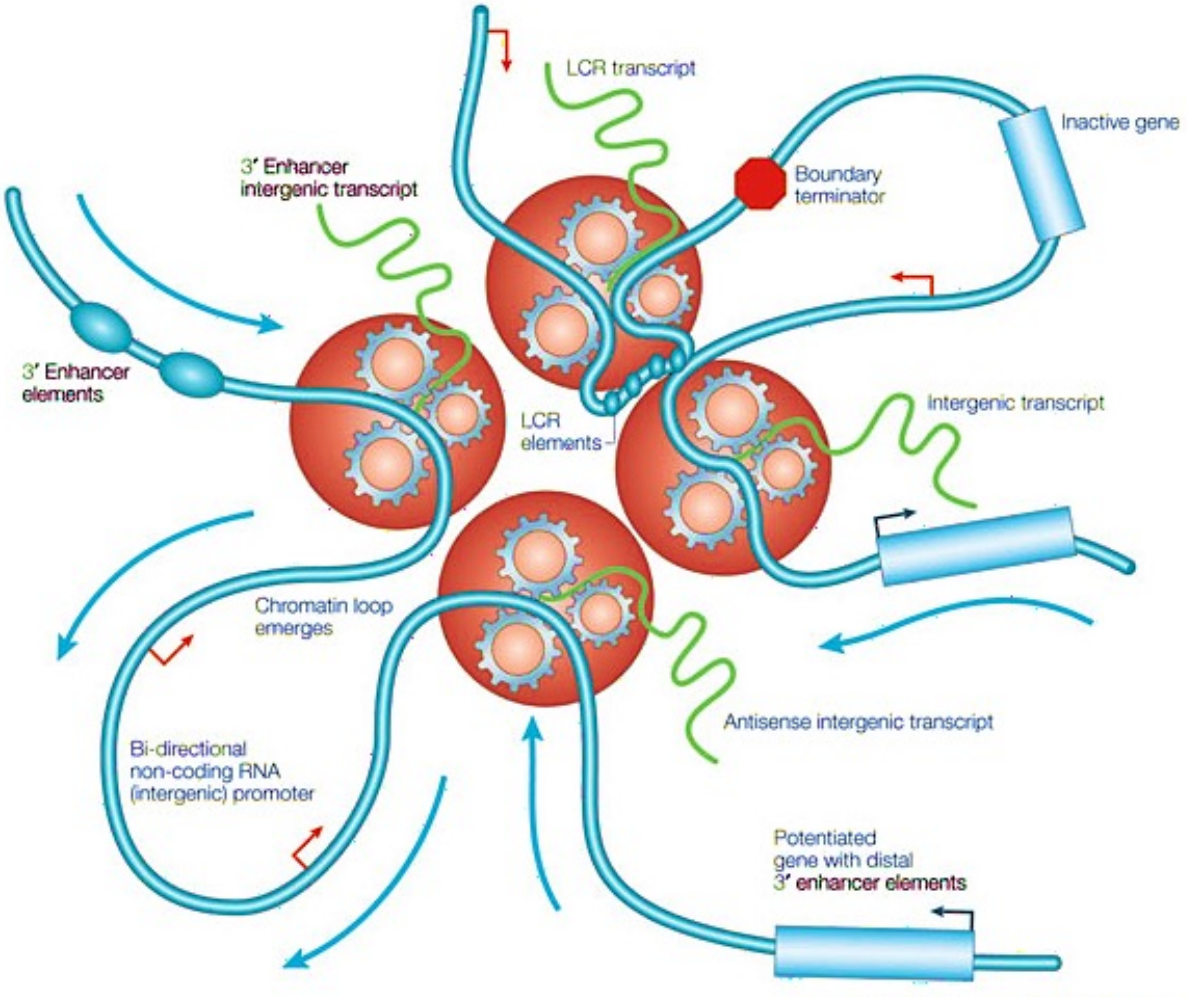
GENE EXPRESSION REGULATION IN EUKARYOTES

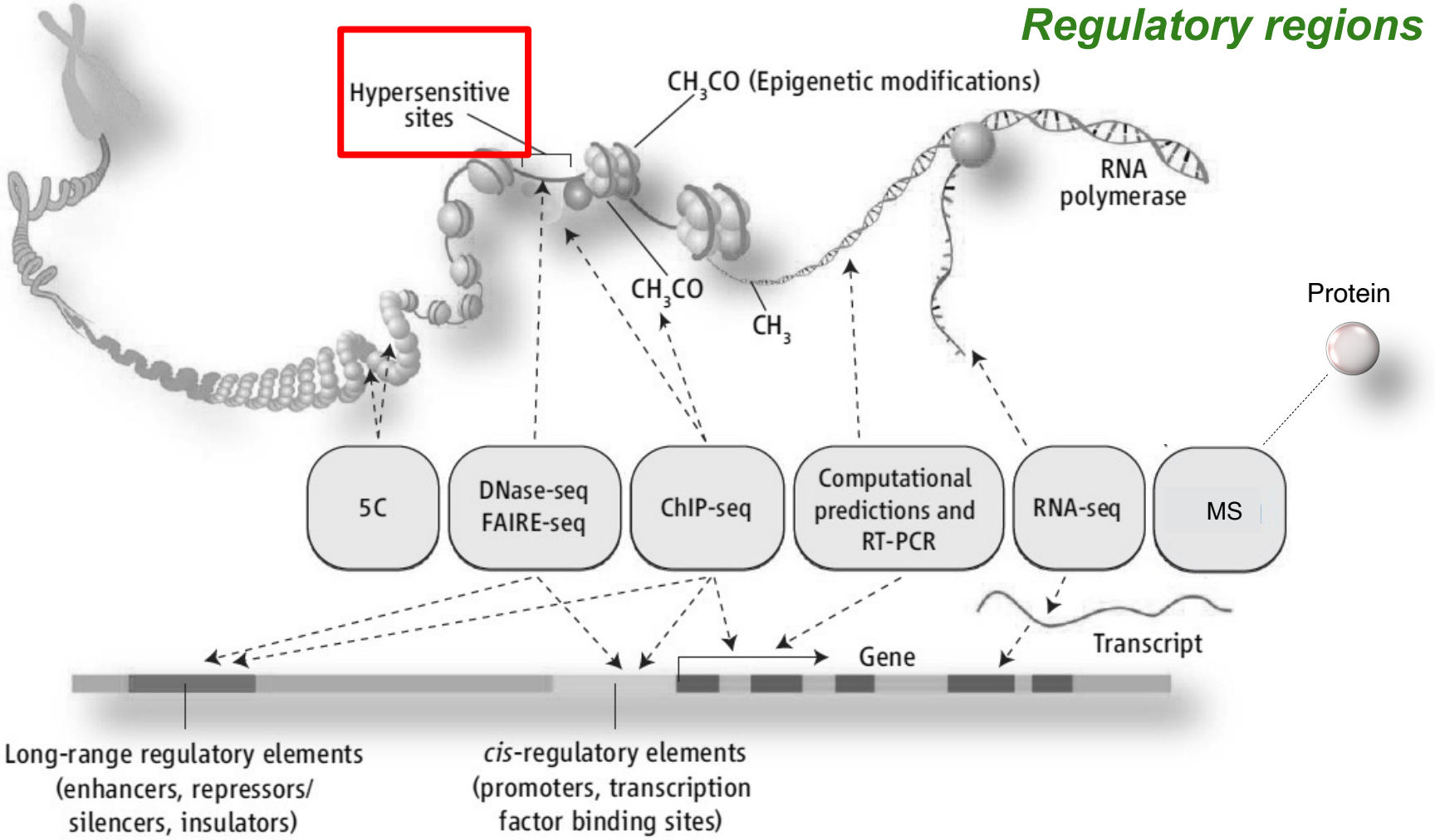


Chakalova *et al*. *Nature Rev Genetics* 6, 669-677.

Chromatin structure and _Transcriptional_ regulation_II Part

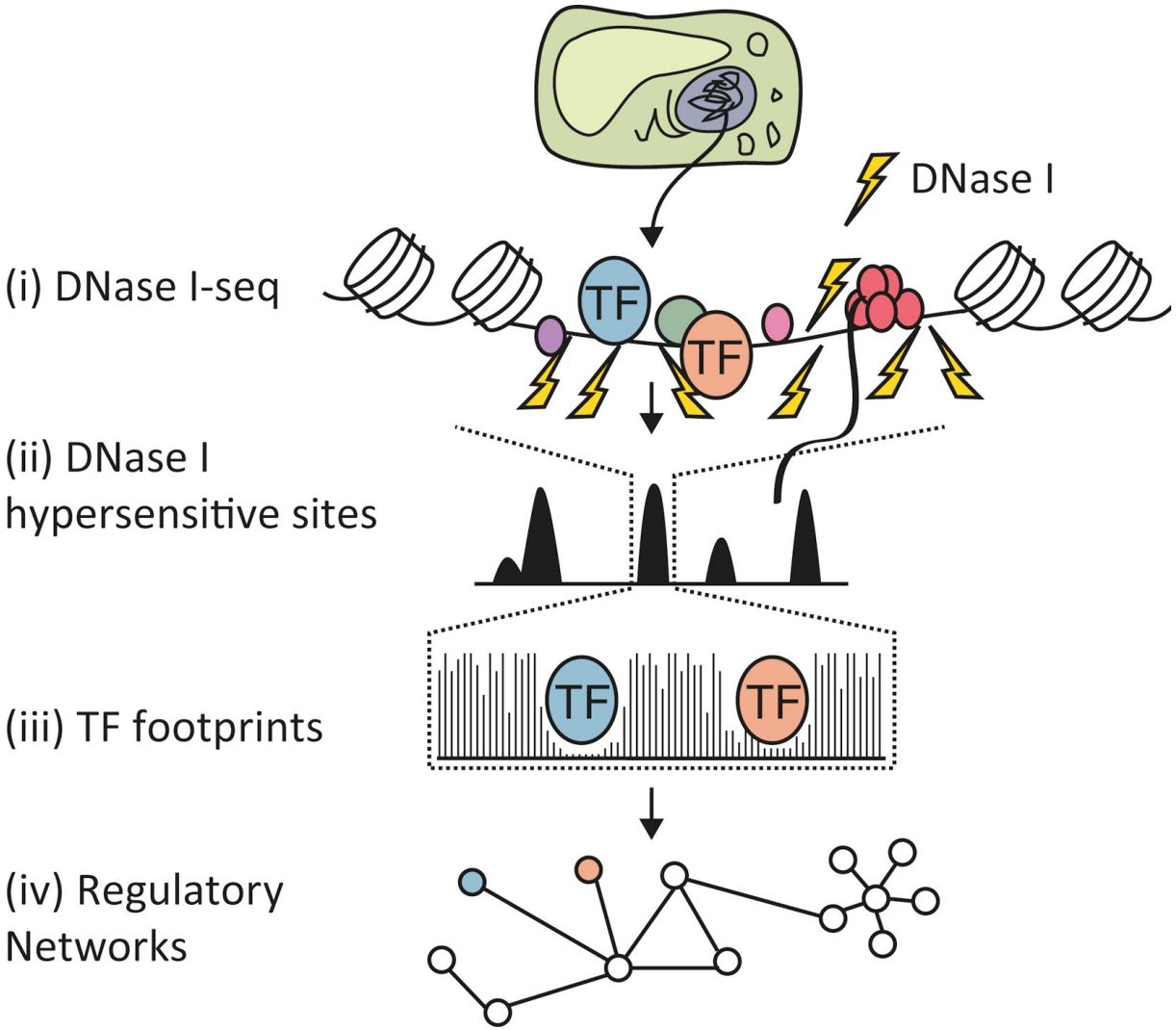
What did we understand from reading the chromatin?

Regulatory regions



modified from *Science 337:1159-60, 2012*

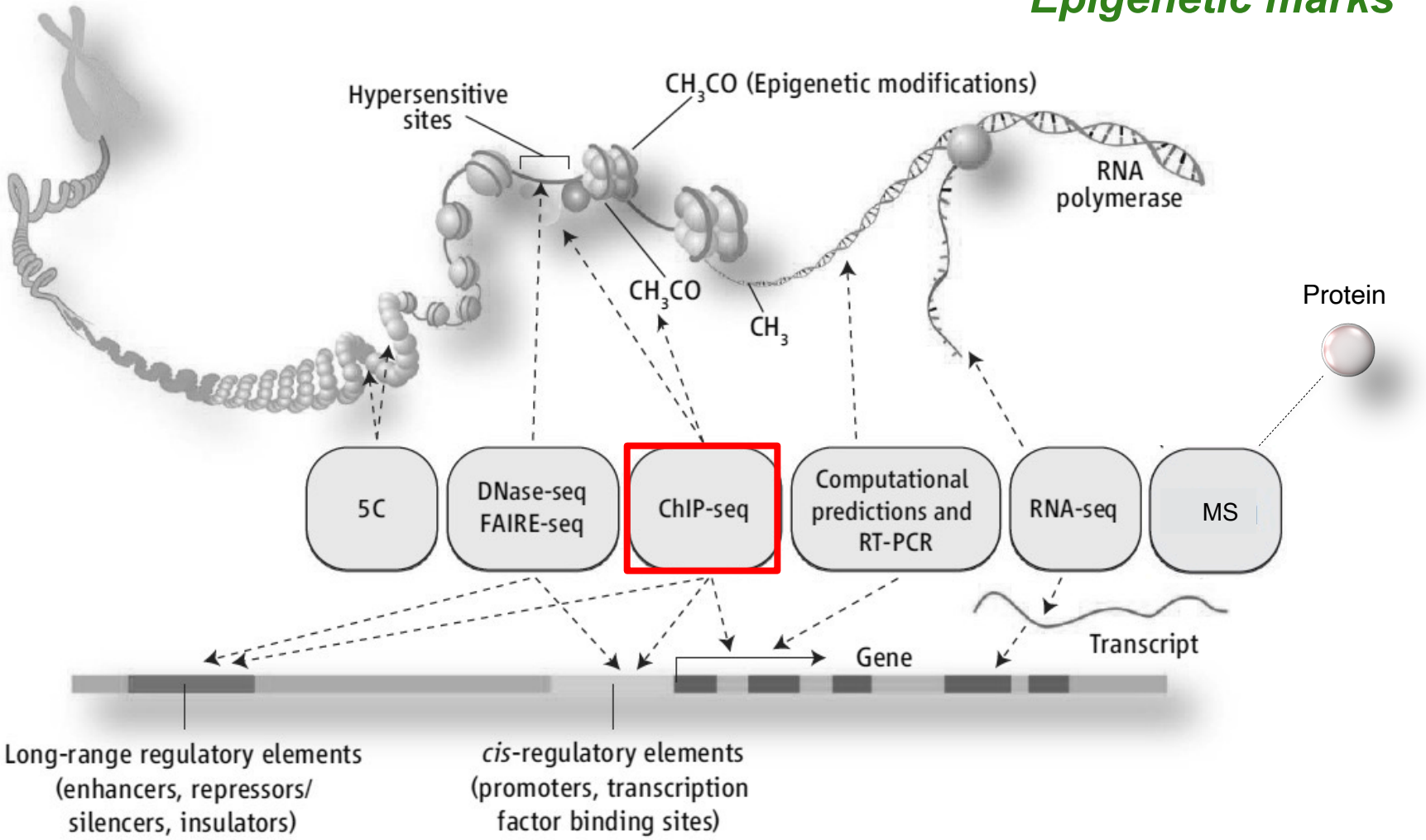
Schematic of DNase I-seq-derived data



(i) Nuclei are harvested from plant tissues and treated with the endonuclease DNase I. (ii) Regulatory regions are hypersensitive to cleavage by DNase I. (iii) Protein-bound regions within DNase I hypersensitive sites are protected from DNase I cleavage leaving detectable "footprints." (iv) Footprint and TF motif information can be integrated to generate TF-to-TF regulatory networks.

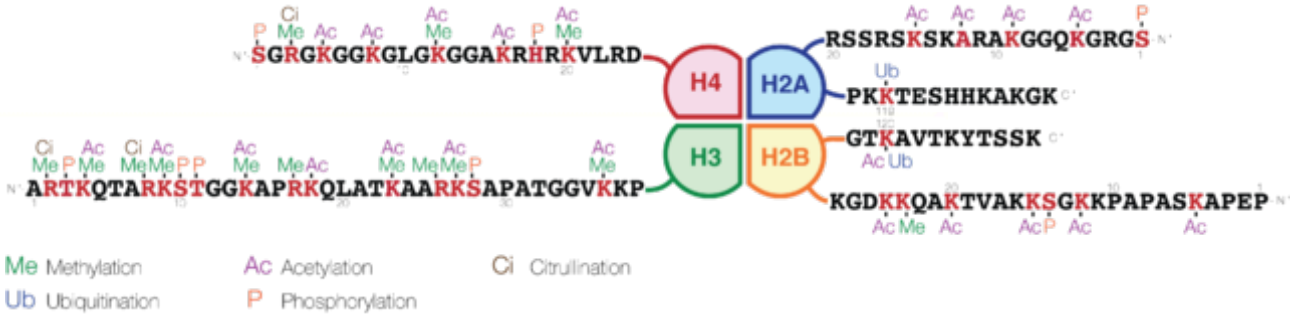
What did we understand from reading the chromatin?

Epigenetic marks



modified from *Science 337:1159-60, 2012*

The readout of the histone post-transcriptional modifications



Type of modification

- Which amino-acid
- Number of modifications (me)

Position in genome

- Promoter: H3K36me, H3K9me are repressive
- Coding region: H3K36me, H3K9me are activating and prevent cryptic initiation of transcription in ORF

Other histone modifications

- combinatorial (occur together)
- H3K4me + H3K9me: transcriptional activation
- H4K20me + H3K9me: heterochromatin formation
- H3K27me + H3K4me: "bivalent" mark in stem cells

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

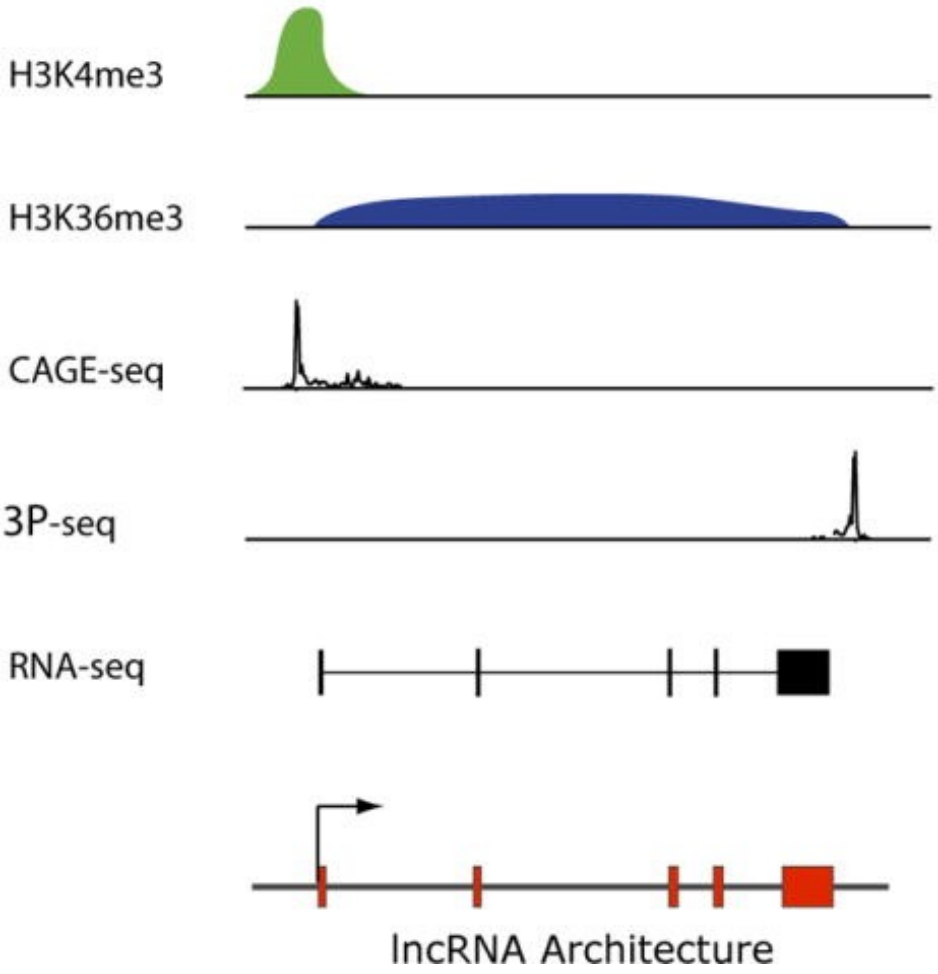
Cycles of modifications

- H2Bub → H2B required for transcriptional elongation

What did we understand from reading the chromatin?

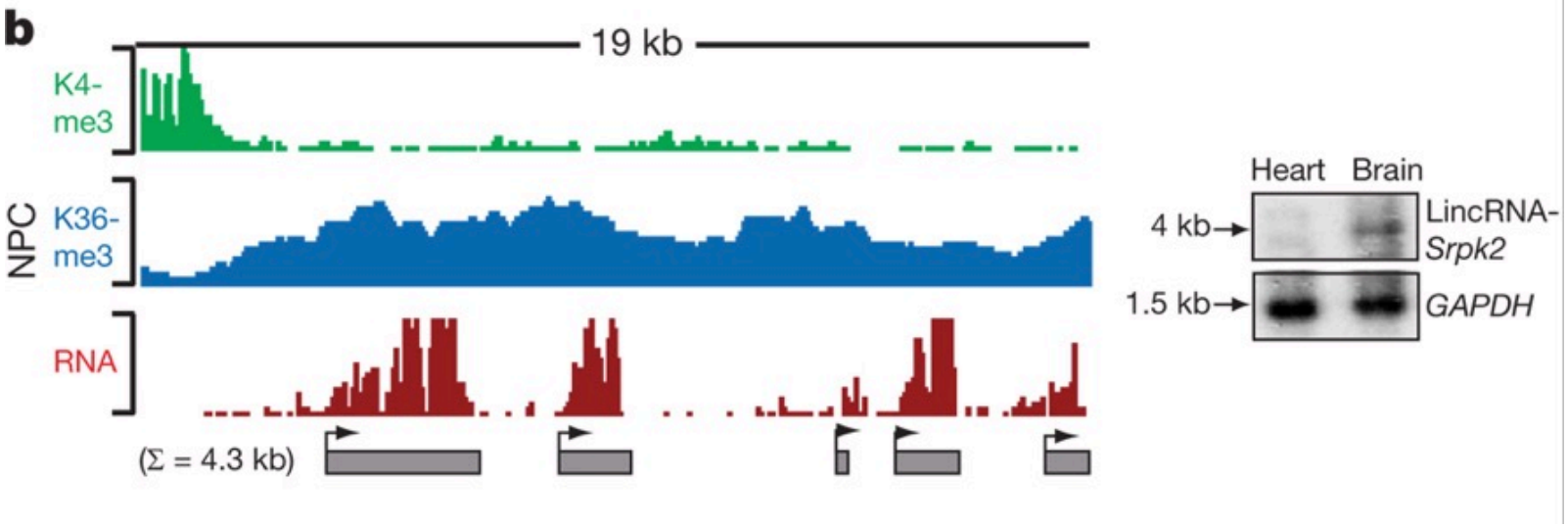
RNA genes

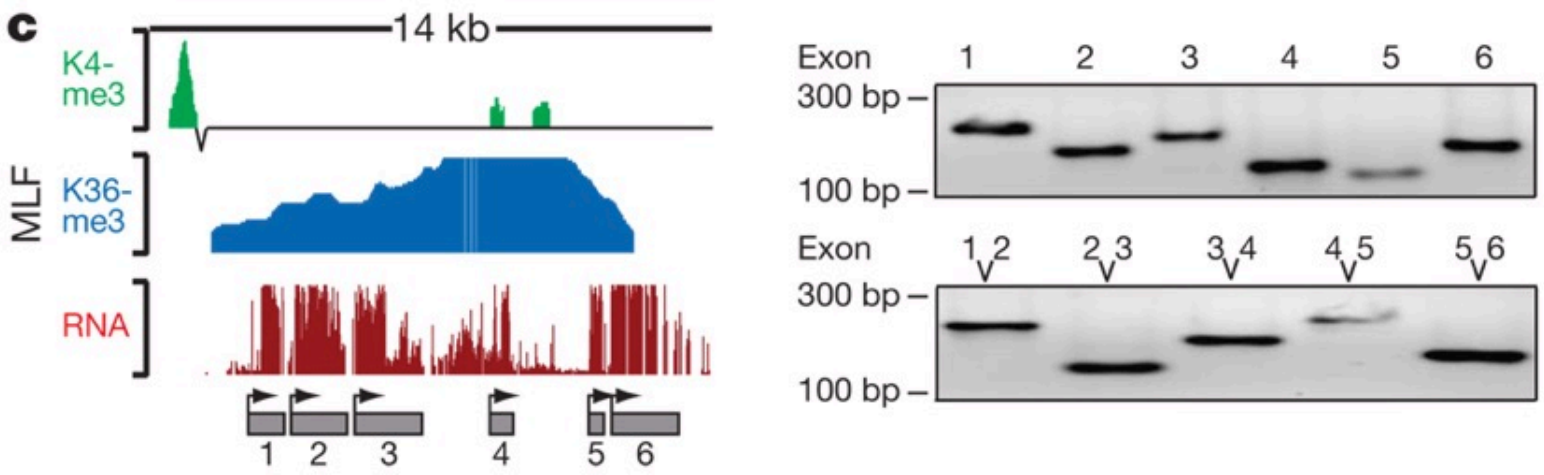
A critical clue for hunting RNA genes came from chromatin



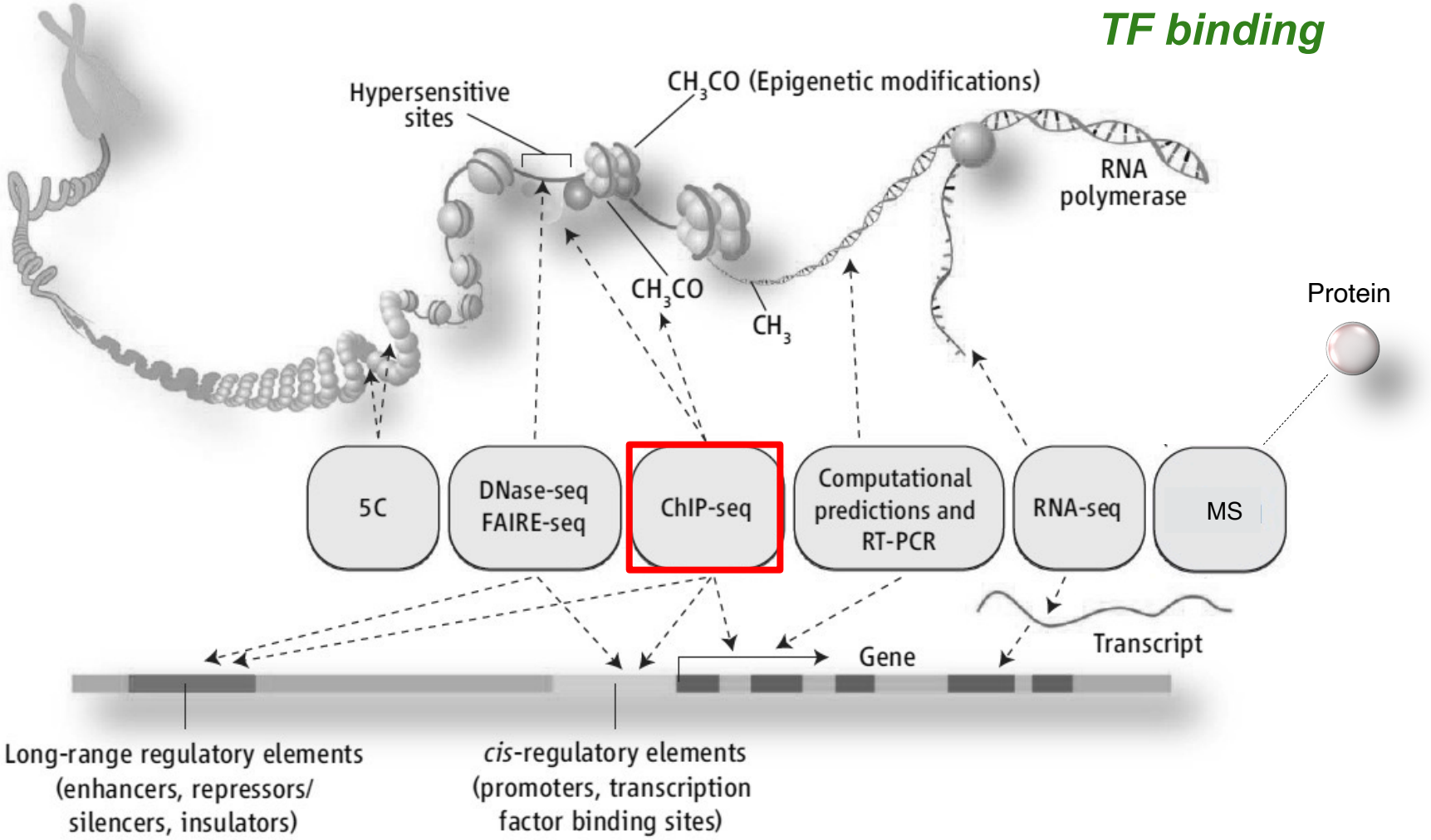
Chromatin marks of transcription initiation (**H3K4me3**) and elongation (**H3K36me3**) define whole transcribed regions of the genome, while sequencing of capped RNA fragments (**CAGE-tag**) or poly-adenylation ends (**3P-seq**) defined the precise beginning and ends of transcripts.

Intergenic K36-K4 domains produce multiexonic RNAs





What did we understand from reading the chromatin?



modified from *Science 337:1159-60, 2012*

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*)

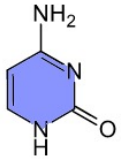
- Regulatory sequences (enhancers, silencers)
- Transcription factors (activators, repressors)

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

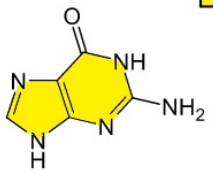
- Small and long non-coding RNAs

Substrates

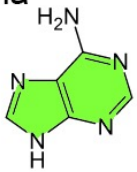
Citosina **C**



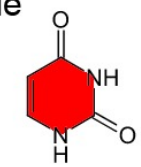
Guanina **G**



Adenina **A**



Uracile **U**



Nitrogenous bases

Enzyme

RNA Polymerase

Gene

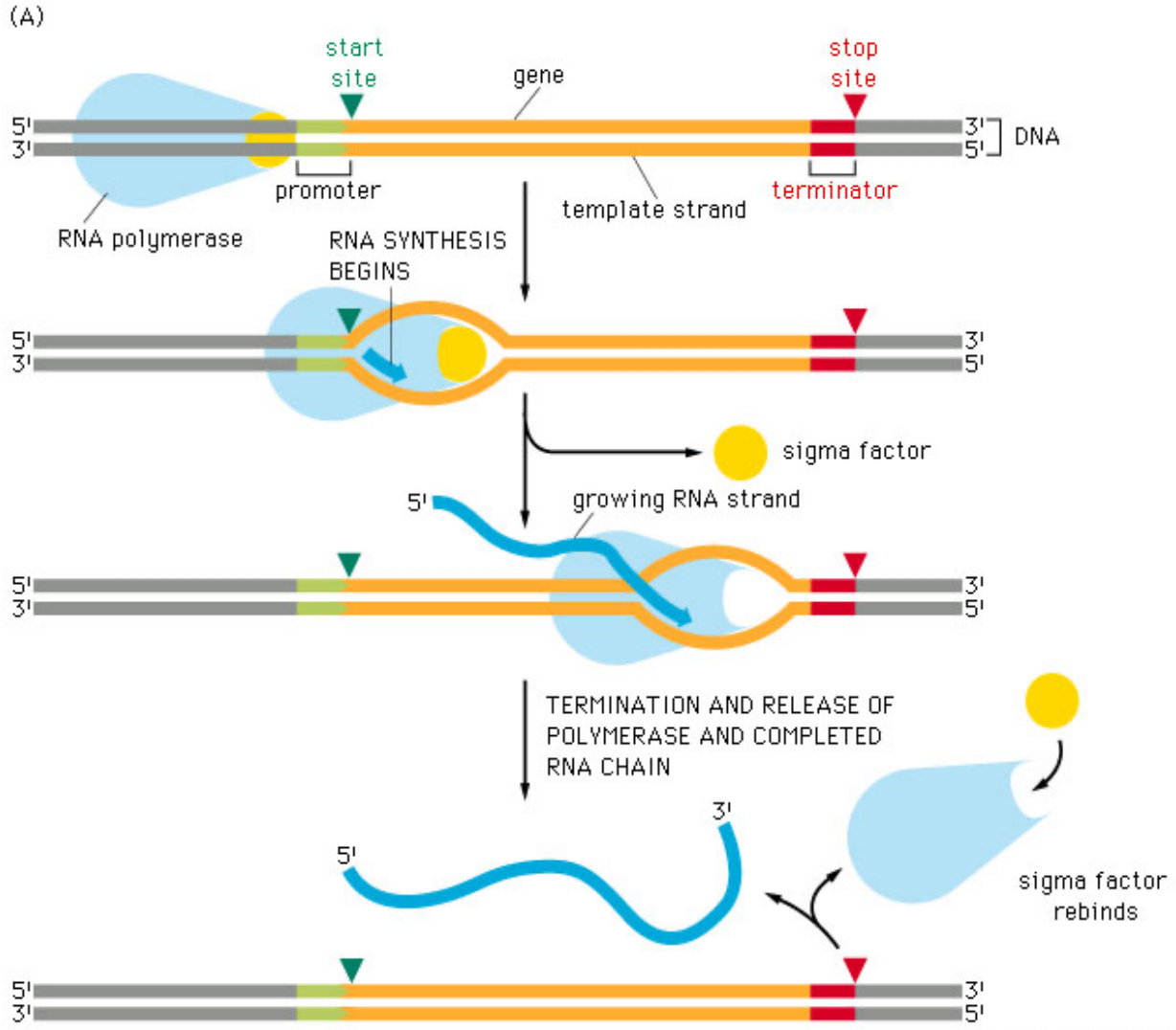
Products

tRNA
rRNA
mRNA

Template

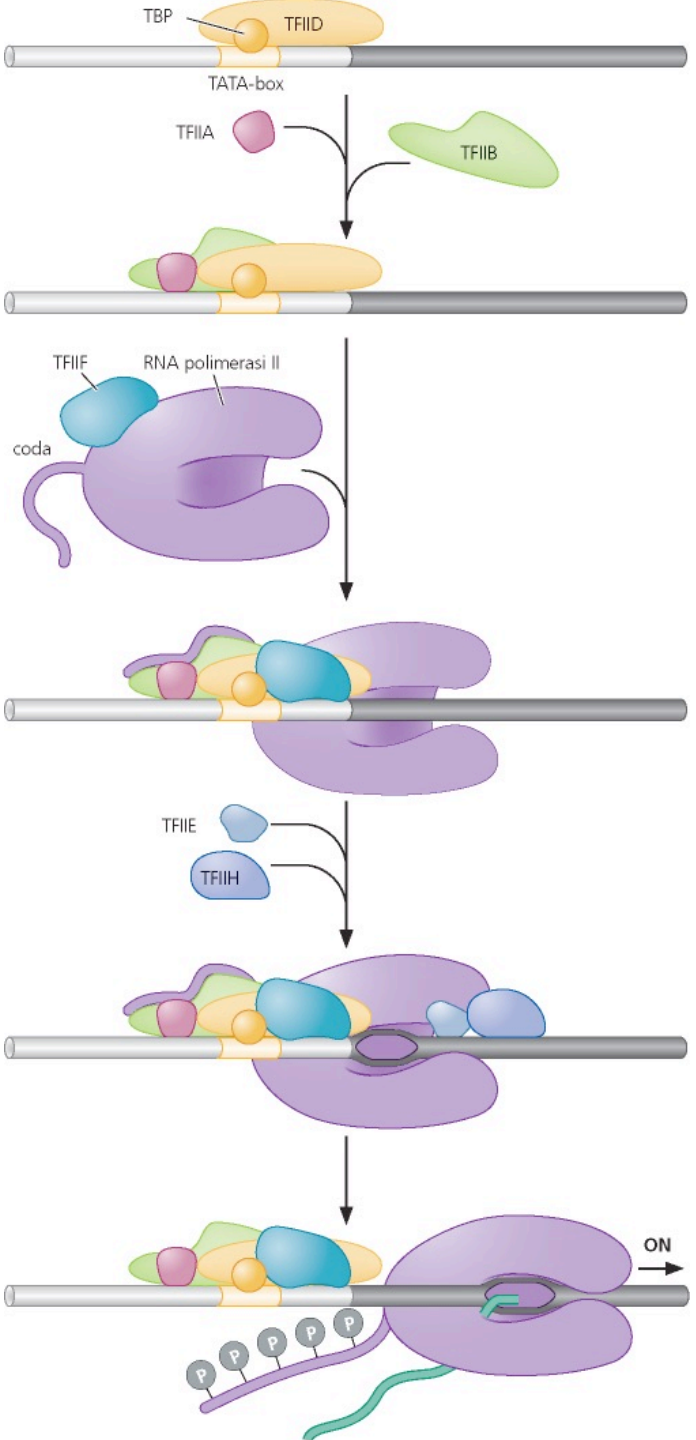
The Transcription UNIT

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 initiation in eukaryotes

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**



General Transcription Factors

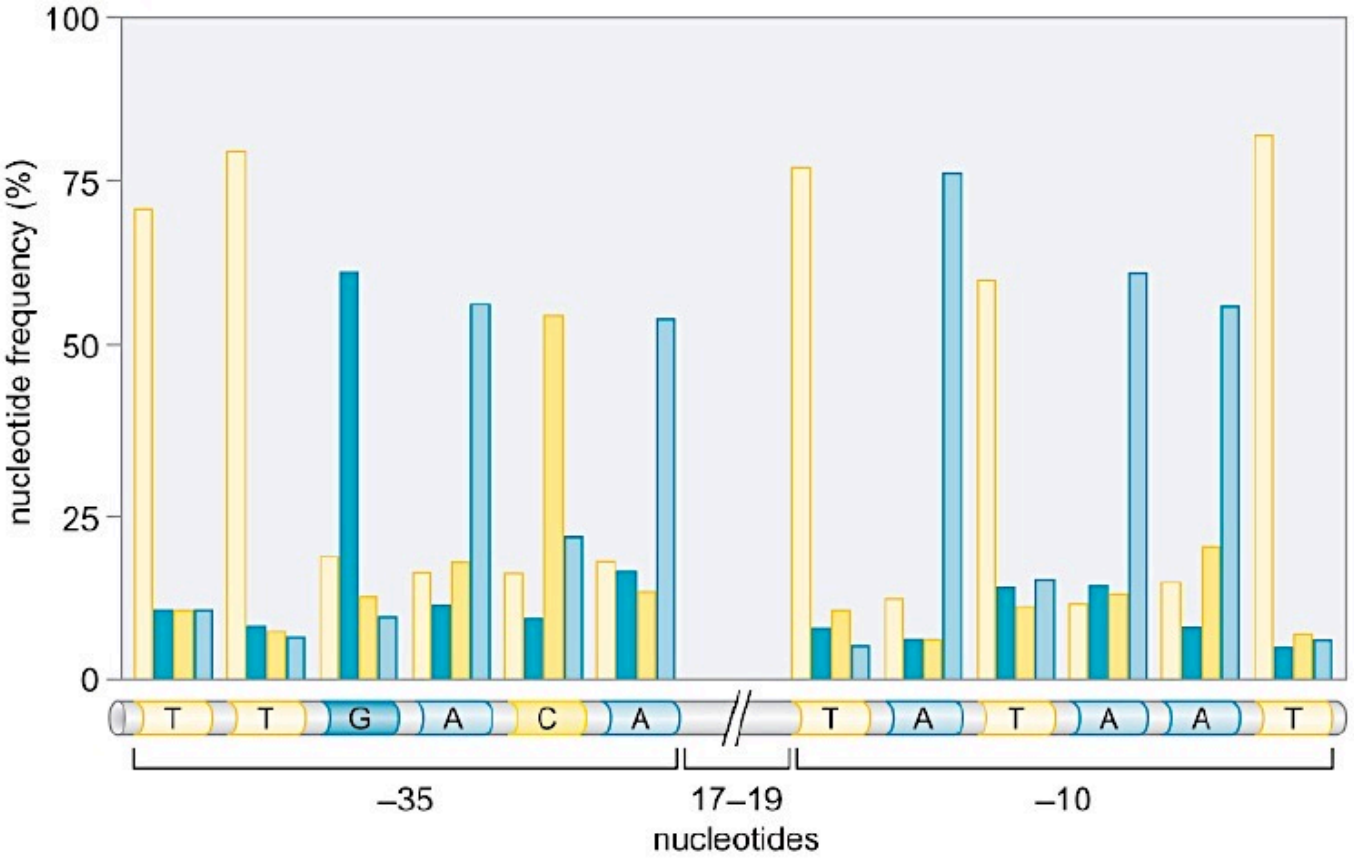
Factor	Gene name		Mass (kDa)		Uniprot accession number		Copies
	Yeast	Human	Yeast	Human	Yeast	Human	
TFIIA[§]: TBP stabilization and counteracts repressive effects of negative co-factors							
Large subunit	*TOA1*	*GTF2A1*	32.2	41.5	P32773	P52655	1
Small subunit	*TOA2*	*GTF2A2*	13.5	12.5	P32774	P52657	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			

Promoter = *Consensus*

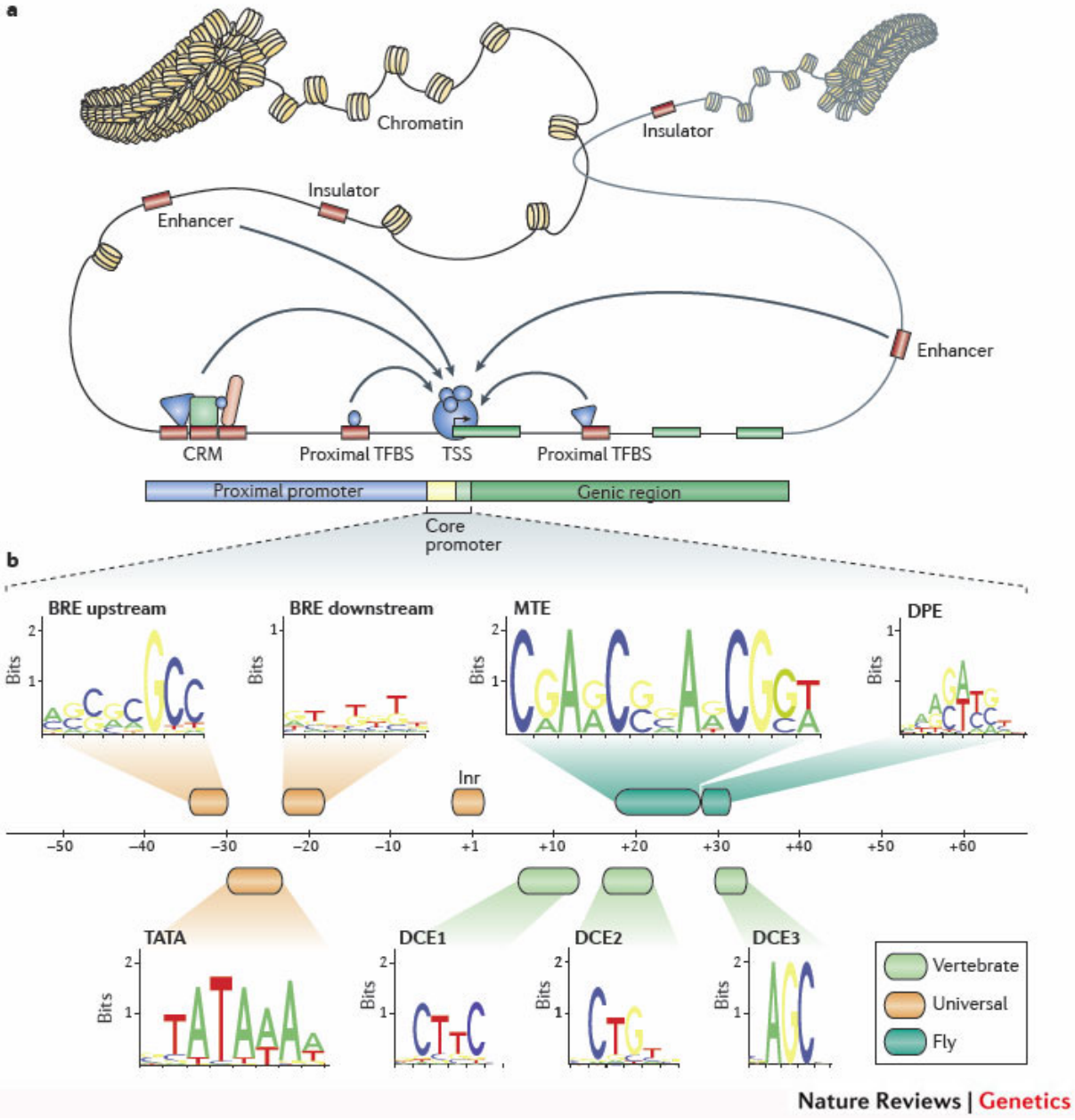
A ***consensus* sequence** is a genetic sequence found in *widely divergent organisms* or *genetic locations* with ***minor*** variations and (probably) *similar* functions.

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}



Regulatory signals and promoter elements in Metazoa



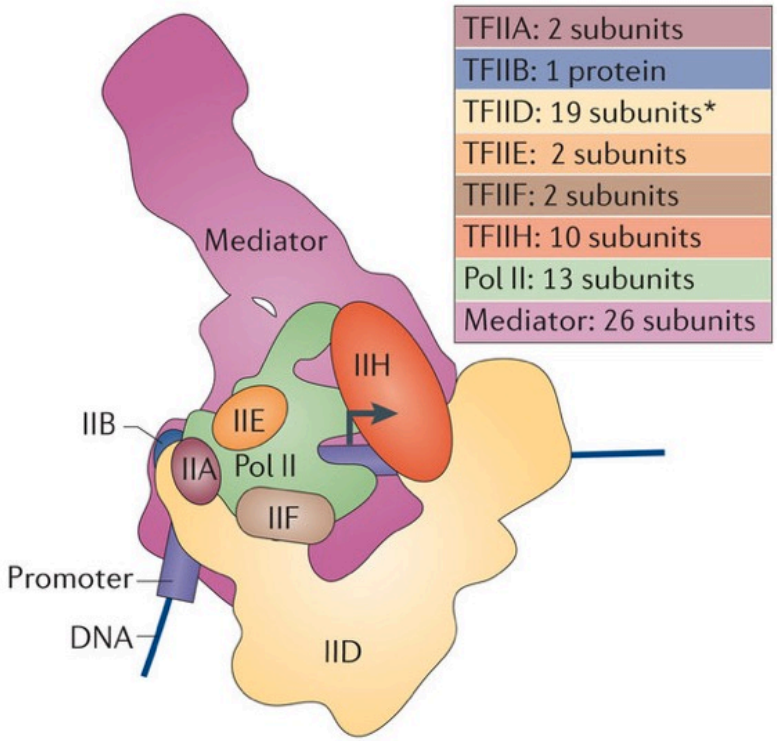
- The term '**core promoter**' is often used to focus on the DNA region in the immediate vicinity of the TSS, which is assumed to dock the **pre-initiation complex (PIC)**
- It consists of several interchangeable sequence elements around the TSS, which bind core components of the PIC.
- The region around the TSS has several over-represented sequence patterns; the **TATA box** and **initiator** (**Inr**) are the most studied and most prevalent.

BRE, B recognition elements
DCE, downstream core element;
DRE, DNA recognition element;
MTE, motif ten element.

Mediator has key roles in the assembly of the PIC

For gene-specific transcription, *Pol II must be recruited* to specific sites on the genome. This is generally controlled by *sequence-specific*, DNA-binding **TFs**.

Although TFs do not directly bind to Pol II, one mechanism by which they can promote Pol II recruitment is by *binding* to the Mediator complex. **Mediator enables Pol II recruitment via interaction with the CTD of the Pol II.** The large size of Mediator is likely to promote stable PIC formation by allowing the complex to directly interact with multiple PIC factors. Moreover, Mediator helps to regulate the recruitment and/or the activity of the PIC components.



Nature Reviews | Molecular Cell Biology

How

to study the binding of a trascription factor (TF) on a specific promoter *in vivo*

Target-directed experiment:

1 - Predict possible TF binding sites

2 - Perform ChIP

3 - Validate the assumptions by functional analyses

1 - Predict possible TF binding sites (TFBS)

To be able to predict ***potentially functional*** TFBS is an important first step in promoter analysis.

TFBS prediction programs (Jaspar, MatInspector…) can infer the ***binding potential***, although *not the functionality of a site*.

It is important to note that TFBS only carry the potential to bind their corresponding protein. However, they can occur everywhere in the genome and are by no means restricted to regulatory regions. ***It is the context that differentiates a functional binding site*** *affecting gene regulation* ***from a mere physical binding site***

Functionality can ultimately be proven only by a wet-lab experiment with defined settings, particularly since potential binding sites in a promoter can be functional in certain cells, tissues or developmental stages and non-functional under different conditions.

2 - *Perform* ChIP to validate the assumptions

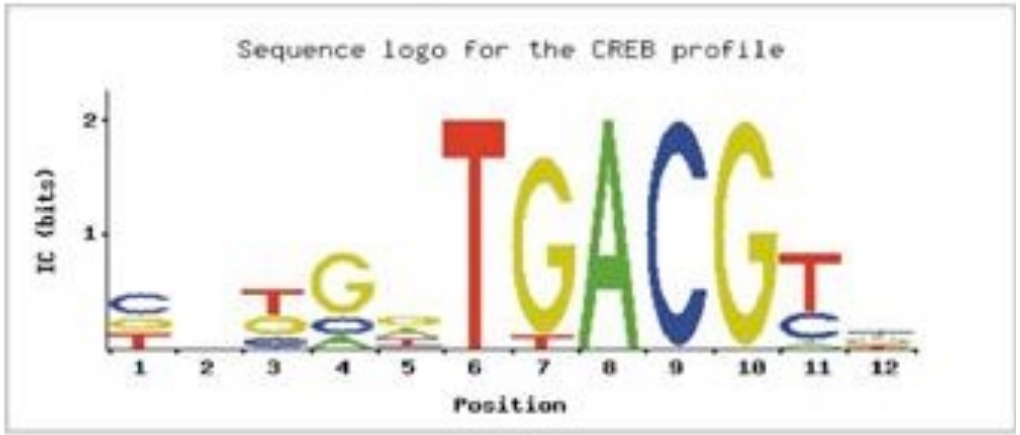
3 - *Validate* the assumptions by functional analyses

Prediction of TF binding sites precedes ChIP

Transcriptional regulation requires the *binding* of transcription factors (TFs) to short sequence-specific DNA motifs, usually located at the gene regulatory regions.

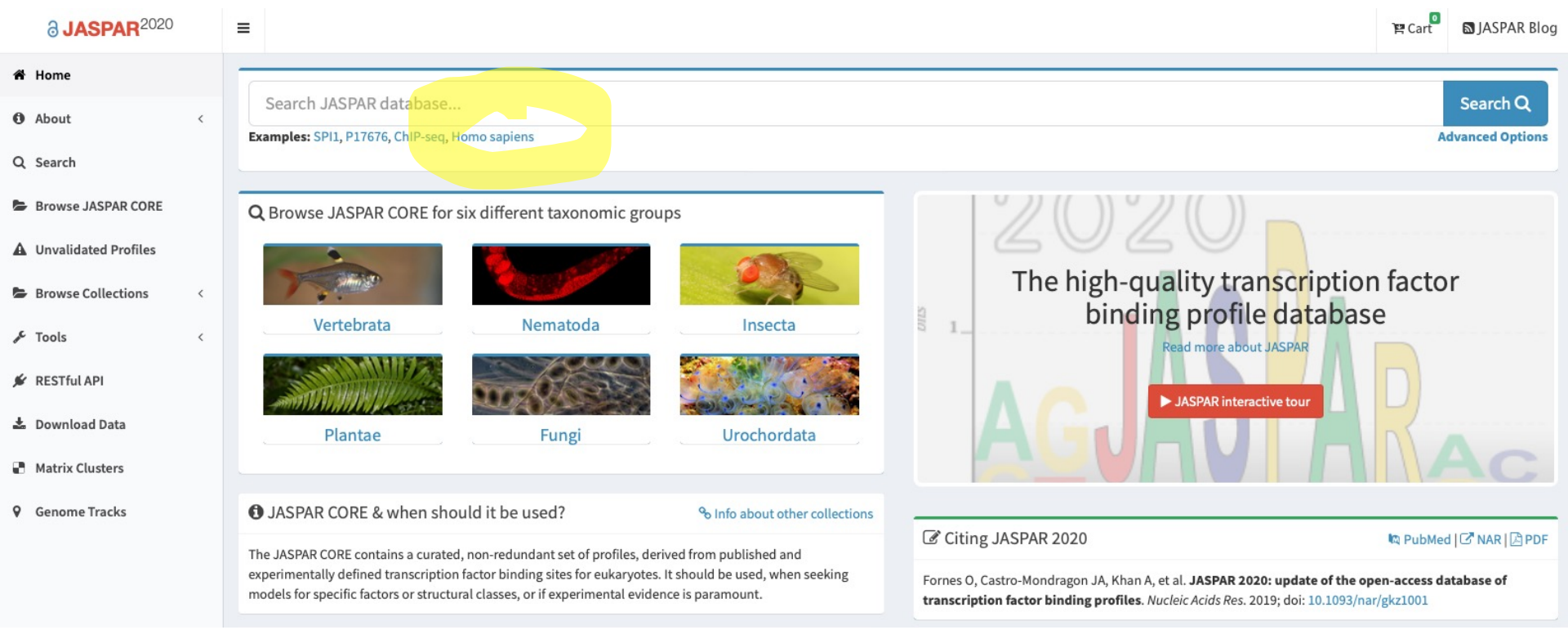
```
CTTGGTGACGTG
GTGAGTGACGTC
CGGGTTGACGCA
CCTACTTACGTA
TATGGTGACGTC
TCGGATGACGAT
TAGGATGACGTC
CCTGGTGACGCC
CGCGGTGACGTA
GCCGTTGACGCC
CGCGATGACGCA
CCTGTTGACGTG
TTGCATGACGTC
GTTGGTGACGTG
GAGGATGACGTT
GGTCGTGACGTA
------------
CTGGGTGACGTC  (Consensus)
------------
CNKGGTGACGTM  (Degenerate Consensus)
```

```
A  [ 0  3  0  2  5  0  0 16  0  0  1  5 ]
C  [ 7  5  3  3  1  0  0  0 16  0  5  6 ]
G  [ 5  4  6 11  7  0 15  0  0 16  0  3 ]
T  [ 4  4  7  0  3 16  1  0  0  0 10  2 ]
```



http://jaspar.genereg.net

Prediction of TF binding sites precedes ChIP



JASPAR2020

Cart 0 | JASPAR Blog

- Home
- About
- Search
- Browse JASPAR CORE
- Unvalidated Profiles
- Browse Collections
- Tools
- RESTful API
- Download Data
- Matrix Clusters
- Genome Tracks

Search profile(s)

Home > search

Homo sapiens

Search

Examples: SPI1, P17676, ChIP-seq, Homo sapiens

Advanced Options

You can search by TF name or ID, species, taxon, UniProt ID or any other keyword.

810 profile(s) found

Display 10 profiles

Filter:

ID	Name	Species	Class	Family	Logo
MA0007.2	AR	Homo sapiens	Nuclear receptors with C4 zinc fingers	Steroid hormone receptors (NR3)	
MA0259.1	ARNT::HIF1A	Mus musculus Rattus rattus Homo sapiens Oryctolagus cuniculus	Basic helix-loop-helix factors (bHLH)::Basic helix-loop-helix factors (bHLH)	PAS domain factors::PAS domain factors	
MA0605.2	ATF3	Homo sapiens	Basic leucine zipper factors (bZIP)	Jun-related factors	
MA0634.1	ALX3	Homo sapiens	Homeo domain factors	Paired-related HD factors	

Analyze selected profiles

Please select matrix profiles on the left side to add to your cart or perform the following analysis.

Add to cart

You have **0** profile(s) in your cart. You can add profiles to the cart to download or perform analysis.

Add to cart | View cart

Scan

Basic sequence analysis

The feature allows the user to submit a DNA sequence and analyze by using a subset of profiles. Sensitivity and specificity of the results will be affected by the relative score threshold, by default 80%. The input sequence should be in fasta-formatted.

JASPAR2020

- Home
- About
- Search
- Browse JASPAR CORE
- Unvalidated Profiles
- Browse Collections
- Tools
- RESTful API
- Download Data
- Matrix Clusters

Search profile(s)

myod

Examples: SPI1, P17676, ChIP-seq, Homo sapiens

10 profile(s) found

Display 10 profiles

Filter:

ID	Name	Species	Class	Family	Logo
MA0499.1	Myod1	Mus musculus	Basic helix-loop-helix factors (bHLH)	MyoD / ASC-related factc	
MA0499.2	MYOD1	Homo sapiens	Basic helix-loop-helix factors (bHLH)	MyoD / ASC-related factc	

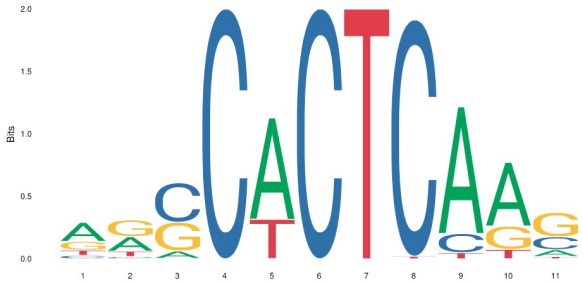
http://jaspar.genereg.net/matrix/MA0503.1/

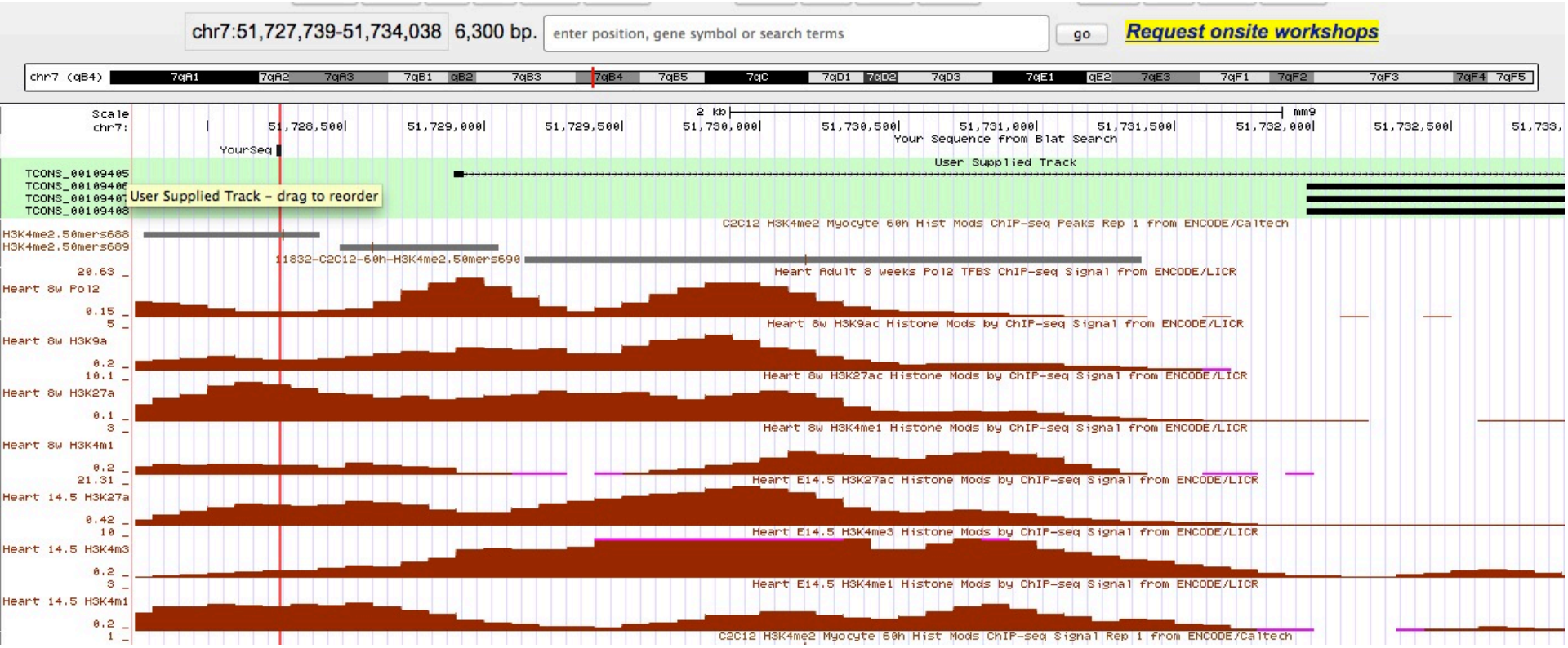
Matrix ID	Name	Score	Relative score	Sequence ID	Start	End
MA0503.1	Nkx2-5(var.2)	16.4337	1.00000000869	prom	1106	1116
MA0461.1	Atoh1	13.7404	0.995554460737	prom	1293	1300
MA0499.1	Myod1	15.9117	0.994762408413	prom	1626	1638
MA0039.2	Klf4	15.1527	0.993950306118	prom	867	876
MA0493.1	Klf1	16.6605	0.993259208783	prom	866	876
MA0521.1	Tcf12	15.0135	0.990487878933	prom	1625	1635
MA0500.1	Myog	14.5801	0.989362224303	prom	1625	1635

http://jaspar.genereg.net/matrix/MA0503.1/

Relative profile score threshold 100 %

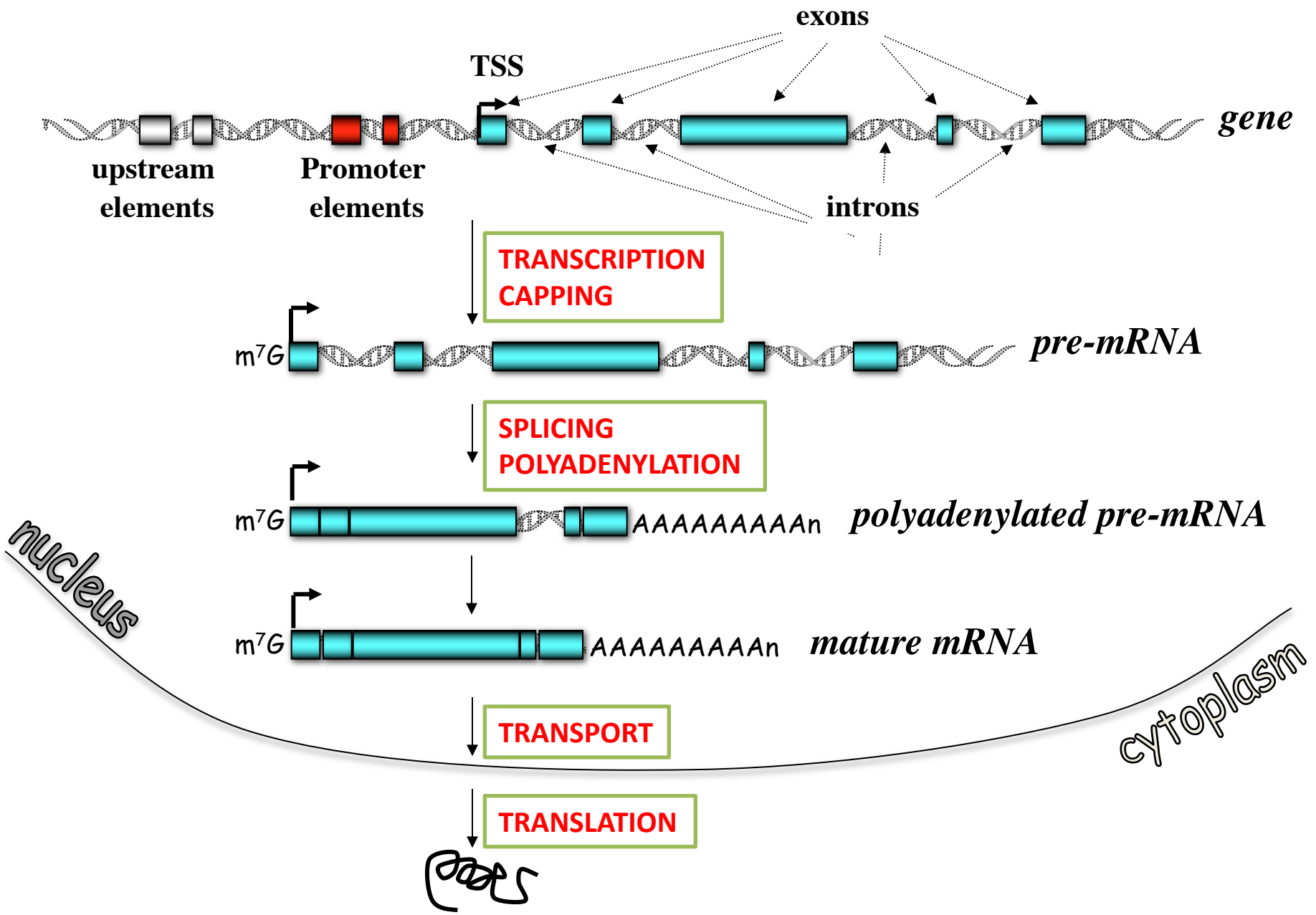
Nkx2-5(var.2)



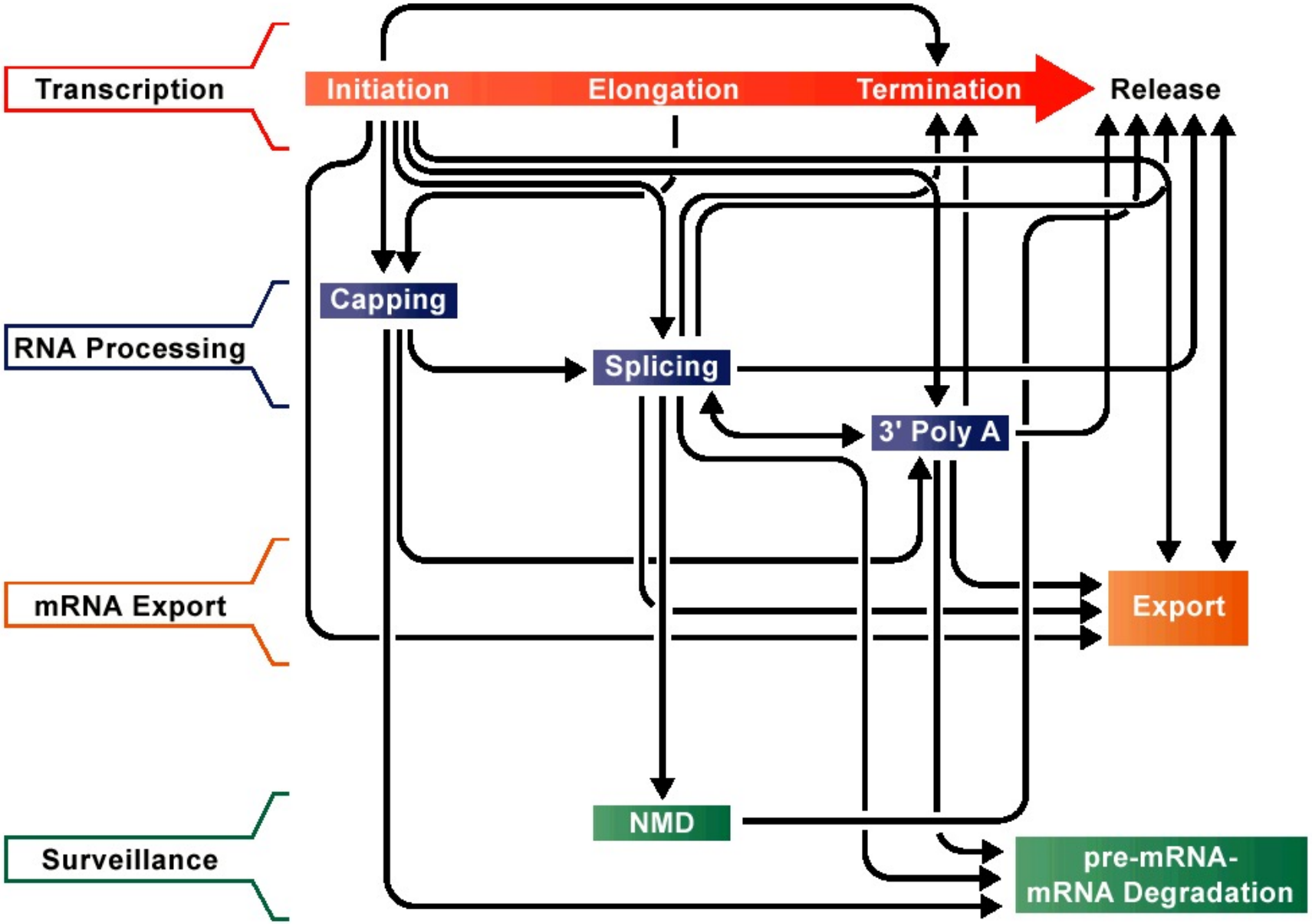


…but then you have to validate the binding peaks! **How?**

Eukaryotic gene expression



Network of *coupled* interactions in gene expression



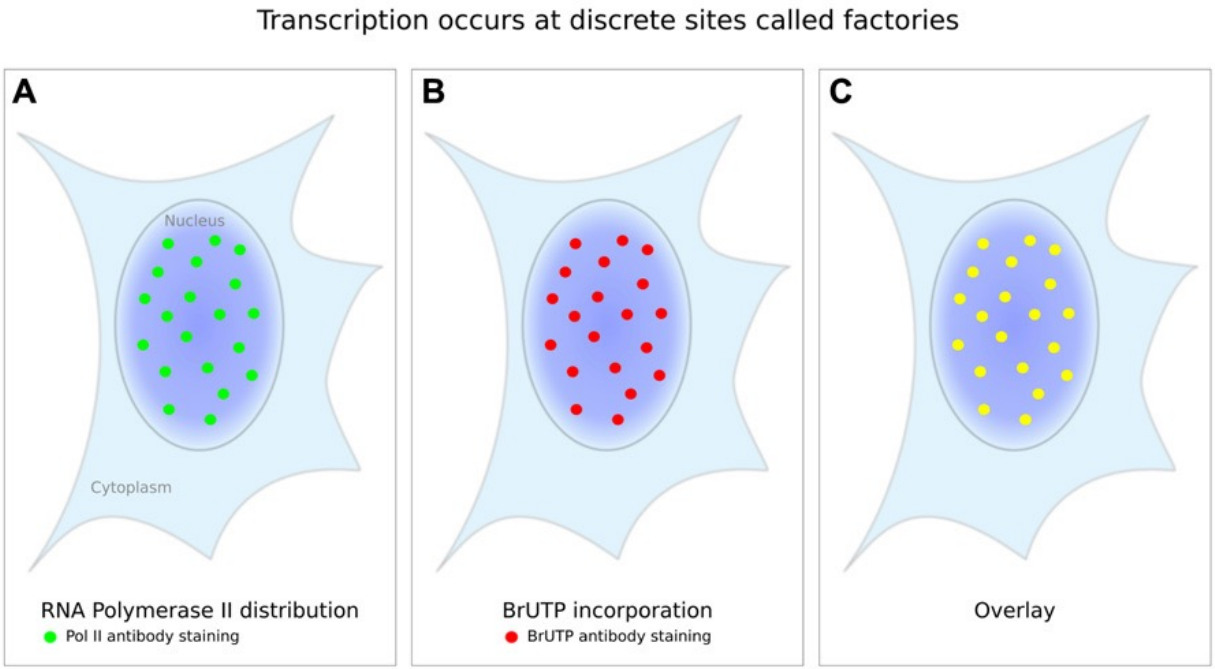
The “mRNA Factory” model

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

2) *spatial* 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).

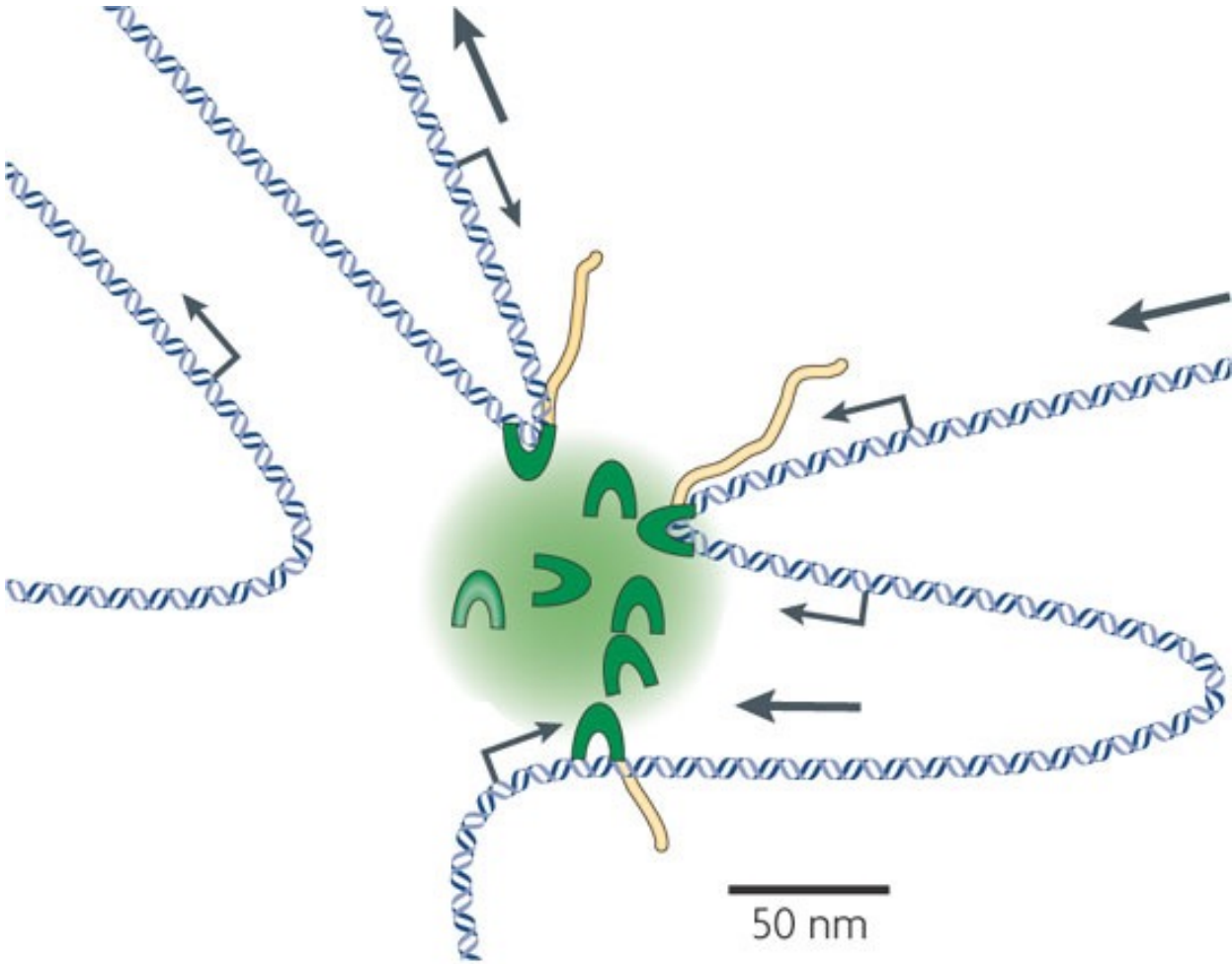
These discrete sites of active transcription are referred to as “***transcription factories***”.



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.

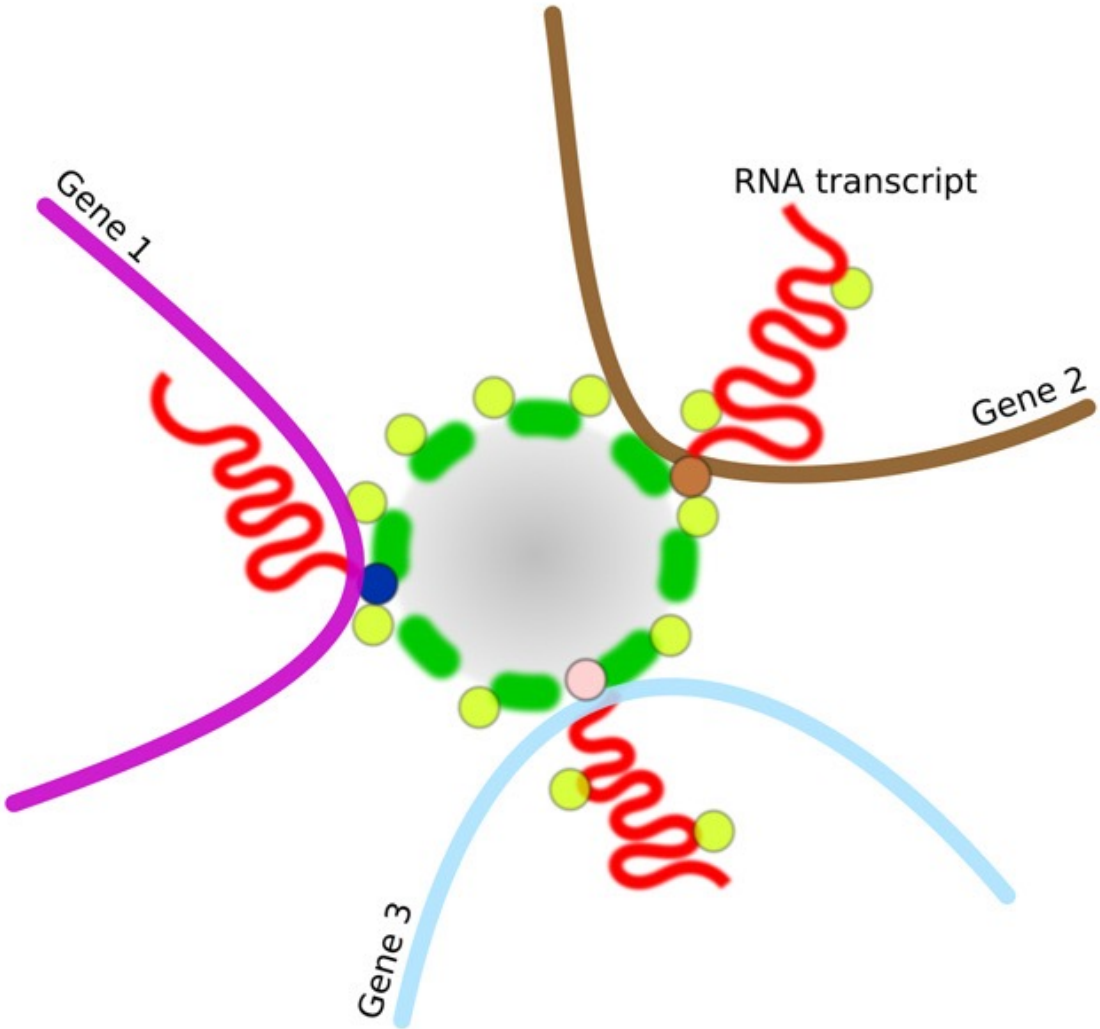
Genes from the *same* or from *different* chromosomes may associate with polymerases in the same factory. Small arrows indicate the direction of transcription at the transcription start site.

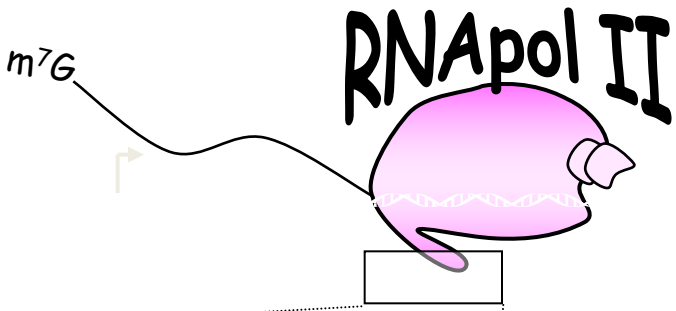


Nature Reviews | Genetics

Structure of a Transcription Factory

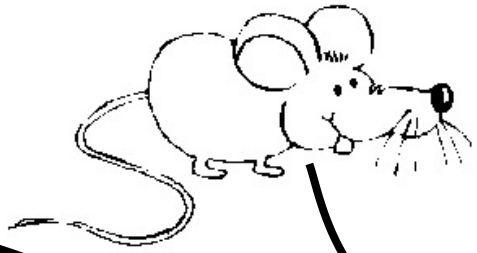
Each factory contains RNA polymerase II molecules which are located on the surface of a *protein-rich core* (87 nm in diameter, as determined by EFTEM in HeLa cells). These proteins include many factors involved in ***transcription*** such as ***co-activators***, ***chromatin remodelers***, ***histone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





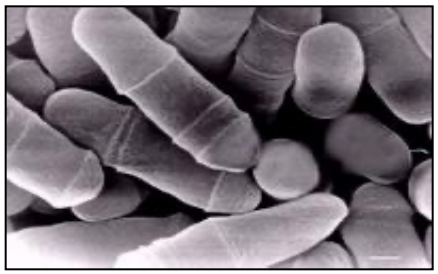
Carbossi Terminal Domain

Mouse



	1	2	3	4	5	6	7
	Glu	Gly	Ala	Met	Ser	Pro	Ser
1.	Tyr	Ser	Pro	Thr	Ser	Pro	Ala
2.	Tyr	Glu	Pro	Arg	Ser	Pro	Gly Gly
3.	Tyr	Thr	Pro	Gln	Ser	Pro	Ser
4.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
5.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
6.	Tyr	Ser	Pro	Thr	Ser	Pro	Asn
7.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
8.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
9.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
10.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
11.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
12.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
13.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
14.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
15.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
16.	Tyr	Ser	Pro	Thr	Ser	Pro	Ala
17.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
18.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
19.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
20.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
21.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
22.	Tyr	Ser	Pro	Thr	Ser	Pro	Asn
23.	Tyr	Ser	Pro	Thr	Ser	Pro	Asn
24.	Tyr	Thr	Pro	Thr	Ser	Pro	Ser
25.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
26.	Tyr	Ser	Pro	Thr	Ser	Pro	Asn
27.	Tyr	Ser	Pro	Thr	Ser	Pro	Asn
28.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
29.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
30.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
31.	Tyr	Ser	Pro	Ser	Ser	Pro	Arg
32.	Tyr	Thr	Pro	Gln	Ser	Pro	Thr
33.	Tyr	Thr	Pro	Ser	Ser	Pro	Ser
34.	Tyr	Ser	Pro	Ser	Ser	Pro	Ser
35.	Tyr	Ser	Pro	Thr	Ser	Pro	Lys
36.	Tyr	Thr	Pro	Thr	Ser	Pro	Ser
37.	Tyr	Ser	Pro	Ser	Ser	Pro	Glu
38.	Tyr	Thr	Pro	Ala	Ser	Pro	Lys
39.	Tyr	Ser	Pro	Thr	Ser	Pro	Lys
40.	Tyr	Ser	Pro	Thr	Ser	Pro	Lys
41.	Tyr	Ser	Pro	Thr	Ser	Pro	Thr
42.	Tyr	Ser	Pro	Thr	Thr	Pro	Lys
43.	Tyr	Ser	Pro	Thr	Ser	Pro	Thr
44.	Tyr	Ser	Pro	Thr	Ser	Pro	Val
45.	Tyr	Thr	Pro	Thr	Ser	Pro	Lys
46.	Tyr	Ser	Pro	Thr	Ser	Pro	Thr
47.	Tyr	Ser	Pro	Thr	Ser	Pro	Lys
48.	Tyr	Ser	Pro	Thr	Ser	Pro	Thr
49.	Tyr	Ser	Pro	Thr	Ser	Pro	Lys Gly Ser Thr
50.	Tyr	Ser	Pro	Thr	Ser	Pro	Gly
51.	Tyr	Ser	Pro	Thr	Ser	Pro	Thr
52.	Tyr	Ser	Leu	Thr	Ser	Pro	Ala
53.	Ile	Ser	Pro	Asp	Asp	Ser	Asp Glu Glu Asn

Yeast



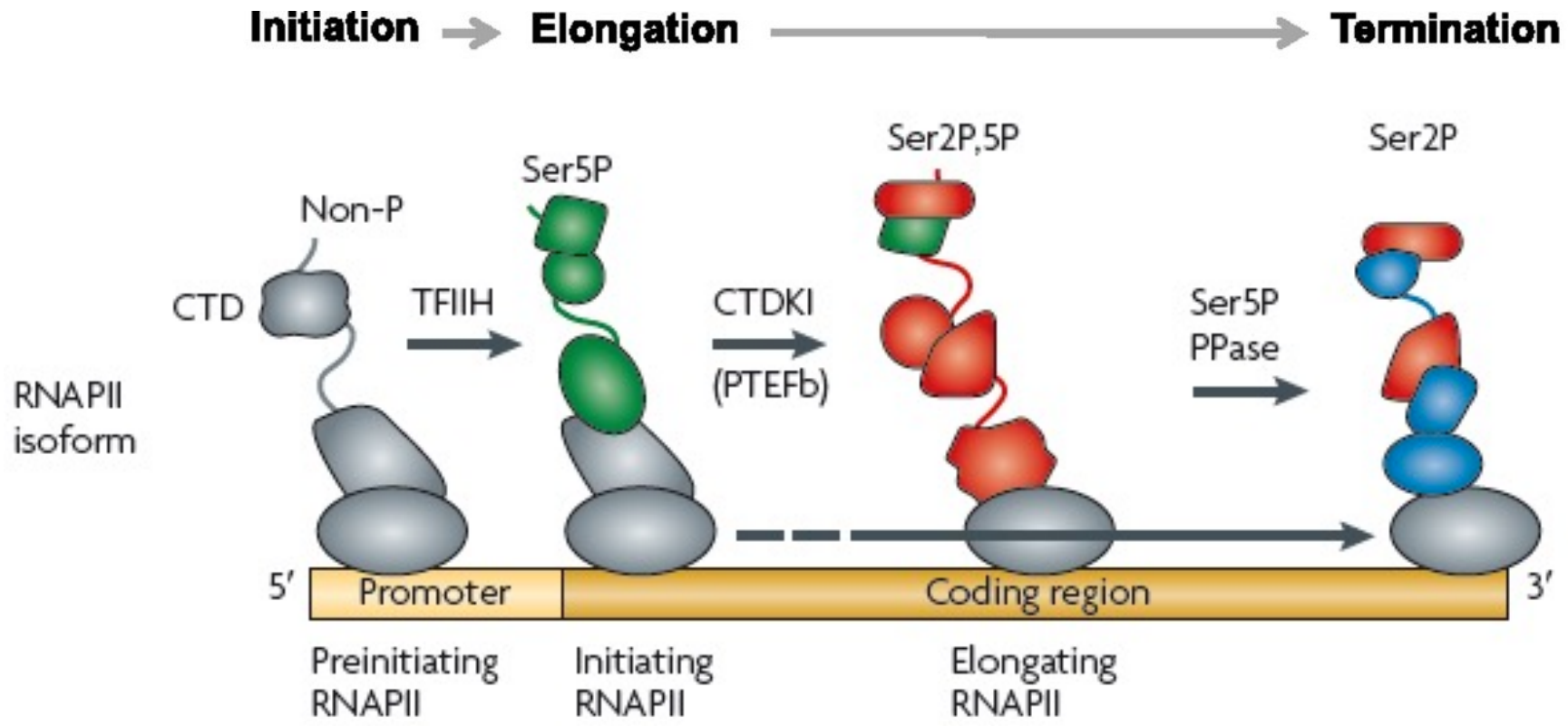
	1	2	3	4	5	6	7
1.	Glu	Ala	Pro	Thr	Ser	Pro	Gly
2.	Phe	Gly	Val	Ser	Ser	Pro	Gly
3.	Phe	Ser	Pro	Thr	Ser	Pro	Thr
4.	Tyr	Ser	Pro	Thr	Ser	Pro	Ala
5.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
6.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
7.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
8.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
9.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
10.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
11.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
12.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
13.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
14.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
15.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
16.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
17.	Tyr	Ser	Pro	Thr	Ser	Pro	Ala
18.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
19.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
20.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
21.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
22.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
23.	Tyr	Ser	Pro	Thr	Ser	Pro	Asn
24.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
25.	Tyr	Ser	Pro	Thr	Ser	Pro	Gly
26.	Tyr	Ser	Pro	Thr	Ser	Pro	Ala

Tyr Ser Pro Lys Gln Asp Glu Gln Lys His Asn Glu Asn Glu Asn Ser Arg

Dynamic modification of the CTD during the transcription cycle

- The pattern of CTD phosphorylation during the transcription cycle is highly *dynamic* and requires the activity of dedicated phosphatases as well as kinases.

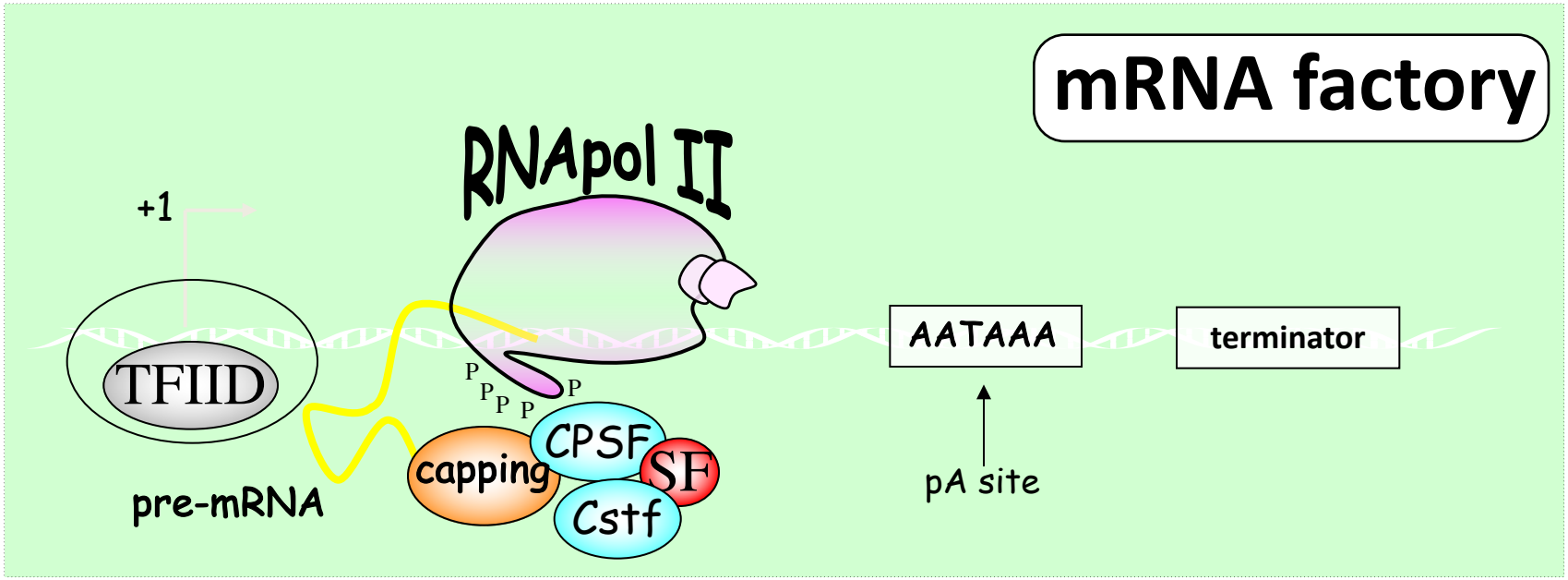
- Transcription *steps* are marked by different modifications of the C-terminal domain



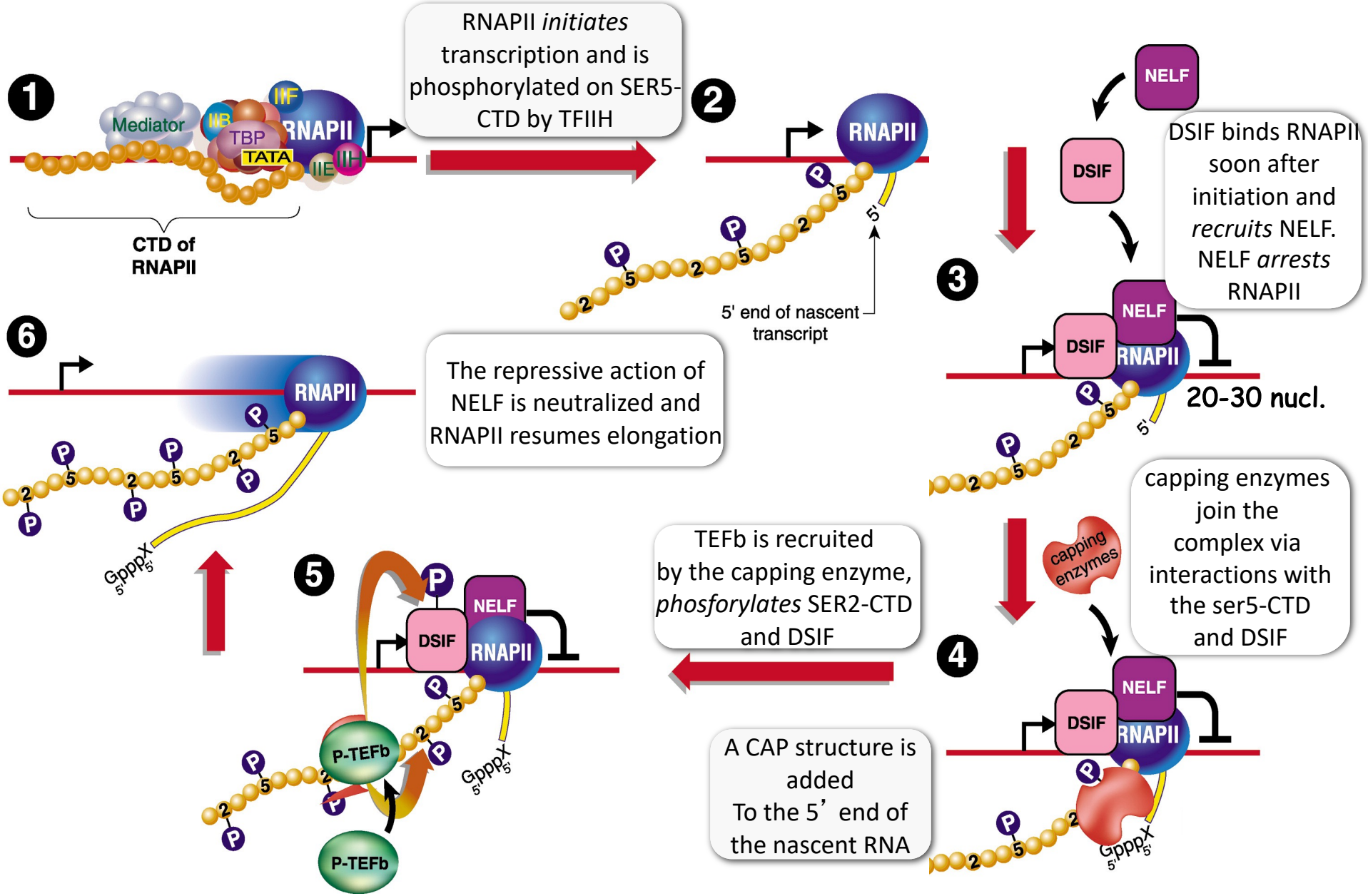
The binding of specific factors starts from the *first steps of gene expression* and directs the nascent ribonucleoprotein complexes along specific pathways of maturation.

By *tethering* machines to each other and to their substrates *coupling* plays a critical role in gene expression dramatically increasing the *specificity* of enzymatic reactions.

The *fate* of a specific RNA is determined at the beginning of transcription



Capping and transcriptional pausing: checkpoint model



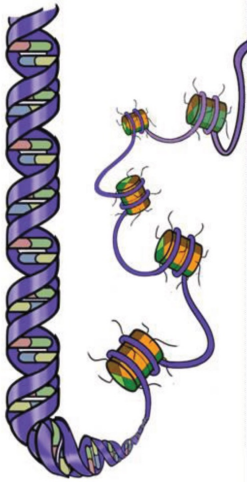
Chromatin organization and structure

Nucleosomal scale

1 pb - ~10 kb

EPIGENETIC MODIFICATIONS

NUCLEOSOMES

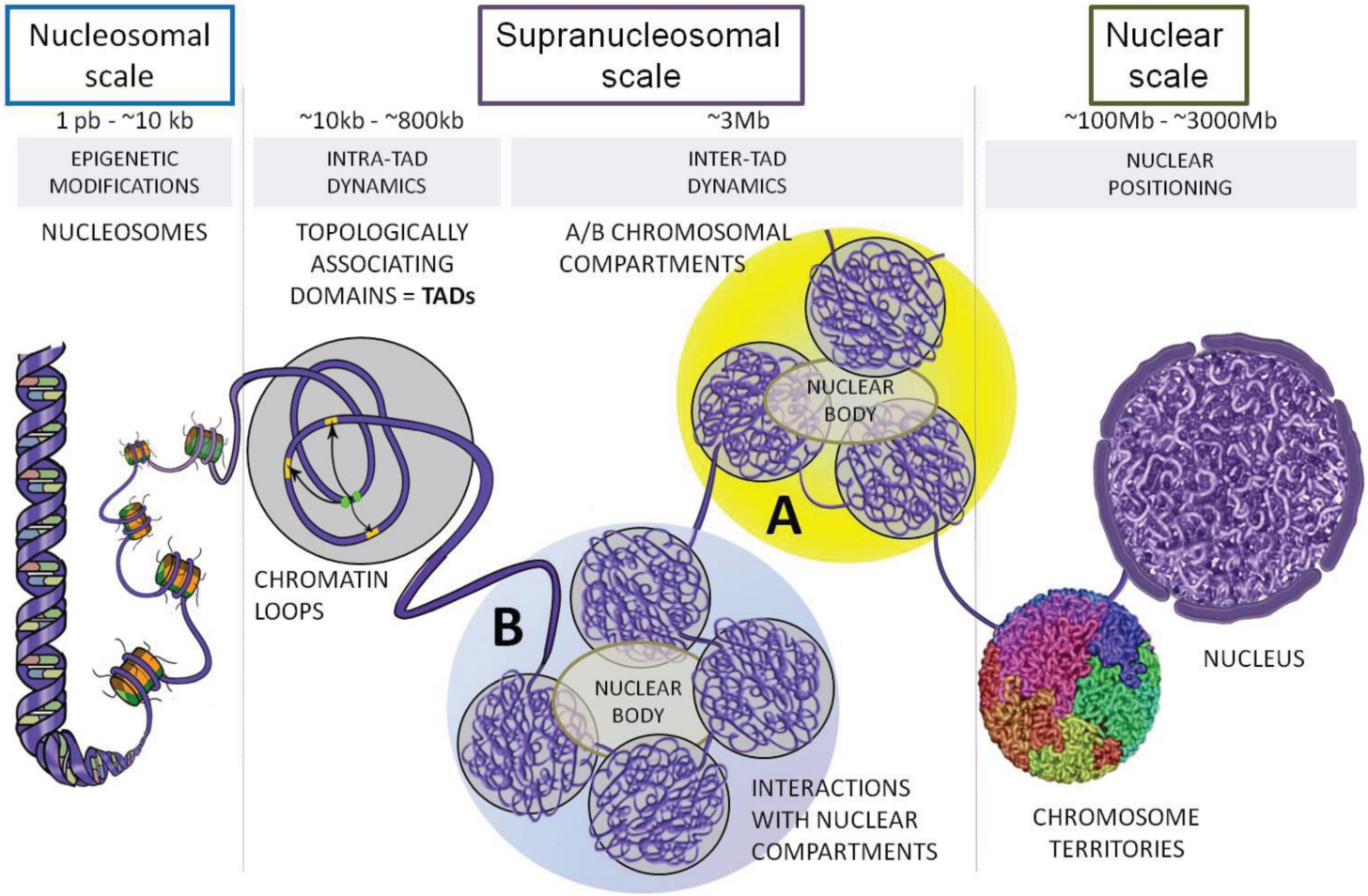


This extremely tight association between DNA and nucleosomes begs the question how DNA is kept accessible to regulatory proteins such as transcription factors. The last two decades have seen tremendous progress in figuring out how chromatin is maintained and remodeled. These mechanisms involved many different modification of histones (histone acetylation) and DNA (methylation). Furthermore, different variants of the histone proteins can be substituted for each other.

One aspect of epigenetic regulation is the positioning of histones on the DNA.

However …this is not enough!

Chromatin organization and structure



COLLECTION | 02 SEPTEMBER 2019

The 3D genome

ALERT!

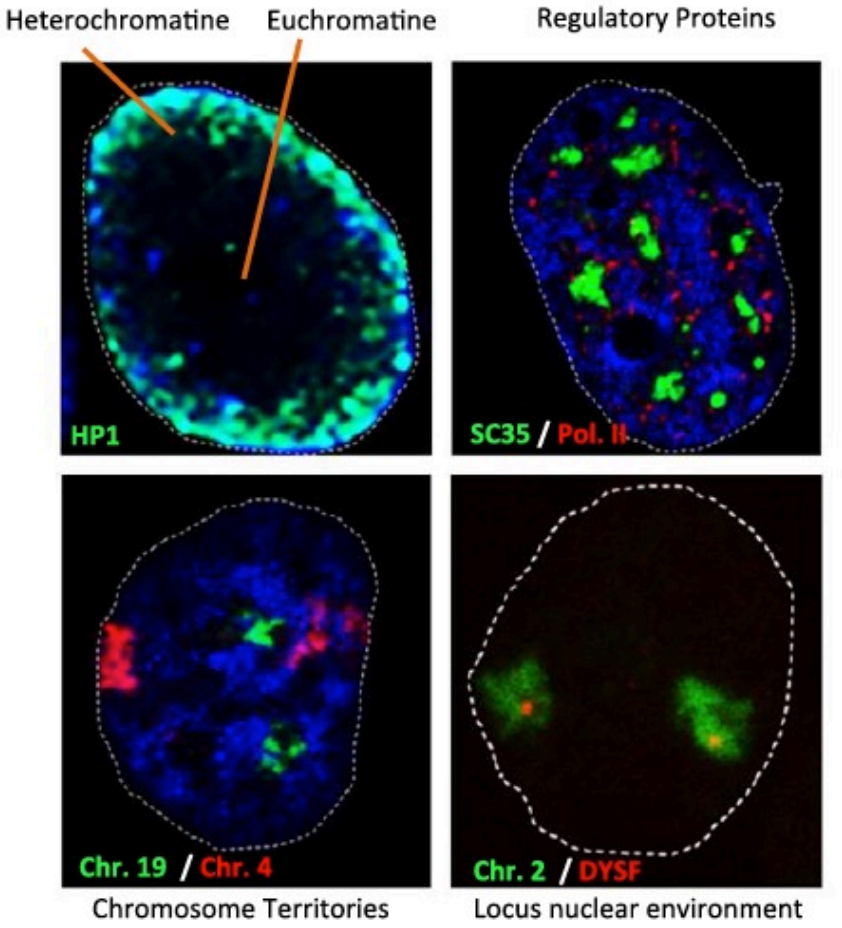
The **organization of the genome** is interconnected with nuclear architecture and and can vary between *cell types* and *during cell differentiation and development*

https://www.nature.com/collections/rsxlmsyslk

Why studying genome 3D structure?

Because it is not random!!!

There is considerable cytological evidence and molecular evidence from chromosome conformation capture approaches (such as 3C and 4C) for the spatial clustering of active genes and genomic regions in the nucleus.



Nuclear position of loci :

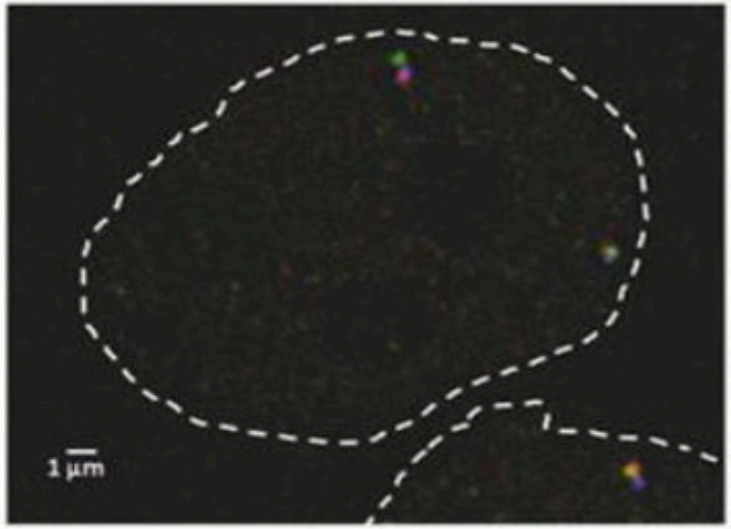
-Differential **disponibility and accessibility** of **regulatory proteins.**

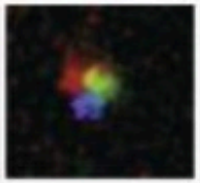
-Distinct **constraints from the surrounding chromatin / nuclear environment.**

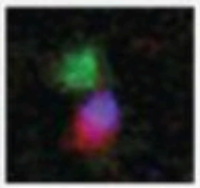
How to study the genome 3D structure?

Microscopy :

Fluorescent In Situ Hybridization (FISH)







- Single Cell
- Low Throughput
- Limited Resolution

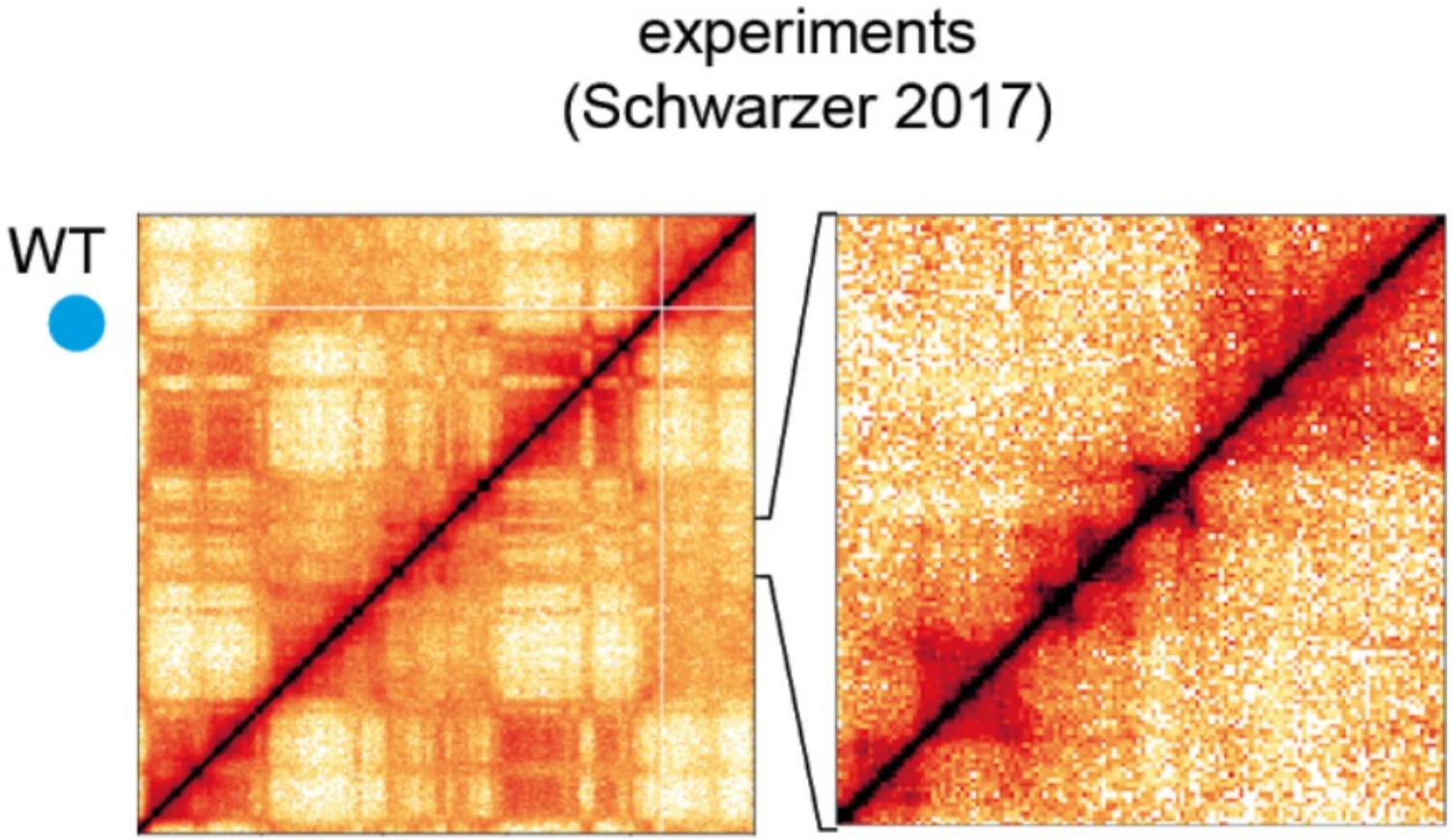
Molecular Biology :

Chromosome Conformation Capture (3C & Co)

*C Methodologies
- 3C
- 4C
- 5C
- ChIA-PET
- Hi-C
- *In situ* Hi-C
- Capture-C
- ChIC
- Single cell Hi-C
- Hi-ChiP
- ...

- Population based
- Access to different resolutions
- Population based
- Resolution cost

These methods generate detailed maps of how likely it is that two points on the chromosome touch.



http://higlass.io/app/?config=MSHhOBbOSW6iIovB5yk6BA

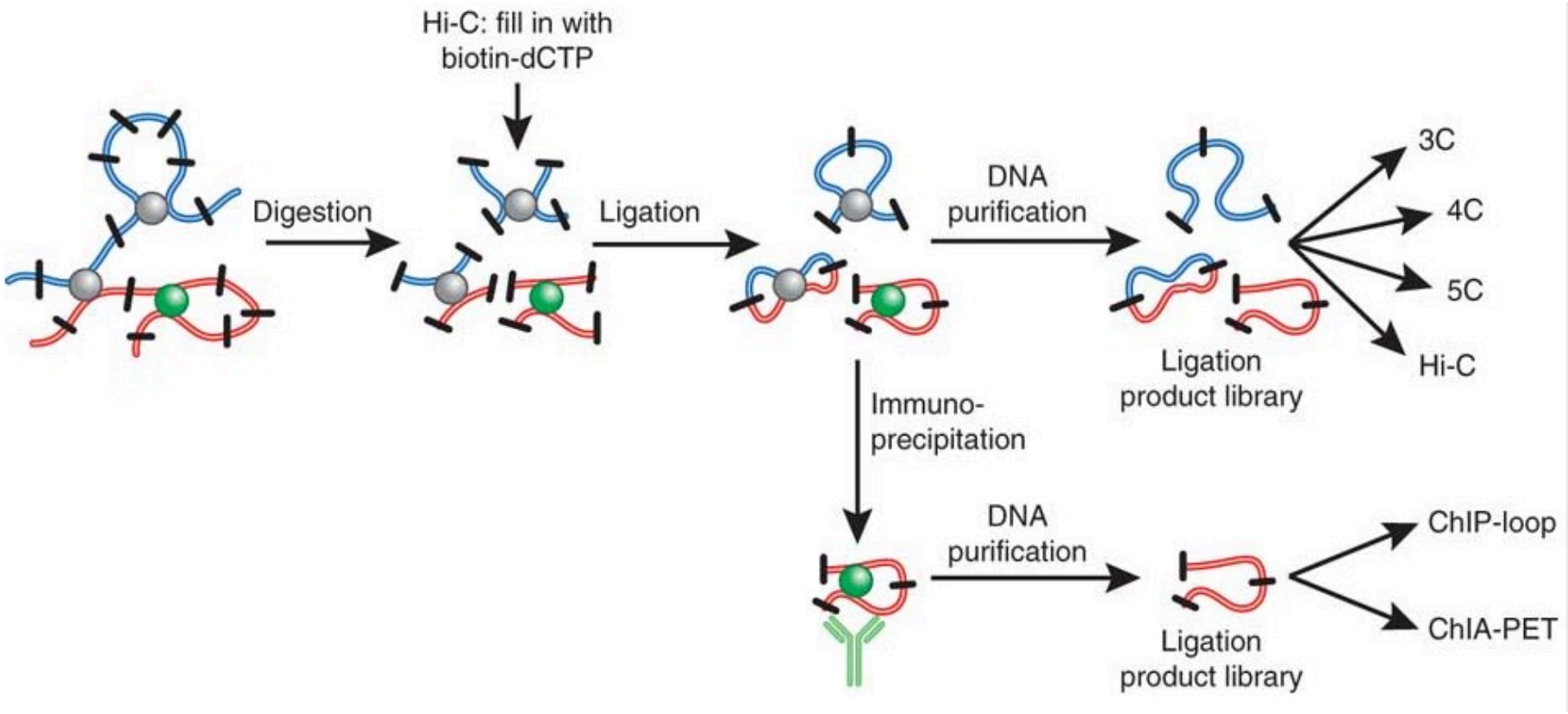
Table 1 | **A tabulation of known chromosome conformation capture technologies**

Assay abbreviation	Full assay name	Refs	Related protocols or guidelines
1 versus 1*			
3C	Chromosome conformation capture	43	97–100
1 versus Many/All*			
Multiplexed 3C-seq	Multiplexed chromosome conformation capture sequencing	101	102
Open-ended 3C	Open-ended chromosome conformation capture	103	–
3C-DSL	Chromosome conformation capture combined with DNA selection and ligation	104	–
4C	Circular chromosome conformation capture	45	105
4C	Chromosome conformation capture-on-chip	51	–
4C-seq	Chromosome conformation capture-on-chip combined with high-throughput sequencing	106	46,72, 107,108
TLA	Targeted locus amplification	30	–
e4C	Enhanced chromosome conformation capture-on-chip	109	110
ACT	Associated chromosome trap	111	112
Many versus Many*			
5C	Chromosome conformation capture carbon copy	52	113–116
ChIA–PET	Chromatin interaction analysis paired-end tag sequencing	23	–
Many versus All*			
Capture-3C	Chromosome conformation capture coupled with oligonucleotide capture technology	25	–
Capture-HiC	Hi-C coupled with oligonucleotide capture technology	58	–
All versus All*			
GCC	Genome conformation capture	–	117
Hi-C	Genome-wide chromosome conformation capture	22	69,70,118
ELP	Genome-wide chromosome conformation capture with enrichment of ligation products	119	–
TCC	Tethered conformation capture	24	–
Single-cell Hi-C	Single-cell genome-wide chromosome conformation capture	38	96
In situ Hi-C	Genome-wide chromosome conformation capture with *in situ* ligation	27	–
DNase Hi-C	Genome-wide chromosome conformation capture with DNase I digestion	49	–
Micro-C	Genome-wide chromosome conformation capture with micrococcal nuclease digestion	50	–

*'1', 'Many' and 'All' indicate how many loci are interrogated in a given experiment. For example, '1 versus All' indicates that the experiment probes the interaction profile between 1 locus and all other potential loci in the genome. 'All versus All' means that one can detect the interaction profiles of all loci, genome-wide, and their interactions with all other genomic loci.

Nuclear clustering of active genes

There is considerable cytological evidence and molecular evidence from chromosome conformation capture approaches (such as 3C and 4C) for the spatial clustering of active genes and genomic regions in the nucleus.



Article ALERT!

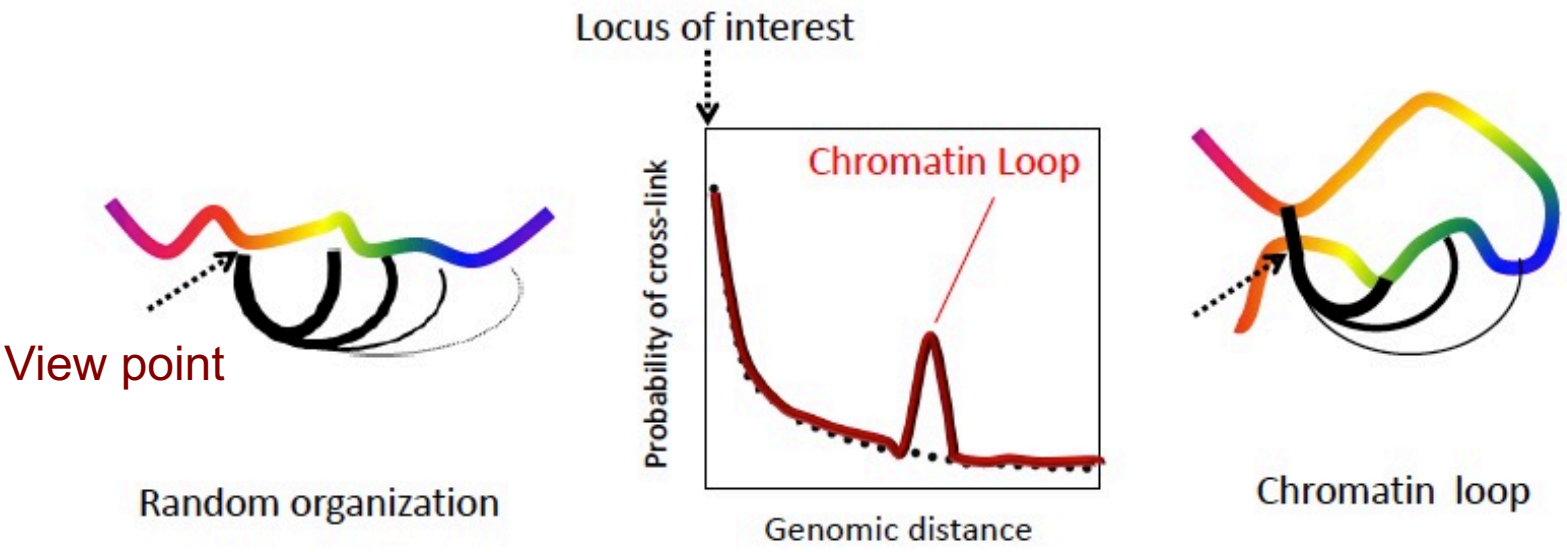
Genomics tools for unraveling chromosome architecture

Bas van Steensel & Job Dekker
Nature Biotechnology **28**, 1089–1095 (2010)

Chromosome Conformation Capture 3C (Dekker J *et al* 2002)

* Take a snapshot of the conformation of the chromatin fiber (Formaldehyde Cross-link).

* Probability of cross-link depend on the distance between two loci.

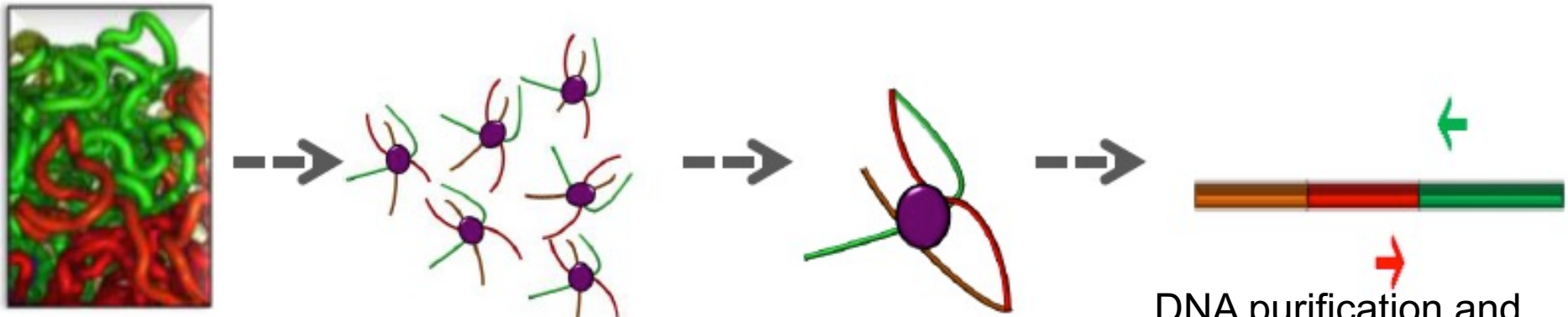


Chromosome Conformation Capture 3C (Dekker J *et al* 2002)

* Take a snapshot of the conformation of the chromatin fiber (Formaldehyde Cross-link).

* Probability of cross-link depend on the distance between two loci.

* Convert the captured spatial contacts in **detectable/quantificable products** .



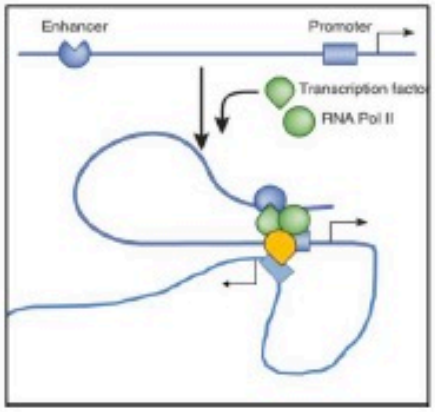
Crosslink chromatin

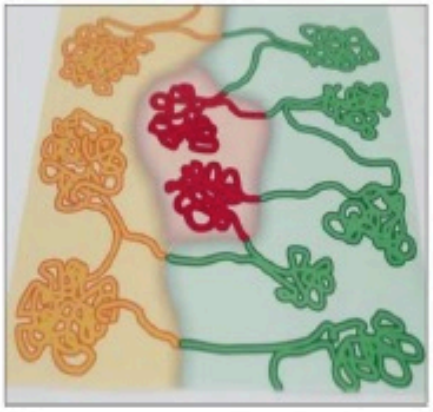
Digest Chromatin with restriction enzyme

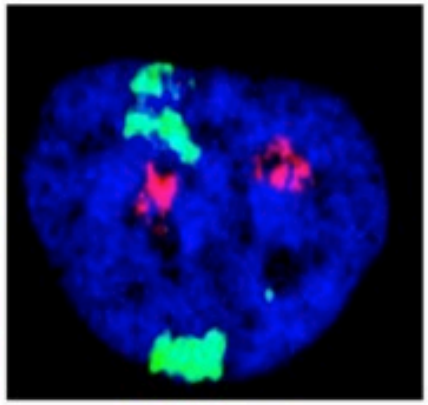
ligation

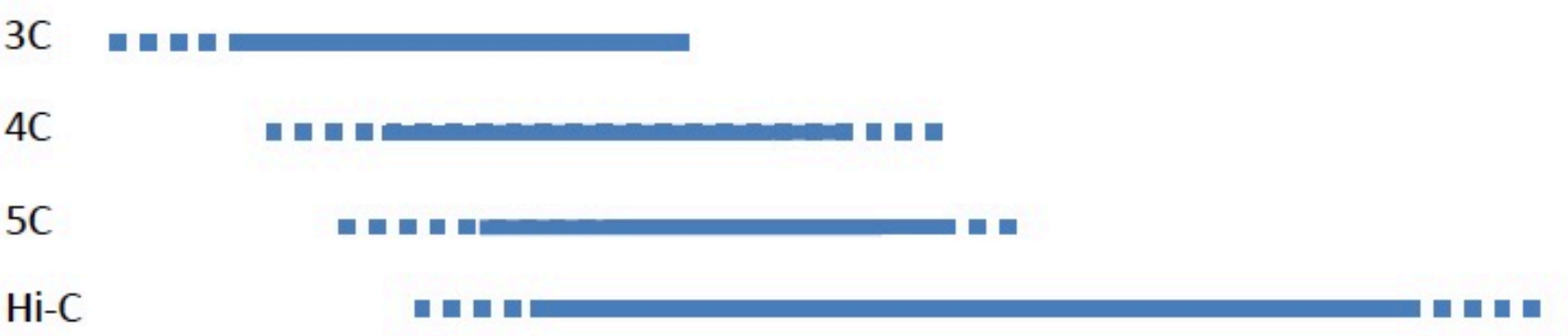
DNA purification and detection of ligation events by PCR

Which *C for which question?









Theorical maximal resolution depends on restriction enzyme choice (6 bp vs 4 bp cutter) and sequencing depth (Increasing the resolution by a factor 10 require increased sequencing depth by a factor 100).