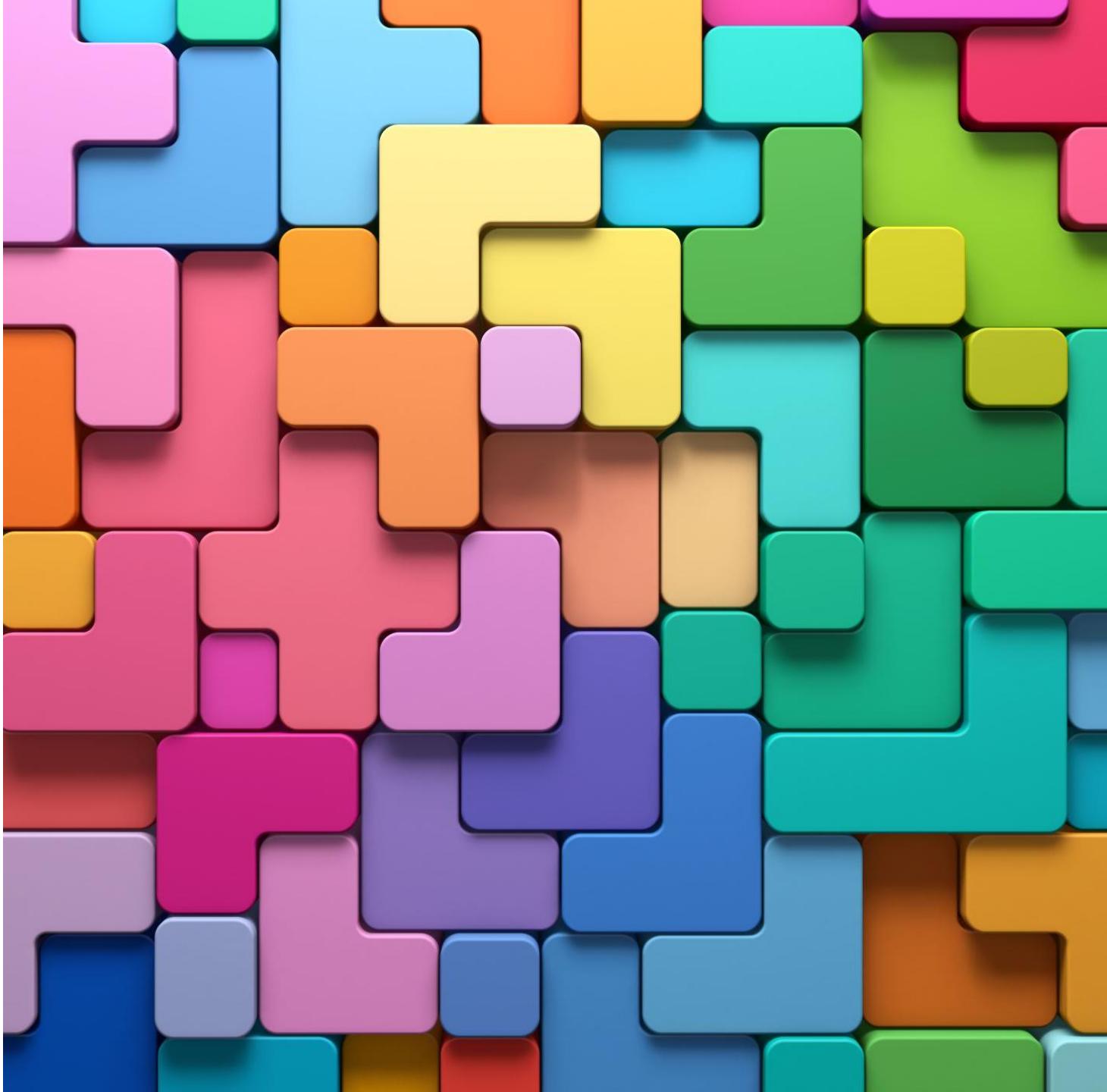


Bulk and Single cell RNA sequencing

Dr. Aatish Thennavan MDS, PhD



Lecture Overview

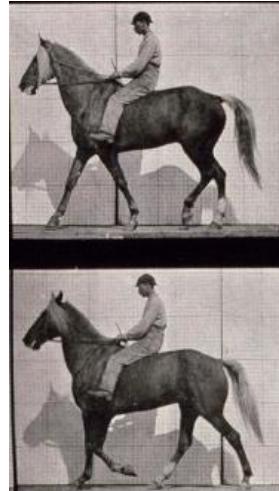
- Introduction – Why RNA sequencing & flowchart of steps
- Bulk RNA sequencing experiment
 - Principles and fundamental concepts
 - Experimental considerations and design
 - Example methodology from a publication
- Single cell sequencing experiment
 - Principles and fundamental concepts
 - Experimental considerations and design
 - Example methodology from a publication
- Next Generation Sequencing (NGS)
 - Principles and fundamental concept
 - Illumina sequencing
 - Sequencing parameters for bulk and single cell sequencing



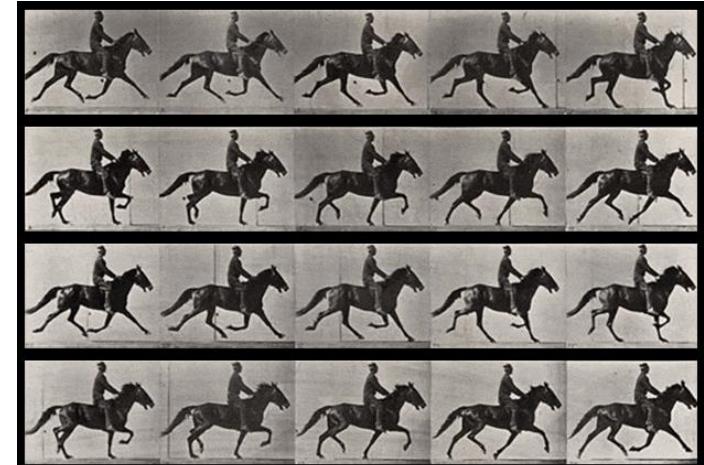
Introduction – Why RNA sequencing

- The cells in our bodies become structurally and functionally diverse by activating different combinations of genes.
- By studying the RNA that is transcribed from these genes, we can find out which genes are active in a particular cell type.
- Measuring DNA gives us the static painting, but RNA measurements gives us the dynamic motion picture

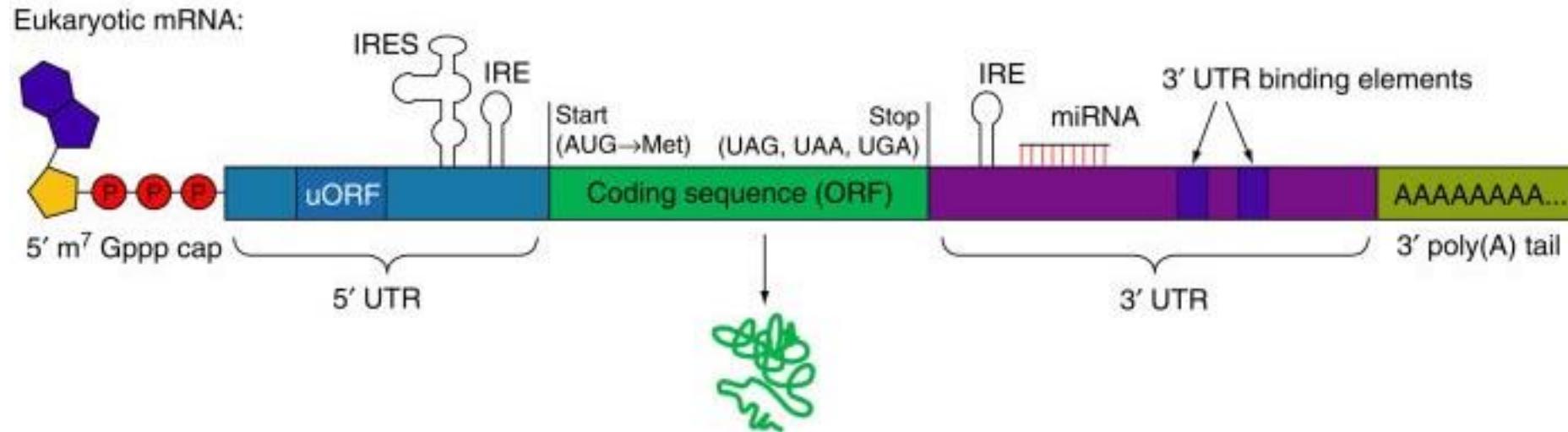
DNA



RNA

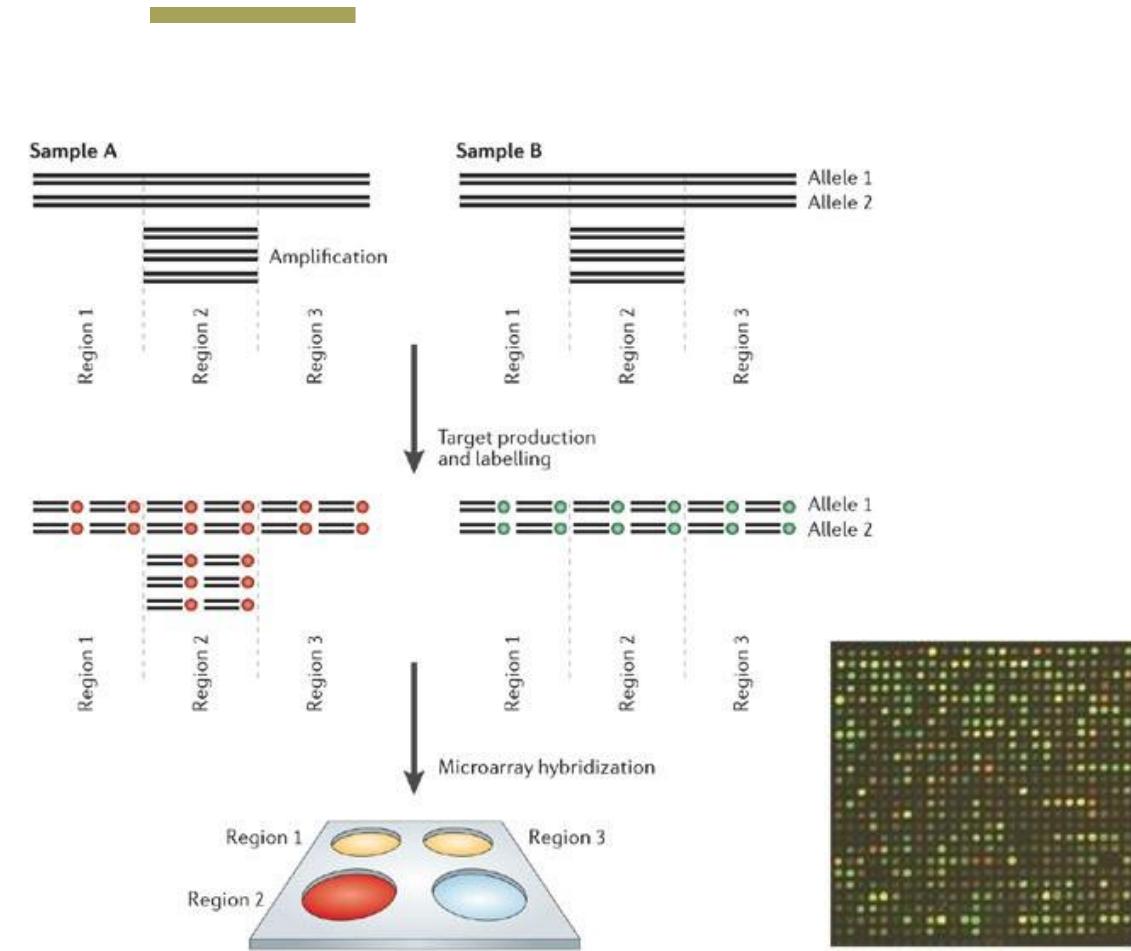


Introduction – What are we sequencing?

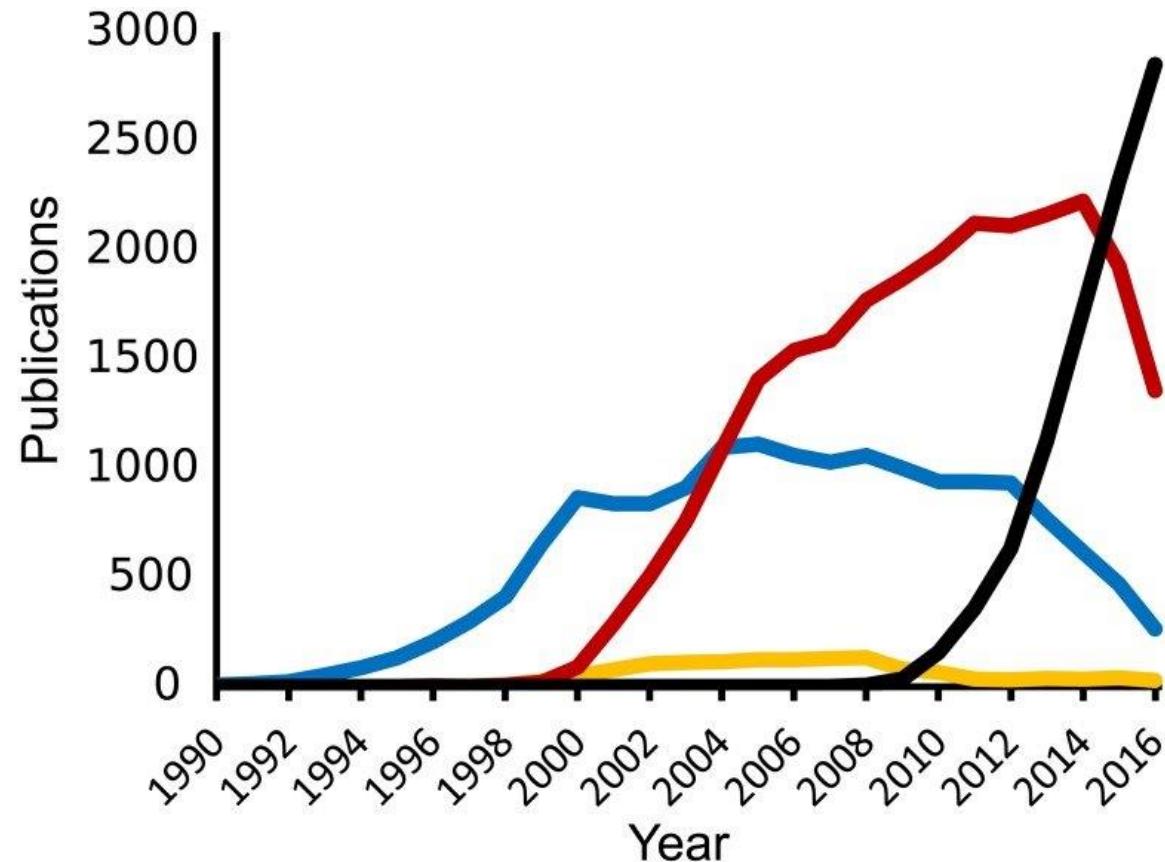


- Poly A tail - regulation of mRNA translation, stability, and export
- Three prime untranslated region (3' UTR) - contains the regulatory regions that post-transcriptionally influence gene expression. Promote proteins and microRNA association with mRNA
- Coding region - codes for proteins
- 5' untranslated region (5' UTR) & 5' Cap

Introduction – What came before sequencing?



Copyright © 2006 Nature Publishing Group
Nature Reviews | Genetics



Red: Microarray
Black: RNA-Sequencing

Introduction – Flowchart of every RNA sequencing

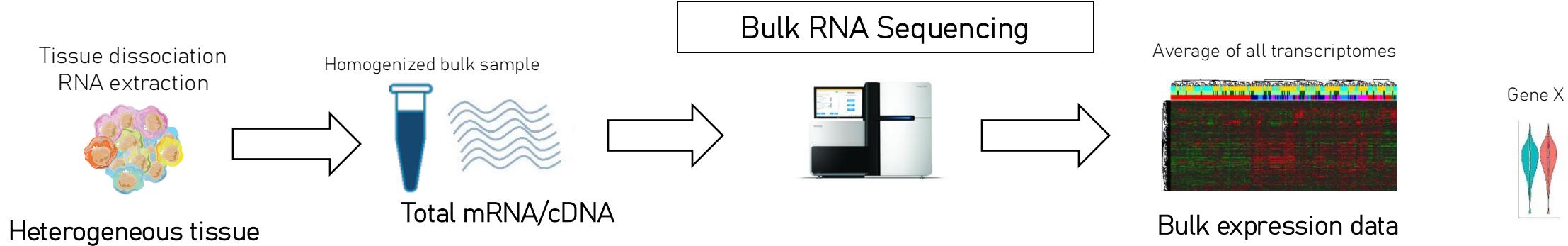
Experimental

1. mRNA isolation/extraction techniques (cell isolation)
2. Quality Check/Quantity
3. Reverse transcription into cDNA
4. Adapted Ligation
5. Amplification
6. Sequencing

Analysis

1. Alignment to a reference genome
2. Quantify transcripts
3. Quality Control
4. Normalization
5. Dimension Reduction
6. Specific analysis –clustering, differential gene expression (DE) analysis etc.

Bulk RNA sequencing - Applications



Coding

mRNA

- Differential expression
- large-scale time-series RNA-seq
- Isoform expression
- Allele specific expression
- Alternative splicing events
- Gene fusion
- Co-expression network analysis
- Meta-analysis (Multi-experimental data)

Evolving applications

- Copy number alteration
- Indel detection
- Gene fusion detection
- Neoantigens prediction
- Transposable elements expression
- Detection of microbial contamination
- Metatranscriptomics
- Cell-type deconvolution
- Variant Analysis
- TWAS and eQTL
- Others analysis
 - Meta-analysis
 - Co-expression analysis

Bulk RNA sequencing – Points to consider before the experiment

- What is the main purpose of my RNA sequencing?
 - DE gene analysis – We need at least 2 conditions to compare; mRNA – 3' end
 - Neoantigen prediction – We need paired DNA data, correct reference genome; mRNA – 5' end
- How many replicates do I need? (Rule of 3)
- What samples will I isolate RNA from – fresh, frozen or fixed tissue/cells
- Testing RNA integrity – RIN & DV200 metrics
- Try to perform library generation once for all samples (Eliminating Batch effect)
- What sequencing parameter do we need – Length of the read, sequencing depth, sequencing platform etc.

Bulk RNA sequencing – Replicates

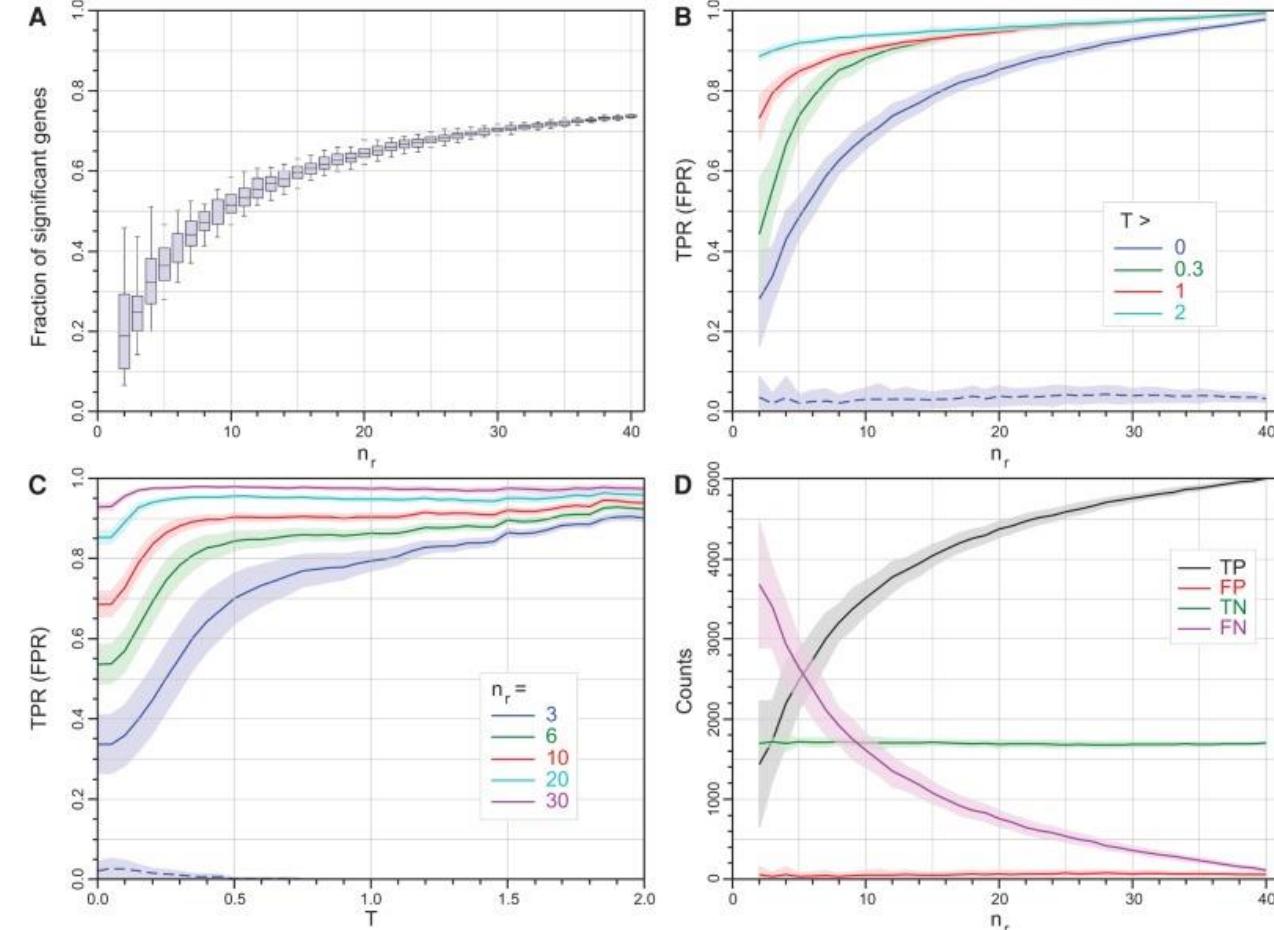


TABLE 2. A summary of the recommendations of this paper

Tool	Agreement with other tools ^a	WT vs. WT FPR ^b	Fold-change threshold (T) ^c			Tool recommended for: (# good replicates per condition) ^d		
			≤3	≤12	>12	≤3	≤12	>12
<i>DESeq</i>	Consistent	Pass	0	-	-	Yes	Yes	Yes
			0.5	-	-	Yes	Yes	Yes
			2.0	Yes	Yes	Yes	Yes	Yes
<i>DESeq2</i>	Consistent	Pass	0	-	-	Yes	Yes	Yes
			0.5	Yes	Yes	Yes	Yes	Yes
			2.0	Yes	Yes	Yes	Yes	Yes
<i>EBSeq</i>	Consistent	Pass	0	-	-	Yes	Yes	Yes
			0.5	-	-	Yes	Yes	Yes
			2.0	Yes	Yes	Yes	Yes	Yes
<i>edgeR (exact)</i>	Consistent	Pass	0	-	-	Yes	Yes	Yes
			0.5	Yes	Yes	Yes	Yes	Yes
			2.0	Yes	Yes	Yes	Yes	Yes
<i>limma</i>	Consistent	Pass	0	-	-	Yes	Yes	Yes
			0.5	-	-	Yes	Yes	Yes
			2.0	Yes	Yes	Yes	Yes	Yes
<i>cuffdiff</i>	Consistent	Fail						
<i>BaySeq</i>	Inconsistent	Pass						
<i>edgeR (GLM)</i>	Inconsistent	Pass						
<i>DEGSeq</i>	Inconsistent	Fail						
<i>NOISeq</i>	Inconsistent	Fail						
<i>PoissonSeq</i>	Inconsistent	Fail						
<i>SAMSeq</i>	Inconsistent	Fail						

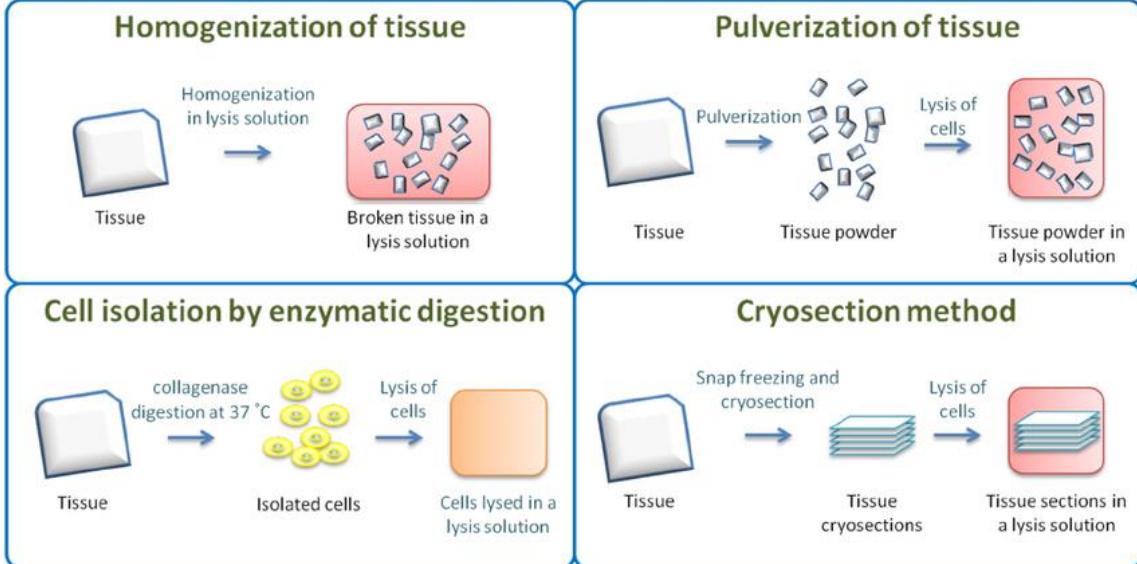
^aFull clean replicate data set, see section “Tool Consistency with High Replicate Data” and Figure 3.

^bSee section “Testing Tool False Positive Rates” and Figure 4.

^cSee section “Differential Expression Tool Performance as a Function of Replicate Number.”

^dSee Figure 2.

Bulk RNA sequencing – RNA isolation strategy - Tissue



Cells



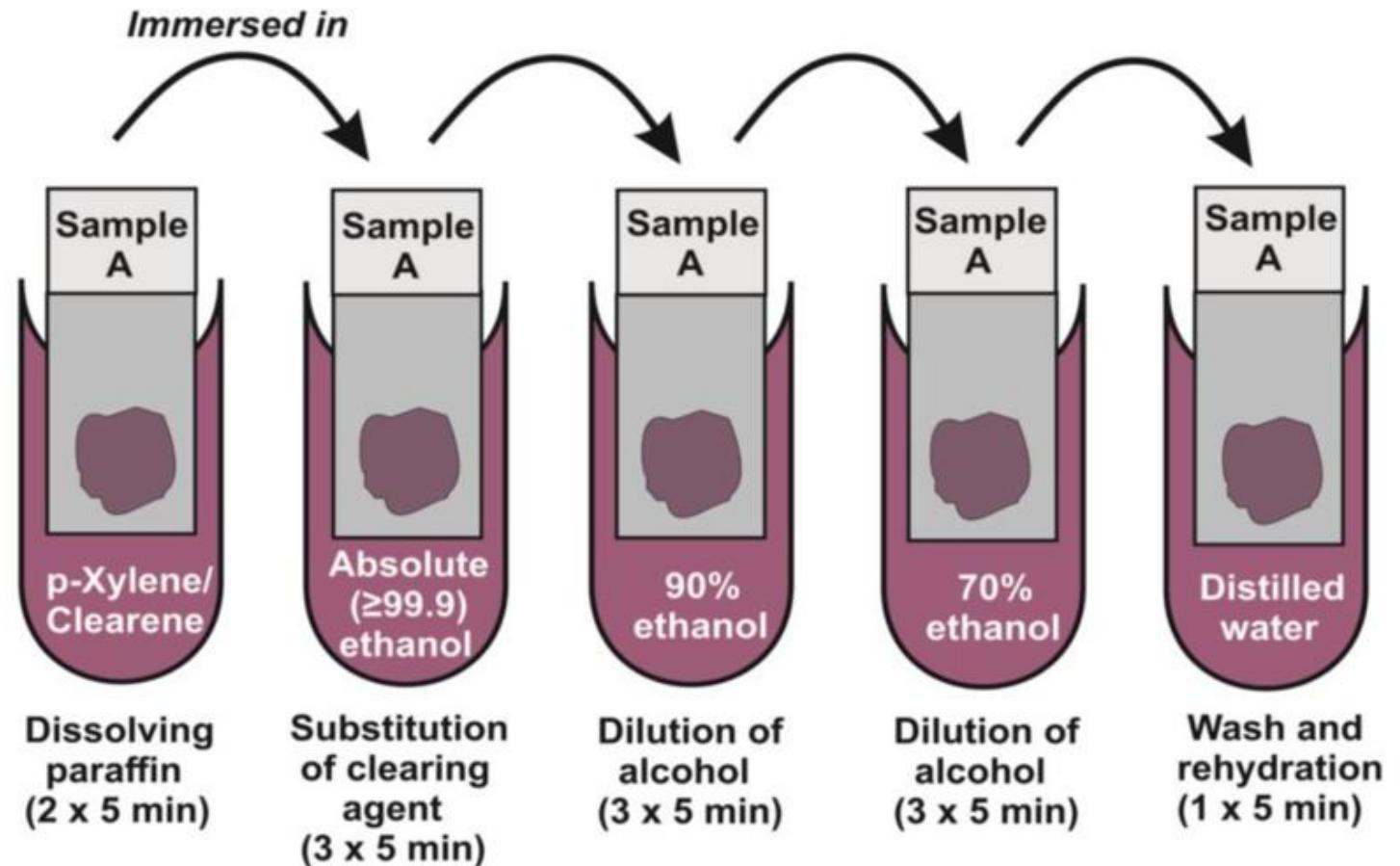
Table 1. Comparison of the advantages and disadvantages of different tissue processing methods for RNA extraction.

Homogenization/ pulverization	Enzymatic method	Cryosection
<ul style="list-style-type: none">No cell isolation procedure	<ul style="list-style-type: none">Can concentrate the cells and remove the extracellular matrix, facilitating RNA extractionCan also obtain cells for culturing	<ul style="list-style-type: none">Can prepare cryosections for histological study at the same timeMinimal time of cells at room temperature or elevated temperatureNo cell isolation procedure
<ul style="list-style-type: none">Heat may be generated, resulting in RNA degradationMay be difficult to break the tough IVD tissuesRequires a homogenizer/ pulverizer	<ul style="list-style-type: none">Relative mRNA levels may change during the cell isolationExtra time is required to isolate cells before RNA extraction	<ul style="list-style-type: none">Requires a cryostat

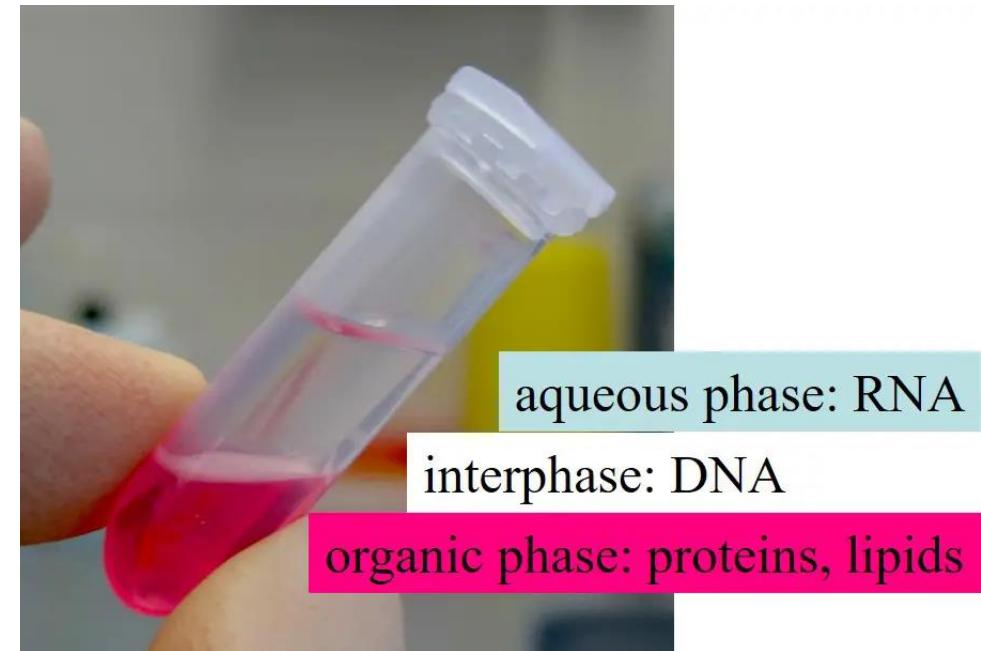
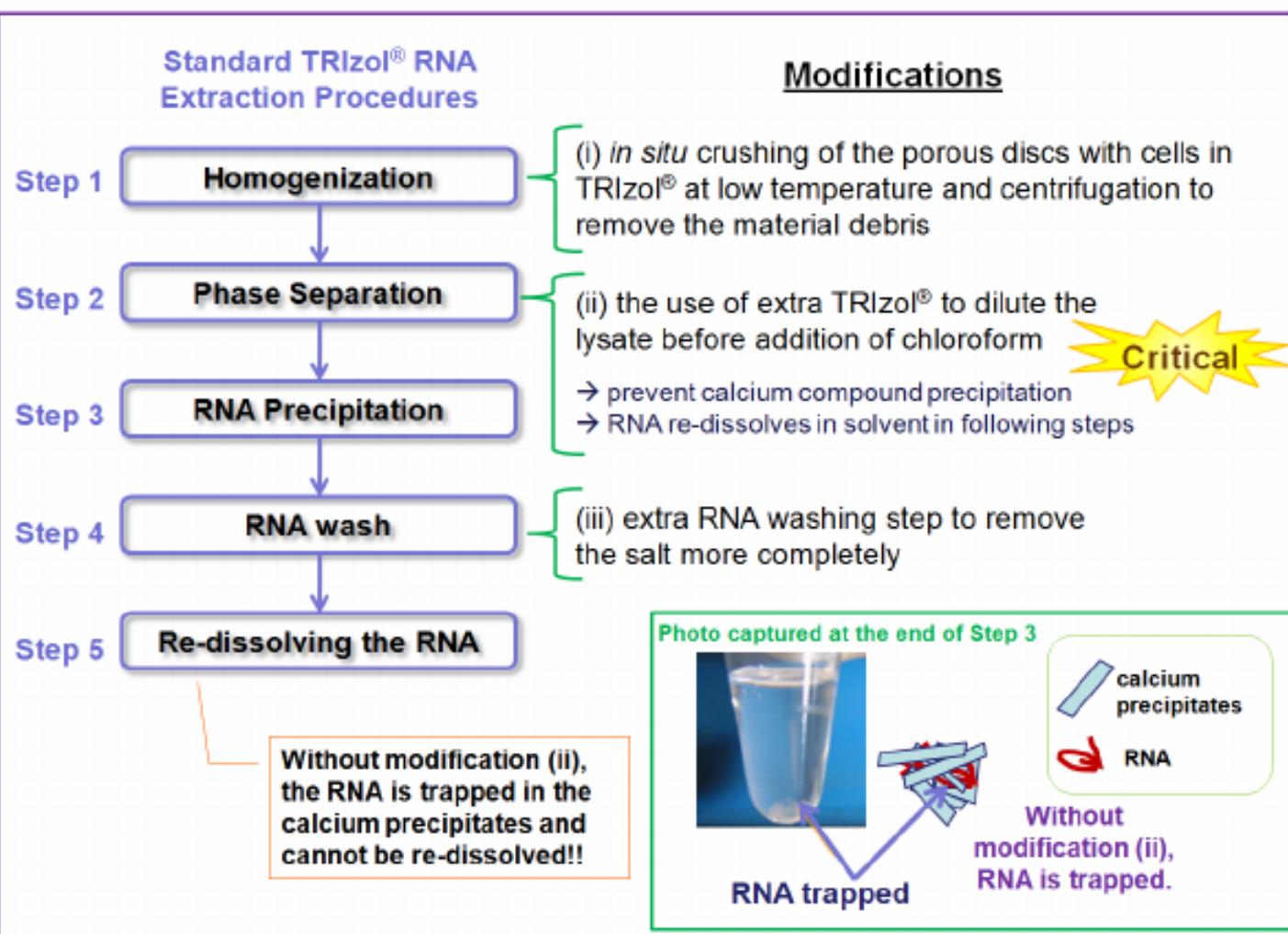
Bulk RNA sequencing – RNA isolation strategy – FFPE Tissue



Formalin fixed paraffin embedded tissue



Bulk RNA sequencing – RNA isolation strategy



Bulk RNA sequencing – RNA isolation strategy

homogenization

lyse open the cells & solubilize all its contents

RNA extraction

let those contents choose their favorite solvent & take out the aqueous phase

salts, water

RNA

DNA, proteins & lipids

phenol, chloroform

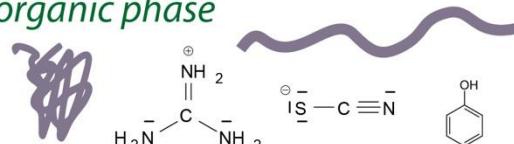
+ isopropanol
helps salts find & bind the RNA's backbone, hiding its charge

wash

remove any excess salt from the pellet

ethanol dissolves the salts but not the RNA

guanidinium thiocyanate & phenol
denature proteins so they'll move into the organic phase



at low pH, DNA will go there too

chloroform helps the phenol & water separate

RNA precipitation

get the RNA out of solution, & remove any leftover phenol

remove the liquid

air drying

evaporate lingering ethanol

resolubilization

dissolve it again - now by itself

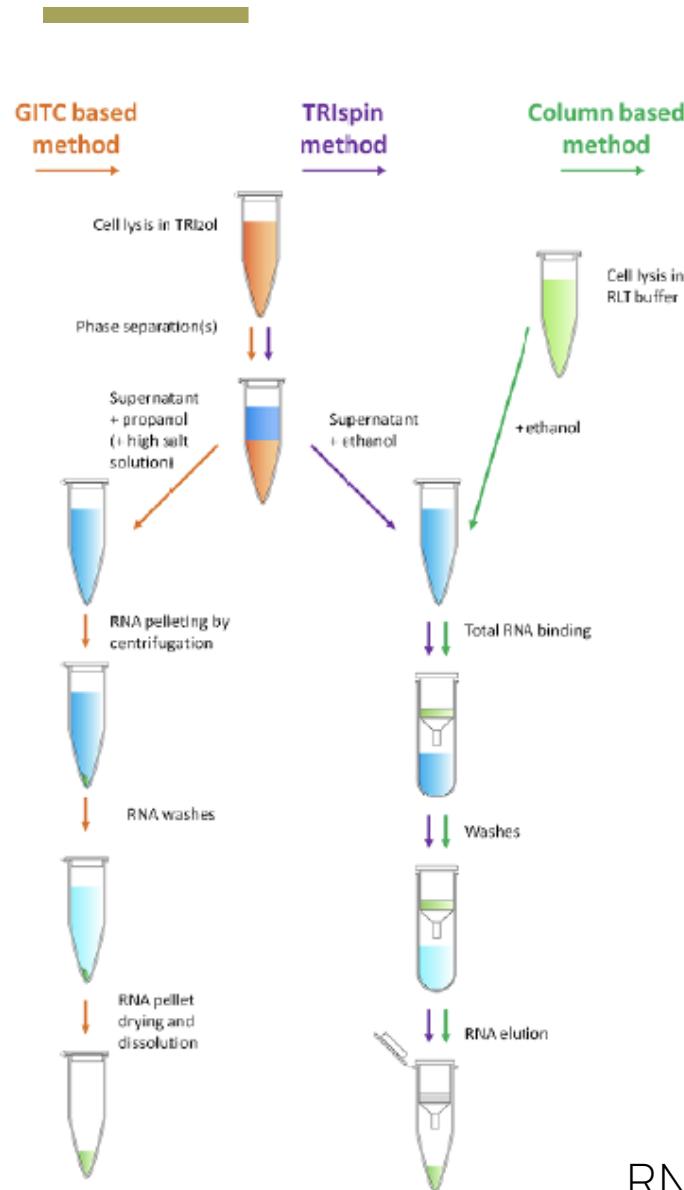


aqueous phase: RNA

interphase: DNA

organic phase: proteins, lipids

Bulk RNA sequencing – RNA isolation strategy



Frozen/Fresh



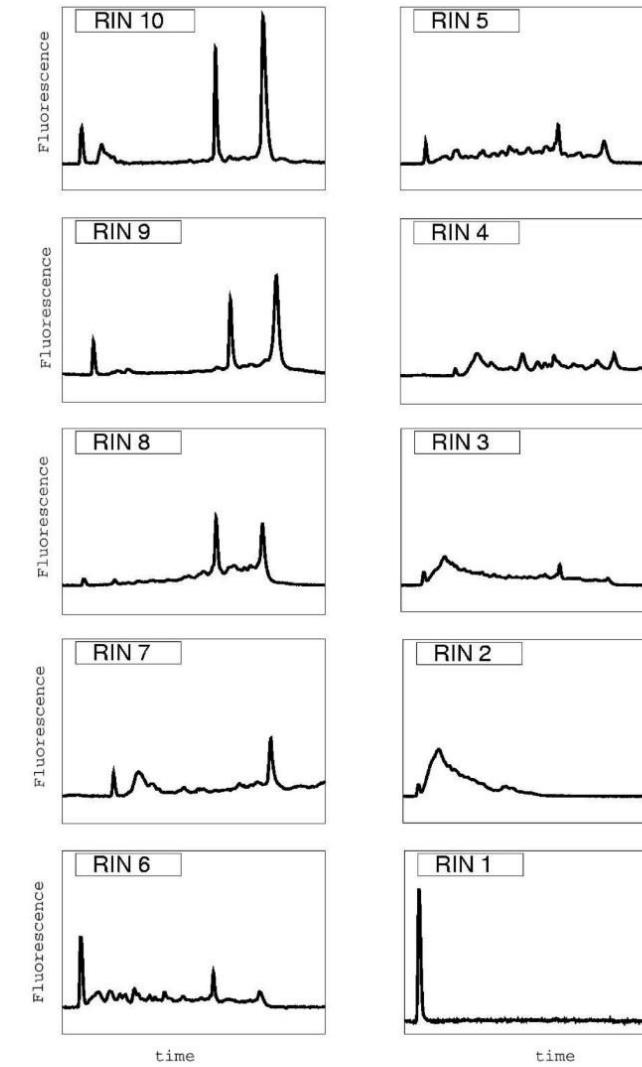
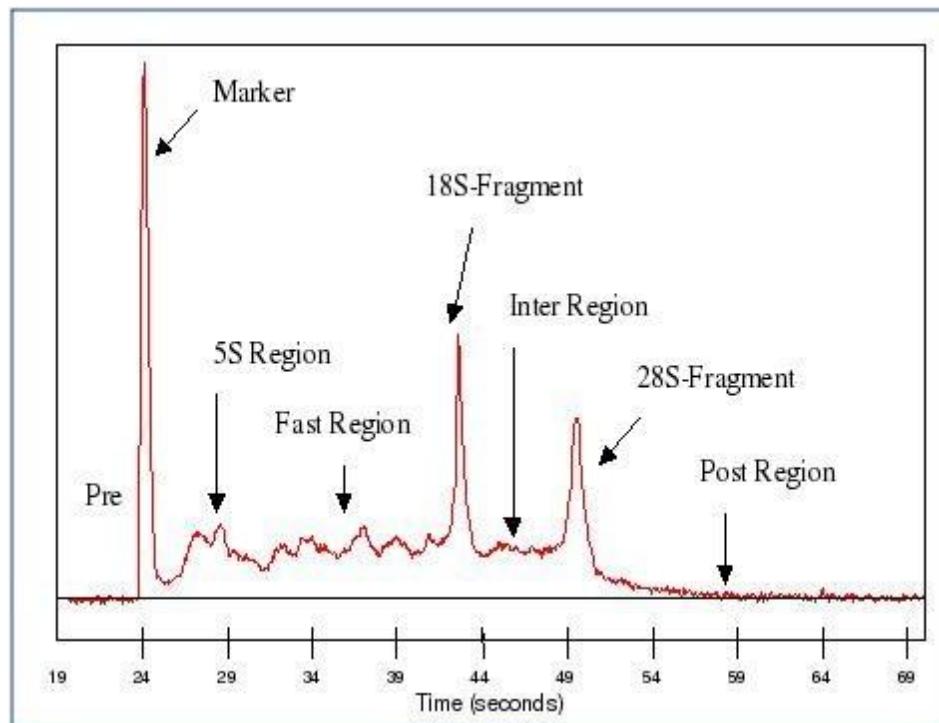
FFPE

RNA column kits for now available for almost all kinds of specimen!

Bulk RNA sequencing – RNA quality check -RIN

RNA Integrity number (RIN) is a metric offered by the Agilent Bioanalyzer as an estimate of the extent of degradation of total RNA.

It gives you a way to identify how much RNA is intact in your sample

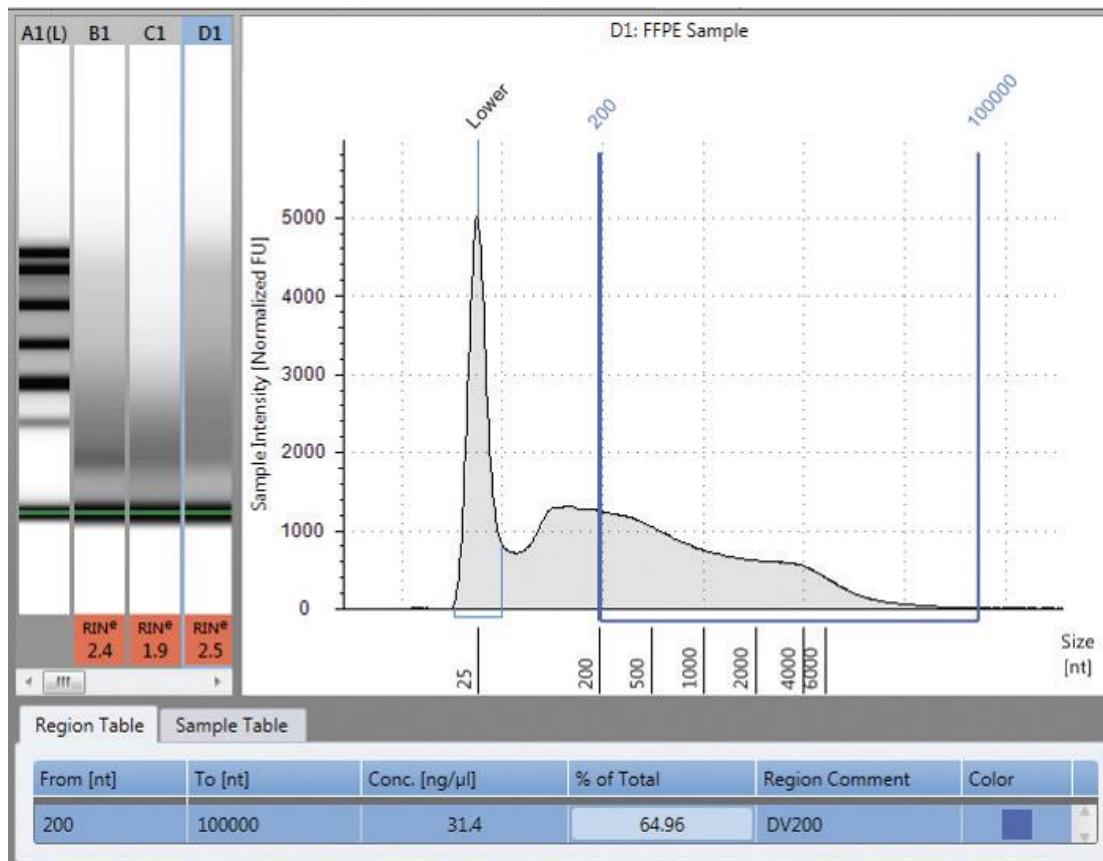


High:10-8
Med:7-3
Low:<3

FFPE
samples
tend to
have low
RIN

Bulk RNA sequencing – RNA quality check – DV200

DV200 represents the percentage of RNA fragments larger than 200 nucleotides with respect to all RNA fragments as provided by the Agilent Tape Station system

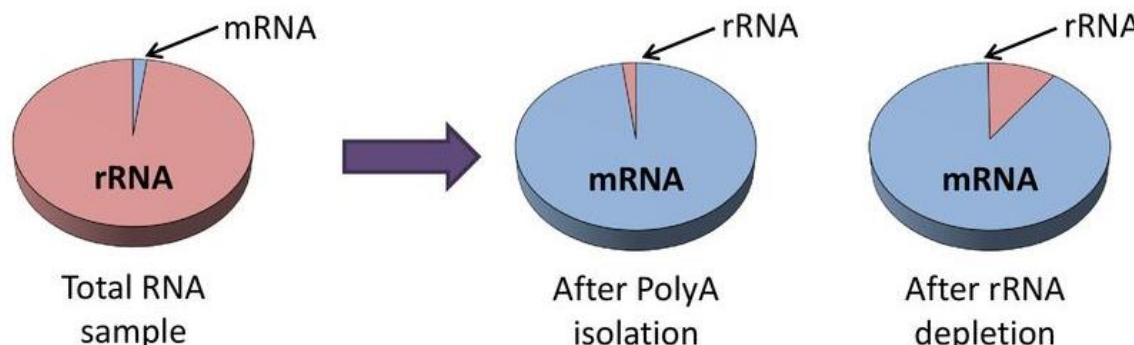


In general, a DV200 of >30% is considered as acceptable for most RNA sequencing

Also calculate quantity of RNA while calculating quality before proceeding to the next step!

Bulk RNA sequencing – How RNA quality influences library prep protocol

- PolyA Selection
 - Oligo-dT, often using magnetic beads
 - Isolates mRNA very efficiently ***unless total RNA is very dilute***
 - Can't be used to sequence non-polyA RNA
- rRNA Depletion
 - RiboZero, RiboMinus
 - Non-polyA RNAs preserved (non-coding, bacterial RNA, etc.)
 - Can be less effective at removing all rRNA



(1) Isolate polyA+ RNAs

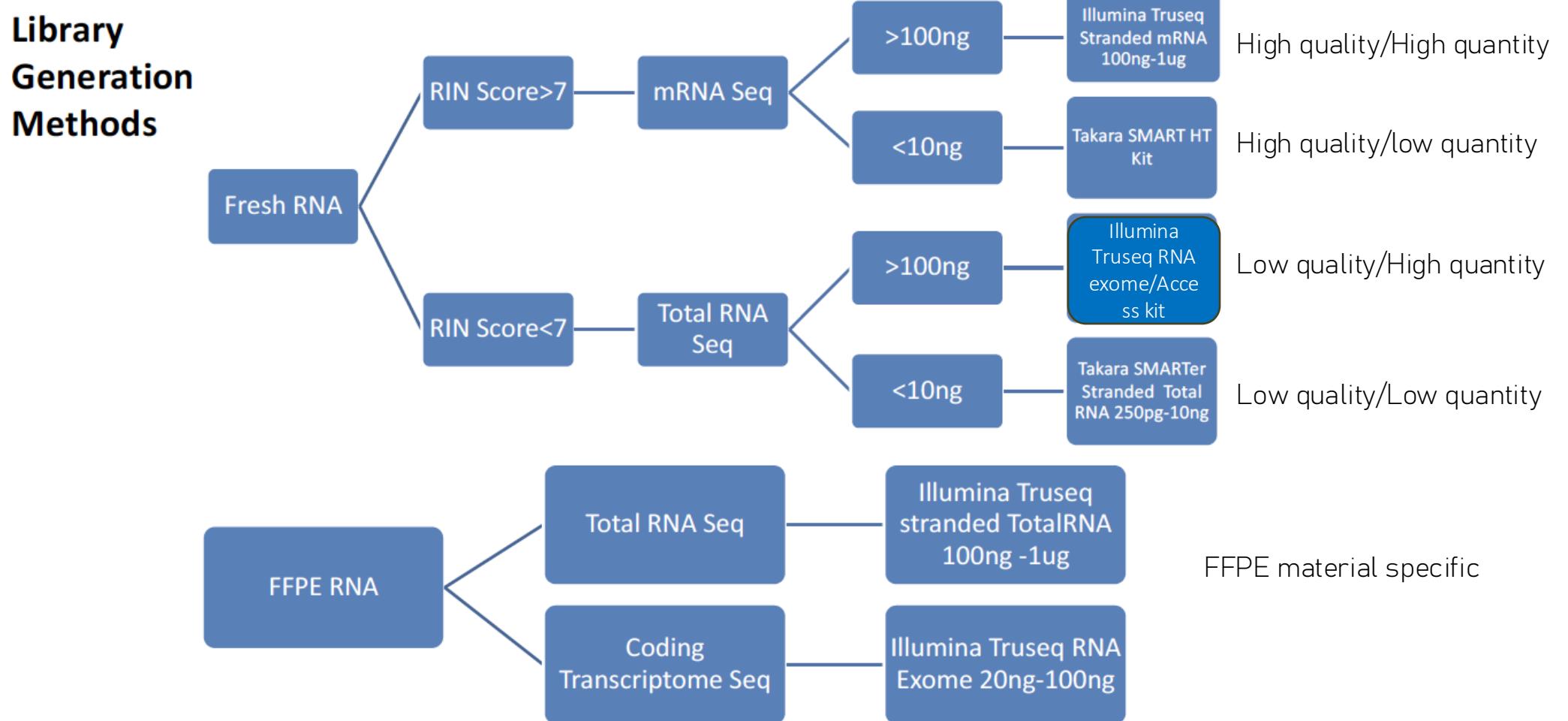
(2) Remove ribosomal by hybridization+immobilization

(3) Prime first strand cDNA synthesis from the polyA tail

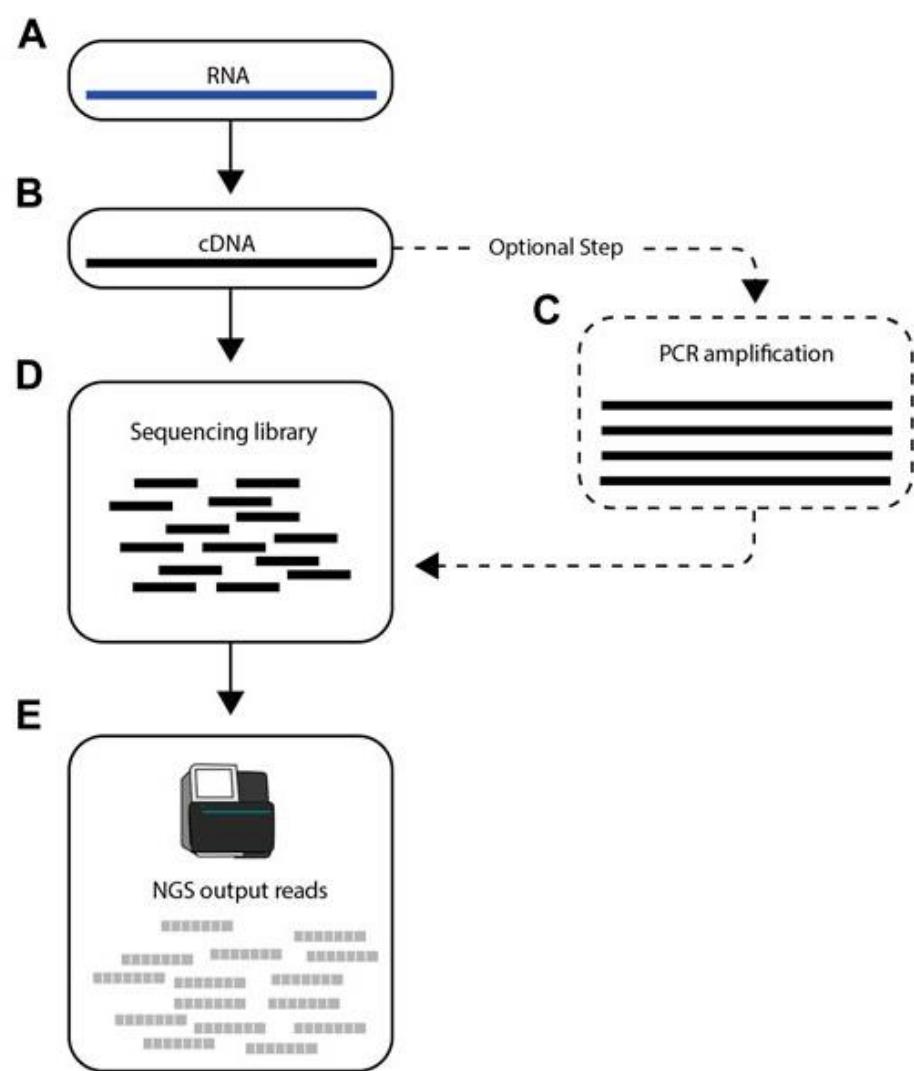
(4) Normalize the total RNA sample to bring the highly expressed RNAs (eg, rRNA) down near to the level of more lowly expressed RNA (eg, messenger RNA)

(5) Degrade all messages without a 5' cap. (Epicentre terminal exonuclease)

Bulk RNA sequencing – RNA library generation framework

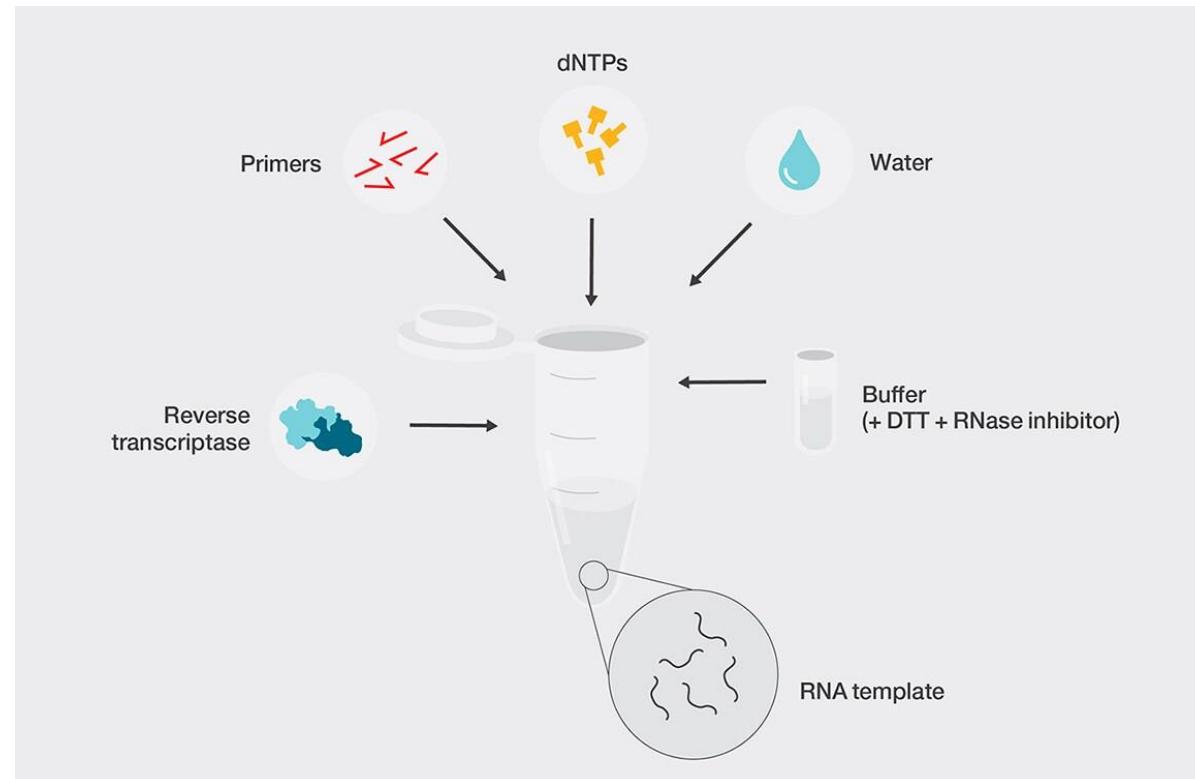


Bulk RNA sequencing – Library generation main step – cDNA formation

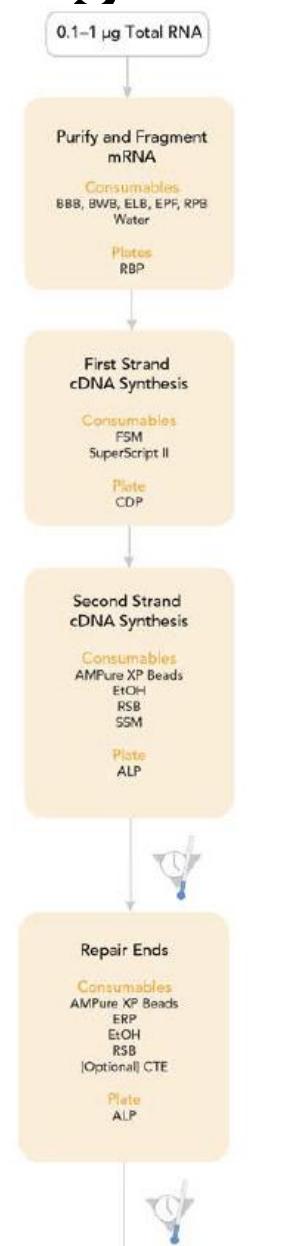
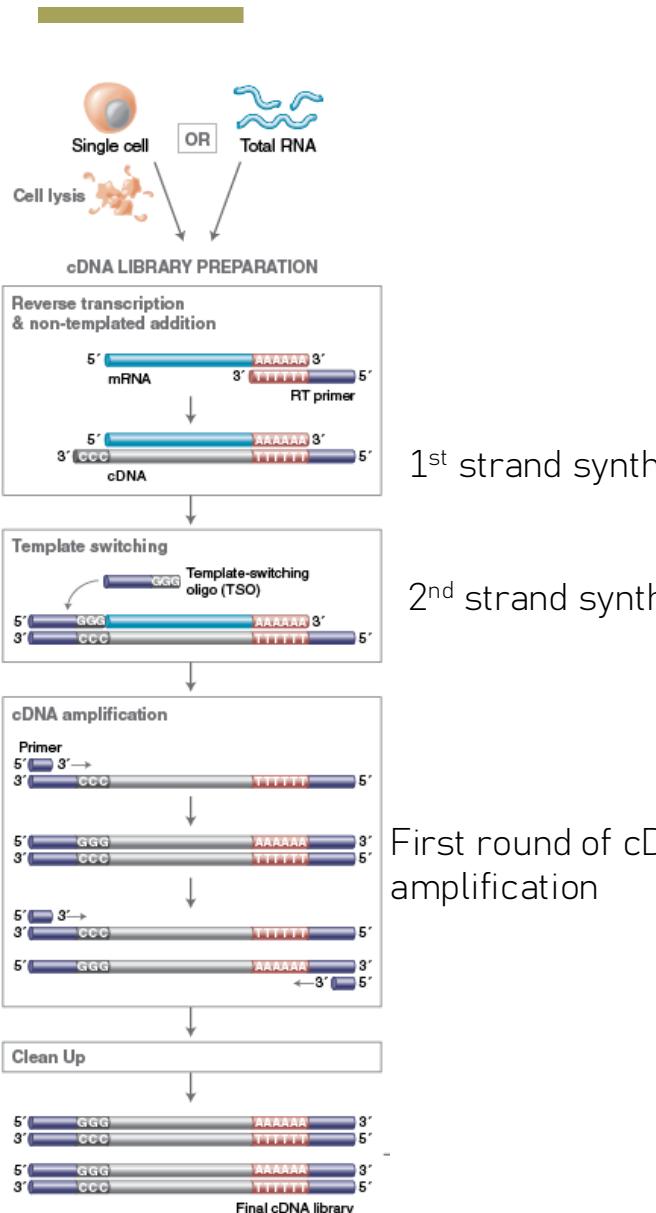


cDNA is a much more stable product to work with

RNases are everywhere!



Bulk RNA sequencing – Library generation addl steps



Reagent	Function	Control	Structure of Control DNA Ends
End Repair Mix	End repair: Generate blunt ended fragments by 3'→5' exonuclease and 5'→3' polymerase activities	End Repair Control 1*	5' overhang at one end, 3' overhang at other end
End Repair Mix	End repair: Add 5'-phosphate groups needed for downstream ligation	End Repair Control 2*	Blunt with 5'-OH group
A-Tailing Mix	A-tailing: Make fragments compatible with adapters and prevent self-ligation by adding a 3'-A overhang	A-Tailing Control	Blunt with 5'-phosphate group
Ligation Mix	Ligation: Join 3'-T overhang adapters to 3'-A overhang inserts	Ligation Control	Single-base 3' 'A' base overhang

Optional Stopping Point
Store after cleanup
at -15° to -25°C

Bulk RNA sequencing – Stranded vs non-stranded

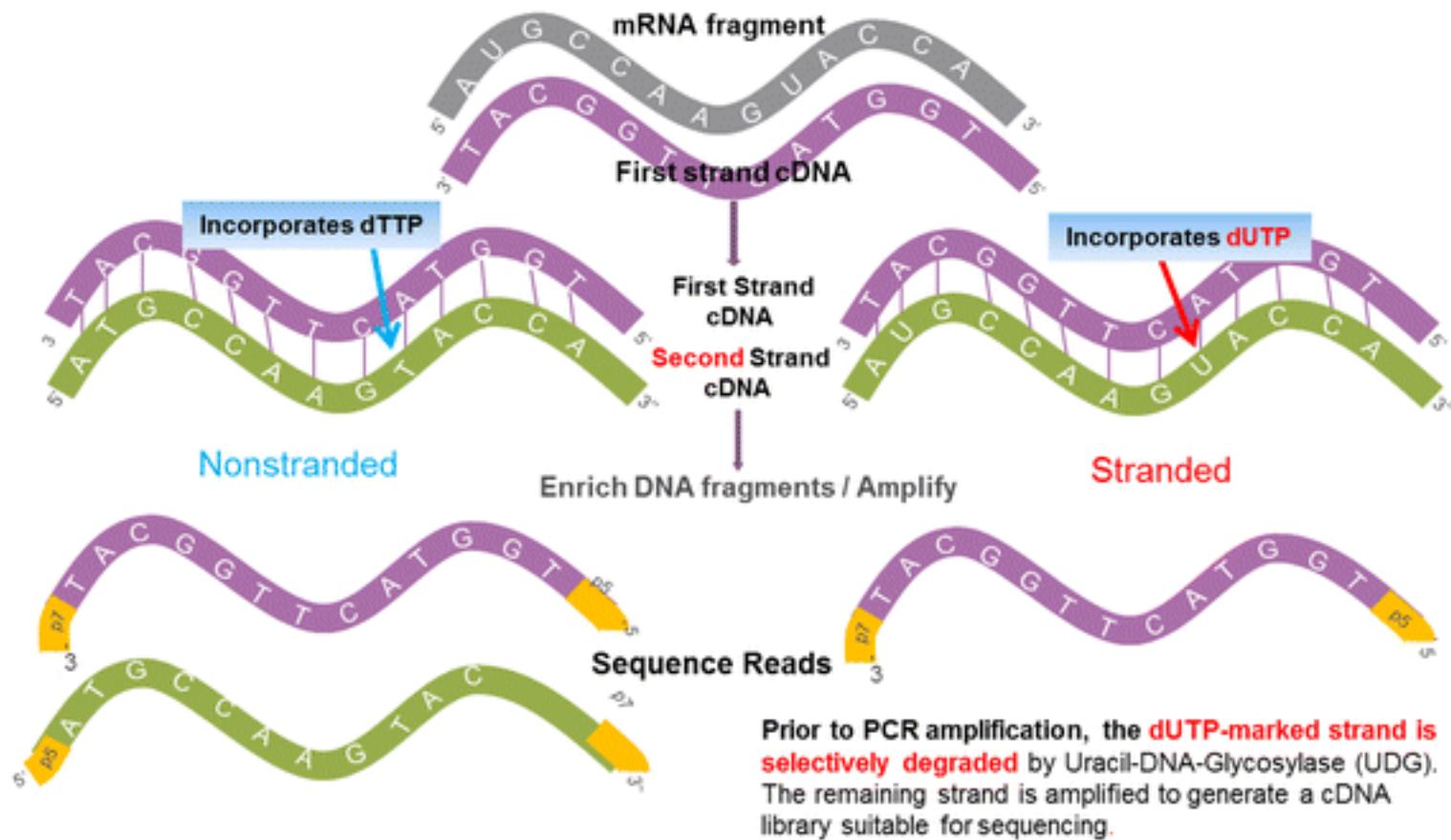
Strand RNA-seq preserves the original orientation of the RNA molecule after reverse transcription

Allows identification of sense and antisense transcripts

All Illumina TruSeq kits are stranded

Stranded RNA is more accurate for gene expression calculations

Stranded RNA sequencing is more expensive



Bulk RNA sequencing – Single ended vs paired-end sequencing

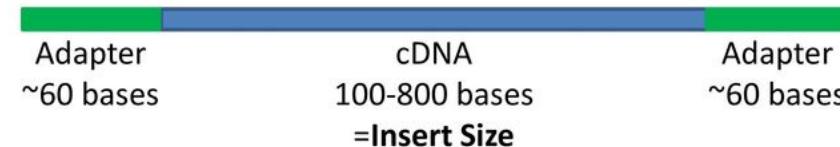
Paired-end sequencing allows reading the RNA insert from both ends

Indicates how many Illumina adapters be added

Provides more information of the RNA and will be especially needed for gene fusions, translocations, isoform detections etc.

Factors to consider

1. Read length
2. Sequencing Depth
3. Cost
4. Application



Batch effect

How to know whether you have batches?

Were all RNA isolations performed on the same day?

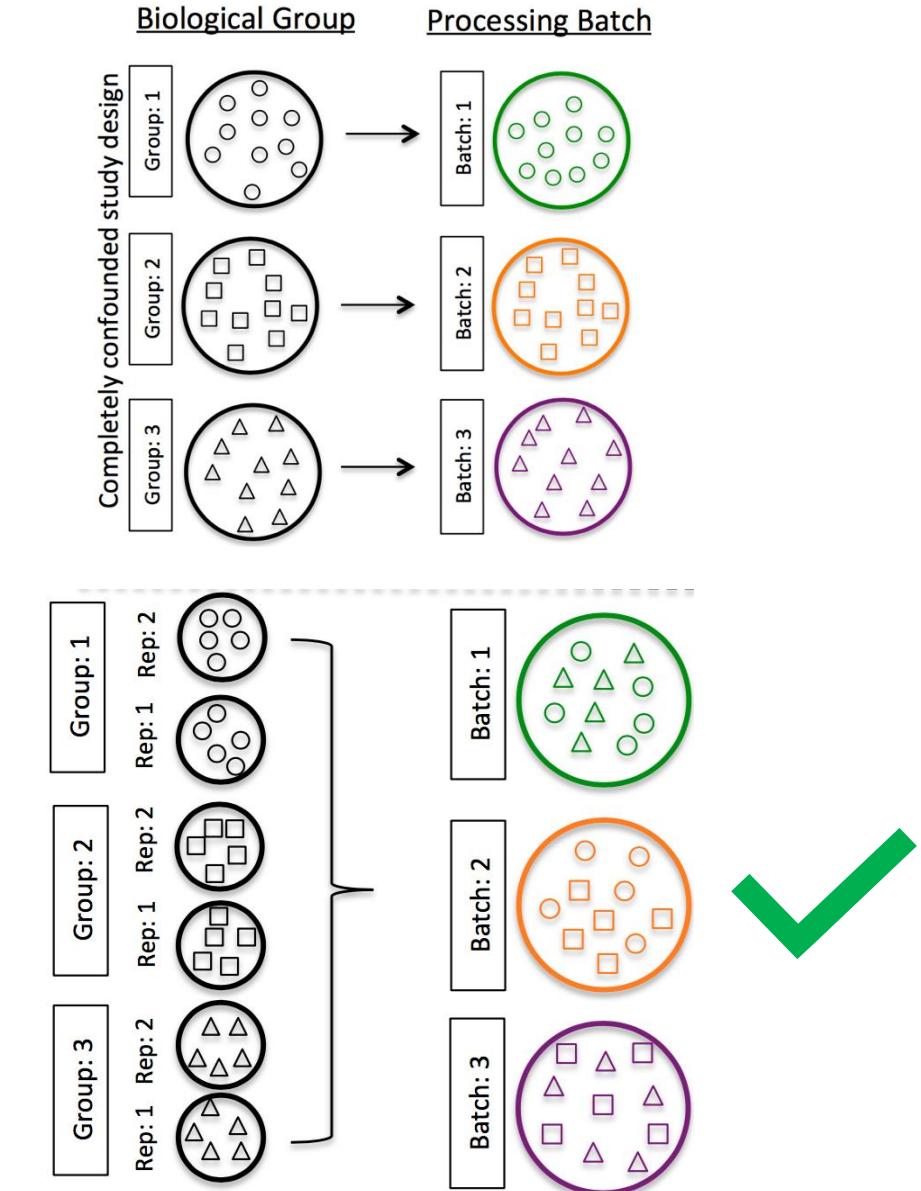
Were all library preparations performed on the same day?

Did the same person perform the RNA isolation/library preparation for all samples?

Did you use the same reagents for all samples?

Did you perform the RNA isolation/library preparation in the same location?

If any of the answers is 'No', then you have batches!

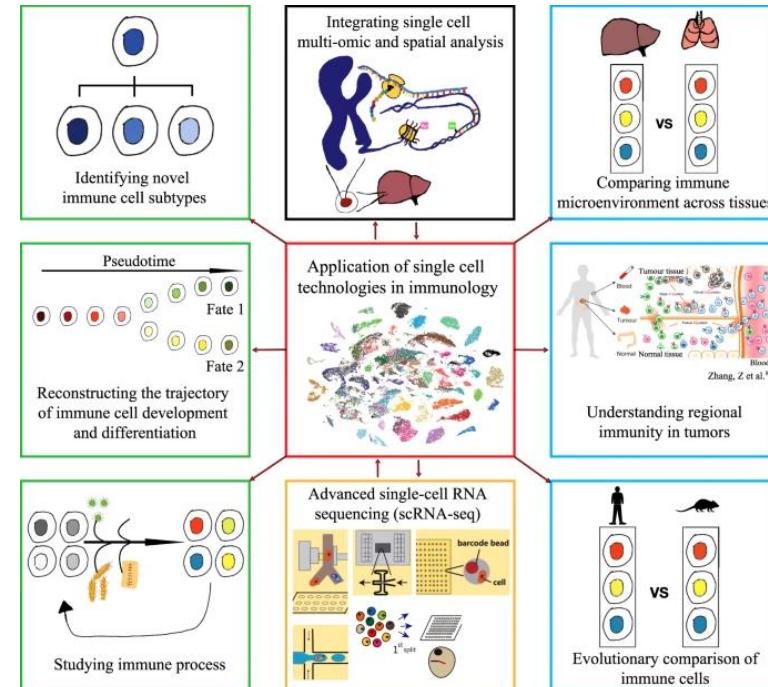
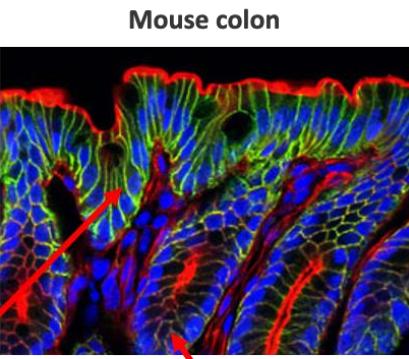
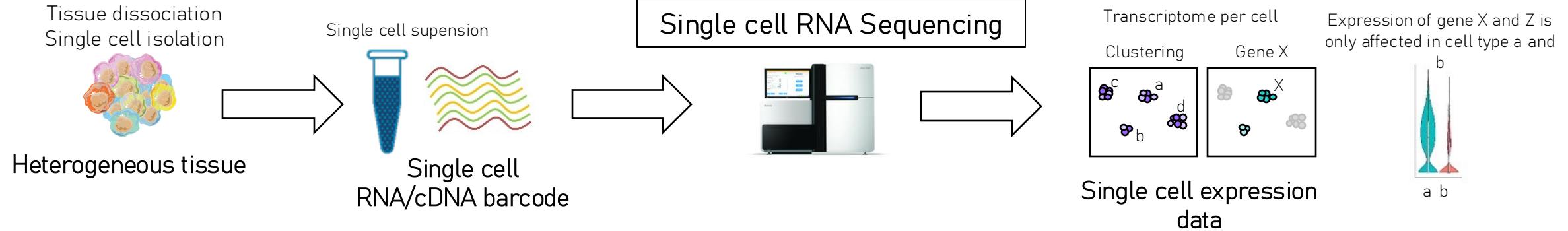


Bulk RNA sequencing - example

DNA and total RNA were extracted from fresh frozen polyps and purified using QIAGEN's AllPrep DNA/RNA/miRNA Universal Kit (QIAGEN), following the manufacturer's instructions. Briefly, the frozen tissue samples were first disrupted and homogenized using Lysing Matrix E (MP Bio) by shaking the tubes on a bead-beater at 5.5 m/sec for 30 s. The lysate was then passed through an AllPrep DNA Mini spin column. This column allows selective and efficient binding of genomic DNA. Following on-column Proteinase K digestion, the column was then washed and pure, ready-to-use DNA was eluted. Flow-through from the DNA Mini spin column was then digested by Proteinase K in the presence of ethanol and applied to the RNeasy Mini spin column, where the total RNA binds to the membrane. Following DNase I digestion, contaminants were efficiently washed away and high-quality RNA was eluted in RNase-free water. The quantity and quality of the DNA/RNA samples were checked by Nanodrop (E260/E280 and E260/E230 ratio) and by separation on an Agilent BioAnalyzer.

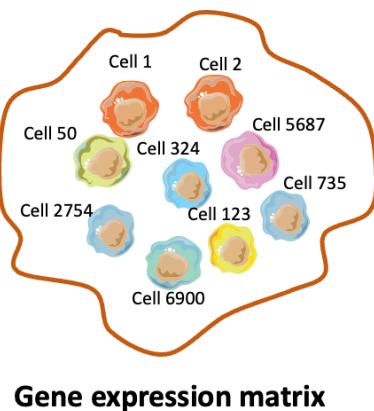
TCPS bulk RNA sequencing and alignment—Bulk RNA-sequencing was performed by Aros Applied Biotechnology A/S. This process involves the initial QC on an Agilent Bioanalyzer, with a minimum quality threshold of the DV200 at 30%. Total RNA-seq libraries which pass this QC threshold are prepared alongside a high-quality human reference RNA control. 100ng of RNA per sample is input to an Illumina TruSeq RNA Access Library Prep Kit, with protocol version 0.2. The yielded libraries undergo another round of QC through qPCR and quantified with a Qubit 2.0 Fluorometer, using its corresponding DNA BR Assay kit (Qubit), and size profiled on an Agilent Bioanalyzer. Pools of 4 libraries in equimolar amounts are created and undergo a final round of QC. These pools are loaded onto paired-end flow cells of a HiSeq2500 equipped with a cBot for sequencing at: 101 read cycles, 7 index cycles, and 101. The samples will be sequenced on a HiSeq2500 using 101 cycles for read 1, 7 index reads, and another 101 cycles for read 2. Following sequencing data generation, the reads are demultiplexed through Illumina's Genome Studio CASAVA software, which detected an average of 120 million reads per 4 sample pool.

Single cell RNA sequencing - Applications



Images courtesy of Cell Applications, Cell Signaling and Perou Lab
<https://doi.org/10.1038/s41423-019-0214-4>

Single cell vs Bulk RNA sequencing

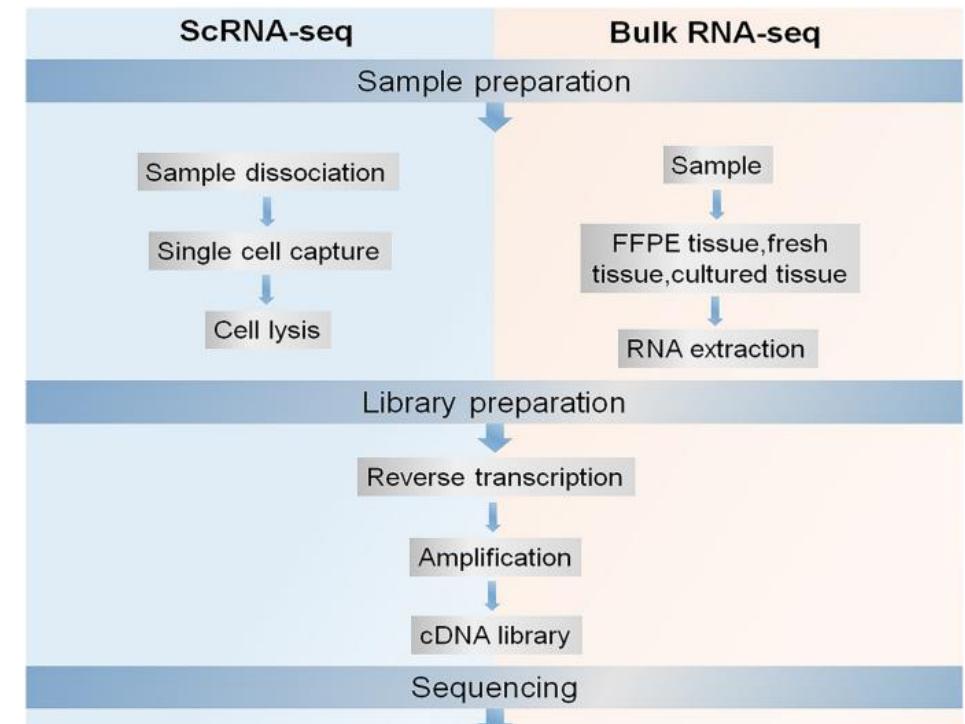


RNAseq counts from bulk RNA

	Sample 1
Gene 1	235
Gene 2	1000
Gene 3	50
Gene 4	354
...	
Gene 1000	0
...	
Gene 20000	156

RNAseq counts from Single cell isolation

	Cell 1	Cell 2	...	Cell 6900
Gene 1	0	3		20
Gene 2	2	0		0
Gene 3	10	6		0
Gene 4	0	0		7
...				
Gene 1000	0	3		2
...				
Gene 20000	35	2		0



Single cell Isolation techniques

- The performance of cell isolation technology is typically characterized by three parameters:
Efficiency or throughput, purity and recovery

1. Physical properties

Size, density, electric charges, and deformability

- Limiting dilution method
- Density gradient centrifugation
- Membrane filtration
- Manual picking/Micromanipulation
- Laser capture microdissection (LCM)

2. Cellular biological characteristics

Beads, chips, plates: labeling

- Fluorescence-activated cell sorting (FACS)
- Magnetic-activated cell sorting (MACS)
- Microfluidics

Single cell Isolation techniques - physical

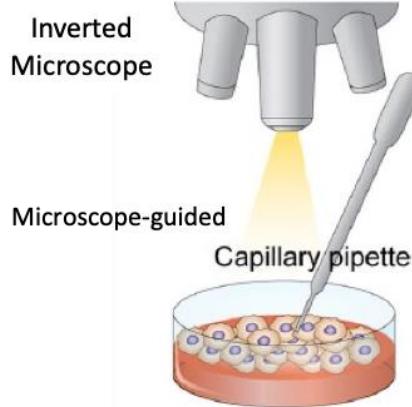
Limiting dilutions



96-well plate

Efficiency: Low
Purity: High
Recovery: Low

Manual picking/Micromanipulation



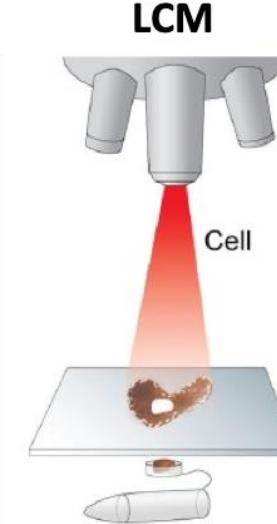
Inverted Microscope

Microscope-guided

Capillary pipette

Efficiency: Low
Purity: High
Recovery: Low

Laser capture microdissection



LCM

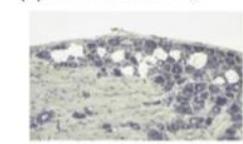
Cell

Laser-capture microdissection

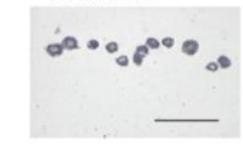
(i) Cell selection



(ii) Laser sectionning



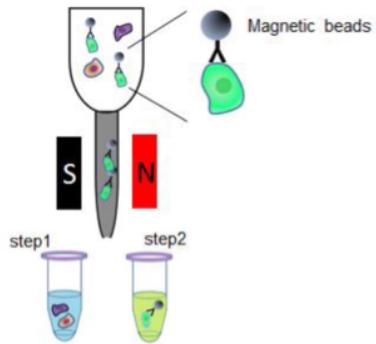
(iii) Cell transfer on a membrane



Efficiency: Low
Purity: Low
Recovery: Low

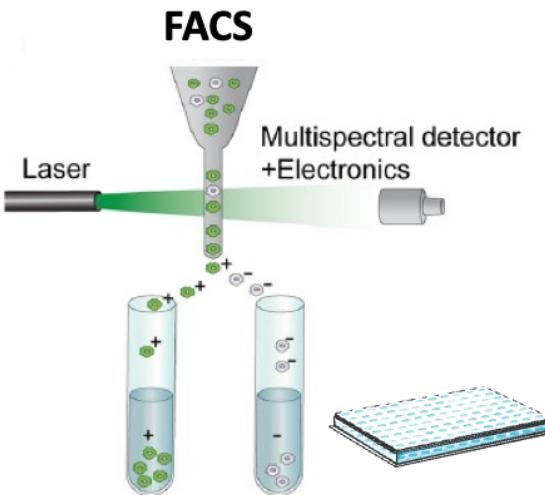
Single cell Isolation techniques - biological

Magnetic beads



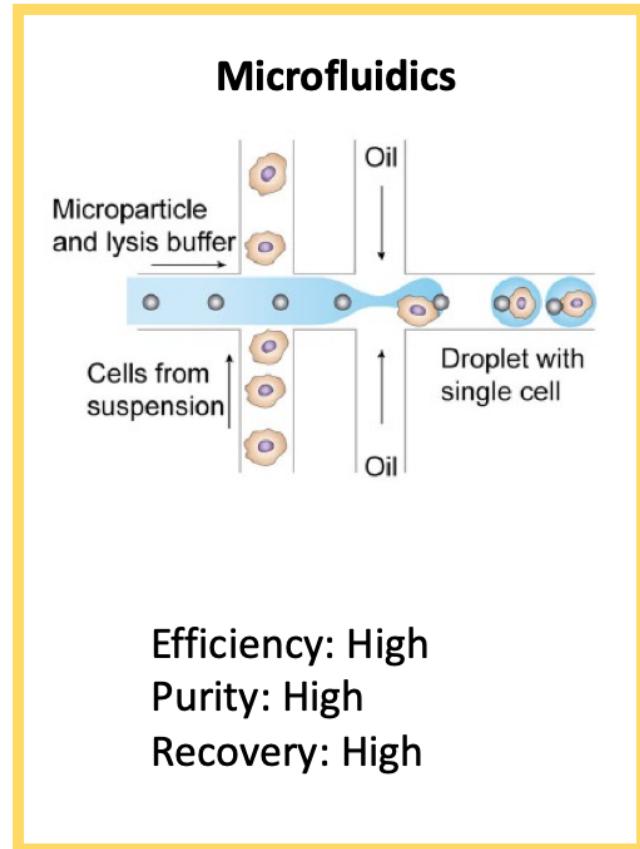
Efficiency: High
Purity: Medium
Recovery: High

Cellular biological characteristics



Efficiency: High
Purity: High
Recovery: High

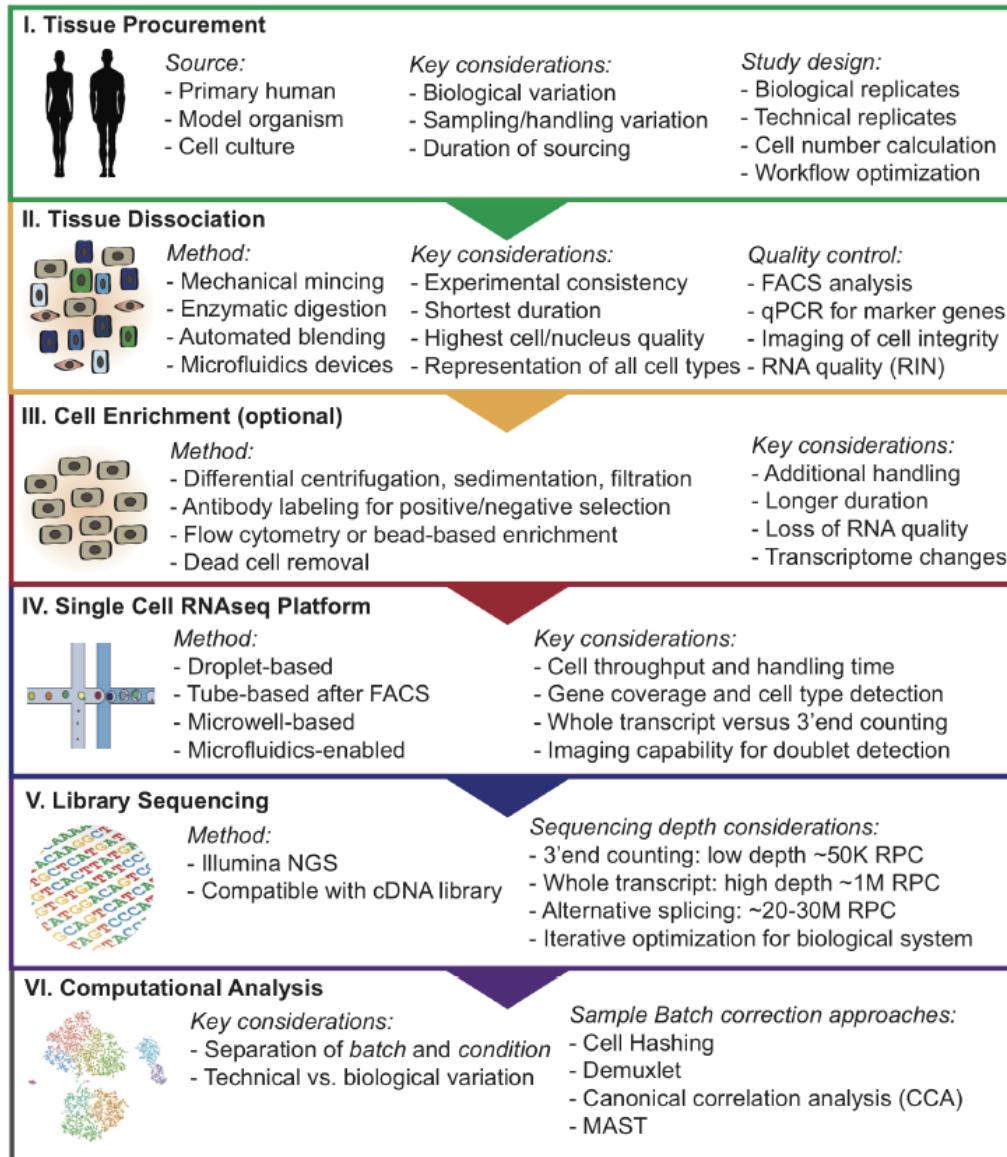
Microdroplet-based microfluidics



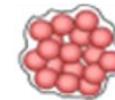
Efficiency: High
Purity: High
Recovery: High

10x Genomics offers high-throughput profiling of 3' ends of RNAs of single cells with high capture efficiency. Consequently, this method enables analysis of rare cell types in a sufficiently heterogeneous biological space.

Single cell RNA sequencing experiment flowchart



Fresh tissue?



Single-cell RNA sequencing methods are optimal when cells can be harvested intact and viable and we can sequence the cytoplasmatic RNA

Frozen?

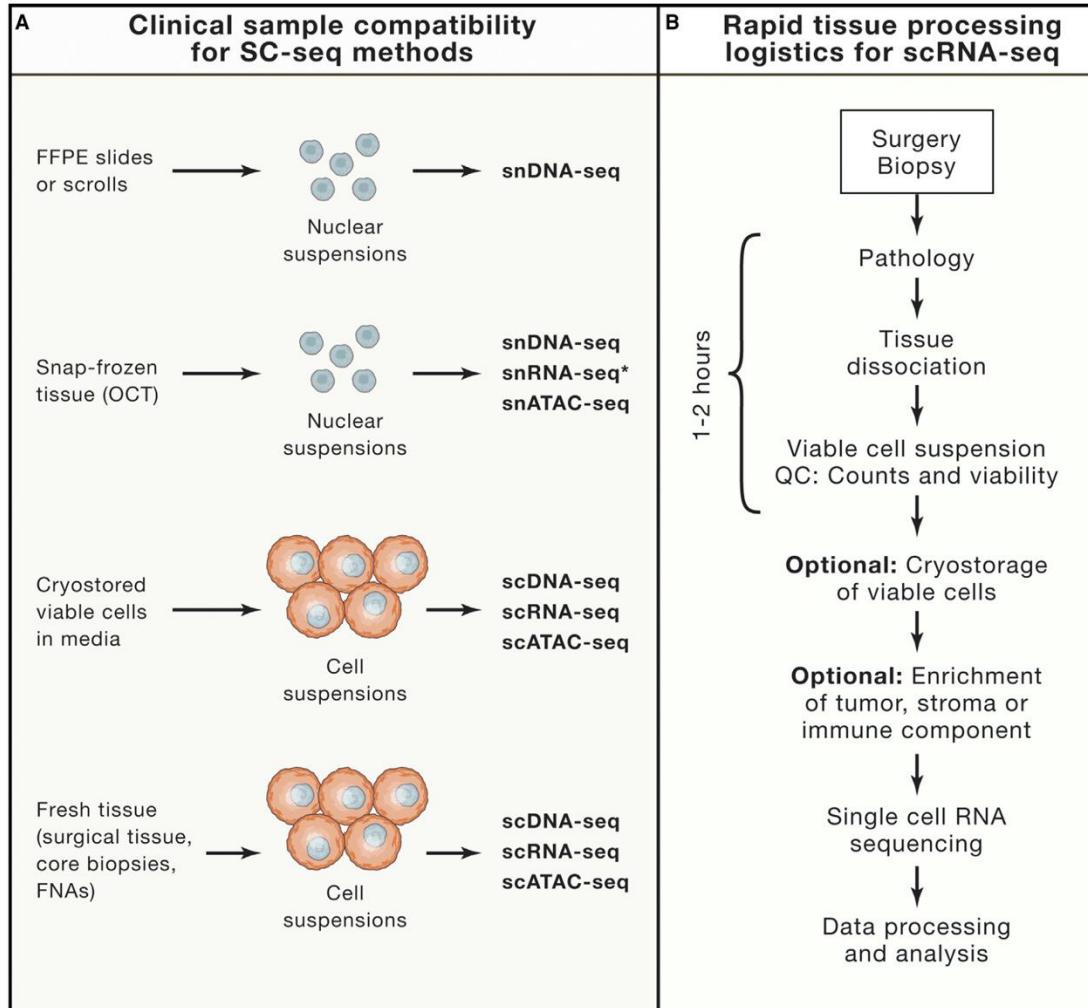


SINGLE NUCLEI ISOLATION

- Enriched for several types of nuclear RNAs
- Higher abundance of intronic sequences in snRNAseq, which ranged between 10–40% of mapped reads

Single cell RNA sequencing- Frozen/FFPE example protocol

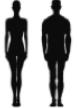
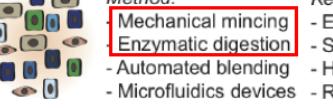
Single cell vs. Single Nuclei



What might you miss if you are using single cell?

- Sometimes entire cell types can be missed, as droplet technology prefers small round cells
 - Neurons – long axon and introns cause issues
 - Adipose – cell size causes problems
- Sample availability
 - Nuclei samples you can use flash frozen
 - Newer technology use FFPE tissues too
 - Archived samples allow larger studies
 - Cell samples require dissociation
 - Need quick processing and high viability.
 - Dissociation strongly affects cell types observed.
 - You can freeze down after dissociation – but is cell type dependent how well cells wake back up

Single cell RNA sequencing- fresh cells dissociation example protocol

I. Tissue Procurement	Key considerations:	Study design:
	Source: <ul style="list-style-type: none">- Primary human- Model organism- Cell culture	<ul style="list-style-type: none">- Biological variation- Sampling/handling variation- Duration of sourcing
		<ul style="list-style-type: none">- Biological replicates- Technical replicates- Cell number calculation- Workflow optimization
II. Tissue Dissociation	Method:	Key considerations:
	Mechanical mincing Enzymatic digestion	<ul style="list-style-type: none">- Experimental consistency- Shortest duration- Highest cell/nucleus quality- Representation of all cell types
	Automated blending Microfluidics devices	<ul style="list-style-type: none">- qPCR for marker genes- Imaging of cell integrity- RNA quality (RIN)



DISSOCIATION OF MAMMARY TISSUE

Gentle Collagenase/Hyaluronidase cocktail

Rotary shaker at 37°C
Magnetic stir O/N

Wash and add ammonium Chloride Solution

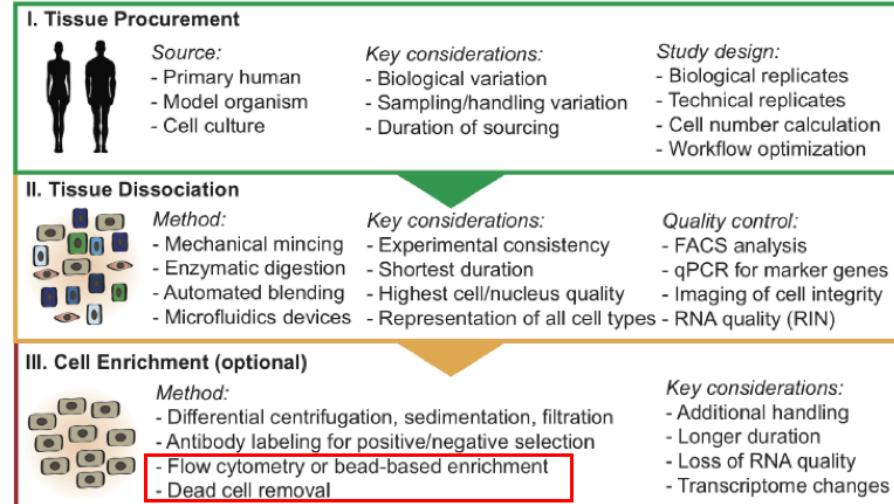
RBC removal

Wash and trypsinize

Wash, filter and count to check cell viability

Resuspend the cells in the appropriate concentration depending on the single cell technique

Single cell RNA sequencing- fresh cells dissociation example protocol



DISSOCIATION OF MAMMARY TISSUE

Gentle Collagenase/Hyaluronidase cocktail

Rotary shaker at 37°C
Magnetic stir O/N

Wash and add ammonium Chloride Solution



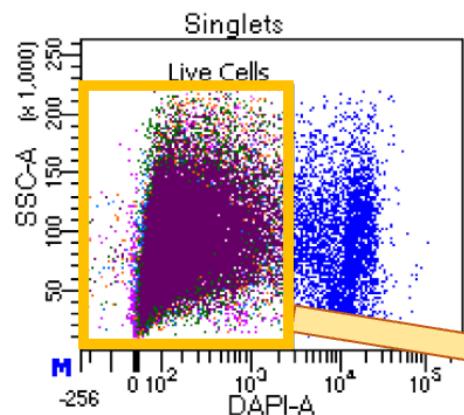
Wash and trypsinize



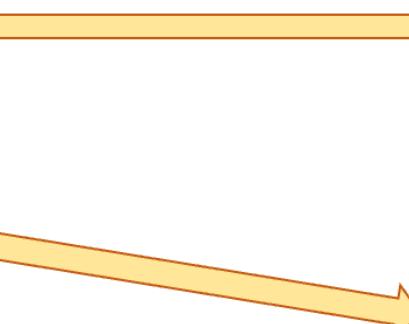
Wash, filter and count to check cell viability



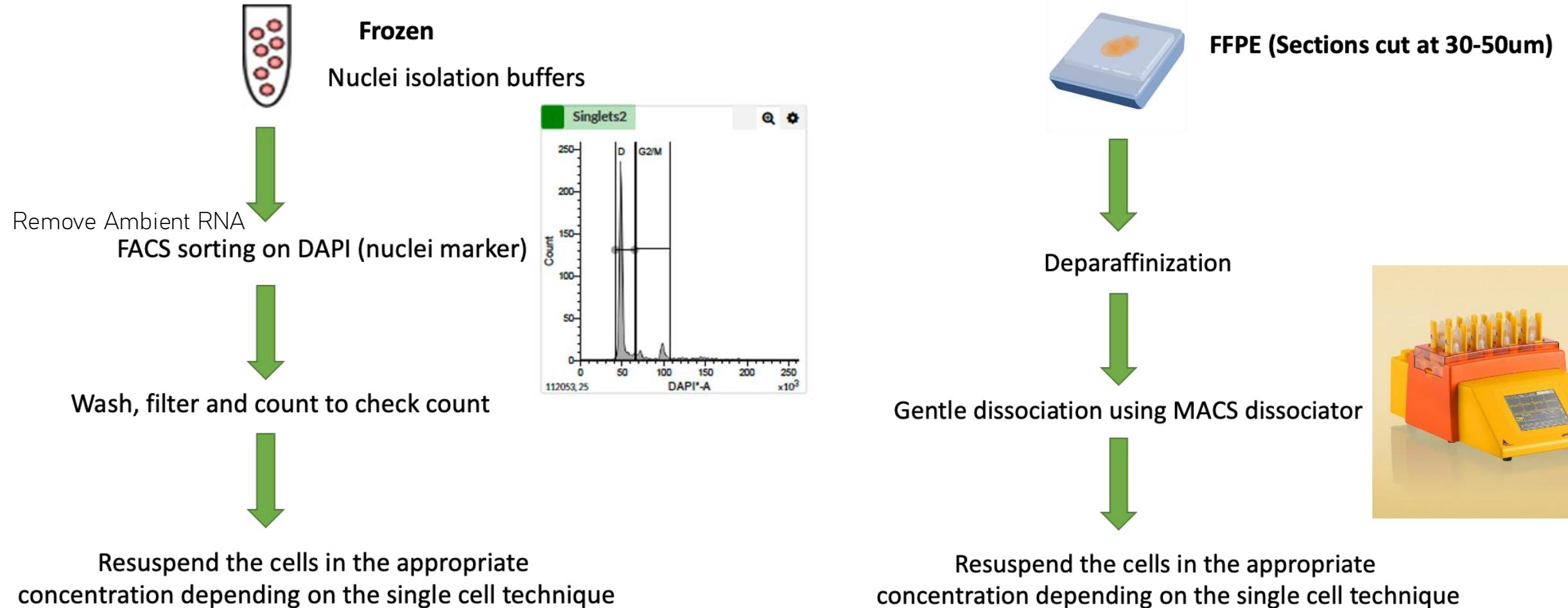
Resuspend the cells in the appropriate concentration depending on the single cell technique



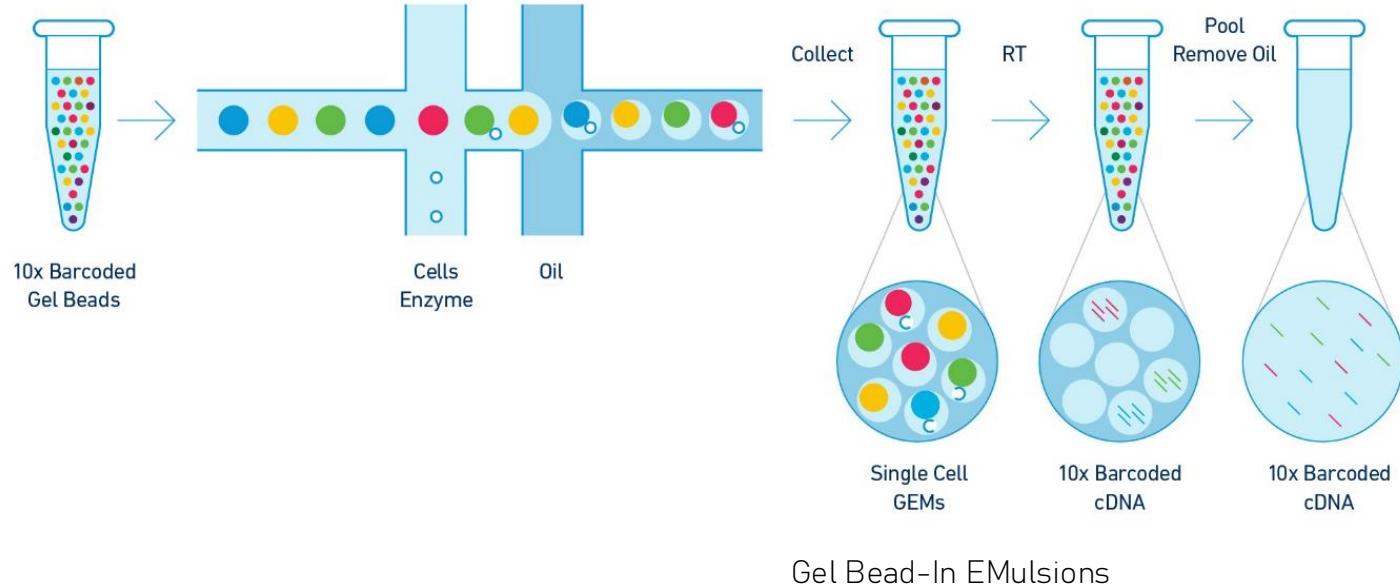
Stain cells with a live/dead cell marker



Single cell RNA sequencing- Frozen/FFPE example protocol



Single cell RNA sequencing- Cell counting is critical



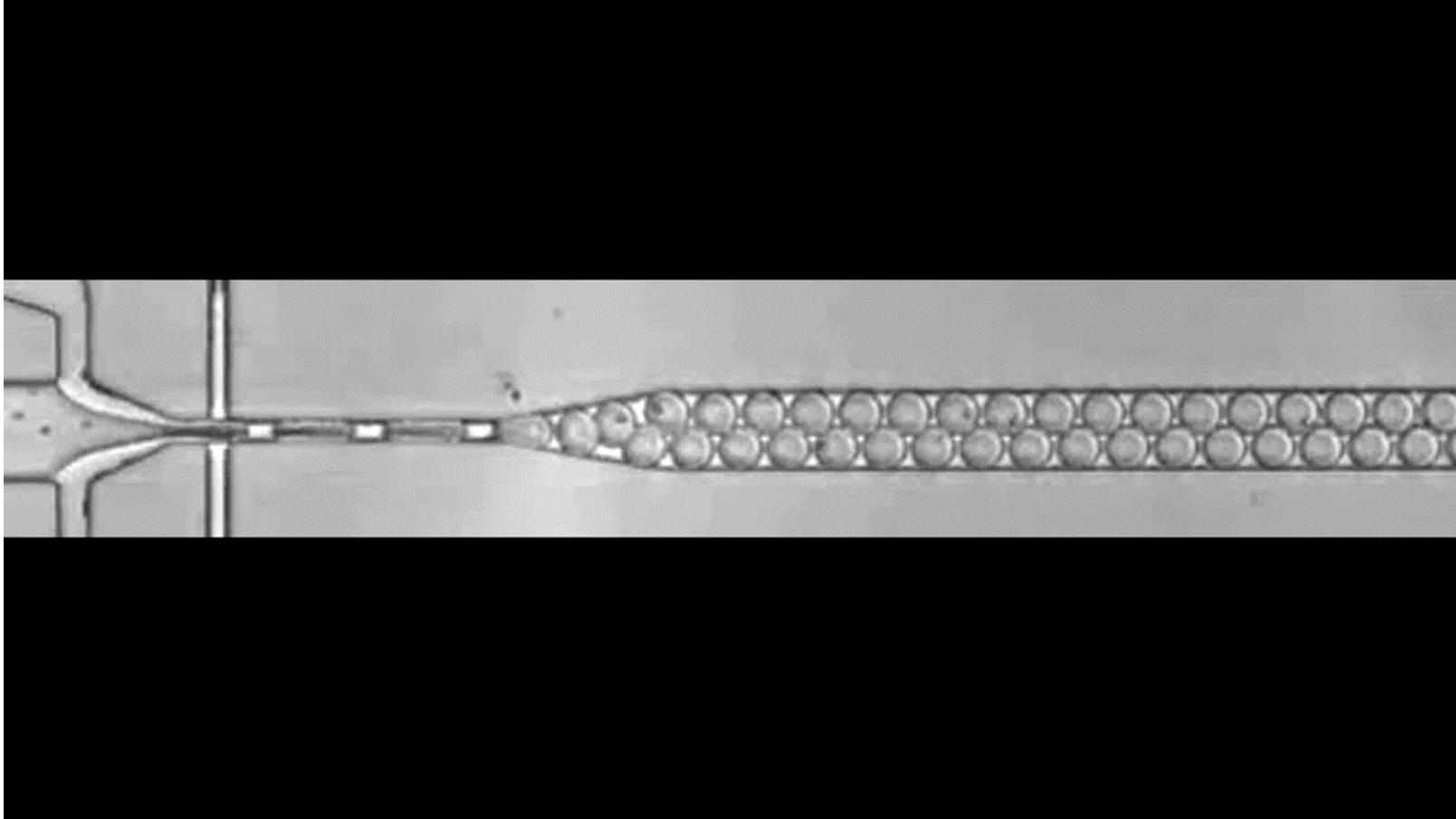
Version 1

GEM Outlet ►
Beads ►
Cells +reagents ►
Oil ►



In most high-throughput droplet based single cell experiments you cannot visualize individual droplets which is why you want to have an optimal density of the sample to prevent cell multiplet/doublet formations in one droplet

Single cell RNA sequencing- Cell counting is critical in Droplet based methods



Single cell RNA sequencing- Cell counting is critical

Step 1 GEM Generation & Barcoding

Cell Suspension Volume Calculator Table
(for step 1.2 of Chromium Next GEM Single Cell 3' v3.1 protocol)

Volume of Cell Suspension Stock per reaction (μ l) | Volume of Nuclease-free Water per reaction (μ l)

Cell Stock Concentration (Cells/ μ l)	Targeted Cell Recovery										
	500	1000	2000	3000	4000	5000	6000	7000	8000	9000	10000
100	8.3 35.0	16.5 26.7	33.0 10.2	n/a							
200	4.1 39.1	8.3 35.0	16.5 26.7	24.8 18.5	33.0 10.2	41.3 2.0	n/a	n/a	n/a	n/a	n/a
300	2.8 40.5	5.5 37.7	11.0 32.2	16.5 26.7	22.0 21.2	27.5 15.7	33.0 10.2	38.5 4.7	n/a	n/a	n/a
400	2.1 41.1	4.1 39.1	8.3 35.0	12.4 30.8	16.5 26.7	20.6 22.6	24.8 18.5	28.9 14.3	33.0 10.2	37.1 6.1	41.3 2.0
500	1.7 41.6	3.3 39.9	6.6 36.6	9.9 33.3	13.2 30.0	16.5 26.7	19.8 23.4	23.1 20.1	26.4 16.8	29.7 13.5	33.0 10.2
600	1.4 41.8	2.8 40.5	5.5 37.7	8.3 35.0	11.0 32.2	13.8 29.5	16.5 26.7	19.3 24.0	22.0 21.2	24.8 18.5	27.5 15.7
700	1.2 42.0	2.4 40.8	4.7 38.5	7.1 36.1	9.4 33.8	11.8 31.4	14.1 29.1	16.5 26.7	18.9 24.3	21.2 22.0	23.6 19.6
800	1.0 42.2	2.1 41.1	4.1 39.1	6.2 37.0	8.3 35.0	10.3 32.9	12.4 30.8	14.4 28.8	16.5 26.7	18.6 24.6	20.6 22.6
900	0.9 42.3	1.8 41.4	3.7 39.5	5.5 37.7	7.3 35.9	9.2 34.0	11.0 32.2	12.8 30.4	14.7 28.5	16.5 26.7	18.3 24.9
1000	0.8 42.4	1.7 41.6	3.3 39.9	5.0 38.3	6.6 36.6	8.3 35.0	9.9 33.3	11.6 31.7	13.2 30.0	14.9 28.4	16.5 26.7
1100	0.8 42.5	1.5 41.7	3.0 40.2	4.5 38.7	6.0 37.2	7.5 35.7	9.0 34.2	10.5 32.7	12.0 31.2	13.5 29.7	15.0 28.2
1200	0.7 42.5	1.4 41.8	2.8 40.5	4.1 39.1	5.5 37.7	6.9 36.3	8.3 35.0	9.6 33.6	11.0 32.2	12.4 30.8	13.8 29.5
1300	0.6 42.6	1.3 41.9	2.5 40.7	3.8 39.4	5.1 38.1	6.3 36.9	7.6 35.6	8.9 34.3	10.2 33.0	11.4 31.8	12.7 30.5
1400	0.6 42.6	1.2 42.0	2.4 40.8	3.5 39.7	4.7 38.5	5.9 37.3	7.1 36.1	8.3 35.0	9.4 33.8	10.6 32.6	11.8 31.4
1500	0.6 42.7	1.1 42.1	2.2 41.0	3.3 39.9	4.4 38.8	5.5 37.7	6.6 36.6	7.7 35.5	8.8 34.4	9.9 33.3	11.0 32.2
1600	0.5 42.7	1.0 42.2	2.1 41.1	3.1 40.1	4.1 39.1	5.2 38.0	6.2 37.0	7.2 36.0	8.3 35.0	9.3 33.9	10.3 32.9
1700	0.5 42.7	0.9 42.2	1.9 41.3	2.9 40.3	3.9 39.3	4.9 38.3	5.8 37.4	6.8 36.4	7.8 35.4	8.7 34.5	9.7 33.5
1800	0.5 42.7	0.9 42.3	1.8 41.4	2.8 40.5	3.7 39.5	4.6 38.6	5.5 37.7	6.4 36.8	7.3 35.9	8.3 35.0	9.2 34.0
1900	0.4 42.8	0.9 42.3	1.7 41.5	2.6 40.6	3.5 39.7	4.3 38.9	5.2 38.0	6.1 37.1	6.9 36.3	7.8 35.4	8.7 34.5
2000	0.4 42.8	0.8 42.4	1.7 41.6	2.5 40.7	3.3 39.9	4.1 39.1	5.0 38.3	5.8 37.4	6.6 36.6	7.4 35.8	8.3 35.0

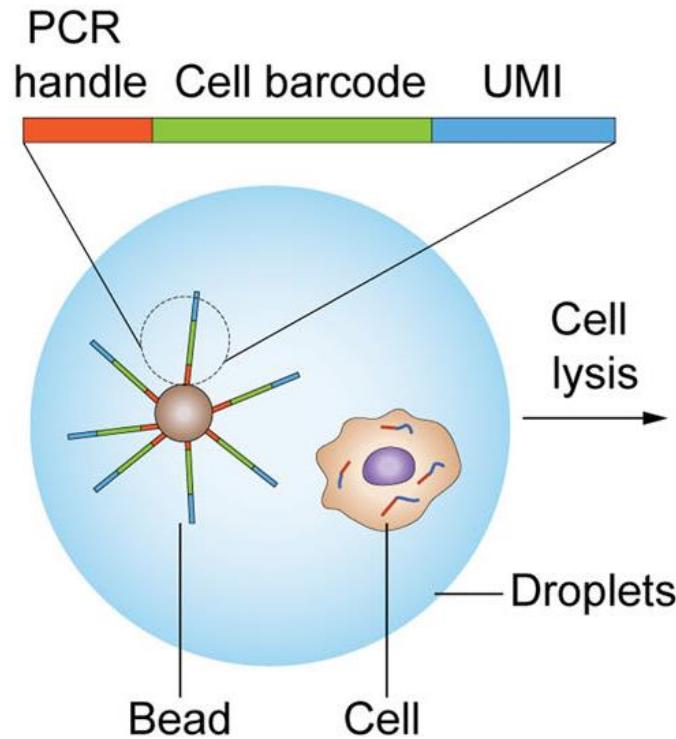
Grey boxes: Volumes that would exceed the allowable water volume in each reaction
Yellow boxes: Indicate a low transfer volume that may result in higher cell load variability
Blue boxes: Optimal range of cell stock concentration to maximize the likelihood of achieving the desired cell recovery target

Microdroplet-based microfluidics using 10x Chromium system

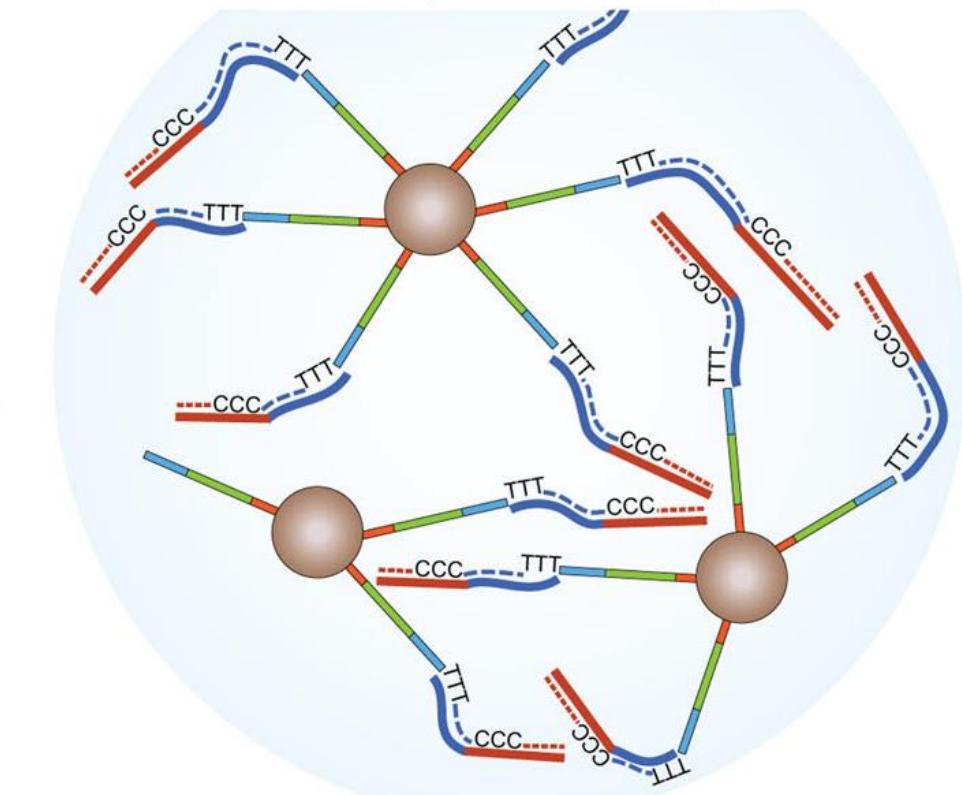
The blue box indicates the optimal range of cell stock concentration
To maximize the likelihood of achieving the desired cell recovery target

Single cell RNA sequencing- cDNA generation

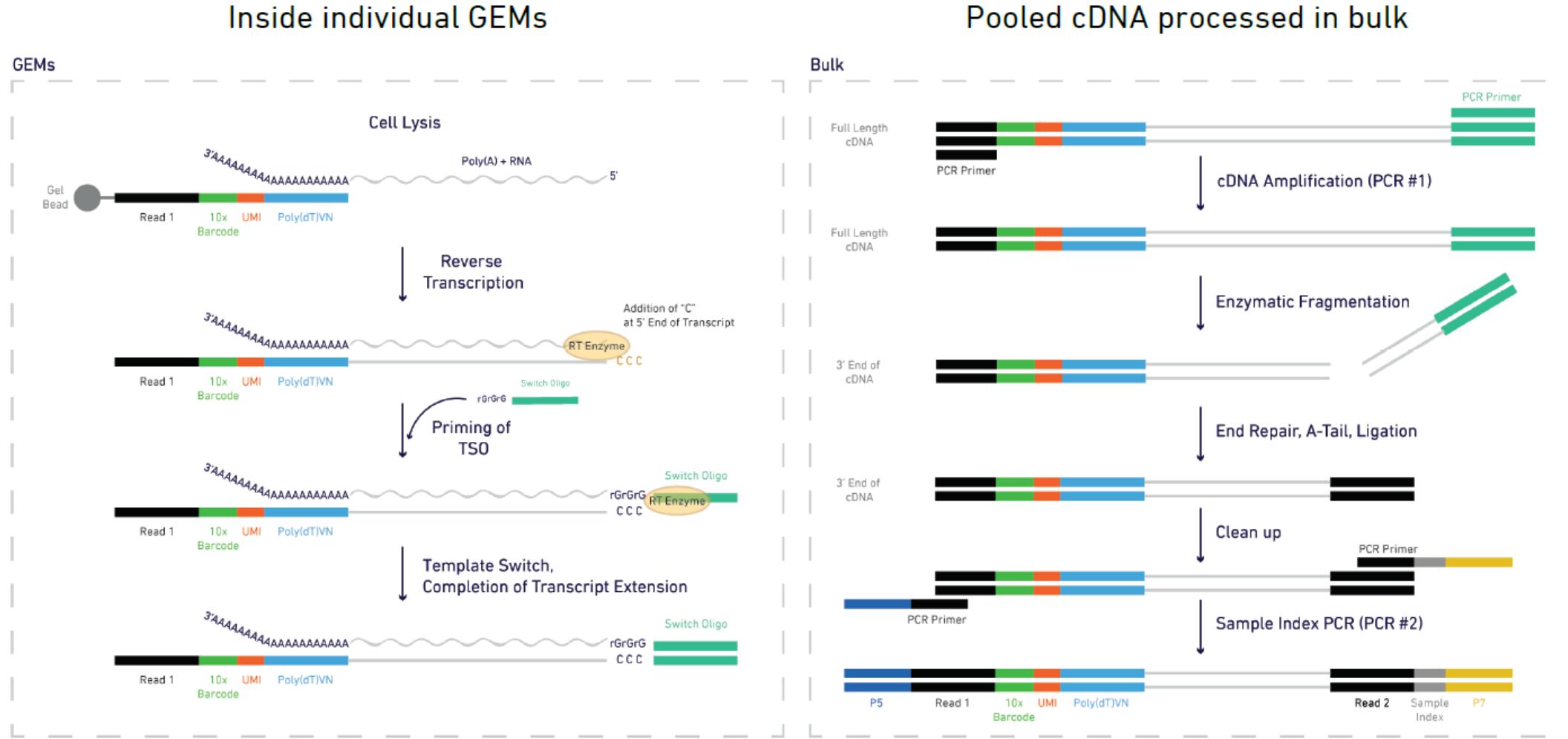
Structure of the barcode primer bead



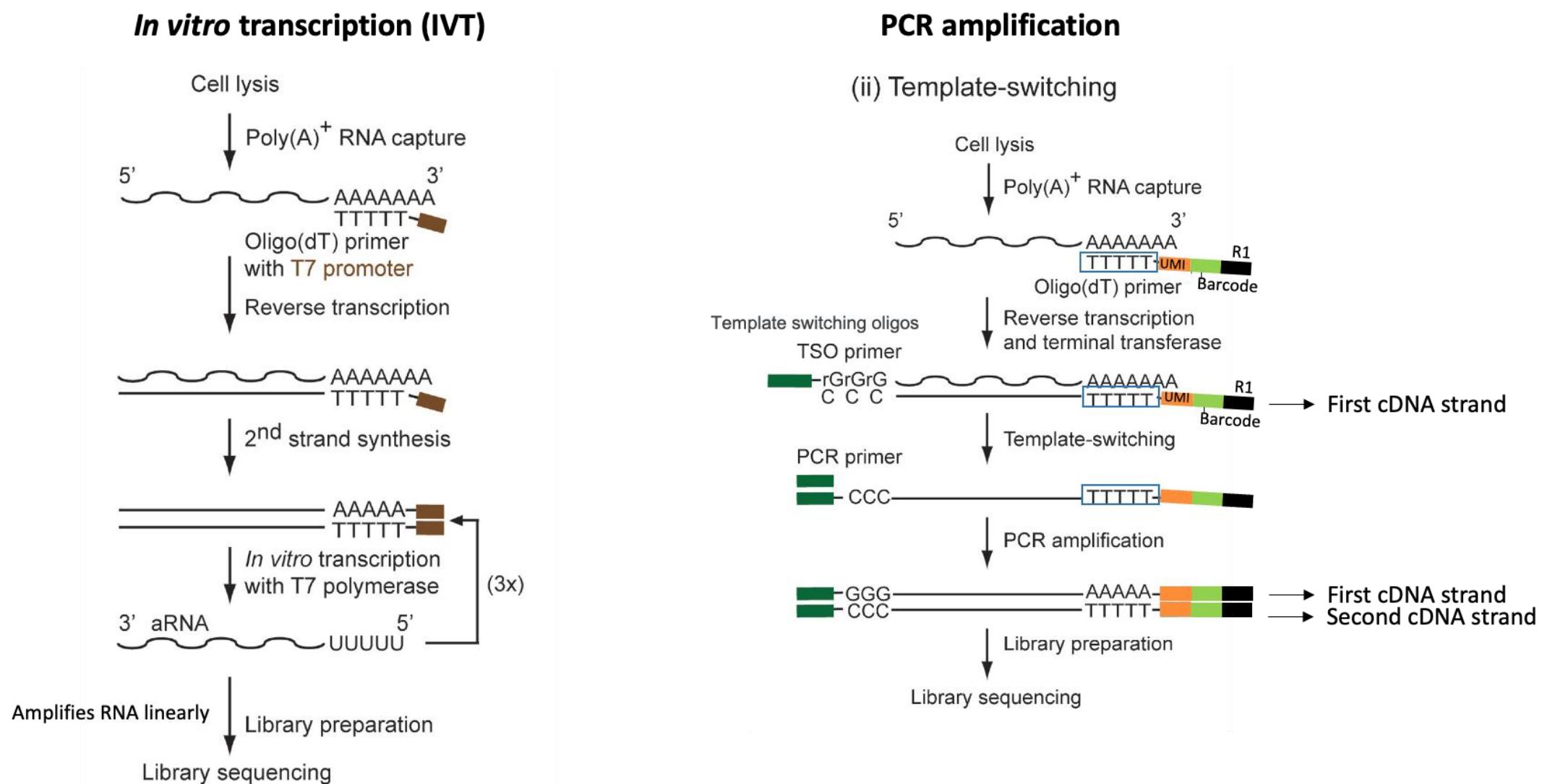
Reverse transcription with template switching



Single cell RNA sequencing- cDNA generation



Single cell RNA sequencing- cDNA generation



Single cell RNA sequencing- commercial single cell platforms

	Micro-manipulation / Automated Pipetting	FACS	Microwell encapsulation	Droplet encapsulation					
Cell Stress	Low	Moderate	Moderate	Moderate					
Selection	Yes	Yes	No* / Yes ⁺⁺	No*					
Doublet	Low	Low	Low-High	Moderate					
Throughput	Low	Moderate	Moderate	High					
Capture efficiency	Low	Moderate	Moderate	Low-Moderate					
Academic / Commercial scRNA workflow	- CellenONE (Cellenion) [†] - Smart-Seq2 (42)	- MARS-Seq (39) - Smart-Seq2 (42)	- C1 (Fluidigm) - ddSeq (Biorad / Illumina) - iCell8 (Clontech) ⁺⁺ - Rhapsody (BD)	- InDrop (1CellBio) - DropSeq (Dolomite-bio) - 10X (Chromium)					
Example of use	Fragile rare cells	Rare cells based on phenotype or marking	Large cell numbers	Large cell numbers					
	FACS	Microwell encapsulation	Droplet encapsulation						
Singlet Capture efficiency	Smart-Seq2	MARS-Seq	C1	ddSeq	iCell8	Rhapsody	InDrop	DropSeq	10X
82%	92%	39%	2.6%	37% ⁺⁺	Not reported	Not reported	7%	Not reported	50%
Not reported	2.27%	3-30%	5.8%	1.3-4%	0.6%		4%	0.36-11.3	1.6-3%
Reference	42	39	37 FWP	PB	PB	PB	36	37	26

[†]Automated pipetting system

^{*}Preselection or enrichment can be performed prior

⁺⁺Only reagents added to wells containing singlets, determined by system

FWP: Fluidigm white paper

PB: Product brochure / manual

Applications	Full length				3' sequencing and barcoding				
	Gene expression Splice variants and BCR and TCR repertoire diversity				Gene expression				
Costs	High			Low					
	Smart-Seq2	Smarter /iCell8/C1	NuGEN Solo	MARS-Seq	ddSeq	Rhapsody	InDrop	DropSeq	10X*
UMI	-	-	/	/	/	/	/	/	/
mRNA priming (1st strand syn)	poly T	poly T	Random priming & poly T	poly T	poly T	poly T	poly T	poly T	poly T
Template Switching	/	/	-	-	-	-	-	/	/
DNase treatment	-	-	/	/	-	-	-	-	-
cDNA preamplification	PCR	PCR	-	In Vitro Transcription	PCR	PCR	In Vitro Transcription	PCR	PCR
Targeted sequencing	-	-	Depletion	-	-	Enrichment	-	-	-
Library generation	Transposon Tagmentation	Transposon Tagmentation	cDNA fragmentation, adapter ligation & library amp	RNA fragmentation & adapter ligation	Transposon Tagmentation	PCR targeted primer panels	RNA fragmentation & adapter ligation	Transposon Tagmentation	cDNA fragmentation, adapter ligation & library amp
Example of use	Sequencing the TCR of tumour-infiltrating lymphocytes				High-throughput sequencing of large cell numbers from solid organ tumours in large patient cohorts				

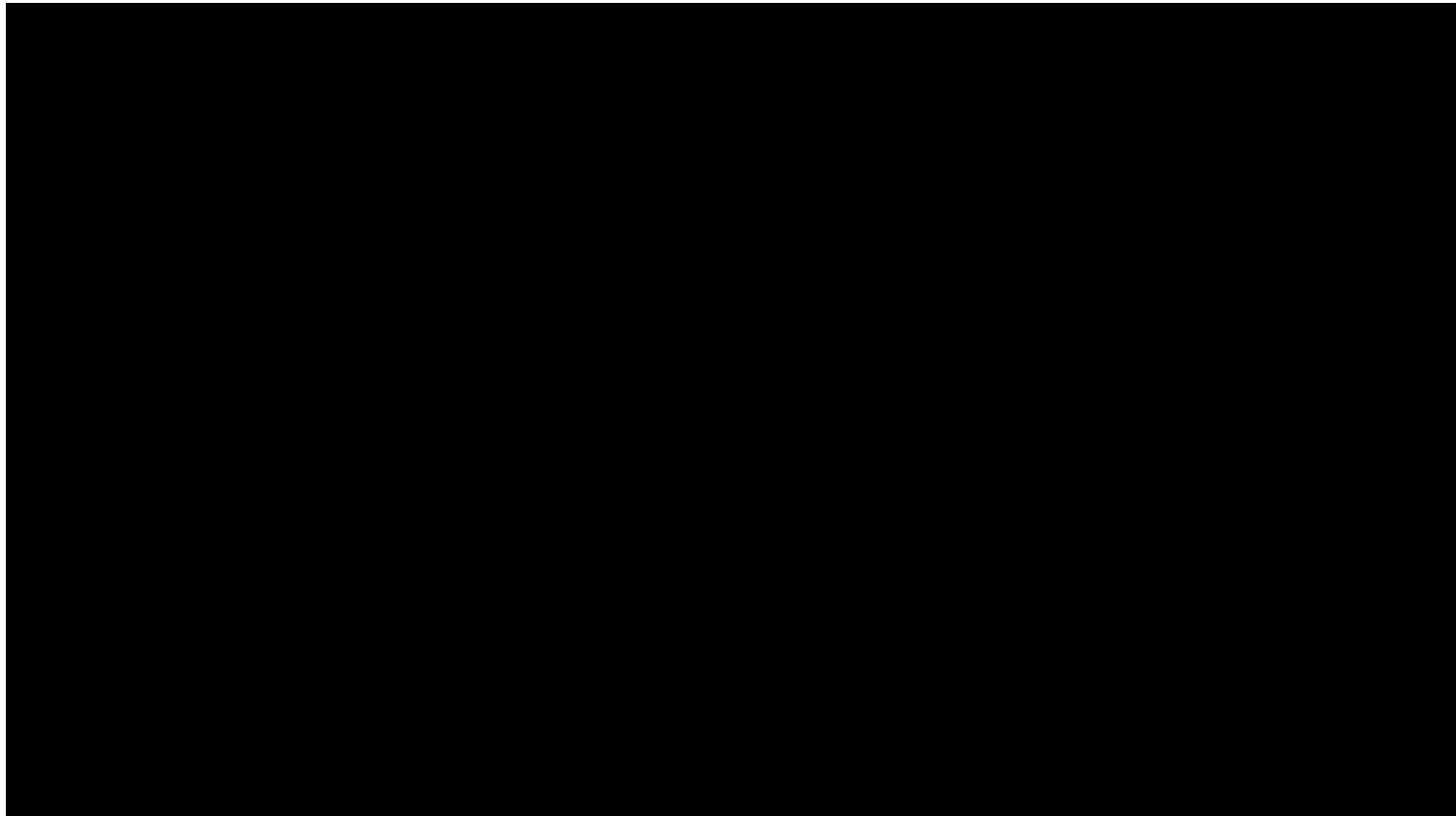
*10X has recently released a 5' barcoding that allows reconstruction of full length idiotypic sequences

Single cell RNA sequencing- iCell8 plate-based demo

Cell Isolation: Nanowells

ICELL8: Demonstration Example

Wu et al. Gigascience (2015)

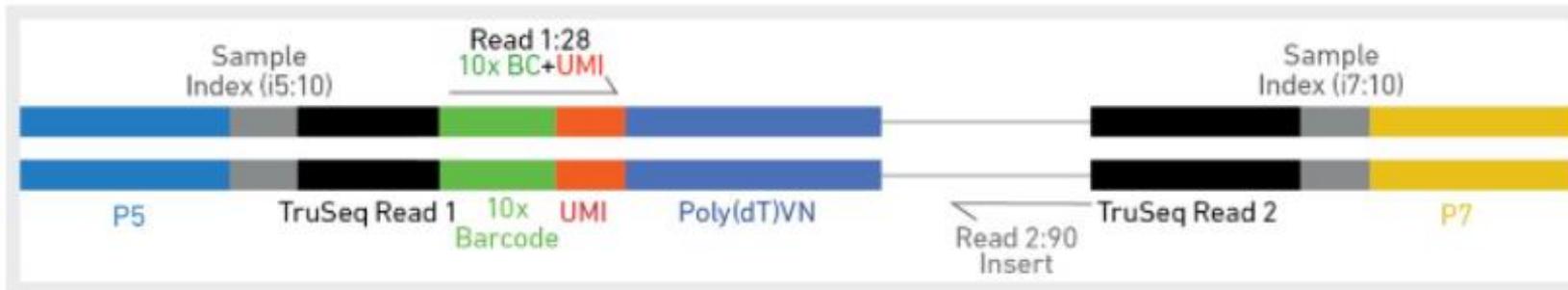


Single cell RNA sequencing – Example

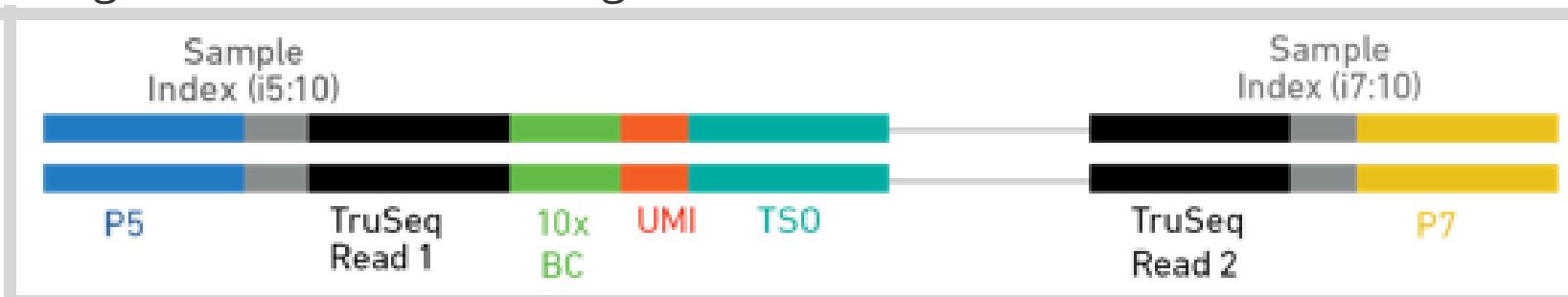
COLON MAP scRNA-seq, single-cell encapsulation and library generation –
Colonic biopsy samples were first placed into RPMI solution, minced to approximately 4mm², and washed with 1x DPBS. These samples were then incubated in chelation buffer (4mM EDTA, 0.5 mM DTT) at 4 °C for 1 h 15 min. Then, the resulting tissue suspension was dissociated with cold protease and DNase I for 25 minutes (Banerjee et al., 2020; Liu et al., 2018). This suspension was titrated throughout the process, every 10 minutes, then washed three times with 1x DPBS before encapsulation. Cells were encapsulated using a modified inDrop platform (Klein et al., 2015), and sequencing libraries were prepared using the TruDrop protocol (Southard-Smith et al., 2020). Libraries were sequenced in a S4 flow cell using a PE150 kit on an Illumina NovaSeq 6000 to a target of 150 million reads.

Single cell RNA sequencing – 10x Genomics is Dual ended

Single Cell 3' v3.1 (Dual Index) Gene Expression Library: Single Cell Gene Expression



Single Cell 5' v2 Gene Expression Library: Single Cell Immune Profiling

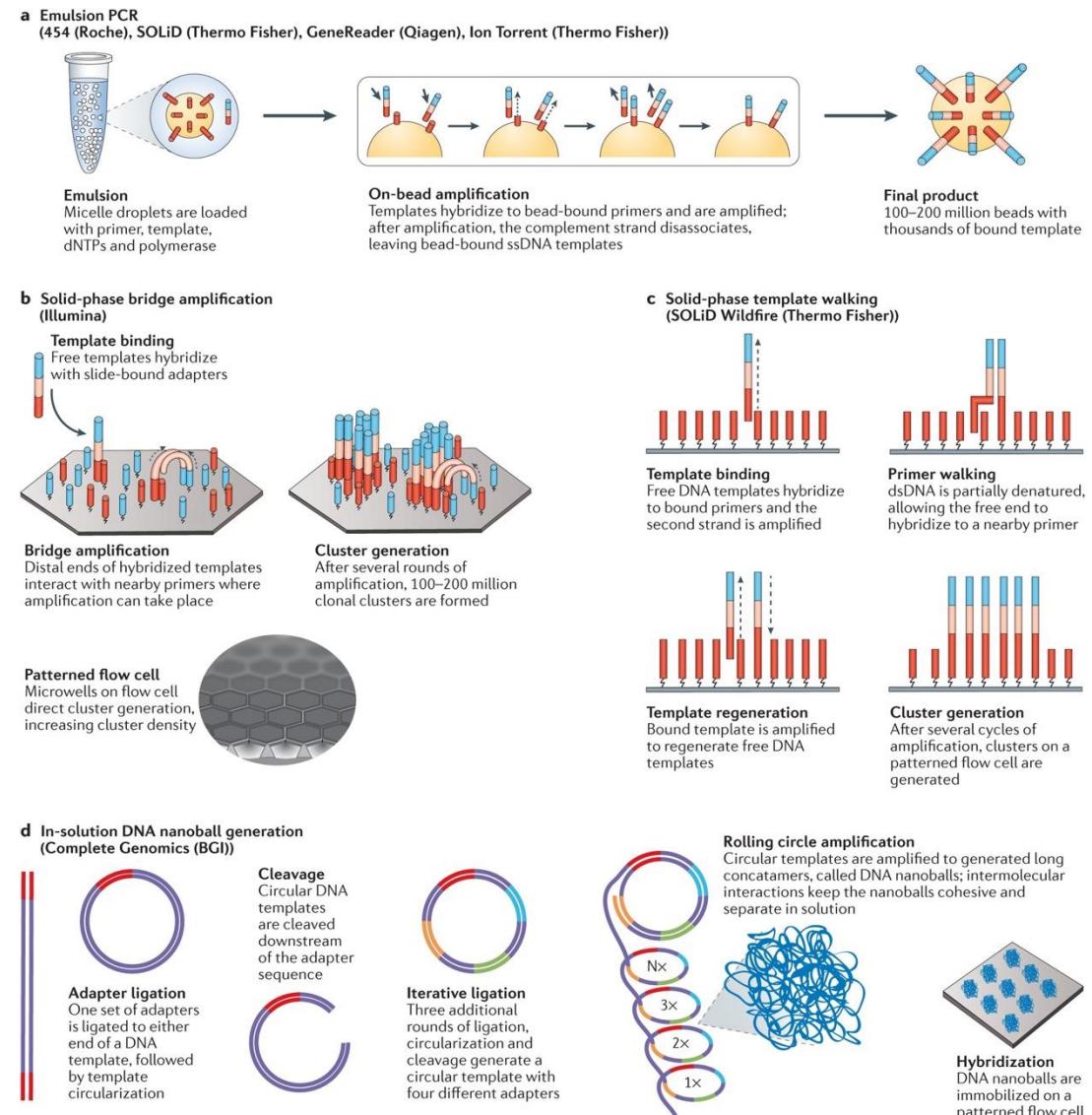


For T cell and B cell
clonotypes

Next Generation Sequencing (NGS)

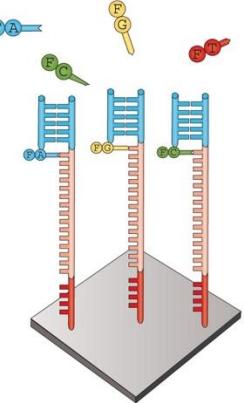
Are either 1) Short read sequencing
or 2) Long read sequencing

Short-read sequencing approaches provide lower-cost, higher-accuracy data that are useful for population-level research and clinical variant discovery. By contrast, long-read approaches provide read lengths that are well suited for de novo genome assembly applications and full-length isoform sequencing.

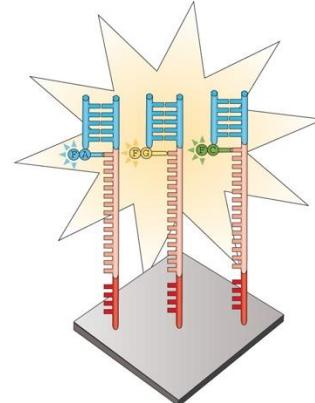


Illumina sequencing

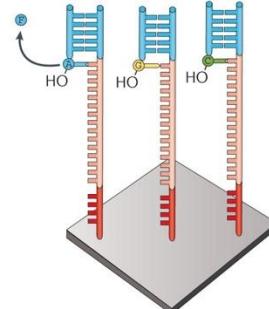
a Illumina



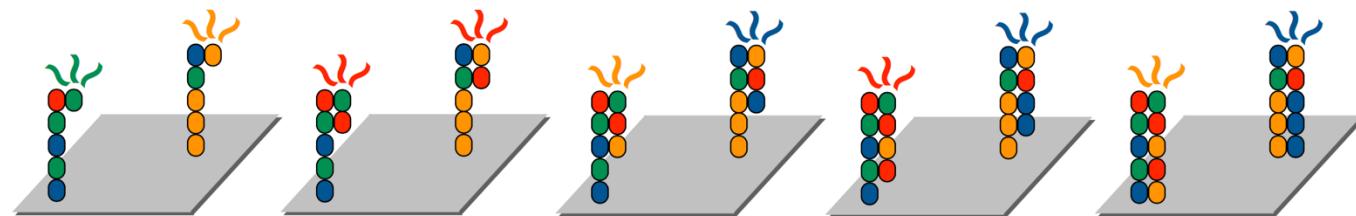
Nucleotide addition
Fluorophore-labelled, terminally blocked nucleotides hybridize to complementary base. Each cluster on a slide can incorporate a different base.



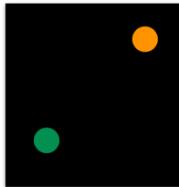
Imaging
Slides are imaged with either two or four laser channels. Each cluster emits a colour corresponding to the base incorporated during this cycle.



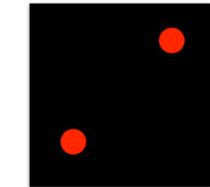
Cleavage
Fluorophores are cleaved and washed from flow cells and the 3'-OH group is regenerated. A new cycle begins with the addition of new nucleotides.



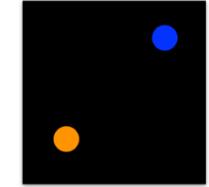
Cycle 1



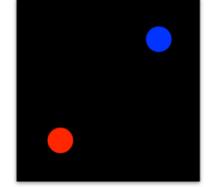
Cycle 2



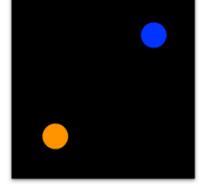
Cycle 3



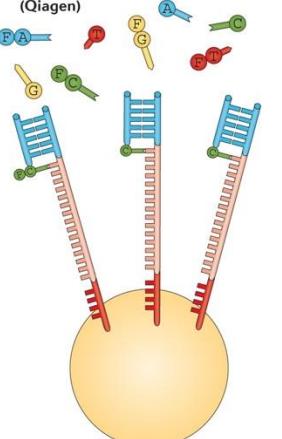
Cycle 4



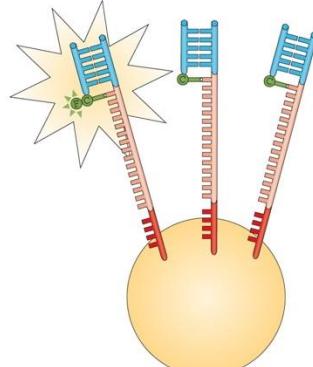
Cycle 5



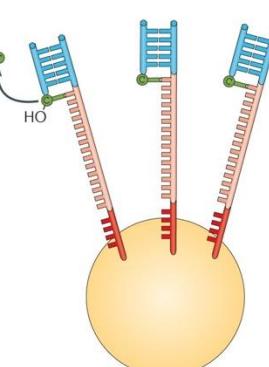
b GeneReader
(Qiagen)



Nucleotide addition
A mixture of fluorophore-labelled, terminally blocked nucleotides and unlabelled, blocked nucleotides hybridize to complementary bases. Each bead on a slide can incorporate a different base.



Imaging
Slides are imaged with four laser channels. Each bead emits a colour corresponding to the base incorporated during this cycle, but only labelled bases emit a signal.



Cleavage
Fluorophores are cleaved and washed from flow cells and the 3'-OH group is regenerated. A new cycle begins with the addition of new nucleotides.

Sequencing Read depth

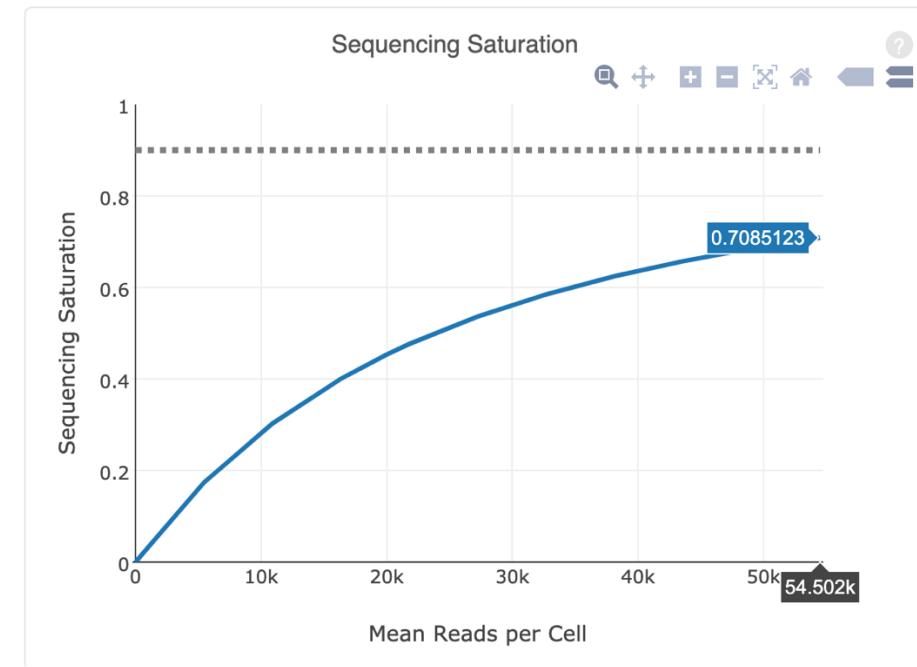
Read depth is defined as total number of reads per sample or per cell desired after sequencing for your experimental purpose

Most experiments require 5 million to 200 million reads per sample (bulk) and 30,000-150,000 reads per cell (single cell)

More read depth = More cost!

Typically for RNAseq experiments, you can do shallow sequencing in most cases

Sequencing depth can be measured by sequencing saturation metric (You want to reach 90% saturation limit)



```
> unique_confidently_mapped_reads = 10,196,940  
duplicate_reads = 24,785,461  
  
x = 1 - (unique_confidently_mapped_reads/(unique_confidently_mapped_reads + duplicate_reads))  
x = 1 - (10,196,940/(10,196,940 + 24,785,461))  
x = 1 - (10,196,940/34,982,401)  
x = 1 - 0.29148771  
x = 0.70851229
```

Illumina sequencers



Miseq

~20M of reads total



Nextseq

~500M of reads total



HiSeq 4000

~4 billion of reads total

How to plan your RNA sequencing experiment

What is the biological question?

What type of sample do you have as input material?

How many replicates do I need?

What library preparation protocol should I use?

How much sequencing depth do I need?

Do I need single-end or dual-end sequencing?

What sequencing platform is available to me?

How do I eliminate batch effect?

What is my budget?



Suggested Readings and Learning links

Gresham, D., Dunham, M. & Botstein, D. Comparing whole genomes using DNA microarrays. *Nat Rev Genet* 9, 291–302 (2008).
<https://doi.org/10.1038/nrg2335>

Stark, R., Grzelak, M. & Hadfield, J. RNA sequencing: the teenage years. *Nat Rev Genet* 20, 631–656 (2019). <https://doi.org/10.1038/s41576-019-0150-2>

Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10, 57–63 (2009). <https://doi.org/10.1038/nrg2484>

Thind AS, Monga I, Thakur PK, Kumari P, Dindhoria K, Krzak M, Ranson M, Ashford B. Demystifying emerging bulk RNA-Seq applications: the application and utility of bioinformatic methodology. *Brief Bioinform.* 2021 Nov 5;22(6):bbab259. doi: 10.1093/bib/bbab259. PMID: 34329375.

Chen JW, Shrestha L, Green G, Leier A, Marquez-Lago TT. The hitchhikers' guide to RNA sequencing and functional analysis. *Brief Bioinform.* 2023 Jan 19;24(1):bbac529. doi: 10.1093/bib/bbac529. PMID: 36617463; PMCID: PMC9851315.

Conte MI, Fuentes-Trillo A, Domínguez Conde C. Opportunities and tradeoffs in single-cell transcriptomic technologies. *Trends Genet.* 2024 Jan;40(1):83–93. doi: 10.1016/j.tig.2023.10.003. Epub 2023 Nov 10. PMID: 37953195.

De Jonghe J, Opzoomer JW, Vilas-Zornoza A, Nilges BS, Crane P, Vicari M, Lee H, Lara-Astiaso D, Gross T, Morf J, Schneider K, Cudini J, Ramos-Mucci L, Mooijman D, Tiklová K, Salas SM, Langseth CM, Kashikar ND; scTrends Consortium; Schapiro D, Lundeberg J, Nilsson M, Shalek AK, Cribbs AP, Taylor-King JP. scTrends: A living review of commercial single-cell and spatial 'omic technologies. *Cell Genom.* 2024 Dec 11;4(12):100723. doi: 10.1016/j.xgen.2024.100723. PMID: 39667347; PMCID: PMC11701258.

Satam H, Joshi K, Mangrolia U, Waghoo S, Zaidi G, Rawool S, Thakare RP, Banday S, Mishra AK, Das G, Malonia SK. Correction: Satam et al. Next-Generation Sequencing Technology: Current Trends and Advancements. *Biology* 2023, 12, 997. *Biology (Basel)*. 2024 Apr 24;13(5):286. doi: 10.3390/biology13050286. Erratum for: *Biology (Basel)*. 2023 Jul 13;12(7):997. doi: 10.3390/biology12070997. PMID: 38785841; PMCID: PMC11107263.

<https://data-science-sequencing.github.io/Win2018/lectures/lecture2/>

<https://www.cd-genomics.com/resource-single-read-vs-paired-end-sequencing.html>