

# RNA Sequencing Library Preparation

Sarah Boswell  
(she/hers)

Director of Sequencing Technologies, Laboratory of Systems Pharmacology

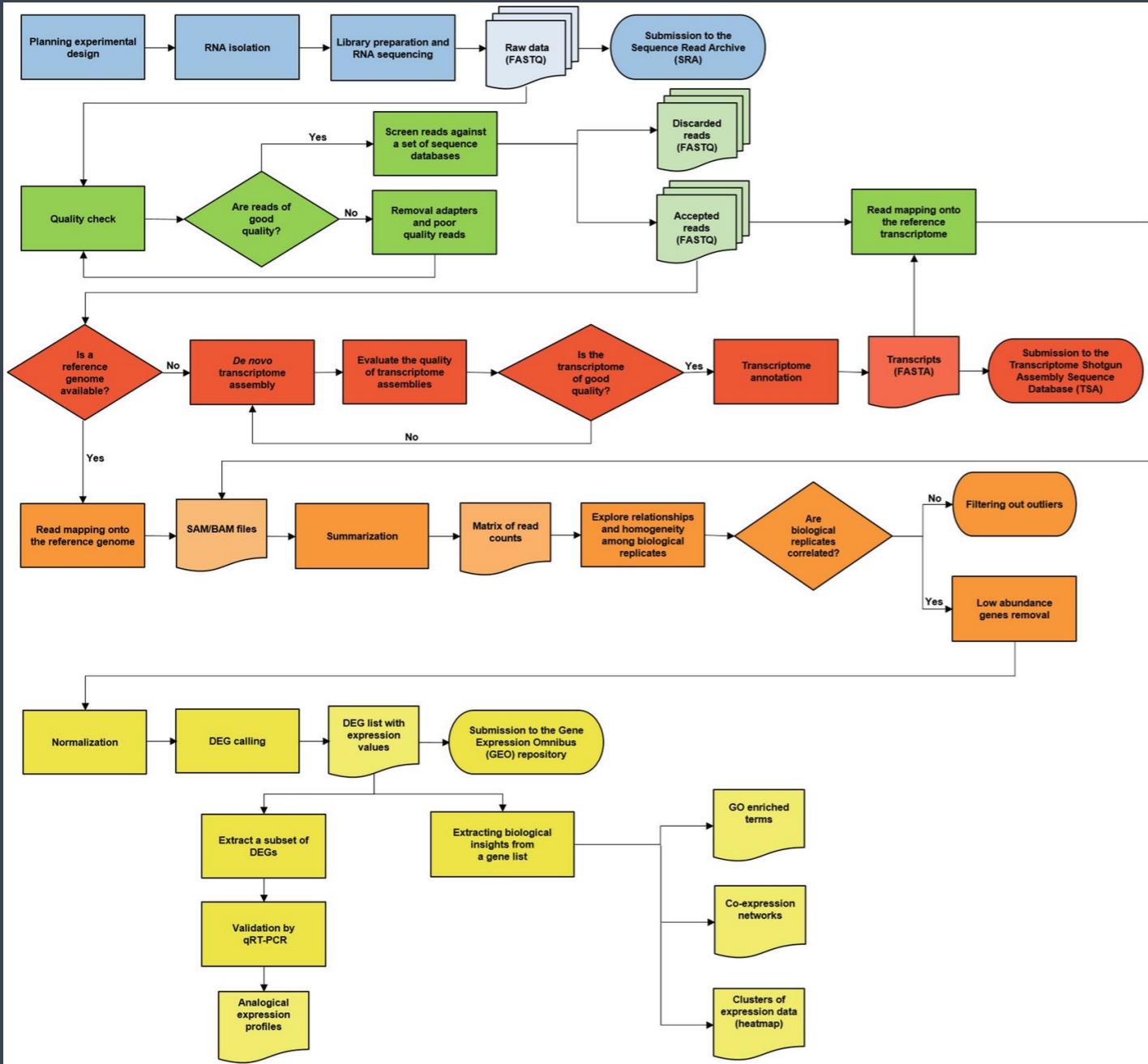
Director of the Single Cell Core, Harvard Medical School

Staff Scientist/Sequencing Specialist, Systems Biology, Springer Lab

[sarah\\_boswell@hms.harvard.edu](mailto:sarah_boswell@hms.harvard.edu)  
<http://scholar.harvard.edu/saboswell>



# RNA-seq Library Preparation



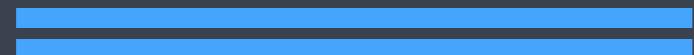
Experimental design &  
Sample prep are key



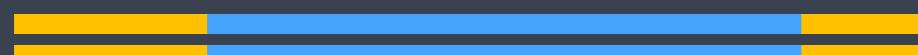
# Sequencing Methods

ALEXA-seq, Amp-seq Apopto-seq, ATAC-seq, AutoMeDip-seq, Bind-n-seq, Bisulfite-seq, CaptureSeq, CAP-seq, CAPP-seq, CEL-seq, ChIA-PET, ChIP-seq, ChiRP-seq, CiIP-seq, ClickSeq, CNV-seq, CLIP-seq, Degradome-seq, Div-seq, DGE-seq, DNA-seq, DNase-seq, d-seq, Exo-seq, FAIRE-seq, Fis-seq, FRAG-seq, FRT-seq, HITS-CLIP, Immune-seq, Gro-seq, indel-seq, MBD-seq, MeDIP-seq, Methyl-seq, microRNA-seq, mRNA-seq, Mnase-seq, MAINE-seq, NA-seq, NET-seq, NOME-seq, NSR-seq, Nuc-seq, PARE-seq, PAS-seq, Peak-seq, PhIP-seq, Purturb-seq, Protein-seq, QUARTZ-seq, RASL-seq, ReChIP-seq, Ribo-seq, RIP-seq, RIT-seq, RNA-seq, rSW-seq, SAGE-seq, SCRB-seq, Seq-Array, SMART-seq, Sono-seq, Sort-seq, STRT-seq, Tag-seq, Tn-seq, TRAP-seq

# Basic Library Preparation



Small piece of DNA or cDNA  
INSERT



INSERT + Sequencing Adapter  
Library

# RNA-seq: What is Transcriptomics?

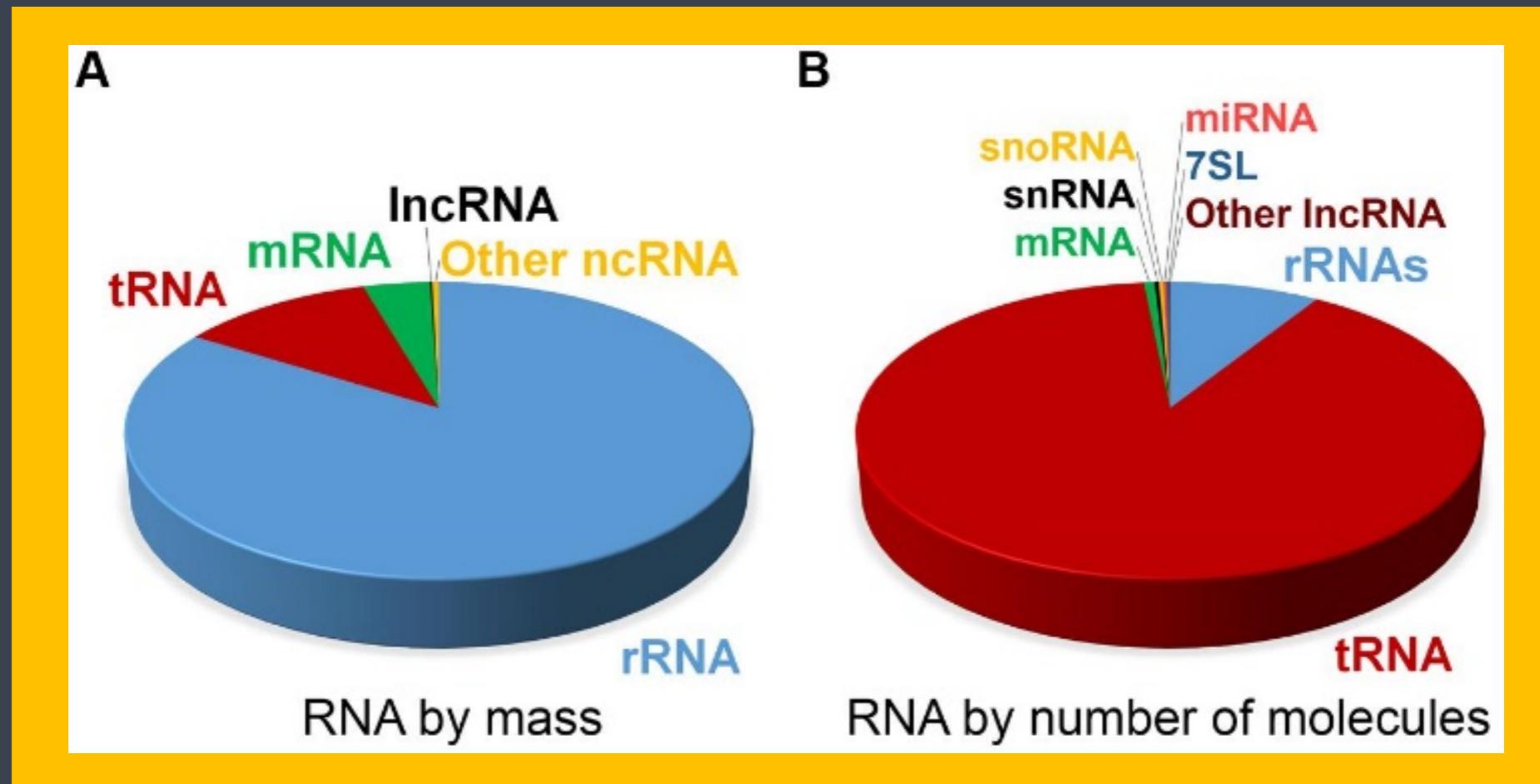
- Wikipedia

“The transcriptome is the set of all **messenger RNA** molecules in one cell or a population of cells.”

- National Cancer Institute Definition of Terms

“The study of **all RNA molecules** in a cell. Transcriptomics is used to learn more about how genes are turned on in different types of cells and how this may help cause certain diseases, such as cancer.”

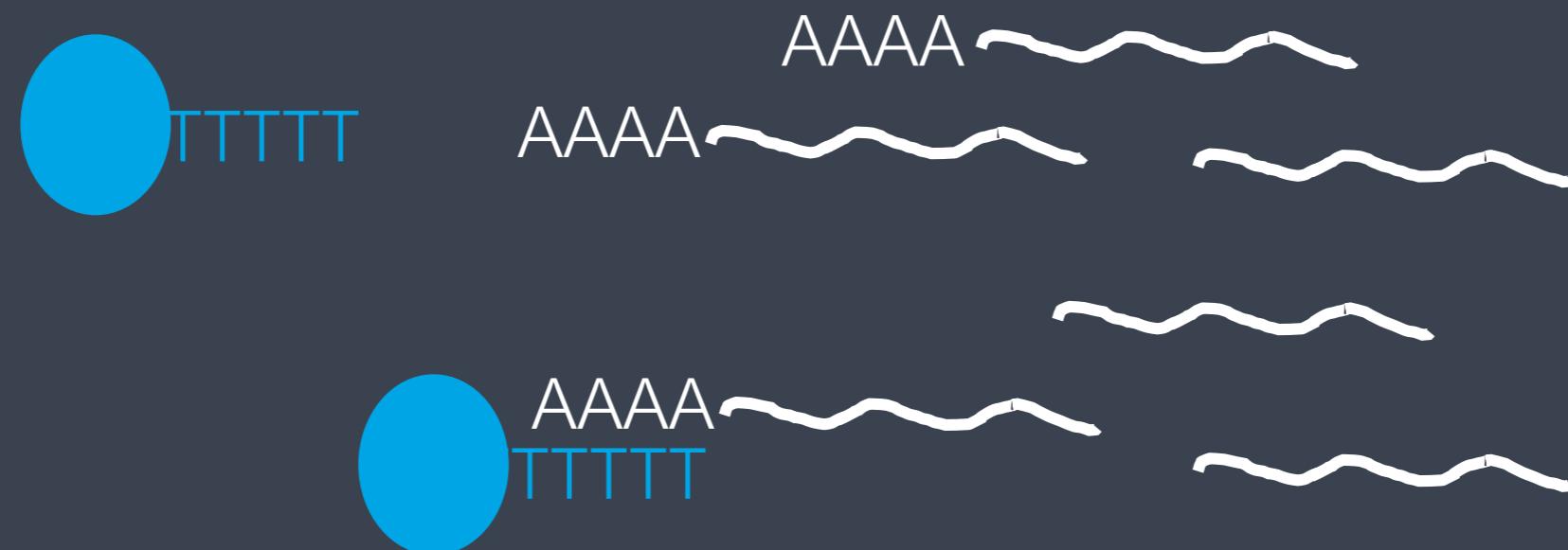
# Messenger RNA Enrichment



- PolyA tailed messenger RNA: mRNA-seq
- Total RNA (rRNA removed): “total” RNA-seq

# mRNA-seq: polyA Purification

- mRNA enrichment
- mRNA binds beads coated with oligo dT primer
- Non-polyadenylated transcripts are washed away

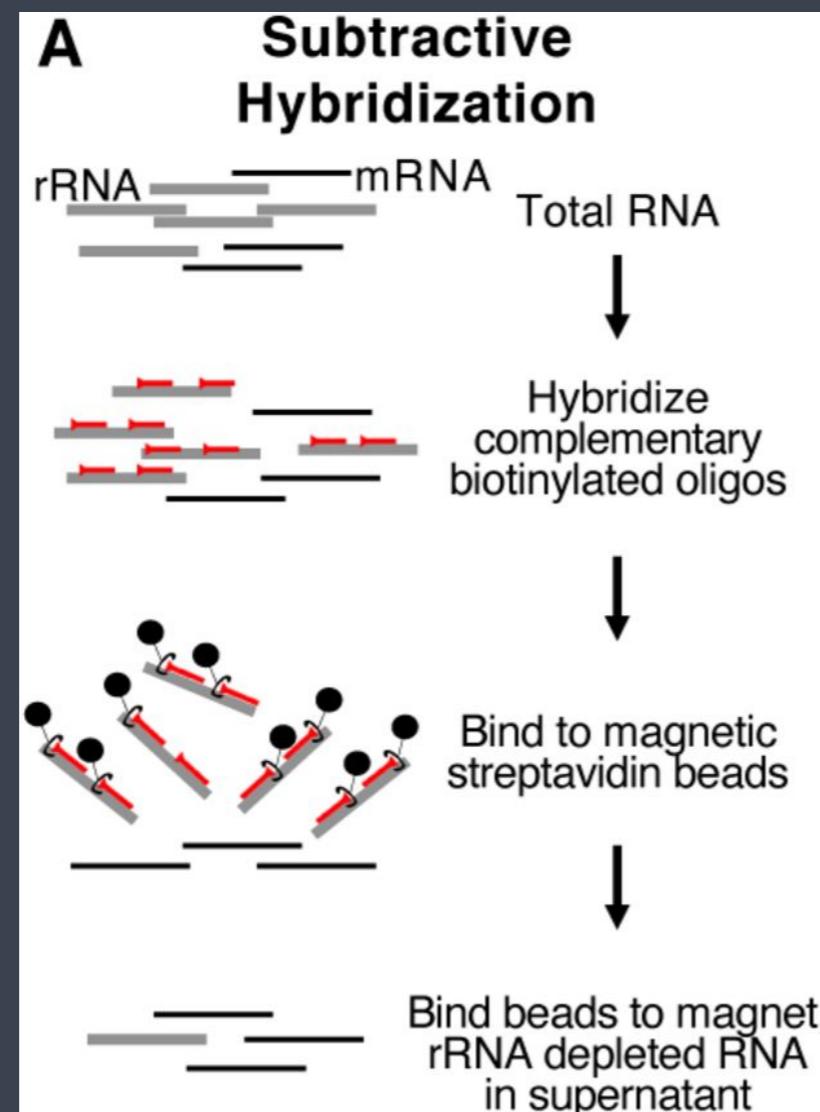


# Transcripts Lost in polyA Purification

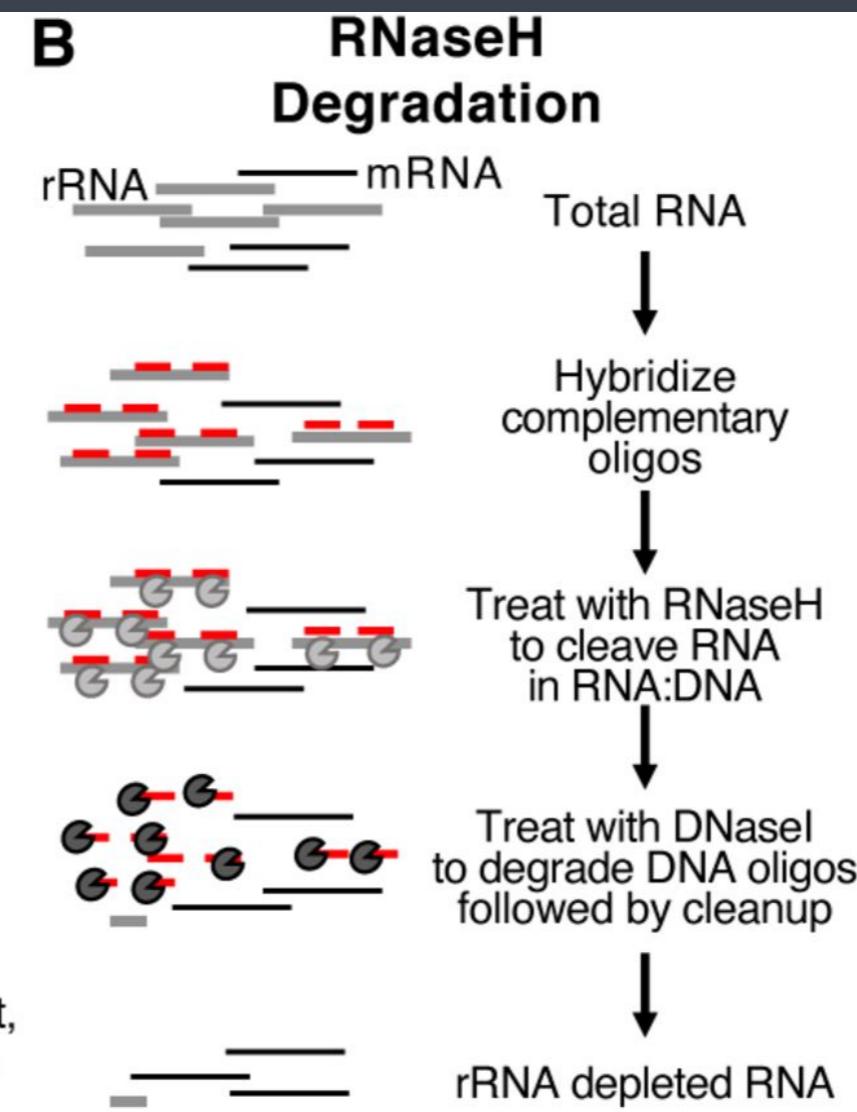
- Ribosomal/Transfer RNA
- Histone mRNA
- Long-noncoding RNA
- Nascent intron containing transcripts
- Micro RNA
- Degraded RNA
- Many viral transcripts
- Prokaryote/Bacterial transcripts
  - polyA is the degradation signal

# “Total” RNA-seq: rRNA Depletion

Illumina: TruSeq



KAPA: RiboErase



# Sequencing Library Preparation

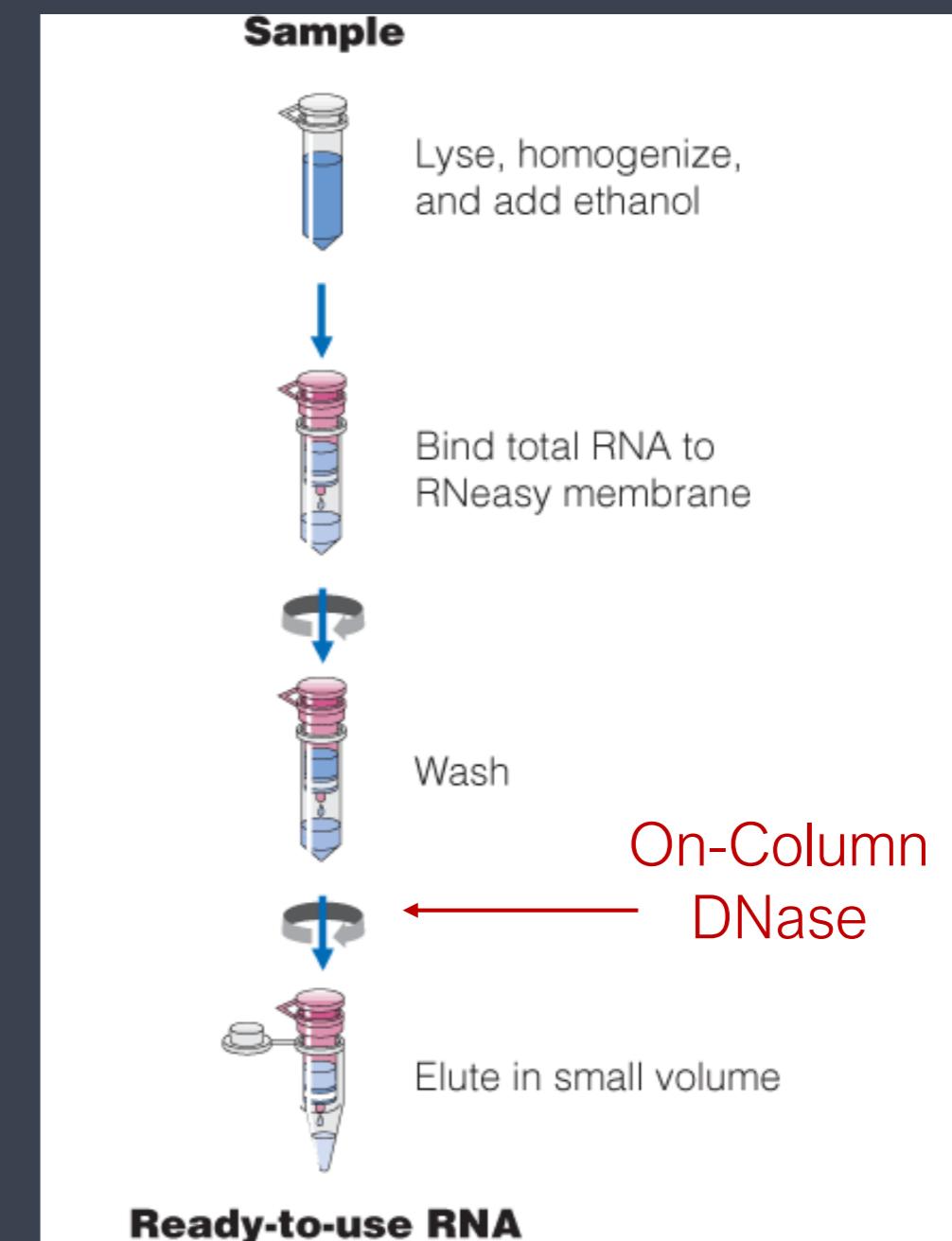
- Starting Material: purifying RNA
- RNA-seq library preparation
- Multiplexing & Sequencing
- Single cell / low input methods
- Capture sequencing

# RNA Extraction

- Cultured cells
  - Use your favorite column kit – Qiagen, Invitrogen, Zymo
  - For high throughput suggest bead based – MagMax
    - 96 well format column plates are also available
- Tissue samples
  - Dissect in cold room if possible
  - Use RNAlater solution to store tissue sample
  - Upon extraction perform homogenization in cold room

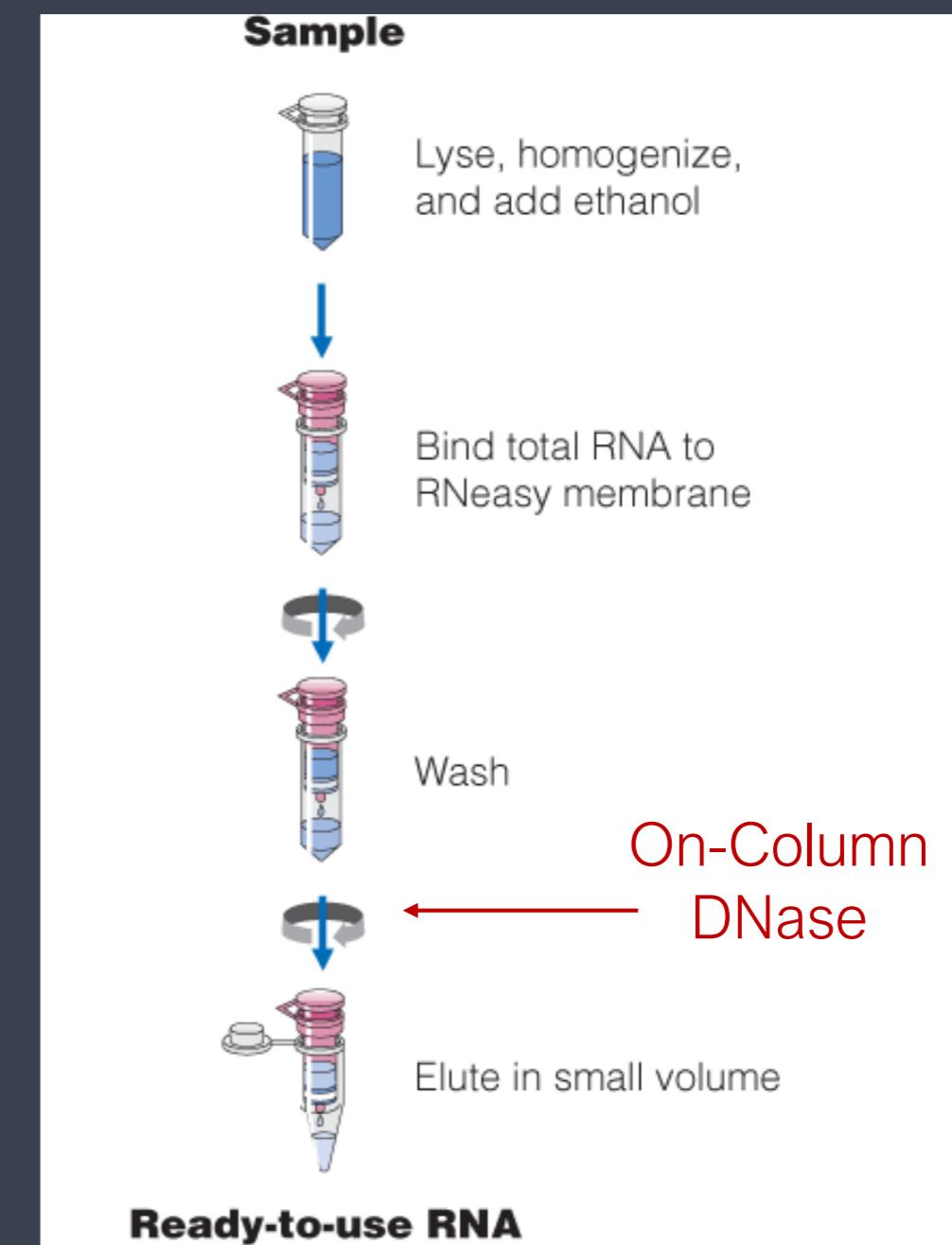
# RNA Extraction: Column Purification

- On-column DNase digestion
- Dry to remove excess ethanol
- Elute with warm water to increase yields



# Micro RNA Extraction

- Pay attention to the MW cut off of the purification column.
- Many of these are the same columns as used for total RNA purification.
- Key is using higher amounts of ethanol to bind smaller material.
  - Standard RNeasy protocol binds at 35% ethanol. Micro RNA will bind at 60% ethanol

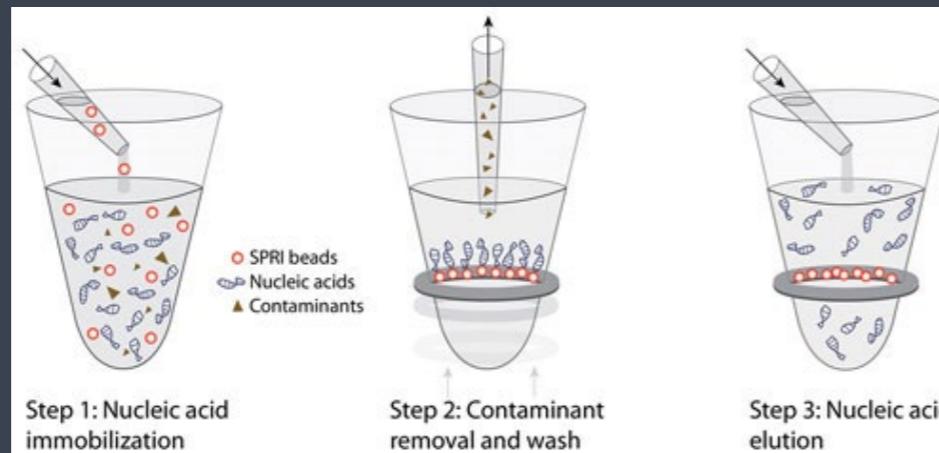


# RNA Extraction: Column Purification

- Standard columns/protocols have 100-200bp cut off.
  - Will result in loss of small RNA species
- Specialized columns and kits are available
  - microRNA
  - FFPE
  - Blood

HACK - Cut off size can be adjusted by changing the percent ethanol used for sample binding

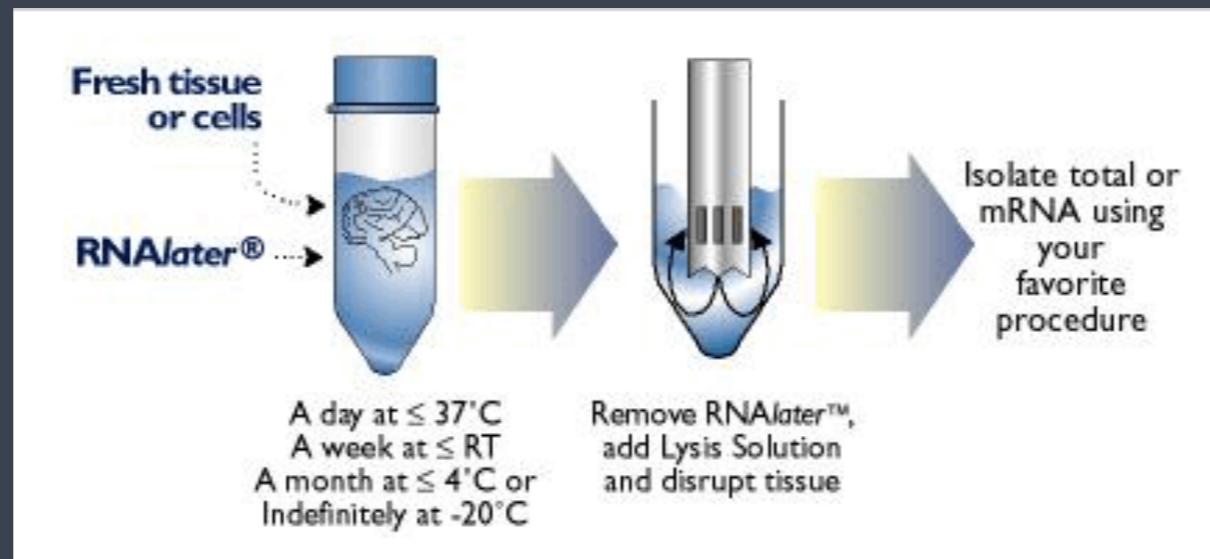
# RNA Extraction: Bead Purification



1. Binding
2. Wash
3. DNase
4. Re-bind
5. Wash
6. Elute

- Magnetic based purification good for high-throughput applications (MagMax kit)
  - Can use oligo dT beads or total RNA beads
- DNase step:
  - DNase I digestion step is sometimes skipped (polyA libraries only)
  - Best practice is to keep this step

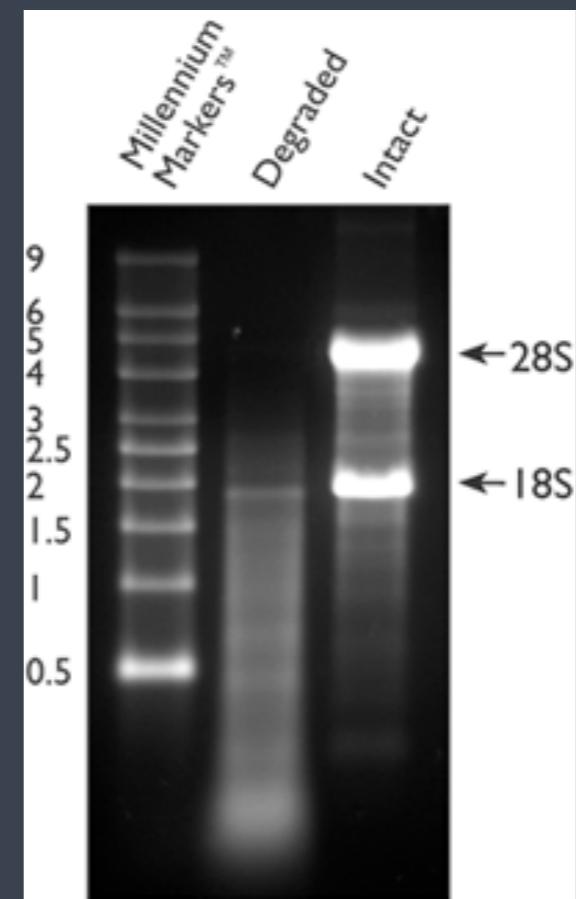
# RNA Extraction: Tissues / Trizol



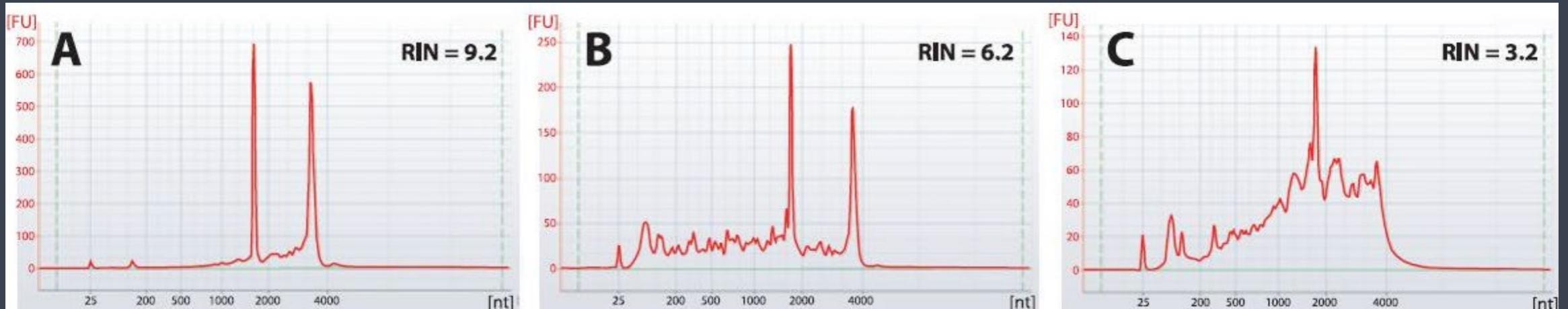
- Keep tissues as cold as possible
  - Work in cold room
- After homogenization suggest column based cleanup
  - Particularly important if used Trizol for lysis
  - DNase 10ug then column cleanup

# RNA Quantitation & Quality

- Quantitation
  - Absorbance: Nano-drop (50-500 ng/ul)
    - Theoretically should read to 3000 ng/ul. Empirically find it is only accurate within range above.
  - Dye based
    - RiboGreen
    - Qubit / Quant-IT  
(specific RNA and miRNA kits)
- Quality
  - Visualize on gel
  - Agilent Bioanalyzer (RIN)

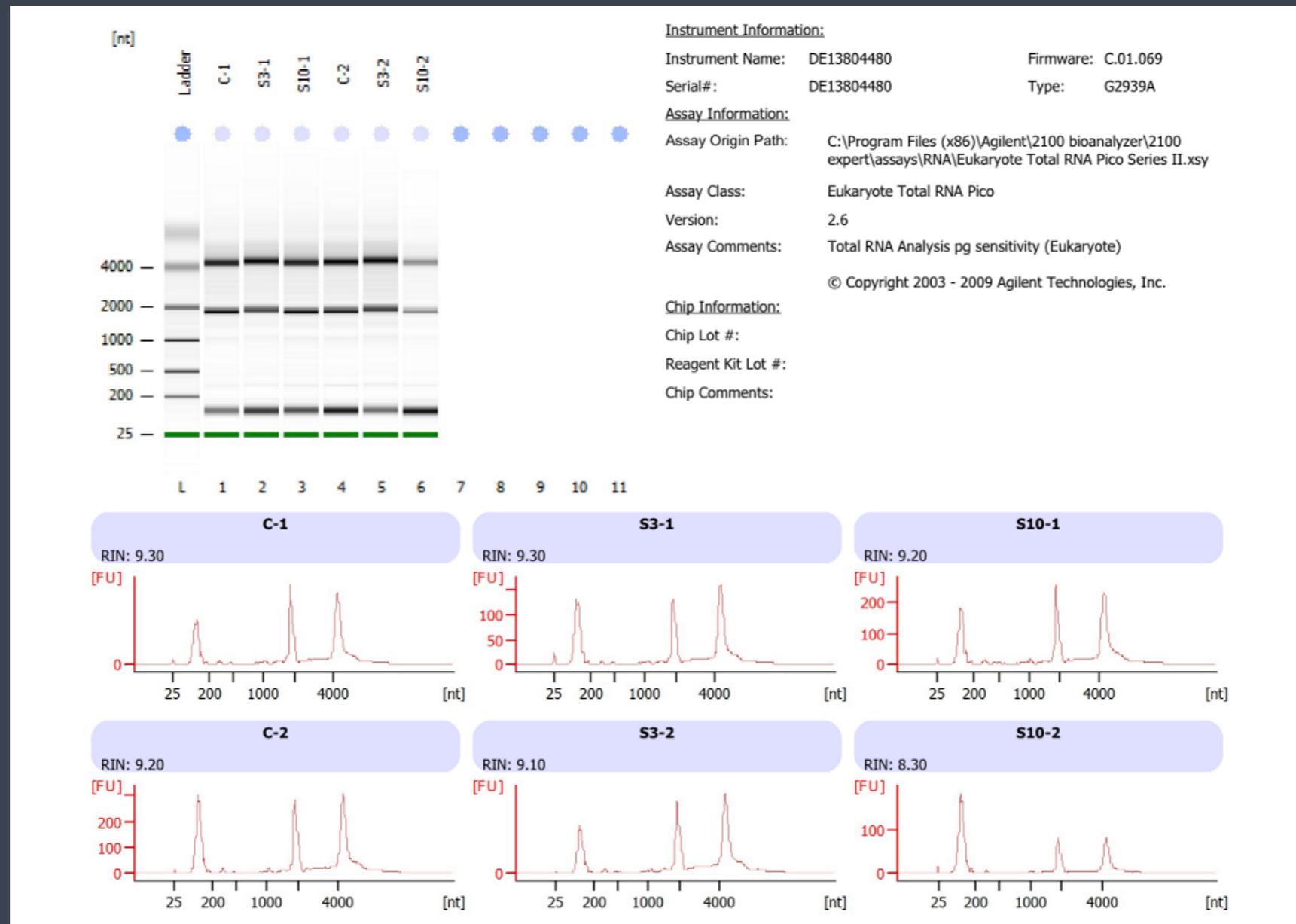


# RNA quality



- High quality RNA needed for mRNA libraries
- Degraded samples should only be used to make a “total” RNA-seq library – rRNA removal
- FFPE & Archival Samples

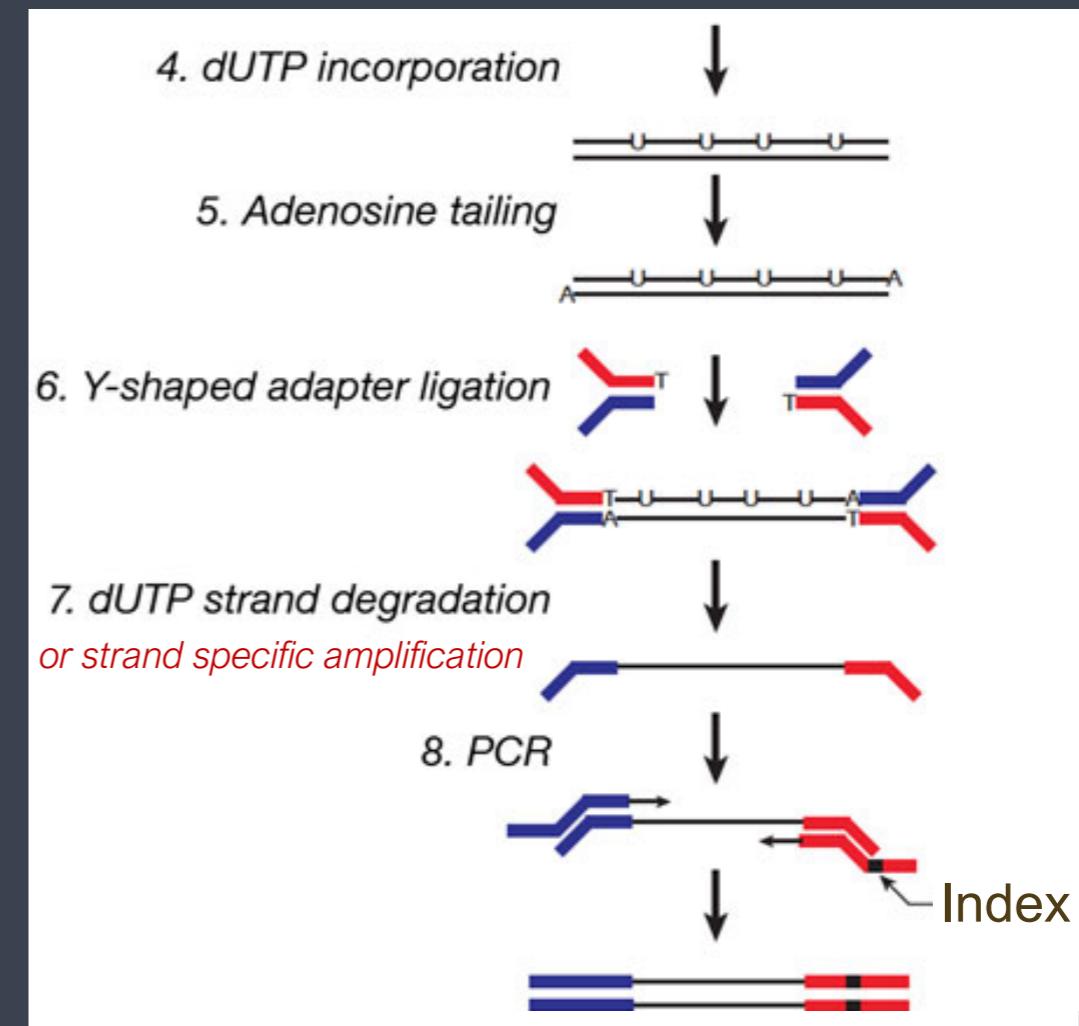
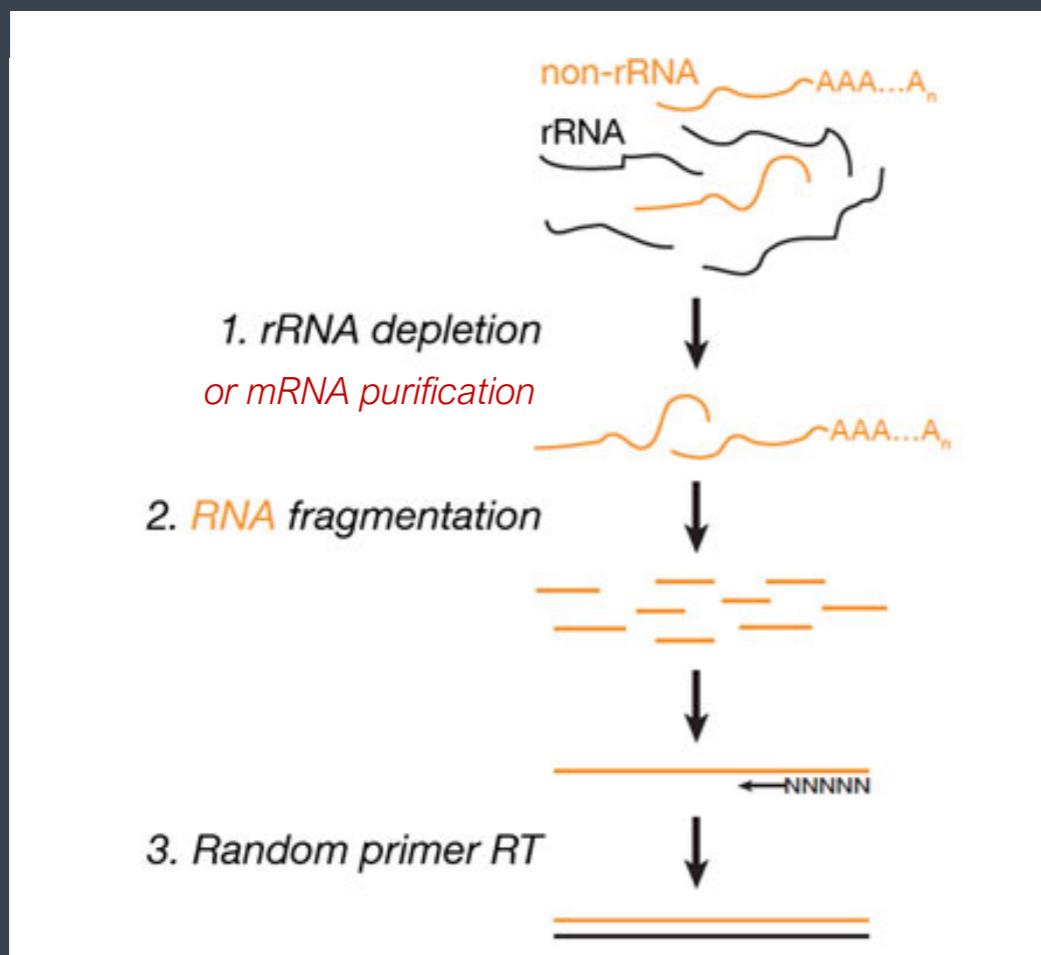
# Did you get your micro RNA?



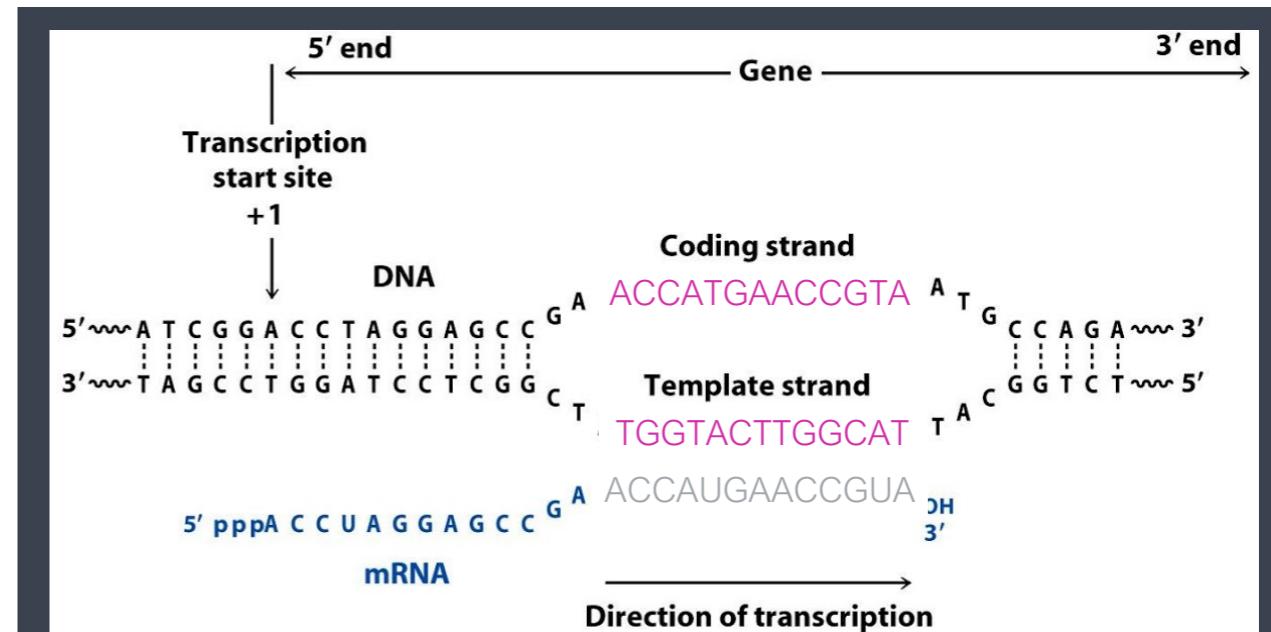
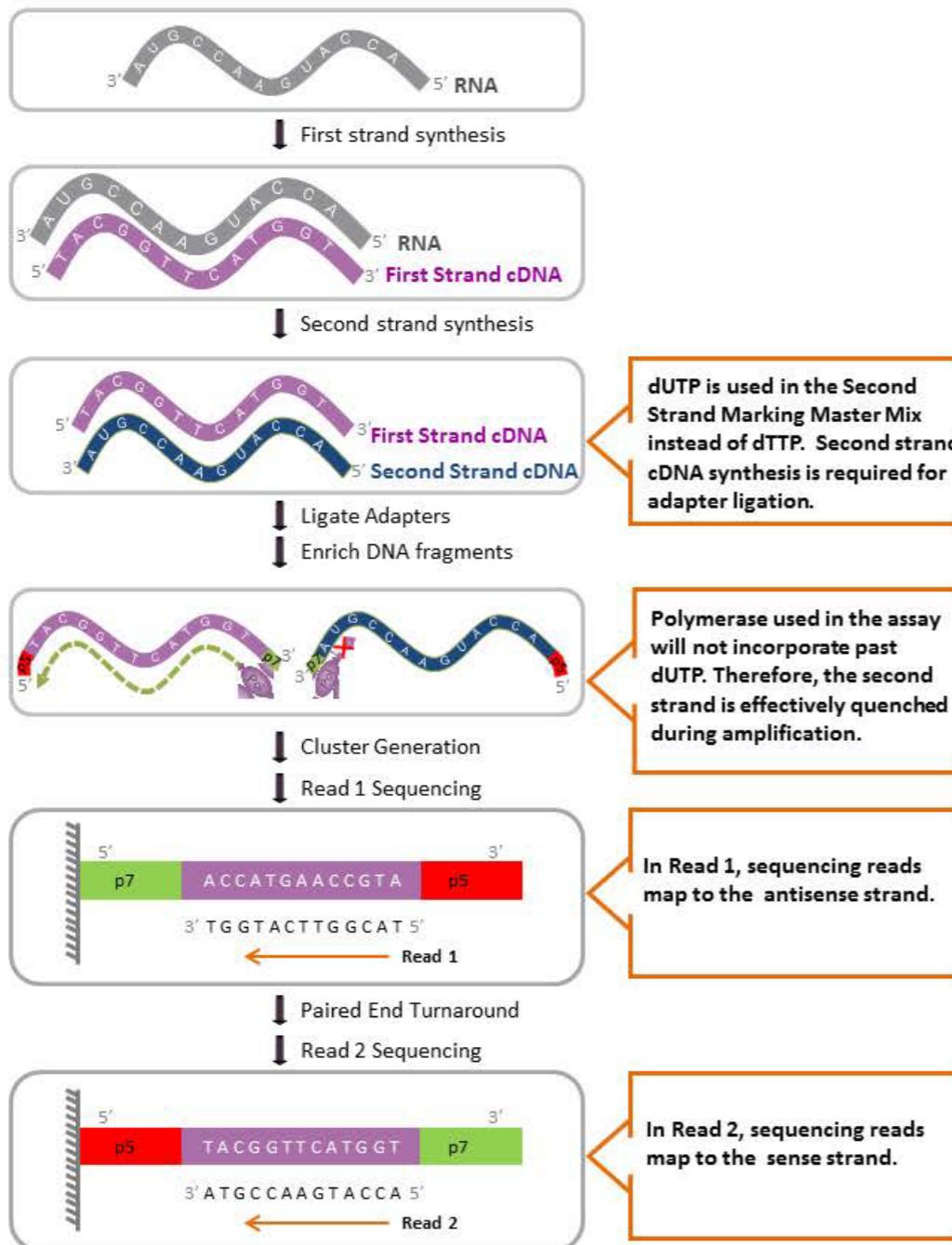
# Sequencing Library Preparation

- Starting Material: purifying RNA
- RNA-seq library preparation
- Multiplexing & Sequencing
- Single cell / low input methods
- Capture sequencing

# RNA-seq Stranded Library Prep (dUTP method)

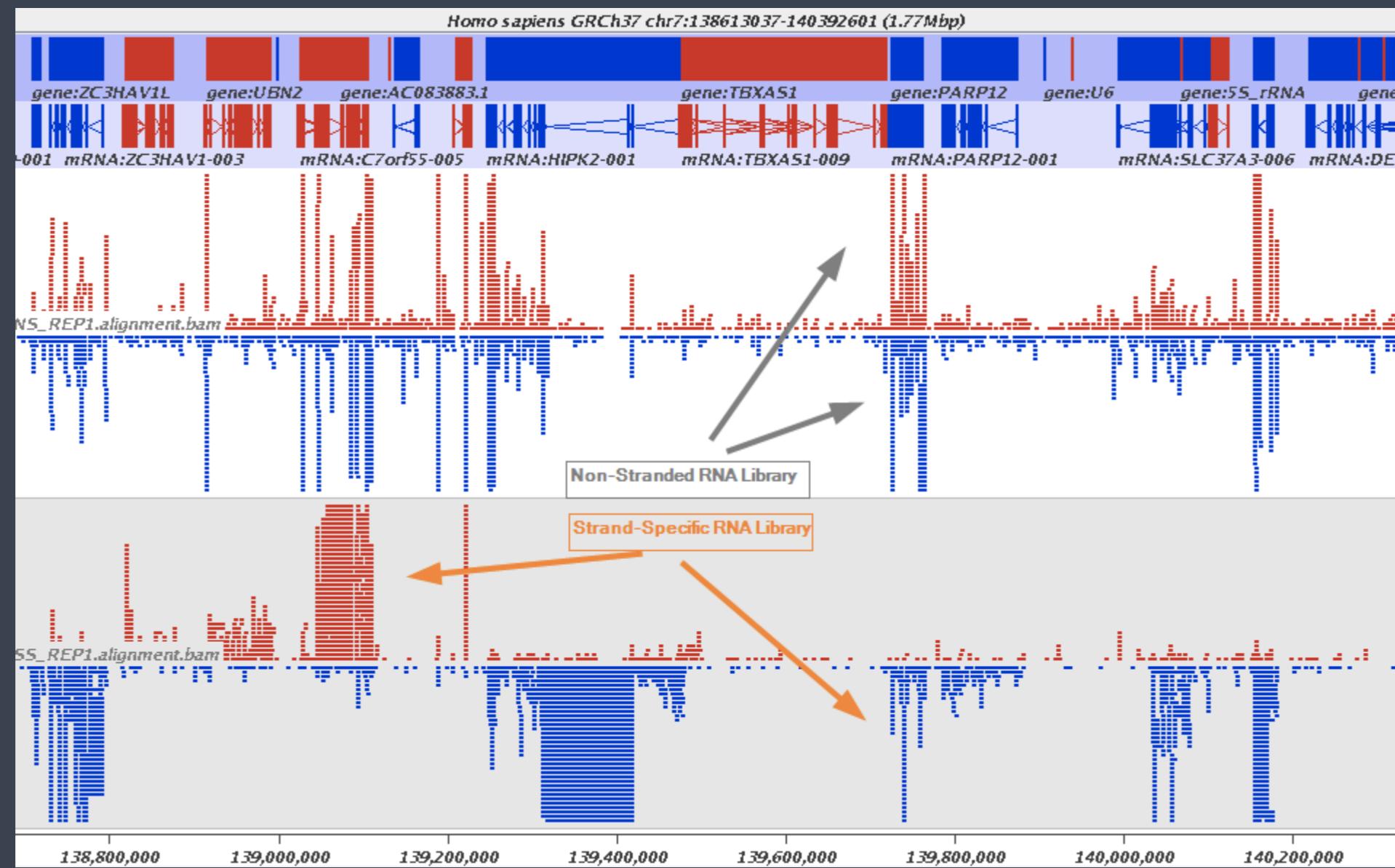


# Library Strandedness



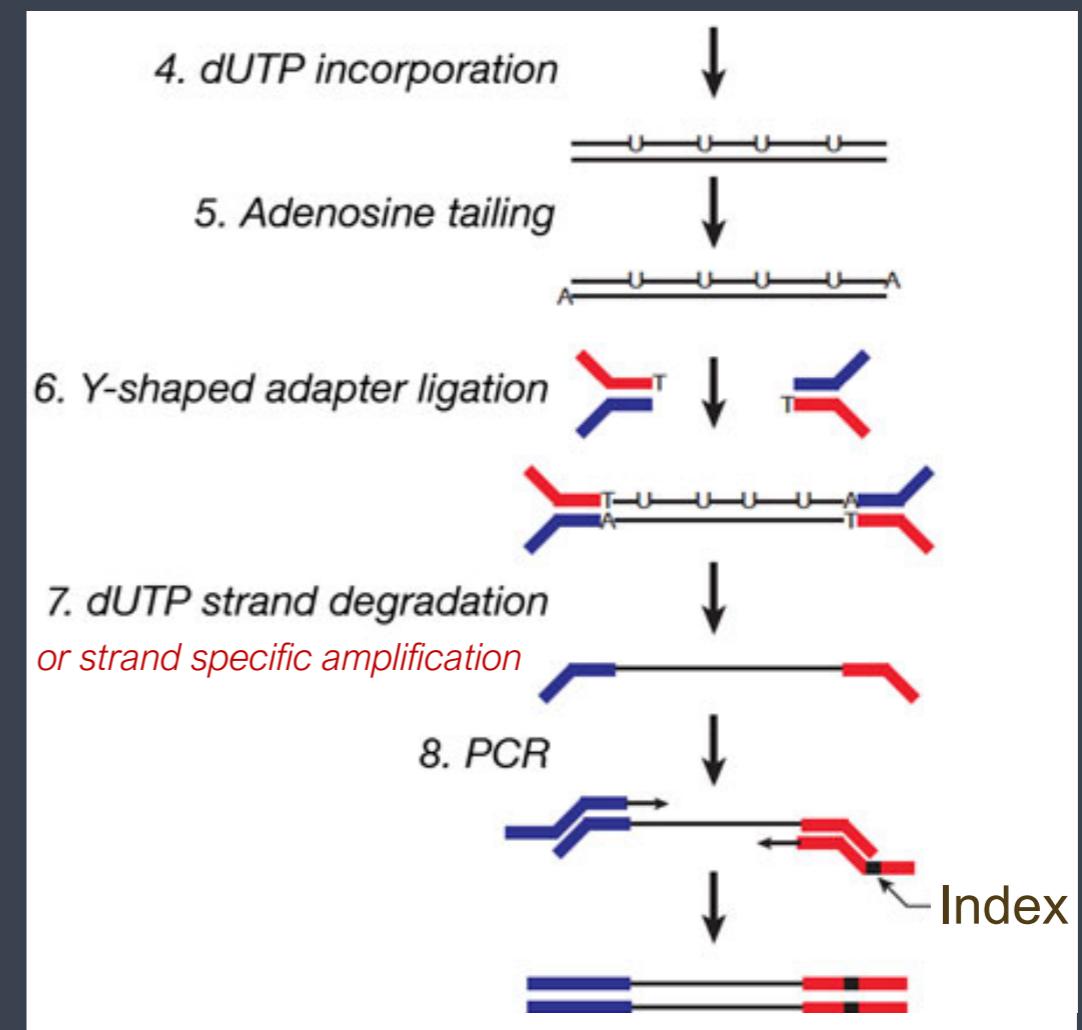
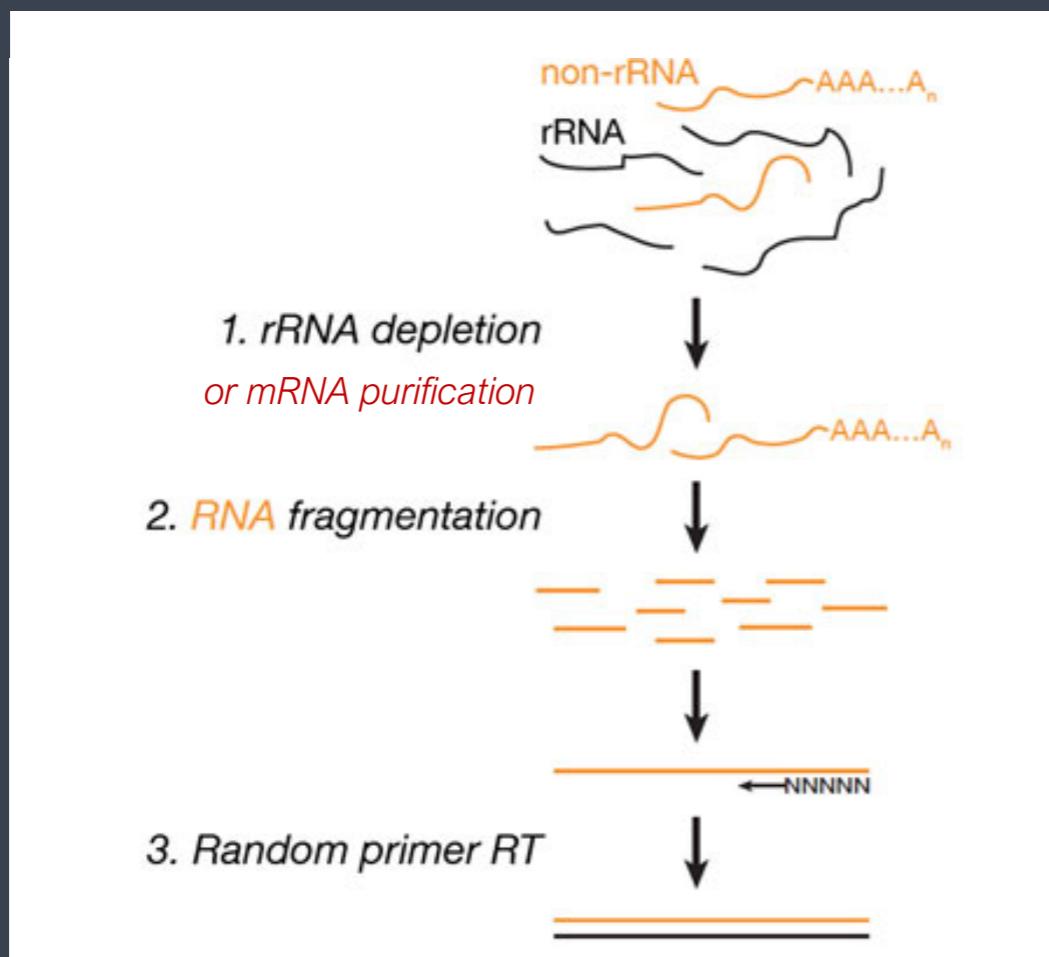
- Read alignment depends on direction of transcription
- “sense” strand of transcript can be on either the sense or antisense strand of the DNA
- Read 2 maps the sense RNA

# Library Strandedness



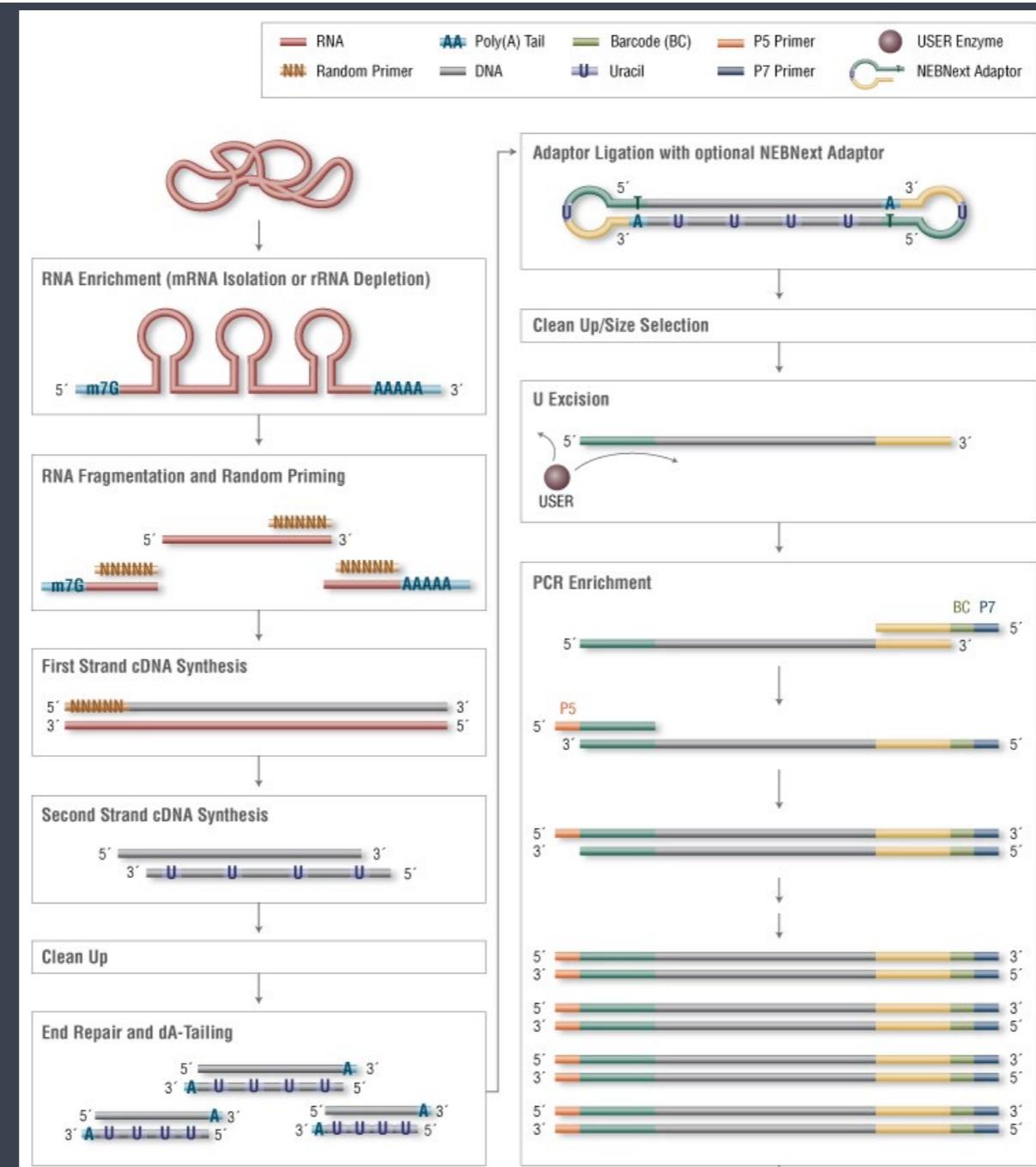
# RNA-seq Stranded Library Prep

## Illumina TruSeq / dUTP method



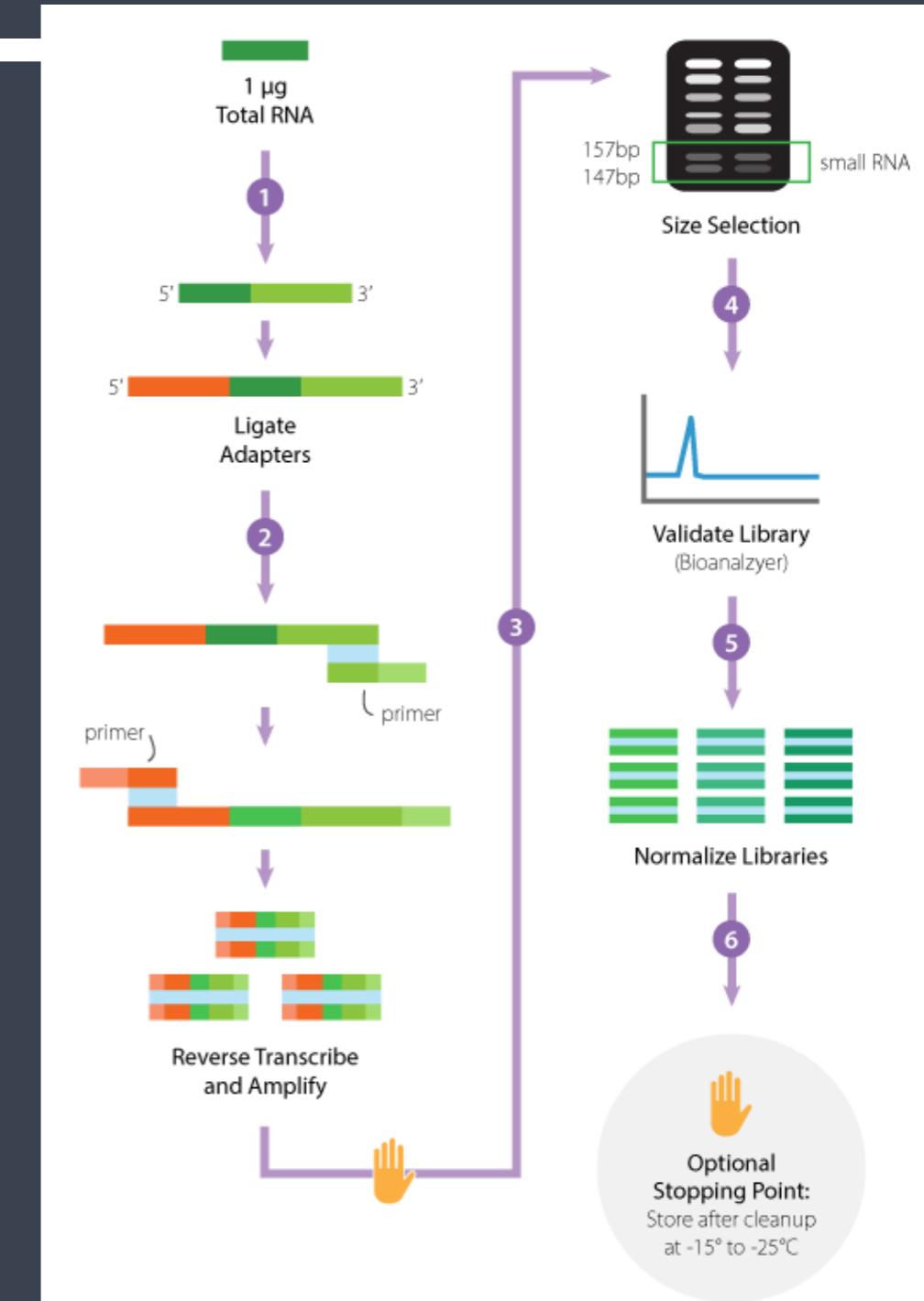
# RNA-seq Stranded Library Prep

## NEBNext Ultra II / dUTP method



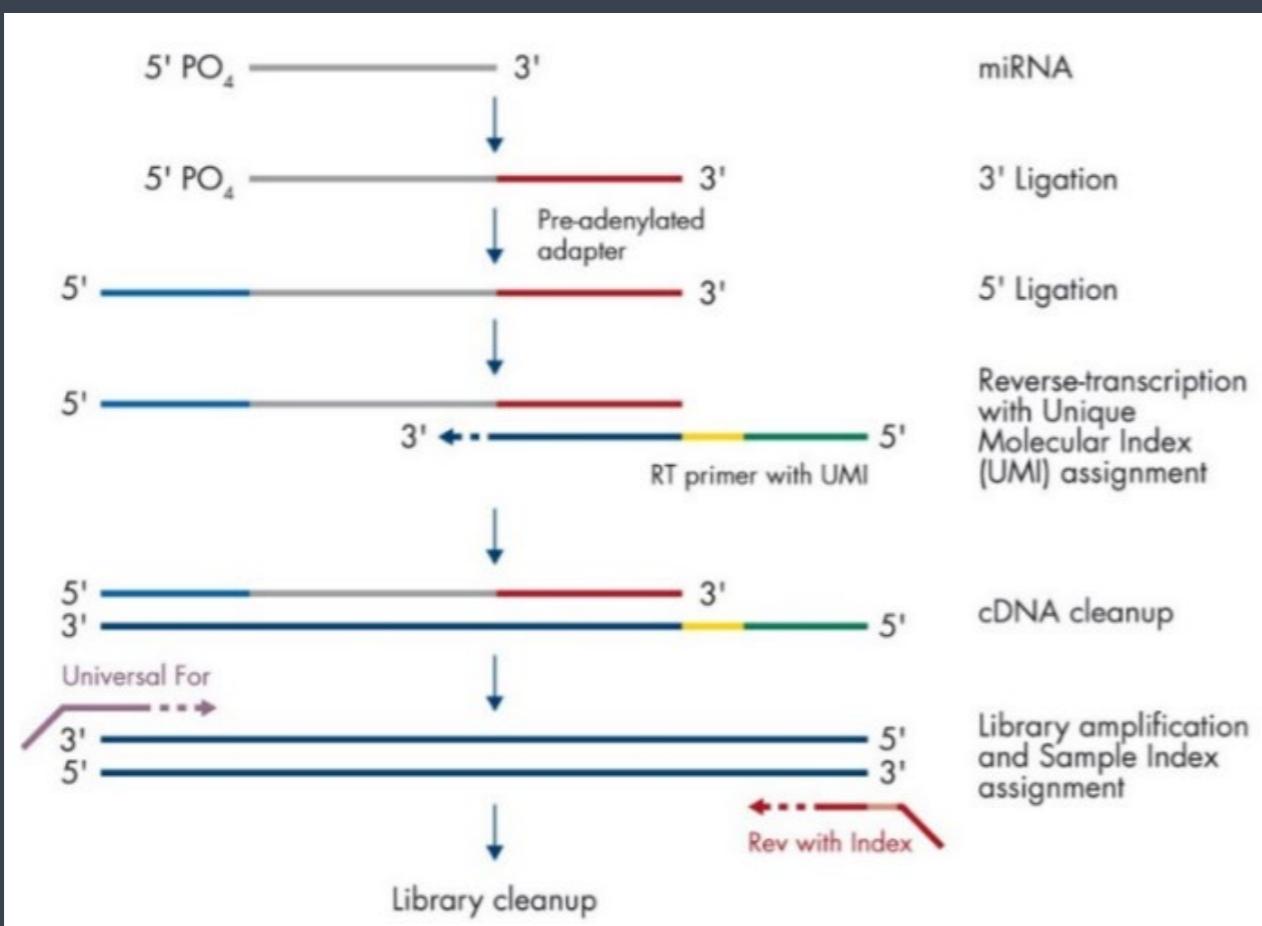
# Micro RNA Library Prep

- Ligate 3' and 5' adaptors
  - 5' Phosphate
  - 3' OH
- RT PCR – 1<sup>st</sup> strand synthesis
- Amplify
- Gel purify
- Validate library



# Gel Free Micro RNA Library Prep

## QIAseq miRNA

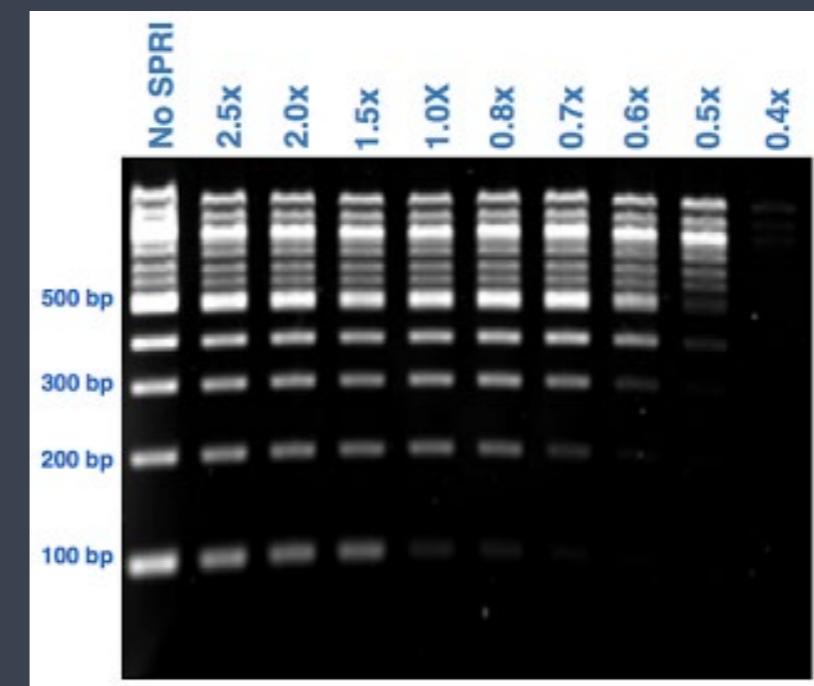
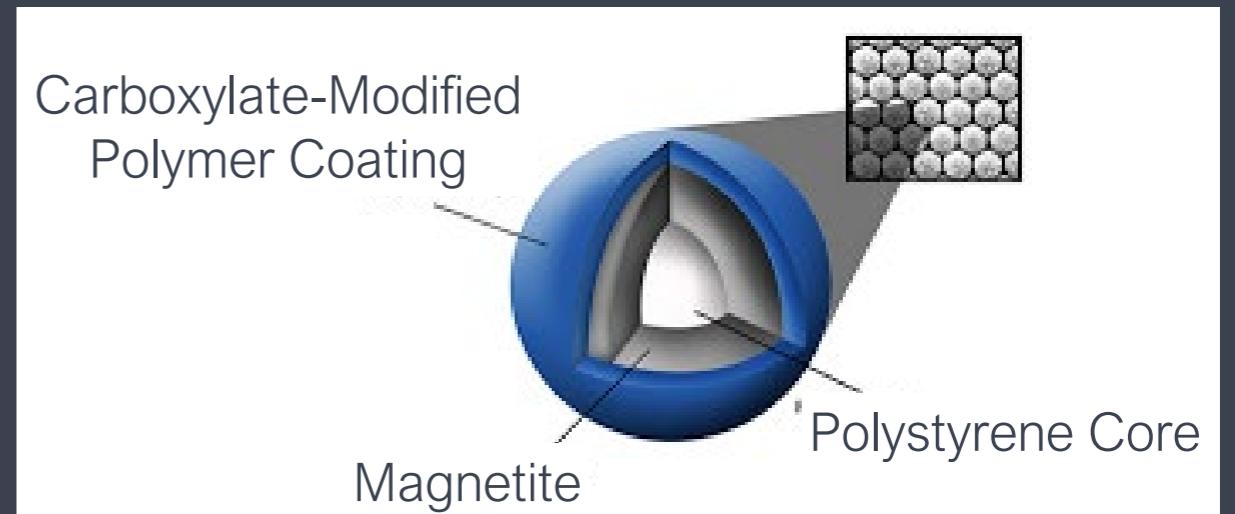


Several vendors offer Gel-Free kits

- QIAseq miRNA (QIAGEN)
- NEXTflex (PerkinElmer)
- CleanTag (Trilink)

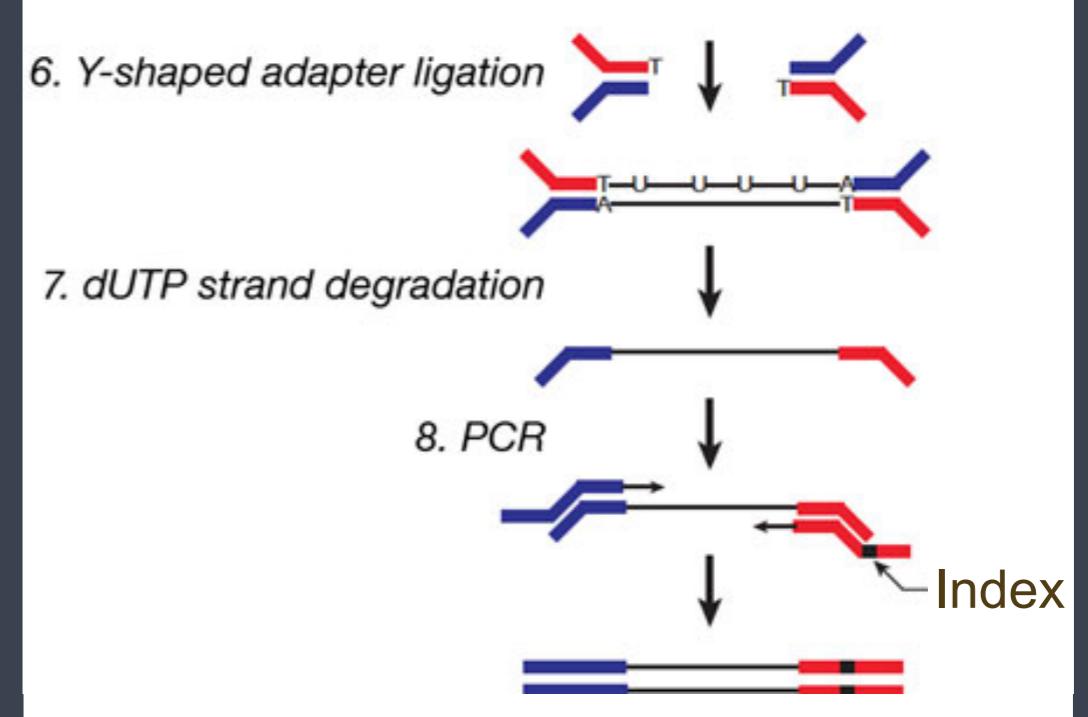
# Size selection with SPRI beads

- Solid Phase Reverse Immobilization beads
- Carboxyl groups on surface bind DNA in the presence of crowding agents (PEG & NaCl)

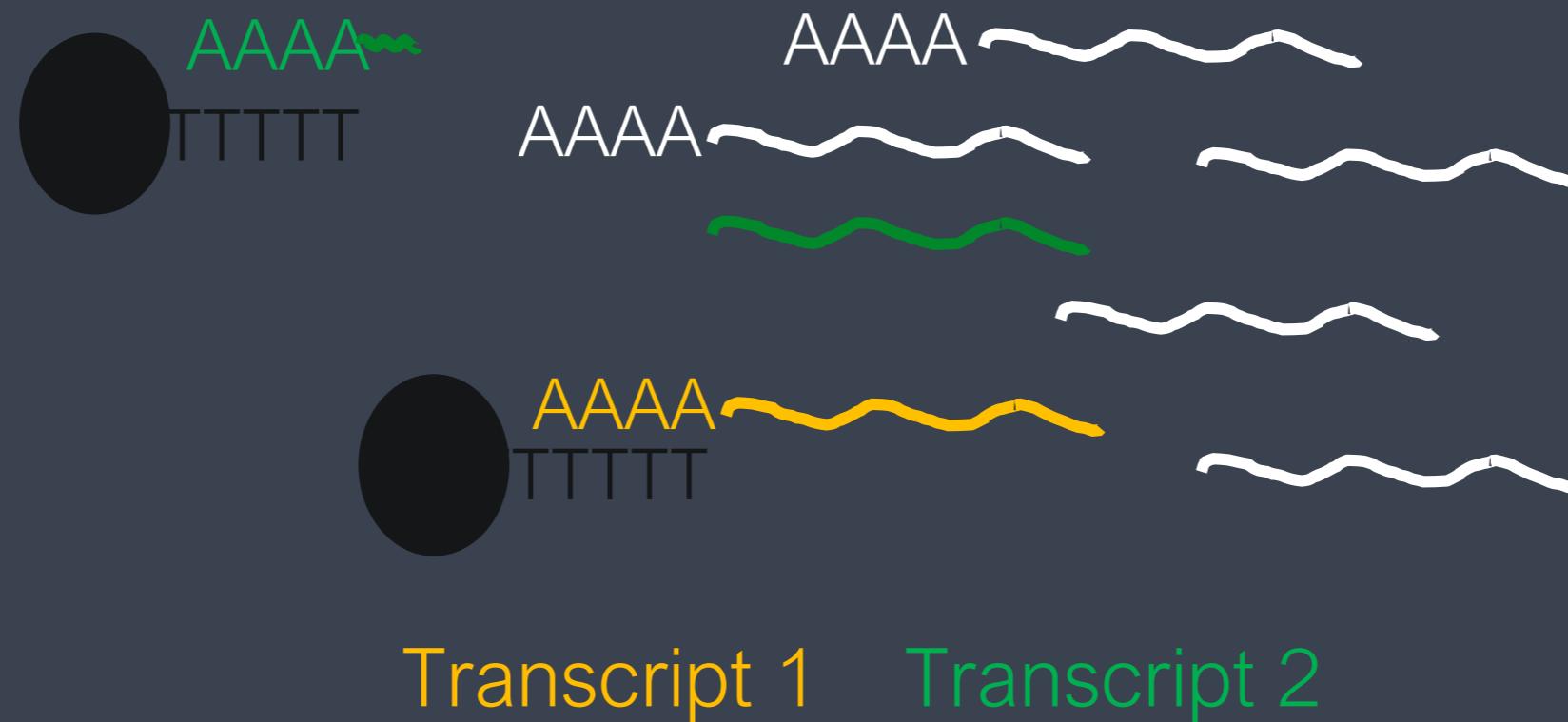


# Key steps in library preparation

- Quality (RIN) of starting RNA
- Library amplification bias
- Library QC & cleanup
- qPCR quantitation



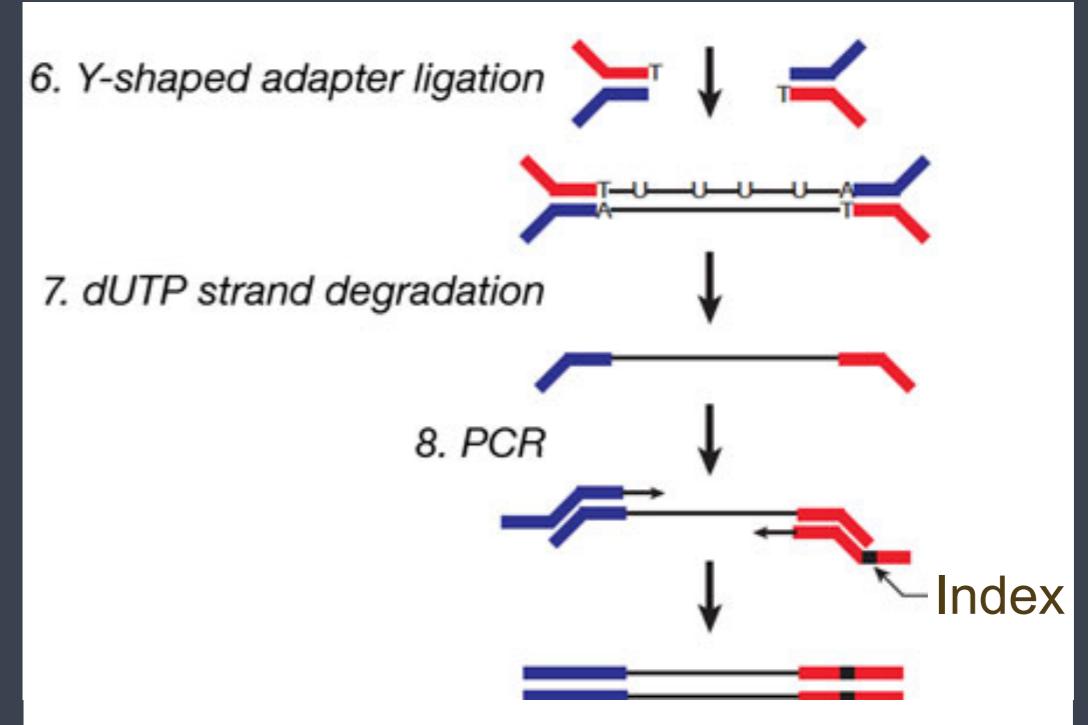
# mRNA Purification of Degraded Samples



- PolyA tail no longer attached to transcript.
- Results in differential loss of transcripts between samples.
- Need to perform rRNA removal not mRNA purification.

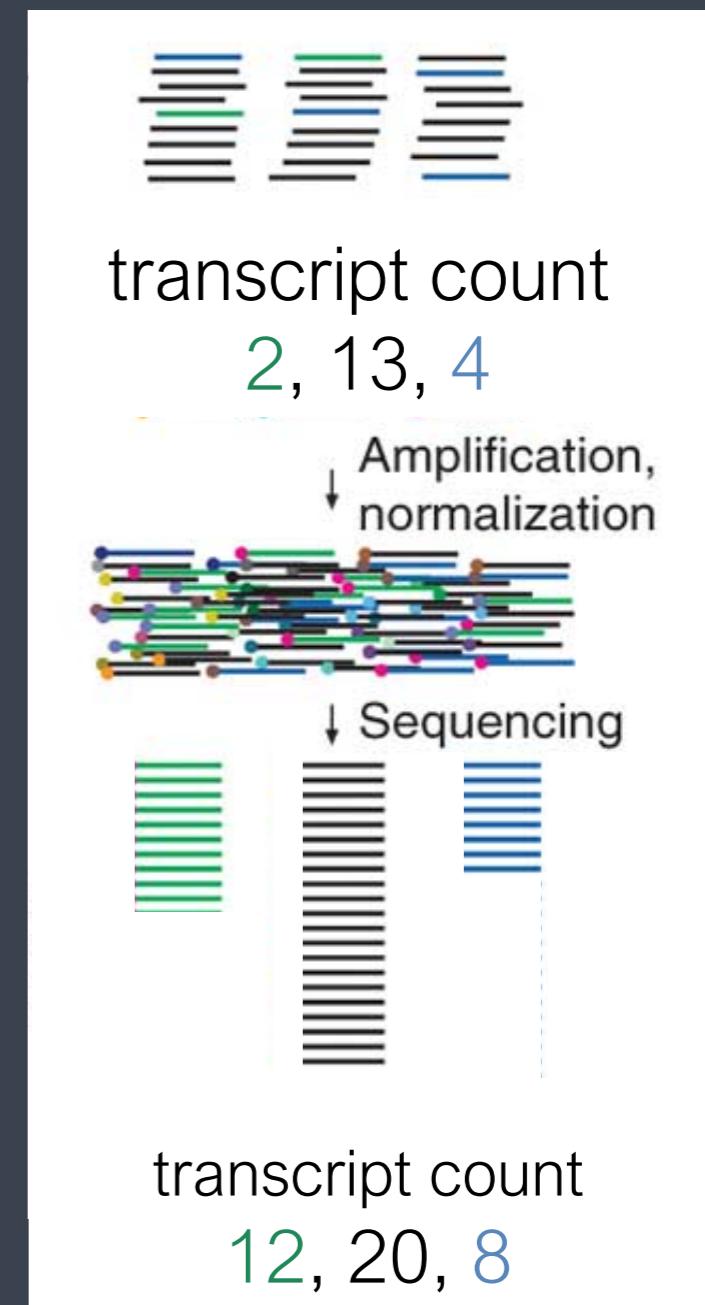
# Key steps in library preparation

- Quality (RIN) of starting RNA
- Library amplification bias
- Library QC & cleanup
- qPCR quantitation



# Library Amplification Bias

- Final step of library prep is amplification
- Introduces library bias
  - Some products preferentially amplified (~40-60% GC)
- Fewer cycles = less bias

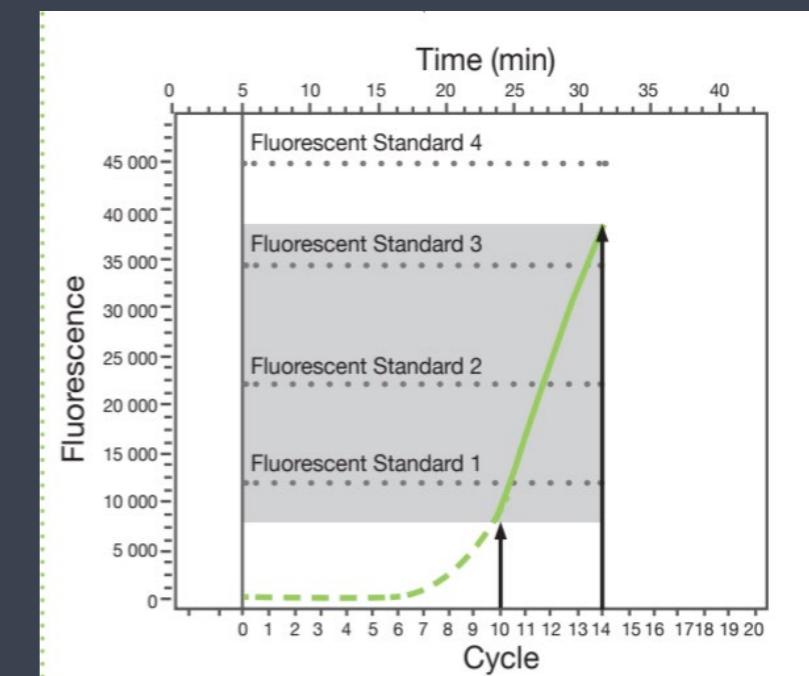
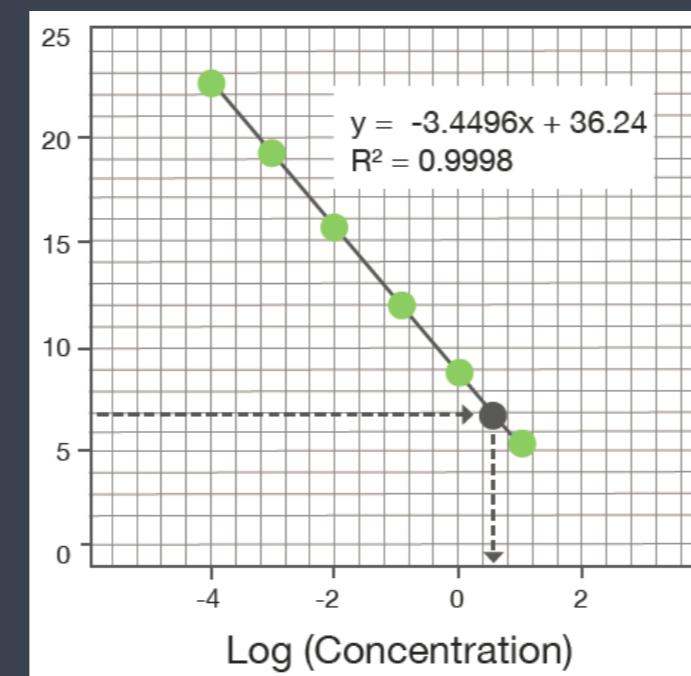
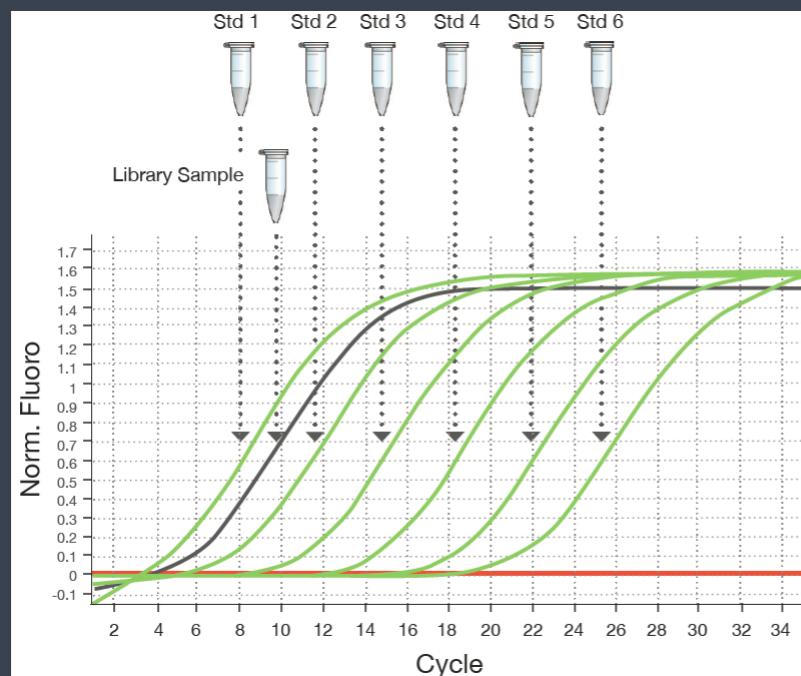


# Limited Cycle Library Amplification

- Number of cycles needed is proportional to amount of input RNA.
- Library prep kits will recommend a certain number of cycles.
  - This is usually optimized for the mid/lower input.
- Test how many cycles will give you enough product.
- Fewer cycles = less bias

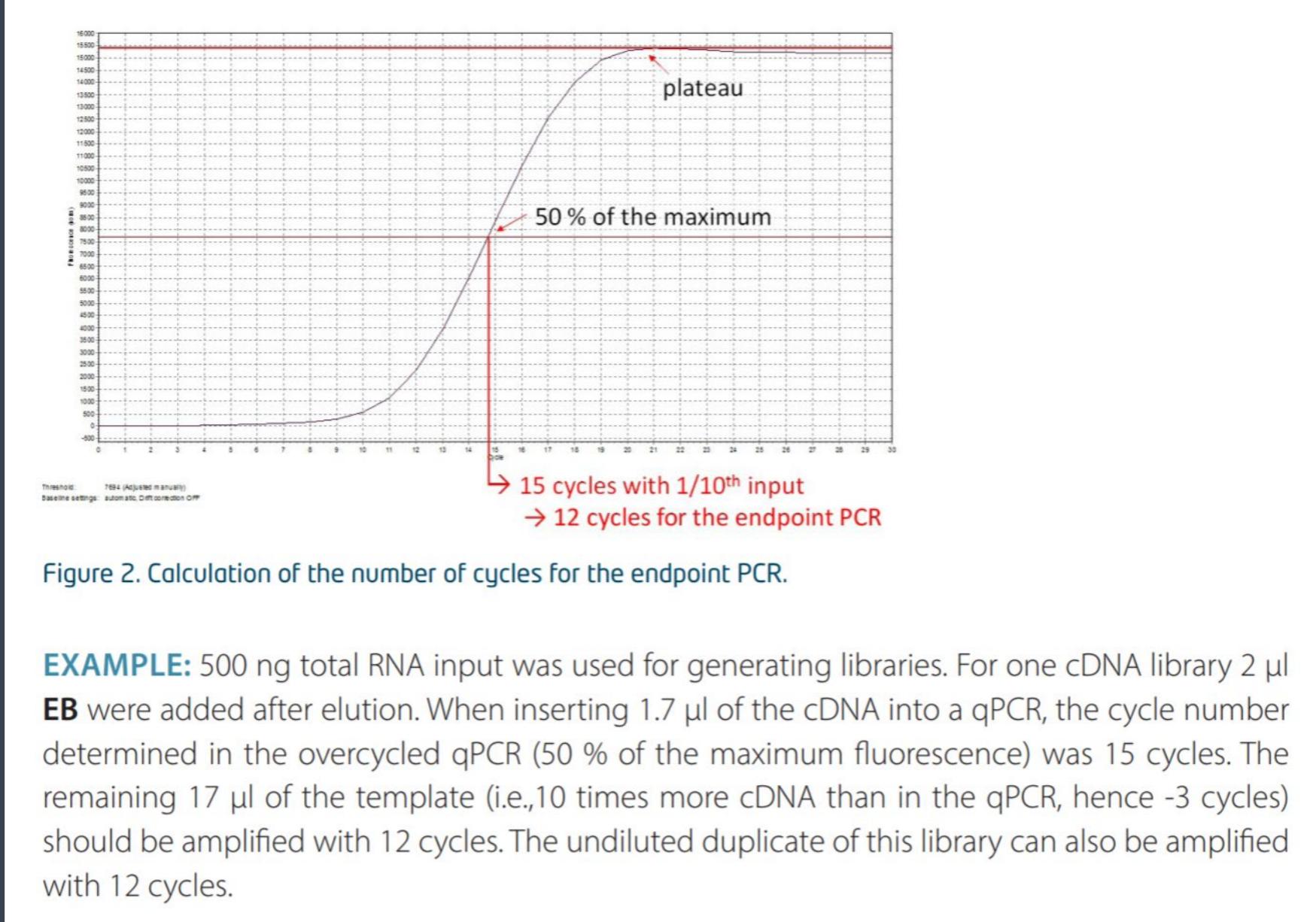
# Limited Cycle Library Amplification

- Perform micro qPCR reaction on small amount of pre-amplification library (Kapa)

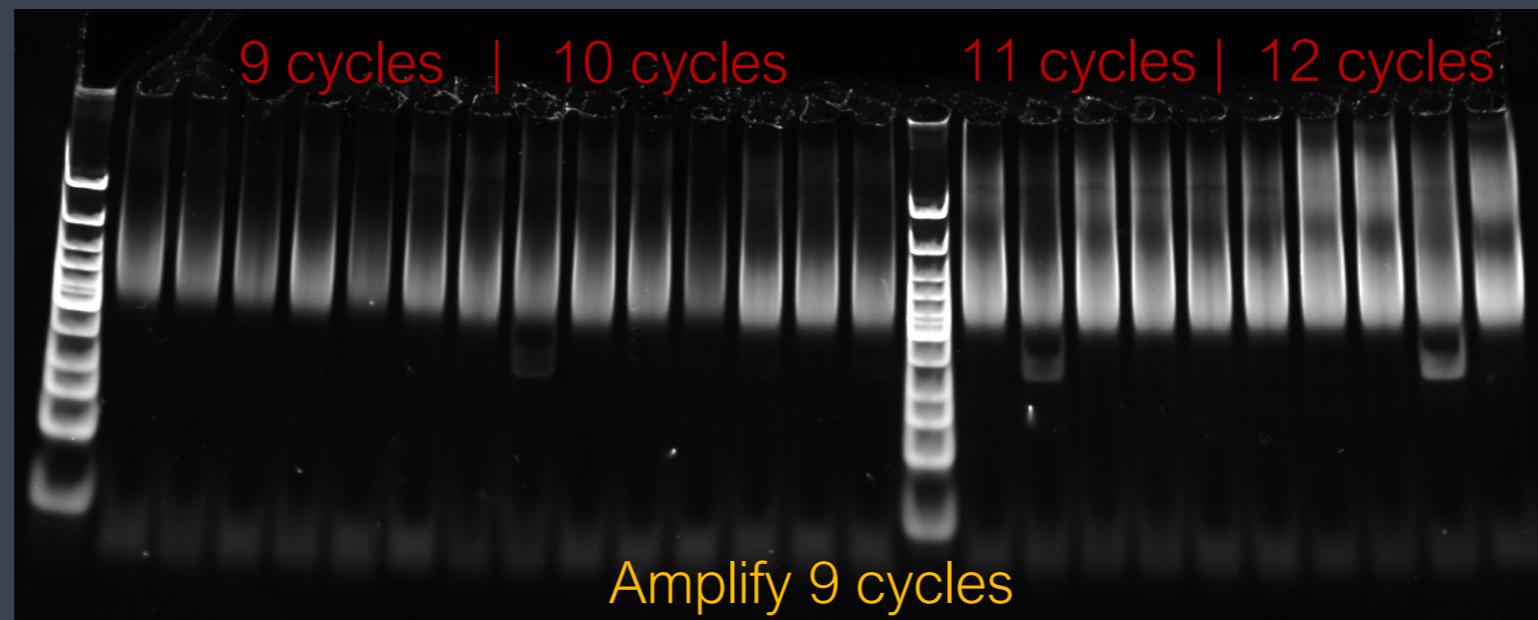


- Amplify only the number cycles needed to get enough product for sequencing (20ul of 4nM product)
- Stop reaction when fluorescence between standard 1-3

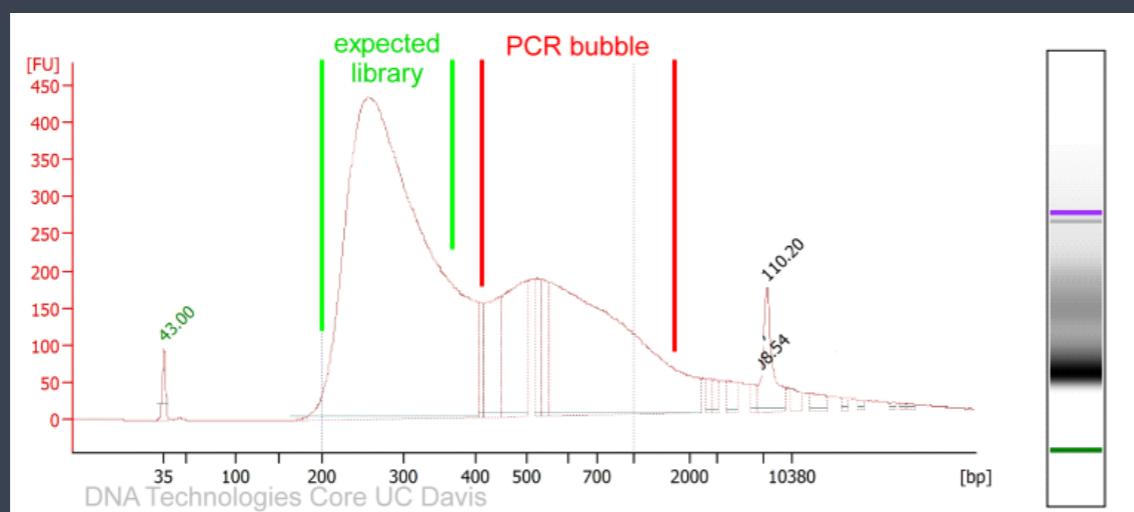
# Limited Cycle Library Amplification



# Limited Cycle Library Amplification

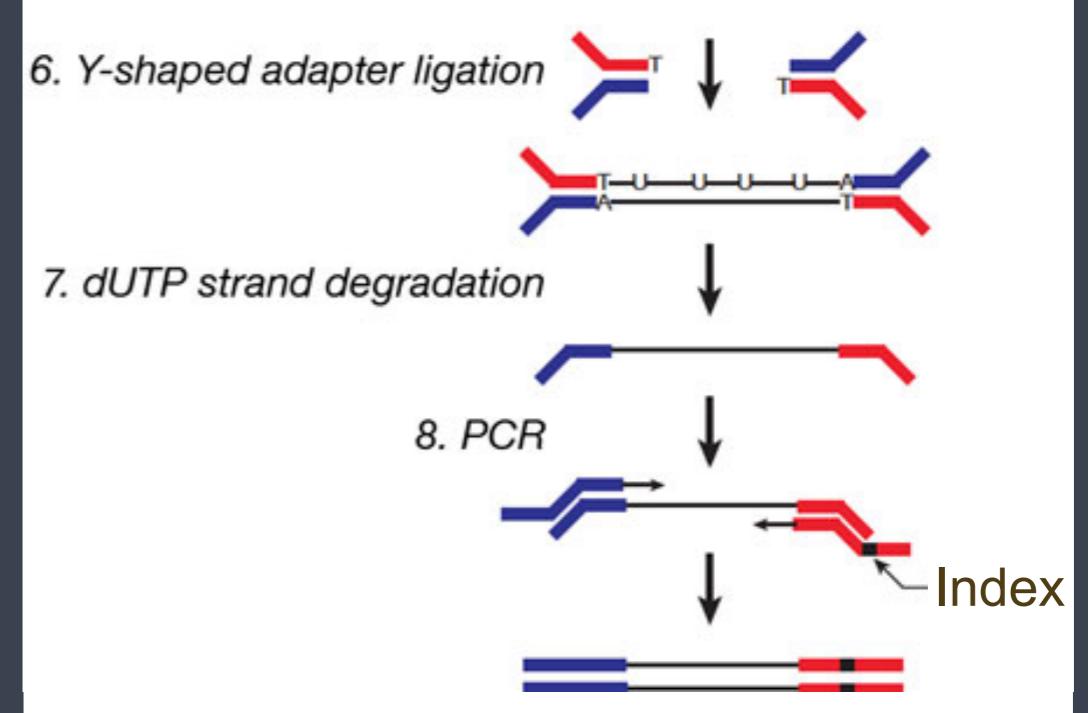


PCR  
bubble



# Key steps in library preparation

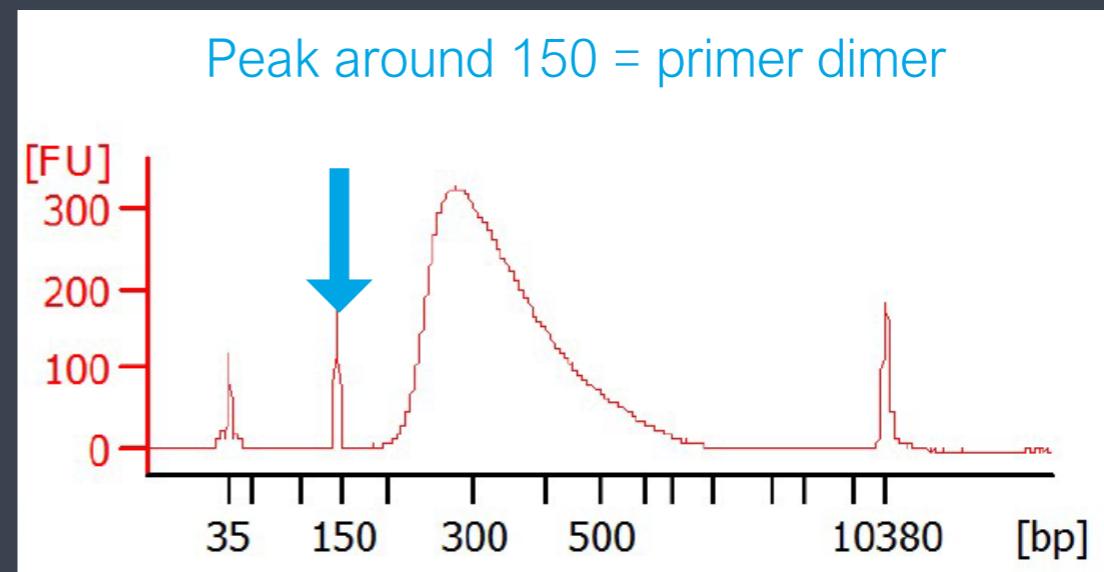
- Quality (RIN) of starting RNA
- Library amplification bias
- Library QC & cleanup
- qPCR quantitation



# Library QC

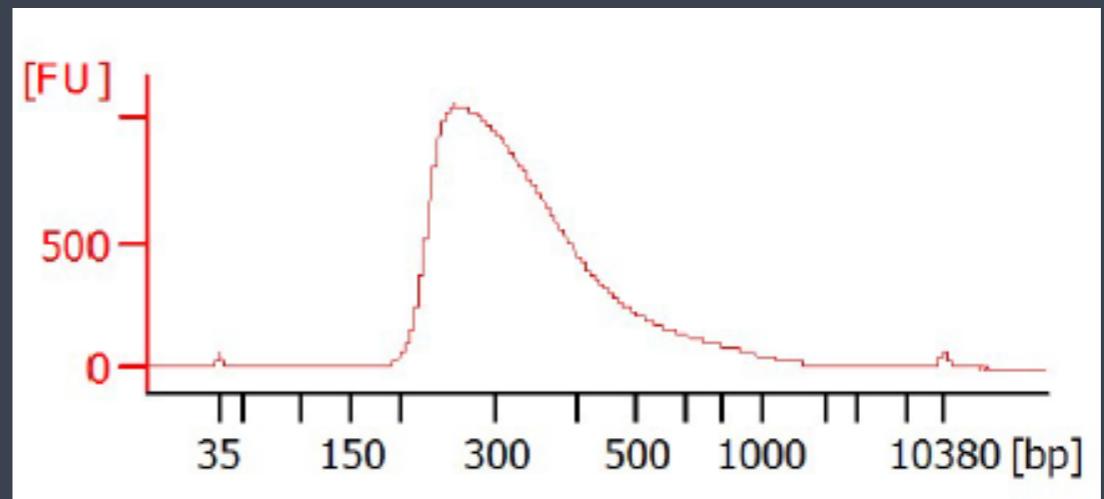
## ➤ Quantitation

- Dye based
  - SYBR Green
  - Qubit / Quant-IT



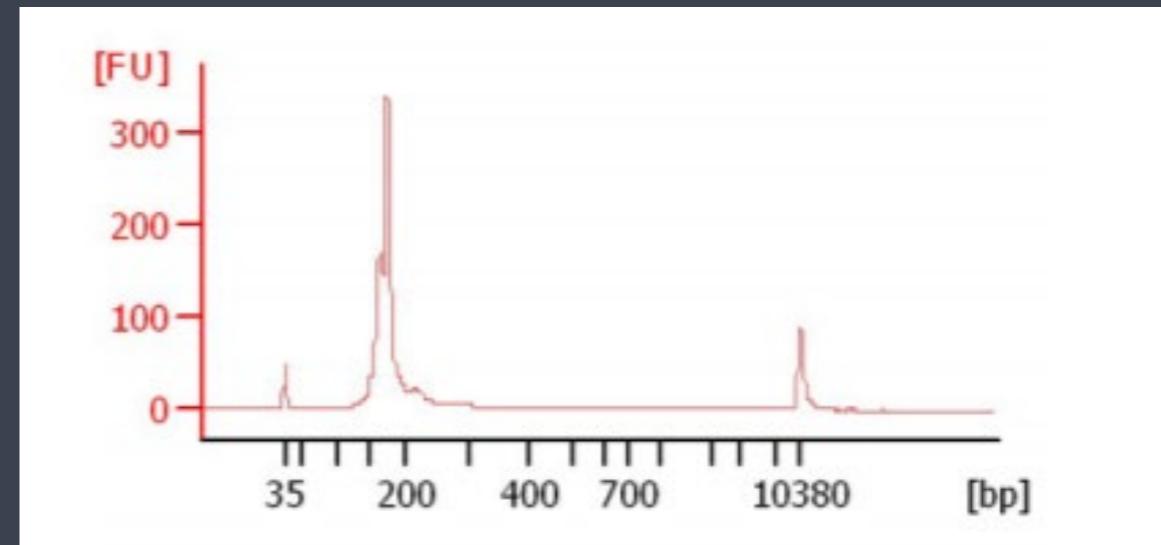
## ➤ Size & Quality

- Agilent Bioanalyzer
- Size determination
- Do not use for quantitation  
(except for micro RNA library)



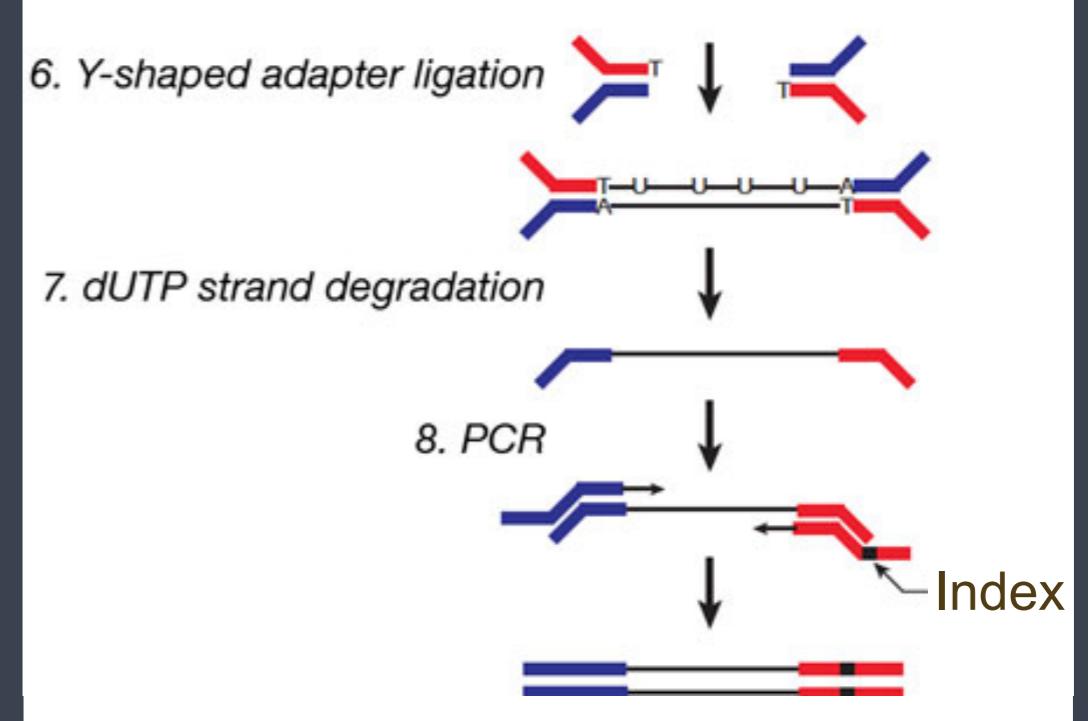
# Library QC miRNA

- miRNA library should have one clear peak.
- Quantitate and pool libraries based on concentration from Bioanalyzer.



# Key steps in library preparation

- Quality (RIN) of starting RNA
- Library amplification bias
- Library QC & cleanup
- qPCR quantitation



# Sequencing Library Preparation

- Starting Material: purifying RNA
- RNA-seq library preparation
- Multiplexing & Sequencing
- Single cell / low input methods
- Capture sequencing

# Multiplexing

- Multiplexing allows optimal use of reads you will get
- Charges for sequencing are usually per lane of the flow cell
  - HiSeq generates ~150 million reads per lane
  - NextSeq generates ~ 450 million reads (one lane instrument)
- For RNA-seq number of reads you need will depend on your experiment
  - 10 million standard for transcriptome
  - 20 million standard for total RNA (rRNA depleted)

**Make sure multiplexing libraries of similar size**

# Sequencing Read Order

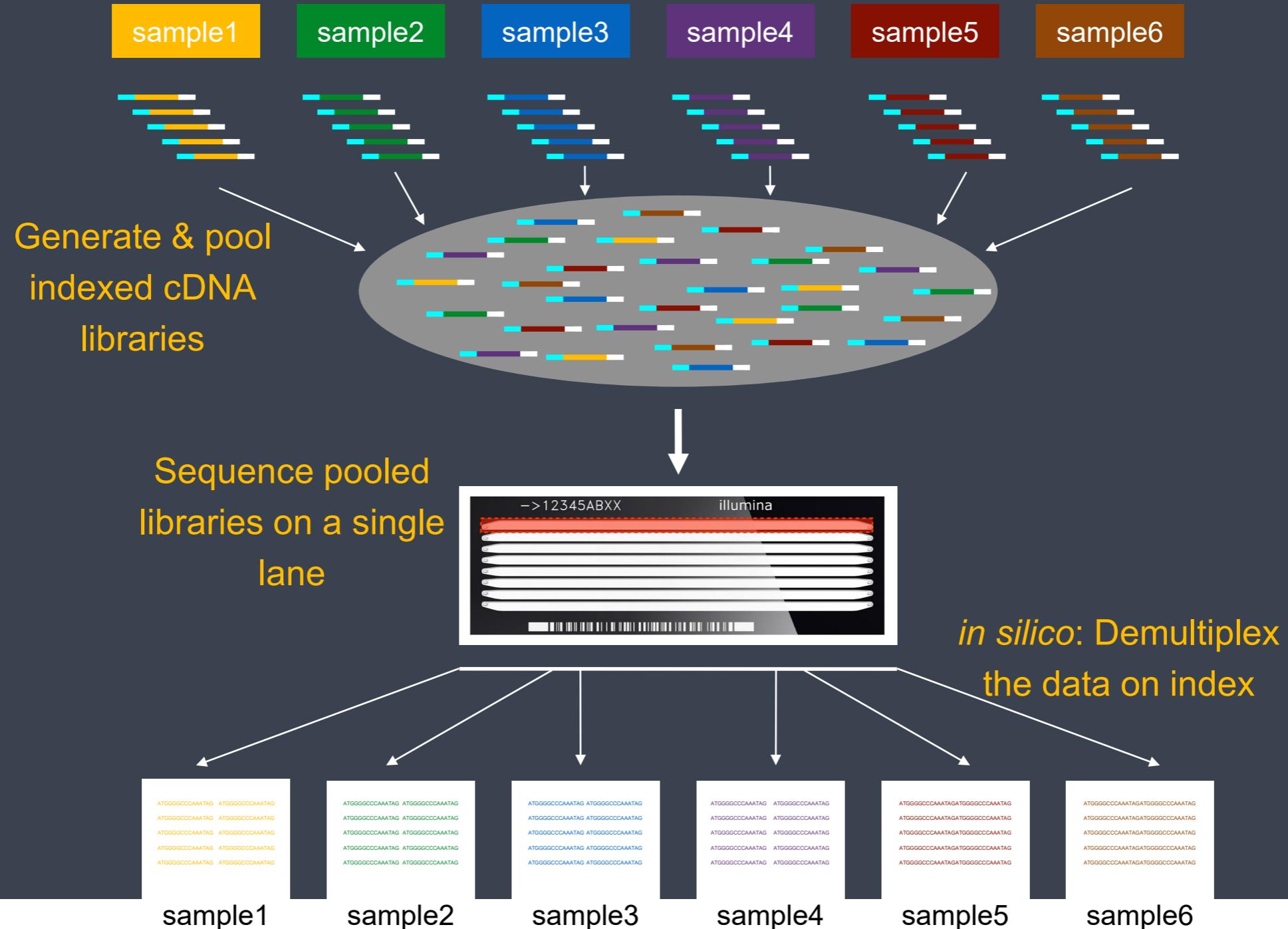
## Index vs Barcode



1. Read 1
2. Index Read 1 (i7)
3. Index Read 2 (i5)
4. Read 2

- HiSeq/MiSeq (4 color)
- A&C read on one camera
  - G&T read on other
- NextSeq (2 color)

# Multiplexing



# Multiplexing

Single index



Unique dual index



xGen UDI-UMI adapter



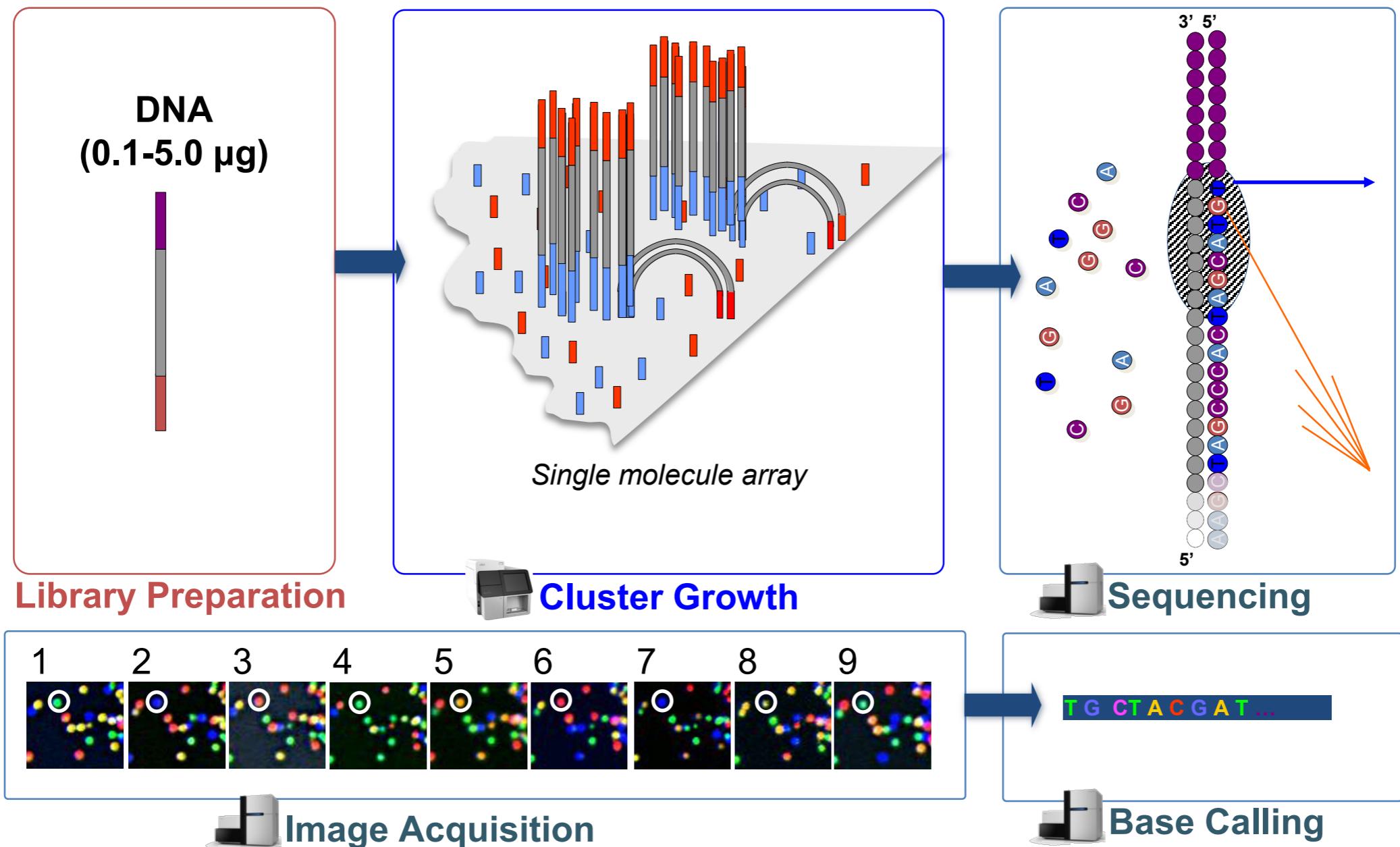
**Flow cell binding sequence:** Platform-specific sequences for library binding to instrument

**Sequencing primer sites:** Binding sites for general sequencing primers

**Sample indexes:** Short sequences specific to a given sample library

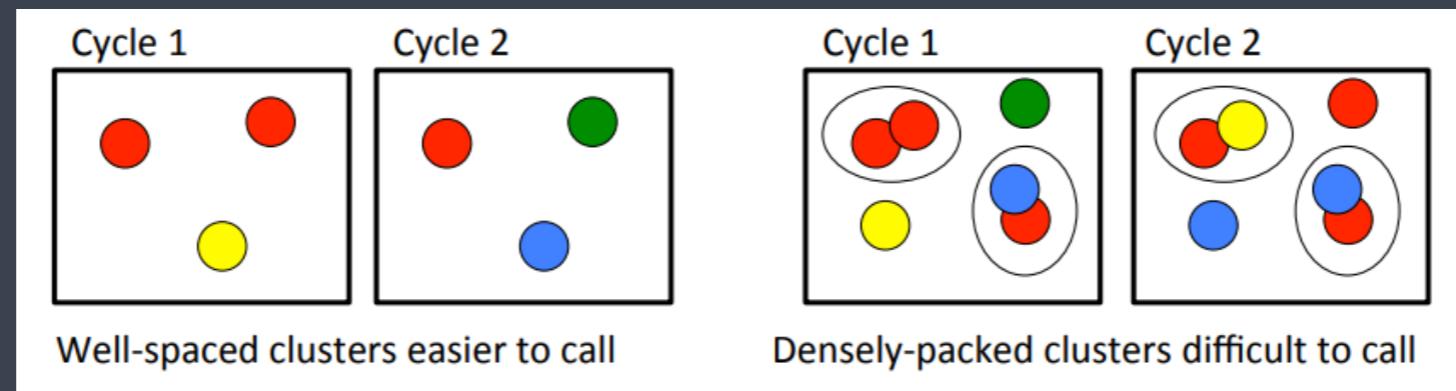
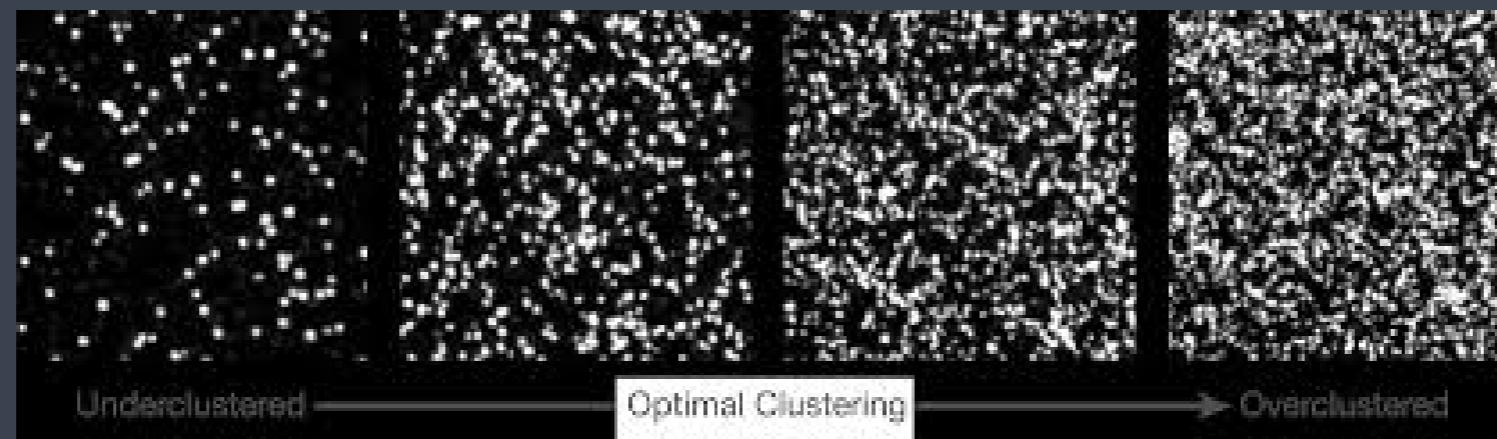
**Molecular index/barcode:** Short sequence used to uniquely tag each molecule in a given sample library

**Insert:** Target DNA or RNA fragment from a given sample library



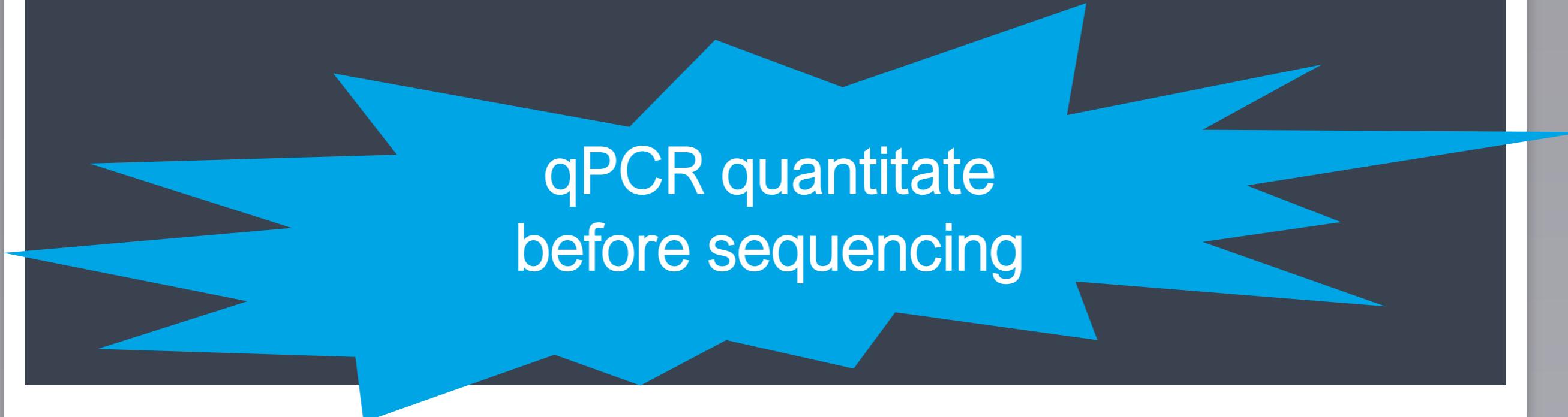
# Consider Cluster Size in Multiplexing

# Cluster / Fragment Size



# Multiplex then qPCR

- Pool samples based on dye based quantitation (qubit or QuantIT)
- Submit pool to core facility for sequencing.
- Make all sequencing libraries in one batch

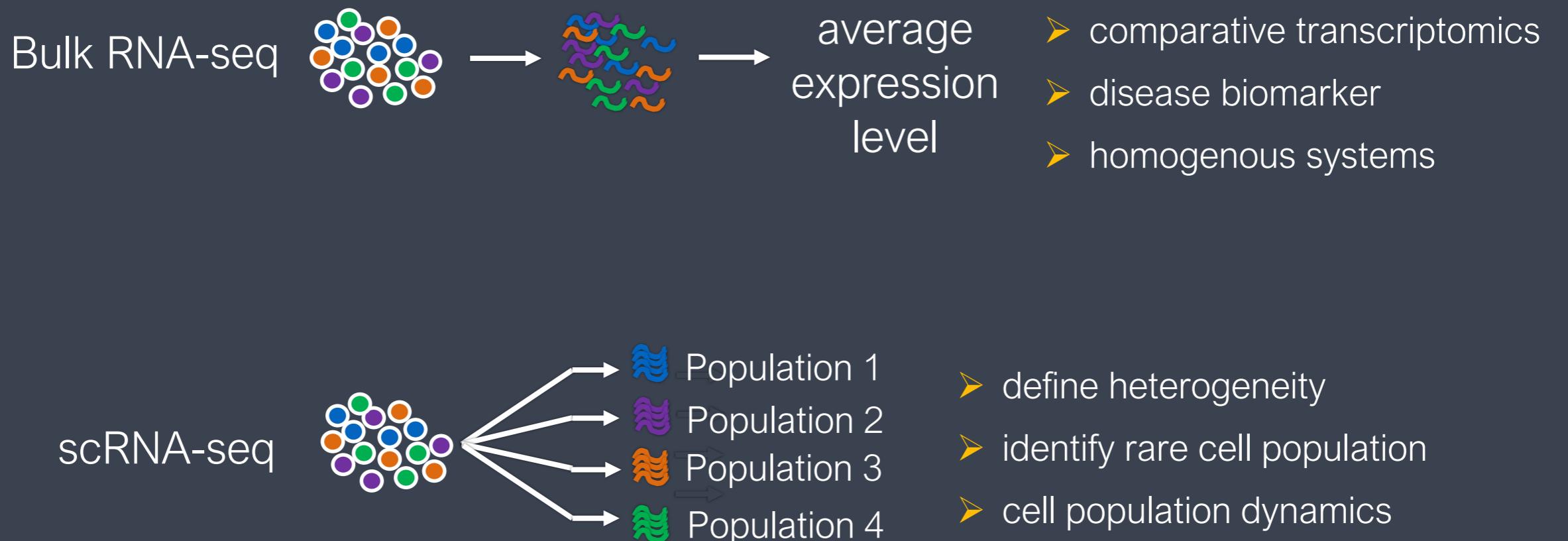


qPCR quantitate  
before sequencing

# Sequencing Library Preparation

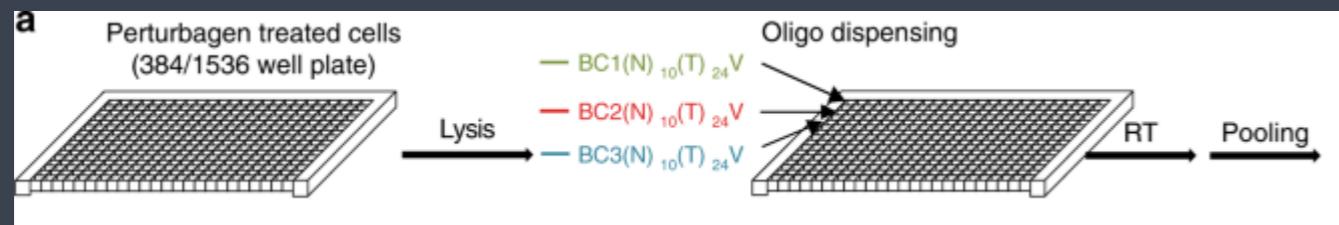
- Starting Material: purifying RNA
- RNA-seq library preparation
- Multiplexing & Sequencing
- Single cell / low input methods
- Capture sequencing

# Bulk vs Single Cell RNA-seq (scRNA-seq)

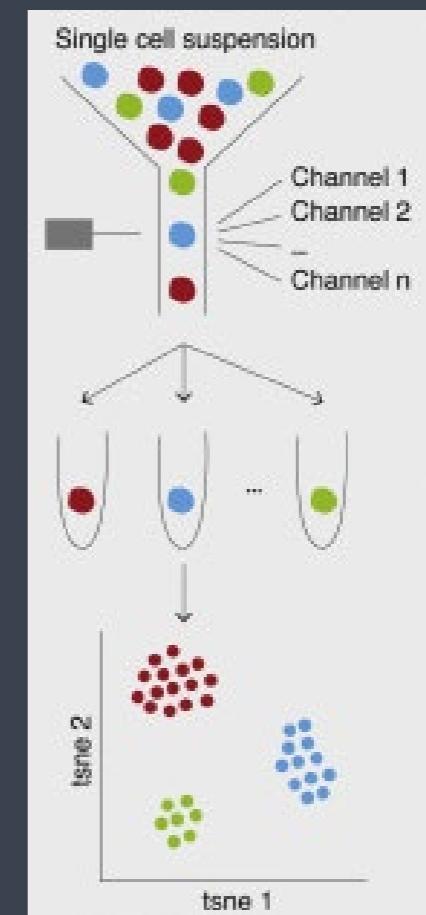


# Bulk vs Single Cell RNA-seq (scRNA-seq)

## Plate based low input RNA-seq



## Sorting to RNA-seq



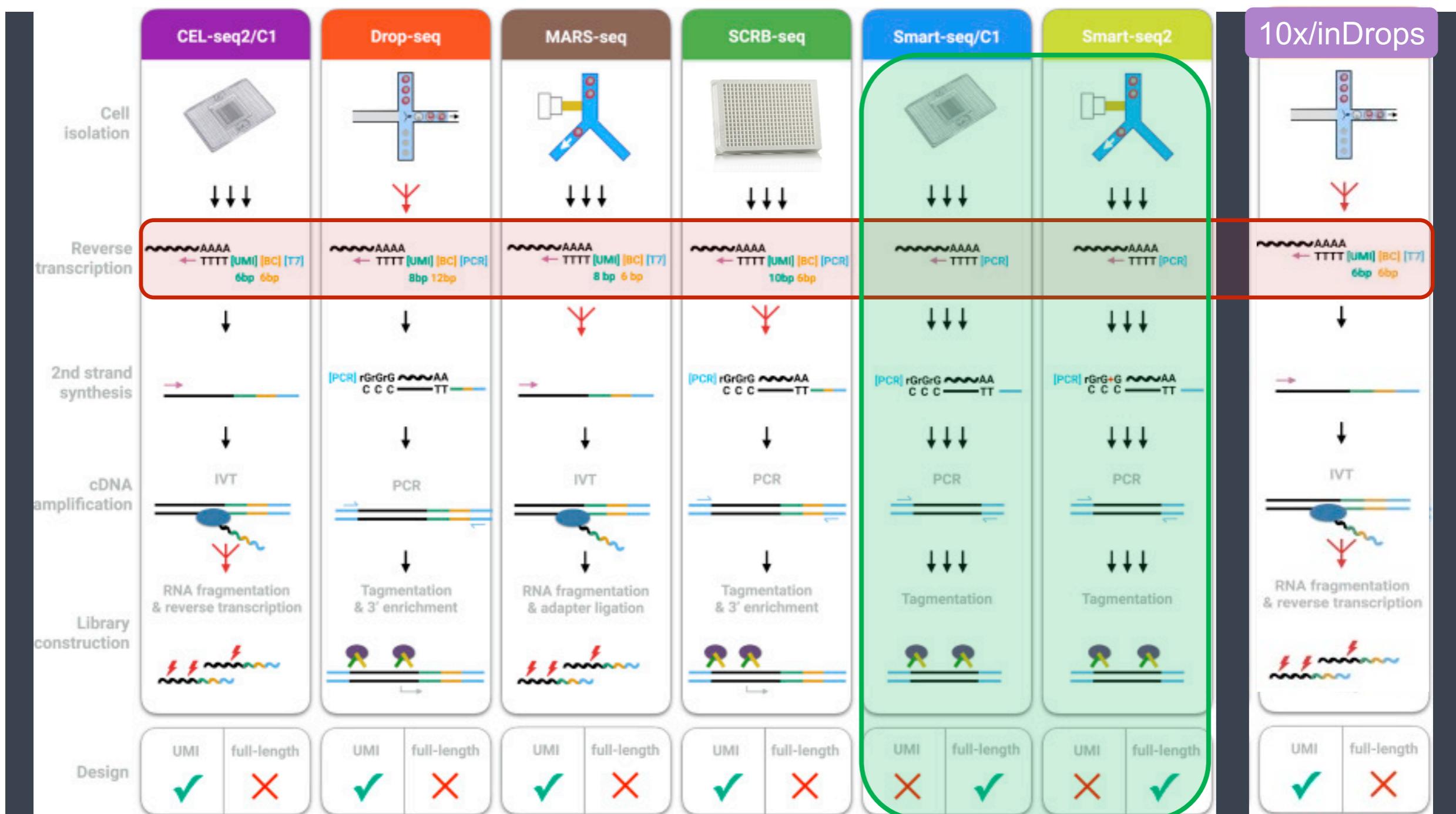
# Low Input RNA-seq

- Lower input = less chance to see mRNA of interest
- Sampling error & high technical variation
- Single cell methods will only capture 10-40% of expected mRNA per cell
- You will observe zero counts in even highly abundant RNA species
- Low input has fewer zeros than single cell methods.

# Between Bulk RNA-seq and scRNA-seq

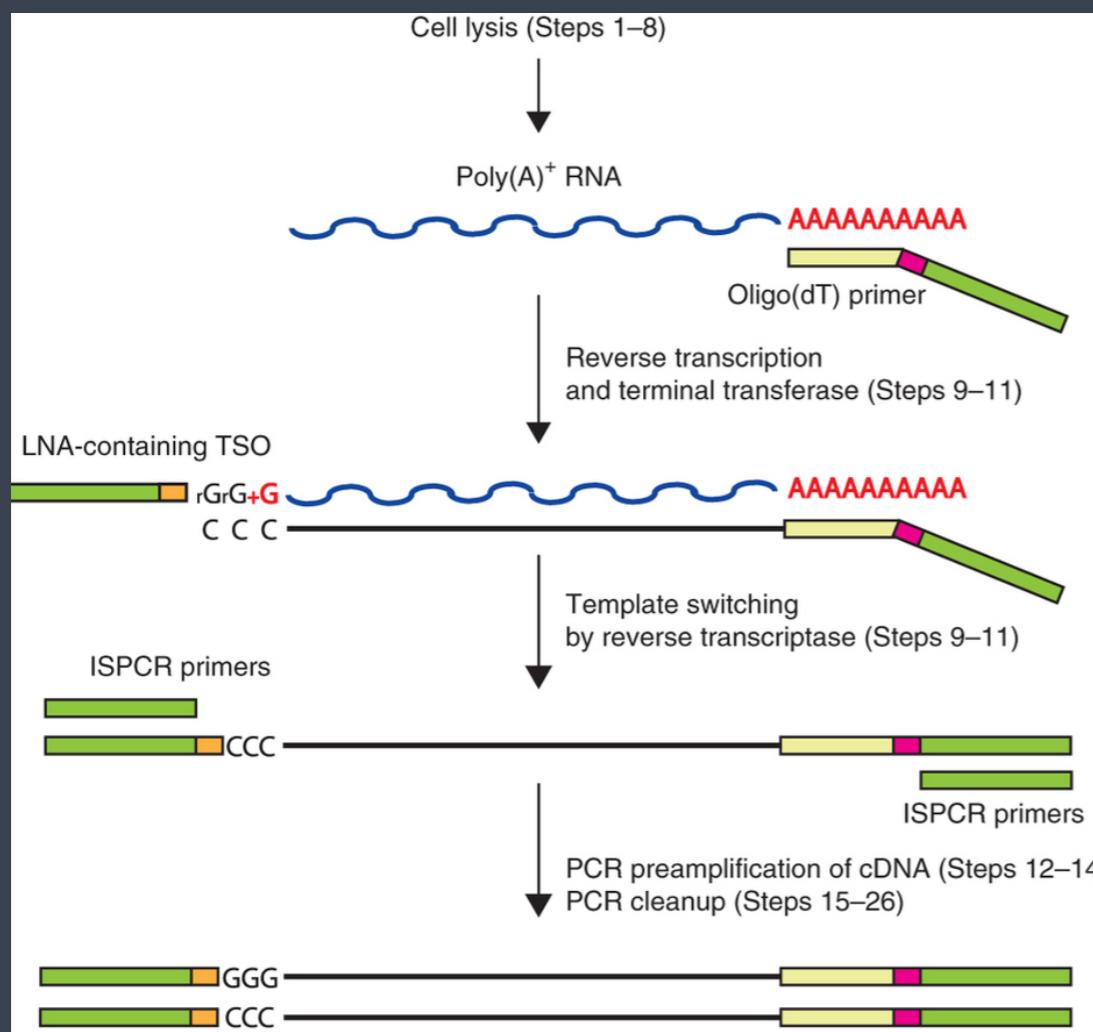
	Deep RNA-seq	Sort-seq	Low Input (SCRB-seq)	scRNA-seq
Transcriptome Coverage	High	High	Moderate	Low
Throughput	Moderate	Low	High	Low
Cell Subtype Information	None	Moderate	None	High
Sequencing Depth	Moderate	Moderate	Low	High
Cost per Sample	Moderate	Moderate	Low	High

# Single Cell / Low Input Methods

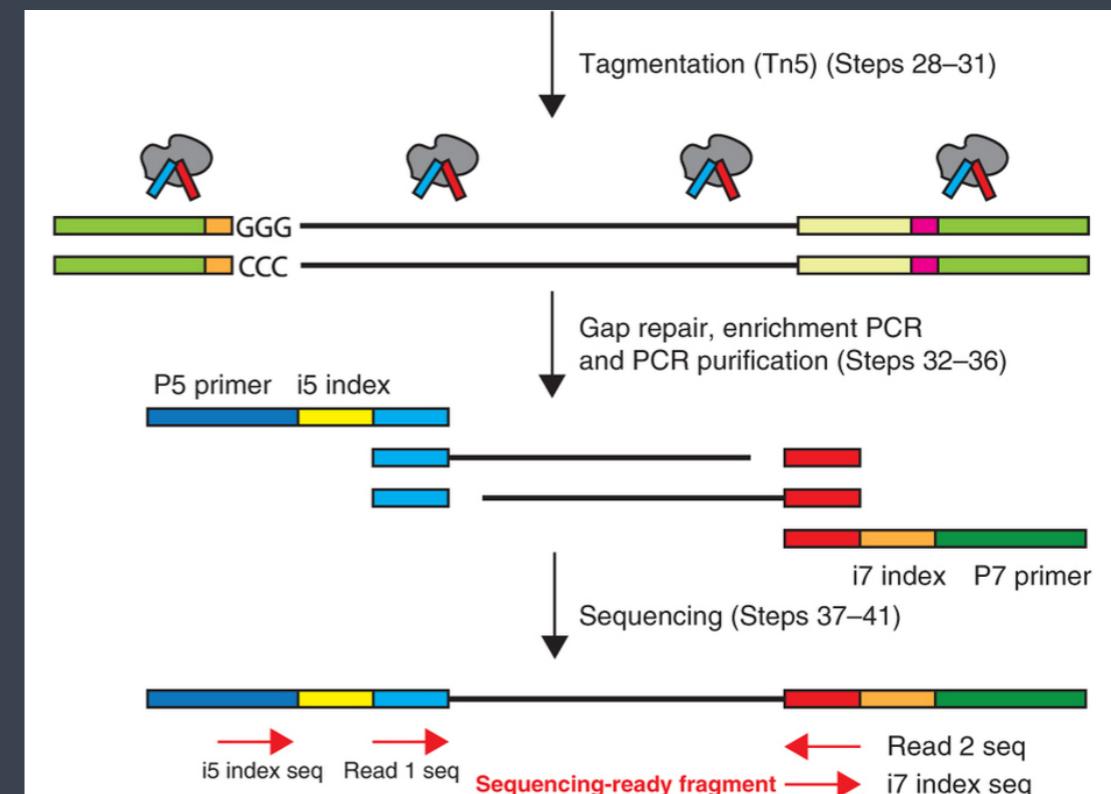


C. Ziegenhain. Comparative Analysis of Single-Cell RNA Sequencing Methods, *Molecular Cell* (2017).

# Smart-seq is gold standard for low input



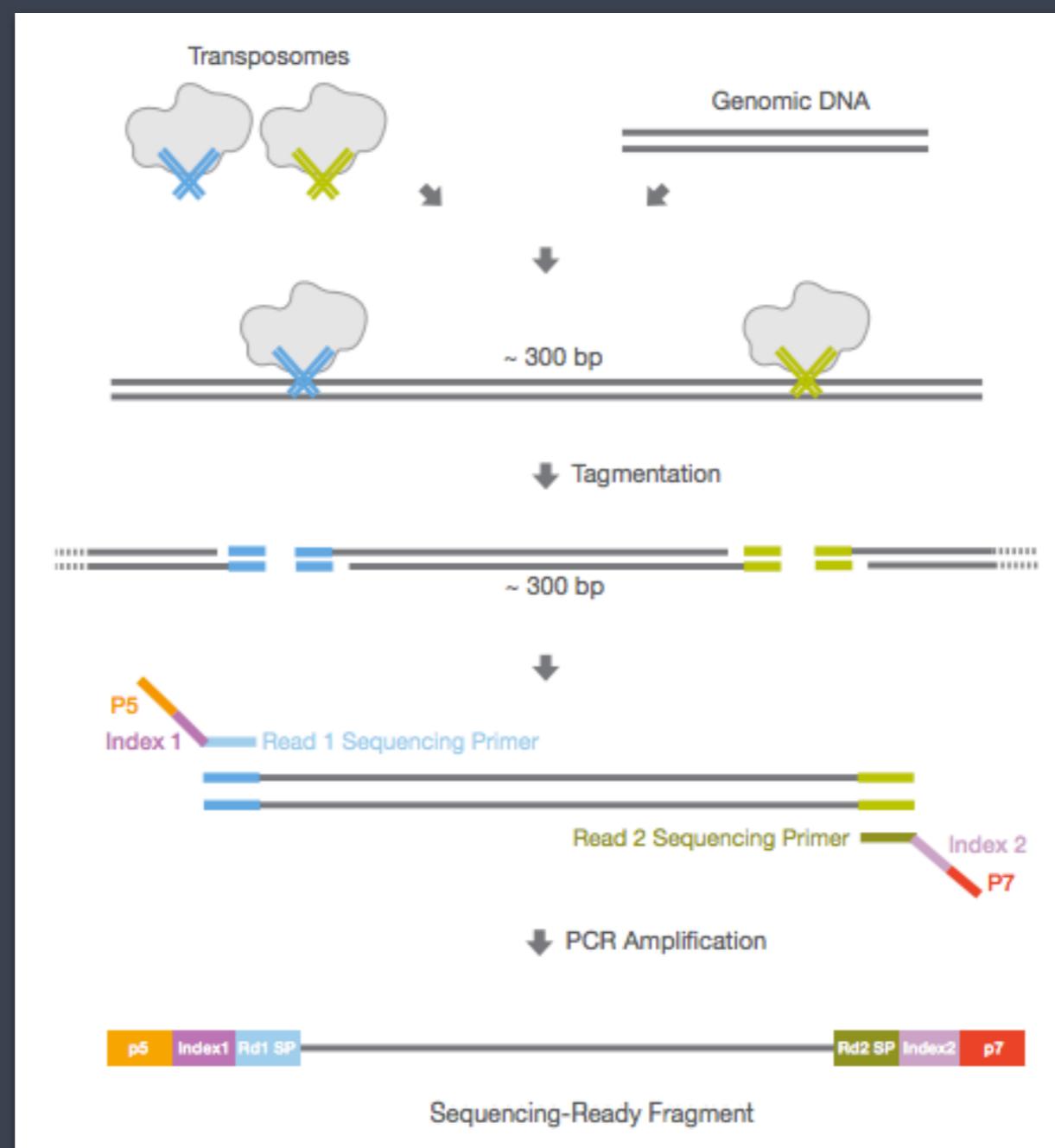
Another option is to shear the full length cDNA using Covaris



Samples identified by well index added at tagmentation or after shearing

# Tagmentation

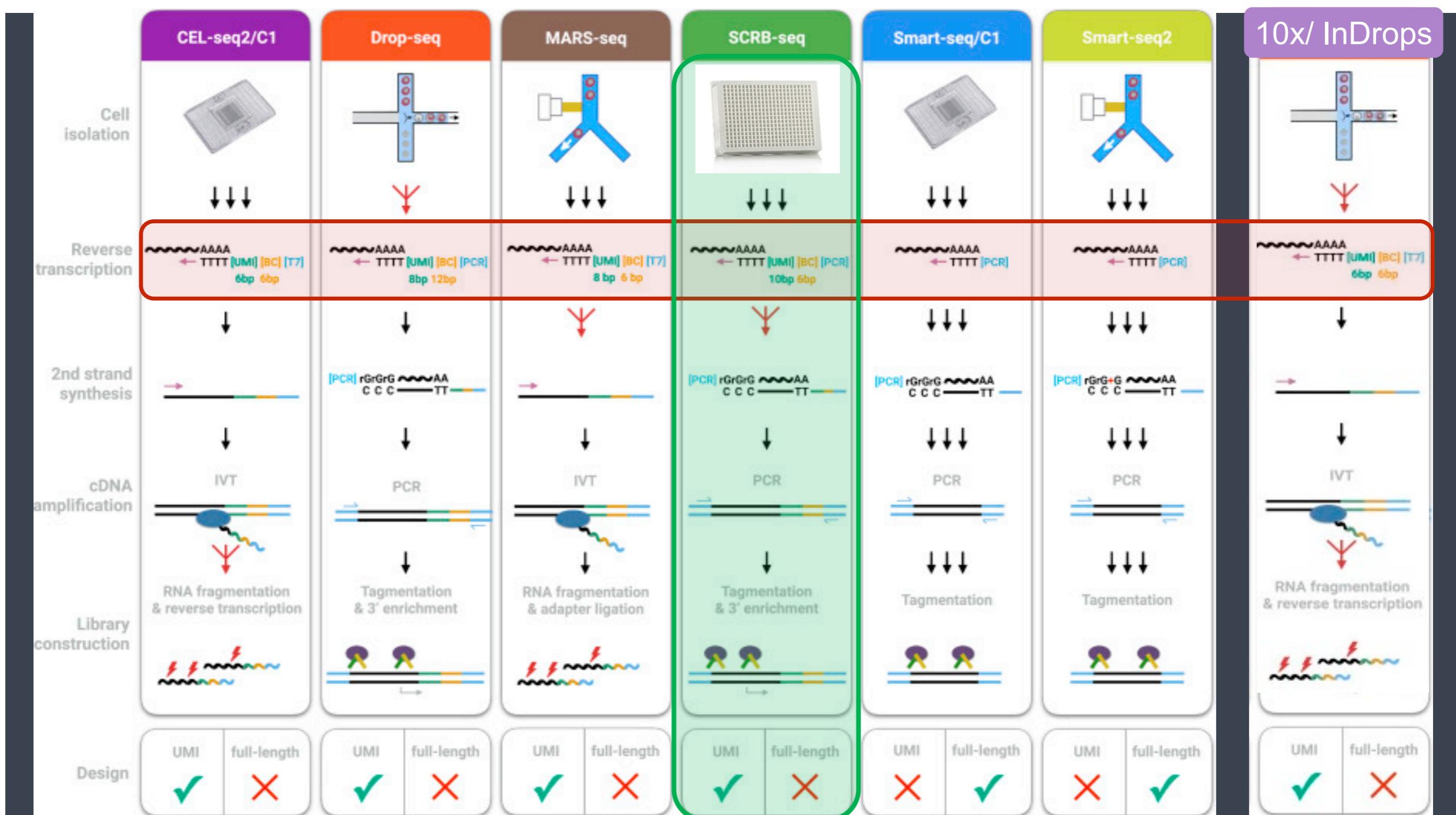
(DNA fragmentation facilitated by transposon activity)



# Tagmentation Approach

- *Nextera* from Illumina
- Very fast and efficient for DNA library preps
  - Works with small amounts of DNA
  - Important to RNase treat your sample
  - Needs precise DNA quantitation (Qubit)
- Often used as last stage of low input RNA-seq library protocols

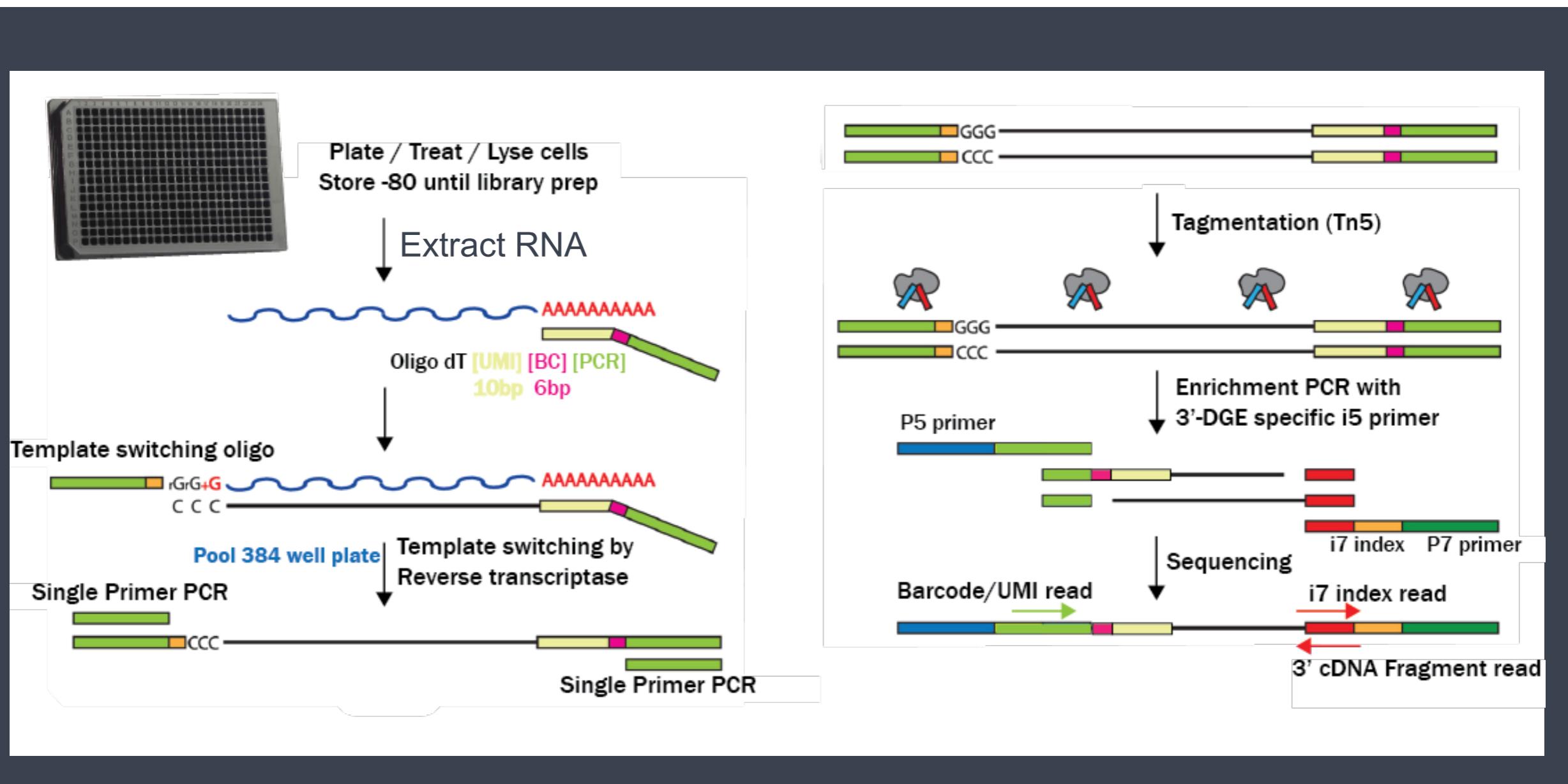
# Single Cell / Low Input Methods



C. Ziegenhain. Comparative Analysis of Single-Cell RNA Sequencing Methods, *Molecular Cell* (2017).

# 3'-DGE: Digital Gene Expression

## Single Cell RNA Barcoding (SCRB-seq)



# Unique Molecular Index

UMI  
Unique N-mer per transcript

cDNA ← ----- TTTTTT(20) – NNNNNN(6-8) – Barcode – Adapter



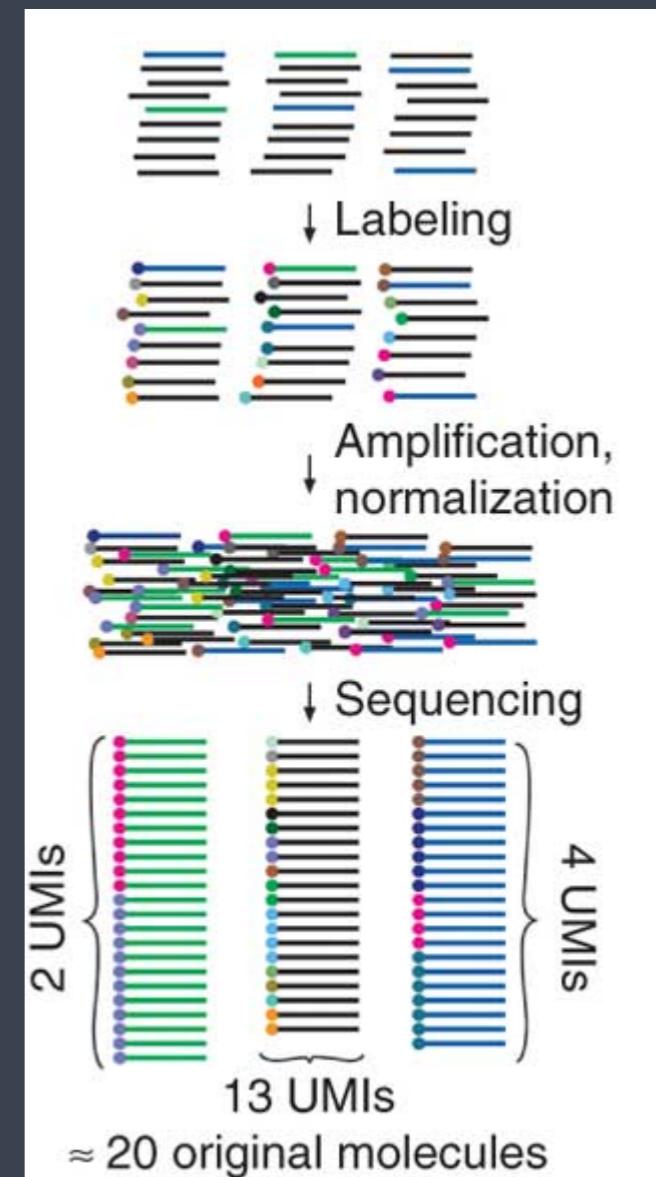
AAAAAAA

Cell or Well Barcode

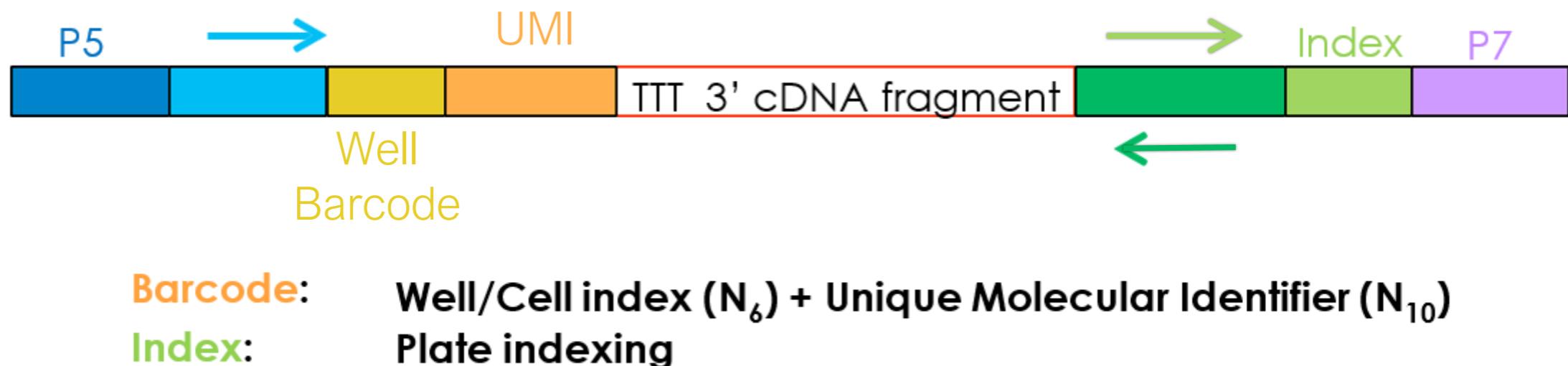
- Plate based assay need 96 or 384 well plate of barcoded/UMI-tagged primers for RT.
- Droplet based methods have this pool in each drop.

# Library Amplification Bias

- Final step of library prep is amplification
- Introduces library bias
  - Some products preferentially amplified
- Fewer cycles = less bias
- Unique Molecular Index (UMI)
  - Accurately quantitates samples

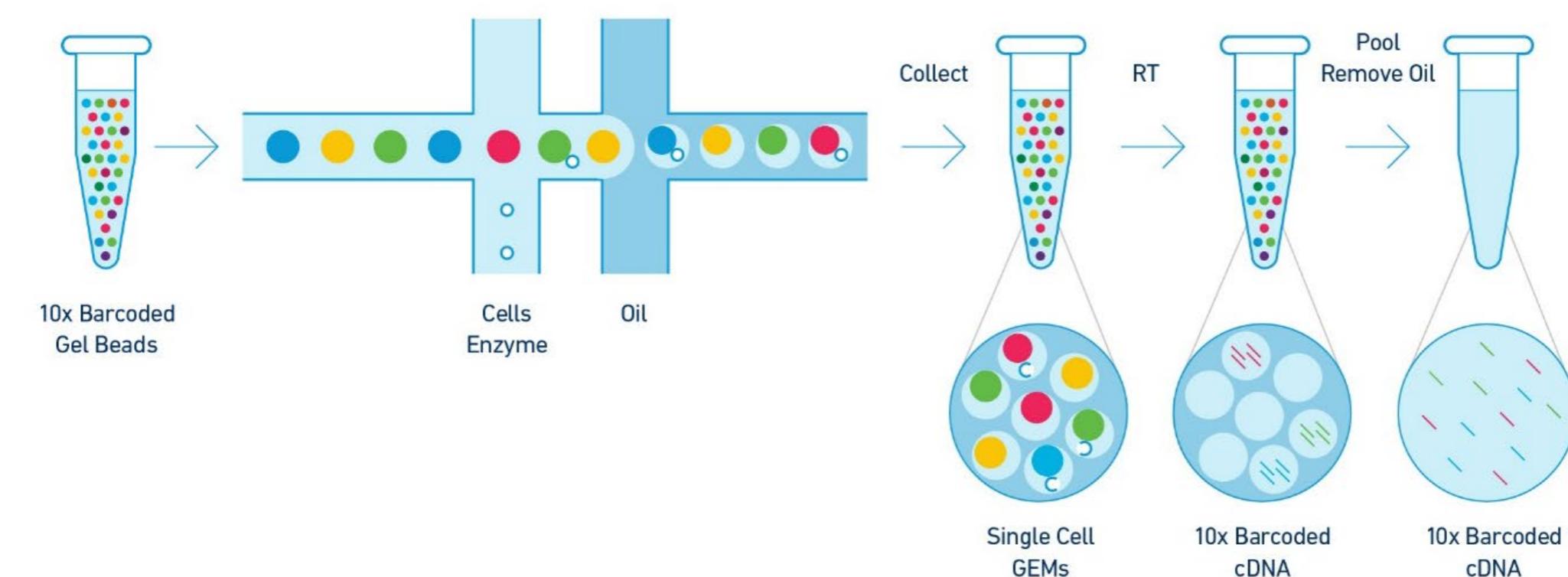


# Final 3'-DGE Library Structure

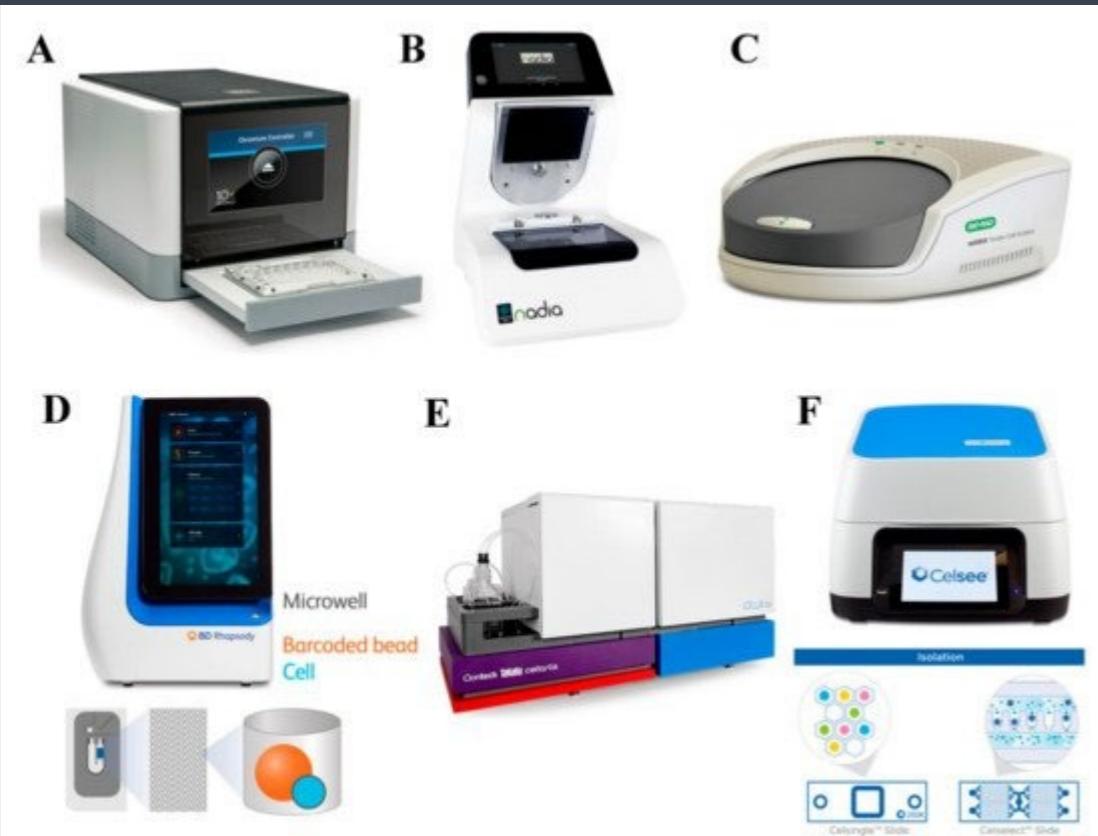


- 17 cycles on **Read 1**      Well Barcode & UMI
- 8 cycles on **Index 1**      Plate Index for multiple plates
- 46 cycles on **Read 2**      cDNA fragment

# 10x Genomics



# Single Cell Platforms



- (A) 10x Chromium
- (B) Nadia (Dolomite Bio)
- (C) ddSEQ (BioRad)
- (D) BD Rhapsody
- (E) ICell8 (Takara)
- (F) Celselect (Celsee)

And more.....

# Single Cell Core Workflow

- Good sample prep is the key to success.
- Well planned pilot experiment key to evaluating sample preparation.
- Do not rush to the final experiment.



# Key to Success: Sample Preparation

- High cell viability (>90-95% preferred)
- Minimal free-floating RNA
- Single cell suspension
- Dissociation protocol is cell type dependent
- Primary samples are much more difficult
- Cryopreservation or Nuc-seq works on some sample types

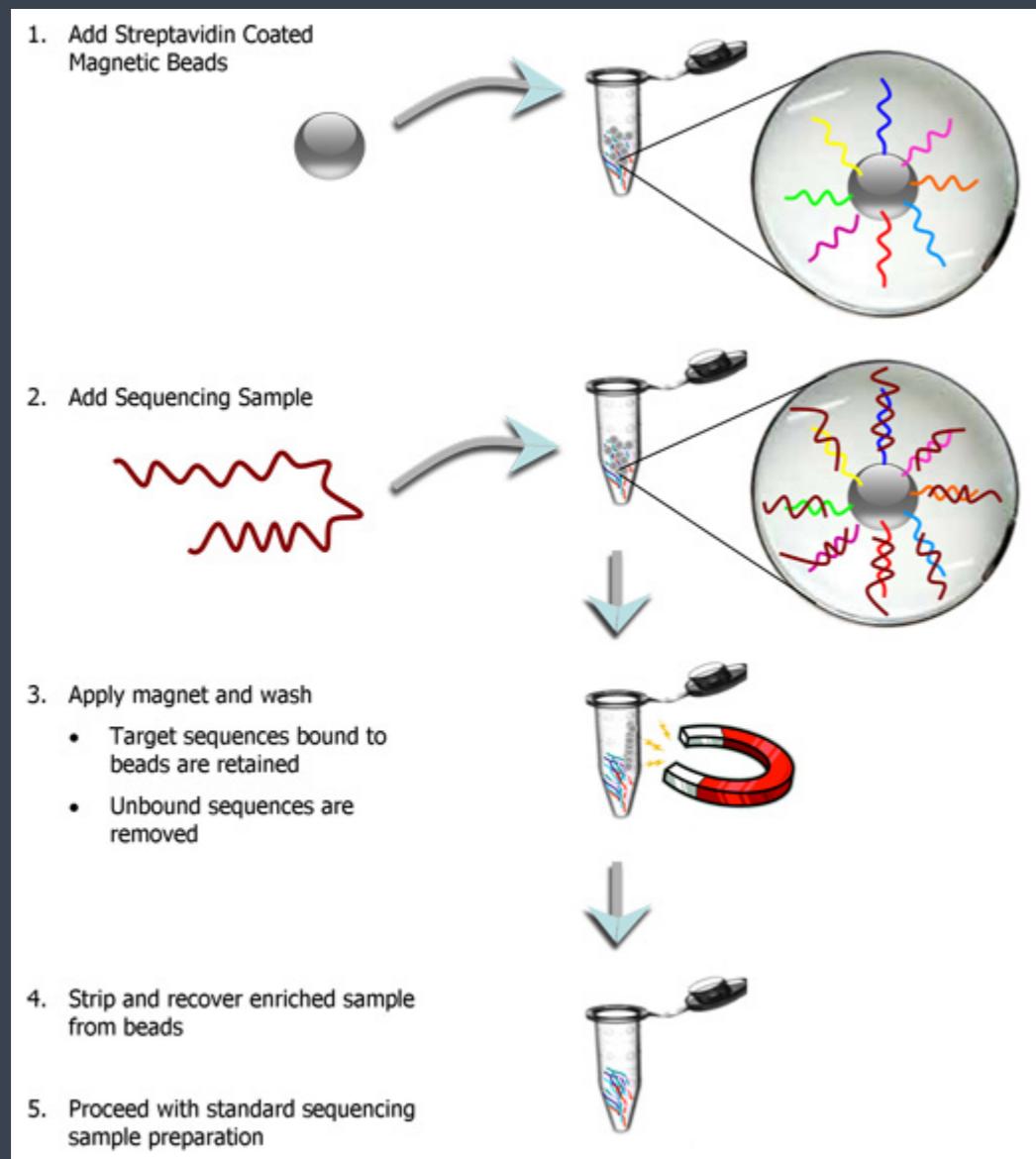
# Single Cell / Low Input Methods

- Only use if needed for experiment.
- Maximal transcriptome coverage ~60%
- All commonly used methods all rely on PolyA tail.
- If you can get more starting material then you will get better results.
- Plan a small scale starter experiment to see if protocol will give useful results.

# Sequencing Library Preparation

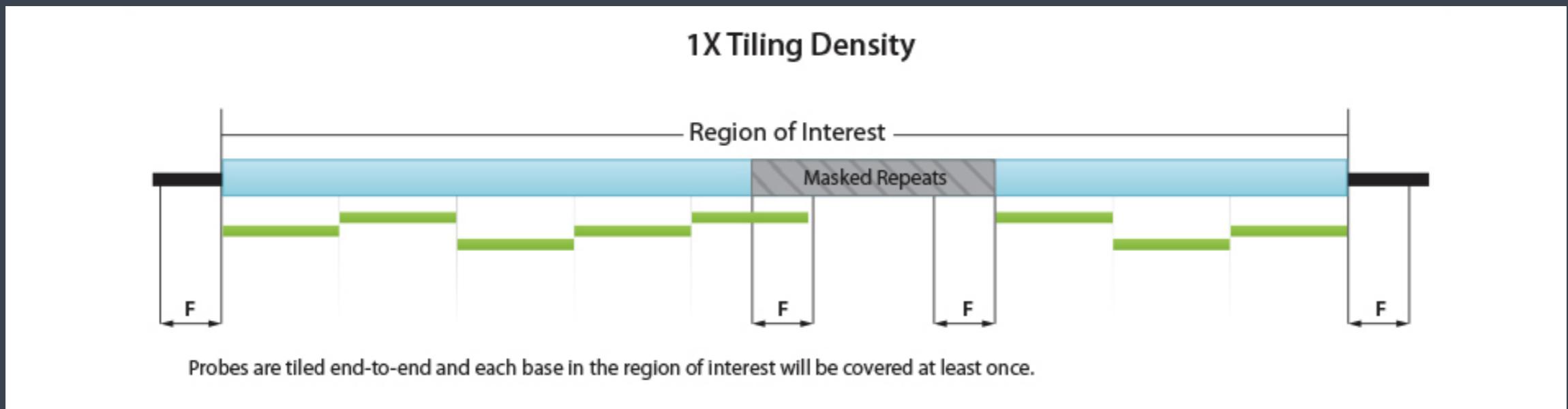
- Starting Material: purifying RNA
- RNA-seq library preparation
- Multiplexing & Sequencing
- Single cell / low input methods
- Capture sequencing

# Transcript Enrichment/Depletion: Capture Sequencing



- Capture targeted sequence using biotinylated RNA or DNA bait
- Library pool applied to beads
- Retain only library covering genes of interest
- Reduces total number of reads needed for sequencing
- IDT lockdown probes expensive but good for small number genes

# Capture Probes



- Tiling is the number of times a base is covered by a different probe.
  - Start with 2x tiling density.
- Difficult to design probes if looking at a single gene family or pseudogenes.

# Transcript Enrichment: Capture Sequencing

- Use to deplete a highly abundant transcript that will swamp out our reads.
- Use to get more reads on a region of interest.
- Can be used in combination with ATAC-seq or other nucleosome positioning methods to focus in on region of interest.
- Critical: use unique dual indexing for capture experiments.

# Final Thoughts

- Practice your library prep on a control sample.
- Be sure you understand each step in library prep.
- Talk to someone who has done the protocol before starting.

qPCR

Precise quantitation is key to effective sequencing!

# Useful Websites

- [support.illumina.com/](http://support.illumina.com/)
- [seqanswers.com/](http://seqanswers.com/)
- [core-genomics.blogspot.com/2012/04/how-do-springerbeads-work.html](http://core-genomics.blogspot.com/2012/04/how-do-springerbeads-work.html)
- [www.broadinstitute.org/files/shared/illuminavids/SamplePrepSlides.pdf](http://www.broadinstitute.org/files/shared/illuminavids/SamplePrepSlides.pdf)