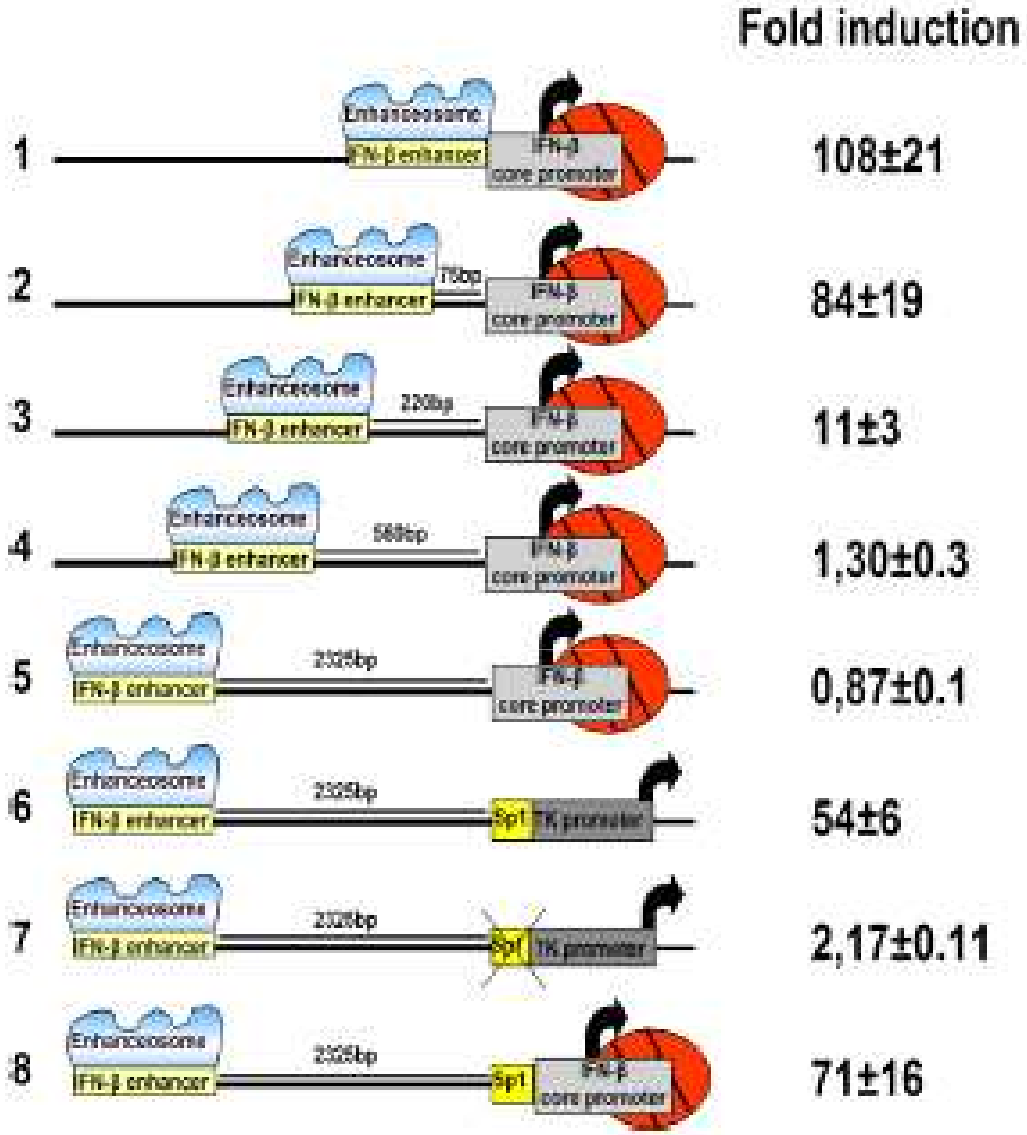
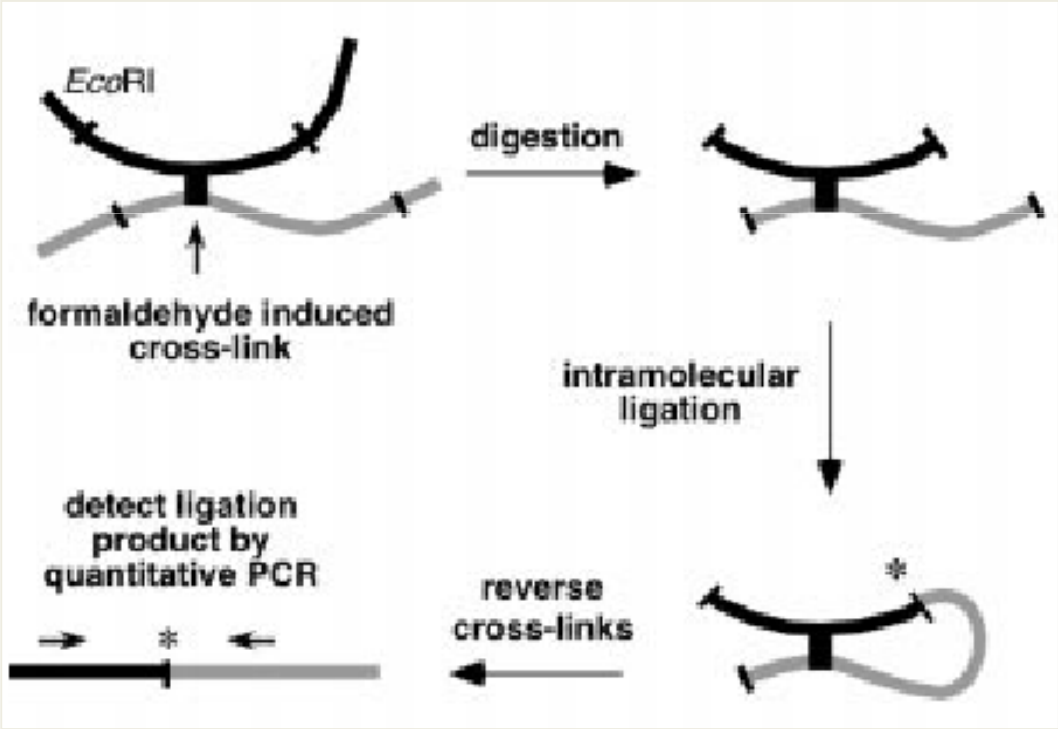
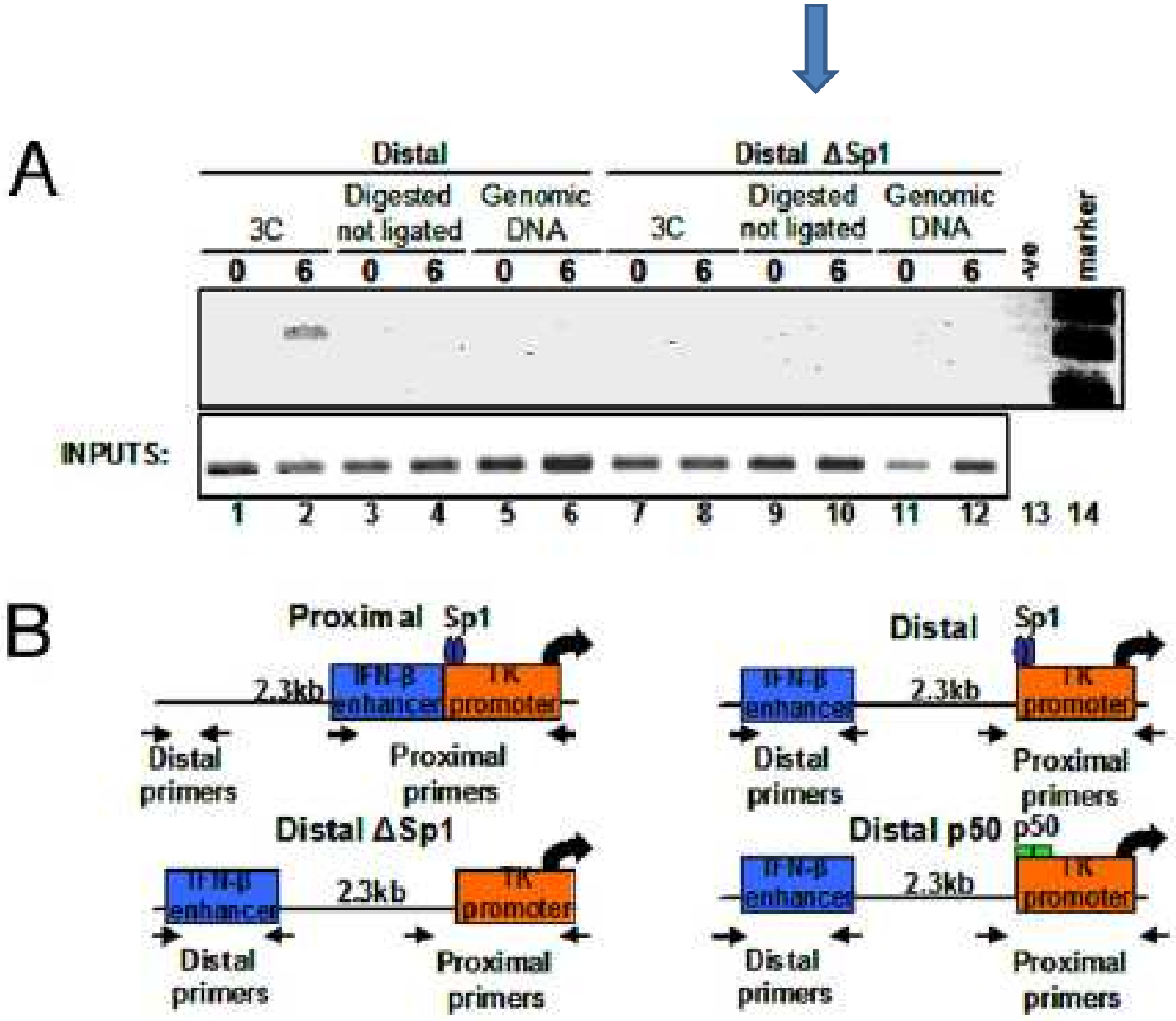


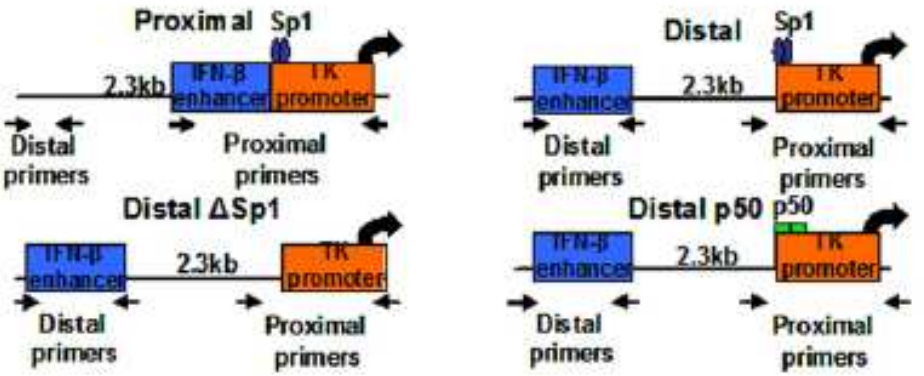
Fig. 1. Enhancer action from a distance requires upstream promoter elements. HeLa cells were transfected with the indicated chloramphenicol acetyl transferase (CAT) reporter plasmids. The cells were mock or virus infected for 24 h before being harvested. Then CAT activity was determined.

3C assay = chromosome conformation capture

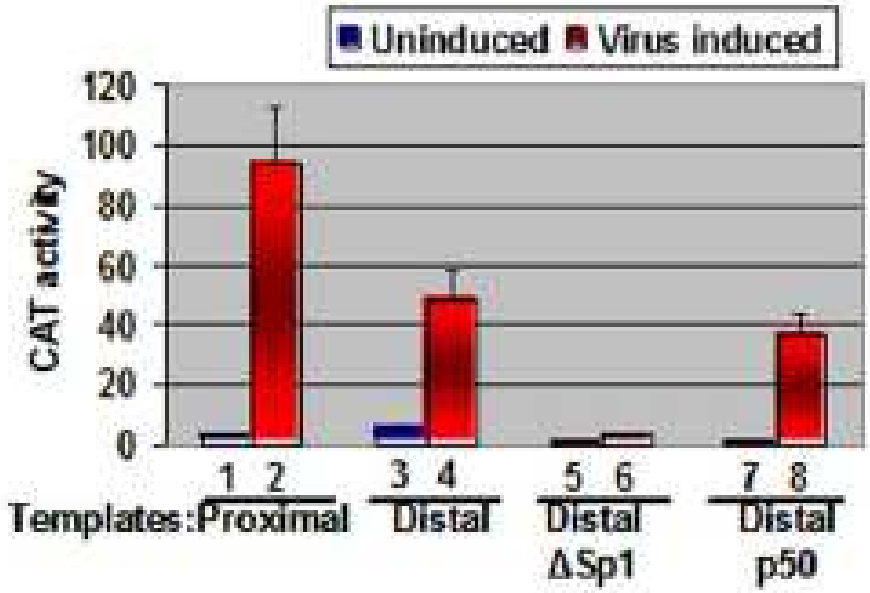



A
Distal
Distal ΔSp1
3C
Digested not ligated
Genomic DNA
3C
Digested not ligated
Genomic DNA
0 6 0 6 0 6 0 6 0 6 0 6
-ve
marker
INPUTS:
1 2 3 4 5 6 7 8 9 10 11 12 13 14
B
Proximal Sp1
2.3kb
IFN-β enhancer
TK promoter
Distal primers
Proximal primers
Distal ΔSp1
IFN-β enhancer
2.3kb
TK promoter
Distal primers
Proximal primers
Distal Sp1
IFN-β enhancer
2.3kb
TK promoter
Distal primers
Proximal primers
Distal p50 p50
IFN-β enhancer
2.3kb
TK promoter
Distal primers
Proximal primers

B



C



stable transfection

Molecular Cell, Vol. 10, 1453–1465, December, 2002,

Looping and Interaction between Hypersensitive Sites in the Active β-globin Locus

Bas Tolhuis,[2] Robert-Jan Palstra,[2] Erik Splinter,
Frank Grosveld, and Wouter de Laat[1]
Department of Cell Biology and Genetics
Faculty of Medicine
Erasmus University, Rotterdam
P.O. Box 1738
3000DR Rotterdam
The Netherlands

Summary

Eukaryotic transcription can be regulated over tens or even hundreds of kilobases. We show that such long-range gene regulation in vivo involves spatial interactions between transcriptional elements, with intervening chromatin looping out. The spatial organization of a 200 kb region spanning the murine β-globin locus was analyzed in expressing erythroid and nonexpressing brain tissue. In brain, the globin cluster adopts a seemingly linear conformation. In erythroid cells the hypersensitive sites of the locus control region (LCR), located 40–60 kb away from the active genes, come in close spatial proximity with these genes. The intervening chromatin with inactive globin genes loops out. Moreover, two distant hypersensitive regions participate in these interactions. We propose that clustering of regulatory elements is key to creating and maintaining active chromatin domains and regulating transcription.

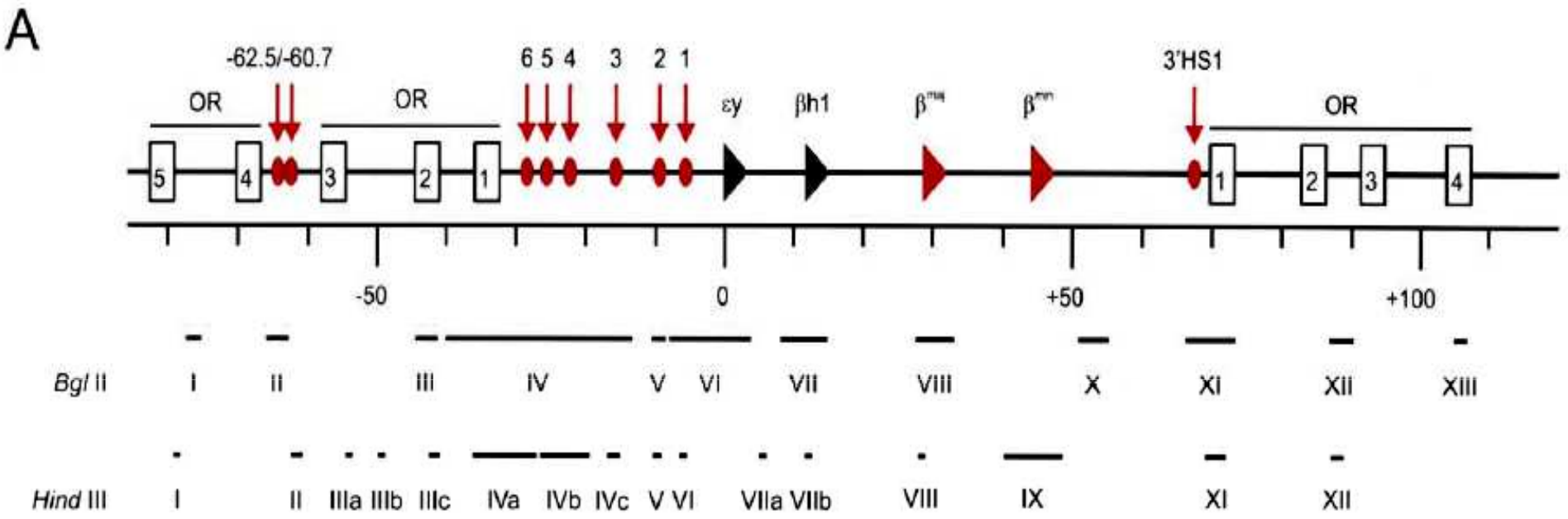
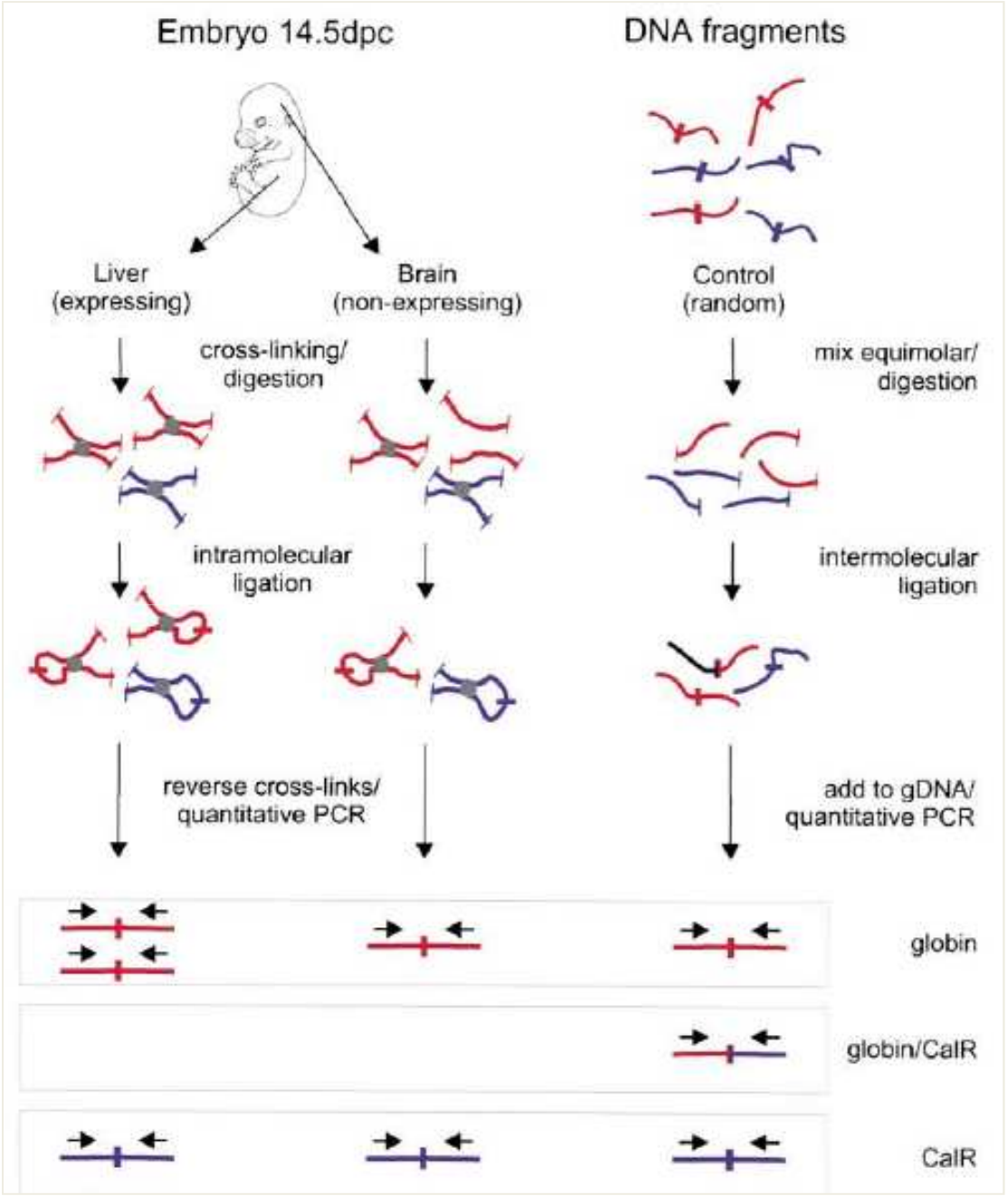
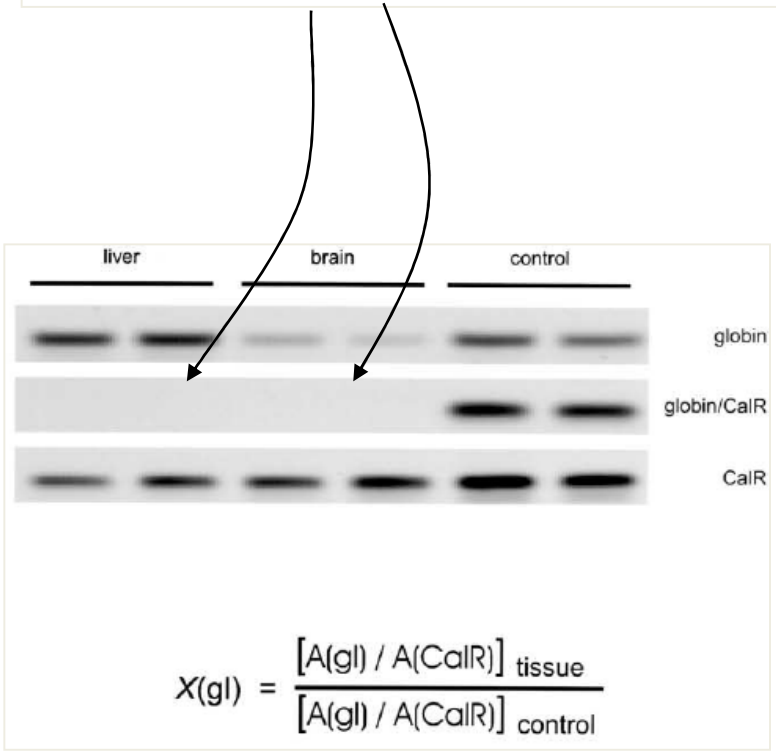


Figure 1. 3C Technology in the Murine beta-globin Locus

(A) Schematic presentation of the murine **beta-globin** locus. Red arrows and ellipses depict the individual HS. The globin genes are indicated by triangles, with **active genes (maj and min) in red** and **inactive genes (y and h1) in black**. The white boxes indicate the olfactory receptor (OR) genes (5OR1-5 and 3OR1-4). The two sets of restriction fragments (BglII and HindIII) that were used for 3C analysis are shown below the locus. The individual fragments are indicated by Roman numerals. Identical numbering between BglII and HindIII indicates that two fragments colocalize. Distances are in kb counting from the site of initiation of the y gene.



No intermolecular ligation detected



$$X(\mathrm{gl}) = \frac{[A(\mathrm{gl}) / A(\mathrm{CalR})]_{\mathrm{tissue}}}{[A(\mathrm{gl}) / A(\mathrm{CalR})]_{\mathrm{control}}}$$

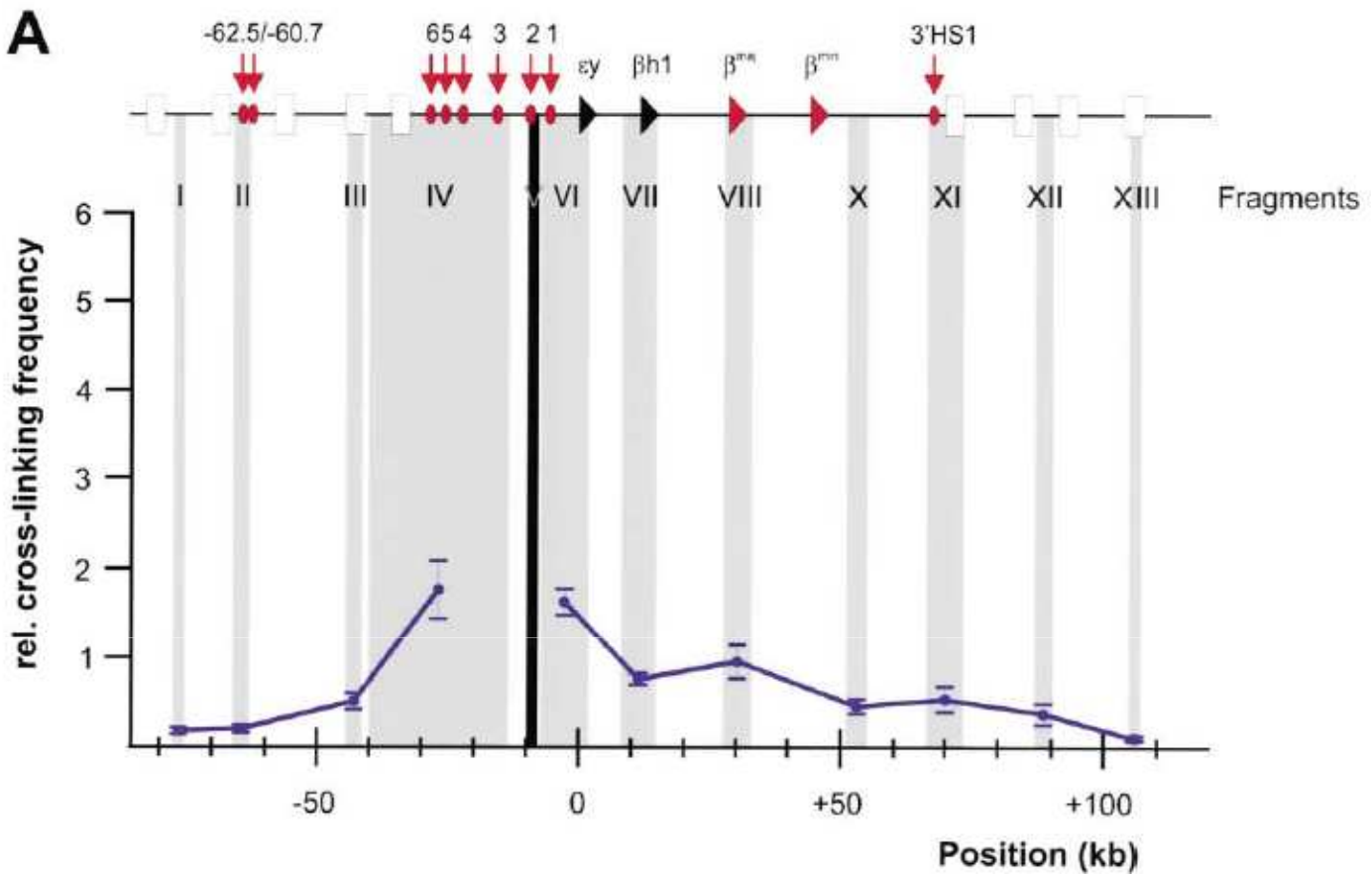


Figure 2. Linear Conformation of the beta-globin Locus in **Nonexpressing Brain Cells**.
The murine -globin locus is depicted on top of each graph (for explanation of symbols, see Figure 1A). The *x* axis shows the position in the locus. The black shading shows the position and size of the fixed fragment. The gray shading indicates the position and size of other fragments. Standard error of the mean is indicated. Crosslinking frequency with a value of 1 arbitrarily corresponds to the crosslinking frequency between two neighboring CalR control fragments (with restriction sites analyzed being 1.5 kb apart). Scaling on the *y* axis (from 0 to 6) allows direct comparison with Figures 3–6.
(**A**) Relative crosslinking frequencies between fixed BglII fragment V (5HS2 in LCR) and the rest of the locus.

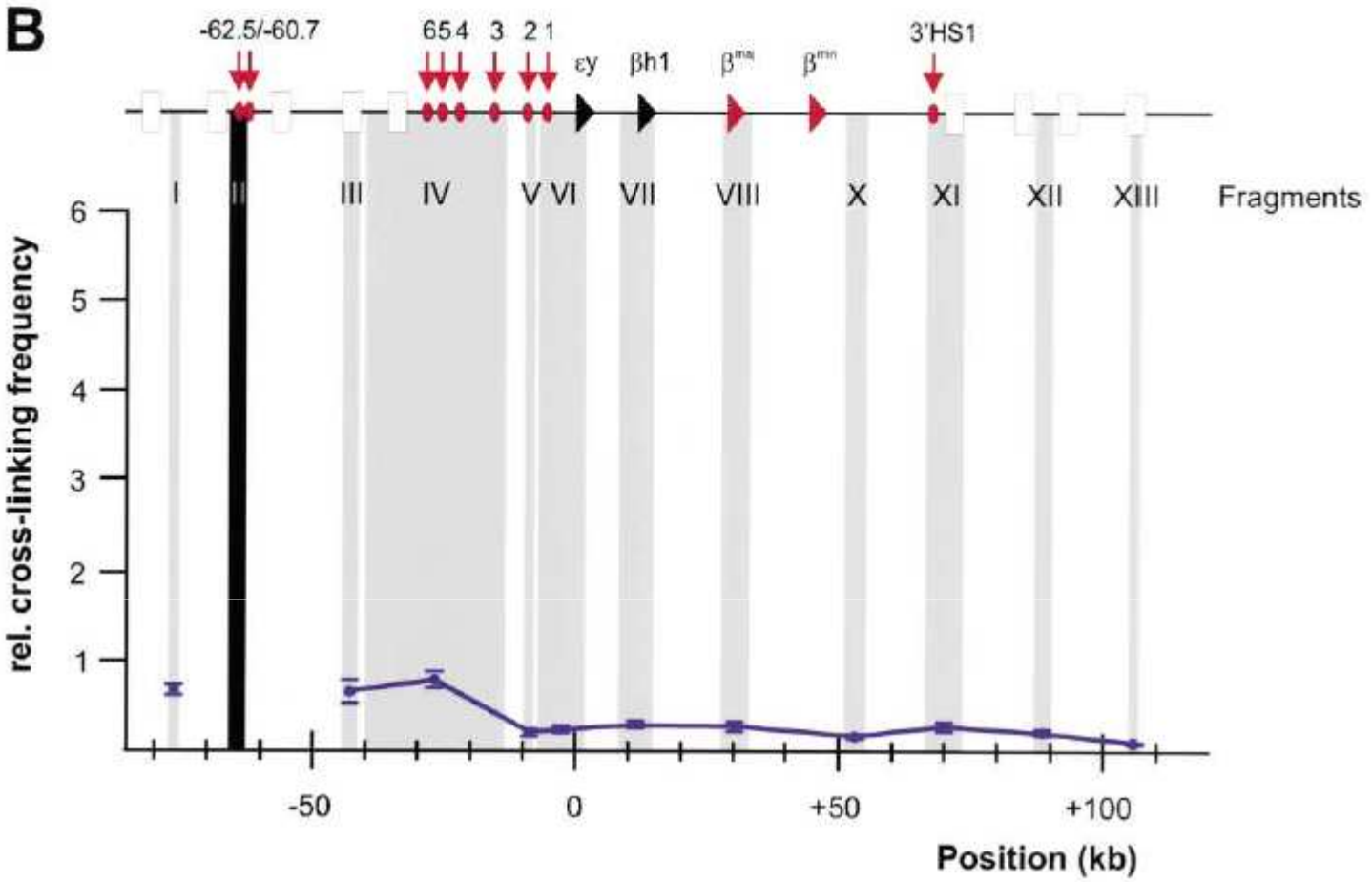
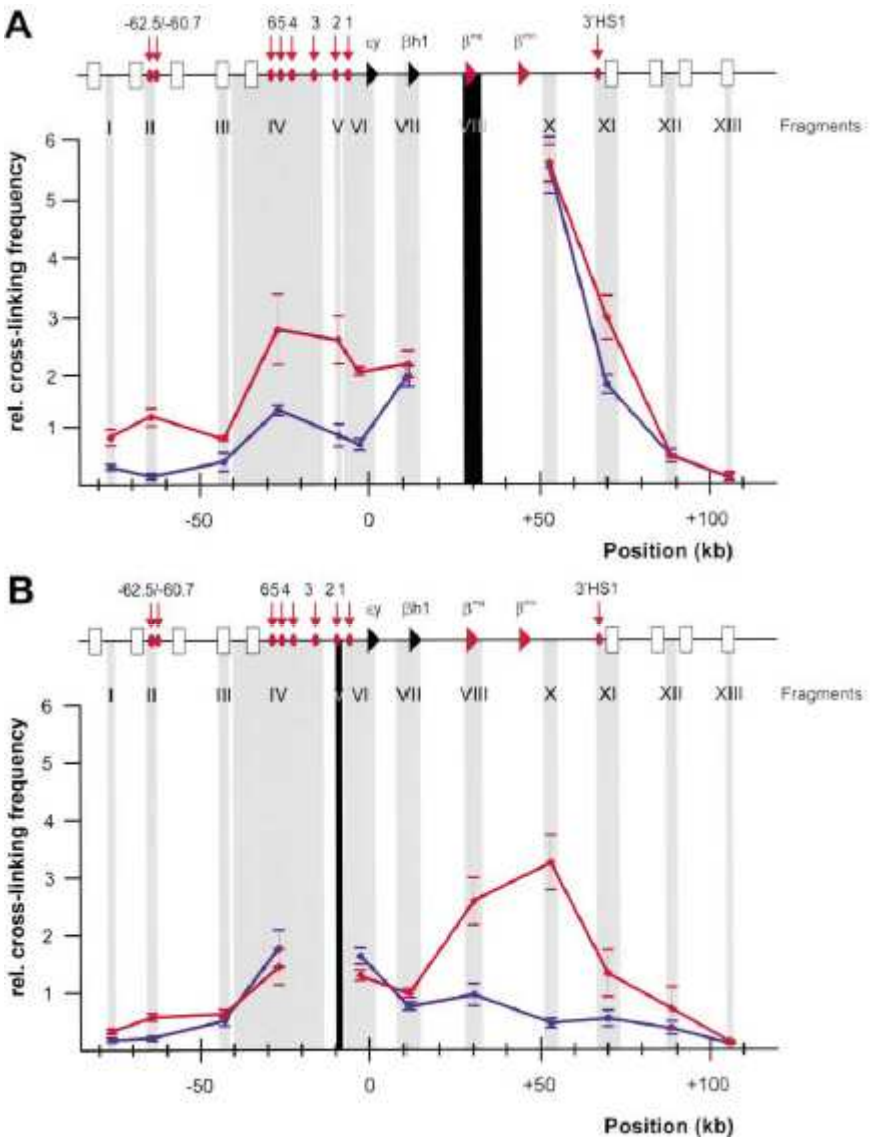


Figure 2 (**B**) Relative crosslinking frequencies between fixed BglII fragment II (5HS62.5/60.7) and the rest of the locus.



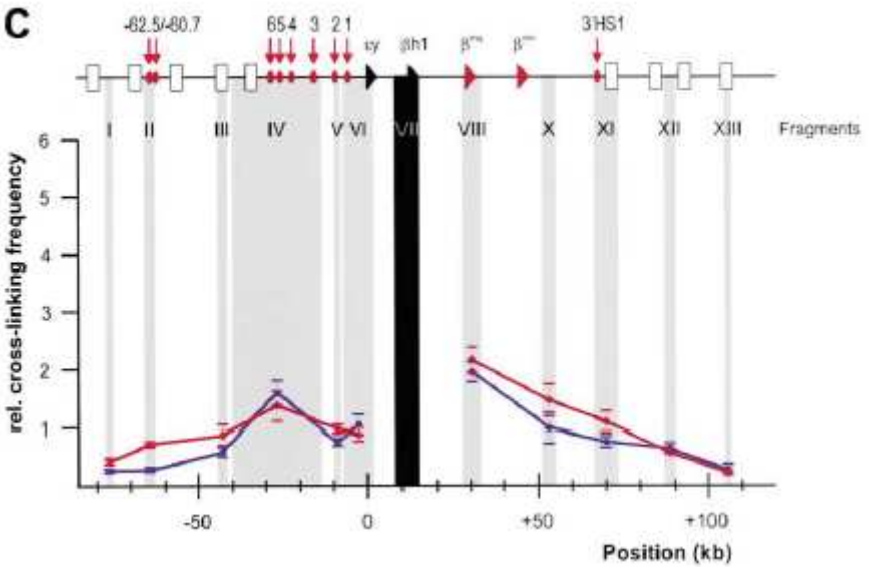


Figure 3. **Erythroid-Specific** Interaction and Looping between the LCR and an Active beta--globin Gene. Relative crosslinking frequencies observed in fetal liver are shown in red. For comparison, data obtained in brain are depicted in blue. Standard error of the mean is indicated.

Crosslinking frequency with a value of 1 arbitrarily corresponds to the crosslinking frequency between two neighboring CalR control fragments (with restriction sites analyzed being 1.5 kb apart). Scaling on the *y* axis (from 0 to 6) allows direct comparison with Figures 2 and 4–6.

(**A**) Fixed BglII fragment VIII (maj) versus the rest of the locus. (**B**) Fixed BglII fragment V (5HS2) versus the rest of the locus. (**C**) Fixed BglII fragment VII (h1) versus the rest of the locus.

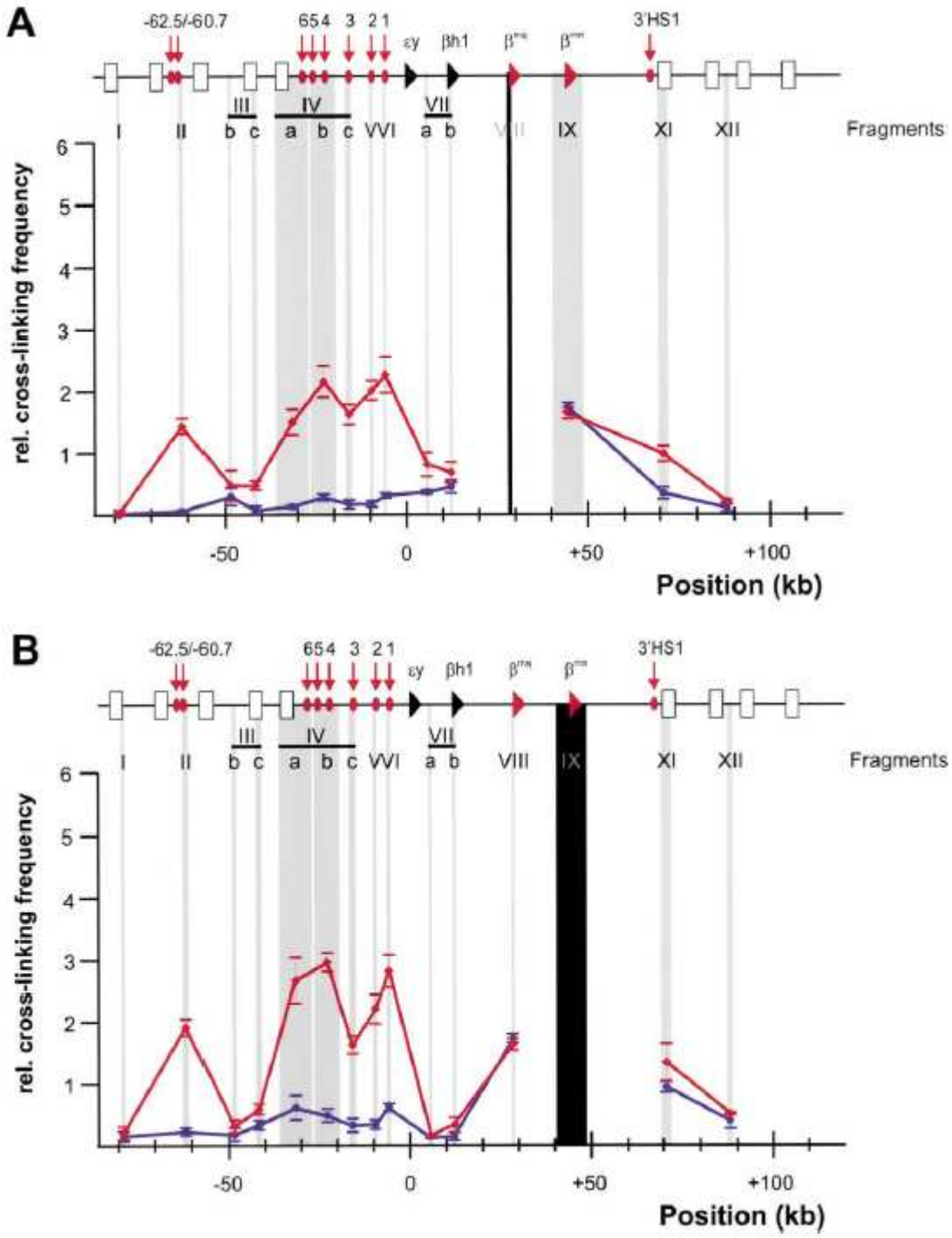


Figure 4. **Erythroid**-Specific Interactions between the Active beta-globin Genes and Individual Hypersensitive Sites in the LCR. Relative crosslinking frequencies observed in fetal liver (red) and brain (blue) are shown. Standard error of the mean is indicated. Crosslinking frequency with a value of 1 arbitrarily corresponds to the crosslinking frequency between two neighboring CalR control fragments (with restriction sites analyzed being 1.5 kb apart). Scaling on the y axis (from 0 to 6) allows direct comparison with other figures.

(A) Fixed HindIII fragment VIII Bmaj versus the rest of the locus.

(B) Fixed HindIII fragment IX (Bmin) versus the rest of the locus.

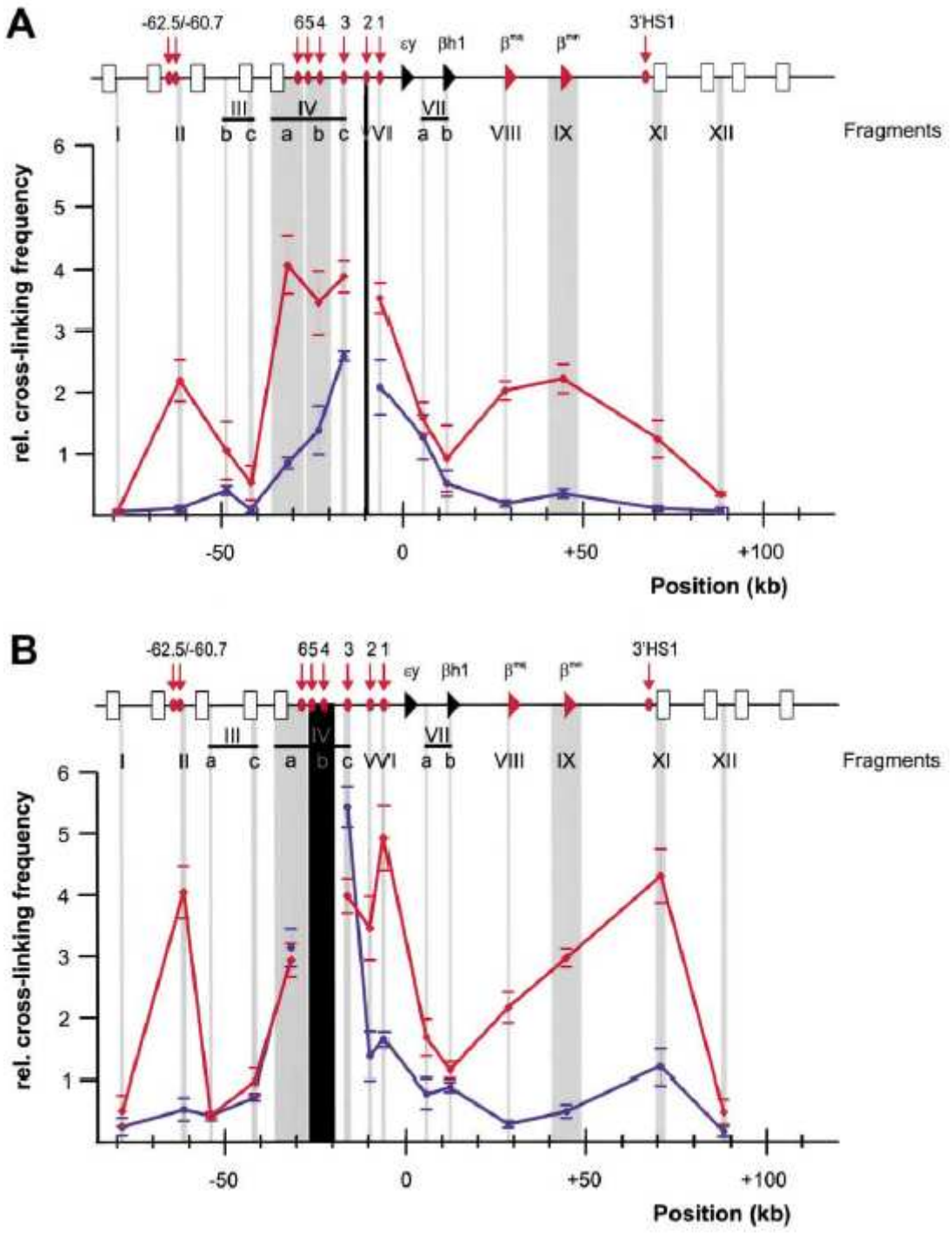


Figure 5. **Erythroid**-Specific High Crosslinking Frequencies among the Individual Hypersensitive Sites of the LCR and Two Distal Hypersensitive Sites

Relative crosslinking frequencies observed in fetal liver (red) and brain (blue) are shown. Standard error of the mean is indicated. Crosslinking frequency with a value of 1 arbitrarily corresponds to the crosslinking frequency between two neighboring CalR control fragments (with restriction sites analyzed being 1.5 kb apart). Scaling on the y axis (from 0 to 6) allows direct comparison with other figures.

(A) Fixed HindIII fragment V (5'-HS2 of the LCR) versus the rest of the locus.

(B) Fixed HindIII fragment IV-b (5'-HS4-5 of the LCR) versus the rest of the locus.

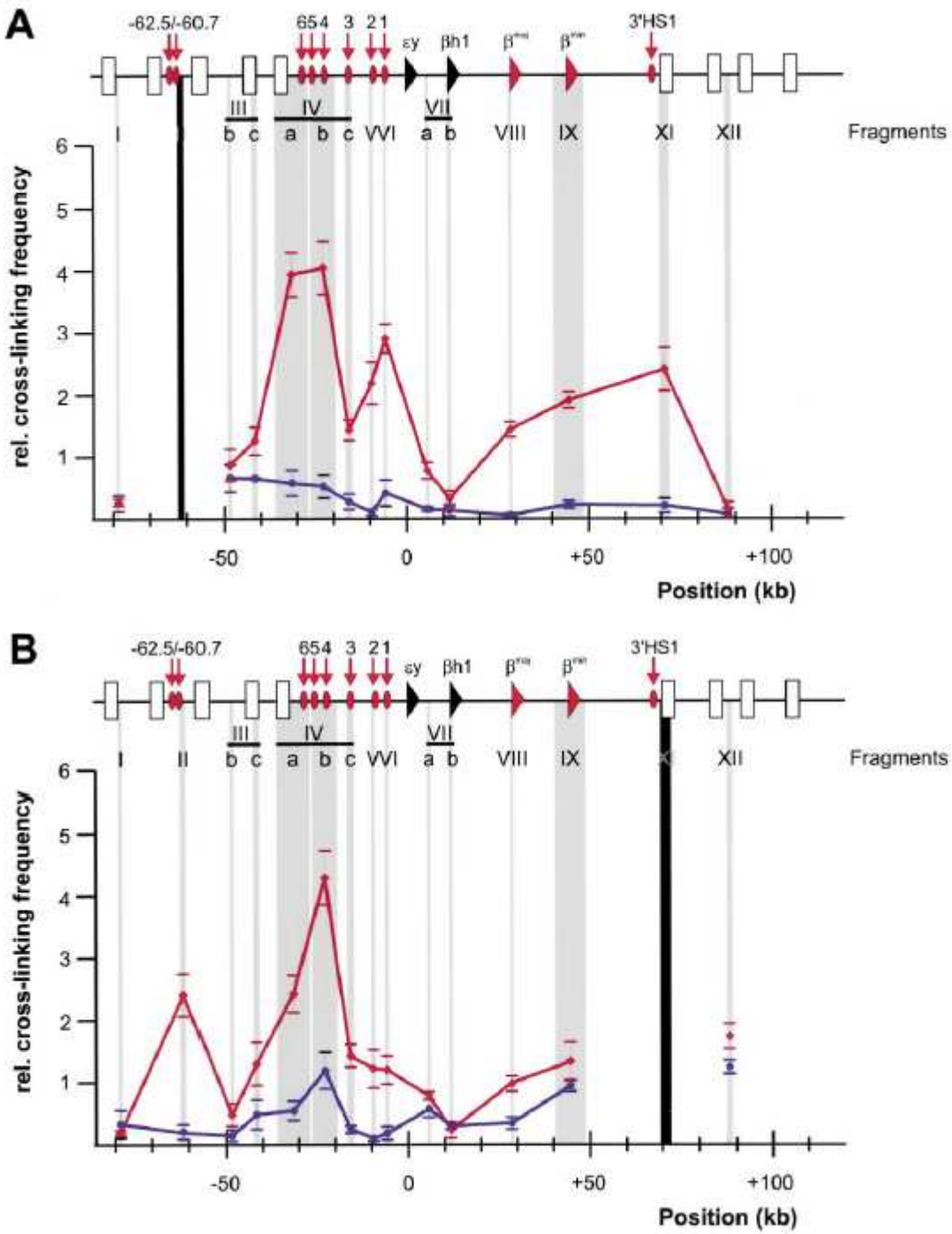


Figure 6. Two Distal Hyper-sensitive Sites at Each Side of the Locus Cluster with the LCR and the Genes Relative crosslinking frequencies observed in **fetal liver** (red) and **brain** (blue) are shown. Standard error of the mean is indicated. Crosslinking
frequency with a value of 1 arbitrarily corresponds to the crosslinking frequency between two neighboring CalR control fragments (with restriction sites analyzed being 1.5 kb apart). Scaling on the *y* axis (from 0 to 6) allows direct comparison with other figures.
(**A**) Fixed HindIII fragment II (5HS62.5/60.7) versus the rest of the locus.
(**B**) Fixed HindIII fragment XI (3HS1) versus the rest of the locus.

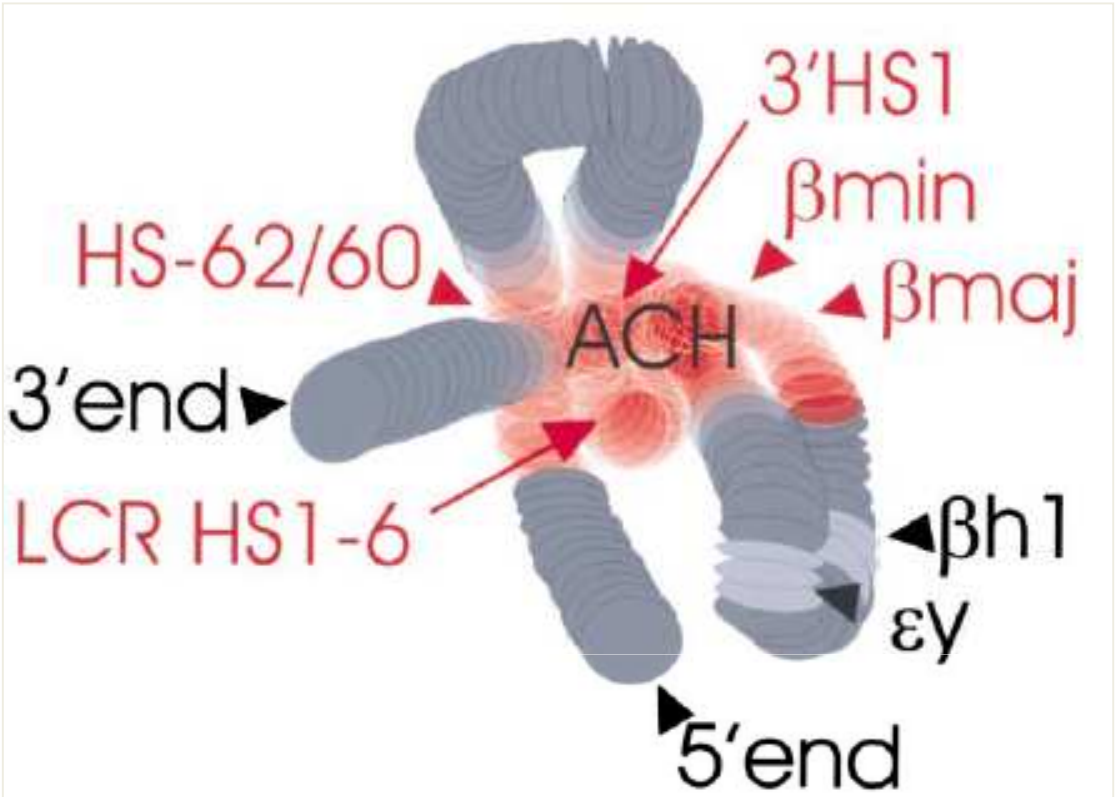


Figure 7. A 3D Model of the ACH

A hypothetical model of the active chromatin hub (ACH) is shown to illustrate the 3D nature of the ACH (not to scale), not the actual position of the elements relative to each other in vivo. Red indicates the active regions (hypersensitive sites and active genes) of the locus forming a hub of hyperaccessible chromatin (ACH). The inactive regions of the locus, having a more compact chromatin structure, are indicated in gray, with the inactive βh1 and εγ genes in lighter gray. The olfactory genes are not shown. The interactions in the ACH would be dynamic in nature, in particular with the active genes (βmaj and βmin), which are alternately transcribed.

This is quite close to what we understand of a enhancer placed at some Kb distance from the promoter

i.e. intrachromosomal interaction

but do we have long-range intrachromosomal interactions

or even inter-chromosomal interactions ?

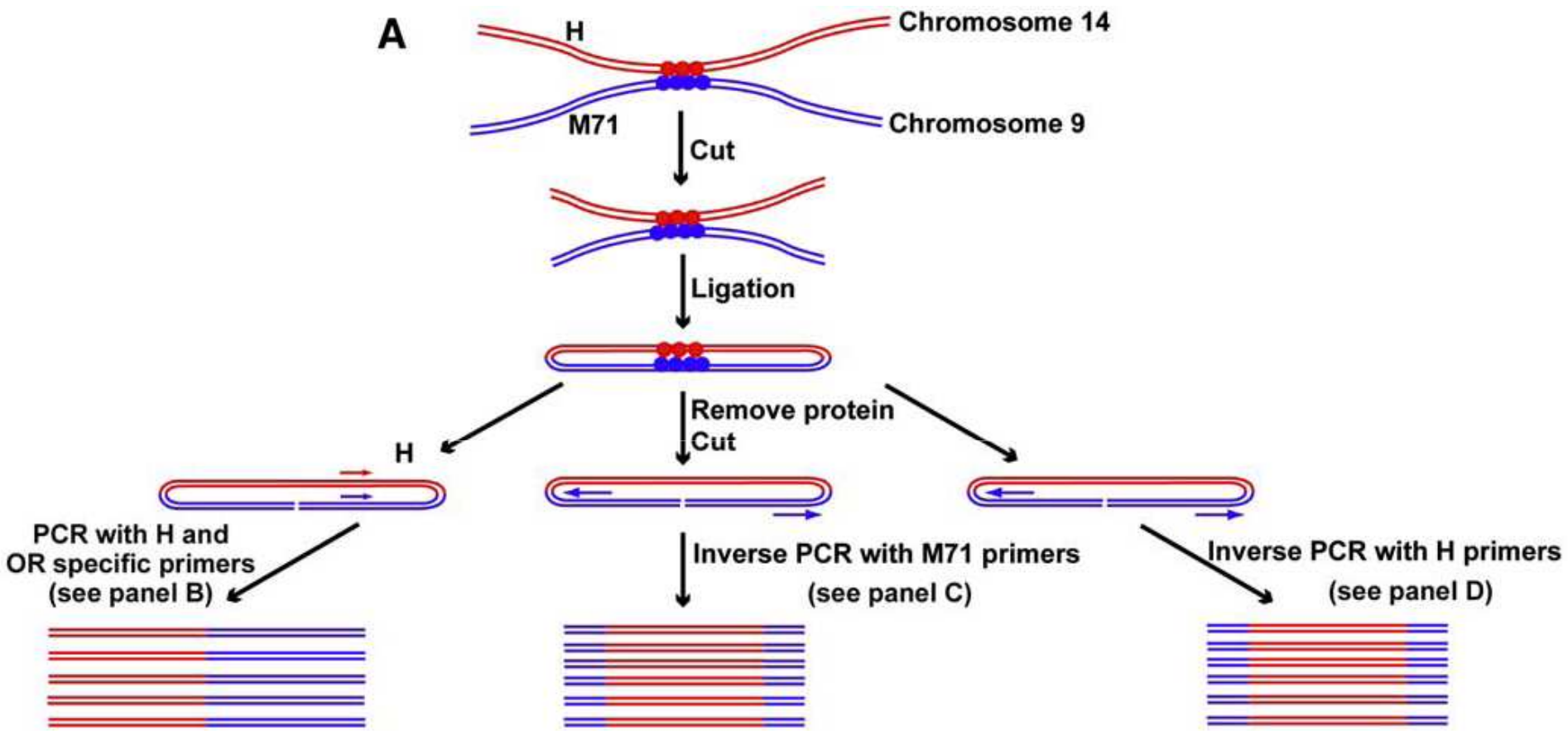
Interchromosomal Interactions and Olfactory Receptor Choice

Stavros Lomvardas,[1] Gilad Barnea,[1] David J. Pisapia,[1] Monica Mendelsohn,[1] Jennifer Kirkland,[1] and Richard Axel[1,*]
[1]Department of Biochemistry and Molecular Biophysics and Howard Hughes Medical Institute, College of Physicians and Surgeons, Columbia University, New York, NY 10032, USA
*Contact: ra27@columbia.edu
DOI 10.1016/j.cell.2006.06.035

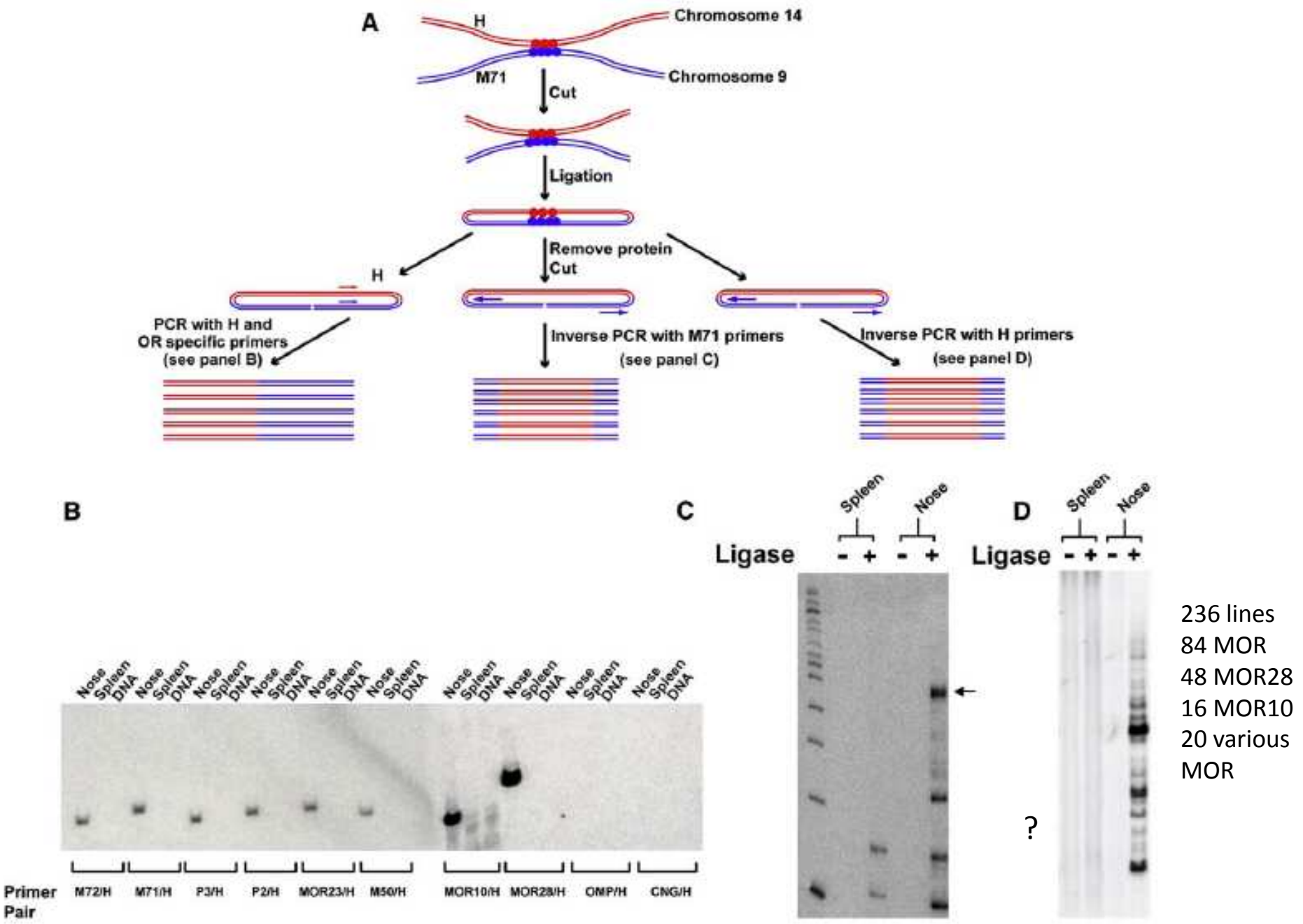
Inter-chromosomal interaction!

SUMMARY

The expression of a single odorant receptor (OR) gene from a large gene family in individual sensory neurons is an essential feature of the organization and function of the olfactory system. We have used chromosome conformation capture to demonstrate the specific association of an enhancer element, *H*, on chromosome 14 with multiple OR gene promoters on different chromosomes. DNA and RNA fluorescence in situ hybridization (FISH) experiments allow us to visualize the colocalization of the *H* enhancer with the single OR allele that is transcribed in a sensory neuron. In transgenic mice bearing additional *H* elements, sensory neurons that express OR pseudogenes also express a second functional receptor. These data suggest a model of receptor choice in which a single *trans*-acting enhancer element may allow the stochastic activation of only one OR allele in an olfactory sensory neuron.


A
H
Chromosome 14
M71
Chromosome 9
Cut
Ligation
Remove protein
Cut
H
PCR with H and
OR specific primers
(see panel B)
Inverse PCR with M71 primers
(see panel C)
Inverse PCR with H primers
(see panel D)

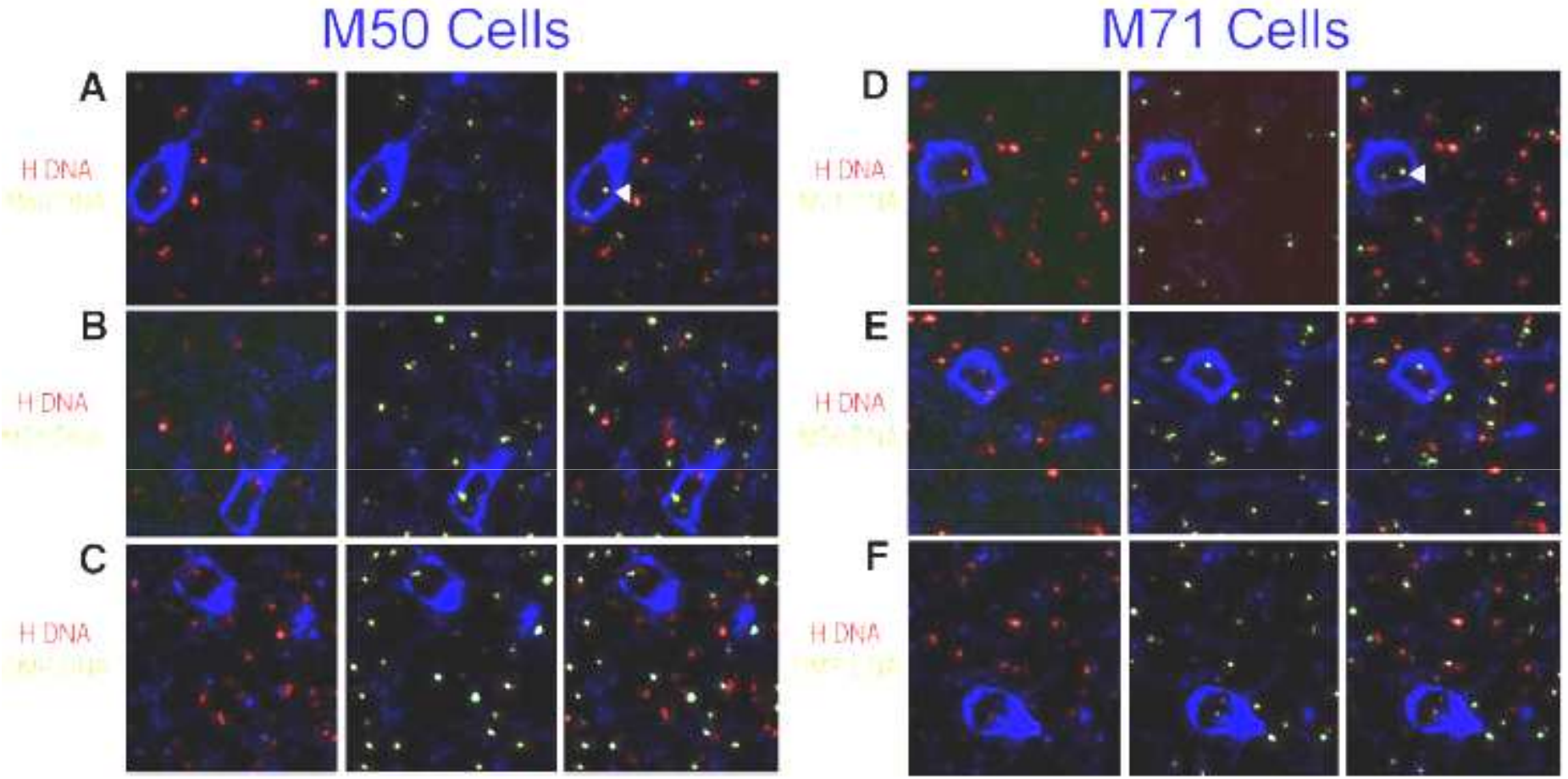
A 3C scheme


A
H
Chromosome 14
M71
Chromosome 9
Cut
Ligation
Remove protein
Cut
H
PCR with H and
OR specific primers
(see panel B)
Inverse PCR with M71 primers
(see panel C)
Inverse PCR with H primers
(see panel D)
B
Nose
Spleen
DNA
Primer
Pair
M72/H
M71/H
P3/H
P2/H
MOR23/H
M50/H
MOR10/H
MOR28/H
OMP/H
CNG/H
C
Ligase
Spleen
Nose
D
Ligase
Spleen
Nose
?
236 lines
84 MOR
48 MOR28
16 MOR10
20 various
MOR

48 MOR28 and 16 MOR10

MOR28 and MOR1o are close to H on chr 14

and are expressed 10 times more frequently than other MOR genes

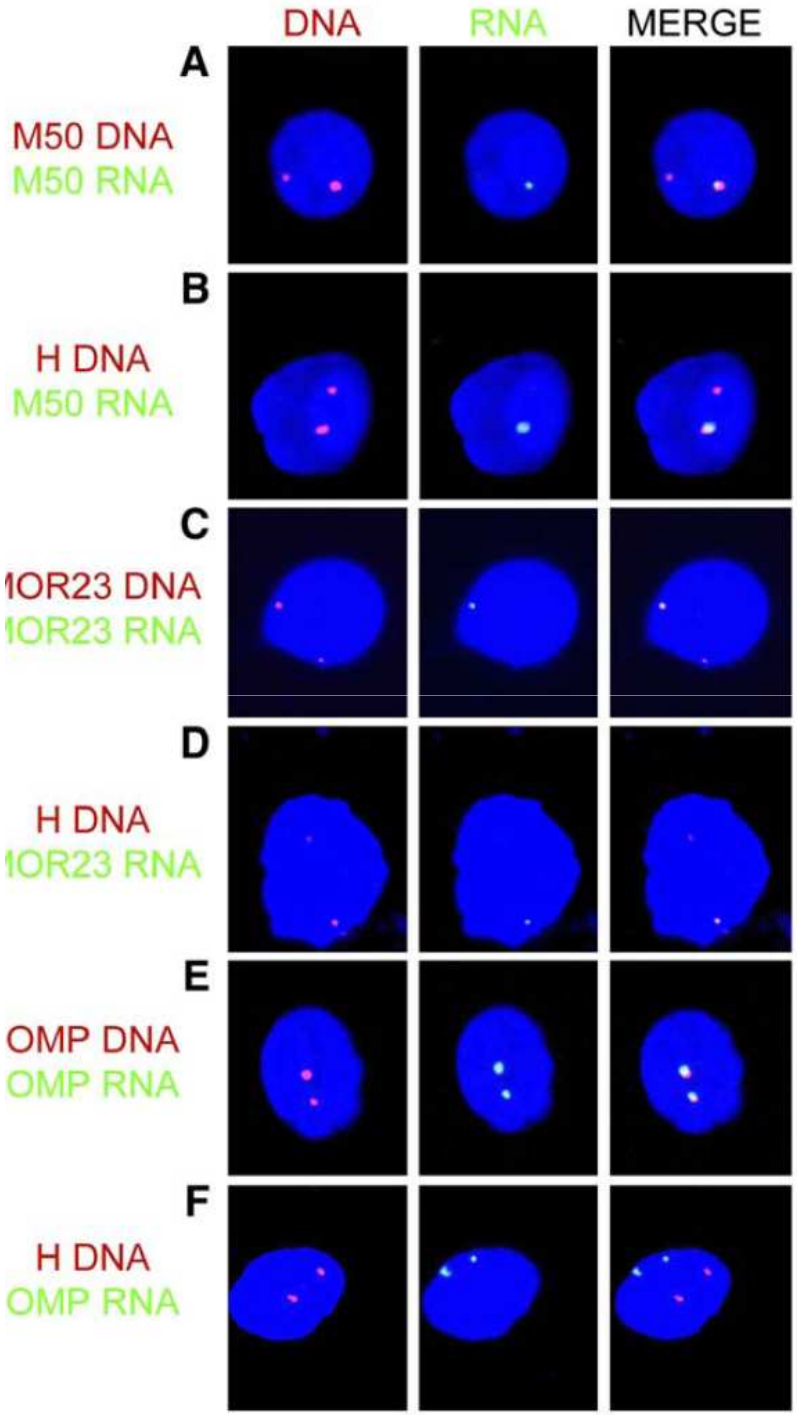


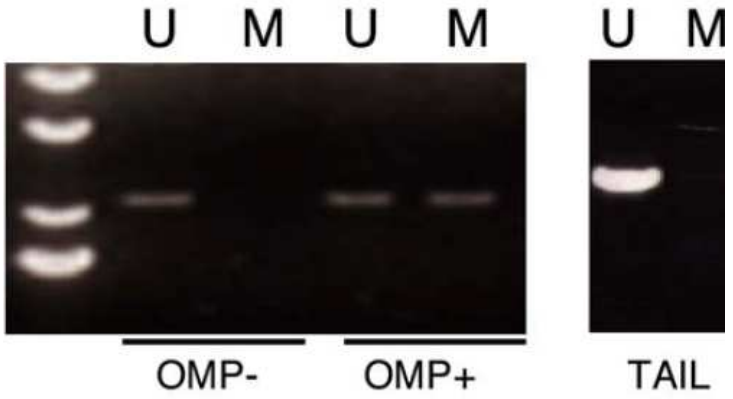
FISH in sections of the olfactory epithelium, stained with Ab-M71

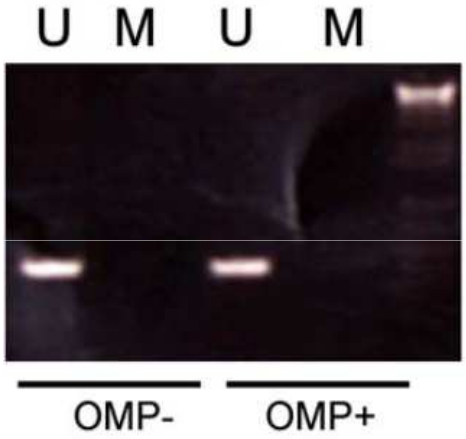
Table 1. Frequency of Colocalization of *H* with Different Loci in *M50*- or *M71*-Expressing Cells

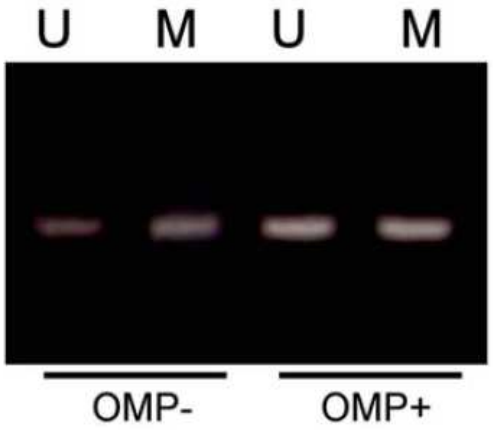
	M50 Cells	*M71* Cells
H/M50	103/356 (29%)	10/275 (3.6%)
H/M71	5/203 (2.4%)	90/312 (28.8%)
H/OMP	5/305 (1.6%)	9/278 (3.2%)

RNA FISH demonstrates transcription from **only one** allele









One H enhancer allele is indeed inactivated by DNA methylation.
As assayed by allele-specific PCR after bisulfite conversion

Enhancing nuclear receptor-induced transcription requires nuclear motor and LSD1-dependent gene networking in interchromatin granules

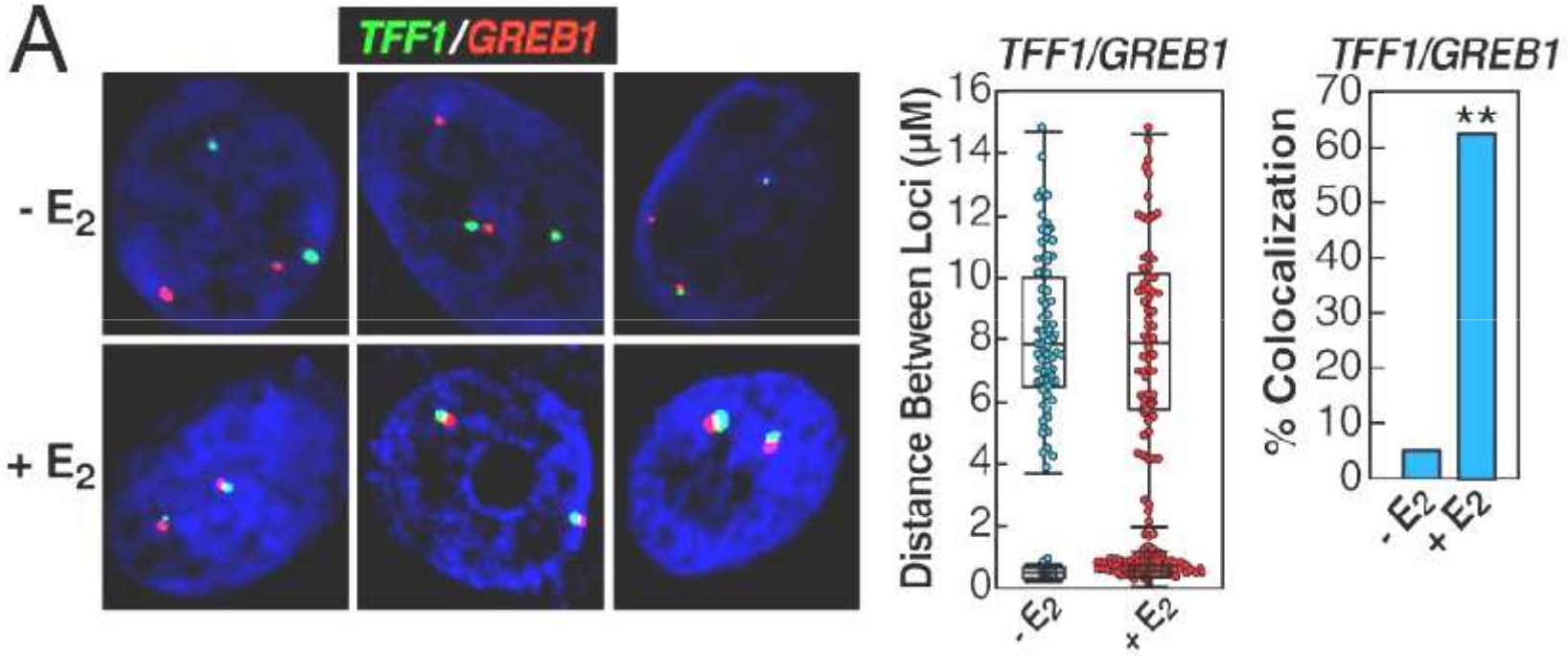
Qidong Hu[a,1], Young-Soo Kwon[b,1], Esperanza Nunez[a,c,1], Maria Dafne Cardamone[a,d], Kasey R. Hutt[a,e], Kenneth A. Ohgi[a], Ivan Garcia-Bassets[a], David W. Rose[f], Christopher K. Glass[b], Michael G. Rosenfeld[a,2], and Xiang-Dong Fu[b,2]

[a]Department of Medicine, Howard Hughes Medical Institute, [b]Department of Cellular and Molecular Medicine, [c]Biomedical Sciences Graduate Program, [e]Bioinformatics Graduate Program, [f]Department of Medicine, Division of Endocrinology and Metabolism, University of California at San Diego School of Medicine, La Jolla, CA 92093; and [d]Department of Oncological Sciences, University of Turin, 10060 Turin, Italy

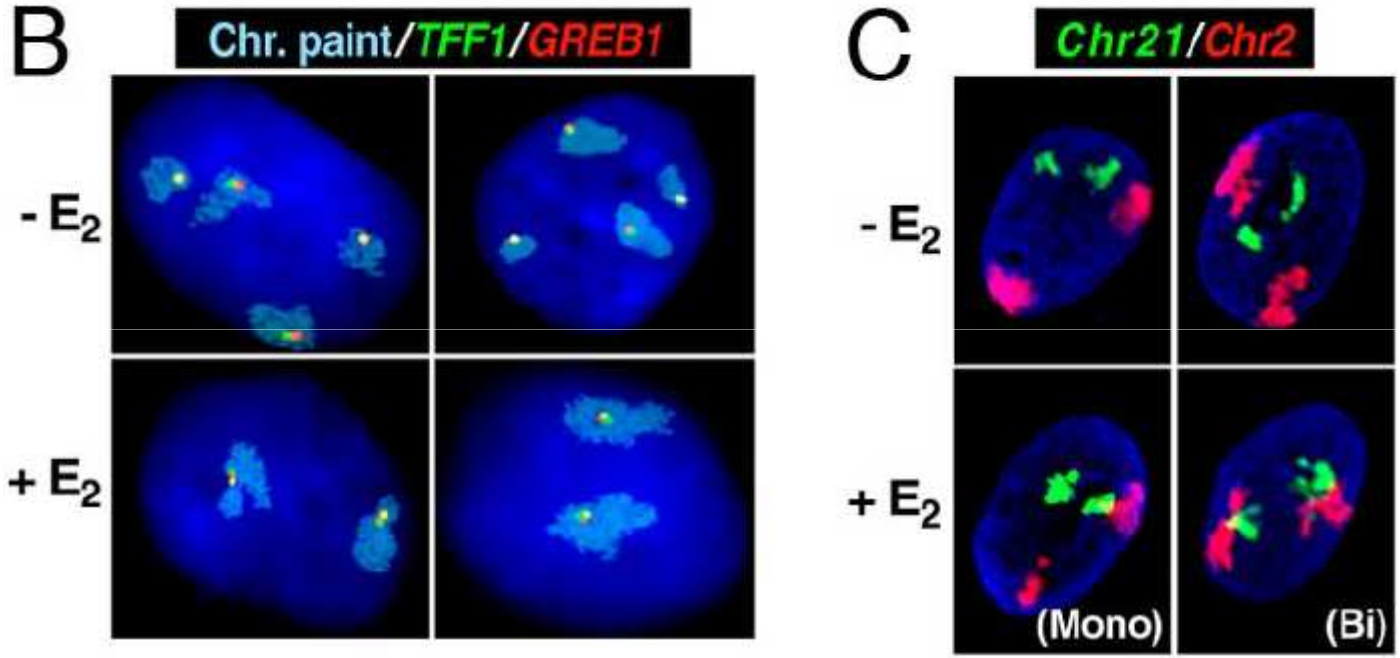
Contributed by Michael G. Rosenfeld, October 22, 2008 (sent for review October 11, 2008)

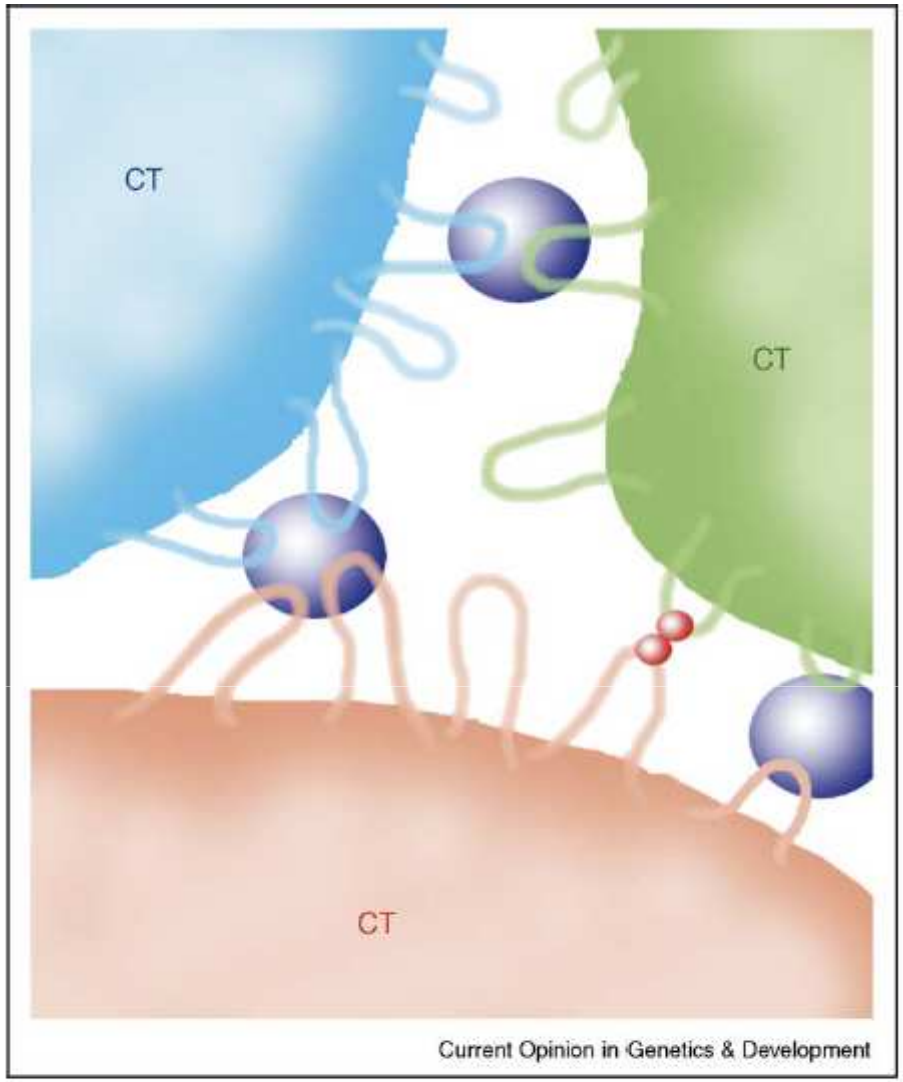
Although the role of liganded nuclear receptors in mediating coactivator/corepressor exchange is well-established, little is known about the potential regulation of chromosomal organization in the 3-dimensional space of the nucleus in achieving integrated transcriptional responses to diverse signaling events. Here, we report that ligand induces rapid interchromosomal interactions among specific subsets of estrogen receptor α-bound transcription units, with a dramatic reorganization of nuclear territories, which depends on the actions of nuclear actin/myosin-I machinery and dynein light chain 1. The histone lysine demethylase, LSD1, is required for these ligand-induced interactive loci to associate with distinct interchromatin granules, long thought to serve as ''storage'' sites for the splicing machinery, some critical transcription elongation factors, and various chromatin remodeling complexes. We demonstrate that this 2-step nuclear rearrangement is essential for achieving enhanced, coordinated transcription of nuclear receptor target genes.

After E2 stimulus, NCF7 cells were subjected to FISH with probes for TFF1 (green) and for GREB1 (red), two genes that are stimulated by estrogen



These genes are brought in close contact in response to the stimulus.


B
Chr. paint/TFF1/GREB1
- E2
+ E2
C
Chr21/Chr2
- E2
+ E2
(Mono)
(Bi)



Nuclear organization. Three different chromosome territories (CTs) are shown. Potentially active genes in *cis* and *trans* dynamically engage transcription factories (blue spheres). Most 'active' genes spend the majority of their time outside transcription factories and are transcriptionally inactive. Other types of regulatory interactions (red spheres) might possibly sequester genes or inhibit mobility and factory contact.

The 4C technique: the 'Rosetta stone' for genome biology in 3D?

Rolf Ohlsson and Anita Göndör

Despite considerable efforts, the spatial link between the nuclear architecture and the genome remains enigmatic. The 4C method, independently innovated in four different laboratories, might in combination with other methods change that. As this method is based on the unbiased identification of sequences interacting with specific baits, there are unique opportunities for unravelling the secrets of how the genome functions in 3D.

Addresses
Department of Development & Genetics, Norbyvägen 18A, Evolution Biology Centre, Uppsala University, S-752 36 Uppsala, Sweden

Corresponding author: Ohlsson, Rolf (rolf.ohlsson@ebc.uu.se) and

Current Opinion in Cell Biology 2007, **19**:321–325

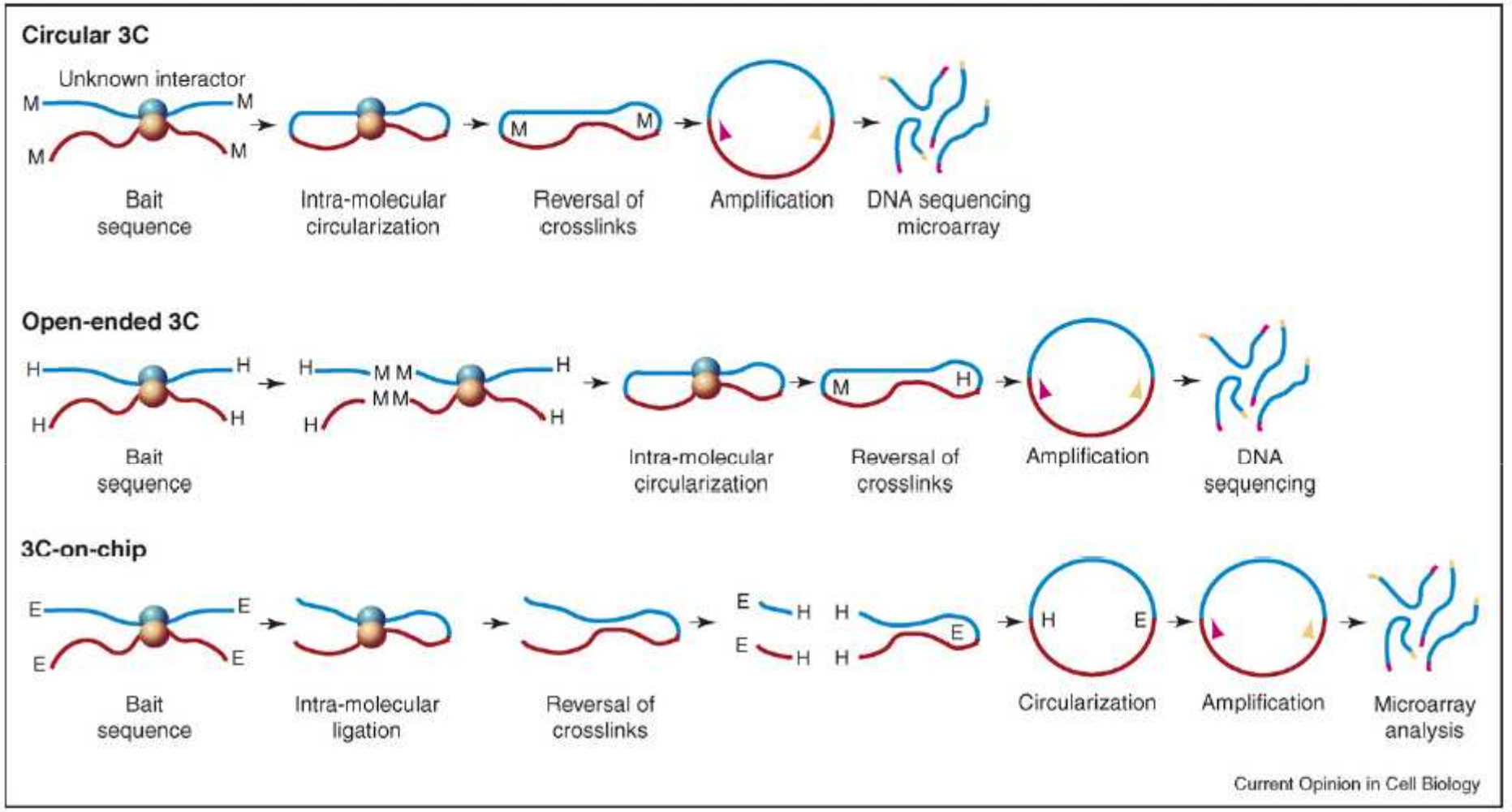
also study unknown interac
Although an interesting optio
to be tested in genome-wide sc
failed to document numerous
actions that both 3C [8] an
documented, it clearly has li
with the 4C method.

Here we review the 4C tech
resolve some of the old and ne
the genome function in the 3

The 4C technique: oppor

All the 4C methods have on
generation of circular DNA m

Figure 1



A schematic comparison of the 4C approaches. All three methods shown use a circular DNA as a vehicle to identify unbiased interactors by means of inverse PCR primers. The 3C-on-chip generates the circular DNA structure after the reversal of crosslinking step, whereas the other methods generate circular DNA while the interacting sequences remain in a crosslinked DNA/protein complex. The methods also differ with respect to the use of restriction enzymes (M, *Msp* I; E, *Eco* RI; H, *Hind* III). A fourth 4C method, 'olfactory receptor' 3C [12•], did not specify the restriction enzymes used, but appears to be most closely related to the circular 3C principle.