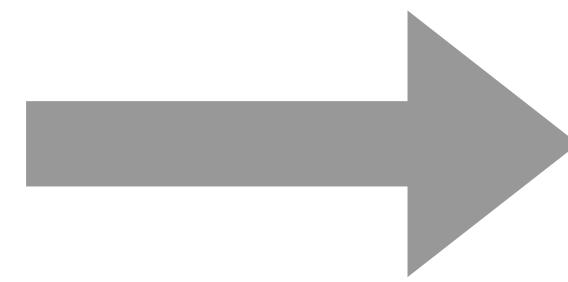
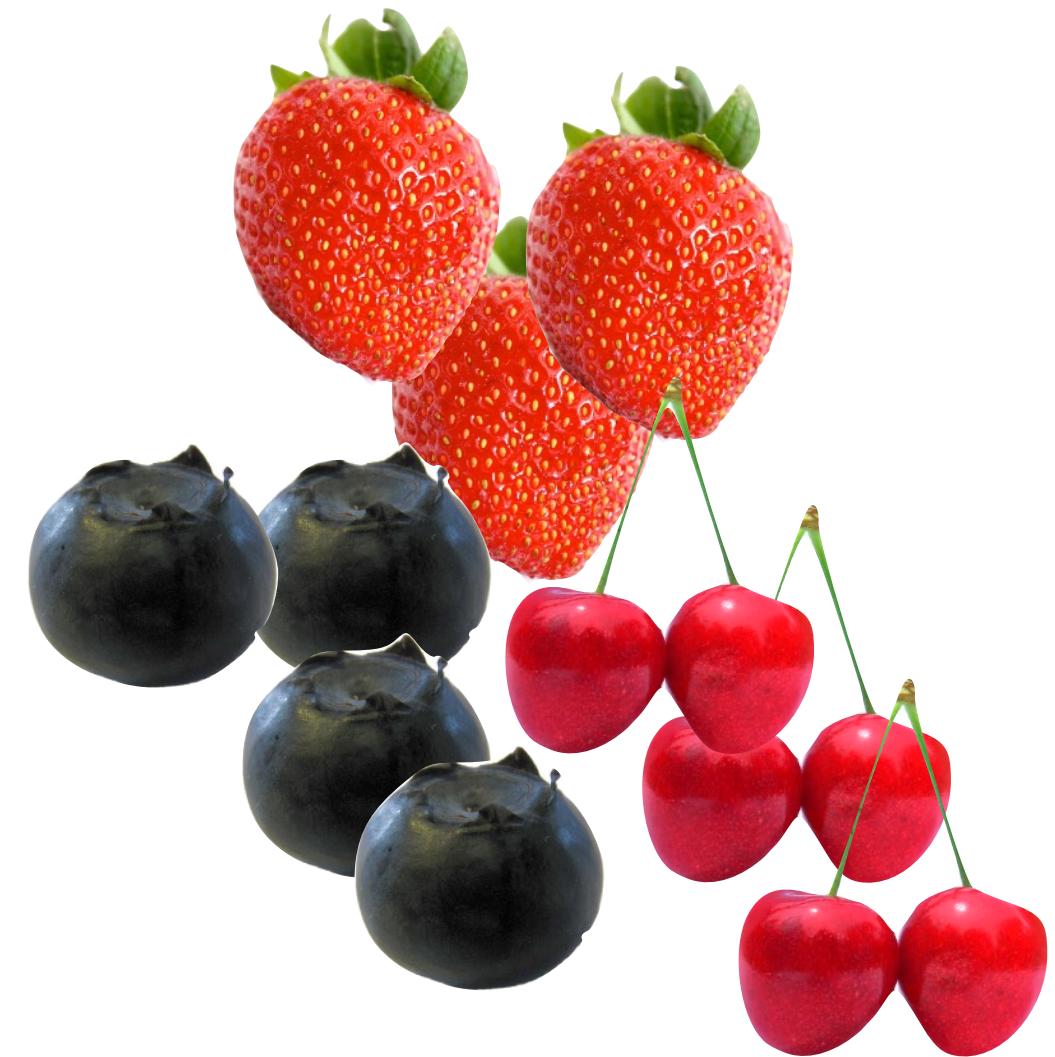


A selection of current challenges in scRNA-seq analysis

Charlotte Soneson
2020-12-01

Bulk RNA-seq

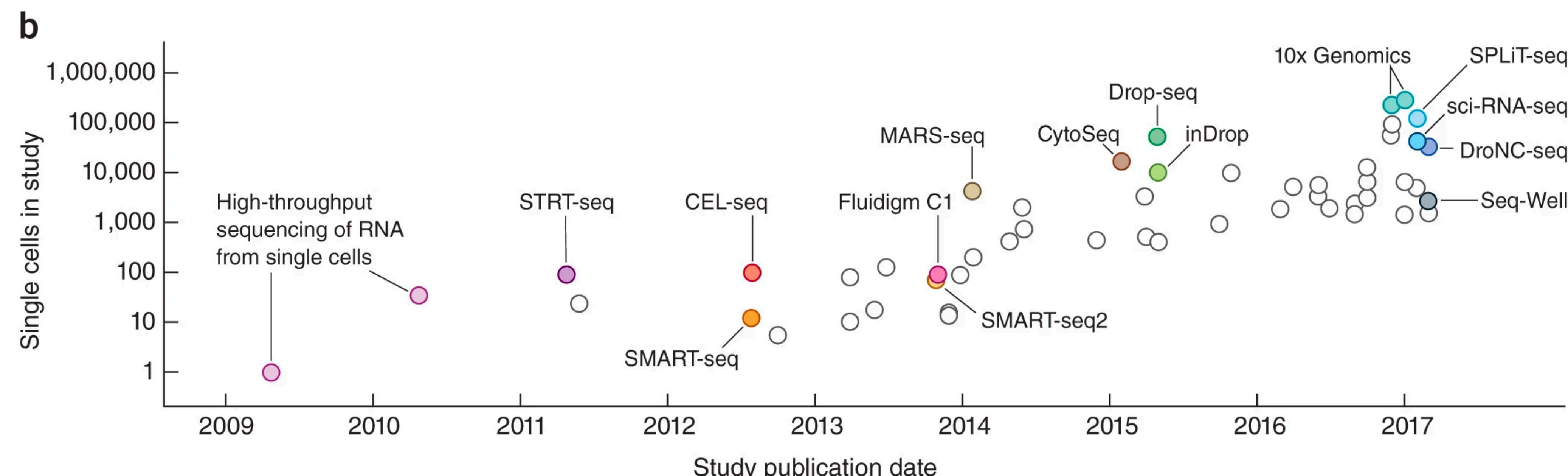
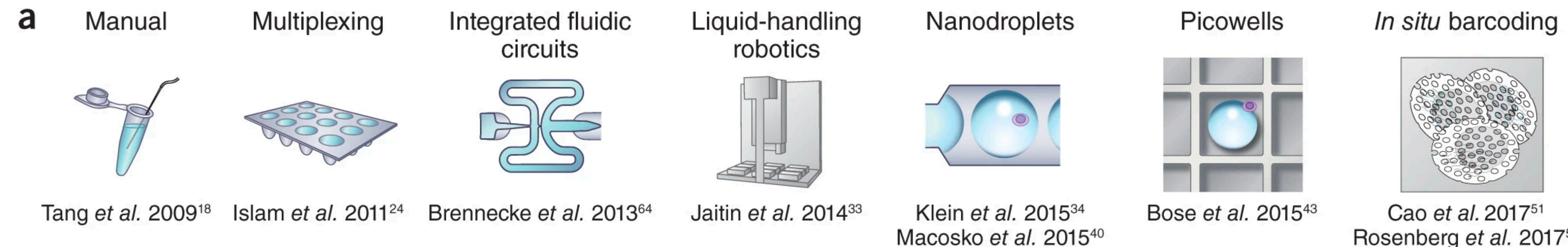
Thousands to millions of individual cells



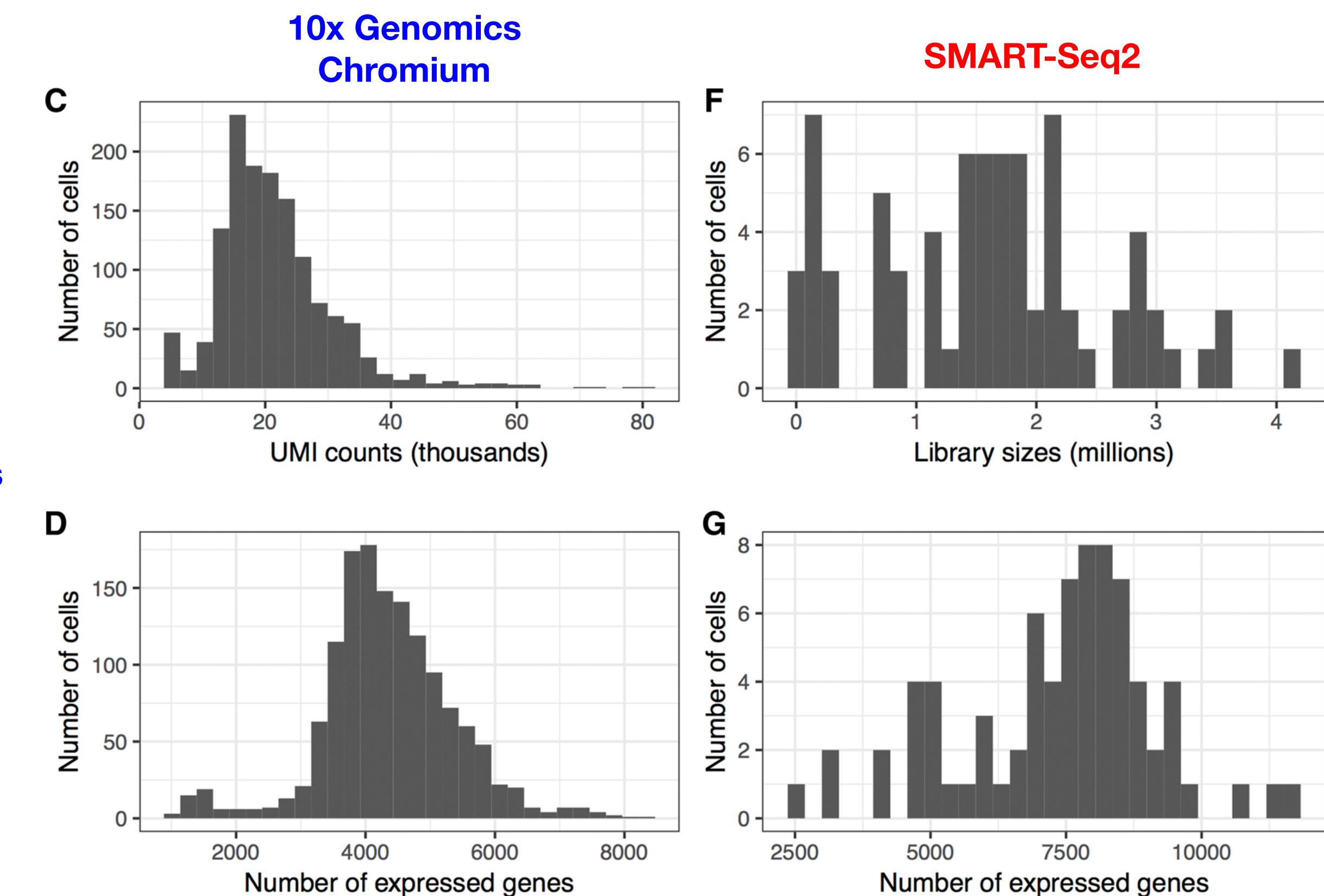
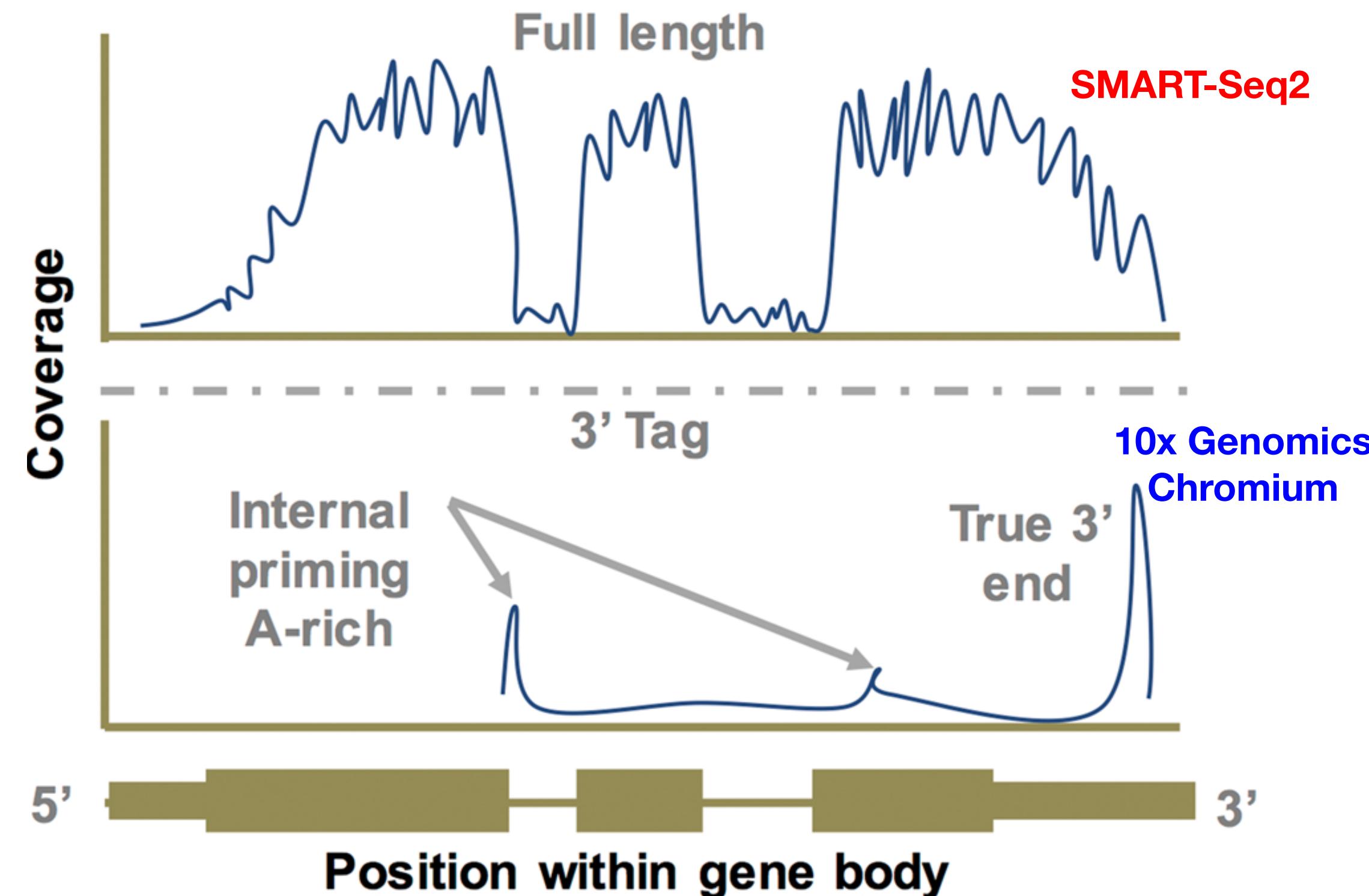
Averaged abundances of genes

- Many samples consist of several different cell types; composition may be different between conditions
- Not all cells of the same type are identical
- Rare cell types may be missed

scRNA-seq protocols

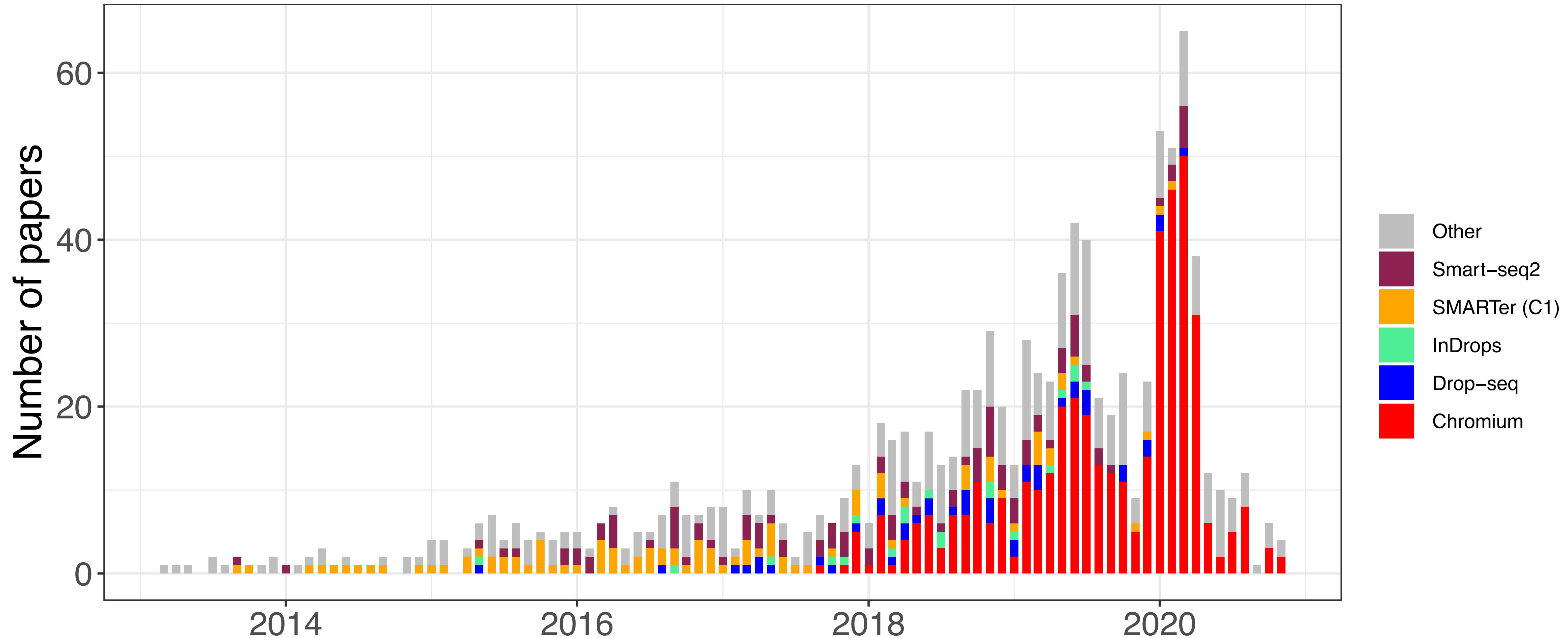


SMART-Seq2 vs 10x Genomics Chromium

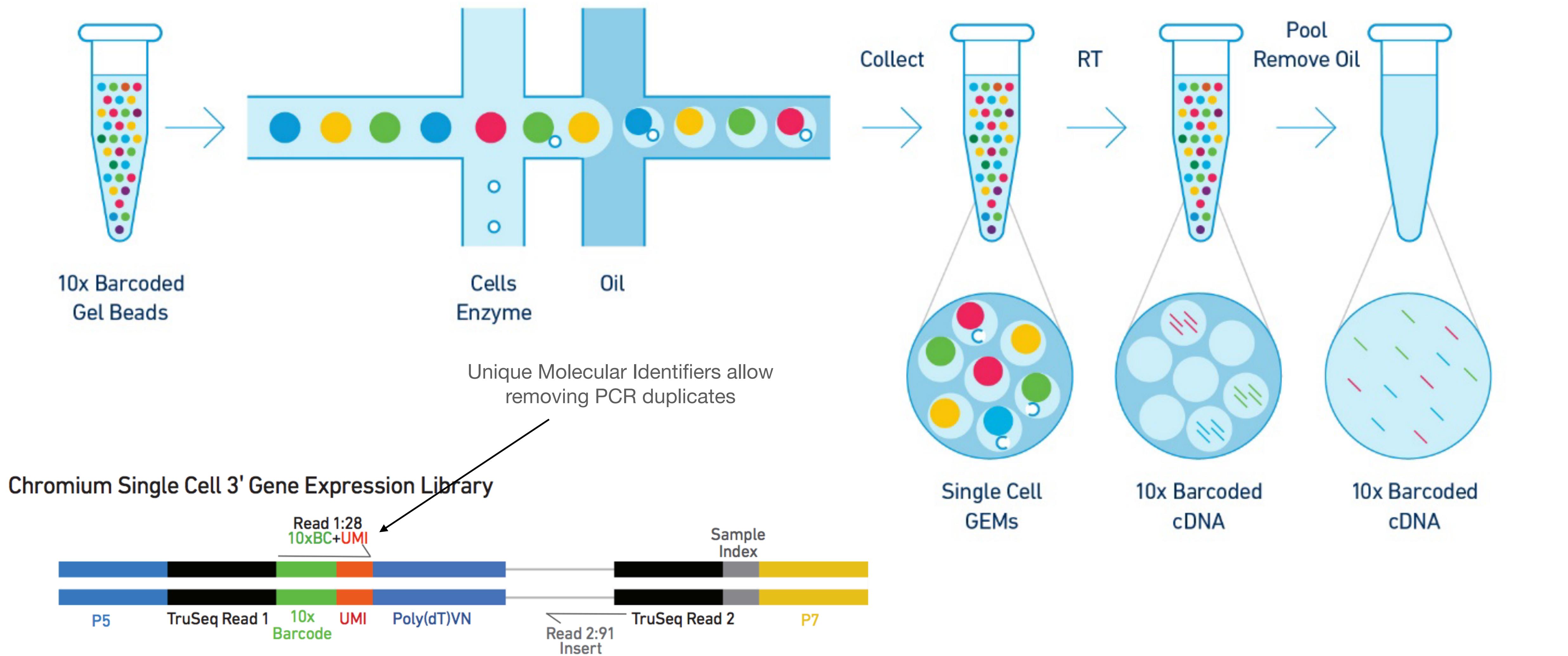


Protocol use over time

Data from <http://www.nxn.se/single-cell-studies/gui>

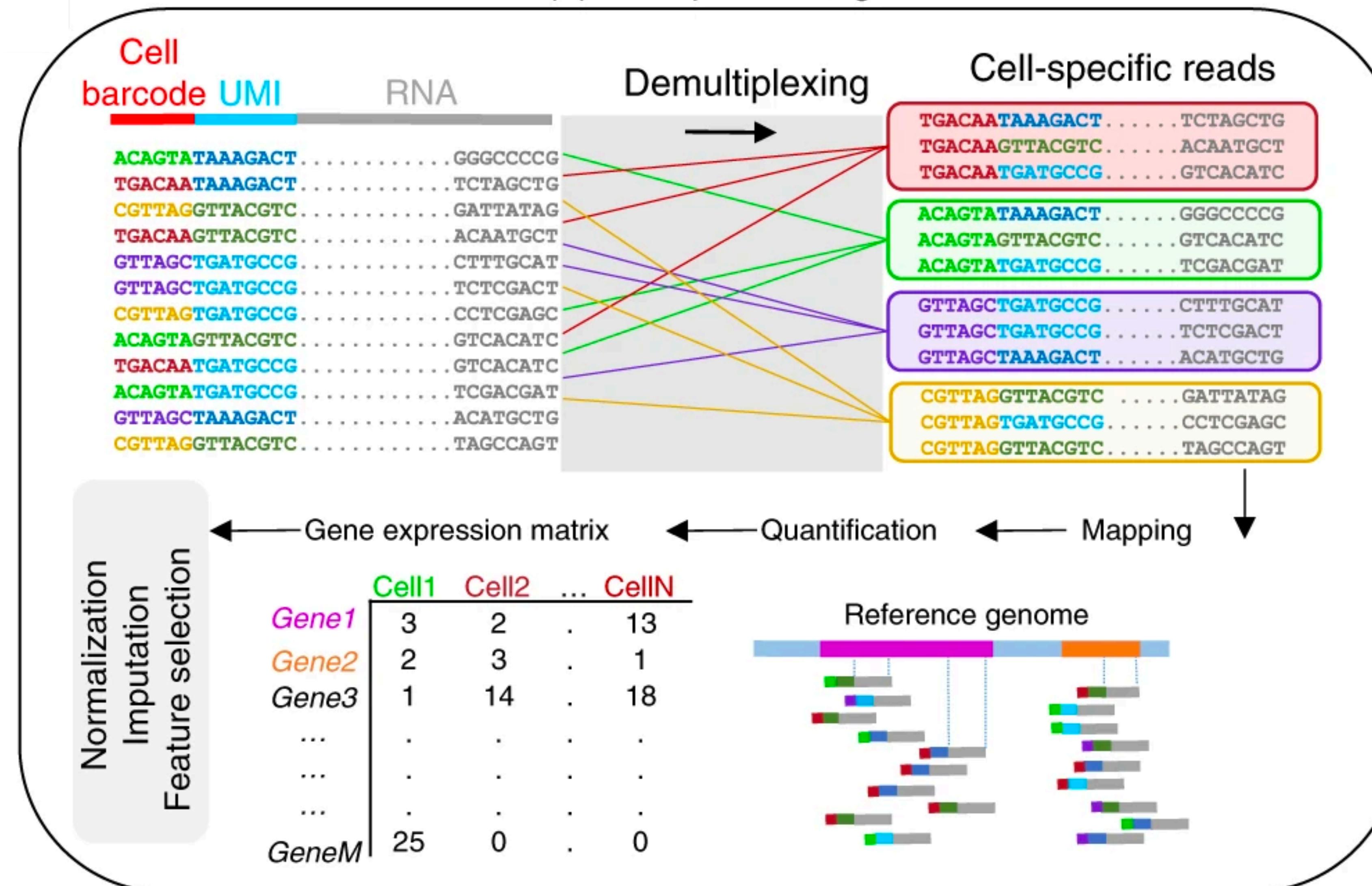


10x Genomics Chromium

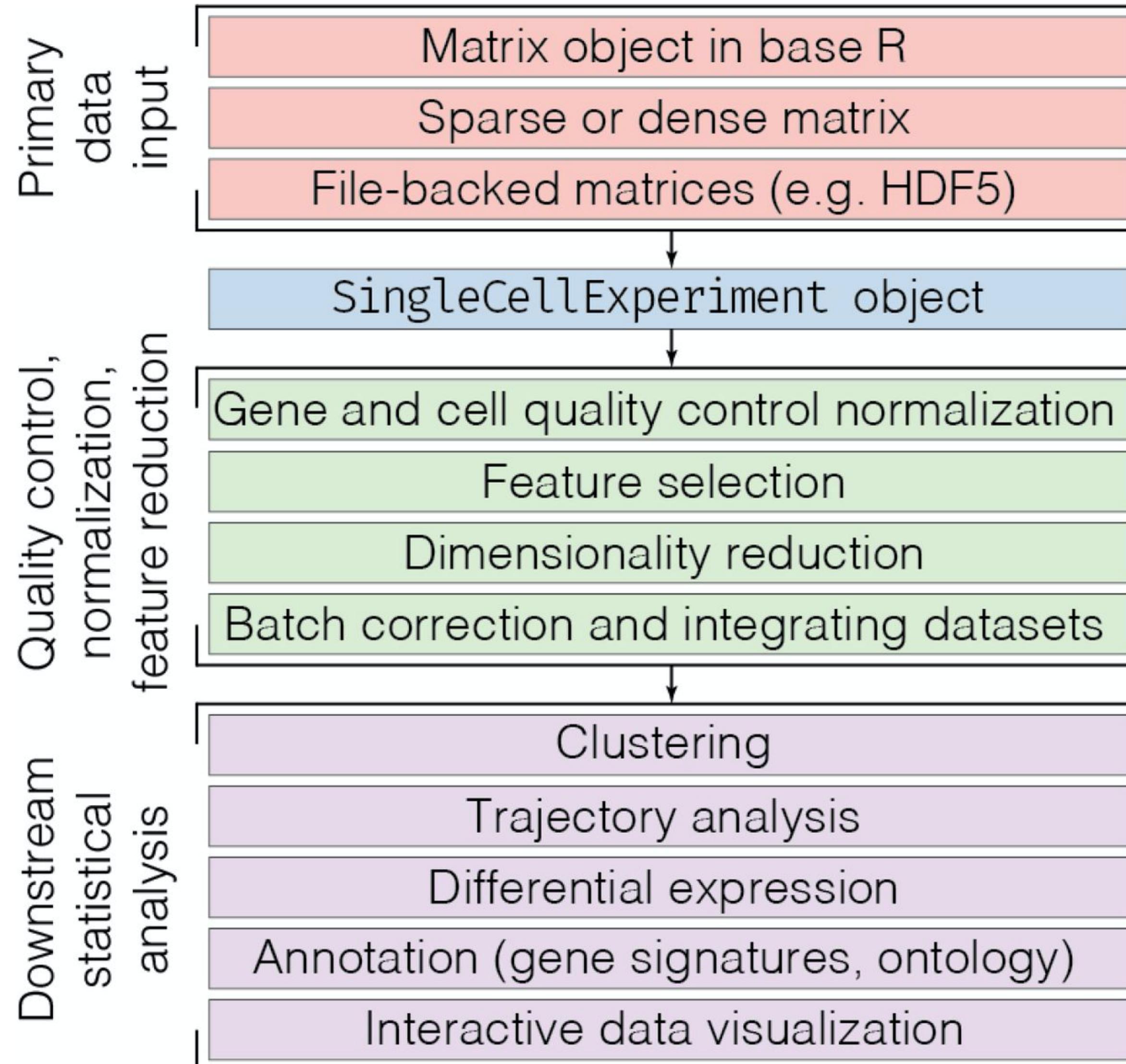


A 'typical' scRNA-seq workflow - preprocessing

(3) Data processing



A 'typical' scRNA-seq workflow - analysis



Orchestrating single-cell analysis with Bioconductor

Robert A. Amezquita, Aaron T. L. Lun, Etienne Becht, Vince J. Carey, Lindsay N. Carpp, Ludwig Geistlinger, Federico Marini, Kevin Rue-Albrecht, Davide Risso, Charlotte Soneson, Levi Waldron, Hervé Pagès, Mike L. Smith, Wolfgang Huber, Martin Morgan, Raphael Gottardo [✉](#) & Stephanie C. Hicks [✉](#)

Nature Methods 17, 137–145(2020) | [Cite this article](#)

<http://bioconductor.org/books/release/OSCA/>

Orchestrating Single-Cell Analysis with Bioconductor

2020-05-04

Welcome

This is the website for “**Orchestrating Single-Cell Analysis with Bioconductor**”, a book that teaches users some common workflows for the analysis of single-cell RNA-seq data (scRNA-seq). This book will teach you how to make use of cutting-edge Bioconductor tools to process, analyze, visualize, and explore scRNA-seq data. Additionally, it serves as an online companion for the manuscript “**Orchestrating**



Lots of challenges remaining!

Review | **Open Access** | Published: 07 February 2020

Eleven grand challenges in single-cell data science

[David Lähnemann](#), [Johannes Köster](#), [Ewa Szczurek](#), [Davis J. McCarthy](#), [Stephanie C. Hicks](#), [Mark D. Robinson](#), [Catalina A. Vallejos](#), [Kieran R. Campbell](#), [Niko Beerenwinkel](#), [Ahmed Mahfouz](#), [Luca Pinello](#), [Pavel Skums](#), [Alexandros Stamatakis](#), [Camille Stephan-Otto Attolini](#), [Samuel Aparicio](#), [Jasmijn Baaijens](#), [Marleen Balvert](#), [Buys de Barbanson](#), [Antonio Cappuccio](#), [Giacomo Corleone](#), [Bas E. Dutilh](#), [Maria Florescu](#), [Victor Guryev](#), [Rens Holmer](#), [Katharina Jahn](#), [Thamar Jessurun Lobo](#), [Emma M. Keizer](#), [Indu Khatri](#), [Szymon M. Kielbasa](#), [Jan O. Korbel](#), [Alexey M. Kozlov](#), [Tzu-Hao Kuo](#), [Boudewijn P.F. Lelieveldt](#), [Ion I. Mandoiu](#), [John C. Marioni](#), [Tobias Marschall](#), [Felix Mölder](#), [Amir Niknejad](#), [Lukasz Raczkowski](#), [Marcel Reinders](#), [Jeroen de Ridder](#), [Antoine-Emmanuel Saliba](#), [Antonios Somarakis](#), [Oliver Stegle](#), [Fabian J. Theis](#), [Huan Yang](#), [Alex Zelikovsky](#), [Alice C. McHardy](#), [Benjamin J. Raphael](#), [Sohrab P. Shah](#) & [Alexander Schönhuth](#)  -Show fewer authors

Valentine Svensson
@vallens

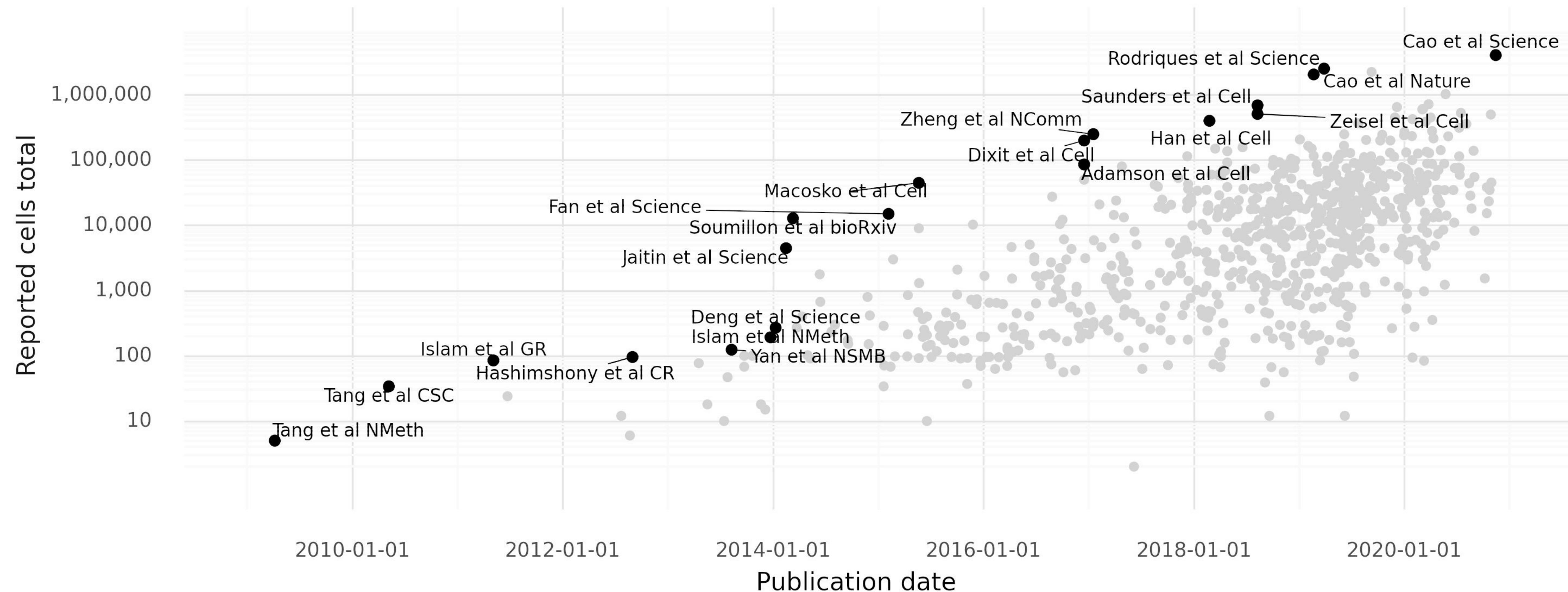
A human cell atlas of fetal gene expression -
science.science.org/content/370/65...

New record-breaking scRNA-seq study with 4,000,000 cells

6:59 AM · Nov 13, 2020 · Twitter Web App

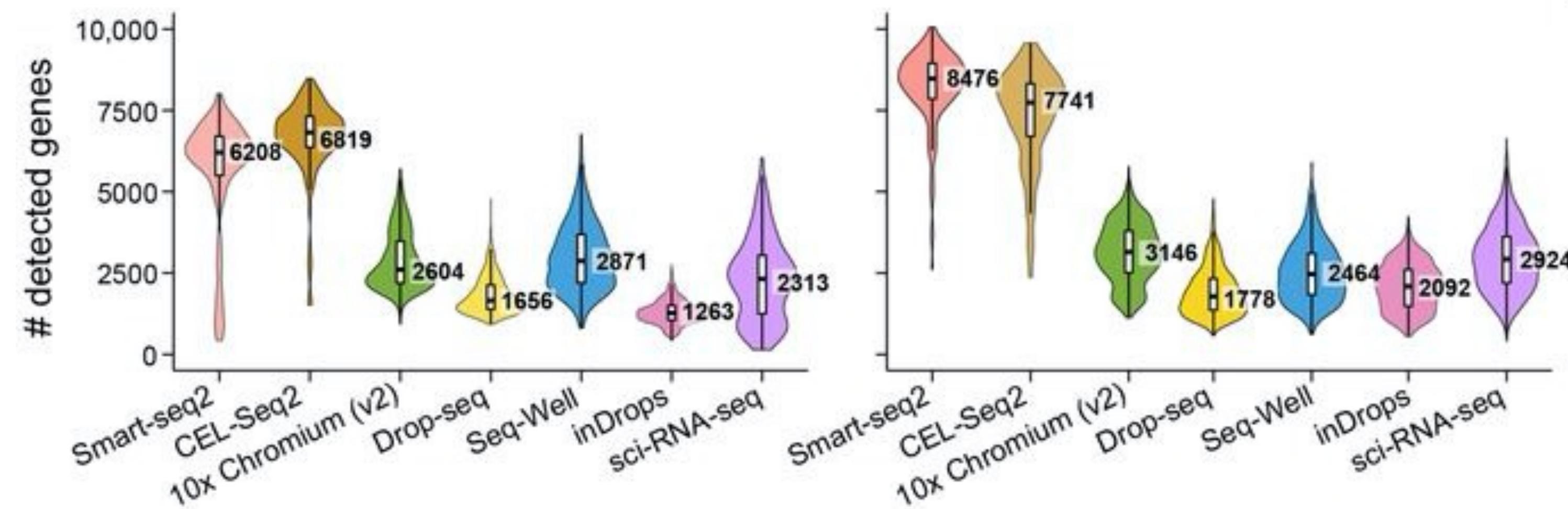
Challenge: scalability

Cells per study over time



Challenge: sparsity and varying detection rates

- Lots of zeros in the count matrix (protocol-dependent); also strongly associated with the number of genes that can be detected in a single cell.

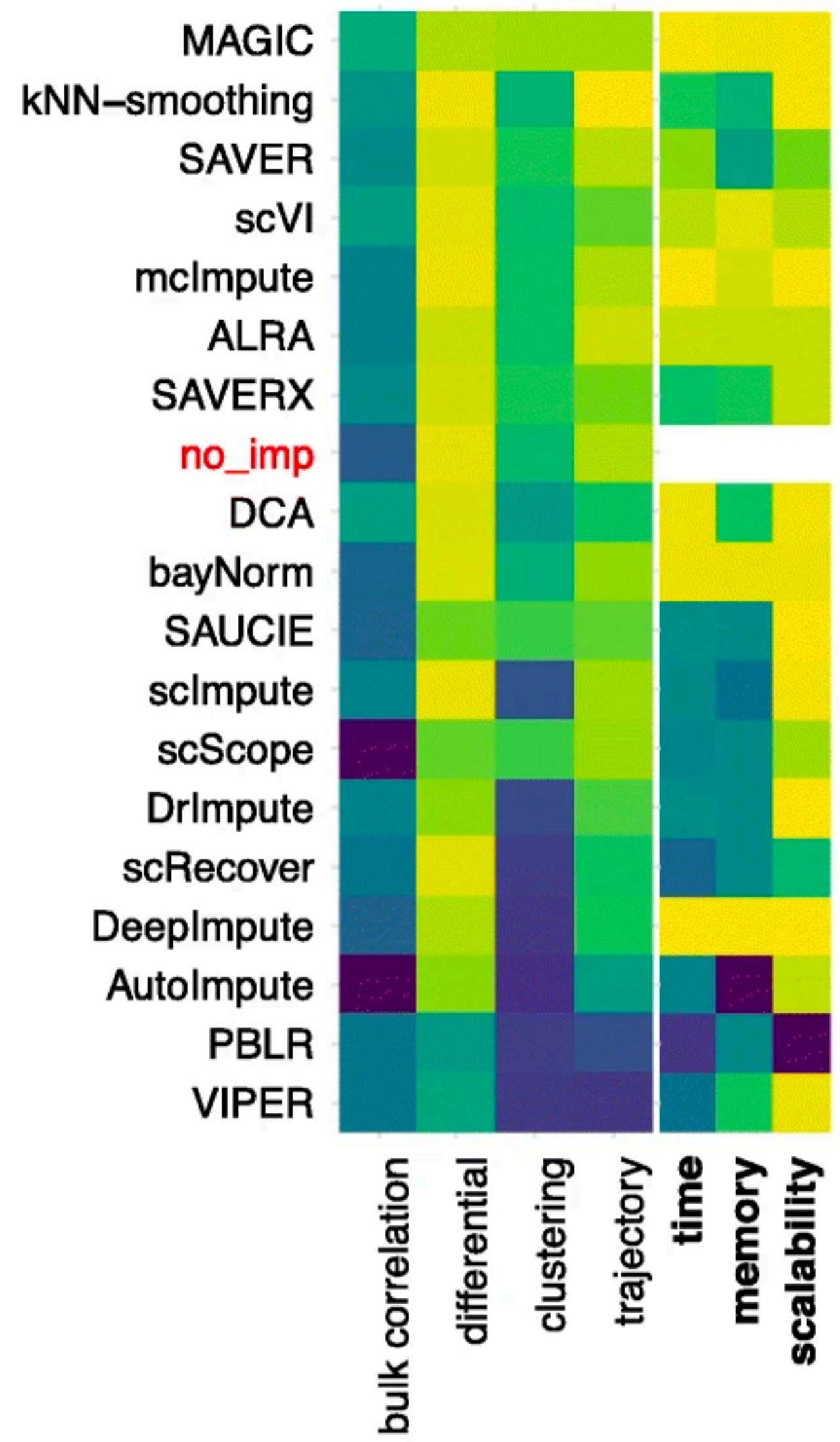
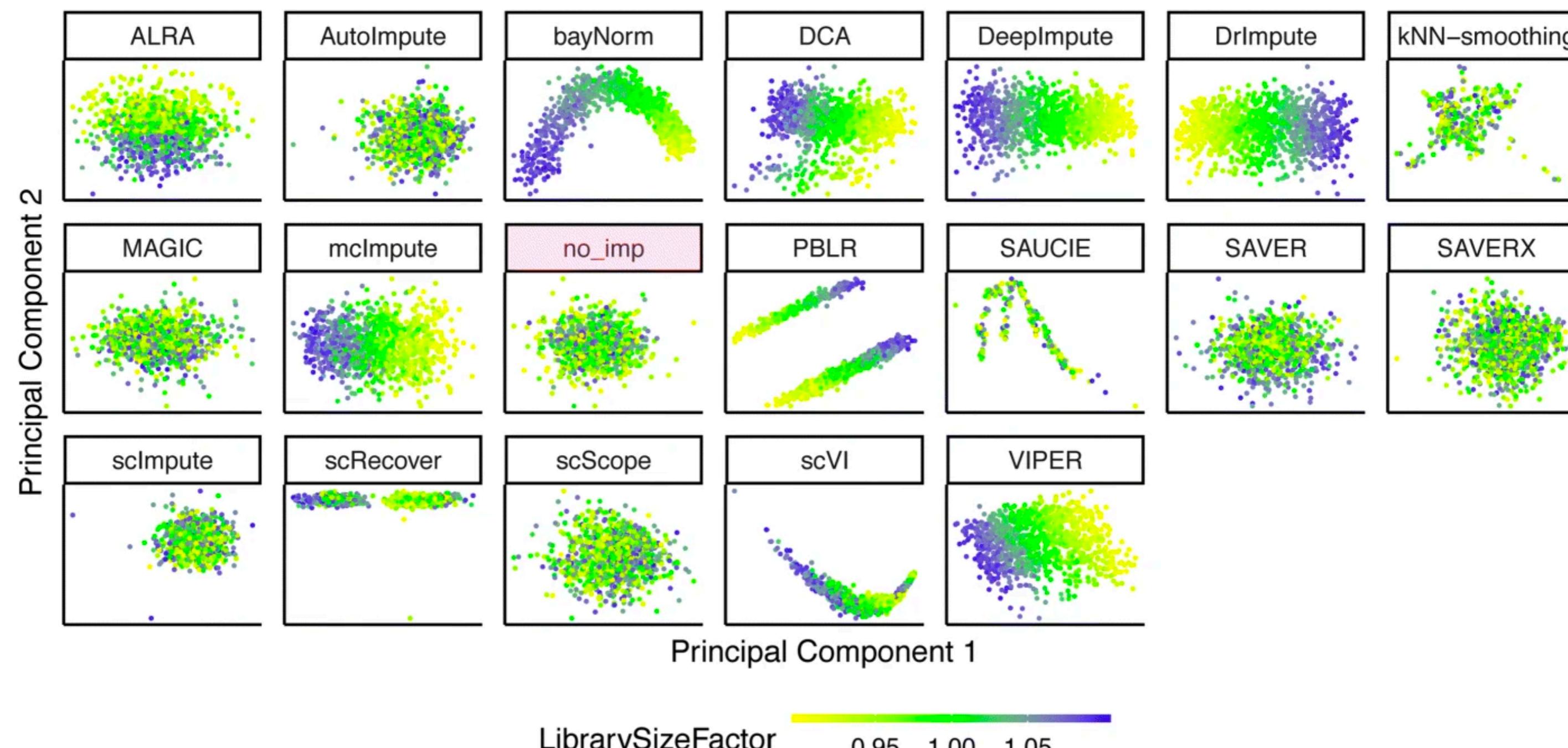


- Problematic for normalization, log-transformation, interpretation.
- Directly modeling the counts may provide a way around the issue.

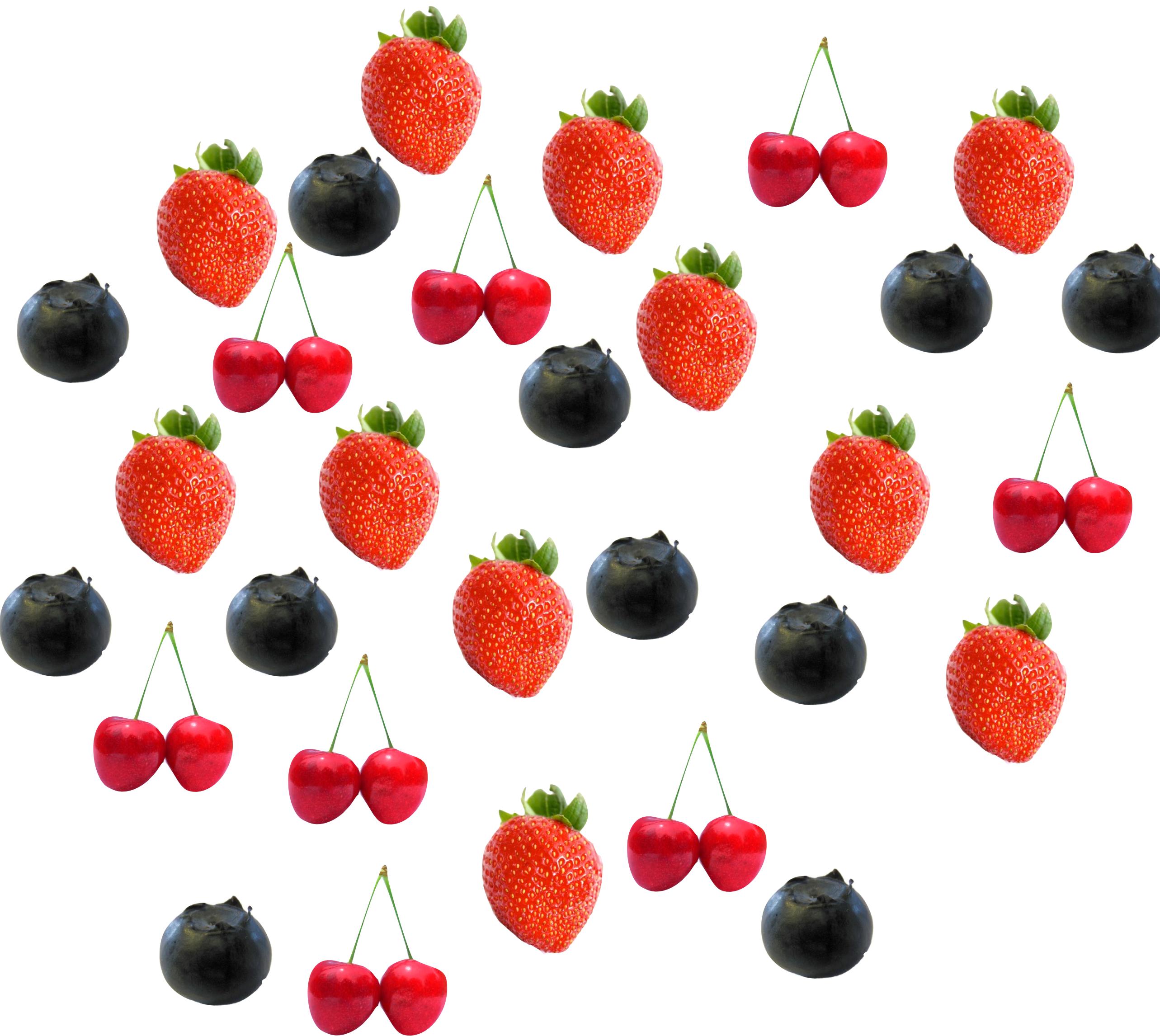


Imputation - yes or no?

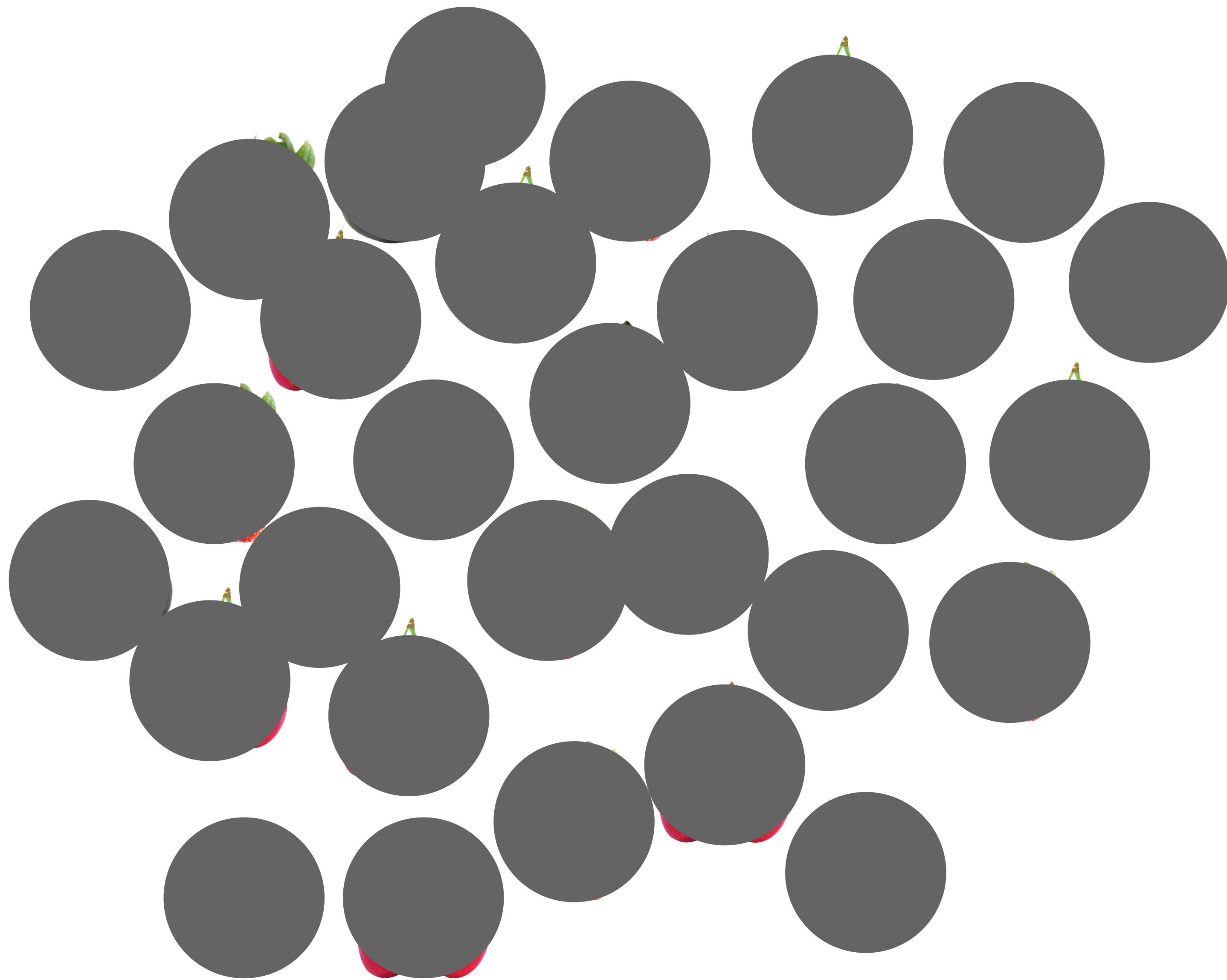
"We found that the majority of scRNA-seq imputation methods outperformed no imputation in recovering gene expression observed in bulk RNA-seq. However, the majority of the methods did not improve performance in downstream analyses compared to no imputation, in particular for clustering and trajectory analysis, and thus should be used with caution."



Challenge: cell type annotation

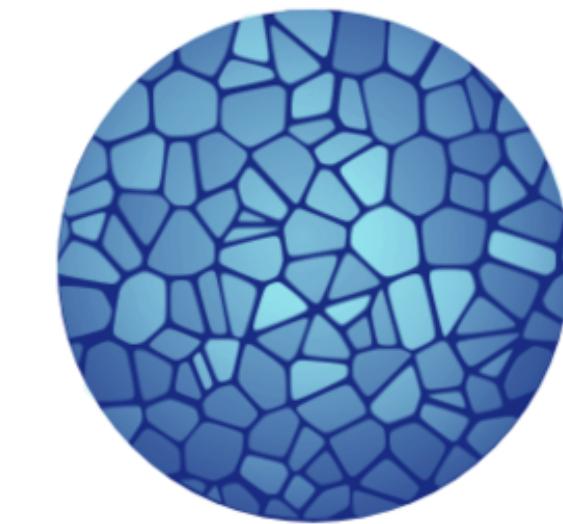


Challenge: cell type annotation

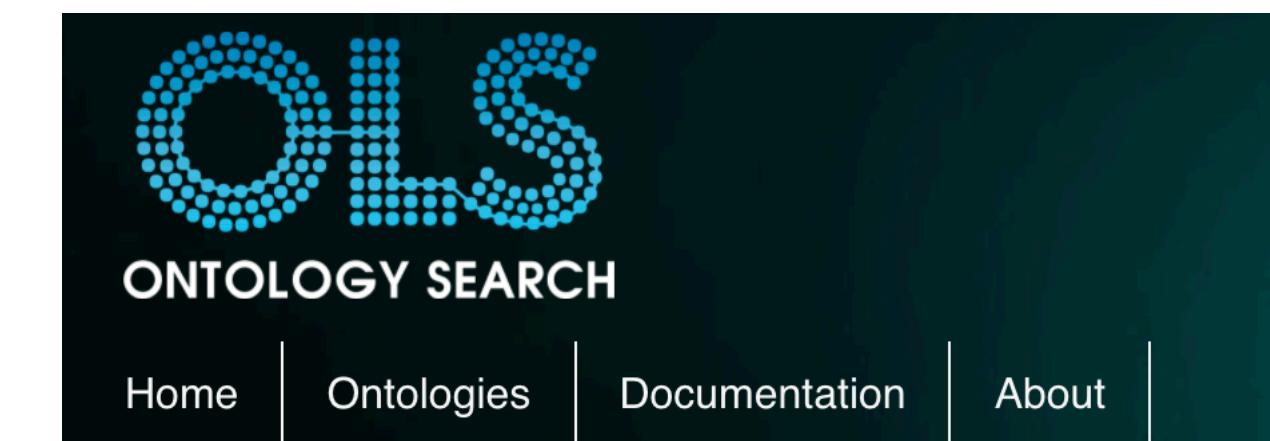


Challenge: cell type annotation

- Compare to large, stable reference data sets and automatically assign a cell type label
- Batch effects, protocol differences
- What is a cell type?
- Make sure that reference atlas is as comprehensive as possible

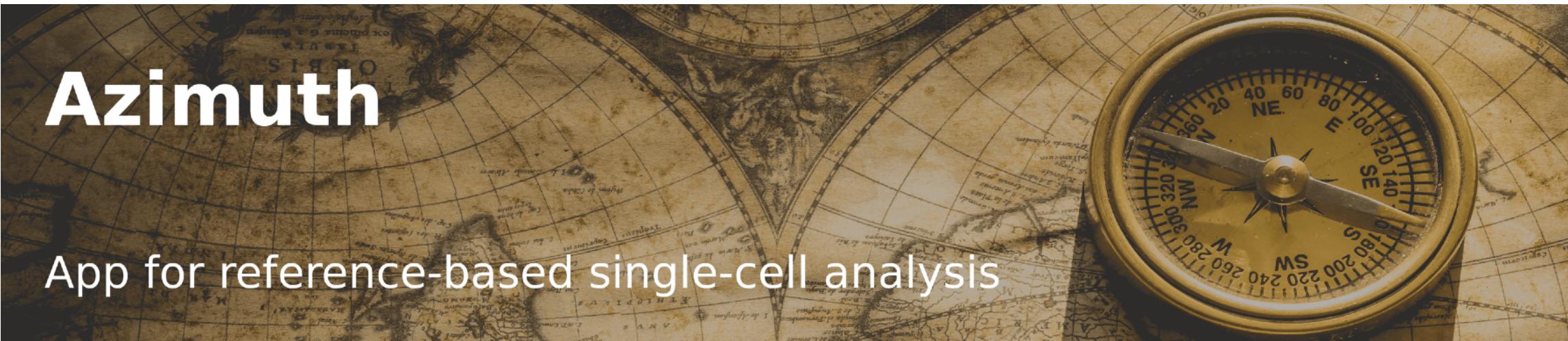


HUMAN
CELL
ATLAS



Cell Ontology

An ontology of cell types.

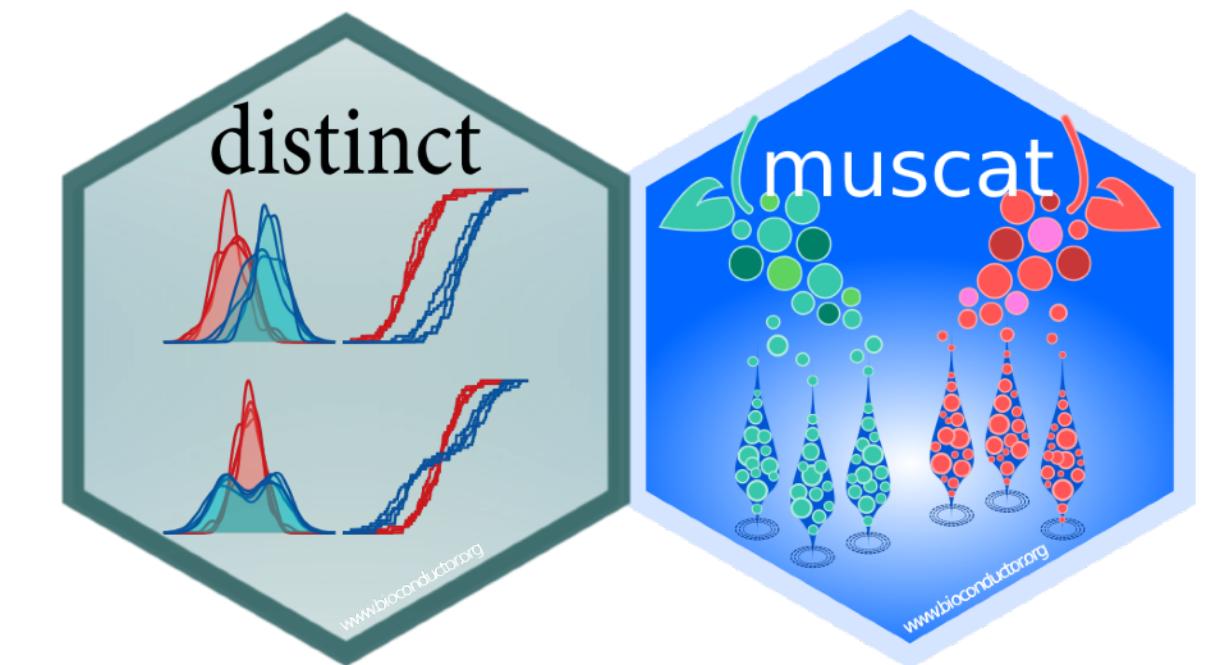


⌚ wangshenguiuc / OnClass

⌚ immunogenomics / symphony

Challenge: multi-sample analysis

- Hierarchical variance structure - cells *within* a biological sample are more highly correlated than cells from *different* biological replicates.
- Cells \neq biological replicates
- For comparison of average expression profiles *between* conditions, *within* a cell type, either
 - create pseudobulk samples, or
 - use mixed models
- Single-cell data allow comparison of full distributions (not just averages) between conditions

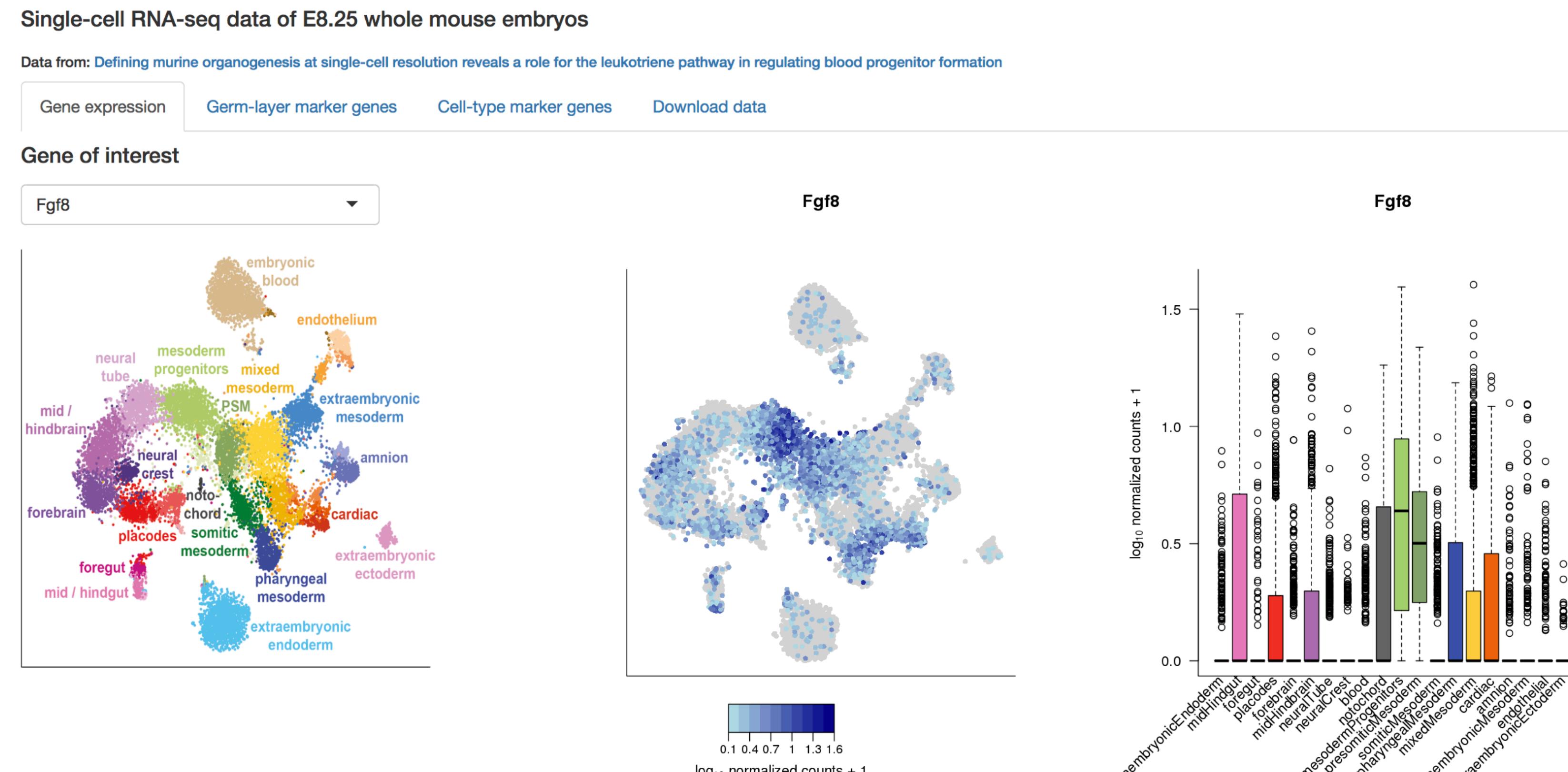


Tiberi et al: *distinct*: a novel approach to differential distribution analyses. *bioRxiv* (2020)

Crowell et al: *muscat* detects subpopulation-specific state transitions from multi-sample multi-condition single-cell transcriptomics data. *Nature Communications* (2020)

Challenge: efficient exploration and sharing

- (Interactive) exploratory analysis is crucial for hypothesis generation, quality control, discussions, dissemination, ...



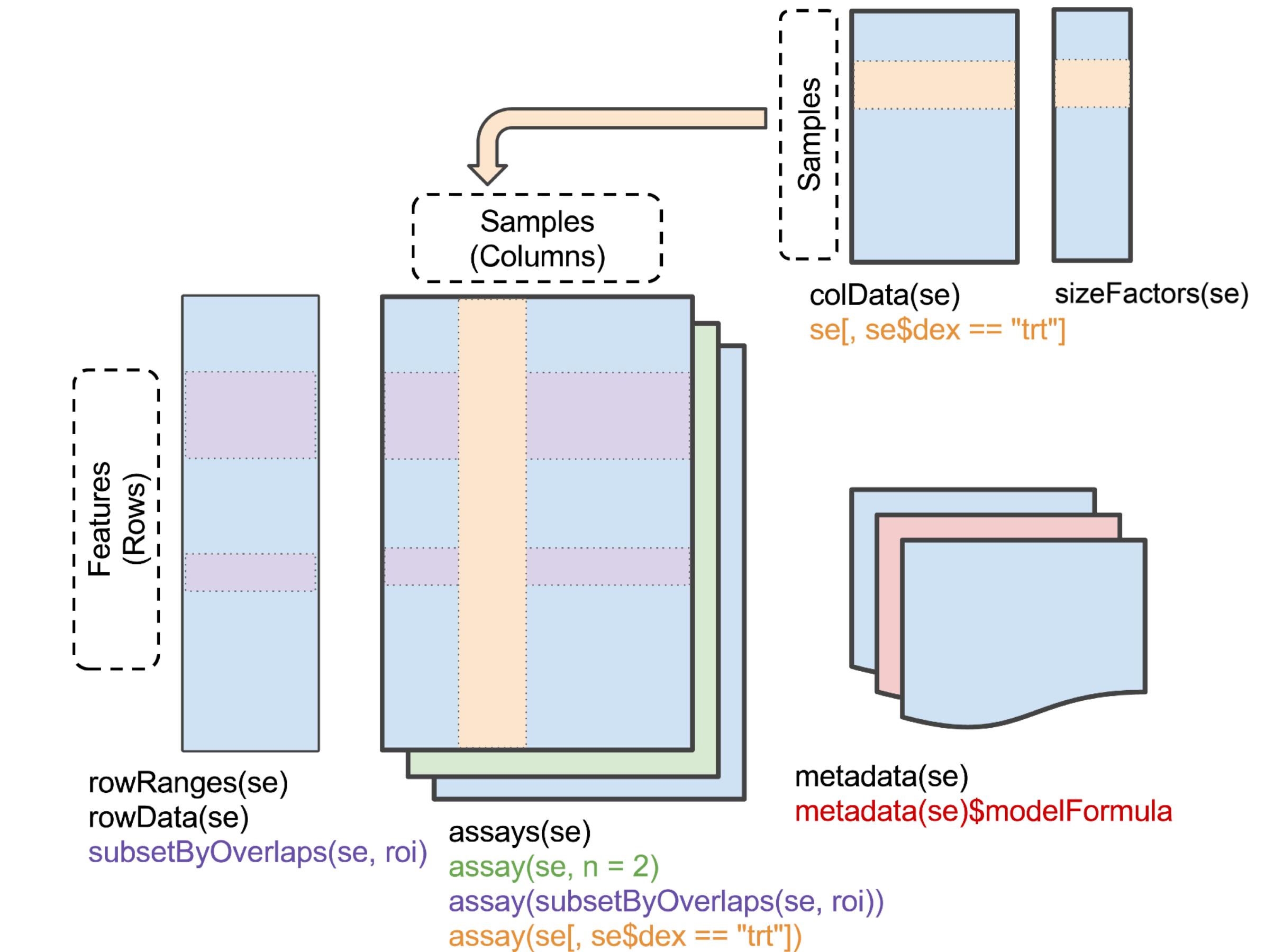


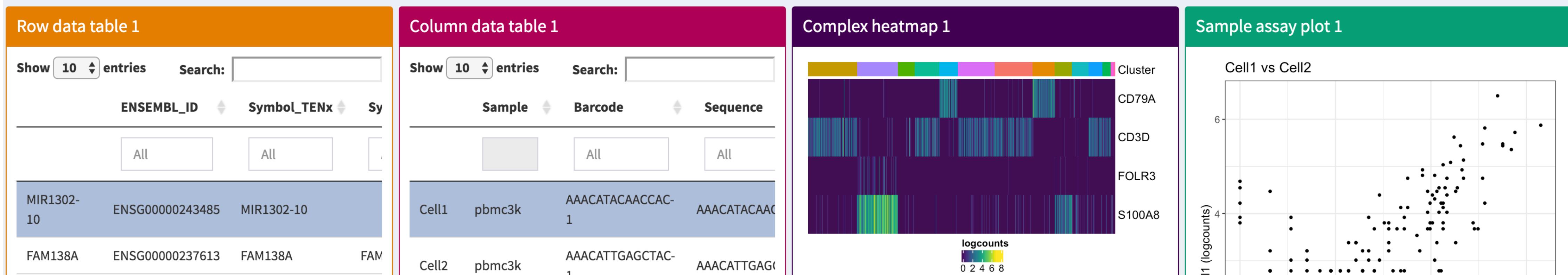
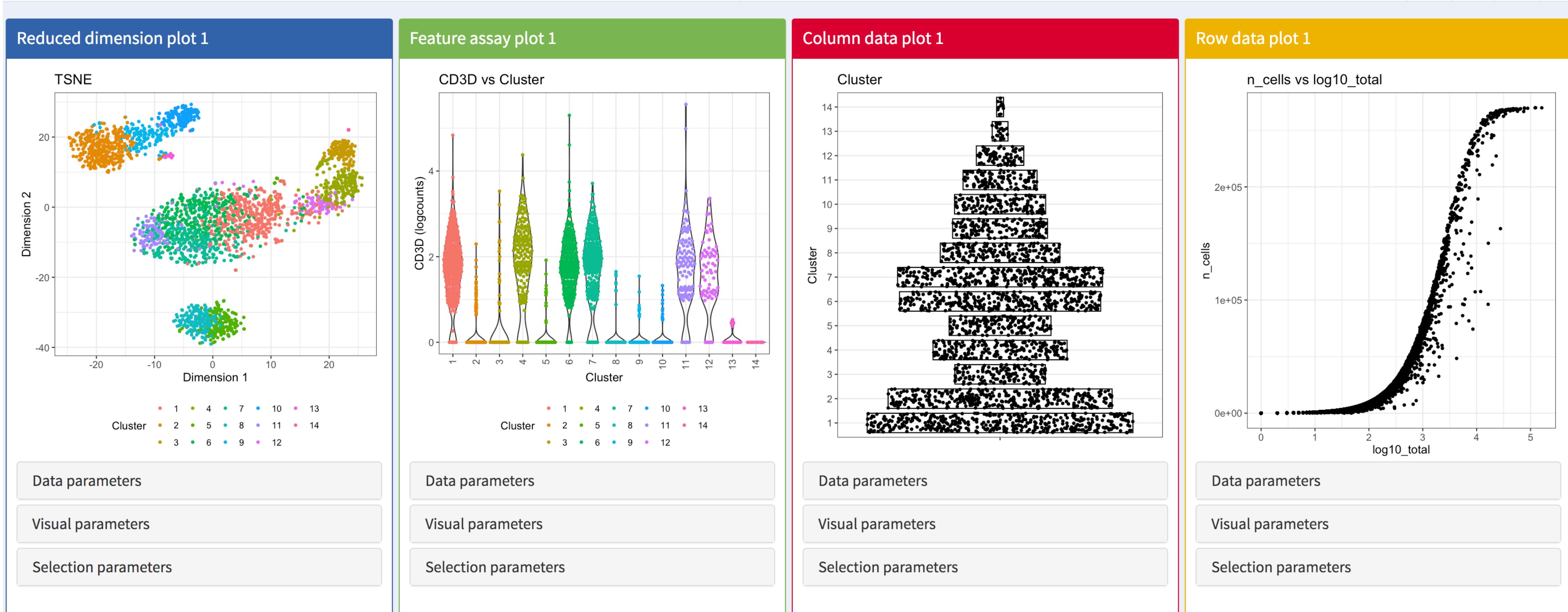
iSEE (interactive SummarizedExperiment Explorer)



- Bioconductor package for interactive exploration of SummarizedExperiment objects (=any 'rectangular' data with row and column annotations).
- Fully customizable (programmatically and interactively) and scalable to large data sets.
- Reproducibility - all code to regenerate figures is provided.
- Tutorial available at

<https://tinyurl.com/isee-tutorial>





Thank you