

Introduction to Single Cell RNA Sequencing

Sarah Boswell

Director of the Single Cell Core, Harvard Medical School

Director of Sequencing Technologies, Laboratory of Systems Pharmacology

Staff Scientist/Sequencing Specialist, Systems Biology, Springer Lab

<https://singlecellcore.hms.harvard.edu/>

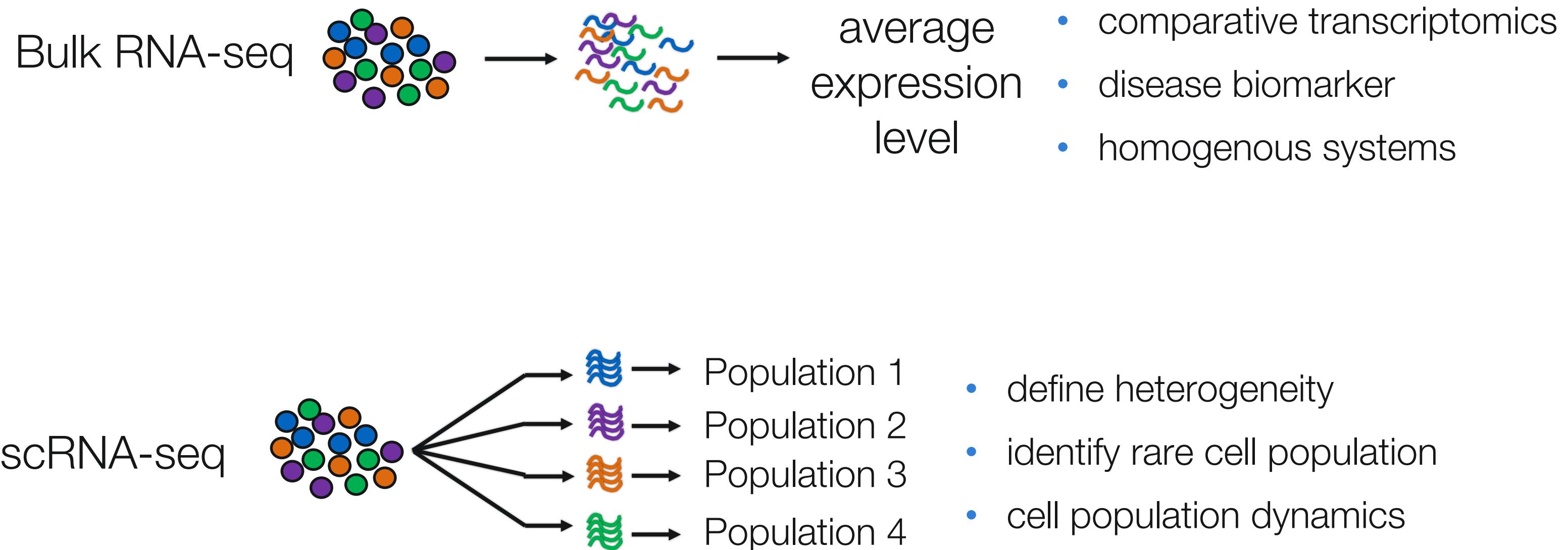
sarah_boswell@hms.harvard.edu



Introduction to Single Cell RNA Sequencing

- Common applications of single cell RNA sequencing.
- Overview of inDrops and 10x platforms.
- Experimental design and sample preparation.
- Effects of sample prep and sample type on analysis.

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



Transcriptome Coverage (mRNA)

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

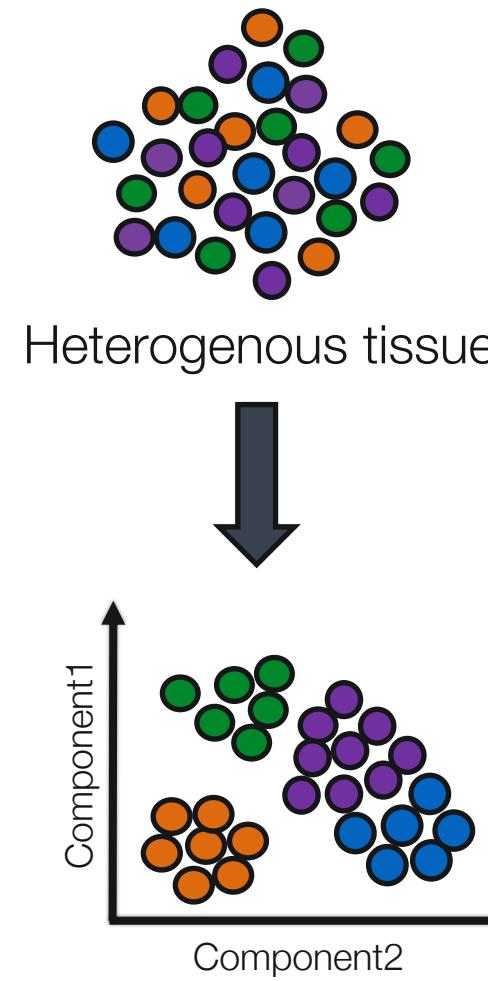
- 4000-6000 transcripts per sample
 - Limiting to transcripts observed across all samples
- Observe 20-60% of the transcriptome

3. Single Cell Methods

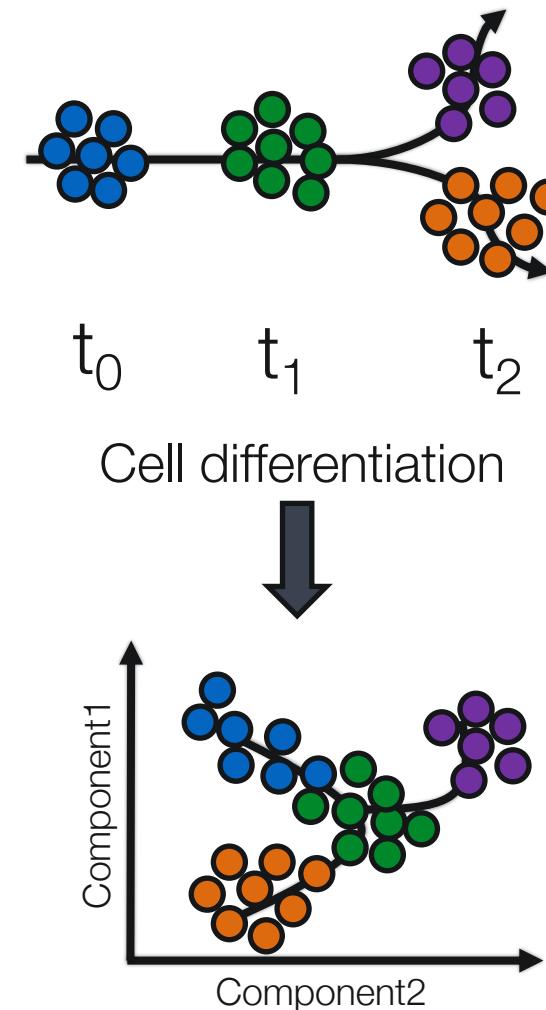
- 200 -10,000 transcripts per cell
- Observe 10-50% of the transcriptome
- Many transcripts will show up with zero counts in every cell. (even GAPDH)
- If you only looked at transcripts observed in all cells numbers drop dramatically.

Common applications of scRNA-seq

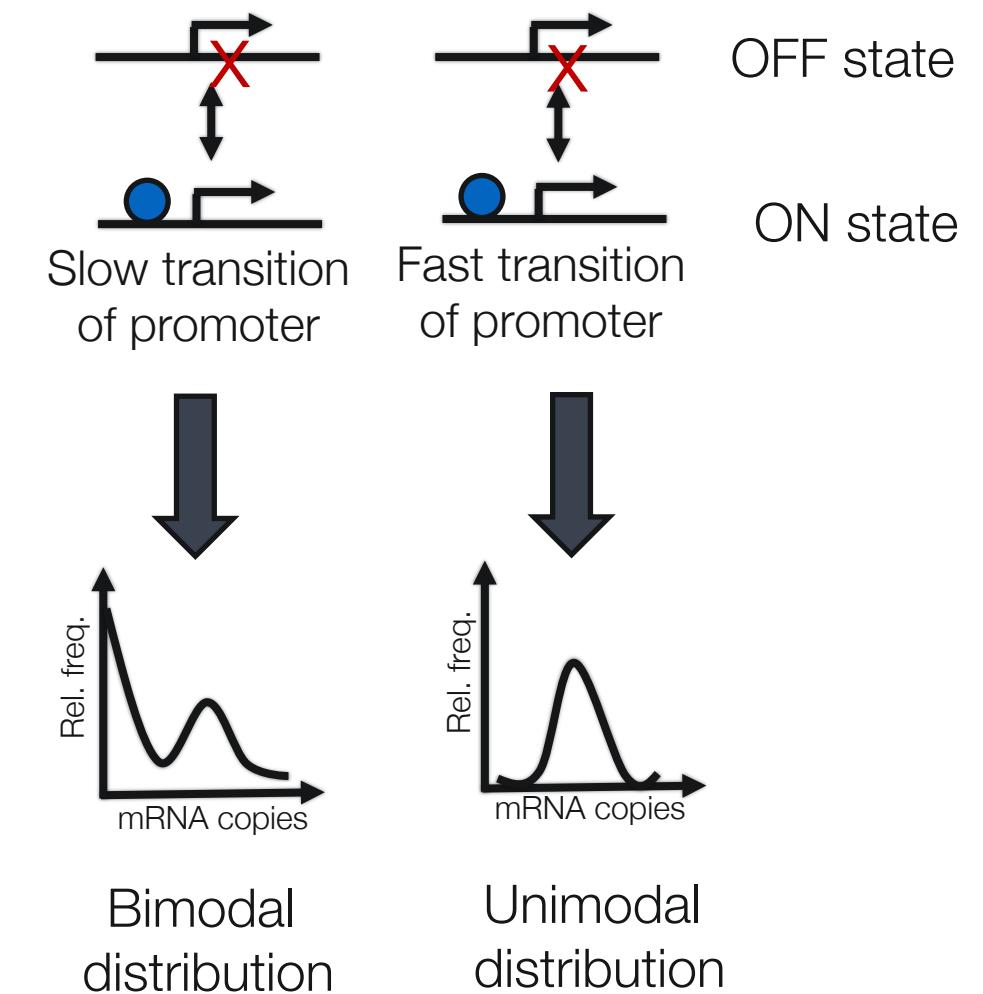
Studying heterogeneity



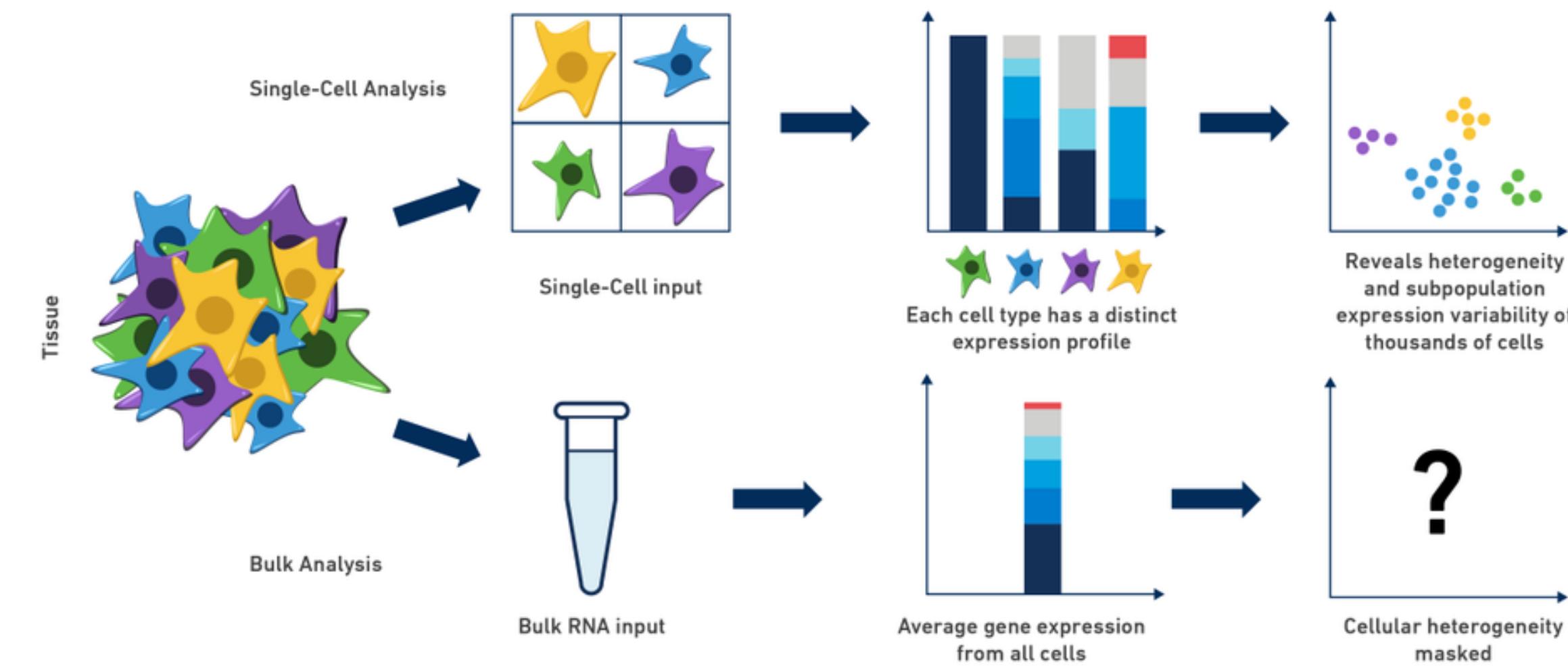
Lineage tracing study



Stochastic gene expression

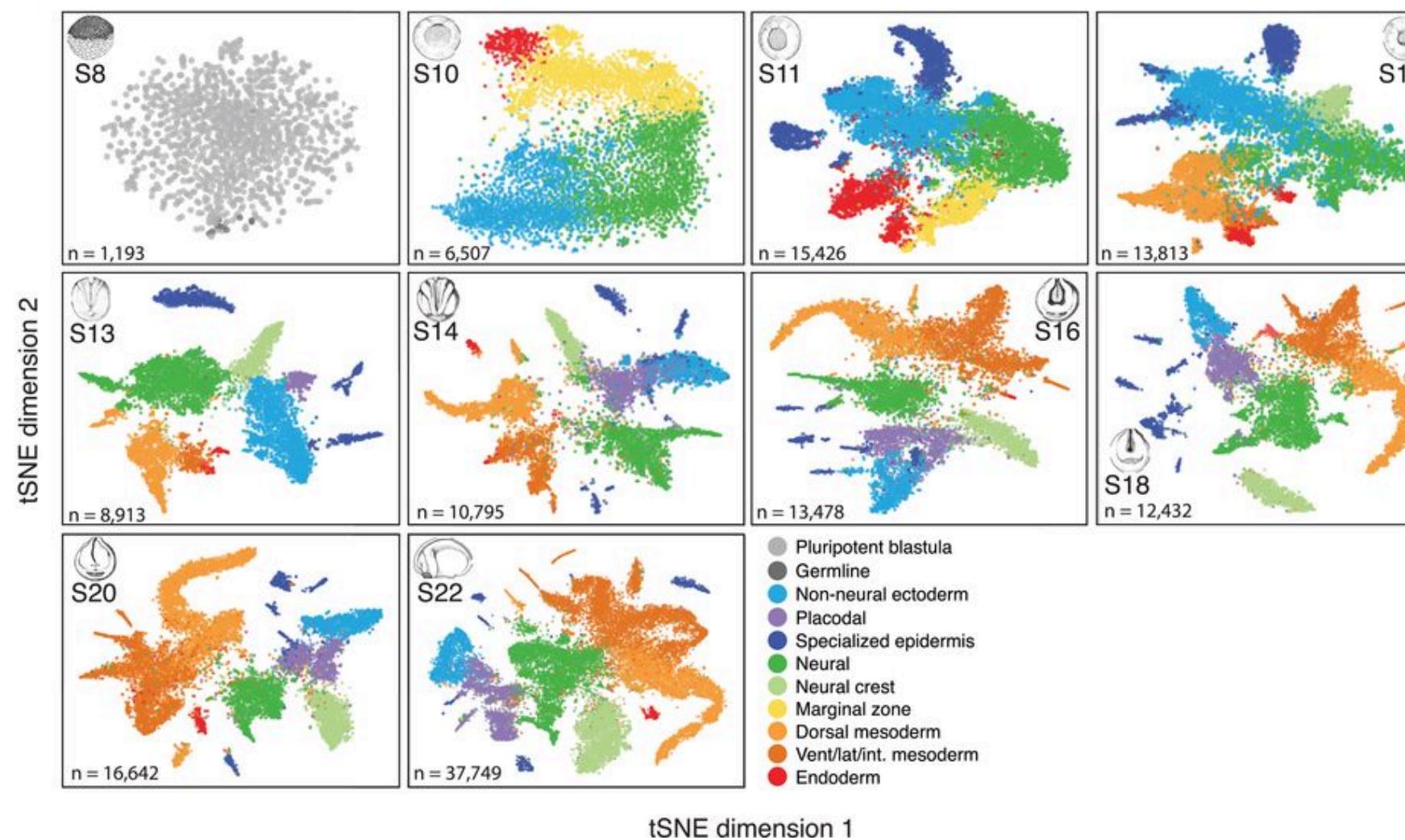


Tumor, Tissue, Organoid Heterogeneity

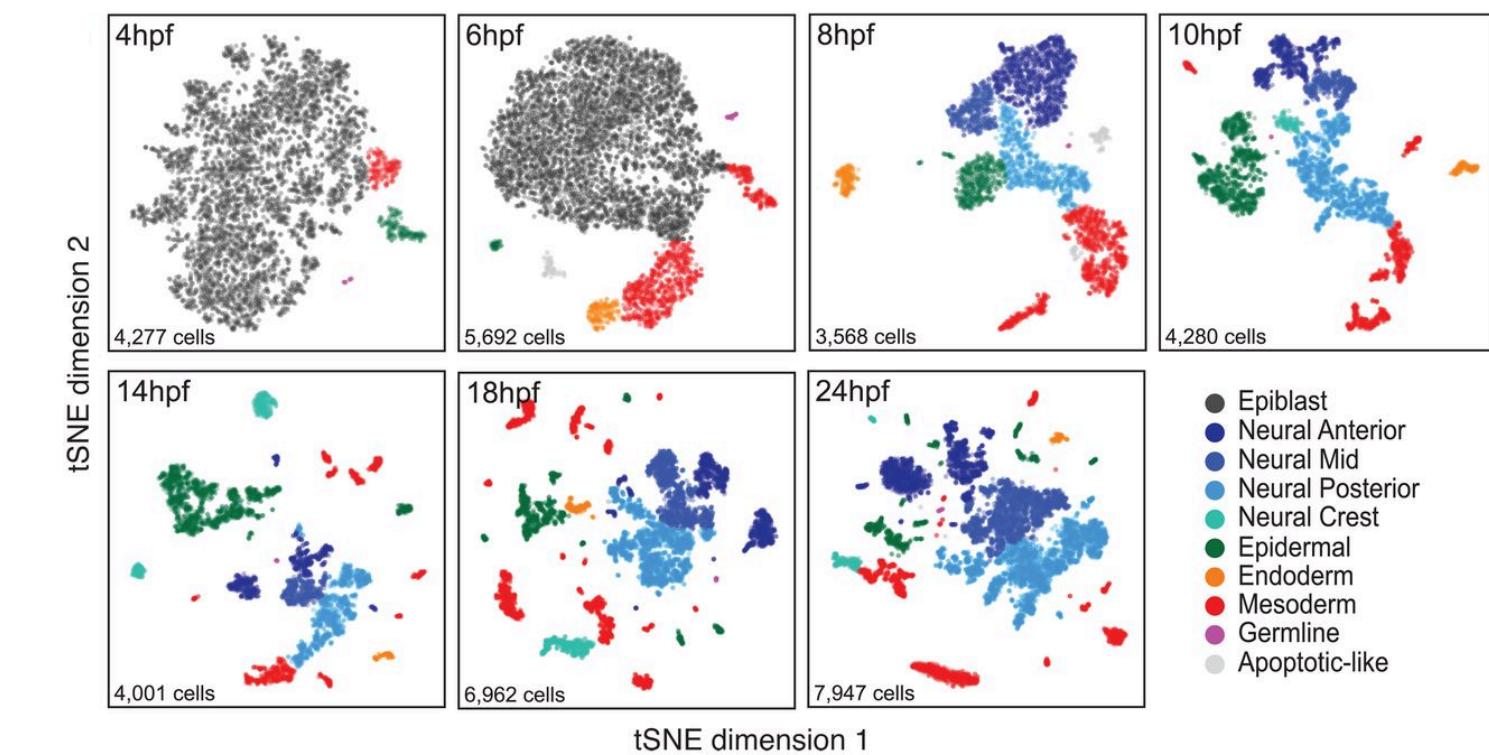


Development Lineage Tracing

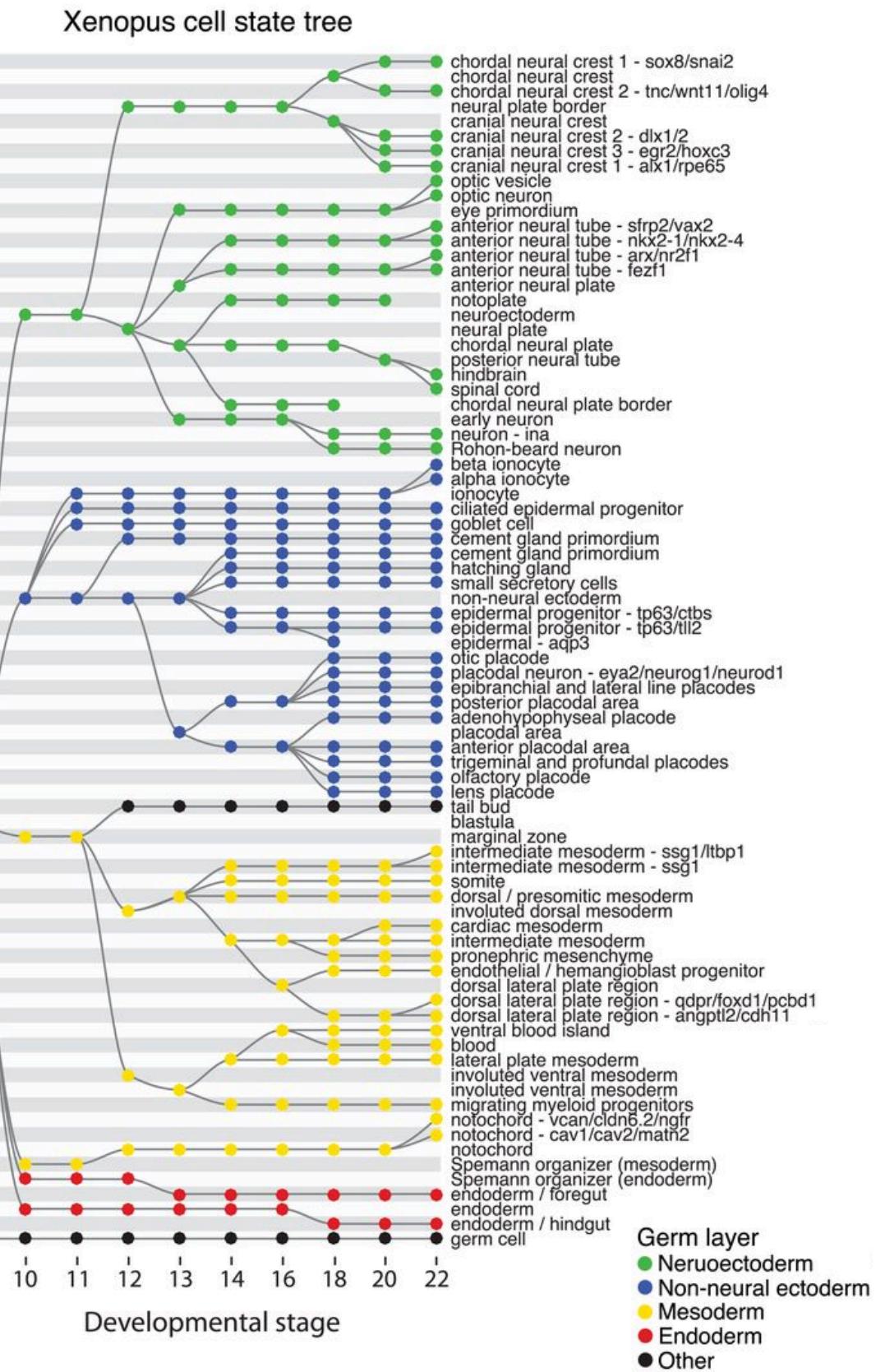
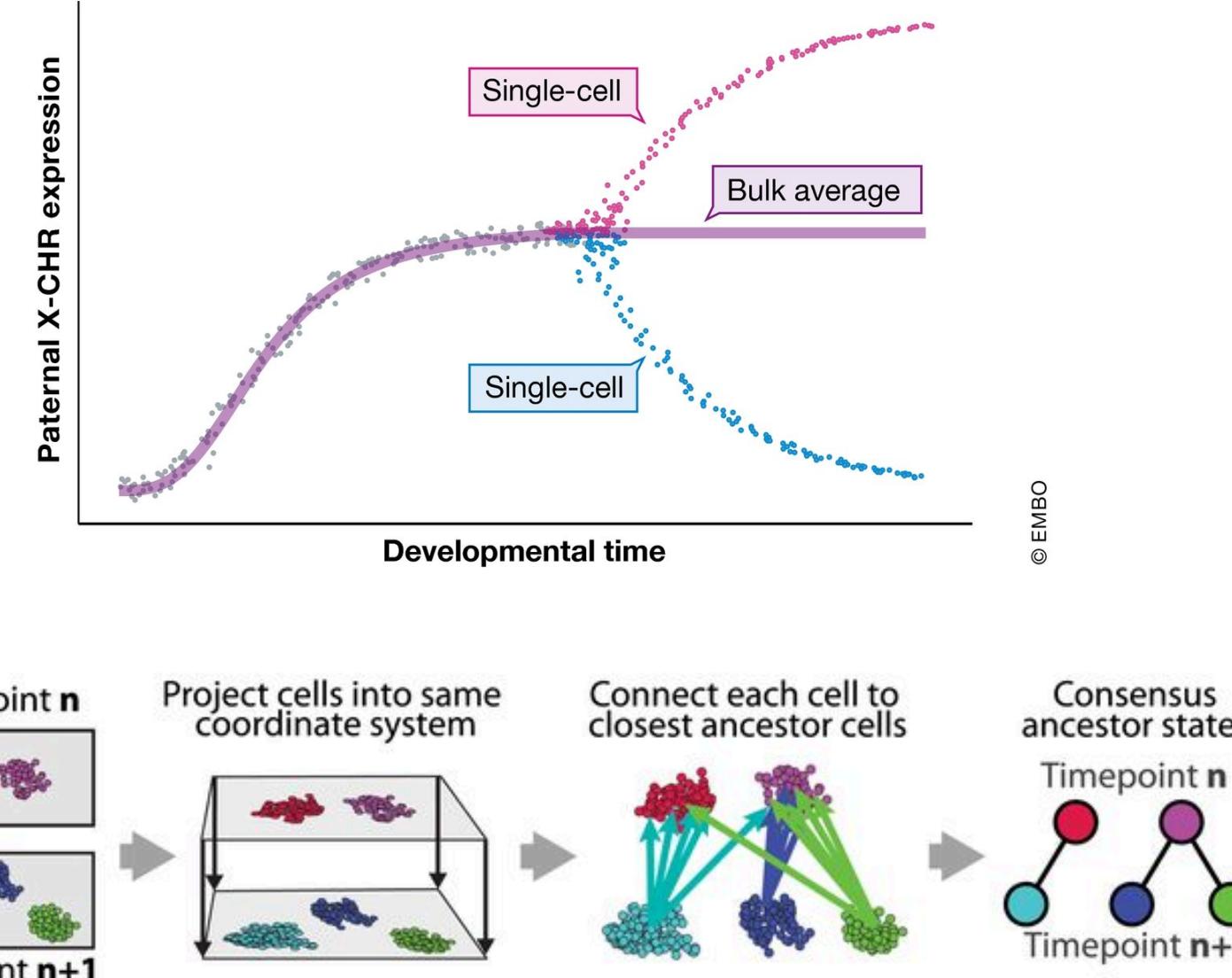
Frog



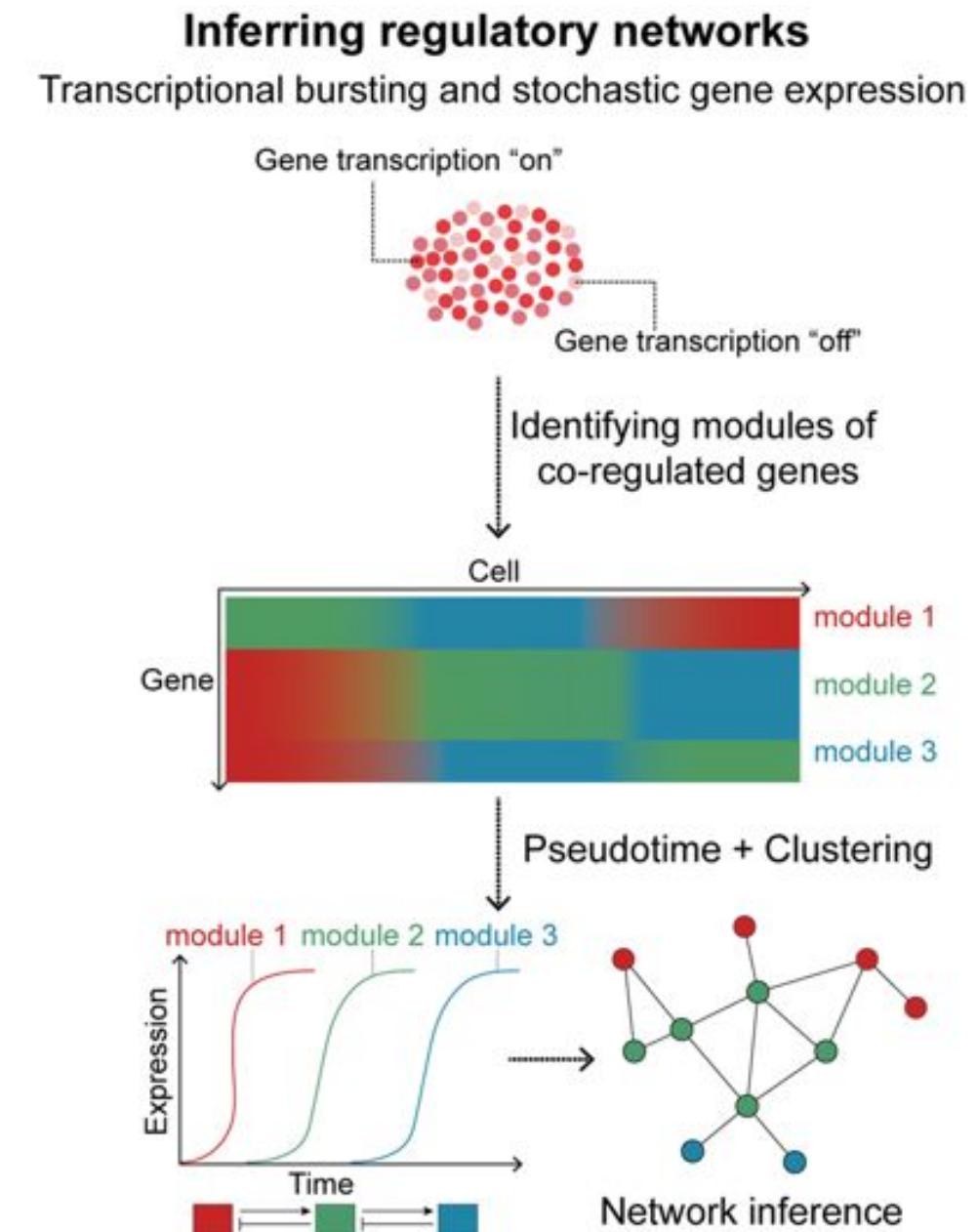
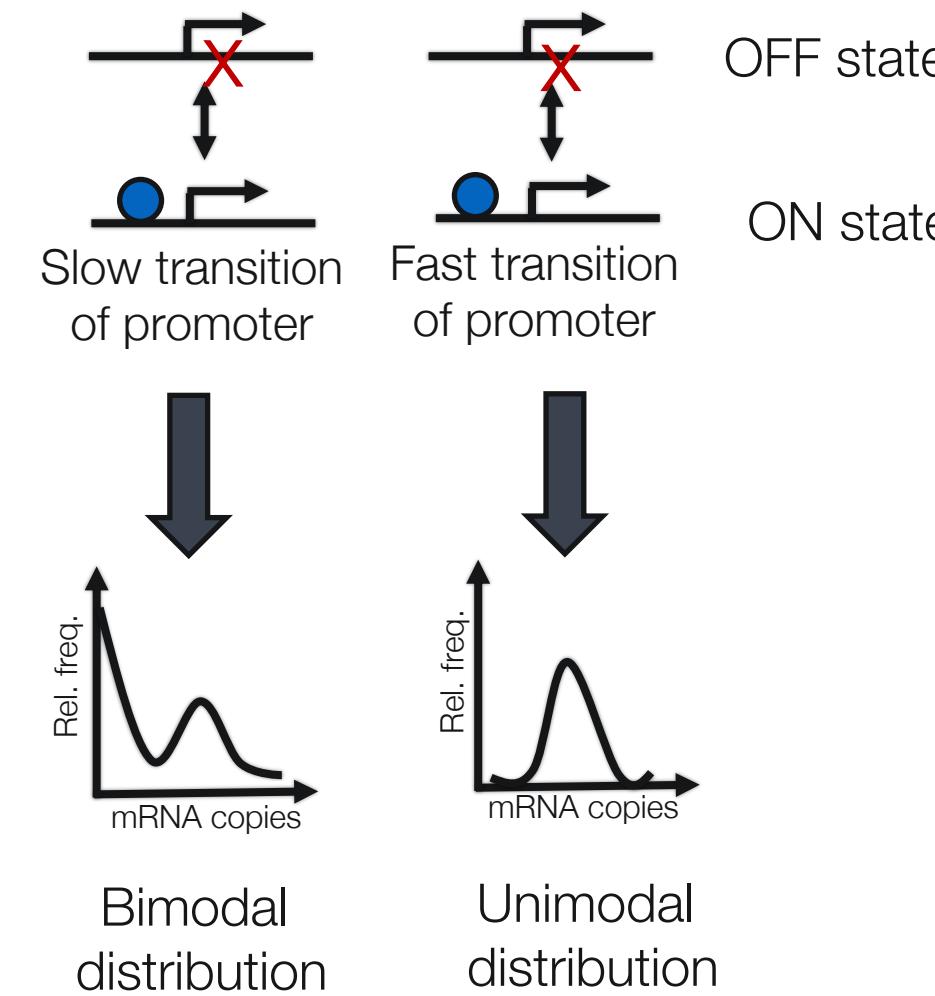
Zebrafish



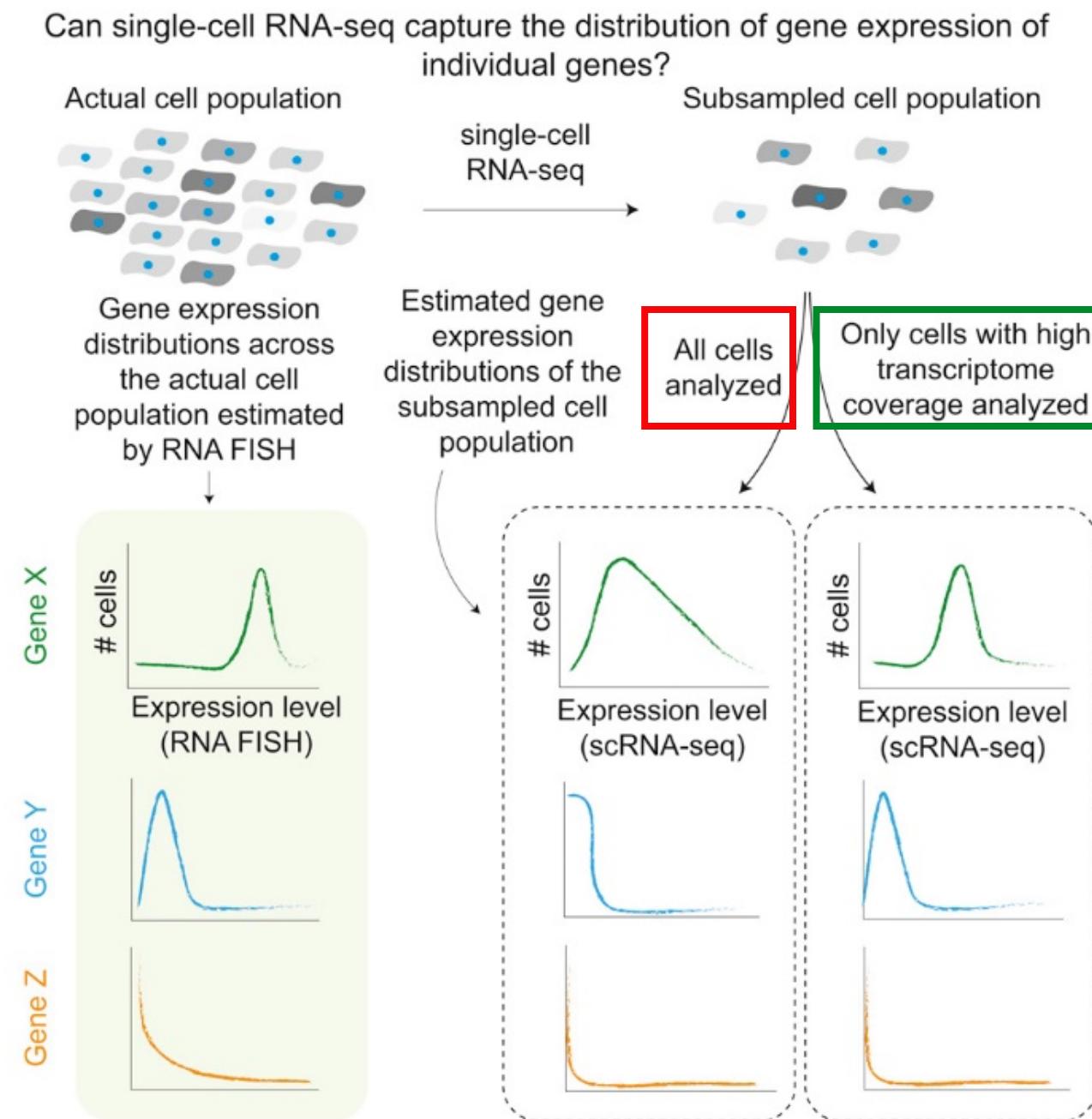
Development Lineage Tracing



Stochastic Gene Expression



Stochastic Gene Expression



- Random fluctuations of the mechanisms underlying mRNA and protein production cause heterogeneity among otherwise-identical cell populations.
- Low mRNA capture efficiency of scRNA-seq makes it difficult to draw definitive conclusions about expression at the single-cell level.
- Number of cells and depth of sequencing critical for understanding rare gene expression phenotypes.

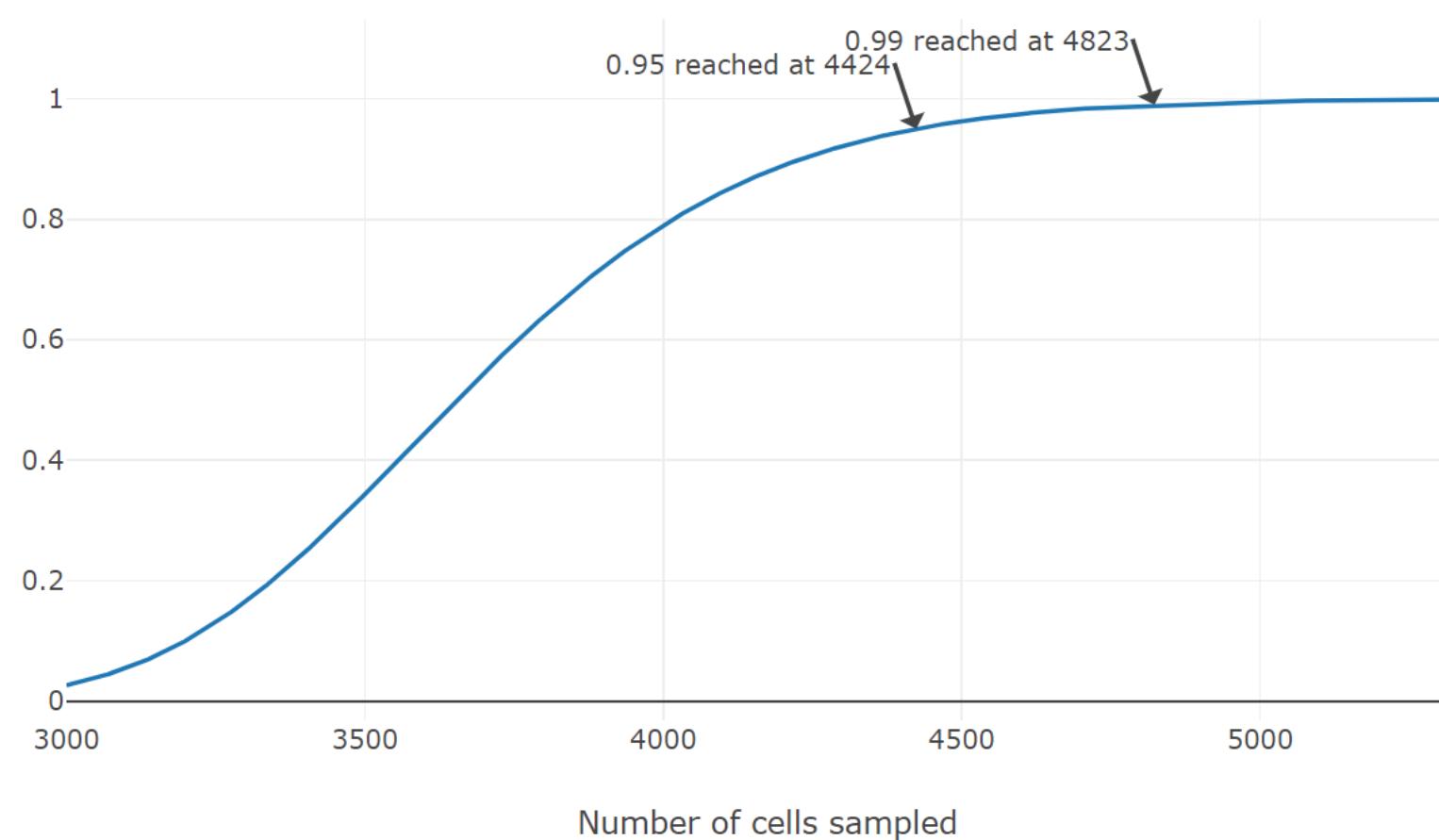
More Cells or More Sequencing Reads?

- Required number of cells increases with complexity of the sample.
- As the number of genes involved in the biology decrease then the coverage requirements increase (more reads).
- Cell-type classification of a mixed population usually requires lower read depth and can be sequenced at 10,000-50,000 reads per cell.
- We typically suggest starting with 25,000-55,000 reads per cell. You can always re-sequence your samples.

<https://satijalab.org/howmanycells>

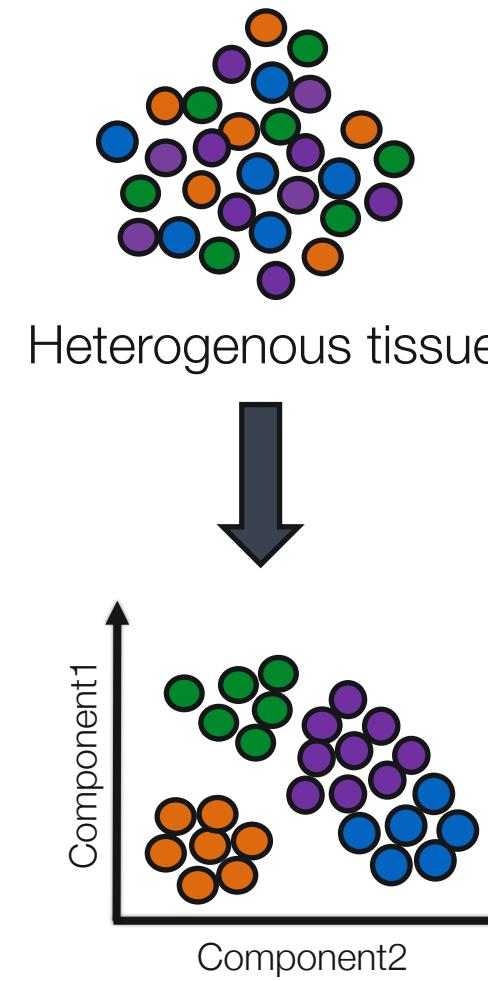
| | | |
|--------------------------------|--|---------------------------------|
| Assumed number of cell types | Minimum fraction (of rarest cell type) | Minimum desired cells per type |
| <input type="text" value="6"/> | <input type="text" value=".01"/> | <input type="text" value="30"/> |

Probability of seeing at least 30 cells from each cluster

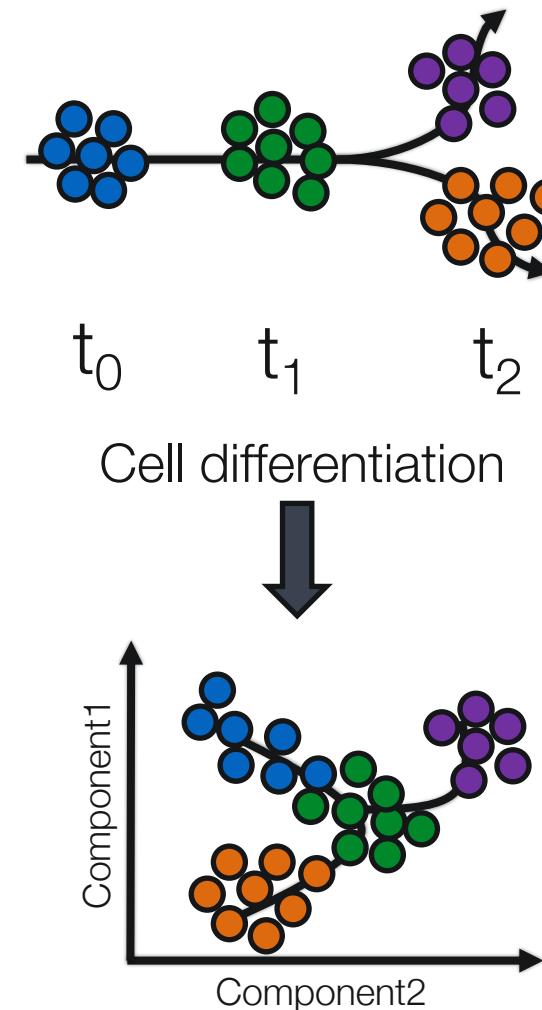


Common applications of scRNA-seq

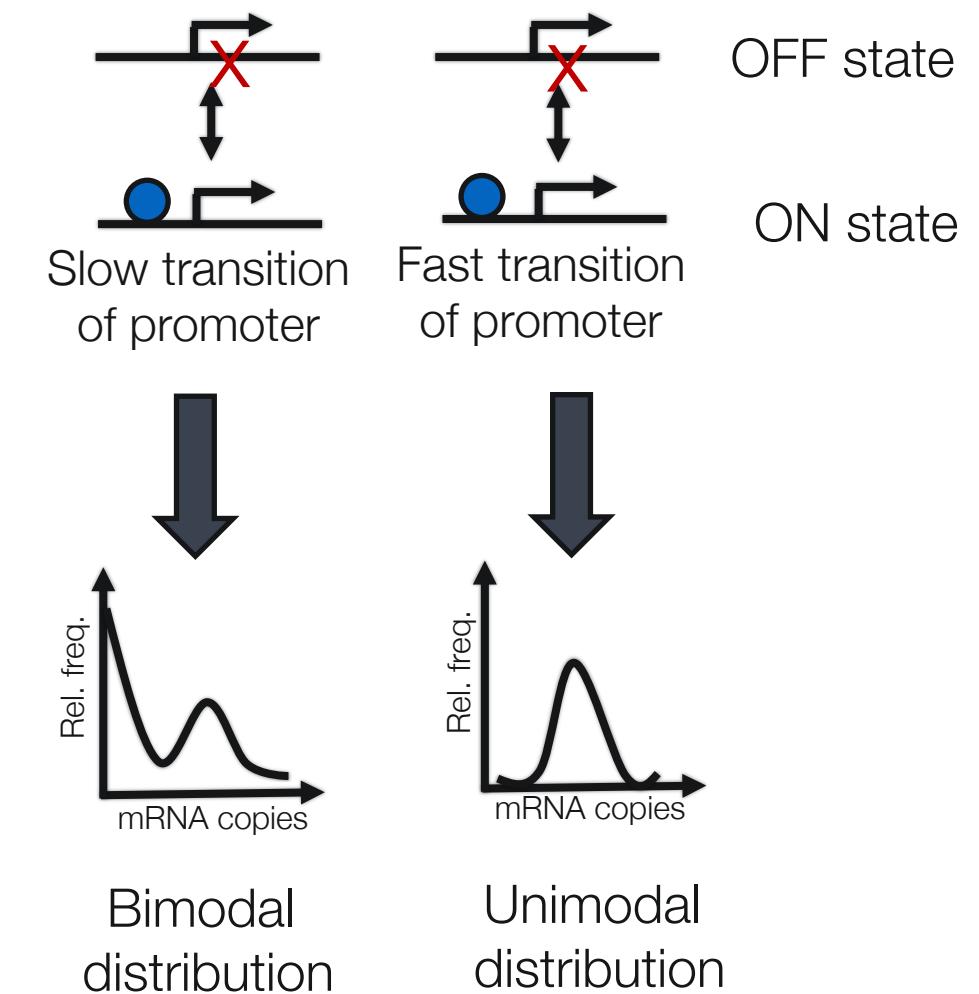
Studying heterogeneity



Lineage tracing study

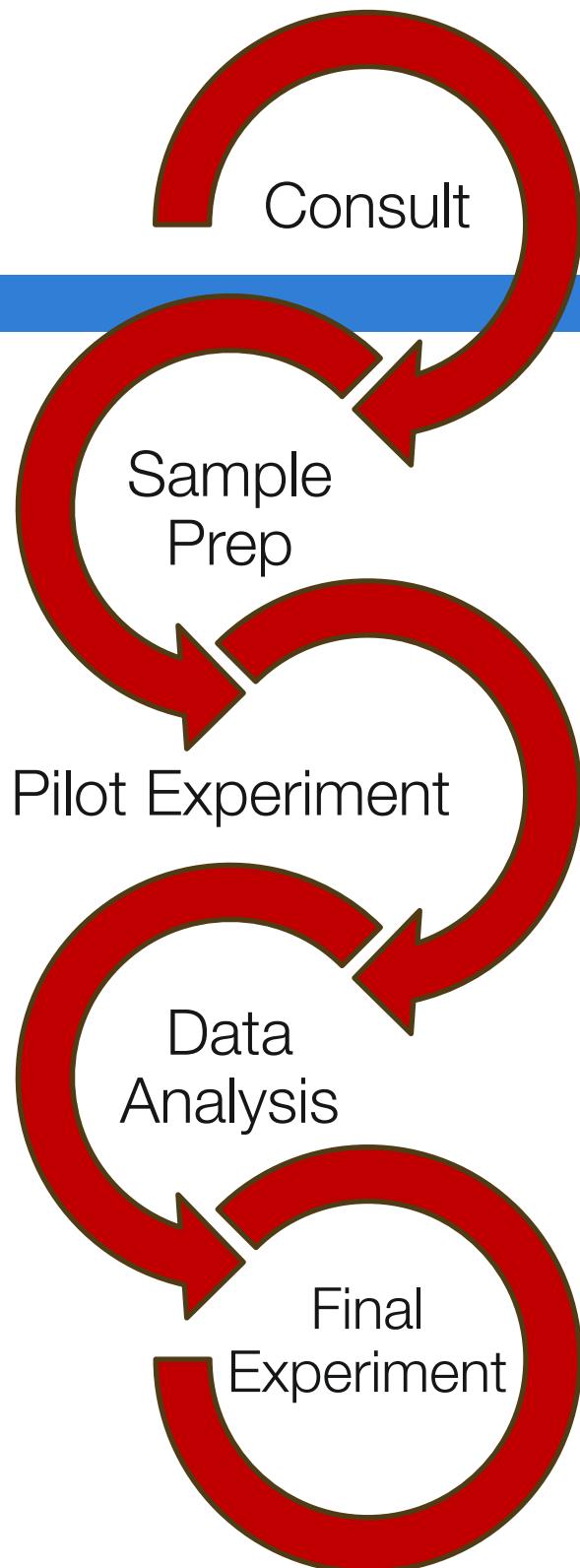


Stochastic gene expression



Single Cell Core Workflow

- Good sample prep is the key to success.
- A well planned pilot experiment is essential for evaluating sample preparation and for understanding the required number of cells.
- Do not rush to the final experiment.



Introduction to Single Cell RNA Sequencing

- Common applications of single cell RNA sequencing.
- Overview of inDrops and 10x platforms.
- Experimental design and sample preparation.
- Effects of sample prep and sample type on analysis.

Comparison of Single Cell Methods

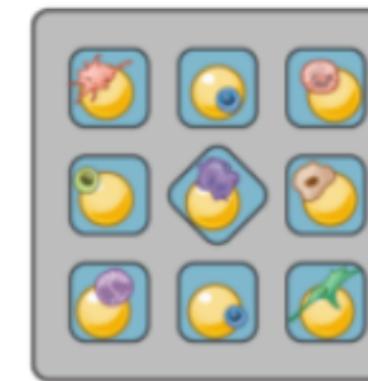
CELL-Seq
MARS-Seq
SMART-Seq
SCRB-Seq



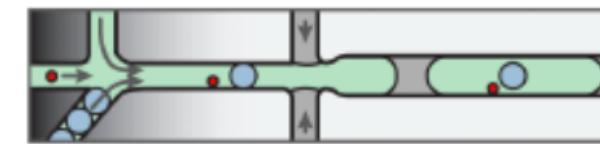
Chromium (10x)



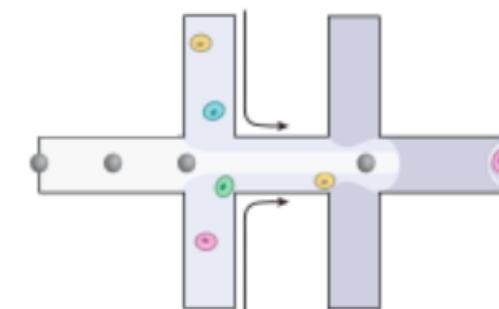
Seq-Well



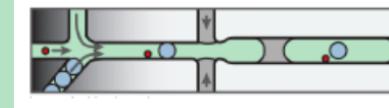
inDrops



Drop-Seq

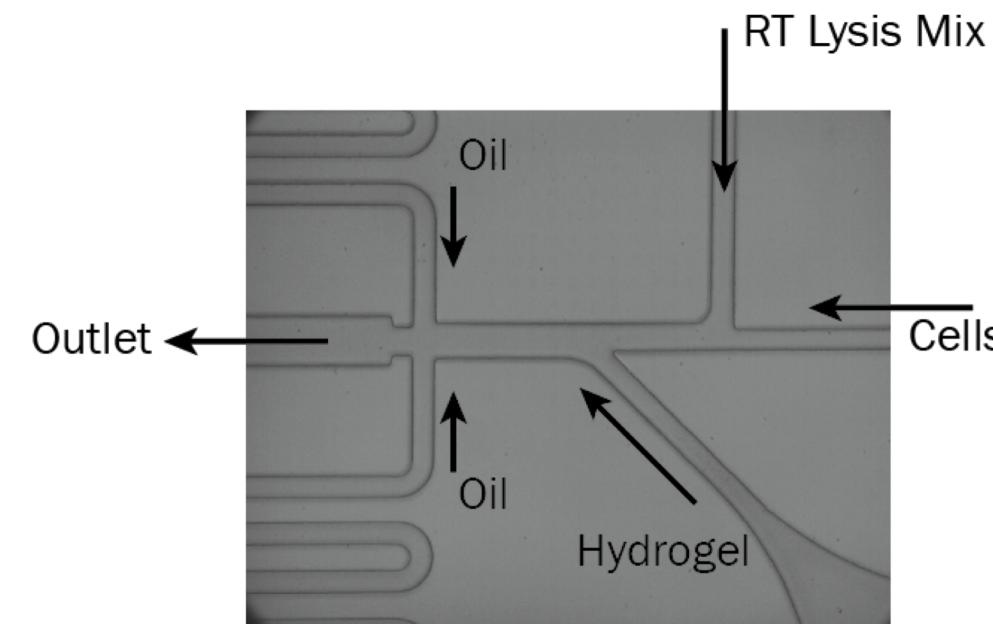


Comparison of Single Cell Methods

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

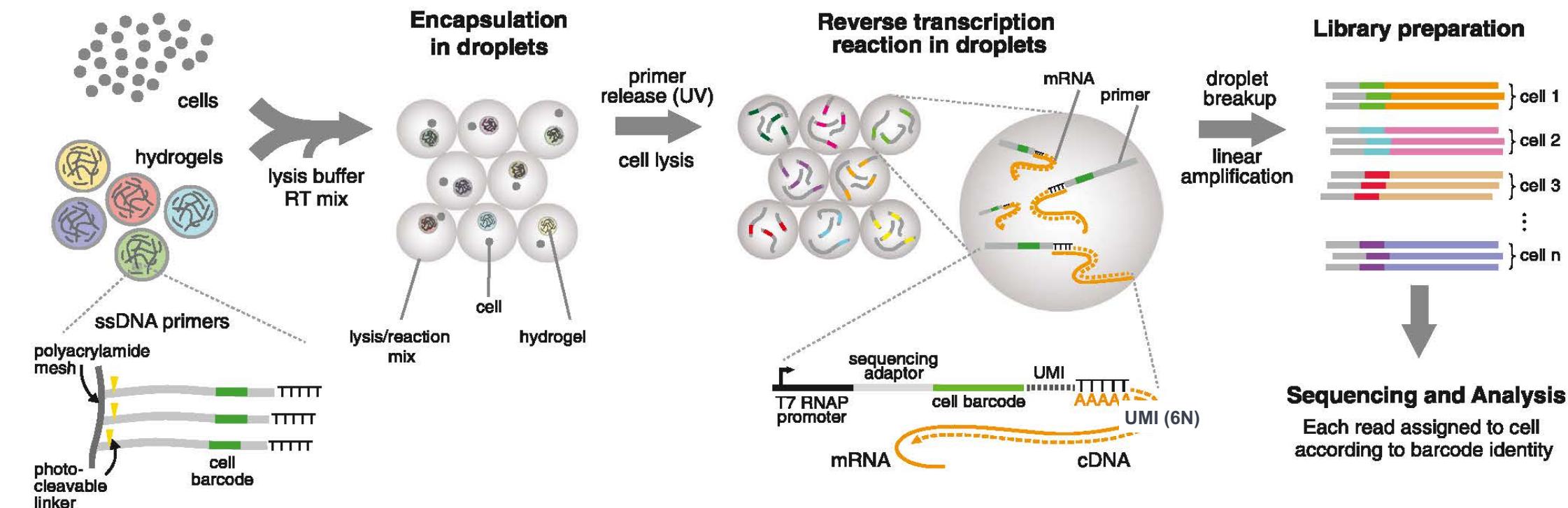
inDrops Method Overview

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



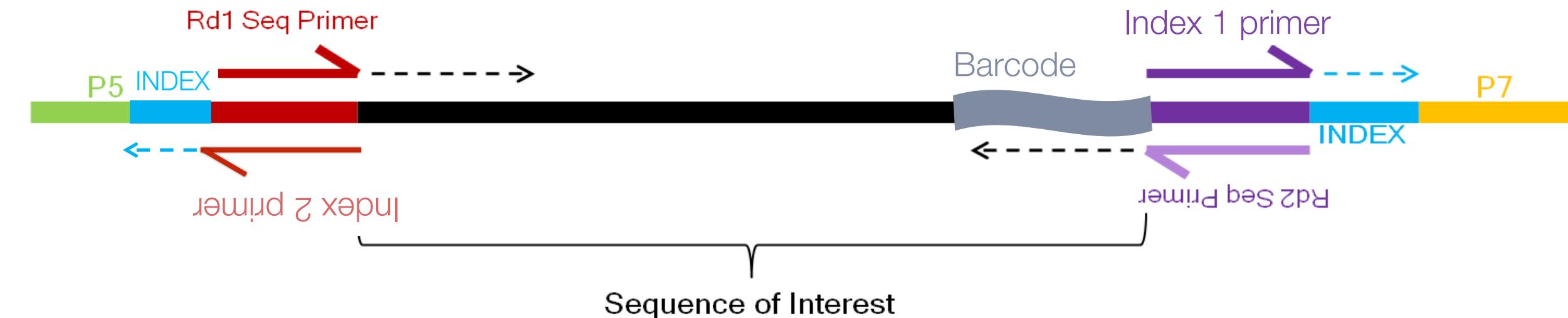
- Matching the speed of bead injection with the speed of droplet generation it is possible to set conditions in which nearly every droplet would be loaded.

inDrops Method Overview



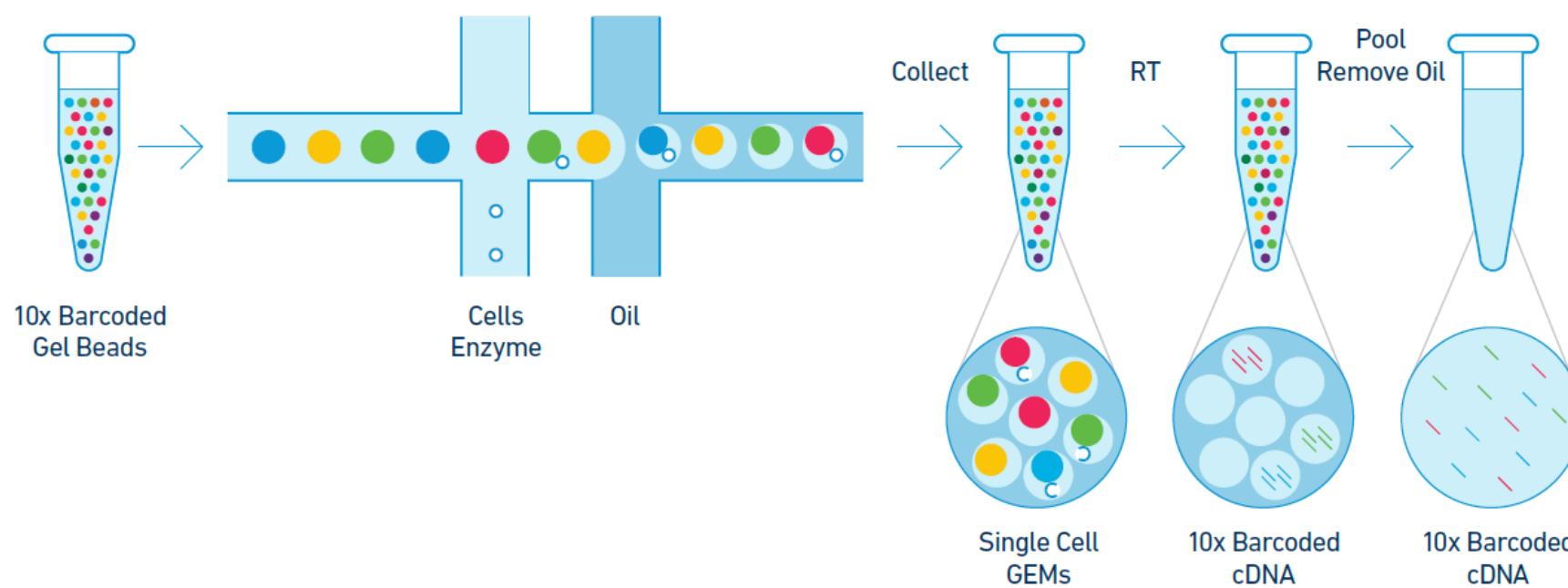
- Lysis and reverse transcription occurs in the beads
- Samples are frozen after RT as RNA:DNA hybrid in gel
- Library prep is based on CEL-Seq method

scRNA-seq Library Structure (inDrops)

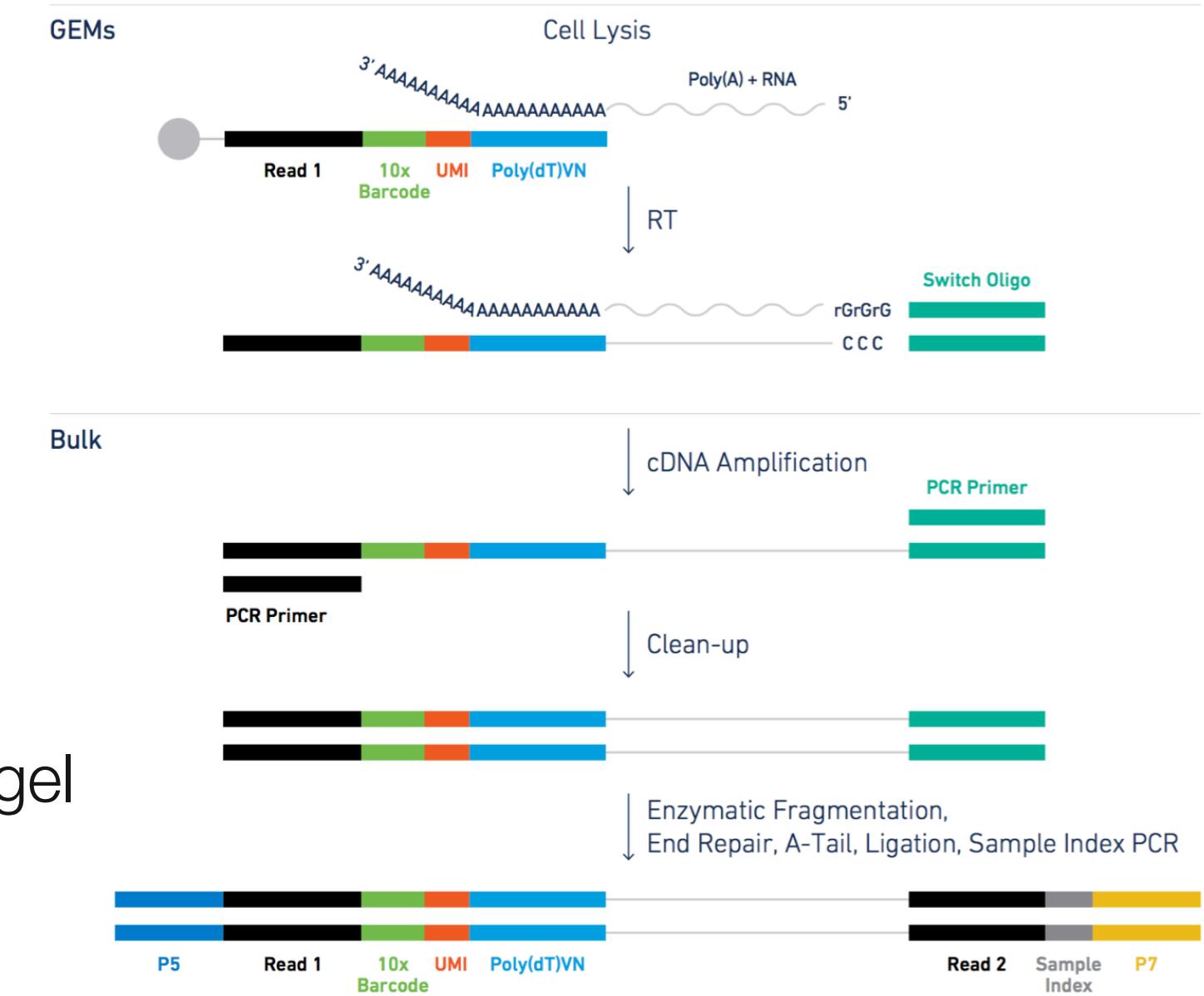


| Sequencing Read | Description | Number of Cycles | Notes |
|-----------------|---------------------------|------------------|--|
| Read1 | Insert (Transcript) | 61 | Can read longer into transcript if desired |
| I7 index | Single Cell Barcode | 8 | Reads first half of barcode |
| I5 index | Library Index | 8 | Distinguish samples |
| Read 2 | Single Cell Barcode & UMI | 14 | Reading longer will read into PolyA tail. |

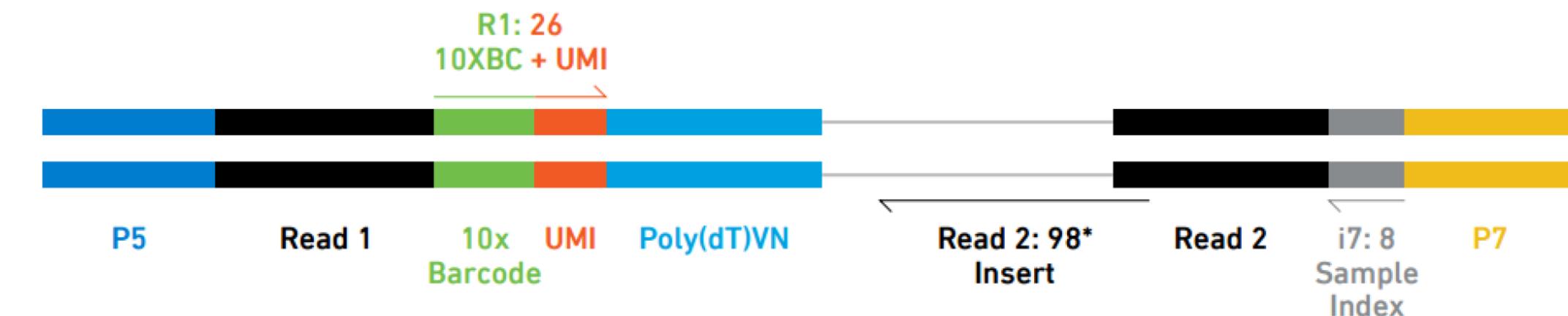
10x Genomics Method Overview



- Lysis and reverse transcription occurs in the beads
- Samples are frozen after RT as RNA:DNA hybrid in gel
- Library prep is similar to SMART-seq method



10x Genomics Method Overview



| Sequencing Read | Description | Recommended Number of Cycles | Notes |
|-----------------|---|------------------------------|---|
| Read 1 | 10x Barcode Read (Cell) and Randomer Read (UMI) | 26 cycles | It cannot be shorter than 26 bp If longer than 26 bp it will be ignored by Cell Ranger |
| i7 Index | Sample Index Read | 8 cycles | If longer than 8 bp it will need to be trimmed during base calling |
| i5 Index | N/A | 0 cycles | N/A |
| Read 2 | Insert Read (Transcript) | 98 cycles | Can be adjusted* |

*User controlled trade-off between read length and mapping rate

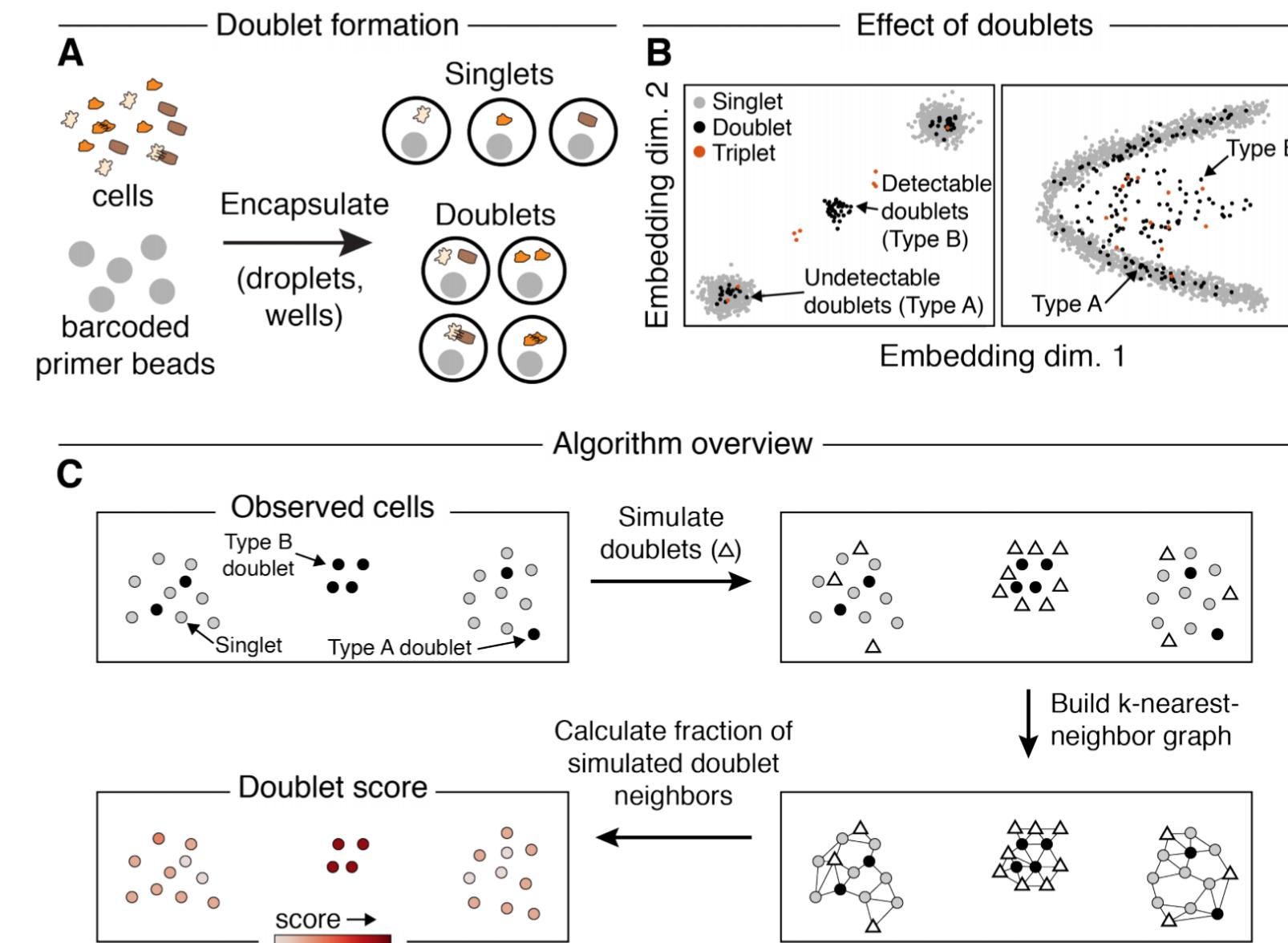
Doublets / Cell Density

- Rate of doublets depends on the cell density and the flow rate used for encapsulation.

| Multiplet Rate (%) | # of Cells Loaded | # of Cells Recovered |
|--------------------|-------------------|----------------------|
| ~0.4% | ~870 | ~500 |
| ~0.8% | ~1700 | ~1000 |
| ~1.6% | ~3500 | ~2000 |
| ~2.3% | ~5300 | ~3000 |
| ~3.1% | ~7000 | ~4000 |
| ~3.9% | ~8700 | ~5000 |
| ~4.6% | ~10500 | ~6000 |
| ~5.4% | ~12200 | ~7000 |
| ~6.1% | ~14000 | ~8000 |
| ~6.9% | ~15700 | ~9000 |
| ~7.6% | ~17400 | ~10000 |

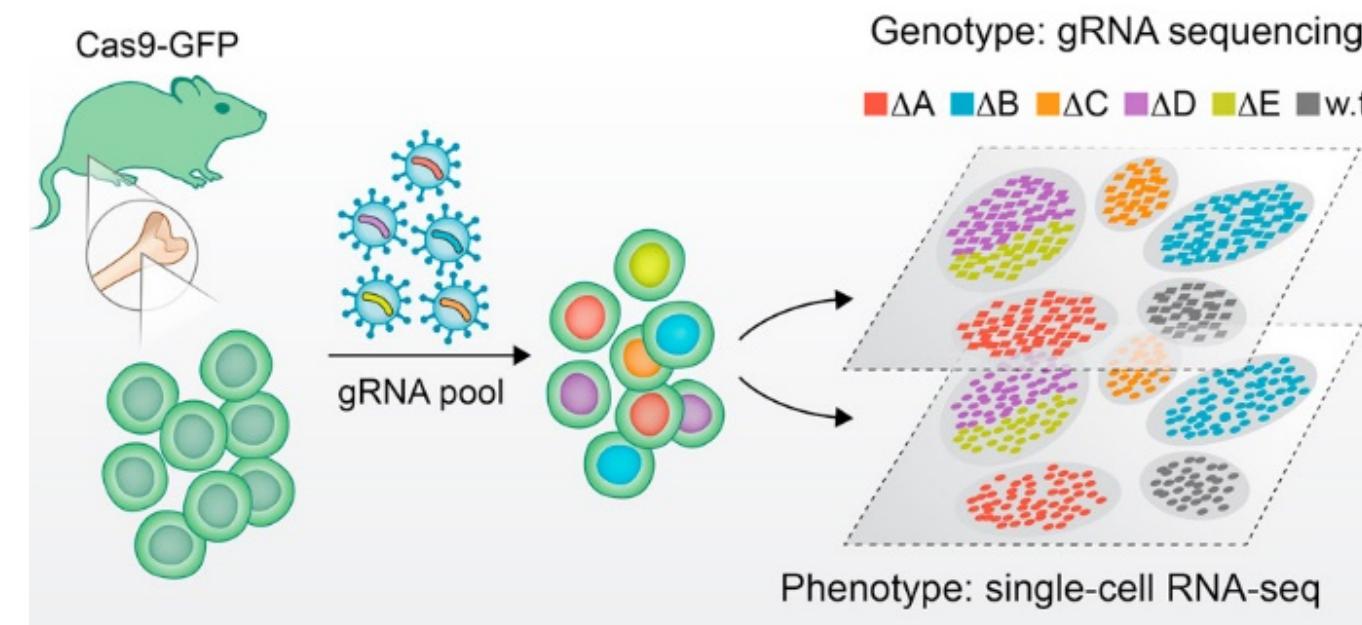
- Easy to claim low doublet rate for a particular method, but be aware of tunable parameters to accurately assess those statements.

Scrublet: Computational Identification of Doublets



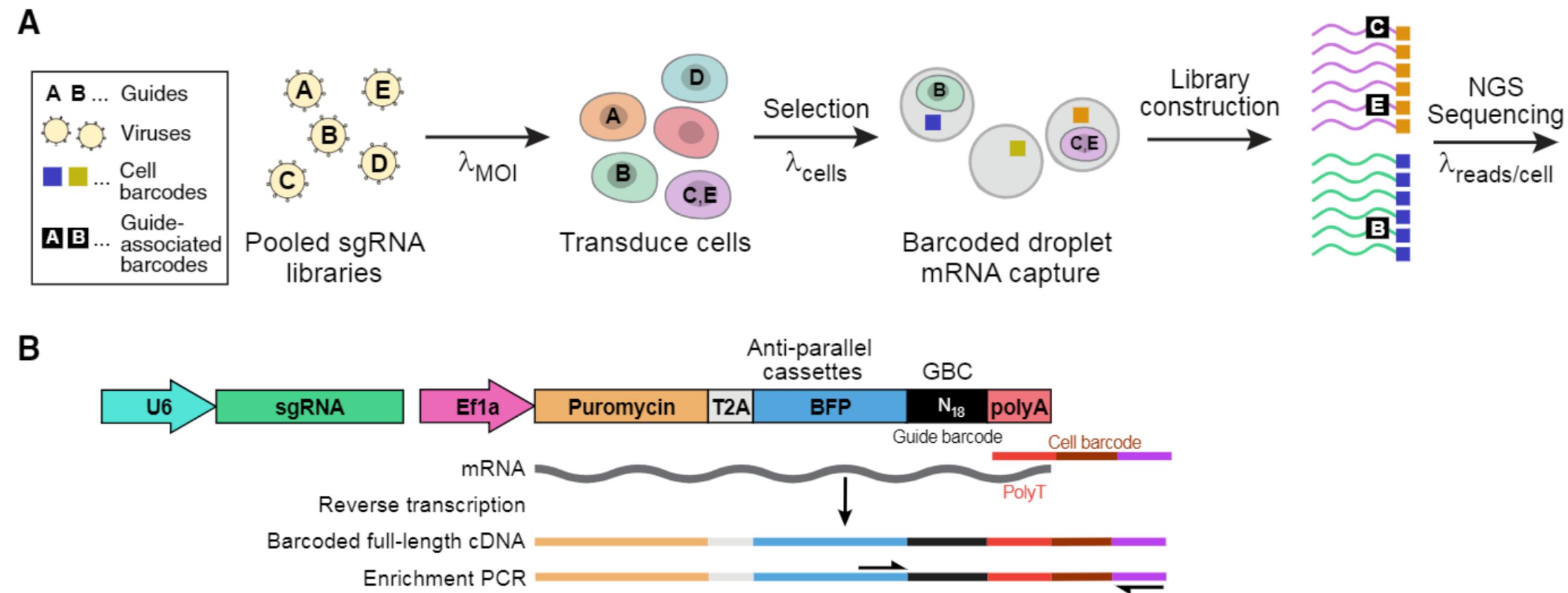
Transcript Specific Library Prep

- CRISPR pool vector backbone must contain a transcribed poly-adenylated unique guide index (UGI), which can include a fluorescent marker



- scRNA-seq library to phenotype cellular transcriptome (NextSeq/HiSeq)
- gRNA-targeted library to ensure proper cell identification (MiSeq)

Transcript Specific Library Prep



Transcript Specific Library Prep

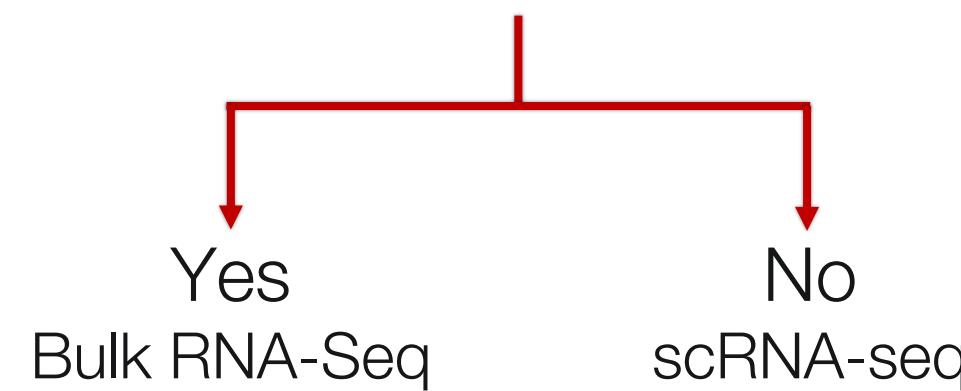
- Make standard library
 - NextSeq/HiSeq sequencing to identify cell barcodes in sample
- Make transcript specific library with aliquot of initial library
 - MiSeq to identify cells with transcripts of interest
- Match barcodes identified in both sequencing runs

Introduction to Single Cell RNA Sequencing

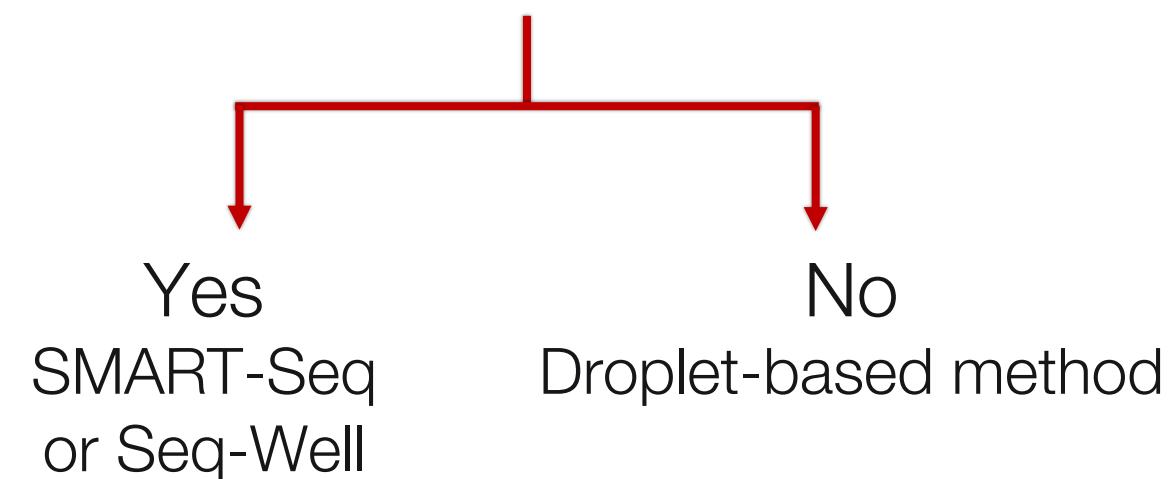
- Common applications of single cell RNA sequencing.
- Overview of inDrops and 10x platforms.
- Experimental design and sample preparation.
- Effects of sample prep and sample type on analysis.

Experimental Decision Making

Is your sample fairly homogeneous?

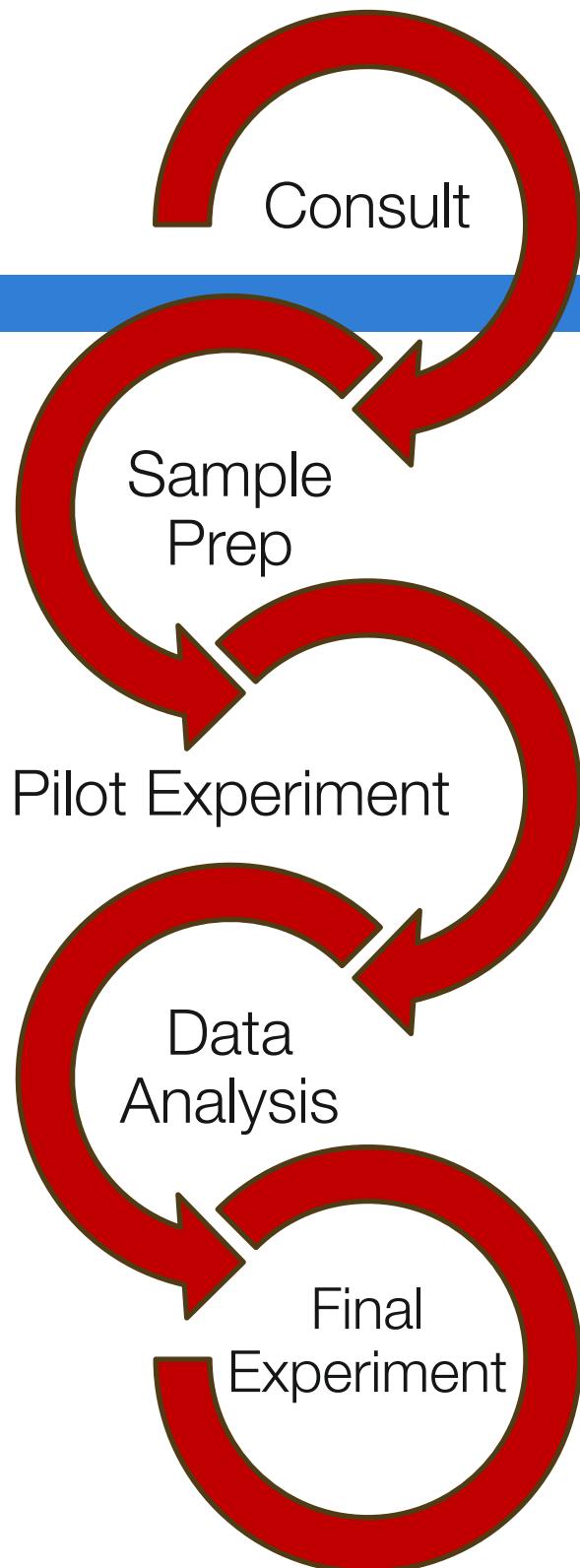


Do you want full length transcripts/splice variants?
Is our sample limited in cell number (<10,000 cells)?



Single Cell Core Workflow

- Good sample prep is the key to success.
- A well planned pilot experiment is essential for evaluating sample preparation and for understanding the required number of cells.
- 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

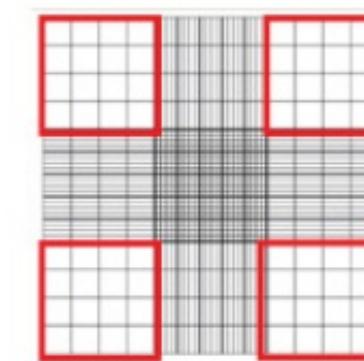
Sample Preparation: increasing cell viability

- Mild dissociation reagent (TrypLE, StemPro, Accutase, Liberase)
- Shorten dissociation time
- Reduce dissociation temperature
- Using ROCK inhibitor/ apoptosis inhibitor (esp. epithelial cells)
- Avoid cell pelleting
- Avoid FACS sorting on more fragile cell types
- Try magnetic activated cell sorting (MACS)



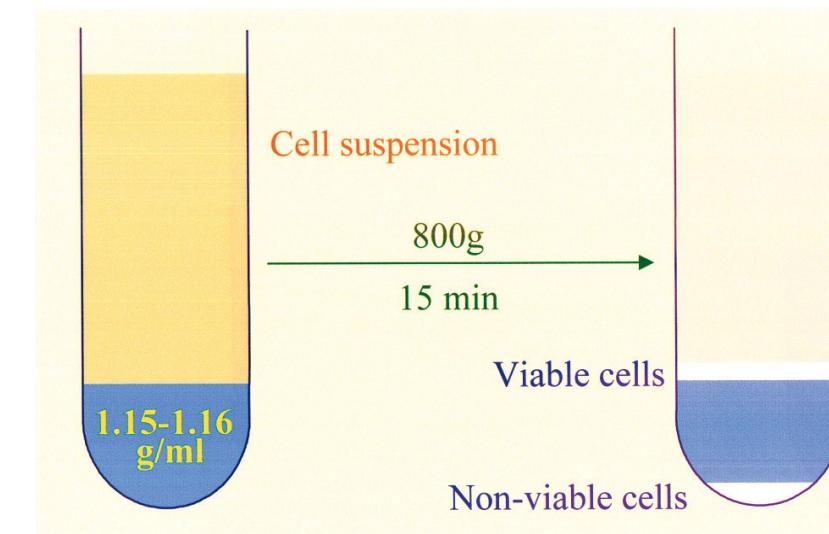
Sample Preparation: cell numbers

- Droplet methods have a 10,000-25,000 cell minimum
 - need ~50-100 cells with a unique transcriptome to identify a population cluster
 - 100-1000 cells per μl = 100,000-1,000,000 cells per ml
- Count cells by hemocytometer – do not trust sorter counts
 - counts from the sorter are often $\frac{1}{2}$ of actual cell counts
- Try negative selection to remove unwanted cells
- Sort on a broader marker to increase cell numbers
- For unavoidably low density samples
 - spike the sample with cells with distinct expression profile



Sample Preparation: single cell suspension

- Use appropriate cell strainer to get rid off clumps.
- Use metabolically inert, non-toxic density gradient media to fractionate cells.
- We routinely use 15% Opti-prep to keep cells in single cell suspension while loading sample, but the concentration might vary from cell to cell.



- Make sure final buffer does not contain calcium, EDTA, or heparin (inhibit RT).

Sample Preparation: buffers

- Suggest final sample prepared in:
 - PBS with 0.1-1% BSA
 - Defined media without calcium or EDTA
- 2% FBS in defined media without calcium or EDTA has worked for some users.
- Make sure buffer does not contain calcium, EDTA, or heparin (inhibit RT).
 - 10x chromium can not have >3mM magnesium.

Sample Preparation: viability checks

- Check viability of sample over time
 - If viability decreases over a short period of time this will be reflected in transcriptional data.
 - Will see high mitochondrial read counts.
- Check single cell suspension supernatant for the presence of free floating RNA (Ribogreen)
 - Creates background noise in all samples and complicates analysis.
- Number of trypan positive cells × number of wasted sequencing reads
 - If 30% of your cells are dead at the time of encapsulation then at most you will be able to use 70% of your sequencing data.

Sample Preparation: cryopreservation

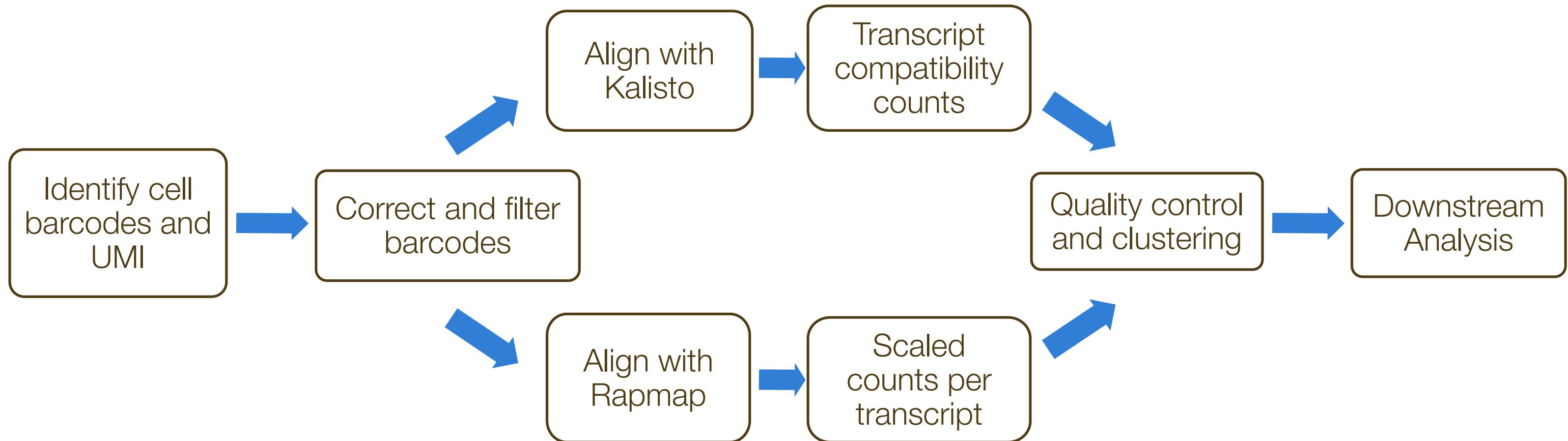
- Several papers have come out using various cryopreservation techniques on samples (PBMC's or cell lines).
- Success of cryopreservation is dependent on the sample type.
- Have seen this worked well on blood and immune cells.
- Key is the viability of the cells upon rehydration.
- Consider Nuc-Seq as an option from cryopreserved cells.

Alles, J. et al. Cell fixation and preservation for droplet-based single-cell transcriptomics. *BMC Biol.* 2017 (doi: 10.1186/s12915-017-0383-5)
Guillaumet-Adkins, A. et al. Single-cell transcriptome conservation in cryopreserved cells and tissues. *Genome Biol.* 2017 (doi: 10.1186/s13059-017-1171-9)
Habib, N. et al. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* 2017 (doi: 10.1038/nmeth.4407)

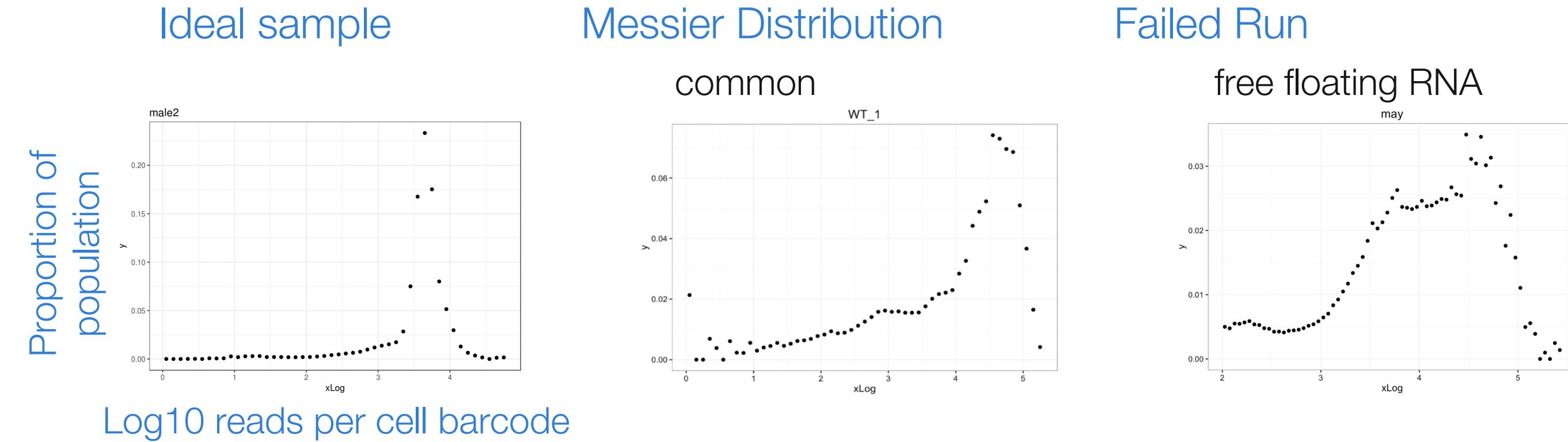
Introduction to Single Cell RNA Sequencing

- Common applications of single cell RNA sequencing.
- Overview of inDrops and 10x platforms.
- Experimental design and sample preparation.
- Effects of sample prep and sample type on analysis.

Single Cell Data Analysis

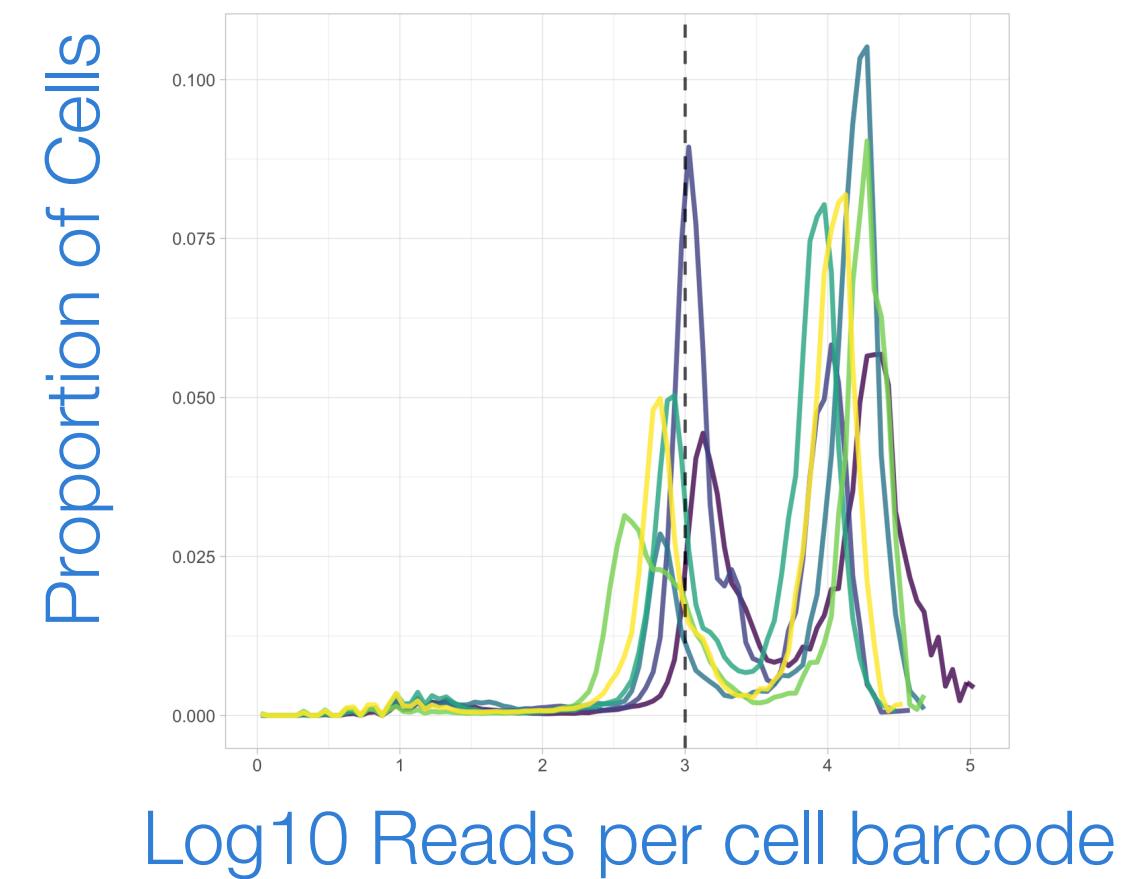


How Sample Prep Effects Data



- Cut off usually remove any cell with < 10,000 / 20,000 barcodes per cell
- It is normal for single cell RNA-seq data to contain a large number of low complexity barcodes.
- Exact threshold will depend on sample

How Sample Type can Effect Data



- Bimodal peak is due to sample type. (infiltrating immune cells in tumor)
- Lower peak can get filtered out in analysis.
- May want to analyze each peak separately.

Data Analysis: Quality Control (QC) metrics

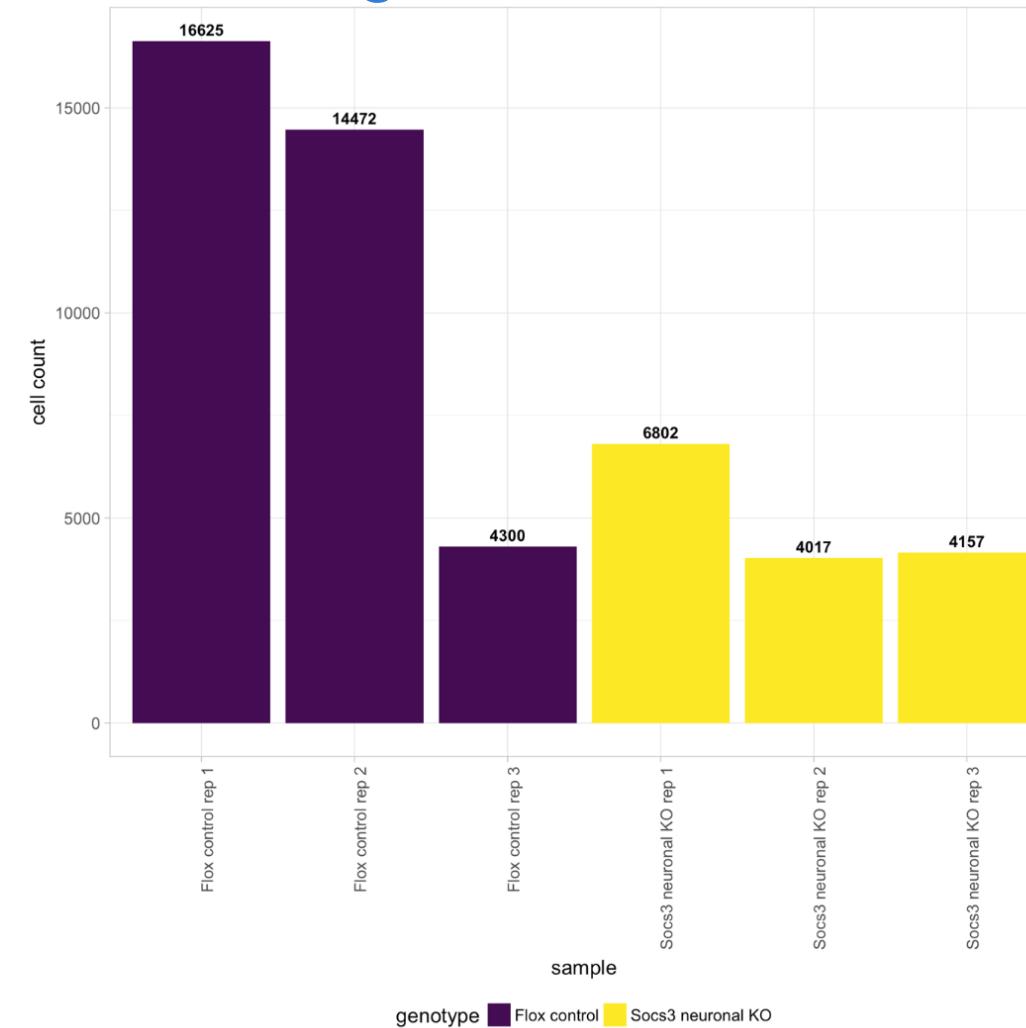
- Reads per cell: How many reads assigned to a given cell barcode
- UMI per cell: “Novelty” score looks for greater diversity genes per UMI
- Genes detected: Genes with a non-zero count measurement per cell
- Mitochondrial counts ratio: Biomarker for cellular stress

Filter parameters

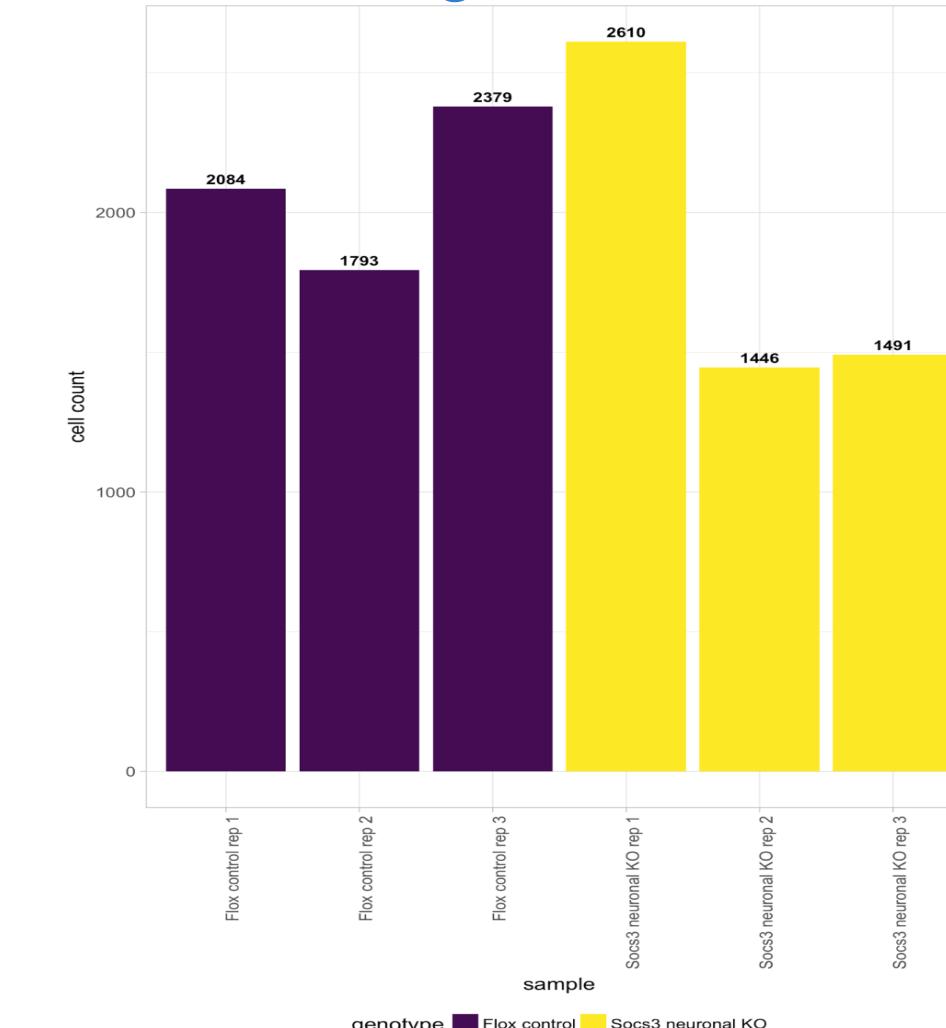
- `>= 500` UMI counts per cell
- `>= 500` genes per cell
- `<= 5000` genes per cell
- `<= 0.1` relative mitochondrial abundance
- `>= 0.8` novelty score

Data Analysis: filtering & correction

Pre-Filtering



Post-Filtering



Libraries were of 3,000 cells. Post-filtering retains 50-80% of cells

Final thoughts on scRNA-seq

- Practice your sample prep protocol. KEY to SUCCESS
- Start with a pilot sample set to ensure your protocol is working.
- Do not make your scRNA-seq run day the first day you run through the whole protocol.
- Be sure sequencing core understands the specific sequencing parameters needed for your scRNA-seq library.

qPCR

Precise quantitation is key to good clustering / sequencing