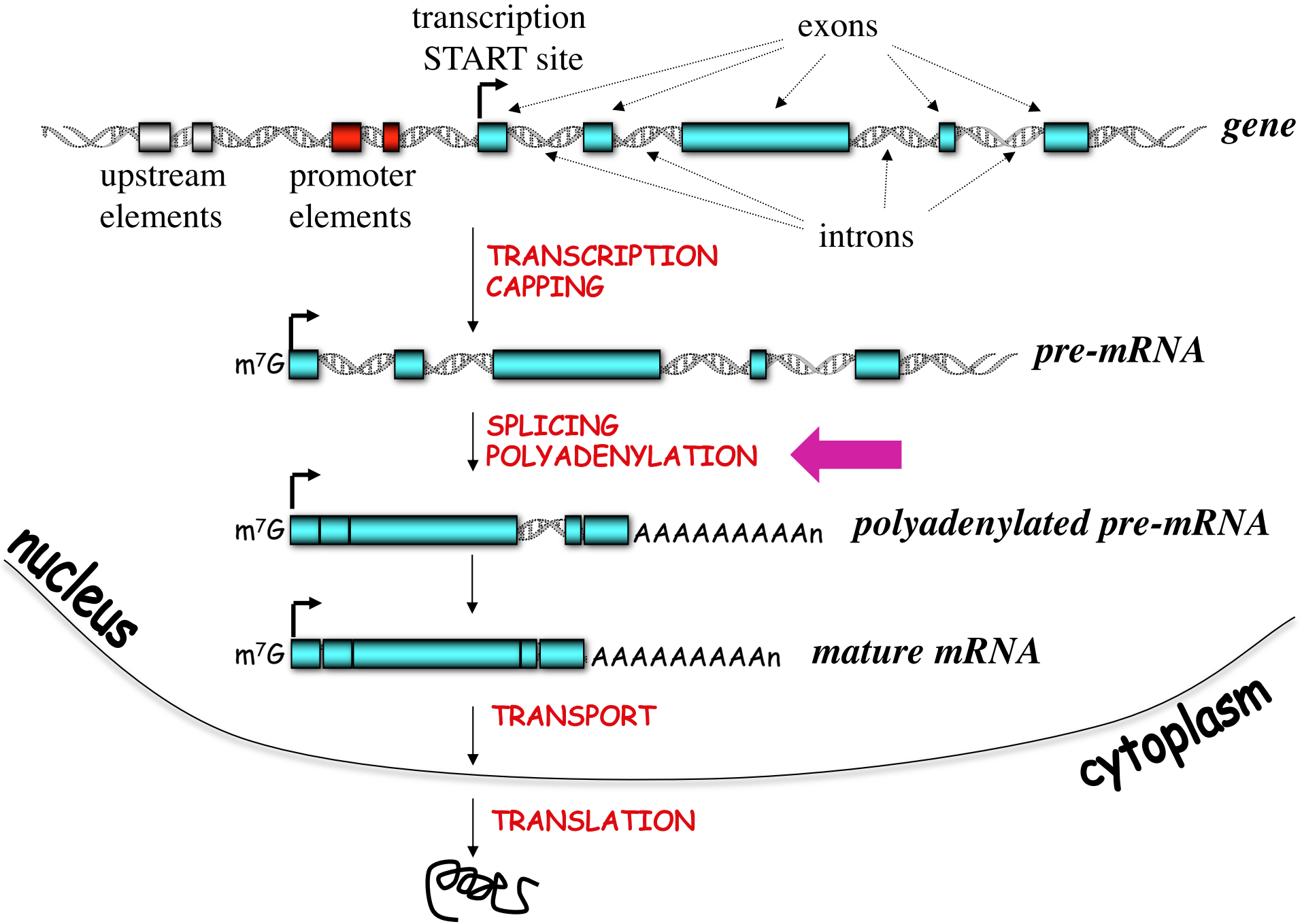
# 3′-end formation

# Eukaryotic gene expression



# What is the function of the PAP-dependendent polyA tail?

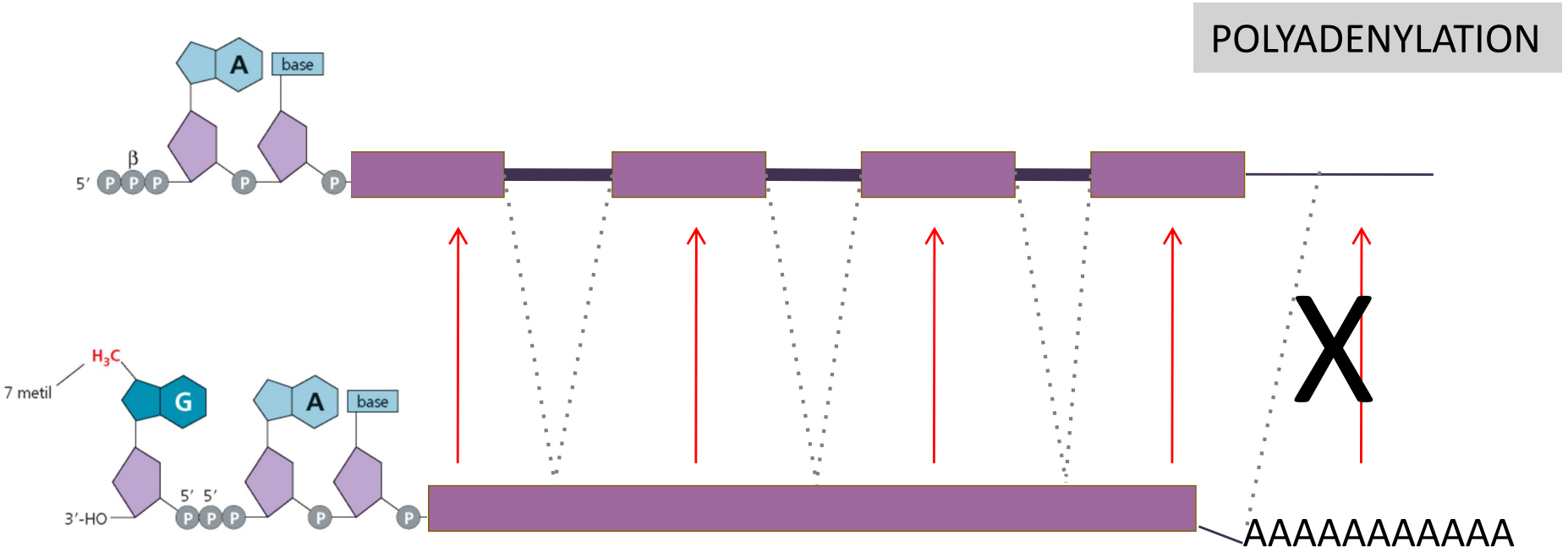
1) Increases RNA stability

2) Favours the mRNA transport to the cytoplasm

3) Increases translation efficiency by favouring the loading of ribosomal 40S subunit

4) Stimulates mRNA transcription termination

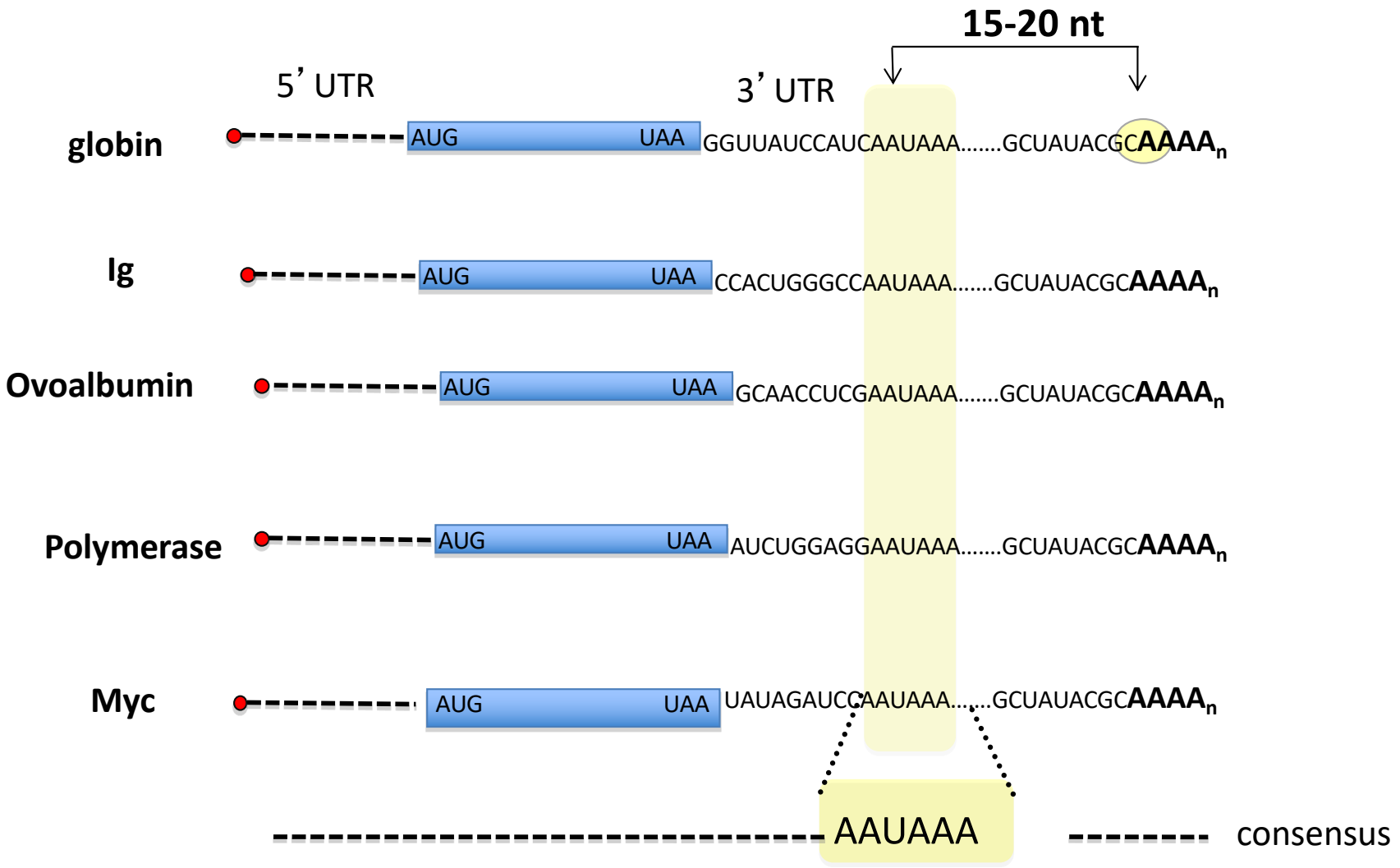
# mRNA maturation




The polyA tail is not transcribed from the DNA but it is added to the mRNA during a process named polyadenylation.

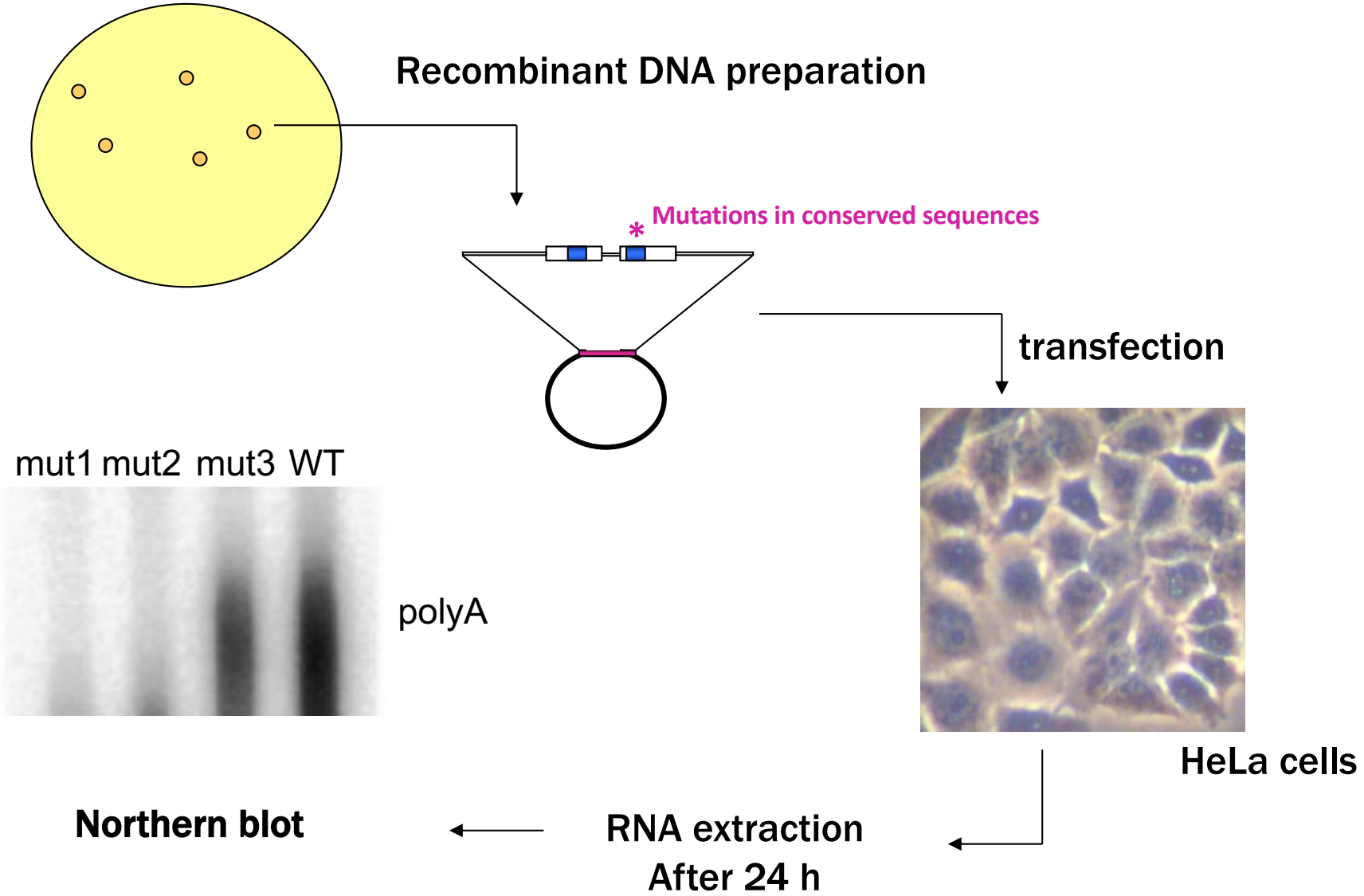
# Looking for consensus sequences in mammals

SEQUENCE ALLINEAMENT OF cDNAs STARTING FROM THE POLYA TAIL

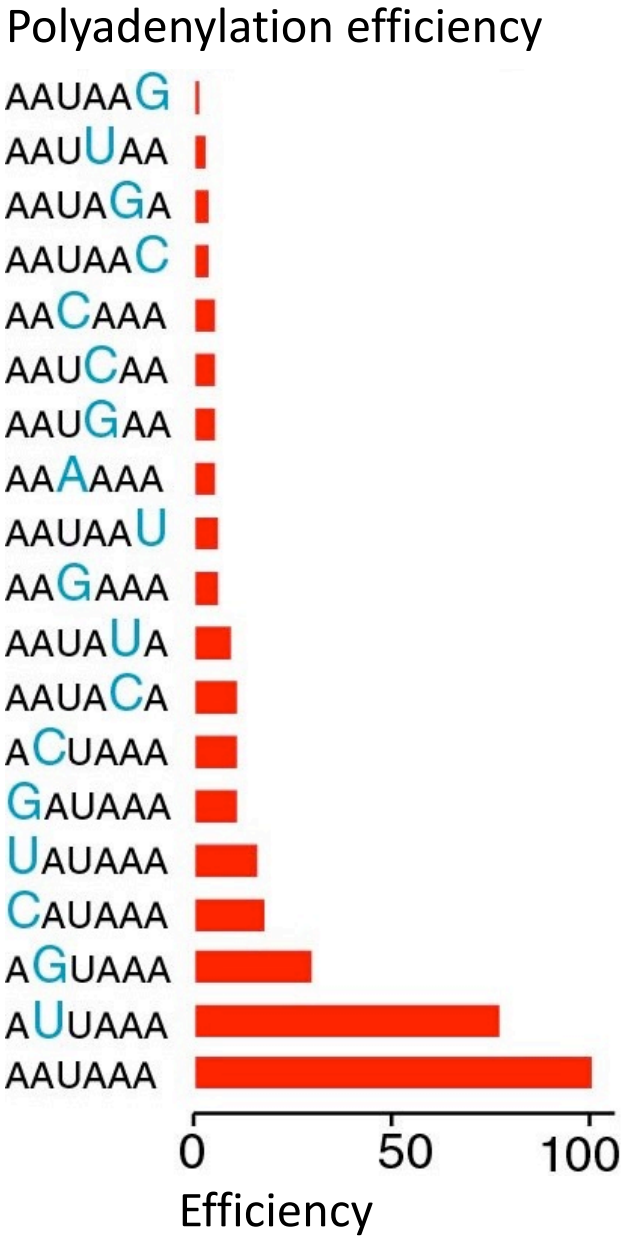


# LOOKING FOR CONSENSUS SEQUENCES IMPORTANT FOR POLYADENYLATION

# *In vivo* polyadelynation analysis



# *In vivo* polyadelynation analysis



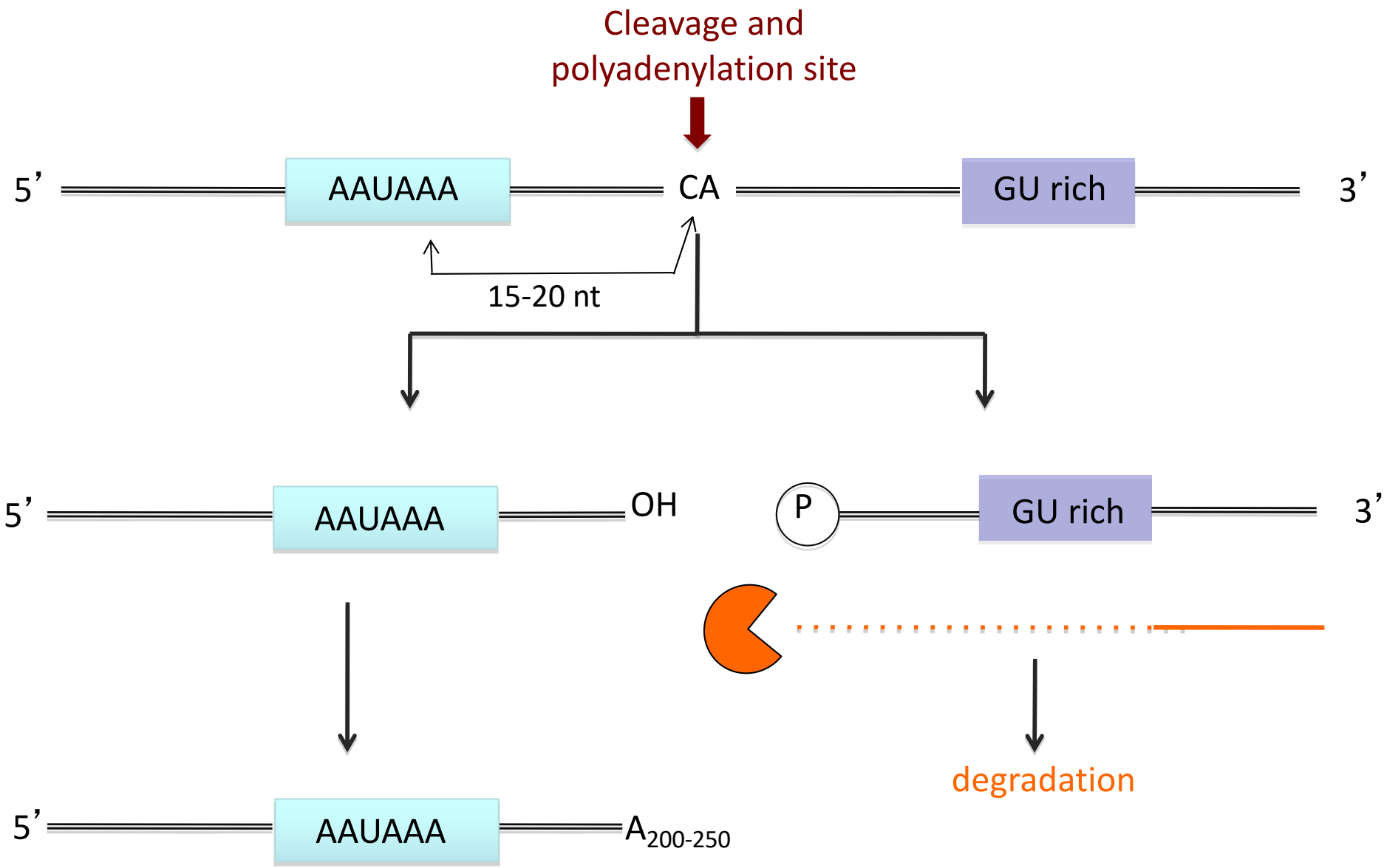

***mammals***

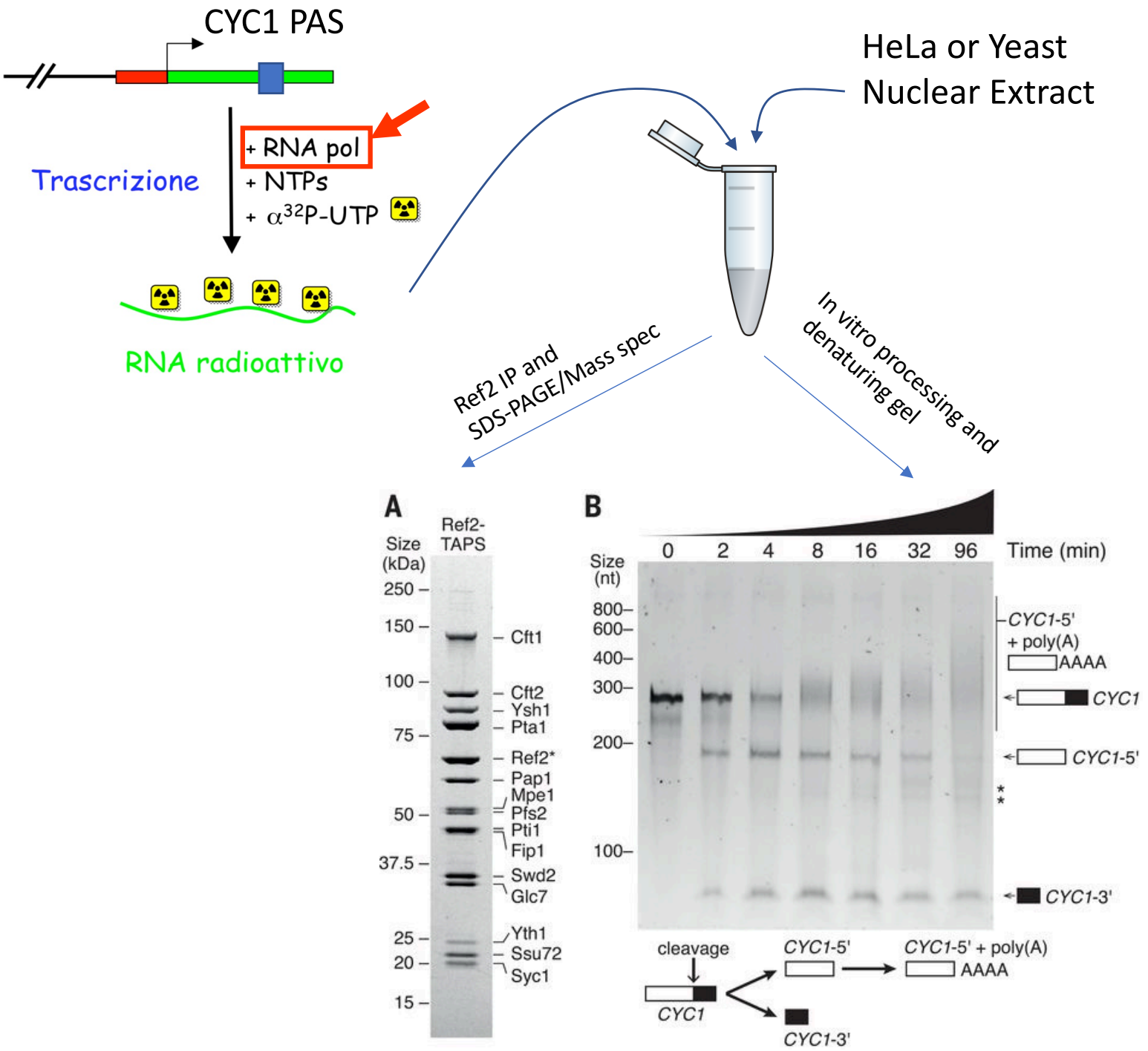
consensus

$A_{98}A_{86}U_{98}A_{98}A_{95}A_{96}$
$U_{12}$

- AAUAAA in mammals Located ~20-30 bp from the polyA site
  - Other hexamers less efficient but are used
- Mutagenesis and *in vivo* expression studies reveal also the importance of the GU-rich downstream of AAUAAA sequence.

# 3′-end formation in mammalian cells




CYC1 PAS
Trascrizione
+ RNA pol
+ NTPs
+ α32P-UTP
RNA radioattivo
HeLa or Yeast
Nuclear Extract
Ref2 IP and
SDS-PAGE/Mass spec
In vitro processing and
denaturing gel
A
Size
(kDa)
Ref2-
TAPS
250
150
100
75
50
37.5
25
20
15
Cft1
Cft2
Ysh1
Pta1
Ref2*
Pap1
Mpe1
Pfs2
Pti1
Fip1
Swd2
Glc7
Yth1
Ssu72
Syc1
B
0
2
4
8
16
32
96
Time (min)
Size
(nt)
800
600
400
300
200
100
CYC1-5'
+ poly(A)
AAAA
CYC1
CYC1-5'
CYC1-3'
cleavage
CYC1
CYC1-5'
CYC1-5' + poly(A)
AAAA
CYC1-3'

# PAS sequences are not conserved in evolution

*Cis* elements around the PAS in simpler species differ from those in mammals:

- downstream GU-rich elements are not present in nematode PASs.
- AAUAAA is prominent in *Schizosaccharomyces pombe*, while a general A-rich sequence is present in the same region in *Saccharomyces cerevisiae*. Yeast PASs do not have downstream GU-rich elements. However, UAUA elements are highly enriched in the upstream region.
- a typical plant PAS is more similar to that of yeast than of metazoan, with a loosely defined upstream AAUAAA element flanked by U-rich sequences.

# The polyadenylation machinery in metazoans

The polyadenylation machinery in metazoans is composed of ~20 core proteins, including four protein complexes and several single proteins:

- **cleavage and polyadenylation specificity factor (CPSF)**, which contains **CPSF160** (also known as CPSF1), **CPSF100** (also known as CPSF2), **CPSF73** (also known as CPSF3), **CPSF30** (also known as CPSF4), **FIP1** (factor interacting with PAP) and **WDR33**

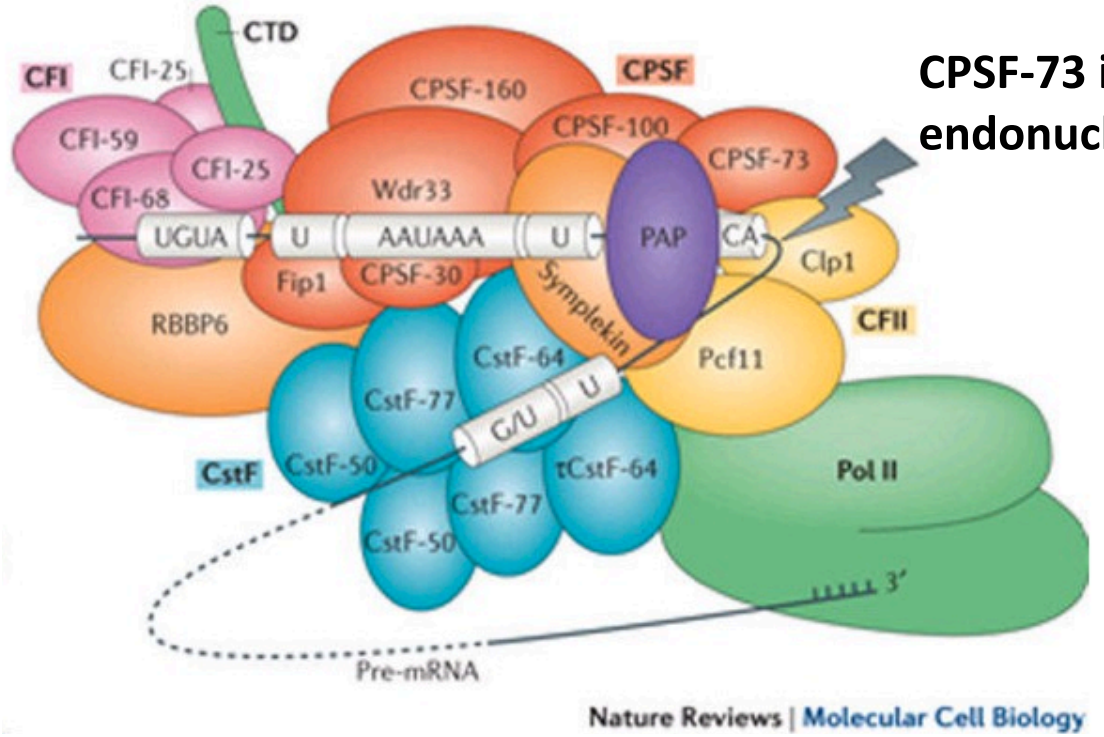
- **cleavage stimulation factor (CSTF)**, which contains **CSTF 77**, **CSTF50** (also known as CSTF1) and **CSTF64**

- **cleavage factor I (CFI)**, which contains **CFI25** and either **CFI68** or **CFI59**;

- **cleavage factor II (CFII)**, which contains **PCF11** and **CLP1**

- Single proteins include **symplexin**, **poly(A) polymerase (PAP)**, **retinoblastoma-binding protein 6 (RBBP6)**.

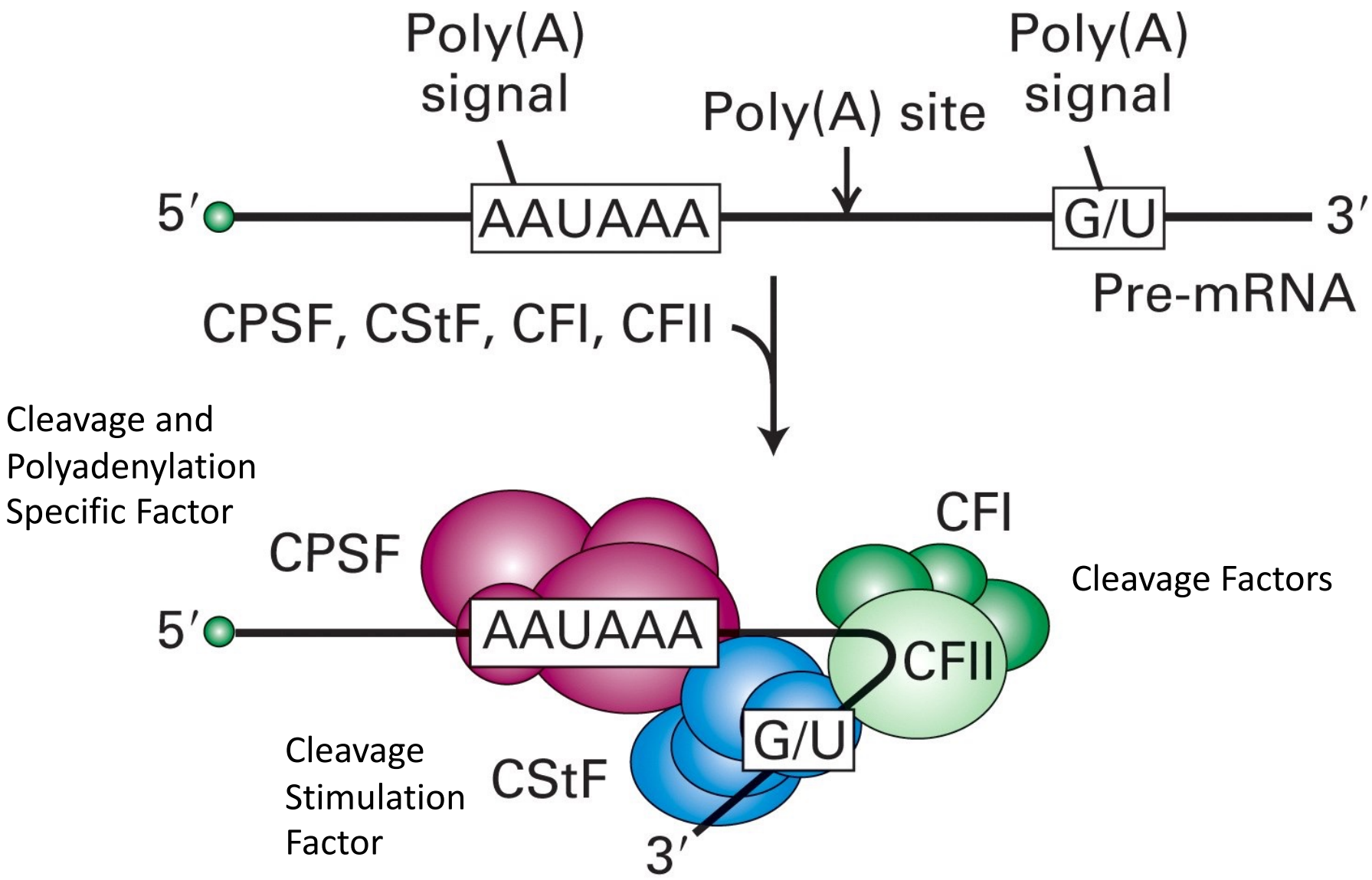
# Specific complexes are involved in mRNA 3'end-processing

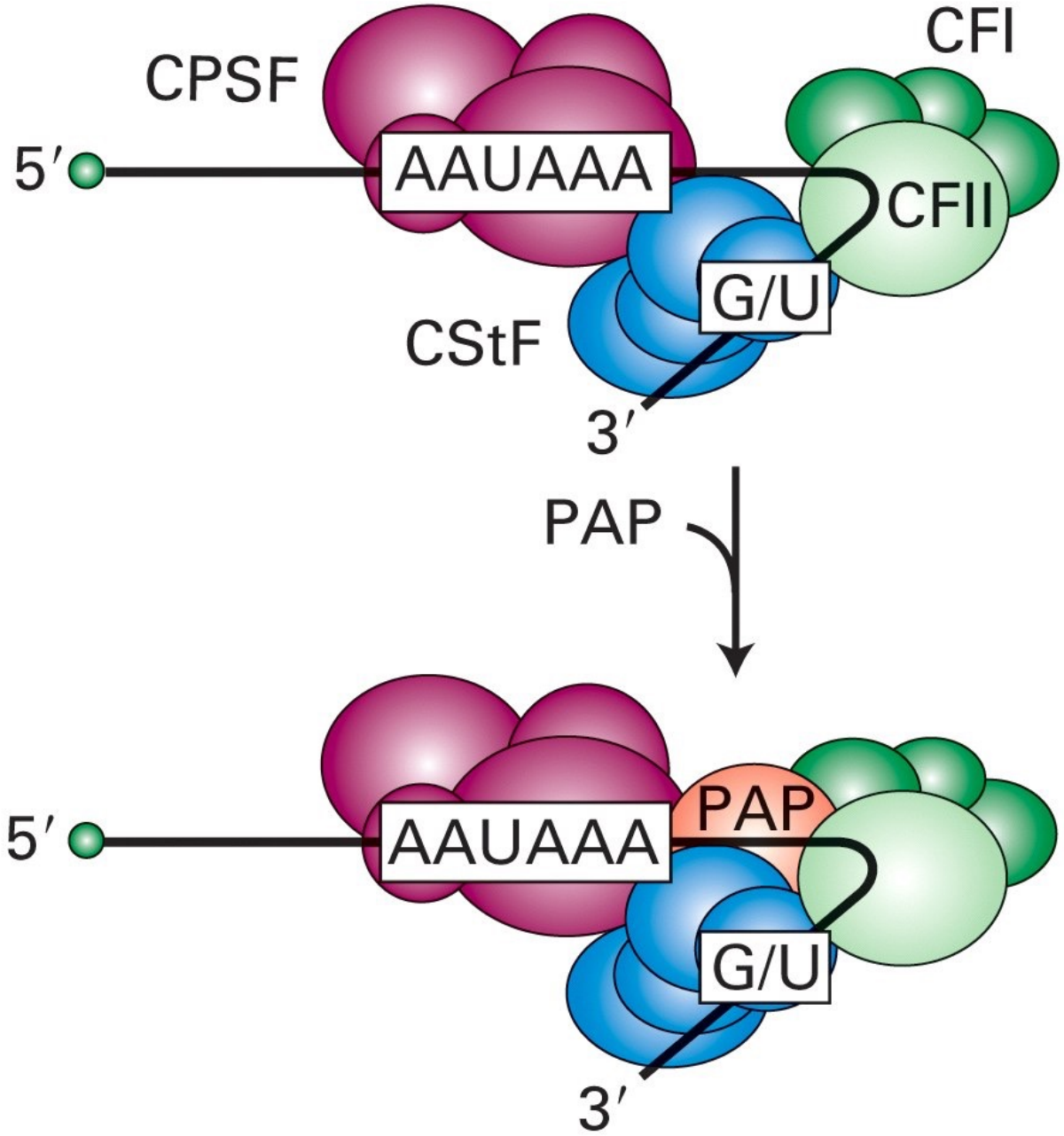


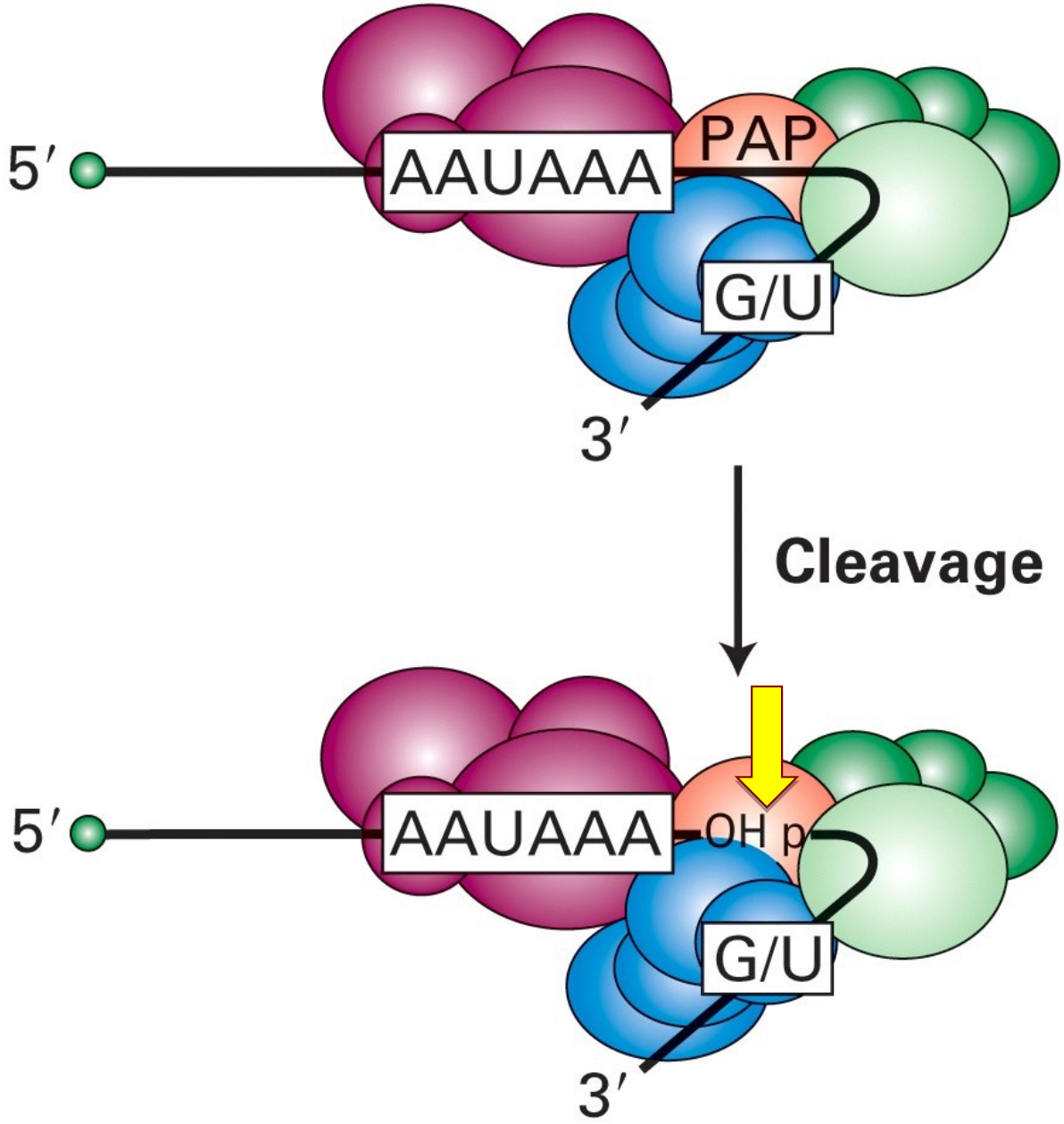

**CPSF-73 is the endonuclease**

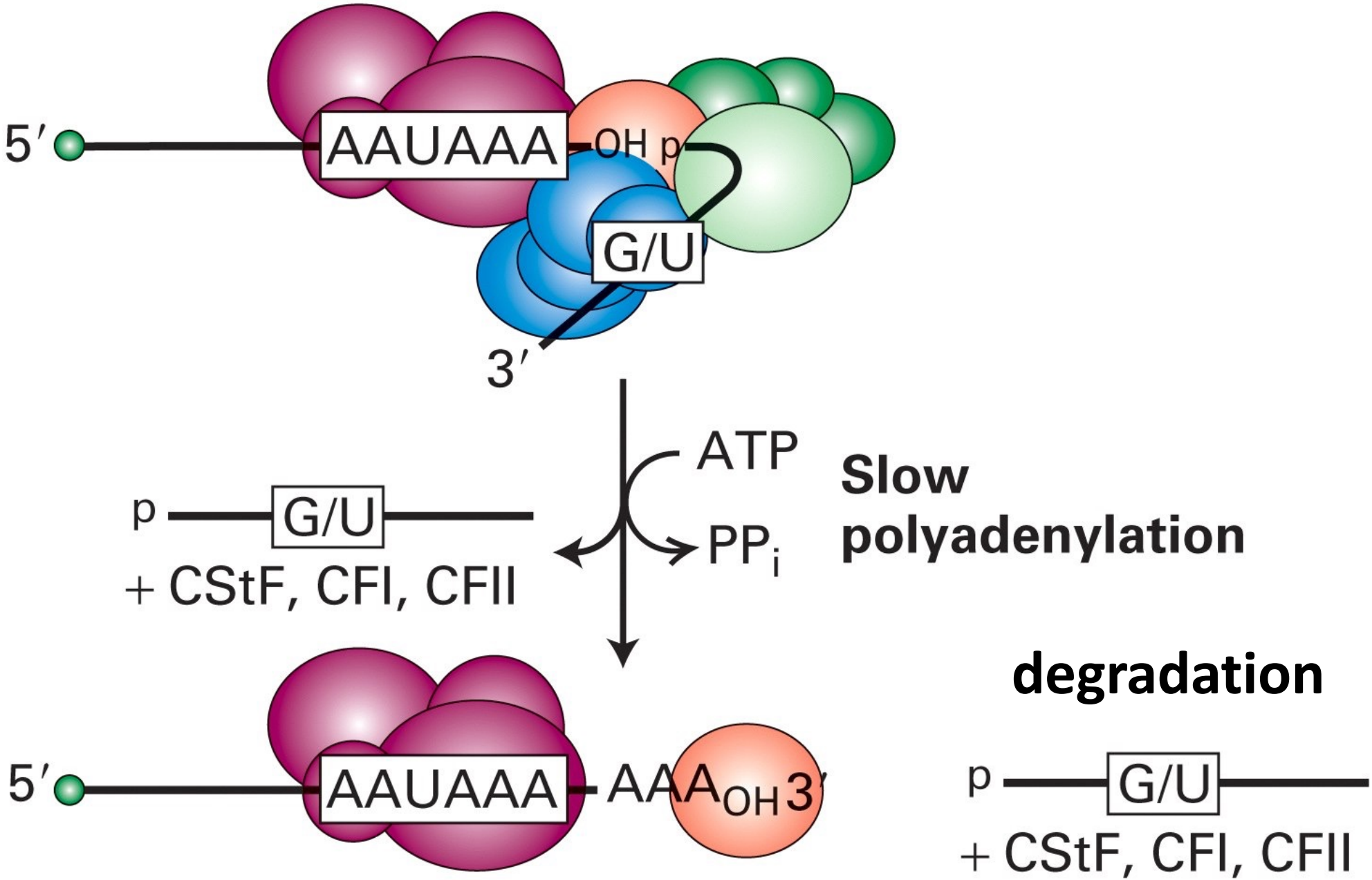
Despite considerable divergence between yeast and mammals in the core RNA sequences that constitute the PAS, nearly all mammalian polyadenylation factors have homologues in yeast, with the exception of the CFI proteins and CSTF50. Moreover, the yeast polyadenylation factor Hrp1p, which interacts with UA-rich elements, is missing from metazoans. The polyadenylation machinery in plants is similar to that in metazoans, but with substantial gene (and thus protein) duplications.

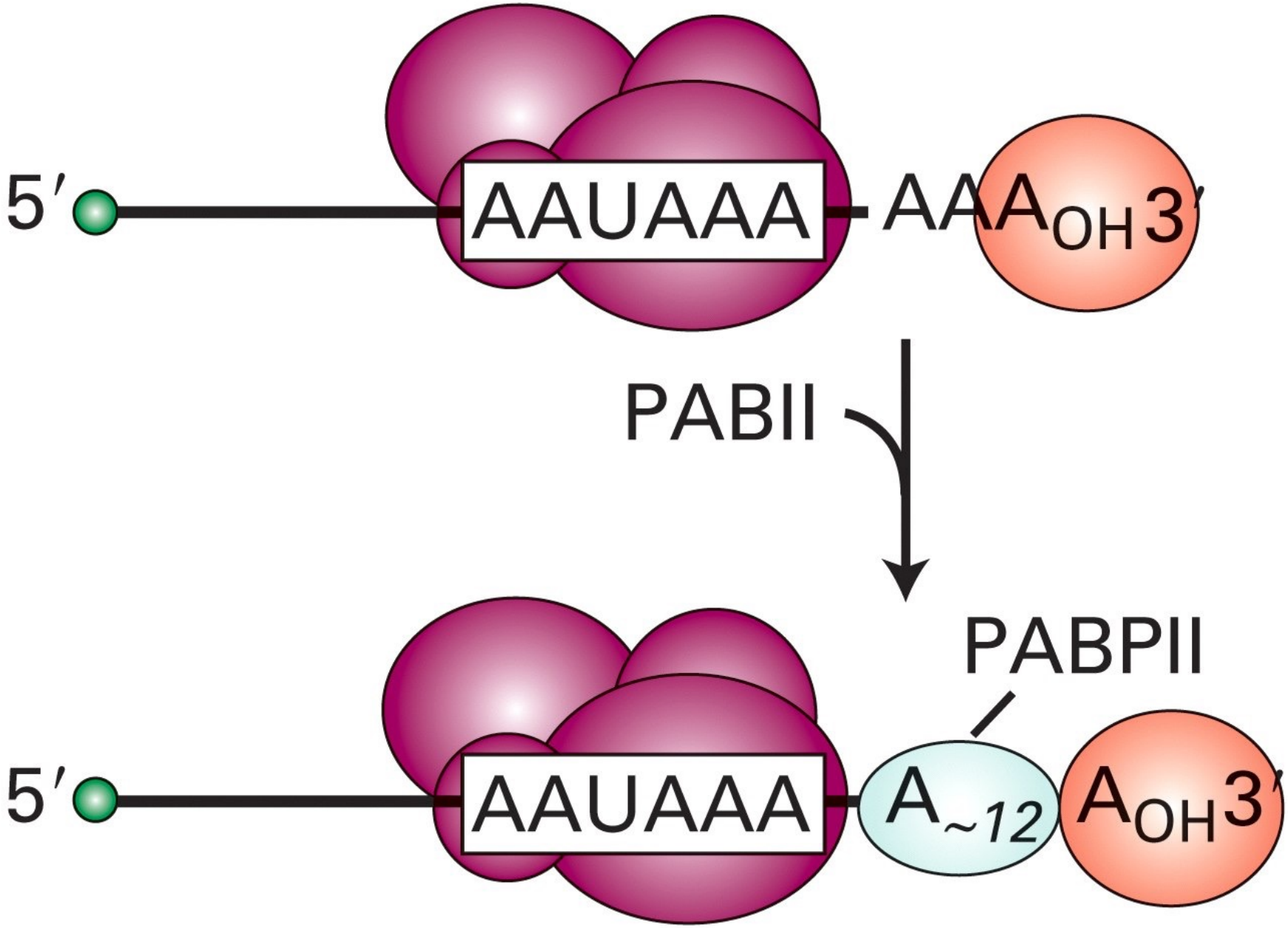
# 3′-end formation: RNA Processing

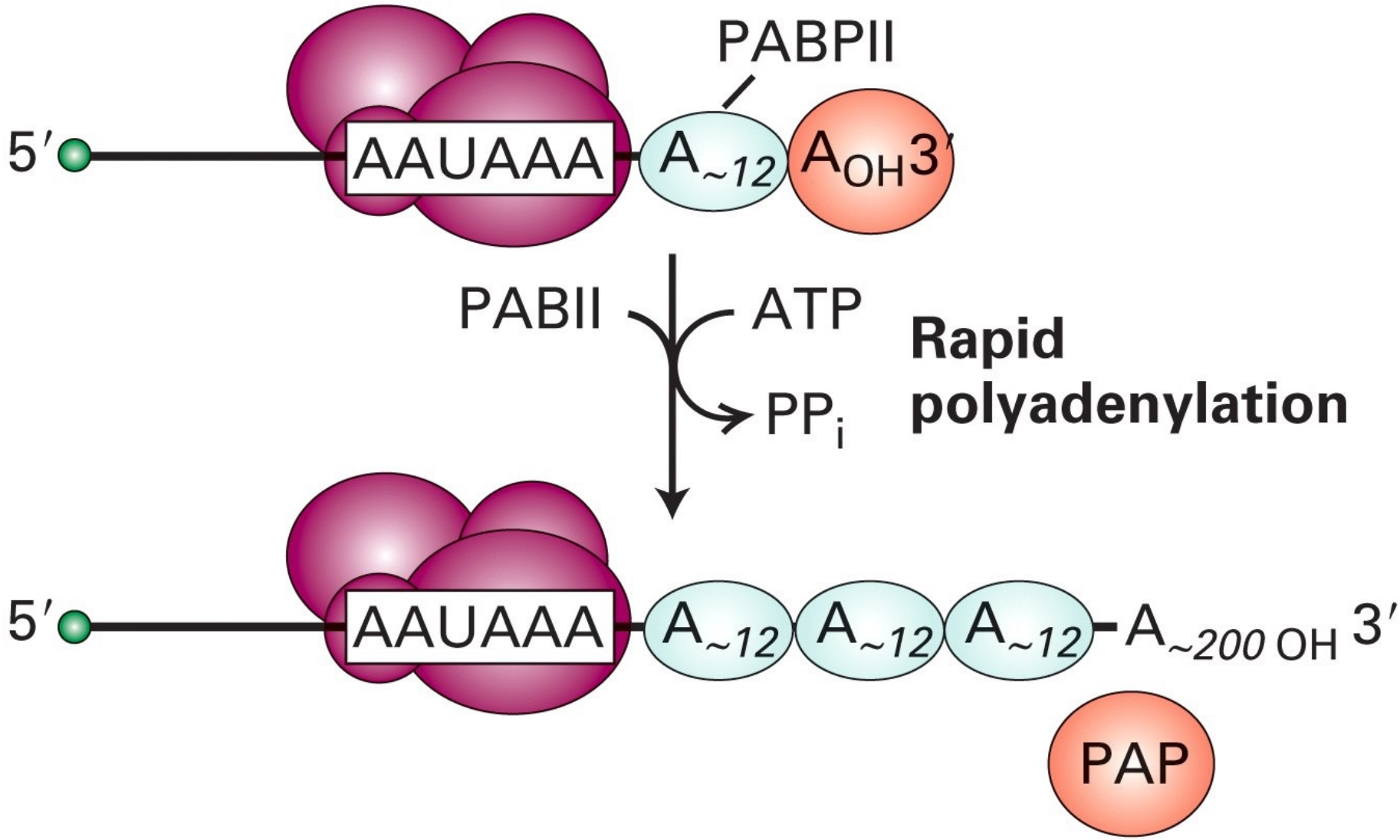



CPSF
CFI
5′
AAUAAA
CFII
G/U
CStF
3′
PAP
5′
AAUAAA
PAP
G/U
3′

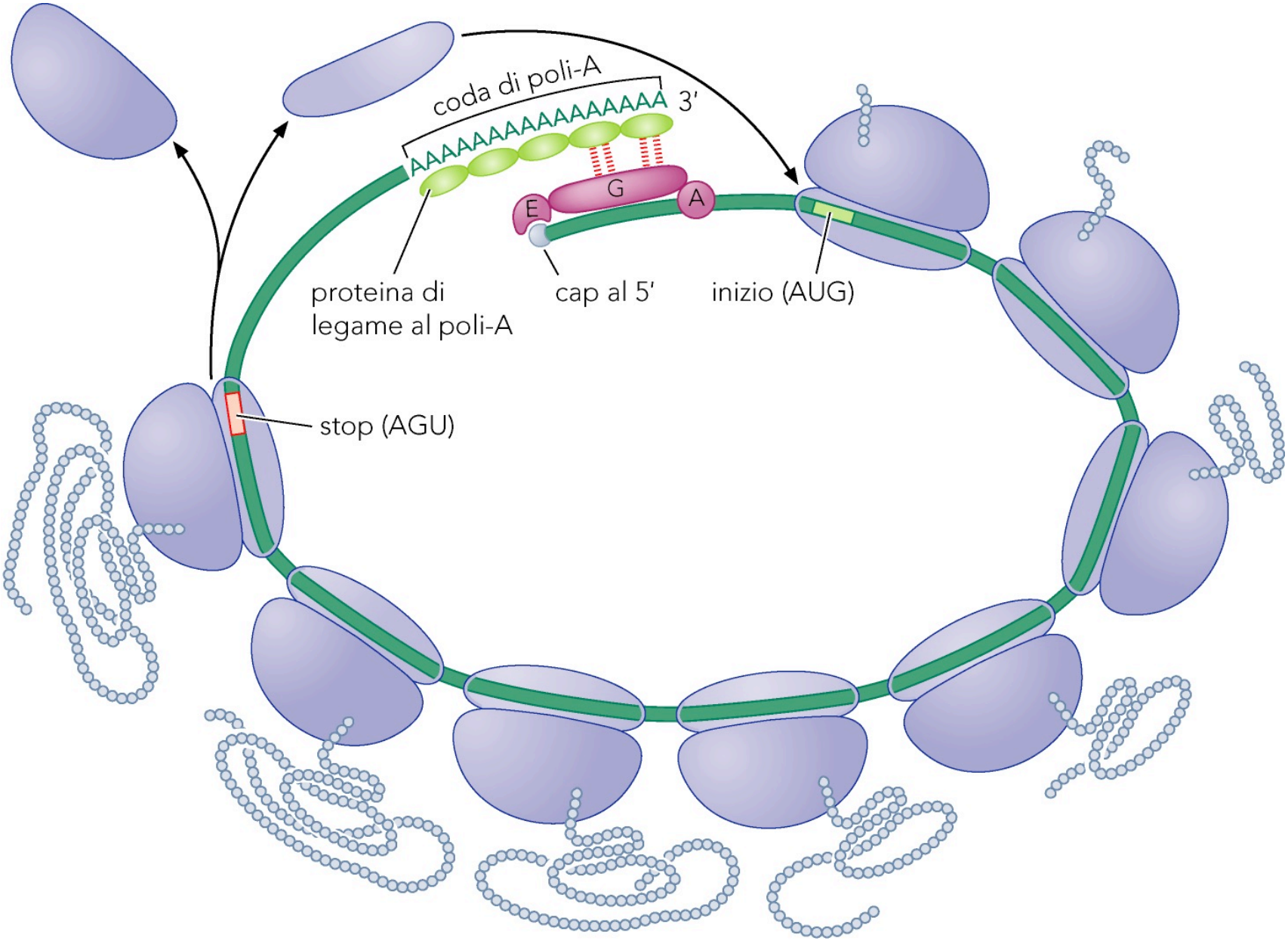

5′
AAUAAA
PAP
G/U
3′
Cleavage
5′
AAUAAA
OH p
G/U
3′


5′
AAUAAA
OH p
G/U
3′
ATP
PPi
Slow
polyadenylation
p
G/U
+ CStF, CFI, CFII
degradation
5′
AAUAAA
AAAOH3′
p
G/U
+ CStF, CFI, CFII


5′
AAUAAA
AAA$_{OH}$3′
PABII
PABPII
5′
AAUAAA
A$_{\sim 12}$
A$_{OH}$3′

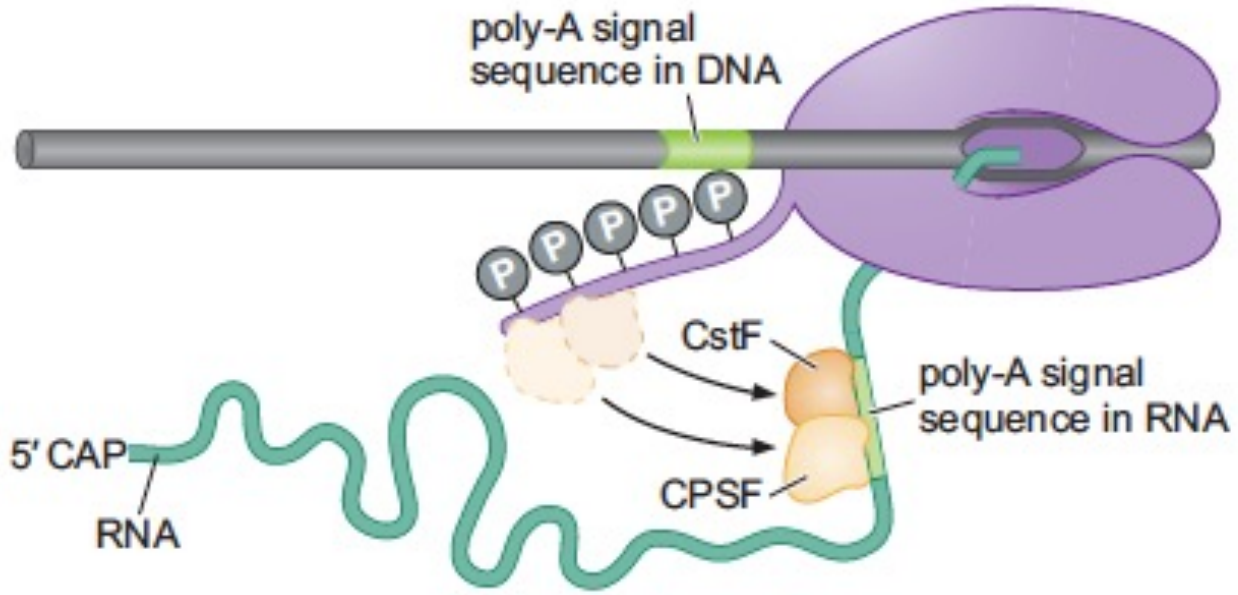

PABPII
5′
AAUAAA
A~12
AOH3′
PABII
ATP
PPi
Rapid
polyadenylation
5′
AAUAAA
A~12
A~12
A~12
A~200 OH 3′
PAP

# CAP and polyA tail promote mRNA translation

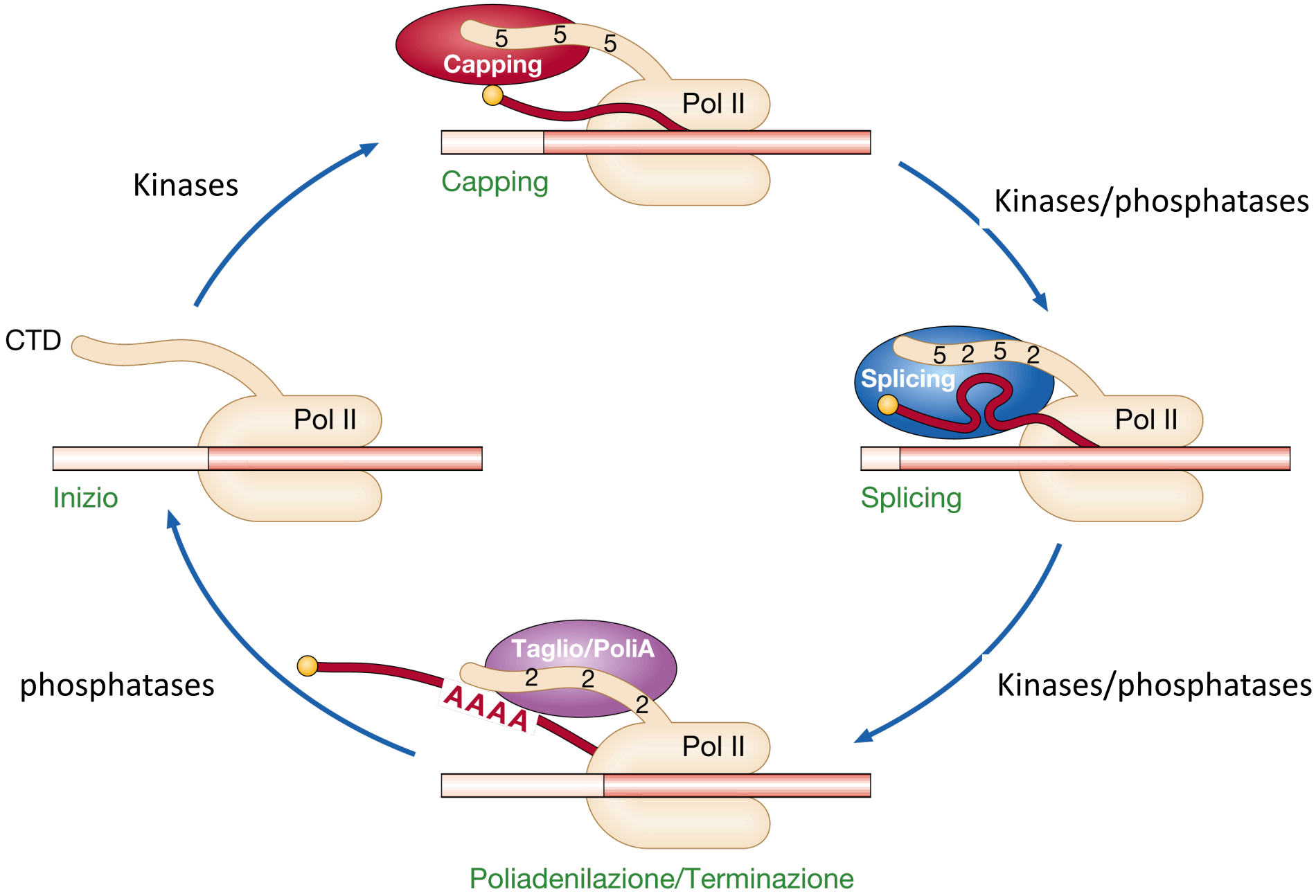


# 3′-end formation occurs during transcritpion

- TFIID associates with CPSF in the preinitiation complex (PIC)
- After transcription initiation CPSF dissociates from the PIC and associates with the CTD of elongating RNA Pol II  together with CstF



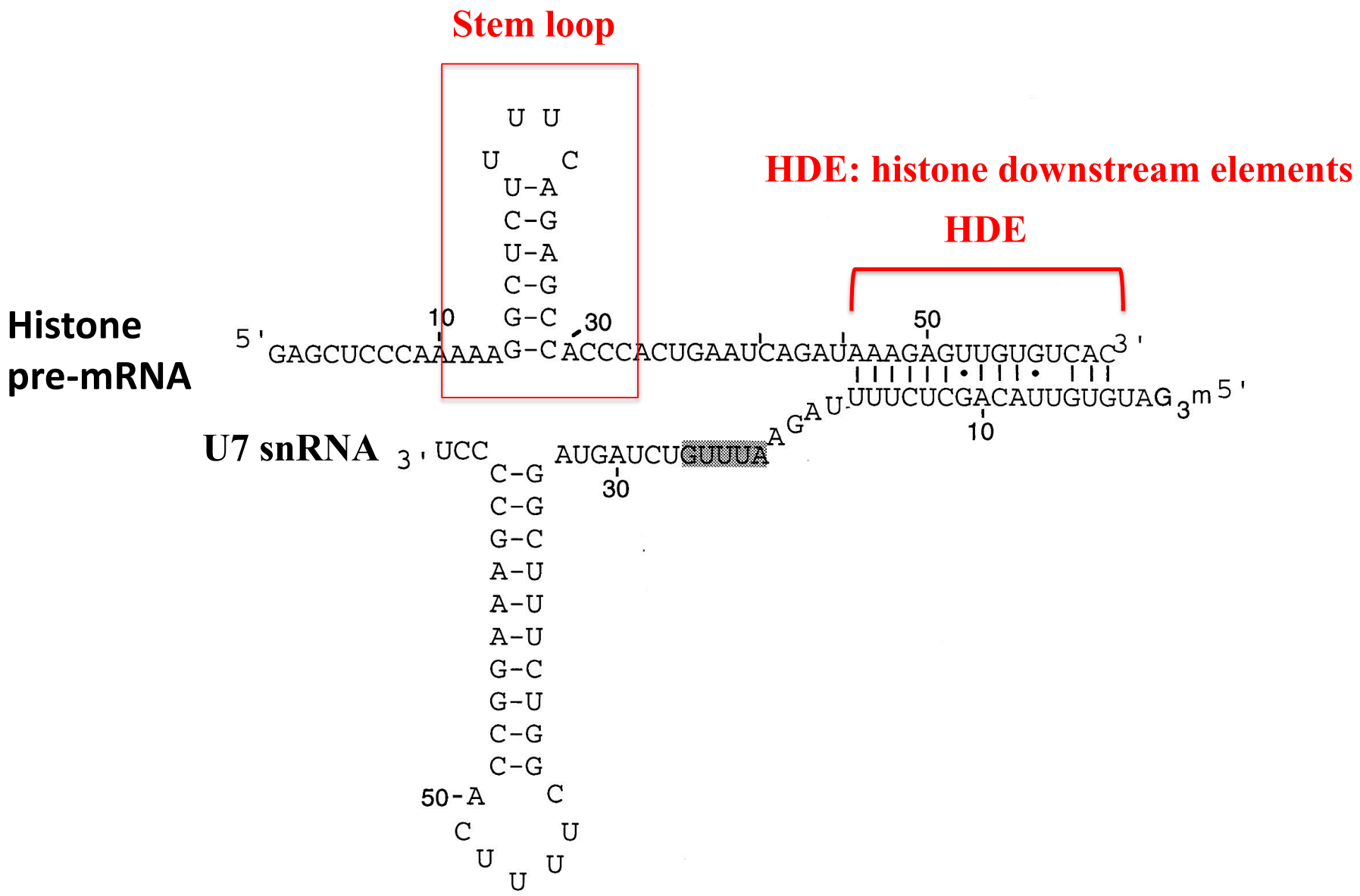
# capping, splicing and polyadenylation processes are tightly coupled with to transcription



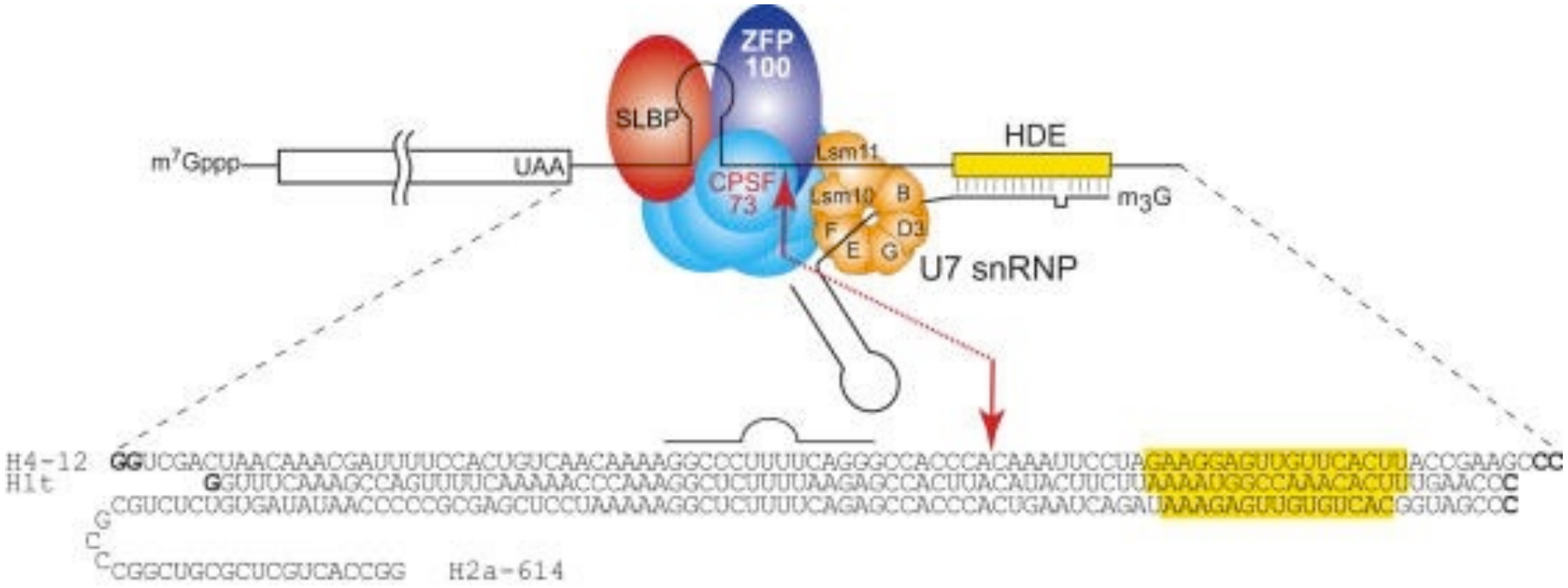
# Characteristics of Histone Messenger RNA

- Lacks introns: No splicing
- Lacks a polyA tail: cleavage not followed by polyadenylation
- Export to the cytoplasm is independent of the polyA tail
- Degradation occurs via uridylation
- Regulation of expression is dependent on the cell cycle

# The maturation of histone mRNA requires cis elements and trans factors, including U7 snRNA

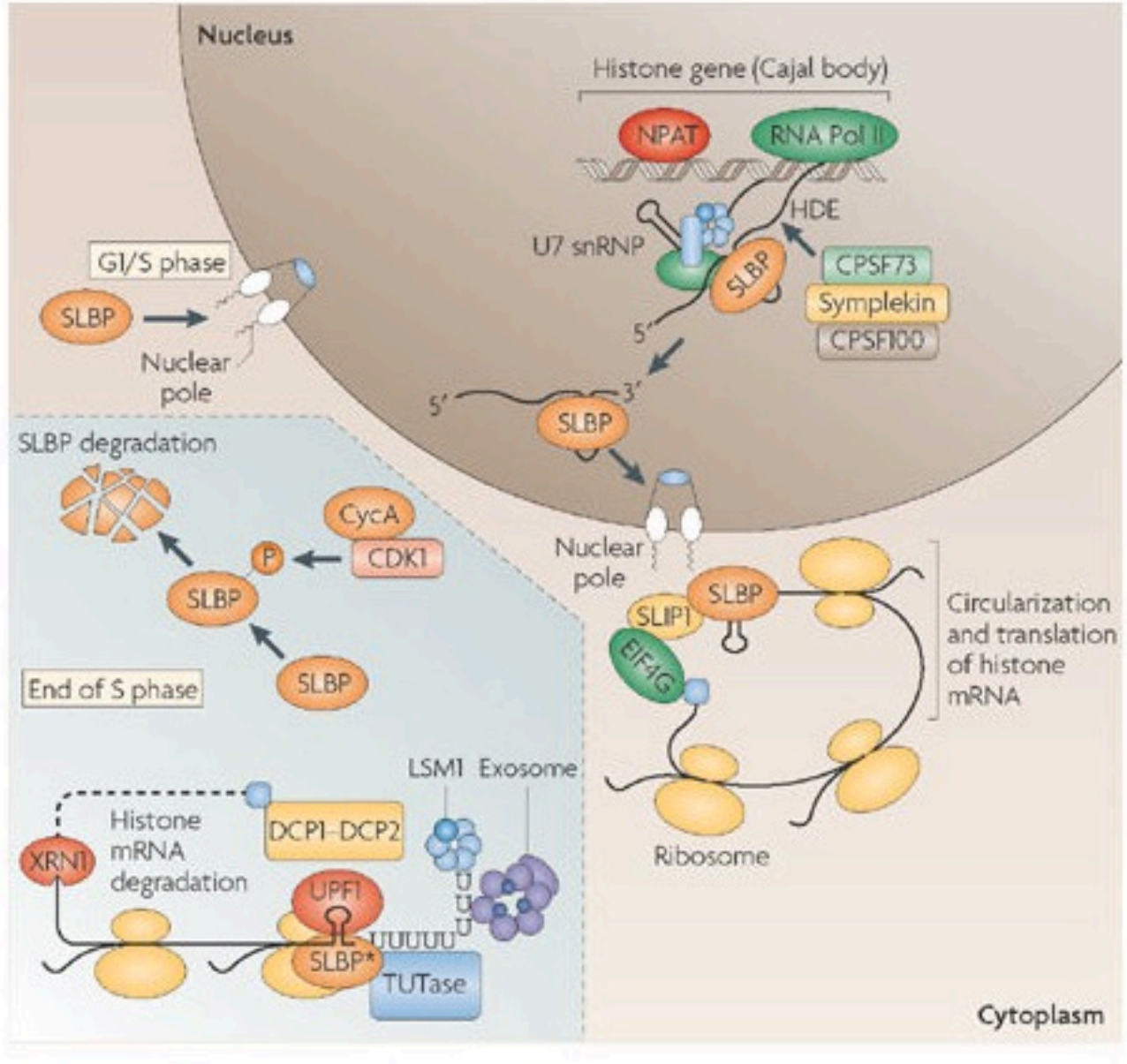


# Histone mRNA processing is polyadenylation independent




The U7 snRNP and stem–loop binding protein (SLBP), and the cleavage complex are responsible for the cleavage of the pre-mRNA from the DNA template, forming the mature histone mRNA

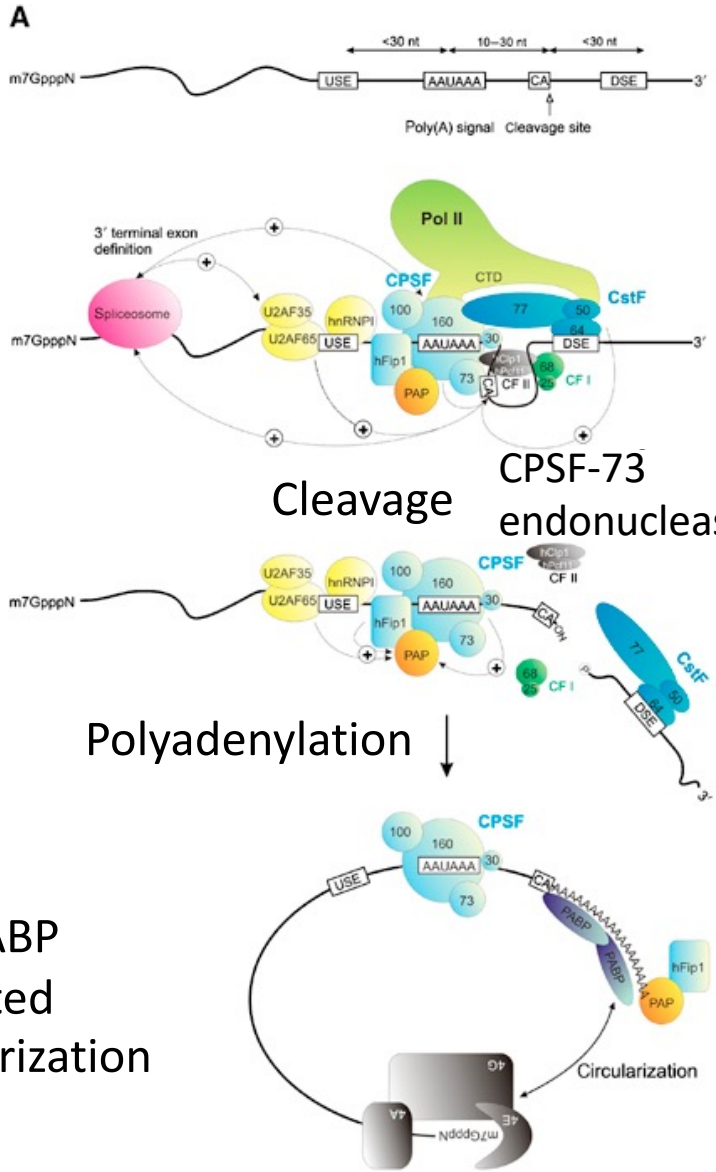
# Histone mRNAs: life without a poly(A) tail




SLBP remains bound to the histone mRNA as it goes to the cytoplasm, where histone mRNA is circularized through a complex of proteins mediating translation of histone mRNA. At the end of S phase, a short U tail is added to histone mRNA in the cytoplasm. The LSM1–7 ring binds the oligo(U) to cooperate in the recruitment of the decapping complex and the exosome to degrade the mRNA. In addition, cyclin A (CycA)–CDK1 (cyclin-dependent kinase 1) phosphorylates SLBP to trigger its degradation at the end of S phase, preventing further histone mRNA synthesis.

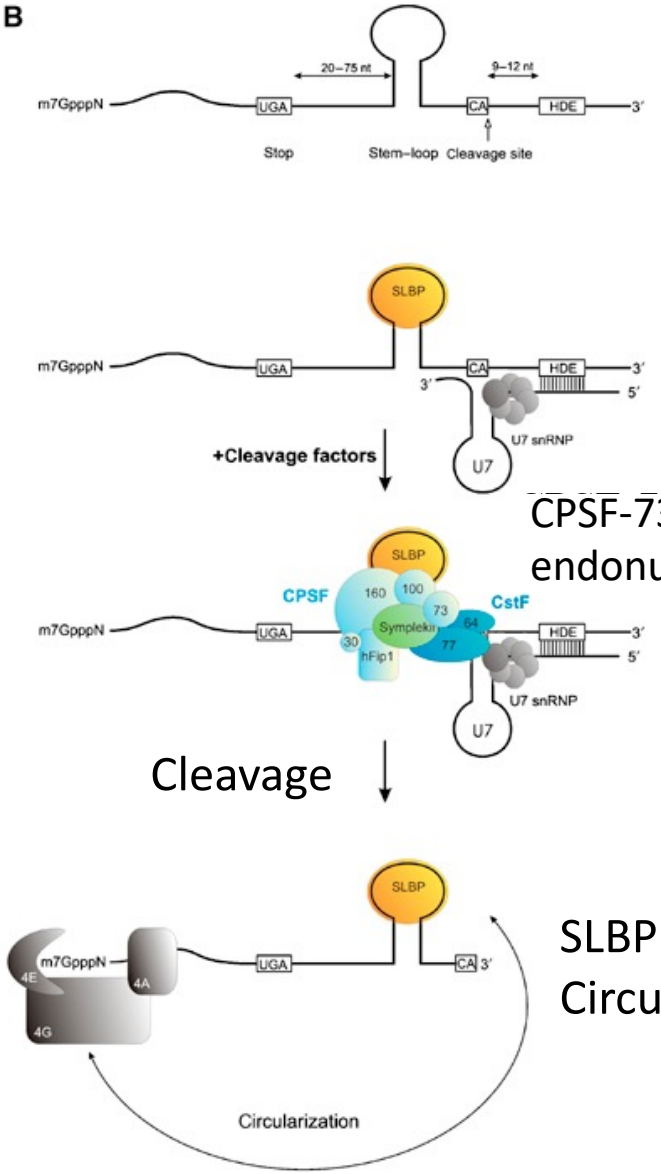
# Processing and translation of mRNAs

## Poly(A)$^+$ mRNAs




CAP-PABP mediated Circularization

## Histone mRNAs

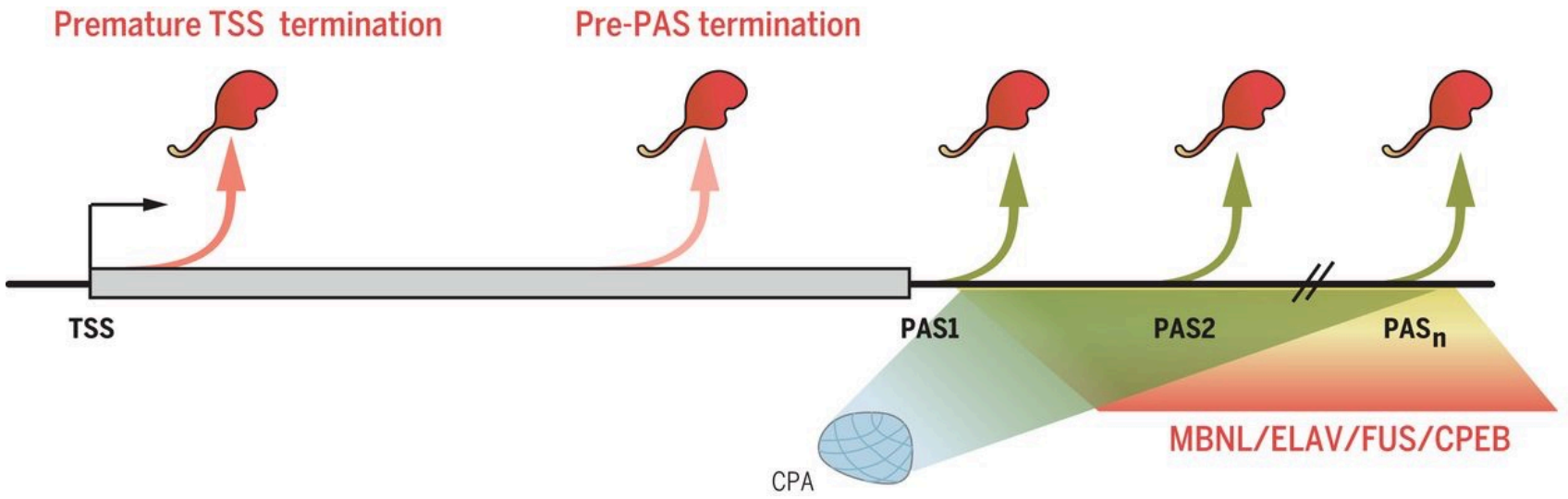



SLBP mediated Circularization

# RNA Pol II alternative 3′-end formation.

Termination at gene 3′-ends is also subject to intense regulation. Many mRNAs possess variable lengths of 3′-untranslated sequence defined by the selective usage of different PASs. Because mRNA 3′UTRs define mRNA cytoplasmic functions, the use of alternative poly(A) sites (APA) can constitute a key regulatory process in gene expression.

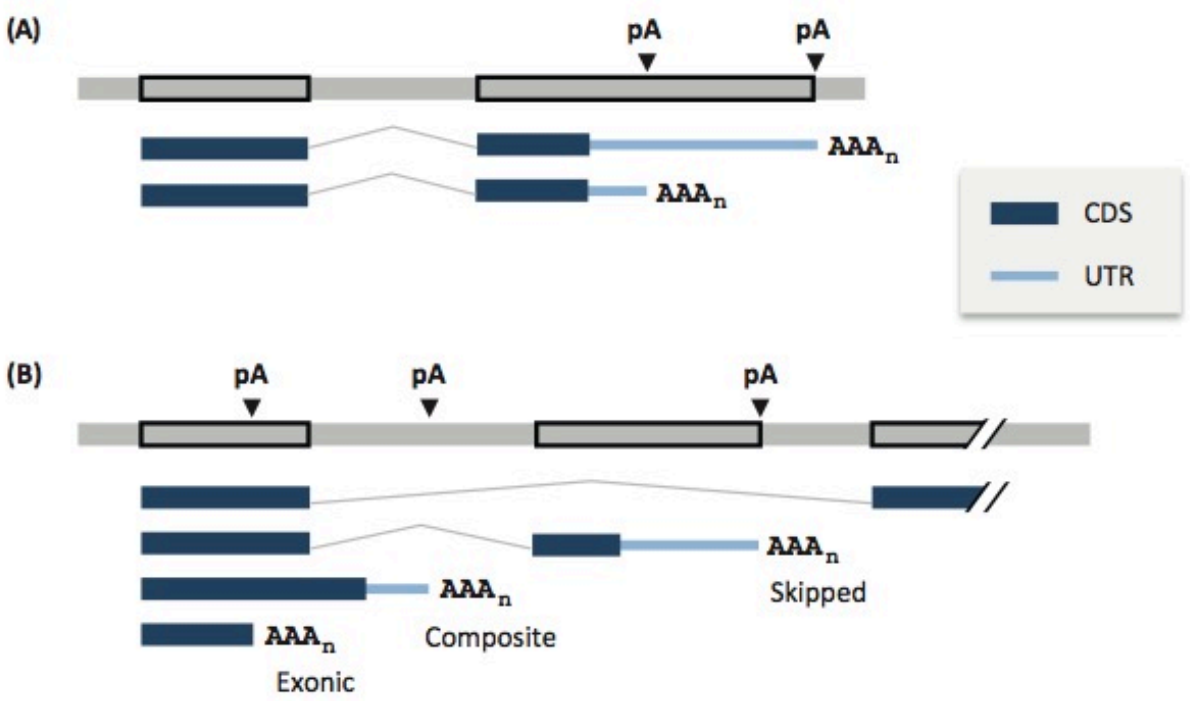
Polymerase speed and in trans elements can affect APA choice




Science AAAS

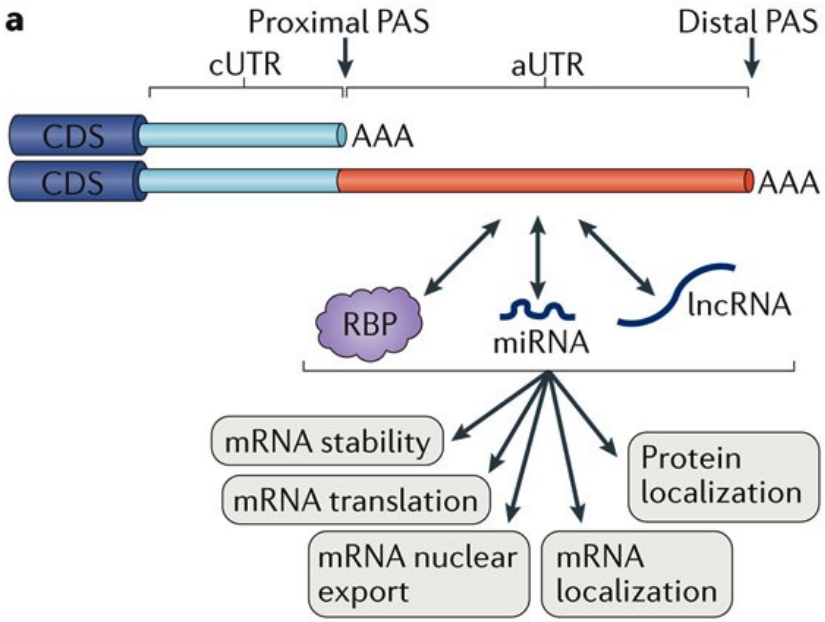
# Alternative polyadenylation

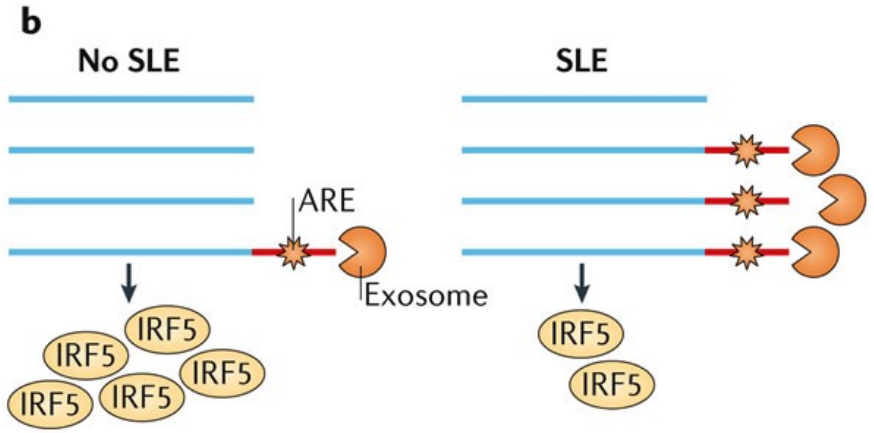
- 70–80% of mammalian genes have been reported to display alternative cleavage and polyadenylation site (APA) ,
- alternative pAs in the 3′-most exon (A) typically leads to variable 3′ UTRs, whereas pAs in upstream introns and exons (B) cause both coding sequence (CDS) and 3′ UTR changes,
- APA is dynamic under different biological conditions (cell growth and development, cancer)



# Alternative polyadenylation

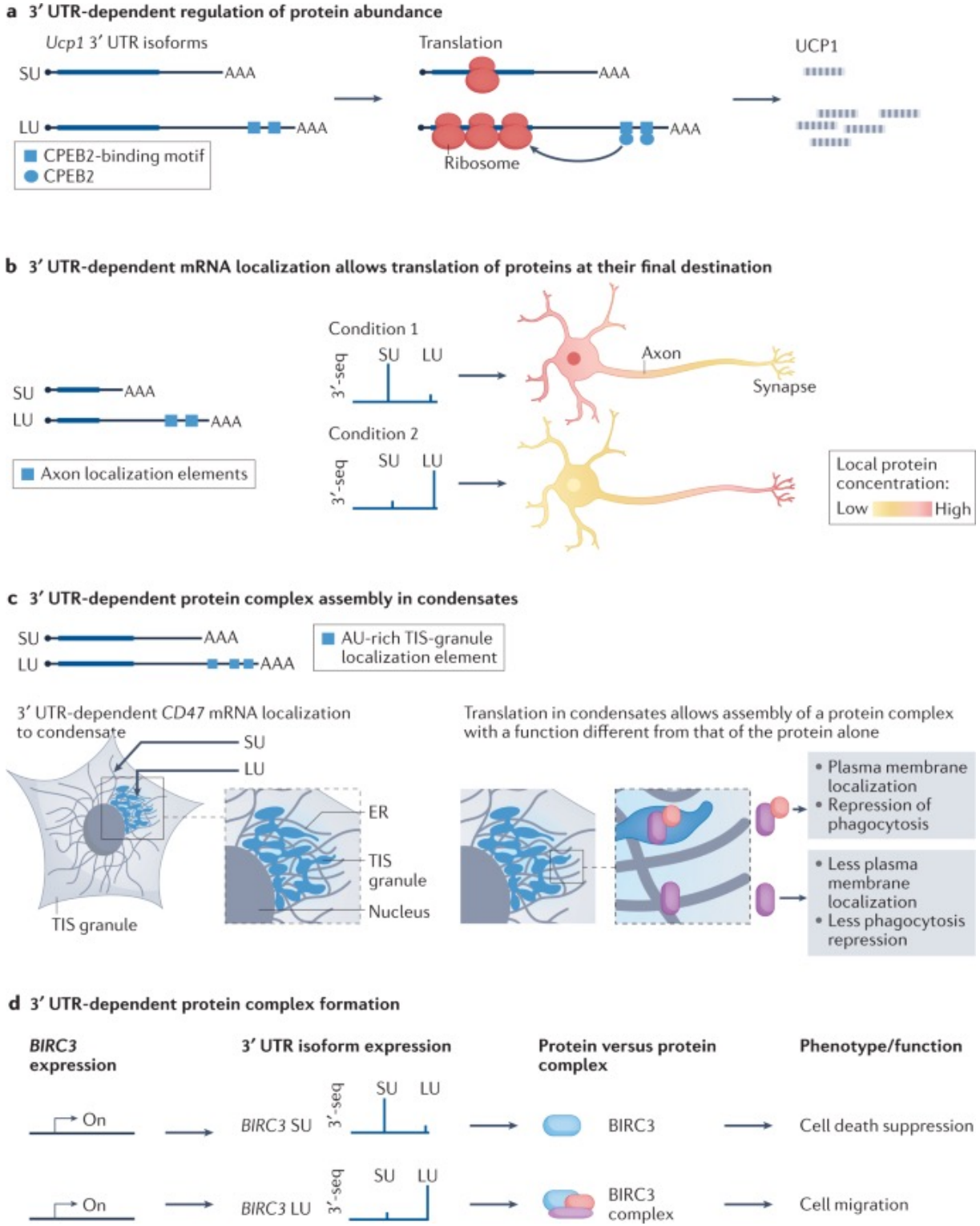
Most alternative polyadenylation (APA) sites are located in 3′ UTRs. As 3′ UTRs contain *cis* elements that are involved in various aspects of mRNA metabolism, 3′ UTR-APA can considerably affect post-transcriptional gene regulation in various ways, including through the modulation of mRNA stability, translation, nuclear export and cellular localization, and even through effects on the localization of the encoded protein.




In patients with systemic lupus erythematosus (SLE), a single nucleotide polymorphism reducing the use of the proximal PAS leads to the production of long isoforms at the expense of short isoforms, which results in reduced IFN-regulatory factor 5 (IRF5) levels.

Nature Reviews | Molecular Cell Biology


a 3′ UTR-dependent regulation of protein abundance
Ucp1 3′ UTR isoforms
SU
AAA
LU
AAA
CPEB2-binding motif
CPEB2
Translation
AAA
AAA
Ribosome
UCP1
b 3′ UTR-dependent mRNA localization allows translation of proteins at their final destination
SU
AAA
LU
AAA
Axon localization elements
Condition 1
3′-seq
SU
LU
Axon
Synapse
Condition 2
3′-seq
SU
LU
Local protein concentration:
Low
High
c 3′ UTR-dependent protein complex assembly in condensates
SU
AAA
LU
AAA
AU-rich TIS-granule localization element
3′ UTR-dependent CD47 mRNA localization to condensate
SU
LU
ER
TIS granule
Nucleus
TIS granule
Translation in condensates allows assembly of a protein complex with a function different from that of the protein alone
• Plasma membrane localization
• Repression of phagocytosis
• Less plasma membrane localization
• Less phagocytosis repression
d 3′ UTR-dependent protein complex formation
BIRC3 expression
3′ UTR isoform expression
Protein versus protein complex
Phenotype/function
On
BIRC3 SU
3′-seq
SU
LU
BIRC3
Cell death suppression
On
BIRC3 LU
3′-seq
SU
LU
BIRC3 complex
Cell migration

# Regulation of Alternative cleavage and polyadenylation

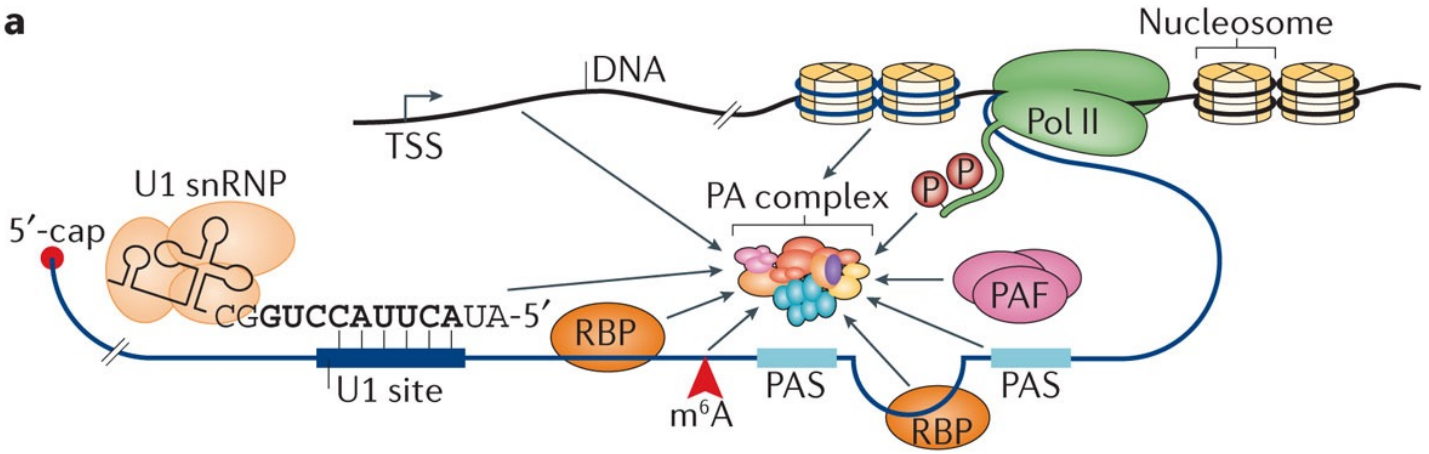
1. U1 can exert a negative role in pA usage

2. RNA binding proteins may inhibits pA usage by occluding the binding of core C/P factors, and some enhance pA usage by recruiting core factors.

3. Transcription activity impact pA choice by :

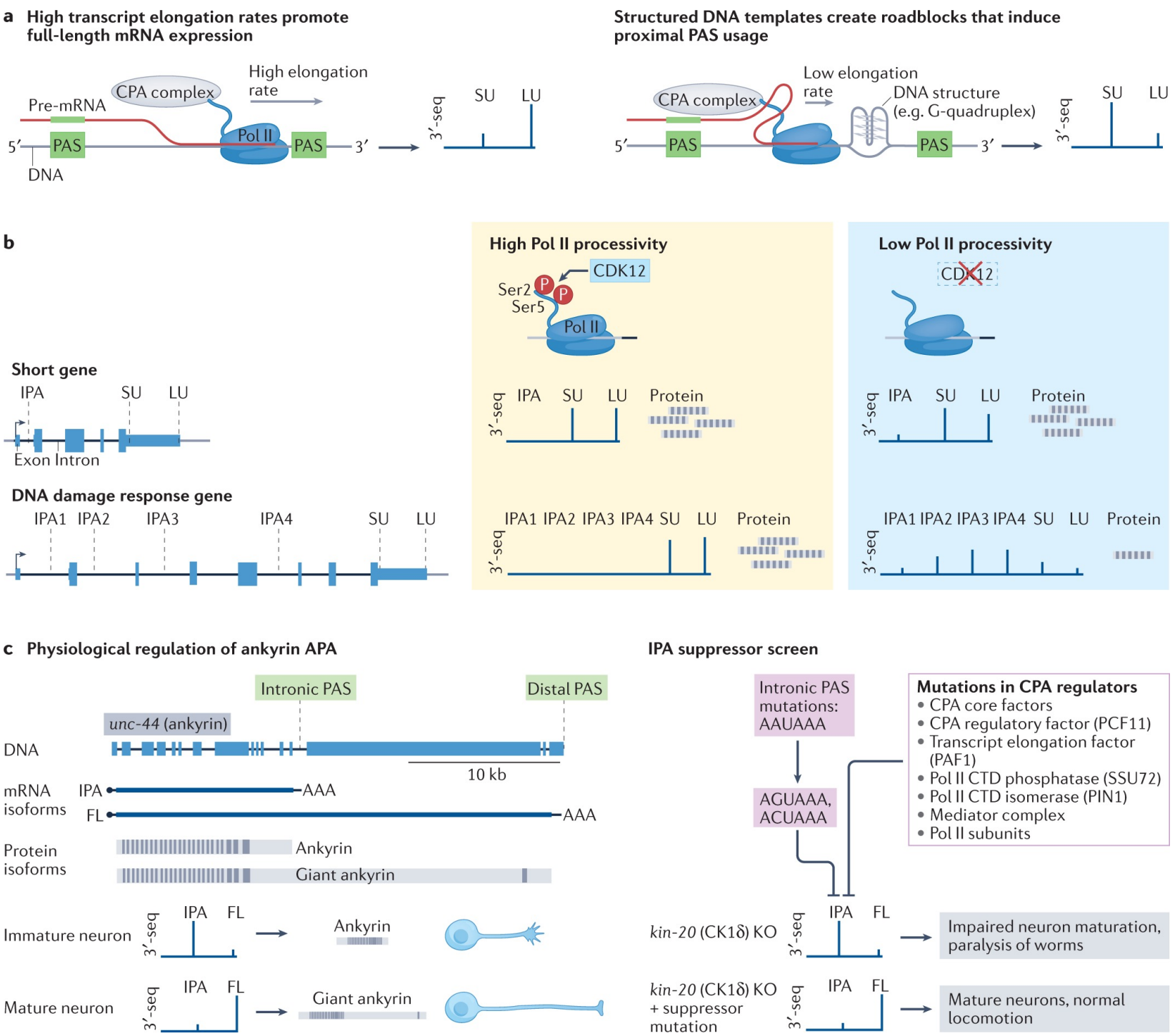
   - **cis elements** that cause pausing of RNA polymerase (RNAP) II, such as G-rich elements, facilitate pA usage,

   - **factor recruitment** at the promoter can influence downstream pA usage.




Nature Reviews | Molecular Cell Biology

# Co-transcriptional regulators of alternative polyadenylation

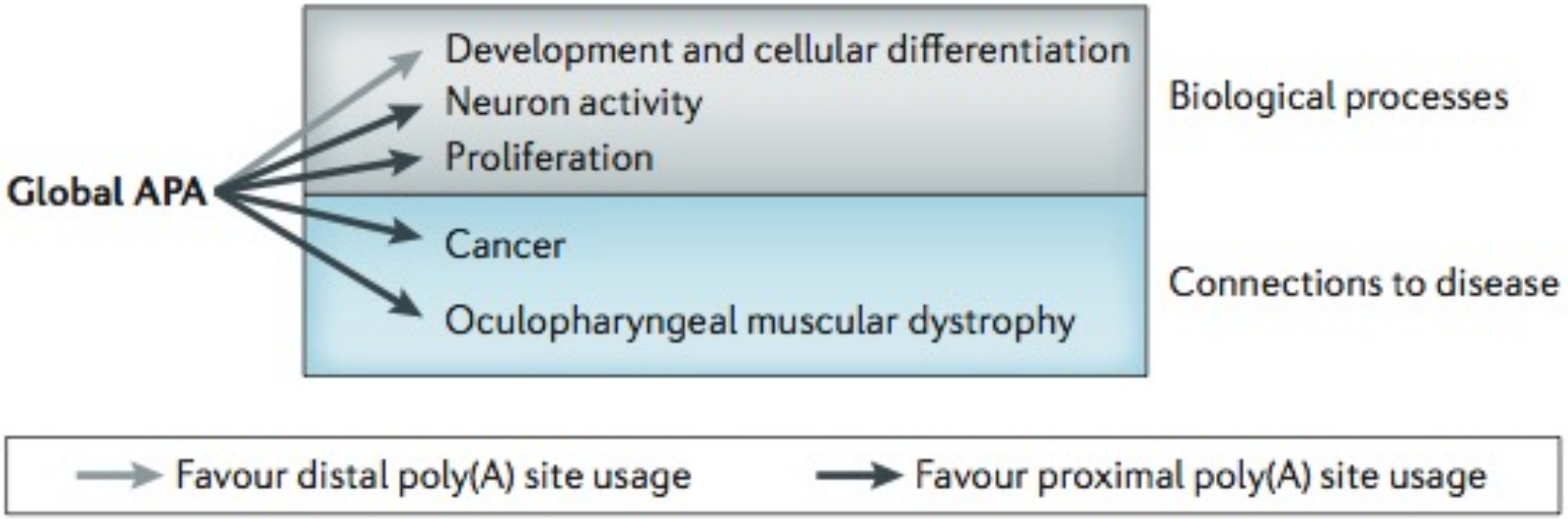


# Mutations in poly(A) *cis*-elements that cause or contribute to human diseases

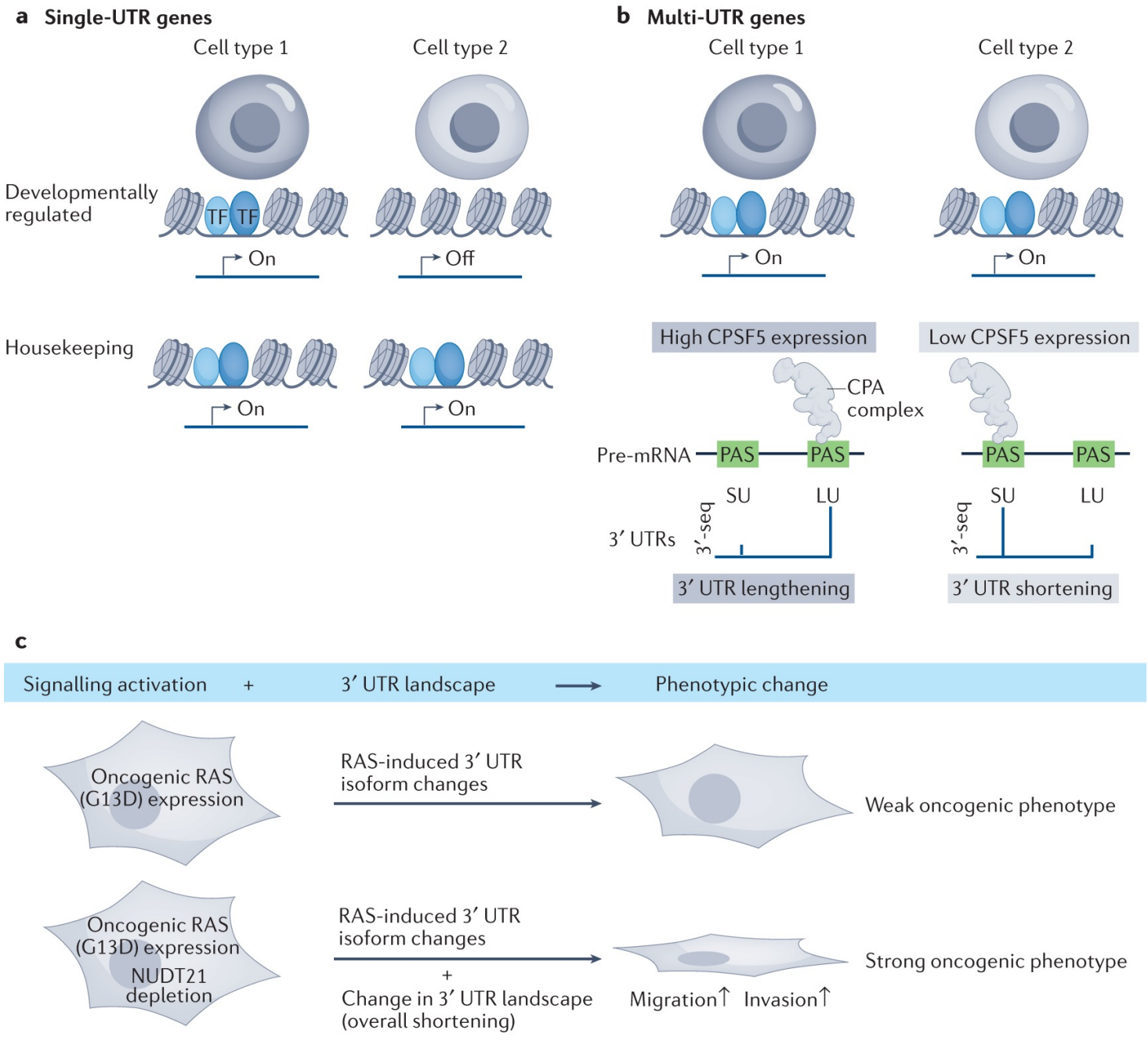
| Disease | Disease description | Poly(A) site mutation | Affected gene | Comments | Refs |
|---|---|---|---|---|---|
| α-Thalassaemia | Thalassaemias are a group of common human genetic diseases that result from defects in haemoglobin production | AATAAA to AATAAG | *HBA2* | | 84 |
| β-Thalassaemia | | AATAAA to AACAAA | *HBB* | The mutation results in generation of an unstable transcript that is ~900 nt longer | 85 |
| Metachromatic leukodystrophy | A neurodegenerative disorder caused by null mutations in *ARSA* | AATAAC to AGTAAC | *ARSA* | Carriers of this mutation show reduced mRNA levels and enzyme activity but do not develop the disease symptoms; this condition is termed 'pseudodeficiency' | 86,87 |
| IPEX | A rare multifaceted and fatal disease | AATAAA to AATGAA | *FOXP3* | | 88 |
| Fabry's disease | A rare and severe X-linked lysosomal storage disease caused by mutations in *GLA* | An AA dinucleotide deletion within the poly(A) site | *GLA* | *GLA* is one of the unusual mammalian genes that lack a 3′UTR and has its PAS within the CDS. The mutation in the poly(A) site results in aberrant 3′ end formation and multiple unstable transcripts | 89 |

*ARSA*, arylsulfatase A; CDS, coding DNA sequence; *FOXP3*, forkhead box P3; *GLA*, alpha-galactosidase; *HBA2*, haemoglobin, alpha 2 (also known as alpha-2-globin); *HBB*, haemoglobin, beta (also known as beta-globin); IPEX, immune dysfunction, polyendocrinopathy, enteropathy, X linked; PAS, poly(A) signal; UTR, untranslated region. Only selected examples are discussed here. For a more thorough discussion, see REF. 64.

# Biological processes that have been linked with broad APA modulation



# Biological processes that have been linked with broad APA modulation



# Manipulation of alternative 3′ UTR expression.