

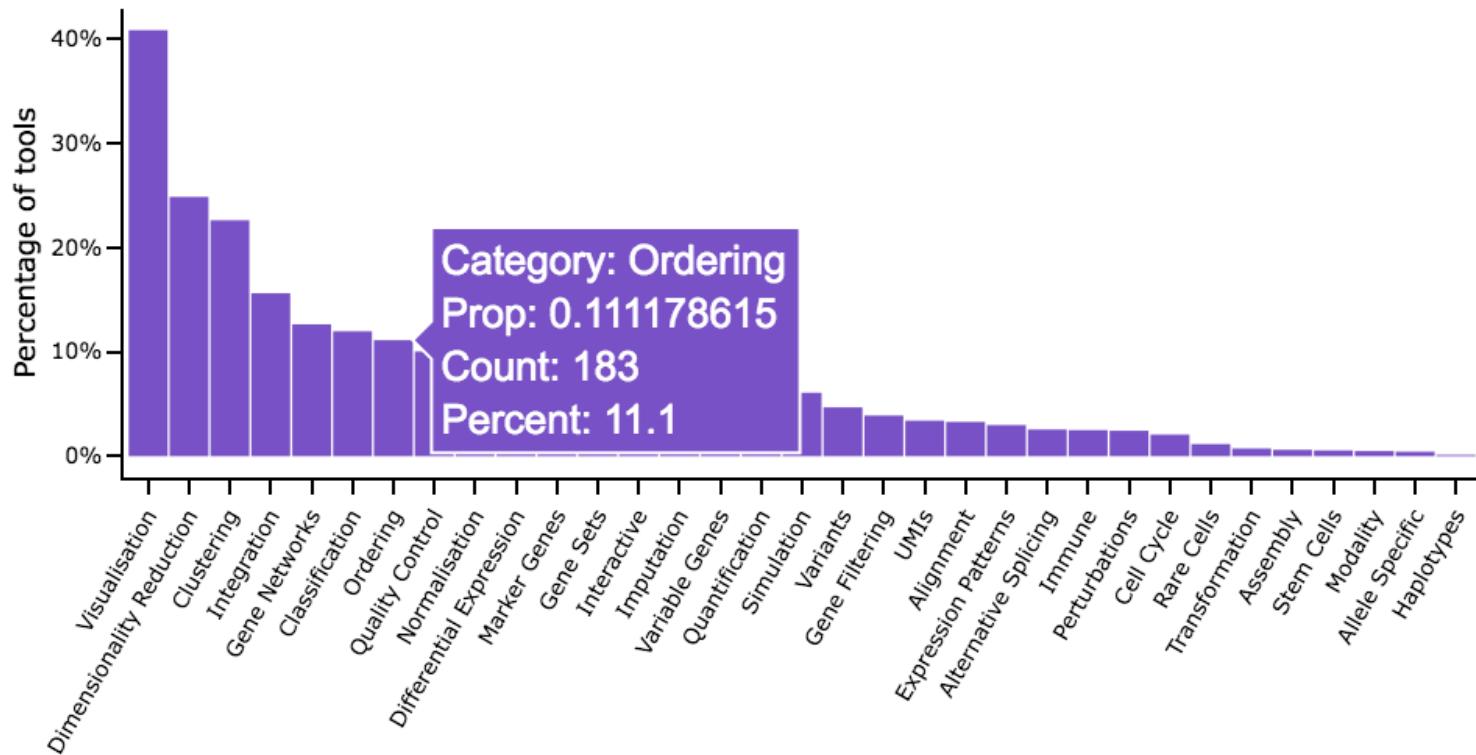
Post hoc analysis

# Possible interest after cell classification and DGE

- Trajectory Analysis
- Ordering of cells or clusters
- MetaCell analysis
- Cell-cell communication
- Transcription factor activity analysis
- Deep Neural Network, e.g. for biomarker discovery

# Trajectory/Ordering

# Trajectory analysis - Again a long list of possible tools...



<https://www.scrna-tools.org/>

# Trajectory analysis

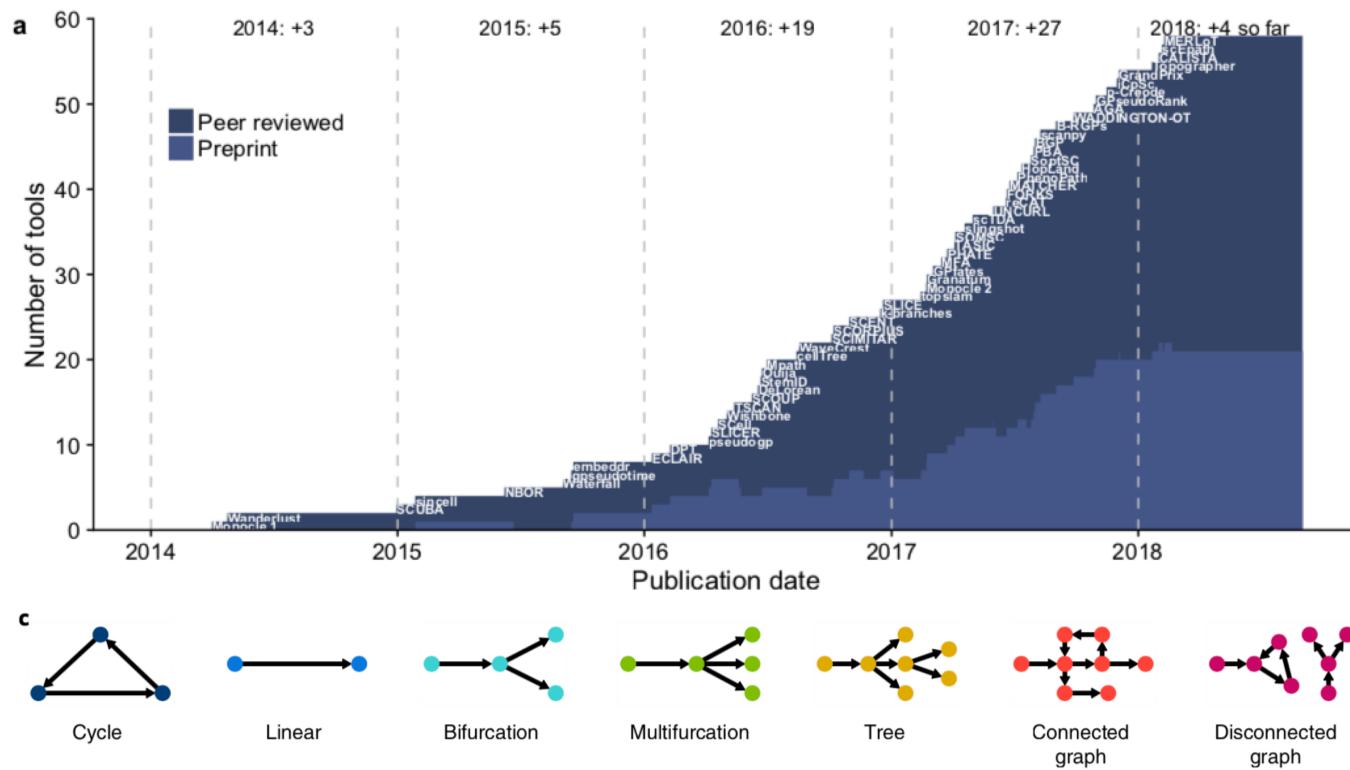
- Differences in gene expression between cells, might be attributed to dynamic processes:
  - Cell cycle
  - Cell differentiation
  - Response to an external stimuli
- Trajectory inference can order a set of individual cells along a path / trajectory / lineage
- Some methods project cells onto a **pseudotime axis** others project each cell along a path.
- This can be a starting point for further analysis to determine gene expression programs driving interesting cell phenotypes or changes in conditions that drive cells towards more or less differentiated states (Number of cells in the beginning vs end of the pseudotime axis).

# Should you run trajectory inference ?

- Are you sure that you have a developmental trajectory?
- Do you have intermediate states?
- Do you believe that you have branching in your trajectory?
- Do you have a time scale on your cells ?
- Do you have a starting state or an ending state ?

Be aware, any dataset can be forced into a trajectory without any biological meaning!

# Different types of trajectory analysis

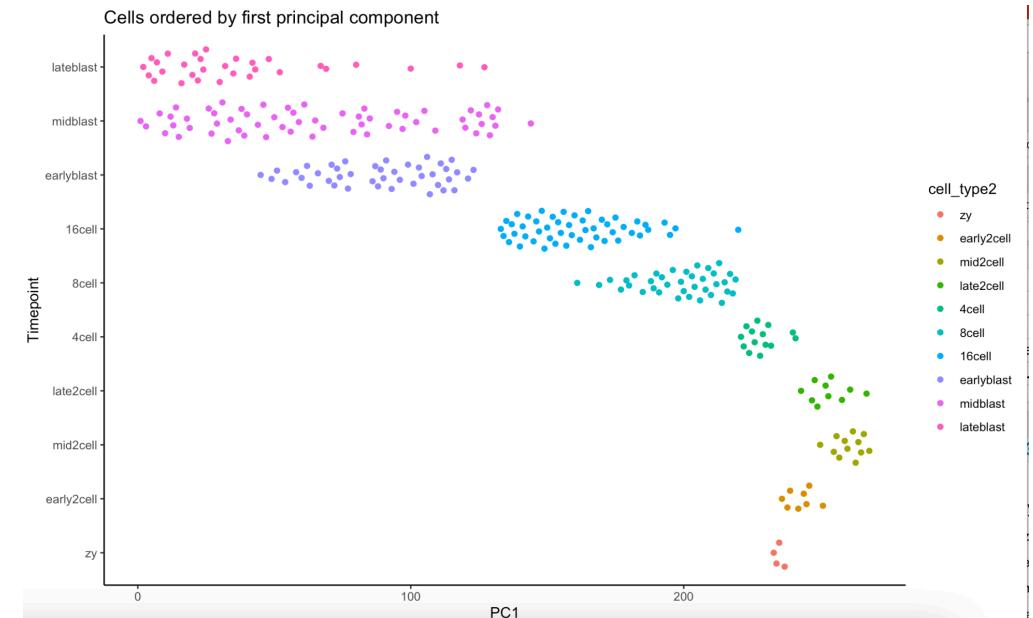
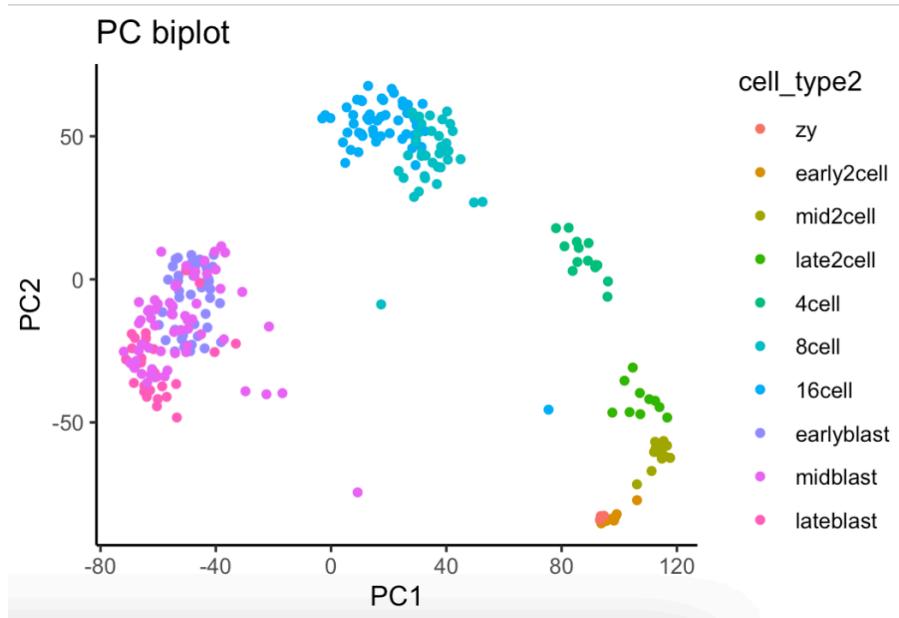


Saelens et al (2019) Nat Biotechnology

## Example of application

- From the paper Single-Cell RNA-Seq Reveals Dynamic, Random Monoallelic Gene Expression in Mammalian Cells (Deng et al. 2014)
- « To investigate allele-specific gene expression at single-cell resolution, we isolated 269 individual cells dissociated from *in vivo* F1 embryos [...] from oocyte to blastocyst stages of mouse preimplantation development»
- Here finding a trajectory makes biological sense.

# A trajectory could be to order the cells according to PC1

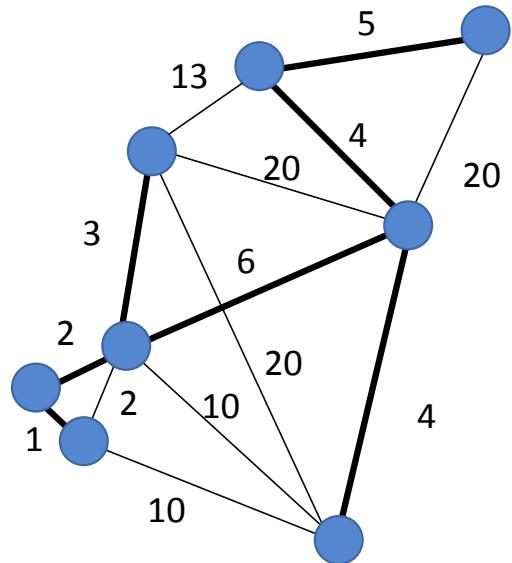


But can we do better ?

## Standard process of trajectory analyses

- Take a weighted graph
- Take a spanning tree
- Take the minimum of all spanning tree.

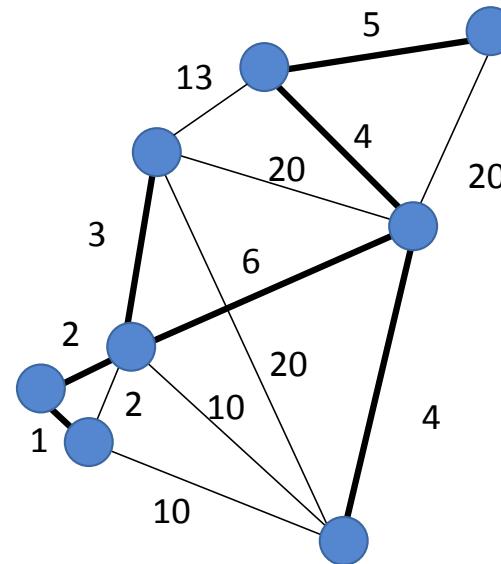
## Example of a weighted graph



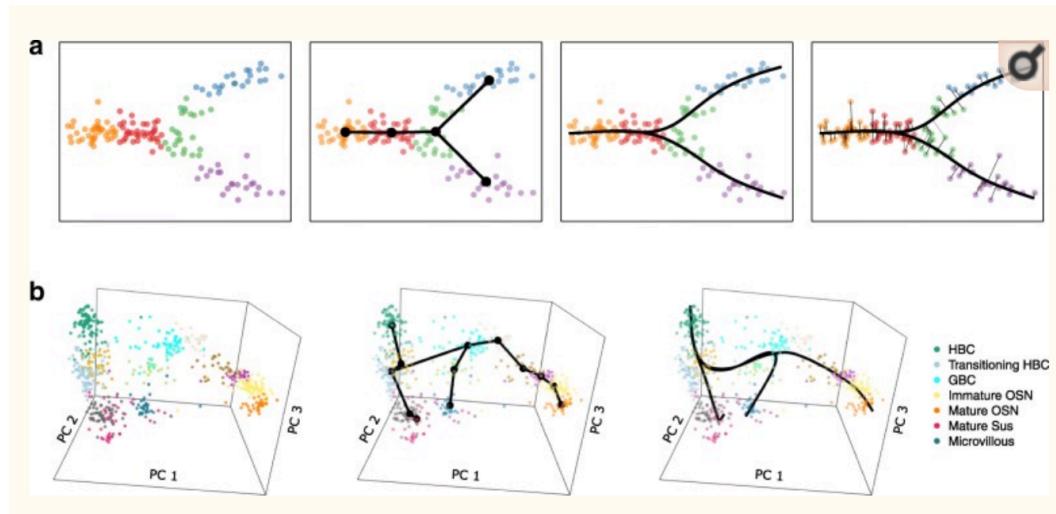
**Bold edges:**  
1 way to connect the cells,  
i.e. one of the spanning trees

# Minimum spanning tree (MST)

- **Sum of all distances in the tree (graph) is at its minimum**
- Having more transitional cells improves the definition of the tree
- The weights can be a distance in the dimensionality reduction space (ICA, T-SNE, UMAP, diffusion maps) or a correlation between cells, etc.
- MST has no cycles, cell cycles will not work in here



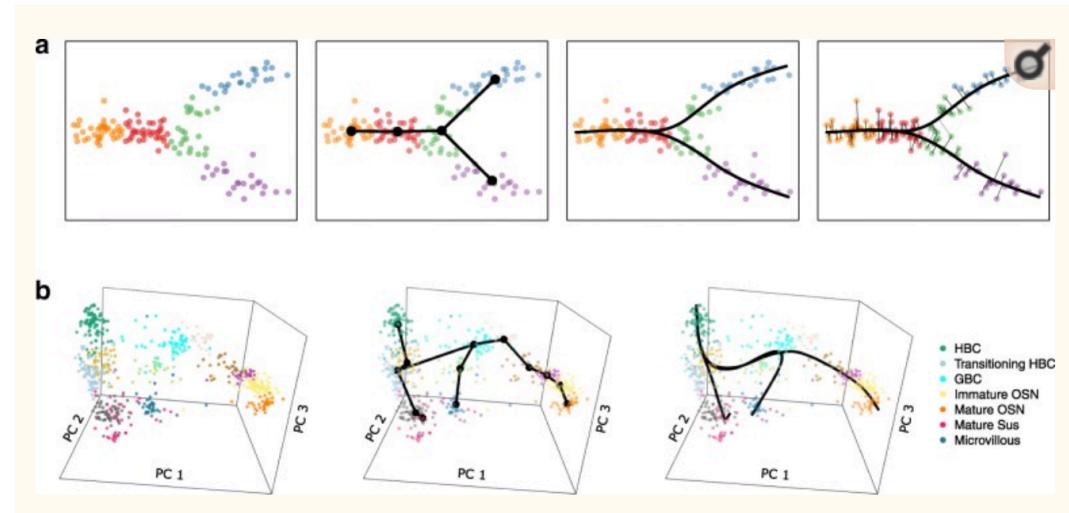
# Slingshot (Street et al 2018)



1. Distance between clusters

$$d^2(\mathcal{C}_i, \mathcal{C}_j) \equiv (\bar{X}_i - \bar{X}_j)^T (S_i + S_j)^{-1} (\bar{X}_i - \bar{X}_j),$$

# Slingshot (Street et al 2018)



1. Distance between clusters
2. Infer lineages by ordering cell clusters and construct MST
3. Construct principal curves\*

\*Principal curves are smooth one-dimensional curves that pass through the middle of a p-dimensional data set, providing a nonlinear summary of the data. They are nonparametric, and their shape is suggested by the data

# Slingshot on Deng data

- ```
sce <-  
  slingshot::slingshot(deng_SCE  
  , clusterLabels =  
  'Seurat_clusters', reducedDim  
  = 'PCA', start.clus = "2")
```

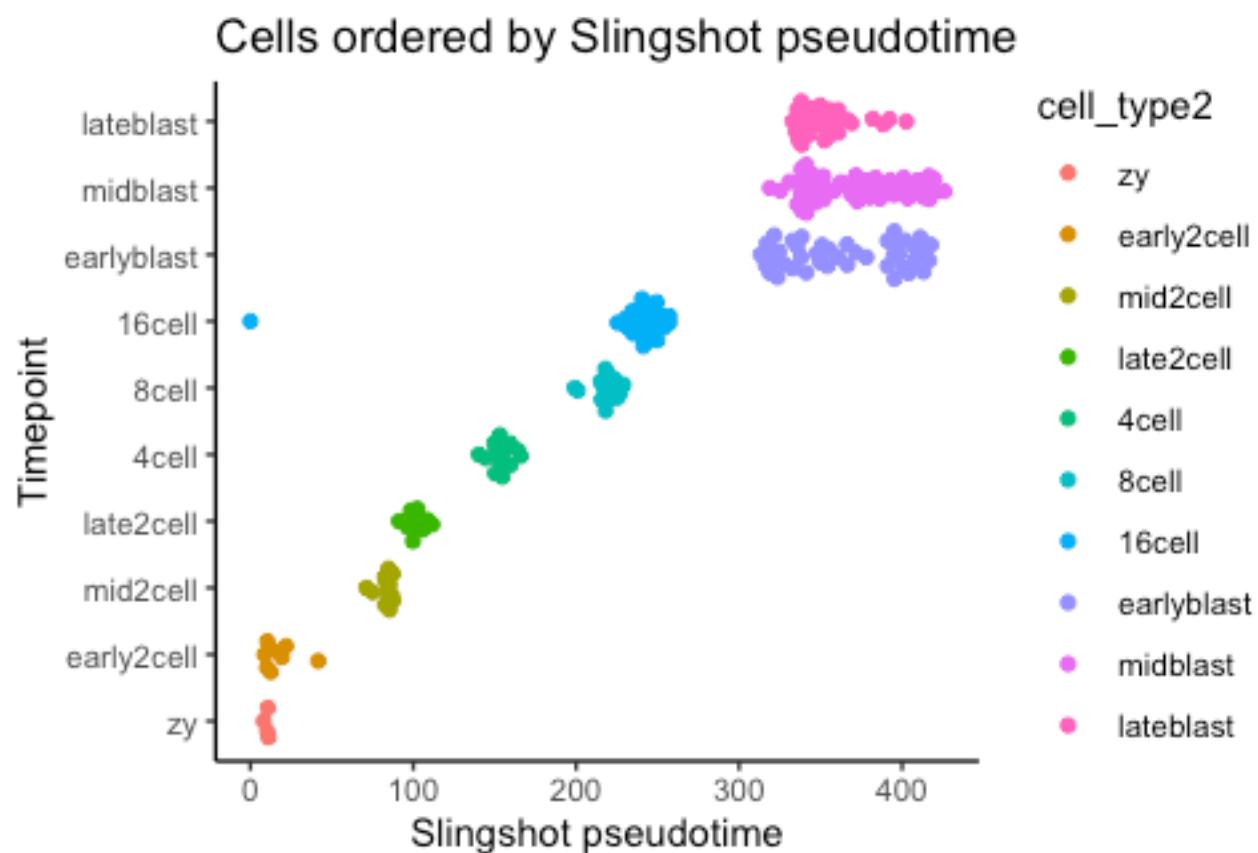
```
> SlingshotDataSet(sce)  
class: SlingshotDataSet
```

| Samples | Dimensions |
|---------|------------|
| 268     | 50         |

| lineages: | 2         |
|-----------|-----------|
| Lineage1: | 2 4 0 5 3 |
| Lineage2: | 2 4 1     |

| curves: | 2                              |
|---------|--------------------------------|
| Curve1: | Length: 425.93 Samples: 234.68 |
| Curve2: | Length: 340.91 Samples: 132.37 |

# Slingshot on Deng data



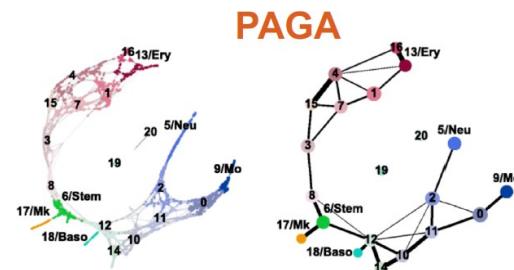
## Reverse Graph Embedding (RGE), DDRTree

- MST relies on each cell, and changes in their distance will change significantly the trajectory outcome
- RGE is clustering prior to MST so you can take the average of the cells in a particular cluster (Monoclev2-v3)

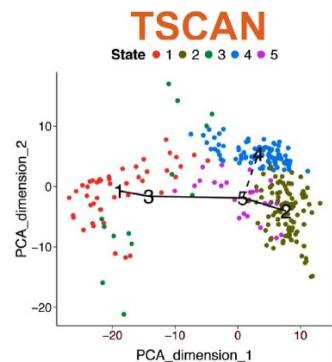
## Monocle3 uses an algorithm based on PAGA (python)

- **PAGA** constructs a **k-nearest neighbour** graph on cells and then identifies ‘communities’ of cells via the Louvain method.
- Two vertices (**Louvain communities**) are linked with an edge, when the cells in the respective communities are neighbours in the *k*-nearest neighbour graph.
- **Monocle 3** constructs a *k*-nearest neighbour graph (*k* = 20) on cells in the UMAP space, then grouping them into Louvain communities, and testing each pair of communities for a **significant number** of links between their respective cells.
- Those communities that have more links than expected under the null hypothesis of spurious linkage (FDR <1%) remain connected in the PAGA graph, and those links that fail this test are severed (correction of **spurious linkage**)

# Some additional tools

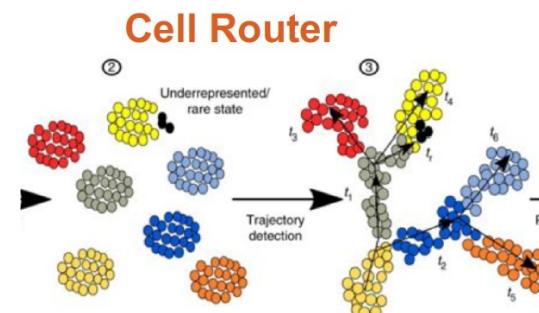


Street et al (2019) Genome Biology



Zhicheng et al (2016) Nuc Acid Res

Spade, StemID 2, Eclair, TSCAN and Mpath use different clustering algorithms such as k-means, k-medoids, hierarchical clustering or DBSCAN in a dimensionality-reduced space, **Psupertime** for time-series data.



Da Rocha et al (2018) Nat Commun

# RNA Velocity: quite a different algorithm

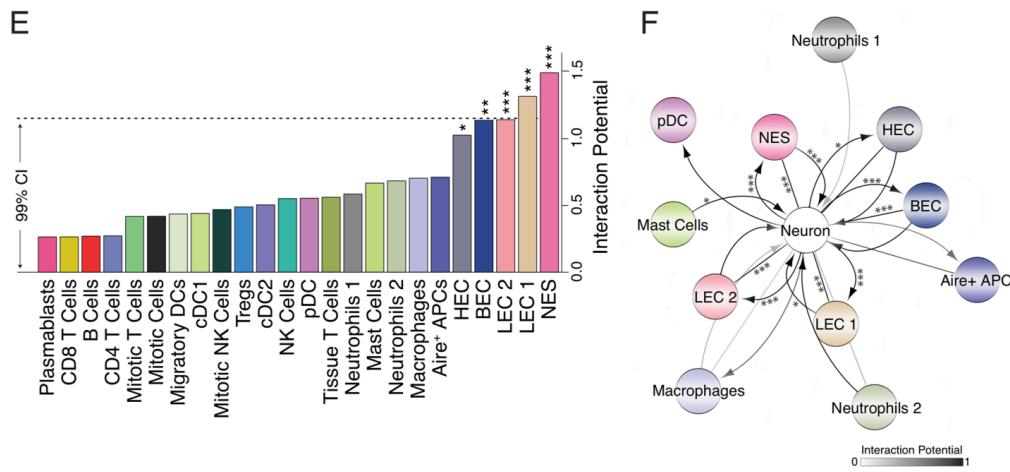
- « RNA velocity is a high-dimensional vector that **predicts** the future state of individual cells on a timescale of hours»
- «aid the analysis of **developmental lineages** and **cellular dynamics**»
- Method : calculate the relative abundance of nascent (unspliced) and mature (spliced) mRNA to estimate the rates of gene splicing and degradation
- During a dynamic process:
  - increase in the transcription rate=> rapid increase in unspliced mRNA=> increase in spliced mRNA until a new steady state is reached.
  - a drop in the rate of transcription => drop in unspliced mRNA => reduction in spliced mRNAs.
- During induction of gene expression: => unspliced mRNAs are present in excess
- During repression: => unspliced mRNAs are present in lower amounts.
- Hence: The balance of unspliced and spliced mRNA abundance is, therefore, an indicator of the future state of mature mRNA abundance, and thus the future state of the cell.

Gioele La Manno (2018) Nature

# Cell-cell communication

# Cell-cell communications

- LRIP – Ramilowski et al, Nat comm, 2015- bioarxiv, Huang et al, bioarxiv
- CellphoneDB - <https://www.cellphonedb.org/> - online « clickable » Mirjana Efremova, Nat protocols, 2020.
- CellChat- <http://www.cellchat.org/>
- NicheNet – needs apriori knowledge, Robin Browaeys, Nat met, 2020.

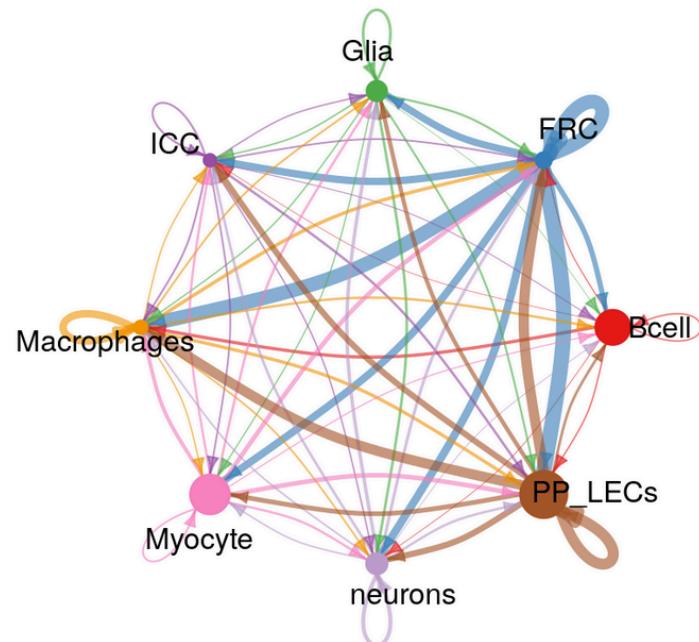


Ligand-receptor interaction potential between LN cells and neurons: LRIP  
Huang et al, Bioarxiv, <https://doi.org/10.1101/833509>

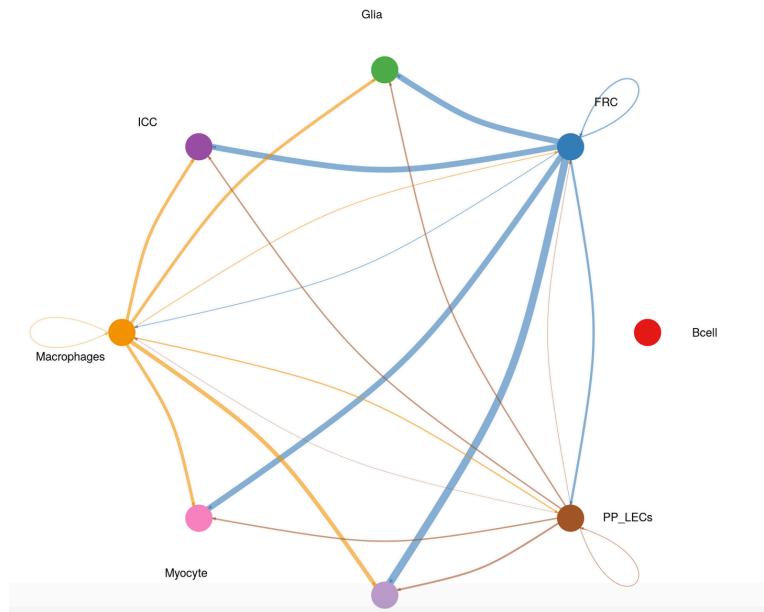
# CellChat

CellChat quantifies the signaling communication probability between two cell groups.  
Can be used for both scRNAseq and spatial transcriptomics.

Number of ligand-receptor interactions



IGF signaling pathway network



<https://github.com/jinworks/CellChat>

<https://doi.org/10.1101/2023.11.05.565674> BioRxiv update Nov 5 2023, CellChat v2

Liana

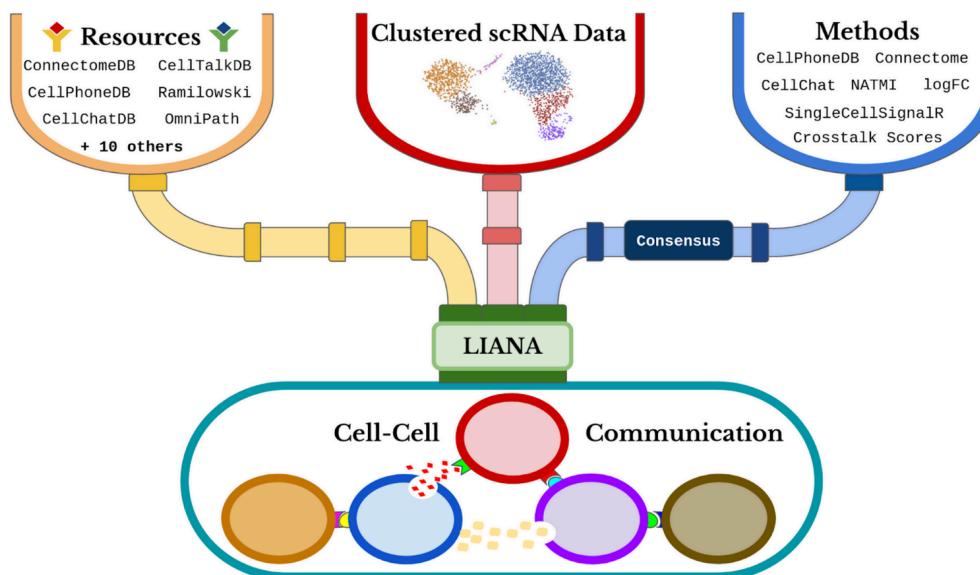


# LIANA: a Ligand-receptor ANalysis frAmework



LIANA enables the use of any combination of ligand-receptor methods and resources, and their consensus. A faster and memory efficient Python implementation is available [here](#).

Importance of the L-R resource



# NicheNet- Ligand receptor analysis

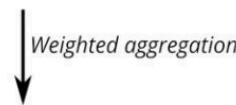
- Question : In your analysis you have a certain list of differentially expressed genes. Can one associate a pair of Ligand and Receptor responsible for the change in expression observed?
- This is extremely useful as it will point biologist to possible pathways to target to block the changes observed and which cell types communicate more or less in response to a condition.

**NicheNet infers active ligand–target links between interacting cells by combining their expression data with a prior knowledge model on ligand–target links**

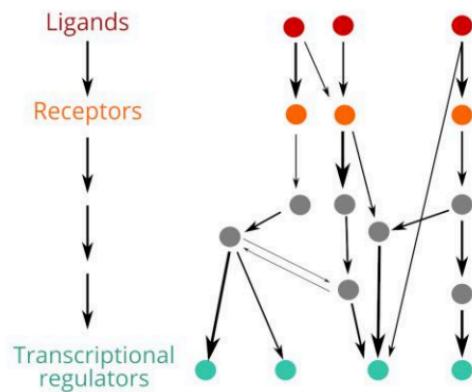
**Ligand-receptor and signaling data sources**

- Ligand-receptor DBs (e.g. Guide2Pharma)
- PPI (e.g. InWeb\_InBioMap)
- PTM (e.g. PhosphoSite)
- Text Mining (e.g. EVEX)
- Pathways (e.g. PathwayCommons)

*Weighted aggregation*



**Integrated ligand-signaling network**

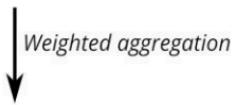


Weighted adjacency matrix  
ligand-signaling network

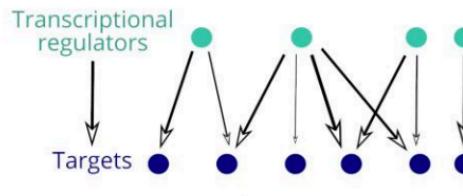
**Gene regulatory data sources**

- ChIP-seq (e.g. ENCODE)
- Text Mining (e.g. EVEX)
- Pathways (e.g. PathwayCommons)
- Motifs (e.g. TRANSFAC)
- Perturbations (e.g. TF KO GEO)

*Weighted aggregation*

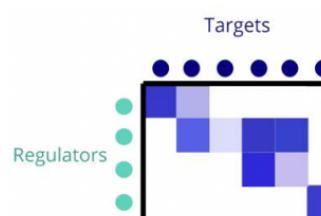
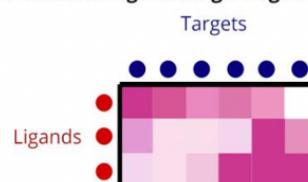


**Integrated gene regulatory network**



Weighted adjacency matrix  
gene regulatory network

**Prior model of ligand-target regulatory potential**



# NicheNet- Ligand receptor analysis-How it works

Prior model of ligand-target regulatory potential

|          | Target_1 | Target_2 | Target_3 | ... | Target_n |
|----------|----------|----------|----------|-----|----------|
| Ligand_1 | P11      | P12      | P13      | ... | P1n      |
| Ligand_2 | P21      |          |          |     |          |
| ...      | ...      |          |          |     |          |
| Ligand_m | Pm1      | Pm2      | Pm3      | ... | Pmn      |

# NicheNet- Ligand receptor analysis-How it works

Prior model of ligand-target regulatory potential

Calculate correlation scores

|          | Target_1 | Target_2 | Target_3 | ... | Target_n |
|----------|----------|----------|----------|-----|----------|
| Ligand_1 | P11      | P12      | P13      | ... | P1n      |
| Ligand_2 | P21      |          |          |     |          |
| ...      | ...      |          |          |     |          |
| Ligand_m | Pm1      | Pm2      | Pm3      | ... | Pmn      |

Select highest scores (scores are not meant to be so high due to high number of 0s)

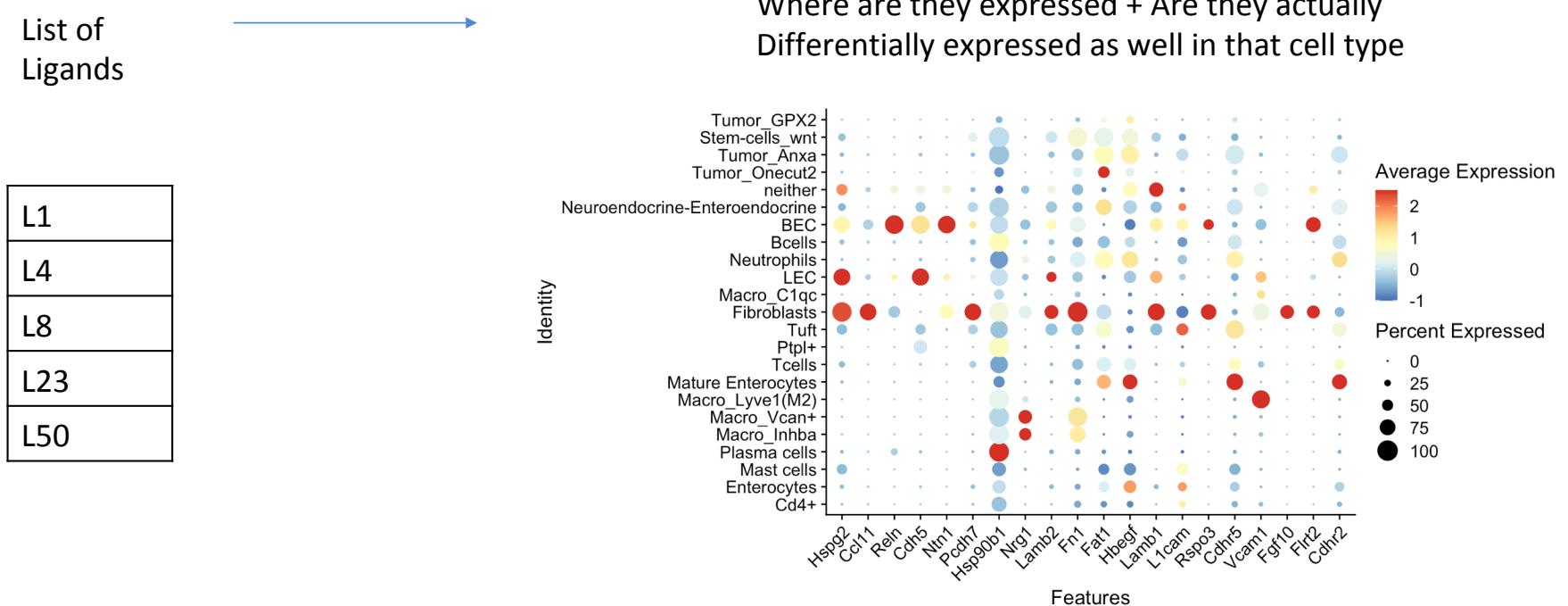
|     | Target_1 | Target_2 | Target_3 | ... | Target_n |
|-----|----------|----------|----------|-----|----------|
| DGE | 1        | 1        | 0        |     | 1        |

Select only expressed ligands

# Assumptions and calculation

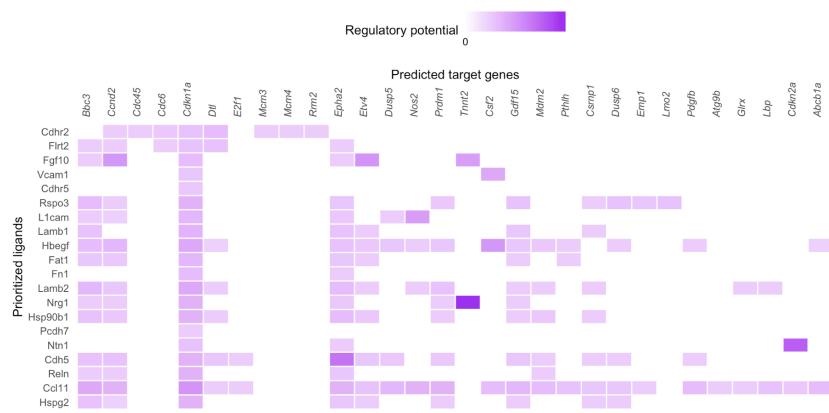
- How likely ligands in sender cells affected the gene expression in interacting receiver cells.
- Under the assumption that true downstream target genes of a ligand will be **differentially expressed after treatment** of cells with this ligand
- **Ligand activity scores :**
  - AUROC (area under receiver operating characteristic curve)
  - AUPR (area under the precision–recall curve)
  - **Pearson correlation**
  - Spearman's rank correlation
  - ...

# NicheNet- Ligand receptor analysis-How it works

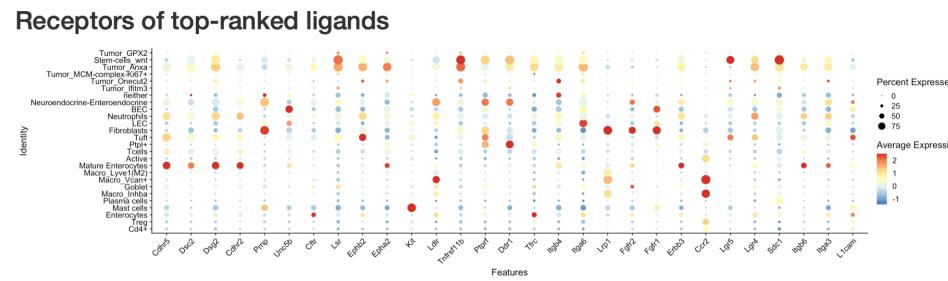


# NicheNet- Ligand receptor analysis-How it works

List of Receptors associated to the potential ligands



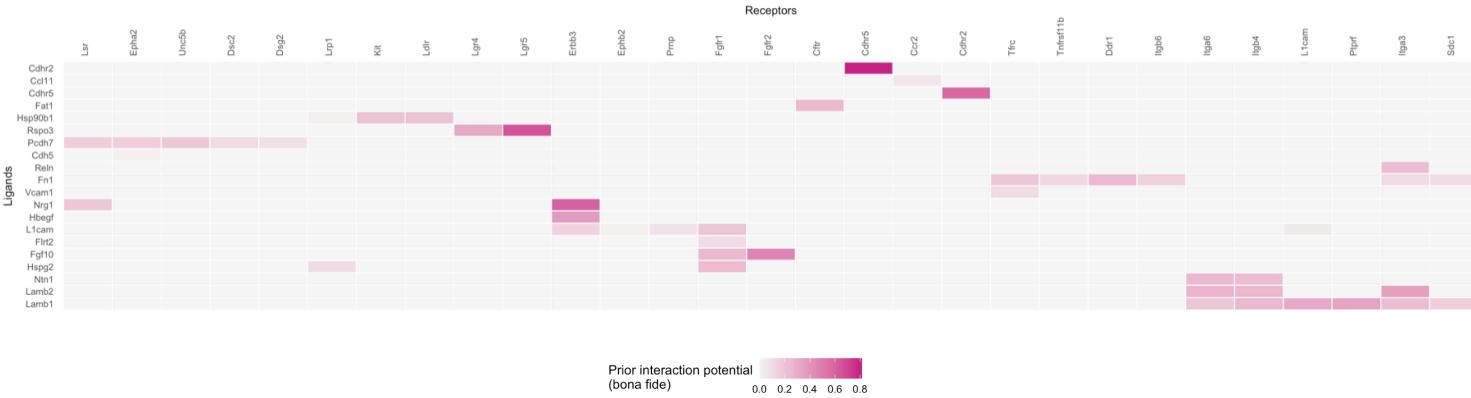
Where are they expressed + Are they actually Differentially expressed as well in that cell type



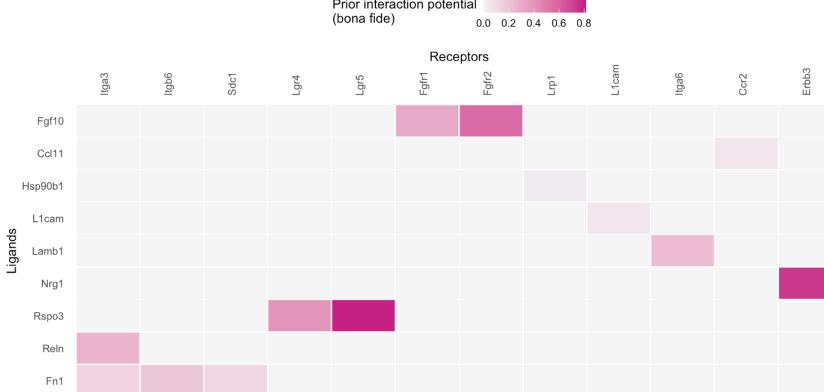
\*Courtesy Amber Bowler

# NicheNet- Ligand receptor analysis-How it works

List of Ligand  
Receptors with  
interaction  
potential



Bona fide



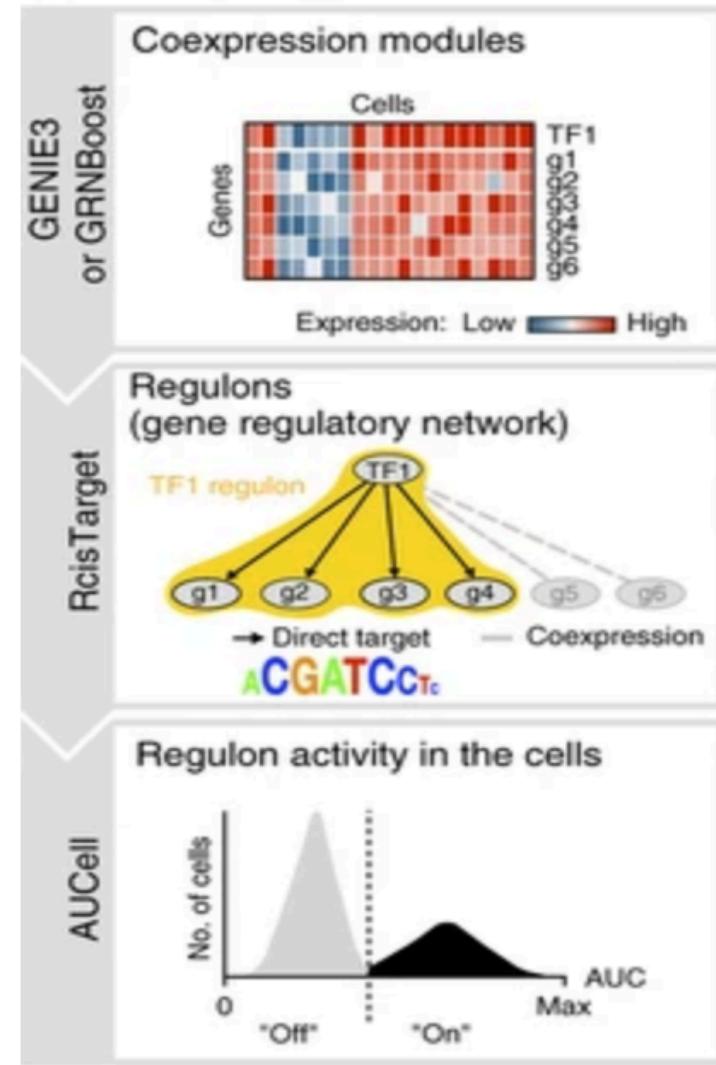
\*Courtesy Amber Bowler

# TF activity inference

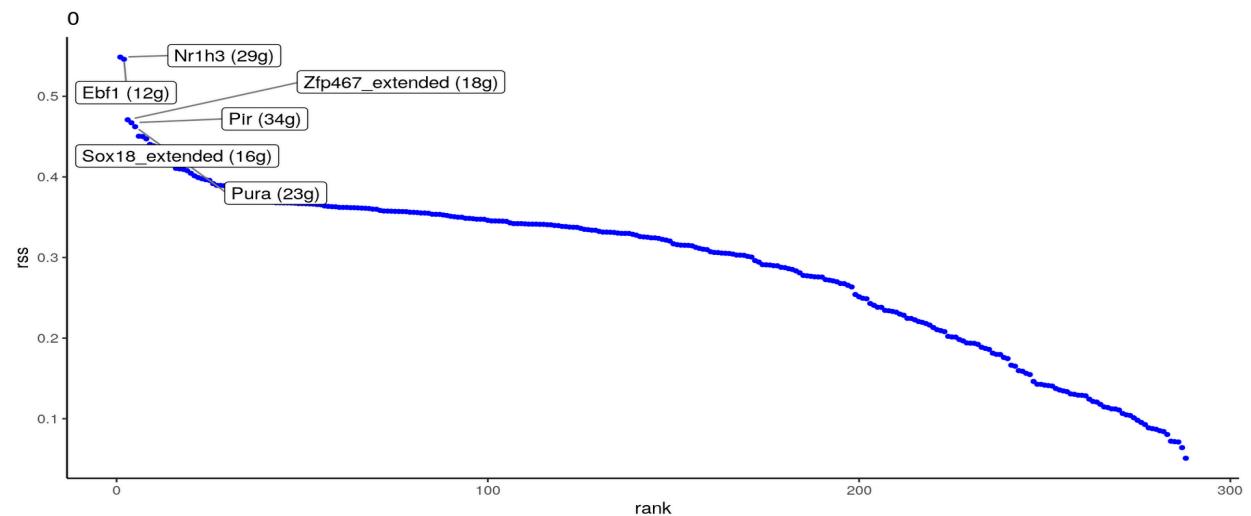
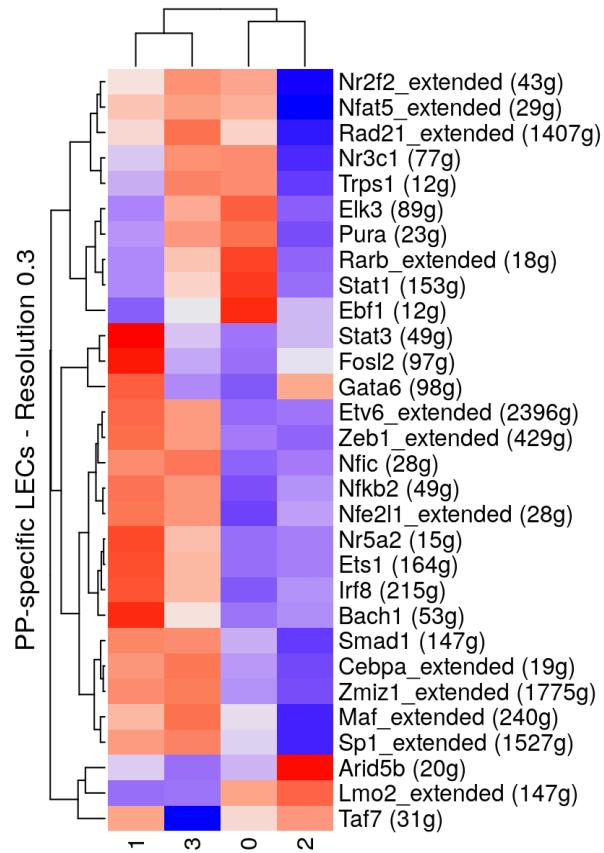
**SCENIC:** single-cell regulatory network inference and clustering

- **GENIE3:** GEne Network Inference with Ensemble of trees (modified)
- **RcisTarget**
- **AUCell**

R implementation is computationally slow!  
Use python implementation



# Which TFs are active within cell clusters ?



Regulon specificity score (RSS) rank plot

Suo et al, Cell Reports, 2018

Question on trajectory analysis