cis regulatory elements:

Switches to modulate the expression level of genes

Sebastiaan Meijsing
Transcriptional regulation group
Berlin, Germany
meijsing@molgen.mpg.de

Max Planck Institute
for Molecular Genetics

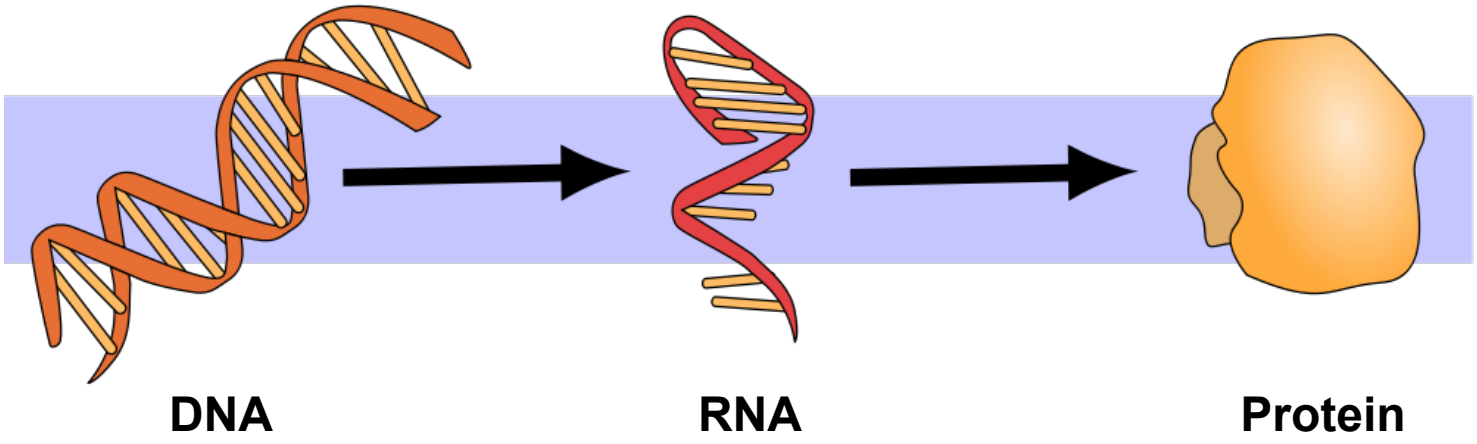
Outline

- Transcription
- Cis regulatory elements
- Transcription factors
- Chromatin

Outline

- **Transcription**
- Cis regulatory elements
- Transcription factors
- Chromatin

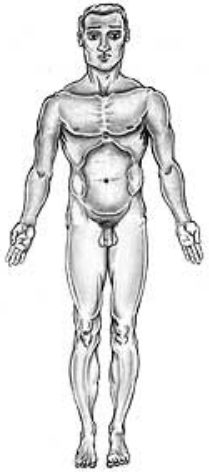
Transcription

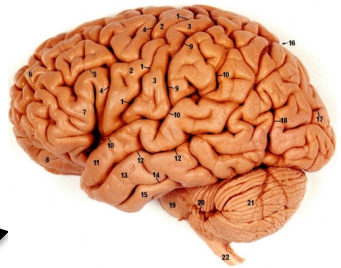


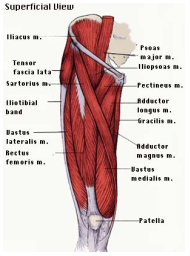
Central Dogma of Biology: DNA is transcribed into RNA which is translated into protein

Only ≈ 1% of genome codes for proteins

Transcription & being multicellular

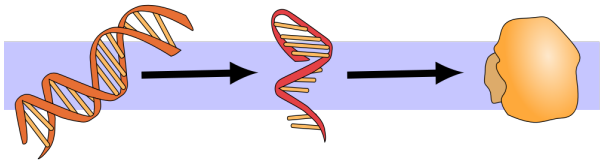


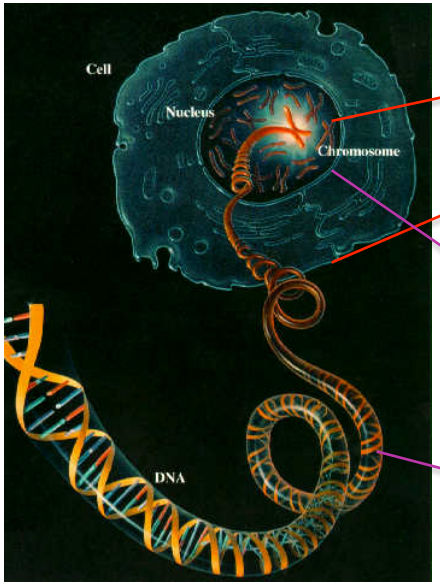


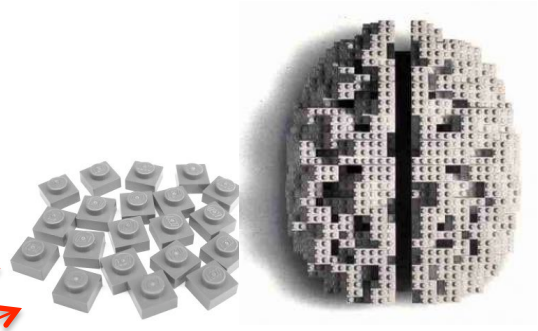


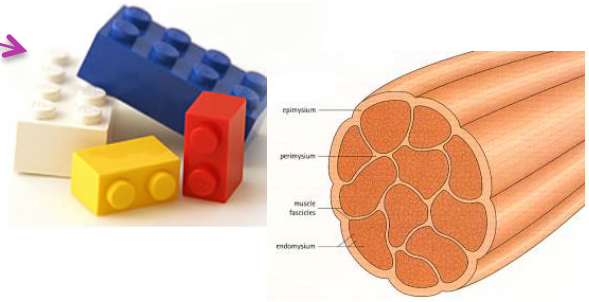
One genome: → **Different cell-types**

Transcription & being multicellular



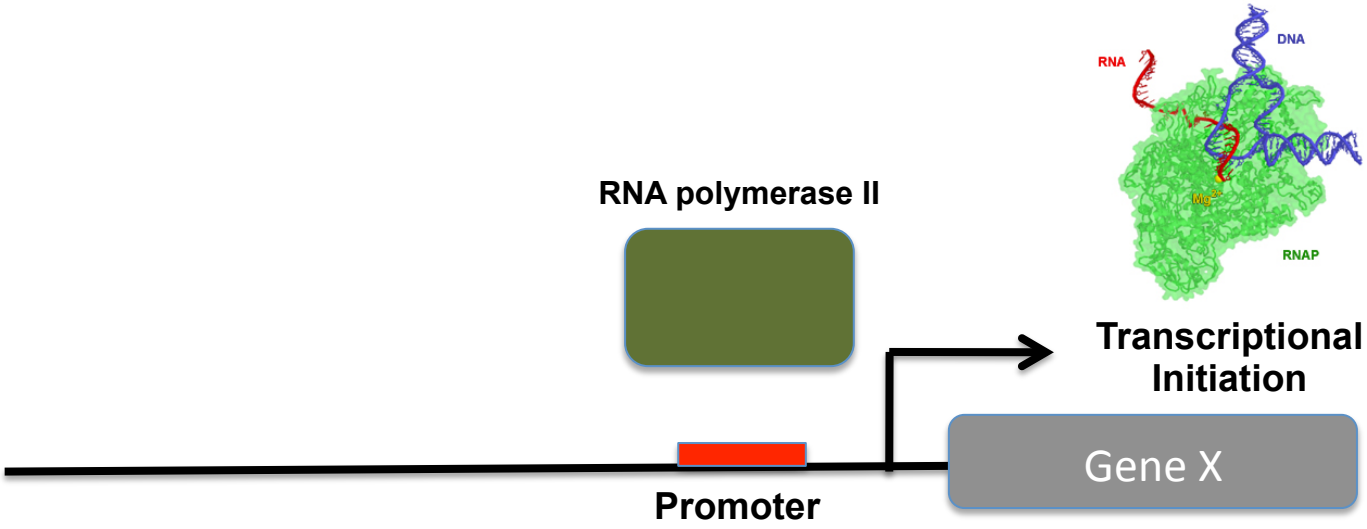






One genome: → **Different transcriptomes**

Transcription performed by RNA polymerases



Eukaryotic RNA polymerase-II alone is unable to bind DNA and relies on transcription factors & cis regulatory elements to initiate transcription

<u>2 flavors:</u>

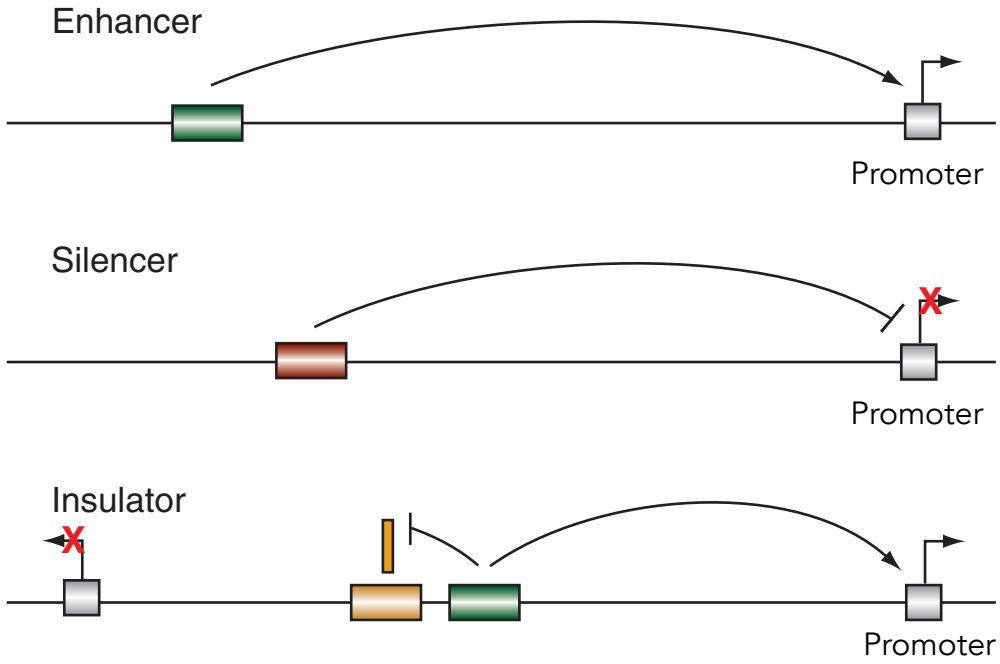
- General transcription factors (e.g. TATAA box binding protein (TBP)

- Transcriptional regulatory factors that regulate the expression of individual genes

Outline

- Transcription
- **Cis regulatory elements**
- Transcription factors
- Chromatin

Cis-regulatory elements: Different flavors



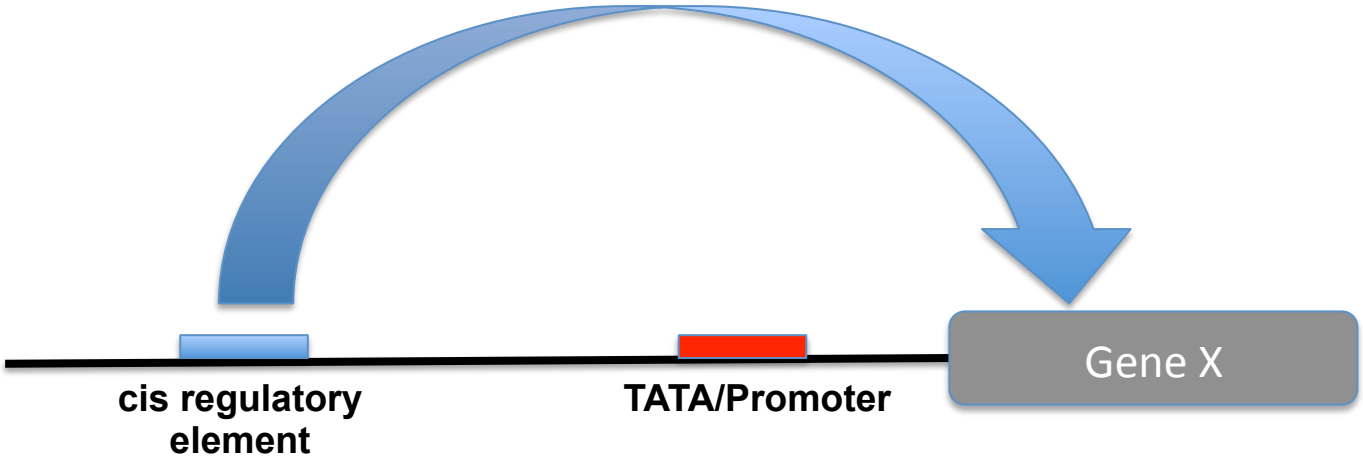
Maston *et al.*, Annu. Rev. Genom. Human Gent. (2006)

Today we mostly focus on cis regulatory elements that act as enhancers

Cis-regulatory elements

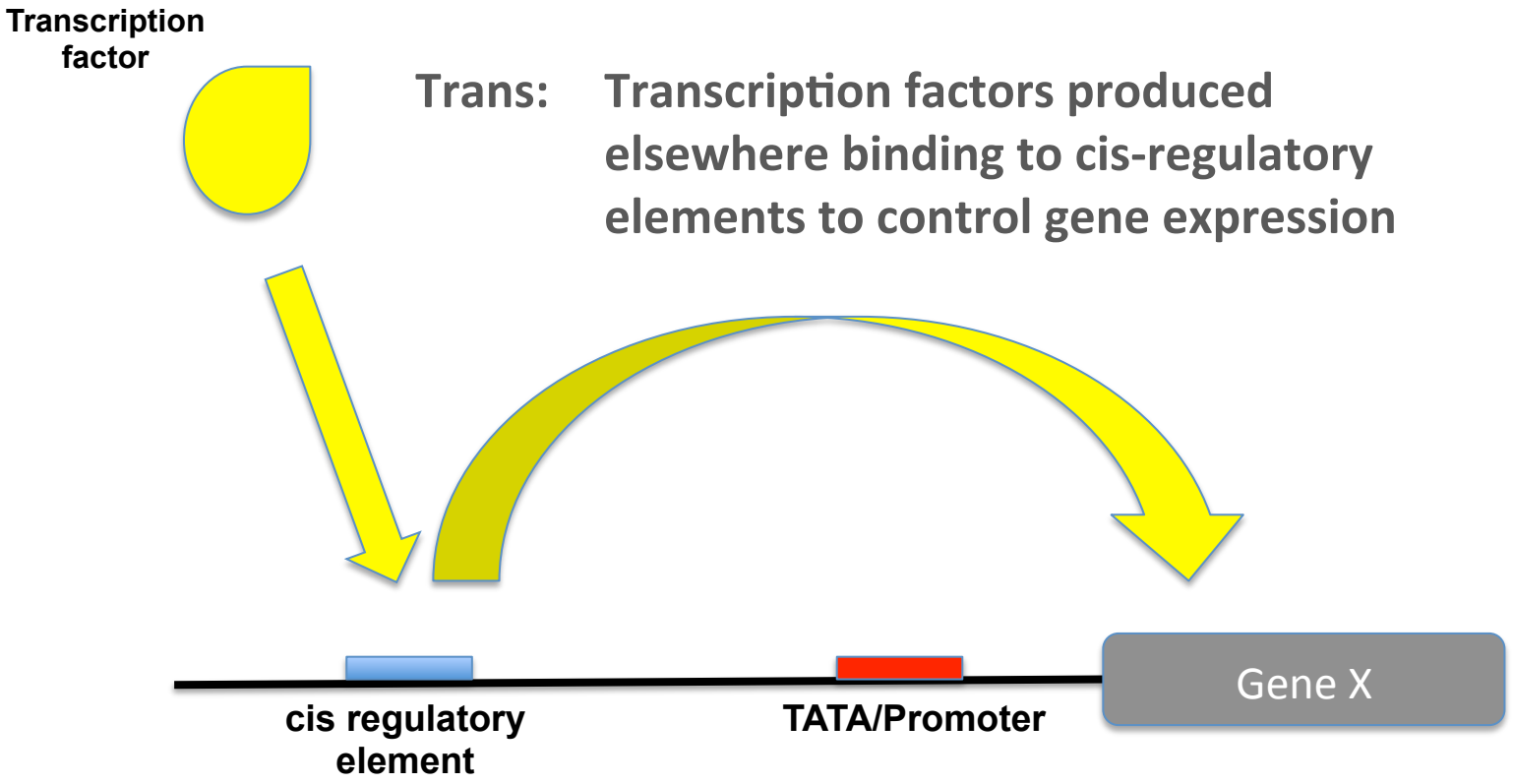
cis vs trans:

Cis: Directly acting on the nearby gene

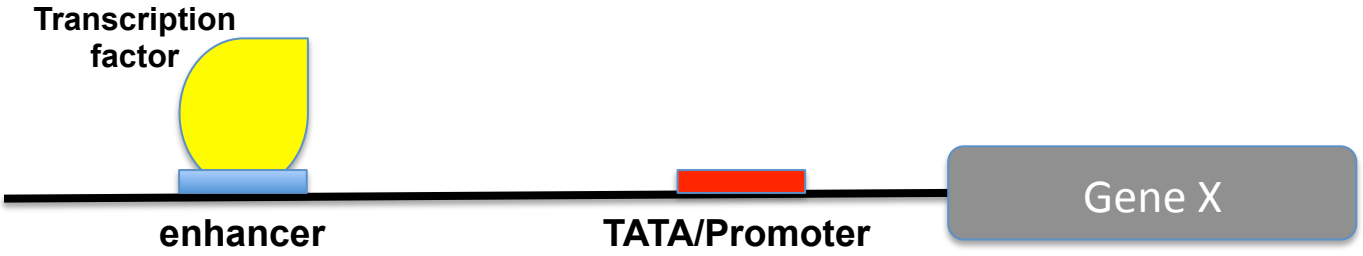


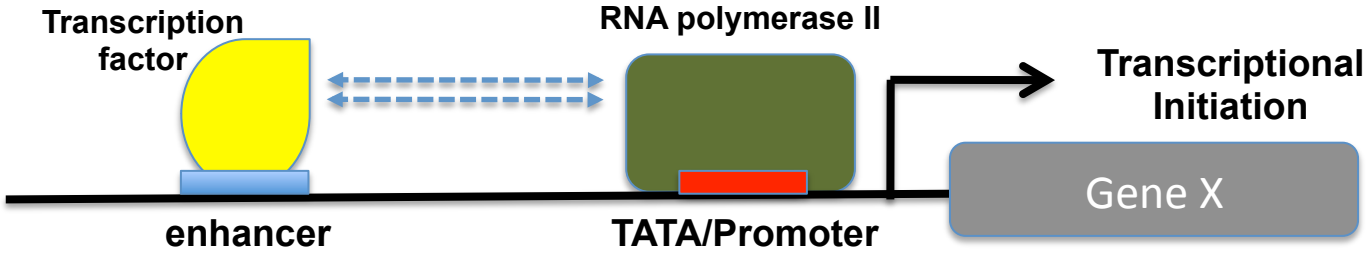
Cis-regulatory elements

cis vs trans:

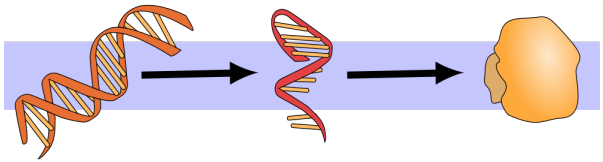


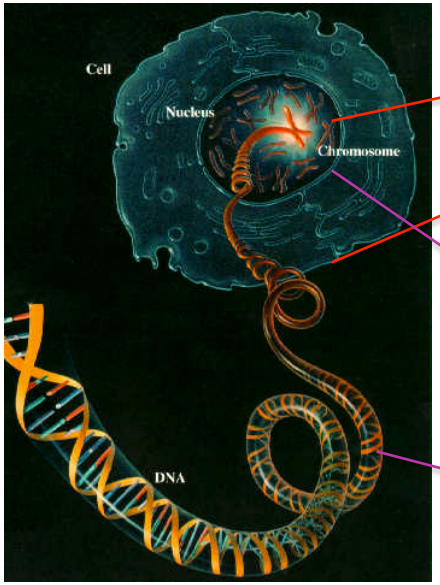
Transcription factors and transcriptional regulation

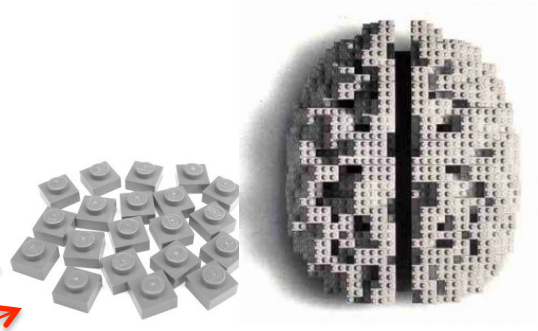


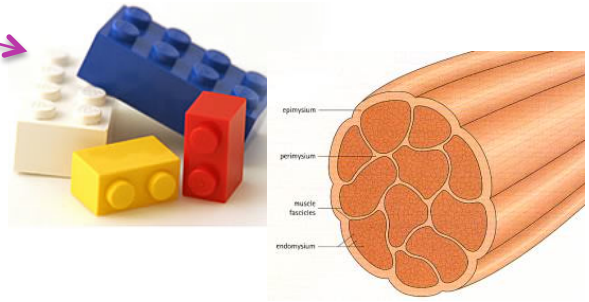


Transcription & being multicellular



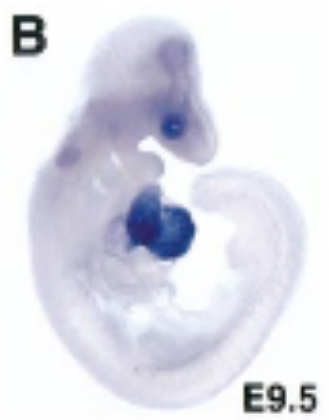




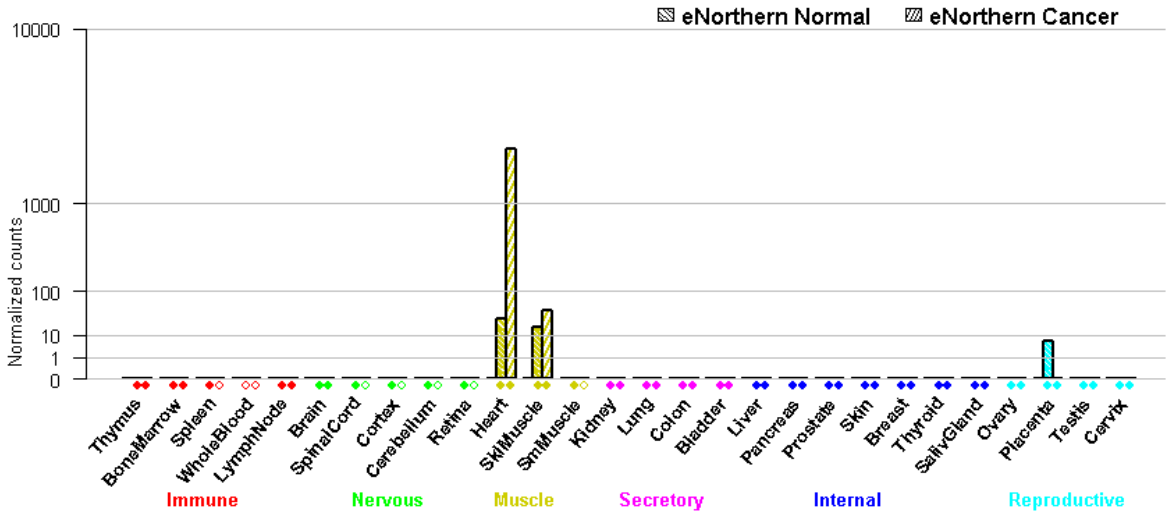


One genome: → **Different transcriptomes**

How to explain tissue-specific expression

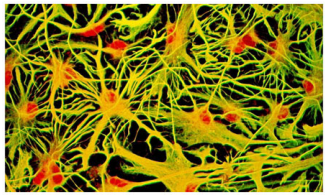


Liu et al., Dev. Bio. 2001

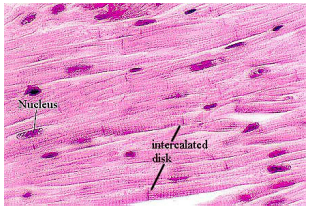


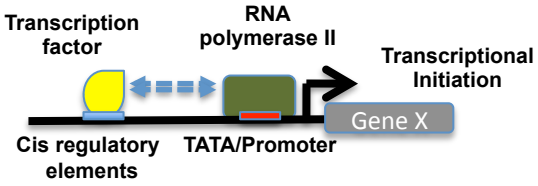
http://bioinfo2.weizmann.ac.il

Brain cell(s)



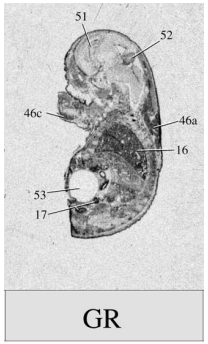
Muscle cell



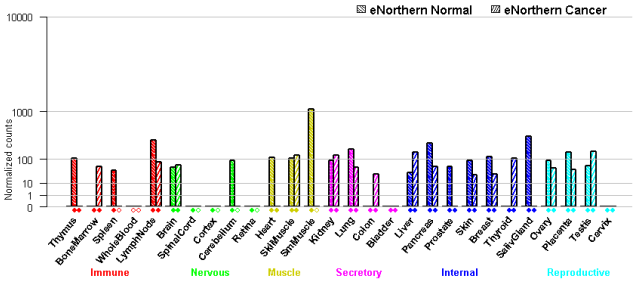


How to explain tissue-specific expression

Some factors are ubiquitously expressed yet target genes are tissue specific

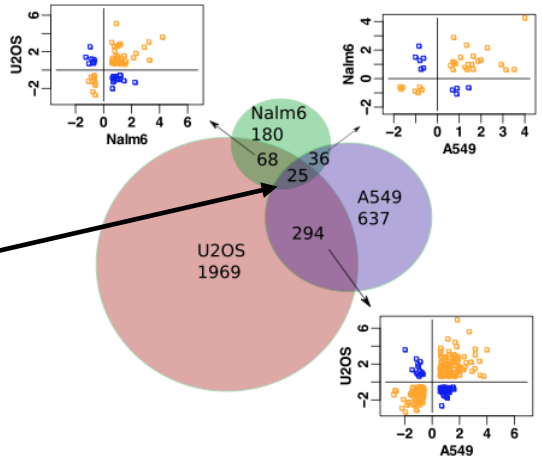


http://bioinfo2.weizmann.ac.il



Liu et al., Dev. Bio. 2001

Microarrays show little overlap in transcriptional regulation between different cell types

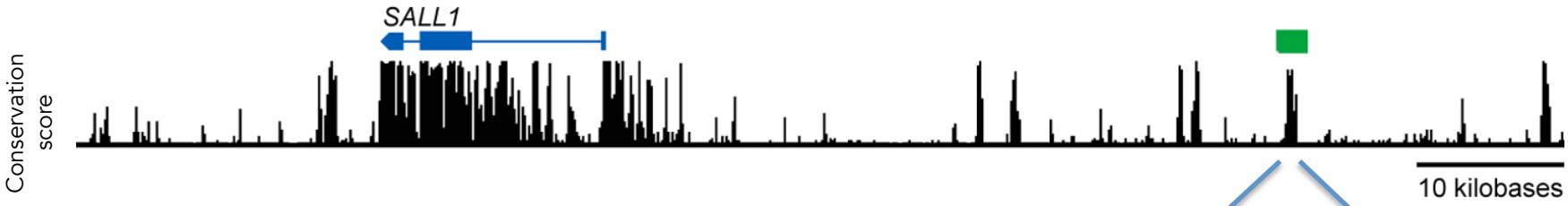


TF binding sites often clustered

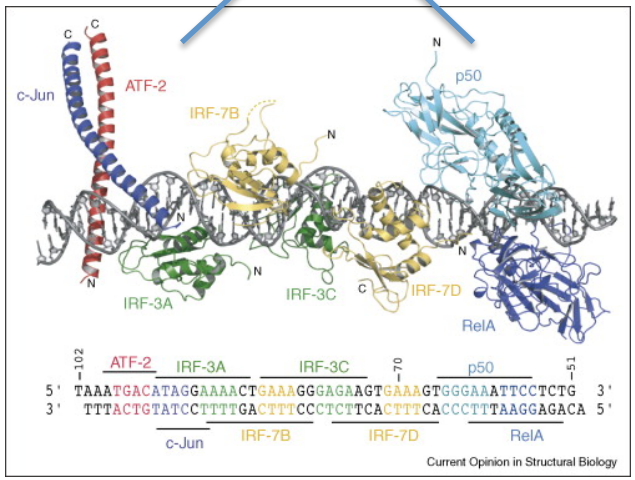
Conservation analysis can be used to identify enhancers.

Typically larger stretches of sequence conserved

Cis-regulatory modules (enhanceasomes)

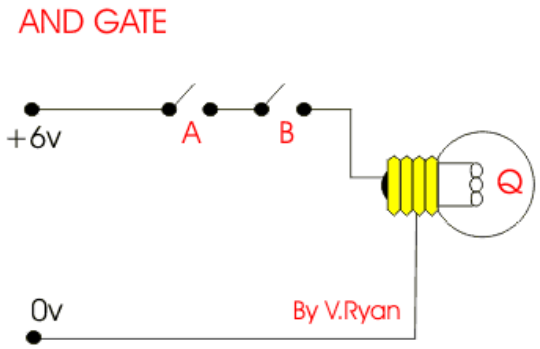


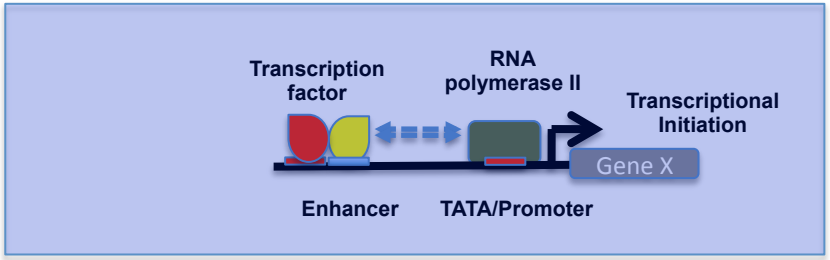
Petersen *et al.*, Plos One (2009)



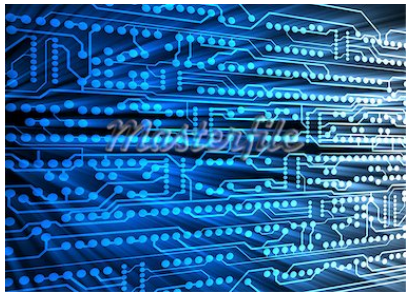
Panne Curr. Opinion in Struc. Biol. (2008)

How to explain tissue-specific expression-II



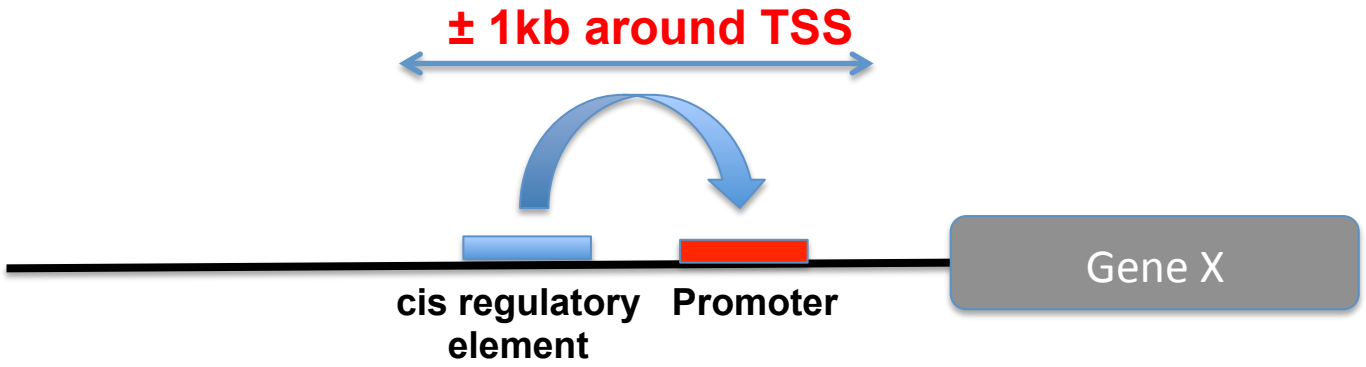


Modular nature of CRM
→ Different "switches"



Distal vs promoter proximal

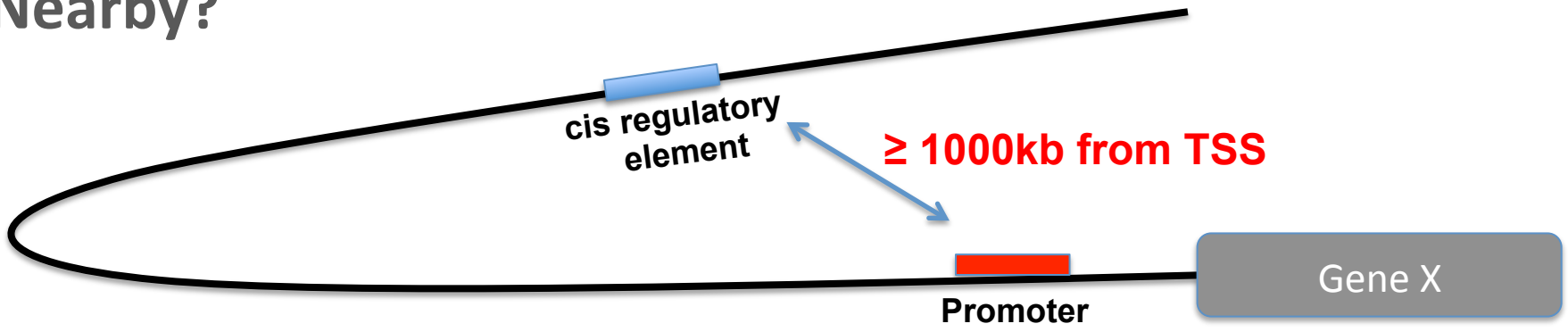
Nearby…..



Traditionally people looked at the promoter…….

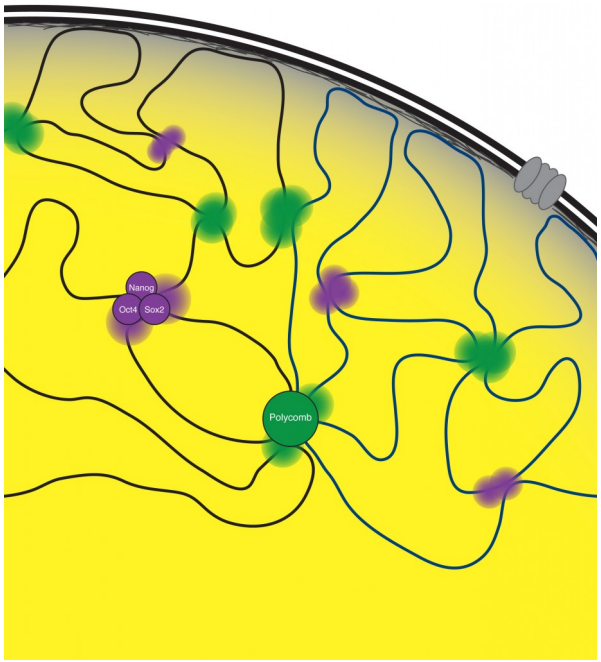
For many TFs majority of binding occurs >10kb away from promoters……

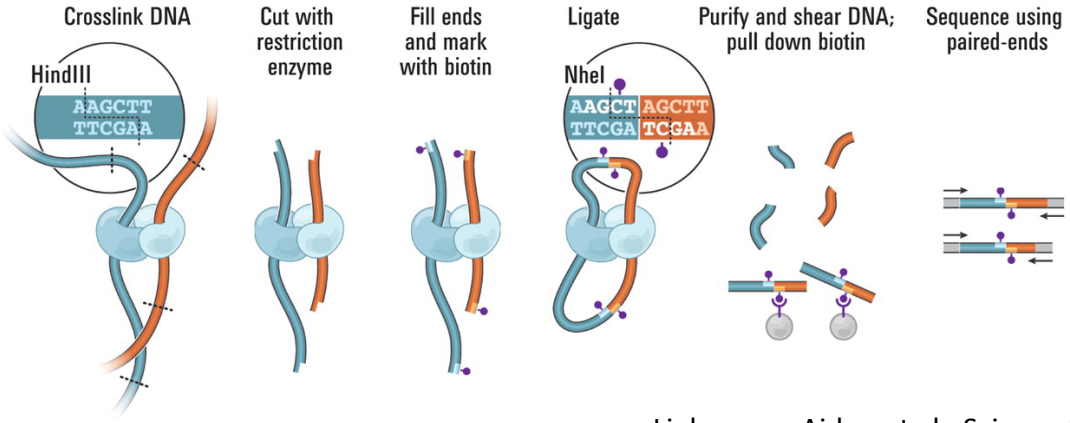
Nearby?



Distal vs promoter proximal

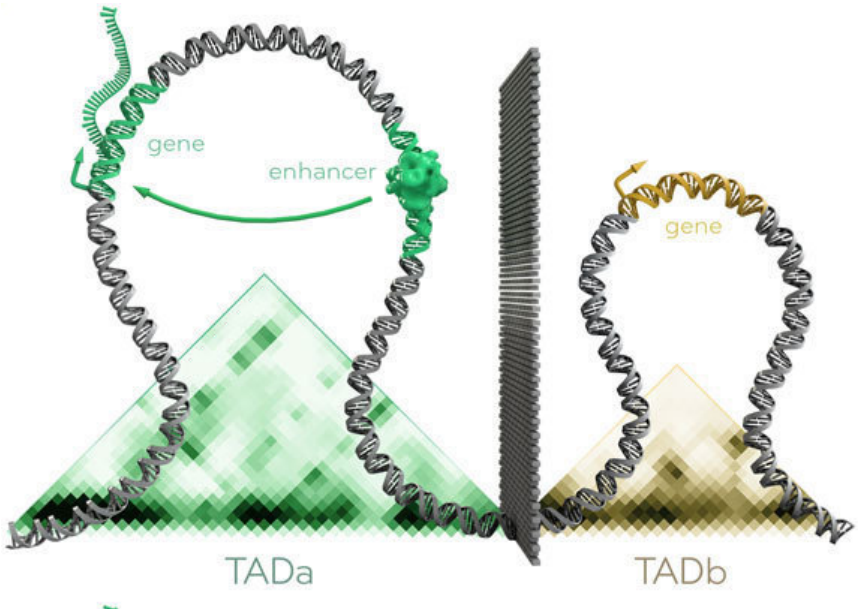
Chromatin Conformation capture (3C & related techniques):

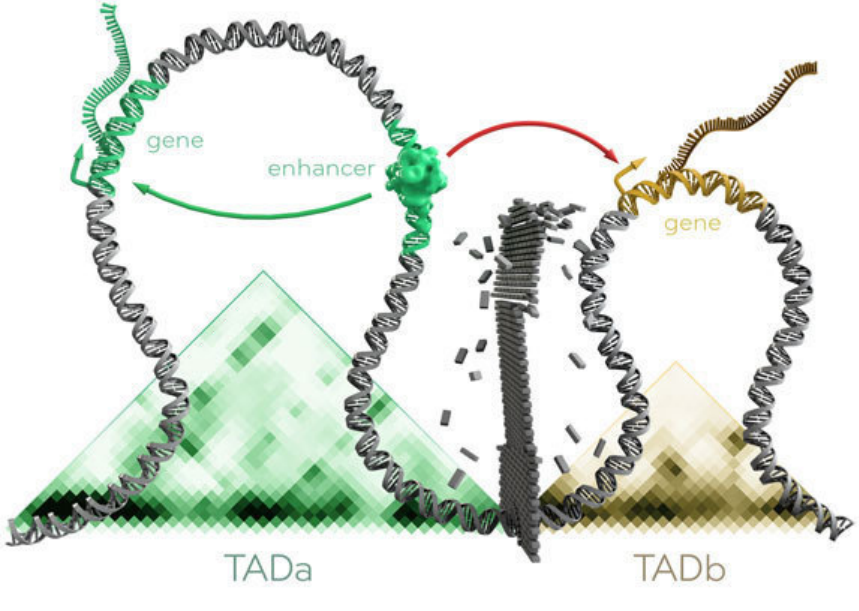




Lieberman-Aiden et al., Science 2012

Distal vs promoter proximal





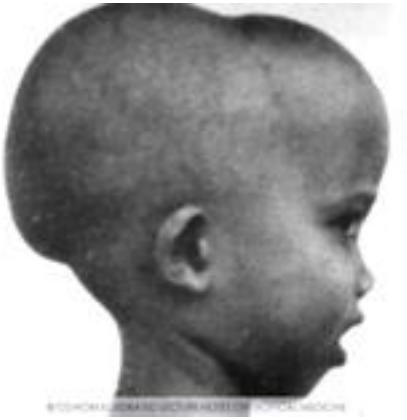
Lupianez et al., Cell 2015

Mutations in cis regulatory elements & disease

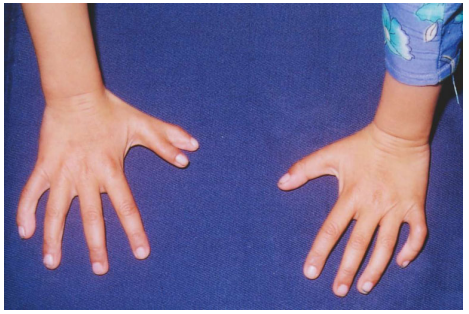
Table 1 Transcriptional regulatory elements involved in human diseases

Regulatory Element	Disease	Mutation (bound factor)	Affected Gene	Reference
Core promoter	β-thalassemia	TATA box, CACCC box, DCE	*β-globin*	(4, 94, 109)
Proximal promoter	Bernard-Soulier Syndrome	133 bp upstream of TSS (GATA-1)	*GpIbβ*	(117)
	Charcot-Marie-Tooth disease	215 bp upstream of TSS	*connexin-32*	(187)
	Congenital erythropoietic porphyria	70, 90 bp upstream of TSS (GATA-1, CP2)	uroporphyrinogen III synthase	(167)
	Familial hypercholesterolemia	43 bp upstream of TSS (Sp1)	*low density lipoprotein receptor*	(92)
	Familial combined hyperlipidemia	39 bp upstream of TSS (Oct-1)	*lipoprotein lipase*	(195)
	Hemophilia	CCAAT box (C/EBP)	*factor IX*	(43)
	Hereditary persistence of fetal hemoglobin	~175 bp upstream of TSS (Oct-1, GATA-1)	*Aγ-globin*	(62)
	Progressive myoclonus epilepsy	Expansion ~70 bp upstream of TSS	*cystatin B*	(96)
	Pyruvate kinase deficient anemia	72 bp upstream of TSS (GATA-1)	*PKLR*	(120)
	β-thalassemia	CACCC box (EKLF)	*β-globin*	(130)
	δ-thalassemia	77 bp upstream of TSS (GATA-1)	*δ-globin*	(125)
	Treacher Collins syndrome	346 bp upstream of TSS (YY1)	*TCOF1*	(123)
Enhancer	Preaxial polydactyly	1 Mb upstream of gene	*SHH*	(107)
	Van Buchem disease	Deletion ~35 kb downstream of gene	*sclerostin*	(116)
	X-linked deafness	Microdeletions 900 kb upstream	*POU3F4*	(46)
Silencer	Asthma and allergies	509 bp upstream of TSS (YY1)	*TFG-β*	(78)
	Fascioscapulohumeral muscular dystrophy	Deletion of D4Z4 repeats	4q35 genes	(66)
Insulator	Beckwith-Wiedemann syndrome	CTCF binding site (CTCF)	*H19/Igf*	(147)
LCR	α-thalassemia	62 kb deletion upstream of gene cluster	*α-globin* genes	(75)
	β-thalassemia	~30 kb deletion removing 5′HS2–5	*β-globin* genes	(52)

Maston *et al.*, Annu. Rev. Genom. Human Gent. (2006)



ß-Thalassemia

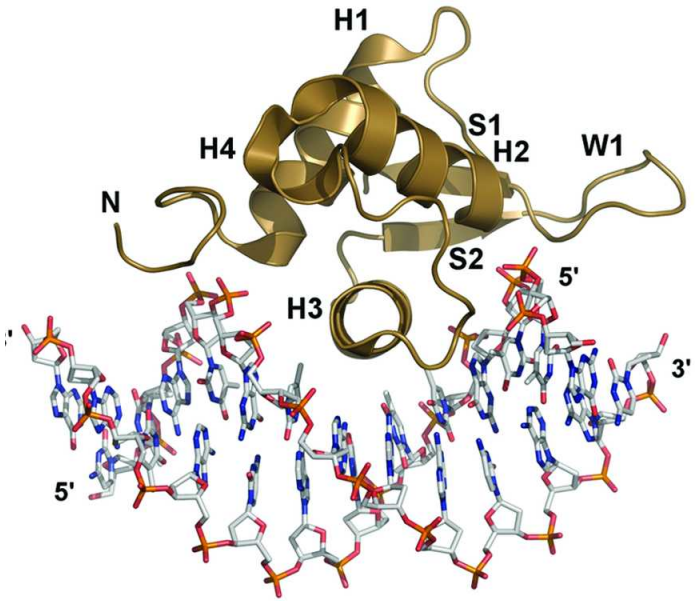


Polydactyly

Outline

- Transcription
- Cis regulatory elements
- **Transcription factors**
- Chromatin

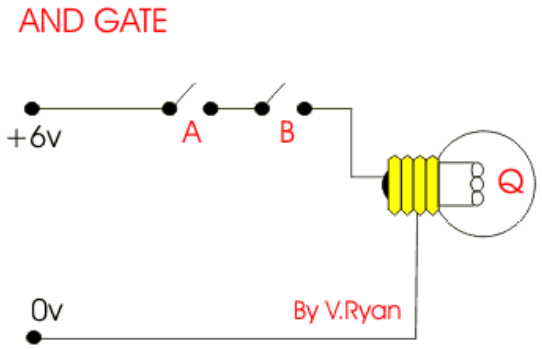
Cis-regulatory elements: contain zip codes for TFs

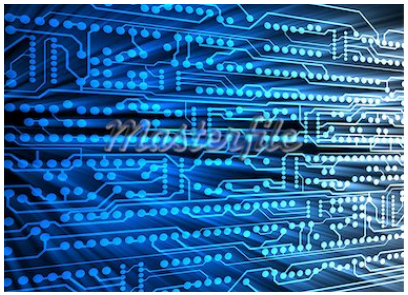


A A A T A A A C A n n

Transcription factors

- **Approx. 3000 in human genome (approx 1/6th of all coding genes)**





Lots of switches allowing cells to run all sorts of different "programs".

Transcription factors

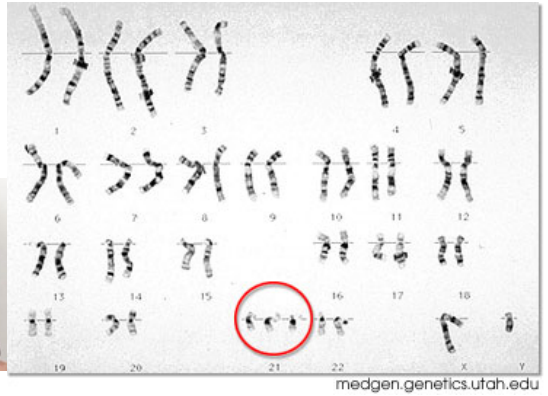
Fine tuning: Getting gene dosage just right is important

Too much: Trisomy 21 (down syndrome)

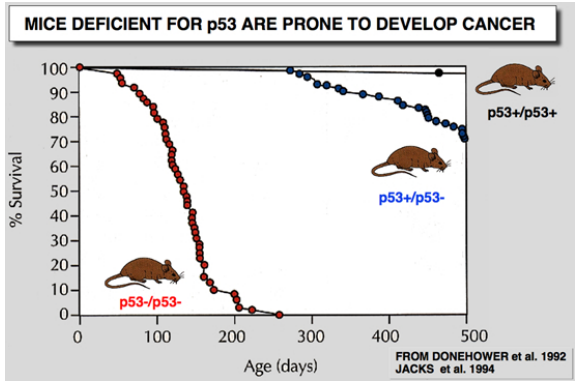
(X-inactivation woman)

Copy number variation linked to various disease:

autism, schizophrenia, systemic lupus erythematosis, Crohn's disease and psoriasis



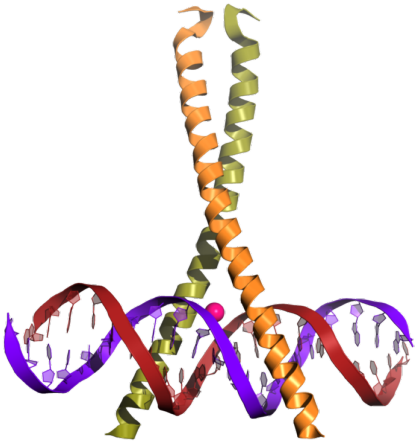
Too little: p53 and cancer



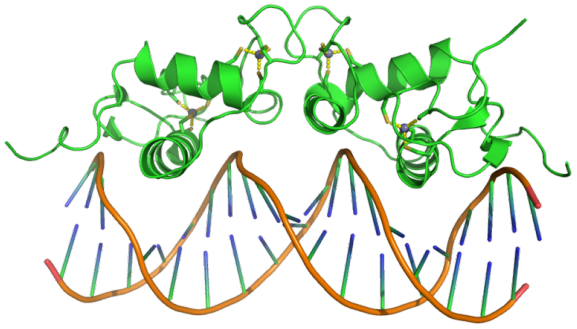
Transcription factors

- **Approx. 3000 in human genome (approx 1/6th of all coding genes)**

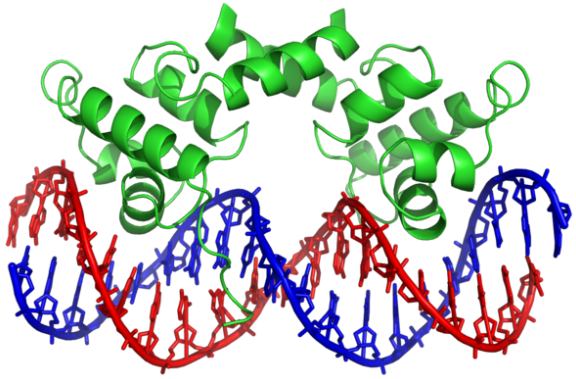
Common feature: TFs recognize DNA using different types of DNA binding domains



Basic leucine zipper domain



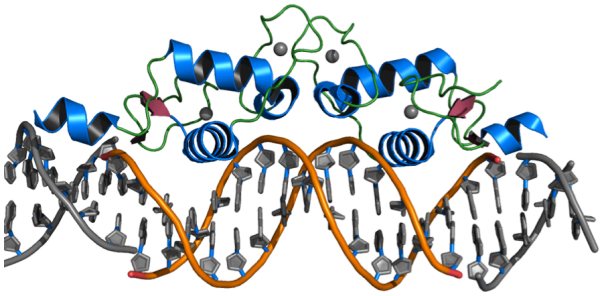
Zinc finger domain



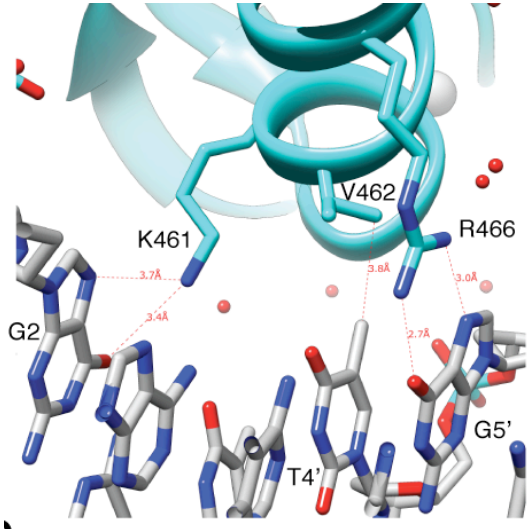
Helix-turn-Helix

Transcription factors

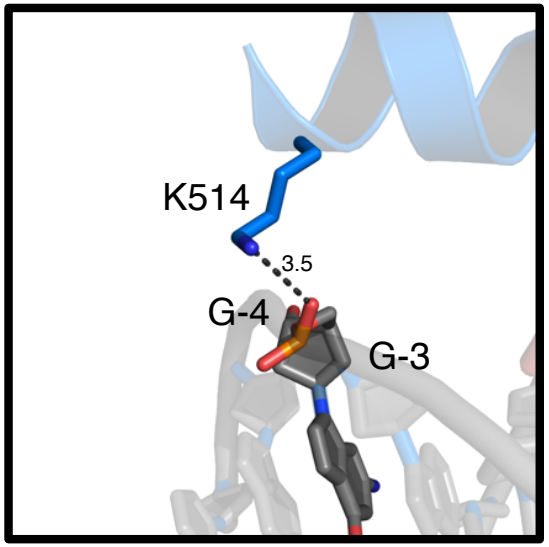
DNA recognition: Specific contacts & non-specific contacts



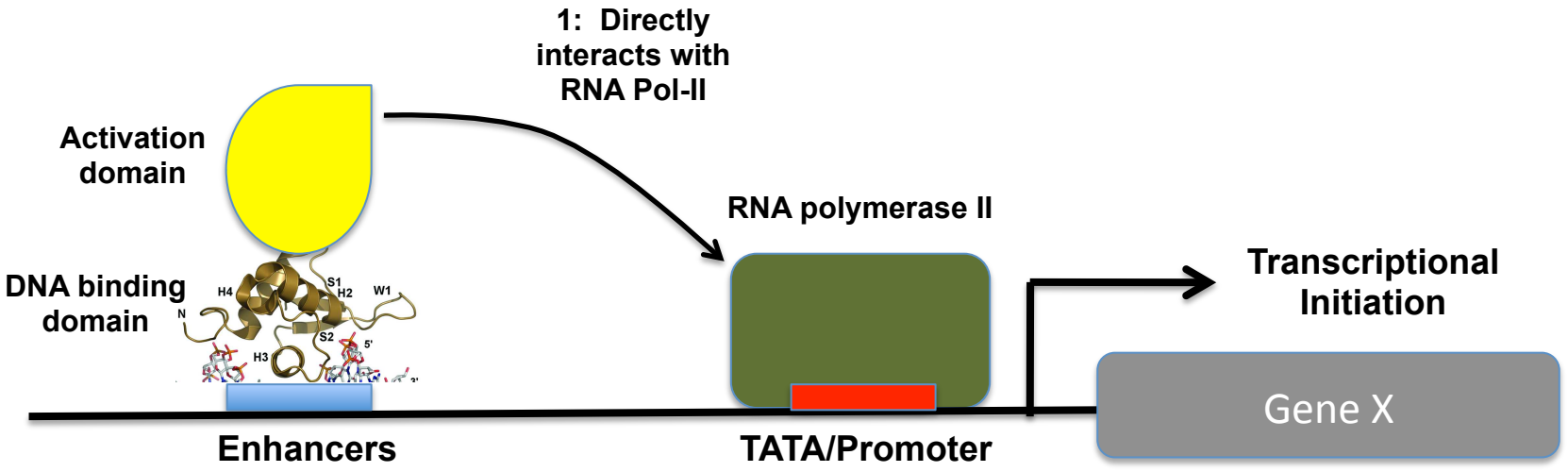
Specific contacts
(recognize bases in major groove)
Sequence specific



Non-specific contacts
(e.g. DNA backbone contacts)
Not sequence specific

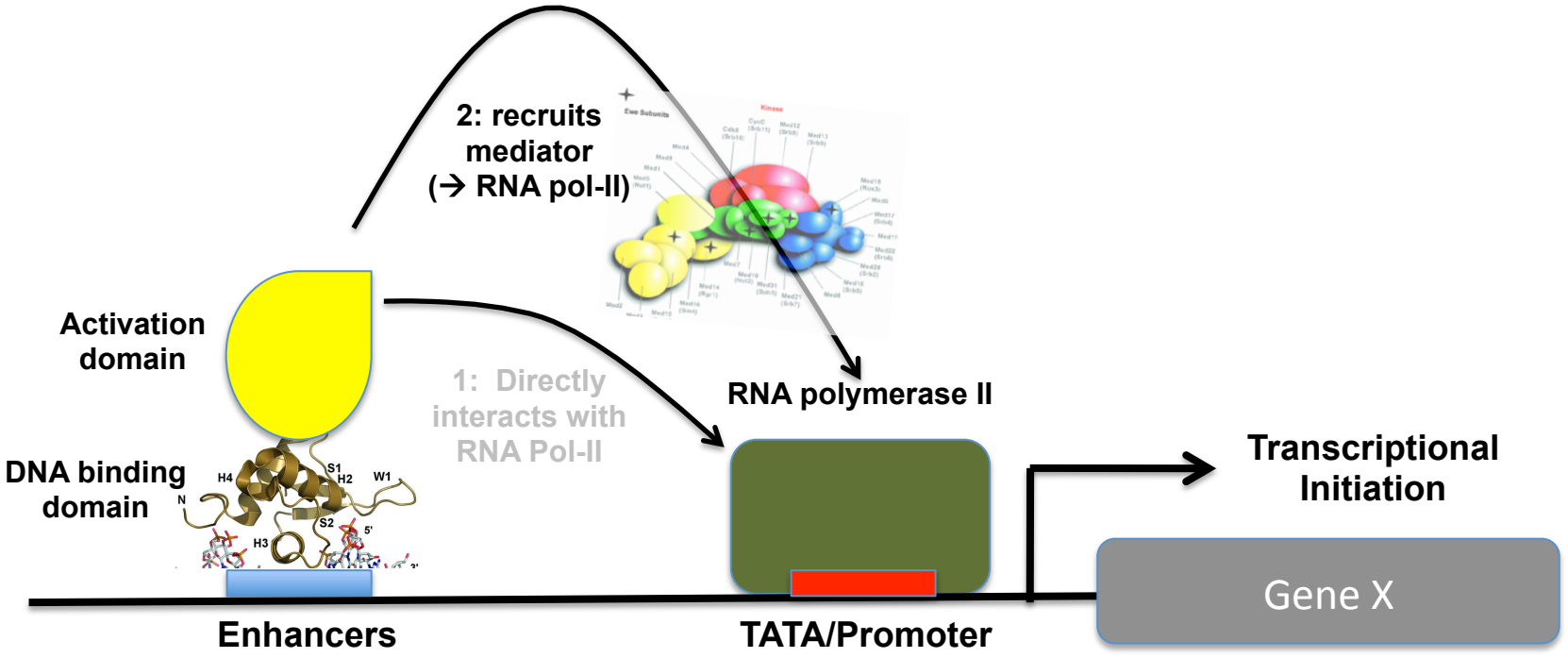


TFs in turn facilitate RNA pol-II recruitment



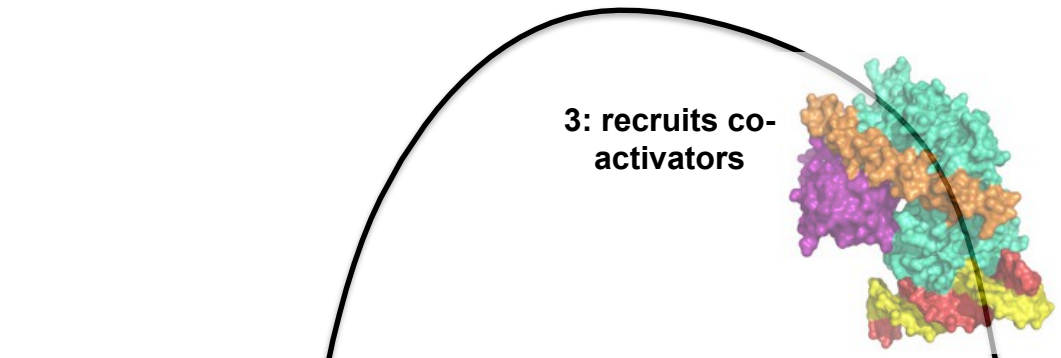
Maston *et al.*, Annu. Rev. Genom. Human Gent. (2006)

TFs in turn facilitate RNA pol-II recruitment-2



Maston *et al.*, Annu. Rev. Genom. Human Gent. (2006)

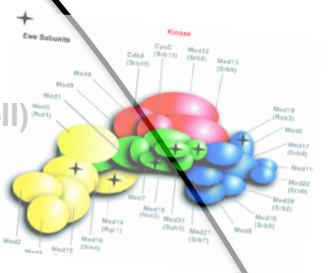
TFs in turn facilitate RNA pol-II recruitment-3



3: recruits co-activators

Co-activators:

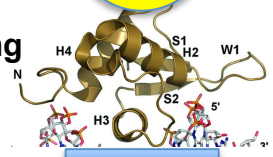
- Interact with Mediator
- Interact with Pol-II
- Chromatin modifying enzymes



2: recruits mediator (→ RNA pol-II)

Activation domain

DNA binding domain



1: Directly interacts with RNA Pol-II

RNA polymerase II

Transcriptional Initiation

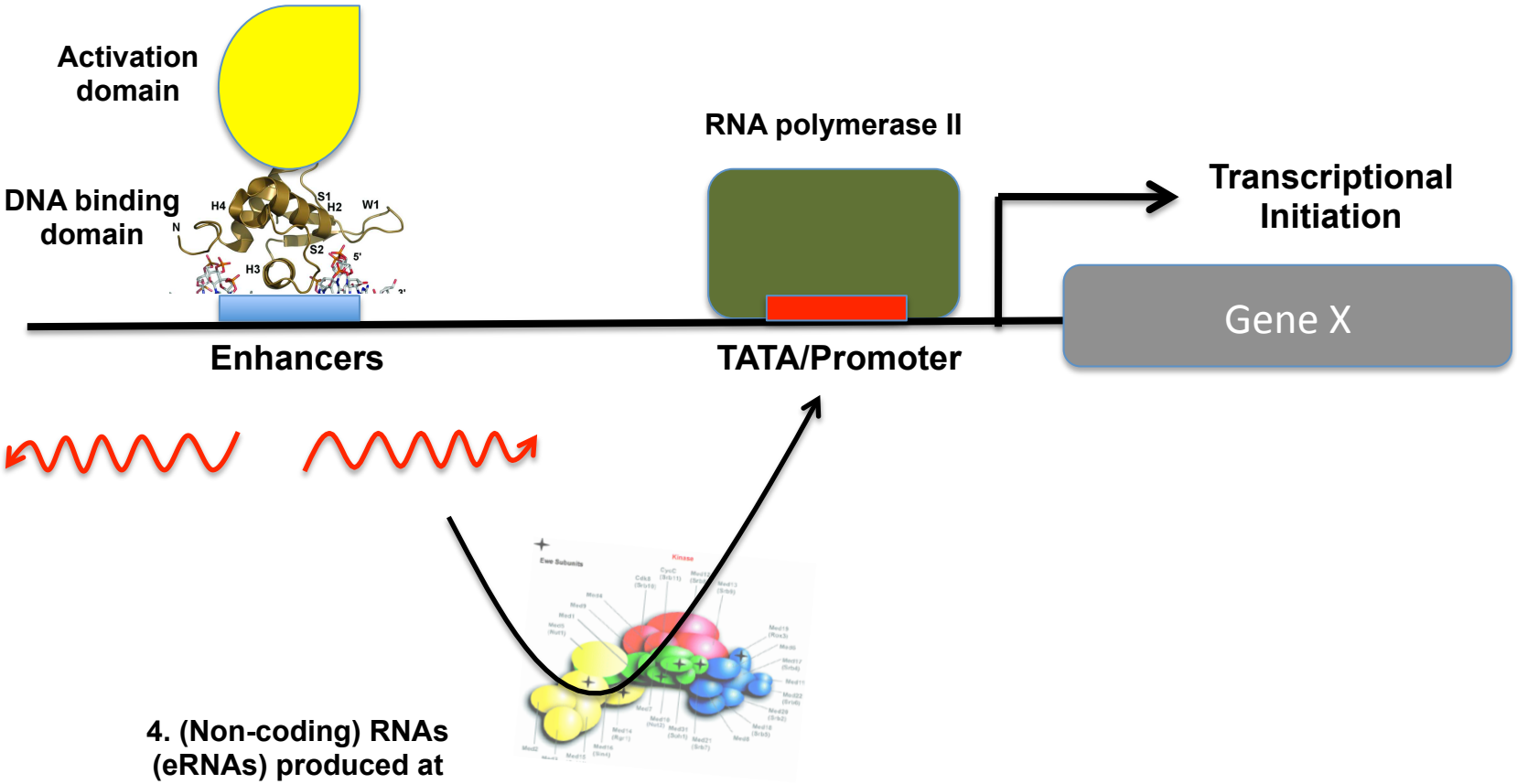
Gene X

Enhancers

TATA/Promoter

Maston *et al.*, Annu. Rev. Genom. Human Gent. (2006)

TFs in turn facilitate RNA pol-II recruitment-4



4. (Non-coding) RNAs (eRNAs) produced at enhancers facilitate interaction with promoter (e.g. by interacting with mediator or keeping chromatin "open")

Li et al., Nature 2013
Lai et al., Nature 2013

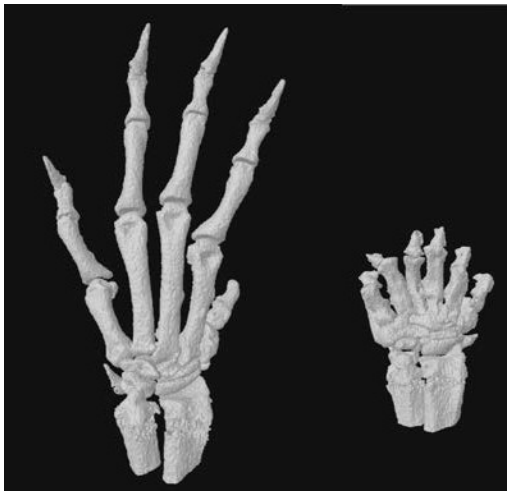
Mutations in transcription factors & disease

1: **Mutations resulting in loss of expression**

2: **Mutations resulting in loss/change of function**

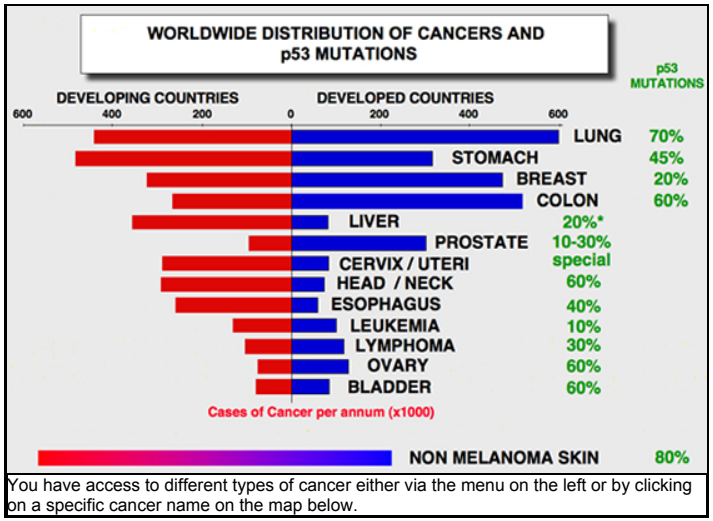
3: **Translocations directing TFs to wrong genomic location**

Developmental defects



Digits in wt (left) and Hoxd13 mutant (right) mice. In the mutant the N-terminal repeat has been expanded by 21 alanines

Cancer:



http://p53.free.fr/

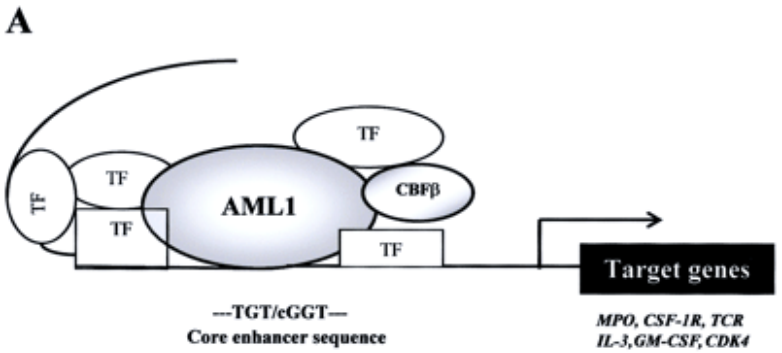
Mutations in transcription factors & disease

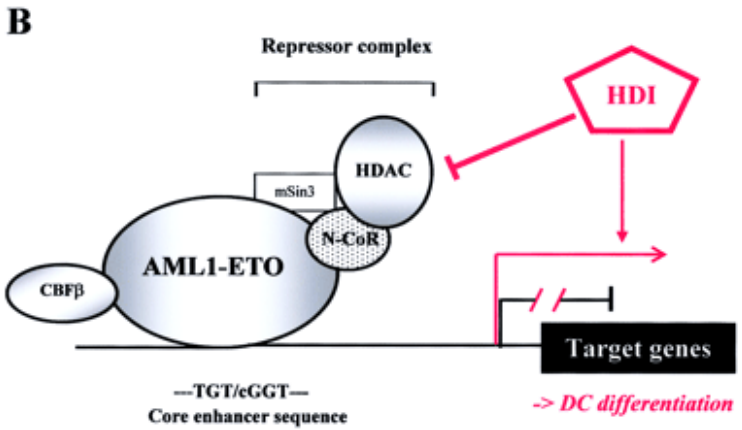
1: Mutations resulting in loss of expression

2: Mutations resulting in loss of function

3: Translocations can mess up the normal program induced by TF
(& result in disease)

Various types of cancer (MLL gene fusion proteins due to translocations result in various types of leukemia)





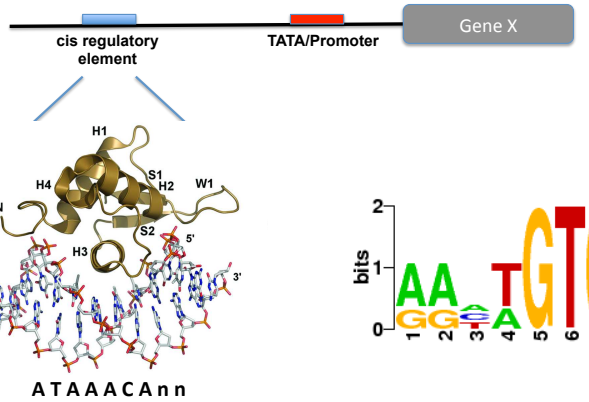
Moldenhauer et al., JLB. (2004)

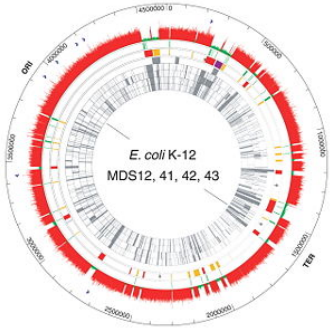
Outline

- Transcription
- Cis regulatory elements
- Transcription factors
- **Chromatin**

How do transcription factors “know” where to go?

Prokaryotes:





LacI (transcriptional repressor of lac operon)

Size genome: approx. 5 x 10^6 bp

Motif frequency $1/4^{10}$

Binding sites / genome: ≈5

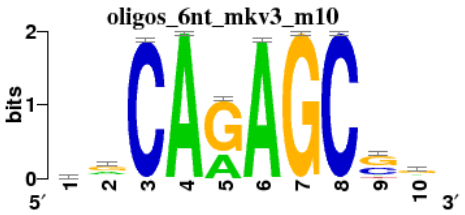
Binding site sequence accurately predict where TFs bind

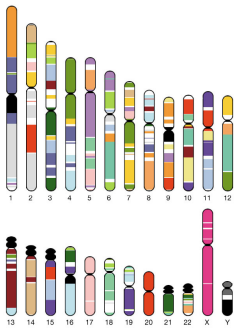
How do transcription factors “know” where to go?

Eukaryotes:

Typical recognition sequence for eukaryotic transcription factors is shorter

Genome is larger……. (x1000)





Smad3 (transcriptional factor)

Size genome: approx. 3 x 10^9 bp

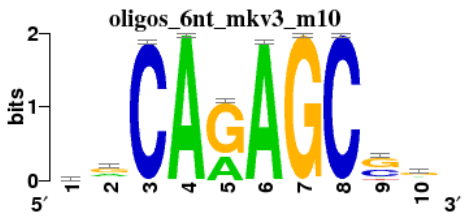
Motif frequency 1/4^5 (1 every kb)

Binding sites / genome: > 3.000.000

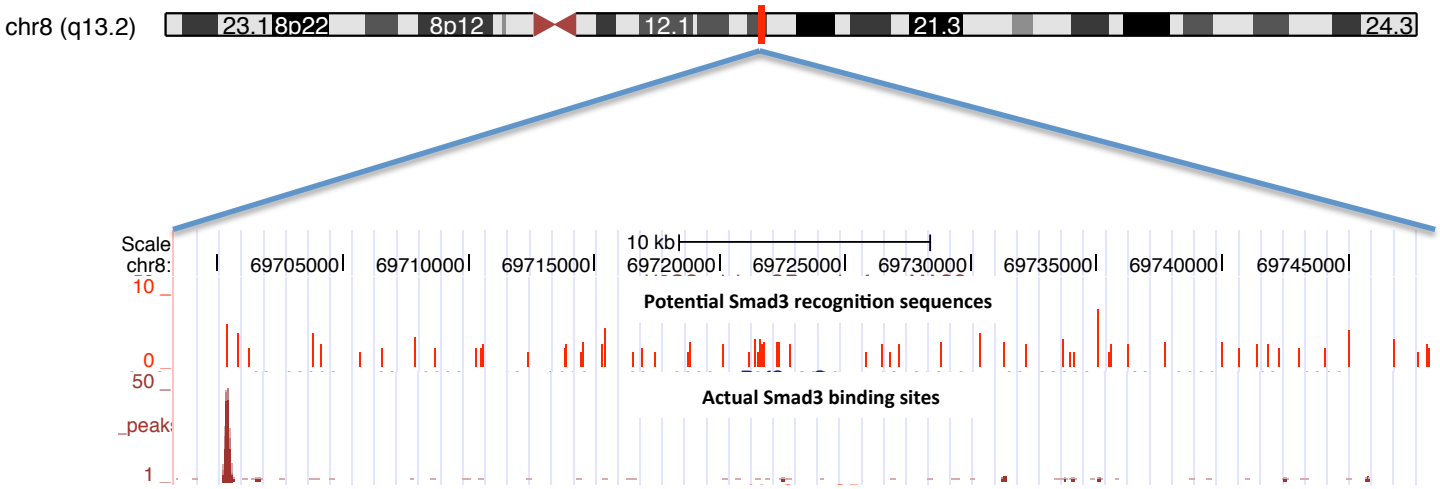
Transcription factors/ cell: 20.000

How do transcription factors “know” where to go?

Smad3



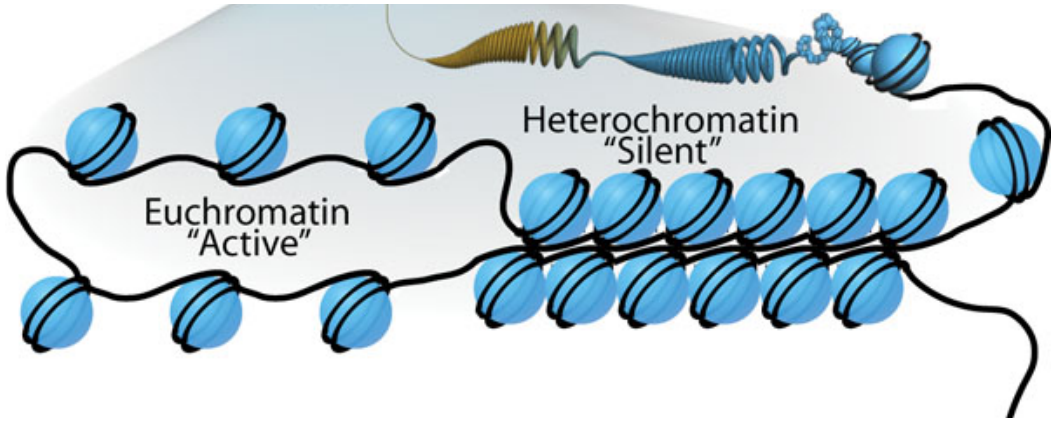
Sequence found in genome more than once every 1000bp



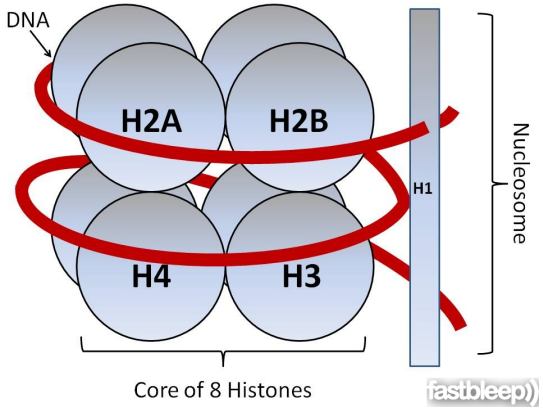
Binding site sequence alone insufficient to predict where TFs bind

Something is missing……..

Chromatin!

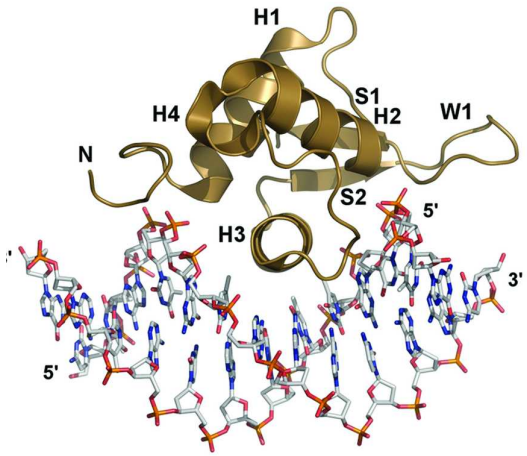


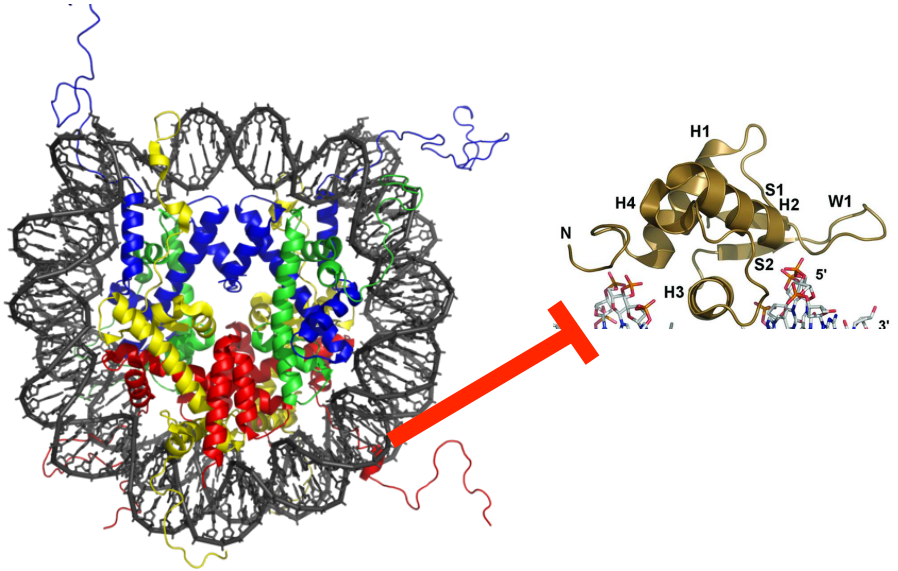
In Eukaryotes DNA is „packaged“ into nucleosomes



Nucleosomes interfere with TF binding

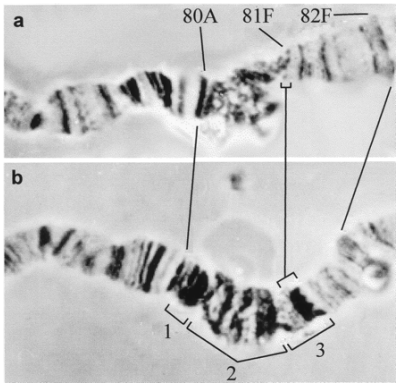
Barrier 1: DNA is „packaged“ into nucleosomes



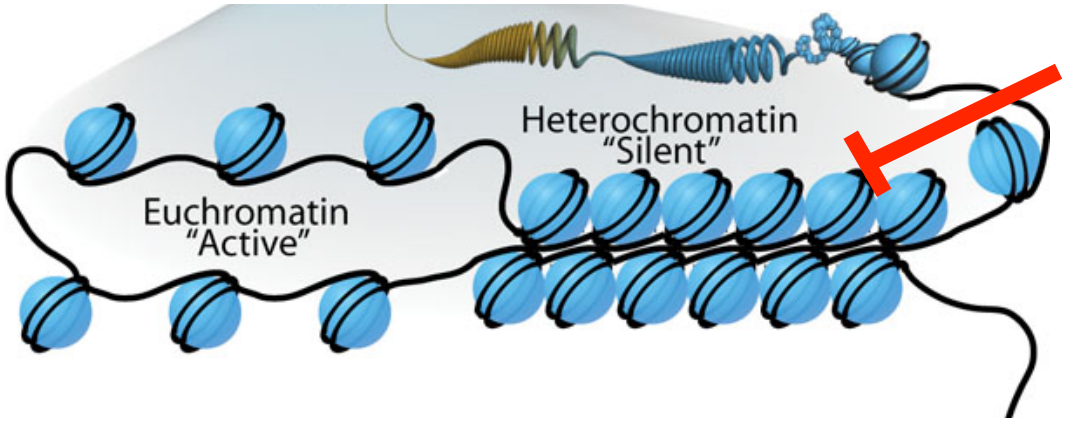


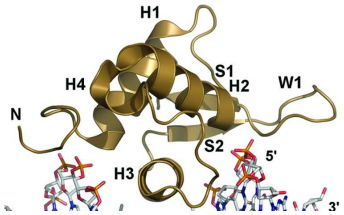
Chromatin interferes with TF binding

Barrier 2: Genome is partitioned into „open“ & „closed chromatin“ called Heterochromatin or Euchromatin



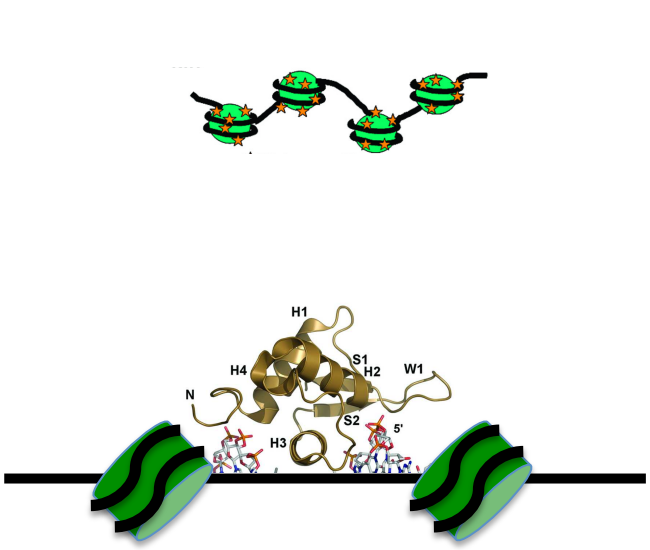
Heintz (1928) / Belyaeva et al., *PNAS* (1998)

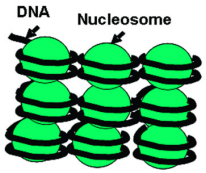


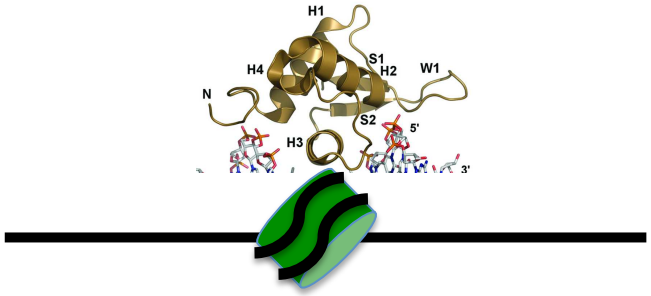


Cis-regulatory element function

Most transcription factors bind to „open“ chromatin

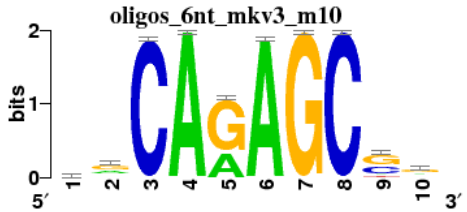




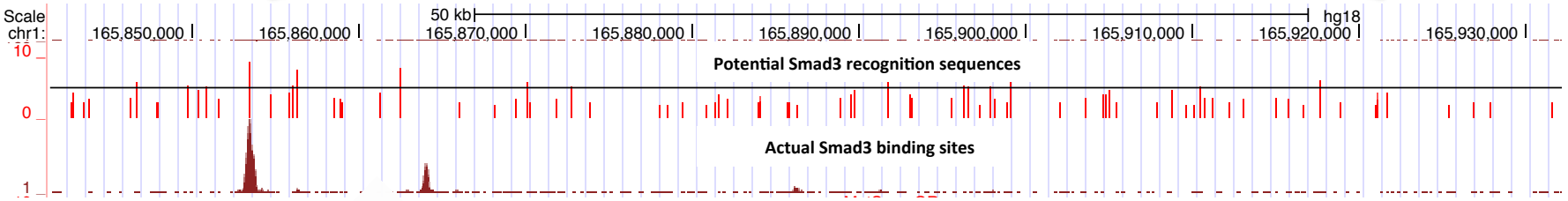


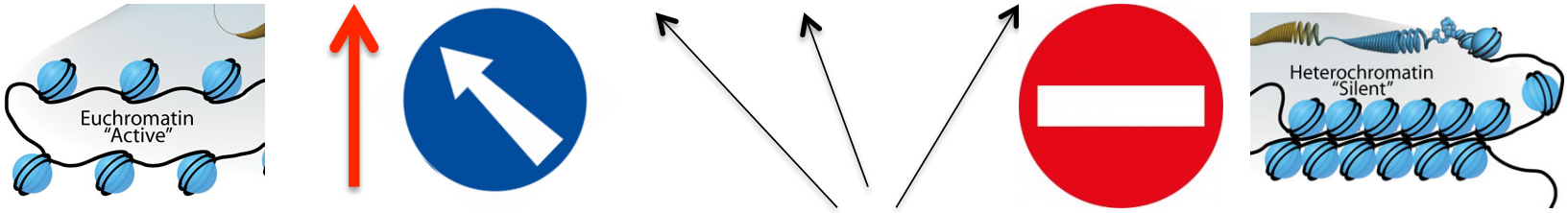
How do transcription factors “know” where to go?

Smad3



Sequence found in genome more than once every 1000bp



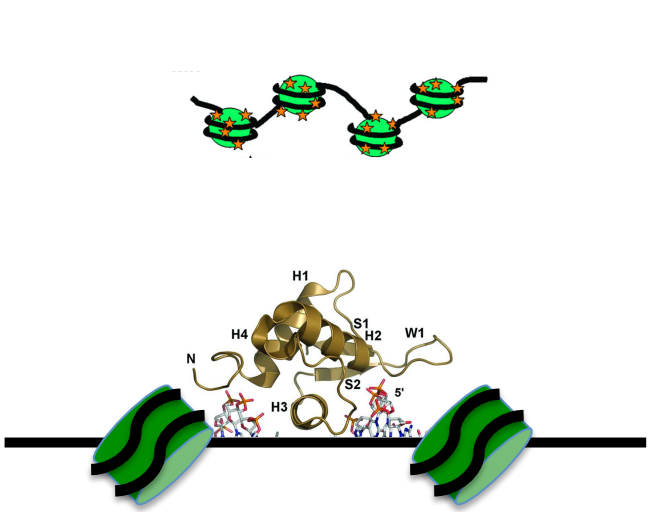


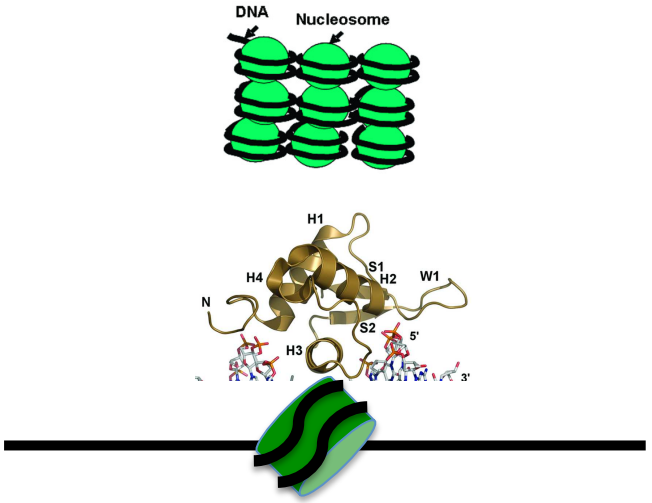
Question: Why binding to this site (and not to others) ?

Pioneering factors

- Most transcription factors bind to „open“ chromatin

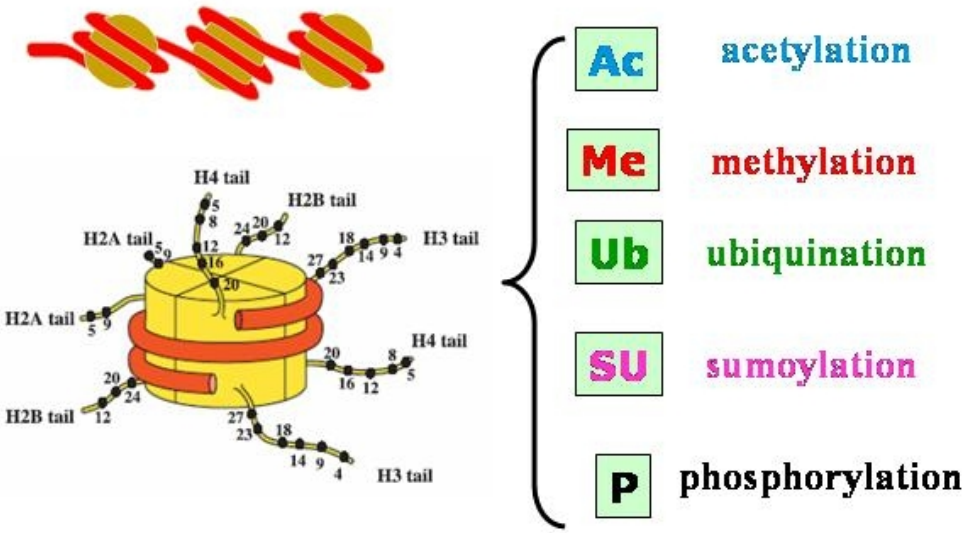
- So called pioneering factors can bind to closed chromatin and open cis regulatory elements for „business“ example: FoxA1





How do so-called pioneering factors gain excess to closed regions????

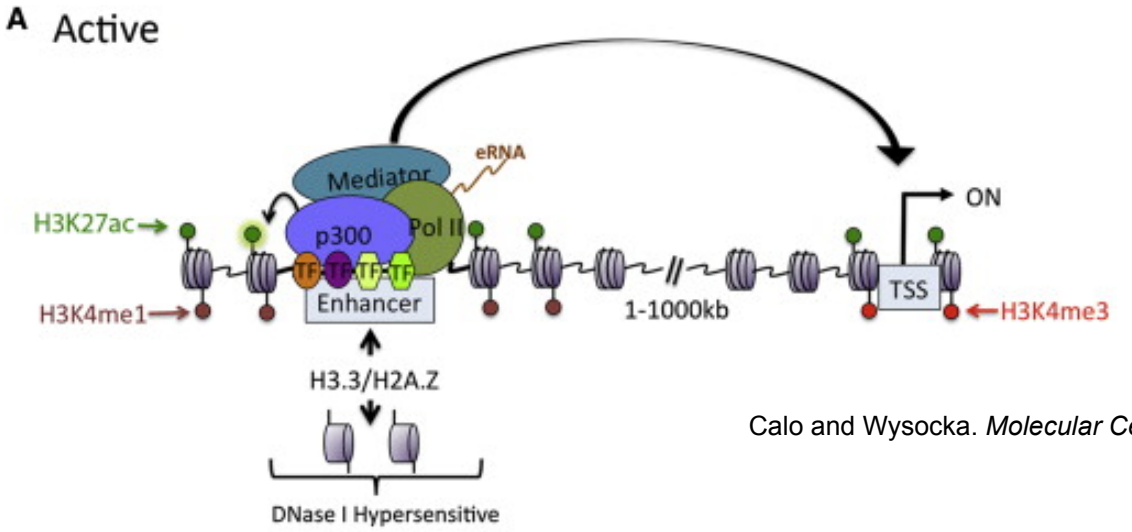
Histone modifications mark different classes of cis-regulatory elements



Function(s) of these histone modifications?

Histone modifications mark different classes of cis-regulatory elements

Different cis regulatory elements are marked with specific histone modifications

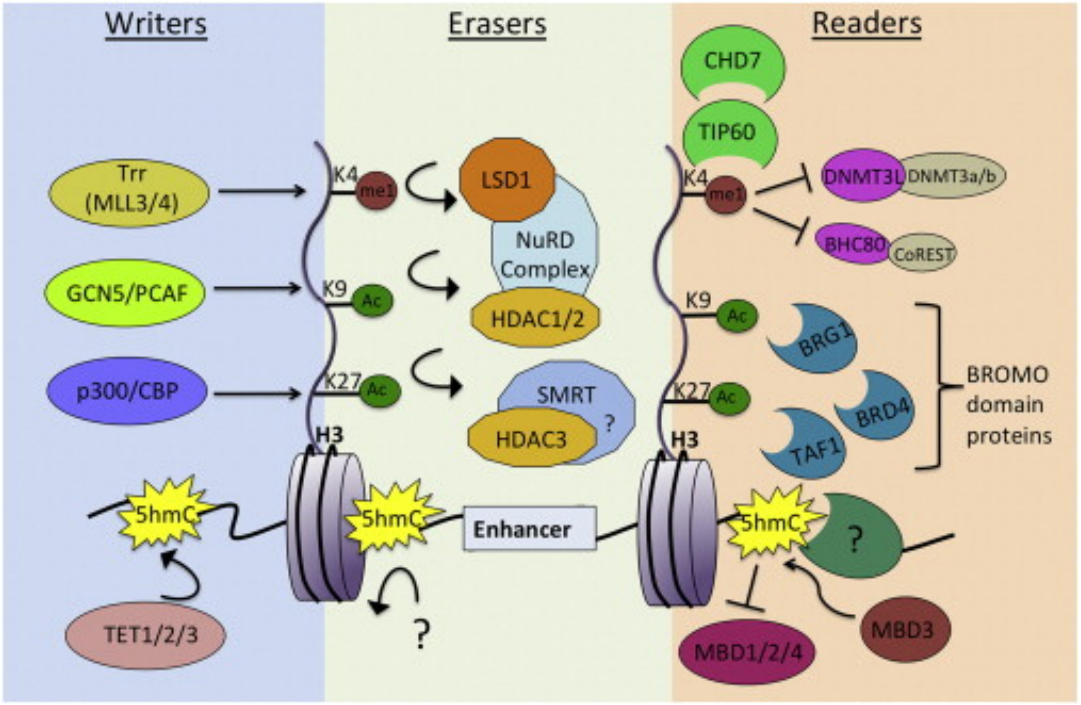


Calo and Wysocka. *Molecular Cell* (2013)

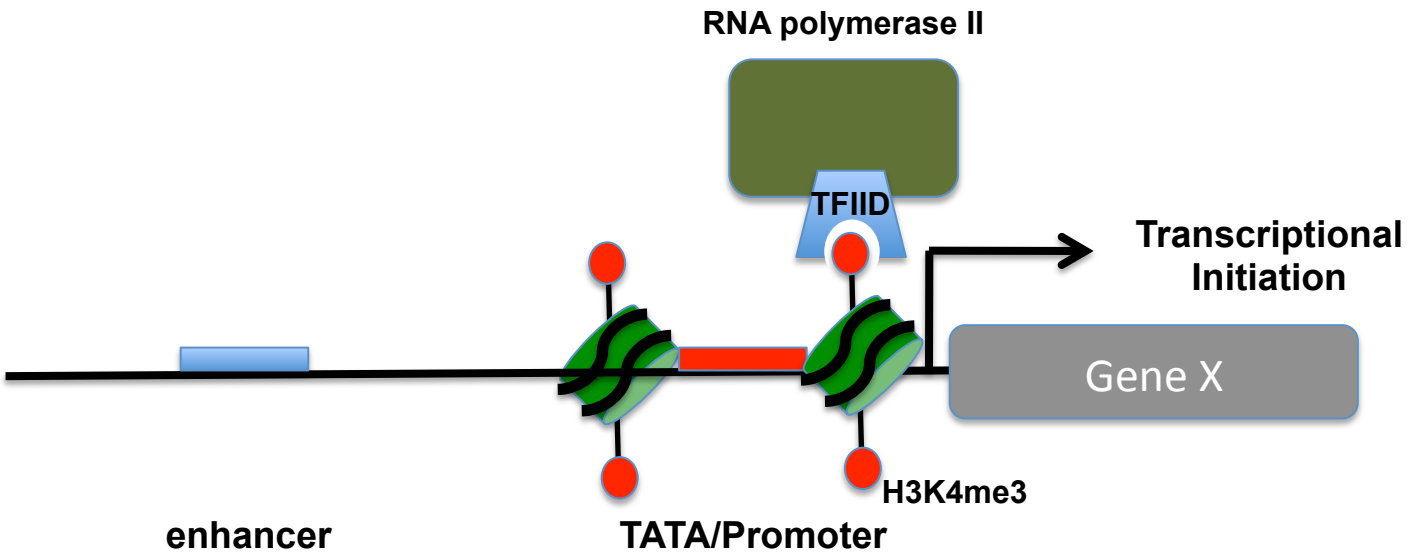
Function(s) of these histone modifications?

Histone modifications provide information ("Epigenetic")

Network of enzymes deposit / erase or recognize histone modifications



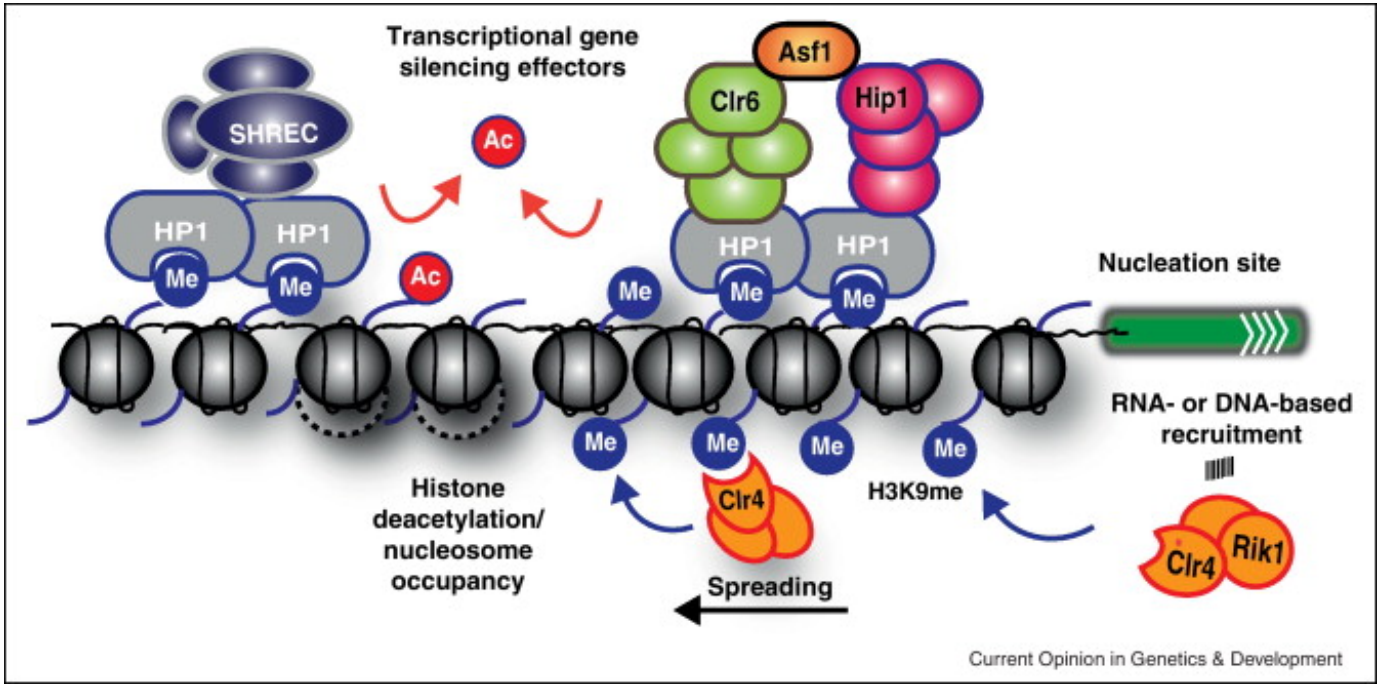
Histone modifications & activation



Vermeulen et al., *Cell* (2007)

Example Reader: TFIID binds to H3K4me3 mark found at promoter

Histone modifications & repression



Example Reader: HP1 (heterochromatin protein 1) binds H3K9me to assemble heterochromatic regions that are not transcribed

Take-home messages:

- **Transcriptional regulation allows cells with essentially the same genome to have very different functions**

  (tissue-specific expression / combinatorial regulation)

- **RNA polymerase critically depends on the help of transcription factors to initiate, elongate & terminate transcription**

- **Chromatin plays an important role in specifying where in the genome TFs & RNA polymerase can bind**

(Big) unknowns out there

- **Linking binding to transcriptional regulation in genomic context**
- **Integration of 3000 TFs & transcriptional output…?**
- **Inheritance of transcriptional programs (e.g. though cell cycle)**
- **Large part of genome encodes ncRNAs what is their function**
- **TF are not acting as on off switches but modulate expression levels quite precisely how is this accomplished**
- **Role of histone modifications in gene regulation (moving beyond correlations….)**
- **Dynamics**
- **How can TF that recognize the same sequence bind to different genomic regions & regulate different sets of genes**
- **……**

TF specificity & recognition sequences….

Low Affinity Binding Site Clusters Confer Hox Specificity and Regulatory Robustness

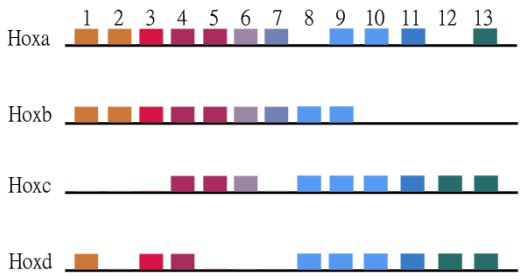
Justin Crocker,[1] Namiko Abe,[2] Lucrezia Rinaldi,[2] Alistair P. McGregor,[3] Nicolás Frankel,[4] Shu Wang,[5] Ahmad Alsawadi,[6,7] Philippe Valenti,[6,7] Serge Plaza,[6,7] François Payre,[6,7] Richard S. Mann,[2,*] and David L. Stern[1,*]
[1]Janelia Research Campus, Howard Hughes Medical Institute, 19700 Helix Drive, Ashburn, VA 20147, USA
[2]Columbia University Medical Center, 701 West 168th Street, HHSC 1104, New York, NY 10032, USA
[3]Department of Biological and Medical Sciences, Oxford Brookes University, Gipsy Lane, Oxford OX3 0BP, UK
[4]Departamento de Ecología, Genética y Evolución, IEGEBA-CONICET, Facultad, de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad, Universitaria, Pabellón 2, 1428 Buenos Aires, Argentina
[5]New Jersey Neuroscience Institute, 65 James Street, Edison, NJ 08820, USA
[6]Centre de Biologie du Développement, Université de Toulouse, UPS, 31062 Cedex 9, France
[7]CNRS, UMR5547, Centre de Biologie du Développement, Toulouse, 31062 Cedex 9, France
*Correspondence: rsm10@columbia.edu (R.S.M.), sternd@janelia.hhmi.org (D.L.S.)
http://dx.doi.org/10.1016/j.cell.2014.11.041

Model system:

Hox-cluster:

TFs that are expressed at specific positions along the anterior-posterior axis depending on their location within the cluster

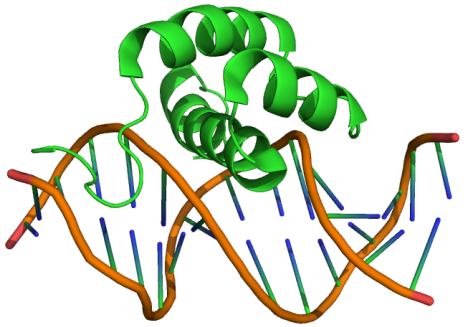
TFs in turn are responsible for the regional identity of distinct anatomical domain by regulating Hox-subtype-specific target gene expression



The Hox paradox: Same motif, different function….

UBX

Scr



Ubx motif from Mann et al. (2009)

Same concensus recognition sequence.....

Yet different target genes & functions.....

TF specificity & recognition sequences….

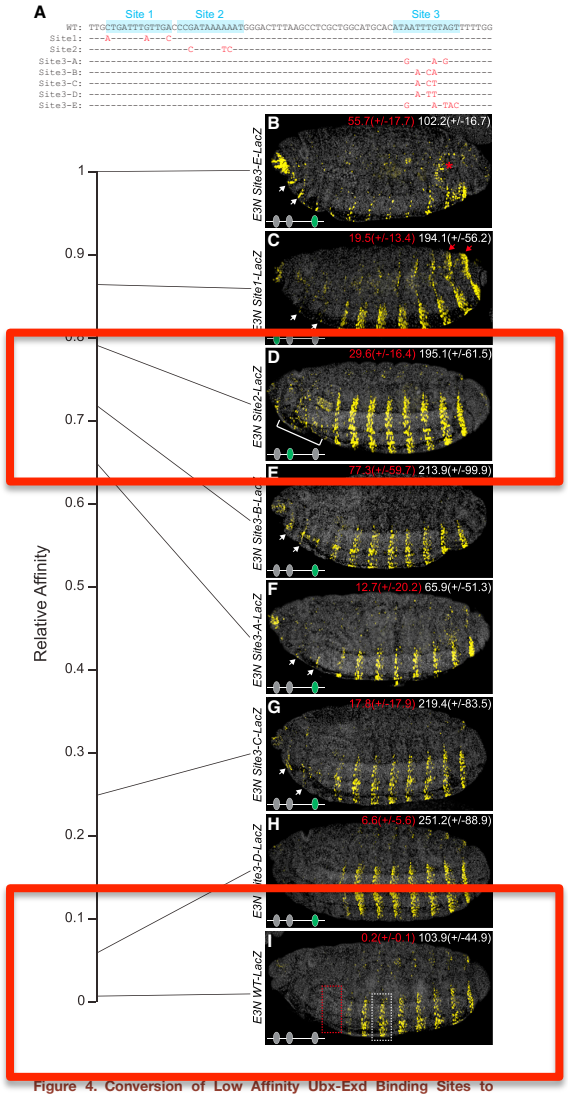


Figure 4. Conversion of Low Affinity Ubx-Exd Binding Sites to Higher Affinity Sites Results in Ectopic Expression

(A) Aligned *E3N* sequences from wild-type and mutated sequences. Dashes and red letters indicate unaltered and modified sequence, respectively.

(B–I) Embryos carrying *E3N::lacZ* constructs, with Ubx-Exd sites altered as indicated in (A). The numbers in the top right of each panel indicate the average levels of expression in the regions outlined in (I) (n = 10 for each genotype), measured in arbitrary units of fluorescence intensity. Numbers in parantheses

- **High affinity (based on SELEX-seq): increased expression & now also mis-expressed in anterior section (specificity lost….)**

- **Low affinity (based on SELEX-seq): selective expression where Ubx & AbdB are present**

TF specificity & recognition sequences….

Role for low affinity binding sites to drive Hox specificity???

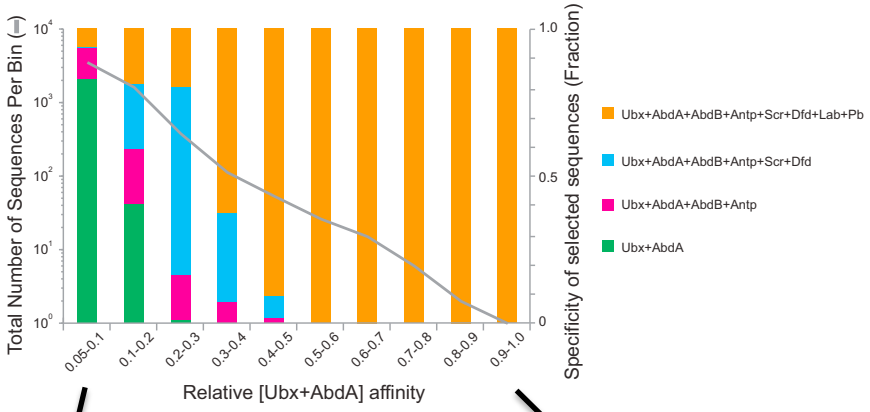


Figure 3. Inverse Correlation between Sequence Affinity and Specificity
The proportion of 12mer sequences bound by various Hox-Exd complexes versus relative affinity of these 12mers for Ubx/AbdA-Exd is shown as colored bars (specificity groups). The number of 12mers in each affinity bin is plotted as a gray line. Average relative affinities of 12mers were calculated for four pairs of Hox-Exd complexes with similar binding profiles: (1) Labial and Pb, (2) Dfd and Scr, (3) AbdB and Antp, and (4) Ubx and AbdA. Sequences specific for Ubx/AbdA-Exd (green bars) are more prevalent in lower affinity bins than in higher affinity bins.

- **High affinity (based on SELEX-seq): all HOX TFs tested bind**

- **Low affinity (based on SELEX-seq): Ubx & AbdA selectively bind**

- **Low affinity binding sites selectively bound by Ubx (& AbdA)!**
- **High affinity binding sites: bound by all Hox proteins...**