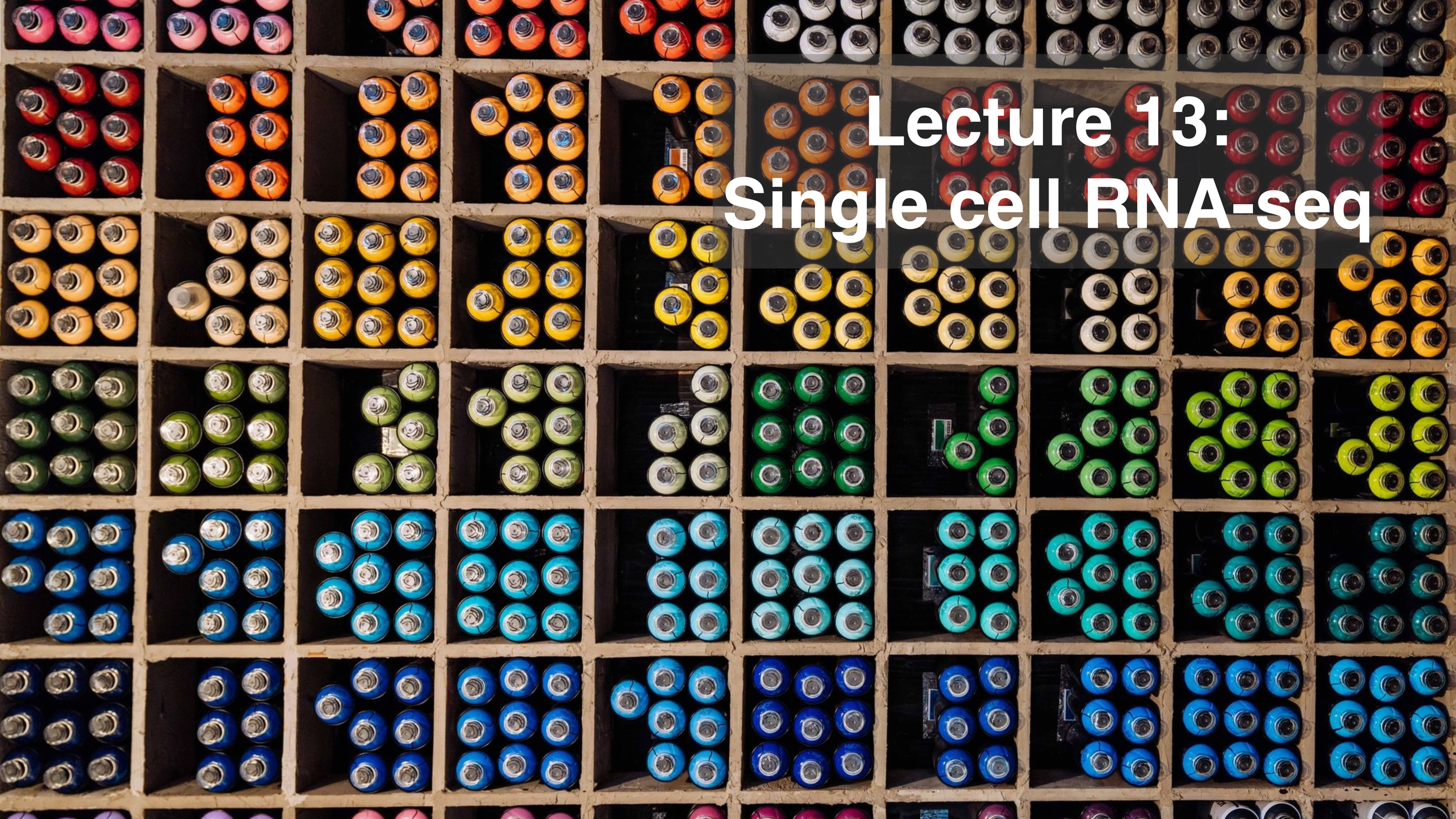


Lecture 13: Single cell RNA-seq

You already have a strong foundation for learning scRNA-seq

Read mapping

Filtering

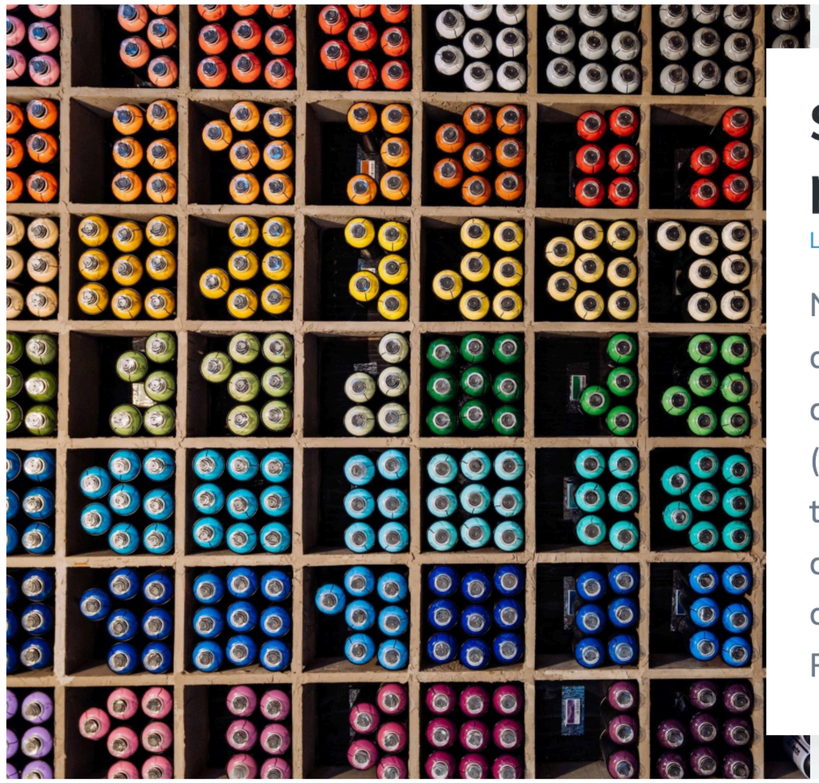
Normalization

Dimensional reduction

Differential gene testing

You already have a strong foundation for learning scRNA-seq

https://github.com/Bioconductor/BiocStickers



Single cell RNA-seq principles and proc

Lecture 13 • watch by December 1, 2021

Now that you're comfortable w
data analysis, we'll shift our fo
developing landscape of singl
(scRNA-seq). In this lecture, yo
the underlying technology and
demonstrate how to process r
data directly on your laptop (!)
R/bioconductor.

Single cell RNA-seq – principles and processing

Lecture 13 • watch by December 1, 2021

Overview

Now that you're comfortable with bulk RNA-seq data analysis, we'll shift our focus to the rapidly developing landscape of single cell RNA-seq (scRNA-seq). In this lecture, you'll learn about the underlying technology and we'll demonstrate how to process raw single cell data directly on your laptop (!) for importing into R/bioconductor.

Learning objectives

- Understand droplet-based scRNA-seq technology
- Be able to compare and contrast single cell and bulk RNA-seq methods
- Understand cost and experimental design considerations for scRNA-seq experiments.
- Familiarity with multiplexed single cell assays (CITE-seq, 'multiome', TEA-seq)
- Be able to define common terms and concepts in single cell genomics
- Use Kallisto-BUStools to preprocess raw scRNA-seq data (via **kb-python**)

What you need to do

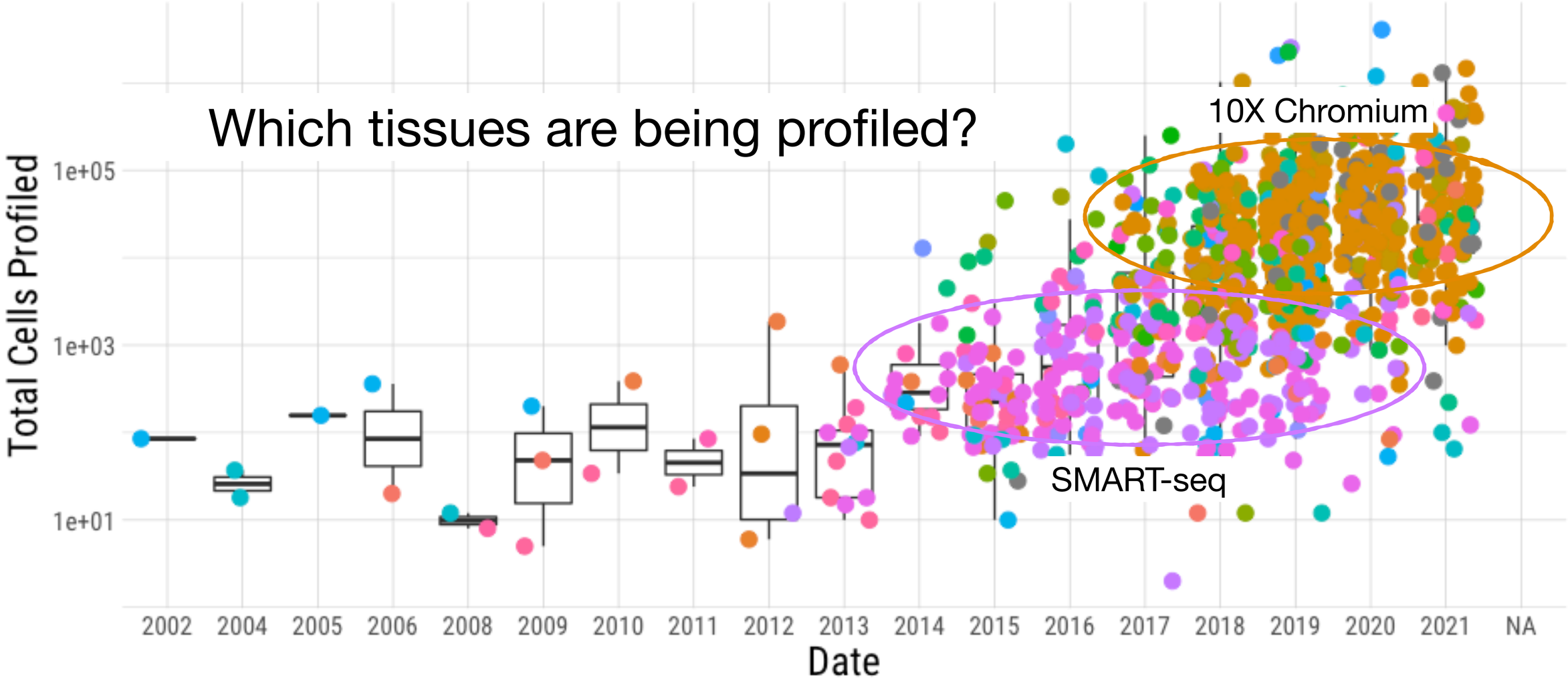
Download raw files. You will need about **5Gb** of storage space on your harddrive to accomodate this download. *please do not uncompress these files (leave them as .gz files).* This is data from 1000 peripheral blood mononuclear cells (PBMCs) and is one of the sample datasets provided by 10X Genomics here. I merged the separate lane files to make this simpler to work with for the course.

Human transcriptome reference index file - this is the index you created using Kallisto way back in lecture 2. If you don't have this, remember it's easy to create using `kallisto index`.

t2g.txt - this is a human transcript-to-gene mapping file that we will use with Kallisto-Bustools to preprocess our data. This file is easy to generate with `kb ref`, but downloading it now will save you some time.

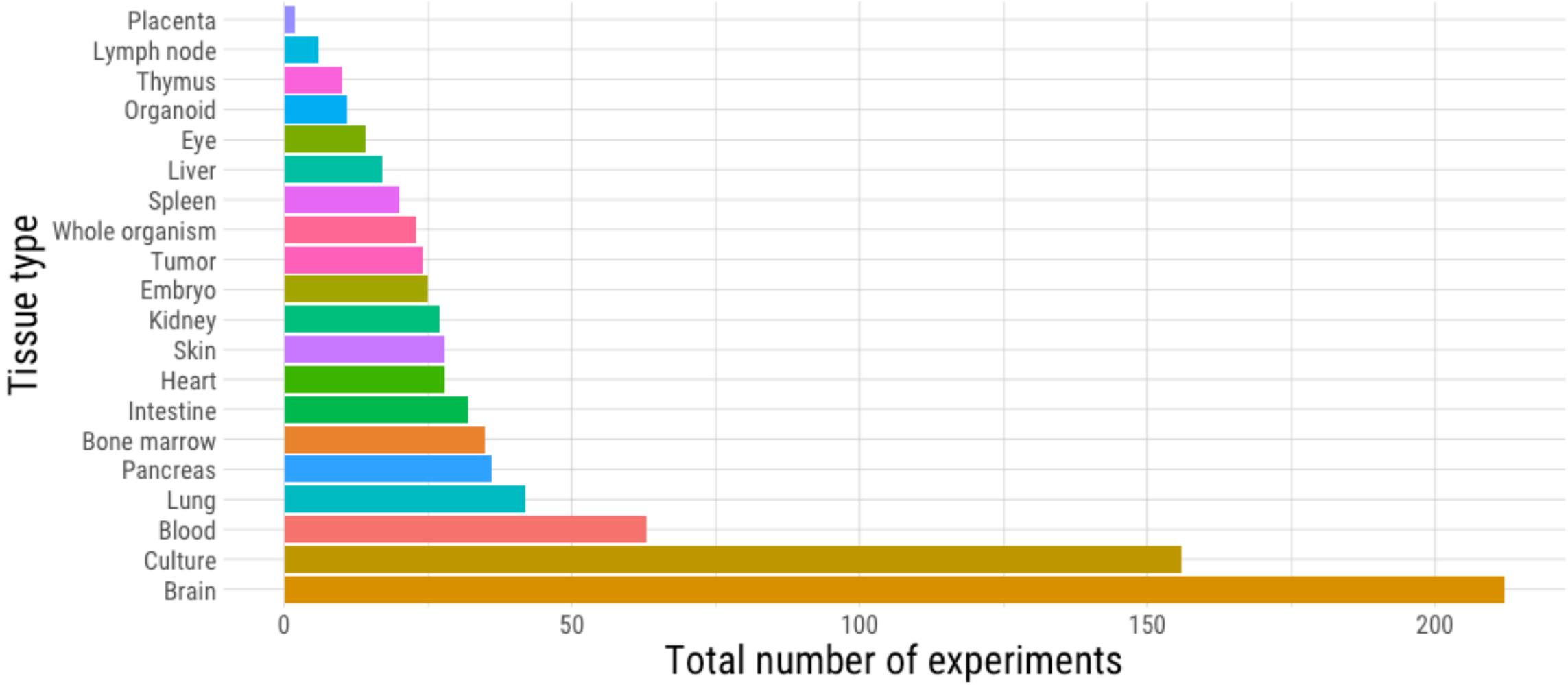
kb-python - You will need to have this software installed in a Conda environment on your laptop. We did this way back in lecture 1. If you are unable to install or use kb-python, just follow along with the lecture so you understand the concepts.

The landscape of single cell sequencing (scRNA-seq) experiments



Data source: Svensson et al., PMID 33247933; www.nxn.se/single-cell-studies/gui

scRNAseq experiments by tissue type

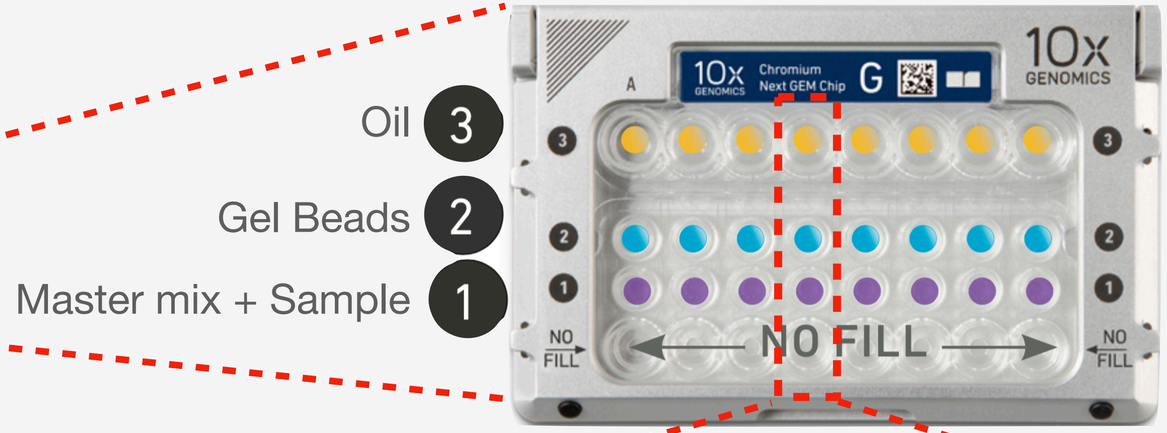


Data source: Svensson et al., PMID 33247933; www.nxn.se/single-cell-studies/gui

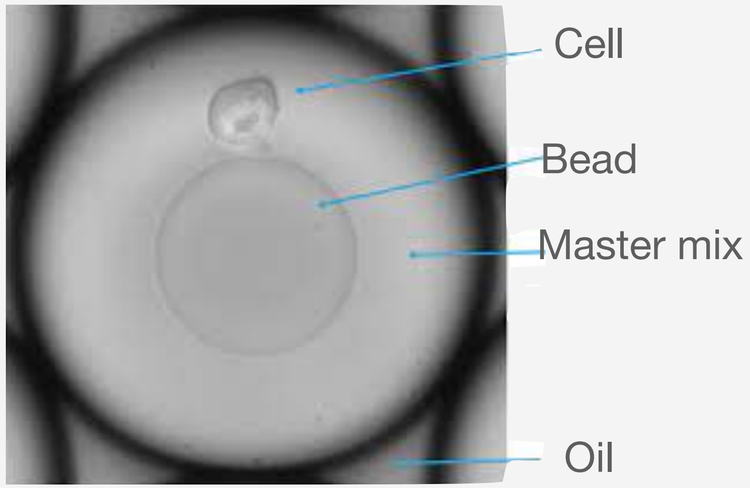
10x workflow for scRNA-seq

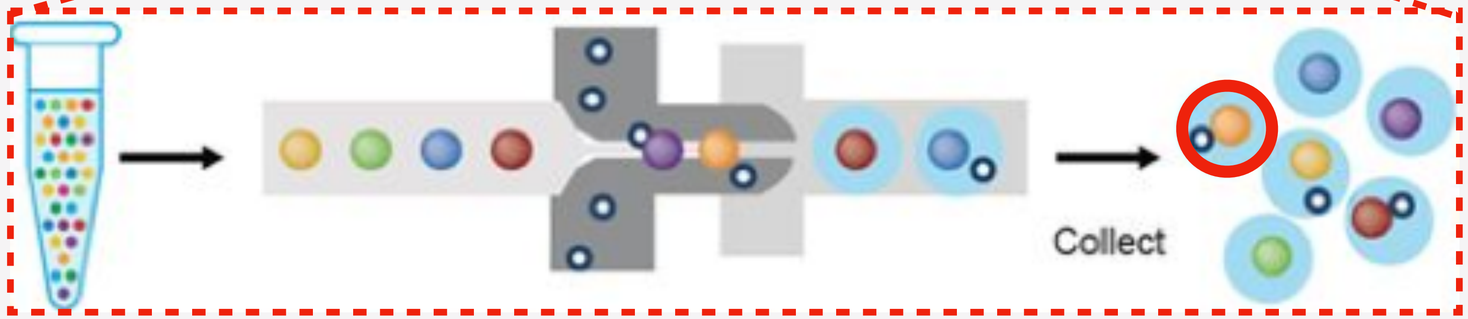
Chromium Controller

Microfluidic chip

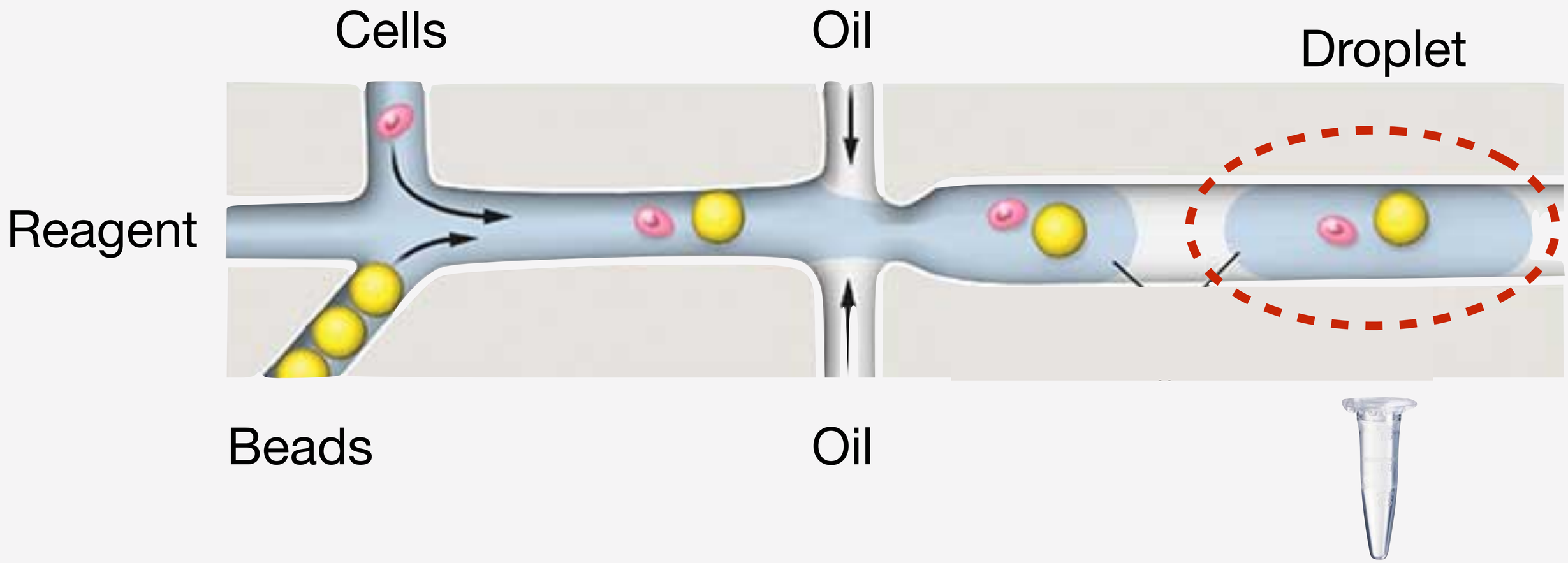


Droplet (a.k.a. GEM)

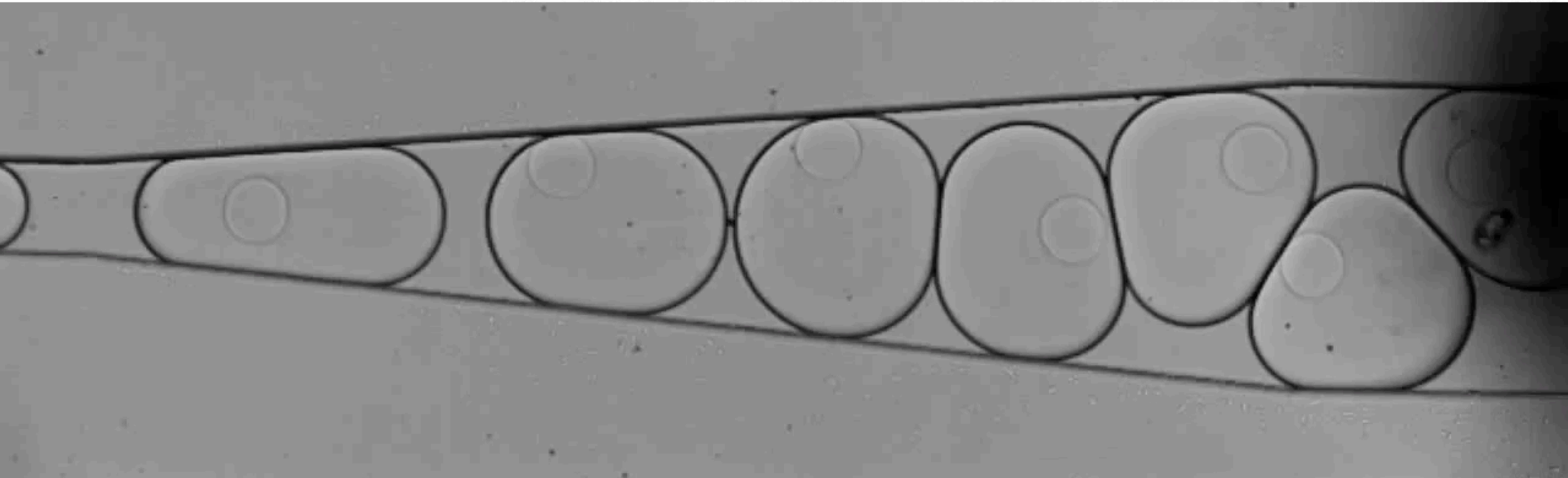




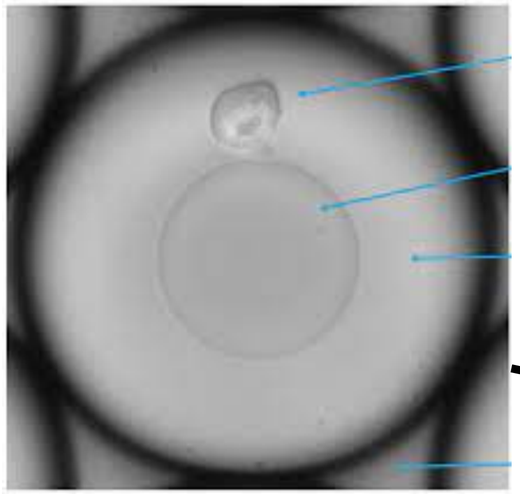
Droplet-based single cell RNAseq (scRNAseq, drop-seq)

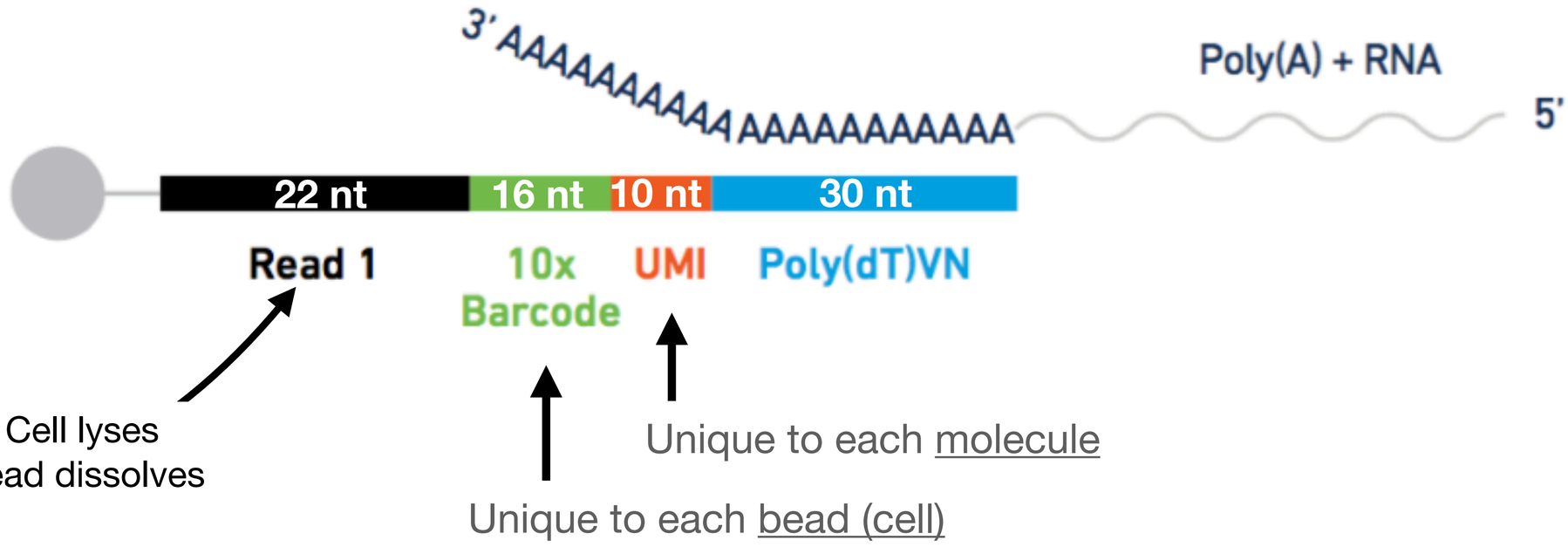


Droplet-based single-cell RNAseq
(e.g., drop-seq, 10X)

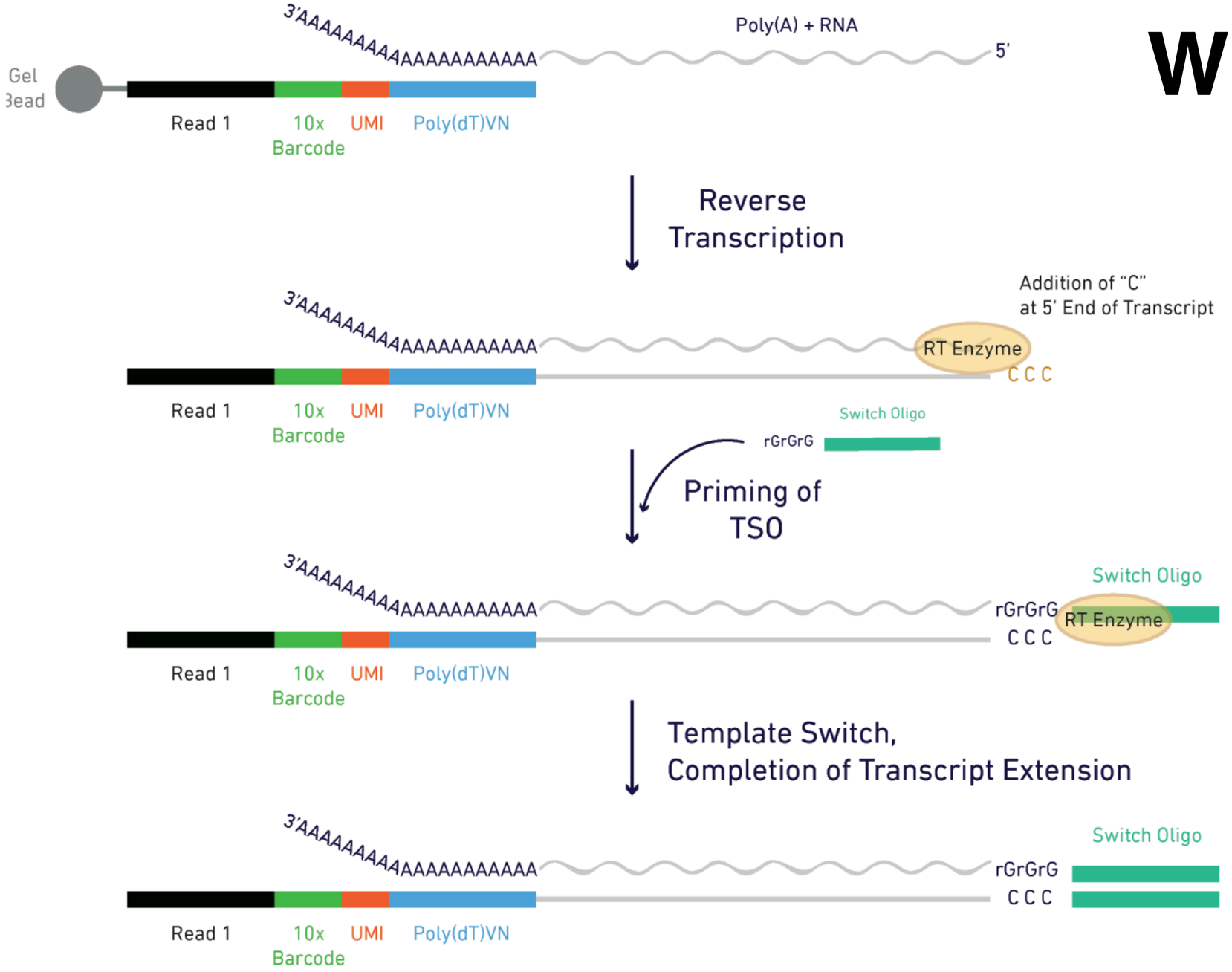


What happens in a droplet?



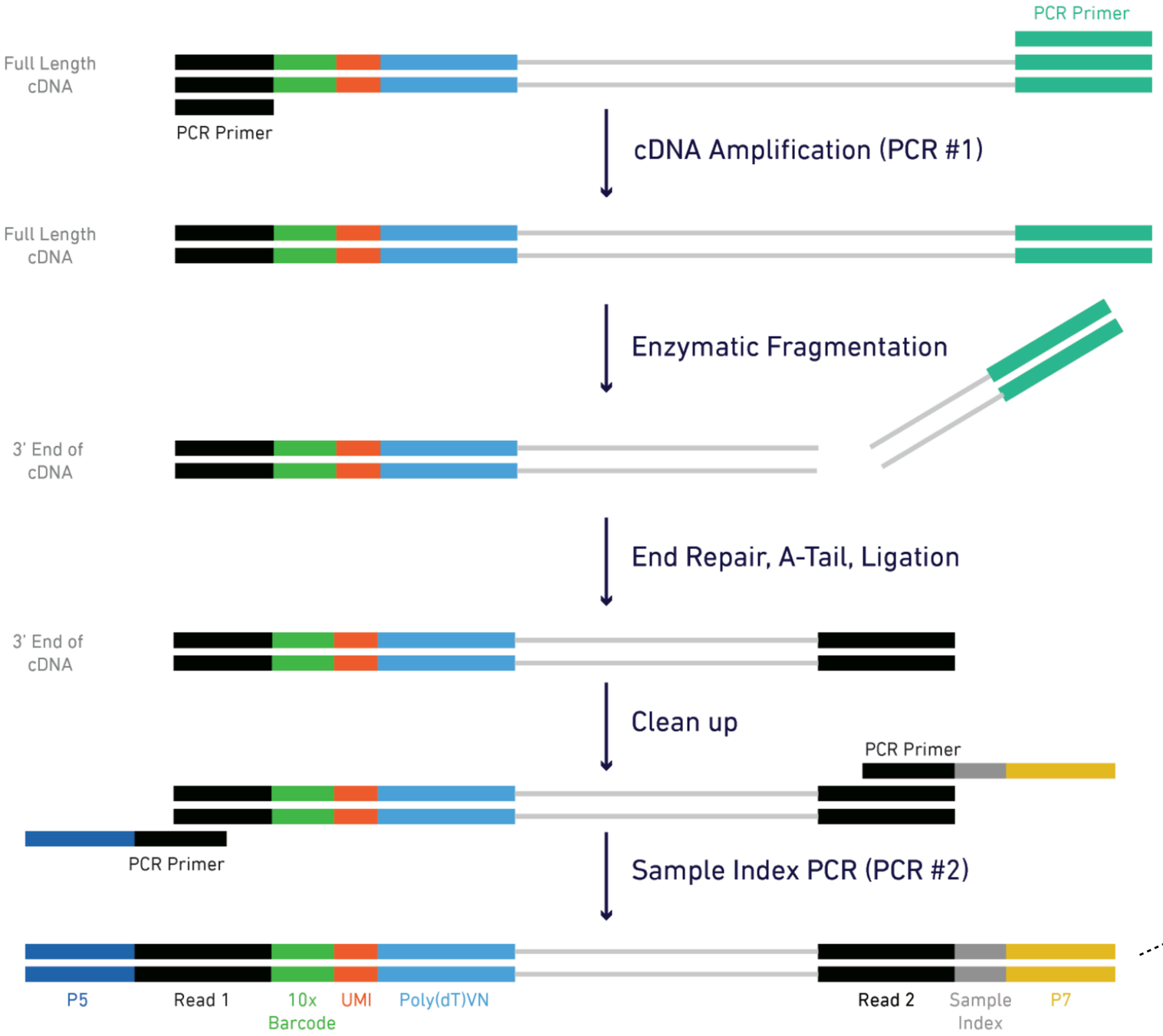


What happens in a droplet?

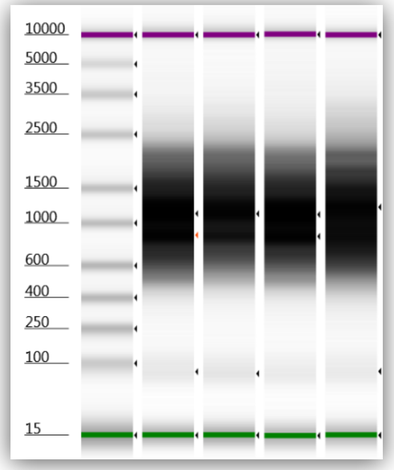


Droplets can now be broken, and remaining steps carried out in bulk

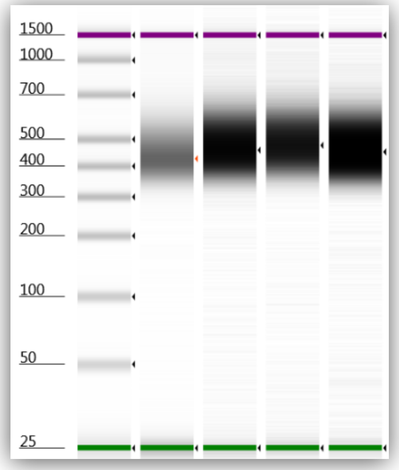
Library prep.
(3’ scRNA-seq)



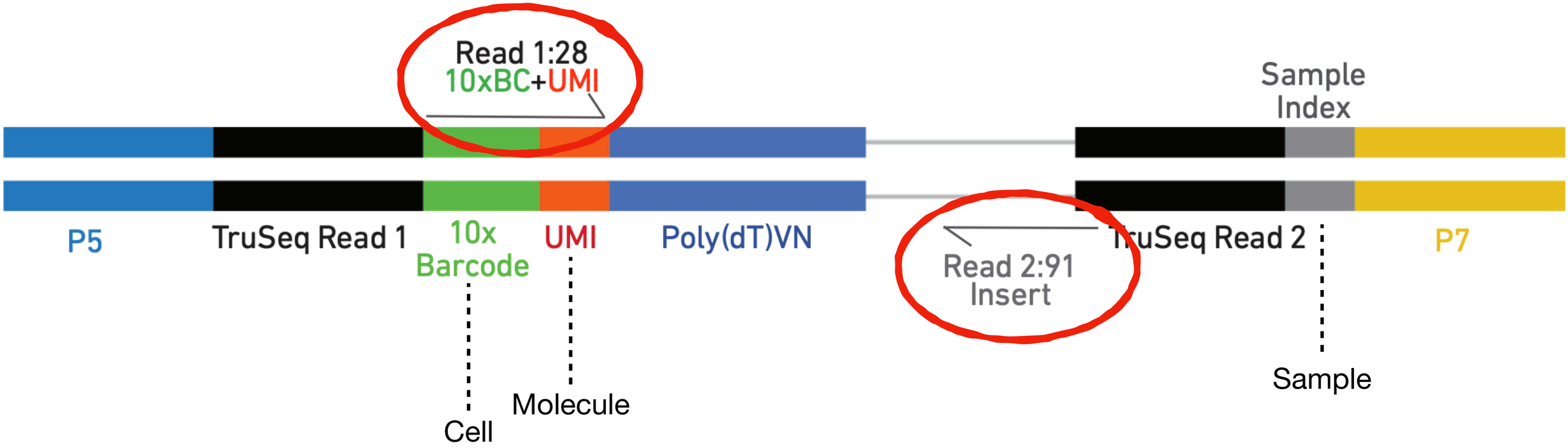
QC cDNA



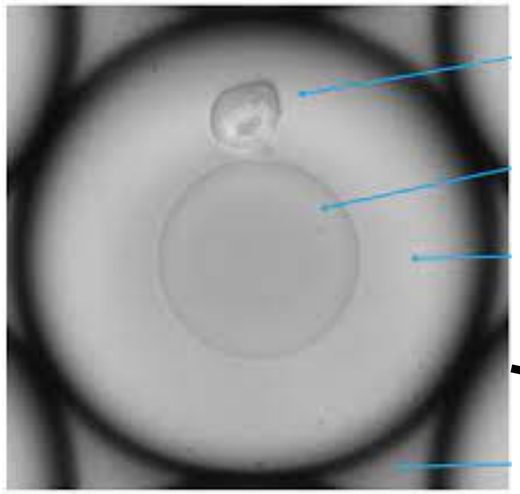
QC Libraries



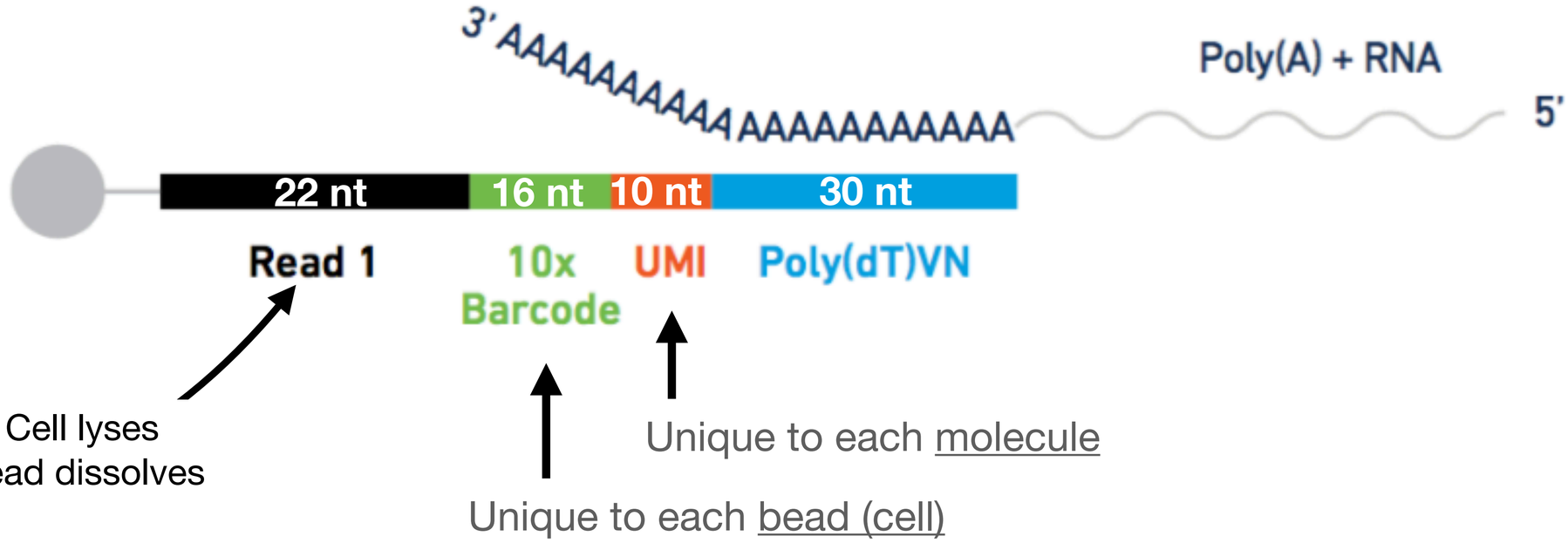
Sequencing


Read 1:28
10xBC+UMI
Sample
Index
P5
TruSeq Read 1
10x
Barcode
UMI
Poly(dT)VN
Read 2:91
Insert
TruSeq Read 2
P7
Cell
Molecule
Sample

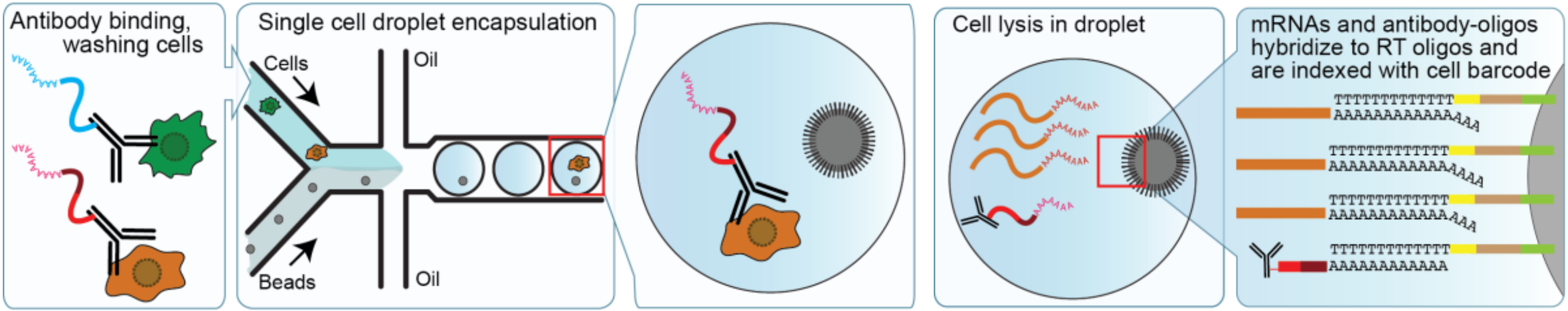
Barcoding is a flexible way to build new single cell assays



Cell lyses
Bead dissolves

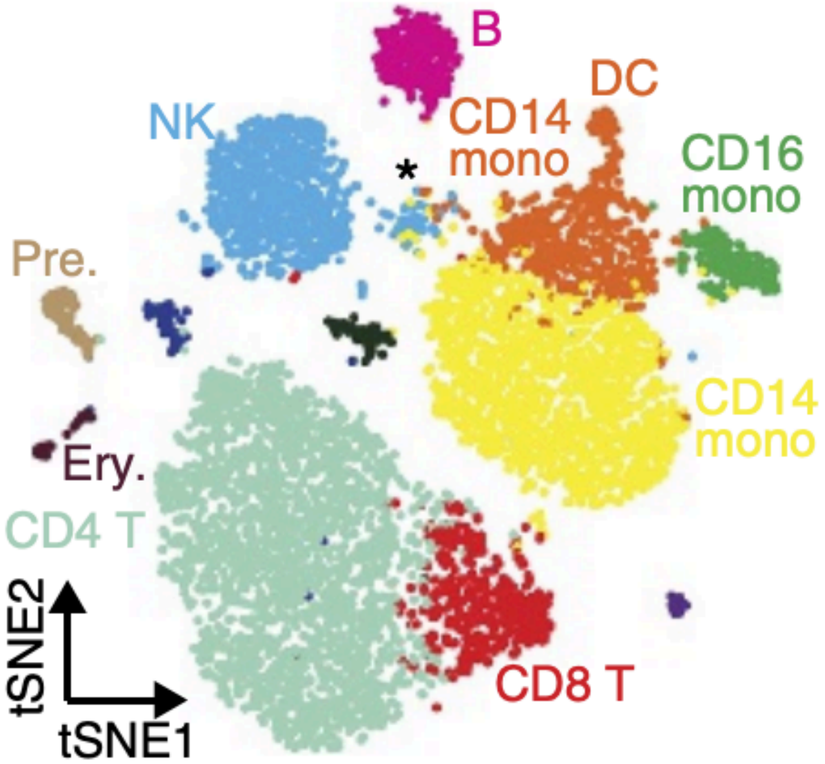


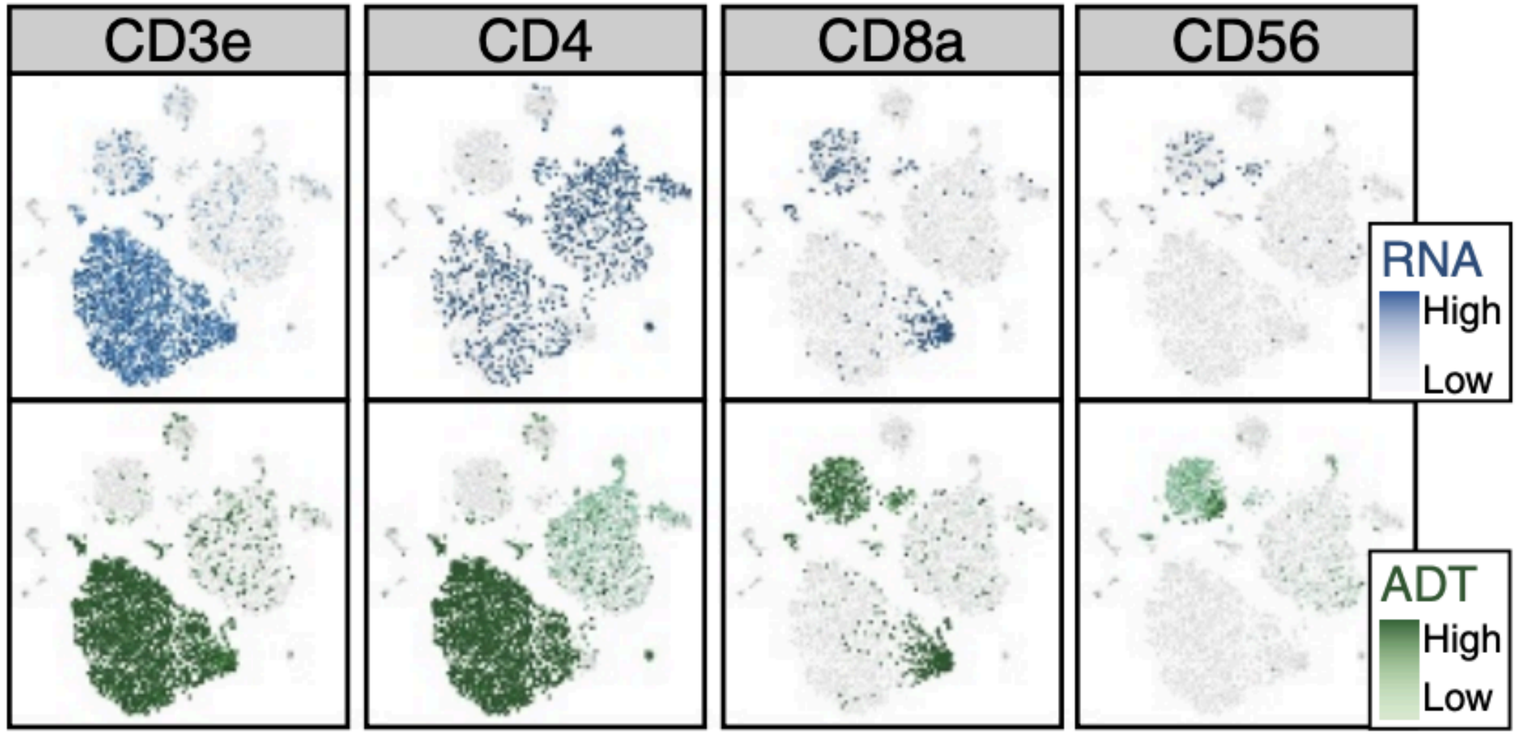
Cellular Indexing of Transcriptomes and Epitomes by Sequencing (CITE-seq)



Stoeckius et al., Nature Methods, 2017

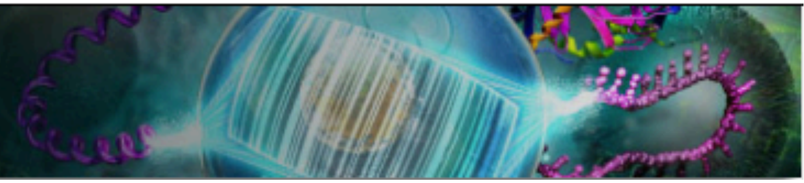
Cellular Indexing of Transcriptomes and Epitomes by Sequencing (CITE-seq)





Stoeckius et al., Nature Methods, 2017

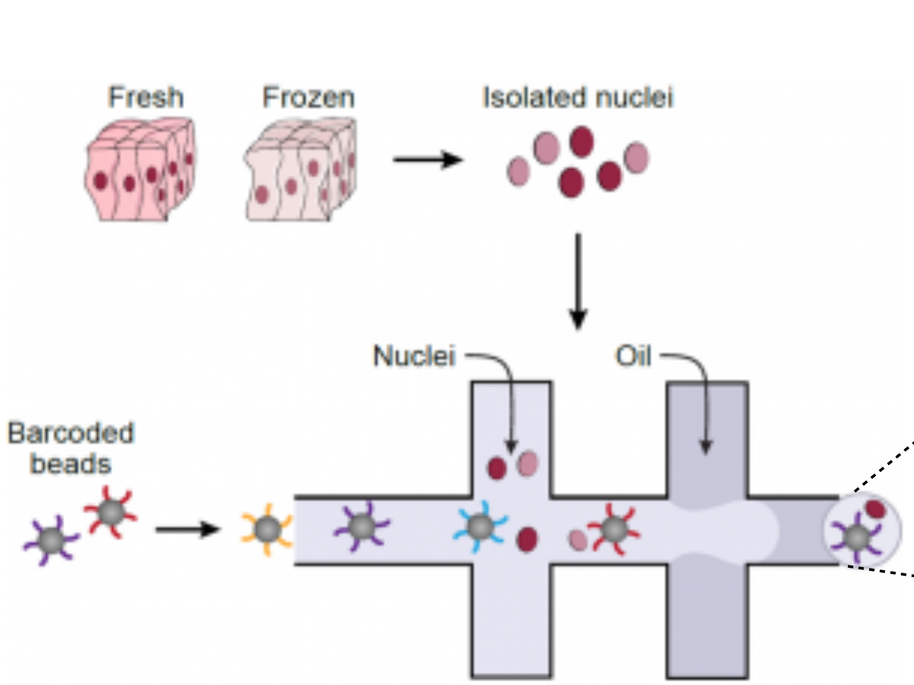
TotalSeq™ and Proteogenomic Analysis

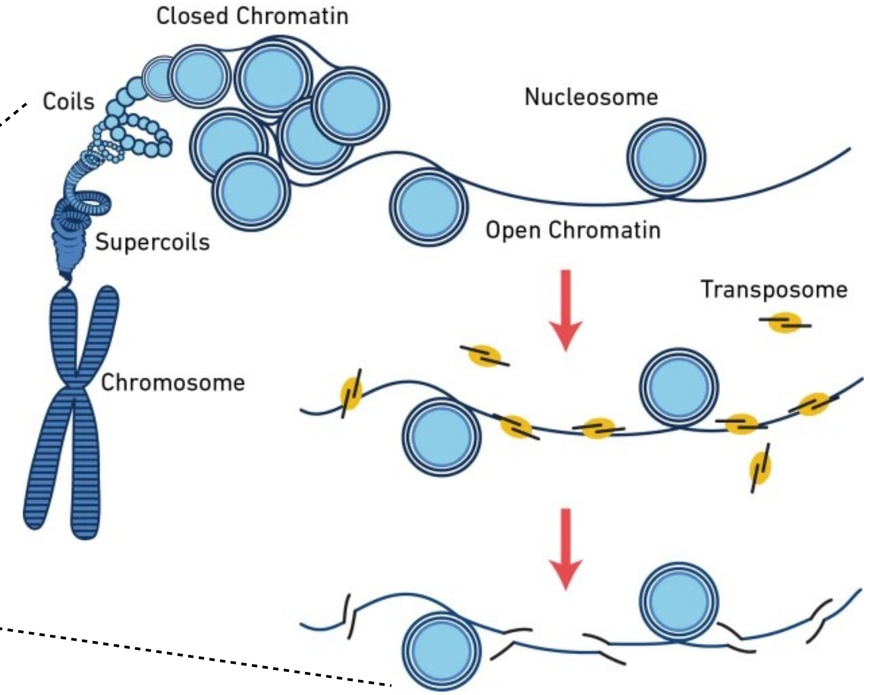


	Catalog	Barcode	Description	Clone	Reactivity	Sequence	Ensembl Gene Id
☐	100569	0001	CD4	RM4-5	Mouse	AACAAGACCCTTGAG	ENSMUSG00000023274
☐	100773	0002	CD8a	53-6.7	Mouse	TACCCGTAATAGCGT	ENSMUSG00000053977
☐	119729	0003	CD366 (Tim-3)	RMT3-23	Mouse	ATTGGCACTCAGATG	ENSMUSG00000020399
☐	109123	0004	CD279 (PD-1)	RMP1-30	Mouse	GAAAGTCAAAGCACT	ENSMUSG00000026285
☐	305239	0005	CD80	2D10	Human	ACGAATCAATCTGTG	ENSG00000121594
☐	305443	0006	CD86	IT2.2	Human	GTCTTTGTCAGTGCA	ENSG00000114013
☐	329743	0007	CD274 (B7-H1, PD-L1)	29E.2A3	Human	GTTGTCCGACAATAC	ENSG00000120217

Droplet methods are extremely versatile

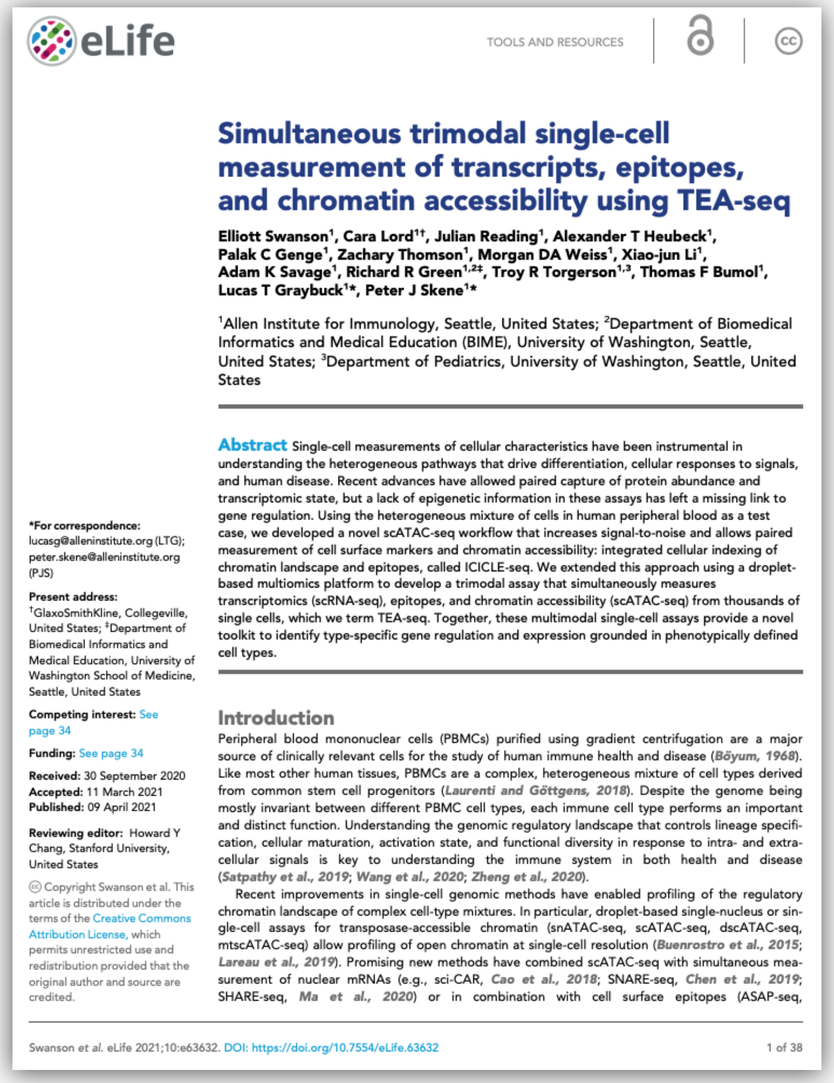
Single <u>nuclei</u> encapsulation for frozen or fixed tissue,
or for scATAC-seq





Droplet methods are extremely versatile

Multiplexed assays



eLife

TOOLS AND RESOURCES

Simultaneous trimodal single-cell measurement of transcripts, epitopes, and chromatin accessibility using TEA-seq

Elliott Swanson[1], Cara Lord[1†], Julian Reading[1], Alexander T Heubeck[1], Palak C Genge[1], Zachary Thomson[1], Morgan DA Weiss[1], Xiao-jun Li[1], Adam K Savage[1], Richard R Green[1,2‡], Troy R Torgerson[1,3], Thomas F Bumol[1], Lucas T Graybuck[1*], Peter J Skene[1*]

[1]Allen Institute for Immunology, Seattle, United States; [2]Department of Biomedical Informatics and Medical Education (BIME), University of Washington, Seattle, United States; [3]Department of Pediatrics, University of Washington, Seattle, United States

Abstract Single-cell measurements of cellular characteristics have been instrumental in understanding the heterogeneous pathways that drive differentiation, cellular responses to signals, and human disease. Recent advances have allowed paired capture of protein abundance and transcriptomic state, but a lack of epigenetic information in these assays has left a missing link to gene regulation. Using the heterogeneous mixture of cells in human peripheral blood as a test case, we developed a novel scATAC-seq workflow that increases signal-to-noise and allows paired measurement of cell surface markers and chromatin accessibility: integrated cellular indexing of chromatin landscape and epitopes, called ICICLE-seq. We extended this approach using a droplet-based multiomics platform to develop a trimodal assay that simultaneously measures transcriptomics (scRNA-seq), epitopes, and chromatin accessibility (scATAC-seq) from thousands of single cells, which we term TEA-seq. Together, these multimodal single-cell assays provide a novel toolkit to identify type-specific gene regulation and expression grounded in phenotypically defined cell types.

***For correspondence:** lucasg@alleninstitute.org (LTG); peter.skene@alleninstitute.org (PJS)

Present address: †GlaxoSmithKline, Collegeville, United States; ‡Department of Biomedical Informatics and Medical Education, University of Washington School of Medicine, Seattle, United States

Competing interest: See page 34

Funding: See page 34

Received: 30 September 2020
Accepted: 11 March 2021
Published: 09 April 2021

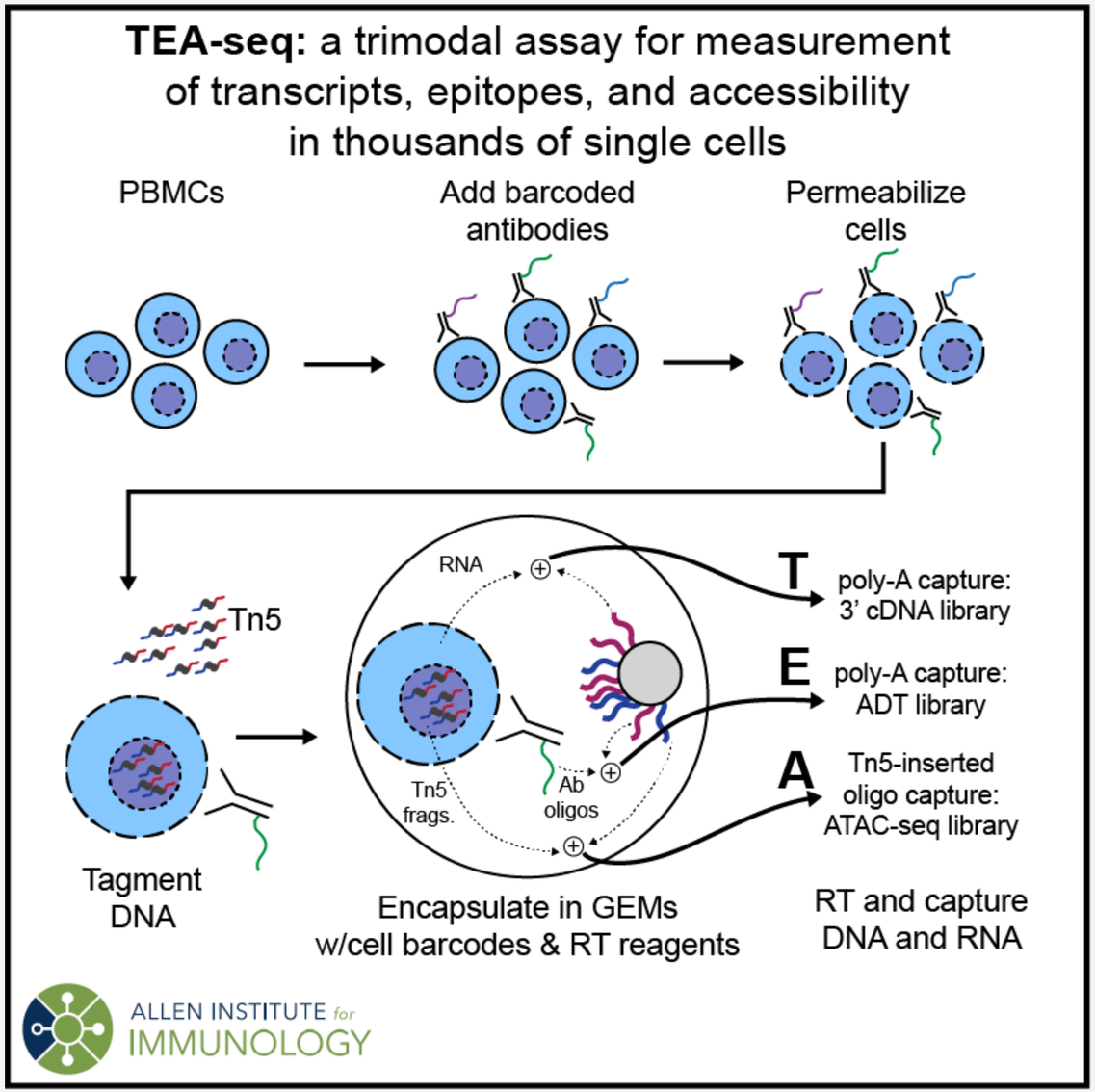
Reviewing editor: Howard Y Chang, Stanford University, United States

Introduction

Peripheral blood mononuclear cells (PBMCs) purified using gradient centrifugation are a major source of clinically relevant cells for the study of human immune health and disease (***Böyum, 1968***). Like most other human tissues, PBMCs are a complex, heterogeneous mixture of cell types derived from common stem cell progenitors (***Laurenti and Göttgens, 2018***). Despite the genome being mostly invariant between different PBMC cell types, each immune cell type performs an important and distinct function. Understanding the genomic regulatory landscape that controls lineage specification, cellular maturation, activation state, and functional diversity in response to intra- and extracellular signals is key to understanding the immune system in both health and disease (***Satpathy et al., 2019***; ***Wang et al., 2020***; ***Zheng et al., 2020***).

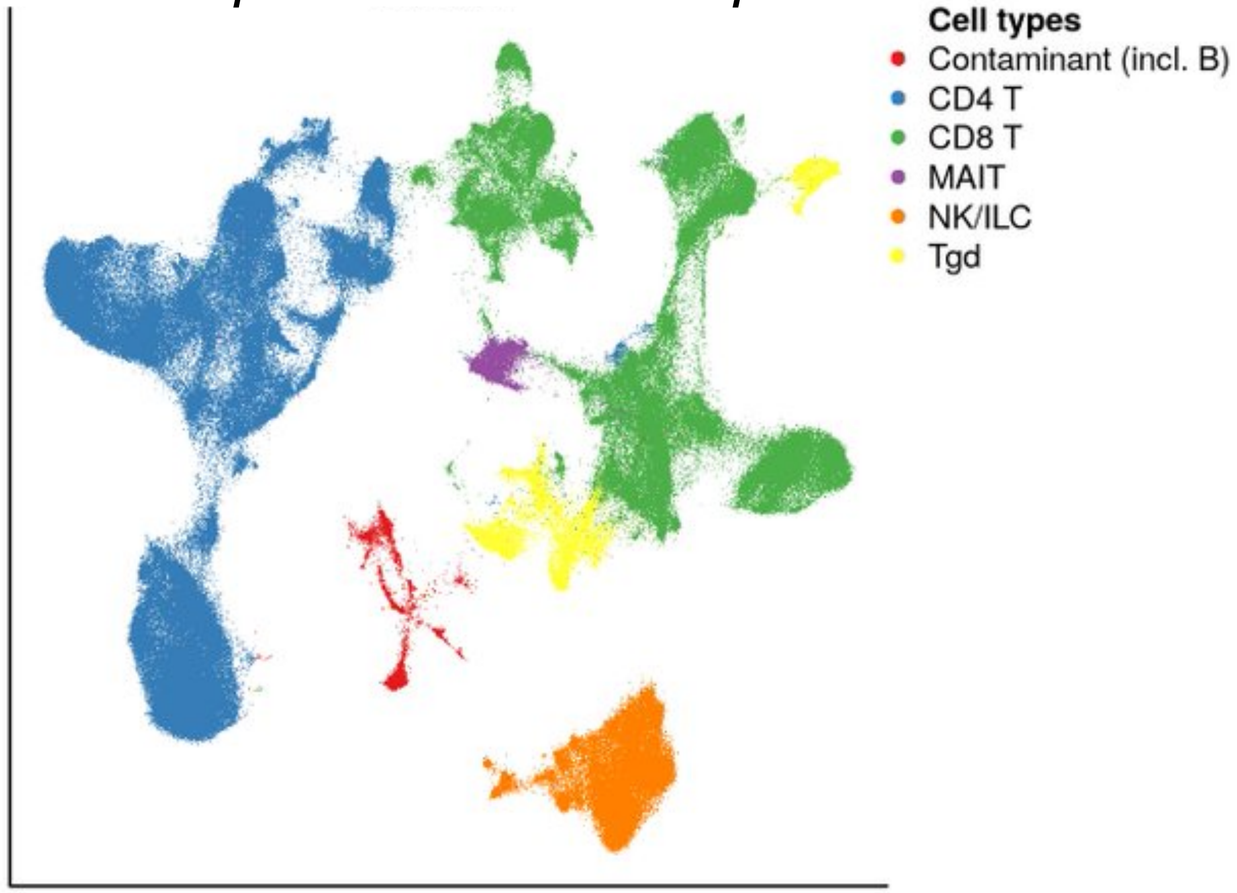
Recent improvements in single-cell genomic methods have enabled profiling of the regulatory chromatin landscape of complex cell-type mixtures. In particular, droplet-based single-nucleus or single-cell assays for transposase-accessible chromatin (snATAC-seq, scATAC-seq, dscATAC-seq, mtscATAC-seq) allow profiling of open chromatin at single-cell resolution (***Buenrostro et al., 2015***; ***Lareau et al., 2019***). Promising new methods have combined scATAC-seq with simultaneous measurement of nuclear mRNAs (e.g., sci-CAR, ***Cao et al., 2018***; SNARE-seq, ***Chen et al., 2019***; SHARE-seq, ***Ma et al., 2020***) or in combination with cell surface epitopes (ASAP-seq,

Swanson et al., eLife, 2021

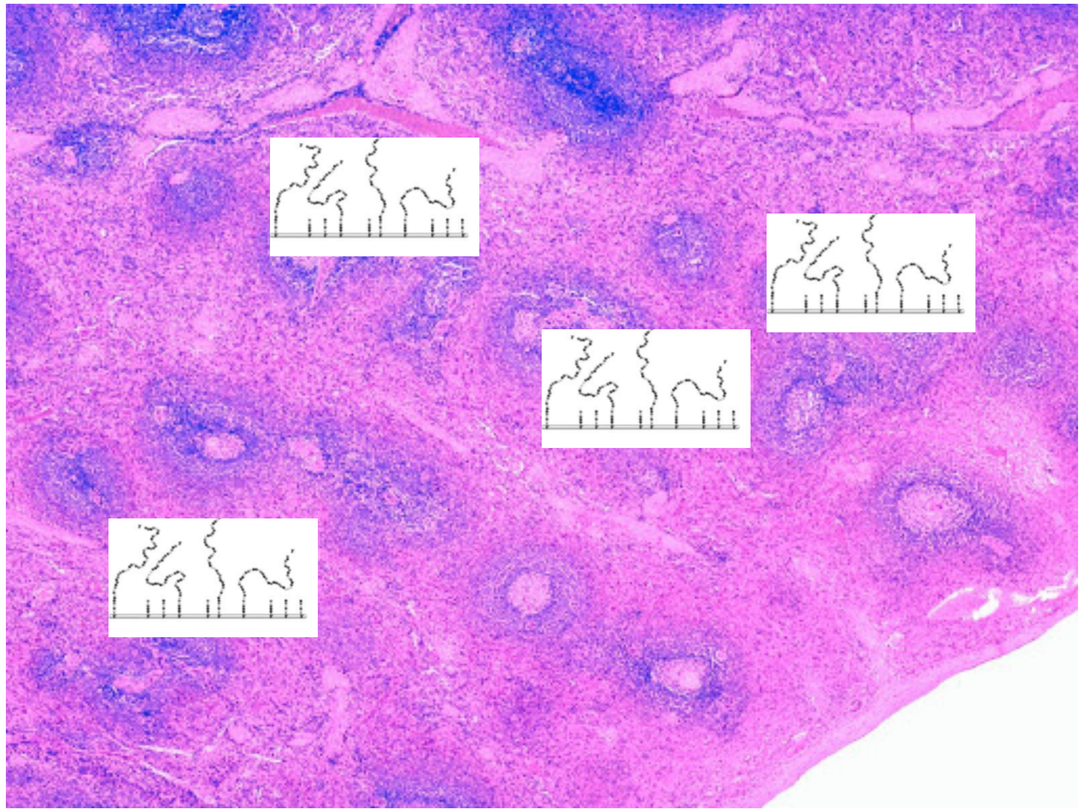


Spatial information is lost in scRNA-seq

Spleen scRNA-seq

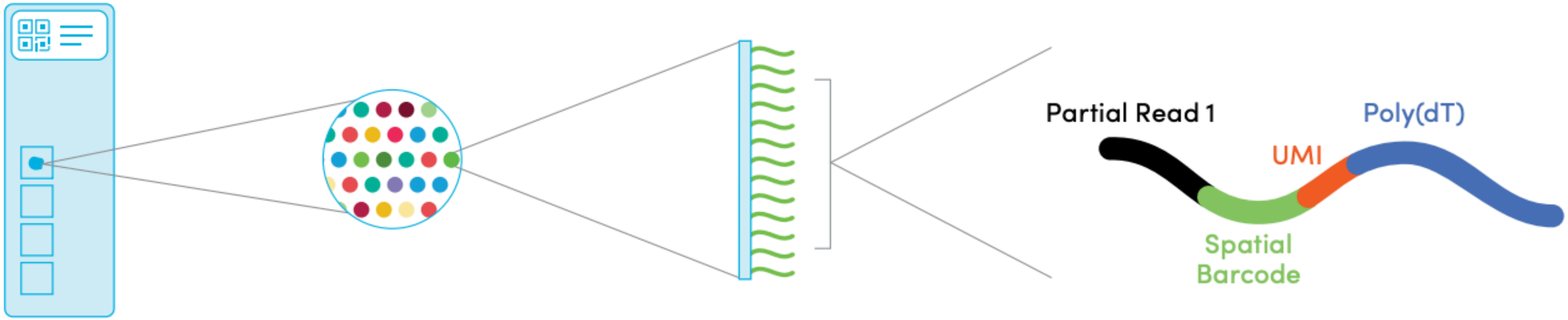


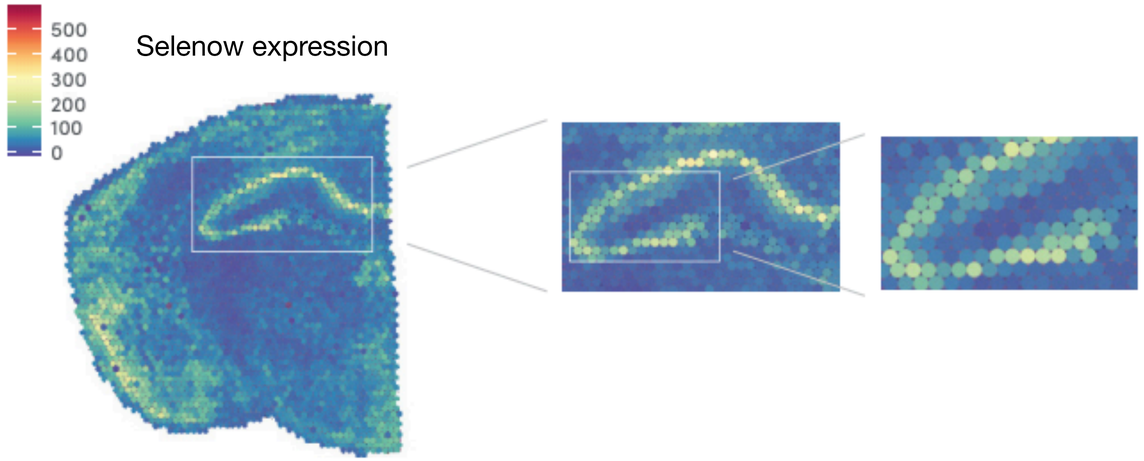
Actual spleen

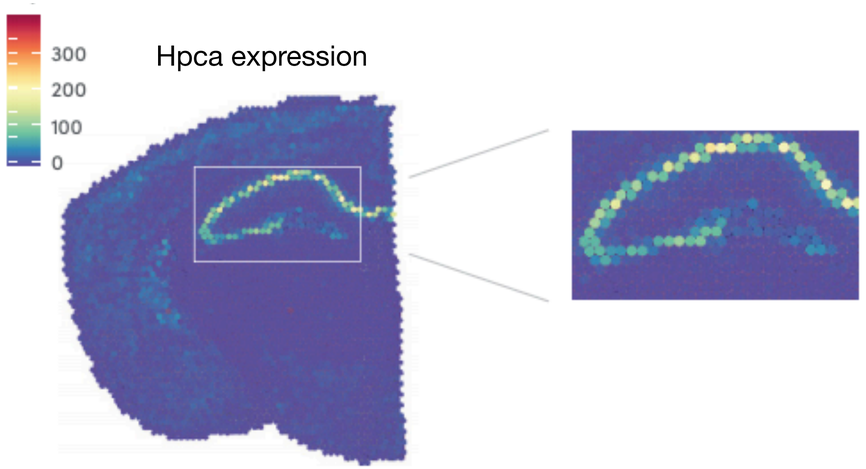


Becht et al., Nature Methods, 2018

Spatial transcriptomics with ‘zipcoded’ beads on a slide

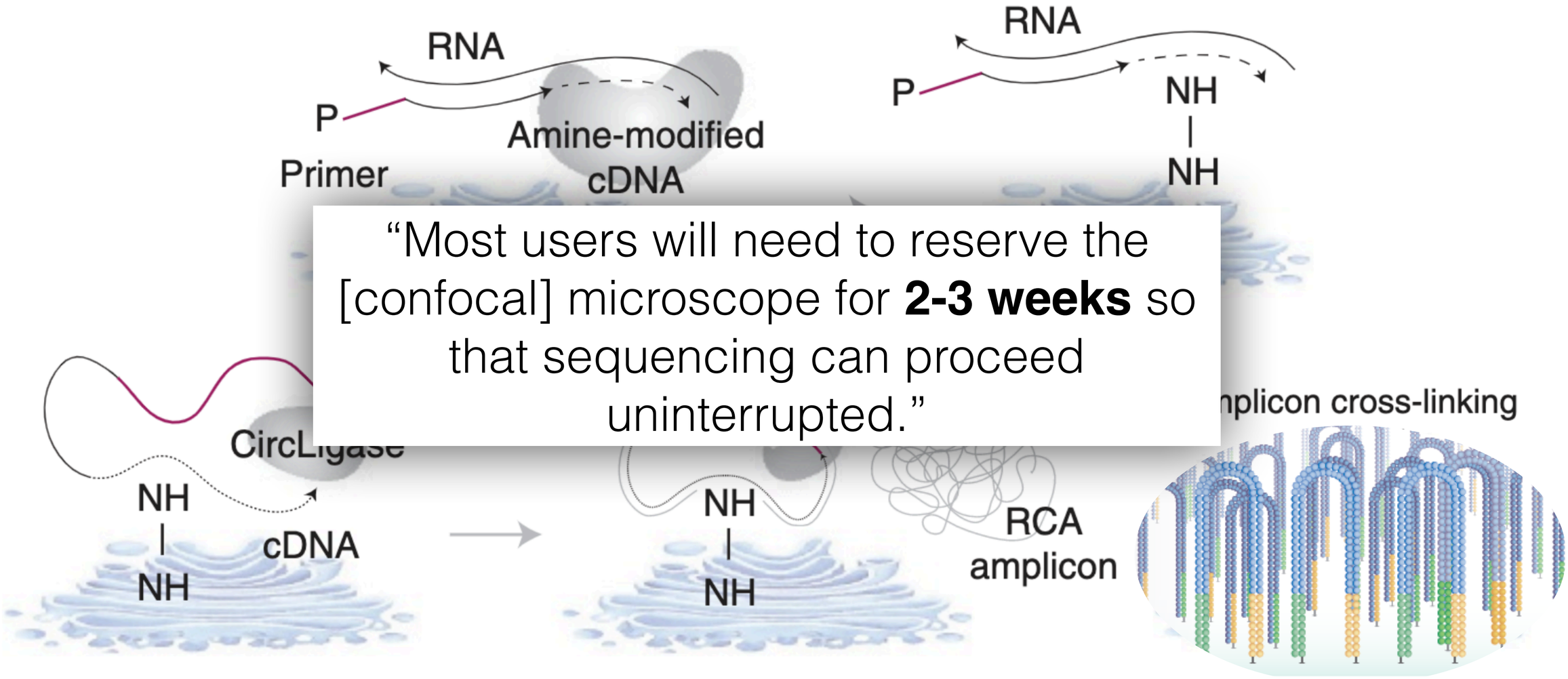






10X Genomics Visium technical note

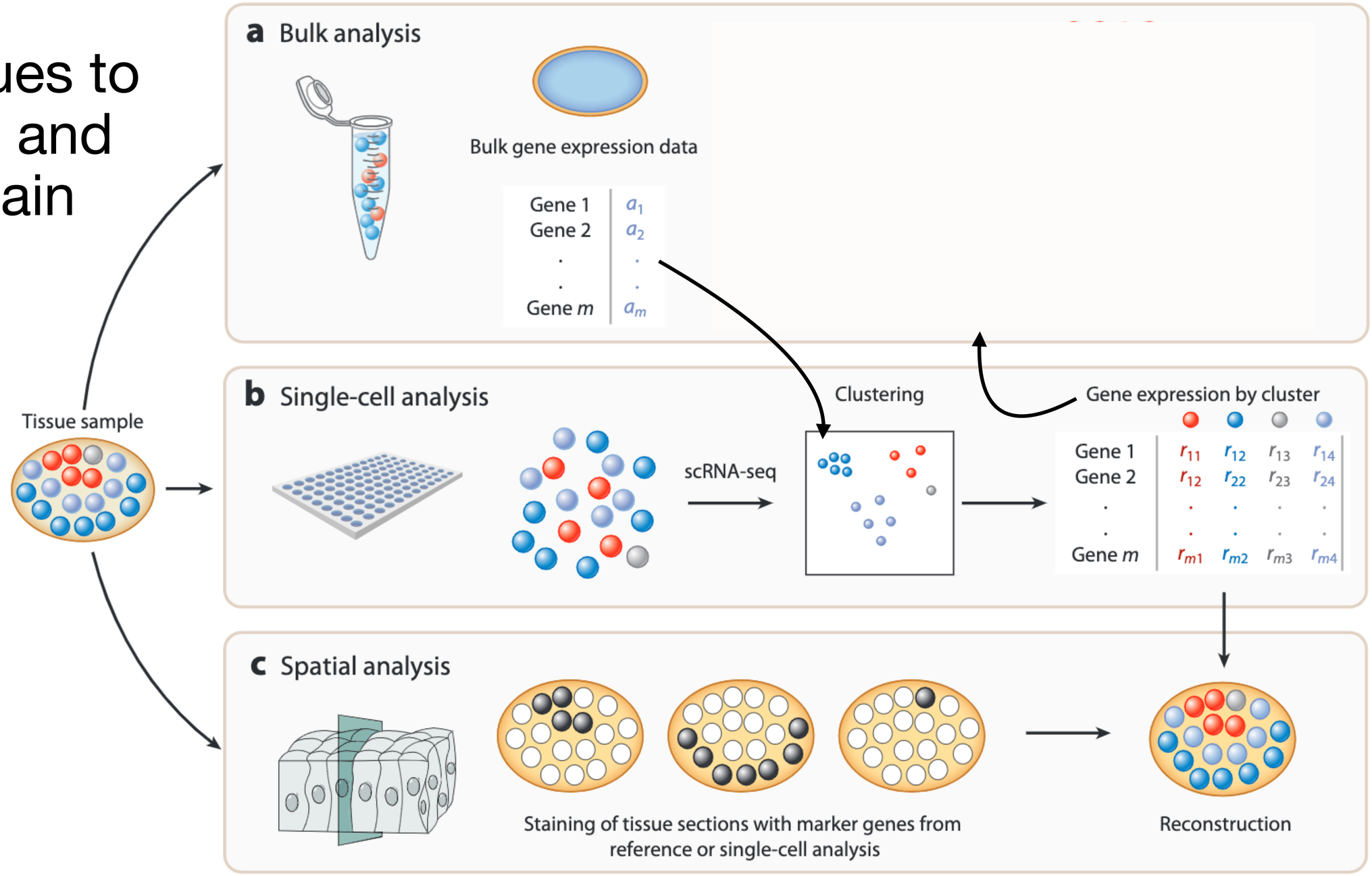
Fluorescent In Situ Sequencing (FISSEQ)



"Most users will need to reserve the [confocal] microscope for **2-3 weeks** so that sequencing can proceed uninterrupted."

Lee JH, et al., 2015

From Tissues to cell types and back again



Chen, Teichmann and Meyer; Annu. Rev. Biomed. Data Sci. 2018.

End of part 1

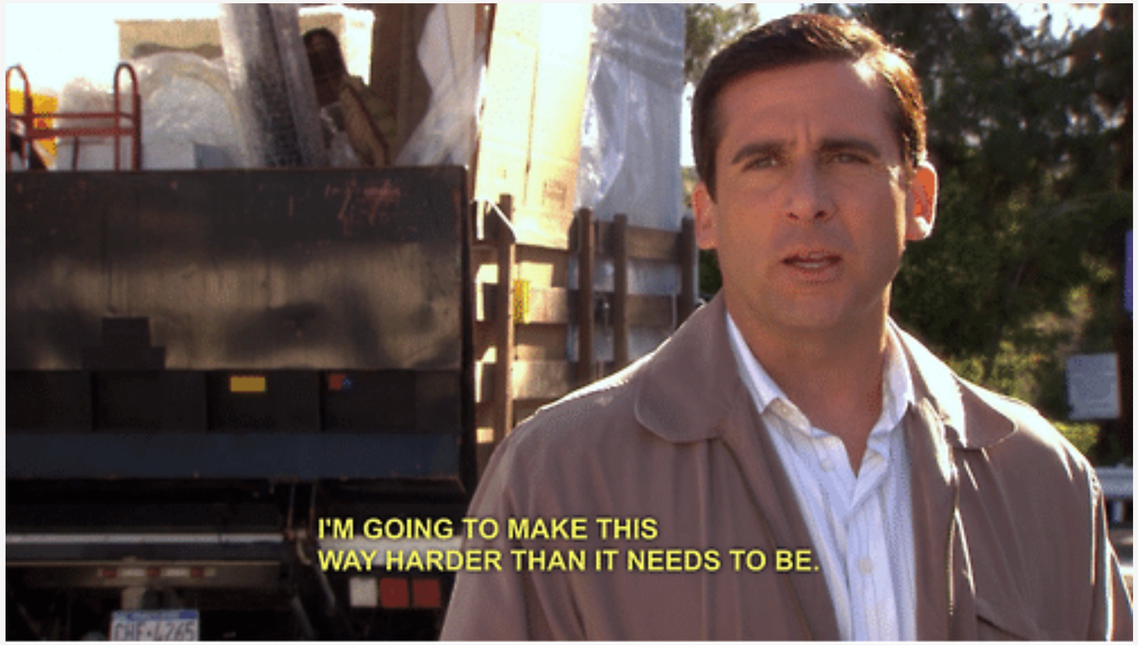
Practical considerations for single cell studies

Practical considerations for single cell studies

Do you *really need* single cell resolution,
or would flow cytometry or bulk RNA-seq be sufficient to answer your question?

scRNA-seq ⊁ bulk RNA-seq

	Single cell	Bulk
Heterogeneity	✅	❌
Trajectory	✅	❌
Multiplexing	✅	❌
Depth	+	+++
Cost	+++++++++	+
Analytic burden	+++	+



Costs associated with bulk* sequencing

From pre-course lecture

Library Preparation

TruSeq kit	cat#	cost
stranded total RNA LT (w/ RiboZero)	RS-122-2201	$5,067
stranded mRNA LT	RS-122-2101	$2,430

all kits process 48 samples

Sequencing

mid output
130M clusters

high output
400M clusters

cycles	cat#	cost
300	FC-404-2004	$4,222
150	FC-404-2002	$2,635
75	FC-404-2005	$1,374

12 samples
library prep = $51/sample
sequencing = $115/sample
data output = 30M reads/sample

**Costs associated with scRNAseq are much more complicated*

Costs associated with scRNA-seq

Library Preparation

Sequencing

<u>P2 flowcell</u>
400M clusters

<u>P3 flowcell</u>
1.1B clusters

10x Genomics	cat#	cost
Chromium Next GEM Single Cell 3' kit v3.1	1000130 (4rxn)	~$2000/sample
Chromium Next GEM Single Cell 3' LT kit v3.1	1000322 (4rxn)	~$900/sample

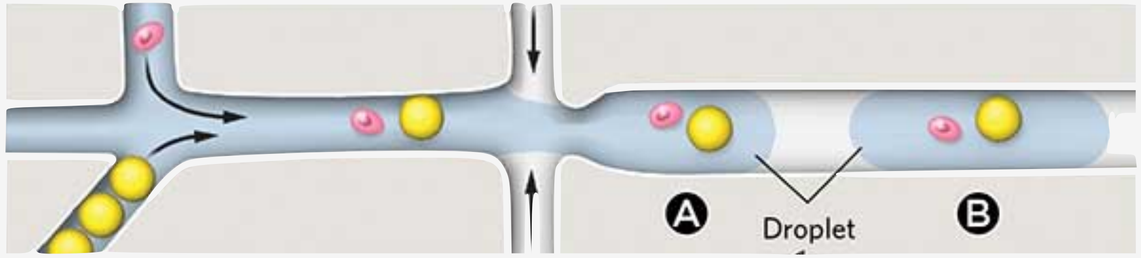
Flow cell	cycles	cat#	cost	samples
P2	100 (+38)	20046811	$1420	~2
P3	100 (+38)	20040559	$3200	~6

2 samples; 10,000 cells/sample; sequenced on P2
library prep = ~$2000/sample
sequencing = ~$700/sample
data output = ~20,000 reads/cell

Practical considerations for single cell studies

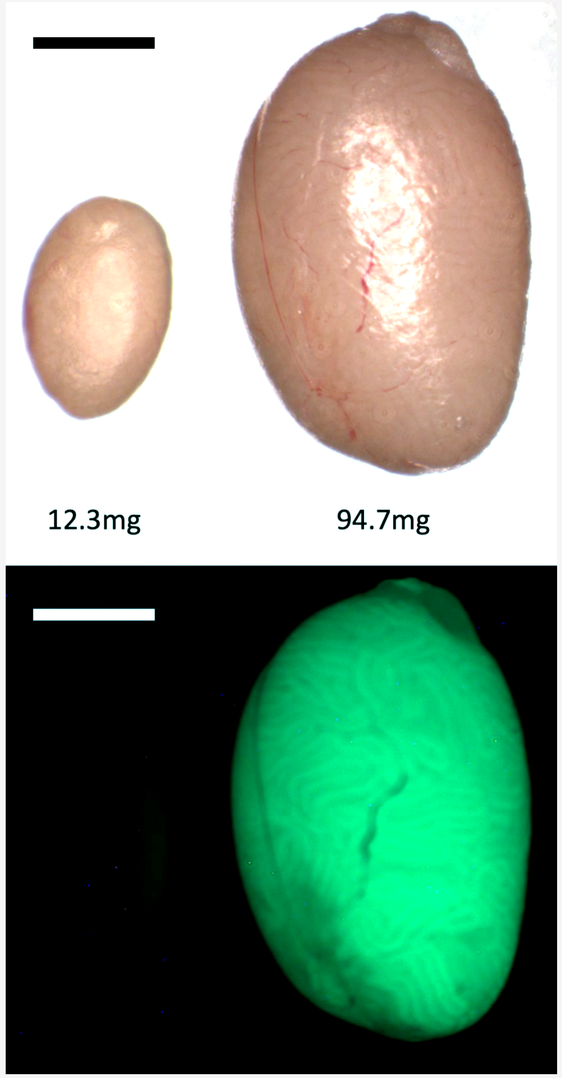
Cell viability is critical to the success of scRNA-seq.

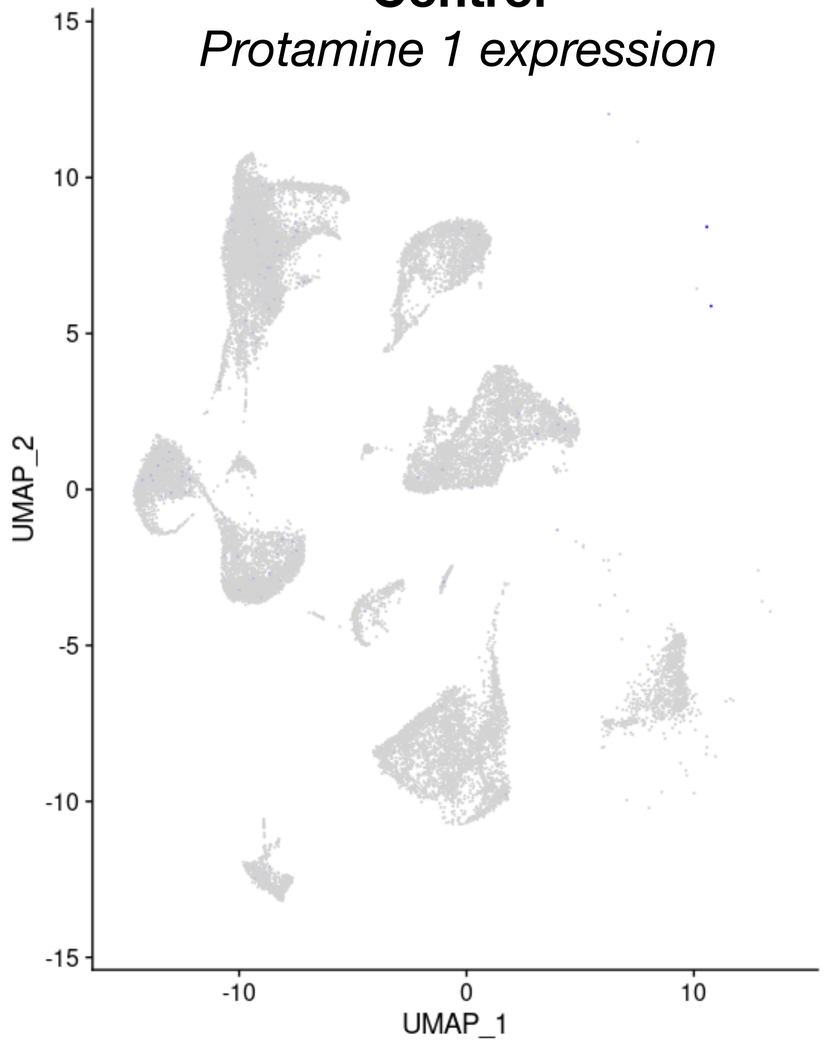
Viability should be >90%

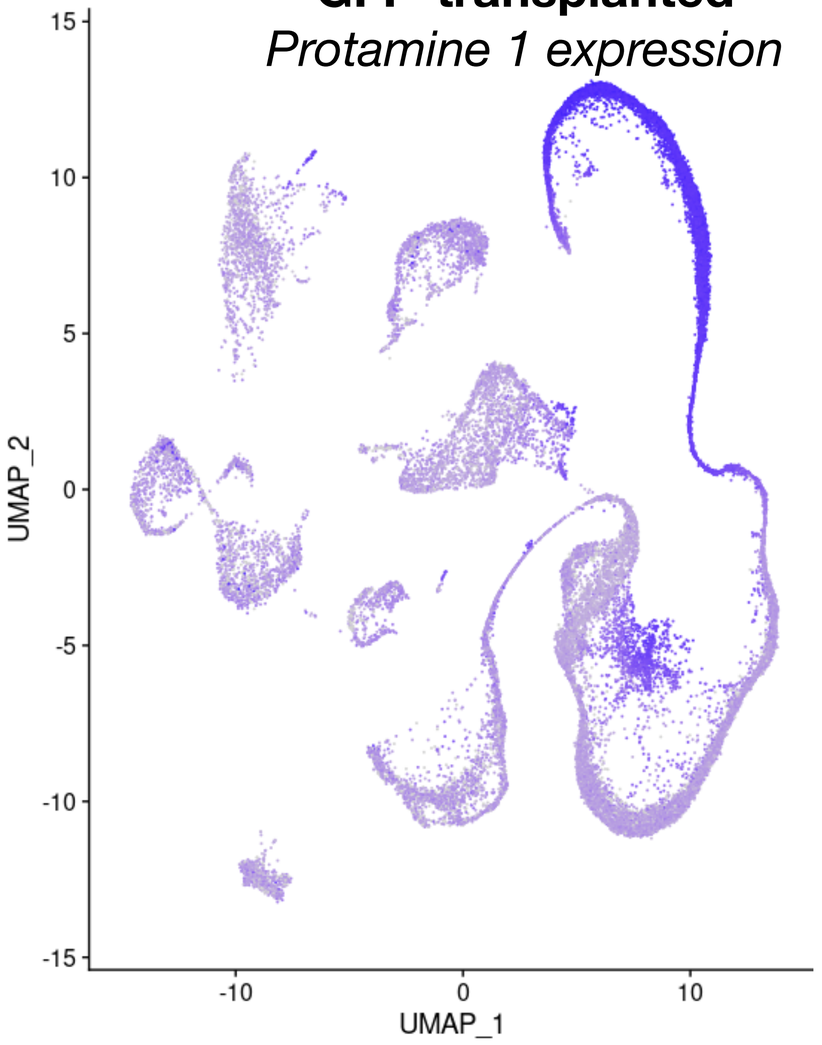


Video source: Instagram account of MicrobialEcology
https://www.instagram.com/microbialecology/

Ambient RNA is your enemy!

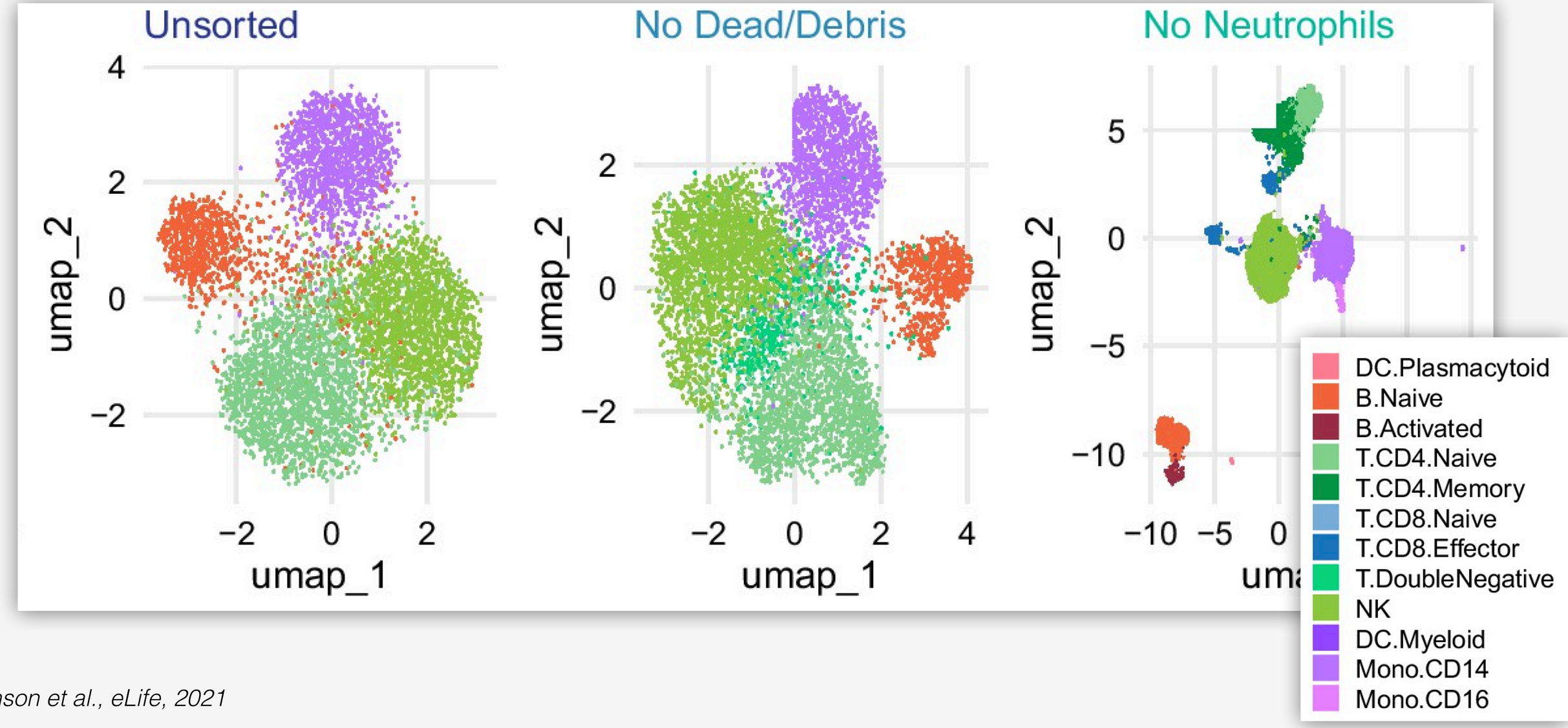






Images courtesy of Eoin Whelan, PhD (Brinster Lab, PennVet)

Certain cell types can be problematic in single cell assays

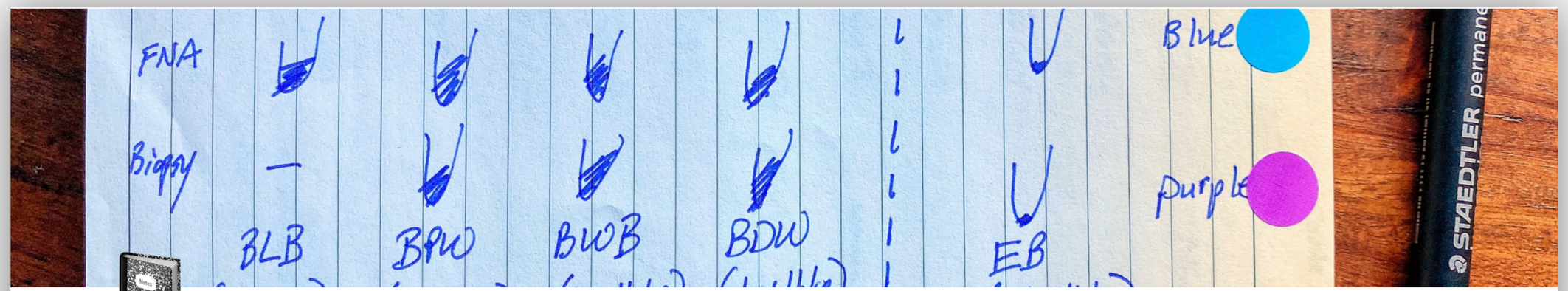


Swanson et al., eLife, 2021

Recommendations for high quality cell preps

- Cell isolation protocols optimized for flow cytometry or *in vitro* functional assays may be insufficient for scRNA-seq experiments.
- If you are accustomed to digesting tissues at 37ºC, consider switching to a 30min digestion at 6ºC with a cold active protease. See **this paper**
- If you need a starting point for tissue dissociation protocols, use the **Worthinton guide**
- Avoid mechanical dissociation (e.g. bashing up lymph nodes with a syringe plunger).
- Regardless of whether you use bead-based purification or FACS, you should carry out a dead cell removal using this kit (or a similar one). Viability should be > 90%.
- Bead-based purification of cells is preferred over flow sorting, because it is usually faster and results in happier cells. Depending on your experimental goals, it may be totally fine if purity isn't great, since you'll get separation of cell populations in the scRNA-seq
- Keep everything cold, cold, cold. Cells should **always** be kept on ice....even when walking from the bench to the centrifuge. Buffers should be pre-chilled on ice. Centrifuges should be pre-chilled to 4deg. This will help to limit cell death and clumping.
- Avoid using EDTA in your tissue dissociation buffers. Excessive amount of EDTA (> 0.1mM) or magnesium (> 3mM) will inhibit reverse transcription reactions and lead to reduced complexity in the scRNA-seq dataset.

scRNA-seq protocols



Protocols

This website provides access to all standard protocols and procedures used at the PennVet Center for Host-Microbial Interactions (CHMI). If you're affiliated with CHMI or a member of the Beiting Lab, then you can collaboratively edit any of these protocols. This wiki-style approach gives us a central place to develop our lab protocols, while always keeping the most up-to-date protocol in one convenient and publicly accessible place. If you're not part of CHMI, no worries, you can still peruse the site to see our protocols. Comments and feedback are welcome!

Bulk RNA-seq and ATAC-seq

- 📌 RNA-seq FAQS
- Stranded mRNA Prep, Ligation (Updated TruSeq)
- Stranded Total RNA Prep, Ligation with Ribo-Zero Plus (Updated TruSeq)
- Low input RNAseq (SMART-Seq HT PLUS Kit)
- RNA-seq on FFPE tissue (Takara-Clontech total RNA pico mammalian v2/v3)
- DRAFT miRNA Library Prep Protocol

Microbial genomics

- 📌 Microbiome FAQs
- 🧬 16S seq
- NexteraFLEX
- 16S Sanger
- 🖥️ Bioinformatics for microbial genomics
- 🏴 Legacy micro protocols

protocols.hostmicrobe.org

Practical considerations for single cell studies

Data analysis

System Requirements

Hardware

Cell Ranger pipelines run on Linux systems that meet these minimum requirements:

- 8-core Intel or AMD processor (16 cores recommended)
- 64GB RAM (128GB recommended)
- 1TB free disk space
- 64-bit CentOS/RedHat 6.0 or Ubuntu 12.04

The minimum requirement of 64GB RAM will allow `cellranger aggr` to aggregate up to 250k cells, more memory will be required beyond that.

"Another challenge in scRNA-seq preprocessing is the amount of data that must be processed. A single-cell experiment can generate 10^6–10^{10} reads from 10^3–10^6 cells. This is leading to bottlenecks in analysis: for example, the current standard program for preprocessing 10x Genomics Chromium scRNA-seq, the Cell Ranger software, requires approximately 22 h to process 784 million reads using 1.5 Tb of disk space"

nature biotechnology

LETTERS

https://doi.org/10.1038/s41587-021-00870-2

Check for updates

Modular, efficient and constant-memory single-cell RNA-seq preprocessing

Páll Melsted[1,8], A. Sina Booeshaghi[2,8], Lauren Liu[3], Fan Gao[4,5], Lambda Lu[4], Kyung Hoi (Joseph) Min[6], Eduardo da Veiga Beltrame[4], Kristján Eldjárn Hjörleifsson[3], Jase Gehring[7] and Lior Pachter[3,4]

We describe a workflow for preprocessing of single-cell RNA-sequencing data that balances efficiency and accuracy. Our workflow is based on the kallisto and bustools programs, and is near optimal in speed with a constant memory requirement providing scalability for arbitrarily large datasets. The workflow is modular, and we demonstrate its flexibility by showing how it can be used for RNA velocity analyses.

The quantification of transcript or gene abundances in individual cells from a single-cell RNA-sequencing (scRNA-seq) experiment is a task referred to as preprocessing[1]. The preprocessing steps for scRNA-seq bear some resemblance to those used for bulk RNA-seq[2] and are in principle straightforward: cDNA reads originating from transcripts must be partitioned into groups according to cells of origin and aligned to reference genomes or transcriptomes to determine molecules of origin, and the reads, which originate from PCR-duplicated molecules, must be 'collapsed' so that they are counted only once during quantification[3]. This is facilitated with unique molecular sequences that serve as barcodes [4]. The preprocessing scRNA-seq lie in [5] considered in determining choices for [6] ...ate the various steps. For example, [7] protocols, collapsing UMIs to [8] be performed naively by associating [9] same gene, with the same UMI, [10] computationally efficient procedure is [11] reads with identical UMIs that [12] PCR duplication of a single molecule [13] ...on can be relaxed, resulting in a [14] ...complete (that is, computationally [15]).

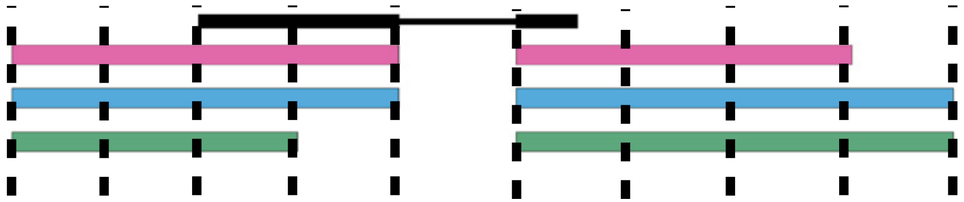
...q preprocessing is the amount of [16] ...ngle-cell experiment can generate [17] ...ls[6]. This is leading to bottlenecks [18] ...rent standard program for preprocessing [19] ...um scRNA-seq, the Cell Ranger [20] ...2 h to process 784 million reads[5] [21] ...mentary Table 1). For this reason [22] ...lows for scRNA-seq preprocessing [23] ing based on pseudoalignment[8] have recently been developed[8,9]. However, despite improvements in running time, current workflows have memory requirements that increase with data size[5], a situation that is untenable given the pace of improvement in technology and the corresponding increase in data volume.

In recent work, we introduced a format for scRNA-seq data that makes possible the development of efficient workflows by virtue of decoupling the computationally demanding step of associating reads to transcripts and genes (alignment) from the other steps required for scRNA-seq preprocessing[10]. This format, called BUS (barcode, UMI, set), can be produced by pseudoalignment, and rapidly manipulated by a suite of tools called bustools (Supplementary Table 2). To illustrate the utility, efficiency and flexibility of this approach for scRNA-seq preprocessing, we describe a Chromium preprocessing workflow based on reasoned choices for the key preprocessing steps. While we focus on Chromium, our workflow is general and can be used with other technologies. We show that our preprocessing workflow is faster and has lower memory requirements than existing methods, and we demonstrate the power of modular processing with the BUS format by developing a fast RNA velocity analysis workflow[11]. We also validate the design decisions underlying the Cell Ranger workflow. Our benchmarking and testing is comprehensive, comprising analysis of almost two dozen datasets and surpassing the scale of testing that has been performed for current workflows. Documentation and tutorials for the kallisto bustools workflow are available at http://pachterlab.github.io/kallistobustools.

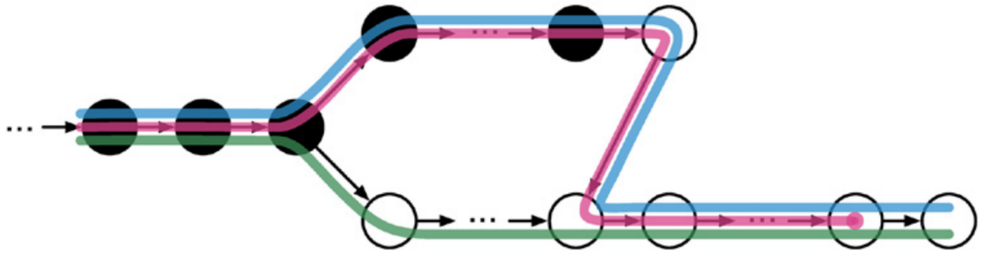
In designing an scRNA-seq preprocessing workflow, we began by investigating each required step: correction of barcodes, collapsing of UMIs and assignment of reads to genes. To achieve single-cell resolution, the Chromium technology produces barcode sequences that are used to associate cDNA reads to individual cells. We began by considering the efficiency–accuracy trade-offs involved in grouping reads with the same, or similar, barcodes to define the contents of individual cells. The Chromium barcodes arise from a 'whitelist', a set of pre-defined sequences that are included with the Cell Ranger software. Grouping reads by barcode is therefore straightforward, except for the fact that barcodes may contain sequencing errors. The Cell Ranger workflow corrects all barcodes that are one base-pair change away (Hamming distance 1) from barcodes in the whitelist. An examination of a benchmark panel of 20 datasets revealed that this error correction approach can be expected to rescue, on average, 0.8% of the reads in an experiment (Fig. 1a,b), a calculation based on an inferred error rate per base for each dataset (Methods and Supplementary Table 3). Thus, correction

Kallisto pseudoalignment

From lecture 2

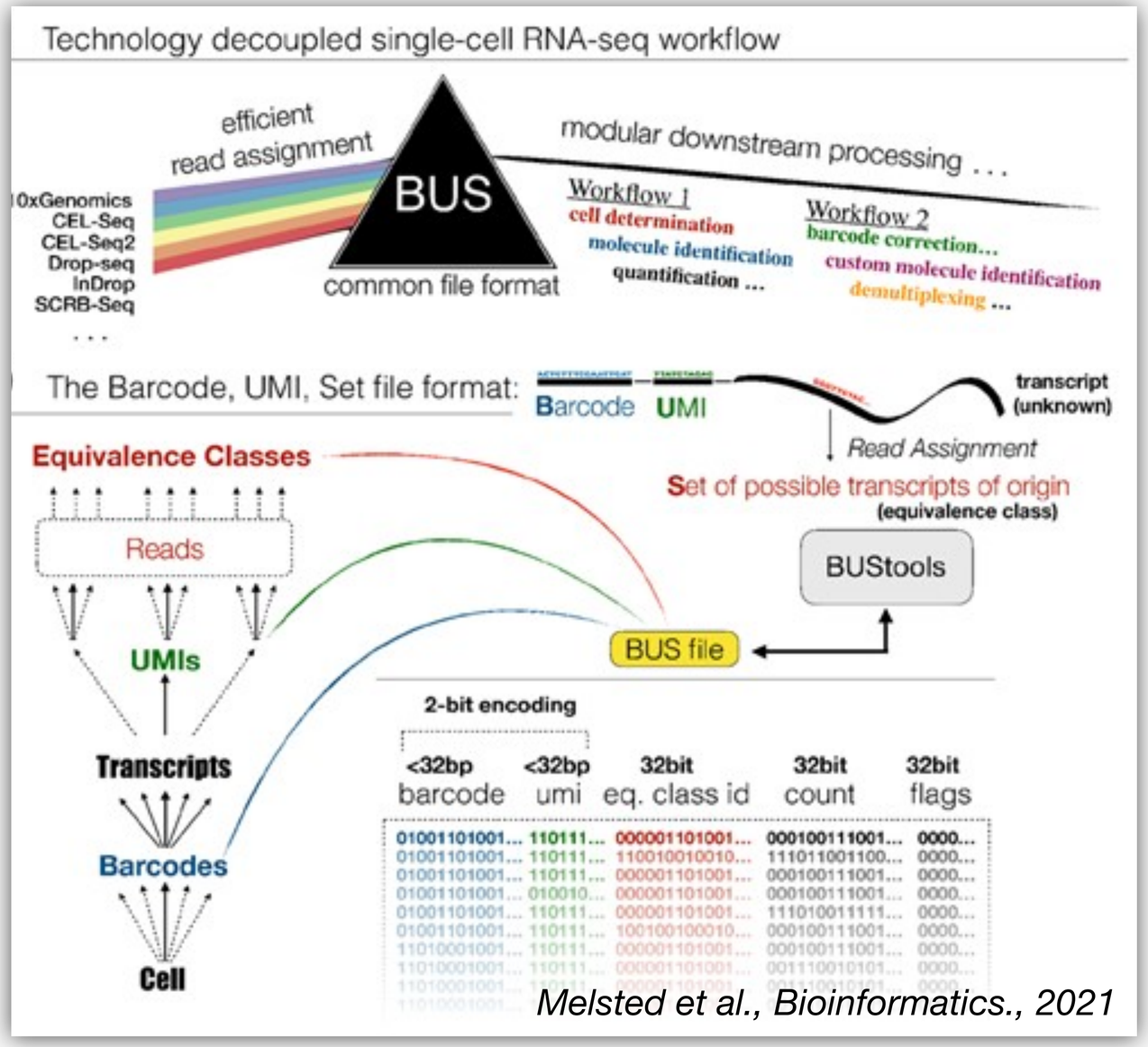


Don’t align, instead find transcripts that are compatible with the read



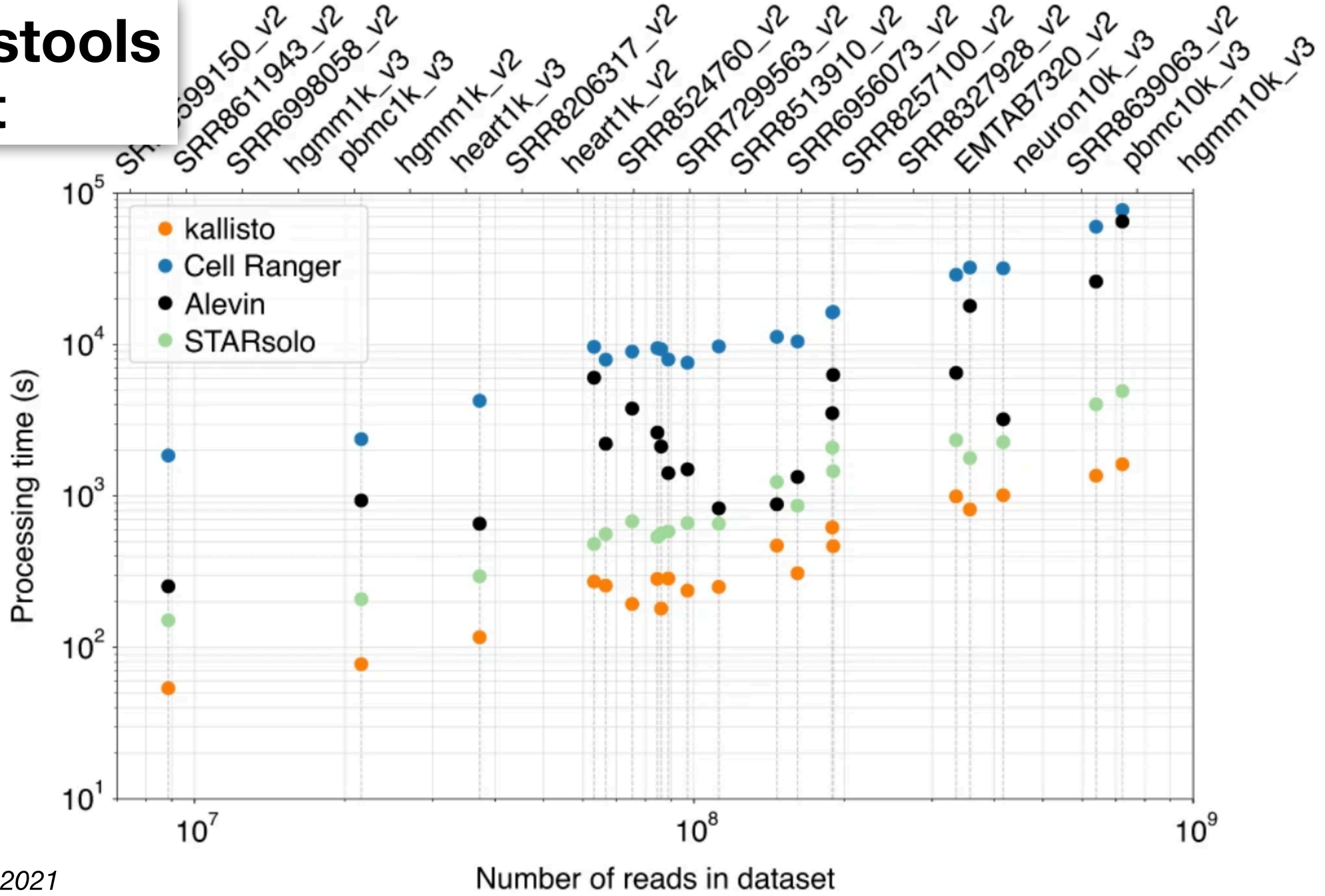
Bray et al., Nat. Biotech., 2016

Barcode, UMI, Set (BUS)



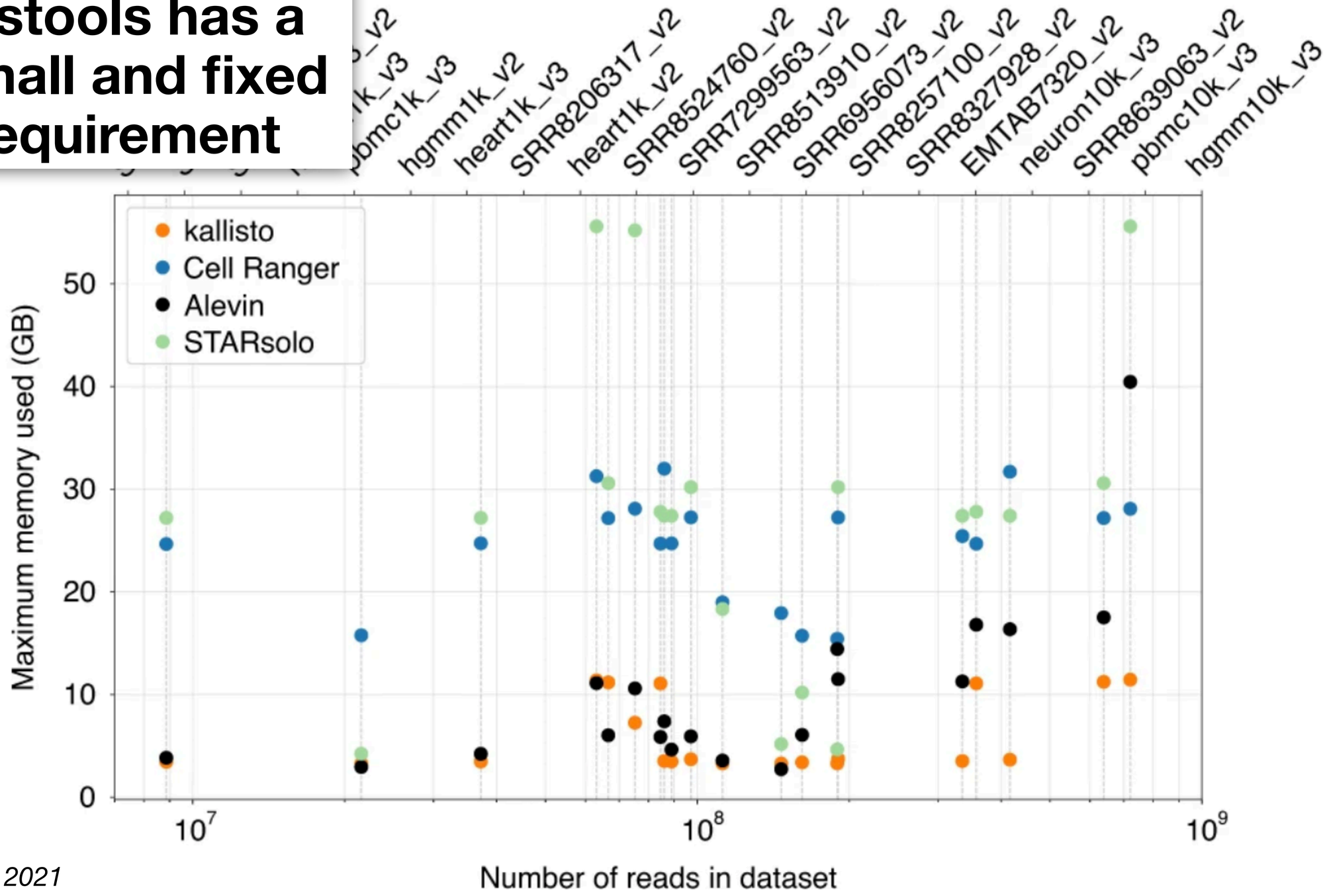
Melsted et al., Bioinformatics., 2021

Kallisto Bustools is fast



Melsted et al., Nat. Biotech., 2021

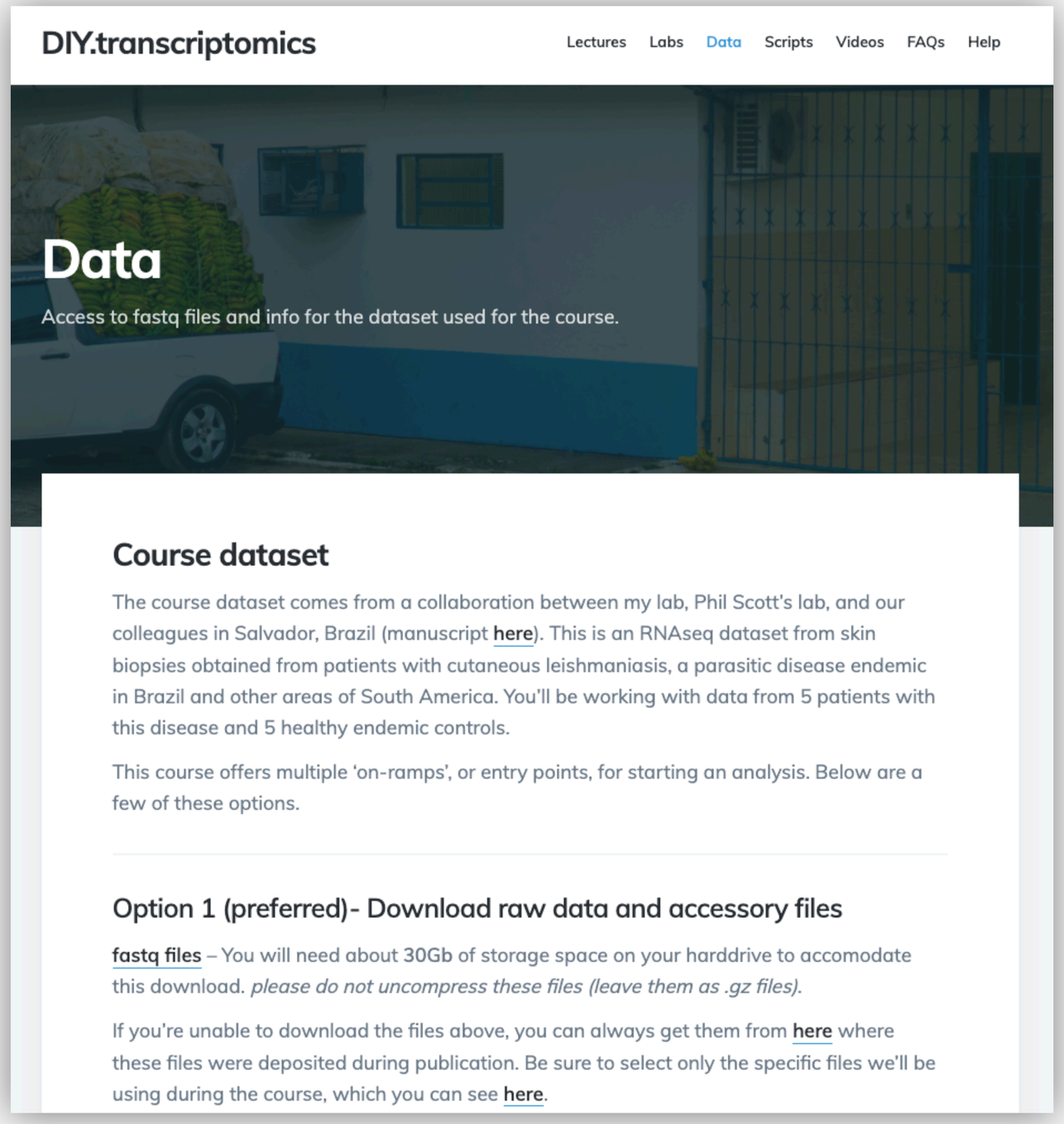
Kallisto Bustools has a relatively small and fixed memory requirement

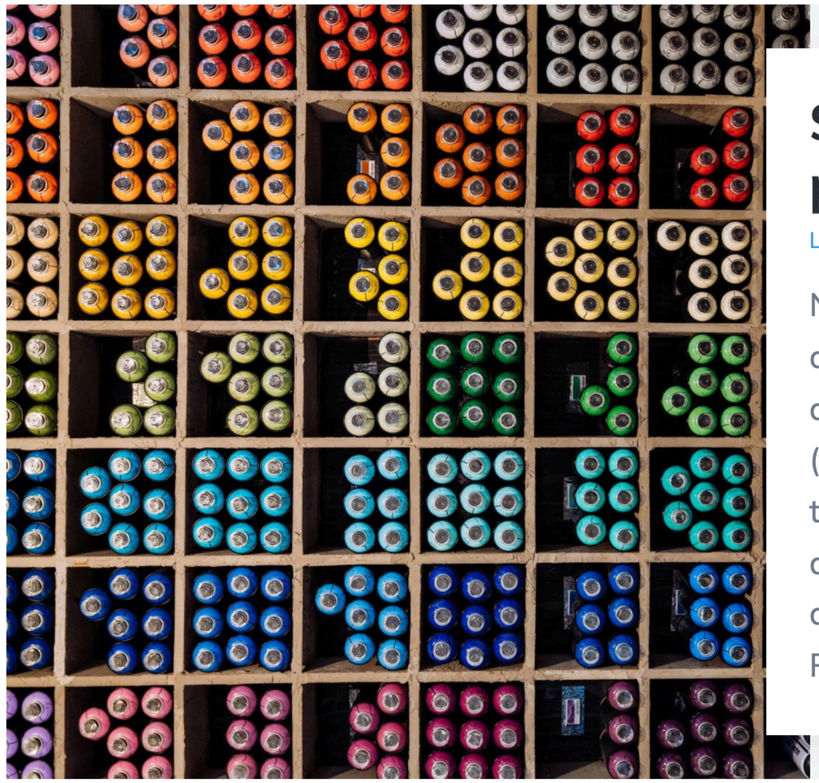


Melsted et al., Nat. Biotech., 2021

Let’s get started using Kallisto bustools to preprocess raw scRNA-seq data directly on our laptop

Say goodbye to our course dataset





Single cell RNA-seq principles and proc

Lecture 13 • watch by December 1, 2021

Now that you're comfortable w
data analysis, we'll shift our fo
developing landscape of singl
(scRNA-seq). In this lecture, yo
the underlying technology and
demonstrate how to process r
data directly on your laptop (!)
R/bioconductor.

Single cell RNA-seq – principles and processing

Lecture 13 • watch by December 1, 2021

Overview

Now that you're comfortable with bulk RNA-seq data analysis, we'll shift our focus to the rapidly developing landscape of single cell RNA-seq (scRNA-seq). In this lecture, you'll learn about the underlying technology and we'll demonstrate how to process raw single cell data directly on your laptop (!) for importing into R/bioconductor.

Learning objectives

- Understand droplet-based scRNA-seq technology
- Be able to compare and contrast single cell and bulk RNA-seq methods
- Understand cost and experimental design considerations for scRNA-seq experiments.
- Familiarity with multiplexed single cell assays (CITE-seq, 'multiome', TEA-seq)
- Be able to define common terms and concepts in single cell genomics
- Use Kallisto-BUStools to preprocess raw scRNA-seq data (via **kb-python**)

What you need to do

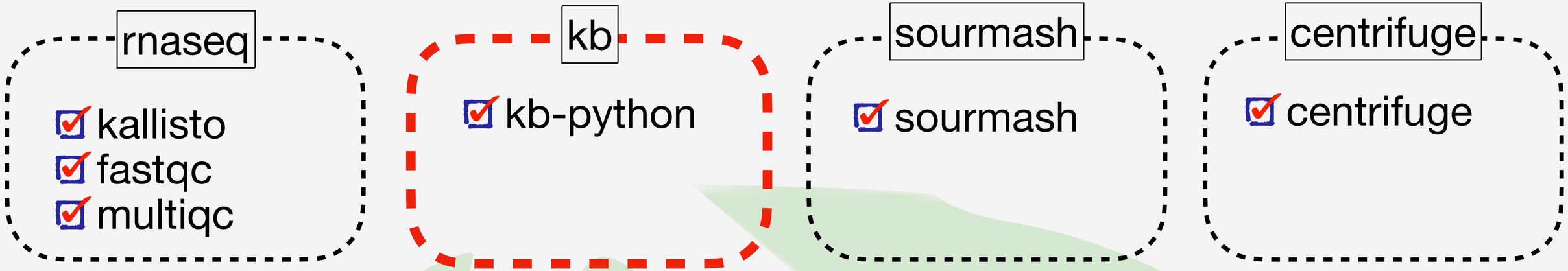
Download raw files. You will need about **5Gb** of storage space on your harddrive to accomodate this download. *please do not uncompress these files (leave them as .gz files).* This is data from 1000 peripheral blood mononuclear cells (PBMCs) and is one of the sample datasets provided by 10X Genomics here. I merged the separate lane files to make this simpler to work with for the course.

Human transcriptome reference index file - this is the index you created using Kallisto way back in lecture 2. If you don't have this, remember it's easy to create using `kallisto index`.

t2g.txt - this is a human transcript-to-gene mapping file that we will use with Kallisto-Bustools to preprocess our data. This file is easy to generate with `kb ref`, but downloading it now will save you some time.

kb-python - You will need to have this software installed in a Conda environment on your laptop. We did this way back in lecture 1. If you are unable to install or use kb-python, just follow along with the lecture so you understand the concepts.

Our Conda software environments



First, activate the environment with `conda activate kb`

Then, call the program with `kb`

protocols.hostmicrobe.org/conda

Kallisto-Bustools: supported technologies

kb --list

```
List of supported single-cell technologies

Positions syntax: `input file index, start position, end position`
When start & end positions are None, refers to the entire file
Custom technologies may be defined by providing a kallisto-supported technology string
(see https://pachterlab.github.io/kallisto/manual)

name         whitelist   barcode                 umi         cDNA
---------    ---------   --------------------    -------     ----------------------
10XV1        yes         0,0,14                  1,0,10      2,None,None
10XV2        yes         0,0,16                  0,16,26     1,None,None
10XV3        yes         0,0,16                  0,16,28     1,None,None
CELSEQ                   0,0,8                   0,8,12      1,None,None
CELSEQ2                  0,6,12                  0,0,6       1,None,None
DROPSEQ                  0,0,12                  0,12,20     1,None,None
INDROPSV1                0,0,11 0,30,38          0,42,48     1,None,None
INDROPSV2                1,0,11 1,30,38          1,42,48     0,None,None
INDROPSV3    yes         0,0,8 1,0,8             1,8,14      2,None,None
SCRUBSEQ                 0,0,6                   0,6,16      1,None,None
SURECELL                 0,0,6 0,21,27 0,42,48   0,51,59     1,None,None
SMARTSEQ                                                     0,None,None 1,None,None
```

Preparing reference files scRNA-seq

Calling the program and function

Indicates code continues on next line

Download a prebuilt index (human or mouse)

Path/name of index file to be created

Path/name of transcript-to-gene mapping file to be created

```
kb ref \
-d human \
-i Homo_sapiens.GRCh38.cdna.all.index \
-g t2g.txt
```

You do not need to run this code!
I provide the t2g.txt file in the downloads for this lecture

Preprocessing scRNA-seq data