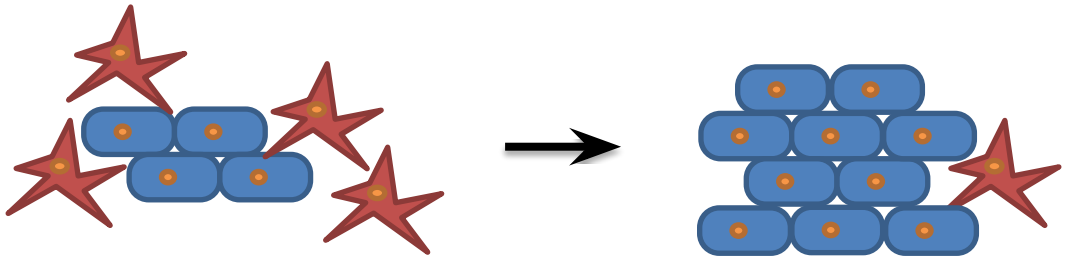
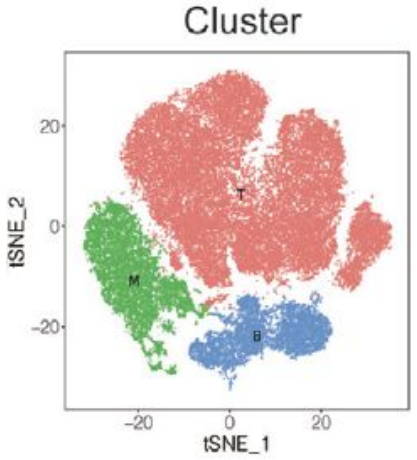
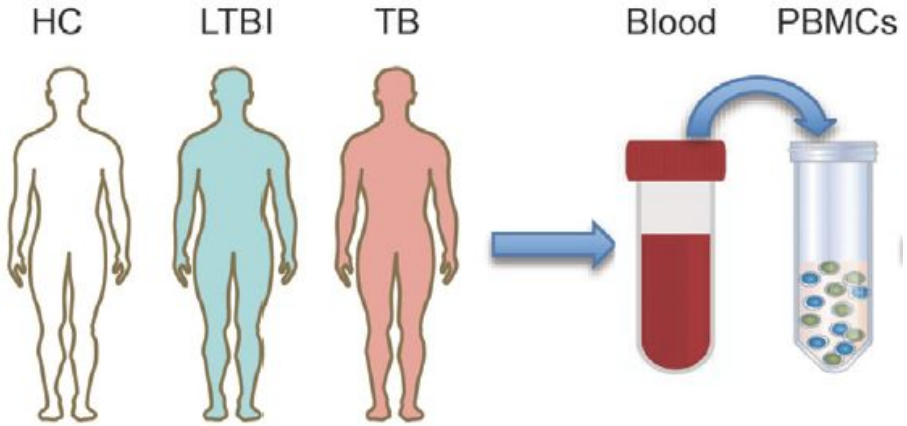
Single-cell RNA-seq

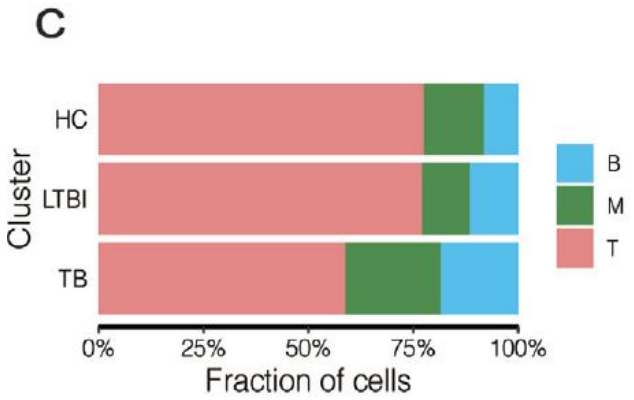
Artem Artemov
22.12.2020

Motivation 1: cell composition

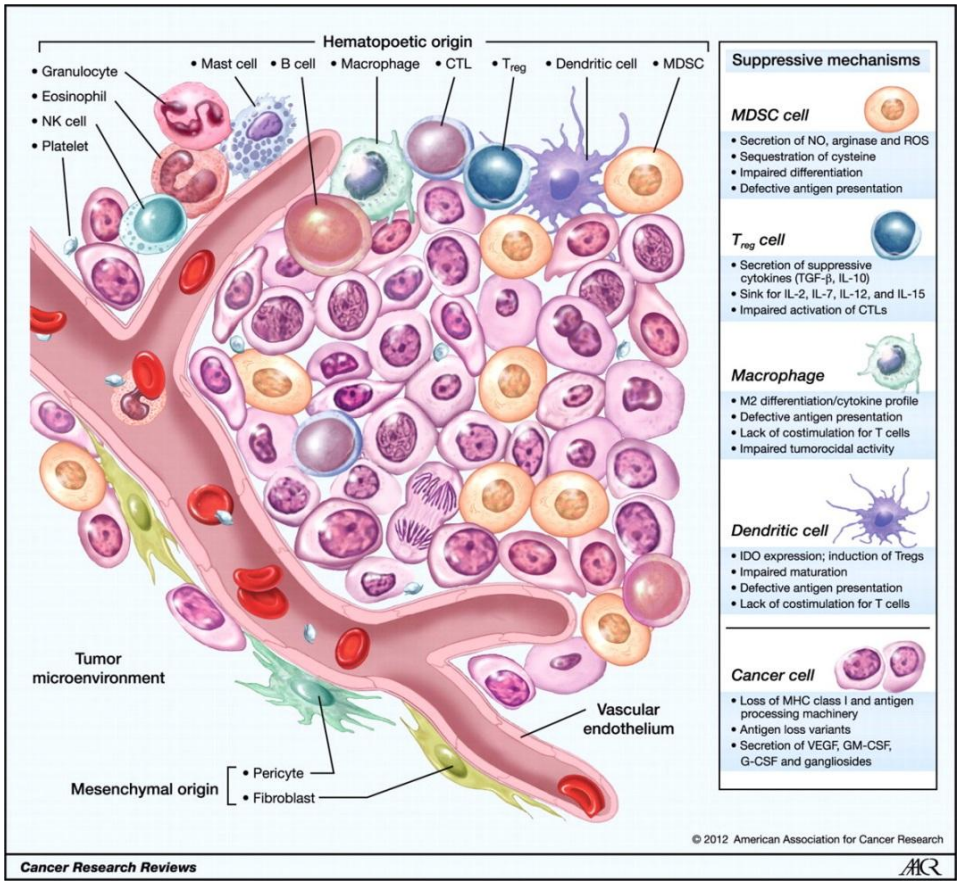


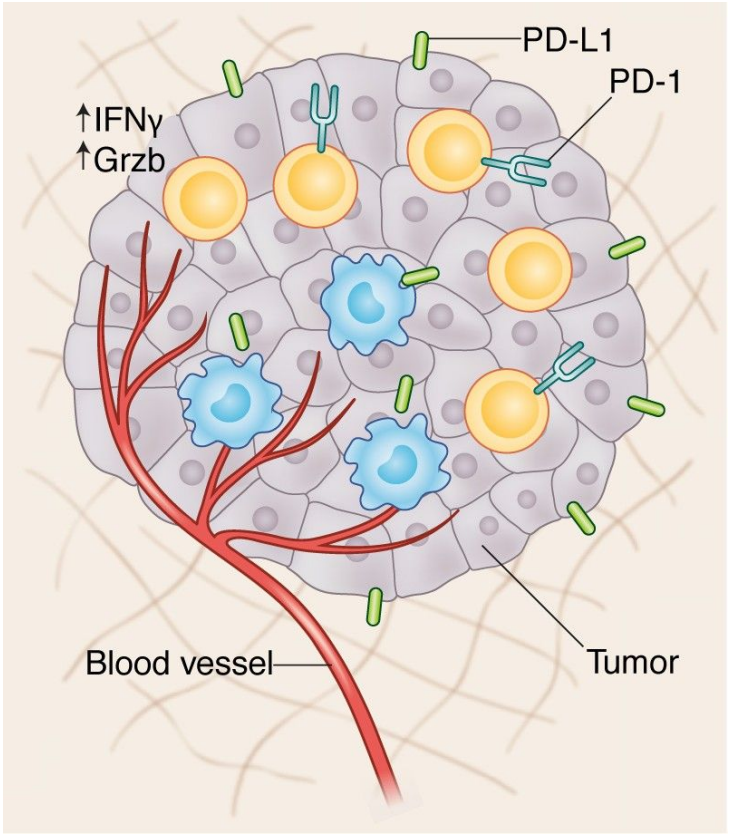






Cell composition: tumor microenvironment



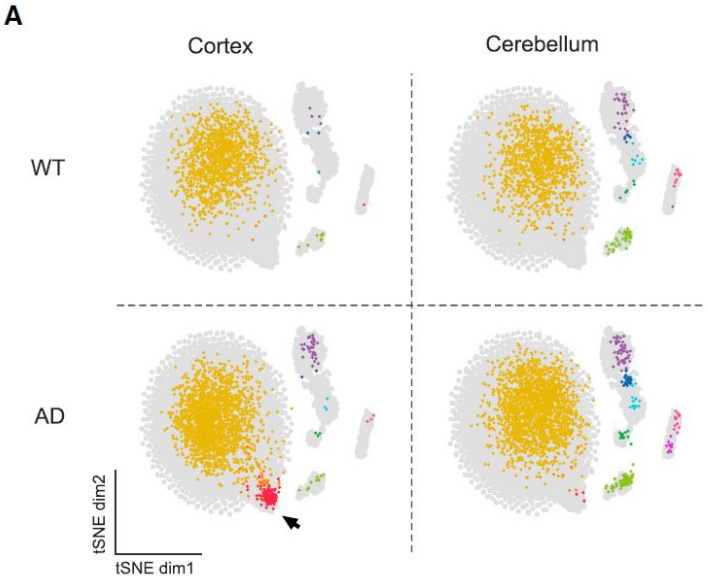


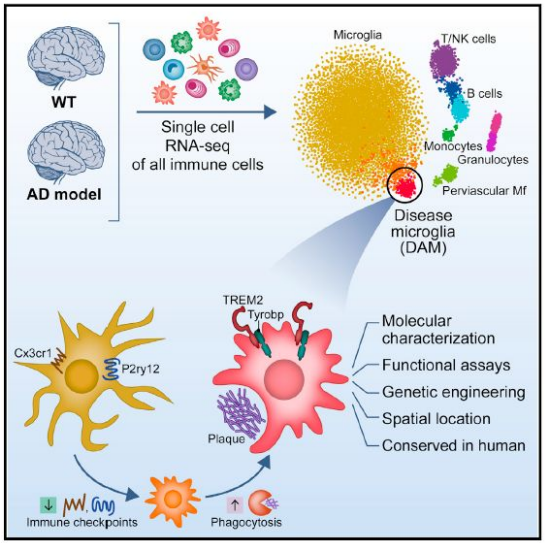
Motivation 2: Rare cell population, stem cell niches

Cell

A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease

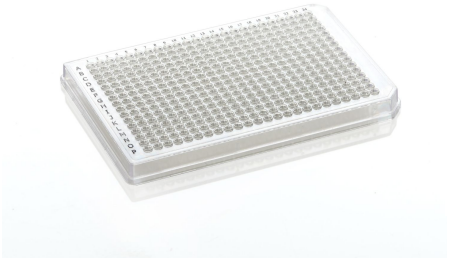
Hadas Keren-Shaul, Amit Spinrad, Assaf Weiner, ..., Marco Colonna, Michal Schwartz, Ido Amit



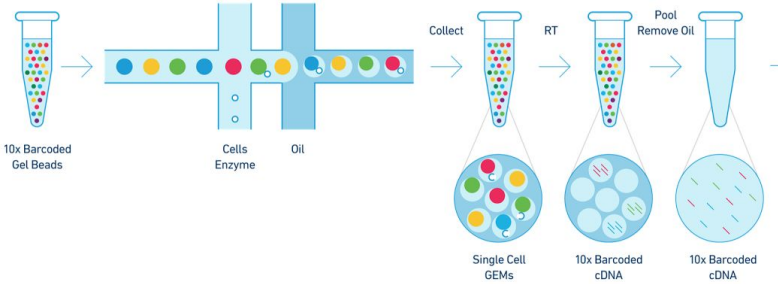


Methods of single-cell RNA-seq. How to barcode RNA from each individual cell

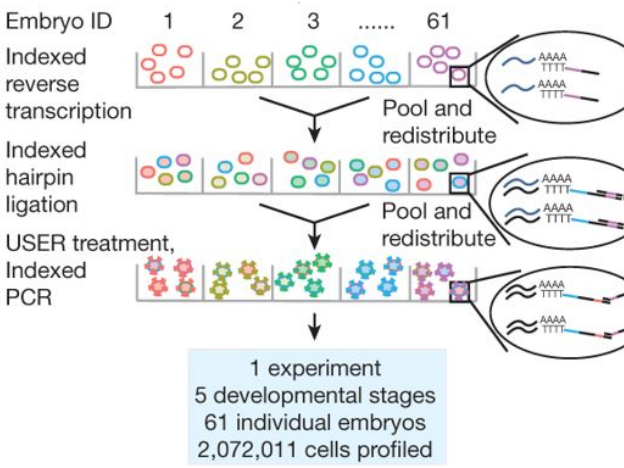
Smart-seq2
cell-sorting-based



10X
droplet-based



sci-RNA-seq
Combinatorial indexing



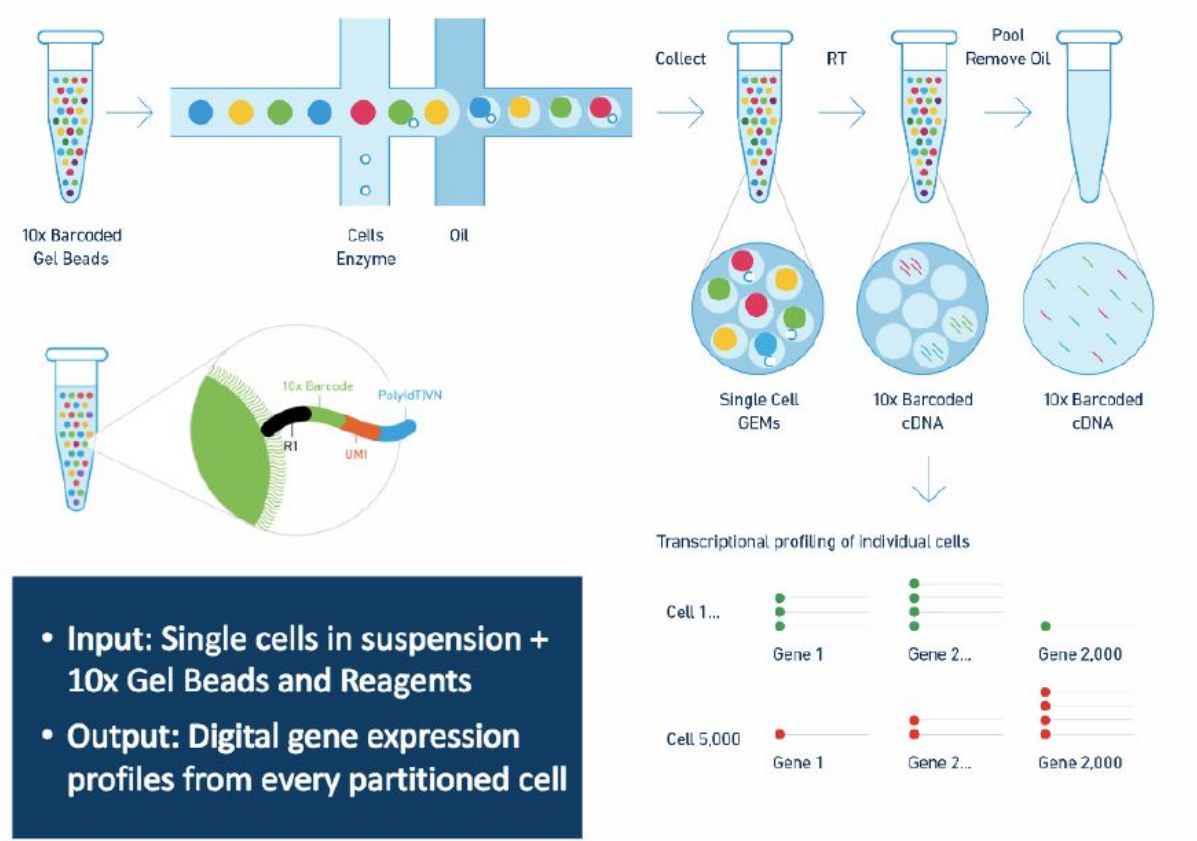
coverage

10K 1-3K 500

number of cells

384 6K-10K 1-2M

10X Chromium Single Cell 3′ Protocol



- Input: Single cells in suspension + 10x Gel Beads and Reagents
- Output: Digital gene expression profiles from every partitioned cell

Cell Preparation

↓

GEM Creation & Thermal Cycling

↓

Post Cycling Cleanup

↓

cDNA Amplification

↓

Library Preparation & Quantification

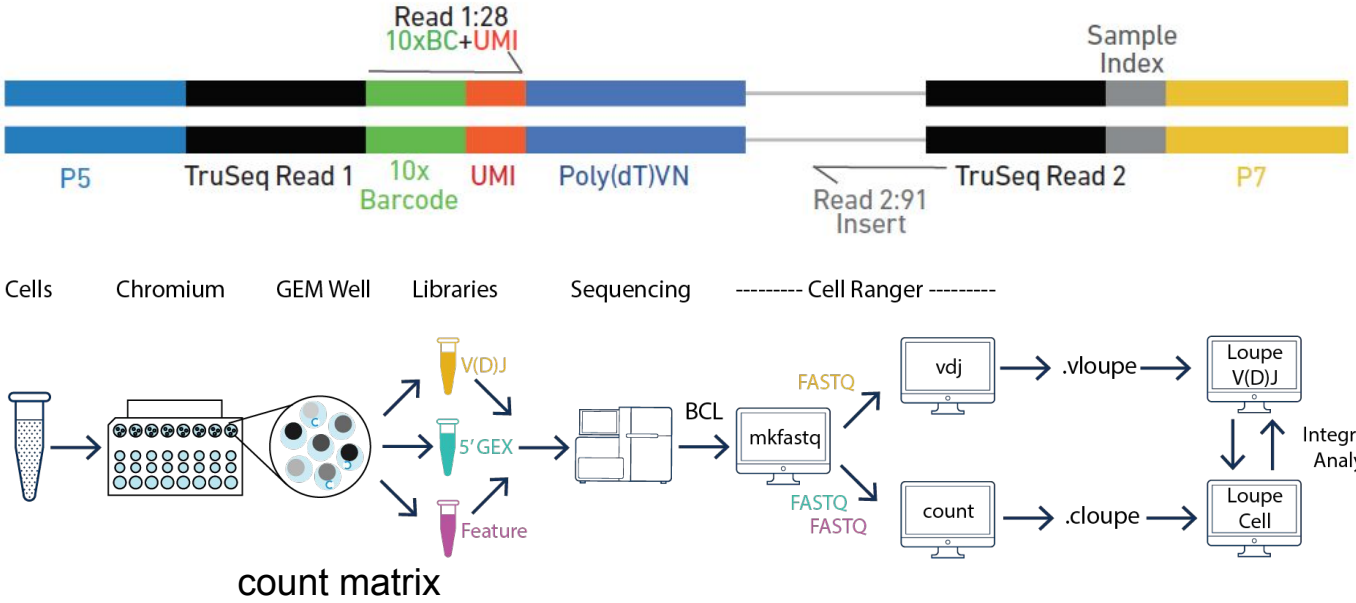
↓

Sequencing

Demultiplexing, mapping, counting

Automated for 10X:

CellRanger

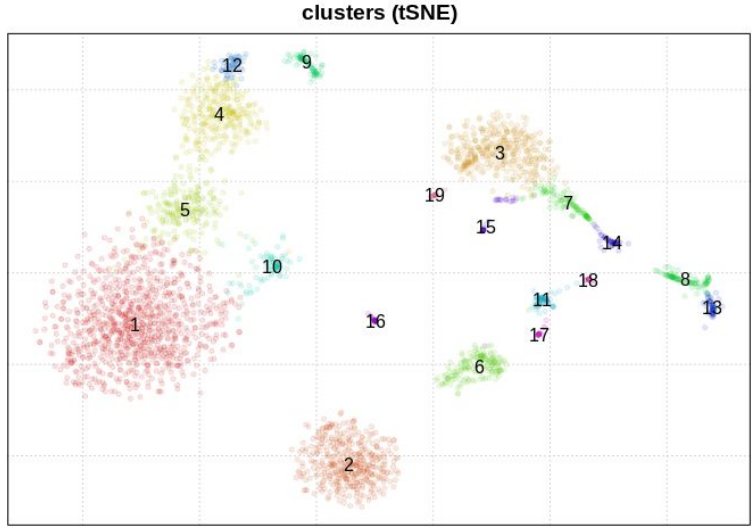



```
Xkr4     .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Rp1      .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Sox17    .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Mrpl15  28 11 44  8 22 44 46 45 112 30 20 18 25 27  . 45 15 14 31 46 53 84 32  9 23
Rgs20    .  .  .  .  .  .  7  .  .  8  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Npbwr1   .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
```

Visualization / embedding

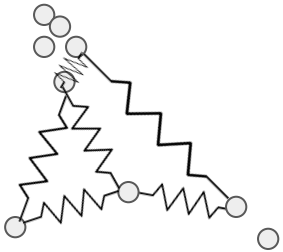
Similar cells together in 2D or 3D

PCA, tSNE, UMAP



count matrix

```
Xkr4     .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Rp1      .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Sox17    .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Mrpl15  28 11 44  8 22 44 46 45 112 30 20 18 25 27  . 45 15 14 31 46 53 84 32  9 23
Rgs20    .  .  .  .  .  .  7  .  .  8  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Npbwr1   .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
```



tSNE

Analogy: cells connected by springs

Force ~ similarity / local density

keep data structure

avoid overcrowding

Typical workflow

`cellranger count`

- Align reads, filter out empty beads, assign reads to cells, count reads per gene per cell

`Seurat`

- Filter cells (low #reads, high % reads mapped to mitochondrial genome)
- (*) Filter doublets
- Find variable genes
- Regress out: coverage, %mitochondrial reads
- Reduce dimensions (i.e. group correlated genes): PCA
- UMAP: visualize
- Find clusters
- Find marker genes for each cluster

cellranger web_summary

Estimated Number of Cells

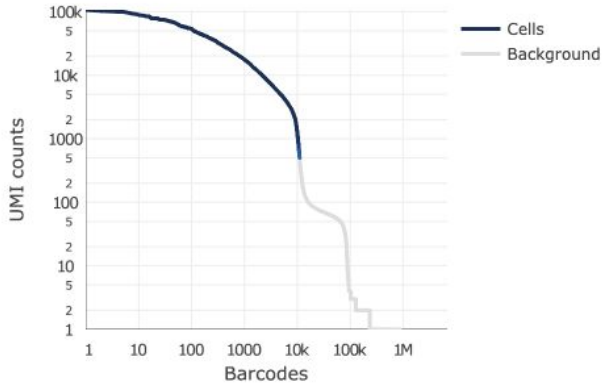
10,866

Mean Reads per Cell	Median Genes per Cell
18,588	1,936

Sequencing

Number of Reads	201,978,484
Valid Barcodes	93.8%
Sequencing Saturation	27.0%
Q30 Bases in Barcode	97.4%
Q30 Bases in RNA Read	86.6%
Q30 Bases in Sample Index	96.3%
Q30 Bases in UMI	97.3%

Cells



Estimated Number of Cells	10,866
Fraction Reads in Cells	93.0%
Mean Reads per Cell	18,588
Median Genes per Cell	1,936
Total Genes Detected	20,803
Median UMI Counts per Cell	4,470

web_summary.html

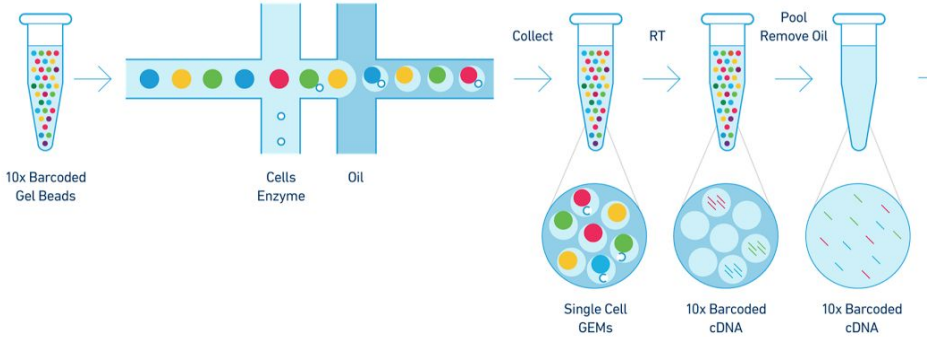
cells

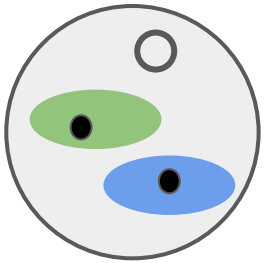
Empty beads

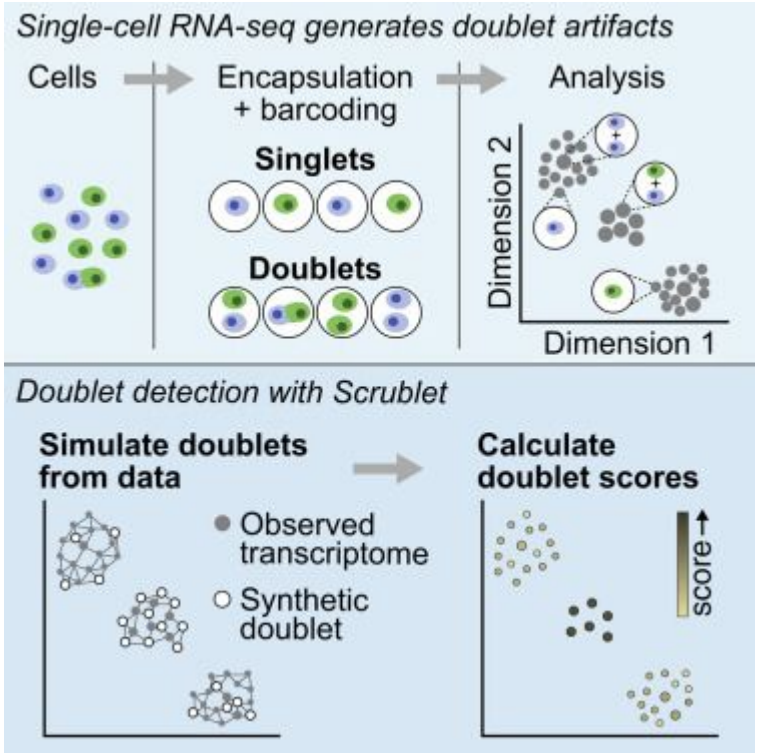
Genes per cell

...

Doublets

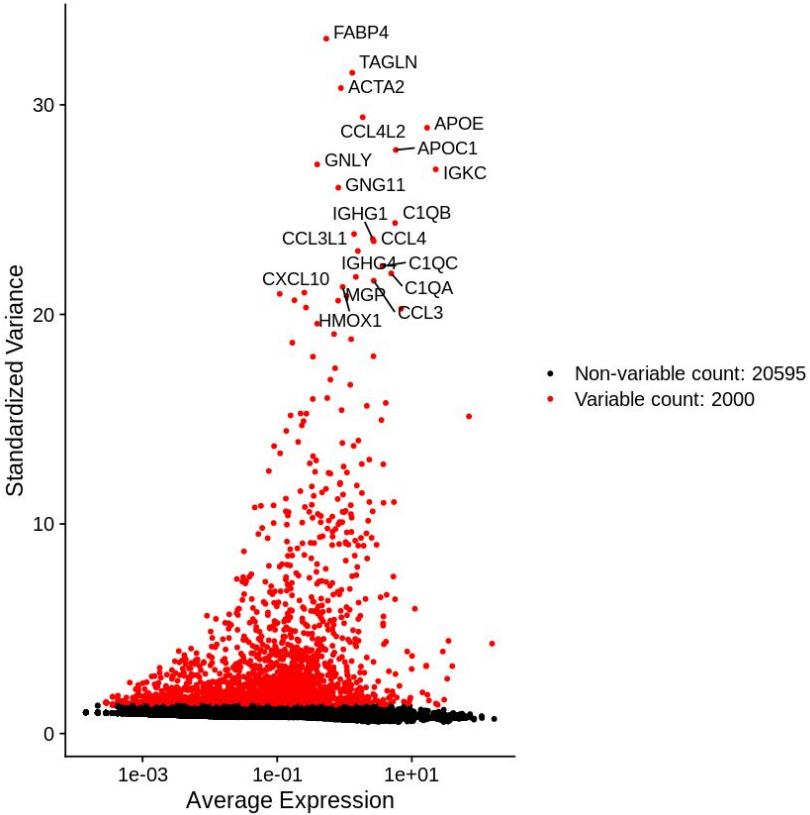




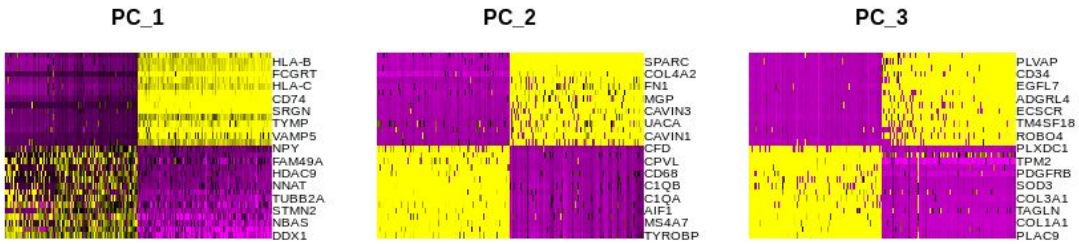


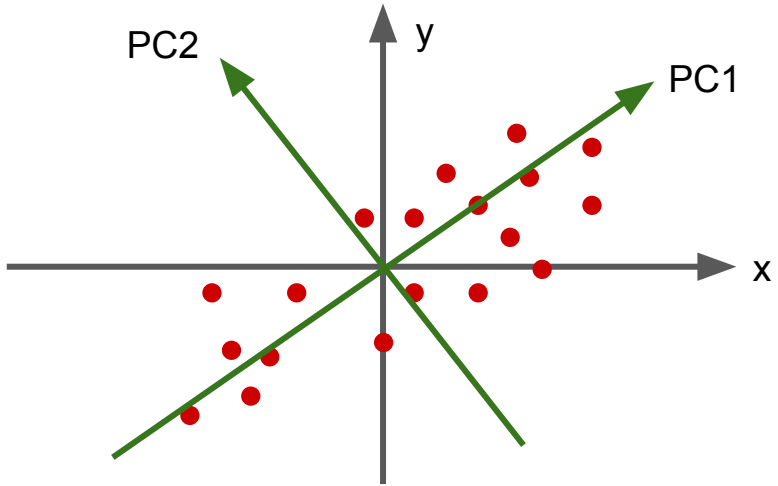
Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data

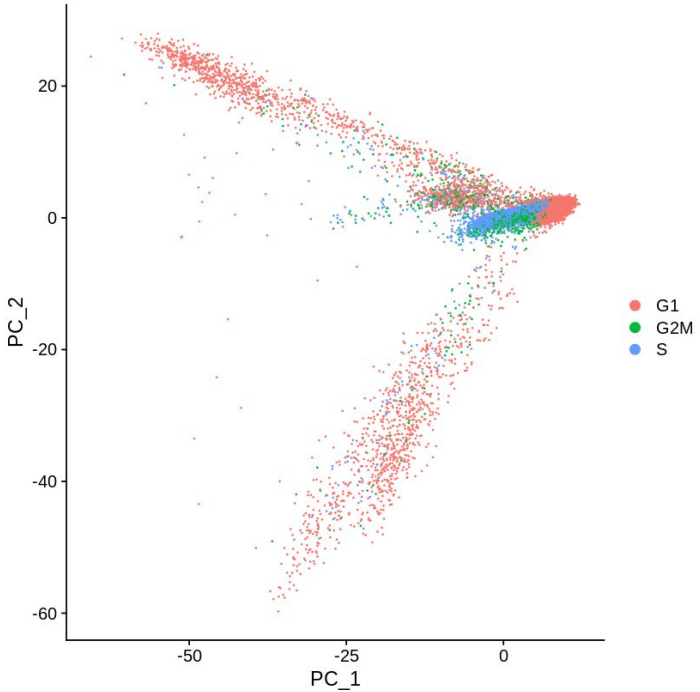
Variable genes



PCA: reduce dimensions

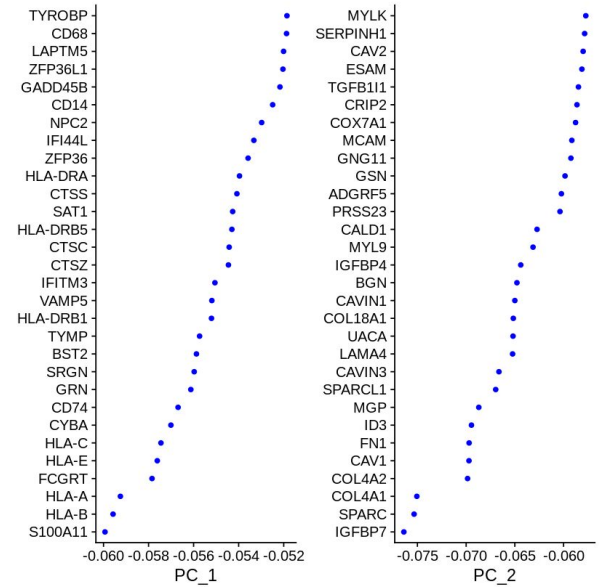


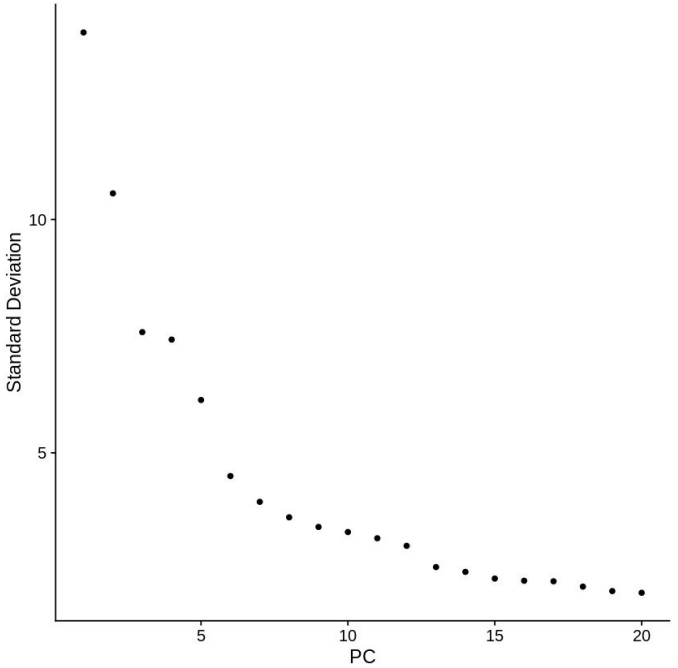




PCA: select first *N* components

```
SR <- RunPCA(SR, features = VariableFeatures(object = SR))
SR <- JackStraw(SR, num.replicate = 100)
SR <- ScoreJackStraw(SR, dims = 1:20)
plot(ElbowPlot(SR))
```



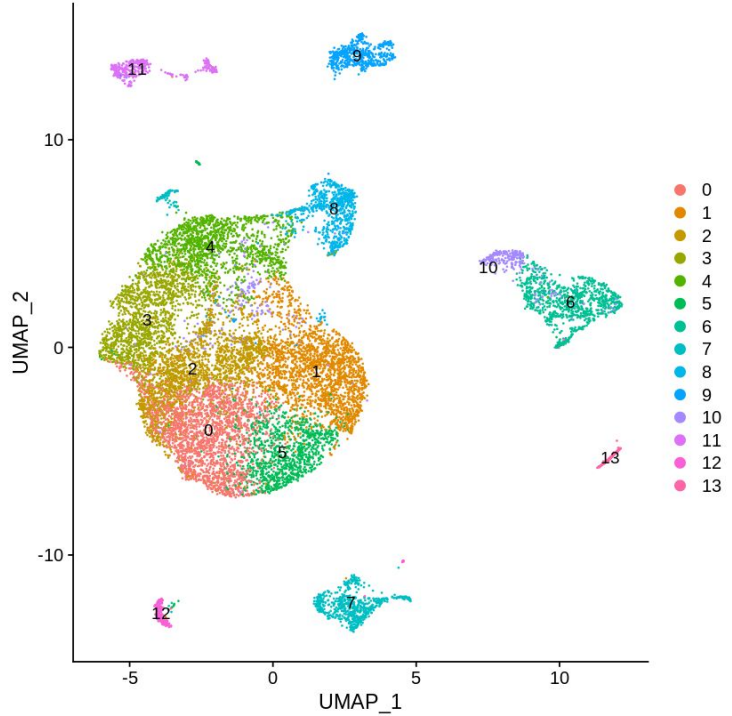


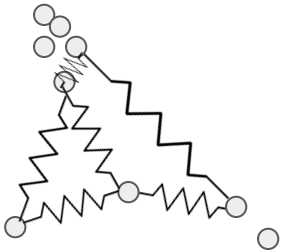
Embedding: PCA, tSNE, UMAP

Similar cells together in 2D or 3D

PCA, tSNE, UMAP

Xkr4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Rp1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sox17	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Mrpl15	28	11	44	8	22	44	46	45	112	30	20	18	25	27	.	45	15	14	31	46	53	84	32	9	23
Rgs20	.	.	.	.	.	.	7	.	.	8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Npbwr1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.





tSNE

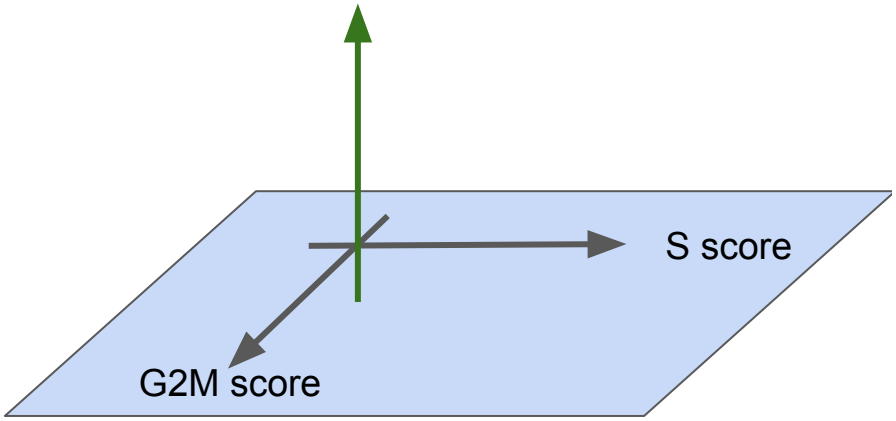
Analogy: cells connected by springs

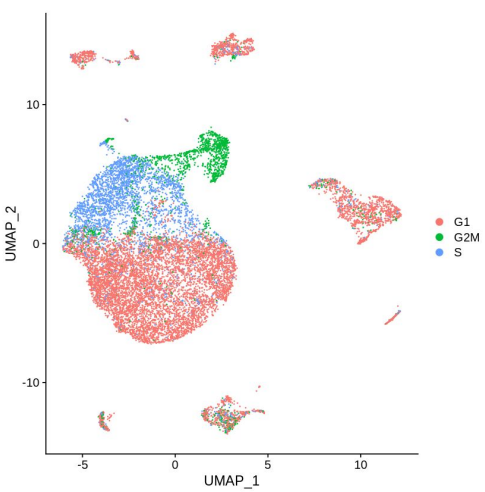
Force ~ similarity / local density

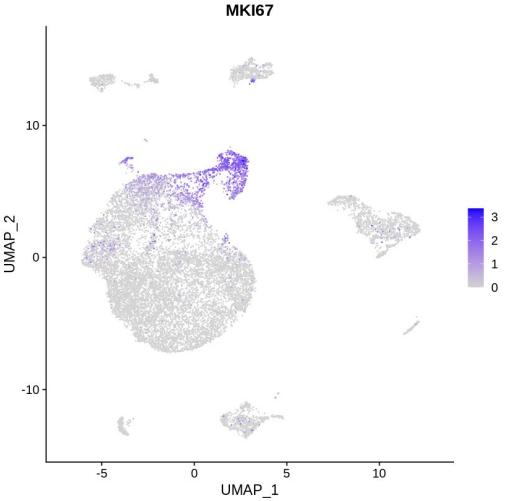
keep data structure

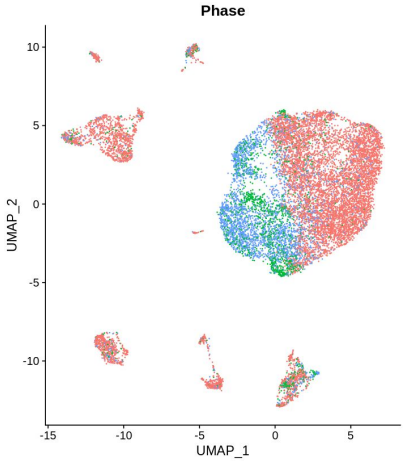
avoid overcrowding

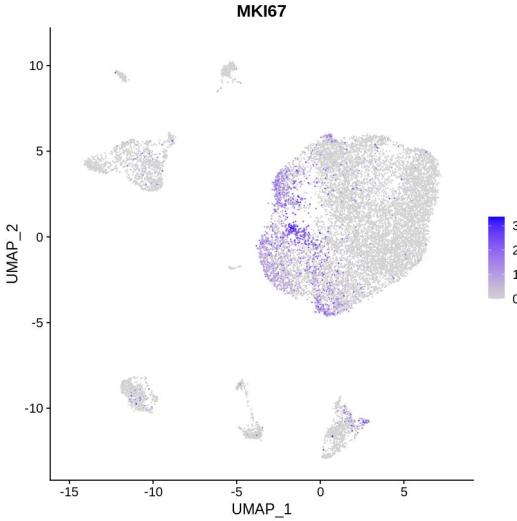
Regressing out the cell cycle





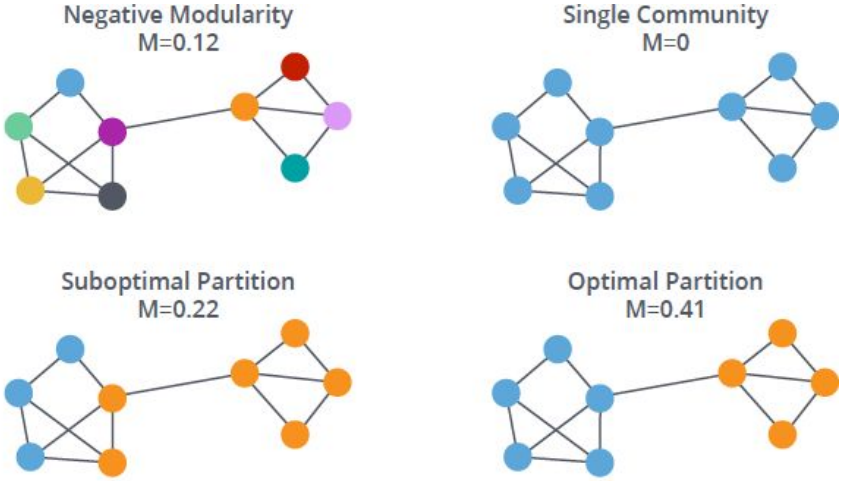






Extra topics

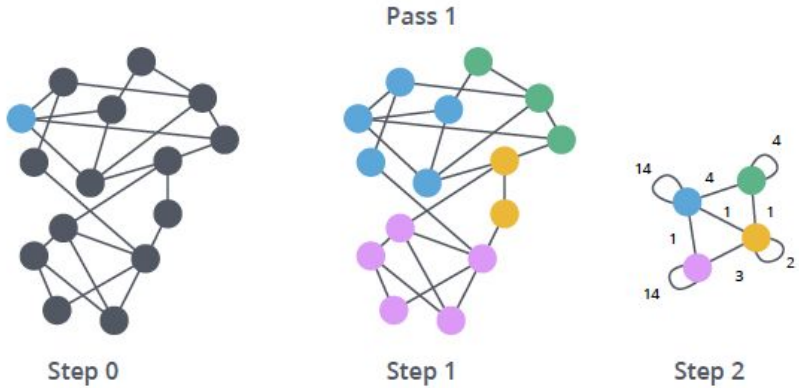
Graph-based clustering: Leuven



Modularity

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j),$$

modularity



Step 0

Choose a start node and calculate the change in modularity that would occur if that node joins and forms a community with each of its immediate neighbors.

Step 1

The start node joins the node with the highest modularity change. The process is repeated for each node with the above communities formed.

Step 2

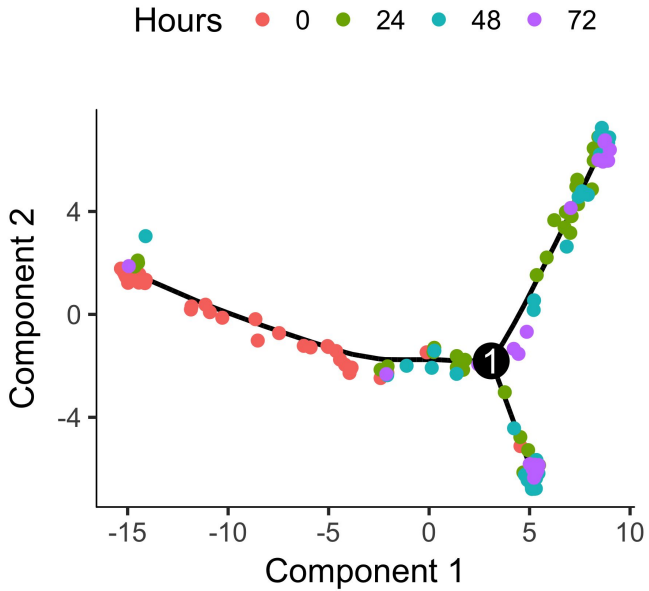
Communities are aggregated to create super communities and the relationships between these super nodes are weighted as a sum of previous links. (Self-loops represent the previous relationships now hidden in the super node.)

Trajectory / pseudotime analysis

Trajectory / pseudotime analysis

Minimal spanning tree,
Selecting reference point so that:
min:
(*Deviation of points from a tree*) +
λ * (*total tree length*)

$$\min_{G \in G_b} \min_{f_G \in \mathcal{F}} \min_{Z} \sum_{i=1}^{N} \| x_i - f_G(z_i) \|^2 + \frac{\lambda}{2} \sum_{(V_i, V_j) \in \varepsilon} b_{i,j} \| f_G(z_i) - f_G(z_j) \|^2$$

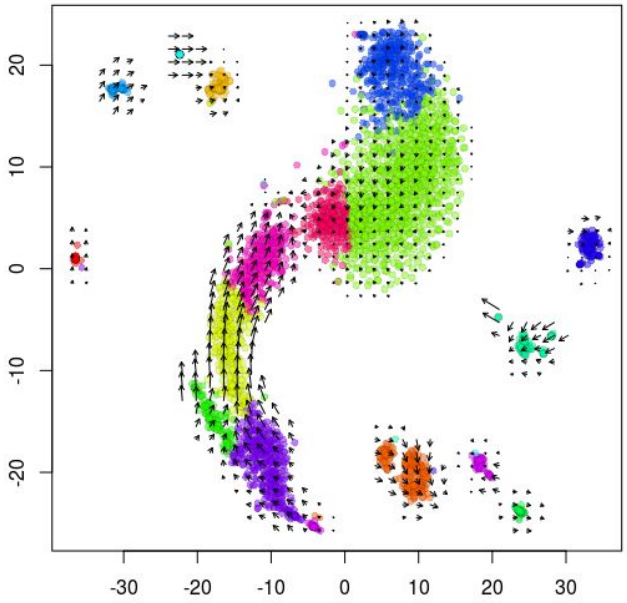


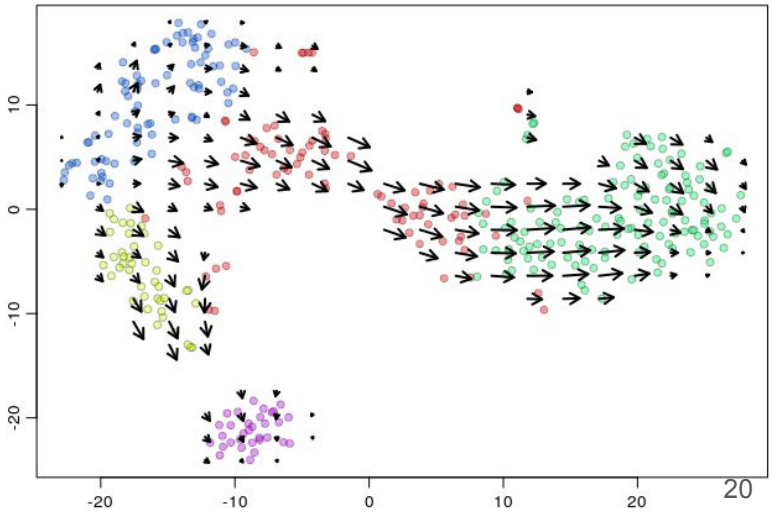
Qiu 2017. Reversed graph embedding resolves complex single-cell trajectories

RNA velocity: predicted future of each individual cell

Trajectory / pseudotime analysis

RNA velocity





la Manno 2018. RNA velocity of single cells

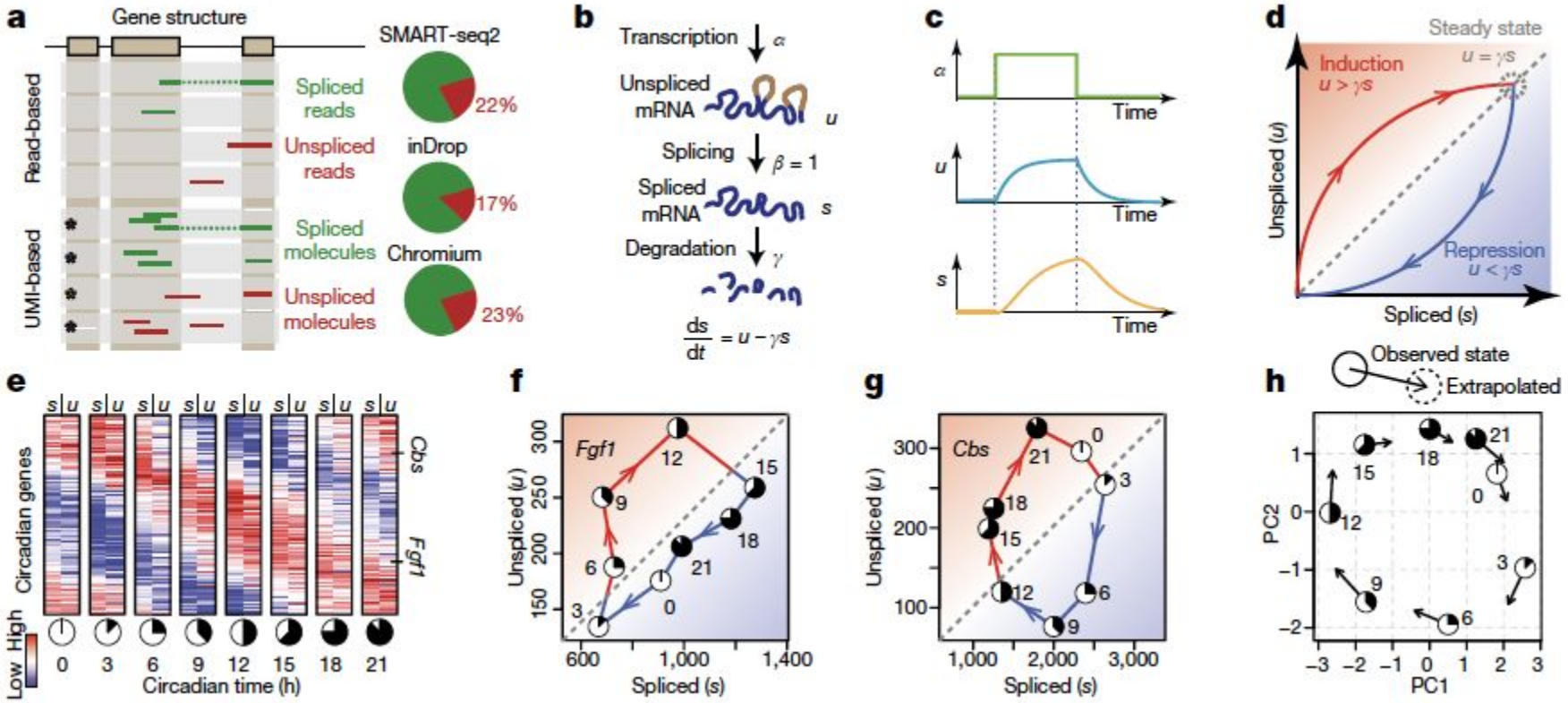
RNA velocity

$$\frac{du}{dt} = \alpha(t) - \beta(t)\,u(t)$$

$$\frac{ds}{dt} = \beta(t)\,u(t) - \gamma(t)s(t)$$

$$\frac{du}{dt} = \alpha - u(t)$$

$$\frac{ds}{dt} = u(t) - \gamma s(t)$$



la Manno, Soldatov 2018. RNA velocity of single cells

Label transfer: conos