



# Trajectory inference and other analyses

Åsa Björklund

[asa.bjorklund@scilifelab.se](mailto:asa.bjorklund@scilifelab.se)

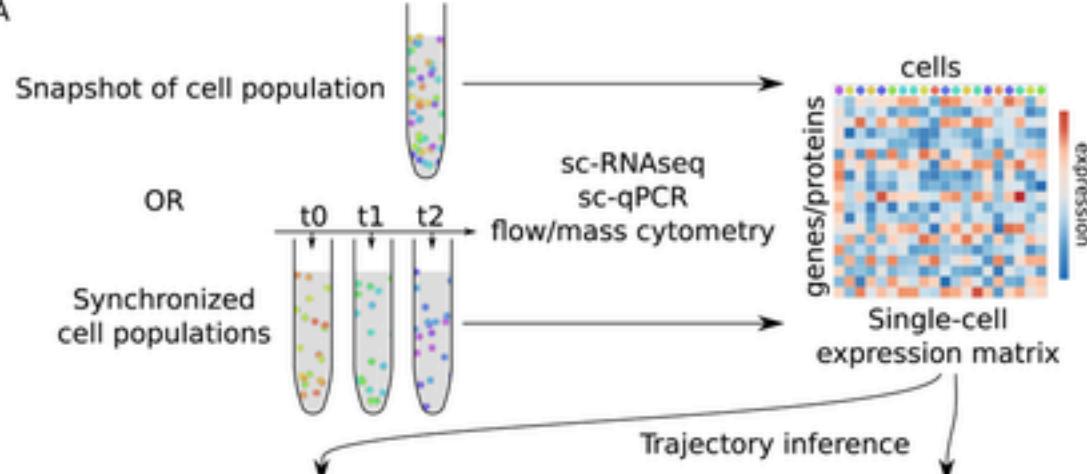
# Outline

- Trajectory analysis
- Other things you can do with scRNASeq data
  - Multi-omics data
  - Other types of analysis

# Trajectory Inference (TI)

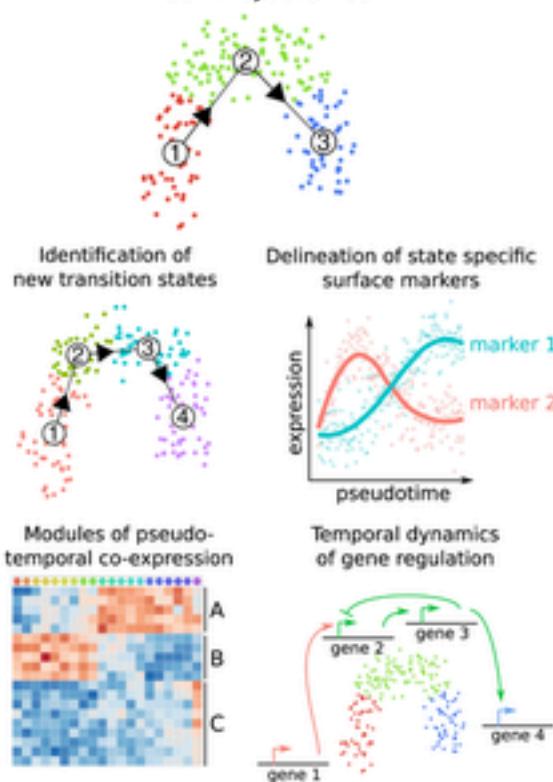
- Cells that differentiate display a **continuous spectrum of states** – transcriptional program for activation and differentiation
- Individual cells will differentiate in an **unsynchronized** manner – each cell is a snapshot of differentiation time
- **Pseudotime** – abstract unit of progress: distance between a cell and the start of the trajectory

A

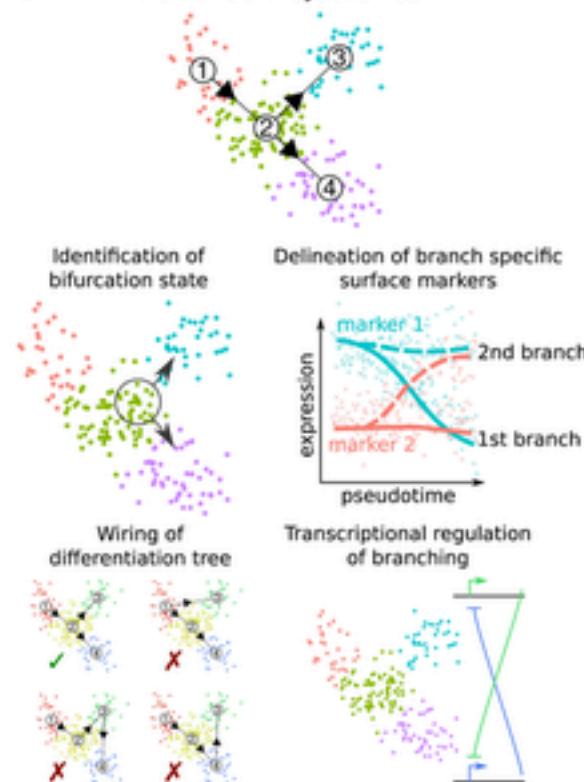


B

## Linear trajectories



## Branched trajectories

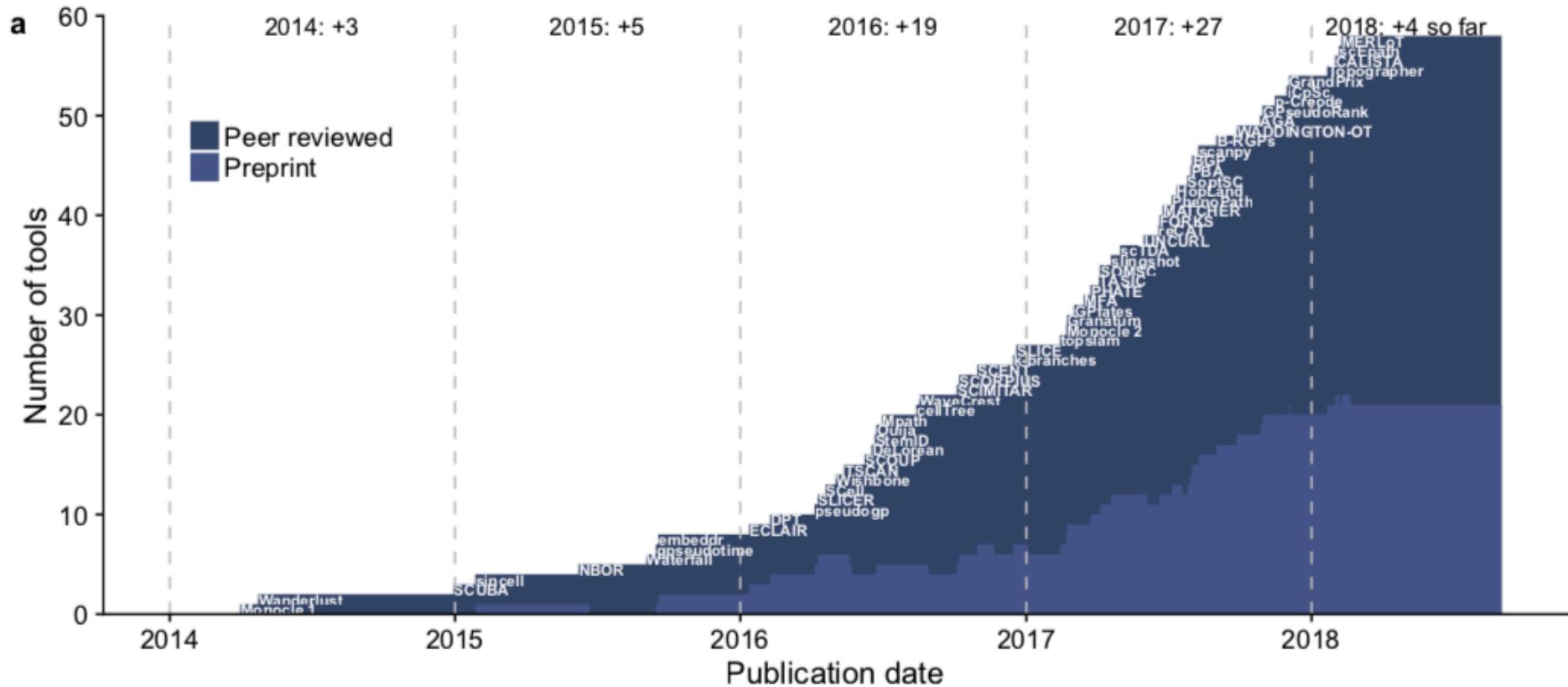


(Cannoodt et al. EJ 2016)

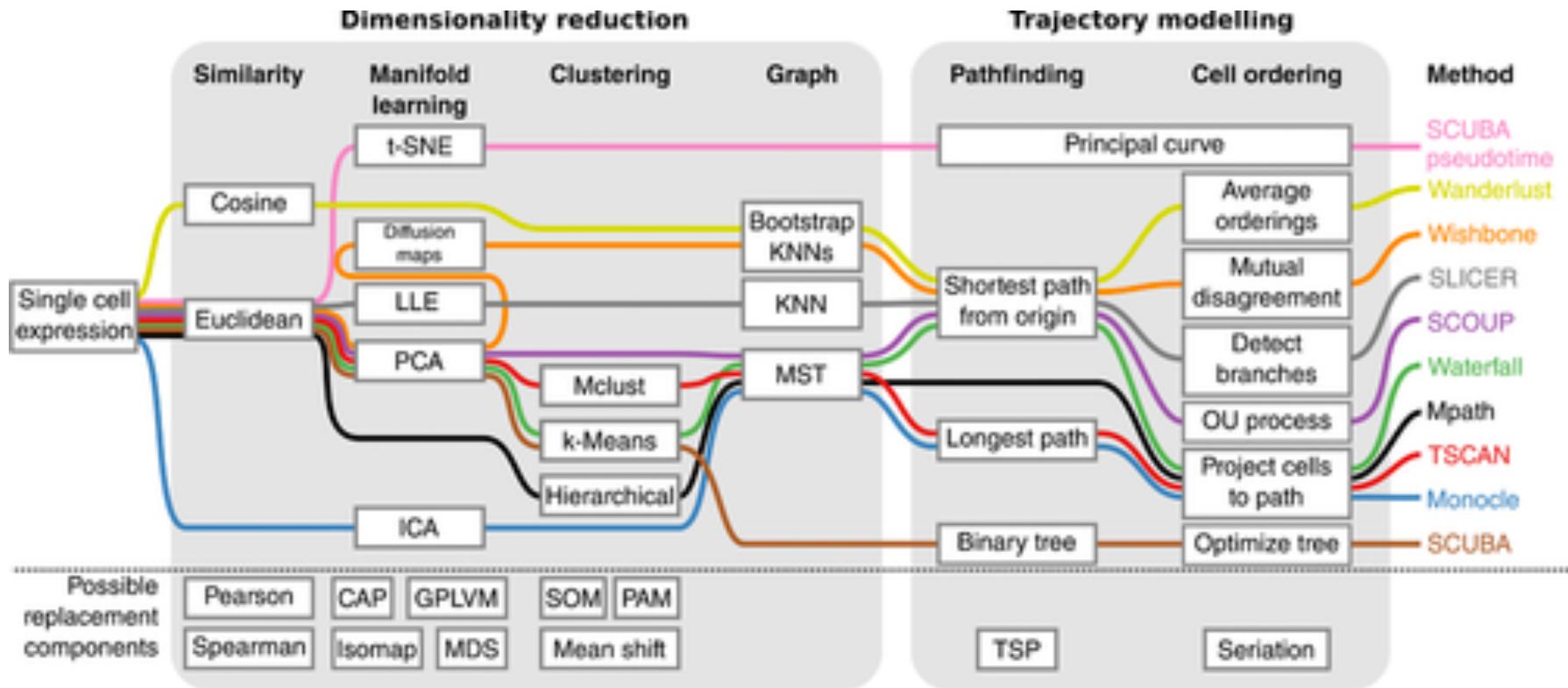
# Should you run Tl?

- Are you sure that you have a developmental trajectory?
- 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.

# TI tools – fast development!



# TI general overview



## TI – main steps

1. Gene set selection
2. Dimensionality reduction
3. Infer trajectories (branched or straight)
4. Order cells
5. Discover interesting gene patterns

# 1. Gene set selection

- Variable genes
- Differentially expressed genes between clusters
- Prior knowledge
- Be careful how you select genes – a more unbiased approach is always better!

## 2. Dimensionality reduction

- Linear: PCA, ICA etc.
- Non-linear: tSNE, Diffusion maps, UMAP
- Graph based

### 3. Infer trajectories

- Many TI methods use graph-based techniques
  - Simplified graph representation as input to find a path through a series of nodes (i.e. individual cells or groups of cells)
  - Different path-finding algorithms are used by different programs
  - Find longest connected path in a sparsified graph
- “starting cell” often defined by the user (e.g. the most immature cell in the case of a cell developmental process)

## 4. Order cells

- Define pseudotime based on cells projection/  
position along trajectory

## 5. Discover gene patterns

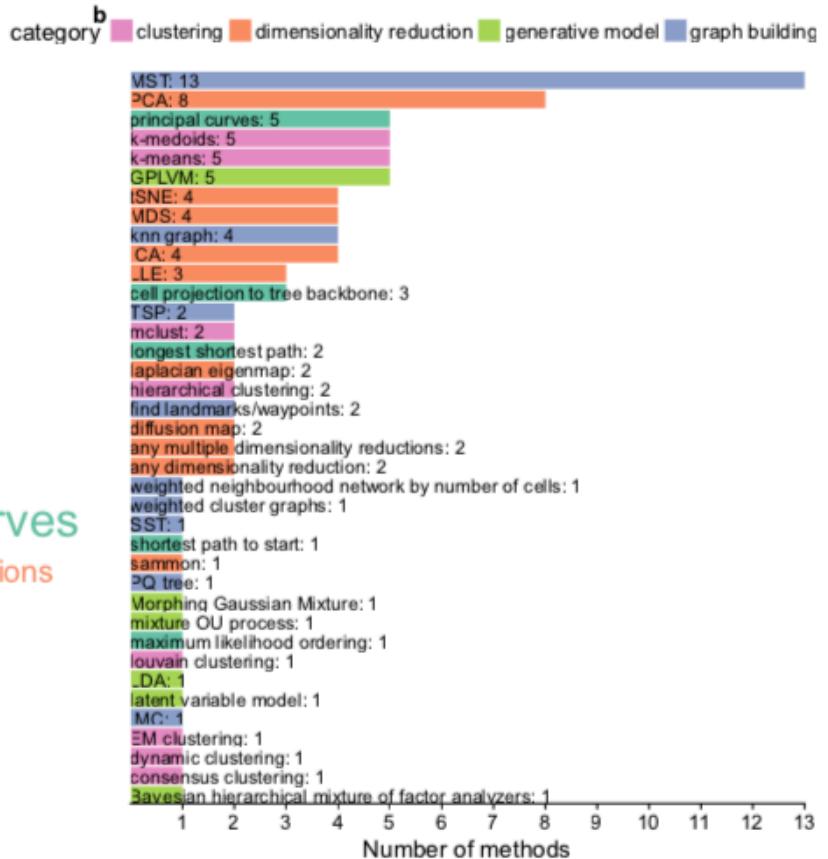
- Statistical tests for regulation along pseudotime
- Branch point analyses

# Methods overview

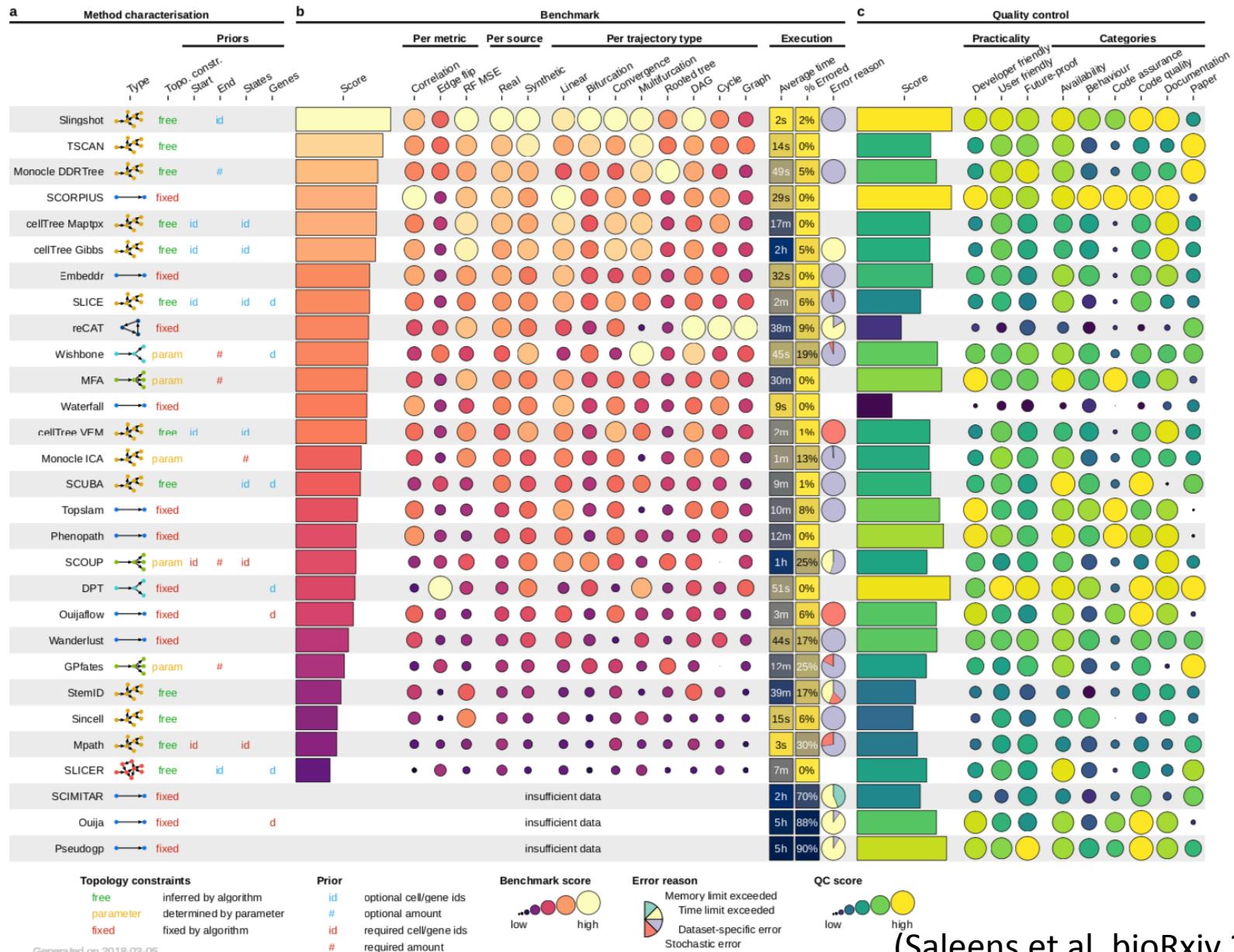
a



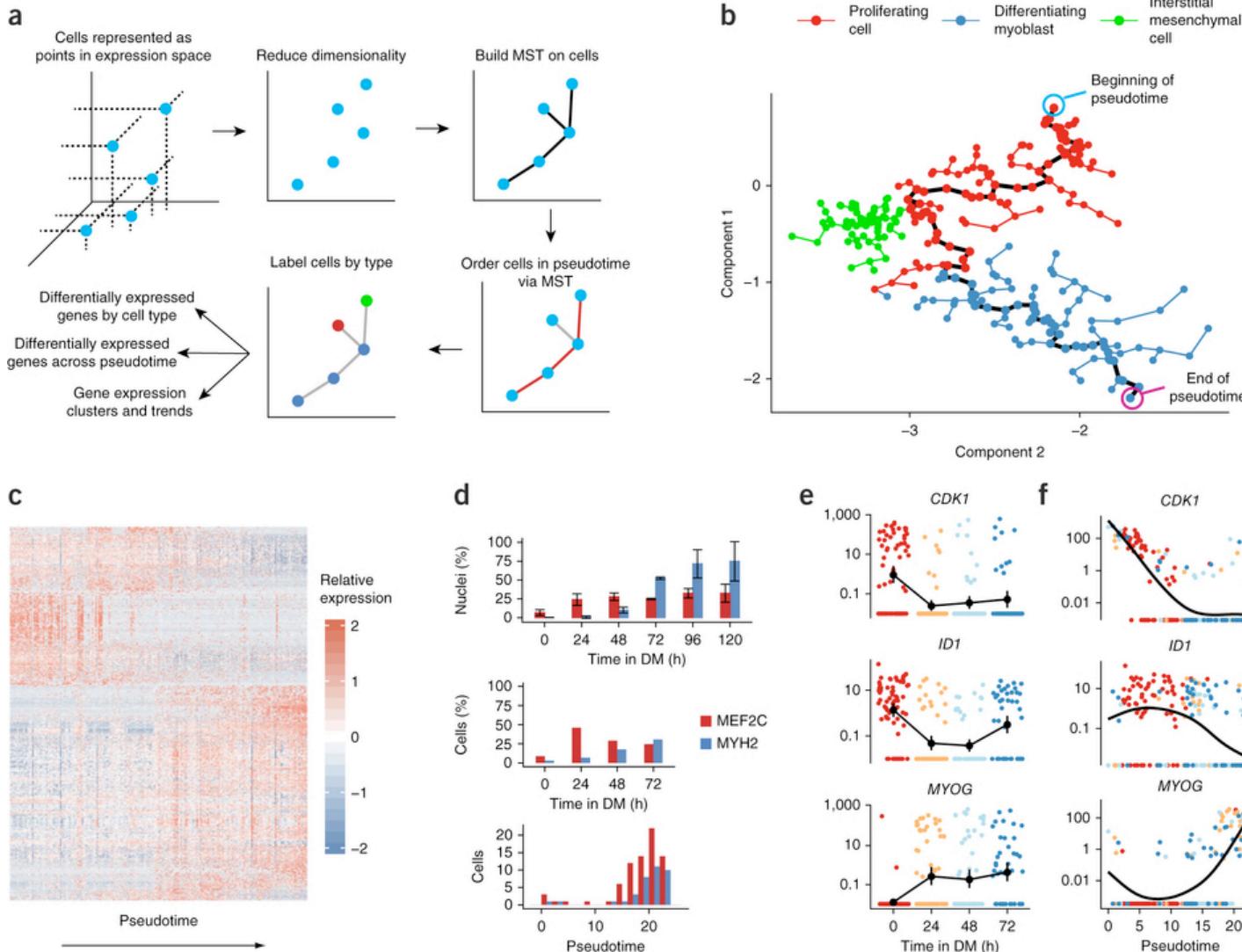
b



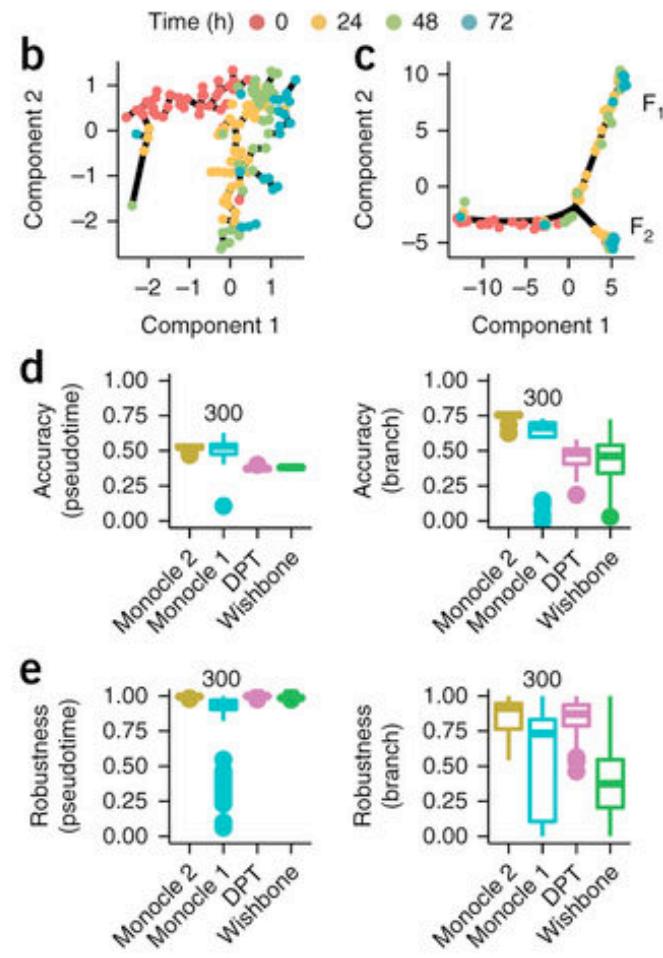
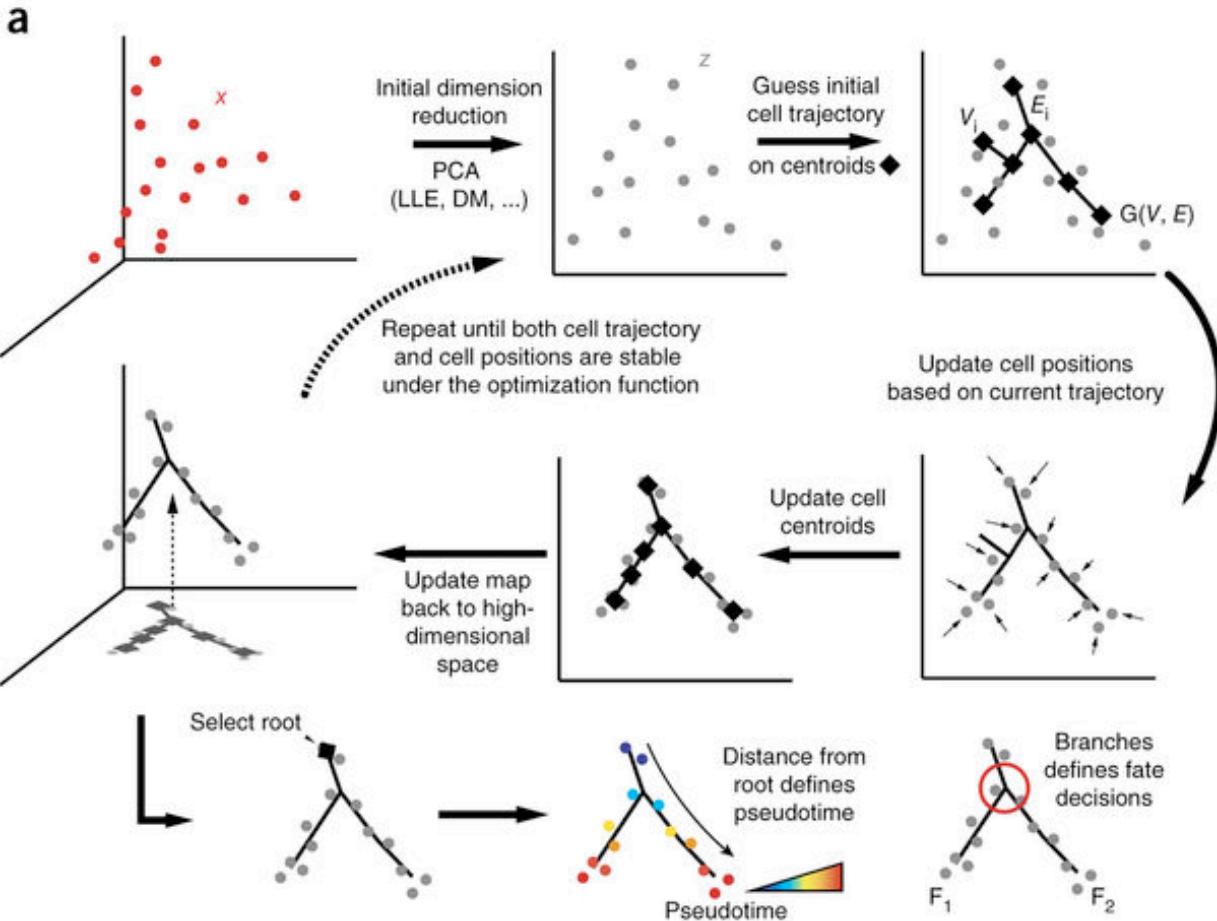
# Tool evaluation



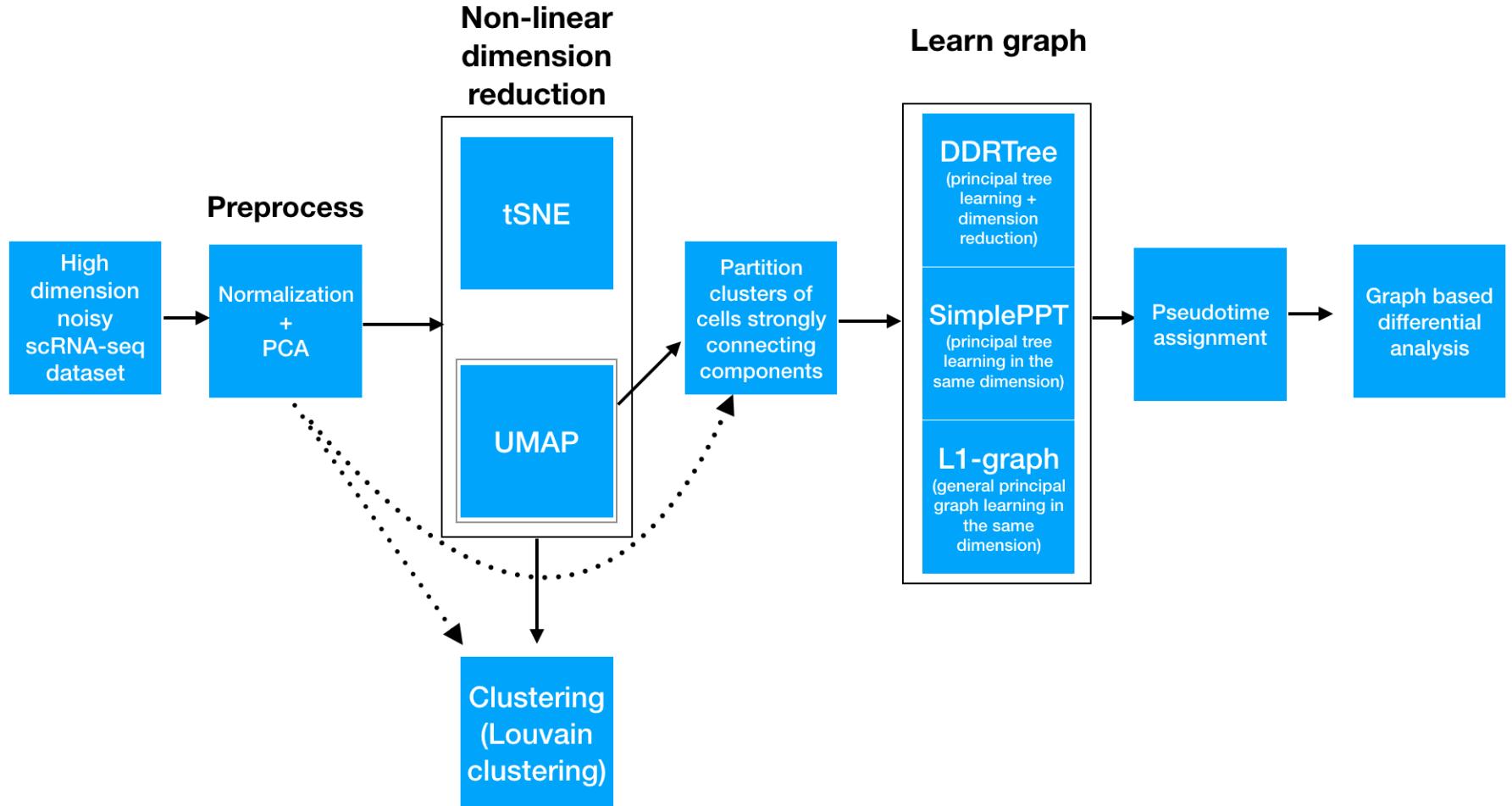
# Pseudotime ordering – Monocle1



# Monocle2 – reversed graph embedding

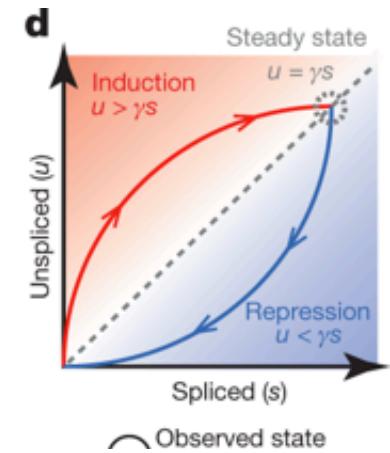
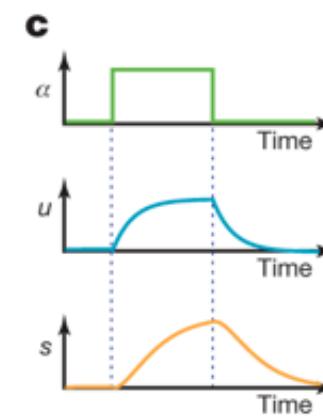
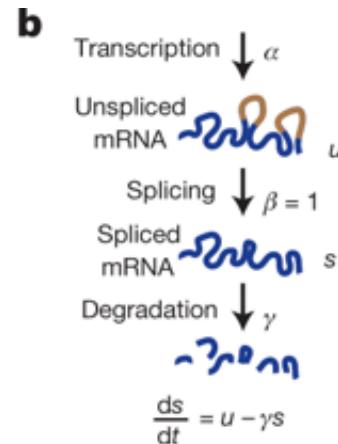
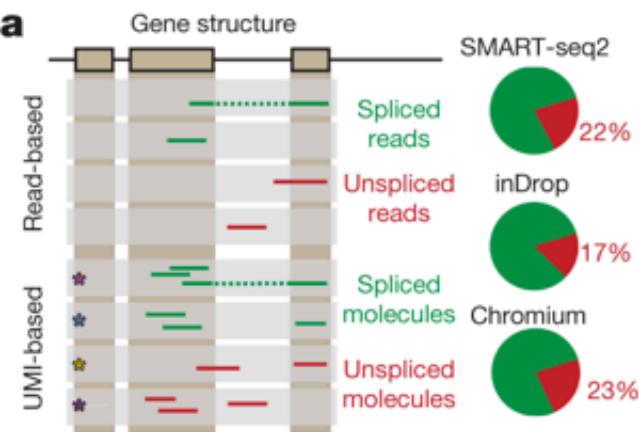


# Monocle 3



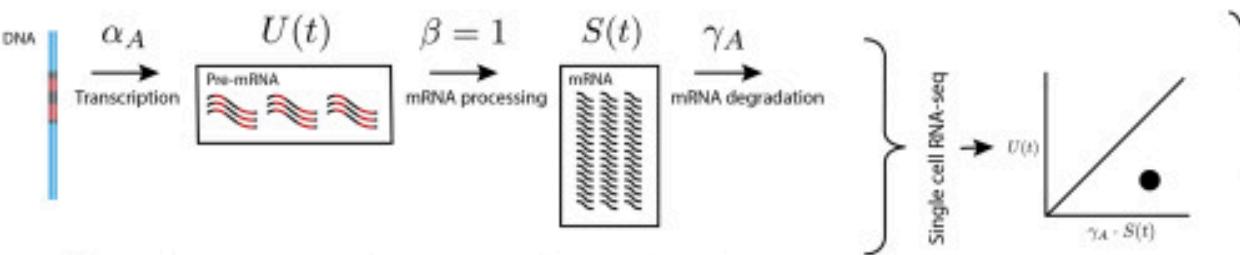
# RNA Velocity

Use proportion spliced/unspliced reads to predict the future state of a cell

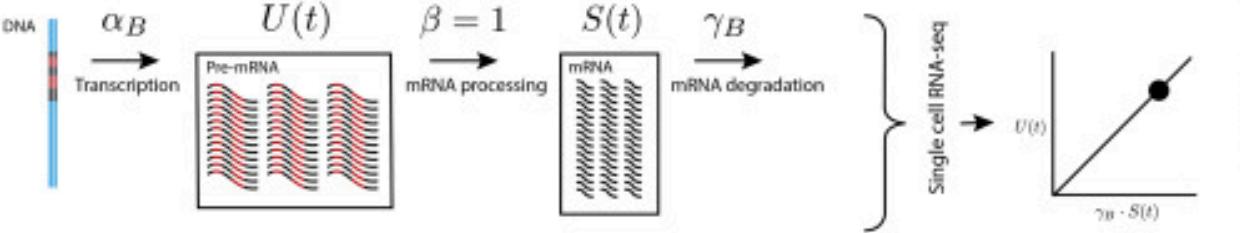


# RNA Velocity

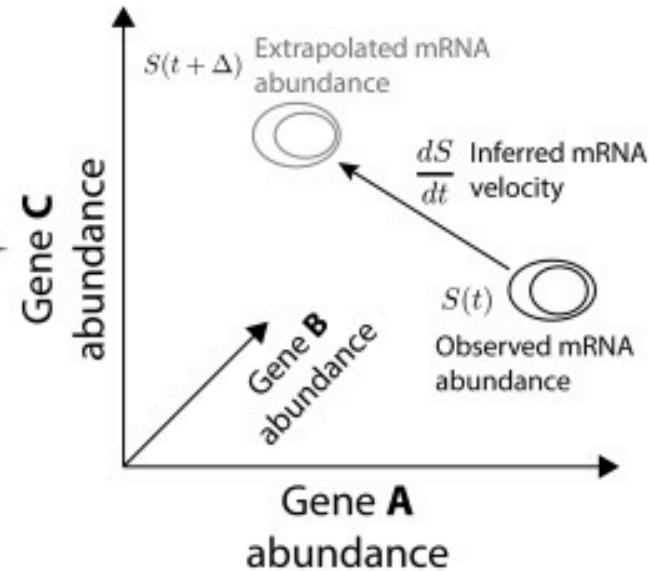
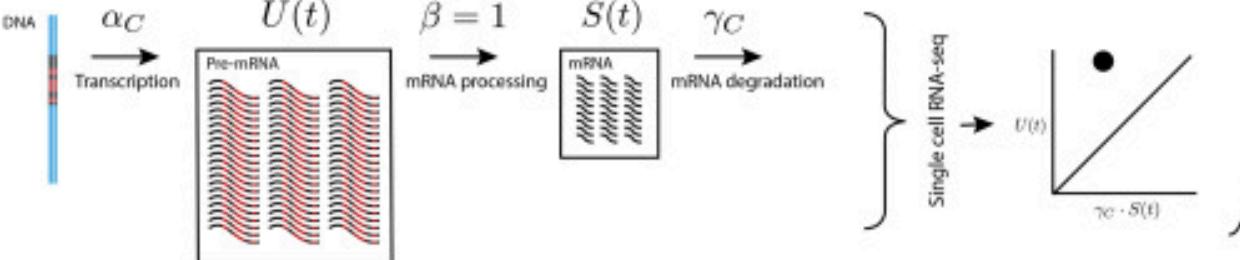
Gene **A** Downregulation, pre-mRNA pool refills slower than mRNA degradation.



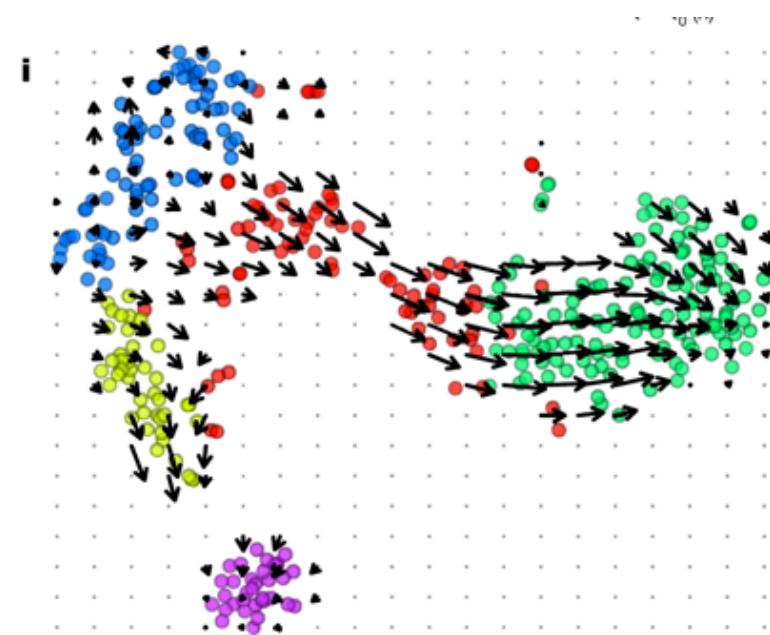
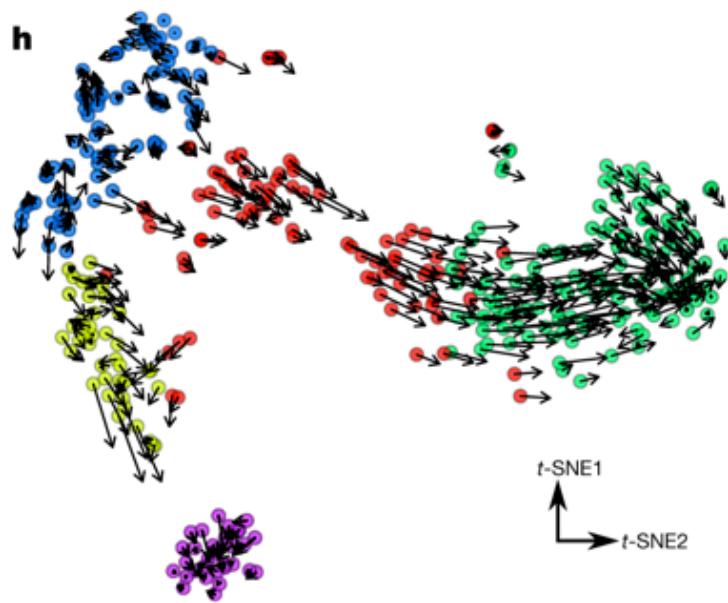
Gene **B** Steady, pre-mRNA pool keeps up with mRNA degradation.



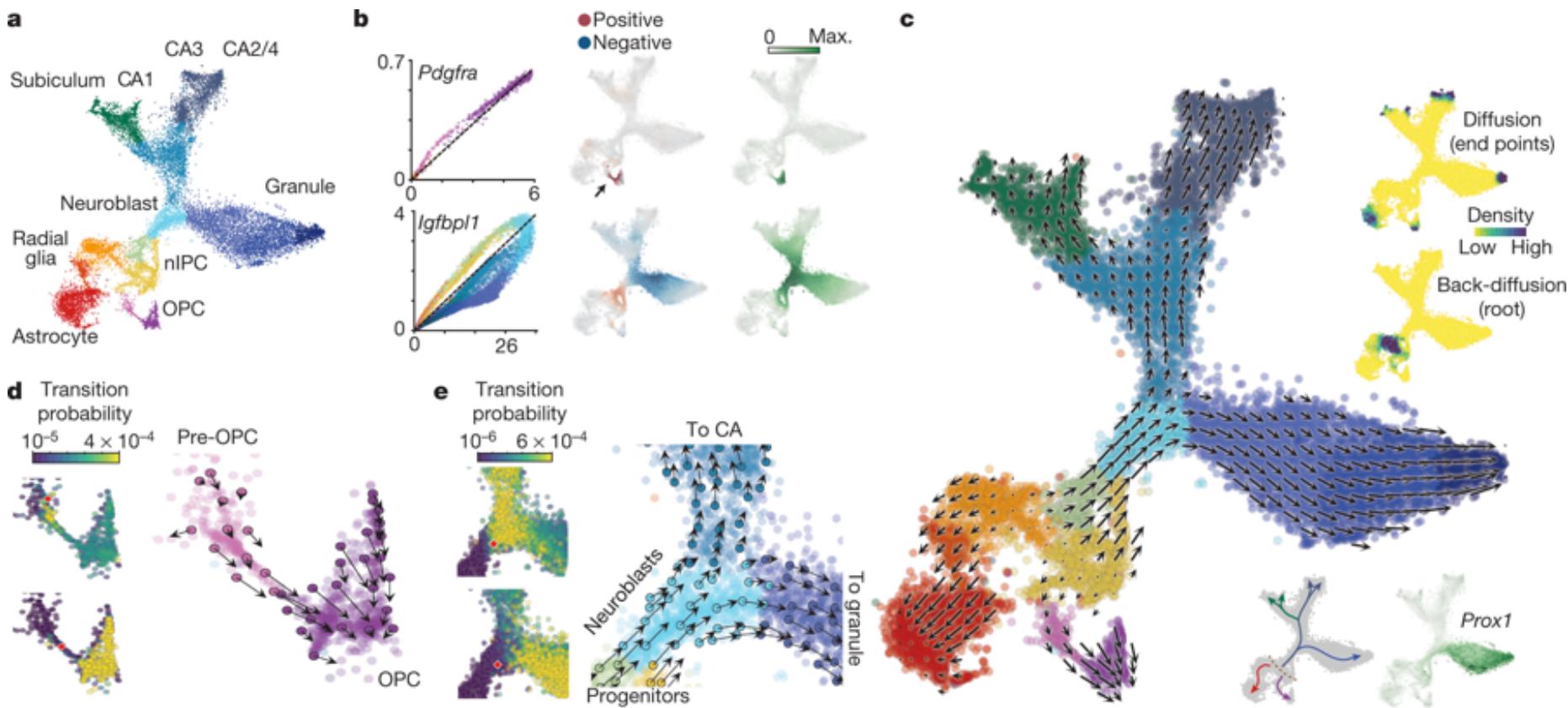
Gene **C** Upregulation, pre-mRNA pool grows faster than mRNA degradation.



# RNA Velocity



# RNA Velocity



# Trajectories - summary

- In reality distance in multidimensional space reflects difference in **transcriptional landscape**, not actual time.
- Necessary to have a **continuum** of states among your cells – will not work with 2 distinct clusters.
- May work with single time-point if ongoing differentiation process – better with multiple time points.



# Additional analyses/data types

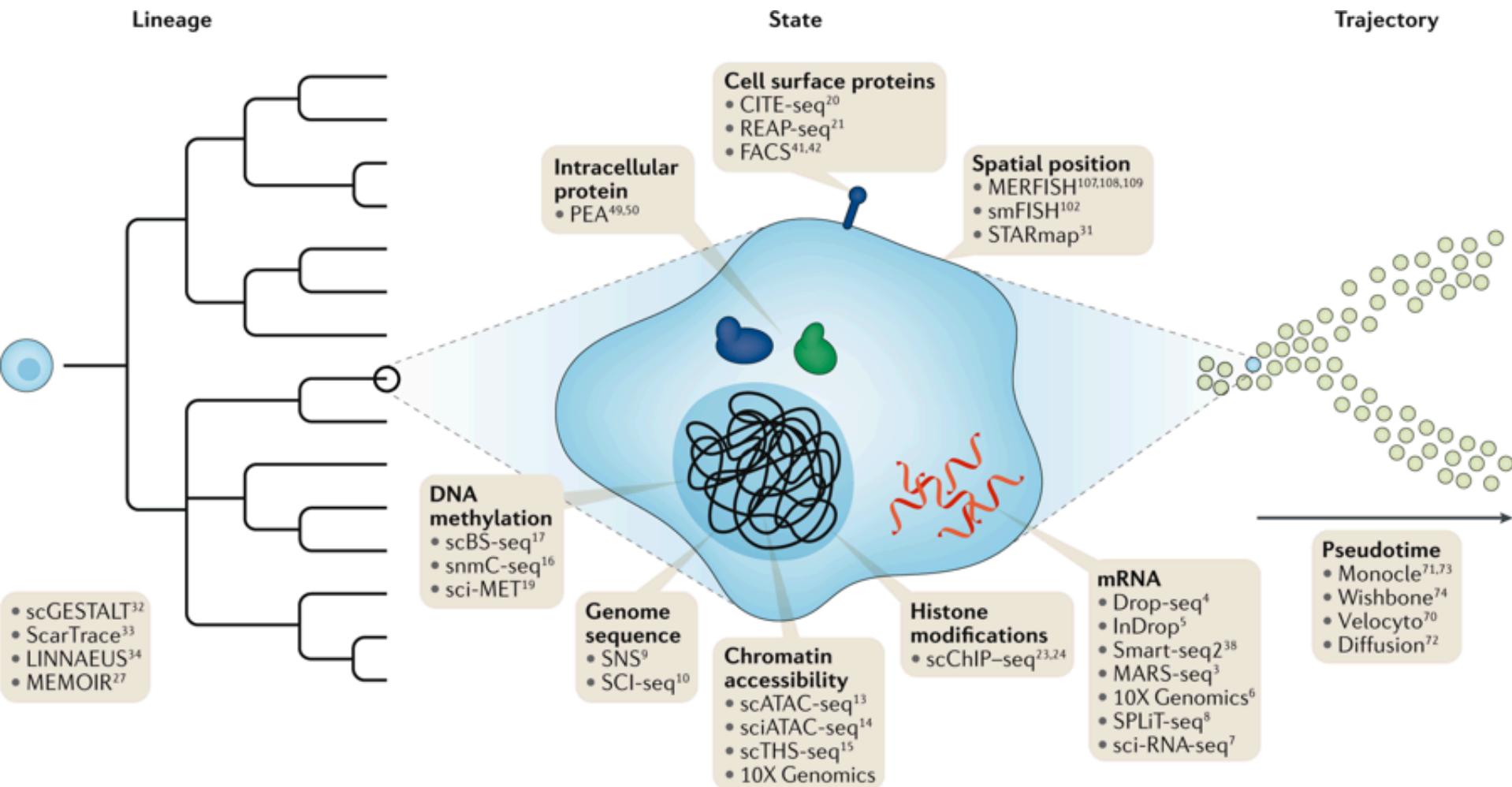
Åsa Björklund

[asa.bjorklund@scilifelab.se](mailto:asa.bjorklund@scilifelab.se)

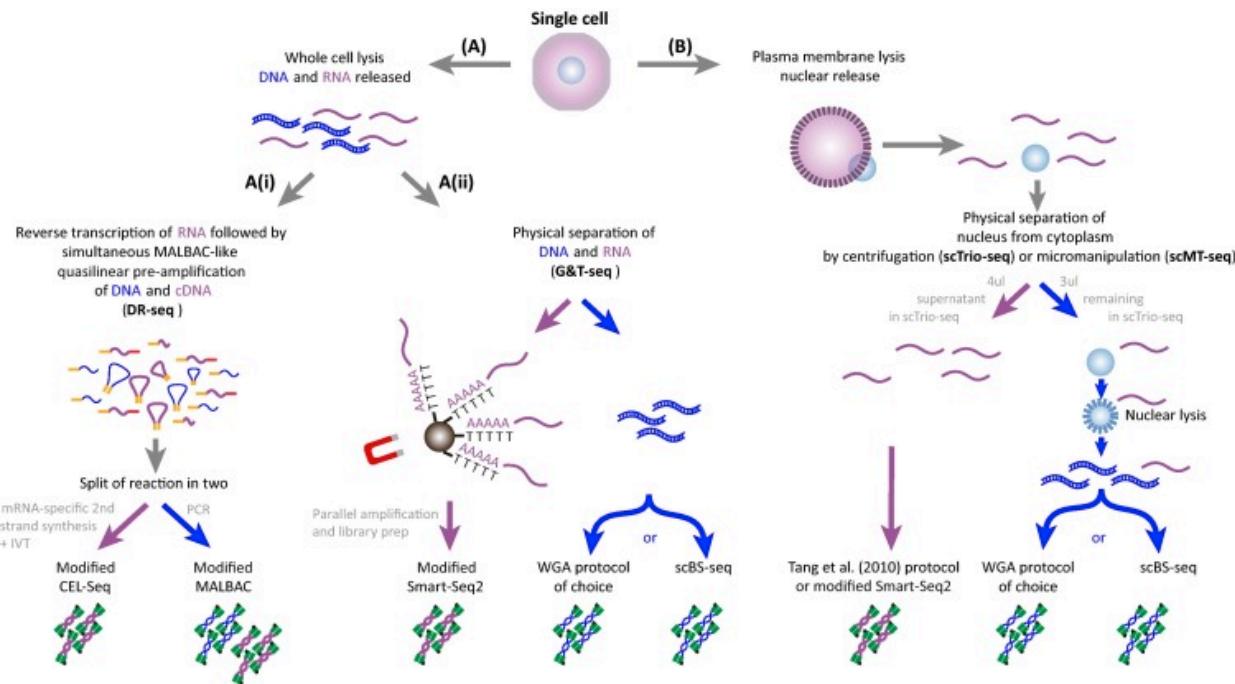
# Brief overview of topics

- Single cell multi-omics
- Allelic expression
- Variant calling
- Alternative splicing
- Copy-number variation
- CRISPR-editing

# Single cell omics

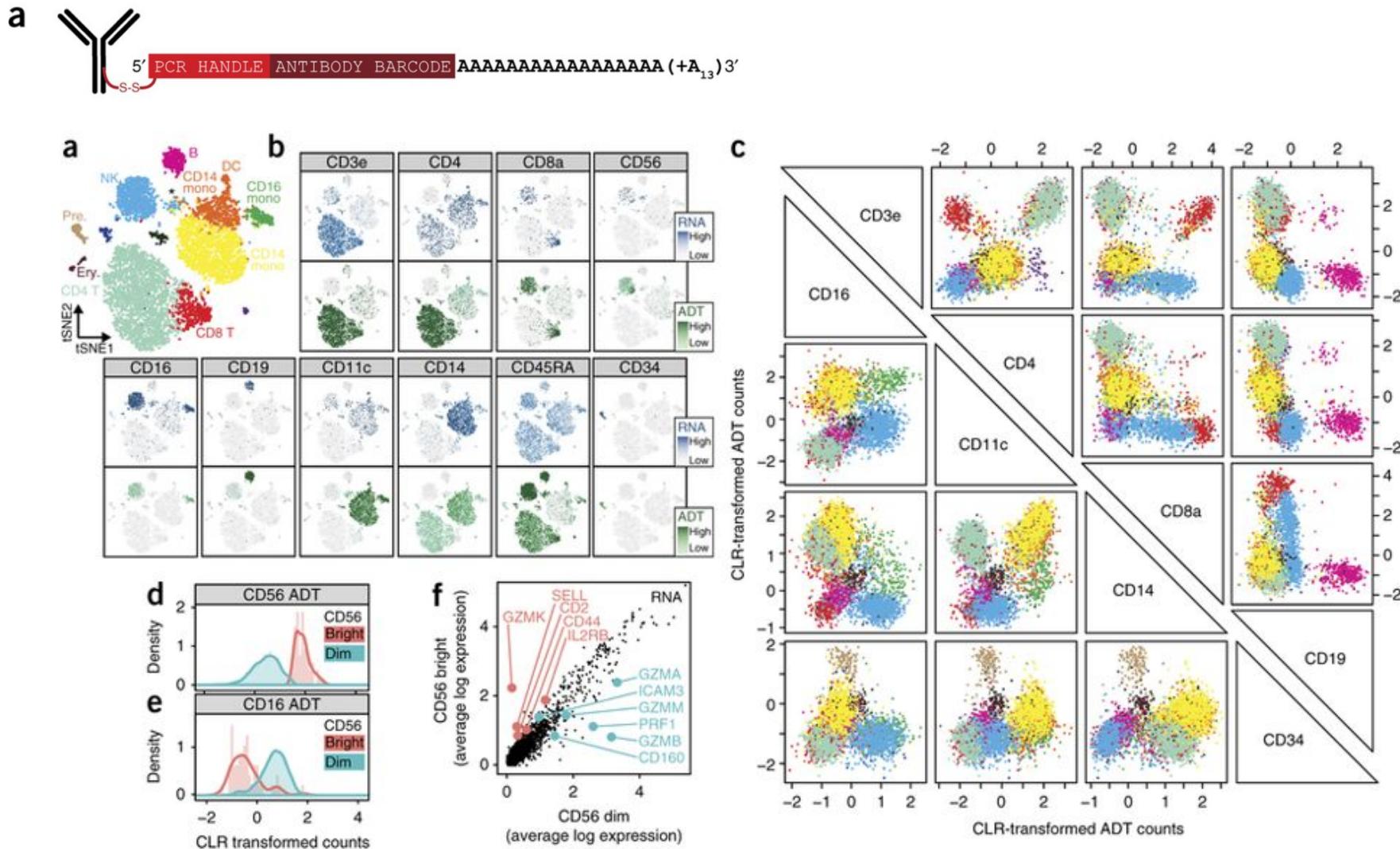


# Genome/Methylome + transcriptome

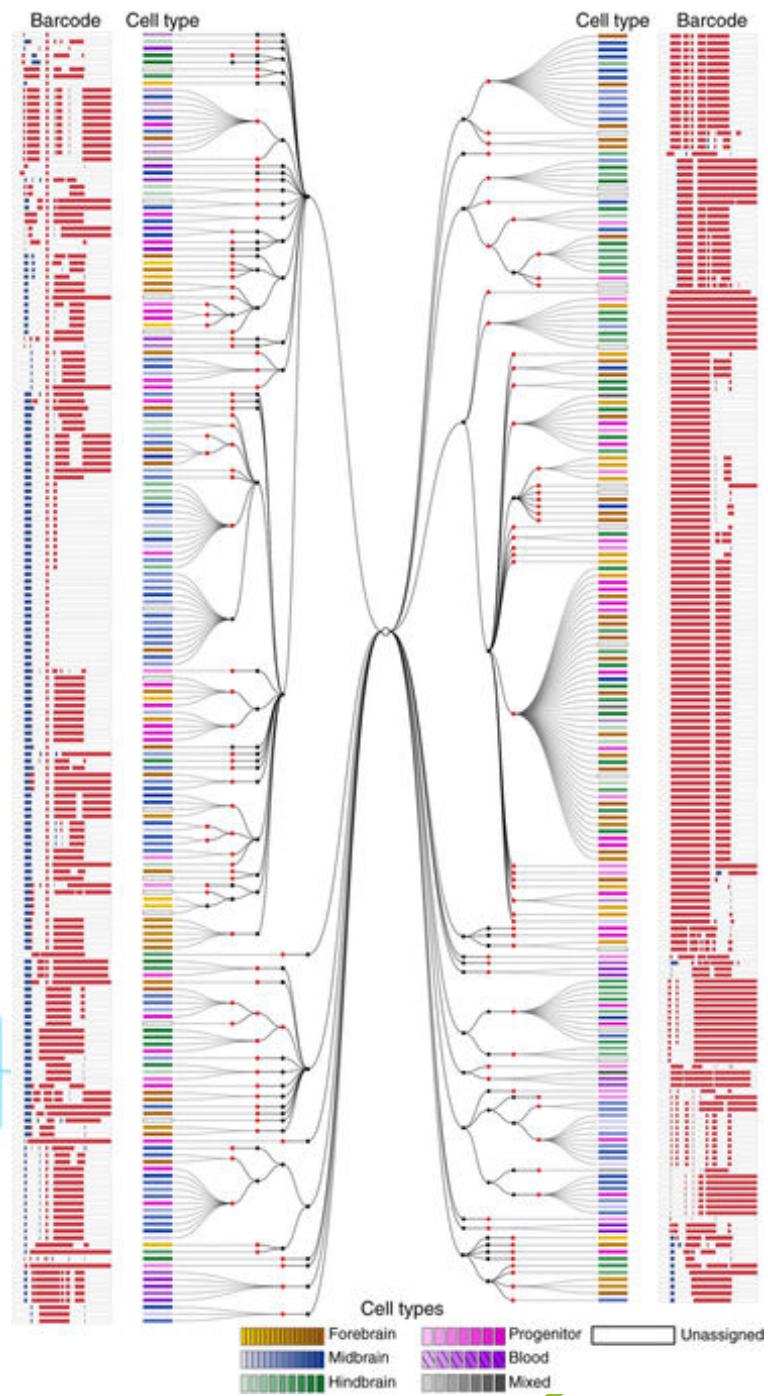
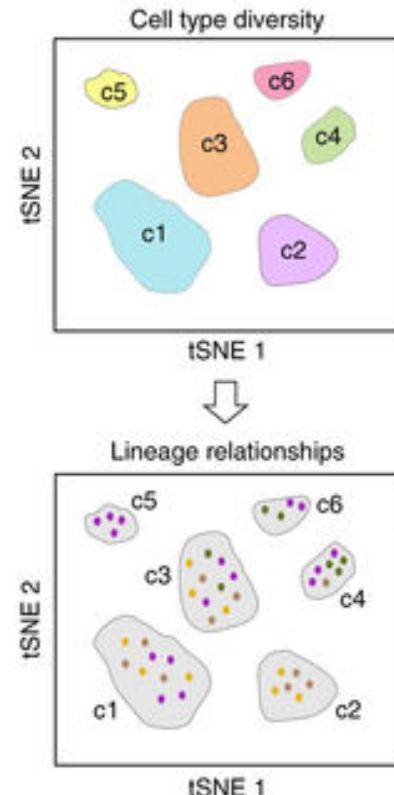
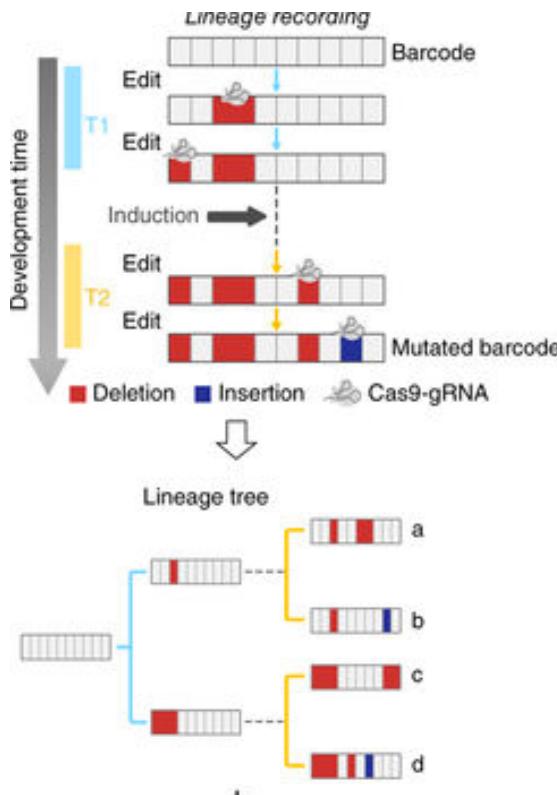


(C)	Loss of nucleic acids	Nature of RNA-seq	Nature of gDNA-seq	Shown amenable to bisulphite-sequencing
DR-seq	Minimal risk of loss	3'end tag transcript seq	MALBAC-like amplified gDNA, contaminated with co-amplified cDNA	no
G&T-seq (like)	Potential loss of mRNA and DNA molecules	Full-length transcript seq	In line with chosen WGA	yes
scTrio-seq	Loss of nearly half of cytoplasmic and all nuclear mRNA-molecules	Full-length transcript seq	reduced representation bisulphite-seq	yes
scMT-seq	Loss of some cytoplasmic and all nuclear mRNA-molecules during micromanipulation	Full-length transcript seq	reduced representation bisulphite-seq	yes

# CITE-seq – epitope + RNAseq

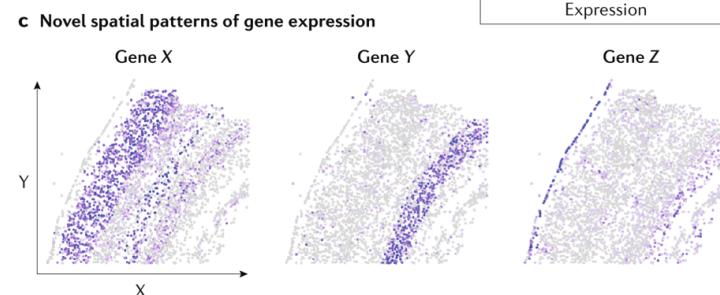
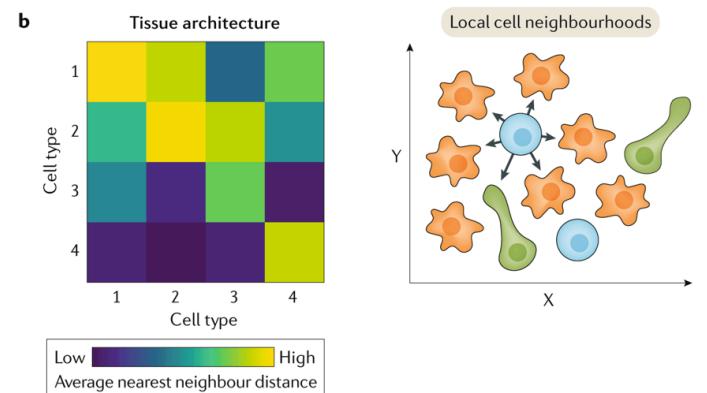
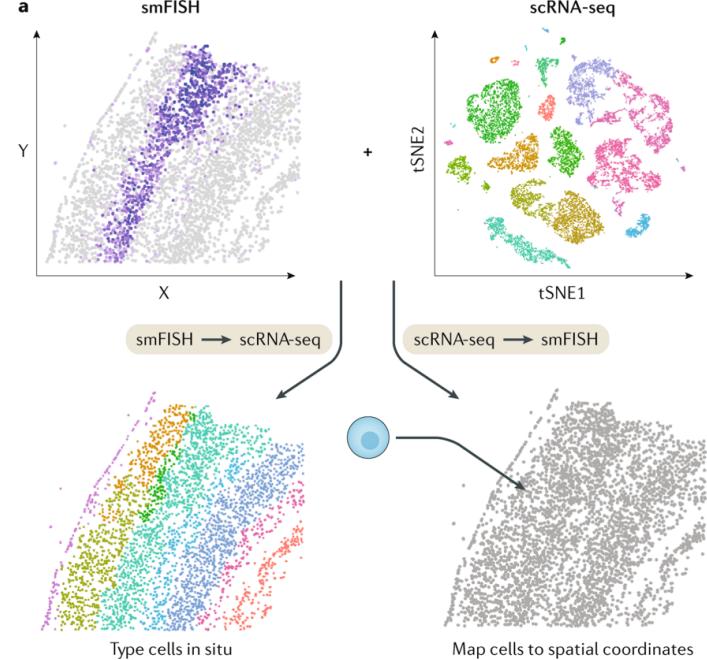


# scGESTALT – lineage tracing and cell profiling with CRISPR-Cas9 editing of barcodes



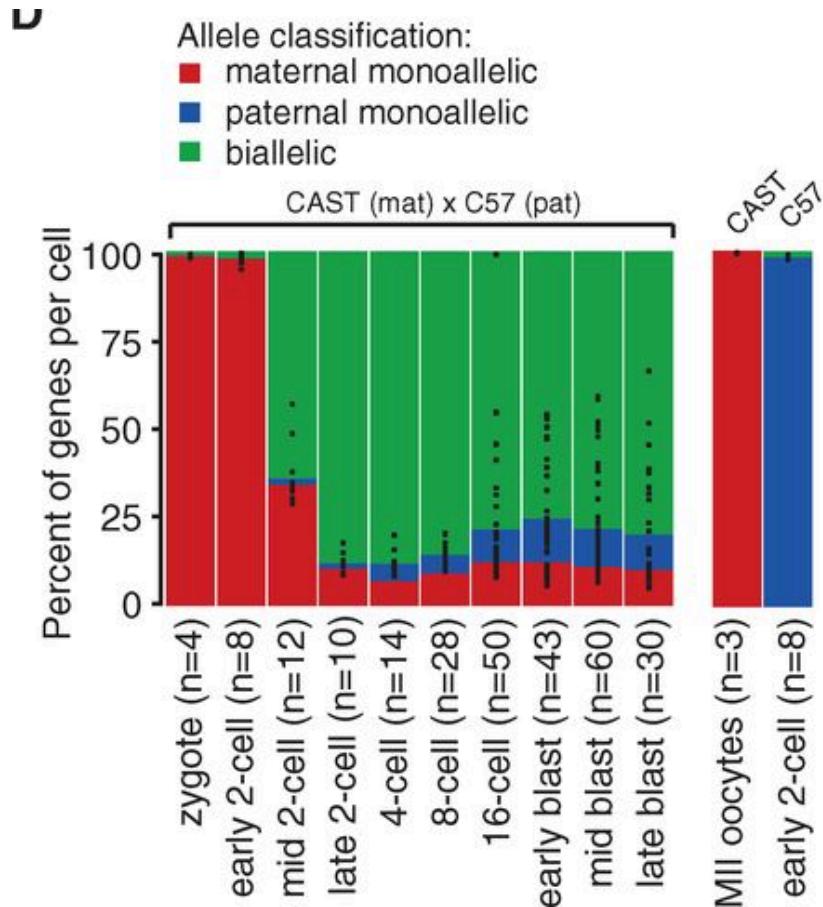
# Spatial integration

- Spatial Transcriptomics
- smFISH
- *In situ* sequencing
- starMAP
- MERFISH

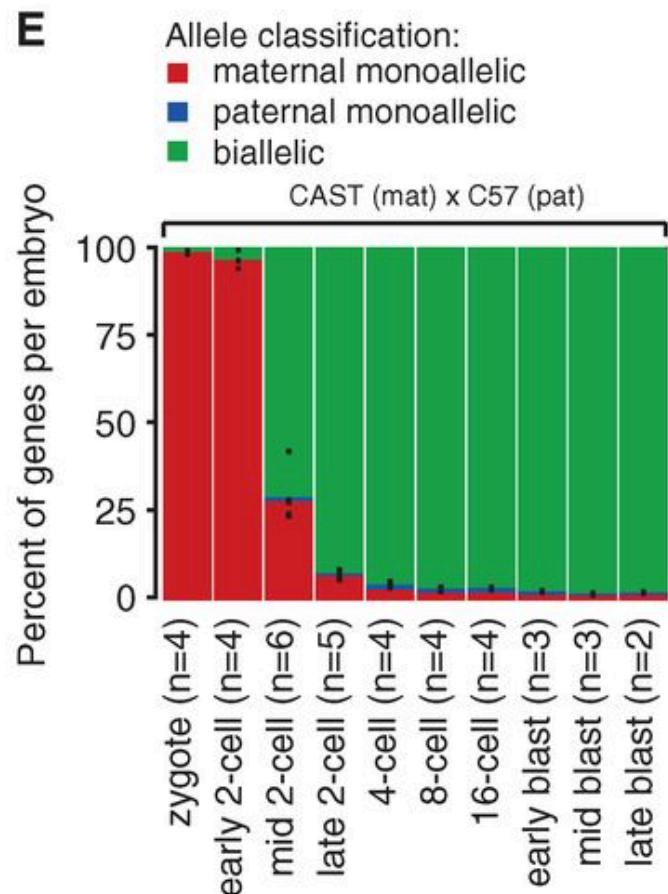


# Single-Cell RNA-Seq Reveals Dynamic, Random Monoallelic Gene Expression in Mammalian Cells

## Single cells

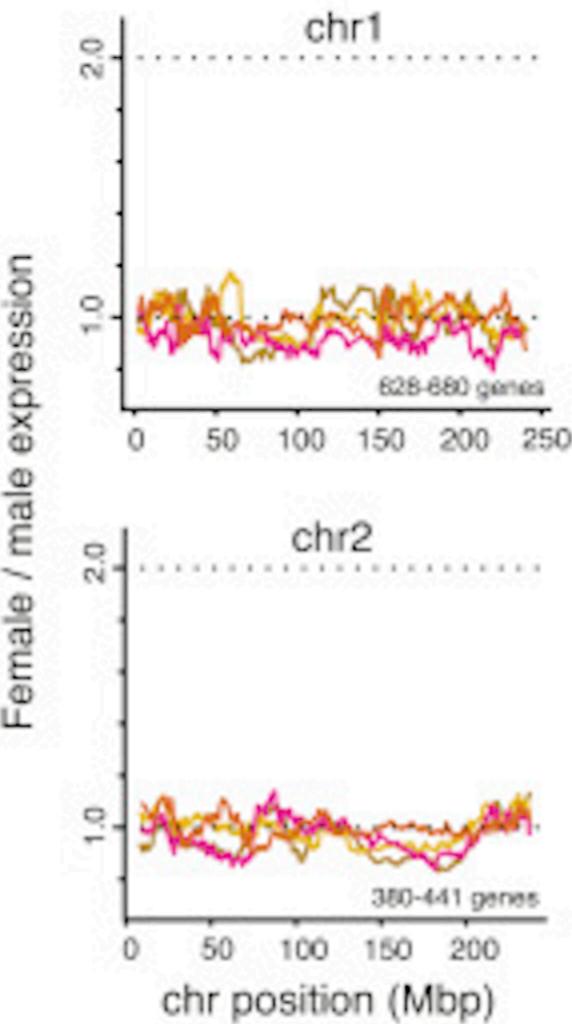
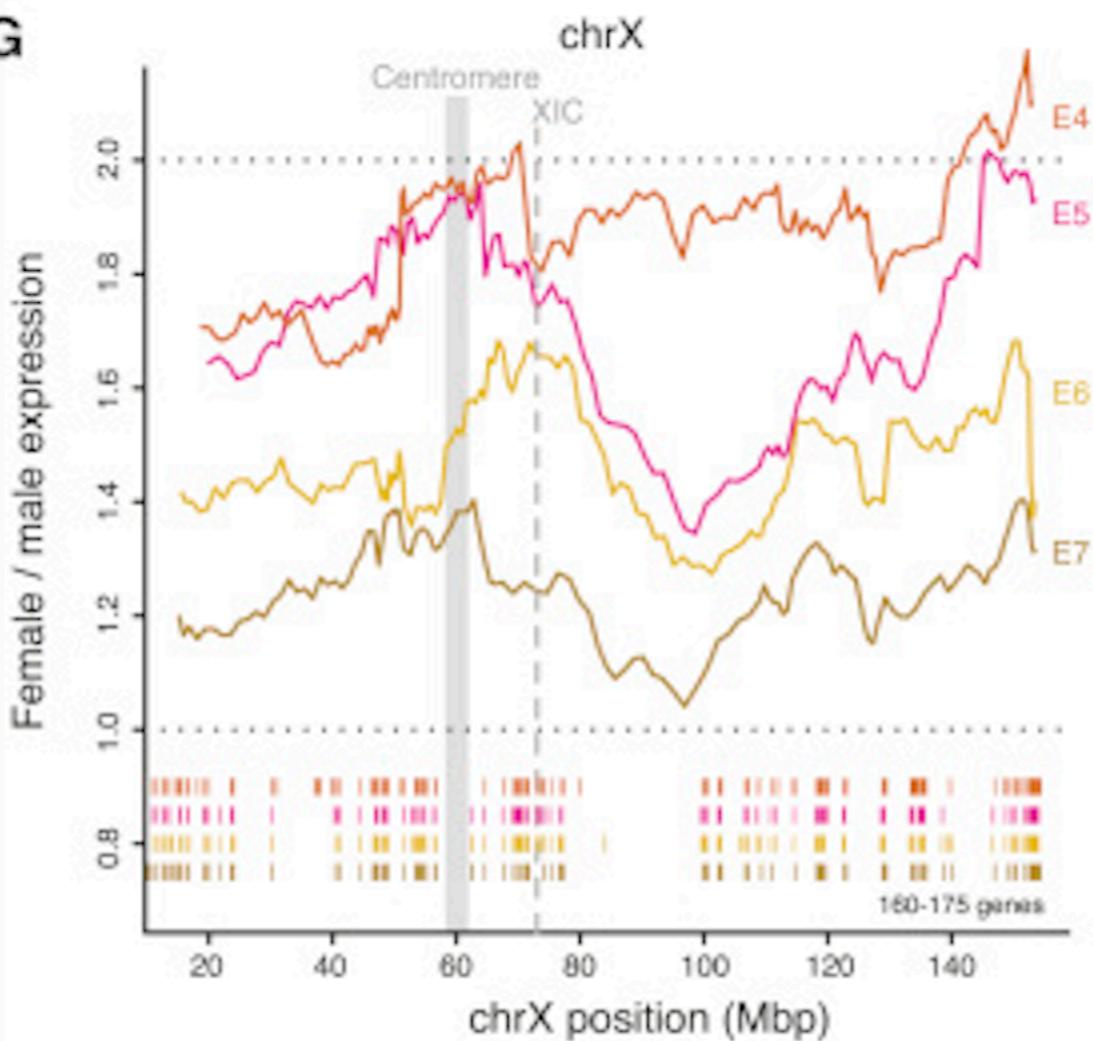


## Pooled embryos

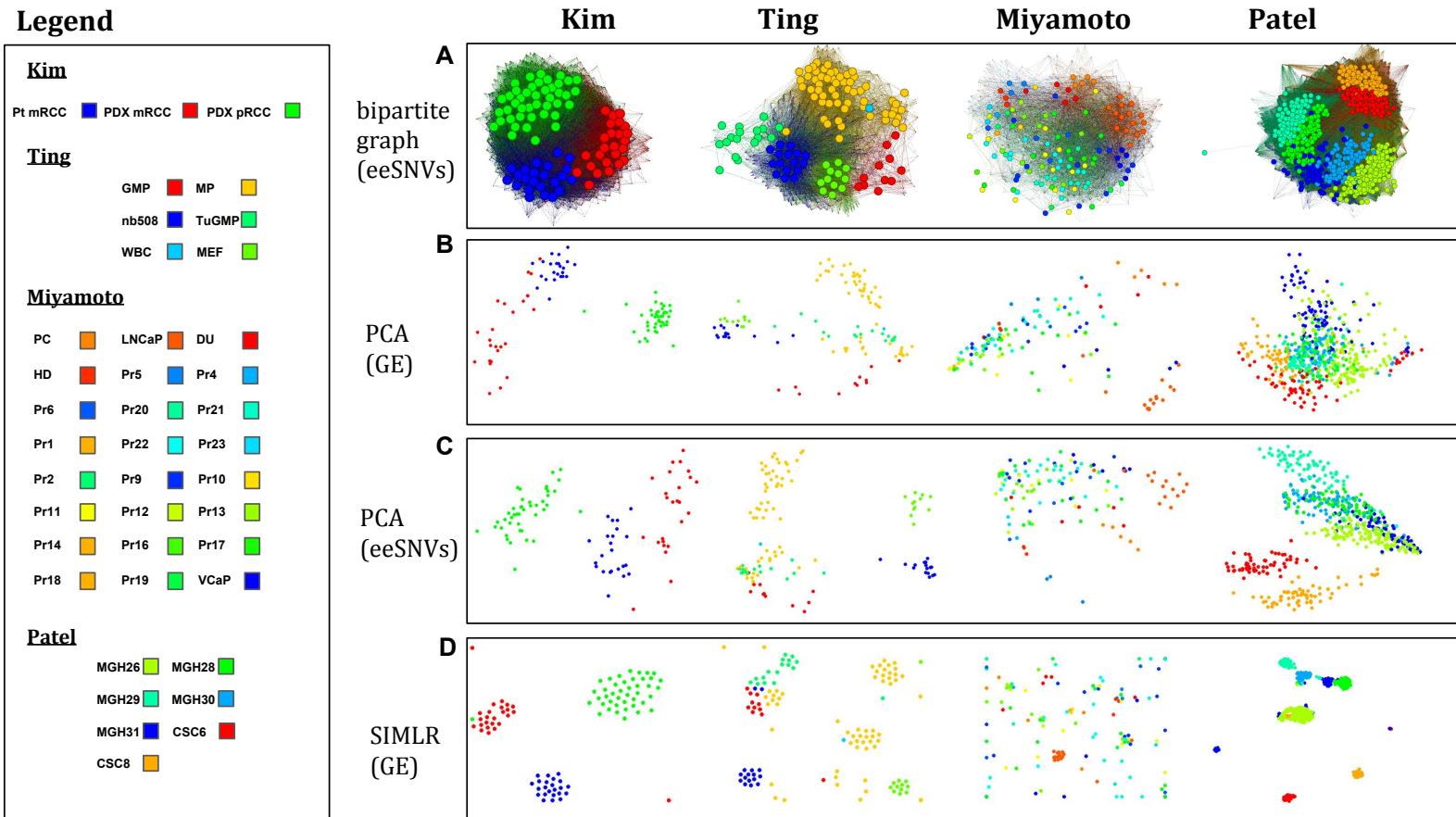


# X Chromosome inactivation

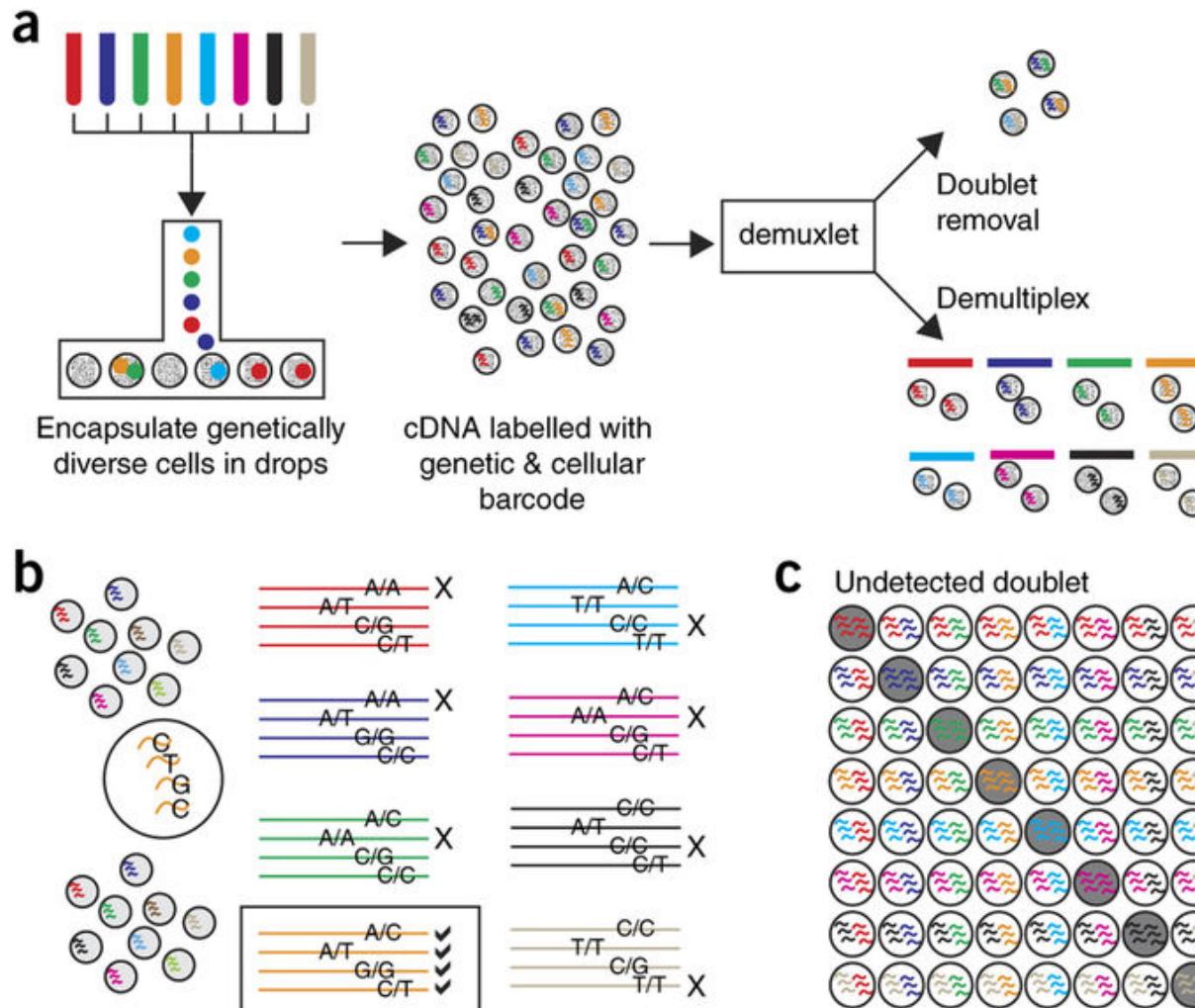
G



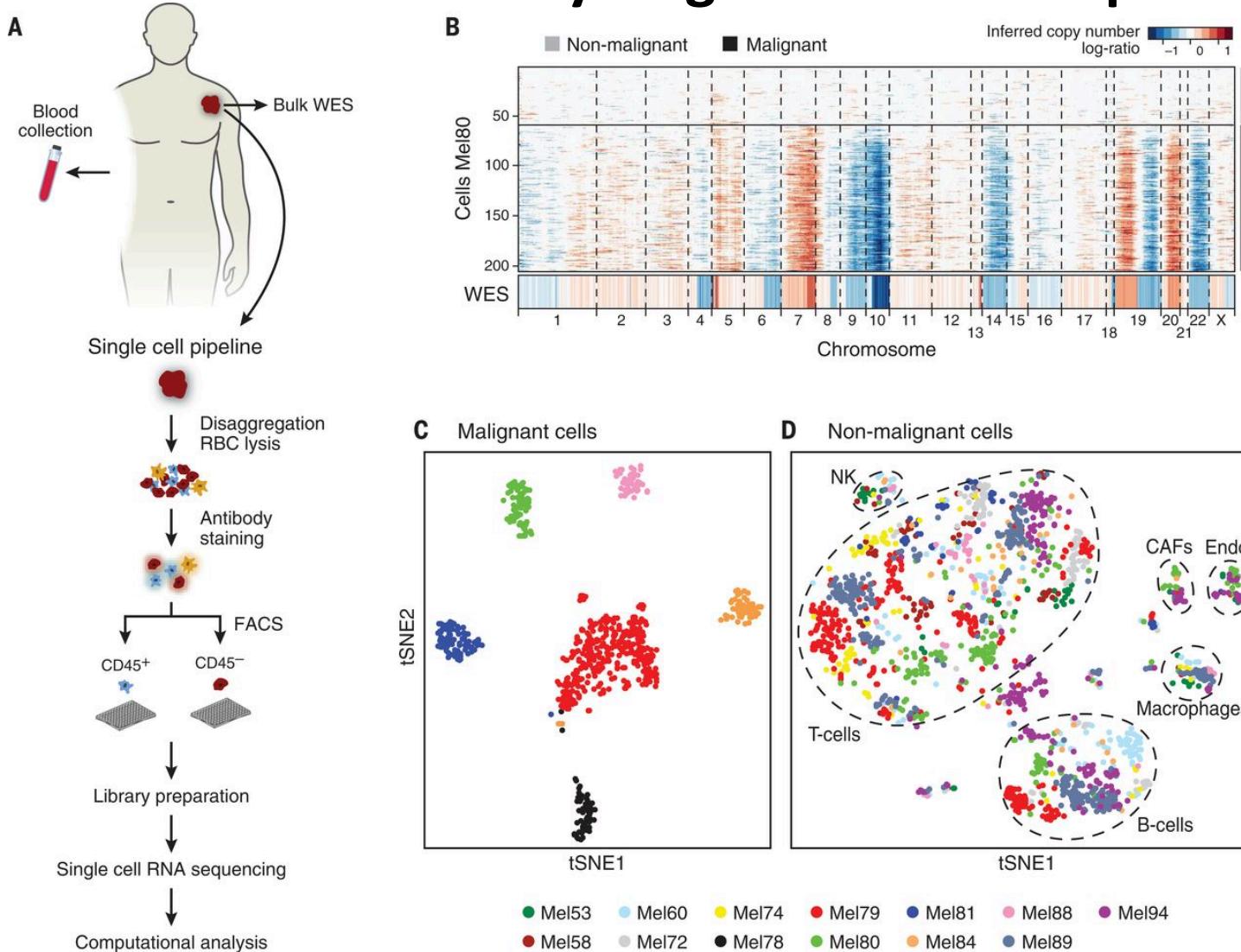
# Using Single Nucleotide Variations in Cancer Single-Cell RNA- Seq Data for Subpopulation Identification and Genotype- phenotype Linkage Analysis



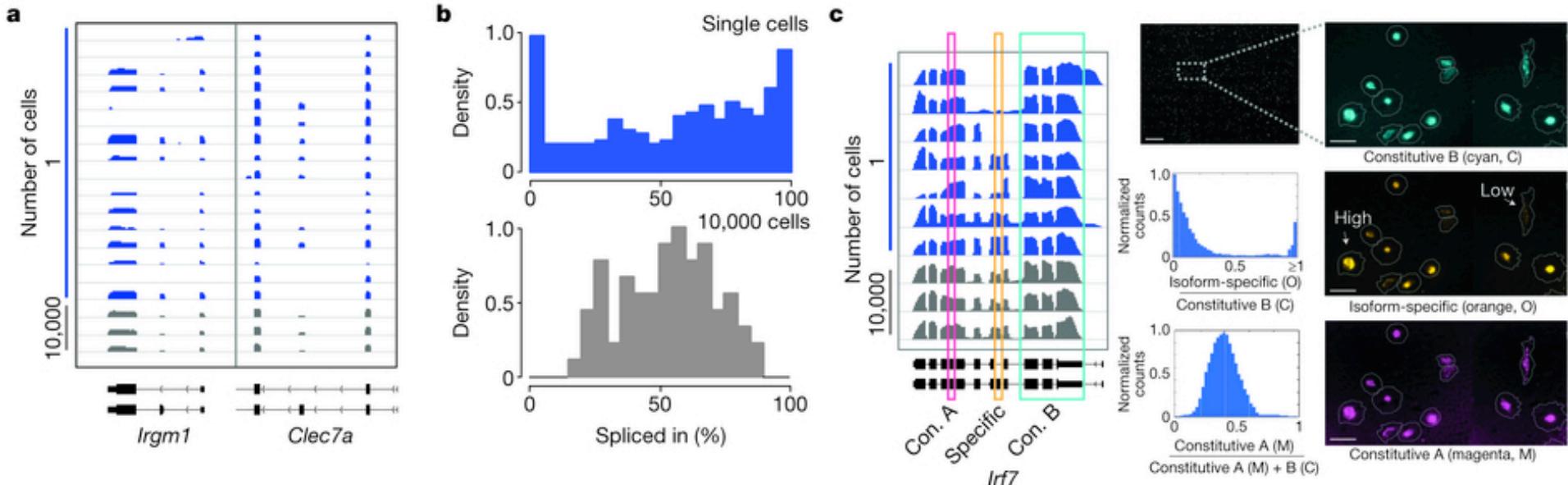
# Multiplexed droplet single-cell RNA-sequencing using natural genetic variation



# Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq



# Cell specific alternative splicing

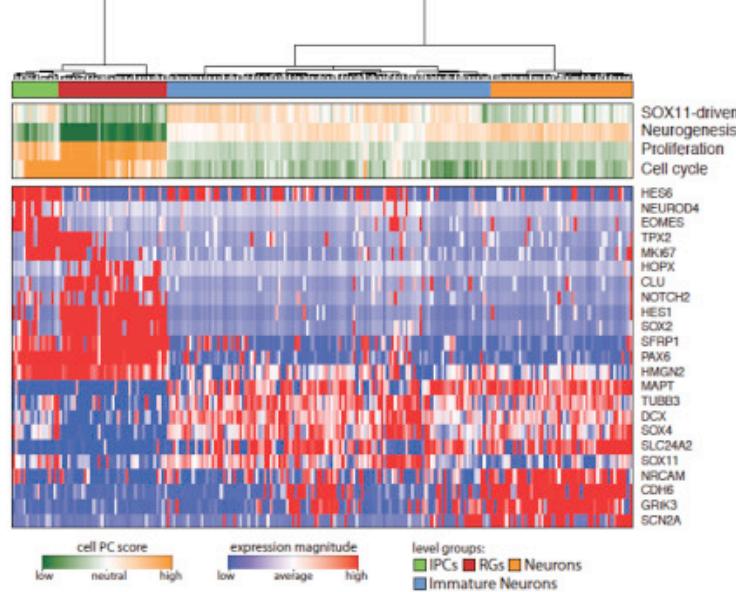


D

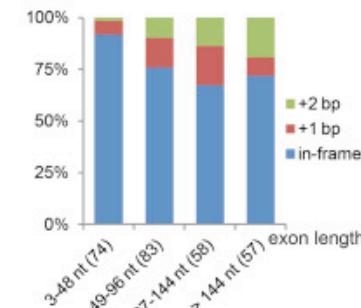
Alternative transcript event types	$ \Delta\text{PSI}  \geq 10\% \text{ BF} \geq 5$		
	Neuron	NPC	Total
Skipped exon (SE)	198	74	272
Retained intron (RI)	11	4	15
Alternative 5' splice site (A5SS)	10	6	16
Alternative 3' splice site (A3SS)	9	15	24
Mutually exclusive exon (MXE)	7	8	15
Alternative first exon (AFE)	104	77	181
Alternative last exon (ALE)	89	85	174
Tandem 3' UTRs (UTR)	34	11	45

Inclusive/extended isoform      Exclusive isoform

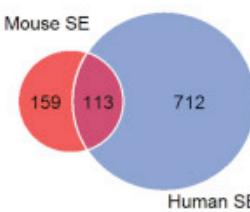
G



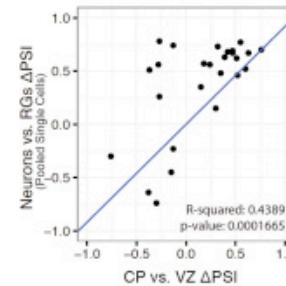
E



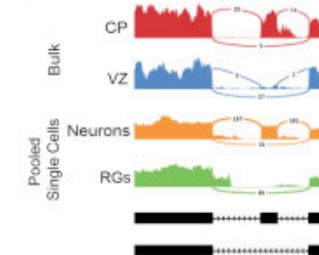
F



H

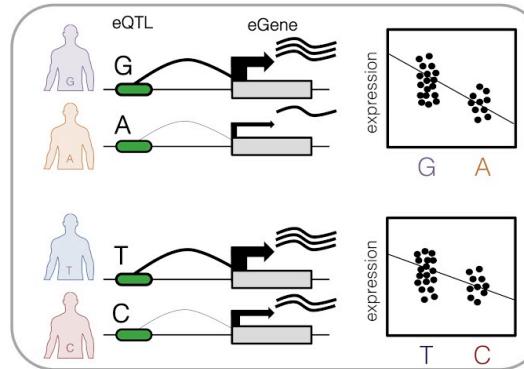
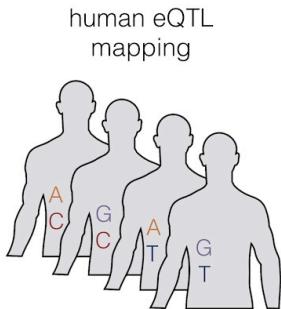


I

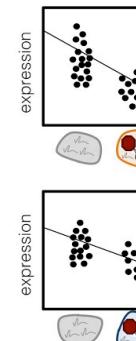
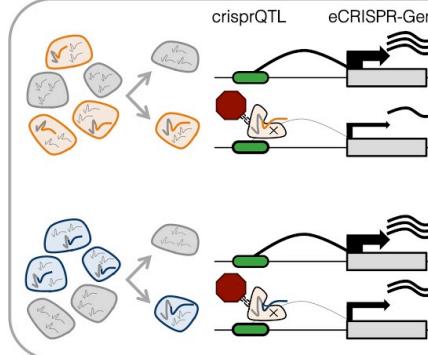
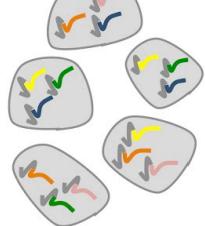


# crisprQTL mapping

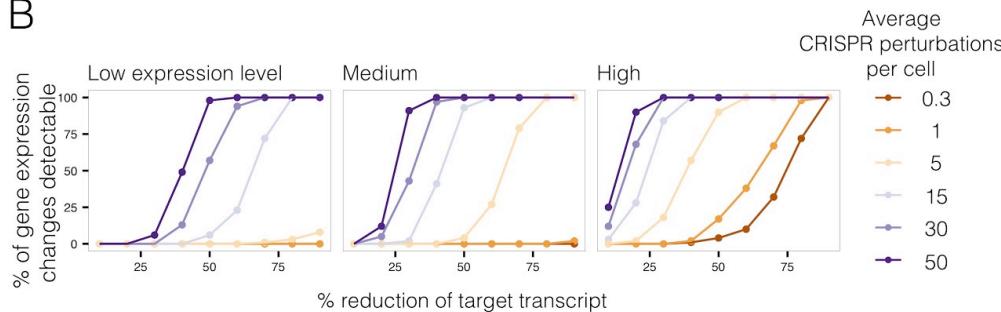
1A



crisprQTL  
mapping

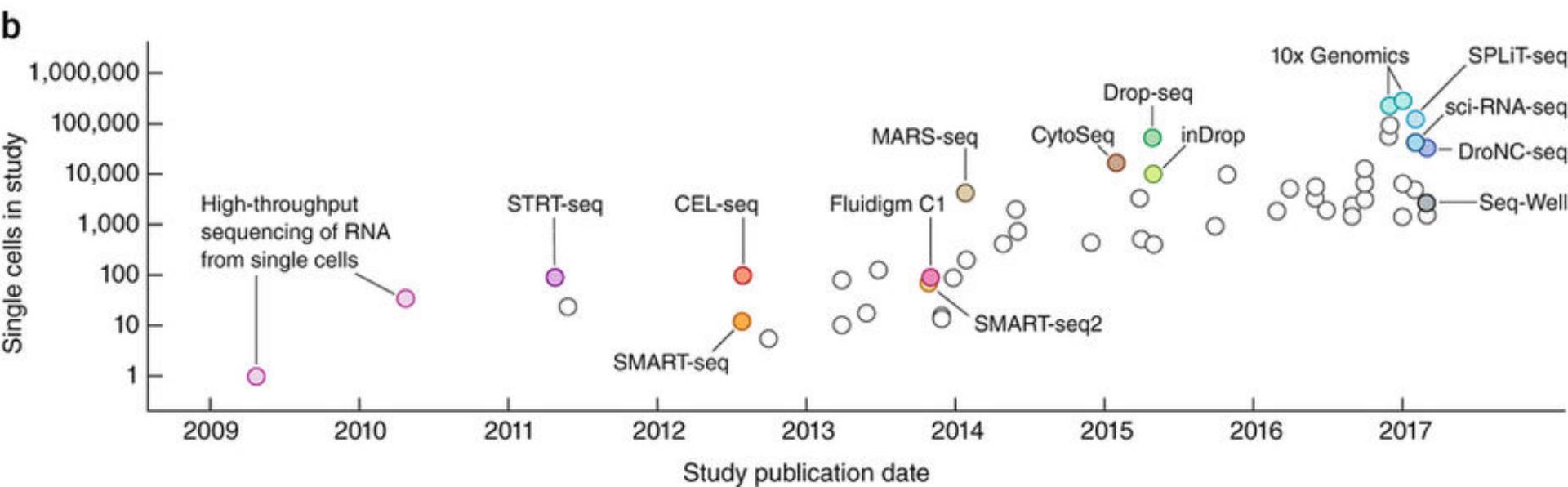
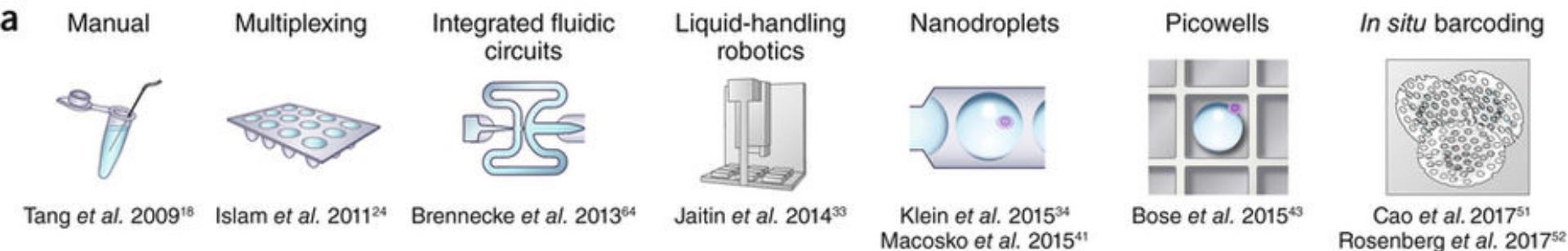


B



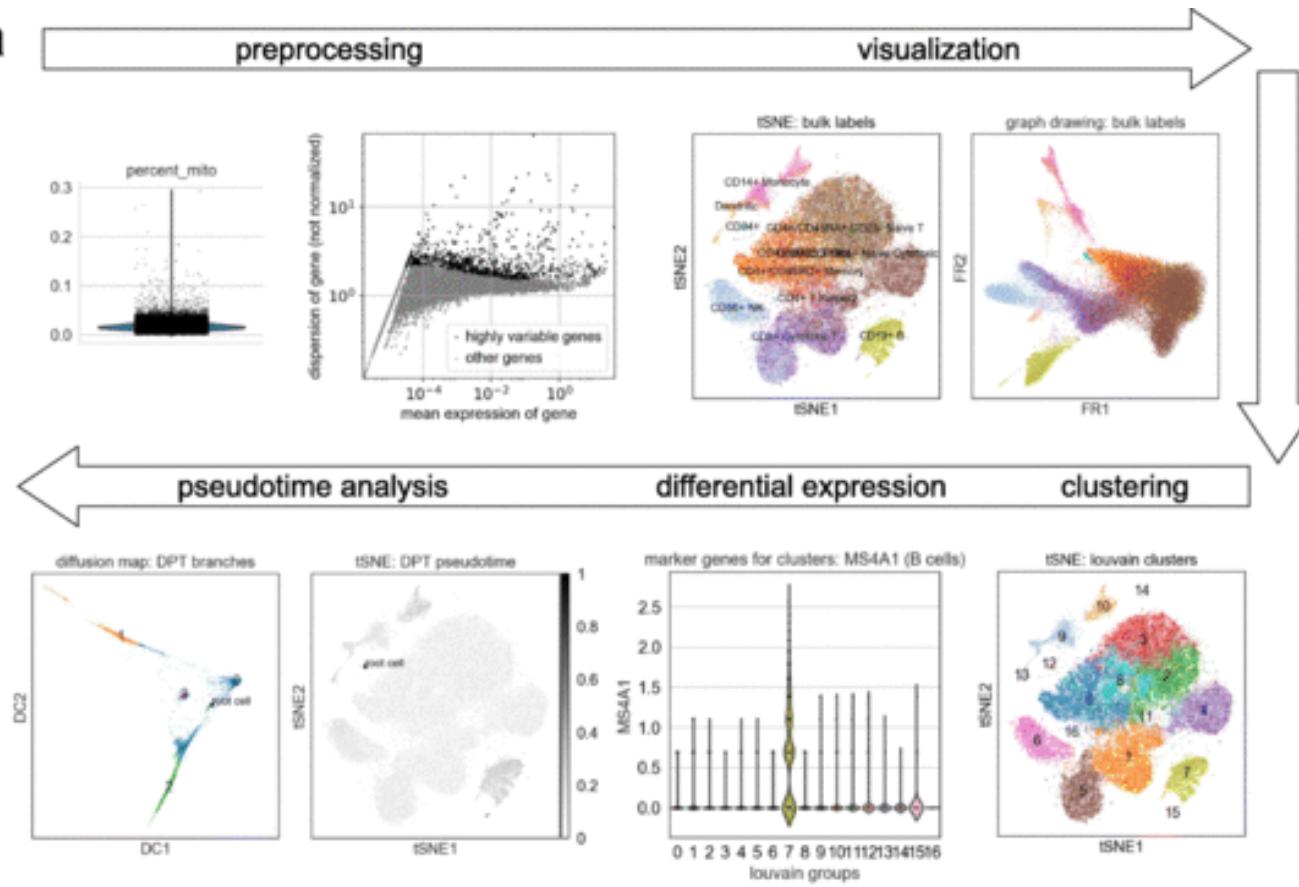
(Gasperini et. al. BioRxiv 2018)

# Large scale analysis

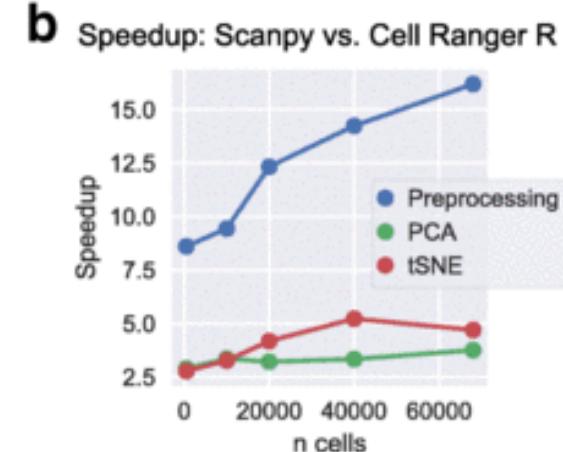


# Large scale analysis

a



b



c

