Metodi di Immunoprecipitazione di RNA e cromatina

Corso di Ingegneria Genetica e Terapia Genica

Ballarino Monica

Immunoprecipitation

RNA immunoprecipitation (RIP)

CrossLinking and ImmunoPrecipitation (CLIP)

Chromatin Immunoprecipitation (ChIP)

Chromatin Isolation by RNA Purification (ChIRP)

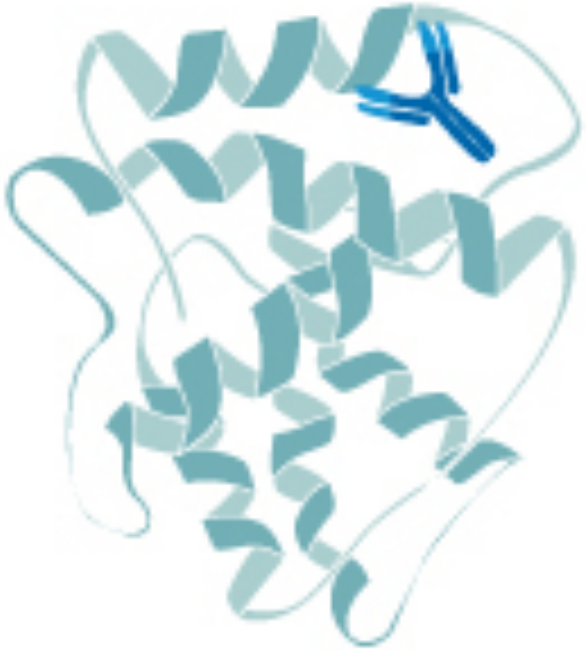
Immunoprecipitation (IP)

is the technique of precipitating a protein *antigen* out of solution using an *antibody* that specifically binds to that particular protein. This process can be used to *isolate* and *concentrate*:

- a particular protein from a sample containing many thousands of different proteins.
- protein partners (i.e. other proteins/RNA/chromatin)

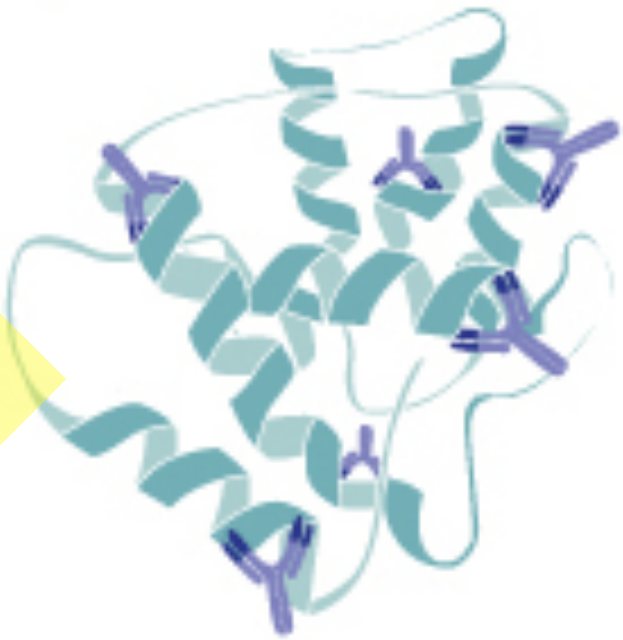
Immunoprecipitation requires that the **antibody** be coupled to a solid substrate at some point in the procedure.

Antibodies



Access to a desired epitope can be compromised by:

- interactions with other proteins
- post-translational modifications
- temperature, pH, fixation, and salt concentration.



Monoclonal

Monoclonal antibodies, produced from a single B cell clone, represent a *homogeneous population* that bind with high affinity and specificity to a ***single*** *epitope*.

Polyclonal

Polyclonal antibodies are *heterogeneous* and can recognize ***multiple*** epitopes they are less likely to be affected by changes in protein conformation.

(1) *Immunization* of a mouse

(2) Isolation of *B cells* from the spleen

(3) Cultivation of *myeloma* cells (immortal)

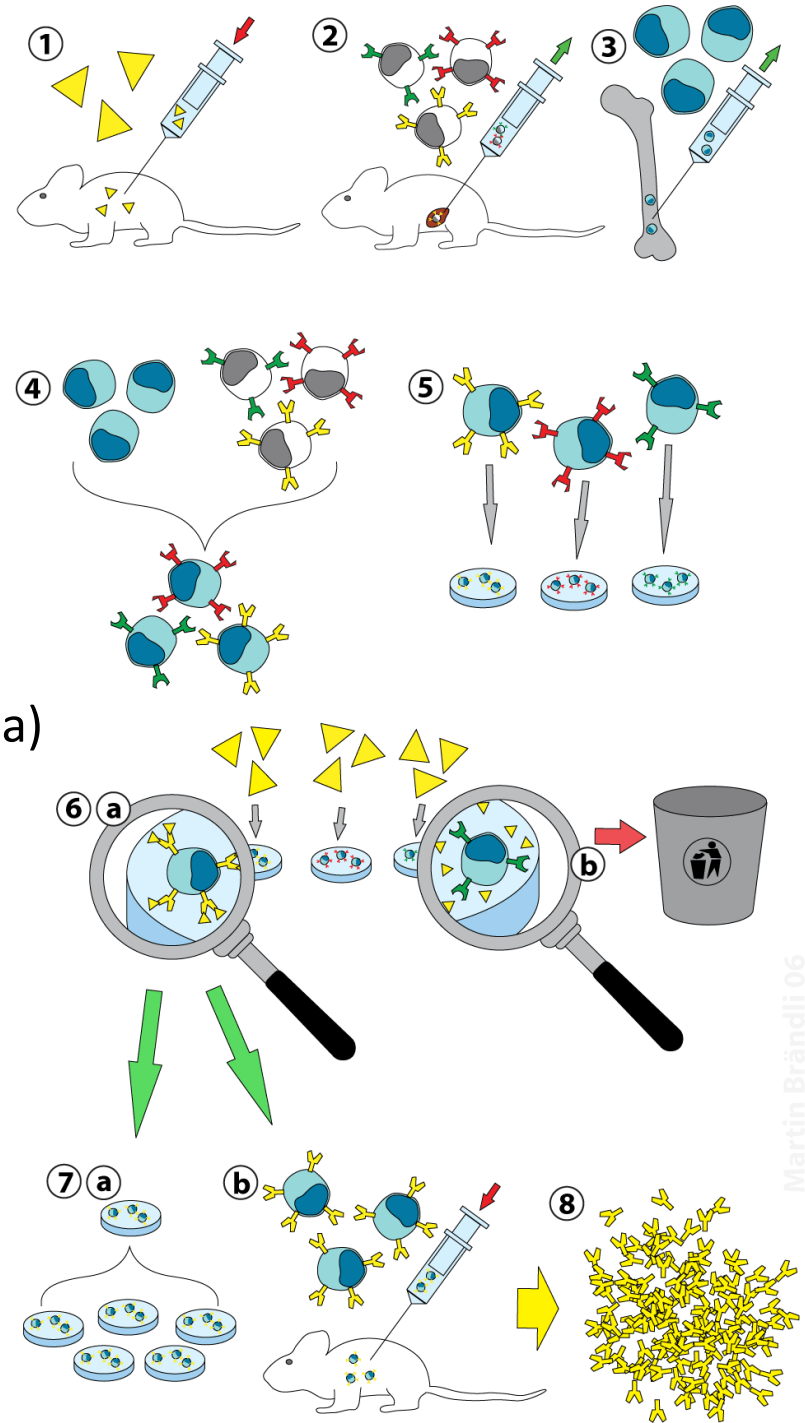
(4) *Fusion* of myeloma and B cells (Hybridoma)

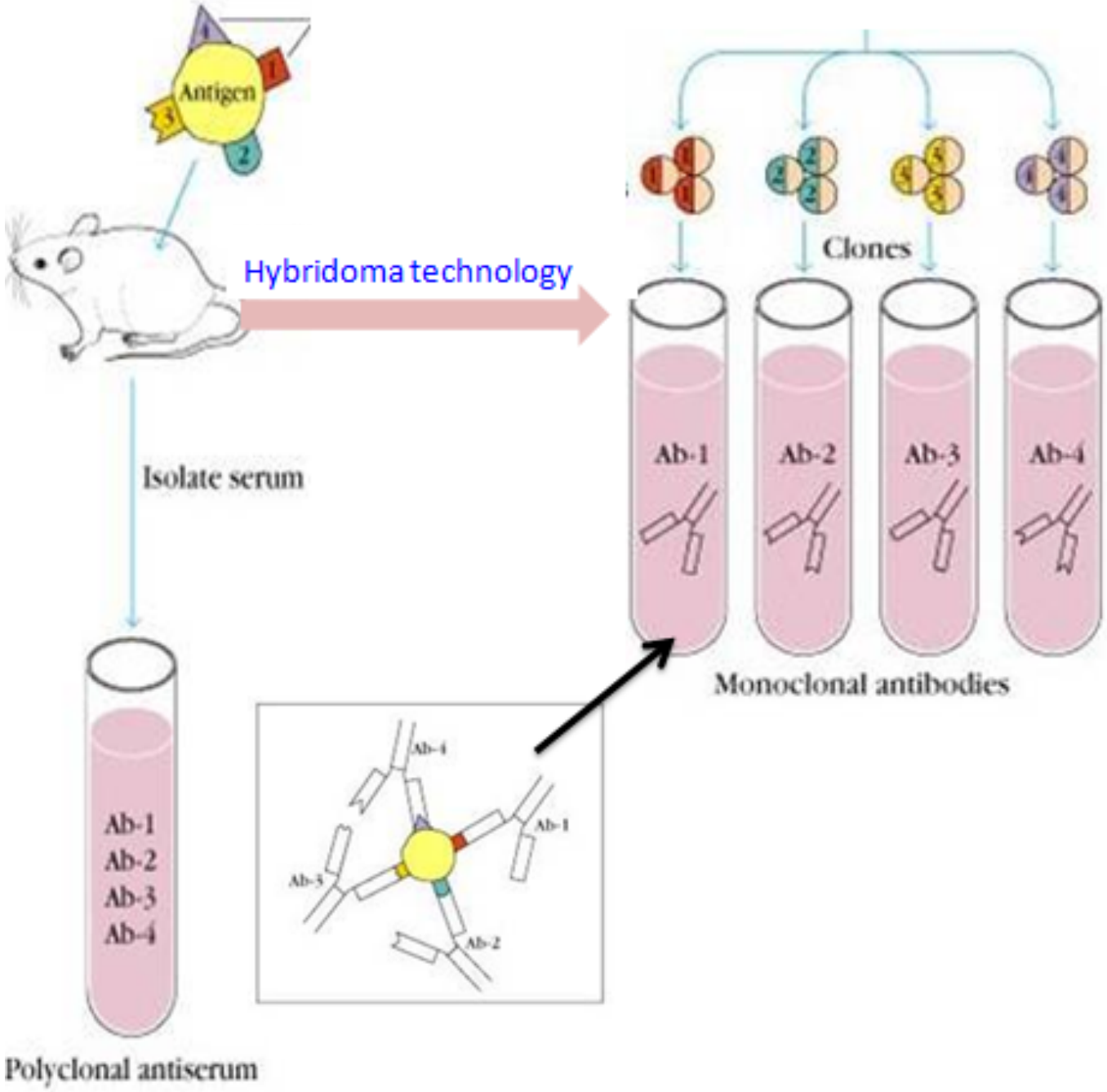
(5) Separation of cell lines

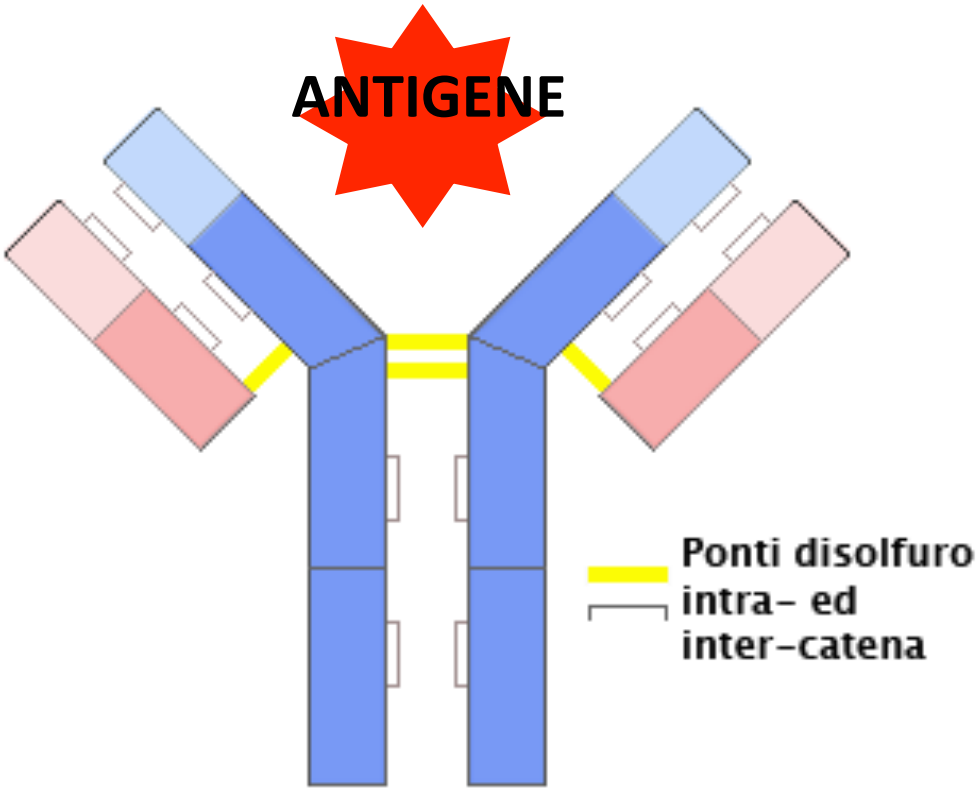
(6) Screening of suitable cell lines

(7) *in vitro* **(a)** or *in vivo* **(b)** multiplication

(8) Harvesting

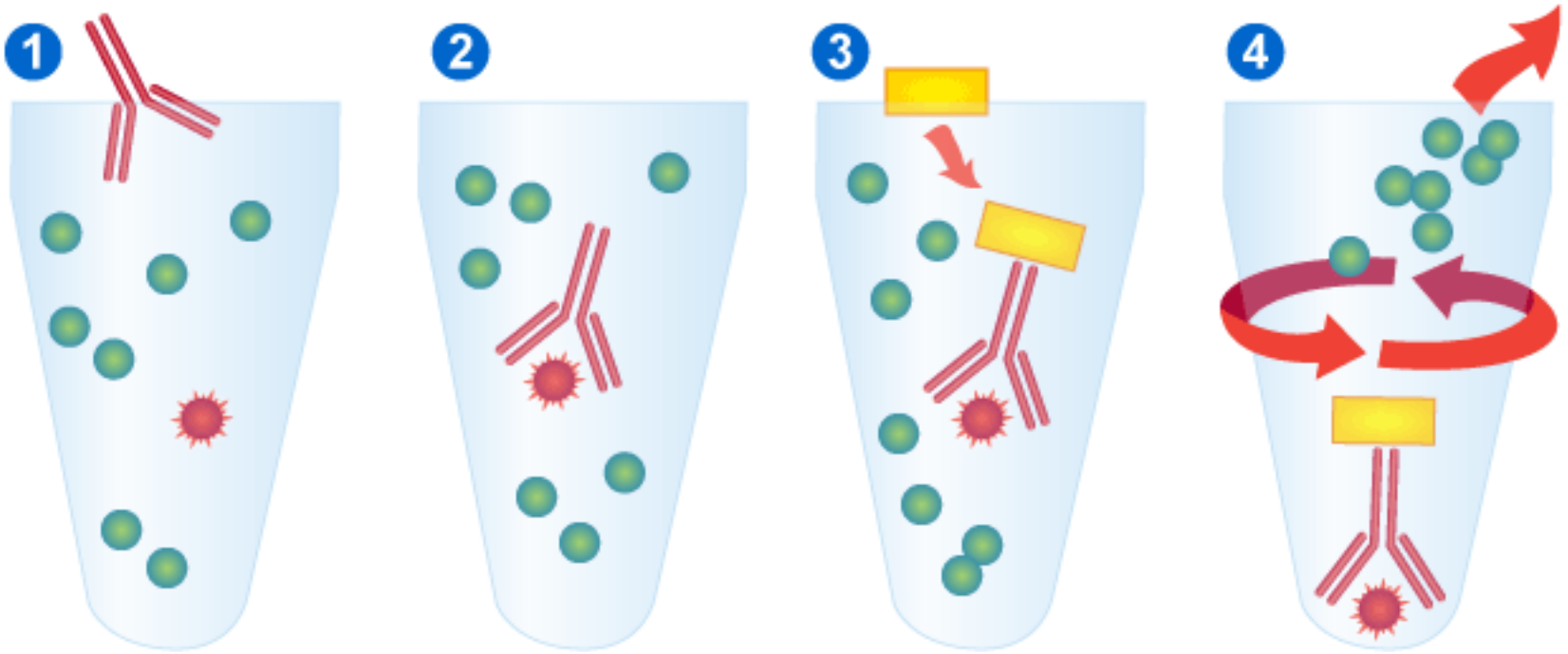



Antigen
Hybridoma technology
Clones
Ab-1
Ab-2
Ab-3
Ab-4
Monoclonal antibodies
Isolate serum
Ab-1
Ab-2
Ab-3
Ab-4
Polyclonal antiserum

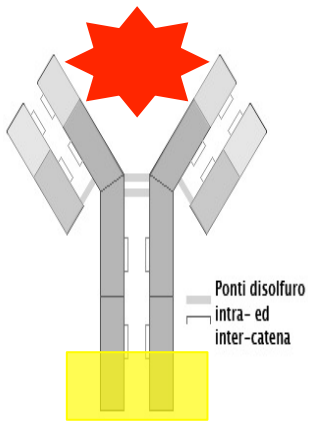


Struttura degli anticorpi. La figura mostra in rosa le due catene leggere, in azzurro le due catene pesanti. I domini più scuri corrispondono alle regioni costanti, quelli chiari indicano, invece, le porzioni variabili, responsabili del riconoscimento con l'antigene.

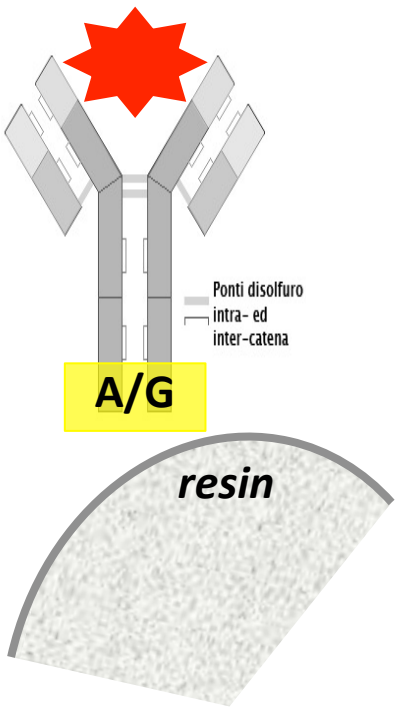
Immunoprecipitation (IP)



1. Suitable antibody is added.
2. Antibody binds to protein of interest.
3. Protein A or G added to make antibody-protein complexes insoluble.
4. Centrifugation of solution pellets antibody-protein complex. Removal of supernatant and washing.



Immunoprecipitation: the role of protein A or G



Binding Characteristics of Some Immunoglobulins			
Immunoglobulin		Protein A	Protein G
Mouse	IgG1	+	++
	IgG2a	+++	+++
	IgG2b	++	++
	IgG3	+	+++
	IgM	-	-
	IgA	-	-
	IgE	-	-
Rat	IgG1	+	+
	IgG2a	-	+++
	IgG2b	-	++
	IgG2c	+	++
Human	IgG1	+++	+++
	IgG2	+++	+++
	IgG3	-	+++
	IgG4	+++	+++

Protein “tagging”

consiste nella *fusione* tra la proteina d’interesse ed un un peptide noto.
Il peptide o “**TAG**” ha il vantaggio di poter essere facilmente *purificato*, permettendo di isolare grandi quantità di proteina, complessi multiproteici, complessi acidi nucleici/proteine

Le proteine di fusione sono ottenute previo **clonaggio in vettori di espressione.**

DNA codificante la proteina + DNA codificante il Tag

La proteina di fusione espressa è definita *proteina ricombinante*

Proteine ricombinanti

- Proteine per la ricerca di base e applicata
- Proteine di interesse commerciale (enzimi)
- Proteine ad uso terapeutico

Espressione eterologa

permette l'espressione di proteine specifiche in un organismo che normalmente non le produrrebbe.

L'espressione dei geni nei sistemi eterologhi nasce:

1. dalla necessità generale ad ottenere *grandi quantità* di peptidi di interesse pratico o scientifico

2. dalla *difficoltà* di esprimere geni, di qualunque natura, in organismi superiori assai complessi e soggetti a regolazioni non sempre pienamente comprese.

3. dalla *facilità* di ottenere in maniera diretta e comoda da organismi "semplici" le proteine purificate

Per esprimere una proteina in un sistema eterologo occorrono:

VETTORE D'ESPRESSIONE + **OSPITE D'ESPRESSIONE**

VETTORE D'ESPRESSIONE + OSPITE D'ESPRESSIONE

E' virtualmente possibile esprimere geni in sistemi di ogni tipo.

La scelta di quale sistema utilizzare dipende dal *fine* e dalle *proprietà* della proteina che deve essere prodotta

Batteri	*Escherichia coli* *Bacillus subtilis*	
Funghi	*Saccaromyces cerevisiae* *Aspergillus nidulans*	
Piante	*Arabidopsis thaliana,* *Nicotiana tabacco*	cellule in coltura protoplasti piante transgeniche
Insetti	*Dorifera californica* *Drosophila melanogaster*	cellule in coltura *org*anismi interi
Animali		oociti cellule in coltura organismi interi

	VANTAGGI	SVANTAGGI
Batteri	• Semplicità, • Corti tempi di generazione • Elevata resa del prodotto • Bassi costi • Secrezione delle proteine	• Misfolding proteico • Formazione di *inclusion bodies* • Tossicità delle proteine • Mancanza di modificazioni post-traduzionali
Lievito	• Semplicità, • Corti tempi di generazione • Elevata resa del prodotto • Bassi costi • Modificazioni post-traduzionali	• Presenza di proteasi attive che degradano le proteine prodotte

Insetti

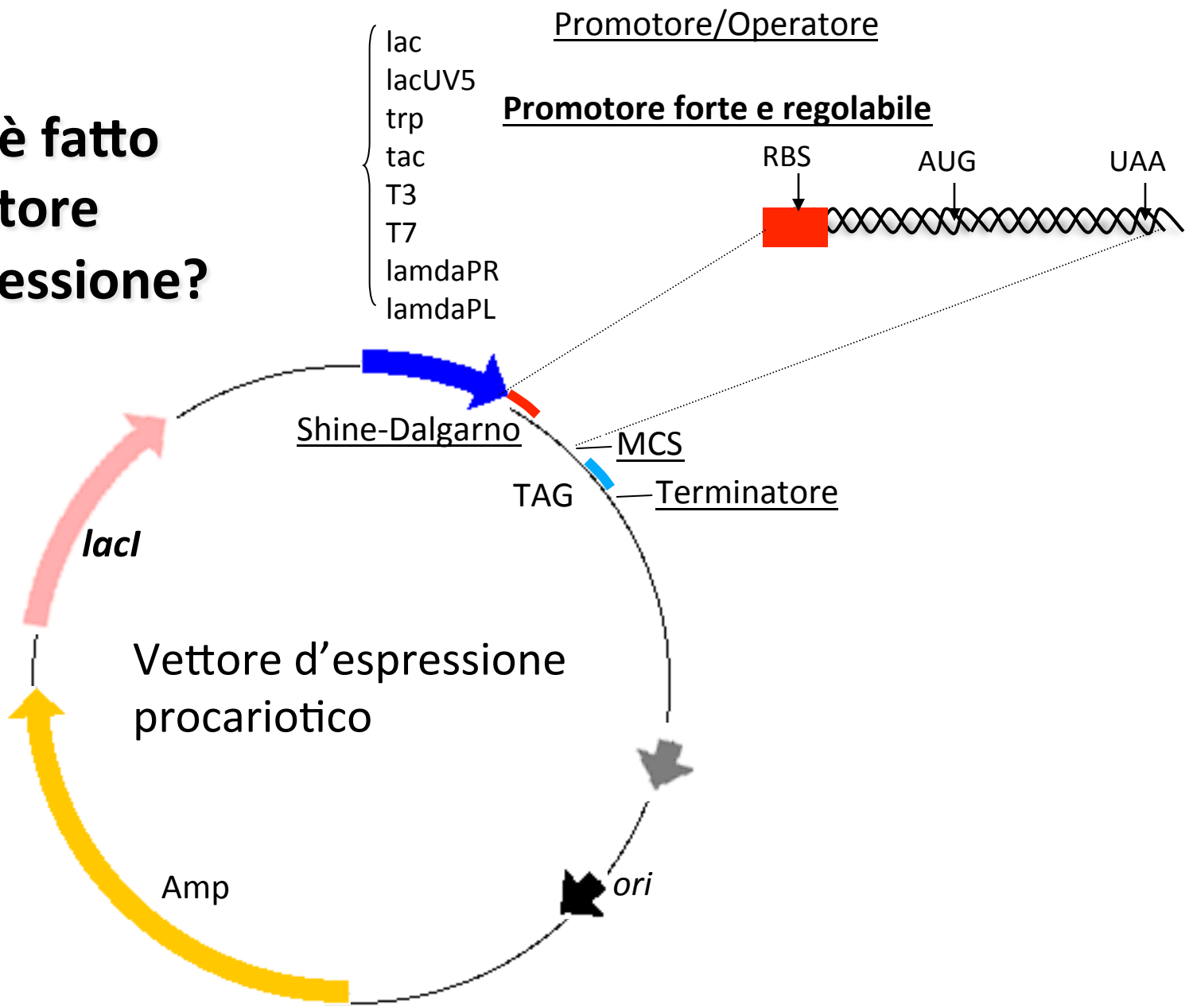
Piante

Animali

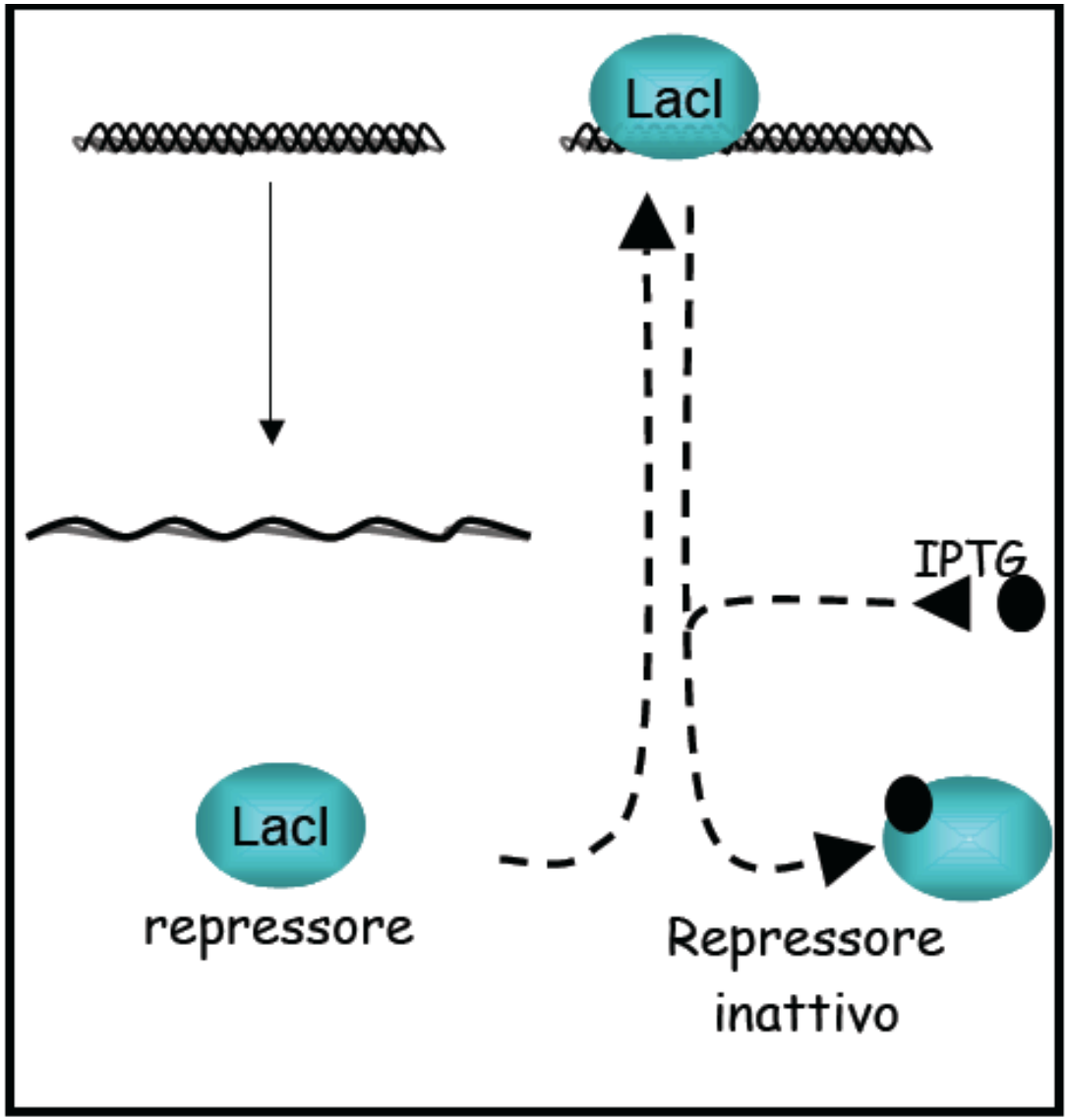
Costi elevati!!

VETTORE D'ESPRESSIONE + OSPITE D'ESPRESSIONE

Come è fatto un vettore d'espressione?



La maggior parte dei vettori d'espressione, come la maggior parte dei geni, sono *regolati*.



Esempio di utilizzo: *espressione* e *purificazione* di una proteina di fusione PROTEINA-TAG

1. *Clonaggio* della proteina (cDNA) taggata nel vettore d'espressione e *trasformazione* del vettore d'espressione (codificante per la proteina di fusione) in ospite.

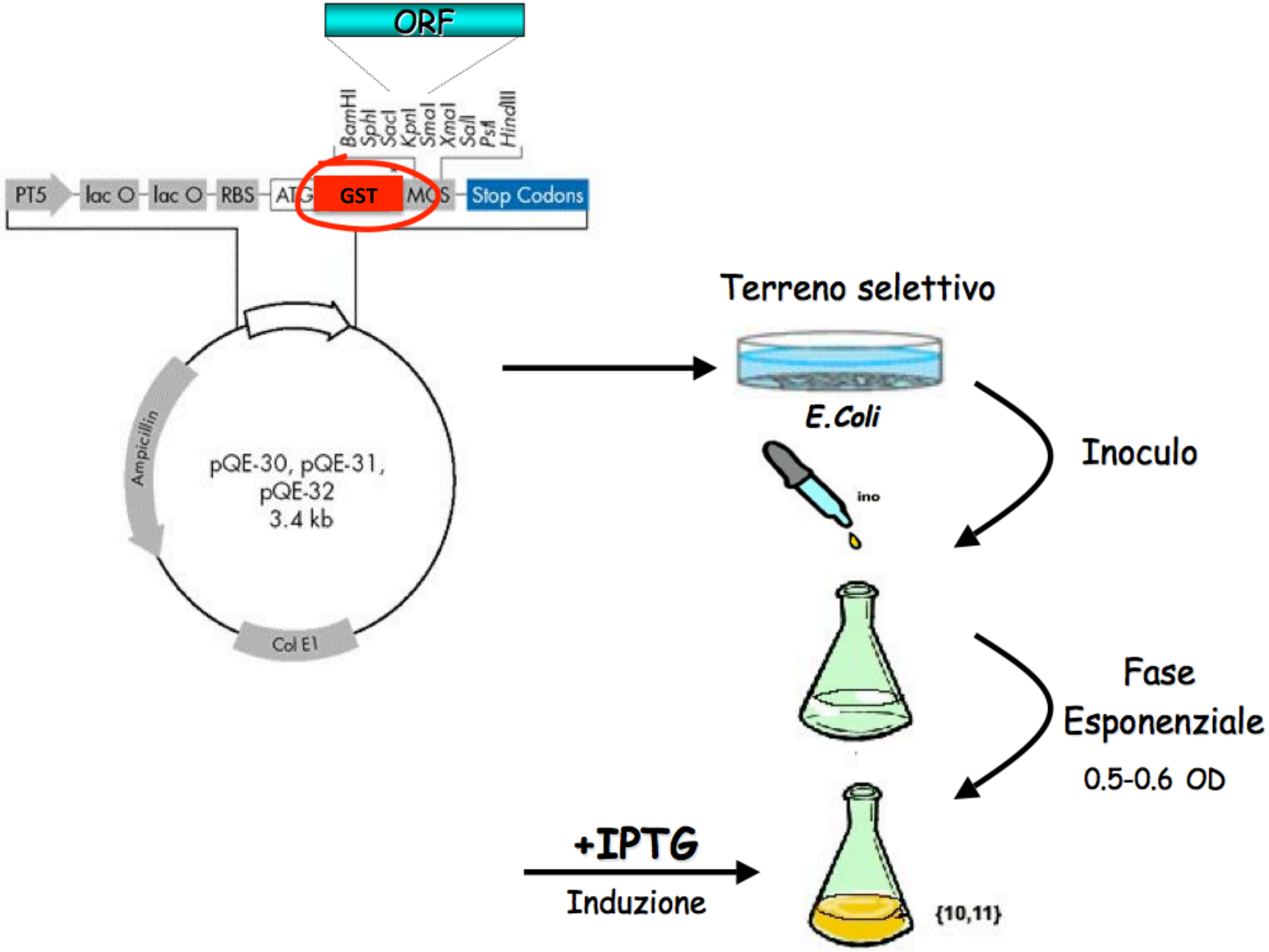
2. *Amplificazione* del ceppo batterico trasformato.

3. *Induzione* dell'espressione della proteina di fusione (lac).

4. *Purificazione* della proteina di fusione.

Esistono diversi tipi di TAG

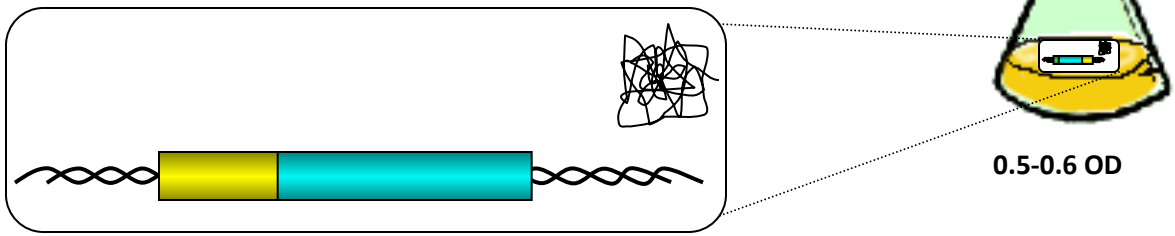
TAG	Description	Affinity ligand
His_6	6 histidines	Ni++, Co++, Cu++
GST	glutathion-S-transferase	glutathion
TAP		
Tioredoxina		
MBP	Maltose binding protein	Maltose
Protein A	Protein A	IgG
CBP (40 kDa)	Calmodulin binding protein	Calmodulin
Epitopi biotinilati		


ORF
BamHI
SphI
SacI
KpnI
SmaI
XmaI
SalI
PstI
HindIII
PT5
lac O
lac O
RBS
ATG
GST
MCS
Stop Codons
Ampicillin
pQE-30, pQE-31,
pQE-32
3.4 kb
Col E1
Terreno selettivo
E.Coli
ino
Inoculo
Fase
Esponenziale
0.5-0.6 OD
+IPTG
Induzione
{10,11}

Il Sistema di Purificazione GST

GST

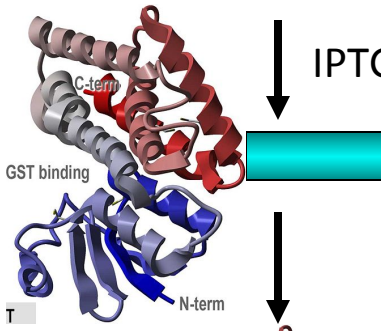
peptide d' interesse



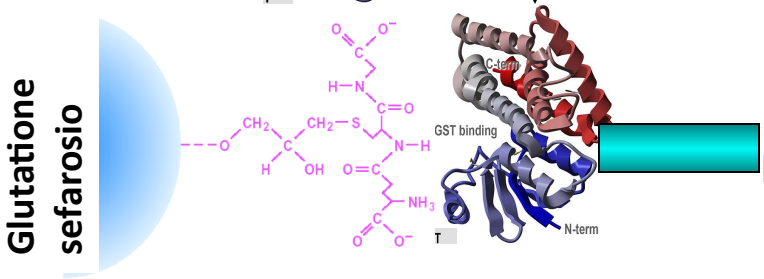
Induzione della proteina di fusione

IPTG

La proteina (blu) è taggata con GST (giallo)

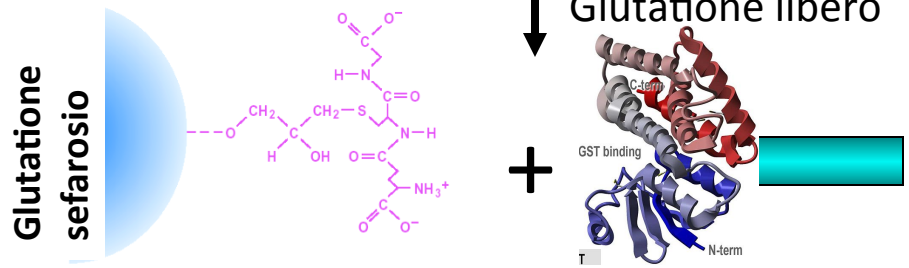


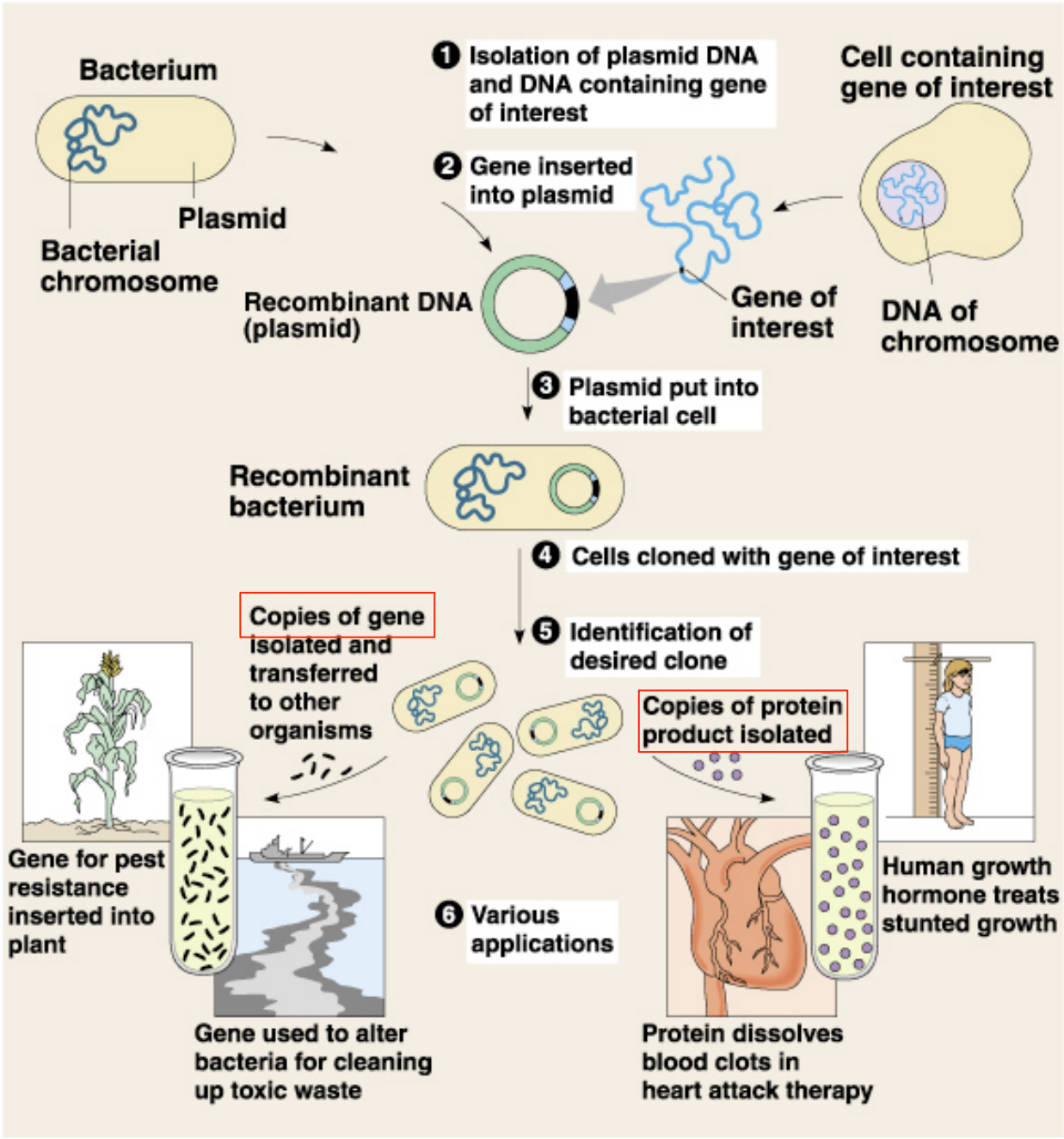
Purificazione della proteina per Cromatografia di Affinità

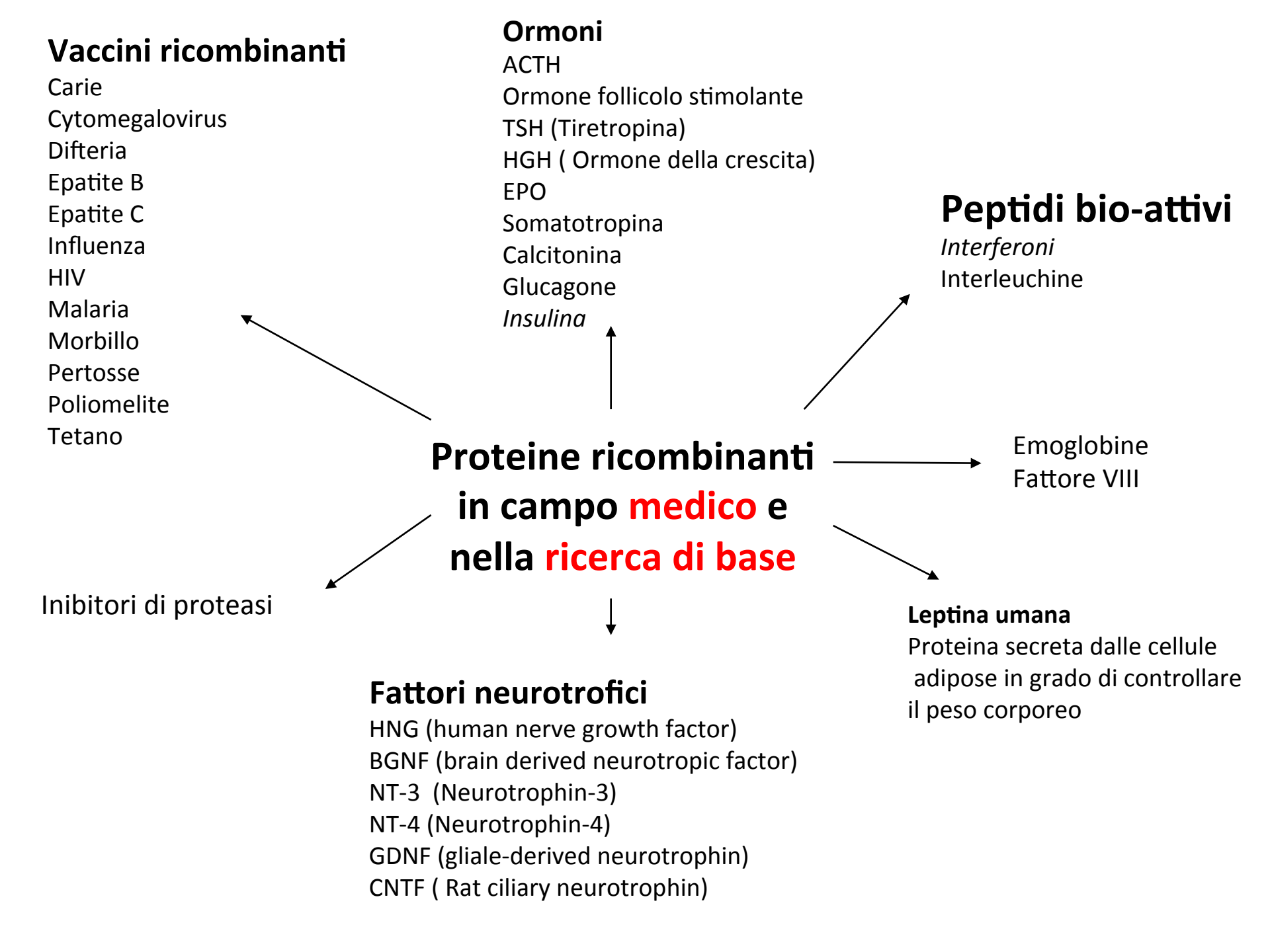


Il GST interagisce con il glutatione legato alla colonna

Eluizione con un eccesso di Glutatione libero






Vaccini ricombinanti
Carie
Cytomegalovirus
Difteria
Epatite B
Epatite C
Influenza
HIV
Malaria
Morbillo
Pertosse
Poliomelite
Tetano
Ormoni
ACTH
Ormone follicolo stimolante
TSH (Tiretropina)
HGH (Ormone della crescita)
EPO
Somatotropina
Calcitonina
Glucagone
Insulina
Peptidi bio-attivi
Interferoni
Interleuchine
Proteine ricombinanti in campo medico e nella ricerca di base
Emoglobine
Fattore VIII
Inibitori di proteasi
Fattori neurotrofici
HNG (human nerve growth factor)
BGNF (brain derived neurotropic factor)
NT-3 (Neurotrophin-3)
NT-4 (Neurotrophin-4)
GDNF (gliale-derived neurotrophin)
CNTF (Rat ciliary neurotrophin)
Leptina umana
Proteina secreta dalle cellule adipose in grado di controllare il peso corporeo

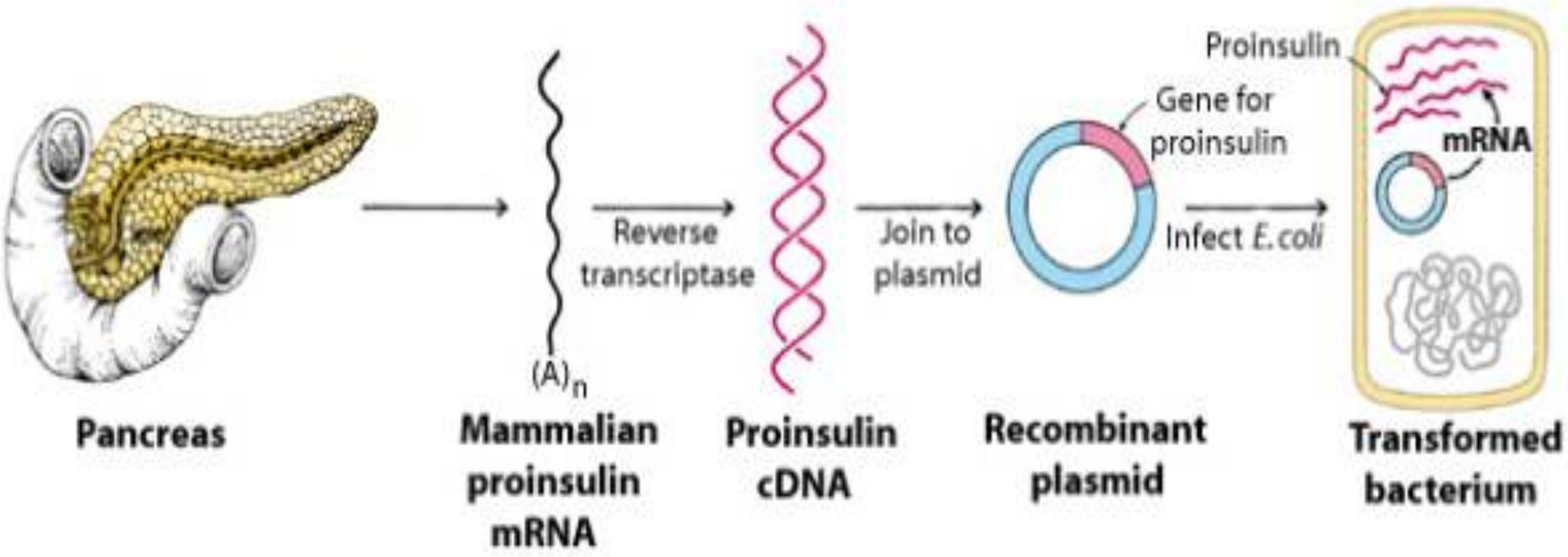
Important Biotherapeutics

- Insulin- hormone which lowers blood sugar; used by diabetics
- Interferon- class of cytokines effective *vs* viral infections
- Factor VIII- blood protein necessary for clotting; missing in hemophiliacs
- Streptokinin - bacterial enzyme to dissolve blood clots in coronary arteries
- Beta endorphins- pain suppressors

Insulin is the first recombinant protein to be produced

- **Insulin** is an important *hormone* which regulates sugar metabolism
- inability to produce insulin results in *diabetes*, this disease needs to be treated by daily injections of insulin
- Historically, insulin from pigs or cows is used, but known to produce immune reactions in some patients
- **Challenge: how to make human insulin to be used as a drug in cell systems or microbes?**

General strategy for insulin production



1. *take the gene* of human insulin, *clone* into a plasmid, introduce the plasmid into *E. coli or cells*, and use them *E.coli* as "Biological Factory" for insulin production

2. insulin produced (which contains 51 amino acids) is *identical* to the "natural human protein" and it will not cause any immune reactions

3. much more *economical* than chemical synthesis

Human growth hormone (hGH, or Somatotropin)

Human Growth Hormone (hGH) also known as somatotrophin is classified as a polypeptide hormone.

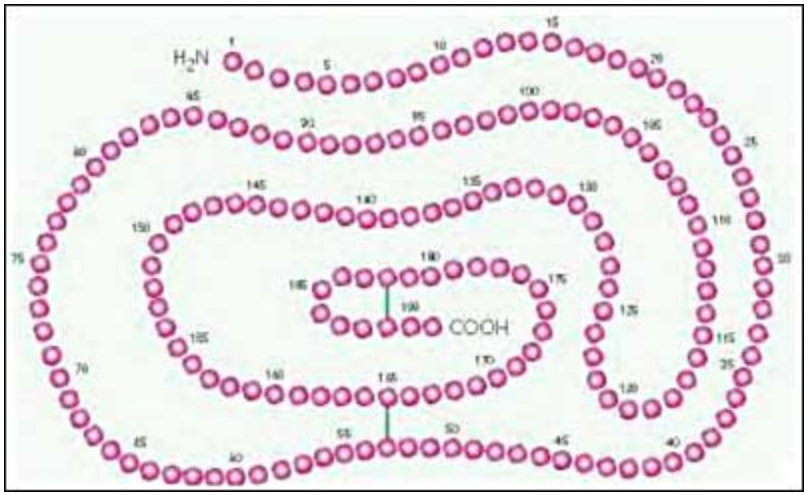
- Secreted by the pituitary gland, and is responsible for *normal body growth* and *development*, by stimulating protein production in muscle cells, energy release from the breakdown of fats and stimulates the development of bones

- These processes together are responsible for *longitudinal* growth. Inadequate production of GH results in short stature, defined as a below normal height for a given age

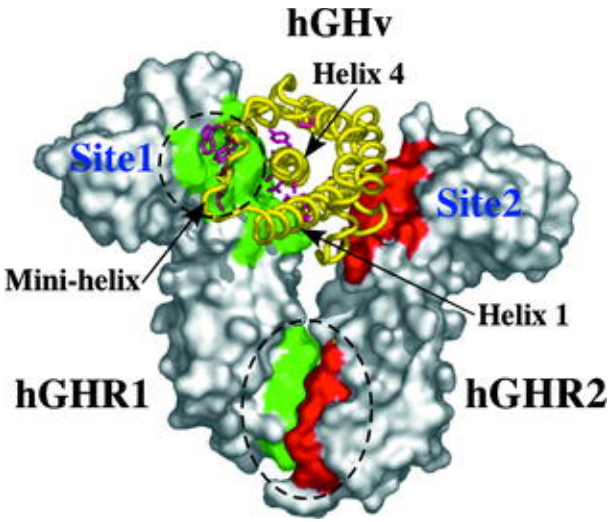
- In children and adolescents, the rate of growth in height is primarily determined by the rate at which endogenous GH is secreted

- Under normal conditions, GH secretion and growth rate remain increased until final height is reached, after which GH secretion is reduced to a steady state

Structure of human Growth hormone

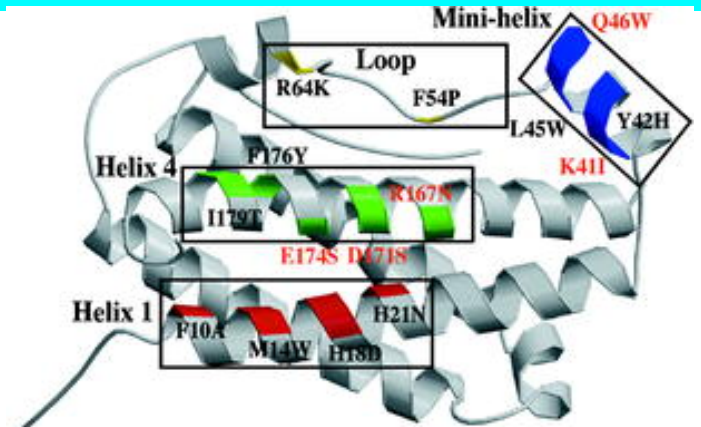


Primary structure



Hormone binding to receptors

Growth Hormone: 191 amino acids, single chain



Teritiary structure

Growth hormone deficiency

GH deficiency is one of the many causes of *short* stature and *dwarfism*. It results primarily from damage to the hypothalamus or to the pituitary gland during fetal development (*congenital*) or following birth (*acquired*). GH deficiency may also be caused by mutations in genes that regulate its synthesis and secretion, as for PIT-1 (pituitary-specific transcription factor-1) and POUF-1 (prophet of PIT-1). The one thing every Dwarf has in common is of course shortness but other problems can become apparent because of this stature including:

- delayed development of motor skills
- breathing problems caused by small cheats
- weight problems
- curvature in spine
- bowed legs
- trouble with joint flexibility and early arthritis
- lower back pain and led numbness
- crowding of teeth

Growth hormone deficiency

GH deficiency is most often treated with *injections* of GH. For decades, however, availability of the hormone was limited, because it was obtained solely from *human cadaver* pituitaries.

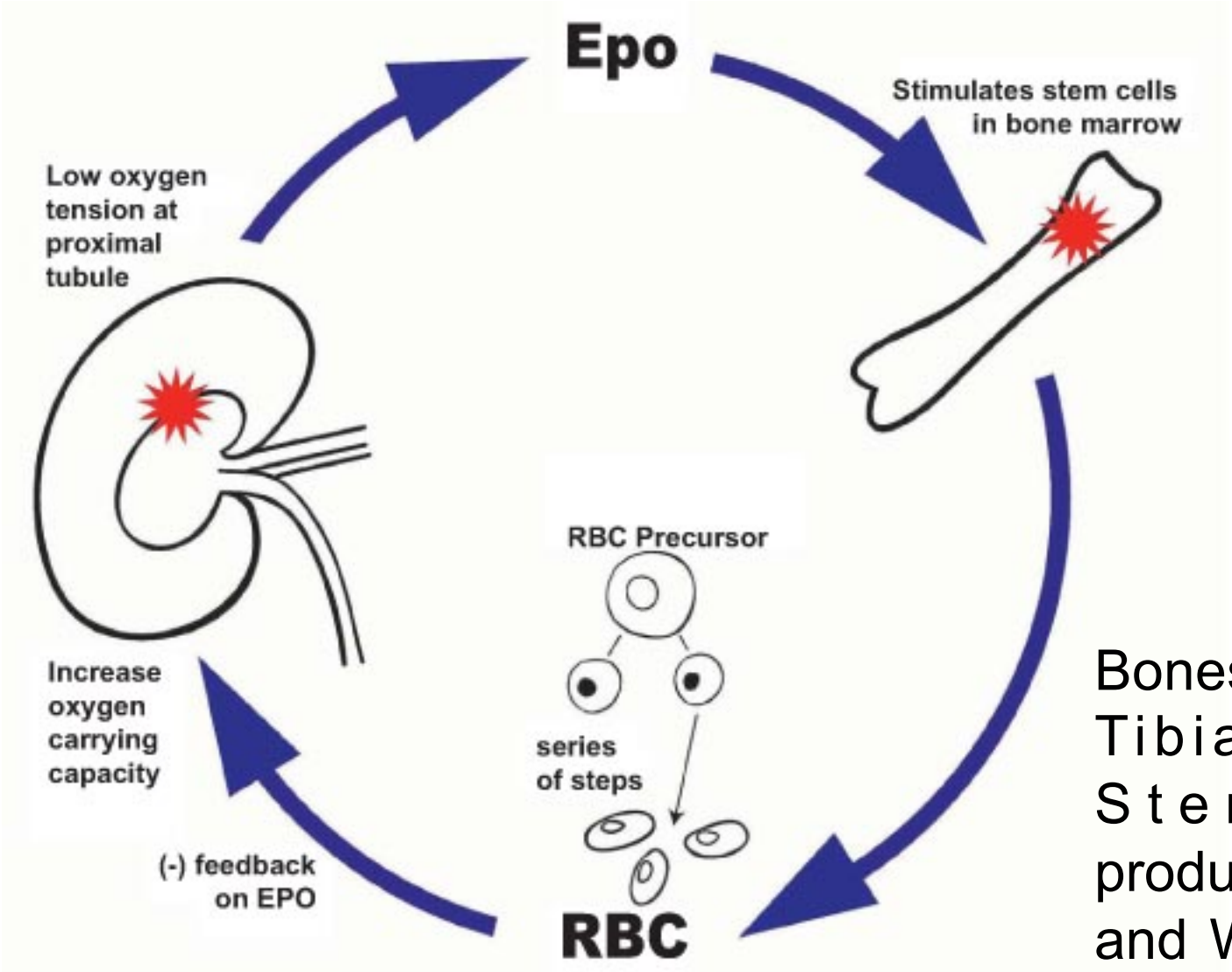
In 1985, use of natural GH was halted in the US and several other countries because of the possibility that the hormone was *contaminated* with a type of pathogenic agent known as a prion, which causes a fatal condition called Creutzfeldt-Jakob disease.

That same year, by means of recombinant DNA technology, scientists were able to produce a *biosynthetic human* form, which they called **somatrem**, thus assuring a virtually unlimited supply of this once-precious substance.

Erythropoietin

- Human **Erythropoietin** is produced in kidney
- A glycoprotein, acts on the bone marrow to increase the production of red and white blood cells. Stimuli such as bleeding or moving to high altitudes (where oxygen is scarce) trigger the release of erythropoietin
- Known as EPO, MW 30400 Kda, 165 amino acids in human (192 Mouse)

Kidney is the principal production site of Erythropoietin



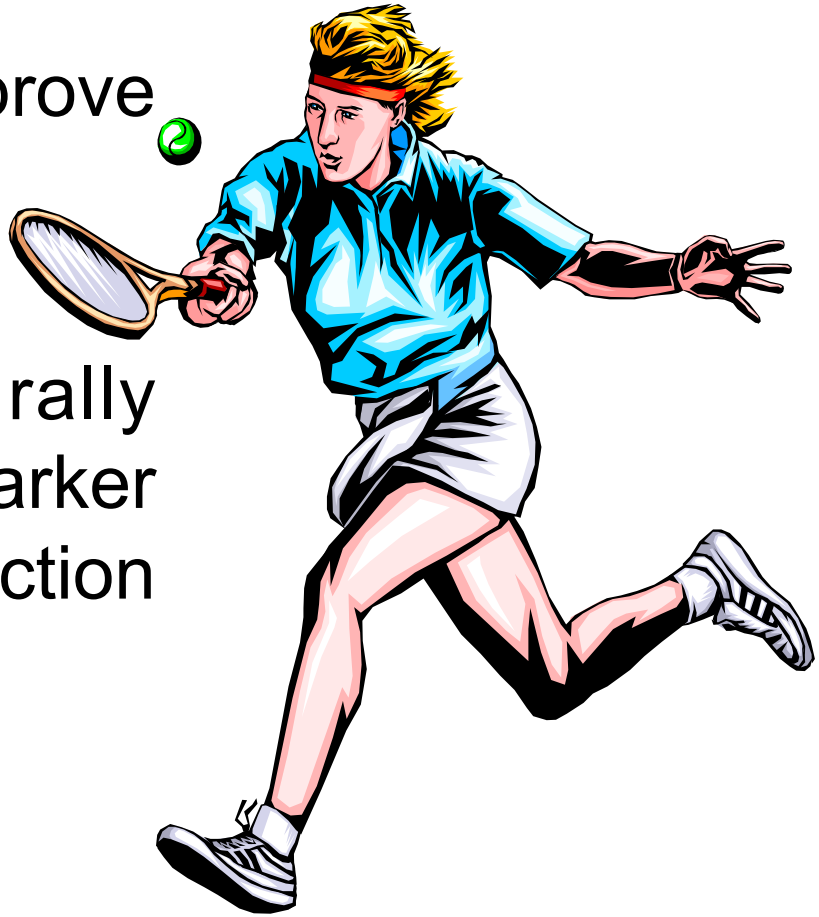
Bones like Femur, Tibia, Vertebra, Sternum, Rib produce most RBC and WBC under the influence of EPO

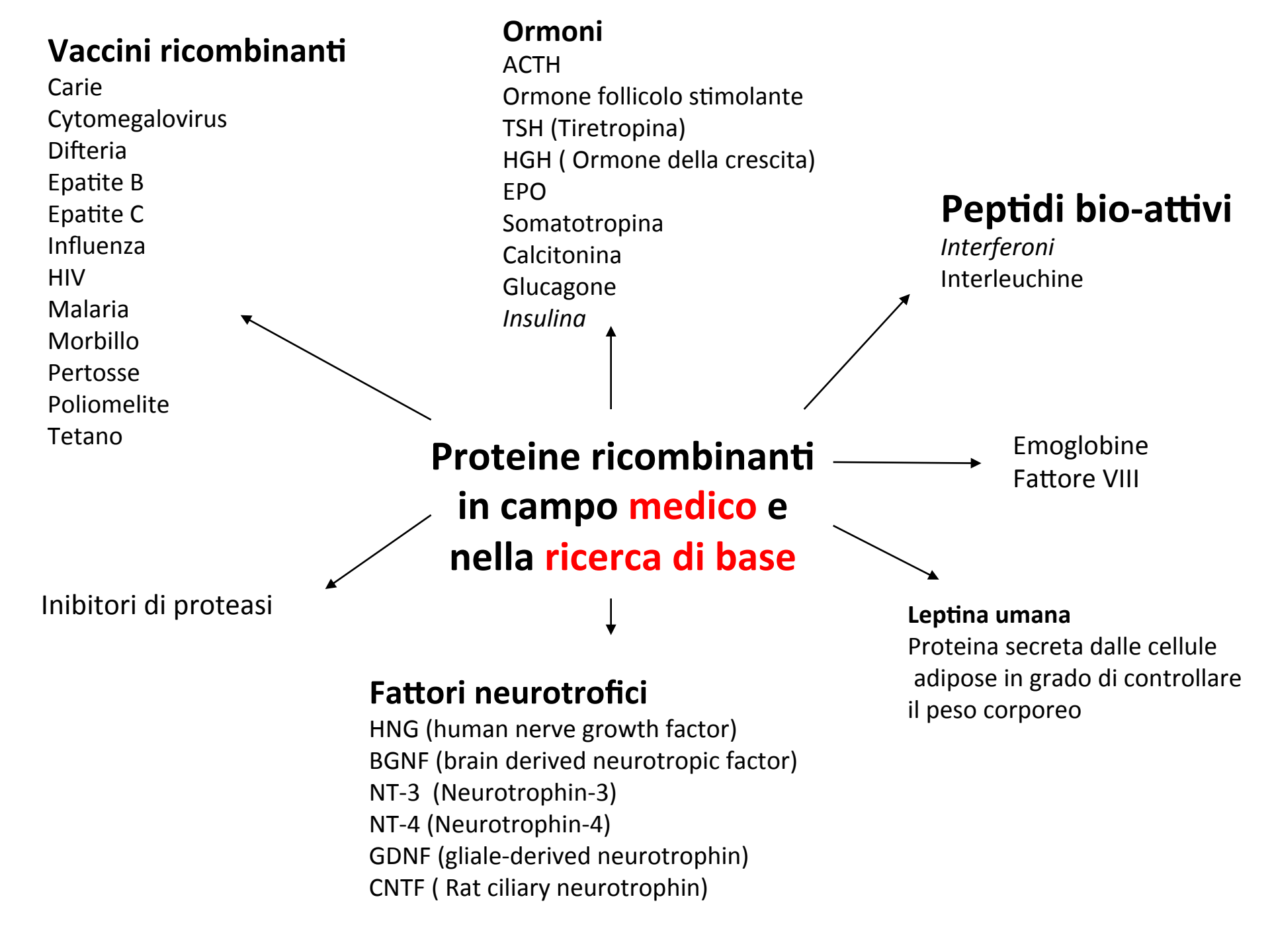
Flow Chart of Production Process

Vial (Cells) → Flask/ Spinner Bottle → Roller Bottle / Bio-reactor → Production → Purification → **Final Bulk EPO concentrate**

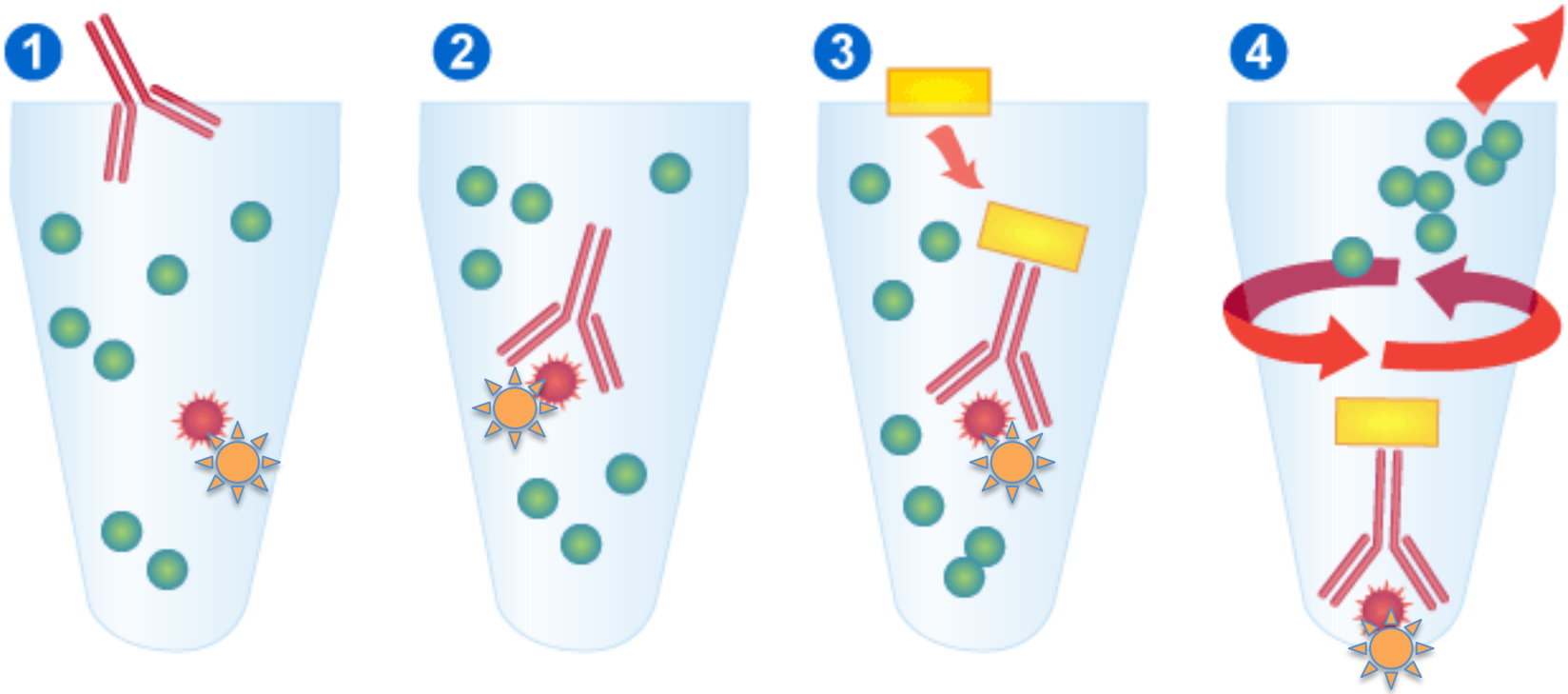
EPO also has therapeutic abuses

- Used in sports to improve endurance
- Now detected from naturally occurring EPO by protein marker produced during post injection phase




Vaccini ricombinanti
Carie
Cytomegalovirus
Difteria
Epatite B
Epatite C
Influenza
HIV
Malaria
Morbillo
Pertosse
Poliomelite
Tetano
Ormoni
ACTH
Ormone follicolo stimolante
TSH (Tiretropina)
HGH (Ormone della crescita)
EPO
Somatotropina
Calcitonina
Glucagone
Insulina
Peptidi bio-attivi
Interferoni
Interleuchine
Proteine ricombinanti in campo medico e nella ricerca di base
Emoglobine
Fattore VIII
Inibitori di proteasi
Fattori neurotrofici
HNG (human nerve growth factor)
BGNF (brain derived neurotropic factor)
NT-3 (Neurotrophin-3)
NT-4 (Neurotrophin-4)
GDNF (gliale-derived neurotrophin)
CNTF (Rat ciliary neurotrophin)
Leptina umana
Proteina secreta dalle cellule adipose in grado di controllare il peso corporeo

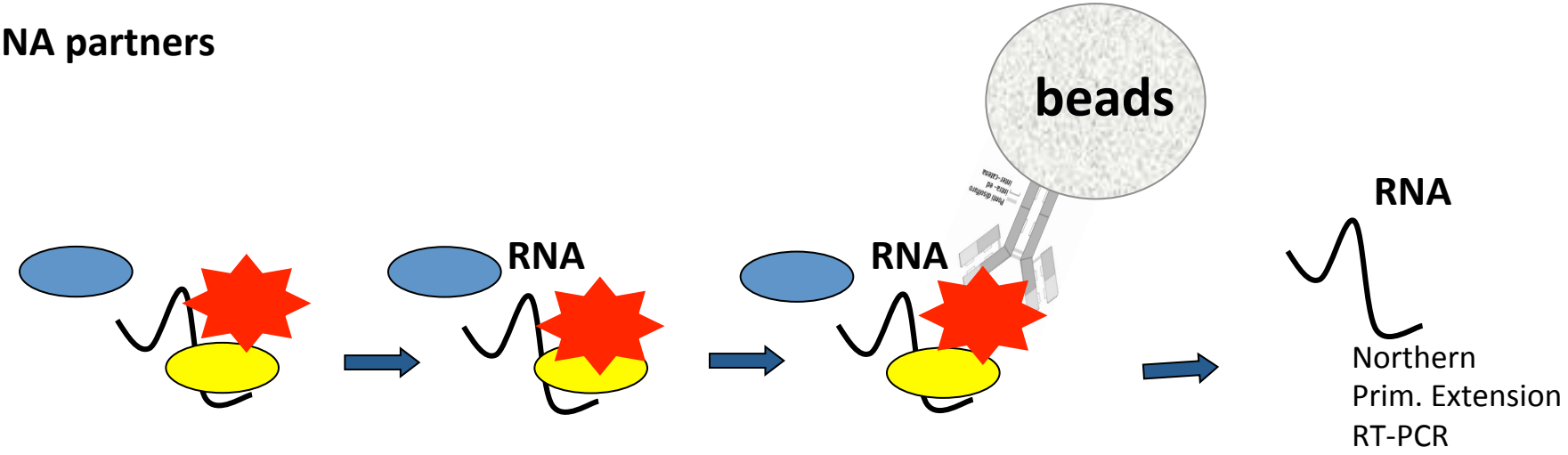
Co-Immunoprecipitation (IP)



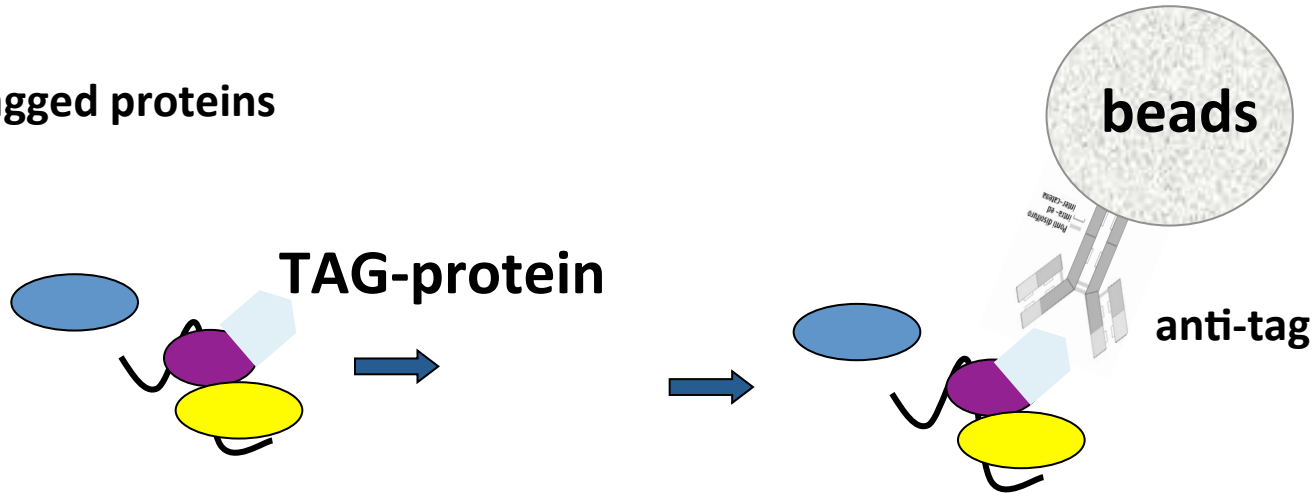
1. Suitable antibody is added.
2. Antibody binds to protein of interest.
3. Protein A or G added to make antibody-protein complexes insoluble.
4. Centrifugation of solution pellets antibody-protein complex.
   Removal of supernatant and washing.

Co-immunoprecipitation themes

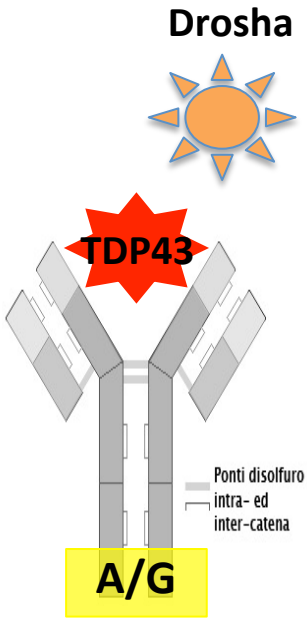
- **RNA partners**

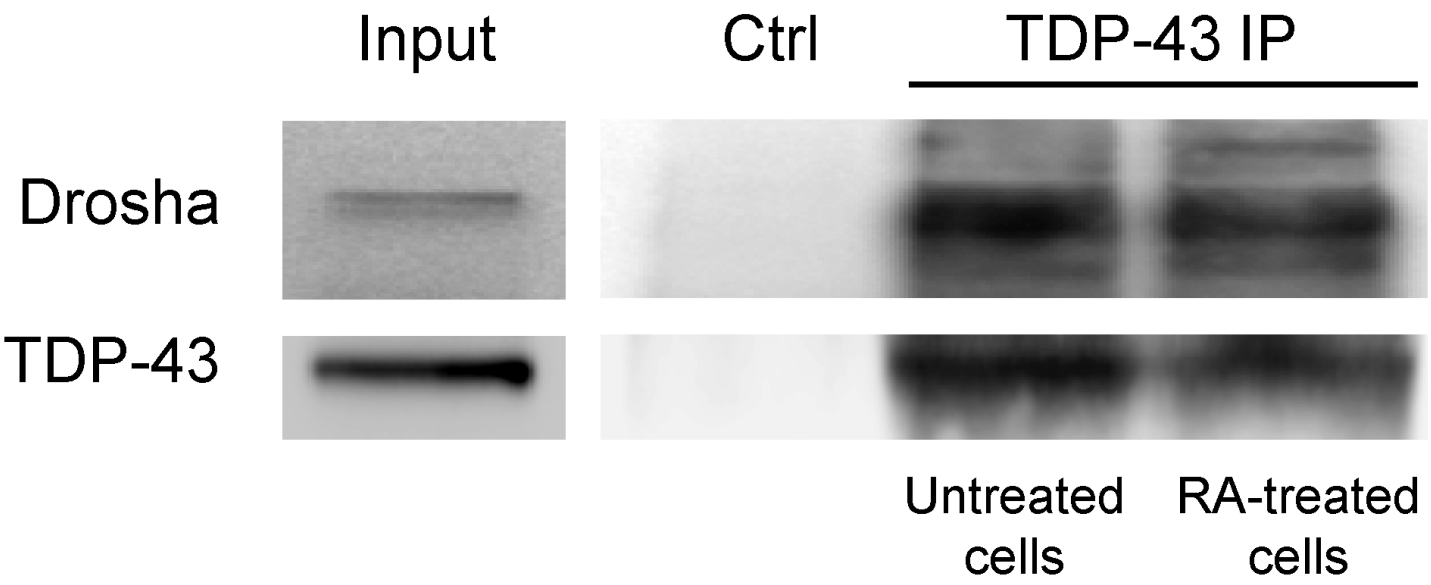


- **Tagged proteins**



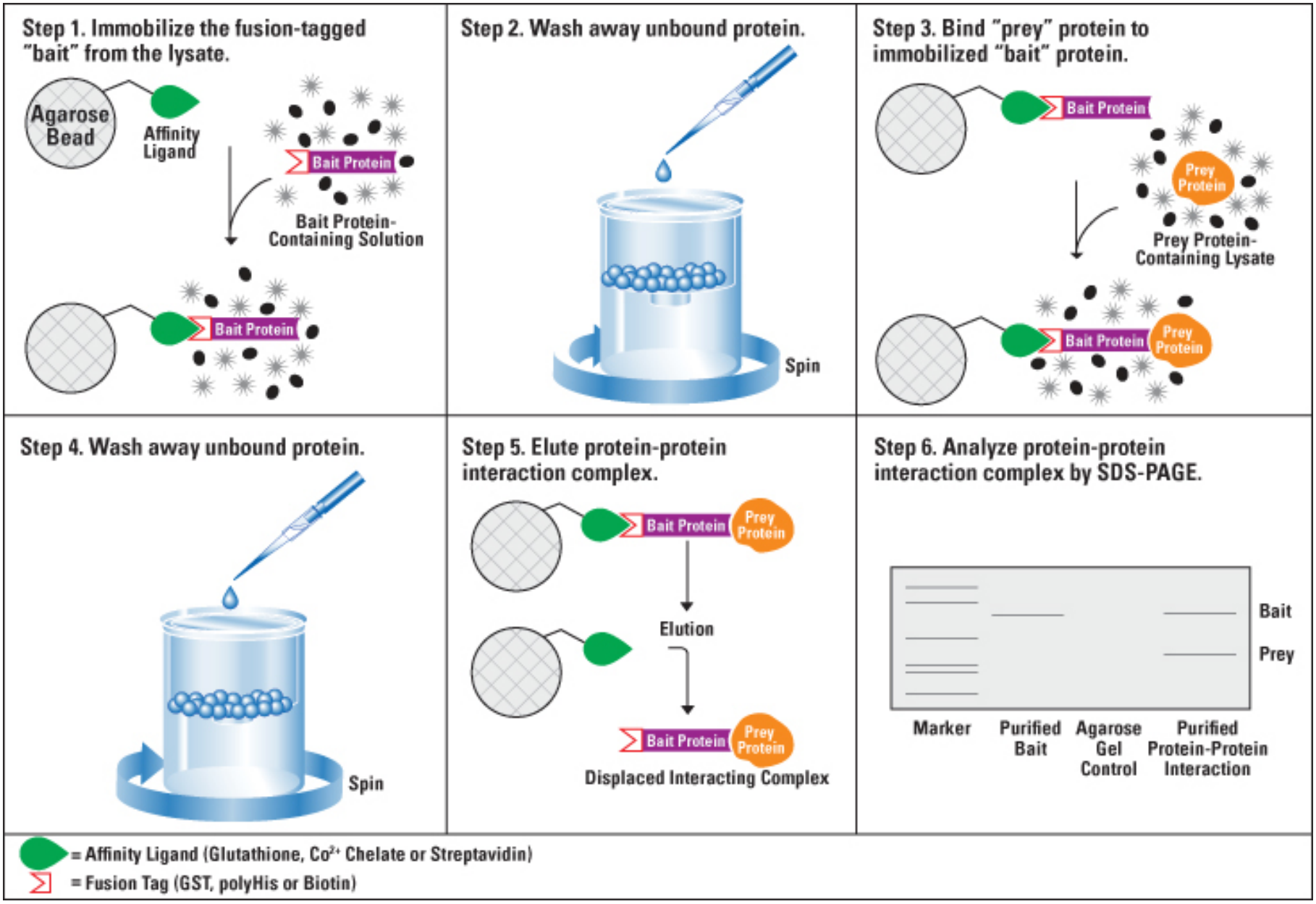
Starting material: sknBE TOTAL EXTRACT



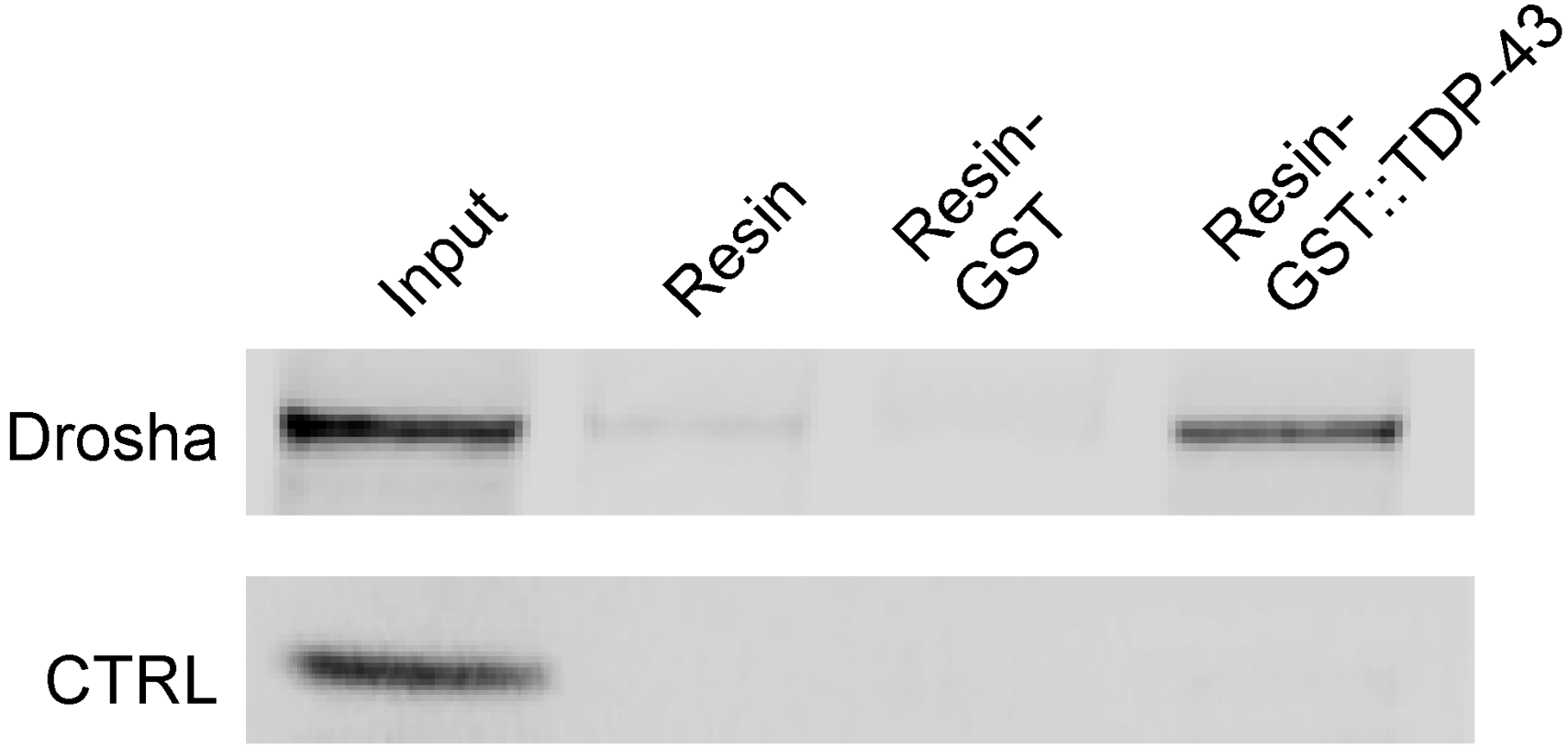


Di Carlo V. et al, Molecular Neurob. 2013

Pull-down assay



Pull-down assay



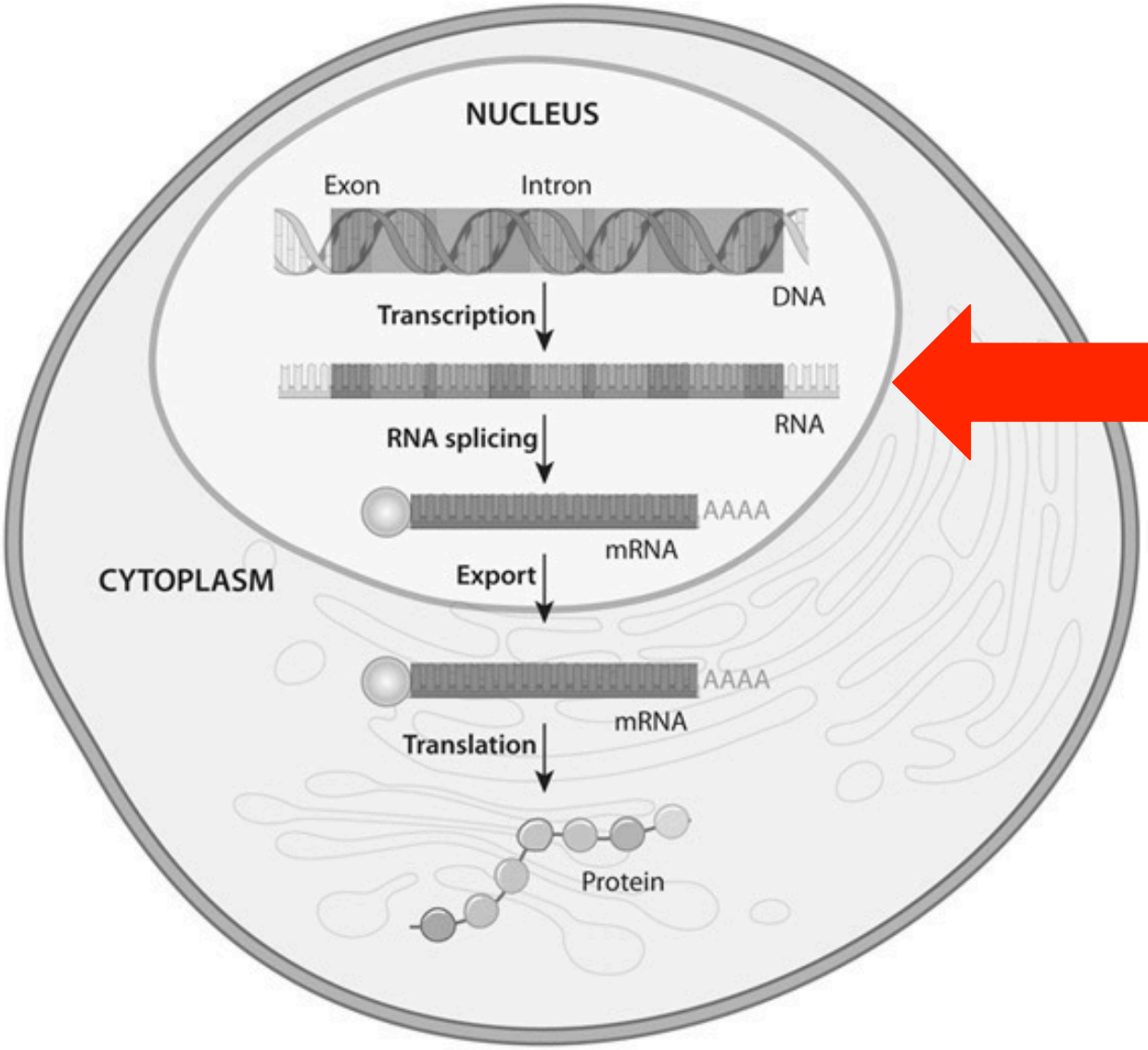
Di Carlo V. et al, Molecular Neurob. 2013

RNA immunoprecipitation (RIP)

CrossLinking and ImmunoPrecipitation (CLIP)

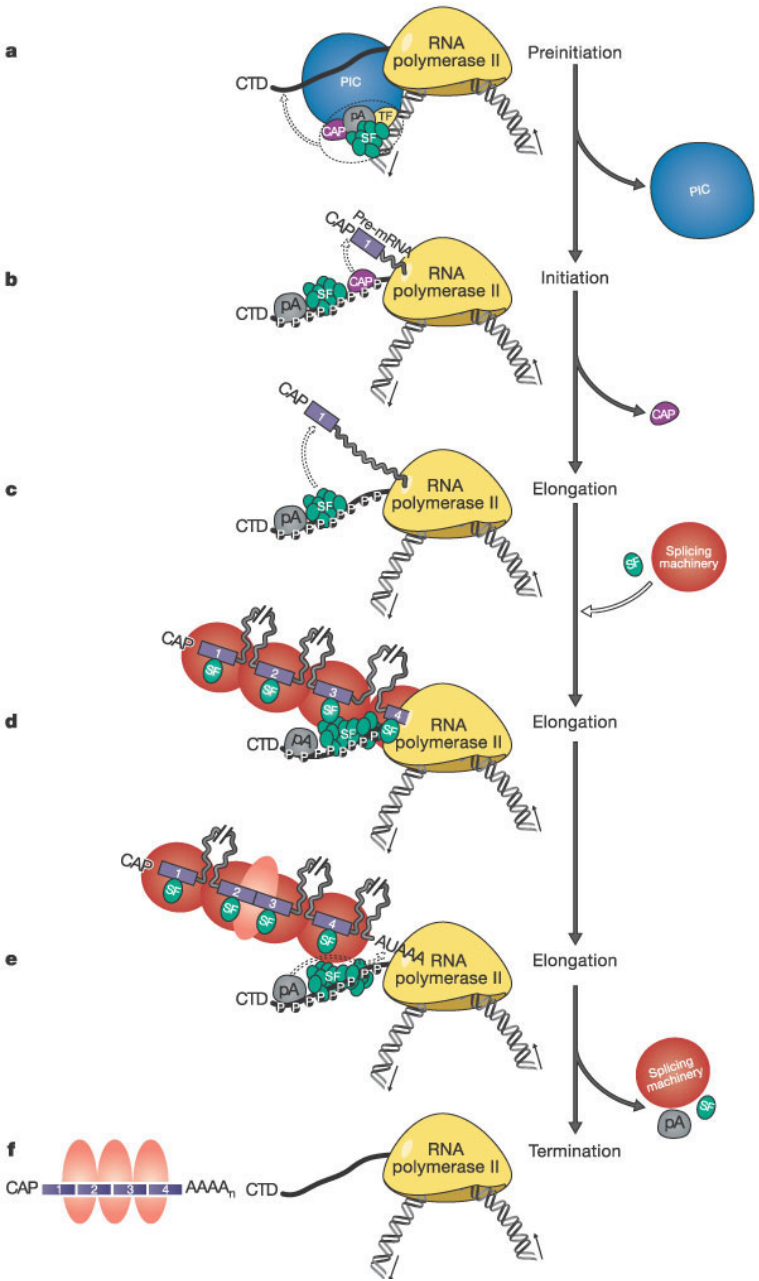
Chromatin Immunoprecipitation (ChIP)

Chromatin Isolation by RNA Purification (ChIRP)



La **Trascrizione** come prima tappa di espressione dell'informazione contenuta nella sequenza di DNA

During and after transcription, RNAs are subject to multiple processing steps coordinated by RNA-binding proteins…

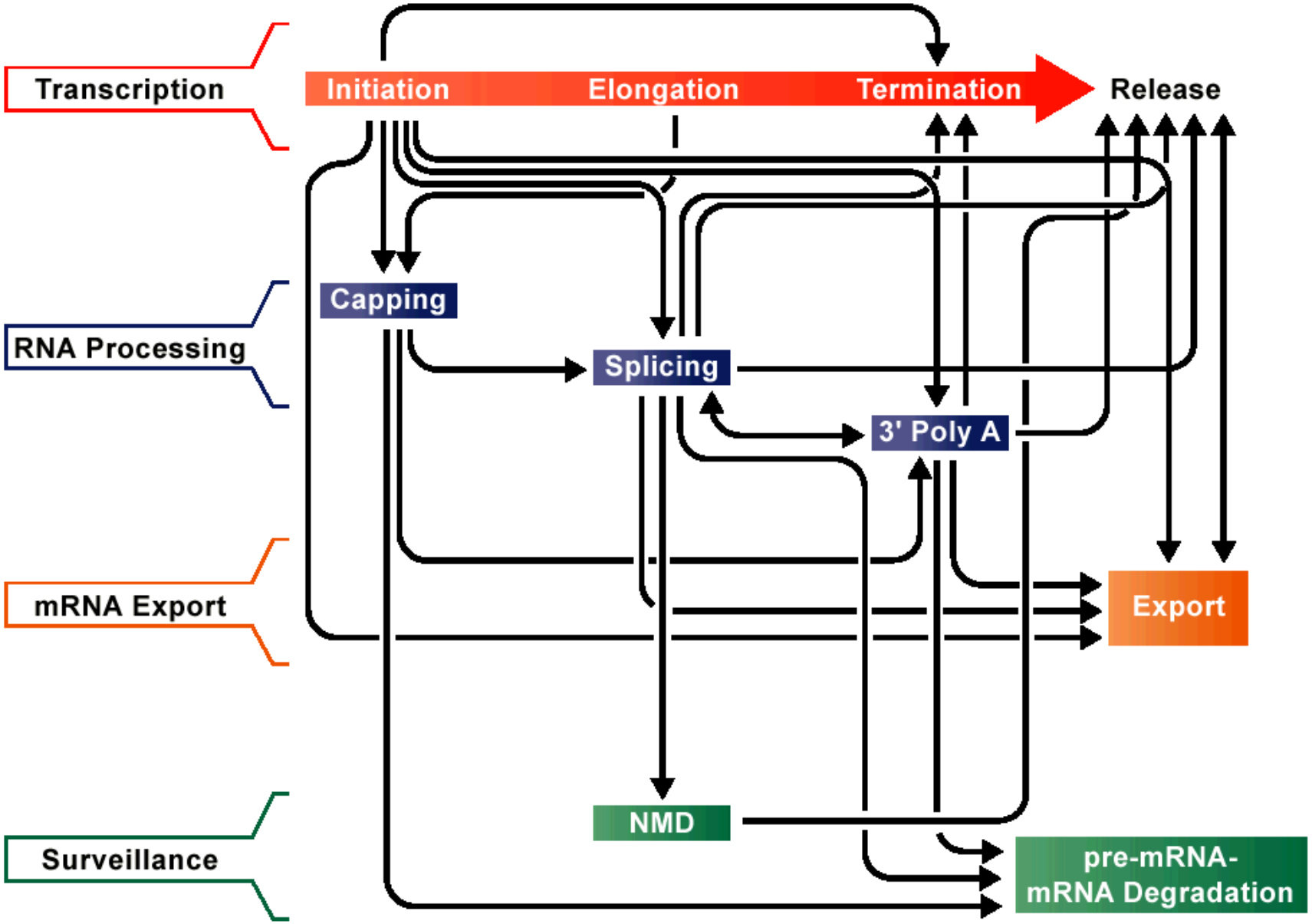


*…therefore a key task is to **map protein–RNA interactions** and to determine their effects on the transcriptome (**Ribonomics**)*

Nature Reviews Genetics **13**, 77-83

Tom Maniatis and Robin Reed
Nature **416**, 499-506 (4 April 2002)

Network of *coupled* interactions in gene expression



The “mRNA Factory” model

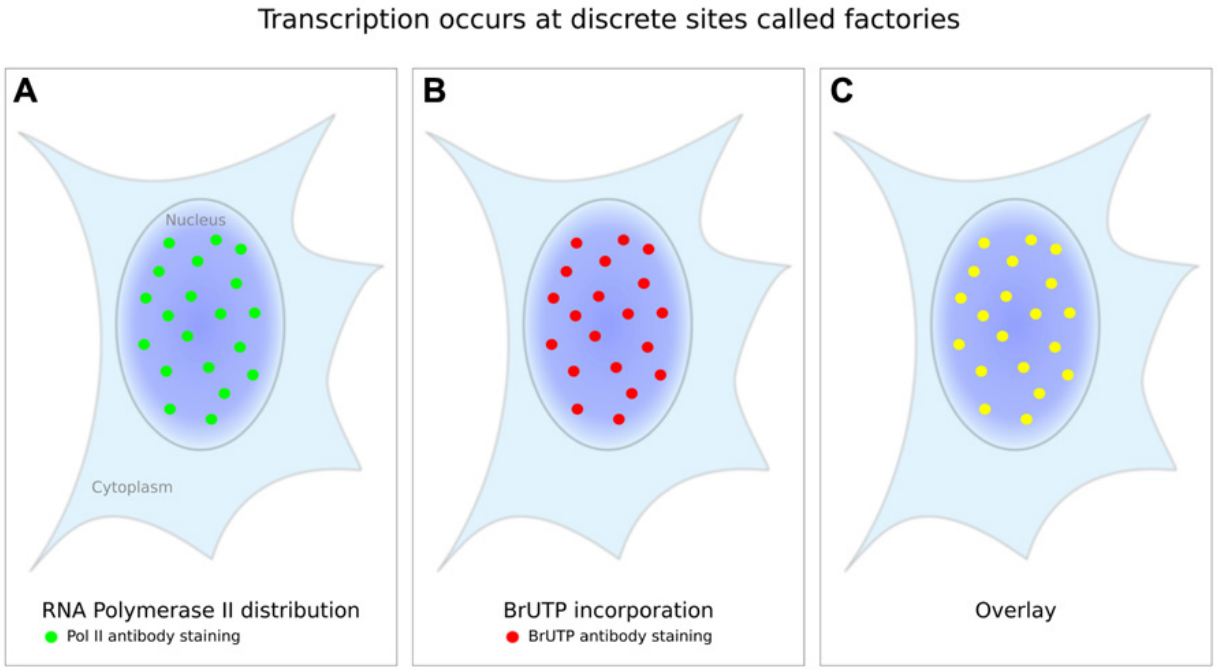
High throughput microarrays and next-generation sequencing technologies have revealed:

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).

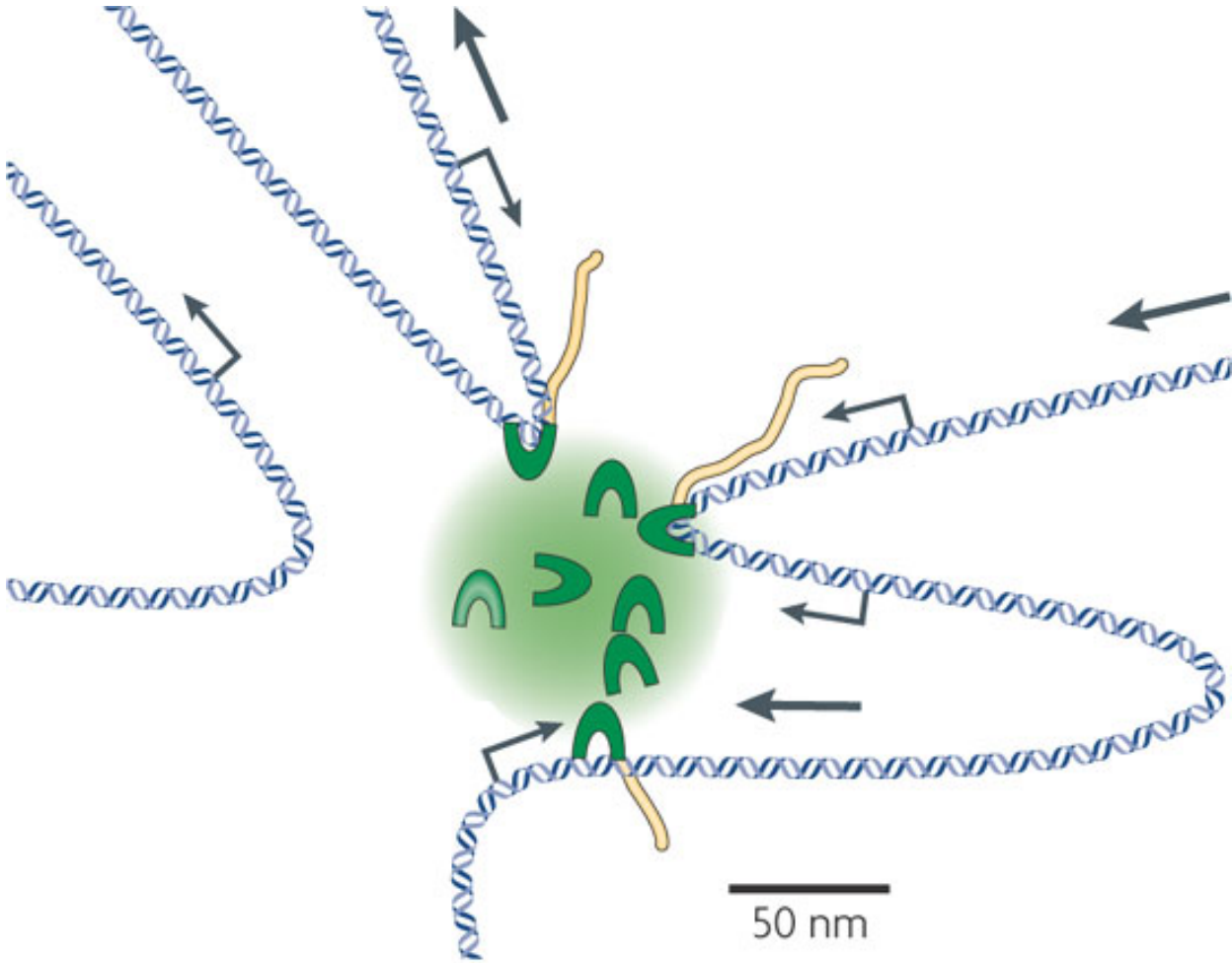
These discrete sites of active transcription are referred to as “***transcription factories***”.



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.

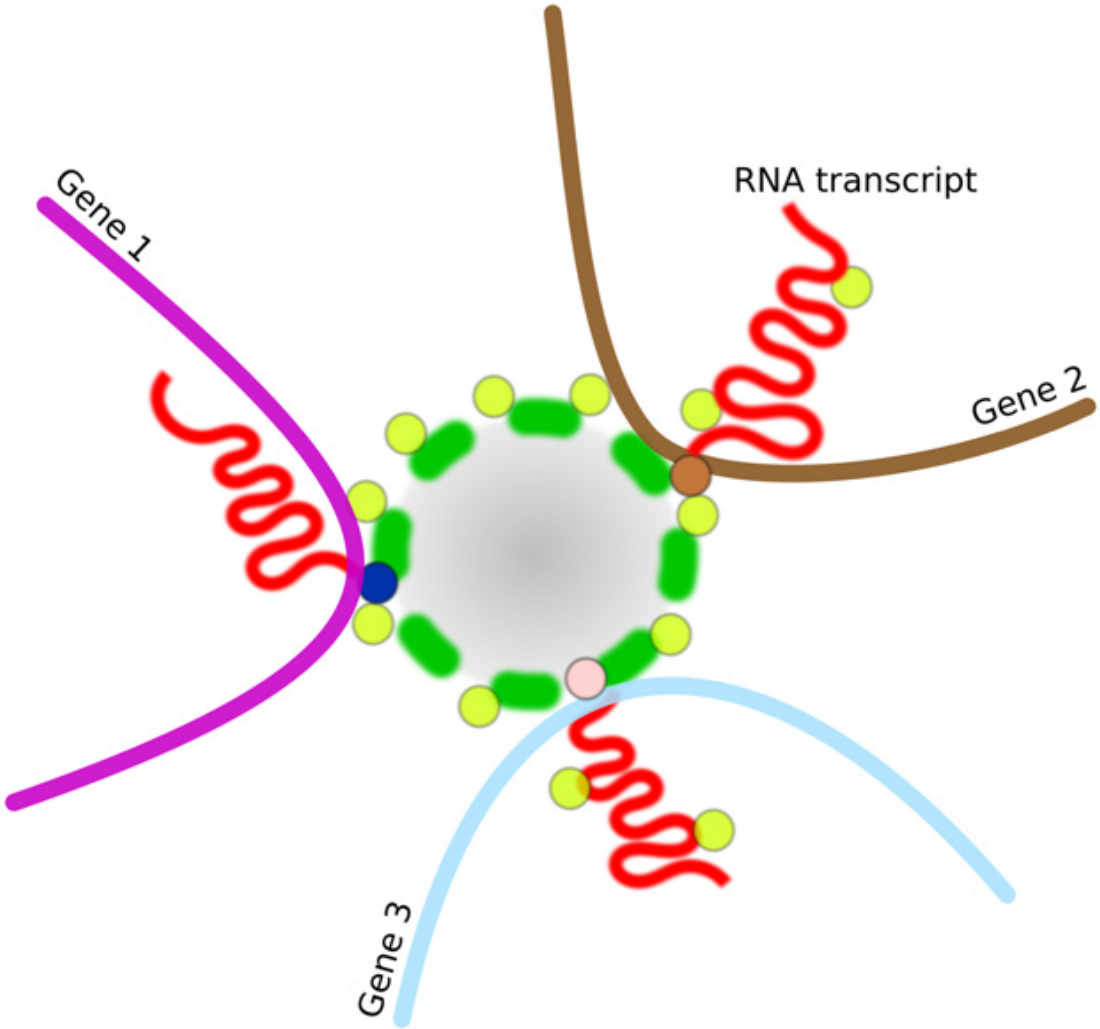
Genes from the *same* or from *different* chromosomes may associate with polymerases in the same factory. Small arrows indicate the direction of transcription at the transcription start site.



Nature Reviews | Genetics

Structure of a transcription factory

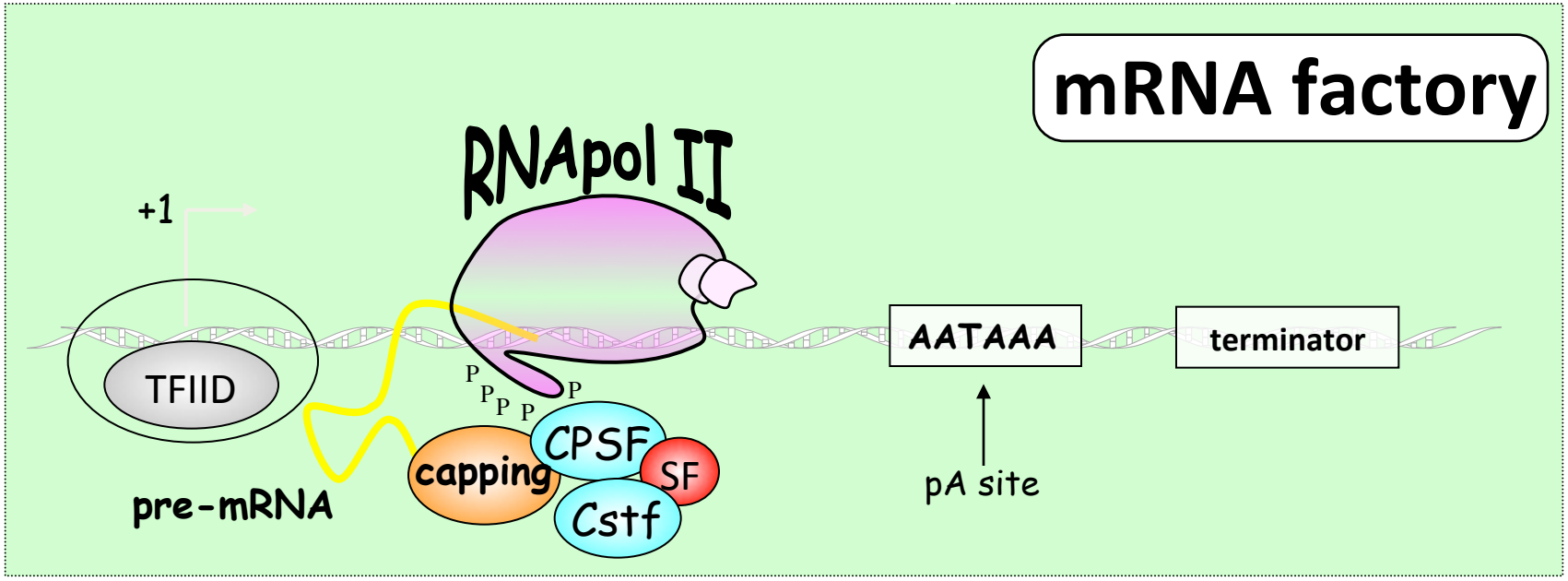
Each factory contains RNA polymerase II molecules which are located on the surface of a *protein-rich core* (87 nm in diameter, as determined by EFTEM in HeLa cells). These proteins include many factors involved in *transcription* such as co-activators, *chromatin remodelers*, transcription factors,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

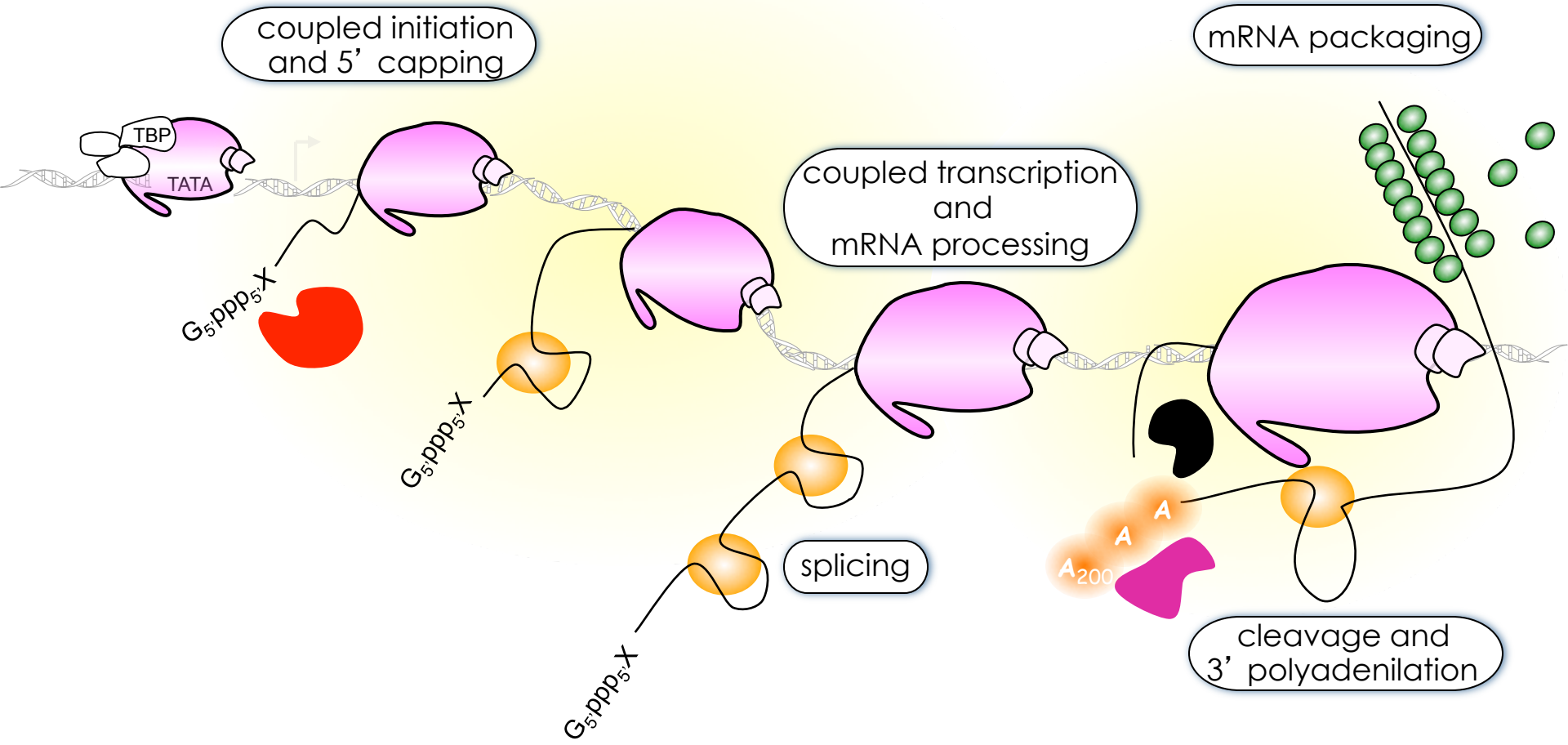


By *tethering* machines to each other and to their substrates *coupling* plays a critical role in gene expression dramatically increasing the *specificity* of enzymatic reactions.

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

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



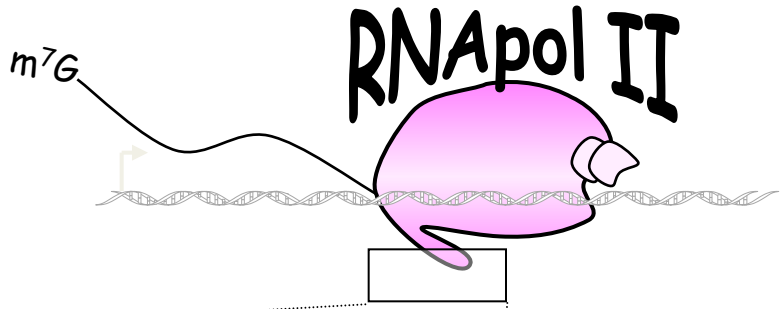


The fate of a specific mRNA is determined at the beginning of

transcription

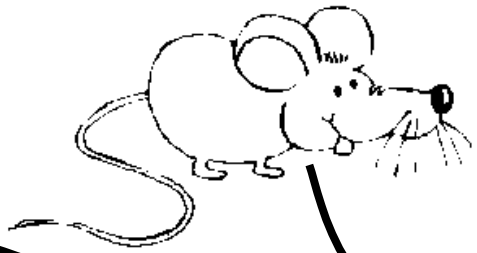
The **RNA polymerase II** has developed a structure composed of repeats of a 7 amino-acid sequence.
In humans this structure – termed “carboxyterminal domain” or CTD of the Po lII largest subunit
– is composed of 52 such repeats. It is placed exactly at the position where RNA emerges from RNA polymerase II.

In less complex organisms the CTD is much shorter: a worm has 36 repeats, and yeast as few as 26, but many single-cell organisms and bacteria have never developed an obvious CTD structure.



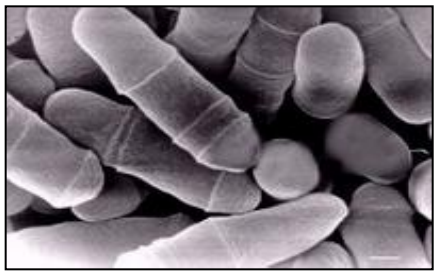
Carbossi Terminal Domain

Mouse



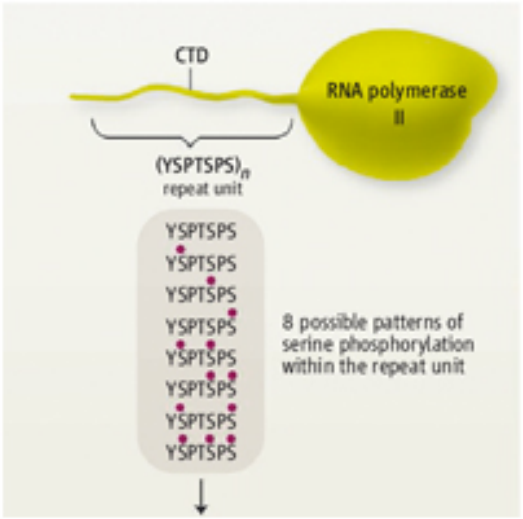
	1	2	3	4	5	6	7
	Glu	Gly	Ala	Met	Ser	Pro	Ser
1.	Tyr	Ser	Pro	Thr	Ser	Pro	Ala
2.	Tyr	Glu	Pro	Arg	Ser	Pro	Gly Gly
3.	Tyr	Thr	Pro	Gln	Ser	Pro	Ser
4.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
5.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
6.	Tyr	Ser	Pro	Thr	Ser	Pro	Asn
7.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
8.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
9.	Tyr	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

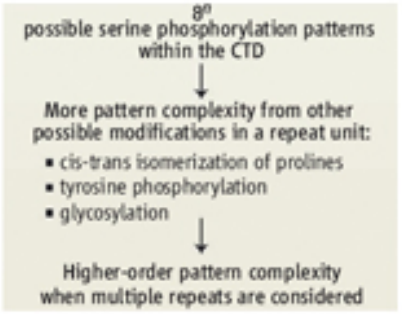
Yeast

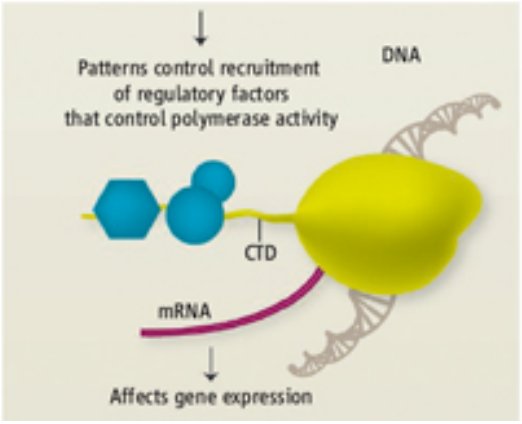


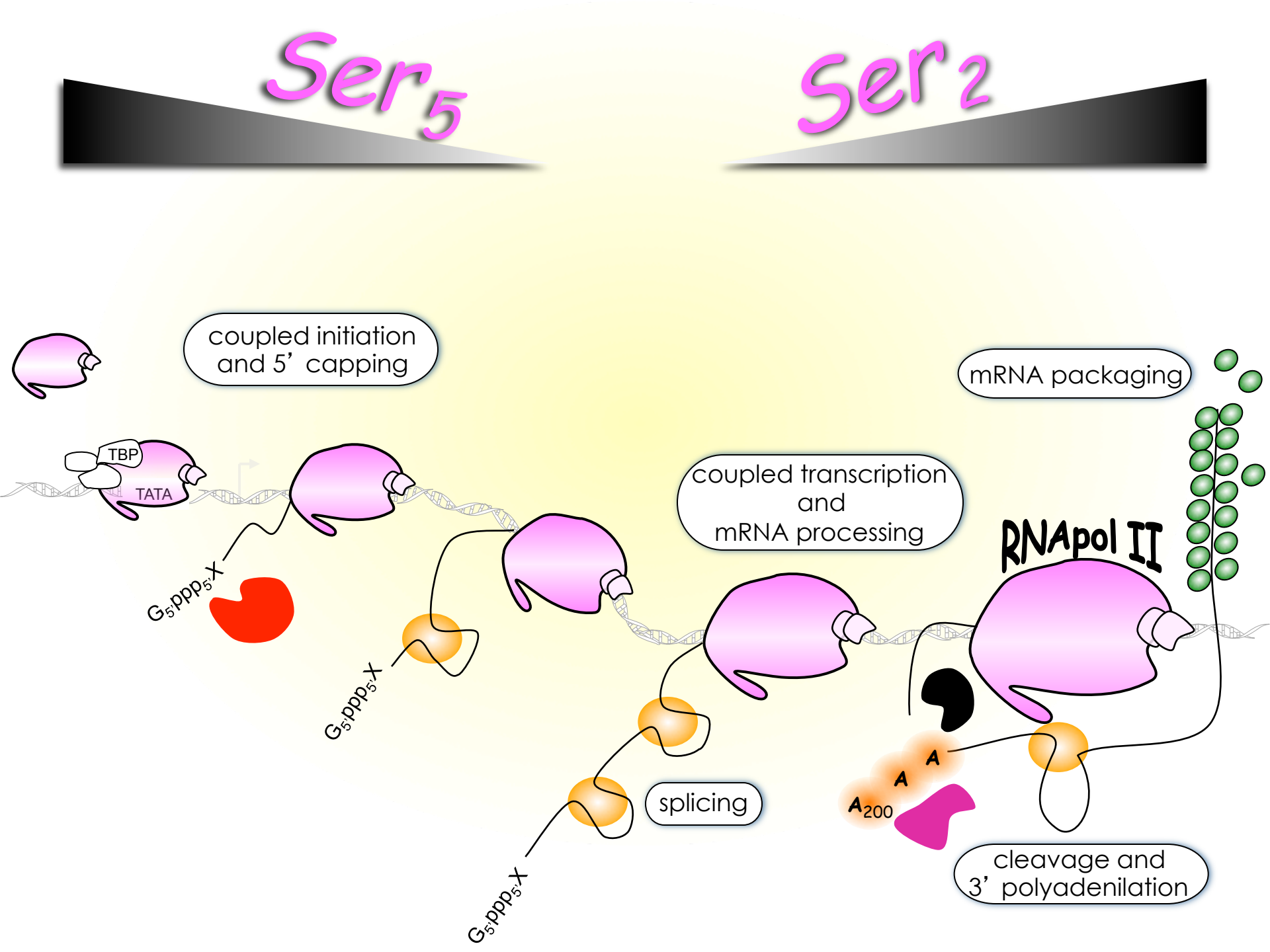
	1	2	3	4	5	6	7
	Glu	Ala	Pro	Thr	Ser	Pro	Gly
1.	Phe	Gly	Val	Ser	Ser	Pro	Gly
2.	Phe	Ser	Pro	Thr	Ser	Pro	Thr
3.	Tyr	Ser	Pro	Thr	Ser	Pro	Ala
4.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
5.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
6.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
7.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
8.	Tyr	Ser	Pro	Thr	Ser	Pro	Ser
9.	Tyr	Ser	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

Patters of *phosphorylation* in the CTD domain of RNA polymerase II may constitute a ***code*** that controls the recruitment of regulatory factors to control gene expression.

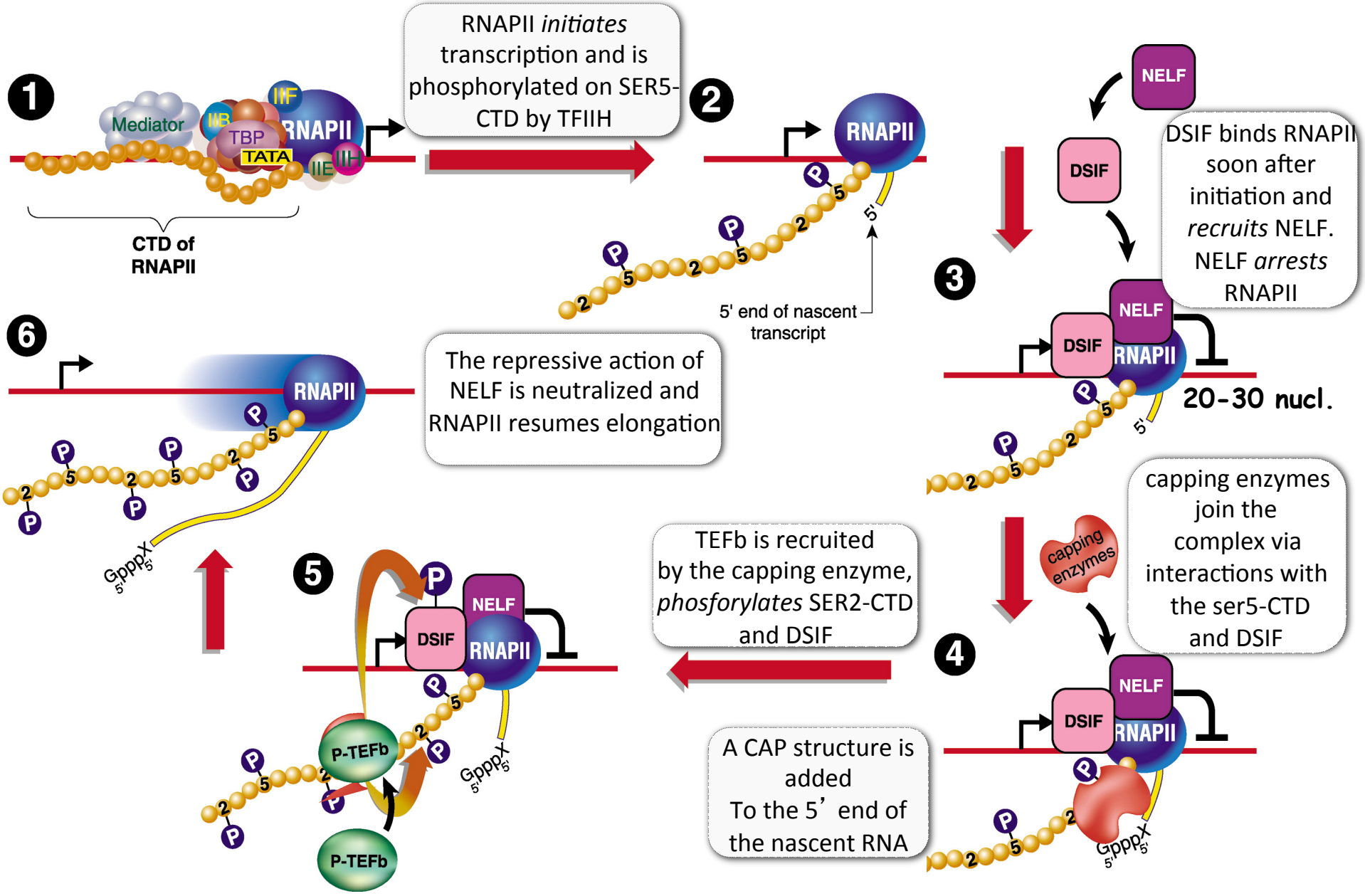







Ser5
Ser2
coupled initiation and 5' capping
TBP
TATA
G5'ppp5'X
G5'ppp5'X
G5'ppp5'X
coupled transcription and mRNA processing
RNApol II
mRNA packaging
splicing
A
A
A200
cleavage and 3' polyadenilation

Checkpoint model for coupling 5' pre-mRNA capping and transcription elongation



RNA immunoprecipitation (RIP)

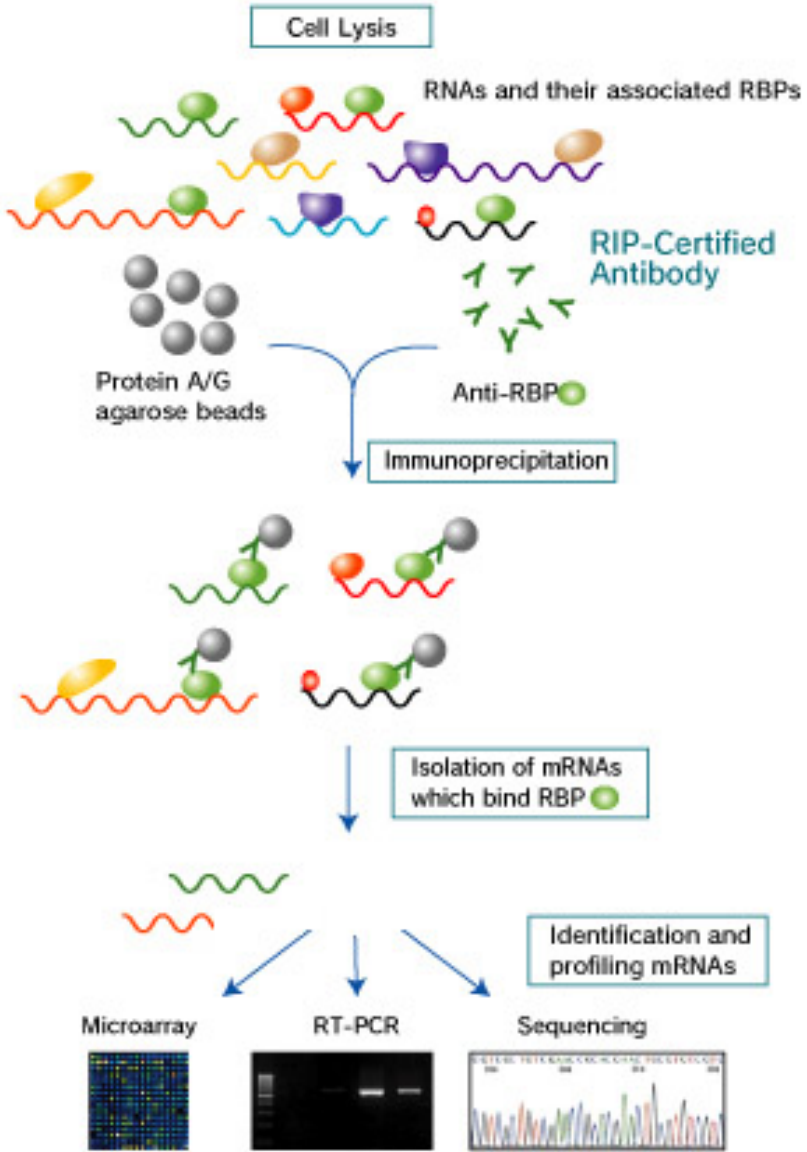
CrossLinking and ImmunoPrecipitation (CLIP)

Chromatin Immunoprecipitation (ChIP)

Chromatin Isolation by RNA Purification (ChIRP)

*The first approach used to identify **RNAs that are bound by specific ribonucleoproteins (RBPs)** was RNA-IP (RIP)*

RNA-Immunoprecipitation (RIP)

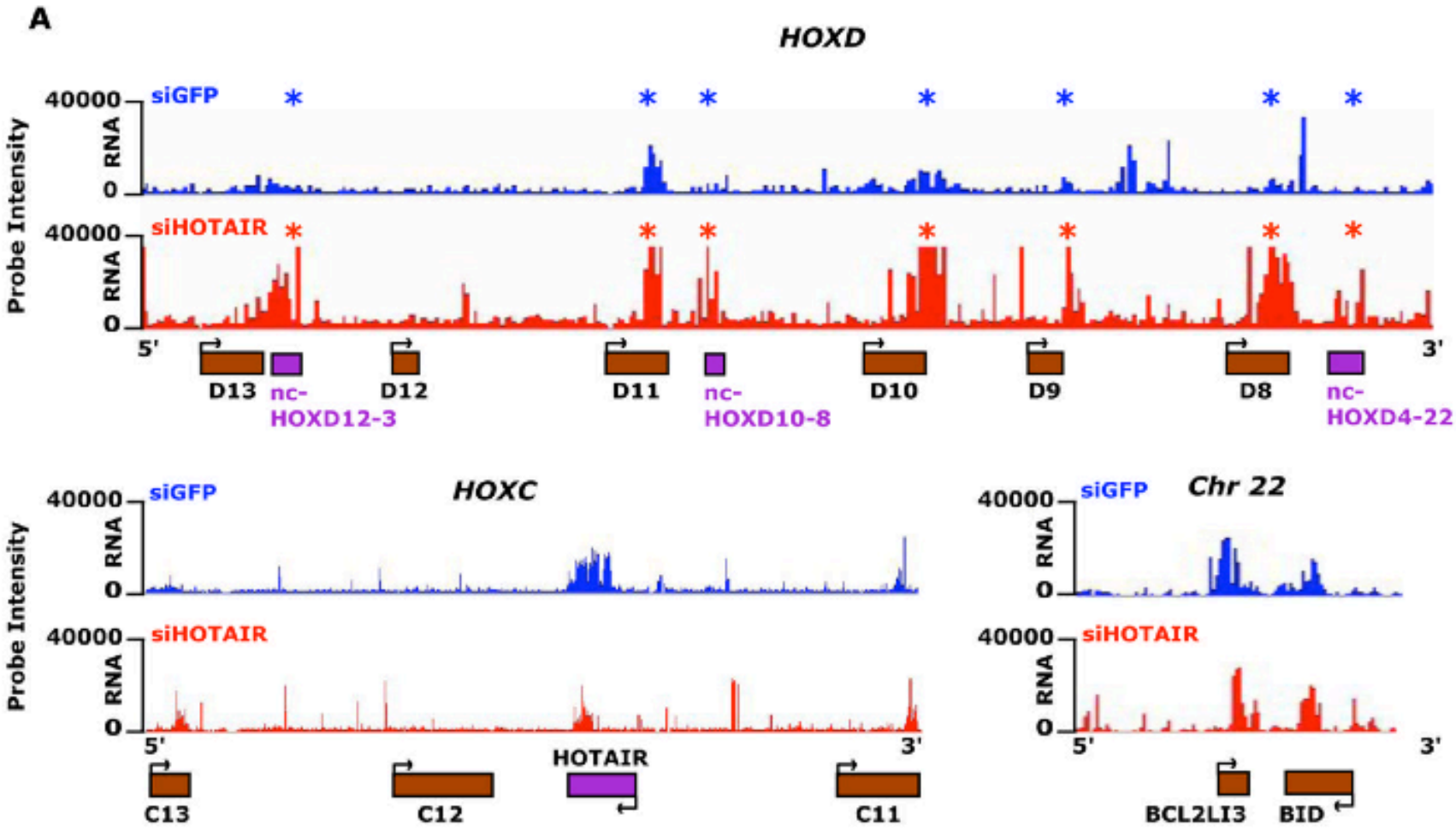


*limited to **stable** RBPs*

*prone to detect **non-specific interactions***

***low resolution**: the binding site in the co-purified RNA molecule remained unresolved*

Loss of HOTAIR results in transcriptional induction of HOXD locus

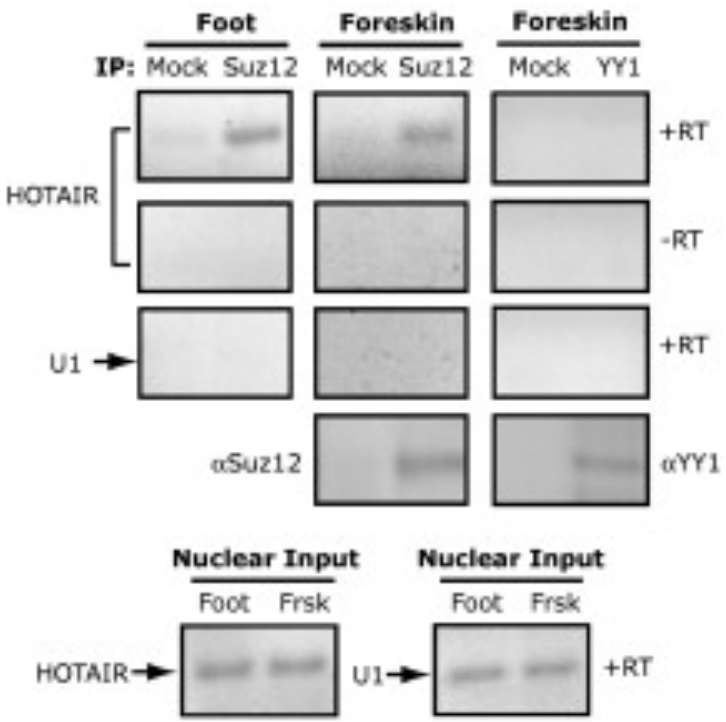


RNAi against HOTAIR in primary fibroblast led to *dramatic* transcriptional activation of the HOXD locus on chromosome 2 spanning over 40 kb, including HOXD8, HOXD9, HOXD10, HOXD11 and multiple ncRNAs

HOTAIR ncRNA Binds PRC2

RNA Immunoprecipitation (RIP)

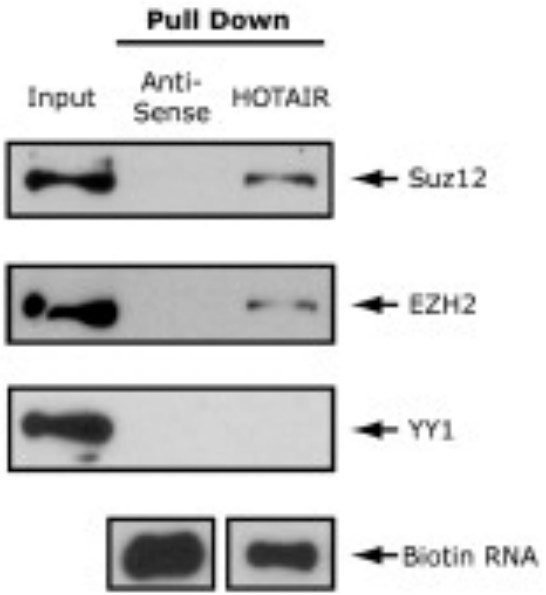
IP of Suz12 retrieves endogenous HOTAIR



usually followed by…

RNA Pull Down

In vitro-transcribed HOTAIR retrieves PRC2 subunits



Nuclear extracts of fibroblasts were immunoprecipiated by IgG (Mock), anti-Suz12, or anti-YY1. Co-precipitated RNAs were detected by RT-PCR using primers for HOTAIR or U1 small nuclear RNA.

Cell. 2007 June 29; 129(7): 1311–1323.

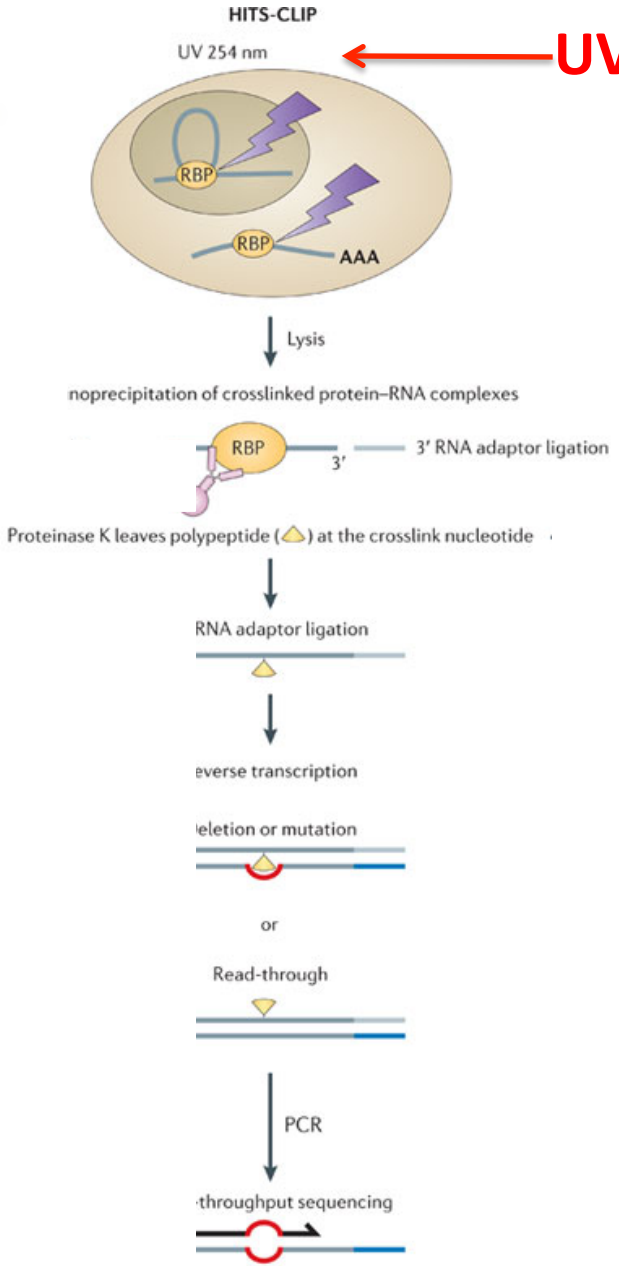
RNA immunoprecipitation (RIP)

CrossLinking and ImmunoPrecipitation (CLIP)

Chromatin Immunoprecipitation (ChIP)

Chromatin Isolation by RNA Purification (*ChIRP*) and Capture Hybridization Analysis of RNA Targets (*ChART*)

in vivo CrossLinking and ImmunoPrecipitation (CLIP)



UV CROSSLINKING

To identify the **positions of protein–RNA interactions** with high resolution and specificity

UV irradiation induces the formation of covalent crosslinks *only* at sites of **direct contact** between *proteins* and *RNA*.

Used in: König, J. *et al.* The fungal RNA-binding protein Rrm4 mediates long-distance transport of *ubi1* and *rho3* mRNAs. *EMBO J.* **28**, 1855–1866 (2009).

in vivo CrossLinking and ImmunoPrecipitation (CLIP)

1- Starting Material

***Option 1*: *UV Cross-Linking* of Tissues**

***Option 2*: *UV Cross-Linking* of Cultured Cells**

covalently capture close (near-covalent bond distances) association of RNA with protein

cross-linking for each protein should be optimized separately, because each RNA-binding protein will cross-link with different efficiencies. For a *preliminary* experiment, try 100, 200, and 400 mJ/cm^2

2- Prepare the Dynabead–antibody complexes

paramagnetic beads can be collected against the side of the tube using a magnetic rack

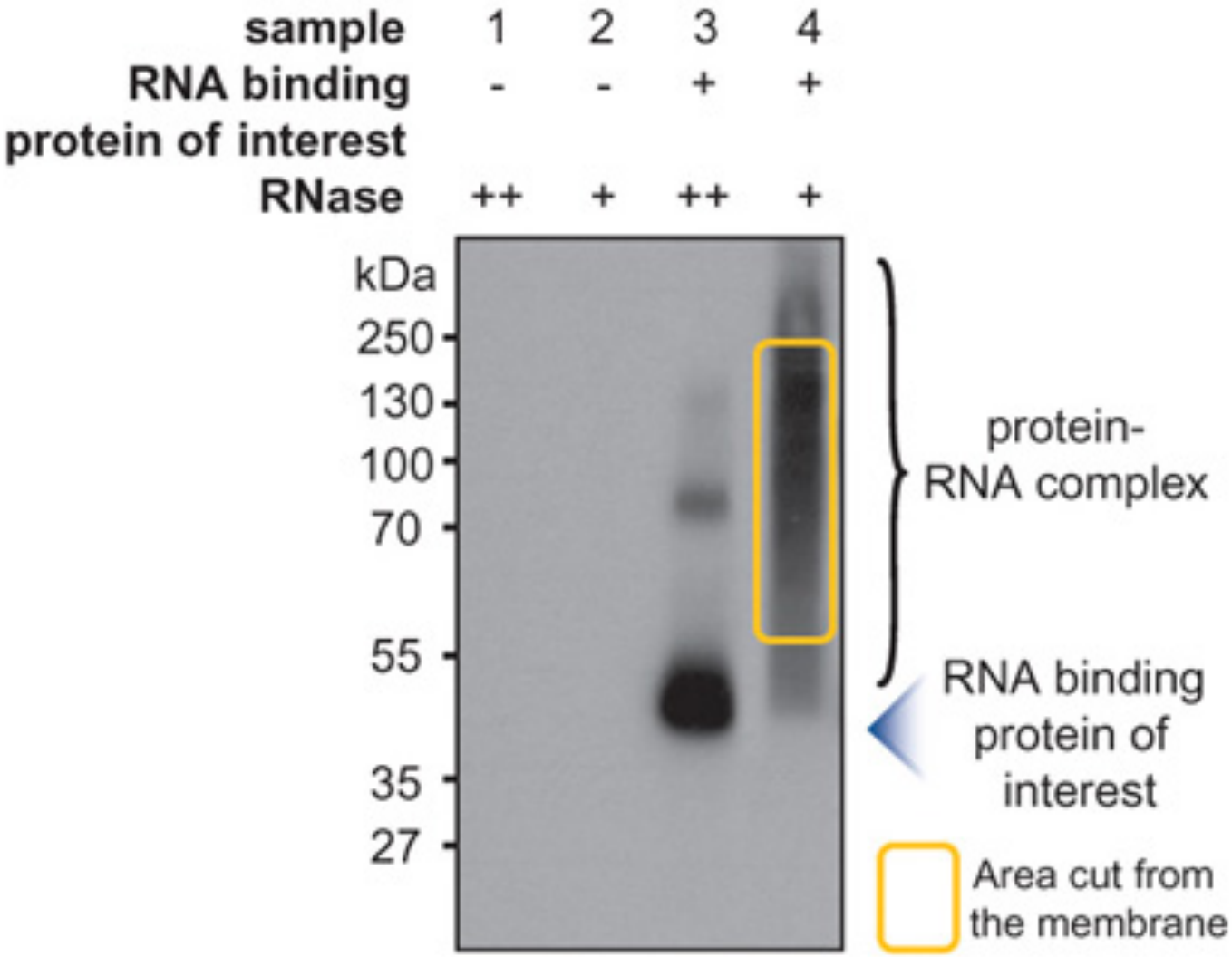
3- PARTIAL RNase Treatment and Immunoprecipitation

4- Remove Phosphates from the 3' RNA Ends and add RNA Linkers to the 3' End of the Fragments with RNA Ligase (on Bead)

5- Phosphorylate the 5' Ends of the RNAs with PNK (on Bead). Labelling

6- Denature and Resolve the RNA–Protein Cross-Linked Complexes on Gel

A sample CLIP experiment for hnRNP A1, a 36-kDa RNA-binding protein



8- **Transfer on nitrocellulose, cut out the RNA-protein crosslinked materials and recover RNA**

9- **Add RNA Linkers to the 5' End of the Fragments and digest with DNAseI**

10- **Amplify by RT-PCR and ri-amplify for sequencing**

CLIP (Cross-Linking and Immunoprecipitation) Identification of RNAs Bound by a Specific Protein

Robert Darnell

Adapted from *RNA: A Laboratory Manual*, by Donald C. Rio, Manuel Ares Jr, Gregory J. Hannon, and Timothy W. Nilsen. CSHL Press, Cold Spring Harbor, NY, USA, 2011.

CLIP was used to identify a number of targets

S. Guil, J.F. Caceres The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a Nat. Struct. Mol. Biol., 14 (2007), p. 591

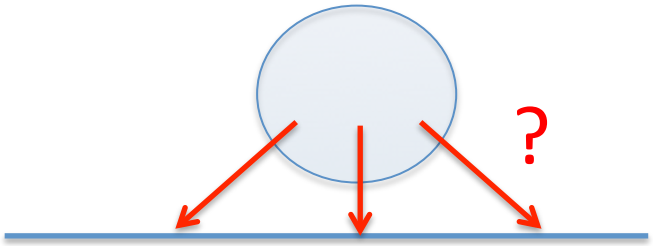
D.D. Licatalosi, et al. HITS-CLIP yields genome-wide insights into brain alternative RNA processing Nature, 456 (2008), pp. 464–469

J.R. Sanford, et al. Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts Genome Res., 19 (2009), pp. 381–394

G.W. Yeo, et al. An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells Nat. Struct. Mol. Biol., 16 (2009), pp. 130–137

CLIP is <u>limited</u> by:

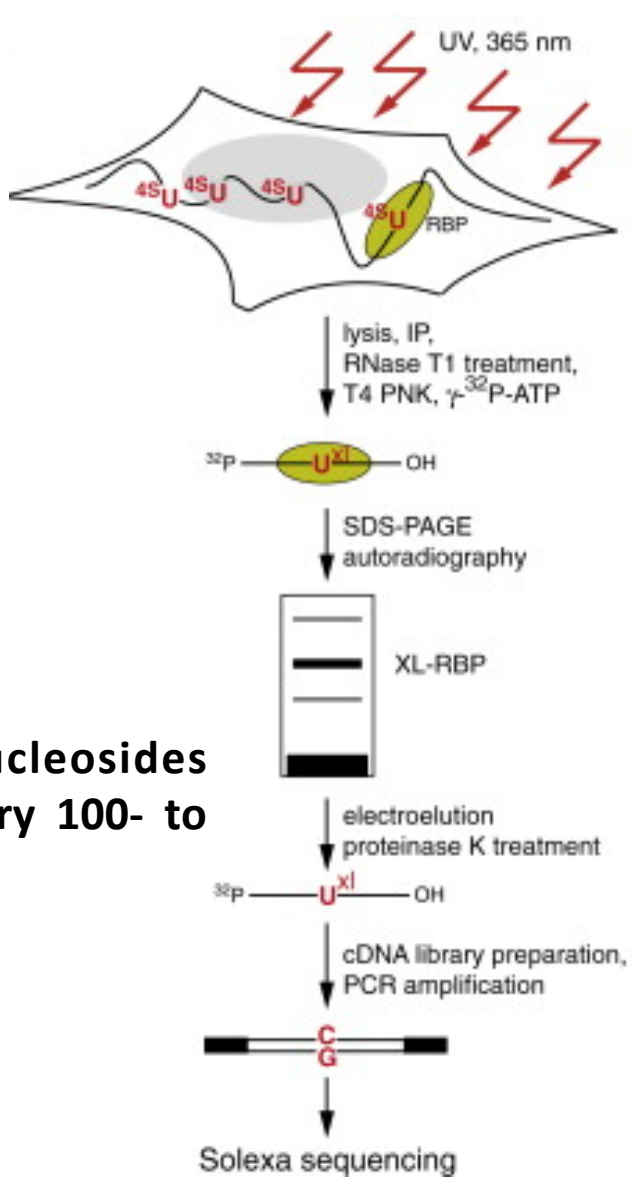
- the ***low*** efficiency of UV 254 nm RNA-protein crosslinking
- the ***location*** of the crosslink is not readily identifiable within the sequenced crosslinked fragments, raising the question of how to separate UV-crosslinked target RNA segments from *background* non crosslinked RNA fragments also present in the sample.



PAR-CLIP

(Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation)

1) Cells are cultured with a photoreactive ribonucleoside analogue, typically **4-thiouridine** (4SU), to boost RNA-protein crosslinking. The nucleotide analogues are become incorporated into newly synthesized transcripts.



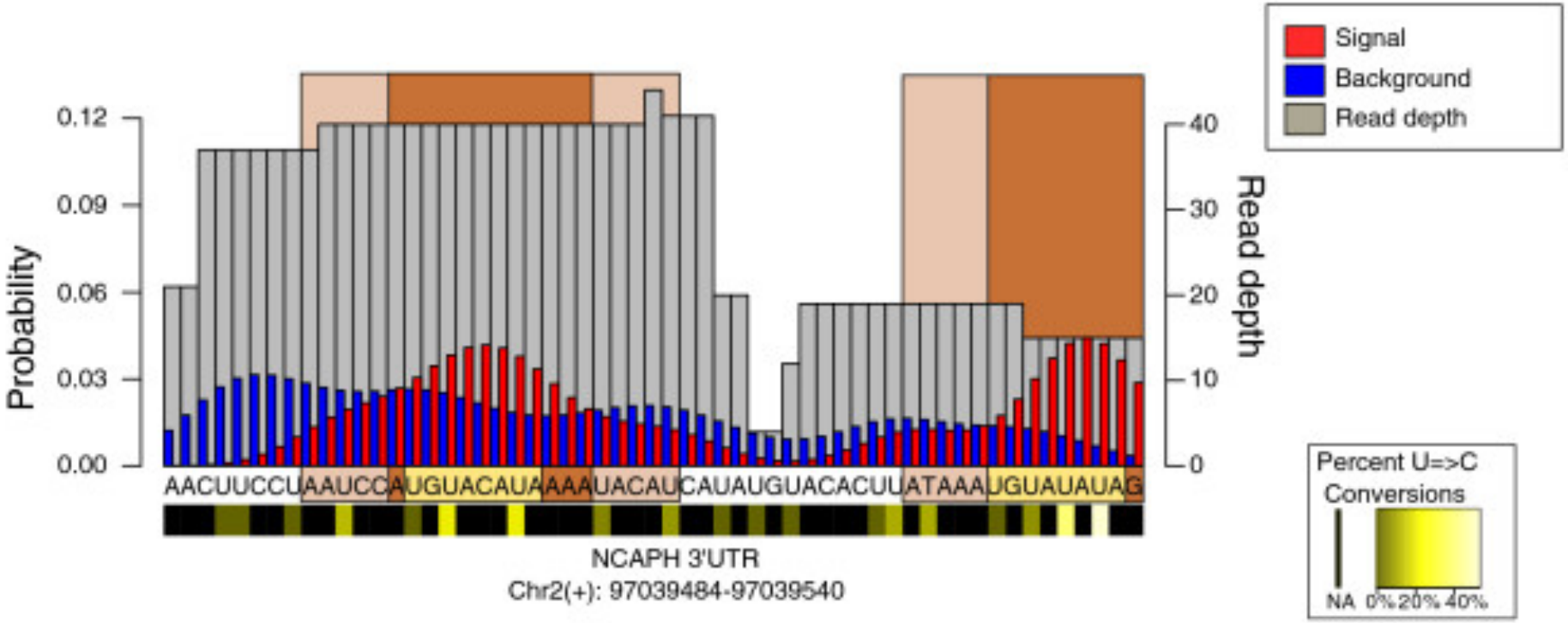
Photoactivatable **nucleosides improved RNA recovery 100- to 1000-fold!!!**

2) Irradiation of the cells by UV light of 365 nm induces **efficient crosslinking** of photoreactive nucleoside-labeled cellular RNAs to interacting RBPs.

3) During cDNA generation base pairing of the 4S(**U**) to a guanine (instead of an adenine) results in a thymine (T) to cytosine (**C**) transition in the PCR-amplified sequence, serving as a diagnostic mutation at the site of contact.

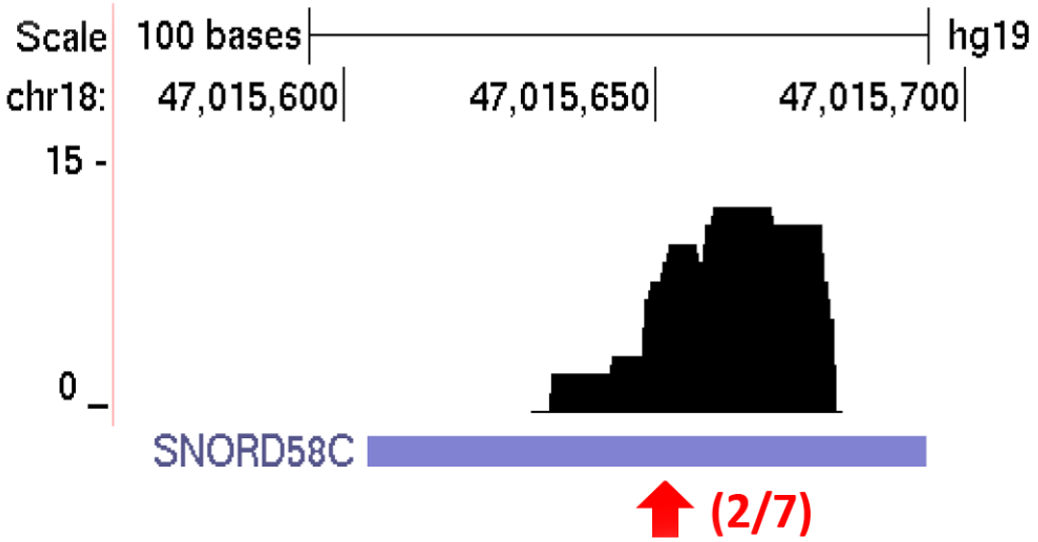
Therefore, mutation analysis of the resulting cDNA sequences can be used to pinpoint crosslink sites at nucleotide resolution.

Example of PAR-CLIP interaction site identification

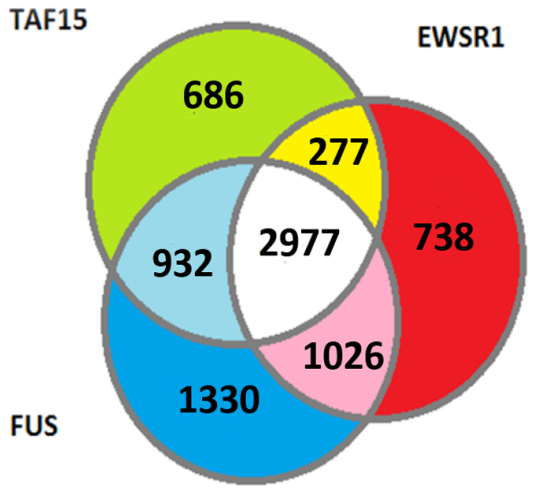


PUM2 binding sites

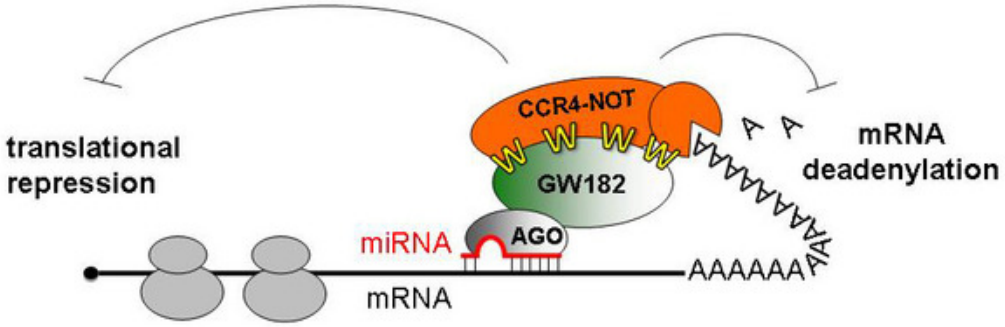
Results produced on the FET family PAR-CLIP dataset



Protein	# of clusters	# of genes associated with clusters
TAF15	15896	4872
EWSR1	18261	5018
FUS	41632	6265



miRNA activity requires base pairing with only 6-8 nucleotides of messenger RNA



Nature. 2009 Jul 23;460(7254):479-86. doi: 10.1038/nature08170. Epub 2009 Jun 17.

Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps.

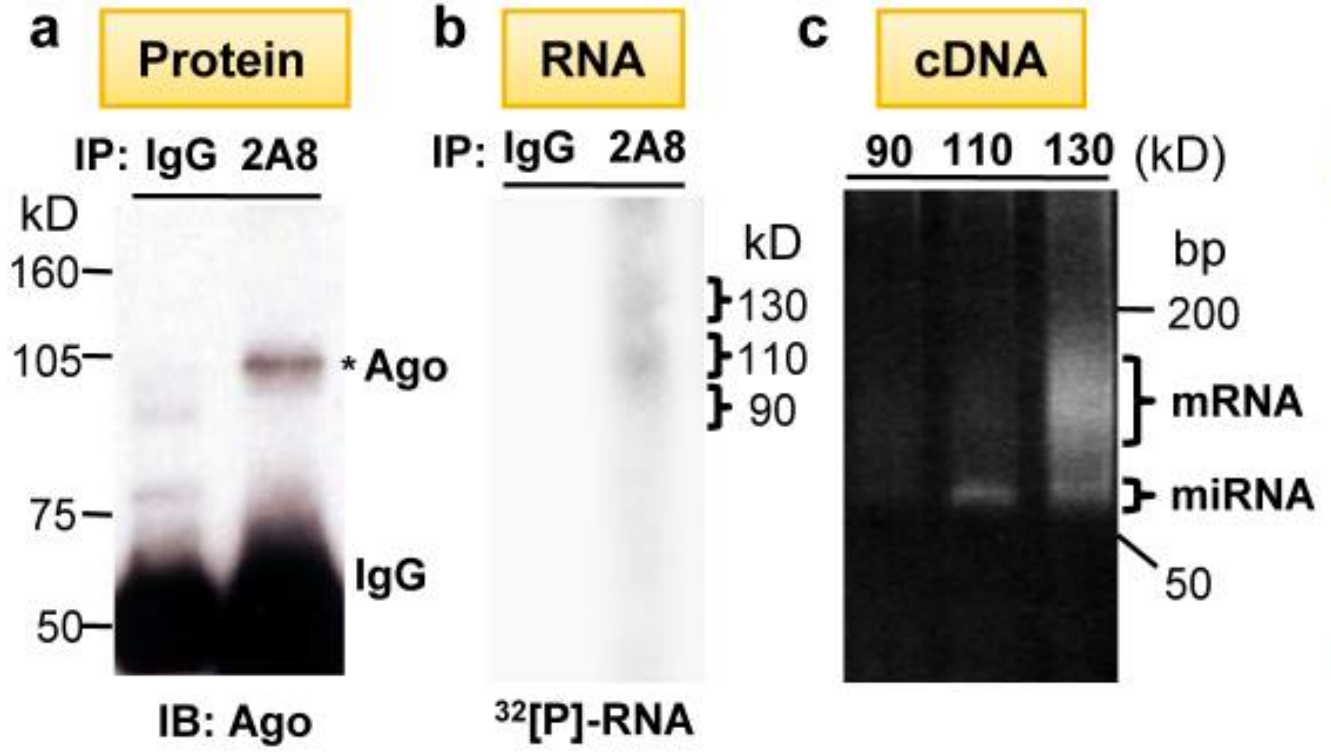
Chi SW[1], Zang JB, Mele A, Darnell RB.

⊕ Author information

Abstract

MicroRNAs (miRNAs) have critical roles in the regulation of gene expression; however, as miRNA activity requires base pairing with only 6-8 nucleotides of messenger RNA, predicting target mRNAs is a major challenge. Recently, high-throughput sequencing of RNAs isolated by crosslinking immunoprecipitation (HITS-CLIP) has identified functional protein-RNA interaction sites. Here we use HITS-CLIP to covalently crosslink native argonaute (Ago, also called Eif2c) protein-RNA complexes in mouse brain. This produced two simultaneous data sets-Ago-miRNA and Ago-mRNA binding sites-that were combined with bioinformatic analysis to identify interaction sites between miRNA and target mRNA. We validated genome-wide interaction maps for miR-124, and generated additional maps for the 20 most abundant miRNAs present in P13 mouse brain. Ago HITS-CLIP provides a general platform for exploring the specificity and range of miRNA action in vivo, and identifies precise sequences for targeting clinically relevant miRNA-mRNA interactions.

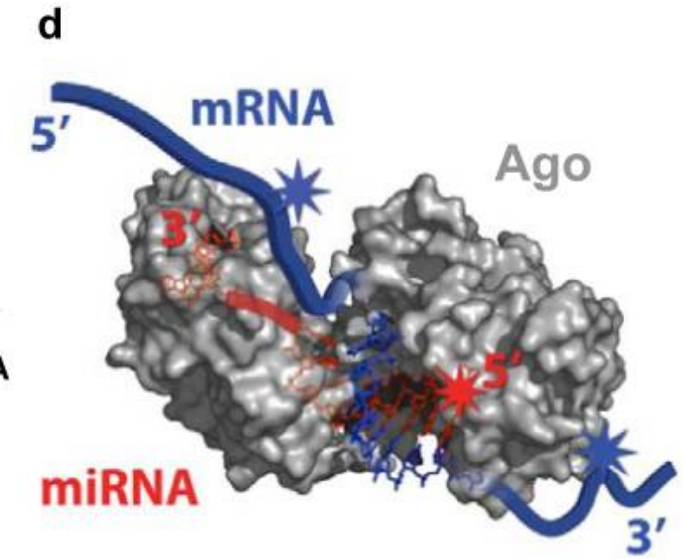
HITS-CLIP has identified functional protein-RNA interaction sites. HITS-CLIP to covalently crosslink native argonaute (Ago, also called Eif2c) protein-RNA complexes in mouse brain.



a) Immunoblot (IB) analysis of Ago immunoprecipitates (IP) from P13 mouse *neocortex* using pre-immune IgG as a control or anti-Ago monoclonal antibody

b) Autoradiogram of 32P-labeled RNA crosslinked to mouse brain Ago purified by IP.

c) PCR products amplified after linker (36 nt) ligation to RNA products excised from (b).



This produced two simultaneous data sets-Ago-miRNA and Ago-mRNA binding sites-that were combined with bioinformatic analysis to identify interaction sites between miRNA and target mRNA.

Nat Commun. 2014 Nov 4;5:5248. doi: 10.1038/ncomms6248.

PAR-CLIP analysis uncovers AUF1 impact on target RNA fate and genome integrity.

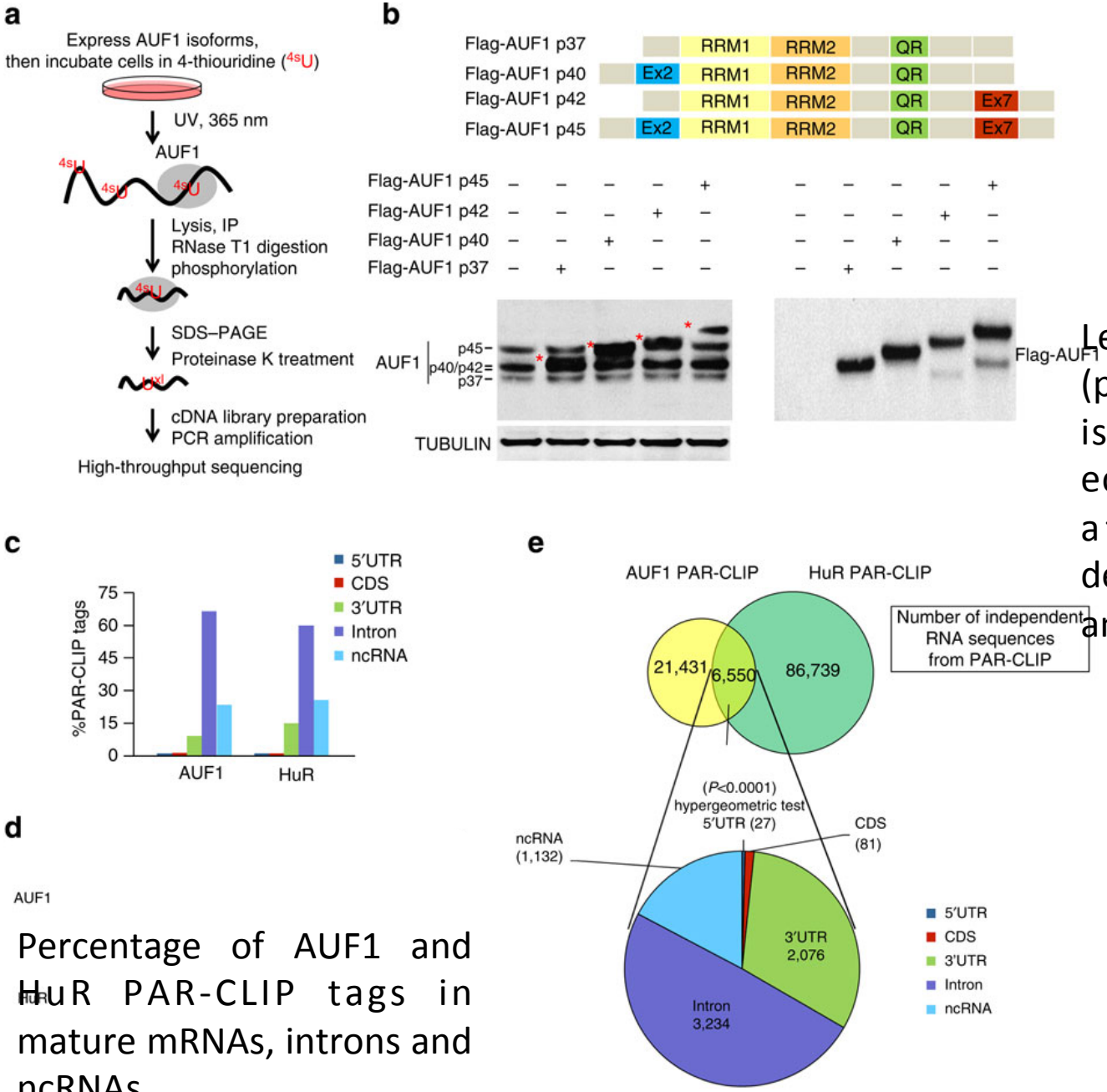
Yoon JH[1], De S[1], Srikantan S[1], Abdelmohsen K[1], Grammatikakis I[1], Kim J[1], Kim KM[1], Noh JH[1], White EJ[2], Martindale JL[1], Yang X[1], Kang MJ[1], Wood WH 3rd[1], Noren Hooten N[3], Evans MK[3], Becker KG[1], Tripathi V[4], Prasanth KV[4], Wilson GM[2], Tuschl T[5], Ingolia NT[6], Hafner M[7], Gorospe M[1].

⊕ Author information

Abstract

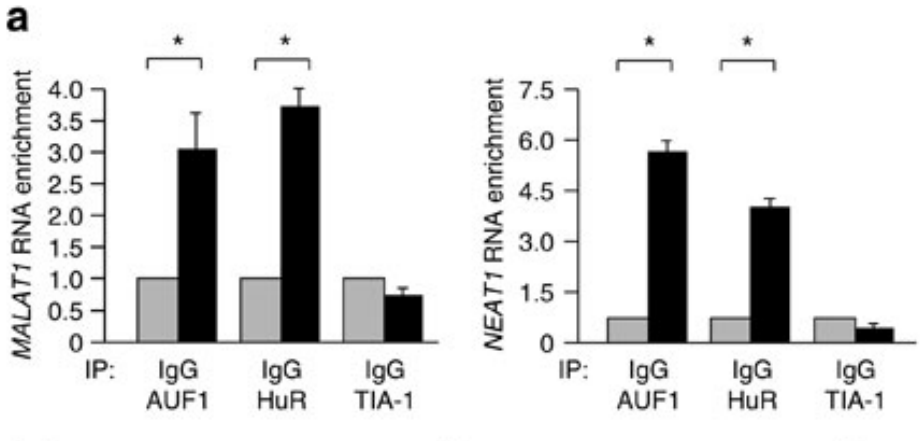
Post-transcriptional gene regulation is robustly regulated by RNA-binding proteins (RBPs). Here we describe the collection of RNAs regulated by AUF1 (AU-binding factor 1), an RBP linked to cancer, inflammation and aging. Photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) analysis reveals that AUF1 primarily recognizes U-/GU-rich sequences in mRNAs and noncoding RNAs and influences target transcript fate in three main directions. First, AUF1 lowers the steady-state levels of numerous target RNAs, including long noncoding RNA NEAT1, in turn affecting the organization of nuclear paraspeckles. Second, AUF1 does not change the abundance of many target RNAs, but ribosome profiling reveals that AUF1 promotes the translation of numerous mRNAs in this group. Third, AUF1 unexpectedly enhances the steady-state levels of several target mRNAs encoding DNA-maintenance proteins. Through its actions on target RNAs, AUF1 preserves genomic integrity, in agreement with the AUF1-elicited prevention of premature cellular senescence.

AUF1 (AU-binding factor 1) is an RBP linked to *cancer*, *inflammation* and *aging*.

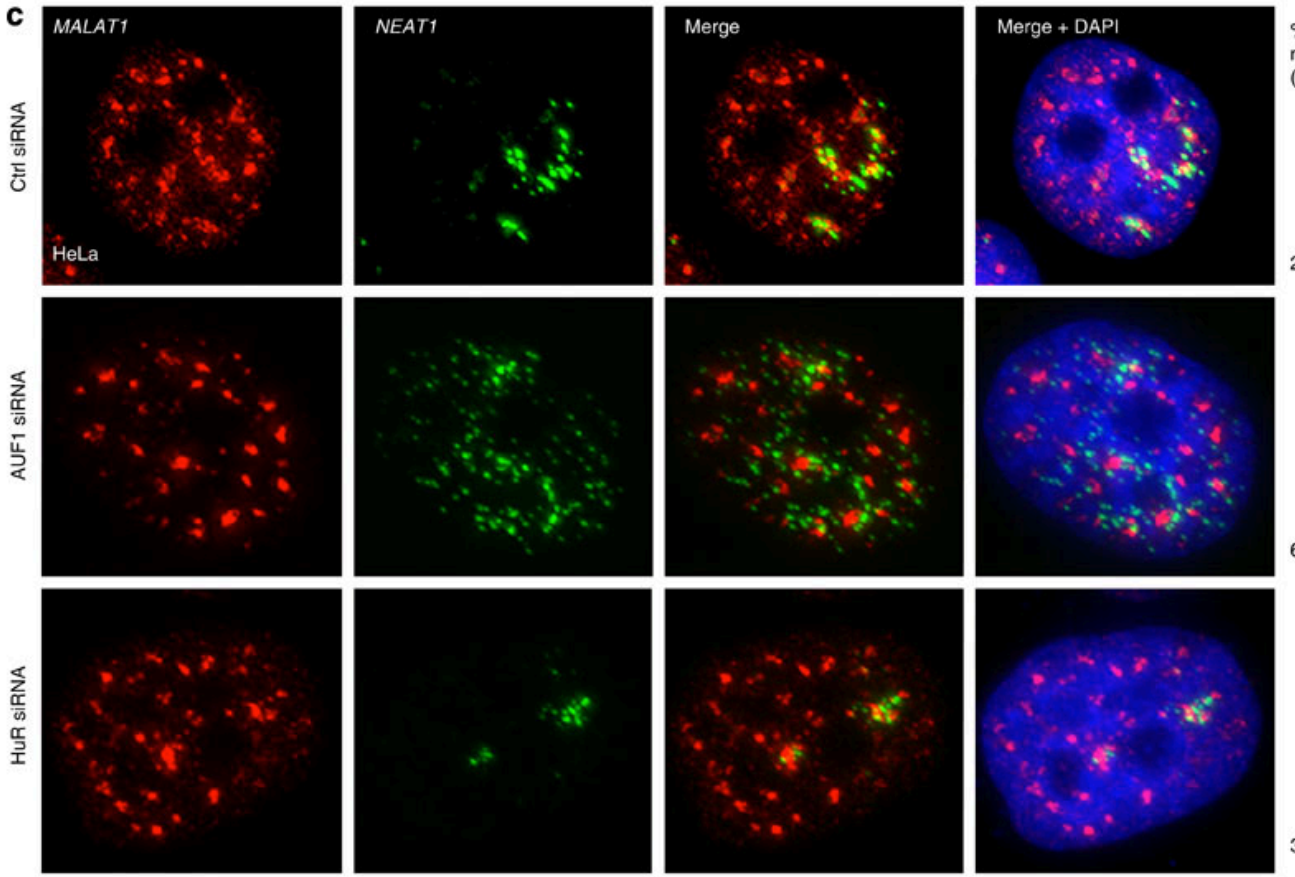


Levels of endogenous AUF1 (p37, p40, p42 and p45 isoforms indicated) and ectopic **Flag- AUF1** 48h after transfection as detected by western blot analysis

Percentage of AUF1 and HuR PAR-CLIP tags in mature mRNAs, introns and ncRNAs.



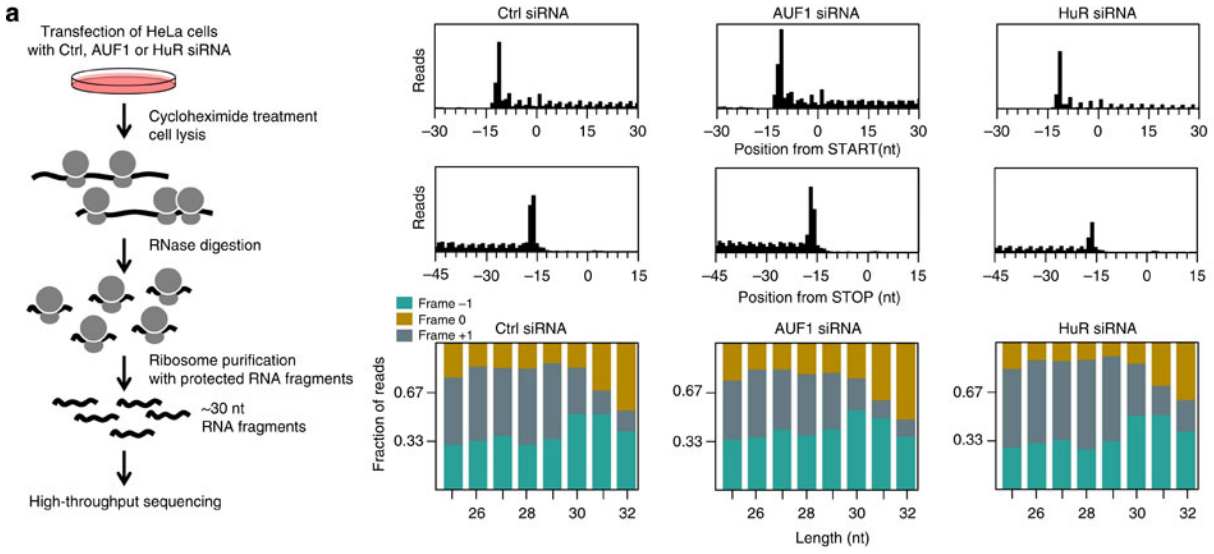
RIP analysis was carried out to measure the relative enrichment of lncRNAs *MALAT1* and *NEAT1*, as measured by RT-qPCR analysis in AUF1, HuR and TIA-1 RNPs or IgG

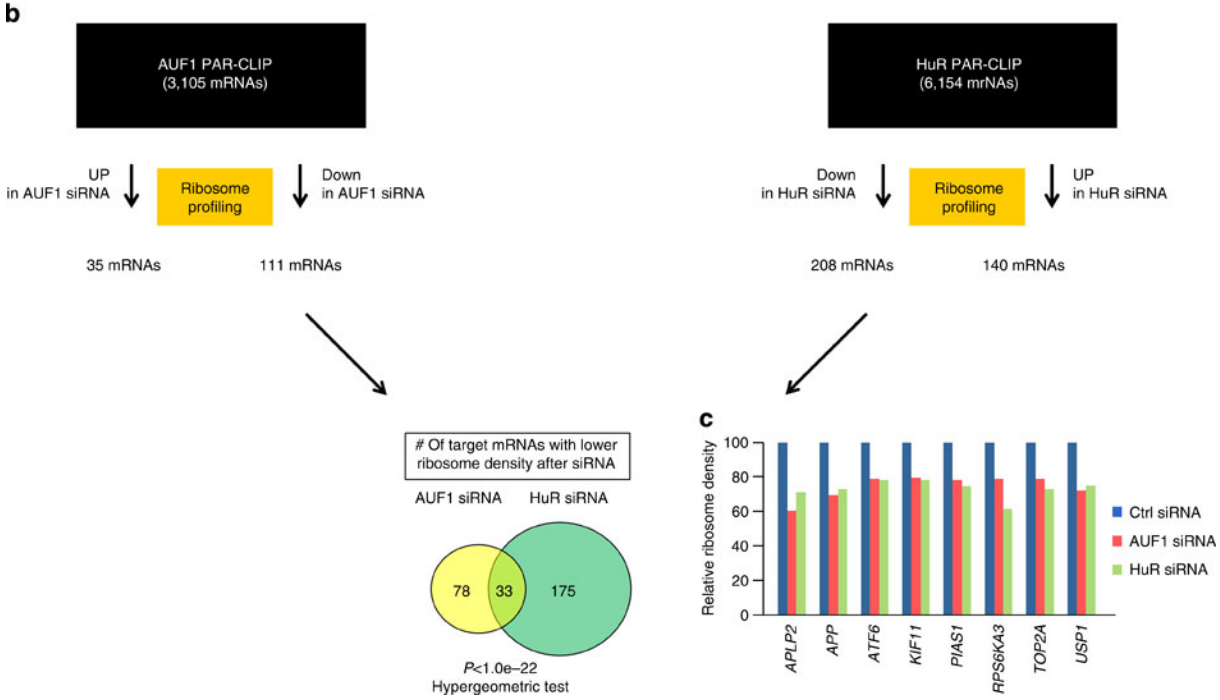


FISH analysis of ***NEAT1*** and ***MALAT1*** RNAs in HeLa cells 48 h after transfection of control, AUF1 or HuR siRNAs. DNA was counterstained with DAPI (blue).

AUF1 *lowers* the steady-state levels of numerous target RNAs, including long noncoding RNA NEAT1, in turn *affecting the organization of nuclear paraspeckles*

AUF1 cooperates with HuR for mRNA translation.





Relative ribosome density on representative mRNAs after AUF1 or HuR silencing.

RNA immunoprecipitation (RIP)

CrossLinking and ImmunoPrecipitation (CLIP)

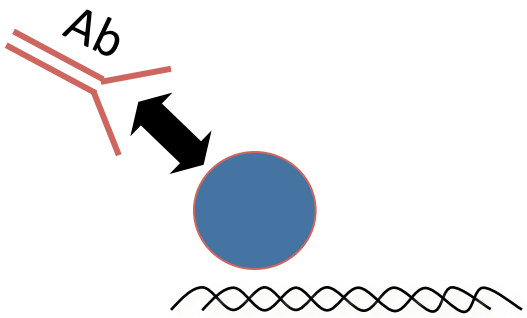
Chromatin Immunoprecipitation (ChIP)

Chromatin Isolation by RNA Purification (*ChIRP*) and Capture Hybridization Analysis of RNA Targets (*ChART*)

Chromatin ImmunoPrecipitation (ChIP)

A powerful method used to determine where and when a particular protein interaction occurs on specific DNA regions in chromatin context

The principle of ChIP: the selective enrichment of a chromatin fraction containing a specific protein



Two main ChIP procedures:

N-ChIP: **native** chromatin is used as substrate

-only proteins tightly associated with DNA can be immunoprecipitated

-antigens cannot be oscured or modified by chemical cross-linking

-the specificity of the antibody binding to unfixed chromatin is more predictable

X-ChIP: **cross-linked** chromatin is used as substrate

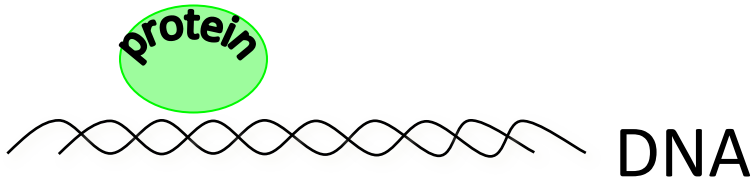
-also proteins weakly or not directly associated with DNA

-antigens can be obscured or modified by the formaldehyde cross-linking

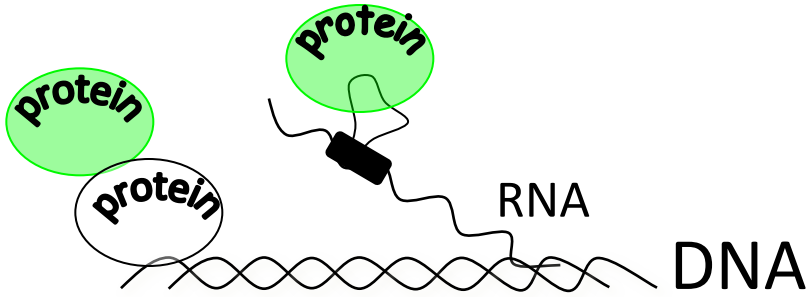
-more widely used than N-ChIP

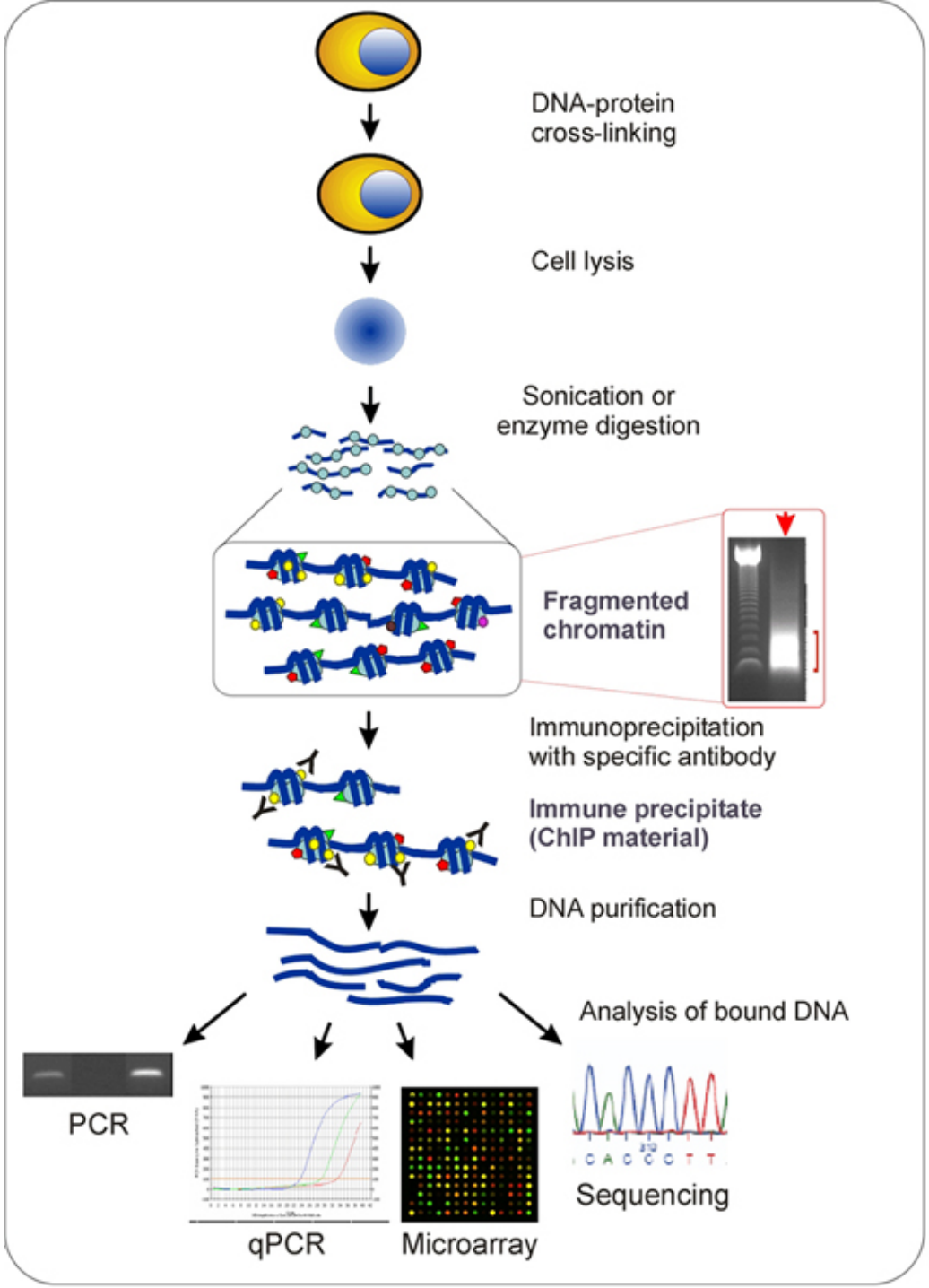
X-ChIP is used to map:

1. Proteins **directly** bound to DNA
   (i.e. transcription, replication, modification)



2. Proteins **not directly** bound to DNA
   (i.e. mRNA processing factors).

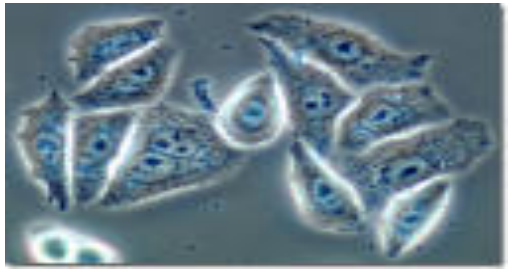



DNA-protein
cross-linking
Cell lysis
Sonication or
enzyme digestion
Fragmented
chromatin
Immunoprecipitation
with specific antibody
Immune precipitate
(ChIP material)
DNA purification
Analysis of bound DNA
PCR
qPCR
Microarray
Sequencing

Starting material

Chromatin from most sources is a suitable substrate for ChIP:

yeast



cell cultures

tissues

For analysis of recovered IP material at least 10-50ng (10^{-9}) DNA is needed

Each "starting ChIP sample" should contain equivalent of 25-50 μg DNA

X-ChIP consists of the following steps:

1- cell crosslinking with formaldehyde

2- cell lysis, chromatin extraction and shearing

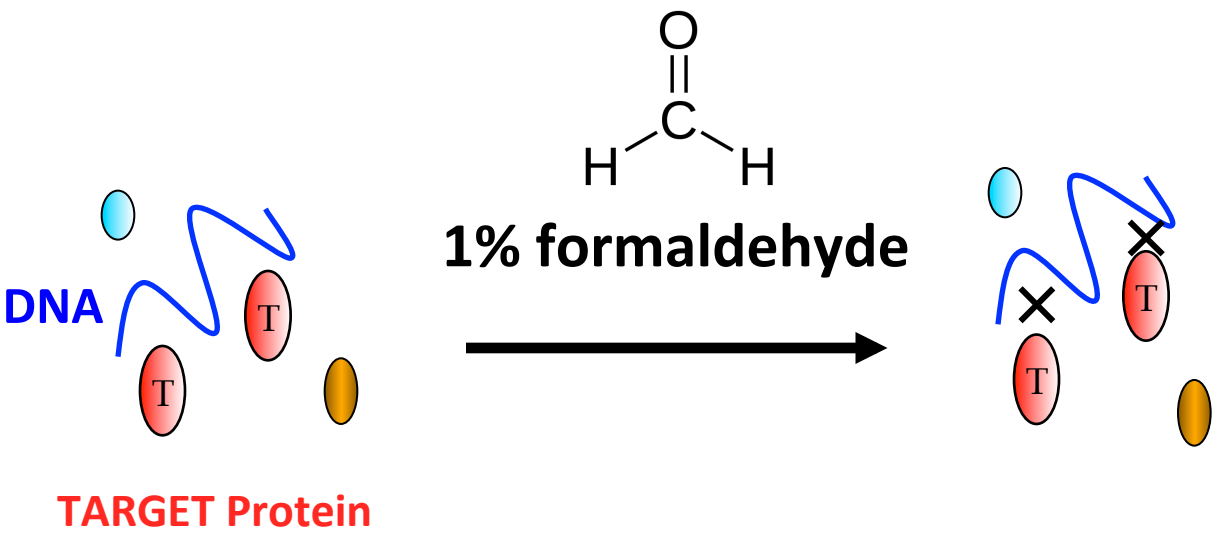
3- immunoprecipitation of resulting chromatin

4- reverse cross-linking (heat treatment)

5- proteins digestion (Proteinase K) and DNA isolation

6- quantification of the immunoprecipitated fraction to determine the level of DNA target sequence

1- Crosslinking with formaldehyde



Formaldehyde is an **organic compound**. It is water **soluble** and penetrates biological membranes

It targets primary **aminogroups** (i.e. lysines in proteins, side chains of A,C,G in DNA)

It crosslinks both protein-nucleic acids, nucleic acids-nucleic acids and protein-protein

Reaction is stopped by providing an excess of primary amino groups (0.125M glycine)

The crosslinking is **reversible** (65ºC reverse protein-DNA; 100ºC reverse protein-protein)

The application of formaldehyde as a cross-linker has several ***advantages****:*

1- only closely associated proteins can be cross-linked due to the small size of formaldehyde

2- its high permeability towards cell membranes enables cross-linking in the intact cell

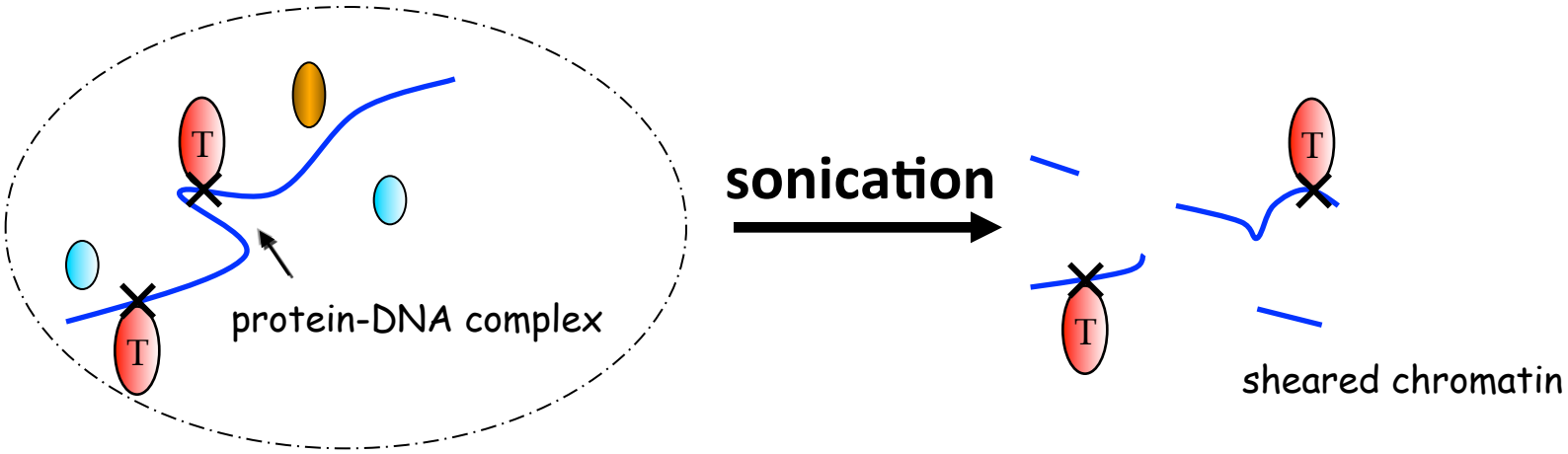
3- very fast cross-linking and the stabilization of transient interactions

4- Finally, formaldehyde is available in almost every laboratory at costs that amount to only a fraction of other cross-linkers.

However…

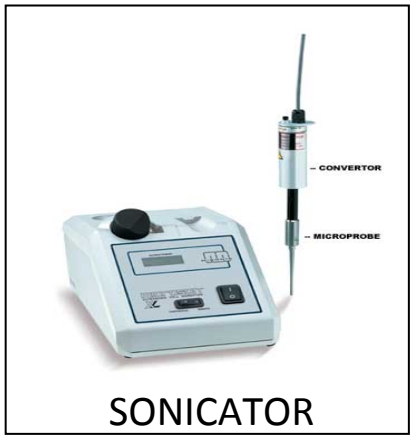
epitopes recognized by antibodies raised against endogenous proteins could be destroyed by formaldehyde modification, which would prevent their application

2- Cell lysis and shearing of chromatin by sonication



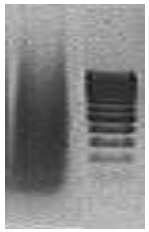
Cell lysis is performed and chromatin is isolated

Shearing is required since shorter DNA fragments provide a higher mapping resolution and less background in IP



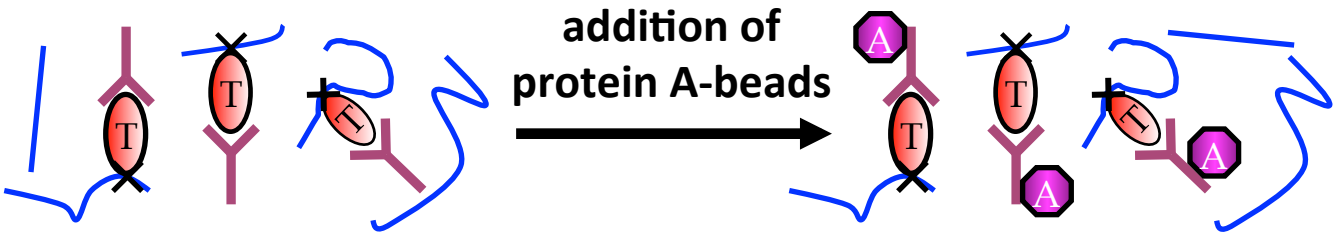
SONICATOR

agarose gel



fragment range

3- ImmunoPrecipitation



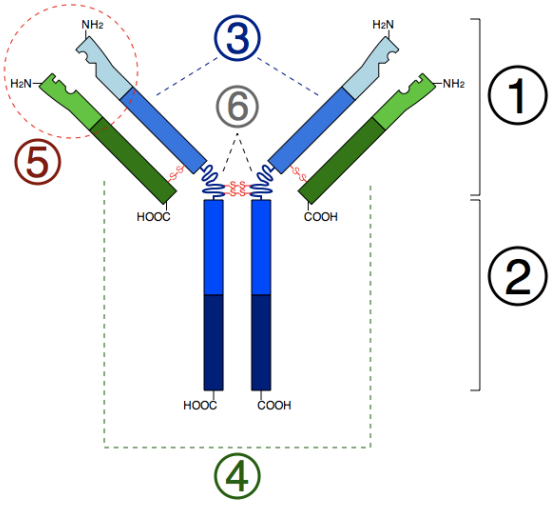
Protein/DNA/Ab COMPLEX

Protein/DNA/Ab/beads COMPLEX

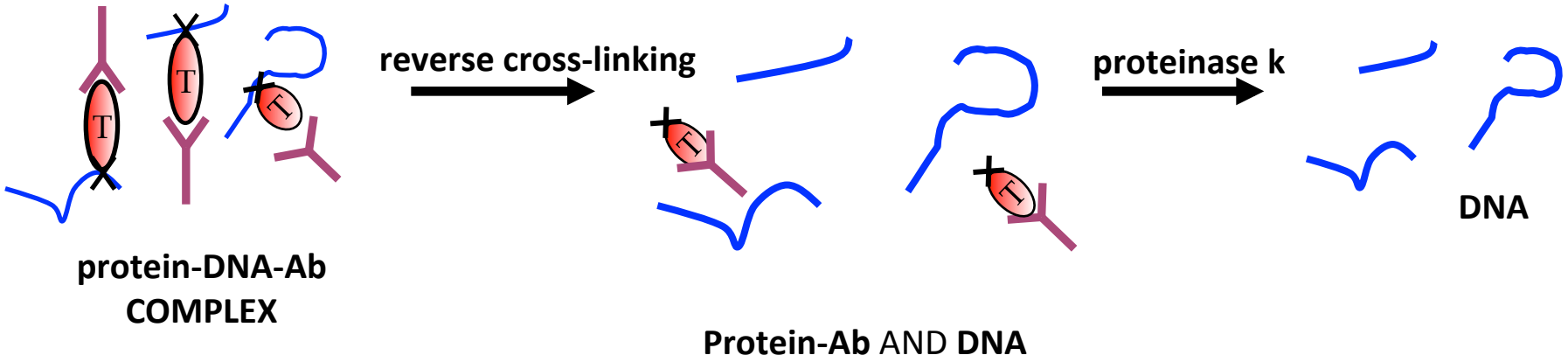
Antibody quality is essential for this step:

It has to:
-Recognize fixed protein

-polyclonal antibodies are preferred to monoclonals to avoid potential epitope masking problems in crosslinked



Reversal of cross-linking, proteinase K digestion and DNA isolation



Reverse of cross-linking — 65ºC incubation

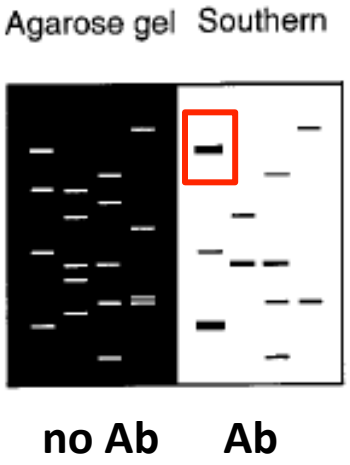
Protein digestion — ProteinaseK

DNA purification → Phenol/Chloroform + EtOH ppt

DNA purification → Affinity column

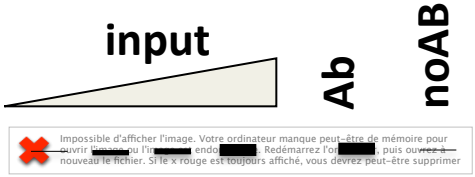
Analysis of the immunoprecipitated fraction

1. The genomic fraction obtained by immunoprecipitation, containing all the *in vivo* binding sites of a given chromatin protein, is radiolabelled and used as a probe against DNA fragments encompassing large genomic regions (**Southern Blot**)



Orlando et al, 1997

2. PCR with specific primer for the region of interest



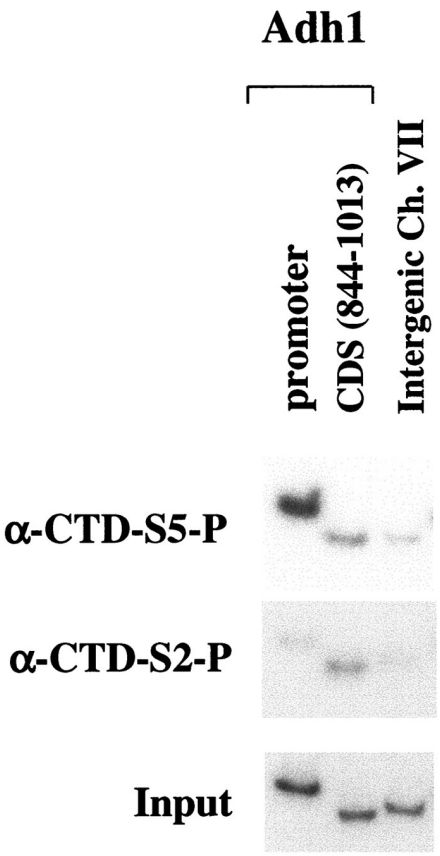
ChIP applications

1. Stato di fosforilazione dell' **RNA polimerasi II** durante la trascrizione

2. Legame di specifici **fattori di trascrizione** al promotore nel differenziamento cellulare

3. Reclutamento co-trascrizionale di **fattori di processamento** mediato dall'RNA nascente

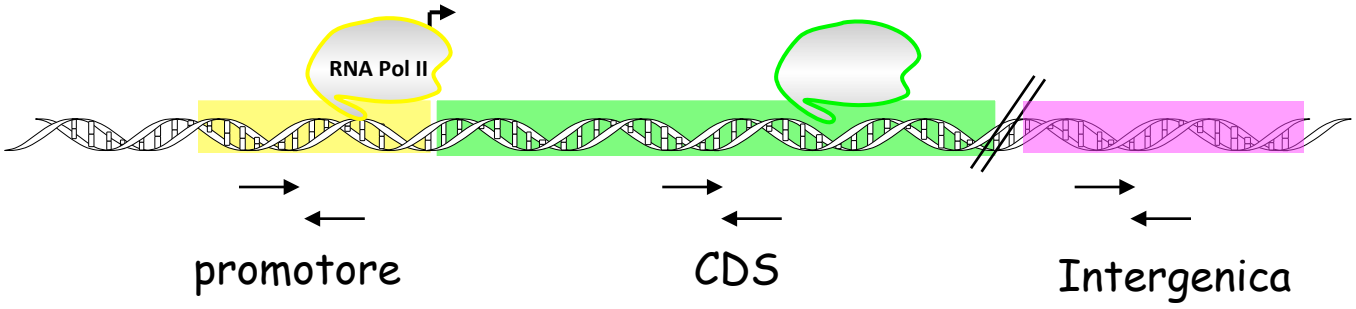
Stato di fosforilazione del CTD della RNA pol II durante la trascrizione



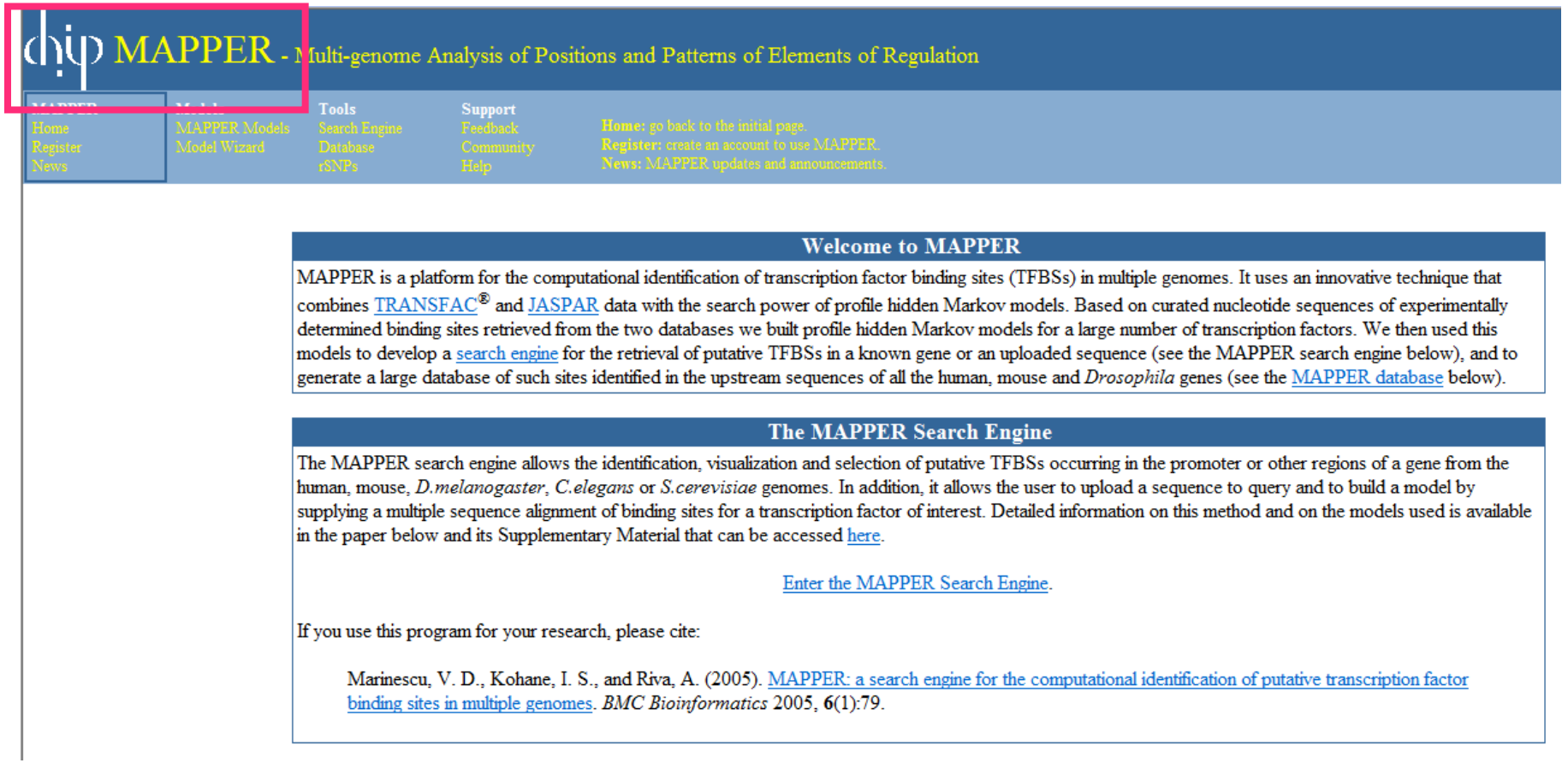
Buratowski e colleghi eseguirono un'analisi mediante ChIP dell'associazione delle due forme fosforilate del CTD con la cromatina vicina o lontana ai promotori dei geni ADH1 e PMA1.

La cromatina venne precipitata con anticorpi diretti contro il CTD fosforilato sia sulla Ser2 che sulla Ser5 dell'eptamero (aCTD-ser2, aCTD-ser5).
Sul DNA estratto venne poi eseguita una amplificazione per PCR con inneschi specifici per regioni vicine o lontane dal promotore, e una regione intergenica.

La cromatina di partenza non immunoprecipitata è indicata come INPUT in basso

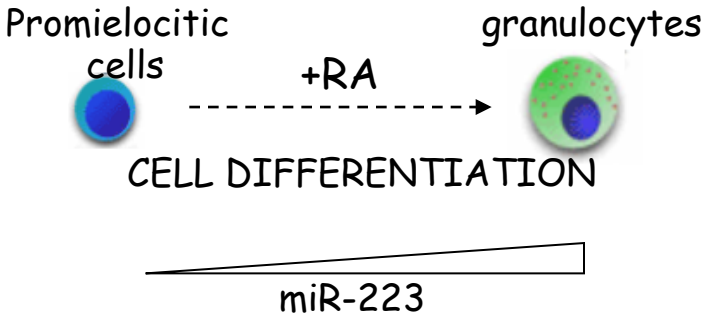


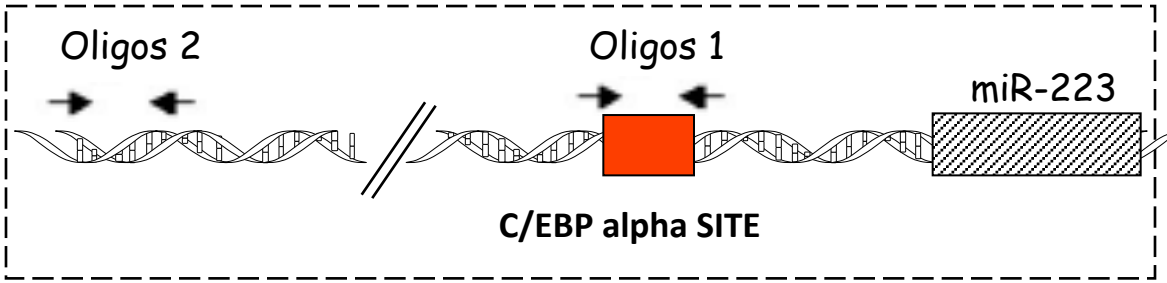
Legame di specifici fattori di trascrizione a livello del promotore

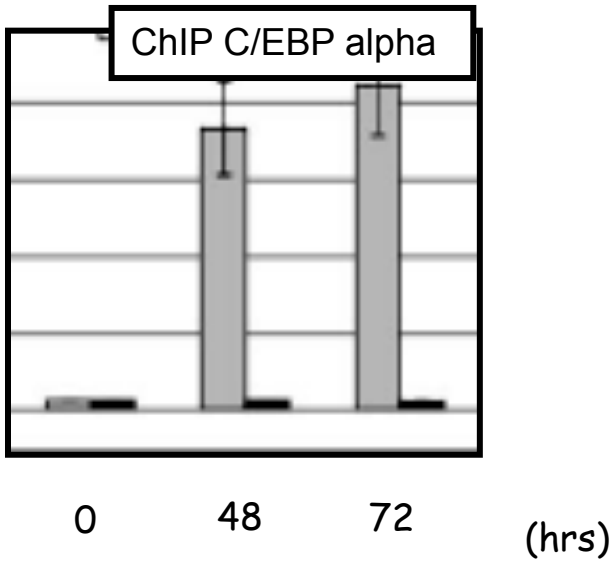


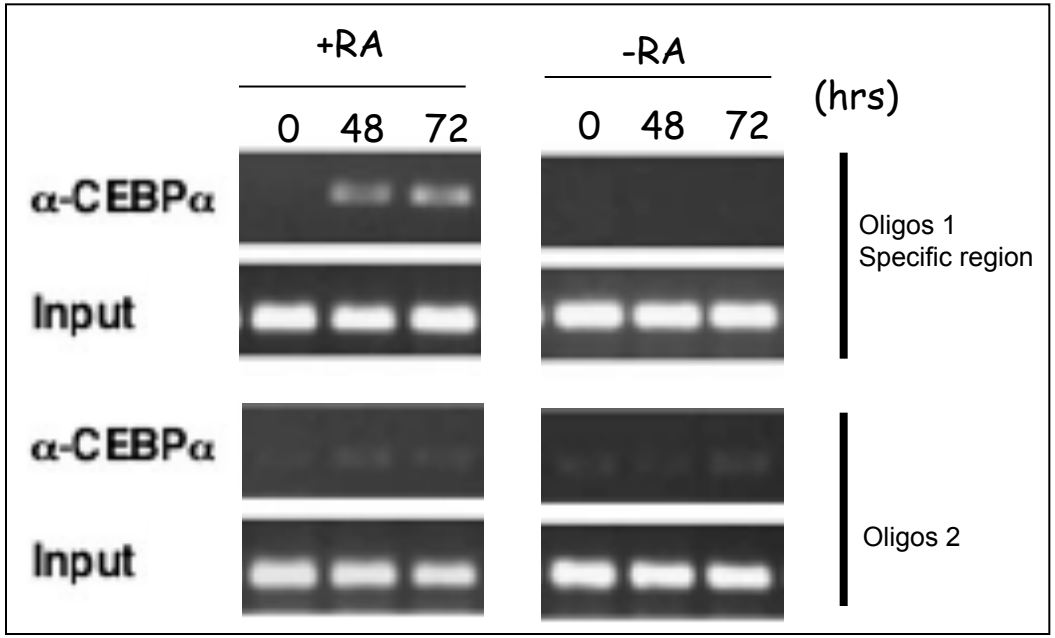
http://mapper.chip.org

Legame di un fattore di trascrizione tessuto-specifico durante il differenziamento

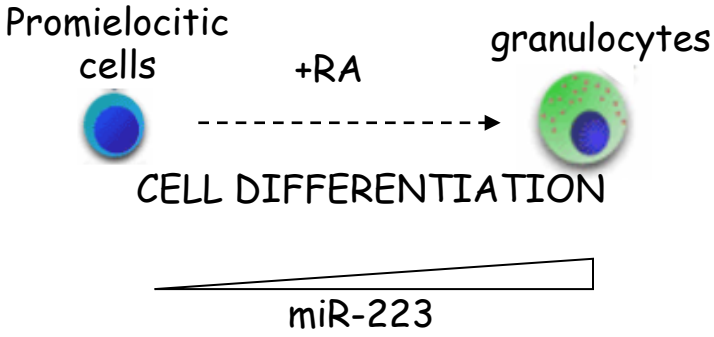


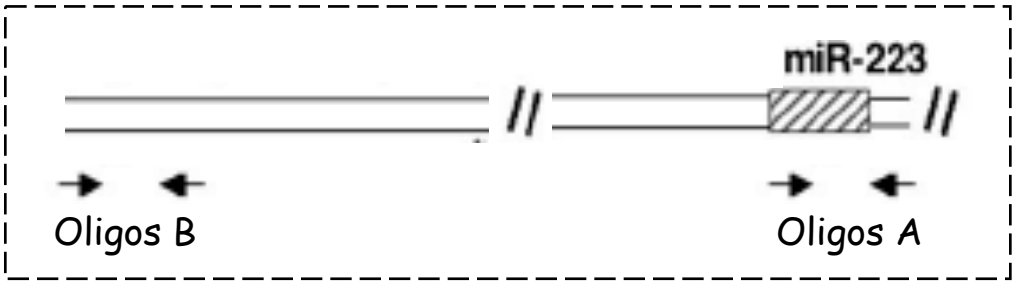


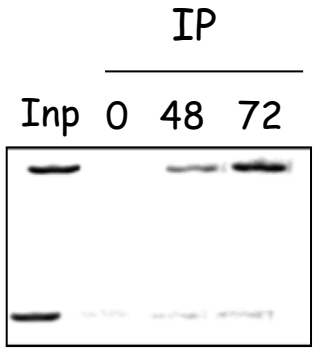




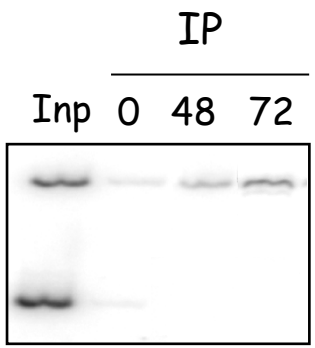
Reclutamento co-trascrizionale di Drosha sui geni dei miRNA



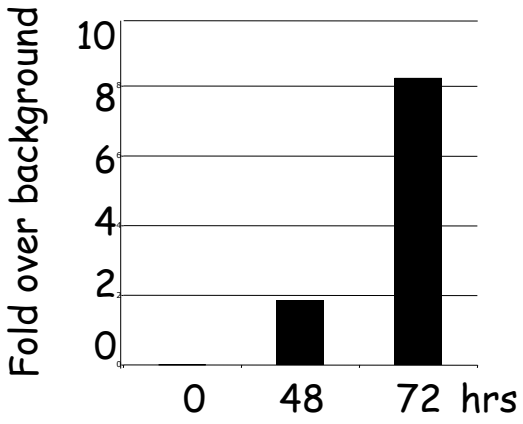


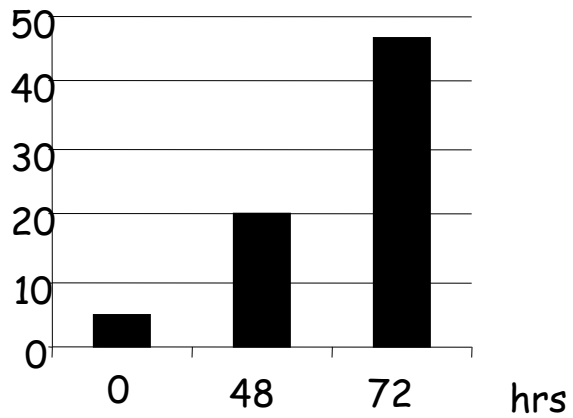


ChIP Drosha



ChIP Pol II





ChIP on chip

…on a genome-wide scale, ChIP on chip takes the strategy of combining chromatin immunoprecipitation with microarray.

Two major areas of research can be studied using ChIP on chip:

1- Identification of sequences with specific histone modifications

- distribution of histone modifications
- Localization of the histone modification enzymes

2- Identification of binding targets for nuclear protein factors

ChIP on chip flowchart

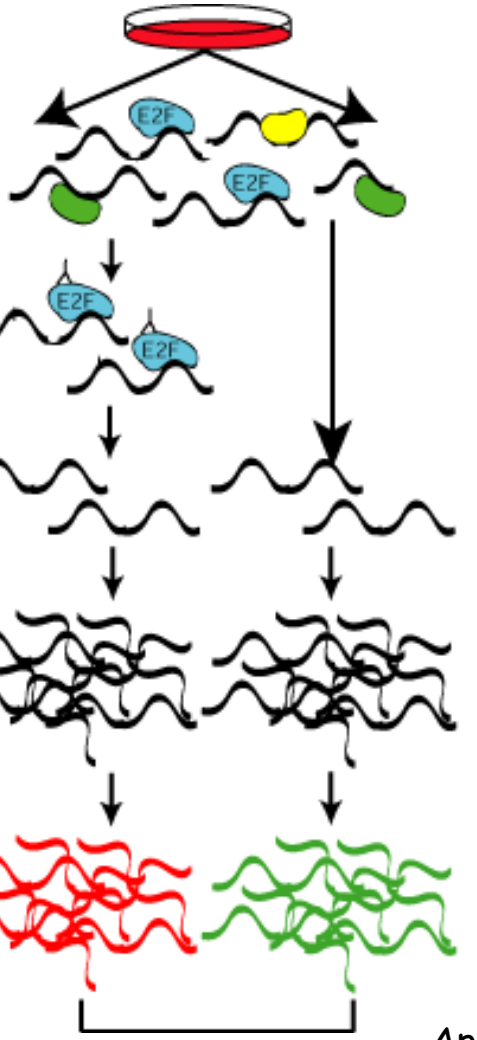
cross-linking

immunoprecipitation

reverse cross-linkng

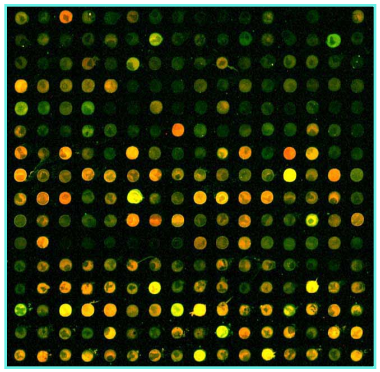
PCR amplification of ChIP DNA

Label amplicons with cyanin dye



Apply labeled samples to a microarray

Compare Red/Green signal intensities to identify binding sites

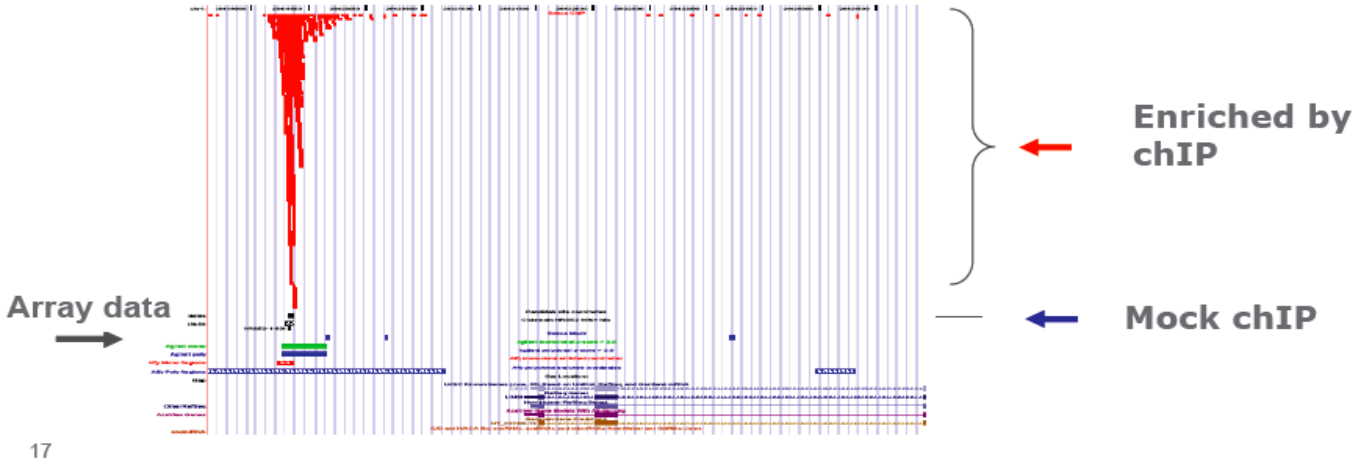


Oberley et al.2003

ChIP-Sequencing

Discovering transcription factor binding sites using chIP

- Treat chromatin with antibody specific to TF and recover trapped DNA
- Generate 6M tags by Solexa sequencing
- Identify clustered alignments in human genome
- Compare to microarray data in ENCODE region
- Genome-wide dataset



- Unlike microarray assay, the vast majority of single copy sites in the genome is accessible for ChIP Seq assay, rather than a subset of selected to be array features.
- ChIP Seq counts sequences and **avoids constraints imposed by array hybridization** chemistry
- ChIP Seq is feasible for any sequenced genome rather than being restricted to species for which whole-genome tiling arrays have been produced

RNA immunoprecipitation (*RIP*)

CrossLinking and ImmunoPrecipitation (*CLIP*)

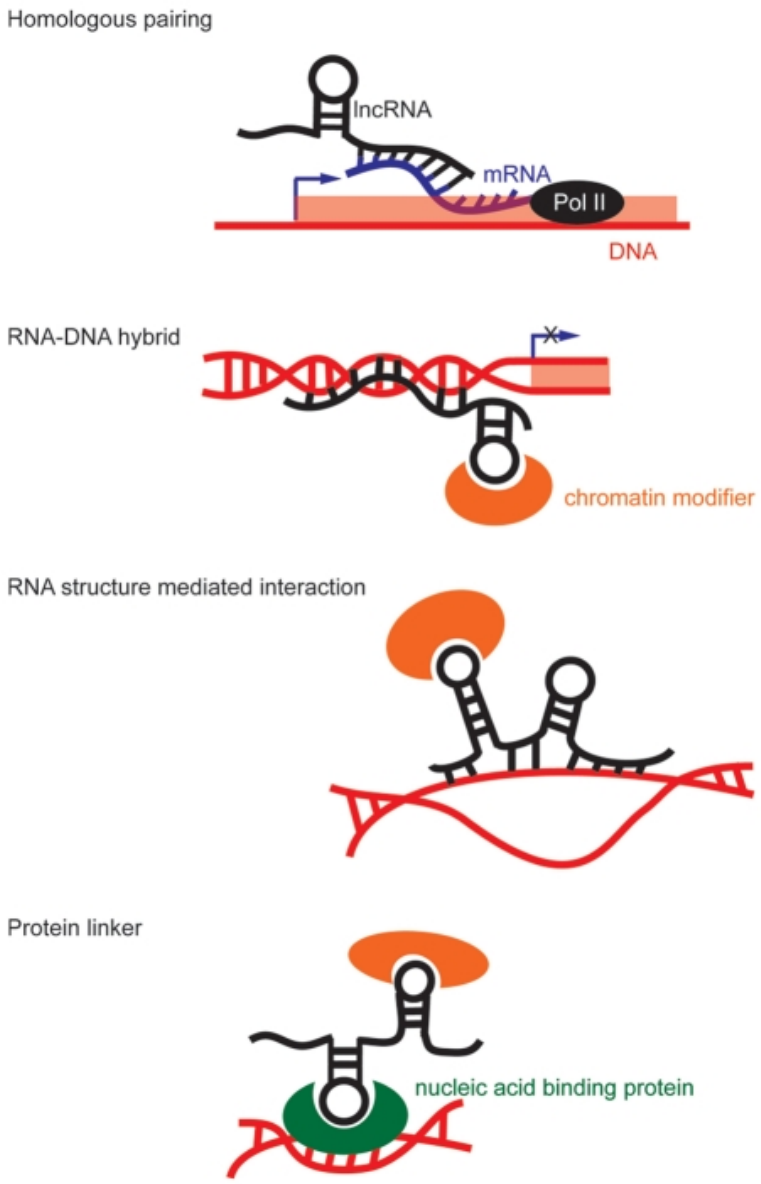
Chromatin Immunoprecipitation (*ChIP*)

Chromatin Isolation by RNA Purification (*ChIRP*) and Capture Hybridization Analysis of RNA Targets (*ChART*)

ChIRP, **CHART** and **RAP** are all based on the same basic idea—using biotinylated oligonucleotides complementary to the RNA of interest as a handle with which to pull down associated proteins, or more accurately, *chromatin*.

How does a lncRNA interface with selective regions of the genome?

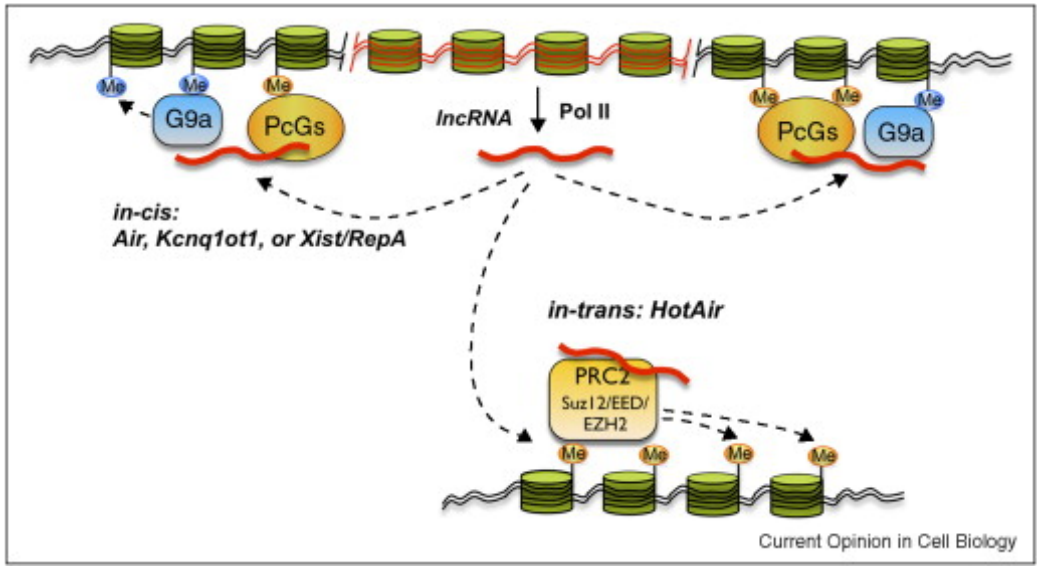
- RNA:RNA hybrid of lncRNA with a nascent transcript
- formation of a RNA: DNA:DNA triplex
- RNA: DNA hybrid that displaces a single-strand of DNA (so called R-loop)
- RNA binding to a sequence-specific DNA binding protein



- **lncRNAs** are *key* regulators of chromatin state

  The recent discovery of thousands of lncRNAs in association with specific chromatin modification complexes, such as PRC2, suggests broad roles for numerous lncRNAs in managing chromatin states.

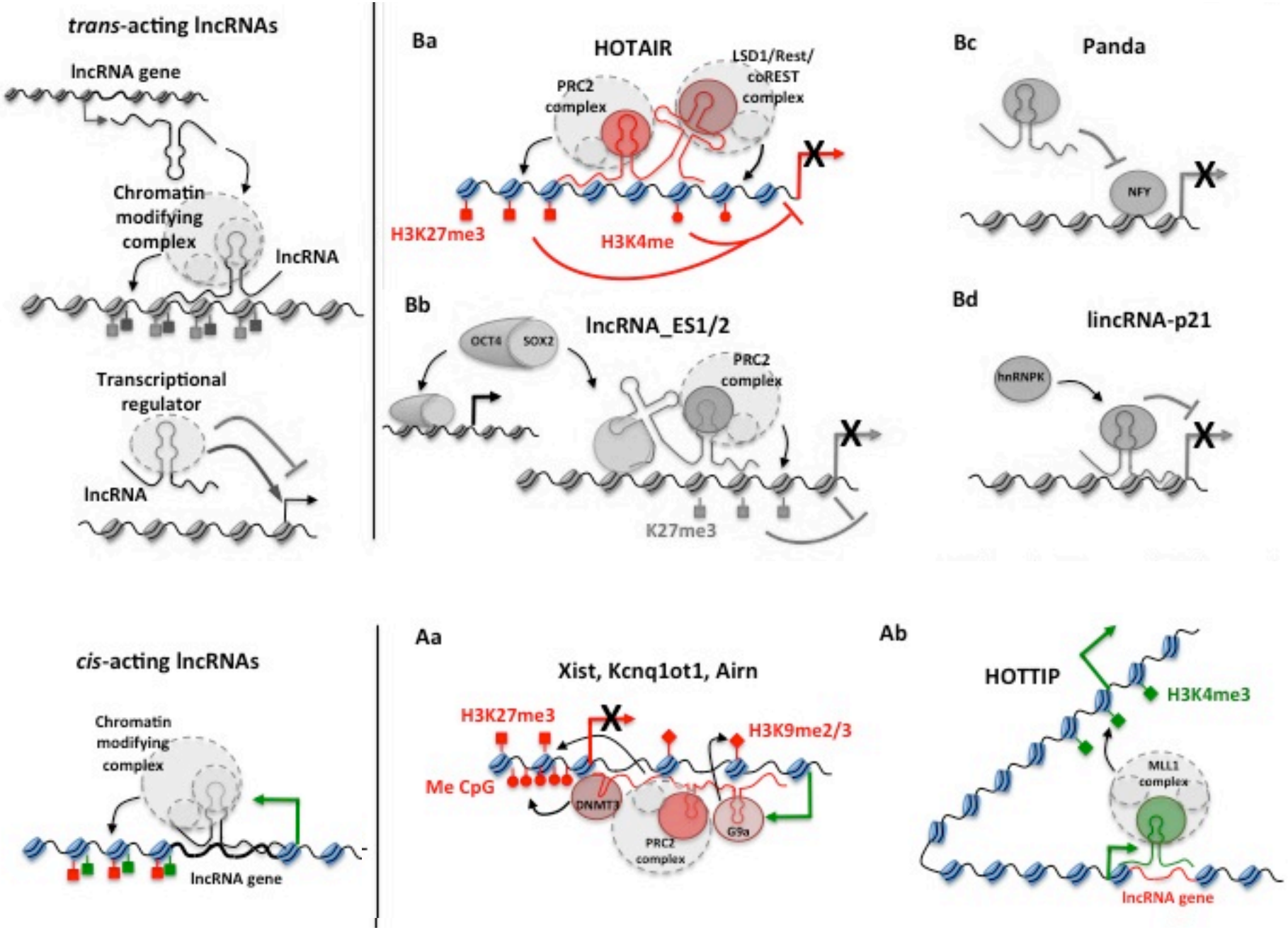
- some **lncRNAs** are thought to work in ***cis*** on neighboring genes, other lncRNAs work in ***trans*** to regulate distantly located genes.



How to discover

- The ***nature*** and ***sites*** of RNA-chromatin interaction are mostly unknown

?


trans-acting lncRNAs
lncRNA gene
Chromatin modifying complex
lncRNA
Transcriptional regulator
lncRNA
Ba
HOTAIR
PRC2 complex
LSD1/Rest/ coREST complex
X
H3K27me3
H3K4me
Bb
lncRNA_ES1/2
OCT4
SOX2
PRC2 complex
X
K27me3
Bc
Panda
NFY
X
Bd
lincRNA-p21
hnRNPK
X
cis-acting lncRNAs
Chromatin modifying complex
lncRNA gene
Aa
Xist, Kcnq1ot1, Airn
H3K27me3
X
H3K9me2/3
Me CpG
DNMT3
PRC2 complex
G9a
Ab
HOTTIP
H3K4me3
MLL1 complex
lncRNA gene

Chromatin Isolation by RNA Purification (ChIRP)

allows to map *in vivo* the RNA occupancy genome-wide at high resolution. ChIRP-seq has enabled the first genome-wide views of ncRNA occupancy on the human genome.

Mol Cell. 2011;44(4):667-78. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. Chu C, Qu K, Zhong FL, Artandi SE, Chang HY.

ChIRP applications

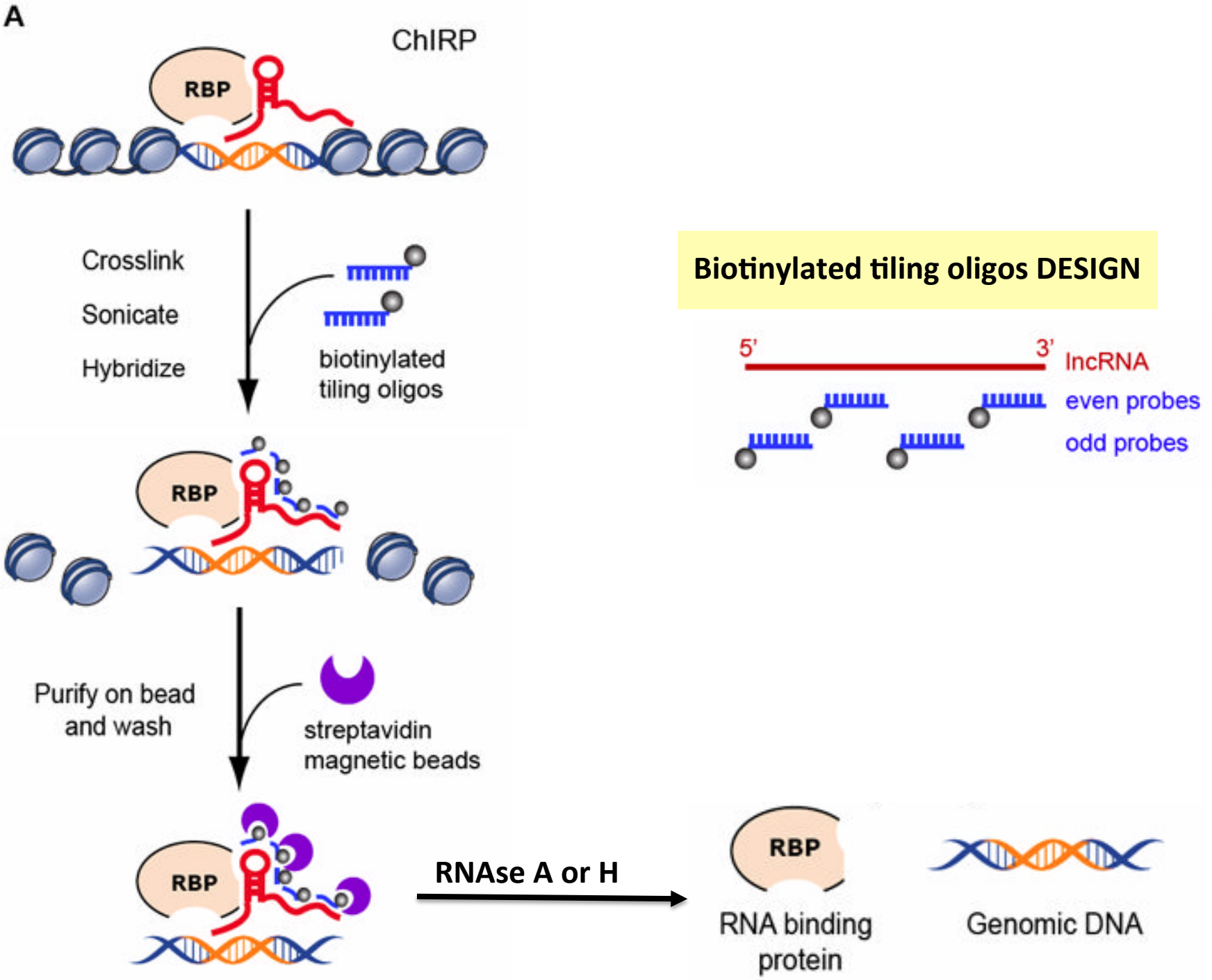
ChIRP-seq of these lncRNAs reveal that RNA occupancy sites in the genome are:

- **focal**
- **sequence-specific**
- **numerous**

- Human lincRNA **HOTAIR** can affect PRC2 occupancy on hundreds of genes genome-wide (Gupta et al., 2010)

- **TERC** RNA that serves as the template and scaffold for the telomerase complex (Zappulla and Cech, 2006)

- Drosophila ncRNAs **roX1** and **roX2** bind numerous regions on the X chromosome of male cells, and are critical for dosage compensation (Franke and Baker, 1999).


A
ChIRP
RBP
Crosslink
Sonicate
Hybridize
biotinylated
tiling oligos
RBP
Purify on bead
and wash
streptavidin
magnetic beads
RBP
RNAse A or H
RBP
RNA binding
protein
Genomic DNA
Biotinylated tiling oligos DESIGN
5'
3'
lncRNA
even probes
odd probes

PROBES DESIGN tool
Stellaris website

Stellaris® Probe Designer version 4.0

Probe Designer

This program takes an input sequence (such as an mRNA coding sequence) and will give as output a set of probes that are designed for optimal binding properties to the target RNA sequence. It will generate a probe list as well as a graphical representation of where each probe binds along the target sequence.

* Indicates a required field for user input

Probe Set Name *
(Maximum 22 characters.)
LncRNA name

Gene Name
Please specify to assist with our technical support (optional)

Organism *
For masking, to improve probe specificity. Masking levels 3-5 use organism specific information, and are unavailable if 'other' is selected.
Choose

Masking Level (0-5)
Genomic information of selected organism used for masking (except "Other")

0

Level 0: No masking
Level 1-2: Non-species specific. Avoids general problematic RNA sequences.
Level 3-5: Improves probe specificity by using genome information from the selected organism

Max. Number of Probes
48

Min. Spacing Length (nt)
2

Target Sequence *
Sense strand of the target sequence should be entered since the program will design probes that are complementary to the input sequence. It is recommended that the target sequence is well-defined, and that the existence of RNAs with similar sequence from related genes, as well as target gene-specific RNA variants has been assessed. RNA variants may arise from alternative transcription start site use, alternative splicing and poly-adenylation.

lncRNA sequence

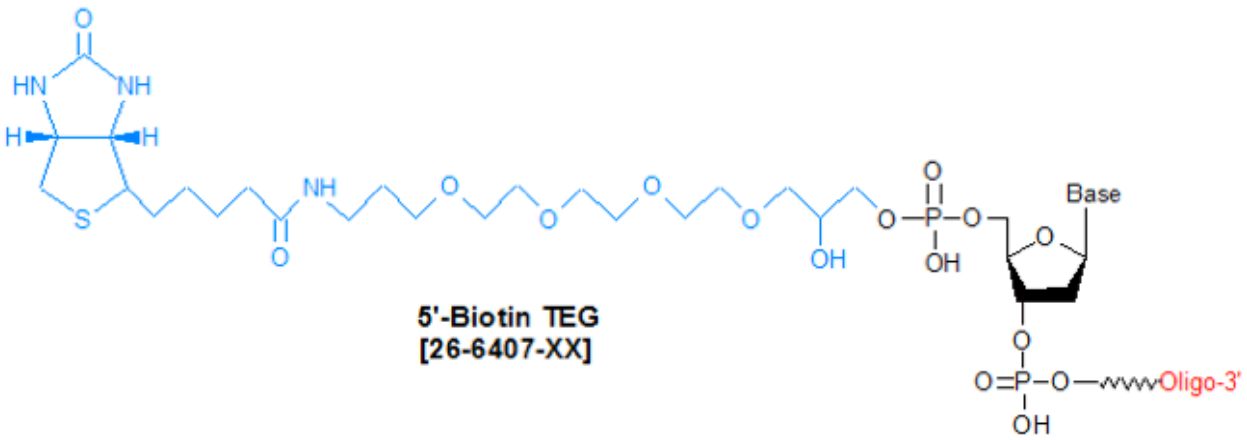
Stellaris Probe Designer version 4.0 designs probes for Stellaris RNA FISH.

Design Notes

- Sequence input is stripped of all non-sequence characters such as FASTA headers and line numbers.
- Please represent uracil bases (U) with a T in the target sequence input.
- Typically, the coding sequences of the target RNA is used as input.
- Probes are designed to minimize deviations in Tm.
- Probes will not be designed across any "N" in the sequence (case insensitive). The "N" is treated as a base for the purpose of meeting probe spacing requirements. Replacing a base with an "N" can be used to prevent design across splice junctions.

FIND PROBES

Reset

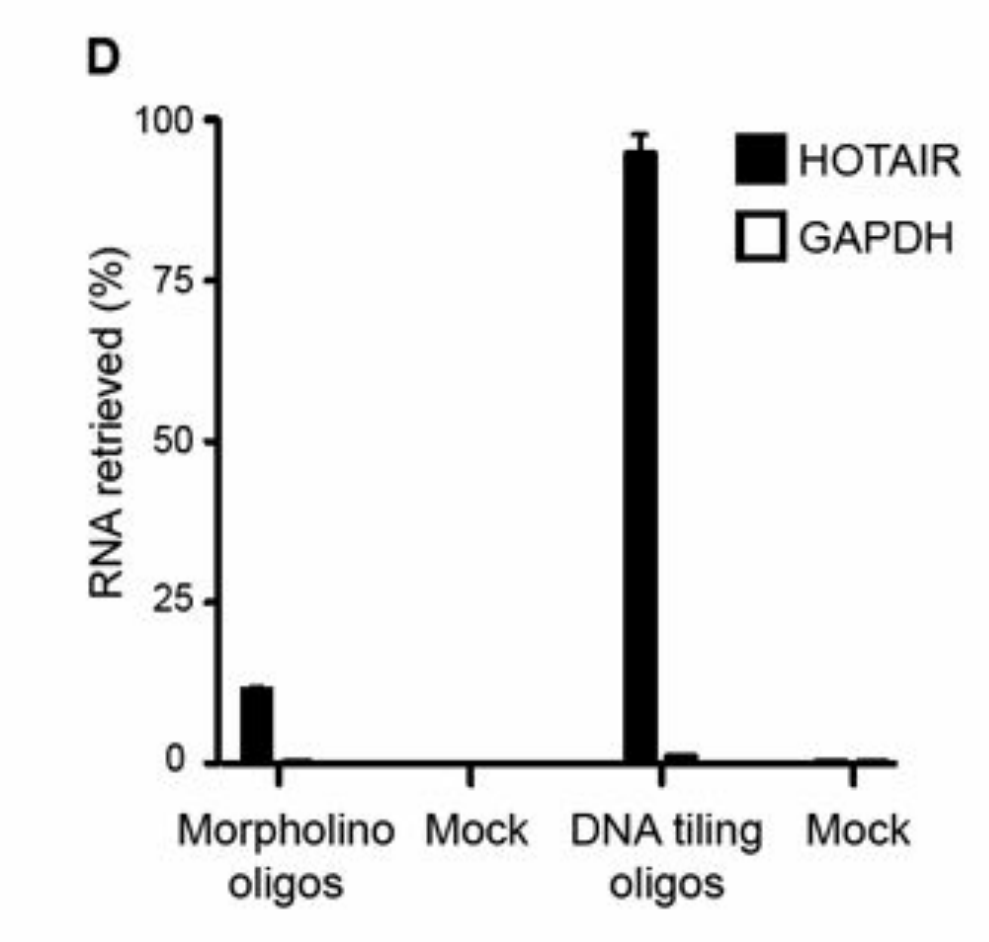


5'-Biotin TEG
[26-6407-XX]

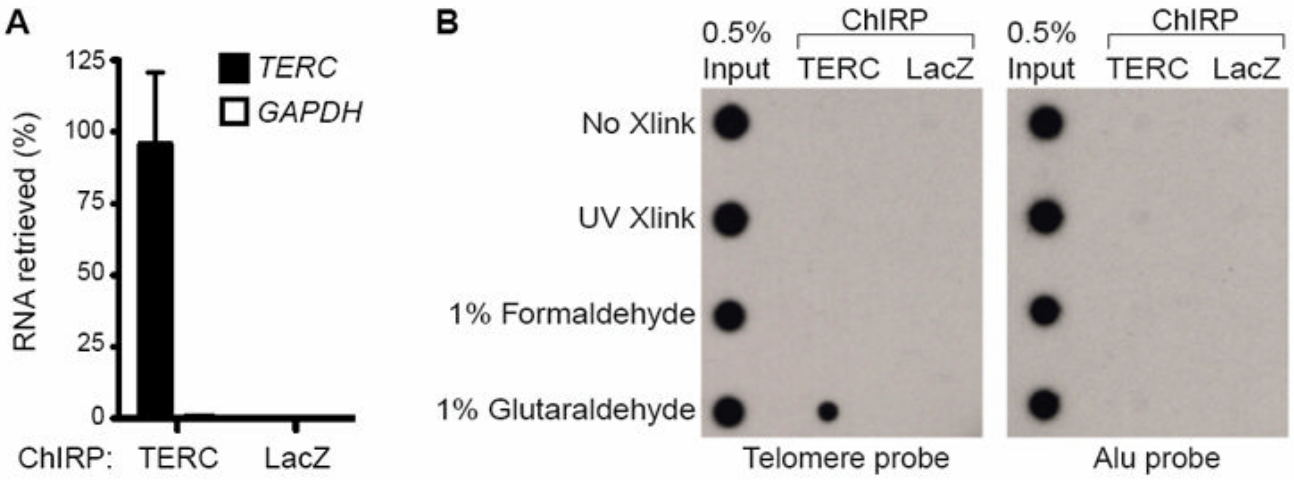
Biotin-TEG is a biotin attached to a 15-atom mixed polarity triethylene glycol spacer

HOTAIR-ChIRP-seq

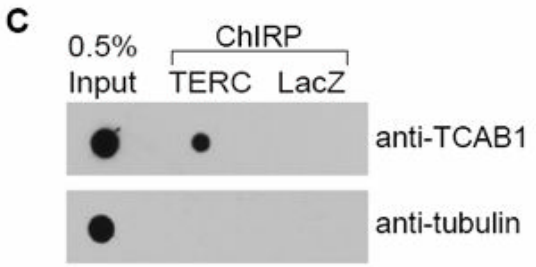
Complementary DNA tiling oligonucleotides effectively retrieve ~95% of HOTAIR RNA from chromatin, as compared to ~10% by morpholino probes.



TERC-ChIRP-seq



Effect of different crosslinking agents on ChIRP-southern. After 1% glutaraldehyde crosslinking, TERC retrieval co-purifies telomeric repeats, but not Alu repeats.

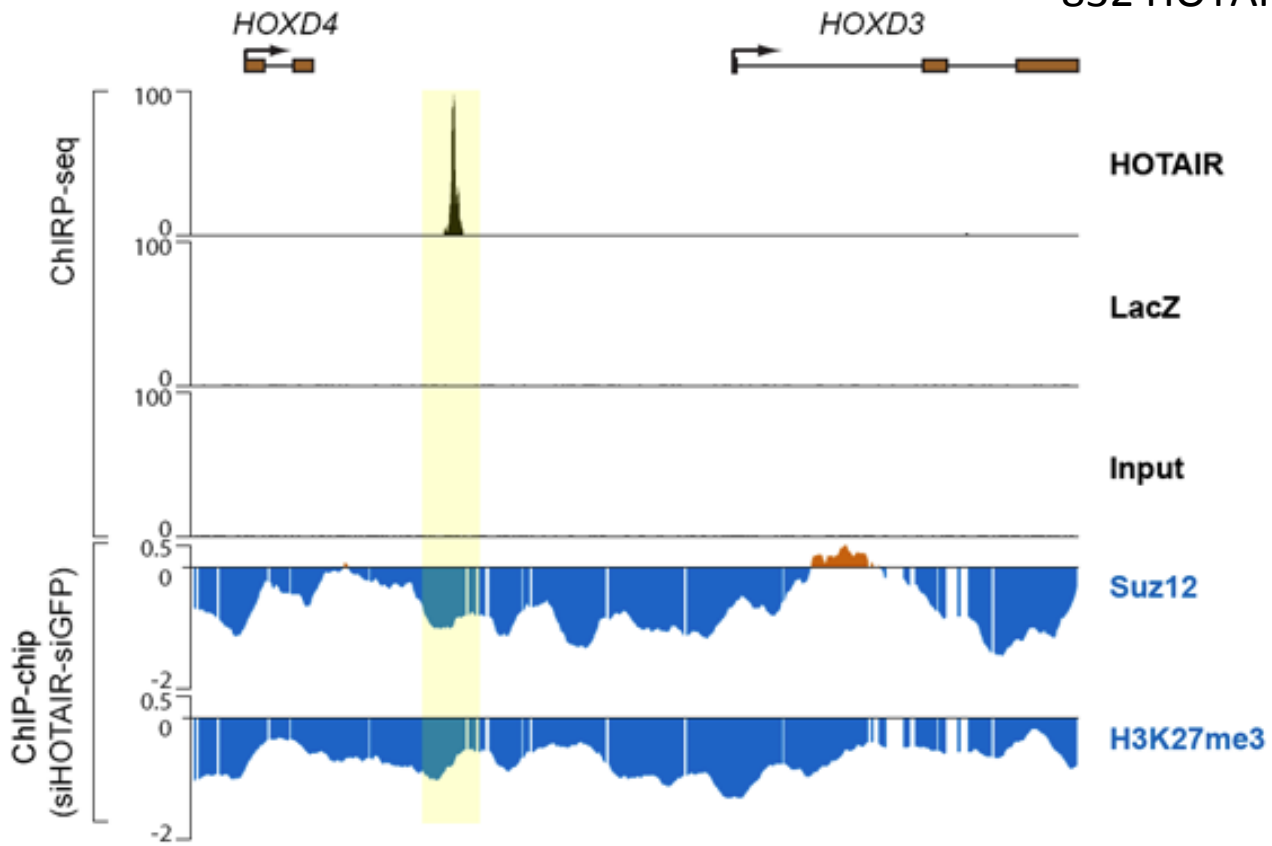


TERC ChIRP retrieves TCAB1, a known telomerase holocomplex chaperone proteins.

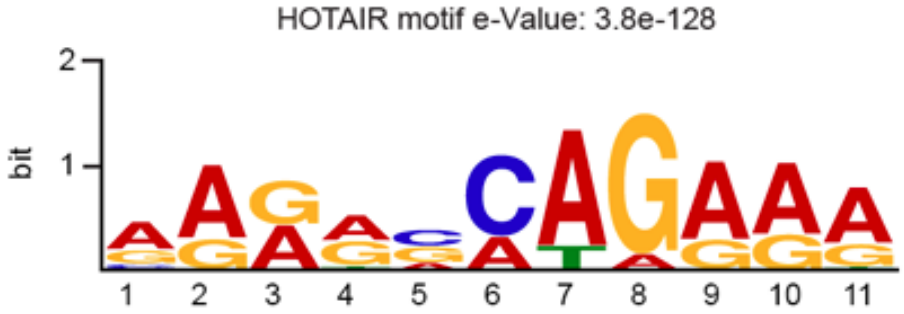
ChIRP is compatible with the simultaneous analysis of *DNA* and *proteins* associated with specific RNAs

ChIRP-seq: genome-wide applications

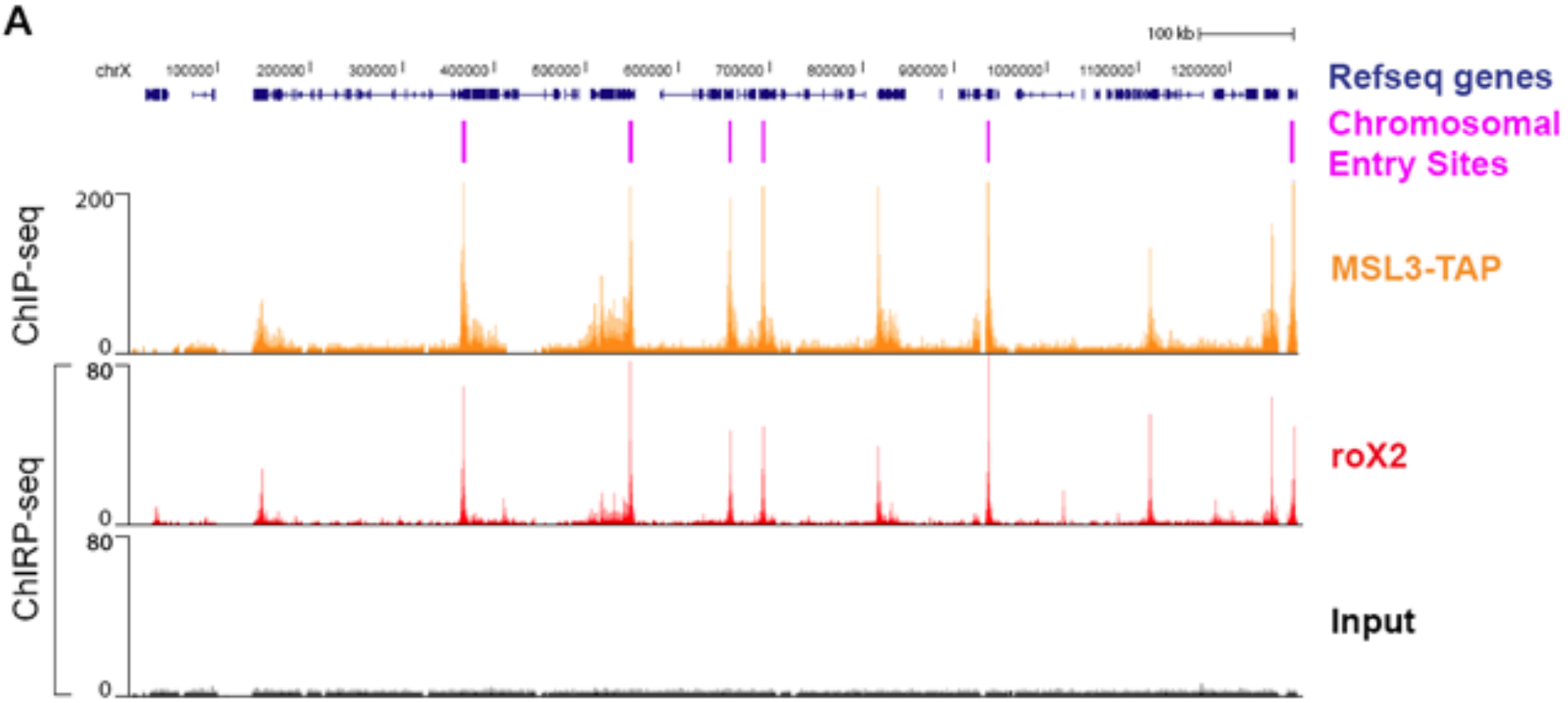
832 HOTAIR occupancy sites



D

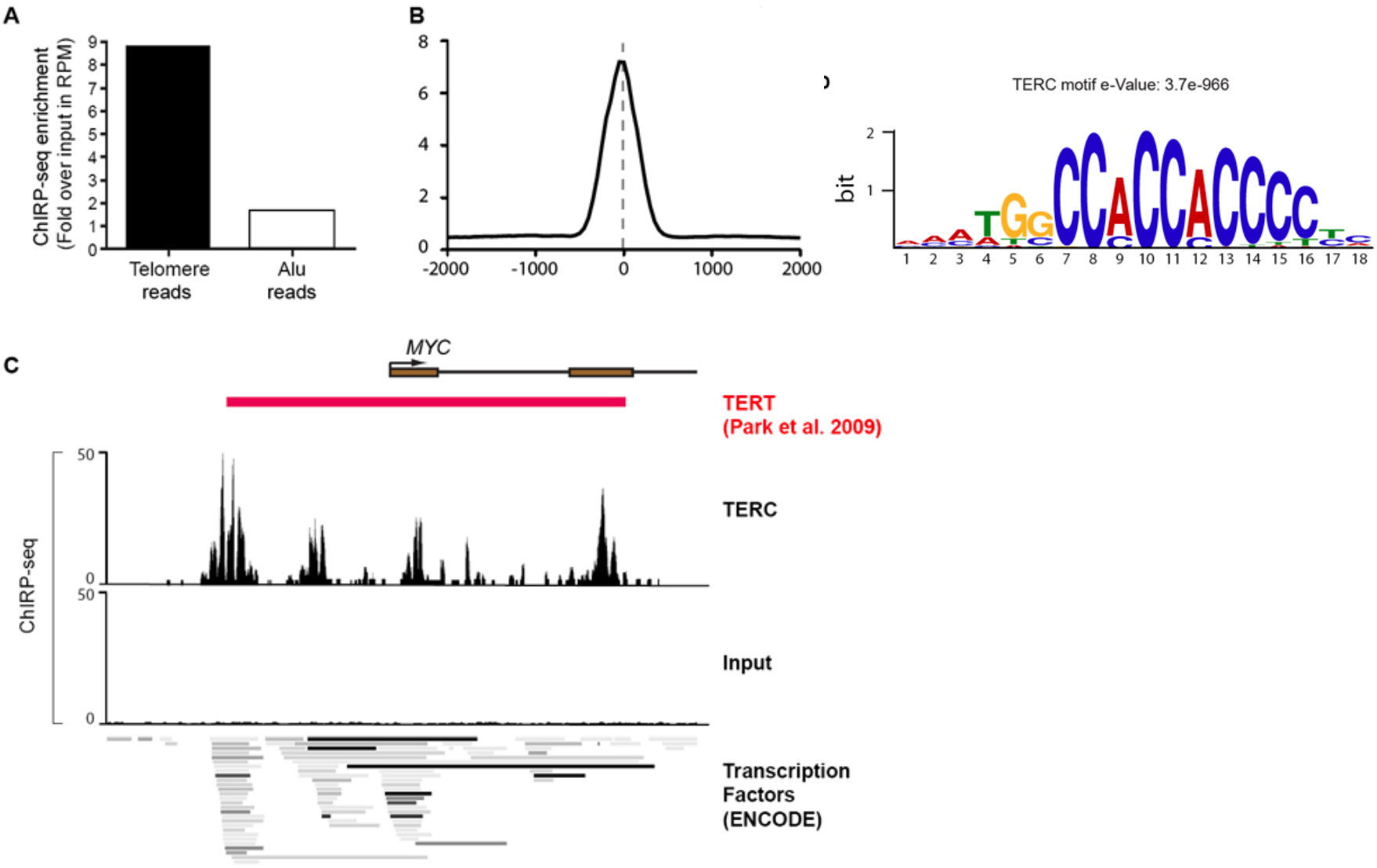


Rox2-ChIRP-seq



- 308 roX2 binding sites--all of them are on the X chromosome
- roX2 occupancy is enriched over gene bodies of X chromosome genes,
- roX2 occupancy increases from 5’ to 3’ end of each gene

TERC-ChIRP-seq



Capture Hybridization Analysis of RNA Targets (*ChART*)

The genomic binding sites of a noncoding RNA

Matthew D. Simon[a], Charlotte I. Wang[b], Peter V. Kharchenko[c], Jason A. West[a], Brad A. Chapman[a], Artyom A. Alekseyenko[b], Mark L. Borowsky[a], Mitzi I. Kuroda[b], and Robert E. Kingston[a,1]

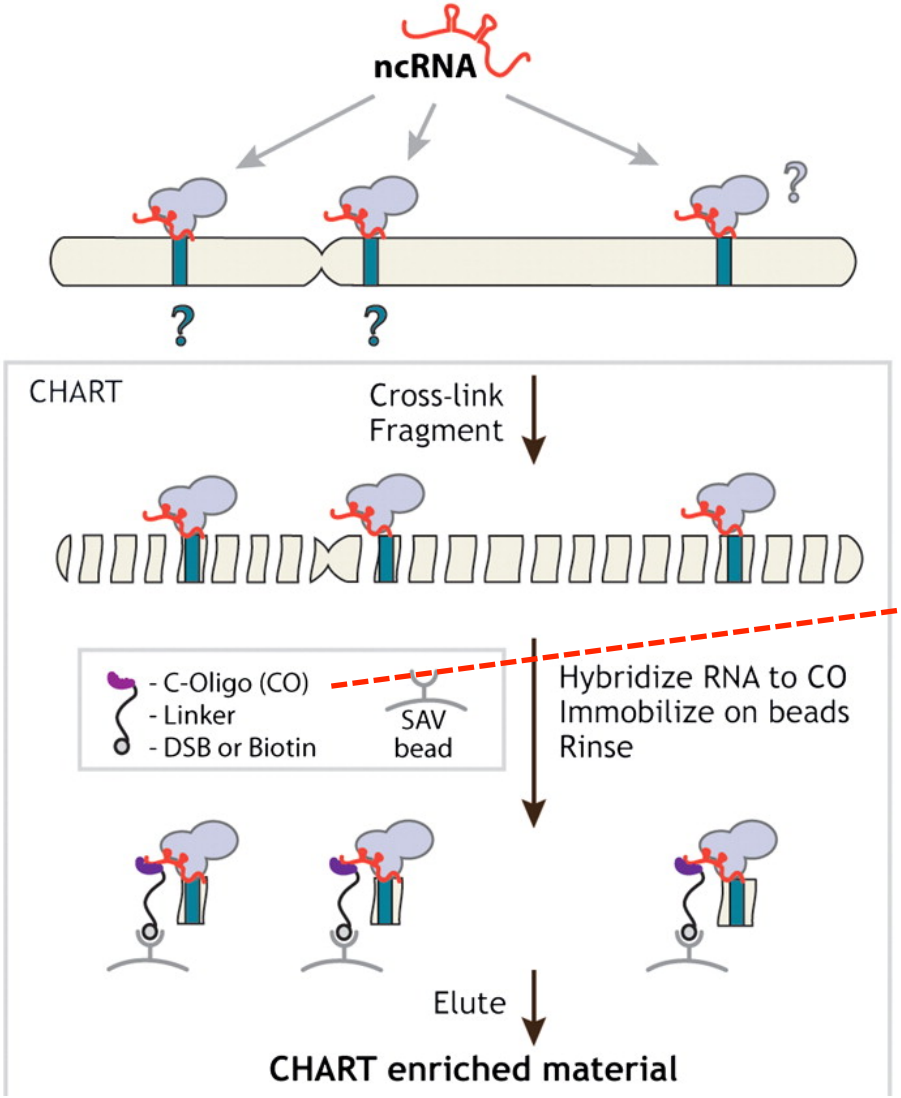
Author Affiliations

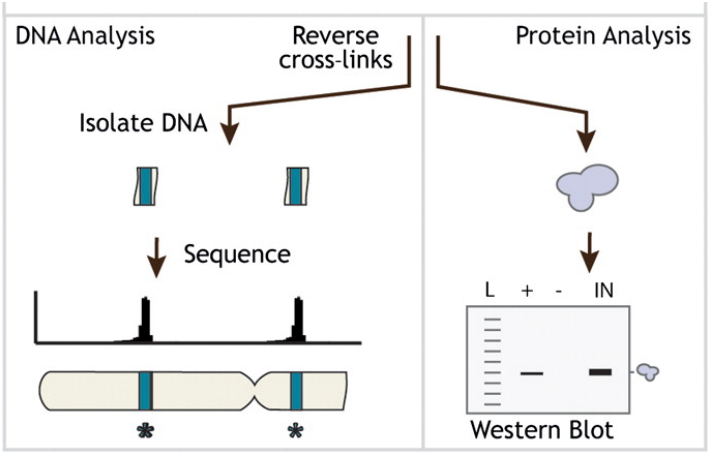
Edited by* Keith R. Yamamoto, University of California, San Francisco, CA, and approved October 19, 2011 (received for review August 17, 2011)

Abstract

Long noncoding RNAs (lncRNAs) have important regulatory roles and can function at the level of chromatin. To determine where lncRNAs bind to chromatin, we developed capture hybridization analysis of RNA targets (CHART), a hybridization-based technique that specifically enriches endogenous RNAs along with their targets from reversibly cross-linked chromatin extracts. CHART was used to enrich the DNA and protein targets of endogenous lncRNAs from flies and humans. This analysis was extended to genome-wide mapping of *roX2*, a well-studied ncRNA involved in dosage compensation in *Drosophila*. CHART revealed that *roX2* binds at specific genomic sites that coincide with the binding sites of proteins from the male-specific lethal complex that affects dosage compensation. These results reveal the genomic targets of *roX2* and demonstrate how CHART can be used to study RNAs in a manner analogous to chromatin immunoprecipitation for proteins.

Capture hybridization analysis of RNA targets





Capture Oligonucleotides.

RNAseH mapping

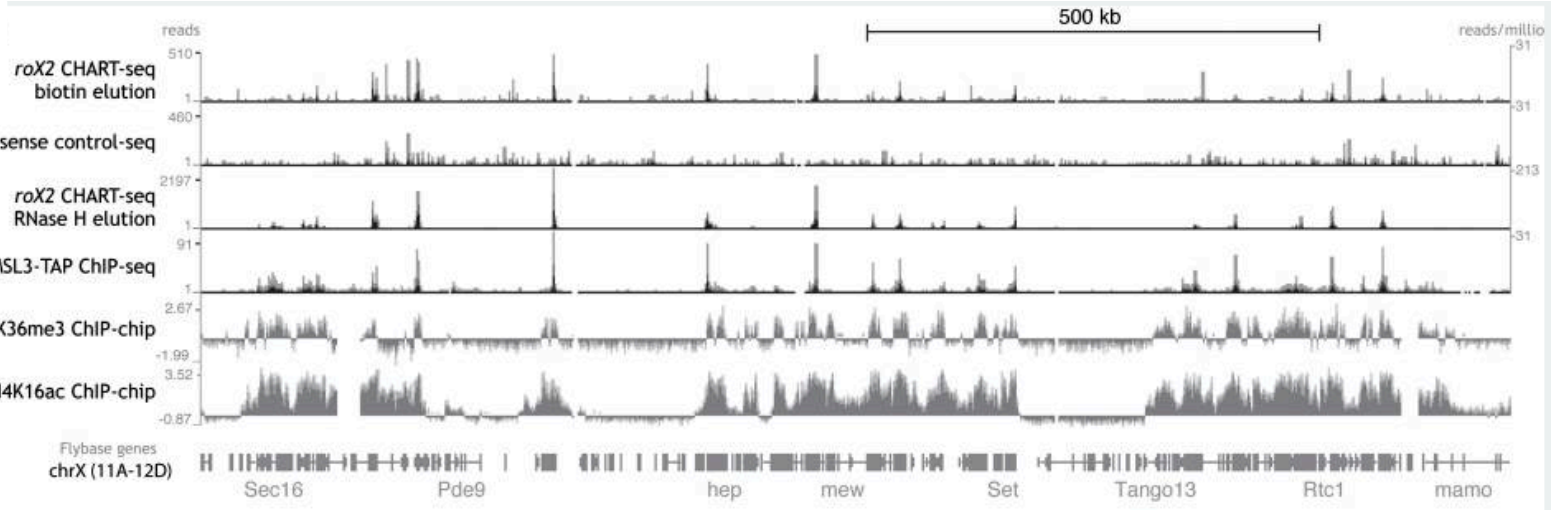
to ensure that these C-oligos would target stretches of *roX2* RNA available for hybridization and not occluded by protein binding or secondary structure.

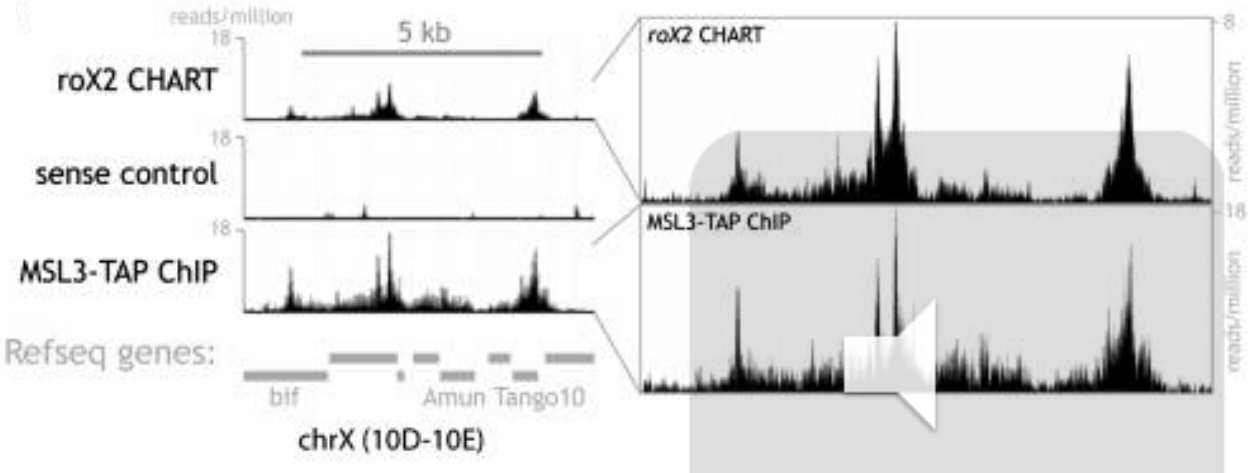
RNase-H mapping assay to probe sites on *roX2* available to hybridization in the context of cross-linked chromatin extracts. RNase-H specifically hydrolyzes the RNA strand of a DNA-RNA hybrid

Simon M D et al. PNAS 2011;108:20497-20502

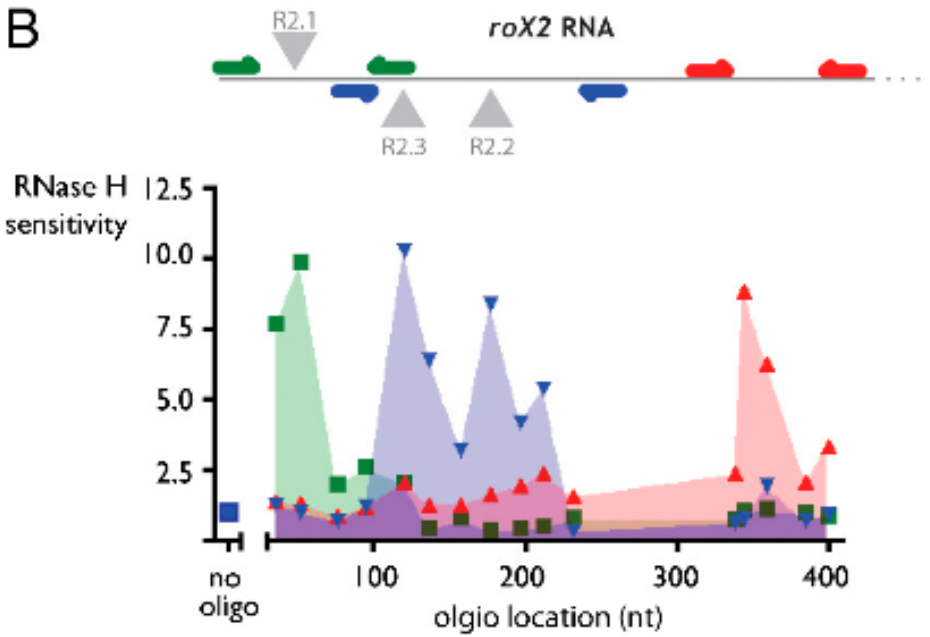
roX2 CHART-seq reveals enrichment of *roX2* on chrX and localization to sites of MSL binding

- *roX2* is known to localize to the X chromosome (chrX), where it acts together with the MSL complex (including protein subunits MSL1, MSL2, MSL3, MLE, and MOF).
- The MSL complex affects dosage compensation, at least in part, by regulating *acetylation of histone H4 lysine 16 (H4K16)* in the bodies of active genes and influencing transcriptional *elongation*.



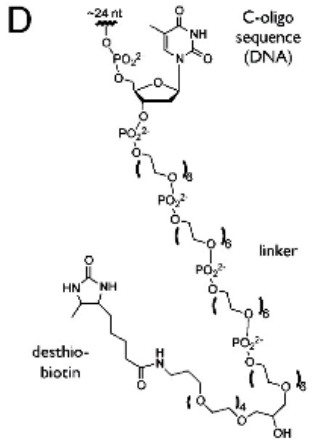


Development of C-oligos for CHART: RNAseH mapping



biotin-eluted CHART

RNase-H–eluted CHART



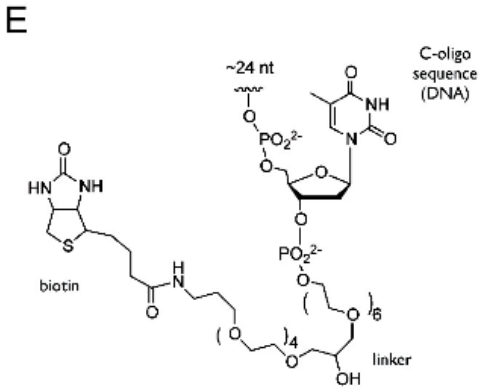
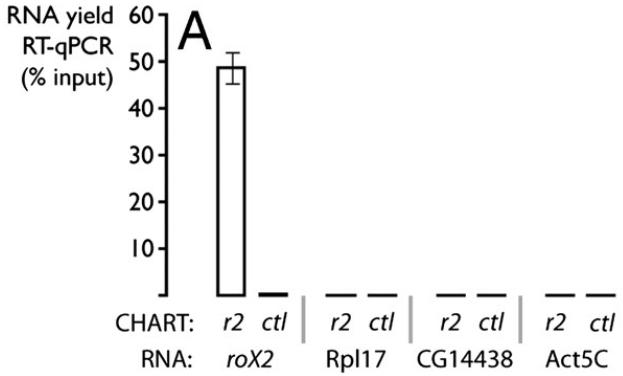


CHART allows specific enrichment of roX2 along with its associated targets

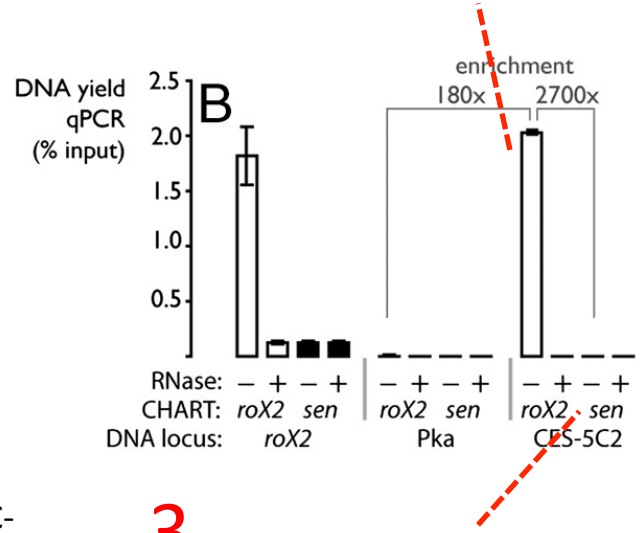


1

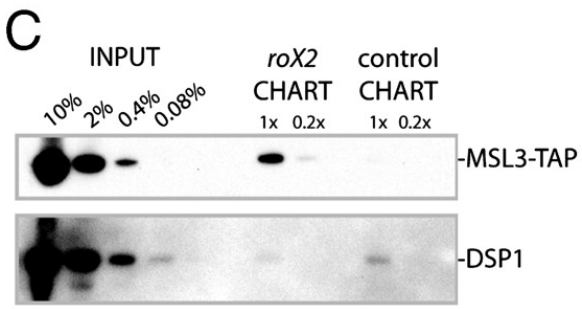
A) *Three* 25-mer DNA-based C-oligos provided a low background signal and high specific yields of *roX2*

B) This enrichment was specific; CHART using a scrambled control C-oligo did not enrich *roX2*, and control RNAs were not enriched by *roX2* CHART

2 candidate genomic sites: known regulatory site of dosage compensation, chromatin entry site 5C2 (CES-5C2)



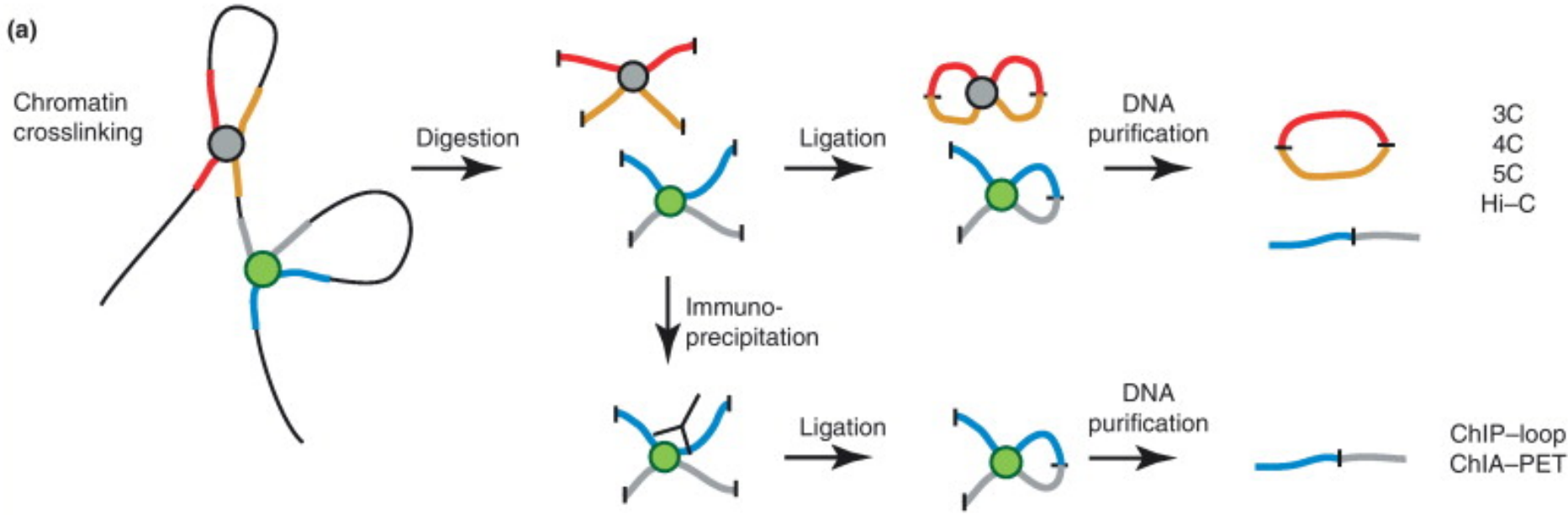
3 RNA-dependence: CHART enrichment from extracts pre-treated with RNase to eliminate RNA-mediated signal



In addition to DNA, also known binding partners are enriched by *roX2* CHART as the subunit of the male-specific lethal complex (MSL3) relative to a scrambled control by *roX2* CHART

Chromosome conformation capture (3C) technologies

The library represents the *sum* of DNA–DNA interactions over the cell population



Nature. 2014 Nov 20;515(7527):402-5. doi: 10.1038/nature13986.

Topologically associating domains are stable units of replication-timing regulation.

Pope BD[1], Ryba T[2], Dileep V[1], Yue F[3], Wu W[4], Denas O[5], Vera DL[1], Wang Y[6], Hansen RS[7], Canfield TK[8], Thurman RE[8], Cheng Y[9], Gülsoy G[10], Dennis JH[1], Snyder MP[9], Stamatoyannopoulos JA[8], Taylor J[5], Hardison RC[4], Kahveci T[10], Ren B[11], Gilbert DM[1].

Author information

Abstract

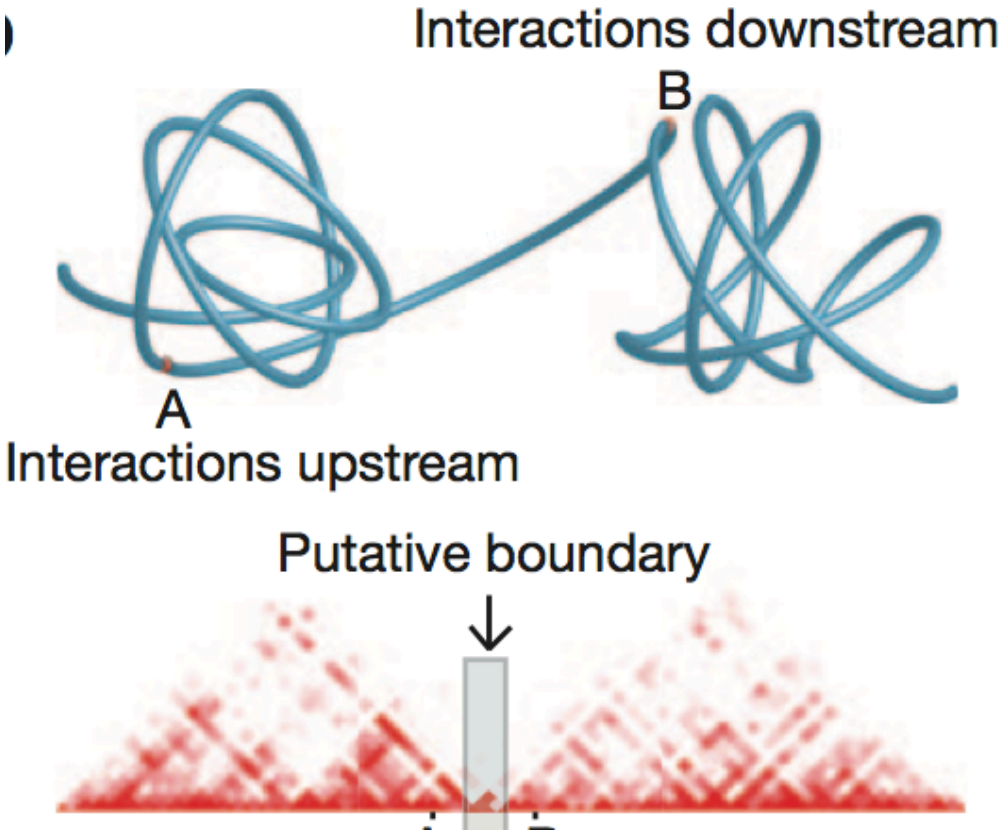
Eukaryotic chromosomes replicate in a temporal order known as the replication-timing program. In mammals, replication timing is cell-type-specific with at least half the genome switching replication timing during development, primarily in units of 400-800 kilobases ('replication domains'), whose positions are preserved in different cell types, conserved between species, and appear to confine long-range effects of chromosome rearrangements. Early and late replication correlate, respectively, with open and closed three-dimensional chromatin compartments identified by high-resolution chromosome conformation capture (Hi-C), and, to a lesser extent, late replication correlates with lamina-associated domains (LADs). Recent Hi-C mapping has unveiled substructure within chromatin compartments called topologically associating domains (TADs) that are largely conserved in their positions between cell types and are similar in size to replication domains. However, TADs can be further sub-stratified into smaller domains, challenging the significance of structures at any particular scale. Moreover, attempts to reconcile TADs and LADs to replication-timing data have not revealed a common, underlying domain structure. Here we localize boundaries of replication domains to the early-replicating border of replication-timing transitions and map their positions in 18 human and 13 mouse cell types. We demonstrate that, collectively, replication domain boundaries share a near one-to-one correlation with TAD boundaries, whereas within a cell type, adjacent TADs that replicate at similar times obscure replication domain boundaries, largely accounting for the previously reported lack of alignment. Moreover, cell-type-specific replication timing of TADs partitions the genome into two large-scale sub-nuclear compartments revealing that replication-timing transitions are indistinguishable from late-replicating regions in chromatin composition and lamina association and accounting for the reduced correlation of replication timing to LADs and heterochromatin. Our results reconcile cell-type-specific sub-nuclear compartmentalization and replication timing with developmentally stable structural domains and offer a unified model for large-scale chromosome structure and function.

LETTER

doi:10.1038/nature11082

Topological domains in mammalian genomes identified by analysis of chromatin interactions

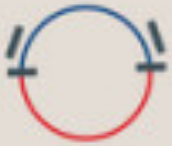
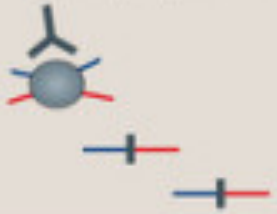
Jesse R. Dixon[1,2,3], Siddarth Selvaraj[1,4], Feng Yue[1], Audrey Kim[1], Yan Li[1], Yin Shen[1], Ming Hu[5], Jun S. Liu[5] & Bing Ren[1,6]



The spatial organization of the genome is intimately linked to its biological function, yet our understanding of higher order genomic structure is coarse, fragmented and incomplete. In the nucleus of eukaryotic cells, interphase chromosomes occupy distinct chromosome territories, and numerous models have been proposed for how chromosomes fold within chromosome territories[1]. These models, however, provide only few mechanistic details about the relationship between higher order chromatin structure and genome function. Recent advances in genomic technologies have led to rapid advances in the study of three-dimensional genome organization. In particular, Hi-C has been introduced as a method for identifying higher order chromatin interactions genome wide[2]. Here we investigate the three-dimensional organization of the human and mouse genomes in embryonic stem cells and terminally differentiated cell types at unprecedented resolution. We identify large, megabase-sized local chromatin interaction domains, which we term 'topological domains', as a pervasive structural feature of the genome organization. These domains correlate with regions of the genome that constrain the spread of heterochromatin. The domains are stable across different cell types and highly conserved across species, indicating that topological domains are an inherent property of mammalian genomes. Finally, we find that the boundaries of topological domains are enriched for the insulator binding protein CTCF, housekeeping genes, transfer RNAs and short interspersed element (SINE) retrotransposons, indicating that these factors may have a role in establishing the topological domain structure of the genome.

a **3C: converting chromatin interactions into ligation products**

b **Ligation product detection methods**

3C	4C	5C	ChIA–PET	Hi-C
One-by-one All-by-all	One-by-all	Many-by-many	Many-by-many	All-by-all
			• DNA shearing • Immunoprecipitation 	• Biotin labelling of ends • DNA shearing 
PCR or sequencing	Inverse PCR sequencing	Multiplexed LMA sequencing	Sequencing	Sequencing

PENNSTATE

Welcome to YUE Lab

Computational and Functional Genomics/Epigenomics