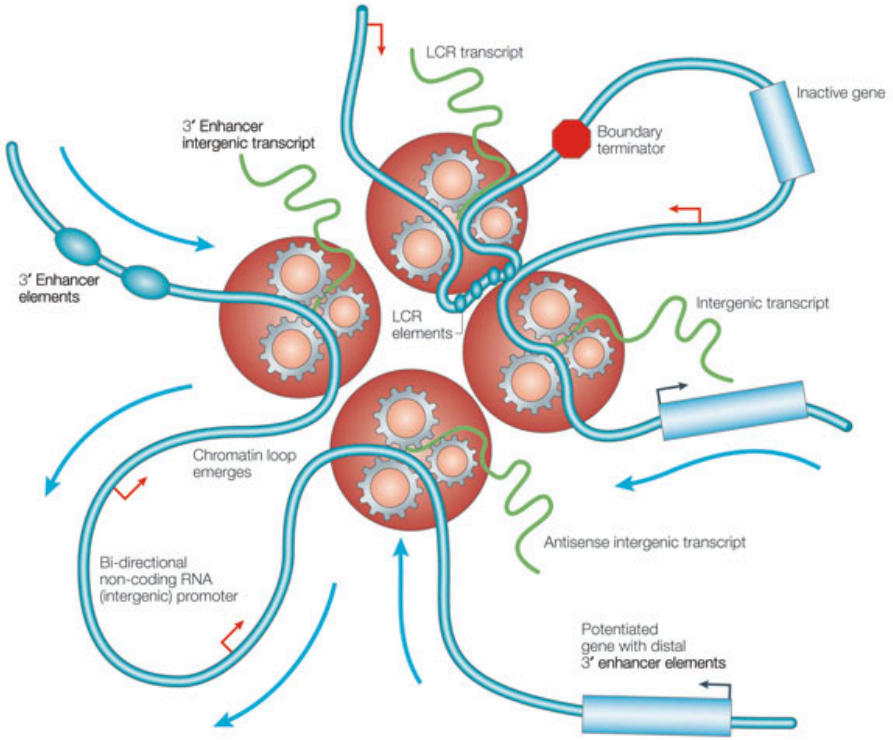
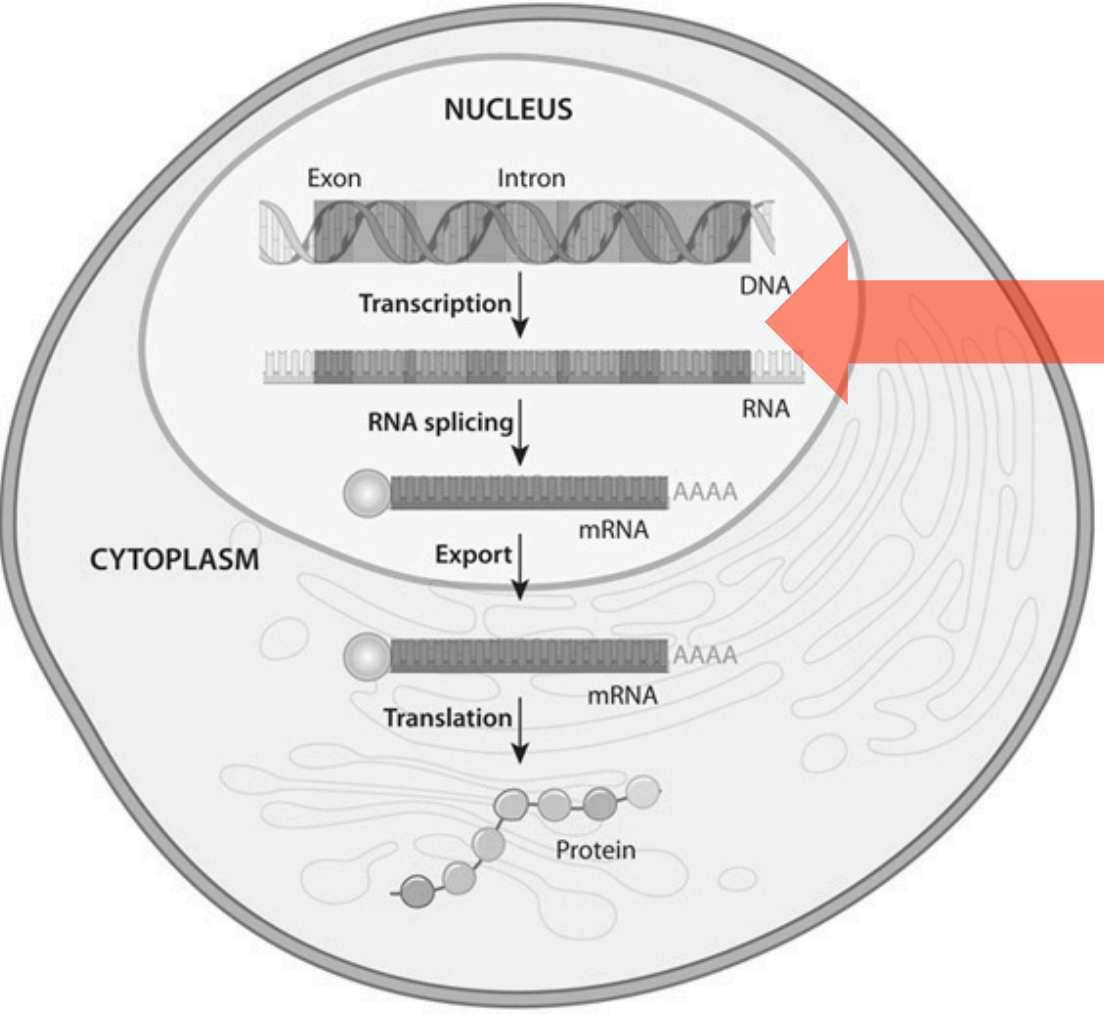
Molecular Biology
LM-FISICA aa 2017-18

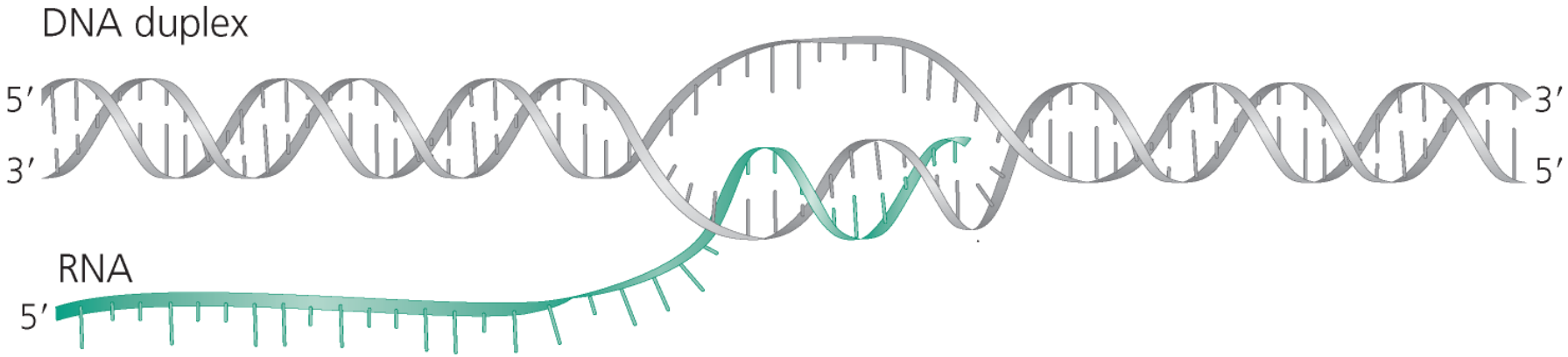


Chakalova et al. *Nature Reviews Genetics* 6, 669-677

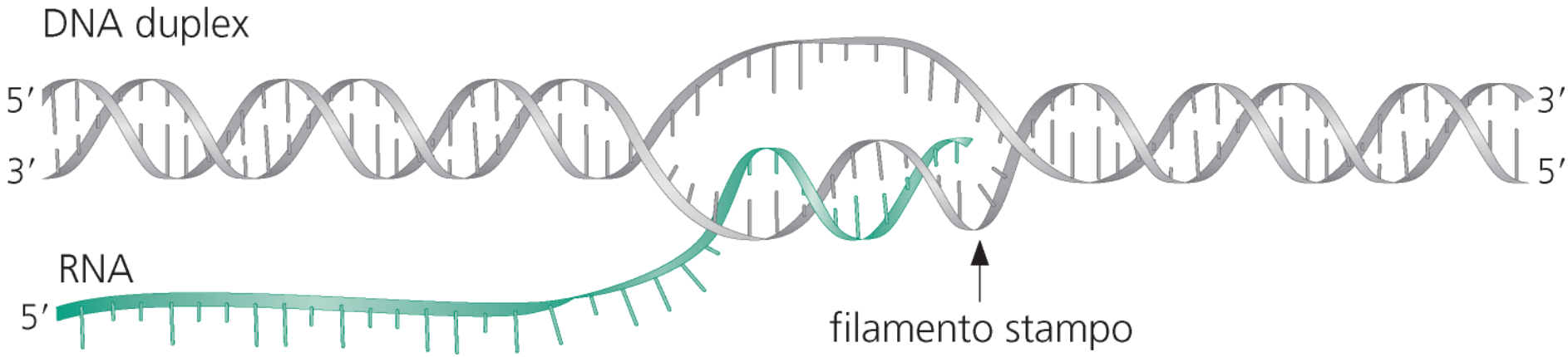
Transcriptional regulation and Chromatin structure



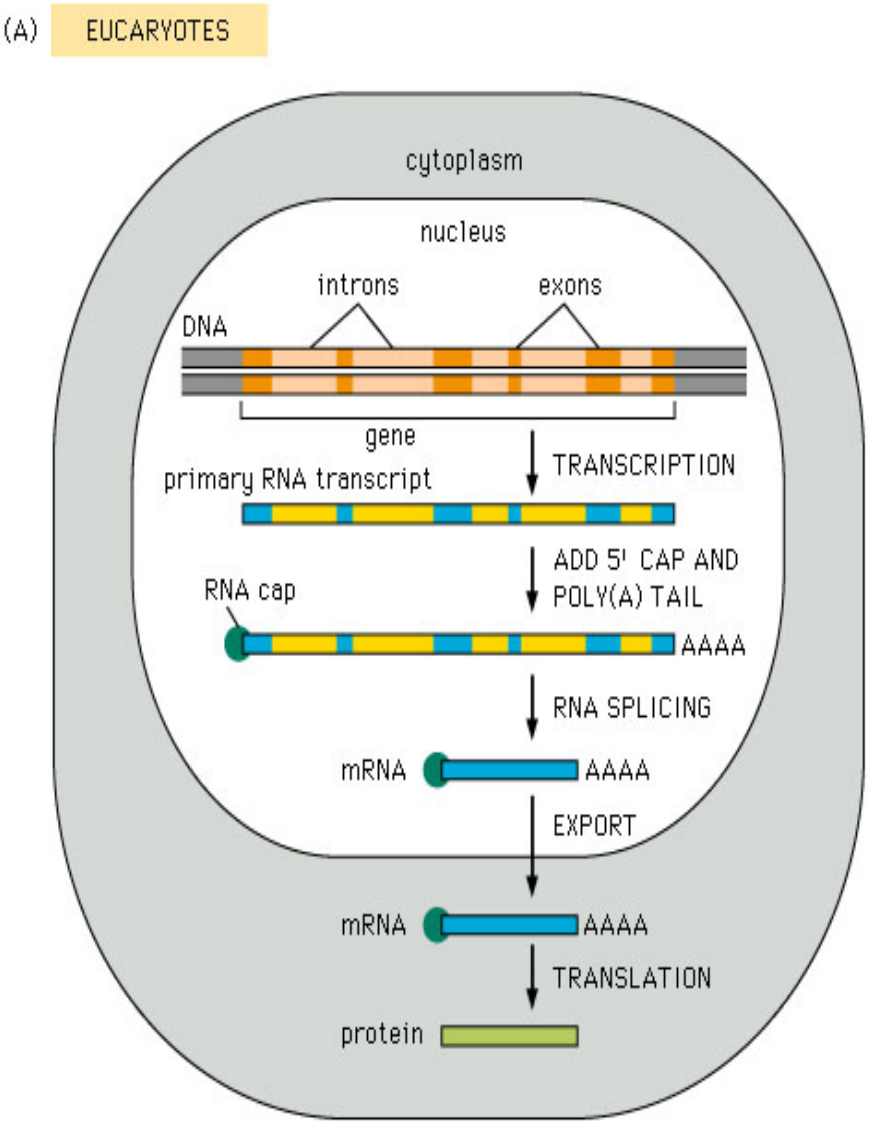
Transcription is the *first* step of gene expression and it is the process by which the information contained into DNA is *converted* in RNA.

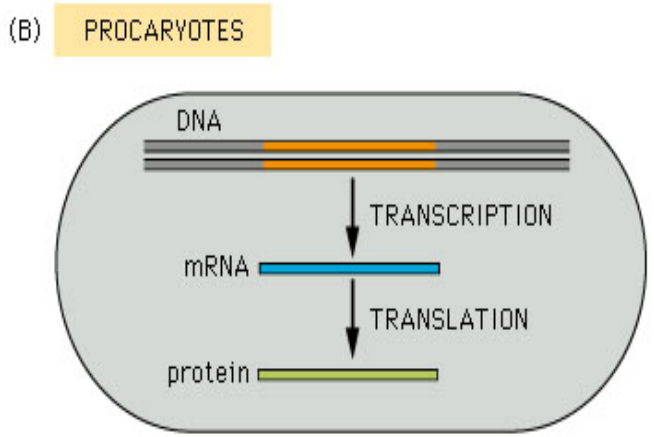


Transcription consists in the *synthesis* of an RNA chain from a DNA template.



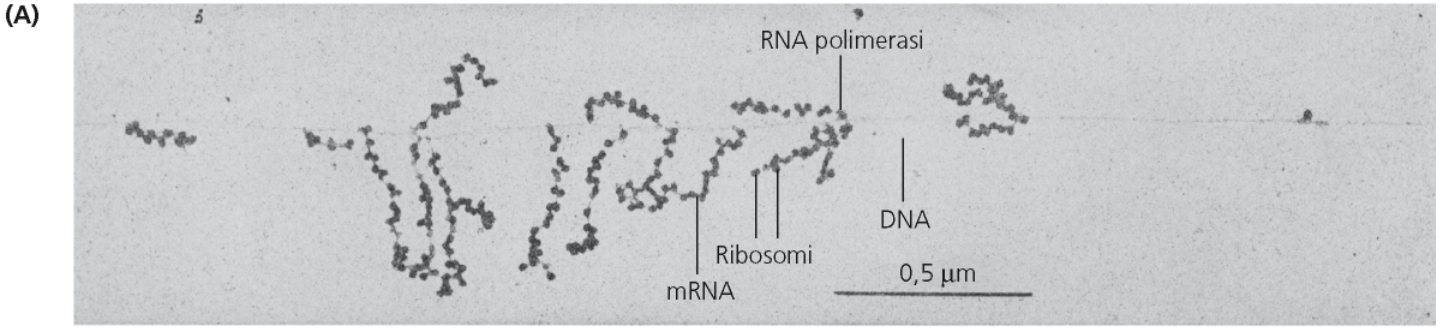
Prokaryotic and Eucaryotic Transcription

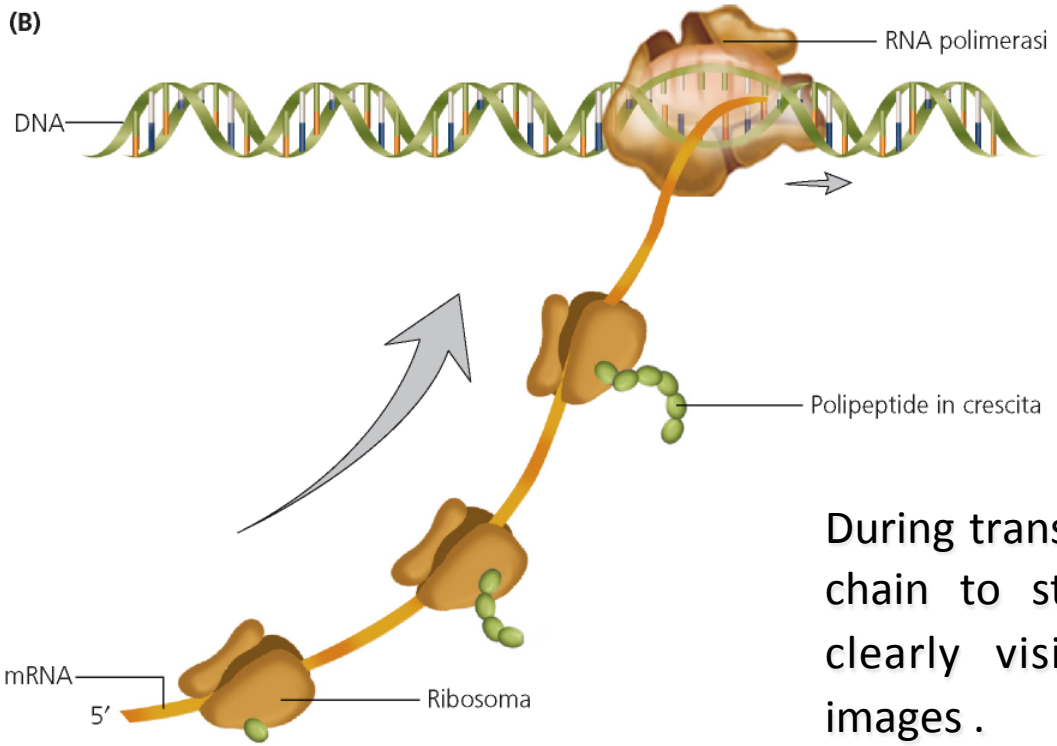




- In Eukaryotes **transcription** and **translation** are temporally and spatially defined events.
- In Prokaryotes **transcription** and **translation** occurr inside the *same* cellular compartment and are coupled.

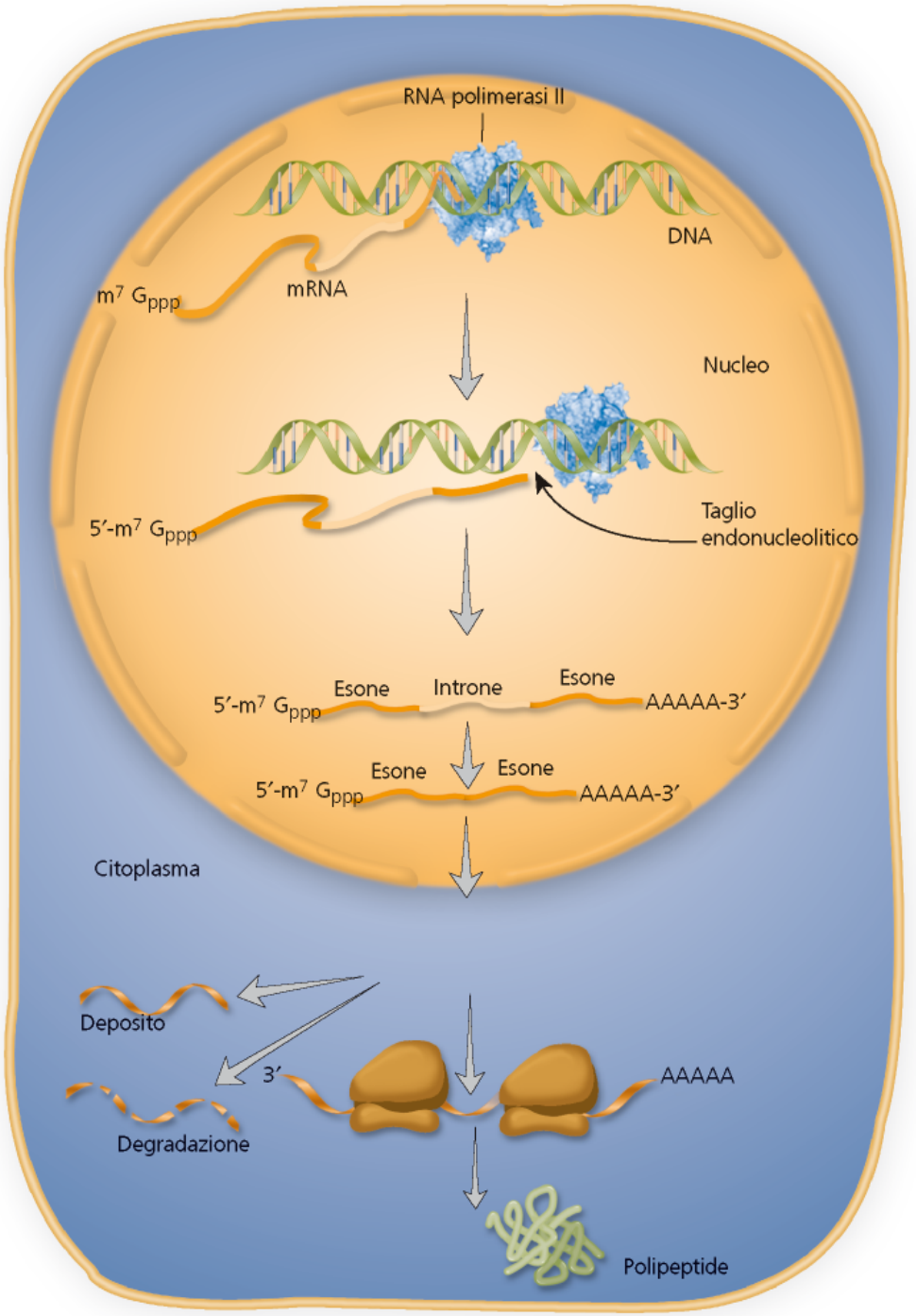
In Prokaryotes …





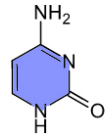
During transcription ribosomes reach the RNA chain to start protein translation, as it is clearly visible from electron microscopy images .

In Eukaryotes…

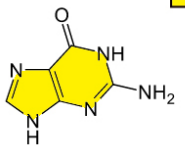


Substrates

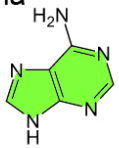
Citosina **C**



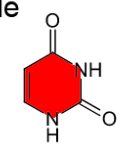
Guanina **G**



Adenina **A**



Uracile **U**



Nitrogenous bases

Enzyme

RNA Polymerase

Gene

Template

Products

tRNA

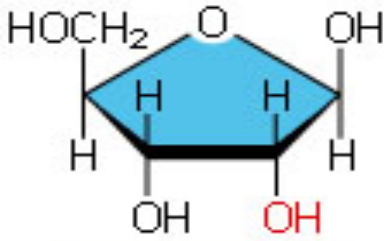
rRNA

mRNA

The RNA chain is chemically different from DNA!!!

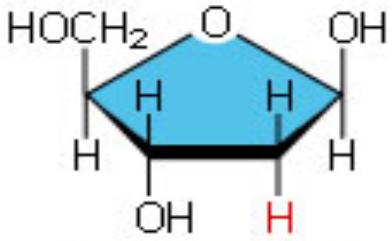
Why?

- *It is single-stranded.*
- Like DNA, RNA is composed of its **phosphate group**, five-carbon **sugar** (the less stable ribose), and four nitrogen-containing **nucleotides** which contain *Uracile* (U) instead of Timine (T). Uracil links to Adenine (A-U) and cytosine links to guanine (C-G).
- *nucleotides* contain **ribose** instead of deoxyribose. Ribose sugar is *more reactive* because of C-OH (hydroxyl) bonds. Not stable in alkaline conditions. *See next…*



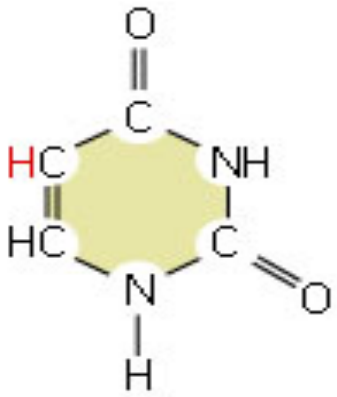
ribose

used in ribonucleic acid (RNA)



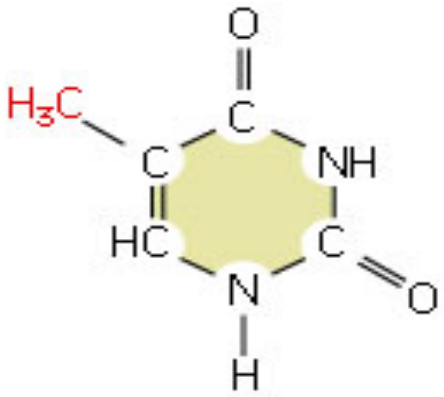
deoxyribose

used in deoxyribonucleic acid (DNA)



uracil

used in RNA

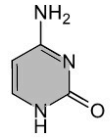


thymine

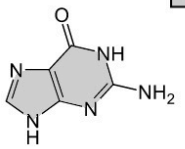
used in DNA

Substrates

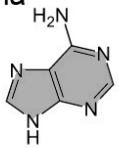
Citosina **C**



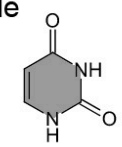
Guanina **G**



Adenina **A**



Uracile **U**



Nitrogenous bases

Enzyme

RNA Polymerase

(The Transcription UNIT)

Gene

Template

Products

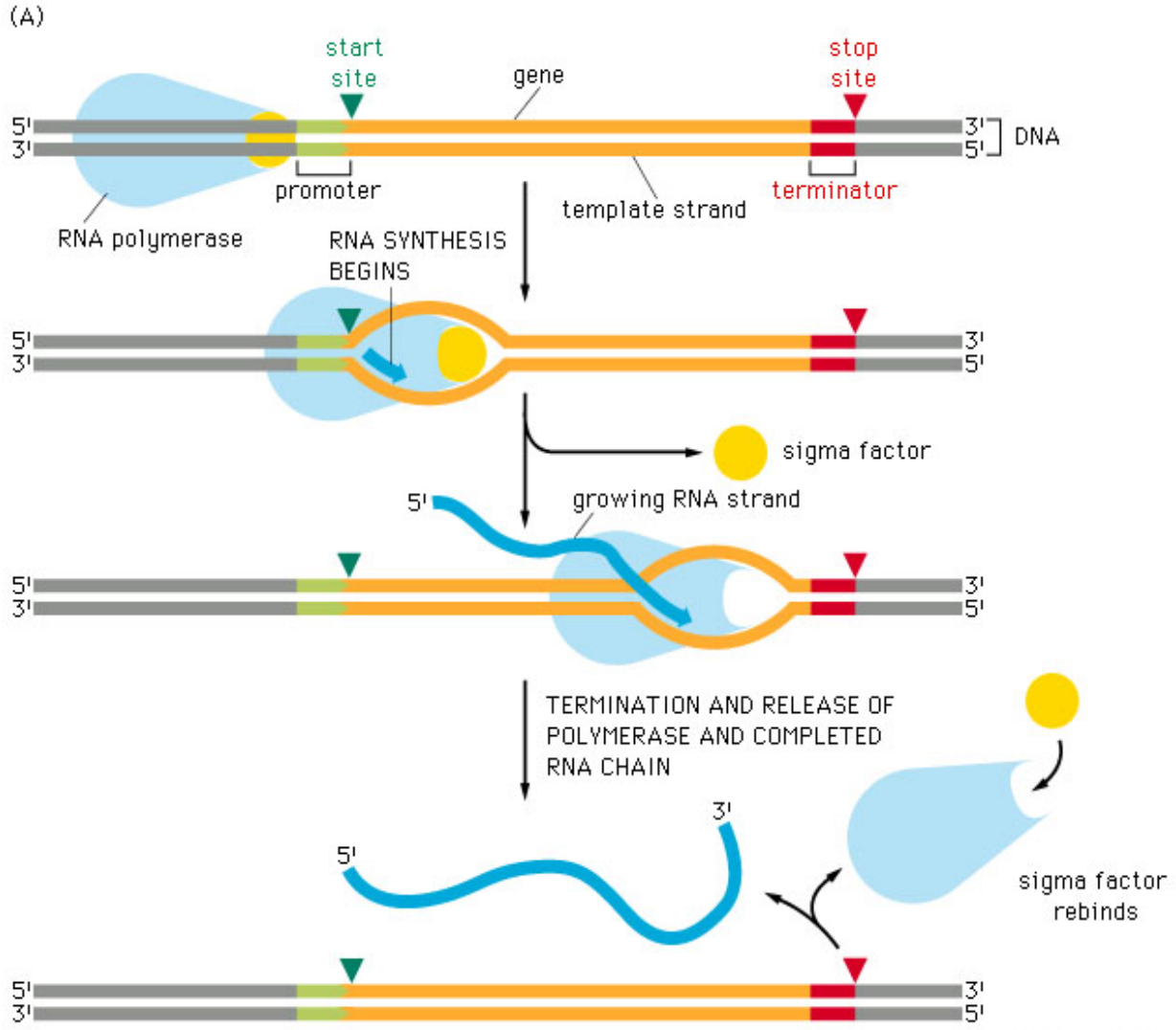
tRNA

rRNA

mRNA

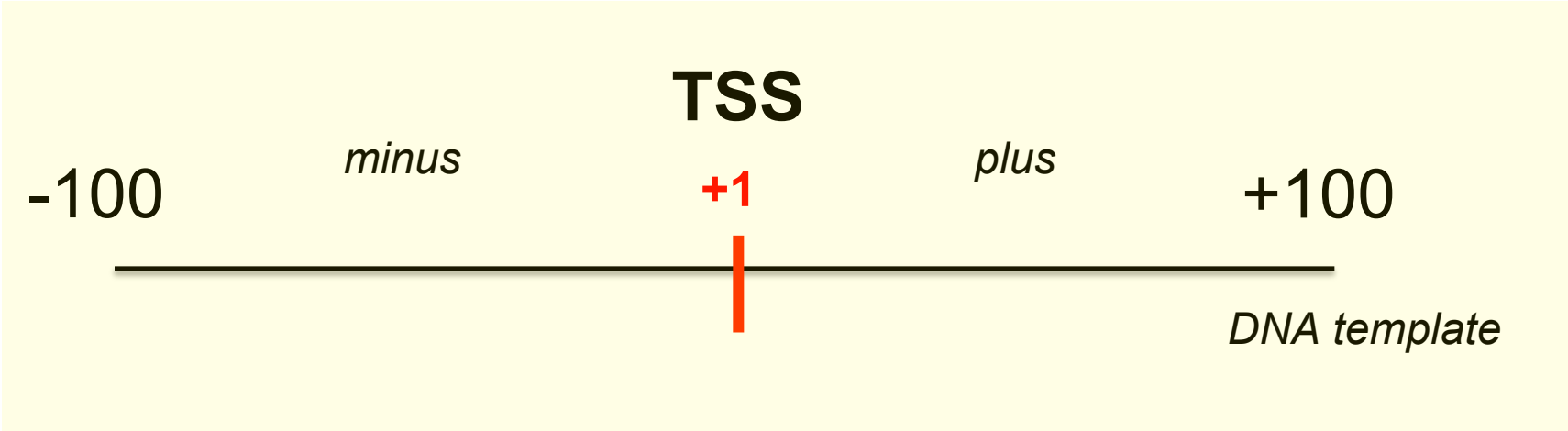
The Transcription UNIT

Sequence of nucleotides in DNA that codes for a single RNA molecule, along with the sequences necessary for its transcription; normally contains a promoter, an RNA-coding sequence, and a terminator.



Transcription starts by the recognition by the enzymatic machinery of specific regions of DNA located at the 5'-end of a gene (UPSTREAM). These sequences generally identify the ***PROMOTER*** of a gene.

Molecular Biologist use a numbering system which has no zero! The first nucleotide of the RNA transcript is numbered **+1** and correspond to the Transcription Start Site or **TSS**; the nucleotide immediately upstream from that is numbered -1.

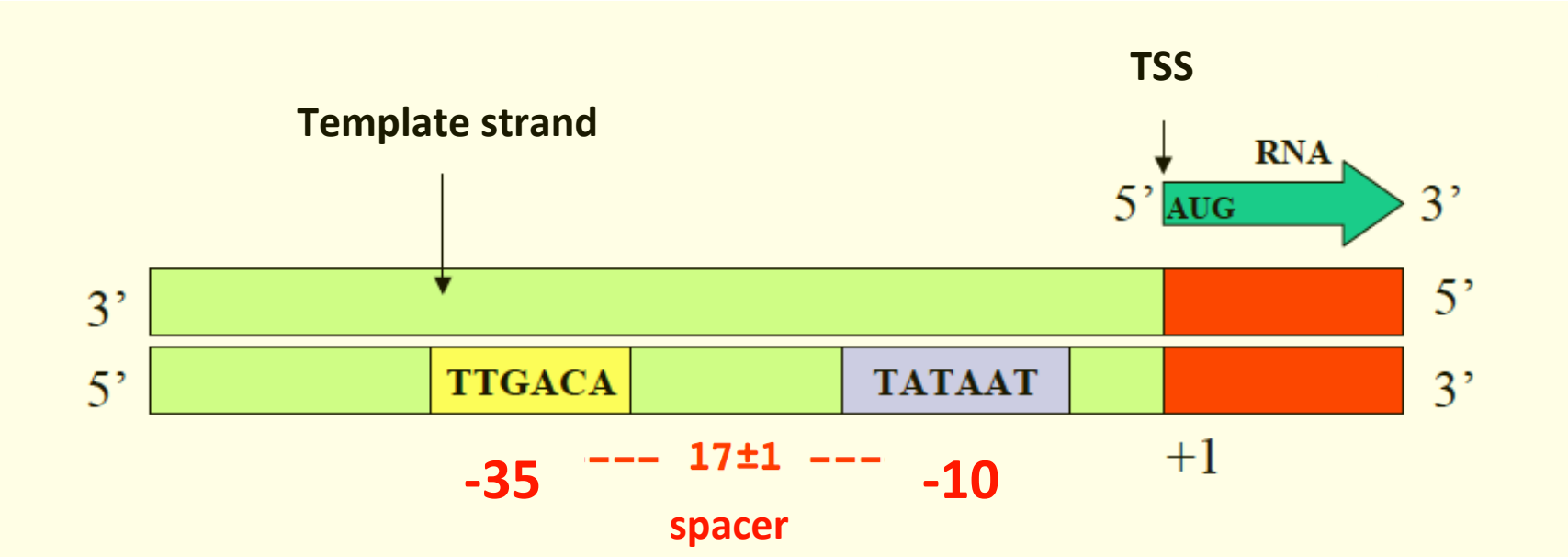


cis elements: DEFINITION

PROMOTERS contain *cis elements* which are important to guide the RNA polymerases to recognize the TSS and to start transcription from the right place.

For instance: the comparison of many E. coli promoters has revealed three main *conserved* boxes (or *consensus* sequences): **-35**, **-10**, and the **spacer**.

Structure of a canonical prokaryotic Promoter

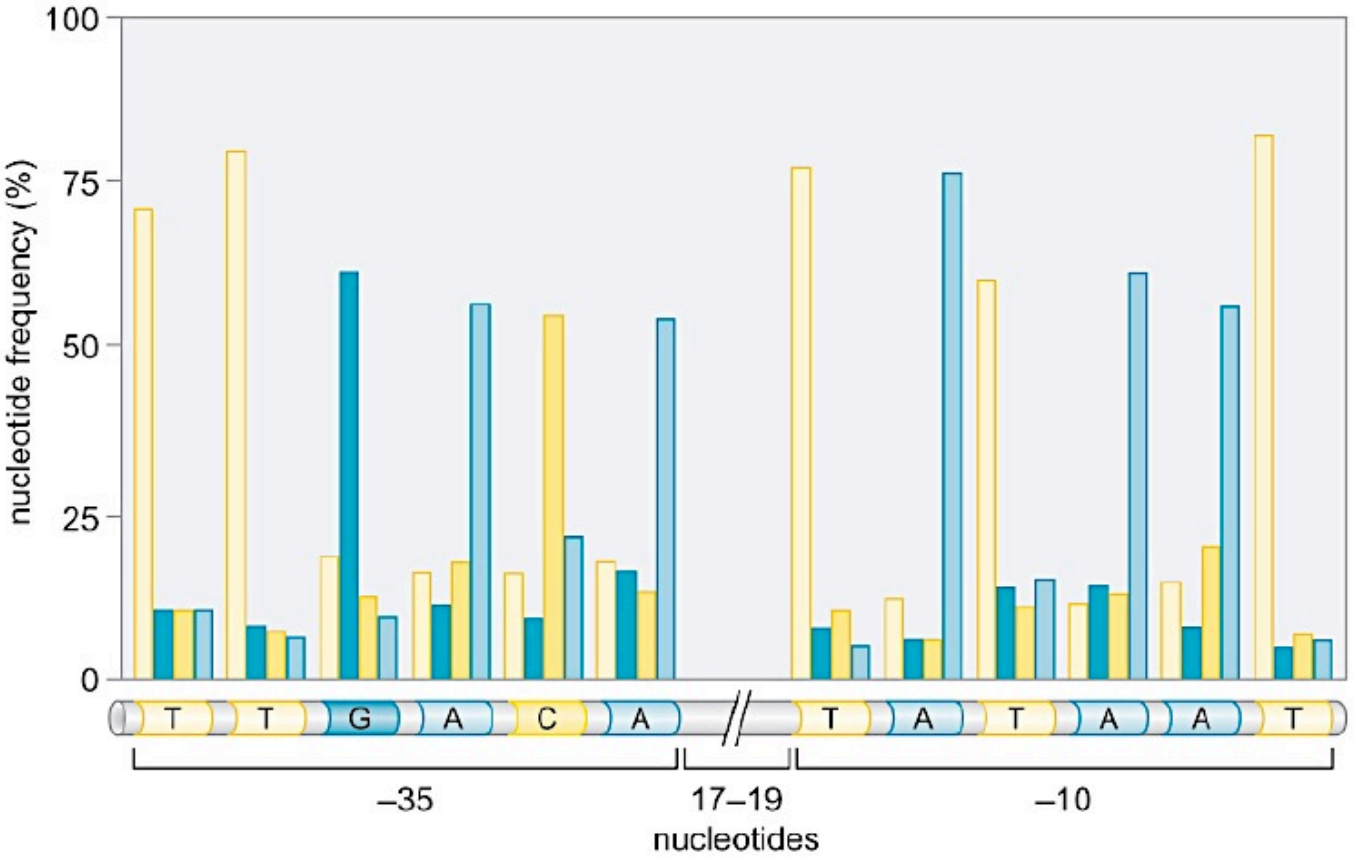


Promoter = *Consensus*

A ***consensus* sequence** is a genetic sequence found in widely divergent organisms or genetic locations with *minor* variations and (probably) *similar* functions.

It represents the residues which are more represented when a lot of sequences are aligned. Take care!!! The consensus sequence is NOT a real sequence but represents the most common nucleotides: it is a **statistical creature**!!!

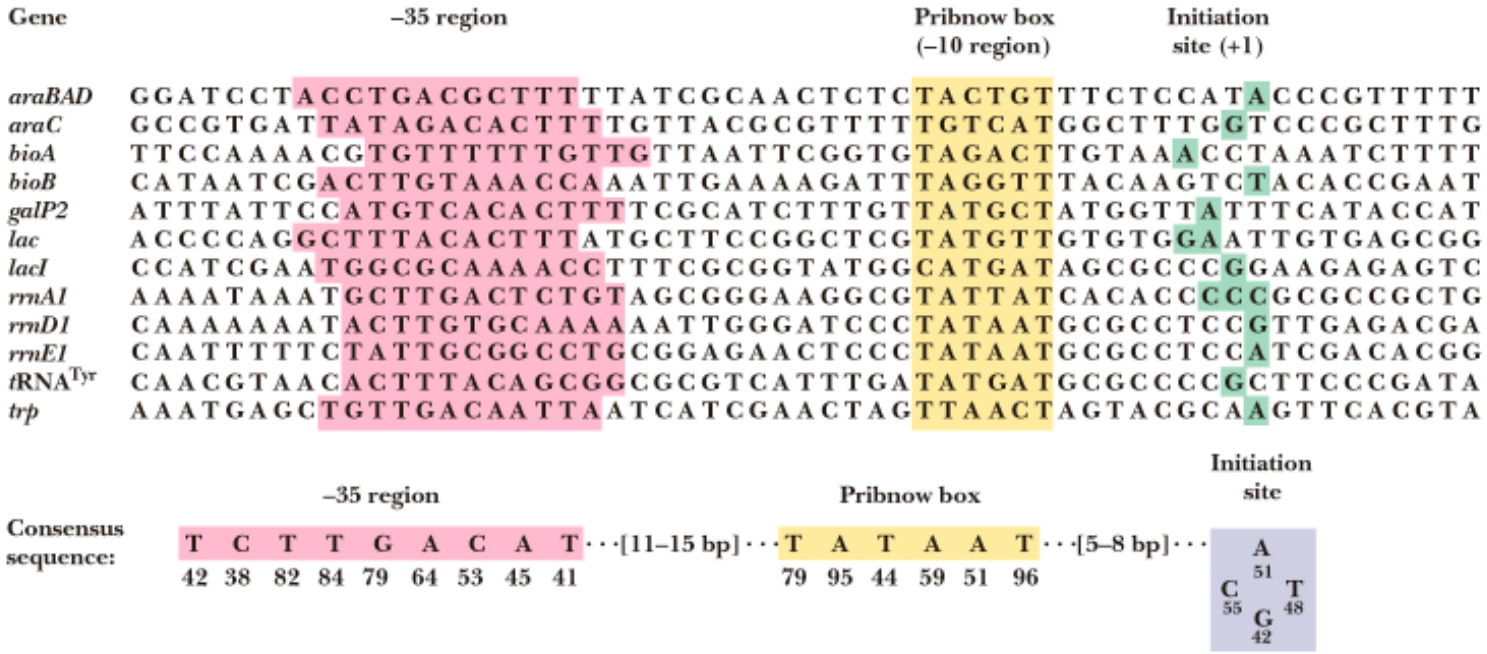
The *consensus sequence* of *E. coli* promoter was found by alignment of 300 sequences interacting with σ^{70}



Based on their *strenght* Promoters can be **STRONG** or **WEAK**

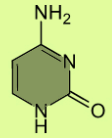
Promoter "*strength*" is defined as the number of transcripts made/unit of time.
It is generally a matter of:

- How *tightly* RNA polymerase binds Promoter (which depends on the consensus sequences)
- *Isomerization* efficiency
- How rapidly the RNA polymerase leaves the Promoter

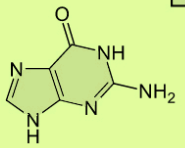


Substrates

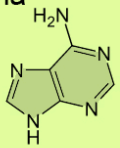
Citosina **C**



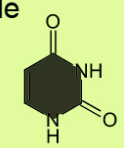
Guanina **G**



Adenina **A**



Uracile **U**



Basi azotate

Enzyme

RNA Polymerases

Products

tRNA

rRNA

mRNA

Gene

Template

The RNA polymerase enzymes

RNA polymerases are able to *recognize* and *selectively* transcribe genes by interacting, with the help of other proteins (*trans* elements), with specific sites of DNA (*cis* elements) inside the Promoters.

Once recognized the *cis* elements, the RNA polymerases begins the synthesis of RNA.

- they *synthetize* RNA in 5'→3' direction by using the 4 *riboNTP* as precursors
- they copy the template DNA
- they do not need start primers
- they have no proofreading activity

Procaryotes		*Eucaryotes*		
Bacteria	**Archei**	**RNAP I**	**RNAP II**	**RNAP III**
Core	Core	Pol I	Pol II	Pol III
β	A’ /A’ ’	RPA1	RPB1	RPC1
β’	B	RPA2	RPB2	RPC2
α^{I}	D	RPC5	RPB3	RPC5
α^{II}	L	RPC9	RPB11	RPC9
ω	K	RPB6	RPB6	RPB6
	[+ other 6]	[+ other 9]	[+ other 7]	[+ other 11]
		transcribes **rRNA**	transcribes **mRNA**	transcribes **tRNA**, small RNA and the **RNA 5S**

Transcription in Eukaryotes

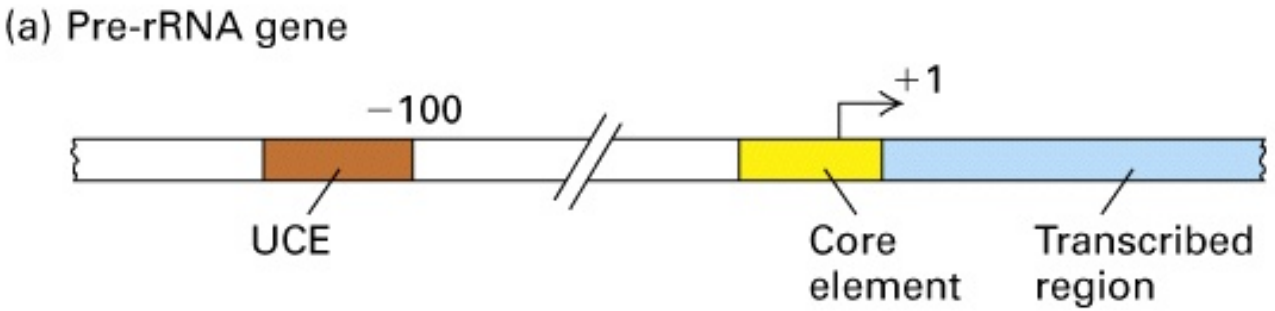
Nucleus of Eucaryotic cells contains **3** different kind of **RNA polymerases** DNA-dependent, all of them are omologs to bacterial polymerases.

Each kind of eukaryotic RNA polymerases recognise *different* promoters (sequence and position) to synthetize specifically *different* kind of RNA.

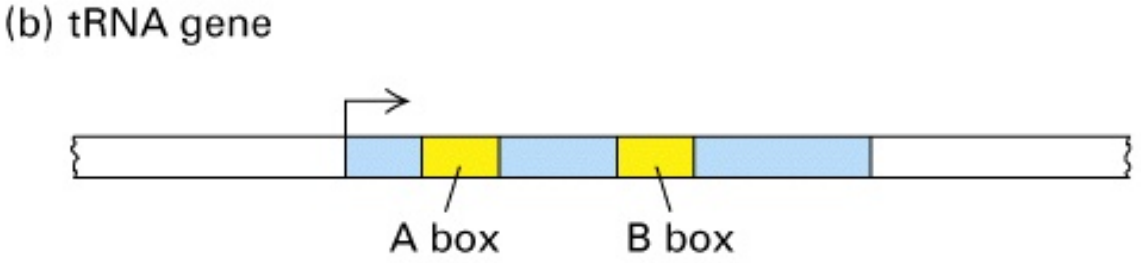
		α-amanitin
RNA polymerase I	**rRNA**	-
RNA polymerase II	**mRNA**	++
RNA polymerase III	**tRNA**	-/+

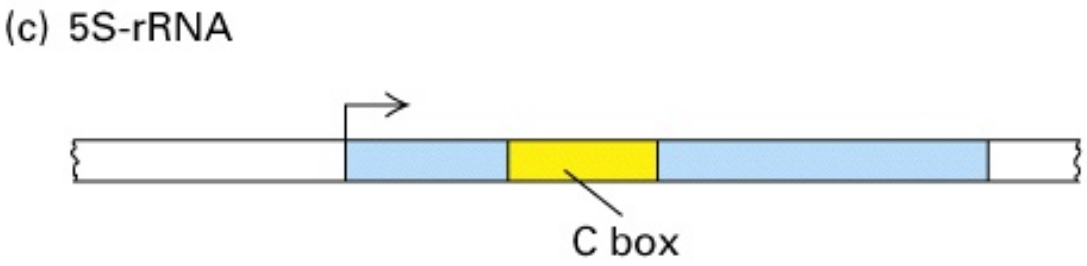
General structure of Pol I e Pol III Promoters

RNA Pol I



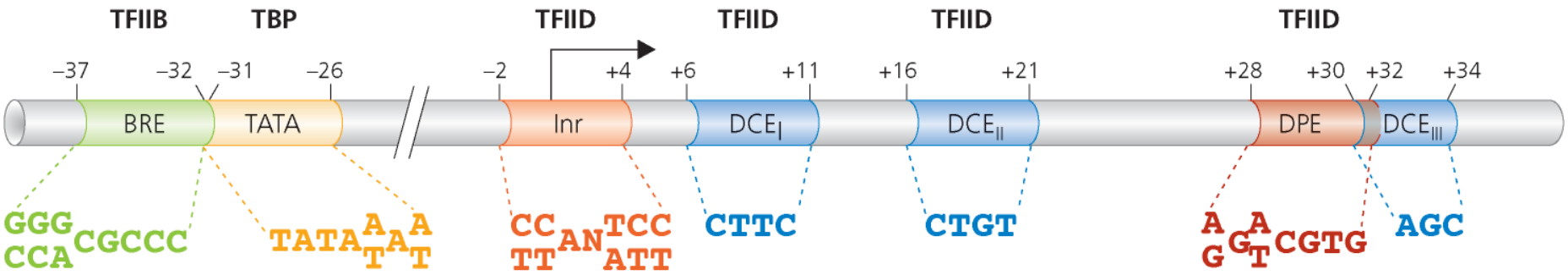
RNA Pol III





RNA Pol II promoters

Each gene contains specific DNA elements at its promoter: the **core promoter elements** (**CPEs**)



-25: **TATA** box, similar to bacterial -10. The presence or absence of a TATA box is used broadly to classify genes as TATA-containing or TATA-less promoters.

The sequences immediately flanking the TATA box can contain the elements recognized by the general transcription factor TFIIB. These elements contact general transcription factors (**GTFs**):

-35: **BRE** (TFIIB Recognition Element)

-2+4: **Inr** element (Initiator)

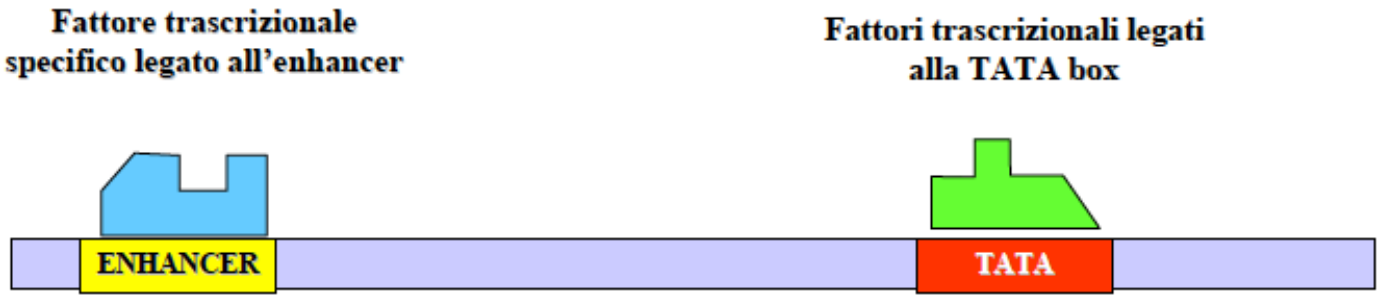
+28-+30: **DPE** (Downstream Promoter Element)

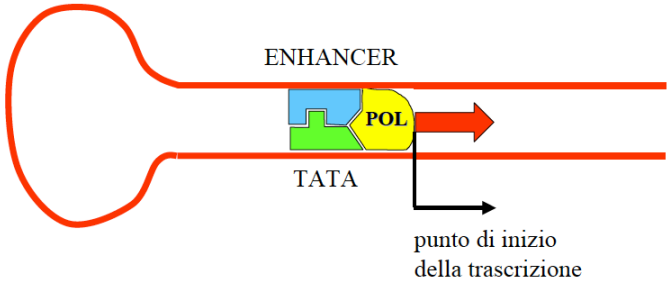
Different positions: **DCE** (Downstream Core Element)

- They are not always present *together* in the Pol II promoters
- There are also other regulatory elements differently located upstream them (enhancers, LCR, insulators)

Il modello sinora illustrato garantisce un *livello basale* di attività trascrizionale. Il processo puo' venir favorito dall'interazione di GTFs *specifici* con sequenze a monte del Promotore.

es. Gli **Enhancer**





Whether prokaryotic or eukaryotic…**transcription** consists of three main events:

1) ***Initiation*** - binding of RNA polymerase to double-stranded DNA; this step involves a transition to *single-strandedness* in the region of binding; RNA polymerase binds at a sequence of DNA called the promoter.

2) ***Elongation*** - the covalent *addition* of nucleotides to the 3' end of the growing polynucleotide chain; this involves the development of a short stretch of DNA that is transiently single-stranded

3) ***Termination*** - the recognition of the transcription termination sequence and the release of RNA polymerase

RNA Polymerases search for specific sequences (promoters)

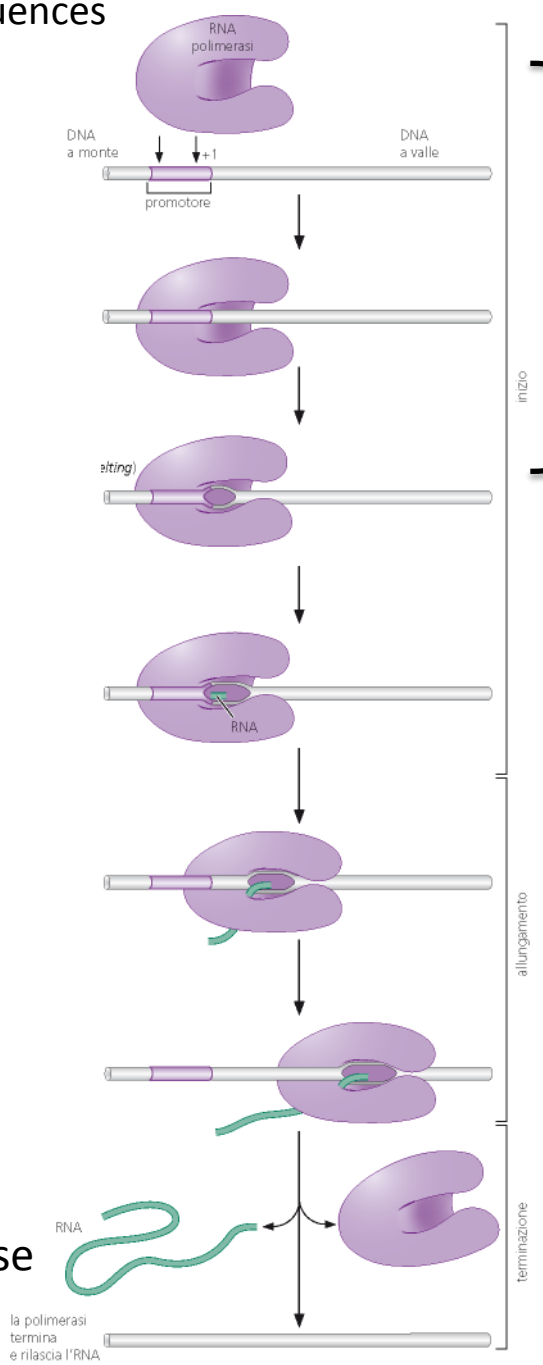
RNA Polymerase binds promoter: *closed* complex

Promoter melting: *open complex*

INITIATION

ELONGATION

TERMINATION and RNA release



PRE-INITIATION

- Only one of the two strands acts as template.
- Transcription only proceeds in 5'-3' direction.
- Transcripts of less than 5 nt are unstable, resulting in a high frequency of *abortive* initiation.

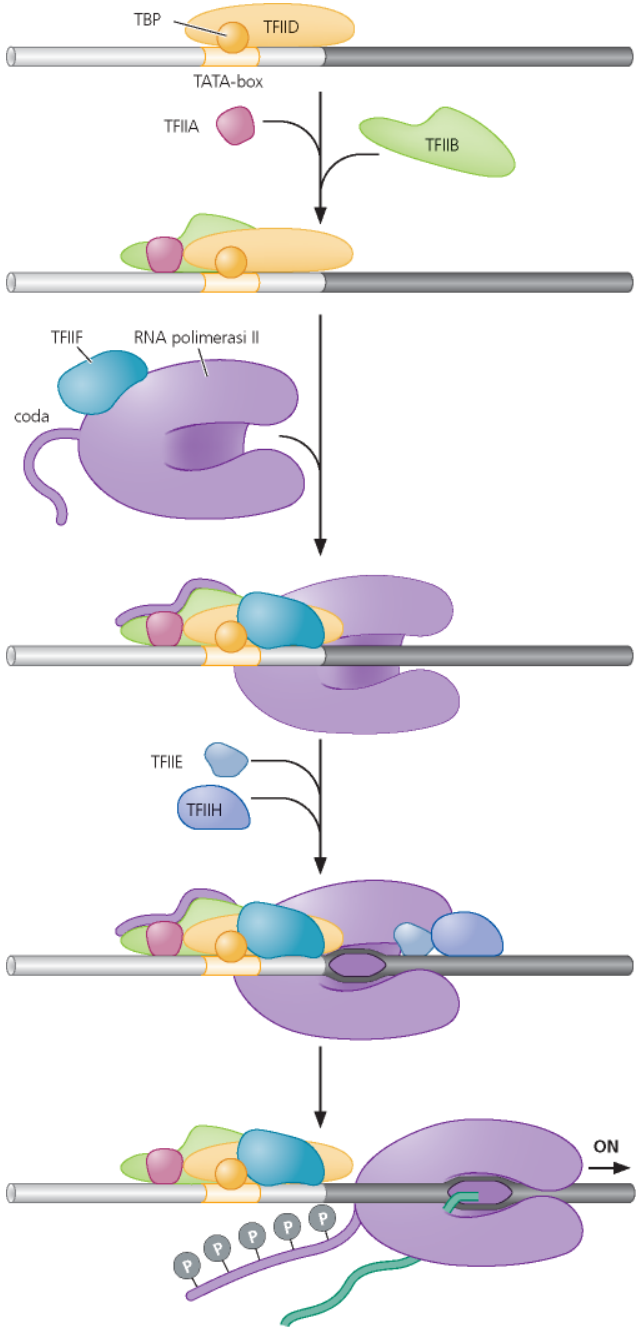
Thus…the General Transcription Factors (**GTFs**) *help* RNA polimerase to bind the promoter

At the transcription start site, Pol II initiation is regulated by a protein assembly known as the **pre-initiation complex (PIC)** containing **TFIIA**, **TFIIB**, **TFIID**, **TFIIE**, **TFIIF**, **TFIIH**, **Pol II** and **Mediator**

TFIID is the first complex which binds promoter and contains:

- **TBP** (TATA-Binding Protein)
- **TAFs** (TBP-associated factors). TAFs bind Initiator and DPE.

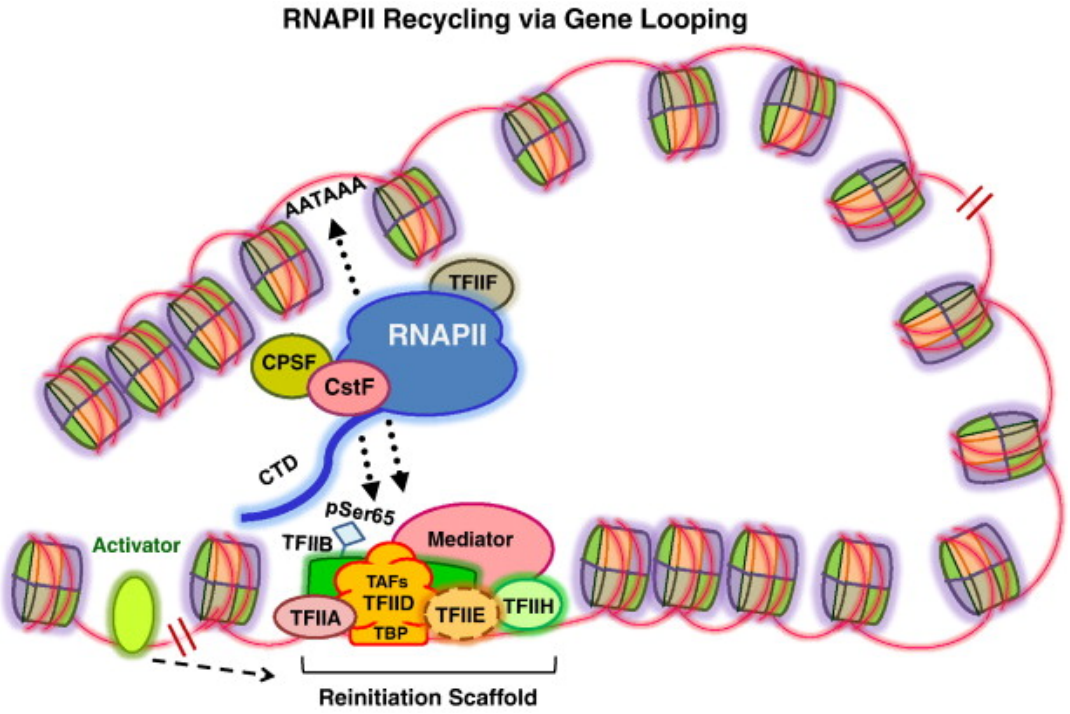
TFIIA and **B** are recruited later. **TFIIB** acts as a bridge between TBP and the Polimerase that is arriving.



General Transcription Factors

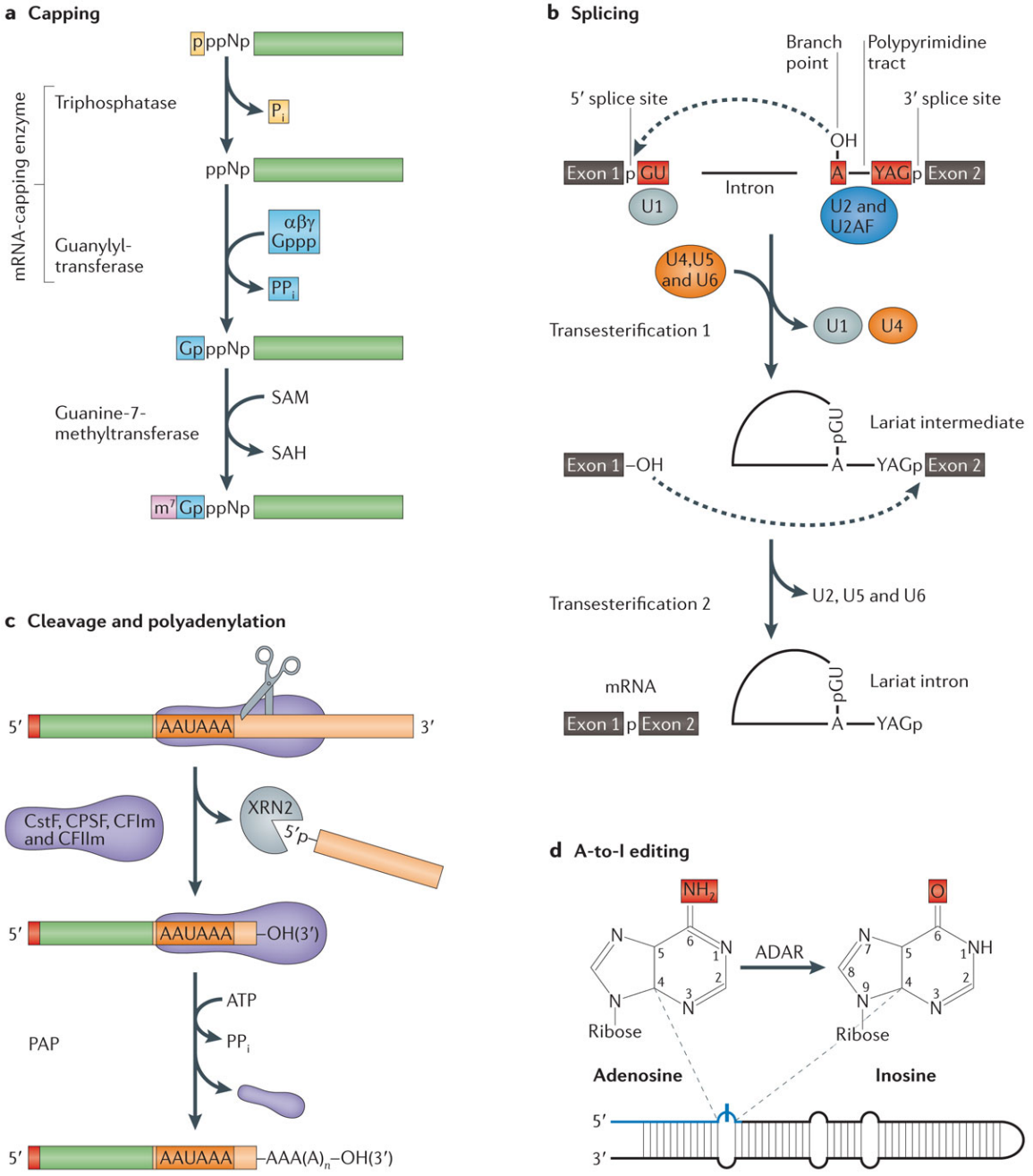
Factor	Gene name		Mass (kDa)		Uniprot accession number		Copies
	Yeast	Human	Yeast	Human	Yeast	Human	
TFIIA[§]: TBP stabilization and counteracts repressive effects of negative co-factors							
Large subunit	*TOA1*	*GTF2A1*	32.2	41.5	P32773	P52655	1
Small subunit	*TOA2*	*GTF2A2*	13.5	12.5	P32774	P52657	1
Total (2 subunits)			45.7	54.0			
TFIIB: Pol II recruitment, TBP binding and TSS selection							
TFIIB (TFB*)	*SUA7*	*GTF2B*	38.2	34.8	P29055	Q00403	1
TFIID: Pol II recruitment and promoter recognition							
TBP (TBP*): recognition of the TATA box	*TBP*	*TBP*	27.0	37.7	P13393	P20226	1
TAF1	*TAF1*	*TAF1*	120.7	212.7	P46677	P21675	1
TAF2	*TAF2*	*TAF2*	161.5	137.0	P23255	Q6P1X5	1
TAF3	*TAF3*	*TAF3*	40.3	103.6	Q12297	Q5VWG9	1
TAF4	*TAF4*	*TAF4*	42.3	110.1	P50105	O00268	2
TAF5	*TAF5*	*TAF5*	89.0	86.8	P38129	Q15542	2
TAF6	*TAF6*	*TAF6*	57.9	72.7	P53040	P49848	2
TAF7	*TAF7*	*TAF7*	67.6	40.3	Q05021	Q15545	1
TAF8	*TAF8*	*TAF8*	58.0	34.3	Q03750	Q7Z7C8	1
TAF9	*TAF9*	*TAF9*	17.3	29.0	Q05027	Q16594	2
TAF10	*TAF10*	*TAF10*	23.0	21.7	Q12030	Q12962	2
TAF11	*TAF11*	*TAF11*	40.6	23.3	Q04226	Q15544	1
TAF12	*TAF12*	*TAF12*	61.1	17.9	Q03761	Q16514	2
TAF13	*TAF13*	*TAF13*	19.1	14.3	P11747	Q15543	1
TAF14[‖]	*TAF14*	NA	27.4	NA	P35189	NA	3
Total (14–15 subunits)			1,200[¶]	1,300[¶]			
TFIIE: recruitment of TFIIH and open DNA stabilization							
TFIIEα (TFE*)	*TFA1*	*GTF2E1*	54.7	49.5	P36100	P29083	1
TFIIEβ	*TFA2*	*GTF2E2*	37.0	33.0	P36145	P29084	1
Total (2 subunits)			91.7	82.5			

The end is a new beginning: gene looping



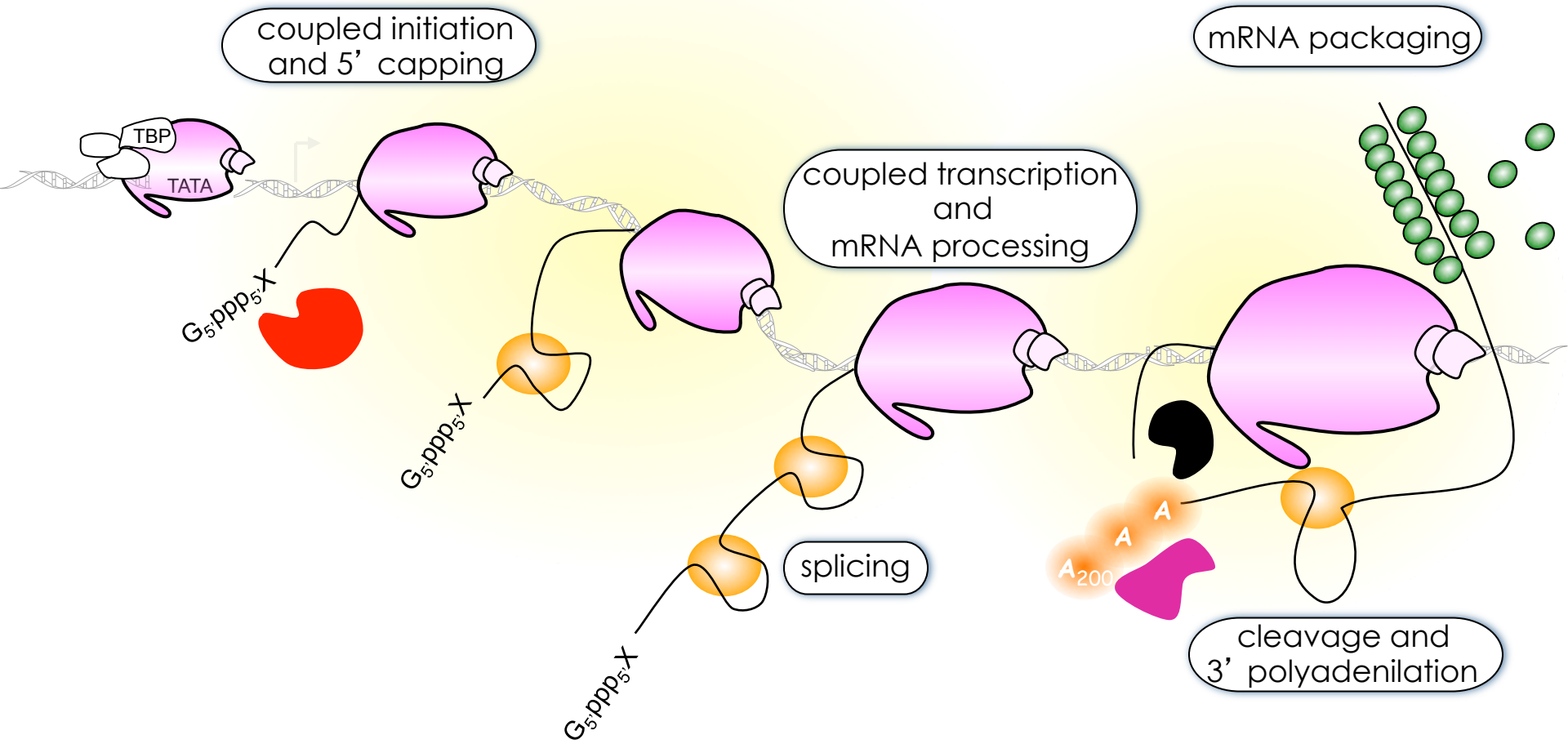
The terminator and promoter regions of a gene juxtapose during active transcription in order to facilitate RNAPII recycling and multiple rounds of transcription. The prebound GTFs and mediator in the reinitiation scaffold stabilized by the activator along with phospho-TFIIB interact with RNAPII and the termination complexes such as CPSF, CstF, and mediate such promoter–terminator contacts known as gene looping and thereby increase the efficiency of reinitiation by RNAPII.

Textbooks often describe mRNA biogenesis as a pathway in which **transcription** is followed by ***capping***, ***3' end formation*** and finally ***splicing***.



This (*sequential*) scheme is consistent with the biochemical reconstitution of these reactions *in vitro* independently of one another. However, in living cells, transcription and processing are mostly not sequential but simultaneous, that is…

..RNA processing is co-transcriptional!



The “mRNA Factory” model

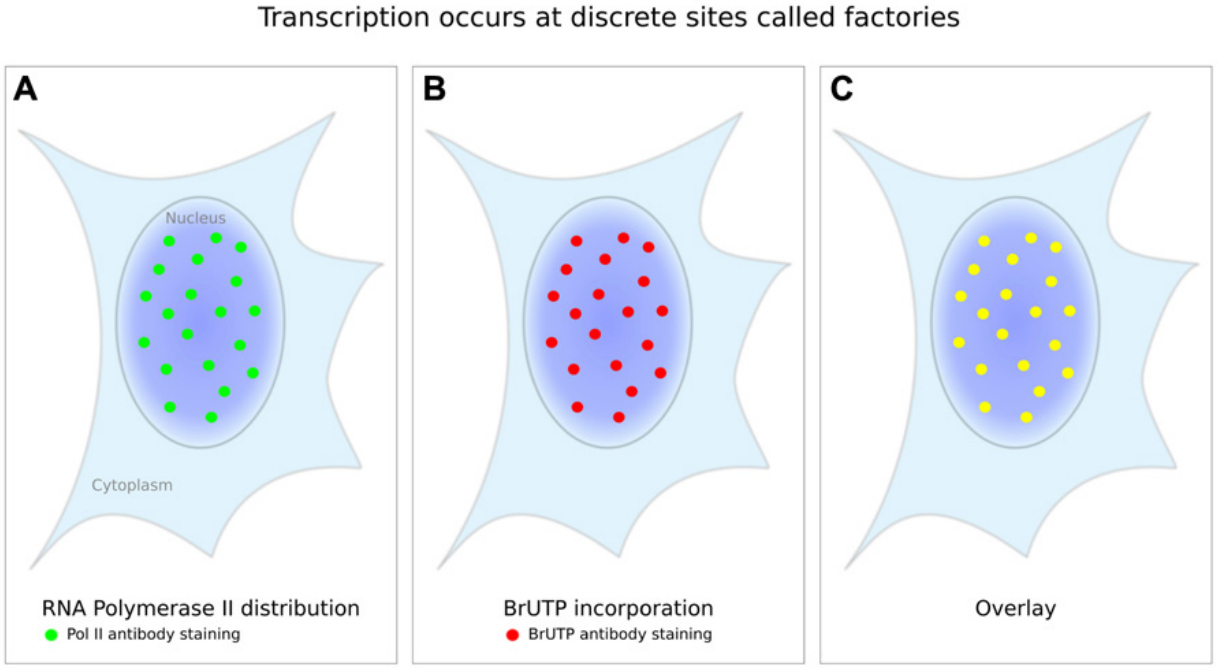
High throughput microarrays and next-generation sequencing technologies have revealed:

1) *temporal* coordination of gene transcription in response to developmental or environmental changes.

2) *spatially* coordination of gene transcription within each cell nucleus

When **RNA polymerase II** is detected by immunofluorescence a non-uniform staining pattern can be observed (*green* dots). **(B)** Labeling of **nascent RNA** by Br-UTP incorporation and subsequent immuno-staining (*red* dots) reveals a staining pattern that matches the polymerase staining as an overlay **(C)** shows (*yellow* dots).

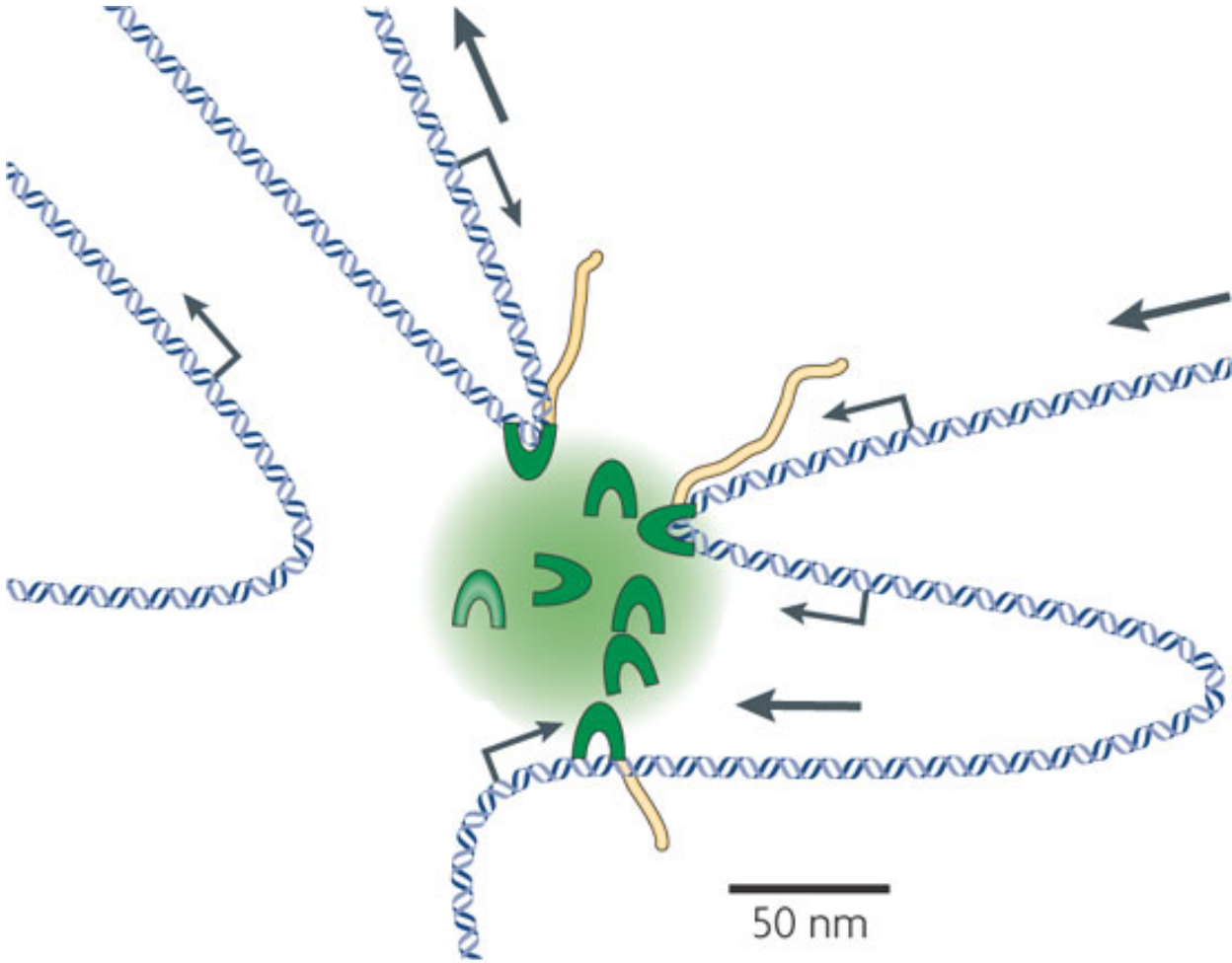
These discrete sites of active transcription are referred to as “***transcription factories***”.



Transcription occurs at discrete sites in the nucleus
termed "***transcription factories***"
where multiple active RNA polymerases are concentrated and anchored to a nuclear substructure.

It shows a ***transcription factory*** with a diameter of 70 nm that *contains eight RNA polymerase II* enzymes (green crescents). Genes are reeled through these polymerases (in the direction of the large arrows) as they are transcribed, and the nascent RNA (yellow) is extruded.

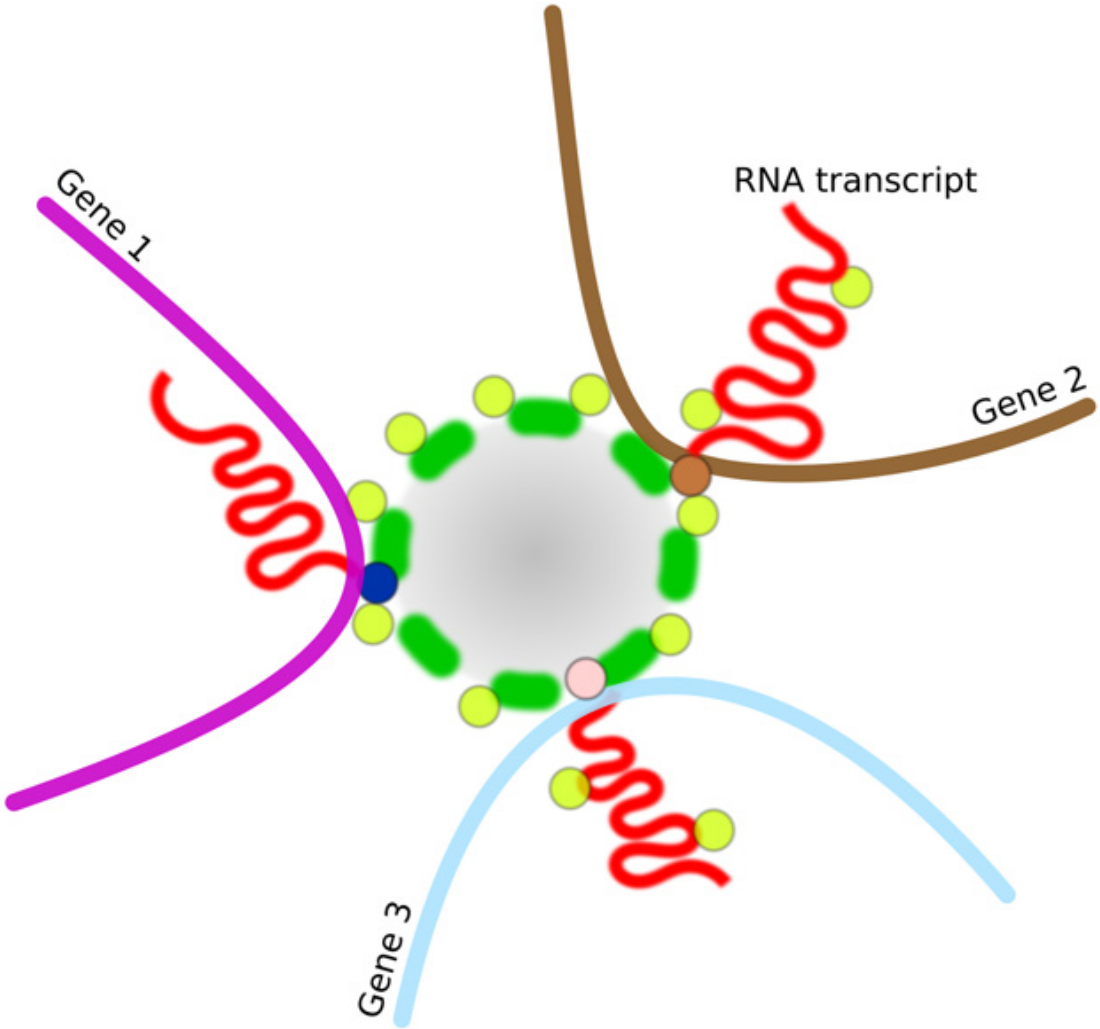
Genes from the *same* or from *different* chromosomes may associate with polymerases in the same factory. Small arrows indicate the direction of transcription at the transcription start site.



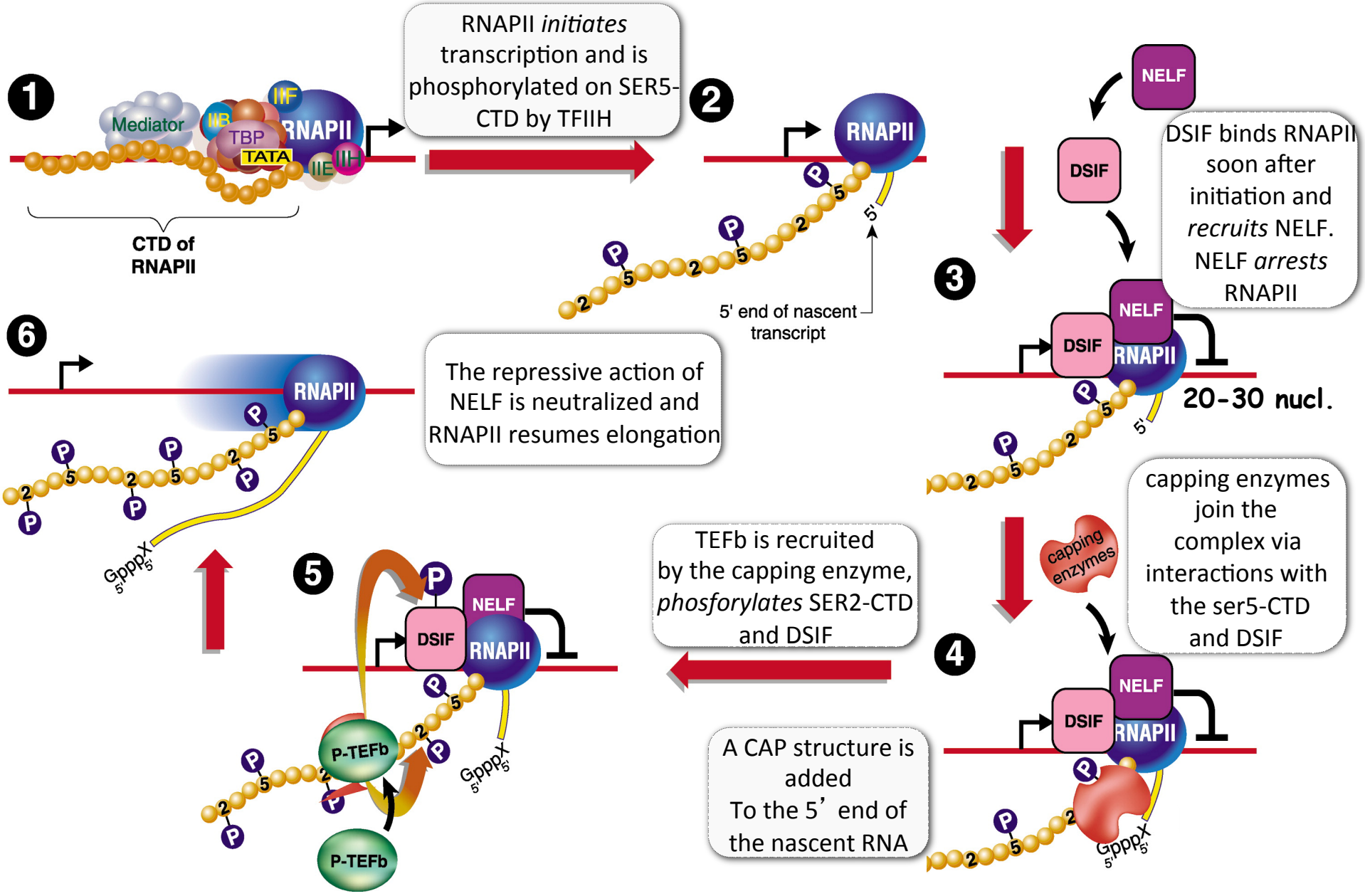
Nature Reviews | Genetics

Structure of a transcription factory

Each factory contains **RNA polymerase II** molecules which are located on the surface of a *protein-rich core* (87 nm in diameter, as determined by EFTEM in HeLa cells). These proteins include many factors involved in *transcription* such as co-activators, *chromatin remodelers*, transcription factors, histone modification enzymes, RNPs, RNA helicases, and *splicing* and *processing factors*. Multiple genes can be processed by the same factory (three are shown).



Capping and transcriptional pausing: checkpoint model



Eukaryotic Transcriptional Regulation

1. Level of Chromatin (*DNA accessibility*)

- Histone modifications
- Histone modifying enzymes & remodeling complexes
- Nucleosome composition
- DNA methylation

2. Level of DNA (*Interaction with basal transcription machinery*)

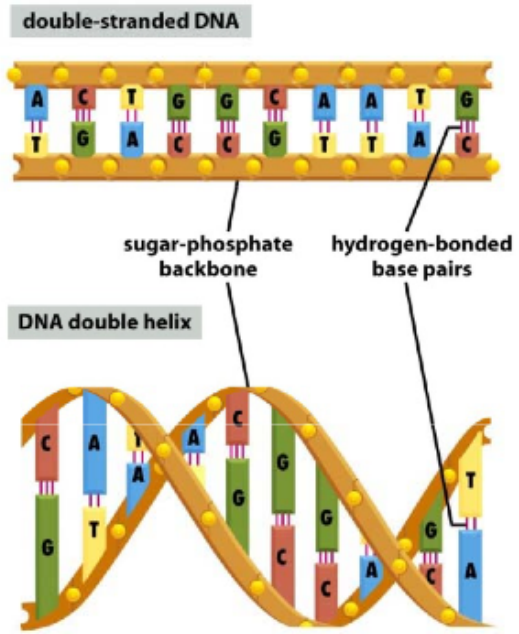
- Regulatory sequences (enhancers, silencers)
- Transcription factors (activators, repressors)

3. Level of *Regulatory RNA* (Interaction with DNA, RNA or protein)

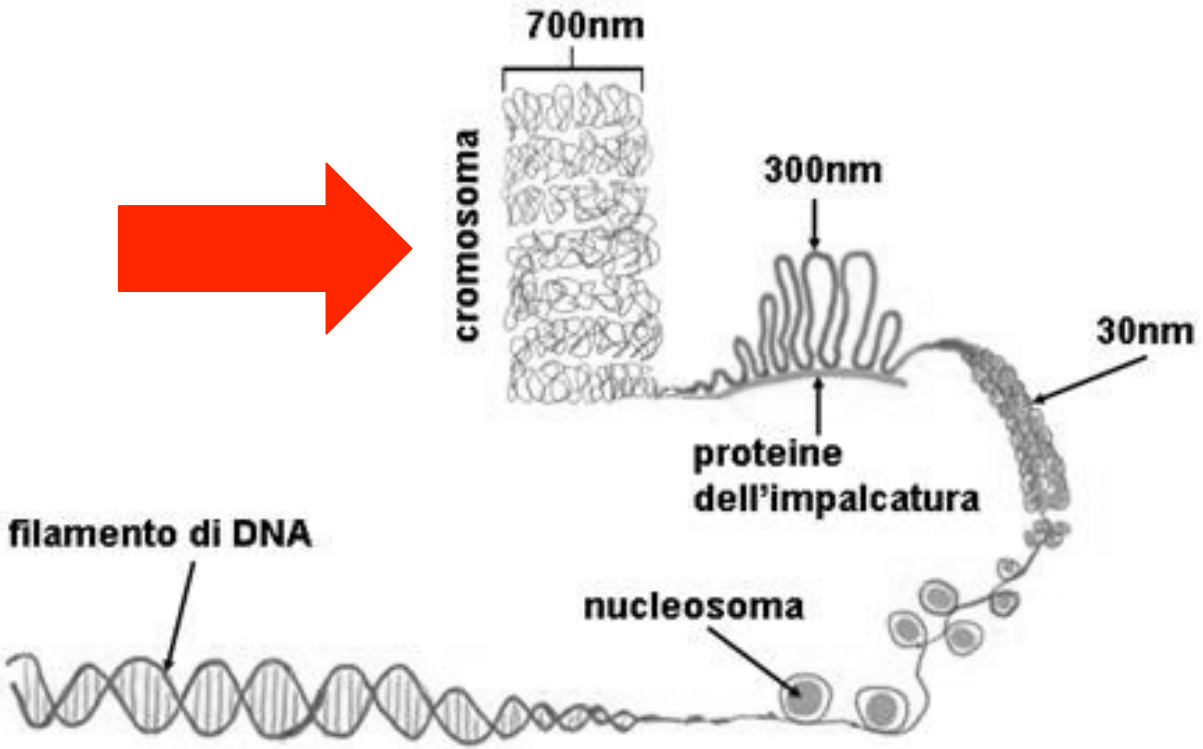
- Small and long non-coding RNAs

The problem: DNA lenght is always higher than the dimension of the compartment in which it is stored

Compartment	Shape	Dimensions	Type of nucleic acid	Length
TMV	filament	0.008 × 0.3 μm	1 single-stranded RNA	2 μm = 6.4 kb
Phage ϕd	filament	0.006 × 0.85 μm	1 single-stranded DNA	2 μm = 6.0 kb
Adenovirus	icosahedron	0.07 μm diameter	1 double-stranded DNA	11 μm = 35.0 kb
Phage T4	icosahedron	0.065 × 0.10 μm	1 double-stranded DNA	55 μm = 170.0 kb
E. coli	cylinder	1.7 × 0.65 μm	1 double-stranded DNA	1.3 μm = 4.2×10^3 kb
Mitochondrion (human)	oblate spheroid	3.0 × 0.5 μm	~10 identical double-stranded DNAs	50 μm = 16.0 kb
Nucleus (human)	spheroid	6 μm diameter	46 chromosomes of double-stranded DNA	1.8 m = 6×10^6 kb



- Spacing between base pairs ≈3.4Å
- For human genome, approximately 3.2 billion base pairs
- Total length ≈ $3.4\times10^{-10}\times3.2\times10^{9}\times2 \approx 2.2$m
- Diameter of a nucleus: $5\sim10\times10^{-6}$m

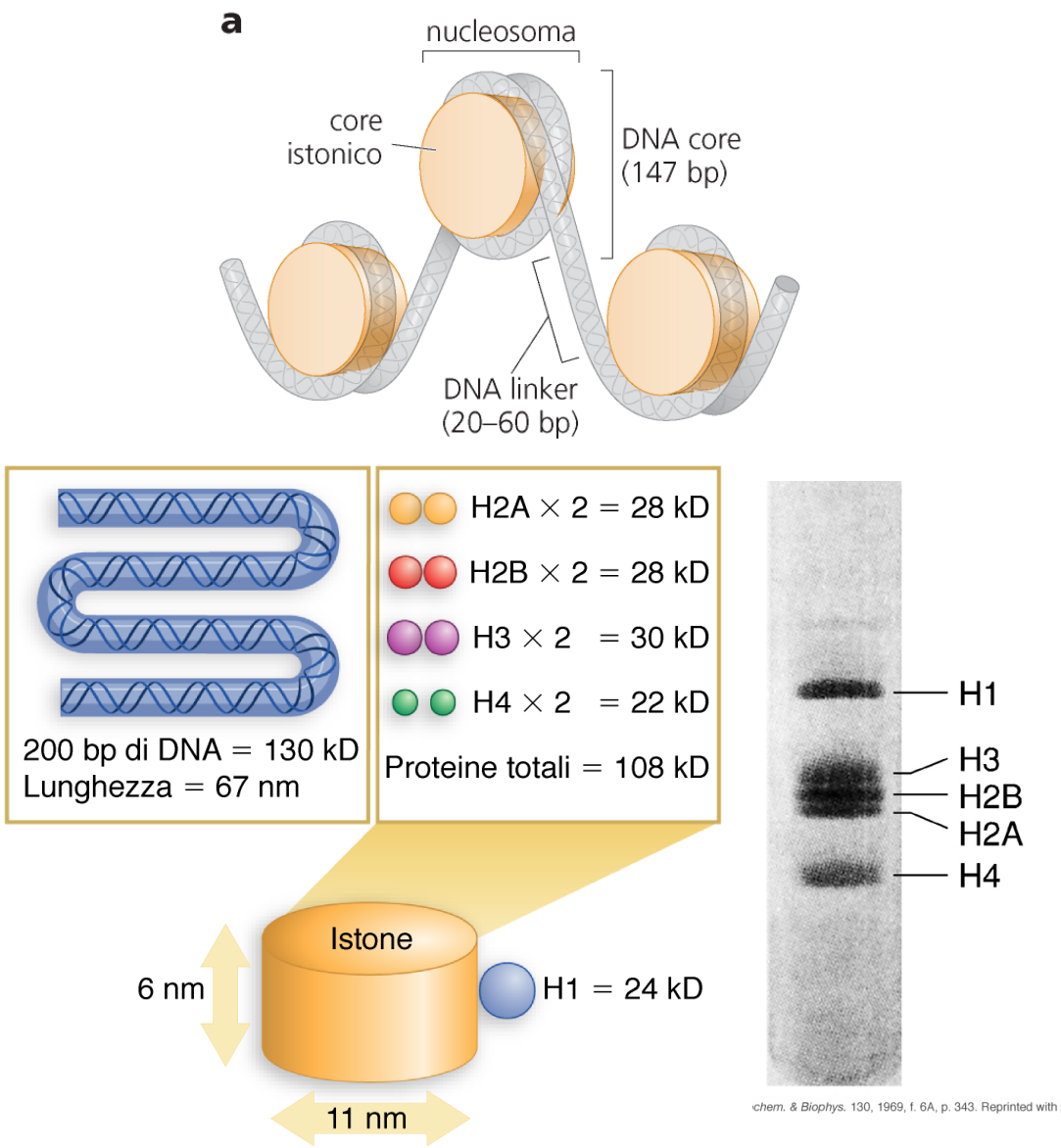


Chromatin is a nucleoprotein complex whose primary function is to pack DNA and to organize eukaryotic genomes.

Packaging is essential for several reasons:

- without such packaging, DNA molecules would be ***too long*** to fit inside cells
- ***damage*** protection
- during cell division, it is essential that DNA remains intact and evenly distributed among cells. Chromosomes are a key part of the process that ensures DNA is accurately *copied* and distributed in the vast majority of cell divisions.

Packaging is accomplished *via* highly conserved proteins called **histones**, which are central components of chromatin. Nucleosome is the fundamental *repeating* unit of chromatin



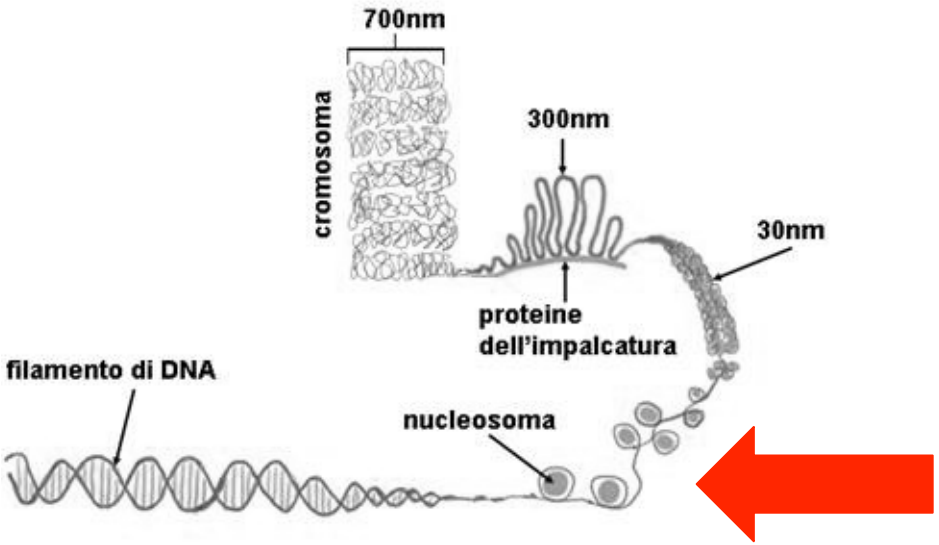
chem. & Biophys. 130, 1969, f. 6A, p. 343. Reprinted with

- two each of the histones **H2A**, **H2B**, **H3**, and **H4** come together to form a histone octamer, which binds and wraps approximately 1.7 turns of DNA, or about **146** base pairs.

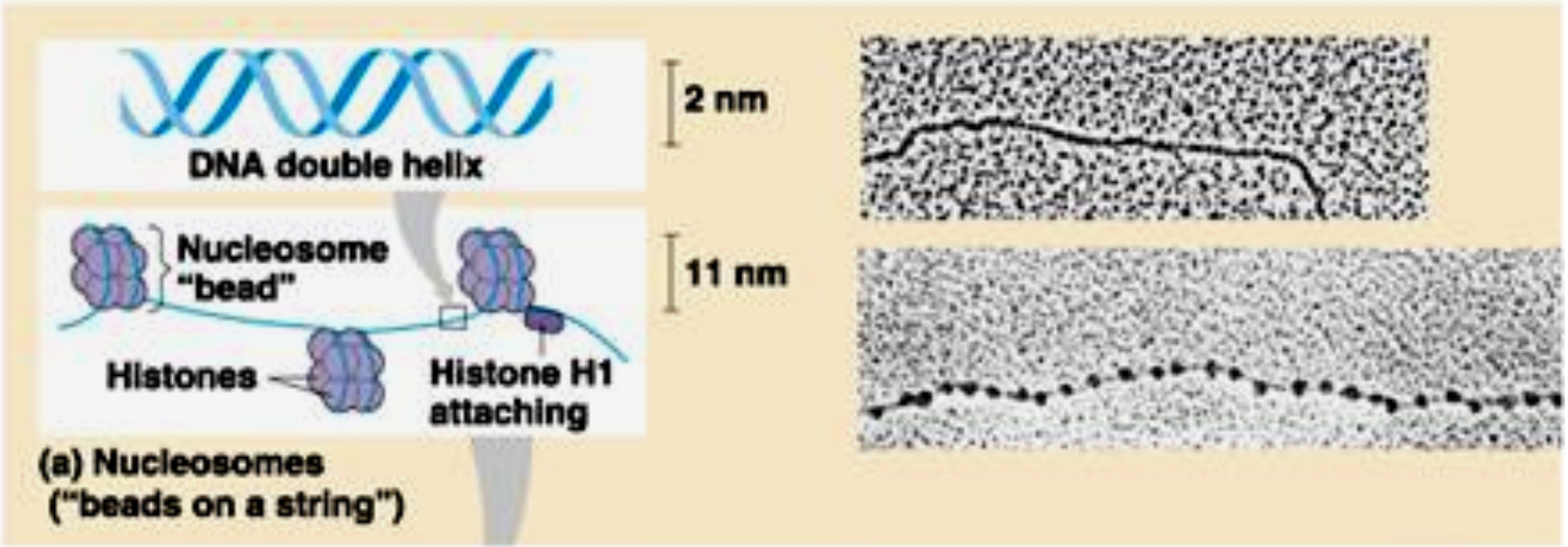
- the *addition* of one **H1** protein *wraps another 20 base pairs*, resulting in two full turns around the octamer, and forming a structure called a *chromatosome*. This joining DNA is referred to as *linker* DNA.

- **Histones** are a family of *small, positively* charged proteins (Van Holde, 1988). *As DNA is negatively charged, due to the phosphate groups in its phosphate-sugar backbone, histones bind with DNA very tightly.*

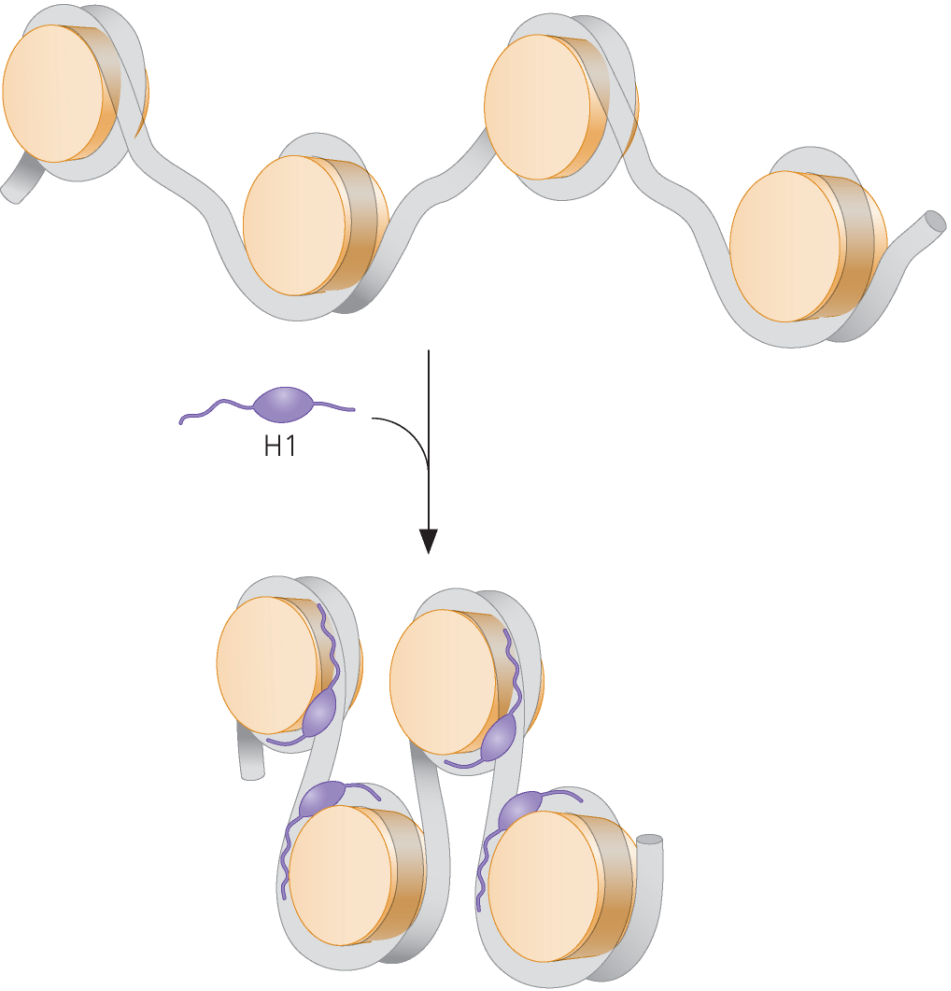
The 10 nm fiber



The observation by electron microscopists that chromatin appeared similar to **beads** on a string



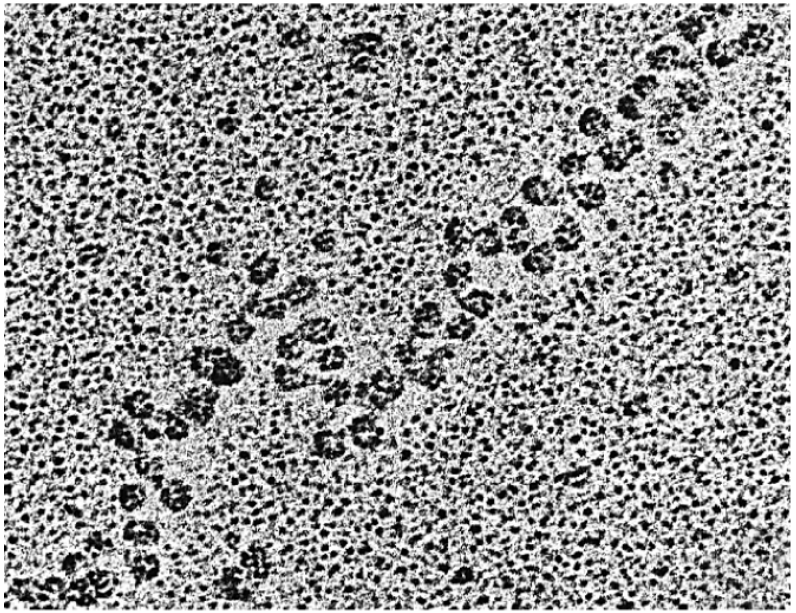
With the help of the linker histone H1 that binds the linker DNA connecting nucleosomes, chromatin forms higher-order structures that enable eukaryotic cells to accommodate and organize genomic DNA inside their nucleus.



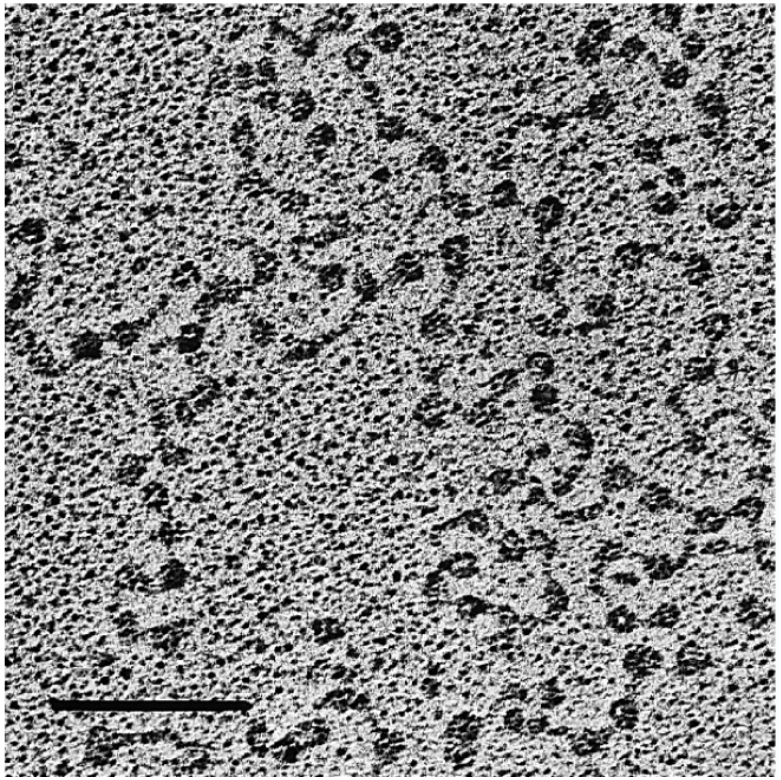
- **H1** is outside the nucleosome and binds DNA at the level of entry/exit points
- **H1** contributes to the formation of higher ordered structures: the 30 nm fiber.

Effect of H1 absence on nucleosomes packaging (Electron Microscopy)

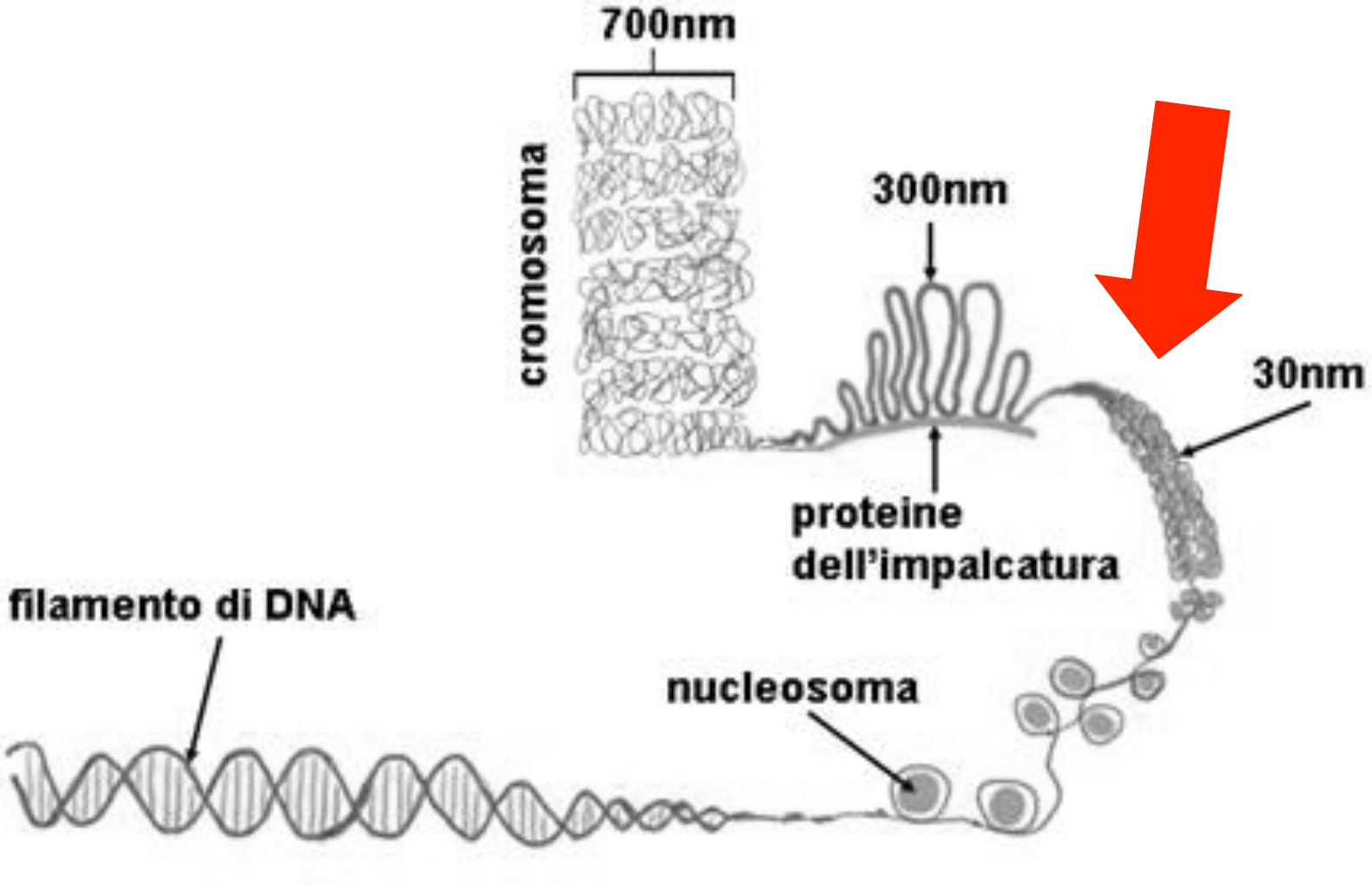
With H1



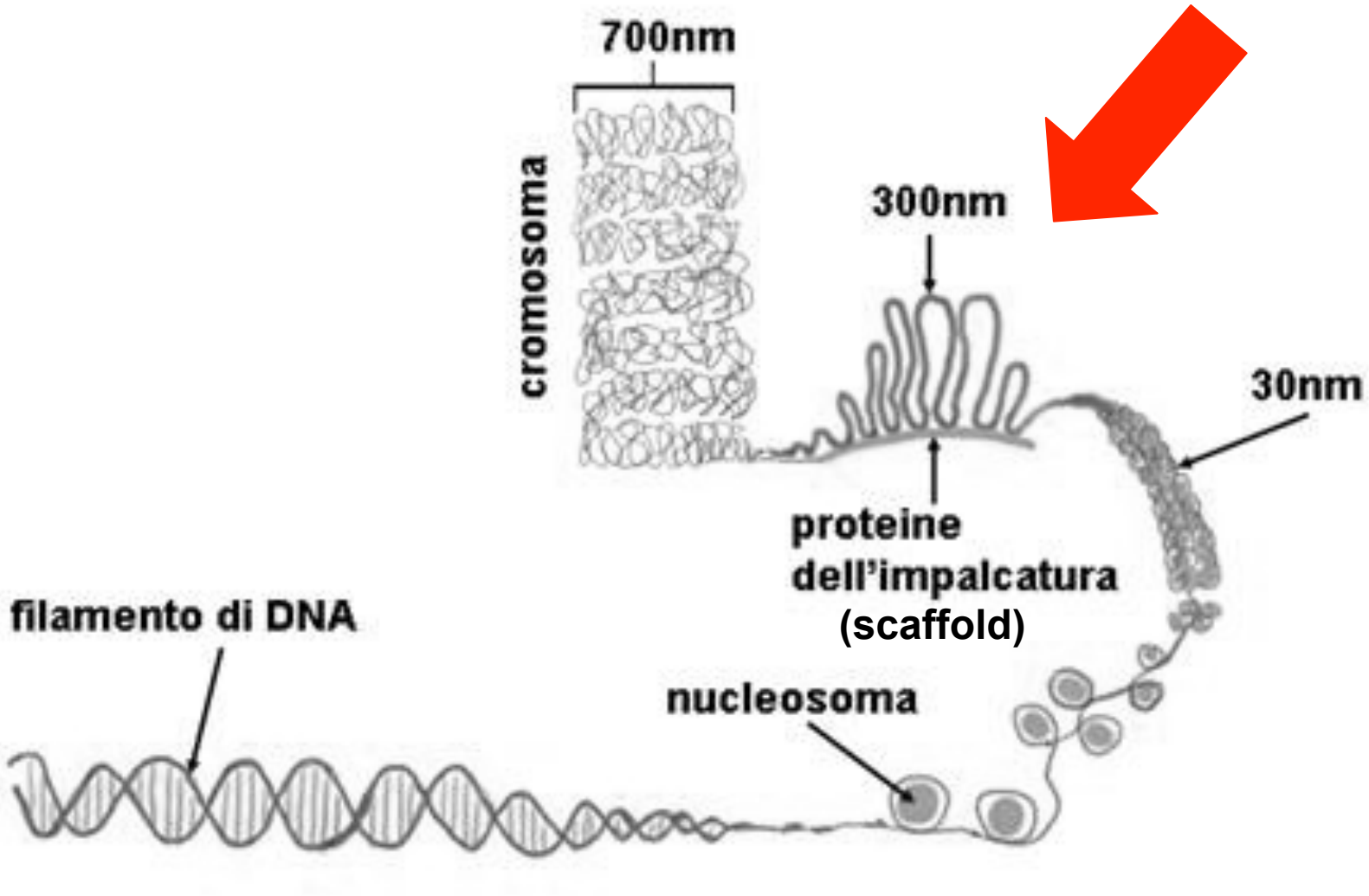
Without di H1



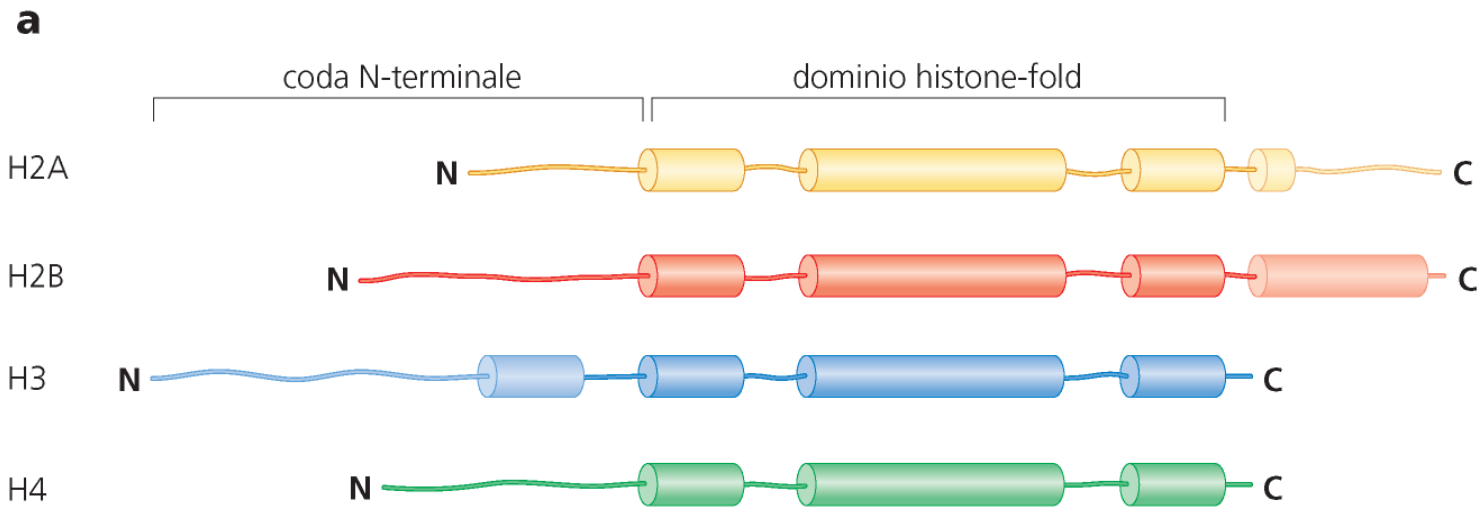
The 30 nm fiber

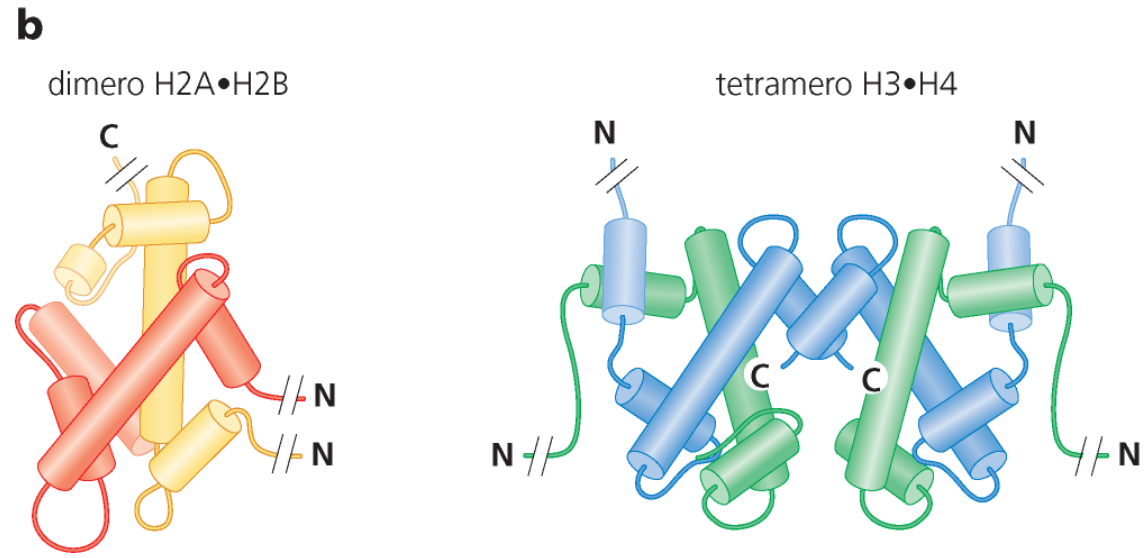


Higher-order structures: the chromatin loops



Histones consist of a flexible tail and a globular core domain that folds into the characteristic histone fold. The ***N-term tails*** protrude outside and interact with DNA

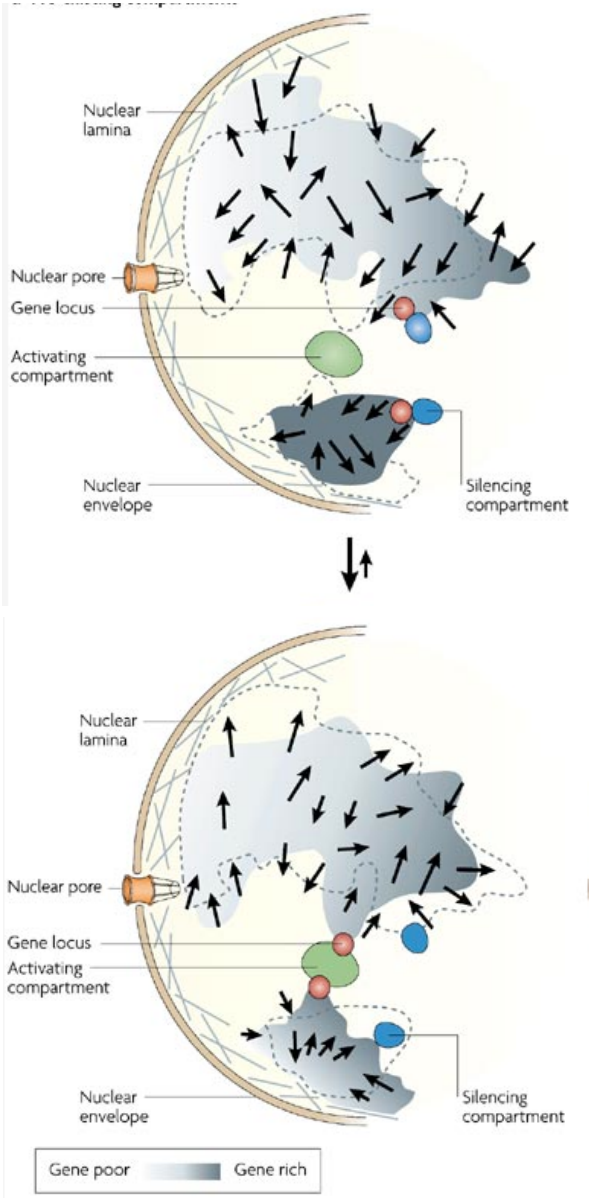




Chromatin is an highly dynamic structure

Chromatin can no longer be considered as merely the sum of independent regions but rather should now be considered as a *flexible* and *interconnected* web in which neighbouring, as well as distant, domains can interact

Chromatin mobility allows **dynamic interactions** between genomic loci and between loci



Gene **repression**

Gene **activation**

FROM THE FOLLOWING ARTICLE:
Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions
Christian Lanctôt, Thierry Cheutin, Marion Cremer, Giacomo Cavalli & Thomas Cremer
Nature Reviews Genetics 8, 104-115 (February 2007)

Different ways to modify the chromatin exists which regulates gene expression, DNA repair, replication and recombination.

1) Histone variants (i.e. H2A e H3)

2) Post-translational histone modifications (PTM): acetilation, methylation etc.…

3) DNA modifications

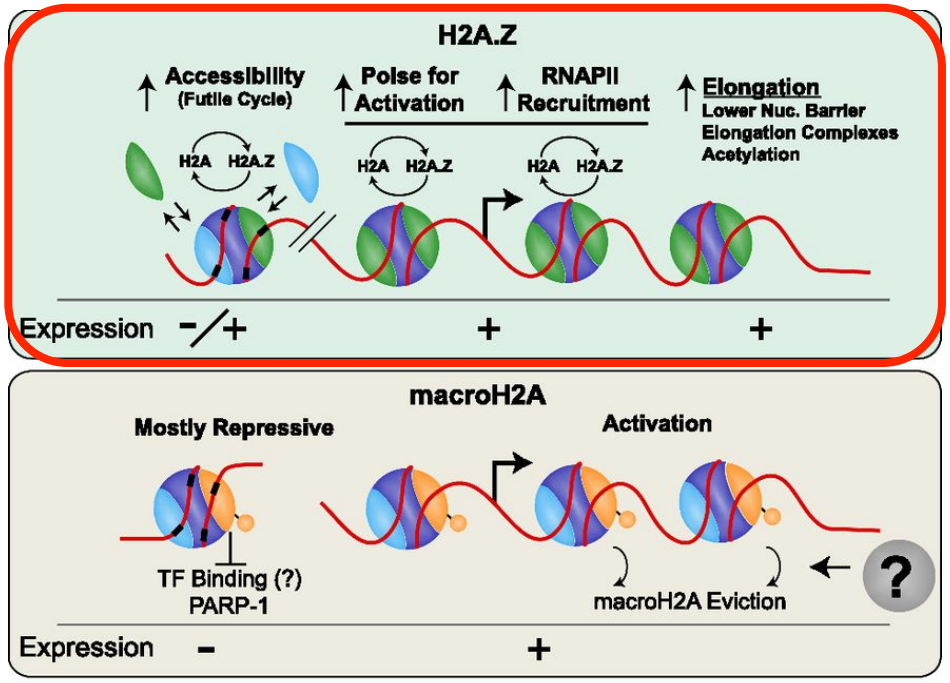
4) Non coding RNA

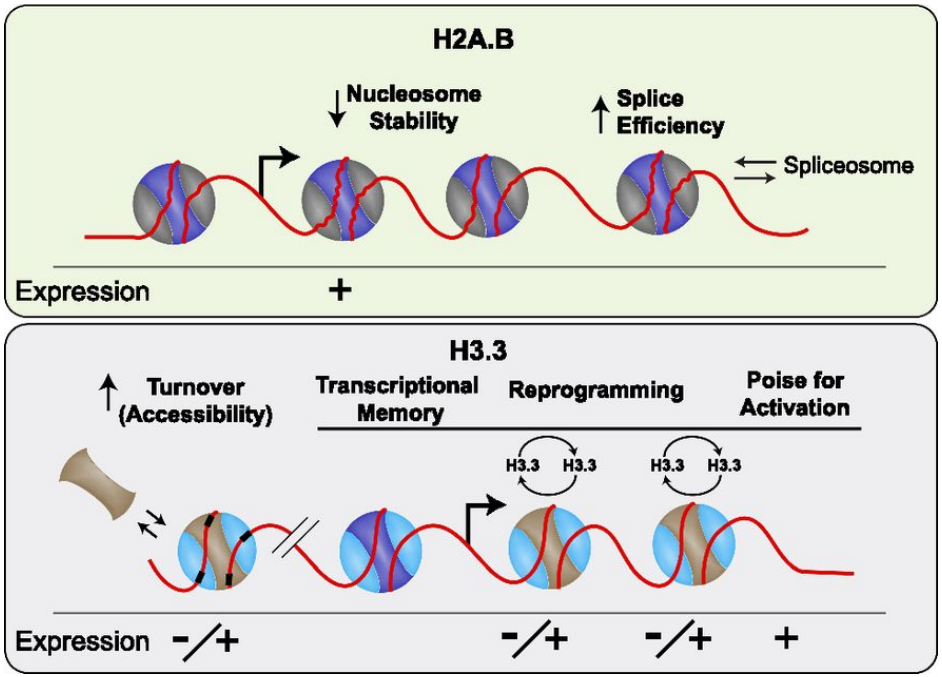
1) Histone variants

- Histone variants have ***distinct*** **amino acid sequences**
- histone variants alter nucleosome structure, stability, dynamics, and, ultimately, DNA accessibility.
- Canonical histones are deposited in a *replication-coupled* manner to package the newly replicated genome. In contrast, histone variants are expressed *throughout the cell cycle* and *replace* canonical histones when nucleosomes are evicted

During transcription, histone variants shape the chromatin landscape of *cis*-regulatory and coding regions in support of specific transcription programs.

General role of histone variants on transcriptional regulation





Christopher M. Weber, and Steven Henikoff Genes Dev. 2014;28:672-682

generally positive role for H2A.Z in transcription

H2A variants are the most diverse, perhaps reflective of relaxed structural constraint within the nucleosome.

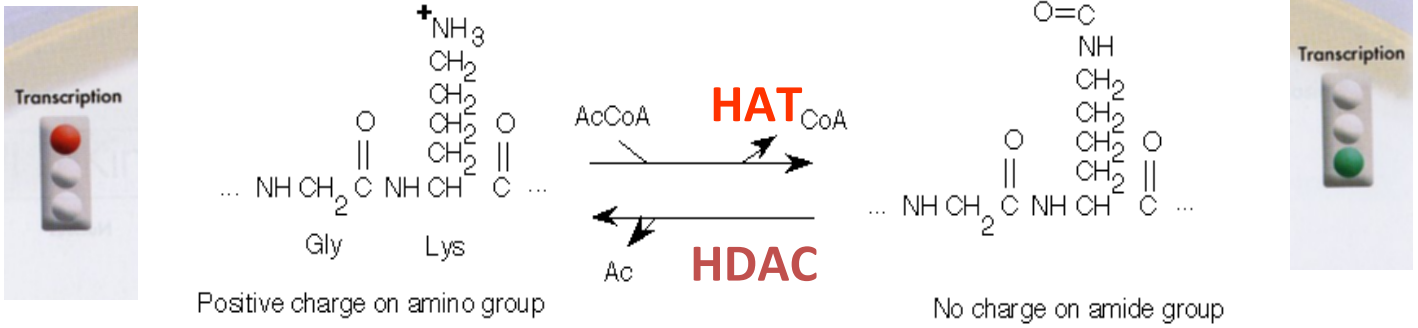
One such variant, H2A.Z, is only ~60% identical to H2A

2) Post-translational histone modifications (PTM)

PTM have a number of different functions

- histone modifications result in a change in the net charge of nucleosomes, which could loosen inter- or intranucleosomal DNA-histone interactions.

This is supported by the observation that acetylated histones are easier to displace from DNA

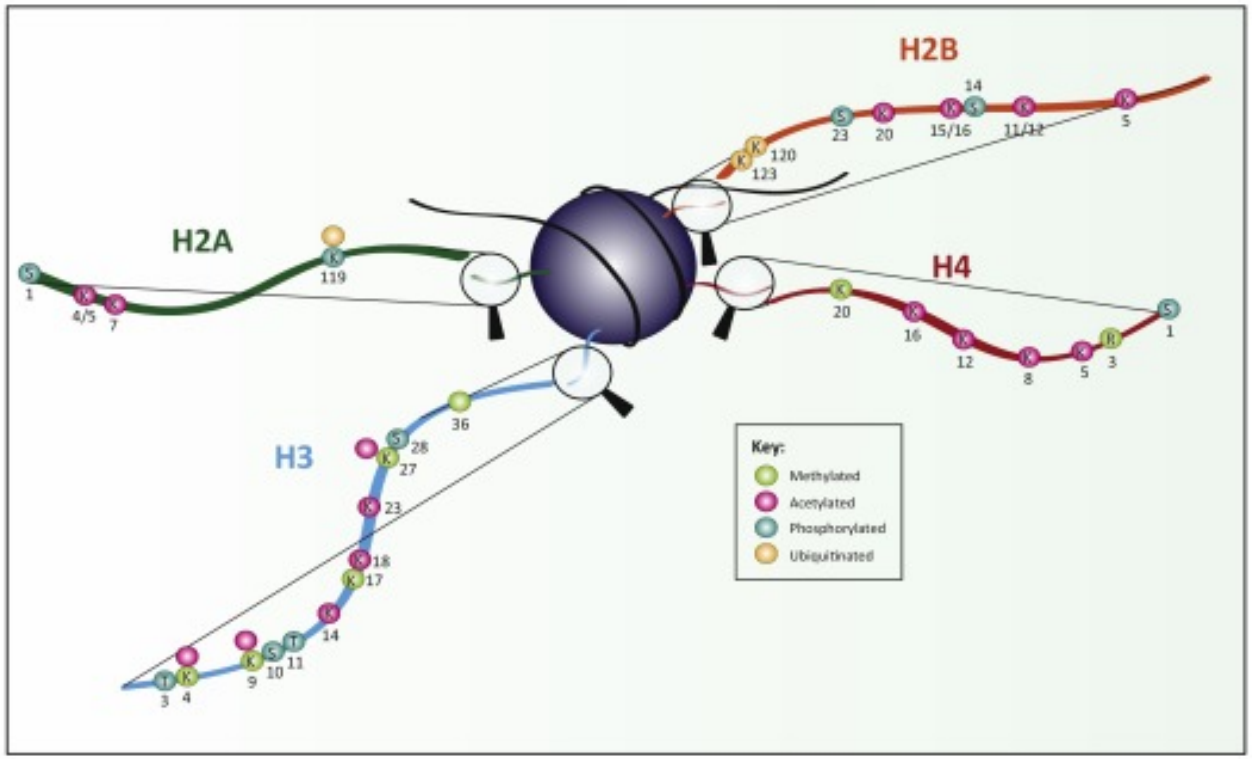


- individual histone modifications or modification patterns are read by other proteins that influence chromatin dynamics and function

- some modifications directly influence higher-order chromatin structure.
  H4 K16 inhibits the formation of compact 30 nm fibers

2) Post-translational histone modifications (PTM)

The best-studied modifications are those occurring on the N-terminal 'tail' regions of the histones, which project from the nucleosome and are accessible on its surface



These modifications include:

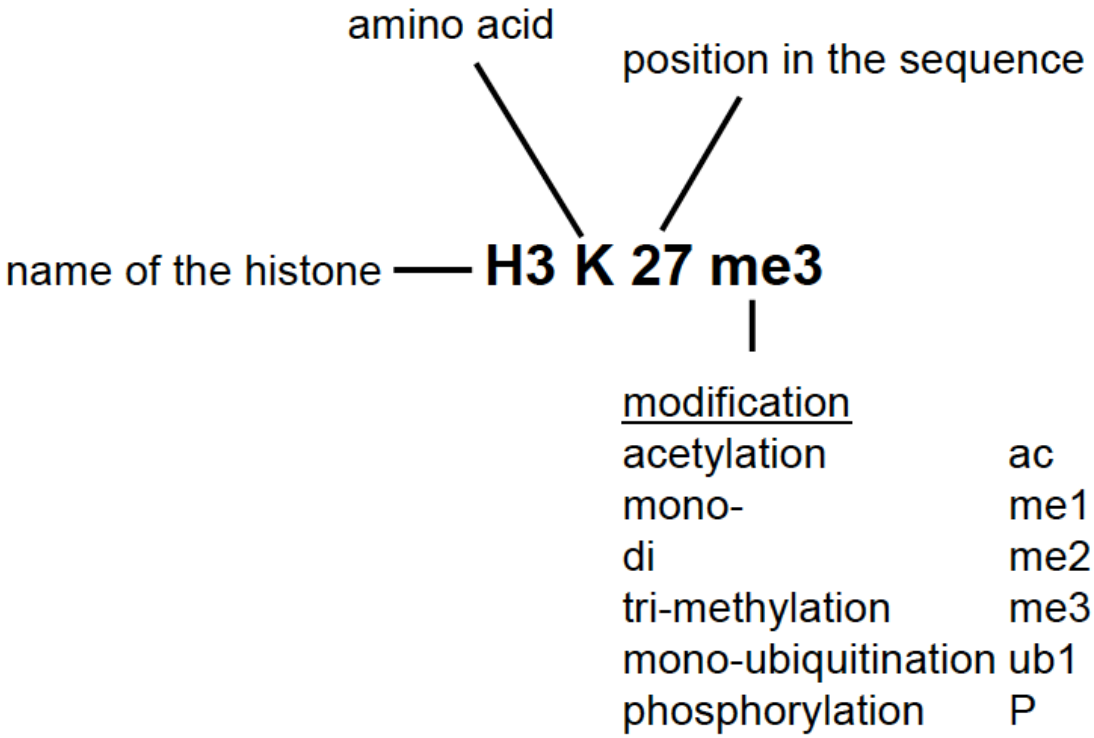
- Acetylation
- Methylation
- Phosphorylation
- Ubiquitinylation
- Sumoylation
- ADP ribosylation
- Deamination

Lateral Thinking: How Histone Modifications Regulate Gene Expression

Moyra Lawrence, Sylvain Daujat, Robert Schneider

2) Post-translational histone modifications (PTM)

Nomenclature



Covalent modifications of N-term tails: NOMENCLATURE

1) Histone acetylases (HATs)
2) Histone de-acetylases (HDACs)
3) Histone methylases (HMTs)
4) Histone de-methylases (HDMs)
5) Histone ubiquitin ligases

Some of the modifications in these tails can directly affect the interactions between nucleosomes.

The **H4K16ac** has been shown to *reduce* chromatin compaction and increase transcription

Histone tail modifications can also do the reverse and increase DNA compaction; for example, **H4K20** di- and tri-methylation, which have been shown to *enhance* chromatin condensation

Table 1. Histone Tail Modifications

Histone	Modification	Role
H2A	H2AS1P	Mitosis; chromatin assembly
	H2AK4/5ac	Transcriptional activation
	H2AK7ac	Transcriptional activation
	H2AK119P	Spermatogenesis
	H2AK119uq	Transcriptional repression
H2B	H2BS14P	Apoptosis
	H2BS33P	Transcriptional activation
	H2BK5ac	Transcriptional activation
	H2BK11/12ac	Transcriptional activation
	H2BK15/16ac	Transcriptional activation
	H2BK20ac	Transcriptional activation
	H2BK120uq	Spermatogenesis/meiosis
	H2BK123uq	Transcriptional activation
H3	H3K4me2	Permissive euchromatin
	H3K4me3	Transcriptional elongation; active euchromatin
	H3K9me3	Transcriptional repression; imprinting; DNA methylation
	H3R17me	Transcriptional activation
	H3K27me3	Transcriptional silencing; X-inactivation; bivalent genes/gene poising
	H3K36me3	Transcriptional elongation
	H3K4ac	Transcriptional activation
	H3K9ac	Histone deposition; transcriptional activation
	H3K14ac	Transcriptional activation; DNA repair
	H1K18ac	Transcriptional activation; DNA repair; DNA replication
	H3K23ac	Transcriptional activation; DNA repair
	H3K27ac	Transcriptional activation
	H3T3P	Mitosis
	H3S10P	Mitosis; meiosis; transcriptional activation
	H3T11/S28P	Mitosis
H4	H4R3me	Transcriptional activation
	H4K20me1	Transcriptional silencing
	H4K20me3	Heterochromatin
	H4K5ac	Histone deposition; transcriptional activation; DNA repair
	H4K8ac	Transcriptional activation; DNA repair; transcriptional elongation
	H4K12ac	Histone deposition; telomeric silencing; transcriptional activation; DNA repair
	H4K16ac	Transcriptional activation; DNA repair
	H4S1P	Mitosis

Lateral Thinking: How Histone Modifications Regulate Gene Expression

Moyra Lawrence, Sylvain Daujat, Robert Schneider

As well as the tails, other regions of the histone can also be modified. The **central globular** domain of the histones, which together form the core of the nucleosome, also contain a large number of modification sites

The first core modification to be discovered, **H3 lysine 79**, has also been the most extensively characterized. It plays a fundamental role in the regulation of chromatin structure

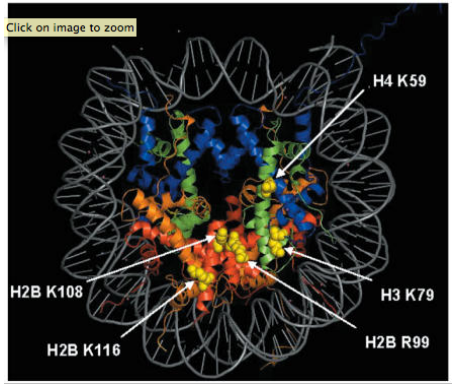


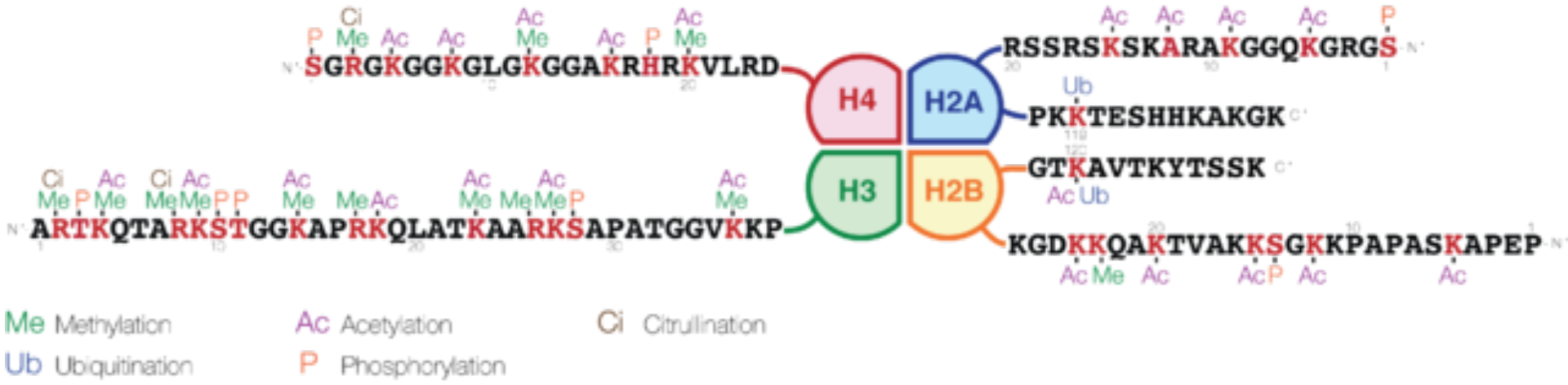
Table 2. Histone Globular Domain Modifications

Histone	Site	Modification	Refs
H2A	H2AK36	Acetylation	[24,96]
	H2AK99	Methylation	[24]
	H2AQ105	Methylation	[89]
	H2AK119	Acetylation	[24]
	H2AK119	Ubiquitylation	[102]
H2B	H2BK40	Methylation	[24]
	H2BK82	Acetylation	[24]
	H2BR96	Methylation	[105]
	H2BK105	Acetylation	[100]
	H2BK113	Acetylation	[100]
	H2BK117	Acetylation	[105]
H3	H3Y41	Phosphorylation	[91]
	H3R42	Methylation	[88]
	H3T45	Phosphorylation	[92]
	H3R53	Methylation	[24]
	H3K56	Acetylation	[64]
	H3K56	Methylation	[56]
	H3K64	Acetylation	[25]
	H3K64	Methylation	[81]
	H3K79	Methylation	[26–28]
	H3K115	Acetylation	[24]
	H3T118	Phosphorylation	[94–96]
	H3K122	Acetylation	[23]
H4	H4K31	Acetylation	[24]
	H4S47	Phosphorylation	[116]
	H4K59	Methylation	[96]
	H4K77	Acetylation	[24]
	H4K79	Acetylation	[117]
	H4K91	Acetylation	[117,118]
	H4R92	Methylation	[24]

The histone CODE

A huge catalogue of histone modifications have been described.

Collectively, it is thought that histone modifications may underlie a **histone code**, whereby combinations of histone modifications have specific meanings.

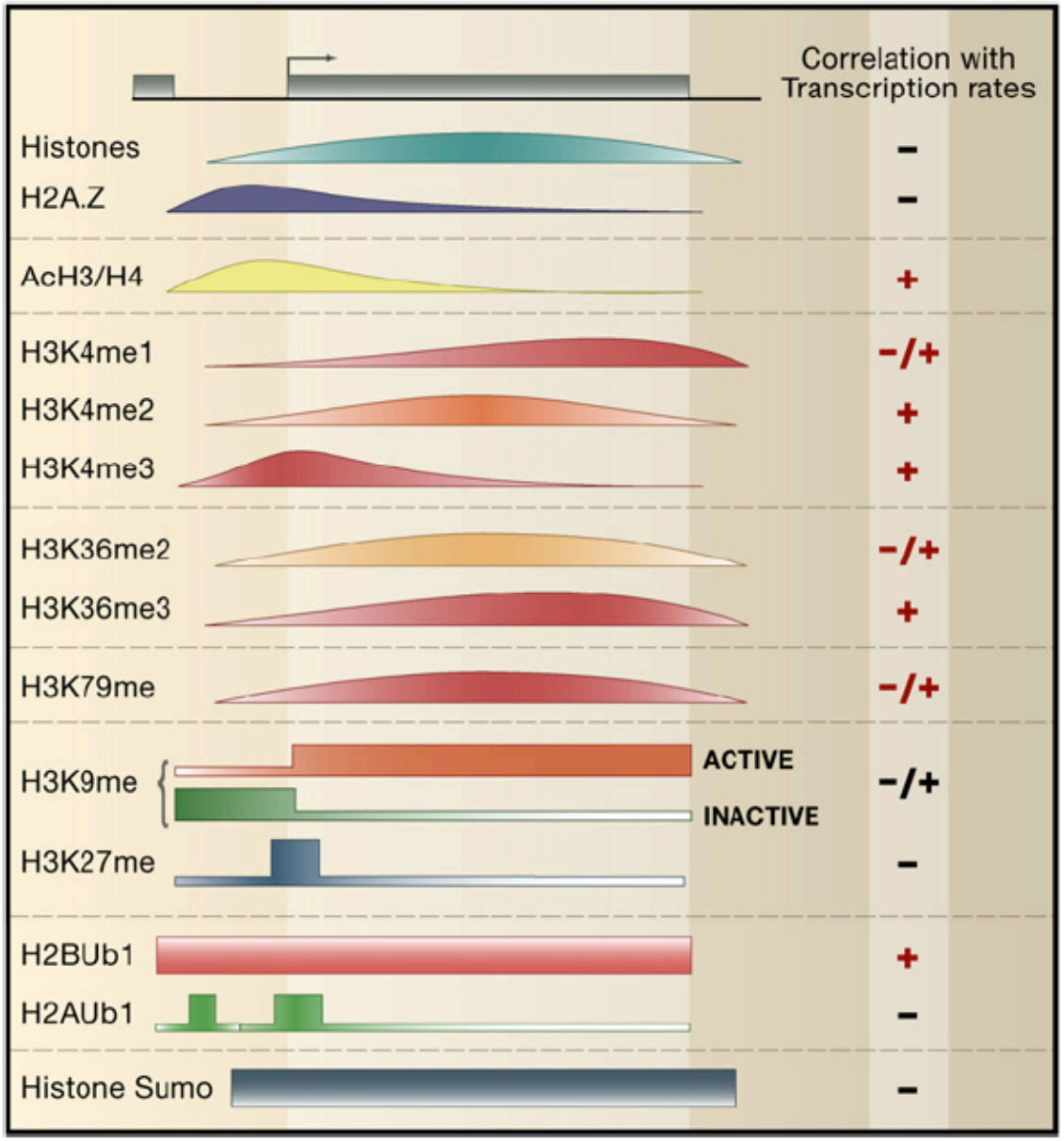


Histone post-transcriptional modifications and readout

1. Type of modification
   - Which amino-acid
   - Number of modifications (me)
2. Position in genome
   - Promoter: H3K36me, H3K9me are repressive
   - Coding region: H3K36me, H3K9me are activating and prevent cryptic initiation of transcription in ORF
3. Other histone modifications
   - combinatorial (occur together)
   - H3K4me + H3K9me: transcriptional activation
   - H4K20me + H3K9me: heterochromatin formation
   - H3K27me + H3K4me: "bivalent" mark in stem cells
4. Size of histone modification domain
   - large: heritable (can be copied more easily)
     - H3K27me can recruit PRC2 has H3K27me3 activity
     - H3K4me recruits WDR5 (MLL thrithorax): H3K4me
5. Cycles of modifications
   - H2Bub → H2B required for transcriptional elongation

Histone code

Most modifications are distributed in **distinct localized patterns** within the *upstream region*, the *core promoter*, *the 5' end* of the open reading frame (ORF) and the *3' end* of the ORF.



The *location* of a modification is *tightly* regulated and is crucial for its effect on transcription.

Histone modifications associated with transcription

Modifications	Position		Enzymes				Recognition Module(s)[a]	Functions in Transcription
			S. cerevisiae	*S. pombe*	*Drosophila*	Mammals		
Methylation	H3	K4	Set1	Set1	Trx, Ash1	MLL, ALL-1, Set9/7, ALR-1/2, ALR, Set1	PHD, Chromo, WD-40	Activation
		K9	n/a	Clr4	Su(var)3-9, Ash1	Suv39h, G9a, Eu-HMTase I, ESET, SETBD1	Chromo (HP1)	Repression, activation
		K27				E(Z)	Ezh2, G9a	Repression
		K36	Set2			HYPB, Smyd2, NSD1	Chromo(Eaf3), JMJD	Recruiting the Rpd3S to repress internal initiation
		K79	Dot1			Dot1L	Tudor	Activation
	H4	K20		Set9	PR-Set7, Ash1	PR-Set7, SET8	Tudor	Silencing
Arg Methylation	H3	R2				CARM1		Activation
		R17				CARM1		Activation
		R26				CARM1		Activation
	H4	R3				PRMT1	(p300)	Activation
Phosphorylation	H3	S10	Snf1				(Gcn5)	Activation
Ubiquitination	H2B	K120/123	Rad6, Bre1	Rad6		UbcH6, RNF20/40	(COMPASS)	Activation
	H2A	K119				hPRC1L		Repression
Acetylation	H3	K56					(Swi/Snf)	Activation
	H4	K16	Sas2, NuA4		dMOF	hMOF	Bromodomain	Activation
	Htz1	K14	NuA4, SAGA					Activation

[a] The proteins that are indicated within the parentheses are shown to recognize the corresponding modifications but specific domains have yet to be determined.

Polycomb and MLL/Trithorax Complexes

The ON and OFF states of key developmental genes are maintained by the **polycomb group (PcG)** and **MLL/Trithorax (Trx)** proteins, which mediate H3K27me3 to repress genes or H3K4me3 to activate genes.

Polycomb-group Proteins

- Maintains a silenced state
- Prevents chromatin remodelling

Trithorax-group Proteins

- Maintains an active state
- Counteracts the action of PcG proteins

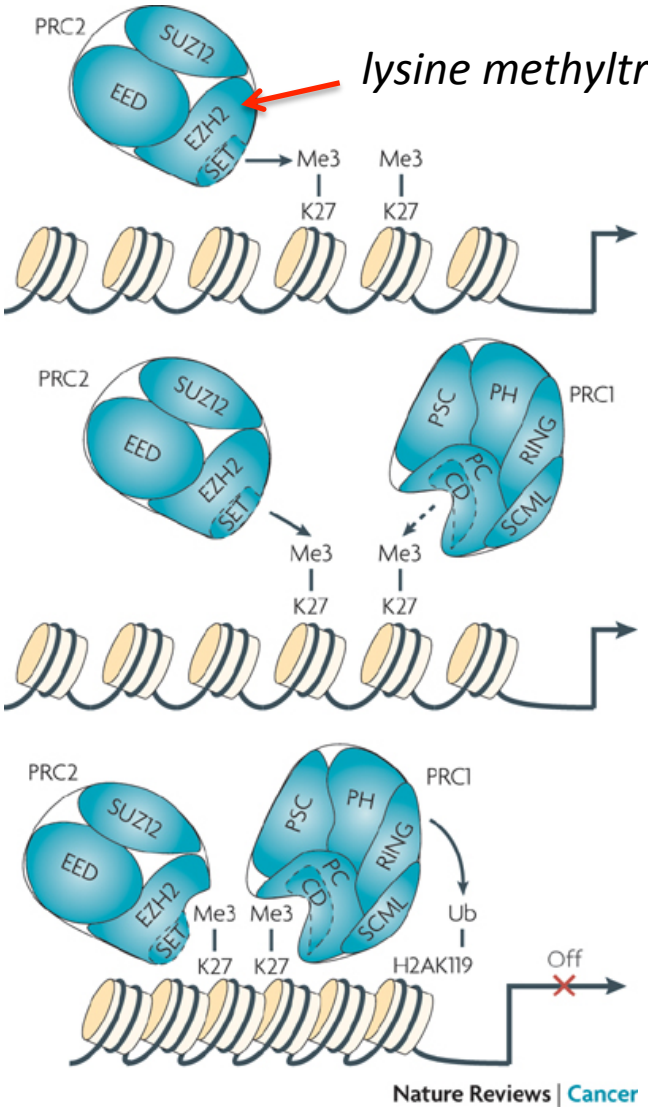
Polycomb group proteins (PcG)

Polycomb repressive complexes (PRCs), **repress** transcription by a mechanism that involves chromatin *modification*. *Two* major Polycomb repressive complexes (PRCs) have been described:

•The **PRC2** contains the histone methyltransferase **EZH2**, which together with **EED** and **SUZ12** catalyses the **H3K27me3.**

•The **PRC1** complexes are recruited by the affinity of chromodomains in chromobox (Cbx) proteins to the H3K27me3 mark.

PRC1 recruitment results in the **ubiquitylation of H2A on lysine 119** *via* the ubiquitin ligases Ring1a or Ring1b, which is thought to be important for transcriptional repression.



How mammalian PRC2 is recruited to chromatin is not clear

In *Drosophila*, DNA sequences called Polycomb Response Elements (PRE) are targets for PcG protein recruitment when inserted at exogenous locus.

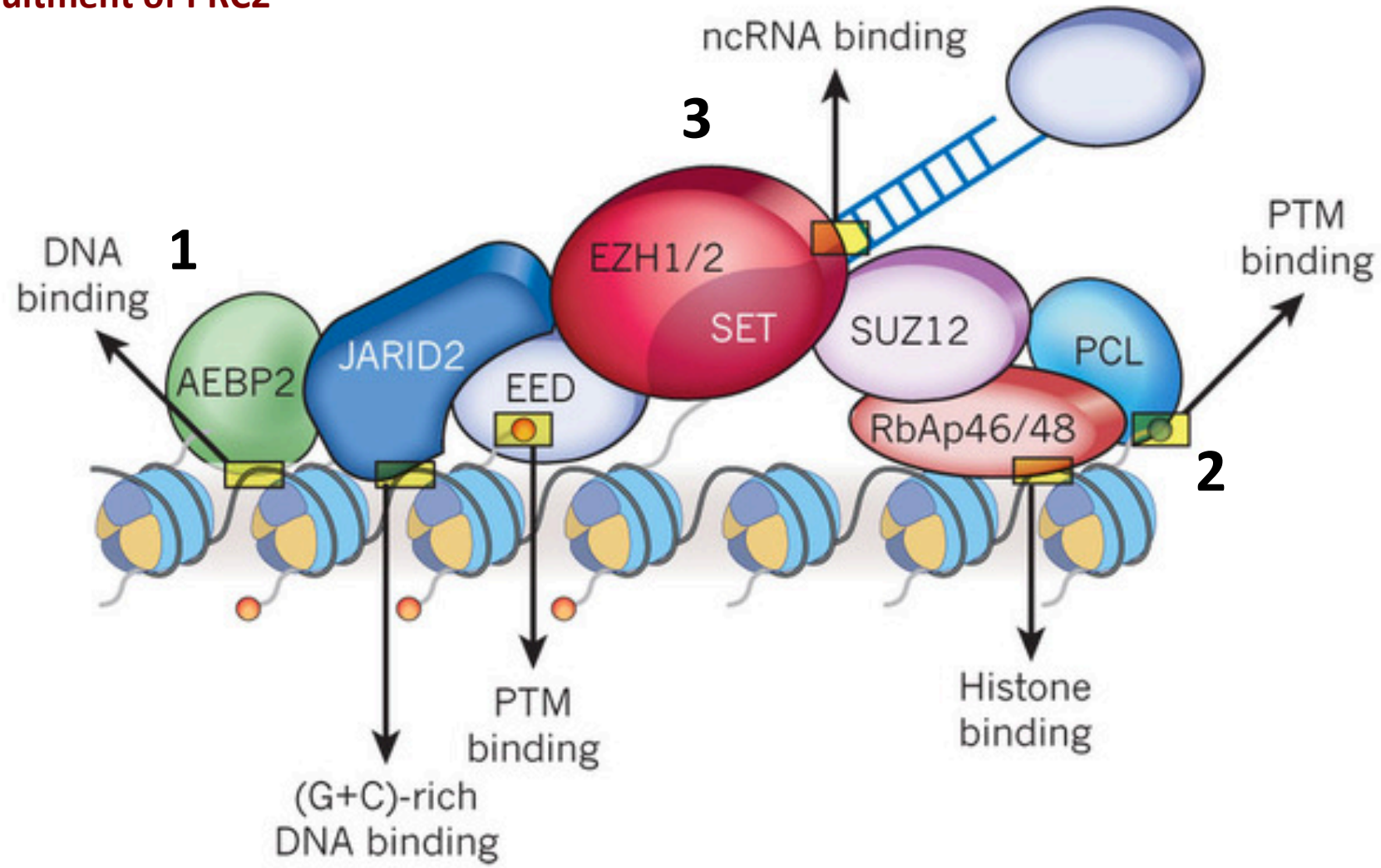
In *mammals*, PRC2 occupies chromatin *enriched* in CpG, but these sequences alone do not indicate a consensus response element

How do these enzymes

which *lack* DNA binding capacity

recognise their target genes in the various cell types

Model which predicts the steps, not necessarily consecutive, that result in the successful recruitment of PRC2



Schematic representation of the PRC2 holoenzyme at chromatin. Putative interactions with either DNA or histones that could explain PRC2 recruitment are highlighted.

The Polycomb complex PRC2 and its mark in life

Raphaël Margueron & Danny Reinberg *Nature* **469**, 343–349 (20 January 2011)

Eukaryotic Transcriptional Regulation

1. Level of Chromatin (*DNA accessibility*)

- Histone modifications
- Histone modifying enzymes & remodeling complexes
- Nucleosome composition
- DNA methylation

2. Level of DNA (*Interaction with basal transcription machinery*)

- Regulatory sequences (enhancers, silencers)
- Transcription factors (activators, repressors)

3. Level of *Regulatory RNA* (Interaction with DNA, RNA or protein)

- Small and long non-coding RNAs

Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression

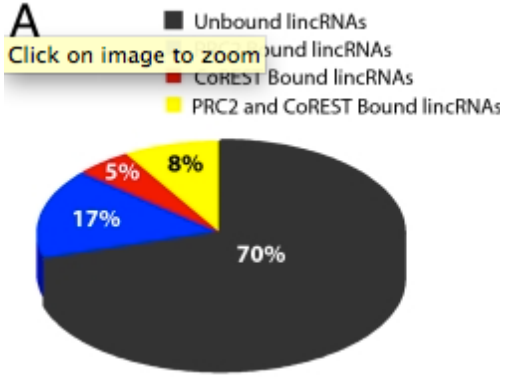
Ahmad M. Khalil,[a,b,1] Mitchell Guttman,[a,c,1] Maite Huarte,[a,b] Manuel Garber,[a] Arjun Raj,[d] Dianali Rivea Morales,[a,b] Kelly Thomas,[a,b] Aviva Presser,[a] Bradley E. Bernstein,[a,e] Alexander van Oudenaarden,[d] Aviv Regev,[a,c] Eric S. Lander,[a,c,f,1,2] and John L. Rinn[a,b,1,2]

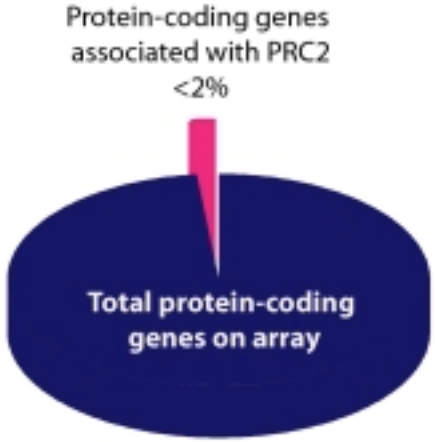
Author information ▶ Article notes ▶ Copyright and License information ▶

ABSTRACT

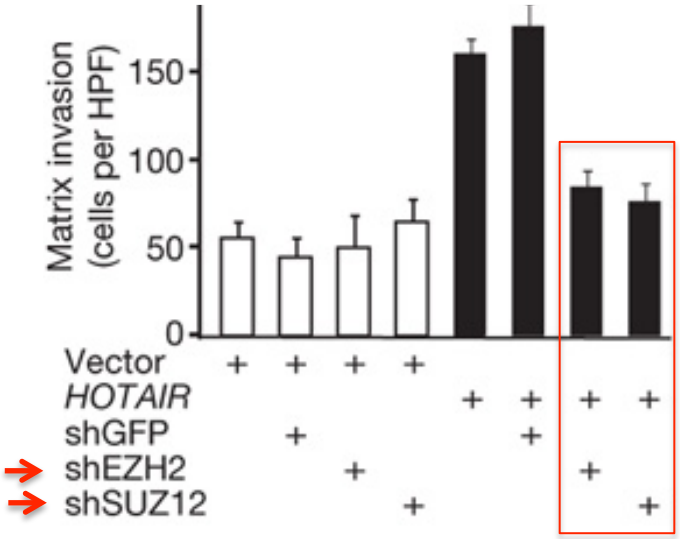
Go to:

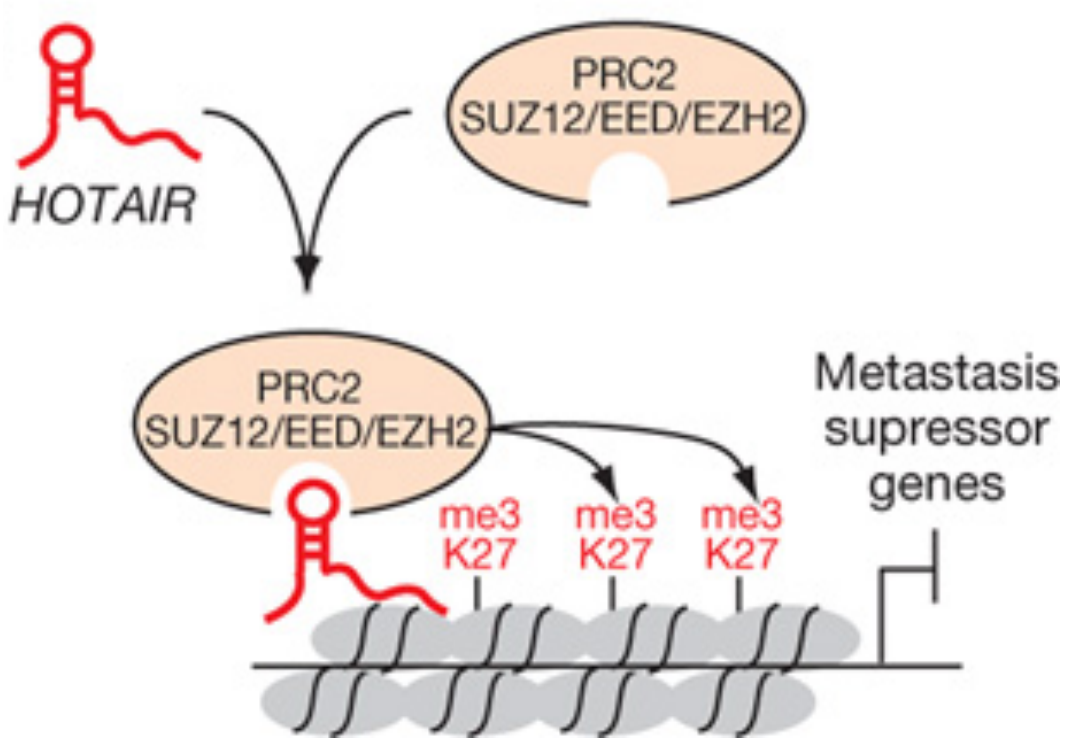
We recently showed that the mammalian genome encodes >1,000 large intergenic noncoding (linc)RNAs that are clearly conserved across mammals and, thus, functional. Gene expression patterns have implicated these lincRNAs in diverse biological processes, including cell-cycle regulation, immune surveillance, and embryonic stem cell pluripotency. However, the mechanism by which these lincRNAs function is unknown. Here, we expand the catalog of human lincRNAs to ≈3,300 by analyzing chromatin-state maps of various human cell types. Inspired by the observation that the well-characterized lincRNA HOTAIR binds the polycomb repressive complex (PRC)2, we tested whether many lincRNAs are physically associated with PRC2. Remarkably, we observe that ≈20% of lincRNAs expressed in various cell types are bound by PRC2, and that additional lincRNAs are bound by other chromatin-modifying complexes. Also, we show that siRNA-mediated depletion of certain lincRNAs associated with PRC2 leads to changes in gene expression, and that the up-regulated genes are enriched for those normally silenced by PRC2. We propose a model in which some lincRNAs guide chromatin-modifying complexes to specific genomic loci to regulate gene expression.





HOTAIR requires PRC2 for function

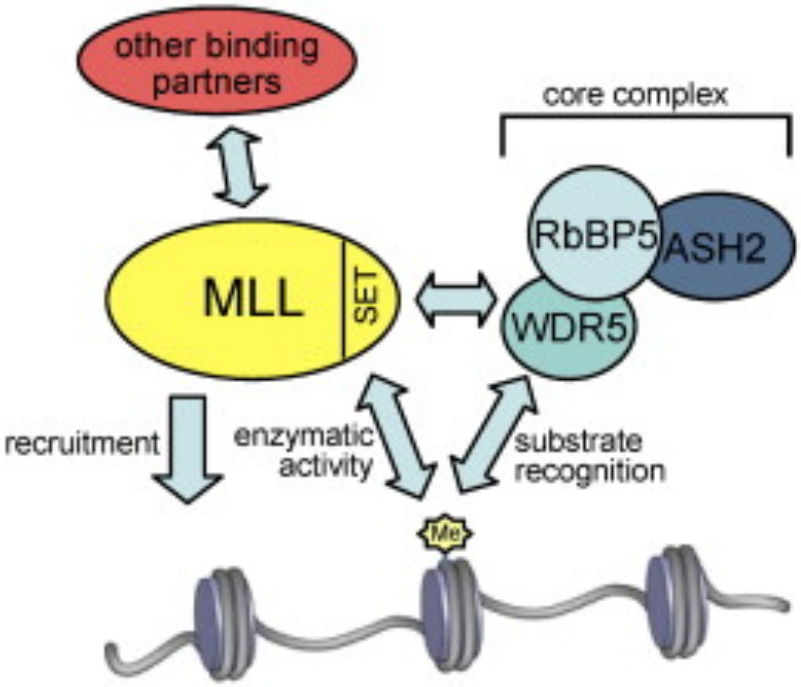




L'overespressione di HOTAIR senza PRC2 non aumenta la crescita tumorale

Writing the H3K4 Methylation Mark

- TrxC (MLL in mammals) *methylates* H3K4 and recruits HAT and remodelling complexes

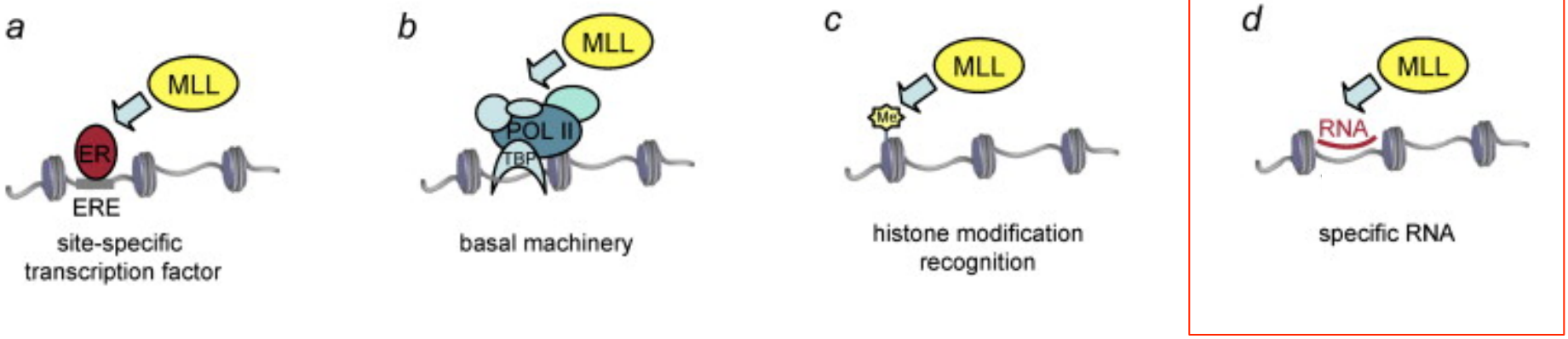


MLL-family HMTs *associate* with the core complex containing **RbBP5**, **WDR5**, and **ASH2**. The core complex cooperates with the catalytic SET domain to *methylate H3K4*, whereas other regions of the MLL protein are involved in association with other protein partners and in recruitment of the MLL complex to the target genes.

WDR5 plays a role in *substrate recognition* and presentation, with preferential, but not exclusive, binding to the H3K4me2 substrate.

Mechanisms of H3K4 methyltransferase recruitment to the target genes

Although precise mechanisms of recruitment remain to be determined, the existing literature suggests that H3K4 methyltransferases are recruited to and/or stabilized on chromatin by a combination of mechanisms involving **association with site-specific transcription factors** (a), **basal machine**ry (b), **histone modification recognition** (c), and **specific RNAs** (d).



HOTTIP appears to regulate genes *in cis*, due to its:

- low copy number
- distance dependence of *HOXA* target gene activation on endogenous *HOTTIP*
- the physical proximity of *HOTTIP* and its target genes as seen in 5C

in cis