Single Cell RNA-seq

Amelia Weber Hall

February 2020

Computational Workgroup

Outline

- History of the method, older methods
- 10X genomics single cell RNA-seq
  - How it works
  - What it's good for
- Computational processing of single cell RNA-seq data
  - Overview: challenges of processing this data
  - Complete processing pipelines
    - Cellranger/cellbender
    - Kallisto bustools
    - Salmon alevin
  - Downstream processing pipelines
    - Seurat (Satija lab)
    - Scanpy (with interactive tutorial)

History of the method: 1992!

› Proc Natl Acad Sci U S A, 89 (7), 3010-4 1992 Apr 1

Analysis of Gene Expression in Single Live Neurons

J Eberwine [1], H Yeh, K Miyashiro, Y Cao, S Nair, R Finnell, M Zettel, P Coleman

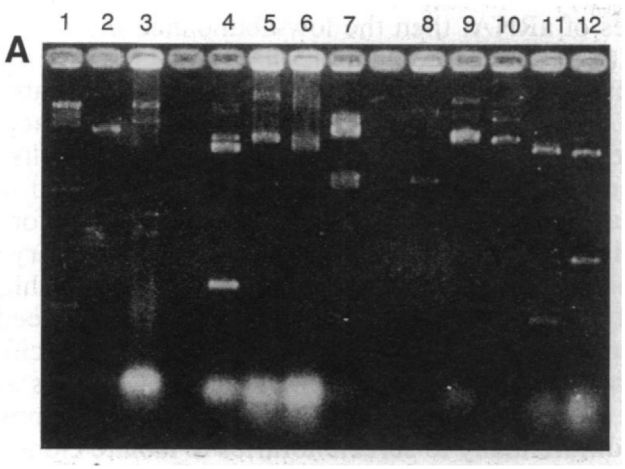
Affiliations + expand

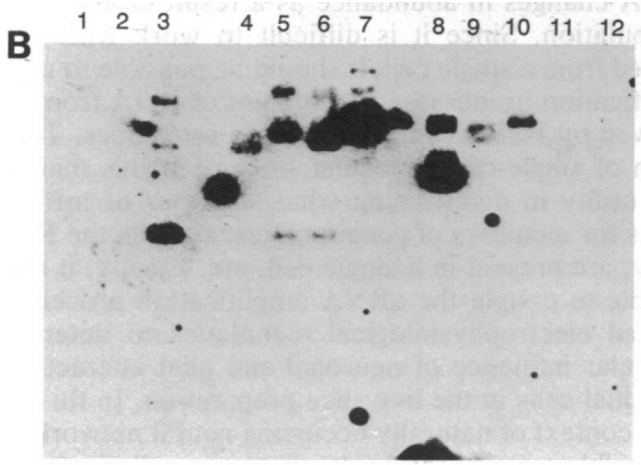
PMID: 1557406 PMCID: PMC48793 DOI: 10.1073/pnas.89.7.3010

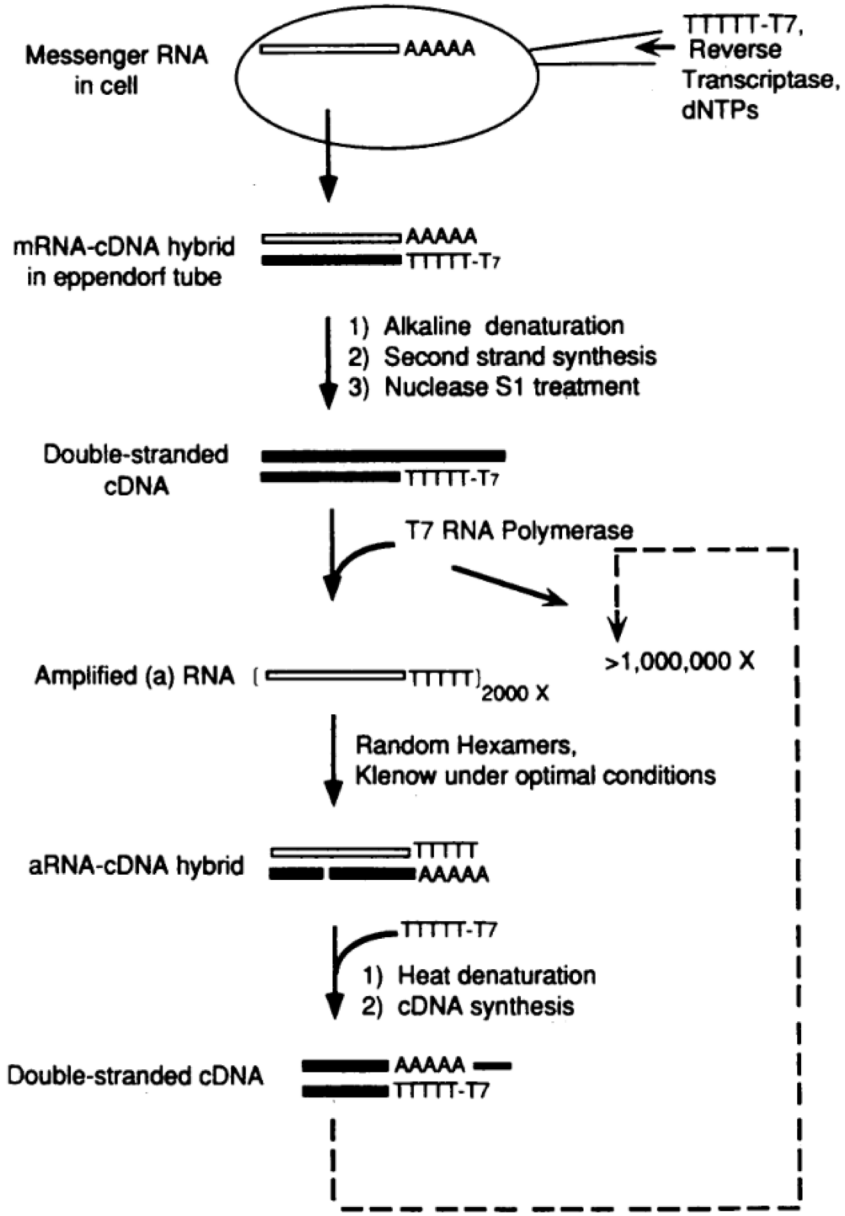
Electrophysiological studies of dissociated rat neurons:

patched the neurons, then did cDNA synthesis ***in the patch electrode***

Looked at just a few genes/transcripts, obviously not NGS at the time







History of the method: NGS (2009)

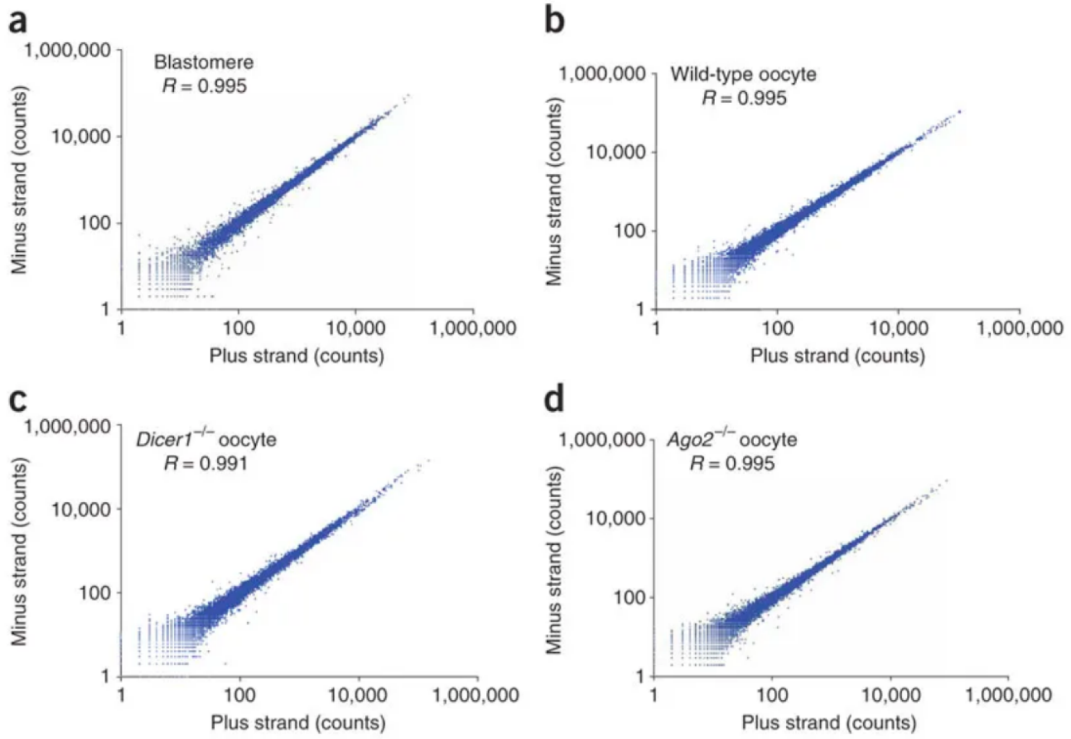
› Nat Methods, 6 (5), 377-82 May 2009

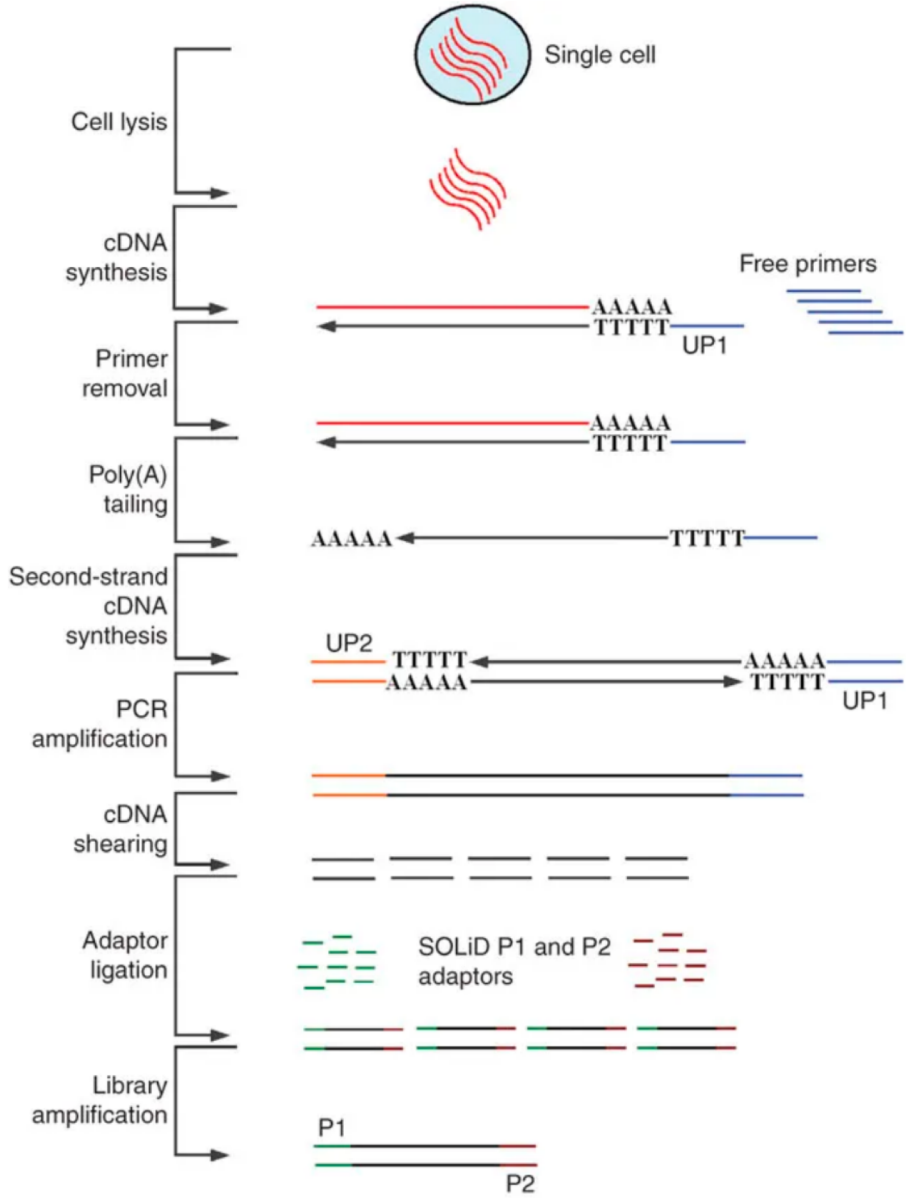
mRNA-Seq Whole-Transcriptome Analysis of a Single Cell

Fuchou Tang [1], Catalin Barbacioru, Yangzhou Wang, Ellen Nordman, Clarence Lee, Nanlan Xu, Xiaohui Wang, John Bodeau, Brian B Tuch, Asim Siddiqui, Kaiqin Lao, M Azim Surani

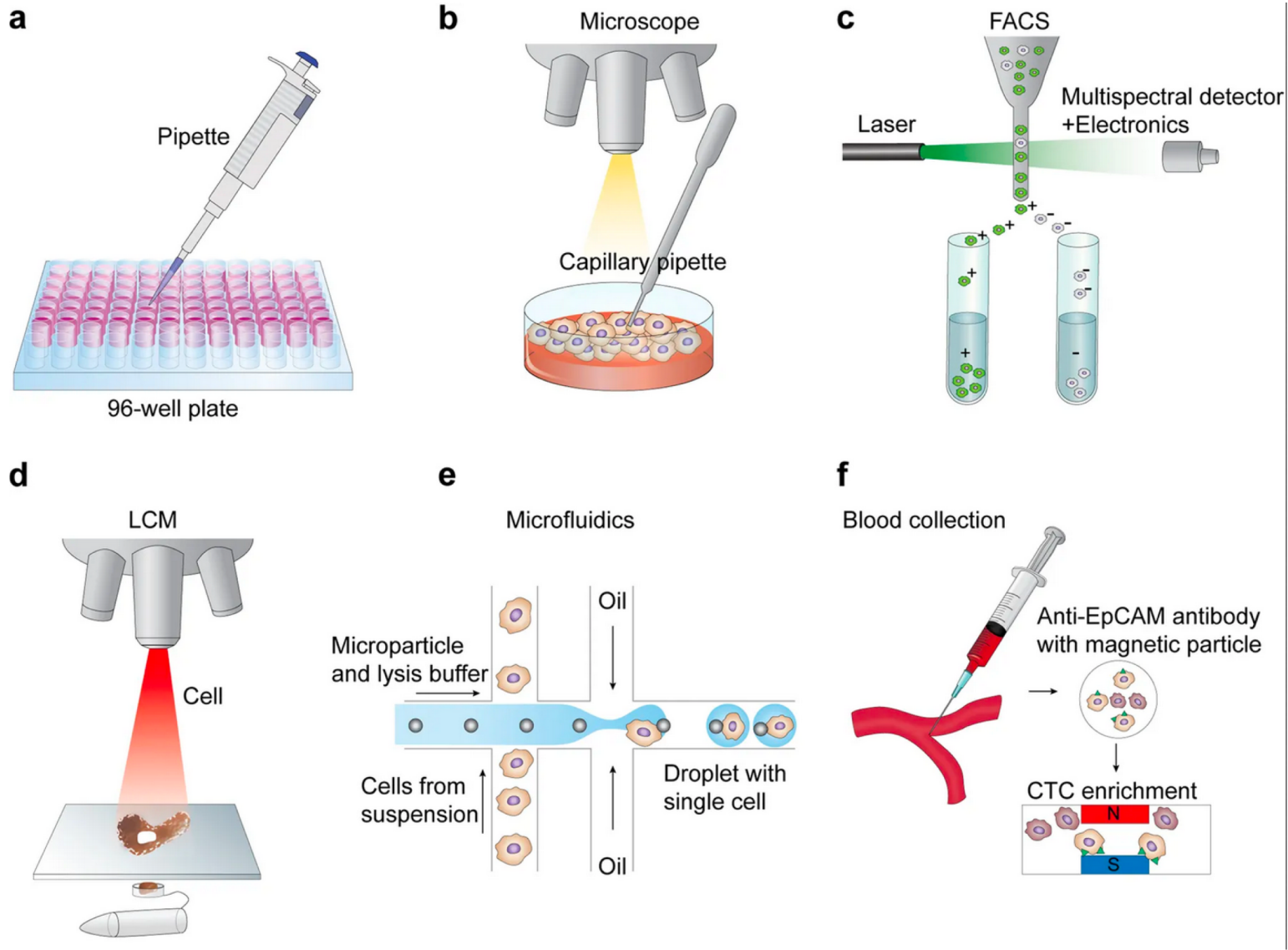
Affiliations + expand

PMID: 19349980 DOI: 10.1038/nmeth.1315





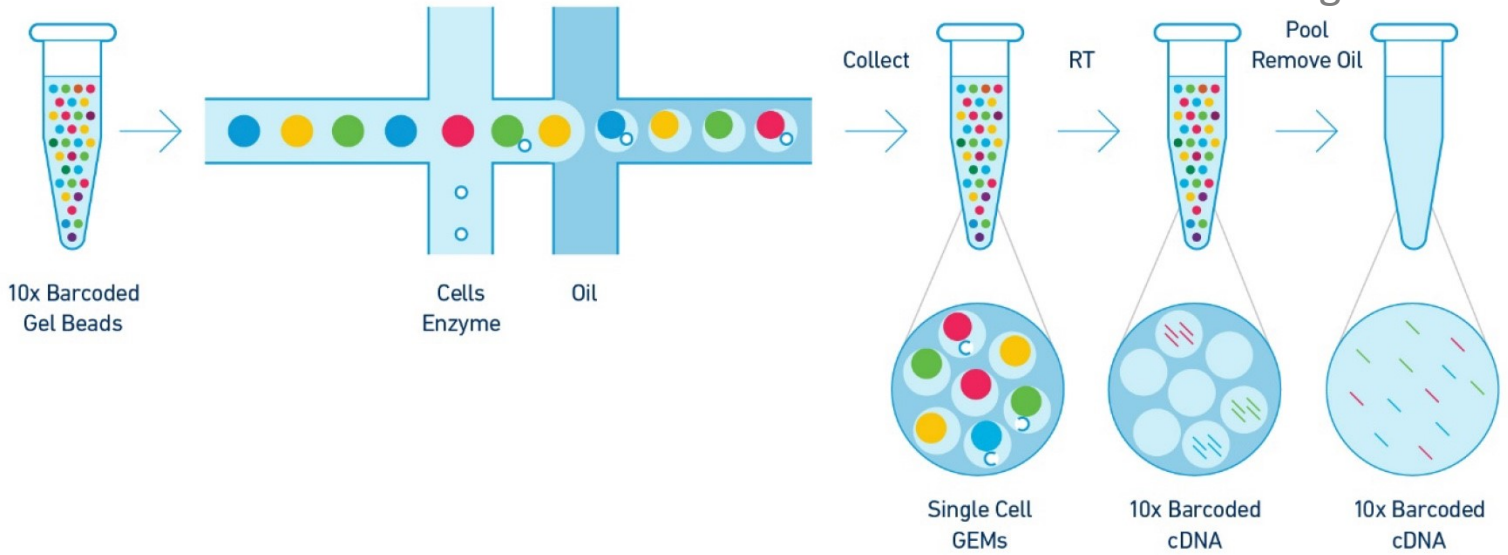
Cell isolation methods

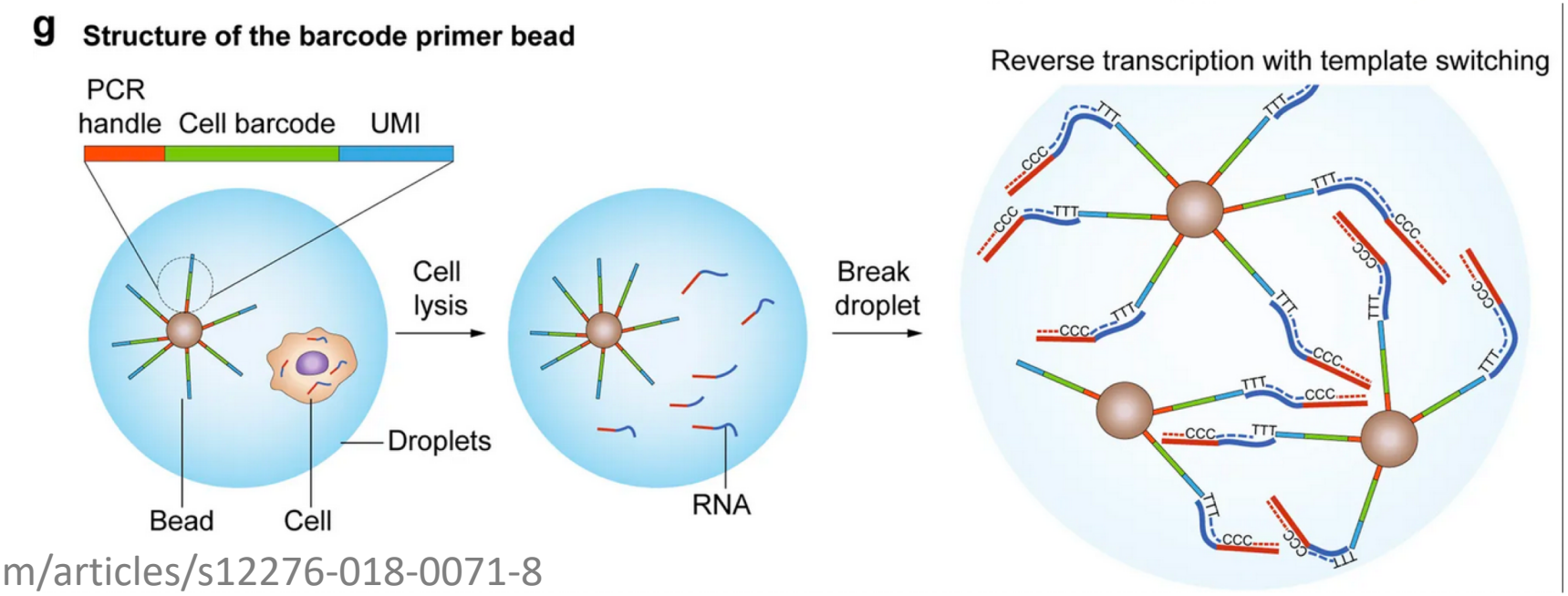


https://www.nature.com/articles/s12276-018-0071-8

10x genomics: overview

Via 10x genomics





https://www.nature.com/articles/s12276-018-0071-8

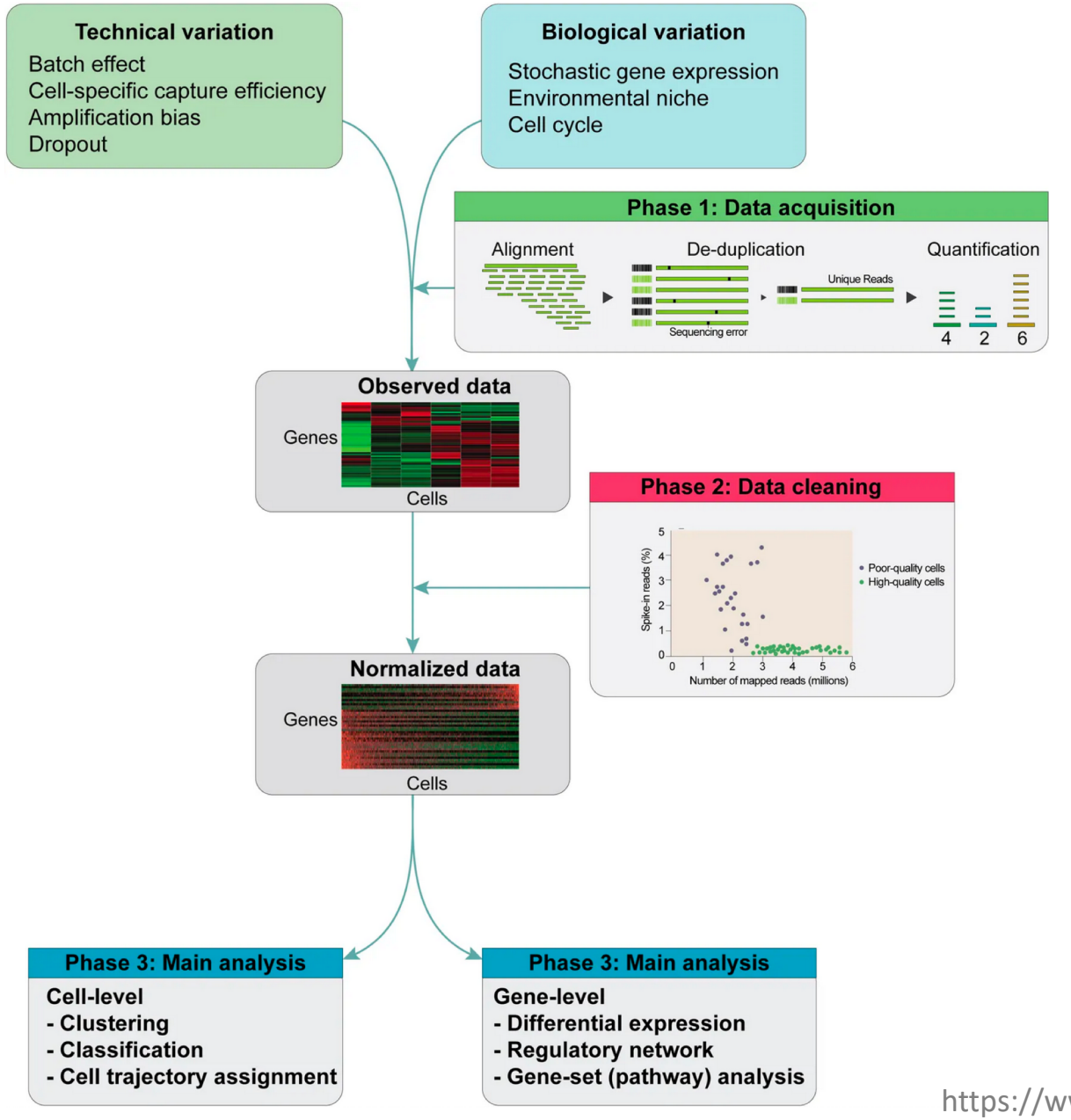
10x genomics: what it's good for

- Looking for rare cell populations in heterogeneous tissue
  - Retina: https://www.nature.com/articles/s41467-019-12780-8
  - Lung: https://www.nature.com/articles/s41586-018-0394-6
- Tracking cell abundance/frequency in progressions
  - Development/embryonic
    - https://www.nature.com/articles/s41586-019-0969-x
  - Health and disease
    - https://www.nature.com/articles/s41467-019-12464-3
- With genotyping, can identify subpopulation-specific eQTLs and other genomic features
  - https://pubmed.ncbi.nlm.nih.gov/29610479-single-cell-rna-sequencing-identifies-celltype-specific-cis-eqtls-and-co-expression-qtls/

Computational processing of single cell RNAseq data: overview and challenges I

- For most cells that are not PBMCs
  - Cells are too large to fit will CLOG the 10x chip
  - Many cultured cells will work though
- Need to use single nuclei
  - Nuclear isolation often causes RNA spillover
  - This is termed “ambient RNA”
  - It makes clustering the data (without additional processing) difficult
    - scVI (a neural network method) can model the ambient or background RNA
- Tissue storage/collection heterogeneity
  - Is the tissue a biopsy? Post mortem?
    - RNA degrades really fast!

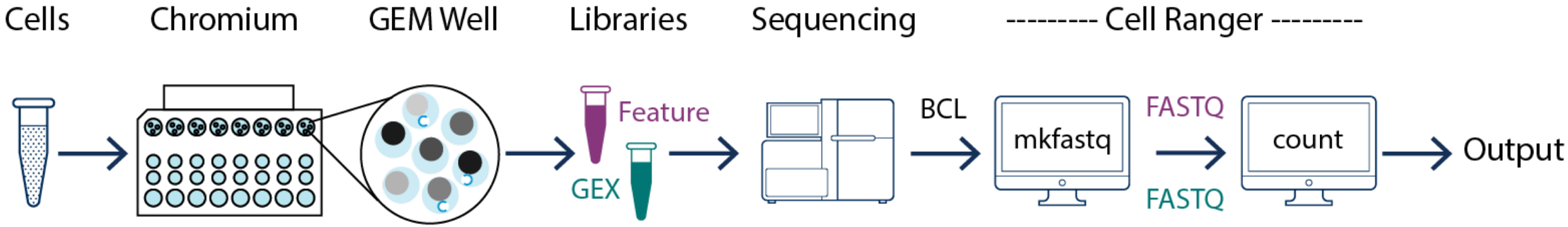
Computational analysis of single cell data: overview



https://www.nature.com/articles/s12276-018-0071-8

Computational processing of single cell RNAseq data: Cellranger/cellbender

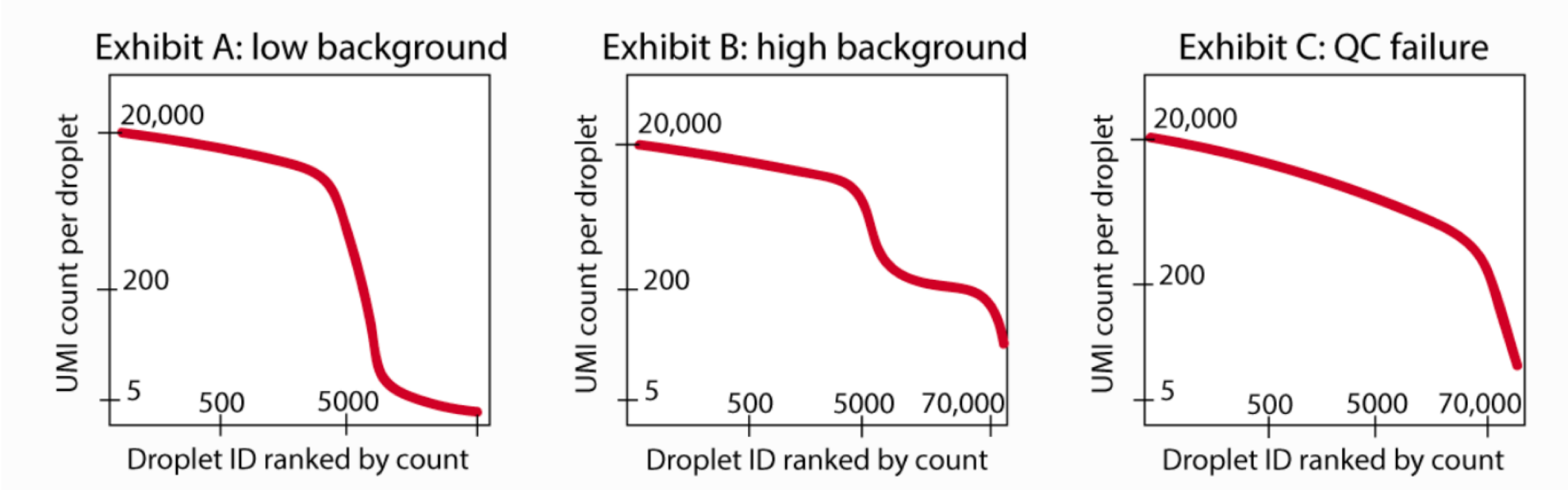
- Cellranger is the 10x genomics provided pipeline
  - Cellranger processes FASTQ files with UMI generated by sequencing 10x genomics libraries
- Cellbender was developed by the Broad DSP (Data Sciences Platform)
  - Cellbender removes the effects of ambient RNA



https://support.10xgenomics.com/single-cell-gene-expression/software/overview/welcome

Computational processing of single cell RNAseq data: Cellranger/cellbender

- Cellranger is the 10x genomics provided pipeline
  - Cellranger processes FASTQ files with UMI generated by sequencing 10x genomics libraries
- Cellbender was developed by the Broad DSP (Data Sciences Platform)
  - Cellbender removes the effects of ambient RNA



https://github.com/broadinstitute/CellBender https://cellbender.readthedocs.io/en/latest/

Computational processing of single cell RNAseq data: Kallisto bustools & Alevin (Salmon)

- Kallisto uses “pseudoalignment” and bustools can process single cell UMI-based data from the fastq phase
  - Available for R and python
    - Kallisto available in prem
- Salmon uses a similar method “quasi-mapping” and Alevin can process:
  - Drop-seq
  - 10x-Chromium v1/2/3

https://salmon.readthedocs.io/en/latest/alevin.html

https://www.kallistobus.tools/about https://www.biorxiv.org/content/10.1101/673285v2

Computational downstream processing of single cell RNAseq data: Seurat

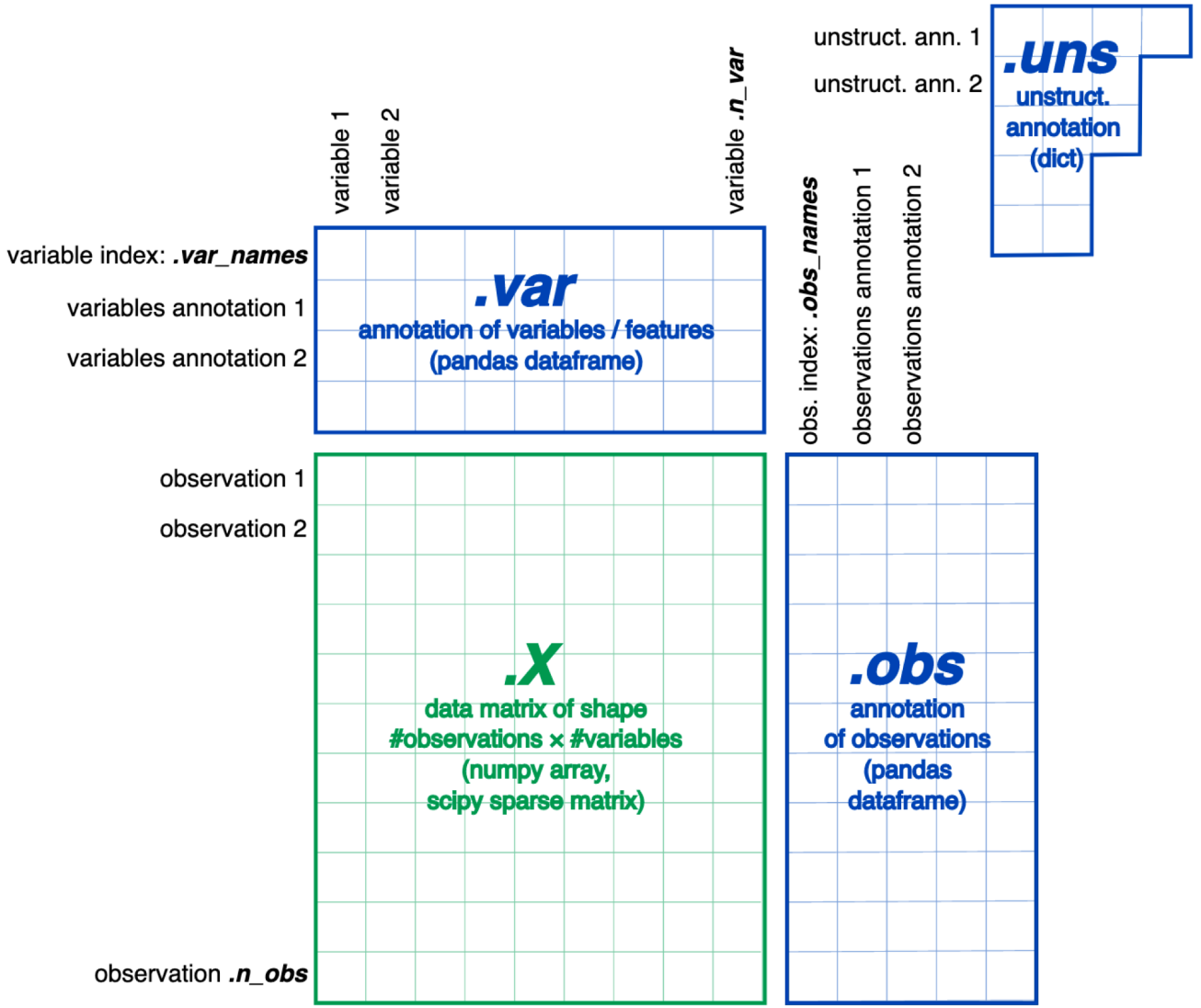
- R package for downstream analysis of processed 10x genomics single cell data
  - Nice plots!
  - Easy to install
  - Will take 10x genomics data from a directory and process into an object
- Well documented and good tutorials
  - https://satijalab.org/seurat/vignettes.html

https://www.cell.com/cell/fulltext/S0092-8674(19)30559-8

https://satijalab.org/seurat/

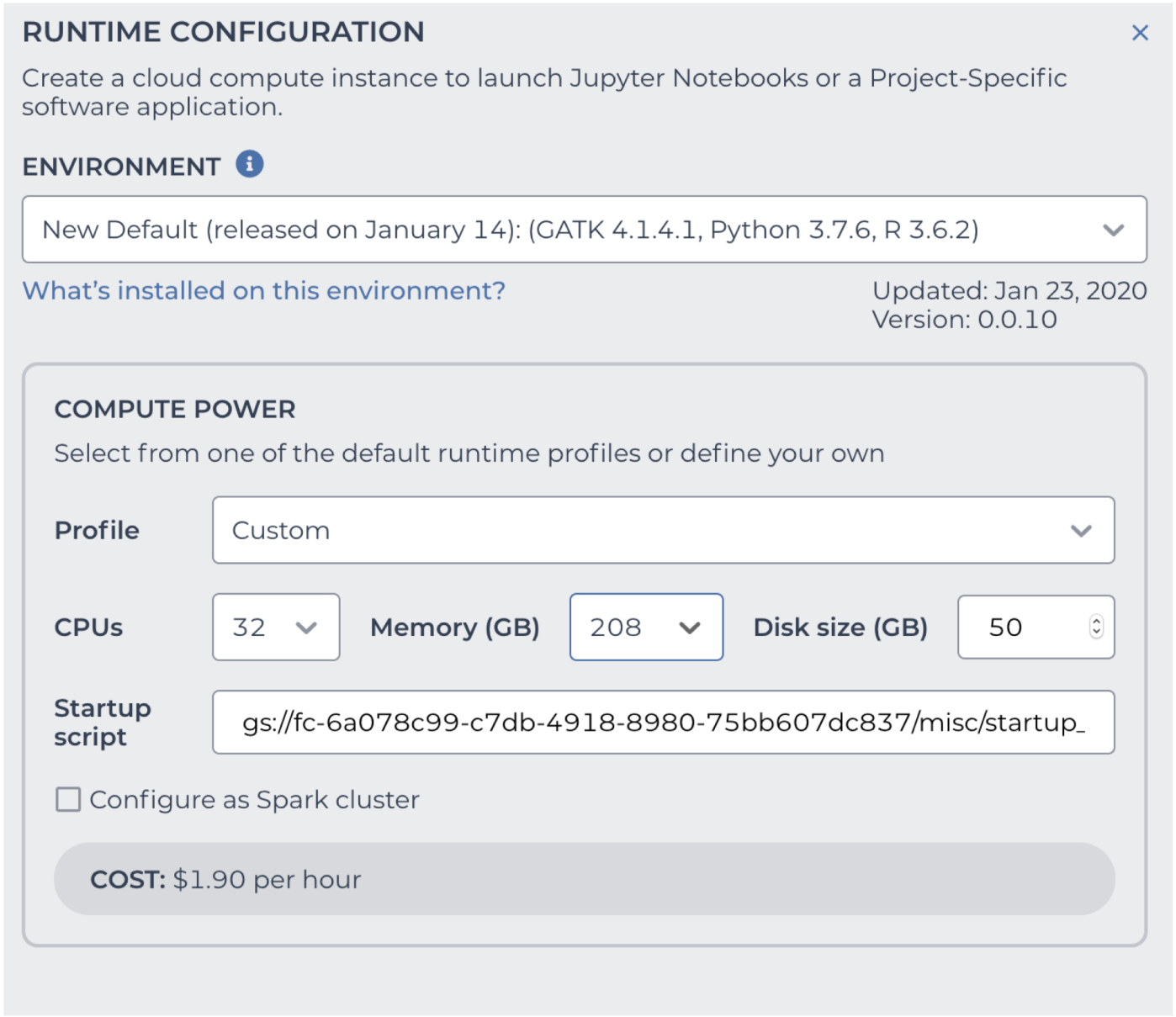
Computational downstream processing of single cell RNAseq data: Scanpy/Anndata

- Scanpy handles analysis of Anndata objects
- Scanpy produces really nice plots and maps of the dimensionality of the data
- Excellent support and tutorials



https://icb-scanpy.readthedocs-hosted.com/en/stable/

Notebook setup: H4C is huge



Workflow of the sandbox

Environment/Setup

- VM setup
- Gsutil cp H4C data file
- Set up python
- Read in h5ad file

Analyses

- Plotting genes w/ highest fraction of counts
- Filtering (basic, MT gene, highly variable gene)
- Compute PCA

Visual outputs

- Plot UMAP/ Louvain_1.0
- Plot by genes of interest
- Plot Louvain marker genes
- Violin plots and dotplots of marker genes

https://app.terra.bio/#workspaces/bayer-pcl-single-cell/single%20cell%20sandbox%20scanpy/notebooks/launch/custom_gene_list_visualization.ipynb?mode=edit

Future Plans/Unsolved Questions

- How much pre-plotting normalization/filtering is required/necessary for this dataset?
  - Different threshold recommendations for different purposes?
- Easy input and plotting for large numbers of genes?
- Subclustering of individual subgroups and clusters?
- Best ways to identify marker genes?
  - AUC calculation?
- Any other useful or highly desired features?