Single cell and single nucleus RNA sequencing experiments: a practical guide to design and analysis

Omics bootcamp I & II: Thursday 5/11/23 and Friday 5/12/23

Amelia Weber Hall, PhD

Gene Regulation Observatory

The Broad Institute of MIT and Harvard

Vascular Discovery:

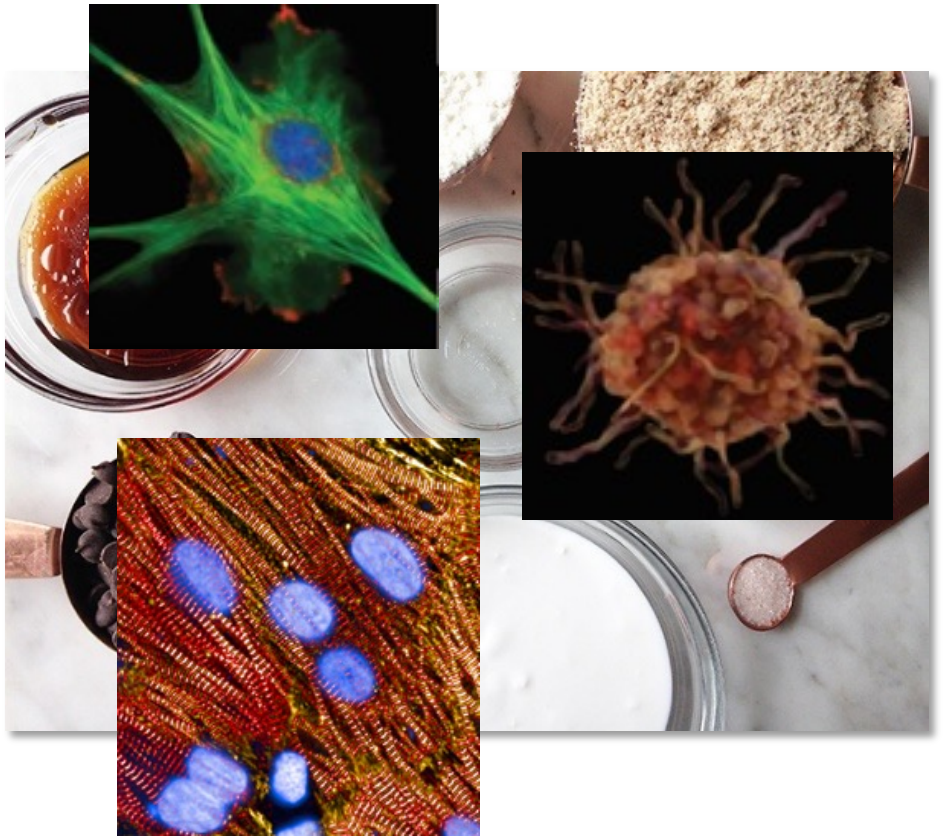
From Genes to Medicine 2023

Outline

- Single cell transcriptomics: a history
- A little intro to ATAC-seq and multi-omic assays
- Specific single cell studies of note in the cardiovascular system
- Experimental design best practices
  - Nuclei isolation methods
  - Tissue storage and validation of single cell findings
- Computational Analysis best practices
- Useful resources
- Limitations
- Brief preview of tomorrow’s coding bootcamp

First, a question: what can we learn from single cell?

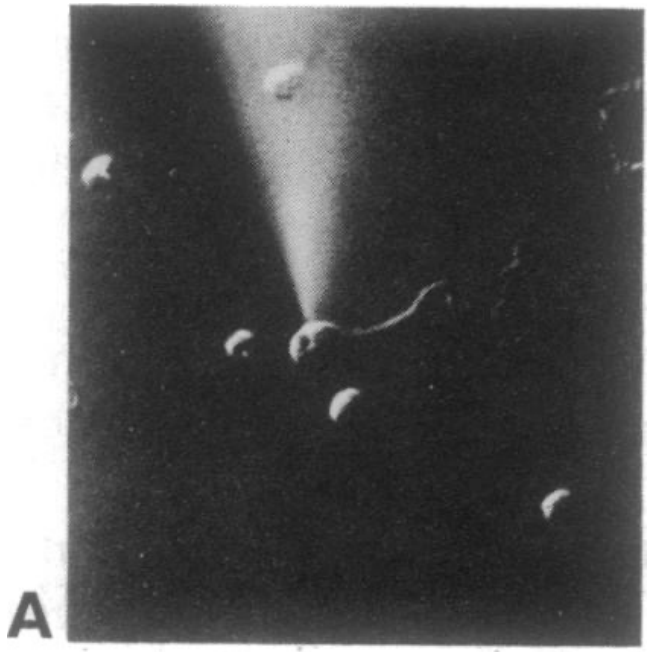
- **We are deconvoluting tissue into cellular subtypes**
- The goal is to learn more about what subtypes make up a complex and heterogeneous tissue



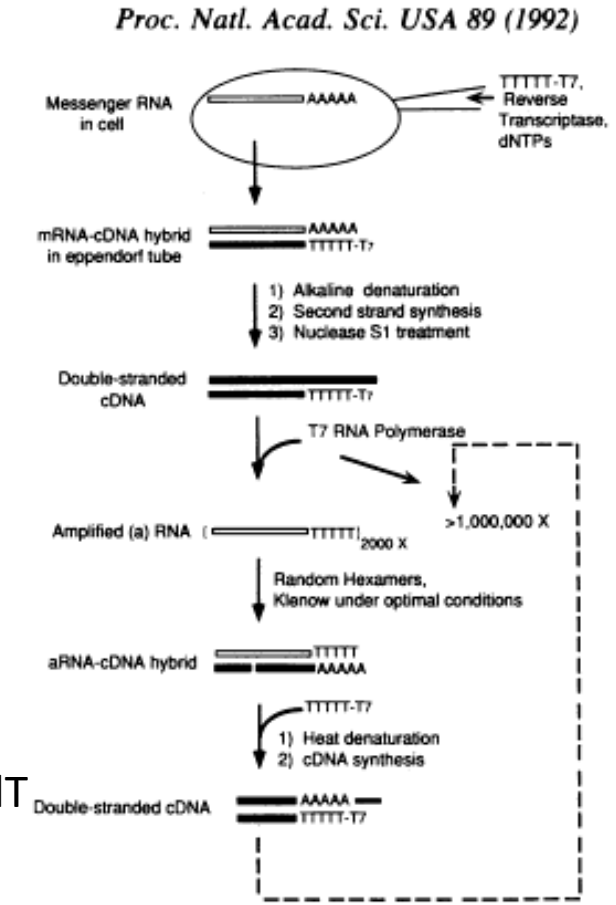
Many thanks to Nathan Tucker for this excellent slide!

Single cell transcriptomics: a History

- Eberwine (1992, neurons) and Iscove (1990, hematopoietic cells)

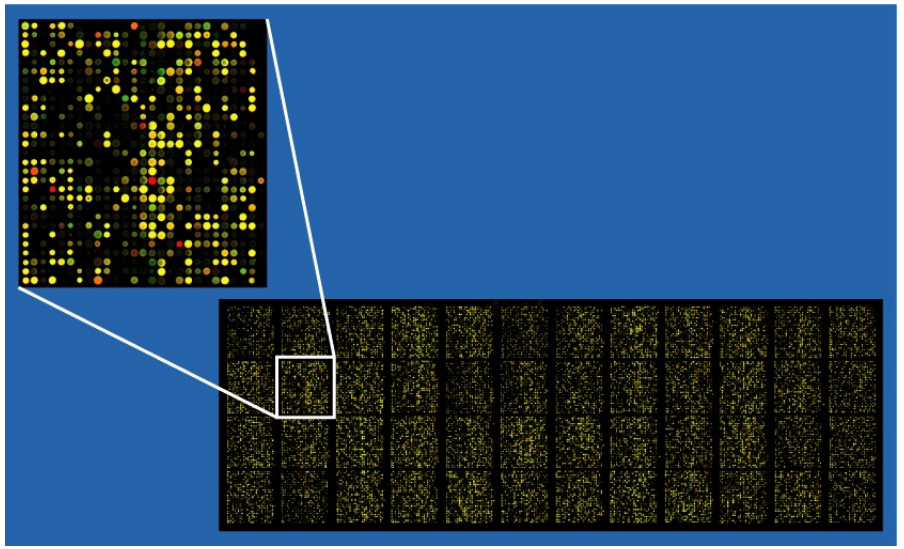


Patch-clamp of a rat hippocampal pyramidal neuron to inject an oligo-dT Nucleotide, dNTPs and RT enzyme



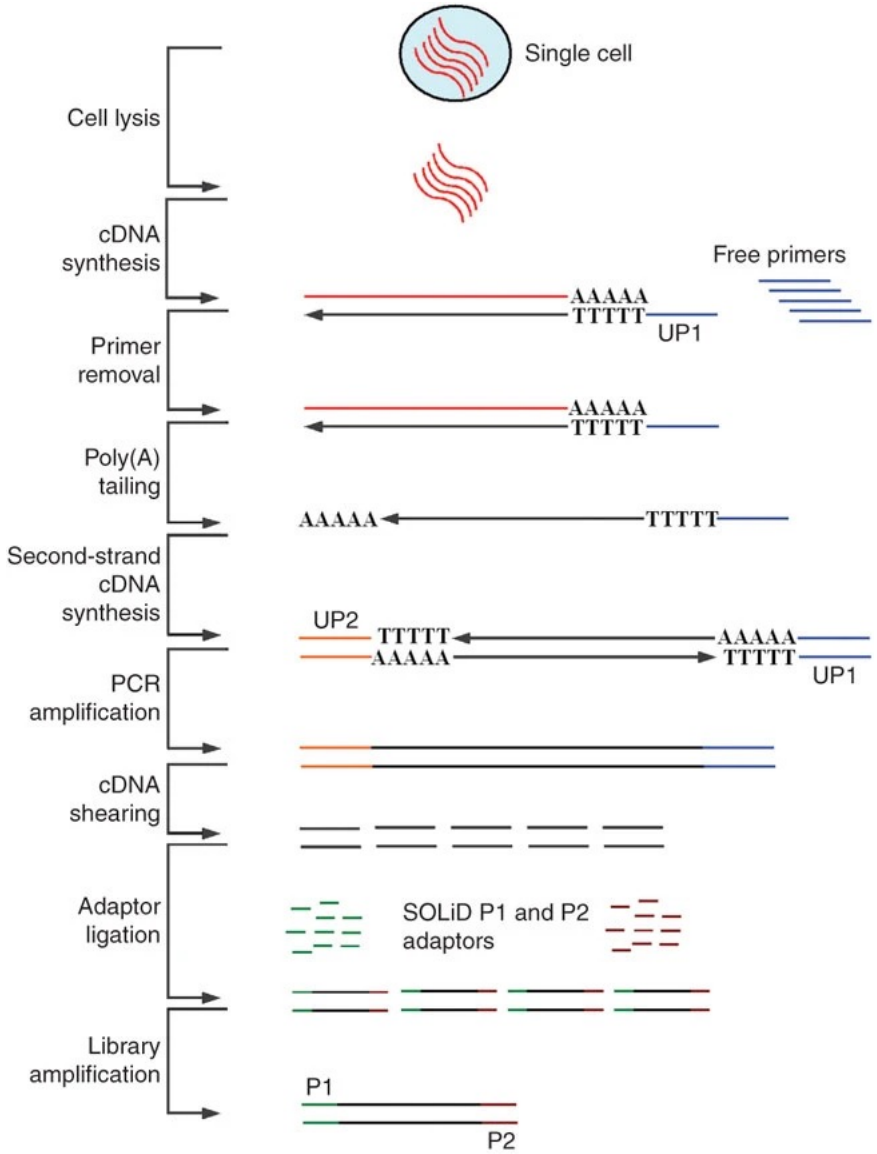
Initially, cDNAs were cloned for further study.

Starting in 1995, the advent of microarrays allowed for chip based study of the transcriptome using hybridization to an array of probes.

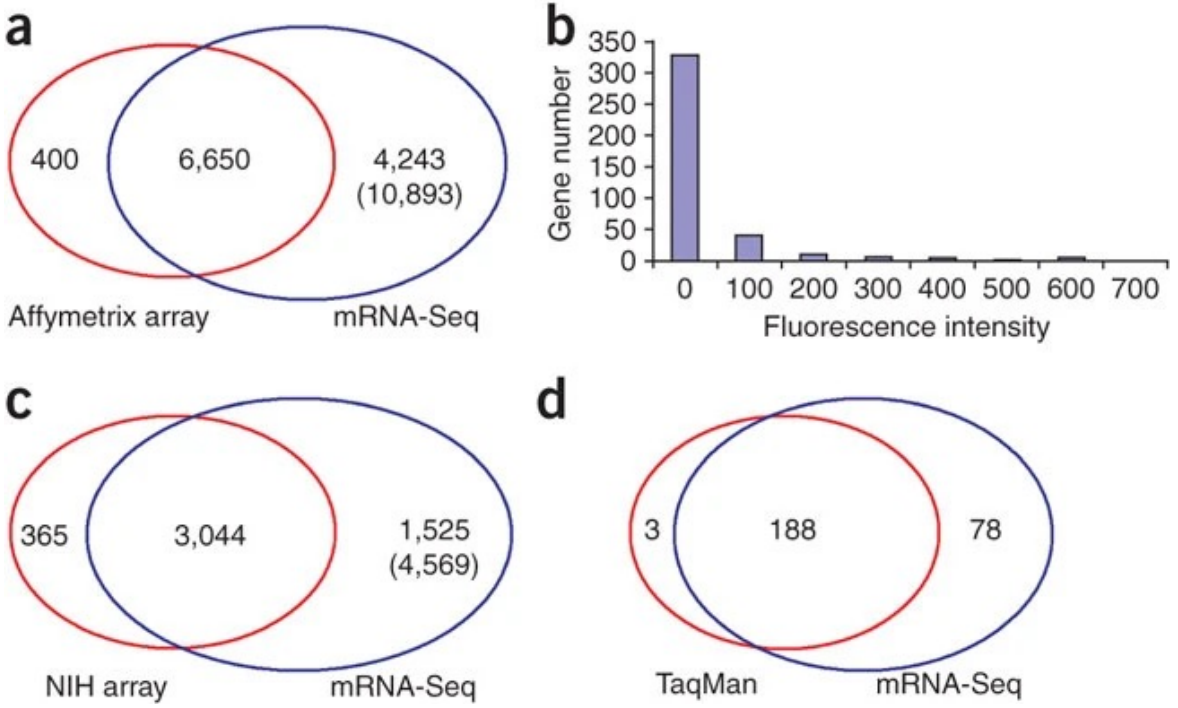


By Paphrag at English Wikipedia - Transferred from en.wikipedia to Commons., Public Domain, https://commons.wikimedia.org/w/index.php?curid=1612185

Single cell transcriptomics: a History



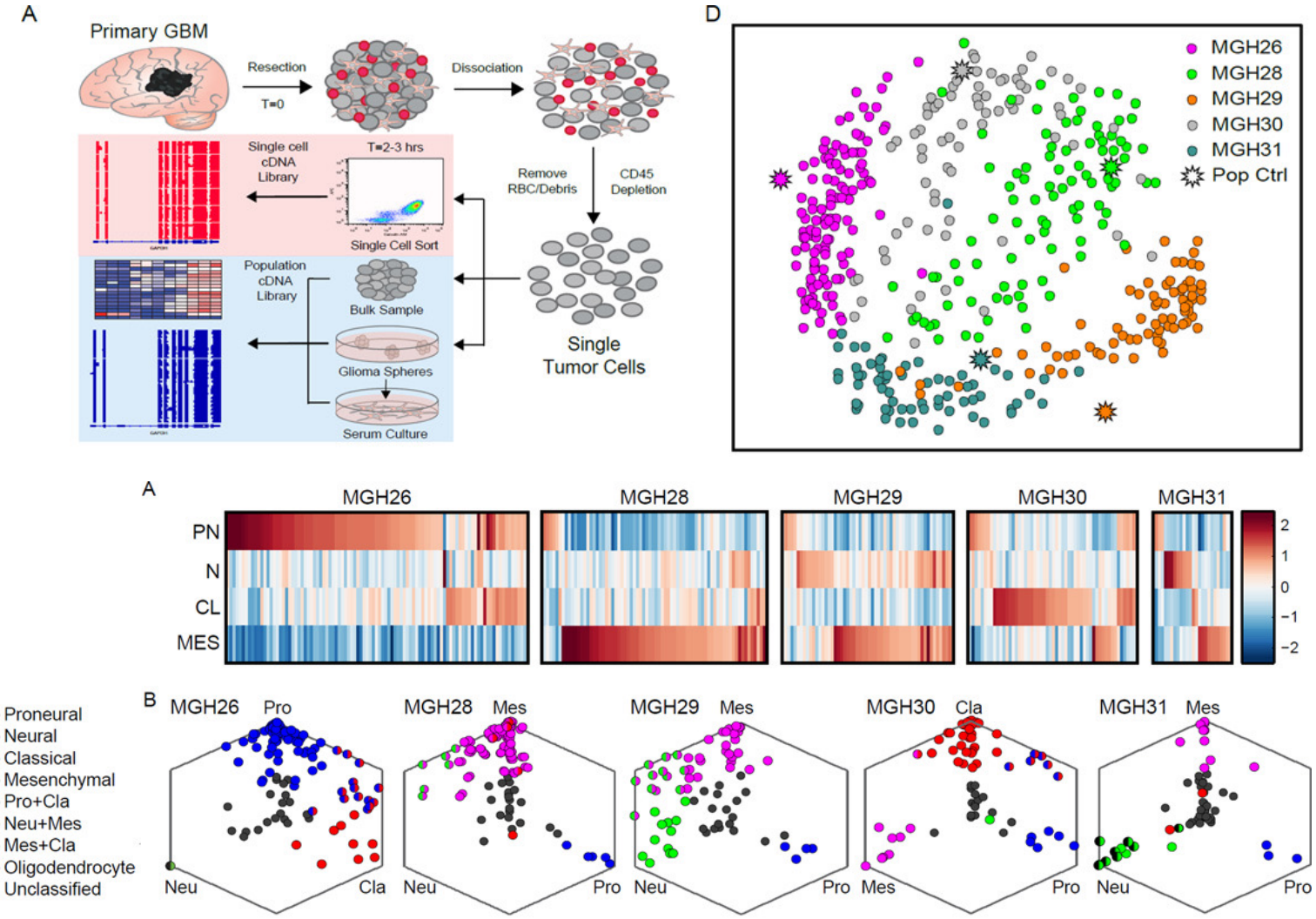
- Tang et al in 2009 were the first to use next generation sequencing on a single cell
  - Sequenced a single mouse blastomere
  - Can detect many more transcripts than using a microarray



Tang, F., Barbacioru, C., Wang, Y. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* **6,** 377–382 (2009). https://doi.org/10.1038/nmeth.1315

Single cell transcriptomics: a History

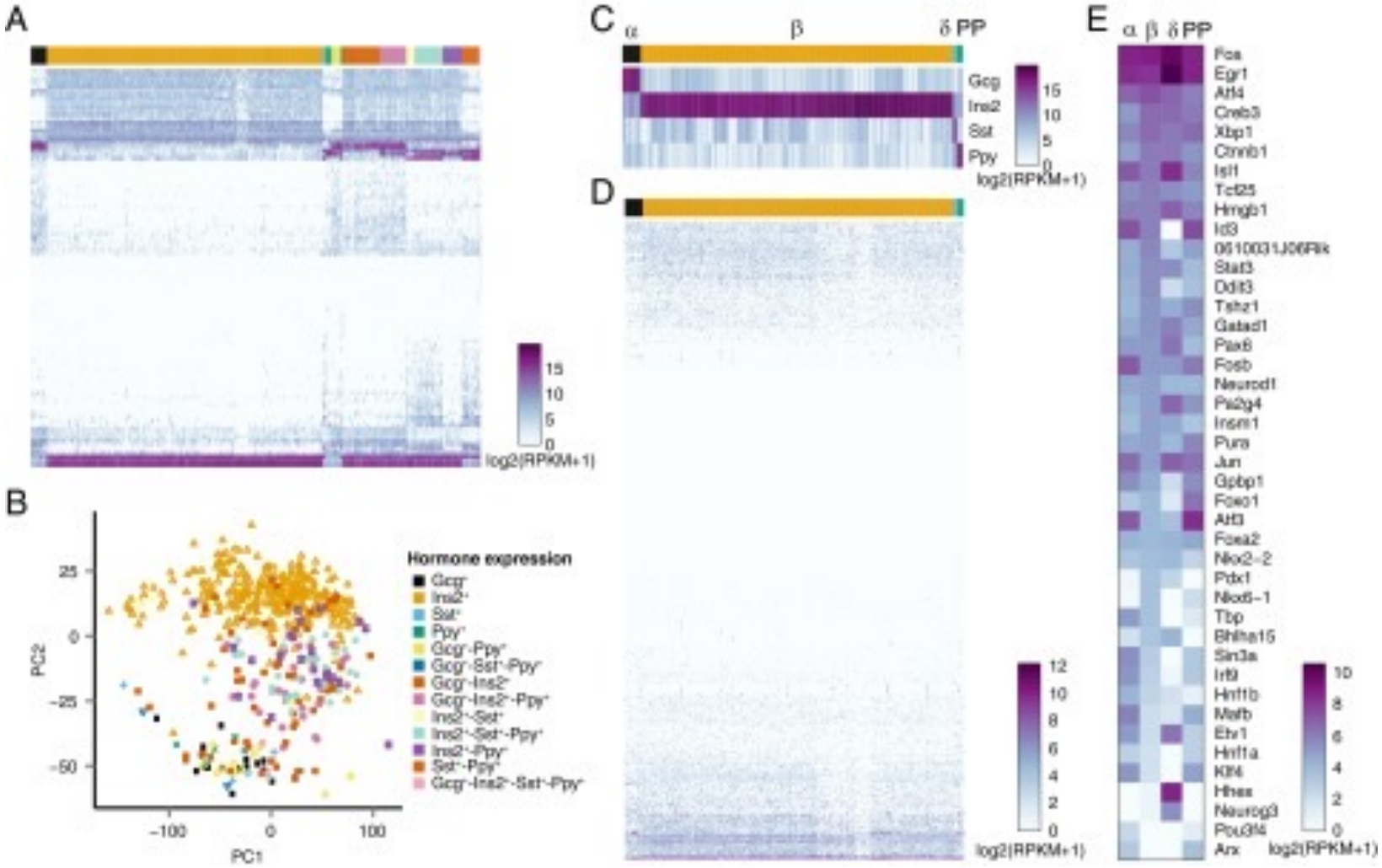
- Patel et al 2014 – sequenced 430 single cells from 5 primary human glioblastoma tumors
  - Glioblastoma has 3 major subtypes, defined by transcription
  - But each tumor contained cells that resembled all 3 subtypes (independent from bulk transcriptional analysis)
- SMART-seq to amplify mRNA



https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4123637/

Single cell transcriptomics: a History

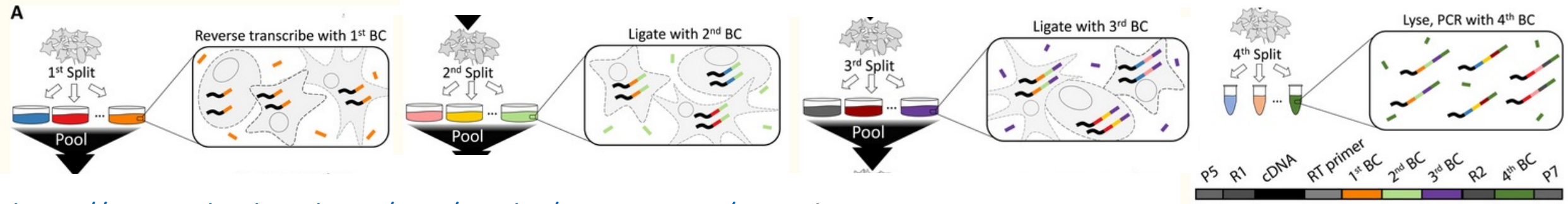
- Xin et al 2016 – sequencing of mouse pancreative islets
- Used the Fluidigm C1 platform to isolate cells for single cell RNA-seq
- Sequenced 622 cells, lost close to 50% due to QC



https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4812709/

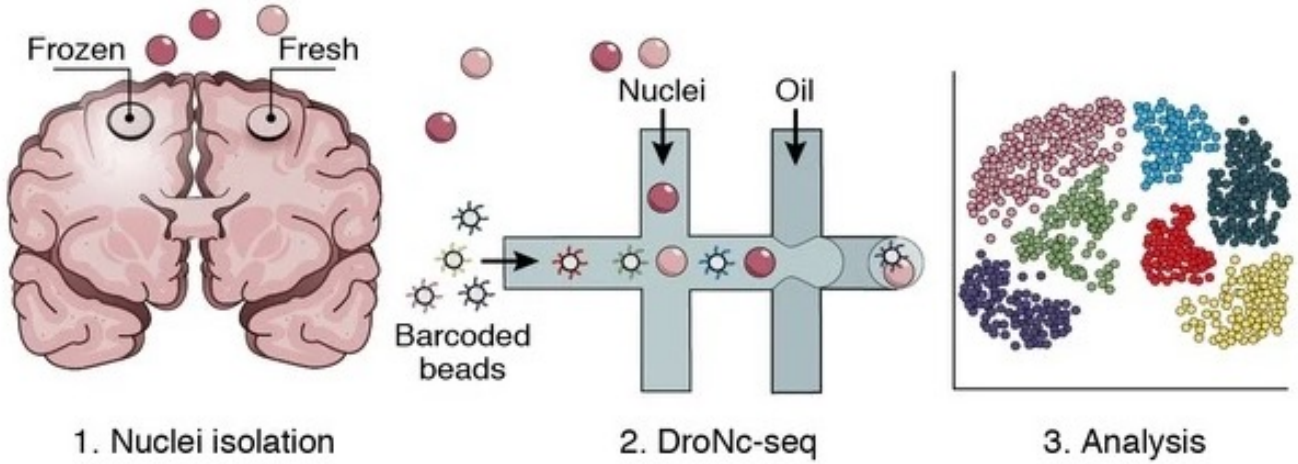
Single cell transcriptomics: a History

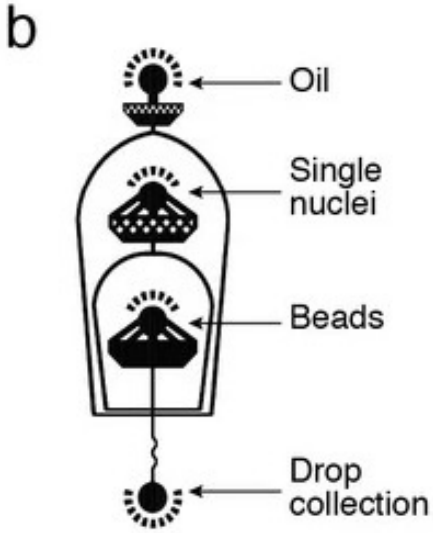
SPLiT-seq



https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7643870/; March 2018

DroNc-seq

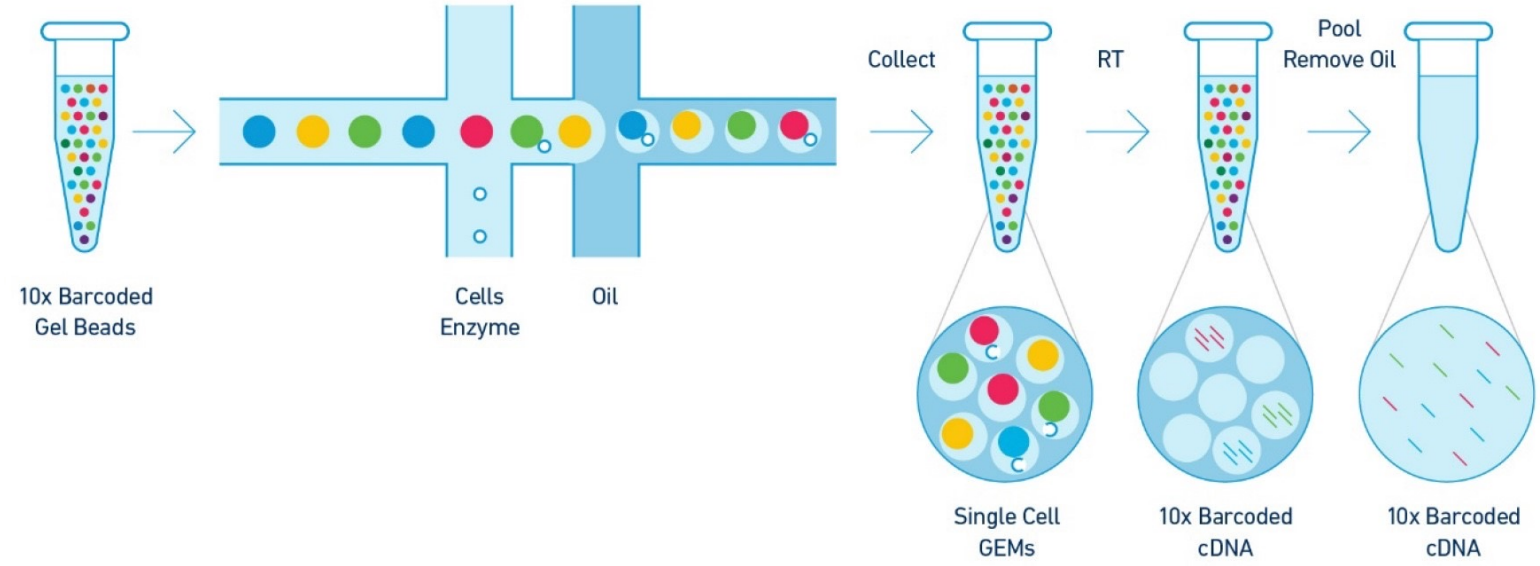


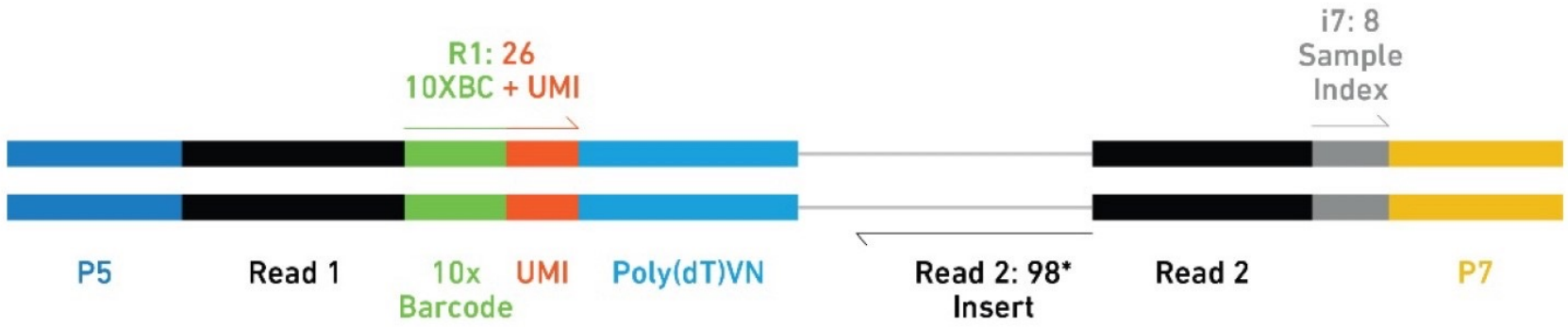


https://www.nature.com/articles/nmeth.4407; August 2017

Single cell transcriptomics: a History

- 10x genomics is effectively the commercial form of droNc-seq
- Uses 3’ capture of transcripts
- Multi-omics assays are being developed





Single cell multi-omic methods: ATAC-seq

- Generally transcriptome and chromatin accessibility (ATAC-seq)
- 10x offers a kit that does this!
- Several other methods: SHARE-seq (SpliT-seq derived), ISSAAC-seq (Xu et al., 2022)

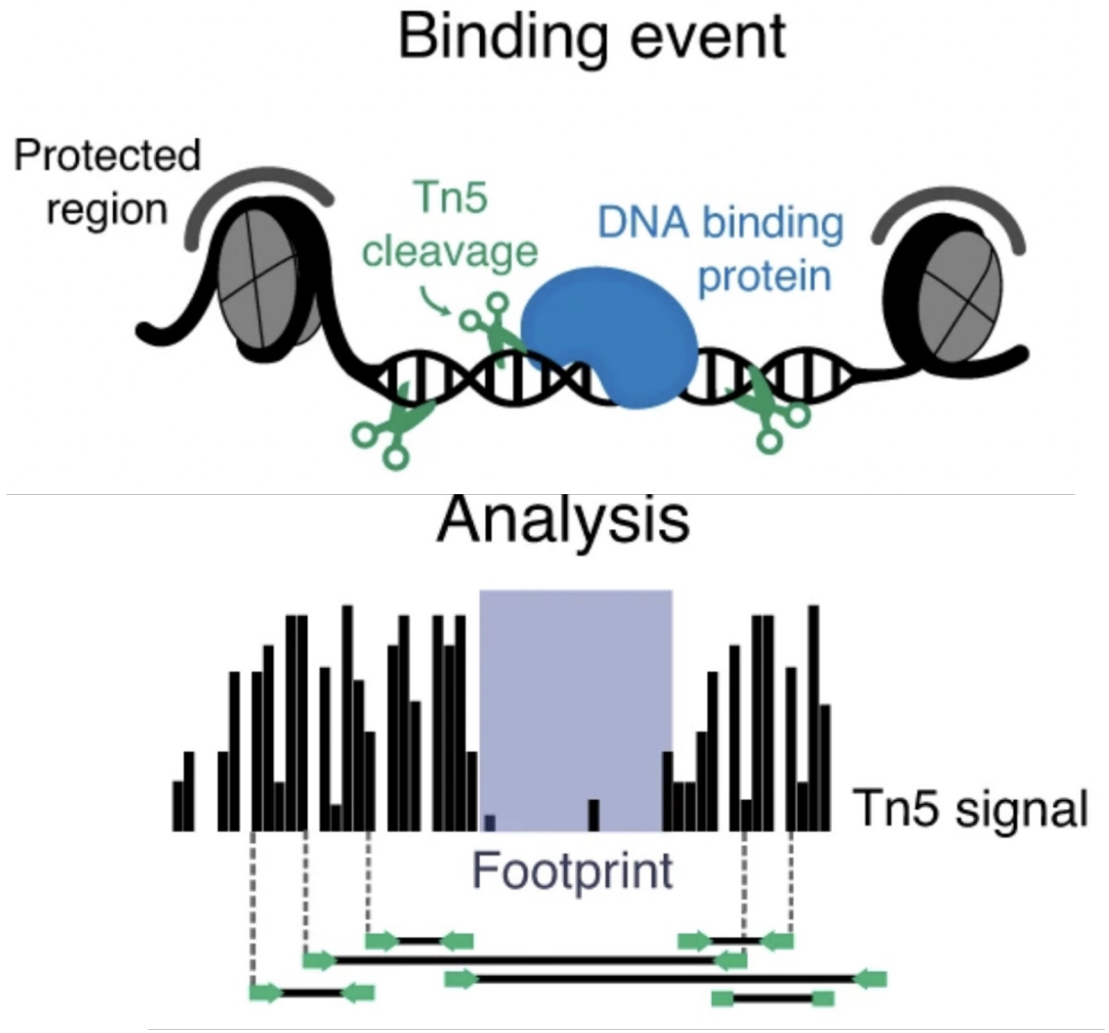


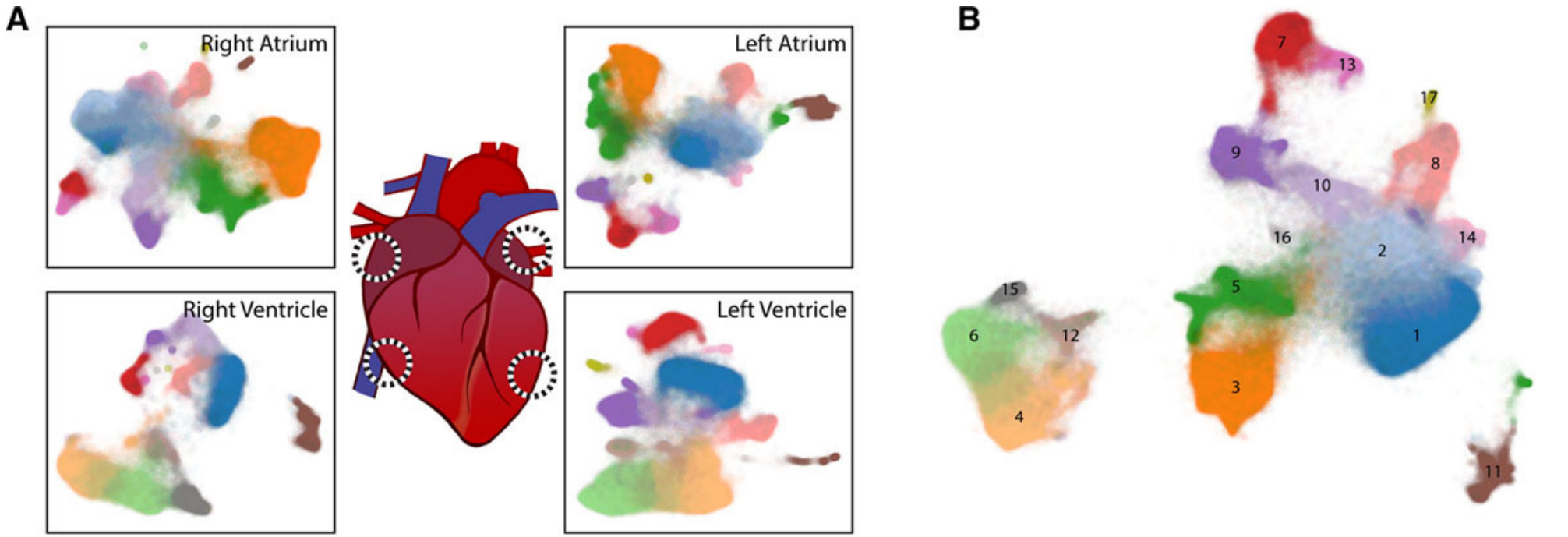
Image credit: Bentsen, Mette, et al. *Nature communications* (2020)

Many thanks to Chuck Epstein and Jason Buenrostro for this slide

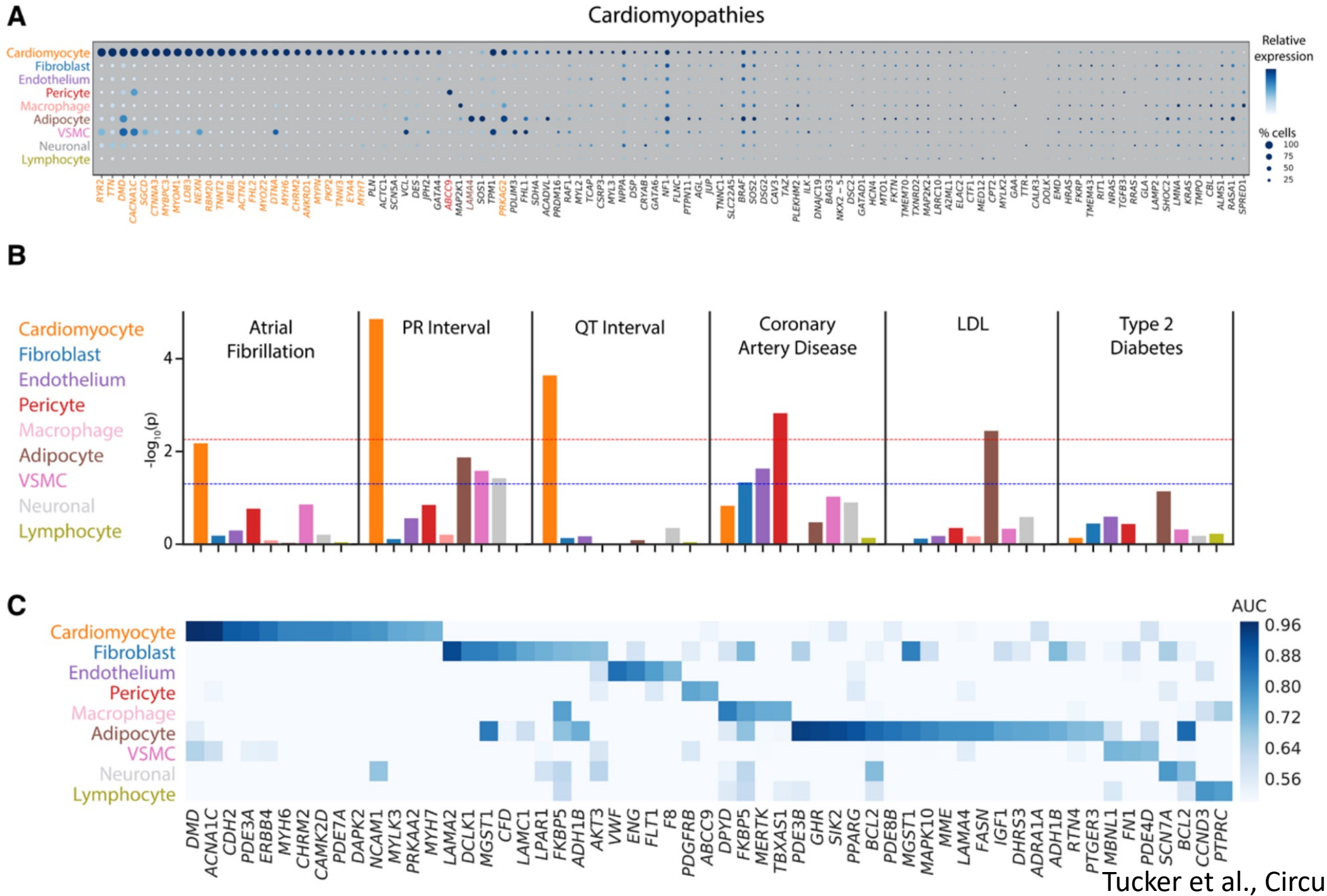
Circulation

ORIGINAL RESEARCH ARTICLE

Transcriptional and Cellular Diversity of the Human Heart



Tucker et al., Circulation 2020



Tucker et al., Circulation 2020

Article

Cells of the adult human heart

https://doi.org/10.1038/s41586-020-2797-4

Received: 10 February 2020

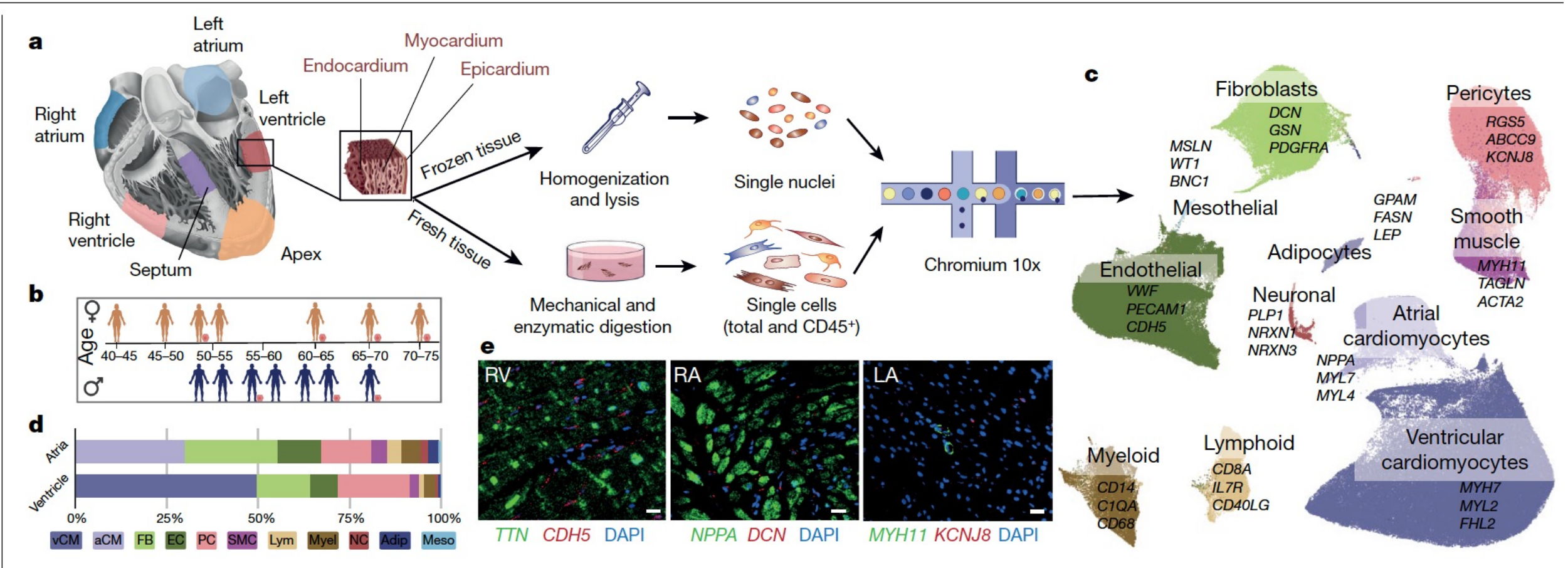
Accepted: 18 September 2020

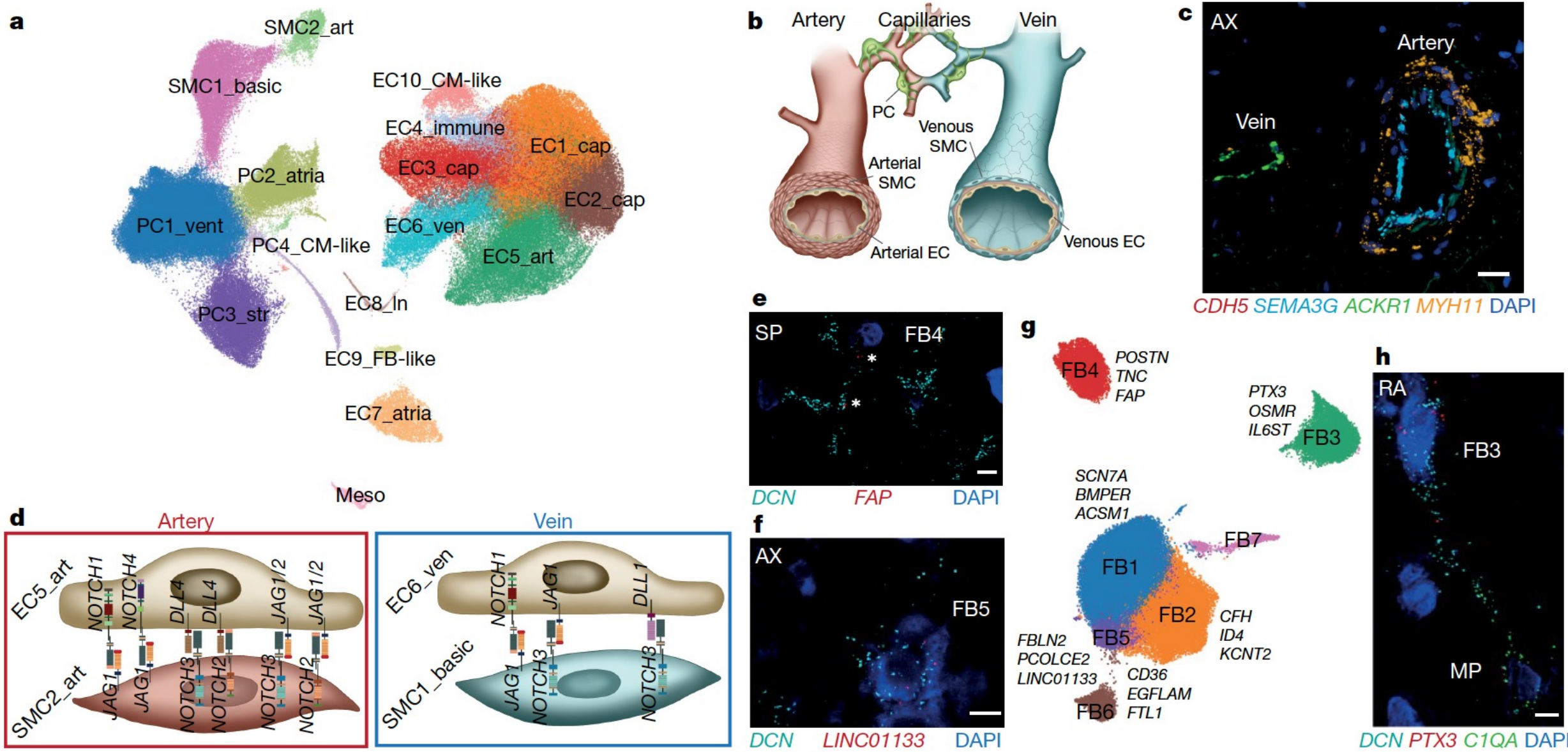
Published online: 24 September 2020

Open access

Check for updates

Monika Litviňuková[1,2,21], Carlos Talavera-López[1,3,21], Henrike Maatz[2,21], Daniel Reichart[4,5,21], Catherine L. Worth[2], Eric L. Lindberg[2], Masatoshi Kanda[2,6], Krzysztof Polanski[1], Matthias Heinig[7,8], Michael Lee[9], Emily R. Nadelmann[4], Kenny Roberts[1], Liz Tuck[1], Eirini S. Fasouli[1], Daniel M. DeLaughter[4], Barbara McDonough[4,11,16], Hiroko Wakimoto[4], Joshua M. Gorham[4], Sara Samari[9], Krishnaa T. Mahbubani[12], Kourosh Saeb-Parsy[12], Giannino Patone[2], Joseph J. Boyle[9], Hongbo Zhang[1,13], Hao Zhang[14,15], Anissa Viveiros[14,15], Gavin Y. Oudit[14,15], Omer Ali Bayraktar[1], J. G. Seidman[4,22]✉, Christine E. Seidman[4,11,16,22]✉, Michela Noseda[9,17,22]✉, Norbert Hubner[2,10,18,19,22]✉ & Sarah A. Teichmann[1,20,22]✉



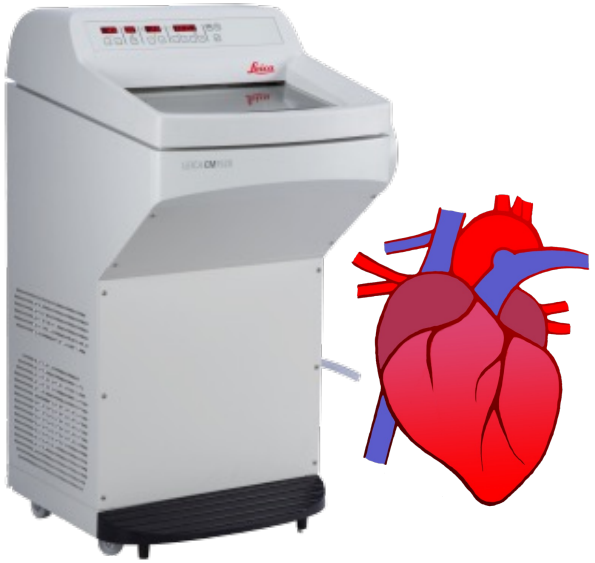


Litvinukova et al., Nature 2020

10x genomics single cell RNA-seq overview

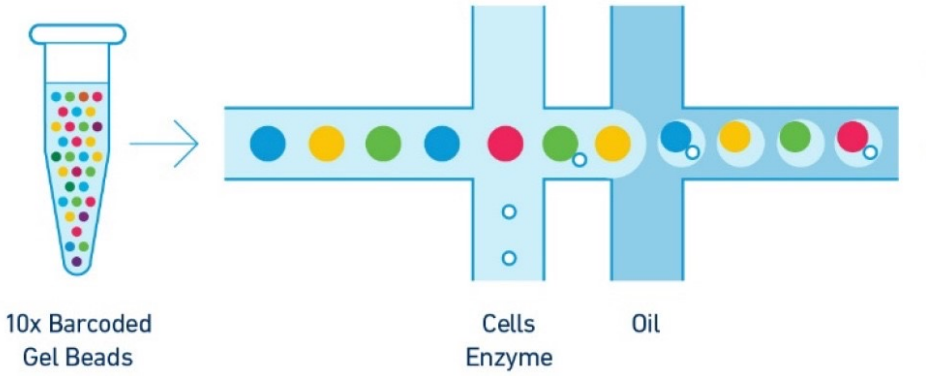
Most user control

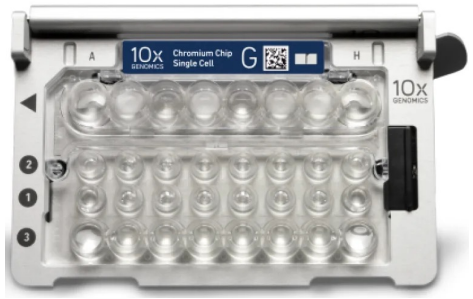
Prepare sample
(isolate cells: fresh tissue,
isolate nuclei: frozen tissue)



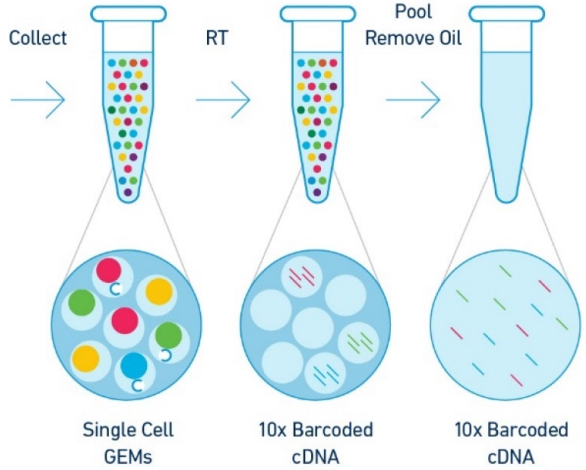
10X genomics protocol control – not much to change

Run sample through chromium controller





Perform RT reaction,
break emulsion



Output standardized

Sequence

some user control

Analysis pipeline

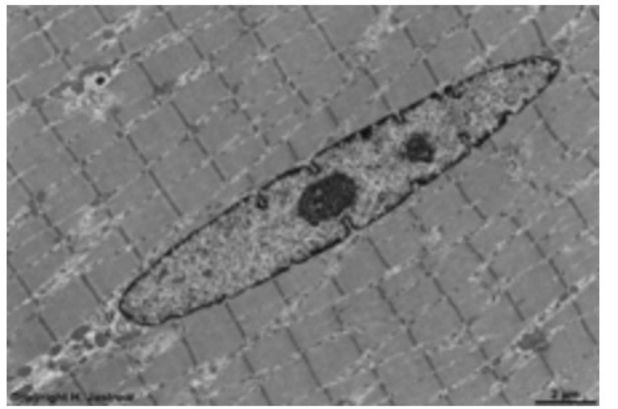
Experimental design best practices: nuclear isolation

- Tissue preparation and nuclear isolation are where you have the most control in this protocol!
  - The Chromium B chip can handle **30um** cells at MOST
- Human nuclei are typically about **10um** in diameter, so nuclear preps always fit
  - And good from frozen tissue
- Debris can clog the chip too, important to have clean nuclear preparations.

Cell type	diameter
PBMC	10um
HeLa	20um
hepatocyte	25-30um
Endothelial cells	50-70um x 10-30um
cardiomyocyte	60-140um x 17-25um
fibroblast	Up to 100um (moving)
nucleus	**10um**

https://bionumbers.hms.harvard.edu/bionumber.aspx

Skeletal muscle nucleus

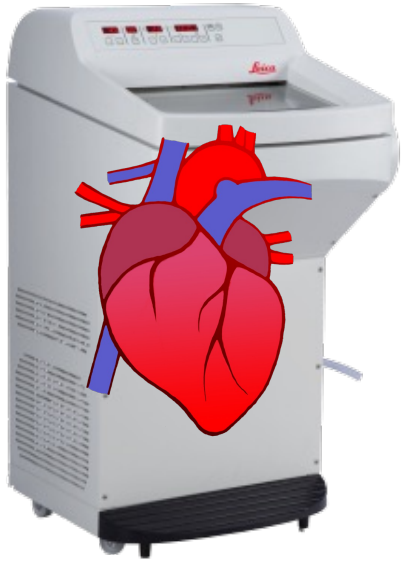


https://kb.10xgenomics.com/hc/en-us/articles/218170543-What-is-the-range-of-compatible-cell-sizes-

http://www.drjastrow.de/WAI/EM/EMKernE.html

Experimental design best practices: nuclear isolation

100um sections



Dounce:
First loose,
then tight pestle

Spin at 40g:
pellet heavy debris

40um filter

10um filter

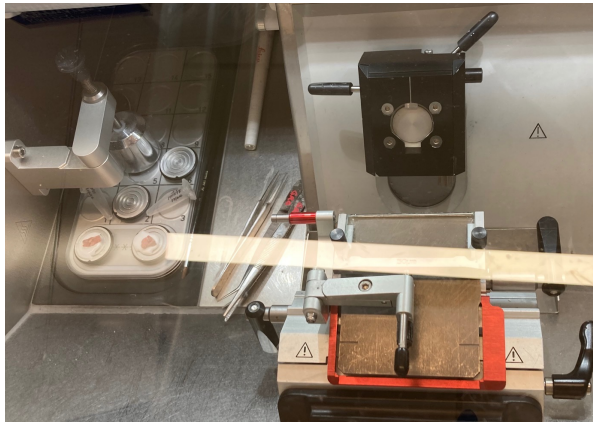
Ultracentrifuge through 2M sucrose gradient
(great for acellular/diseased samples):

45min, 30000xg, 4 deg C

- Sample preparation
  - Fresh tissue? Might be able to isolate cells and freeze viable
  - Frozen tissue? Need to isolate nuclei
    - With vascular/CV derived tissue, sectioning on a cryotome is key
    - **Diseased tissue is often acellular**

Overall workflow of this nuclear isolation protocol

Section tissue in cryostat
100uM sections
6-20 sections/tube



Wide bore triturate
Dounce loose/tight

Crude nuclear prep!!

Spin 40-100g to remove debris

3000xg 4 deg C
20 minutes

Pass supe through filters (40uM,10uM)

10^6 nuclei/tube

Spin to pellet nuclei at 550xg

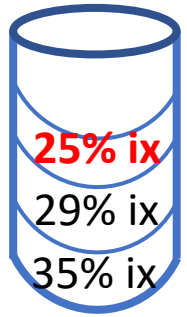
Count and image
PI/trypan

Iodixanol gradient

0.2% fixation
5 minutes RT

Fix and freeze

Count and image
PI/trypan



Nuclei band
At 29%:35%
interface

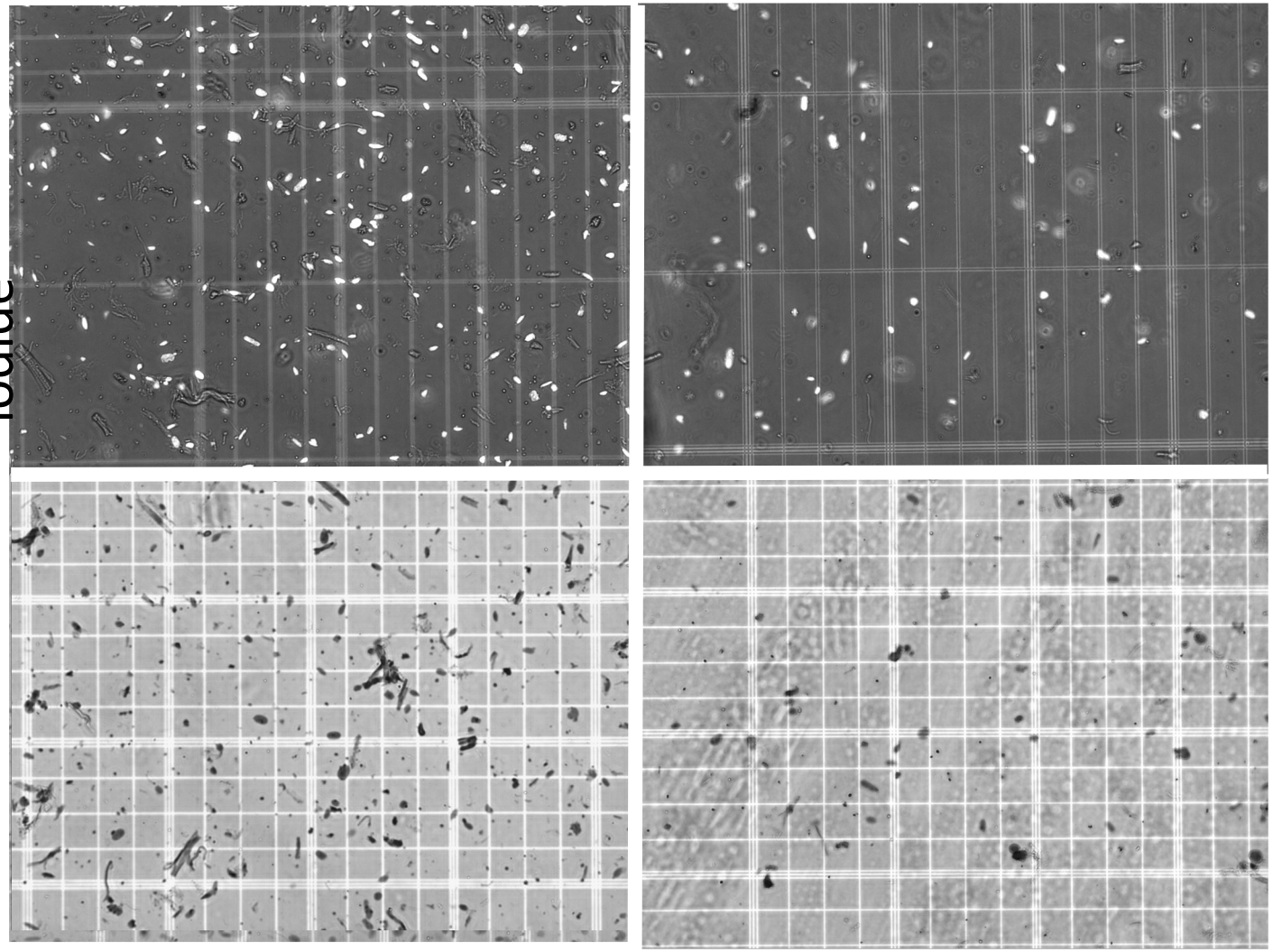
Nuclei are in 25% ix
In NSB buffer

Heart Left Ventricle Nuclear Prep

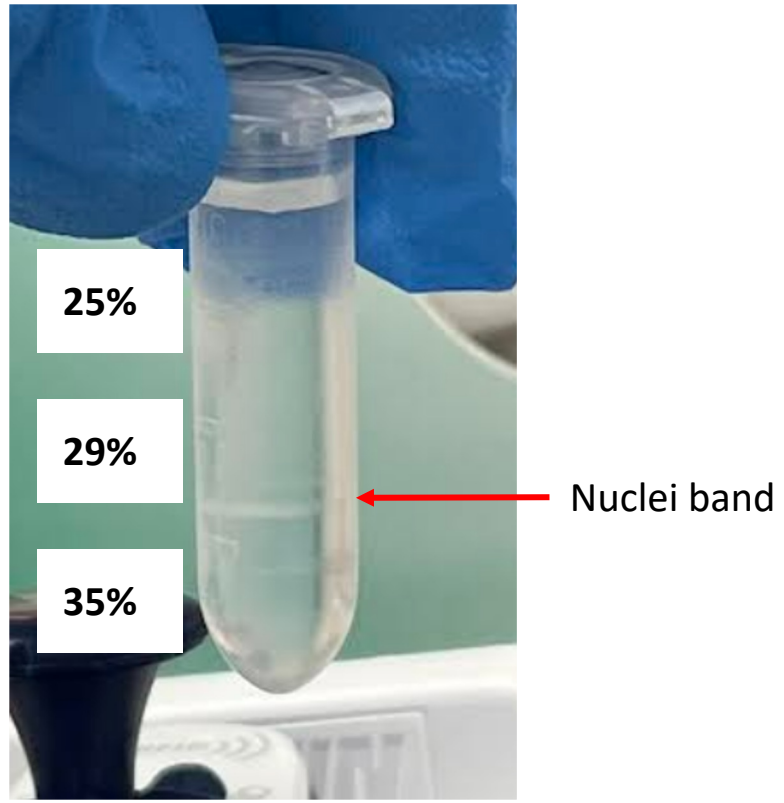
10x pre-iodixanol post-iodixanol

+propidium iodide

+trypan



Pre-iodixanol gradient: 1.2M nuclei
Post-iodixanol gradient: .94M nuclei



IX gradient from a colon prep: shown as an example

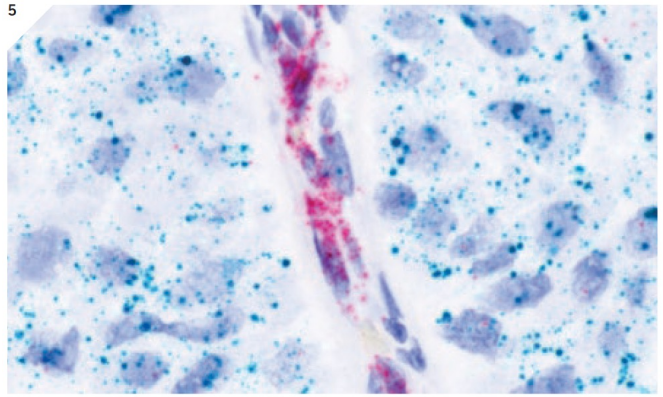
Experimental design best practices: tissue validation

- Identify a novel marker gene using computational analyses?
  - You need to be prepared to validate!!

RNA ISH (RNAscope)

Oligo + fluorophore staining

Breast cancer slide



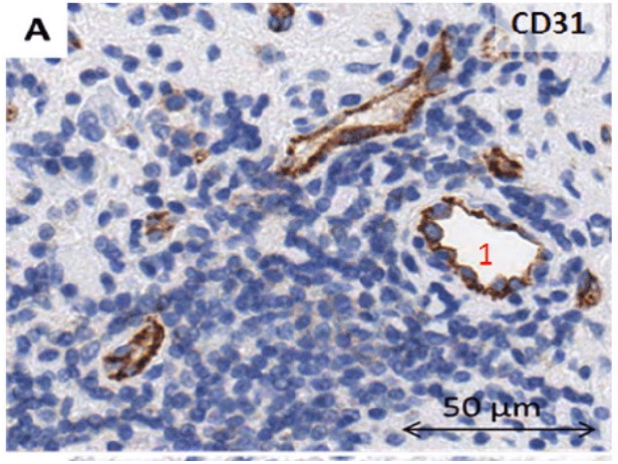
EPCAM in red
EGFR in green

Image via ACB

IHC

immunohistochemistry

Neuroblastoma slide

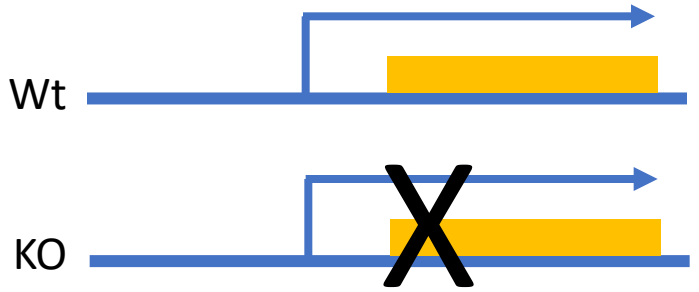


CD31 in brown
Counterstain in blue

Tadeo et al., 2016

In vitro models

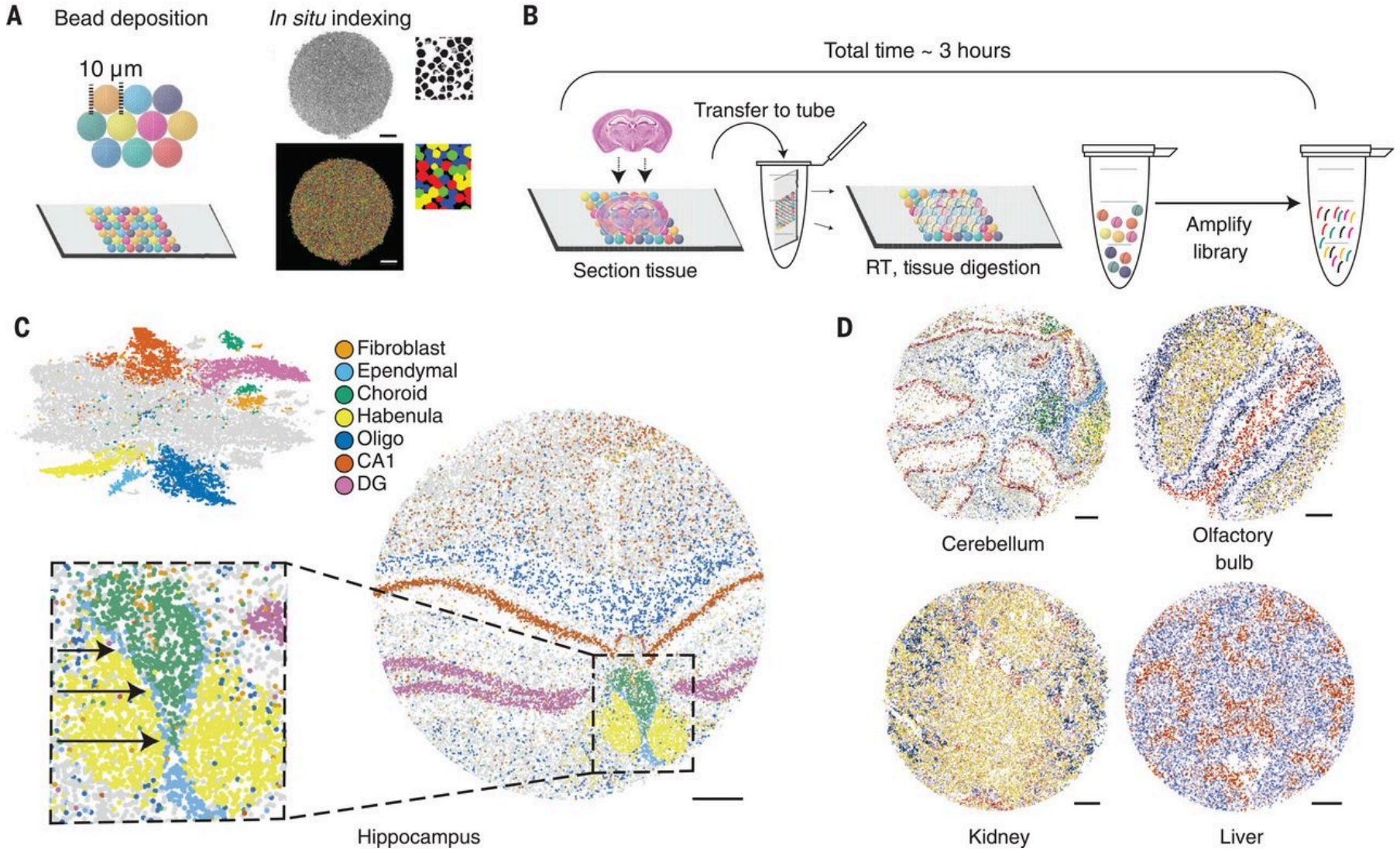
CRISPR KO (cells)
Plus imaging



Need a good *in vitro* cell line
iPSC cells from patients possible

Can also do EP to show effect in cell types

Experimental design best practices: tissue validation



SLIDE-seq

Also see:

MERFISH

Visium

Sci-space

https://www.science.org/doi/full/10.1126/science.aaw1219

Sequencing structure for 10x multiome

R1: 28-50bp
i7: 8-10bp
i5: 10-24bp
R2: 49-90bp

mRNA

Dual Indexed Sequencing Run: Single Cell Multiome Gene Expression libraries are dual-indexed. This means both i5 and i7 reads are used for demultiplexing. We do not recommend sequencing 10x Single Cell Multiome Gene Expression dual index libraries with a single-index configuration.

Read	Read 1	i7 Index	i5 Index	Read 2
Purpose	Cell barcode & UMI	Sample Index	Sample Index	Insert
Length**	28	10	10	90

ATAC

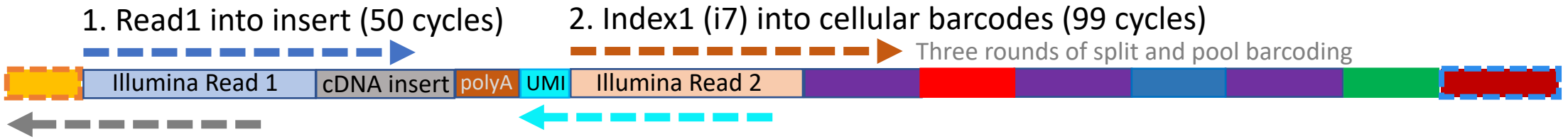
Read	Read 1	i7 Index	i5 Index	Read 2
Purpose	Transposed DNA	Sample Index	10x Barcode	Transposed DNA
Length	50 **	8	24†	49 **

Don’t combine ATAC and mRNA on the same flowcell!!

Sequencing structure for SHARE

R1: 50bp
i7: 99bp
i5: 8bp
R2: 50bp

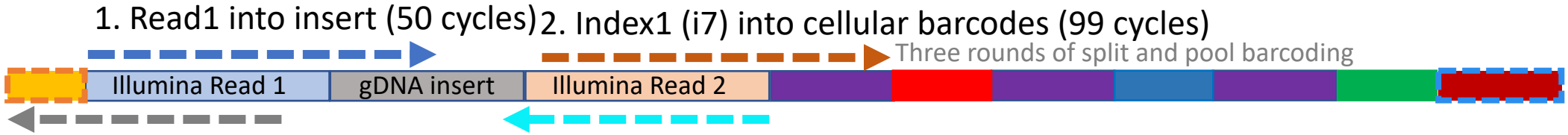
mRNA



3. Index2 to get i5 (library) index (8 cycles)

4. Read2 to get **UMI** (50 cycles)

ATAC



3. Index2 to get i5 (library) index (8 cycles)

4. Read2 to get **paired** ATAC-seq (50 cycles)

Don’t combine ATAC and mRNA on the same lane!!

In brief: sequencing processing pipeline

BCL files (output from sequencer)

Walkupseq submission of library pool to GP

Demultiplex to get fastq files

Align and filter ATAC (bowtie2)

Peak calling

Peak bins X cell barcode matrix

Clustering & annotation

ATAC UMAP

Peaks to genes DORCs

Align and filter RNA (STARsolo)

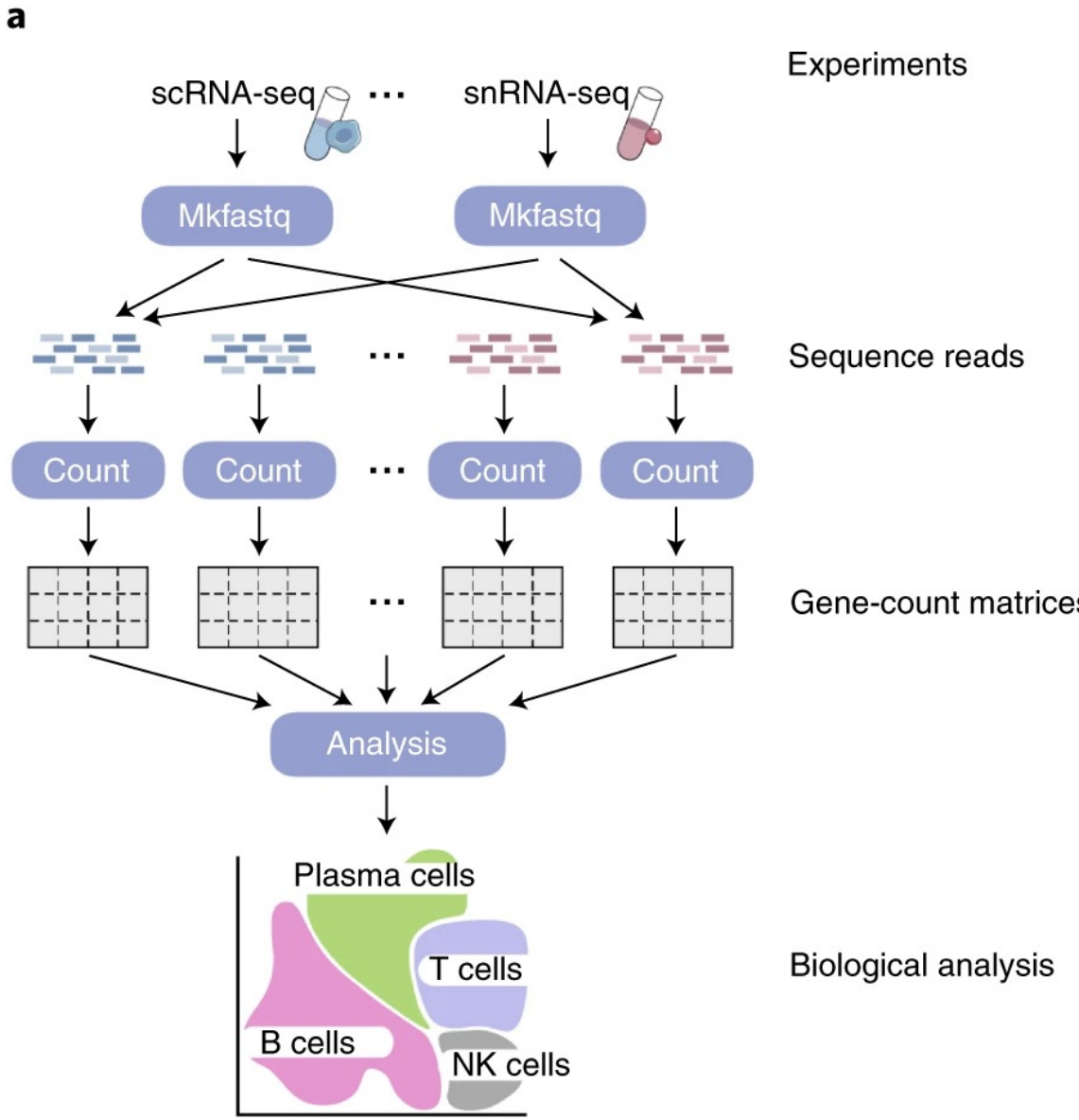
Gene X cell barcode matrix

Clustering & annotation

RNA UMAP

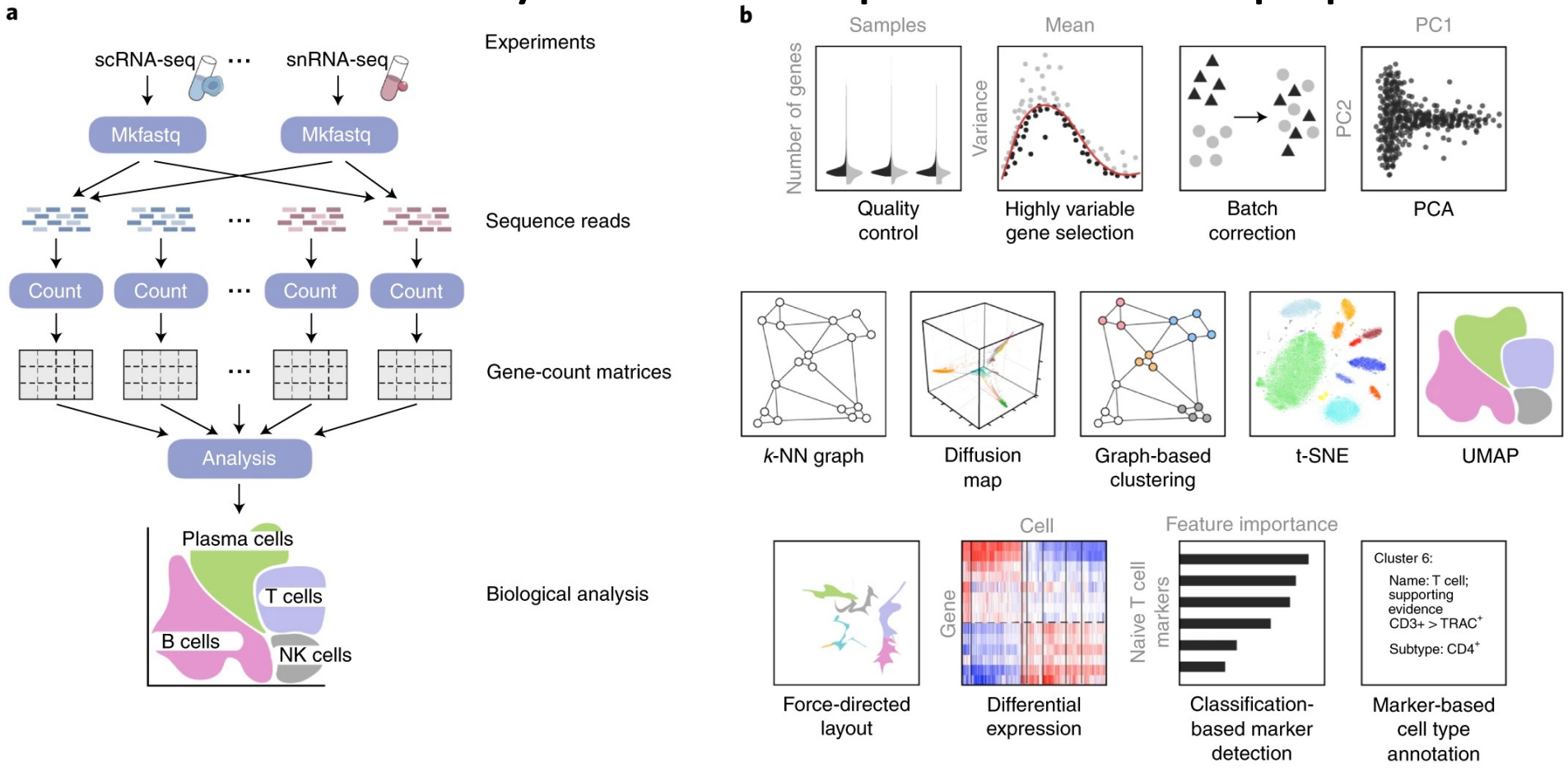
Slide derived from pipeline presentations by Eugenio Mattei

Computational analysis best practices: pipelines



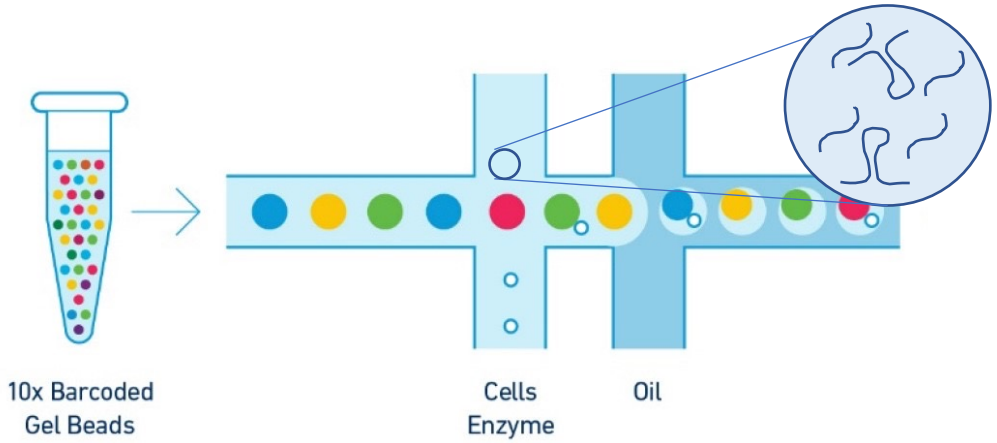
Li et al., Nature Methods, 2020

Computational analysis best practices: pipelines

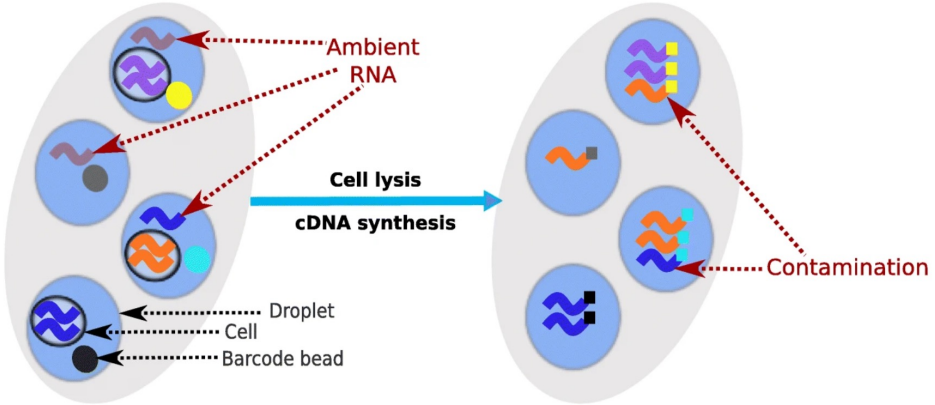


Li et al., Nature Methods, 2020

Computational analysis best practices: ambient RNA



- VERY important for single nucleus RNA-seq
  - Some cytoplasmic material can remain, or organelles attached to the nucleus
  - This leads to some ambient RNA contaminating the aqueous solution around the nucleus + gel bead

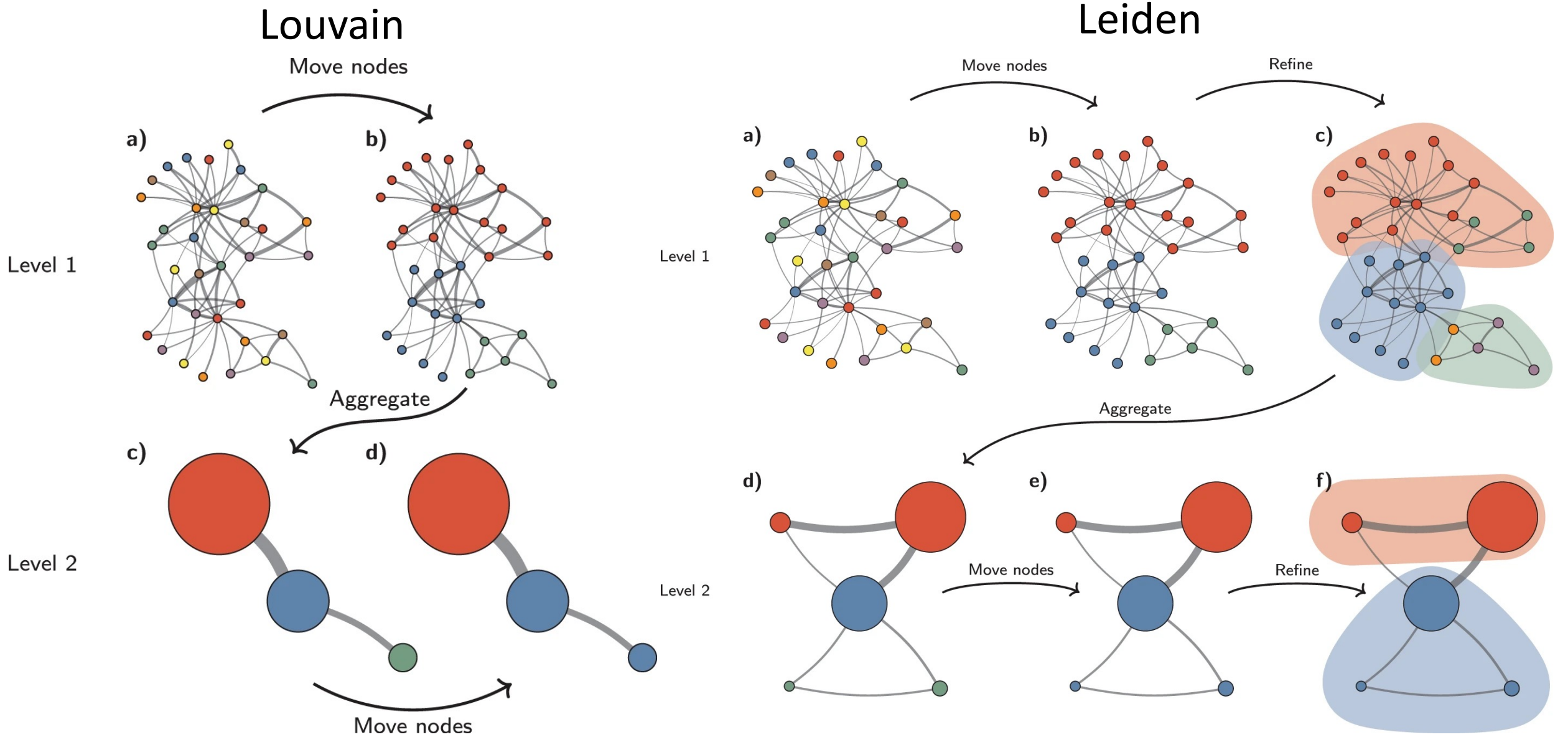


Many methods available to handle this:

Cellbender: https://github.com/broadinstitute/CellBender
SoupX: https://github.com/constantAmateur/SoupX
ScVI: https://docs.scvi-tools.org/en/stable/
DeconX: https://github.com/campbio/celda

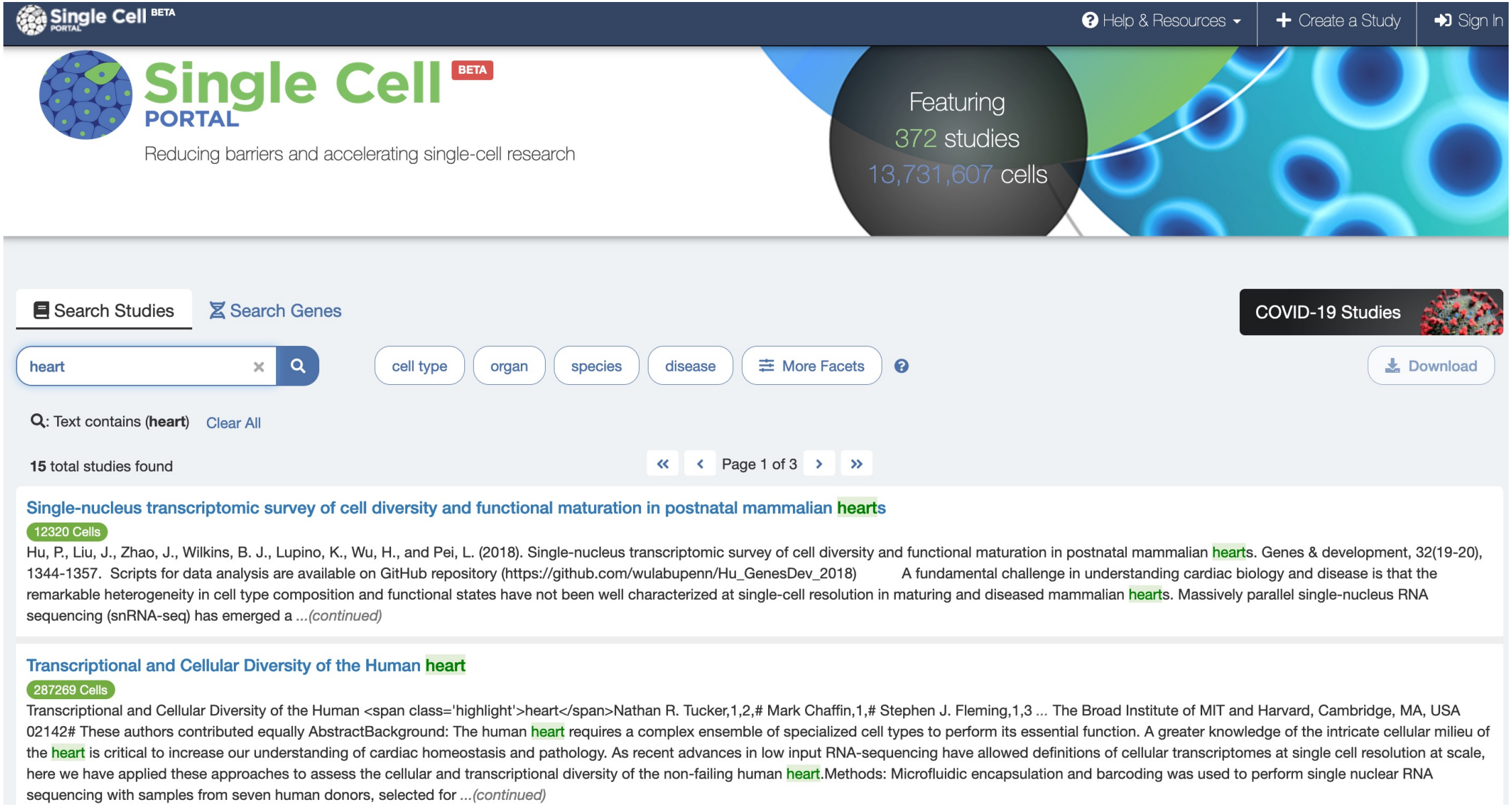
Yang et al., Genome Biology 2020.https://genomebiology.biomedcentral.com/articles/10.1186/s13059-020-1950-6

Computational analysis best practices: clustering



Traag et al., Nature Scientific Reports 2019

Resources: The single cell portal at the Broad Institute



https://singlecell.broadinstitute.org/single_cell

Resources: Azimuth (Satija lab)

Azimuth is a web application that uses an annotated reference dataset to **automate the processing, analysis, and interpretation of a new single-cell RNA-seq experiment**. Azimuth leverages a **'reference-based mapping'** pipeline that inputs a counts matrix of gene expression in single cells, and performs normalization, visualization, cell annotation, and differential expression (biomarker discovery). All results can be explored within the app, and easily downloaded for additional downstream analysis.

The development of Azimuth is led by the New York Genome Center Mapping Component as part of the NIH Human Biomolecular Atlas Project (HuBMAP). Seven molecular reference maps are currently available, with more coming soon.

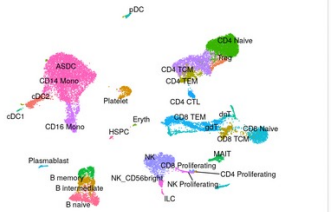
NOW WITH SINGLE CELL ATAC-seq DATA!!!

https://azimuth.hubmapconsortium.org/

Resources: Azimuth (Satija lab)

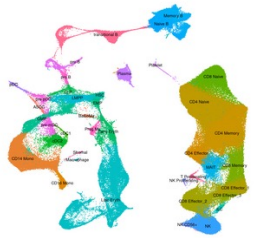
References for scATAC-seq Queries

Human - PBMC



Learn more | Go to App

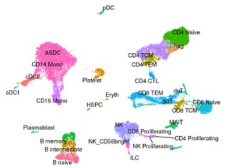
Human - Bone Marrow



Learn more | Go to App

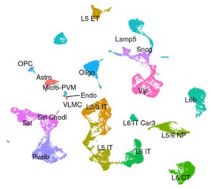
References for scRNA-seq Queries

Human - PBMC



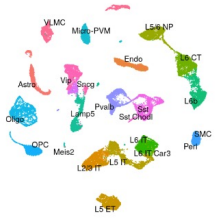
Learn more | Go to App

Human - Motor Cortex



Learn more | Go to App

Mouse - Motor Cortex

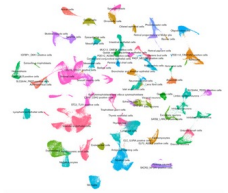


Learn more | Go to App

Human - Pancreas

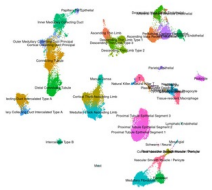
Learn more | Go to App

Human - Fetal Development



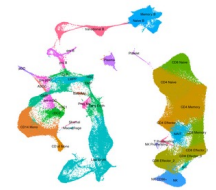
Learn more | Go to App

Human - Kidney



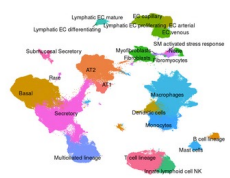
Learn more | Go to App

Human - Bone Marrow



Learn more | Go to App

Human - Lung v2 (HLCA)



Learn more | Go to App

Human - Tonsil

Learn more | Go to App

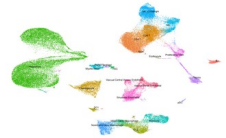
Human - Adipose

Learn more | Go to App

Human - Heart

Learn more | Go to App

Human - Liver

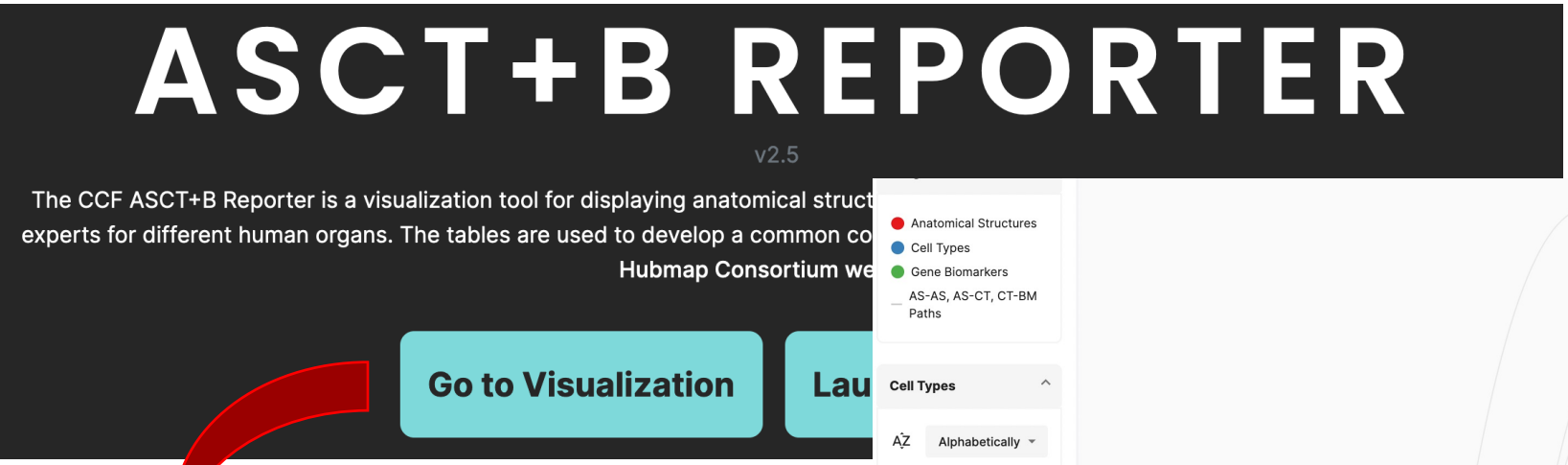


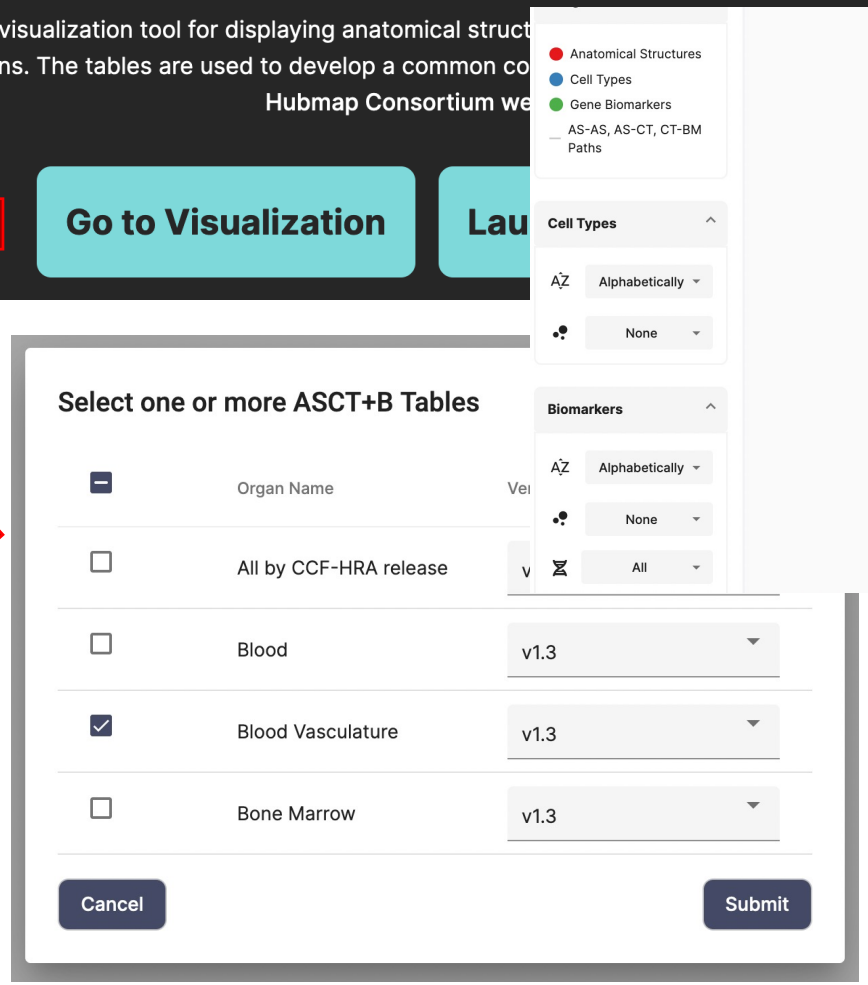
Learn more | Go to App

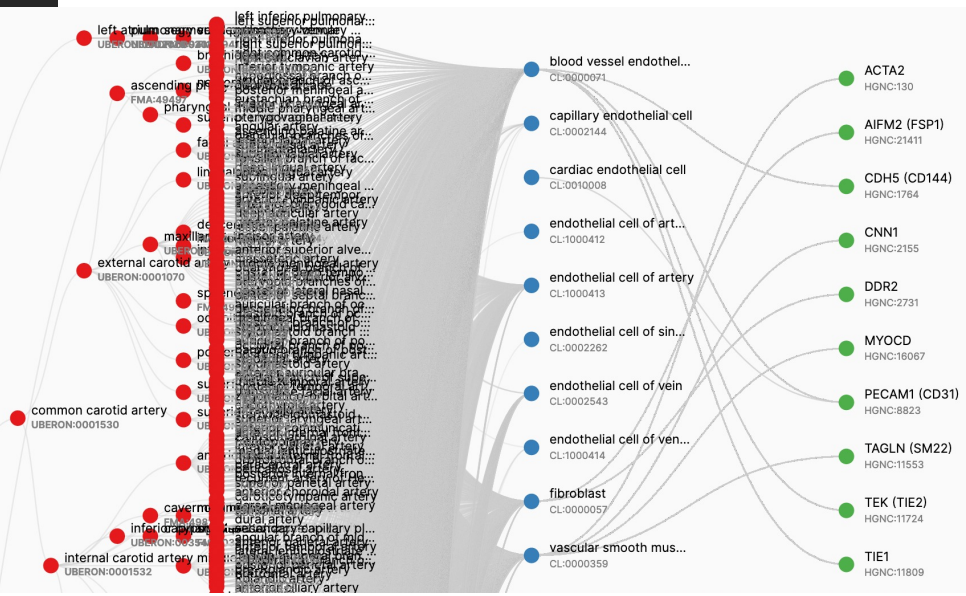
NOW WITH SINGLE CELL ATAC-seq DATA!!!

https://azimuth.hubmapconsortium.org/

Resources: Hubmap! (validated cell type markers!)

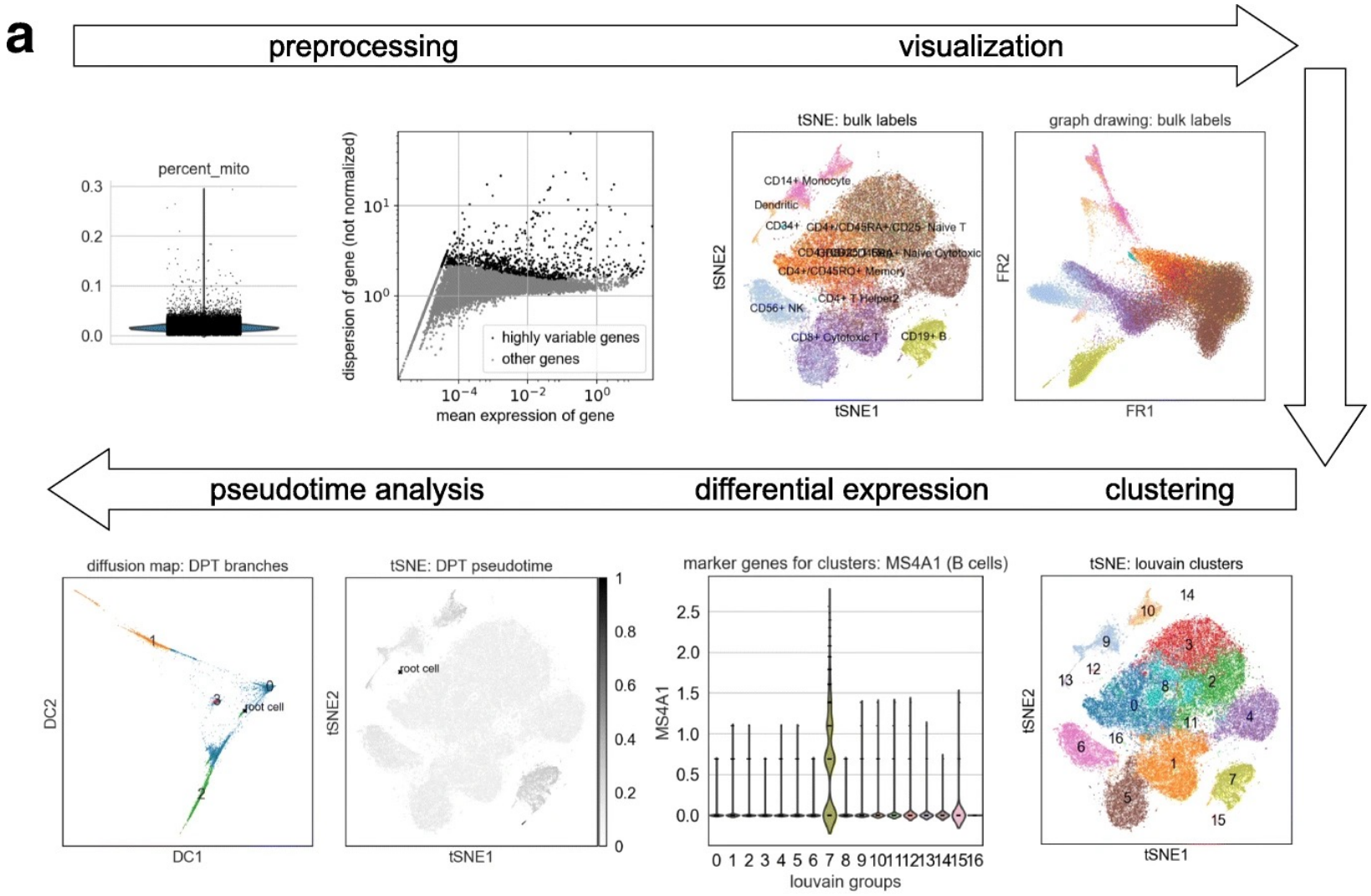


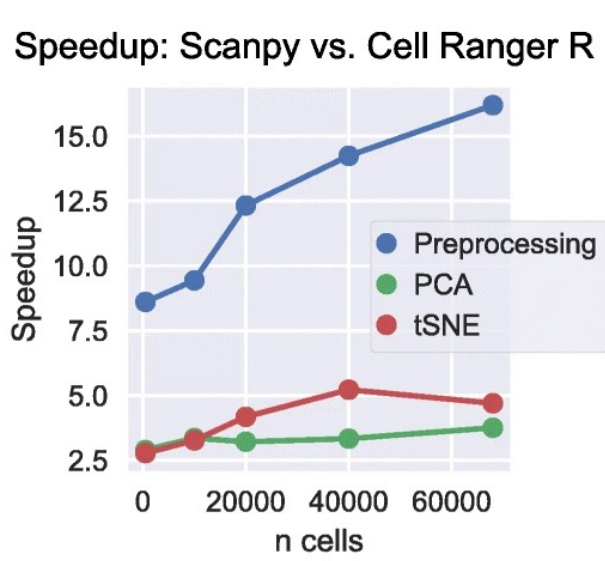


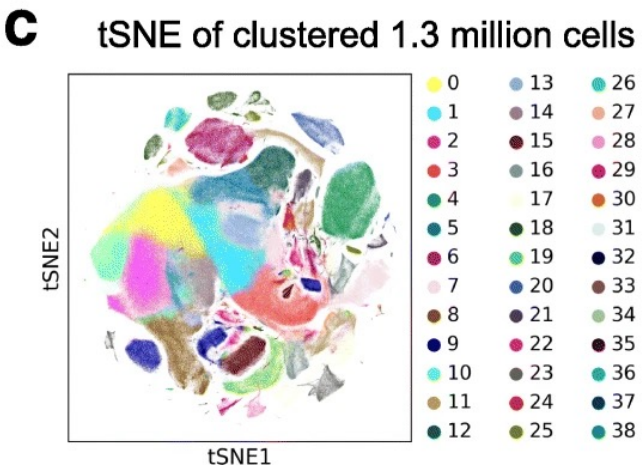


https://hubmapconsortium.github.io/ccf-asct-reporter/

Packages: Scanpy (python)

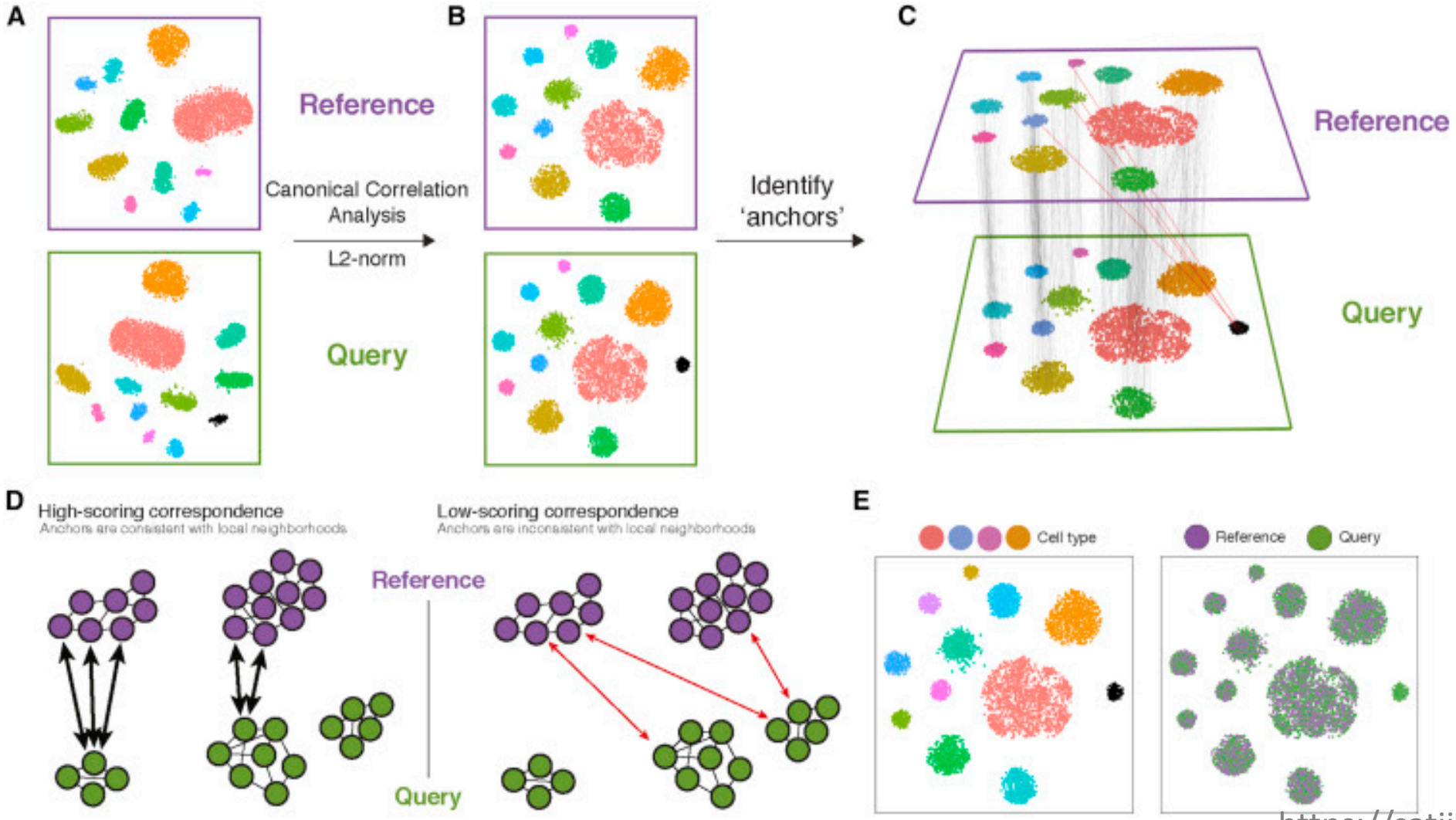






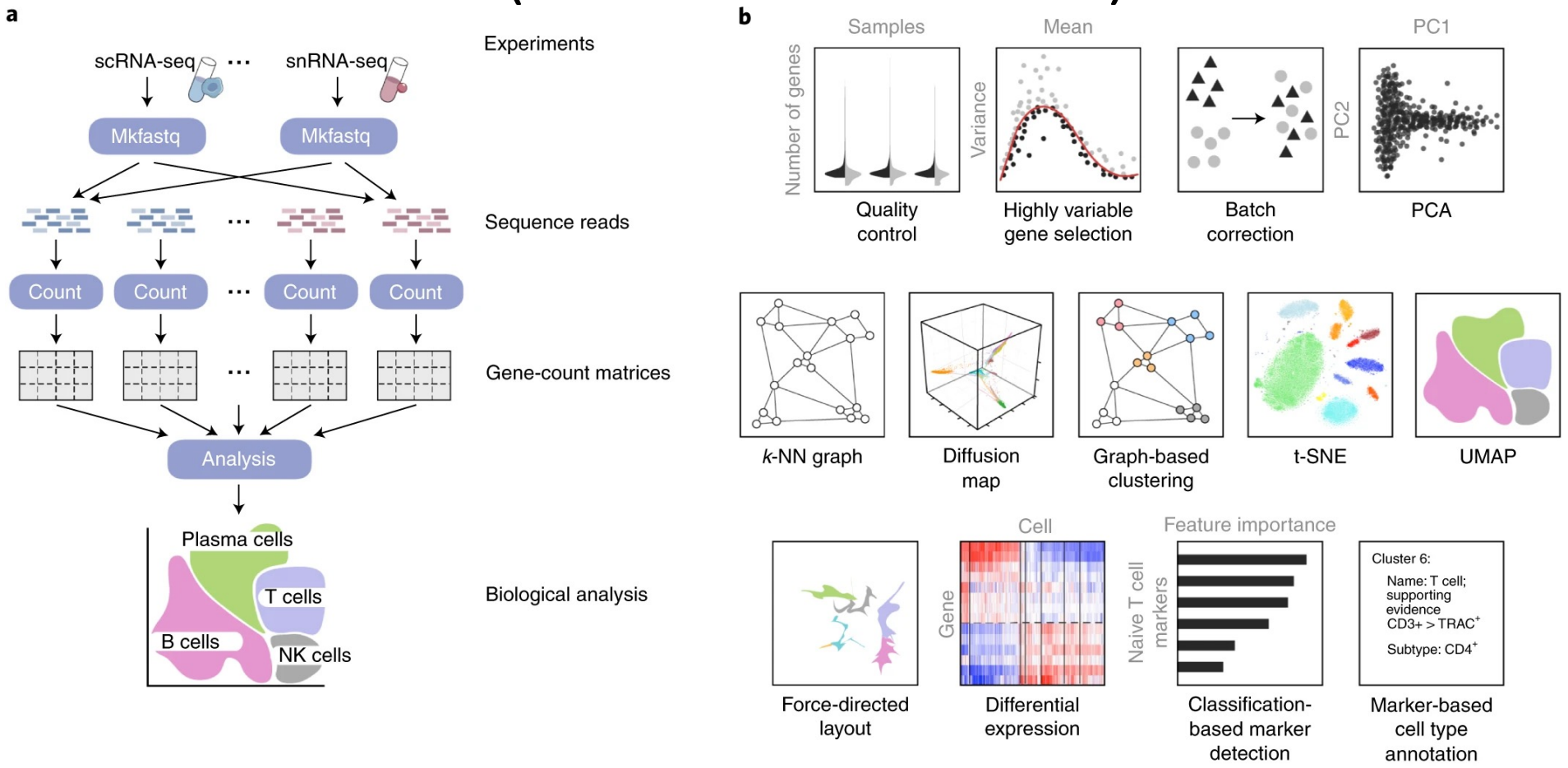
https://scanpy.readthedocs.io/en/stable/

Packages: Seurat (R)



https://satijalab.org/seurat/

Pipelines: Cumulus (Terra workflow)



https://cumulus.readthedocs.io/en/stable/index.html

Li et al., Nature Methods, 2020

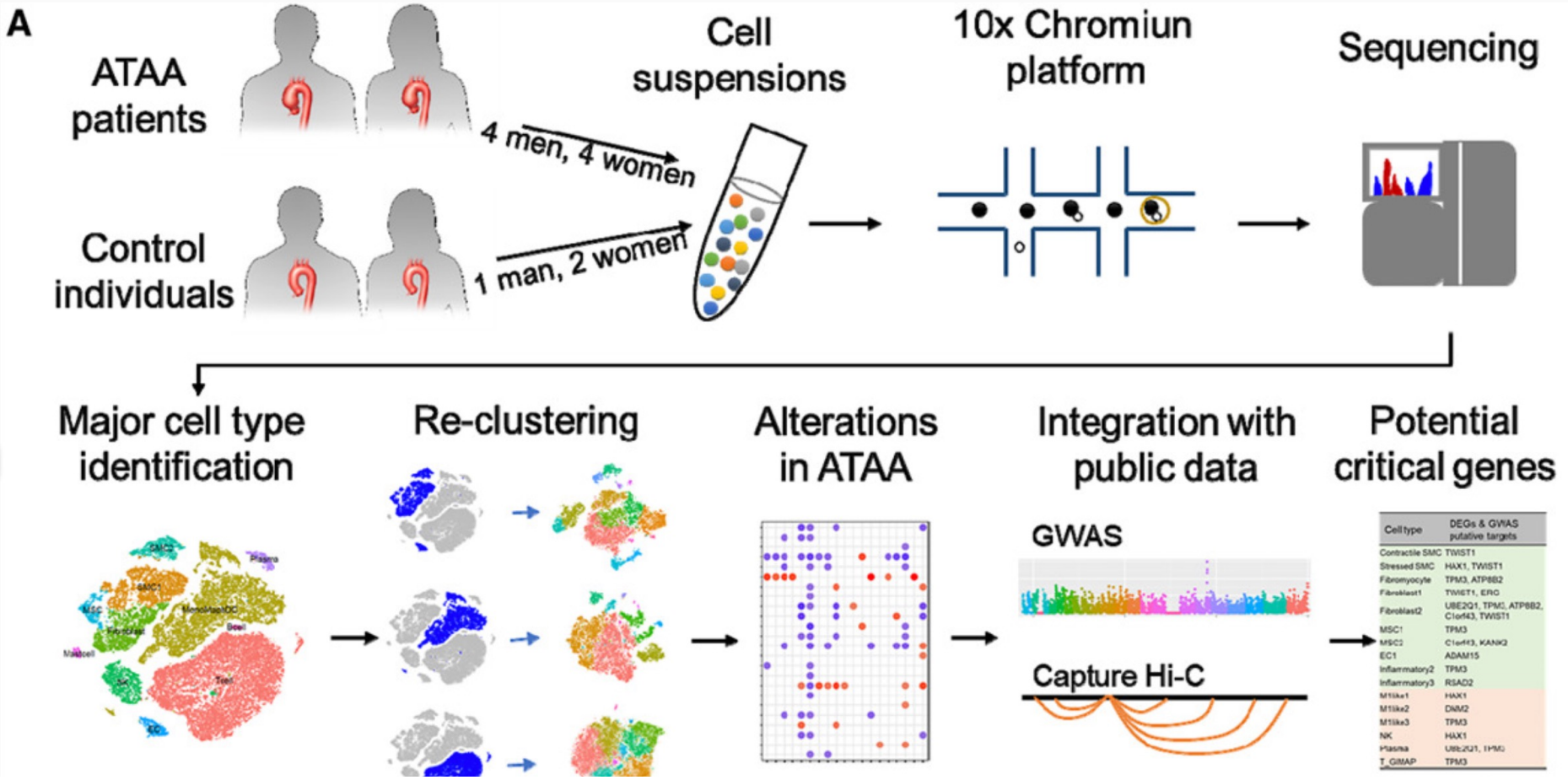
Limitations of 10x/single nucleus RNA-seq

- mRNA is nuclear, so intermediate in spliceosome processing
  - No isoform determination
  - Must map to entire gene body
- RNA is specifically nascent, non-cytoplasmic
  - Limiting factor for some trajectory analysis
- Need good quality isolated nuclei to avoid clogging 10x chip
  - For some tissues, debris removal is necessary to isolate intact nuclei
    - Iodixanol, ultracentrifugation
- For multi-omic assays, it can be hard to balance ATAC-seq quality with RNA-seq quality

Single-Cell Transcriptome Analysis Reveals Dynamic Cell Populations and Differential Gene Expression Patterns in Control and Aneurysmal Human Aortic Tissue

Yanming Li, Pingping Ren, Ashley Dawson, Hernan G. Vasquez, Waleed Ageedi, Chen Zhang, Wei Luo, Rui Chen, Yumei Li, Sangbae Kim, Hong S. Lu, Lisa A. Cassis, Joseph S. Coselli, Alan Daugherty, Ying H. Shen ✉, Scott A. LeMaire ✉

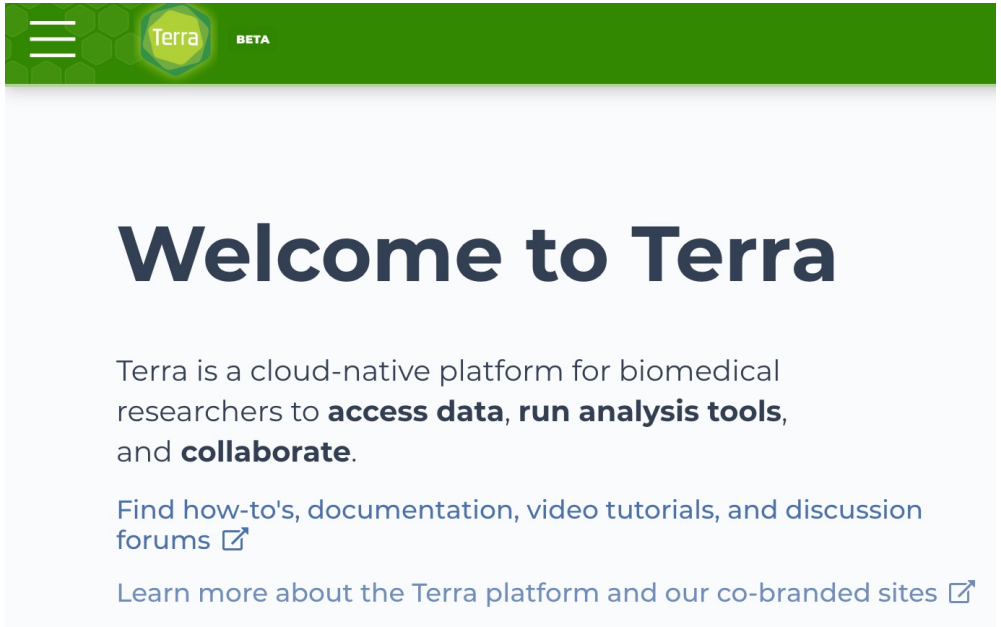
Originally published 5 Oct 2020 | https://doi.org/10.1161/CIRCULATIONAHA.120.046528 | Circulation. 2020;142:1374–1388

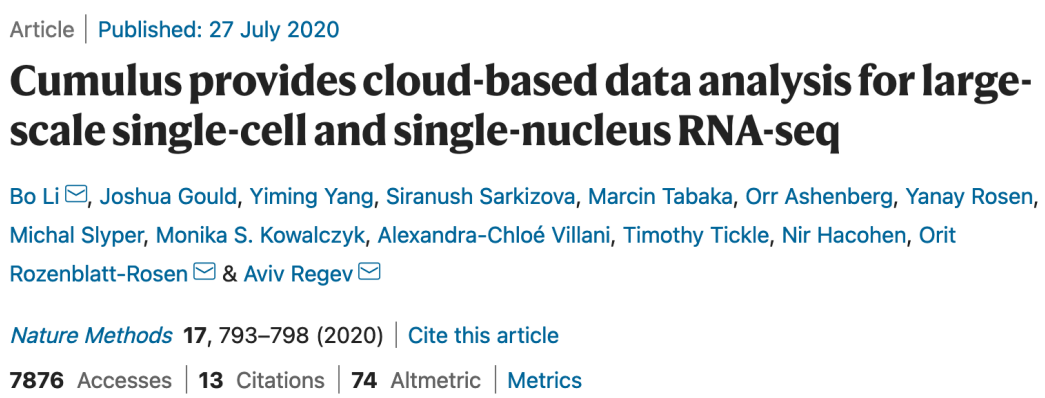


https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.120.046528

Preview/overview of the coding session tomorrow

- We will be using a workspace on Terra →
  - https://app.terra.bio/#workspaces/gro-share-seq-computational/useful_notebooks_for_teaching_public/analysis/launch/2021_09_human_aorta_single_cell_cumulus_scanpy.ipynb
  - We will be examining the human aorta dataset on human aorta reprocessed using the pipeline Cumulus
- We also have a single cell ArchR workspace for ATAC analysis
  - https://app.terra.bio/#workspaces/gro-share-seq-computational/useful_notebooks_for_teaching_public/analysis/launch/2022_02_25_ArchR_tutorial_atac.ipynb





Thank you! And see you tomorrow!

- Many thanks to:
  - The AHA GPM Council
  - **Lisa Wilsbacher**
  - **Matt Nayor**
  - Charles Epstein, for letting me take time to teach

American Heart Association®

Vascular Discovery:
From Genes to Medicine 2023