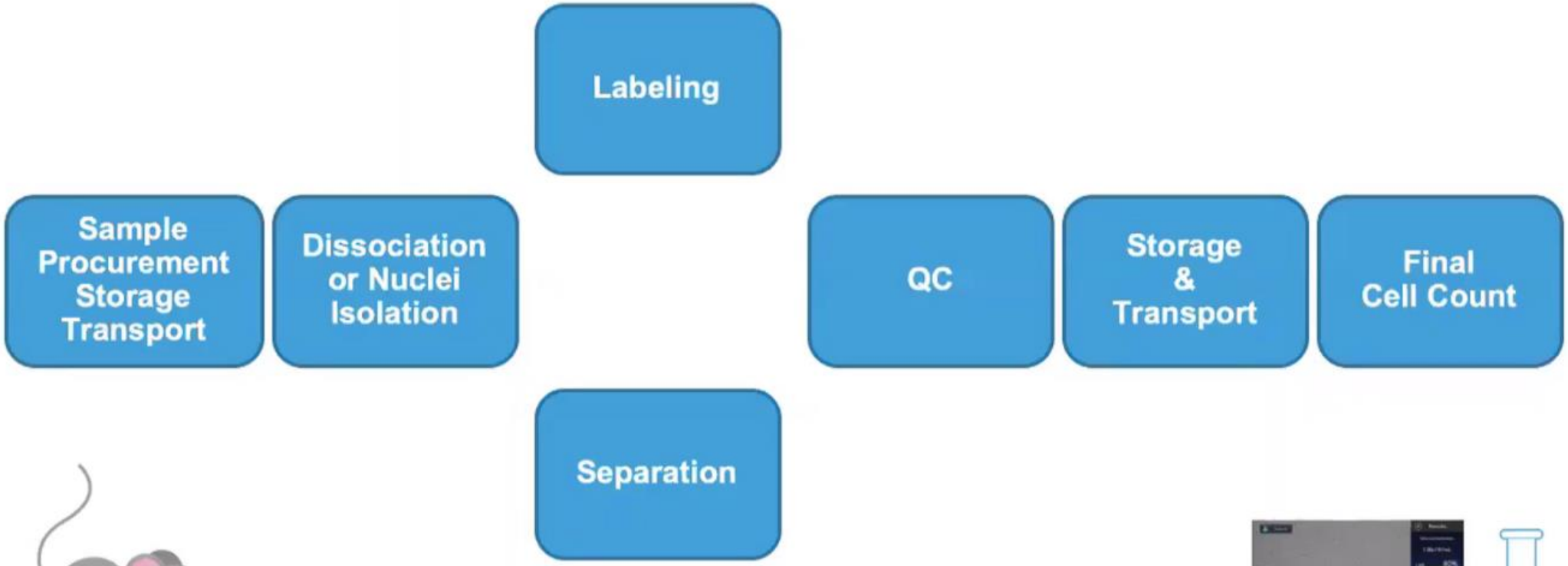
Sample Prep and Experimental Design

Susan Kloet

Leiden Genome Technology Center

24 October 2023

What is sample prep?

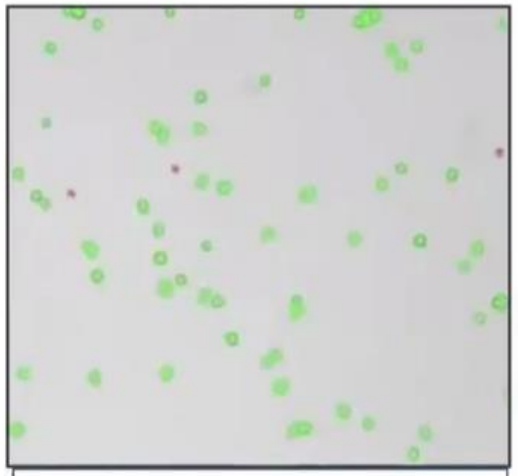
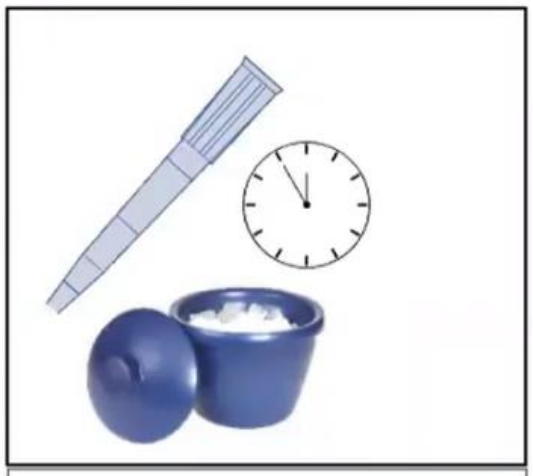


Know what you are working with

Cells Tested	Species	Cell Source	Total RNA (pg/cell)	Cell Size (µm)
PBMC	Human	Extracted from blood	~0.75	~5-10
E18 neuron	Mouse	Brain tissue	~2-3	~9
Jurkat	Human	Suspension	5.5	~12
Raji	Human	Suspension	7.3	~12
293T	Human	Adherent	14.2	~18
3T3	Mouse	Adherent	16.1	~18
HCC1954	Human	Adherent	15.7	~18
HCC38	Human	Adherent	21.6	~30

Quality is critical

Goal is to minimize

Clean	Healthy	Gentle
		
• Aggregates/clumps • Subcellular debris • Free-floating RNA/DNA	• RNA leakage • RNA degradation	• Physical decomposition

Handle with care

Minimize handling, both pipetting and centrifugation

Metric	Control	Wide Bore (Rough)	Narrow Bore (Rough)	Vortex 5s
Number of Cells	1,118	846	1,012	983
Reads per Cell	50,000	50,000	50,000	50,000
Valid Barcodes	95.40%	95.50%	95.30%	95.50%
Fraction Reads in Cells	79.40%	72.80%	54.00%	63.10%
Reads Mapped Confidently to Transcriptome	70.50%	71.40%	71.80%	71.00%
Valid UMIs	99.40%	99.40%	99.40%	99.40%
Median Genes per Cell	3,137	3,180	2,833	2,934

Sample procurement, storage, transport

CNV

Cultured cells, PBMCs
*Fresh and cryopreserved
*10x protocols work as-is

Fresh tissue
*Minimal optimization

Snap-frozen tissue
*Minimal optimization

Fixation?
*Not tested but likely (ethanol or methanol)

ATAC

Cultured cells, PBMC
*Fresh and cryopreserved
*10x protocols work as-is

Fresh tissue
*Optimization

Snap-frozen tissue
*Optimization

Fixation?
*Not tested

GEX

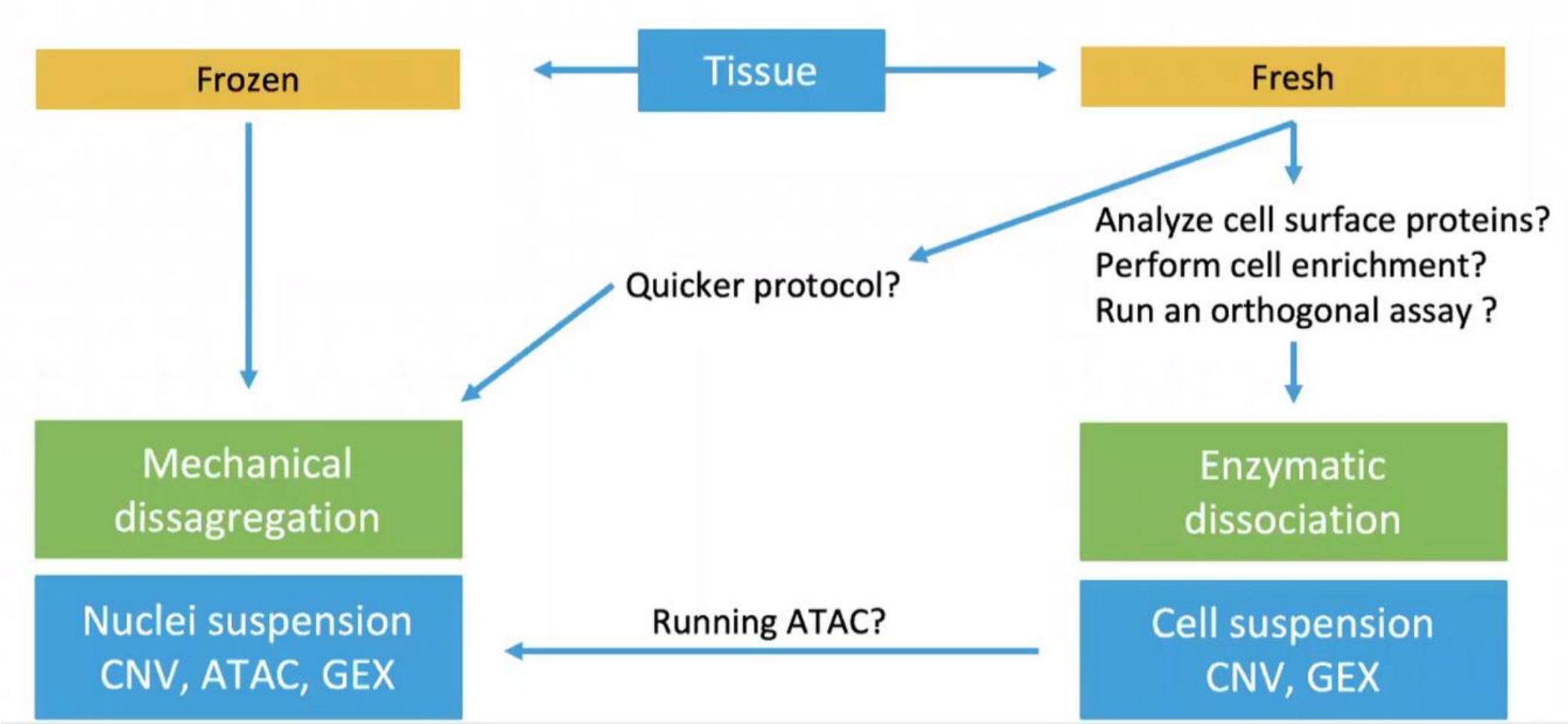
Cultured cells, PBMC
*Fresh and cryopreserved
*10x protocols work as-is

Fresh tissue
*Significant optimization

Snap-frozen tissue
*Nuclei (no cell surface proteins)
*Technically challenging
*Significant optimization

Fixation?
*Methanol (but we prefer cryopreservation)

How to choose a dissociation method



Separation

Separate intact cells and nuclei from

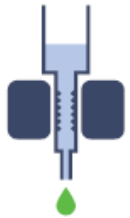
- Aggregates/clumps
- Debris
- Free-floating mRNA
- Dead cells
- Enrichment/depletion

Challenges with separation

- Samples are fragile
- Physical stress
- Buffers
- Time
- Yield

Want the minimum handling necessary. Maintain sample integrity.

There are many separation methods

	Standard Cleanup Methods				Advanced Cleanup Methods	
	Filtering	Washes	Slow Centrifugation	Density Gradient	Dead Cell Removal	Cell Sorting
						
Large Debris (>30 µm)	✓	✓	⊘	✓	⊘	✓
Small Debris (<30 µm)	⊘	✓	✓	⊘	⊘	✓
Dead Cells	⊘	⊘	⊘	⊘	✓	✓
Cell Aggregates	✓	⊘	⊘	✓	⊘	✓

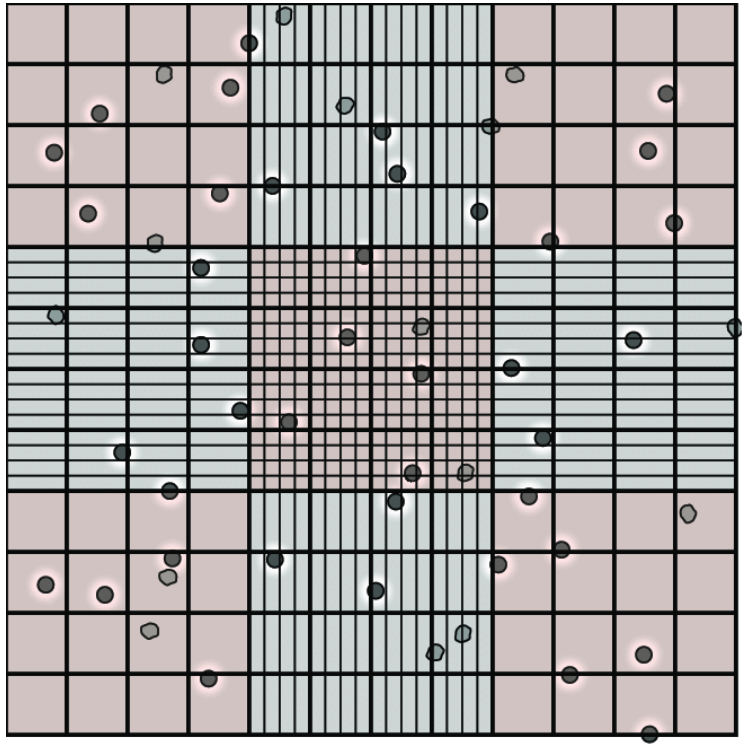
Washing / straining

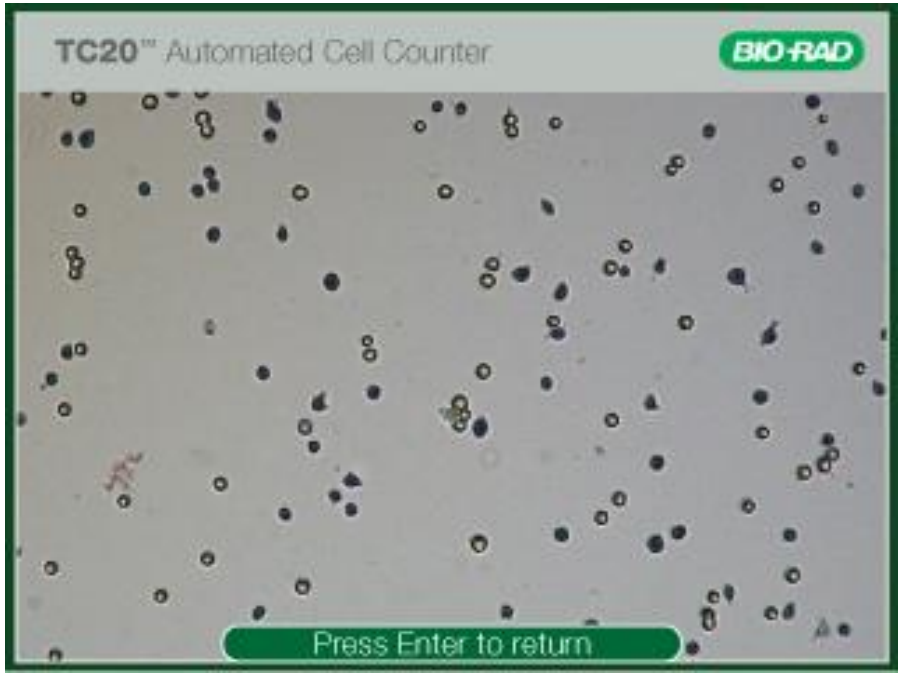
- 3 washes in PBS + 0.04% BSA before loading
- Use a relatively large volume for washes relative to cell number
  - Concentrated cells can clump
- Large cell aggregates or debris can increase the risk of clogs

  => Use a cell strainer, 30 to 40 µm

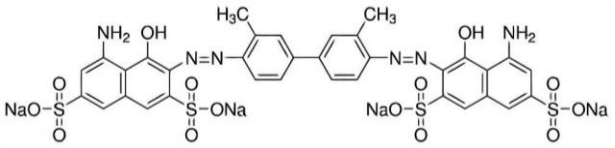
  BUT – this will result in sample loss, volume loss, and possible cell loss

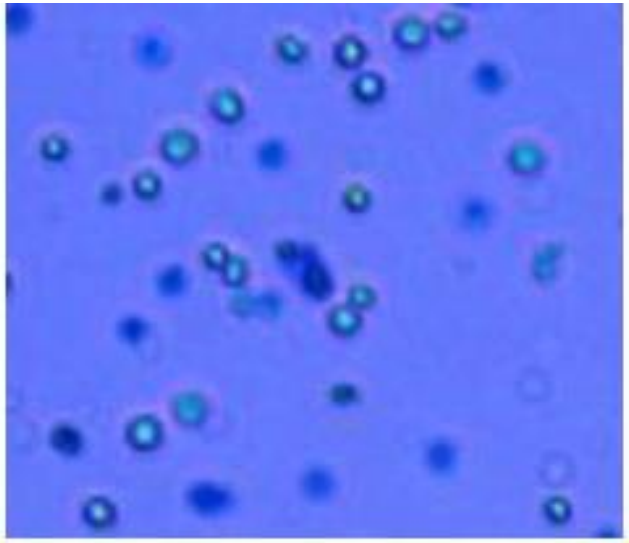
Counting and viability assessment

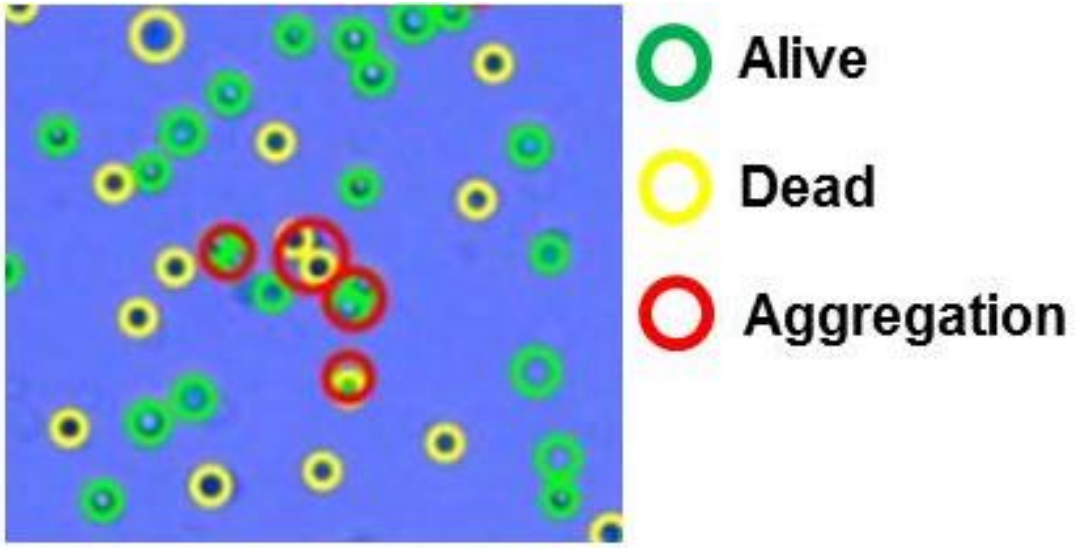




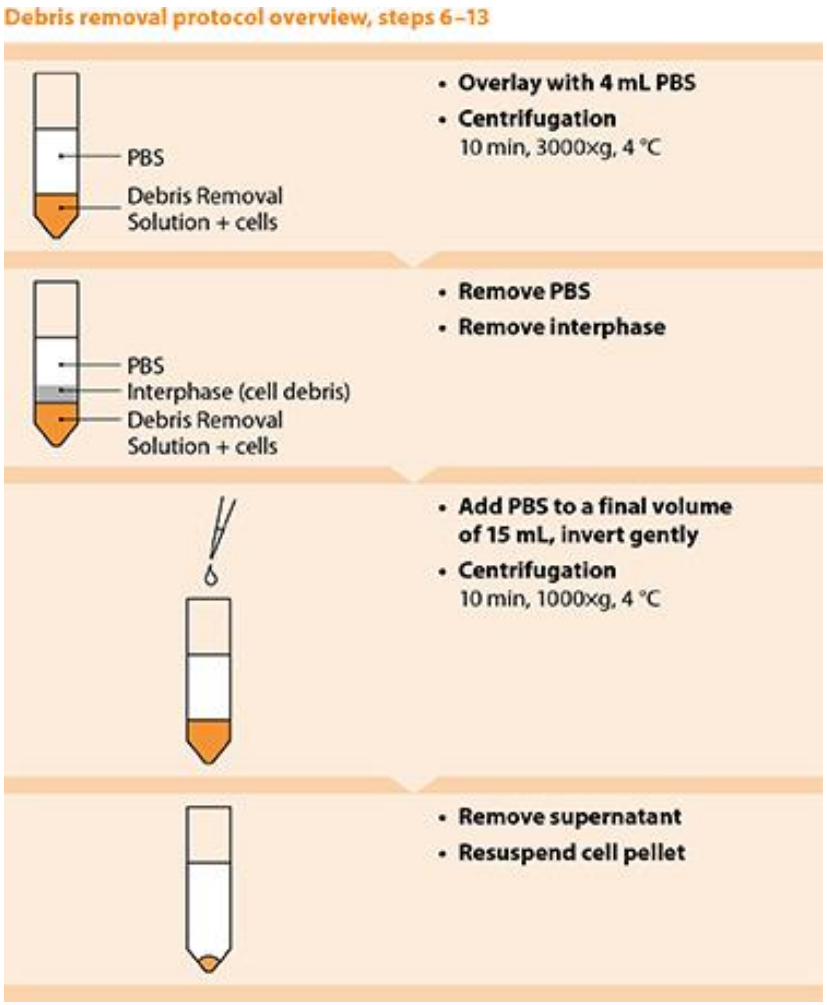
Counting and viability assessment – live/dead



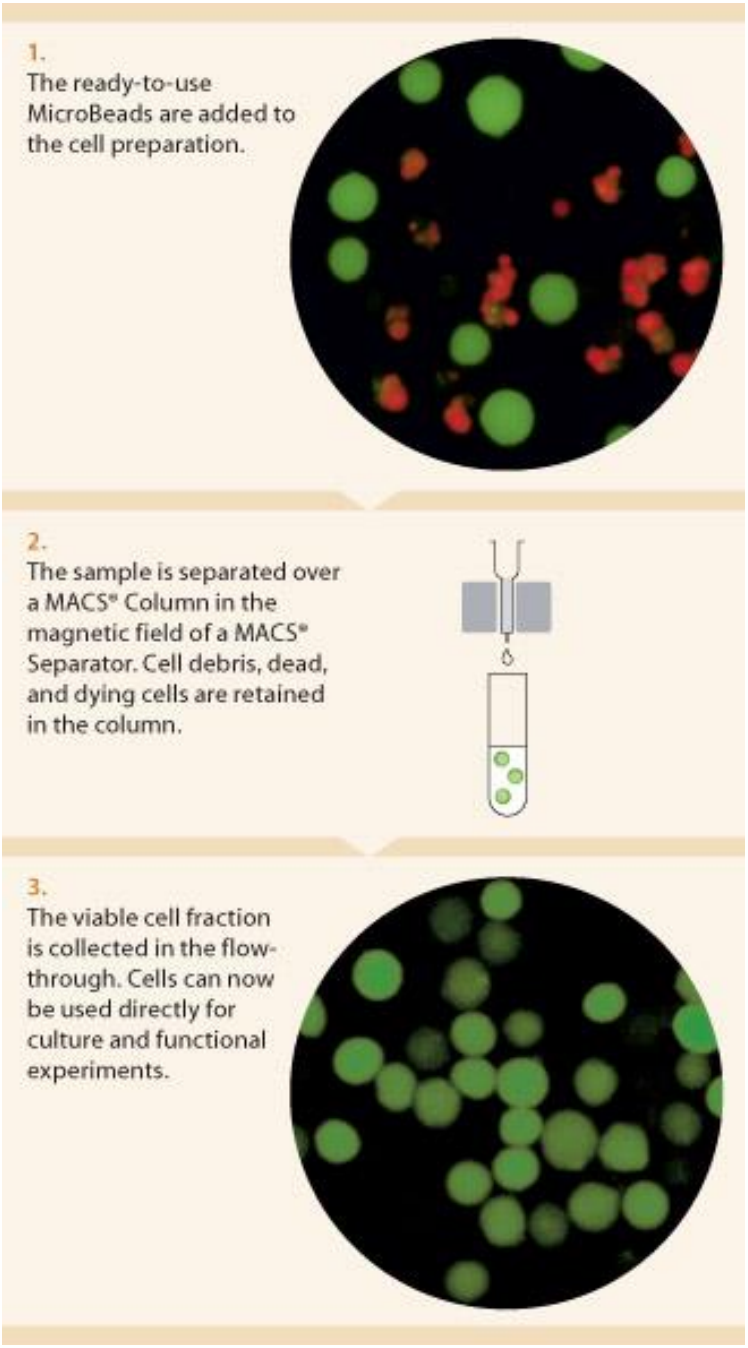




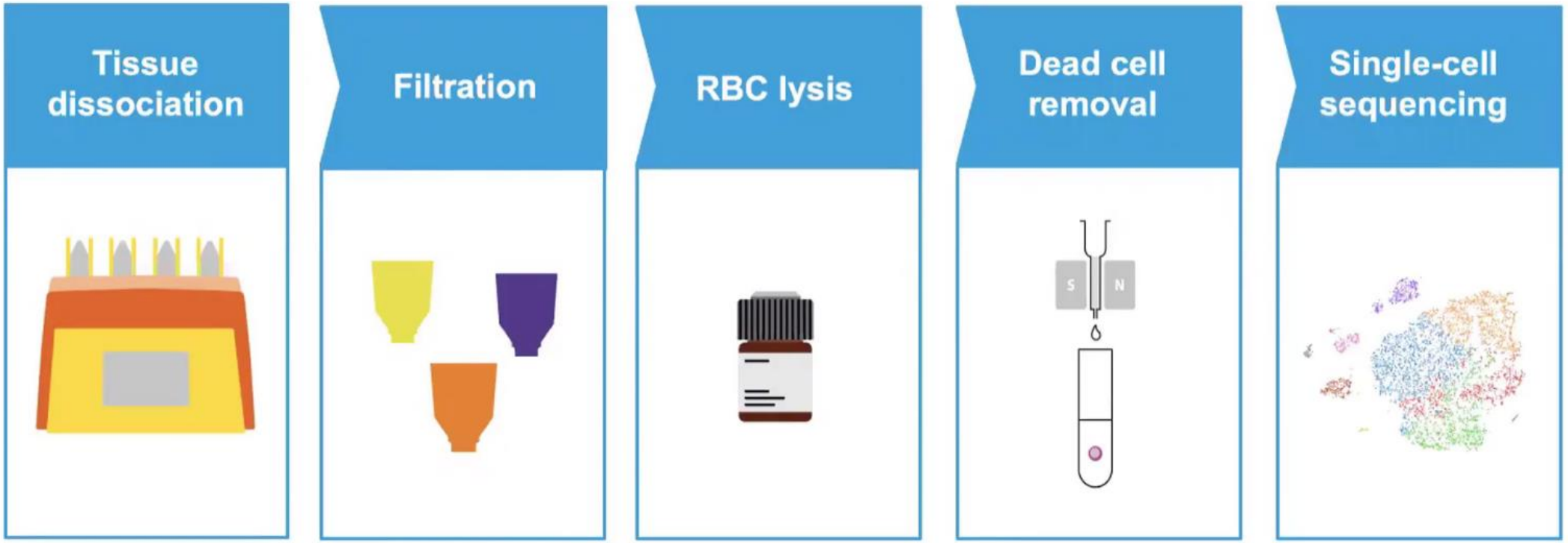
Dead cell and debris removal



Source:
Miltenyi



Example: tumor processing workflow



Summary of key sample prep lessons

- Treat cells gently and keep them happy
  - Try gentle(r) lysis conditions
  - Reduce washing steps
  - Use a swinging bucket centrifuge
  - Keep cells in media + FBS instead of PBS
- Work quickly
  - Consider sorting
  - Minimize handling steps
- *Consider the benefits and drawbacks of different techniques*

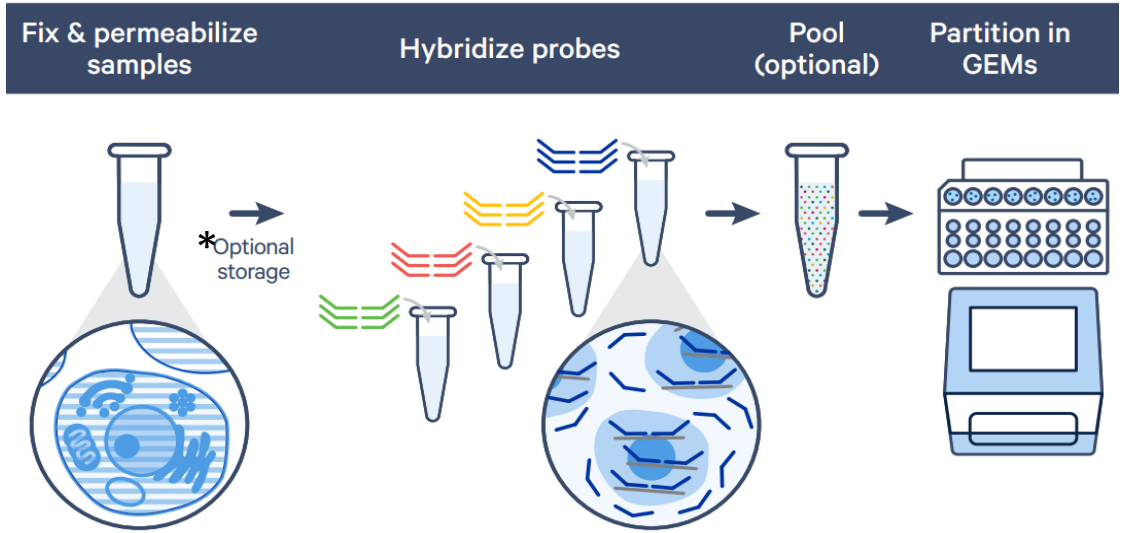
Storage after preparation? Older options

- Cryopreservation (DMSO)
  - No quantifiable difference between fresh and cryopreserved scRNA-seq samples
- Methanol fixation

**NOTE: both approaches have shown a potential bias in cell-type composition, and it is strongly recommended to thoroughly evaluate preservation methods for new cell types that have not been tested.

Storage after preparation? New options!

- Fixed RNA kit from 10x Genomics

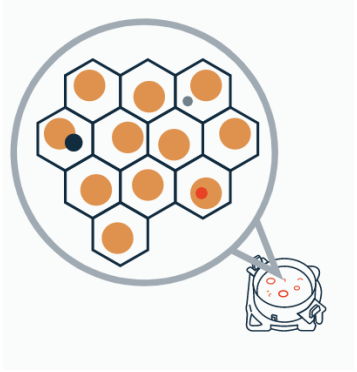


*4% PFA fixation
Storage at 4°C short-term or -20°C long-term

- Parse Biosciences (SPLiT-seq)
- Scale Biosciences (sci-seq)

Storage after preparation? Other options!

- iCell8 and Rhapsody microwells – freeze plate after single cell sorting
- The HIVE from Honeycomb – freeze HIVE after collection



- Low(er)-throughput option – sort into plates using FACS
  - Great for the DIY workflows like SMART-seq, CEL-seq2, VASA-seq, etc.

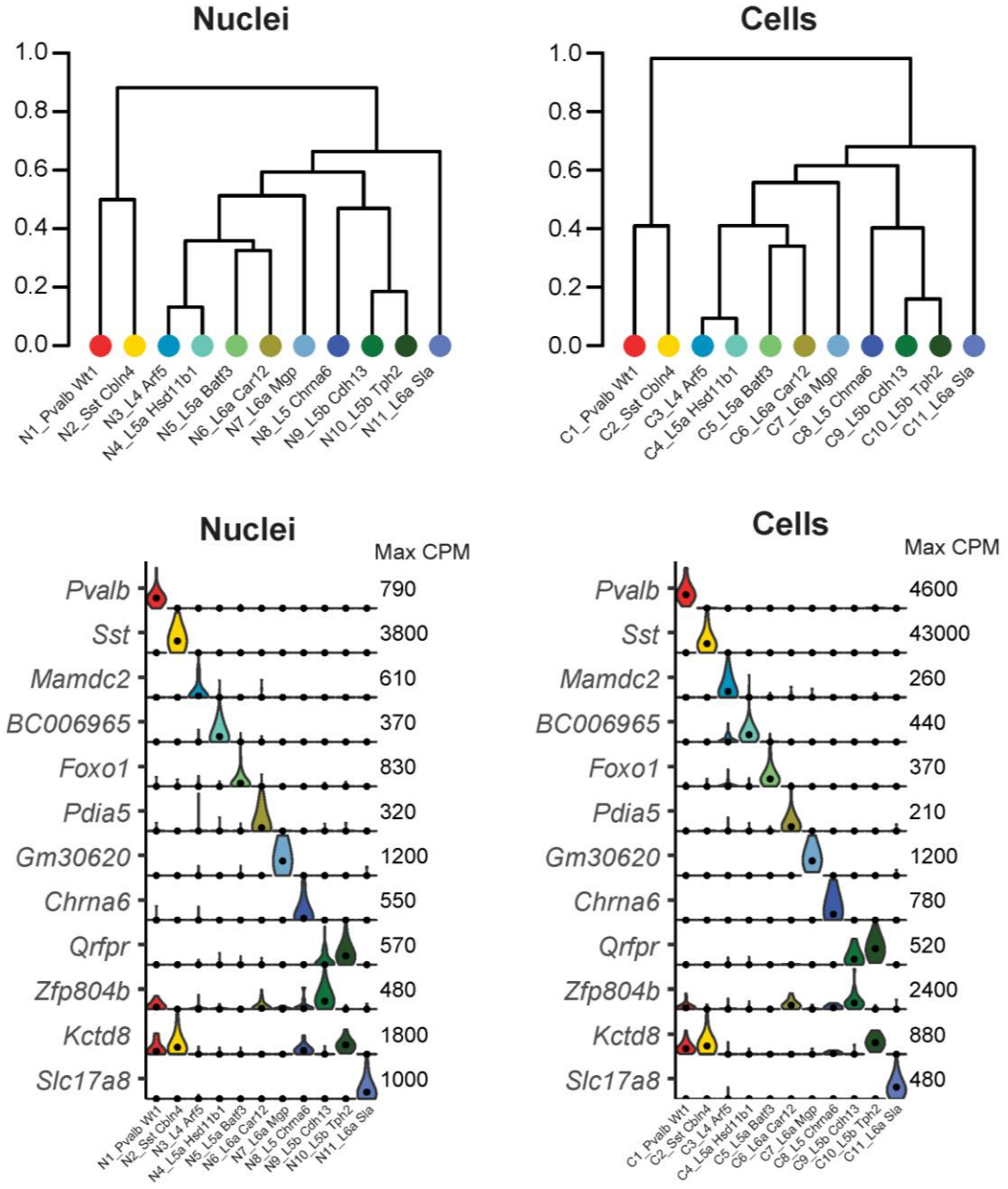
Nuclei vs whole cell

What if cells are not an option?

- Neurons
- Snap-frozen
- FFPE

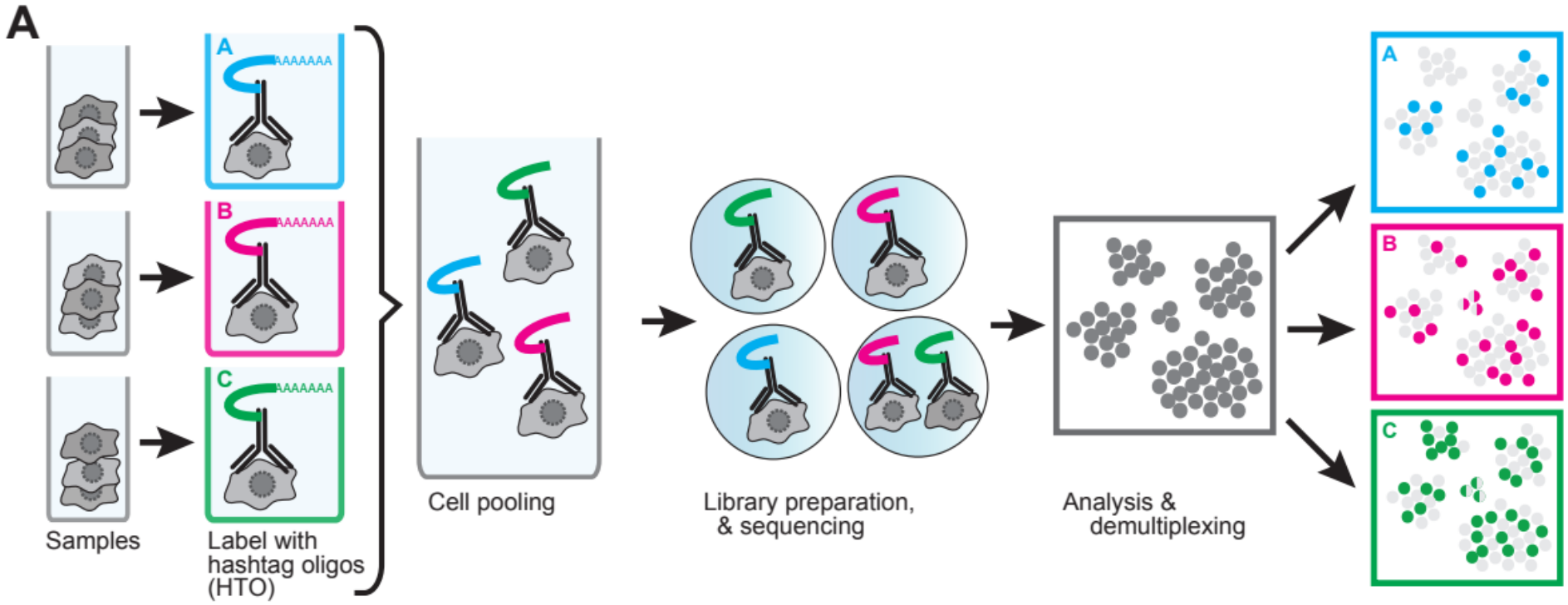
snRNA-seq is nearly identical to scRNA-seq

- Fewer unique transcripts
- More variability in per-nucleus content



Bakken et al PLOS ONE 2018

Cell hashing / multiplexing option 1

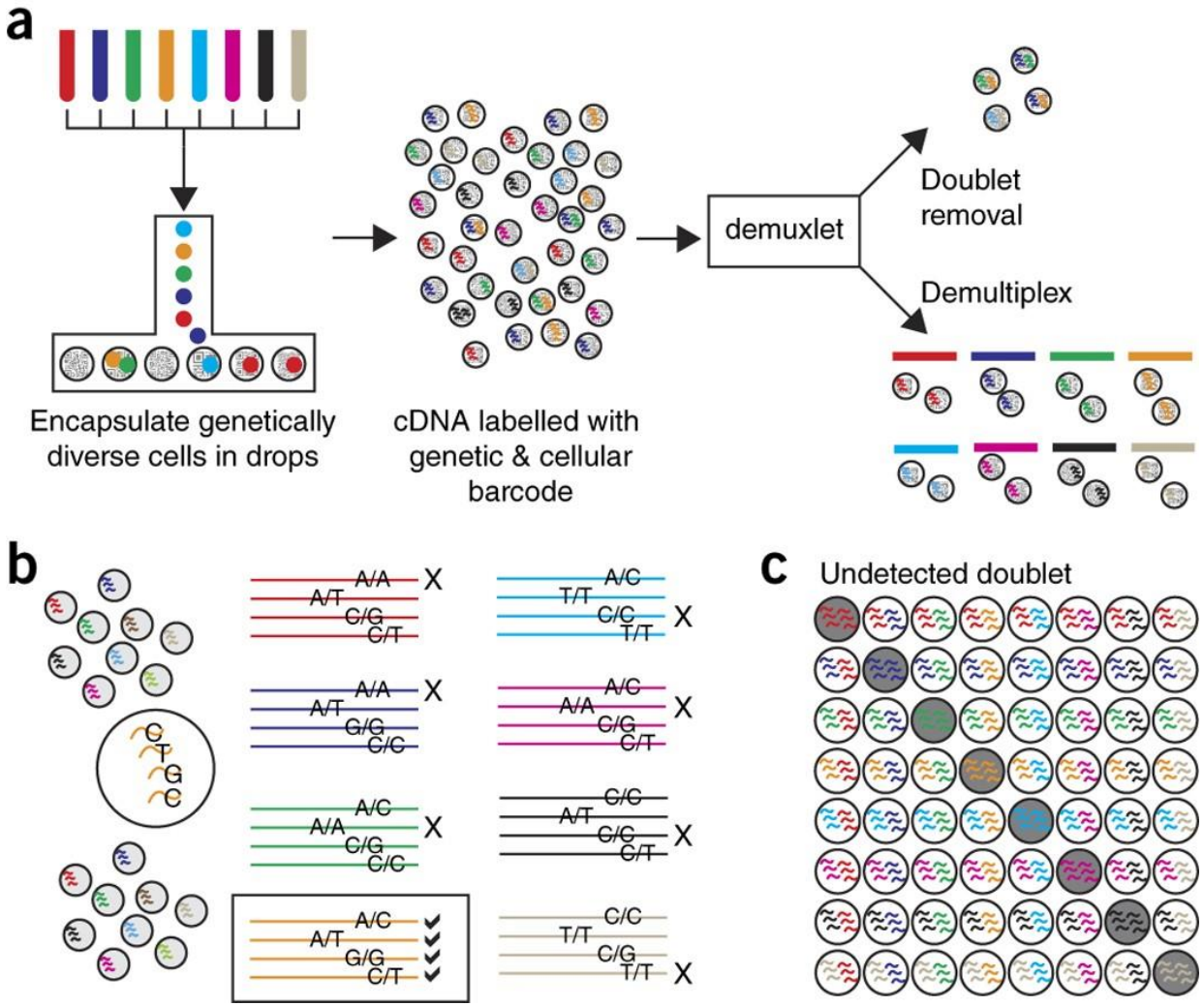


Pooled 8 PBMC samples in 1 reaction, “super-loading” 20k cells per reaction (5k multiplet)

Stoeckius et al Genome Biology Dec 2018

Cell hashing / multiplexing option 2

Demuxlet
Demultiplex samples using natural genetic variation



Kang et al Nature Biotech 2017

How many cells?

Consider sample heterogeneity and subpopulation frequency

General rule of thumb:

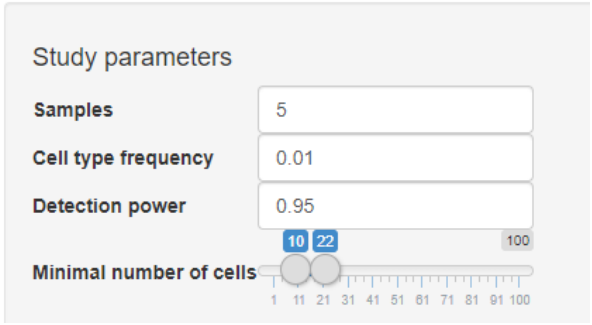
- Heterogenous populations require MORE cells
  - Here, a larger number of cells will provide greater statistical power
  - Heterogeneity can be reduced through experimental design (sorting, transgenics)

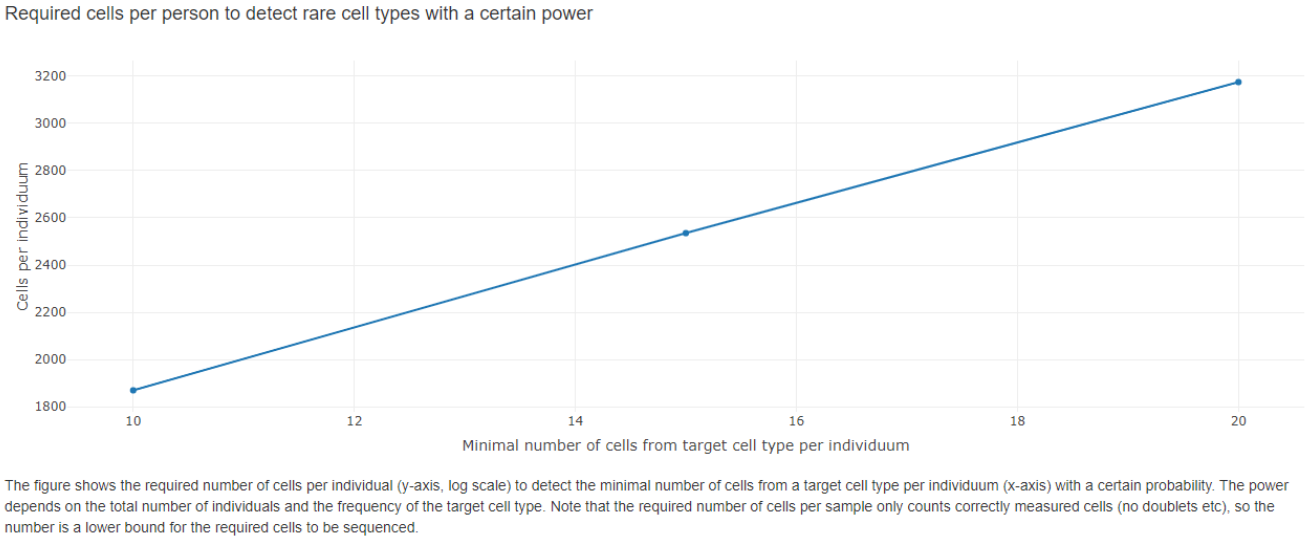
- Identification of rare cells requires MORE cells

How many samples?

Power calculations + group size estimates

- Is other single-cell data available (flow- or mass-cytometric data)?
- Computational tools available:
  - How many cells? – Satija lab
  - SCOPIT – Navin lab
  - scPower – Theis and Hening labs





The figure shows the required number of cells per individual (y-axis, log scale) to detect the minimal number of cells from a target cell type per individuum (x-axis) with a certain probability. The power depends on the total number of individuals and the frequency of the target cell type. Note that the required number of cells per sample only counts correctly measured cells (no doublets etc), so the number is a lower bound for the required cells to be sequenced.

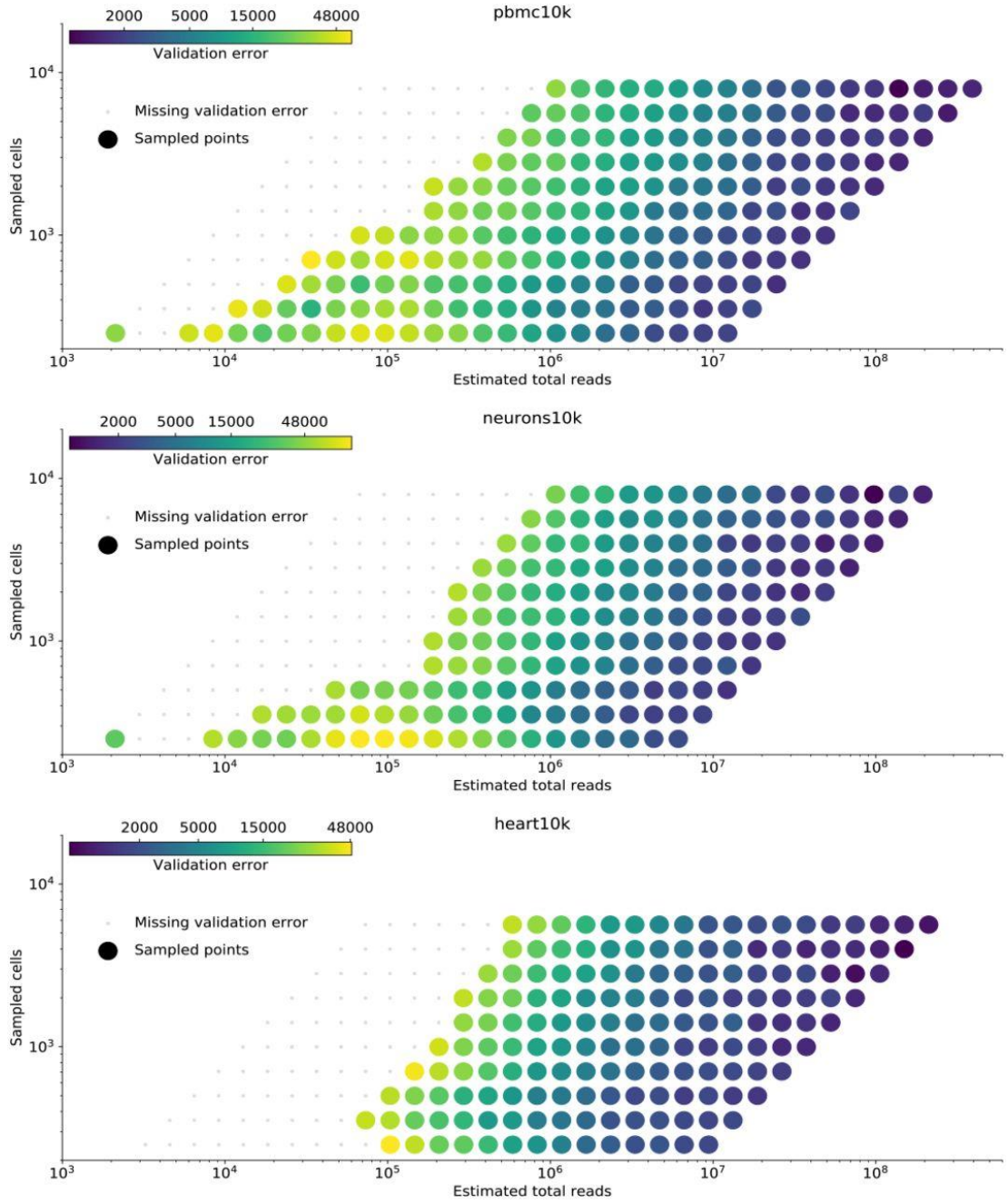
Pilot experiment recommended before large-scale data production

How deep?

For 3’ counting approaches:

<15k reads/cell, doubling reads reduces error by 30-40%

>15k reads/cell, doubling reads or cells has little effect (1-3%)



Svensson et al BioRxiv 2019

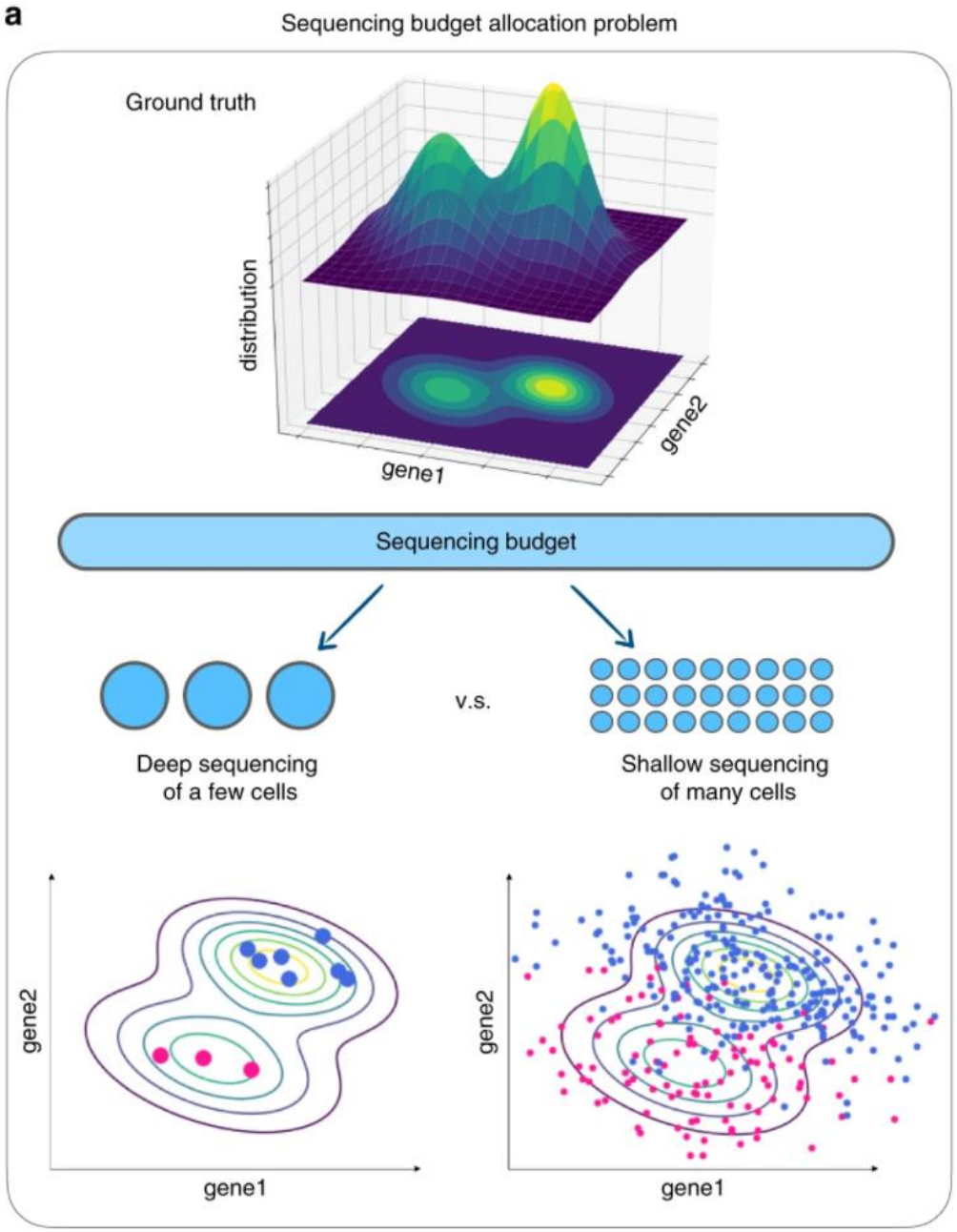
How deep?

For 3’ counting approaches:

One read per cell per gene

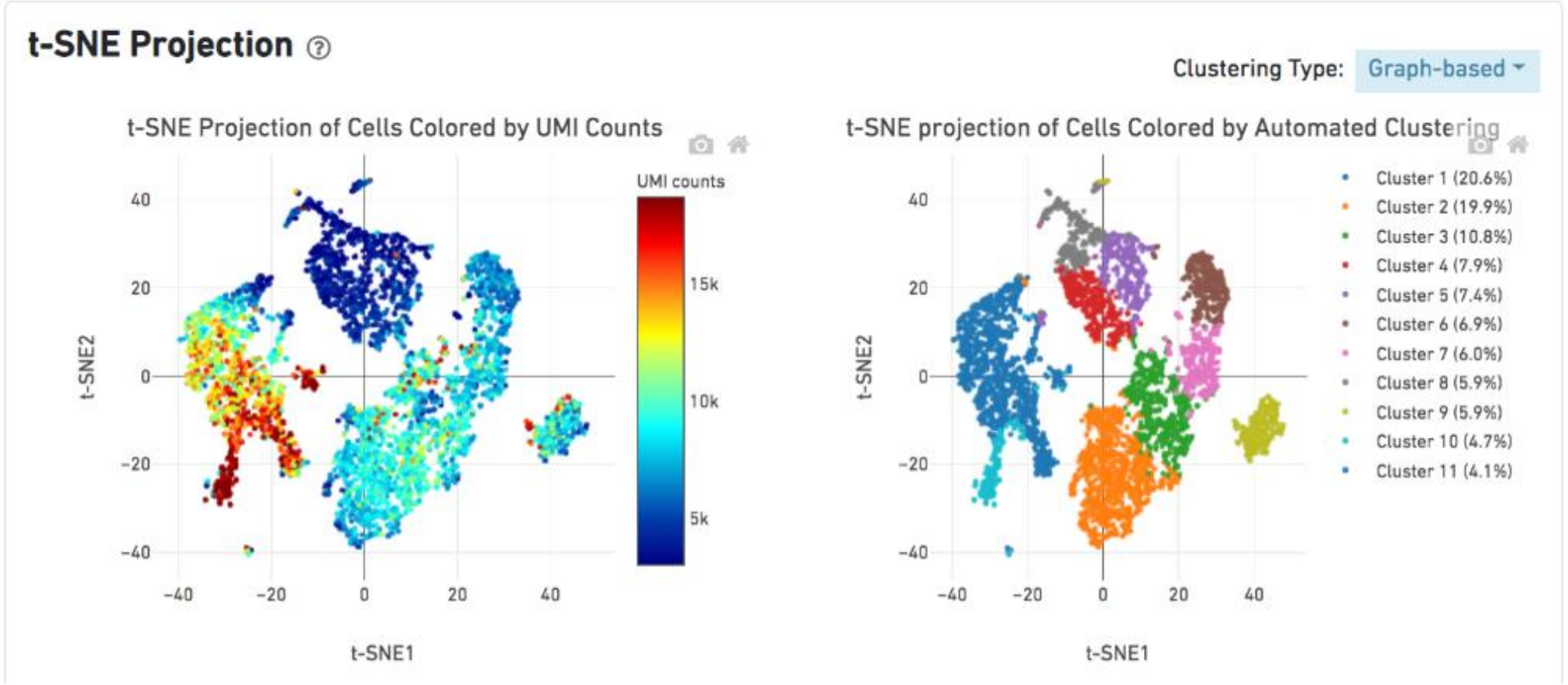
[sceb (single-cell empirical Bayes) estimator (Python package)]

Zhang, Ntranos and Tse Nat Comm 2020



How deep?

- BUT… not all cells will have the same depth of sequencing



Conclusions

- Think before you start
  - Which library prep protocol will you choose?
  - How should you store your samples?
  - How should you prepare your tissue/cells/nuclei?
- Maintain sample integrity
- Minimize batch effects
- Make sure your experiment/study has enough power