La trascrizione è **regolata**.

Le regolazioni possono essere di diverso tipo:

✓Macroscopica: eterocromatina ed eucromatina (silencing)

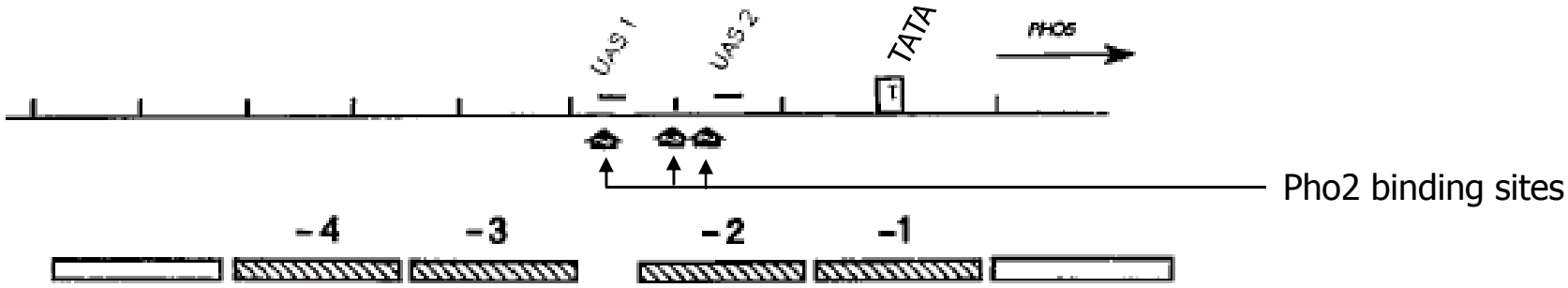
✓Individuale: singoli geni sono attivati o repressi

✓Continua: singoli geni vengono trascritti più o meno (feedback) (modulated genes)

✓Dipendente: singoli geni sono attivati o repressi dauno stimolo (on-off genes).

✓Differenziamento-dipendente: singoli geni o gruppi di geni sono attivati o repressi in funzione del progressivo differenziamento cellulare.

✓Allele-specifica: in alcuni casi, soltanto uno dei due alleli viene utilizzato.



Pho4 is the P-sensitive inducer, whereas Pho2 is constitutive

yeast

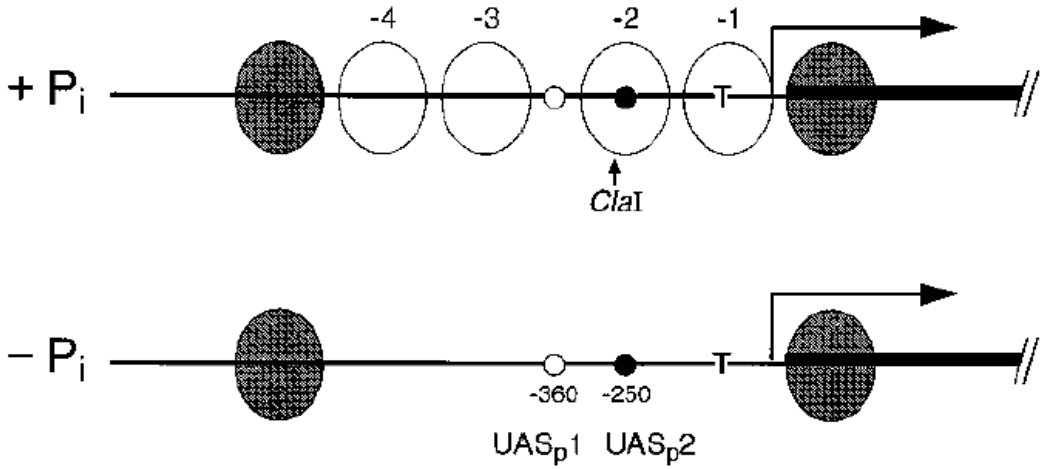
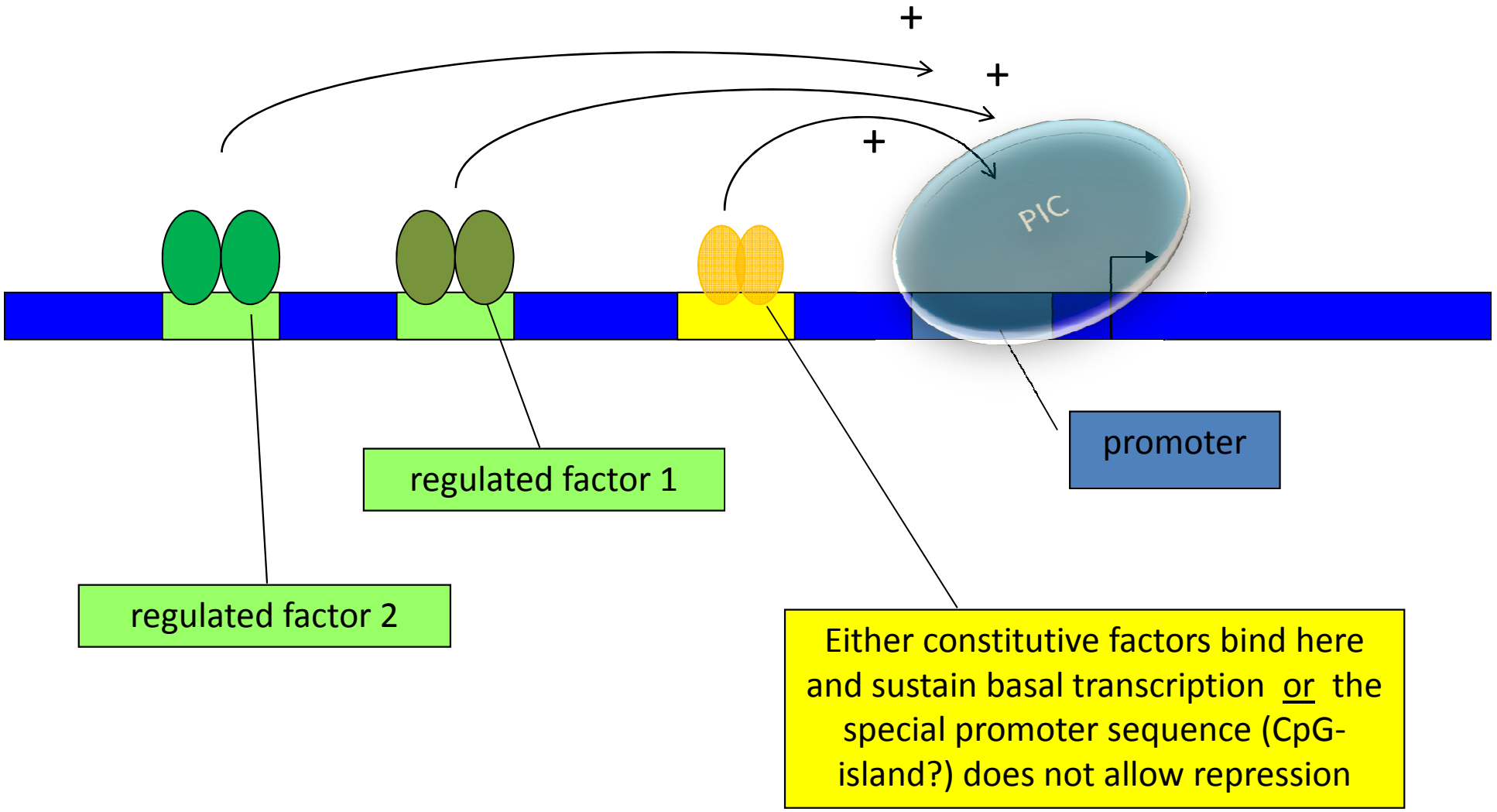


Figure 1. Chromatin Structure at the *PHO5* Promoter

Nucleosomes −1, −2, −3, and −4 are remodeled upon activation of the promoter by phosphate starvation conditions (Almer et al., 1986). The small circles mark UASp1 (open) and UASp2 (solid), which are Pho4-binding sites found by in vitro (Vogel et al., 1989) and in vivo (Venter et al., 1994) footprinting experiments. The positions are listed relative to the coding sequence (solid bar). T denotes the TATA box (Rudolph and Hinnen, 1987). The location of a ClaI site at −275 relative to the coding region is shown.

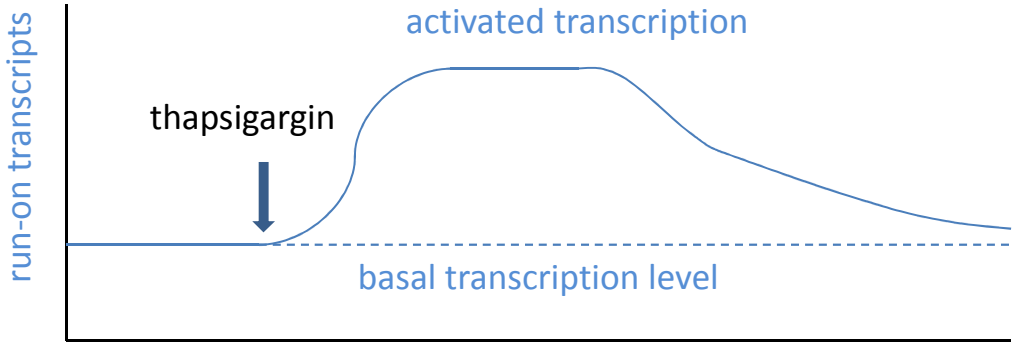
Modulated genes

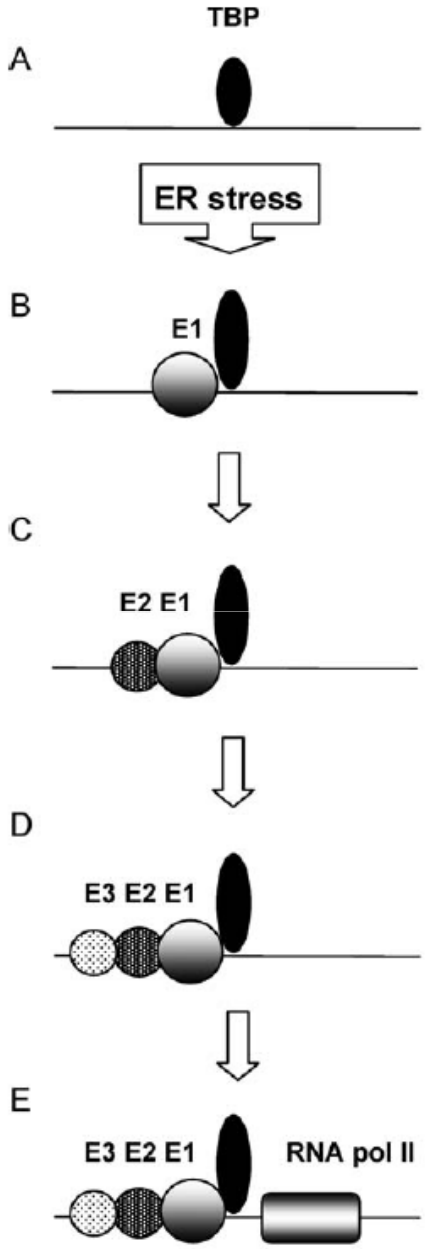


Constitutive Nucleosome Depletion and Ordered Factor Assembly at the *GRP78* Promoter Revealed by Single Molecule Footprinting

Einav Nili Gal-Yam[1,2], Shinwu Jeong[1,2], Amos Tanay[3], Gerda Egger[1,2], Amy S. Lee[2], Peter A. Jones[1,2*]

Authors have studied nucleosome occupancy at the GRP78 promoter after activation by stress.
Note that this is an Housekeeping gene that is always basally activated (A in figure) and transcribed at low level. TBP and basal factors are alway bound at the promoter, and rare alleles show the presence of RNA Pol II molecules traveling on the gene. The promoter is always nucleosome-free.
After induction by thapsigargin, a progressively more extended region gets protected from methylation in vitro. These are activating factors binding to regulatory elements (B, C, D). After awhile, a consistent RNA Pol II footprint is seen in the transcribed region, testifying frequent initiation and increased transcription (E).
This "activated status" is quite stable even when factors are no longer present and is self-sustaining.





As a rule these promoters are often TATA-less, CpG-rich

DNA methylation was measured gene-by-gene in the **past**, using bisulfite conversion, followed by PCR, cloning and sequencing

DNA methylation can **today** be assayed on a genome-wide level, by essentially three methods:

1) methyl-C-DNA immunoprecipitation, followed by promoter arrays, tiling arrays or deep-sequencing

2) bisulfite conversion followed by microarrays containg probes for both unconverted and converted oligos.

3) bisulfite followed by direct re-sequencing (NGS).

nature genetics

ASSIGNED PAPER

Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome

Michael Weber[1], Ines Hellmann[2,3], Michael B Stadler[1], Liliana Ramos[4], Svante Pääbo[2], Michael Rebhan[1] & Dirk Schübeler[1]

To gain insight into the function of DNA methylation at *cis*-regulatory regions and its impact on gene expression, we measured methylation, RNA polymerase occupancy and histone modifications at 16,000 promoters in primary human somatic and germline cells. We find CpG-poor promoters hypermethylated in somatic cells, which does not preclude their activity. This methylation is present in male gametes and results in evolutionary loss of CpG dinucleotides, as measured by divergence between humans and primates. In contrast, strong CpG island promoters are mostly unmethylated, even when inactive. Weak CpG island promoters are distinct, as they are preferential targets for *de novo* methylation in somatic cells. Notably, most germline-specific genes are methylated in somatic cells, suggesting additional functional selection. These results show that promoter sequence and gene function are major predictors of promoter methylation states. Moreover, we observe that inactive unmethylated CpG island promoters show elevated levels of dimethylation of Lys4 of histone H3, suggesting that this chromatin mark may protect DNA from methylation.

Conclusions.

Methylation sufficient to inactivate CpG promoter but not necessary, in fact most inactive CpG prooters are unmethylated

Repression by methyl-CpG requires high CpG density, in fact most active LCP are hypermethylated

Discussion:

Bias to promoters (known promoters)

Measurements not quantitative

low resolution

Vol 462 | 19 November 2009 | doi:10.1038/nature08514 nature

Human DNA methylomes at base resolution show widespread epigenomic differences

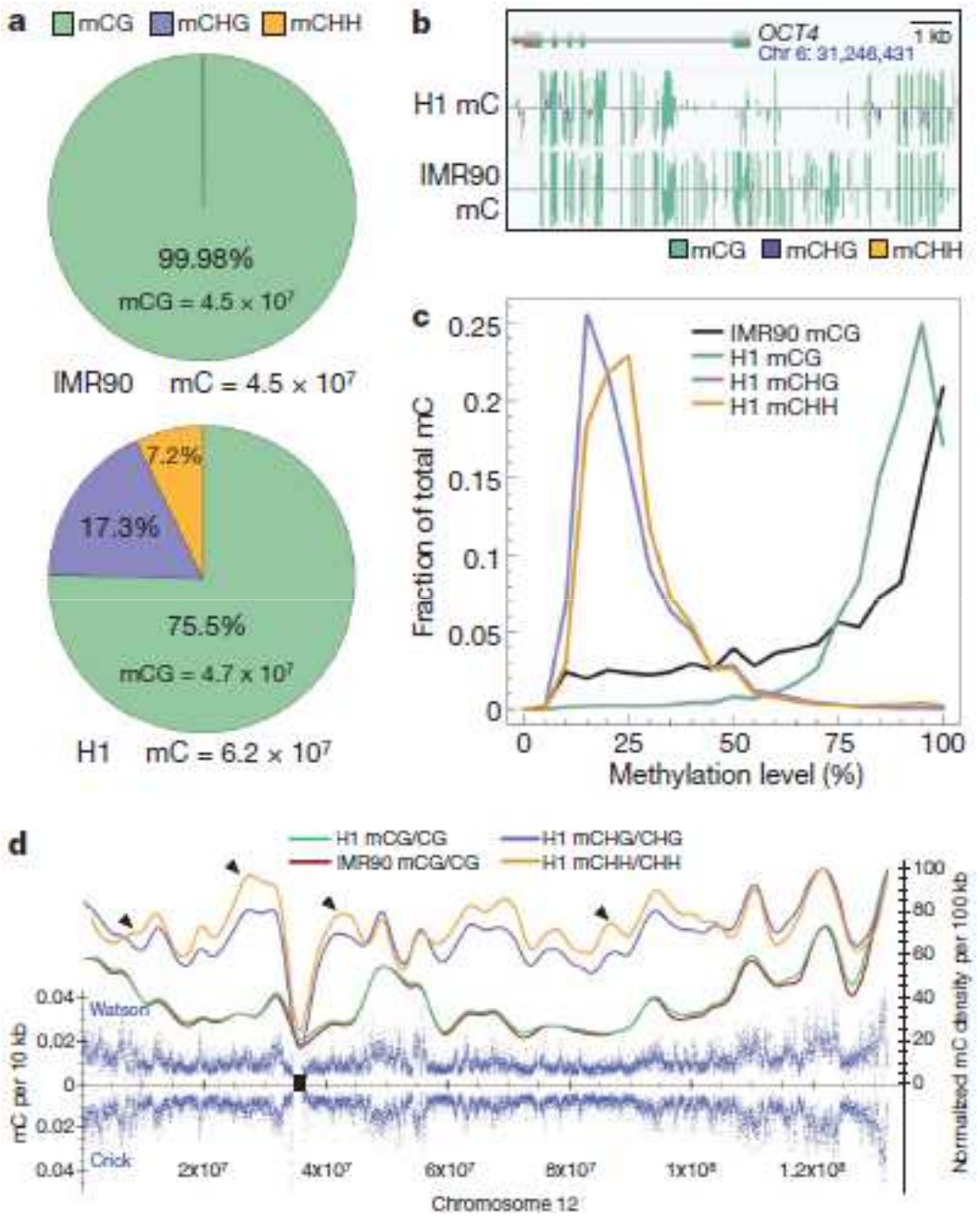
Ryan Lister[1]*, Mattia Pelizzola[1]*, Robert H. Dowen[1], R. David Hawkins[2], Gary Hon[2], Julian Tonti-Filippini[4], Joseph R. Nery[1], Leonard Lee[2], Zhen Ye[2], Que-Minh Ngo[2], Lee Edsall[2], Jessica Antosiewicz-Bourget[5,6], Ron Stewart[5,6], Victor Ruotti[5,6], A. Harvey Millar[4], James A. Thomson[5,6,7,8], Bing Ren[2,3] & Joseph R. Ecker[1]

DNA cytosine methylation is a central epigenetic modification that has essential roles in cellular processes including genome regulation, development and disease. Here we present the first genome-wide, single-base-resolution maps of methylated cytosines in a mammalian genome, from both human embryonic stem cells and fetal fibroblasts, along with comparative analysis of messenger RNA and small RNA components of the transcriptome, several histone modifications, and sites of DNA–protein interaction for several key regulatory factors. Widespread differences were identified in the composition and patterning of cytosine methylation between the two genomes. Nearly one-quarter of all methylation identified in embryonic stem cells was in a non-CG context, suggesting that embryonic stem cells may use different methylation mechanisms to affect gene regulation. Methylation in non-CG contexts showed enrichment in gene bodies and depletion in protein binding sites and enhancers. Non-CG methylation disappeared upon induced differentiation of the embryonic stem cells, and was restored in induced pluripotent stem cells. We identified hundreds of differentially methylated regions proximal to genes involved in pluripotency and differentiation, and widespread reduced methylation levels in fibroblasts associated with lower transcriptional activity. These reference epigenomes provide a foundation for future studies exploring this key epigenetic modification in human disease and development.

pdf

H1 are human embryonic stem cells

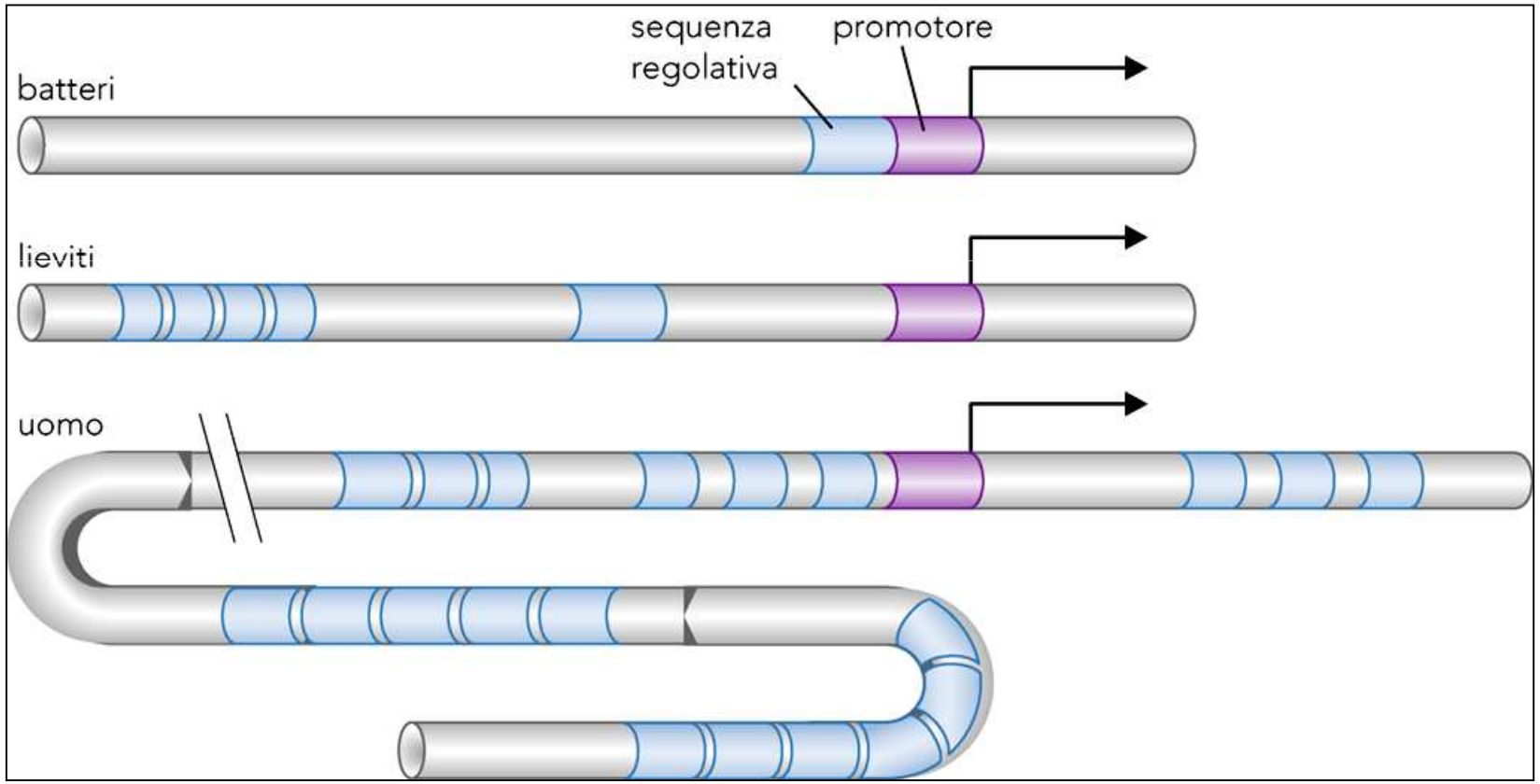
IMR90 are fetal lung fibroblasts

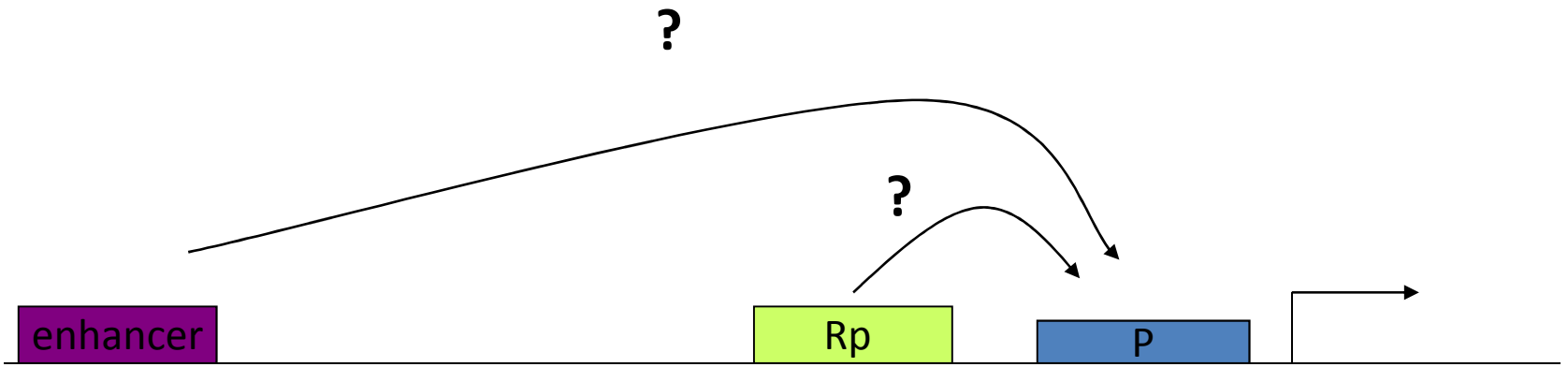


Elementi di regolazione: dove quanti e come
combinatorialità
Enhancers e promoters
Come si studiano (reporter-in vitro transcription –band shift)
I fattori di trascrizione regolatori
famiglie strutture
combinatorialità degli elementi di regolazione – specificità
Sviluppo di Drosophila – paradigma degli enhancers tessuto-specifici

è vero che gli enhancers interagiscono con i promotori?

L’aumento di complessità dei genomi è parallelo all’aumento delle regioni regolatrici in rapporto alle regioni codificanti




?
?
enhancer
Rp
P

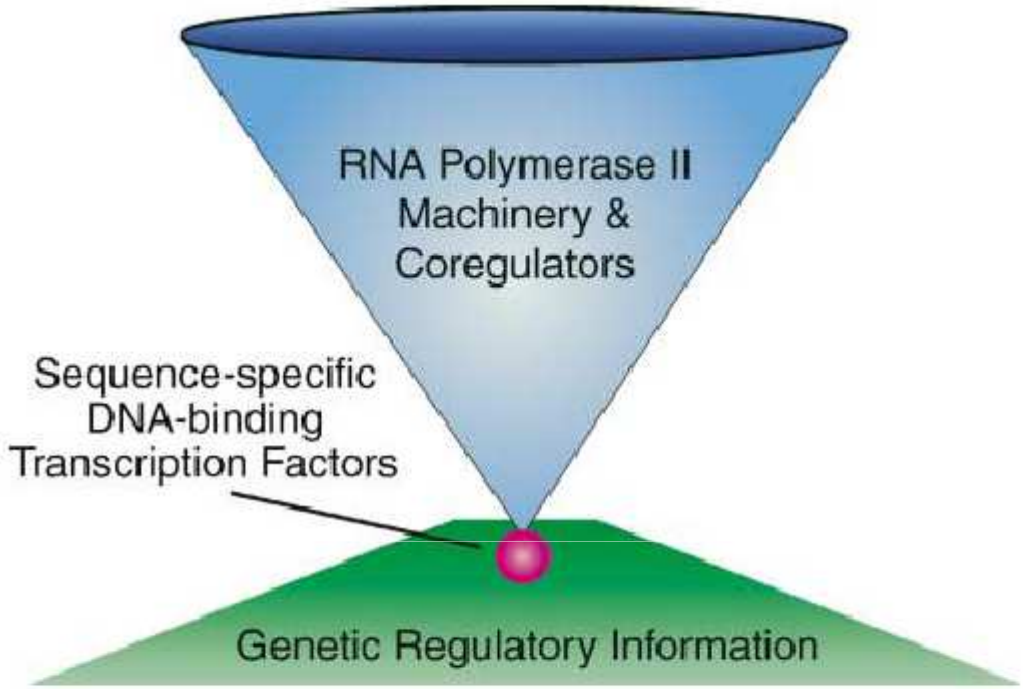
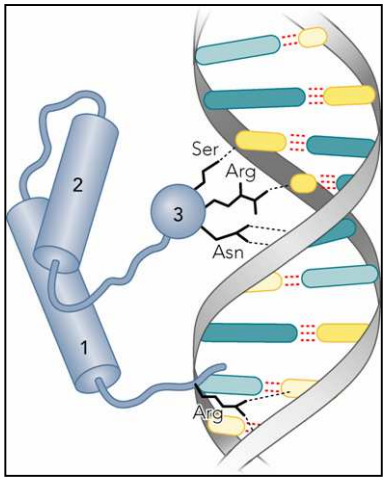
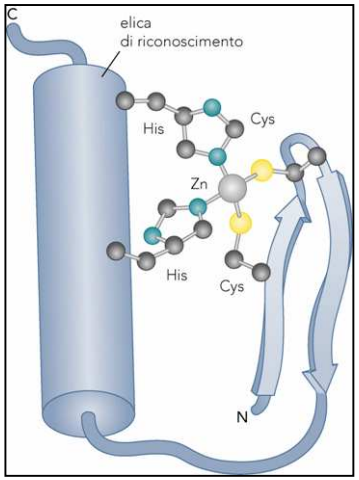
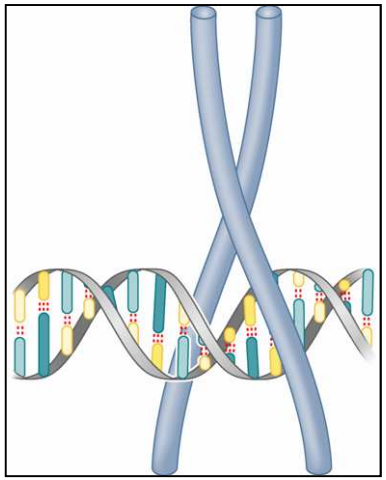


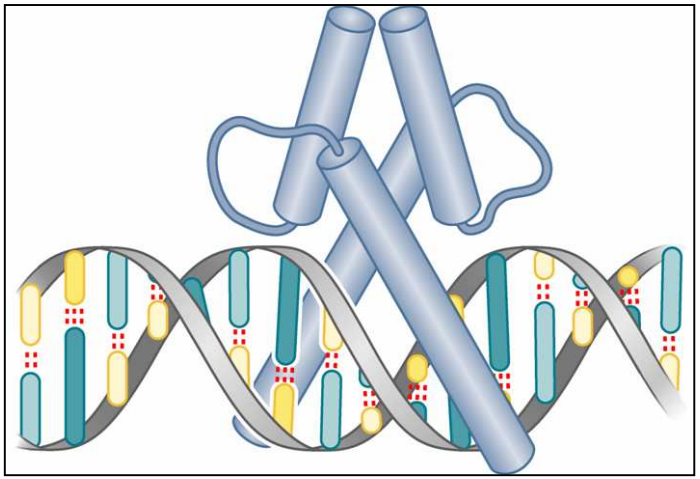
Figure 1. Sequence-Specific DNA Binding Transcription Factors Interpret and Transmit Genetic Regulatory Information

In this diagram, sequence-specific factors are depicted as the apex at the interface of the vast array of genetic regulatory information and the inverted cone of the RNA polymerase II transcriptional machinery and coregulators.









Examples:

Hox proteins	GAL4	GCN4	Myc
Antennapedia	Steroid receptors	fos-jun (AP-1)	Myo-D, Neuro-D
Matα	Nuclear receptors	CREB	SREBP

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

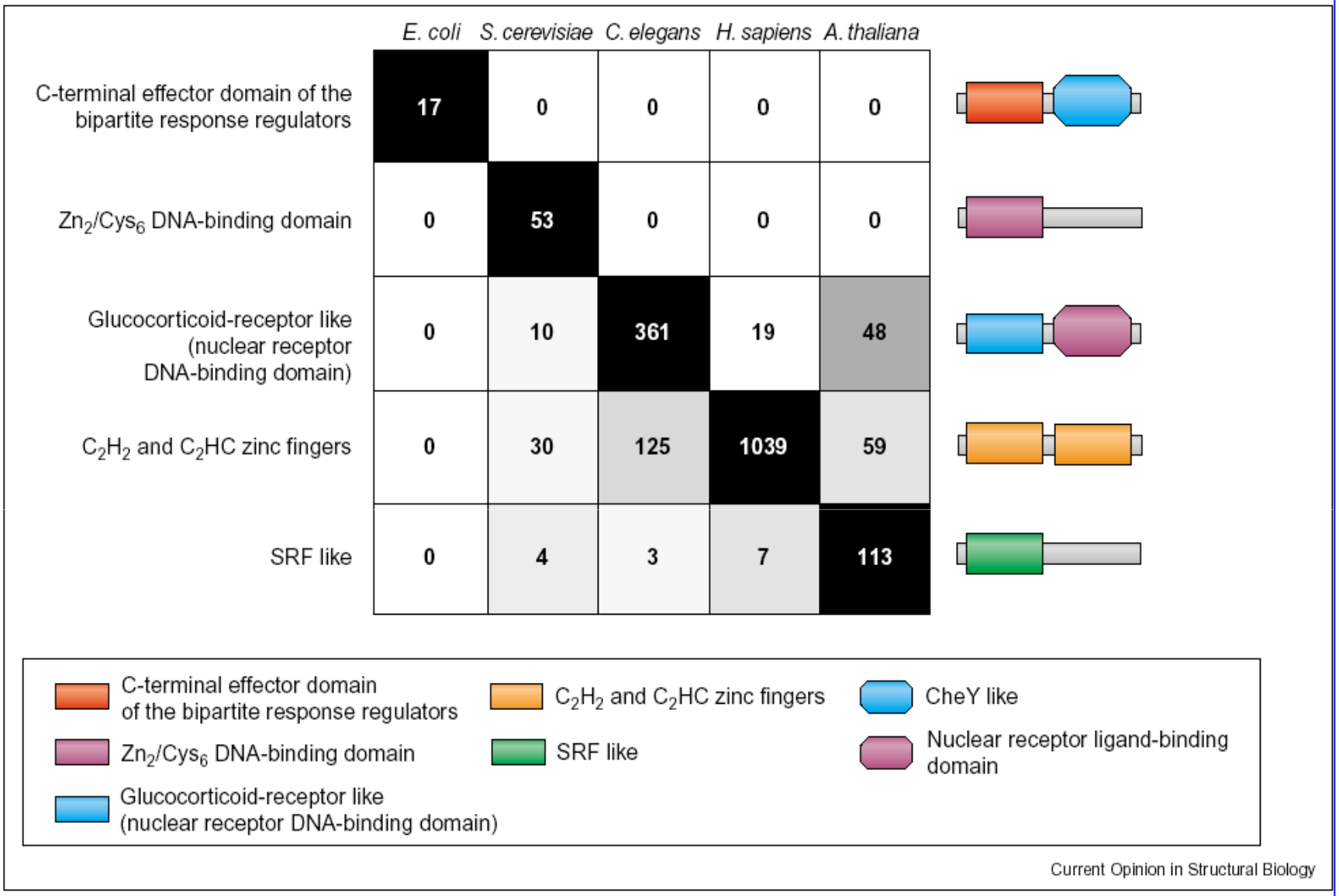
DNA-binding factors

Table 1

Numbers of DNA-binding transcription factors in five organisms[a].

Organism	Number of transcripts	Number of proteins with DNA-binding domains	Percentage of transcripts containing DNA-binding domains
E. coli	4280	267	6.2
S. cerevisiae	6357	245	3.9
C. elegans	31 677	1463	4.6
H. sapiens	32 036[b]	2604	8.1
A. thaliana	28 787	1667	5.7

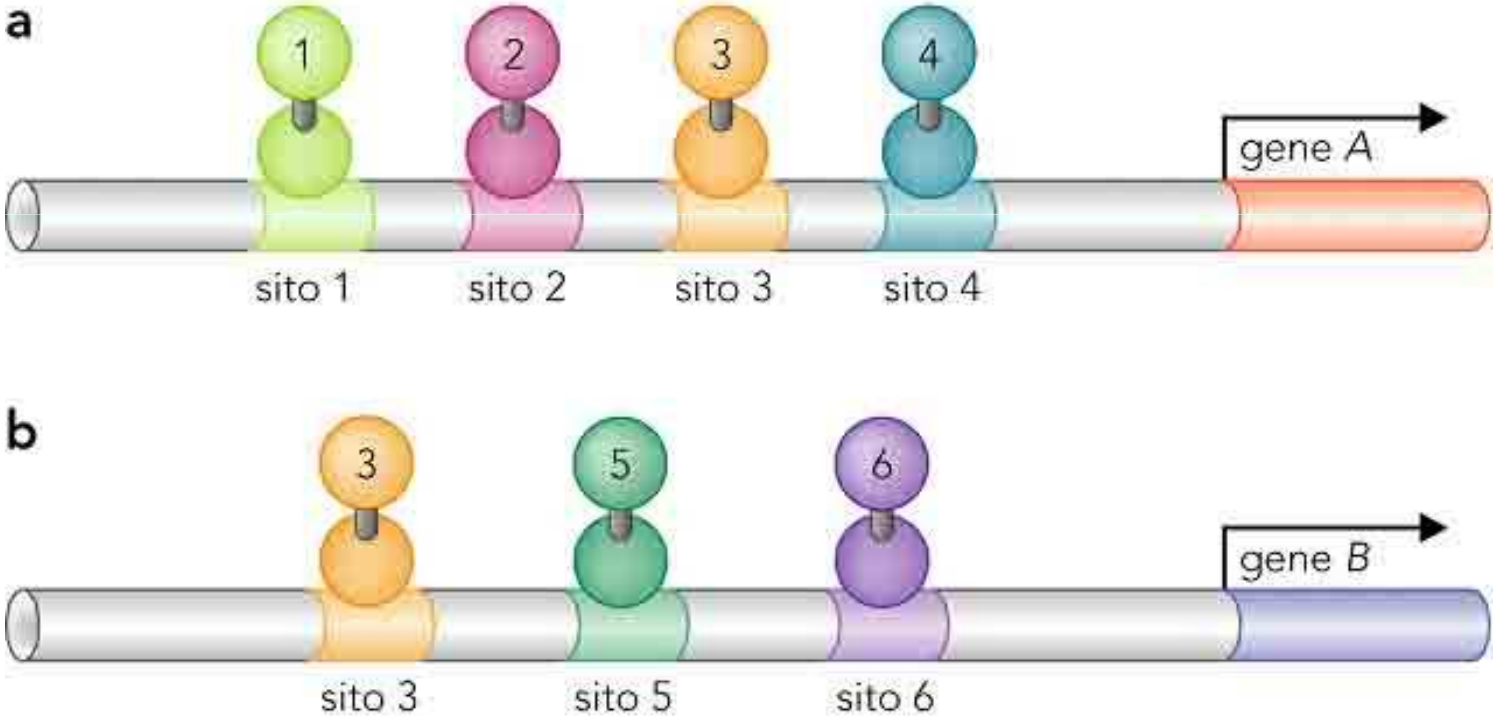
[a]DNA-binding domain assignments from Pfam and SUPERFAMILY are used to establish the repertoire of DNA-binding transcription factors in five model organisms. An expectation value threshold of 0.002 was used in making the assignments. Co-regulators that do not bind DNA directly are excluded. [b]Predicted by Ensembl v19.34a [42].



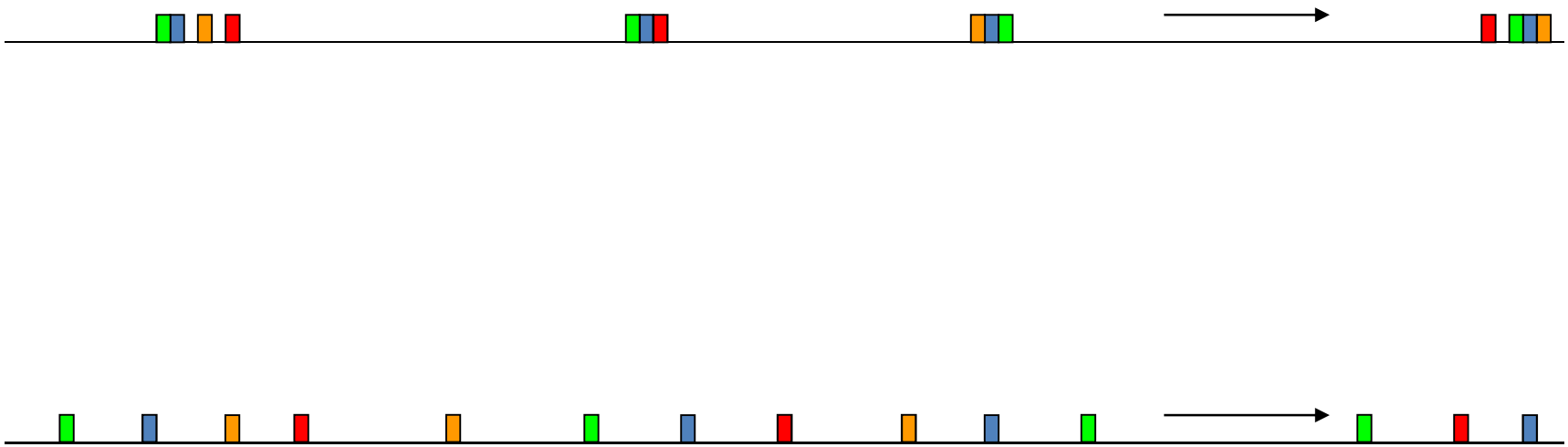
Lineage-specific expansion of DNA-binding domain families. Examples of DNA-binding domain families of transcription factors that are prevalent in one of the five genomes, but are rare in the others. The genomic occurrence of each family is provided in the table and we depict their most common domain architectures alongside. SRF, serum response factor.

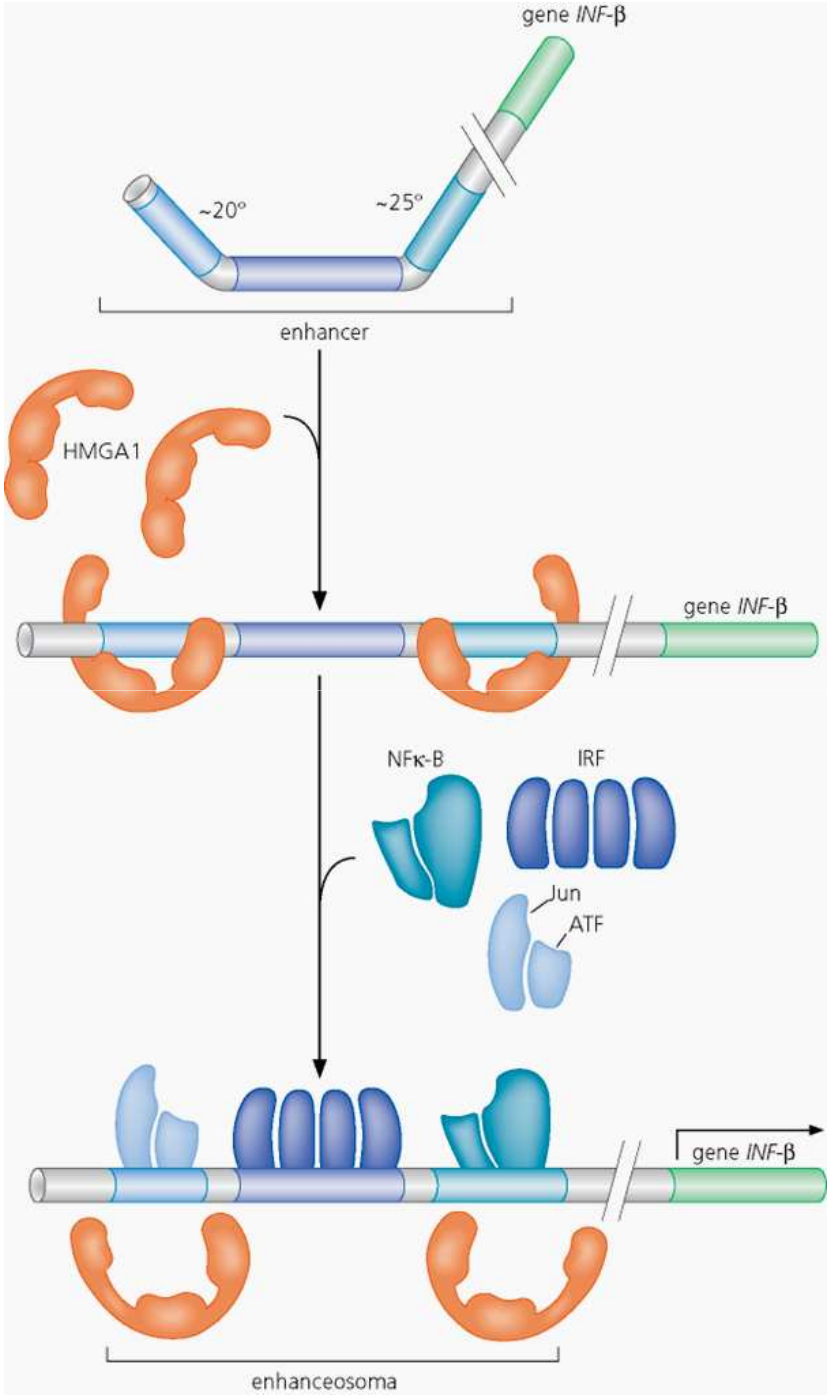
Il controllo trascrizionale viene realizzato con un numero limitato (ancorchè assai grande) di fattori trascrizionali leganti il DNA.

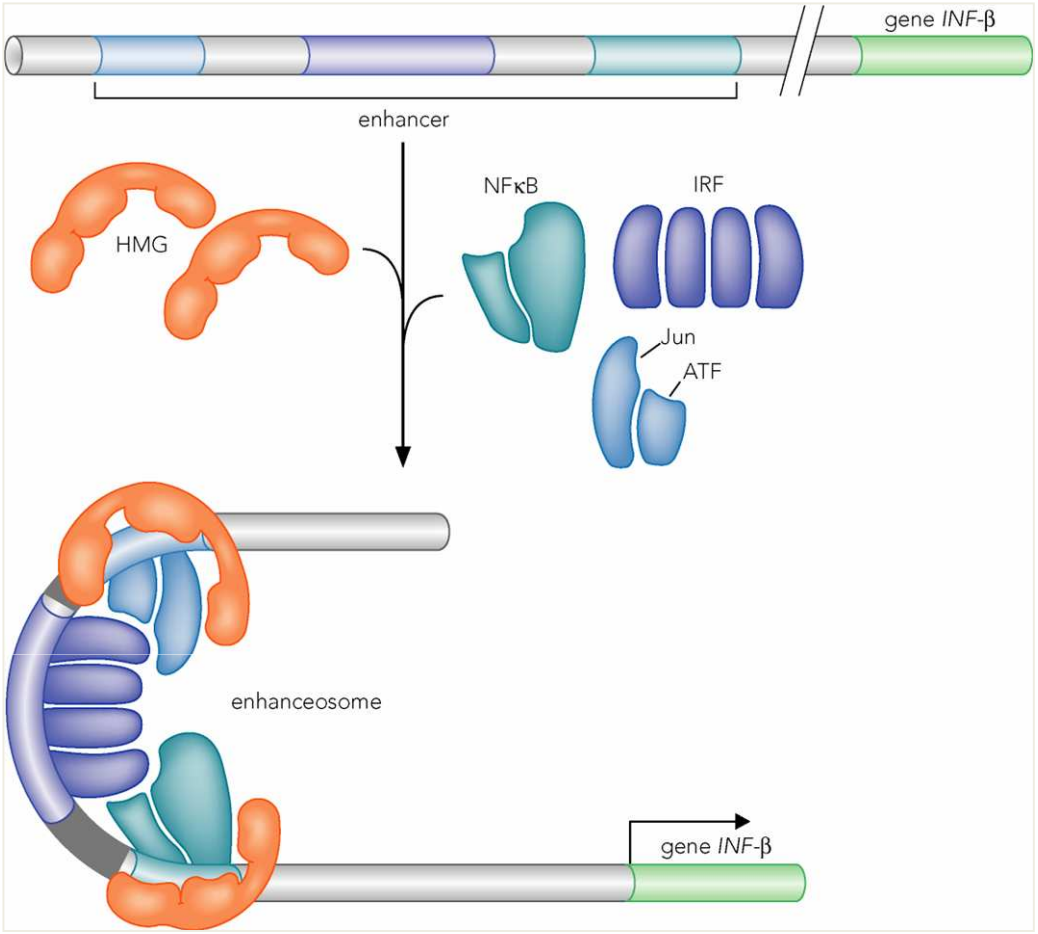
Ogni regione di controllo è formata dalla giustapposizione di diversi elementi in un ordine spaziale specifico.



La cooperazione e composizionalità delle sequenze regolatrici è suggerita anche dal fatto che gli elementi *cis* sono distribuiti in *clusters* e non in modo uniforme.

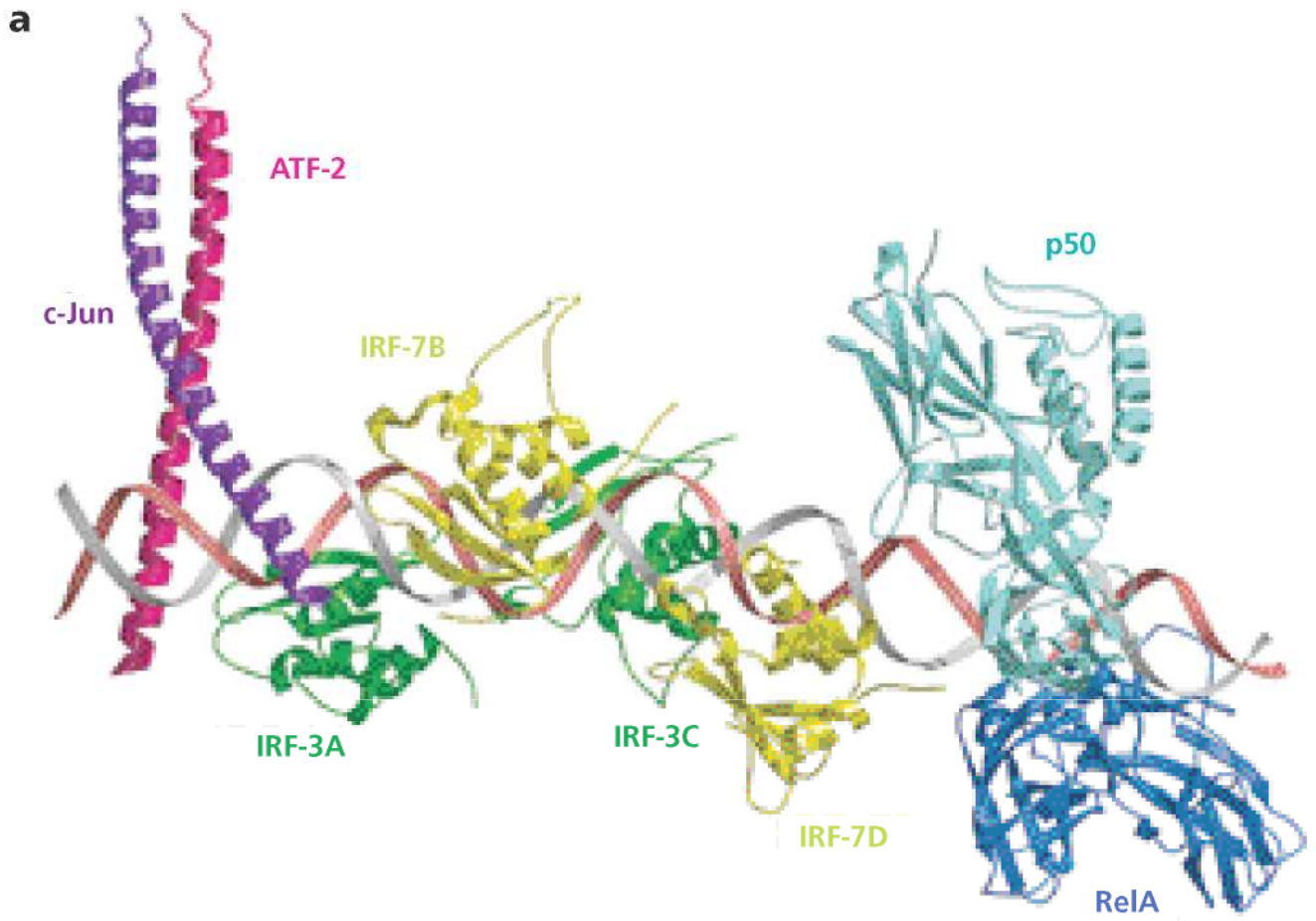



gene *INF*-β
~20°
~25°
enhancer
HMGA1
gene *INF*-β
NFκ-B
IRF
Jun
ATF
gene *INF*-β
enhanceosoma



The INF-β “enhanceosome”

In certain cases, the binding of multiple different proteins to adjacent sites in an enhancer is required to make it working. HMG are DNA-binding proteins with no transactivating domain, but display “architectural” functions, e.g. bending the DNA

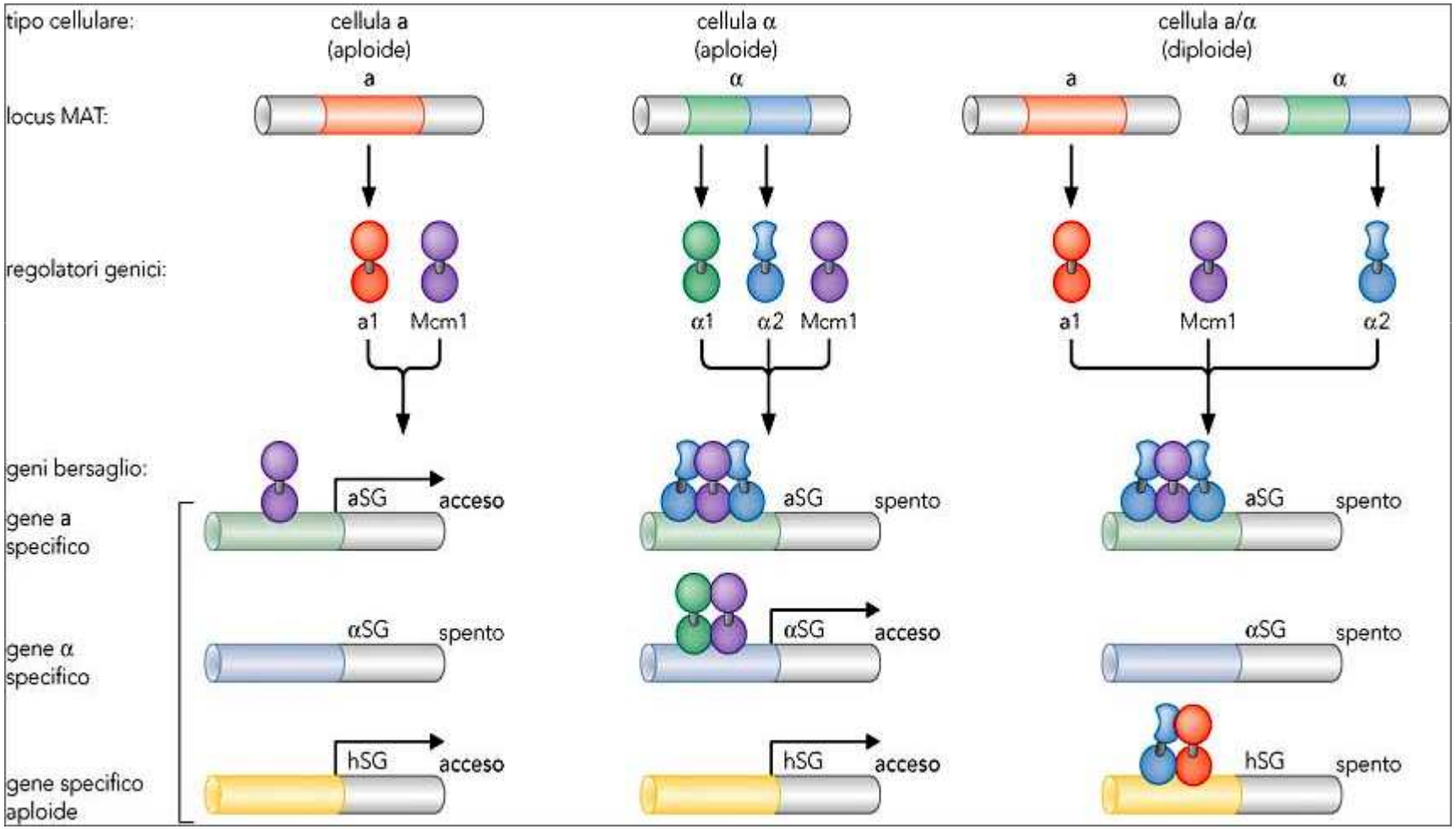


b

	ATF	Jun	IRF	IRF	IRF	IRF	NF-κB
Uomo	1:AAATGTAAATGACATAGGAAAACTGAAAGGGAGAAGTGAAAGTGGGAAATTCCTCTGAAT:60						
Topo	1:.....AAATGACAGAGGAAAACTGAAAGGGAGAACTGAAAGTGGGAAATTCCTCTGA..:52						
Ratto	1:.....AAATGACGGAGGAAAAGTGAAAGGGAGAACTGAAAGTGGGAAATTCCTCTGA..:52						
Suino	1:.....AAATGACATAGGAAAACTGAAAGGGAGAACTGAAAGTGGGAAATTCCTCTGAA.:53						
Cavallo	1:.AATGTAAATGACATAGGAAAACAGAAAGGGAGAACTGAAAGTGGGAAATTCCTCTGAA.:58						
Bovino2	1:.....TAAATGACAAAGGAAAACTGAAAGGGAGAACTGAAAGTGGGAAATCTCTCC....:45						
Bovino	1:.....TAAATGACATGGGAAAAATGAAAGCGAGAACTGAAAGTGGGAAATTCCTCT....:51						

100 milioni di anni

As clearly esemplified in the case of the MAT locus-encoded transcription factors a1, α1, α2, the activatory or repressory result is given by the **combination** and **geometry** of interaction between different factors, on composed DNA elements.



Combinatorial control of gene expression

Attila Reményi[1,2,4], Hans R Schöler[1,3] & Matthias Wilmanns[2]

Revealing the molecular principles of eukaryotic transcription factor assembly on specific DNA sites is pivotal to understanding how genes are differentially expressed. By analyzing structures of transcription factor complexes bound to specific DNA elements we demonstrate how protein and DNA regulators manage gene expression in a combinatorial fashion.

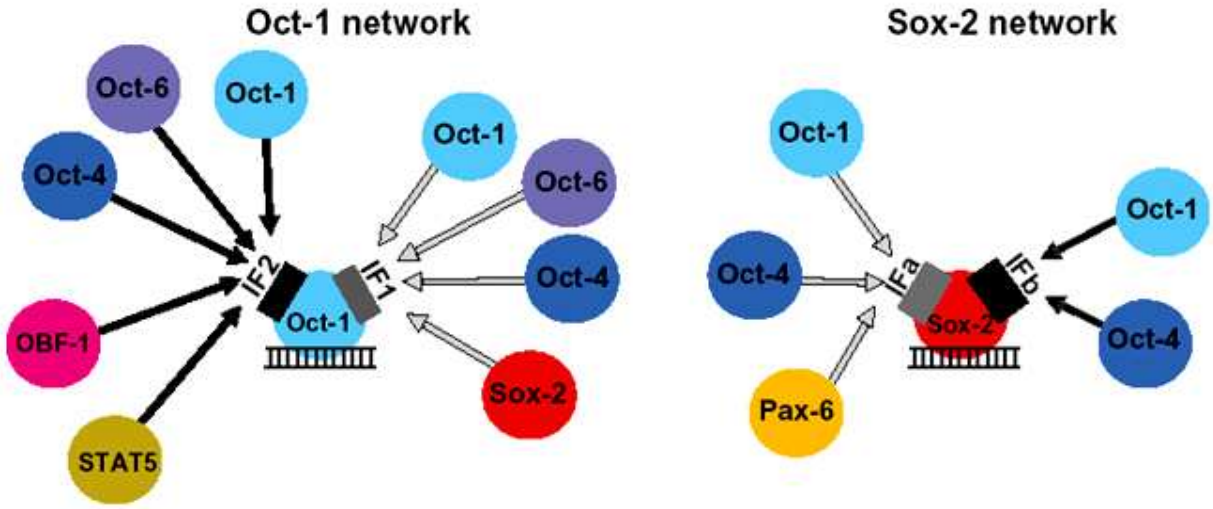


Figure 3 Interaction diagram of Oct-1 and Sox-2. Transcription factors are depicted as protein molecules with surface patches that can interact with a whole array of different partners provided that the protein is bound to a specific DNA element. DNA-bound Oct-1 and Sox-2 are depicted schematically with protein-protein interaction surface patches that are instrumental in binding to other partners. IF1 and IF2 on the Oct-1–DNA complex denote two interfaces of Oct-1 that are accessible and used for interaction on various DNA. Similarly, IFa and IFb designate interfaces of Sox-2 that are used for interaction on different DNA sites.

Historical

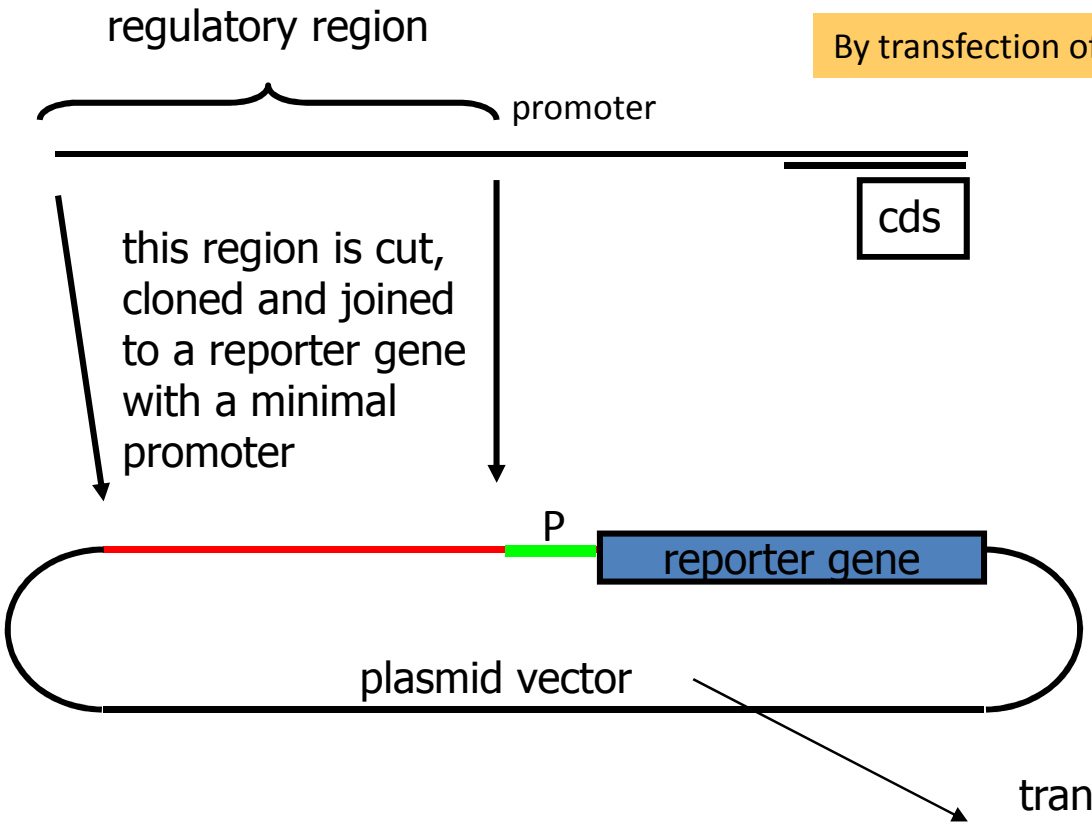
1° route:
isolating a promoter sequence, make deletional mutants and identify regulatory elements.
This is paralleled with Dnase I footprinting experiments using whole N. E.
Once identified, the response elements are further analyzed by Band-shift (EMSA)
Proteins bound are then isolated by DNA affinity chromatography and identified.

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

2° route:
Several putative TFs are identified by homology cloning.
The binding site was often identified by SELEX
Finally, bioinformatic search for the binding site is performed on known genomic sequences.

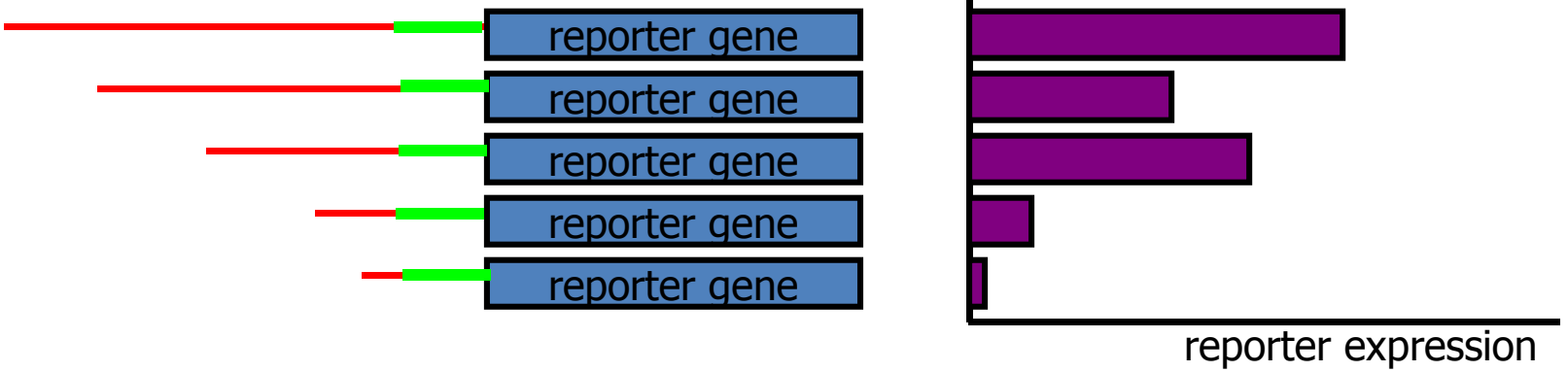
3° route:
Conserved, nontranscribed sequences proximal to known genes are explored statistically to describe over-represented sequence “words” as compared to the whole genome.
Experimental proofs that the identified “words” (or motifs) can bind regulatory factors are needed

By transfection of reporter vectors in cultured cells

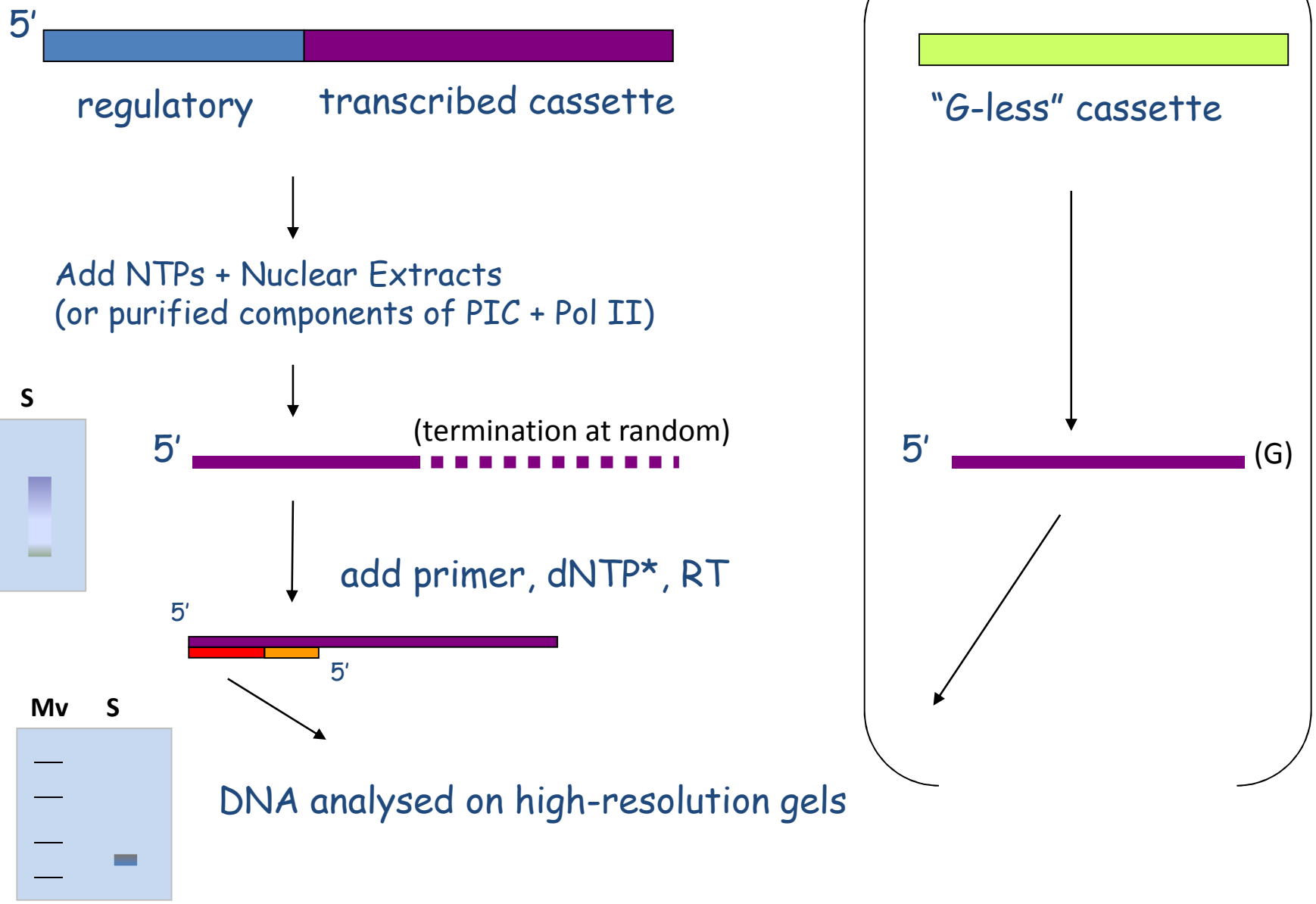


transfected into animal cells, expression of reporter protein measures the activity or regulatory sequences

Progressive deletions may identify different regulatory elements



In vitro transcription assay



When a DNA binding protein is under study, the sequence of DNA it interacts with can be selected using a process called:

SELEX

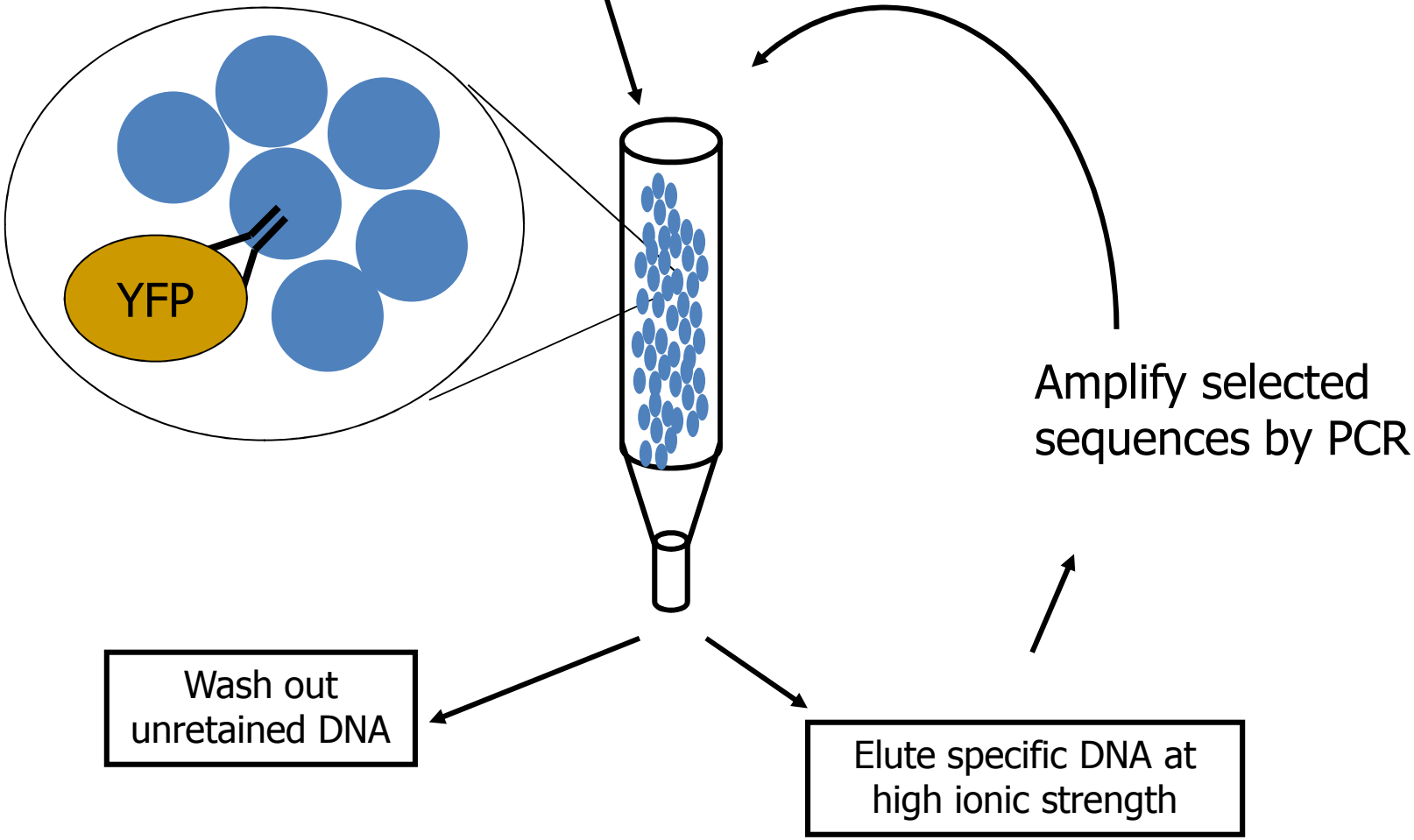
Systematic Evolution of Ligands by Exponential Enrichment

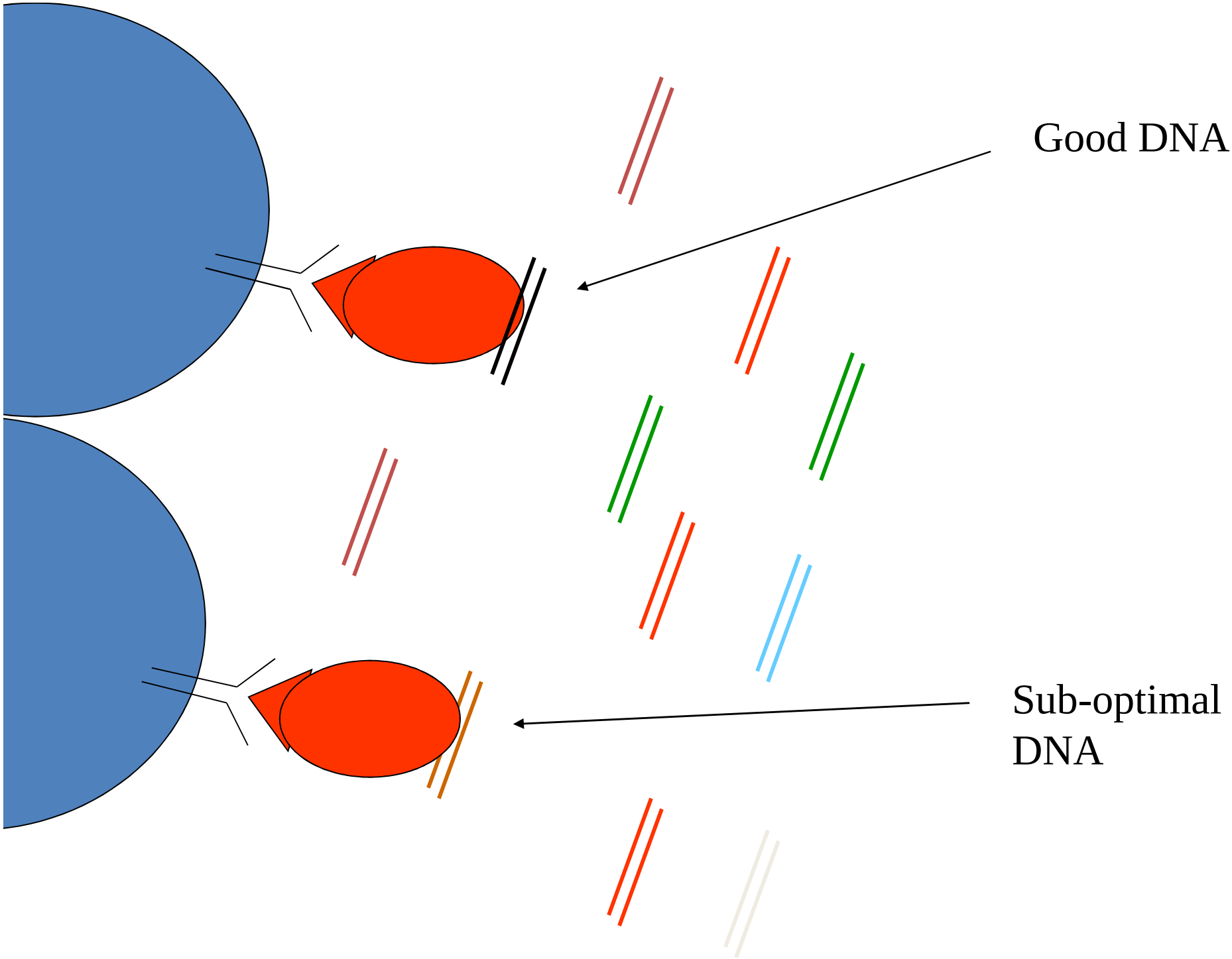
5’sense primer- NNNNNNNNNNNN – reverse primer3’
3’remirp esnes- NNNNNNNNNNNN – remirp esrever5’

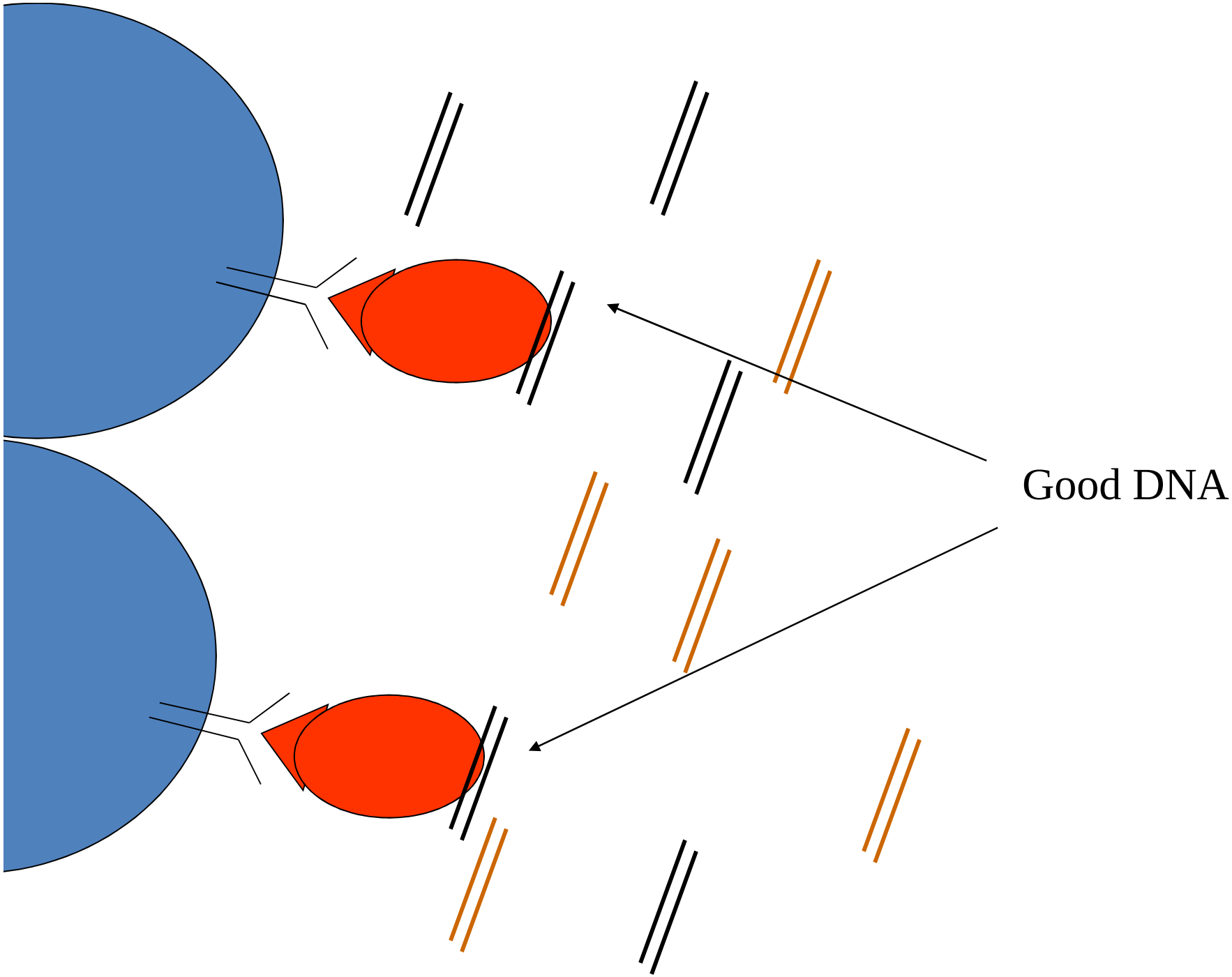
random (4^n sequences)

SELEX

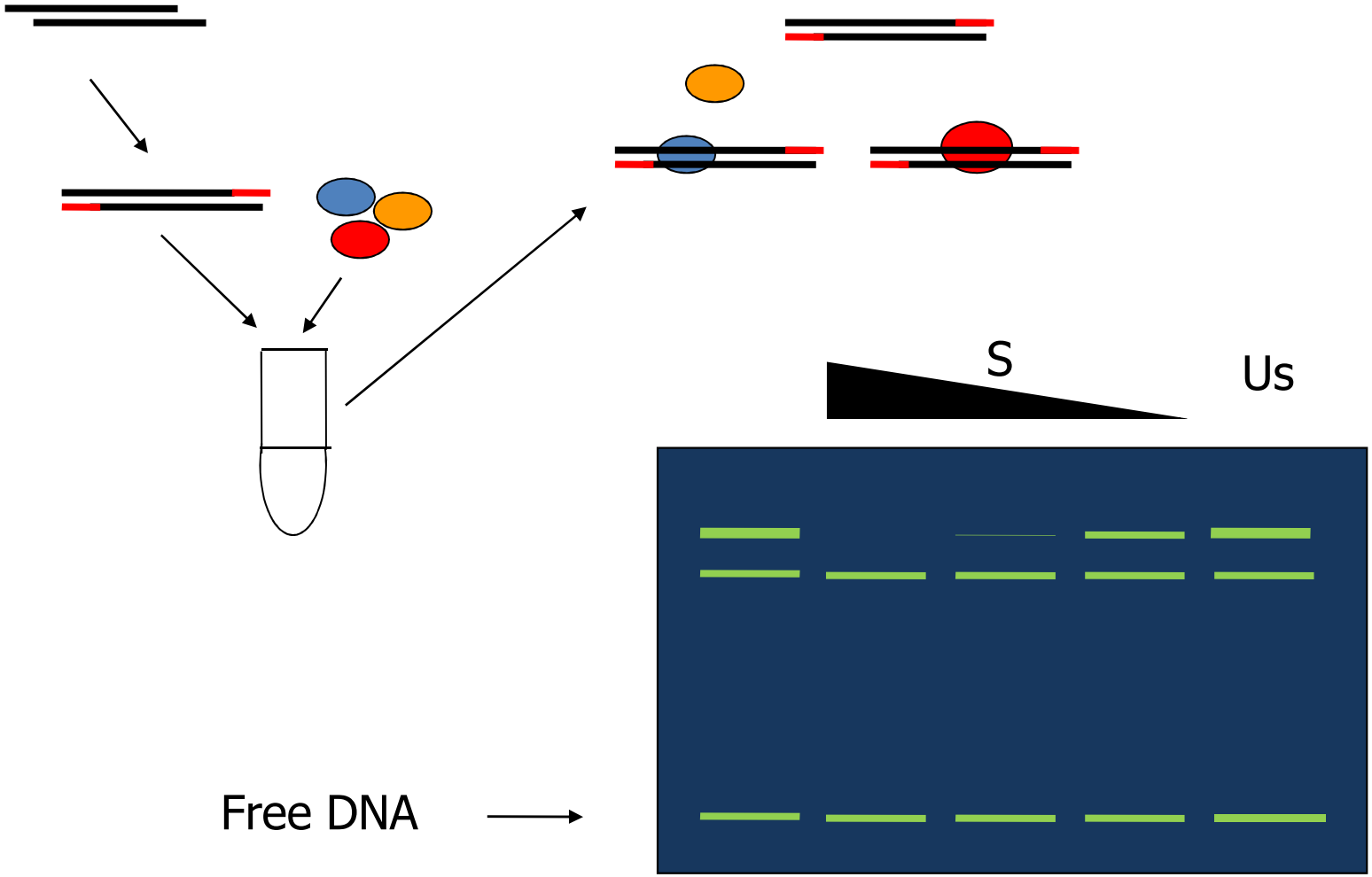
Systematic Evolution of Ligands by Exponential Enrichment




Good DNA
Sub-optimal DNA


Good DNA

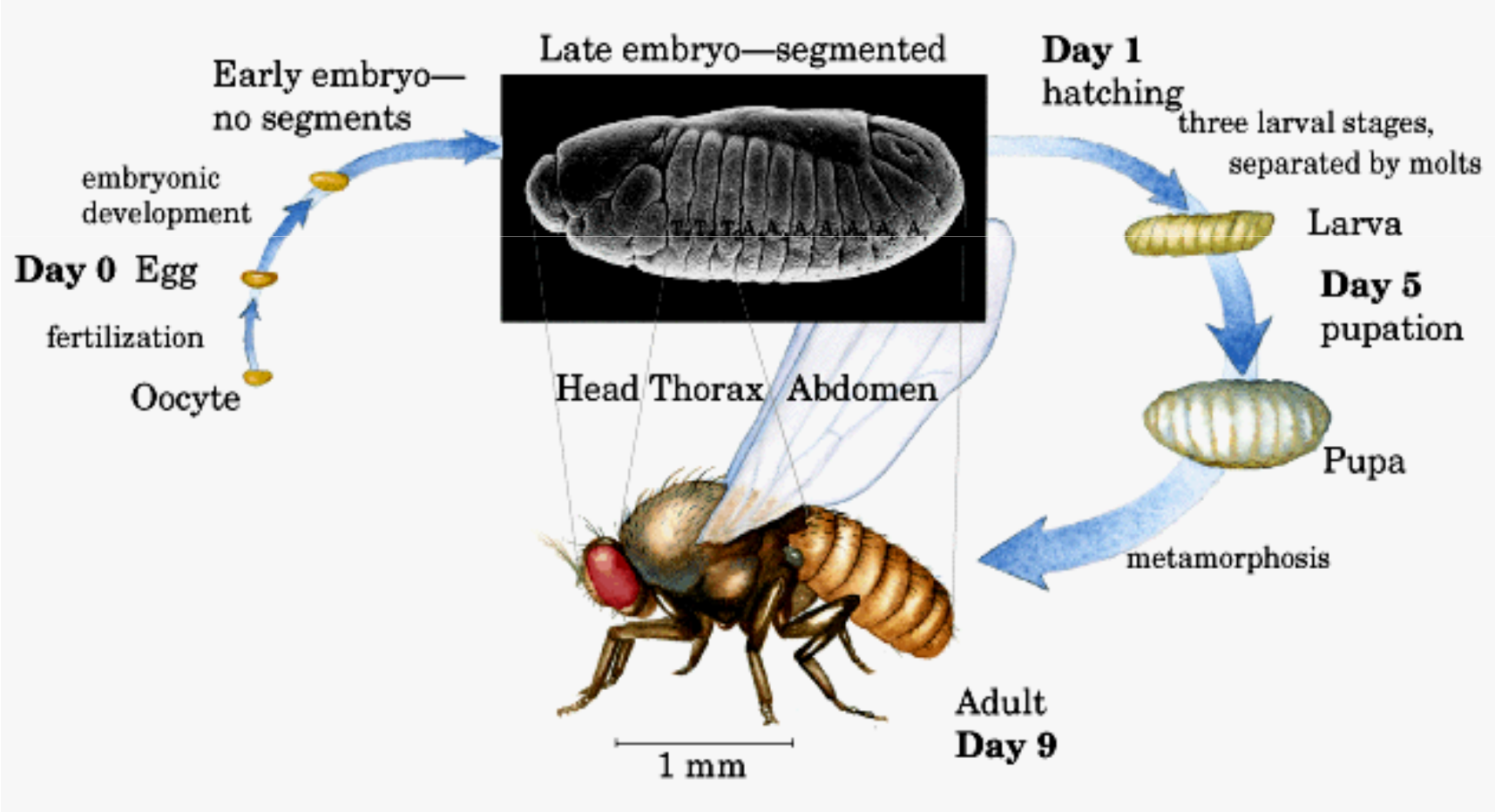
Band-shift assay or Electrophoretic Mobility Shift Assay EMSA

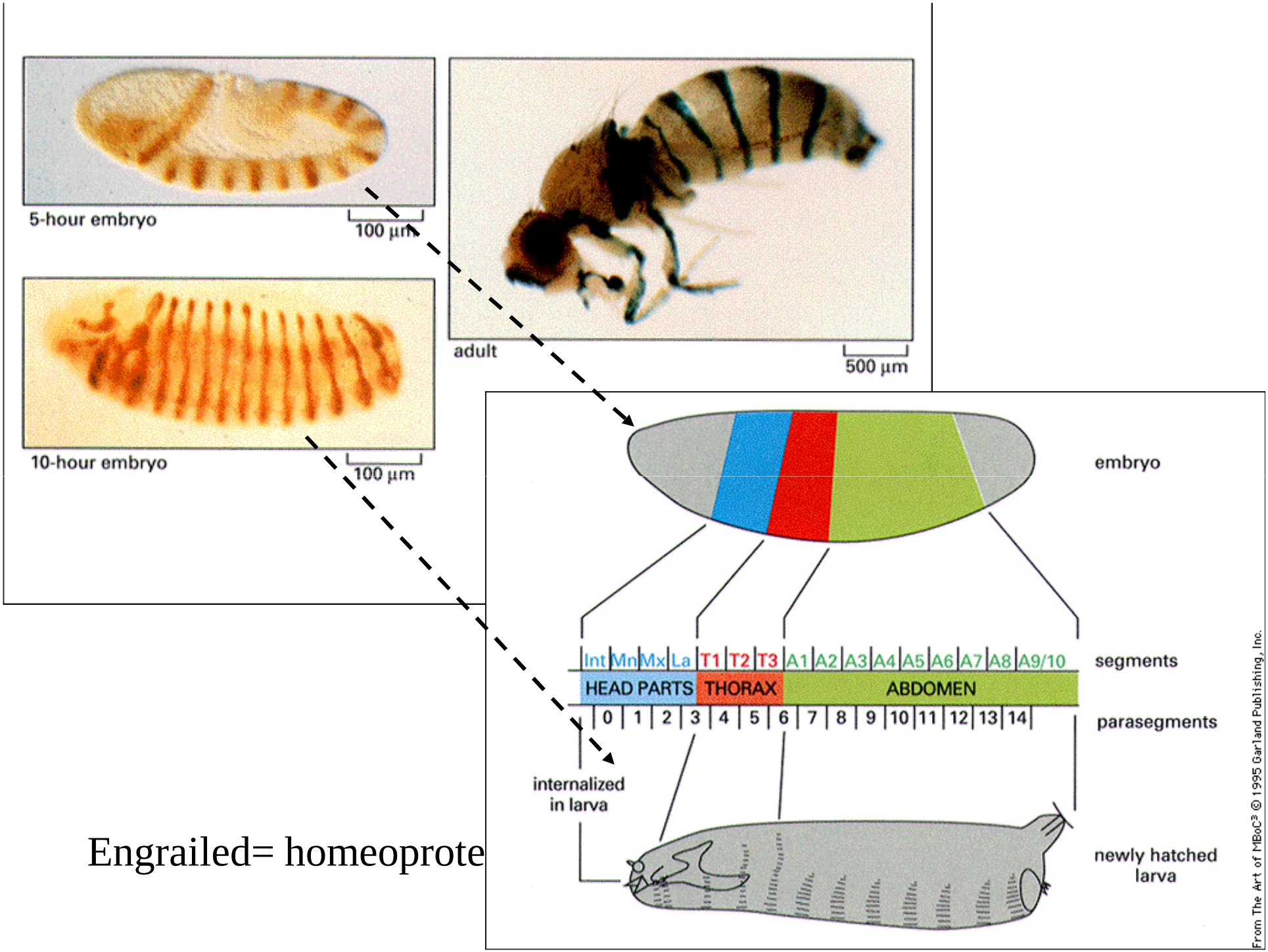


Molti dei concetti di “composizionalità” e “combinatorialità” degli elementi di regolazione trascrizionale deriva da

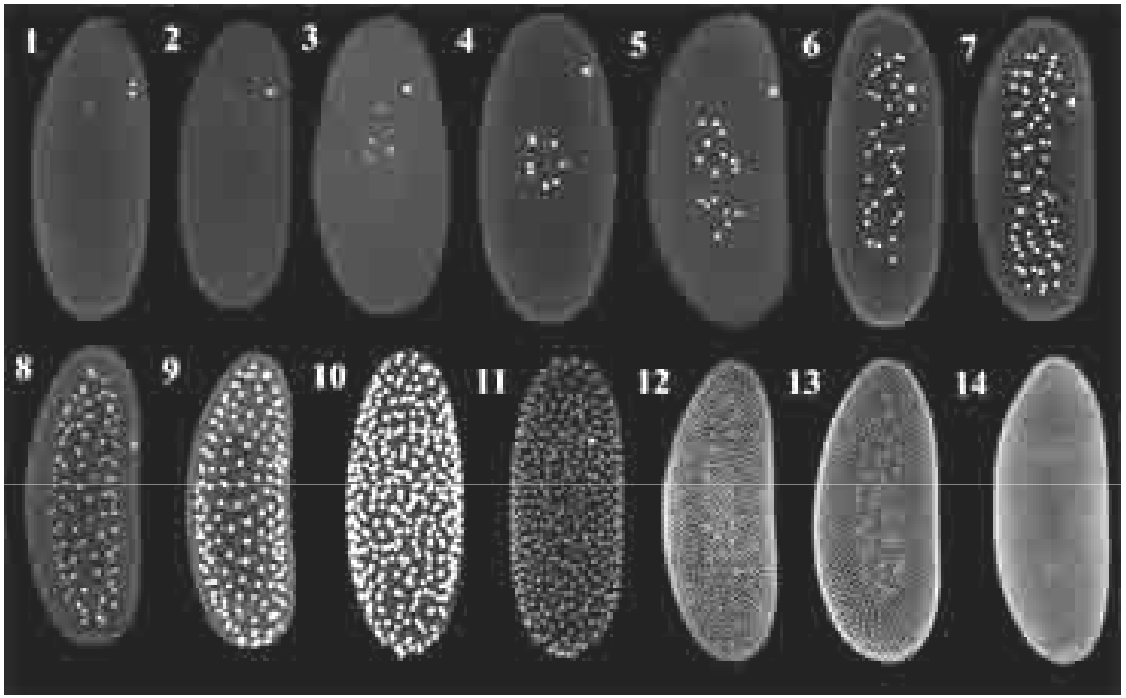
biologia molecolare dello sviluppo

in particolare dal modello di *Drosophila melanogaster*

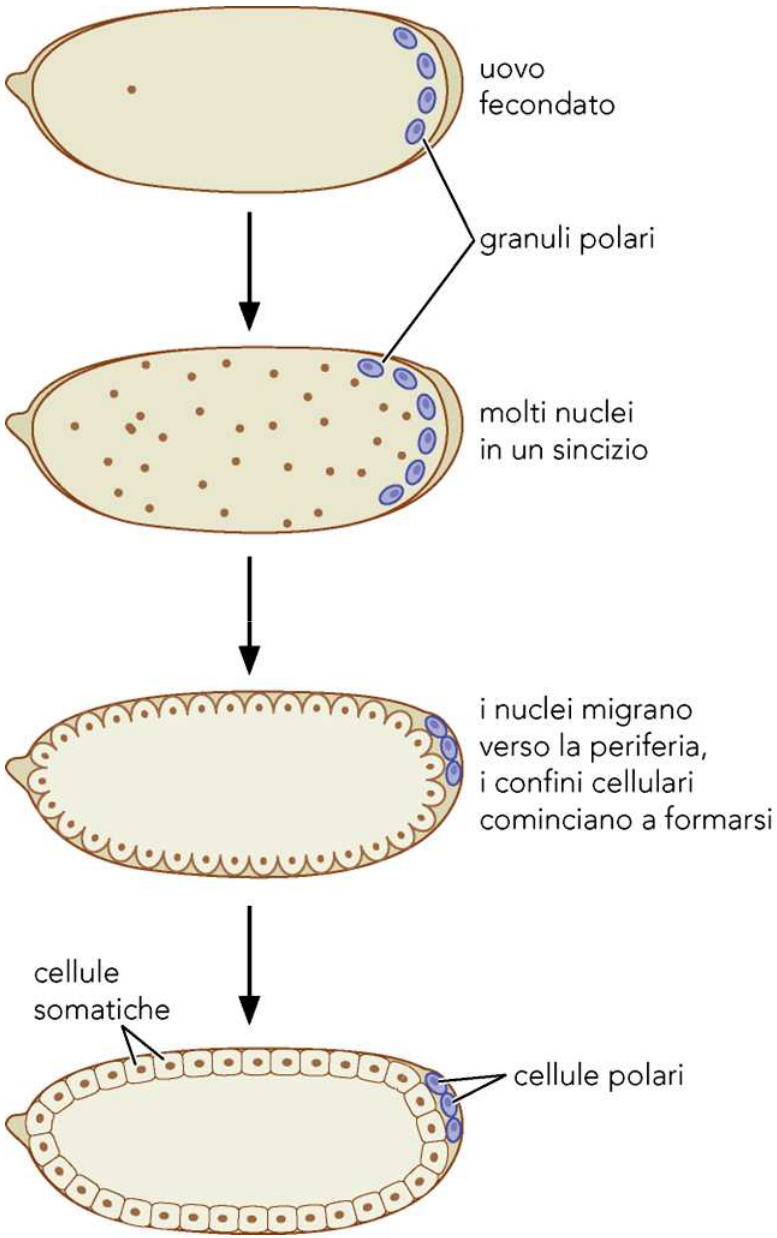




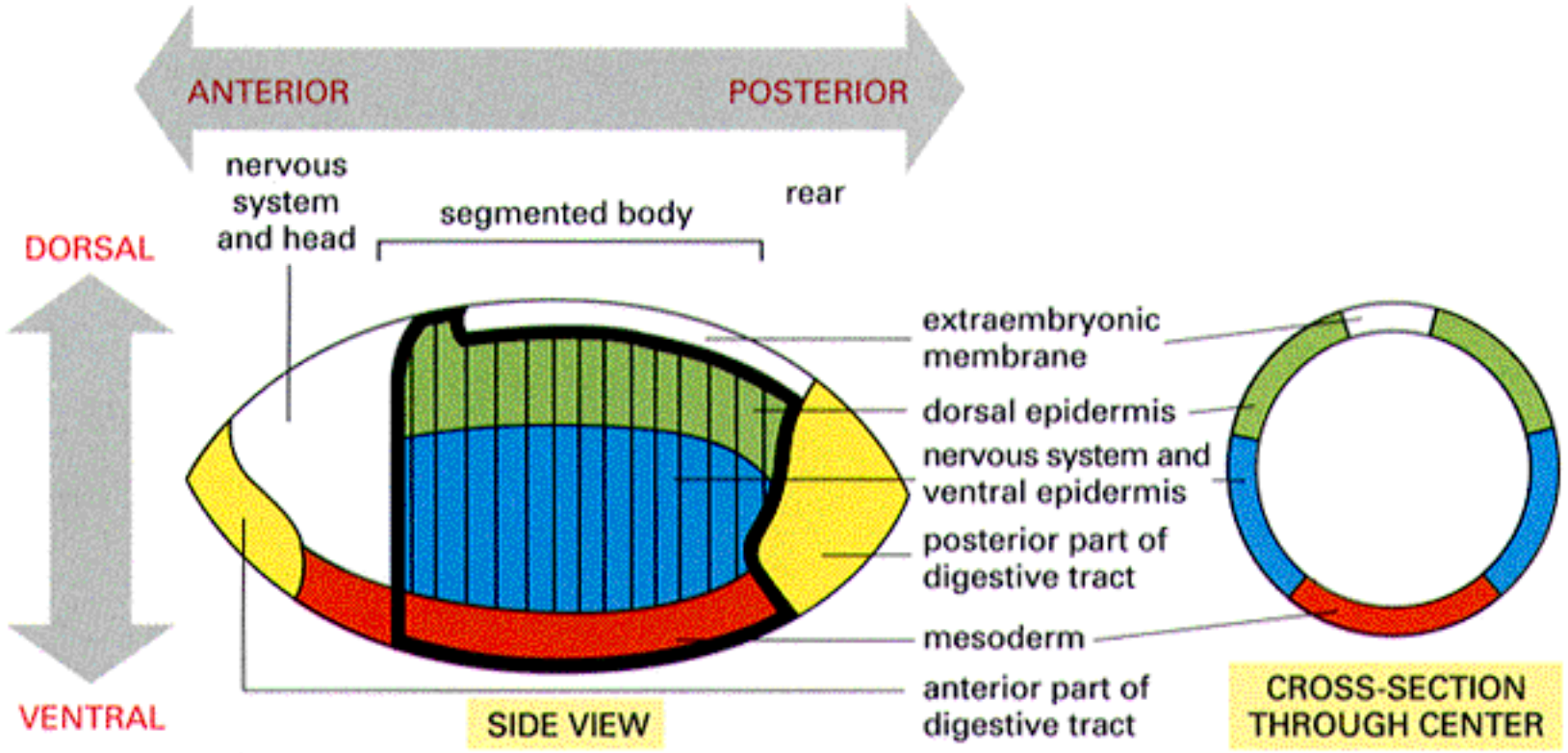
Engrailed= homeoprote

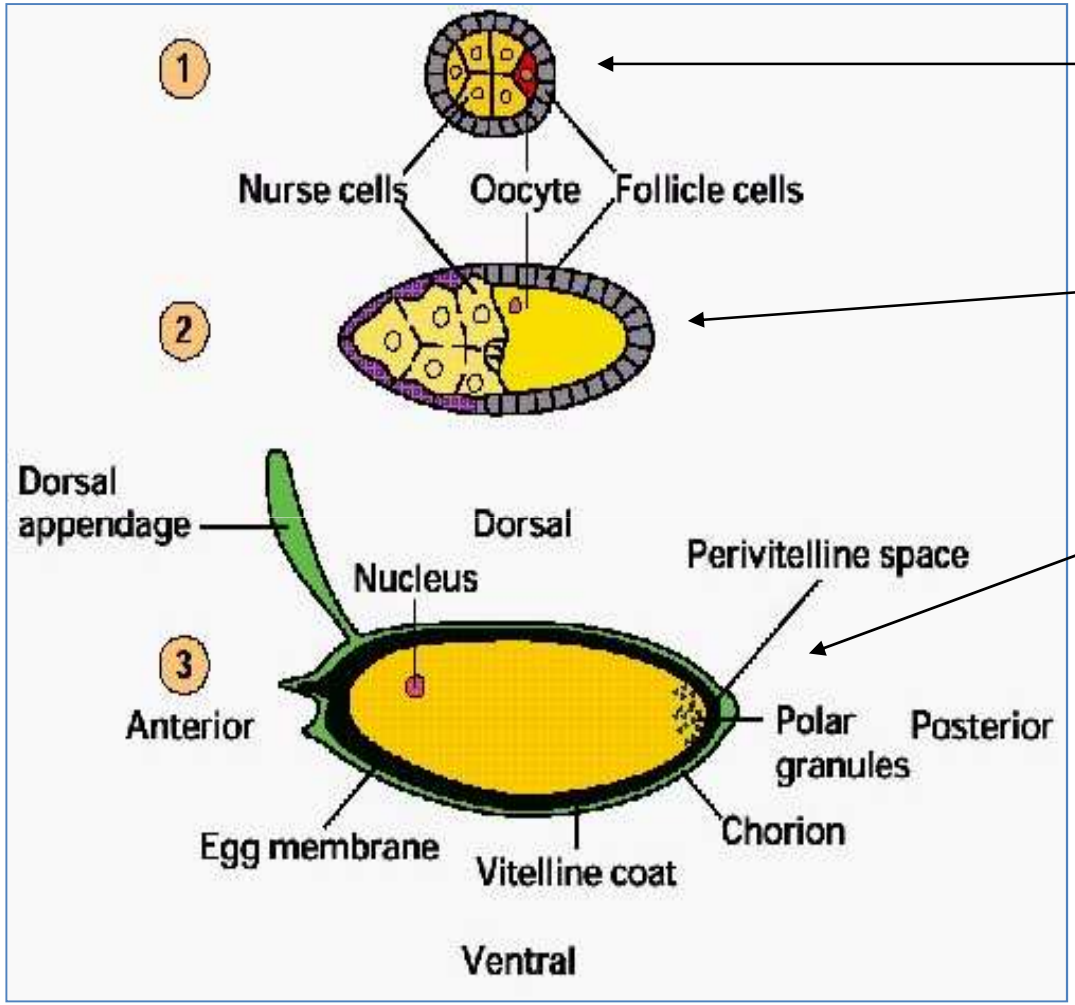

1
2
3
4
5
6
7
8
9
10
11
12
13
14


(a)
(b)
(c)
(d)
Syncytial blastoderm
Nuclei
Cleavage furrow
Yolk
Nuclear elongation and extension of cleavage furrows between cells
Cellularization
Blastula
Yolk


uovo
fecondato
granuli polari
molti nuclei
in un sincizio
i nuclei migrano
verso la periferia,
i confini cellulari
cominciano a formarsi
cellule
somatiche
cellule polari

Quindi nell'embrione precoce abbiamo già le direzionalità dello sviluppo, sia in senso antero-posteriore, sia in senso dorso-ventrale





Each developing unit, or follicle, consists of a developing oocyte, nurse cells and a layer of somatic cells called follicle cells.

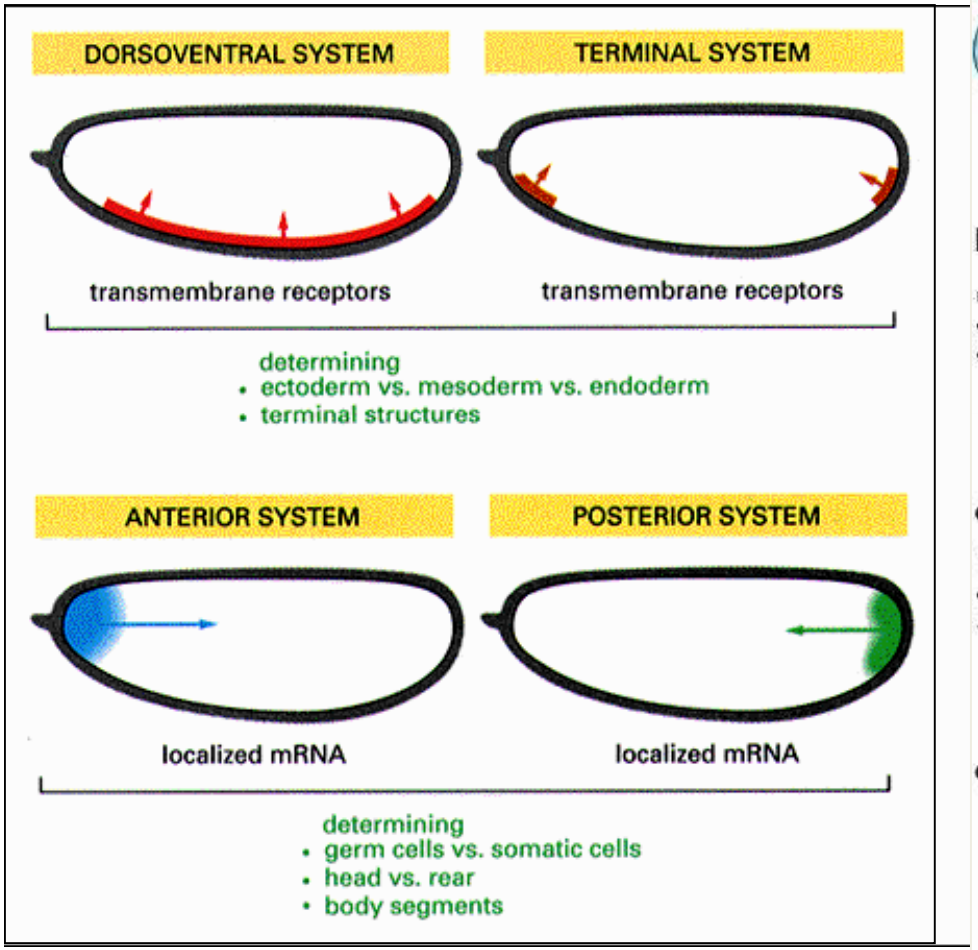
Stage 1: Early in oogenesis, the oocyte is about the same size as the neighboring nurse cells.

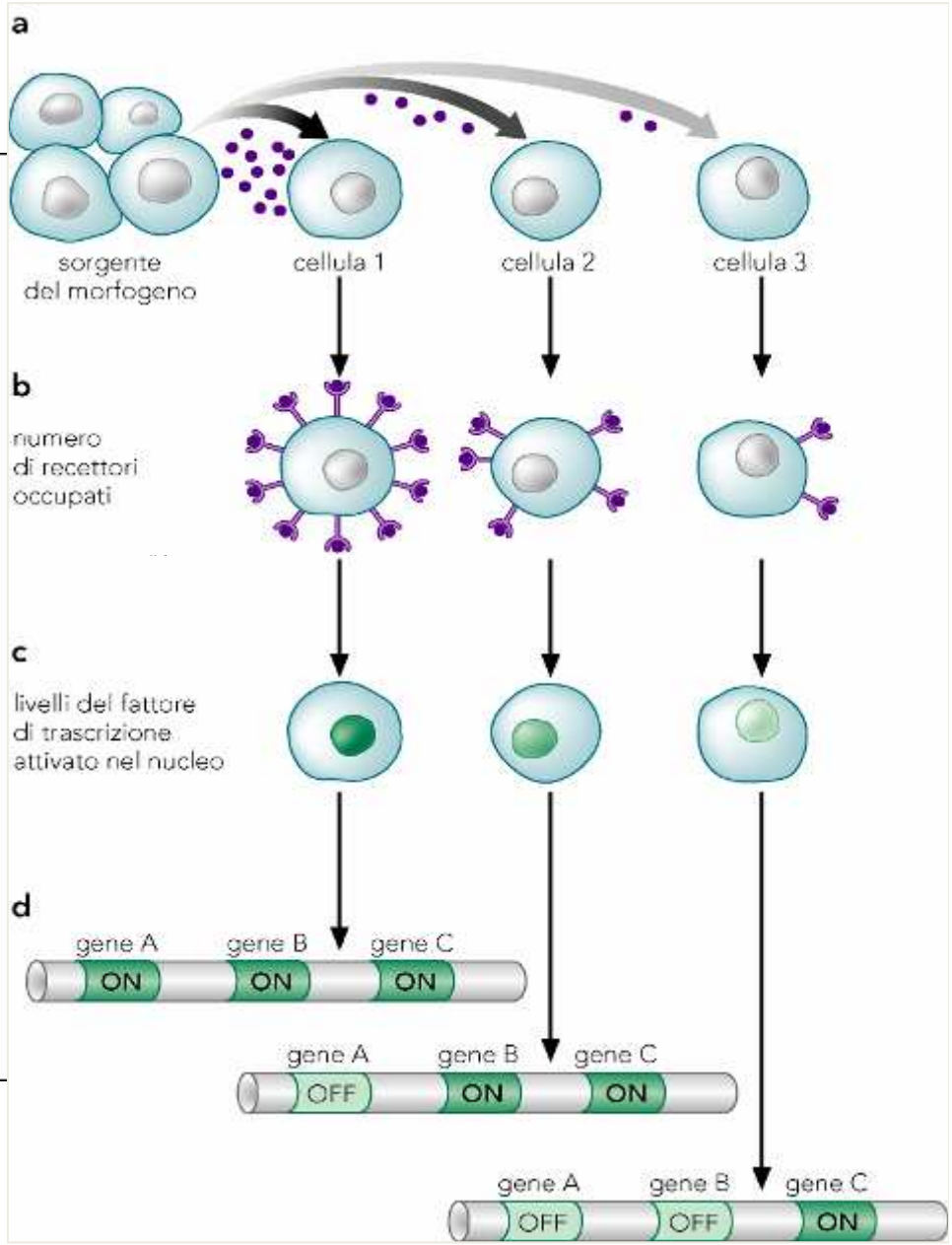
Stage 2: The nurse cells begin to synthesize mRNAs and proteins necessary for oocyte maturation, and the follicle cells begin to form the egg shell.

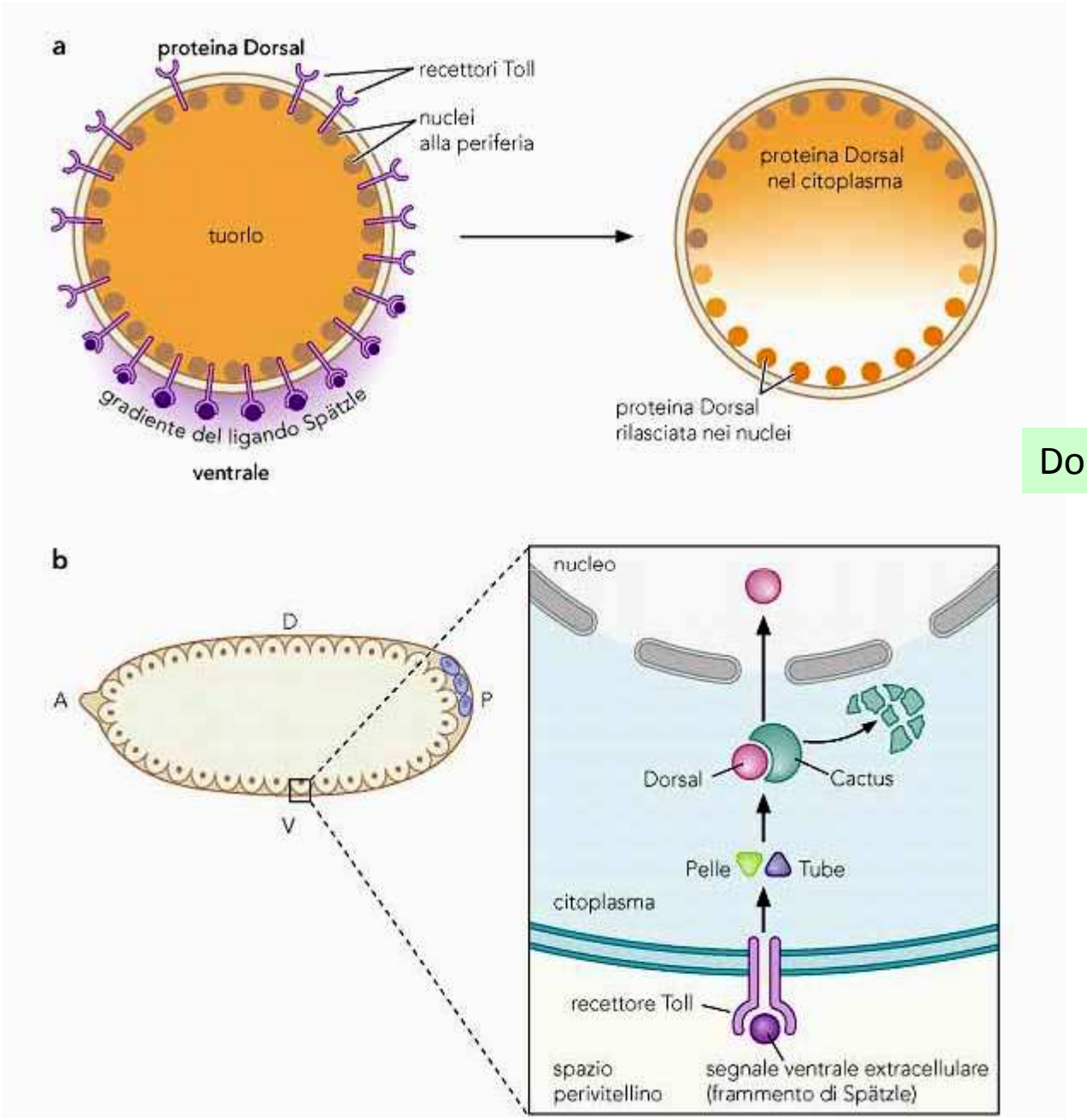
Stage 3: The mature egg is surrounded by the vitelline coat and chorion, which compose the egg shell. The nurse cells and follicle cells have been discarded, but some of the mRNAs synthesized by nurse cells, which become localized in discrete spatial domains of the oocyte, function in early patterning of the embryo.

Polar granules are disinct cytoplasmic structures located in the posterior region of the egg. This is the region in which germ cells arise.

Several morphogenetic gradients are present

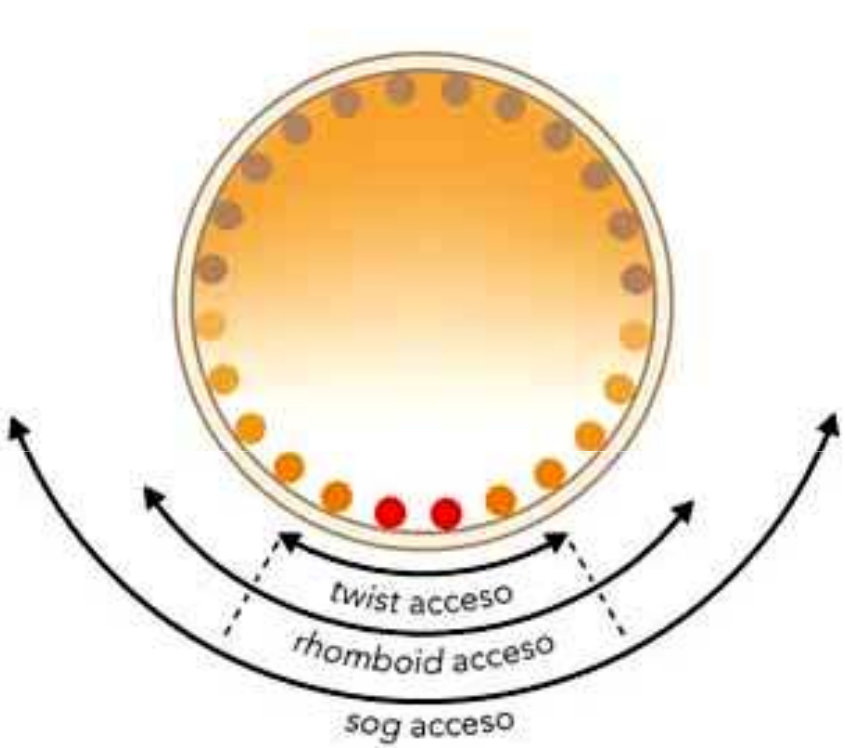


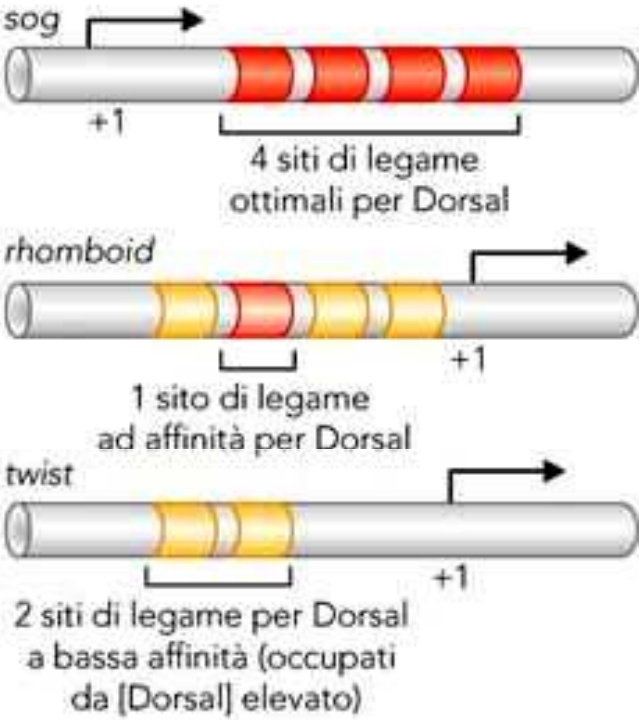




Dorsal è l'omologo di NFkB

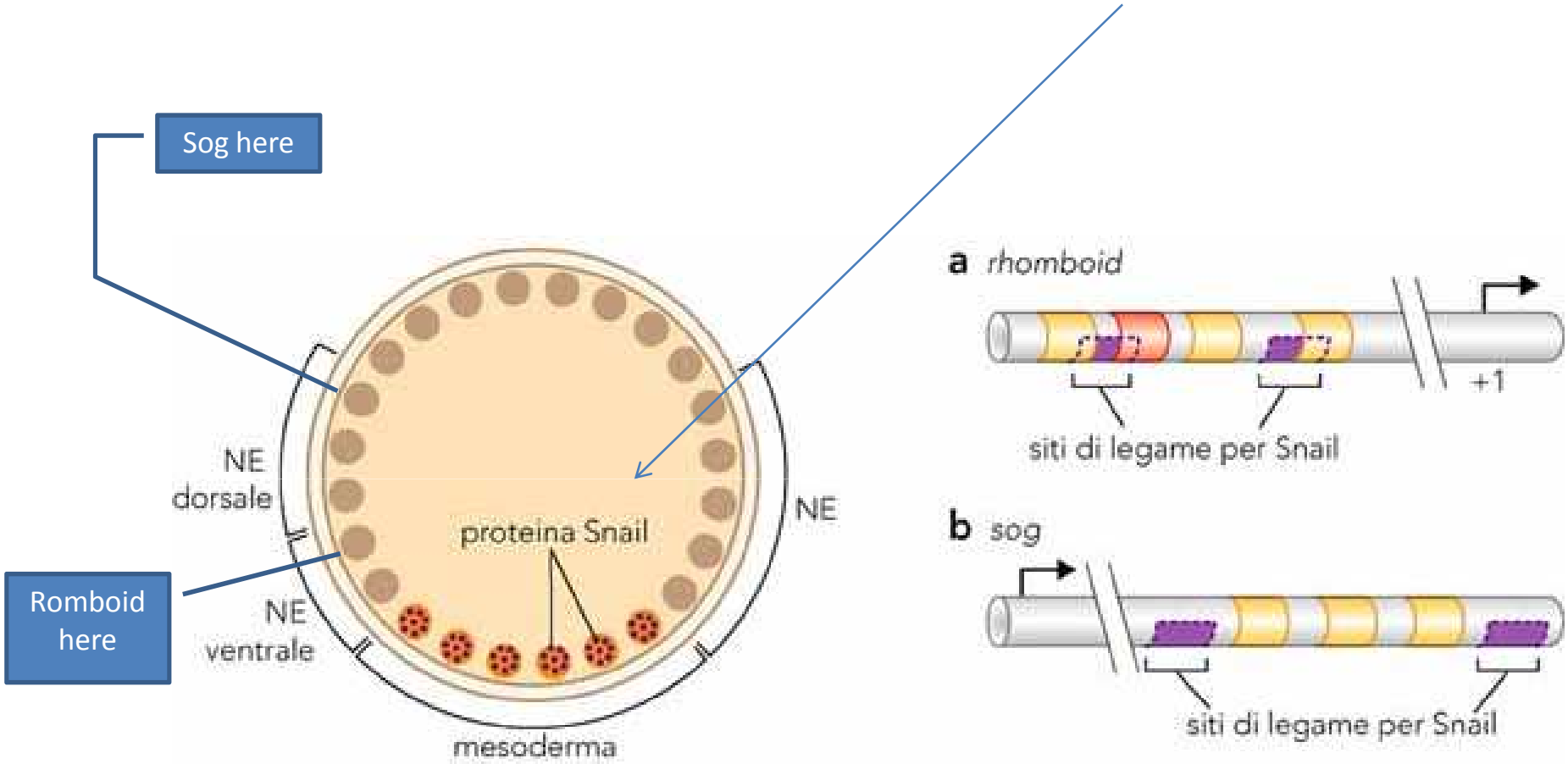
geni regolati da dorsal





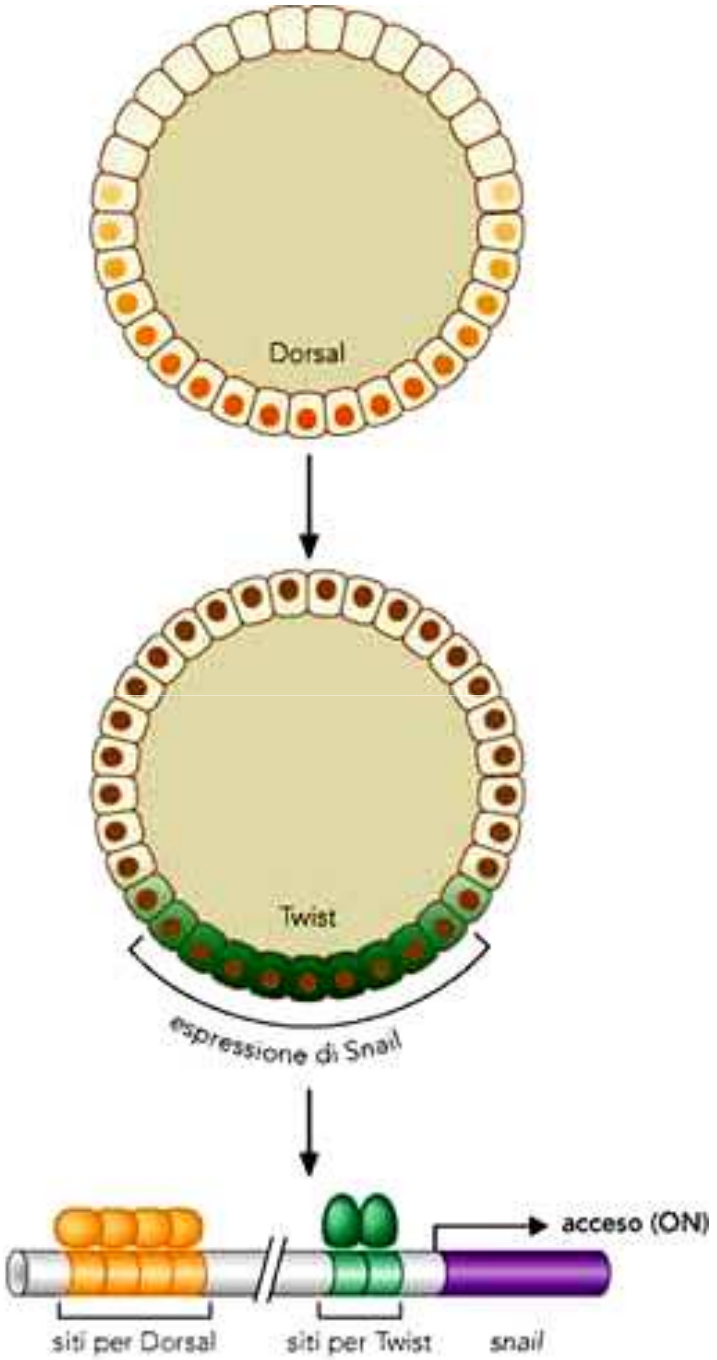
Twist 5’ contains 2 low affinity sites for Dorsal (bound only were Dorsal is higher)
Rhomboid 5’ enhancer cotains several sites: only one is high-affinity: it is on at high or intermediate levels of Dorsal.
Sog intronic enhancer contains 4 high-affinity dorsal sites: **on** in all cells where dorsal is present

In ventral cells, twist also stimulate expression of the Snail repressor:



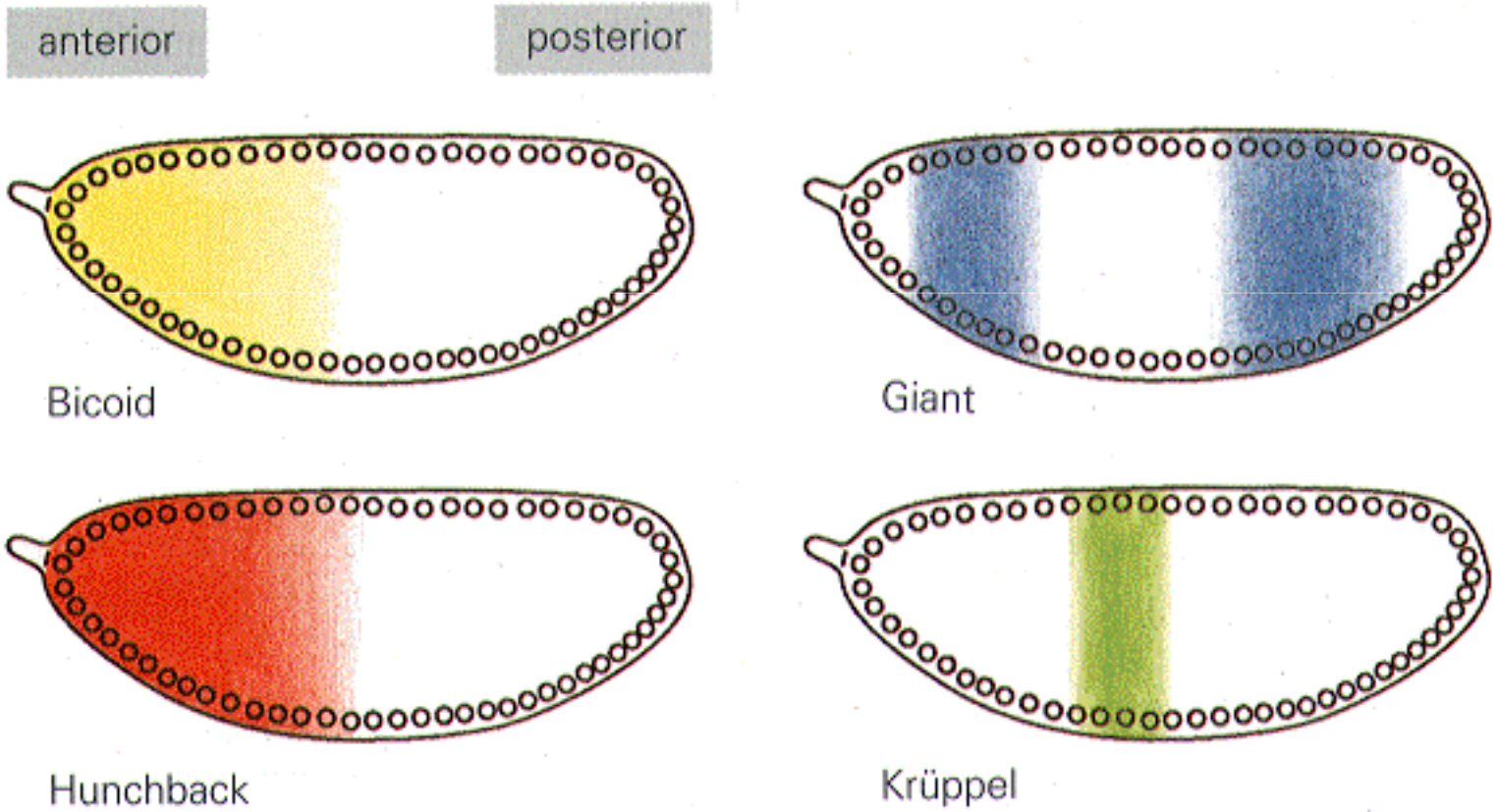
Snail è il repressore primario dei geni epiteliali (es. E-cadherin)

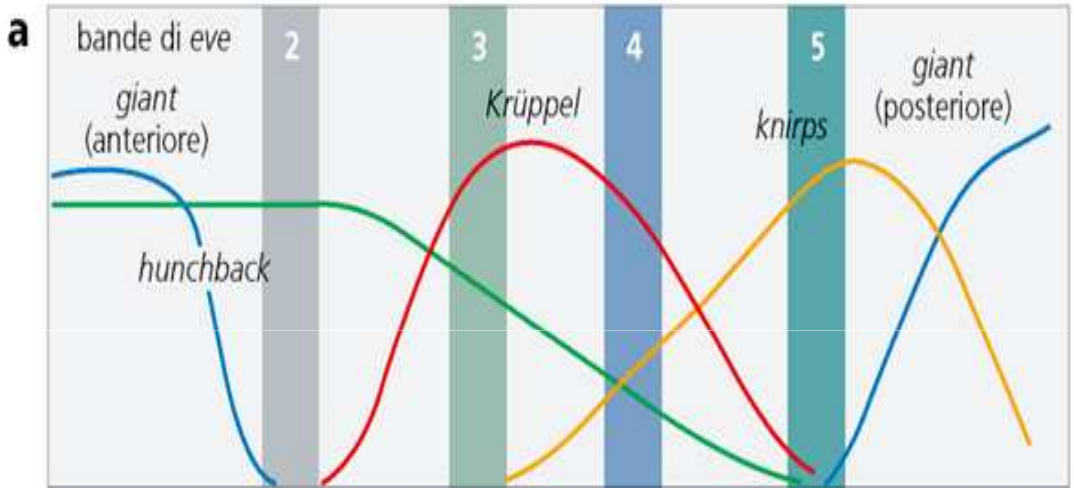
Snail is activated by **synergy** between Twist and dorsal

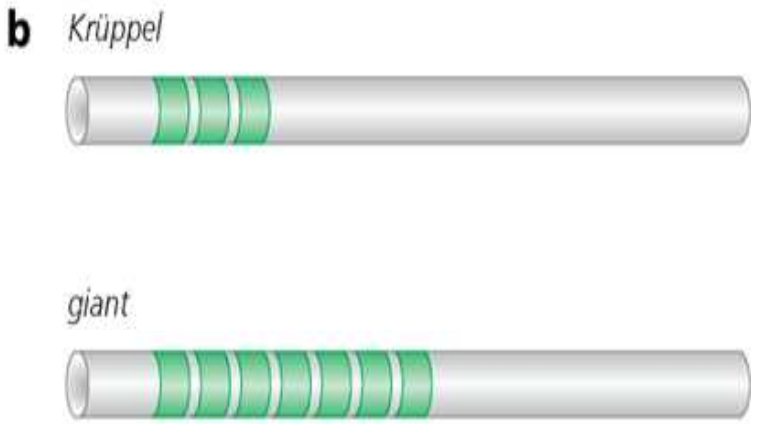


primary genes

Gap genes (Giant, kruppel, Knirps)

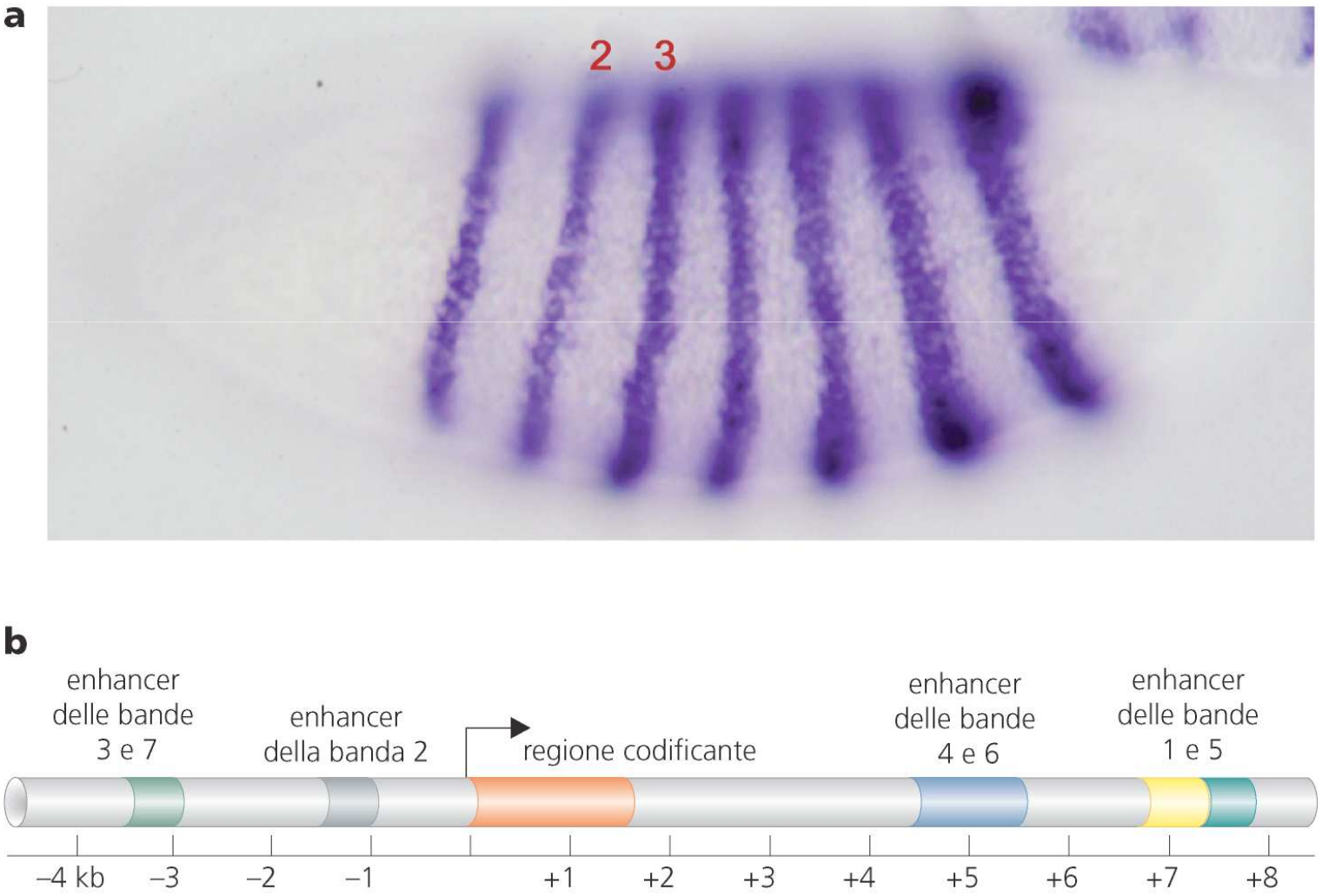


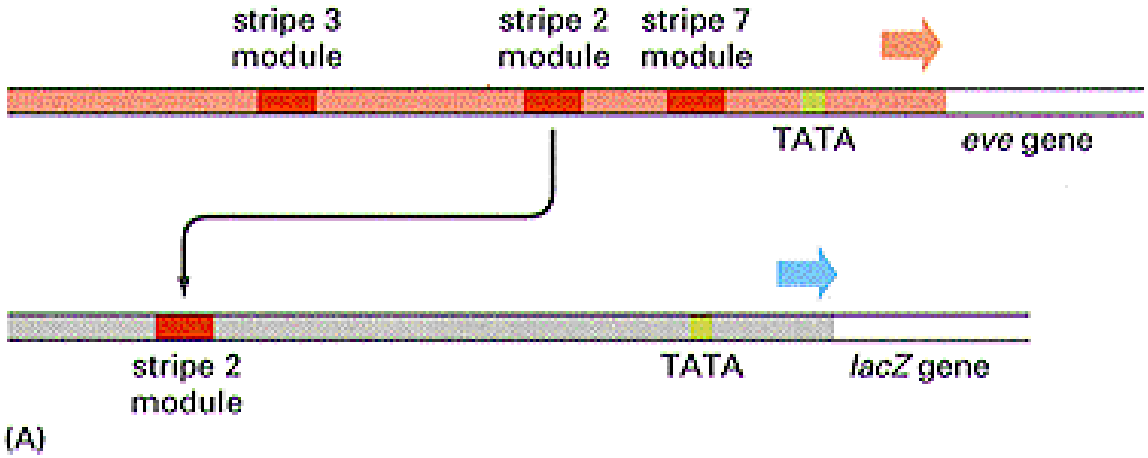


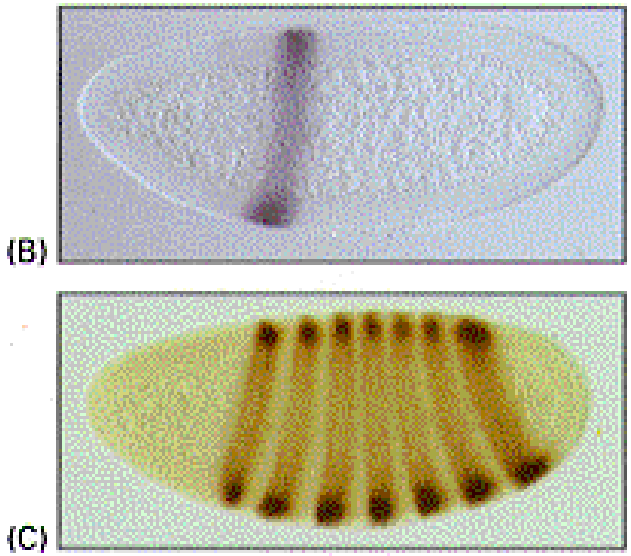


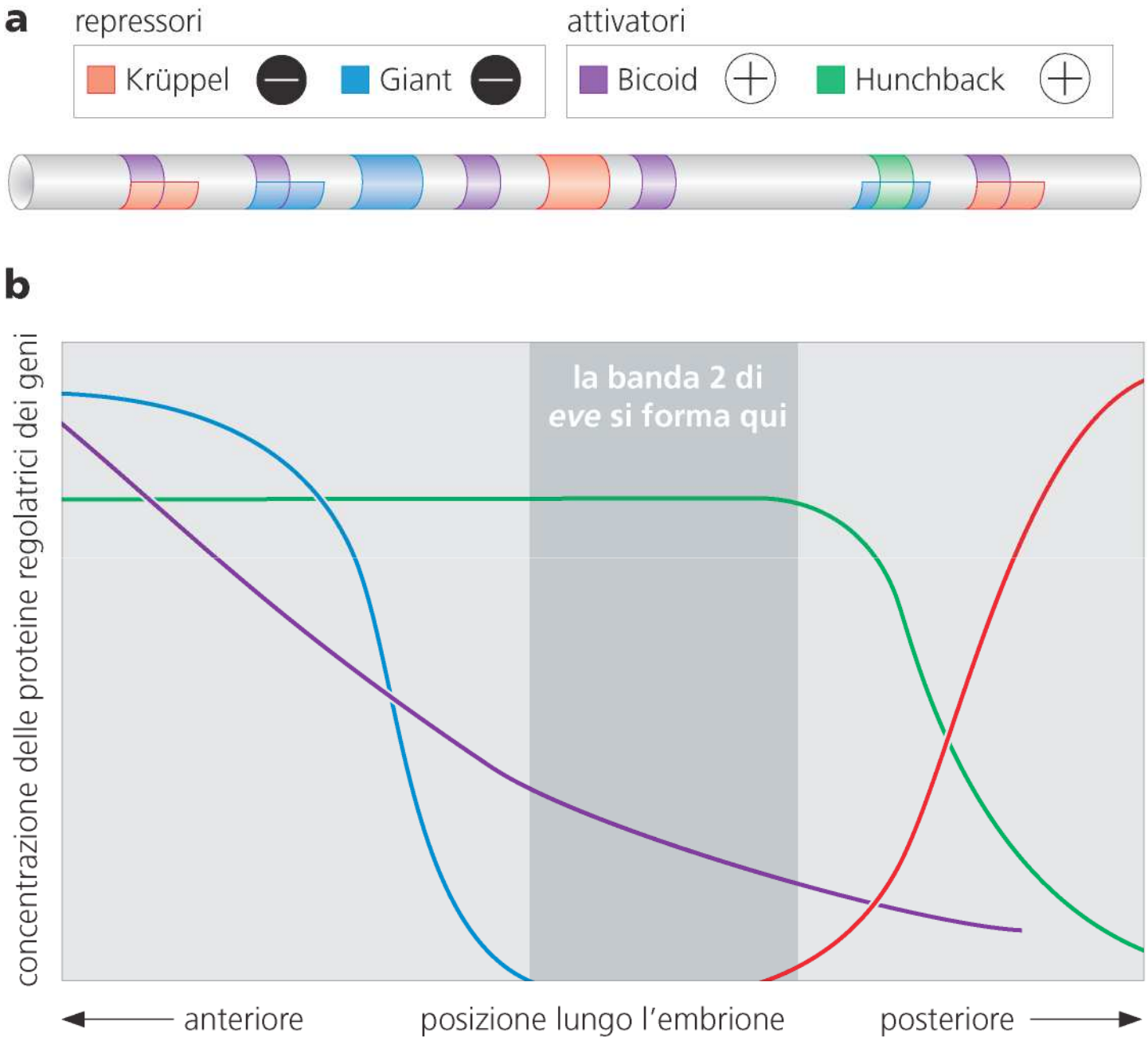
Hunchback is a repressor:
Kruppel has few sites, requires higher levels
Giant has more sites: lower levels are enough

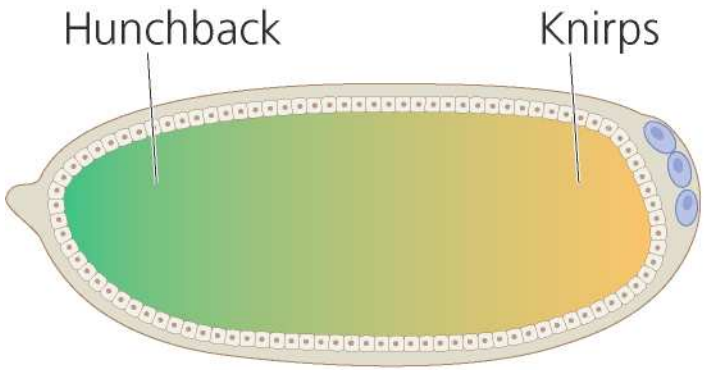
Eve (even-skipped) is the first “pair-rule” segmental gene: it has more than 12Kb essential regulatory sequences
Each enhancer is regulated by the exact combination of factors present in stripes

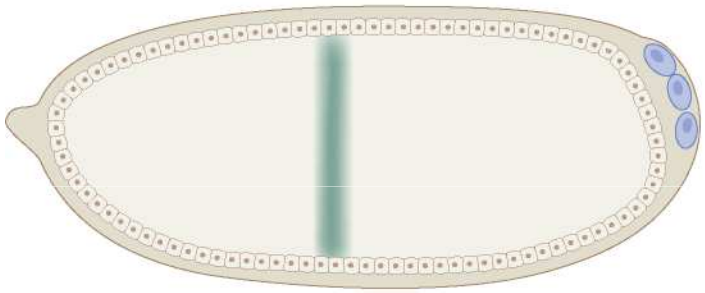






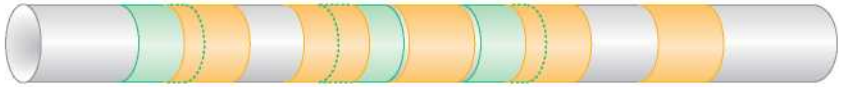

a
repressori
attivatori
Krüppel
Giant
Bicoid
Hunchback
b
la banda 2 di
eve si forma qui
concentrazione delle proteine regolatrici dei geni
anteriore
posizione lungo l'embrione
posteriore

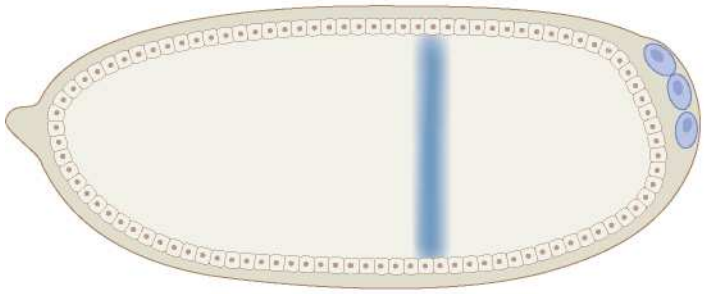




eve-3

enhancer della banda 3





eve-4

enhancer della banda 4