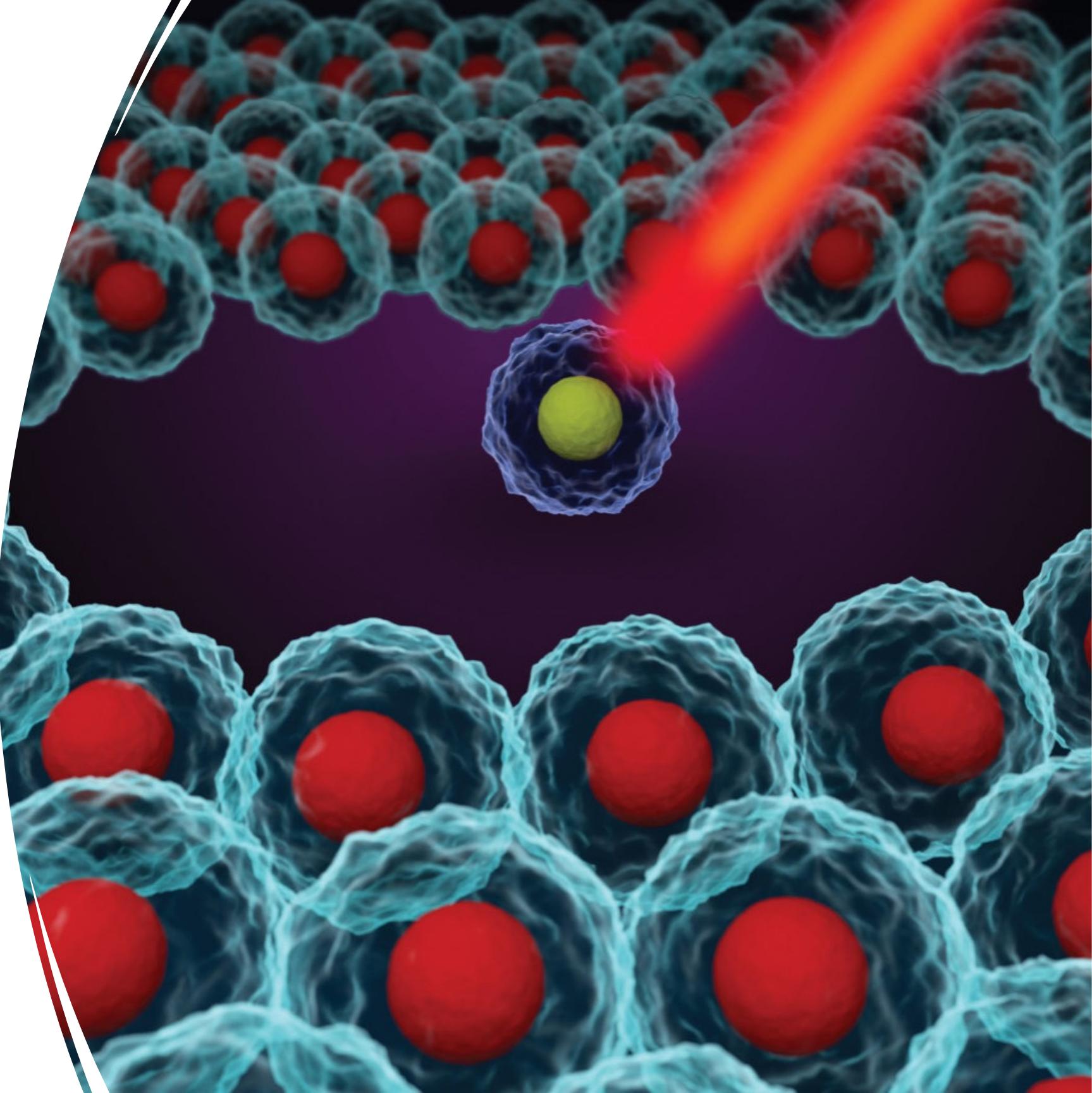


Exploring the world of **Single Cell Technology**

Single Cell Boot Camp 2025

14 July 2025

Simone Puccio



What is Single Cell Technology?

The Basics

Single Cell Technology lets us study the behavior of individual cells, which is critical for understanding things like disease progression and cancer.

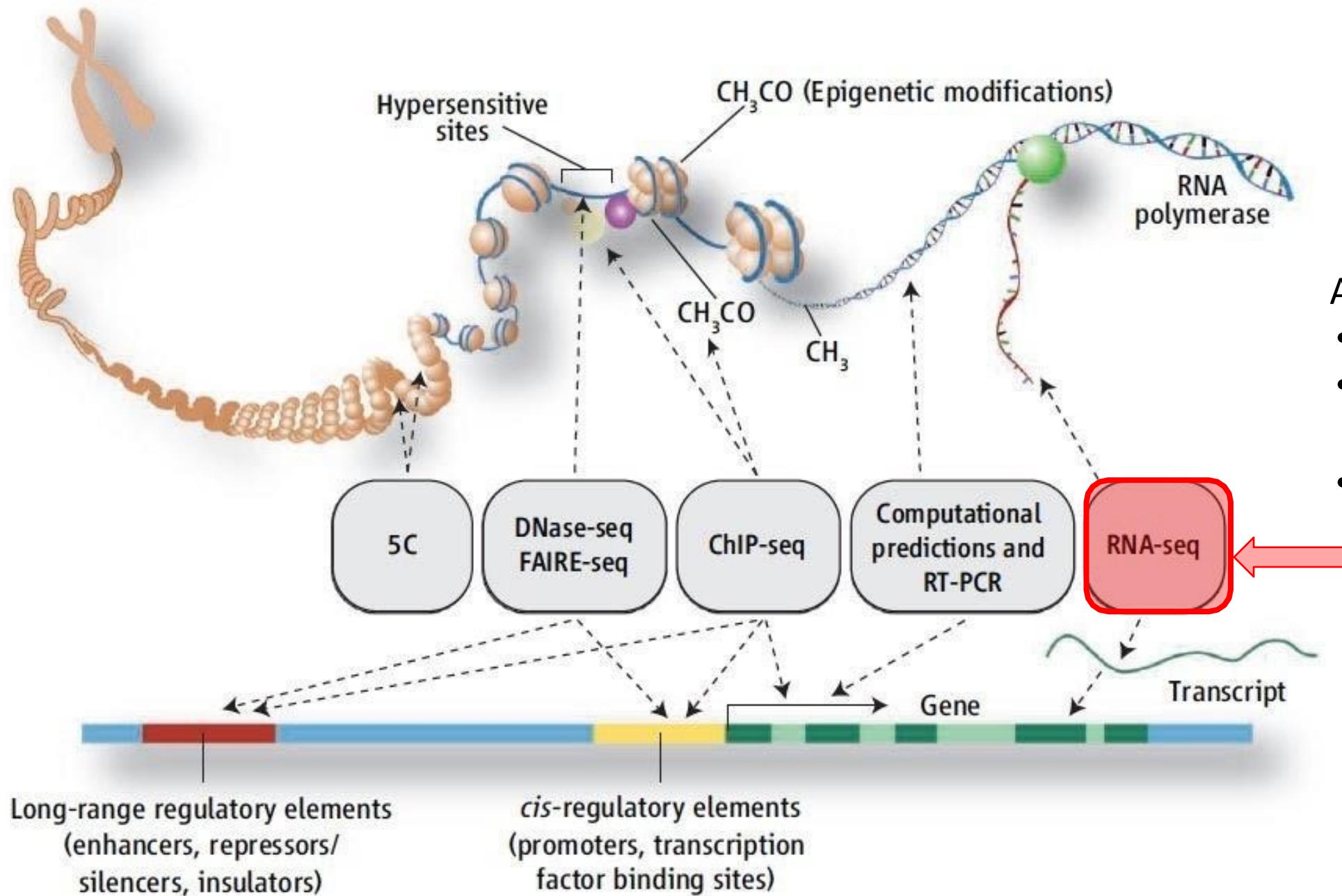
The Process

This technique involves the isolation and analysis of individual cells, allowing researchers to study the differences between cells in the same tissue type.

The Benefits

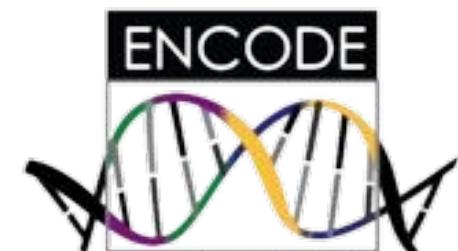
Single Cell Technology can help us understand the intricacies of human biology and unlock new treatments for a variety of diseases.

There are multiple genomic layers that can be measured



And

- DNA Methylation
- Nucleosome position (ATAC-seq)
- etc...



The Impact of Single Cell Technology

Advancing Medical Research

Single Cell Technology has led to new discoveries in cancer biology, developmental biology, and more, and has the potential to unlock new treatments for a variety of diseases.

Improving Agricultural Practices

By better understanding the genetics of crop plants, this technology can help researchers develop new crop varieties that are more resistant to disease and pests.

Informing Biotech Industry

Single Cell Technology has the potential to transform the biotech industry by providing a more nuanced understanding of the function of individual cells in the body.

One genome: diverse functional outputs



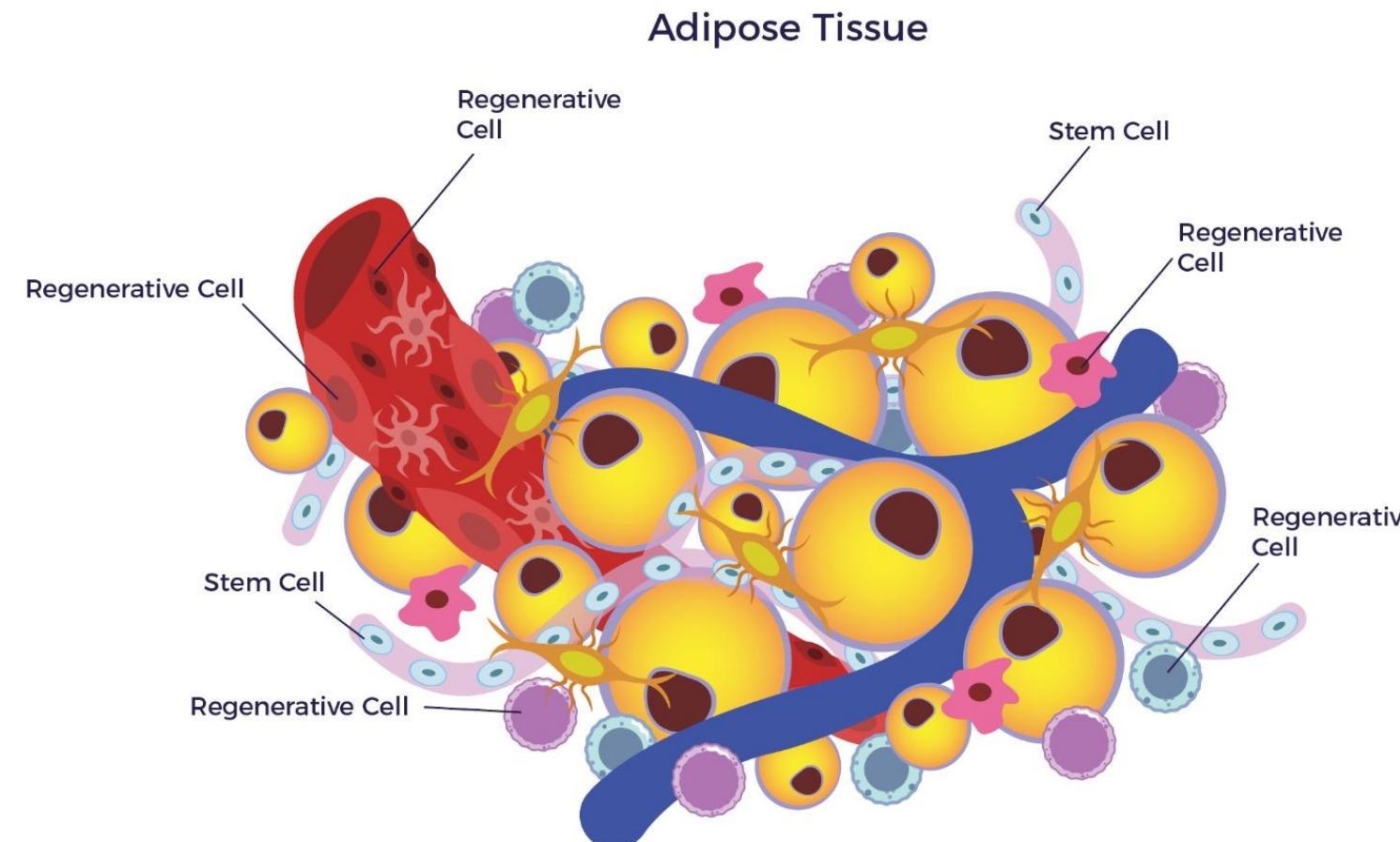
1 genome

One genome gives rise to a multitude of different cell types with highly distinct morphologies & functions

Remarkable
Cellular Diversity
and Specialisation

Human body:
~ 100 trillion cells
~ 200 types of cells
complex
tissue & organ
functions

Tissues are generally heterogeneous



Bulk RNA-seq: estimate expression of transcripts in a sample

How to measure gene expression?

⇒ Bulk RNA-seq

Technique appeared in 2008

The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing

Ugrappa Nagalakshmi^{1,*}, Zhong Wang^{1,*}, Karl Waern¹, Chong Shou², Debasish Raha¹, Mark Gerstein^{2,3}, Michael Snyder^{1,2,3,†}

Abstract

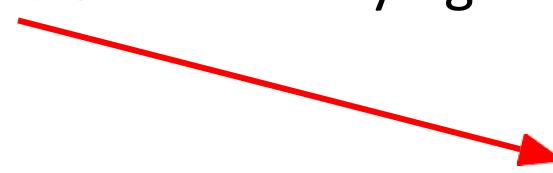
The identification of untranslated regions, introns, and coding regions within an organism remains challenging. We developed a quantitative sequencing-based method called RNA-Seq for mapping transcribed regions, in which complementary DNA fragments are subjected to high-throughput sequencing and mapped to the genome. We applied RNA-Seq to generate a high-resolution transcriptome map of the yeast genome and demonstrated that most (74.5%) of the nonrepetitive sequence of the yeast genome is transcribed. We confirmed many known and predicted introns and demonstrated that others are not actively used. Alternative initiation codons and upstream open reading frames also were identified for many yeast genes. We also found unexpected 3'-end heterogeneity and the presence of many overlapping genes. These results indicate that the yeast transcriptome is more complex than previously appreciated.

Limitations of bulk RNA-seq

Bulk RNA-seq was a major breakthrough in the late 00's (replaced microarrays)
=> Great advances were made through genomics, but ...

Limitations: minimum starting material requirements techniques
applied on millions of cells

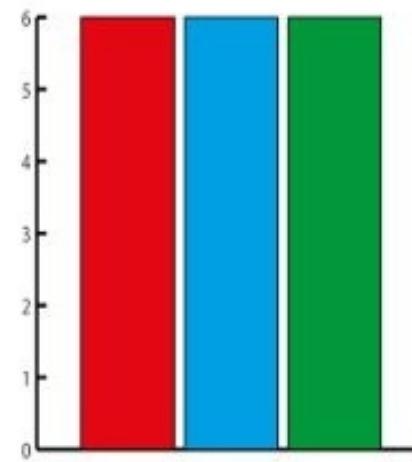
- *rare cell types & states* cannot be analyzed (e.g. transitions, circulating tumor cells, etc..)
- *insufficient* for studying heterogeneous systems (e.g. complex tissues such as brain)



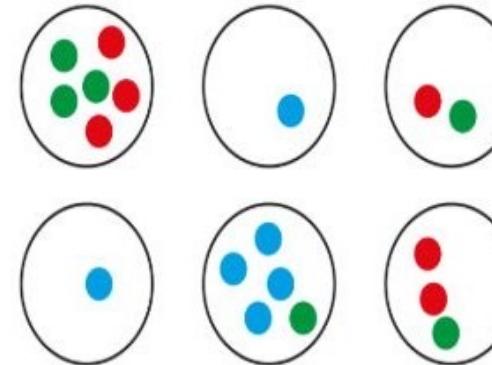
!each sample is an AVERAGE!
no idea of the underlying values in single cells of the
heterogeneity of the tissue

Single-cell RNA-seq

Bulk analysis



Single cell transcriptome analysis



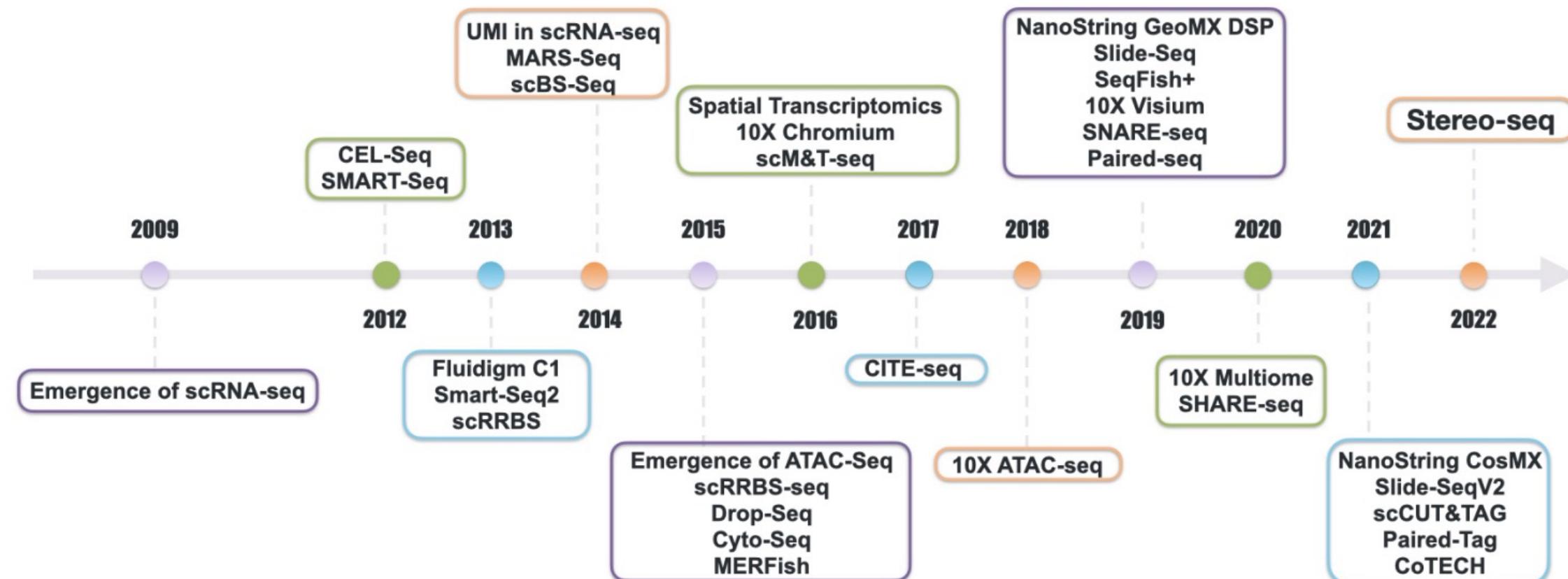
[Macaulay IC, Voet T \(2014\) PLoS Genet](#)

Each black circle is a cell, each colored dot is a transcript, colors encode transcript of the same gene

- **RARE CELL TYPES** (e.g. early development, stem cells, circulating tumor cells)
- **HETEROGENEITY** (e.g. tissue composition, cancer, temporal processes) **GENE**
- **REGULATORY NETWORKS** (non-confounded correlations)
- **SINGLE-CELL PHENOMENA*** (gene expression stochasticity, mono-allelic expression)

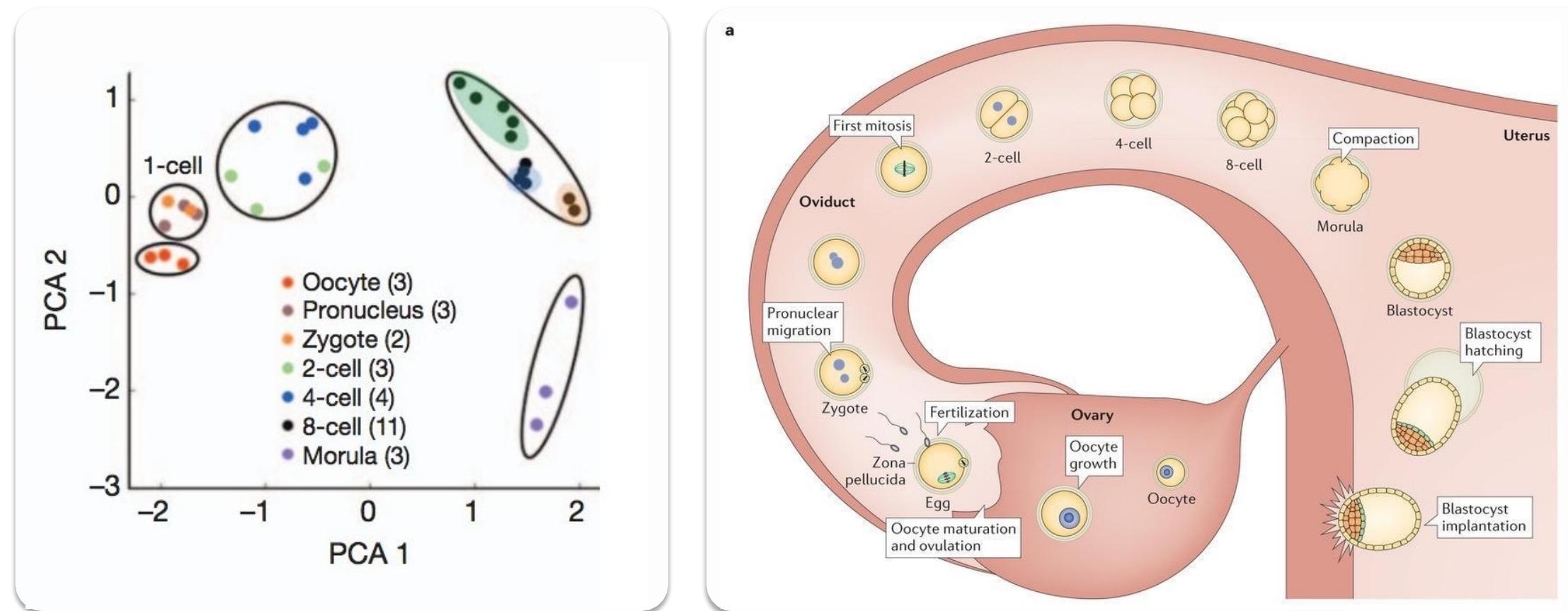
* see also review [CoulonLarsonNatRevGenet2013](#)

Single-cell timeline



Single-cell transcriptomics (scRNA-seq) applications – Development

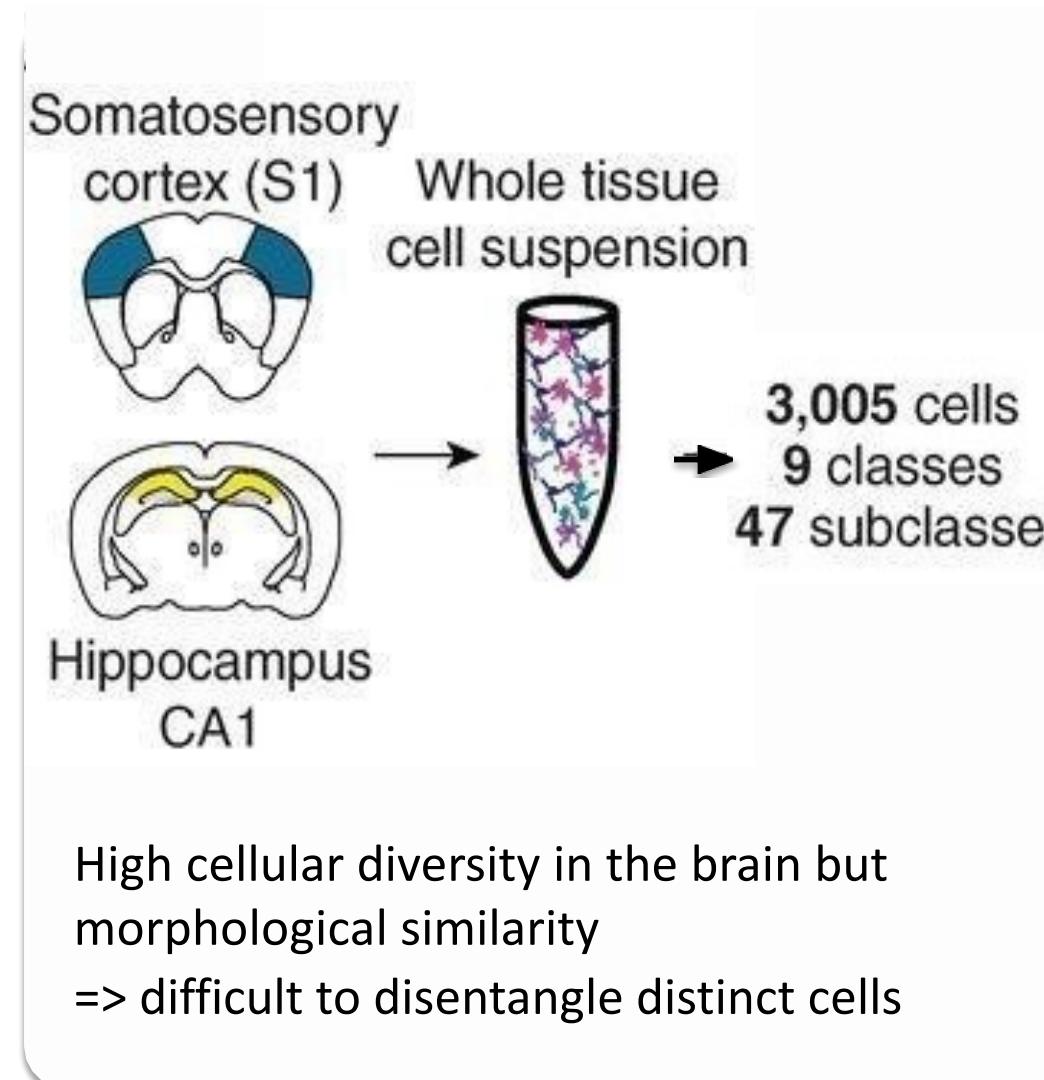
Analyzing transcriptome of cells in human and mouse early embryos



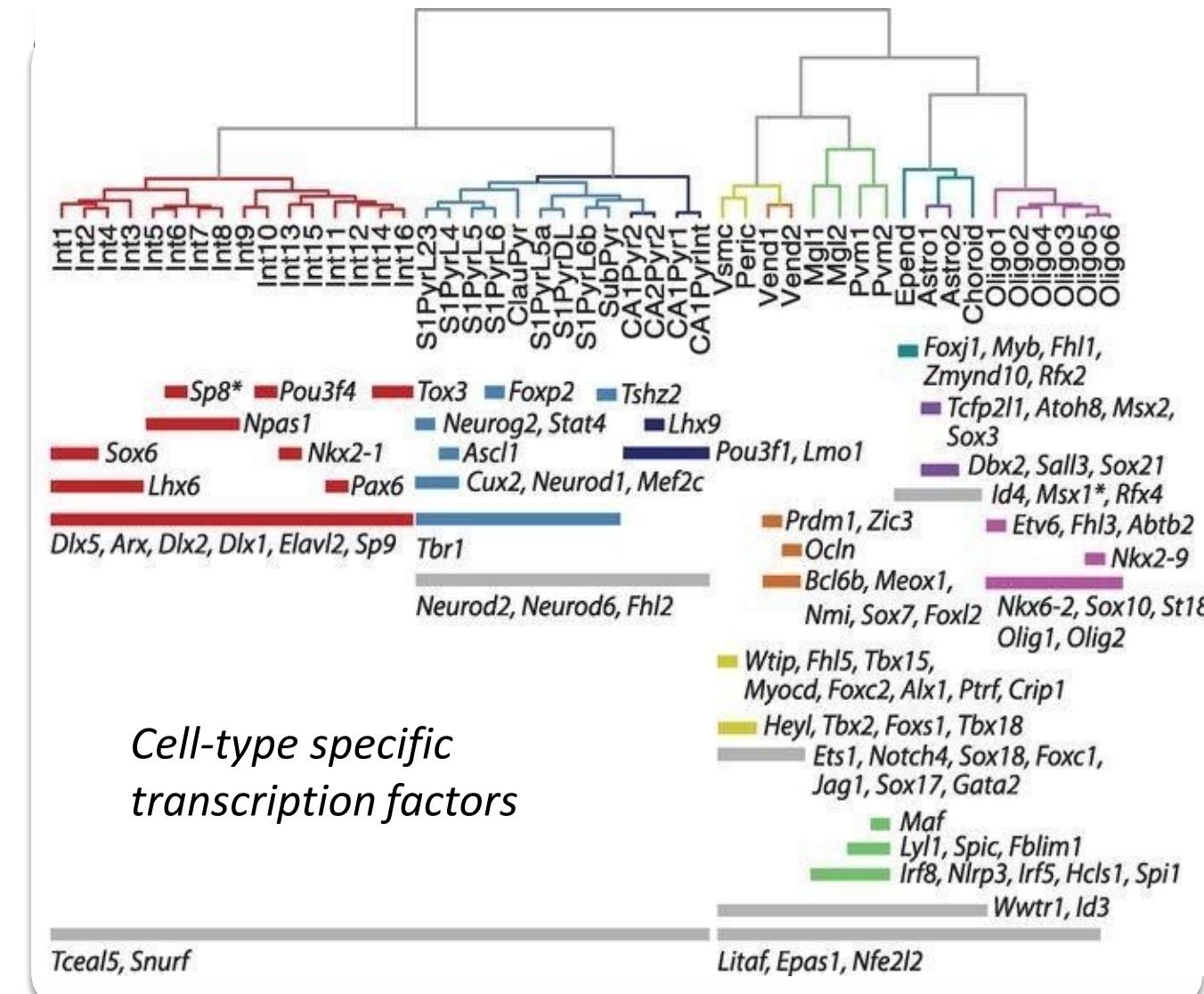
- ⇒ Each developmental stage can be delineated concisely by a small number of functional modules of co-expressed genes
- ⇒ Temporal developmental pattern different mouse-human
- ⇒ Conserved key members of human & mouse networks

scRNA-seq applications – Tissue heterogeneity

Mapping out cell types in the mouse cortex & hippocampus



Neuronal cell types hierarchy



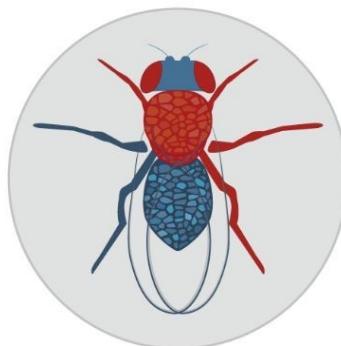
Interneurons of similar type exist in dissimilar regions of the brain

Identification of oligodendrocytes subtype

Microglia associated with blood vessels distinguished from perivascular macrophage

Zeisel et al. (2015) Science

scRNA-seq applications – Creating XXX cell atlases



FLY CELL ATLAS

About

The Fly Cell Atlas will bring together Drosophila researchers interested in single-cell genomics, transcriptomics, and epigenomics, to build comprehensive cell atlases during different developmental stages and disease models.

>> More

nature
neuroscience

Resource | Published: 06 May 2019

A single-cell atlas of mouse brain macrophages reveals unique transcriptional identities shaped by ontogeny and tissue environment

Hannah Van Hove, Liesbet Martens, Isabelle Scheyltjens, Karen De Vlaminck, Ana Rita Pombo Antunes, Sofie De Prijck, Niels Vandamme, Sebastiaan De Schepper, Gert Van Isterdael, Charlotte L. Scott, Jeroen Aerts, Geert Berx, Guy E. Boeckxstaens, Roosmarijn E. Vandenbroucke, Lars Vereecke, Diederik Moechars, Martin Guilliams, Jo A. Van Ginderachter, Yvan Saeys & Kiavash Movahedi

Cell Atlas of Worm

A Cell Atlas of Worm

The *C. elegans* transcriptome at single cell resolution



In Cao et al. (Science, 2017) we reported single cell RNA-seq of *C. elegans* larvae at ~50x 'shotgun cellular coverage' using a combinatorial indexing approach (sci-RNA-seq).



HUMAN CELL ATLAS

nature
International journal of science

Article | Published: 10 July 2019

A human liver cell atlas reveals heterogeneity and epithelial progenitors

Nadim Alzaran, Antonio Saviano, Sagar, Laurent Mailly, Sarah Durand, Josip S. Herman, Patrick Pessaix, Thomas F. Baumert & Dominic Grün

Cell

A Single-Cell Atlas of the Tumor and Immune Ecosystem of Human Breast Cancer

Johanna Wagner,^{1,2,14} Maria Anna Rapsomaniki,³ Stéphane Chevrier,^{1,14} Tobias Anzeneder,⁴ Claus Langwieder,⁵ August Dykgers,⁵ Martin Rees,⁵ Annette Ramaswamy,⁶ Simone Muenst,⁷ Savas Deniz Soysal,^{8,9} Andrea Jacobs,^{1,14} Jonas Windhager,^{1,10,14} Karina Silina,¹¹ Maries van den Broek,¹¹ Konstantin Johannes Dedes,¹² Maria Rodriguez Martinez,^{3,15} Walter Paul Weber,^{9,13,15} and Bernd Bodenmiller^{1,14,16,*}



bioRxiv
THE PREPRINT SERVER FOR BIOLOGY

A molecular cell atlas of the human lung from single cell RNA sequencing

Kyle J. Travaglini, Ahmad N. Nabhan, Lolita Penland, Rahul Sinha, Astrid Gillich, Rene V. Sit, Stephen Chang, Stephanie D. Conley, Yasuo Mori, Jun Seita, Gerald J. Berry, Joseph B. Shrager, Ross J. Metzger, Christin S. Kuo, Norma Neff, Irving L. Weissman, Stephen R. Quake, Mark A. Krasnow

Science

RESEARCH ARTICLE

The Malaria Cell Atlas: Single parasite transcriptomes across the complete *Plasmodium* life cycle

Virginia M. Howick^{1,*}, Andrew J. C. Russell^{1,*}, Tallulah Andrews¹, Haynes Heaton¹, Adam J. Reid¹, Kedar Natarajan², Hellen...

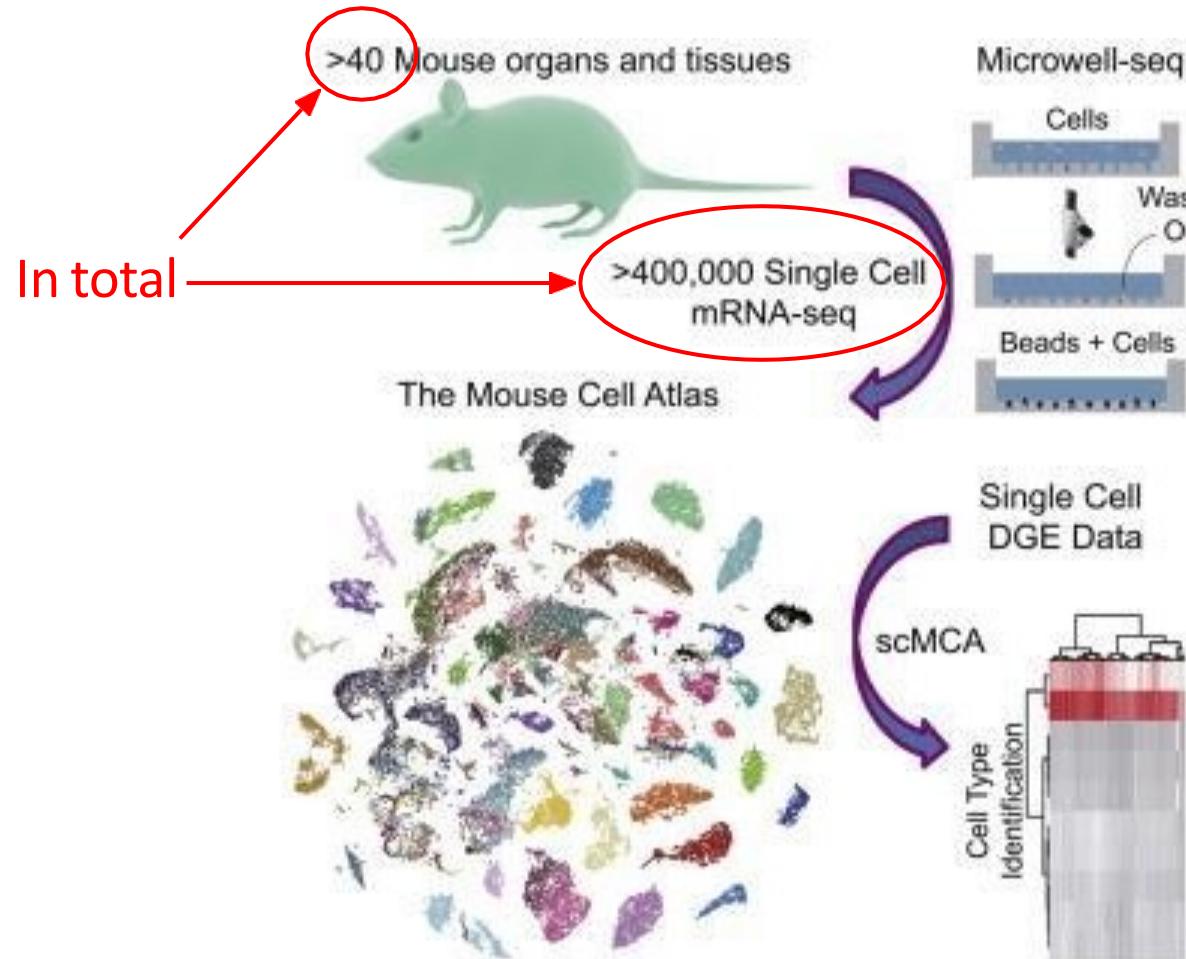
* See all authors and affiliations

Science 23 Aug 2019;
Vol. 365, Issue 6455, eaaw2619
DOI: 10.1126/science.aaw2619

Resource

Amongst first atlases: *Mus musculus*

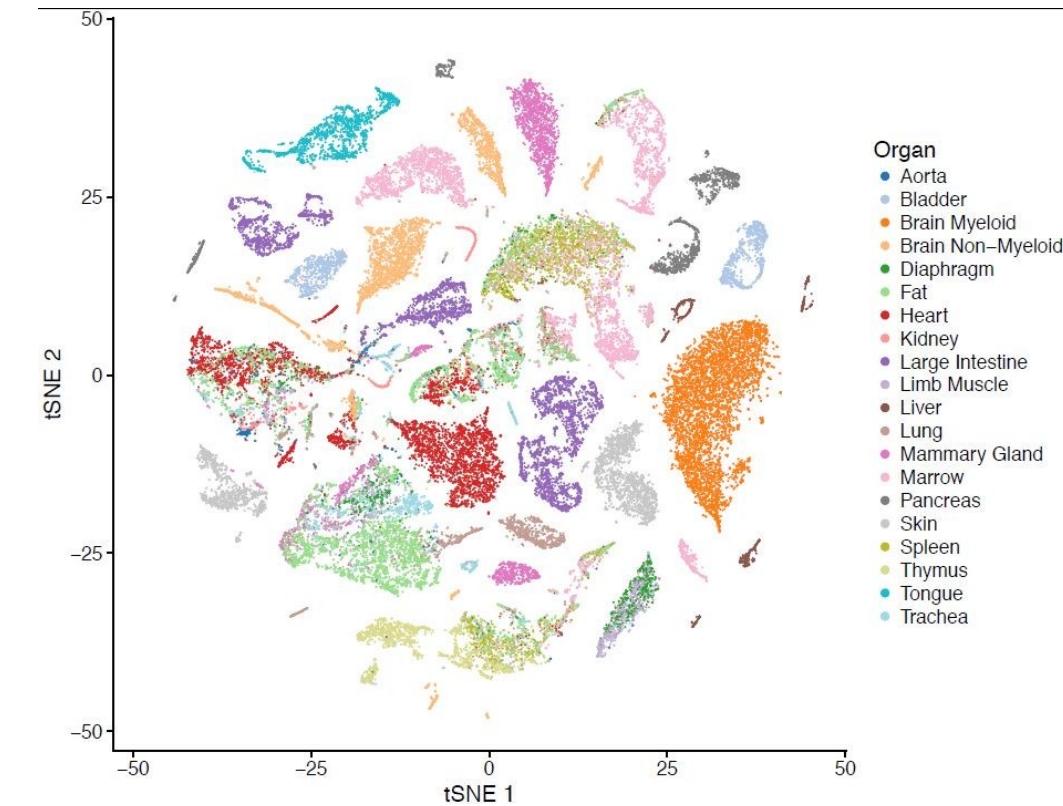
Mouse Cell Atlas (Han, Guo et al., Cell, 2018)



bis.zju.edu.cn/MCA/



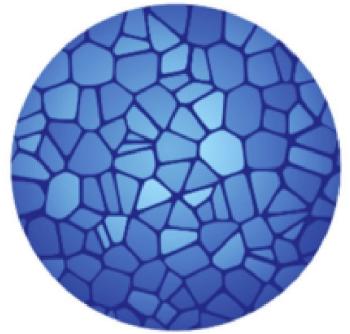
→ ~50'000 cells from 20 organs and tissues (SMART-Seq2)
~50'000 cells from 10x
(Quake, Darmanis et al., Nature, 2018)



⇒ More challenging because needs for reproducibility, data sharing, standardization of nomenclature, integration between techs

tabula-muris.ds.czbiohub.org

The Human Cell Atlas



THE HUMAN CELL ATLAS

MISSION

To create comprehensive reference maps of all human cells—the fundamental units of life—as a basis for both understanding human health and diagnosing, monitoring, and treating disease.

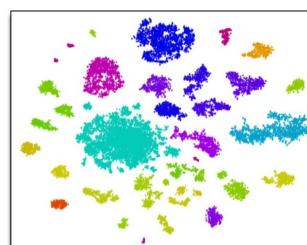
HCA plans to sequence
~1-10 billion cells?
©Dana Pe'er talk in the
past HCA meeting



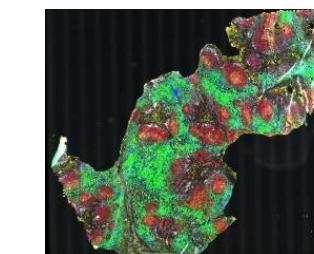
Chan Zuckerberg Biohub (\$600 M Initiative)

<https://www.humancellatlas.org>

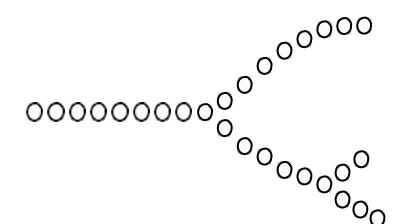
Scope, Scale, Quality and Compatibility



Cell States and Types



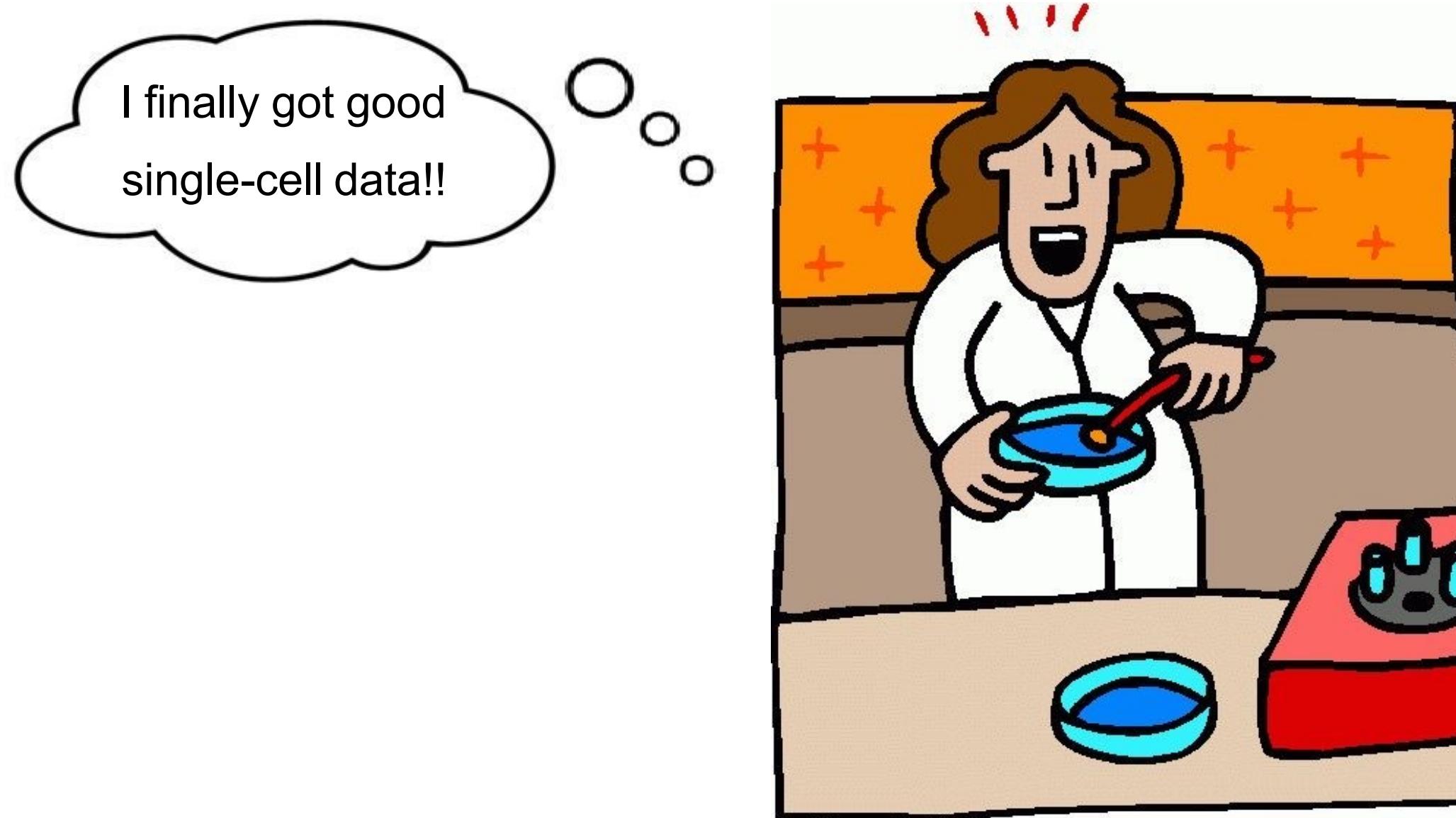
Spatial location and architecture



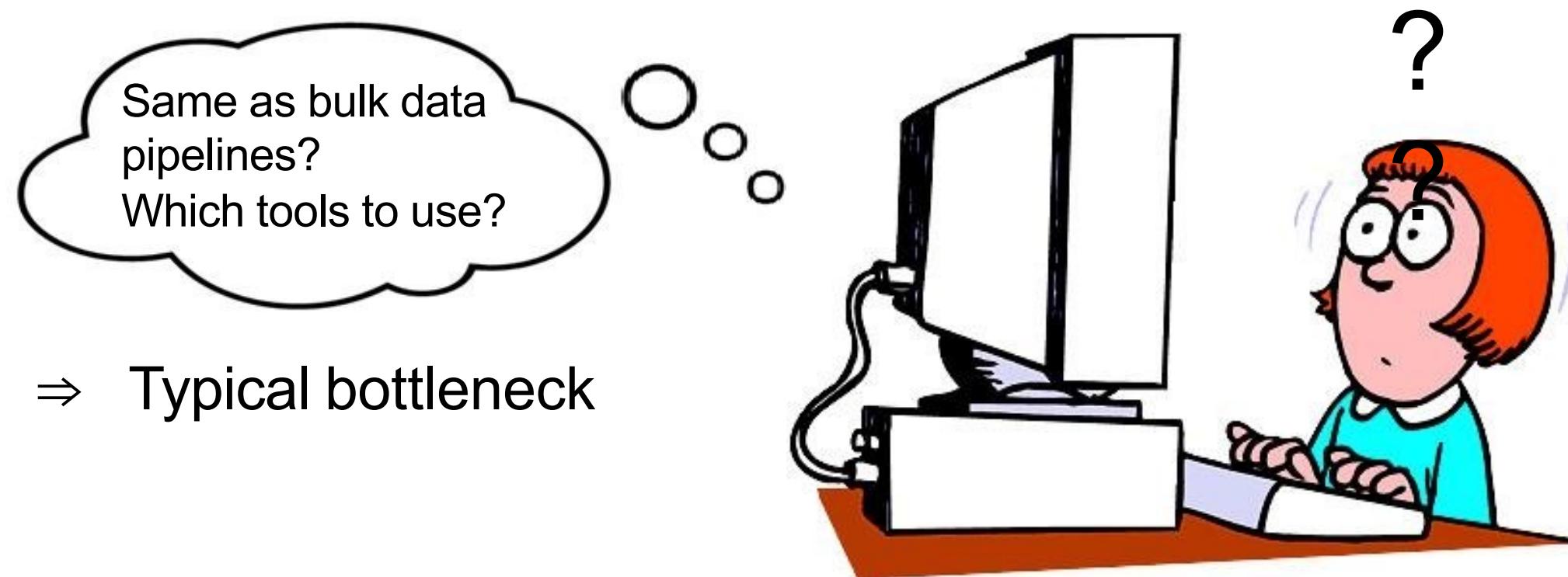
Lineages and transitions

data.humancellatlas.org

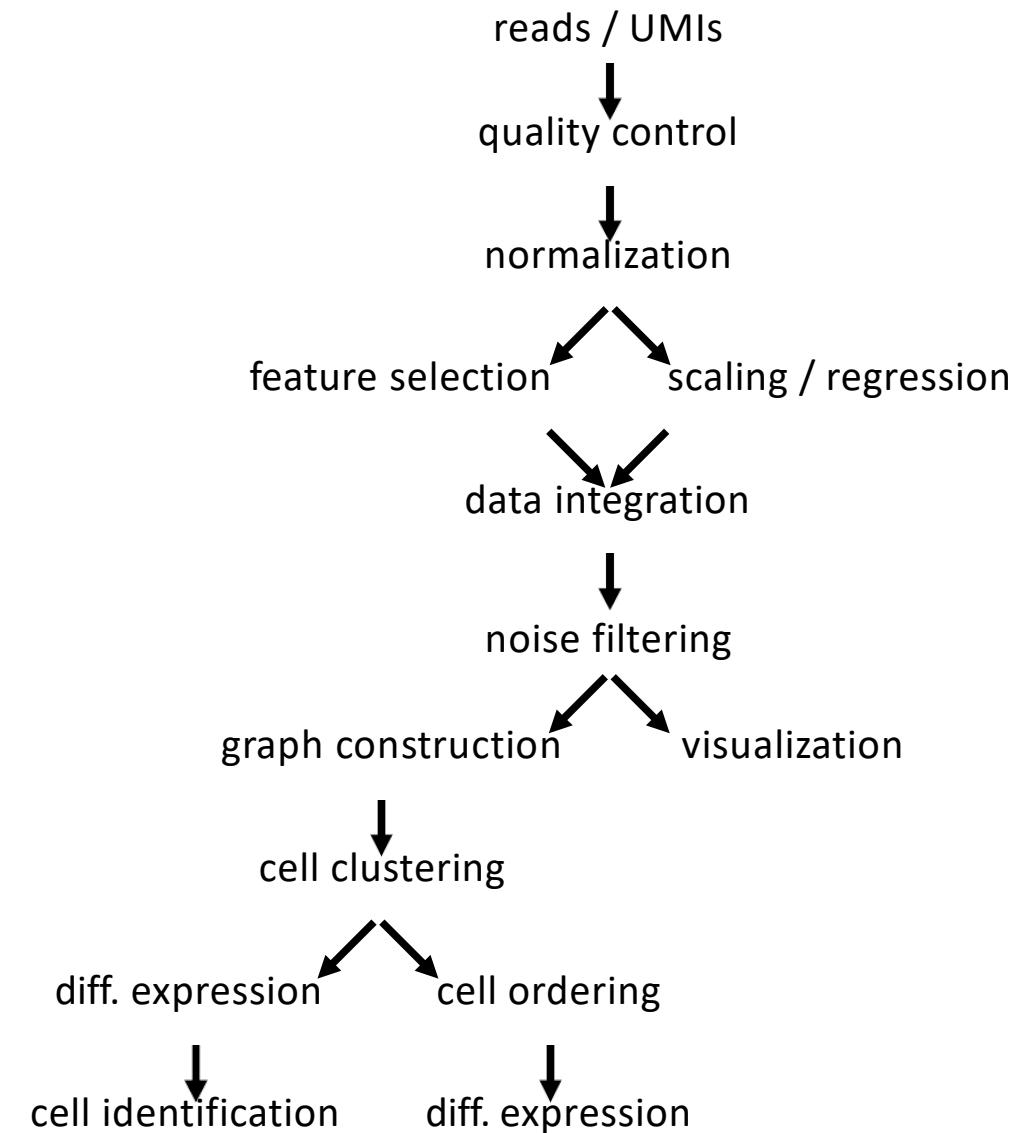
Community needs: « I want to do single-cell ! »



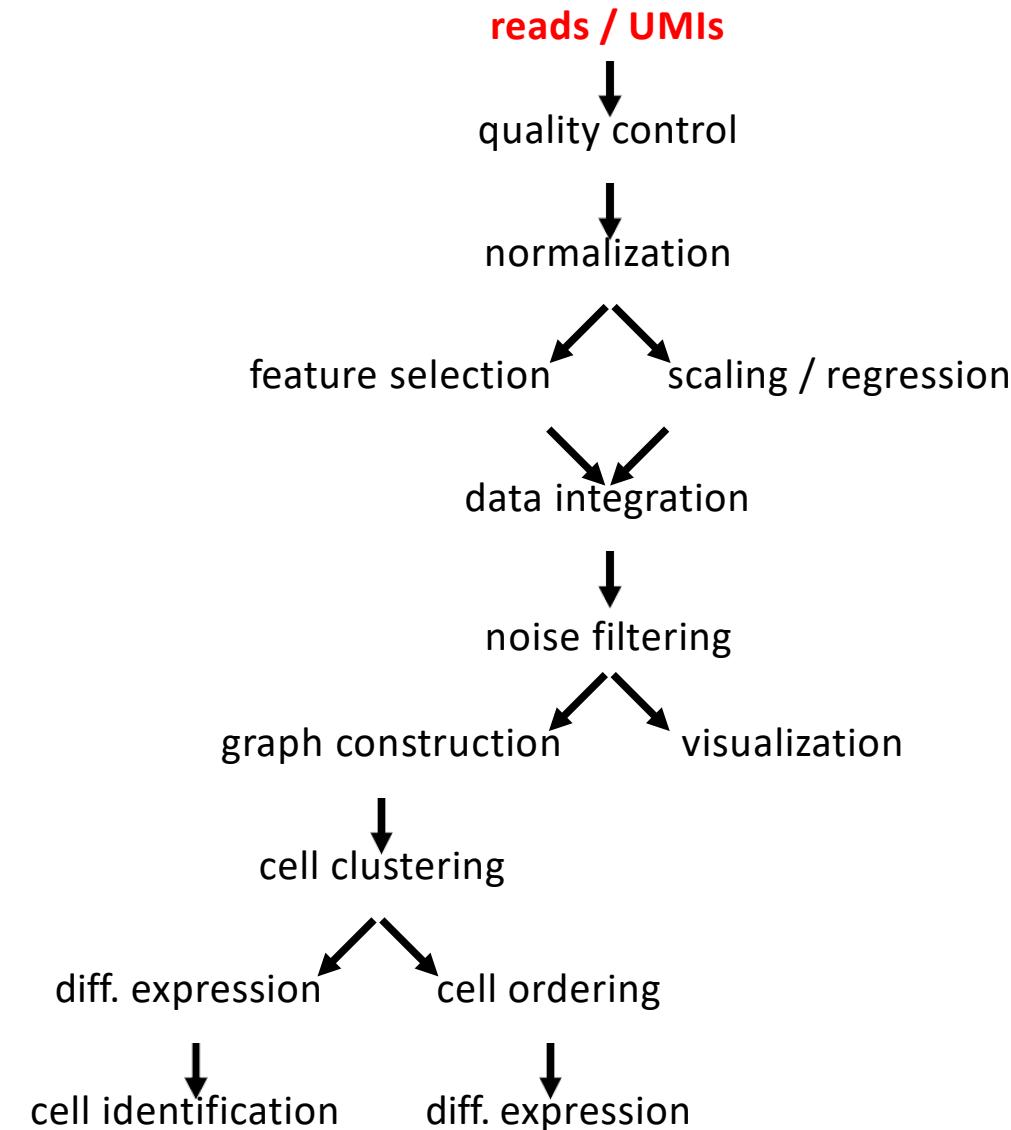
Community needs: « But how do I analyze single-cell data? »



scRNA-seq analysis workflow

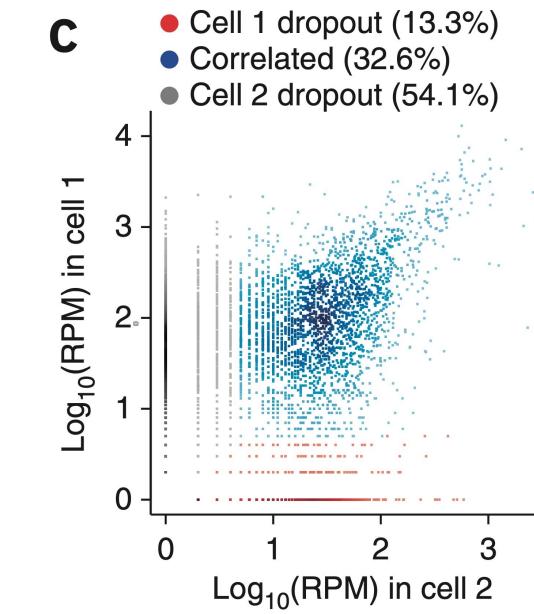
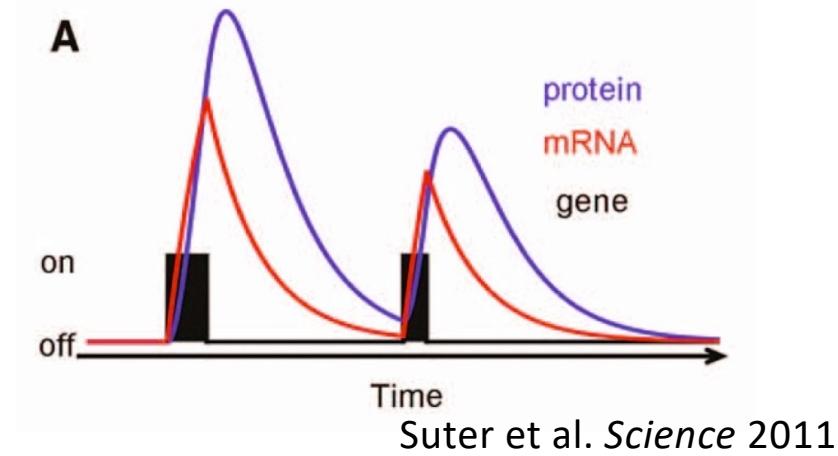


scRNA-seq analysis workflow



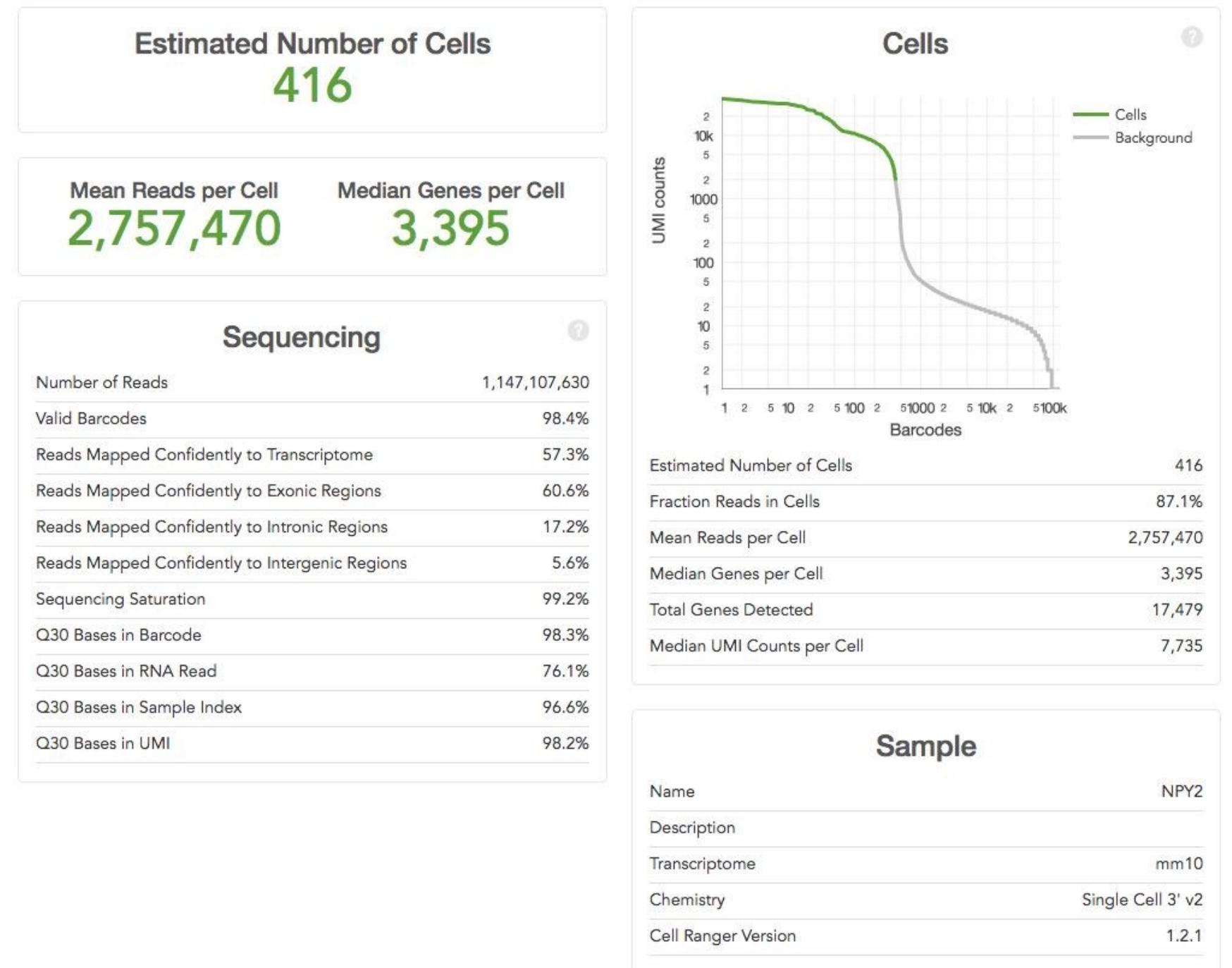
scRNA-seq biases

- Amplification bias
- Drop-out rates
- Transcriptional bursting
- Background noise
- Bias due to cell-cycle, cell size and other factors
- Often clear batch effects
- Dissociation protocols may introduce transcriptional artifacts
- Ambient RNA

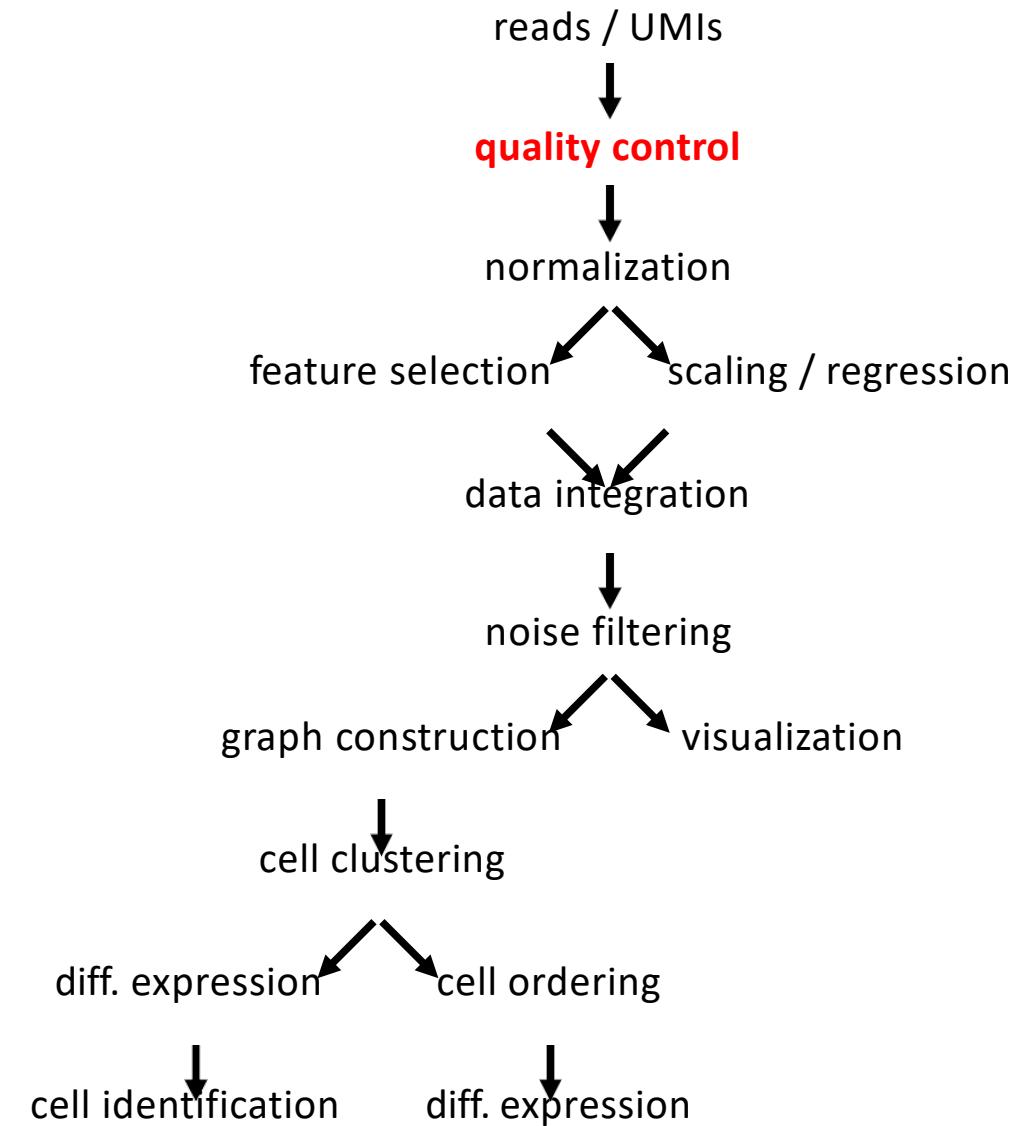


Karchenko et al. *Nature Methods* 2014

Cell Ranger output

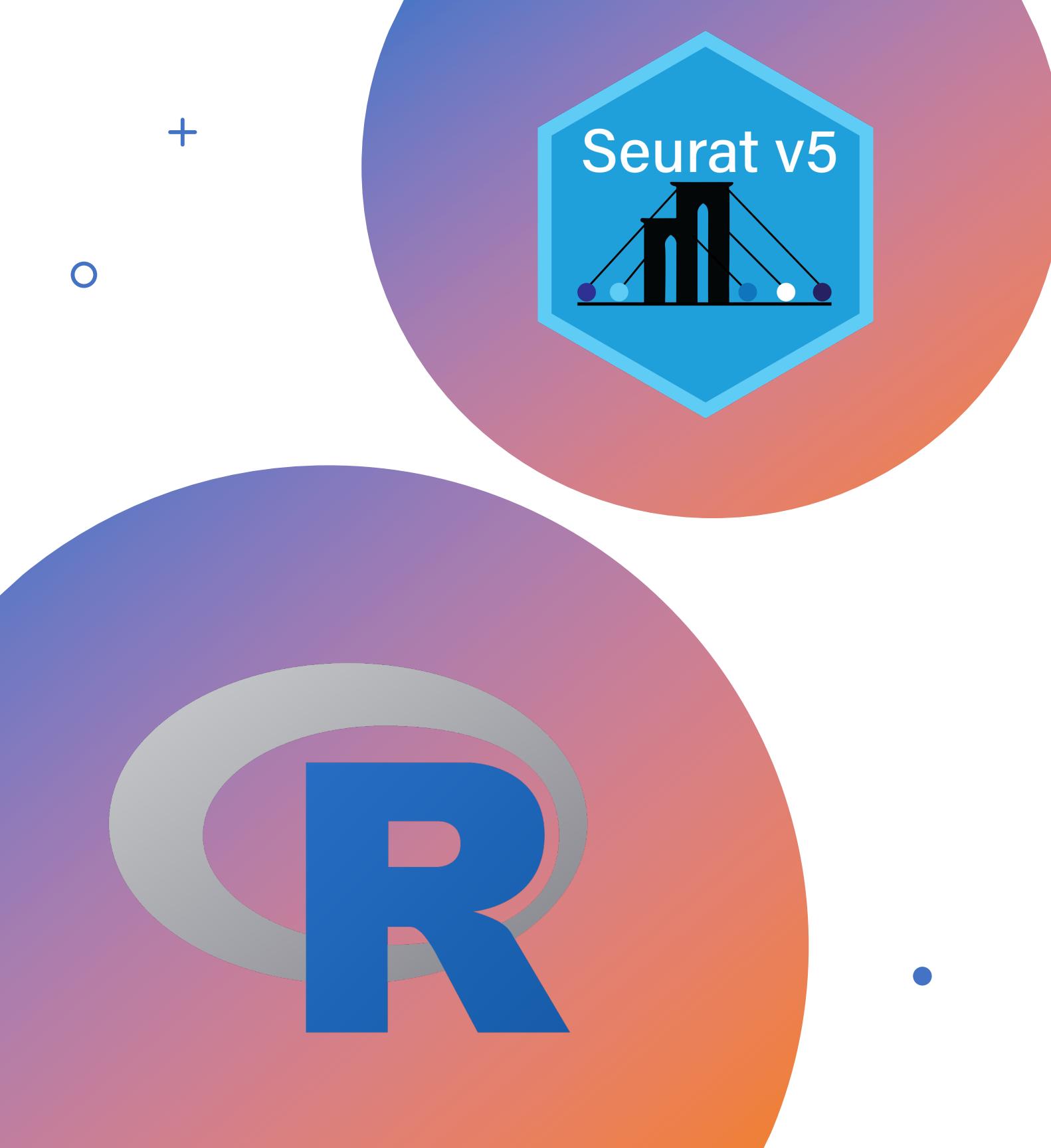


scRNA-seq analysis workflow



Single cell Analysis workflow

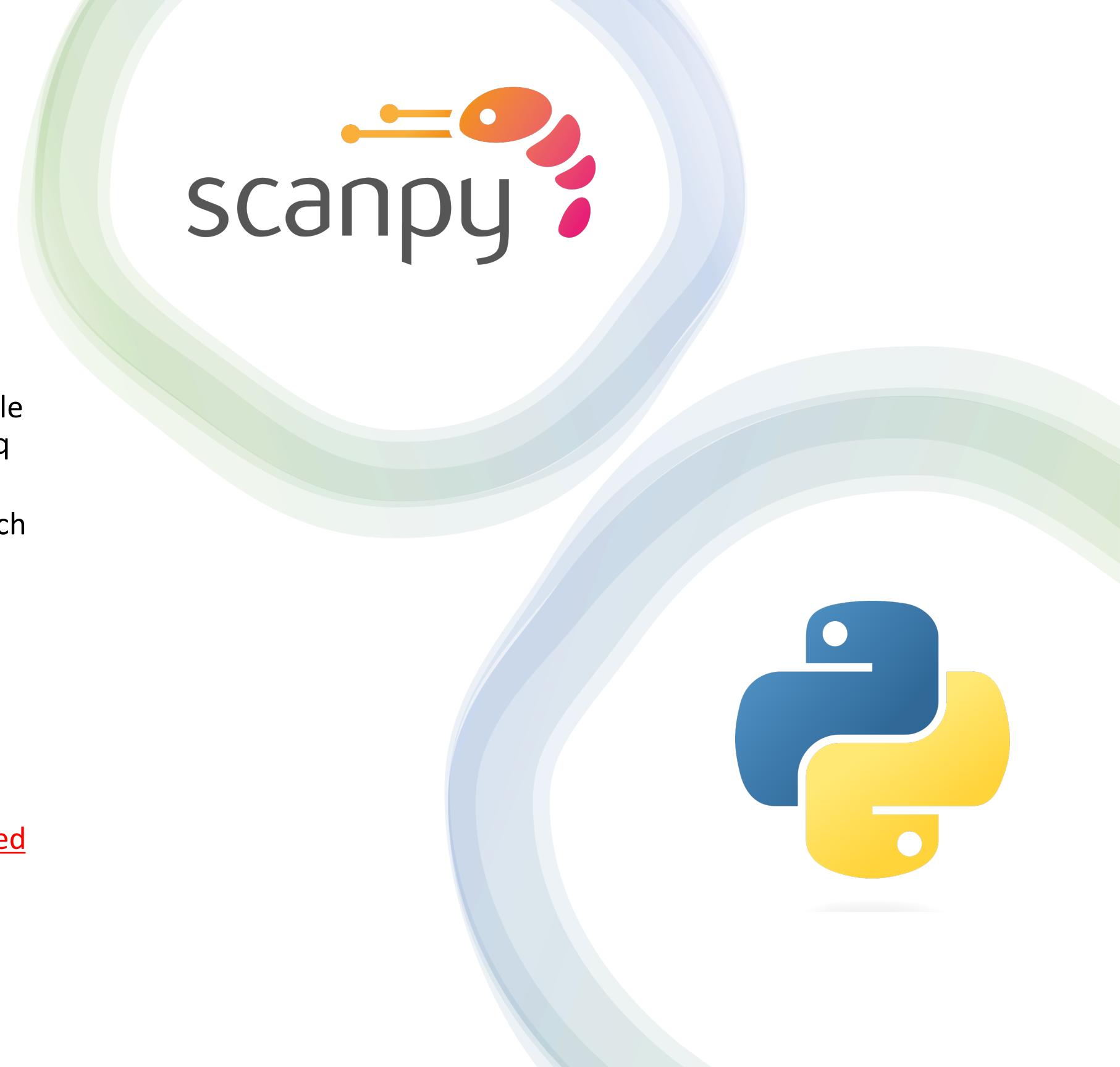
- R-based package developed by the Satija Lab at the New York Genome Center.
- Widely used in the scRNA-seq community and has a strong user base.
- Offers a comprehensive set of functions for preprocessing, dimensionality reduction, clustering, trajectory inference, and differential expression analysis.
- Supports integration of multiple scRNA-seq datasets, enabling cross-dataset analysis.
- Provides extensive visualization capabilities, including feature plots, t-SNE/UMAP embeddings, and trajectory visualization.



Single cell Analysis workflow

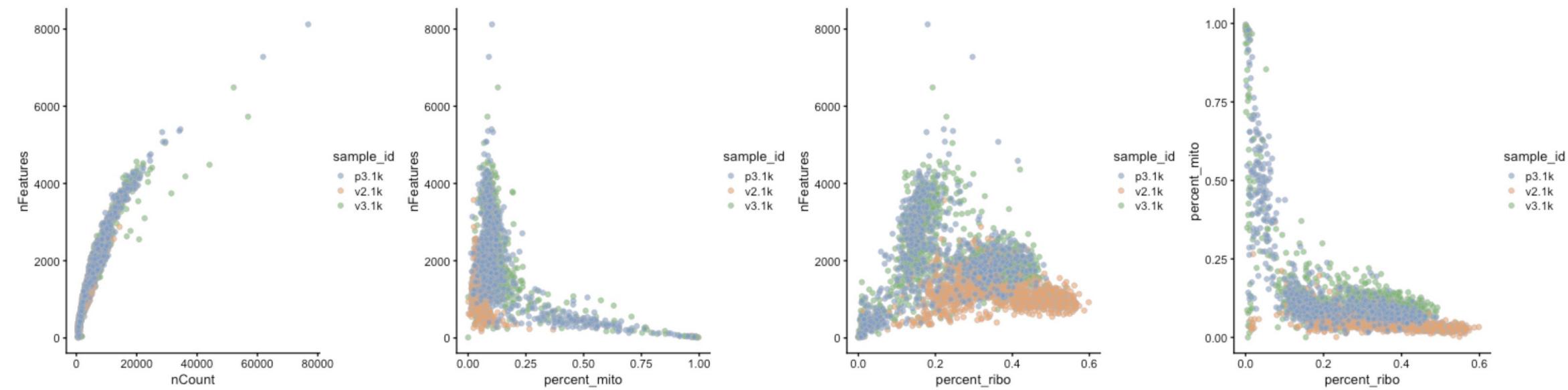
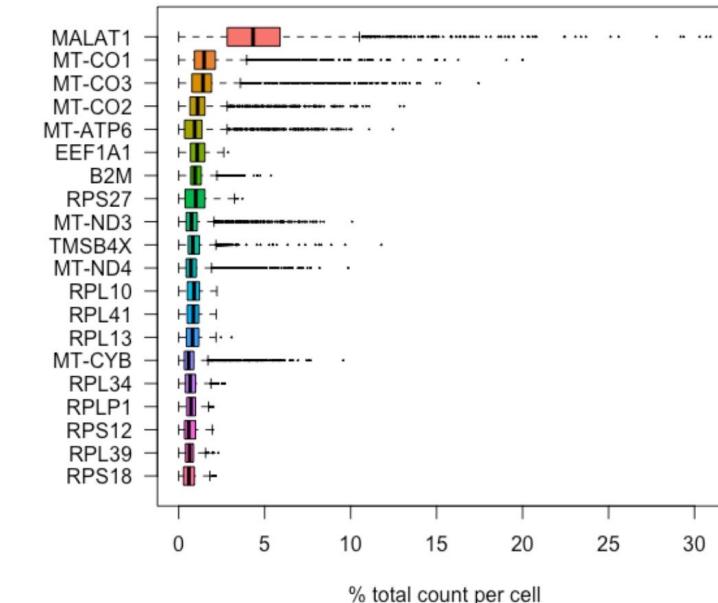
- Python-based package developed by the Genome Biology Unit at the Center for Genomic Regulation.
- As Seurat, Scipy is a user-friendly and scalable framework for analyzing large-scale scRNA-seq datasets.
- Integrates well with other Python libraries, such as numpy, pandas, and matplotlib.
- Supports visualization of results using interactive plots and embedding techniques like UMAP and t-SNE.
- Both Scipy and Seurat are powerful tools for scRNA-seq analysis, offering a wide range of functionality.

Choosing between Scipy and Seurat can be based on factors such as programming language preference, community support, and integration with other analysis pipelines.

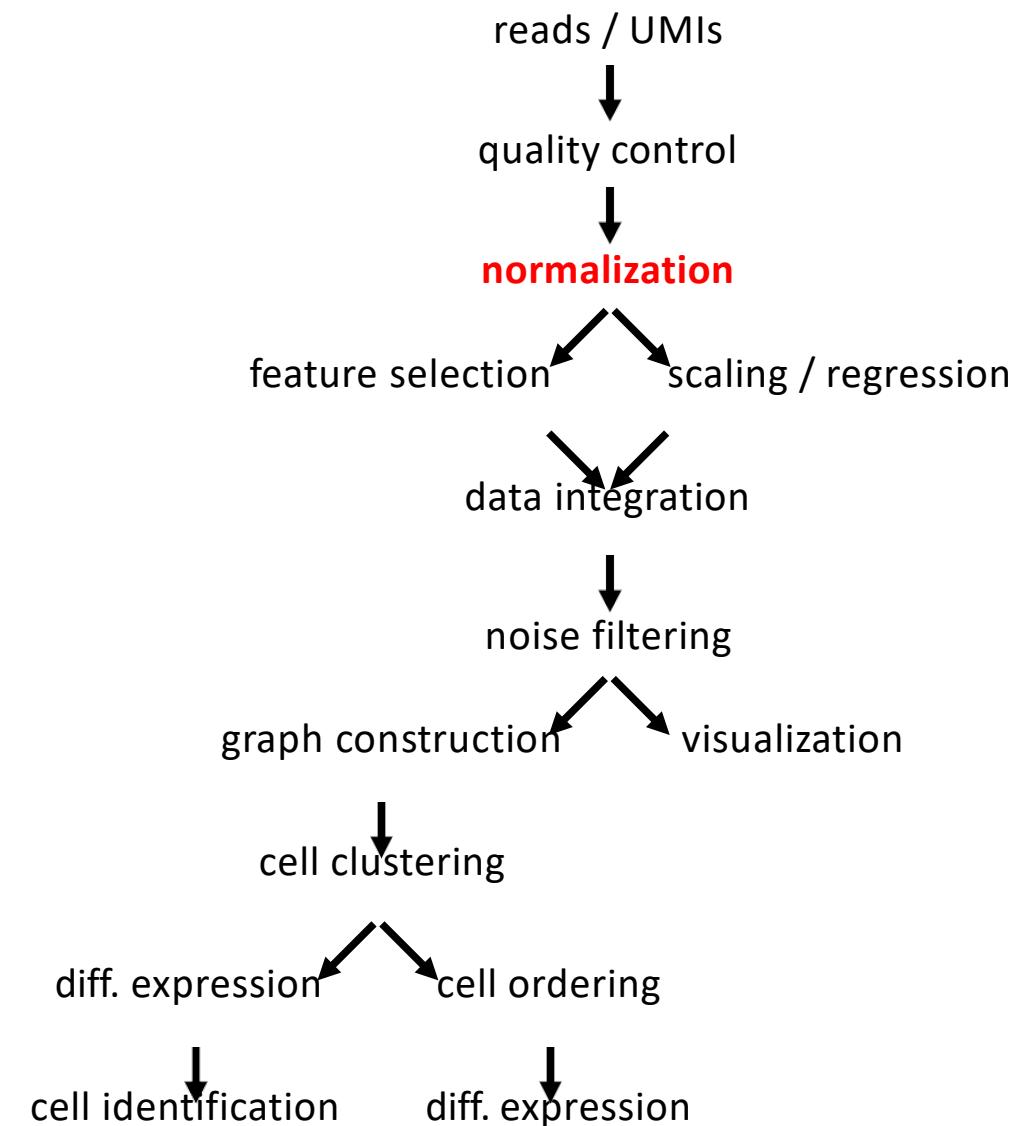


scRNA-seq quality control

- Mapping statistics (% uniquely mapping)
- mRNA-mapping read percentage
- Number of UMIs/read counts
- Protein-coding percentage
- Number of detected genes
- Spike-in detection
- Mitochondrial percentage
- Ribosomal percentage



scRNA-seq analysis workflow



scRNA-seq normalization

Count normalization (UMI and read counts)

for uneven sequencing depth

- CPM - $\log[CP10K+1]$

Gene length normalization (read counts)

for differences in gene detection due to gene length

- TPM (closer to UMI counts)
- FPKM

Drop-out rate normalization (UMI and read counts)

for differences in RNA content / drop-out rates

- Deconvolution/Scran(Pooling-Across-Cells)
- SCnorm(Expression-DepthRelation)
- SCTransform
- Census
- Linnorm
- ZINB-WaVE
- ...

bulk

$$CPM = \log\left(\frac{counts}{library_{size}} \cdot 10^6 + 1\right)$$



single-cell

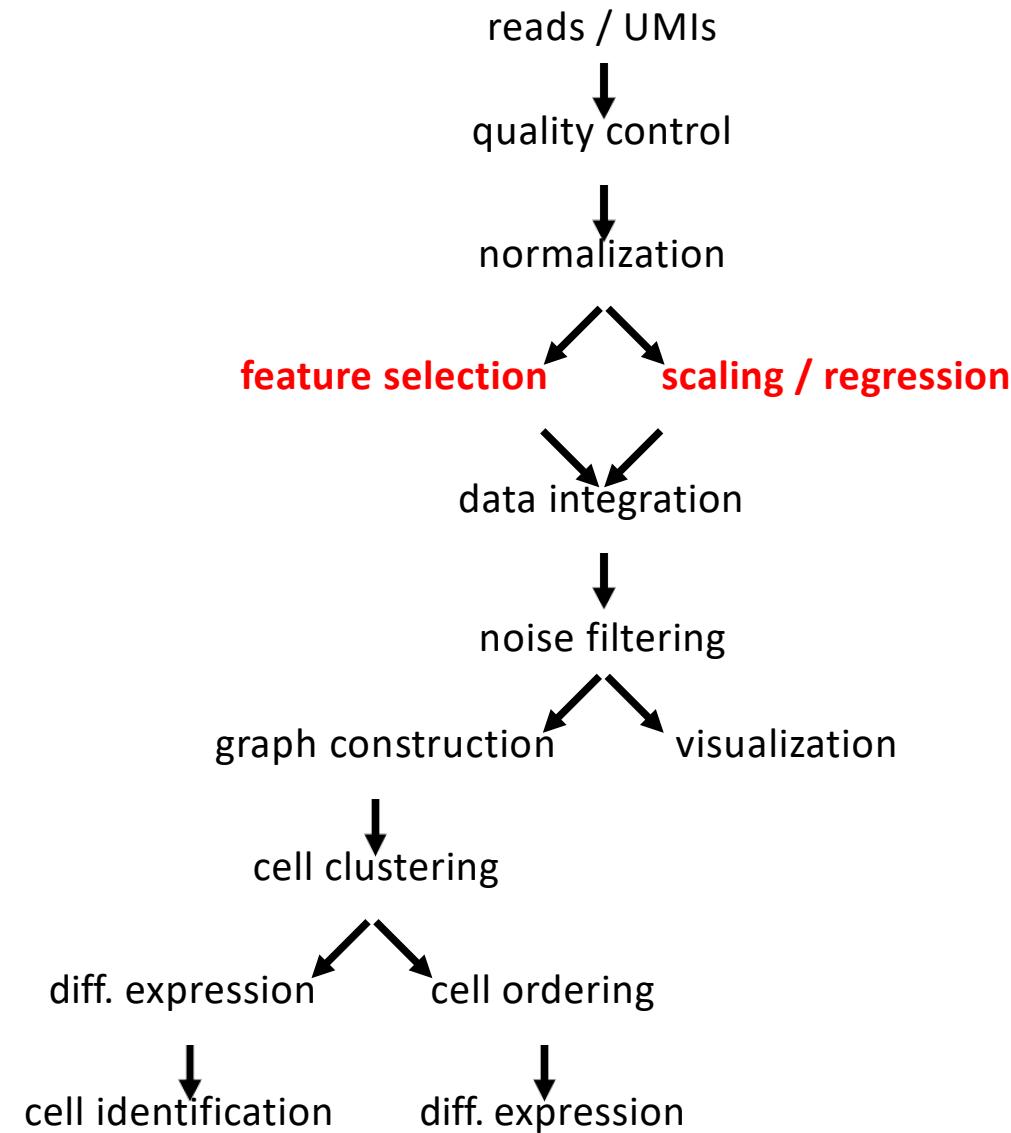
$$\log[TP10K + 1] = \log\left(\frac{counts}{library_{size}} \cdot 10\% + 1\right)$$

Most common for UMI data / fast

$$FPKM = \log\left(\frac{counts}{library_{size} ; transcript_{len}} \cdot 10\% + 1\right)$$

$$TPM = \log\left(\frac{counts}{transcript_{len}} ; \frac{10\%}{\sum \frac{counts}{transcript_{len}}} + 1\right)$$

scRNA-seq analysis workflow

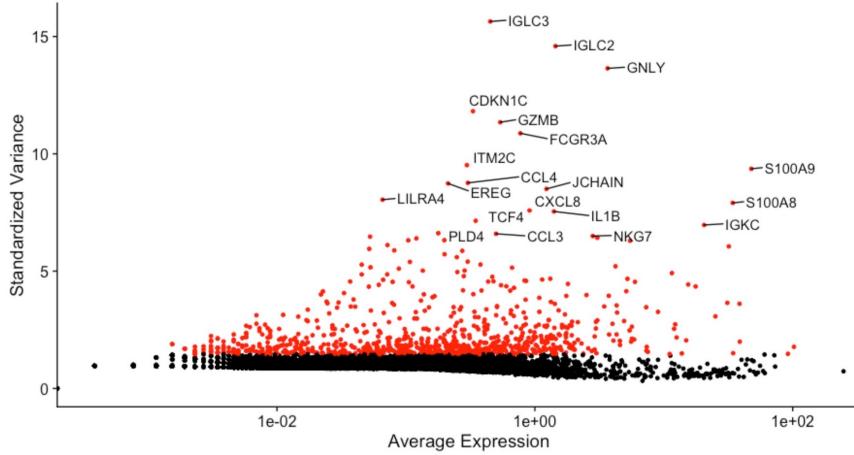


scRNA-seq feature selection

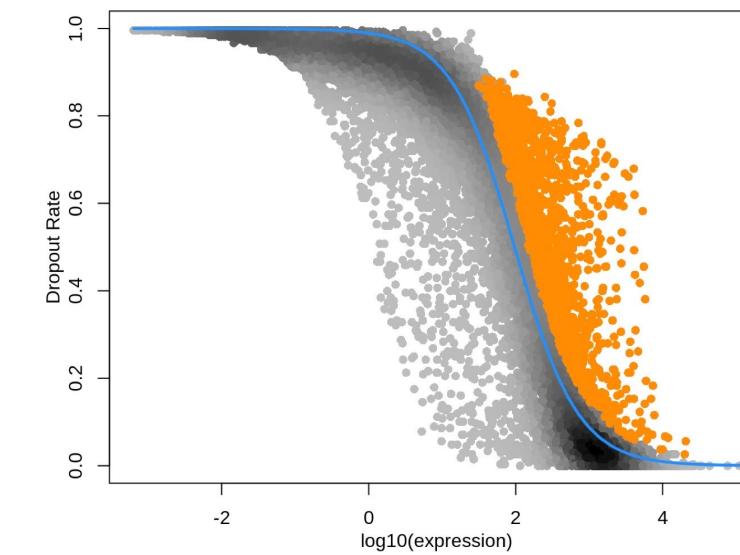
Not all genes are important to define your cell types

Hyper-variable genes are typically characterized by large differences in expression levels between cells, indicating distinct functional roles or cellular states. They can reflect diverse biological processes such as cell cycle stages, cell type-specific markers, or genes associated with cellular responses and regulatory networks.

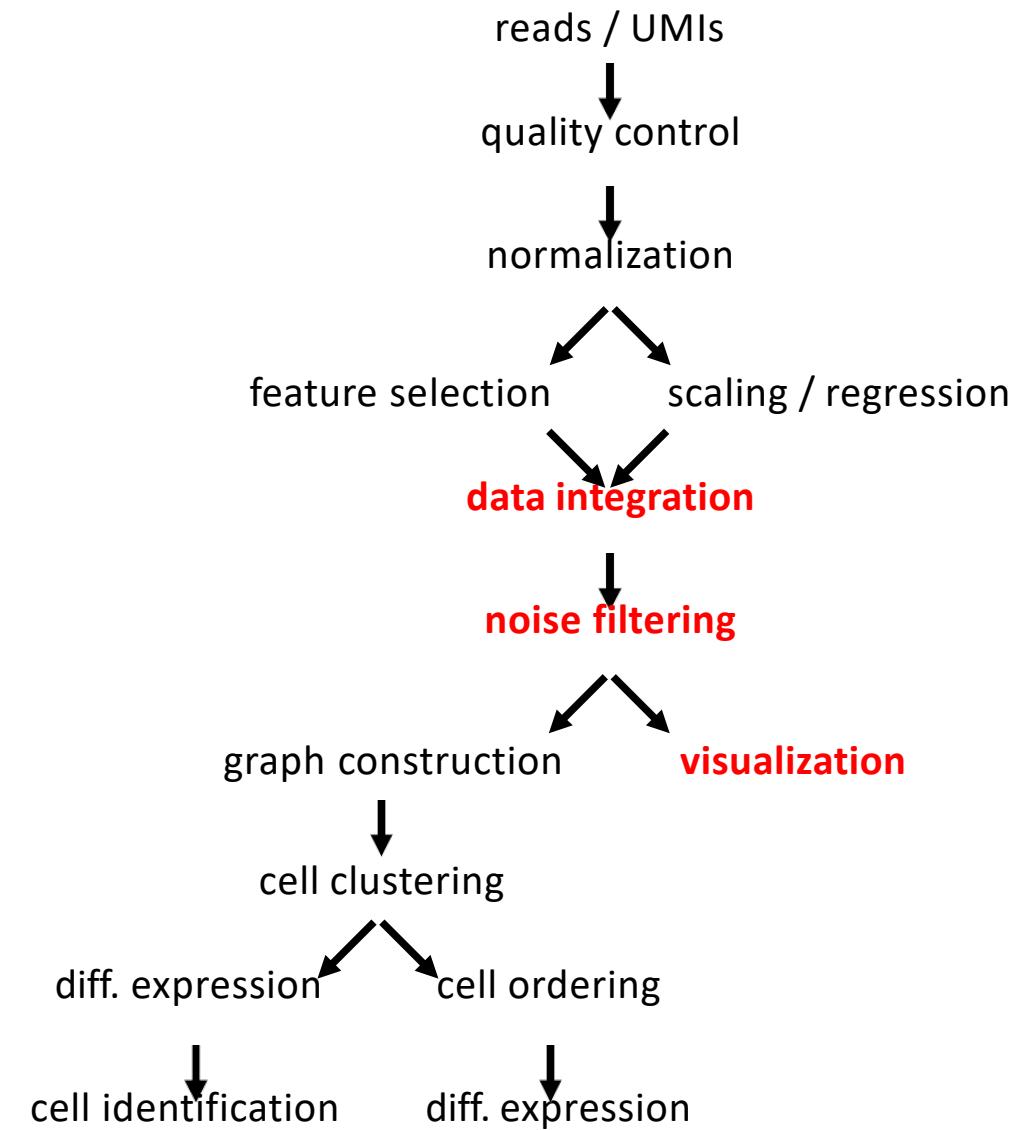
$$HVG = \frac{variance}{\log(meanExpression)}$$



$$HVG = \frac{\log(meanExpression)}{dropoutRate}$$

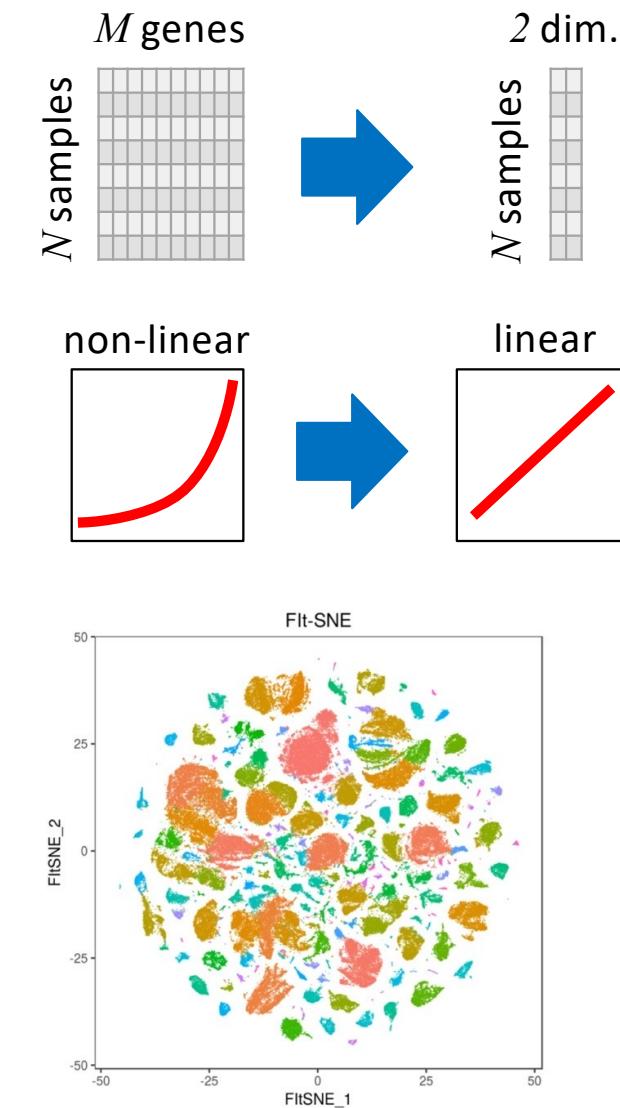


scRNA-seq analysis workflow

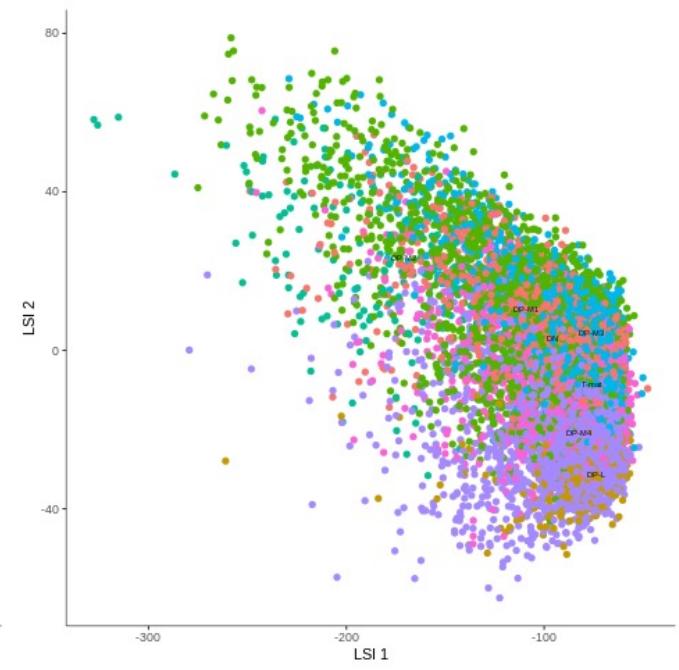
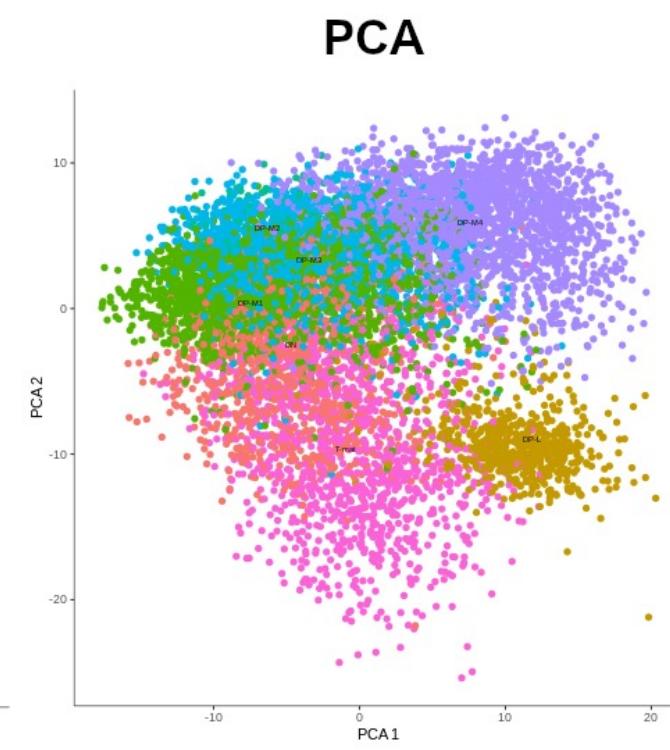
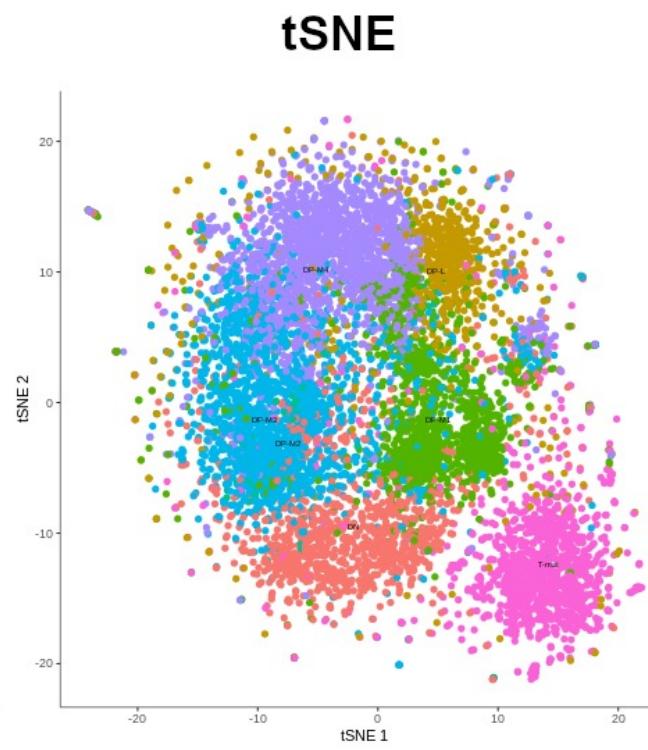
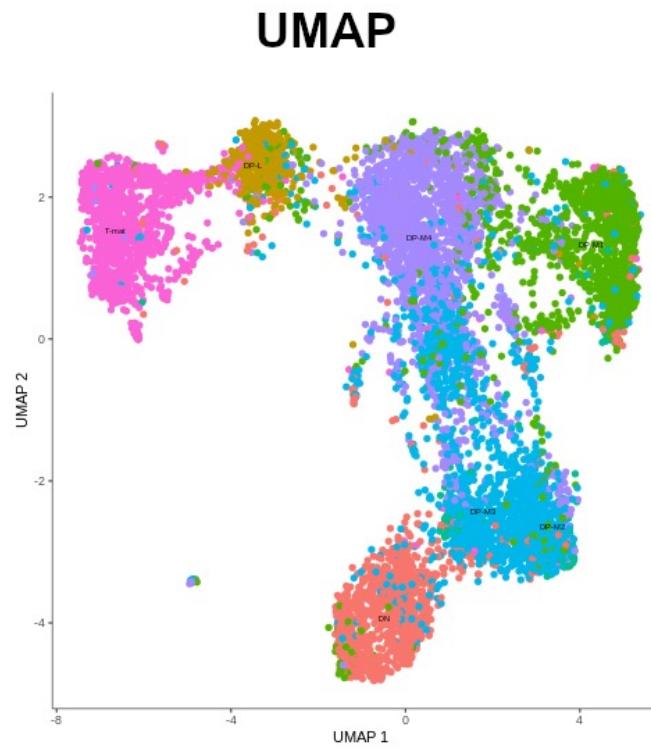


scRNA-seq dimensionality reduction

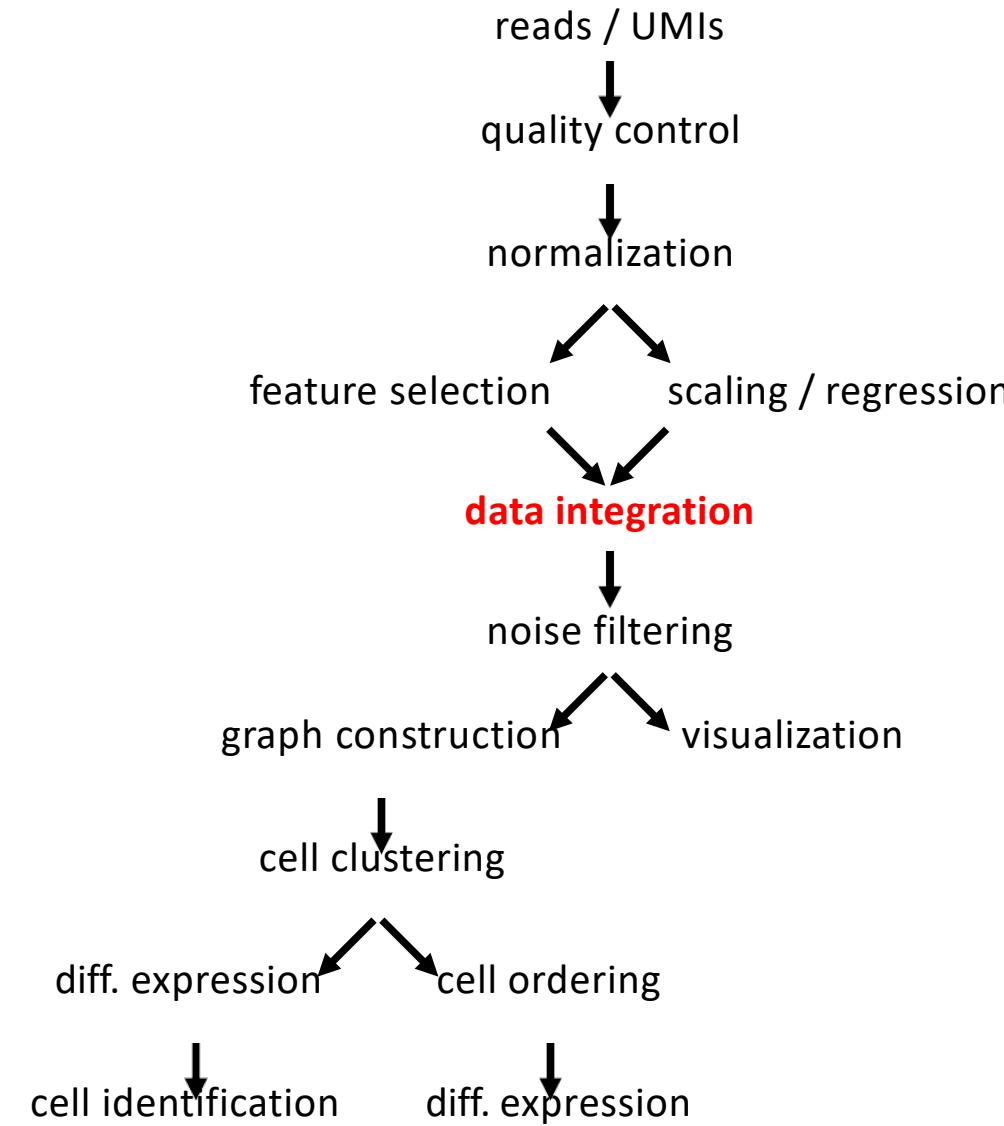
- Simplify complexity, so it becomes easier to work with. Reduce number of features (genes)
In some: Transform non-linear relationships to linear
 - “Remove” redundancies in the data
 - Identify the most relevant information (find and filter noise)
 - Reduce computational time for downstream procedures
 - Facilitate clustering, since some algorithms struggle with too many dimensions
 - Data visualization
- ... and more ...



Some dimensionality reduction methods



scRNA-seq analysis workflow



scRNA-seq data integration

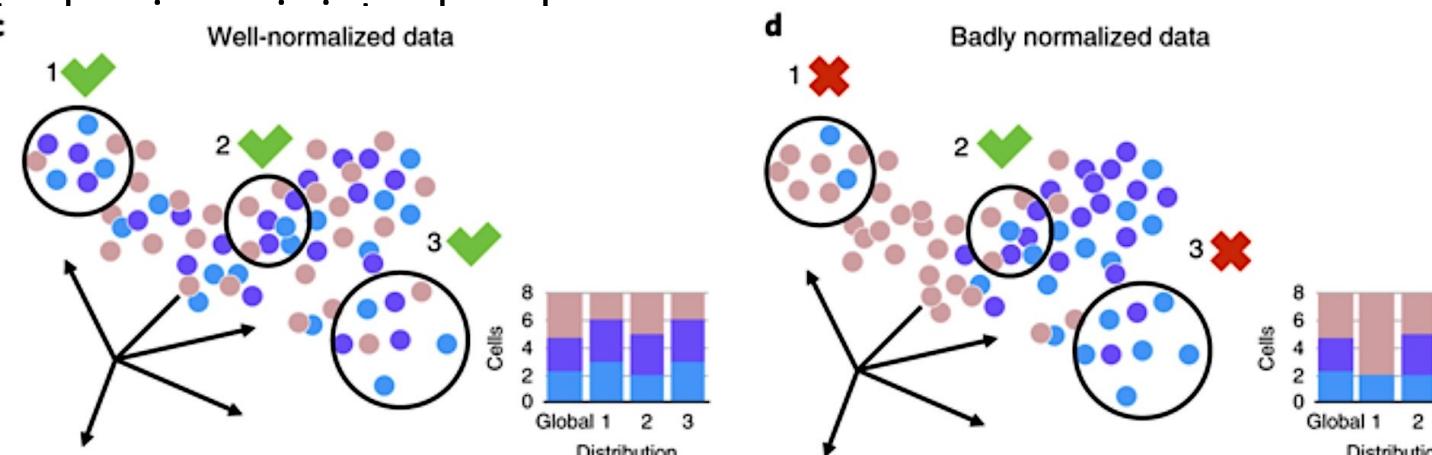
We wish to obtain corrected data where the following goals are met:

Goal:

- 1.The batch-originating variance is erased
- 2.Meaningful heterogeneity is preserved
- 3.No artefacts

What it practically means:

Similar cell types are intermixed across batches
We are not mixing distinct cell types (across or within batches)
We do not separate similar cells within batches

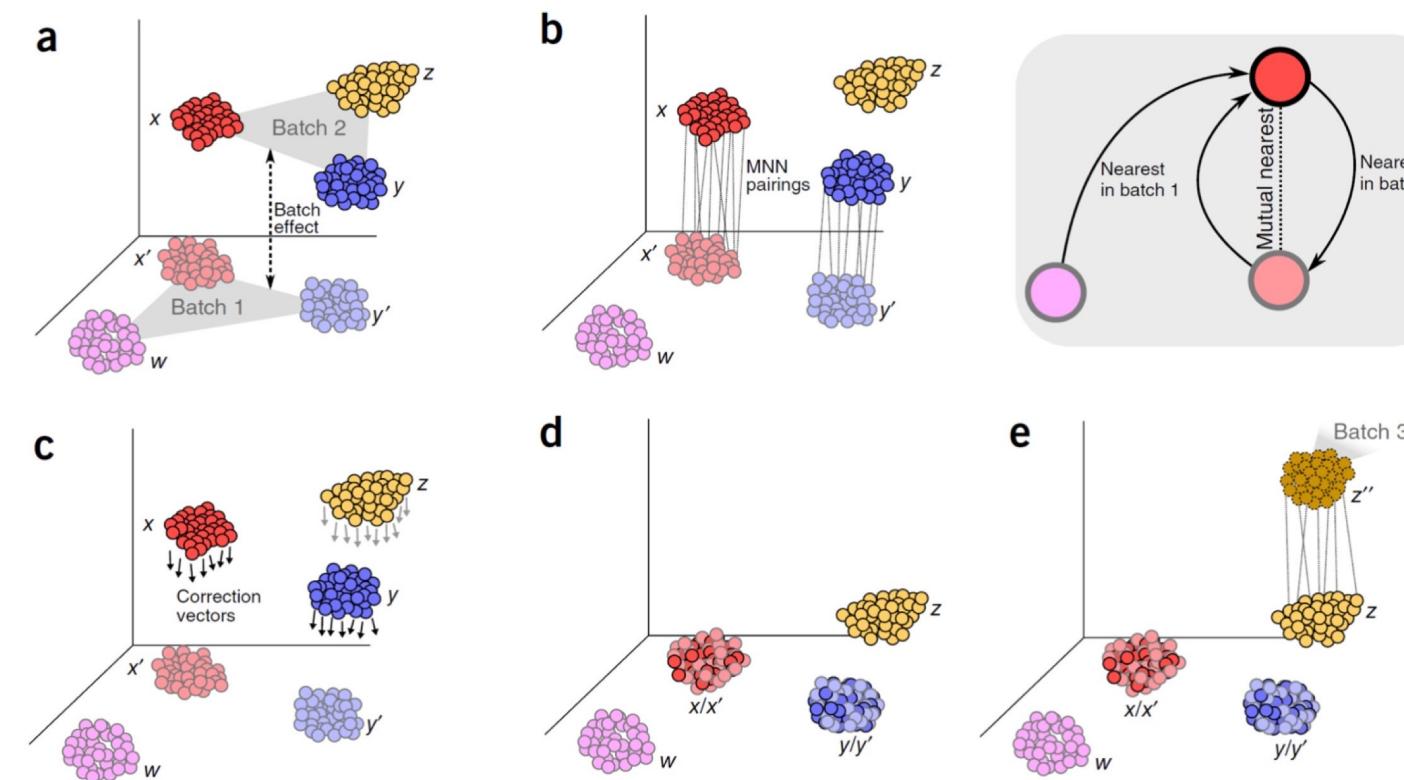


scRNA-seq analysis workflow

Regression based bulk-RNAseq batch correction methods are slow and assume the batch is constant across cells

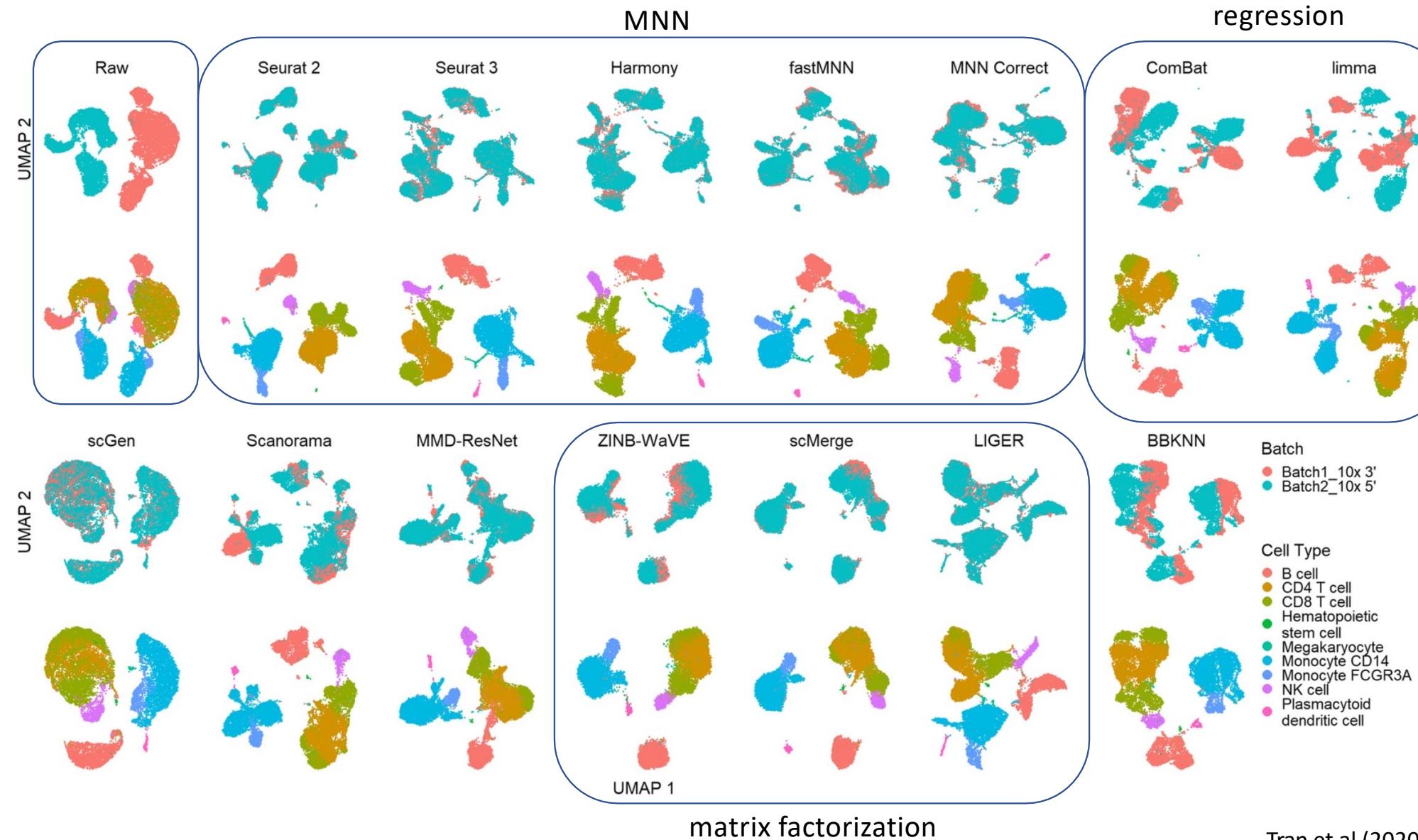
Modern data integration methods are based on the same principle:

- find MNN (mutual nearest neighbours) across datasets and correct each cell individually
- Done on a graph: much faster



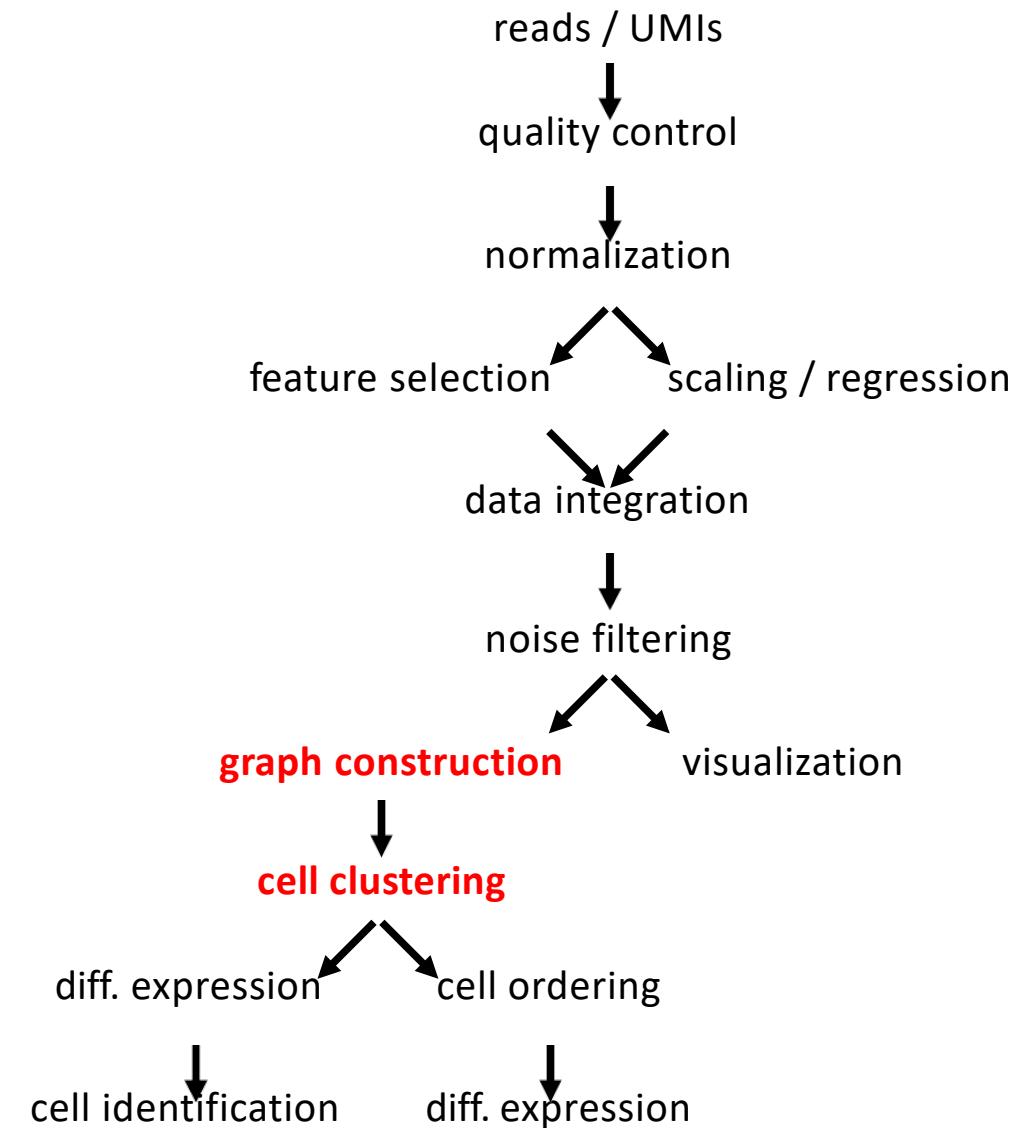
Haghverdi et al (2017) Nat Biotechnology

scRNA-seq analysis workflow

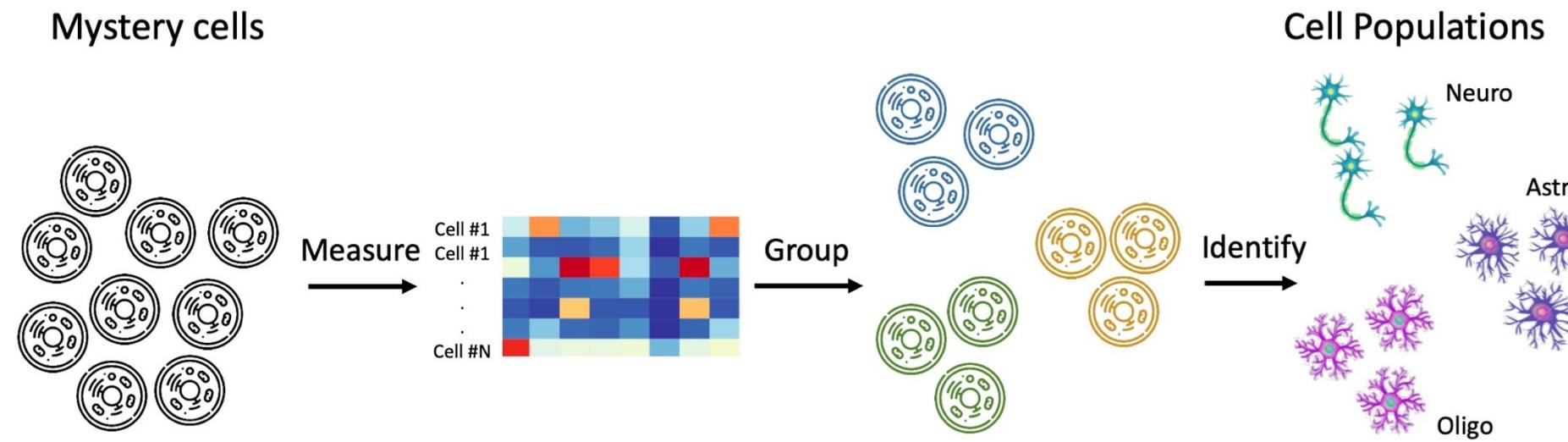


Tran et al (2020) *Genome Biology*

scRNA-seq analysis workflow



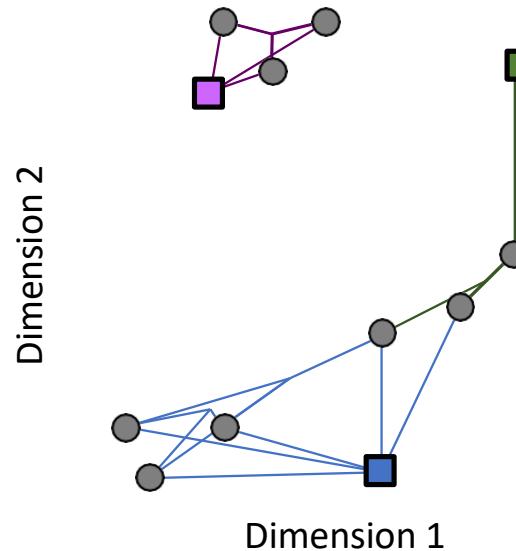
scRNA-seq clustering



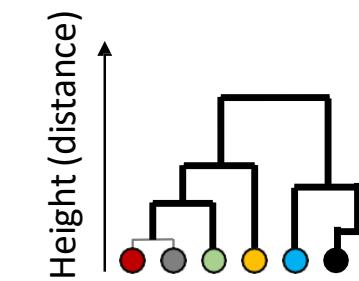
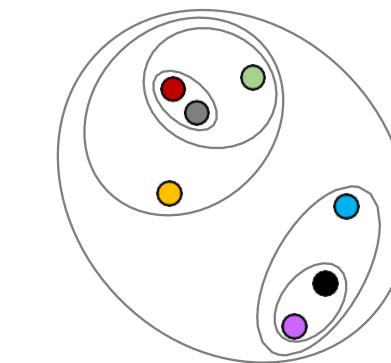
scRNA-seq clustering algorithm

K-means

$K = 3$



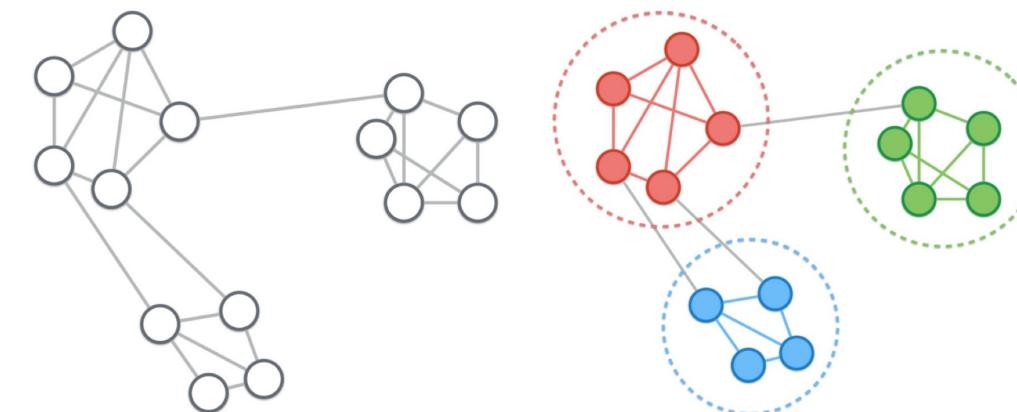
Hierarchical Clustering



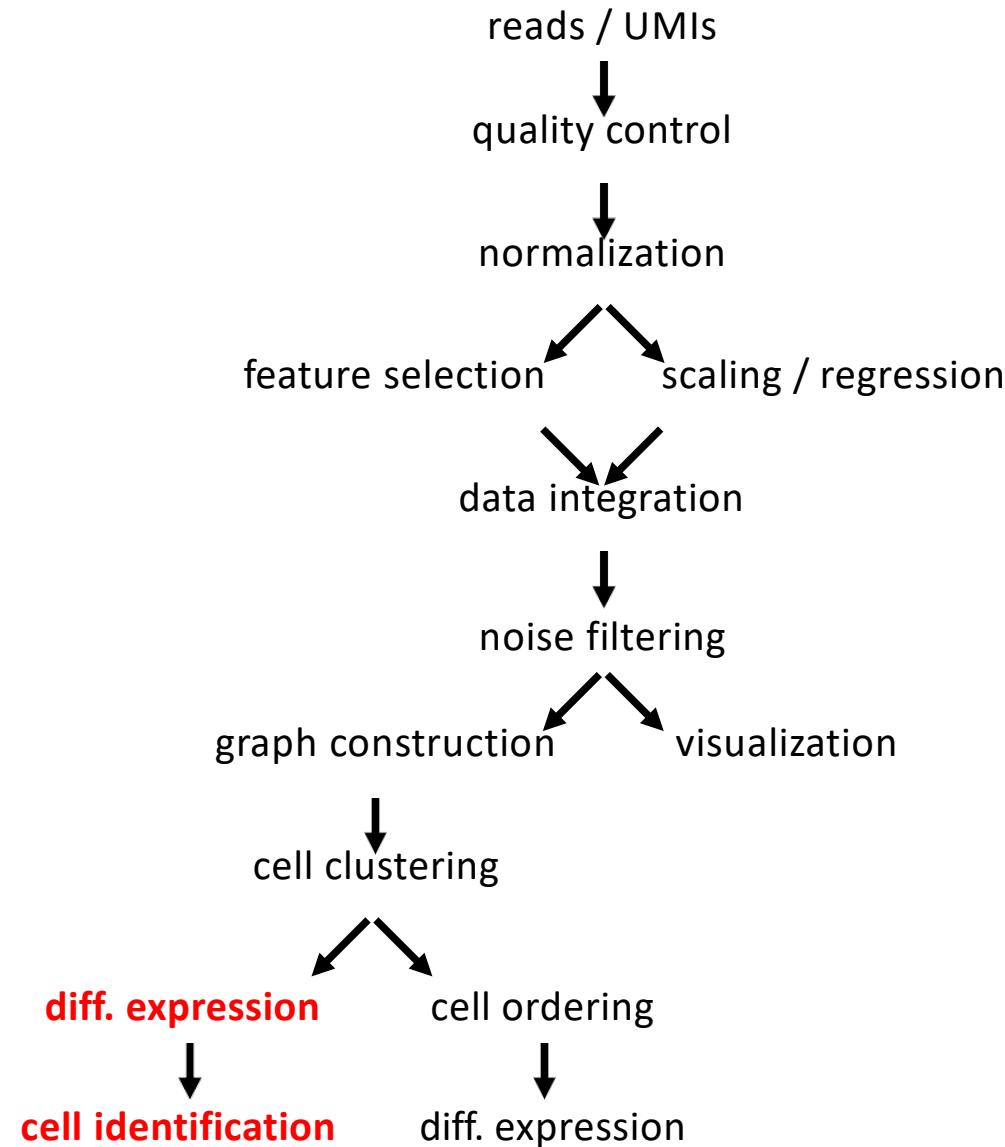
GRAPH

Louvain / Leiden community detection

Communities, or clusters, are usually groups of vertices having higher probability of being connected to each other than to members of other groups.



scRNA-seq analysis workflow

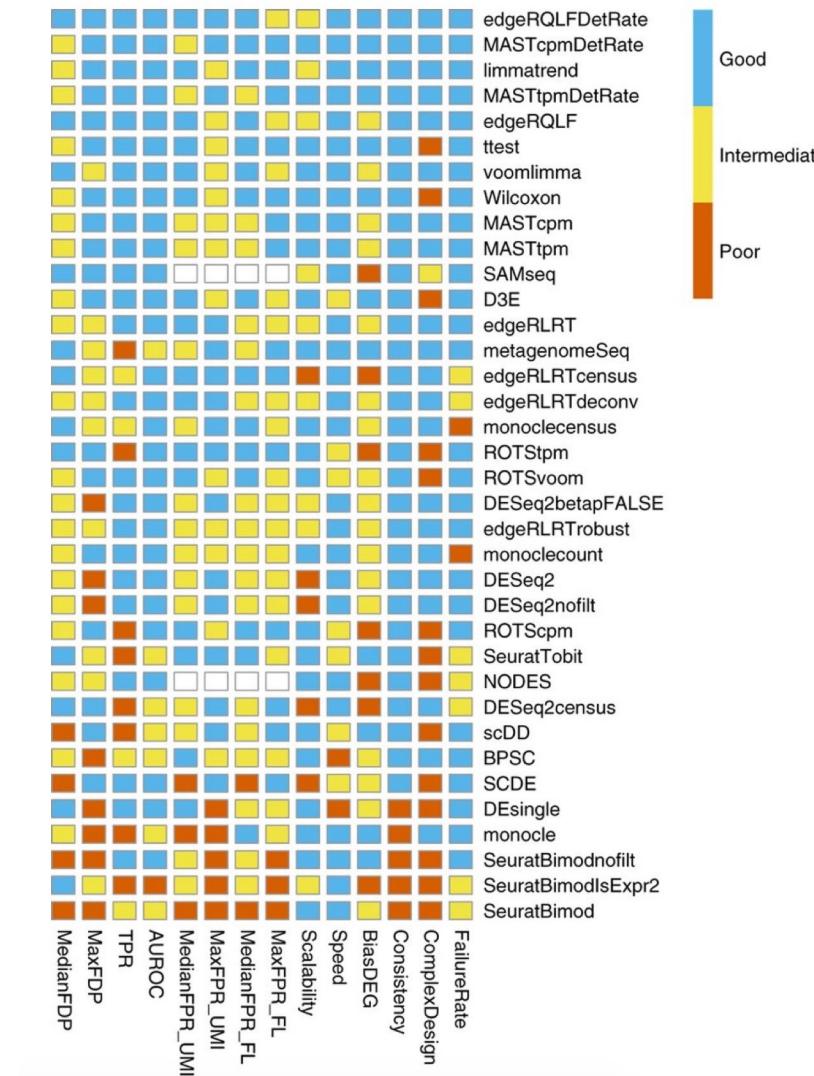


scRNA-seq differential gene expression

Typically we have more than two clusters in a data set

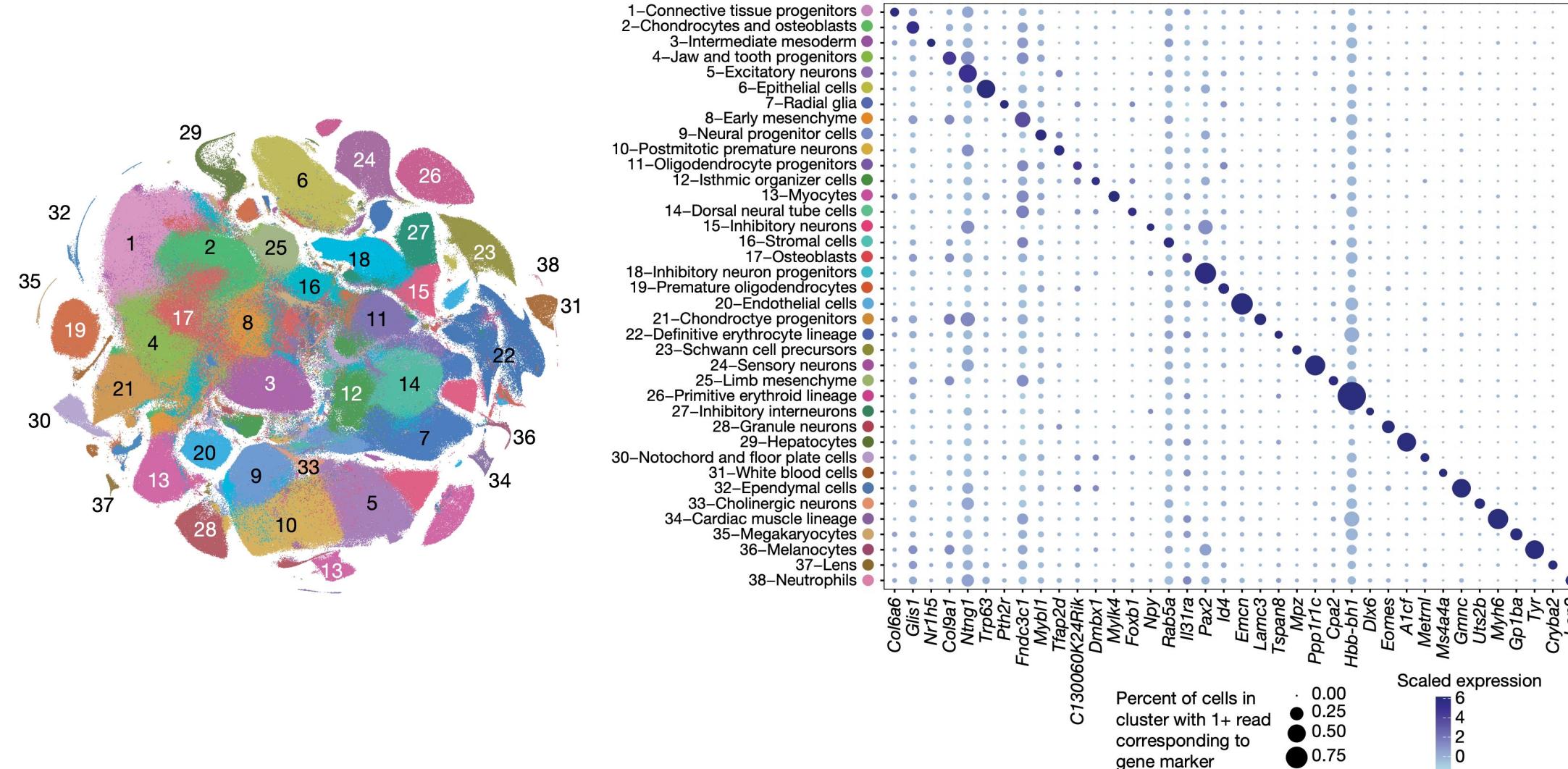
For a given cluster, are we interested in “marker genes” that are:

- **DE compared to all cells outside of the cluster (most common)**
- DE compared to at least one other cluster
- DE compared to *each* of the other clusters
- DE compared to “most” of the other clusters



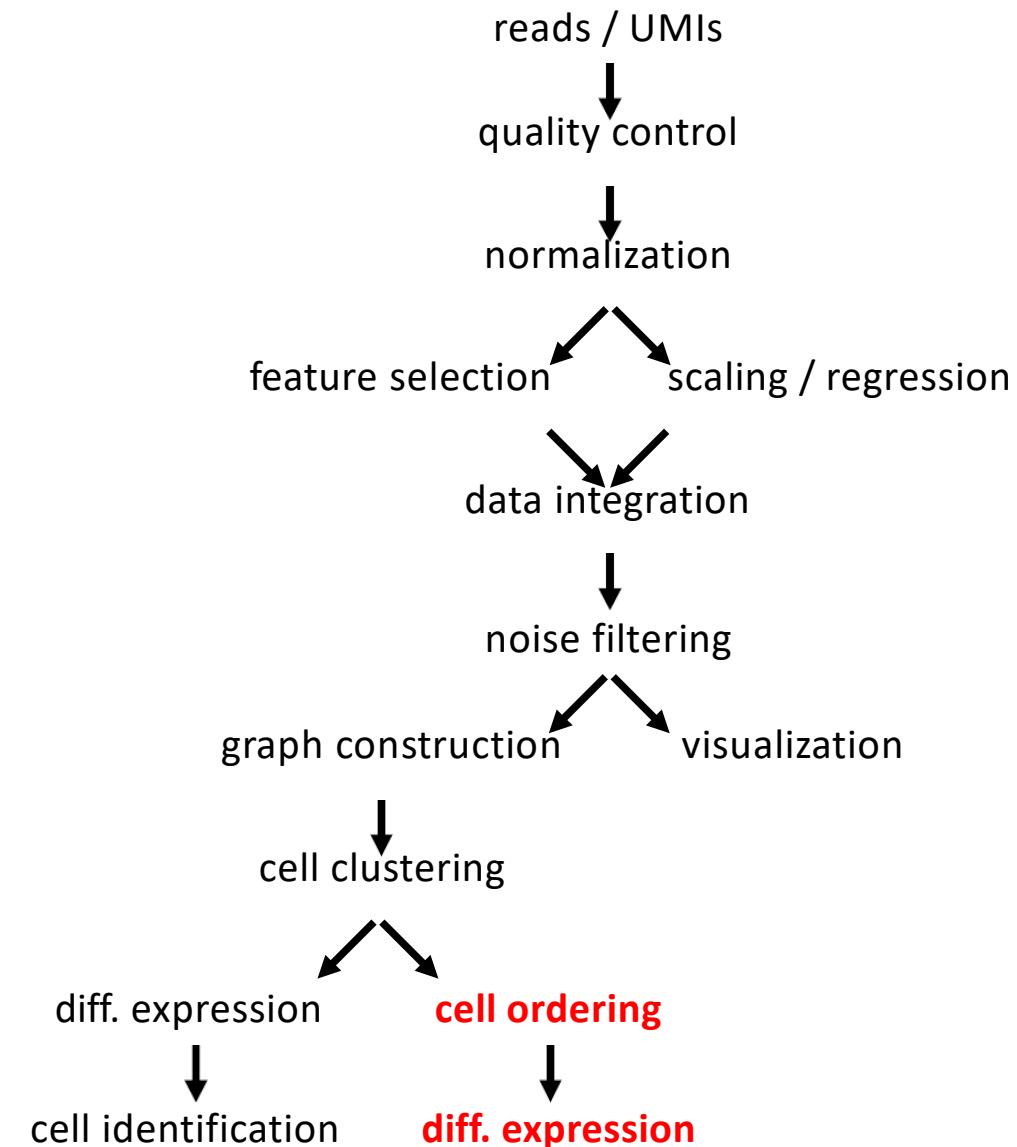
Soneson et al 2018 *Nat Methods*

scRNA-seq differential gene expression



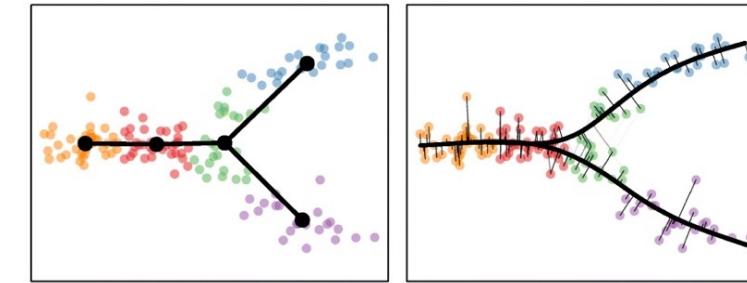
Cao et al 2019 *Nature*

scRNA-seq analysis workflow



scRNA-seq trajectory inference

Are you sure that you have a trajectory?



Street et al (2018) *BMC Genomics*

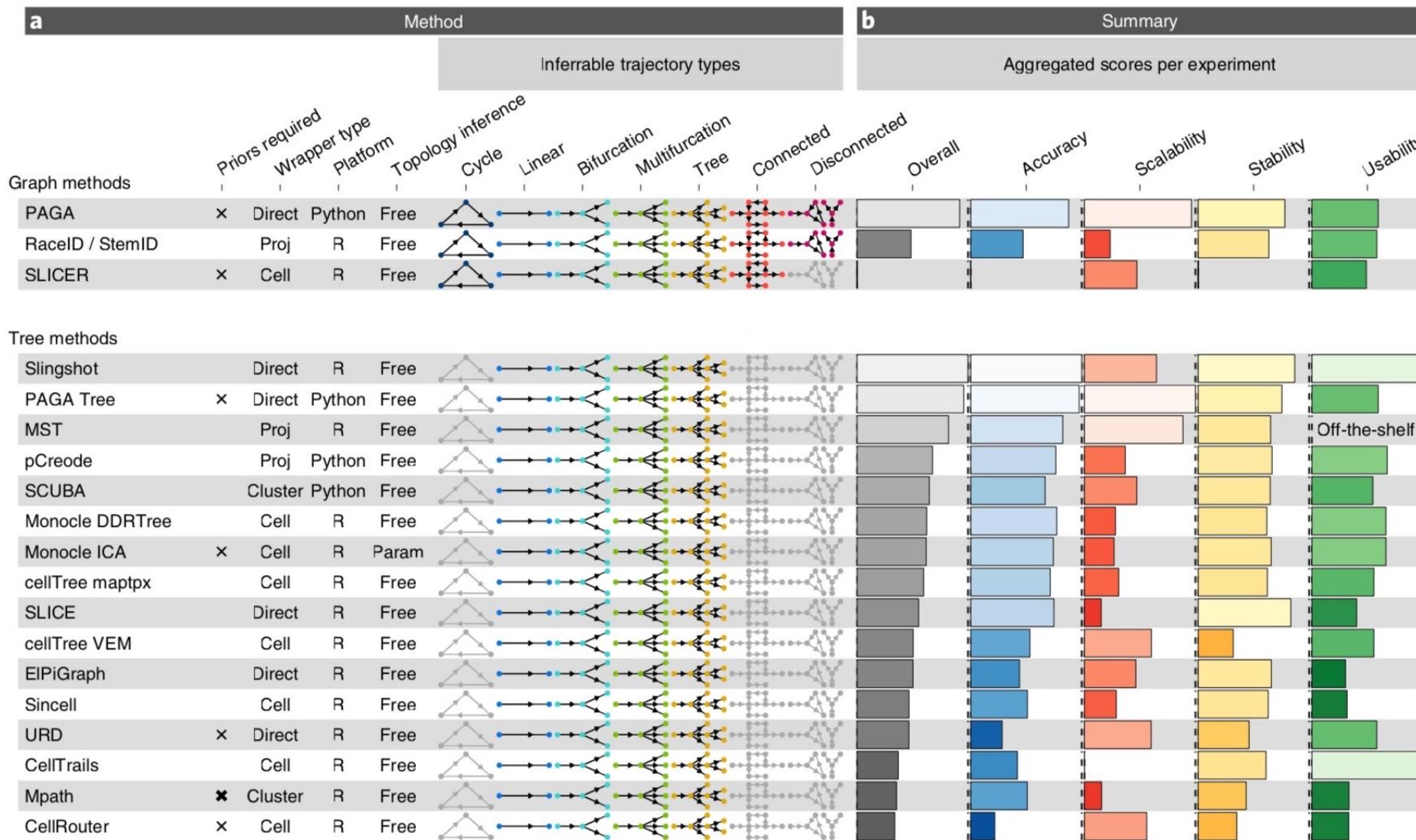
Do you have intermediate states?

Do you believe that you have branching in your trajectory?

! Be aware, any dataset can be forced into a trajectory without any biological meaning!

! First make sure that gene set and dimensionality reduction captures what you expect.

scRNA-seq trajectory inference



Public scRNA-Seq repository

Gene Expression Omnibus (GEO)

GEO is a public repository hosted by the National Center for Biotechnology Information (NCBI). It contains a wide range of genomic datasets, including single-cell datasets. You can search for single-cell datasets using keywords or browse through different studies. The website for GEO is:

www.ncbi.nlm.nih.gov/geo

Single Cell Portal

The Single Cell Portal is a comprehensive resource provided by the Broad Institute that hosts a large collection of single-cell genomics datasets. It includes datasets from various species and experimental techniques. You can search and explore datasets using the portal's interface. The website for the Single Cell Portal is:

<https://singlecell.broadinstitute.org/>

European Nucleotide Archive (ENA)

ENA is a repository for raw sequencing data and related metadata. It contains a significant number of single-cell genomics datasets. You can search for datasets using keywords or browse through different studies. The website for ENA is:

www.ebi.ac.uk/ena/browser/home

Human Cell Atlas Data Portal

HCA is an international effort to create a comprehensive reference map of all human cells. The HCA Data Portal hosts a vast collection of single-cell genomics datasets focused on human cells. You can search for datasets and explore different cell types and tissues. The website for the HCA Data Portal is:

<https://data.humancellatlas.org>

The Future of Single Cell Technology

Cost Reduction

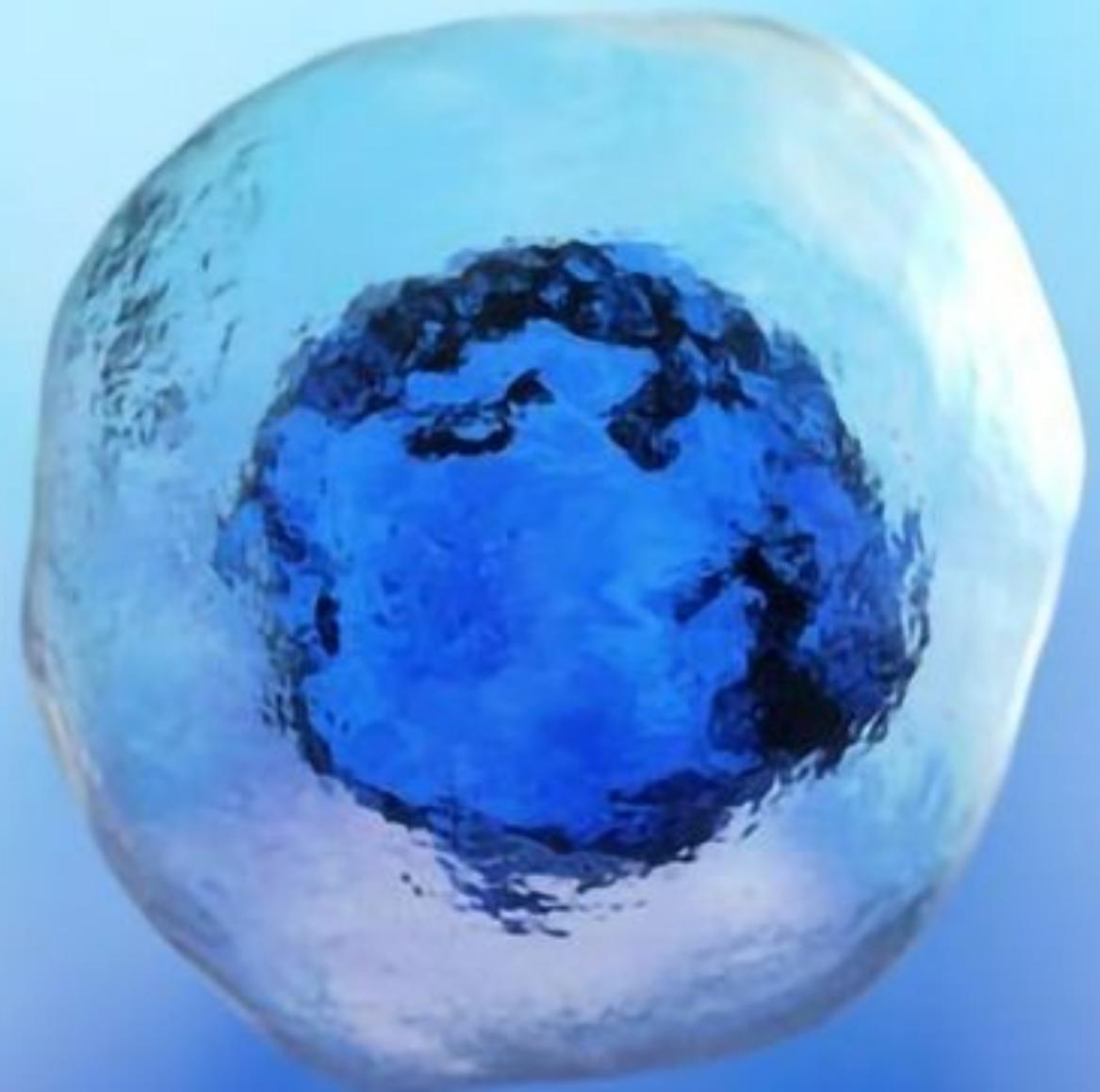
As the technology develops, it is likely that the cost of single-cell sequencing will decrease, making it more accessible for a wider range of researchers.

Data Integration

The integration of single-cell sequencing data with other types of data, such as functional and spatial data, will allow researchers to gain a more comprehensive understanding of cell function.

Real-time Analysis

The development of methods for real-time analysis of single cells in living organisms will allow researchers to better understand dynamic processes such as tissue regeneration and cancer metastasis.



Conclusion

Single Cell Technology is revolutionizing the way we study biology, and has the potential to unlock new treatments for diseases and improve our understanding of the natural world. Keep an eye on this exciting field as it continues to develop!

1. Human Cell Atlas

<https://www.humancellatlas.org/>

2. Celltypist

<https://www.celltypist.org/>

3. Human Lung Cell Atlas

<https://hlca.ds.czbiohub.org/>

4. 3CA – Curated Cancer Cell Atlas

<https://www.weizmann.ac.il/sites/3CA/>

5. Global characterization of T cells in non-small cell lung cancer by single-cell sequencing

<http://lung.cancer-pku.cn/>