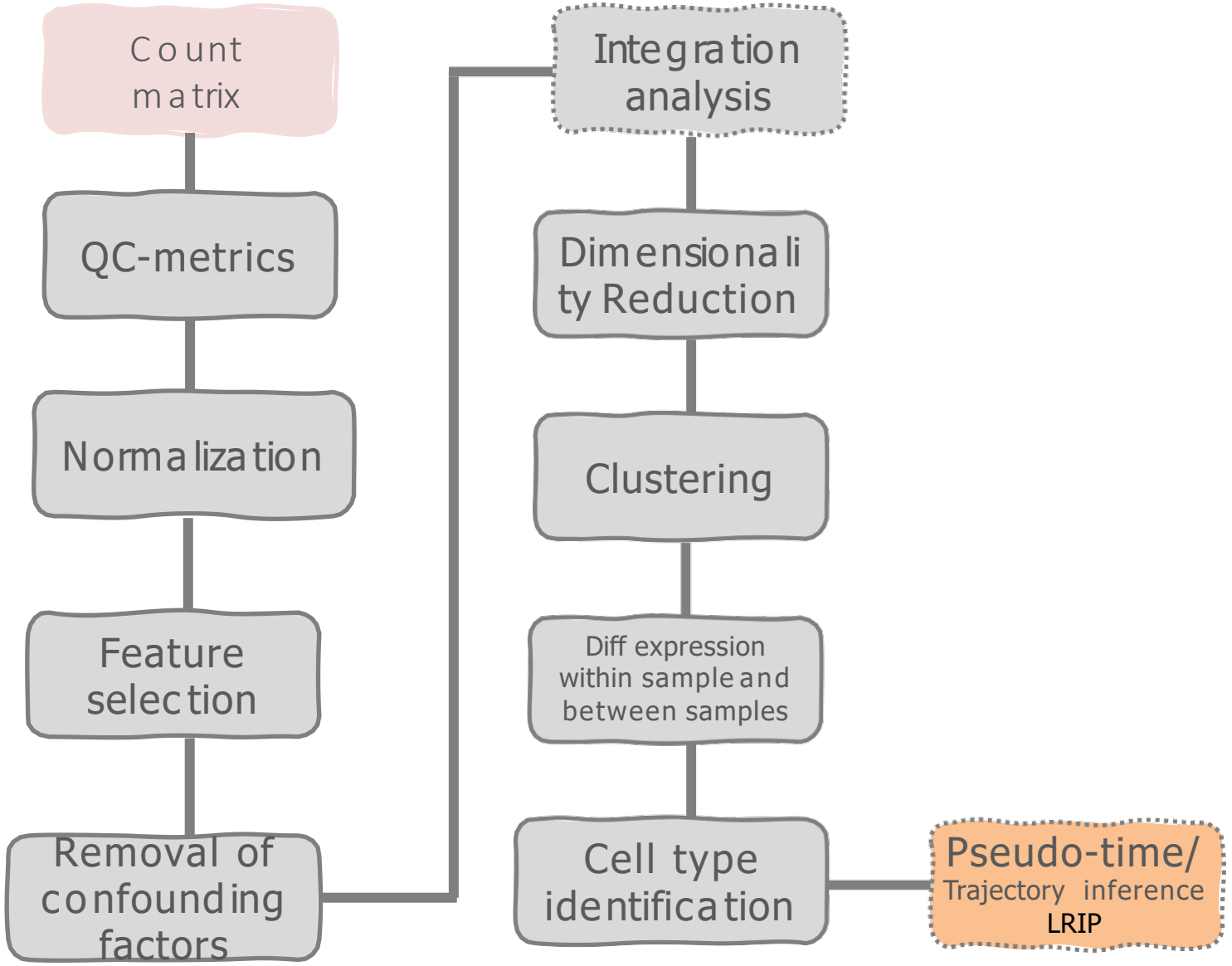
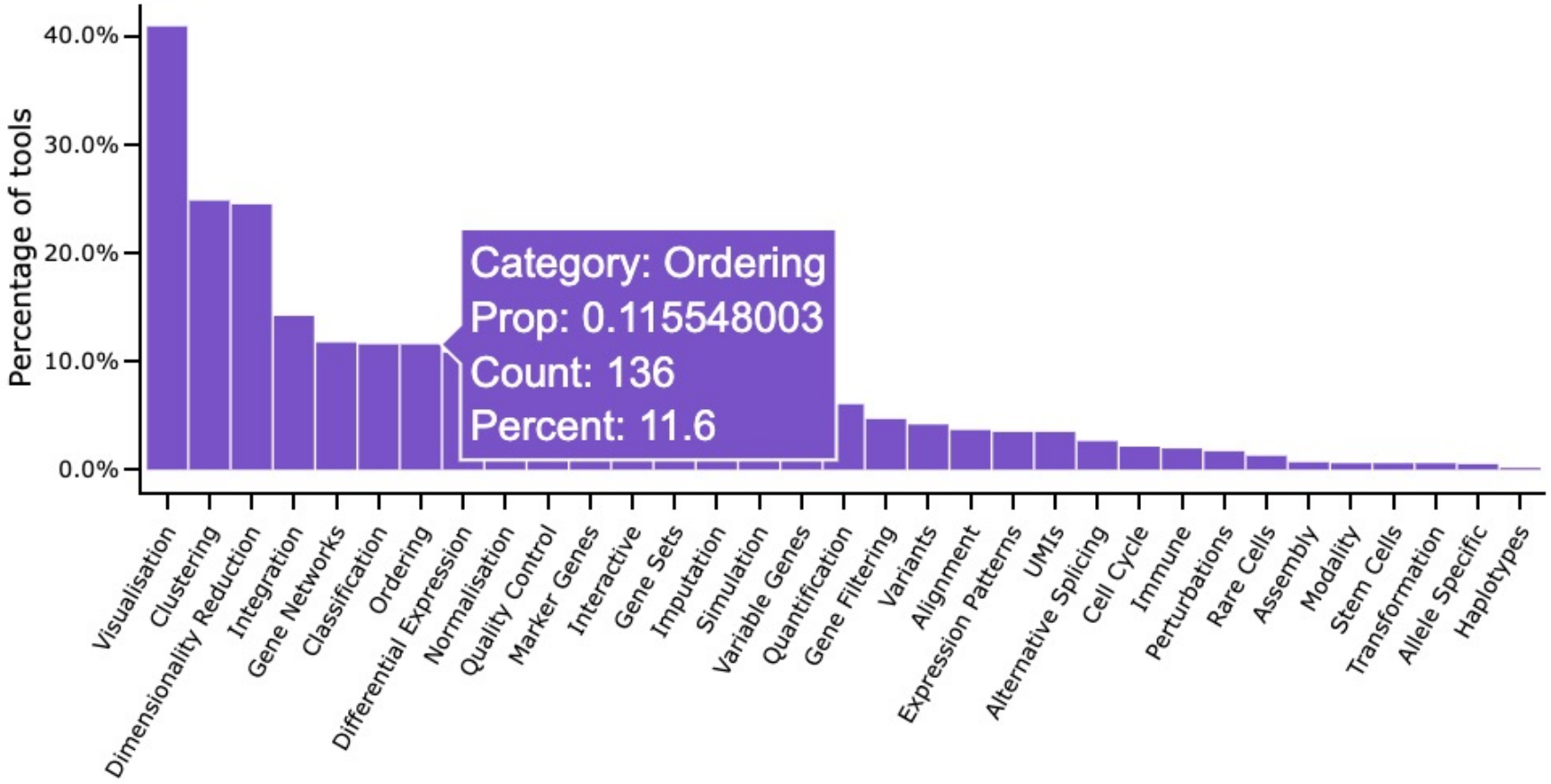
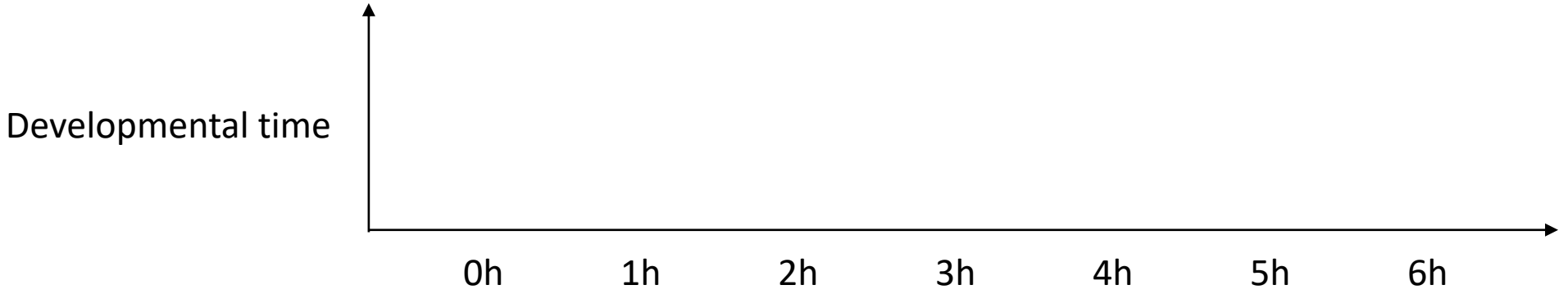

Count matrix
QC-metrics
Normalization
Feature selection
Removal of confounding factors
Integration analysis
Dimensionality Reduction
Clustering
Diff expression within sample and between samples
Cell type identification
Pseudo-time/
Trajectory inference
LRIP

Again long list of possible tools



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

Example of Trajectories



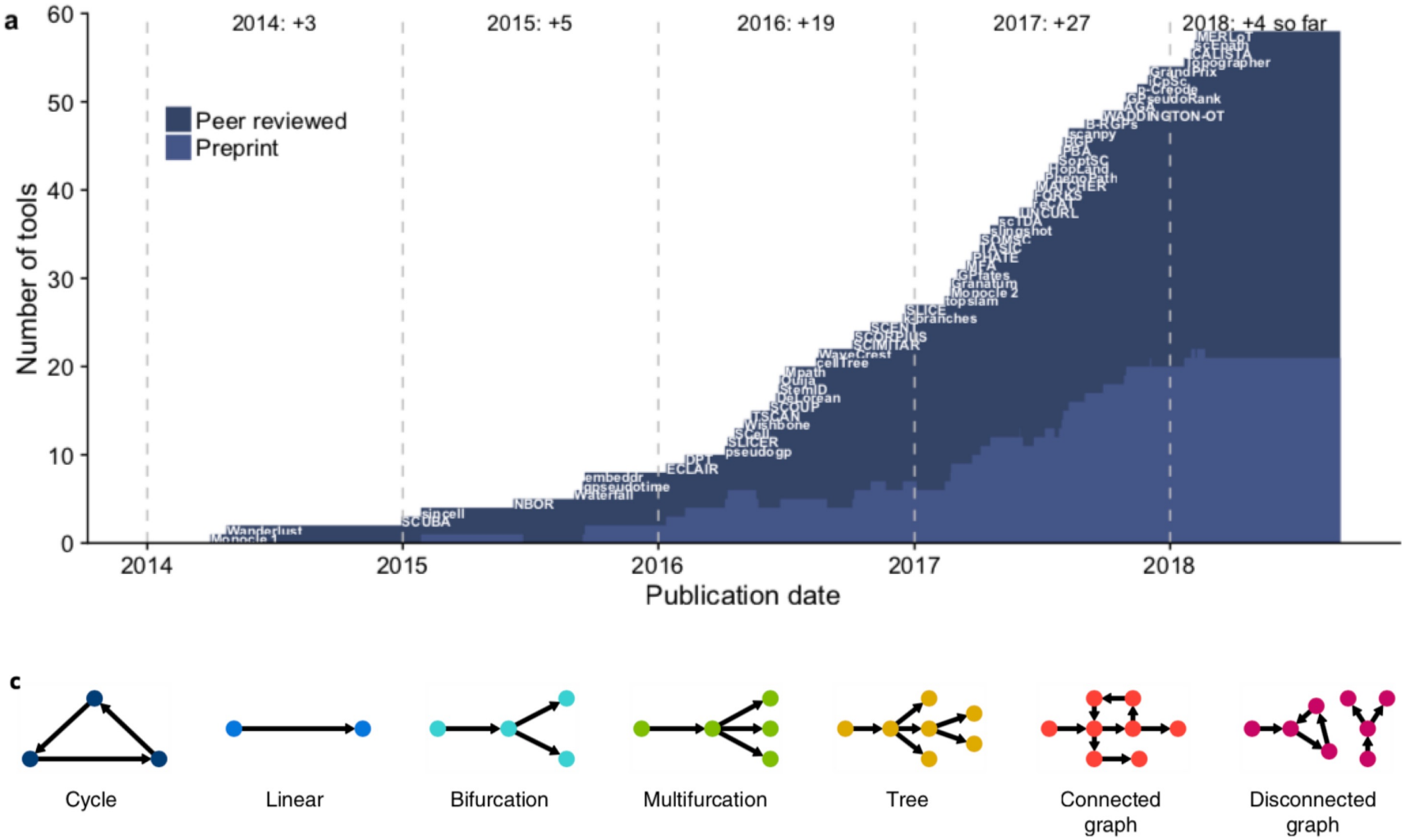
•In the analysed data set one might encounter :
•1. Cells that differentiate display a **continuous spectrum** of states
Transcriptional program for activation and differentiation
•2. Individual cells will differentiate in an unsynchronized manner Each cell is a snapshot of **differentiation** time
•OR 3. Pseudotime – abstract **unit of progress**
Distance between a cell and the start of the trajectory

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?

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.



Saelens et al (2019) Nat Biotechnology

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.

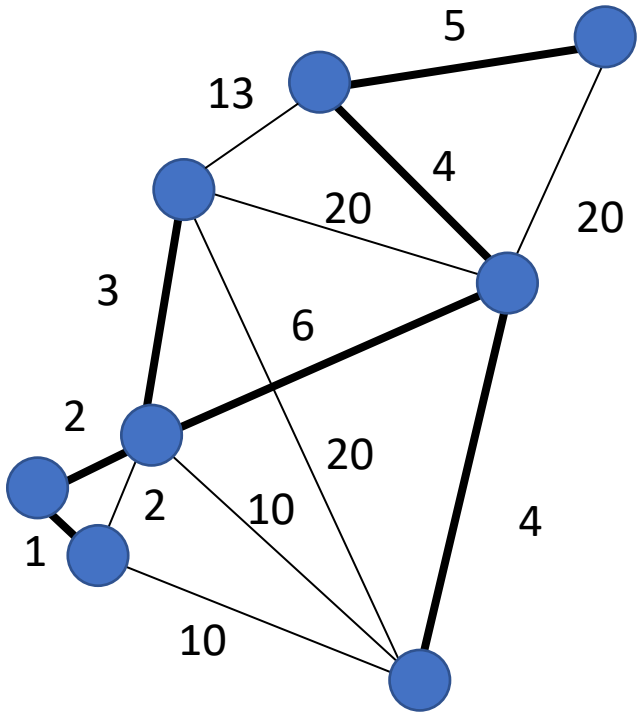
Example of application

- 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 (CAST/EiJ × C57BL/6J, hereafter abbreviated as CAST and C57, respectively) from oocyte to blastocyst stages of mouse preimplantation development (PD)»
- Here finding a trajectory between the cells might be of high interest.

Minimum spanning tree

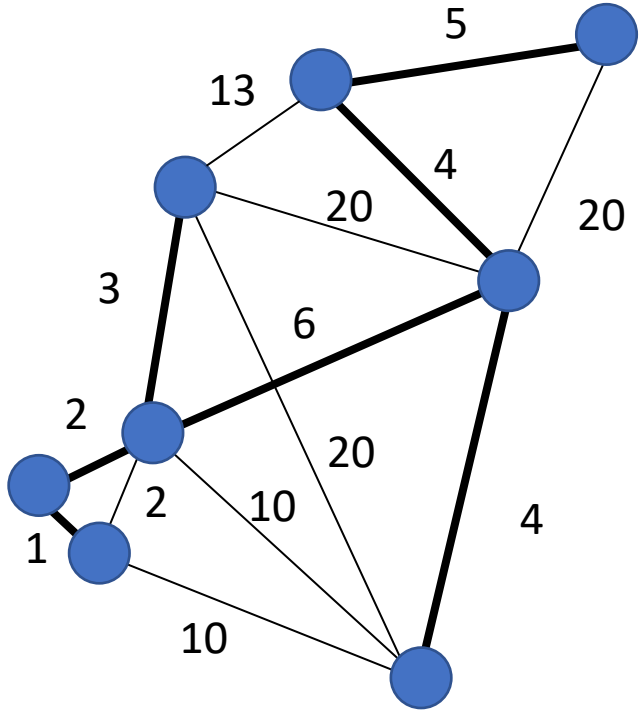
- Take a weighted graph.
- Take a spanning tree
- Take the minimum of all spanning trees.

Example

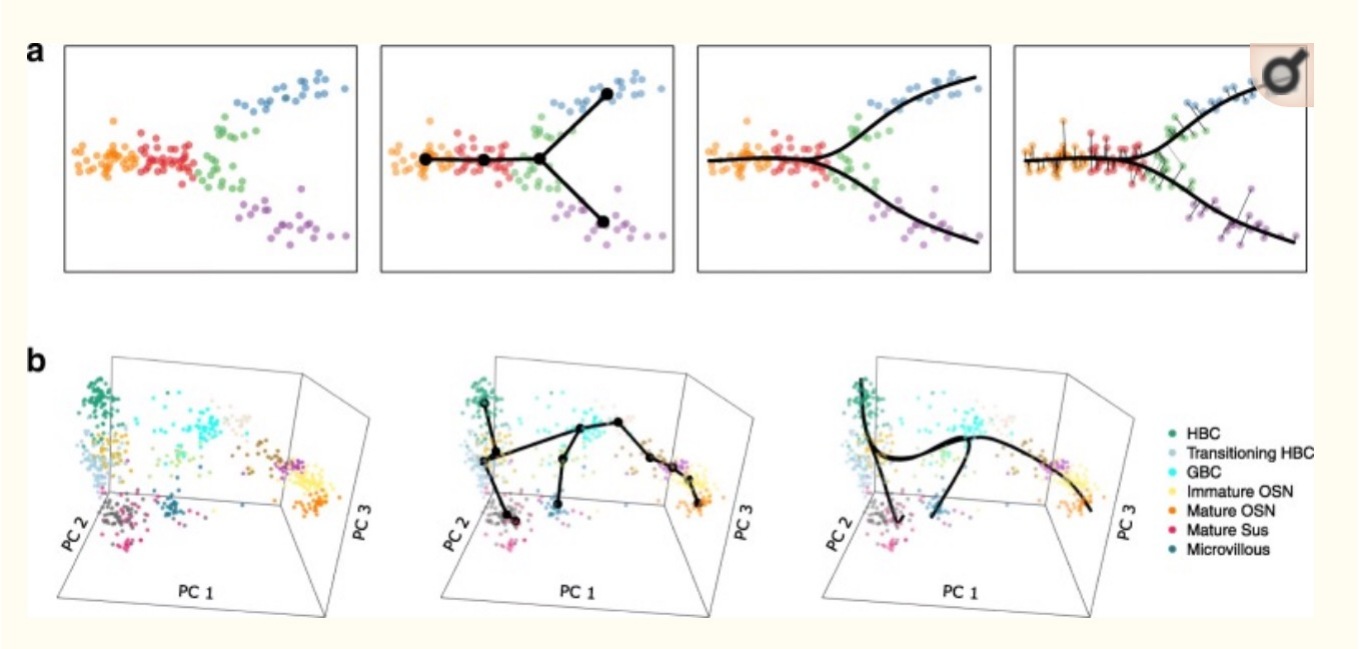


Minimum spanning tree (MST)

- **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



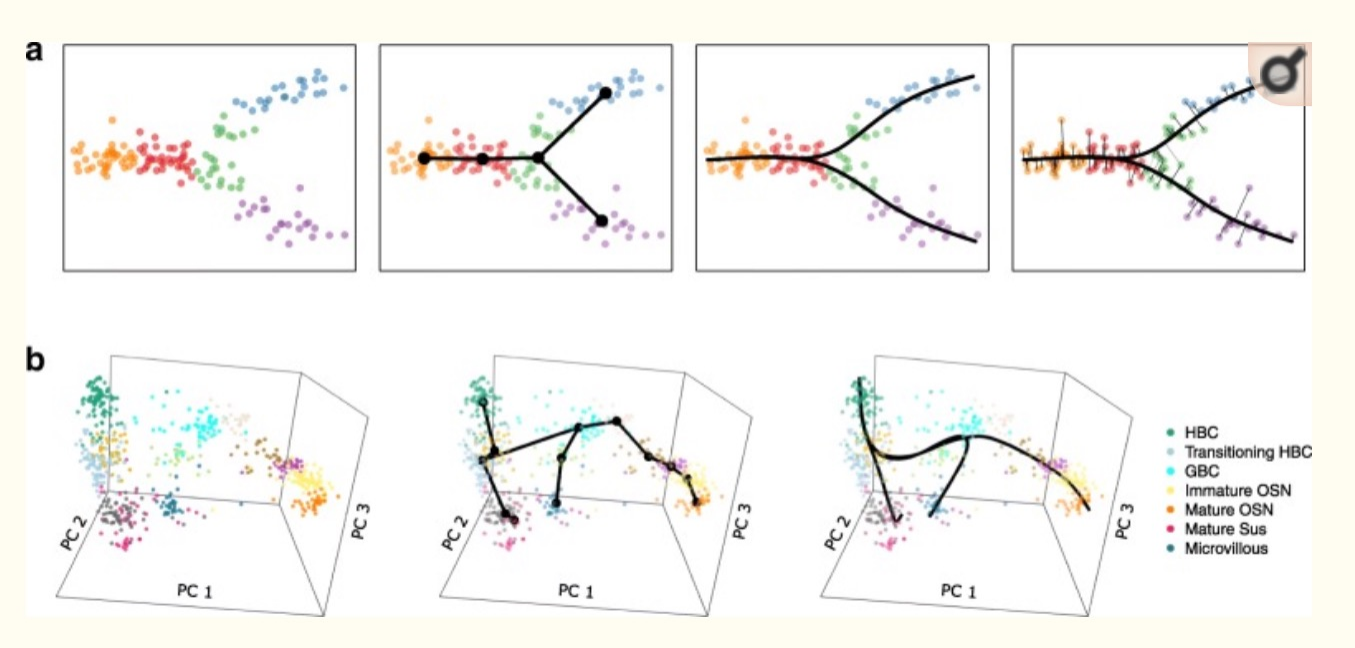
Slingshot (Street et al 2018)



1. Distance between clusters

$$d^2\left(\mathscr{C}_i, \mathscr{C}_j\right) \equiv \left(\overline{X}_i - \overline{X}_j\right)^T \left(S_i + S_j\right)^{-1} \left(\overline{X}_i - \overline{X}_j\right),$$

Slingshot (Street et al 2018)

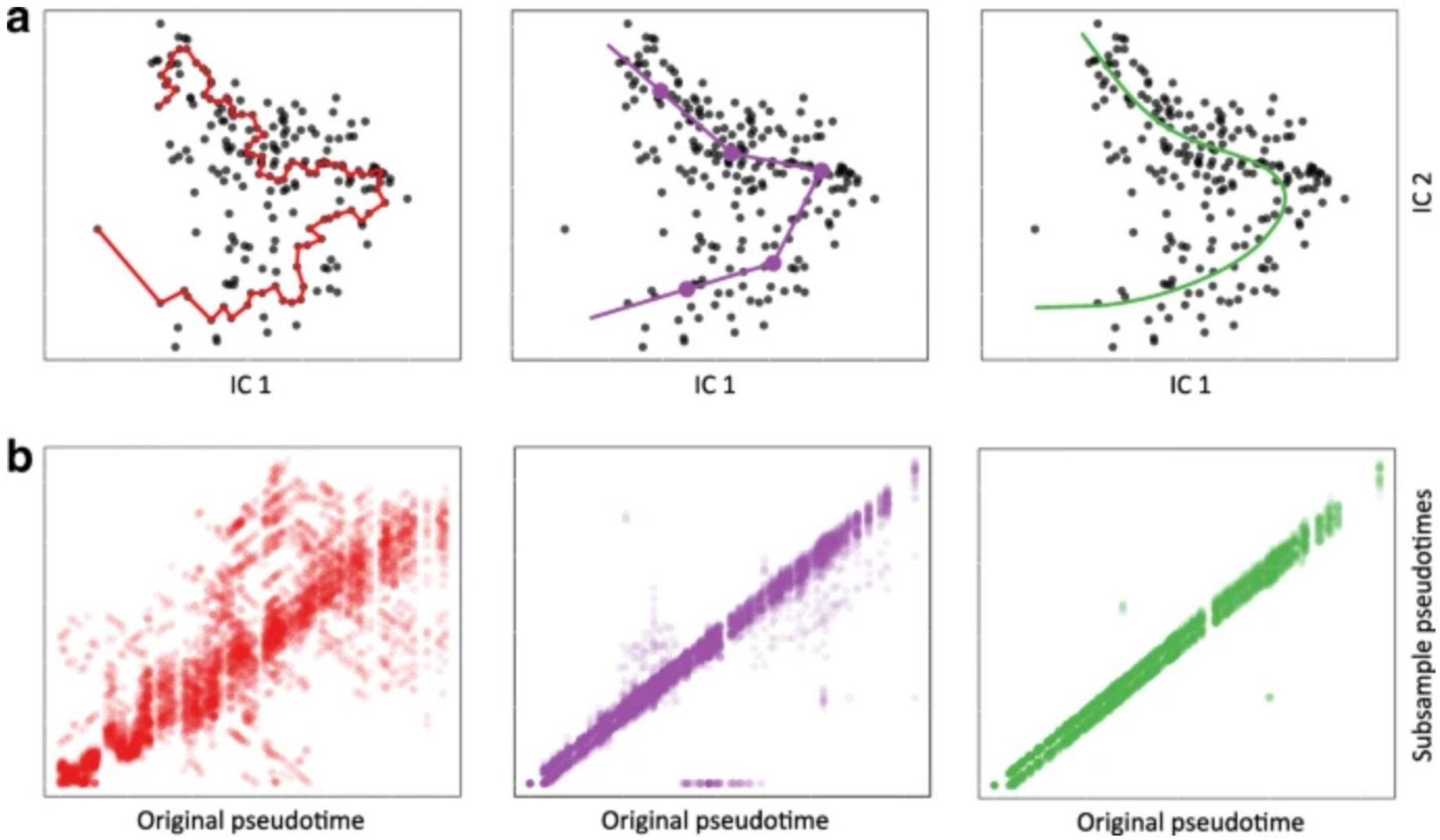


1. Distance between clusters
2. Infer lineages by ordering cell clusters and construct MST
3. Construct principal curves*

*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

Street et al 2018

Slingshot vs others



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.

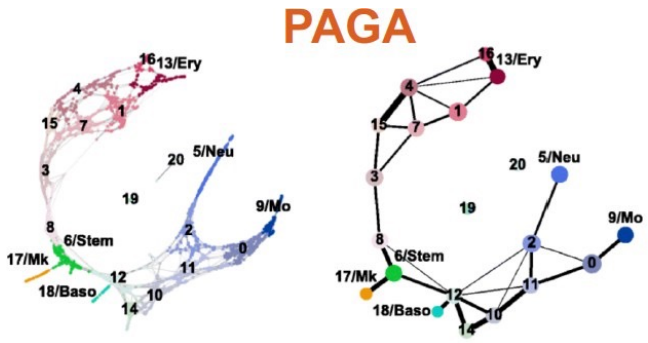
- Those communities that have more links than expected under the null hypothesis of spurious linkage (FDR <1%) remain connected in the PAGA graph, and those links that fail this test are severed. (correction of **spurious linkage**)

RNA Velocity a quite 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.

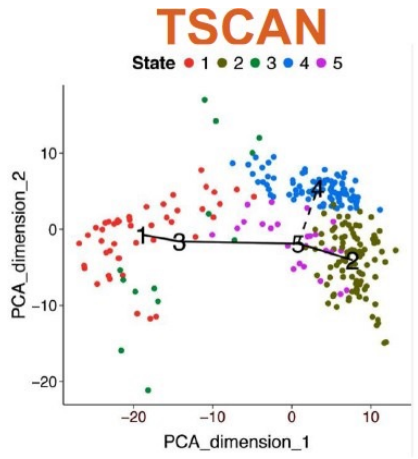
Gioele La Manno (2018) Nature

Some additional tools

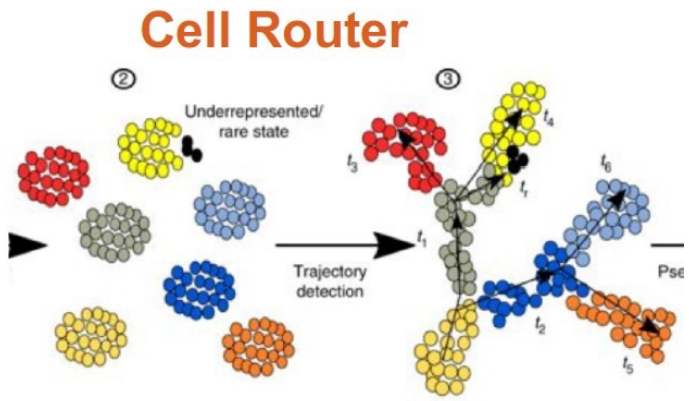


Street et al (2019) Genome Biology

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.



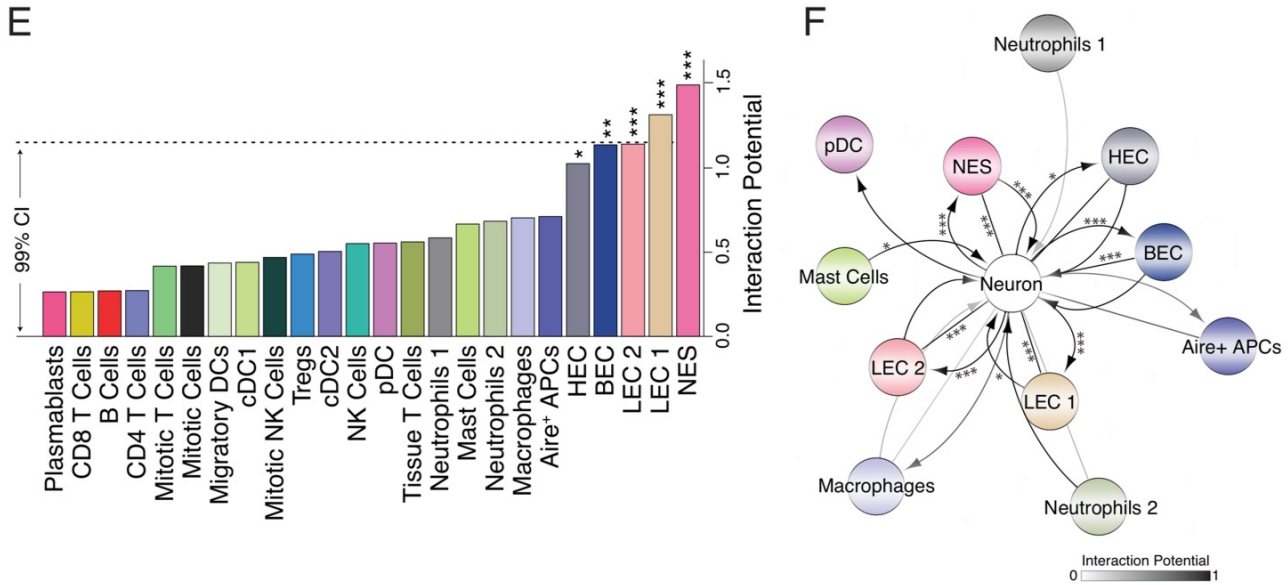
Zhicheng et al (2016) Nuc Acid Res



Da Rocha et al (2018) Nat Commun

Other Post-hoc analysis

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



LRIP – Bioarxiv, https://doi.org/10.1101/833509

NicheNet- Ligand receptor analysis

- Question : In your analysis you have a certain list of DGE genes. Can one associate a pair of Ligand and Receptor responsible for the change in expression of those genes ?

- This is extremely useful as it will point biologist to possible pathways to target.

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

	Target_1	**Target_2**	**Target_3**	**…**	**Target_n**
Ligand_1	P11	P12	P13	…	P1n
Ligand_2	P21				
…	…				
Ligand_m	Pm1	Pm2	Pm3	…	Pmn

Calculate correlation scores

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	0	1	1		0

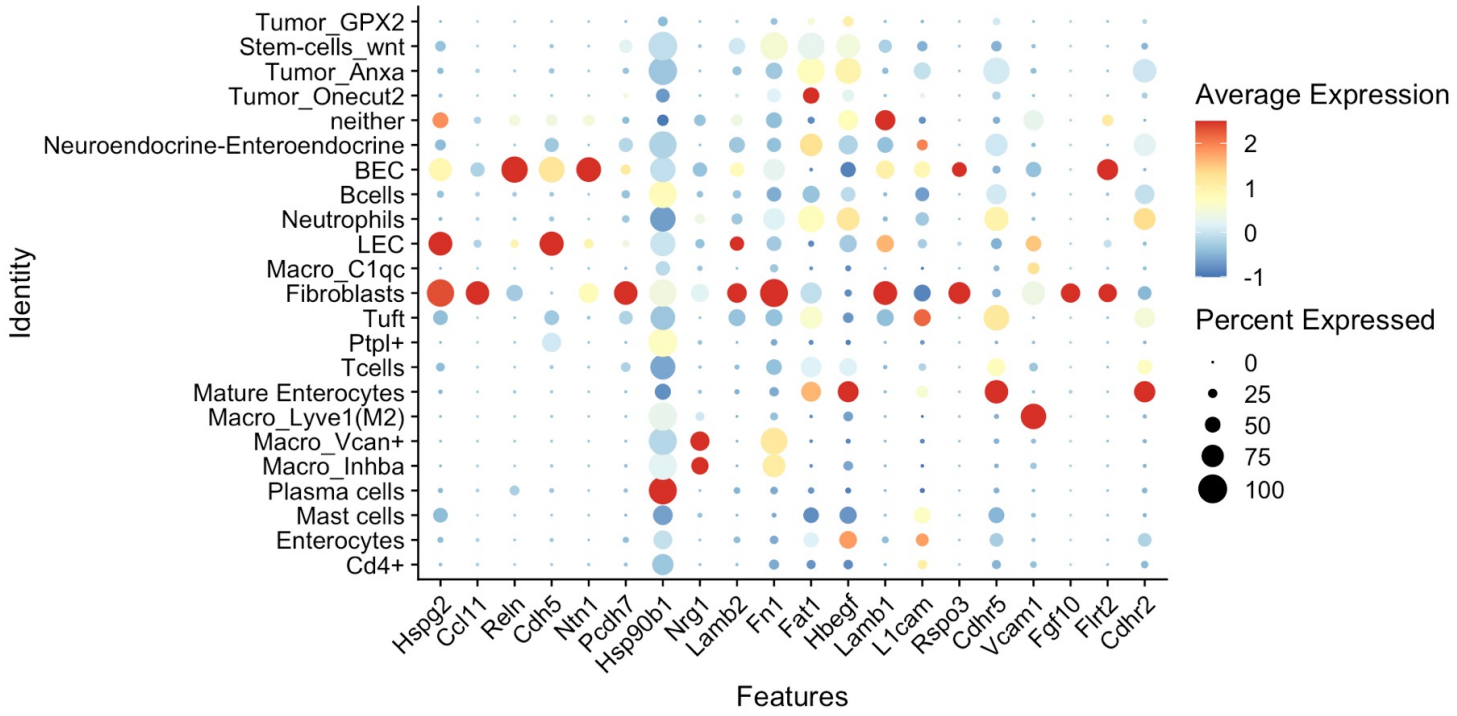
Select only expressed ligands

NicheNet- Ligand receptor analysis-How it works

List of Ligands

L1
L4
L8
L23
L50

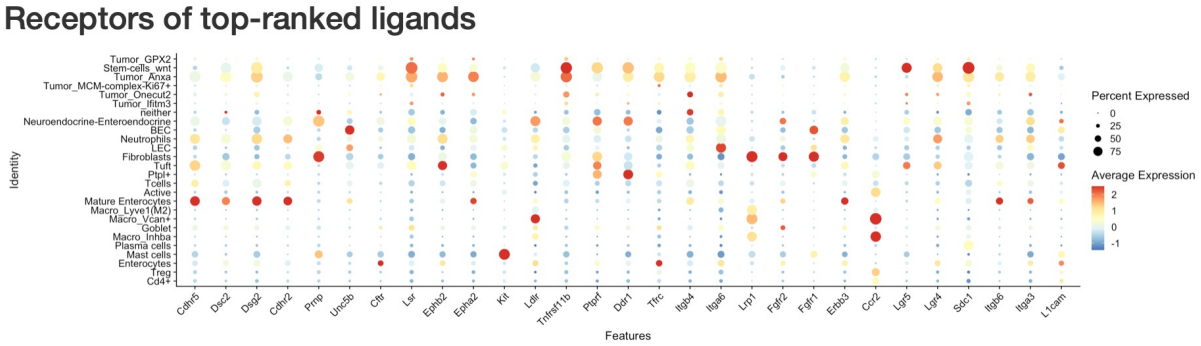
Where are they expressed + Are they actually Differentially expressed as well in that cell type



NicheNet- Ligand receptor analysis-How it works

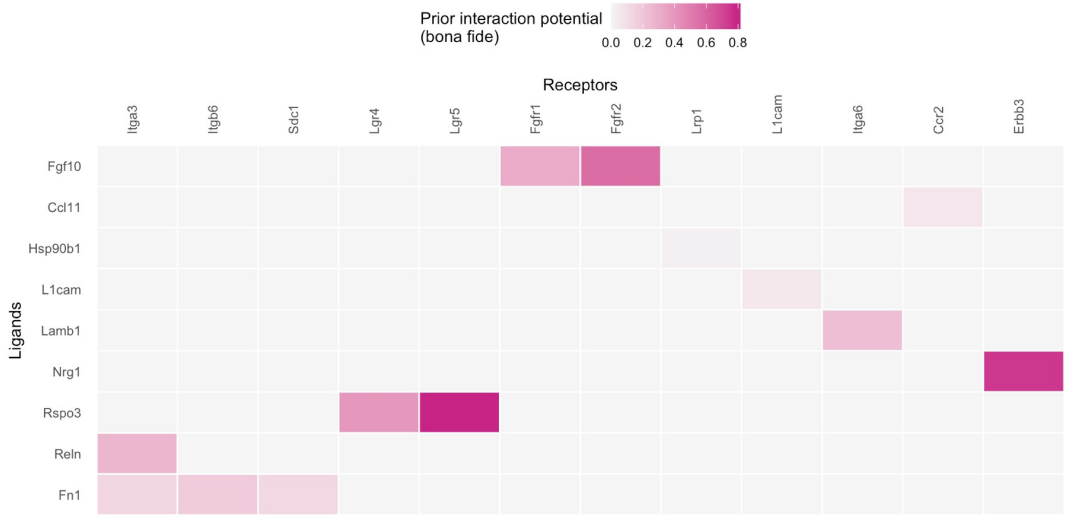
List of Receptors associated to the potential ligands

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



NicheNet- Ligand receptor analysis-How it works

List of Ligand Receptors with interaction potential



Bona fide

In the cluster of cells where the receptor is expressed