Introduction to Single Cell RNA-sequencing (scRNA-seq)

Harvard Chan Core Single Cell Analysis Course

Mandovi Chatterjee, PhD

Director, Single Cell Core

Harvard Medical School

SINGLE CELL CORE (SCC): TEAM

Core Director

Associate Director

Faculty Advisors:
Allon M. Klein, PhD
Jeffrey Moffitt, PhD
Christophe Benoist, MD, PhD

Funding:
HMS Foundry Award Program

HARVARD MEDICAL SCHOOL | RESEARCH CORES AND TECHNOLOGY
Single Cell Core

https://singlecellcore.hms.harvard.edu

SINGLE CELL CORE: SAMPLE REPERTOIRE

Primary Cells and Tissue

- Whole embryo, embryonic stem cells
- Blood and immune cells, hematopoietic stem cells
- Brain and spinal cord
- Retina
- Thymus
- Lung
- Stomach, intestine, colon
- Heart
- Liver
- Adipose tissue

In Vitro Cultures

- Differentiated cells from iPSC's
- Organoids
- Genetically engineered cells (organ-on-chip)

TECHNOLOGIES SUPPORTED BY SCC

10X GENOMICS

Parse BIOSCIENCES

Fluent BioSciences

PacBio

Oxford NANOPORE Technologies

vizgen

Complete GENOMICS

curio BIOSCIENCE

Single modality:

- Single cell and single nuclei RNA-seq
  - 3'- & 5'-gene expression
  - Fixed RNA profiling
- Single cell ATAC-seq

Multiple modalities (Multiome)

- CITE-seq- cell surface receptor w/ scRNA-seq
- Combined scATAC-seq and scRNA-seq

Long-read sequencing

- PacBio (Kinnex) & Oxford Nanopore Technologies

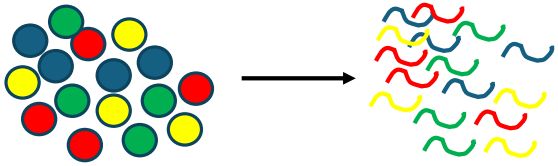
Spatial transcriptomics

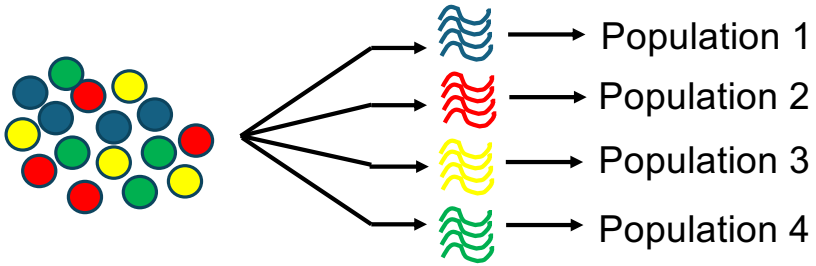
- Image-based
  - MERSCOPE (From Vizgen)
- NGS-based
  - VISIUM/VISIUM HD (From 10X Genomics)
  - STOmics (From Complete Genomics)
  - Curio Seeker (From Curio Bioscience)

TALK OUTLINE

- Overview of the emerging single cell technologies and applications
- Diverse biological questions being addressed by these technologies
- Choosing the correct platform
- Workflow of scRNA-seq experiment- sample prep, library preparation and sequencing
- Challenges- batch effect

WHY SINGLE CELL?

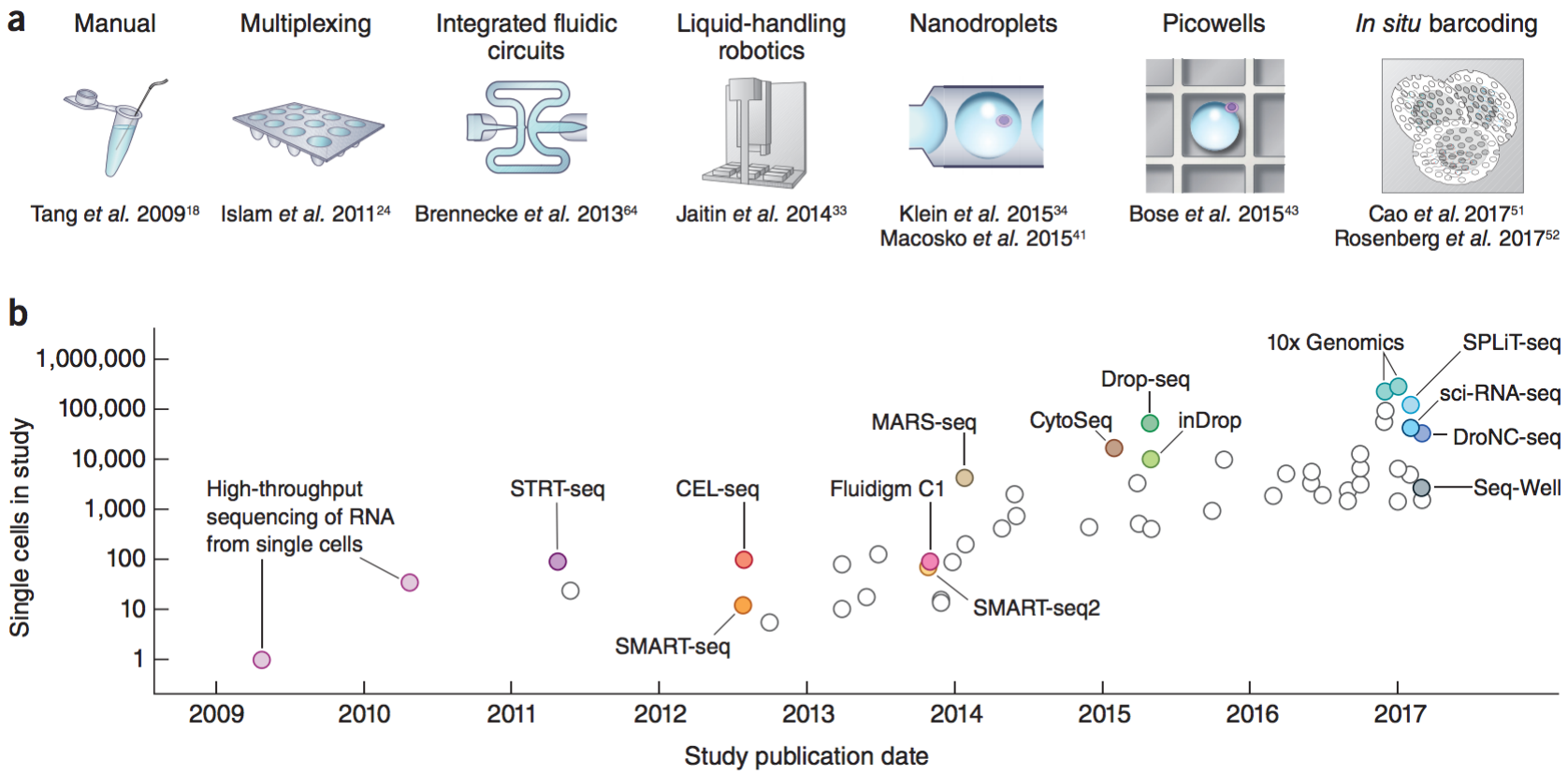




Bulk RNA-seq

scRNA-seq

HISTORY & PROGRESS

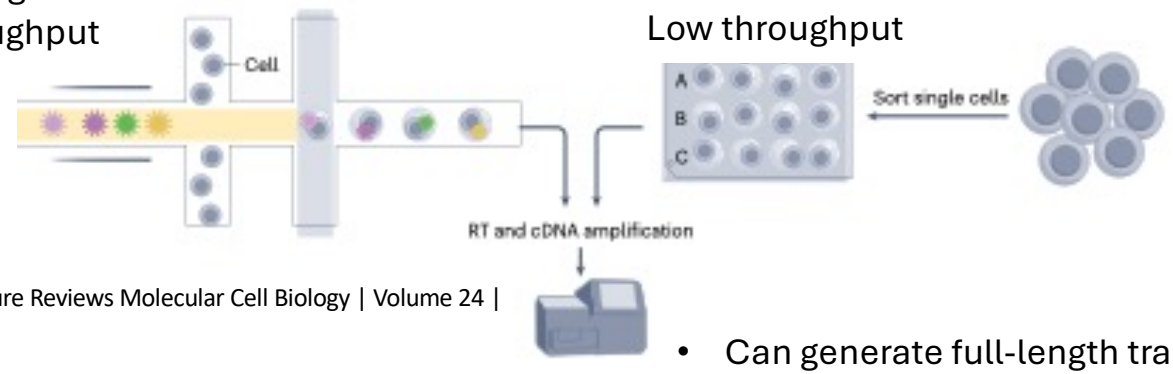


Svennson *et al.*, *Nature Protocols* (2018)

TECHNOLOGIES

High throughput

Low throughput



Adapted from Baysoy et. al; Nature Reviews Molecular Cell Biology | Volume 24 | October 2023 | 695–713

- Throughput is high- can profile many thousands of cells per sample
- Can profile transcripts from either 3' or 5' end. Therefore, allele-specific expression or splice variants cannot be detected (exception: long-read sequencing compatible with single cell workflow)

10X Genomics

BD Rhapsody

Honeycomb Biosciences

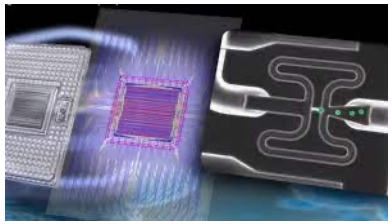
Parse Biosciences

SCIPIO Bioscience

Fluent Biosciences

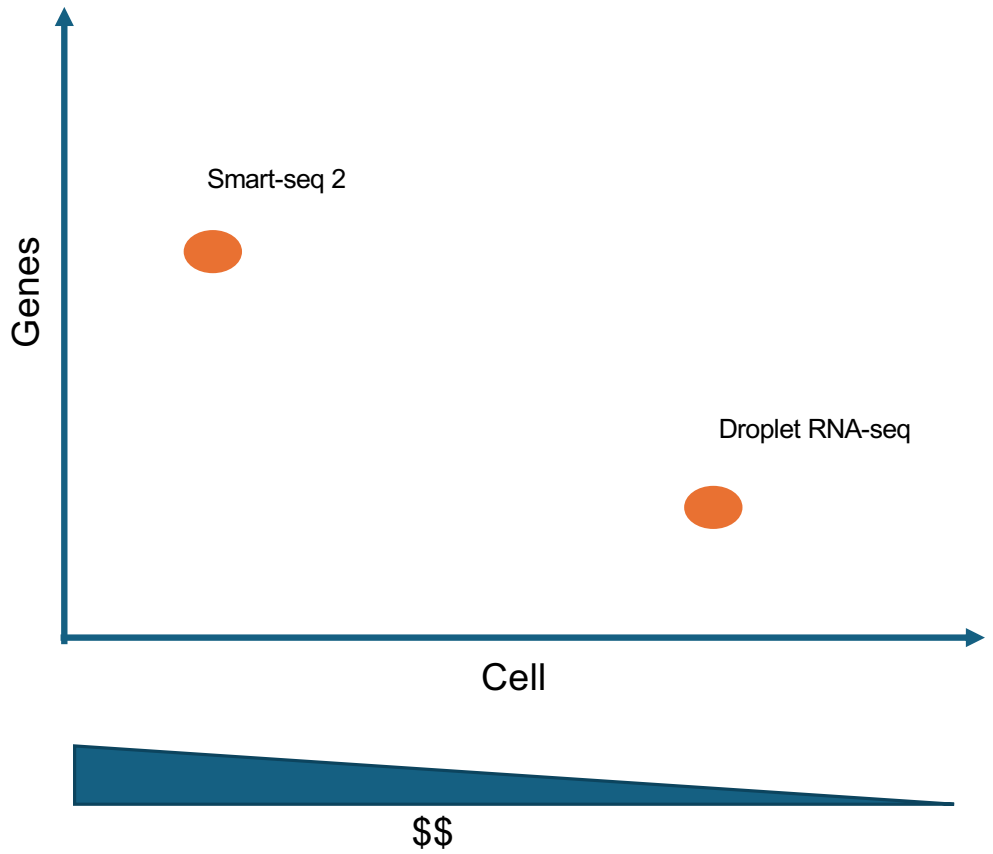
- Can generate full-length transcripts.
- Has high sensitivity
- Throughput is low- limited by plate-size or cell numbe

iCell8 from TakaraBio (SMART-seq)

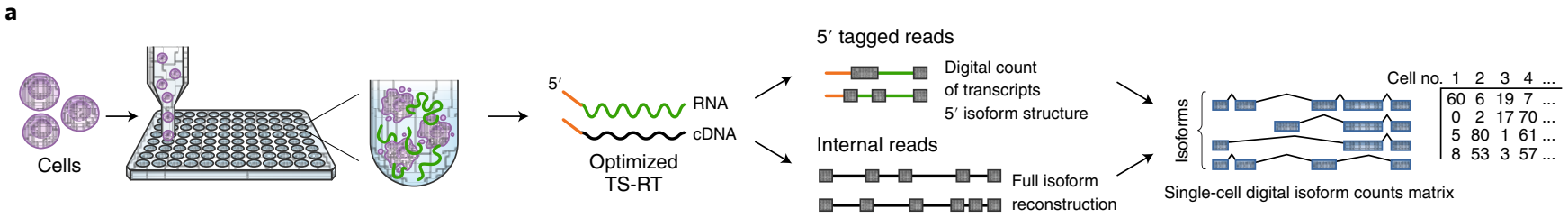


Fluidigm C1

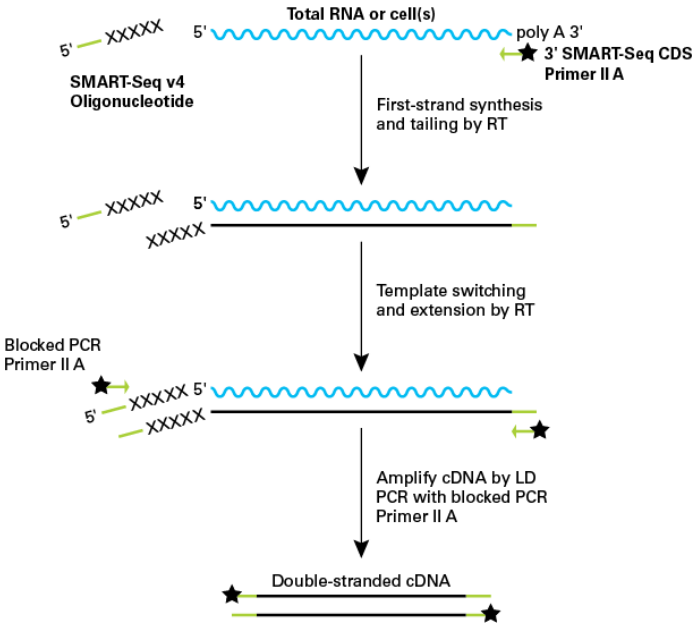
THROUGHPUT VS SENSITIVITY



LOW THROUGHPUT: SMART-SEQ 2/3/4



Macosko; Nature BiotechNology | VOL 38 | JuNe 2020 | 697–700



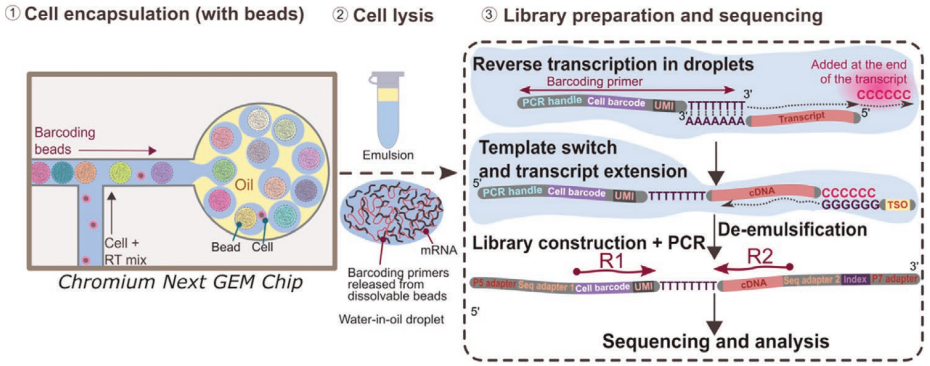
TakaraBio

- Plate-based- individual cells are sorted into single well
- Gives full transcript information. Isoform detection is possible.
- Sensitivity is high.
- Sequencing is costly.

M Hagemann-Jensen et al, Single-cell RNA counting at allele- and isoform-resolution using Smart-seq3 bioRxiv

MICROFLUIDIC DROPLET

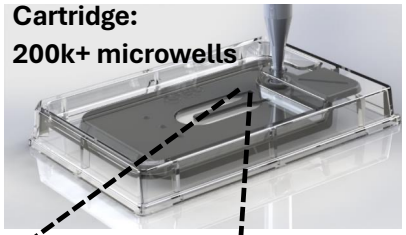
10X GENOMICS®

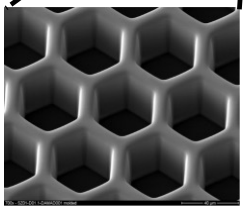


- Barcoding inside the droplet
- Supports a plethora of applications- scRNAseq, multiome, CRISPR screening
- Single cell, nuclei or fixed cells

MICROWELL

BD

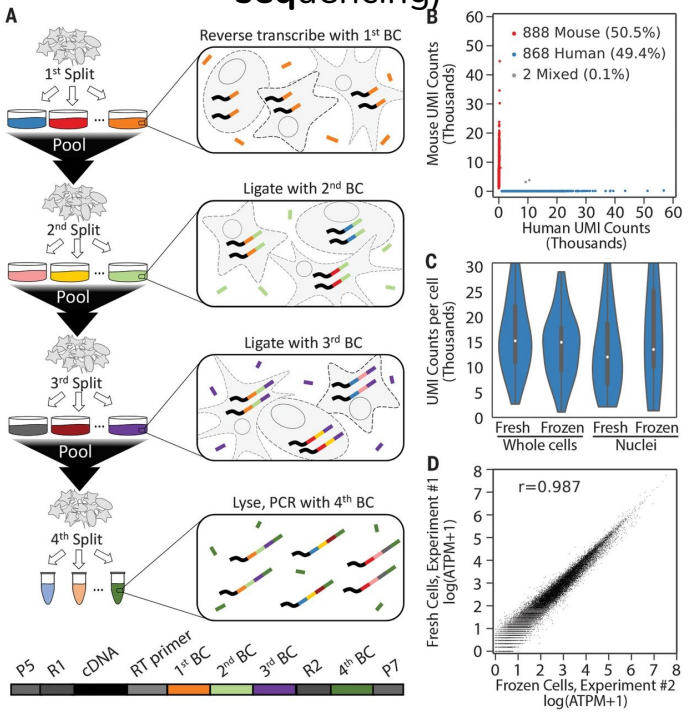




- Magnetic bead with immobilized oligos
- Real-time cell count, viability, true doublet rate
- Supports scRNAseq and multiome
- Single cell and nuclei

COMBINATORIAL BARCODING

SPLiT-seq (**S**plit-**p**ool **li**gation-based **t**ranscriptome **seq**uencing)

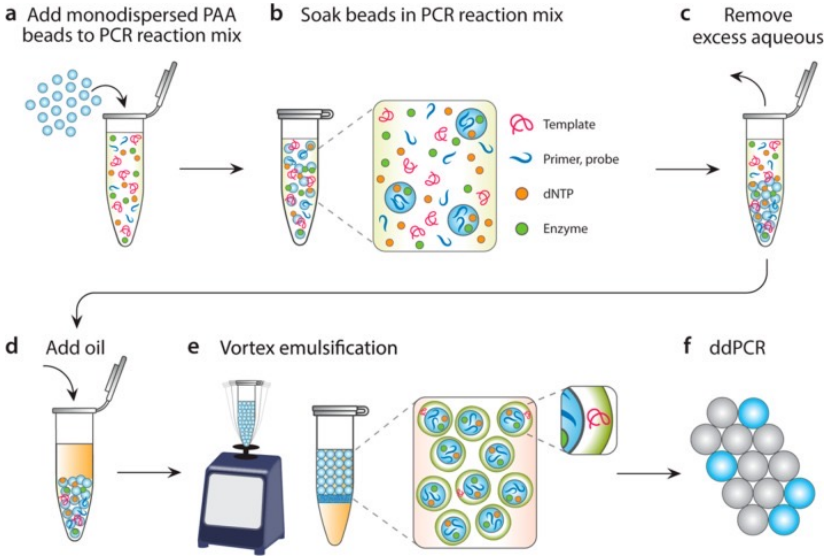


Rosenberg et. al; SCIENCE, Vol. 360, Issue 6385, pp. 176-182

- Combinatorial barcoding by in-situ ligation
- Supports scRNA-seq and CRISPR screening
- Fixed cells and nuclei

MICROFLUIDIC-FREE DROPLET

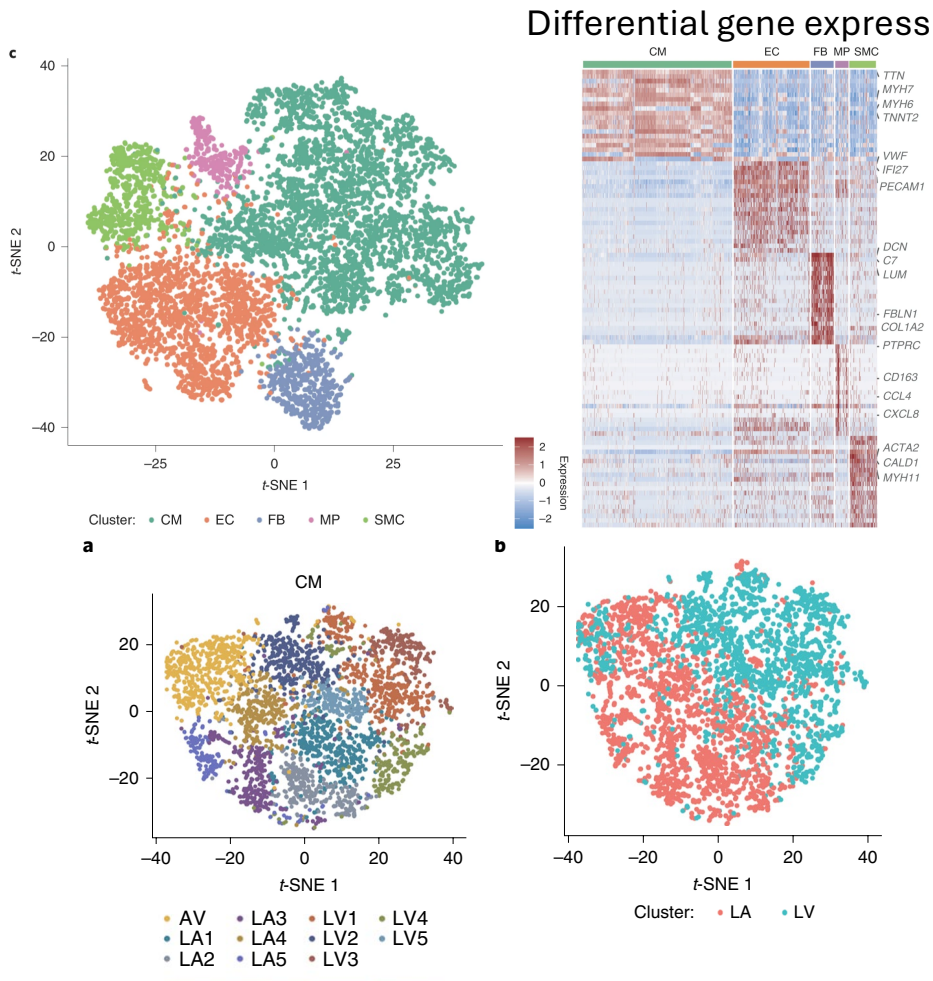
PIP-seq (**P**re-templated **I**nstant **P**artitions)



Hatori et. al; Anal Chem. 2018 Aug 21; 90(16): 9813–9820.

- Barcoding inside the droplet
- Supports scRNA-seq
- Single cell and nuclei

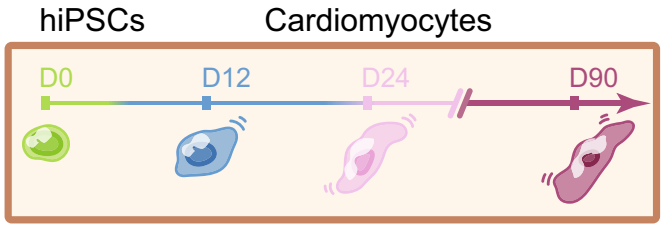
CARDIAC TISSUE PROFILING USING ICELL8

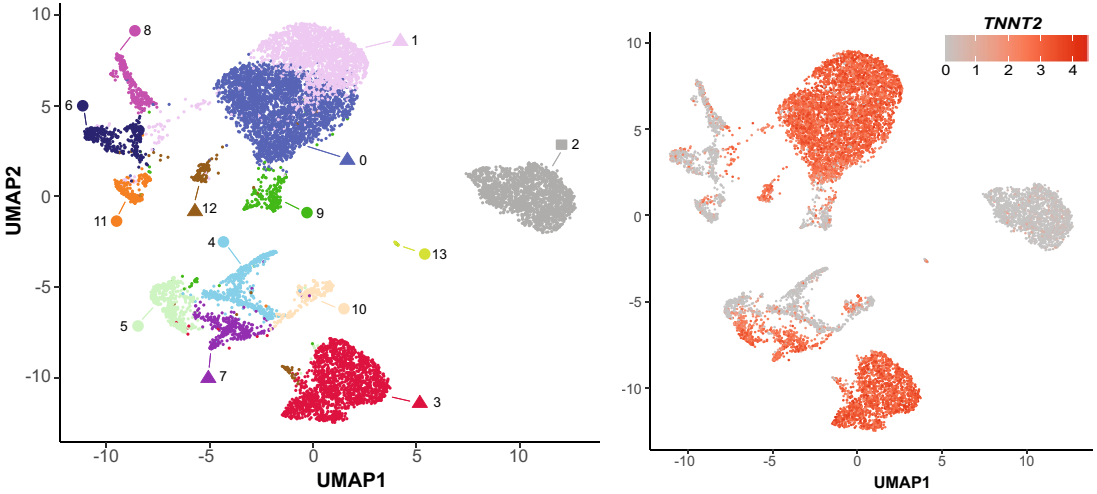


- Tissue type- normal adult human heart
- Both cardiomyocytes and non-cardiomyocytes
- ~2000 genes per cell were detected
- ~7500 cells profiled
- Median sequencing depth- ~300,000 reads per cell

Wang et. Al; Nature CELL BIOLOGY | VOL 22 | JANUARY 2020 | 108–119

CARDIAC TISSUE PROFILING USING SPLiT-SEQ



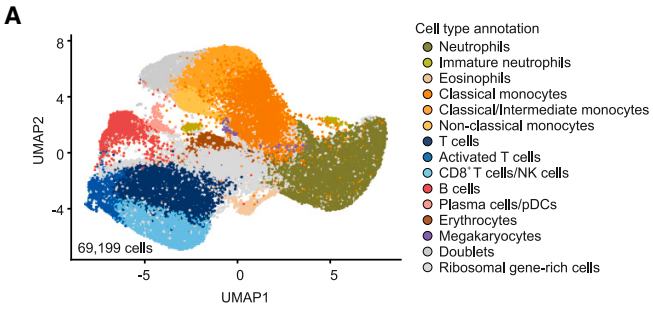


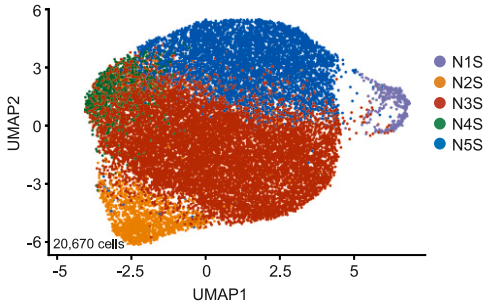
- Sample type- Human iPSC-derived cardiomyocytes
- Time-point study- Cells from a total of 55 samples from multiple independent differentiation experiments with 4 different cell lines, 2 differentiation protocols
- 4 time-points
- ~12,000 cells were clustered here
- Expected seq depth- Min. 20,000 reads per cell

Grancharova et. Al; Sci Rep. 2021; 11: 15845.

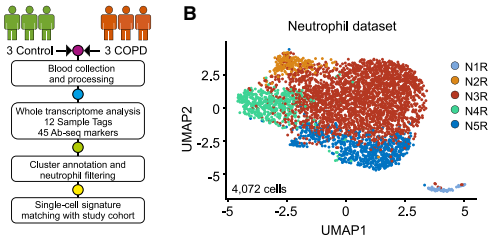
Neutrophil profiling using multiple scRNA-seq methods

Seq-well

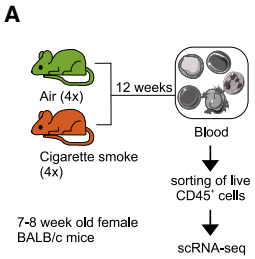


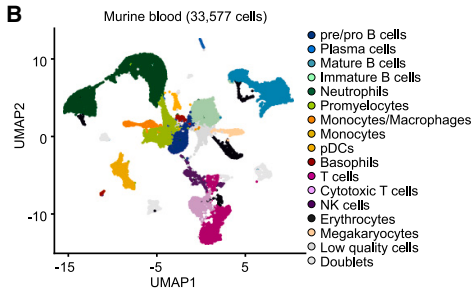


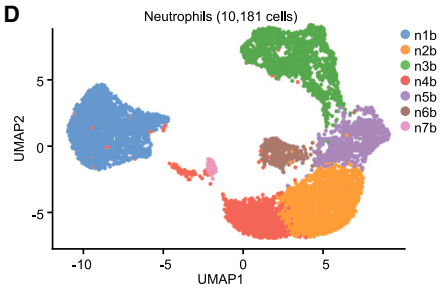
BD Rhapsody



10X Genomics





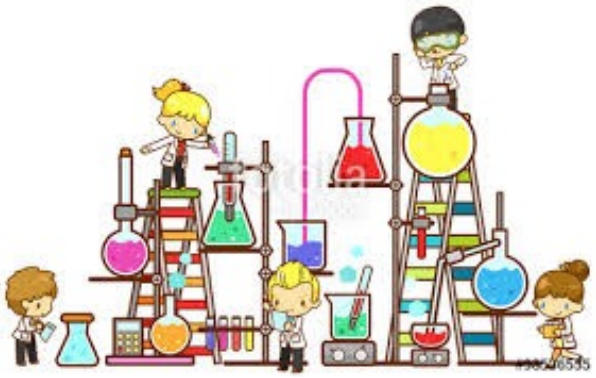


Kapellos et. Al; Cell Reports *42*, 112525, June 27, 2023

KEY CONSIDERATIONS FOR CHOOSING A TECHNOLOGY

- Biological question/ experimental design
  - Sample type- fresh/frozen/fixed; single cell or nuclei
  - Assay type- single cell or multiome or immune profiling
  - Cell type of interest- cell size, fragility of cells etc.
  - Do you need full length transcript information? Long-read sequencing is an option
- Scale
  - How many sample?
  - How many cells per sample?
  - Is it a low input sample?
- Cost
  - Rarer the cell type of interest, more cells one needs to profile
  - Is sample pooling an option?
  - Sequencing depth

scRNA-seq WORKFLOW



Sample prep

10X Genomics

BD Rhapsody

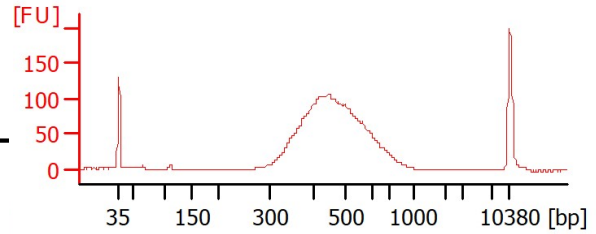
Honeycomb Biosciences

Parse Biosciences

SCIPIO Bioscience

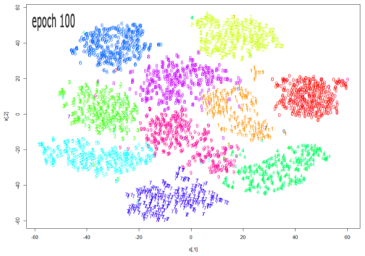
Fluent Biosciences

Barcoding

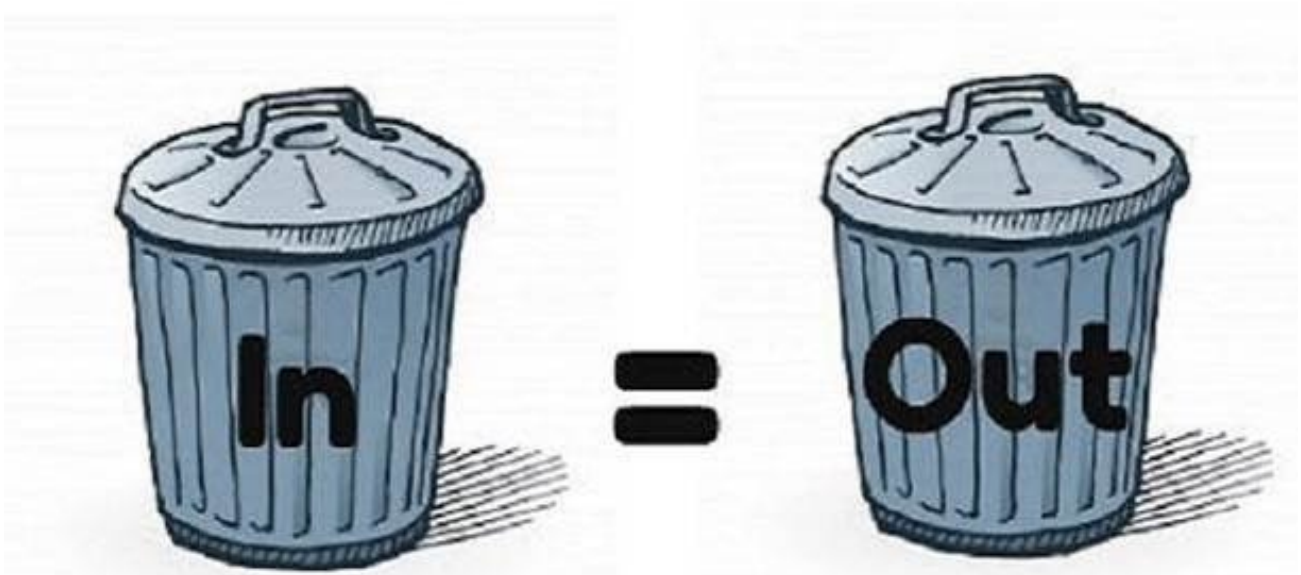


Library

Sequencing

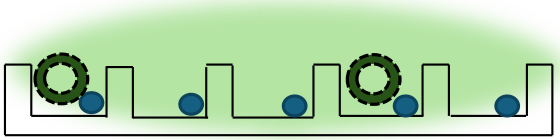


Analysis


In
=
Out

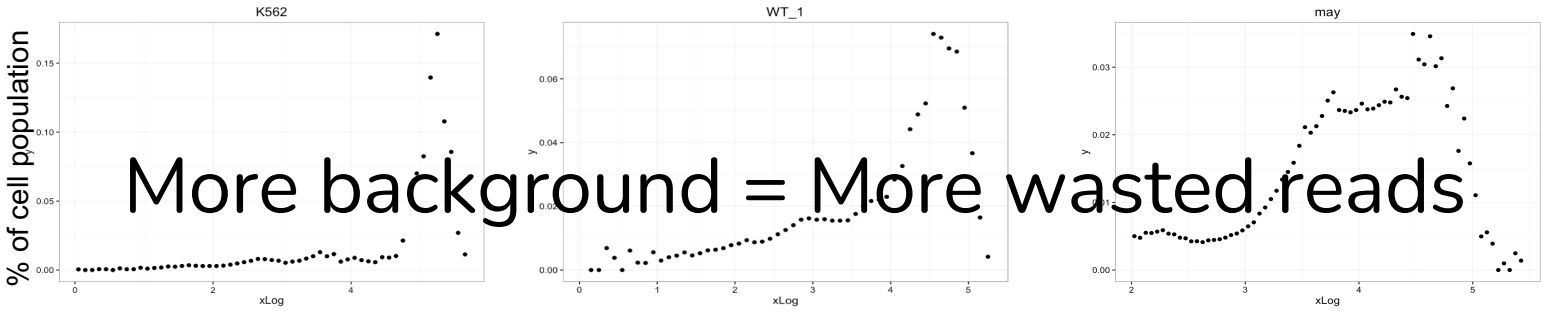
HIGH SAMPLE QUALITY

- High viability
  - More than 90% viability is ideal.
  - Cell membrane integrity is required until they are encapsulated.
- Good single cell suspension. No clumps.
- Clean prep with little or no debris.



Nanowell-based barcoding platform

Droplet-based barcoding platform



More background = More wasted reads

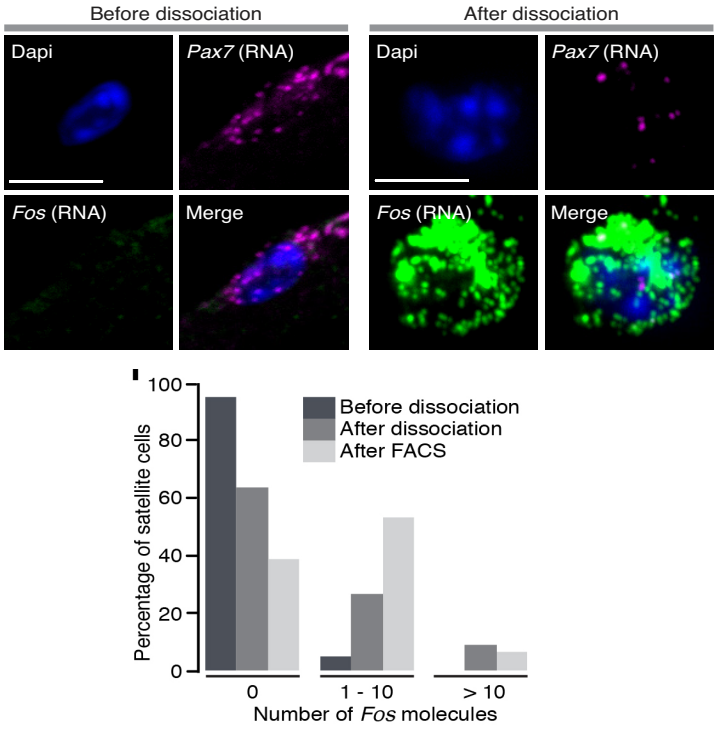
Reads/ cell barcode
Ideal data

Reads/ cell barcode

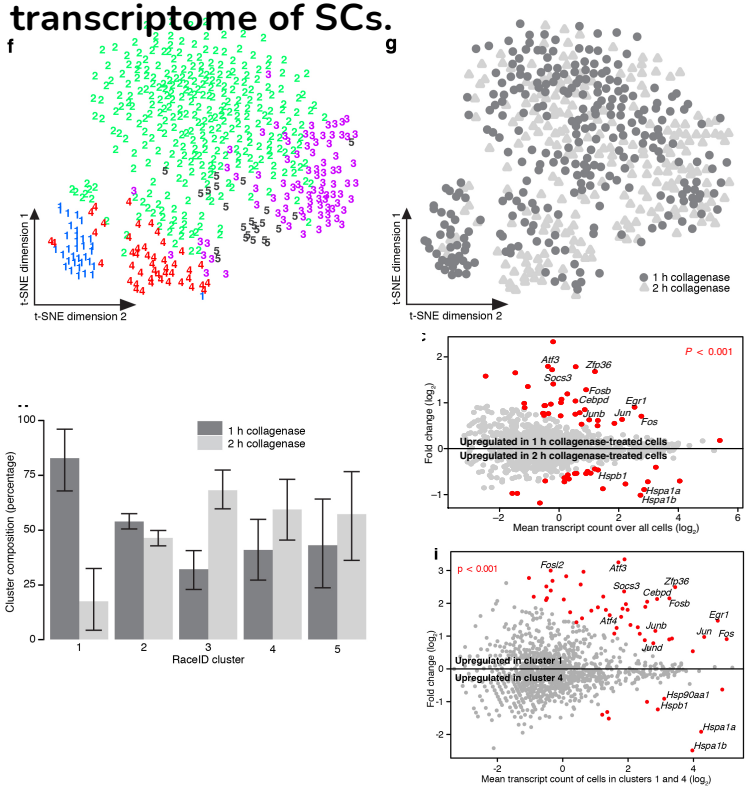
Reads/ cell barcode
Free-floating RNA

TRANSCRIPTOME IS STRESS-SENSITIVE

Single molecule FISH analysis shows that *Fos* expression is induced during the SC isolation procedure.



Dissociation time experiment confirms that the dissociation procedure influences the transcriptome of SCs.



Van Den Brink et.al; *Nature Methods* **volume 14**, pages935–936(2017)

SAMPLE PREPARATION

- Enzyme-based dissociation- trypsin, collagenase, liberase, accutase.
- Gentle washes.
- Dead cell removal kit, filtering out the debris.
- Density gradient (Ficoll, Optiprep)

Solid tissue

Adherent cell culture

Dissociation

Suspended cell culture

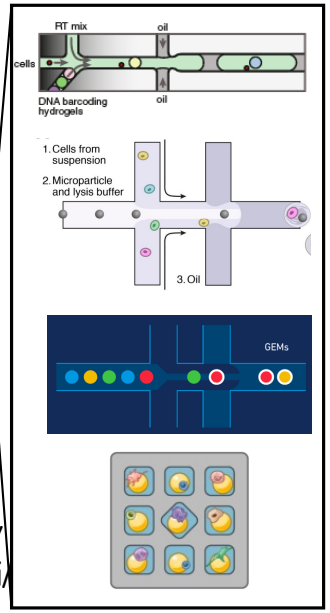
Liquid tissue

Enrichment

- Magnetic bead selection
- FAC sorting (Stain for viability)

Quality check

- Cell conc.
- Cell buoyancy
- Viability (Dapi/ Trypan blue)



ENRICHMENT METHODS: PROS & CONS

FACS (Fluorescence activated cell sorting)

Pros:
- Enrichment is robust. Can be really useful for rare population of cells.
- Yields good single cell suspension.
- Live/dead sorting by DNA stains, eg. DAPI.

Cons:
- Uses high pressure to sort the cells, therefore can be pretty harsh.
- Can introduce bias in the experiment. Using a broad marker is recommended.
- Long sample prep protocol.

MACS (Magnetic-activated cell sorting)

Pros:
- Gentle on cells.
- Fast protocol.
- Greater number of cells can be processed at a time.
- Not limited by FAC sorter availability.

Cons:
- Number of available surface marker-conjugated to magnet is limited.
- Enrichment is not precise. Not applicable for rare population.

SAMPLE PREP CHALLENGES & SOLUTIONS

Single Cell

PROS
- Superior data quality

CONS
- Time-consuming and time-sensitive
- Limits the number of samples that can be run
- Achieving high cell viability is challenging

Single Nuclei

PROS
- Faster protocol
- Amenable for complex experimental design

CONS
- Challenging
- Presence of debris
- Counting is challenging
- Not compatible for certain cell population.

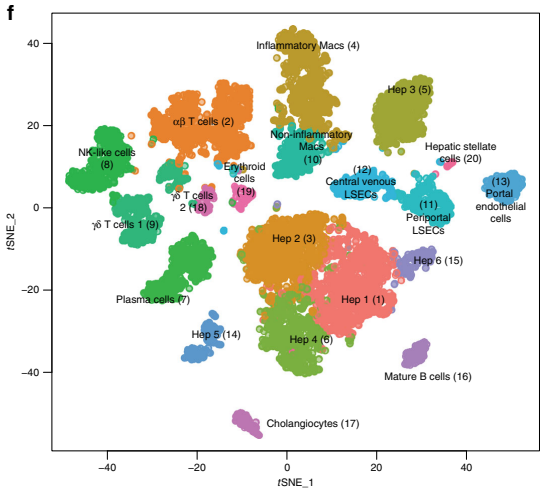
Fixed samples

PROS
- Simplifies experimental design involving many samples
- Choice of fixative depends on experimental goals

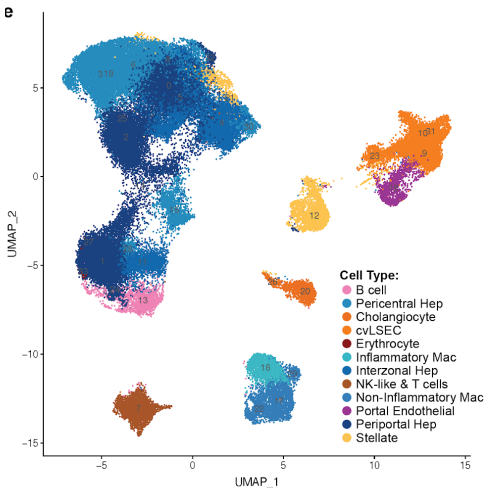
CONS
- High cell loss

SAMPLE PREPARATION VARIES BY QUESTION

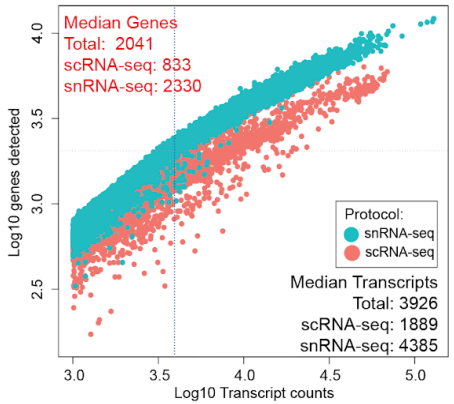
scRNA-seq

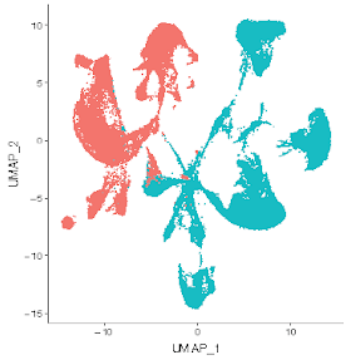


snRNA-seq



- All major hepatic cell types were represented in both scRNA-seq and snRNA-seq, but were captured at different frequencies.
- Cholangiocytes and parenchymal cells were underrepresented.
- Immune cells were more easily detected in the single cell



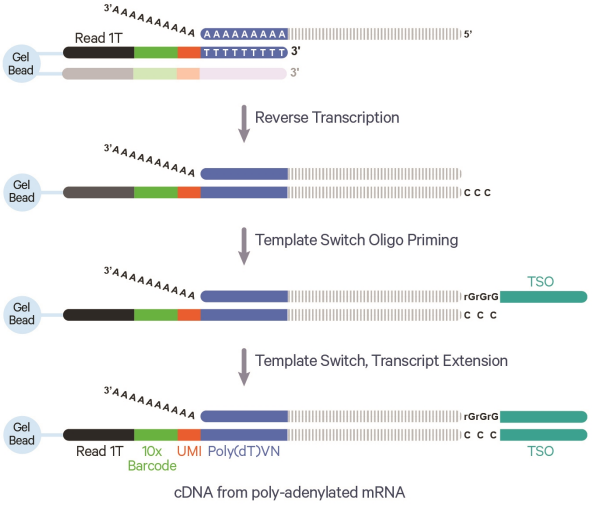


- SnRNA-seq captured a greater diversity of genes than scRNA-seq- high proportion of UMIs in scRNA-seq data are derived from transcripts encoding ribosomal proteins and genes encoded in the mitochondrial genome, which are not present in snRNA-seq data
- scRNA and snRNA don't cluster together due to the systemic difference between the RNA found in the nucleus vs in the cytoplasm. Also the preparation methods for these two are different.

Andrews et. al; bioRxiv; https://doi.org/10.1101/2021.03.27.436882

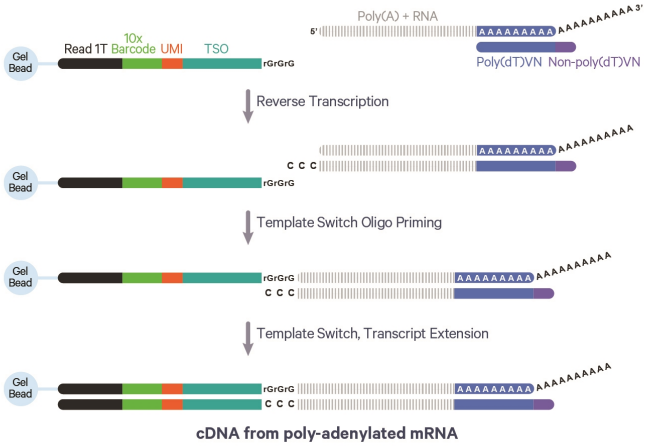
DIFFERENT APPROACHES FOR BARCODING

3’ Capture



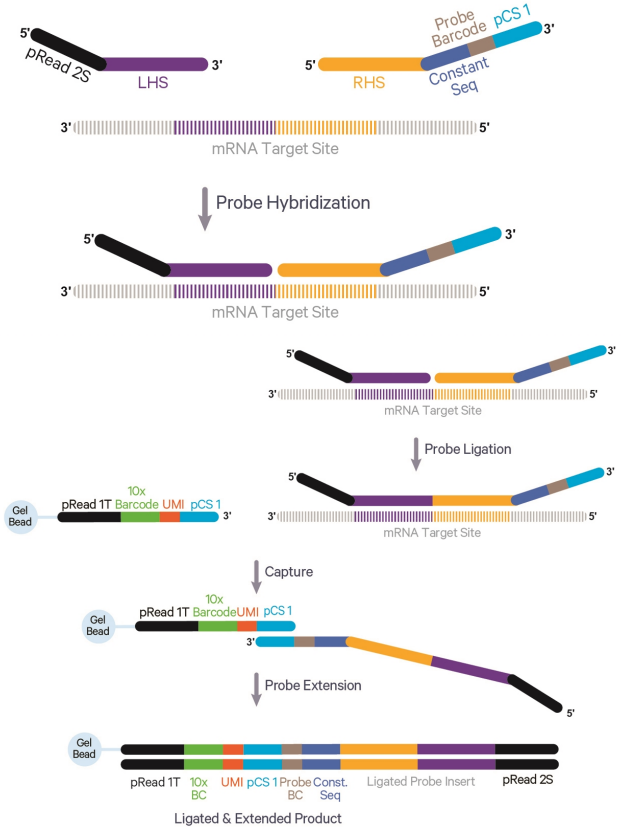
The most common approach for studying single cell biology

5’ Capture



Applicable for immune cell profiling including VDJ recombination and CRISPR

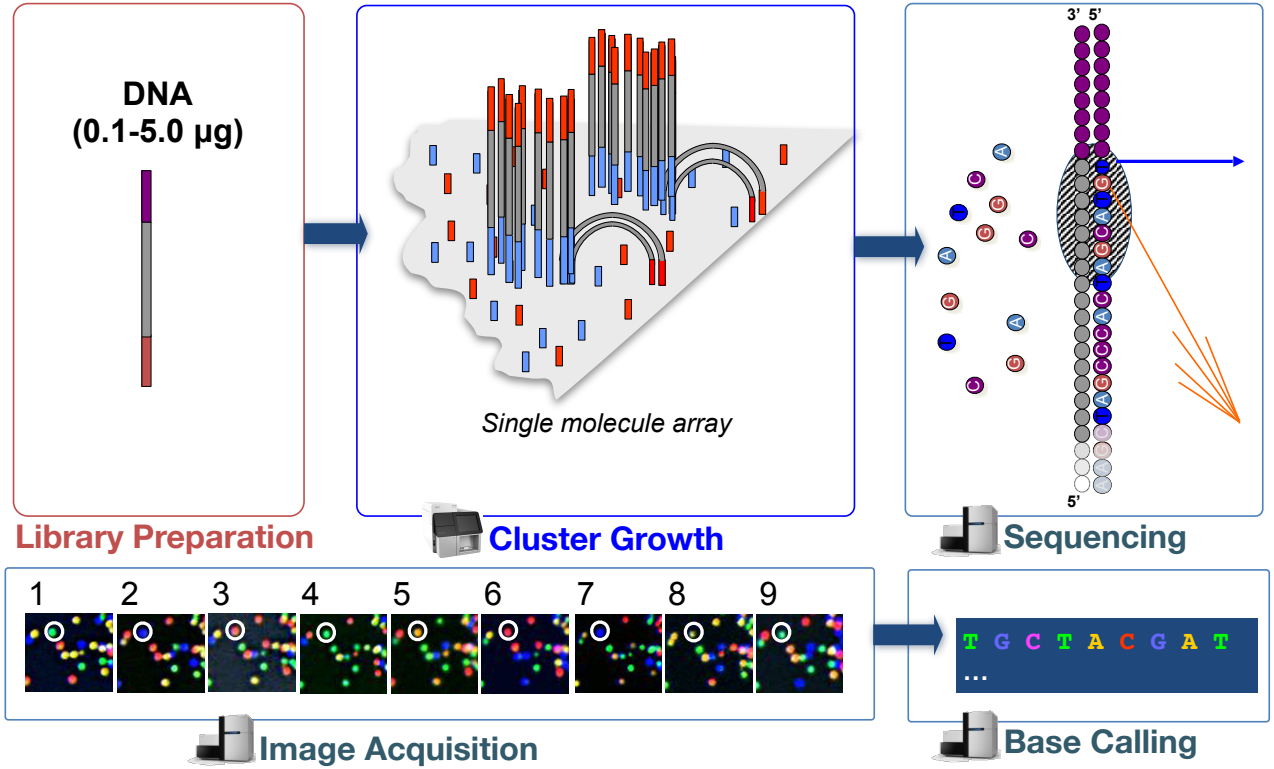
Probe-based



Compatible with fixed cells/nuclei

https://www.10xgenomics.com

SEQUENCING BY SYNTHESIS (ILLUMINA)



https://www.youtube.com/watch?v=fCd6B5HRaZ8&t=3s

Number of clusters = Number of reads
Number of cycles = Length of reads

https://www.illumina.com/systems/sequencing-platforms.html

THROUGHPUT & SEQUENCING DEPTH

The answer is not simple! It depends!

- Two questions:
  - How rare is the cell type of interest?
  - Does it have highly expressed markers?
- Rule of thumb: 50-100 cells with unique transcriptome signature is necessary for forming a distinct cluster in a t-SNE plot.
- Rare cell type of your interest- need to analyze many cells.
- For overall heterogeneity- fewer cells might be enough.
- Sequencing depth- rarer the transcript, higher should be the sequencing depth. Also depends on the RNA content of the cells.

CHALLENGE: BATCH EFFECT

Batch effect represents systematic variation in data, gene expression in this case, resulting from non-biological sources.

Sources

- Sample prep
- Animals from different clutches/ litters
- Different batches of reagents
- Operator
- Library prep performed at different time
- Different sequencing run

Solution

- Wet-lab solution- careful experimental design
  - Use biological replicates. Standard statistical considerations apply. This can become expensive.
  - Same reagent batch, if possible
  - Keeping the operator same
  - Prepping libraries and sequencing together
  - Pooling of samples by sample barcoding (hashtagging/ MULTI-seq)
- Dry-lab solution

- Sometimes batch effects are unavoidable- patient samples, time-course experiments, perturbation experiments
- Batch effect correction works best when individual samples contain sufficient internal complexity to identify shared sources of transcriptional variation

BEST PRACTICES FOR scRNA-SEQ EXPERIMENT

- Good sample prep is the KEY to success.
- Do not rush to the final experiment. Protocol optimization is necessary.
- A well-planned pilot experiment is essential for evaluating sample preparation and other parameters like number of cells, sequencing depth etc.
- Include biological replicates.
- Process drug/ treatment/ mutant and control on the same day.
- Randomize the order of samples, if running on different days.
- Library prep could be a major source of batch effect. All libraries should be prepared together.

THANK YOU