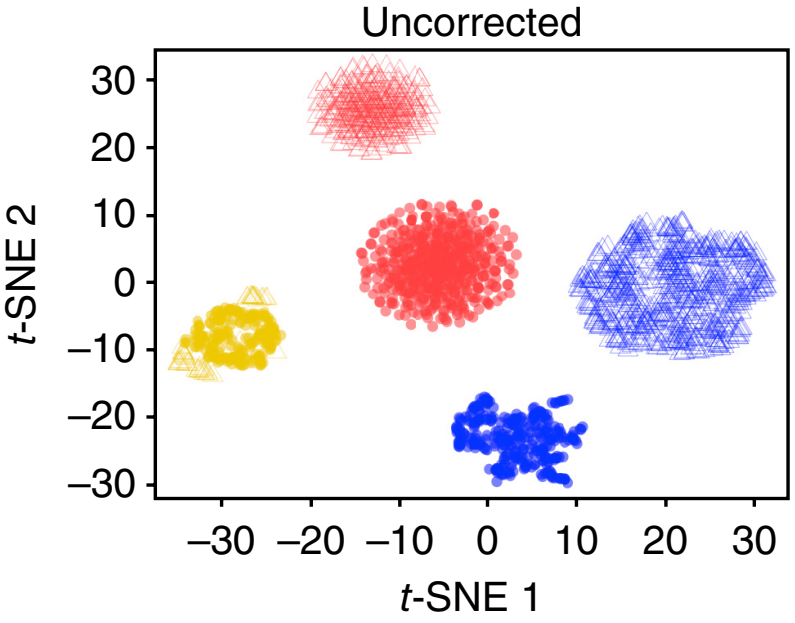
BROAD INSTITUTE

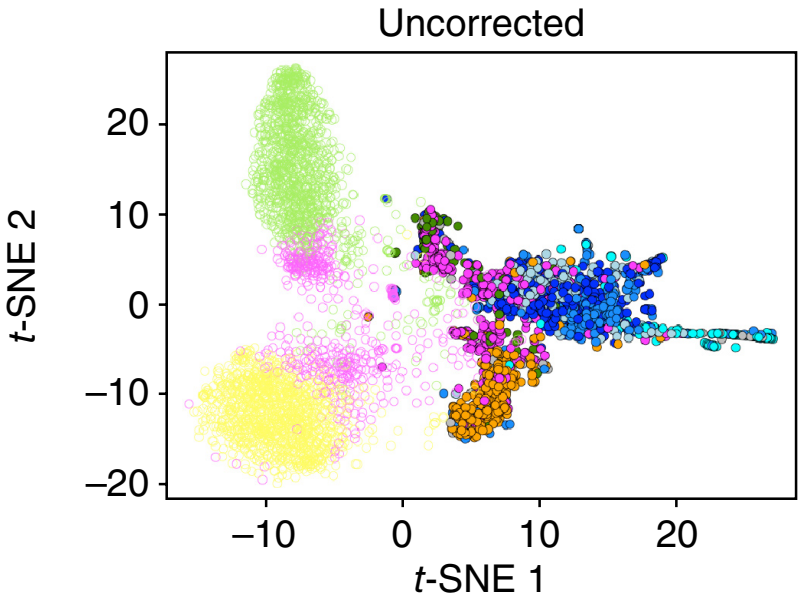
Theory II

Single cell RNAseq tutorial

ISMB 2018, Chicago

Correcting for confounders (batch correction)





Technical/experimental sources of batch difference

- Sequencing related differences
  - Sequencing depth
  - Sequencing saturation
  - Sequencing device (e.g. Nextseq vs HiSeq)

Technical/Experimental sources of batch difference

- Poor choice of experimental design
  - E.g. batch tracks biology
- Differences in capture technology
  - SmartSeq2 vs 10x

Biological confounders

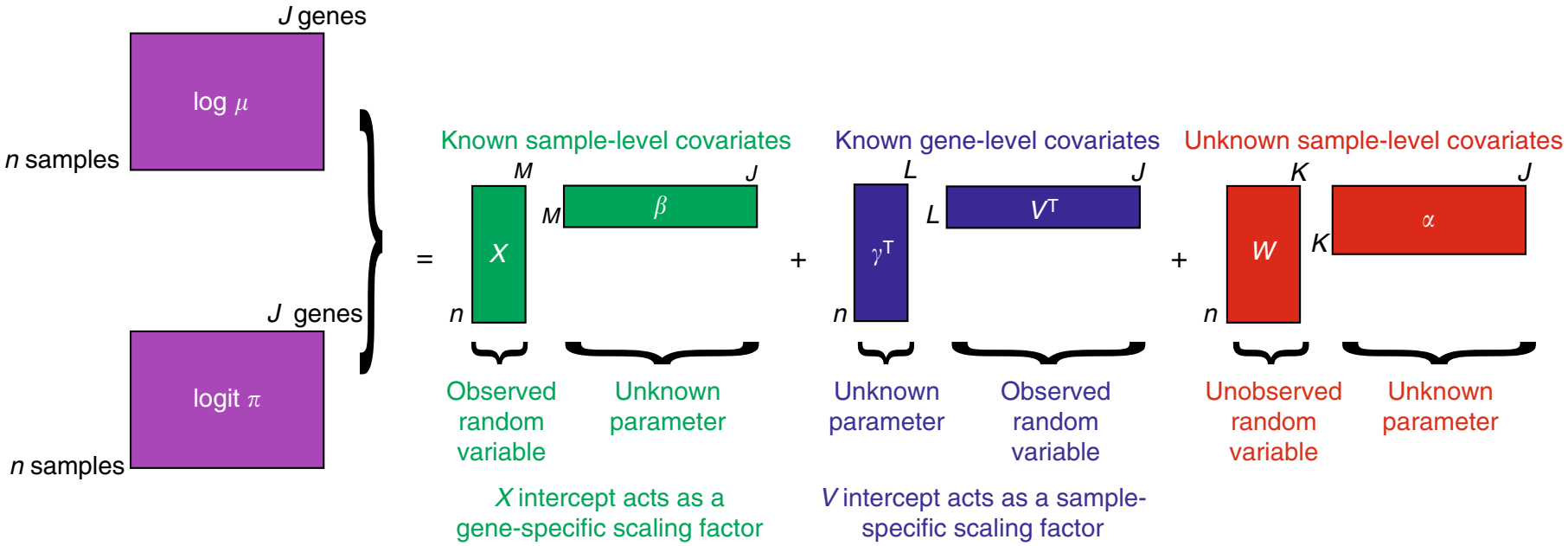
- Sample degradation
- Cell cycle
- Mitochondrial expression (or DNA content)
- Ribosomal fraction
- Copy number differences

Classic approaches to correction

- Down sampling
- Normalization
  (scaling, centering, variance, log)
- Removing genes associated with batch
  - Removing PCs associated with batch
- ComBat
- Regression of residuals (linear or glm)
  - Experimental factor
  - Number of genes
  - Total UMI
  - % mitochondrial

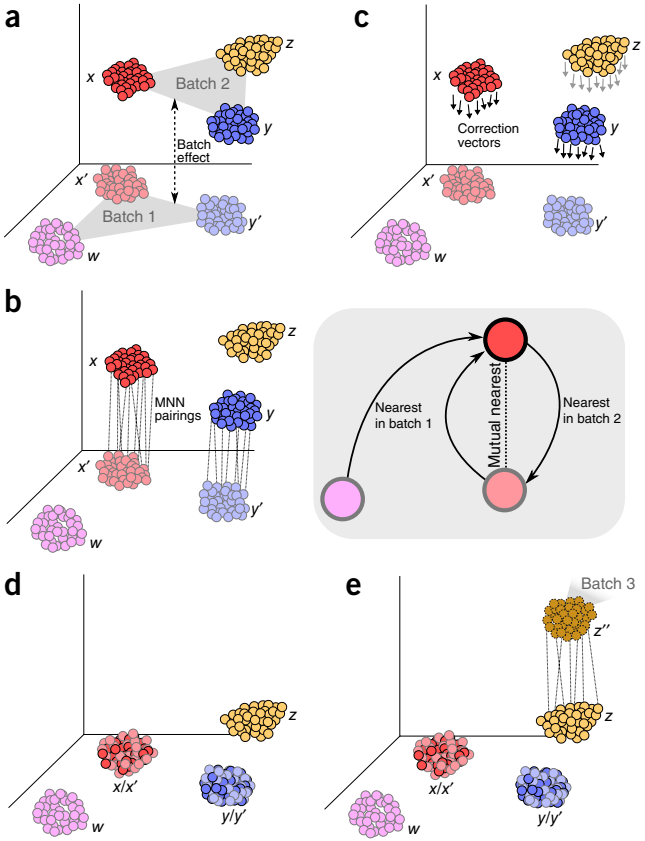
Modeling based approaches

- ZINB-wave



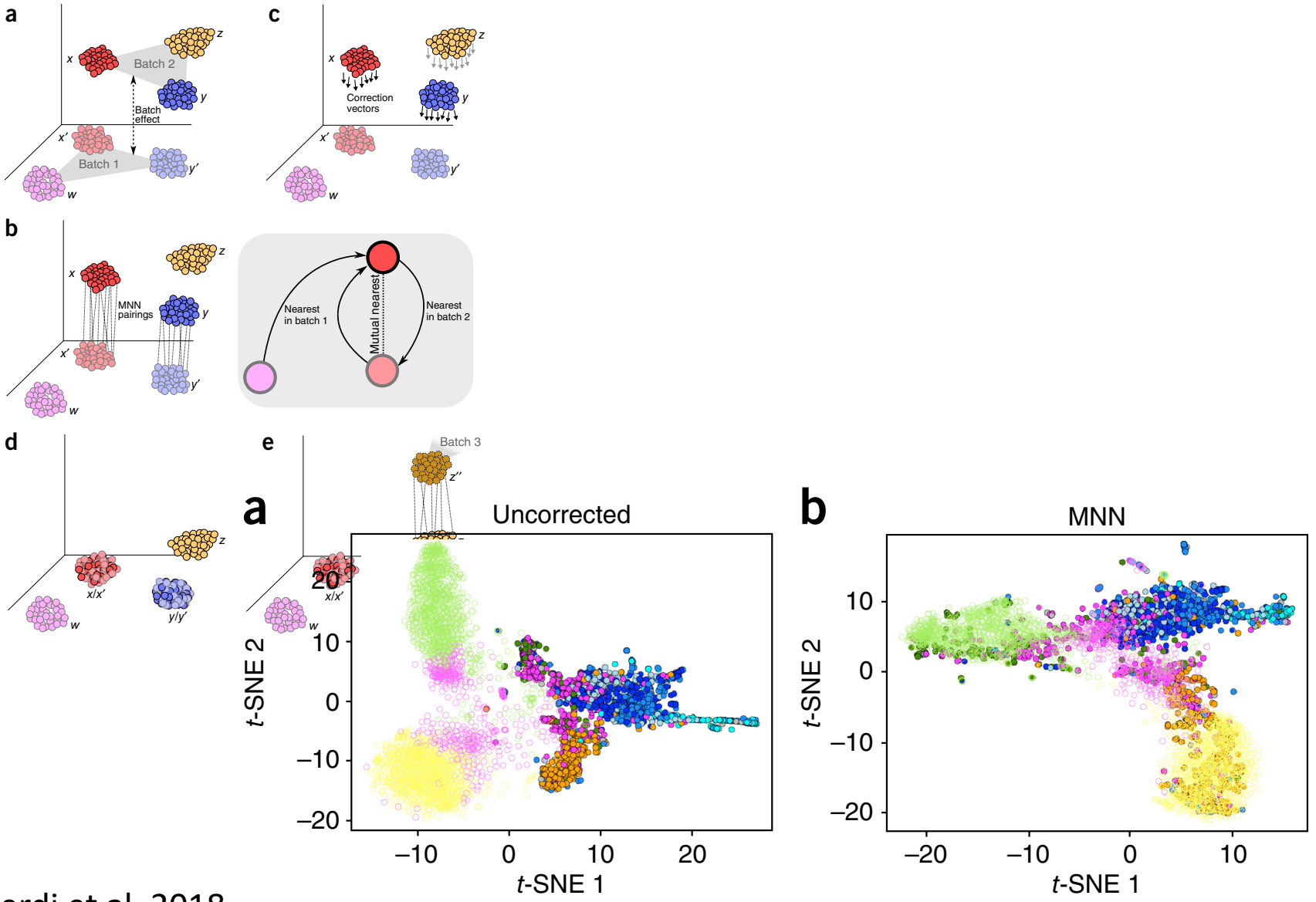
Risso et al. 2017

NN graph correction – mutually nearest neighbors



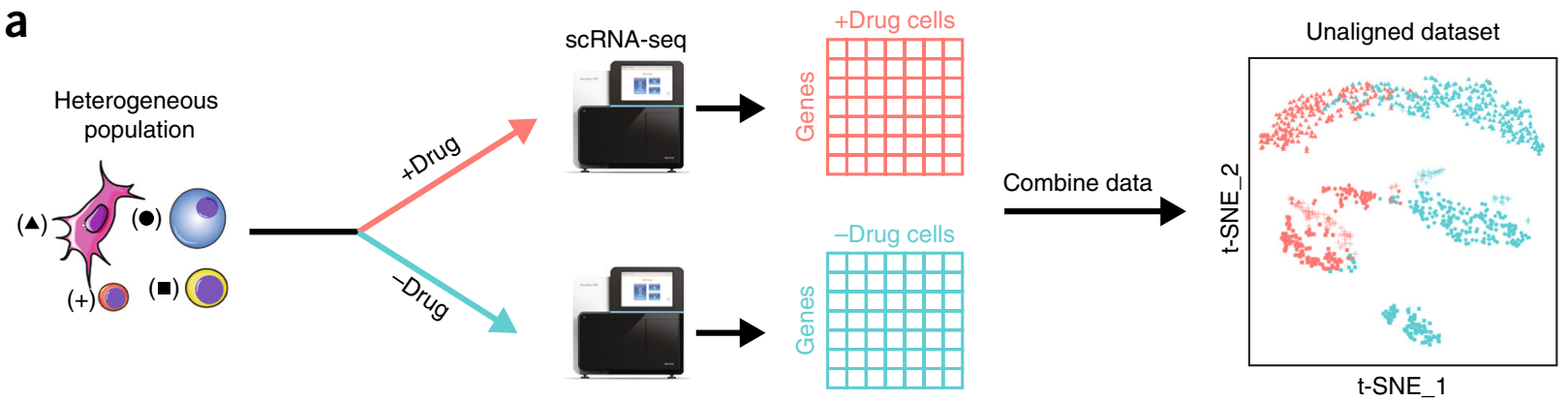
Haghverdi et al. 2018

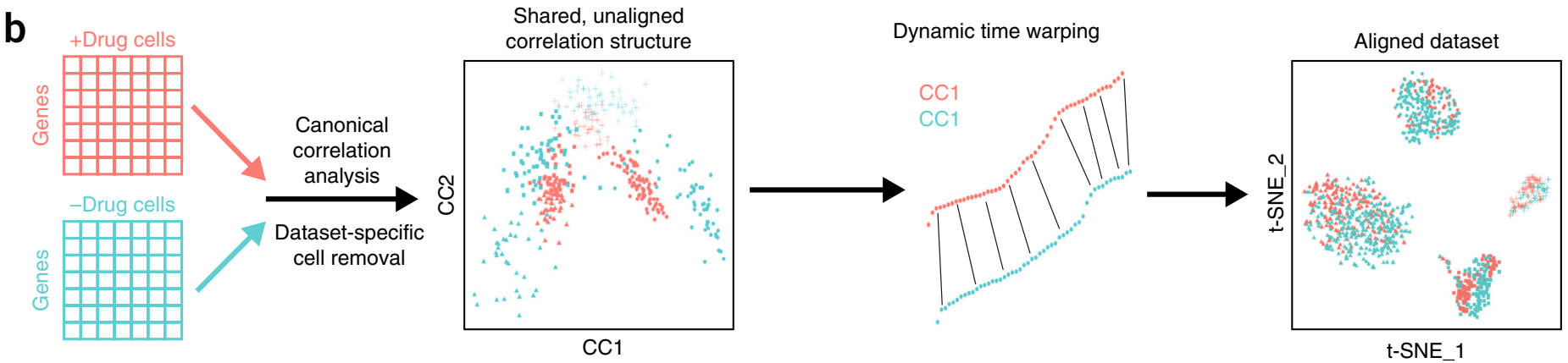
Mutually nearest neighbors



Haghverdi et al. 2018

Canonical correlation analysis



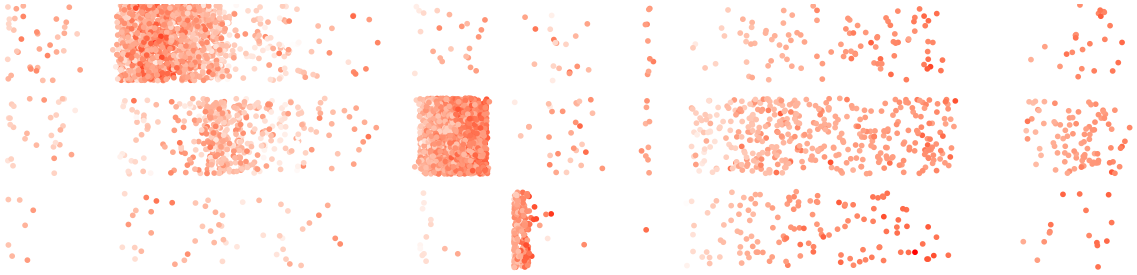


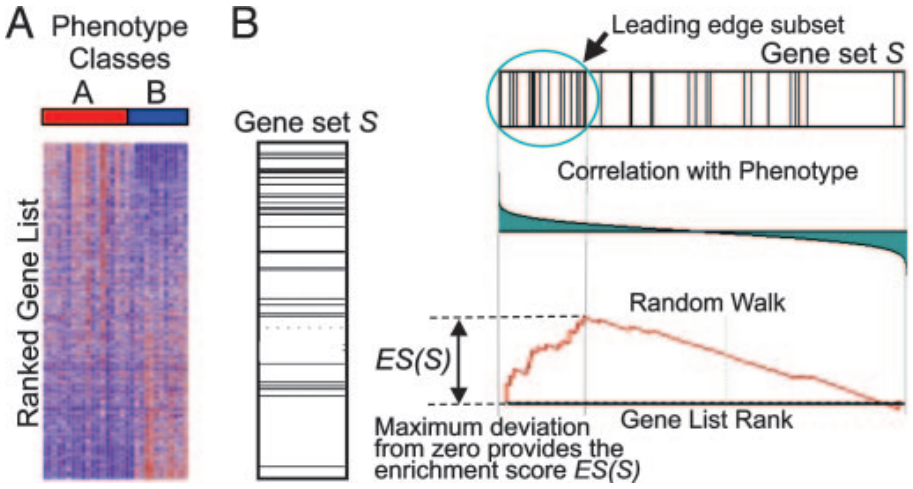
Butler et al. 2018

Geneset and pathway enrichment

Gene set enrichment analysis and scRNAseq

Sparse gene x cell matrix





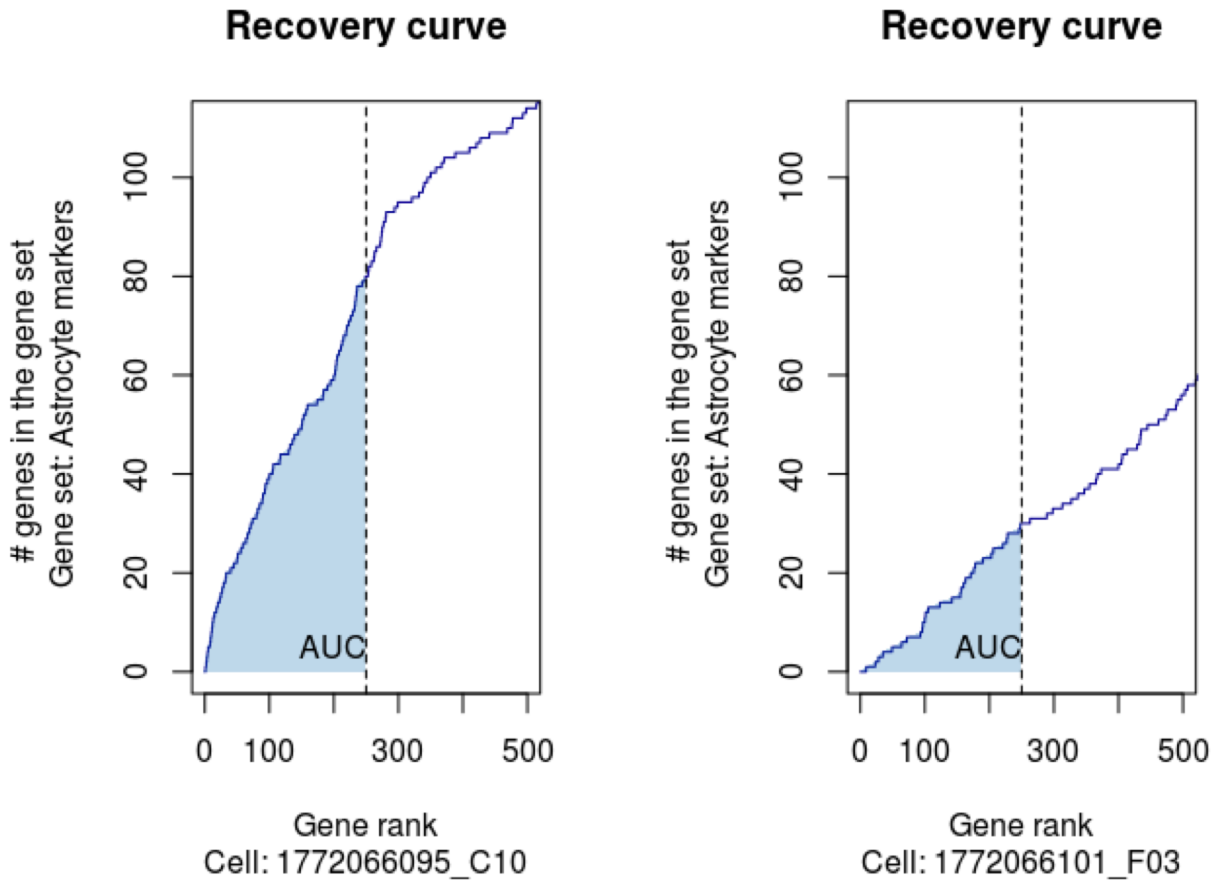
Useful sources for gene set information

- GO/KEGG/BioCarta (old school cool)
- MsigDB (Broad)
- Enricher (Ma’ayan lab, Mount Sinai)
- GSKB – multi species (Ge lab, South Dakota State)
- G:profiler (Vilo lab, University of Tartu, Estonia)

Additional bulk RNA enrichment tools

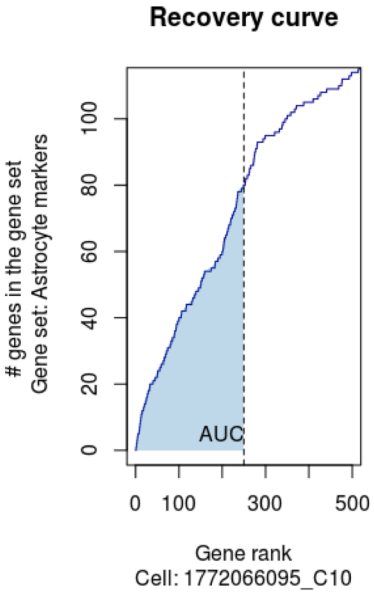
- GSA (R)
- GAGE (R)
- fgsea (R)
- HOMER (Perl)

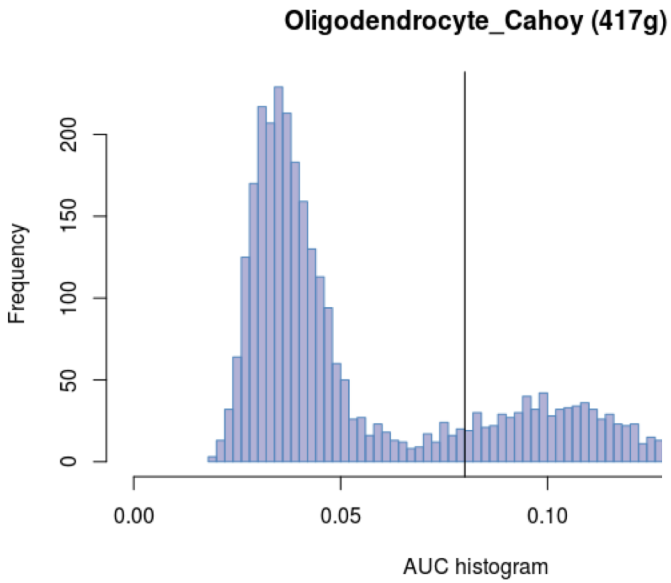
Single cell enrichment score - AUCell

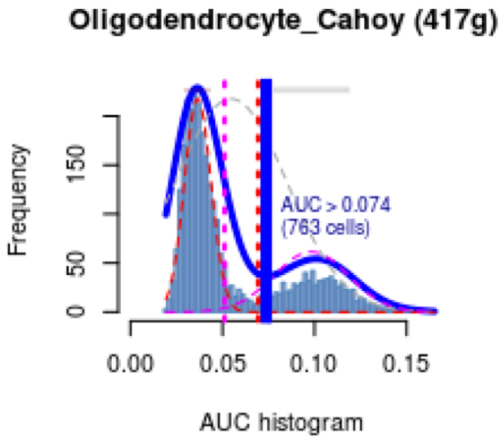


Aibar et al, 2017

Single cell enrichment score - AUCell

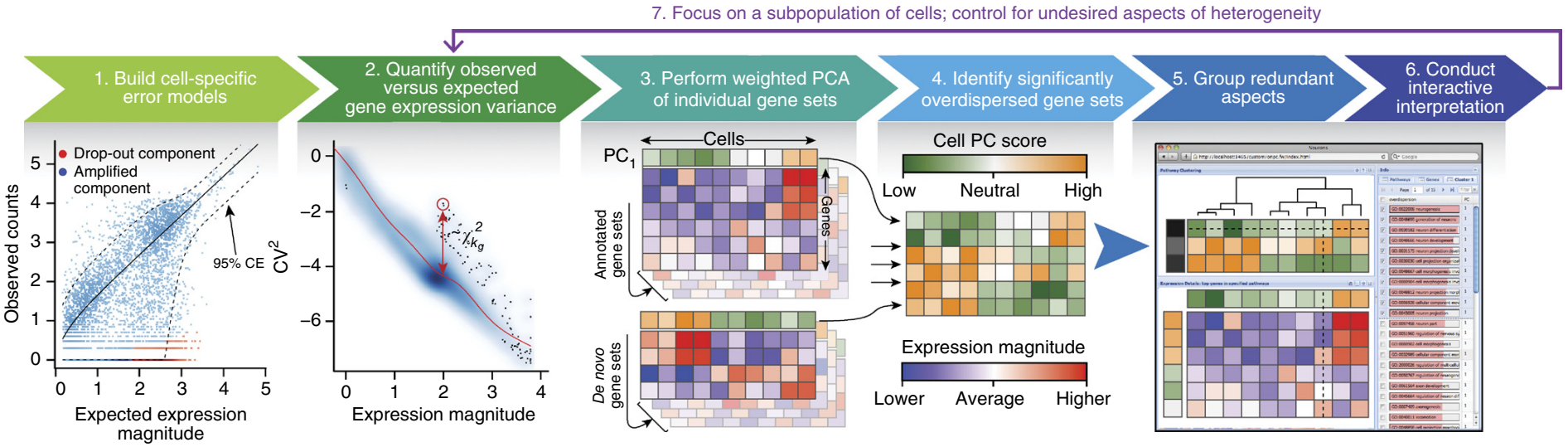






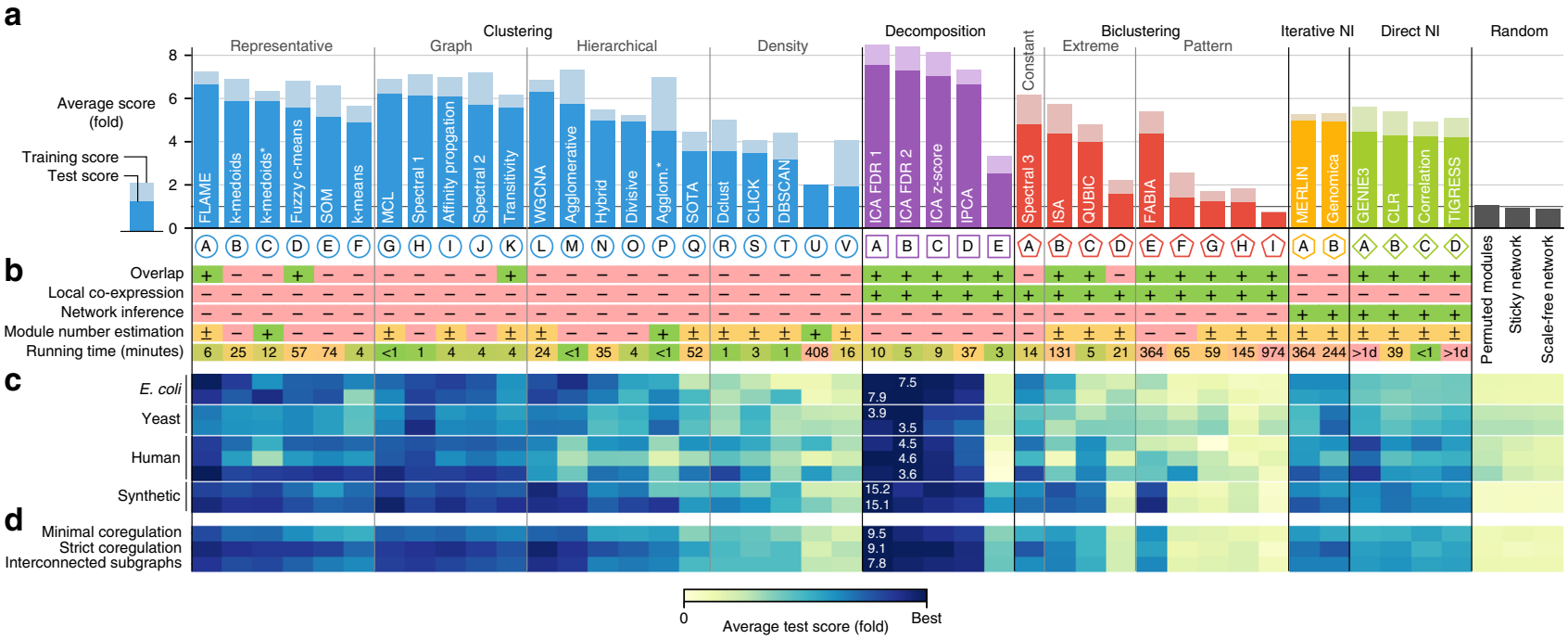
Aibar et al, 2017

Enriched variance within a gene-set - Pagoda



Fan et al, 2017

Survey of module discovery methods



COMMUNICATIONS

ARTICLE

DOI: 10.1038/s41467-018-03424-4 OPEN

A comprehensive evaluation of module detection methods for gene expression data

Wouter Saelens[1,2], Robrecht Cannoodt[1,3] & Yvan Saeys[1,2]

Trajectory/pseudo-time analysis

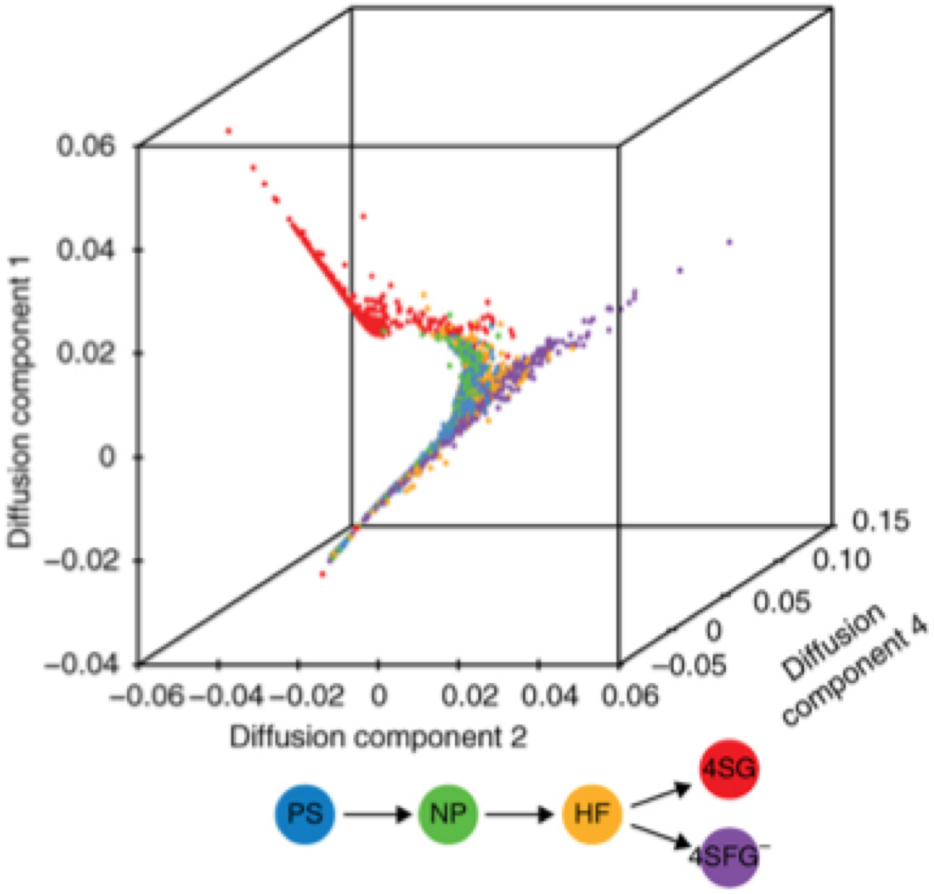
Visualizing Continuous Variation using Diffusion Maps

For many systems we can not assume clear states.

Assumes that there exists a “continuous manifold” of states that cells are diffusing through (e.g. during differentiation).

Diffusion maps construct a random walk (Markov process) on the data, and infer a low dimensional representation (DCs) that describes these paths.

Of the analysis types, trajectory inference is modeling data in a way almost impossible using bulk



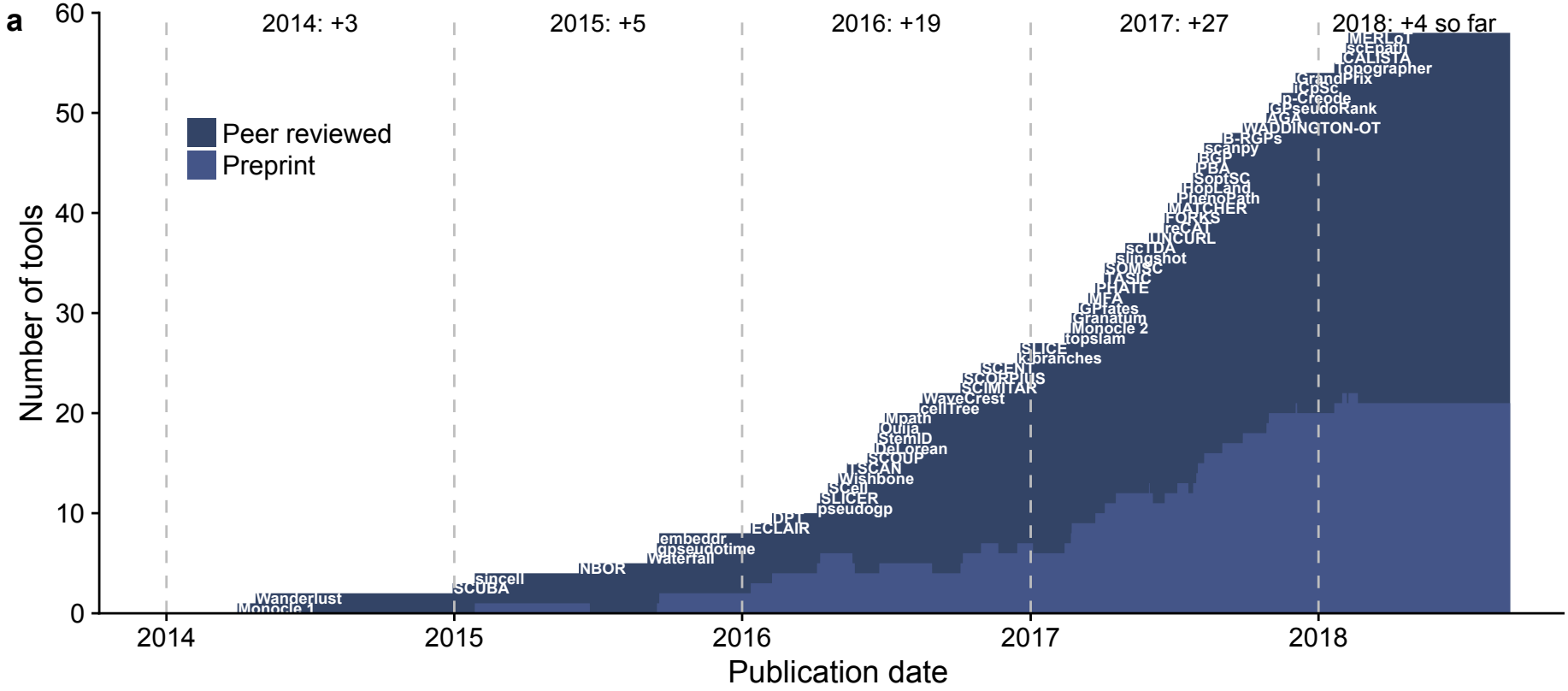
The pseudotime/trajectory estimation problem

Given: single-cell expression measurements for a heterogeneous collection of cells that is transitioning among states.

Return:

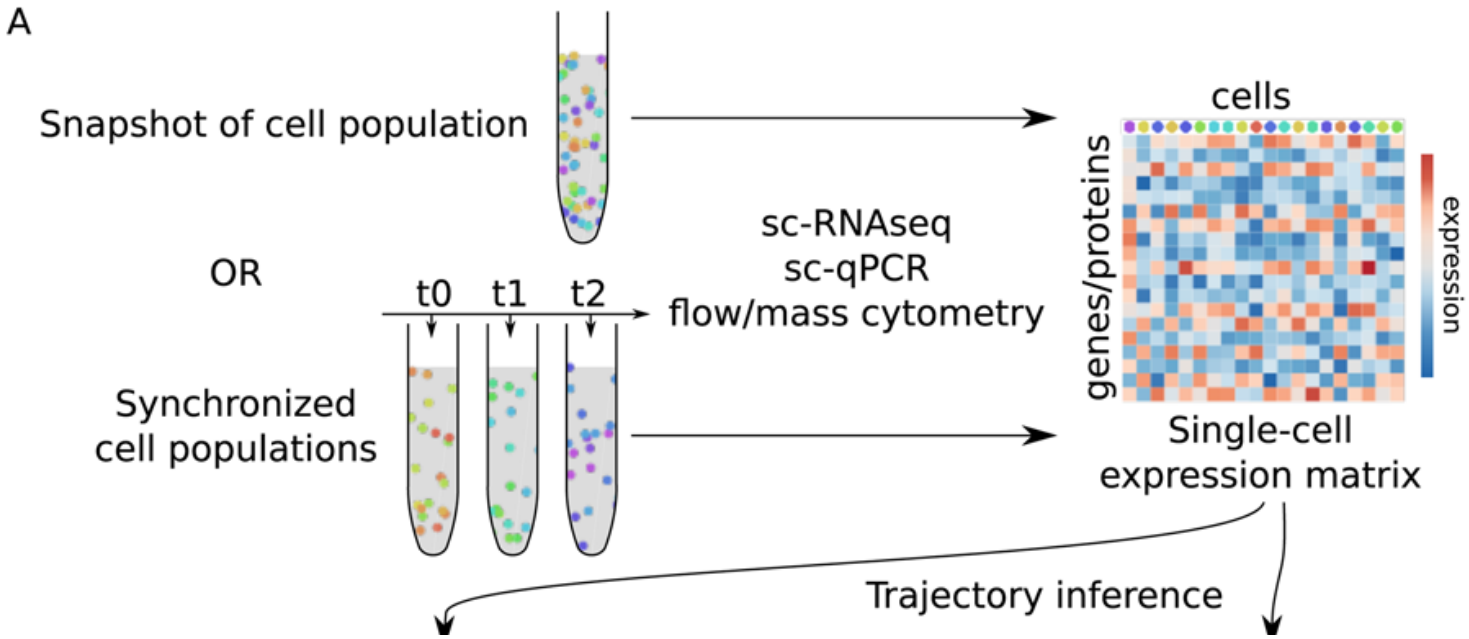
1. A quantitative value for each cell that represents its progress in the transition.
2. A path describing the likely transition between biological states.

The trajectory inference gold rush



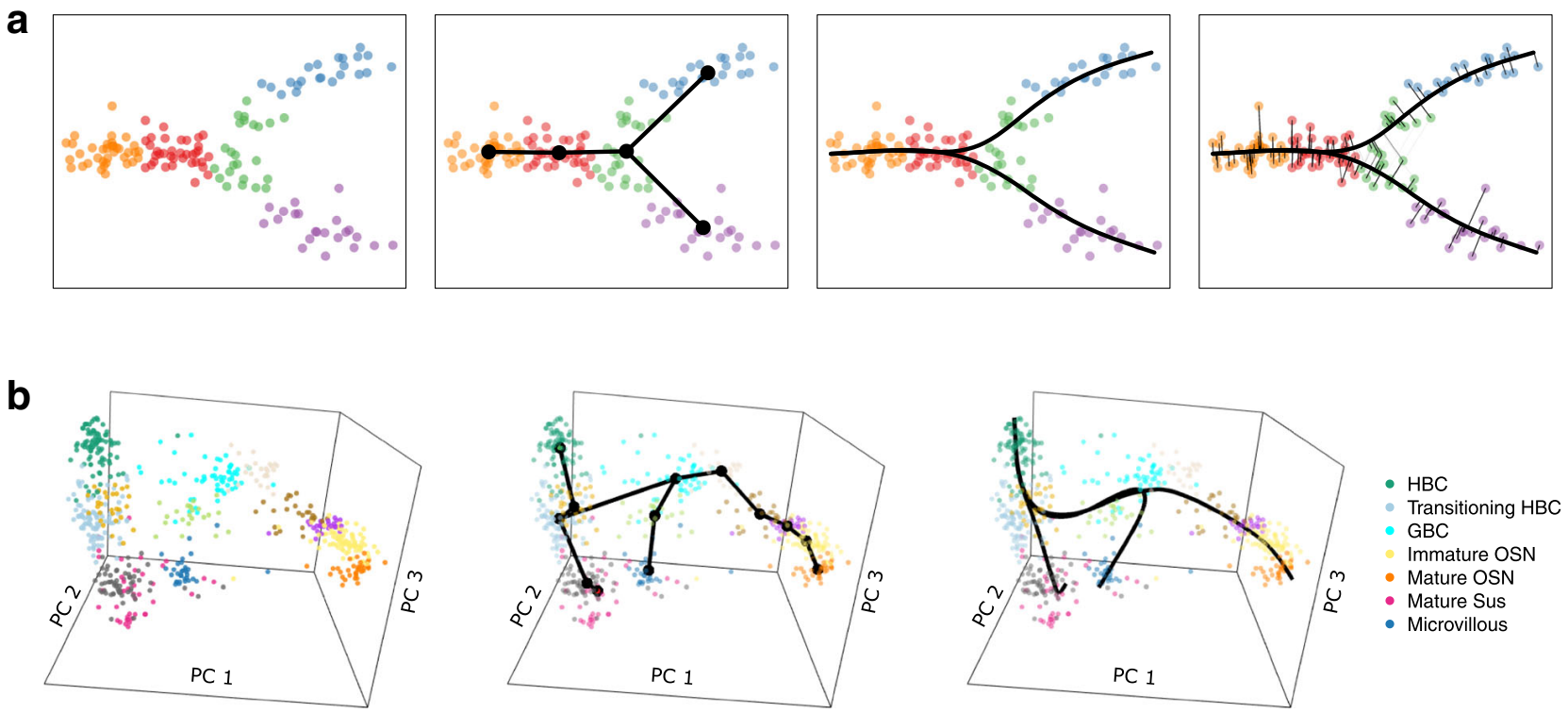
Saelens et al. 2018

Inferring expression dynamics from a snapshot



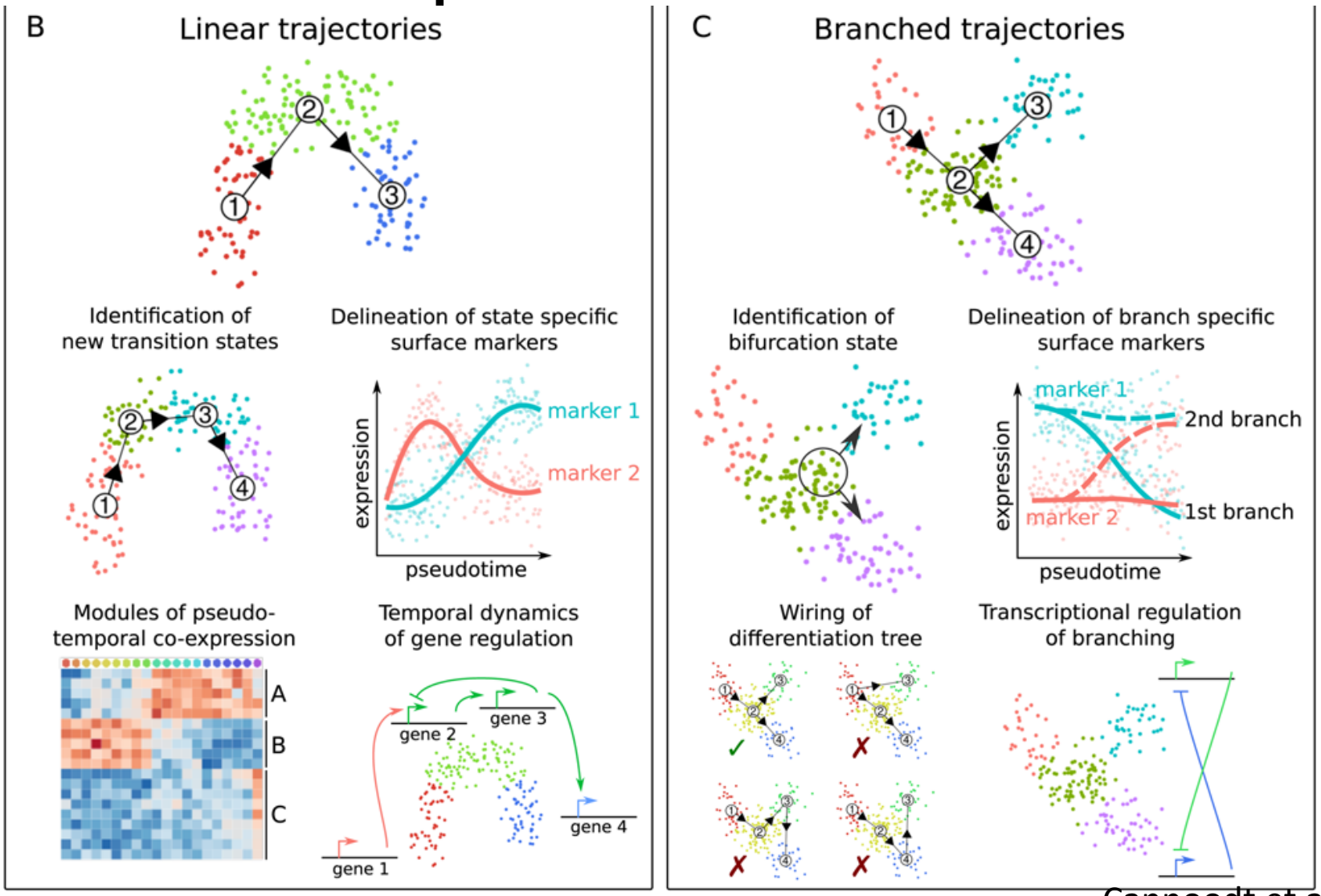
Cannoodt et al. 2016

Inferring expression dynamics from a snapshot



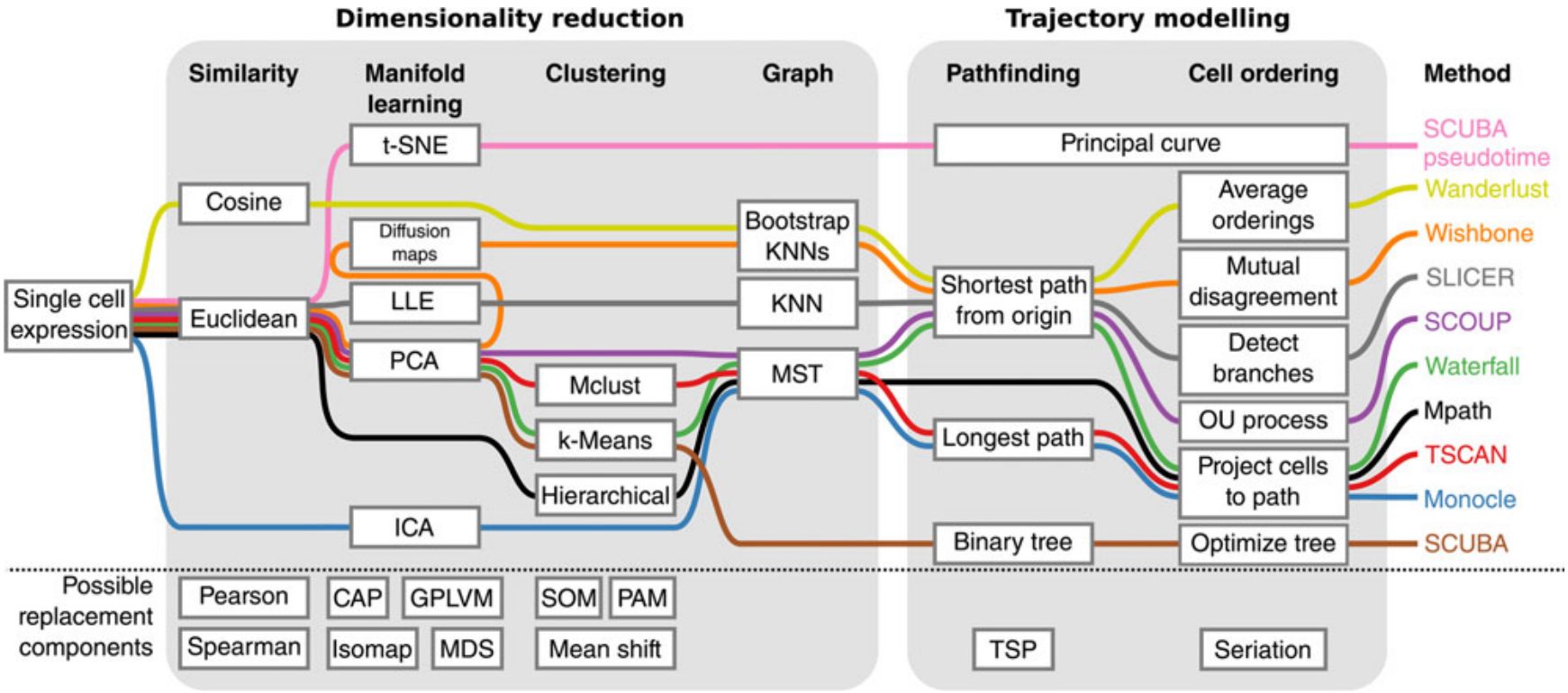
Street et al. 2018

Inferring expression dynamics from a snapshot



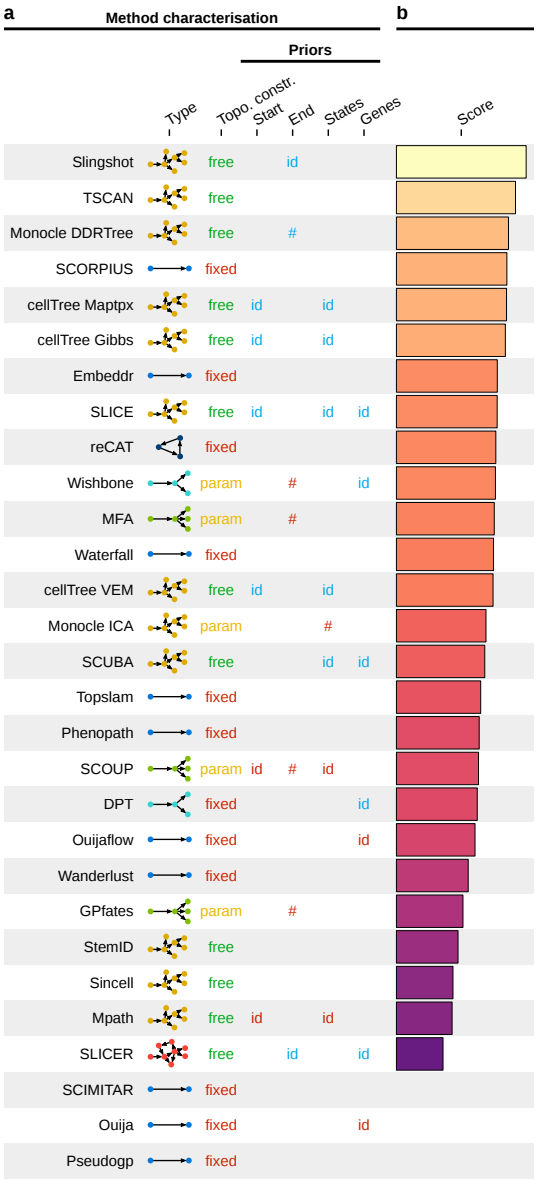
Cannoodt et al. 2016

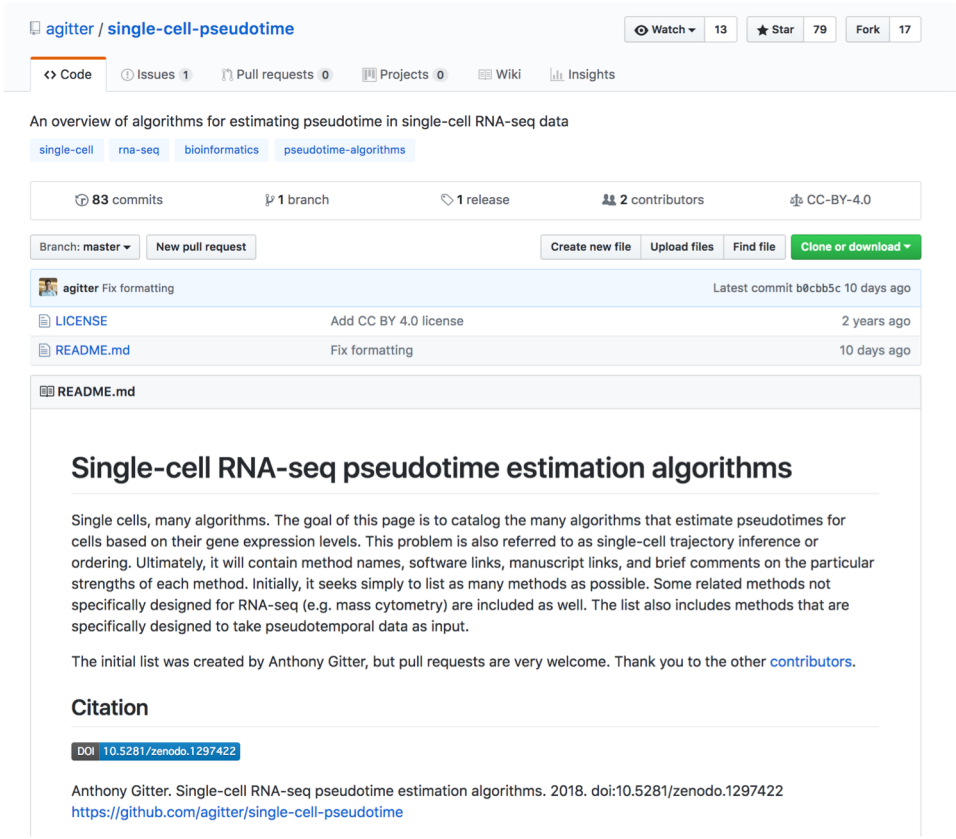
An honest attempt to chart the landscape (circa 2016)



Cannoodt et al. 2016

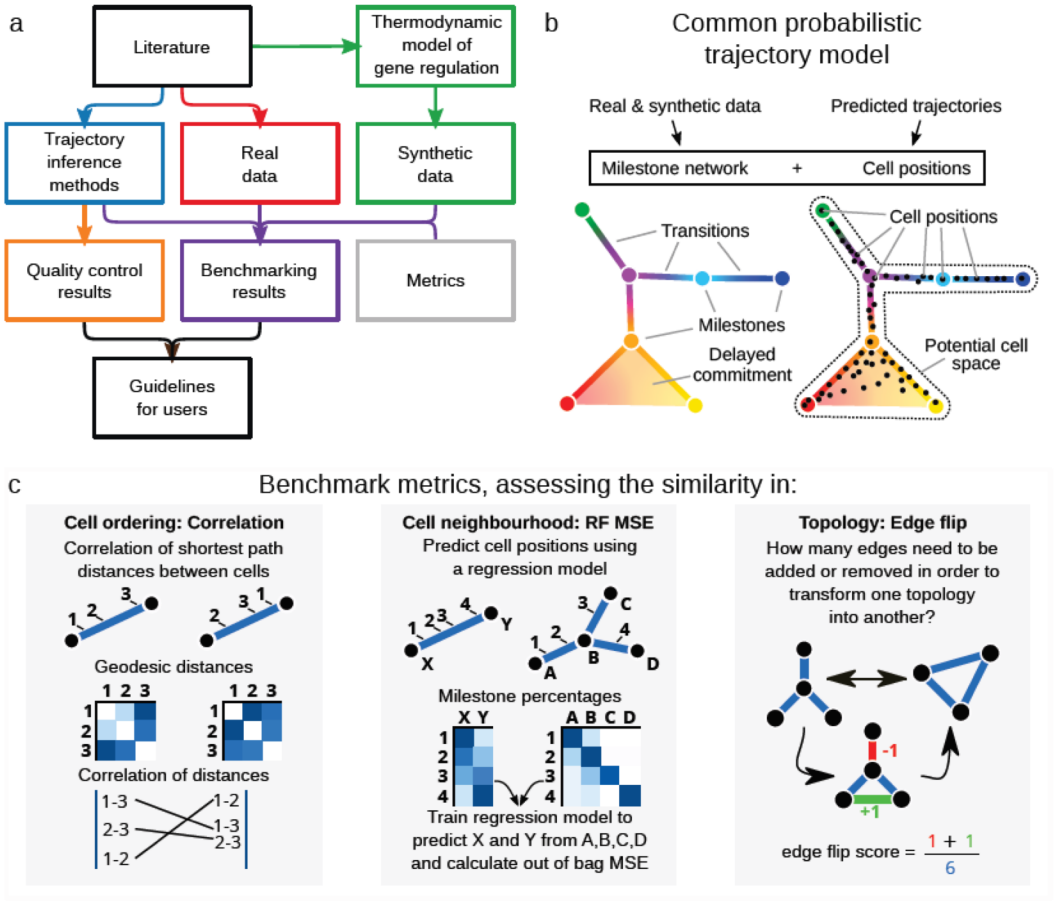
In 2018 …





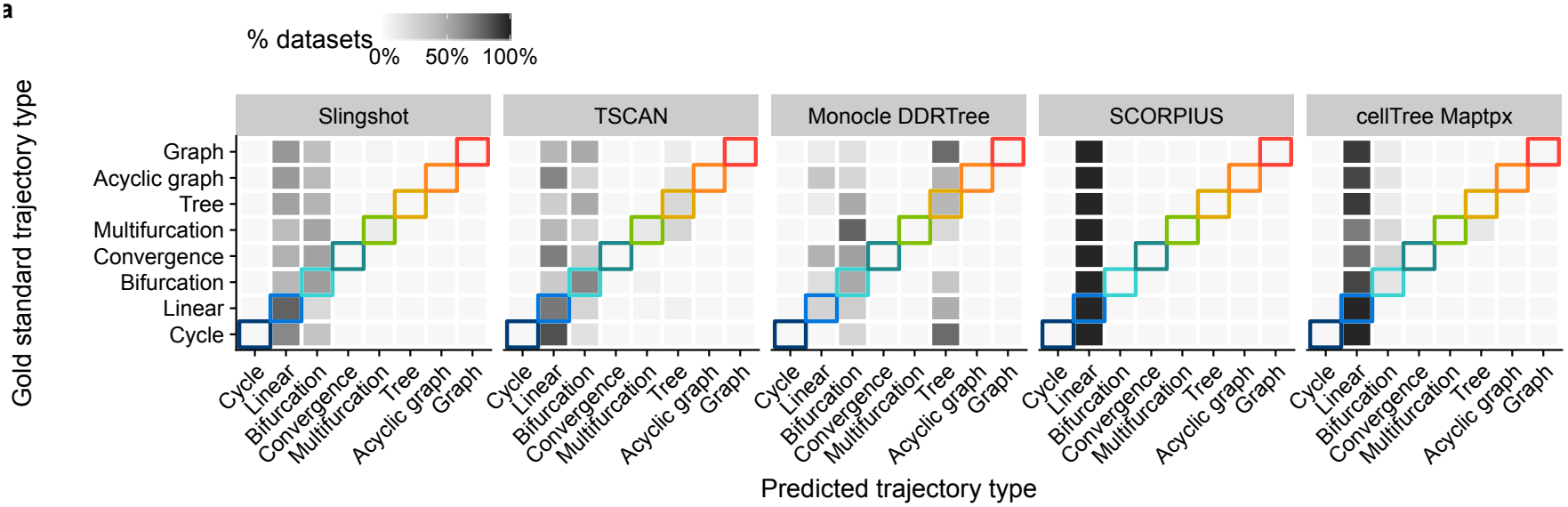
Saelens et al. 2018

TI method evaluation



Saelens et al. 2018

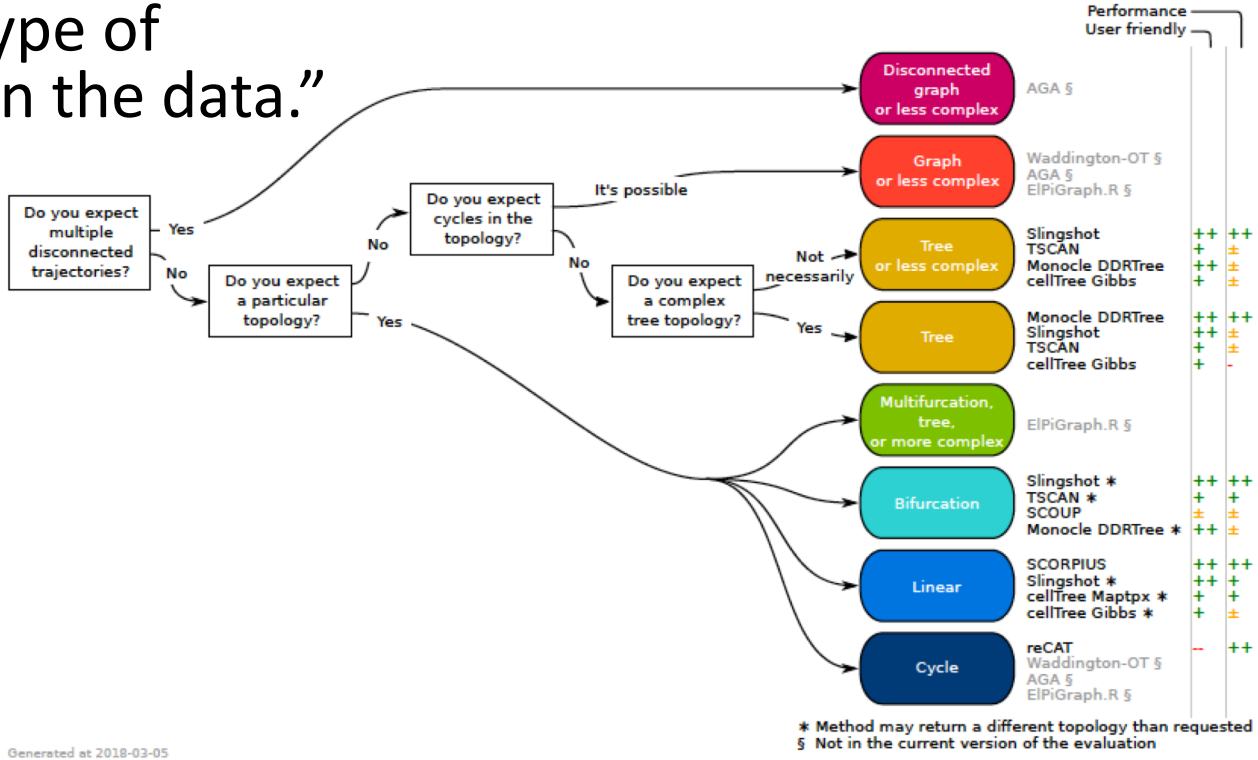
Not quite a solved problem?



Saelens et al. 2018

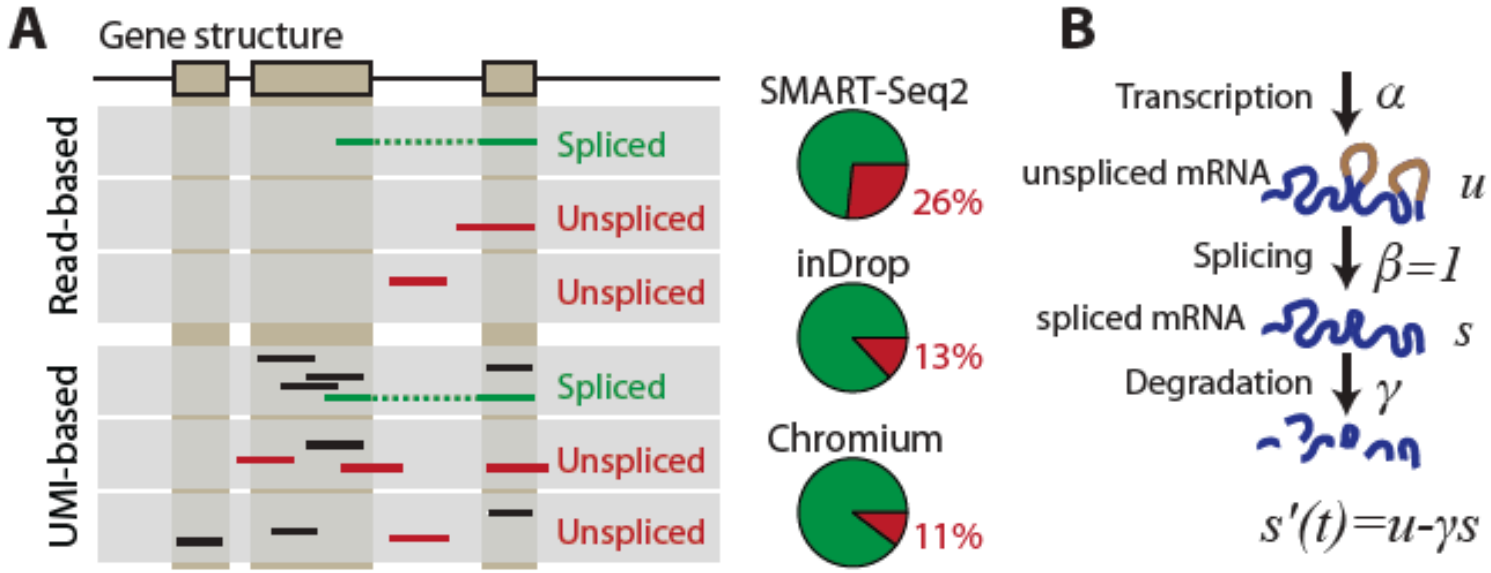
Decision Tree for Method Use

"We found that some methods, including Slingshot,TSCAN and Monocle DDRTree, clearly outperform other methods, although their performance depended on the type of trajectory present in the data."



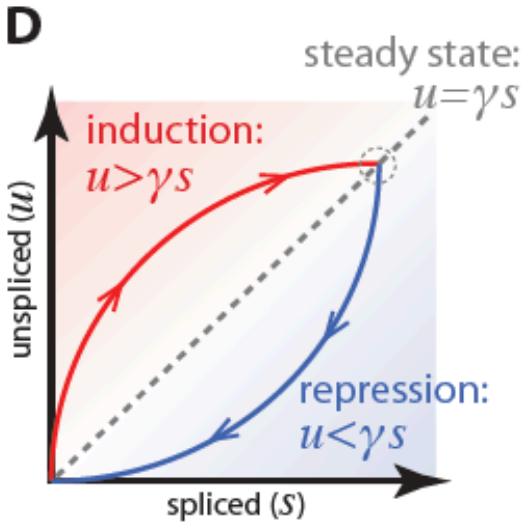
Saelens et al. 2018

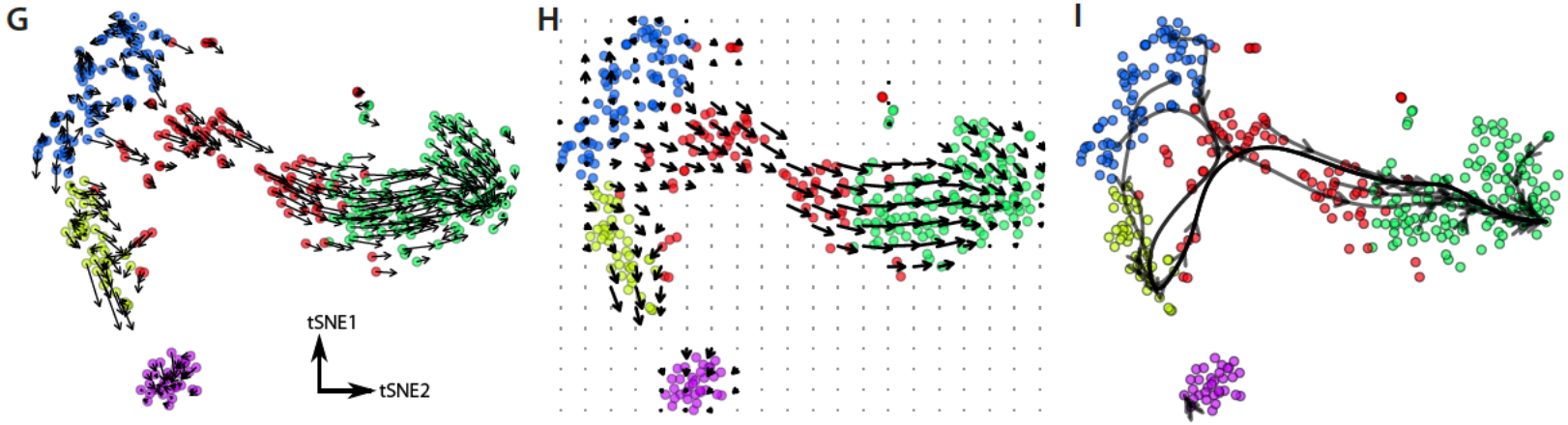
RNA velocity - Inferring dynamics from RNA degradation



Manno et al, preprint 2018

Inferring dynamics -

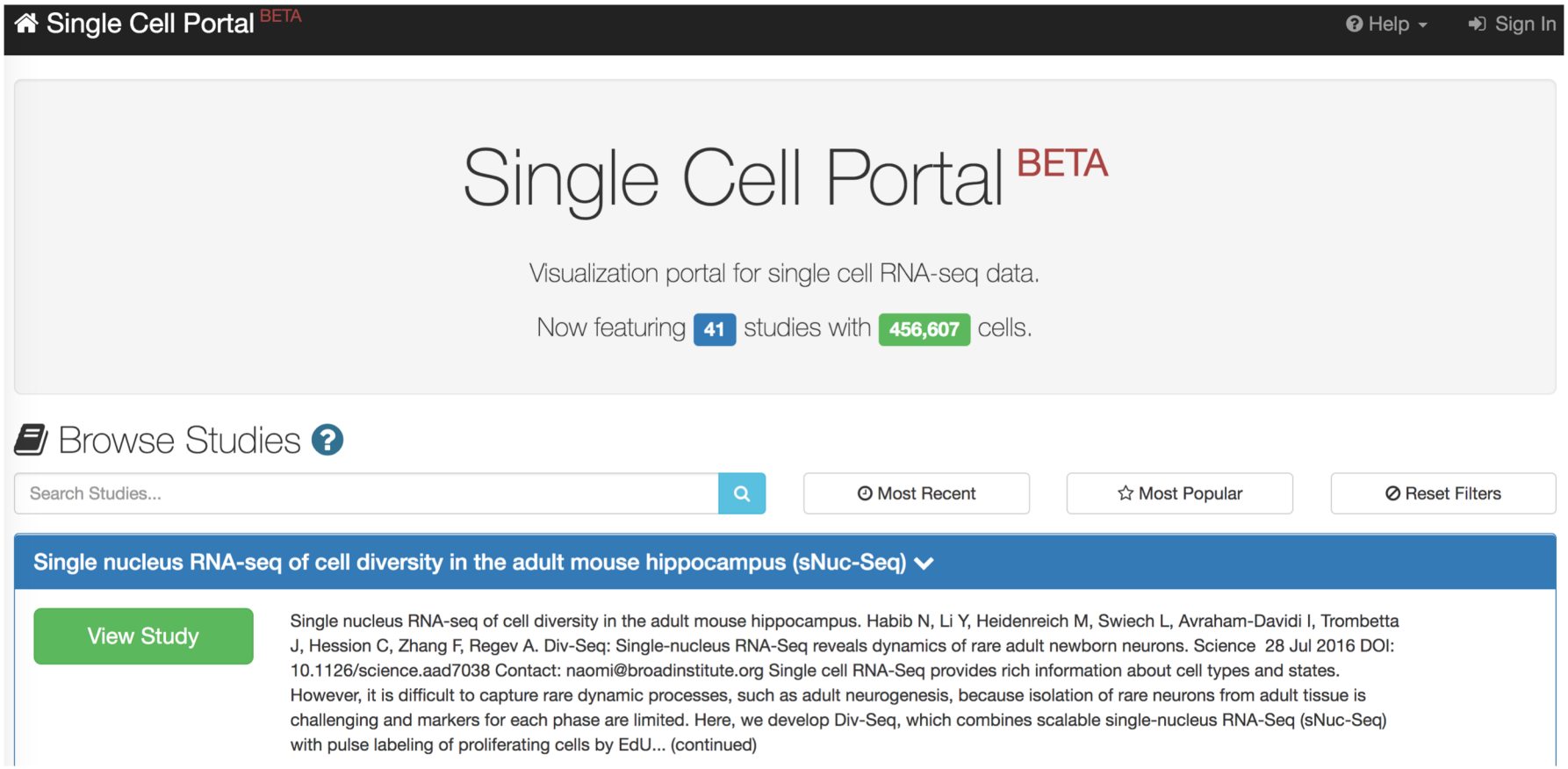




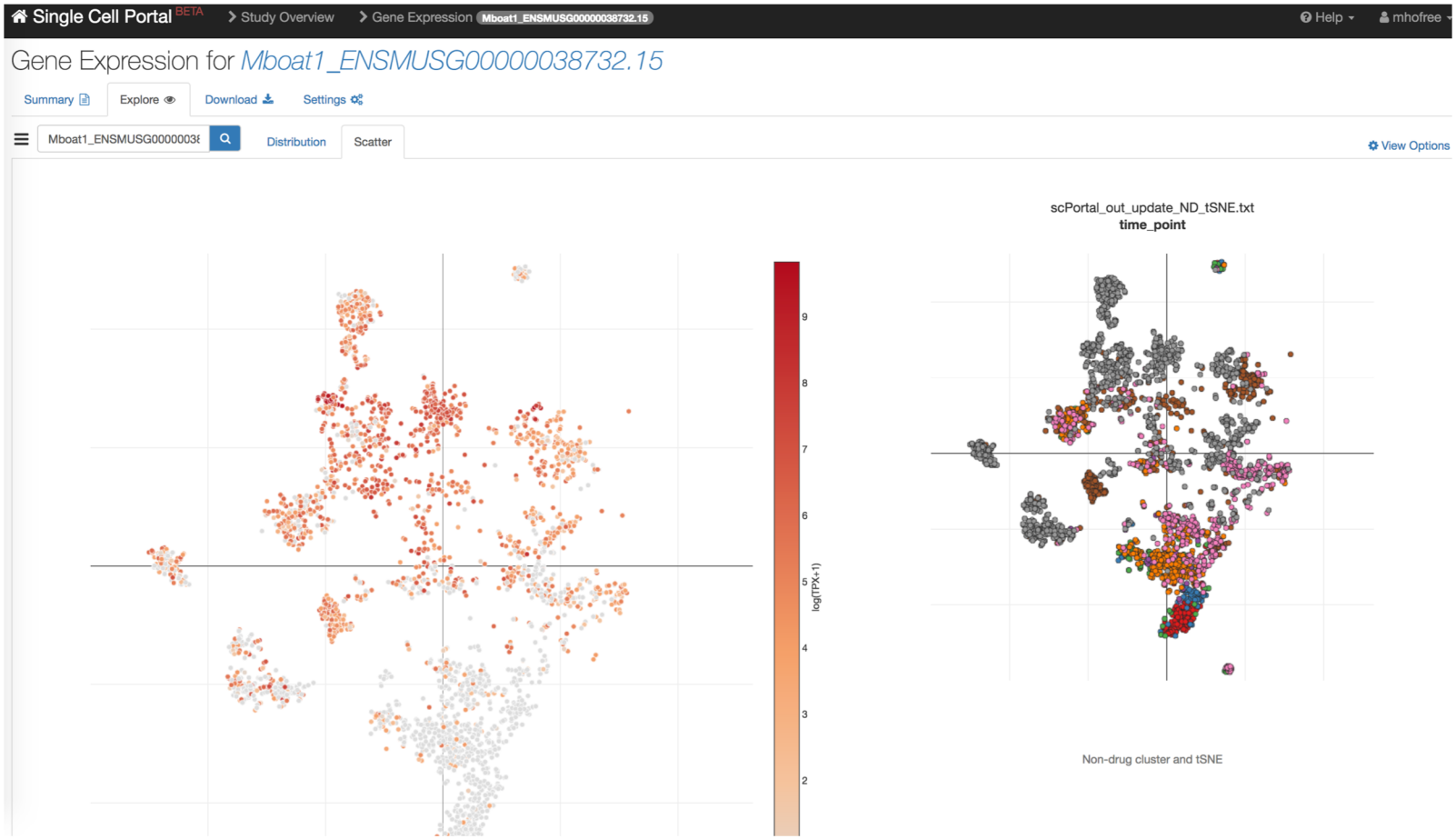
Manno et al, preprint 2018

Interactive tools

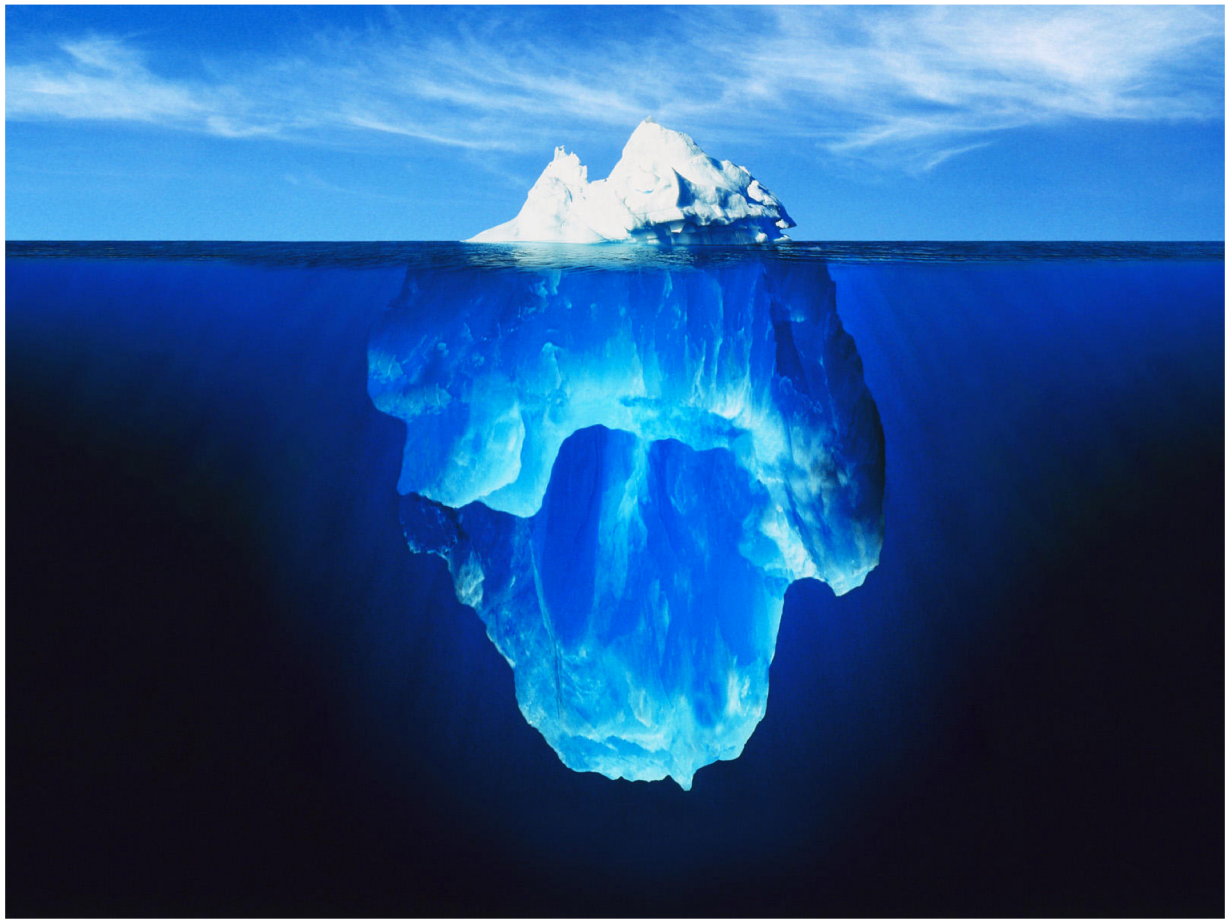
Single Cell Portal



Single cell portal - interactive



What did we cover?



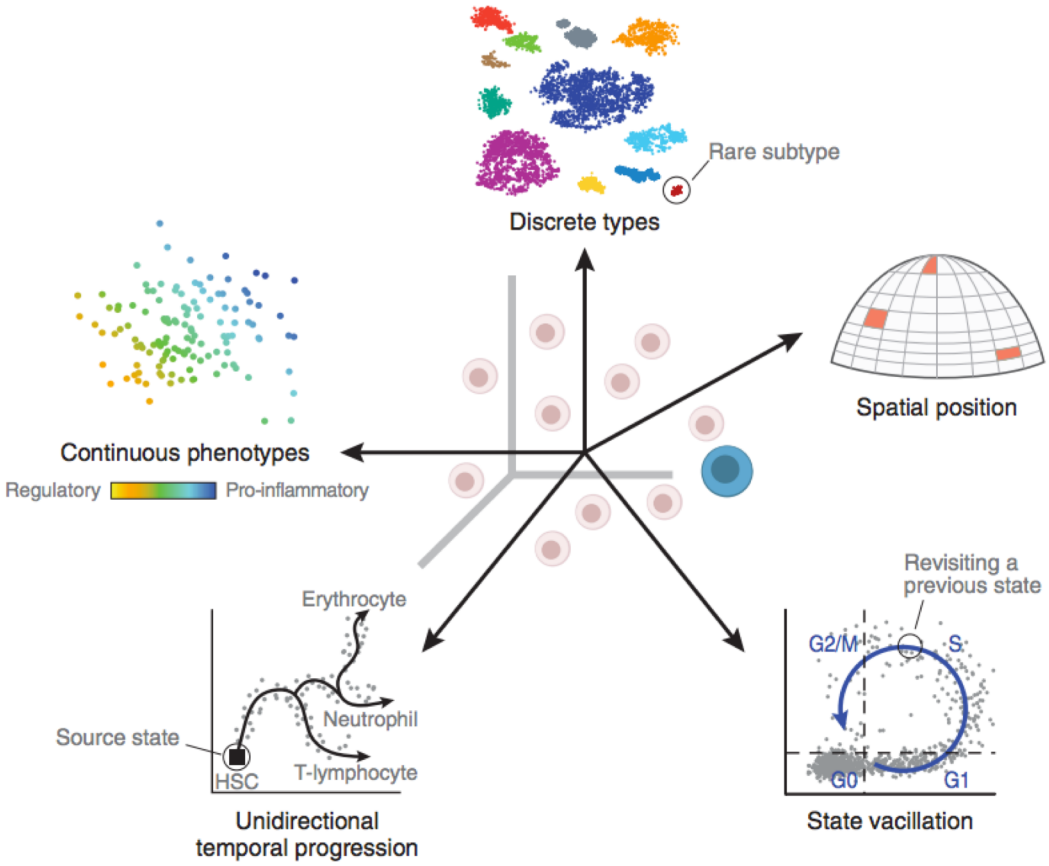
What did we cover…

REVIEW

nature biotechnology

Revealing the vectors of cellular identity with single-cell genomics

Allon Wagner[1], Aviv Regev[2,3,5] & Nir Yosef[1,4,5]



Acknowledgements

Special thanks:

Brian Haas

Timothy Tickle

Karthick Shekar

Aviv Regev

BROAD INSTITUTE

ITCR
Informatics Technology for Cancer Research

NIH NATIONAL CANCER INSTITUTE

KLARMAN CELL OBSERVATORY
AT BROAD INSTITUTE