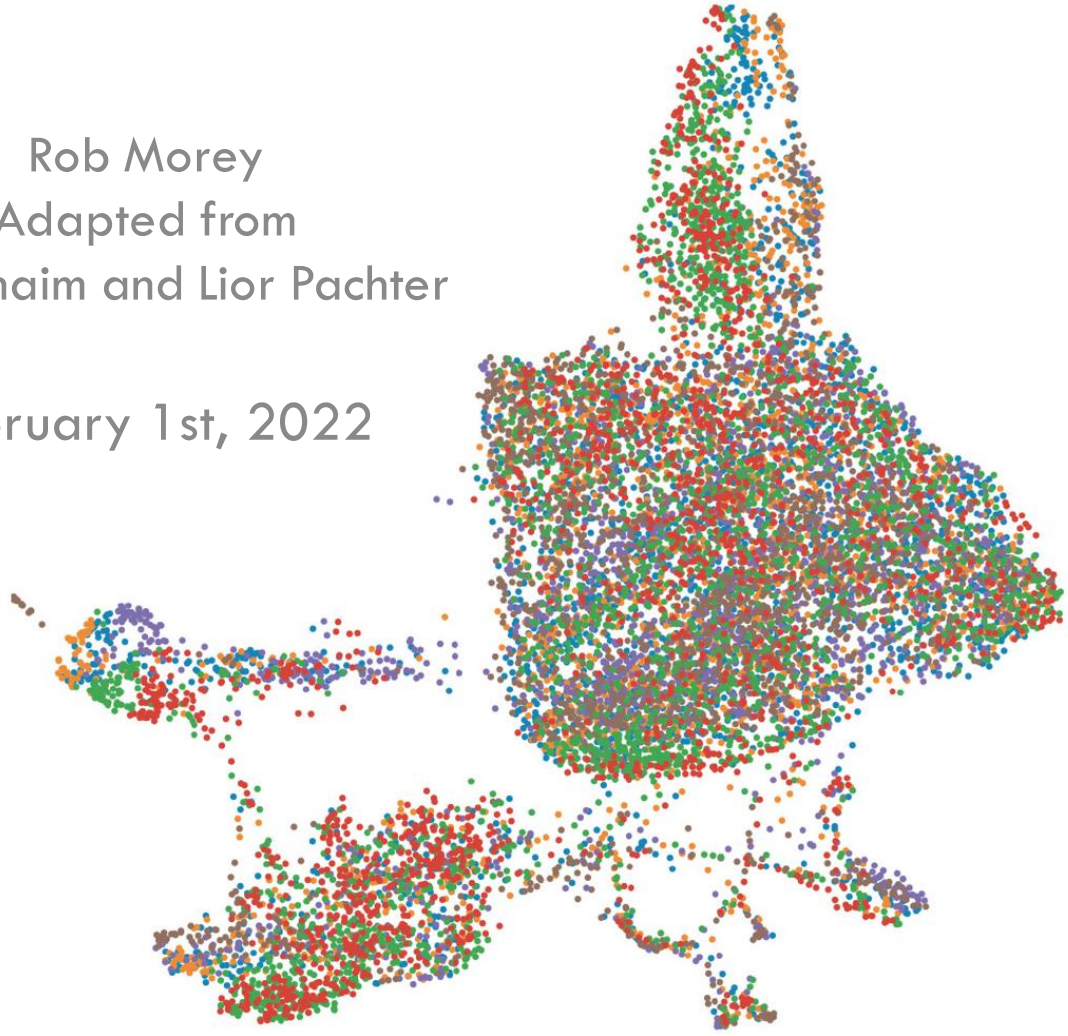
Quantitative Methods in Genetic and Genomics

# Single Cell Module

Rob Morey
Adapted from
Alex Chaim and Lior Pachter

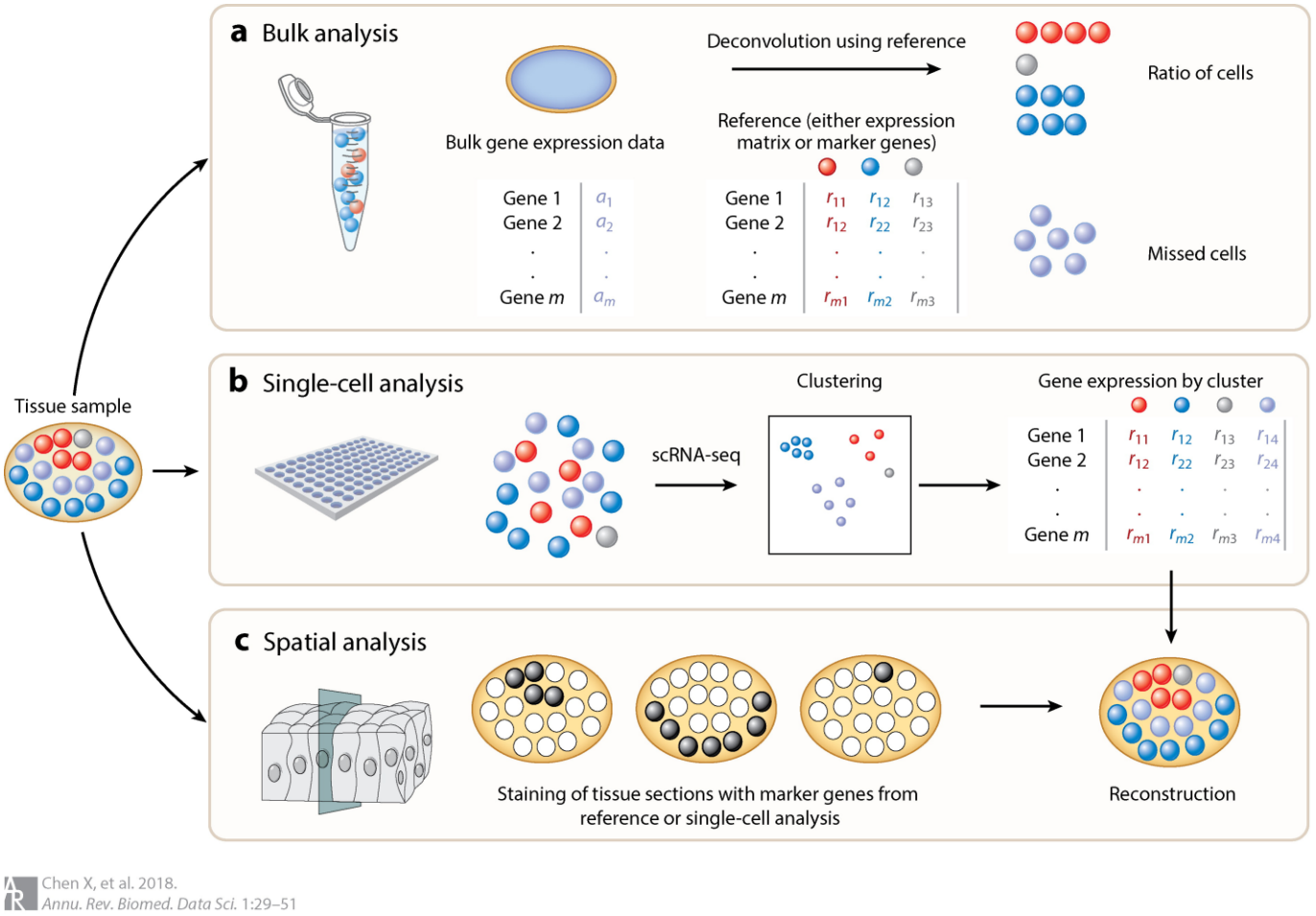
February 1st, 2022



# Module Overview

- Day 1:
  - Single cell RNA-seq methods/biochemistry (lecture)
  - Break
  - Analysis methods and options (lecture)
- Day 2:
  - Data analysis using Liu et al., "*Reprogramming roadmap reveals route to human induced trophoblast stem cells*"; Nature 2020

# Why single cell transcriptomics?

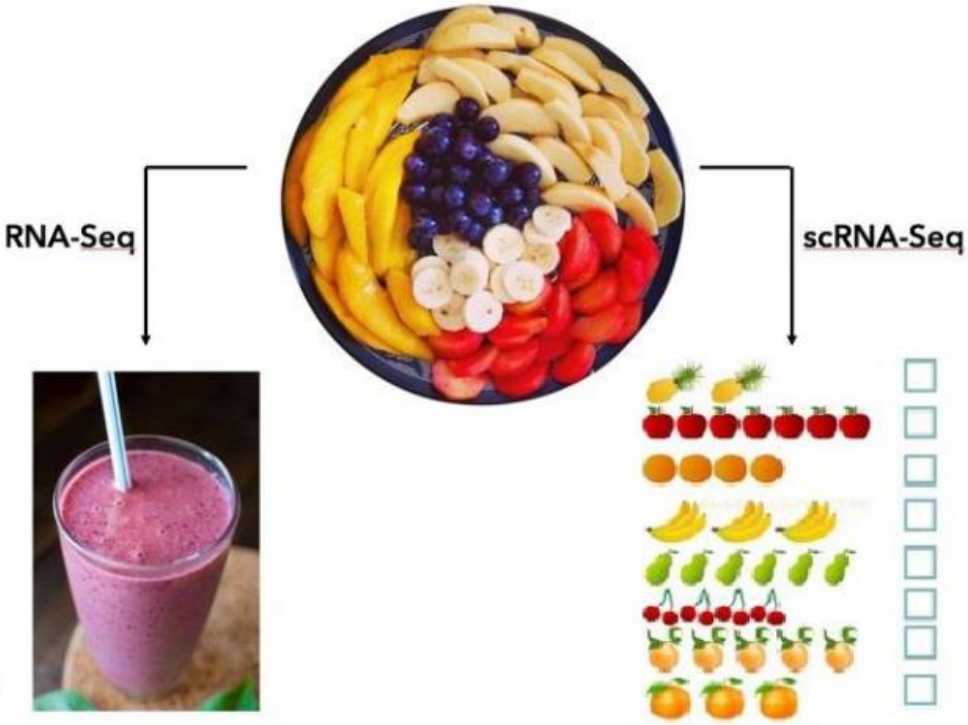



Chen X, et al. 2018.
*Annu. Rev. Biomed. Data Sci.* 1:29–51

Microarrays and bulk RNA sequencing mask the true distribution of gene expression levels across cells.

# BULK VS SINGLE CELL RNA-SEQ

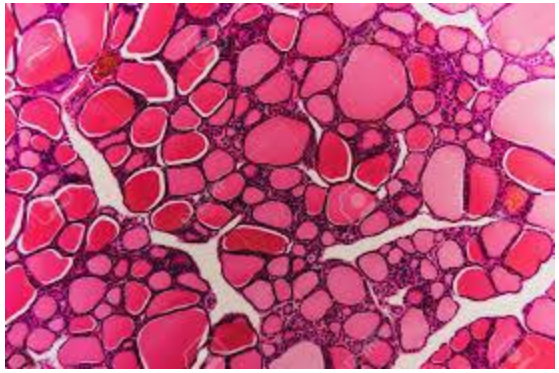
Average expression level
- Comparative transcriptomics
- Disease biomarker
- Homogenous systems




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

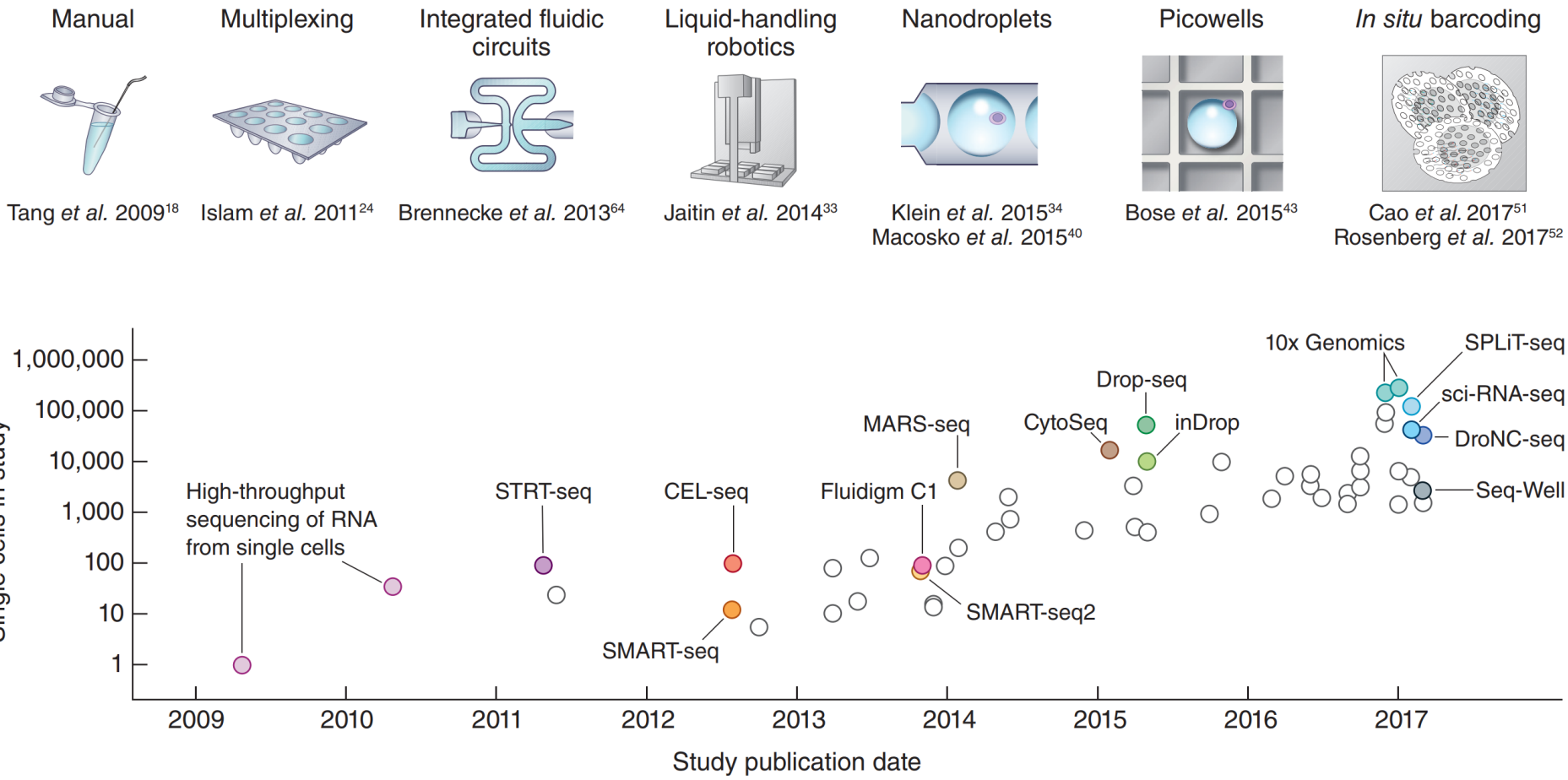
CANCER RESEARCH UK | CAMBRIDGE INSTITUTE

# How do we go from cell to count matrix?



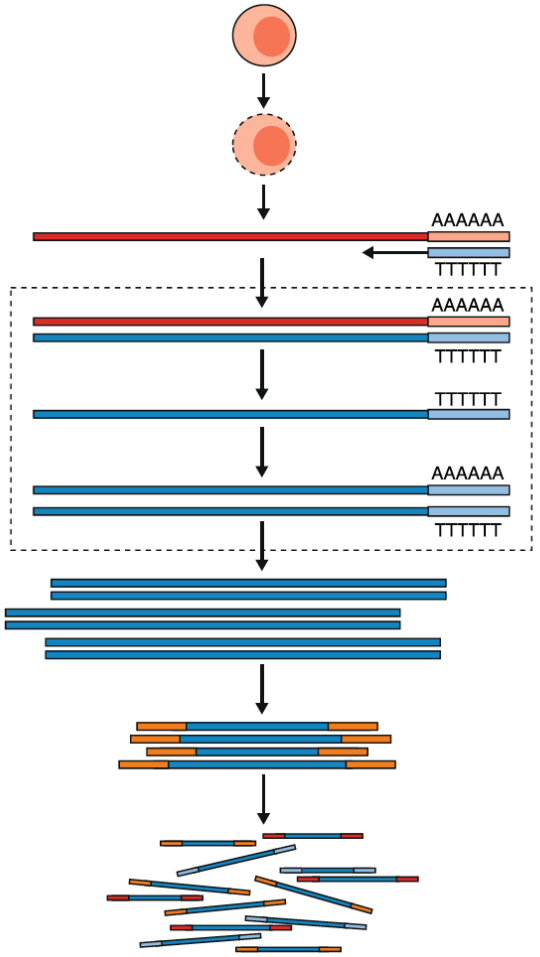
| Cell: | 1 | 2 | ··· | N |
|---|---|---|---|---|
| GENE 1 | 1 | 2 | | 14 |
| GENE 2 | 4 | 27 | | 8 |
| GENE 3 | 0 | 0 | | 1 |
| ⋮ | ⋮ | ⋮ | | ⋮ |
| GENE M | 6 | 2 | | 0 |

# A decade of single-cell RNA-seq




Svensson et al. 2018. DOI:10.1038/nprot.2017.149

# General workflow of a single cell (sc) RNA-Seq experiment

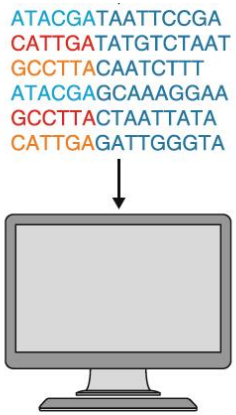



Critical Steps

1. Cell Isolation (cell or nuclei)
2. Cell lysis
3. mRNA capture

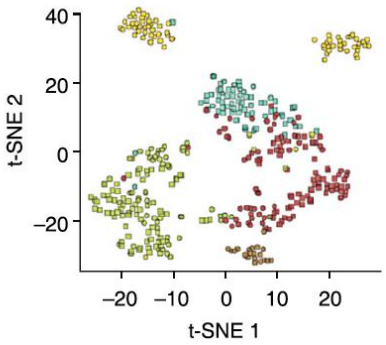
Cell Barcoding

4. Second Strand Generation
5. cDNA amplification
6. Library Construction
7. Pool and sequence




8

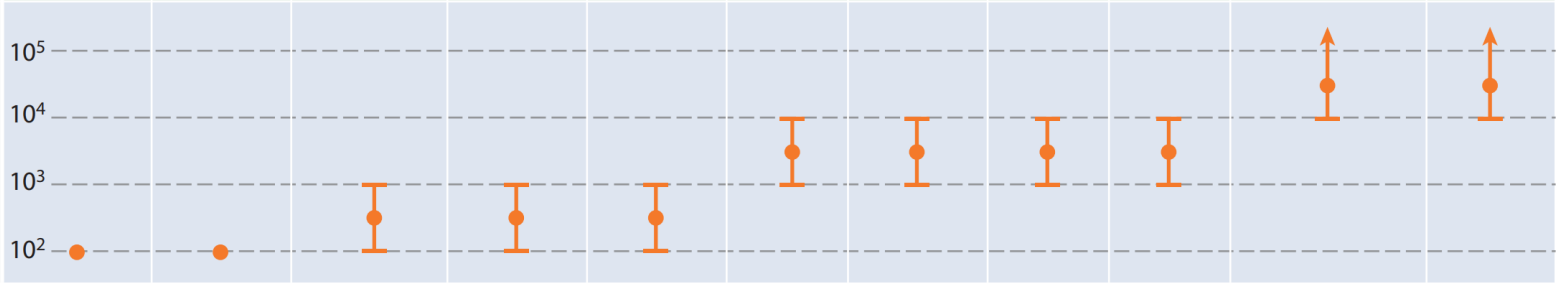
Analysis




9

Haque *et al.*, 2017 Genome Medicine

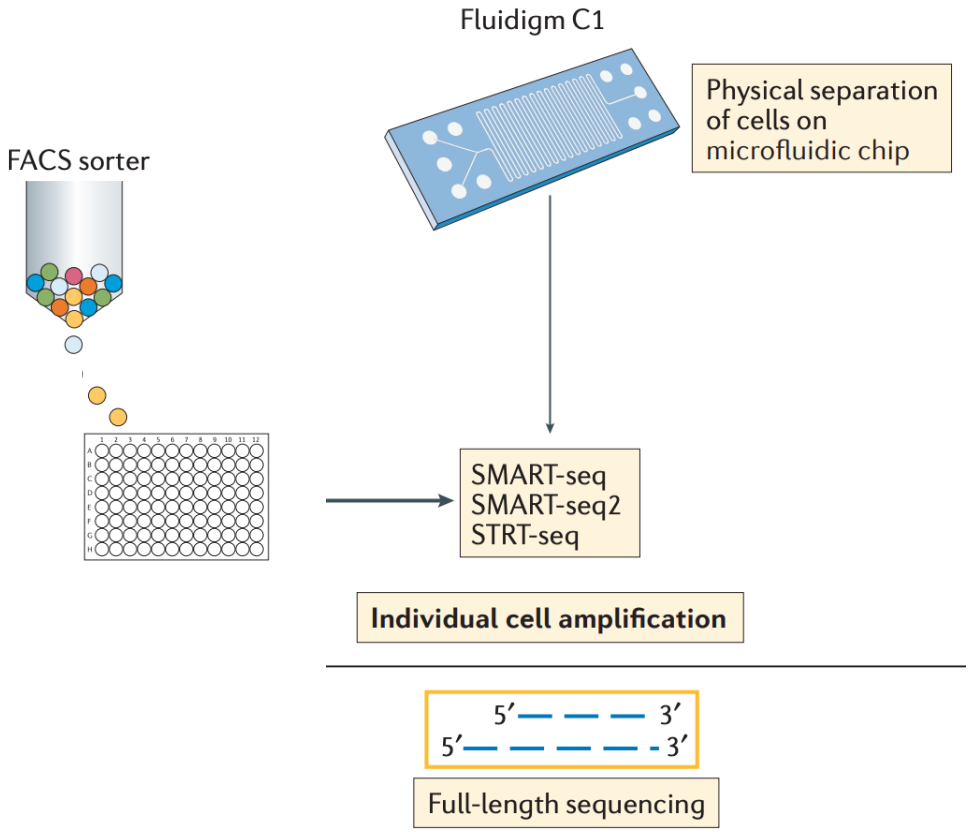
# Popular single-cell RNA-seq protocols

| | SMART-seq2 | CEL-seq2 | STRT-seq | Quartz-seq2 | MARS-seq | Drop-seq | inDrop | Chromium | Seq-Well | sci-RNA-seq | SPLiT-seq |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Single-cell isolation** | FACS, microfluidics | FACS, microfluidics | FACS, microfluidics, nanowells | FACS | FACS | Droplet | Droplet | Droplet | Nanowells | Not needed | Not needed |
| **Second strand synthesis** | TSO | RNase H and DNA pol I | TSO | PolyA tailing and primer ligation | RNase H and DNA pol I | TSO | RNase H and DNA pol I | TSO | TSO | RNase H and DNA pol I | TSO |
| **Full-length cDNA synthesis?** | Yes | No | Yes | Yes | No | Yes | No | Yes | Yes | No | Yes |
| **Barcode addition** | Library PCR with barcoded primers | Barcoded RT primers | Barcoded TSOs | Barcoded RT primers | Barcoded RT primers | Barcoded RT primers | Barcoded RT primers | Barcoded RT primers | Barcoded RT primers | Barcoded RT primers and library PCR with barcoded primers | Ligation of barcoded RT primers |
| **Pooling before library?** | No | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| **Library amplification** | PCR | In vitro transcription | PCR | PCR | In vitro transcription | PCR | In vitro transcription | PCR | PCR | PCR | PCR |
| **Gene coverage** | Full-length | 3′ | 5′ | 3′ | 3′ | 3′ | 3′ | 3′ | 3′ | 3′ | 3′ |
| **Number of cells per assay** |  | | | | | | | | | | |



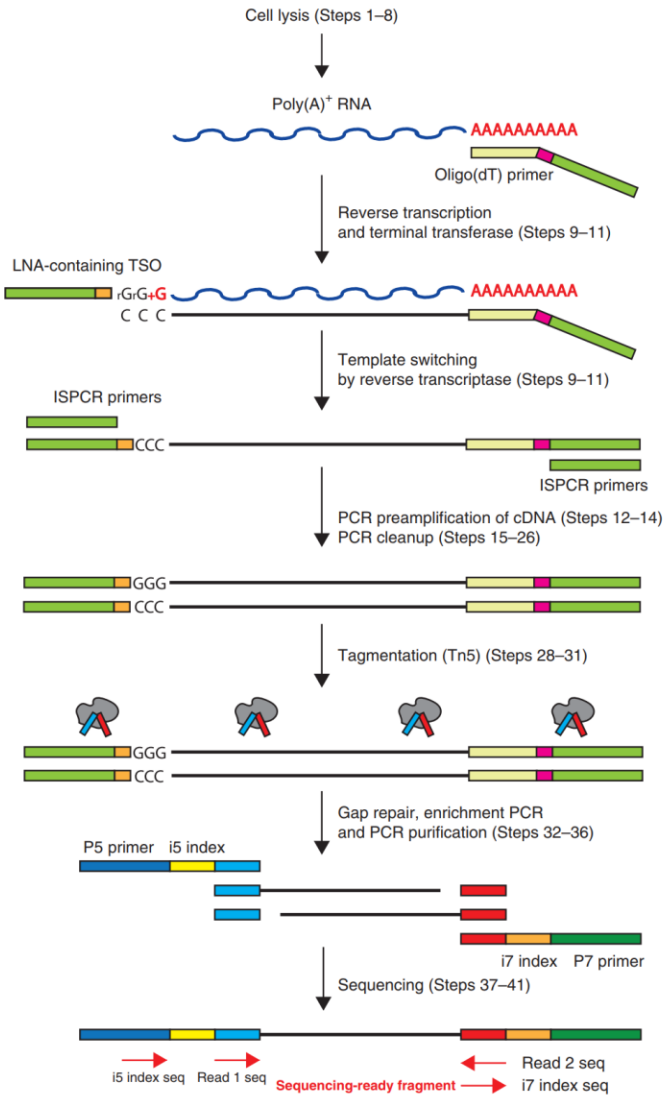
Chen et al. 2018. DOI:10.1146/annurev-biodatasci-080917-013452

# Droplet vs. Physical
# Physical separation of cells in wells




Papalexi and Satija, 2017. DOI:10.1038/nri.2017.76

# Example: library preparation for SMART-Seq2




▲ **CRITICAL STEP** All the experiments must be performed under a UV-sterilized hood with laminar flow, and all the surfaces must be free from RNase to prevent degradation of RNA and from DNA to prevent cross-contamination from previous samples. The hood must be used only for single-cell experiments up to (but excluding) the cDNA amplification step (Step 12). An ideal scenario would be to place the hood in a separate room with a positive air pressure to prevent any contaminants from being carried inside, where they might affect the experiments. The room should be equipped with a garmenting area in which the user changes into a fresh disposable lab coat, hair net, dust mask, shoe covers and vinyl gloves (powder-free).

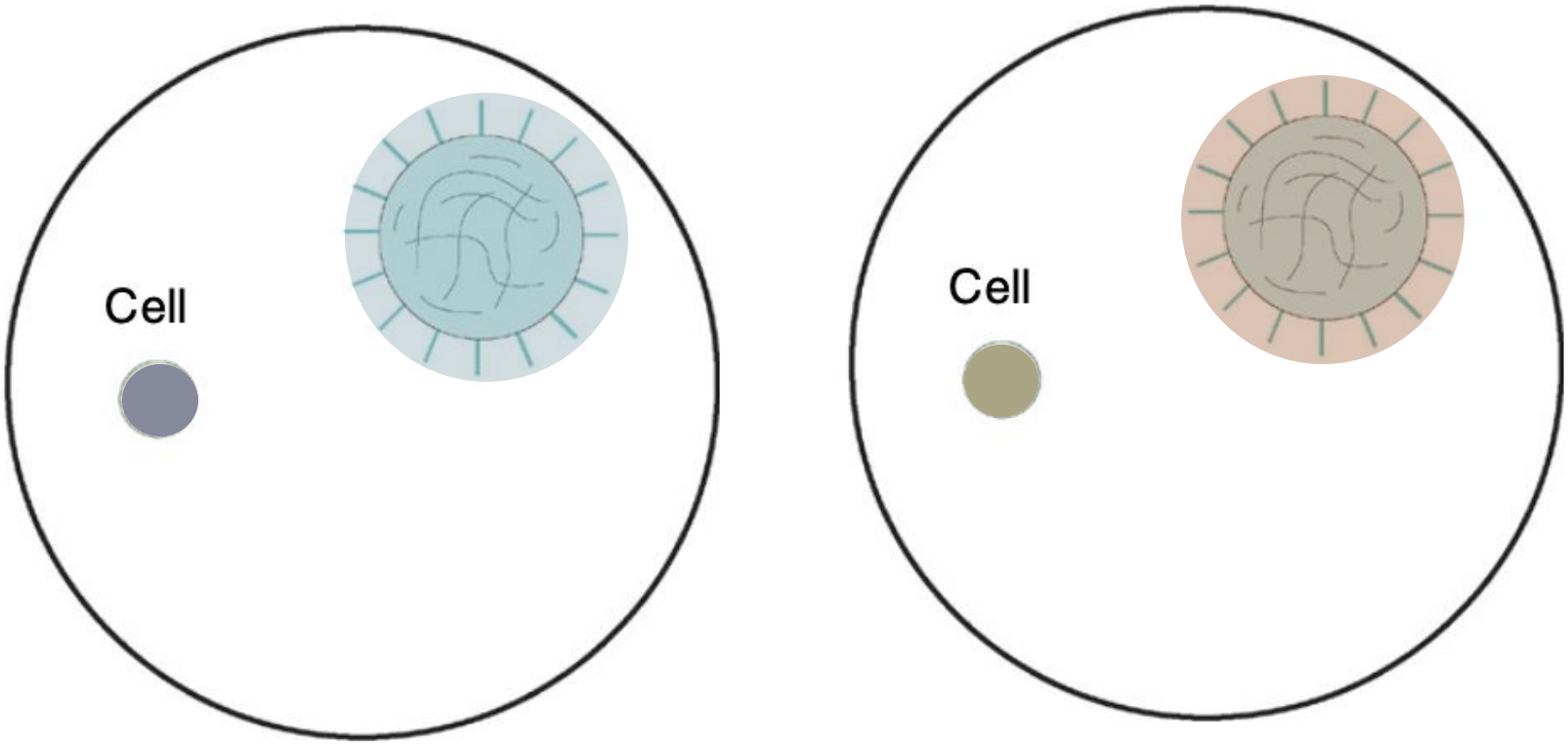
▲ **CRITICAL STEP** Thaw all the reagents in advance and assemble the RT mix while performing denaturation (Step 7) to minimize bias.

▲ **CRITICAL STEP** The number of PCR cycles depends on the input amount of RNA. We typically use 18 cycles for single eukaryotic cells to obtain ~1–30 ng of amplified cDNA. The number of cycles can be increased for smaller cells (with less RNA content) or lowered for large cells (with more RNA).

▲ **CRITICAL STEP** The number of cycles depends on the amount of DNA used for tagmentation. If we are starting from 100 pg of amplified cDNA, we usually perform 12 PCR cycles. The optimal number of cycles depends on the sample and the experiment. It may be helpful to run a range of cycles to determine the best conditions. Above are cycling guidelines on the basis of the input DNA used for tagmentation.

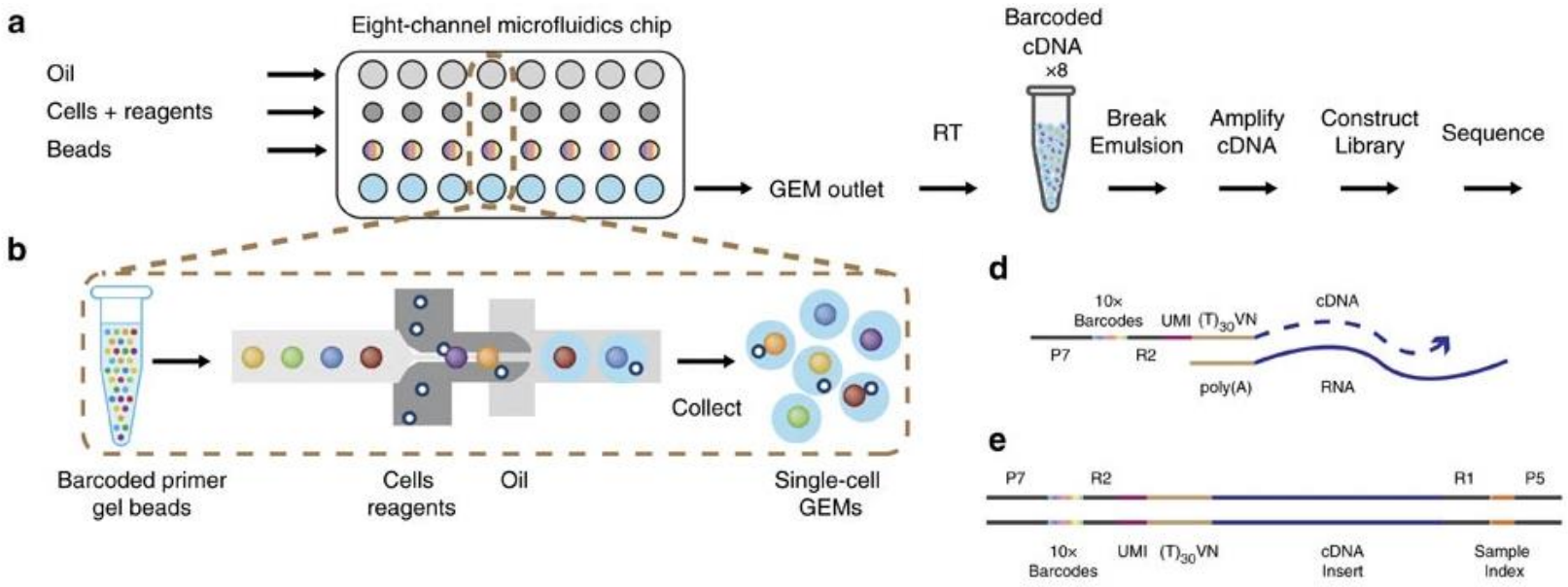
Picelli et al., 2014. DOI:10.1038/nri.2017.76

# Droplet vs. Physical
# Microfluidic methods: Beads, Cells and Droplets



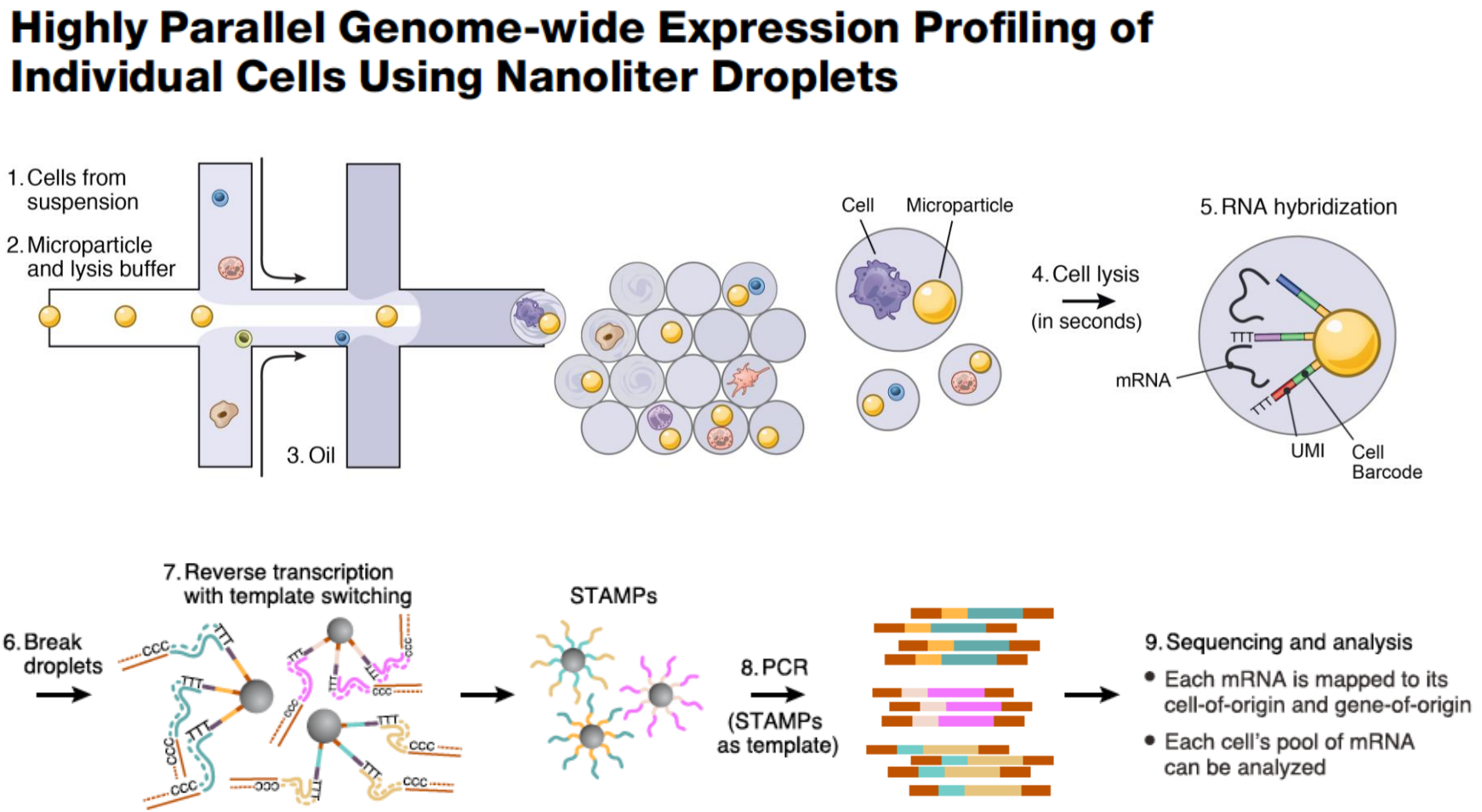

Zilionis et al. 2016. DOI:10.1038/nprot.2016.154

# 10X Genomics Overview



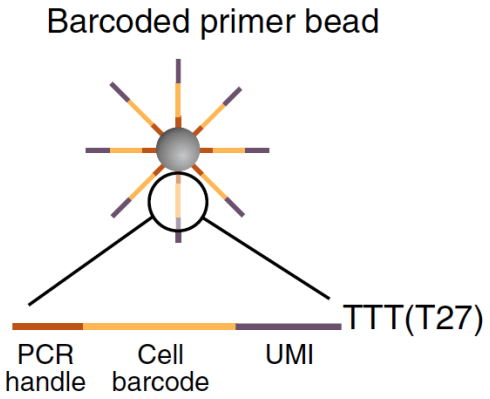

Zheng *et al*., 2017 Nature Communications

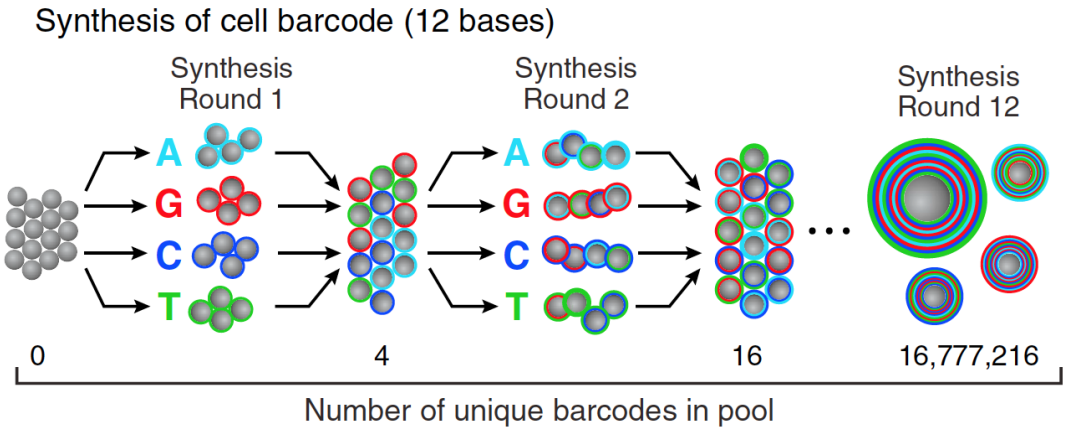
# Drop-Seq



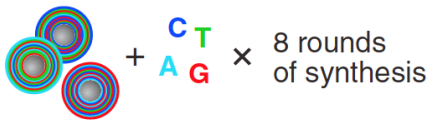

Macosko *et al*., 2015 Cell

# 3’ end Capture Beads - UMIs and Bead Barcodes


Synthesis of UMI (8 bases)




- Millions of the same **cell barcode** per bead
- $4^8$ different **molecular barcodes** (UMIs) per bead

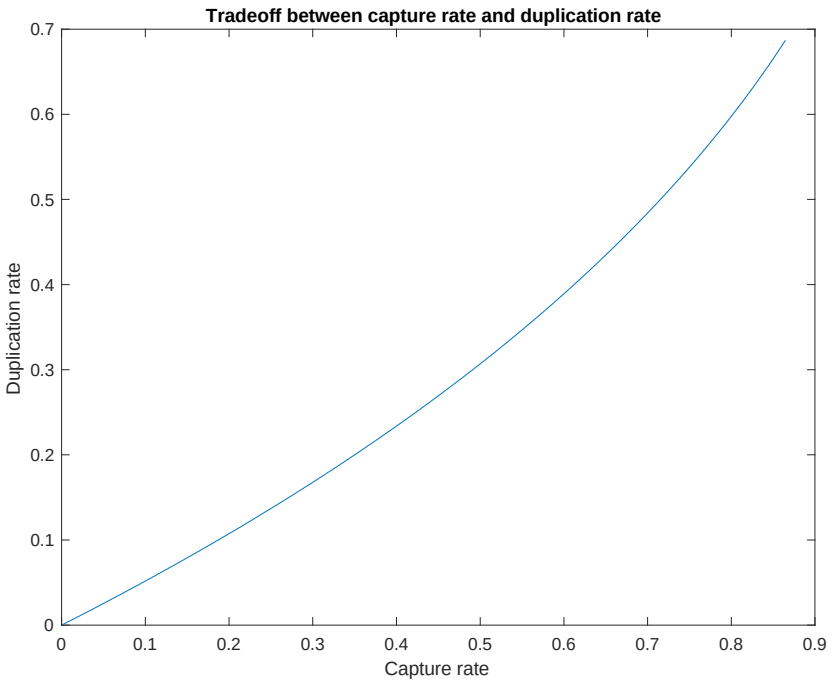
- $4^{12} = 16{,}777{,}216$ barcodes
- $4^8 = 65{,}536$ UMIs
- $> 10^6$ oligos/bead

Macosko *et al*., 2015 Cell

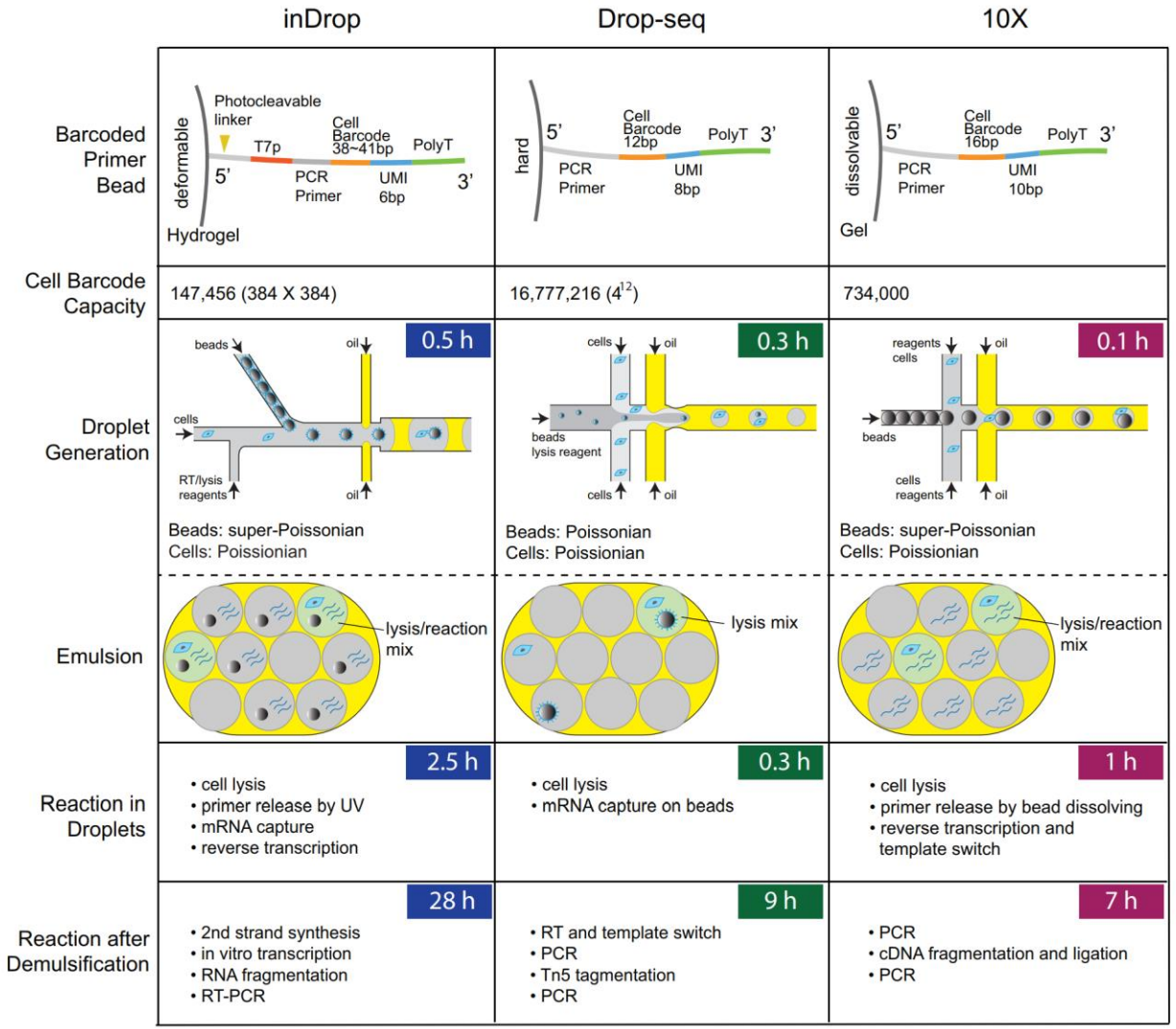
$$\frac{(1 - e^{-\mu}\mu - \mu e^{-\mu})}{1 - e^{-\mu}}$$

# Cell capture and duplication rates

- The cell capture rate is the probability that a droplet has at least one bead = .
- The cell duplication rate is the rate at which captured single cells are associated with two or more different barcodes, which is equal to .
- This leads to a tradeoff:



# Summary of droplet technologies



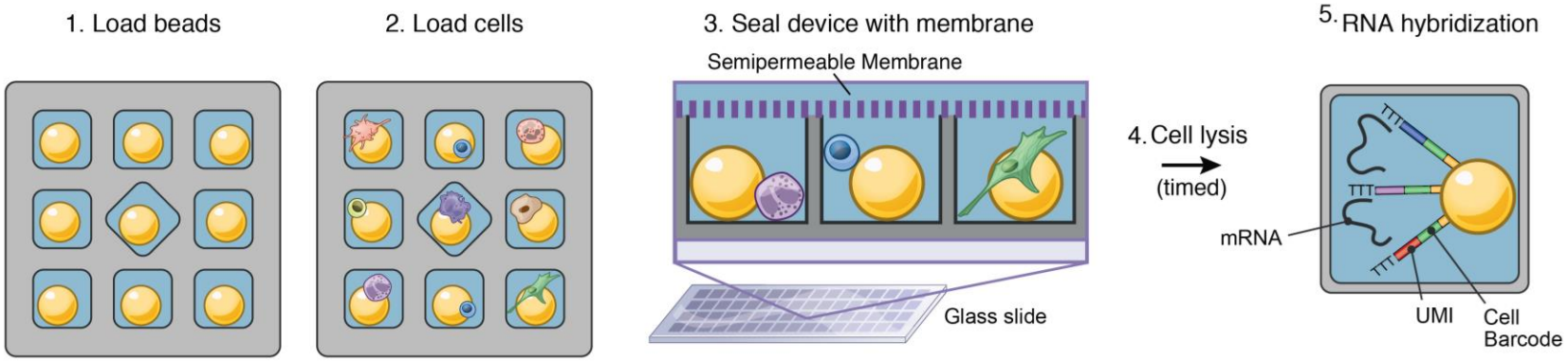

Zhang et al. 2019. DOI:10.1016/j.molcel.2018.10.020

# Seq-Well abrogates the need for droplet-based microfluidic devices

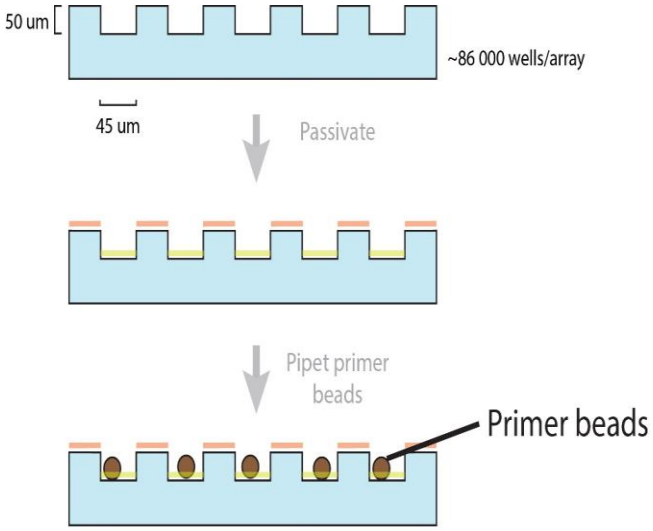
## Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput

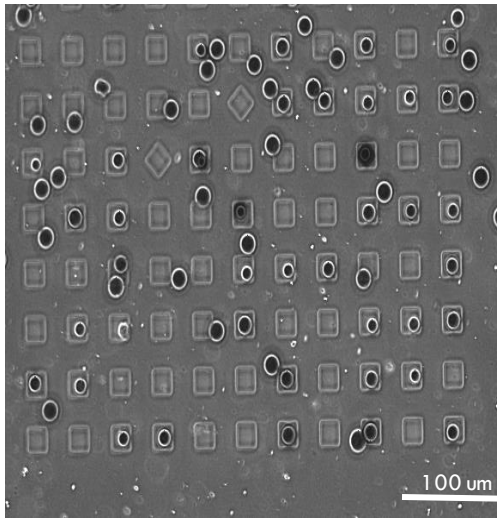
NATURE METHODS | VOL.14 NO.4 | APRIL 2017 |

Todd M Gierahn[1,8], Marc H Wadsworth II[2–4,8], Travis K Hughes[2–4,8], Bryan D Bryson[4,5], Andrew Butler[6,7], Rahul Satija[6,7], Sarah Fortune[4,5], J Christopher Love[1,3,4,9] & Alex K Shalek[2,3,4,9]

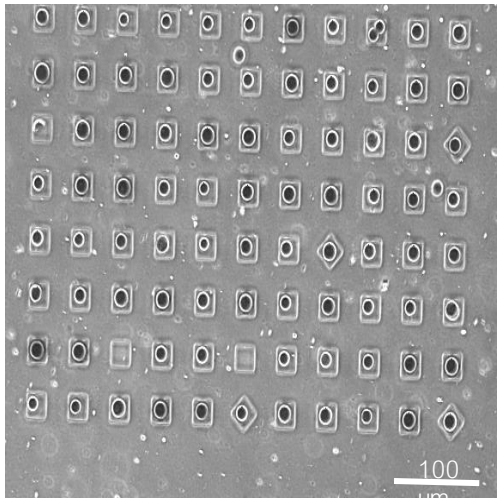


# Seq-Well workflow in Images



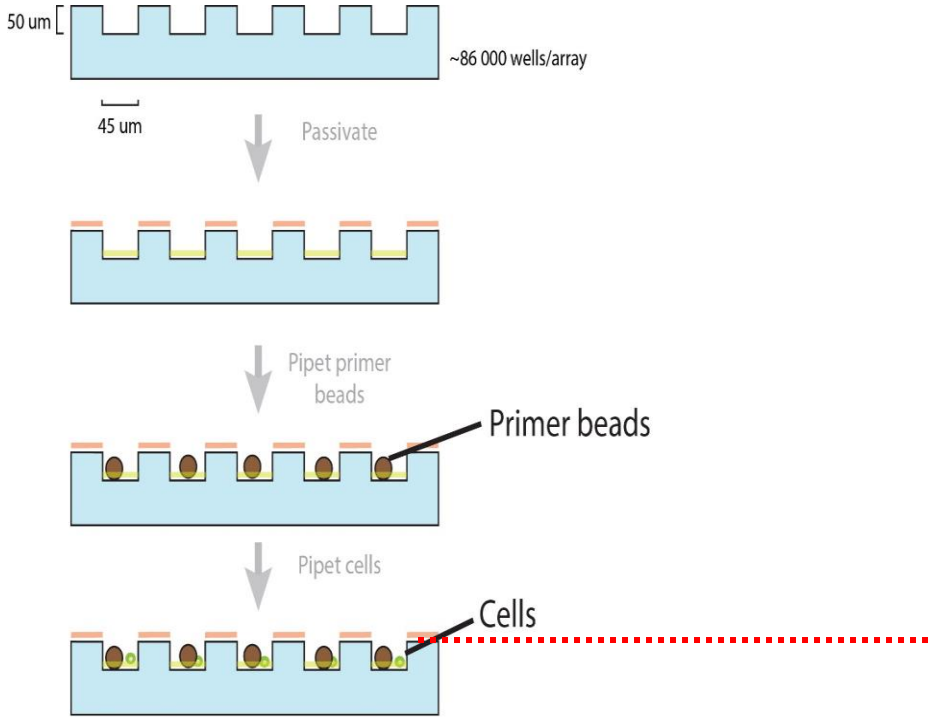



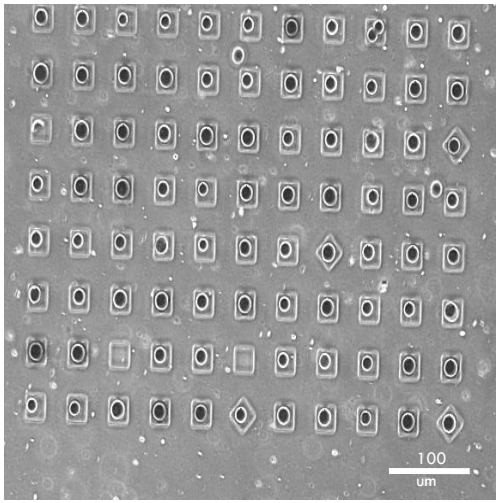

Pre-wash



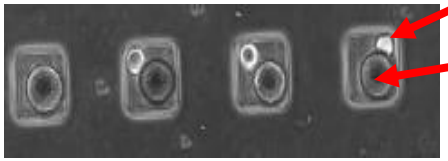

Post-wash + shake

# Seq-Well workflow in Images



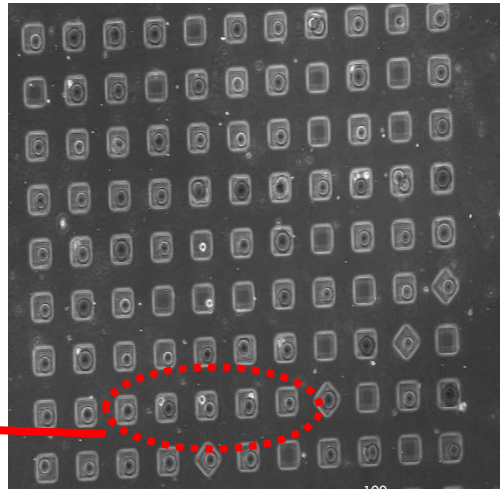




Pre-cell loading



Cell

Bead



Post-cell loading + shake

Capture rate > 50%

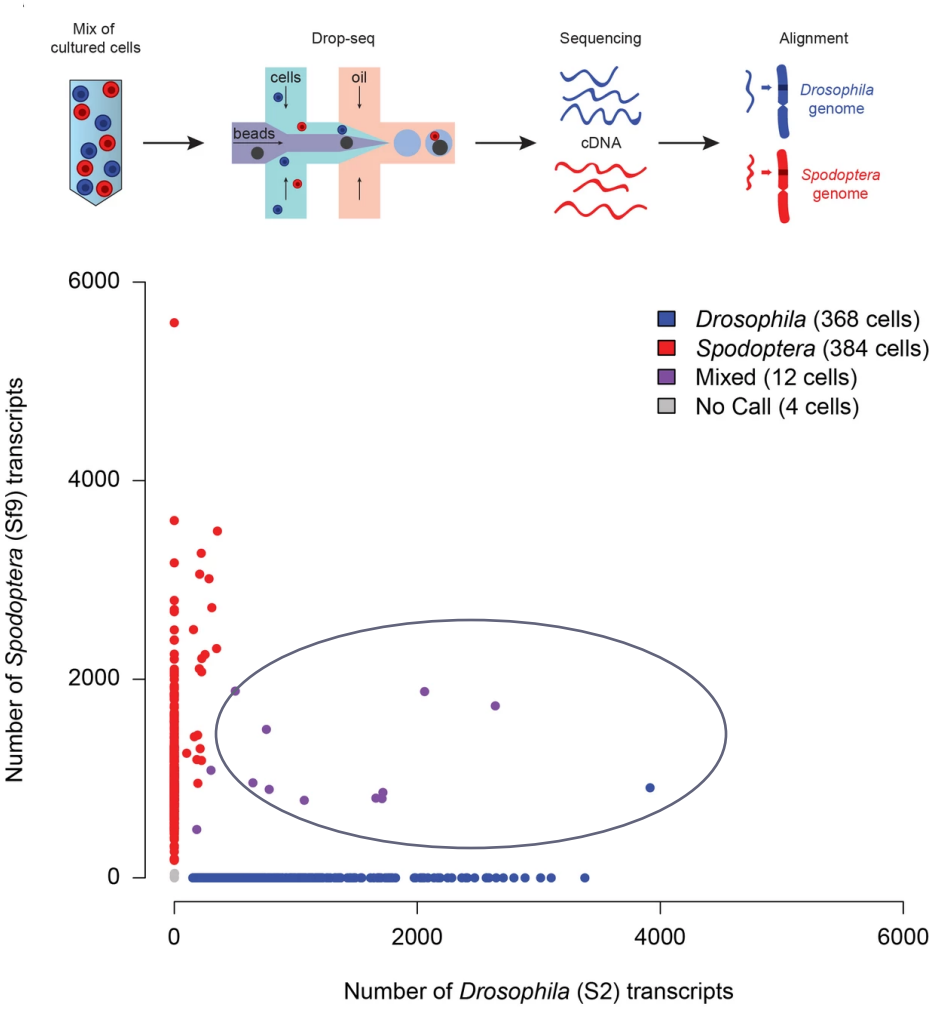
$$\frac{\lambda(1 - e^{-\lambda} - \lambda e^{-\lambda})}{1 - e^{-\lambda}}.$$

# Common problems - Technical doublets

- **Technical doublets** arise when two or more cells are captured in a droplet with a single bead. The technical doublet rate is therefore the probability of capturing two or more cells in a droplet given that at least one cell has been captured in a droplet:

- Note that “overloading” a microfluidics single-cell experiment by loading more cells while keeping flow rates constant will increase the number of technical doublets due to an effective increase in and also the number of synthetic doublets due to a decrease in relative barcode diversity.

# Doublet detection: the barnyard plot




Croset et al. 2018. DOI:10.7554/eLife.34550


# Which is better – full length or 3' end?

# Single cell technologies (3'-end vs full-length)

**Table 1. Cost Efficiency Extrapolation for Single-Cell RNA-Seq Experiments**

| Method | TPR[a] | FDR[a] (%) | Cell per Group[b] | Library Cost ($) | Minimal Cost[c] ($) |
|---|---|---|---|---|---|
| CEL-seq2/C1 | 0.8 | ~6.1 | 86/100/110 | ~9 | ~2,420/2,310/2,250 |
| Drop-seq | 0.8 | ~8.4 | 99/135/254 | ~0.1 | ~1,010/700/690 |
| MARS-seq | 0.8 | ~7.3 | 110/135/160 | ~1.3 | ~1,380/1,030/820 |
| SCRB-seq | 0.8 | ~6.1 | 64/90/166 | ~2 | ~900/810/1,080 |
| Smart-seq/C1 | 0.8 | ~4.9 | 150/172/215 | ~25 | ~9,010/9,440/11,290 |
| Smart-seq2 (commercial) | 0.8 | ~5.2 | 95/105/128 | ~30 | ~10,470/11,040/13,160 |
| Smart-seq2 (in-house Tn5) | 0.8 | ~5.2 | 95/105/128 | ~3 | ~1,520/1,160/1,090 |

See also Figure 6.

[a]True positive rate and false discovery rate are based on simulations (Figure 6; Figure S9).

[b]Sequencing depth of one, 0.5, and 0.25 million reads.

[c]Assuming $5 per one million reads.

Ziegenhain *et al.*, 2017 Molecular Cell

**Table 1. Widely Used Single-Cell Sequencing Methods**

| *Sequencing Method* | Starting Cell No. | Cell Separation | Notes | Cell Capture | Transcript Capture | Representative Library Prep Cost per Cell[a] |
|---|---|---|---|---|---|---|
| *Fluidigm C1*[b] | ~1,000 cells | cells capture in size-specific chambers | must know the size of cells of interest; allows for staining and imaging prior to cell rupture | 96- or 800-chamber units are available | an average of 6,606 genes/cell (no data on percentage) | $1.70 |
| *DropSeq* | ~150,000 cells/run | droplet-based separation | remains the most cost-effective and most customizable | ~5% of cells per run (approximately 7,000 cells) | ~10.7% of the cell's transcripts | $0.06 |
| *Chromium 10X* | ~1,700 cells/run | droplet-based separation | the most commercially successful method; almost fully automated | ~65% of cells per run (approximately 1,000 cells) | ~14% of the cell's transcripts | $0.10 |
| *SCI-Seq* | ~500,000 cells (depends on experimental design) | FACS sorter; cells are never singly isolated | combinatorial indexing of individual methanol-fixed permeable cells | 5%–10% of cells | ~10%–15% of the cell's transcripts | $0.05–$0.14[c] |

All of the methods require the establishment of a cell dissociation technique. The price is highly dependent on the number of cells sequenced, the desired depth of sequencing, and the sequencing platform used. For this table, the prices are at the lower end of the price range for single-cell library prep.

[a]As of July 2018.

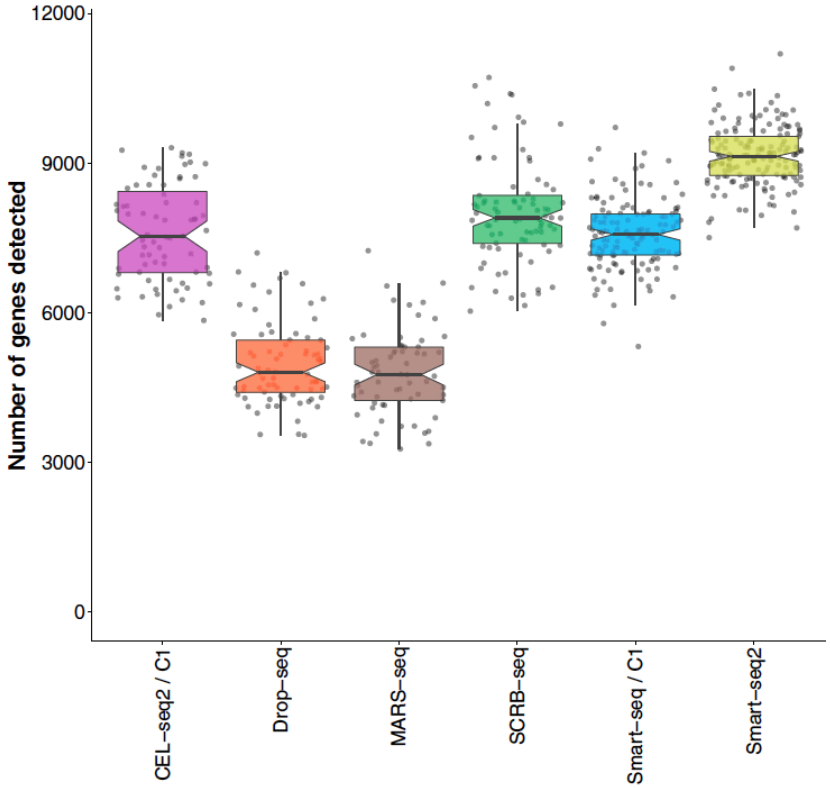
[b]Based on the 800-chamber medium-size isolation unit.

[c]Dependent on how many cells are prepped for sequencing and how many doublets are tolerated.

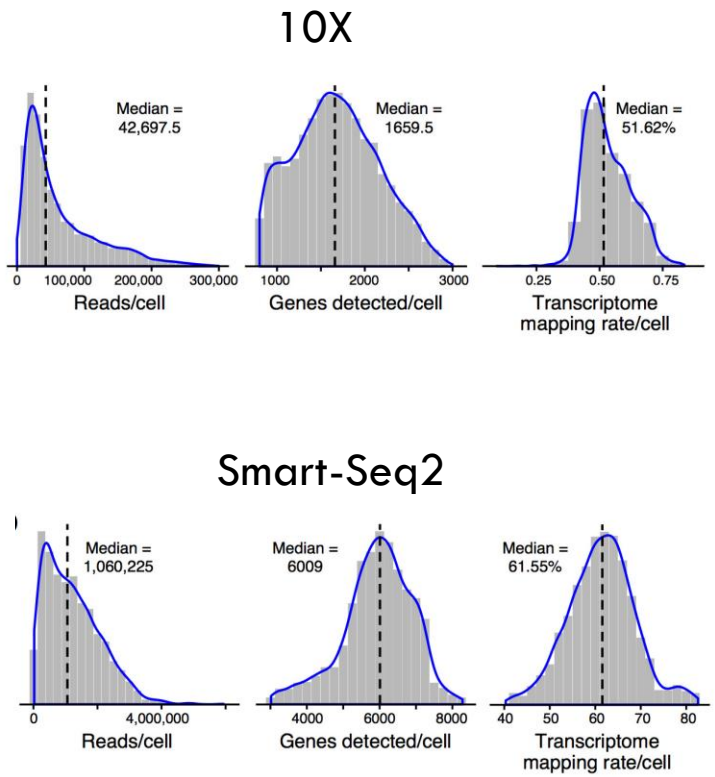
An Introduction to the Analysis of Single-Cell RNA-Sequencing Data

AlJanahi, Aisha A. et al.
Molecular Therapy - Methods & Clinical Development , Volume 10 , 189 - 196

# 3’-end vs full-length Inter-platform comparison

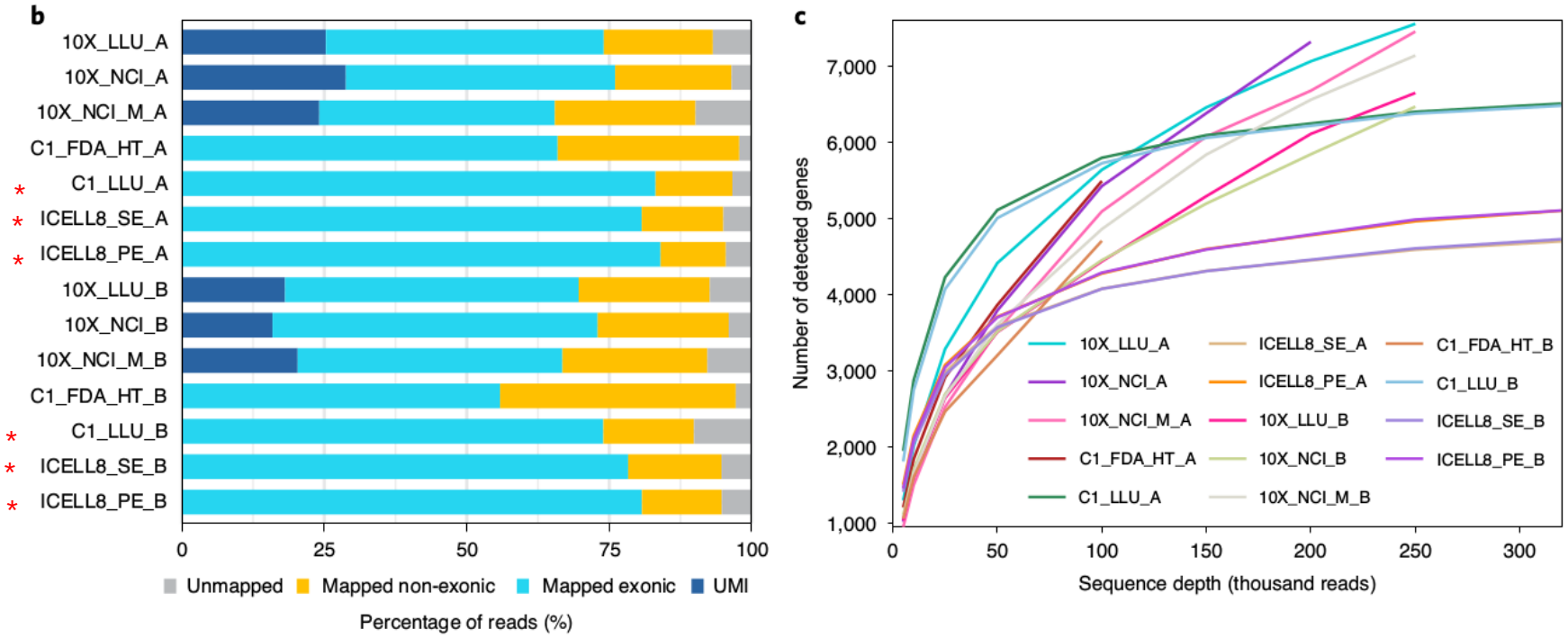



Ziegenhain *et al.*, 2017 Molecular Cell




Haber *et al.*, 2017 Nature

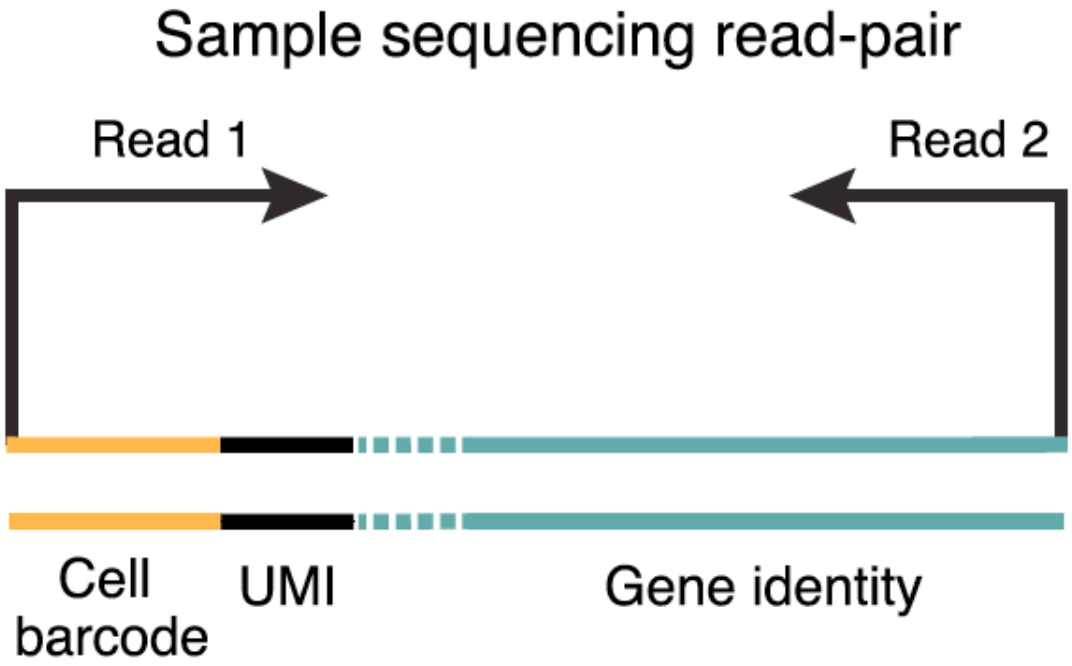
# 3’-end vs full-length Inter-platform comparison



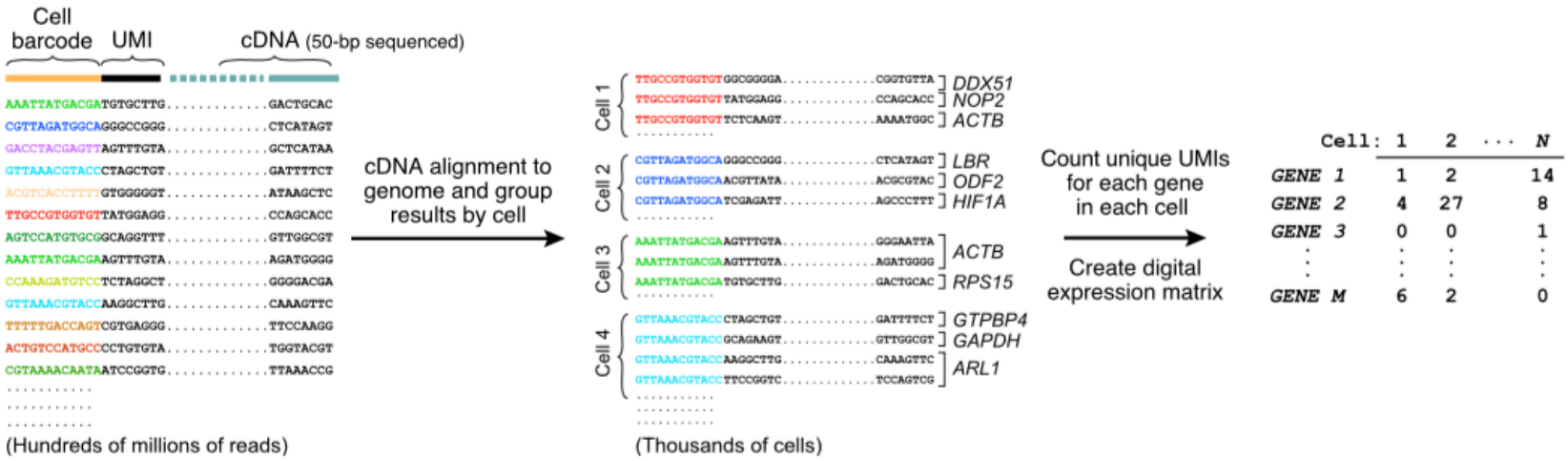

"*" = Full-length

*Chen et al.*, 2020 Nature Biotech

# 3’-end based technologies Post-sequencing

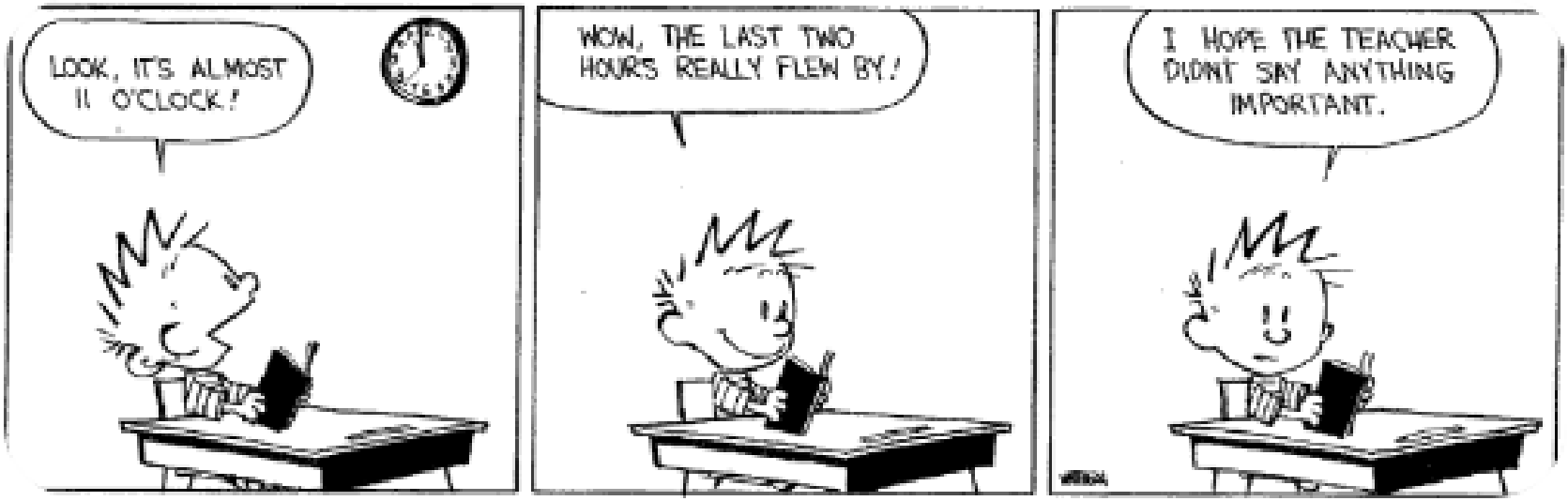


# 3’-end based technologies - Post-sequencing




Macosko *et al*., 2015 Cell

# Break Time


LOOK, IT'S ALMOST 11 O'CLOCK!
WOW, THE LAST TWO HOURS REALLY FLEW BY!
I HOPE THE TEACHER DIDN'T SAY ANYTHING IMPORTANT.