

# RNA Sequencing Library Preparation



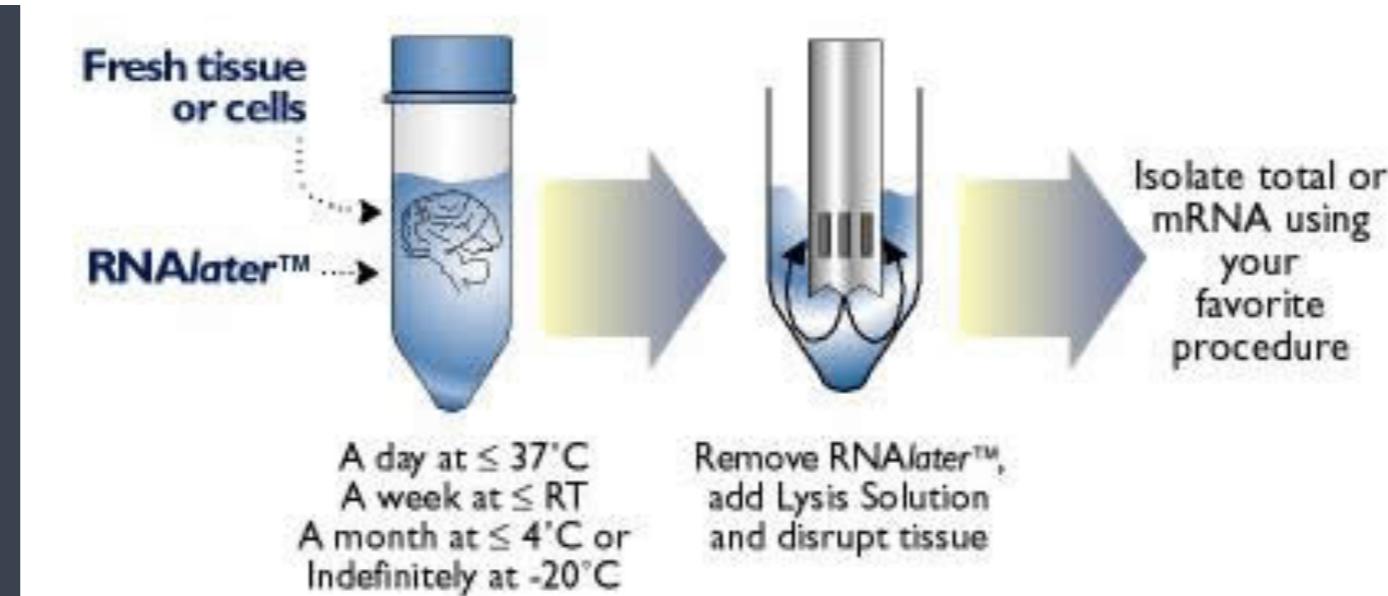
# RNA-seq Library Preparation

- Starting material
- Library preparation
- Multiplexing
- Low input / single cell RNA-seq

# Purification and QC of RNA

- Start with highest quality RNA possible
  - Column or bead purification
  - DNase sample
  - Clean up any Trizol or phenol contamination with secondary column or bead purification
- Accurately quantify RNA
- Assess quality of RNA

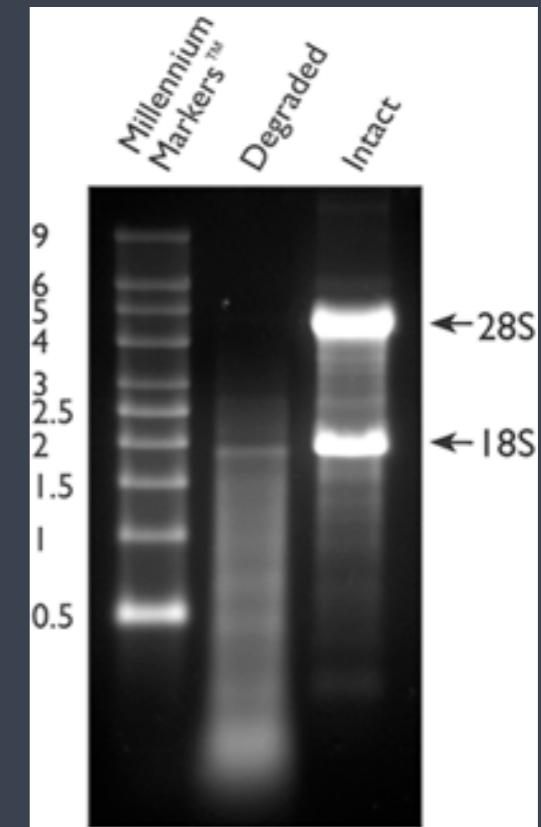
# RNA Extraction: Tissues / Trizol



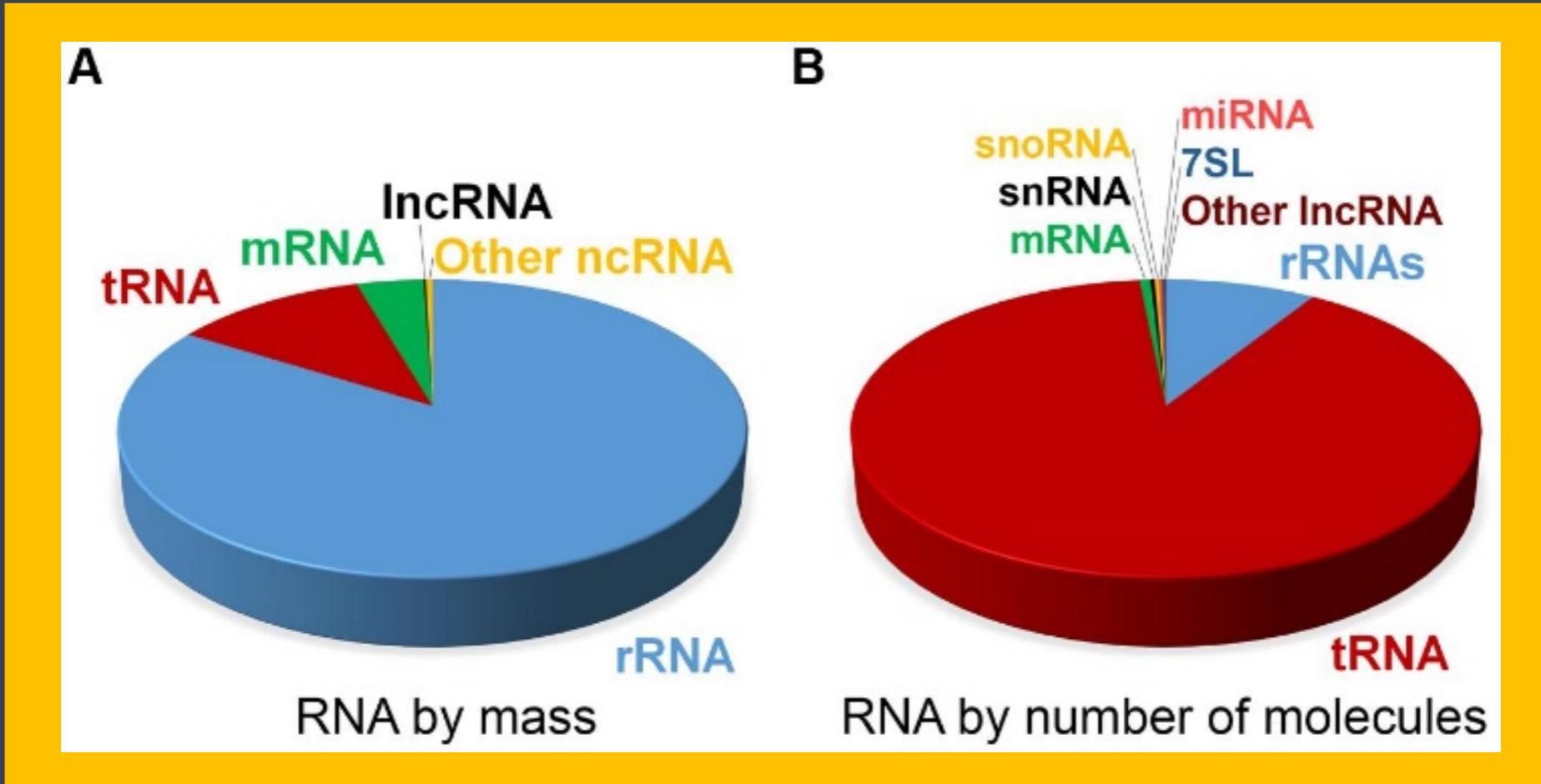
- Store tissue in RNAlater if possible.
- Keep tissues as cold as possible.
  - Work in cold room
- After homogenization suggest column based cleanup.

# 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
- Quality
  - Visualize on gel
  - Agilent Bioanalyzer or Tapestation (RIN)

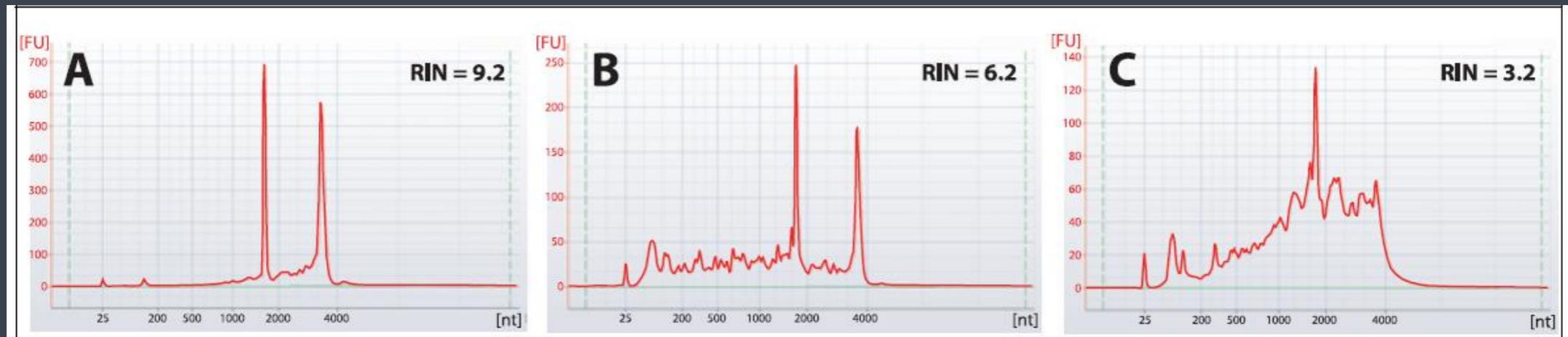


# RNA Enrichment



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

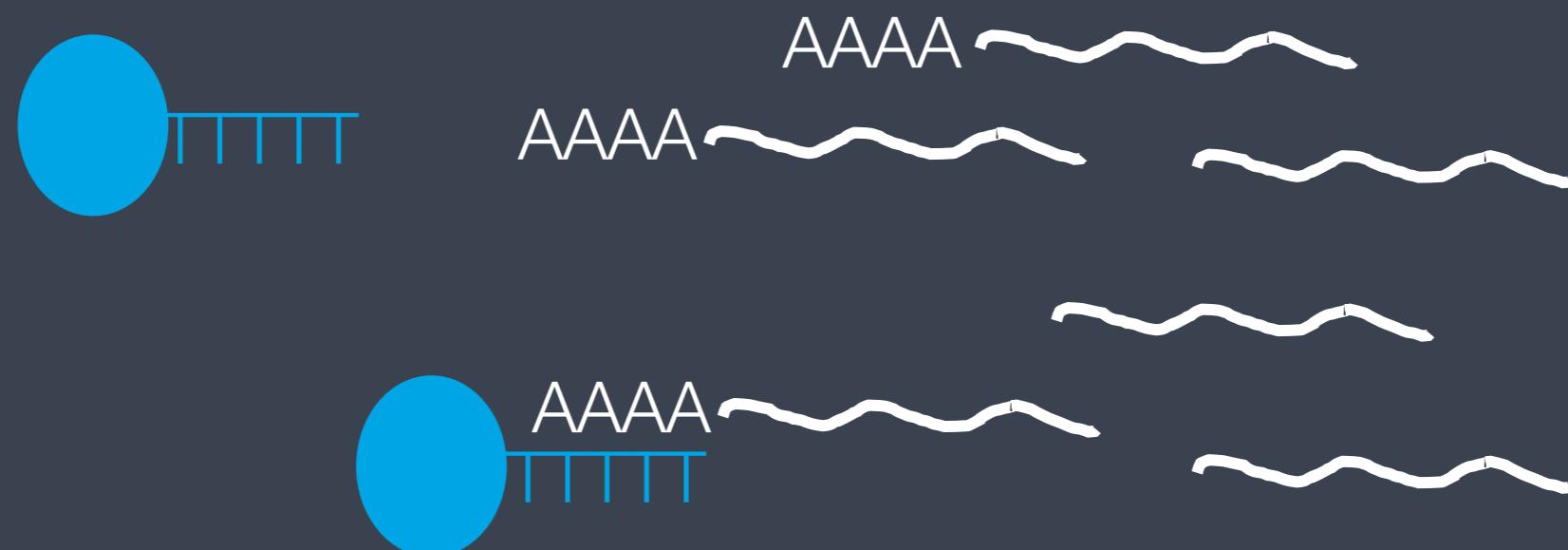
# RNA Quality



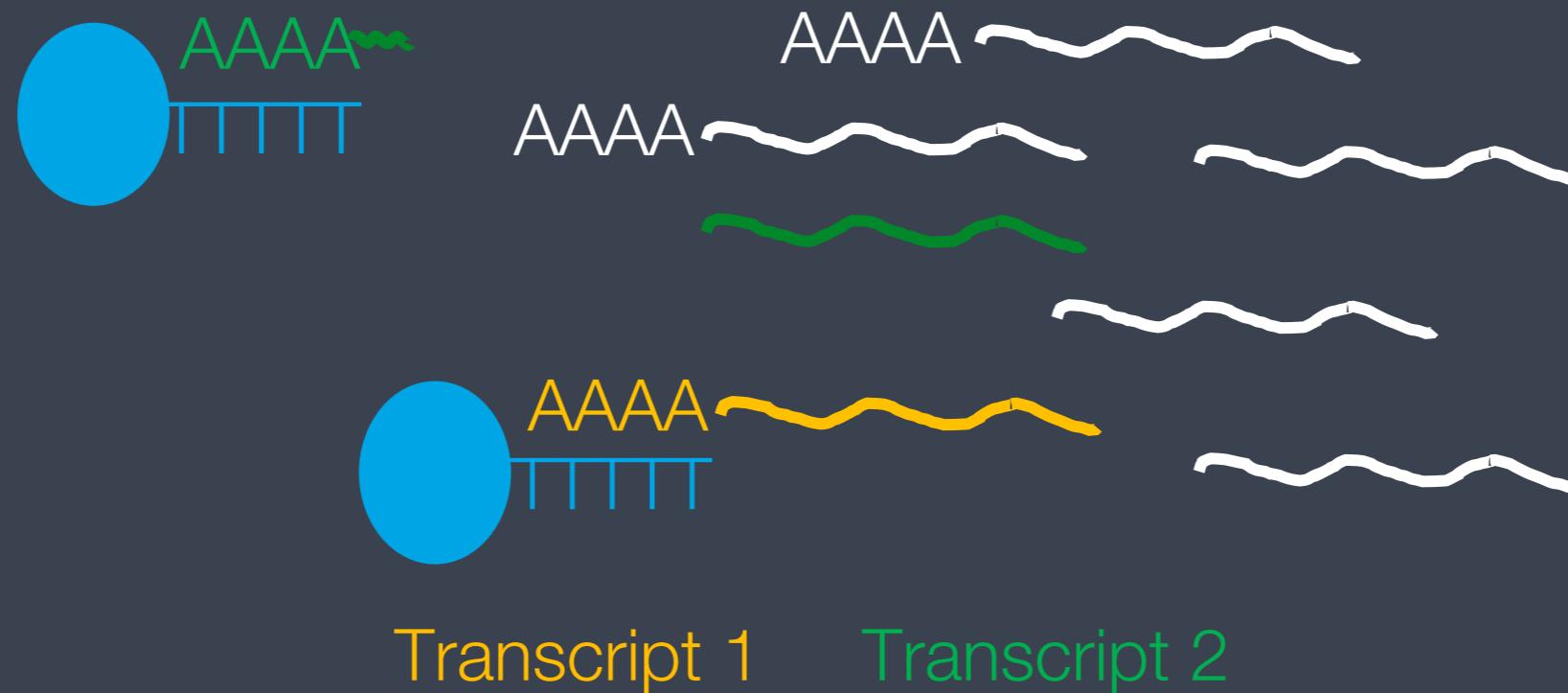
- High quality RNA needed for mRNA libraries
  - RNA integrity number (RIN)  $> 8$
- Degraded samples should only be used to make a “total” RNA-seq library – rRNA removal
- FFPE & Archival Samples

# mRNA (polyA) Purification

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



# mRNA Purification of Degraded Samples



- PolyA tail no longer attached to transcript.
- Results in random loss of transcripts between samples.

# 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 a degradation signal

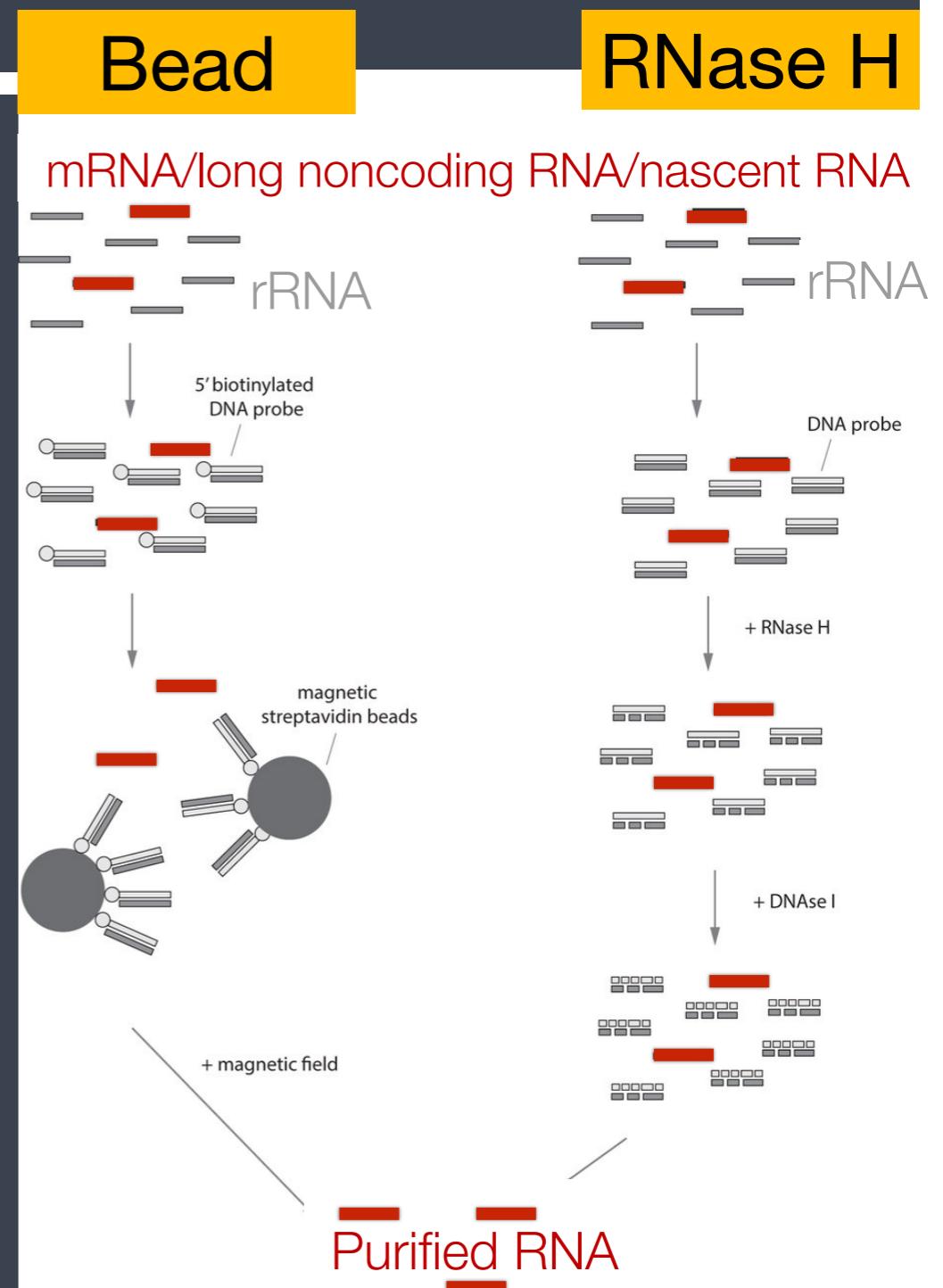
# rRNA Depletion

## ➤ Illumina: TruSeq

- Probes hybridize rRNA on magnetic beads
- RNA of interest remains in supernatant

## ➤ KAPA: RiboErase

- Probes hybridize rRNA in solution
- Hybrids are digested with RNase H
- Probes digested with DNase I



Modified from: Scientific Reports 6, article 37876 (2016)

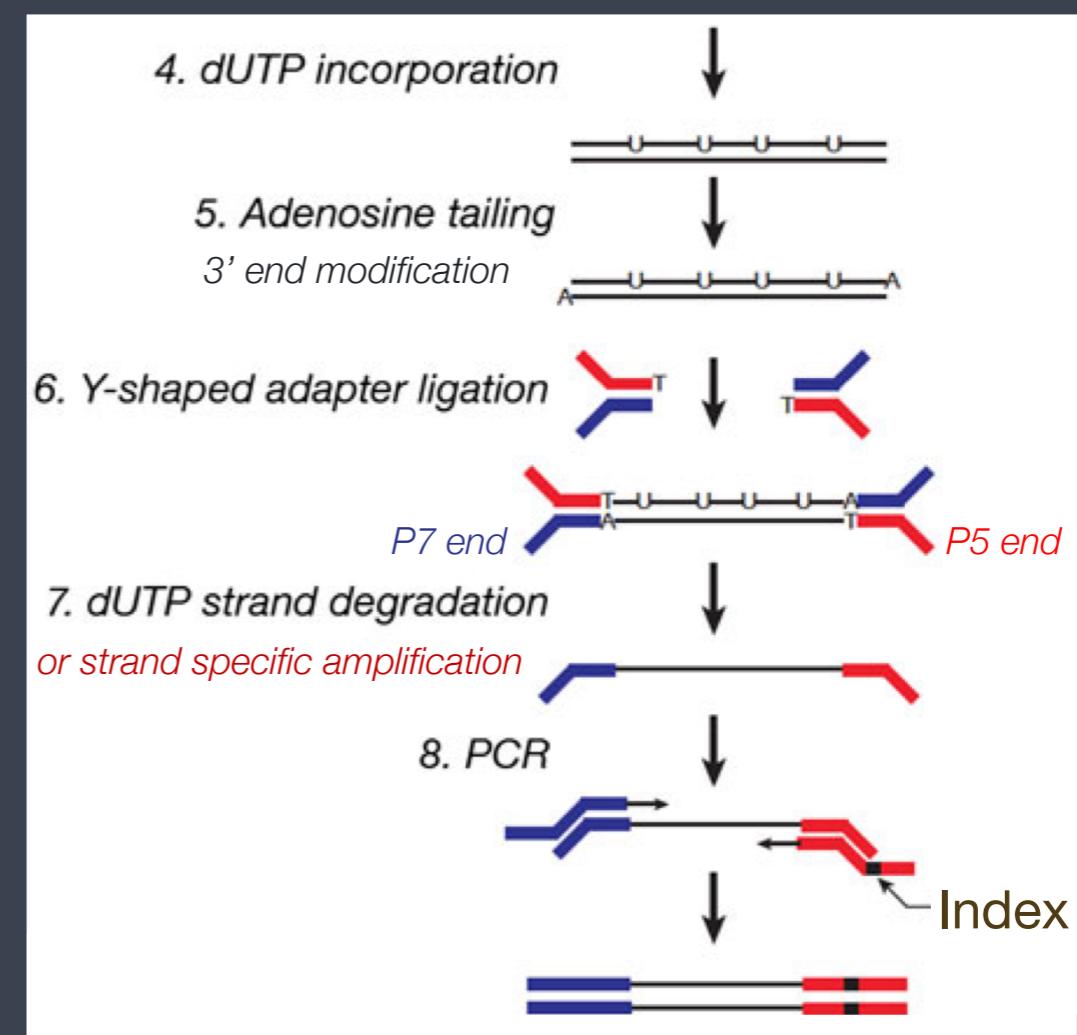
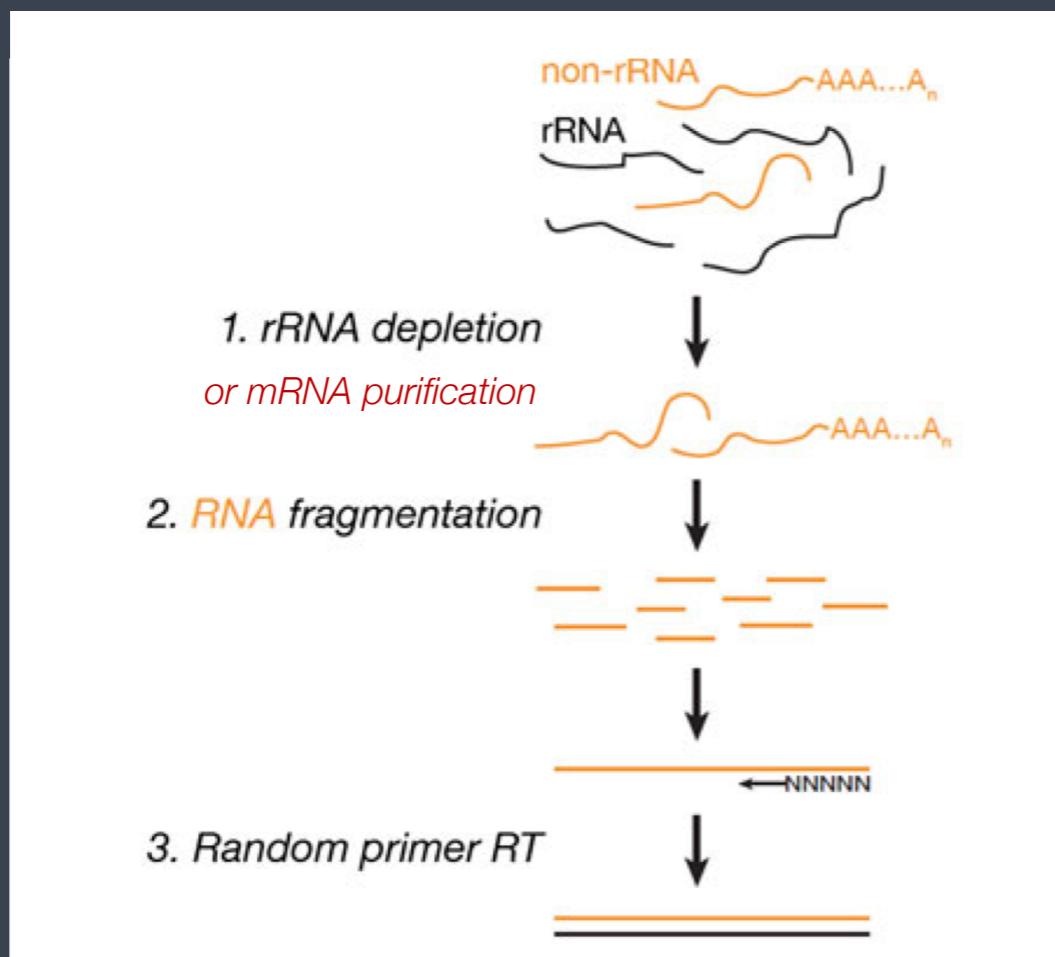
# RNA Purification

- Test your sample storage and extraction method prior to any large-scale experiment.
- Determine the quantity and quality of your sample.
- Decide if you are performing mRNA-seq or total RNA-seq.

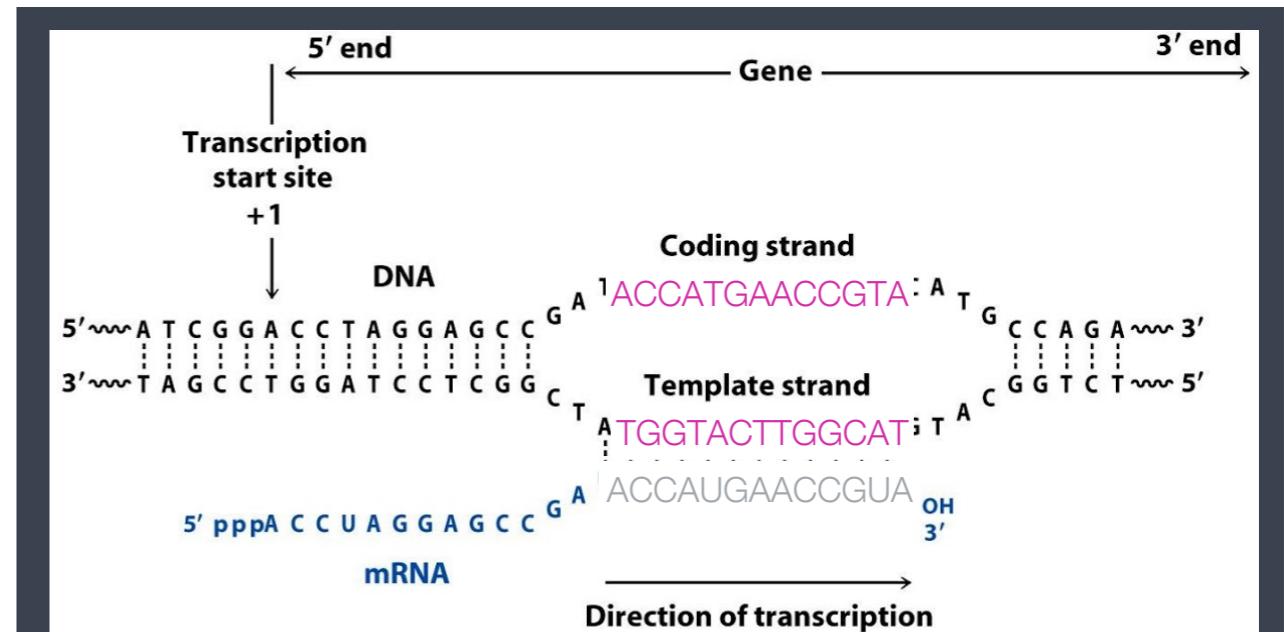
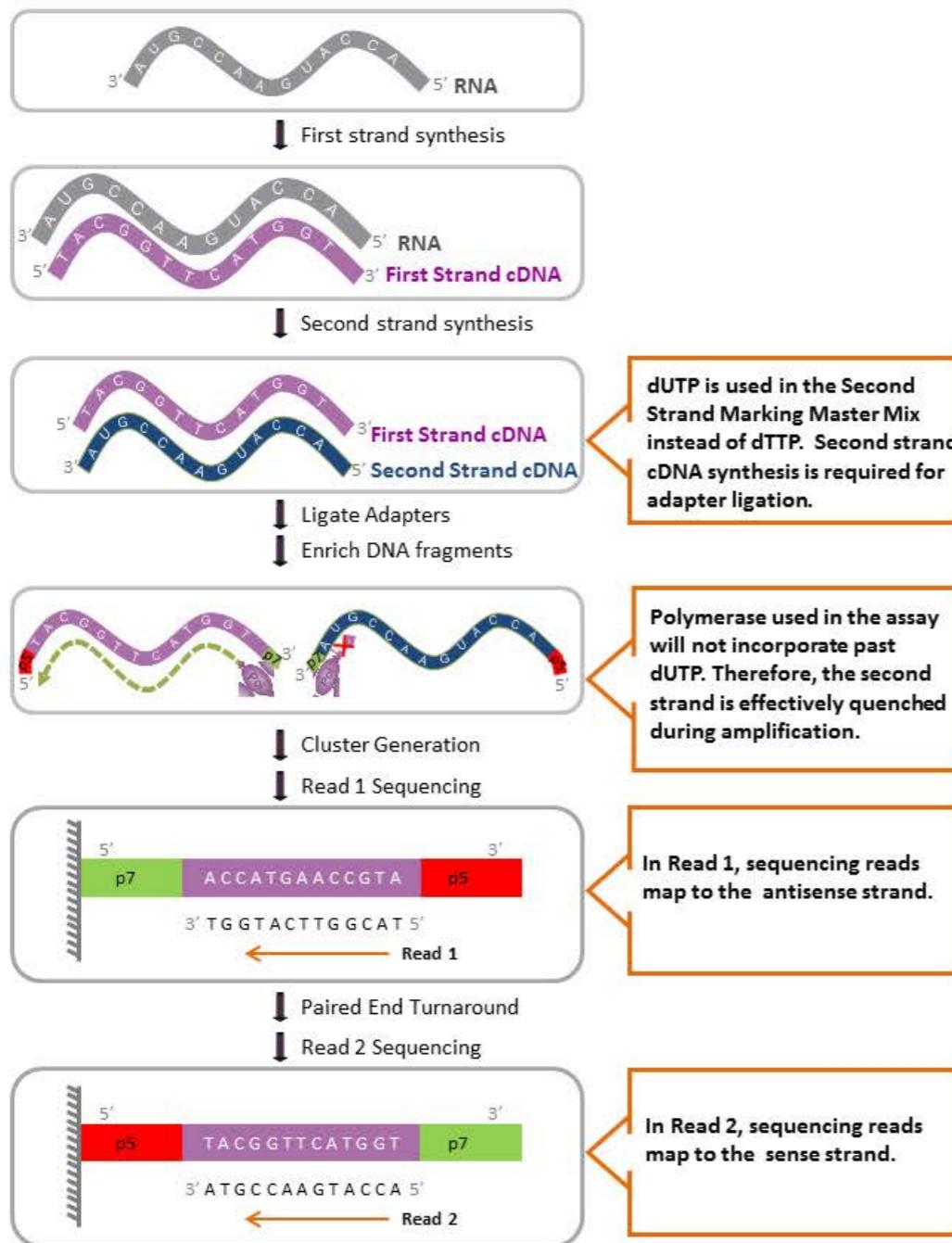
# RNA-seq Library Preparation

- ✓ Starting material
- Library preparation
- Multiplexing
- Low input / single cell RNA-seq

# RNA-seq Stranded Library Prep (dUTP method)

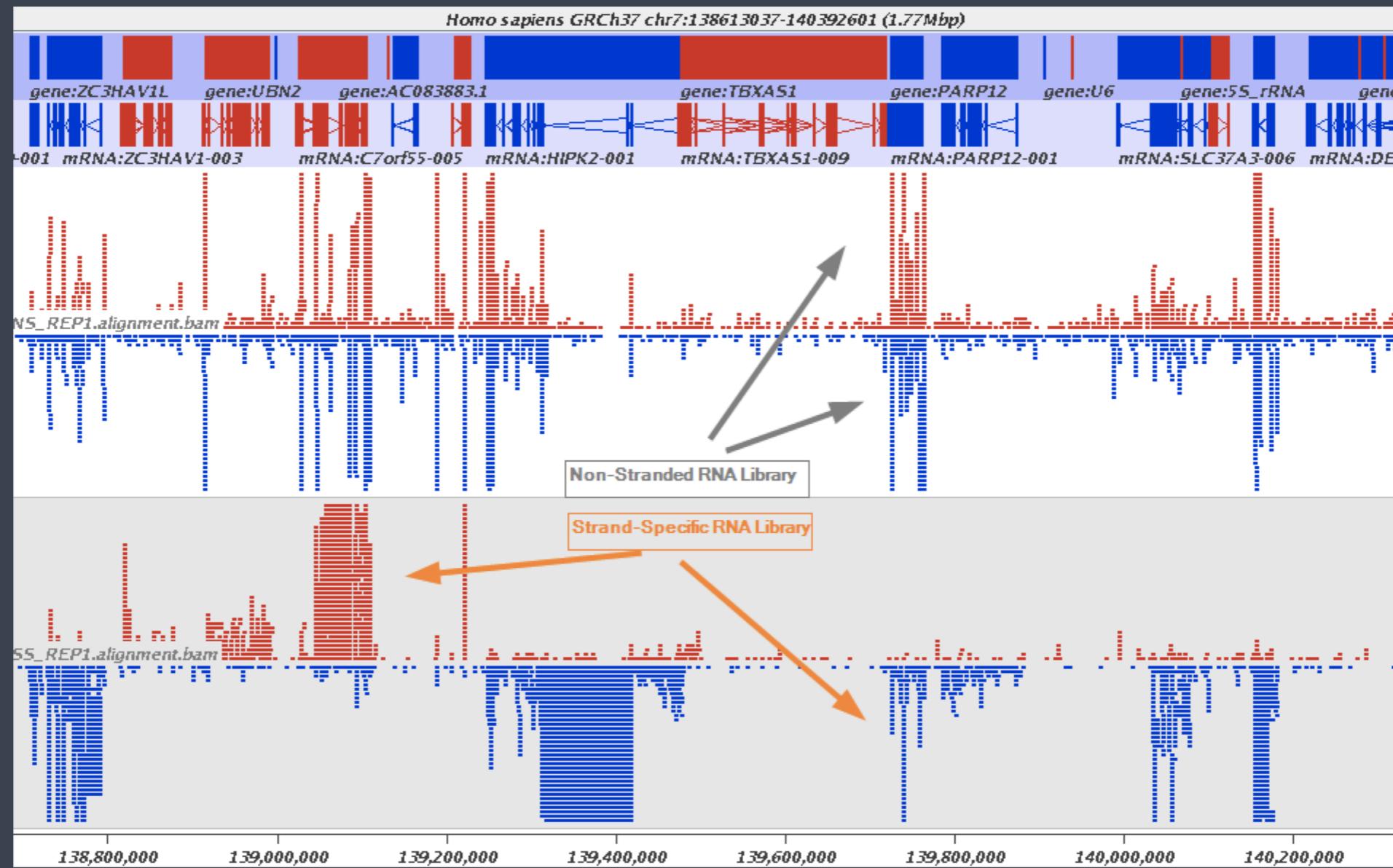


# Library Strandedness



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

# Library Strandedness

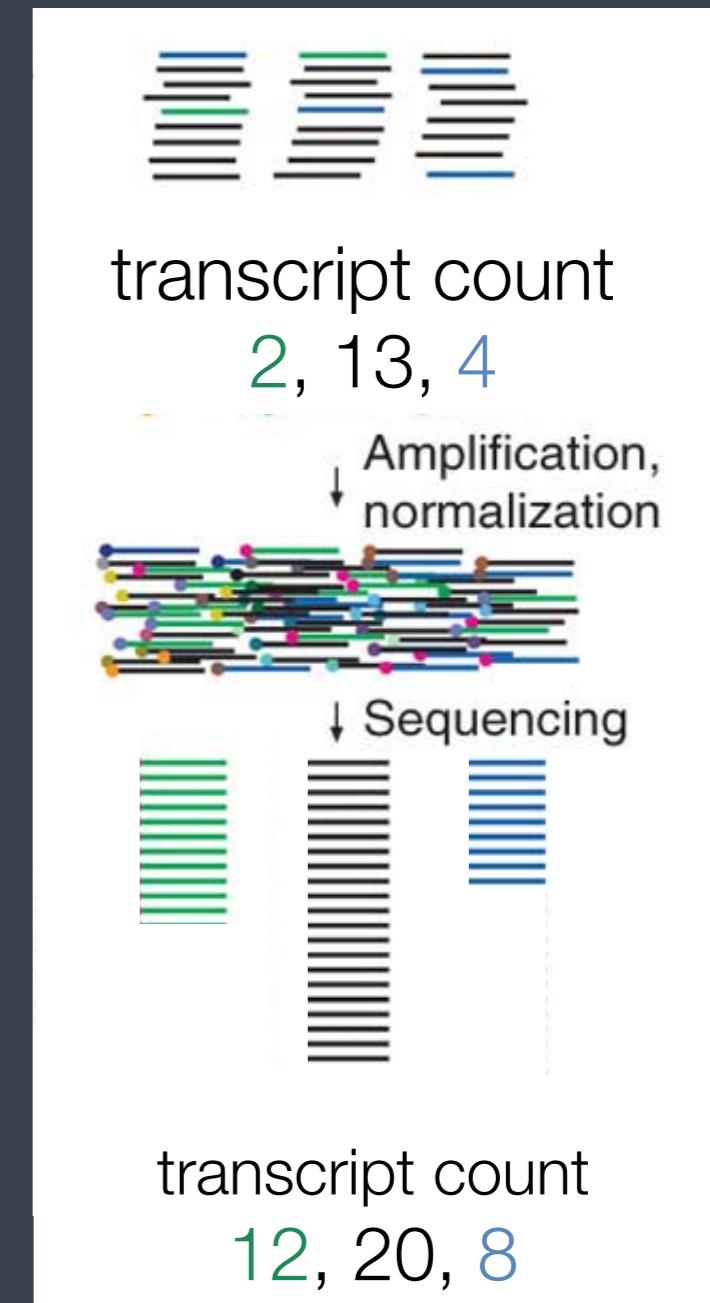


# RNA-seq Library Preparation

- ✓ Starting material
- ✓ Library preparation
  - Amplification bias
  - Library QC
  - Multiplexing
  - Low input / single cell RNA-seq

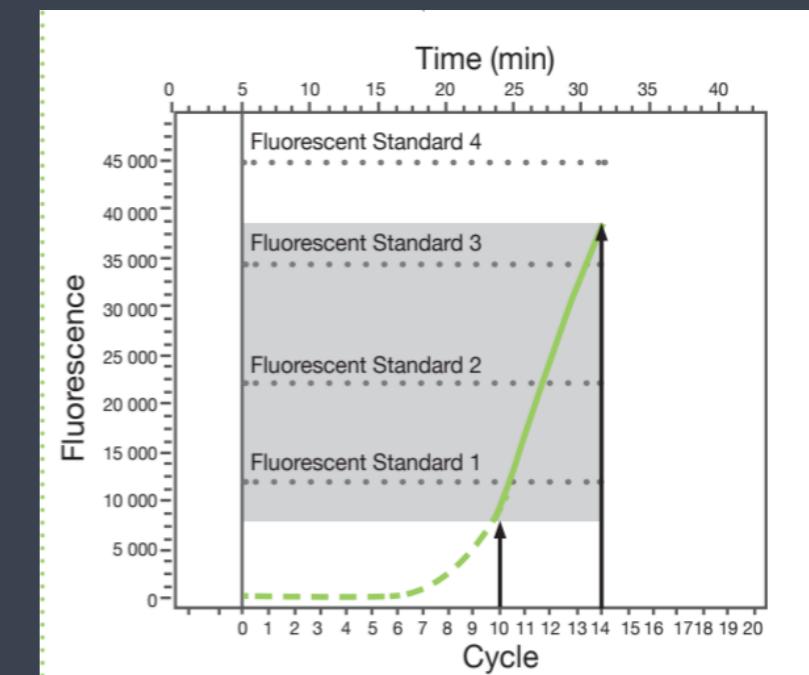
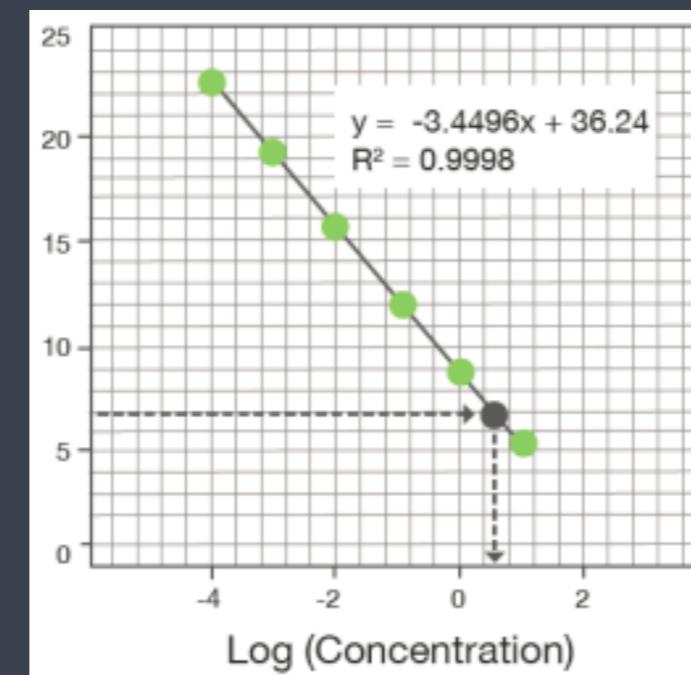
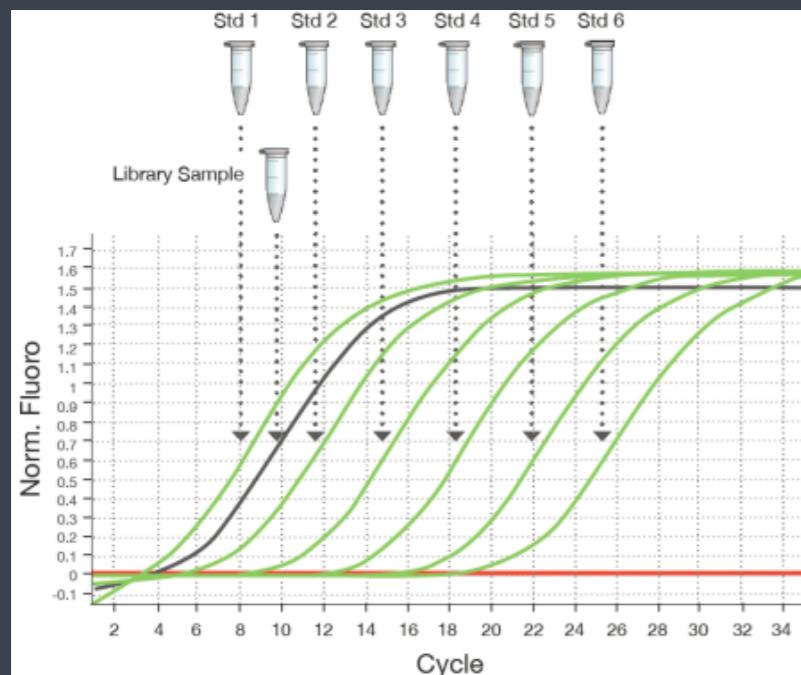
# Library Amplification Bias

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



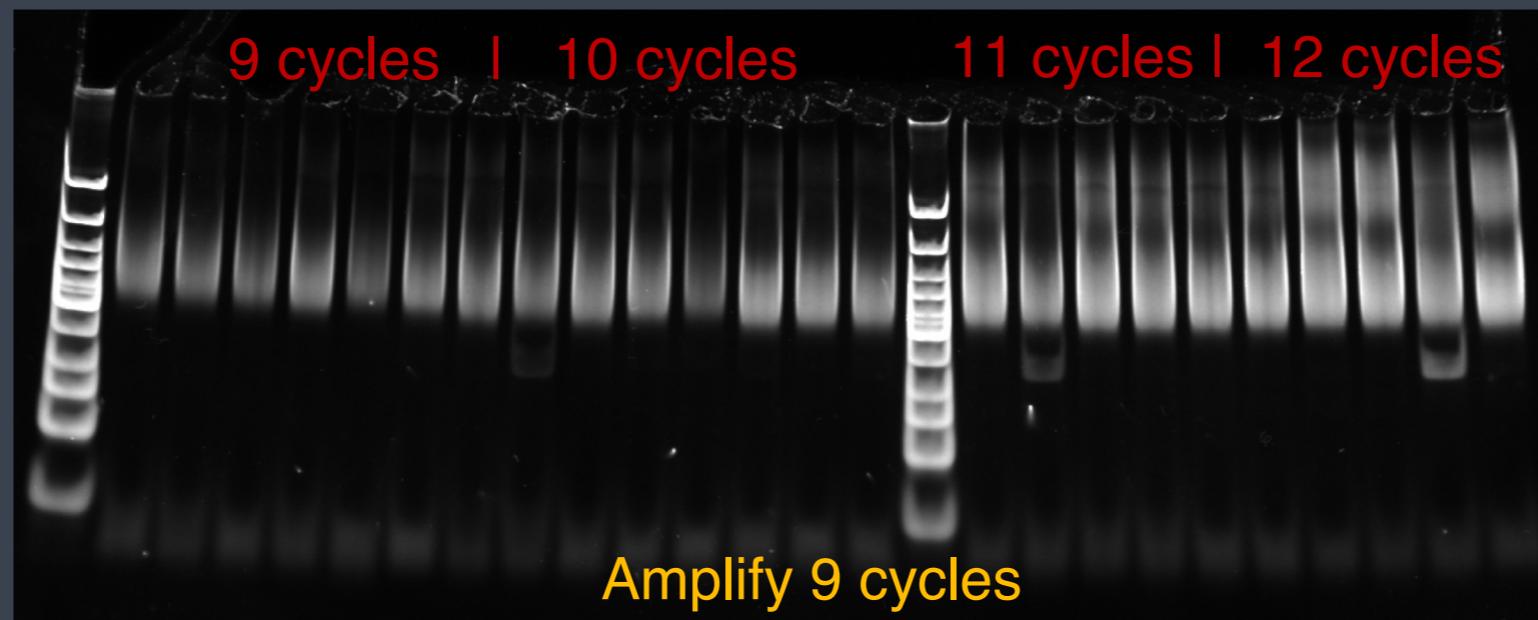
# Limited Cycle Library Amplification

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

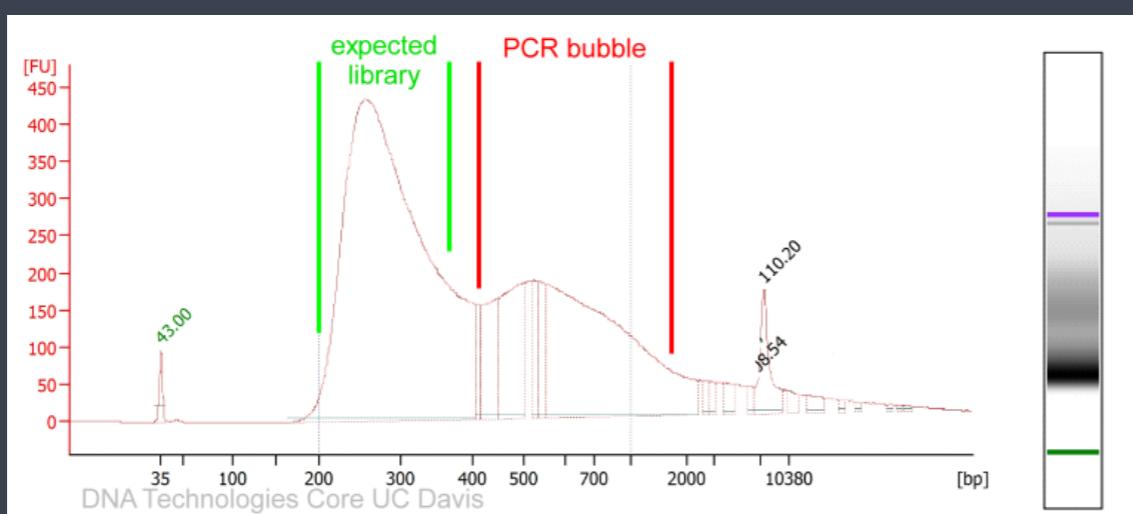


- Amplify only the number cycles needed to get enough product for sequencing (20ul of 4nM product)
- Stop reaction when fluorescence between standard 1-3

# Limited Cycle Library Amplification

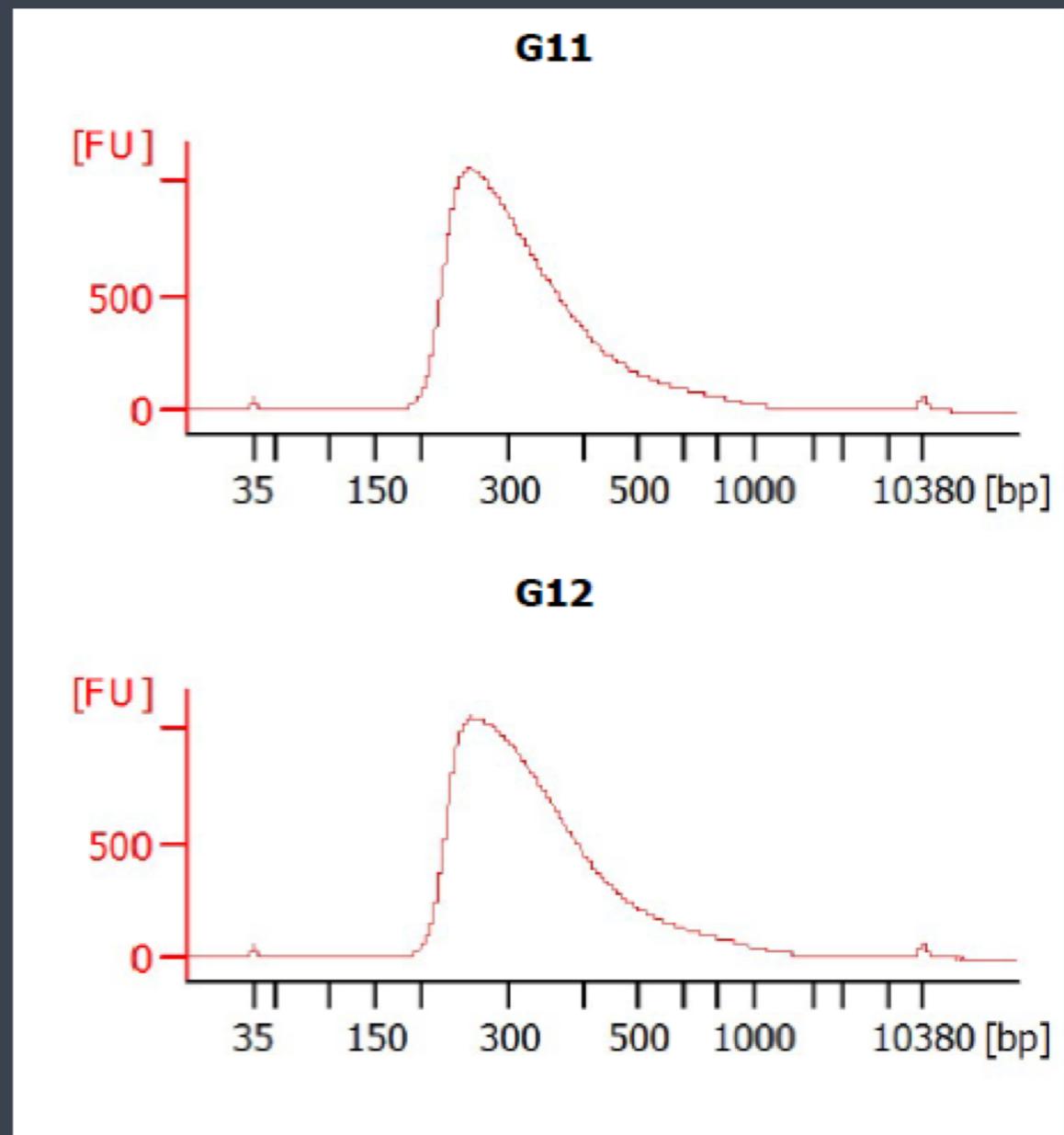


PCR  
bubble



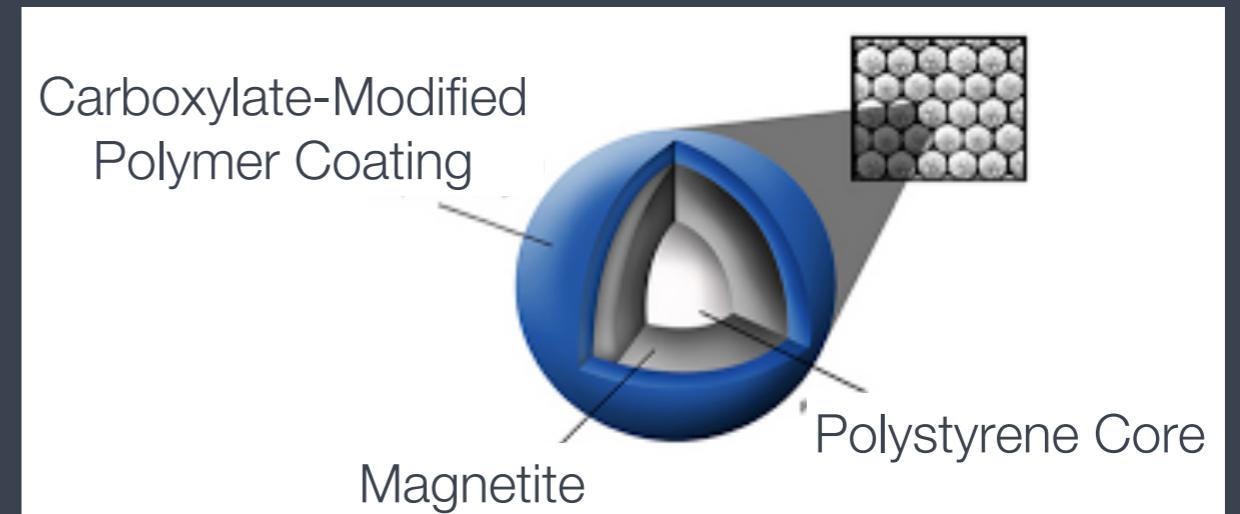
# Library QC

- Quantification
  - Dye based
    - SYBR Green
    - Qubit / Quant-IT
- Size & Quality
  - Agilent Bioanalyzer or Tapestation
  - Size determination
  - Do not use for final quantitation

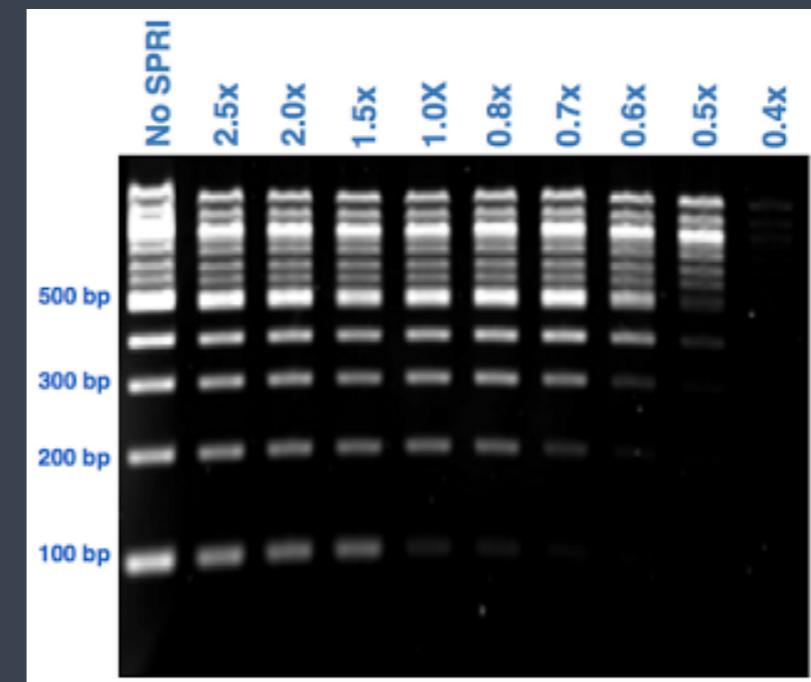


# Size selection with SPRI beads

- Solid Phase Reverse Immobilization beads



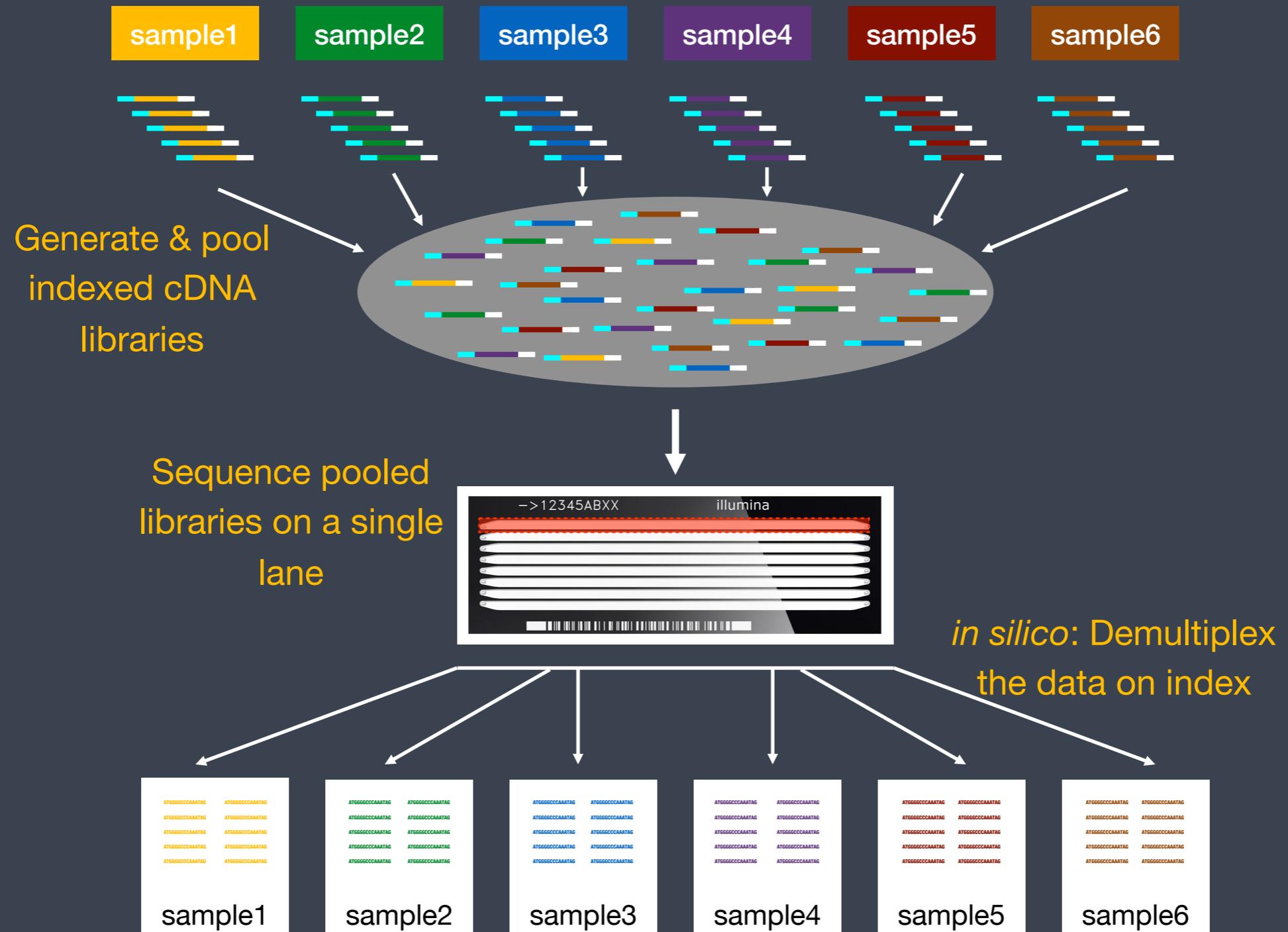
- Carboxyl groups on the bead surface bind DNA in the presence of crowding agents (PEG & NaCl).

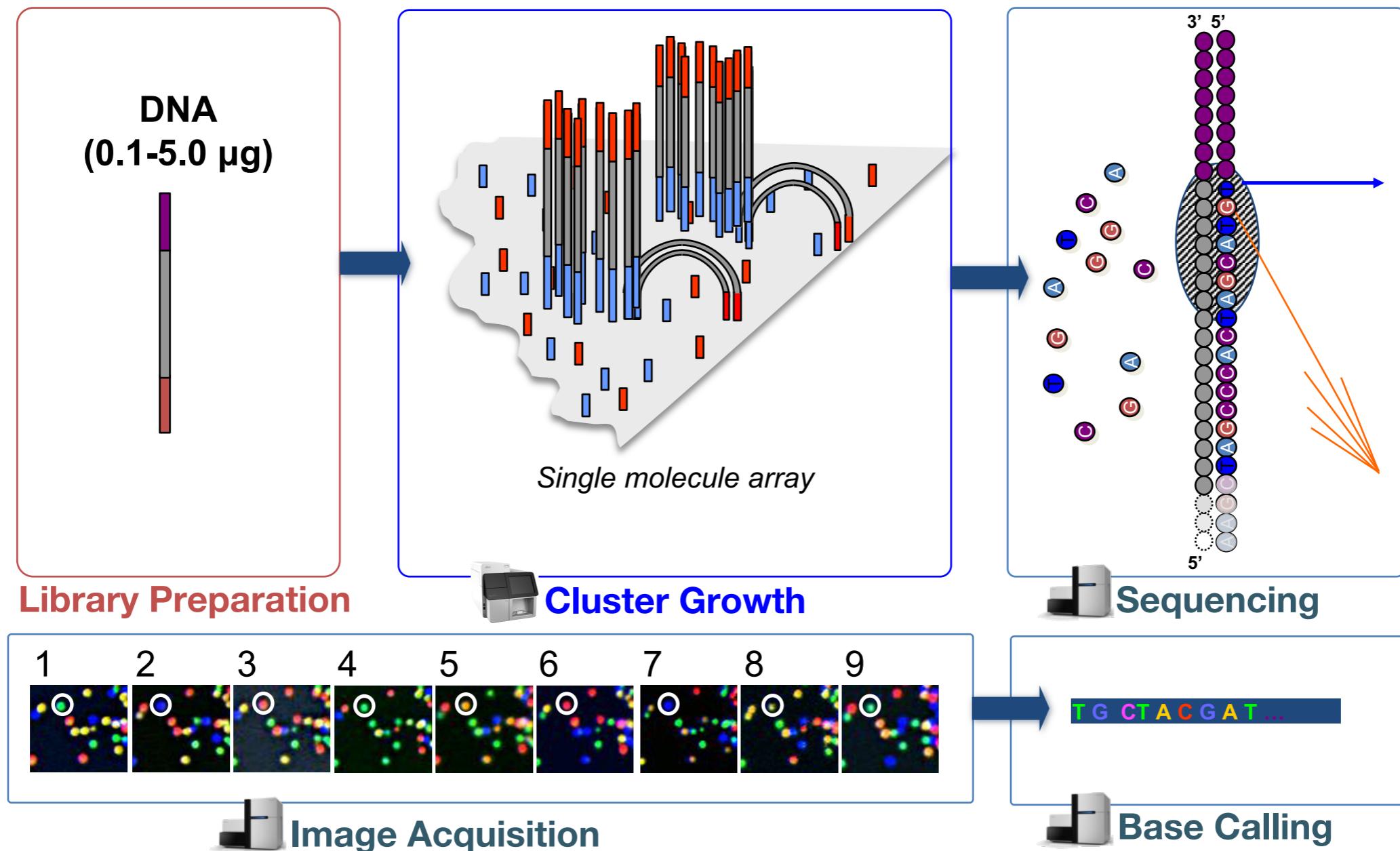


# RNA-seq Library Preparation

- ✓ Starting material
- ✓ Library preparation
  - Multiplexing
    - Sequencing read order/terminology
  - Low input / single cell RNA-seq

# Multiplexing (barcodes and indices)

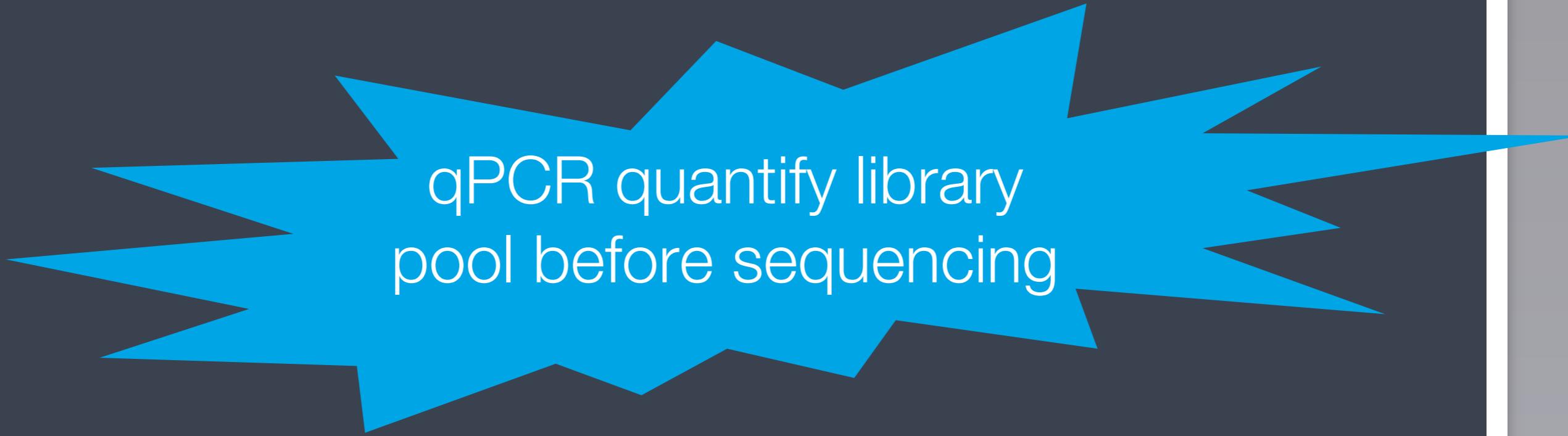




# Consider Cluster Size in Multiplexing

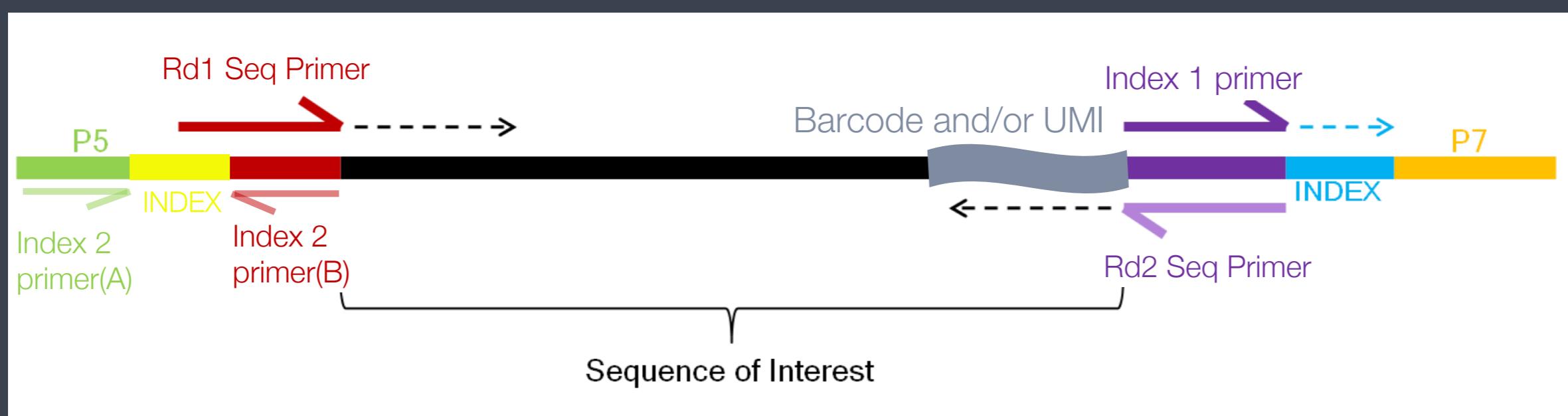
# Multiplexing

- Pool samples based on dye based quantitation.
- Submit pool to core facility for sequencing.



qPCR quantify library  
pool before sequencing

# Sequencing Read Order



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

HiSeq/MiSeq (4 color)

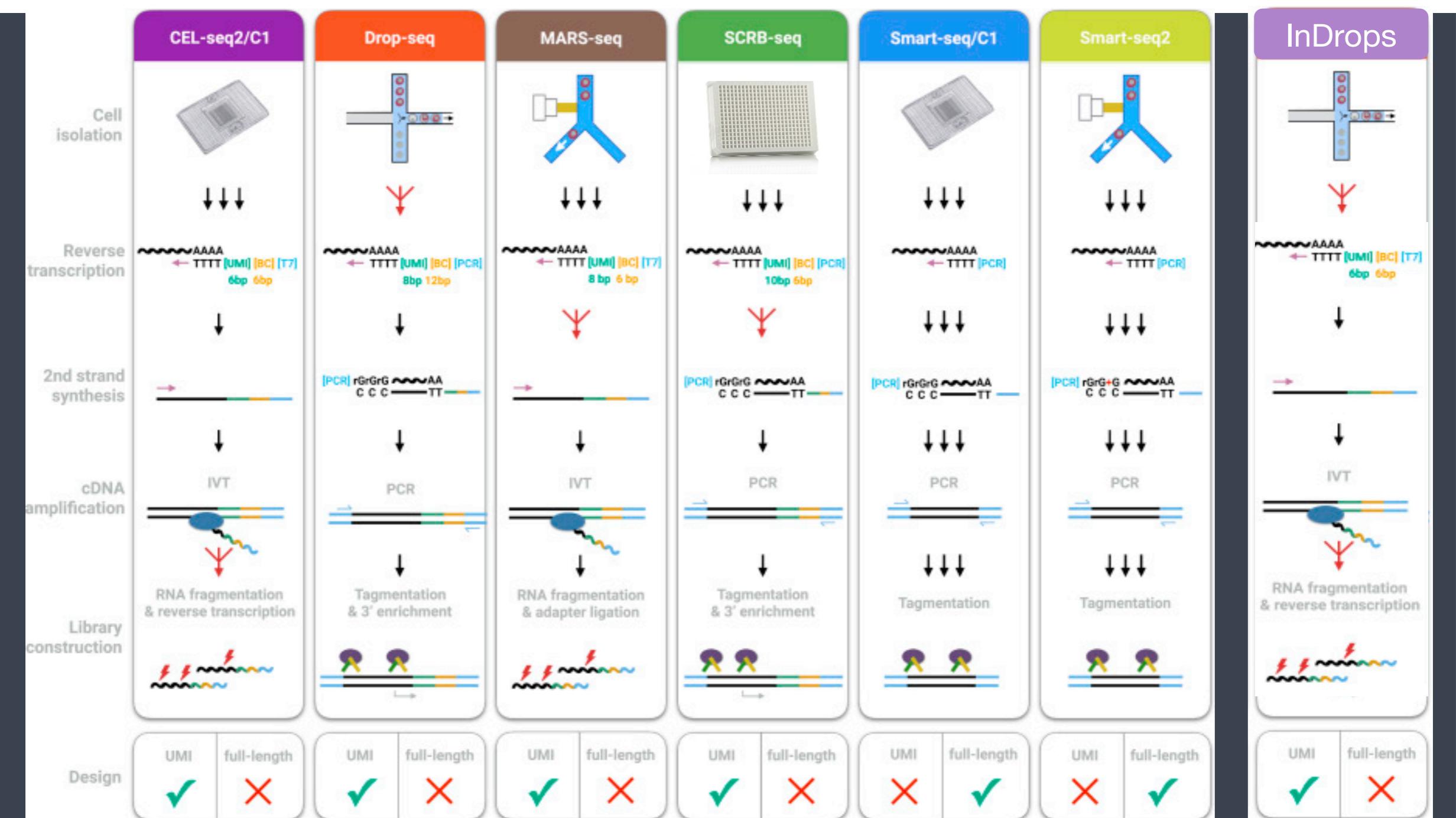
- A&C read on one camera
- G&T read on other

NextSeq/NovaSeq (2 color)

# RNA-seq Library Preparation

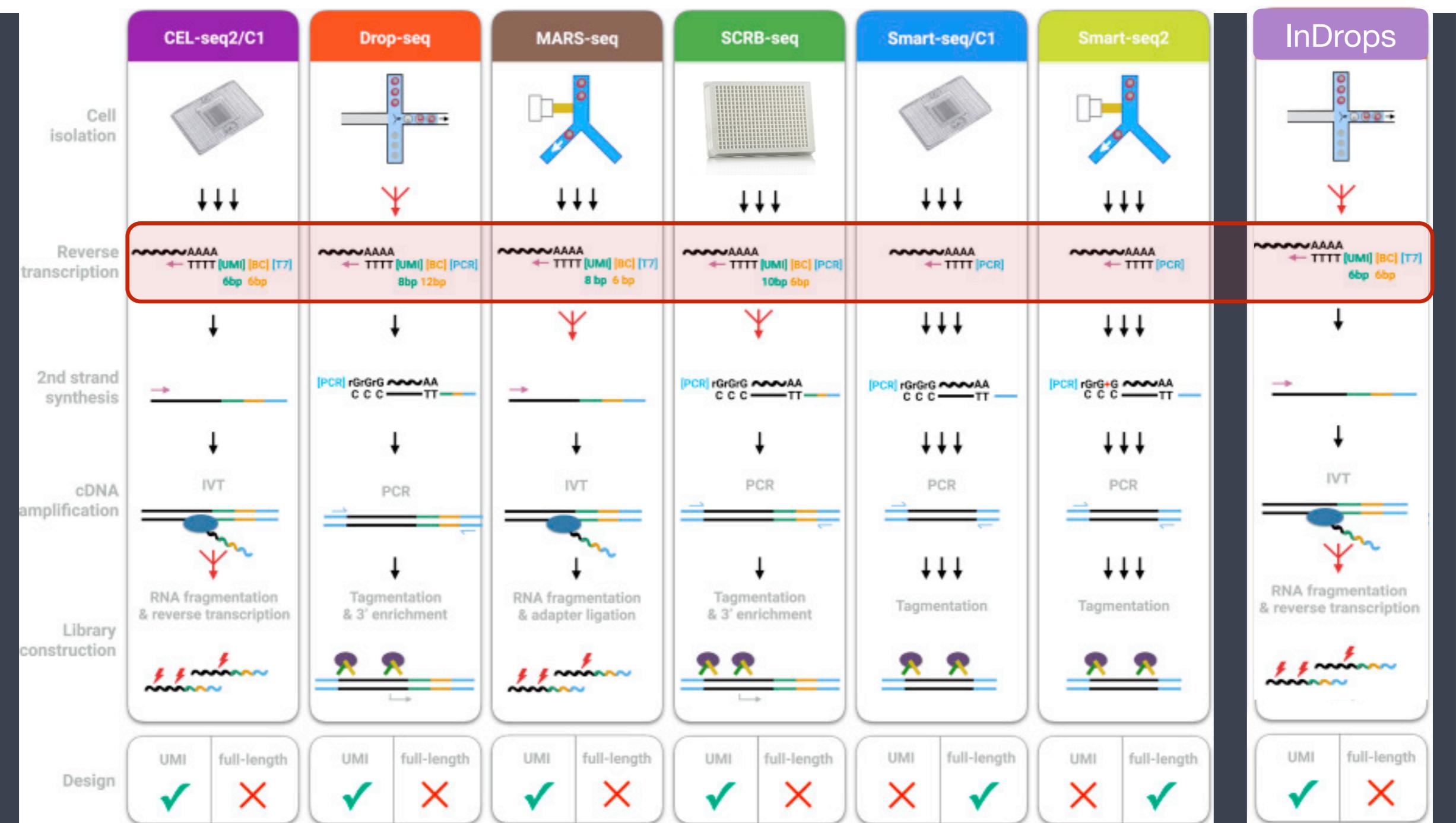
- ✓ Starting Material
- ✓ Library amplification bias
- ✓ Multiplexing
- Low input / single cell RNA-seq

# Single Cell / Low Input Methods



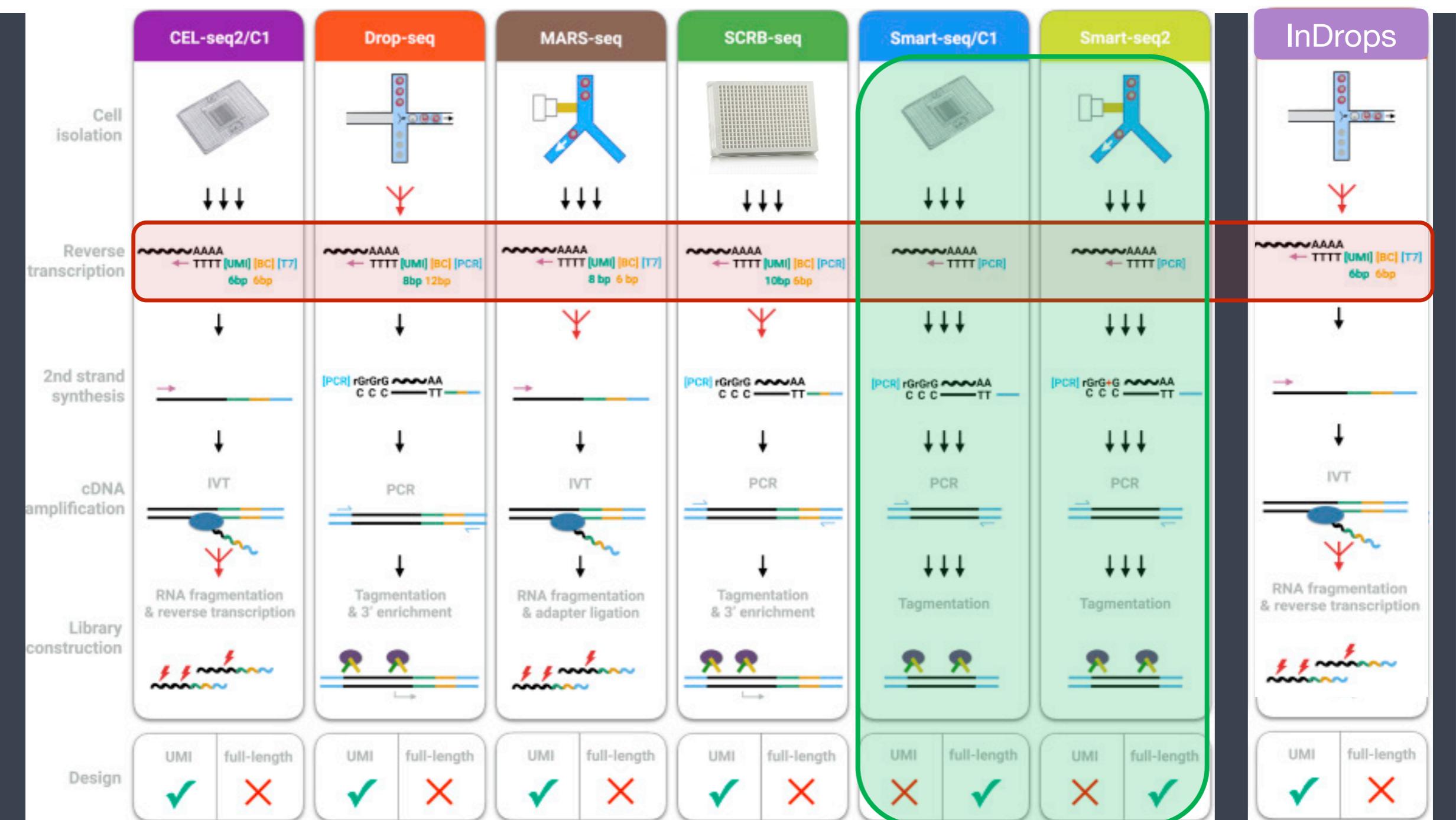
Modified from :C. Ziegenhain et al., Comparative Analysis of Single-Cell RNA Sequencing Methods, Molecular Cell 2017 (doi: 10.1016/j.molcel.2017.01.023)

# Single Cell / Low Input Methods



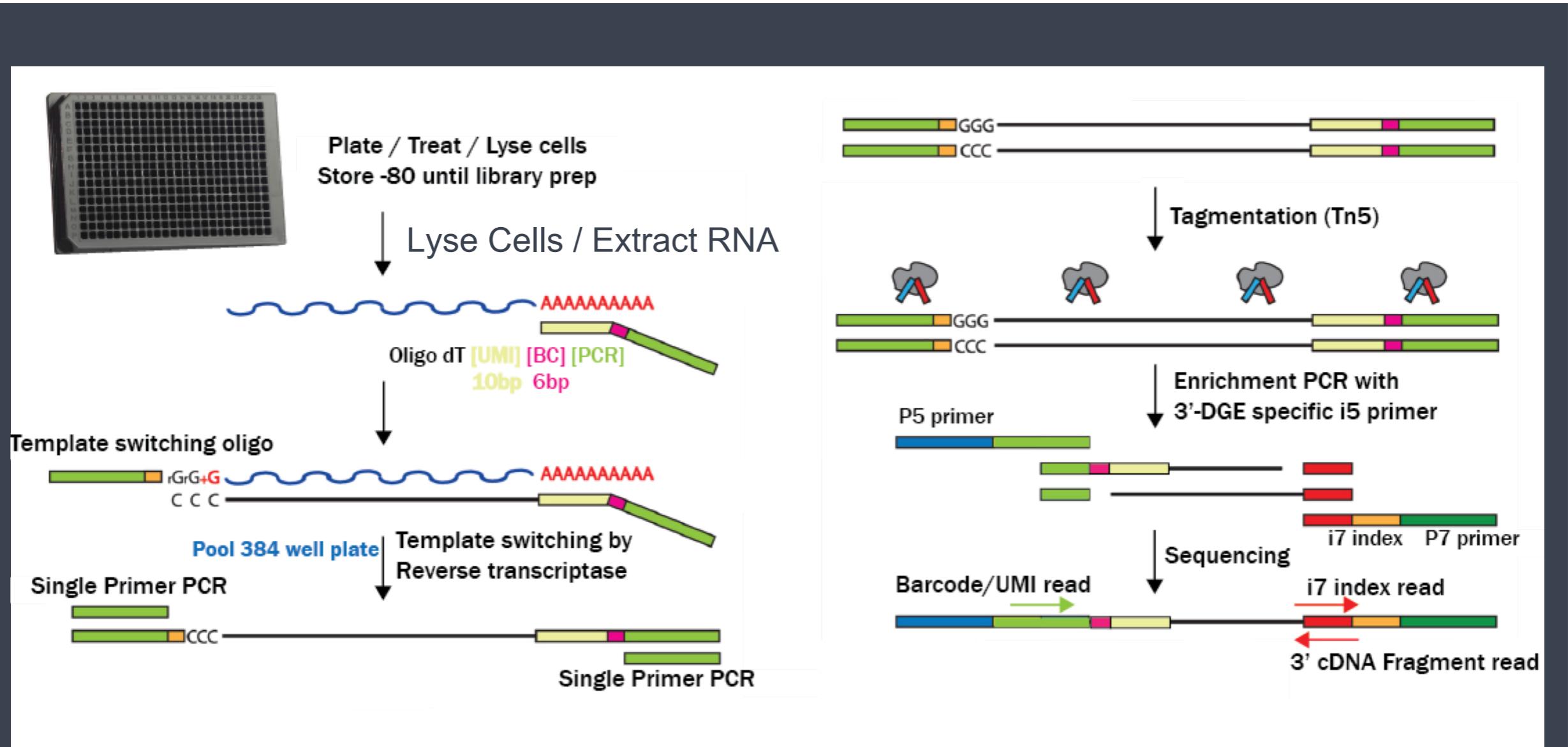
Modified from :C. Ziegenhain et al., Comparative Analysis of Single-Cell RNA Sequencing Methods, Molecular Cell 2017 (doi: 10.1016/j.molcel.2017.01.023)

# Single Cell / Low Input Methods

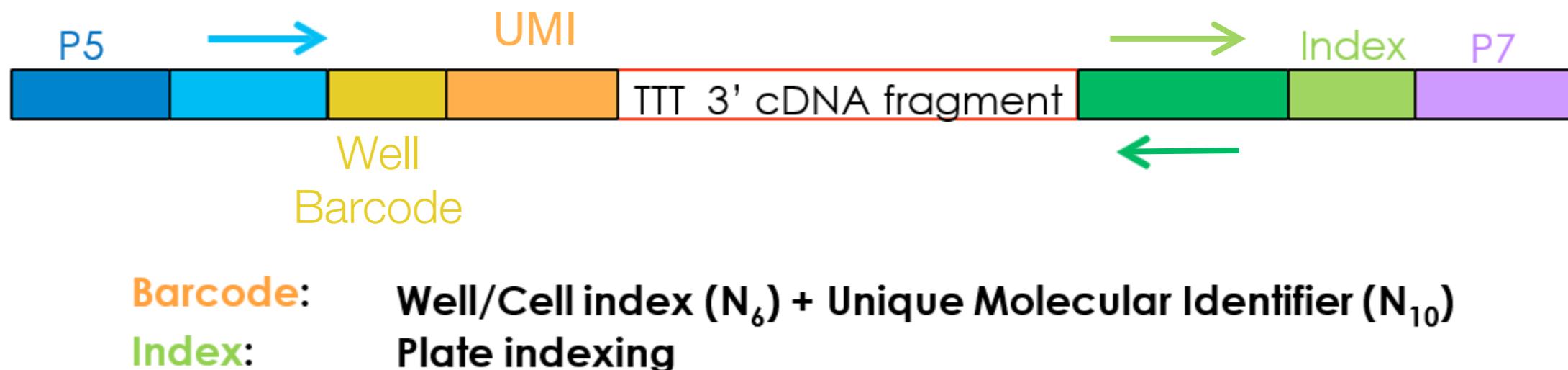


Modified from :C. Ziegenhain et al., Comparative Analysis of Single-Cell RNA Sequencing Methods, Molecular Cell 2017 (doi: 10.1016/j.molcel.2017.01.023)

# SCRB-seq: 3'-bias SMART-seq with UMIs



# Final SCRB-seq Library



- 17 cycles on **Read 1**
- 8 cycles on **Index 1**
- 46 cycles on **Read 2**

Well Barcode & UMI  
Plate Index for  
multiple plates  
cDNA fragment

# Unique Molecular Index

UMI  
Unique N-mer per transcript

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



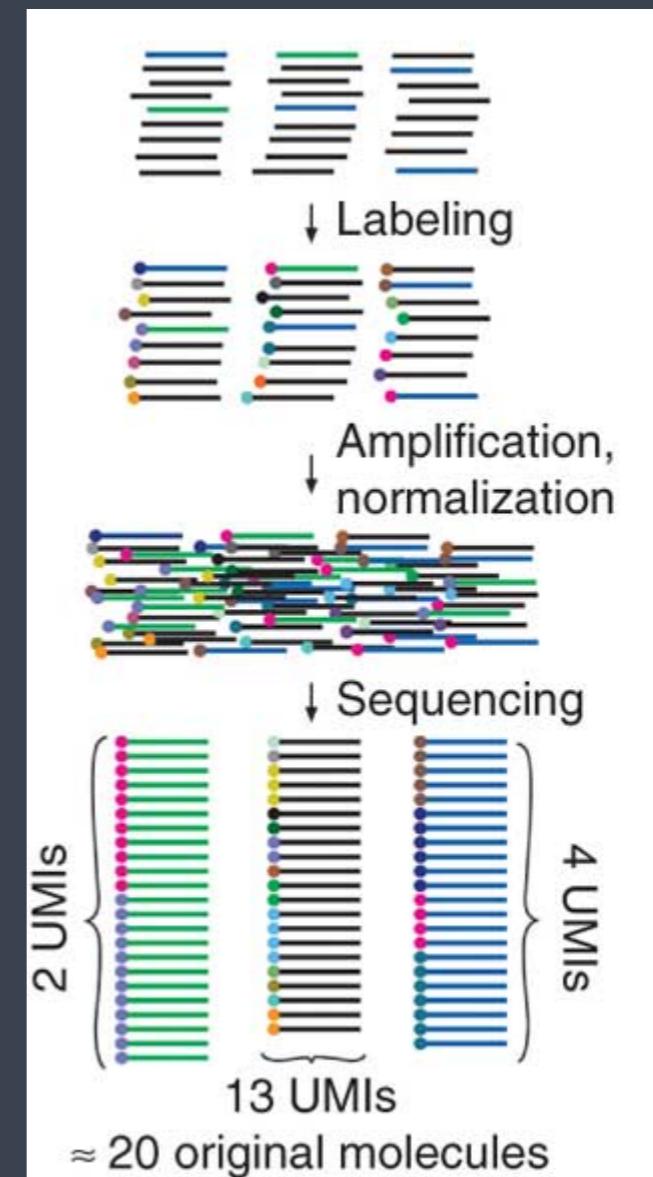
AAAAAAA

Cell or Well Barcode

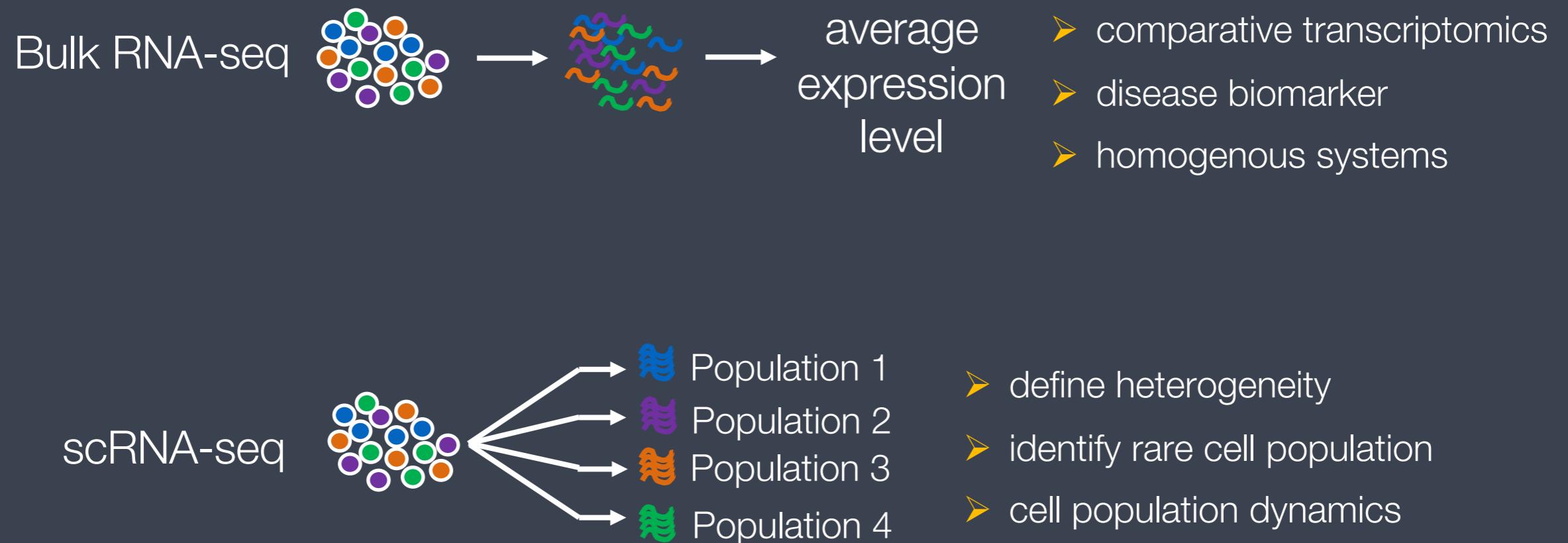
- Plate based assay need 96 or 384 well plate of barcoded/UMI-tagged primers for RT.
- Droplet based methods have a similar pool of primers in each drop.

# Library Amplification Bias

- Most low input preps have several rounds of amplification
- Introduces library bias
  - Some products preferentially amplified
- Unique Molecular Index (UMI)
  - Accurately quantitates samples



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



# Transcriptome Coverage (mRNA)

## 1. mRNA: TruSeq (Gold Standard)

- ~20,000 transcripts
  - More when consider splice variants / isoforms
- Observe 80-95% of transcripts depending on sequencing depth

# Transcriptome Coverage (mRNA)

## 1. mRNA: TruSeq (Gold Standard)

- ~20,000 transcripts
  - More when consider splice variants / isoforms
- Observe 80-95% of transcripts depending on sequencing depth

## 2. Low input methods ~3000 cells / well

- 4000-6000 transcripts per sample
  - Limiting to transcripts observed across all samples
- Observe 20-60% of the transcriptome

# Transcriptome Coverage (mRNA)

## 1. mRNA: TruSeq

- ~20,000 transcripts / 80-95% transcriptome

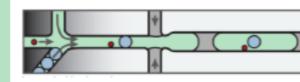
## 3. Single Cell Methods

- 200 -10,000 transcripts per cell
- Observe 10-50% of the transcriptome
- Many transcripts will show up with zero counts in every cell. (even GAPDH)

## 2. Low input methods ~3000 cells / well

- 4000-6000 transcripts / 20-60% of transcriptome

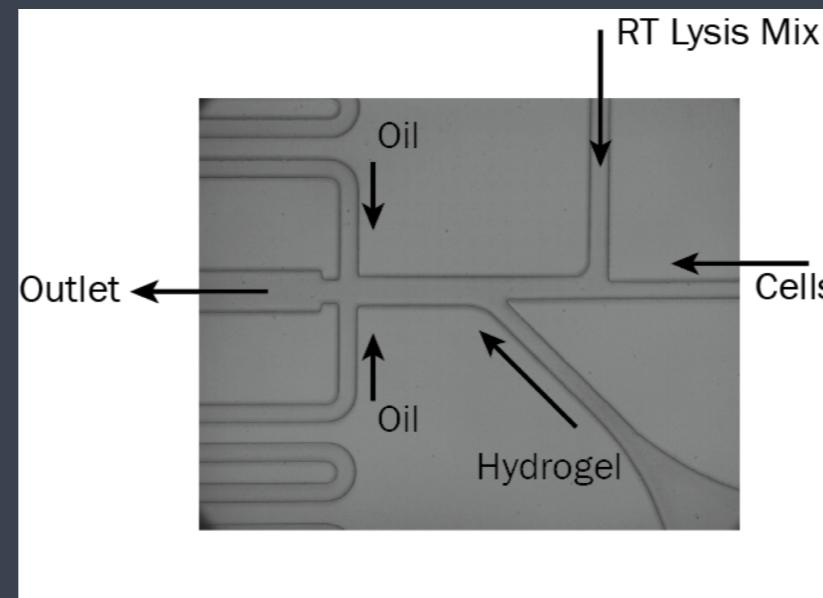
# Comparison of Single Cell Methods

	inDrops	10x	Drop-seq	Seq-well	SMART-seq
<b>Cell capture efficiency</b>	~70-80%	~50-65%	~10%	~80%	~80%
<b>Time to capture 10k cells</b>	~30min	10min	1-2 hours	5-10min	--
<b>Encapsulation type</b>	Droplet 	Droplet 	Droplet 	Nanolitre well 	Plate-based 
<b>Library prep</b>	CEL-seq Linear amplification by IVT	SMART-seq Exponential PCR based amplification	SMART-seq Exponential PCR based amplification	SMART-seq Exponential PCR based amplification	SMART-seq Exponential PCR based amplification
<b>Commercial</b>	Yes	Yes	--	--	Yes
<b>Cost (~\$ per cell)</b>	~0.06	~0.2	~0.06	--	1
<b>Strengths</b>	<ul style="list-style-type: none"> <li>Good cell capture</li> <li>Cost-effective</li> <li>Real-time monitoring</li> <li>Customizable</li> </ul>	<ul style="list-style-type: none"> <li>Good cell capture</li> <li>Fast and easy to run</li> <li>Parallel sample collection</li> <li>Best gene per cell output</li> </ul>	<ul style="list-style-type: none"> <li>Cost-effective</li> <li>Customizable</li> </ul>	<ul style="list-style-type: none"> <li>Good cell capture</li> <li>Cost-effective</li> <li>Real-time monitoring</li> <li>Customizable</li> </ul>	<ul style="list-style-type: none"> <li>Good cell capture</li> <li>Good mRNA capture</li> <li>Full-length transcript</li> <li>No UMI</li> </ul>
<b>Weaknesses</b>	Difficult to run	Expensive	Difficult to run & low cell capture efficiency	Still new!	Expensive

Modified from: C. Ziegenhain et al., Comparative Analysis of Single-Cell RNA Sequencing Methods, Molecular Cell 2017 (doi: 10.1016/j.molcel.2017.01.023)

# inDrops & 10x Single Cell Encapsulation

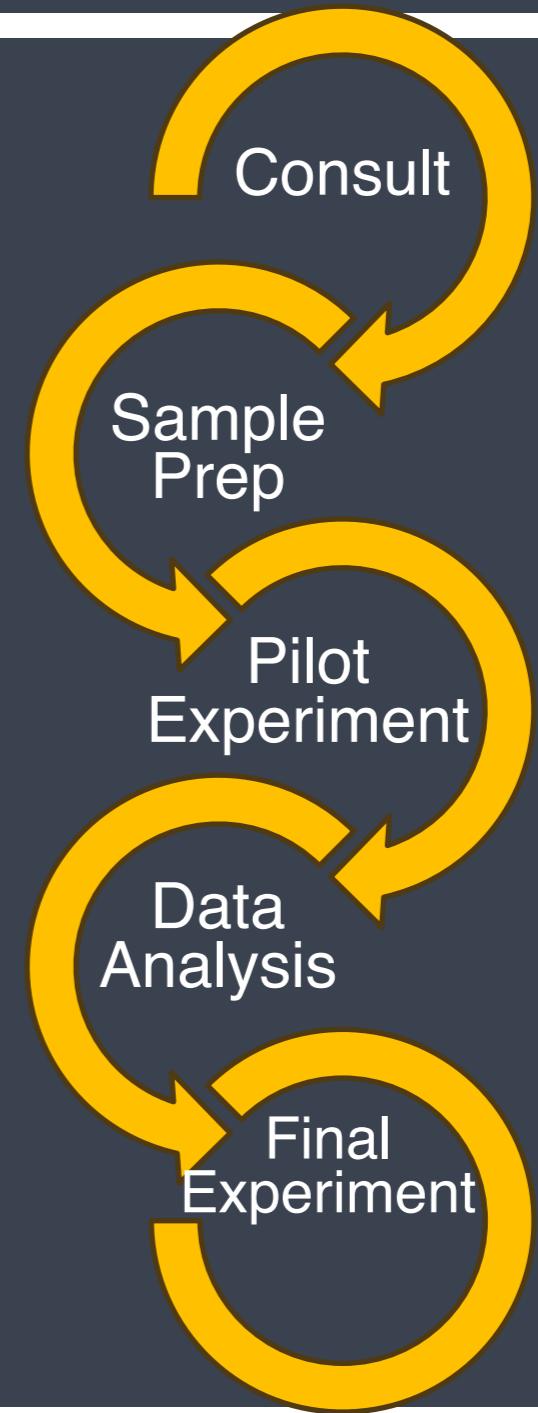
- Single cell suspension injected at density ~80,000 cells/ml



- By matching the speed of bead injection with the speed of droplet generation nearly every droplet is loaded with one hydrogel bead

# 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

# Key to Success: Reduce Batch Effect

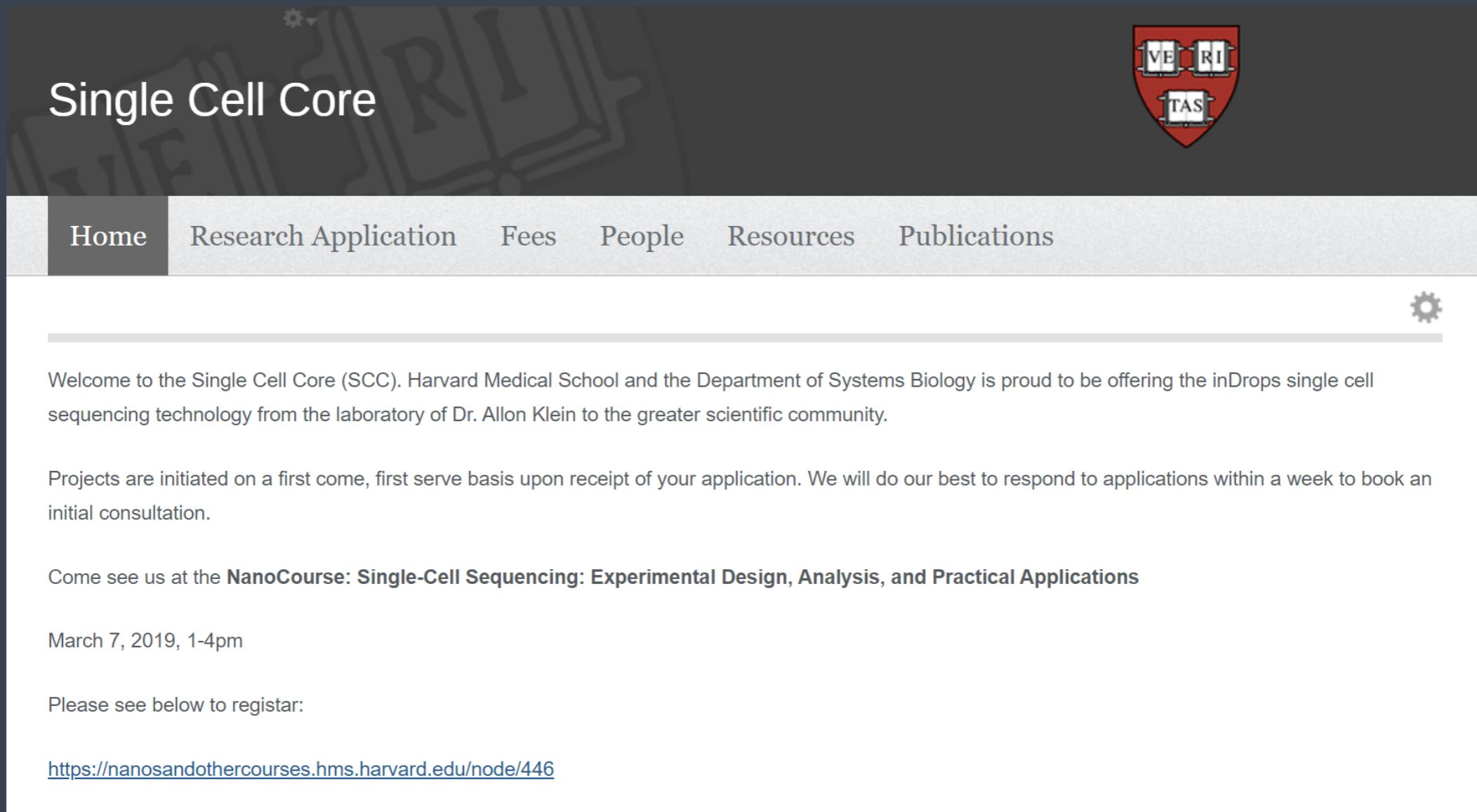
Encapsulate control and test sample on same day.

- Minimally need biological replicates on all samples.
- Use same processing method and times for all samples.

inDrops and 10x samples stored as DNA:RNA hybrids

- Accumulate encapsulated samples until study is finished before processing samples to final libraries.
- Backup samples can be used for batch effect controls.

# Key to Success: Talk to an expert



The image is a screenshot of the Single Cell Core website. The header features a dark background with a faint watermark of a DNA helix. The title "Single Cell Core" is in white on the left, and a red Harvard shield logo is on the right. A navigation bar below has tabs for Home (selected), Research Application, Fees, People, Resources, and Publications. A gear icon is in the top right corner. The main content area has a light gray background. It starts with a welcome message: "Welcome to the Single Cell Core (SCC). Harvard Medical School and the Department of Systems Biology is proud to be offering the inDrops single cell sequencing technology from the laboratory of Dr. Allon Klein to the greater scientific community." It then describes the application process: "Projects are initiated on a first come, first serve basis upon receipt of your application. We will do our best to respond to applications within a week to book an initial consultation." It also mentions a "NanoCourse" and provides a registration link: "Please see below to register:  
<https://nanosandothercourses.hms.harvard.edu/node/446>

<https://singlecellcore.hms.harvard.edu/scc-application>

# 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

# Final Thoughts

- Practice your library prep.
- Be sure you understand each step in library prep.
- Design your experiment to reduce batch effect.
  - Prep all libraries in one batch.
- Talk to someone who has done the protocol before starting.

## qPCR

Precise quantitation is key to effective sequencing!