An overview of the single cell -omics field

Diego A. Espinoza

Laboratory of Amit Bar-Or, MD

Immunology Graduate Group

Perelman School of Medicine

University of Pennsylvania

Aim for today

To provide an introductory overview of the current computational tools used in scRNA-seq, with an emphasis on breadth rather than depth.

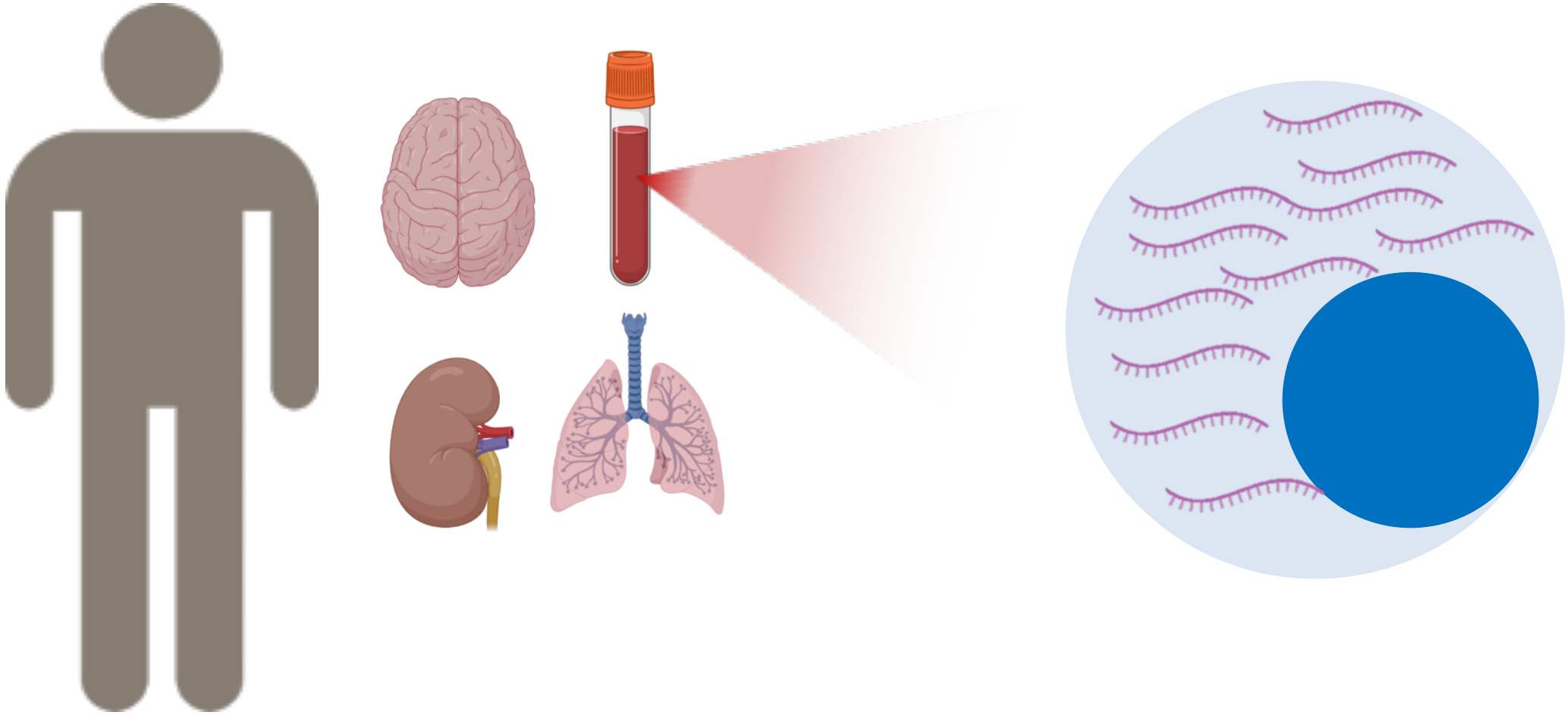
Please feel free to ask questions!

Today's presentation

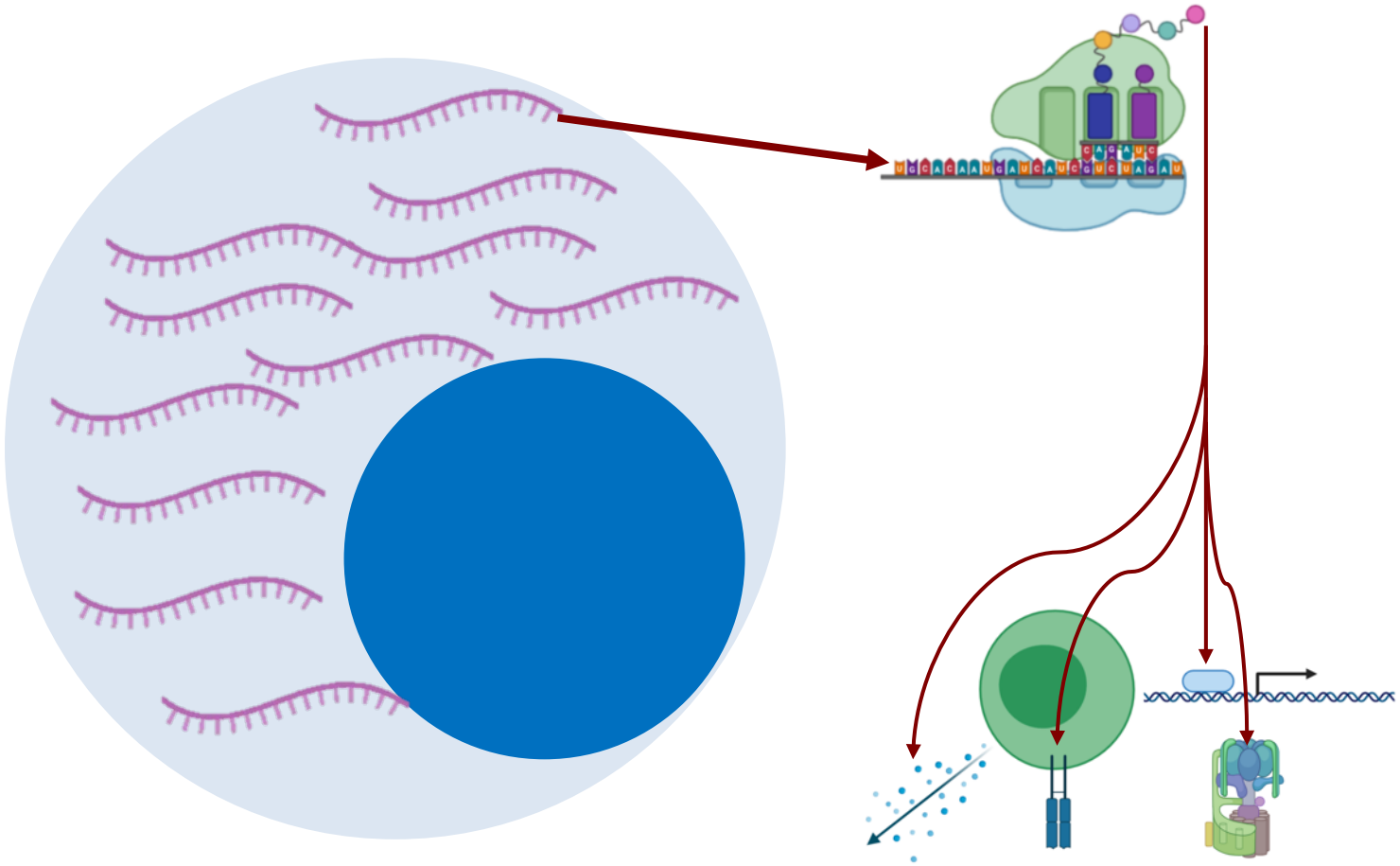
- 1) Introduction to single-cell RNA-sequencing (scRNA-seq)
  - Motivation
  - Technology
- 2) scRNA-seq of healthy tonsillar B cells
  - Motivation
  - Overview of common analysis practices
  - UMAP, t-SNE, and trajectory inference
  - Trajectory inference in B cell maturation
- 3) CITE-seq of peripheral blood cells in MS before/after treatment
  - Multimodal data
  - Integration of multimodal data
- Fin) Further reading and useful references

Introduction to scRNA-seq
Motivation

Why do we use scRNA-seq?



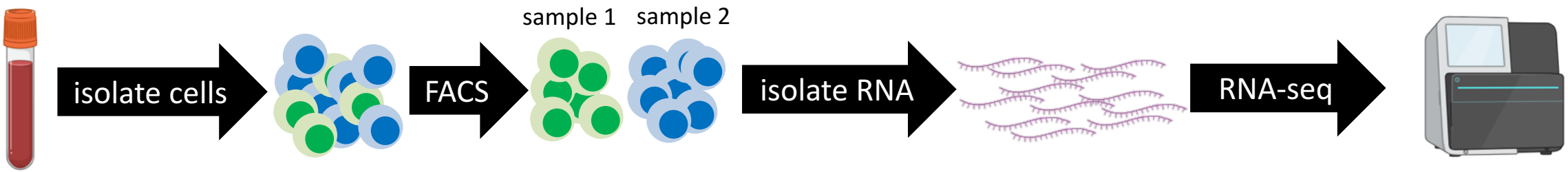
Why do we use scRNA-seq?



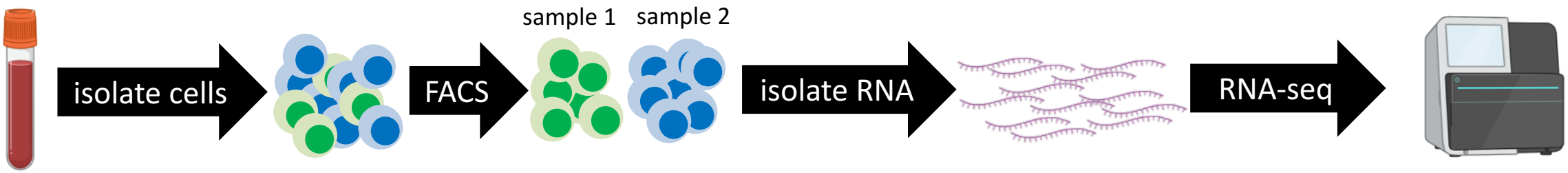
RNA ultimately provides one approach* at gaining insight into cellular function and processes

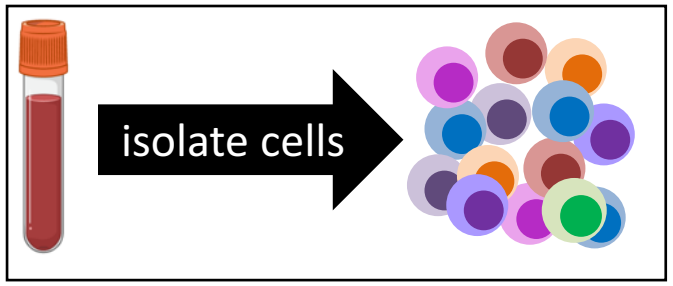
*whether this approach is sufficient to show insight depends on the context… and the reviewer 😉

Why do we use scRNA-seq?



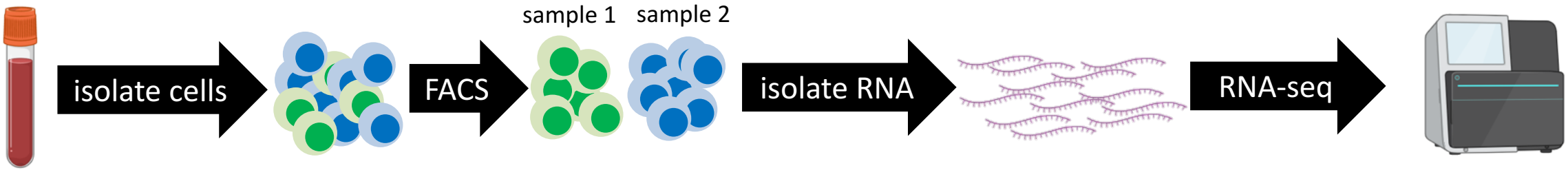
Why do we use scRNA-seq?

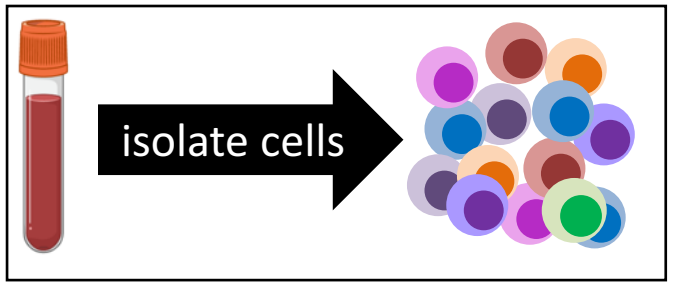




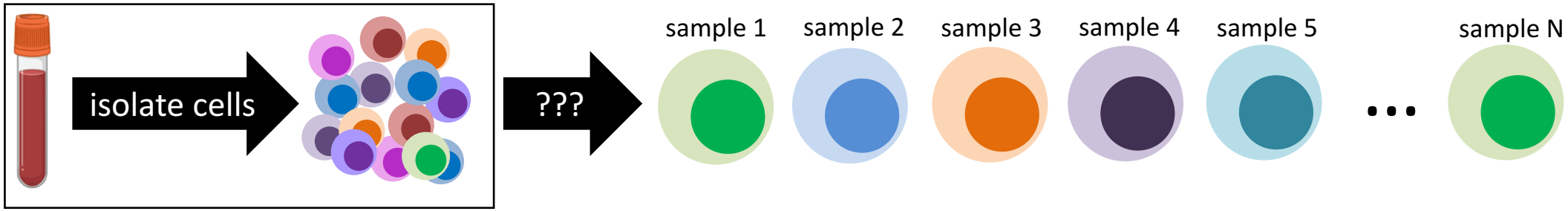
the harsh reality

Why do we use scRNA-seq?



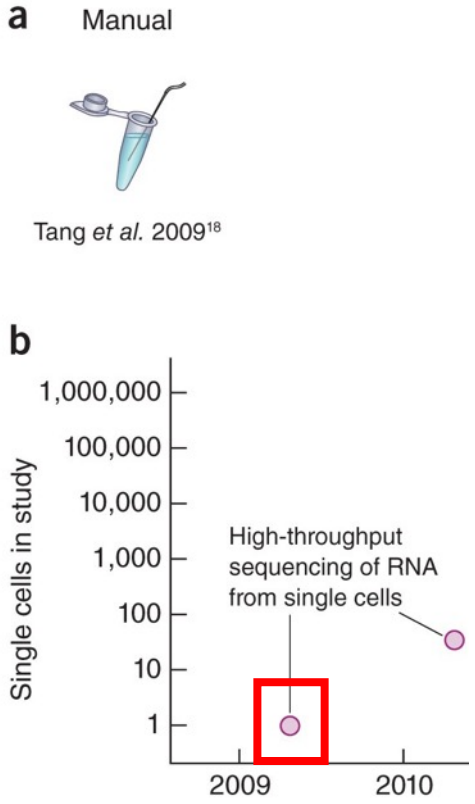


the harsh reality



Introduction to scRNA-seq
Technology

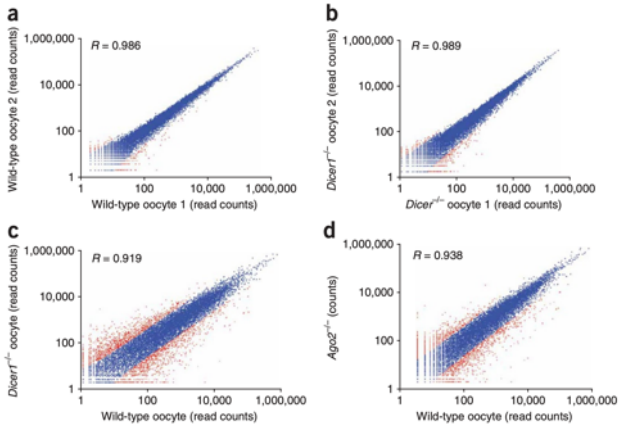
How do we perform scRNA-seq?



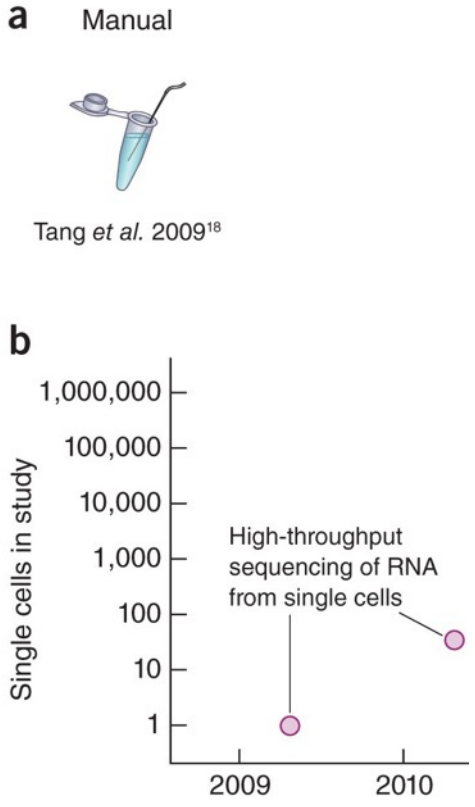
mRNA-Seq whole-transcriptome analysis of a single cell

Fuchou Tang[1,3], Catalin Barbacioru[2,3], Yangzhou Wang[2], Ellen Nordman[2], Clarence Lee[2], Nanlan Xu[2], Xiaohui Wang[2], John Bodeau[2], Brian B Tuch[2], Asim Siddiqui[2], Kaiqin Lao[2] & M Azim Surani[1]

"A single cell is manually picked under a microscope and lysed…"



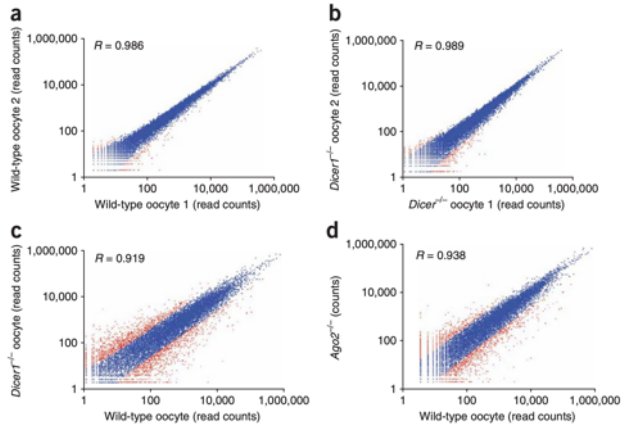
How do we perform scRNA-seq?



mRNA-Seq whole-transcriptome analysis of a single cell

Fuchou Tang[1,3], Catalin Barbacioru[2,3], Yangzhou Wang[2], Ellen Nordman[2], Clarence Lee[2], Nanlan Xu[2], Xiaohui Wang[2], John Bodeau[2], Brian B Tuch[2], Asim Siddiqui[2], Kaiqin Lao[2] & M Azim Surani[1]

"A single cell is manually picked under a microscope and lysed…"



Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 3010–3014, April 1992
Neurobiology

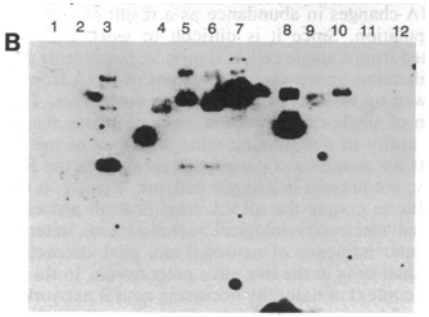
Analysis of gene expression in single live neurons

(amplified, antisense RNA/expression profile/mRNA complexity/pyramidal cell)

JAMES EBERWINE*†‡, HERMES YEH§, KEVIN MIYASHIRO*, YANXIANG CAO*, SURESH NAIR*, RICHARD FINNELL*¶, MARTHA ZETTEL§, AND PAUL COLEMAN§

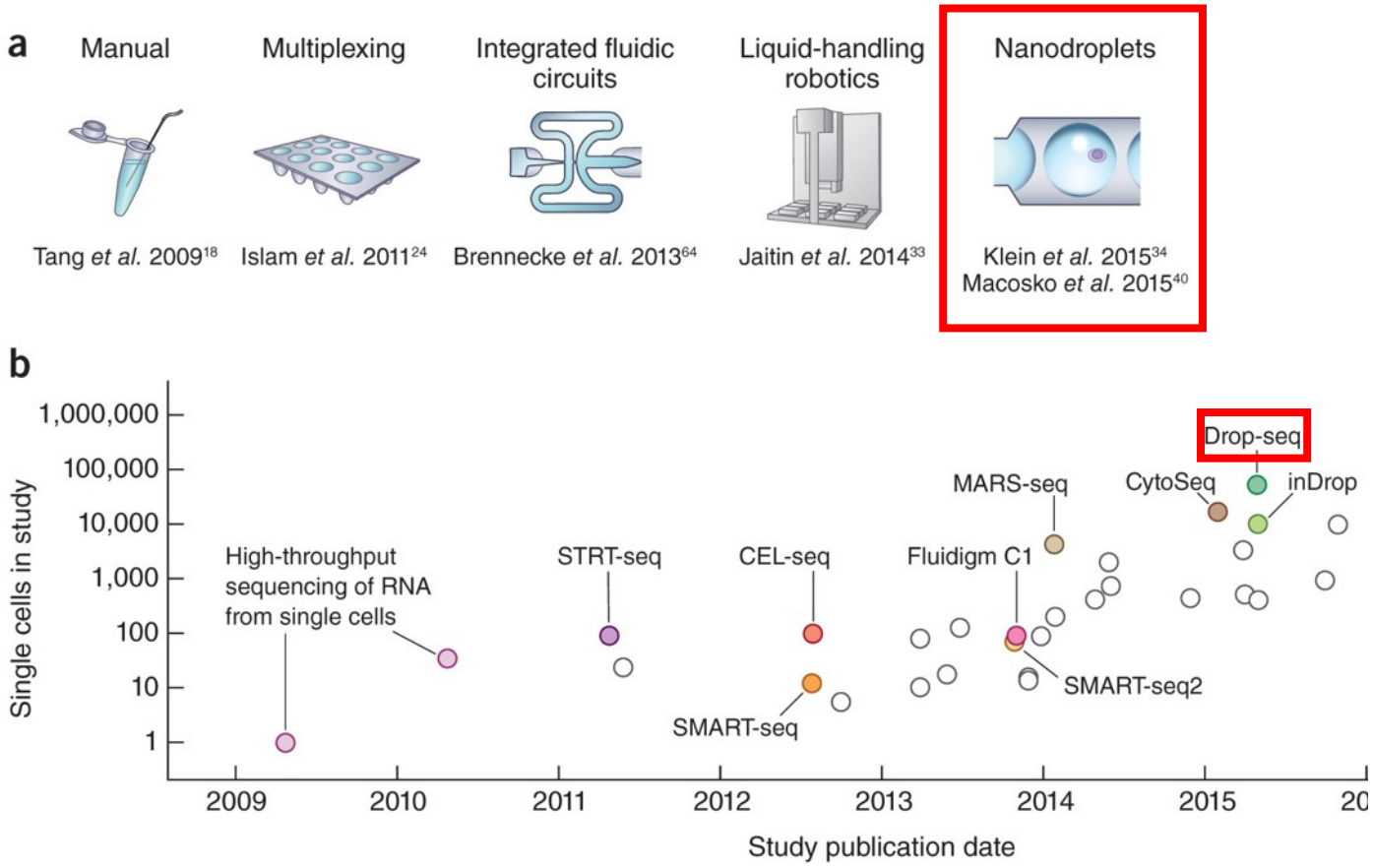
Departments of *Pharmacology and †Psychiatry, University of Pennsylvania Medical School, Philadelphia, PA 19104; and Department of §Neurobiology, Anatomy, University of Rochester Medical Center, Rochester, NY 14642

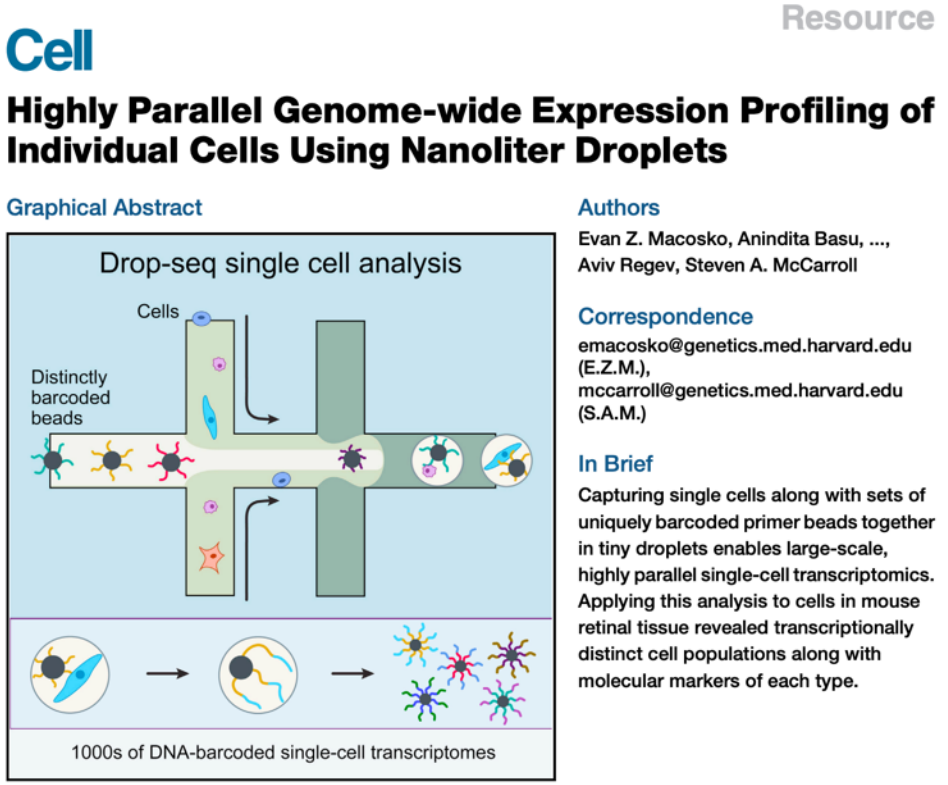
Communicated by George Koelle, December 2, 1991



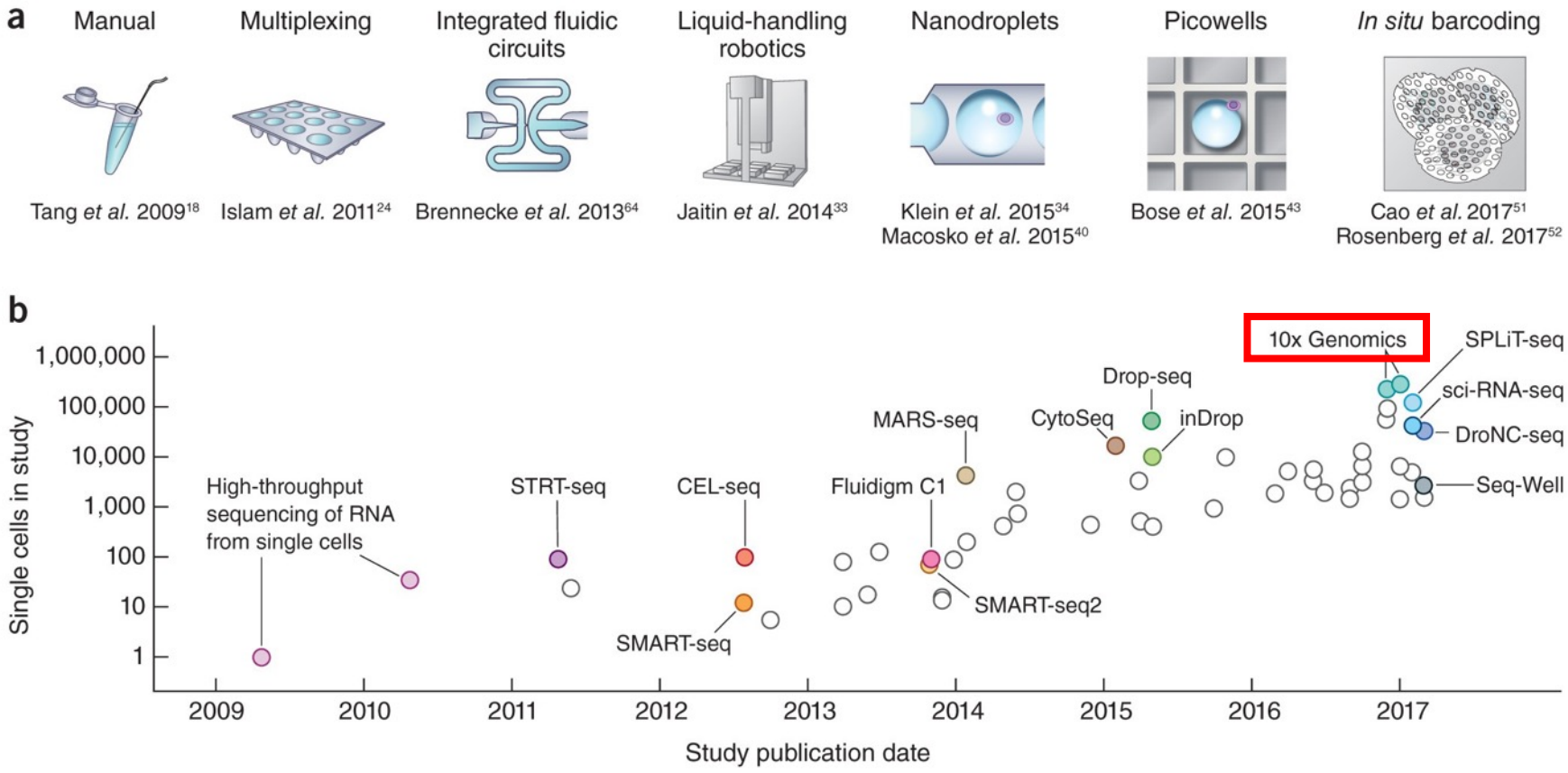
one neuron, Southern blot

How do we perform scRNA-seq?





How do we perform scRNA-seq?



10x Genomics, Inc.

Type	Public
Traded as	NASDAQ: TXG
Founded	2012
Founder	Serge Saxonov, Ben Hindson, Kevin Ness [1]
Headquarters	Pleasanton, California
Key people	Serge Saxonov (CEO) Ben Hindson (President and Chief Scientific Officer)
Revenue	▲ US$246 million (2019)
Number of employees	500
Website	10xgenomics.com

How do we perform scRNA-seq?

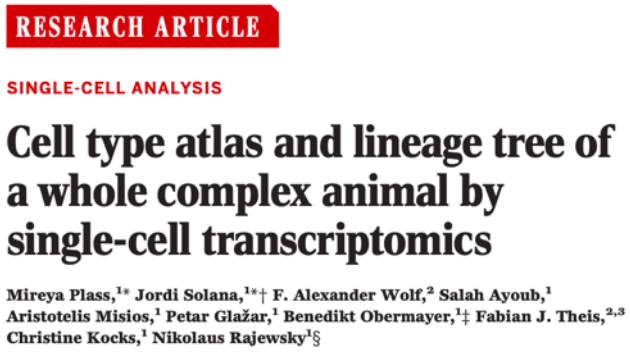
*Initially and now: re-explore difficult to study systems with these tools and build "*atlases*" or* "landscapes" *of these tissues*


RESEARCH ARTICLE SUMMARY

IMMUNOGENOMICS

A cell atlas of human thymic development defines T cell repertoire formation

Jong-Eun Park, Rachel A. Botting, Cecilia Domínguez Conde, Dorin-Mirel Popescu, Marieke Lavaert, Daniel J. Kunz, Issac Goh, Emily Stephenson, Roberta Ragazzini, Elizabeth Tuck, Anna Wilbrey-Clark, Kenny Roberts, Veronika R. Kedlian, John R. Ferdinand, Xiaoling He, Simone Webb, Daniel Maunder, Niels Vandamme, Krishnaa T. Mahbubani, Krzysztof Polanski, Lira Mamanova, Liam Bolt, David Crossland, Fabrizio de Rita, Andrew Fuller, Andrew Filby, Gary Reynolds, David Dixon, Kourosh Saeb-Parsy, Steven Lisgo, Deborah Henderson, Roser Vento-Tormo, Omer A. Bayraktar, Roger A. Barker, Kerstin B. Meyer, Yvan Saeys, Paola Bonfanti, Sam Behjati, Menna R. Clatworthy, Tom Taghon*, Muzlifah Haniffa*, Sarah A. Teichmann*


RESEARCH ARTICLE

SINGLE-CELL ANALYSIS

Cell type atlas and lineage tree of a whole complex animal by single-cell transcriptomics

Mireya Plass,[1*] Jordi Solana,[1*†] F. Alexander Wolf,[2] Salah Ayoub,[1] Aristotelis Misios,[1] Petar Glažar,[1] Benedikt Obermayer,[1‡] Fabian J. Theis,[2,3] Christine Kocks,[1] Nikolaus Rajewsky[1§]


nature

Explore Content | Journal Information | Publish With Us

nature > articles > article

Article | Published: 25 March 2020

Construction of a human cell landscape at single-cell level

Xiaoping Han, Ziming Zhou, [...] Guoji Guo

Nature 581, 303–309(2020) | Cite this article

58k Accesses | 53 Citations | 406 Altmetric | Metrics


nature

Explore Content | Journal Information | Publish With Us

nature > articles > article

Article | Open Access | Published: 24 September 2020

Cells of the adult human heart

Monika Litviňuková, Carlos Talavera-López, [...] Sarah A. Teichmann

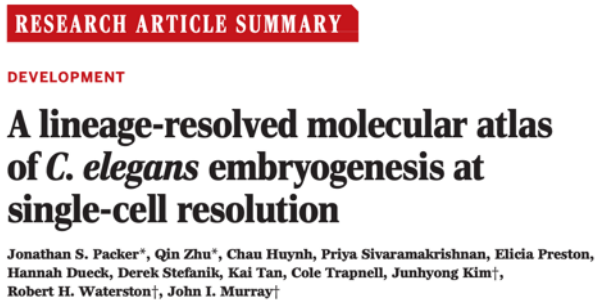
Nature 588, 466–472(2020) | Cite this article

56k Accesses | 11 Citations | 1324 Altmetric | Metrics


Cell

Article

Single-cell landscape of the ecosystem in early-relapse hepatocellular carcinoma


RESEARCH ARTICLE SUMMARY

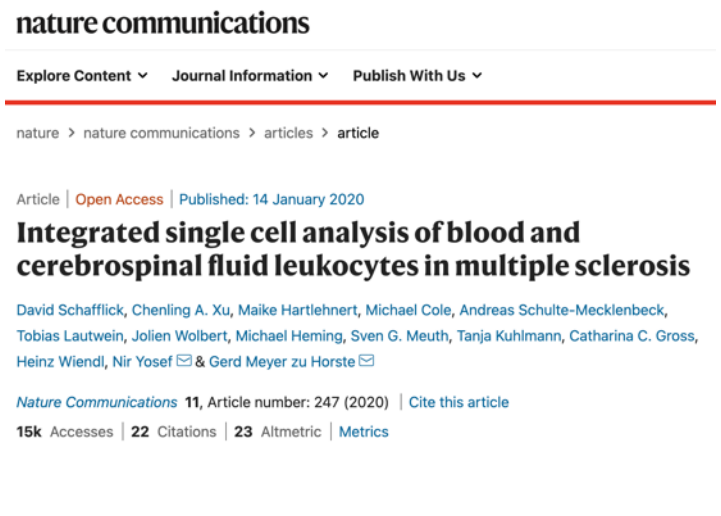
DEVELOPMENT

A lineage-resolved molecular atlas of *C. elegans* embryogenesis at single-cell resolution

Jonathan S. Packer*, Qin Zhu*, Chau Huynh, Priya Sivaramakrishnan, Elicia Preston, Hannah Dueck, Derek Stefanik, Kai Tan, Cole Trapnell, Junhyong Kim†, Robert H. Waterston†, John I. Murray†

How do we perform scRNA-seq?

Emerging: leverage these tools for more traditional "hypothesis-driven" questions



nature communications

Explore Content | Journal Information | Publish With Us

nature > nature communications > articles > article

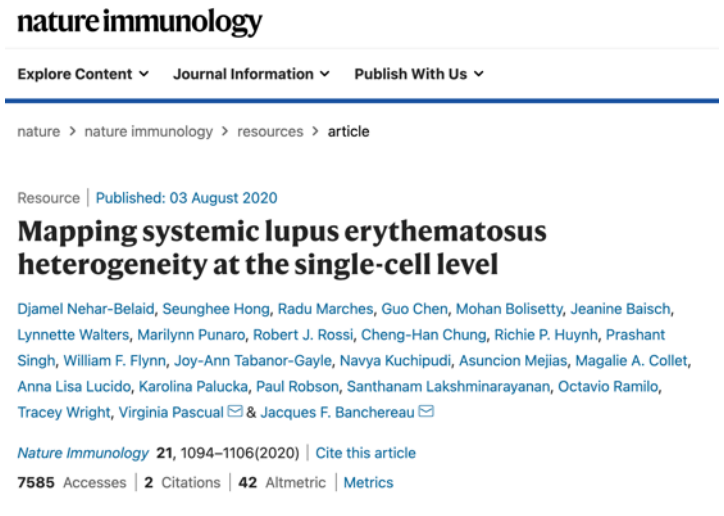
Article | Open Access | Published: 14 January 2020

Integrated single cell analysis of blood and cerebrospinal fluid leukocytes in multiple sclerosis

David Schafflick, Chenling A. Xu, Maike Hartlehnert, Michael Cole, Andreas Schulte-Mecklenbeck, Tobias Lautwein, Jolien Wolbert, Michael Heming, Sven G. Meuth, Tanja Kuhlmann, Catharina C. Gross, Heinz Wiendl, Nir Yosef & Gerd Meyer zu Horste

Nature Communications **11**, Article number: 247 (2020) | Cite this article

15k Accesses | **22** Citations | **23** Altmetric | Metrics



nature immunology

Explore Content | Journal Information | Publish With Us

nature > nature immunology > resources > article

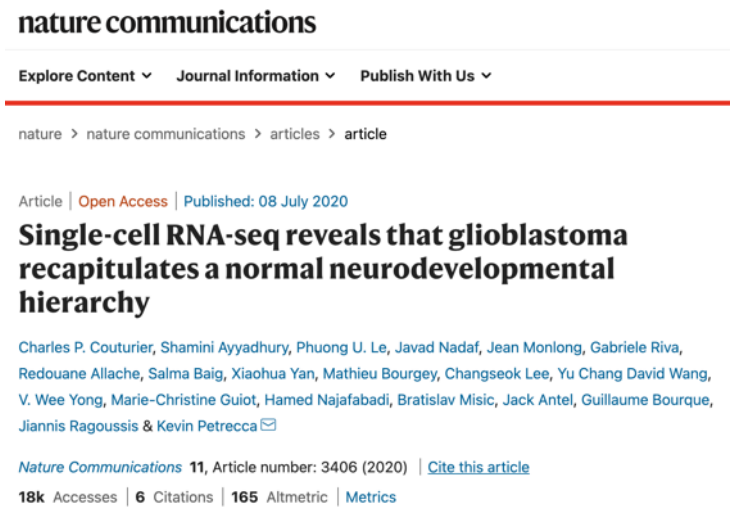
Resource | Published: 03 August 2020

Mapping systemic lupus erythematosus heterogeneity at the single-cell level

Djamel Nehar-Belaid, Seunghee Hong, Radu Marches, Guo Chen, Mohan Bolisetty, Jeanine Baisch, Lynnette Walters, Marilynn Punaro, Robert J. Rossi, Cheng-Han Chung, Richie P. Huynh, Prashant Singh, William F. Flynn, Joy-Ann Tabanor-Gayle, Navya Kuchipudi, Asuncion Mejias, Magalie A. Collet, Anna Lisa Lucido, Karolina Palucka, Paul Robson, Santhanam Lakshminarayanan, Octavio Ramilo, Tracey Wright, Virginia Pascual & Jacques F. Banchereau

Nature Immunology **21**, 1094–1106(2020) | Cite this article

7585 Accesses | **2** Citations | **42** Altmetric | Metrics



nature communications

Explore Content | Journal Information | Publish With Us

nature > nature communications > articles > article

Article | Open Access | Published: 08 July 2020

Single-cell RNA-seq reveals that glioblastoma recapitulates a normal neurodevelopmental hierarchy

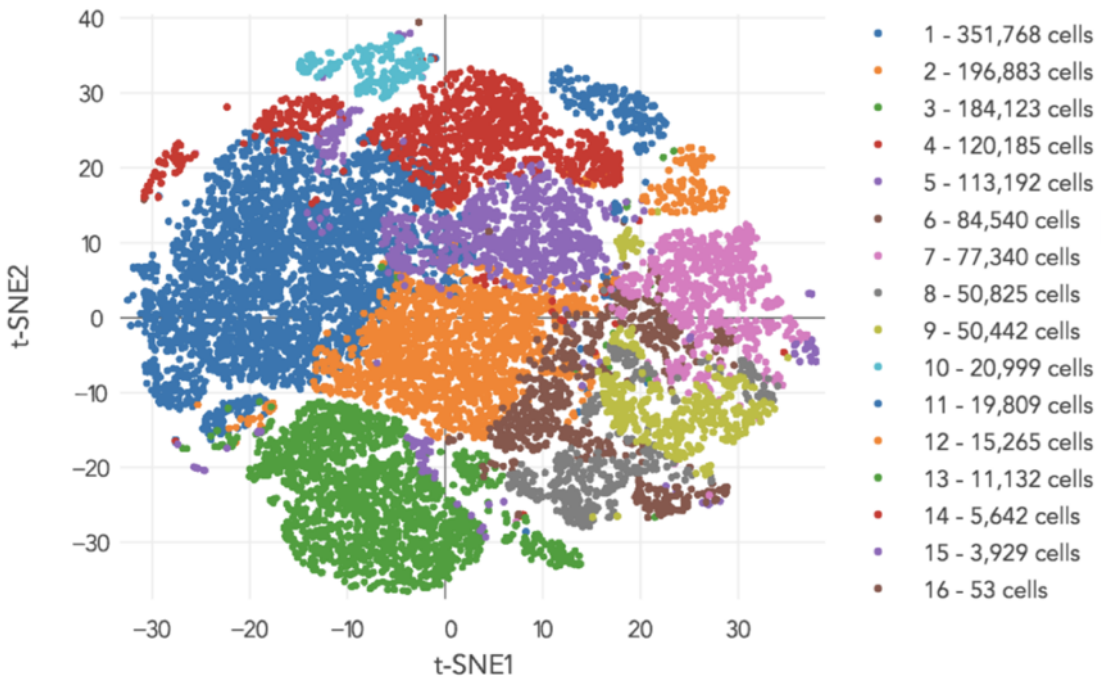
Charles P. Couturier, Shamini Ayyadhury, Phuong U. Le, Javad Nadaf, Jean Monlong, Gabriele Riva, Redouane Allache, Salma Baig, Xiaohua Yan, Mathieu Bourgey, Changseok Lee, Yu Chang David Wang, V. Wee Yong, Marie-Christine Guiot, Hamed Najafabadi, Bratislav Misic, Jack Antel, Guillaume Bourque, Jiannis Ragoussis & Kevin Petrecca

Nature Communications **11**, Article number: 3406 (2020) | Cite this article

18k Accesses | **6** Citations | **165** Altmetric | Metrics

How do we perform scRNA-seq?

Datasets continue to increase in size (more cells detected and better sensitivity)…
and a parallel increase in tool development to make sense of these data



1.3 million murine brain cells

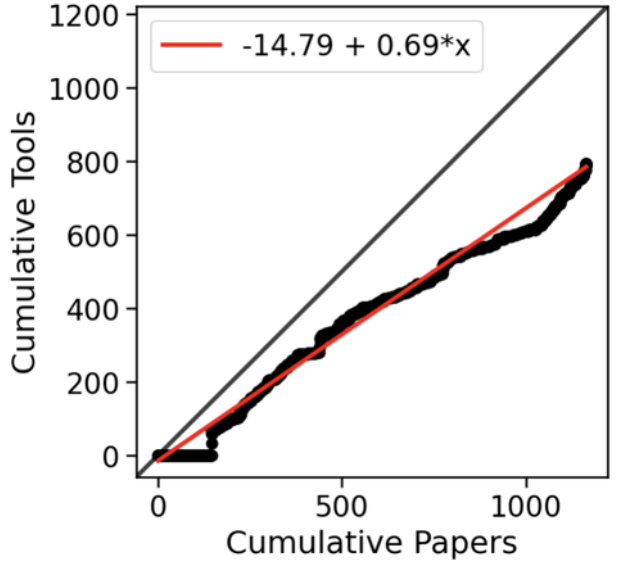
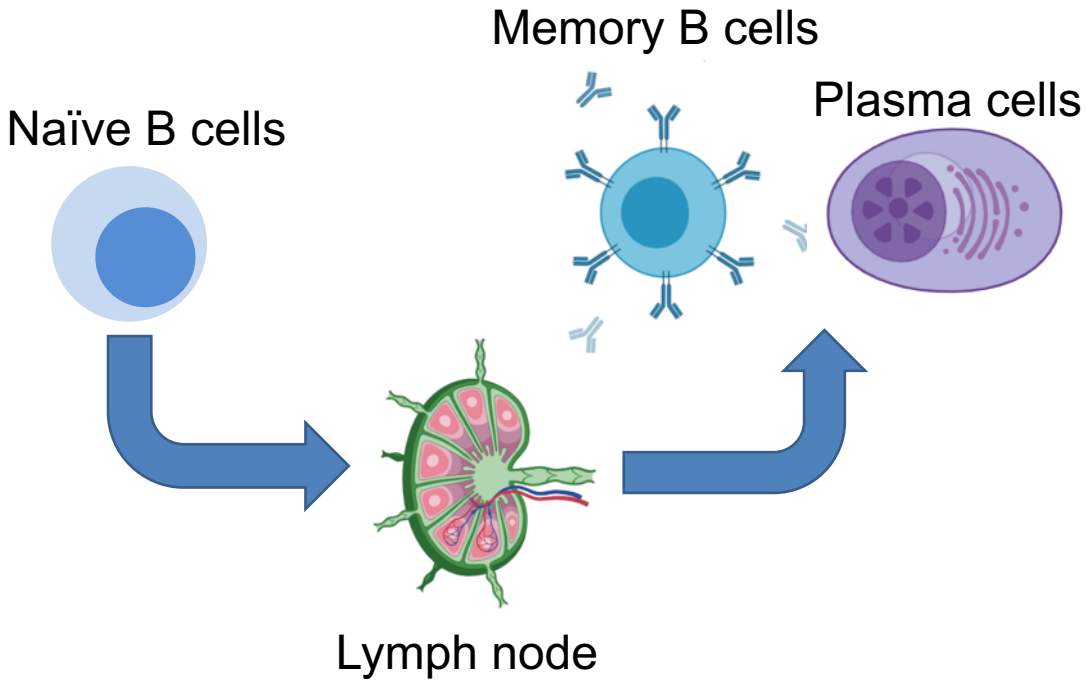


Figure credit: Valentine Svennson and Lior Pachter

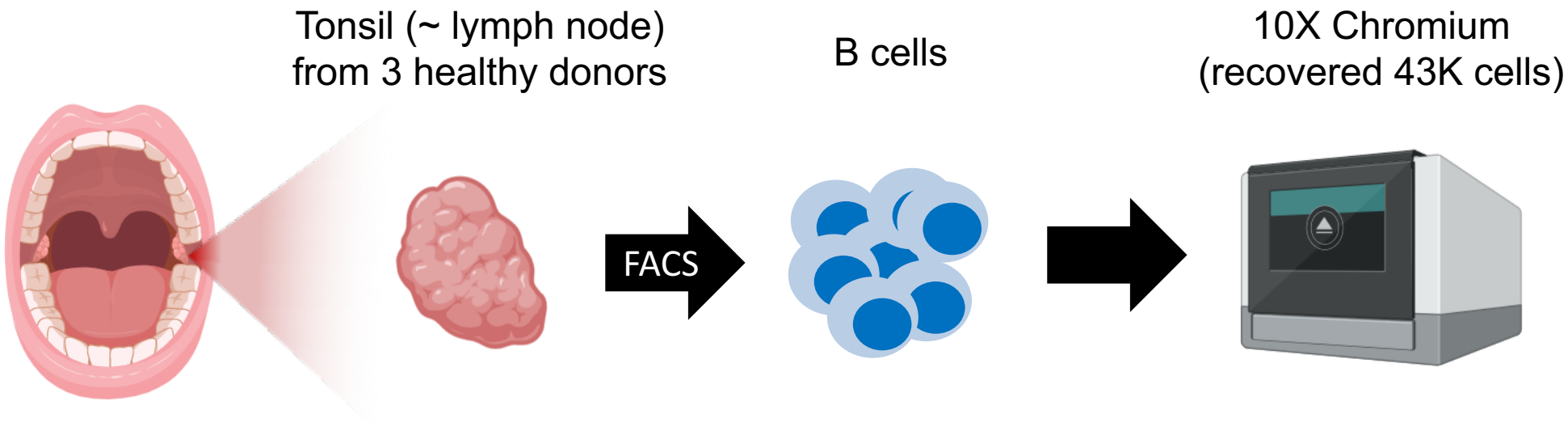
scRNA-seq of healthy tonsillar B cells

Motivation

Tonsillar B cells: an atlas-type approach



Tonsillar B cells: an atlas-type approach



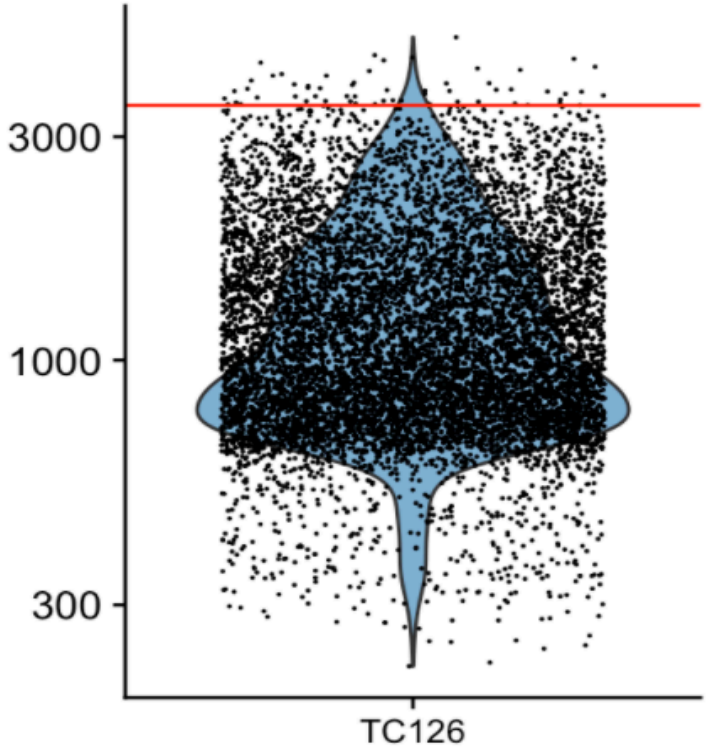
scRNA-seq of healthy tonsillar B cells

Overview of common analysis practices

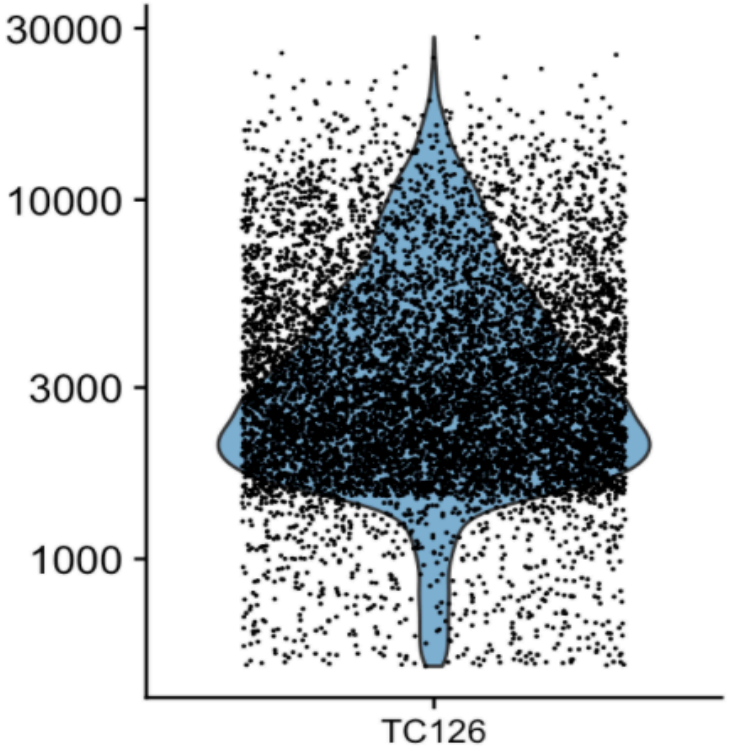
Overview of data processing

An example of one tonsil's dataset (8,828 cells)

Number of unique mRNA species detected per cell (median = 1,066)



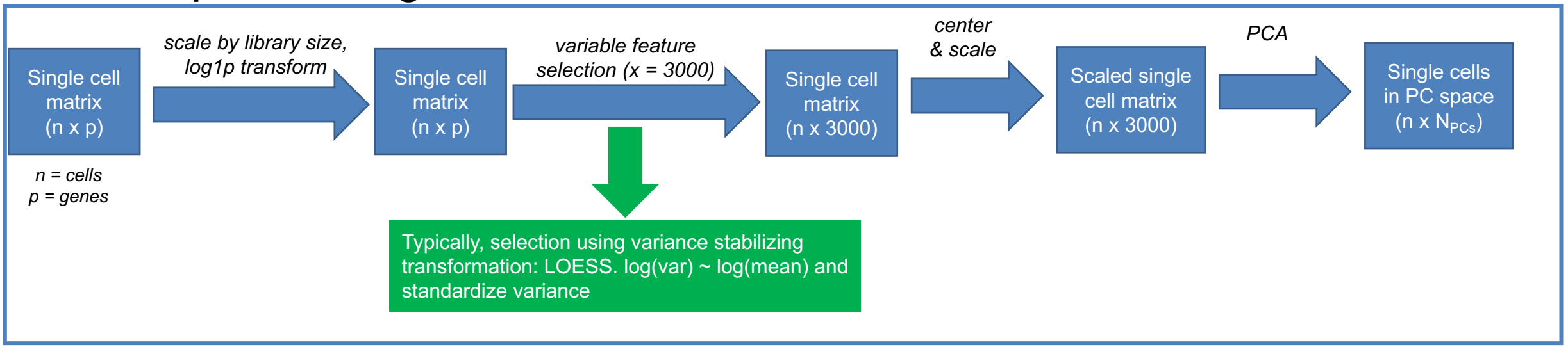
Number of RNA molecules detected per cell (median = 2,860)



Result is a relatively sparse matrix

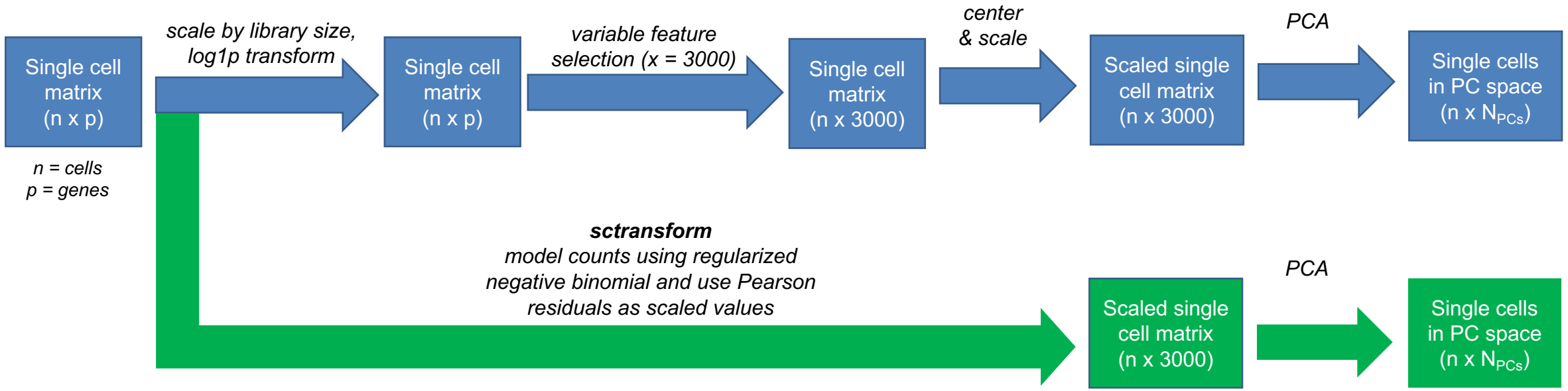
	Cell 1	Cell 2	Cell 3	Cell 4	…	Cell N
Gene 1	10	0	10	0		
Gene 2	0	0	39	0		
Gene 3	87	11	4	0		
Gene 4	0	0	0	41		
…						
Gene 16,000						

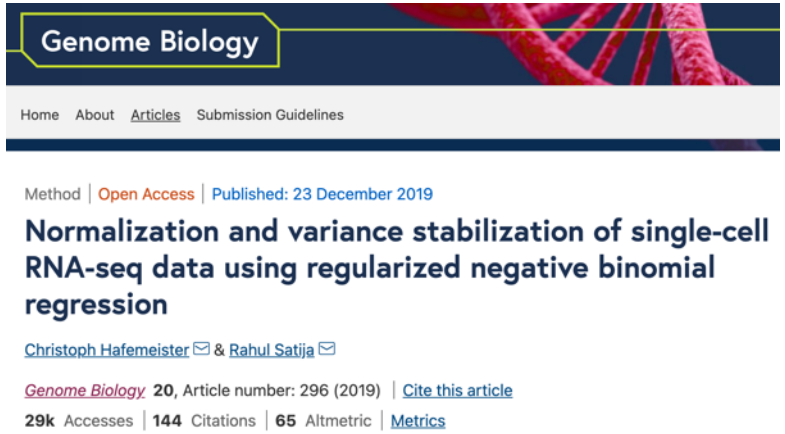
Data processing overview



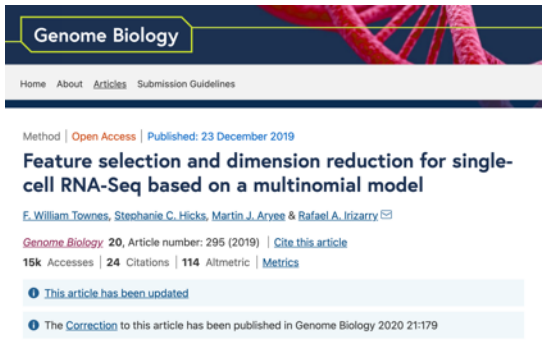
Most common workflow

Data processing overview

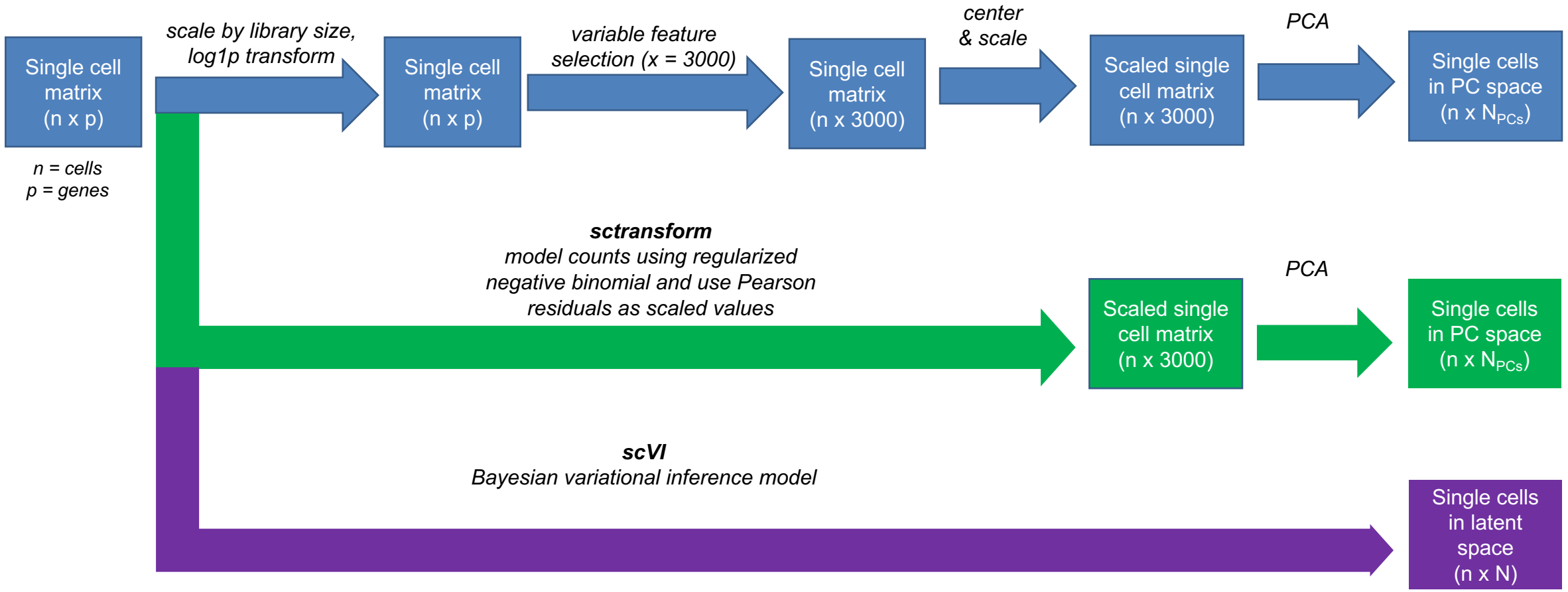




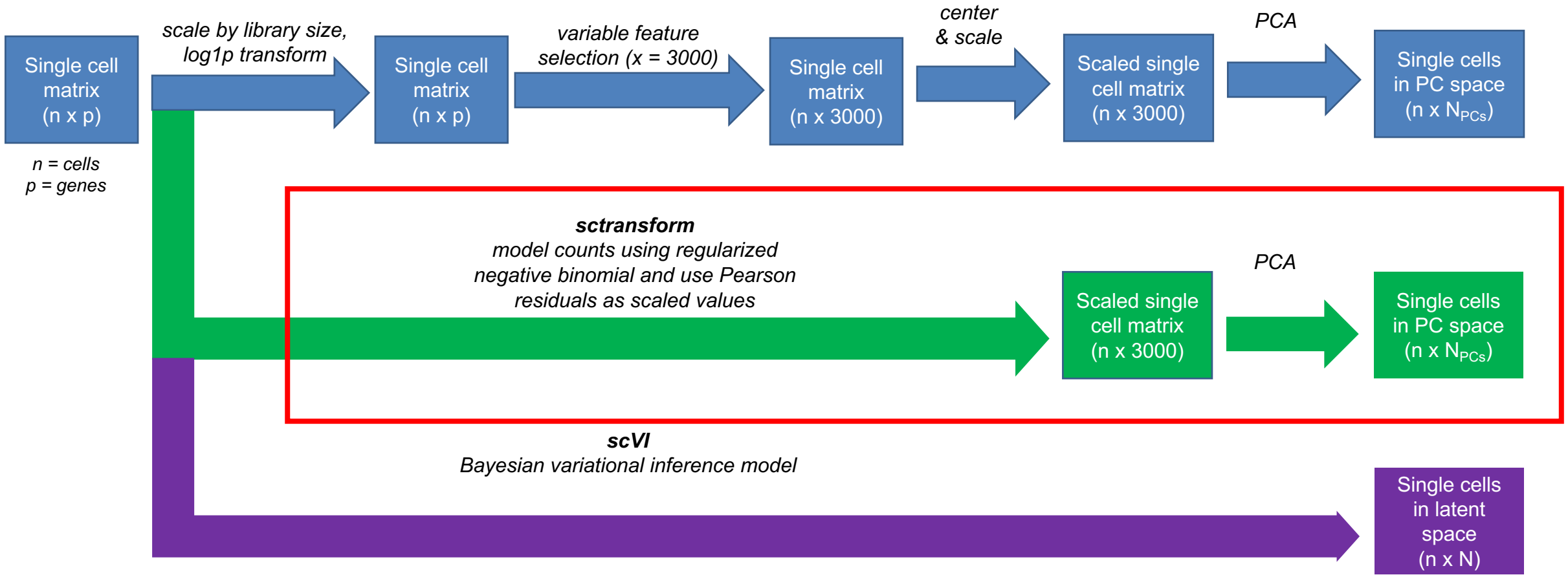
see also



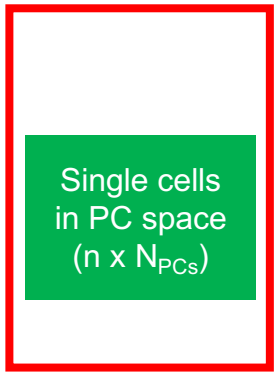
Data processing overview



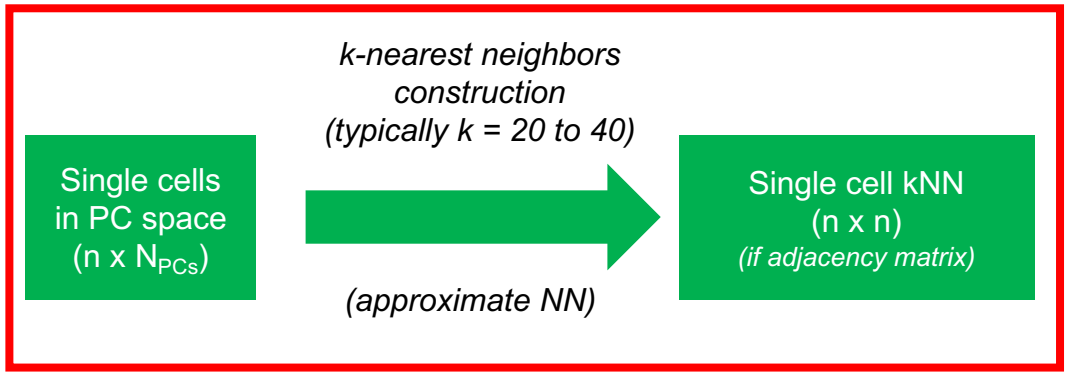
Data processing overview



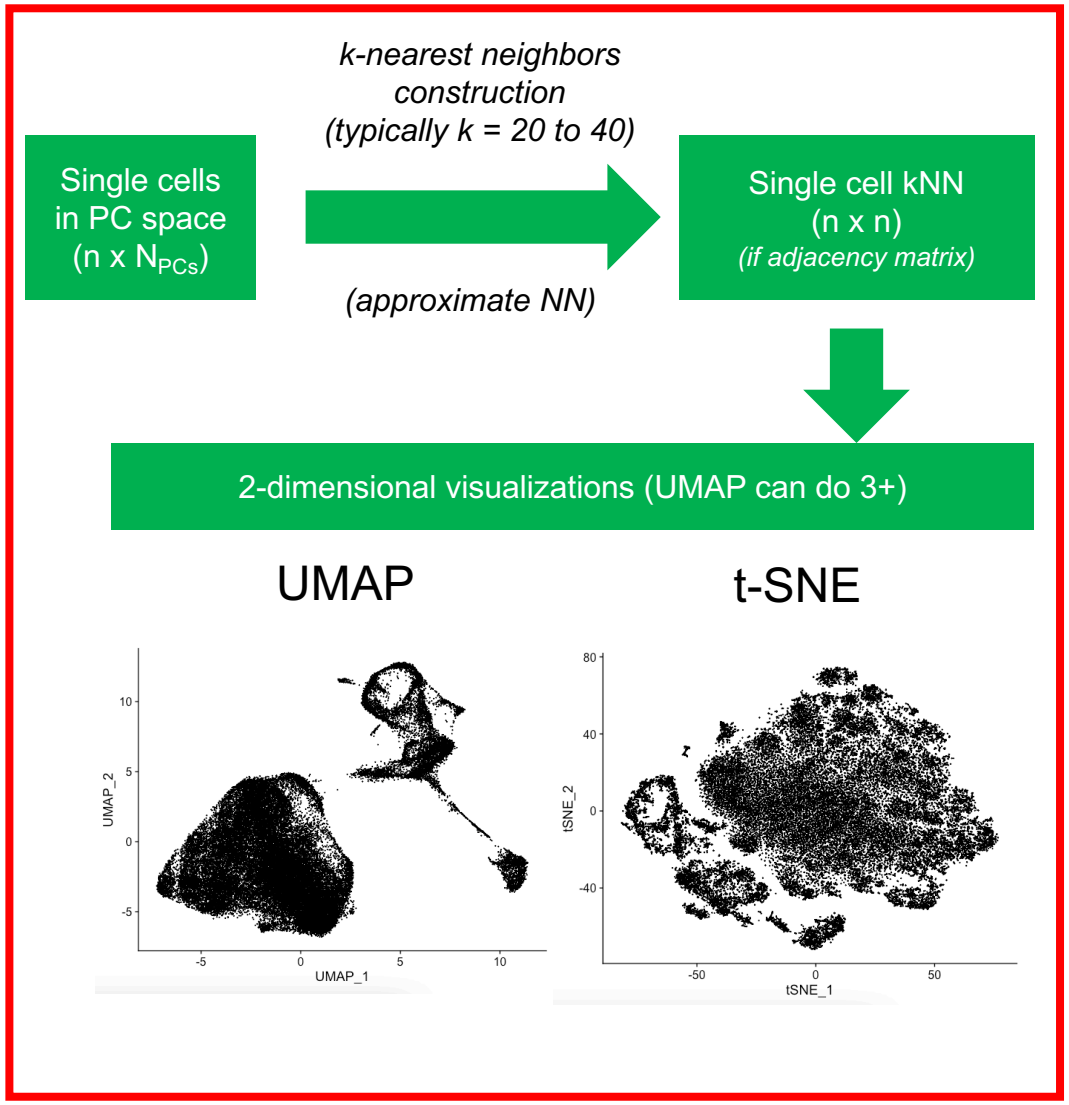
Data processing overview



Data processing overview



Data processing overview



Data processing overview

k-nearest neighbors construction (typically k = 20 to 40)

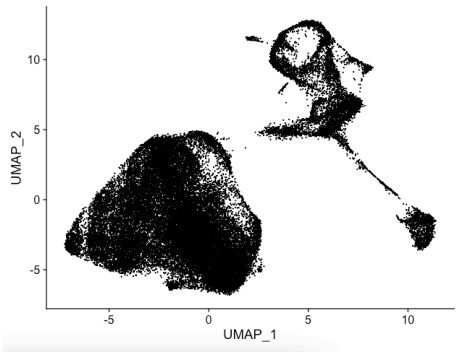
Single cells in PC space (n x N_{PCs})

(approximate NN)

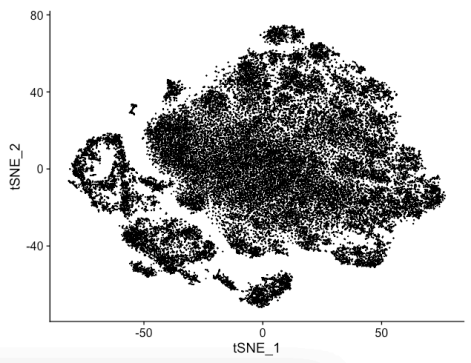
Single cell kNN (n x n) *(if adjacency matrix)*

2-dimensional visualizations (UMAP can do 3+)

UMAP



t-SNE



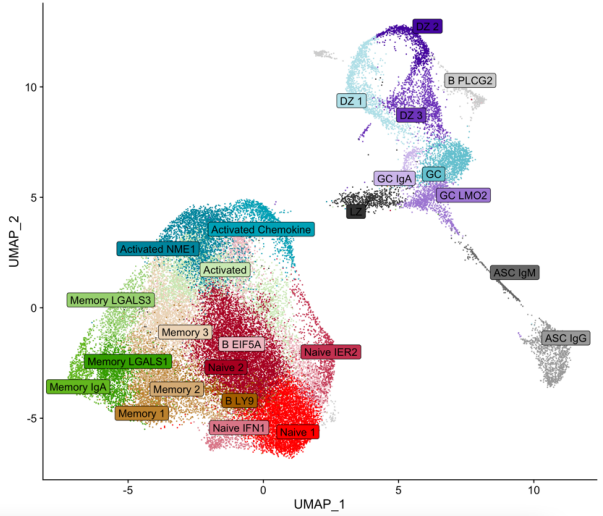
Graph-based clustering methods using an SNN (Louvain, Leiden)

Louvain and Leiden modularity clustering are almost exclusively used in the single-cell literature

Identify genes enriched in clusters

This step requires knowledge of the literature of your system and manual curation. Supervised classifiers are growing in use, however.

Annotate clusters with determined identities



Data processing overview

Single cells in PC space (n x N_{PCs})

k-nearest neighbors construction (typically k = 20 to 40)

(approximate NN)

Single cell kNN (n x n) *(if adjacency matrix)*

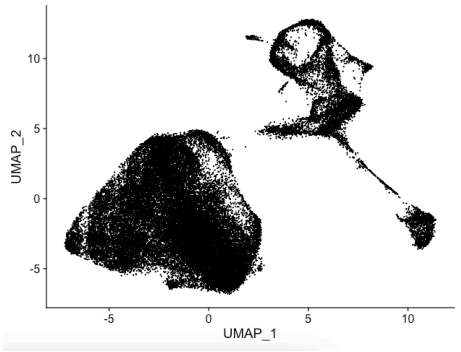
Graph-based clustering methods using an SNN (Louvain, Leiden)

Identify genes enriched in clusters

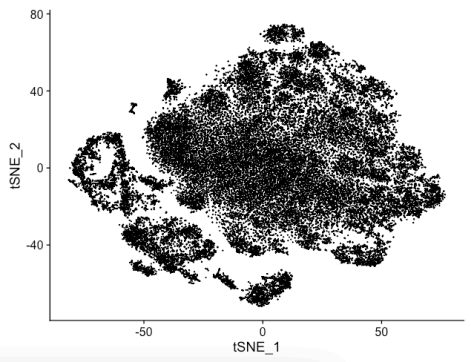
Annotate clusters with determined identities

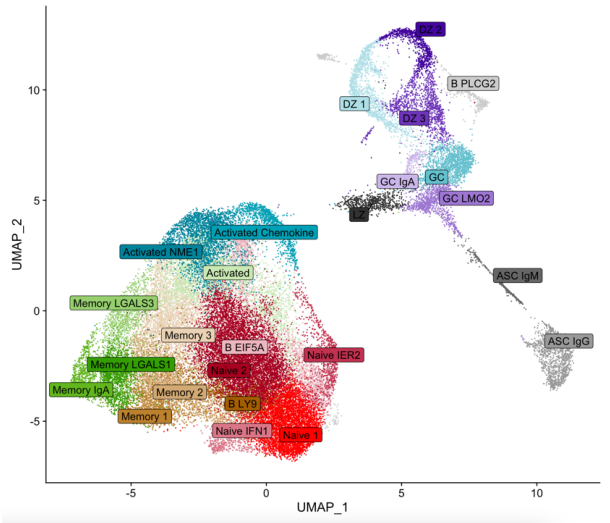
2-dimensional visualizations (UMAP can do 3+)

UMAP



t-SNE





- A. Publishable atlas-level annotation resource
- B. Compositional or differential expression analyses for condition-based experiments
- C. Network analyses: cell-cell and gene-gene interactions network analysis
- D. Trajectory inference and pseudotemporal analyses

Data processing overview

Single cells in PC space (n x N_{PCs})

k-nearest neighbors construction (typically k = 20 to 40)

(approximate NN)

Single cell kNN (n x n) *(if adjacency matrix)*

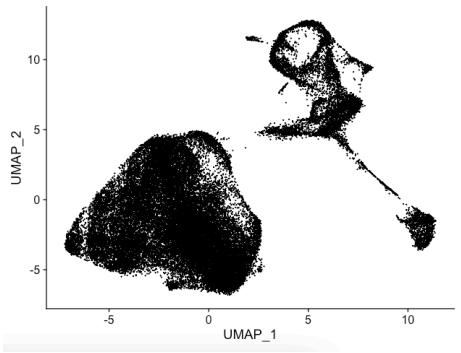
Graph-based clustering methods using an SNN (Louvain, Leiden)

Identify genes enriched in clusters

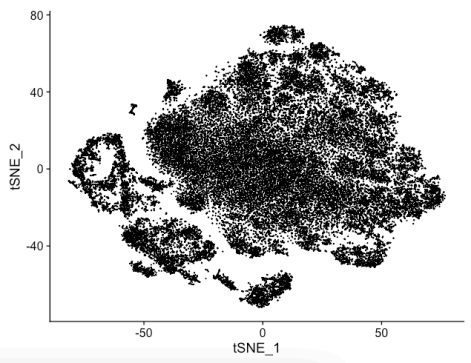
Annotate clusters with determined identities

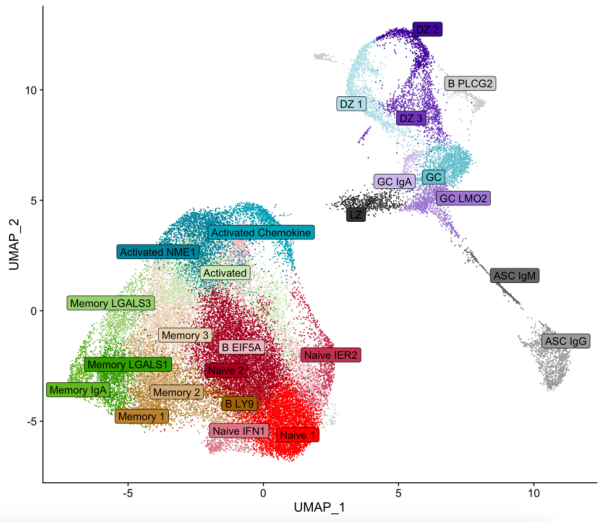
2-dimensional visualizations (UMAP can do 3+)

UMAP



t-SNE





A. Publishable atlas-level annotation resource

B. Compositional or differential expression analyses for condition-based experiments

C. Network analyses: cell-cell and gene-gene interactions network analysis

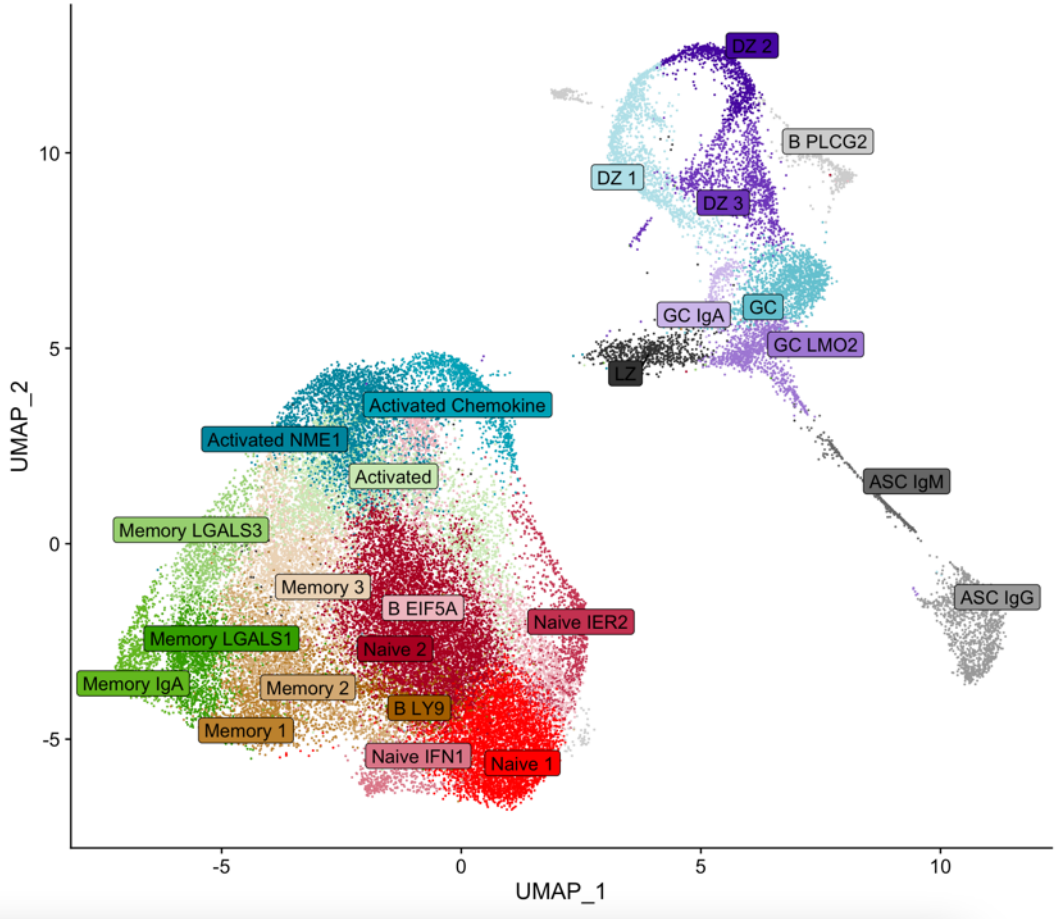
D. Trajectory inference and pseudotemporal analyses

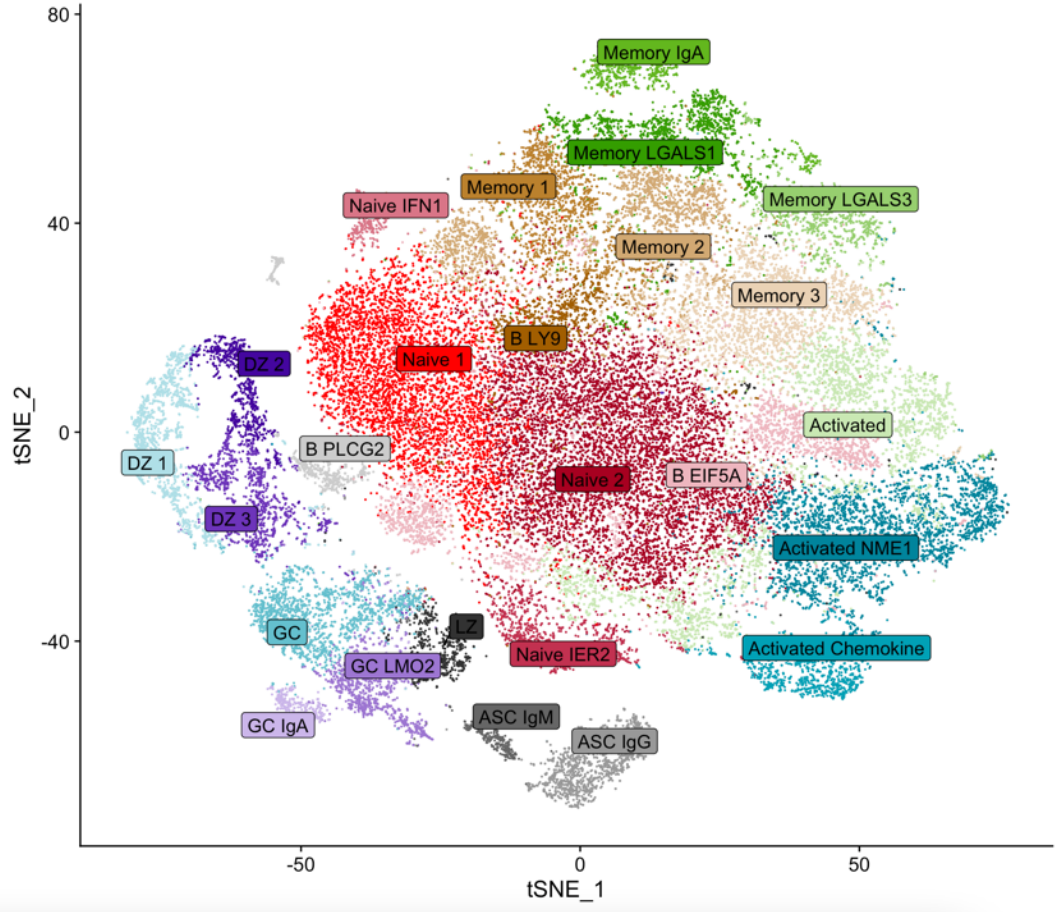
scRNA-seq of healthy tonsillar B cells

UMAP, t-SNE, and trajectory inference

Clusters and dimensional reduction in tonsillar B cell dataset

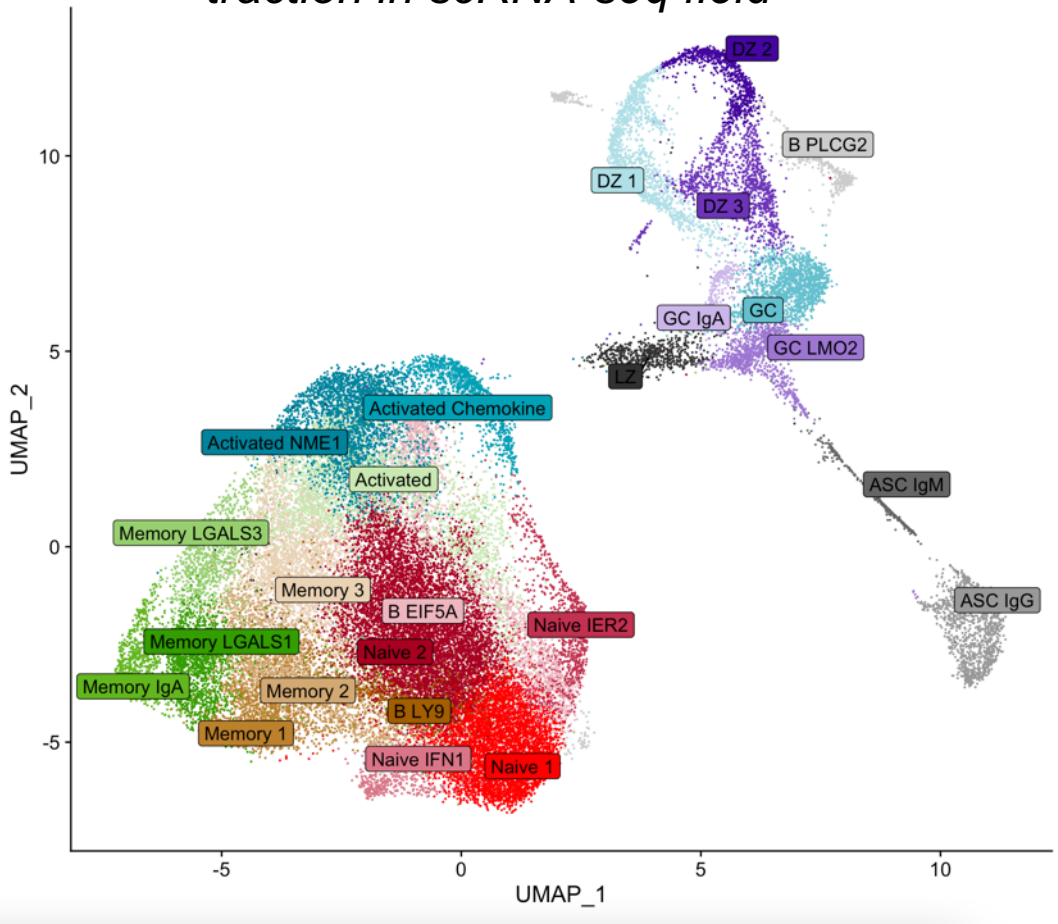
UMAP and t-SNE are traditionally used as visualization tools …





Clusters and dimensional reduction in tonsillar B cell dataset

UMAP in particular has gained most traction in scRNA-seq field

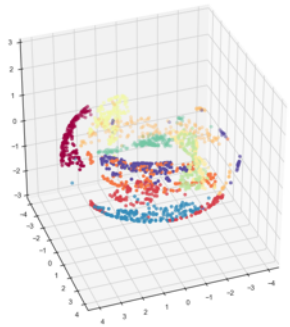


UMAP

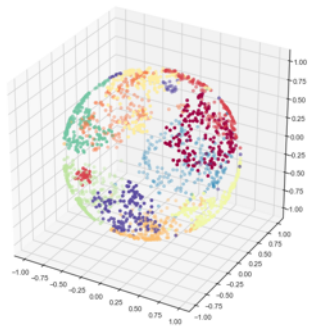
https://umap-learn.readthedocs.io/en/latest/index.html

USER GUIDE / TUTORIAL:

- How to Use UMAP
- Basic UMAP Parameters
- Plotting UMAP results
- UMAP Reproducibility
- Transforming New Data with UMAP
- Inverse transforms
- Parametric Embedding
- UMAP on sparse data
- UMAP for Supervised Dimension Reduction and Metric Learning
- Using UMAP for Clustering
- Outlier detection using UMAP
- Combining multiple UMAP models
- Better Preserving Local Density with DensMAP
- Document embedding using UMAP
- Embedding to non-Euclidean spaces
- How to use AlignedUMAP
- AlignedUMAP for Time Varying Data
- Release Notes
- Frequently Asked Questions

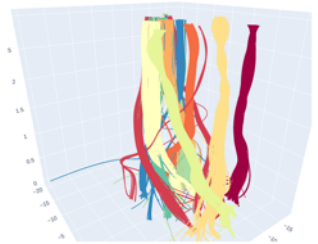


Torus



Sphere

Align UMAP over time

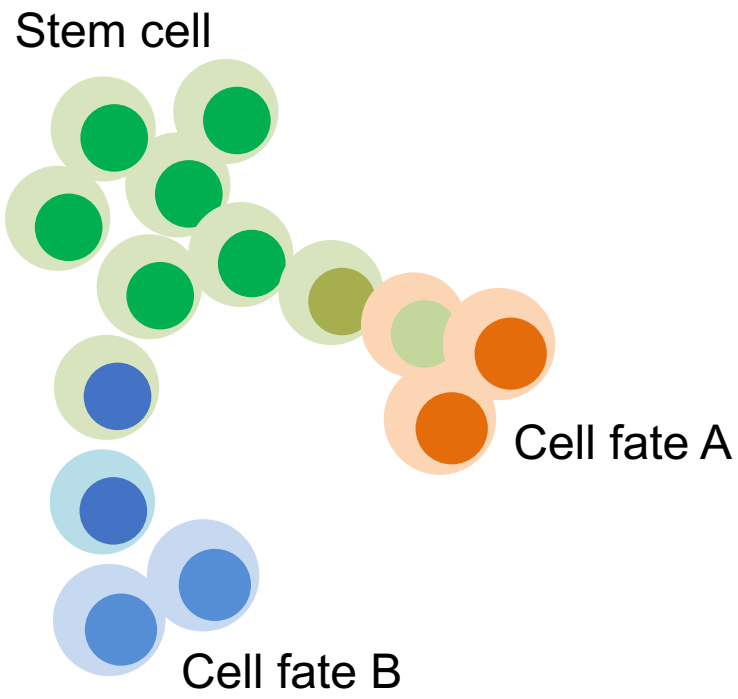


Clusters and dimensional reduction in tonsillar B cell dataset

These reductions are largely for visual use…
with the exception of their utility in trajectory inference methods

Key assumption: chosen dimensionality reduction (UMAP, t-SNE, whatever you use) captures continuum of cell states across paths of maturation, transition, or development

Clusters and dimensional reduction in tonsillar B cell dataset

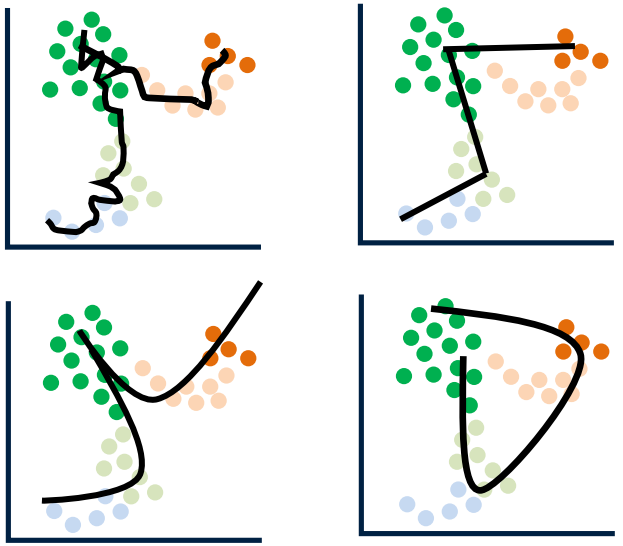


PCA, UMAP, t-SNE, ICA, DDRTree,
PHATE, Diffusion maps, etc…

Clusters and dimensional reduction in tonsillar B cell dataset

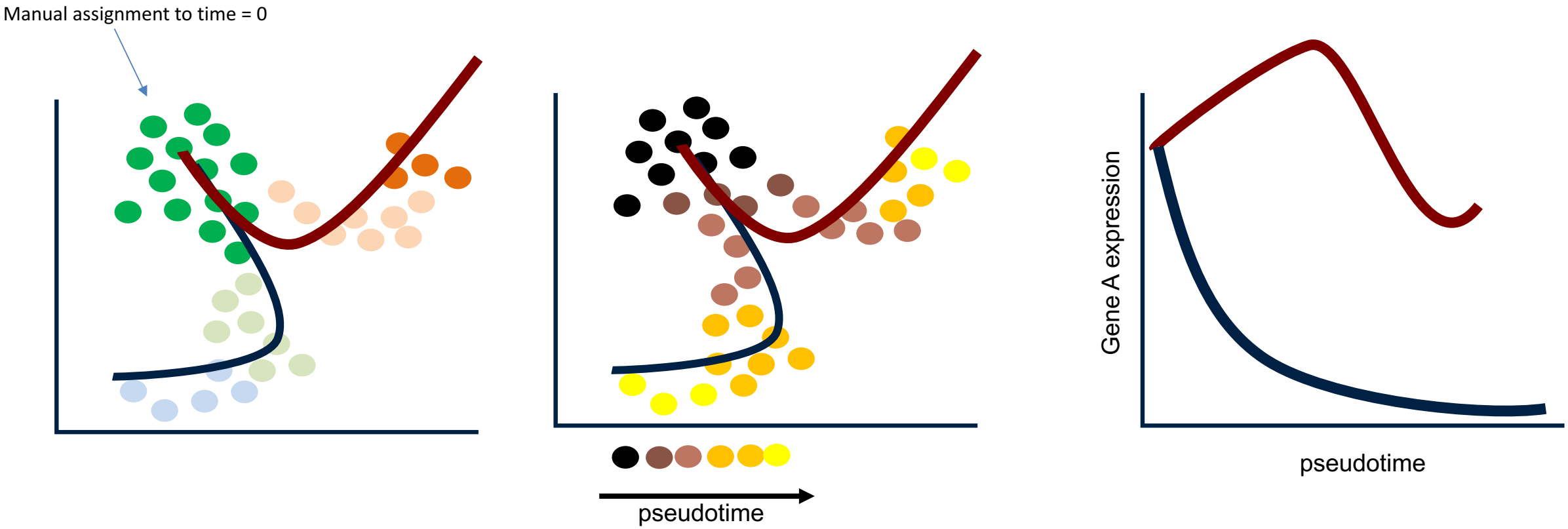
PCA, UMAP, t-SNE, ICA, DDRTree,
PHATE, Diffusion maps, etc…

As of 2019, 70+ TI methods developed,
with differing performance…



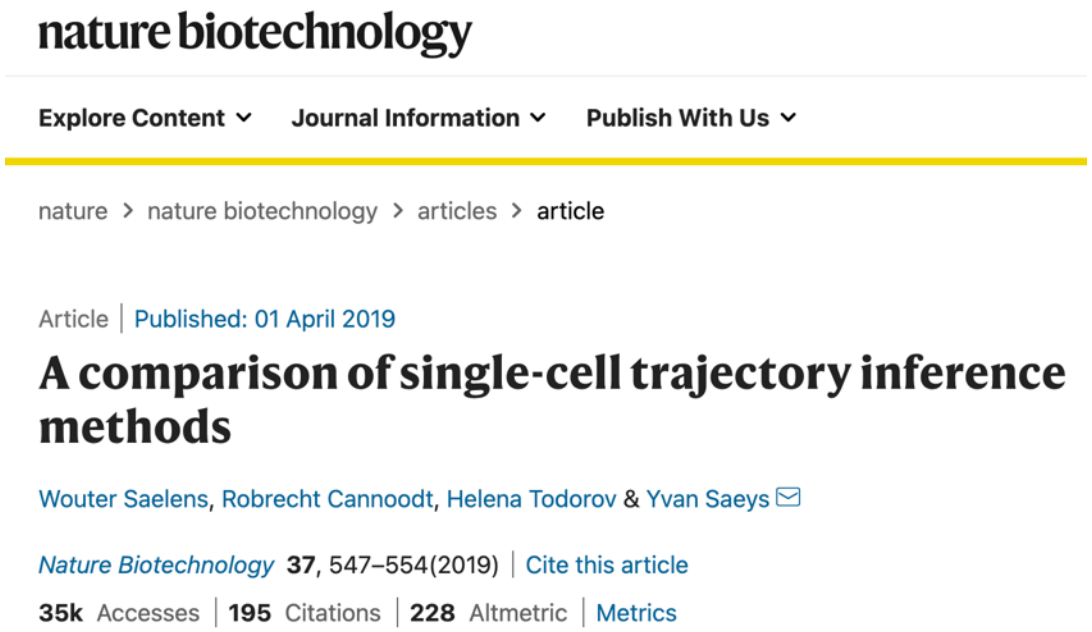
Clusters and dimensional reduction in tonsillar B cell dataset

Use the trajectory as a "pseudo-time" axis and project cells to this time axis:
each cell will now have an associated "pseudotime" (and trajectory membership)



Clusters and dimensional reduction in tonsillar B cell dataset

Unlike previous analyses, very little consensus within the field for choosing a single method



nature biotechnology

Explore Content ⌄ Journal Information ⌄ Publish With Us ⌄

nature › nature biotechnology › articles › article

Article | Published: 01 April 2019

A comparison of single-cell trajectory inference methods

Wouter Saelens, Robrecht Cannoodt, Helena Todorov & Yvan Saeys ✉

Nature Biotechnology **37**, 547–554(2019) | Cite this article

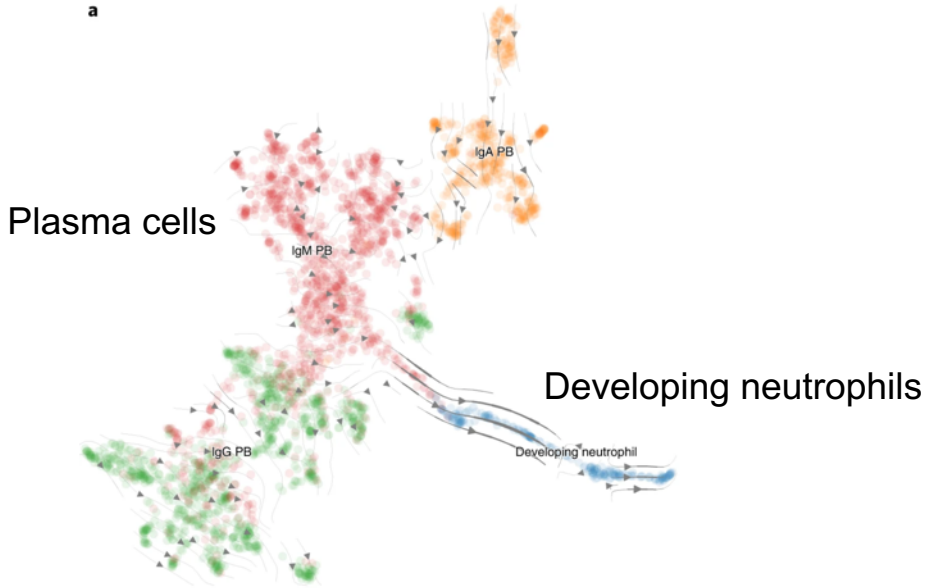
35k Accesses | **195** Citations | **228** Altmetric | Metrics

Clusters and dimensional reduction in tonsillar B cell dataset

Clusters and dimensional reduction in tonsillar B cell dataset

IDENTIFIED TRAJECTORIES DO NOT ALWAYS IMPLY CELL LINEAGE RELATIONSHIPS!

Wilk et al, 2020, Nature Medicine (Figure 4a)



No evidence that plasmablasts transdifferentiate into developing neutrophils in severe COVID-19 disease

José Alquicira-Hernandez[1], Joseph E Powell[1,2]*, Tri Giang Phan[1,3]*

[1] Garvan Institute of Medical Research, 384 Victoria St, Darlinghurst NSW 2010, Sydney, Australia
[2] UNSW Cellular Genomics Futures Institute, University of New South Wales, Sydney, Australia
[3] St Vincent's Clinical School, Faculty of Medicine, UNSW Sydney, 384 Victoria St, Darlinghurst NSW 2010, Sydney, Australia
*equal contribution
Email: j.powell@garvan.org.au or t.phan@garvan.org.au

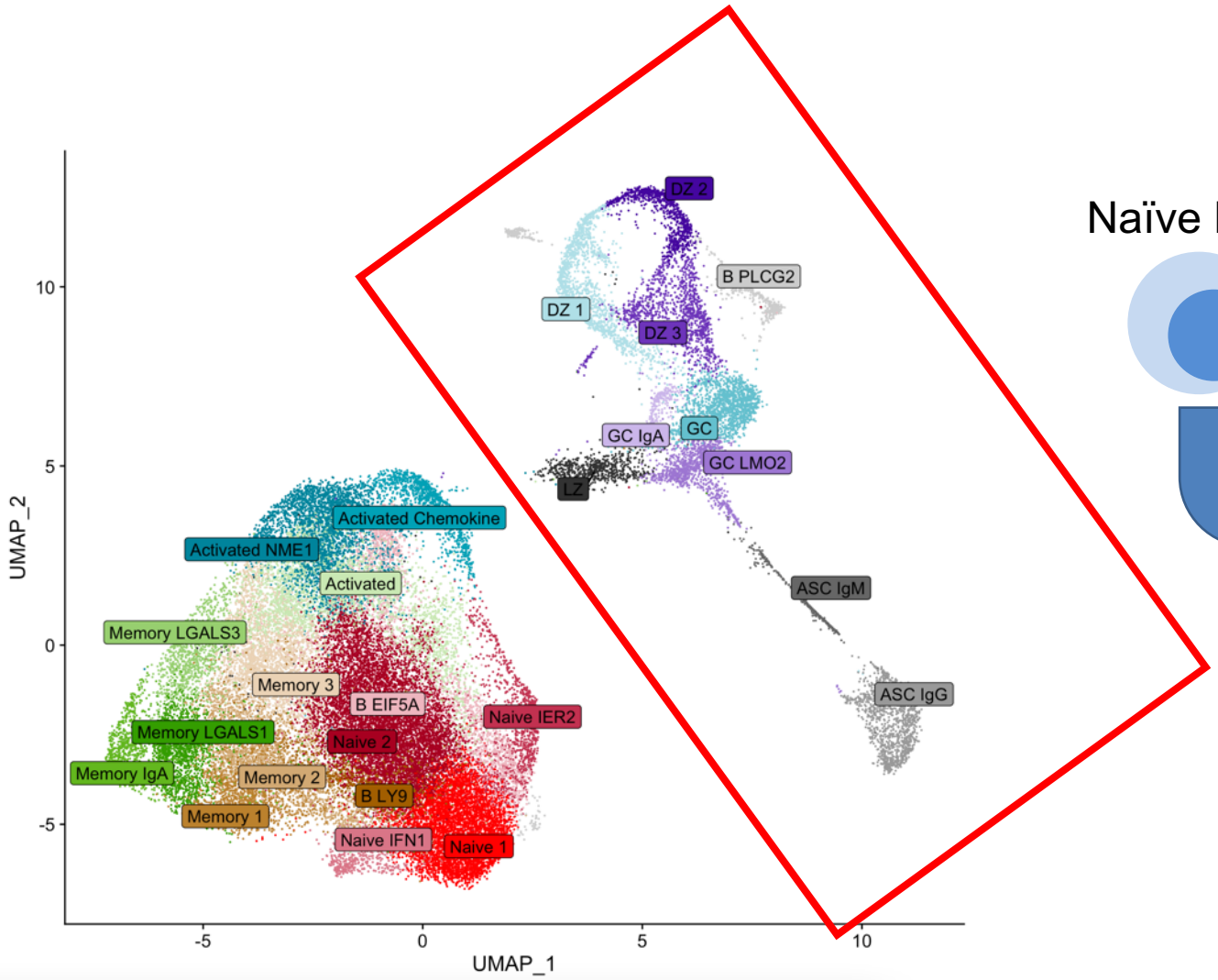
"UMAP embeddings may reflect the expression of similar genes but not necessarily direct cell lineage relationships"

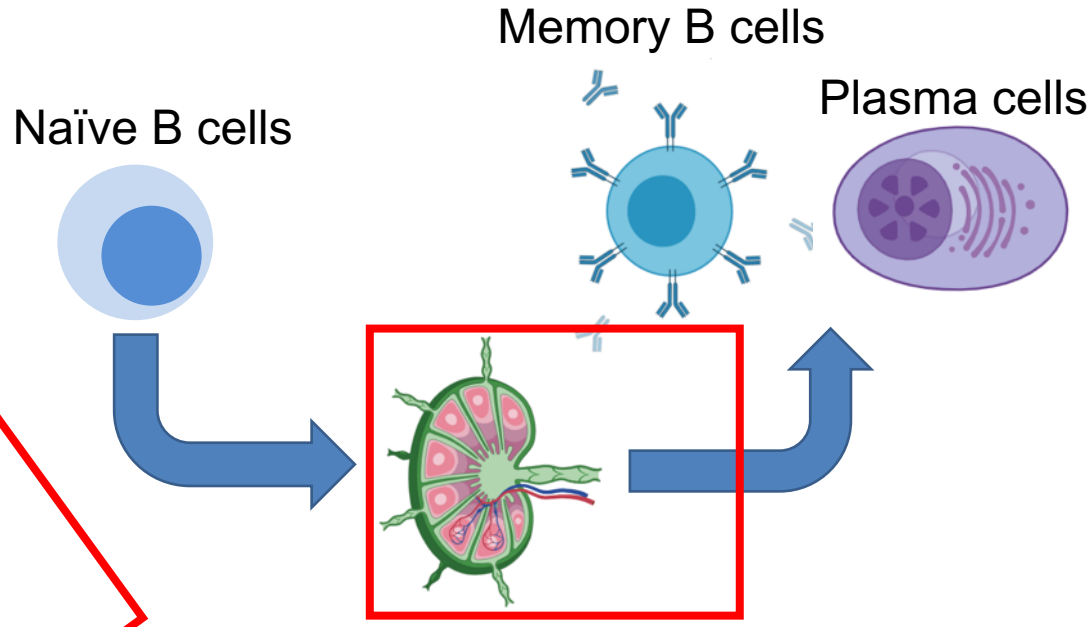
(this should be obvious but there exist plenty of examples of this in the literature 🤦)

scRNA-seq of healthy tonsillar B cells

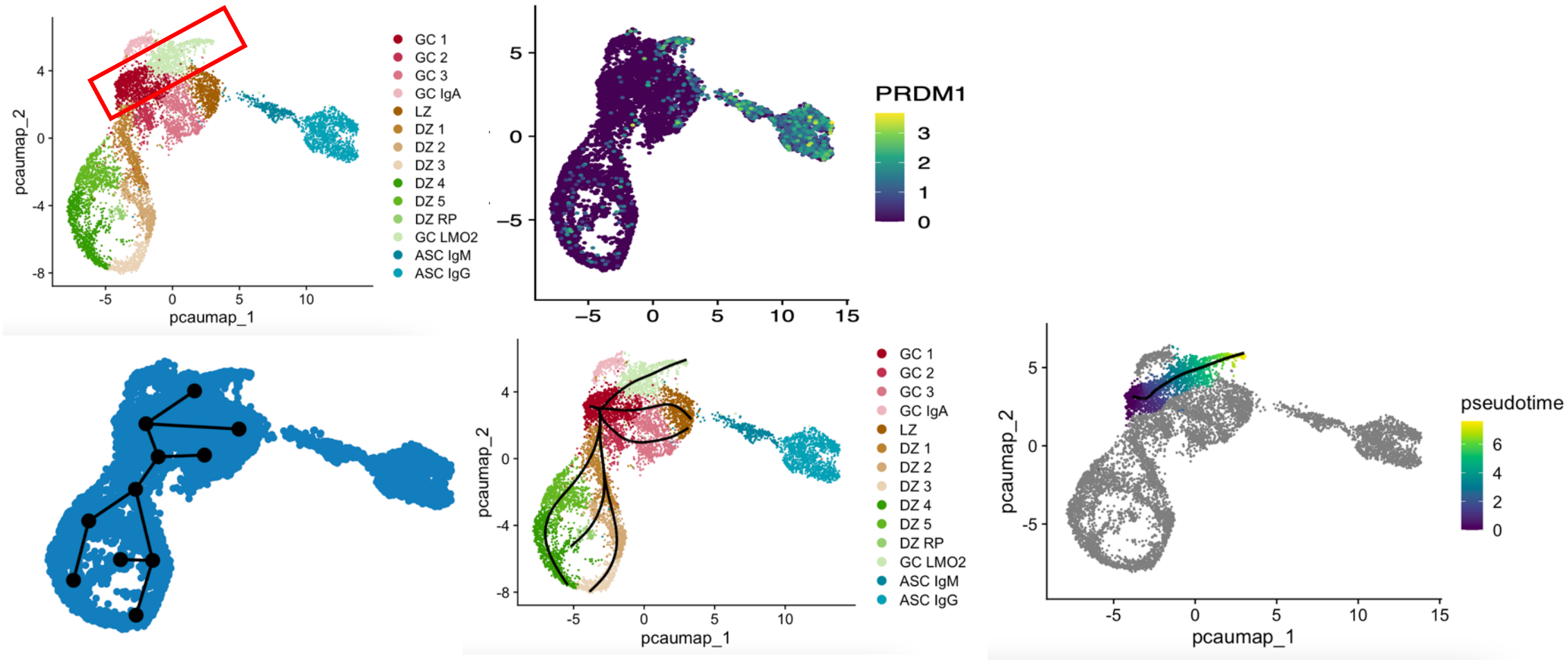
Trajectory inference in B cell maturation

Clusters and dimensional reduction in tonsillar B cell dataset

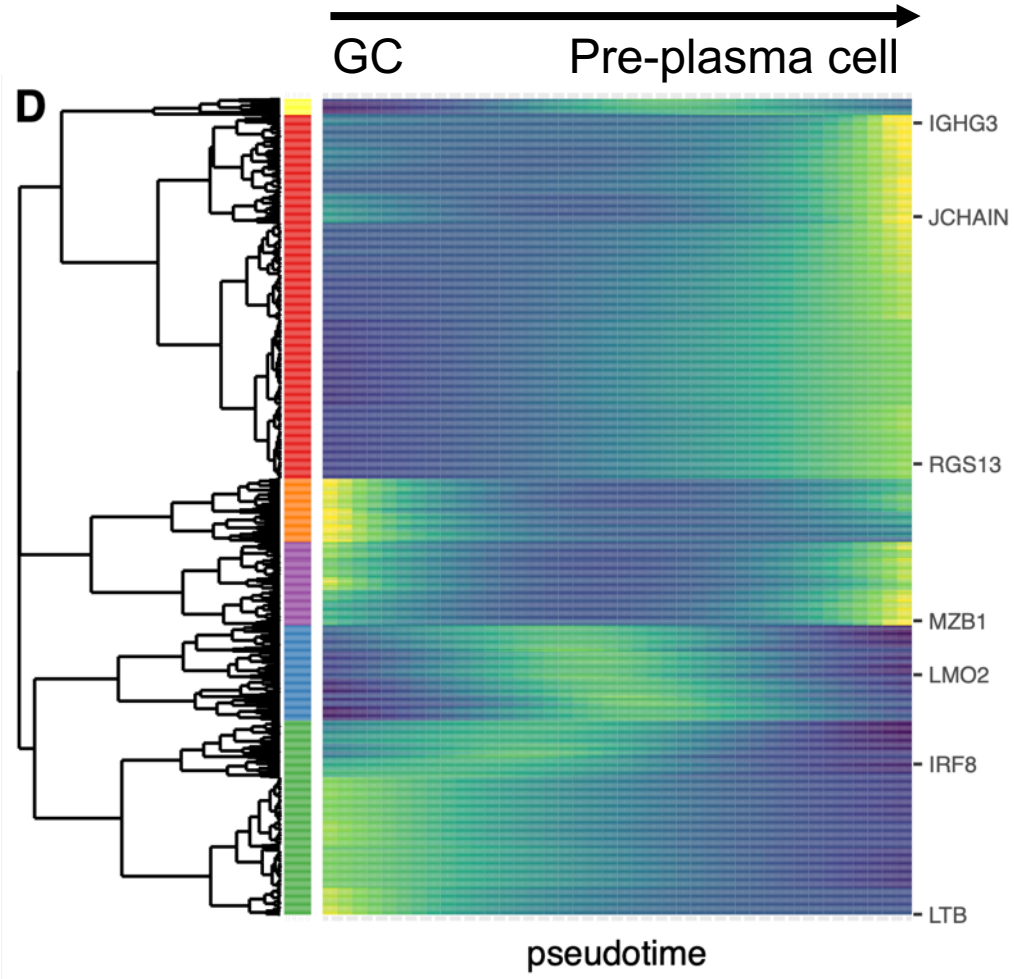


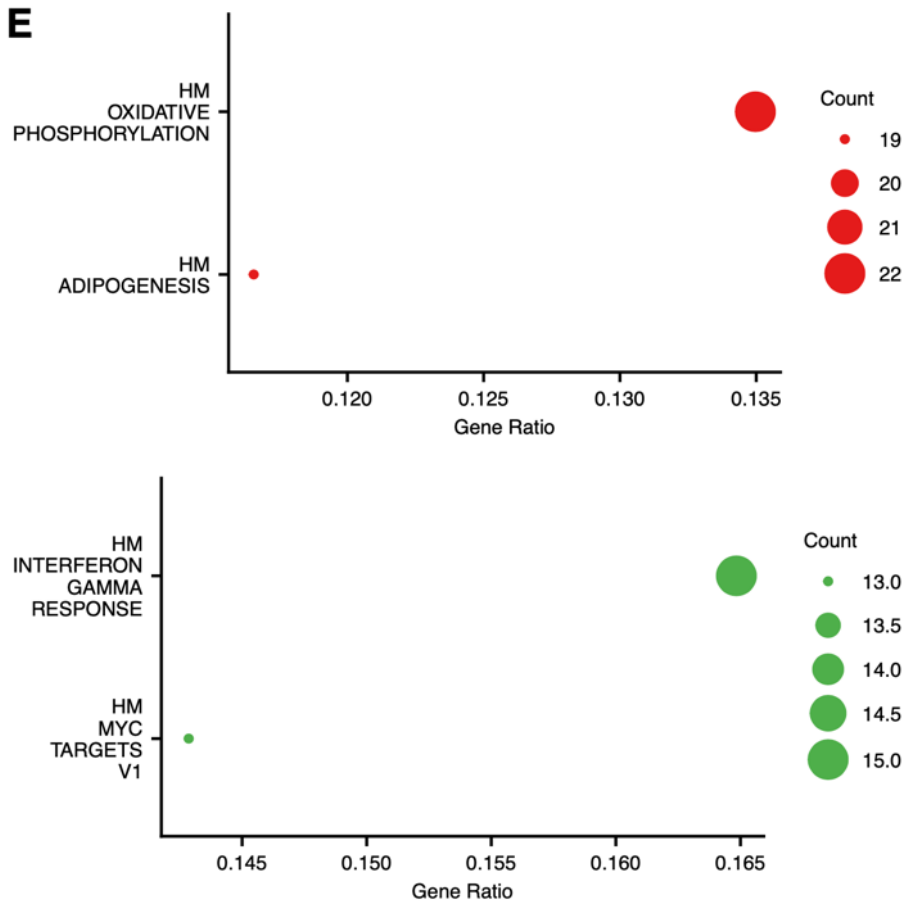


Clusters and dimensional reduction in tonsillar B cell dataset



Clusters and dimensional reduction in tonsillar B cell dataset

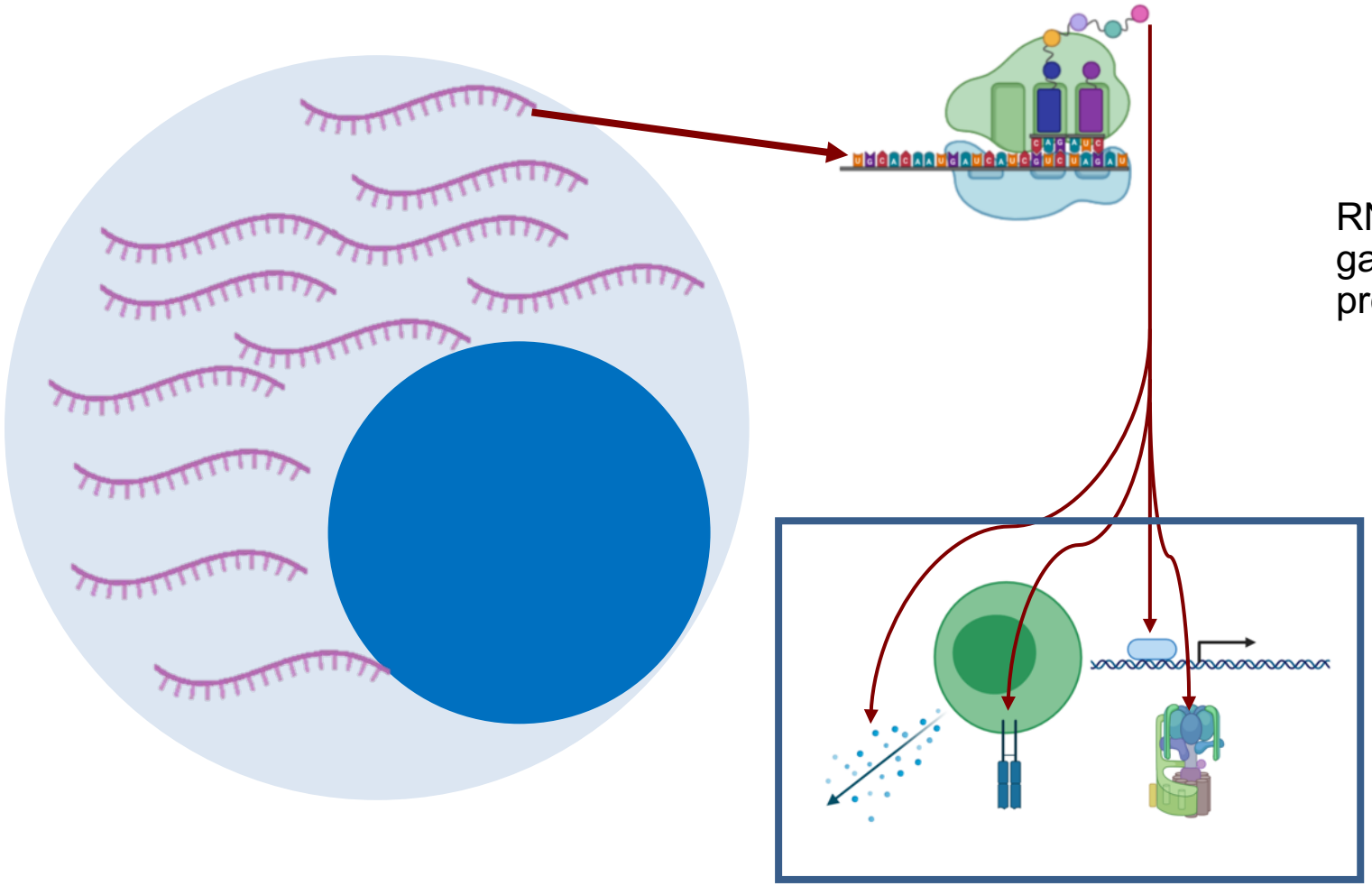




CITE-seq of peripheral blood cells in MS before/after treatment

Multimodal data

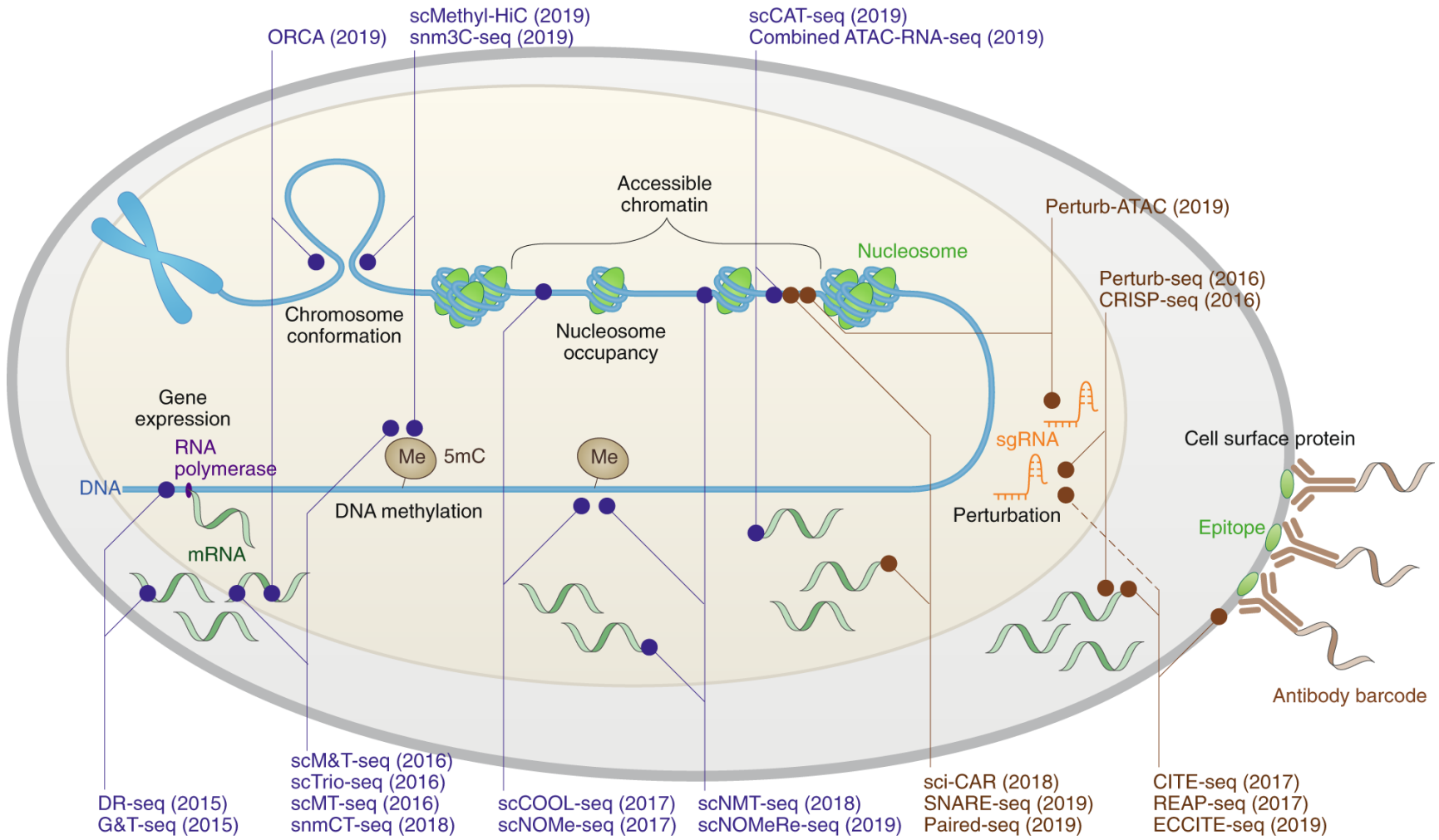
Why do we use scRNA-seq?



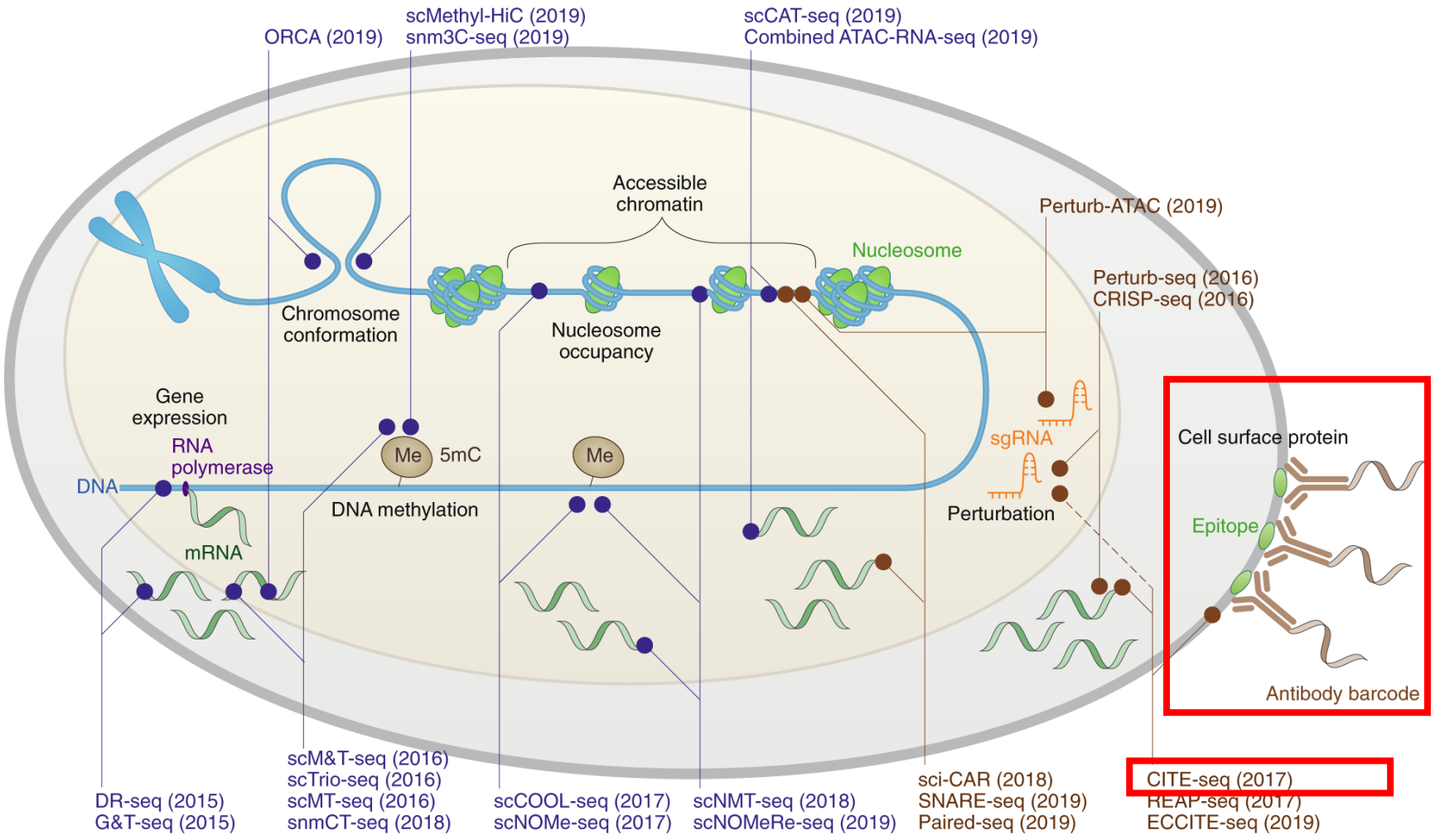
RNA ultimately provides one approach at gaining insight into cellular function and processes

But we know that cell function is defined by more than just RNA…

Multimodal data analyses (slide already dated…)

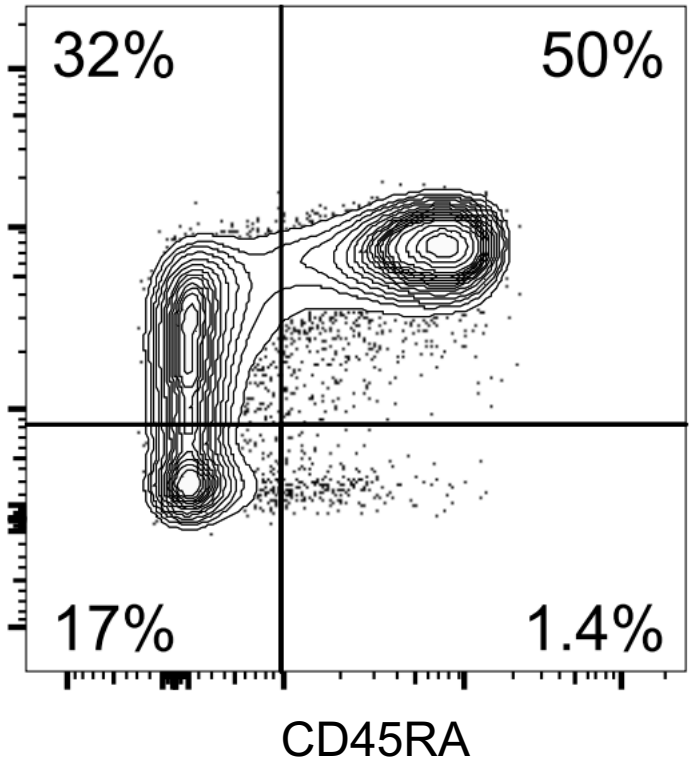


CITE-seq perhaps most commonly used (and commercially available) method

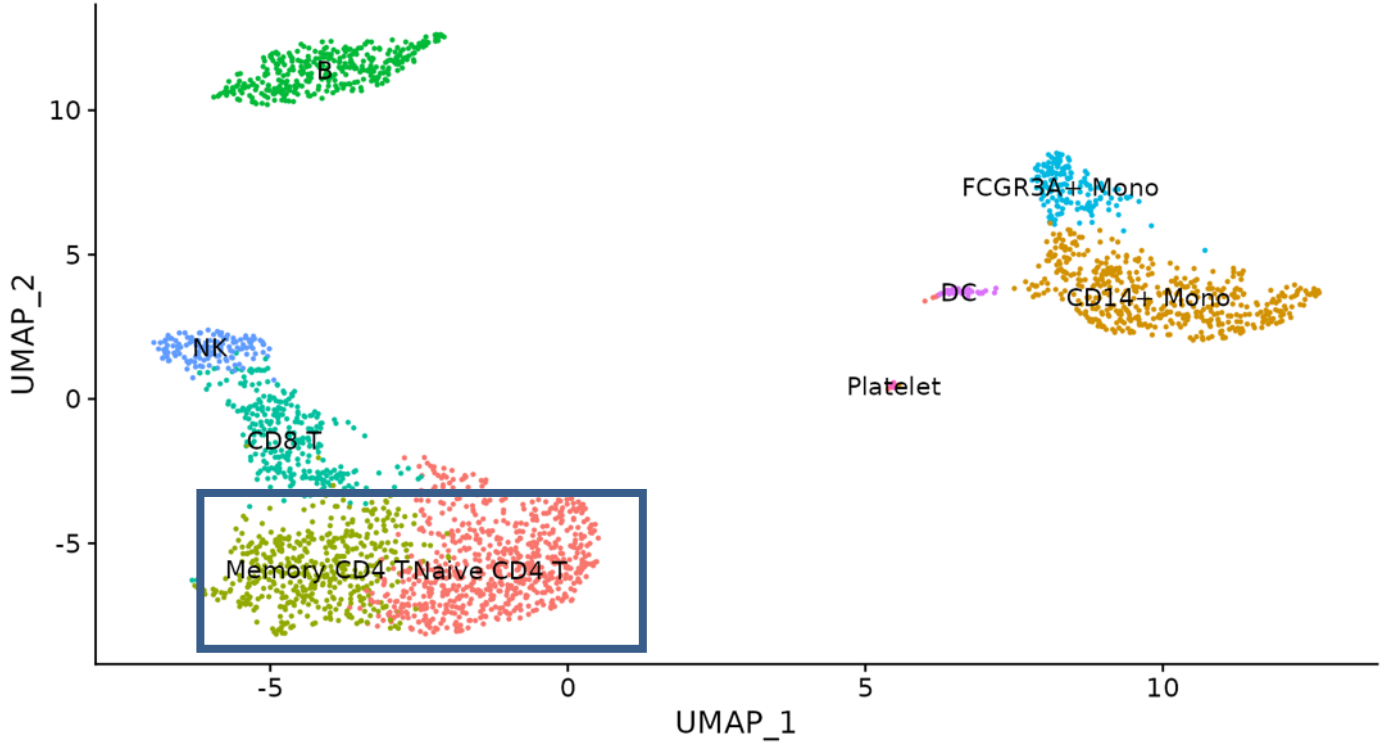


What are the advantages?

These 4 CD4 T cell populations functionally different based on decades of study

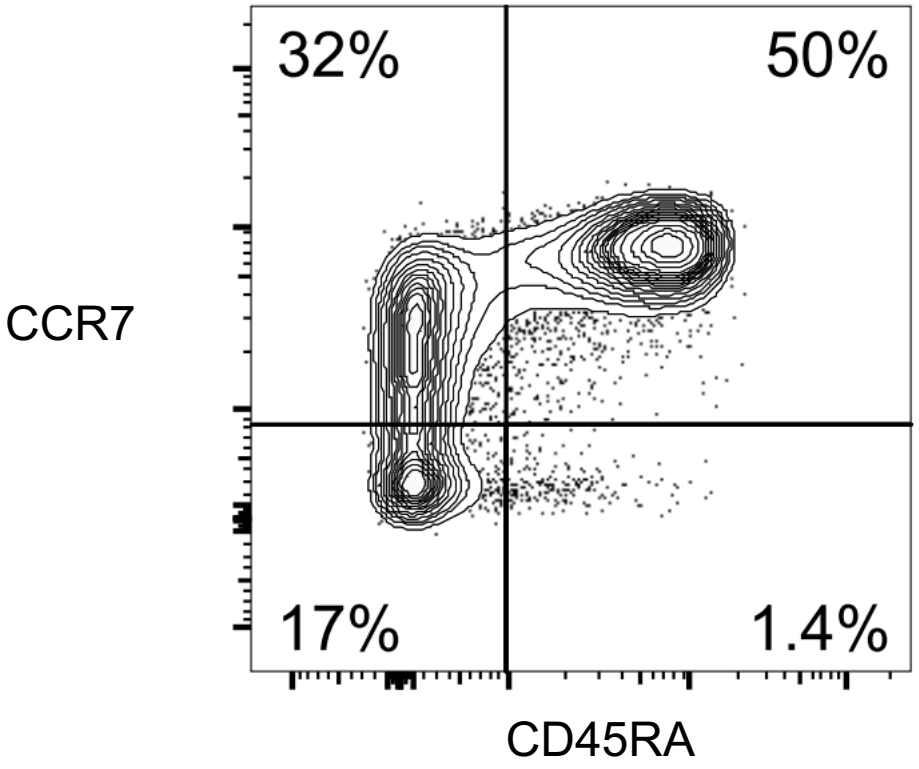


4 populations poorly captured using solely RNA information…

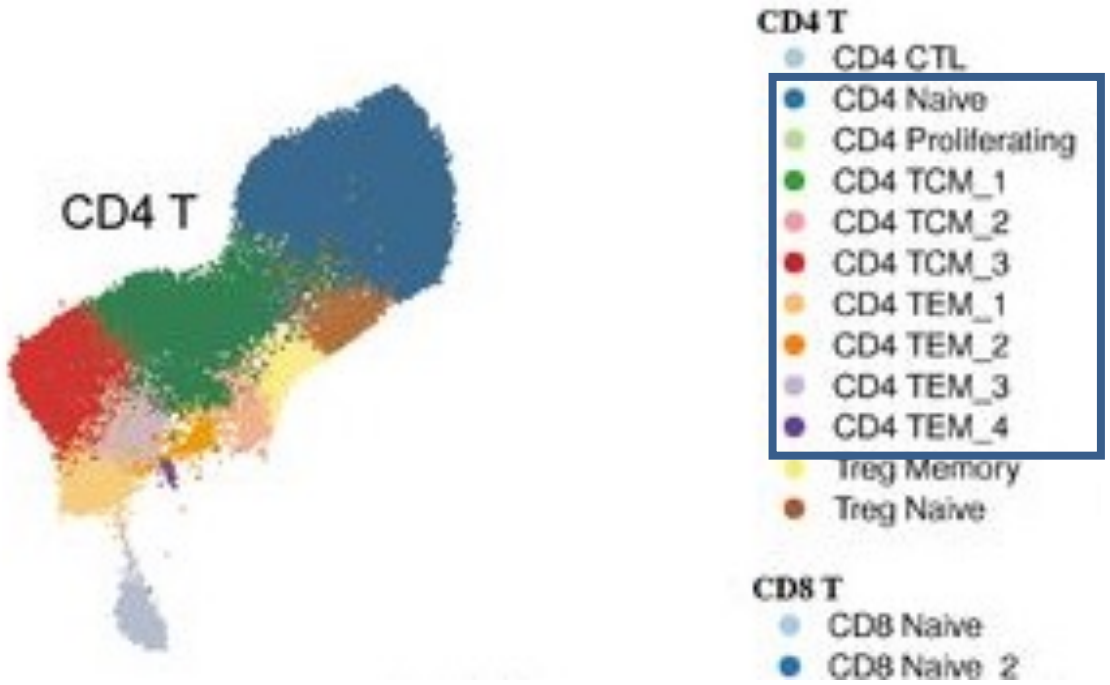


What are the advantages?

These 4 CD4 T cell populations functionally different based on decades of study



Adding surface protein expression information helps in population discrimination.

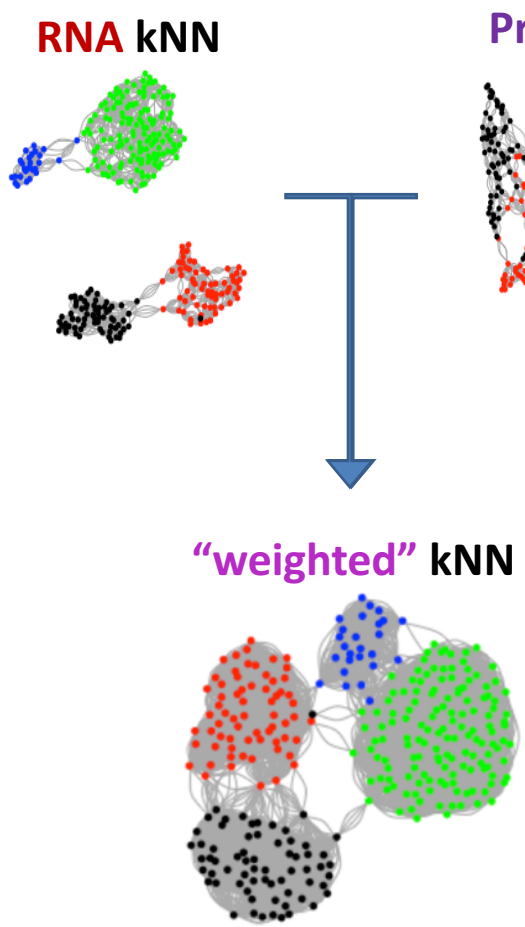


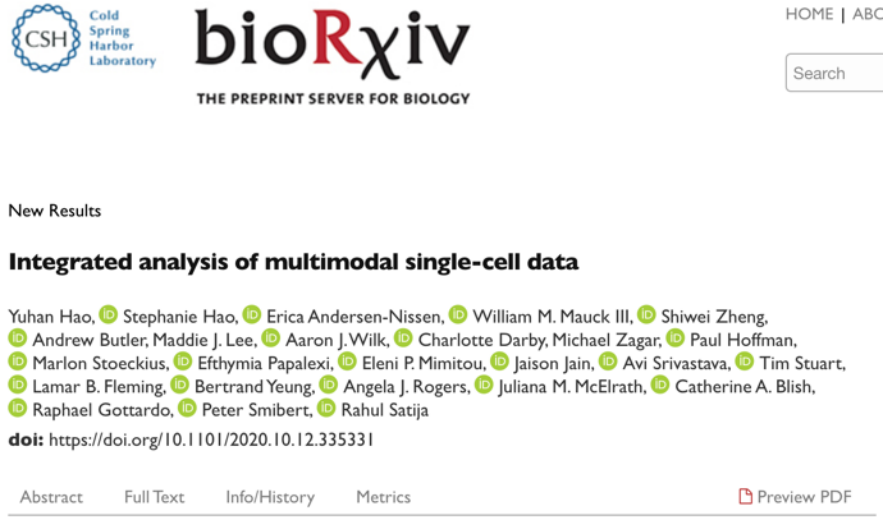
Building in an additional data modality into our analysis

Single cell matrix
(n x m)

Single cell matrix
(n x d)

n = number of cells
m = number of genes
d = number of protein markers





Further reading and useful references

References

- Software packages
  - Seurat (R): https://satijalab.org/seurat
  - scanpy (python): https://scanpy.readthedocs.io/en/stable/
  - scVi (python): https://www.scvi-tools.org/en/stable/
- Single cell technologies
  - Droplet-based 10X profiling (RNA, surface protein, immune repertoire, perturbation) https://www.10xgenomics.com/products/single-cell-gene-expression
  - SeqWell http://shaleklab.com/resource/seq-well/
  - NYGC: https://www.nygenome.org/labs/technology-innovation-lab/
  - Combinatorial indexing: https://cole-trapnell-lab.github.io/projects/sc-rna/
- Preprocessing techniques
  - Doublet detection with scrublet (Wolock SL et al. https://www.cell.com/cell-systems/pdfExtended/S2405-4712(18)30474-5)
  - Background RNA with SoupX (Young MD et al. https://academic.oup.com/gigascience/article/9/12/giaa151/6049831)

References

- Count data transformation
  - sctransform (part of Seurat now, Hafemeister et al. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1874-1)
  - GLM-PCA (Townes FW et al. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1861-6)
- Data integration (aka batch correction)
  - Seurat v3 (Stuart T et al. https://www.cell.com/cell/pdf/S0092-8674(19)30559-8.pdf)
  - Harmony (in Seurat now, Korsunsky I et al. https://www.nature.com/articles/s41592-019-0619-0)
  - bbKNN (in scanpy now, Polanski K et al. https://academic.oup.com/bioinformatics/article/36/3/964/5545955)
- Clustering
  - Louvain & Leiden modularity detection (http://bioconductor.org/books/release/OSCA/clustering.html#clustering-graph)
- Machine learning methods
  - scVI
  - DESC (Li X et al. https://www.nature.com/articles/s41467-020-15851-3)

References

- Data visualization
  - UMAP (in Seurat and scanpy now, also at https://umap-learn.readthedocs.io/en/latest/ and R https://cran.r-project.org/web/packages/uwot/index.html)
  - t-SNE (in Seurat)
  - PHATE (Moon KR et al. https://www.nature.com/articles/s41587-019-0336-3)
  - ForceAtlas2 (in scanpy)
- Trajectory inference
  - Slingshot (Street K et al. https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-018-4772-0)
  - PAGA (Wolf FA et al. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1663-x)
  - RNA velocity (part of scVelo package in python, most recent pub at Bergen V et al. https://www.nature.com/articles/s41587-020-0591-3)
- Multi-modal data anlysis
  - Weighted nearest neighbors (in Seurat now, Hao Y et al. https://www.biorxiv.org/content/10.1101/2020.10.12.335331v1.full)
  - totalVI (part of scVI, Gayoso A et al. https://www.biorxiv.org/content/10.1101/2020.05.08.083337v2)

References

- Useful reference dictionary: http://bioconductor.org/books/release/OSCA/

Most publications have publicly available data you can use!!!

Orchestrating Single-Cell Analysis with Bioconductor

Authors: *Robert Amezquita [aut], Aaron Lun [aut, cre], Stephanie Hicks [aut], Raphael Gottardo [aut]*

Version: *1.0.6*

Modified: *2020-11-13*

Compiled: *2020-12-08*

Environment: *R version 4.0.3 (2020-10-10), Bioconductor 3.12*

License: *CC BY-NC-ND 3.0 US*

Copyright: *Bioconductor, 2020*

Source: *https://github.com/Bioconductor/OrchestratingSingleCellAnalysis*

On cyclical structures, RNA velocity, and trajectory inference

- Gene expression dynamics of cyclical structures in dimension reductions:
  - See Figure 2 of https://www.nature.com/articles/s41590-018-0181-4
- Other packages exist amenable to or specific to cyclical analyses:
  - See Figure 2 of https://www.nature.com/articles/s41587-019-0071-9
- Incorporation of RNA velocity into cell cycle and visualization
  - See preprint at https://twitter.com/lylaatta/status/1355161798845095936
- RNA velocity pre-processing matters
  - https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1008585
- Latest RNA velocity publication
  - https://www.nature.com/articles/s41587-020-0591-3
  - See scVelo package: https://scvelo.readthedocs.io/
- Framework for gene expression analyses in trajectory inference
  - See tradeSeq package: https://www.nature.com/articles/s41467-020-14766-3

Thanks!