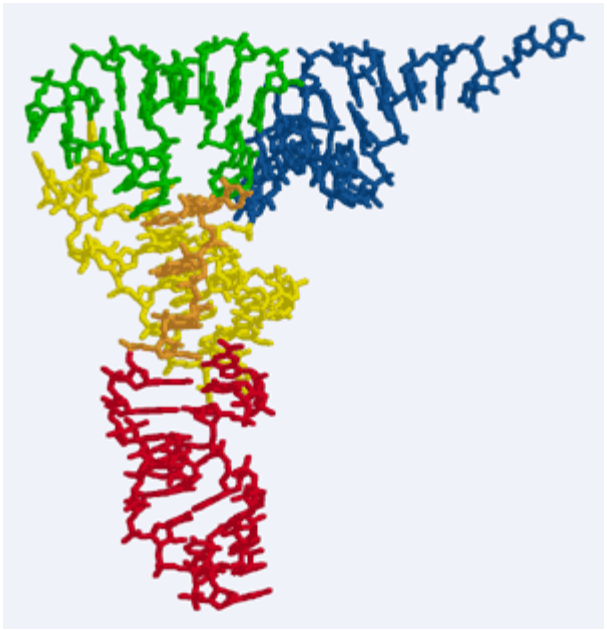
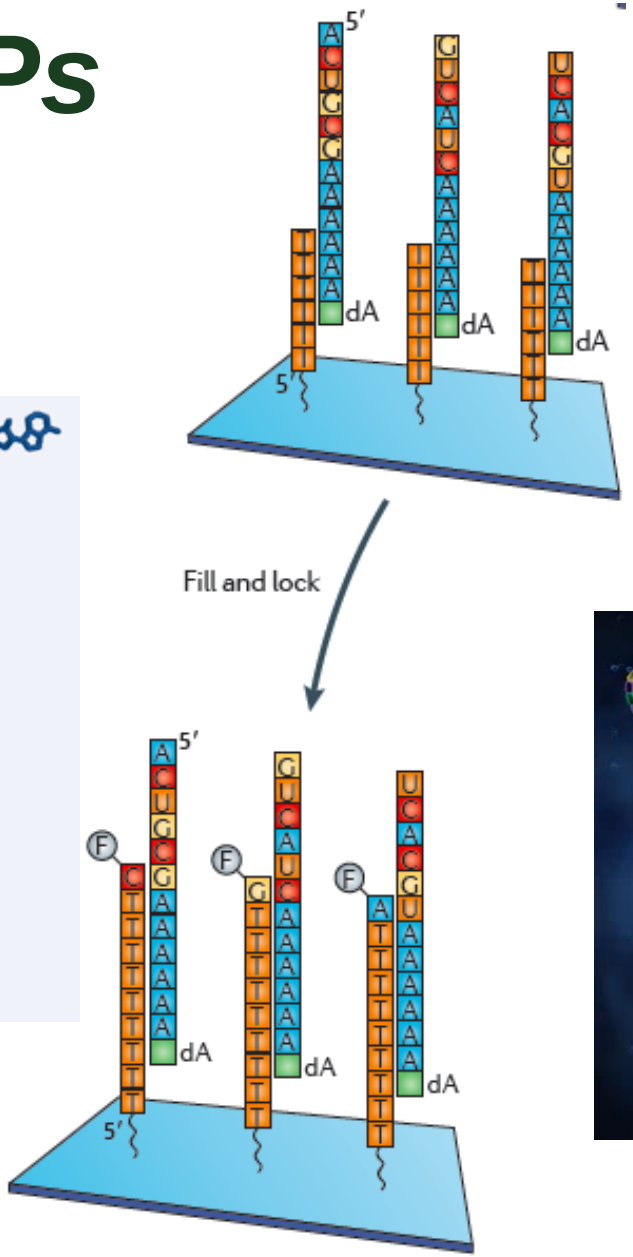
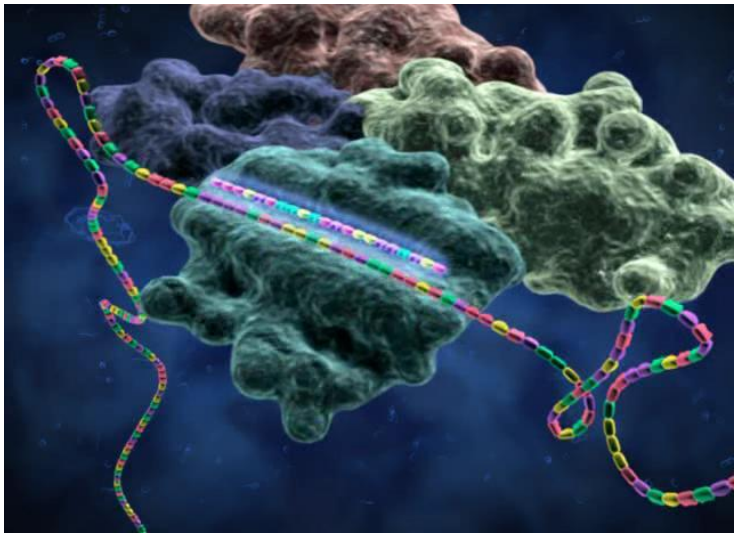
GLOBAL ANALYSES of RNAs and RNPs

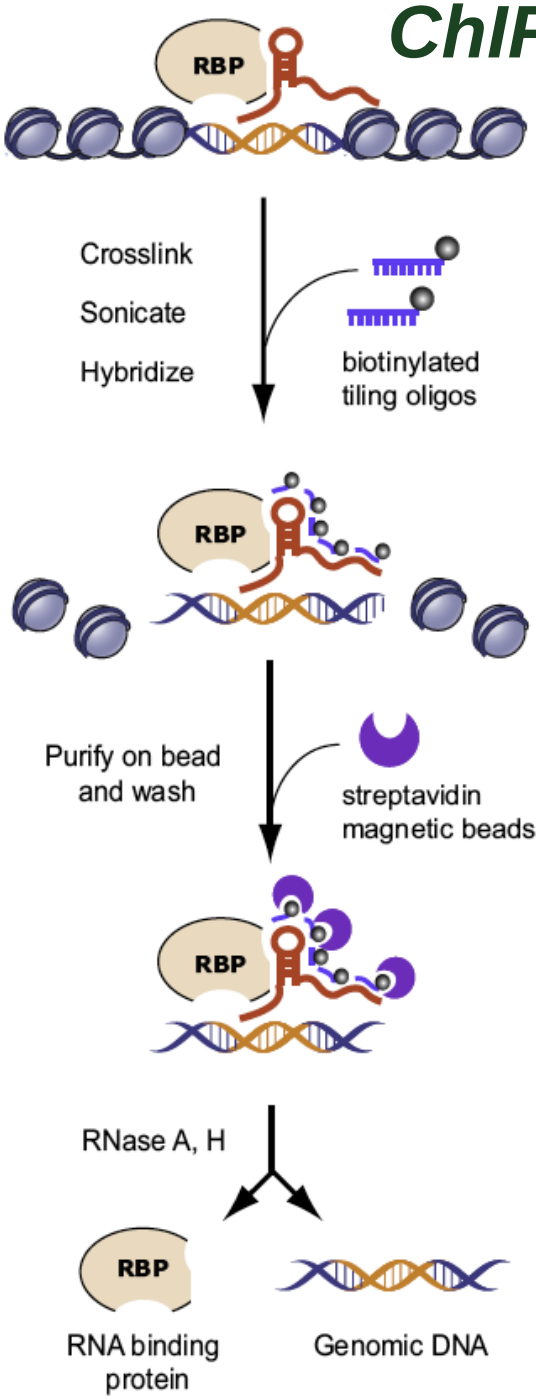






Chromatin Isolation by RNA Purification

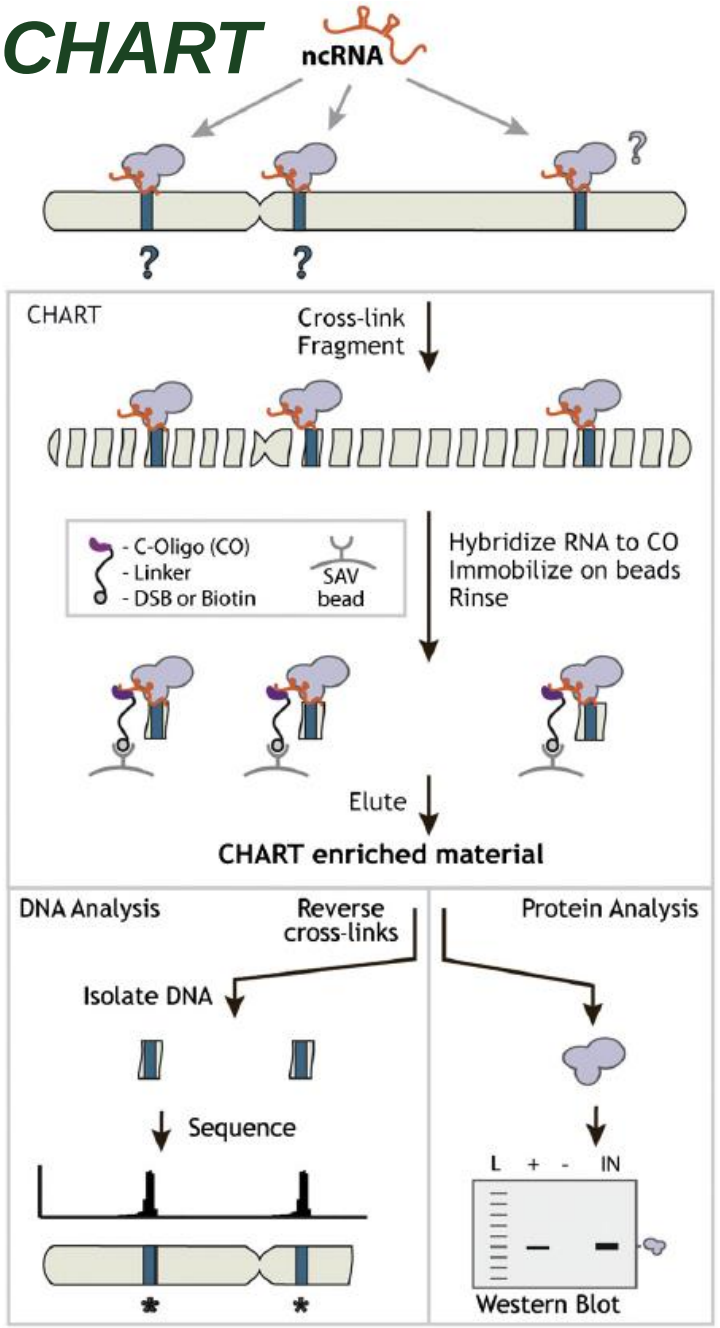
ChIRP



lncRNA
proteins
DNA

Capture Hybridization Analysis of RNA Targets

CHART

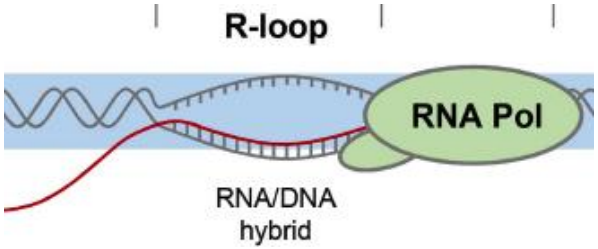


Chu et al., Mol. Cell, 2011; Simon et al., PNAS '11

Detection of RNA-DNA hybrids

R-loop

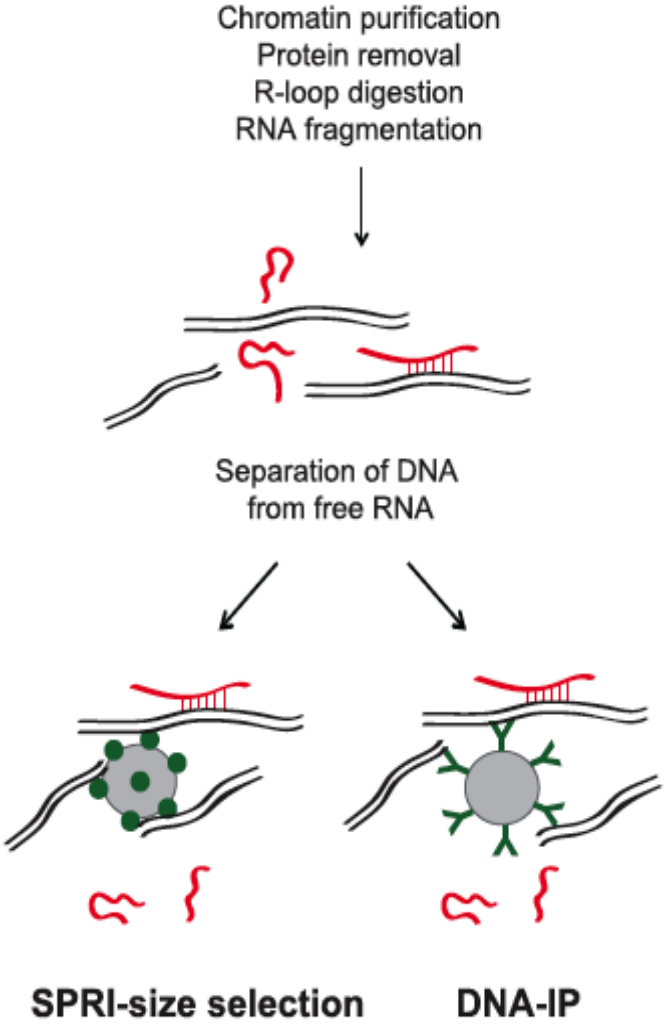
IP with specific RNA/DNA S9.6 Ab followed by RNAseq



Enrichment of DNA-associated RNA

- SPRI (Solid Phase Reversible Immobilization) - based paramagnetic bead size selection
- DNA-IP using anti-DNA Ab

RNA/DNA tripplex

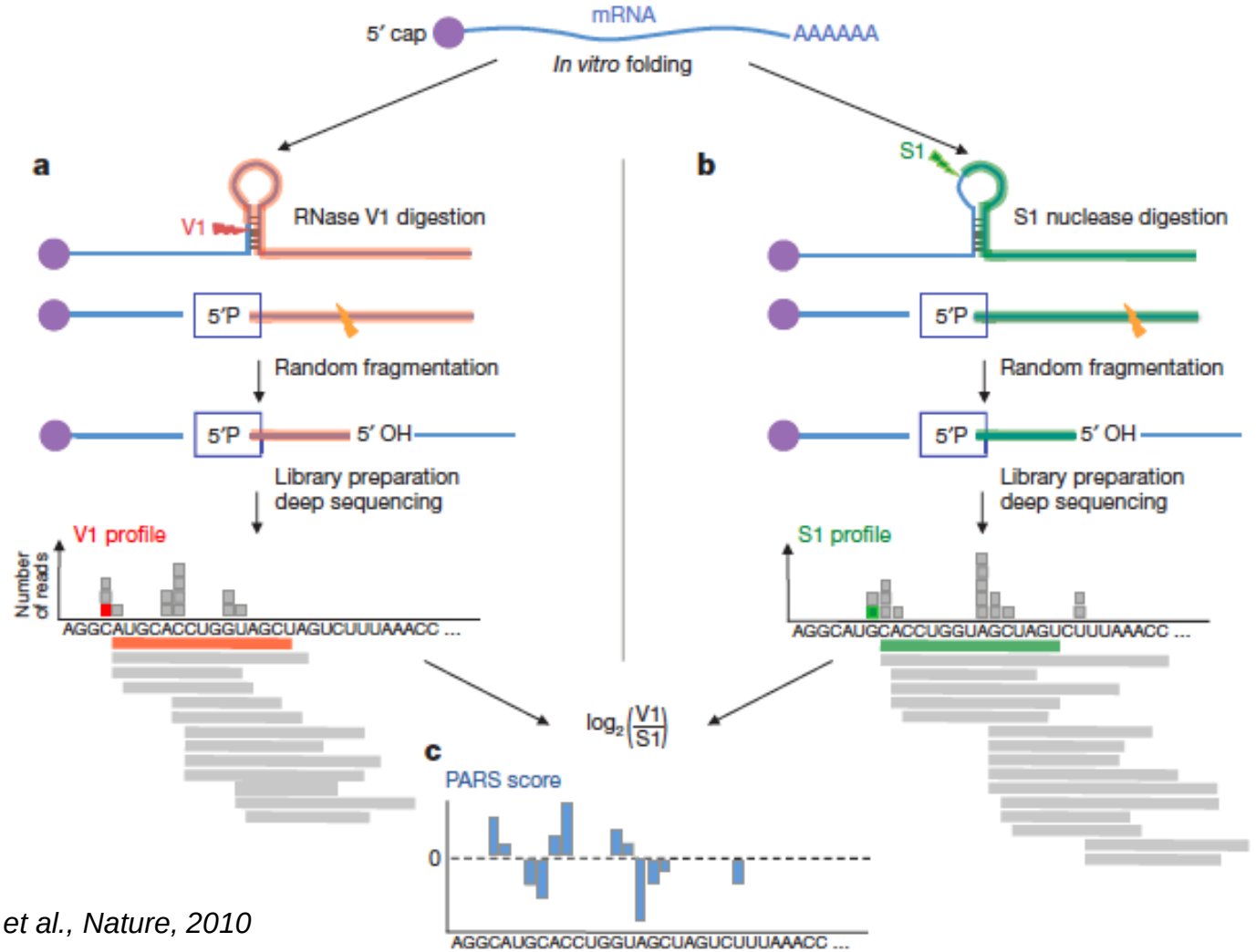


Cetin at al, RNA, 2019

PARS: Parallel Analysis of RNA Structure

measuring RNA structural properties by deep sequencing

- PARS confirmed for known RNA structures
- used to establish structures of > 3000 yeast transcripts
- unexpected conclusion: coding mRNA regions are more structured than UTRs!



Kertesz et al., Nature, 2010

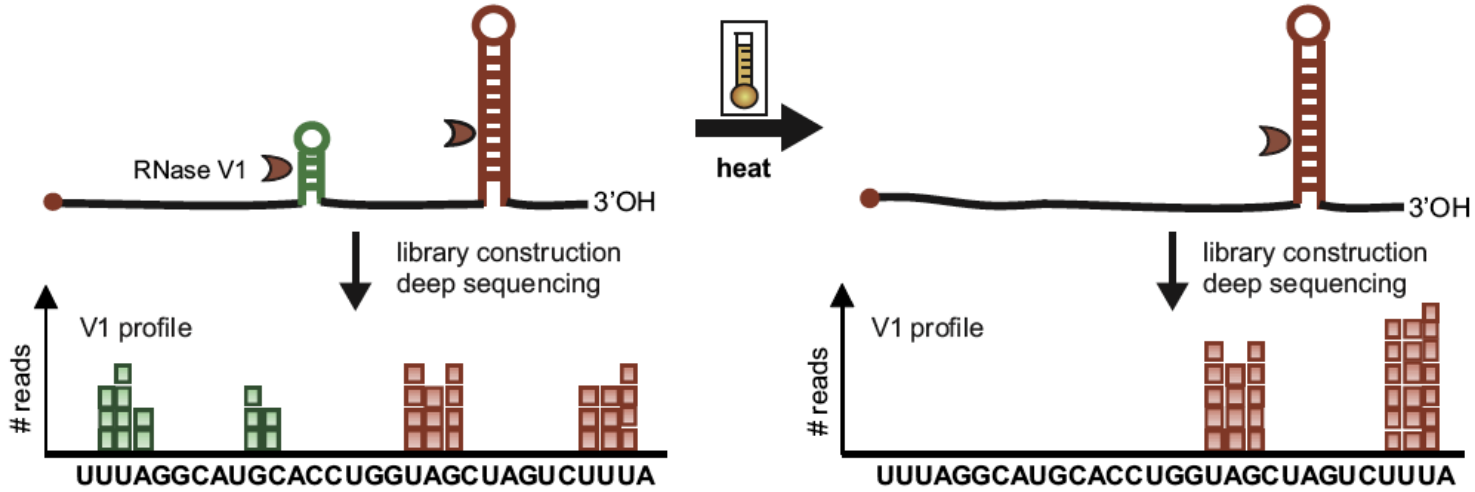
PARS: Parallel Analysis of RNA Structure

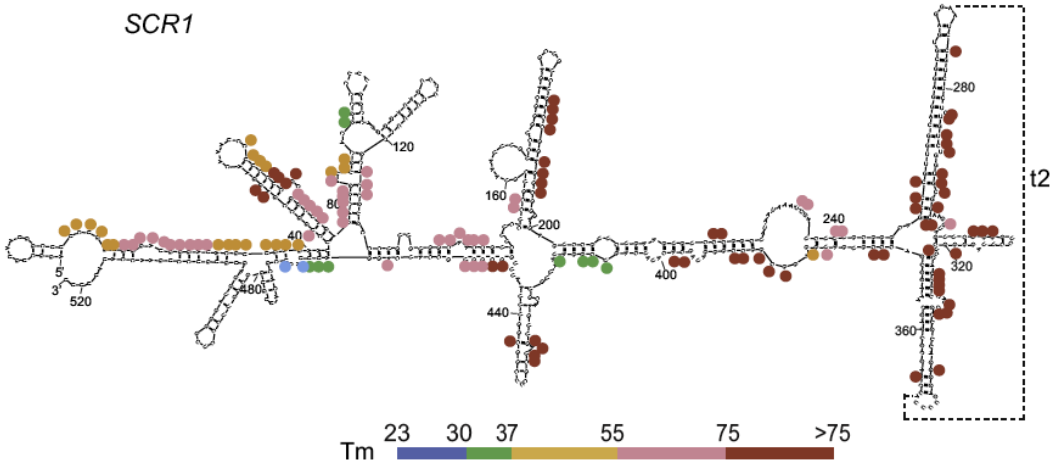
Molecular Cell
Article

Genome-wide Measurement of RNA Folding Energies

Molecular Cell *48*, 169–181, October 26, 2012

Yue Wan,[1] Kun Qu,[1,8] Zhengqing Ouyang,[1,2,8] Michael Kertesz,[3] Jun Li,[4] Robert Tibshirani,[4] Debora L. Makino,[5] Robert C. Nutter,[6] Eran Segal,[7,*] and Howard Y. Chang[1,*]





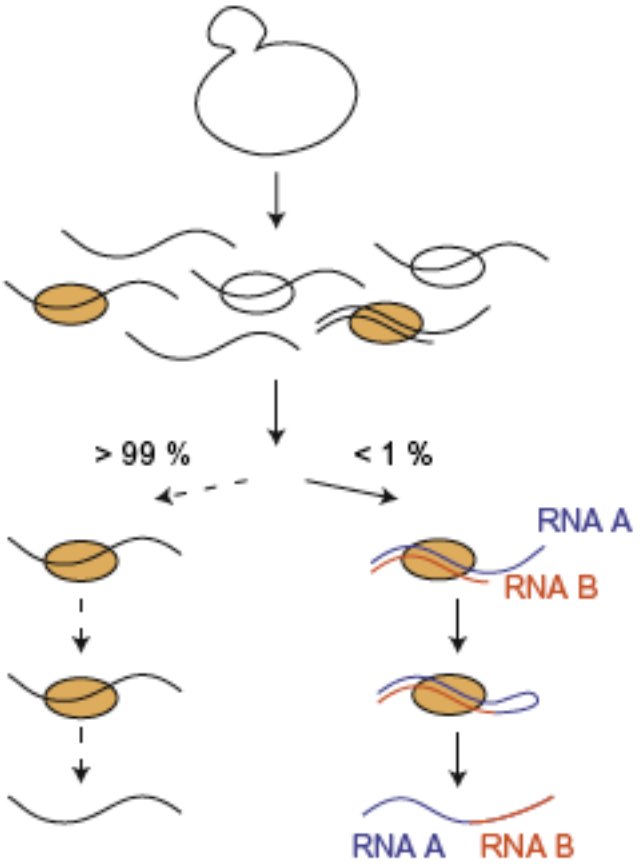
Regulatory Impact of RNA Secondary Structure across the *Arabidopsis* Transcriptome [W][OA]

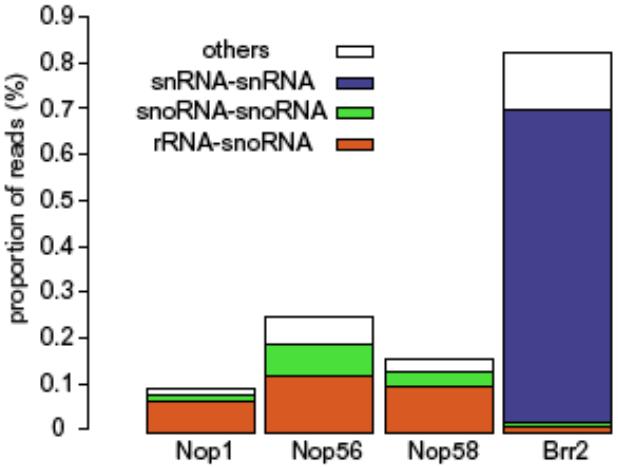
The Plant Cell 2012

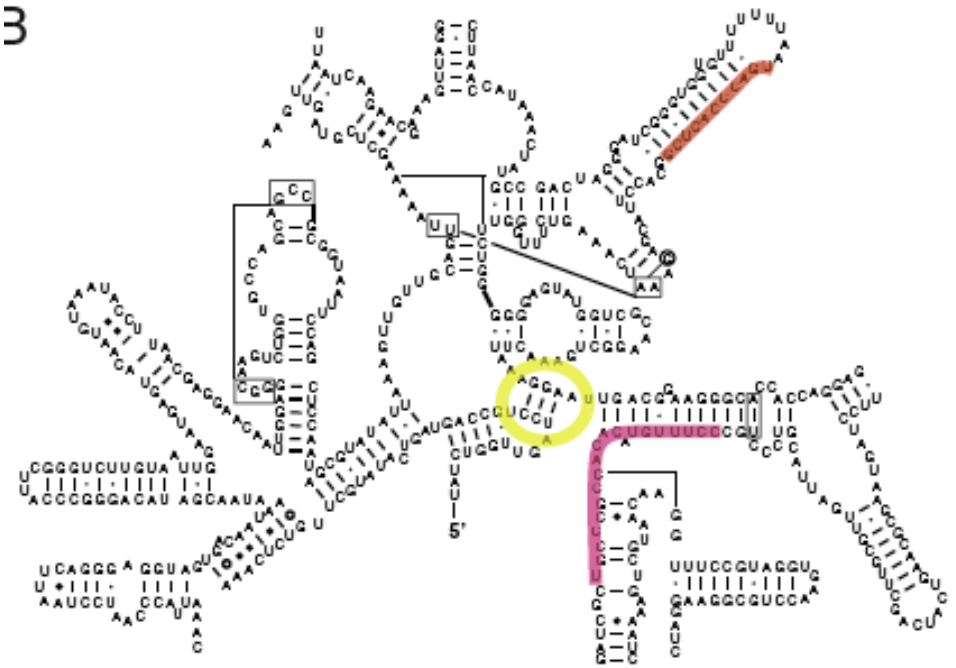
Fan Li,[a,b,c,1] Qi Zheng,[a,b,1] Lee E. Vandivier,[a,b,d] Matthew R. Willmann,[a,b] Ying Chen,[a,b,c] and Brian D. Gregory[a,b,c,d,2]

CLASH (intra- and intermolecular RNA-RNA interactions)

Crosslinking Ligation and Sequencing of Hybrids





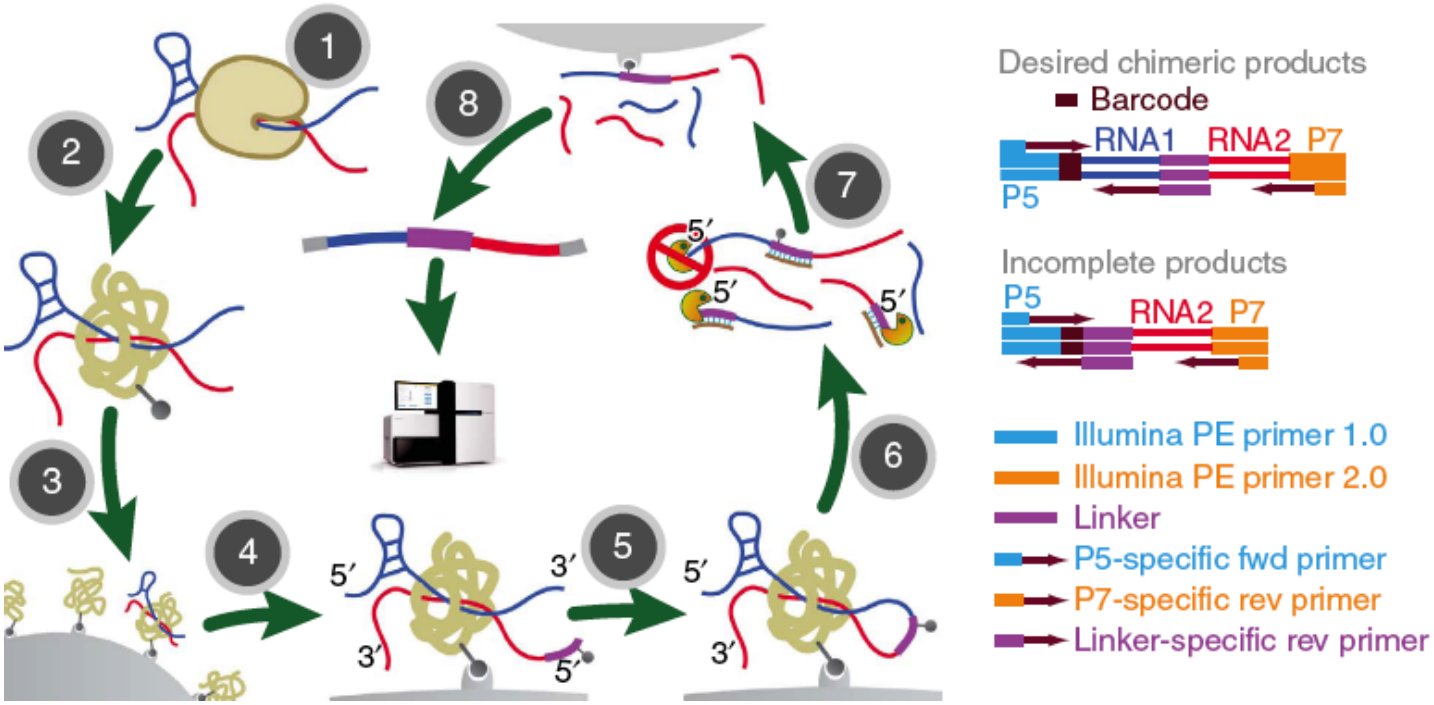


U3- 18S rRNA interactions

Kudla et al., PNAS, 2011

MARIO (intra- and intermolecular RNA-RNA interactions)

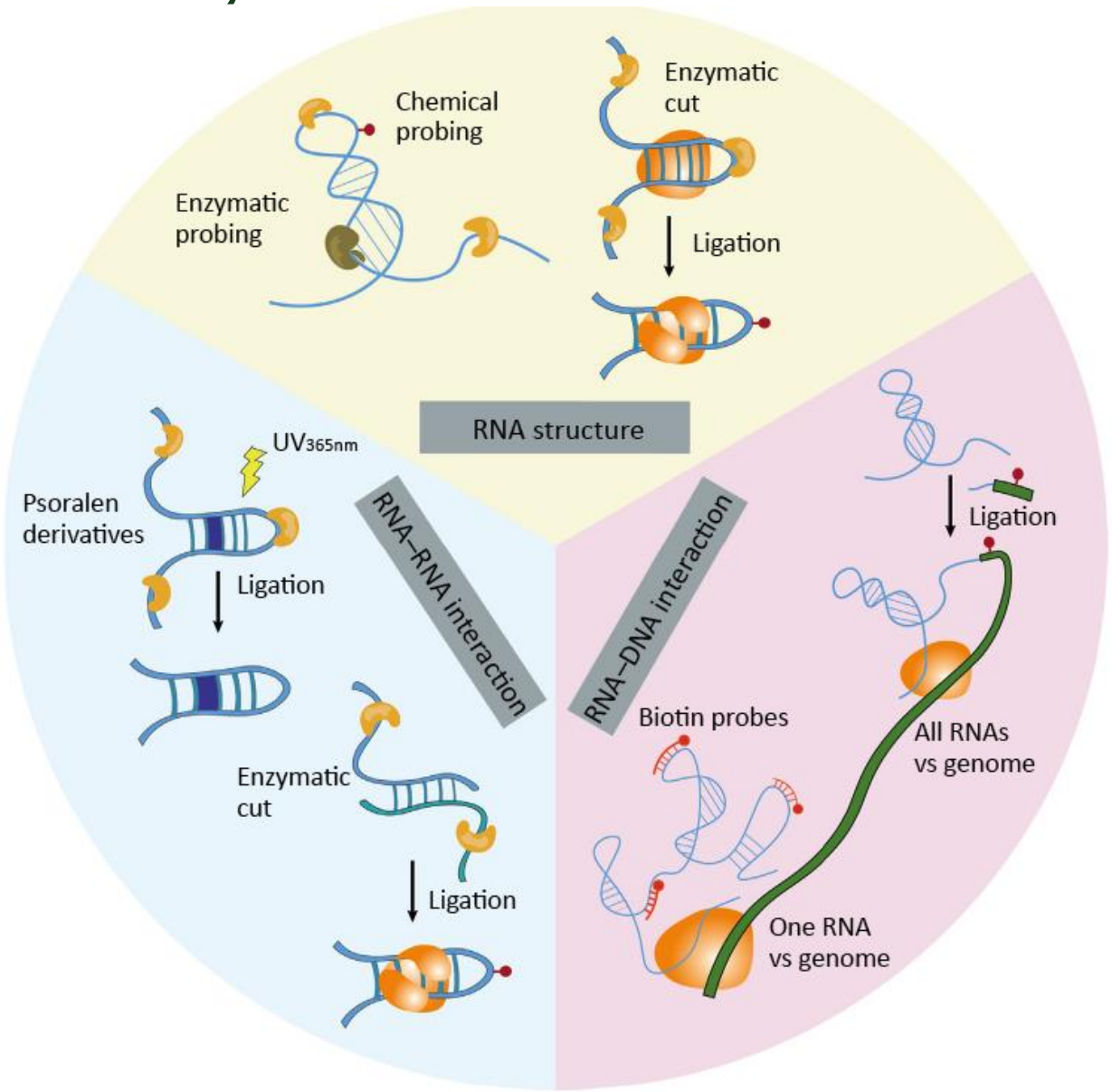
Mapping RNA interactome in vivo



(1) cross-linking RNAs to proteins
(2) RNA fragmentation, protein denaturing and biotinylation
(3) immobilization of RNA-binding proteins at low density
(4) ligation of a biotinylated RNA linker
(5) proximity ligation under a dilute condition
(6) RNA purification and RT
(7) biotin pull-down
(8) construction of sequencing library

Nguyen et al., NatComm, 2016

RNA-seq-based methods for mapping RNA structures, RNA–RNA and RNA–DNA interactions



Nguyen et al, *TiG, 2018*

RNA structure *in vivo*: SHAPE, PARIS/SPLASH/LIGR

Chemical and enzymatical- based structure probing

SHAPE: Selective 2'- Hydroxyl Acylation and Primer Extension

SHAPE-seq: SHAPE followed by RNA-seq

PARIS: Psoralen Analysis of RNA Interactions and Structures

SPLASH: Sequencing of Psoralen crosslinked, Ligated, and Selected Hybrids

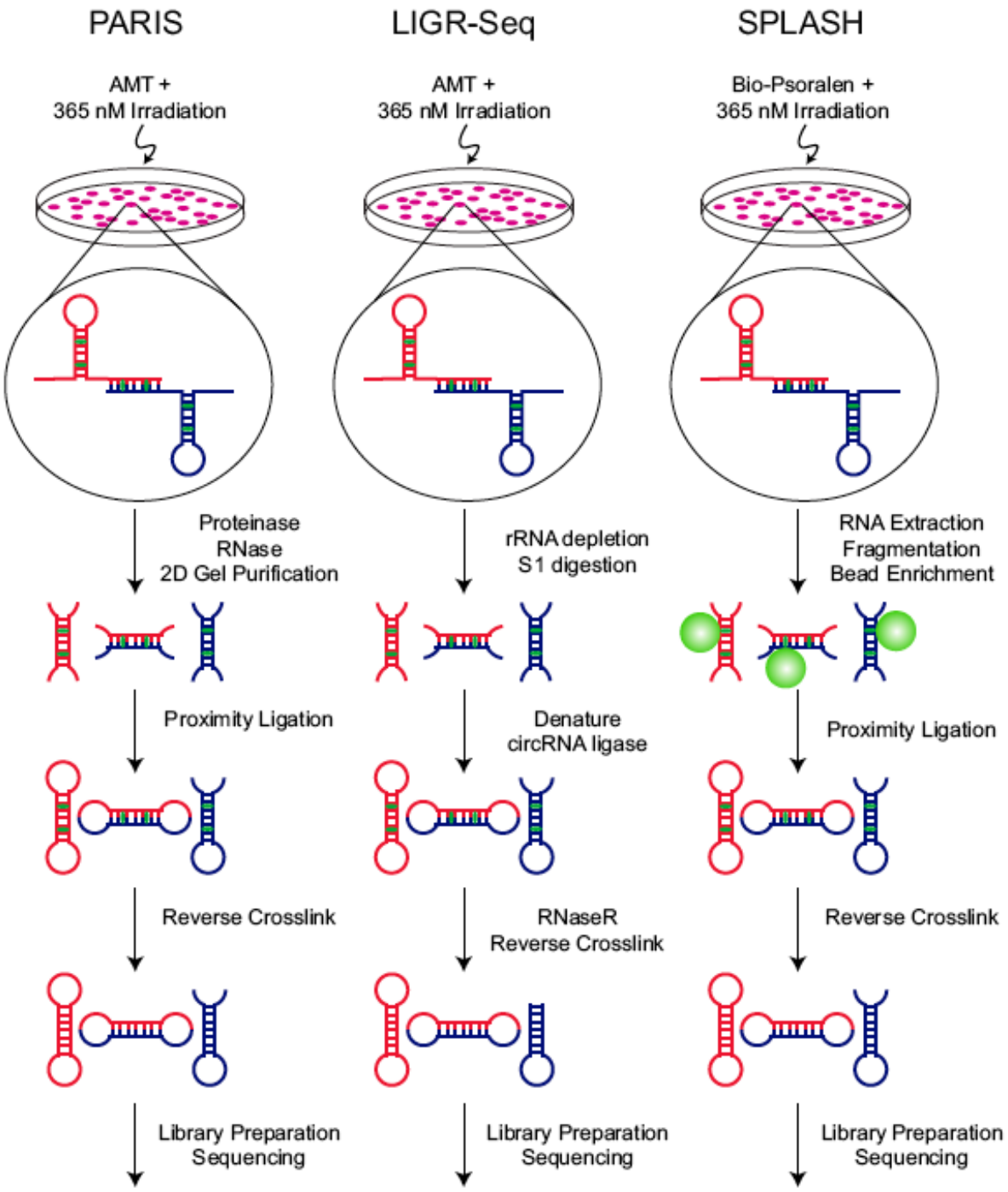
LIGR-seq: LIGation of interacting RNA followed by high-throughput Sequencing

SHAPE chemicals: DMS, dimethyl sulfate; 1M7, 1-methyl-7-nitroisatoic anhydride
SHAPE enzymes: P1 nuclease, RNases V1 and S1
PARIS/SPLASH chemicals: psoralen; AMT, 4'-aminomethyltrioxsalen

Table 1. Transcriptome-wide RNA Structure Probing Methods

Assay	Probing Agent	Detection	In Vitro Probing	In Vivo Probing
FragSeq	P1 nuclease	single-stranded bases	X	
PARS	RNase V1 and S1 nuclease	paired and single-stranded regions	X	
SHAPE-seq	1M7	single-stranded bases	X	
mod-seq	DMS	unpaired A & C		X
DMS-seq	DMS	unpaired A & C	X	X
Structure-seq	DMS	unpaired A & C	X	X
icSHAPE	NAI-N_3	single-stranded bases		X
SHAPE-MaP	1M7	single-stranded or unbound bases	X	X
PARIS	AMT	base-paired sequence partners		X
LIGR-seq	AMT	base-paired sequence partners		X
SPLASH	biotinylated psoralen	base-paired sequence partners		X

Graveley, Mol Cell, 2016



PARIS
SPLASH
LIGR

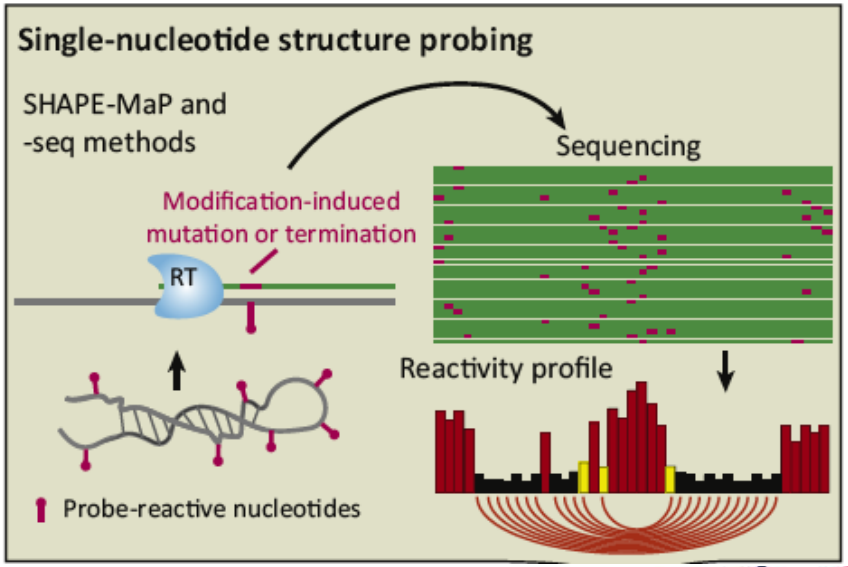
- in vivo psoralen or AMT, intercalate into RNA duplex and generate inter-strand adducts between juxtaposed pyrimidine bases upon 365 nm UV irradiation
- ssRNAse S1 limited digest
- RNA end proximity ligation (circRNA ligase)
- removal of uncrosslinked RNA (ss and structured RNAase R1)
- crossling reversal (254 nm)
- RNAseq

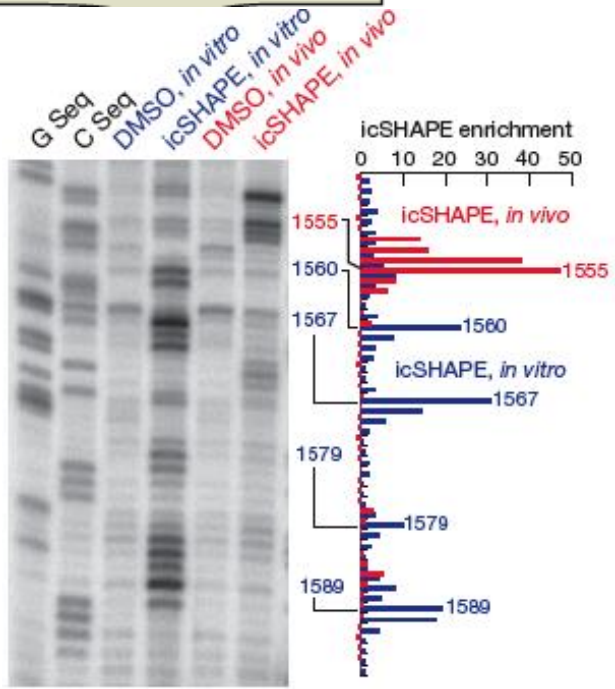
[AMT = psoralen derivative 4'-aminomethyltrioxalen] and

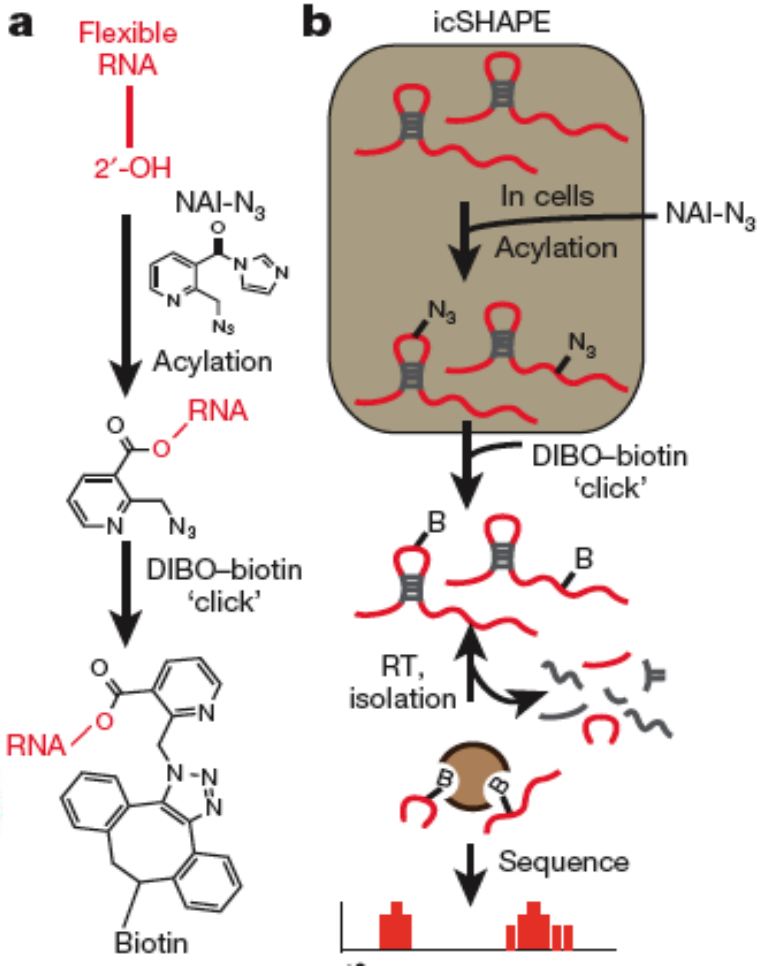
Graveley, Mol Cell, 2016

RNA structure in vivo: SHAPE, icSHAPE

icSHAPE: click selective 2'-hydroxyl acylation and profiling

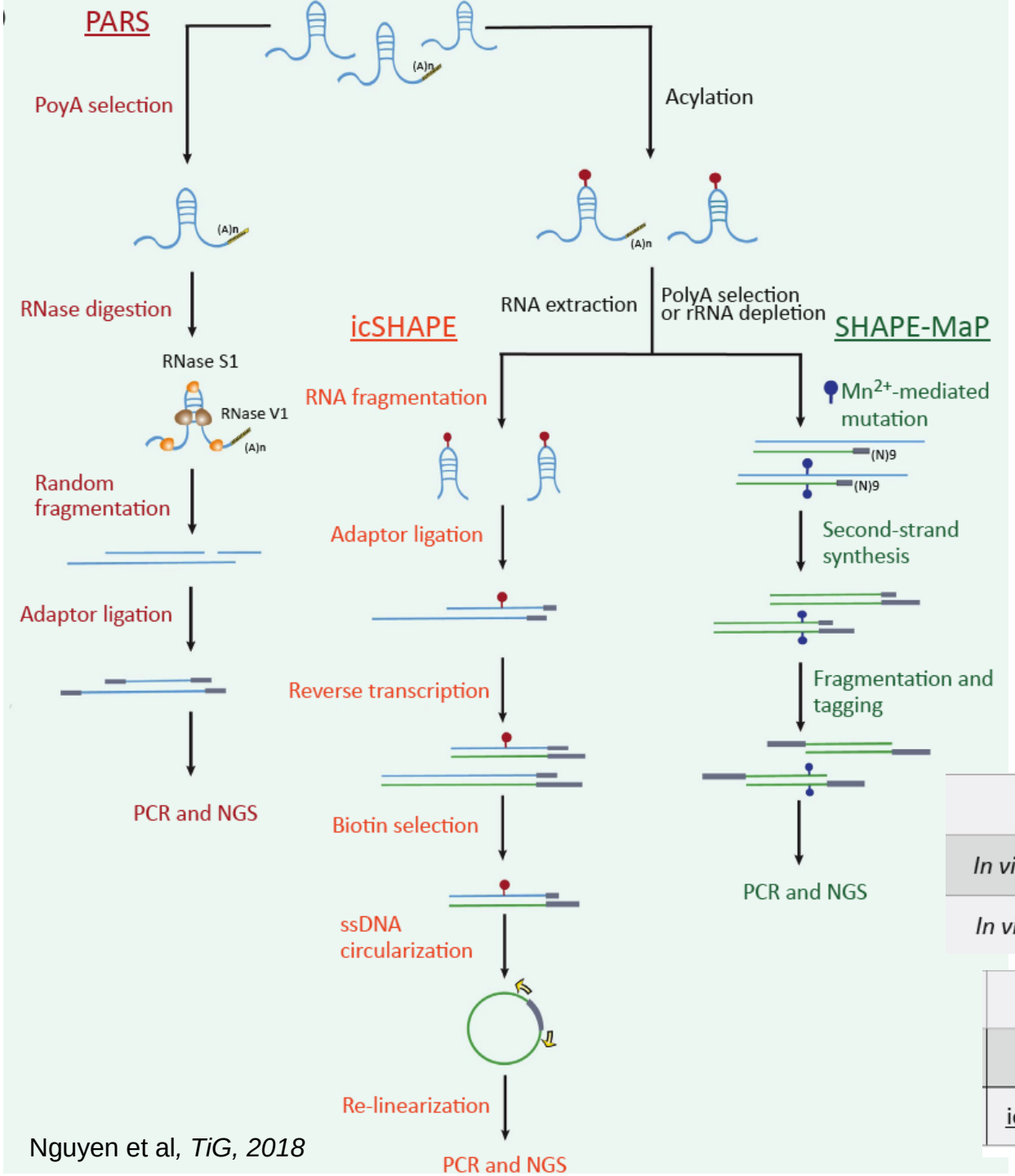






Low and Weeks, Methods 2010
Weidman et al, TiBS, 2016

Mapping RNA structures and RNA-RNA interactions



Nguyen et al, *TiG, 2018*

	Enzyme based			
In vitro	PARS	PIP-seq	PARTE	FragSeq
In vivo	MARIO			

Chemical based			
SHAPE-MaP		CIRS-seq	
icSHAPE	Structure-seq	DMS-seq	Mod-seq

RNA-seq-based mapping of RNA structures

Sequencing based for mapping RNA structures		
Method	Advantages	Limitations
Enzyme-based methods		
PIP-seq	• Reveals both protein-bound RNA regions and RNA secondary structure. • Provides strand-specific information.	• Limited resolution at small nucleotide bulges and loops.
PARS	• Increased sensitivity by sequencing both single- and double-stranded regions.	• RNA was folded *in vitro*.
PARTE	• Measures melting temperature. • Single-nucleotide resolution. • Preserves *in vivo* RNA modifications. • Can infer RNA regulatory motifs.	
FragSeq	• Simple and fast protocol. • Accompanied with modifiable software.	
Chemical-based methods		
DMS-seq	• Identifies RNA structure in native conditions. • Single-nucleotide resolution.	• Limited to the analysis of two bases (As and Cs). • RNA-binding proteins can block DMS activity.
icSHAPE	• Measures base flexibility. • Single-nucleotide resolution.	• Limited to the analysis of relatively short (~300 nt) *in vitro*-transcribed RNAs.
Structure-seq	• Single-nucleotide resolution. • Applicable to both *in vitro* and *in vivo* analyses.	• Limited to the analysis of two bases (As and Cs). • RNA-binding proteins can block DMS activity.
Mod-seq	• Can probe structures of long RNAs *in vivo*. • Single-nucleotide resolution.	• Limited to the analysis of two bases (As and Cs).
CIRS-seq	• Single-nucleotide resolution. • Can identify structural requirements for RNA-binding proteins.	
SHAPE-MaP	• Can be customized for different applications. • Applicable to analysis of long RNAs. • Can infer structural changes of single-nucleotide and other allelic polymorphisms.	• Length of the RNA must be at least ~150 nt for the randomer and native workflow, and at least ~40 nt for the small-RNA workflow.

Nguyen et al, *TiG, 2018*

RNA-seq-based mapping of RNA-RNA interactions

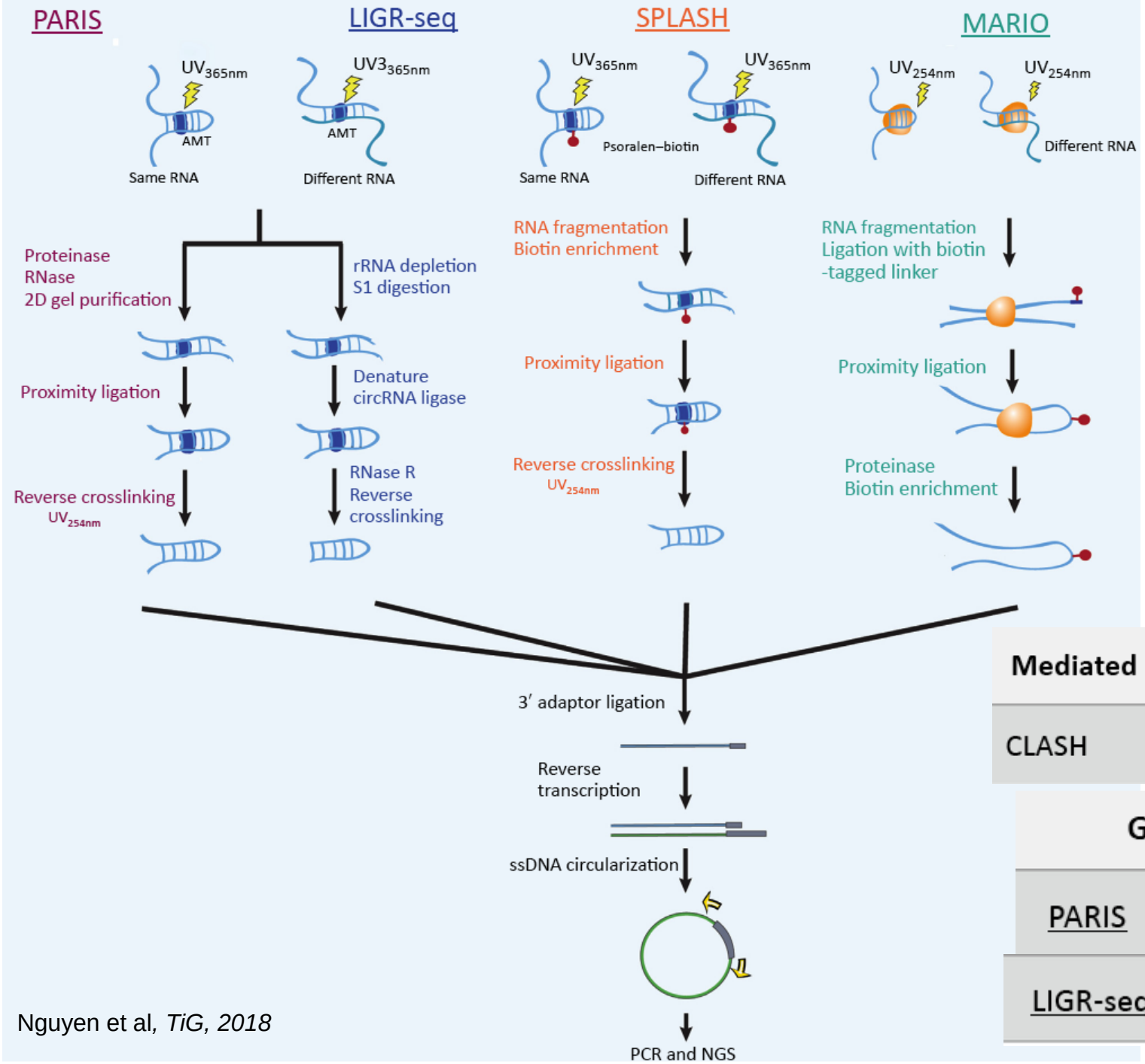
Sequencing based for mapping RNA–RNA interactions		
CLASH	• Stringent purification conditions remove nonphysiological interactions.	• Requires prior knowledge of an RNA-binding protein. • Requires a good antibody.
hiCLIP	• Incorporation of an adaptor between two RNA molecules increases ligation efficiency and improves accuracy in sequence mapping.	• Requires prior knowledge of an RNA-binding protein. • Requires a good antibody. • No *in vivo* crosslinking step may incur challenges in differentiating *bona fide* and spurious RNA attachments.
PARIS	• Many-to-many mapping.	• 4′-Aminomethyl trioxsalen (AMT) preferentially crosslinks pyrimidine bases and may introduce bias.
SPLASH	• Improves signal-to-noise ratio by leveraging biotinylated psoralen. • Many-to-many mapping.	• Psoralen preferentially crosslinks pyrimidine bases and may introduce bias.
LIGR-seq	• Many-to-many mapping.	• AMT preferentially crosslinks pyrimidine bases and may introduce bias.
MARIO	• Many-to-many mapping. • Incorporation of an adaptor between two RNA molecules increases ligation efficiency and improves accuracy in sequence mapping. • Reports both between- and within-molecule interactions.	• Loses RNA duplexes that are not associated with any proteins.
	• Captures proximal regions of an RNA molecule in 3D. • Reveals single-stranded regions of each RNA.	

Nguyen et al, *TiG, 2018*

RNA-seq-based mapping of RNA-RNA interactions

Sequencing based for mapping RNA–DNA interactions		
Method	Advantages	Limitations
	• Captures proximal regions of an RNA molecule in 3D. • Reveals single-stranded regions of each RNA.	
Sequencing based for mapping RNA–DNA interactions		
ChIRP	• Tilling the entire transcript with antisense DNA.	• Limited to analyzing RNA at a time.
CHART	• Tilling the RNase H accessible region by antisense DNA.	• Limited to analyzing RNA at a time.
RAP	• Tilling the entire transcript with complimentary RNA.	• Limited to analyzing RNA at a time. • Limited to analysis of long RNA.
MARGI	• Many-to-many mapping. • Captures interaction at native conditions.	• Require a large number (10^7) of cells.
ChAR-seq	• Many-to-many mapping. • Proximity ligation is performed in nuclei, which reduces nonspecific interactions.	• Only sequencing reads that cover the entire bridge sequence are informative, reducing the number of informative reads.
GRID-seq	• Many-to-many mapping. • Proximity ligation is performed in nuclei, which reduces nonspecific interactions.	• The informative sequence lengths on the RNA side and the DNA side are both limited to ~20 bases, resulting in challenges in unambiguous sequence mapping.

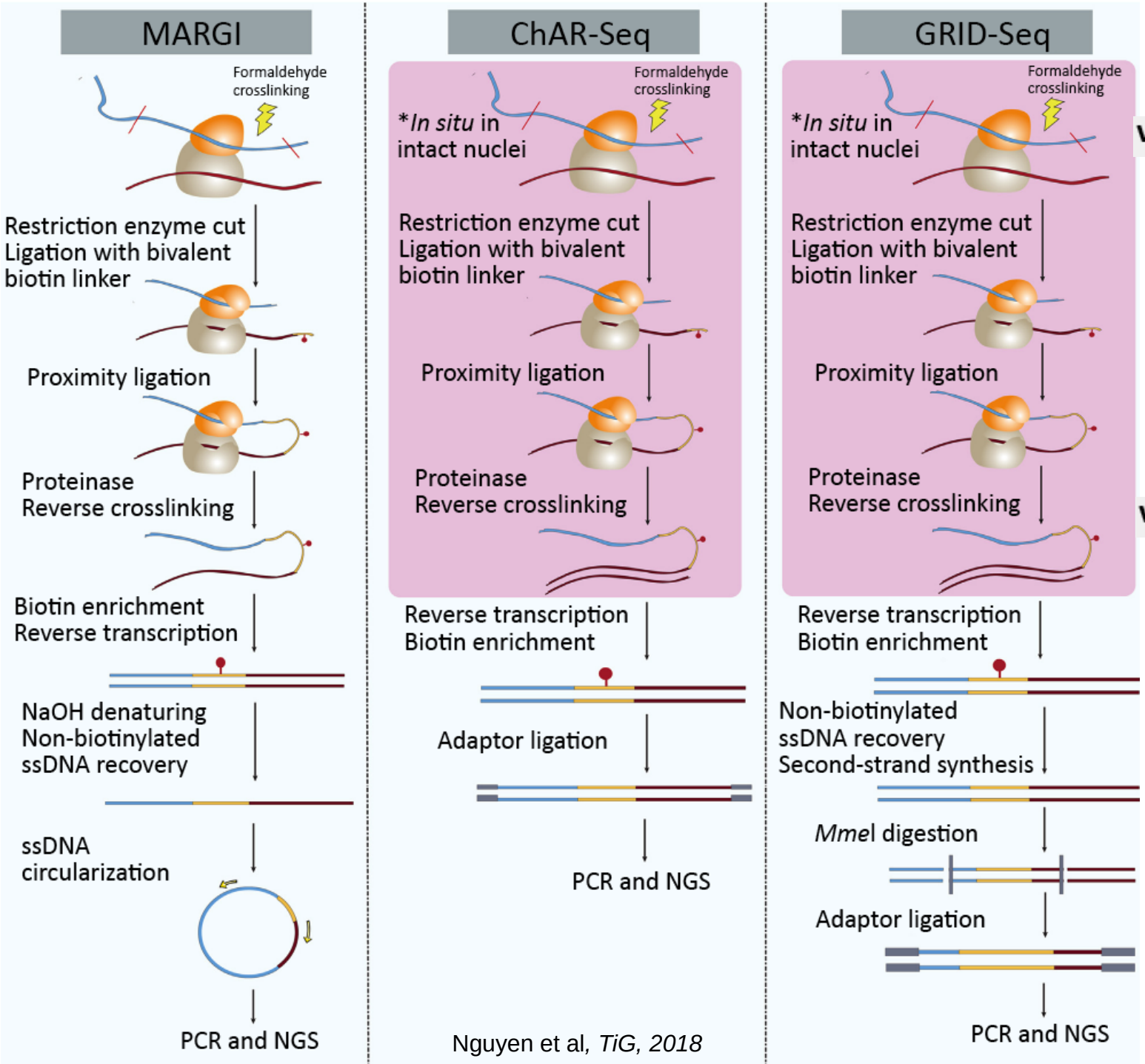
Nguyen et al, *TiG, 2018*



Mediated by a specific protein	
CLASH	hiCLIP

Genome-wide	
PARIS	SPLASH
LIGR-seq	MARIO

Nguyen et al, *TiG, 2018*



One RNA versus the genome

ChIRP-seq

CHART-seq

RAP-seq

All RNAs versus the genome

MARGI

ChAR-seq

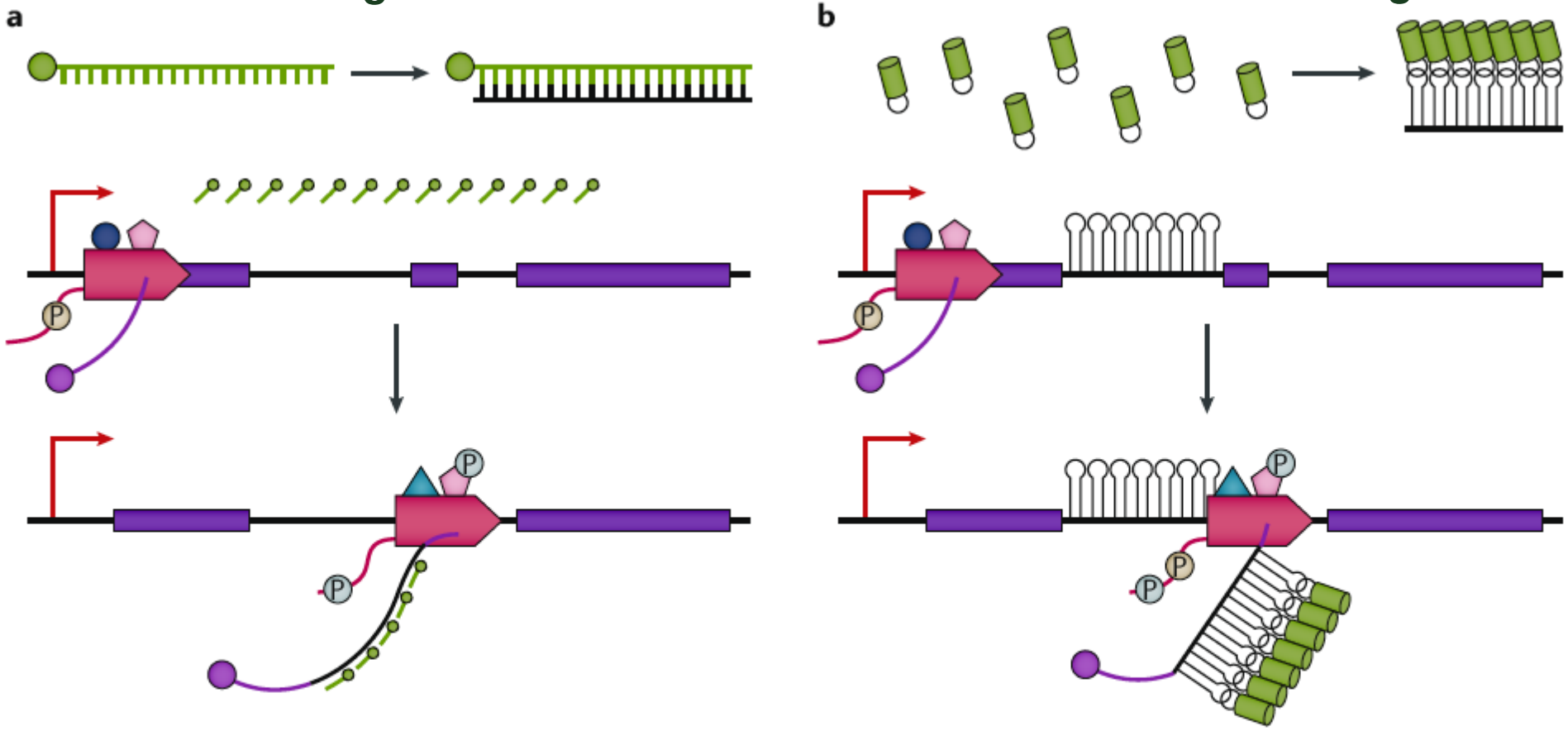
GRID-seq

Nguyen et al, *TiG, 2018*

Imaging of nascent RNA

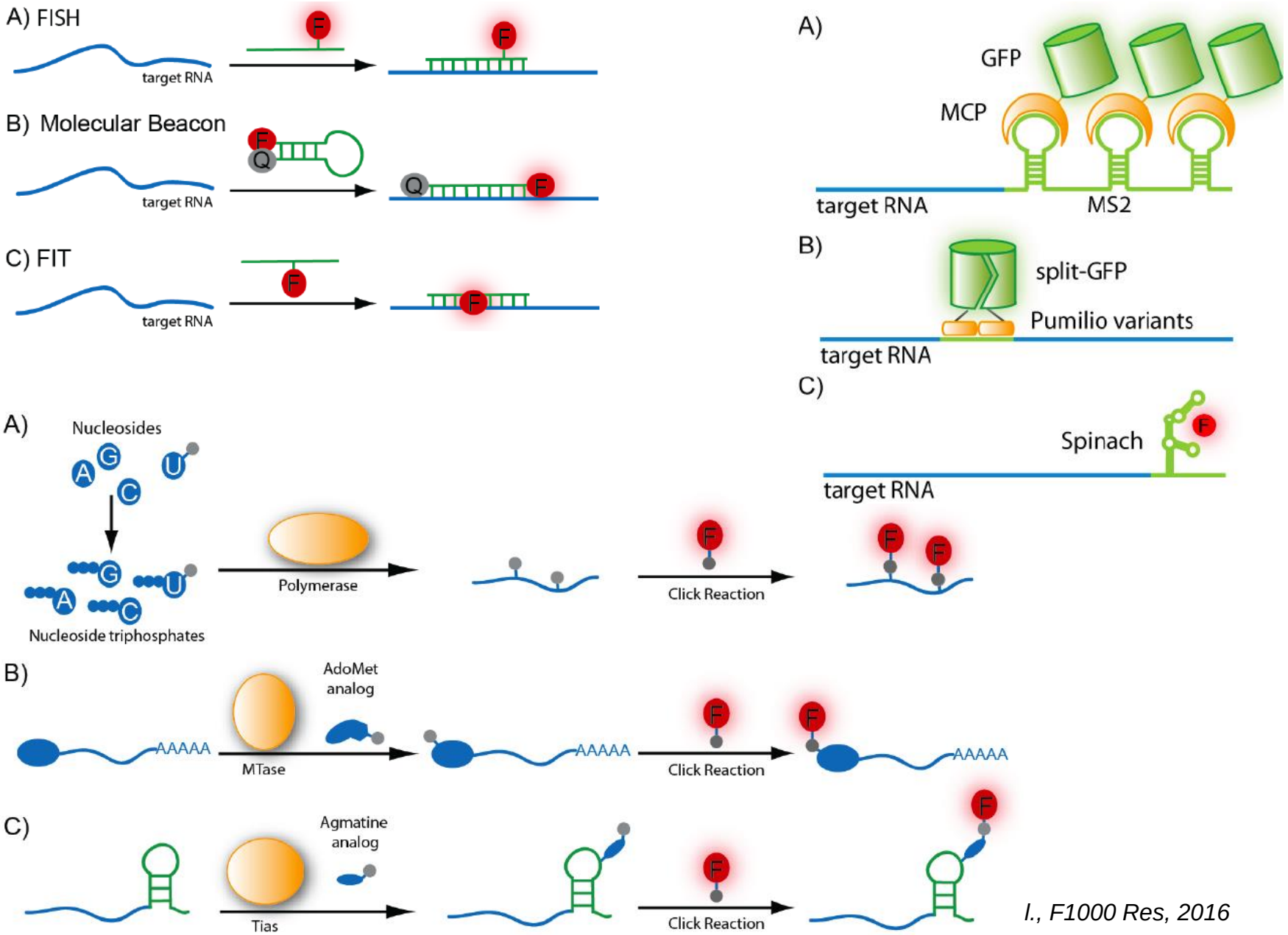
smFISH – single molecules FISH

in living cells
MS2/MCP labeling



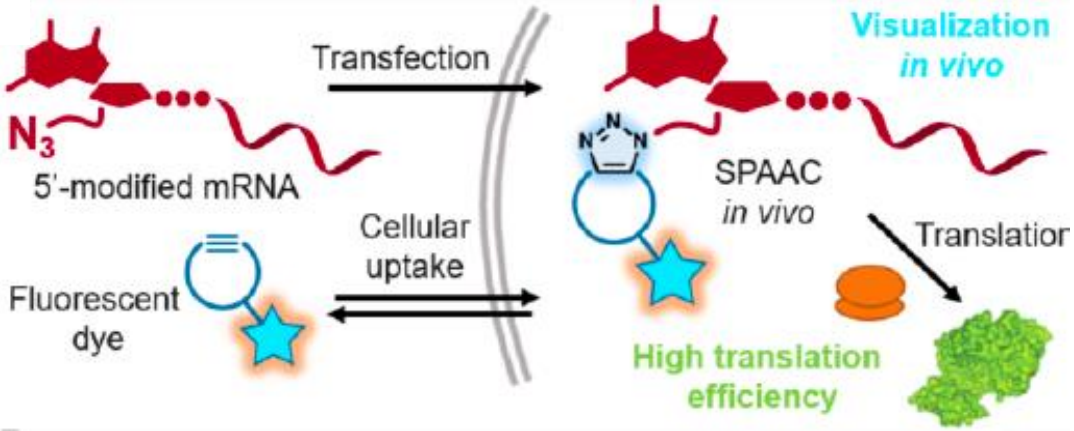
Wissink et al, Nat Rev Genet, 2019

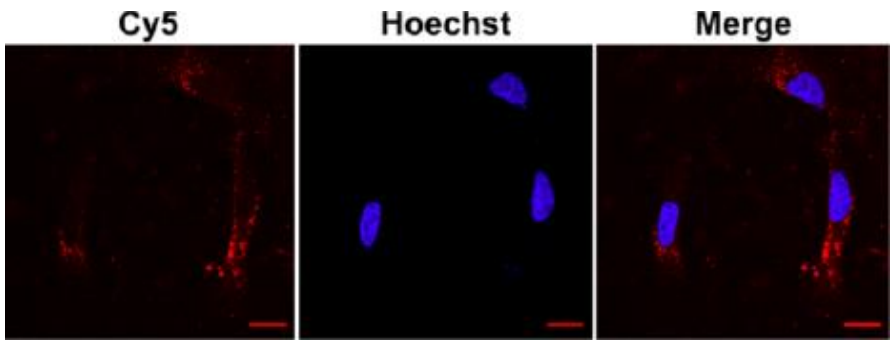
mRNA fluorescent labeling for imaging

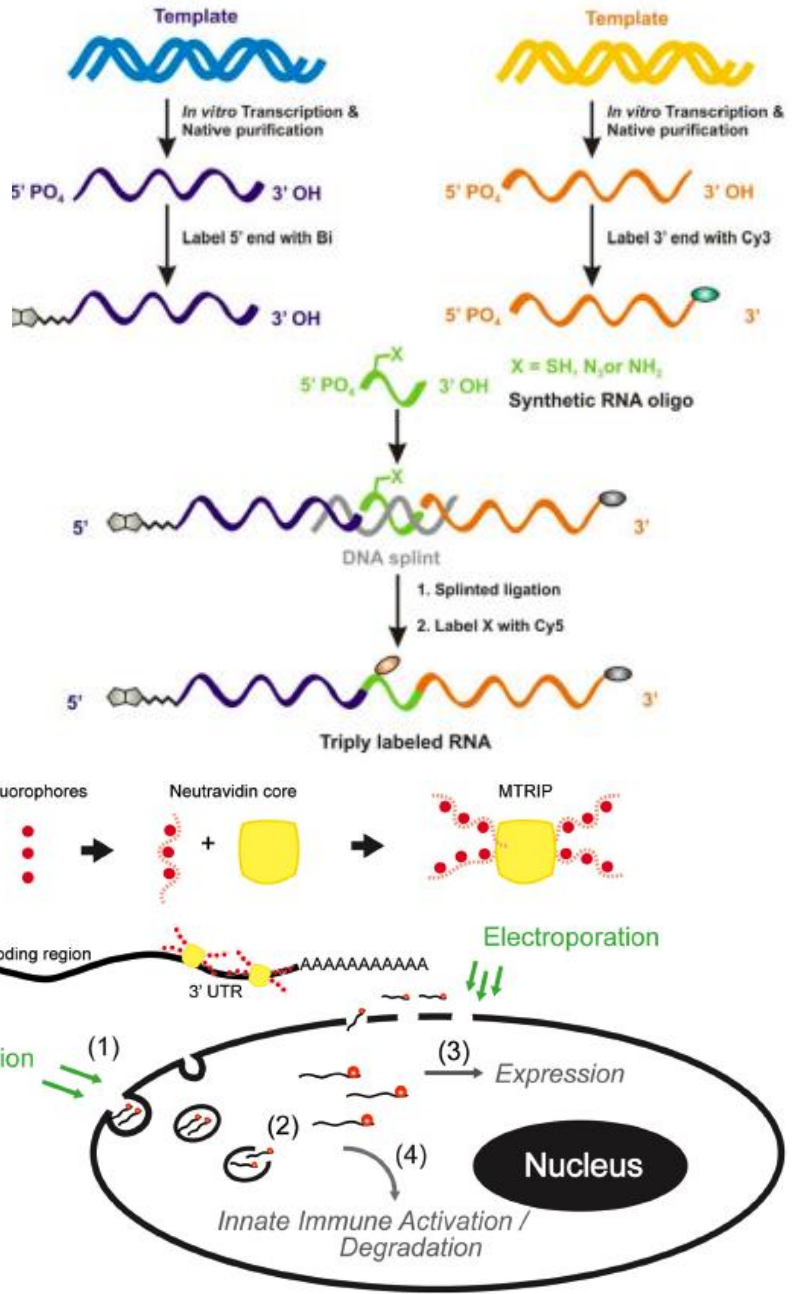


I., F1000 Res, 2016

mRNA fluorescent labeling for imaging

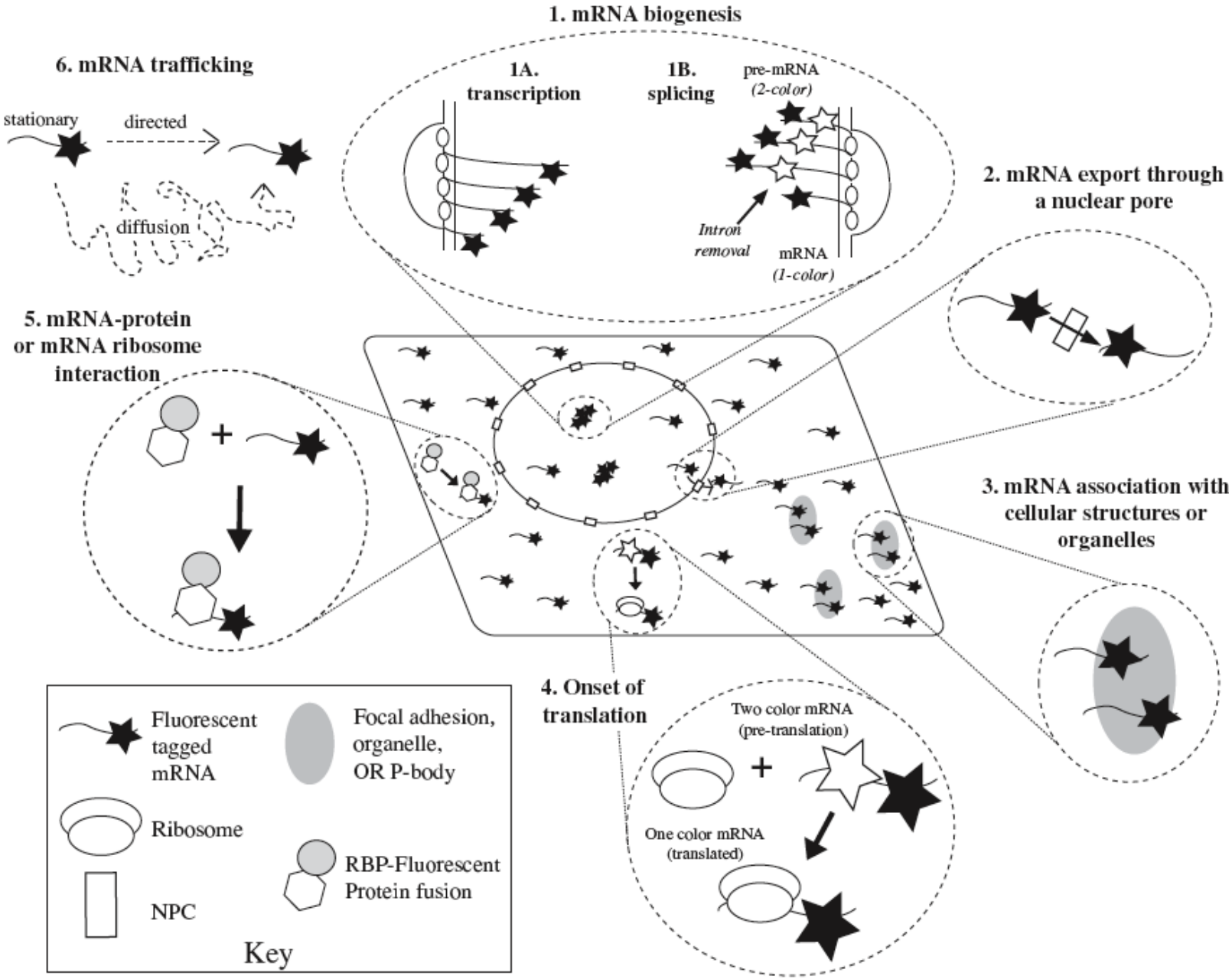






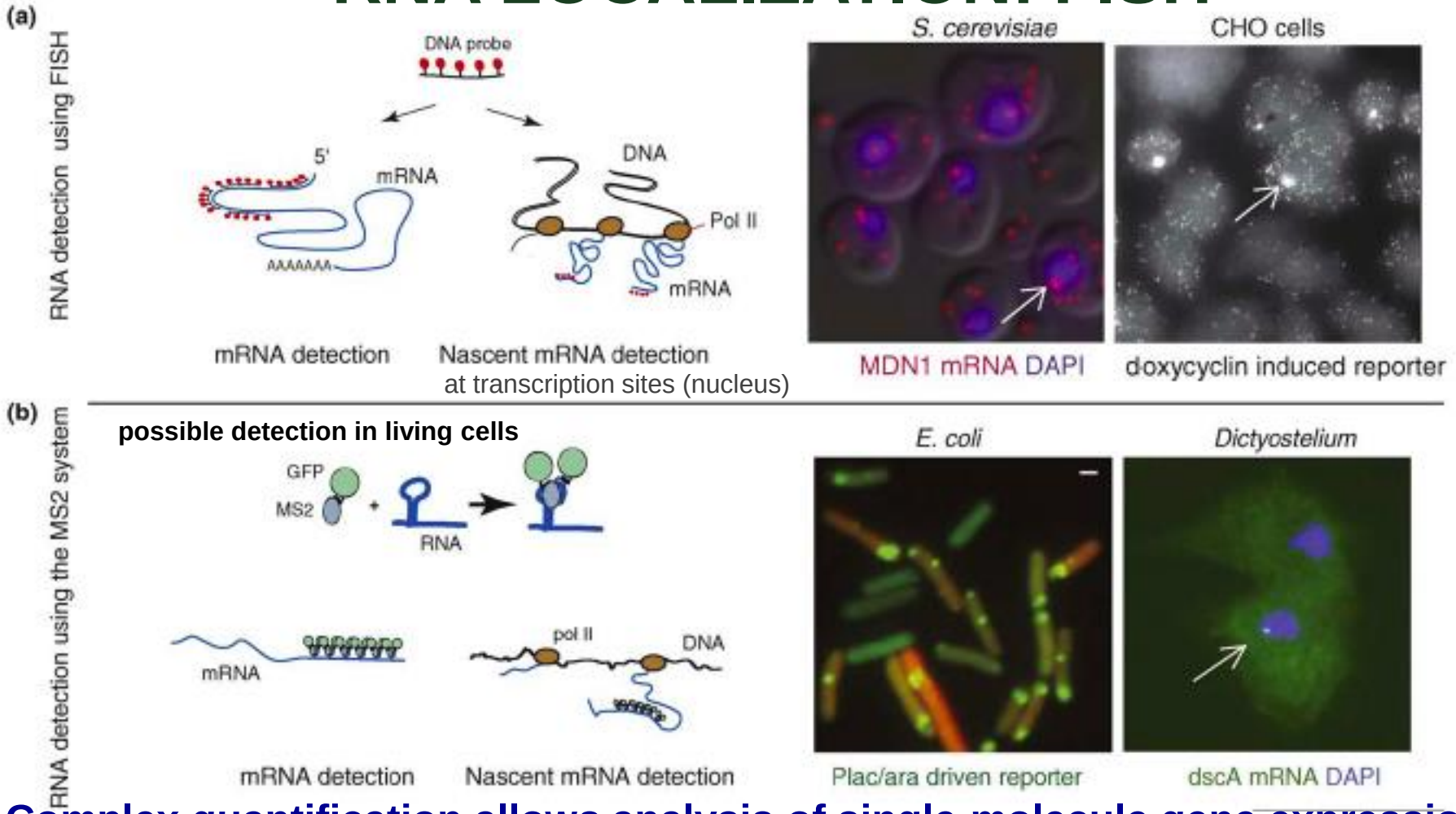
Mamot et al, Angew Chem, 2017
Rinaldi et al., Methods Mol Biol, 2015
Kirschman et al. NAR 2017

FISH: Fluorescent *in situ* hybridization



Abbaszadeh and Gavis, Methods, 2016

RNA LOCALIZATION: FISH



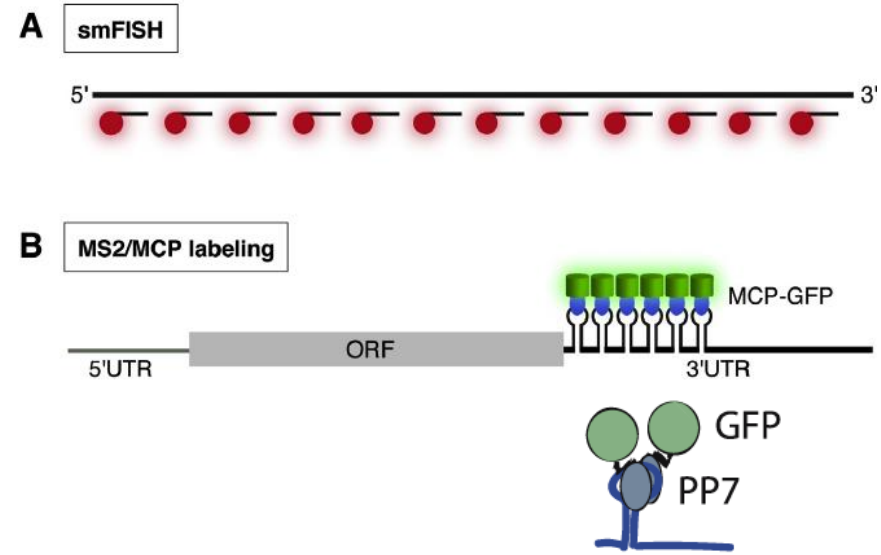
Larson et al., TiCB, 2009

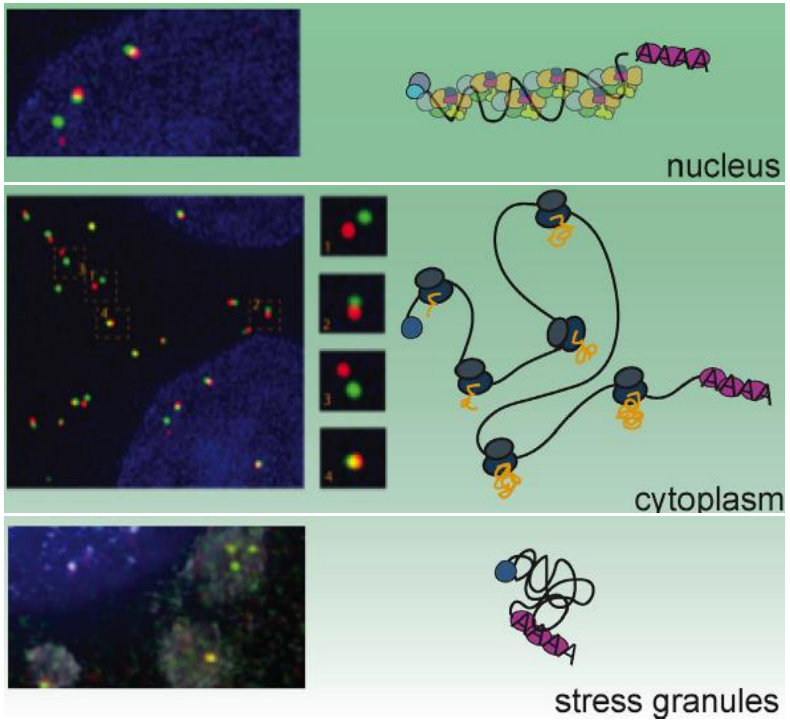
Complex quantification allows analysis of single-molecule gene expression, e.g. transcription/splicing in real time, RNA level in single cells.

- **Constitutively expressed genes are transcribed by single events separated in time; regulated genes (e.g. by SAGA) are expressed by transcriptional bursts**
- **Transcription of functionally related constitutive genes is not coordinated**

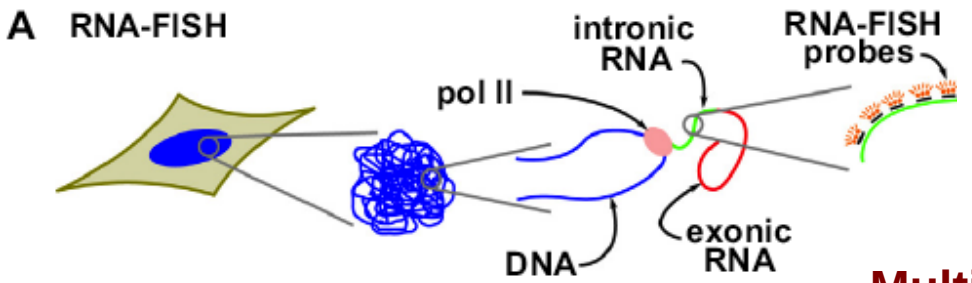
(regulated post-transcriptionally or post-translationally) */Singer lab/*

RNA LOCALIZATION: FISH

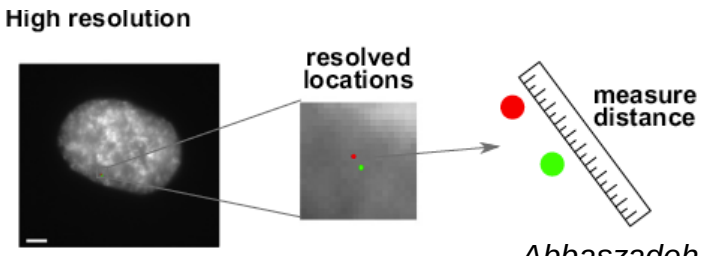




Adivarahan et al., Mol Cell 2018



RNA detection at transcription sites with intronic probes



Abbaszadeh and Gavis, Methods, 2016

Multi-colored smFISH:

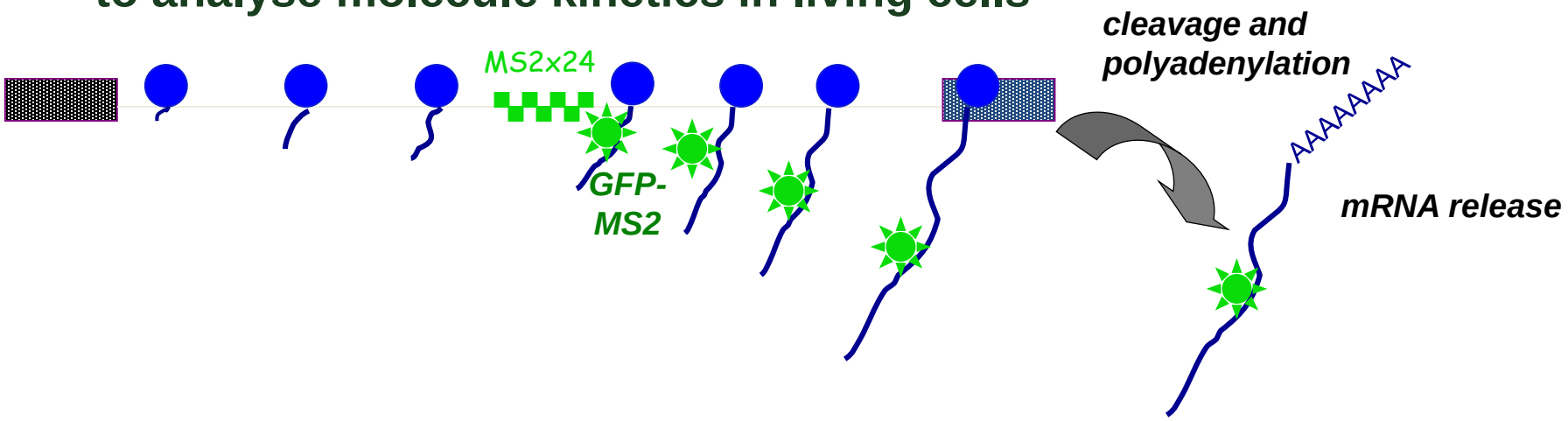
1. Nuclear mRNAs are partially extended

2. Translating mRNAs usually do not have a circular form (no interacting 5' and 3' ends)

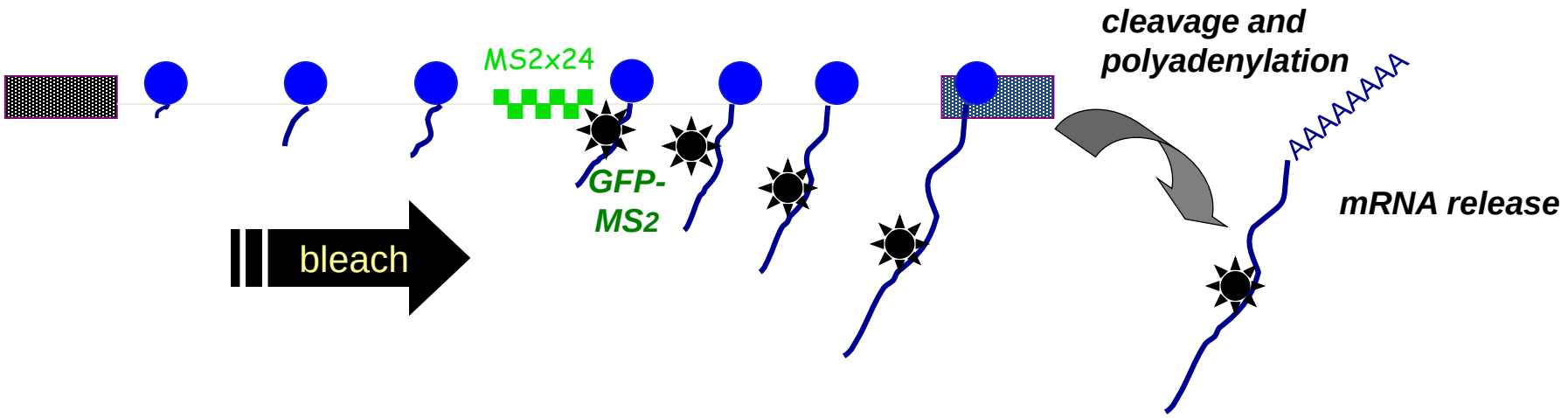
3. mRNAs in stress granules are more compacted than translating mRNAs

RNA LOCALIZATION: FRAP and FLIP

FRAP - fluorescence recovery after photobleaching
FLIP - fluorescence loss in photobleaching
to analyse molecule kinetics in living cells

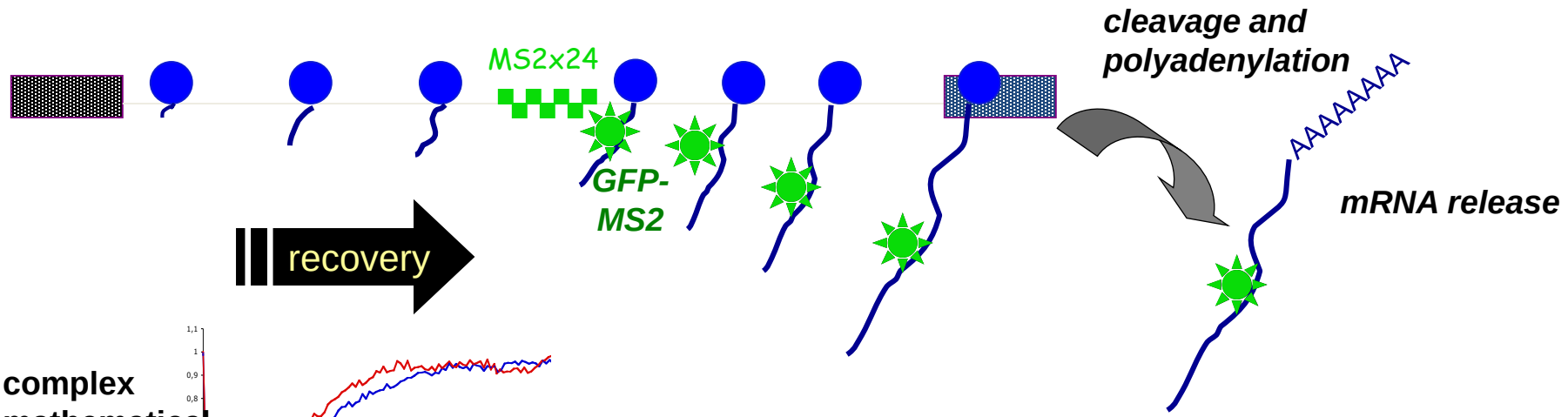


RNA LOCALIZATION: FRAP

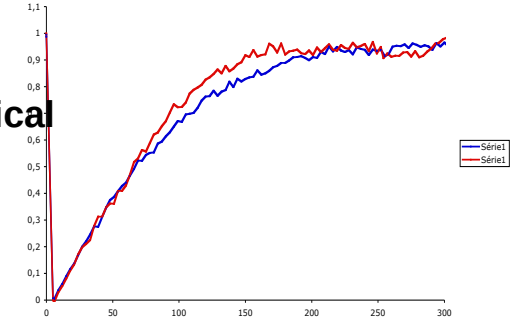


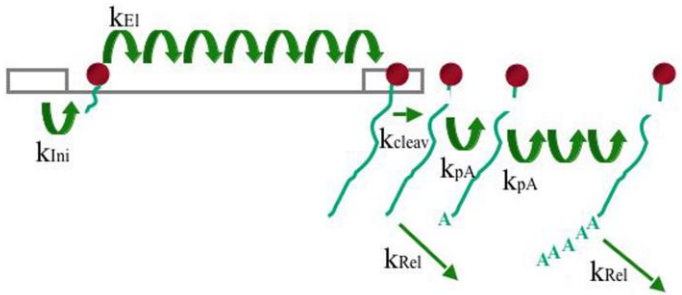
RNA LOCALIZATION: FRAP

Analysis of: - transcription rates
- 3'-end formation
- transcript release

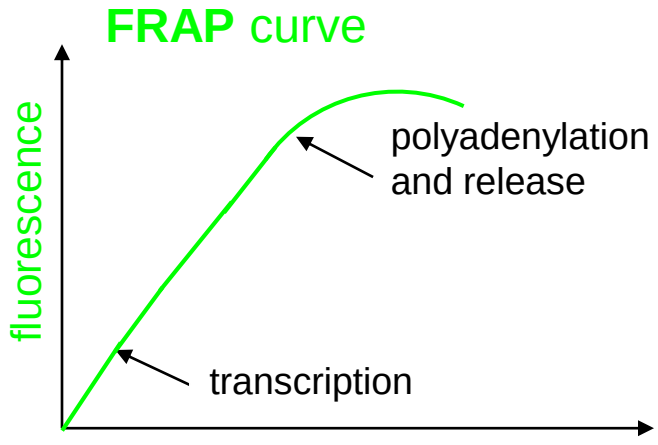


complex mathematical modeling





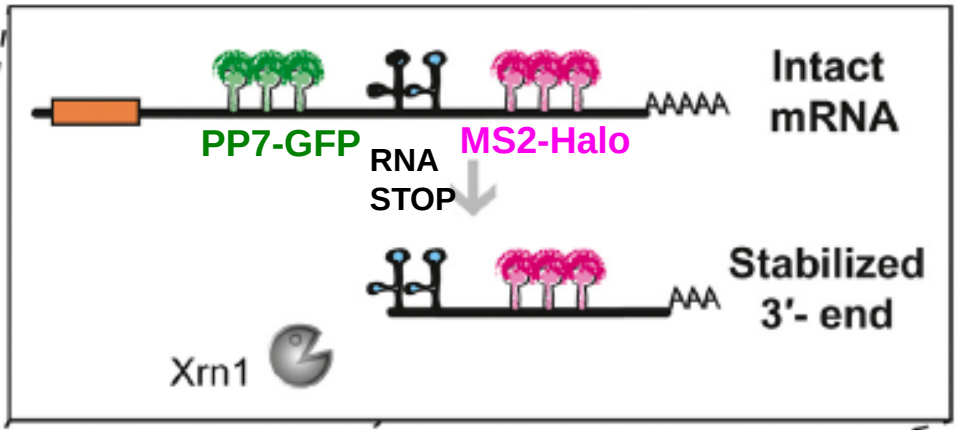
trx longation rate: 2 kb/min

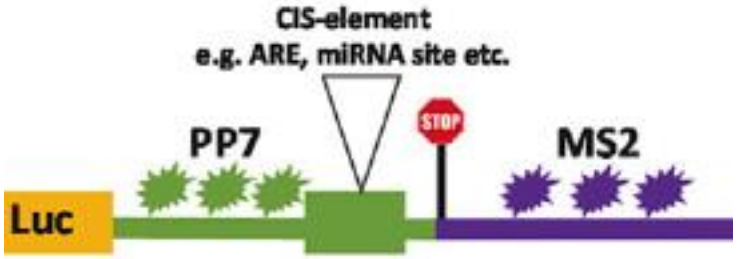


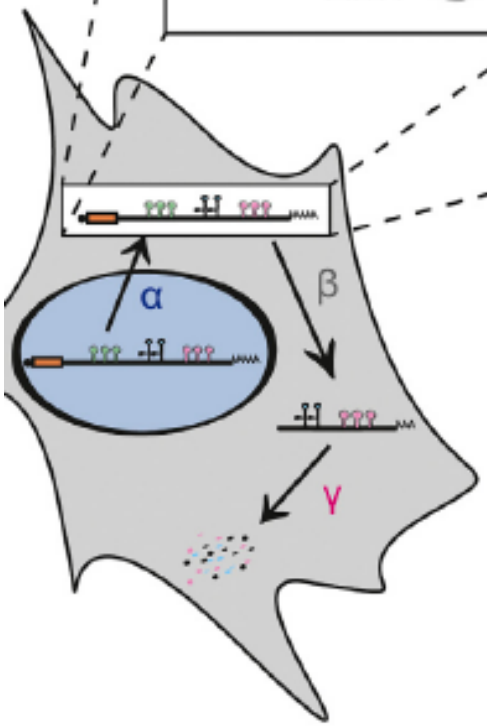
Edouard Bertrand, Montpellier, *RIBOSYS*

TREAT- 3'-RNA end accumulation during turnover

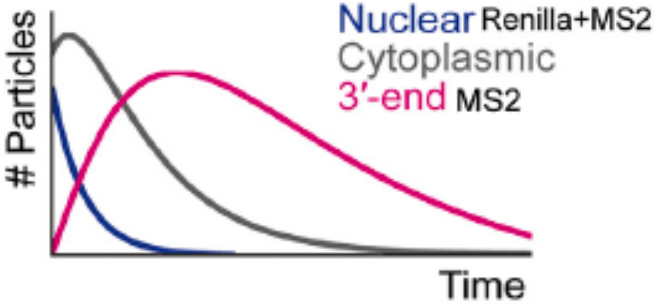
Single-mRNA imaging of RNA degradation in single cells







Real-time imaging of siRNA-mediated slicing

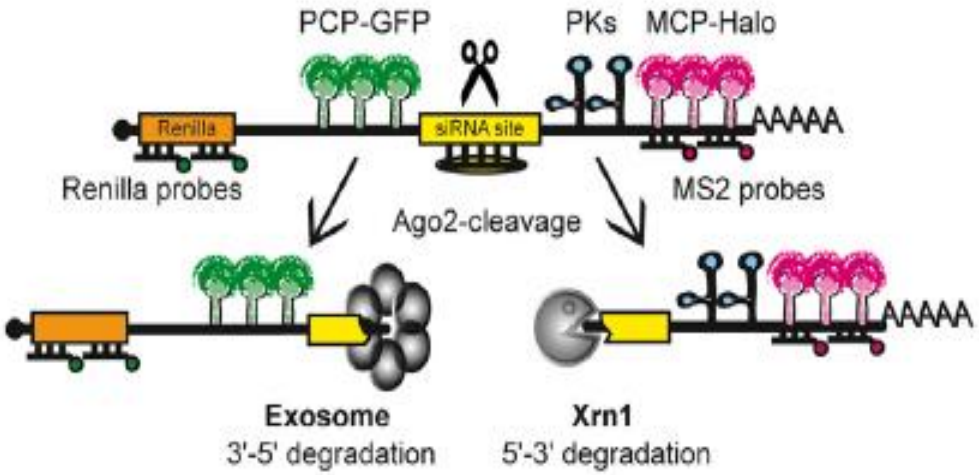


Quantification of mRNA decay dynamics

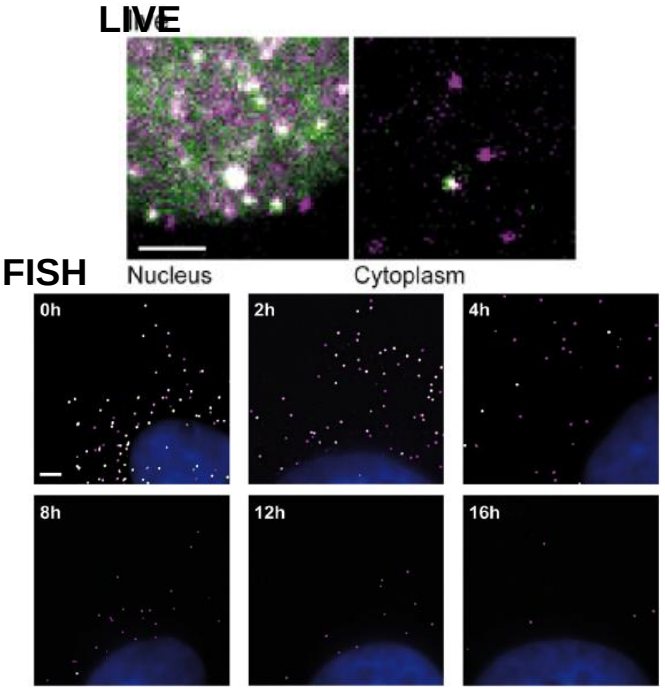
Horvathova et al, Mol Cell, 2017

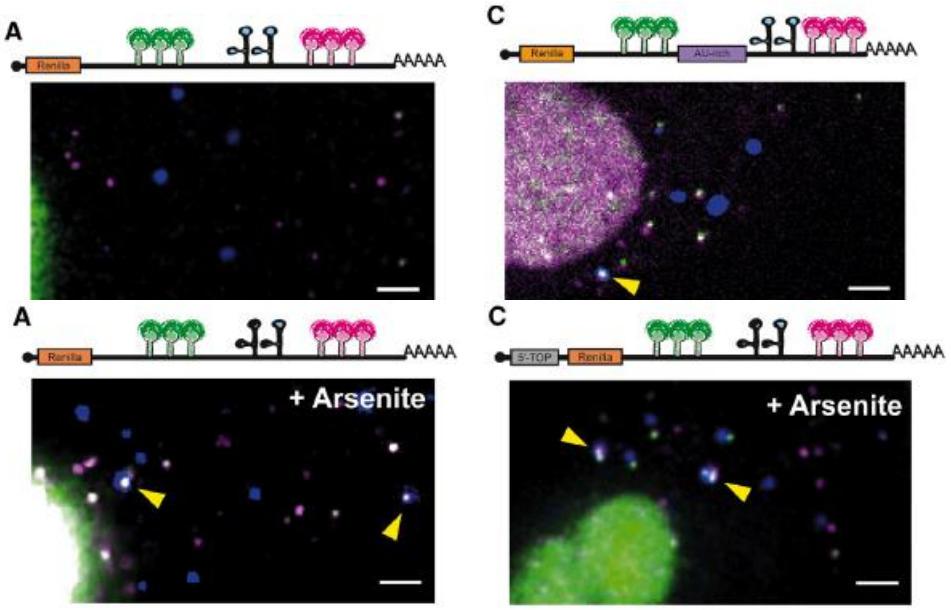
TREAT- 3'-RNA end accumulation during turnover

- **real-time observation of Ago2 slicing of TREAT mRNAs**



- **single-mRNA imaging of RNA degradation in single cells**



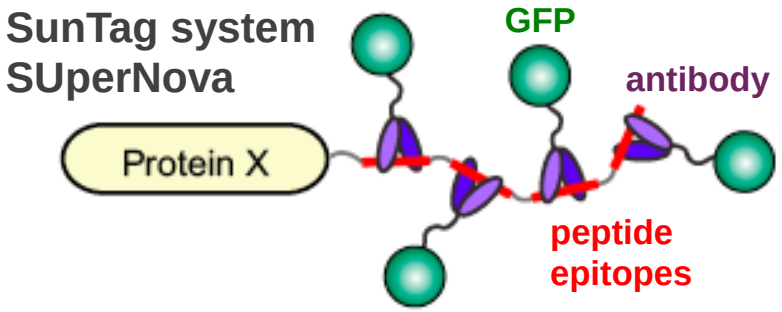


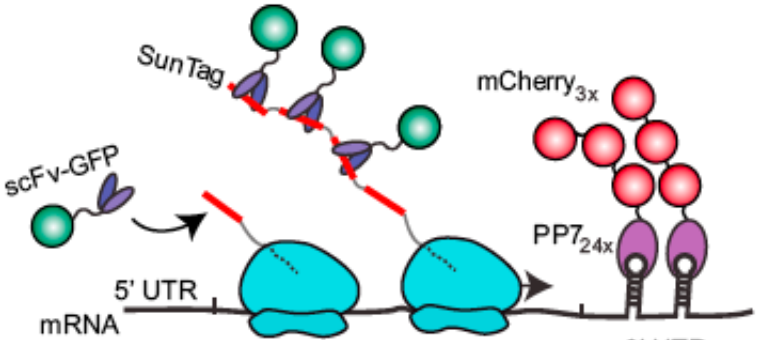
- **TREAT mRNAs are not degraded in P-bodies, also in stress**

DCP1- PB marker

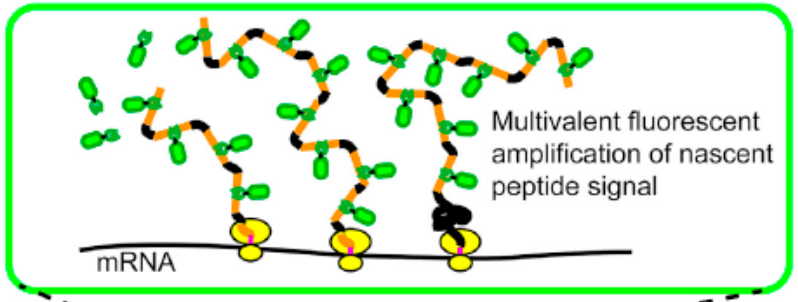
Horvathova et al, Mol Cell, 2017

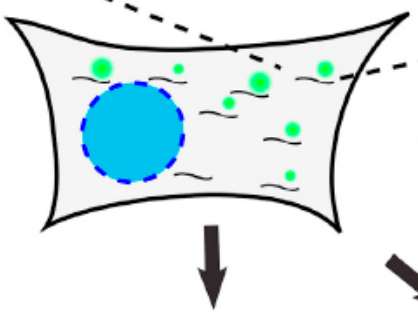
SunTag -Translation of single mRNAs *in vivo*



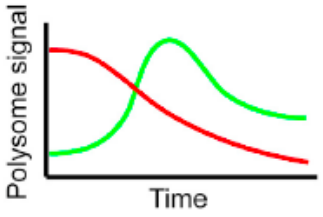


Translation visualized by labeling nascent peptide epitopes with antibody-GFP

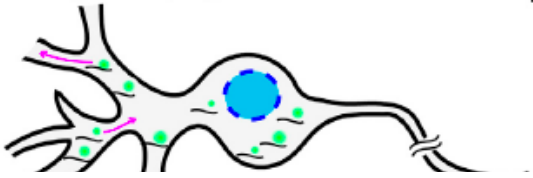




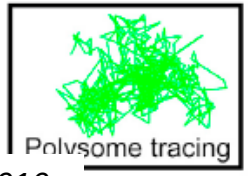
Translational stress response

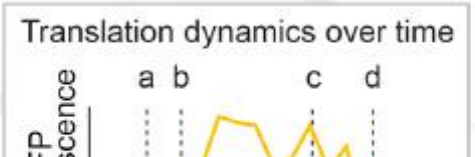


Local translation and active transport of polysomes

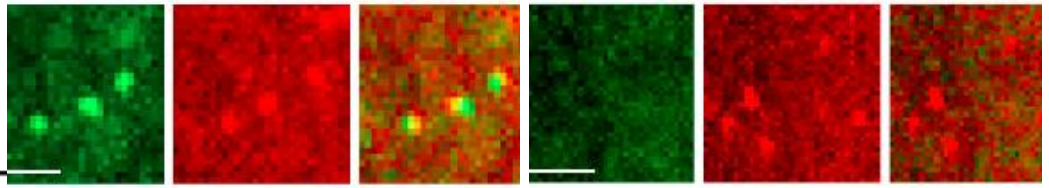


Polysome mobility analysis

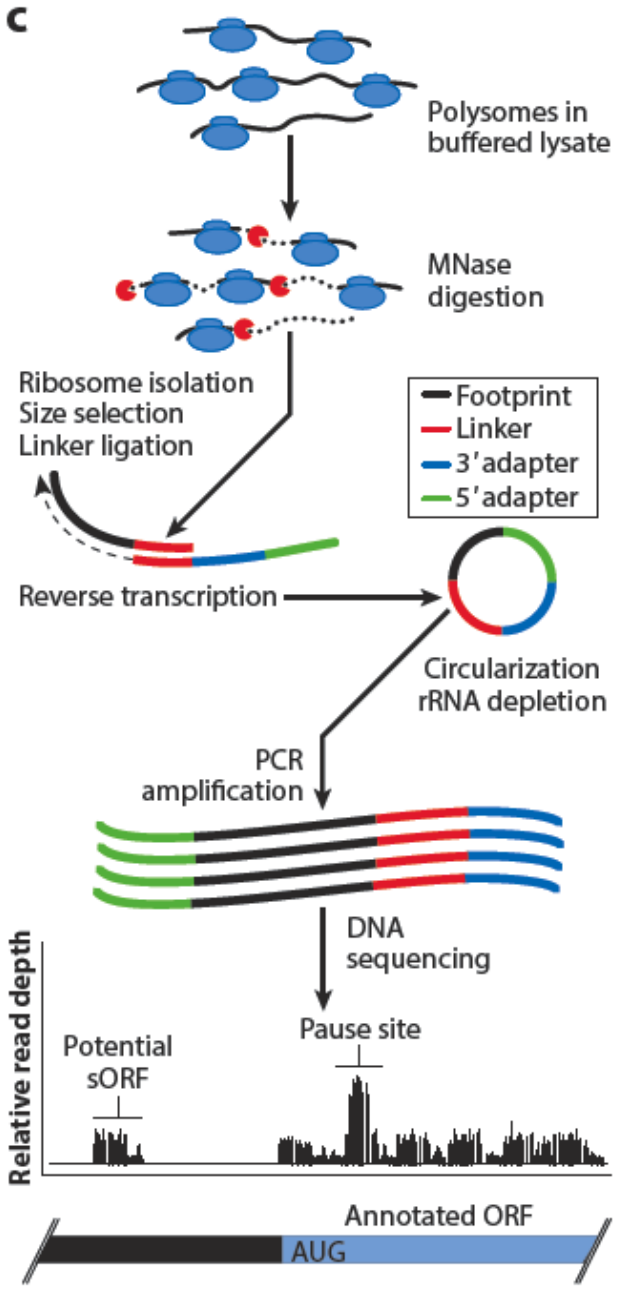




mCherry-labeled mRNA tethered to the membrane



Wang et al, Cell, 2016 Yan et al., Cell 2016



Ribosome Profiling
Ribosome Mapping
Ribo-seq

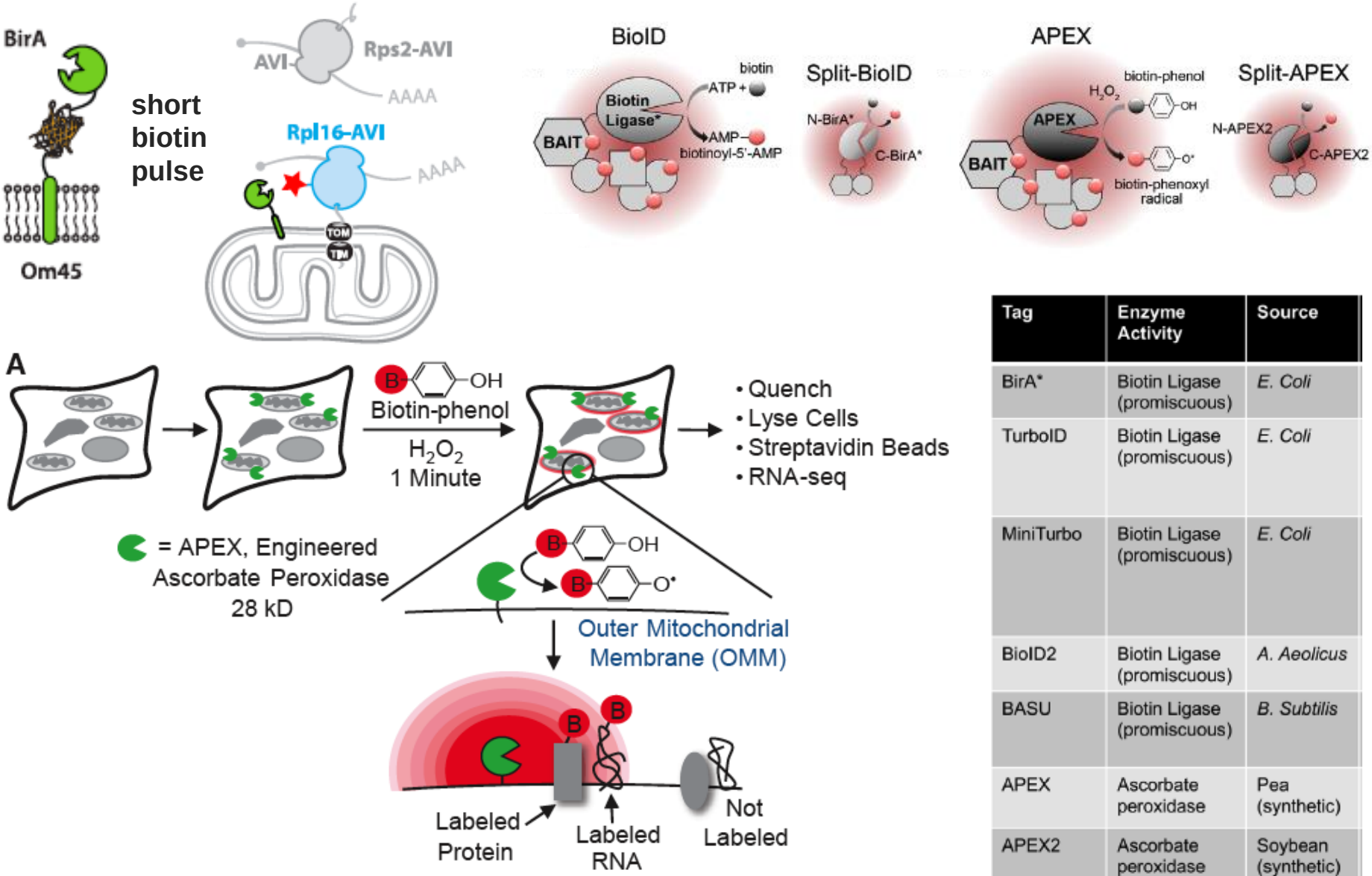
In vivo analysis of translating ribosomes and mRNAs

- transcript abundancy
- translation kinetics and efficiency/protein synthesis rate
- polysome occupancy profiles
- co-translational processes
- combined with RNASeq- correlations between transcription and translation
- short and alternative ORFs

Barquist and Vogel, *Ann Rev Genet, 2015*

APEX and BioID

Proximity biotin-based labeling methods

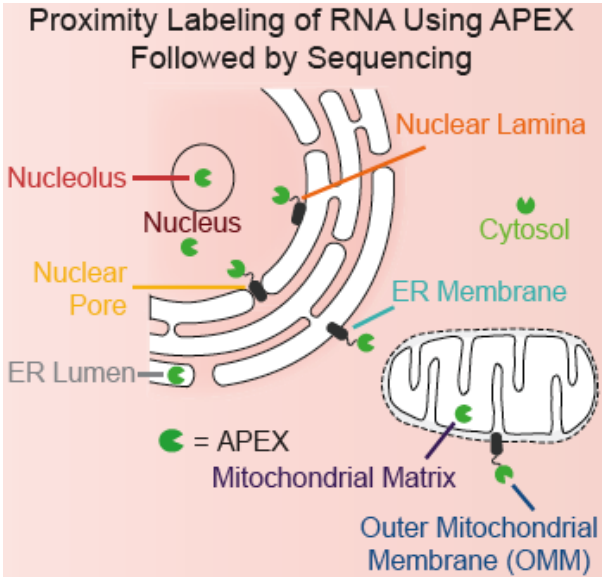


Tag	Enzyme Activity	Source
BirA*	Biotin Ligase (promiscuous)	*E. Coli*
TurboID	Biotin Ligase (promiscuous)	*E. Coli*
MiniTurbo	Biotin Ligase (promiscuous)	*E. Coli*
BioID2	Biotin Ligase (promiscuous)	*A. Aeolicus*
BASU	Biotin Ligase (promiscuous)	*B. Subtilis*
APEX	Ascorbate peroxidase	Pea (synthetic)
APEX2	Ascorbate peroxidase	Soybean (synthetic)

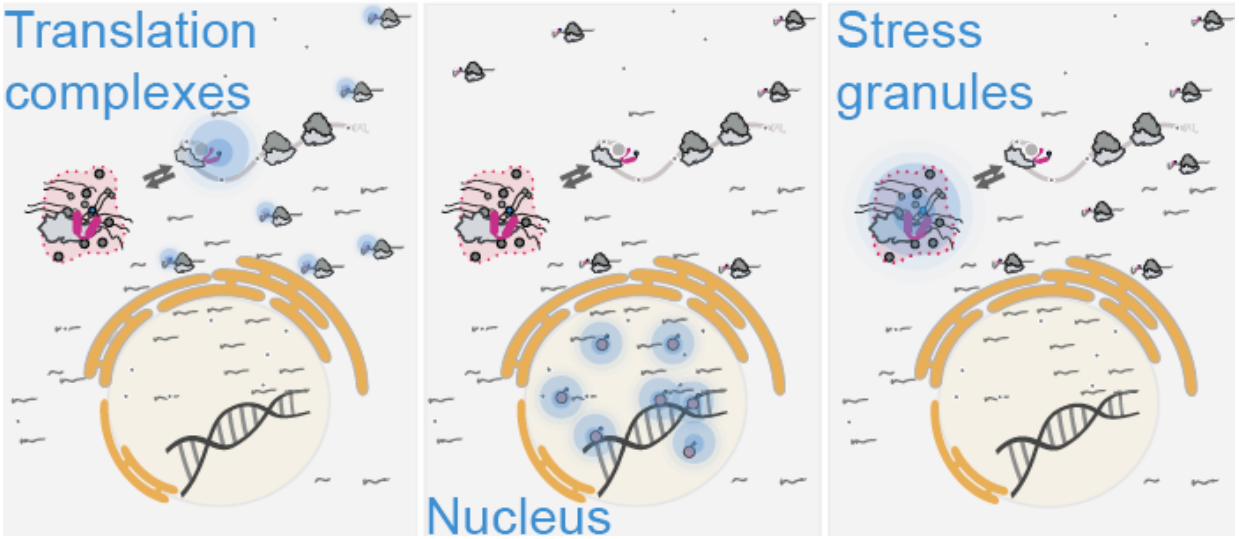
Williams et al, Science, 2014; Fazal et al, Cell, 2019; Trinkle-Mulcahy, F1000Research, 2020

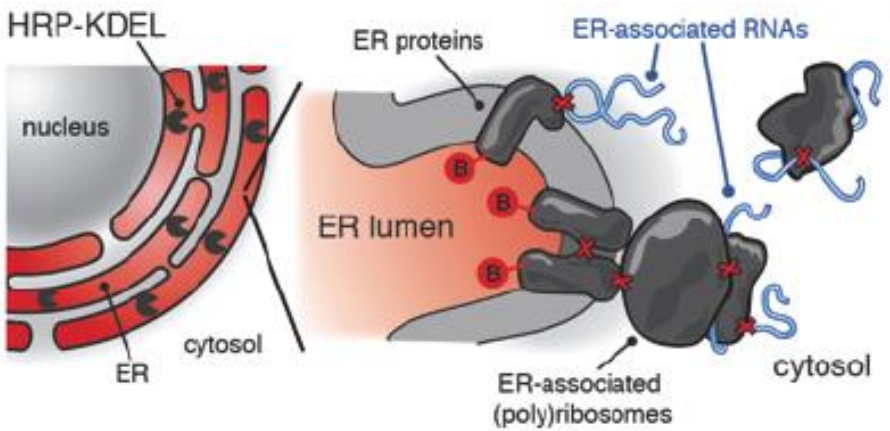
APEX: applications

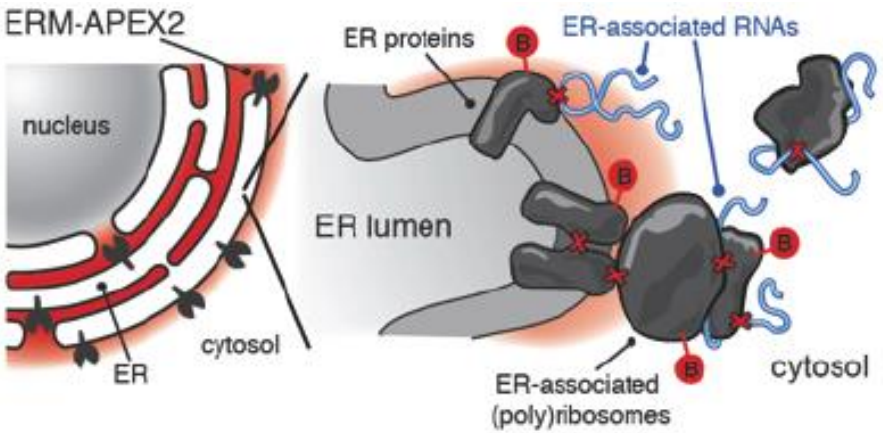
Protein and RNA localization



RNP organization- translation, granules



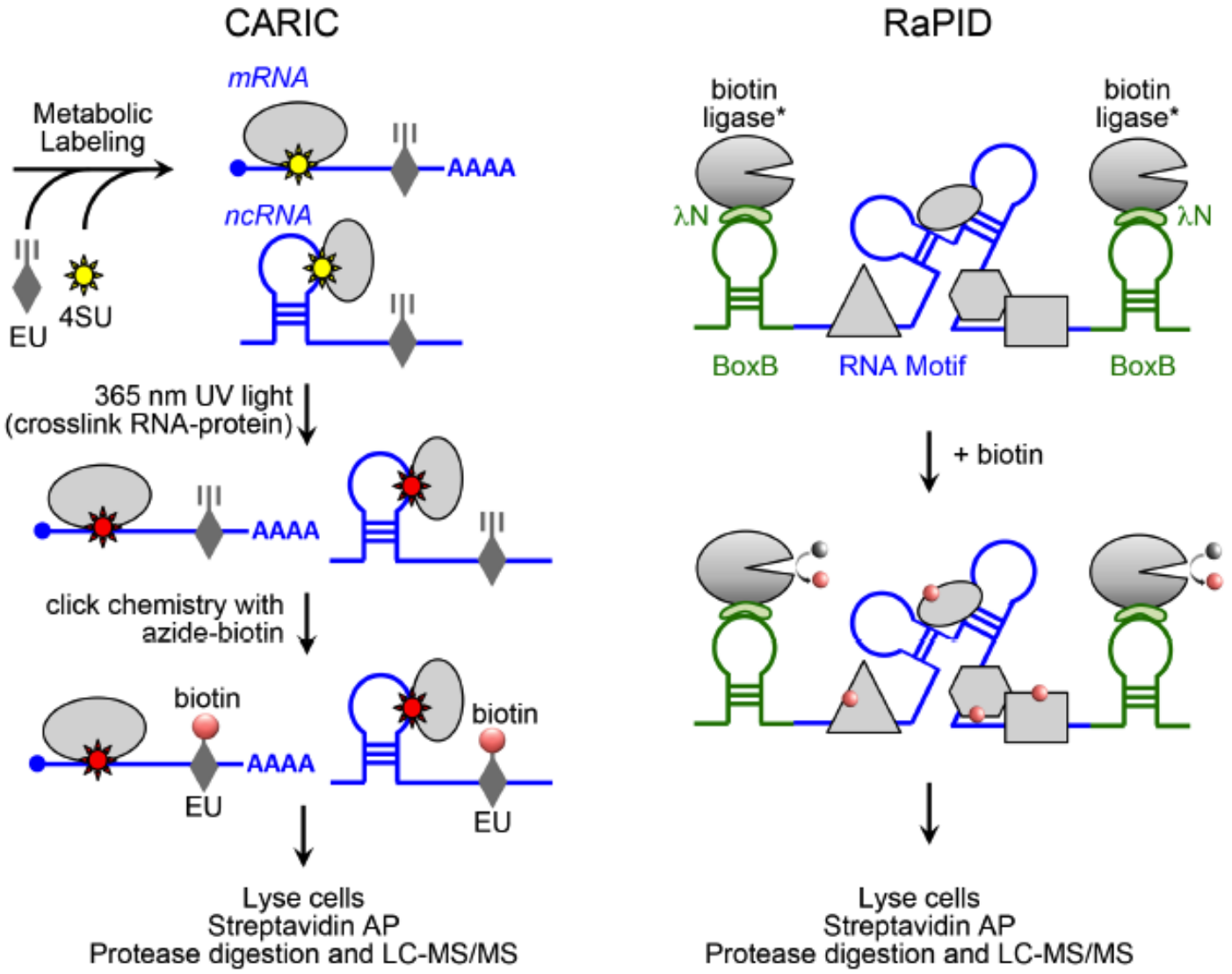




Fazal et al, Cell, 2019; Padron et al, Mol Cell, 2019

APEX: applications

Protein-RNA (RNP) complexes

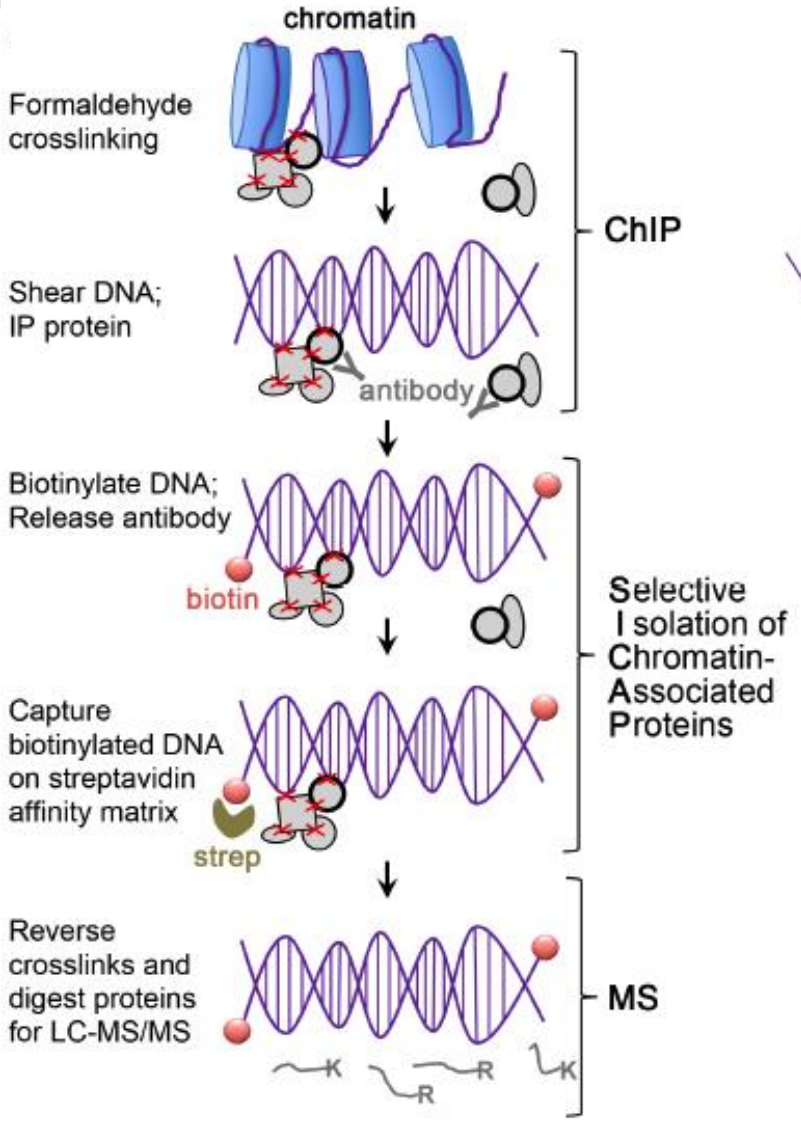


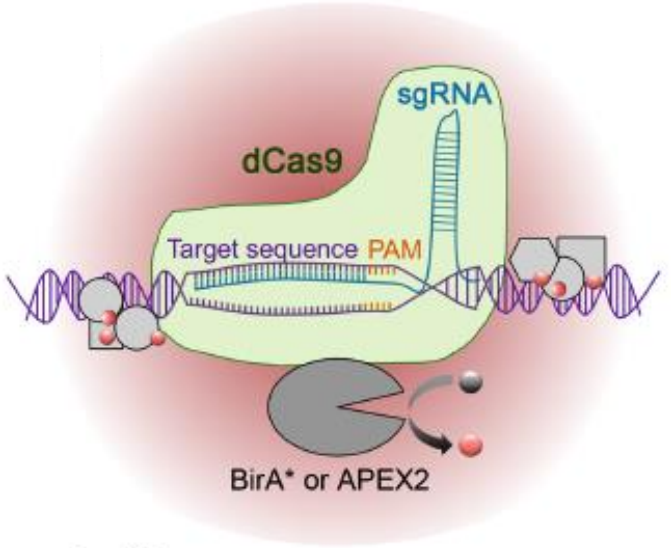
Trinkle-Mulcahy, F1000Research, 2020

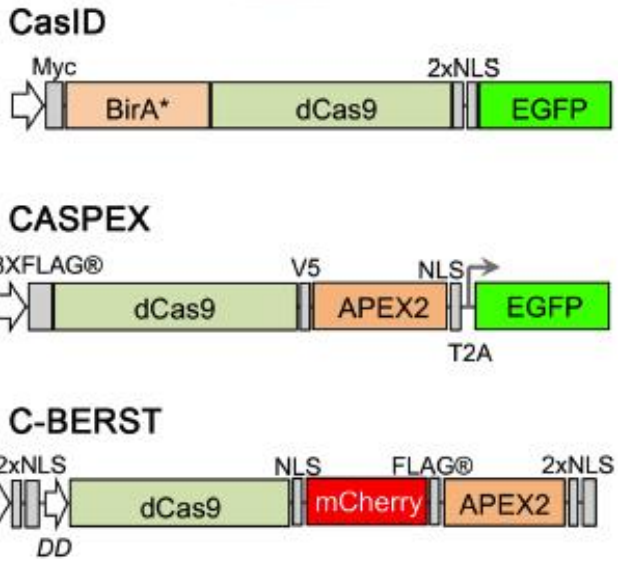
APEX: applications

Chromatin-associated protein complexes

ChIP-SICAP

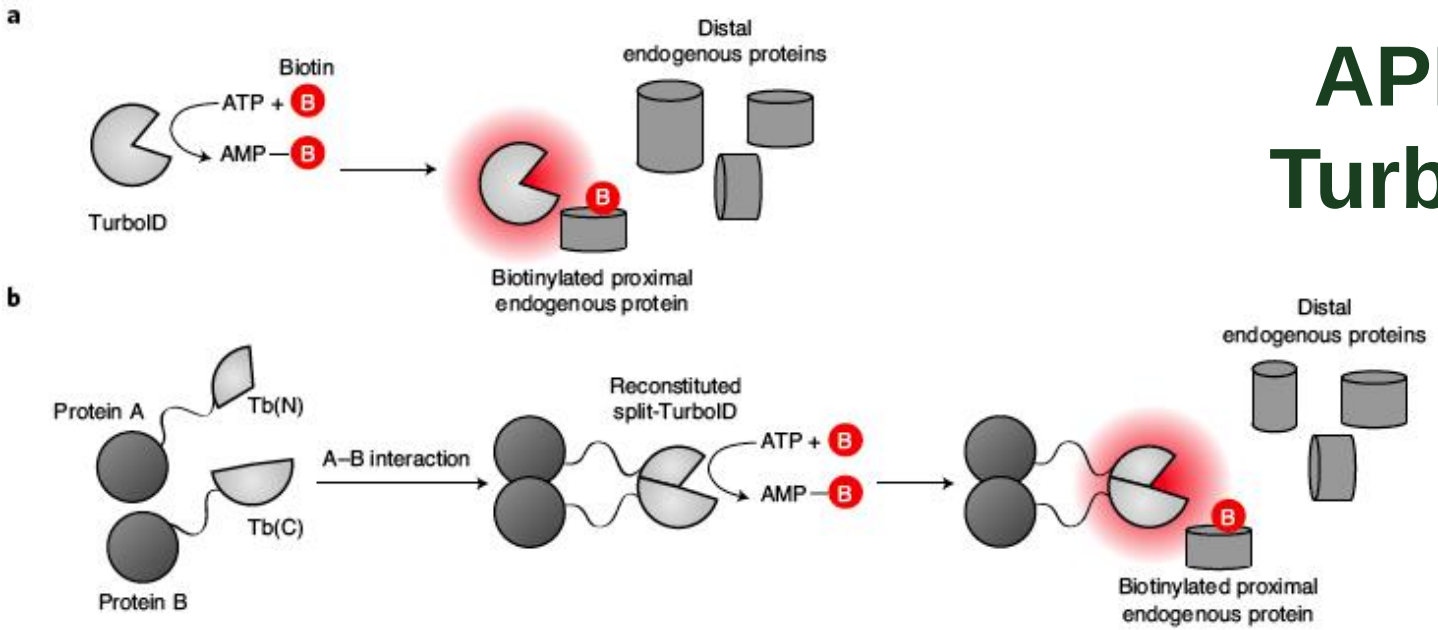


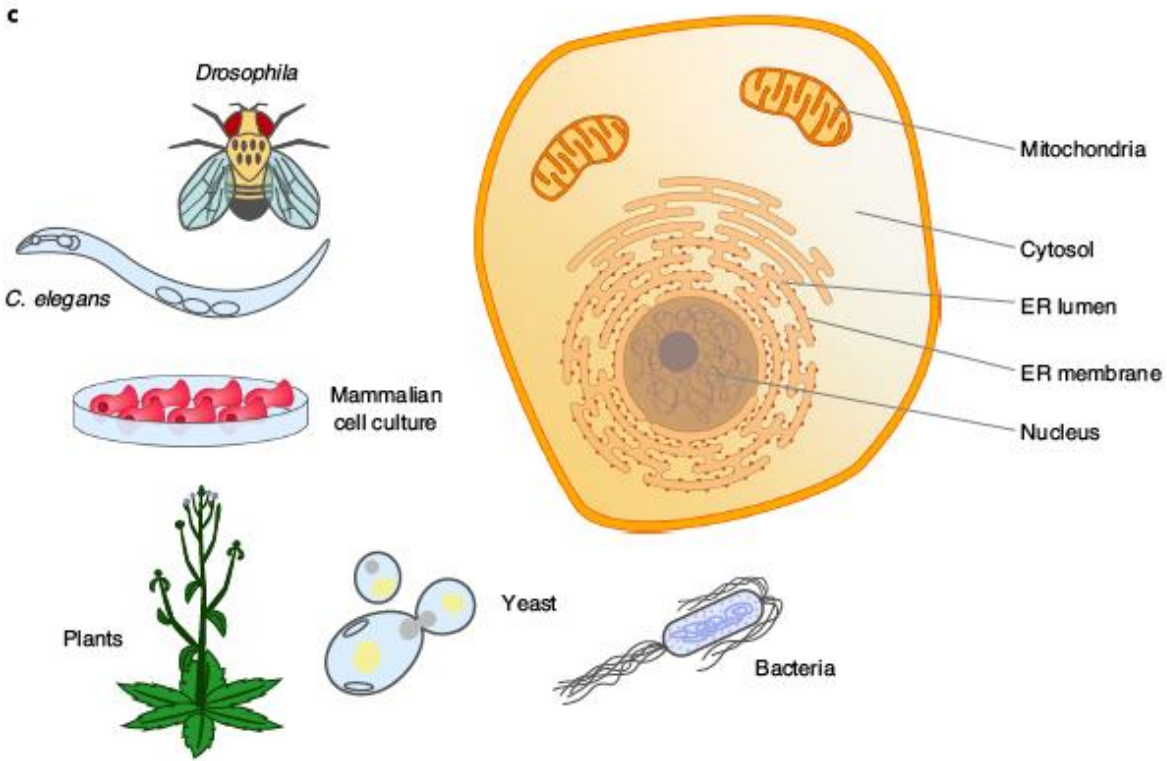




Trinkle-Mulcahy, F1000Research, 2020

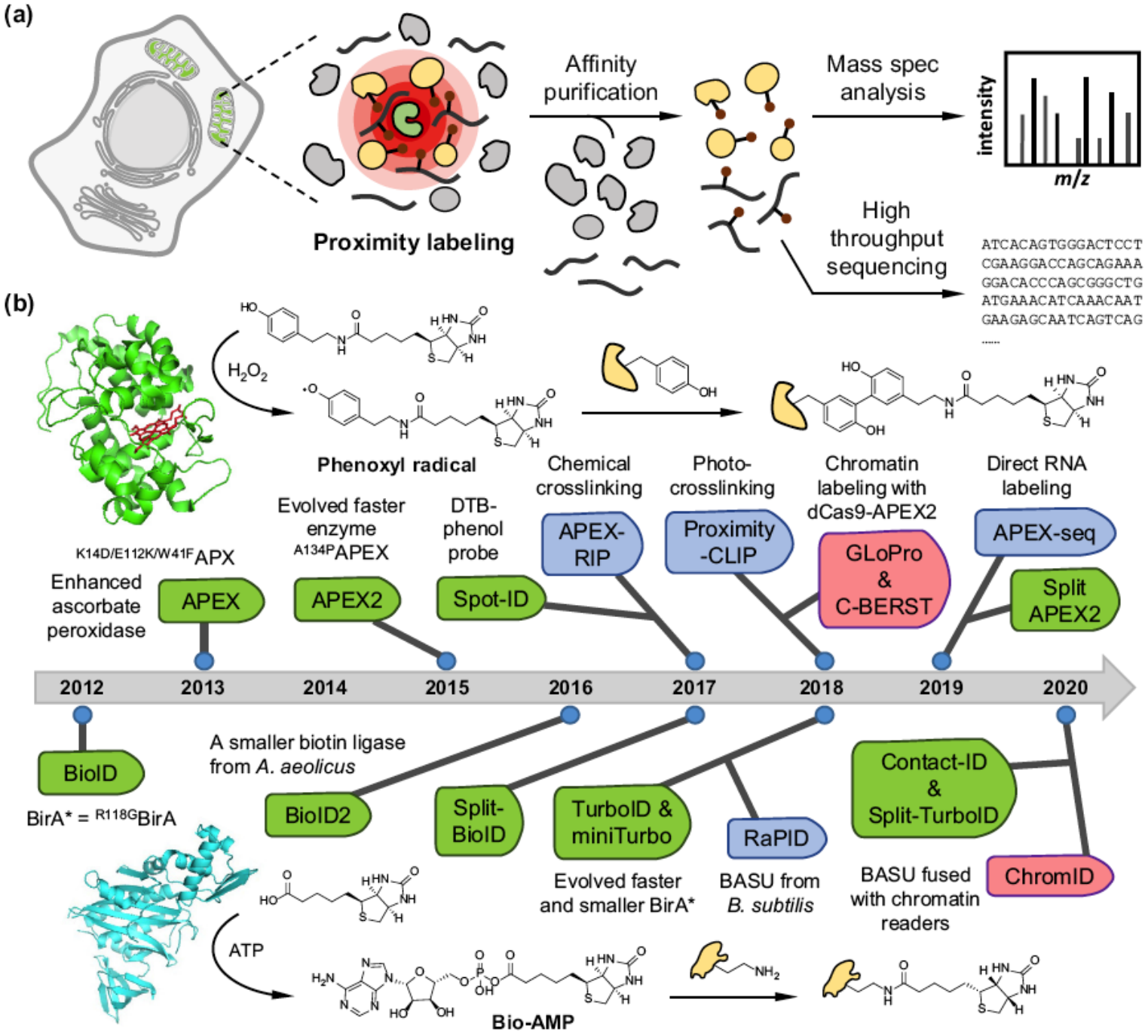
APEX
TurboID





Cho et al, Nat Protocols 2020

APEX

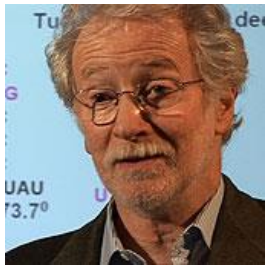


Zhou and Zou Curr Op Chem Biol 2020

SELEX = Systematic Evolution of Ligands by EXponential enrichment

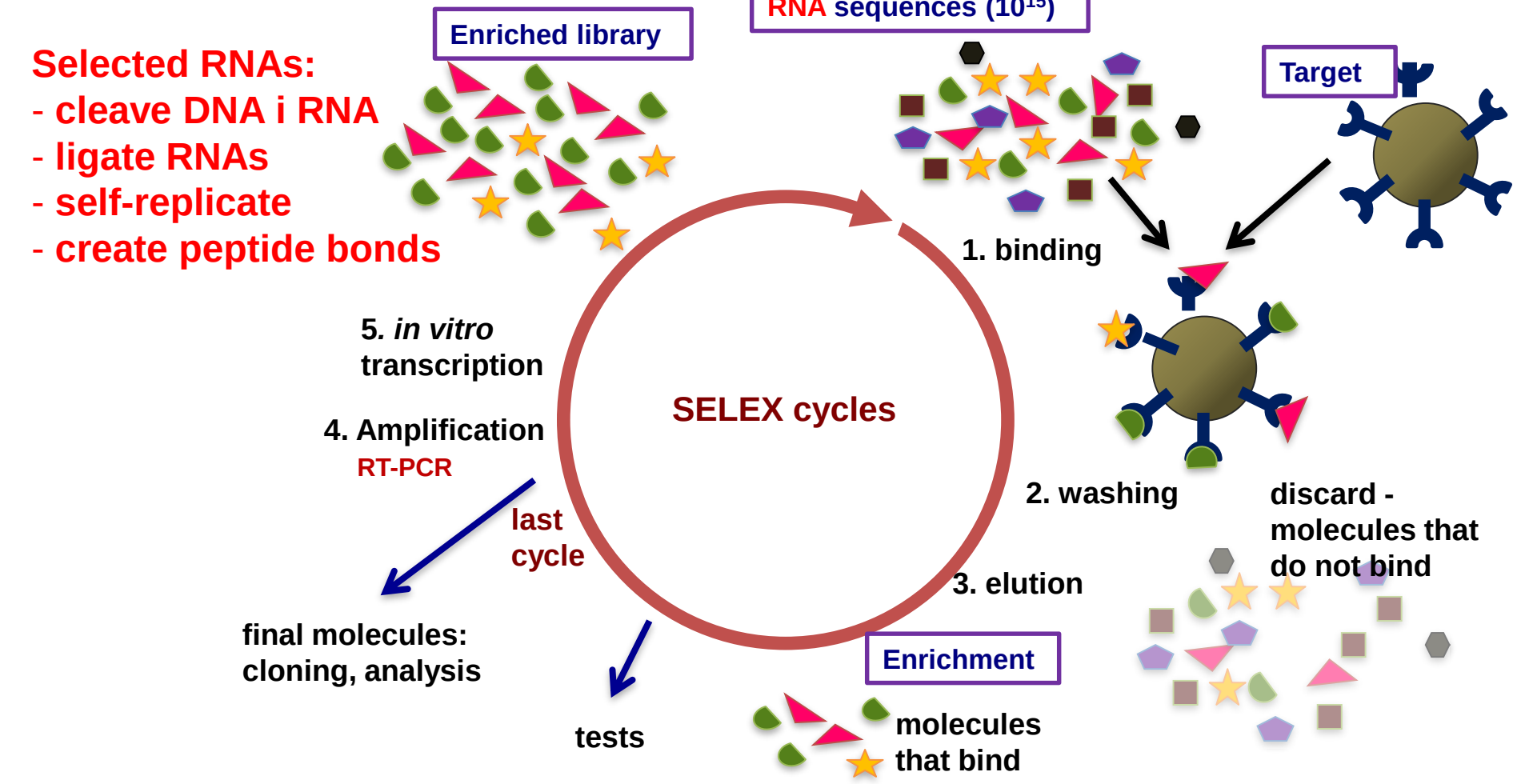
1990


Szostak


Gold

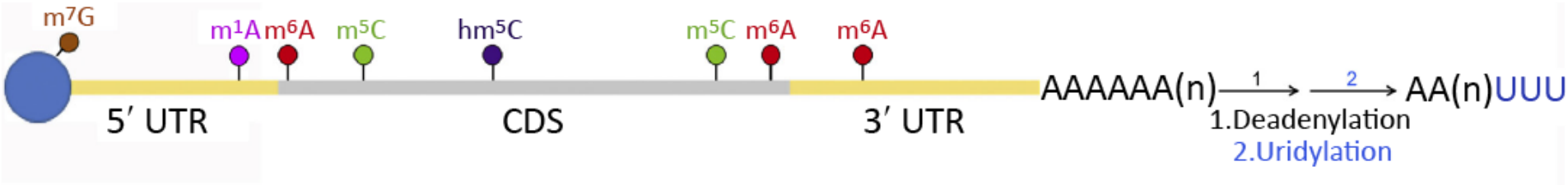
Method of selecting RNA/DNA molecules with desired properties (aptamers, ribozymes) based on cycles of amplification

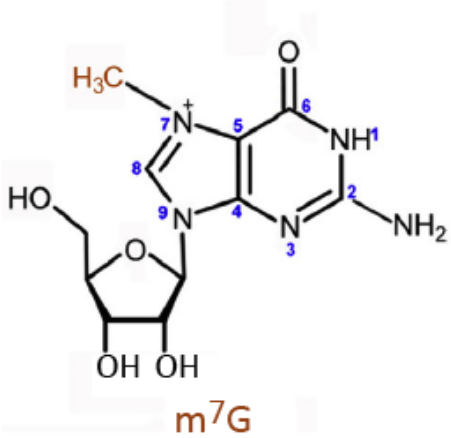
Selected RNAs:
- cleave DNA i RNA
- ligate RNAs
- self-replicate
- create peptide bonds

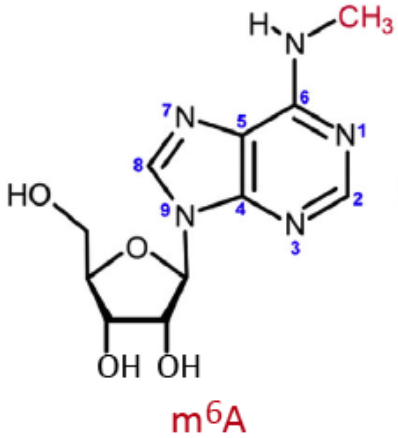


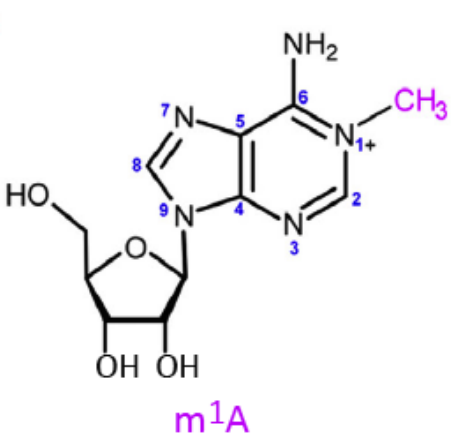
RNA MISHMASH

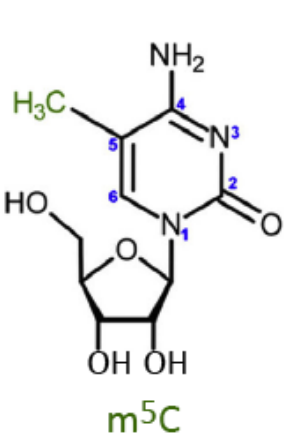
mRNA MODIFICATIONS

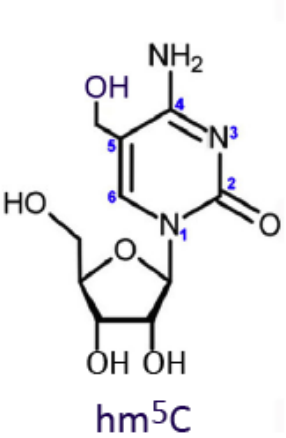




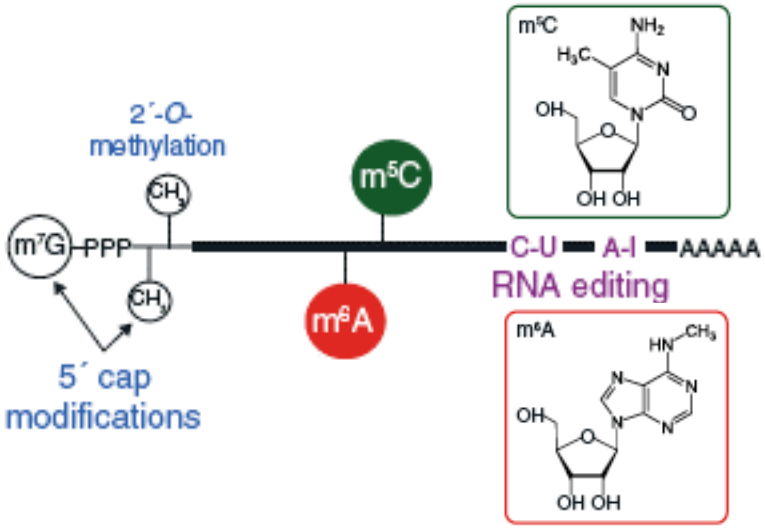


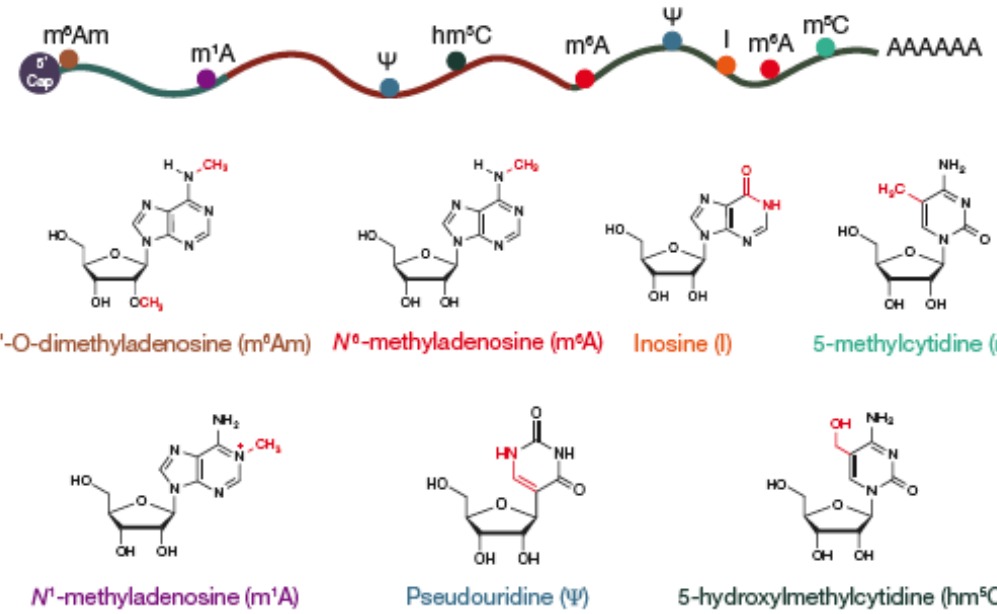




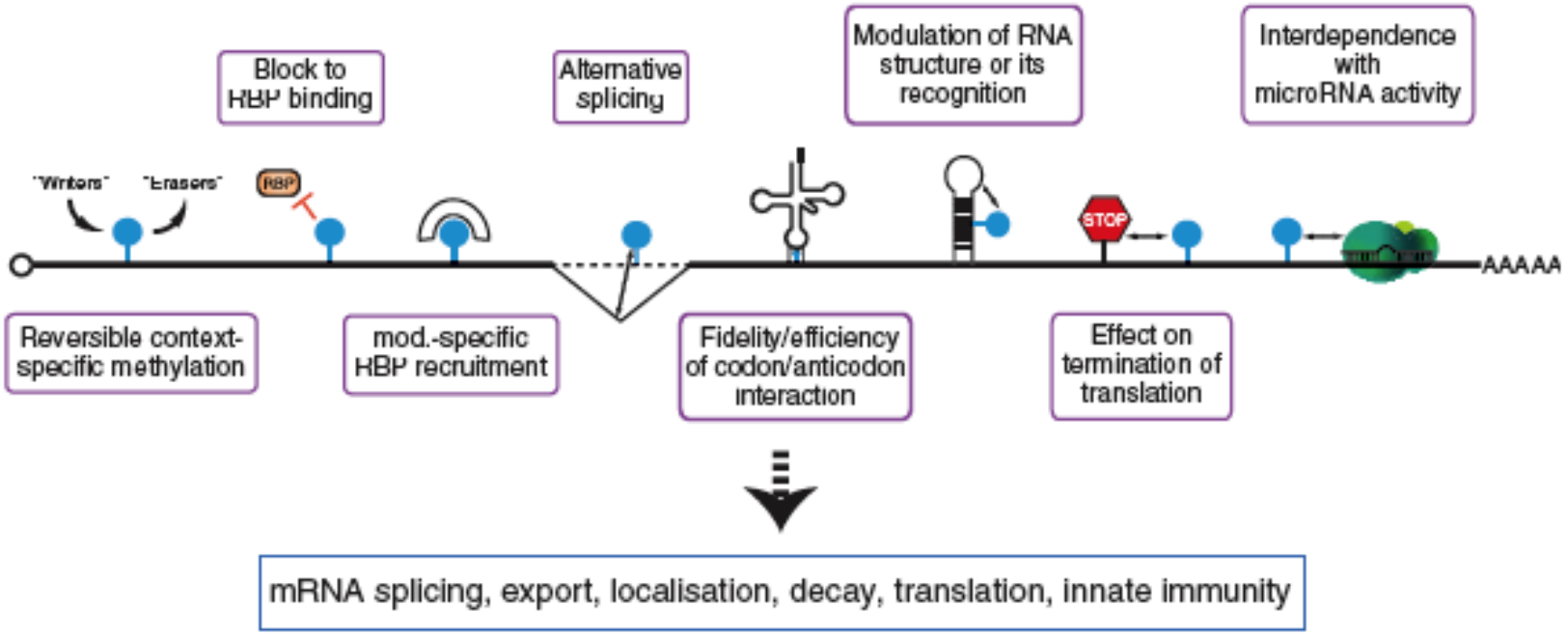


RNA MODIFICATIONS





FUNCTIONS

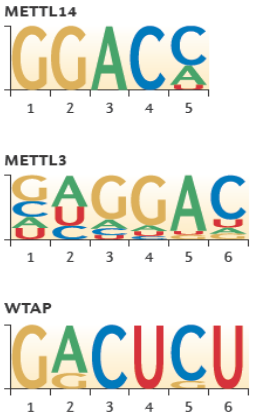


Sibbrrit et al, WIREsRNA 2013

mRNA MODIFICATIONS

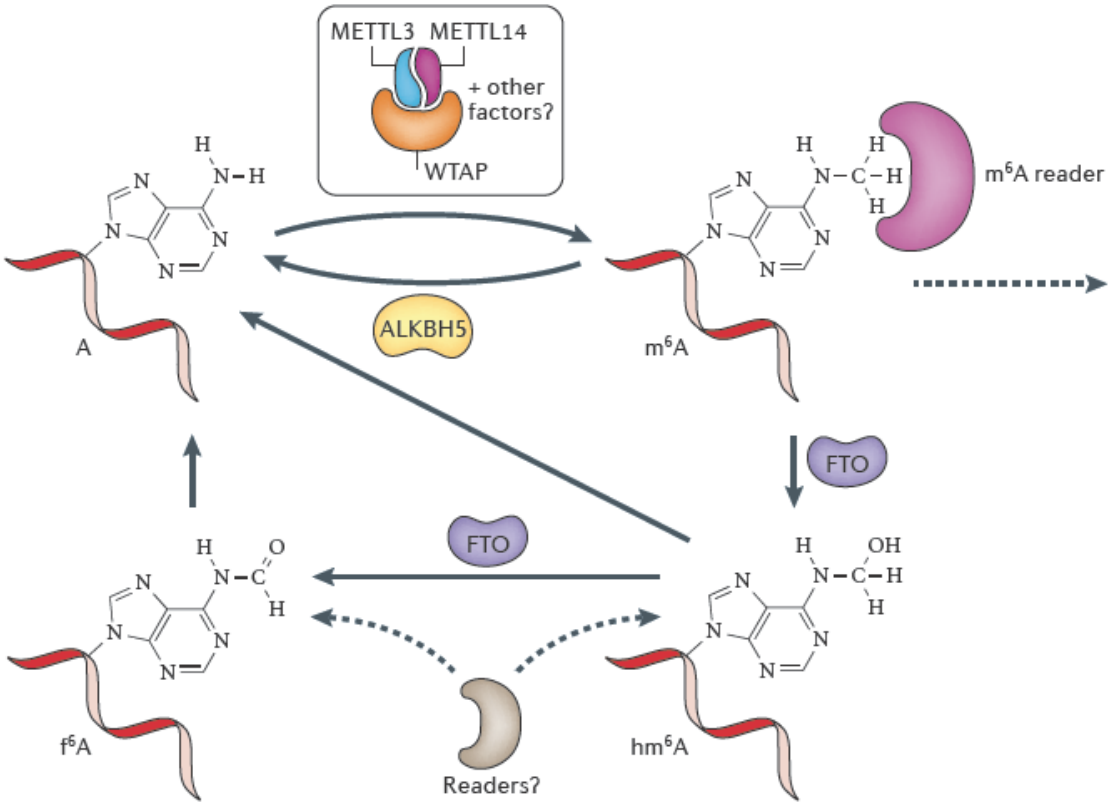
Modification	Method
m^6A	Dot blot[a]
	LC-MS/MS[a]
	TLC[a]
	SCARLET
	SELECT
	MeRIP-qPCR[a]/m^6A-IP-qPCR[a]
	m^6A-seq/MeRIP-seq[a]
	PA-m^6A-seq
	m^6A-CLIP/miCLIP
	m^6A-LAIC-seq
m^1A	LC-MS/MS[a]
	m^1A-seq
	m^1A-ID-seq
	m^1A-MAP
m^5C	Dot blot[a]
	LC-MS/MS[a]
	TLC
	Bisulfite sequencing[a]
	m^5C-RIP-seq[a]
	Aza-IP
	miCLIP
hm^5C	Dot blot
	LC-MS/MS[a]
	hMeRIP-seq

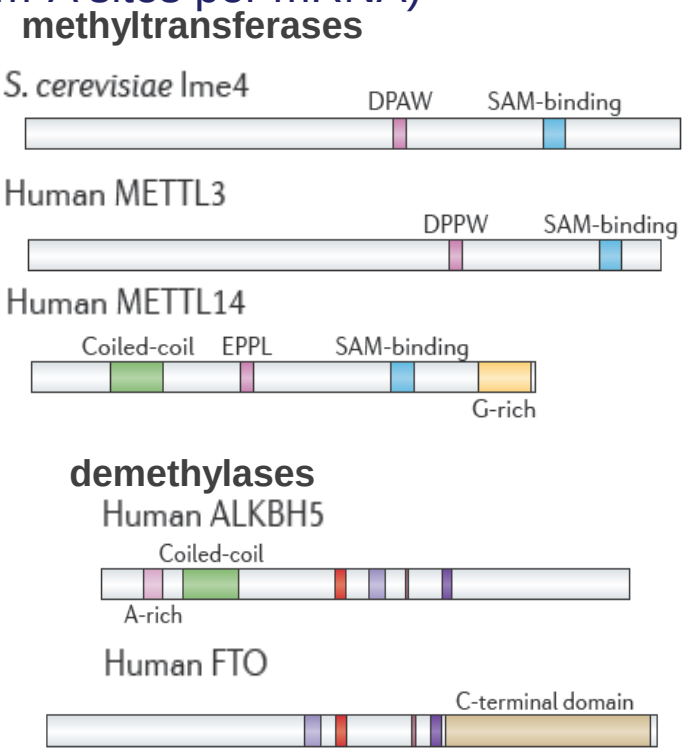
RNA MODIFICATION: mRNA m^{6}A



N^6-methyladenosine:

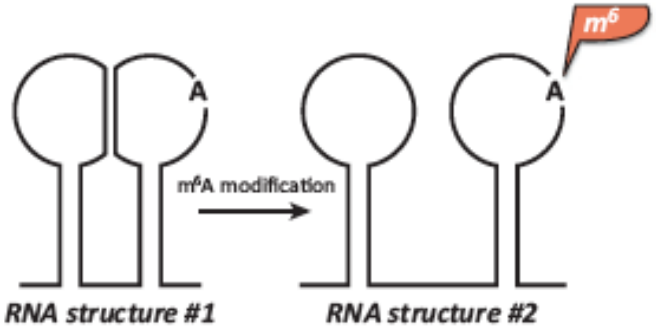
- **in eukaryotic mRNAs and lncRNAs** (discovered in 1970s)
- **reversible, conserved**
- **writers:**
  methyltransferase METTL3 or METTL4-METTL14 complex
  with **WTAP** (yeast Mum2) in a [G/A/U][G>A]m6AC[U>A>C] context
- **erasers: demethylases FTO** and **ALKBH5; readers: YTHDFXx**
- **occurrence 0.1–0.4% of As in mammals** (~3–5 m^{6}A sites per mRNA)

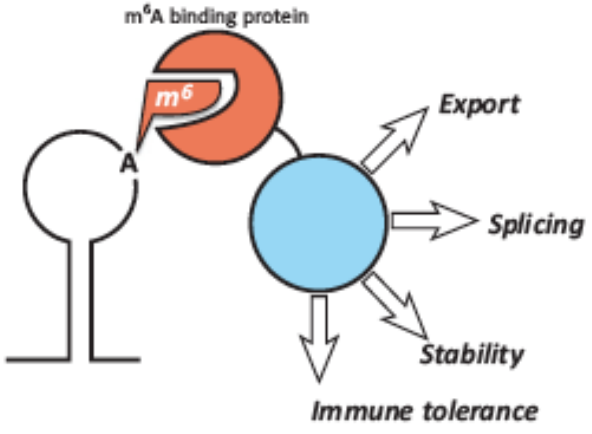




Dominissini at al, Nat.Rev.Genet., 2014

FUNCTIONS of m^{6}A

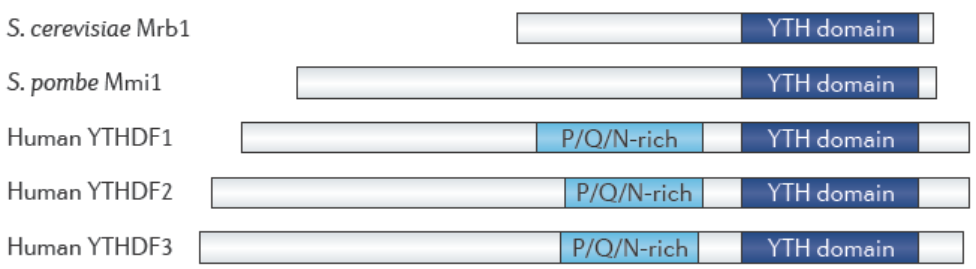


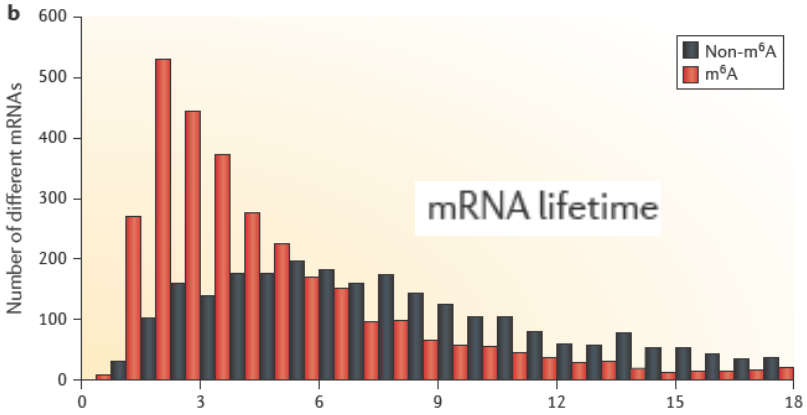


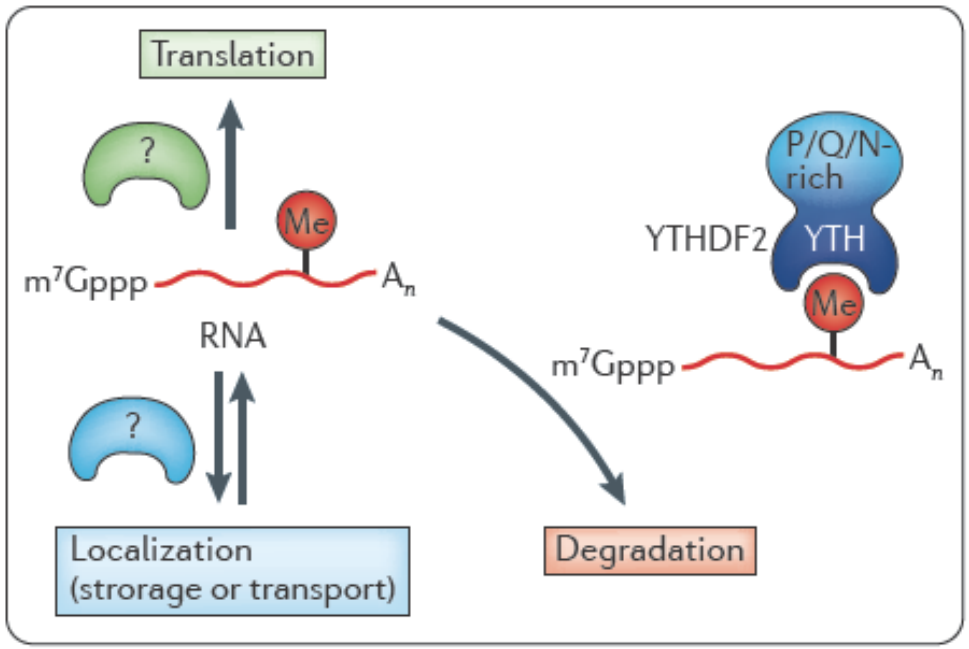
Dominissini at al, Nat.Rev.Genet., 2014;
Pan, TiBS, 2013

FUNCTIONS of mRNA m^{6}A

Readers (or anti-readers)**: YTHDF2 family preferentially recognize m^{6}A RNA**
m^{6}A can be also read by hnRNPs







- **Regulation of mRNA stability and localization**
- **circadian clock**

- inhibition of m6A leads to prolonged nuclear retention of circadian mRNAs and delays their nuclear exit

- **cell cycle**

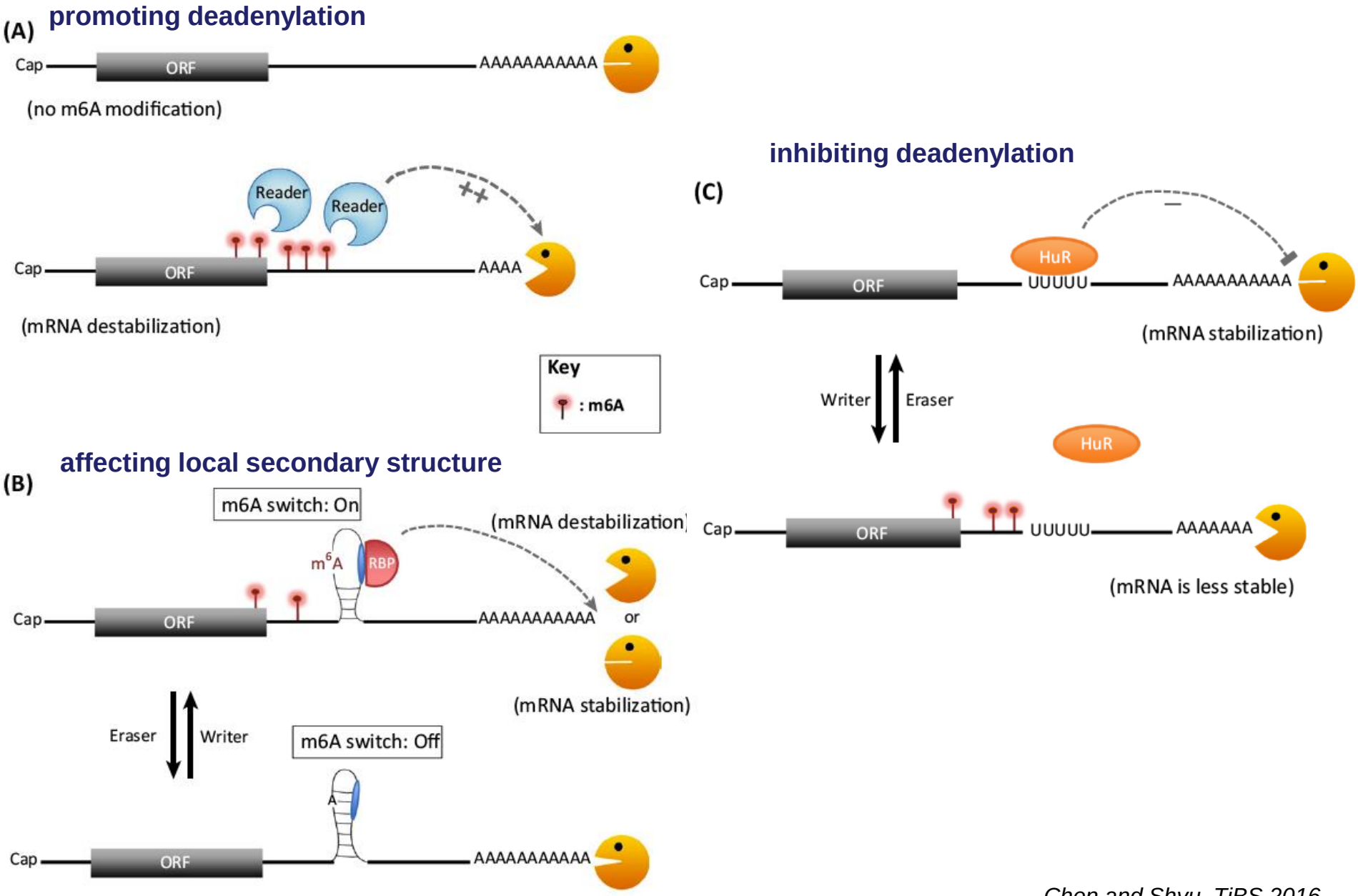
- meiosis in yeast in nitrogen starvation

- **development and differentiation**

- in embryonic stem cells (mESCs)

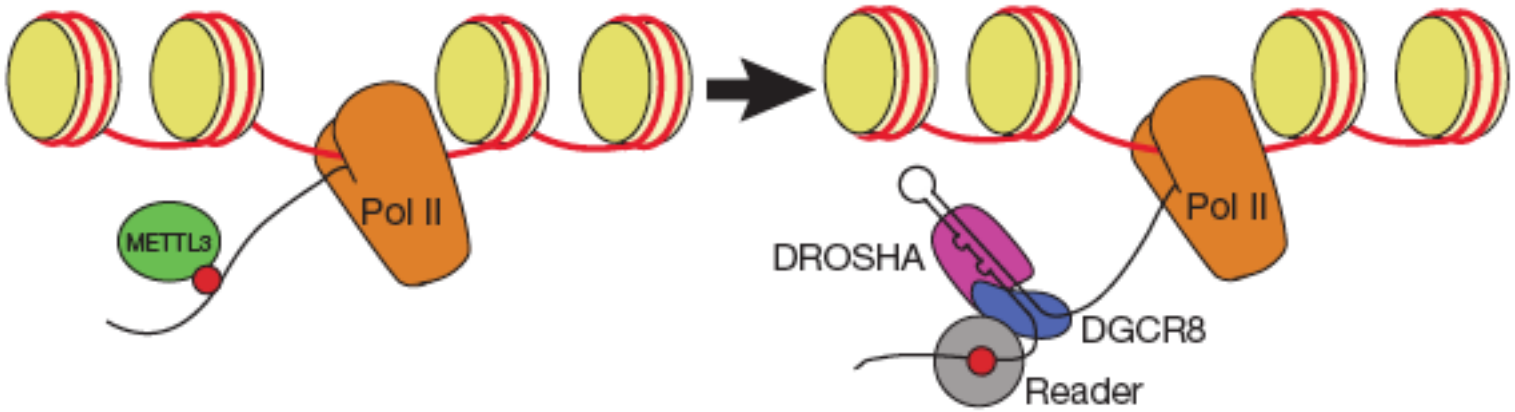
Dominissini at al, Nat.Rev.Genet., 2014

m^{6}A and mRNA STABILITY



Chen and Shyu, TiBS 2016

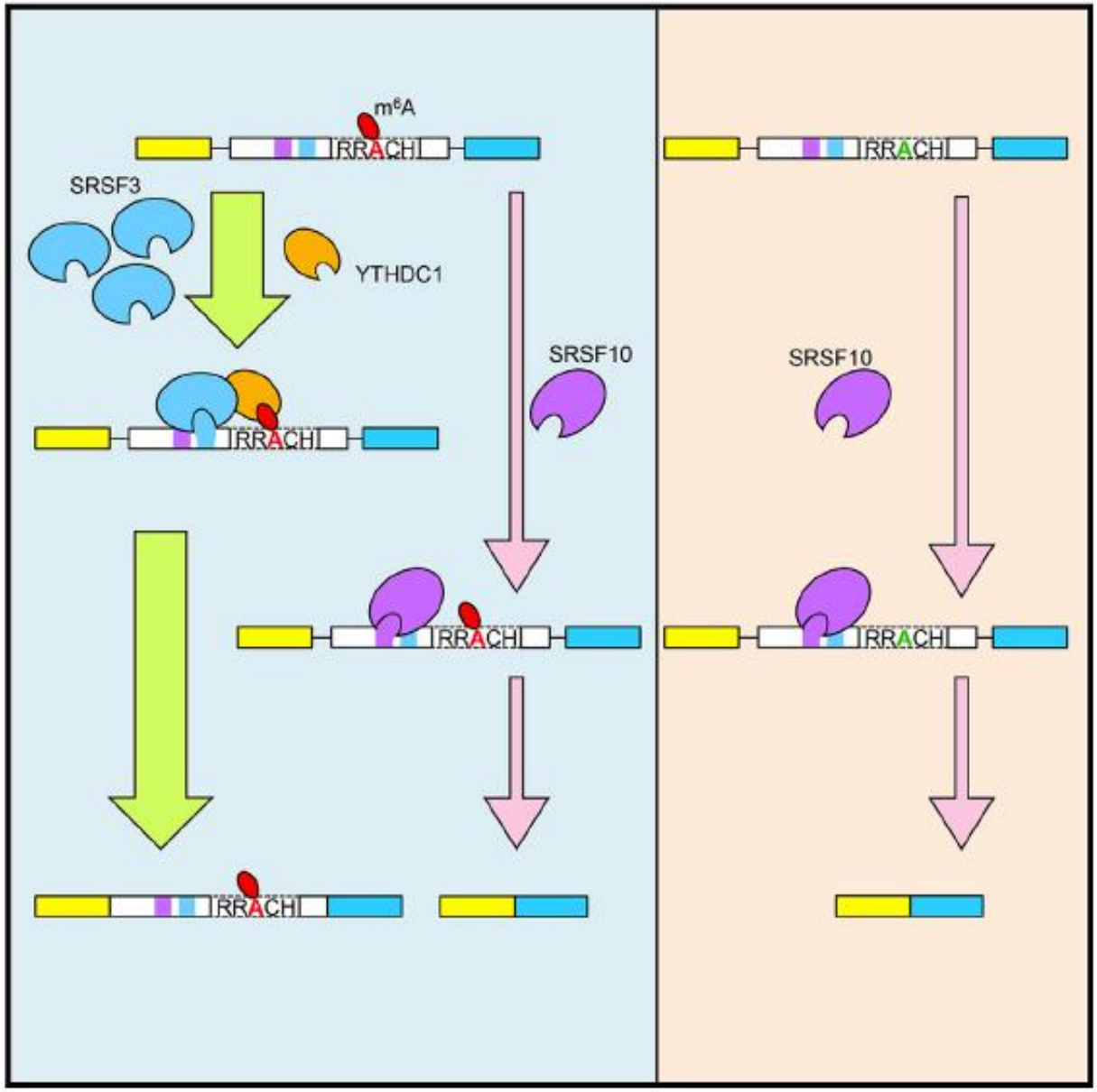
FUNCTIONS of m^{6}A: pri-miRNA PROCESSING



- m^{6}A is present in pri-miRNA regions
- METTL3 modulates miRNA expression level
- METTL3 targets pri-miRNAs for m^{6}A methylation
- m^{6}A in pri-miRNA is required for normal processing by DGCR8
- HNRNPA2B1 RNA-binding protein recognizes m^{6}A sites
- HNRNPA2B1 nuclear reader recruits Microprocessor

Alarcon at al, Nature, 2015

FUNCTIONS of m^{6}A: mRNA SPLICING



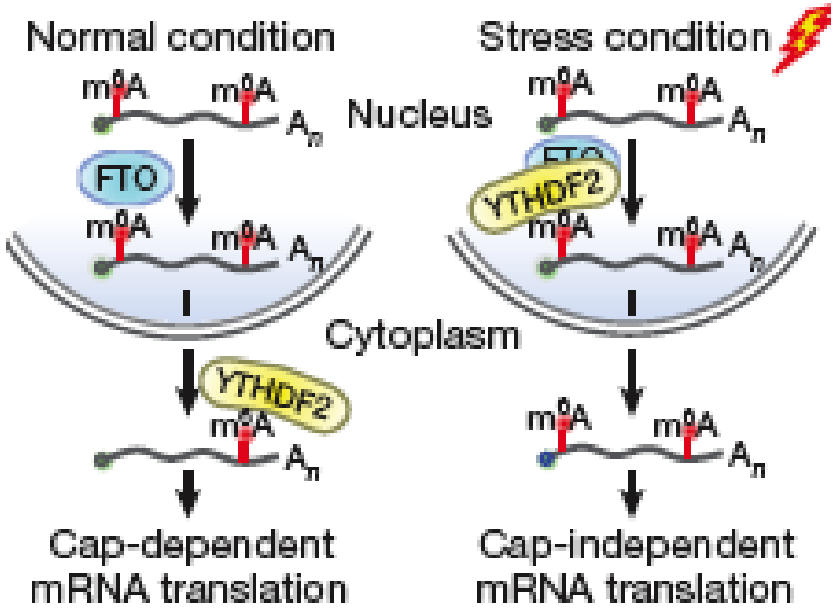
nuclear YTHDC1 m^{6}A reader

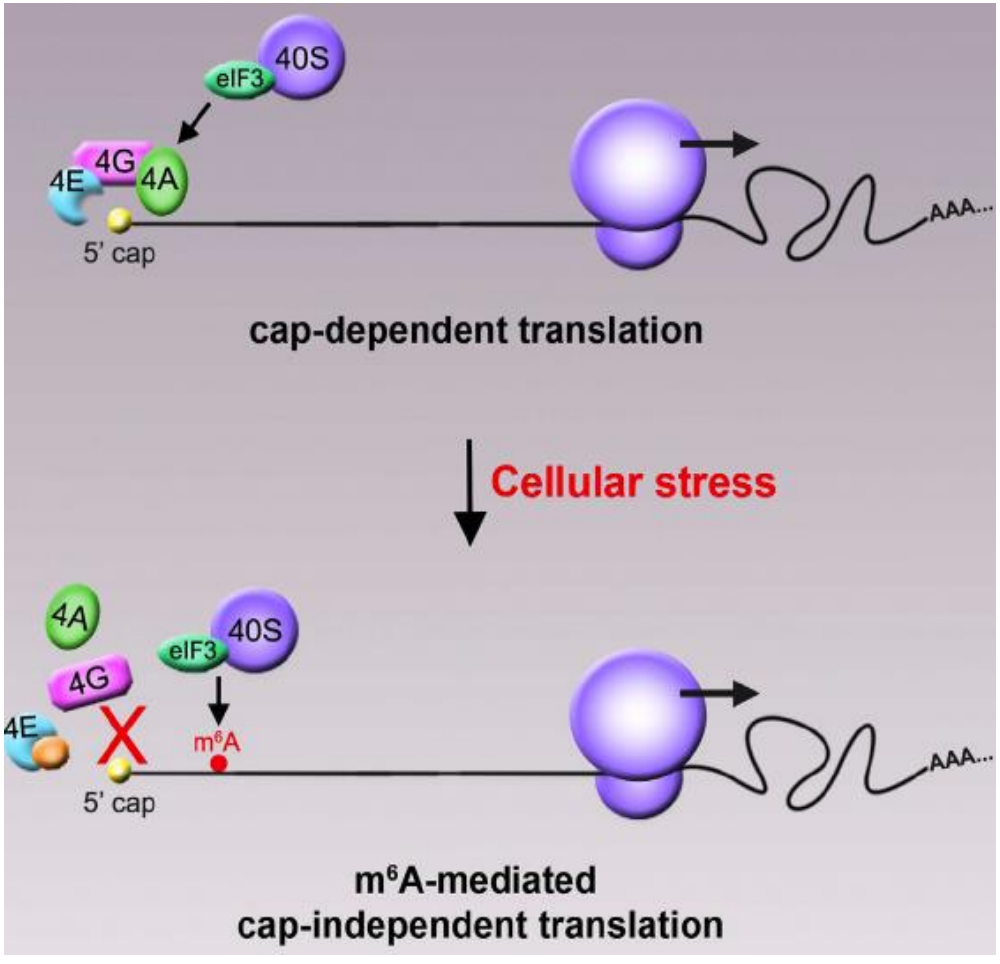
- **interacts with SR proteins SRSF3 and SRSF10**
- **facilitates/blocks binding of SRSF3/SRSF10 to pre-mRNAs**
- **promotes exon inclusion of targeted mRNAs**

Xiao et al, Mol. Cell, 2016

FUNCTIONS of m^{6}A: TRANSLATION

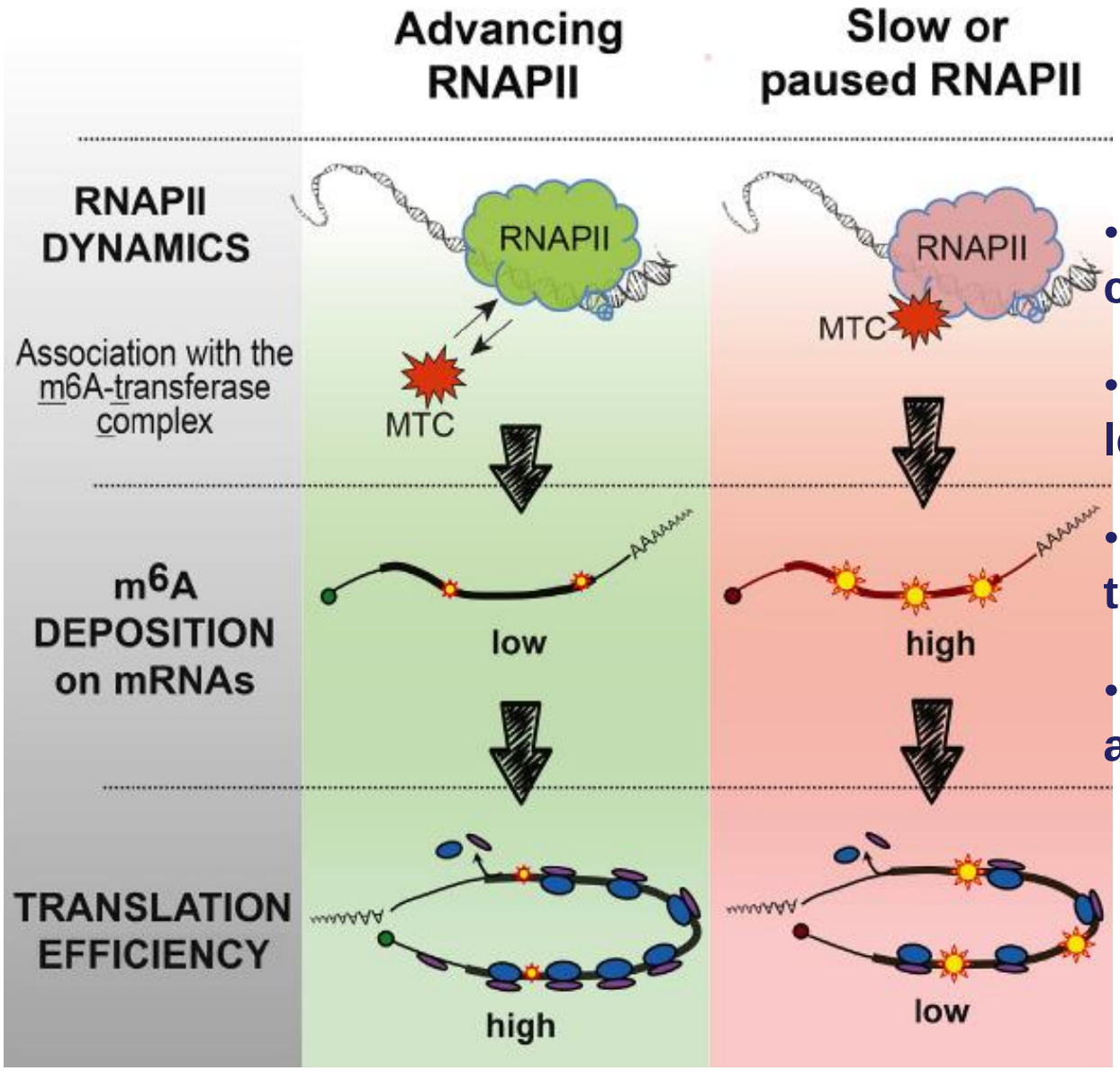
- m^{6}A in 5' UTR promotes cap-independent translation
- m^{6}A in 5' UTR upregulates translation
- cellular stresses increase m^{6}A in 5' UTRs
- YTHDF2 in heat shock induces m^{6}A-dependent translation of HS mRNAs
- m^{6}A in mRNA body disrupts tRNA selection and translation elongation dynamics
- m^{6}A in 3' UTR or near stop codon contributes to alternative polyadenylation





Meyer et al, Cell, 2015;
Zhou et al, Nature, 2015;
Choi et al, Nat. Struct. Mol. Biol.'16

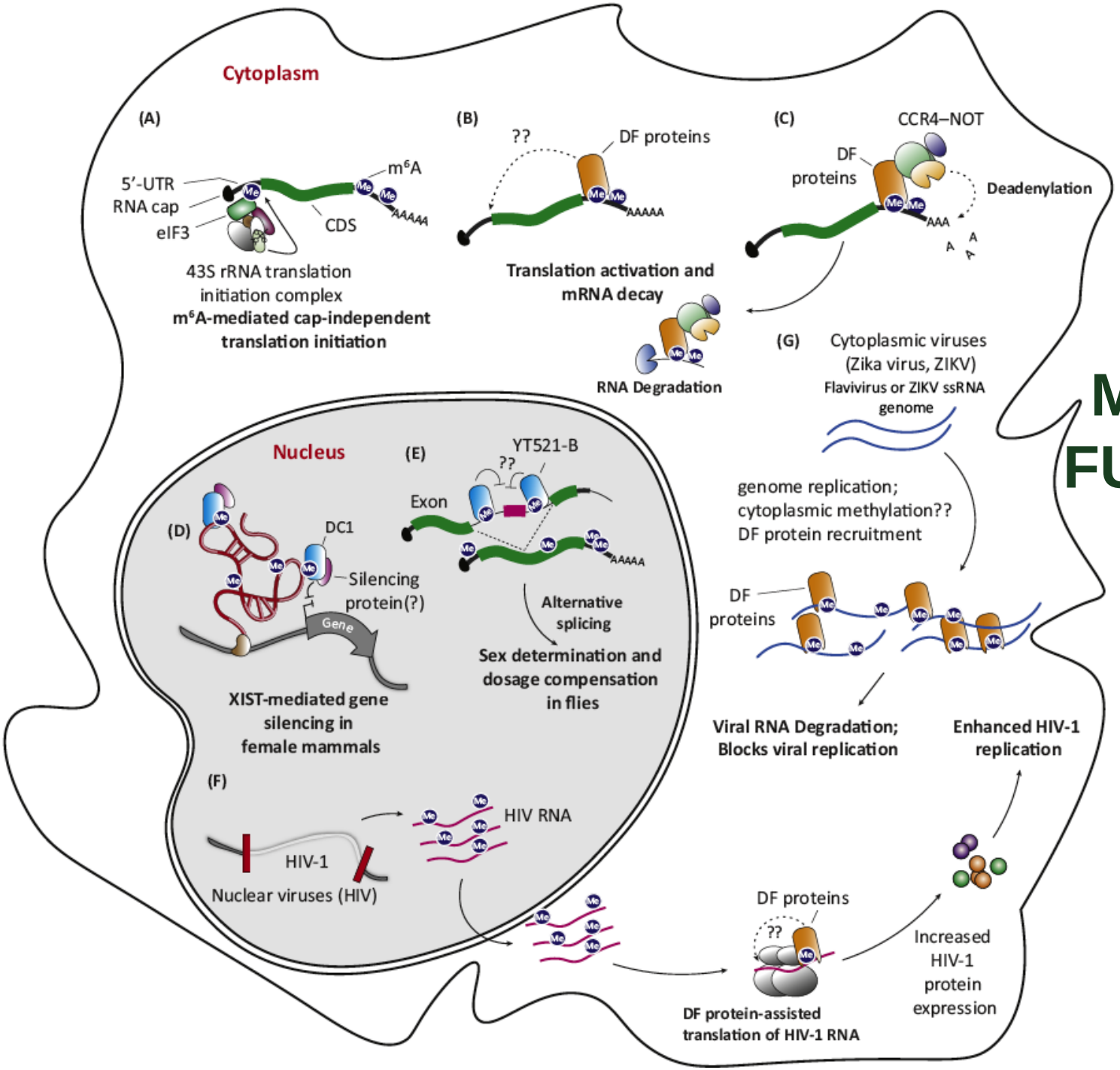
FUNCTIONS of m^{6}A: RNAPII and TRANSLATION



- **mRNA transcription rates correlate with translation**
- **slow PolII results in higher level of m^{6}A in mRNAs**
- **high level m^{6}A reduces translation rate**
- **nuclear control on protein abundance**

Slobodin et al, Cell, 2017

m⁶A MULTIPLE FUNCTIONS

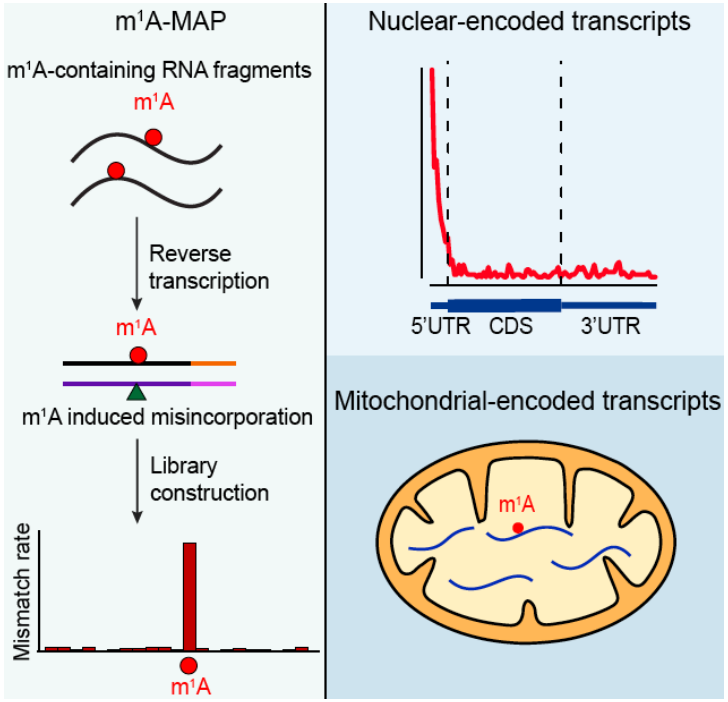


Patil et al, 2017, TiCB

RNA MODIFICATION: mRNA m^1A

N^1-methyladenosine m^1A:

- **in eukaryotic mRNAs** (from yeasts to mammals)
- **modified by TRMT6/TRMT61A** (nuclear) **or TRMT61B, TRMT10C** (mitochondrial)



Dominissini et al, Nature 2017

- **widespread** (20% in humans)
- **enriched around the start codon upstream of the first splice site**
- **preferentially in more structured regions around translation initiation sites**
- **is dynamic in response to different conditions**
- **promotes translation**

Safra et al, Nature 2017

- **in cytosol low in few mRNAs**
- **in tRNA T-loop like structures**
- **present also in mitochondria**
- **leads to translational repression**
- **is disruptive to W-C basepairing**
- **generally avoided by cells**

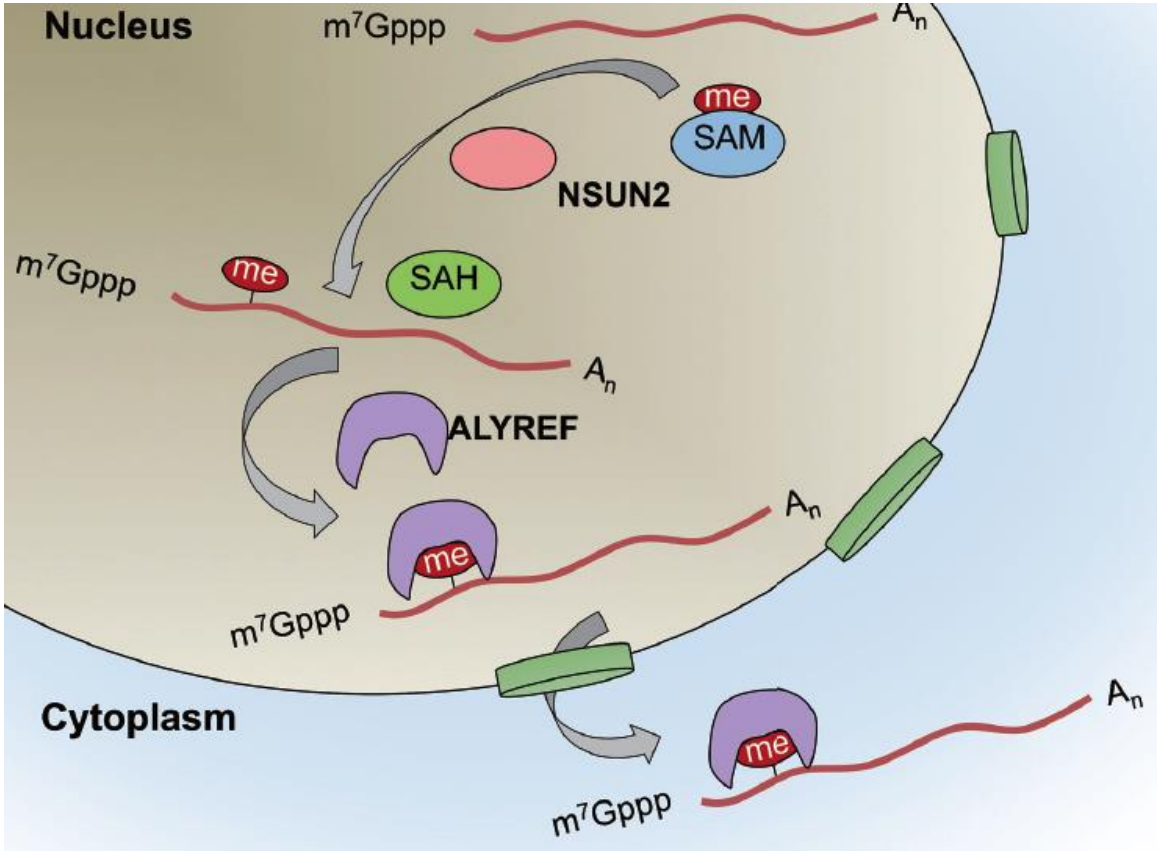
Li at al, Mol Cell 2017

- **at mRNA cap and 5' UTR increases translation**
- **prevalent in mitochondrial-encoded transcripts inhibits translation**
- **in different mRNA regions differentially impacts translation**

FUNCTIONS of m^5C

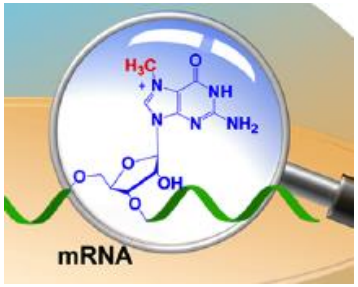
- synthesized by TRM4B methyltransferase
- enriched in the CG context and in downstream of translation initiation sites
- present in mRNAs and ncRNAs, also tRNAs
- tissue specific, acts as a epitranscriptome marker

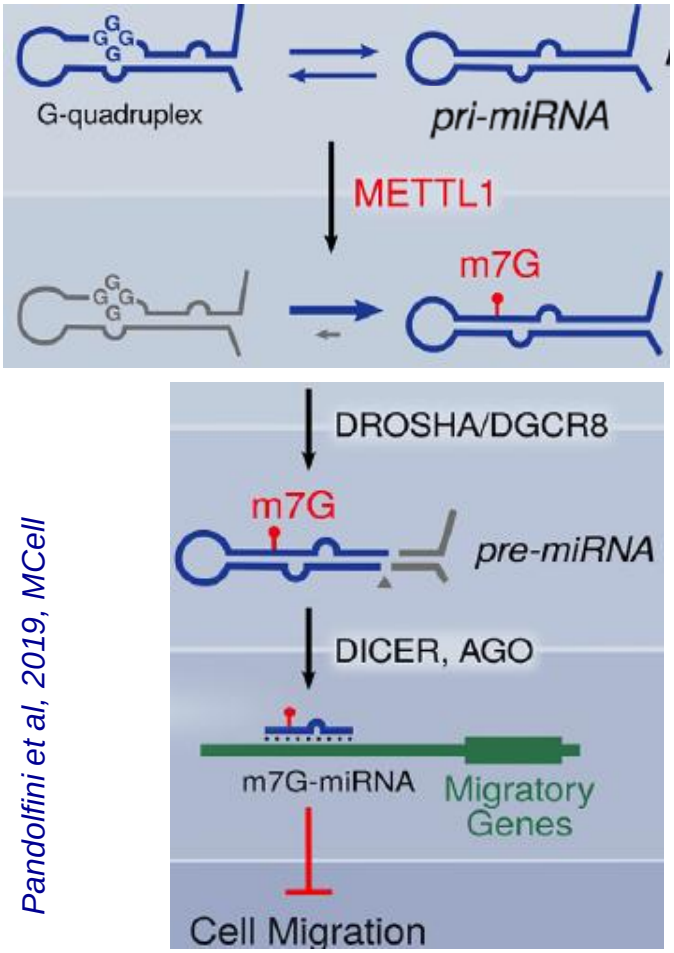
- mRNA export: NSUN2 as the methyltransferase and ALYREF as an m^5C reader



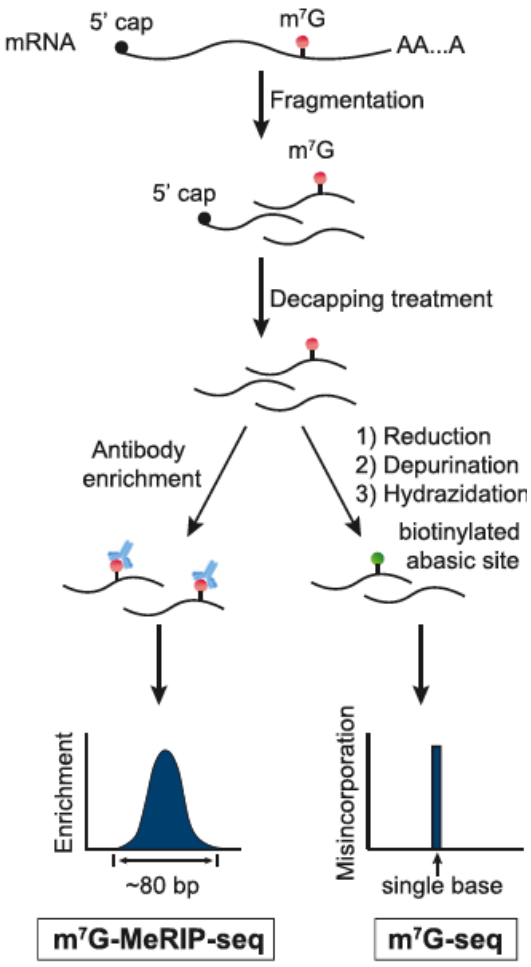
Yang et al, 2017, Cell Res

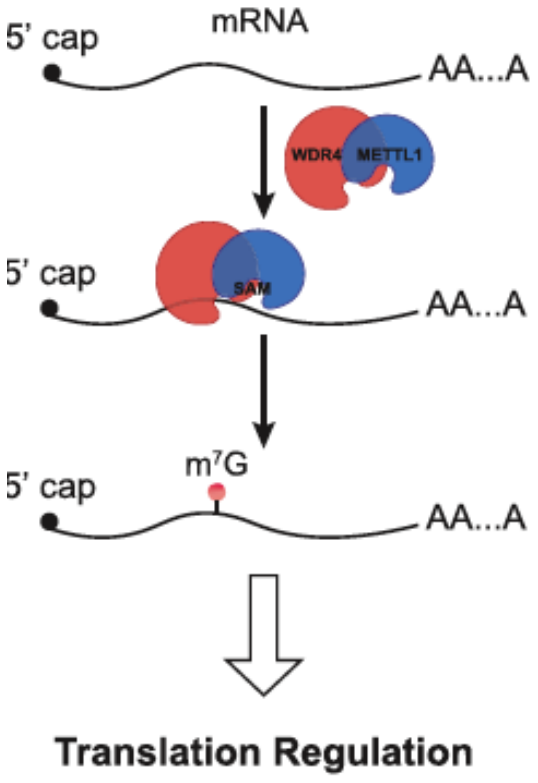
RNA MODIFICATION: internal RNA m^{7}G





Pandolfini et al, 2019, MCell



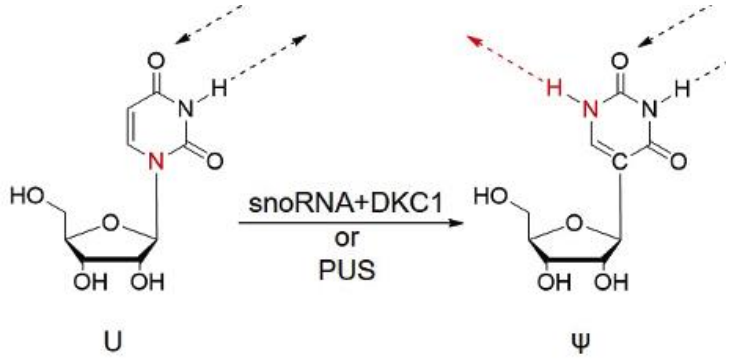


- **specific miRNAs are m^{7}G-modified by METTL1**
- **m^{7}G promotes miRNA processing by antagonizing G4 in pre-miRNAs (G4 inhibit pre-miRNA processing)**

- **human mRNAs contain internal m^{7}Gs**
- **some m^{7}Gs are introduced by METTL1**
- **internal m^{7}G affect translation**

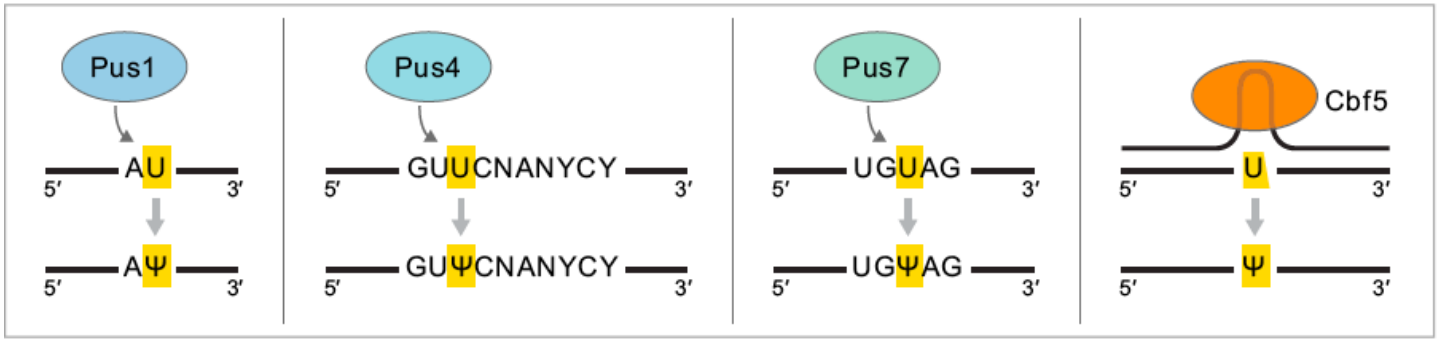
Zhang et al, 2019, MCell

RNA MODIFICATION: mRNA, lncRNA pseudoU

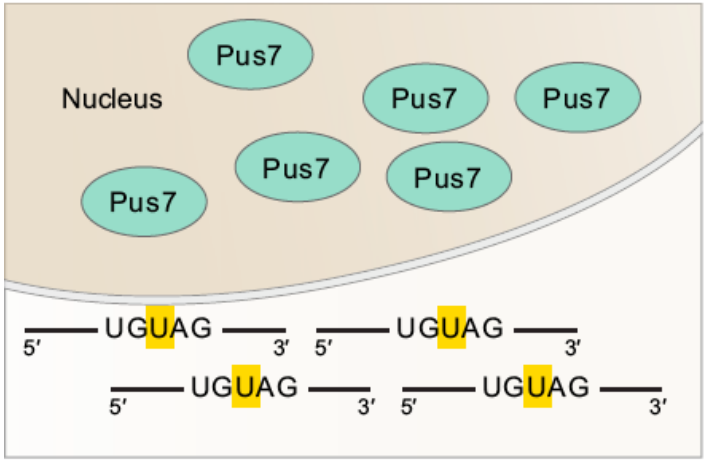


- changes mRNA structure
- facilitates recoding by the ribosome
- stimulates translation
- when added cotranscriptionally - affects splicing

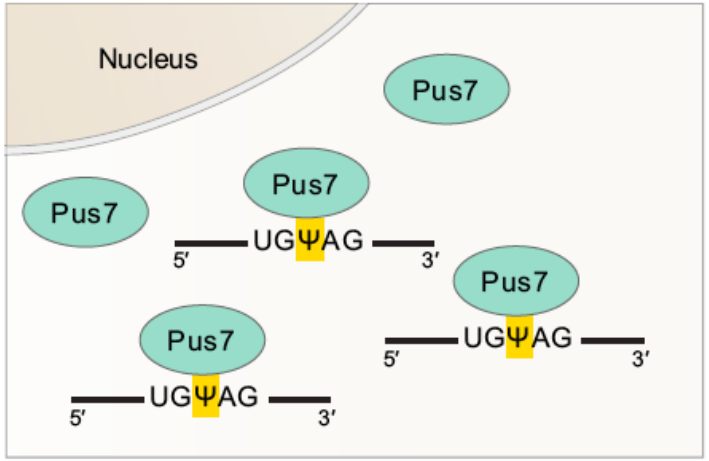
mRNA pseudouridylation



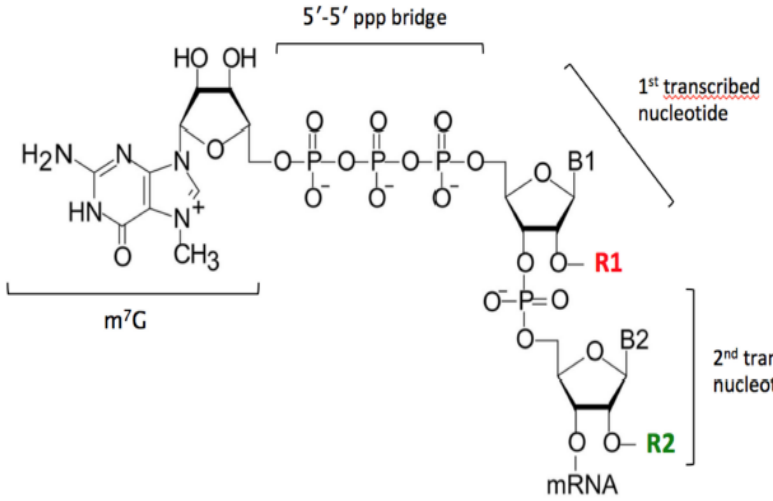
No heat



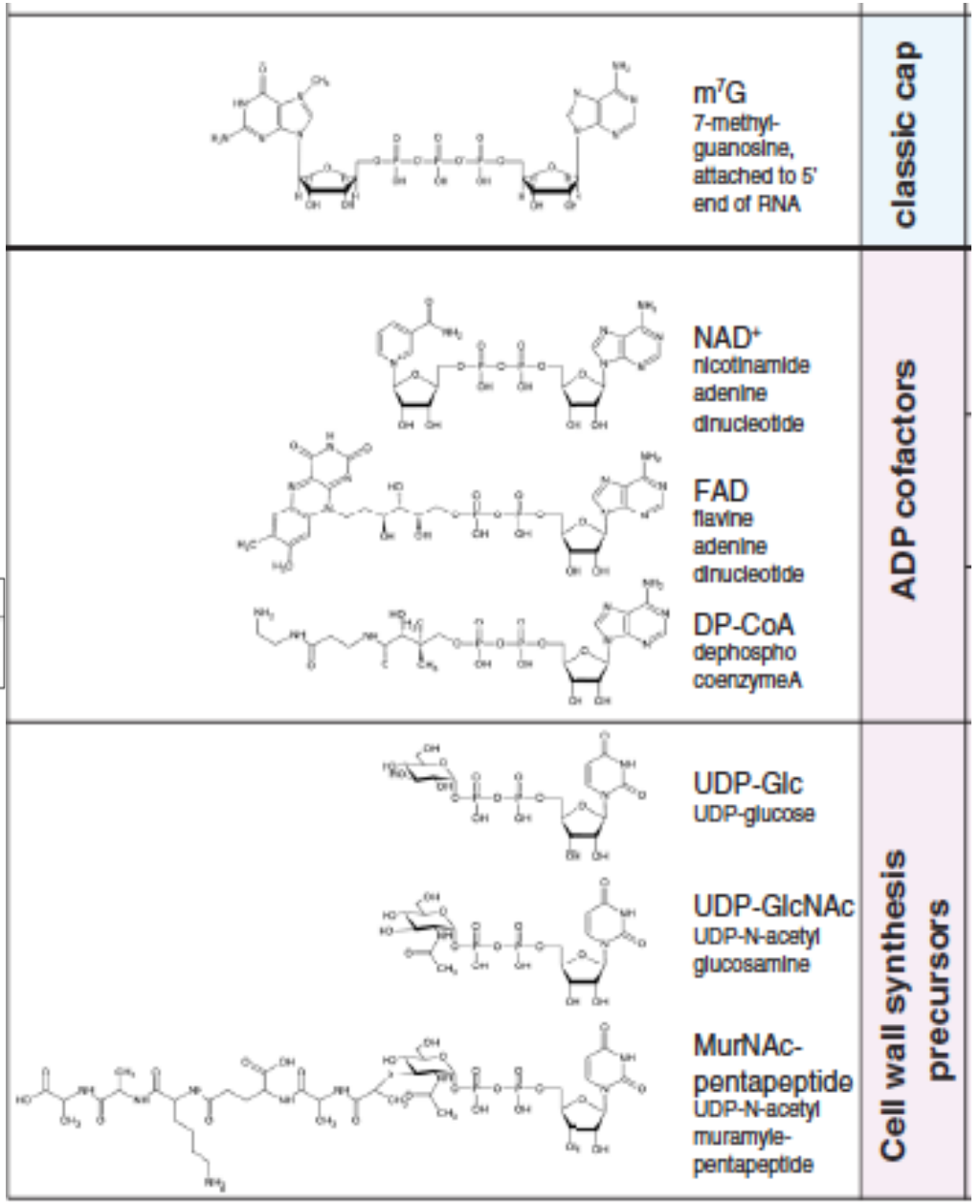
Heat shock



RNA MODIFICATION: alternative caps



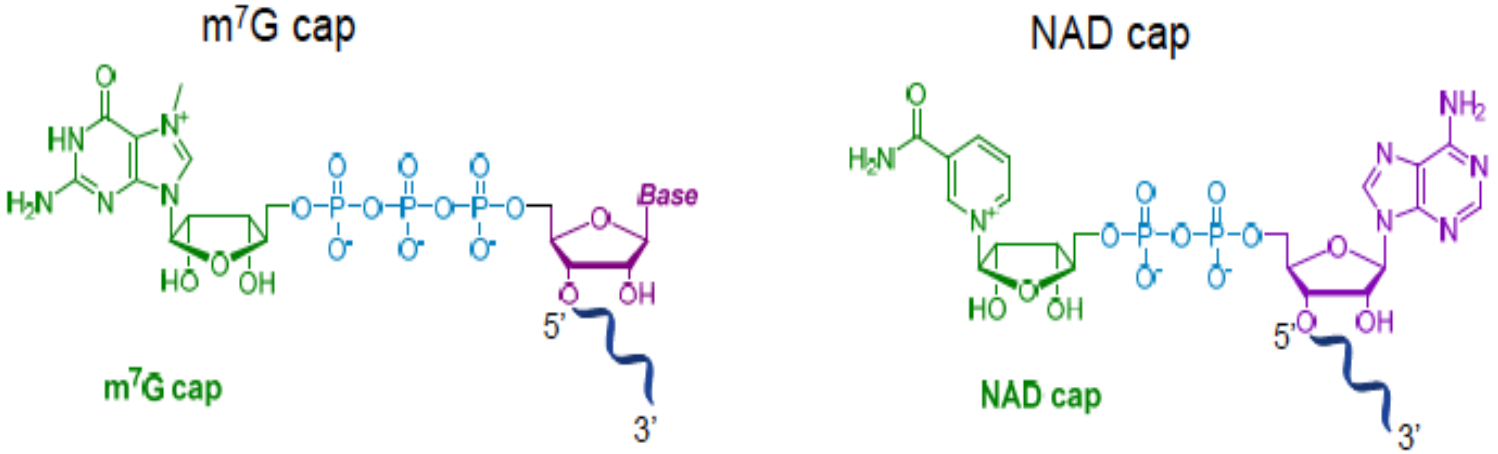
Cap0	Cap1	Cap2
R1 = H R2 = H	R1 = CH_3 R2 = H	R1 = CH_3 R2 = CH_3



Dimitrova et al, Genes 2019;
Jukius and Yuzenkova, WIREsRNA 2018

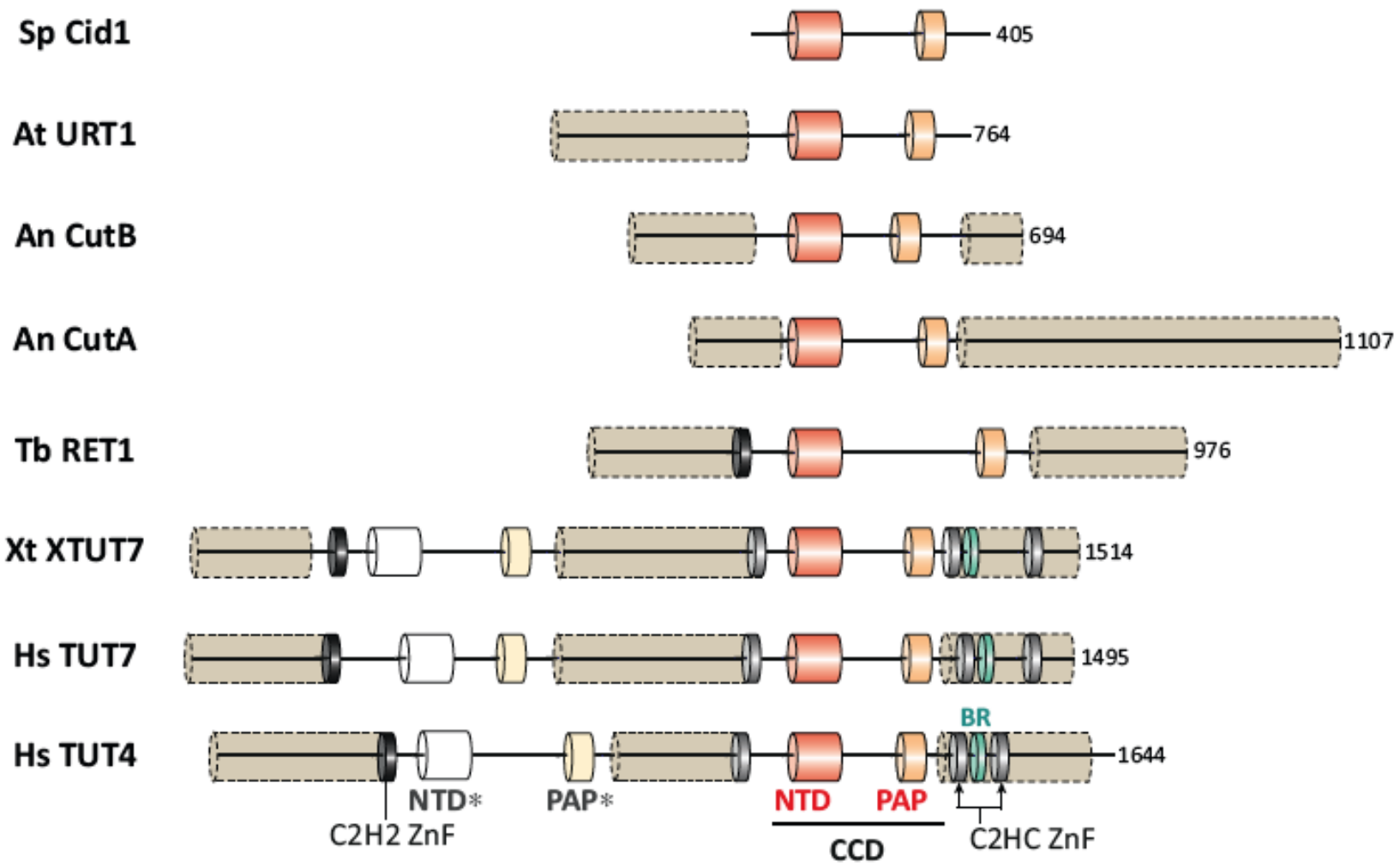
NAD^+ RNA cap

Nicotinamide Adenine Dinucleotide



- **Found in bacteria** (2009)**, *S. cerevisiae*** (2017), **mammalian cells** (2017, **plants** (2019)
- **Added cotranscriptionally by RNAP or posttranscriptinally**
- **Targets mRNA for degradation in eukaryotes, stabilizes mRNA in bacteria**
- **Hydrolyzed by specific enzymes (DXO and NUDT families)**
- **NAD^+ capped RNAs constitute only 1-5% of total RNAs**
- **NAD^+ capped RNAs are more abundant in mitochondria** (15% in humans, 60% in yeast)
- **NAD^+ capped RNAs are present on polysomes**
- **NAD^+ cap supports translation?**

ALTERNATIVE POLYMERASES



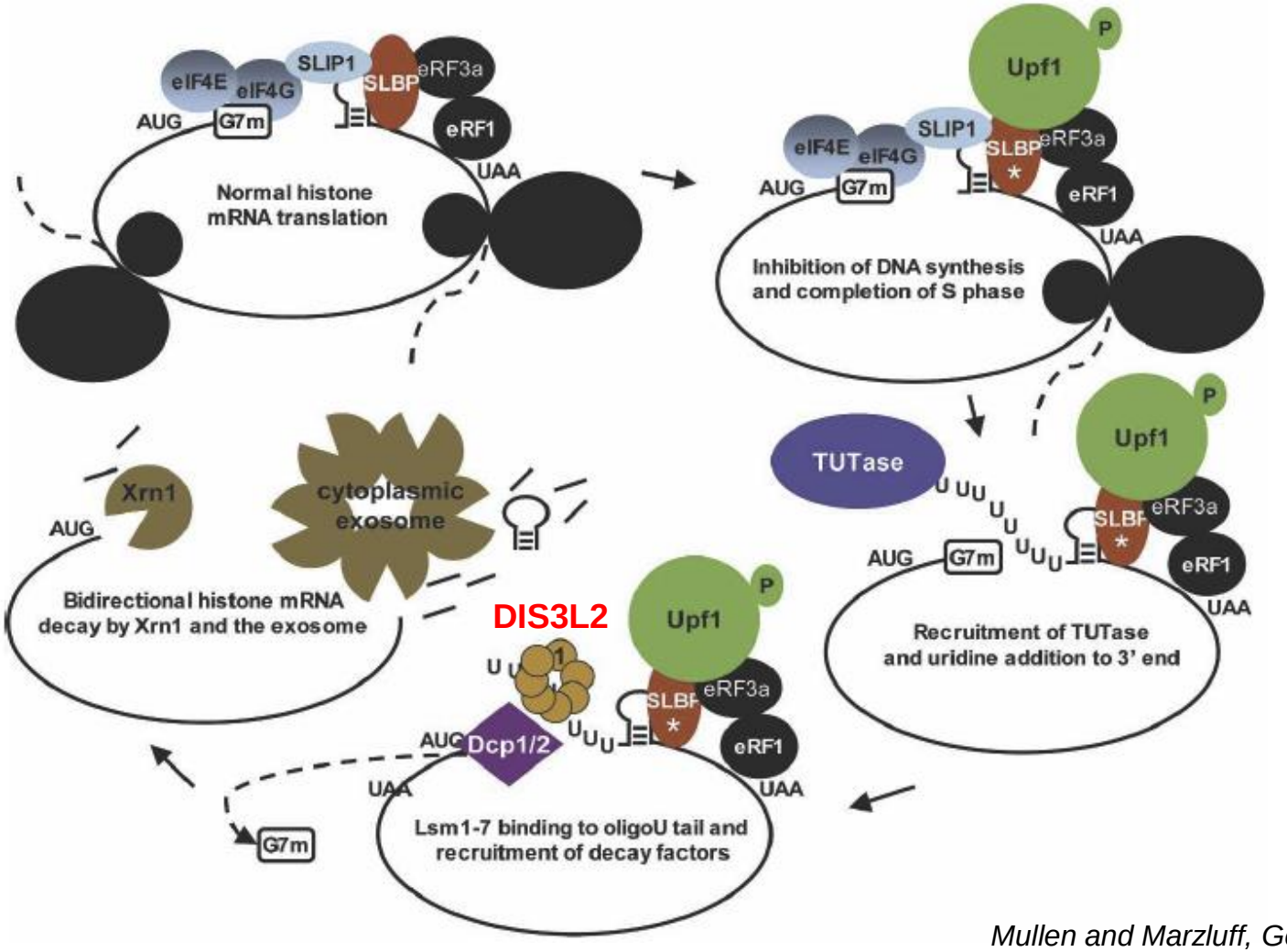
Scheer et al, TiG., 2016

OLIGO-URIDYLATION

PUP Poly(U) Polymerases
TUTase Terminal Uridylyl Transferase

3' oligouridylation

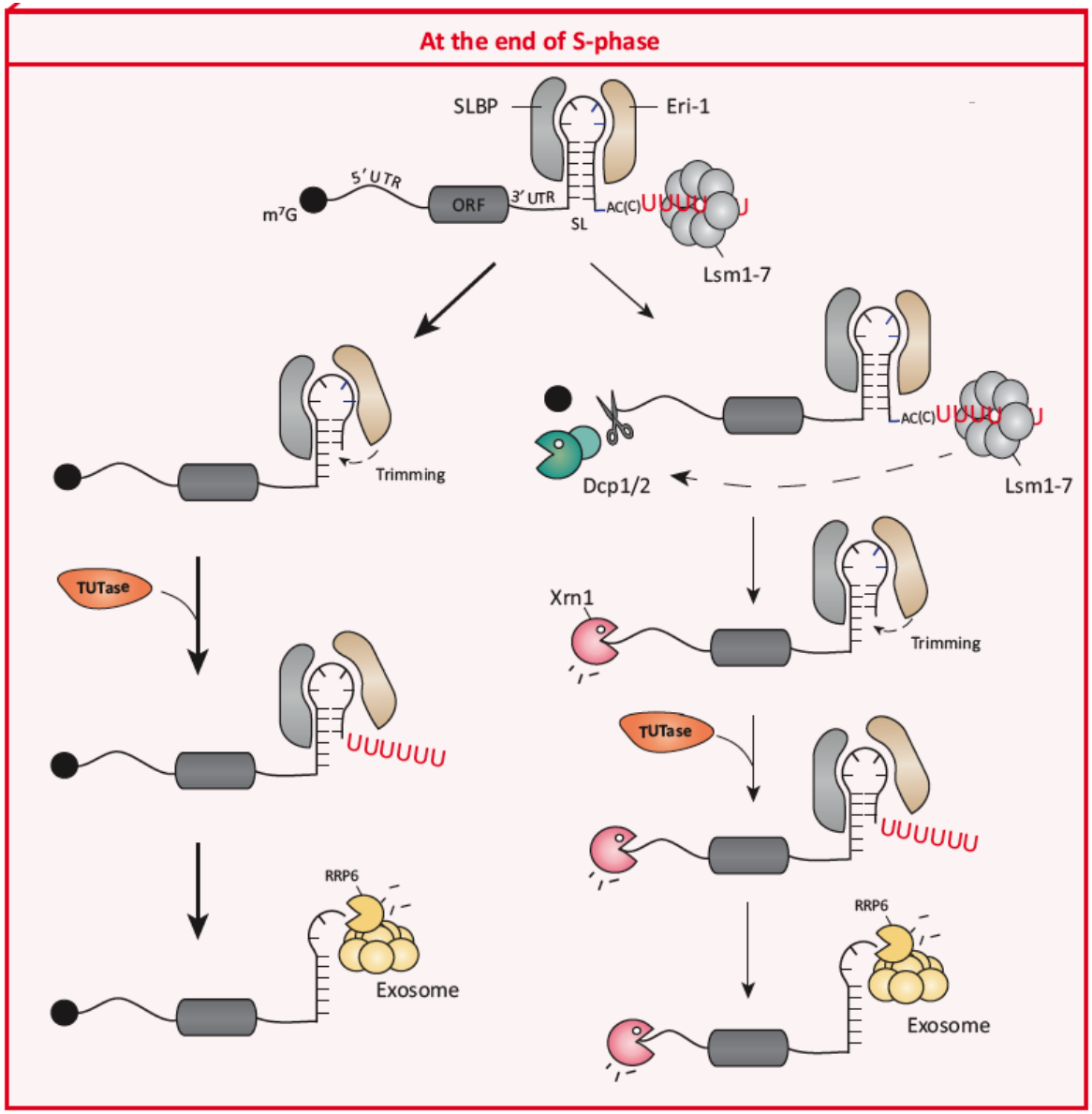
1. Histone mRNA degradation (metazoans)



Mullen and Marzluff, Genes Dev., 2008

OLIGO-URIDYLATION

Histone mRNA degradation (metazoans)

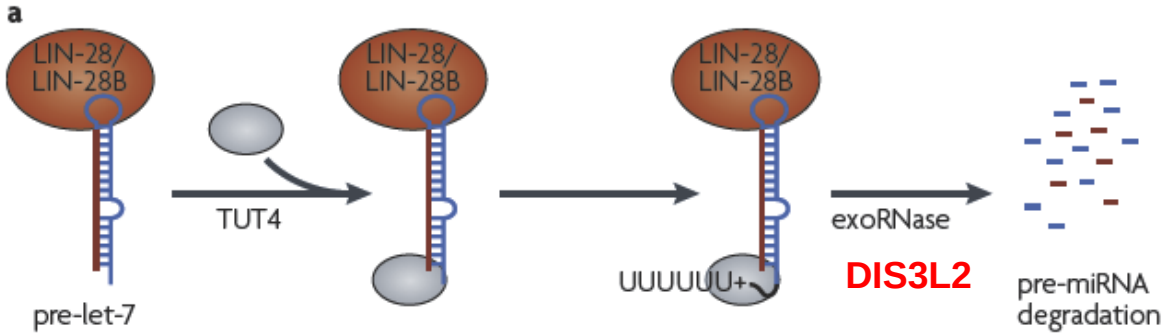


Scheer et al, TiG., 2016

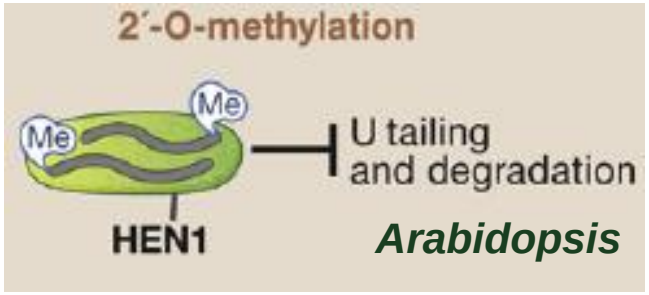
OLIGO-URIDYLATION

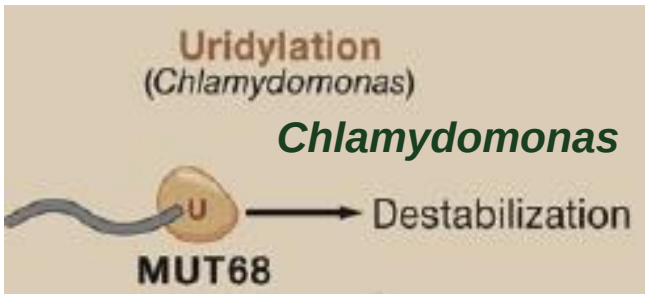
miRNA degradation

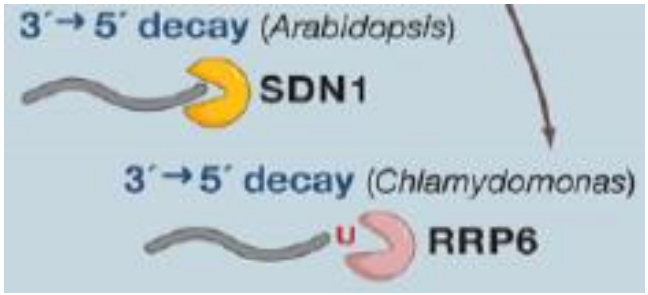
precursors
C. elegans



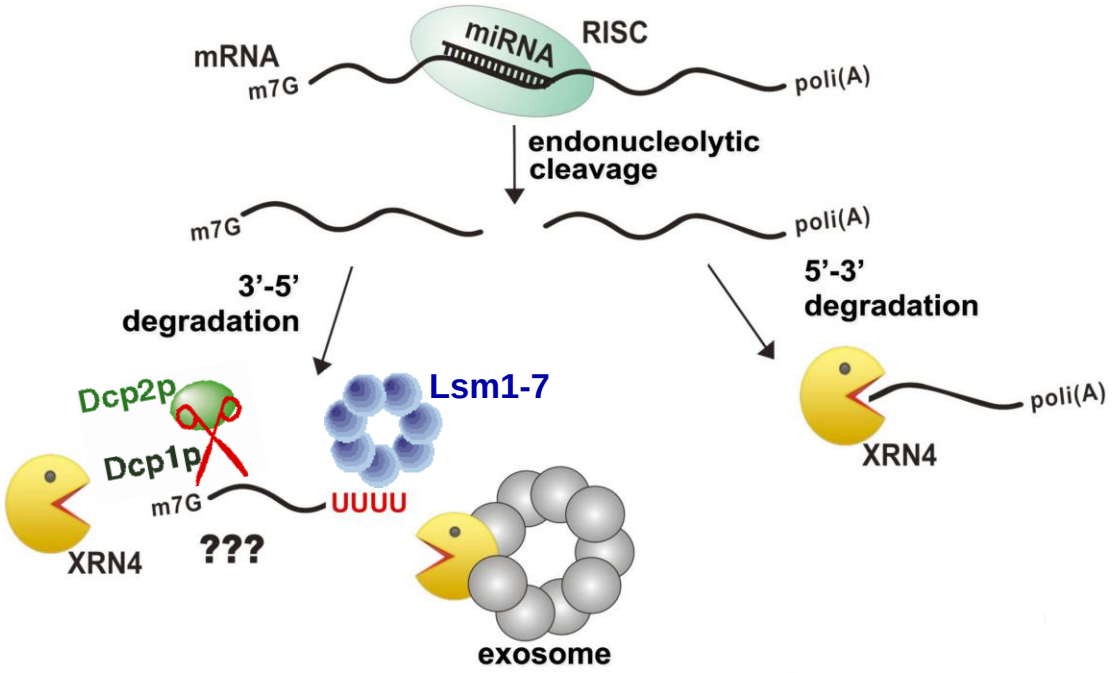
mature







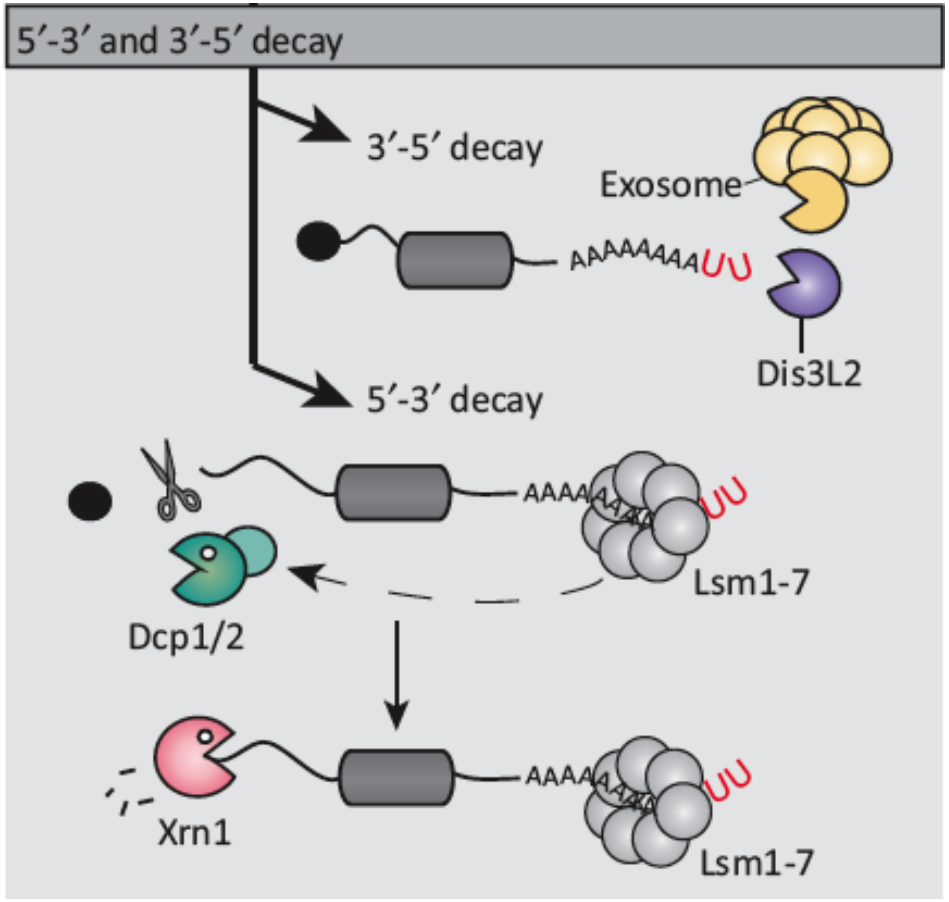
3. mRNA degradation? (plants)



Krol et al., Nat Rev Genet, 2010; Kim et al., Cell, 2010

OLIGO-URIDYLATION

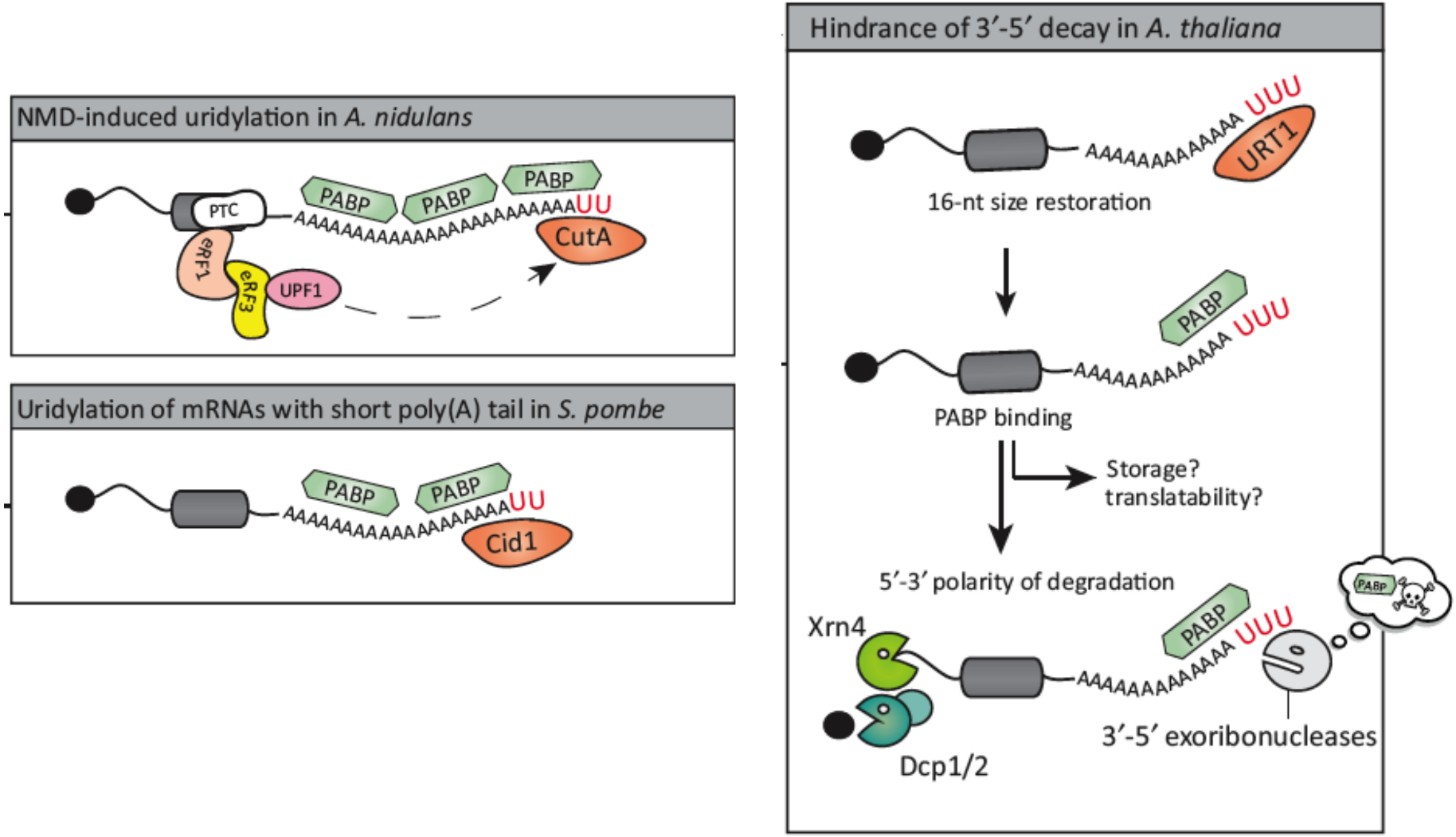
mRNA degradation



Scheer et al, TiG., 2016

OLIGO-URIDYLATION

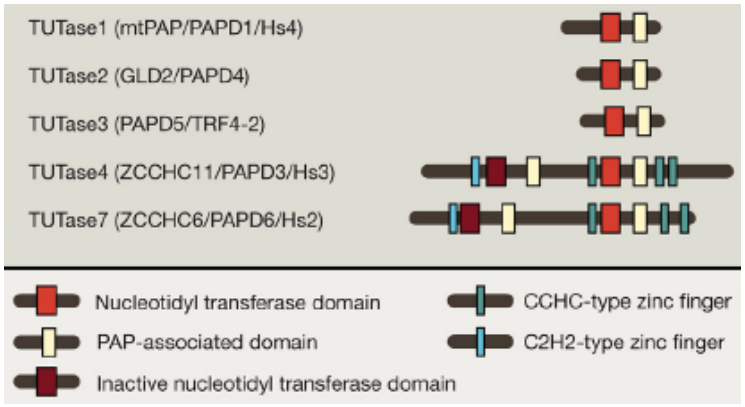
Other



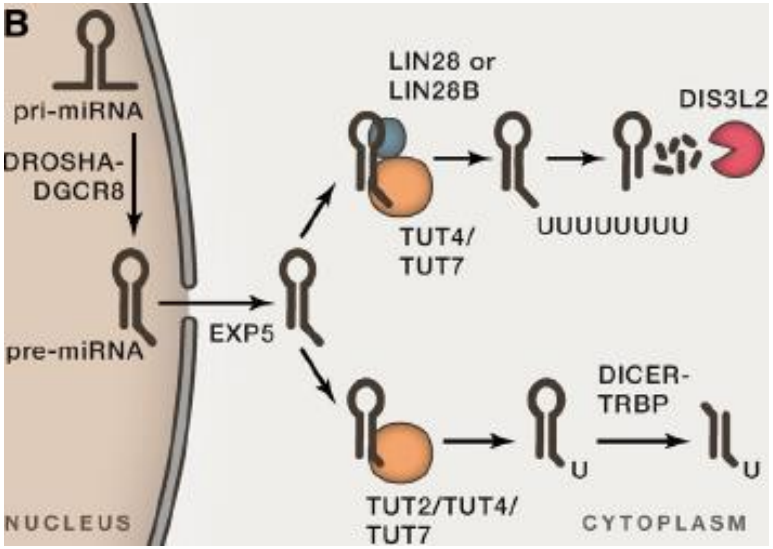
Scheer et al, TiG., 2016

URIDYLATION

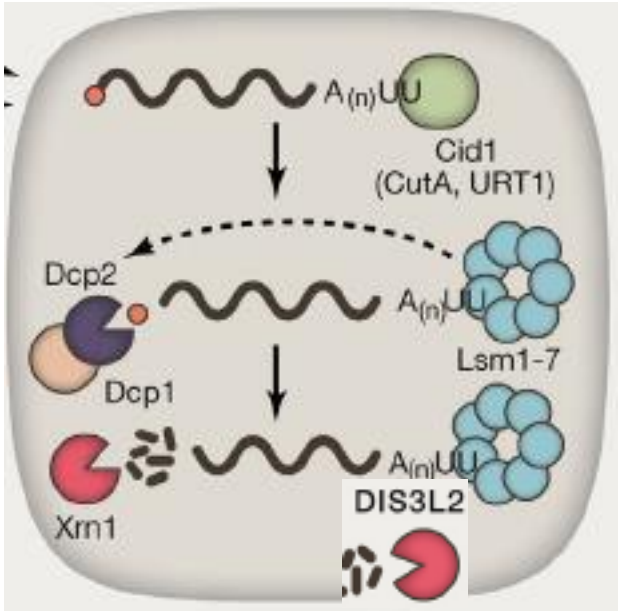
TUTases



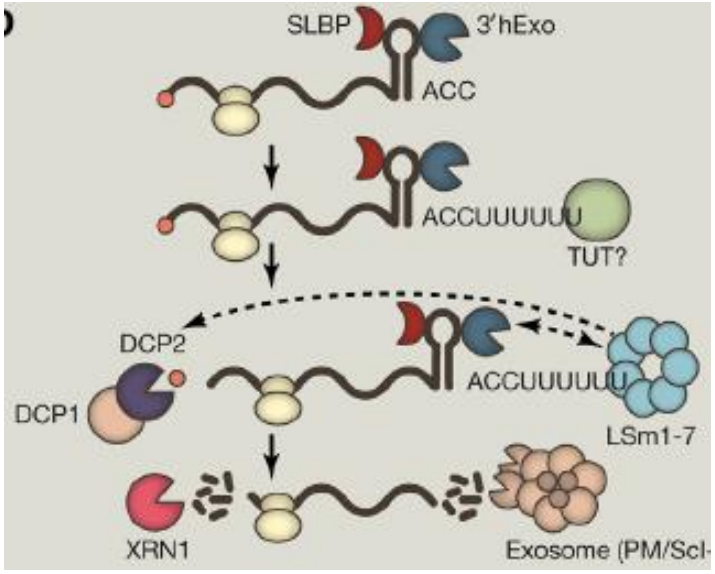
Uridylation of pre-miRNAs and miRNAs



Uridylation-dependet mRNA decay

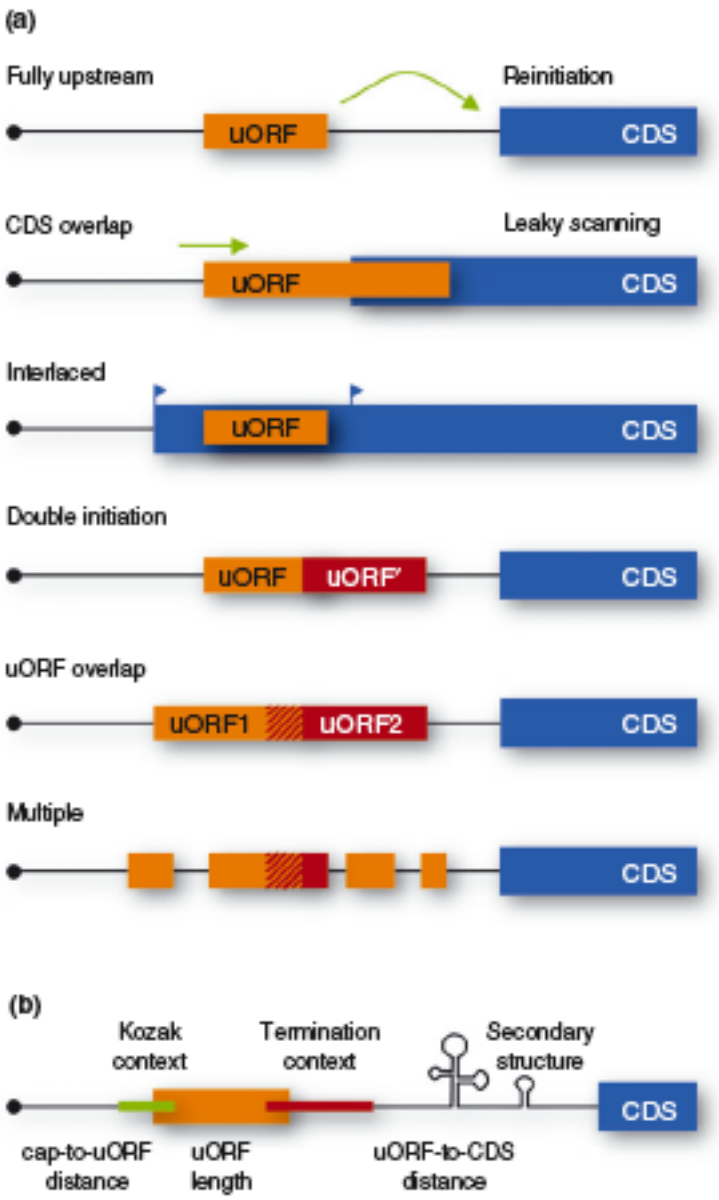


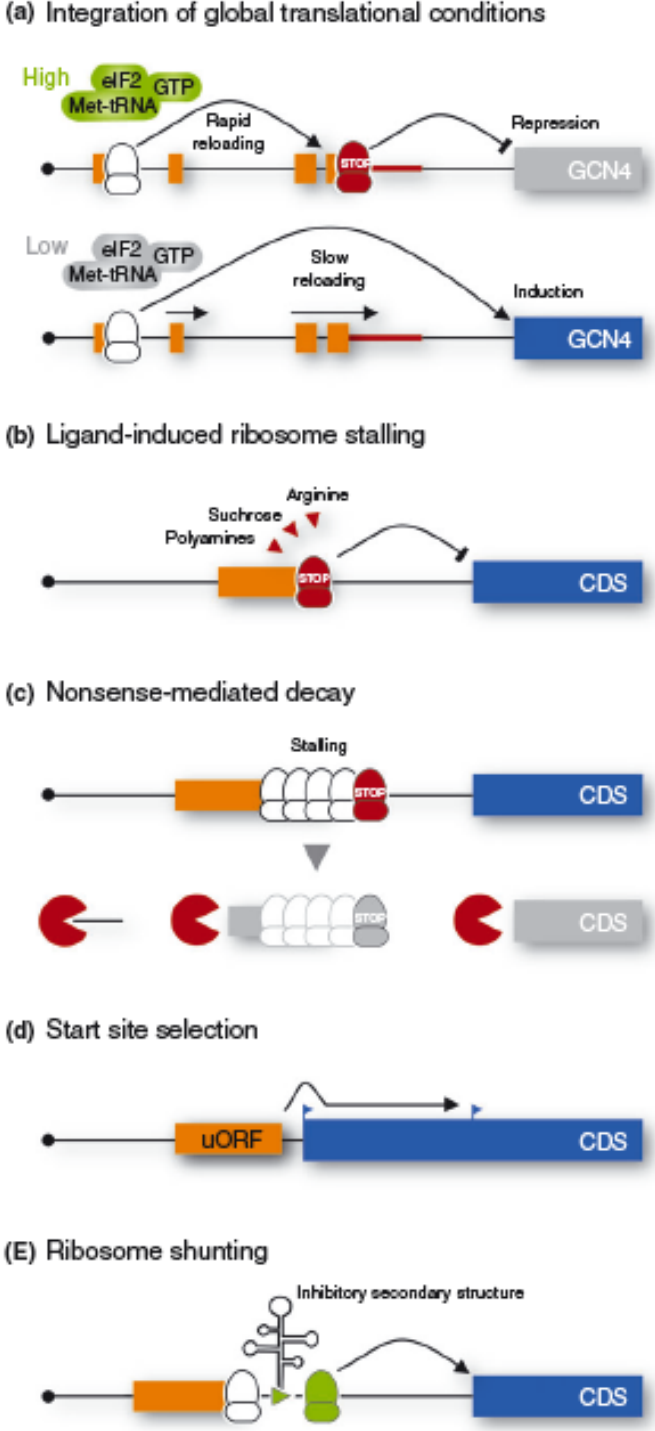
Degradation of histone mRNAs



Lee et al, Cell, 2014

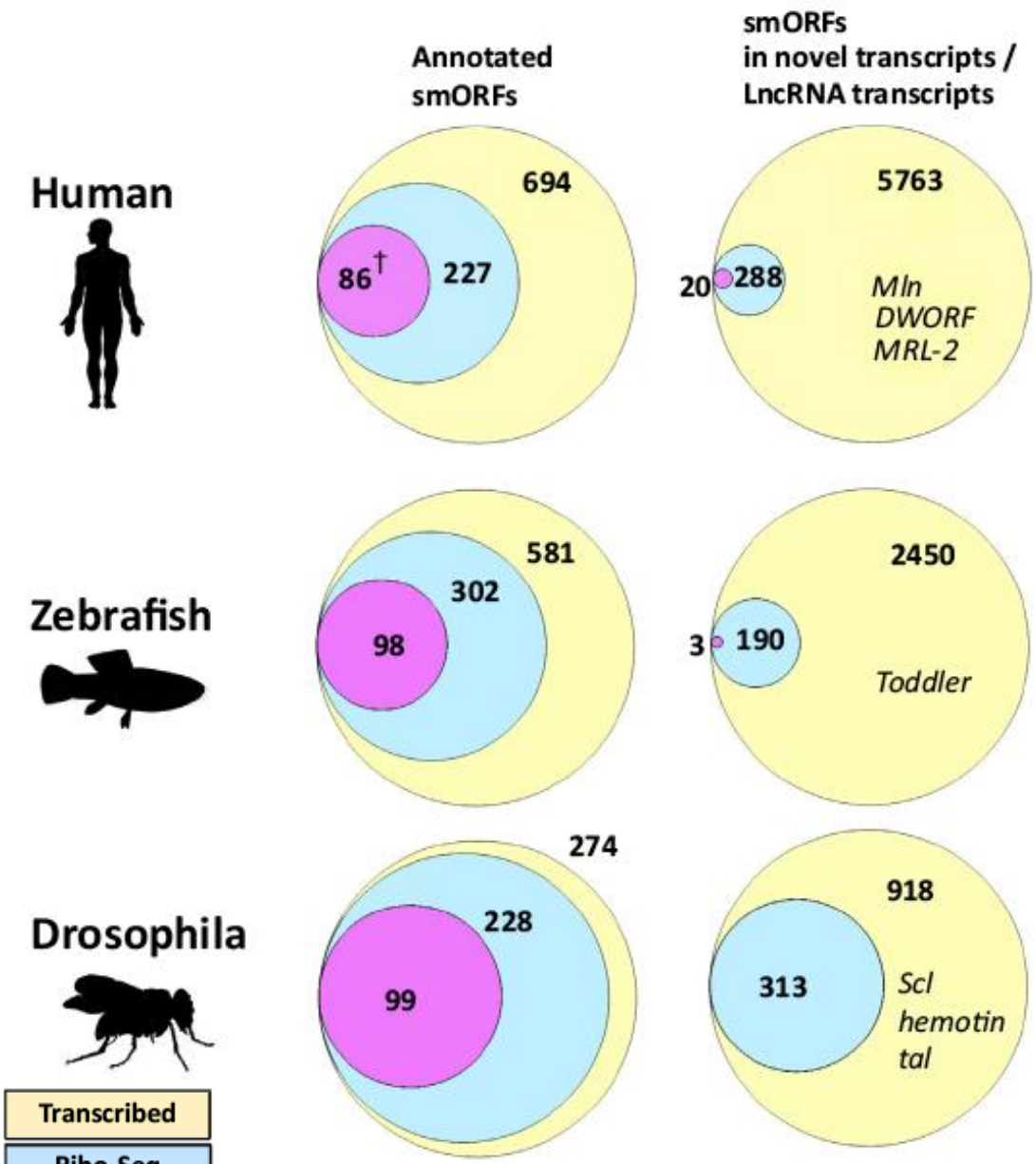
uORFs = upstream ORFs





Wethmar WIREsRNA, 2014

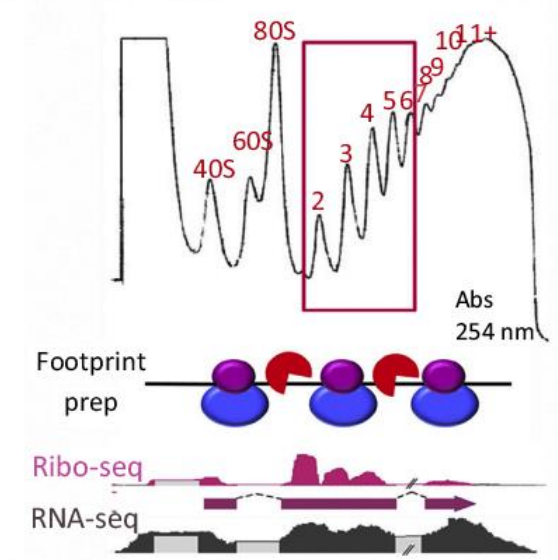
small ORFs = sORFs, sPEPs, smORF



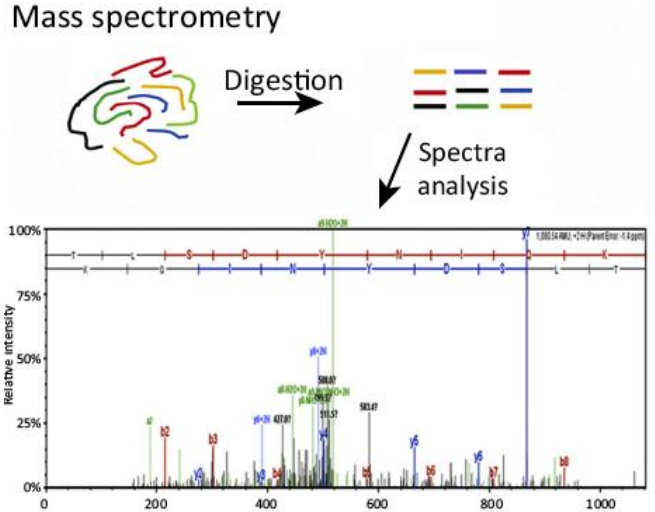
Puyeo et al, TiBS, 2016

Bioinformatics approaches
Sequence composition analysis
Conservation
Biochemical approaches

Ribosome profiling

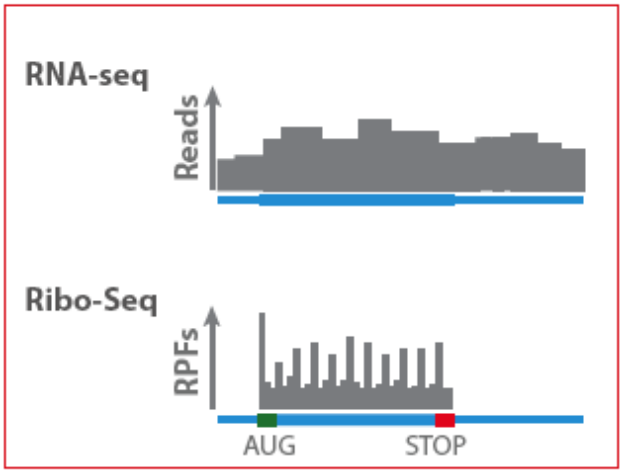


Mass spectrometry

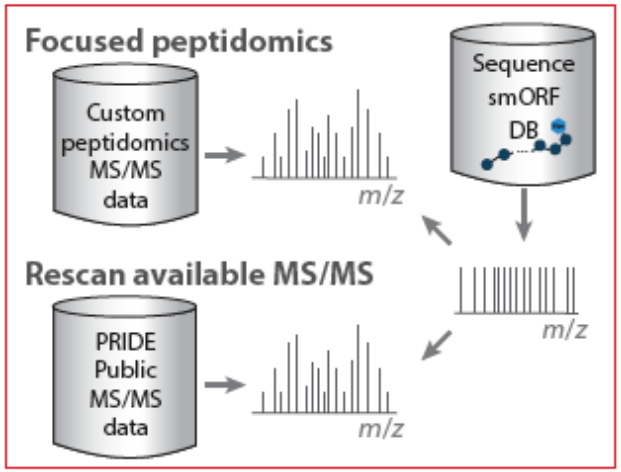


Methods to identify sORFs and sPEPs

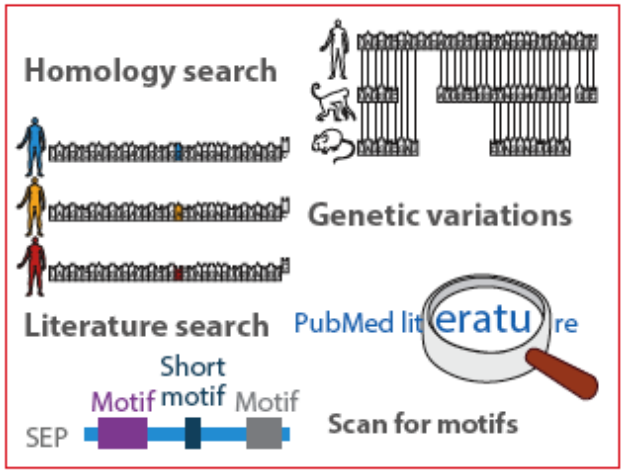
Sequencing techniques



Mass spectrometry

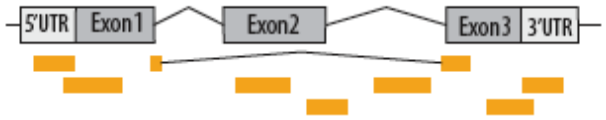


Bioinformatics



Genome-wide discovery of SEPs

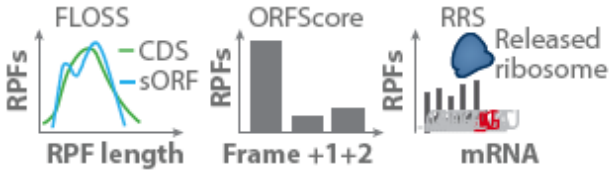
ORF detection (I)



Available prediction tools:
- ORF finder
- PROTEOFORMER
- ORF-RATER
- RiboTaper

ORF coding potential (II)

(a) Ribo-Seq-derived scores:



(b) Phylogenetic conservation scores:
- PhyloCSF, micPDP, CPC

(c) MS validation

SEP annotation (III)

(a) Disease-related genetic variation:
- NCBI dbSNP
- COSMIC (somatic mutations in cancer)

(b) Motif-derived function:
- protein domains (InterPro, Pfam...)
- short linear motifs (ELM)

(c) Homology-derived function:
- NCBI BLAST

Plaza et al, Annu Rev Cell Dev Biol, 2017

Integrative repositories for SEPs

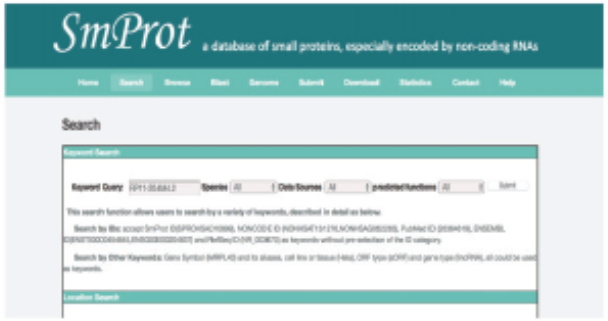
http://www.sorfs.org

Multispecies

Query interface:
simple, advanced (BioMart), genome browser

Contains:
I: PROTEOFORMER
IIa: FLOSS, ORFScore
IIb: PhyloCSF, PhyloP, PhastCons
IIc: global PRIDE-ReSpin
IIIa: NCBI dbSNP
IIIb: InterPro, Pfam, sLiMs, ELM (future)
IIIc: NCBI BLAST

http://www.bioinfo.ibp.ac.cn/SmProt/

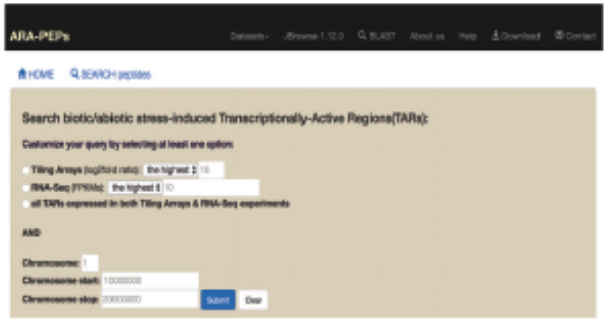


Multispecies

Query interface:
simple, genome browser, BLAST

Contains:
I: RiboTaper
IIa:
IIb: PhyloCSF (partly from Mackowiak et al. 2015)
IIc: specific MS/MS data sets
IIIa:
IIIb: InterPro
IIIc: BLAST search is available

http://www.biw.kuleuven.be/CSB/ARA-PEPs/



Arabidopsis thaliana

Query interface:
simple, genome browser, BLAST

Contains:
I: sORF finder or tiling arrays
IIa:
IIb: d*N*/d*S* ratio
IIc:
IIIa:
IIIb: TM domain, signal sequence
IIIc: BLAST search is available

Plaza et al, Annu Rev Cell Dev Biol, 2017

sORFs

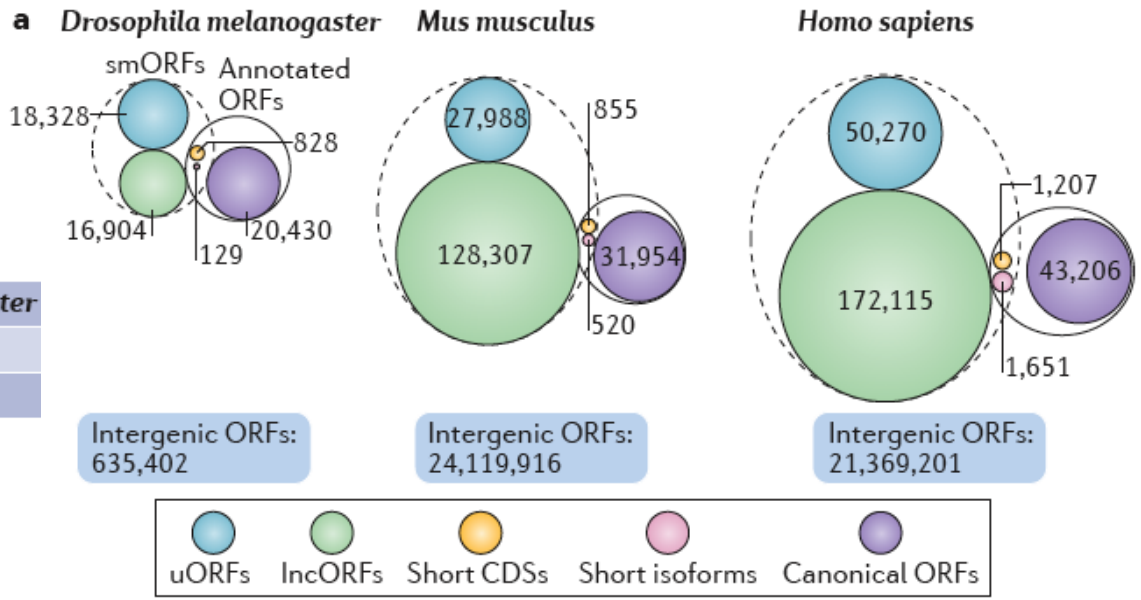
ORF class	RNA type	Median size (codons)	Translation[15]	Conservation	Coding features	Function
Intergenic ORFs	None	22	None	None[6,8]	Non-canonical AA	None
uORFs	5′ UTRs of mRNAs	22	Low	None[8,30]	• Nonrandom AA • No domains	• Non-coding • Translation regulation
lncORFs	lncRNAs	24	Low	None[8,10]	• Nonrandom AA • No domains	Non-coding or coding
Short CDSs	Short mRNAs	79	High	Class	• Positively charged AA • Transmembrane α-helices	• Coding • Regulators of canonical proteins
Short isoforms	Spliced mRNAs	79	High	Kingdom	• Canonical AA • Protein domain loss	• Coding • Small interfering peptides
Canonical ORFs	mRNAs	491	High	Kingdom	• Canonical AA • Multiple protein domains	• Coding[42] • Structural, enzymatic, regulatory

Transcribed smORFs

Untranslated region
ORFs
DNA
Other coding sequences
RNA splicing
Ribosome profiling signal

Median length (codons)

uORFs	lncORFs	Short CDSs	Canonical ORFs	
20	25	79	490	*D. melanogaster*
22	23	81	424	*M. musculus*
23	24	78	421	*H. sapiens*



Couso and Patraquim
Nat Rev Mol Cell Biol, 2017

SUMMARY or HOW TO PASS THE EXAM?

- THEORY
- METHODS
- GENERAL IDEAS, CONCEPTS, SOLUTIONS
- PATHWAYS
- MECHANISMS