

# Ecole-Chercheur 2017

De l'expression des gènes aux réseaux

MODULE 1 : LES PROTOCOLES DE SEQUENCAGE A HAUT DEBIT DES  
ARNs ET LEURS APPLICATIONS

JÉRÉMIE BAZIN (POSTDOC IPS2)



# OUTLINE



- Global transcriptome analysis : Goals, technical approaches and potential bias
- How to capture RNA subpopulations with RNA-sequencing : Concepts and Protocols
- Address biological questions and uncover molecular mechanisms with RNA-seq
- Combine technologies to access multiple gene expression regulations tiers in specific cells

# RNA-SEQ OR TRANSCRIPTOME SEQUENCING

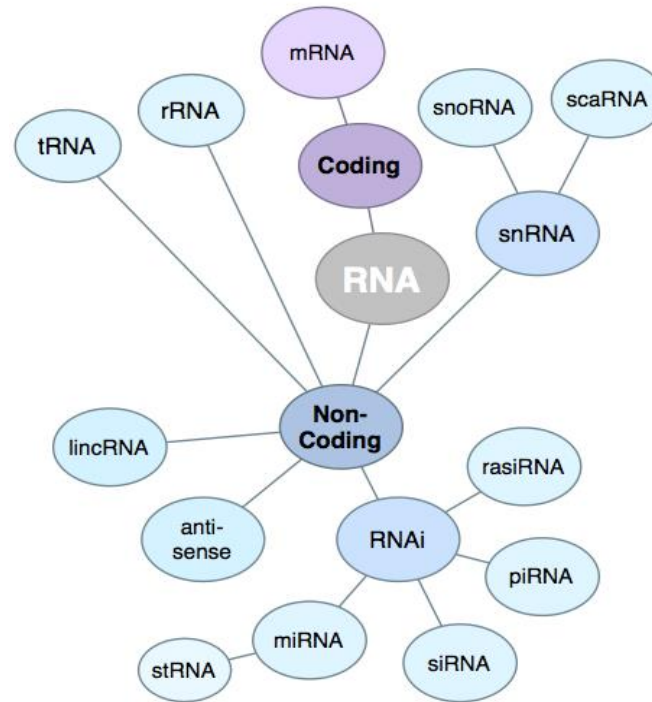


- **Transcriptome**
- It is a set of all transcripts in any given cell, all products of the transcriptional machinery.
- It includes coding (protein encoding) and non-coding RNA (not protein encoding).

# RNA-SEQ OR TRANSCRIPTOME SEQUENCING



## Transcriptome: RNA World



<http://finchtalk.geospiza.com/2009/05/small-rnas-get-smaller.html>

# GOALS TO STUDY THE TRANSCRIPTOME :



- ✓ **Biological question ?**
  - What kind of RNA (mRNA, ncRNA, smallRNA, ...) ?
  
- ✓ **Quantification ?**
  - Abundance of transcripts between different conditions
  
- ✓ **Genome annotation ?**
  - Identify genes, exons, splicing events, ncRNAs, etc.
  - Novel genes or transcripts

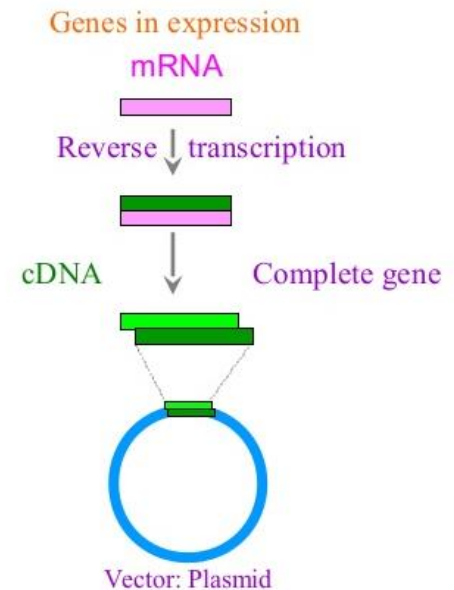
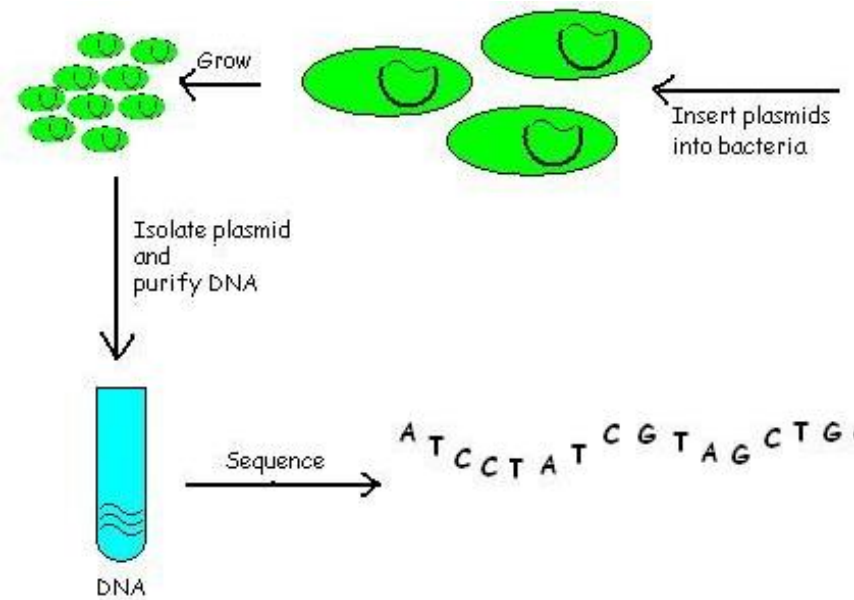
# CAPTURE THE TRANSCRIPTOME : THE cDNA LIBRARY



- cDNA, not RNA sequencing
- Old technique to study gene expression (Sim et al., Cell 1979)

Cell, Vol. 18, 1303-1316, December 1979, Copyright © 1979 by MIT

## Use of a cDNA Library for Studies on Evolution and Developmental Expression of the Chorion Multigene Families

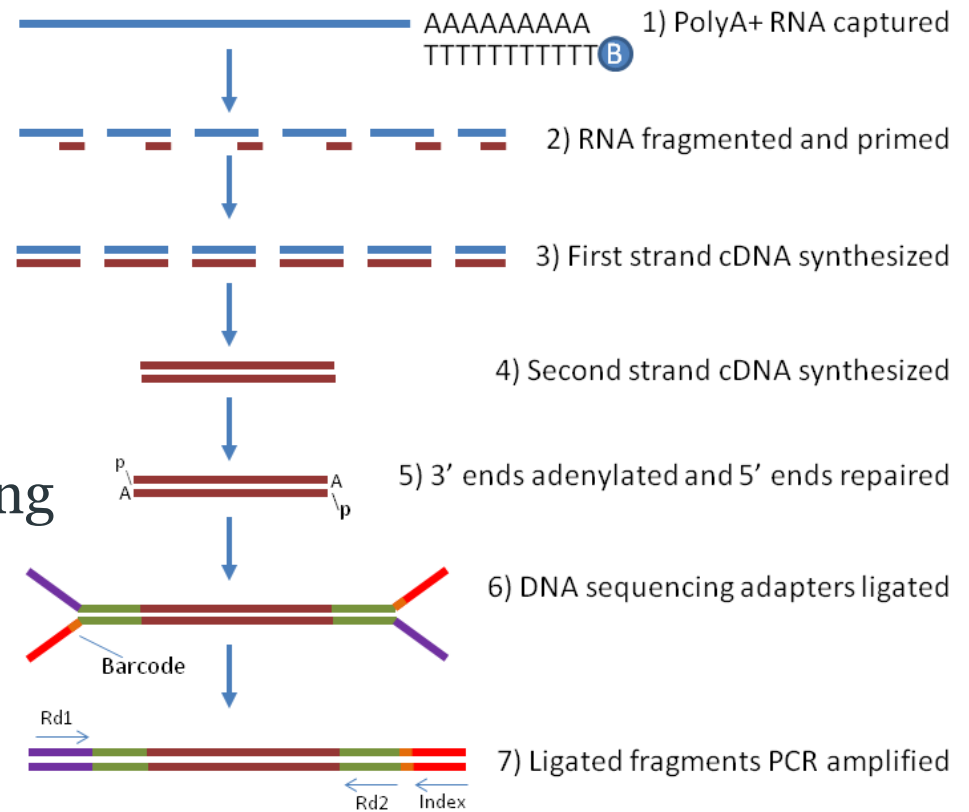


# NEXT GENERATION CDNA LIBRARY



- **Basic steps :**

- RNA Fragmentation
- Reverse transcription
- 5' /3' Adapter Ligation
- PCR amplification
- Library quantification
- High throughput sequencing

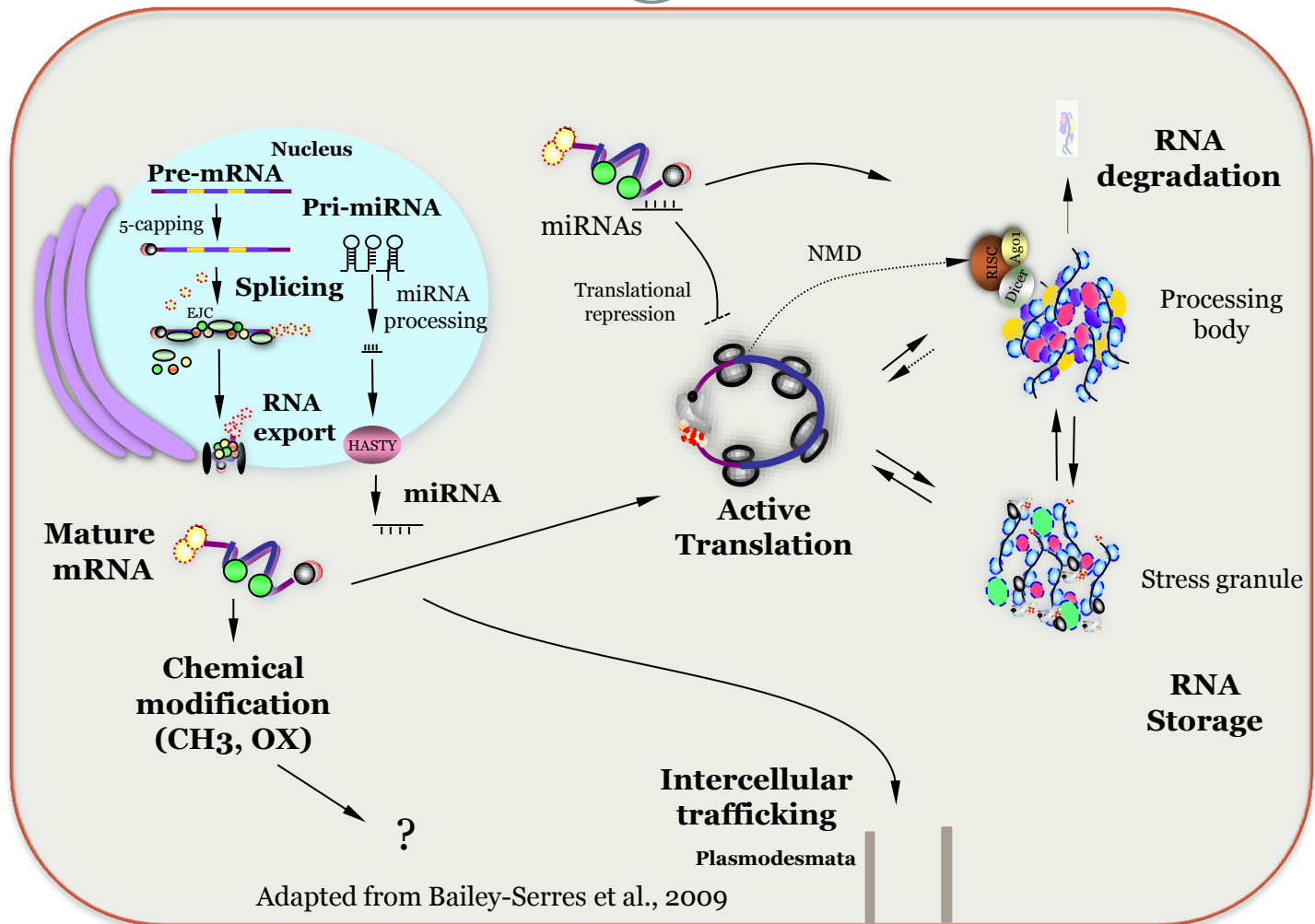


# NEXT GENERATION cDNA LIBRARY



- Possible bias (most protocols)
  - RNA degradation during extraction
  - RNA fragmentation
  - PCR artifacts :
    - ✦ GC content sequences: depend on the DNA polymerase
    - ✦ PCR over-amplification : The number of cycle may need to be optimized : the less the better
  - Adapter dimer contamination

# THE LIFE OF AN RNA, A JOURNEY OF MANY PATHWAYS

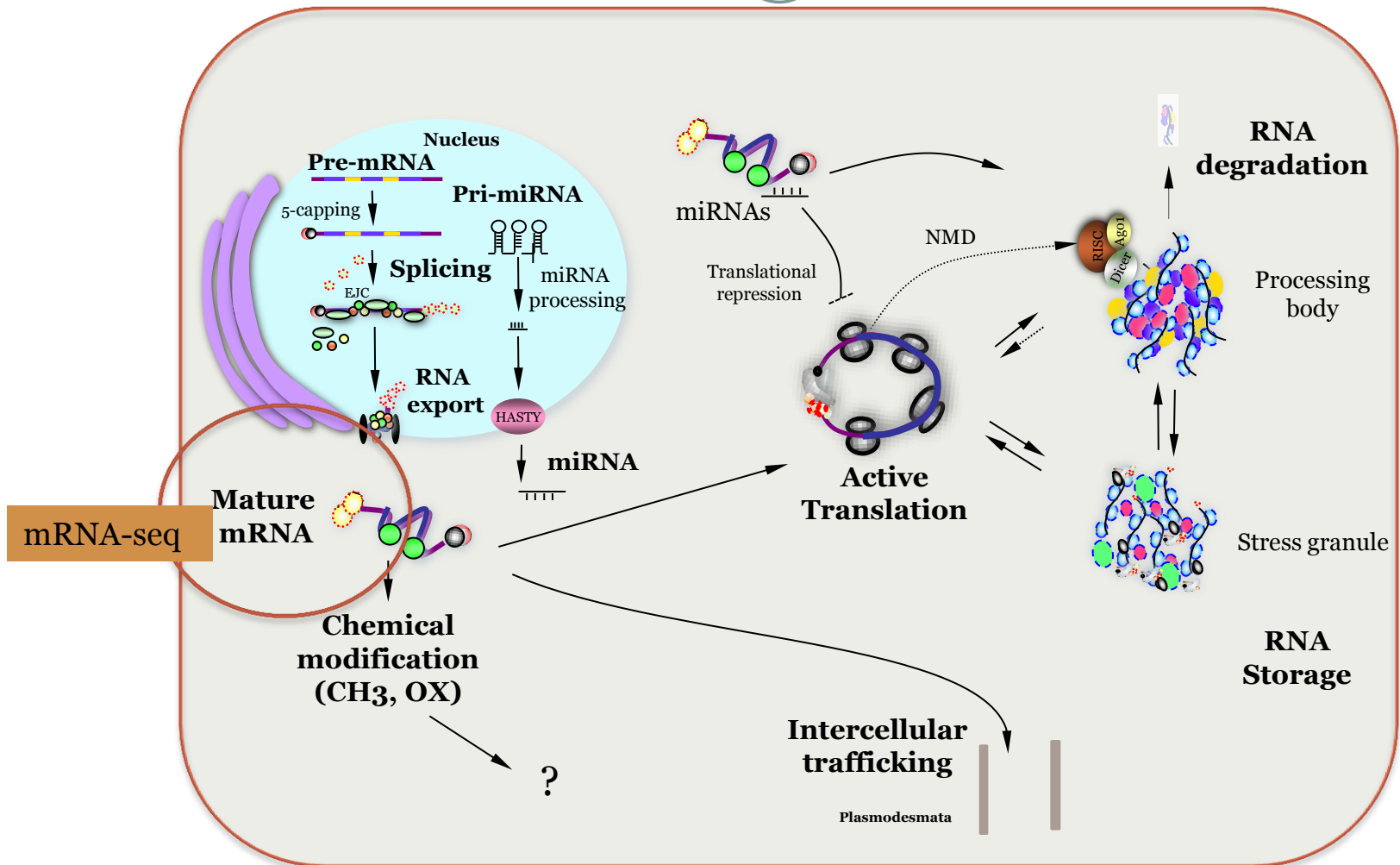


# HOW TO ISOLATE RNA SUBPOPULATIONS



- Based on intrinsic chemical features of the RNA (Size, 5'/3' end)
- Subcellular localization (Nucleus, organelles...)
- Association with proteins complexes
- Combination of these approaches
- Everything is possible but need to be designed carefully !

# Q: What is the effect of my favorite stress, mutation on nuclear gene expression ?



# MRNA-SEQ

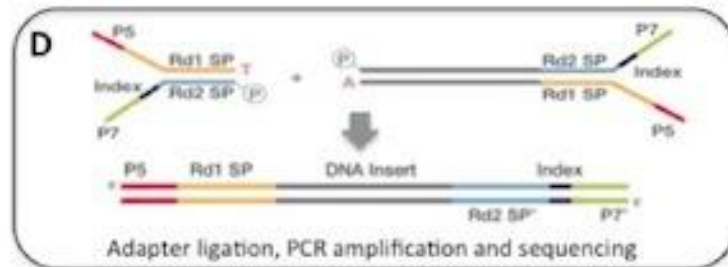
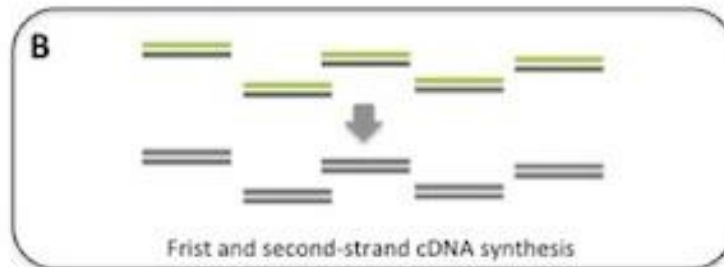
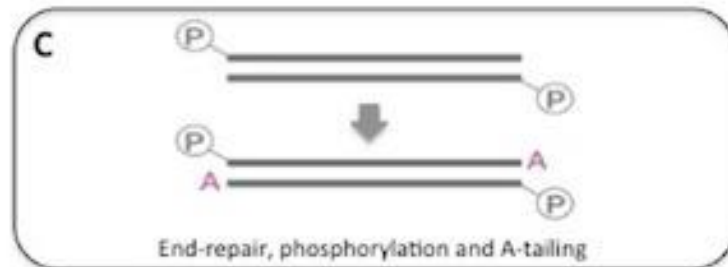
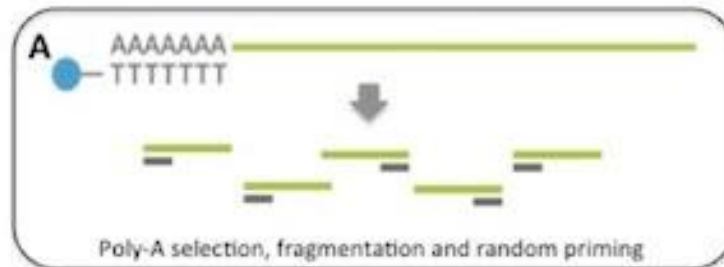


- mRNA features :
  - ✦ Broad size distribution,
    - Will require fragmentation before cloning
  - ✦ polyA tail, 5' Cap
    - PolyA tail can be used to enrich for mRNA among other RNA

# MRNA-SEQ



## Illumina Tru-Seq RNA-seq protocol

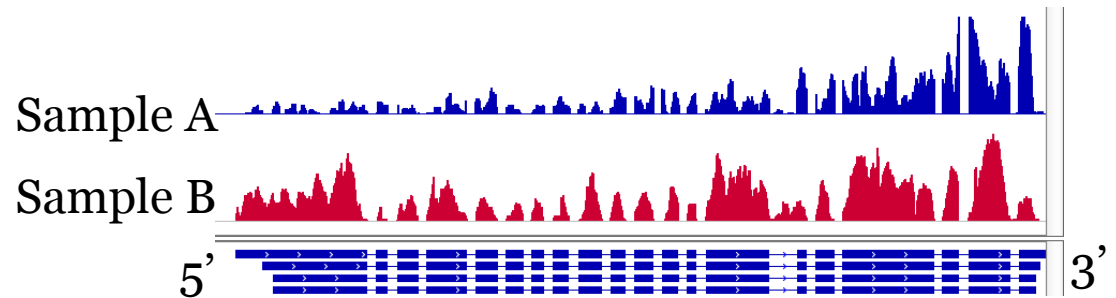
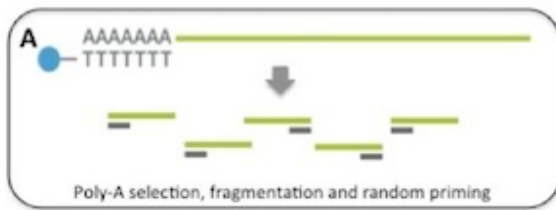


Library prep begins from 100ng-1ug of Total RNA which is poly-A selected (A) with magnetic beads. Double-stranded cDNA (B) is phosphorylated and A-tailed (C) ready for adapter ligation. The library is PCR amplified (D) ready for clustering and sequencing.

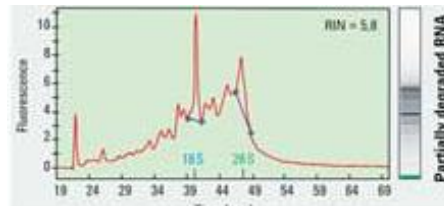
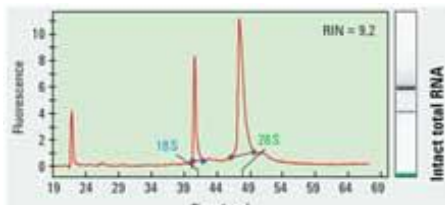
# MRNA-SEQ



- Potential bias I
  - Based on purification of polyA RNA : sensitive to RNA degradation !!
  - Can create a 3' end biased coverage which may compromise quantification



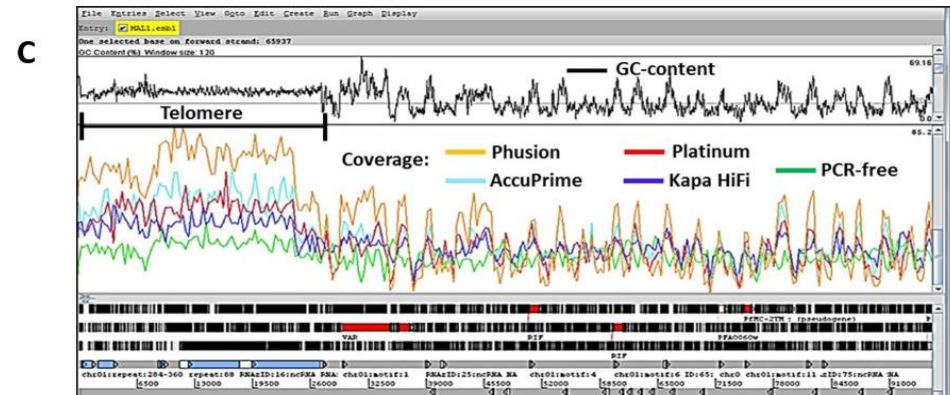
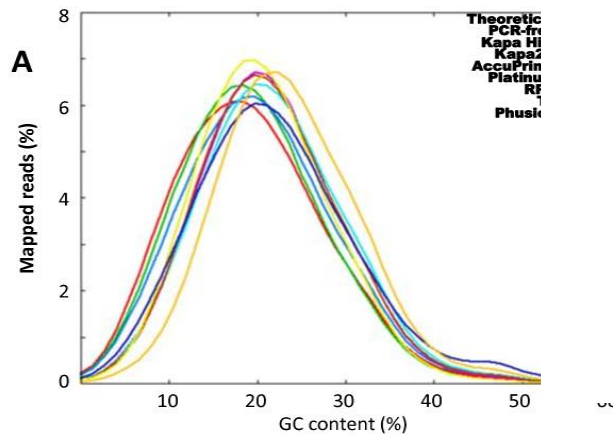
- Solutions : Check RNA integrity
  - Agilent Bioanalyzer (RIN > 8)



# MRNA-SEQ



- Potential bias II
- PCR overamplification :
  - Reduce library complexity, results in overrepresented sequences.
- GC content
  - Polymerase dependent
  - Can be reduced by reducing PCR cycles

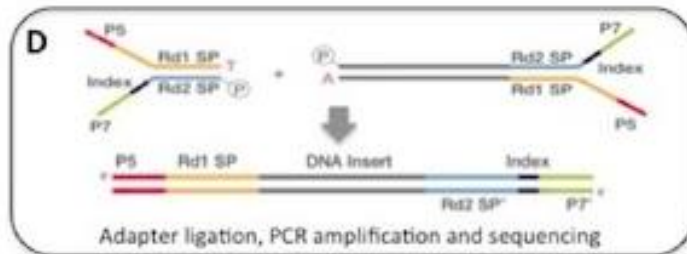


Oyola et al., 2012 BMC Genomics

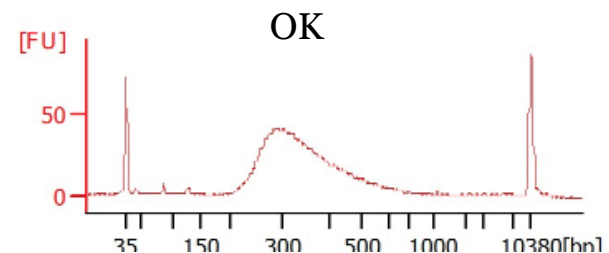
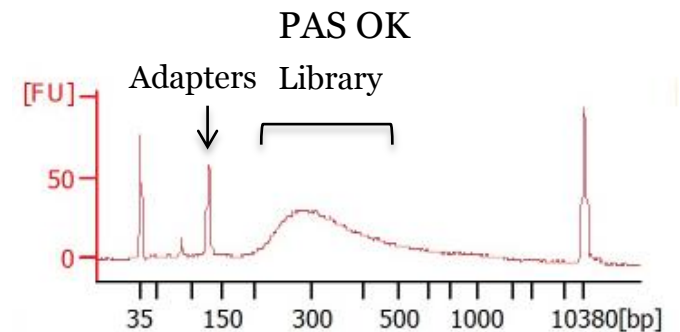
# MRNA-SEQ



- Potential bias III
- Adapter dimer contamination:
  - Shorter fragment are sequenced preferentially = loss of usable reads



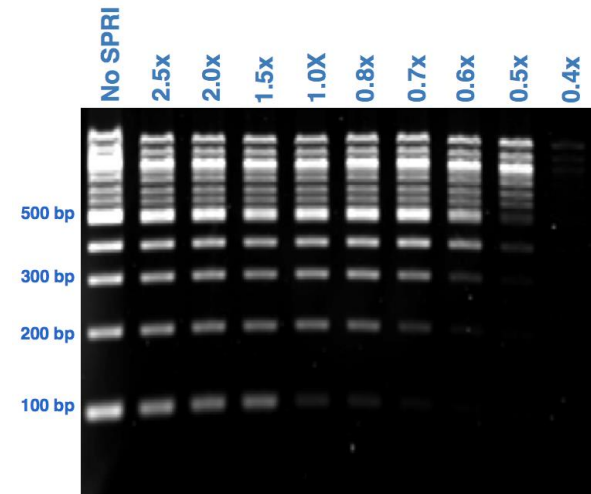
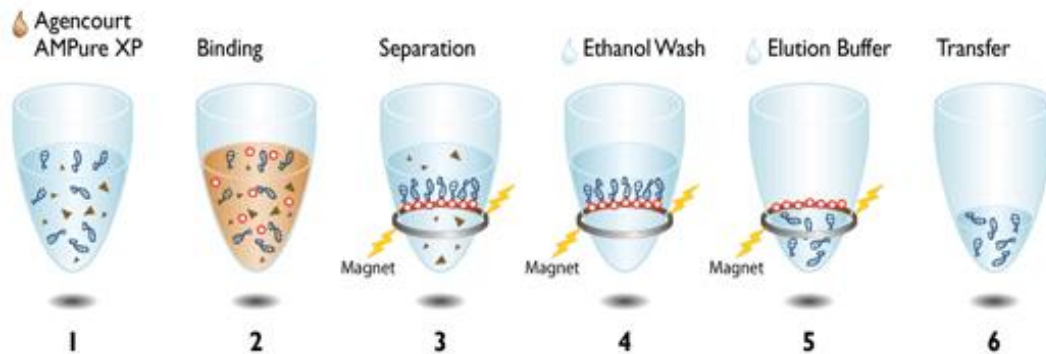
At this step, adapters can be ligated together.  
Especially when RNA input is to low



# MRNA-SEQ



- AMPure XP bead purification
- At each step of the protocol, where buffer or removal of primer is required.
  - Is also used to separate dsDNA by size

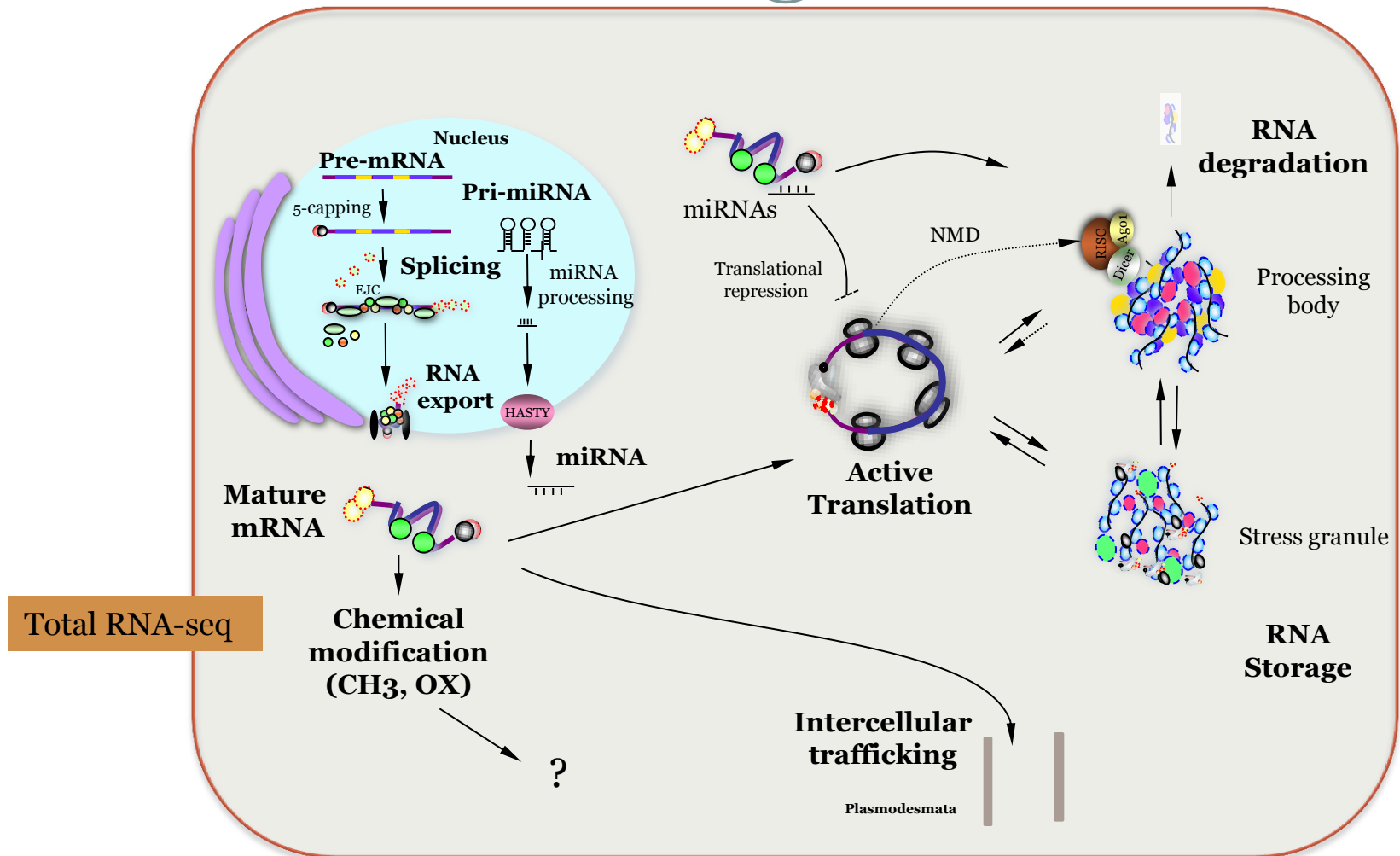


# HOW TO DECREASE THE COST OF mRNA-SEQ



- Design your experiment carefully, use the right depth, more replicates is often better than more depth.
- Prices of sequencing are constantly decreasing but not the cost of library prep kits !
- You can use a home made protocol, this is just basic molecular biology!
  - ✦ Ex : Wang et al., Plos One 2011 : TrueSeq like protocol , 5\$ per library vs 50\$ for Illumina TrueSeq mRNA
- The amount of reagents Illumina TrueSeq and Script-seq protocols can be reduced to half (25\$/Library)

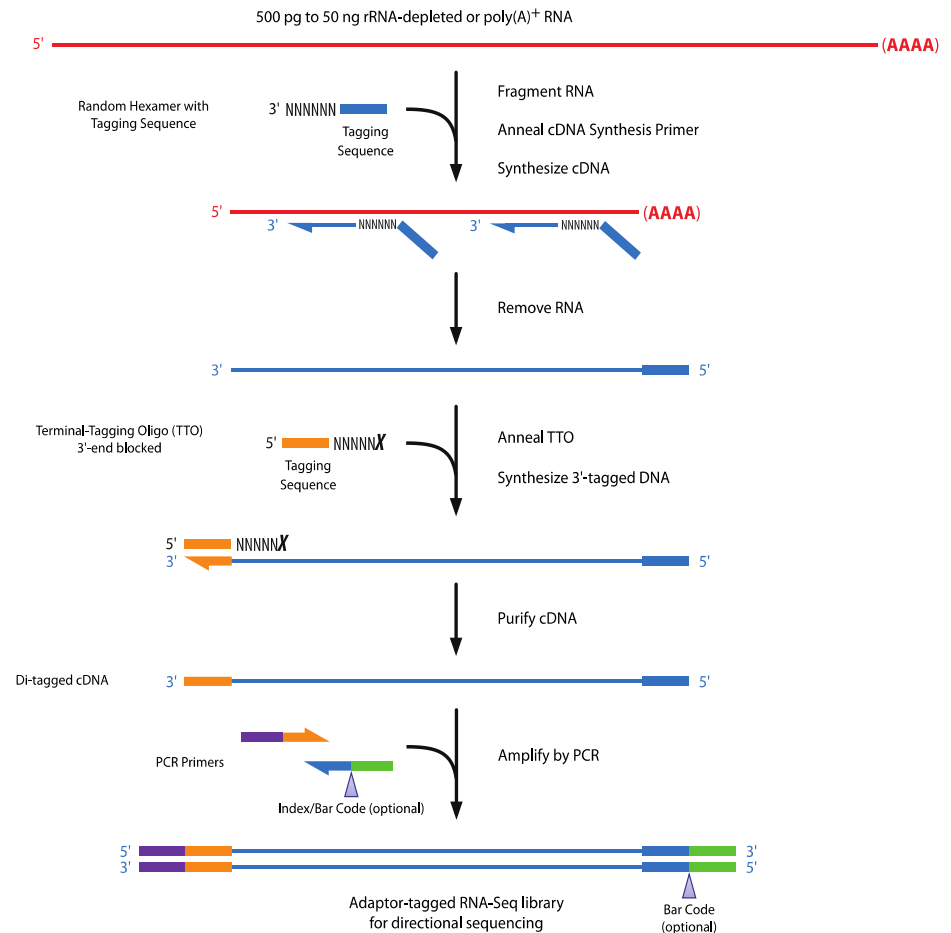
# Q: What is the effect of my favorite stress, mutation on global gene expression ?



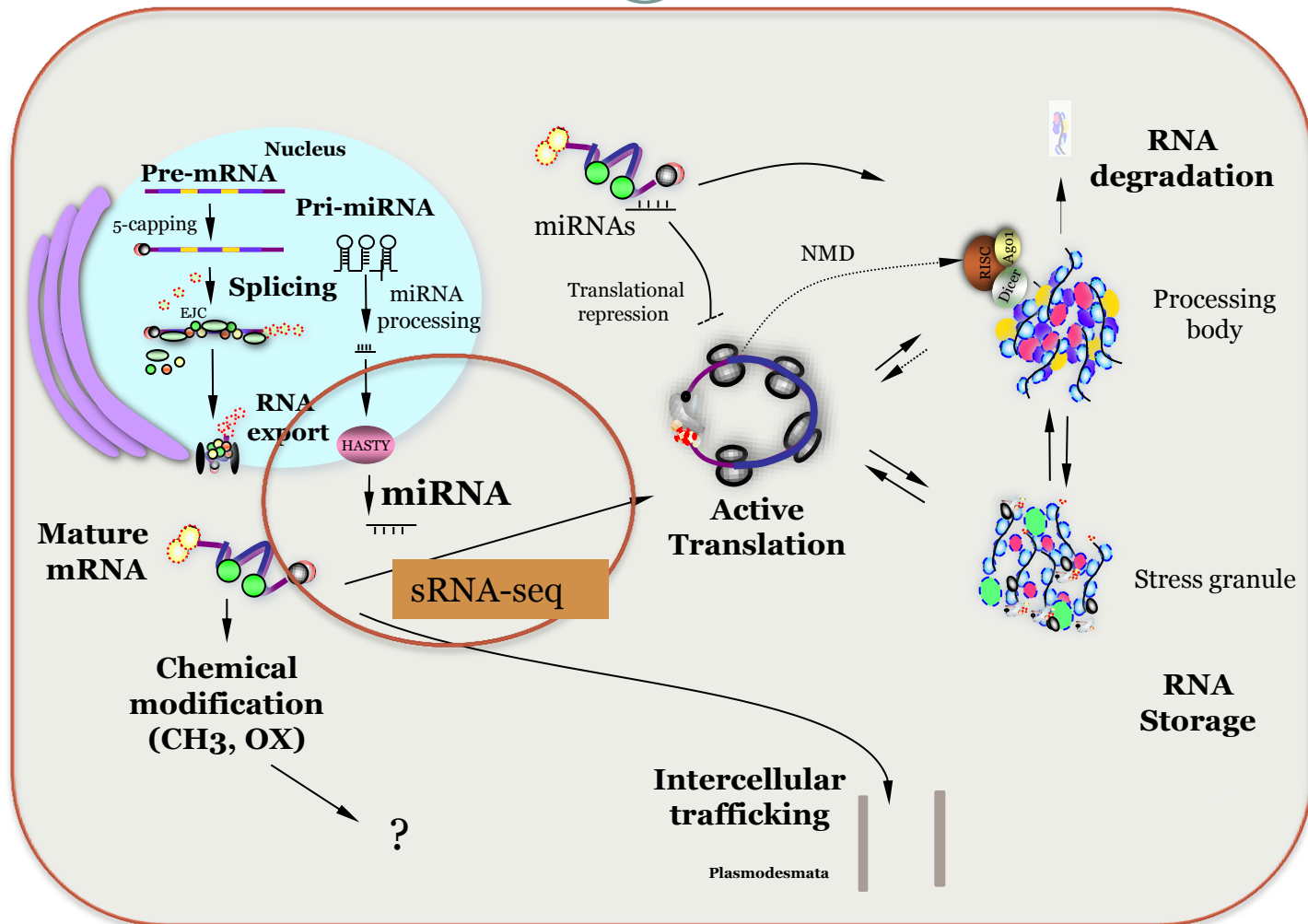
# TOTAL RNA SEQ



- Illumina Script-seq V2
  - Generally used after ribosomal RNA depletion (Ribo-zero kit)
  - Ligation of adapter on single stranded cDNA
  - Requires less RNA (50-500ng)
  - Less sensitive to RNA degradation
  - Sensitive to gDNA contamination
  - Much faster !



# Q2: WHICH miRNAs ARE EXPRESSED IN A PARTICULAR CONDITION/ SPECIES



# sRNA-SEQ



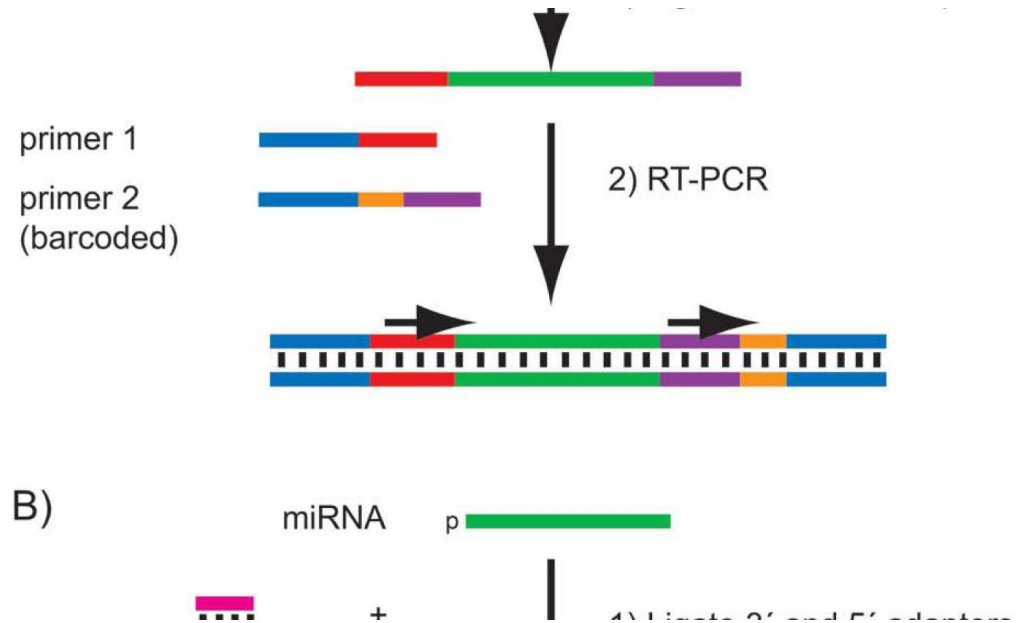
## ○ sRNA features :

- ✦ Narrow size distribution,
  - Do not require fragmentation, can be size selected
- ✦ 5' phosphate end, 3' O-methylated (plant)
  - 5'P end can be used to ligate selectively sRNA

# sRNA-SEQ



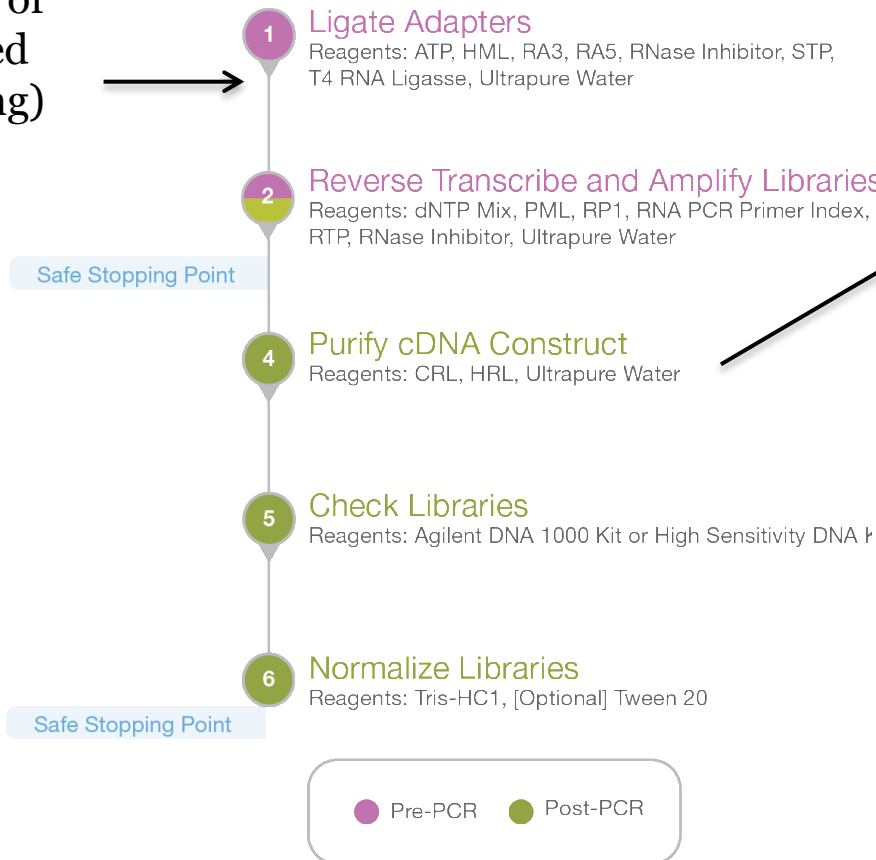
- Ligation of adapter directly on the end of the sRNA
- Uses the 5' phosphate end to select for miRNA
- Very sensitive to RNA degradation !
- Require a high amount of RNA.



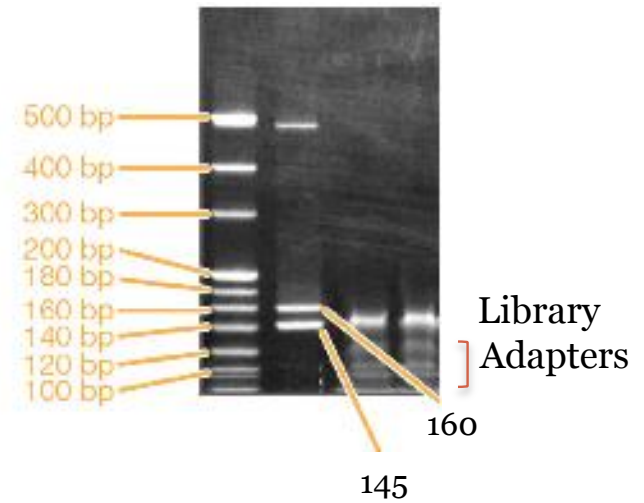
# sRNA-SEQ

## Illumina Small RNA Library Prep kit

Total RNA or  
size selected  
sRNA (50ng)



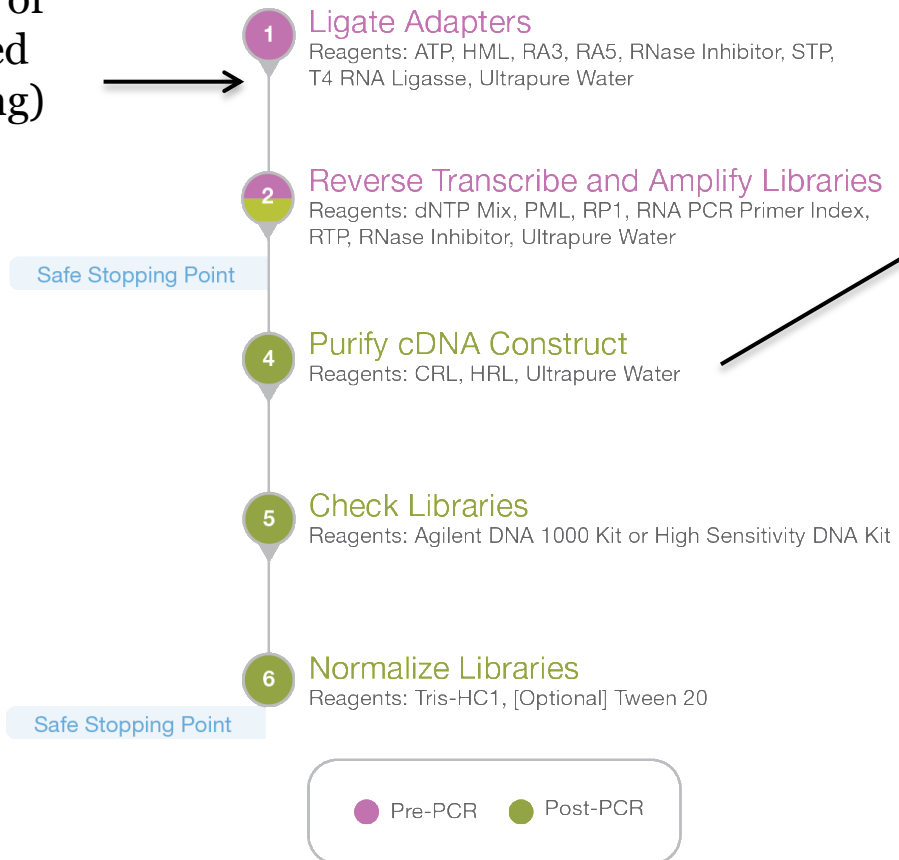
Separate library from  
contaminant on PAGE or  
Pippin prep



# sRNA-SEQ

## Illumina Small RNA Library Prep kit

Total RNA or  
size selected  
sRNA (50ng)



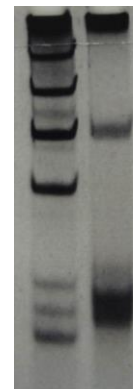
Separate library from  
contaminant on PAGE or  
Pippin prep



# sRNA-SEQ

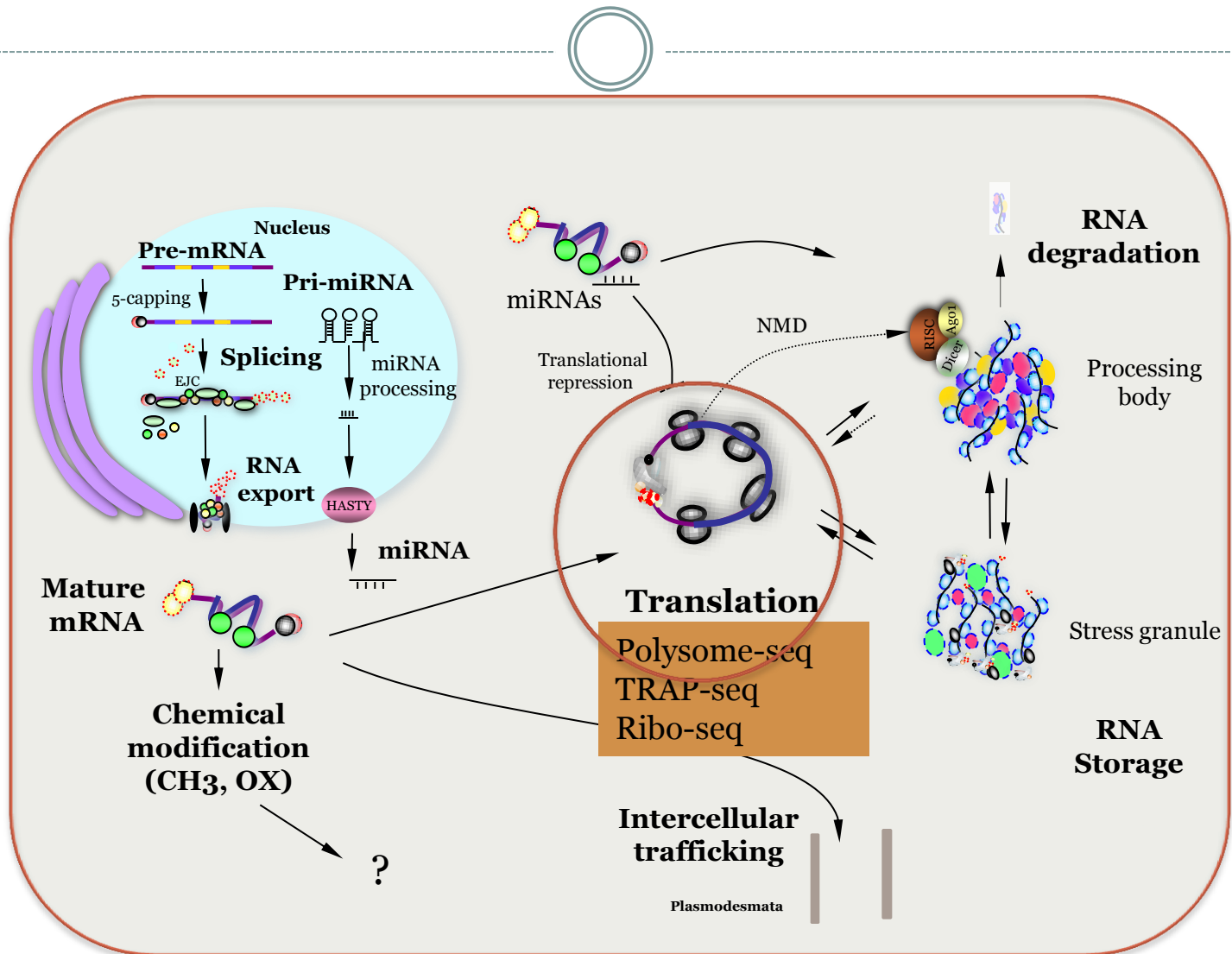


- A few recommendations:
  - Work with high integrity (RIN $\geq$ 9) pure RNA
  - Extract with Trizol or miRNA kit (eg miRVana Ambion)
  - Start with a high amount of material ( $> 10\mu\text{g}$ )
  - Size select on a 15% denaturing UREA TBE PAGE gel to enrich for 18-25nt long RNA

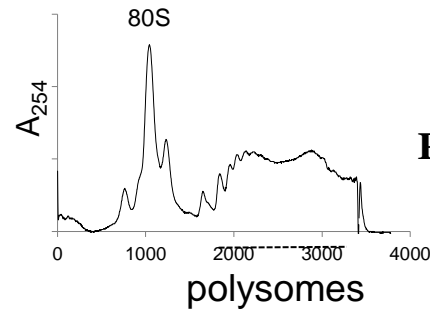
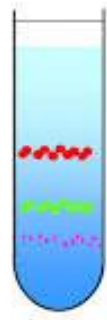
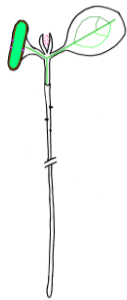


← Small RNA

# Q: Understand translation dynamics in response to an environmental stress



# CAPTURE ACTIVELY TRANSLATING MRNA

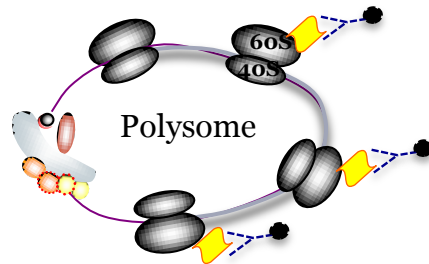


**Polysomal RNA**



**PolyA+ RNA-seq  
=Translatome**

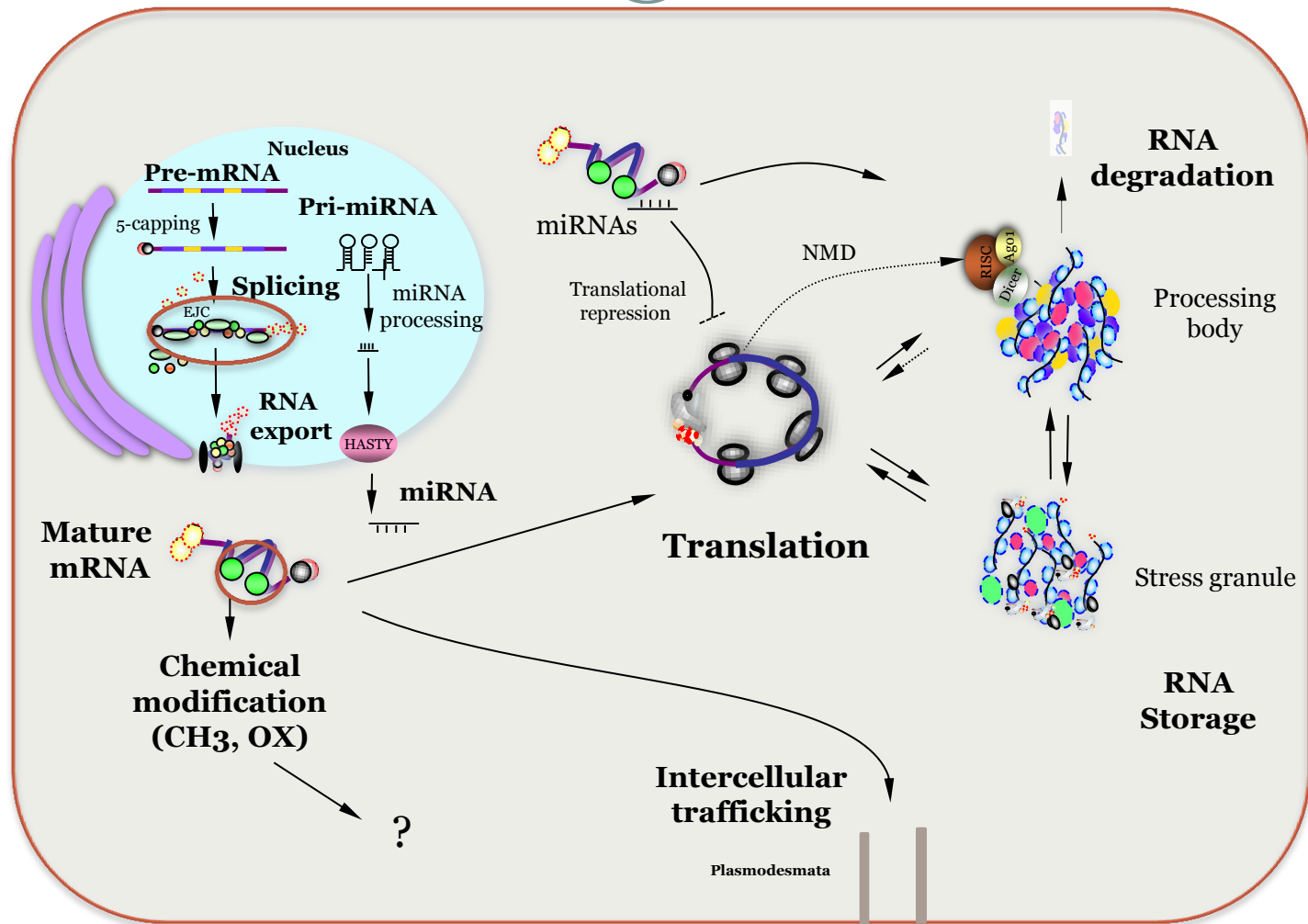
Translating Ribosome Affinity  
Purification  
(TRAP)



**Polysomal RNA**

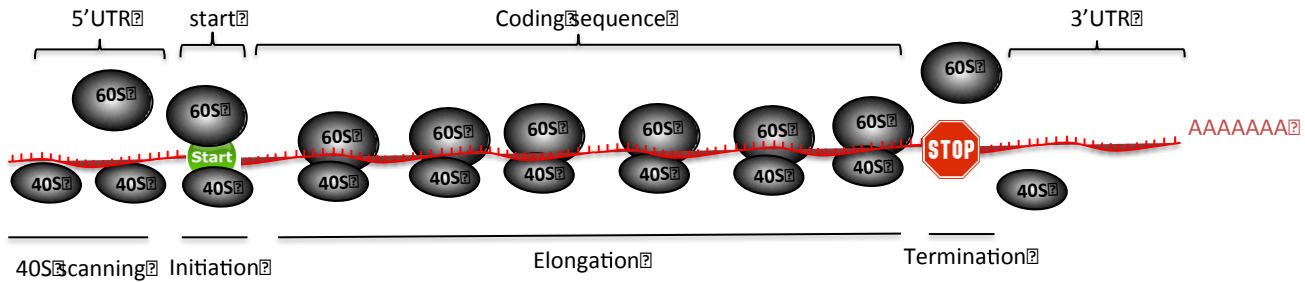


# Q: Which RNA species are bound to an RNA binding protein ?

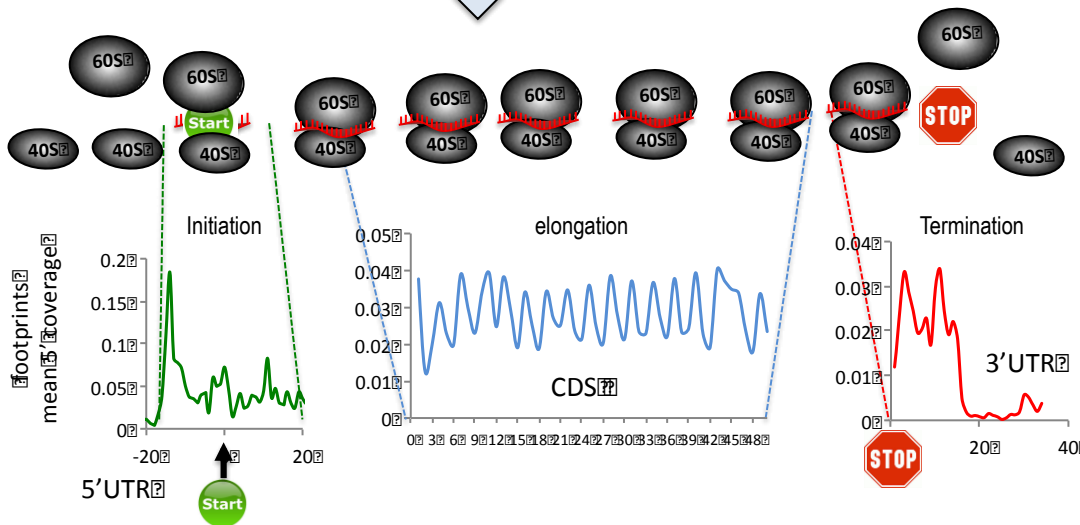


# Q: WHAT ARE THE MECHANISMS OF TRANSLATIONAL REGULATION ?

## Ribo-Seq : Sequencing of ribosome footprints



1. RNase1 digestion
2. Footprint library preparation
3. Sequencing & mapping to genome



Coverage plots in 5'UTR, CDS and 3'UTR  
 Arabidopsis seedling roots  
 35S:FLAG-RPL18

Bazin et al., under review

# RIBOSOME FOOTPRINT SEQUENCING: RIBO-SEQ

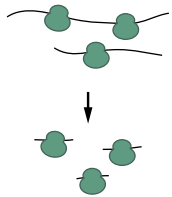


- Ribosome footprint features :
  - ✦ Size distribution : 26-32nt
    - Require optimal RNase1 digestion
  - ✦ 5' phosphate end,
    - 5'P end can be used to ligate RNA adapter

# RIBOSOME FOOTPRINT SEQUENCING: RIBO-SEQ

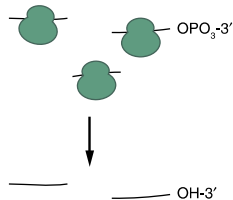
- Protocols :
  - Ingolia et al., 2012 Nat. Prot.
  - Illumina ArtSeq Kits

## RNase1 digestion

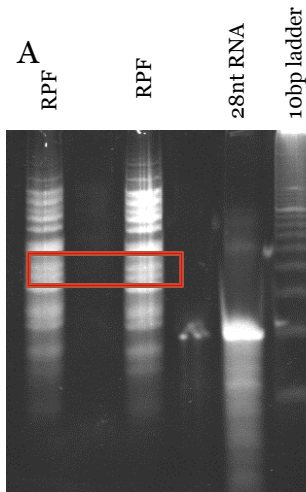


Steps 1–17: cell lysis,  
nuclease footprinting and ribosome recovery

## Footprint purification

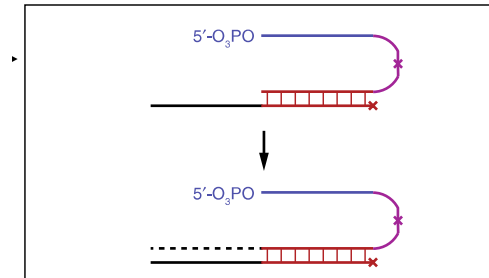


Steps 18–29: footprint fragment purification

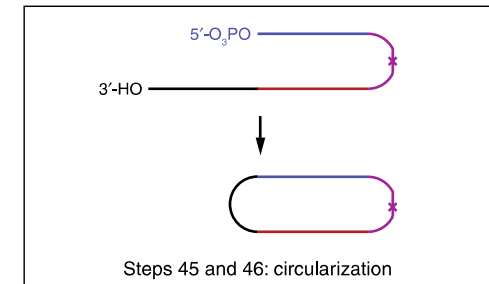


Juntawong et al., 2015

## Reverse transcription

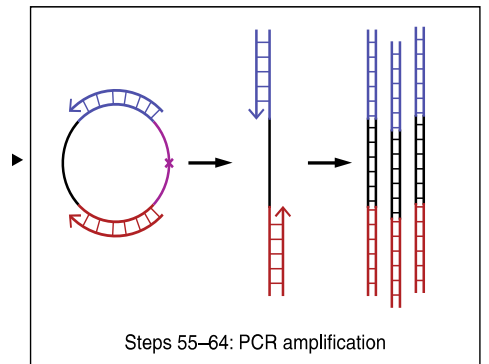


## Circularization



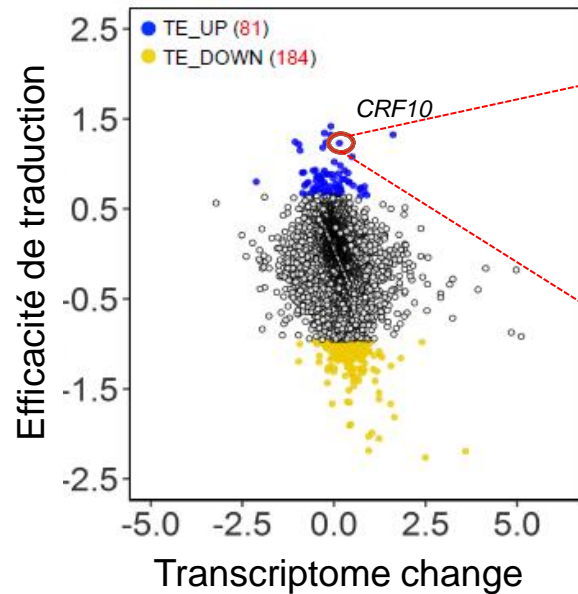
Steps 45 and 46: circularization

## PCR

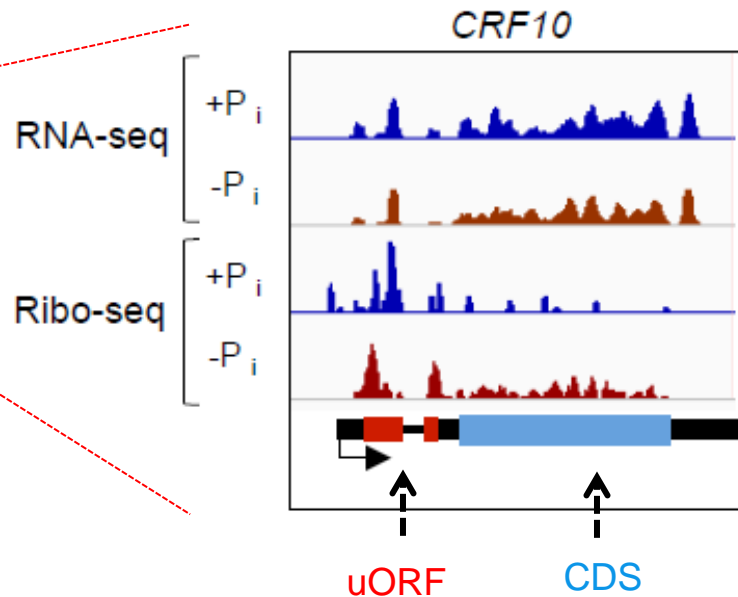


Steps 55–64: PCR amplification

# RIBOSOME FOOTPRINT SEQUENCING: RIBO-SEQ



✓ Régulations traductionnelles des ARNm à l'échelle du génome



✓ Mécanismes de régulation: rôles des uORF

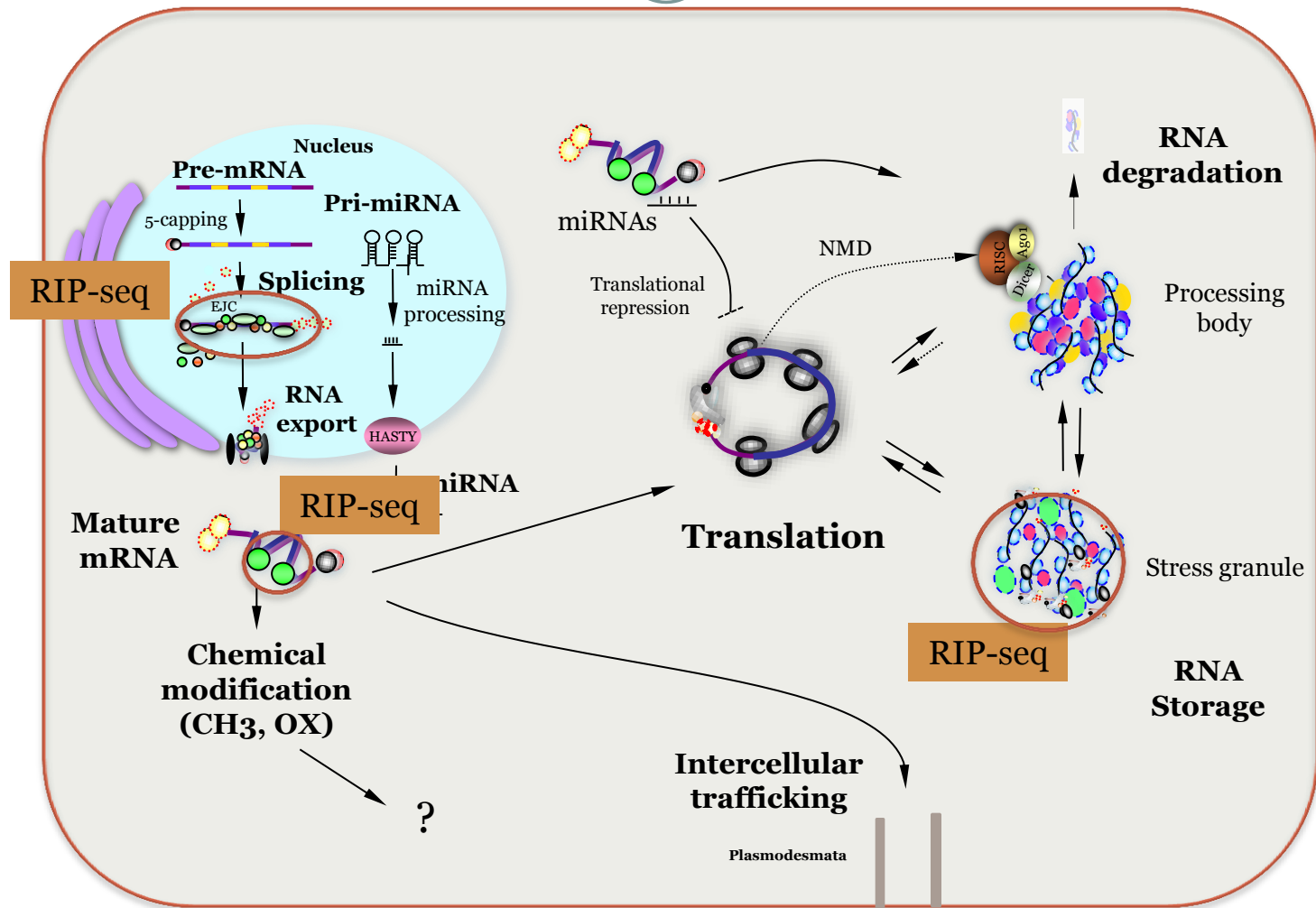
Bazin et al., under review

# RIBOSOME FOOTPRINT SEQUENCING: RIBO-SEQ



- A few recommendations:
  - Adjust RNase1 digestion for each tissue type/Species
  - Use a standardized protocol, always the same reagents (specially RNase1 batch)
  - Start with a relatively high amount of material ( $\geq 10\mu\text{g}$  of RNase1 digested RNA fragment )
  - Use ribo-zero kit to remove rRNA fragments from the ribosome footprints (can make up to 90% of the final read number!)

# Q: Which RNA species are bound to an RNA binding protein ?

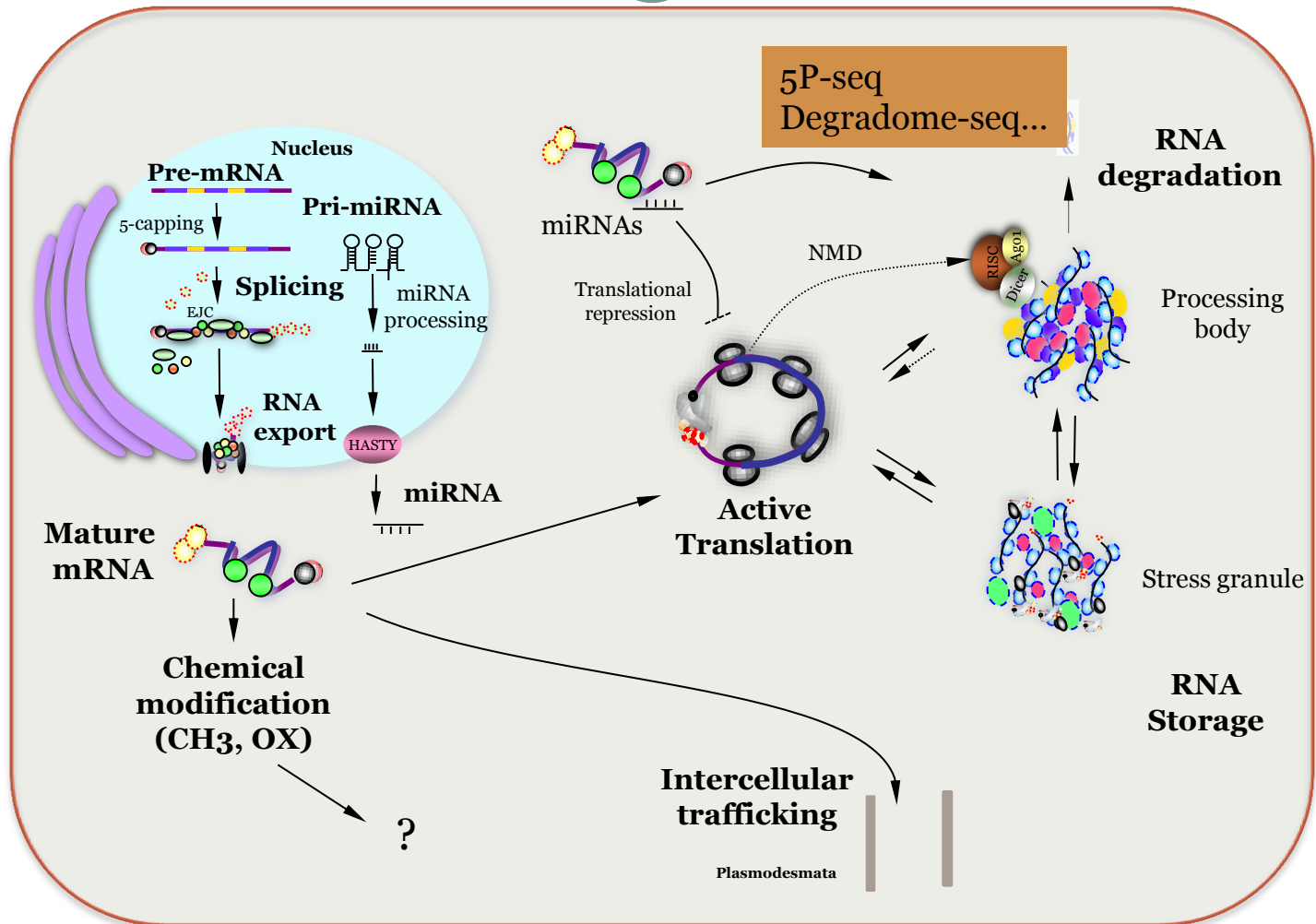


# Q: Which RNA species are bound to an RNA binding protein?



- Full length target RNA sequencing
  - RIP-seq : RNA immunoprecipitation
    - ✦ Analysis by RNA-seq of the RNA which co-immunoprecipitate with the target protein.
- RNA binding protein footprints
  - CLIP-seq : Cross-linking Immunoprecipitation :
    - ✦ Tissue cross-linking, immunoprecipitation and digestion of the unprotected RNA. Sequencing of the protected fragments is done using protocol similar to small RNA seq
  - Small RNA-seq reanalysis : Some small RNA sequencing contains footprint of RNA Binding Protein (eg PPR)

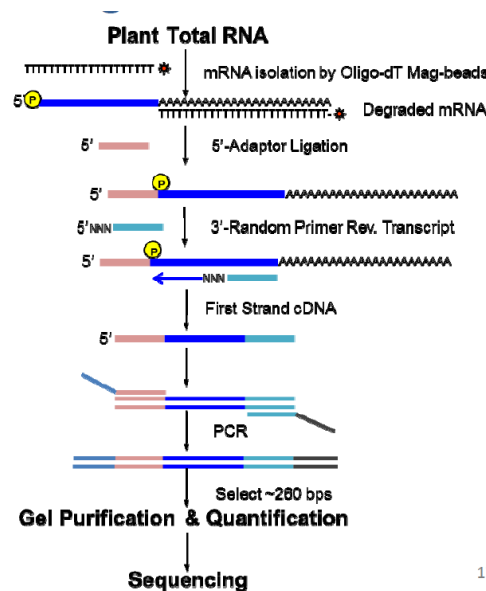
# Q: Find the mRNA targets of microRNAs



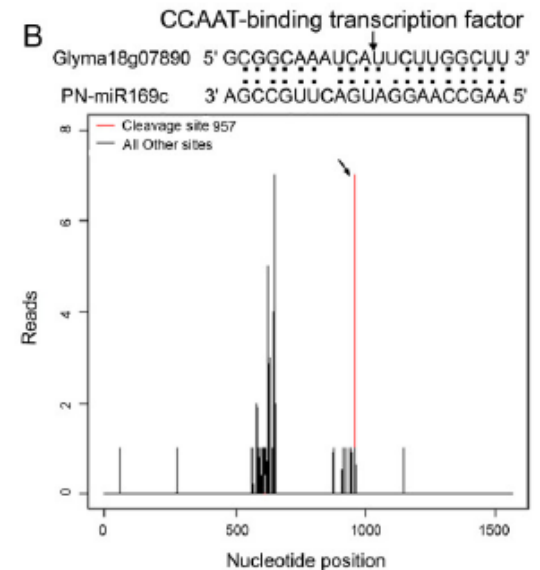
# Q: Find the mRNA targets of microRNAs



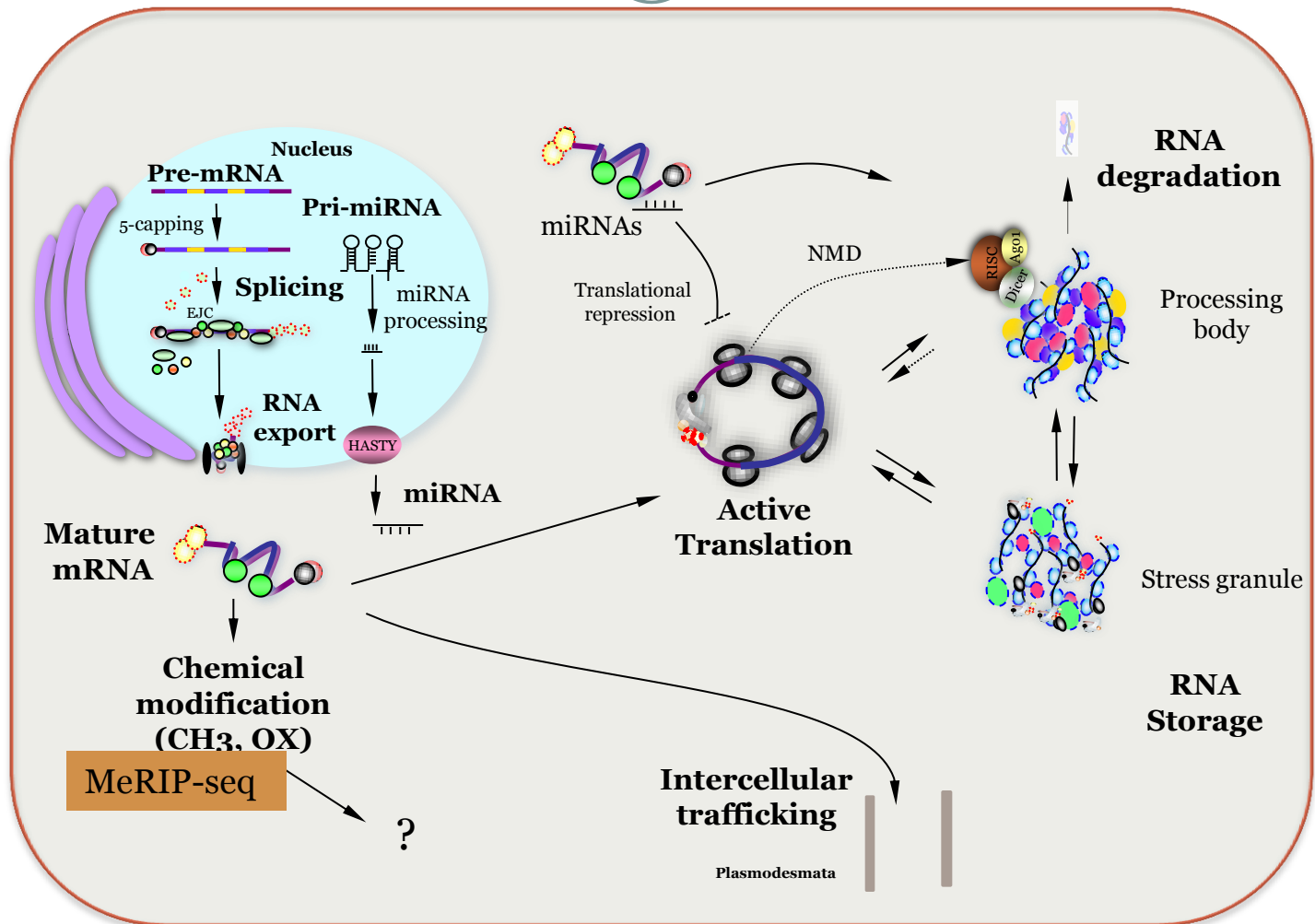
- Degradome sequencing/ 5P seq : Sequencing of mRNA decay intermediate
  - Based on chemical features of mRNA fragments :
    - ✦ PolyA tail / 5'P end



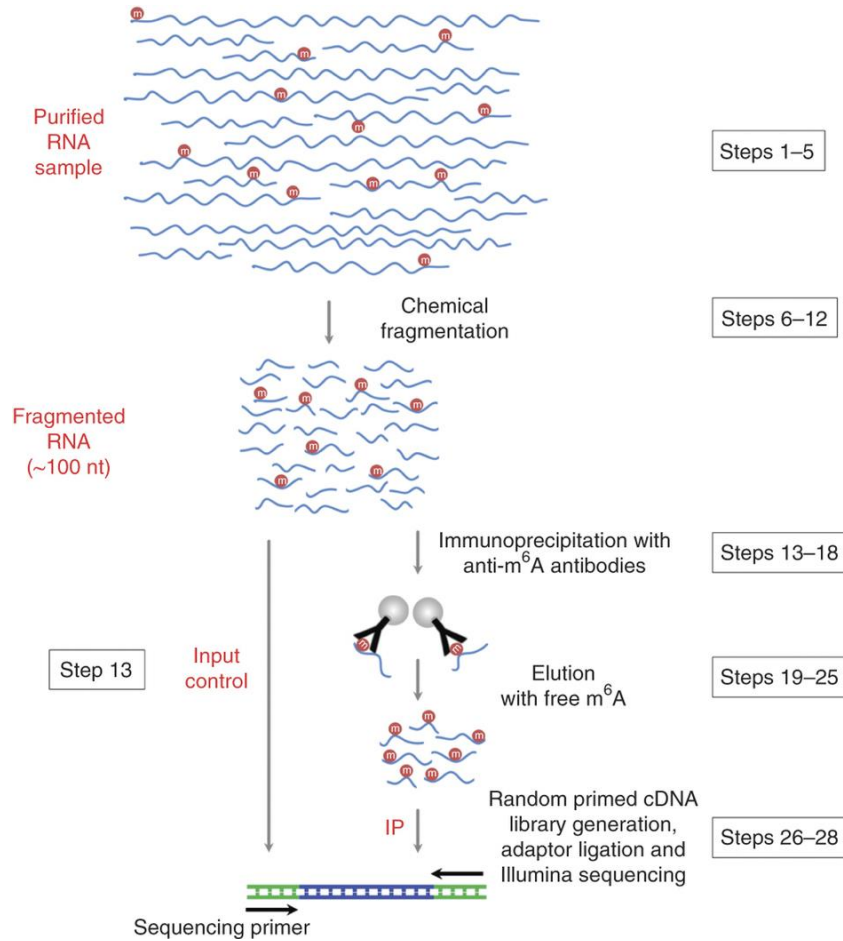
17



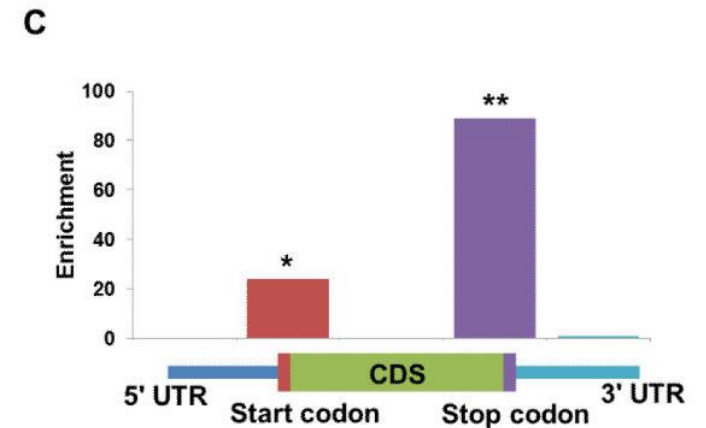
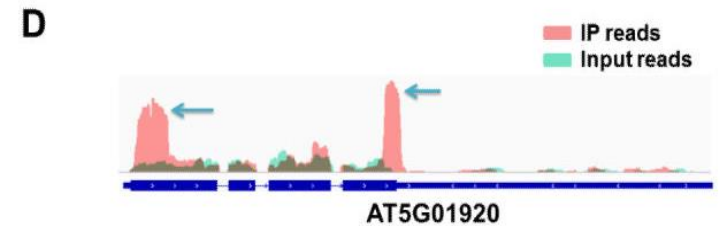
# Identify/quantify methylated RNAs



# MERIP-SEQ



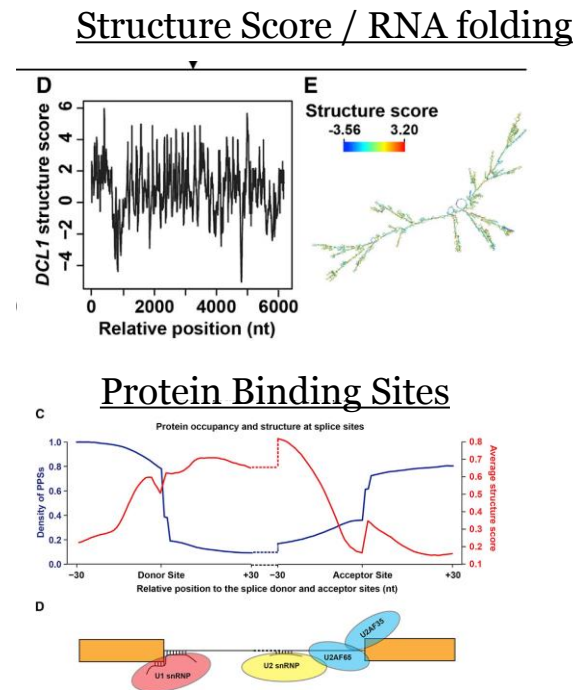
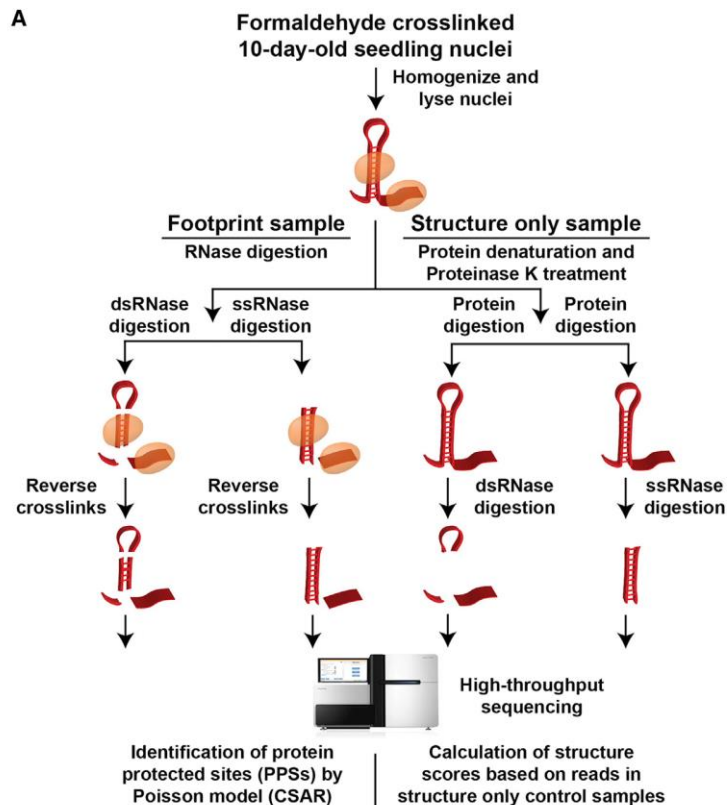
Dominissi et al., Nat. Prot 2013



Luo et al., Nat. Com 2014

# PROBE RNA STRUCTURE AND PROTEIN BINDING SITES ?

- Protein interaction profile sequencing : PIP-seq
  - Based on crosslinking and selective digestion of Protein / RNA complexes

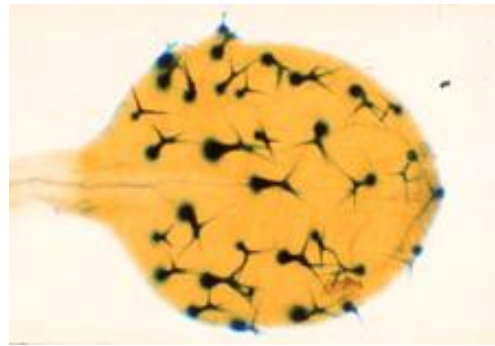
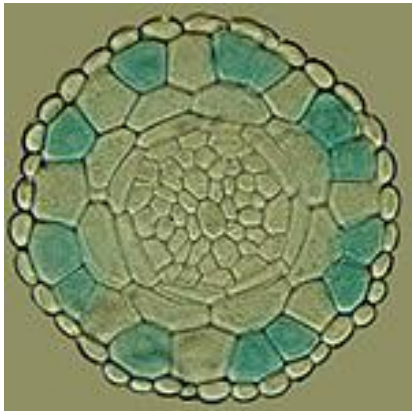


Anderson et al., 2016 (Current Prot. Plant. Biol)  
Gosai et al., 2015 (Mol. Cell)

# RNA POPULATIONS ARE DIFFERENTS IN DISTINCT CELL TYPE



- Gene expression is highly variable in space



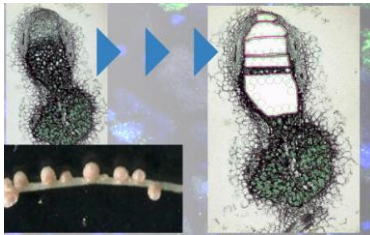
pGL2::GUS (Rerie et al., 1994)

# HIGH-THROUGHPUT METHODS FOR CELL-SPECIFIC GENE REGULATION ASSESSMENT



- Dissection-based and sorting-based methodologies :

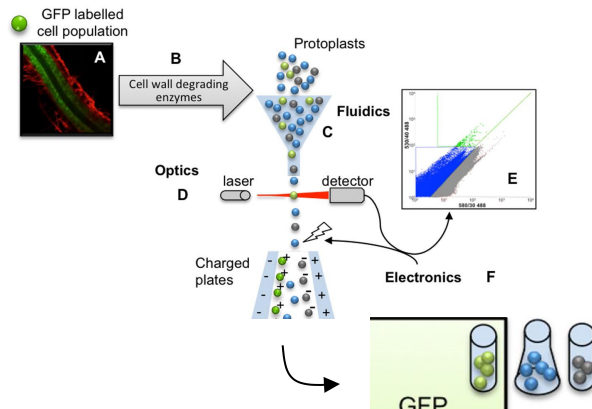
- Laser Capture Microdissection (LCM)



P.Gamas (LIPM, Toulouse)

- Fixed tissue
- Time consuming
- Access to whole cell content

- Fluorescence activated cell sorting (FACS)

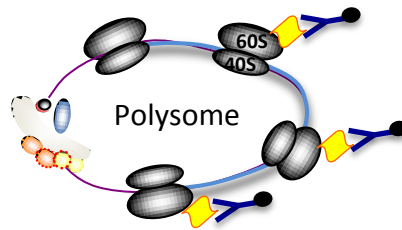
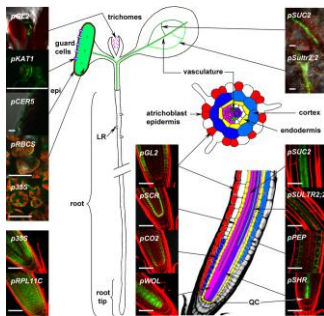


- Live tissue
- Give little RNA
- Stress during protoplast isolation
- Access to whole cell content

# HIGH-THROUGHPUT METHODS FOR CELL-SPECIFIC GENE REGULATION ASSESSMENT



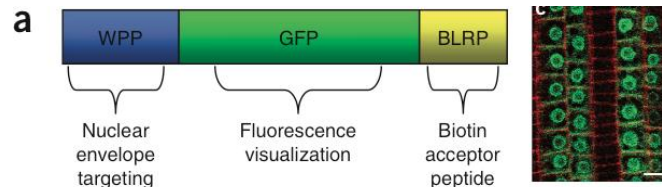
- Affinity purification and immunoprecipitation :
  - Translating Ribosome Affinity Purification (TRAP)



Mustroph et al. PNAS 2009

- Live tissue/*in vivo*
- Ribosome associated RNA
- Require transgenics

- Isolation of Nuclei Tagged in Specific Cell Types (INTACT)



Deal and Henikoff., 2013

- Live tissue/*in vivo*
- Nuclear RNA/Chromatin
- Require transgenics

# RNA-SEQ FROM LIMITED AMOUNT OF RNA

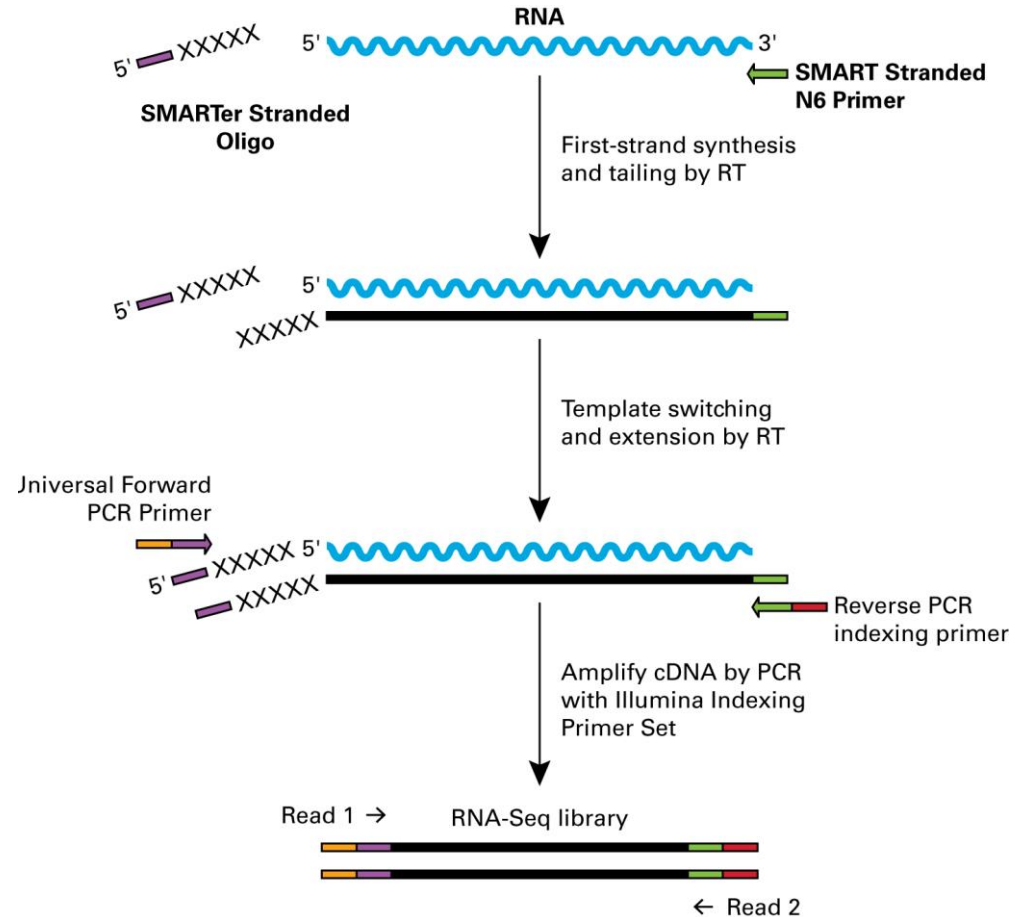


- **Cell-specific Approach = limited amount of RNA**
  - Require optimization of RNA protocols
    - ✦ If total RNA quantity is  $\geq 10$  ng :
      - Protocols / Kits optimized for reduced RNA amount
        - Illumina Script-seq V2
        - Smarter Stranded (Clontech)
        - NEBNext<sup>®</sup> Ultra<sup>™</sup> II Directional RNA Library Prep Kit for Illumina
        - Townsley et al., 2015 (Front. Plant Sciences)

# RNA-SEQ FROM LIMITED AMOUNT OF RNA



- Smarter Stranded (Clonetechn)
- Very simple design (Smart!)
- Based on template switching activity of RT
- Reduced purification steps



# RNA-SEQ FROM LIMITED AMOUNT OF RNA

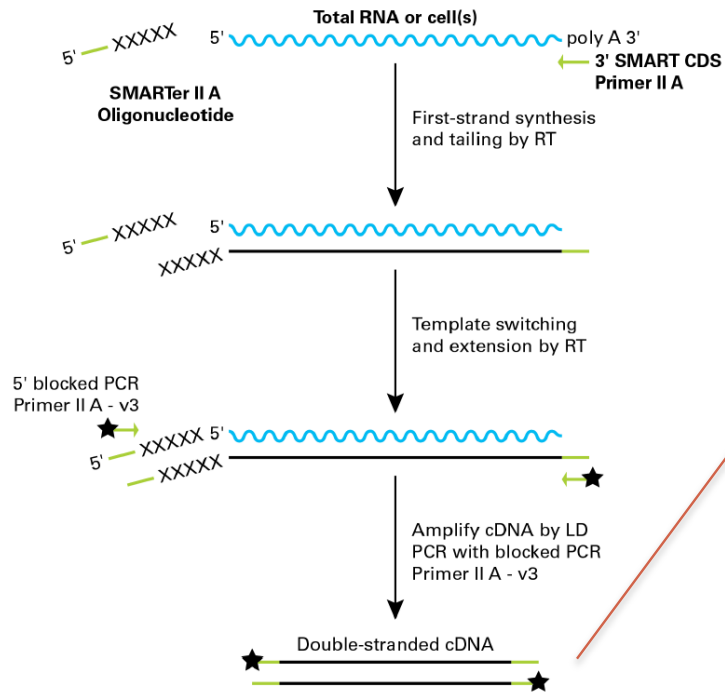


- **Cell-specific Approach = limited amount of RNA**
  - Require optimization of RNA protocols
    - ✦ If total RNA quantity is  $\leq 1$  ng :
      - Protocols / Kits including an additional RNA amplification step
        - Smarter Ultra Low Input RNA (Clontech)
        - Smart-seq2 (Picelli et al., Nat. Prot. 2013)
        - Ovation RNA-seq System V2 (NuGEN)

# RNA-SEQ FROM LIMITED AMOUNT OF RNA



## Smarter Ultra Low Input RNA (Clontech)



## « DNA seq library »

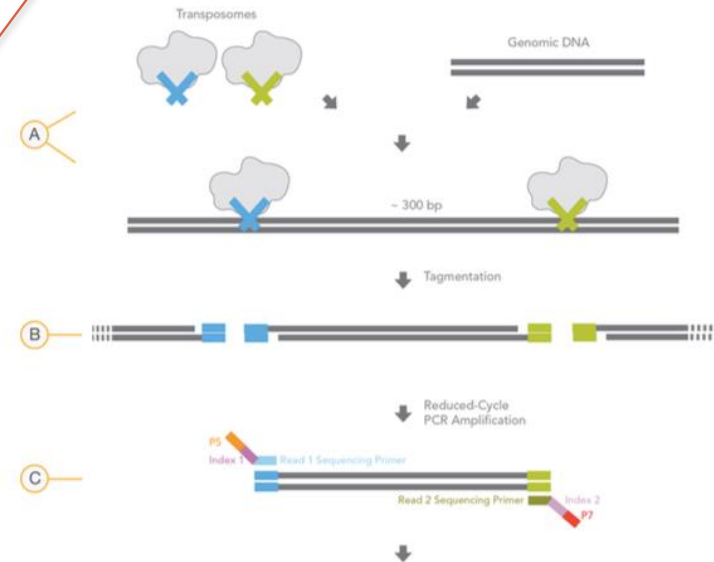
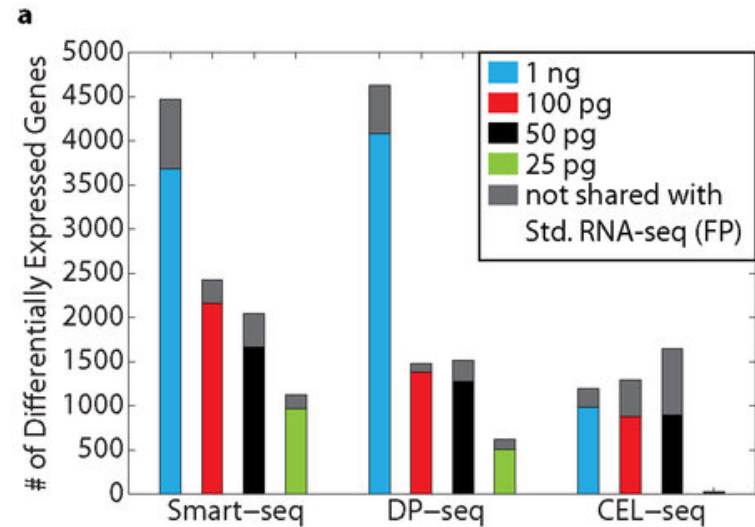
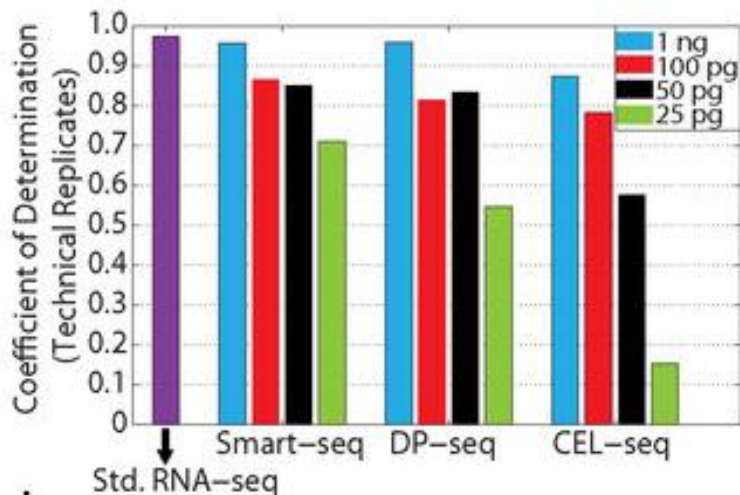


Figure 2. Flowchart of SMARTer cDNA synthesis. The SMARTer II A Oligonucleotide, 3' SMART CDS Primer II A, and PCR Primer II A - v3 all contain a stretch of identical sequence.

# RNA-SEQ FROM LIMITED AMOUNT OF RNA

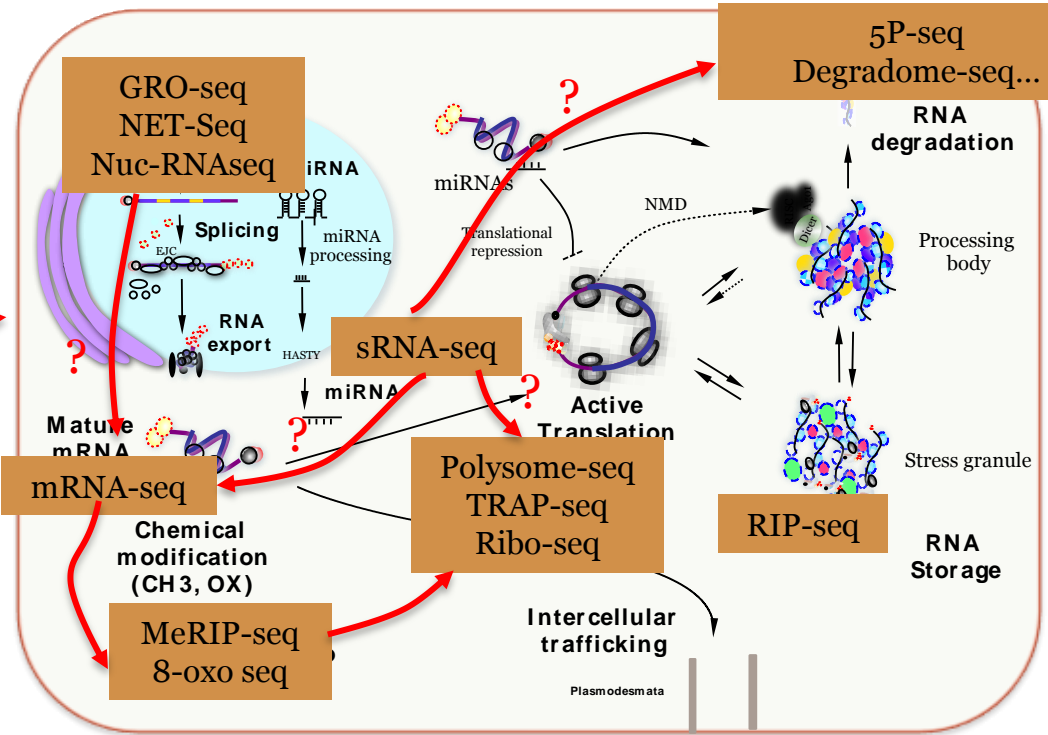
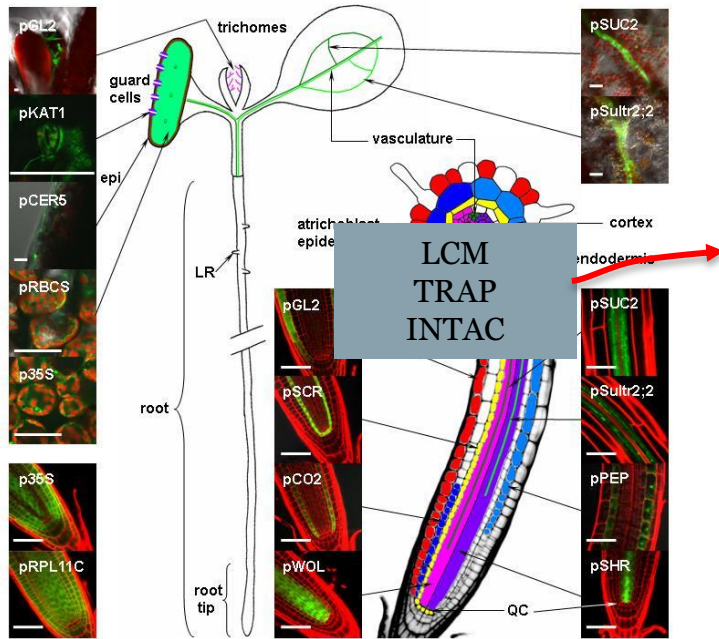


- RNA amplification from very low amount of RNA can introduce variability in the data



Barghava et al., 2014

# GENOME-SCALE, CELL-SPECIFIC MONITORING OF MULTIPLE GENE REGULATION TIERS USING RNA-SEQ



# CONCLUSIONS



- An RNA-seq library is just cDNA fragments between adapters.
- There are always biases in genome-wide technologies including RNA-seq: Just keep the same bias for all samples
- One can sequence subpopulations of RNA based on their chemical feature, localization and protein associations
- By combining RNA-seq approaches and system biology, one can answer complex biological questions