

《分子生物学》第五讲

肖锐

2023.3.24. 珞珈山工学部6教201



肖锐 武汉大学医学研究院 教授/博导

国家高层次青年人才计划

湖北省“百人计划”，“楚天学者”特聘教授

Email: xiaorui9@whu.edu.cn

地址：武汉大学医学部8号楼1301，1302，1304

教育与科研经历： 1999-2003: 武汉大学，生物技术专业，学士
2004-2010: 武汉大学，生物化学与分子生物学，博士
2011-2017: 美国加州大学圣地亚哥分校(UCSD)，博士后
2017.12至今：武汉大学，医学研究院，教授/博导

Research interests: **Functional genomics and RNAomics**

1. RNA-binding proteins in transcription and disease
2. Cryptic splicing and disease

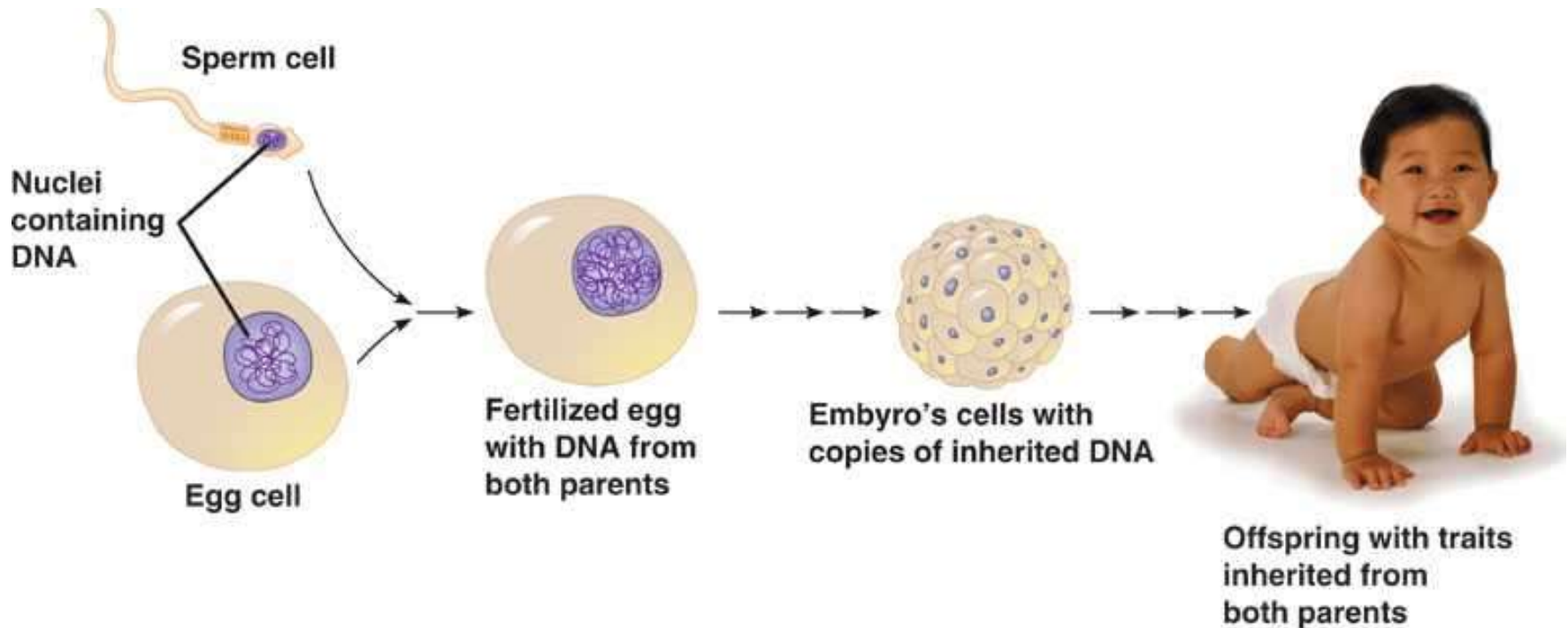
代表性论文： *Cell* 2019; *Nature* 2020; *Nature* 2015;
Mol Cell 2012, 2016;



Brief Contents

1. Transcription in Eukaryotes
2. RNA Splicing
3. Translation and the Genetic Code
4. Transcriptional Regulation in Prokaryotes
5. Transcriptional Regulation in Eukaryotes

Human development



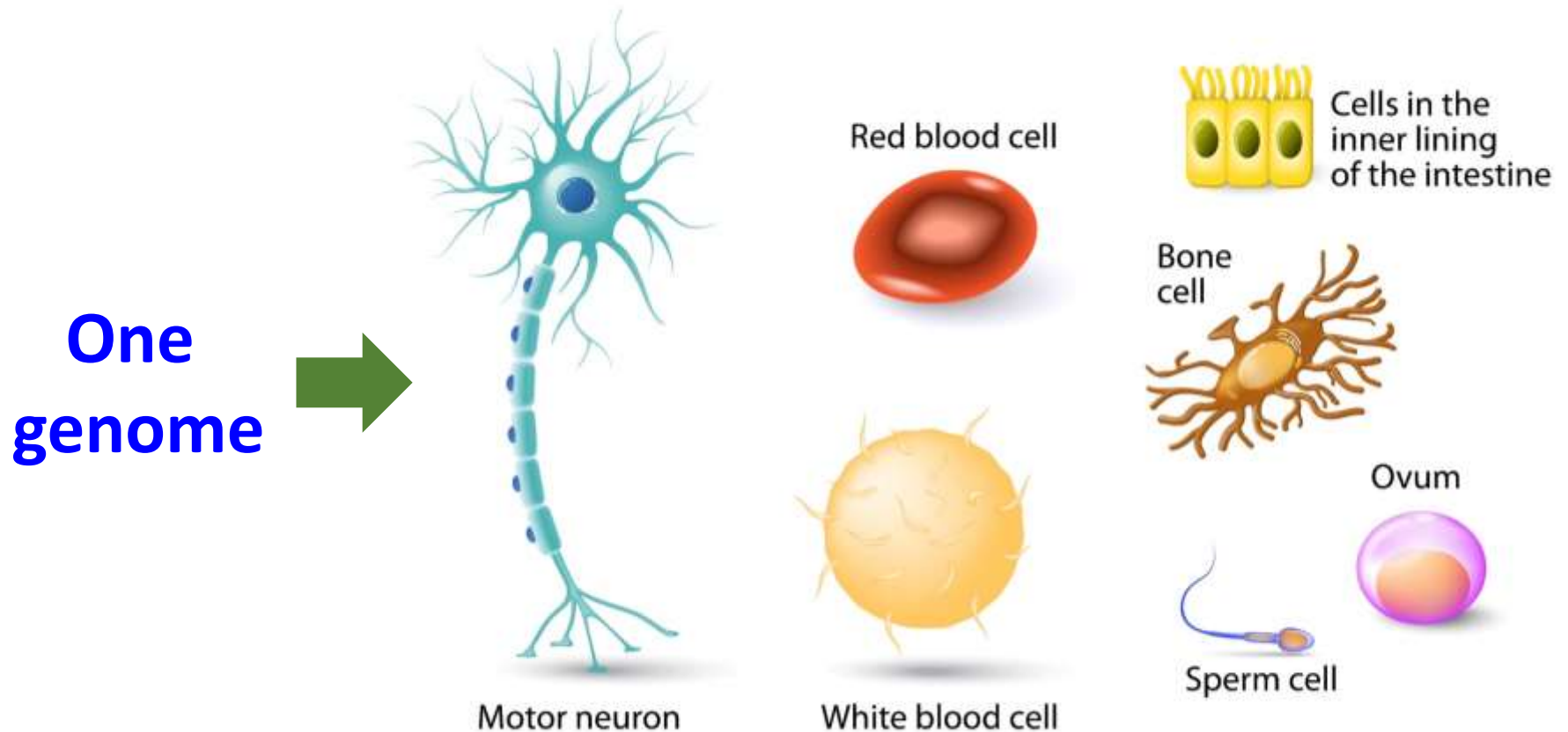
Human tissue # = 4

organ # = 78

cell type # = >233

cell # = 37 trillion

The power of transcription



Different gene expression shapes different cell types.

Differential gene expression

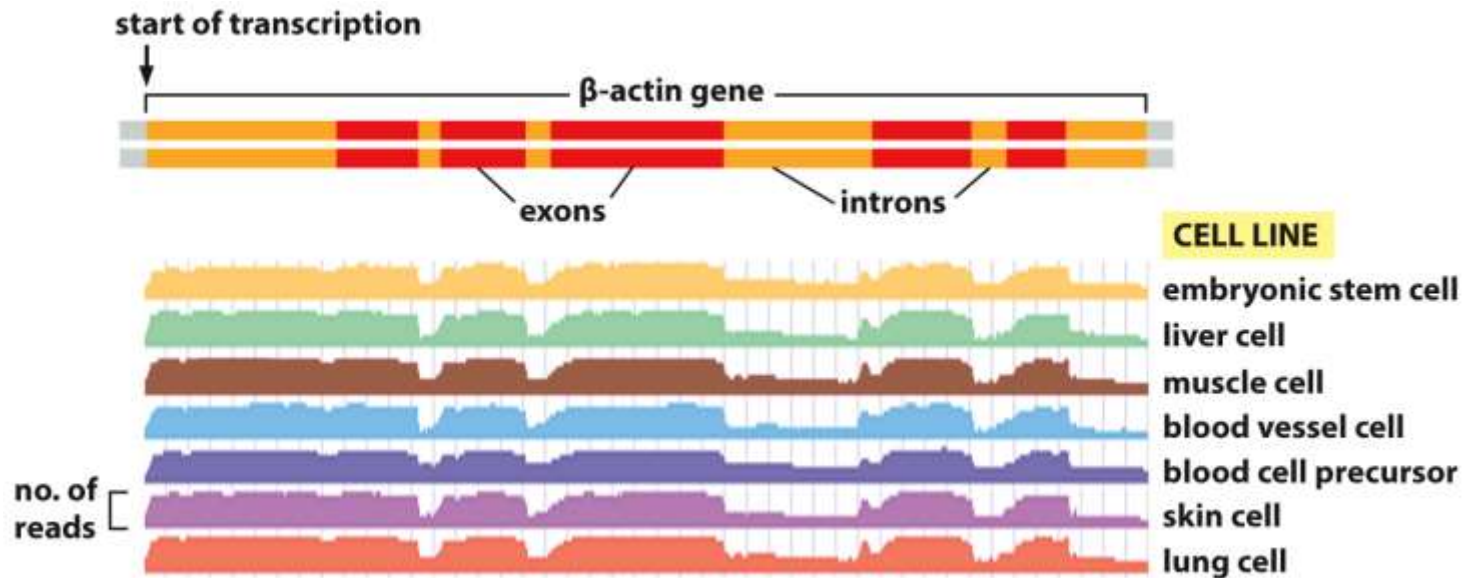


Figure 7-3a Molecular Biology of the Cell 6e (© Garland Science 2015)

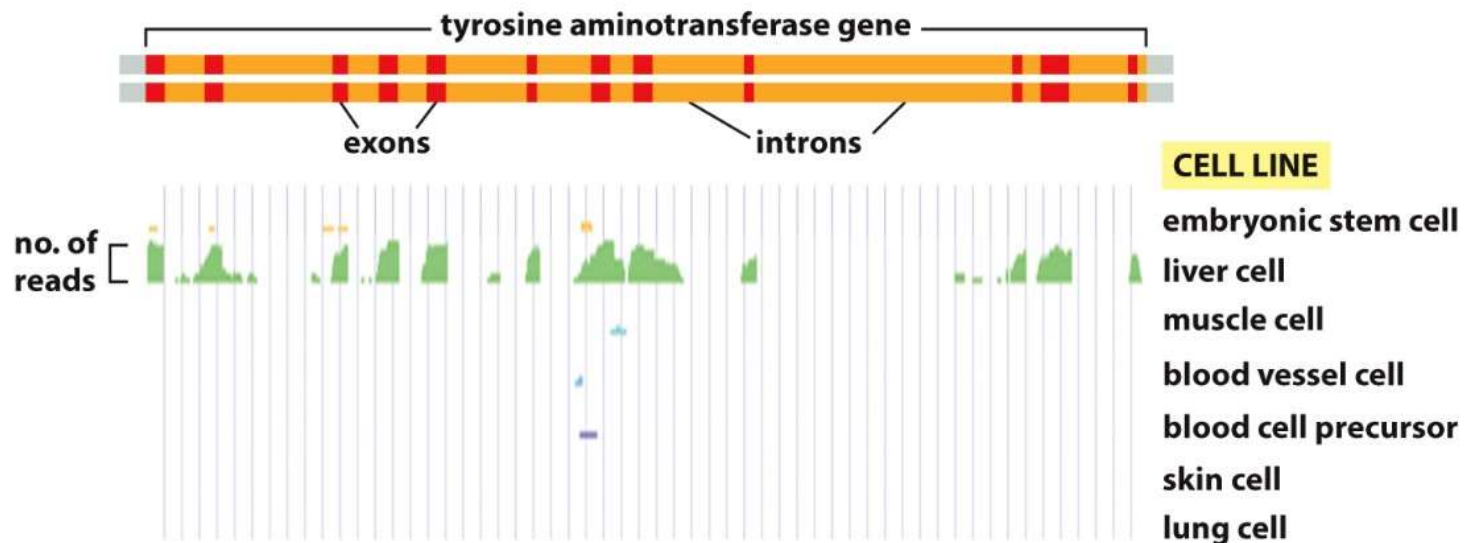


Figure 7-3b Molecular Biology of the Cell 6e (© Garland Science 2015)

Chapter 12-Part II

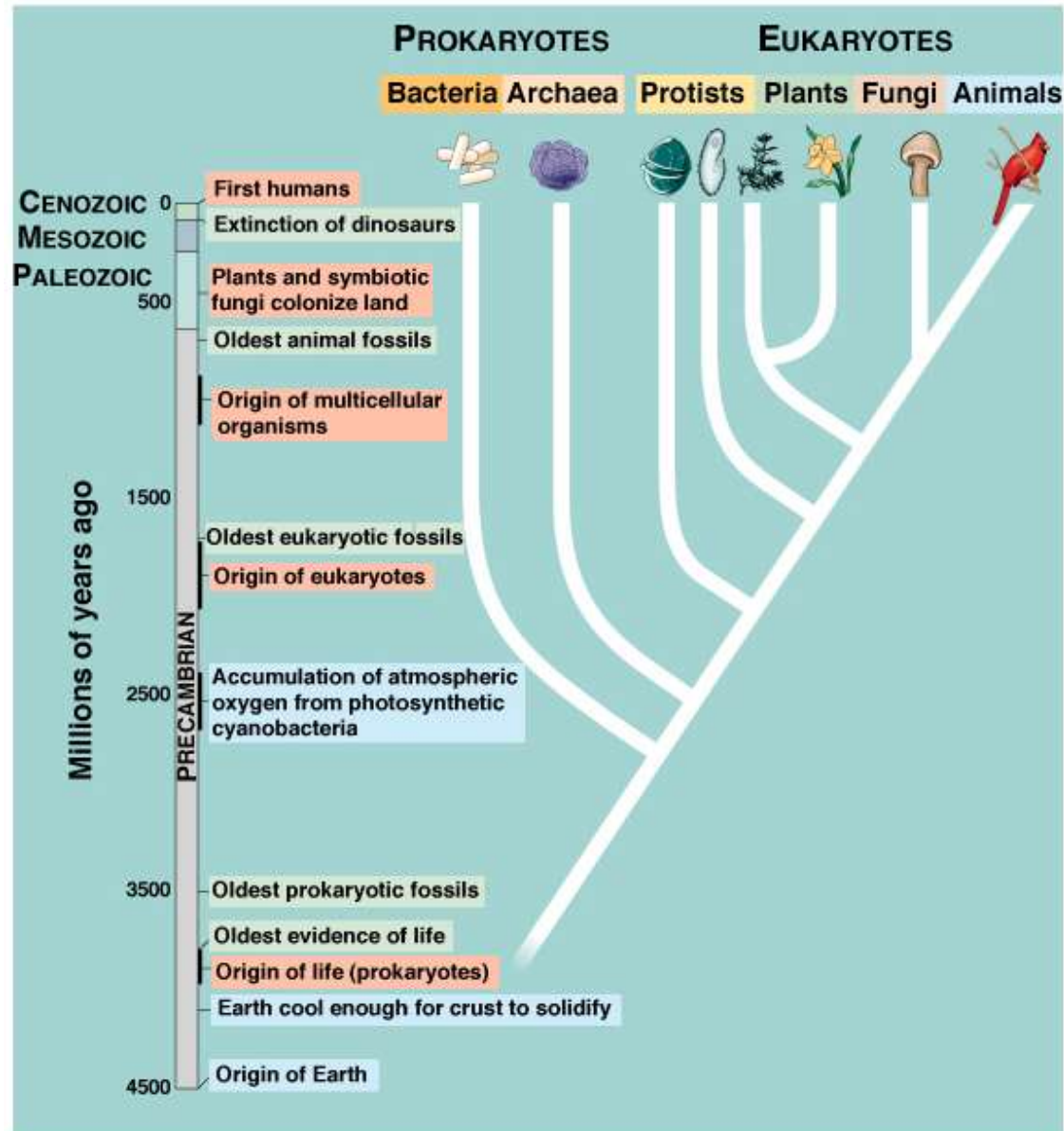
Transcription in Eukaryotes

Rui Xiao

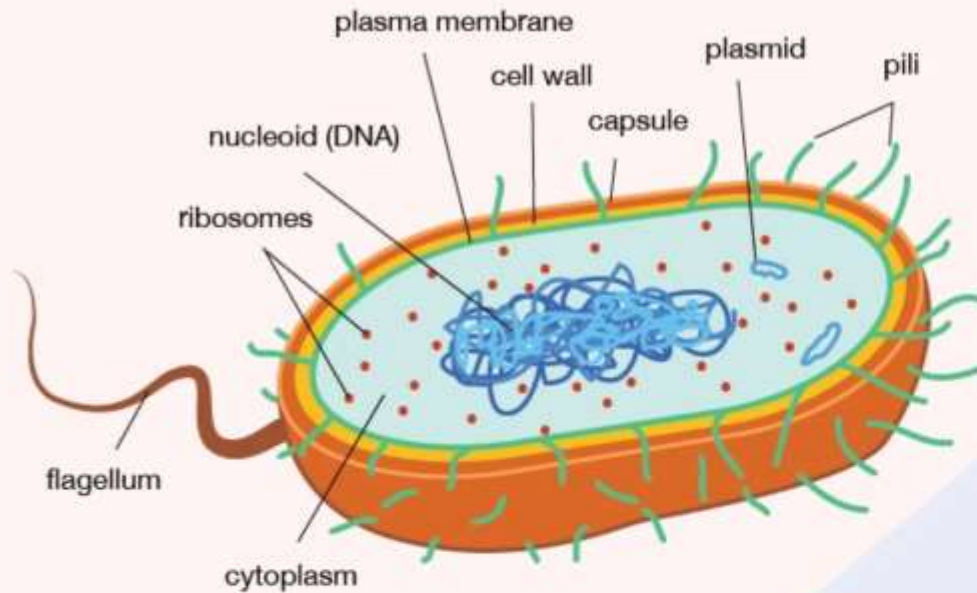
Outline

- 1. Eukaryotic transcription machineries**
- 2. Function elements for eukaryotic transcription**
- 3. Transcription initiation in eukaryotes**
- 4. Transcription-coupled RNA processing**
- 5. Techniques to study eukaryotic transcription**
- 6. Examples for studying eukaryotic transcription**

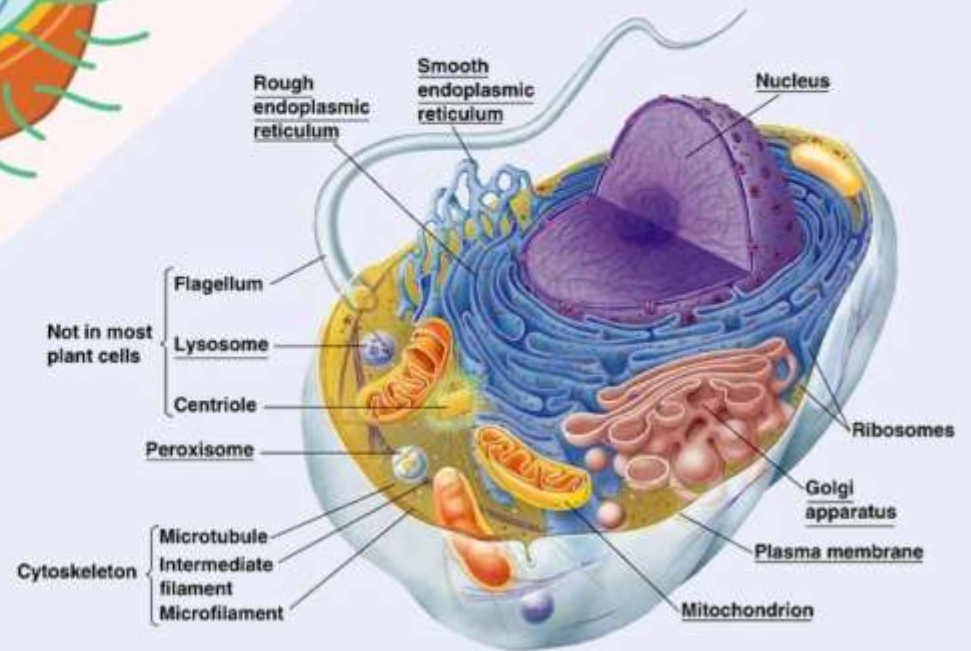
Evolution of life on earth



Eukaryotes VS Prokaryotes



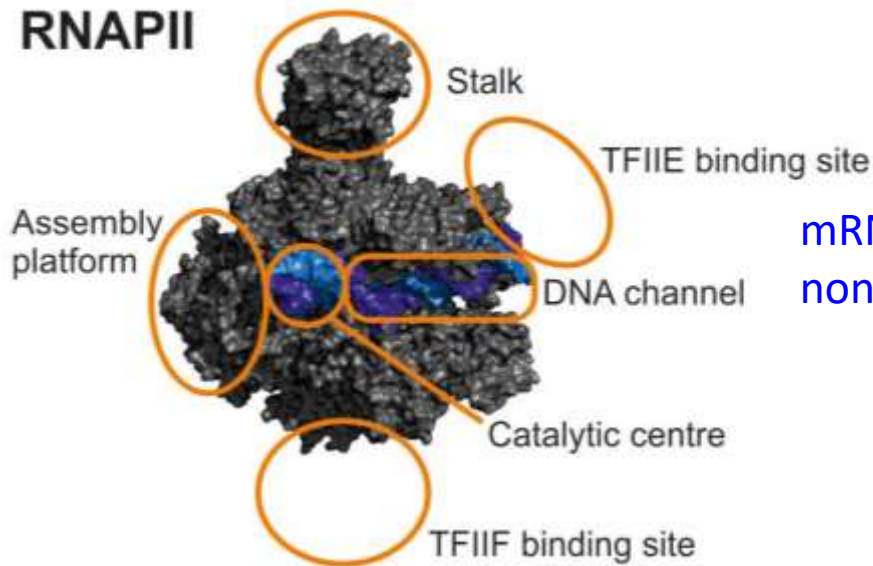
Prokaryotic Cell



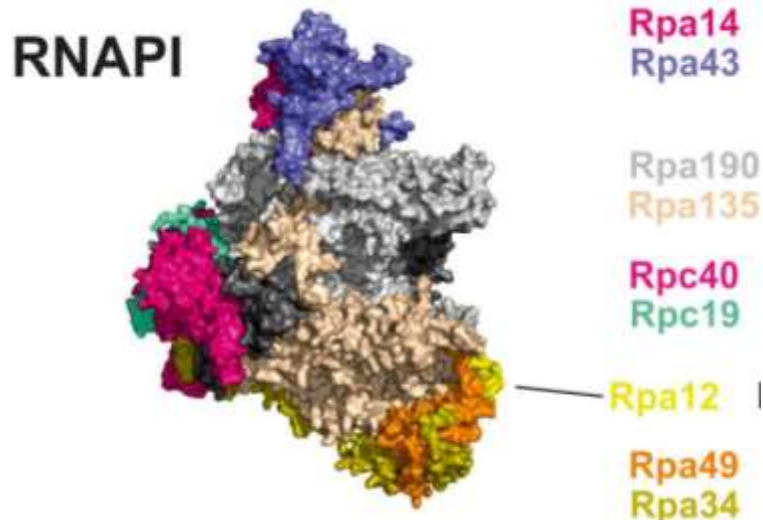
Eukaryotic Cell

How different?

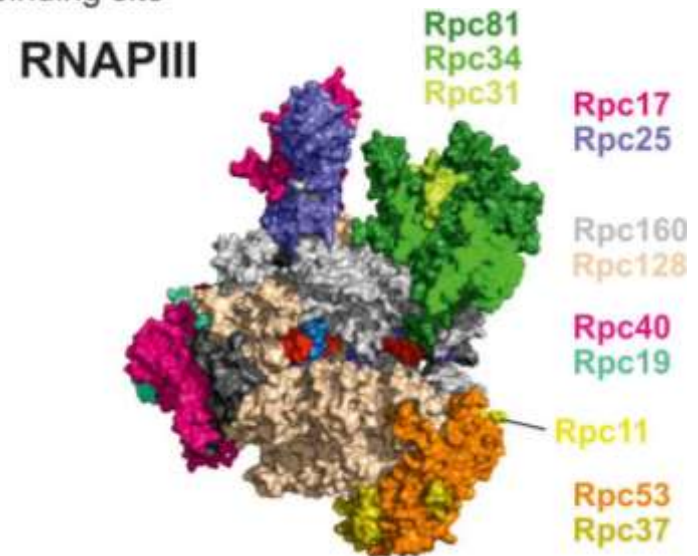
Eukaryotic RNA polymerases



mRNAs and most
non-coding RNAs



45S rRNA, the precursor
of 18S, 5.8S and 28S rRNA

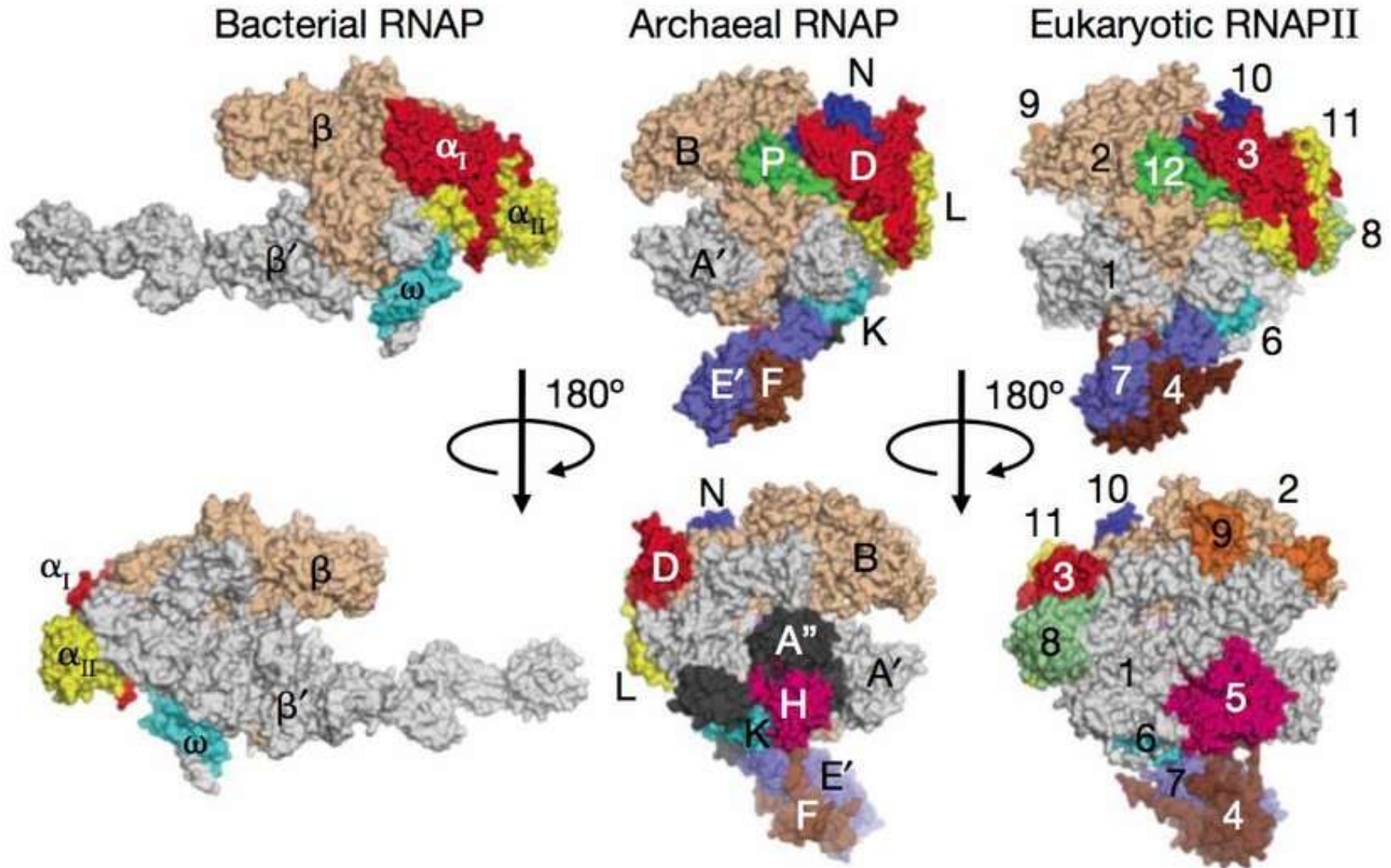


5S rRNA, tRNA,
U6 snRNA and 7SL RNA

Eukaryotic transcription machineries

Bacteria			Eukaryotes				
RNAP	RNAPII		RNAPI		RNAPIII		
	yeast	human	yeast	human	yeast	human	
β-subunit	Rpb1	RPB1	Rpa190	RPA1	Rpc160	RPC1	
β-subunit	Rpb2	RPB2	Rpa135	RPA2	Rpc128	RPC2	
ω-subunit	Rpb6	RPABC2	Rpb6	RPABC2	Rpb6	RPABC2	
	Rpb5	RPABC1	Rpb5	RPABC1	Rpb5	RPABC1	
	Rpb8	RPABC3	Rpb8	RPABC3	Rpb8	RPABC3	
α-subunit	Rpb10	RPABC5	Rpb10	RPABC5	Rpb10	RPABC5	Assembly platform
	Rpb12	RPABC4	Rpb12	RPABC4	Rpb12	RPABC4	
	Rpb3	RPB3	Rpc40	RPAC1	Rpc40	RPAC1	
	Rpb11	RPB11-a	Rpc19	RPAC2	Rpc19	RPAC2	
α-subunit	Rpb4	RPB4	Rpa14	RPA14(?)	Rpa17	RPC9	Stalk
	Rpb7	RPB7	Rpa43	RPA43	Rpc25	RPC8	
	Rpb9	RPB9	Rpa12	RPA12	Rpc11	RPC10	
			Rpa49	RPA49	Rpc53	RPC4	TFIIF-like
			Rpa34	RPA34	Rpc37	RPC5	
					Rpc82	RPC3	TFIIE-like
					Rpc34	RPC6	
					Rpc31	RPC7	

Molecular basis of RNA polymerase



Transcription initiation in eukaryotes

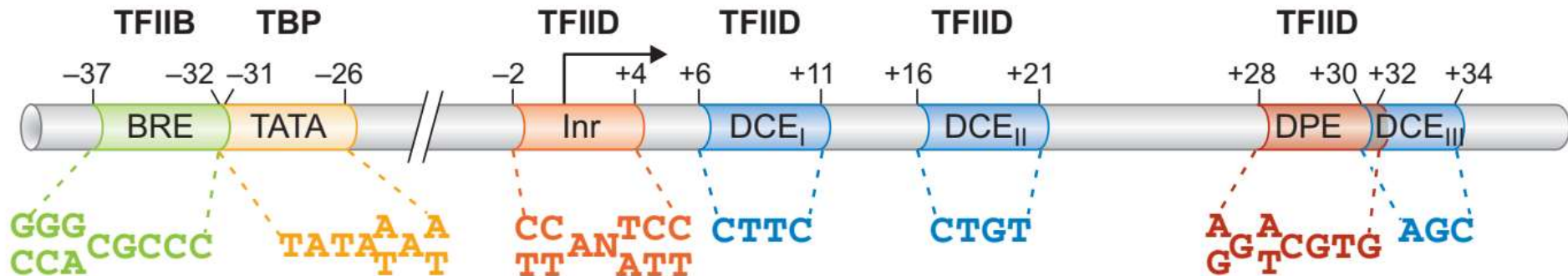
- General transcription factors (GTFs)
- DNA-binding regulatory proteins
(specific transcription factors, TFs)
- Mediator complex
- Chromatin-modifying enzymes
- Core promoter

Outline

1. Eukaryotic transcription machineries
- 2. Function elements for eukaryotic transcription**
3. Transcription initiation in eukaryotes
4. Transcription-coupled RNA processing
5. Techniques to study eukaryotic transcription
6. Examples for studying eukaryotic transcription

Eukaryotic Pol II core promoter

- **Pol II Core promoter**: the minimal set of sequence elements required for accurate transcription initiation by the Pol II machinery

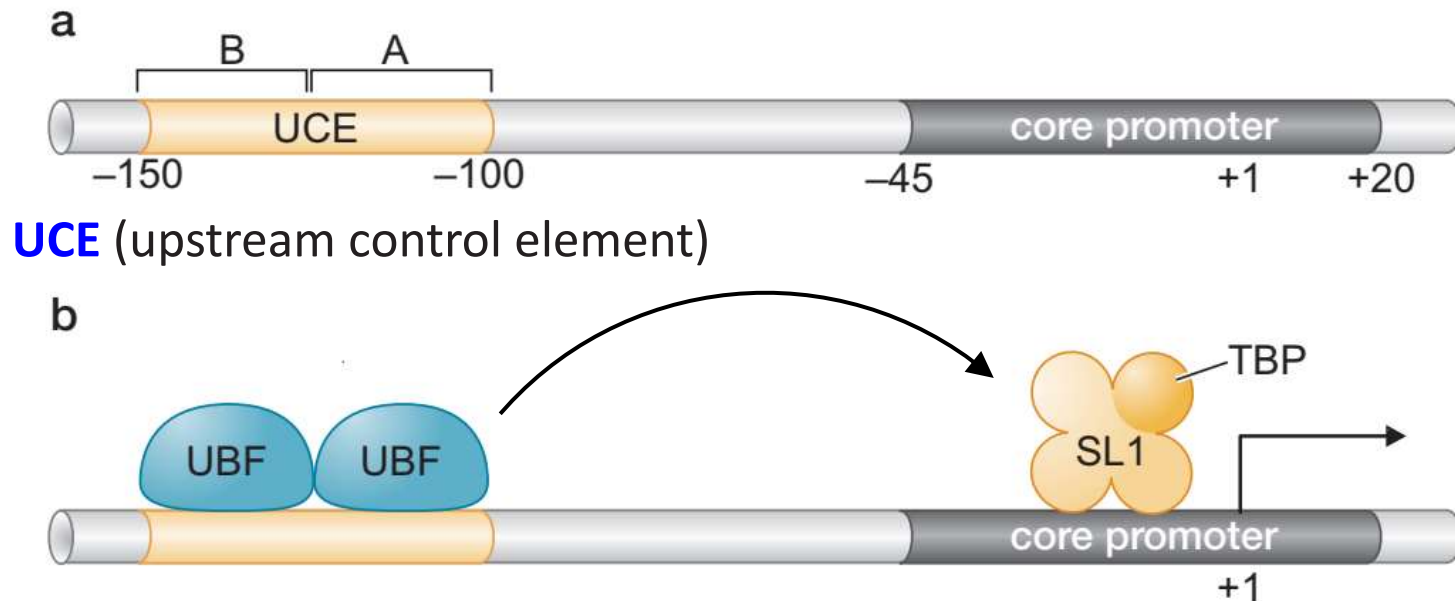


BRE: TFIIB recognition element; **TATA element/box**;

Inr: Initiator element; **DPE**: downstream promoter element

DCE: downstream core element

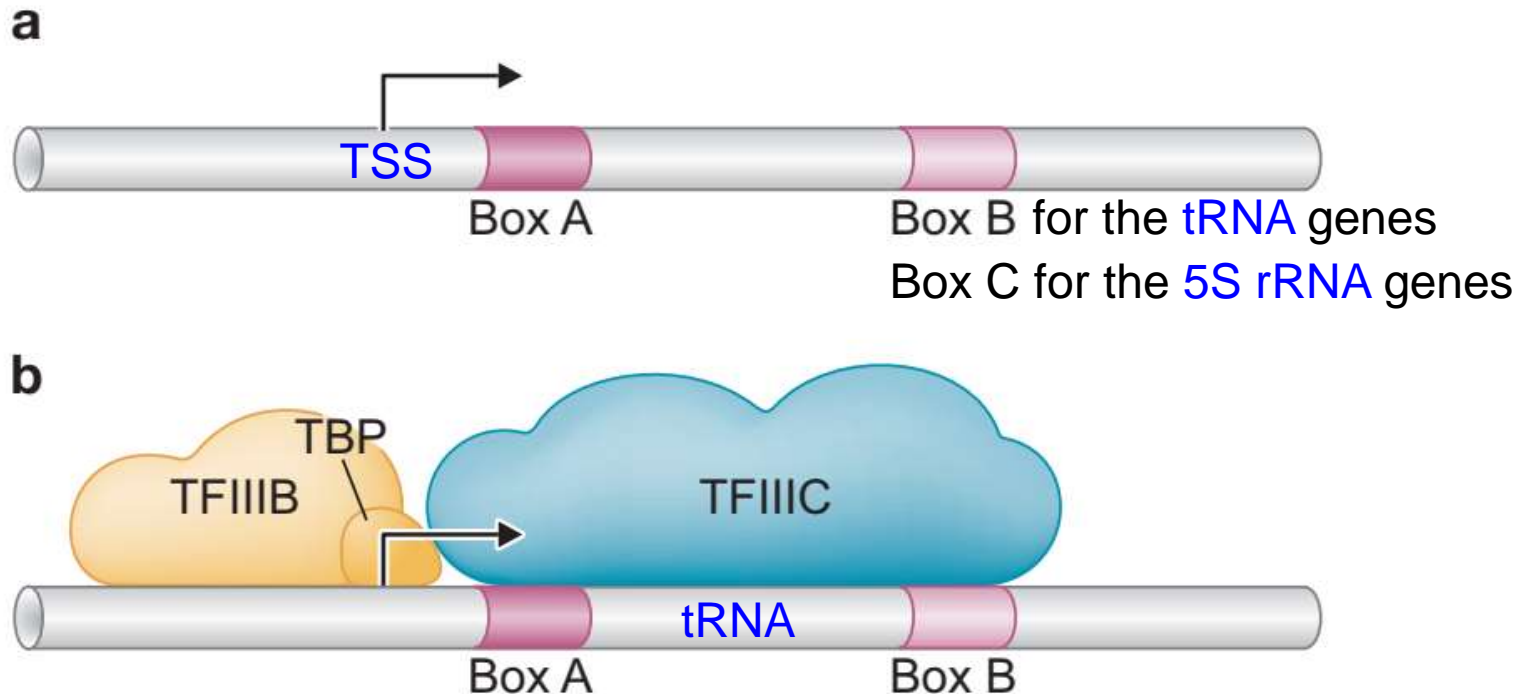
Eukaryotic Pol I core promoter



UCE (upstream control element)

- Initiation requires two other factors, called **SL1** and **UBF**.
- **SL1, which binds to the core promoter**, comprises **TBP** and three **TAFs** specific for Pol I transcription. SL1 binds DNA only in the presence of **UBF**.
- **UBF binds to UCE**, bringing in SL1 and stimulating transcription from the core promoter by recruiting Pol I.

Eukaryotic Pol III core promoter



- The **TFIIIC complex** binds to the promoter region and then recruits **TFIIIB** to the DNA just upstream of the start site, where it, in turn, recruits **Pol III** to the start site of transcription. **Pol III** then initiates, presumably **displacing TFIIIC** from the DNA template as it goes.
- **Pol I, Pol II and Pol III** use **TBP** to initiate transcription.

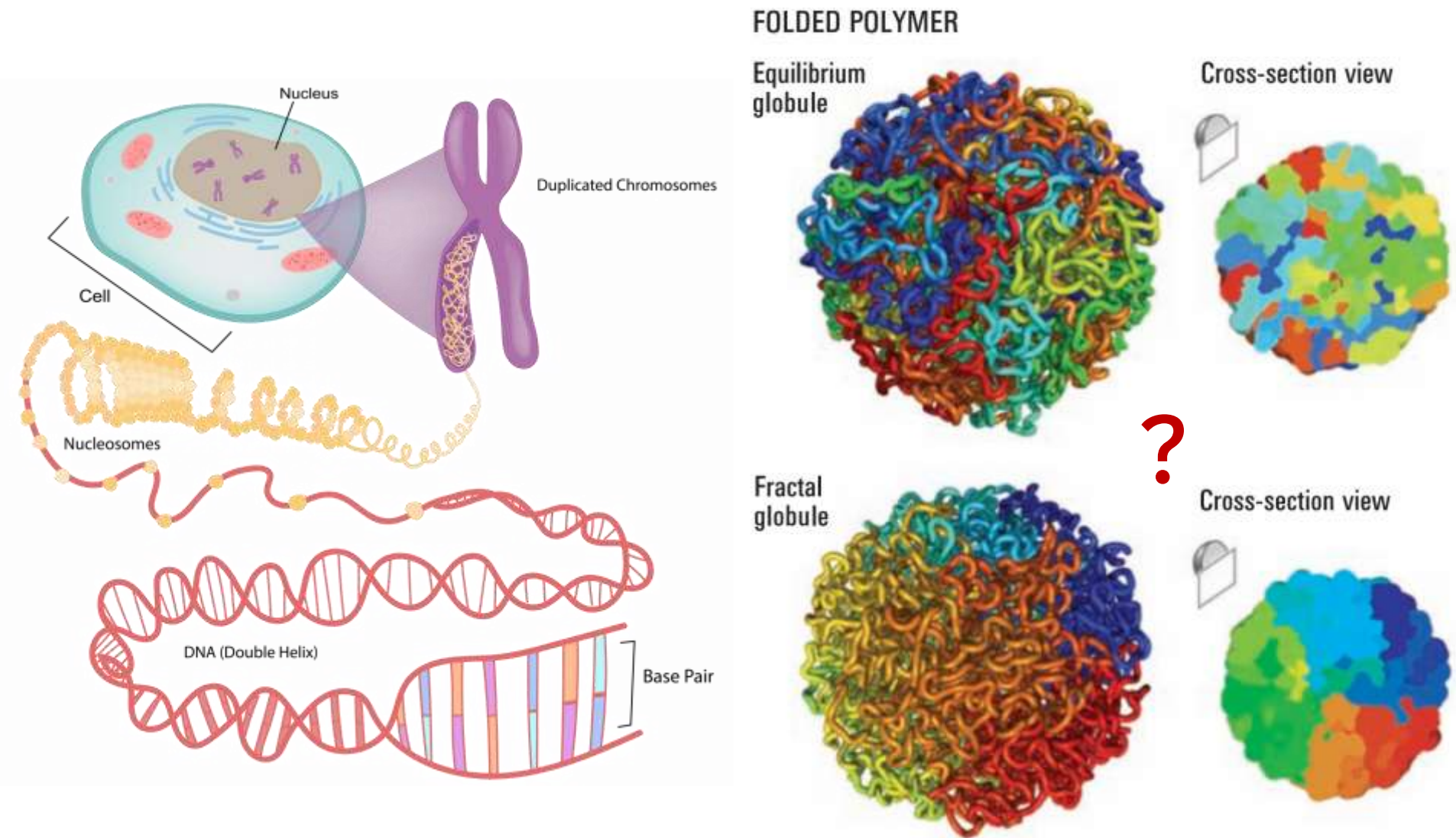
Additional regulatory sequences for Pol II

- Promoter proximal elements
- Upstream activator sequences (UASs)
- Enhancers
- Silencers
- Boundary elements
- Insulators (eg. CTCF binding sites)

Some of these regulatory elements can be located many tens or even hundreds of kilobases from the core promoters on which they act.

How do these elements work in such long distance?

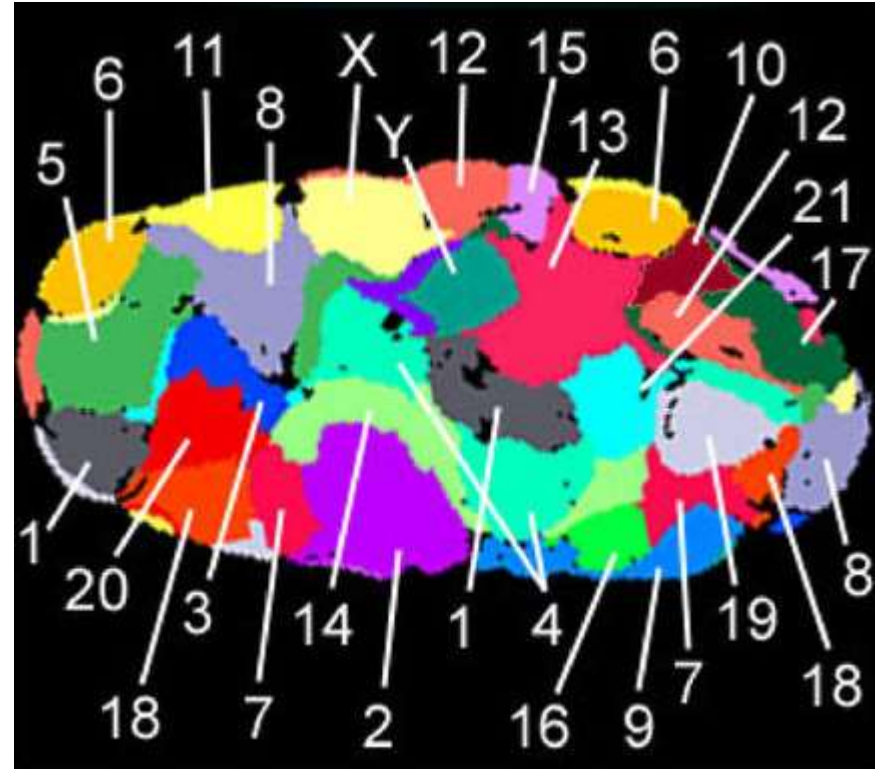
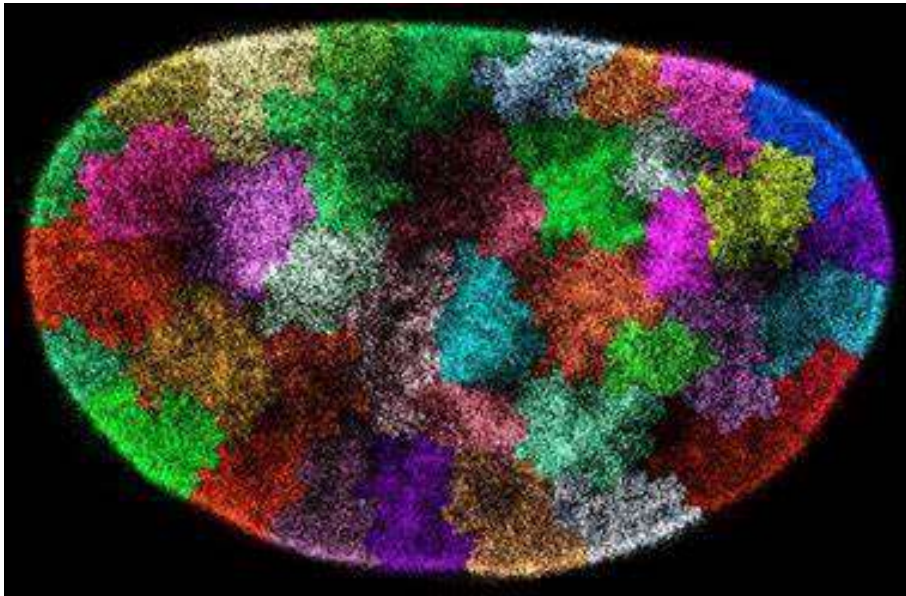
Human genome: equilibrium or fractal?



Chromosome territories

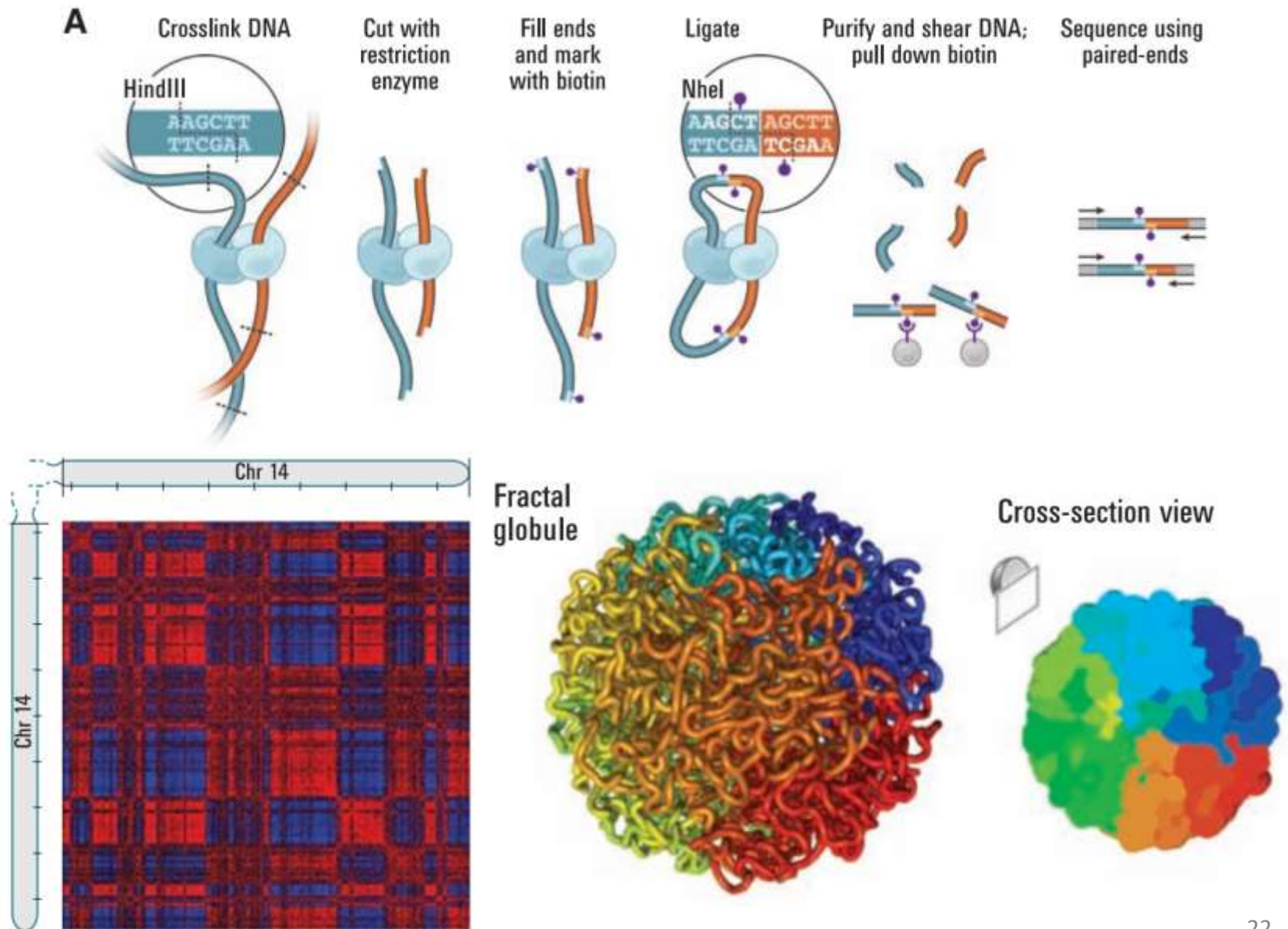
Annotated chromosomes

Chromosome painting



How to study the 3D genome at the molecular level?

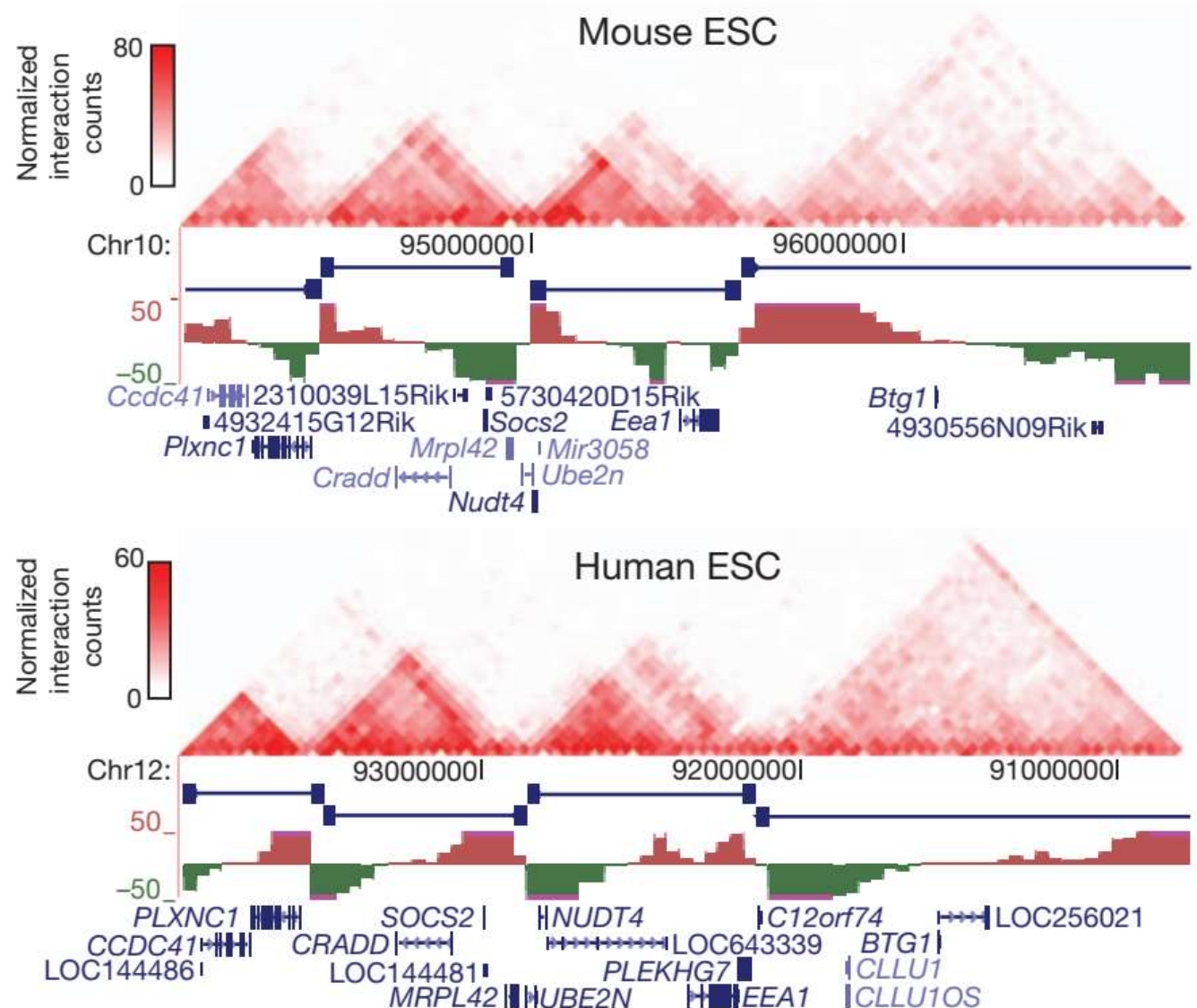
Hi-C, a method to probe the 3D genome architecture



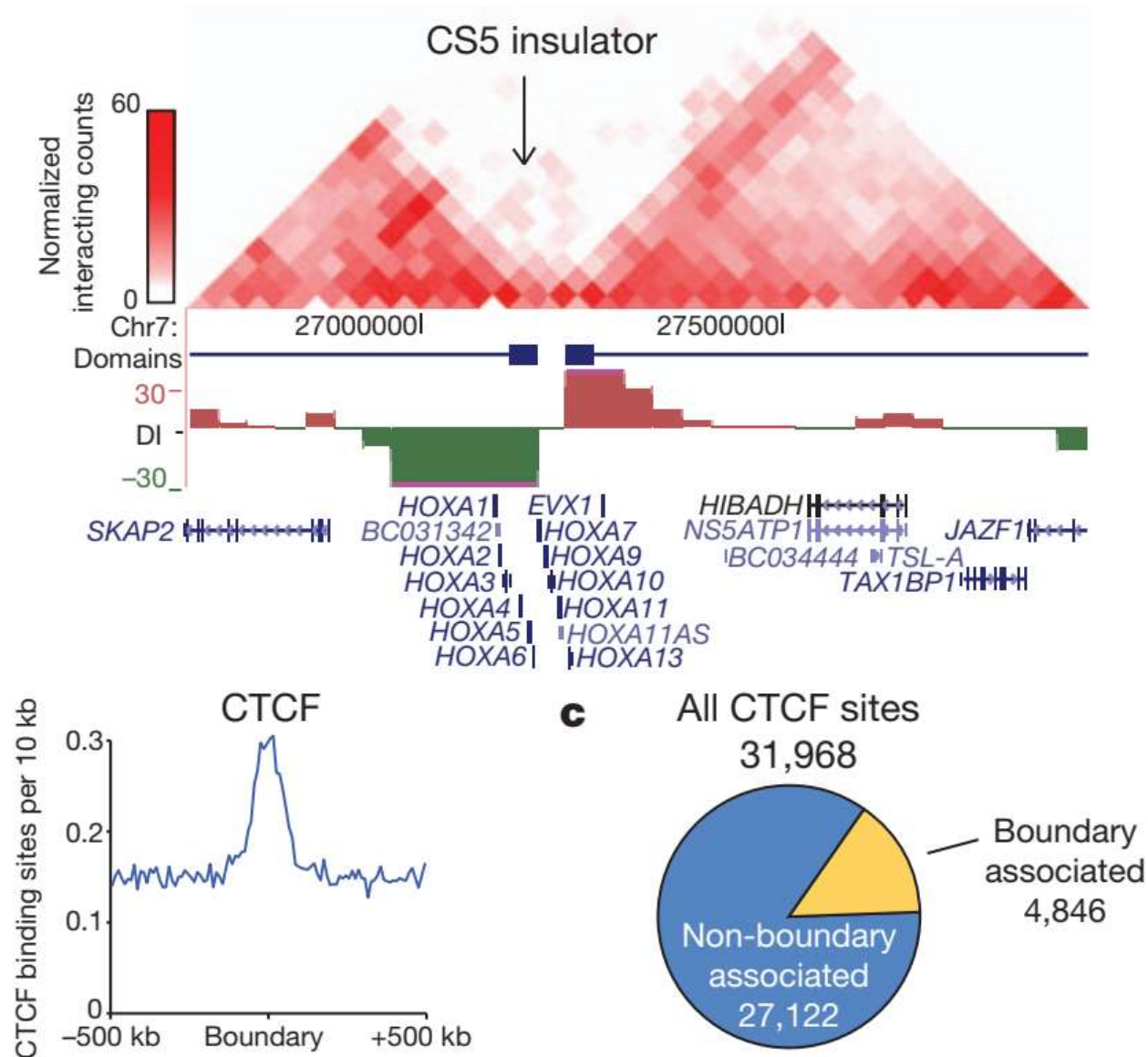
Topological domains in mammalian genomes



Dr. Bing Ren (任兵)
UCSD



Insulators separate the topological domains

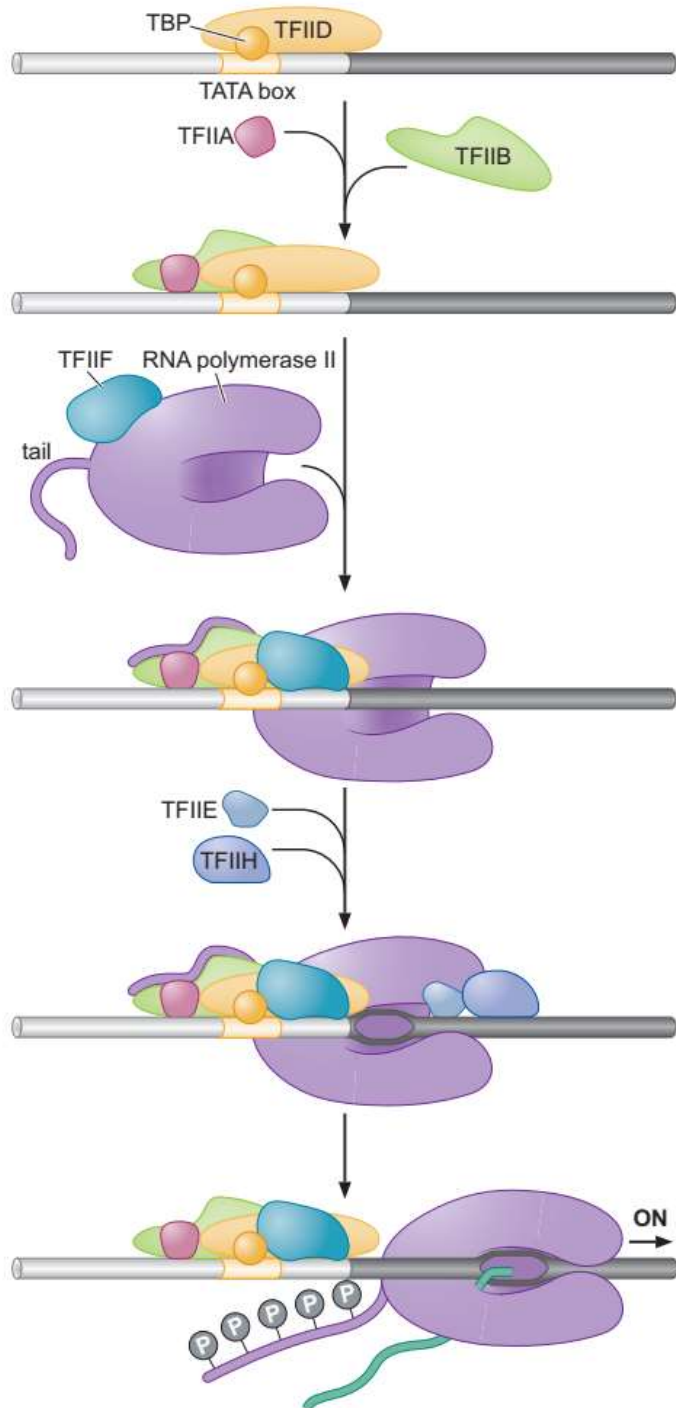


Outline

1. Eukaryotic transcription machineries
2. Function elements for eukaryotic transcription
- 3. Transcription initiation in eukaryotes**
4. Transcription-coupled RNA processing
5. Techniques to study eukaryotic transcription
6. Examples for studying eukaryotic transcription

How to initiate transcription

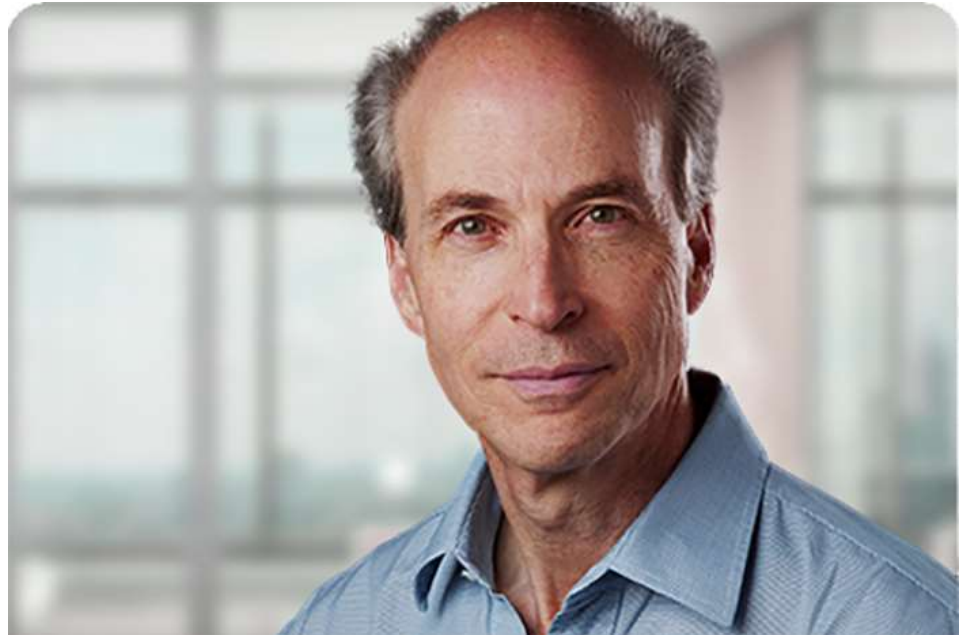
- General transcription factors (GTFs):
TFIID, TFIIA, TFIIB, TFIIIF, TFIIE and TFIIH
- **TFIID:** TBP (TATA-binding protein) & TAFs
(TBP-associated factors)



1. TATA element is recognized by TFIID (TBP + TAFs)
2. TFIIA and TFIIB are sequentially recruited to the promoter.
3. TFIIF-RNA Pol II complex is then recruited
4. TFIIE and TFIIH then assemble at the promoter to form the preinitiation complex
5. TFIIH mediates promoter melting by hydrolysis of ATP
6. Promoter escape is followed by the phosphorylation of the pol II CTD (carboxy-terminal domain)

The 2006 Nobel Price in Chemistry

“for his study in the molecular basis of eukaryotic transcription”

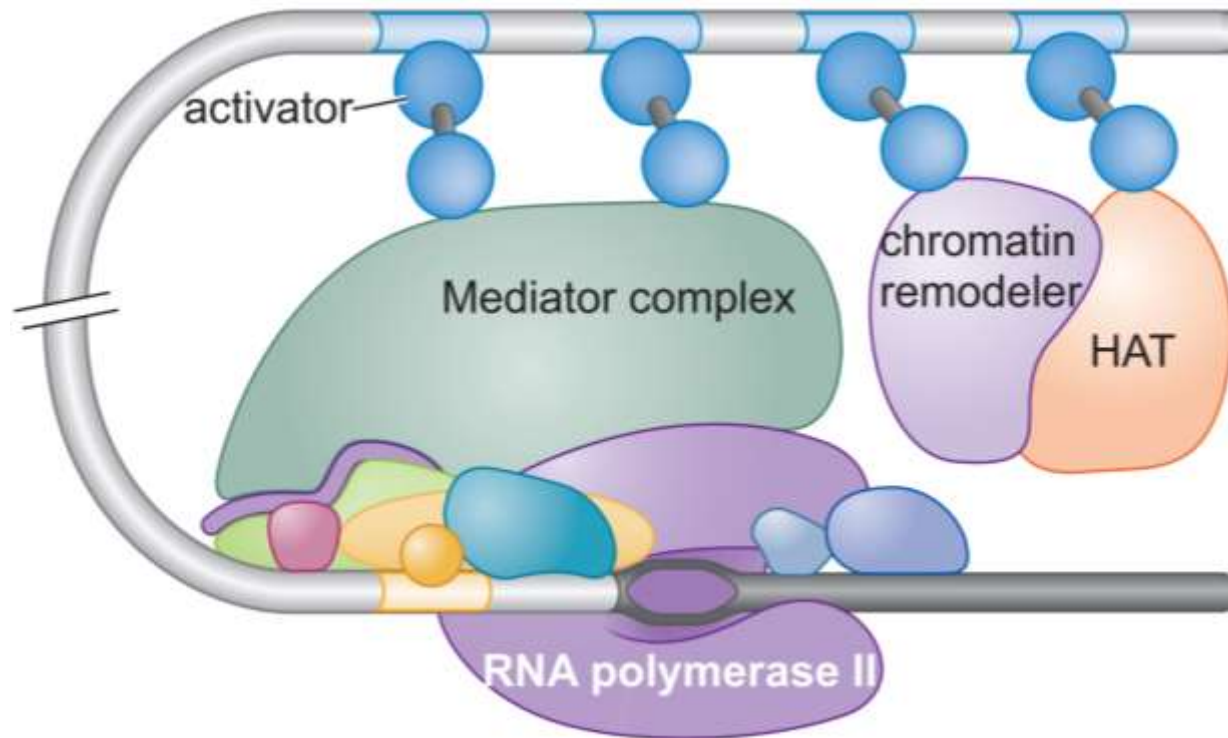


Prof. Roger D. Kornberg

“For the love of science”

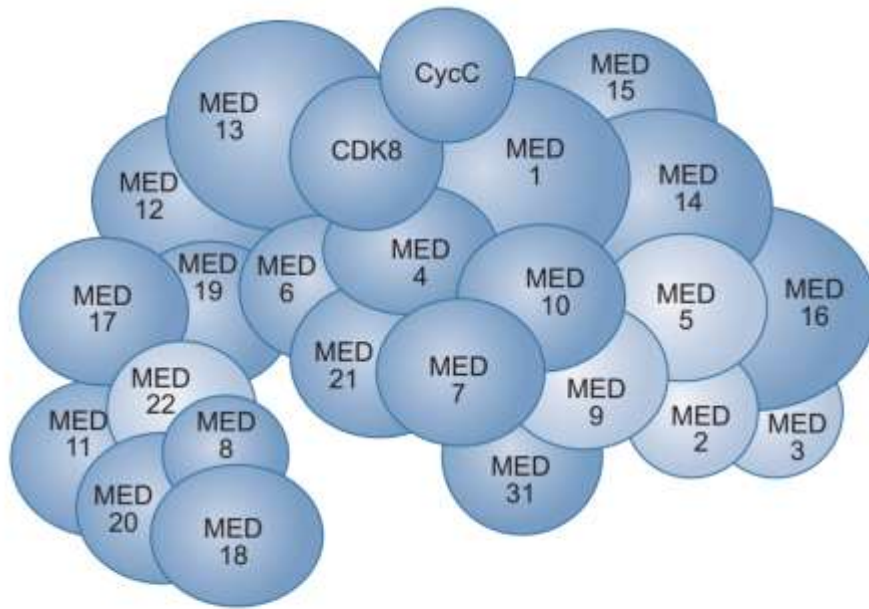
“在座的同学们都可以在自己热爱的科学领域从事研究，能够获得自己的成功，并且能够获得比我更大的发现。这一点无关乎天赋或者出身，而是一个选择和信仰的问题。”

Additional proteins are required for transcription initiation *in vivo*

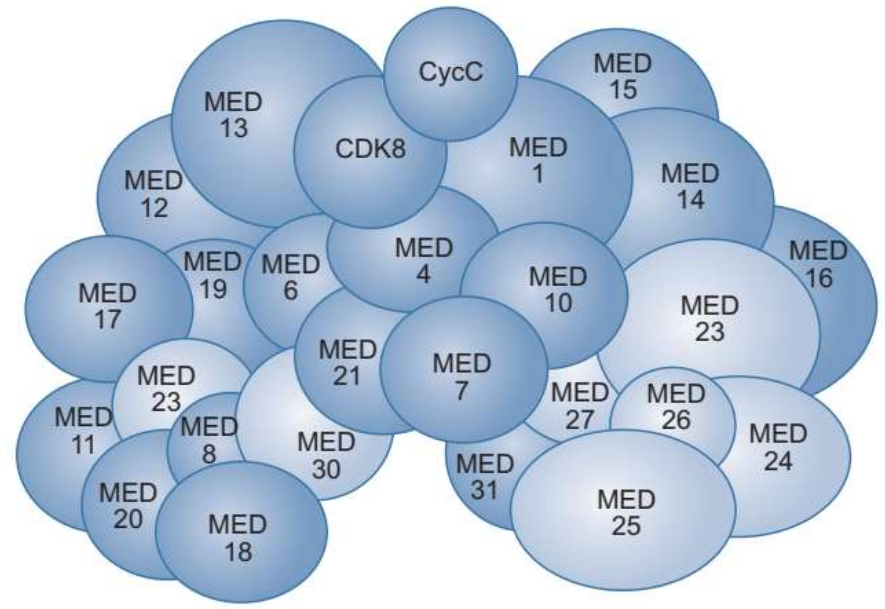


- **Activators** coordinate with chromatin remodeler and mediator to help recruit polymerase to the promoter, stabilizing its binding there
- **Mediator** is associated with the transcription machinery by interacting with the unphosphorylated Pol II CTD, providing surfaces for interaction with DNA-bound activators

Comparison of the yeast and human Mediators



yeast Mediator



human Mediator

- > 20 subunits; a similar shape
- 7 show significant sequence homology
- [Srb4/Med17](#) is essential for transcription of essentially all Pol II genes in vivo

Outline

1. Eukaryotic transcription machineries
2. Function elements for eukaryotic transcription
3. Transcription initiation in eukaryotes
- 4. Transcription-coupled RNA processing**
5. Techniques to study eukaryotic transcription
6. Examples for studying eukaryotic transcription

The composition of the Pol II CTD

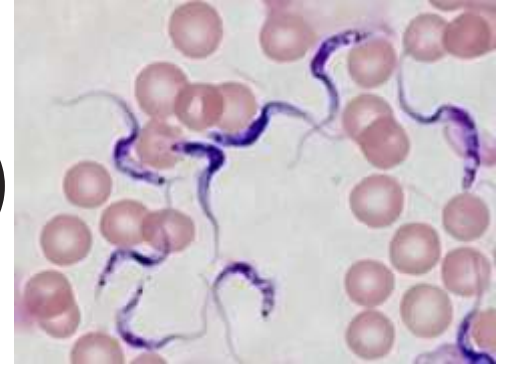
- **CTD**: the carboxy-terminal domain of the large subunit of Pol II
- The CTD contains a series of repeats of the conserved heptapeptide sequence:

Tyr1-Ser2-Pro2-Thr4-Ser5-Pro6-Ser7

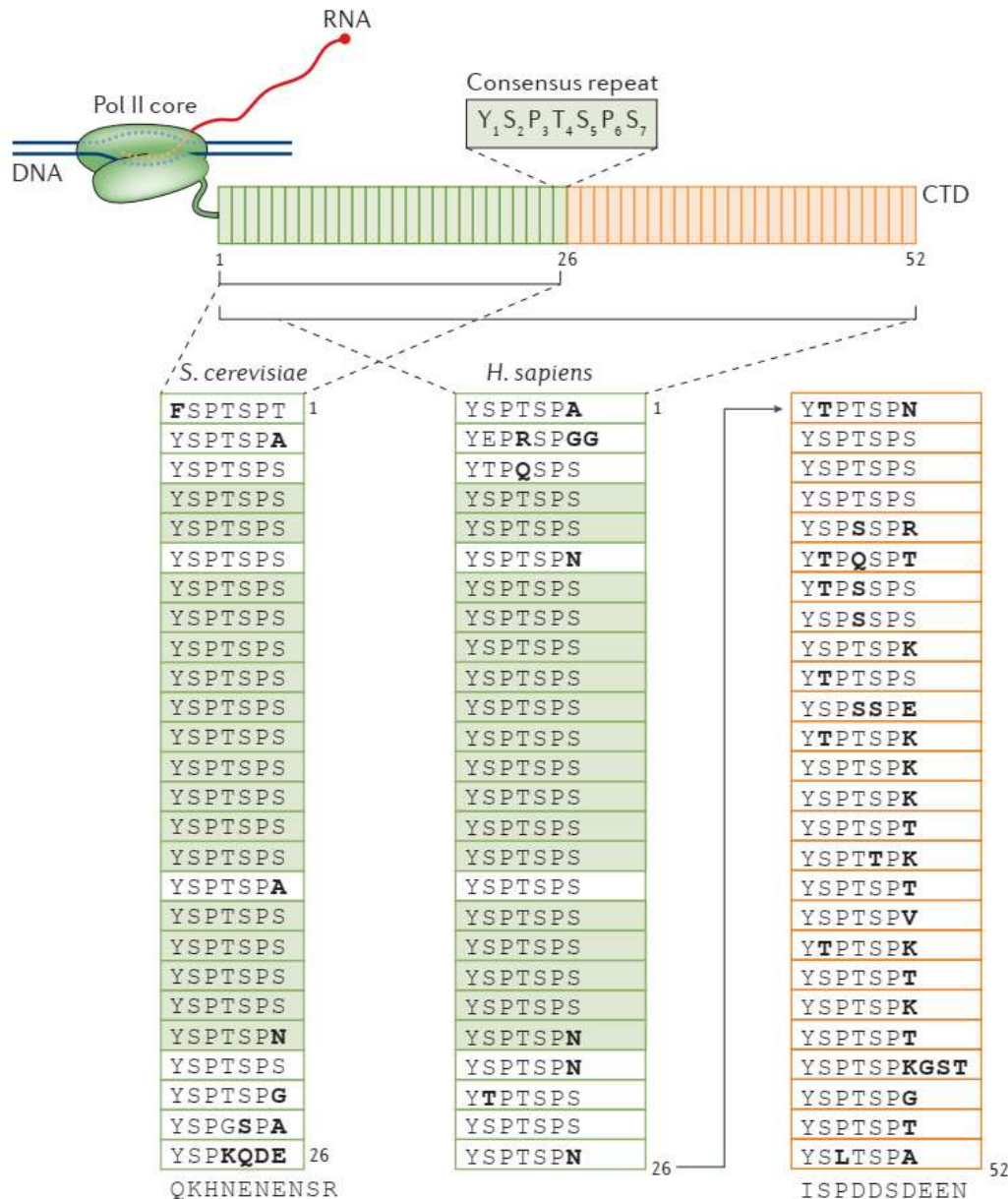
**How many heptapeptide repeats
in the CTD of different species?**

The size of the CTD in different species

- The *Trypanosoma brucei* (布氏锥虫)
Pol II lacks a CTD entirely
- **5** repeats in *Plasmodium yoelii* (约氏疟原虫)
Pol II CTD
- **26** in the yeast Pol II CTD
- **32** in the worm *Caenorhabditis elegans*
- **45** in the fly *Drosophila*
- **52** in humans



The conservation of the CTD



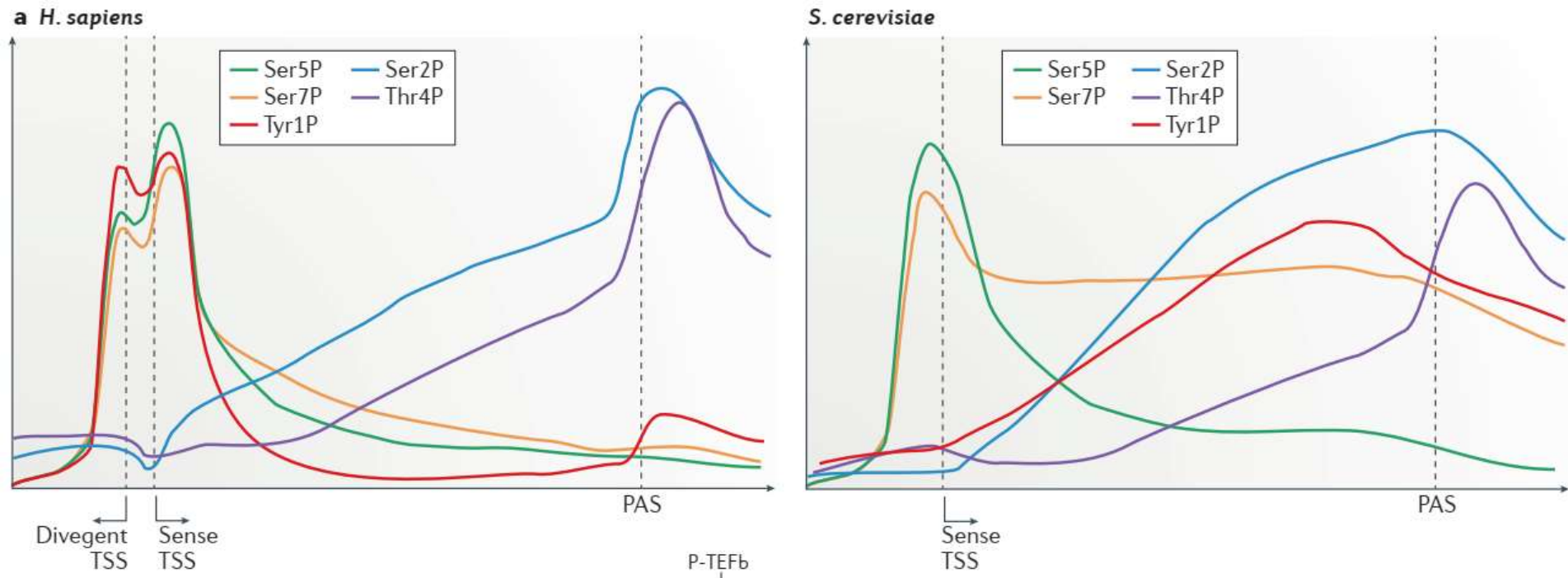
Is the entire CTD
code essential?

no

Phosphorylation of the CTD

Post-translational modification	Position in the CTD	Organisms
Ser5 phosphorylation	Multiple repeats	<ul style="list-style-type: none"> • <i>Saccharomyces cerevisiae</i> • <i>Schizosaccharomyces pombe</i> • <i>Homo sapiens</i>
Ser2 phosphorylation	Multiple repeats	<ul style="list-style-type: none"> • <i>S. cerevisiae</i> • <i>S. pombe</i> • <i>H. sapiens</i>
Ser7 phosphorylation	Multiple repeats	<ul style="list-style-type: none"> • <i>S. cerevisiae</i> • <i>S. pombe</i> • <i>H. sapiens</i>
Thr4 phosphorylation	Multiple repeats	<ul style="list-style-type: none"> • <i>S. cerevisiae</i> • <i>S. pombe</i> • <i>Gallus gallus</i> • <i>H. sapiens</i>
Tyr1 phosphorylation	Multiple repeats	<ul style="list-style-type: none"> • <i>S. cerevisiae</i> • <i>S. pombe</i> • <i>G. gallus</i> • <i>H. sapiens</i>

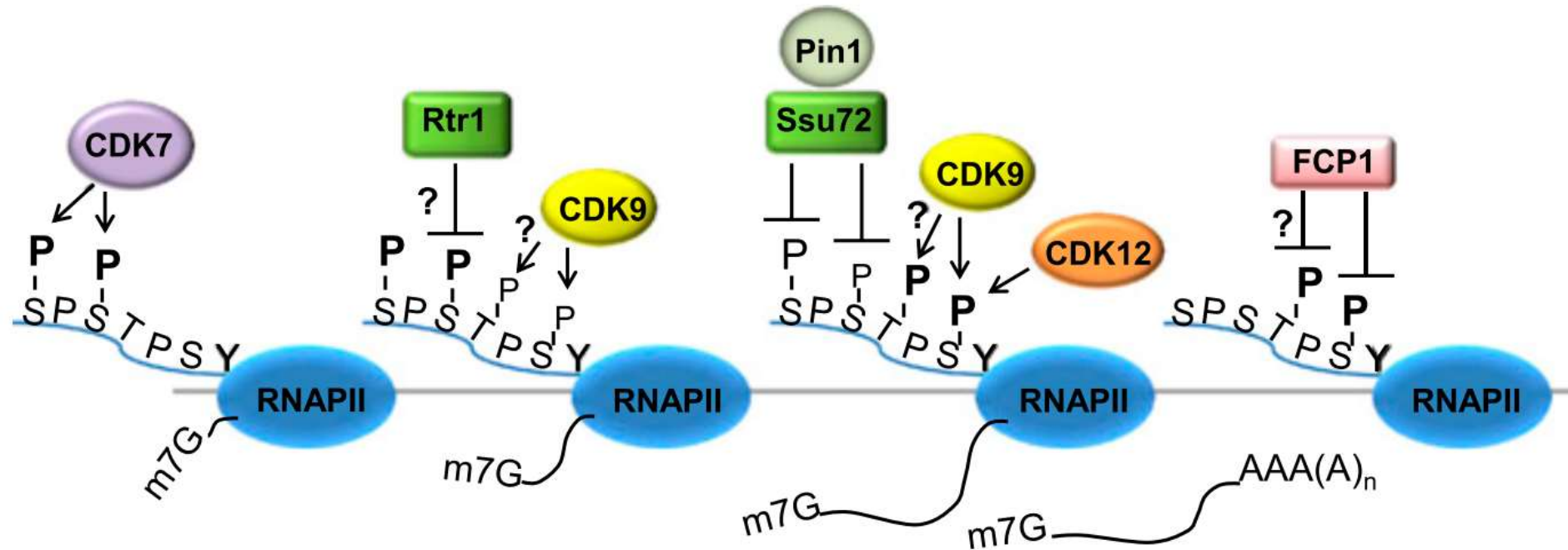
Distribution of the phosphorylated CTD at genes



Harlen and Churchman 2017 NAT REV MOL CELL BIO

CTD Phosphorylation dynamics

Initiation → Elongation → Termination

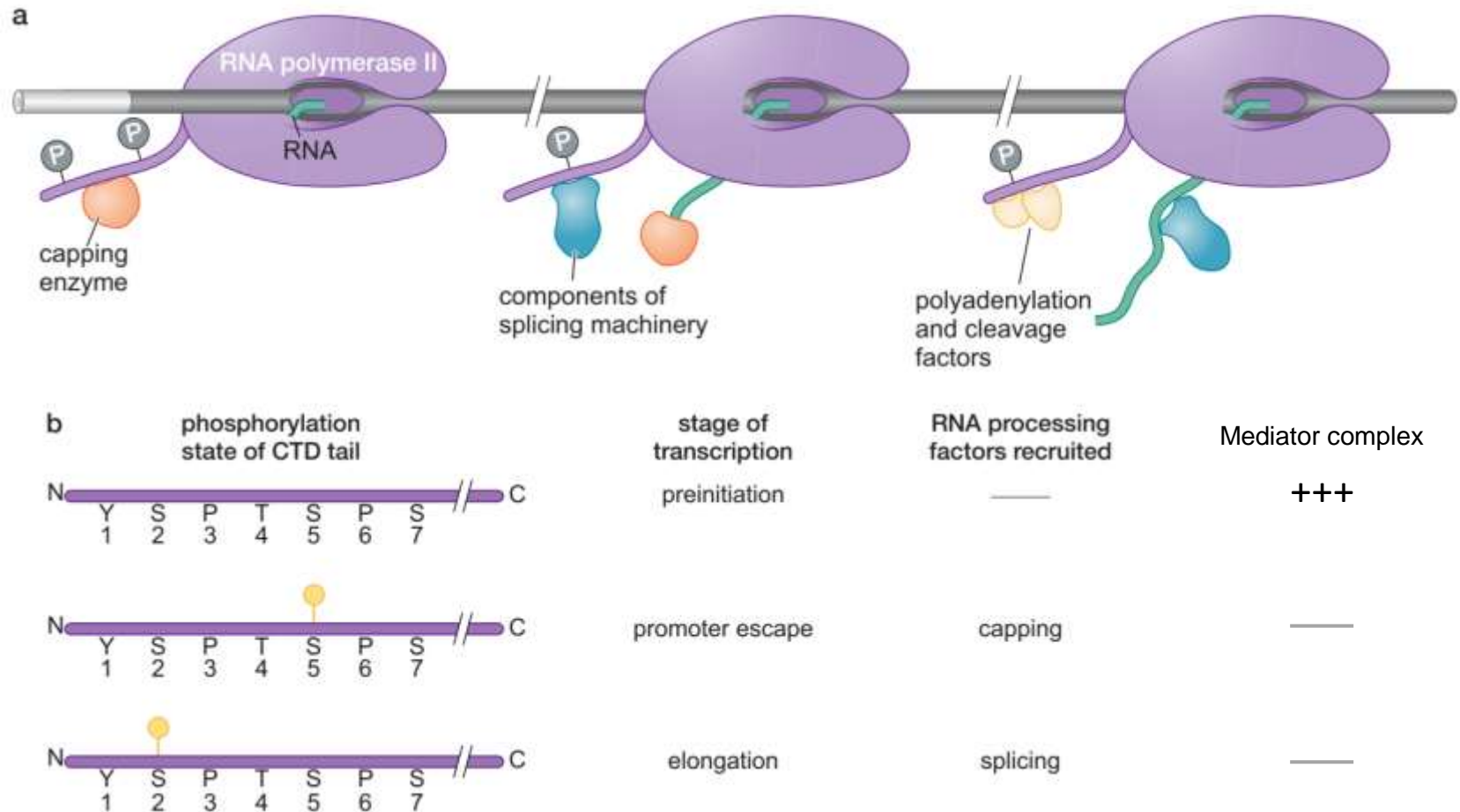


Hisn and Manley 2012 GENE DEV

Transcription regulation by the CTD code

Post-translational modification	Associated process or processes
Ser5 phosphorylation	Transcription initiation, mRNA capping and splicing, non-coding RNA transcription termination and chromatin modification
Ser2 phosphorylation	Transcription elongation, promoter-proximal pause and release, splicing, transcription termination and DNA topology
Ser7 phosphorylation	snRNA expression, interaction with the Integrator complex and P-TEFb recognition
Thr4 phosphorylation	Transcription elongation and termination, post-transcriptional splicing, processing of histone mRNA and chromatin remodelling
Tyr1 phosphorylation	Inhibition of recruitment of transcription termination factors, CTD stability, antisense and enhancer transcription

CTD phosphorylation-mediated co-transcriptional RNA processing



- Phosphorylation of the CTD leads to an exchange of initiation factors for those factors required for elongation and RNA processing.

Transcription elongation factors

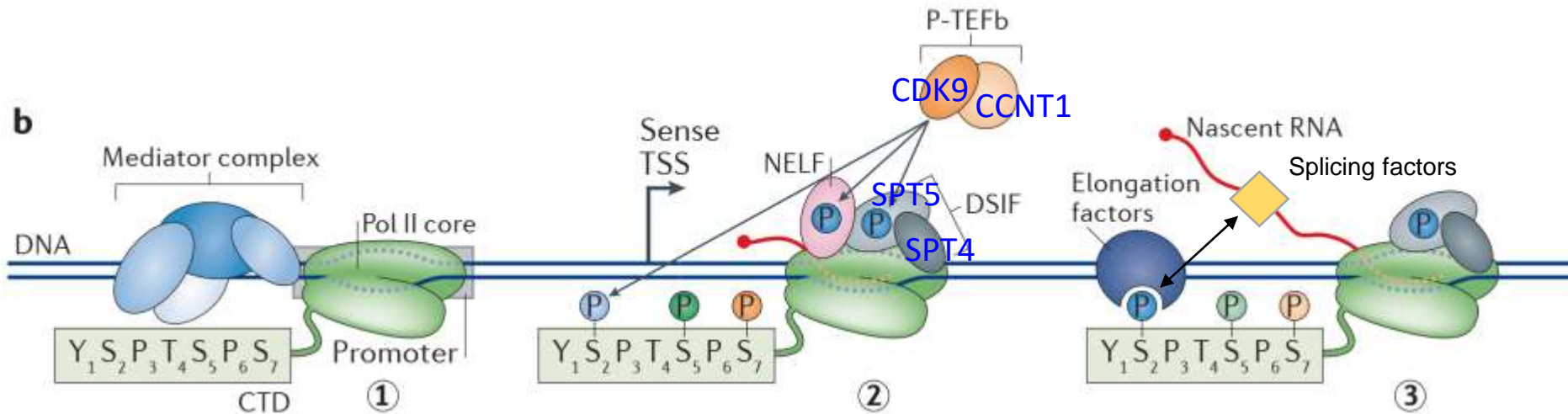
Initiation



Promoter escape

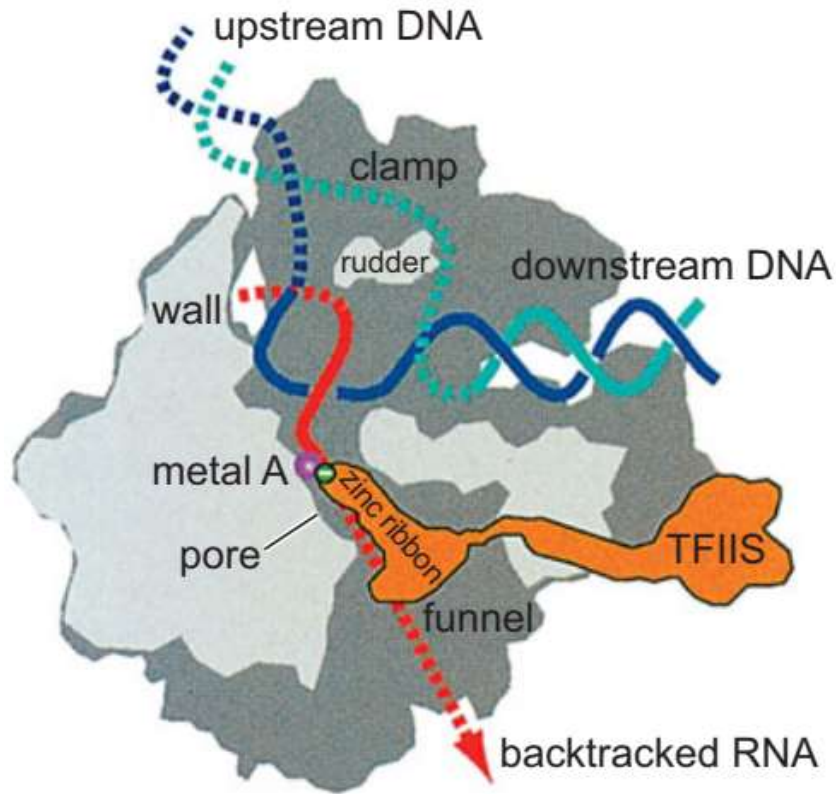


Elongation

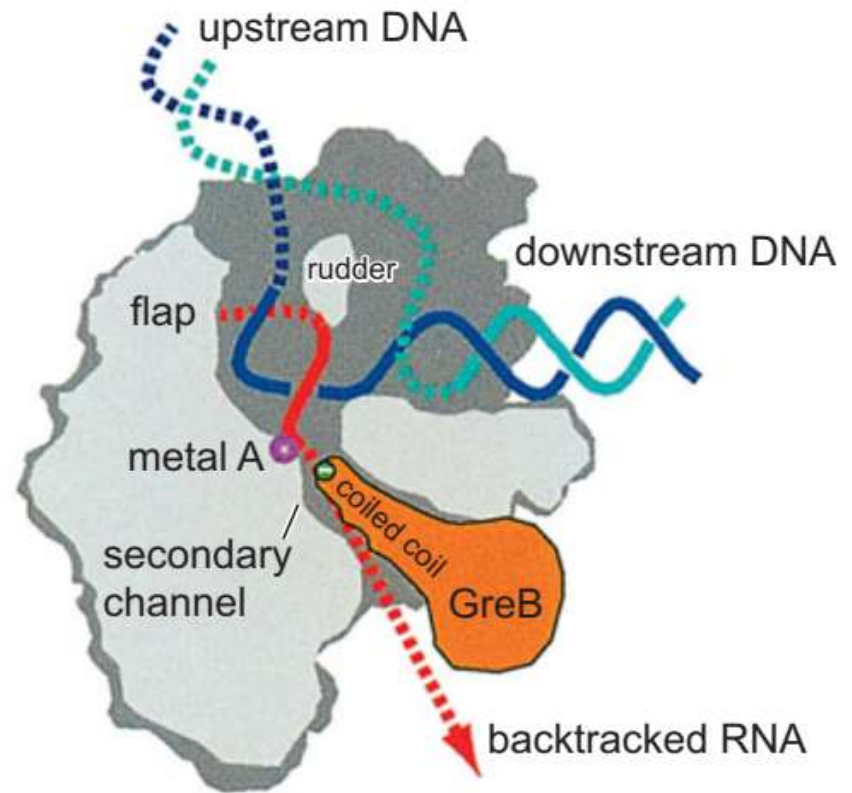


- **ELL family proteins** suppress transient pausing by the Pol II.
- **TFIIS** stimulates the overall rate of elongation by limiting the time of polymerase pausing at some sequences that would otherwise tend to slow the enzyme's progress.
- **TFIIS** contributes to **proofreading by polymerase** by stimulating its inherent RNase activity.

TFIIS and GreB in proofreading by Polymerase

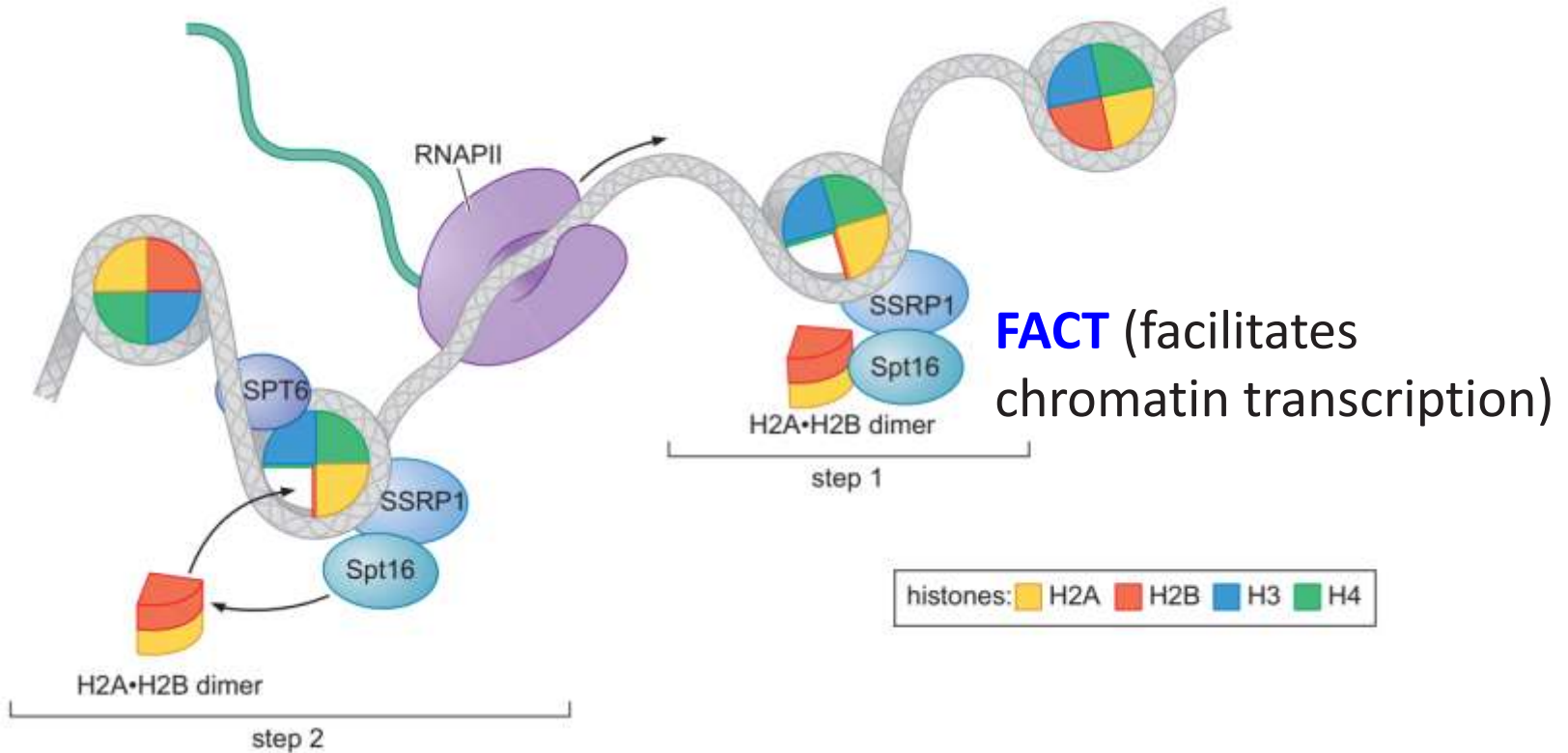


yeast RNA polymerase II



bacterial RNA polymerase

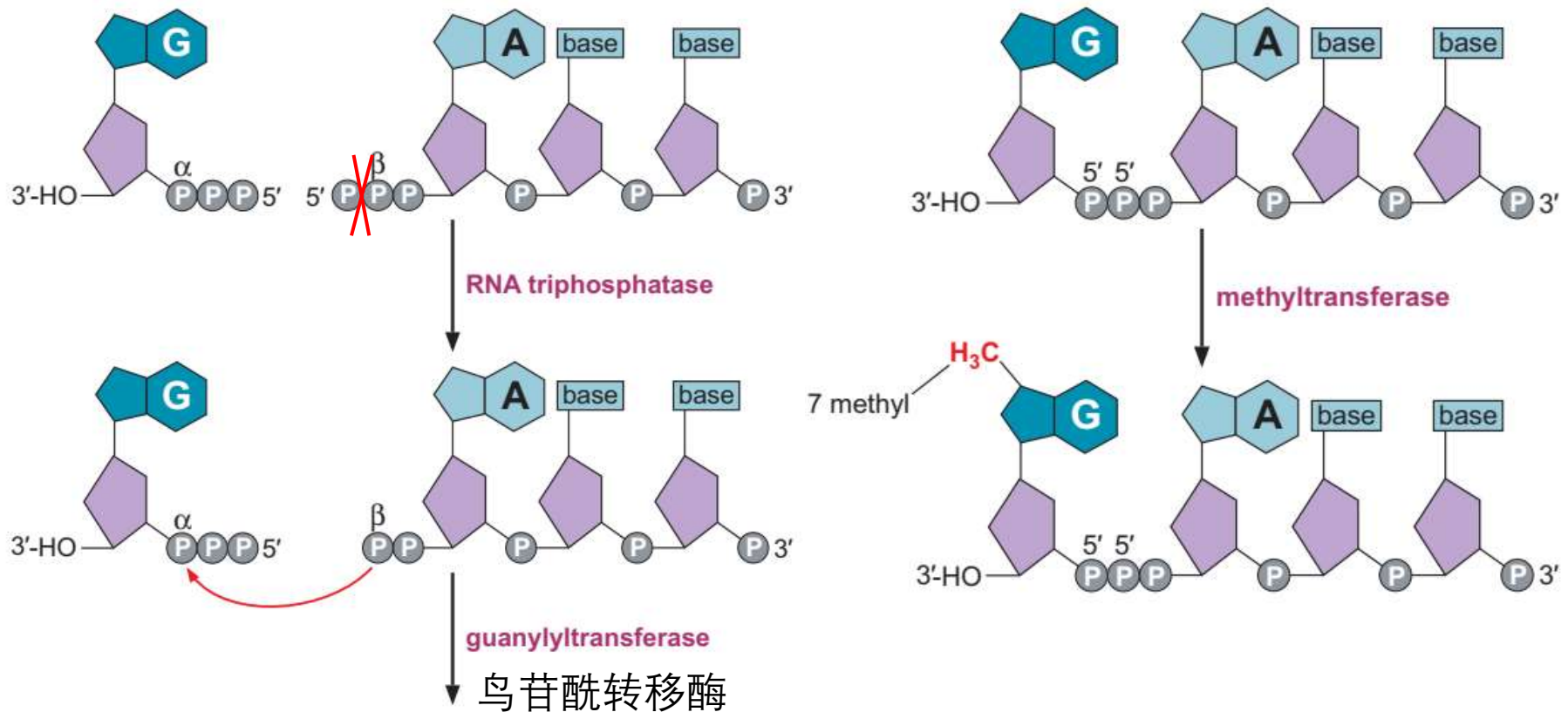
FACT-aided elongation through nucleosomes



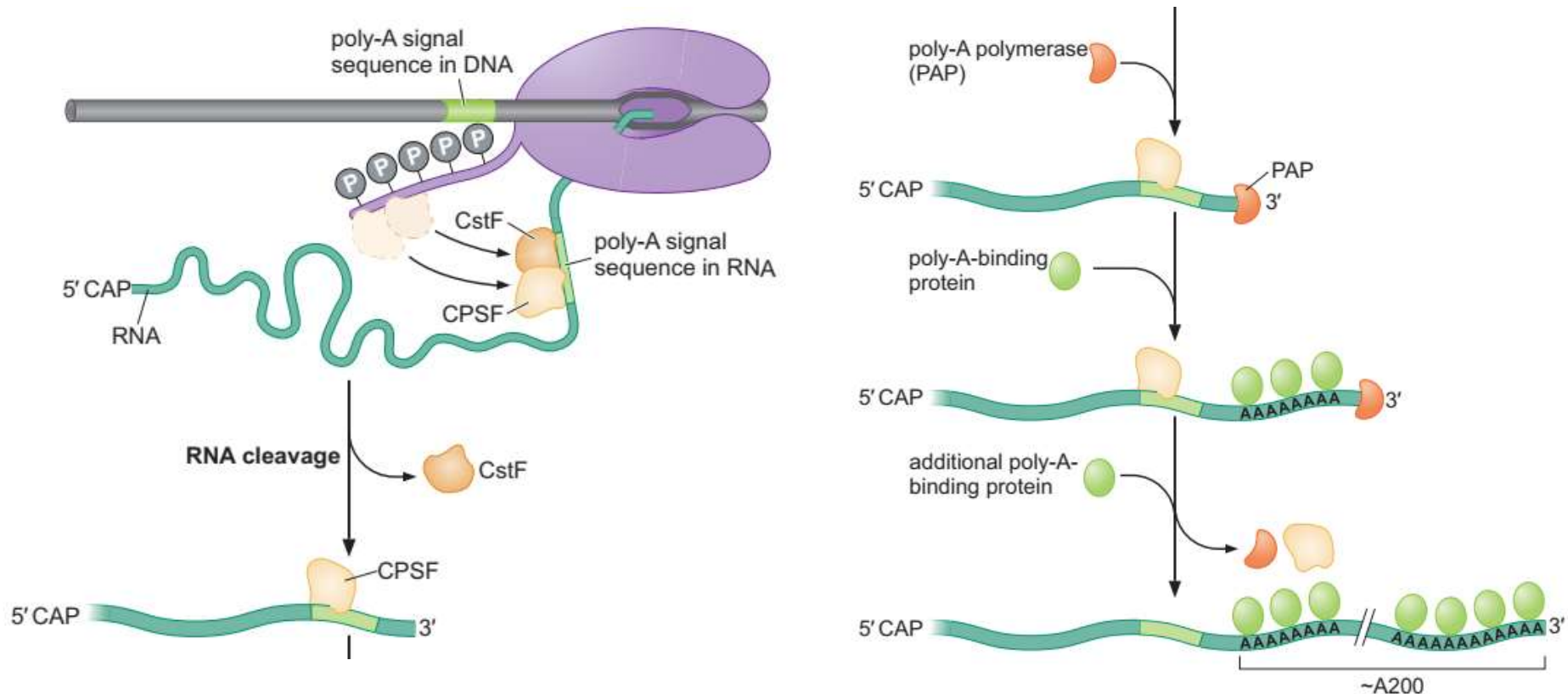
- Ahead of a transcribing Pol II, **FACT removes one H2A.H2B dimer**, and **restores that dimer to the histone hexamer** immediately behind the processing polymerase, allowing polymerase to elongate and at the same time maintains the integrity of the chromatin.

5' capping: the first RNA processing event

The addition of a modified guanine base to the 5' end of the RNA

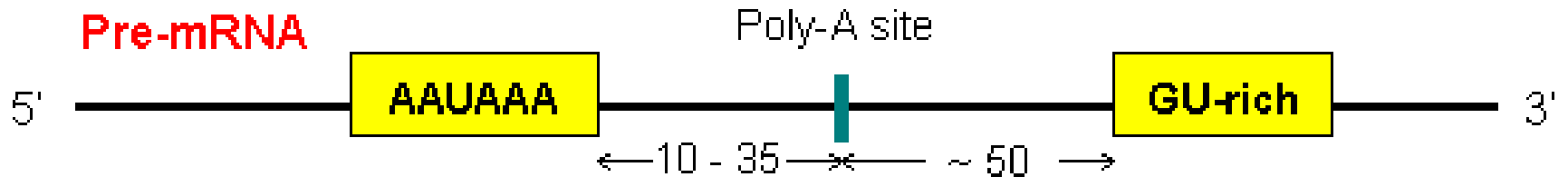


Co-transcriptional RNA cleavage and polyadenylation



- CPSF (cleavage and polyadenylation specificity factor)
- CSTF (cleavage stimulation factor)

Poly-A signals



Cleave and add Poly-A tail



Human

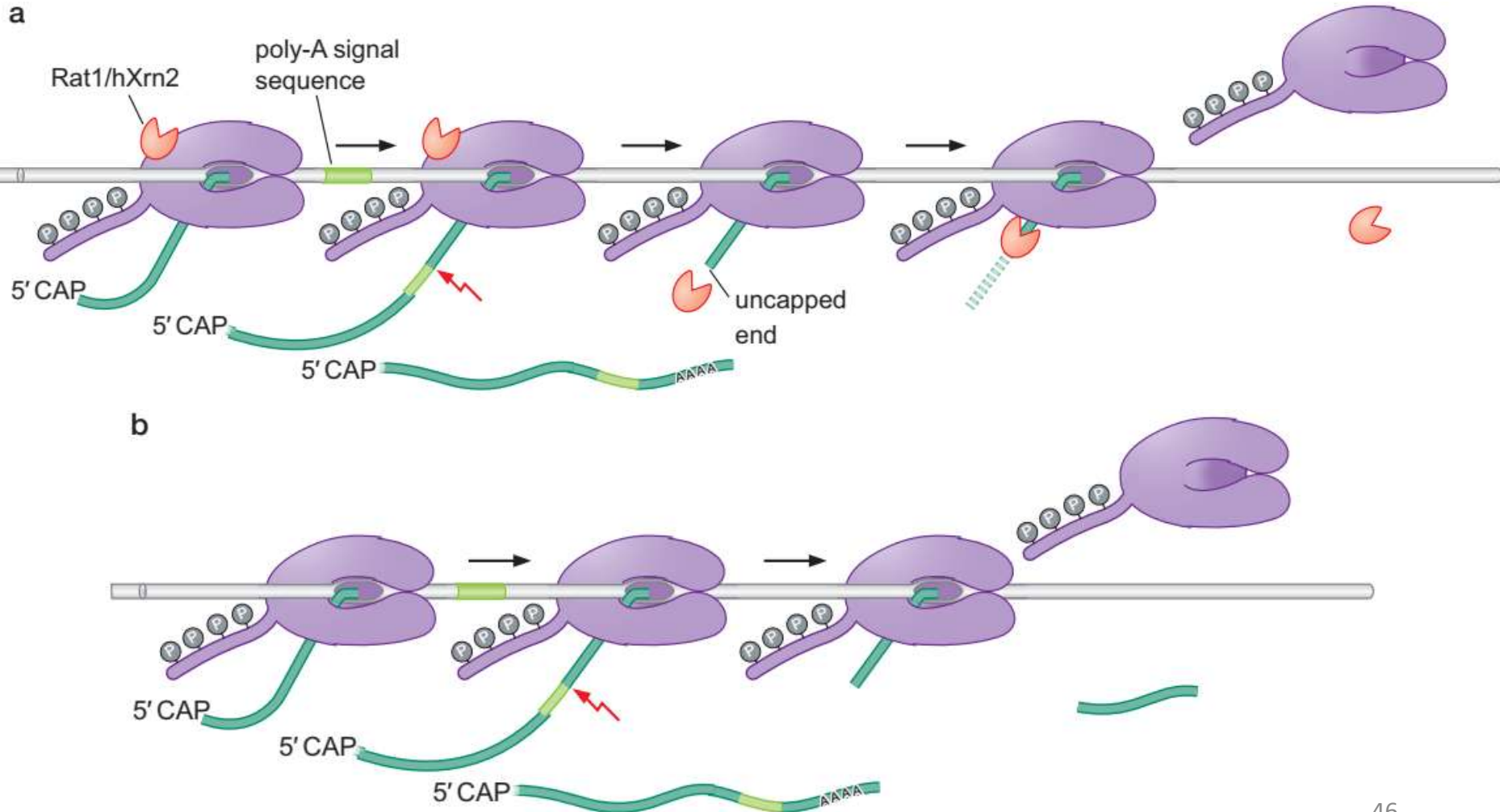


Drosophila

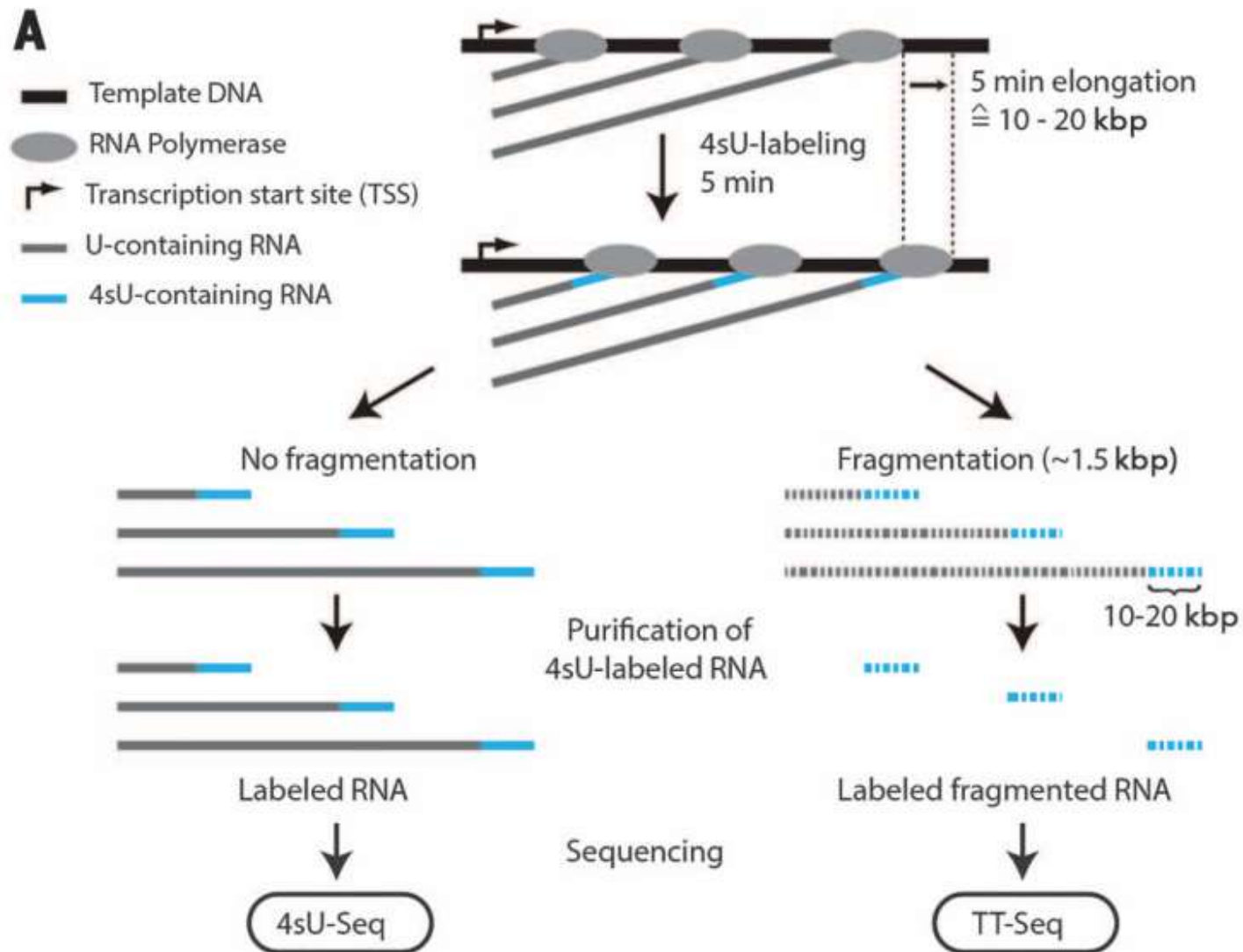


Torpedo & allosteric models of termination

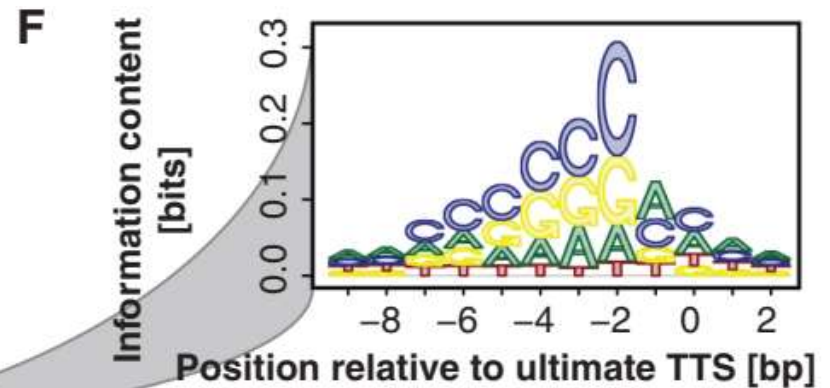
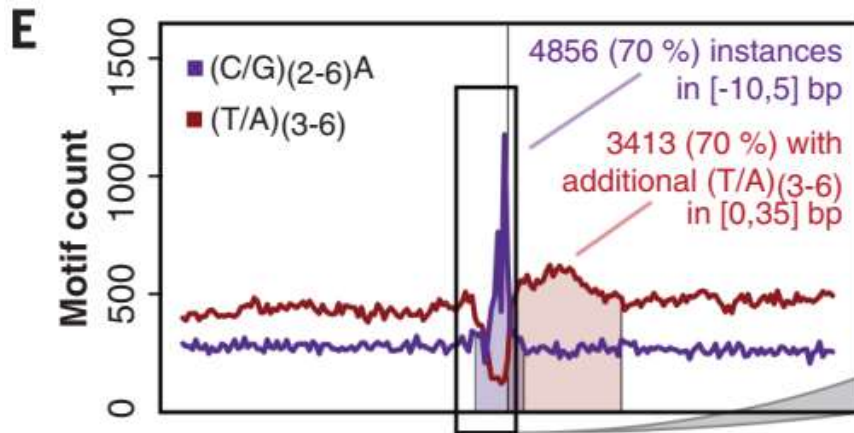
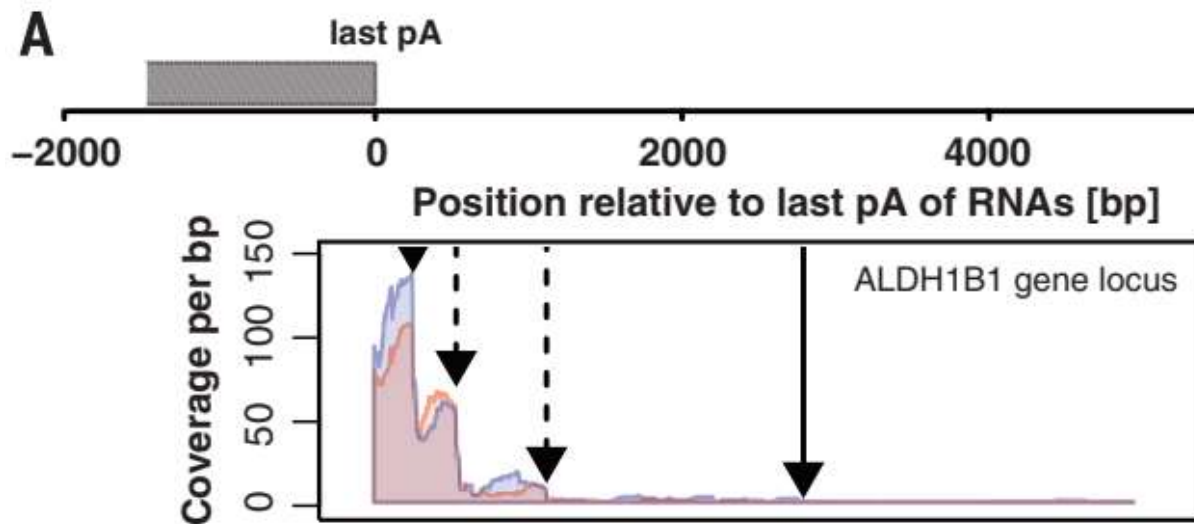
RNA cleavage and then polyadenylation trigger termination of transcription, although the exact mechanism is still unclear.



Transient transcriptome sequencing (TT-seq)



Transcription termination sites



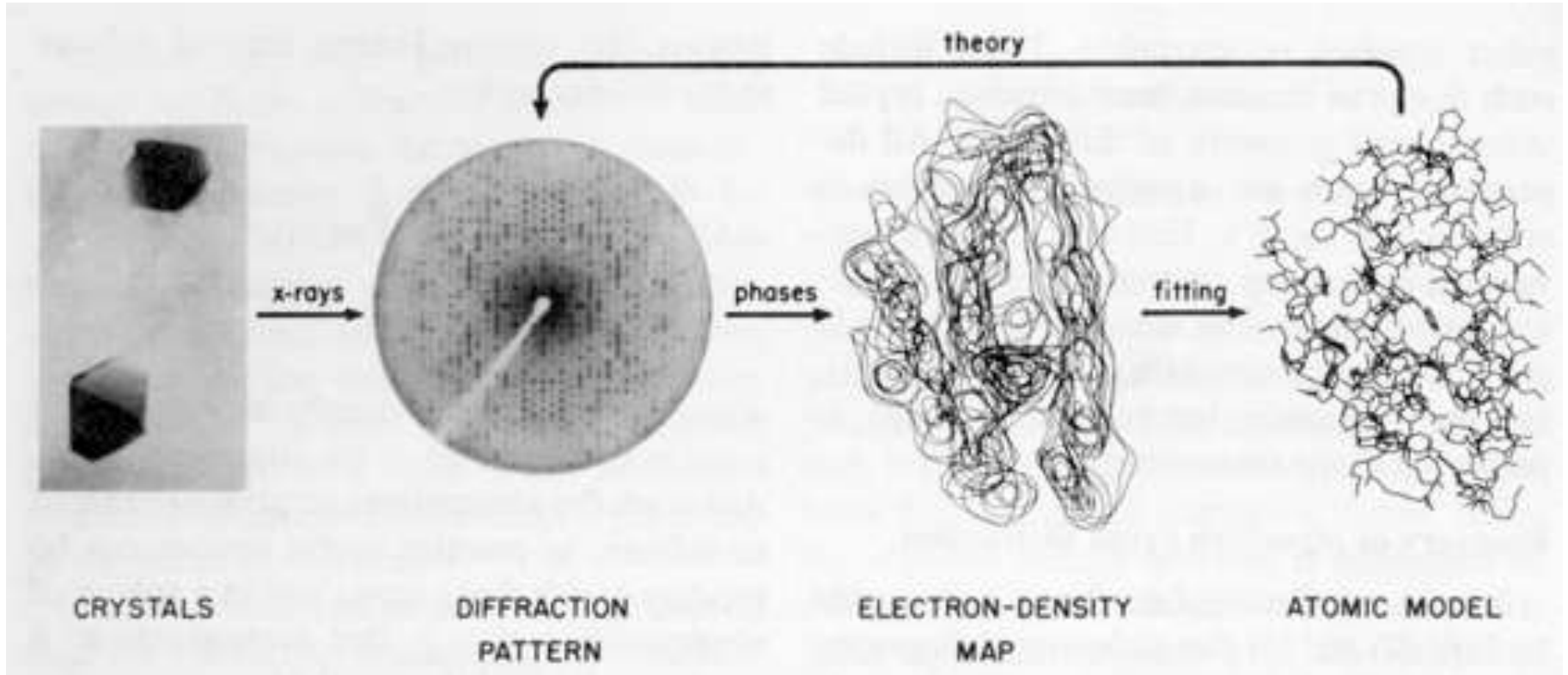
Outline

1. Eukaryotic transcription machineries
2. Function elements for eukaryotic transcription
3. Transcription initiation in eukaryotes
4. Transcription-coupled RNA processing
- 5. Techniques to study eukaryotic transcription**
6. Examples for studying eukaryotic transcription

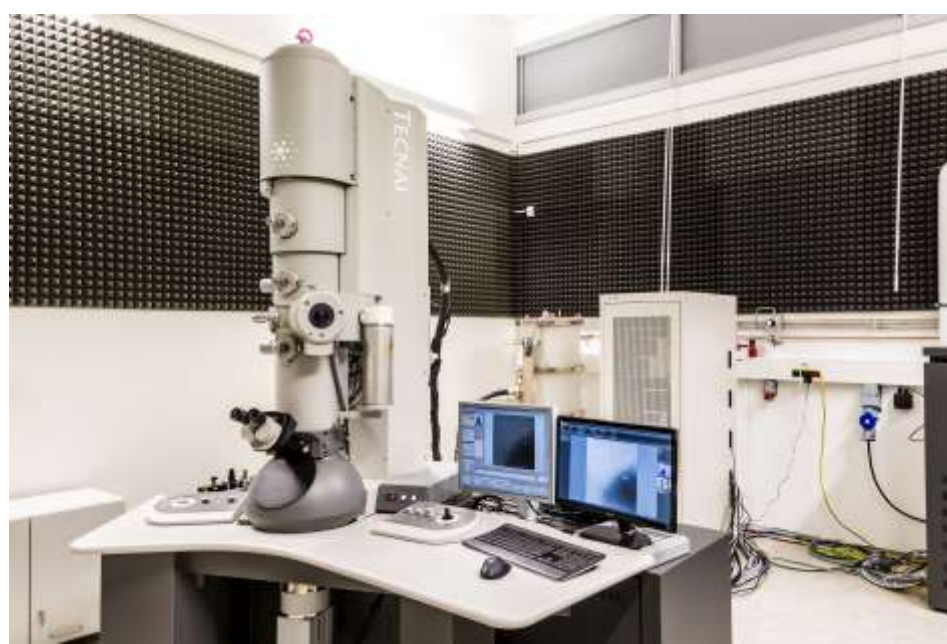
Methodologies to study eukaryotic transcription

- Biochemical analysis, eg. In vitro binding assays, such as EMSA (gel shift)
- Structure analysis, eg. X-ray crystallography or Cryo-Electron Microscopy
- High-throughput sequencing-bases techniques, eg. ChIP-seq, ChIP-exo, Cut&Run, Cut&Tag, GRO-seq, PRO-seq, TT-seq

X-ray crystallography



Cryo-Electron Microscopy



相比于晶体学手段，冷冻电镜技术对样品量和均一度的要求大幅降低；可利用同一套电镜数据，观察到同一蛋白在不同构象状态下的结构，以及不同蛋白的高分辨率结构。

目前，利用单颗粒技术解析的生物大分子结构的整体分辨率可达到3-4埃。

The Nobel Prize in Chemistry 2017



© Nobel Media AB. Photo: A. Mahmoud
Jacques Dubochet

Prize share: 1/3



© Nobel Media AB. Photo: A. Mahmoud
Joachim Frank

Prize share: 1/3

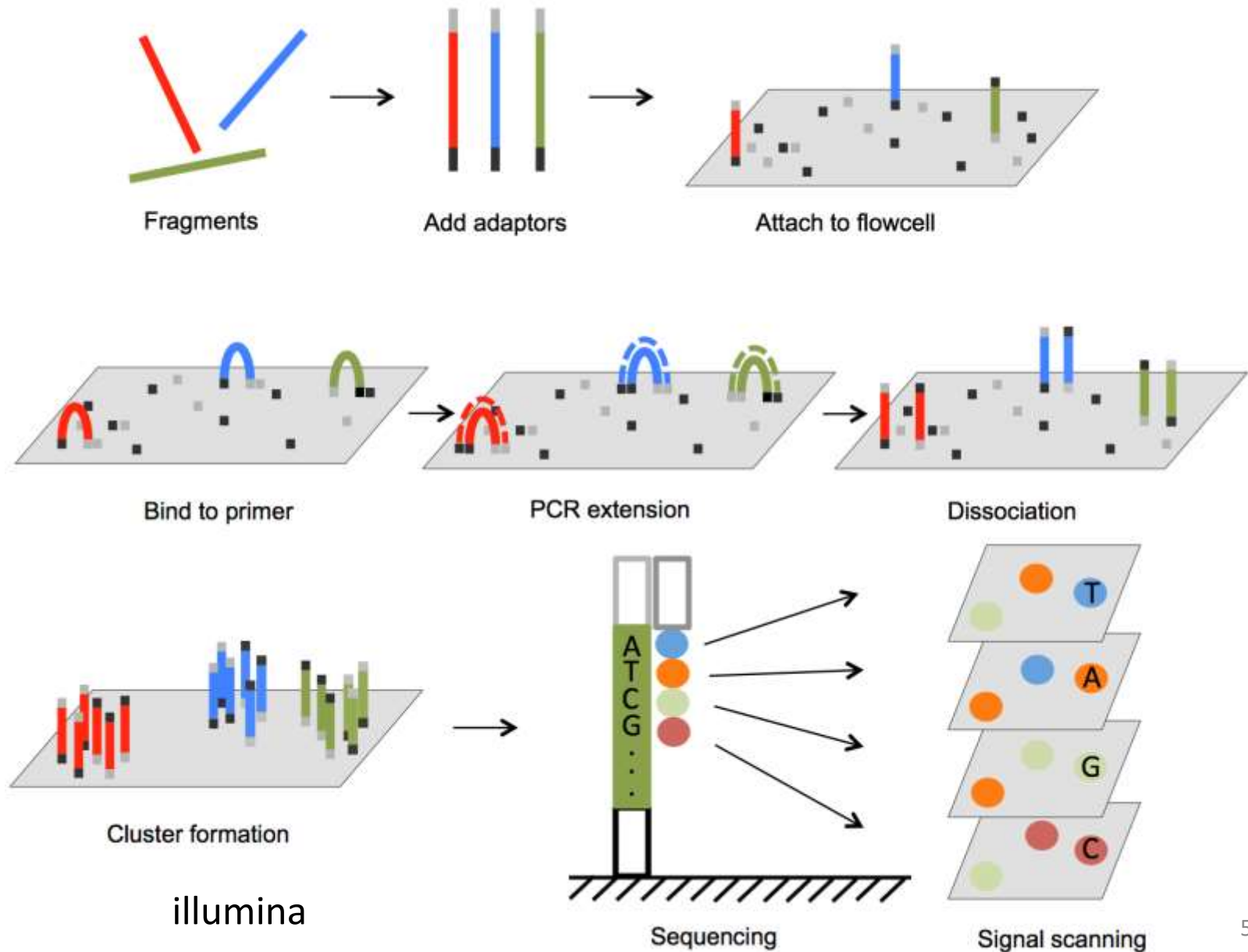


© Nobel Media AB. Photo: A. Mahmoud
Richard Henderson

Prize share: 1/3

The Nobel Prize in Chemistry 2017 was awarded jointly to Jacques Dubochet, Joachim Frank and Richard Henderson "for developing cryo-electron microscopy for the high-resolution structure determination of biomolecules in solution."

High-throughput sequencing

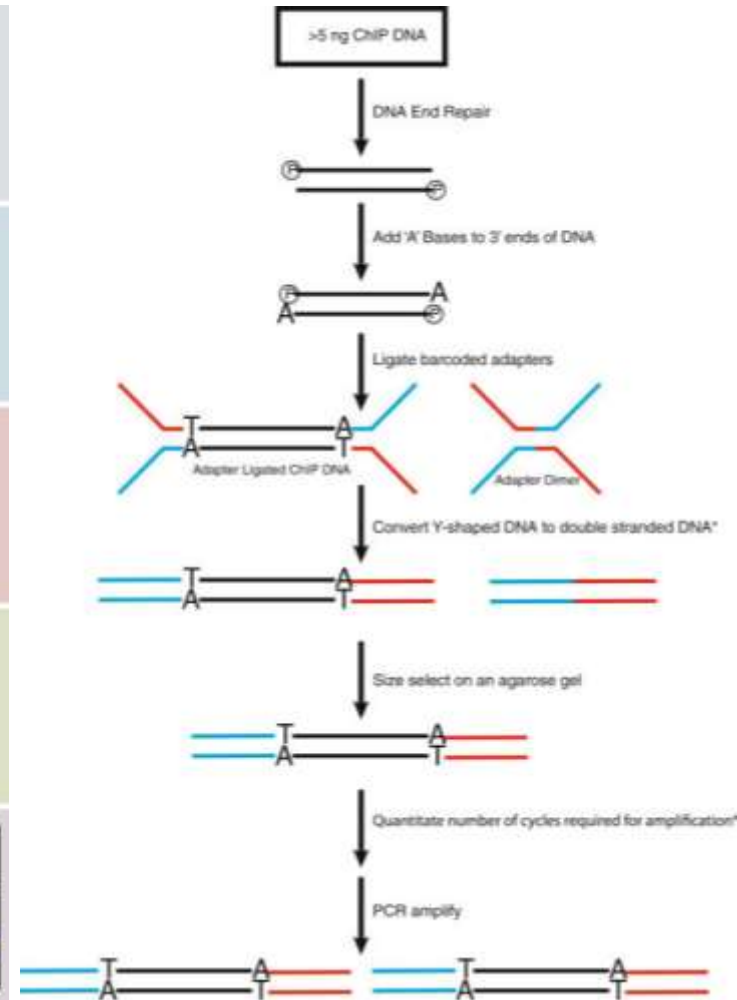
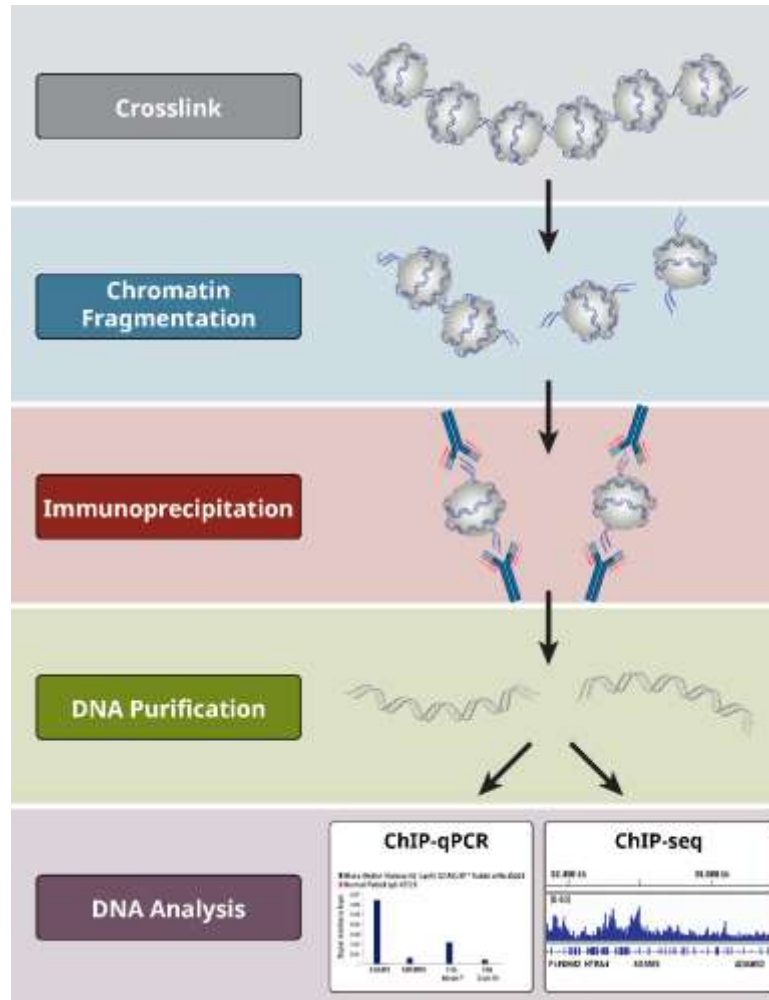


Illumina NGS sequencing principle

Profiling chromatin binding by ChIP-seq

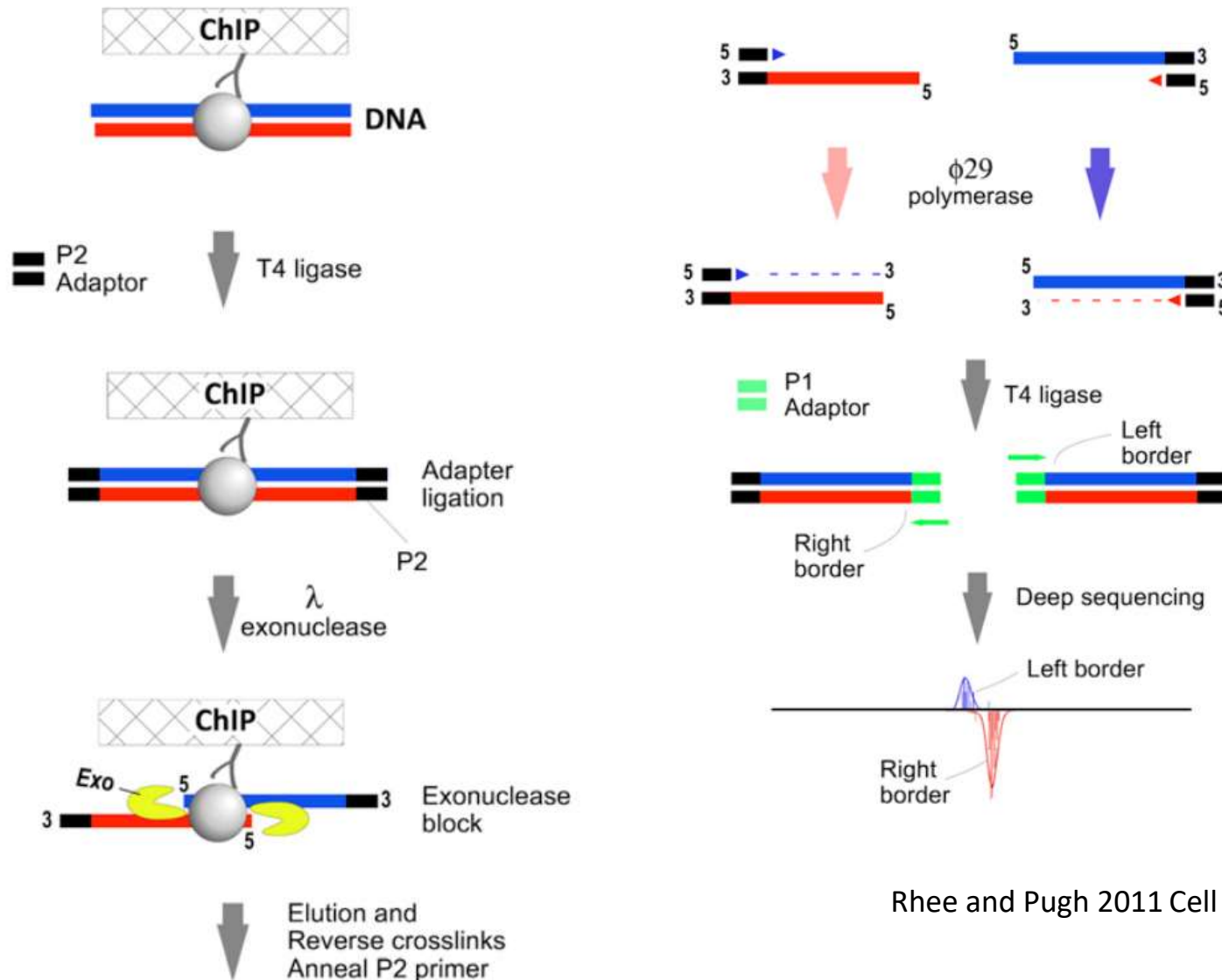


Prof. Keji Zhao

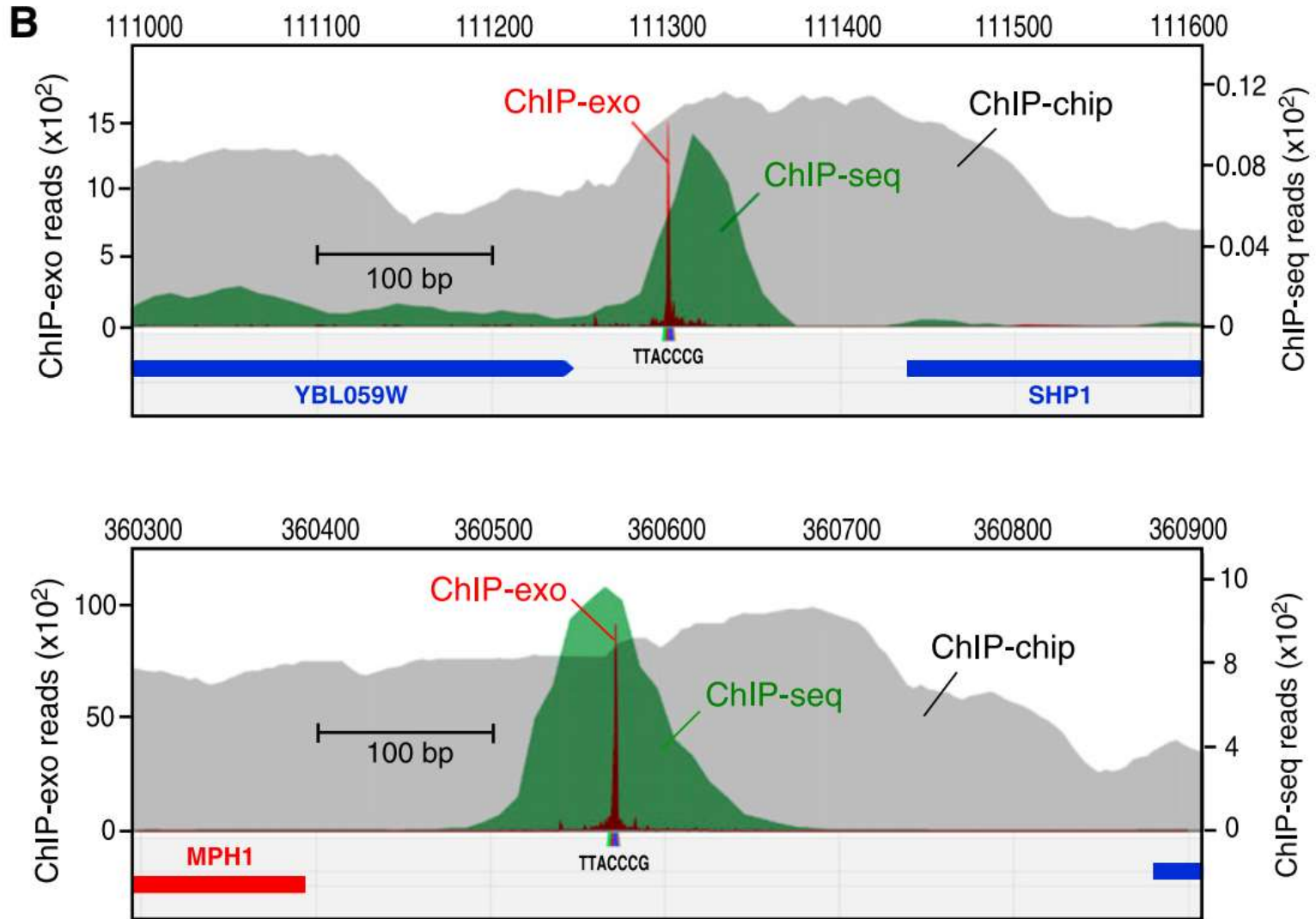


Barski et al. 2007 Cell

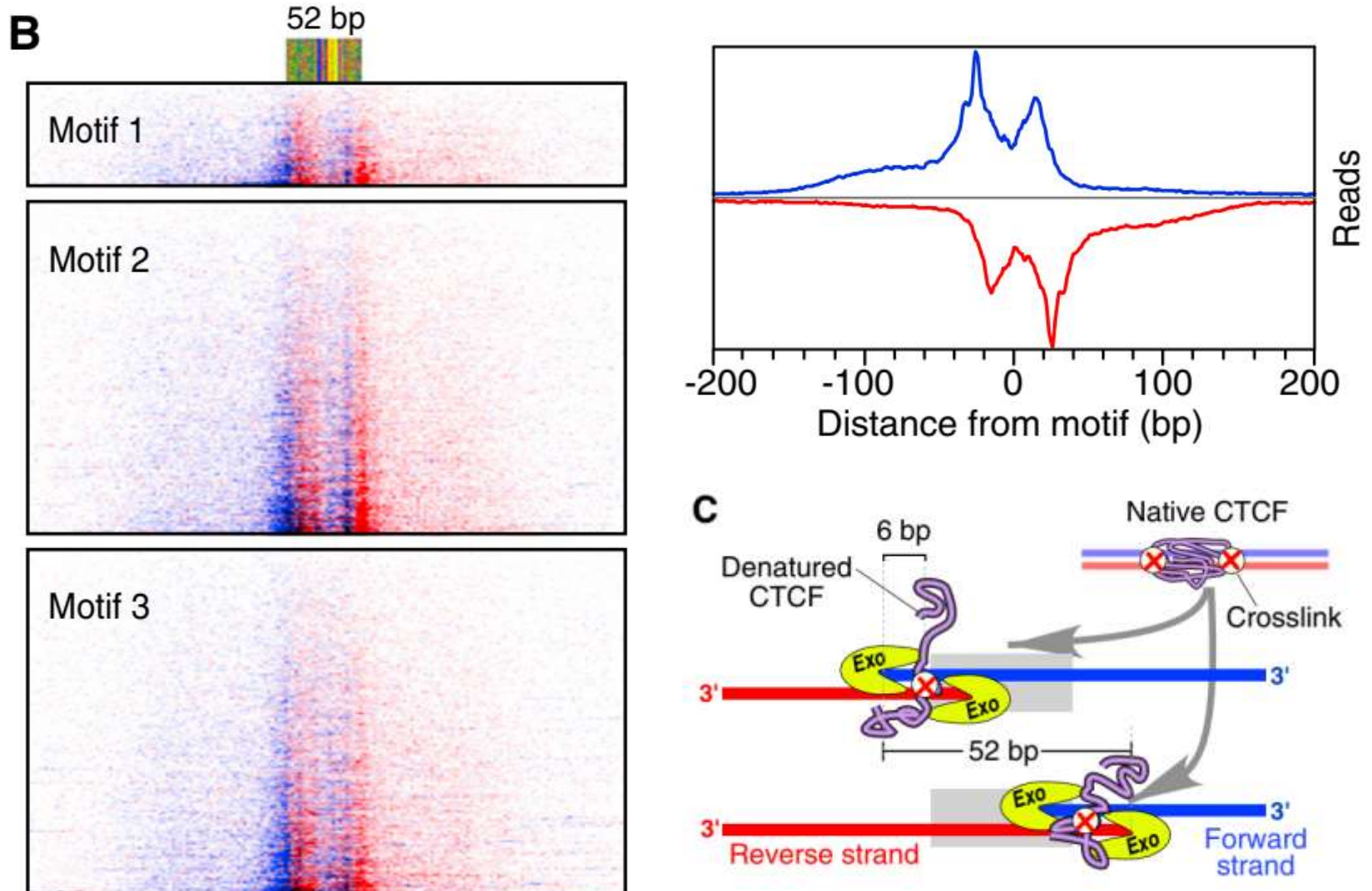
Detect protein-DNA interactions at single nucleotide resolution by ChIP-exo



Detect protein-DNA interactions at single nucleotide resolution by ChIP-exo



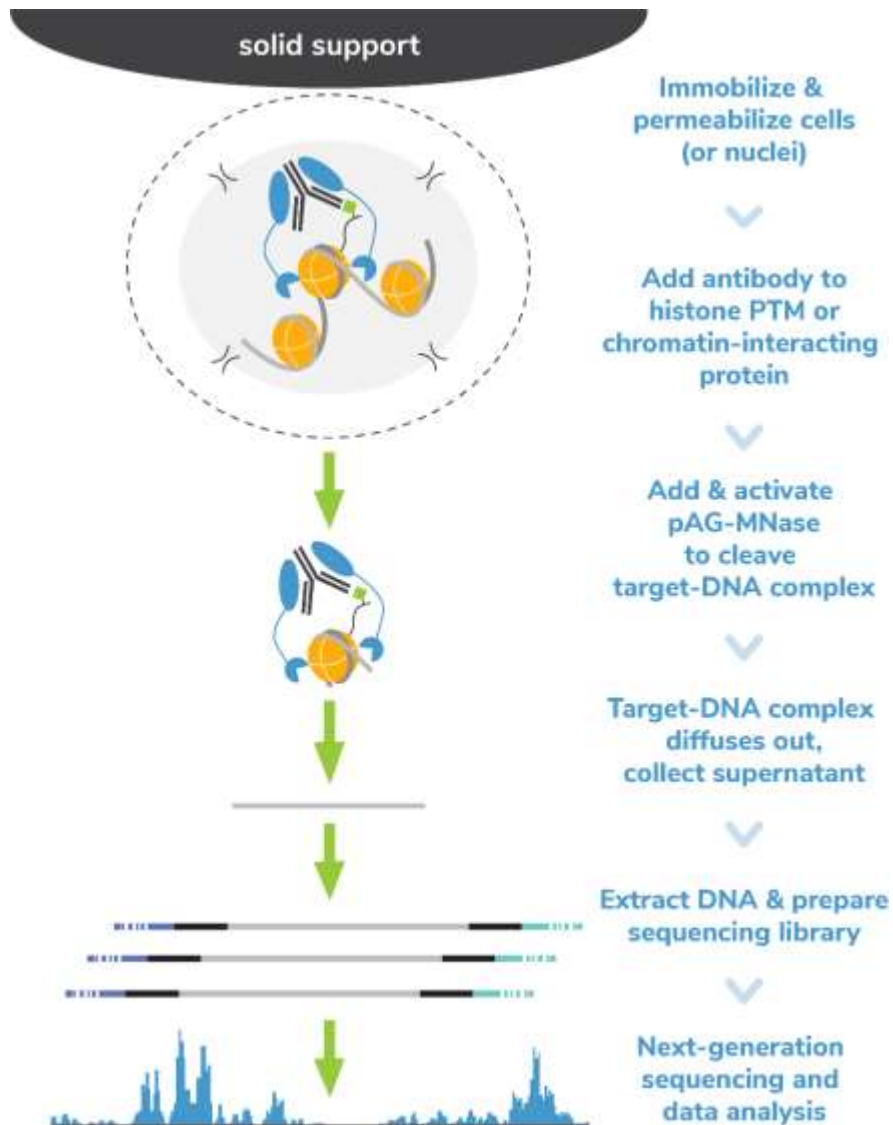
Detect protein-DNA interactions at single nucleotide resolution by ChIP-exo



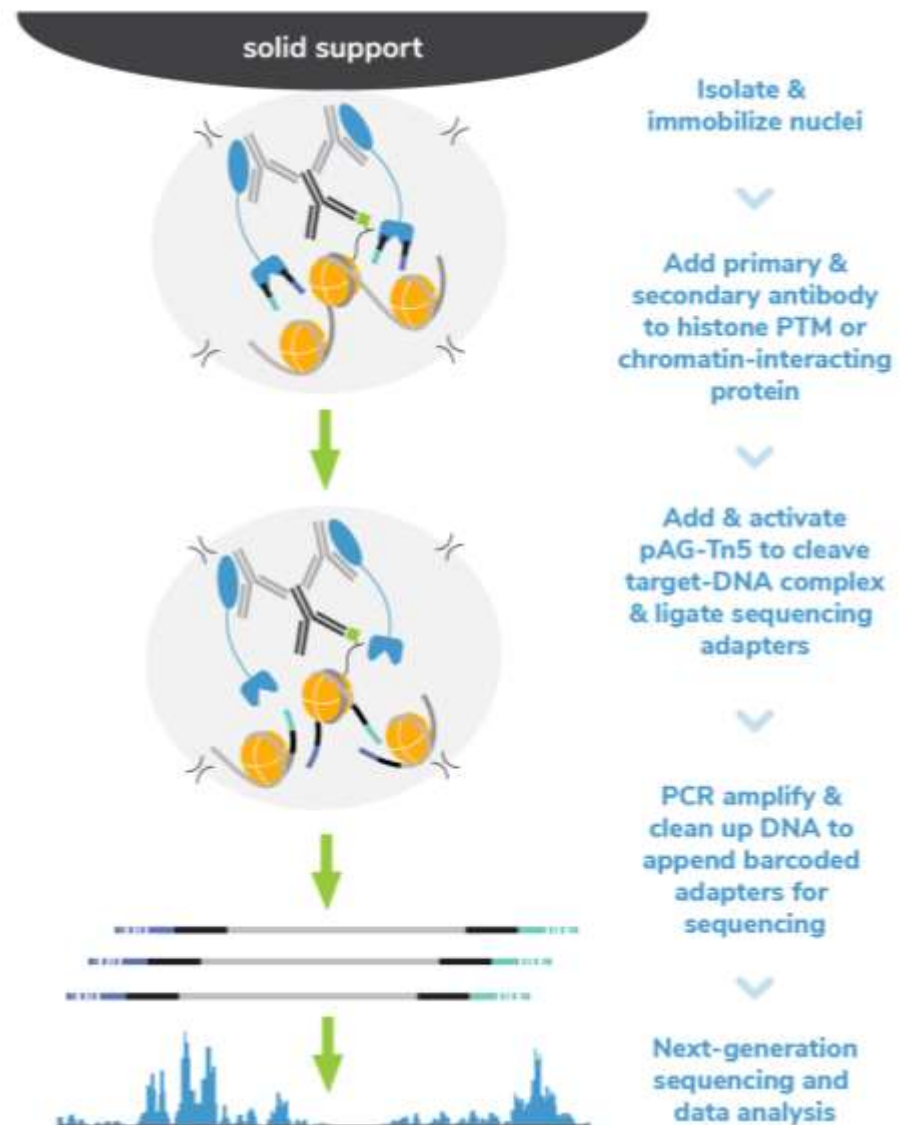
Evolution of ChIP-exo

Descriptive name	ChIP-exo		ChIP-nexus	ChIP-tag-exo		ChIP-SSL-exo		ChIP-exo 5.0
Unique aspect	Original		Circular ligation	Tagmentation		ssDNA, splint adapter		Splint adapter
Hands-on time	12 hours		10.5 hours	9 hours		6 hours		6.5 hours
Version	1.0	1.1	2.0	3.0	3.1	4.0	4.1	5.0
ChIP	ChIP		ChIP	ChIP		ChIP		ChIP
First Adapter Ligation	Polish		Polish	Tagmentation		Polish		A-tailing
	Kinase		Kinase			Lambda exonuclease		
	A-tailing		A-tailing					Adapter ligation/ Kinase
	Adapter ligation (SOLID)	Adapter ligation (Illumina)	Adapter ligation					
Exonuclease Treatment	Fill-in		Fill-in	Fill-in		Adapter ligation (ssDNA)	Adapter ligation (splint)	Fill-in
	Polish		Polish	Polish				Lambda exonuclease
	Kinase			Kinase				
	Lambda exonuclease		Lambda exonuclease	Lambda exonuclease				
	RecJ exonuclease		RecJ exonuclease	RecJ exonuclease		Reverse XL/ProK	Reverse XL/ProK	
	Reverse XL/ProK		Reverse XL/ProK	Reverse XL/ProK				
	AMPure		AMPure	AMPure				
Second Adapter Ligation	Primer extension		Circularization	Primer extension	Splint ligation	Splint ligation	Splint ligation	
	A-tailing			A-tailing				
	Adapter ligation		BamH1 digestion	Adapter ligation				
	AMPure		Ethanol precipitation	AMPure				
PCR	PCR		PCR	PCR		PCR		PCR

Cut&Run



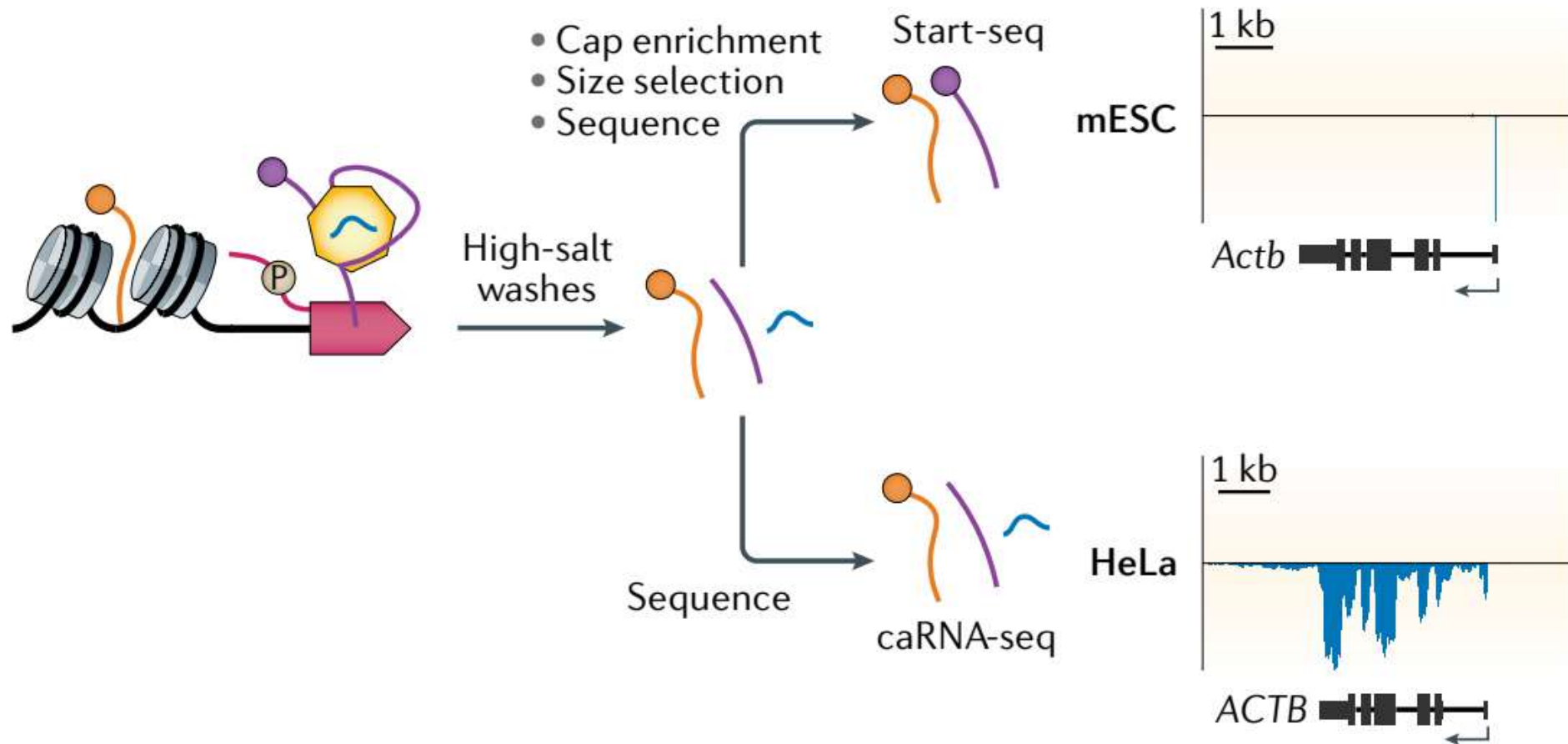
Cut&Tag



Tracking transcription by nascent RNA sequencing

Start-seq/caRNA-seq

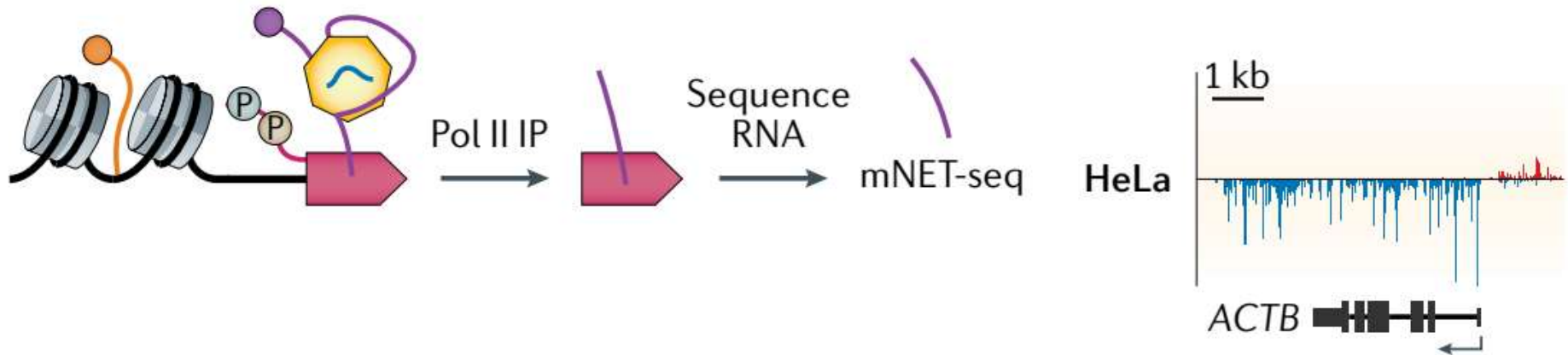
a Chromatin-associated RNA enrichment



Tracking transcription by nascent RNA sequencing

mNET-seq

b Pol II-associated RNA enrichment

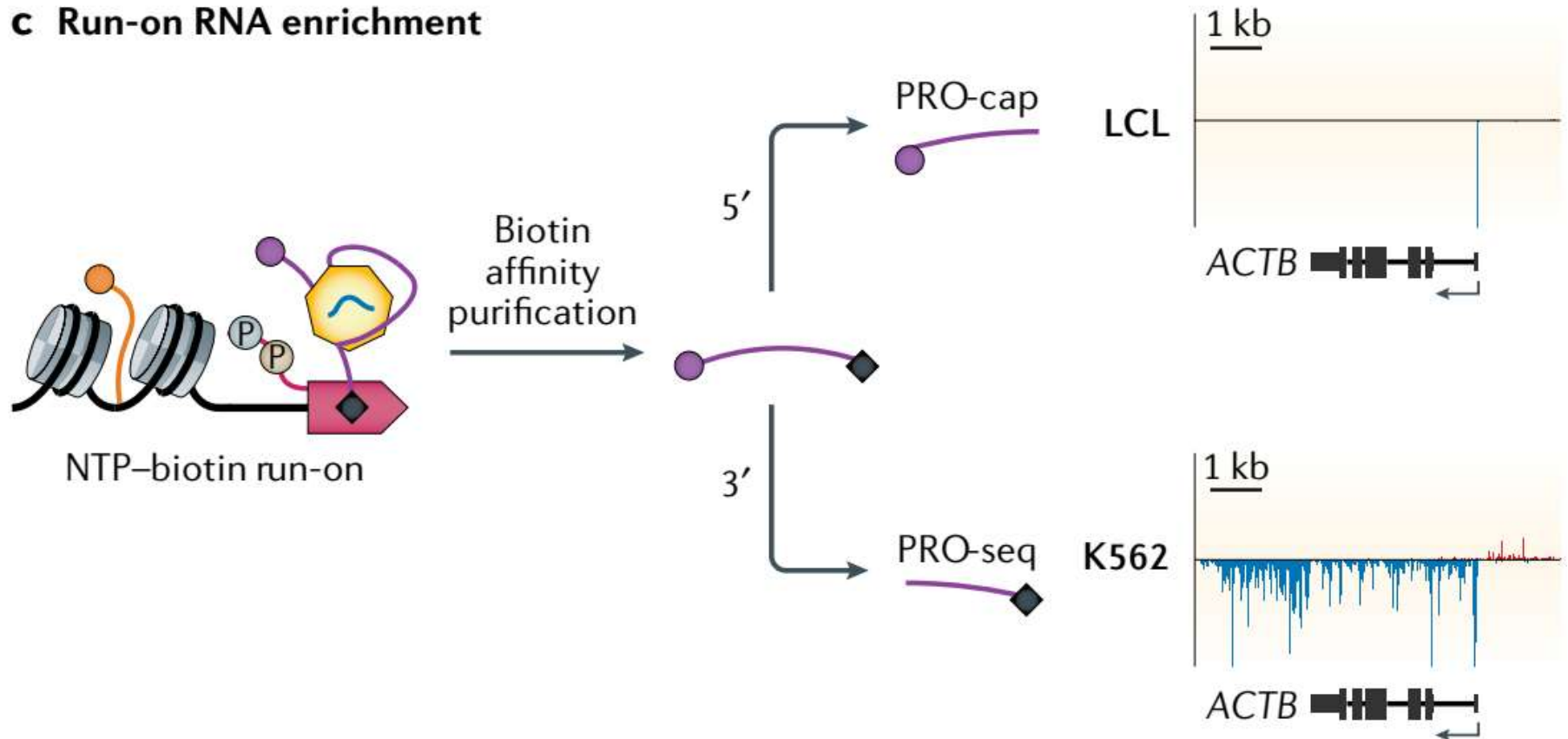


Wissink et al. 2019 NAT REV GENET

Tracking transcription by nascent RNA sequencing

PRO-cap/GRO-seq/PRO-seq

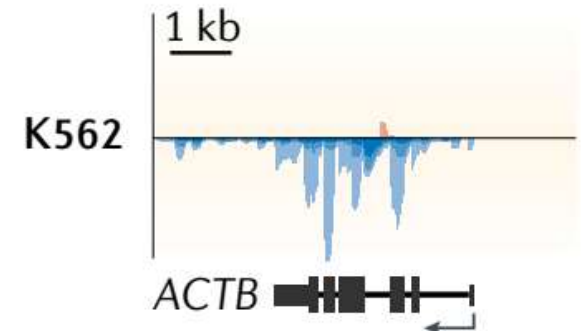
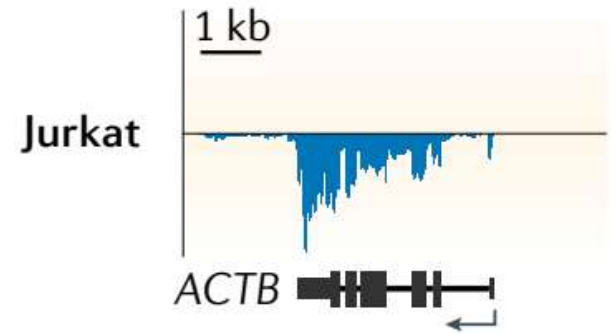
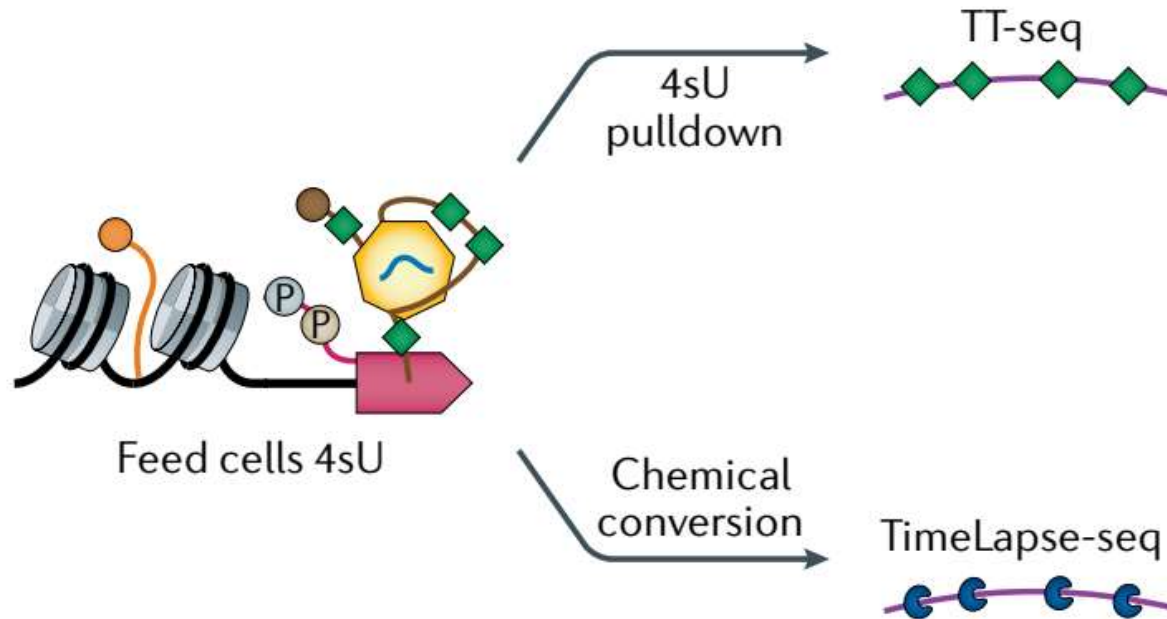
c Run-on RNA enrichment



Tracking transcription by nascent RNA sequencing

TT-seq/TimeLapse-seq

d Metabolic RNA labelling

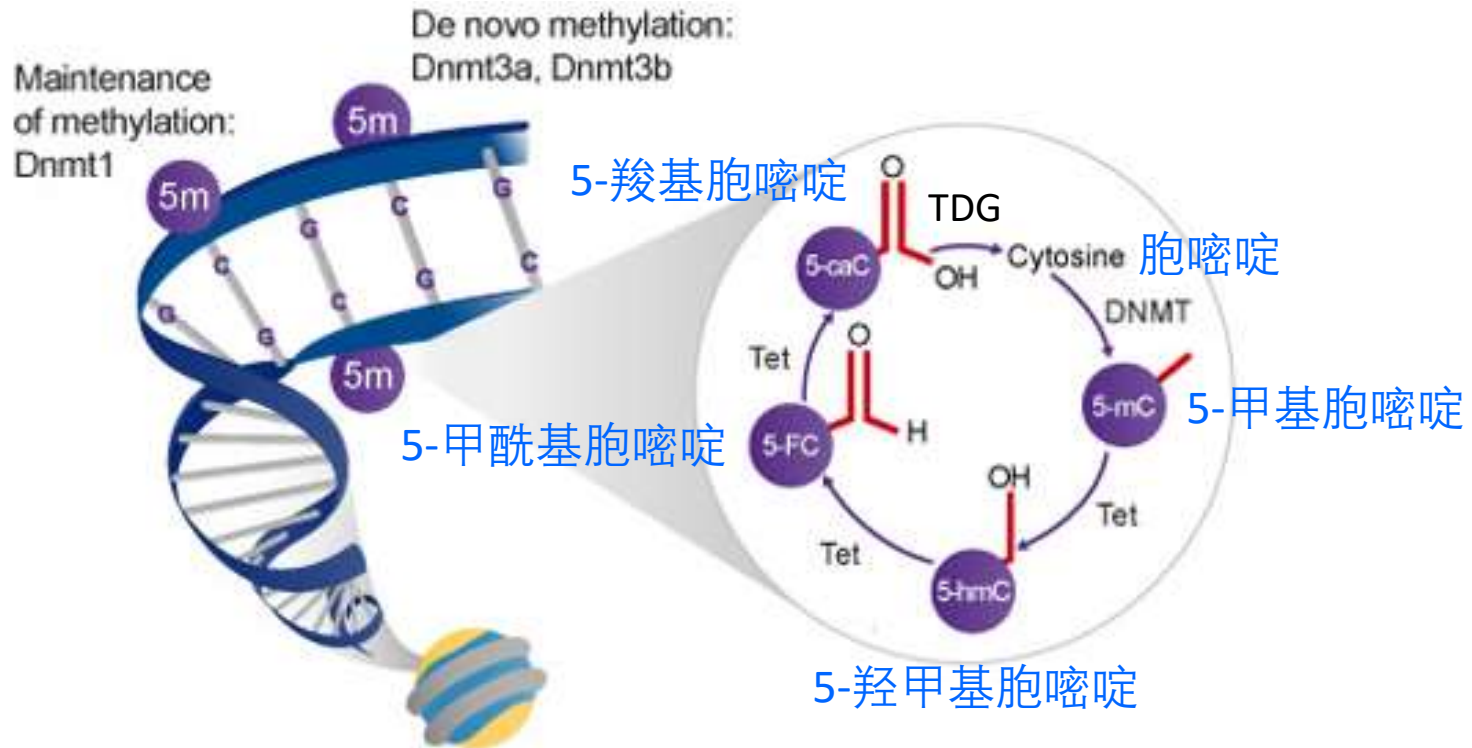


How to employ these techniques to study transcription?

Outline

1. Eukaryotic transcription machineries
2. Function elements for eukaryotic transcription
3. Transcription initiation in eukaryotes
4. Transcription-coupled RNA processing
5. Techniques to study eukaryotic transcription
- 6. Examples for studying eukaryotic transcription**

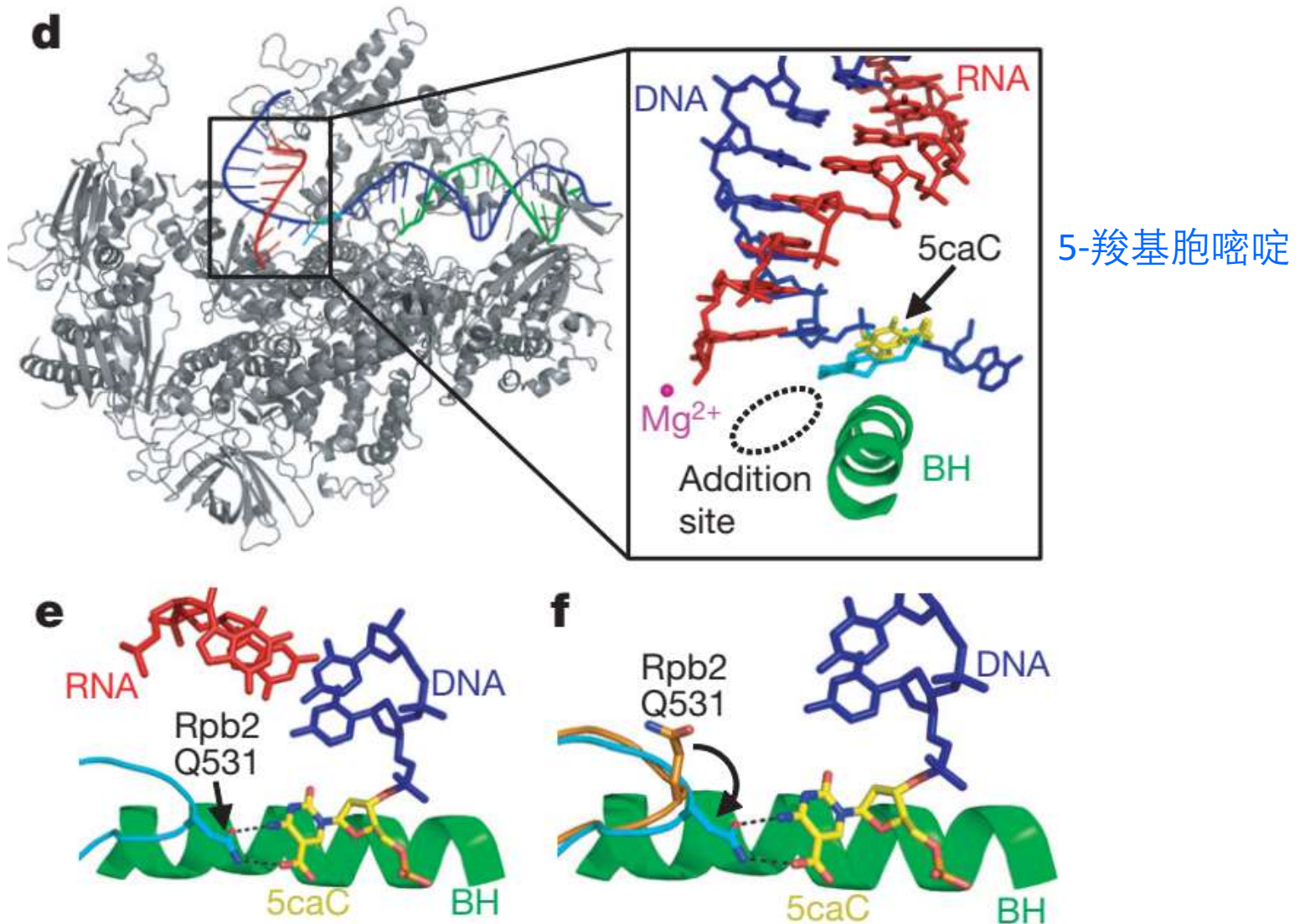
5caC recognition by Pol II elongation complex



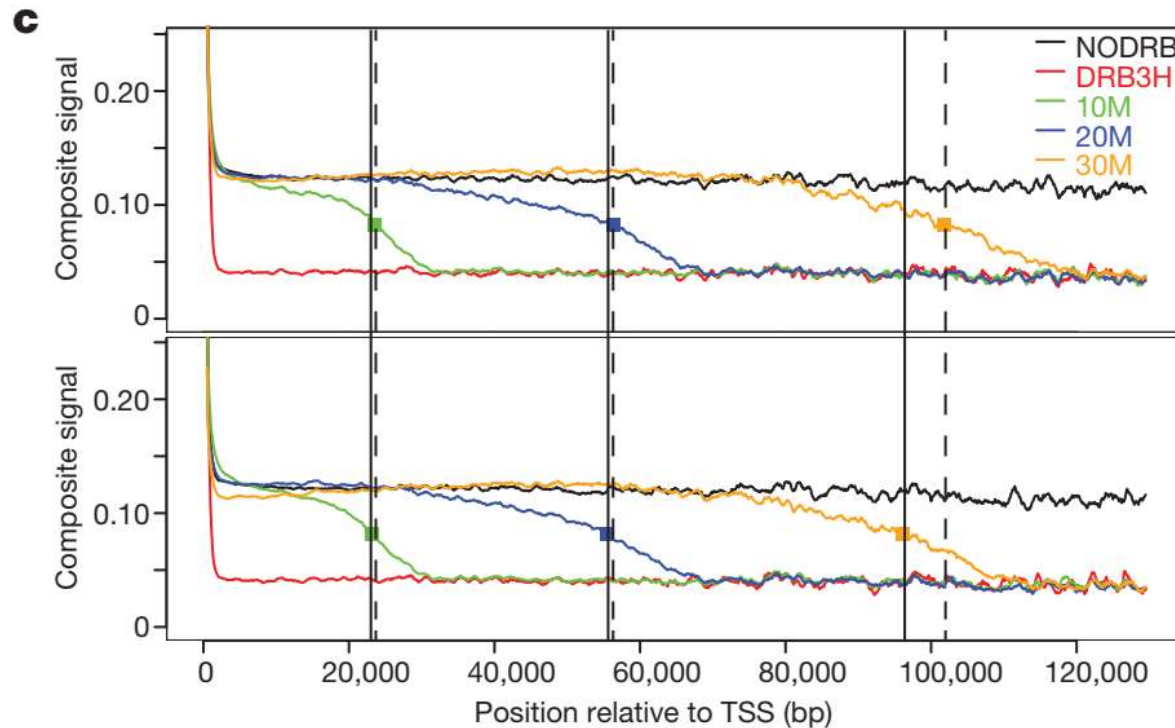
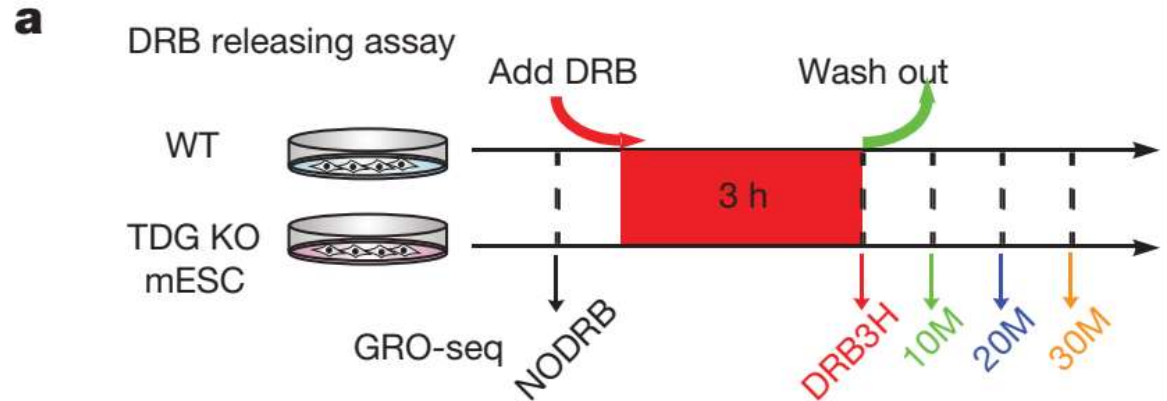
- DNA甲基化是基因组上重要的表观遗传密码，且呈现甲基化-去甲基化的动态修饰。

5caC等去甲基化中间产物是否具有基因表达调控功能？

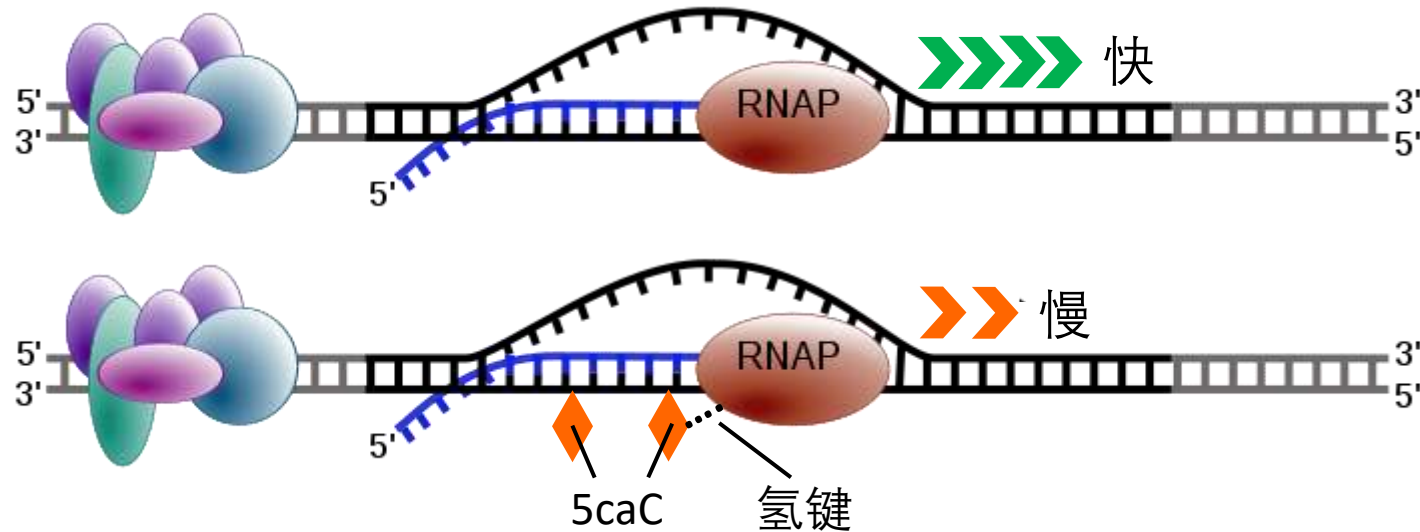
5caC recognition by Pol II elongation complex



5caC reduces Pol II elongation rate



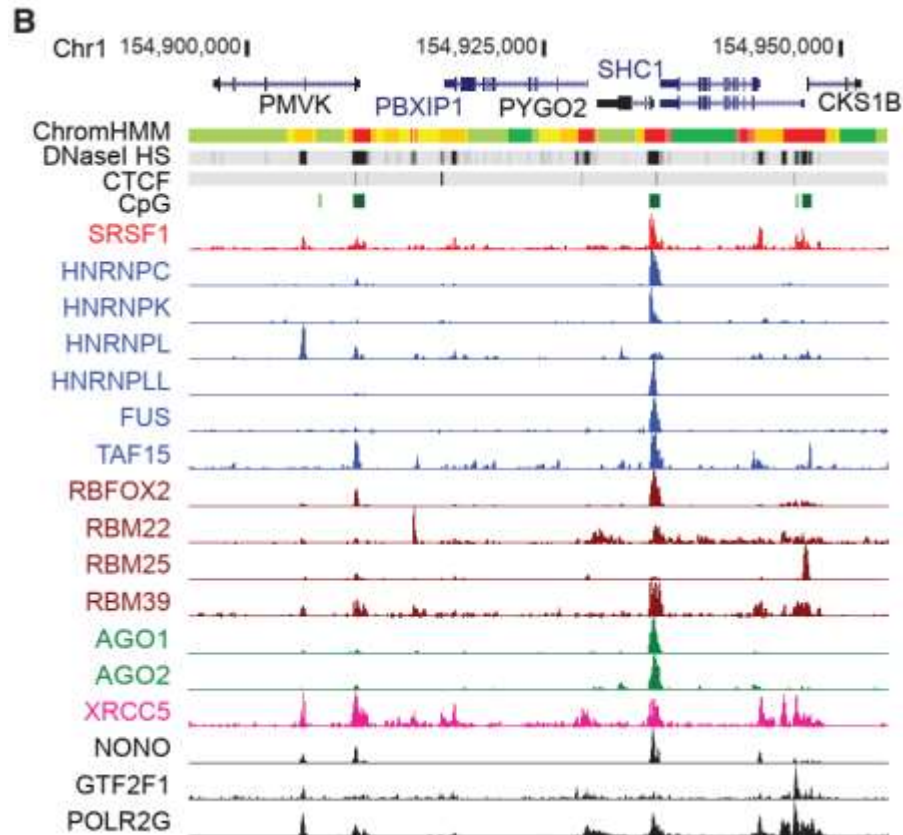
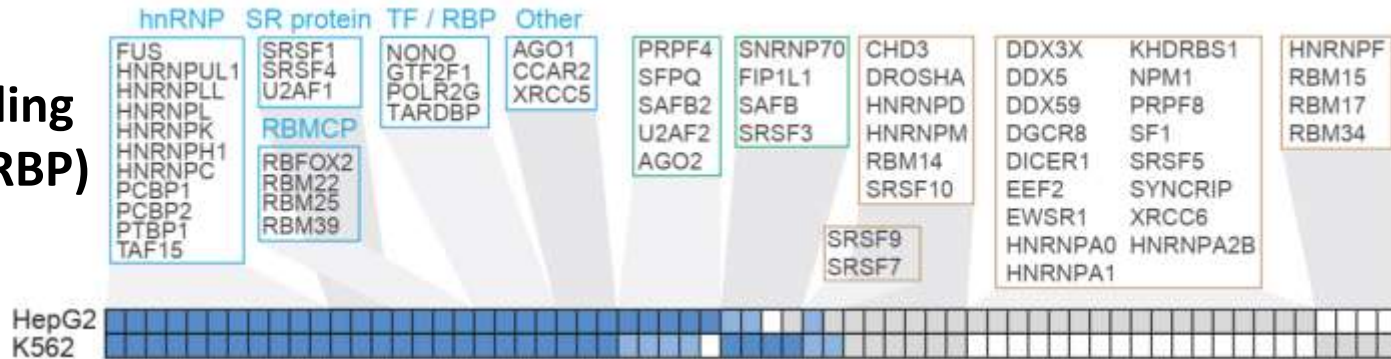
The function of 5caC in transcription



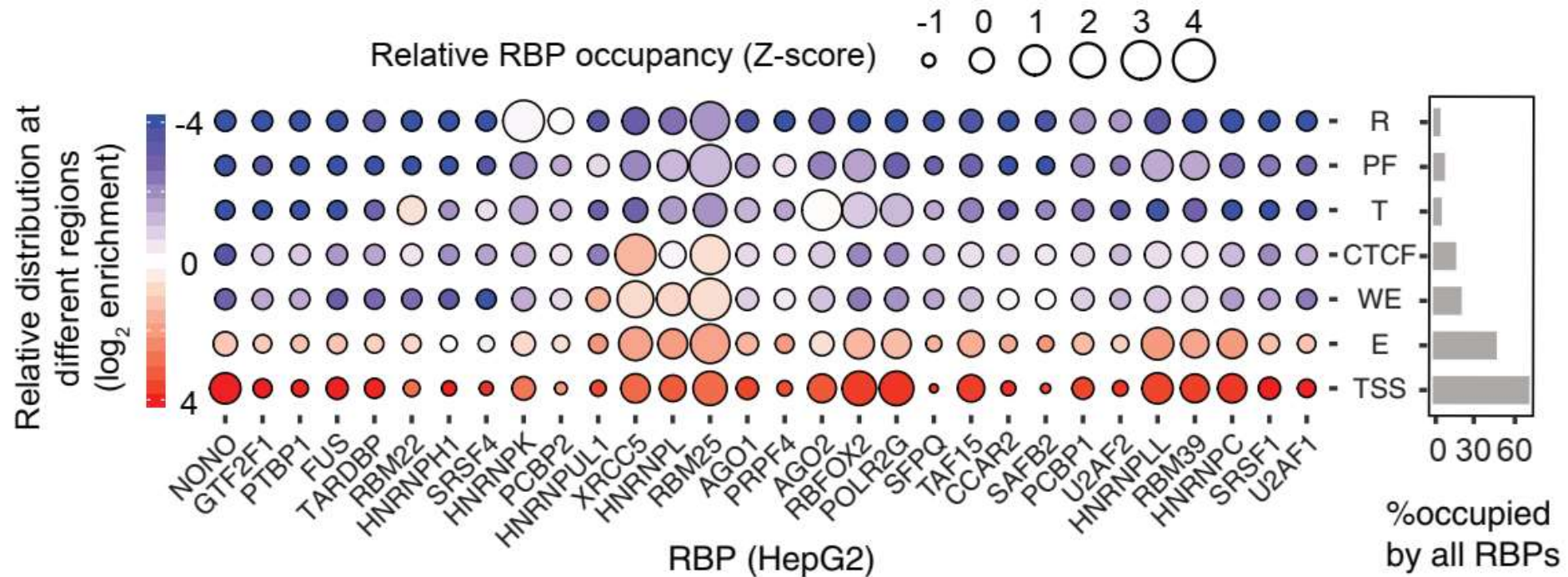
- 揭示了5-羧基胞嘧啶(5caC)的转录调控功能
- 发现5-羧基胞嘧啶可与Rbp2 Q531形成氢键相互作用，且阻碍GTP的配对和掺入
- 开发了时段GRO-seq新技术，实现体内转录测速
- 在体内，5caC阻碍转录延伸

The landscape of RBP-chromatin interactions

RNA-binding
protein (RBP)

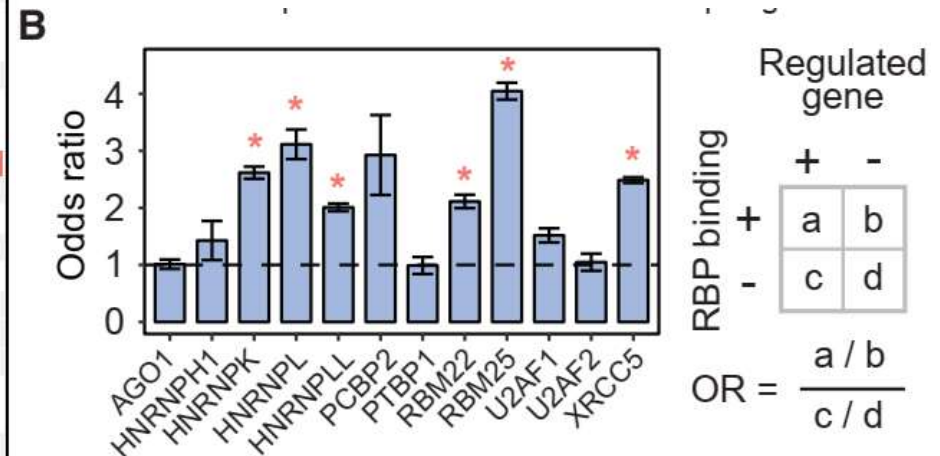
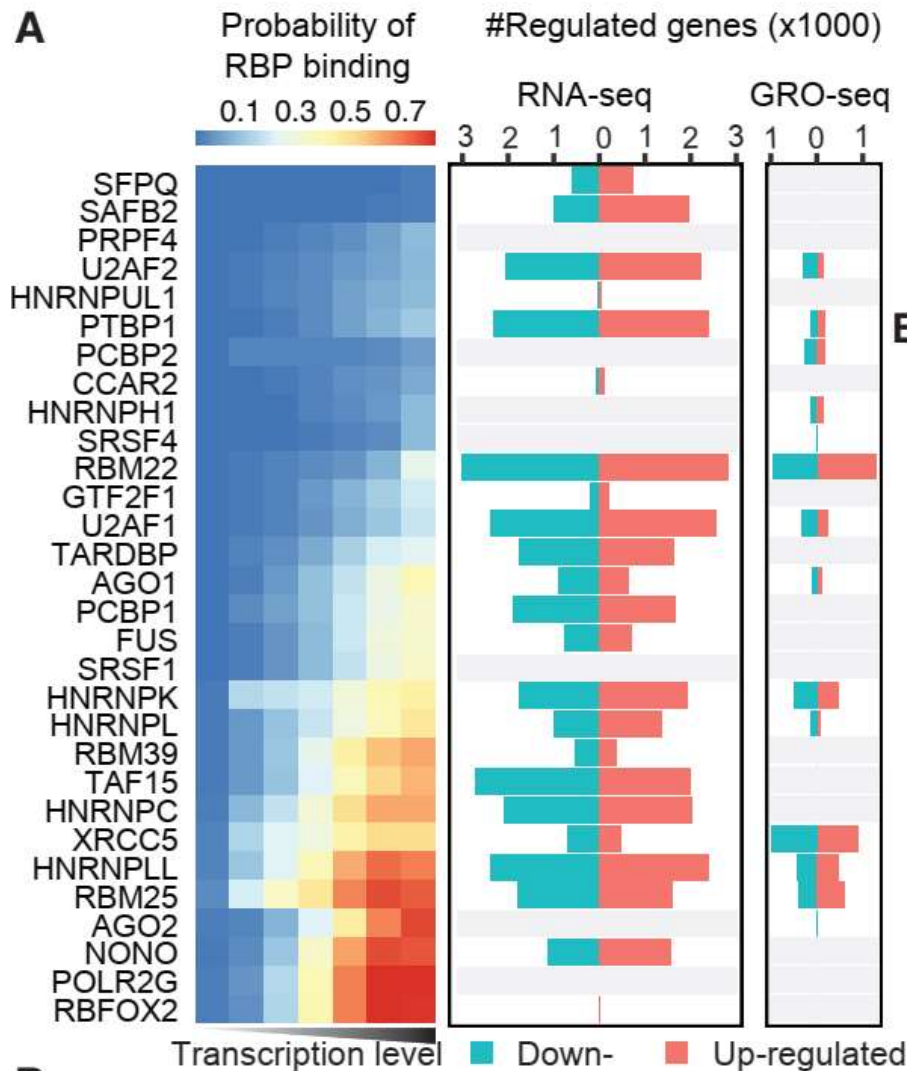


The landscape of RBP-chromatin interactions



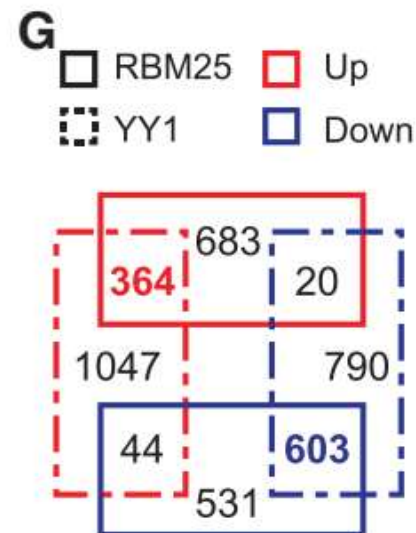
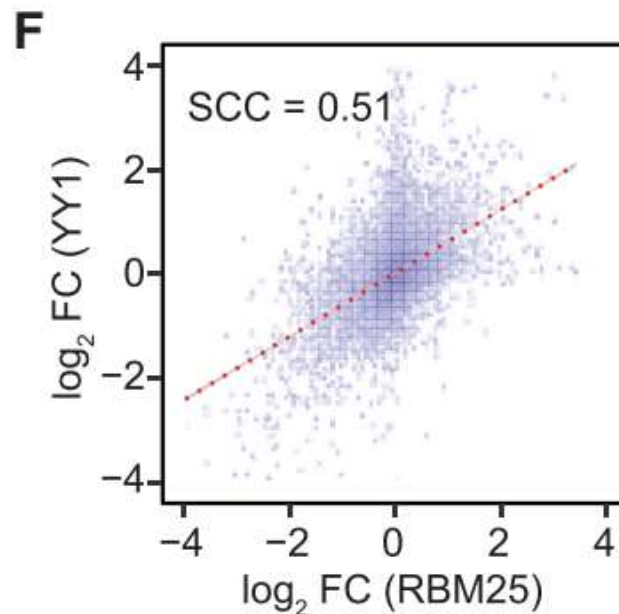
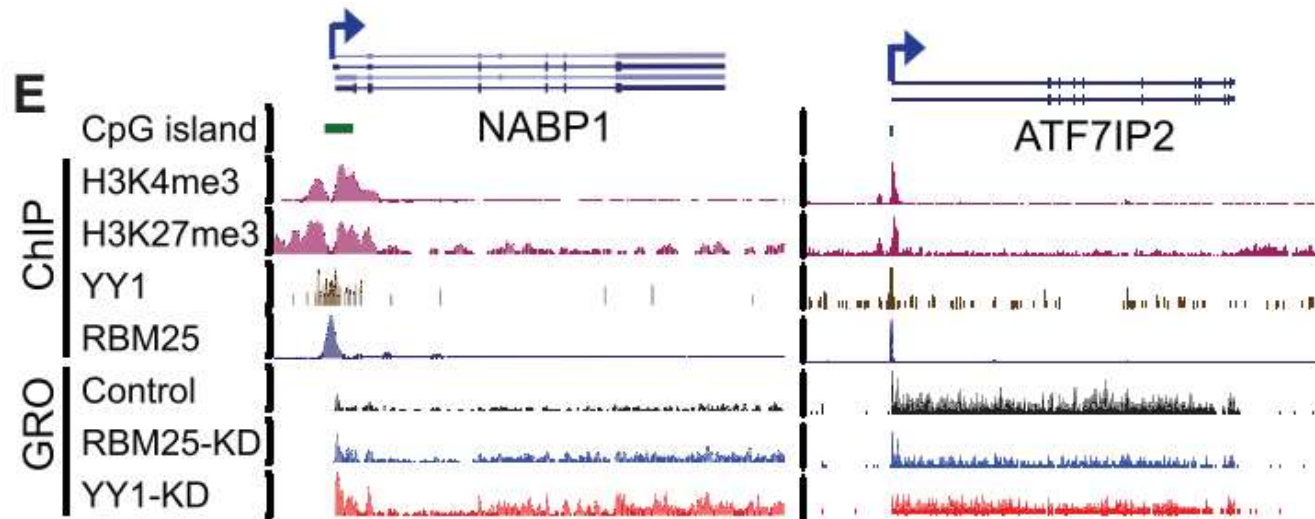
Xiao#* et al. 2019 Cell

RBPs participate in transcription regulation

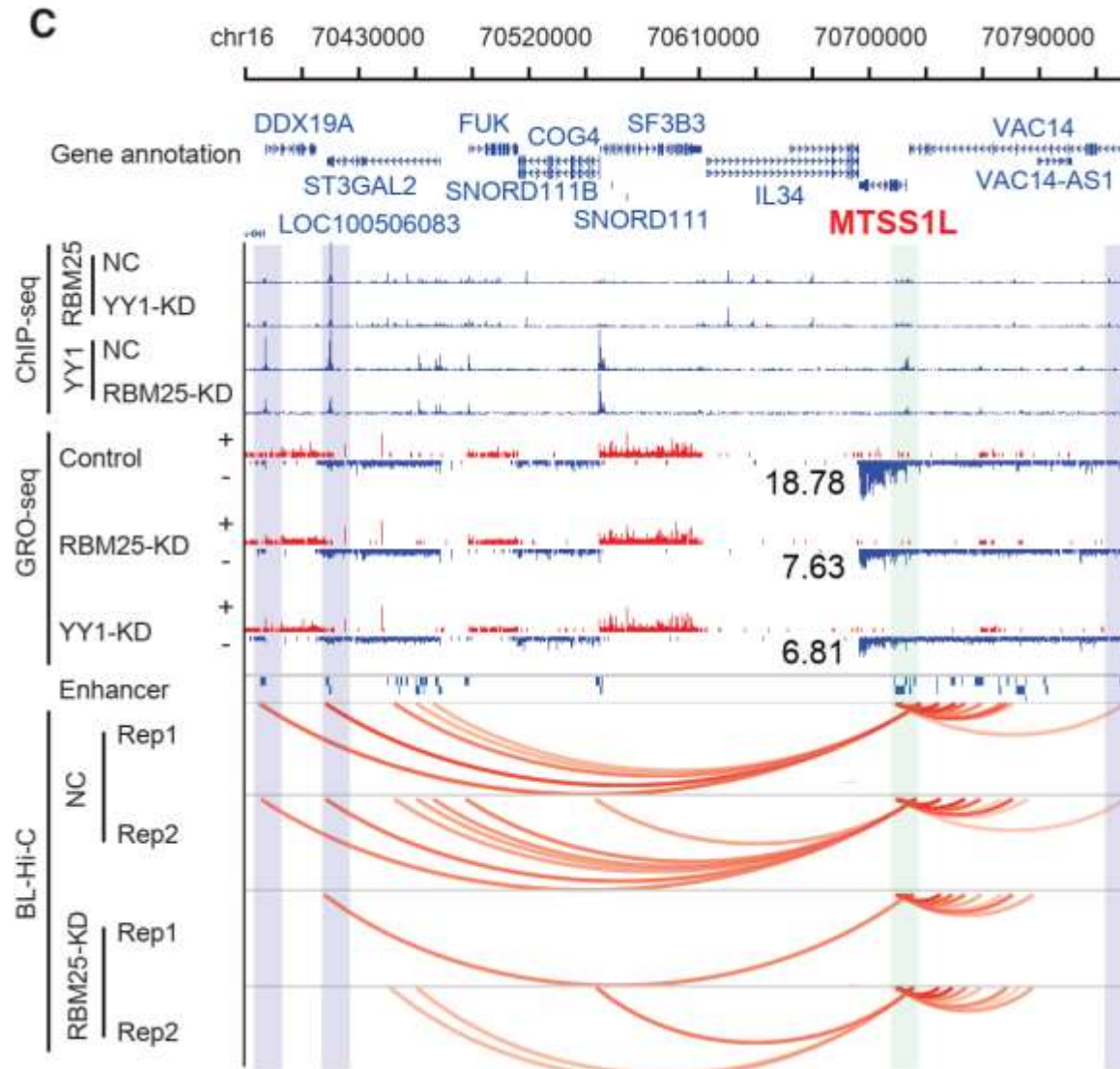


Xiao#* et al. 2019 Cell

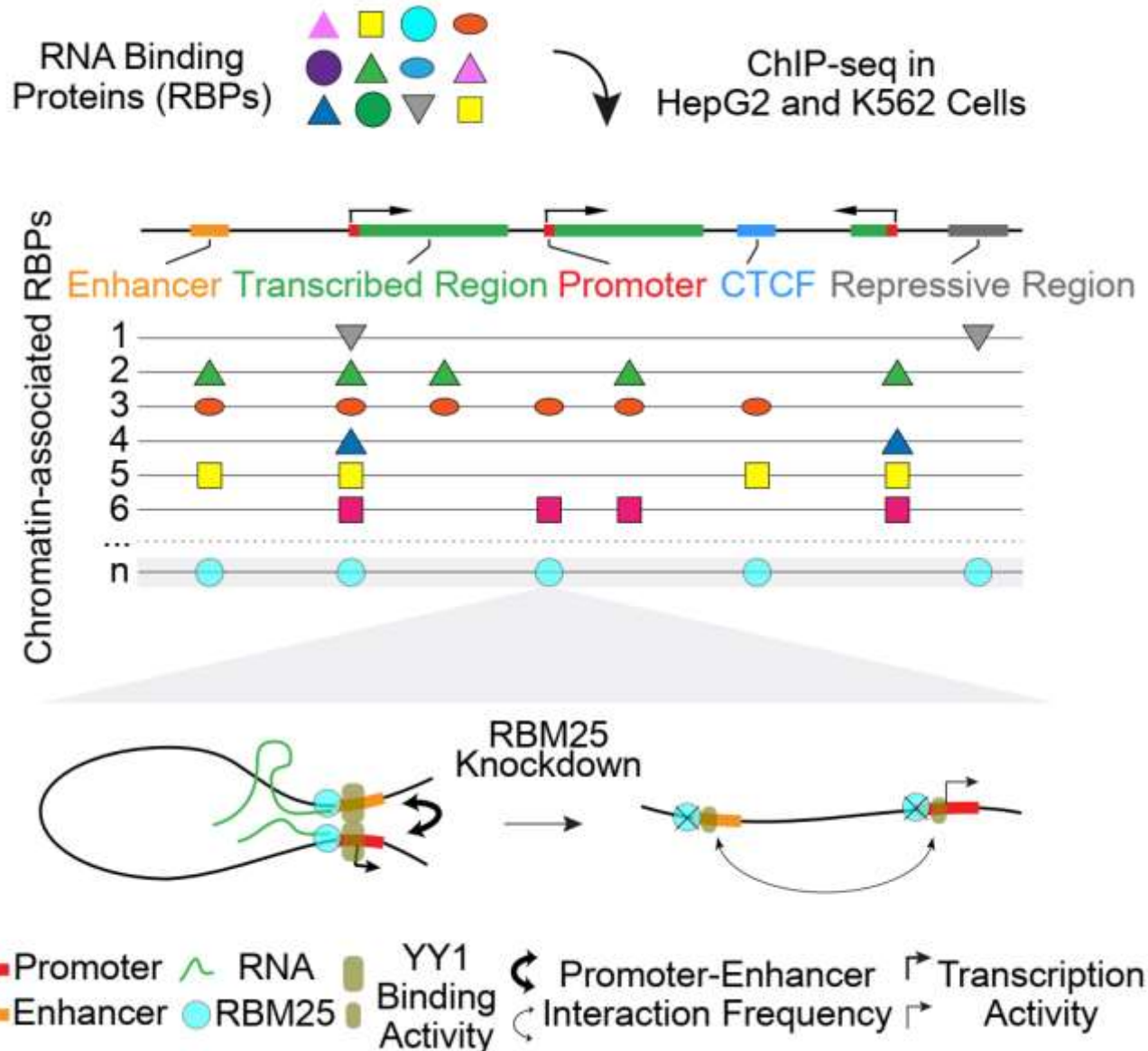
RBM25 coordinates YY1 to regulate transcription



RBM25 regulates YY1-mediated promoter-enhancer looping



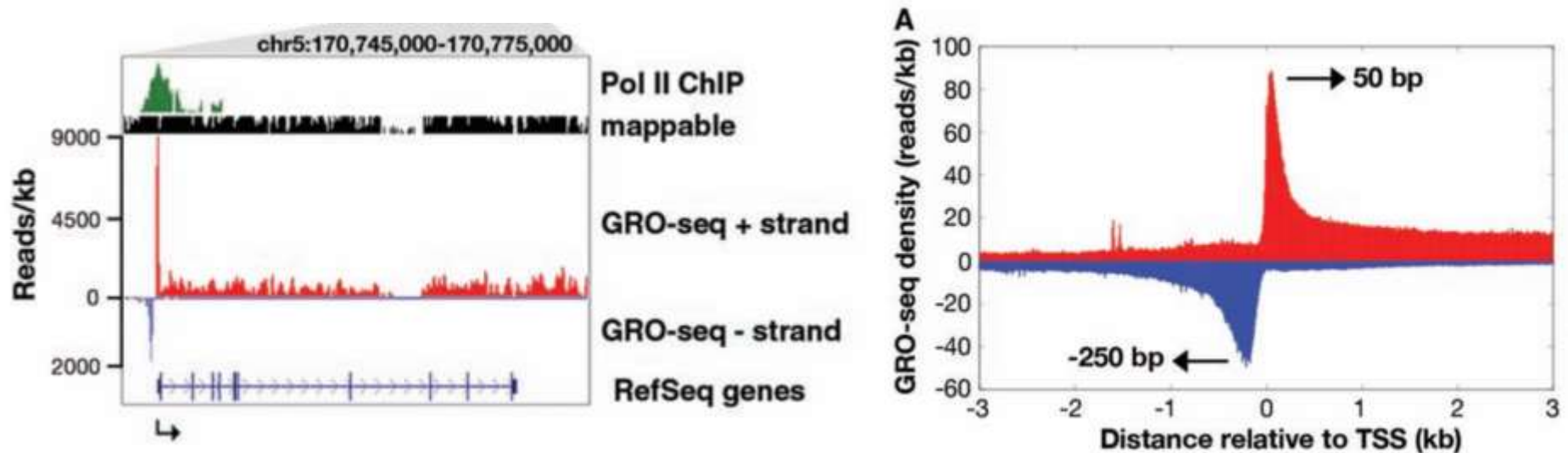
Pervasive RBP-chromatin interactions and functions



Take-home messages

- **How to initiate transcription in eukaryotes?**
- **The RNA processing events coupled with eukaryotic transcription**
- **The techniques to study transcription in eukaryotes**

Think over



Core et al. 2008 Science

**RNA Pol II undergoes bidirectional transcription.
However, what determines the full length transcription?**

Made with **KINEMASTER**

instructions to make proteins are contained in our DNA. DNA contains genes. A gene is a continuous string of