INTRODUCTION TO SINGLE-CELL RNA-SEQ

Gabriel Alencar
gf8kz@virginia.edu

BIOC8145
2020/04/27 to 2020/05/01

OUTLINE

2020/04/27: Lecture 1

- scRNA-seq introduction
- Available Technologies and understanding limitations
- STUDY DESIGN!!!
- 10x FASTQ alignment

2020/04/29: Lecture 2

- Quality control / gene and cell filtering
- Dimensional Reduction and plotting genes and cells

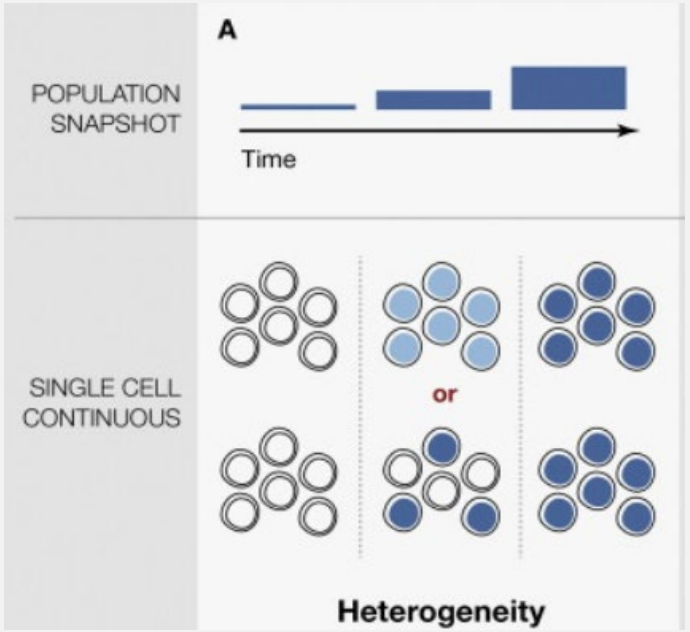
2020/05/01: Lecture 3

- Differential gene expression
- Downstream analysis

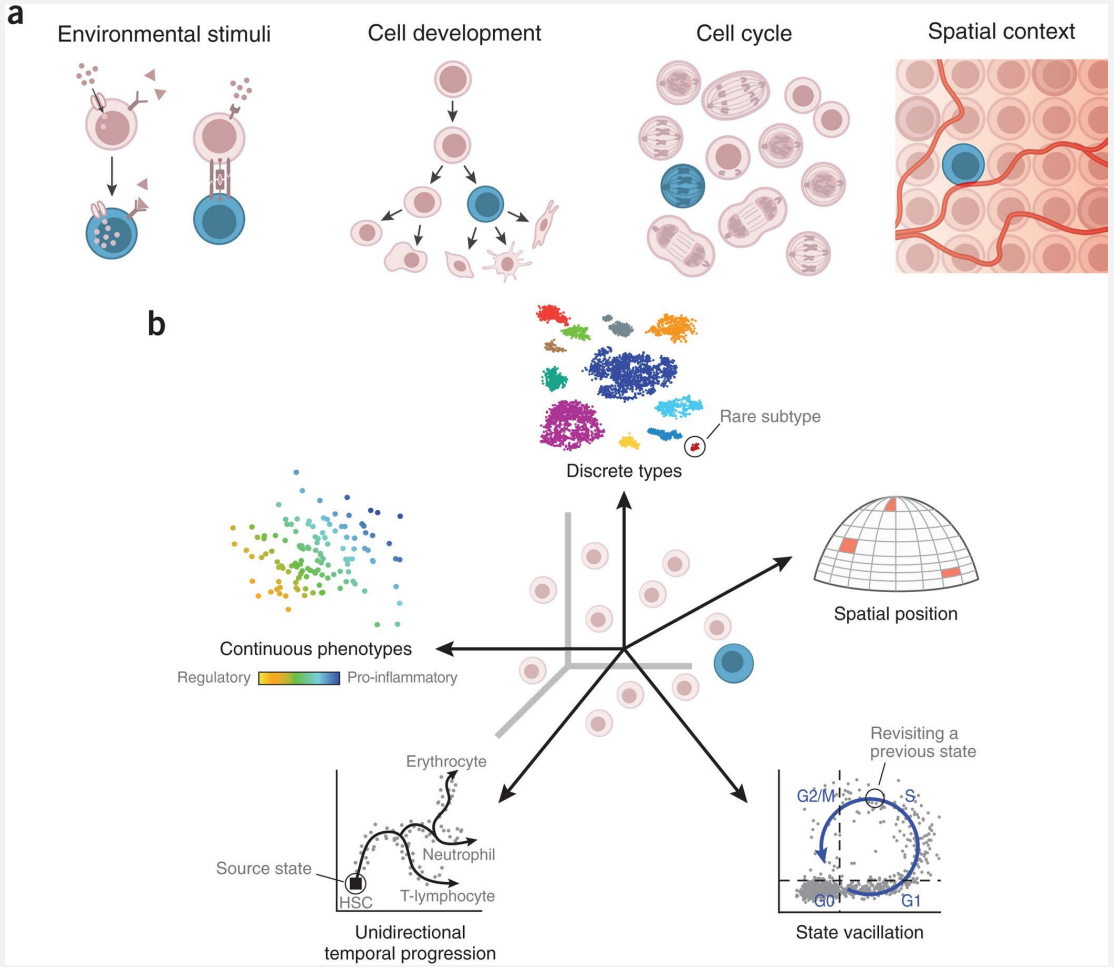
Why do single cell RNA-seq?

scRNA-seq will help you understand what are the “identities” of cells in your specific model.

- Identify expression profiles of individual cells (that may be missed with bulk RNA-seq)



Etzrodt M., Endele M & Schroeder T. **Quantitative Single-Cell Approaches to Stem Cell Research**. *Cell Stem Cell* 2014

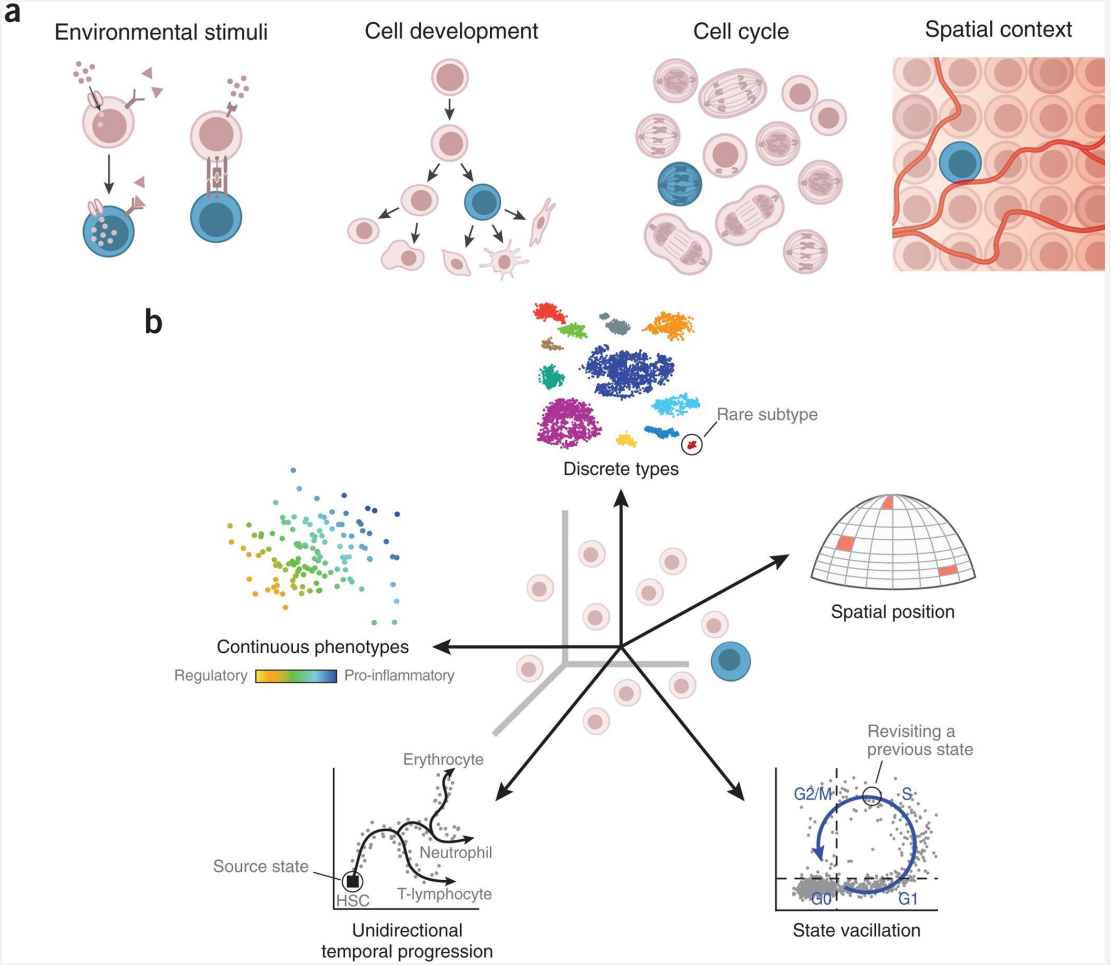


Wagner A., Regev A & Yosef N. **Revealing the vectors of cellular identity with single-cell genomics**. *Nat biotechnology* 2016

Why do single cell RNA-seq?

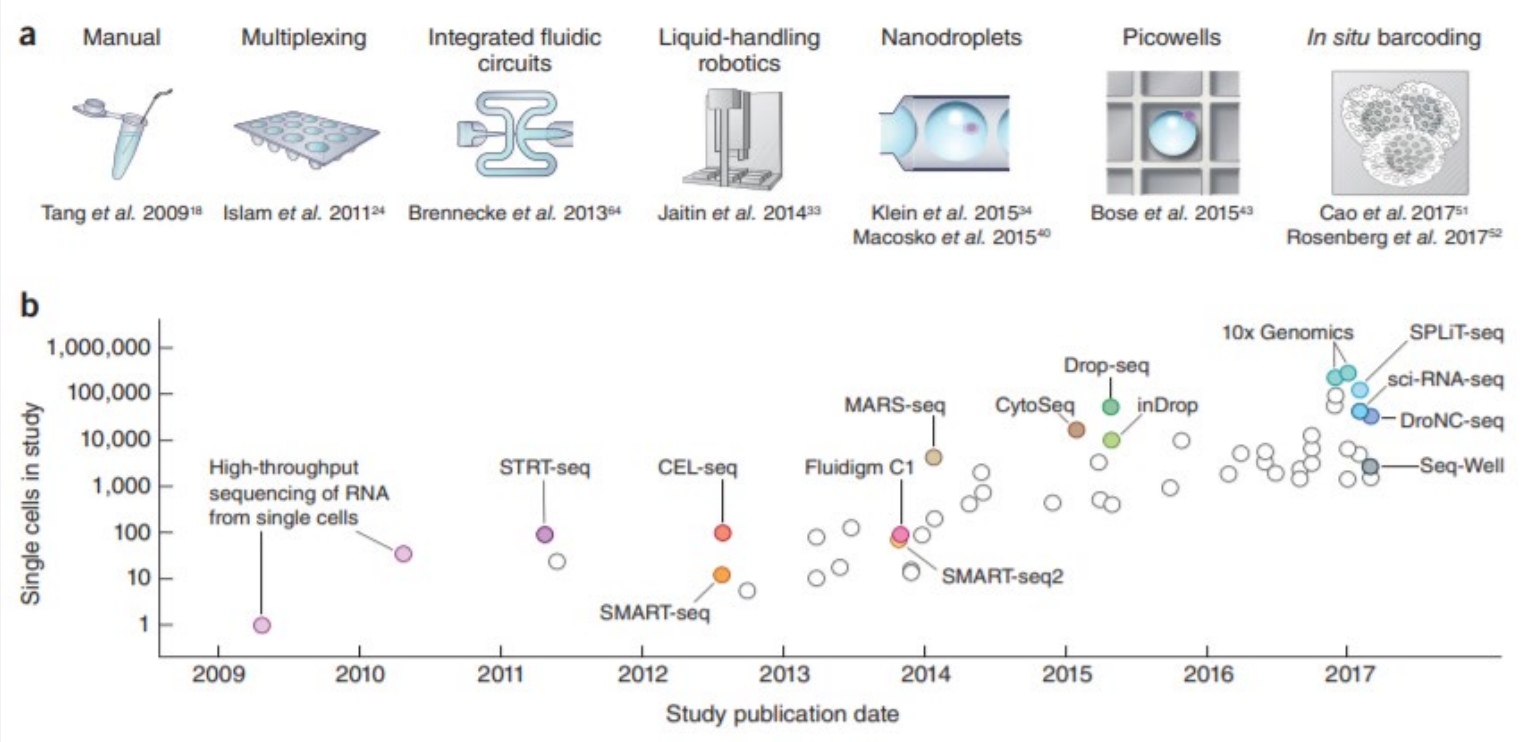
scRNA-seq will help you understand what are the “identities” of cells in your specific model.

- Identify expression profiles of individual cells (that may be missed with bulk RNA-seq)
- Discover new rare populations and/or cell states
- Response to the environment and development trajectory

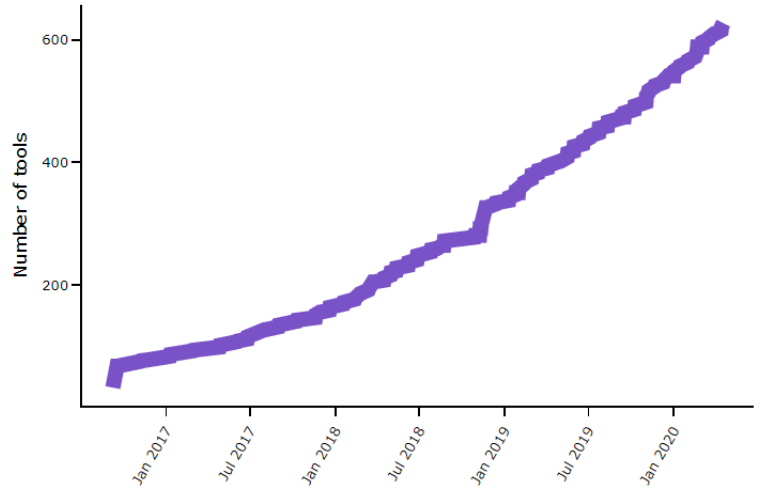


Wagner A., Regev A & Yosef N. **Revealing the vectors of cellular identity with single-cell genomics**. *Nat biotechnology* 2016

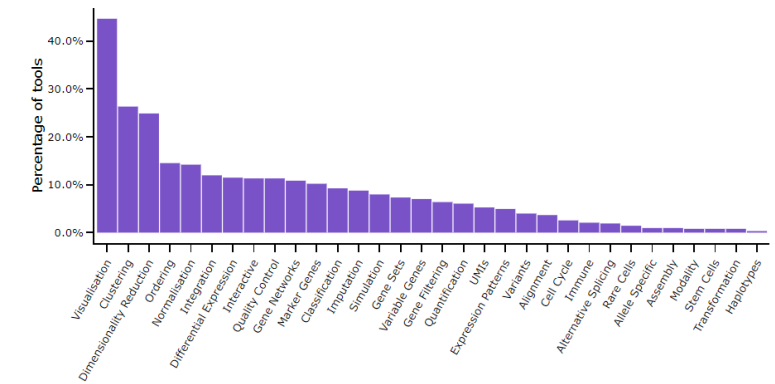
Advances in scRNA-seq assays led to an exponential growth in publications and methods in the last 5 to 10 years



Svensson V., Vento-Tormo R & Teichmann S. **Exponential scaling of single-cel RNA-seq in the past decade**. *Nat protocols* 2018



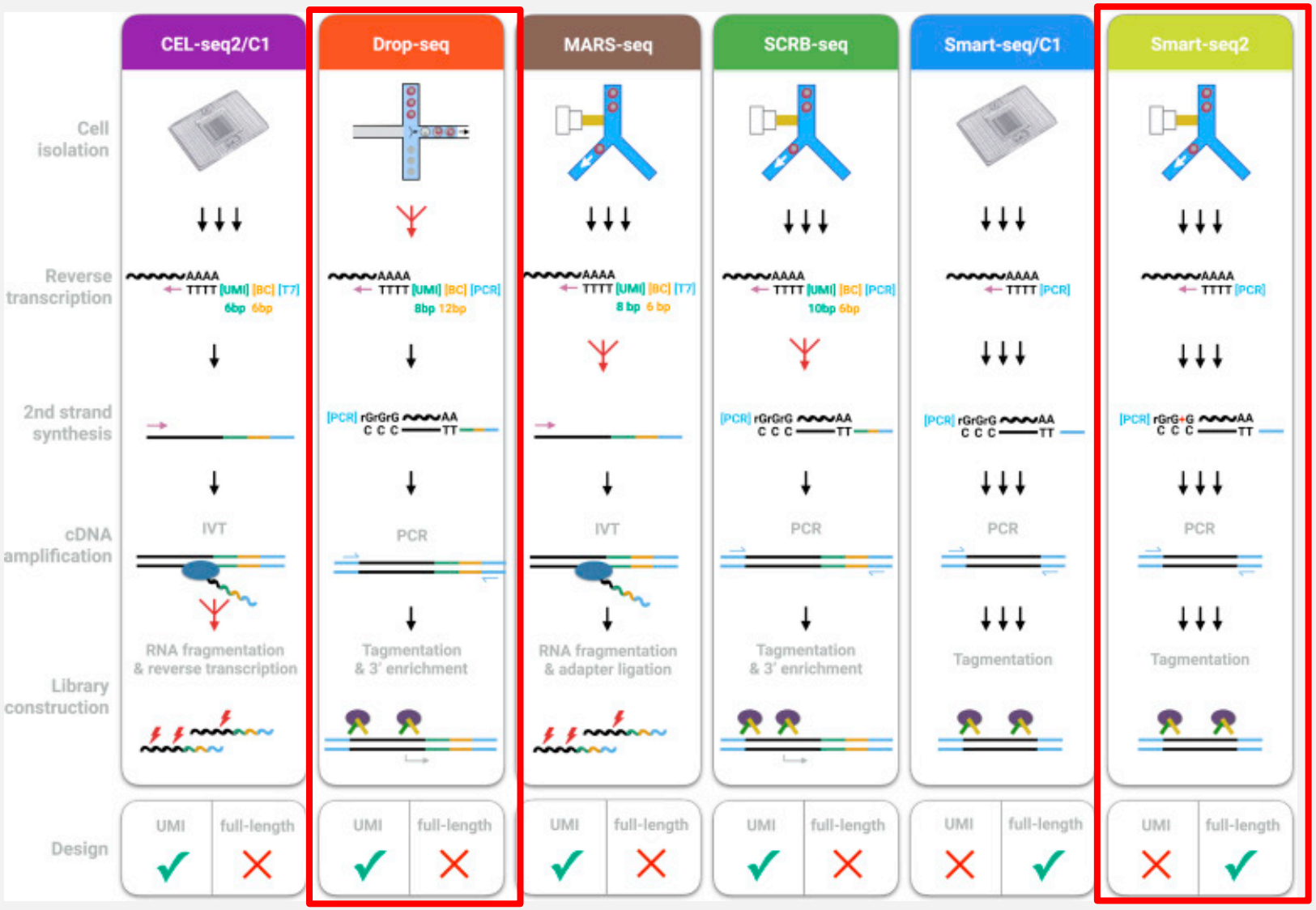
We currently track 627 tools...



...in over 30 categories

www.scrna-seq.org

scRNA-seq assays differs in several areas, including library prep, UMI or not, transcript information, COST



Ziegenhain C. **Comparative Analysis of Single-Cell RNA Sequencing Methods**. *Molecular Cell* 2017

Smart-seq2

nature methods

Brief Communication | Published: 22 September 2013

Smart-seq2 for sensitive full-length transcriptome profiling in single cells

Simone Picelli, Åsa K Björklund, Omid R Faridani, Sven Sagasser, Gösta Winberg & Rickard Sandberg

Nature Methods **10**, 1096–1098(2013) | Cite this article

11k Accesses | **666** Citations | **113** Altmetric | Metrics

nature protocols

Protocol | Published: 02 January 2014

Full-length RNA-seq from single cells using Smart-seq2

Simone Picelli, Omid R Faridani, Åsa K Björklund, Gösta Winberg, Sven Sagasser & Rickard Sandberg

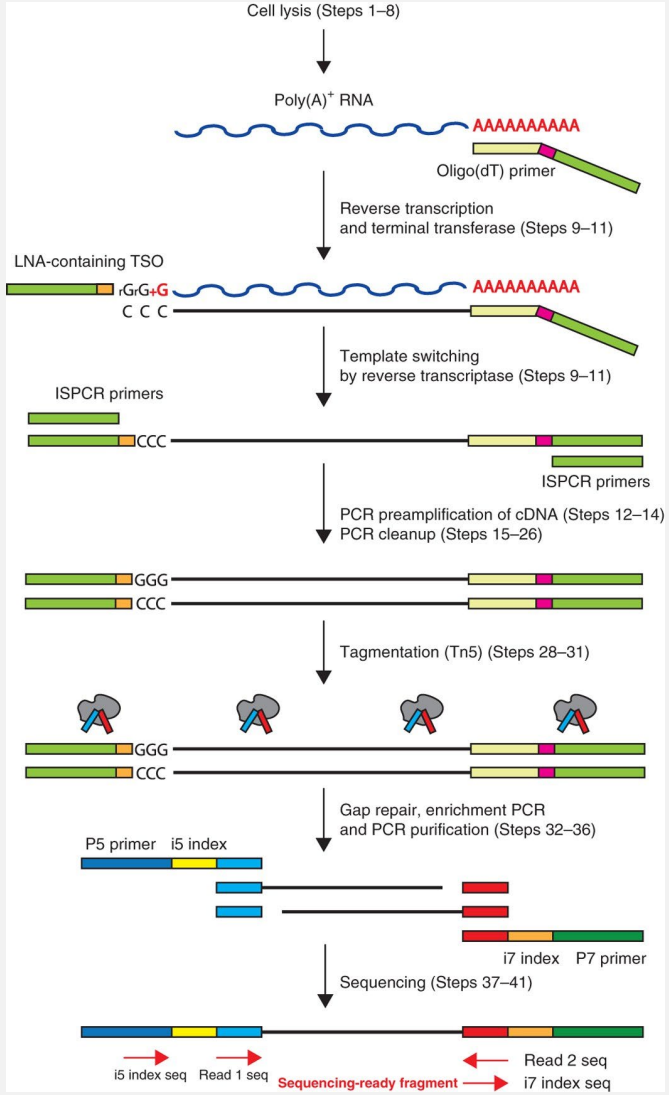
Nature Protocols **9**, 171–181(2014) | Cite this article

28k Accesses | **857** Citations | **53** Altmetric | Metrics

Also developed NASC-seq in 2019

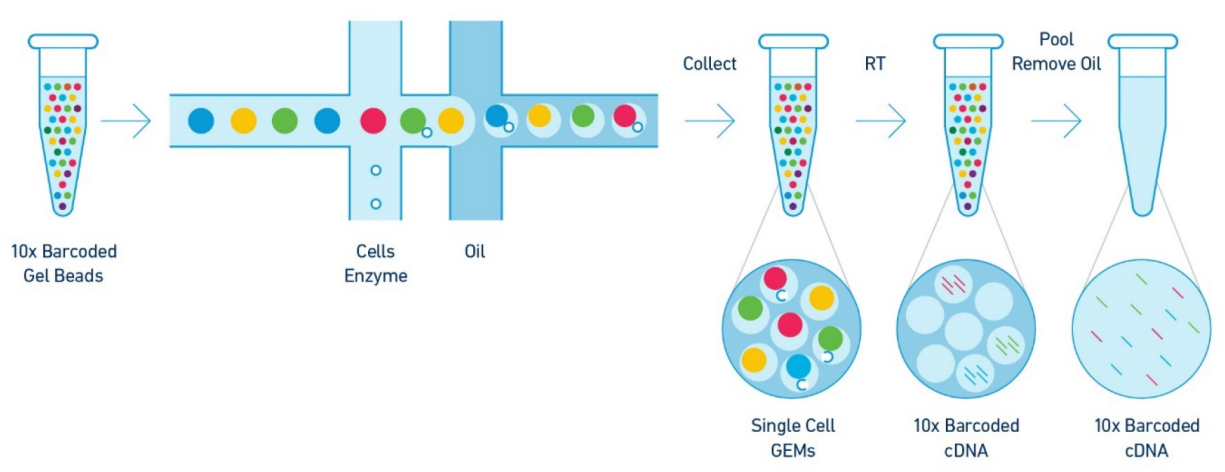
Smart-seq2: Description and technology overview

- Developed for single cell, but has been adapted and used for total RNA (excellent for low input RNA)
- Selects for poly-A tail
- **Full transcript scRNA-seq**
- Off-the-shelf products
- Hundreds of samples
- Most often used without UMIs, but with spike-ins



Picelli S. et al. **Full-length RNA-seq from single cells using Smart-seq2**. *Nat protocols* 2014

Drop-seq: Description and technology overview (also, if you are using UVA cores, this is what they offer)



From www.10xgenomics.com

- Throughput from hundreds to thousands.
- Droplet-based processing using microfluidics
- Nanoliter scale aqueous drops in oil.
- 3' End
- Uses UMI (Unique Molecular Identifier).

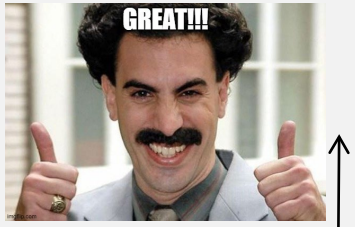
More specifically to 10x:

- GEM (Gel Bead in Emulsion)
- Standardized instrumentation and reagents.
- Scalable to tens of thousands.
- Less processing time.
- Software is available and very supported by company

Drop-seq: Description and technology overview (also, if you are using UVA cores, this is what they offer)

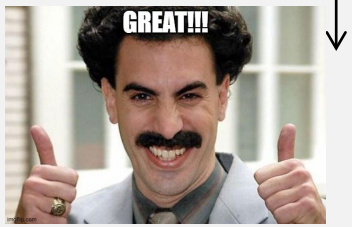
From www.10xgenomics.com

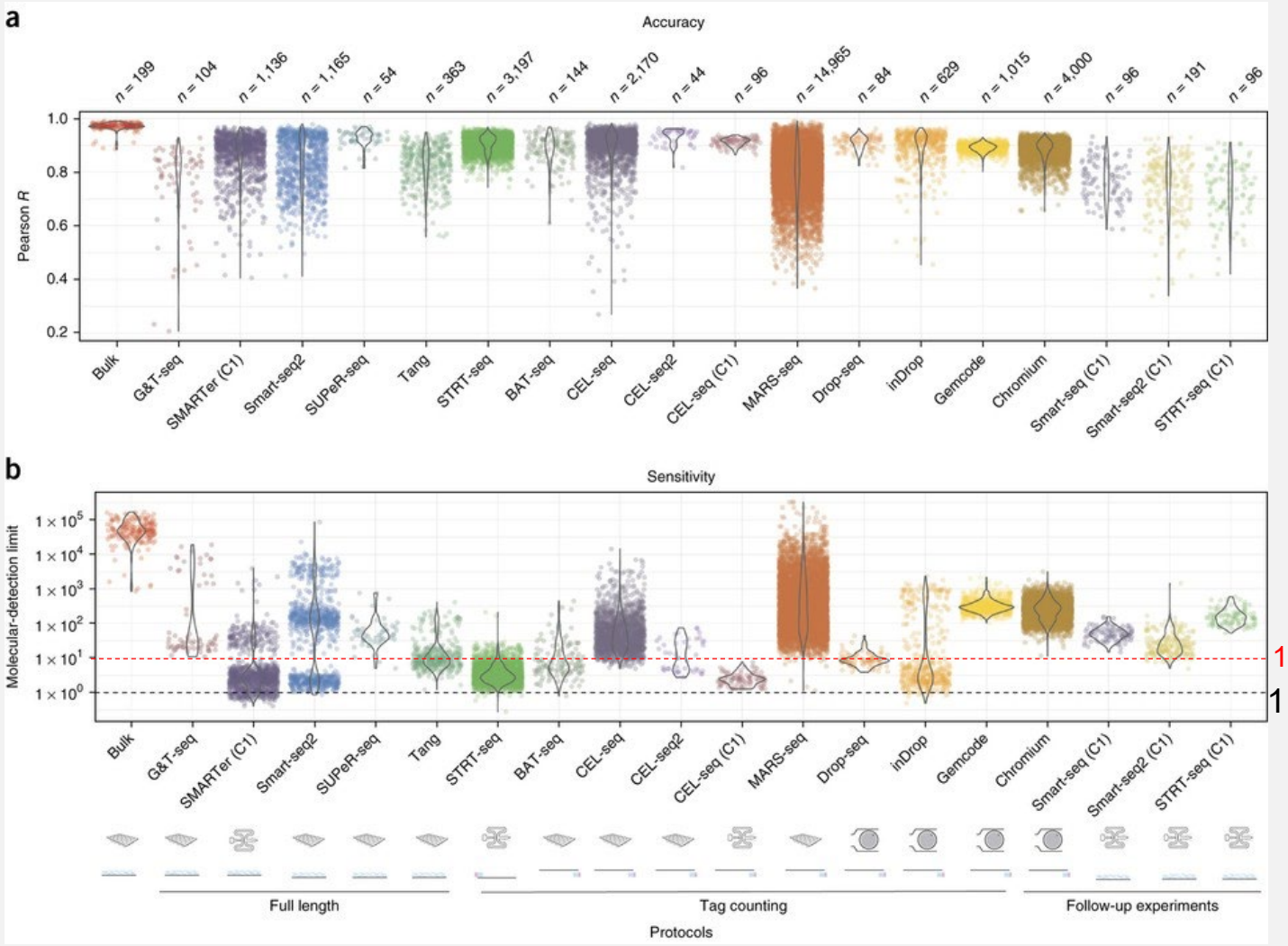
Comparison between different assays



Accuracy

Sensitivity





Svensson V. et al. **Power analysis of single-cell RNA-sequencing experiments**. *Nat methods* 2017

Final thoughts on different technologies/assays

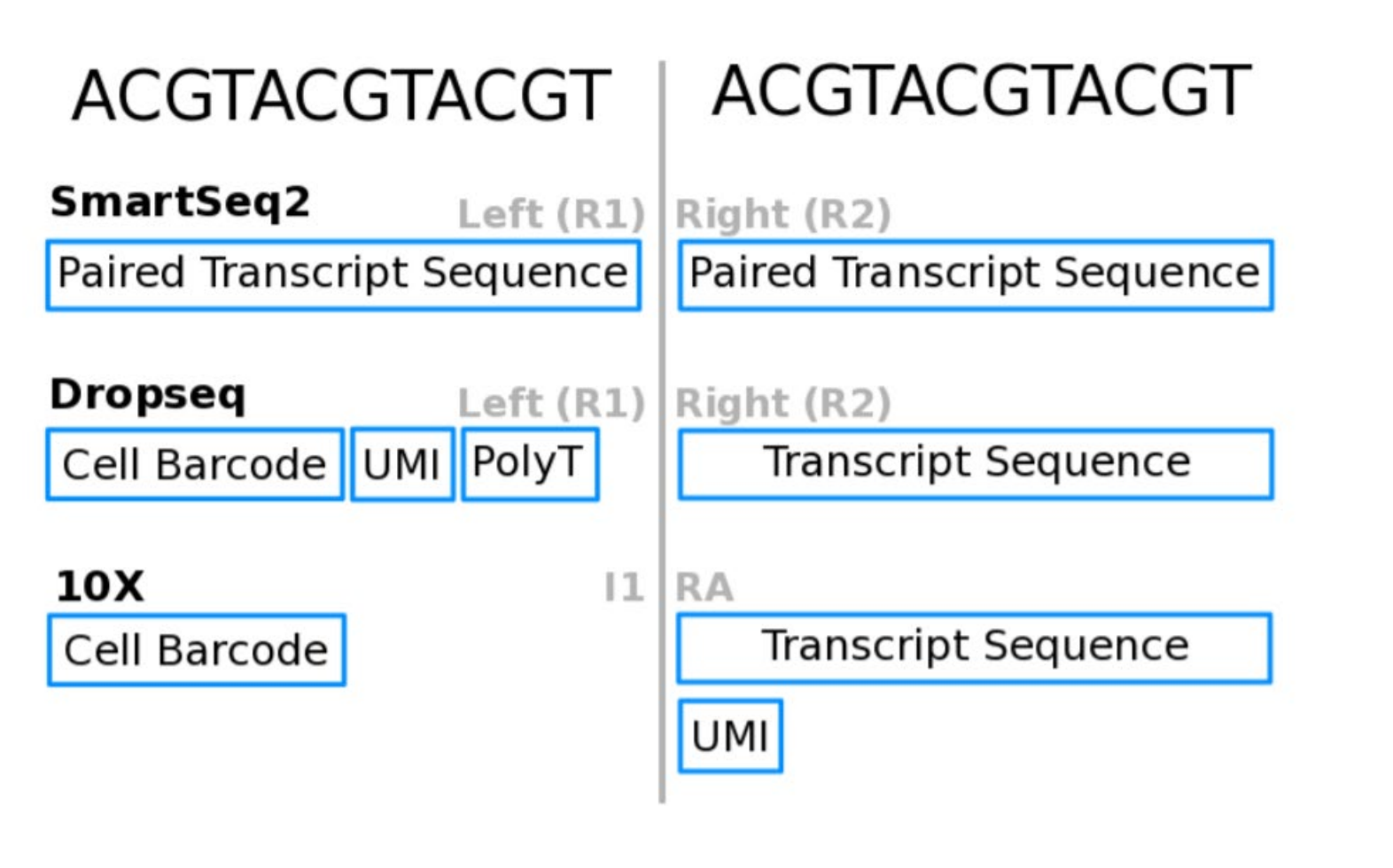
- Different assays have different throughput.
  - Smart-seq2 < Drop-seq < 10X
- Transcript length
  - 10x (only partial) < Smart-seq2 (full)
- Plate-based methods get lysed in wells and so do not leak.
  - Droplet-based can have leaky RNA.
- In Drop-seq assays RT happens outside the droplets

Understanding (or at least knowing the basics) of the assay/technology you selected is crucial for pipeline and downstream analysis

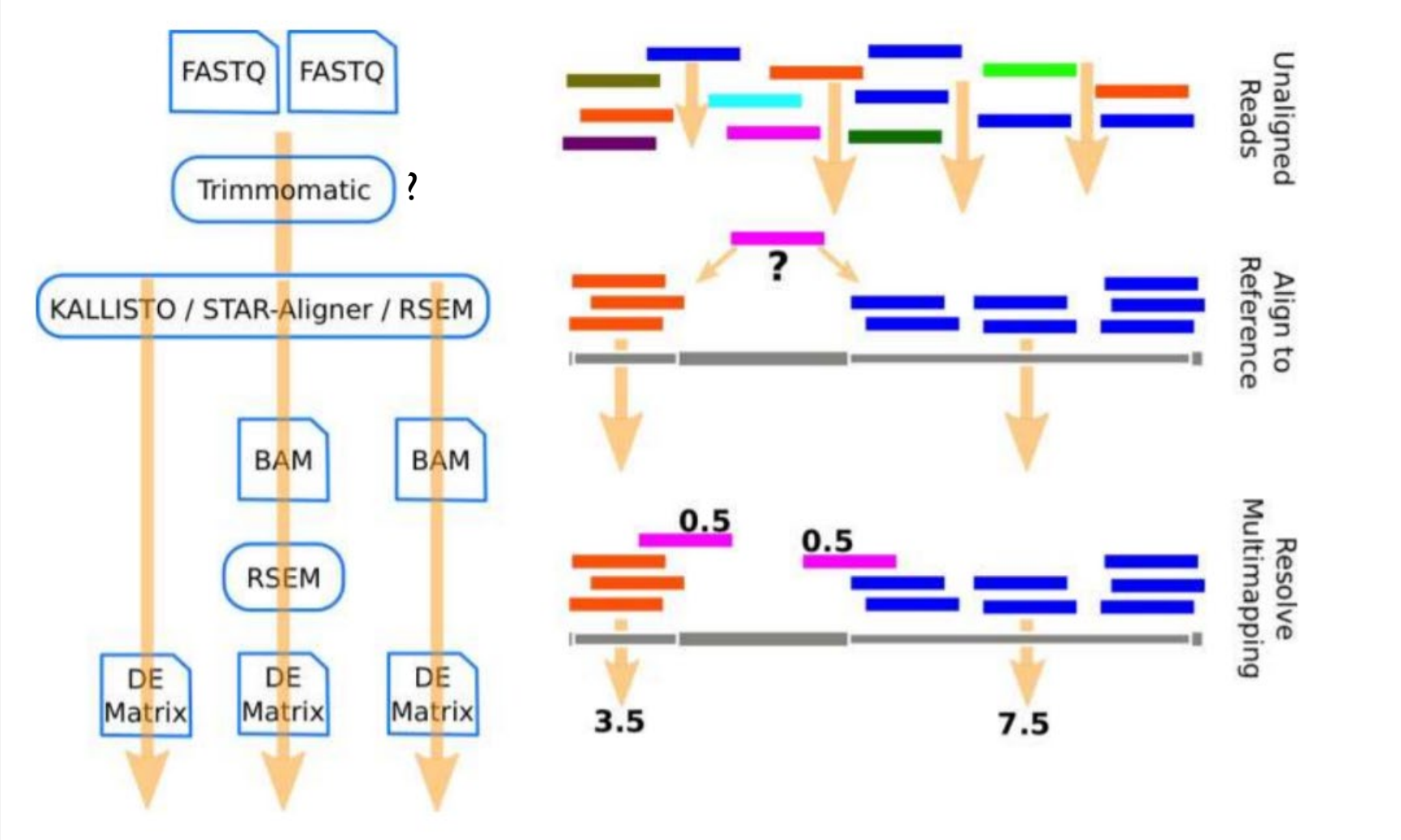
Final thoughts on different technologies/assays

- Different assays have different throughput.
  - Smart-seq2 < Drop-seq < 10X
- Transcript length
  - 10x (only partial) < Smart-seq2 (full)
- Plate-based methods get lysed in wells and so do not leak.
  - Droplet-based can have leaky RNA.
- In Drop-seq assays RT happens outside the droplets
  - Can use harsher lysis buffers.
- 10X is more standardized and comes with a pipeline.
- Drop-seq is more customizable.
- **Cost per library varies greatly**.

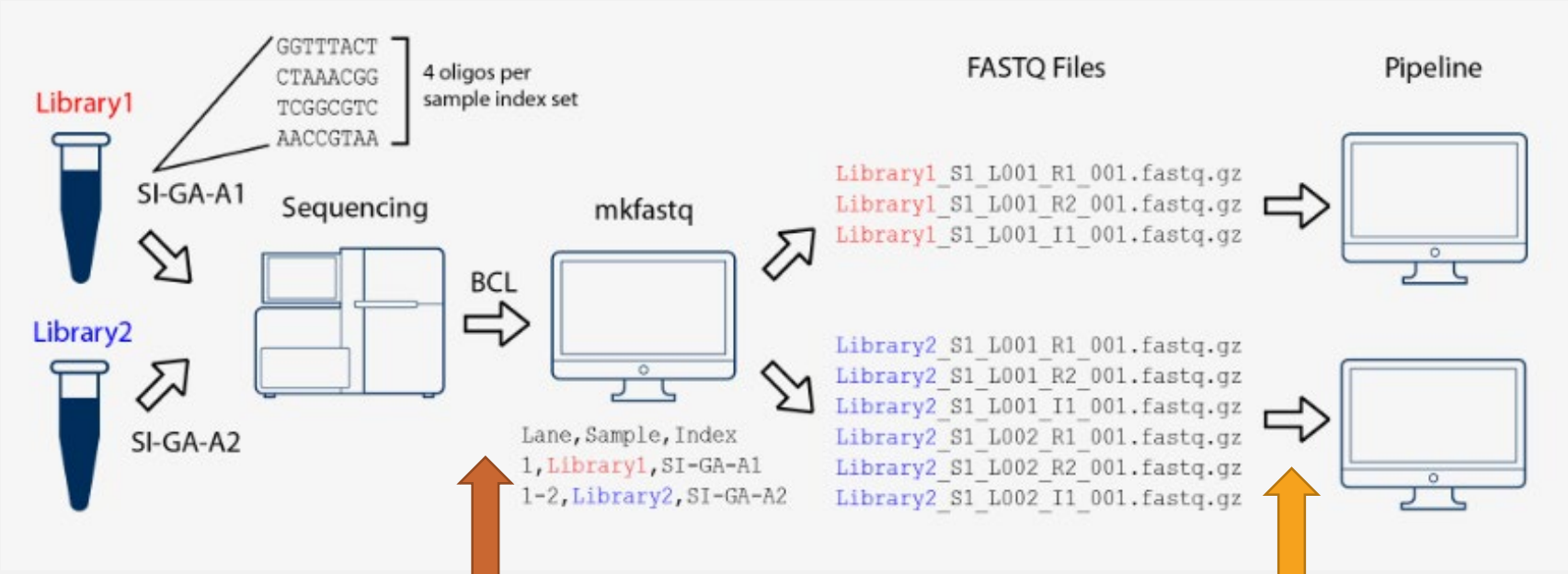
Different technologies have different sequencing FASTQ files



Smart-seq2 Pipeline



10X Genomics Pipeline

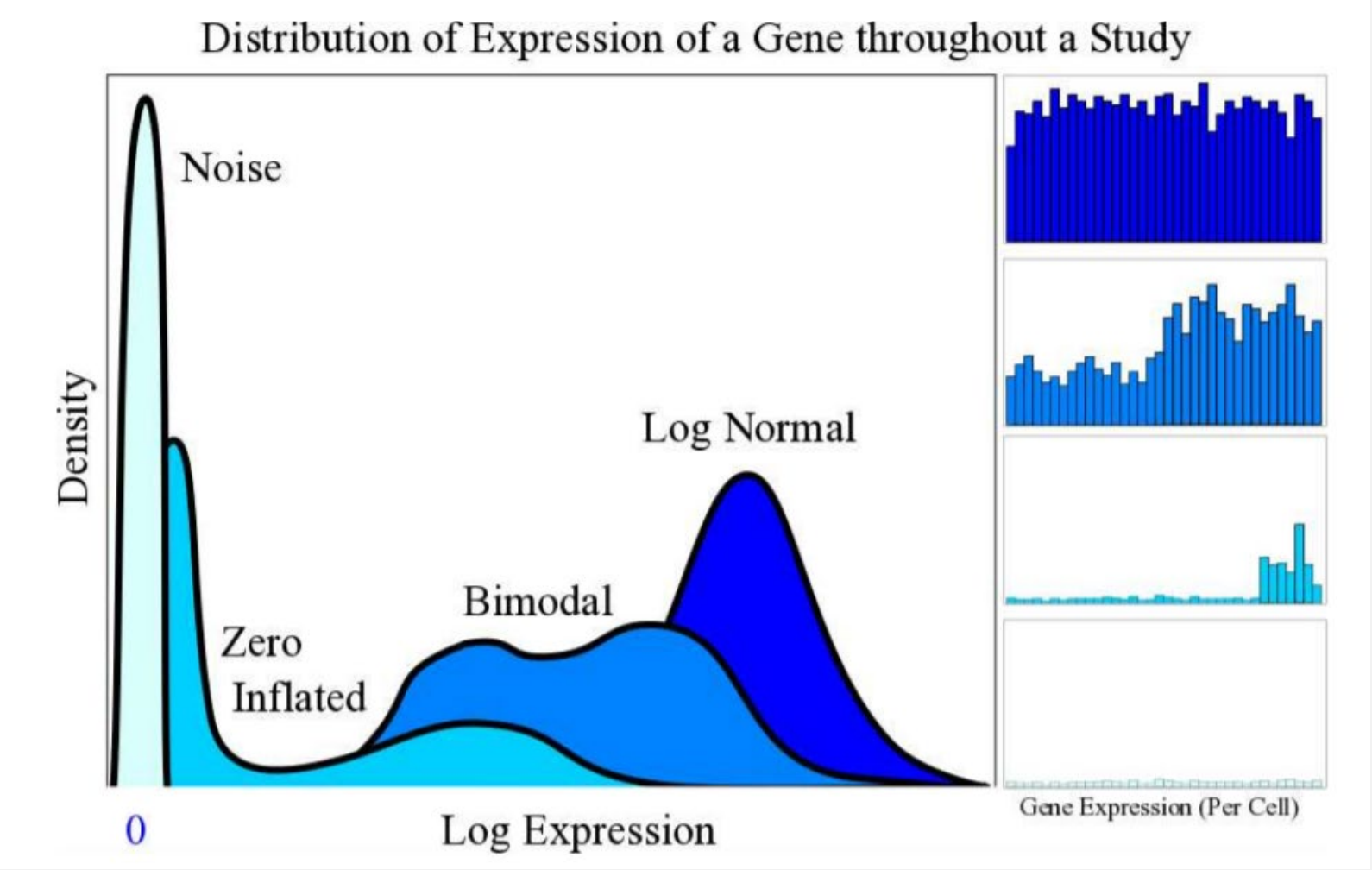


scRNA-seq Study Design

- How many cells should you sequence?
  - HIGHLY dependent on: technology available and/or what type of transcript you are interested in looking at
  - What is your biology and what is the expectation of finding rare populations?
- scRNA-seq is VERY prone to technical batch affects
- Use UMIs or spike-ins

- Satija lab online tool:
  - How many cells?

Genes can have different distributions in a scRNA-seq experiment



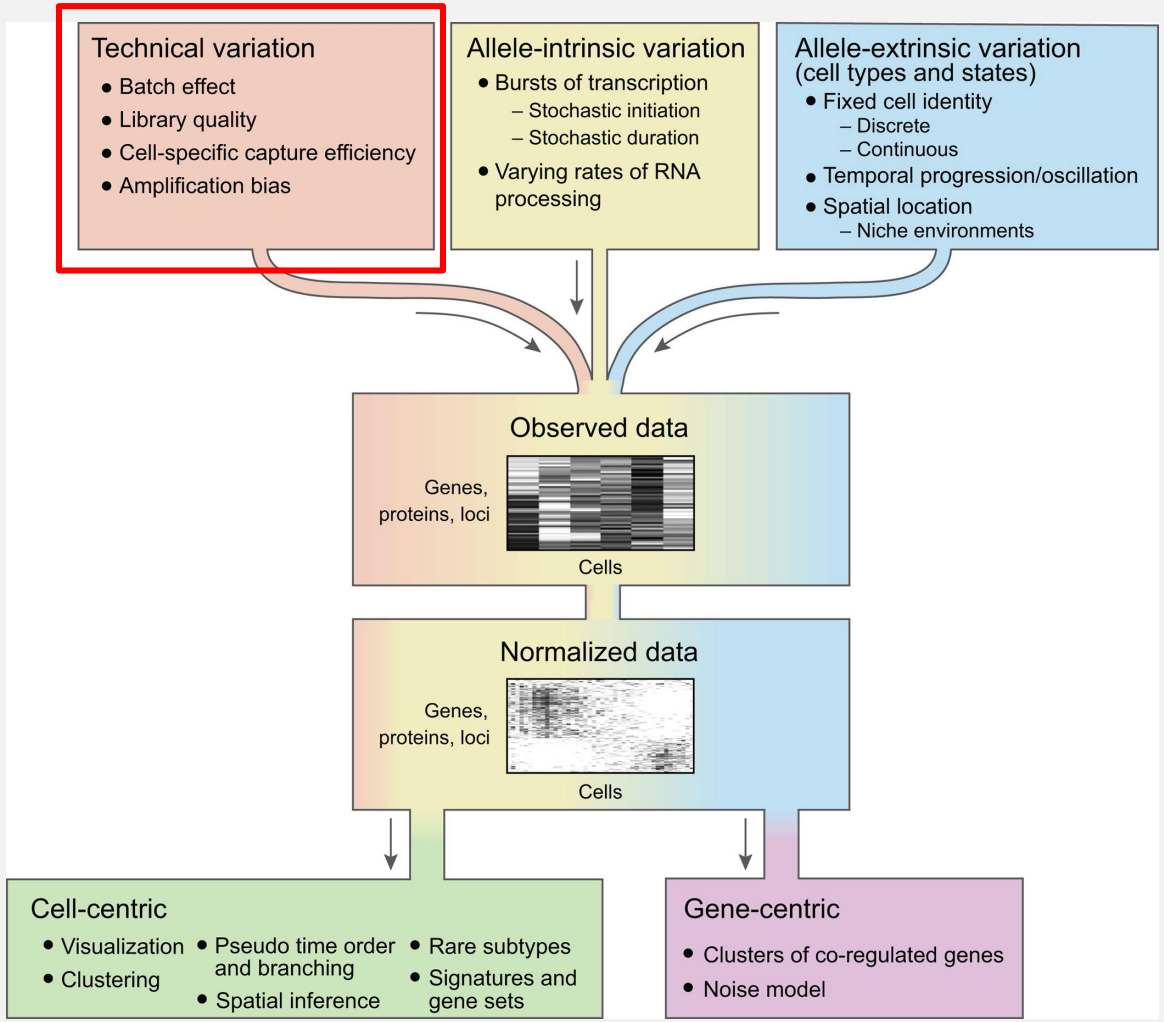
House keeping genes

Bimodal (different populations)

Rare populations

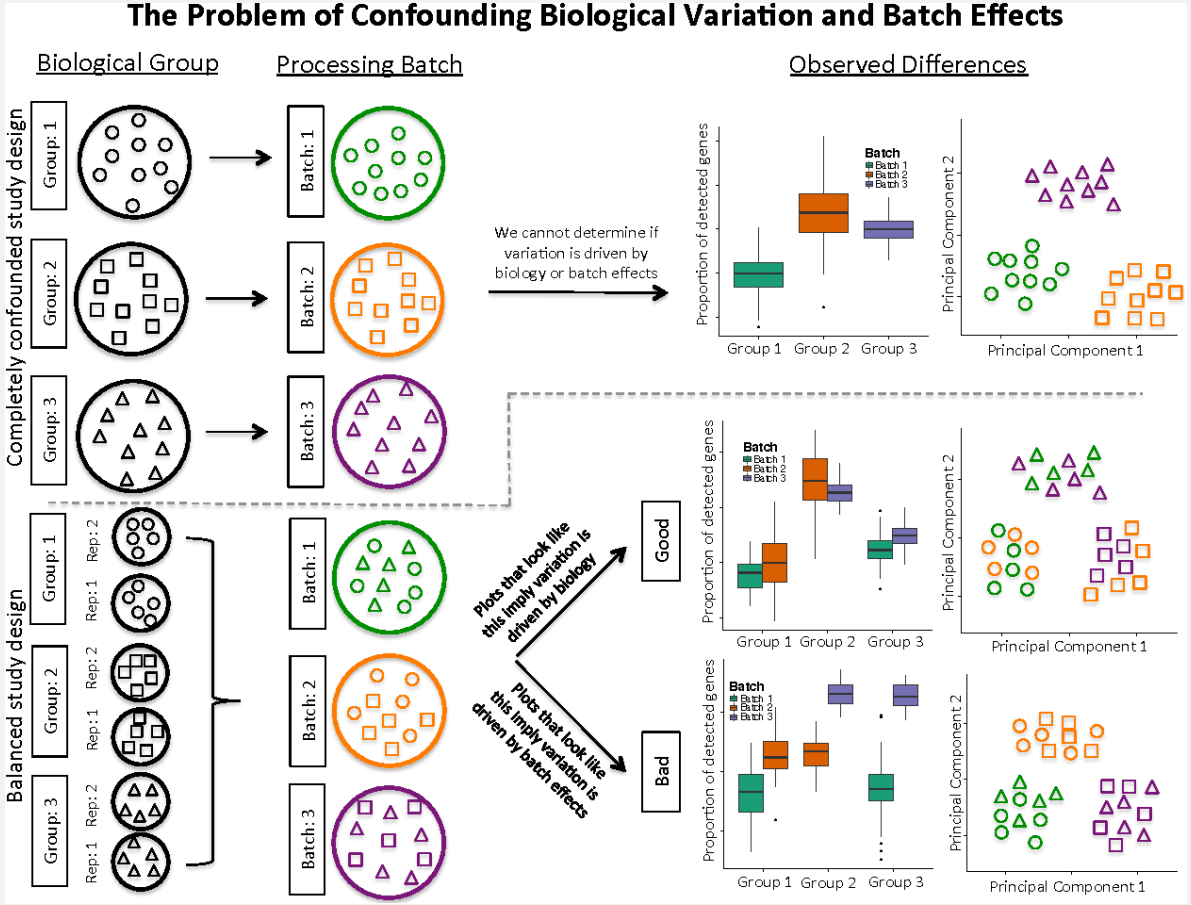
Noise

Expression can have several sources of variation that will affect single-cell

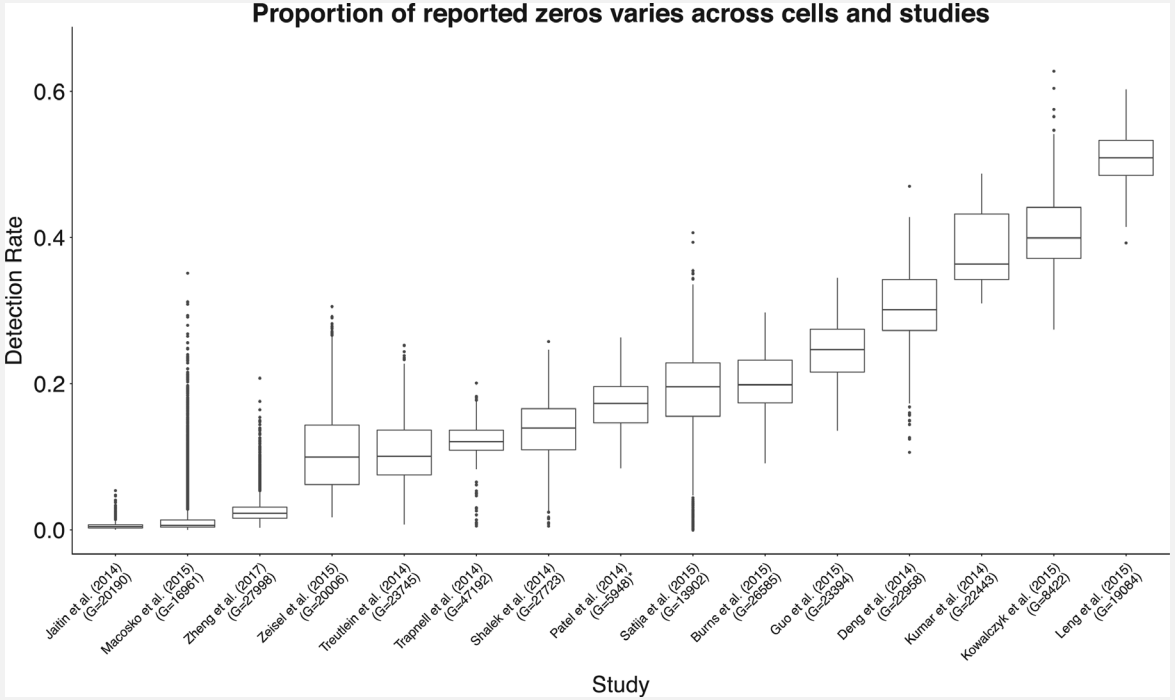


Wagner A., Regev A & Yosef N. **Revealing the vectors of cellular identity with single-cell genomics**. *Nat biotechnology* 2016

Batch Effect and “Missing” data



Hicks S., Townes F., Teng M., Irizarry R. **On the widespread and critical impact of systematic bias and batch effects in single-cell RNA-seq data.** *bioRxiv* 2015



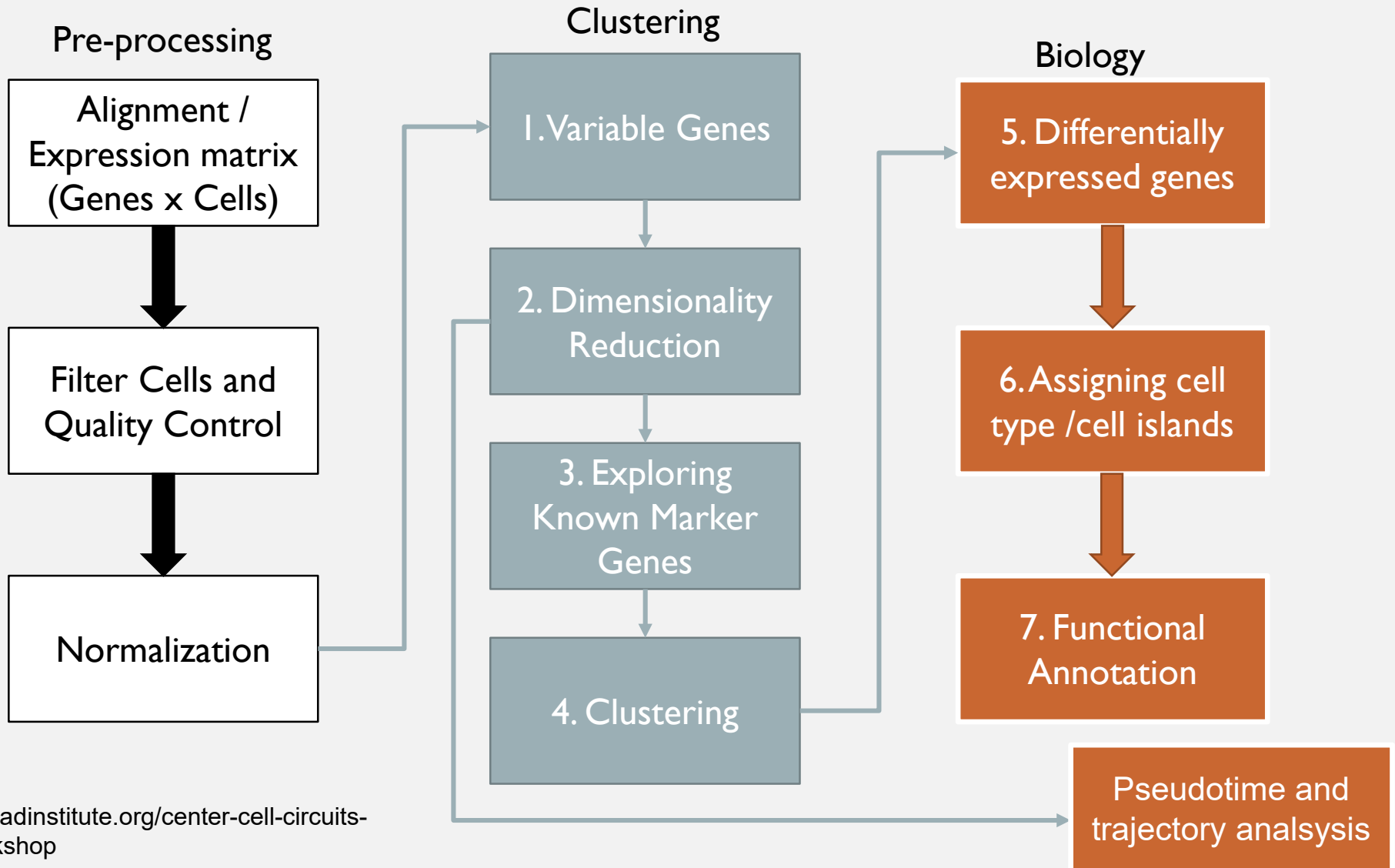
Hicks S., Townes F., Teng M., Irizarry R. **Missing data and technical variability in single-cell RNA-sequencing experiments.** *Biostatistics* 2018

Summary of the data

We are still early in the process in understanding scData and how to apply it:

- Data can be NOT normal
- Data can be Zero-inflated
- Data can be very noisy
- Cells vary in library complexity
- And many other factors

Analysis pipeline



Adapted from https://www.broadinstitute.org/center-cell-circuits-computational-genomics-workshop


I AM
GROOT?