

INTRODUCTION TO SINGLE CELL RNA-SEQ

Analysis of single cell RNA-seq data

Katarzyna Kania

9th February 2026

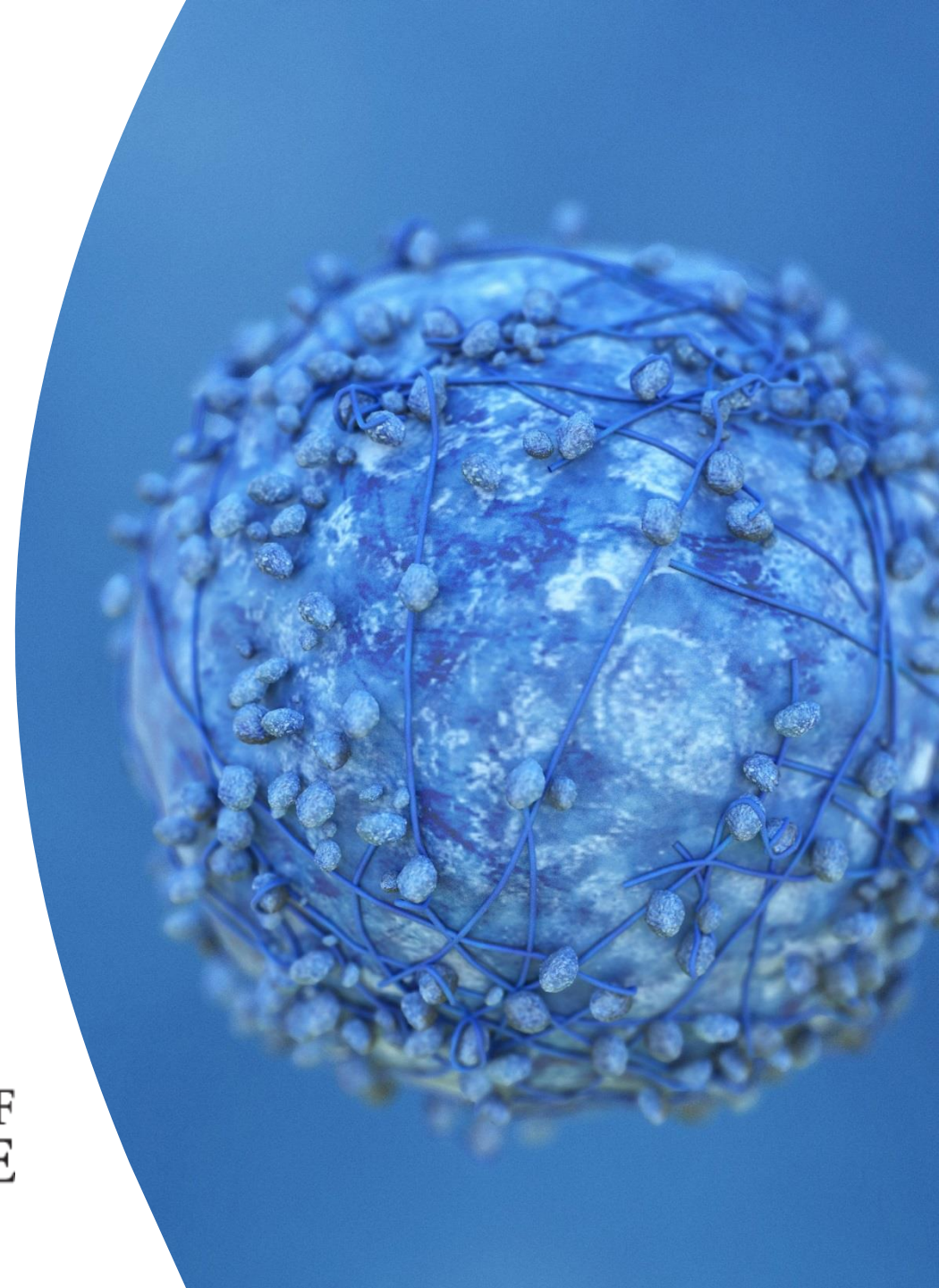


CAMBRIDGE
INSTITUTE

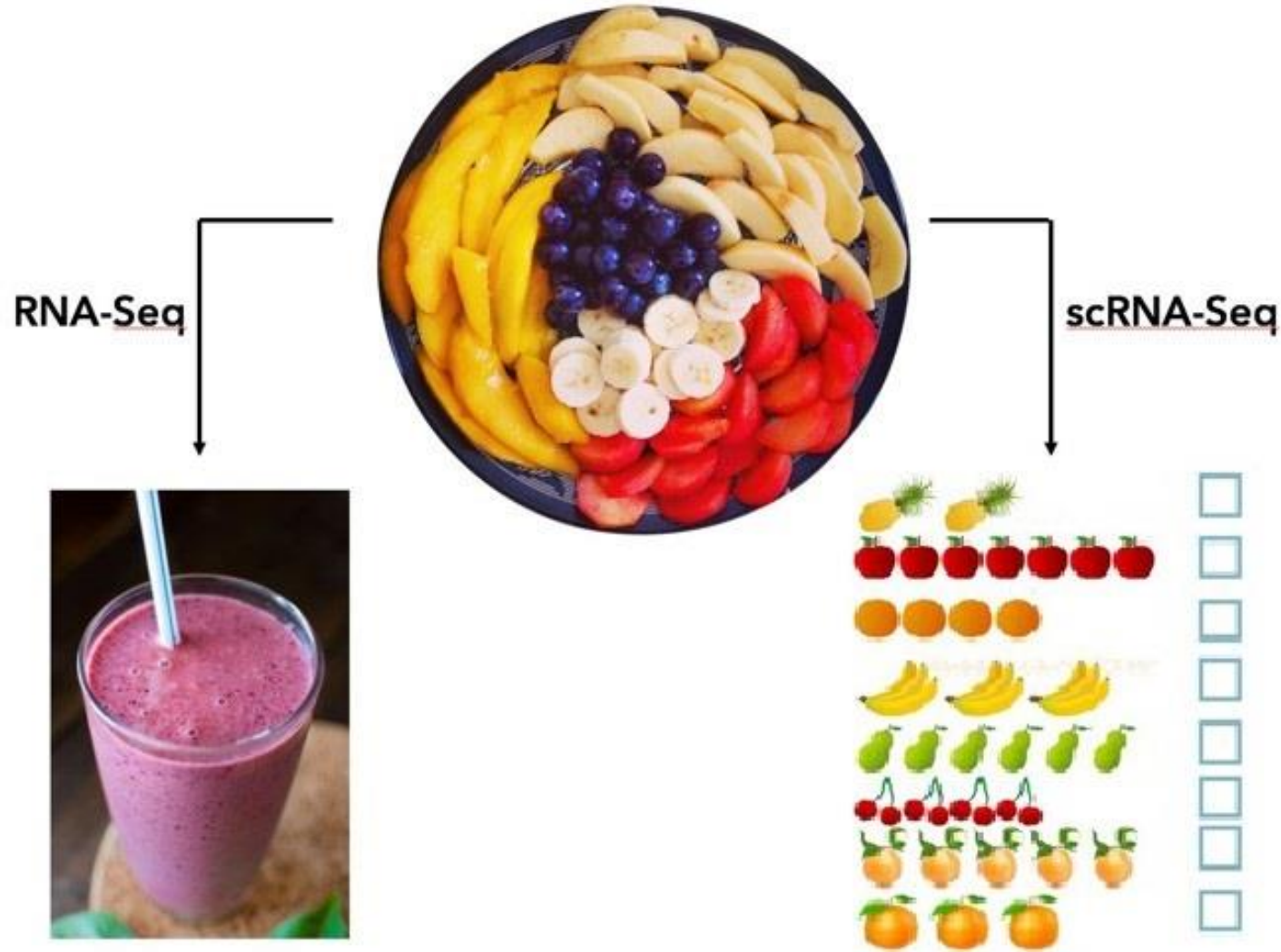


UNIVERSITY OF
CAMBRIDGE

Together we will beat cancer



BULK VS SINGLE CELL RNA-SEQ

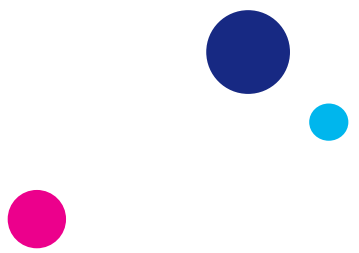


Separate populations

- Define heterogeneity
- Identify rare cell populations
- Cell population dynamics

Average expression level

- Comparative transcriptomics
- Disease biomarker
- Homogenous systems



BULK VS SINGLE CELL RNA-SEQ

1. mRNA: TruSeq RNA-Seq (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

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)
- If you only looked at transcripts observed in all cells numbers drop dramatically.

BULK VS SINGLE CELL RNA-SEQ

	Deep RNA-seq	Sort-seq	Low input	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

Disadvantages of scRNA-seq

- Dropouts and noisy data
- Lowly expressed genes might be undetected
- Samples will contain doublets
- Replicates without batch effect are unlikely
- Expensive

Source: Sarah Boswell, Harvard Medical School, September 2020

APPLICATIONS

naturemedicine

Letter | Published: 08 June 2020

A single-cell atlas of the peripheral immune response in patients with severe COVID-19

Aaron J. Wilk, Arjun Rustagi, Nancy Q. Zhao, Jonasel Roque, Giovanni J. Martínez-Colón, Julia L. McKechnie, Geoffrey T. Ivison, Thanmayi Ranganath, Rosemary Vergara,

LETTER

<https://doi.org/10.1038/s41586-018-0394-6>

A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte

Lindsey W. Plasschaert^{1,3,7}, Rapolas Zilionis^{2,3,7}, Rayman Choo-Wing^{1,5}, Virginia Savova^{2,6}, Judith Knehr⁴, Guglielmo Roma⁴, Allon M. Klein^{2*} & Aron B. Jaffe^{1,3*}

nature

Article | Published: 20 February 2019

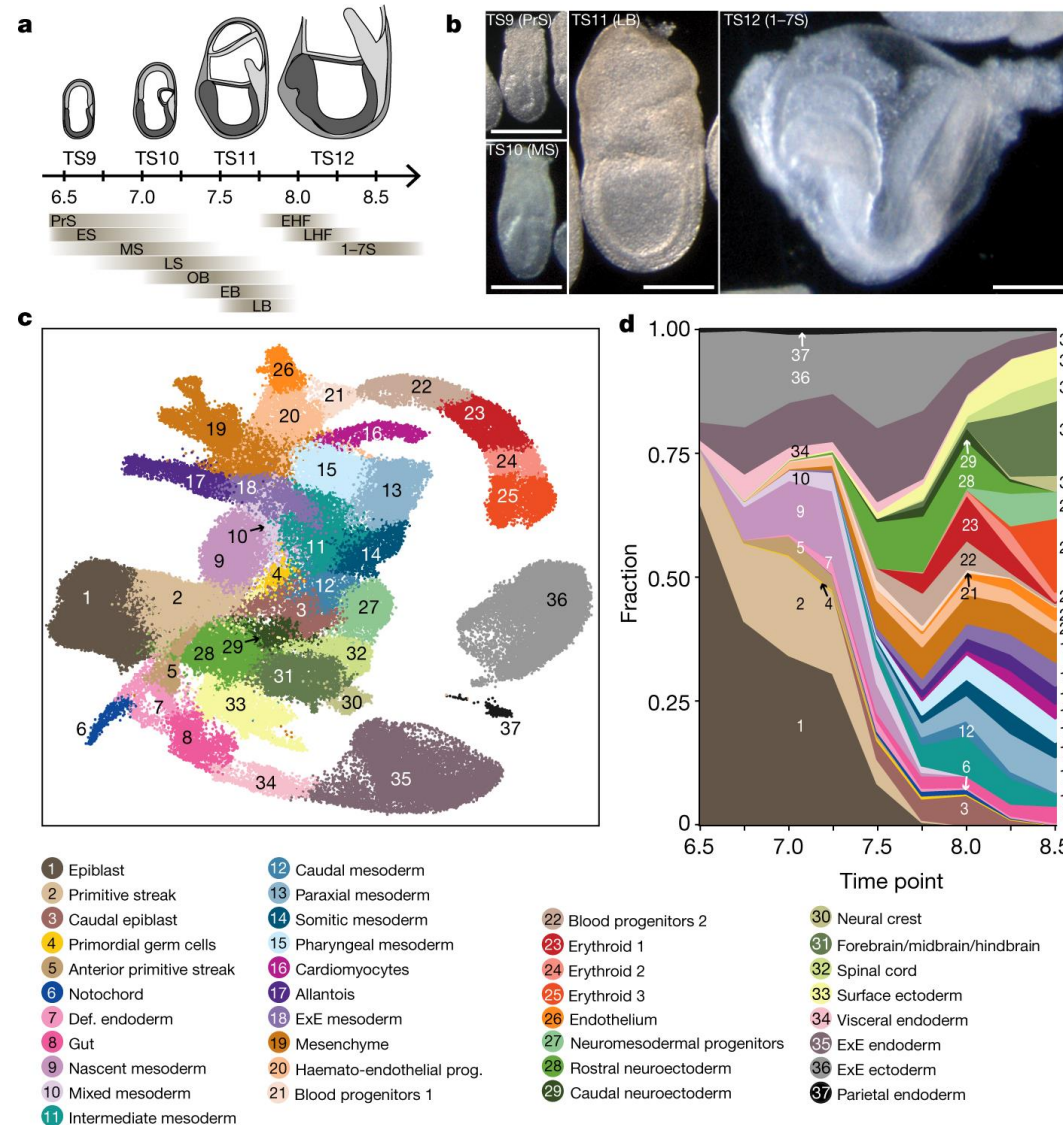
A single-cell molecular map of mouse gastrulation and early organogenesis

Blanca Pijuan-Sala, Jonathan A. Griffiths, Carolina Guibentif, Tom W. Hiscock, Wajid Jawaid, Fernando J. Calero-Nieto, Carla Mulas, Ximena Ibarra-Soria, Richard C. V.



CANCER
RESEARCH
UK

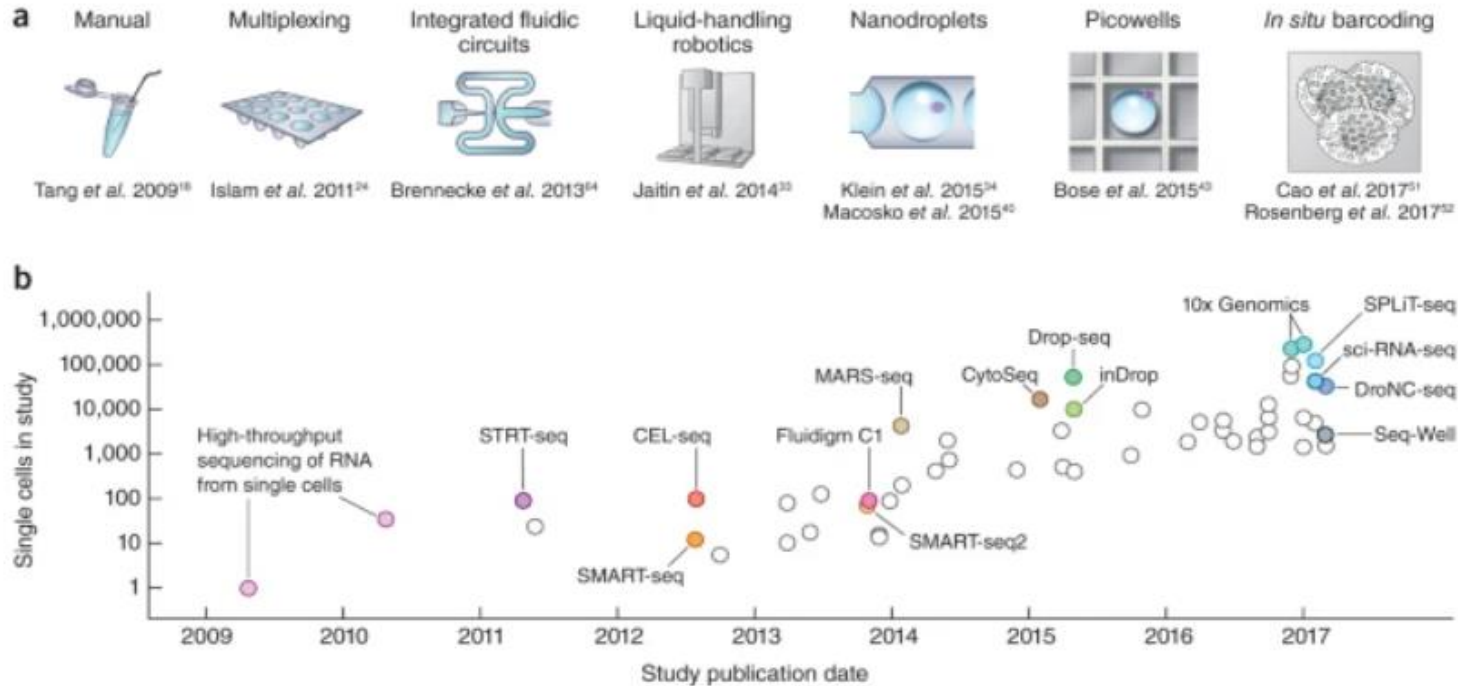
CAMBRIDGE
INSTITUTE



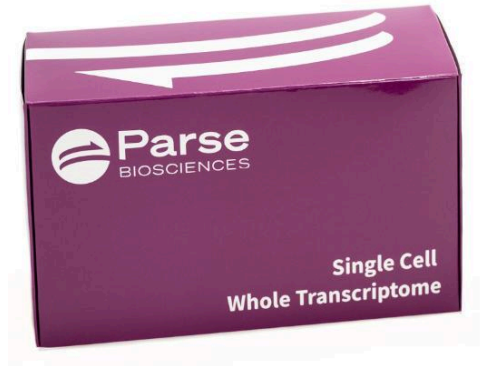
Source: Pijuan-Sala et al. Nature 566, 490–495 (2019)

TECHNOLOGIES

Figure 1: Scaling of scRNA-seq experiments.



Source: Svensson et al. *Nat Protoc* 13, 599–604 (2018)



Parse WT Penta –
5mln cells in one
experiment

Chromium X



Making 1 million cell experiments routine



CANCER
RESEARCH
UK

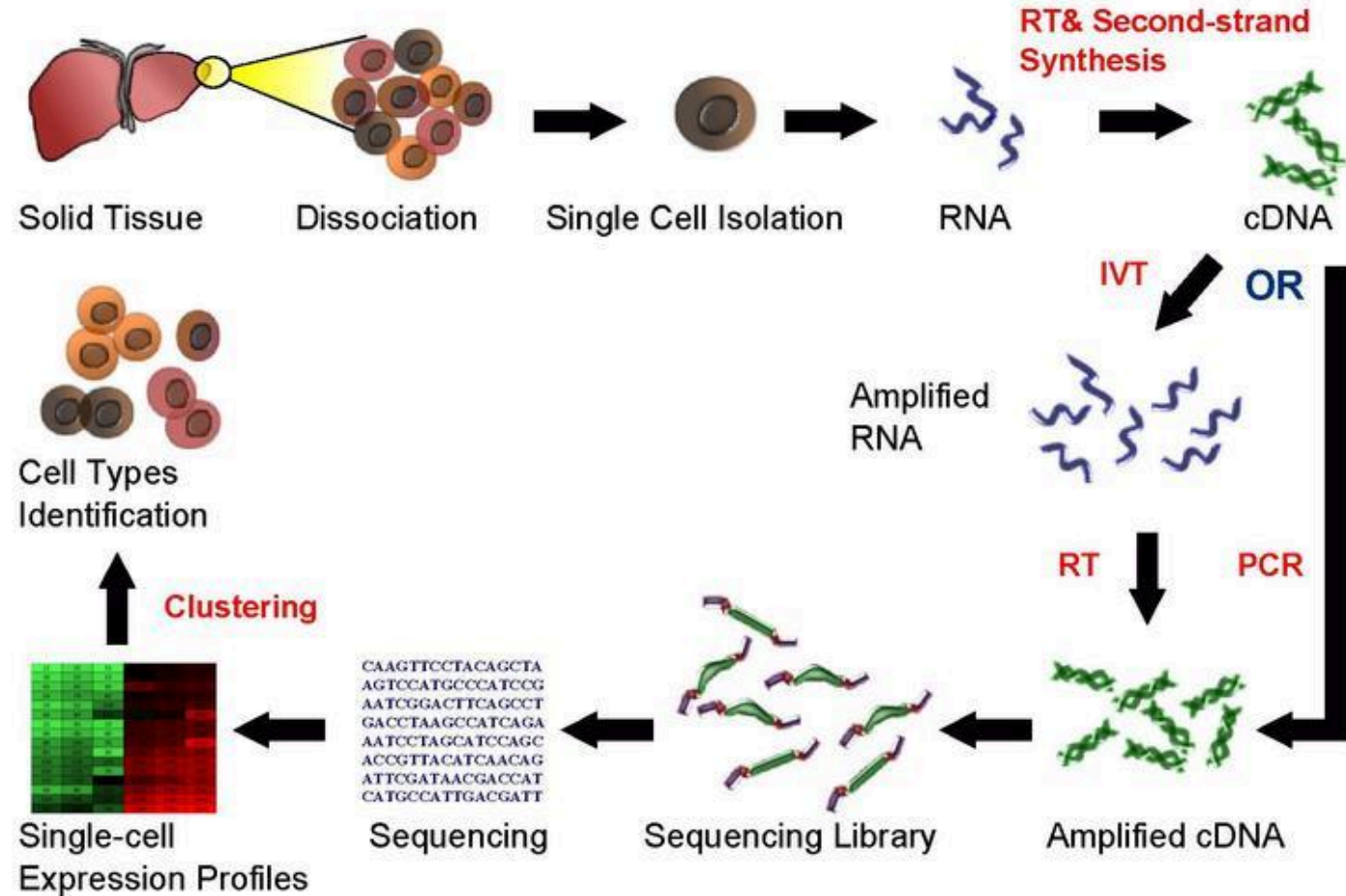
CAMBRIDGE
INSTITUTE

WORKFLOW

Single Cell RNA Sequencing Workflow

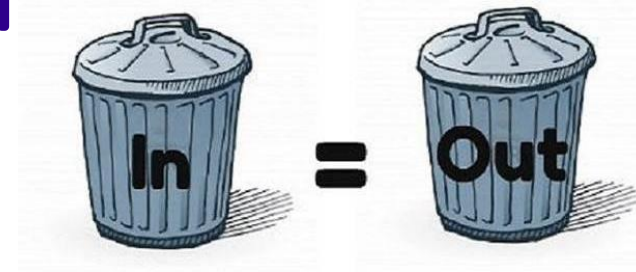


Good sample preparation is key to success!



Source: https://en.wikipedia.org/wiki/Single_cell_sequencing

SAMPLE PREPARATION

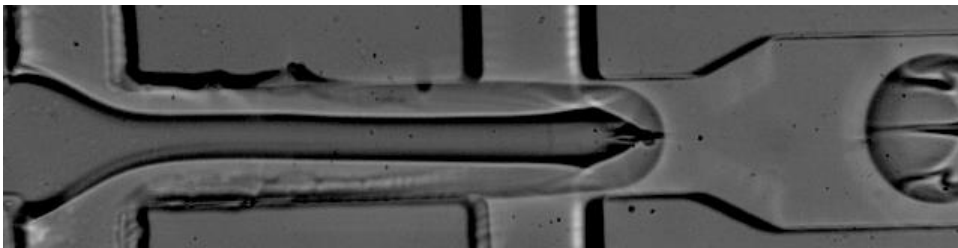


- Understand well the nature of the sample (sampling conditions, preparation, purity)
- Identify the source of technical difficulties in order to resolve them first
- Practice your sample preparation, optimise the protocol well, do not rush to the final experiment
- A well planned pilot experiment is essential for evaluating sample preparation and for understanding the required number of cells.
- You need your cells to be highly viable (>90-95%), have no clumps and no debris. Cell membrane integrity is a must!
- Free-floating RNA will make analysis more challenging
- Be cautious about FACS (especially with more fragile cells). If FACS necessary for enrichment, remember that time is crucial factor
- Count with haemocytometer or cell counter (Countess II Automated Cell Counter) – do not trust sorter counts
- Fixation and cryopreservation are not compatible with many techniques



DROP-SEQ OVERVIEW

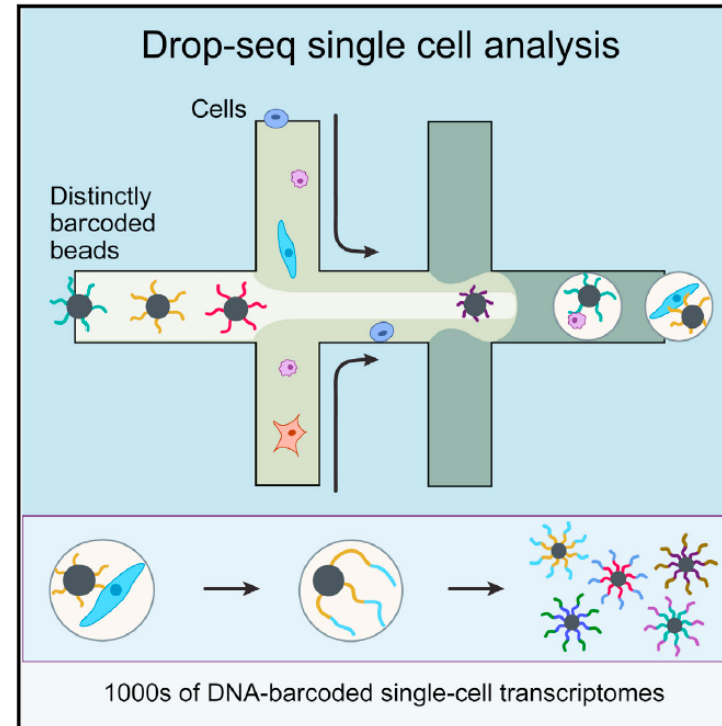
- Moved throughput from hundreds to thousands.
- Droplet-based processing using microfluidics
- Nanoliter scale aqueous drops in oil.
- 3' End
- Bead based (STAMPs).
- Single-cell transcriptomes attached to microparticles.
- Cell barcodes use split-pool synthesis.
- Uses UMI (Unique Molecular Identifier)
- Chance to have two cells within one droplet



Cell

Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets

Graphical Abstract



Resource

Authors

Evan Z. Macosko, Anindita Basu, ...,
Aviv Regev, Steven A. McCarroll

Correspondence

emacosko@genetics.med.harvard.edu
(E.Z.M.),
mccarroll@genetics.med.harvard.edu
(S.A.M.)

In Brief

Capturing single cells along with sets of uniquely barcoded primer beads together in tiny droplets enables large-scale, highly parallel single-cell transcriptomics. Applying this analysis to cells in mouse retinal tissue revealed transcriptionally distinct cell populations along with molecular markers of each type.

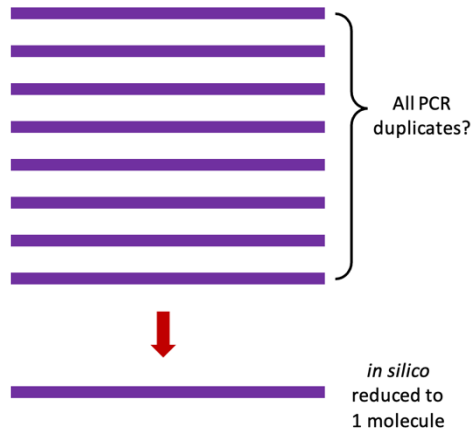
UMI – UNIQUE MOLECULAR IDENTIFIERS

After PCR enrichment, without UMIs, one can not distinguish if multiple copies of a fragment are caused by PCR clones or if they are real biological duplicated.

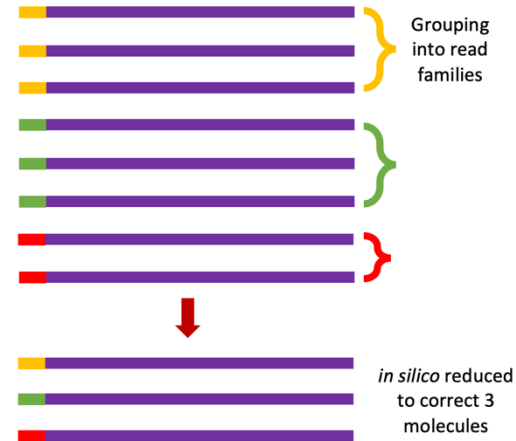
By using UMIs, PCR clones can be found by searching for non-unique fragment-UMI combinations, which can only be explained by PCR clones.

When performing variant analyses, these falsely overrepresented fragments can result in incorrect calls and thus wrong diagnostic findings

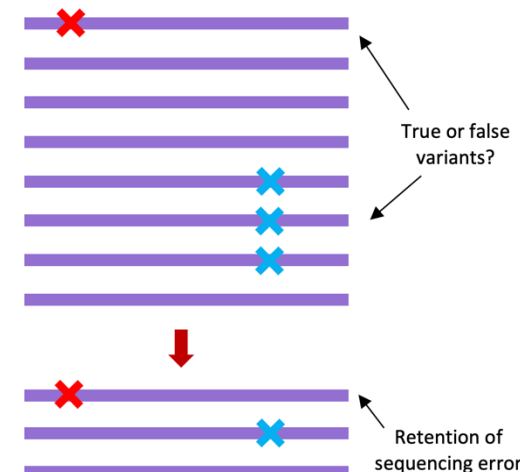
PCR duplicate removal **without** UMIs



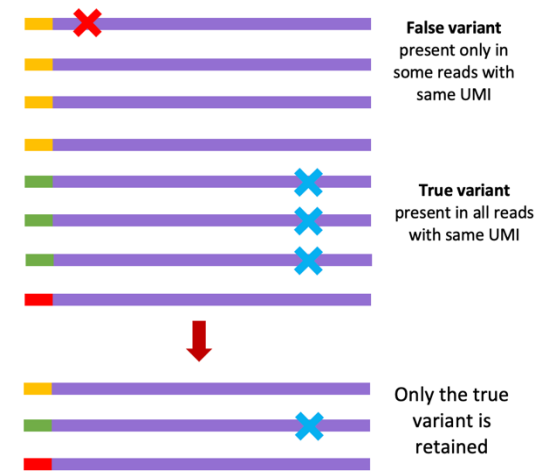
PCR duplicate removal **with** UMIs



Variant calling **without** UMIs

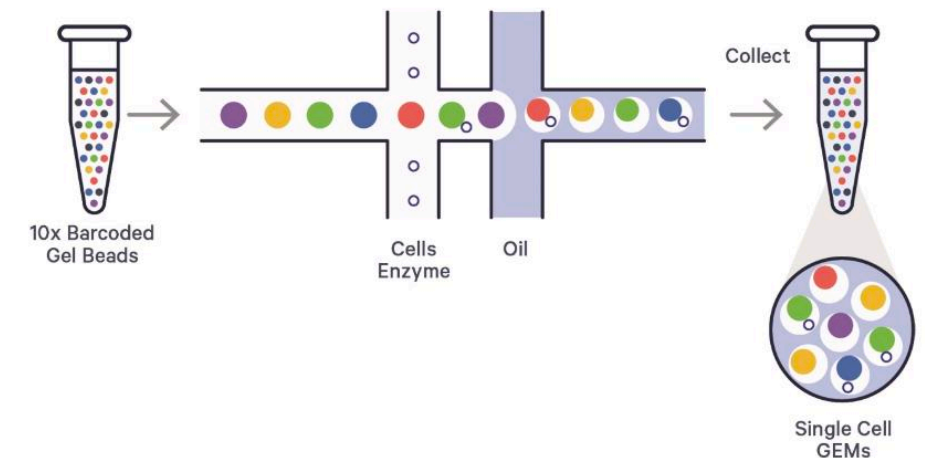
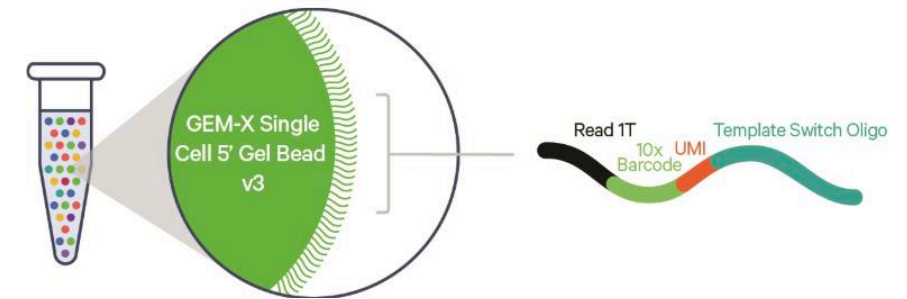
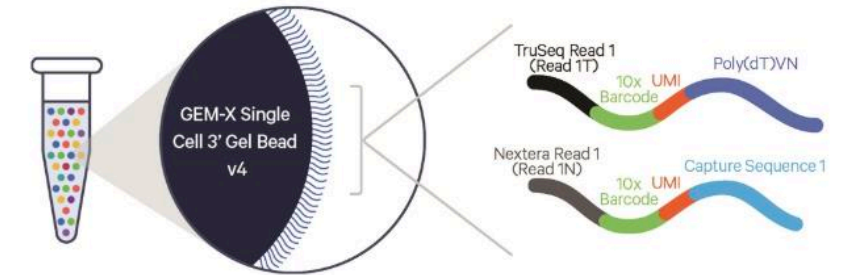


Variant calling **with** UMIs

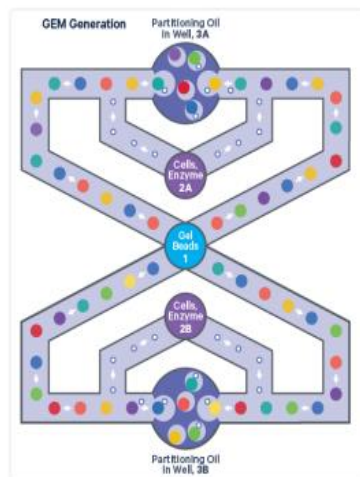


10X GENOMICS OVERVIEW

- Droplet-based similar to Drop-Seq, 3' or 5' mRNA
- In contrast to Drop-seq, where solid beads are used for RNA capture, 10X uses soft hydrogels containing oligos. These enable “single Poisson loading” leading to capture of >60% of input cells.
- Standardized instrumentation and reagents (unhackable so no customisation or control)
- Very easy to use and less processing time
- More high-throughput scaling - 16 samples can be processed simultaneously with up to 20000 cells captured per sample
- The doublet rate increases with number of cells loaded
- CellRanger and CellLoupe software are available and user friendly
- A lot of additional addons possible (CRISPR, TCR & BCR profiling etc.)



10X GENOMICS OVERVIEW

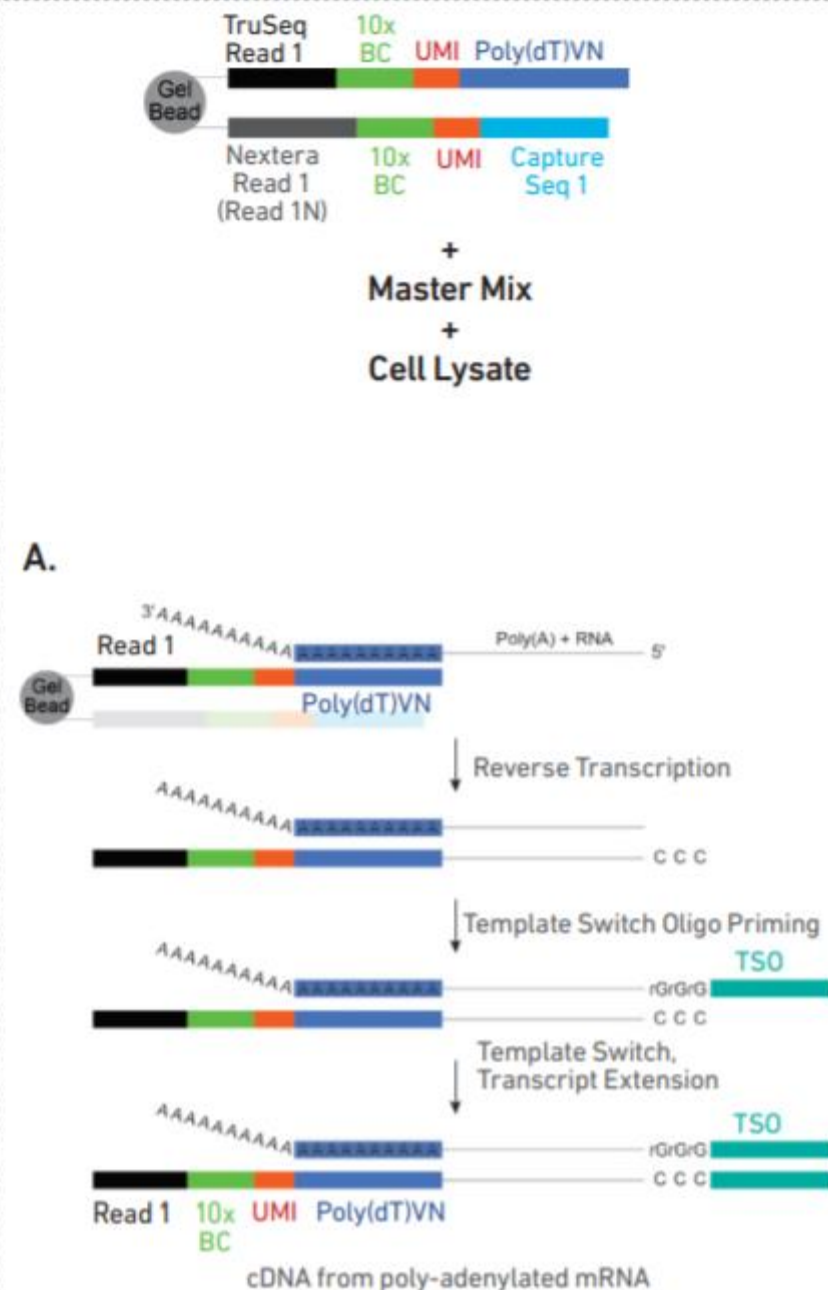


Chromium X



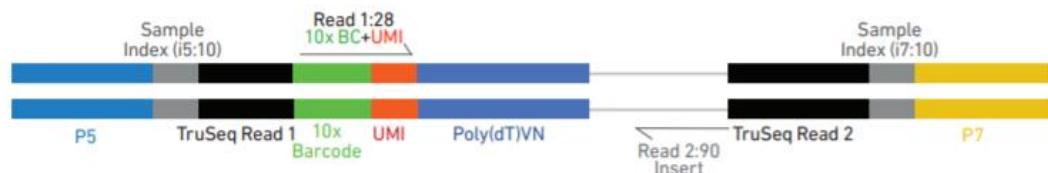
Making 1 million cell experiments routine

Inside individual GEMs

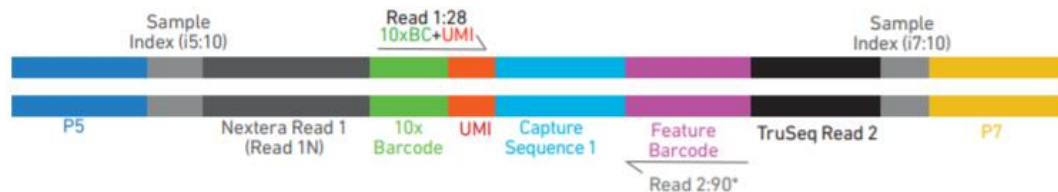


10X GENOMICS LIBRARIES

Chromium Single Cell 3' Gene Expression Dual Index Library



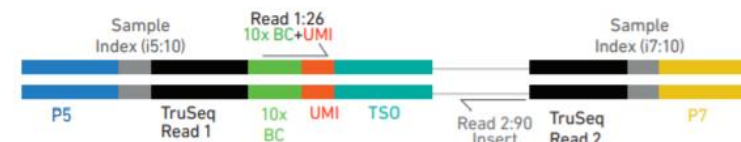
Chromium Single Cell 3' Cell Surface Protein Dual Index Library



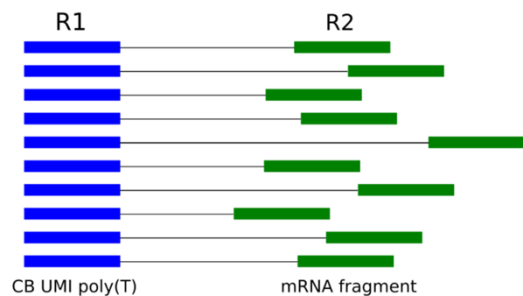
Chromium Single Cell V(D)J Dual Index Library



Chromium Single Cell 5' Gene Expression Dual Index Library

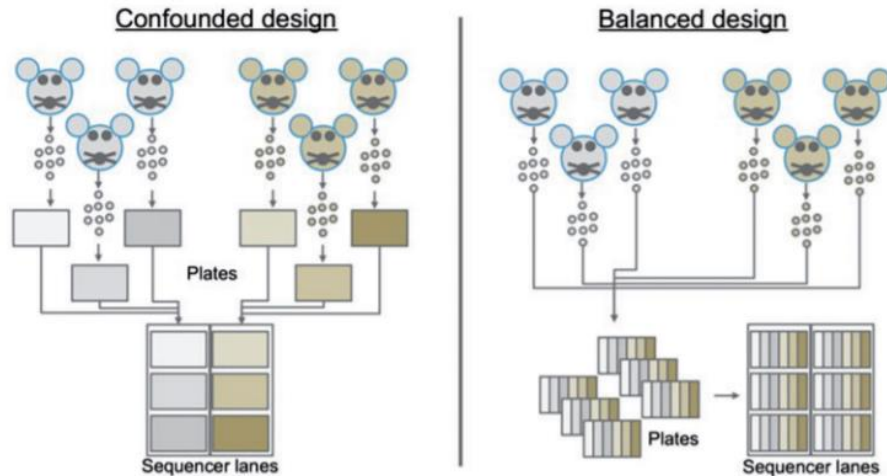


Source: 10x Genomics



Sequencing Read	Description	Number of cycles
Read1	10x Barcode Read (Cell) + Randomer Read (UMI)	28bp
i7 index	Sample index read	10bp
i5 index	Sample index read	10bp
Read2	Insert Read (Transcript)	90bp

EXPERIMENTAL DESIGN



Source: Baran-Gale et al. Brief Func Genomics. 17 (4):233–239. (2018)

I. Tissue Procurement



Source:

- Primary human
- Model organism
- Cell culture

Key considerations:

- Biological variation
- Sampling/handling variation
- Duration of sourcing

Study design:

- Biological replicates
- Technical replicates
- Cell number calculation
- Workflow optimization

II. Tissue Dissociation



Method:

- Mechanical mincing
- Enzymatic digestion
- Automated blending
- Microfluidics devices

Key considerations:

- Experimental consistency
- Shortest duration
- Highest cell/nucleus quality
- Representation of all cell types

Quality control:

- FACS analysis
- qPCR for marker genes
- Imaging of cell integrity
- RNA quality (RIN)

III. Cell Enrichment (optional)



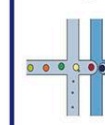
Method:

- Differential centrifugation, sedimentation, filtration
- Antibody labeling for positive/negative selection
- Flow cytometry or bead-based enrichment
- Dead cell removal

Key considerations:

- Additional handling
- Longer duration
- Loss of RNA quality
- Transcriptome changes

IV. Single Cell RNAseq Platform



Method:

- Droplet-based
- Tube-based after FACS
- Microwell-based
- Microfluidics-enabled

Key considerations:

- Cell throughput and handling time
- Gene coverage and cell type detection
- Whole transcript versus 3' end counting
- Imaging capability for doublet detection

V. Library Sequencing



Method:

- Illumina NGS
- Compatible with cDNA library

Sequencing depth considerations:

- 3' end counting: low depth ~50K RPC
- Whole transcript: high depth ~1M RPC
- Alternative splicing: ~20-30M RPC
- Iterative optimization for biological system

VI. Computational Analysis



Key considerations:

- Separation of batch and condition
- Technical vs. biological variation

Sample Batch correction approaches:

- Cell Hashing
- Demuxlet
- Canonical correlation analysis (CCA)
- MAST

Source: Nguyen QH et al. Front Cell Dev Biol 6:108. (2018)