Frederick National Laboratory
for Cancer Research
sponsored by the National Cancer Institute

Single Cell RNA-Seq Workshop – Day1

NCI CCR BTEP

2019-10-03

Single Cell RNA-seq Workshop - Introduction

Frederick National Laboratory for Cancer Research

Day 1 Training Topics

- ❑ Introduction to single cell technologies and applications - Mike Kelly
- ❑ Overview of single cell transcriptomic analysis workflow and software pipelines - Yongmei Zhao
- ❑ scRNA-seq preprocessing and quality control - Vicky Chen and Nathan Wong
- ❑ Feature selection, dimensionality reduction, clustering, marker gene identification, and visualization - Cihan Oguz
- ❑ Single cell RNA-seq cell type classification and annotation - Keyur Talsania

Day 2 Training Topics

- ❑ Trajectory analysis - Abdalla Abdelmaksoud
- ❑ Single cell integration analysis from multiple technologies - Nathan Wong, Cihan Oguz, Vicky Chen, Keyur Talsania
- ❑ Panel Discussion - All workshop speakers

Frederick National Laboratory for Cancer Research

Introduction to Single Cell Technologies and Applications

Mike Kelly
Manager, Single Cell Analysis Facility

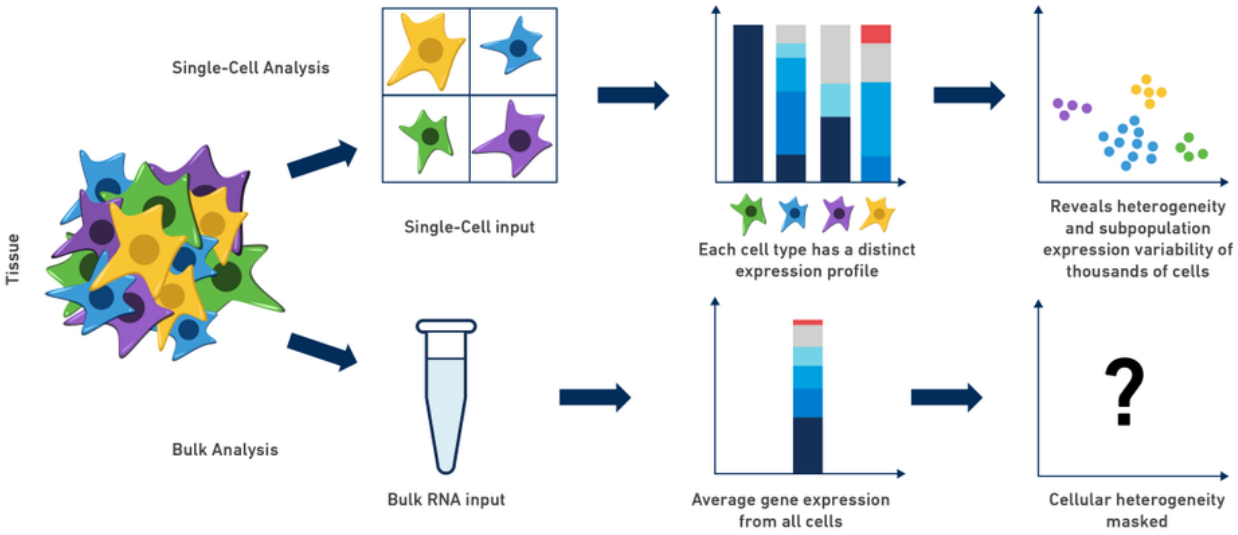
Section Outline

- Why single cell RNA-Seq?
- General method of generating single cell RNA-Seq libraries
- Sample prep considerations for better single cell data
- scRNA-Seq (and now with additional modalities)
- Single cell ATAC-Seq and CNV
- Typical project workflow and how to get support

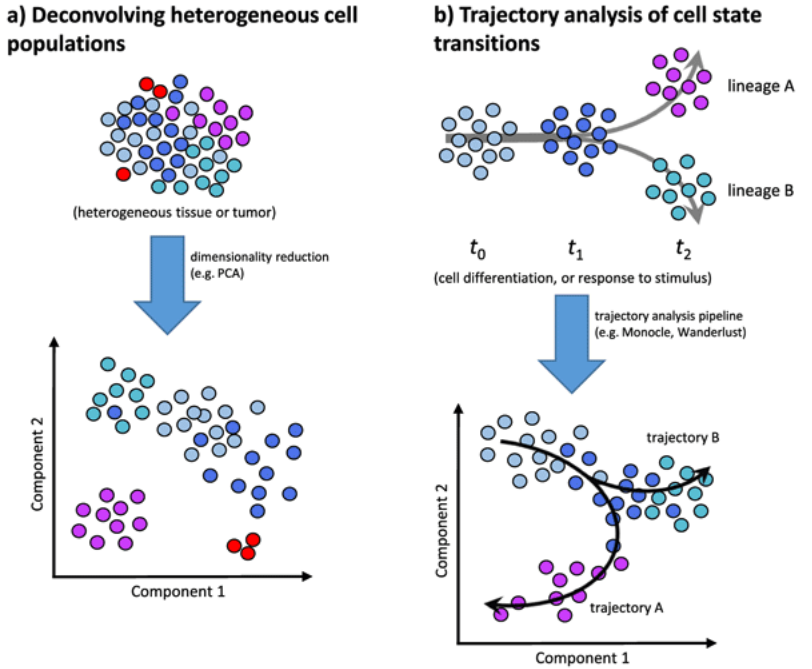
Why single cell RNA-Seq?

Single cell sequencing avoids caveat of bulk averaging & allows inference of dynamic processes

Frederick National Laboratory for Cancer Research



From 10x Genomics Community Website

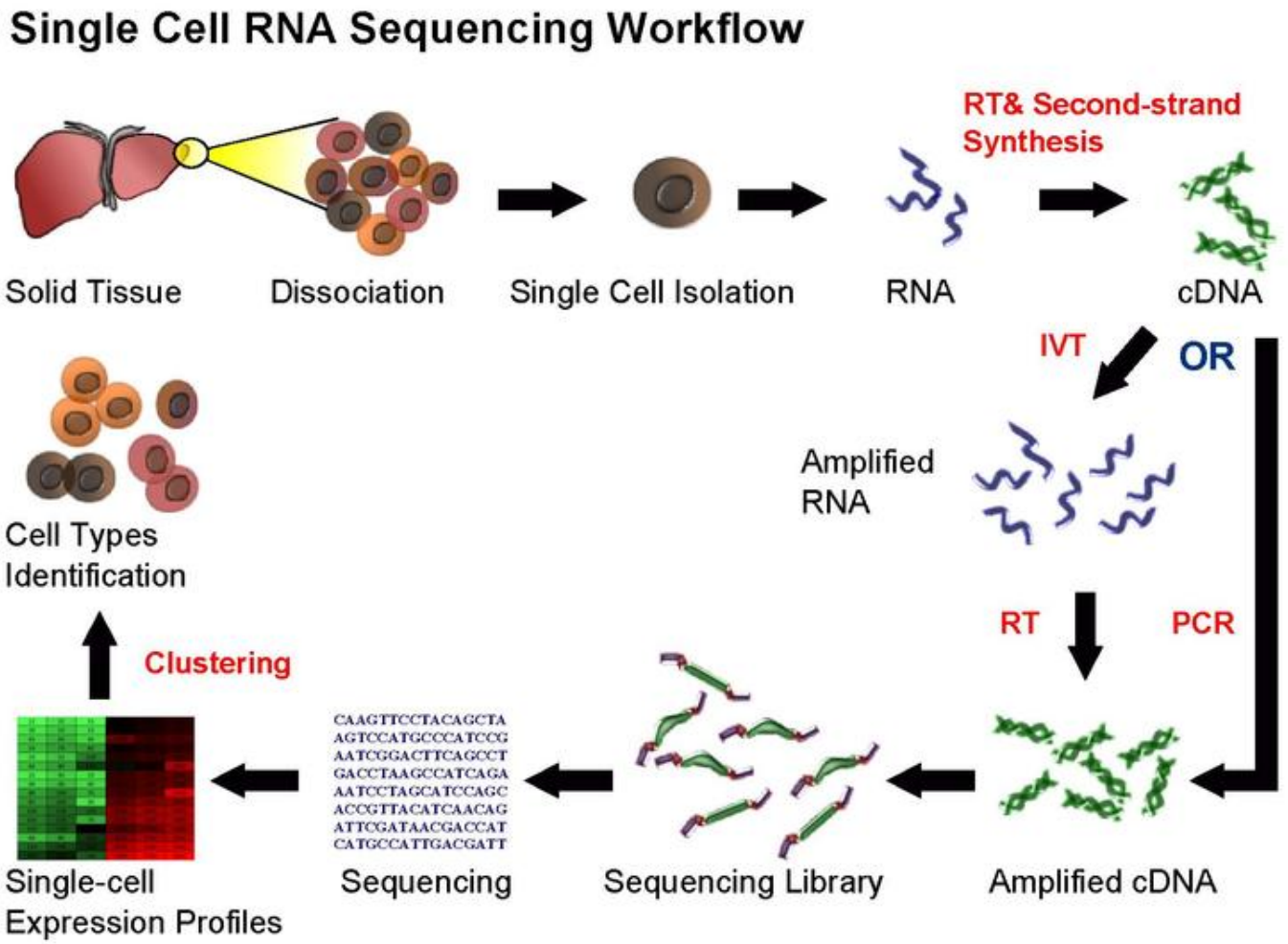


Liu & Trapnell Review – F1000Research 2016

- Survey cells present in a biological system and identify gene signatures associated with cell types
- Compare cell number and/or phenotypic differences between conditions (i.e. healthy vs disease)
- Model dynamic changes representing biological processes
- Determine clonal evolution of cells within a composite tissue / system
- ***Interrogate potential mechanisms at cellular resolution in health and disease***

General method of generating single cell RNA-Seq libraries

Generalized workflow of generating single cell RNA-Seq data

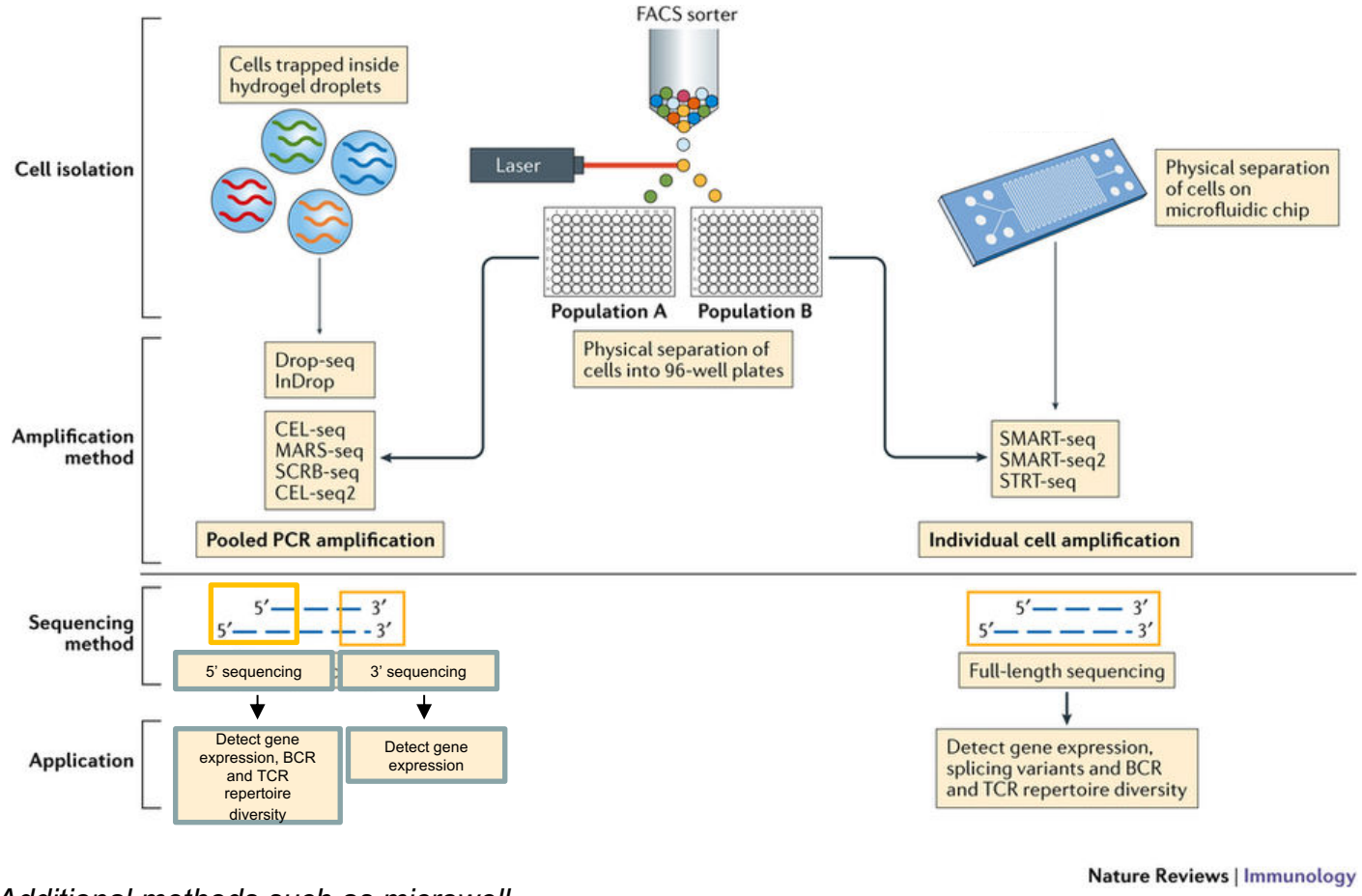


https://hemberg-lab.github.io/scRNA.seq.course/

- Partition single cells
- Convert mRNA into cDNA
- Amplify cDNA
- Generate sequencing library
- Sequence
- Data analysis with identification of what transcripts are expressed by each cell profiled

Single cell barcodes introduced at different step, depending on method

Multiple methods & platforms for performing single cell RNA-Seq



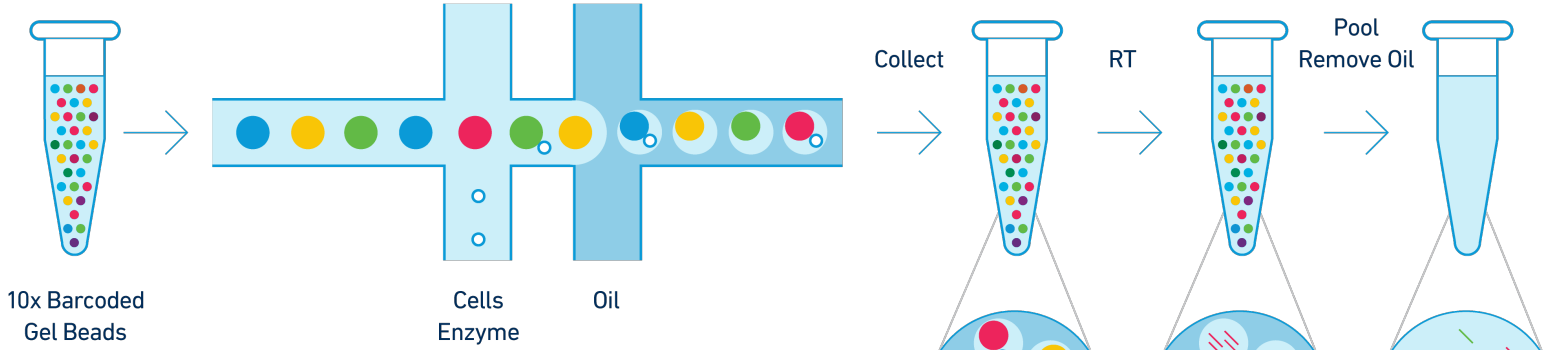
Additional methods such as microwell, combinatorial indexing, not shown

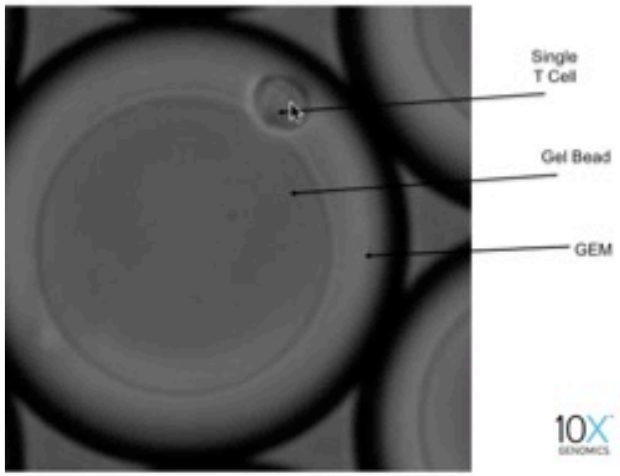
Modified from Papalexi & Satija (2018)

- Single cell per well protocols (middle and right side of diagram) allow profiling of “full-length” mRNA transcripts and generally have higher sensitivity
- Droplet-based approaches having higher throughput and provide “end-counting” of gene-level detection of expression
- Different methods vary considerably in cost, ease of implementation, and requirement for specialized equipment
- Analysis methods depend somewhat on the types of data being generated

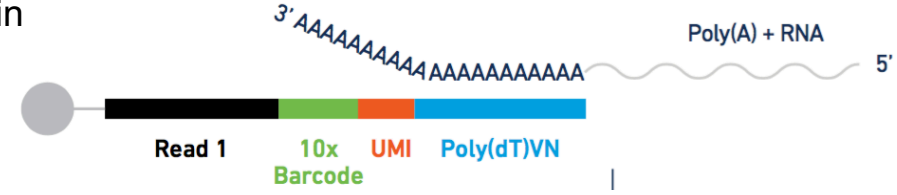
Droplet-based single cell sequencing allows high-throughput profiling with unprecedented ease – *the current standard*

Frederick National Laboratory for Cancer Research





- Cells partitioned with barcoded gel bead in Poisson loading (50-65% input cells productive)
- Molecules barcoded by in-droplet reactions
- Single cell barcoded molecules then handled in bulk for library prep and sequencing
- Barcoding method for scRNA-Seq shown



Sample prep considerations for better single cell data

scRNA-Seq capture input of a high viability single cell suspension is important for good data

Frederick National Laboratory for Cancer Research

Single Cell Suspension for Optimal Performance

10x Genomics® Single Cell Protocols require a suspension of viable single cells as input. Minimizing the presence of cellular aggregates, dead cells, non-cellular nucleic acids and potential inhibitors of reverse transcription is critical to obtaining high quality data.

1.3. Factors Influencing Cell Recovery

To recover the expected number of cells, it is critical to maximize viability, minimize the cell preparation time, accurately measure the input cell concentration and pipette the correct volume into each reaction.

Ideally, input cell suspensions should contain more than 90% viable cells. Non-viable and dying cells generally contain less and more fragmented RNA that may not be efficiently captured by 10x Genomics Single Cell Solutions. The presence of a high fraction of non-viable cells in the input suspension may therefore decrease the apparent efficiency of cell partitioning and recovery.

10x Genomics Cell Preparation Guide

- **10X Genomics recommends loading 90% viable cells or higher**
- **Often the underlying causes of below target # of cells**
- **Can contribute a “background signal” – ambient RNA**

Alternative sample prep methods when viable single cell dissociations are not practical

- **Single nuclei preparation**
  - Fast extraction of nuclei from solid tissue; little dissociation-driven artifact
  - Less RNA content than whole cell; higher pre-mRNA ratio
  - Compatible with frozen tissue or difficult to dissociate tissue
  - More difficult to QC sample; results assessed after sequencing

Matson et al 2018 JoVE

- **Transcriptional inhibition / cold-active proteases**
  - Perform dissociation in transcriptionally-slowed environment
  - Reduces dissociation-driven transcriptional artifact
  - Additional control of dissociation process required

Wu et al 2017 Neuron

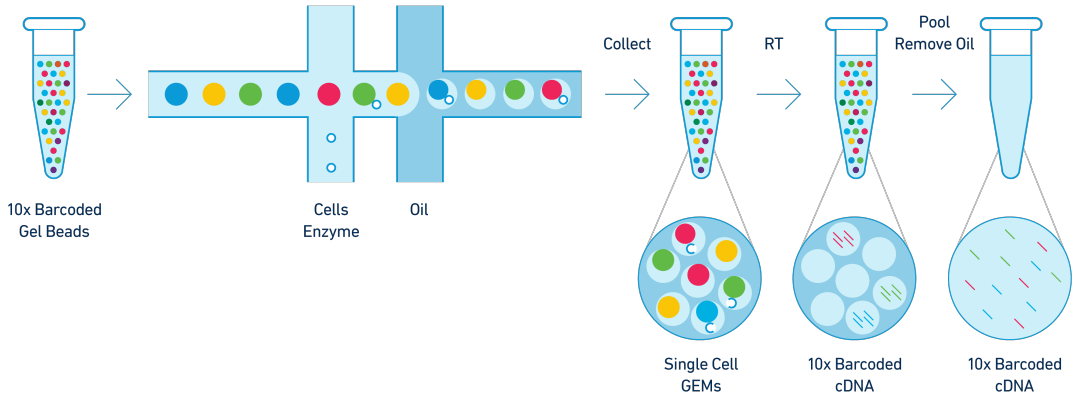
- **Cell fixation with Methanol**
  - Dissociated cells still needed as input, but allows ‘batching’ of samples
  - Preserves cells and transcript content for cold storage
  - May not work for all cell types;

Chen et al 2018 J Transl Med

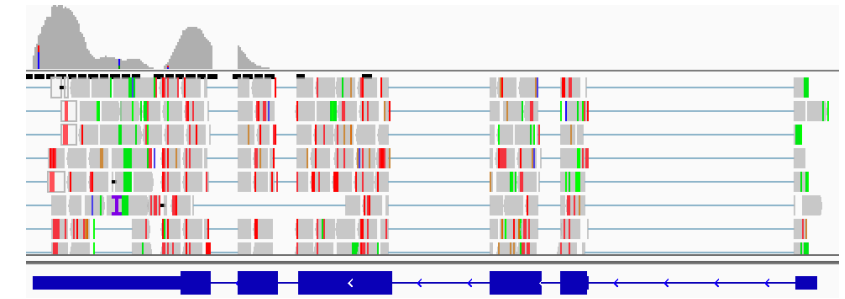
scRNA-Seq (and now with additional modalities)

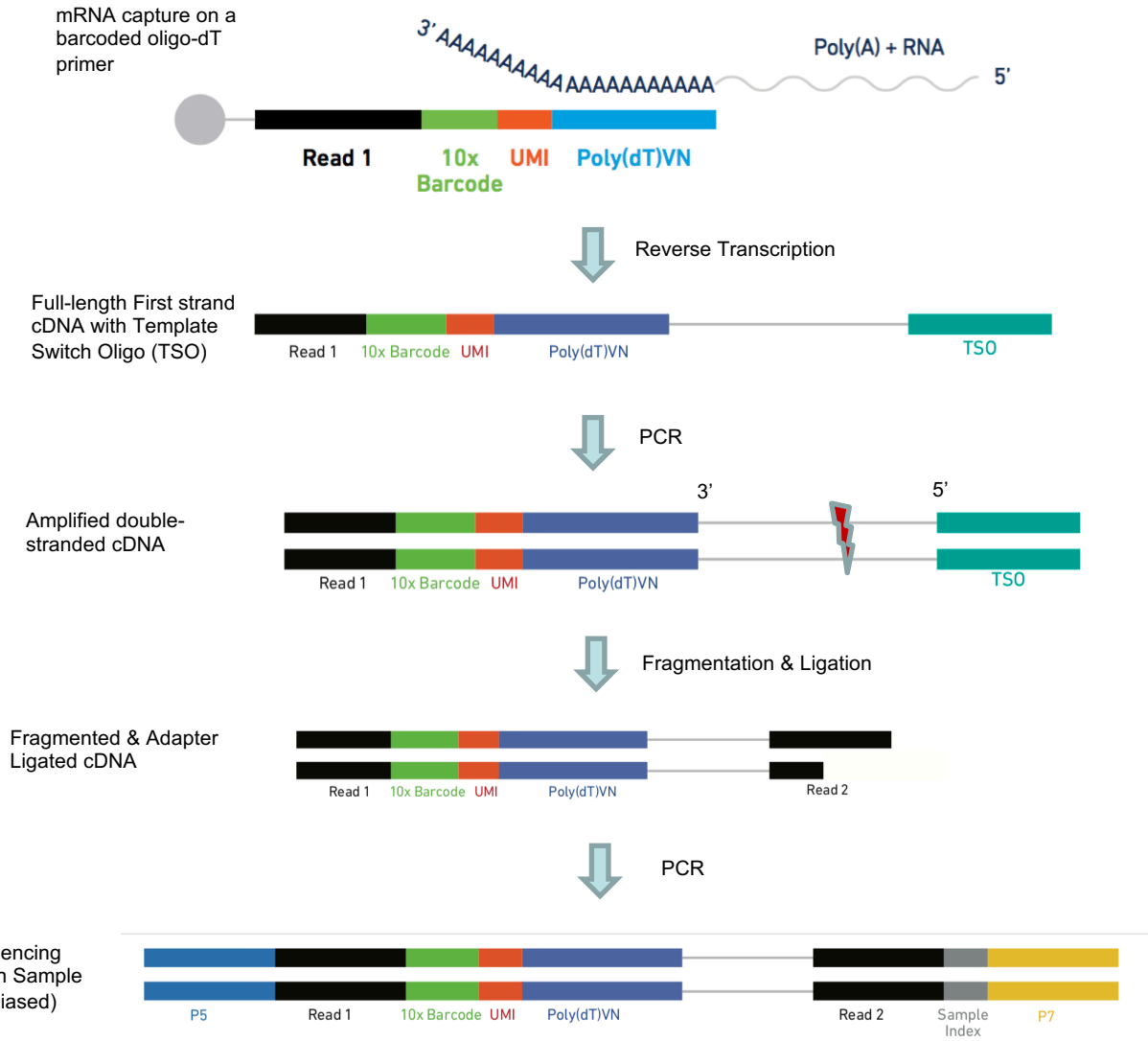
How it's made: the 10x Genomics single cell 3' gene expression library

Frederick National Laboratory for Cancer Research



- Cell barcodes on 3' end of cDNA
- Library prep enriches for 3' end
- Mapped reads are 3' biased

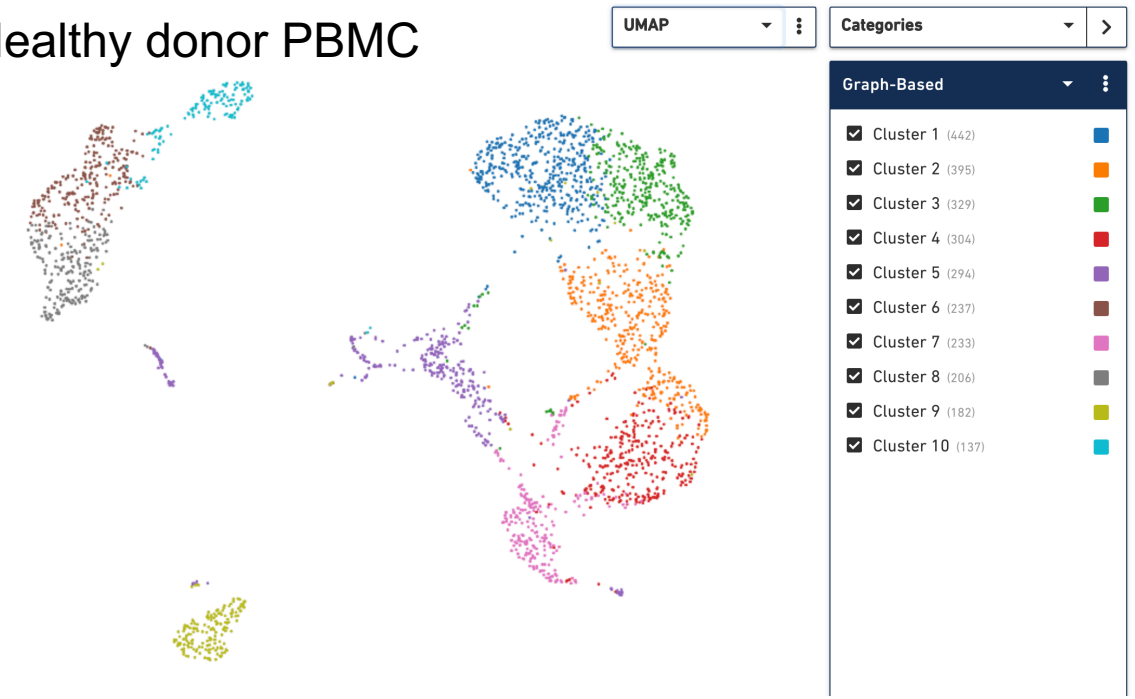


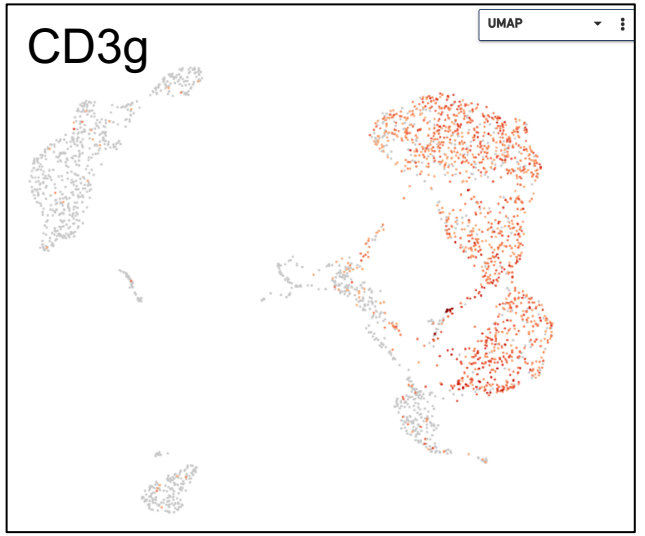


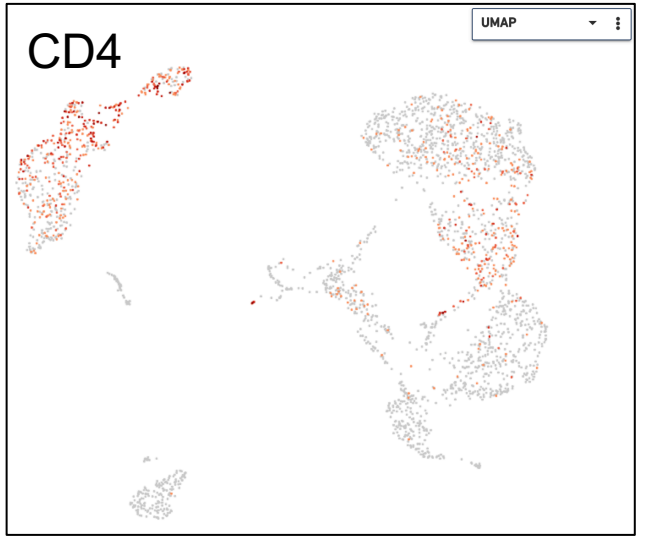
Single cell gene end-counting RNA-Seq provides a rich dataset for many analysis paths

Frederick National Laboratory for Cancer Research

Healthy donor PBMC



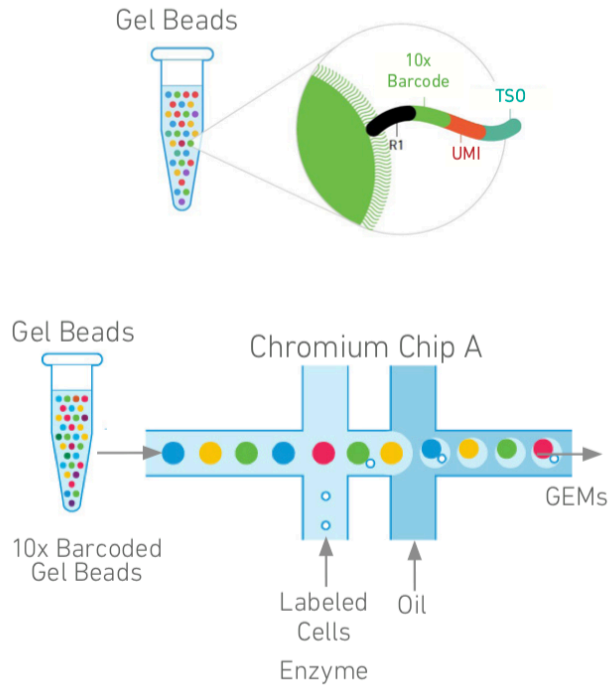


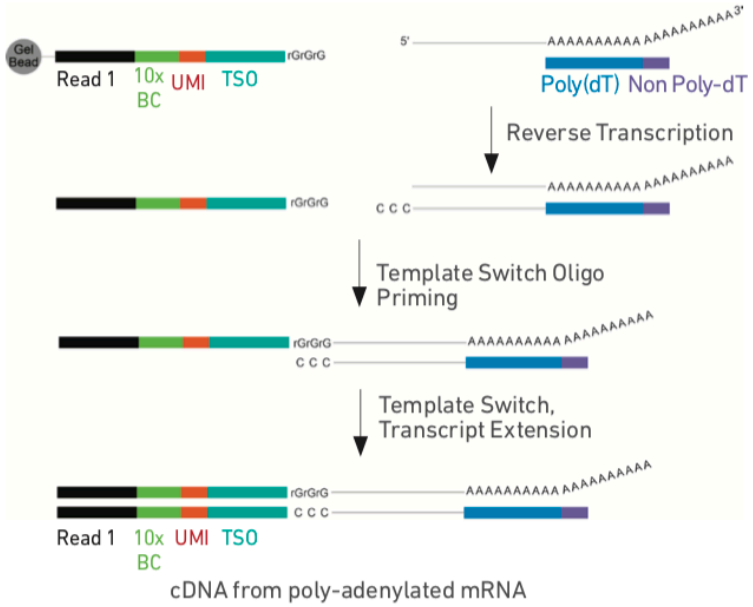


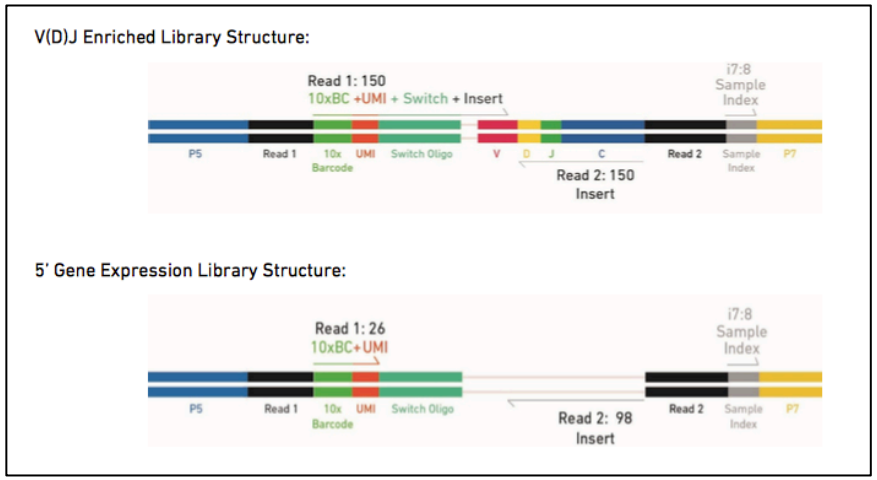
- UMI-corrected gene level counts allows gene expression survey of every cell
- Cells clustered based on expression, 2D projection shows relative transcriptional similarities
- Can query dataset for expression of genes, run differential expression, etc.
- Data is ‘sparse’; some favorite genes may be difficult to detect

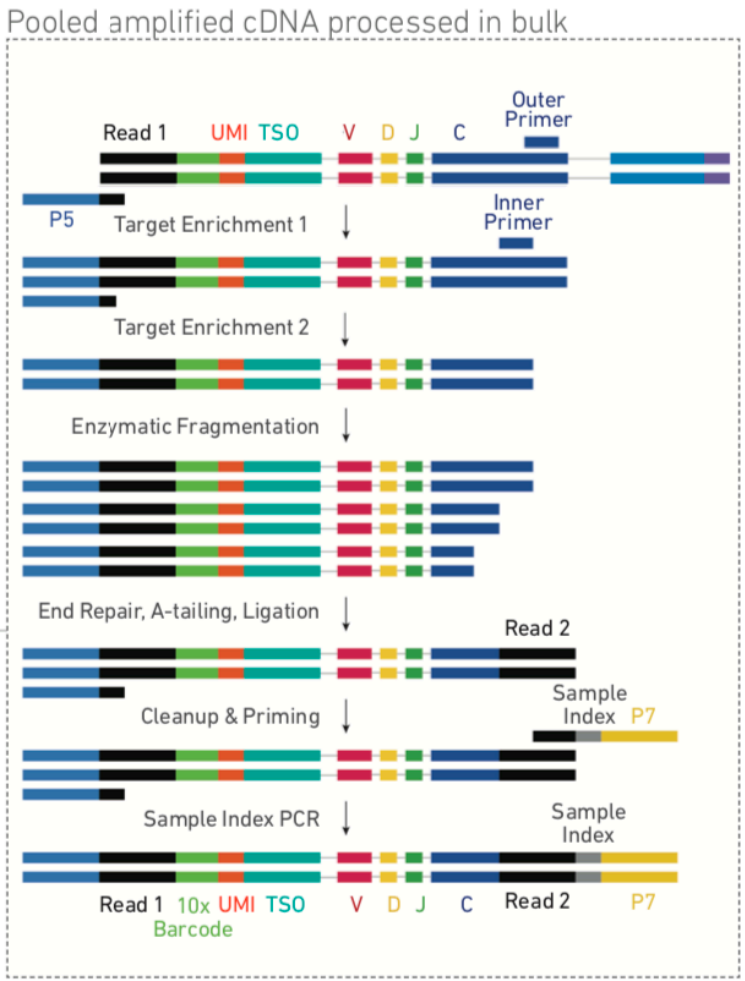
5’ end barcoding single cell allows gene expression and VDJ (T-cell and B-cell receptor) sequencing from the same cells

Frederick National Laboratory for Cancer Research



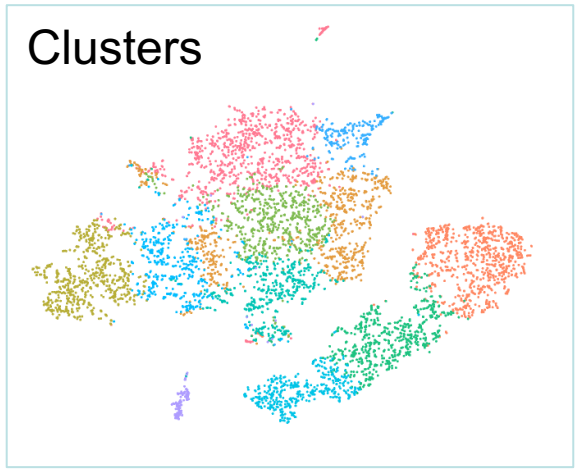


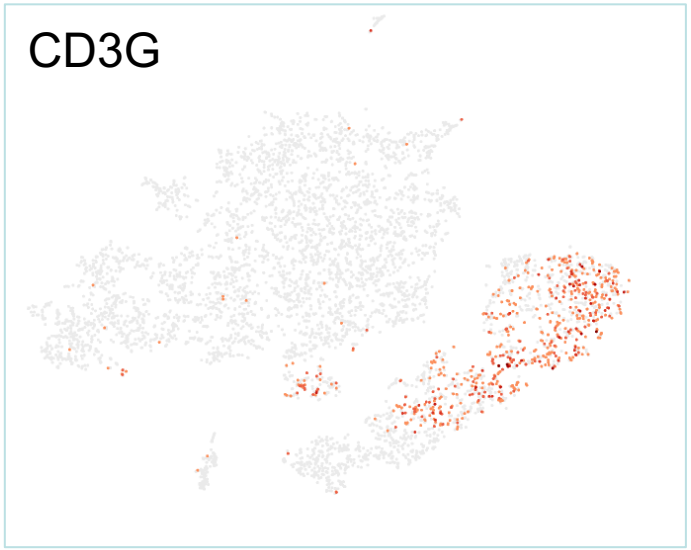


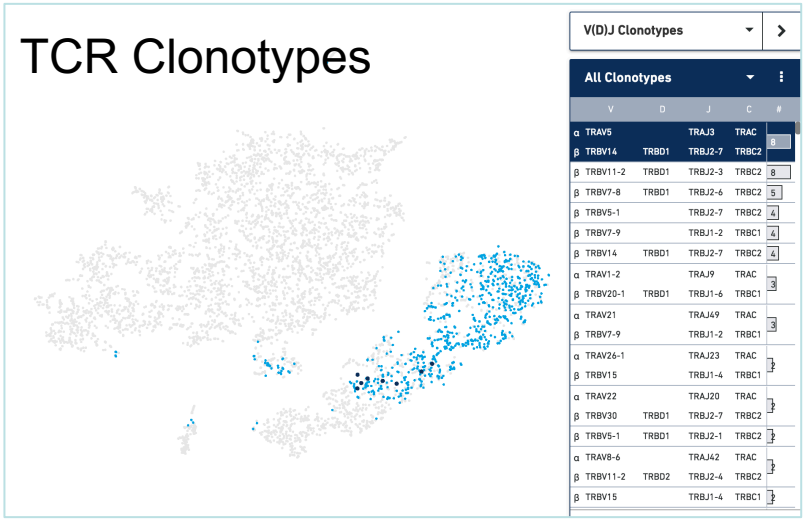


Specific target genes can be amplified from the single cell cDNA library and sequenced – linked by cell barcodes

Frederick National Laboratory for Cancer Research





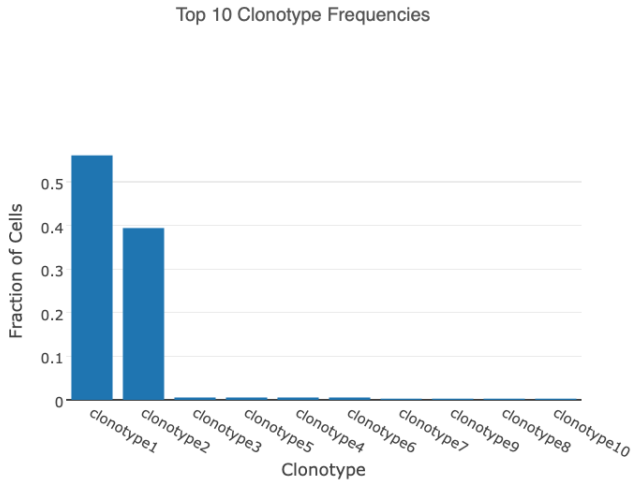


Applications:

-Determine genes associated with an expanded clone (malignancy)

-Determine phenotype of cells activated and responding to target (tumor killing T-cells)

-Determine the TCR of an expanded T-cell clone effective at responding to a target



Notes:

-Higher chance of getting full VDJ from high viability samples

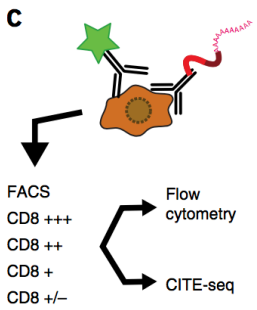
-VDJ sequencing requires longer sequencing reads = more expensive

–Analysis of a diverse repertoire takes some additional work

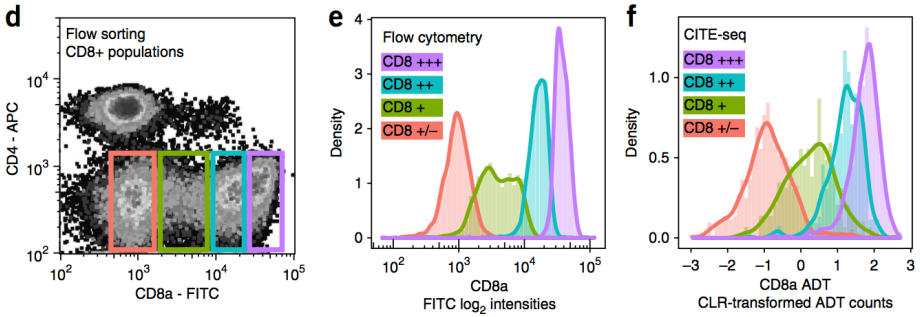
Feature barcoding for cell surface protein measurement in parallel to gene expression profiling

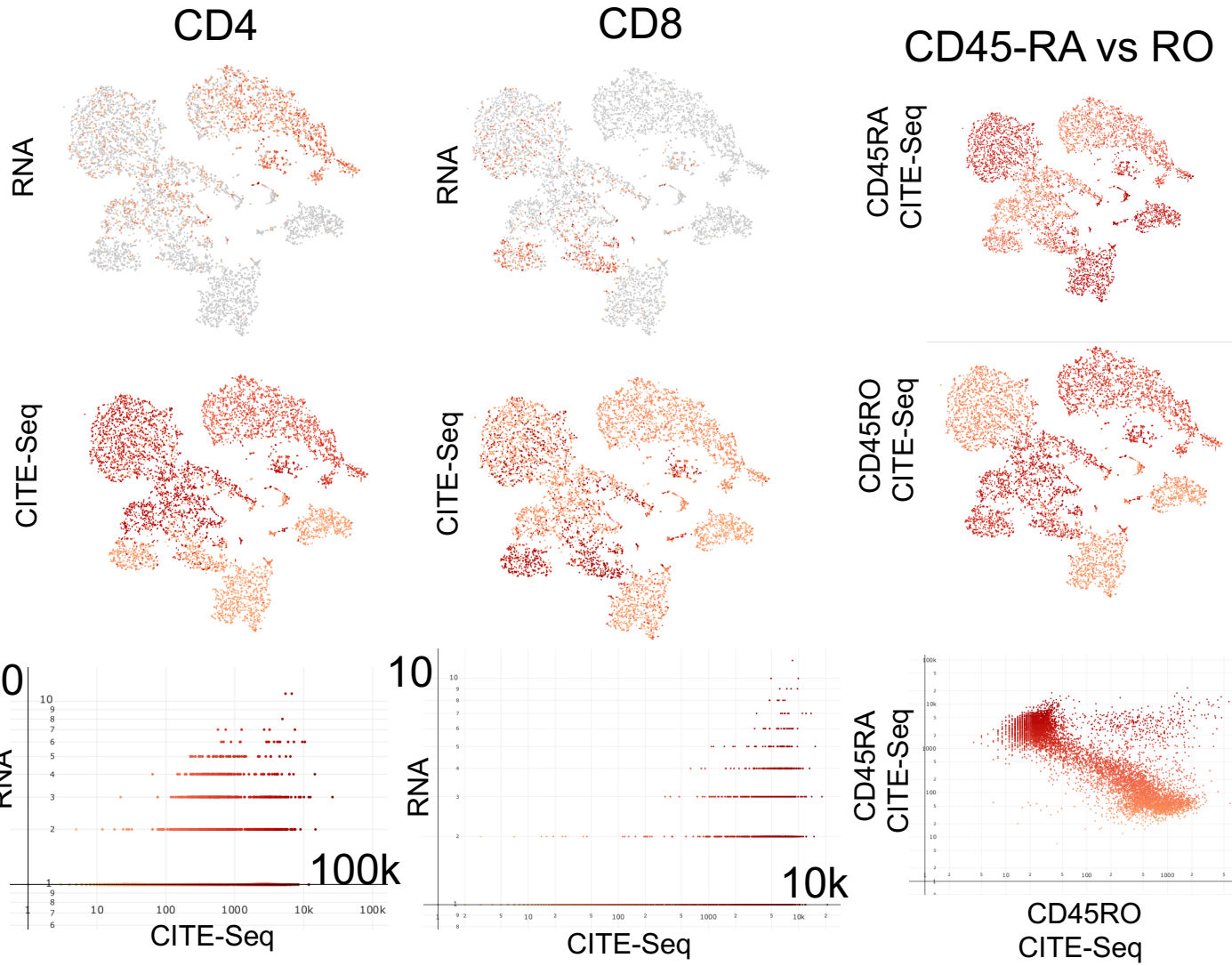
Frederick National Laboratory for Cancer Research

'CITE-Seq' = 'AbSeq' =
'Antibody Feature Barcoding'



Poly-adenylated antibody (or other feature label) labels cells of interest – similar to FACS

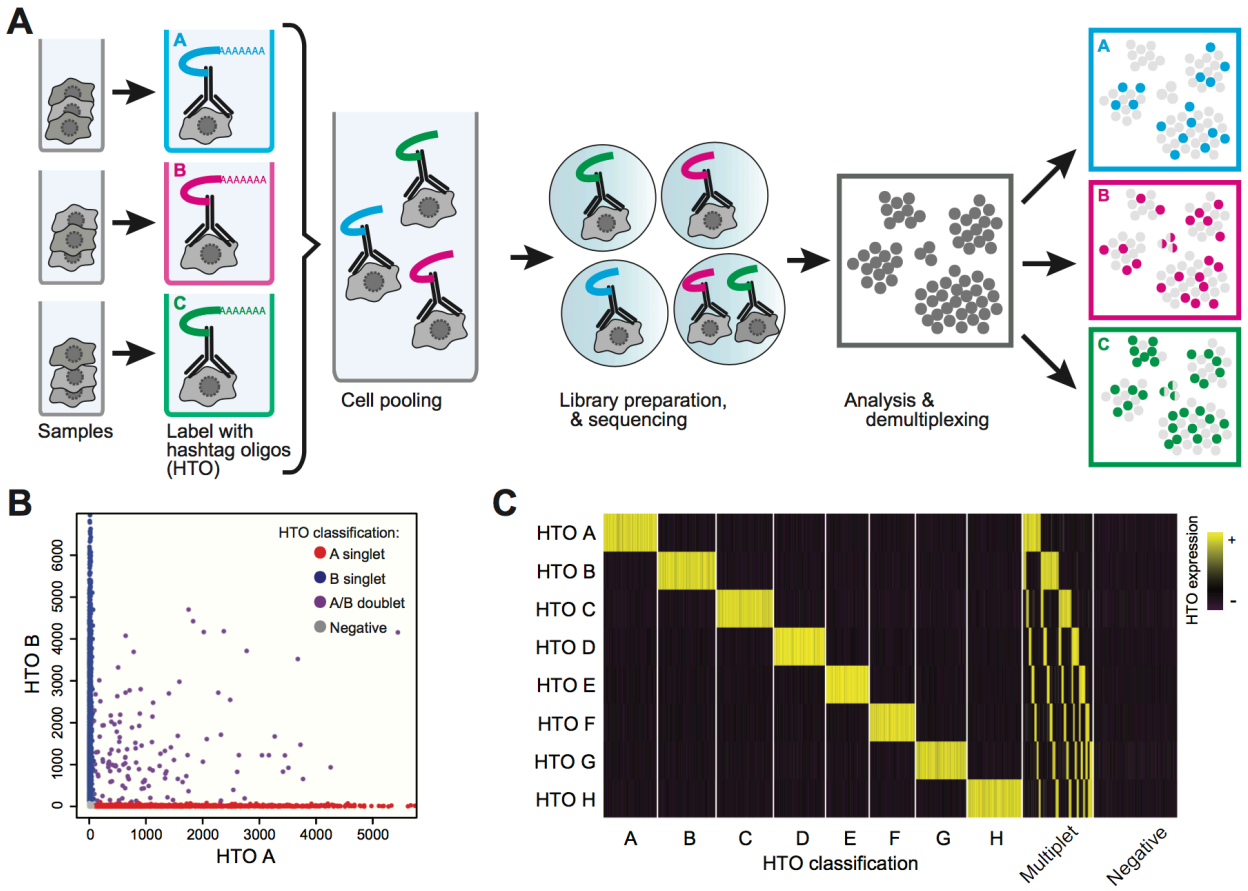




10x Genomics Example Dataset (10k PBMC 3' v3 with Feature Barcodes)

Antibody-based cell labeling allows sample multiplexing with 'superloading' capability

Frederick National Laboratory for Cancer Research



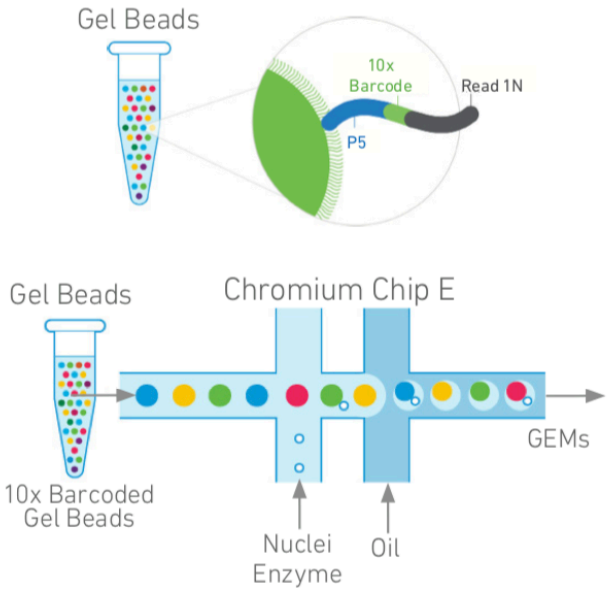
- Allows multiple samples to be combined and run on single capture lane
- Can more cost-effectively increase number of identifiable biological replicates
- Reduce technical batch effects
- Improve detection of doublets

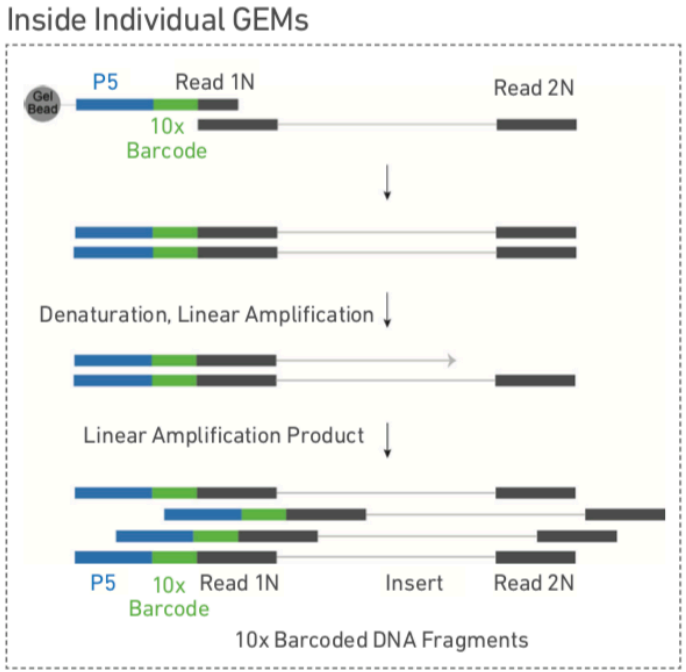
Target # Cells Recovered	Single-plex (standard) NI multiplet rate	2-sample multiplex NI multiplet rate	4-sample multiplex NI multiplet rate	6-sample multiplex NI multiplet rate
5,000	~3.9%	~2.1%	~1.0%	~0.7%
10,000	~7.6%	~4.2%	~2.2%	~1.5%
20,000	~16.1%	~8.9%	~4.7%	~3.2%

- Sample multiplexing sometimes fails (consider for precious samples)
- Different design for single nuclei samples

Single cell ATAC-Seq and CNV

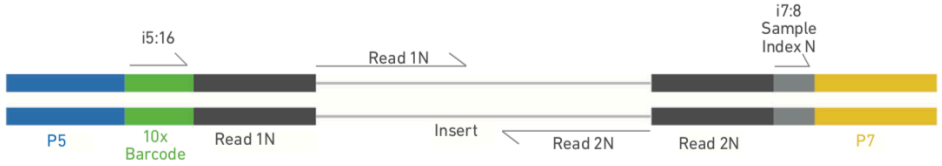
Single Cell ATAC-seq Workflow & Library Format





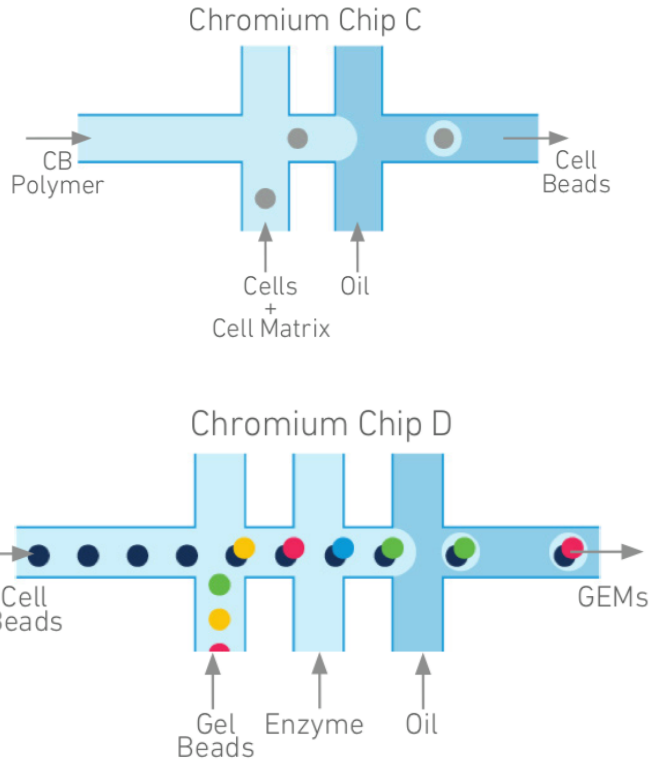
- Requires isolated nuclei (reduces mitochondrial fragment contamination)
- Nuclei incubation with Tn5 transposition mix
- Transposed nuclei captured with barcoded gel beads
- Sequencing library prepared

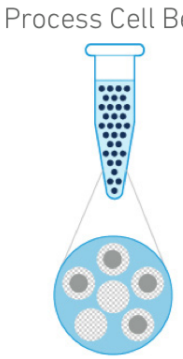
scATAC Sequencing Library

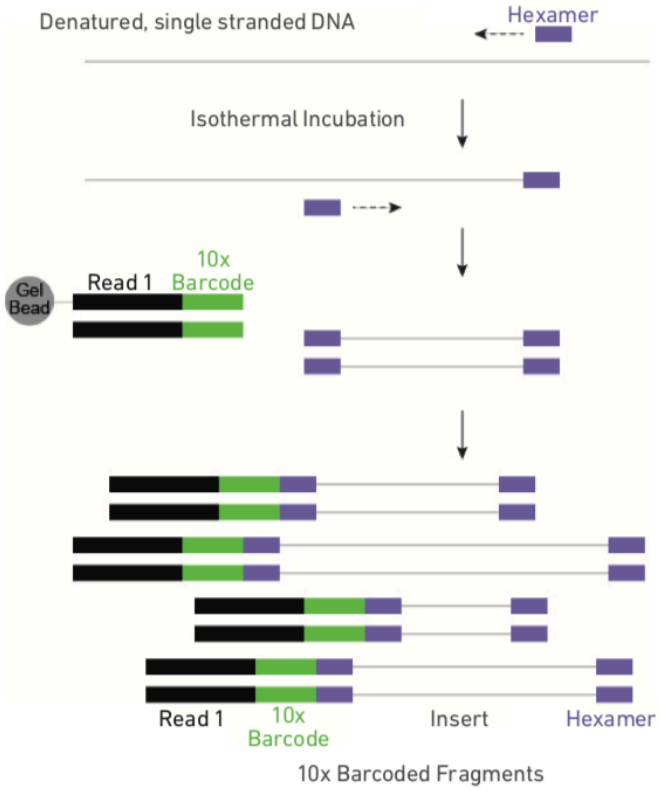


Single Cell DNA (for CNV) Workflow & Library Format

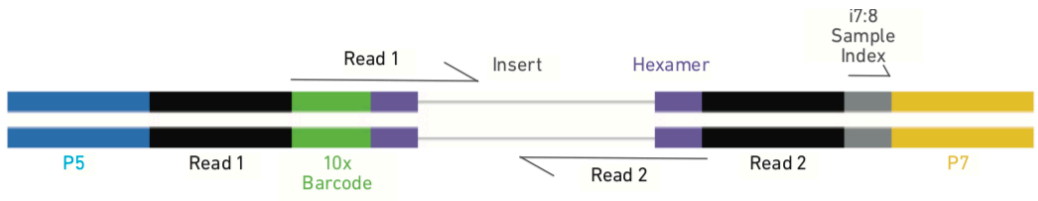
Frederick National Laboratory for Cancer Research







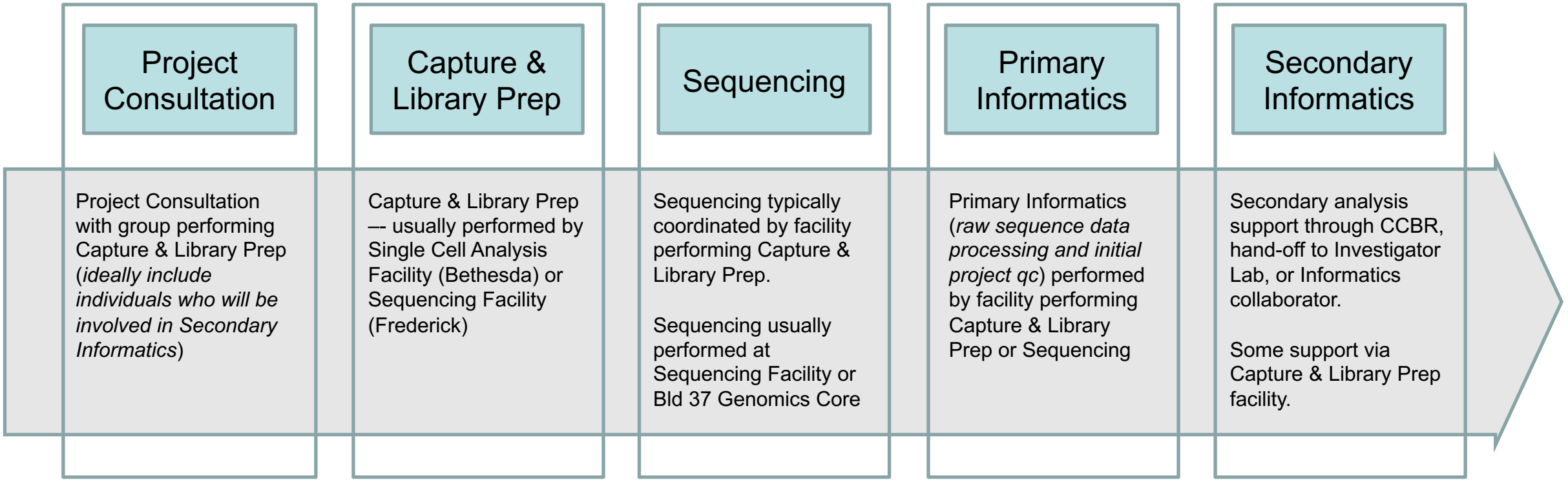
- Cells partitioned into gel beads
- Subsequent capture with gel beads and enzyme
- Whole genome amplification and tagging with cell barcodes
- Sequencing library prepared



Typical project workflow and how to get support

CCR Resources for Integrated Single Cell Analysis

Frederick National Laboratory for Cancer Research



Initiating single cell projects:

Bethesda – Single Cell Analysis Facility
Mike Kelly
michael.kelly3@nih.gov

Frederick – Sequencing Facility
Monika Mehta
monika.mehta@nih.gov

Informatics Support:

Maggie Cam's team (CCBR)
maggie.cam@nih.gov

Yongmei Zhao's team (SF)
yongmei.zhao@nih.gov

Mike Kelly's team (SCAF)
michael.kelly3@nih.gov

Sequencing cores:

Bethesda – Liz Conner's team(GC)
Frederick – Bao Tran's team (SF)

Office of Science & Technology Resource
Dave Goldstein & Mariam Malik

Bioinformatics Training & Education Program
Amy Stonelake