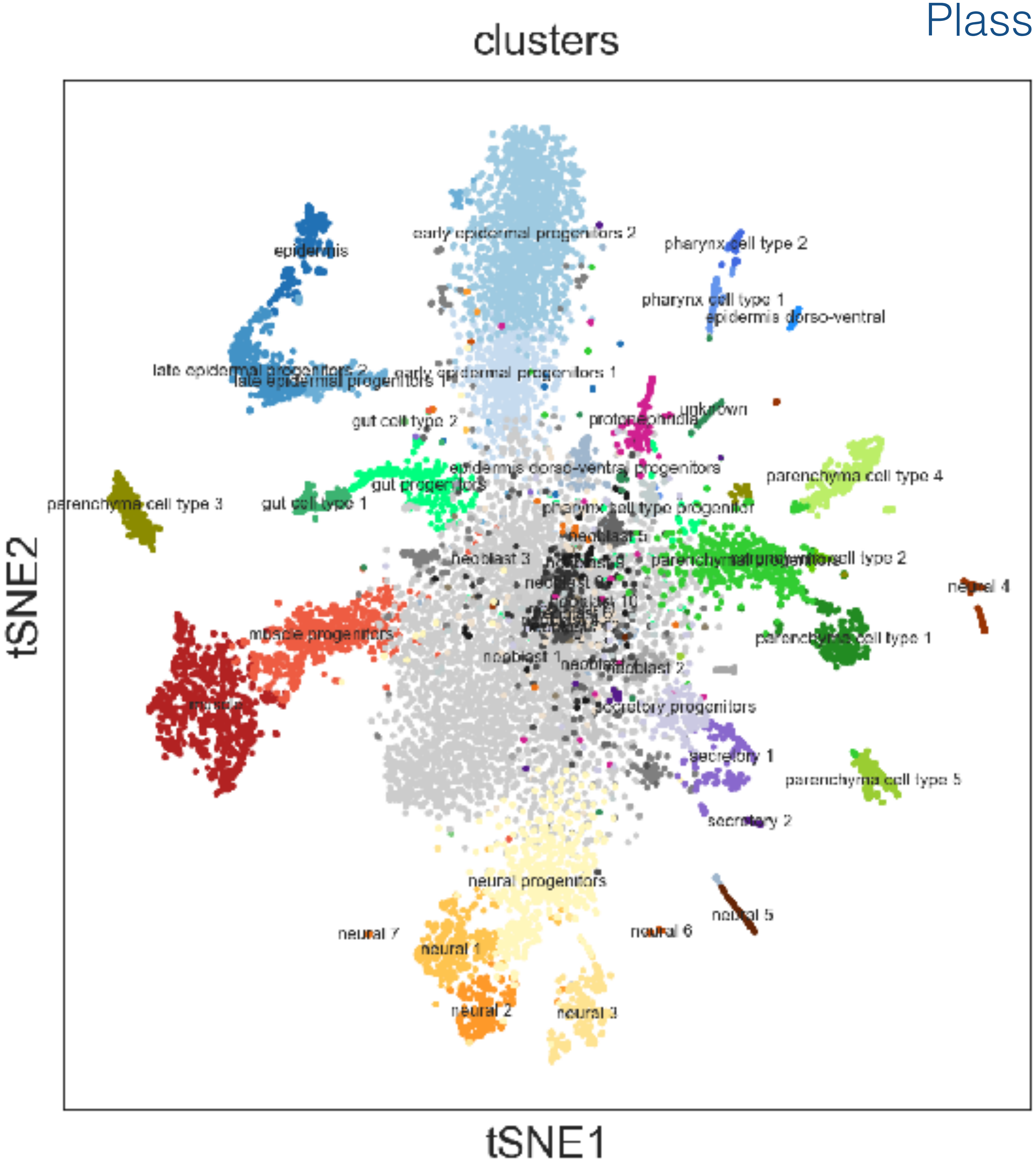
Graph Abstraction

Reconciling clustering with trajectory inference through a topologically consistent map of single cells

F. Alexander Wolf, Institute of Computational Biology, Helmholtz Munich
October 17, 2017 - Single Cell Genomics - Weizmann Institute of Sciences

46 cell types of planaria

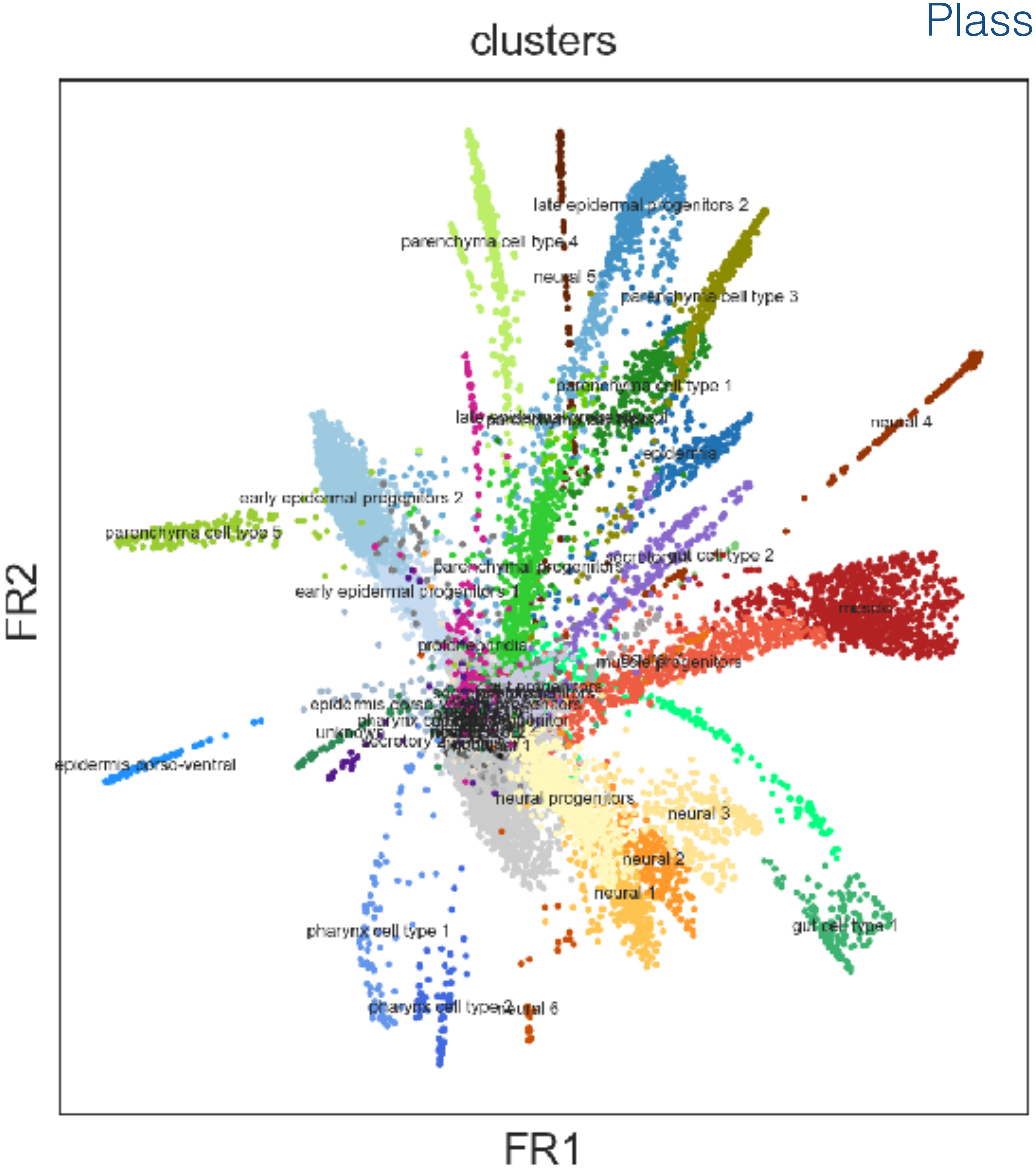
Plass, Solana, …, Rajewski, unpublished (2017)



- Which "cell types"/ clusters are connected?
- Which paths do cells take, where do branchings occur?
- Trace gene "dynamics"/ changes along paths?

46 cell types of planaria

Plass, Solana, …, Rajewski, unpublished (2017)



- Which "cell types"/ clusters are connected?
- Which paths do cells take, where do branchings occur?
- Trace gene "dynamics"/ changes along paths?

Weinreb et al., bioRxiv (2017)

Cell-to-cell variation: overview

Levine et al., Cell (2015), Xu et al. Bioinf (2015) …

clustering

Trapnell et al., Bendall et al. (2014), Haghverdi et al. (2016) …

pseudotemporal ordering

Rizvi et al., Nat Biotechn (2017)

connected topologies

Qiu et al., Nat Meth (2017), Grün et al. Nature (2015) & Cell Stem Cell (2016) …

tree fitting

Leng et al., Nat Meth (2015)

cycle fitting

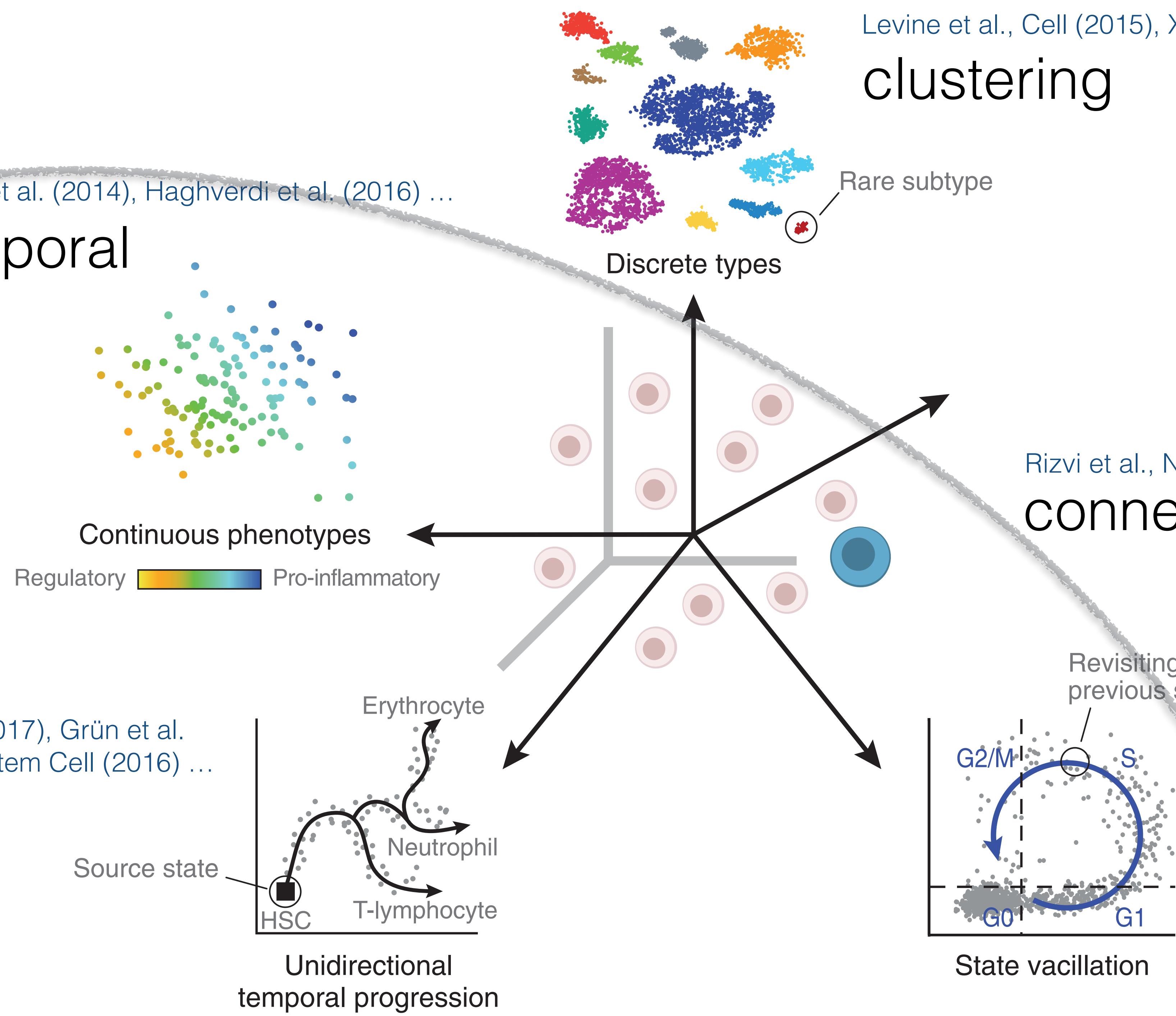


figure from Review Wagner et al., Nat Biotechn (2016)

Cell-to-cell variation: coordinates

Levine et al., Cell (2015), Xu et al. Bioinf (2015) …

one categorical coordinate

Trapnell et al., Bendall et al. (2014), Haghverdi et al. (2016) …

one continuous coordinate

Rizvi et al., Nat Biotechn (2017)

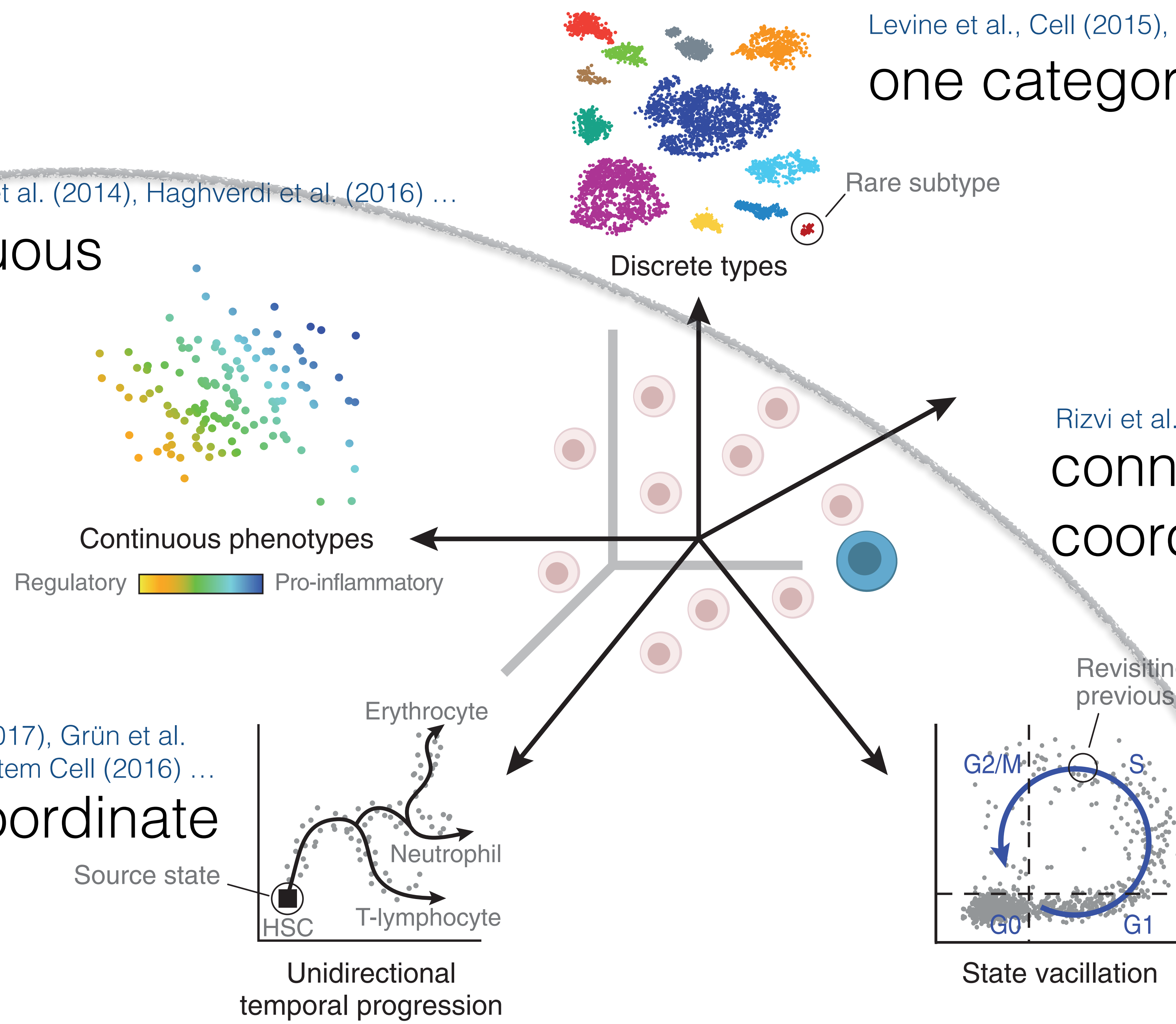
connected-graph like coordinate system

Qiu et al., Nat Meth (2017), Grün et al. Nature (2015) & Cell Stem Cell (2016) …

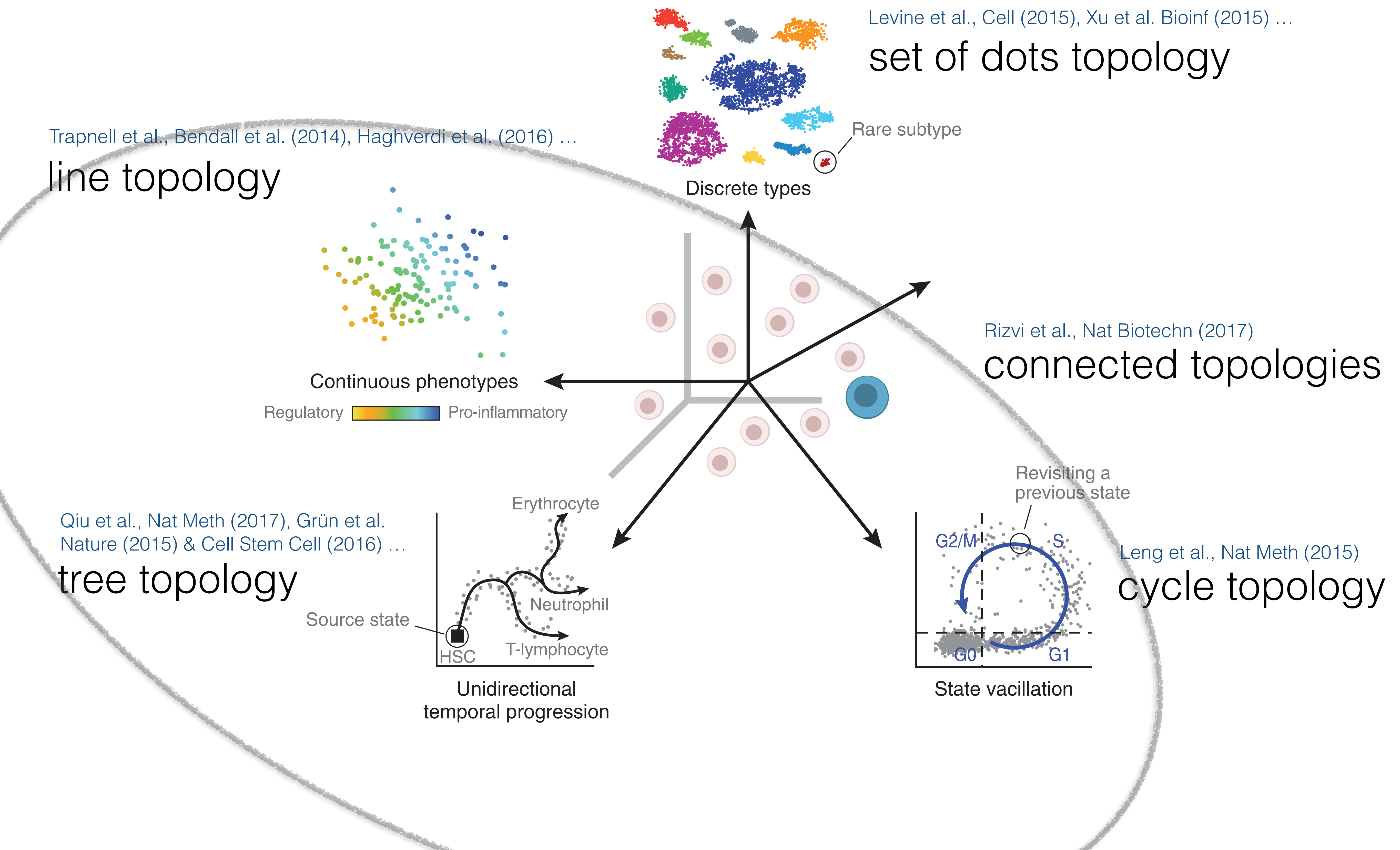
tree-like coordinate system

Leng et al., Nat Meth (2015)

continuous coordinate

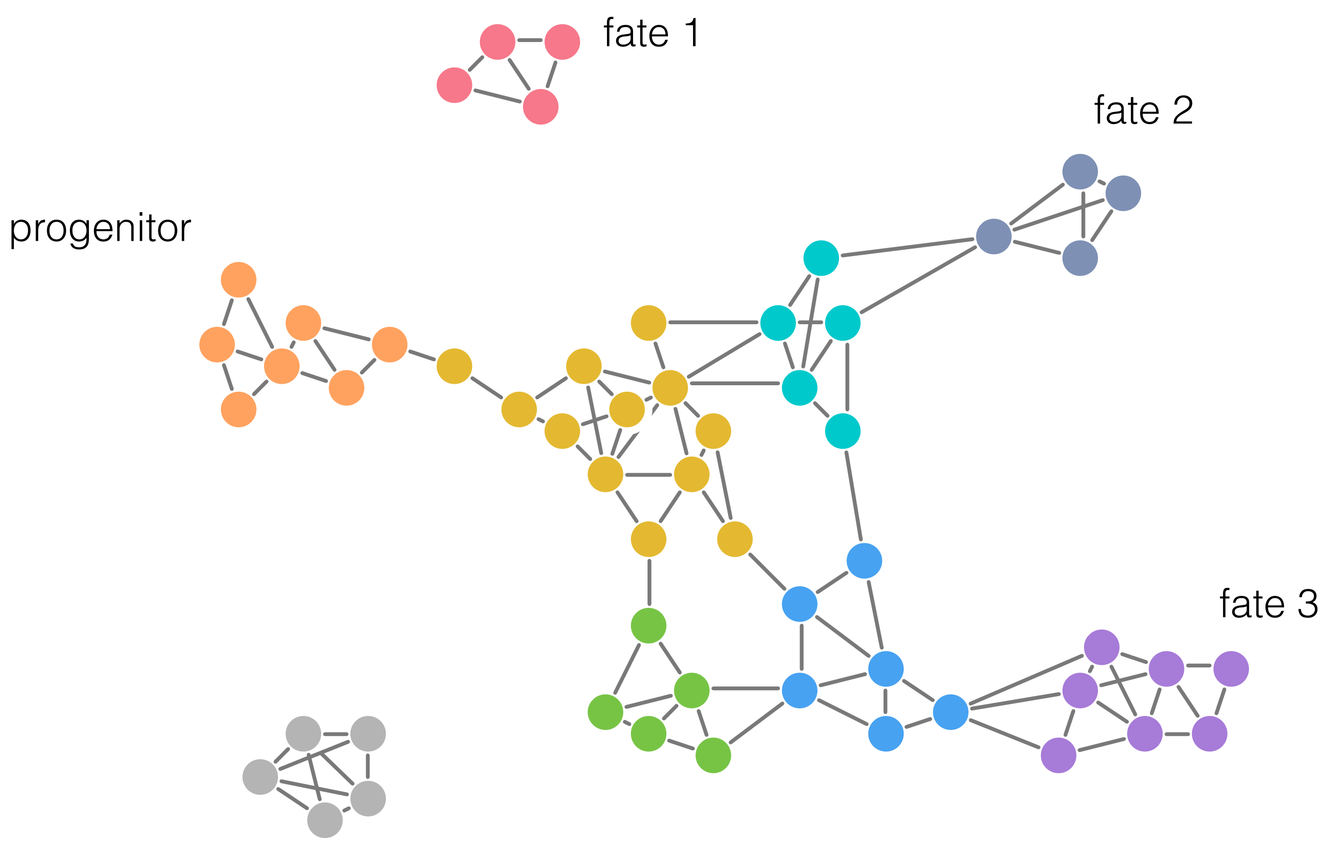


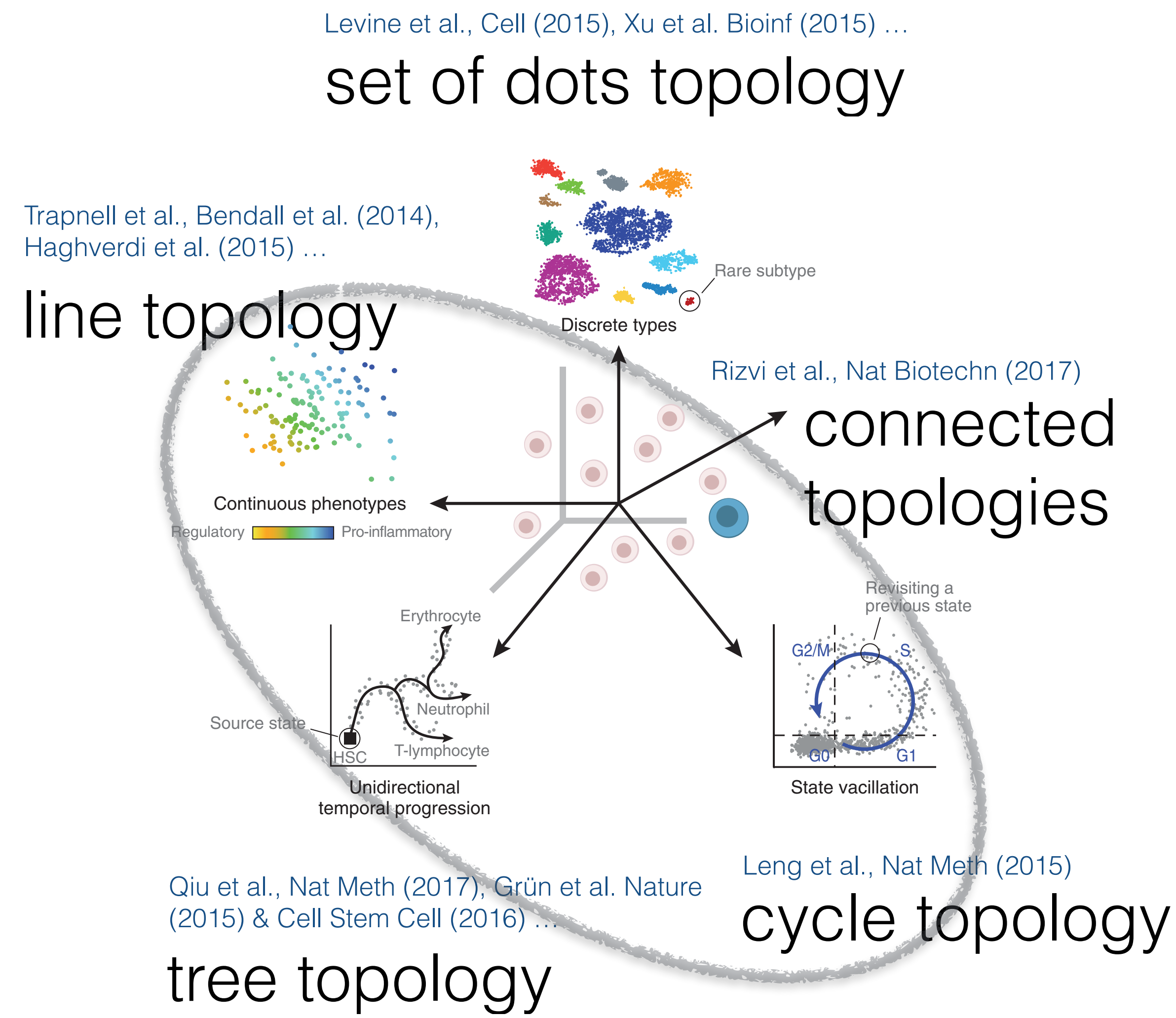
Cell-to-cell variation: topology



Need to unify…

… as single-cell data has complicated connected and disconnected topology





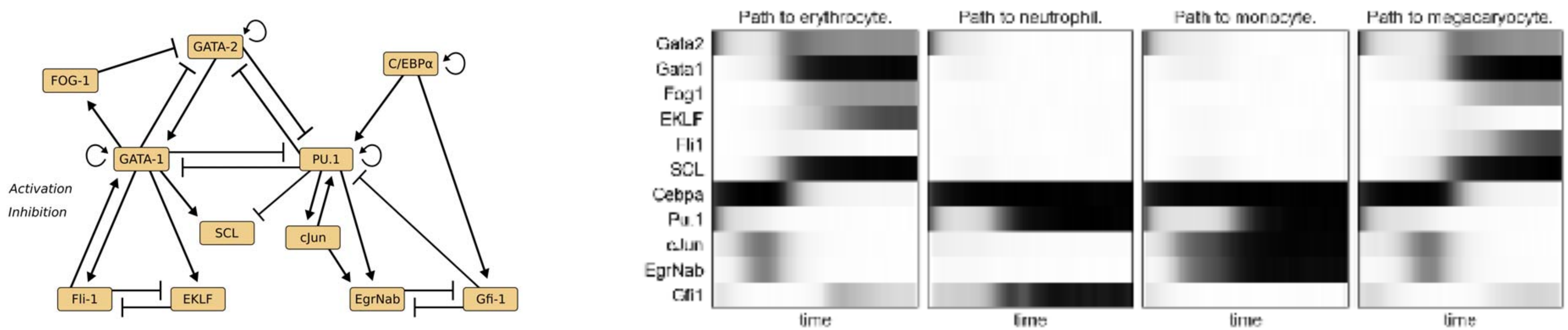
‘single-cell graph’ represents topology at single-cell resolution

Plan

Simplify single-cell graph to generate a cell map that represents topology at a coarse-grained, human-interpretable resolution.

Graph-based visualization

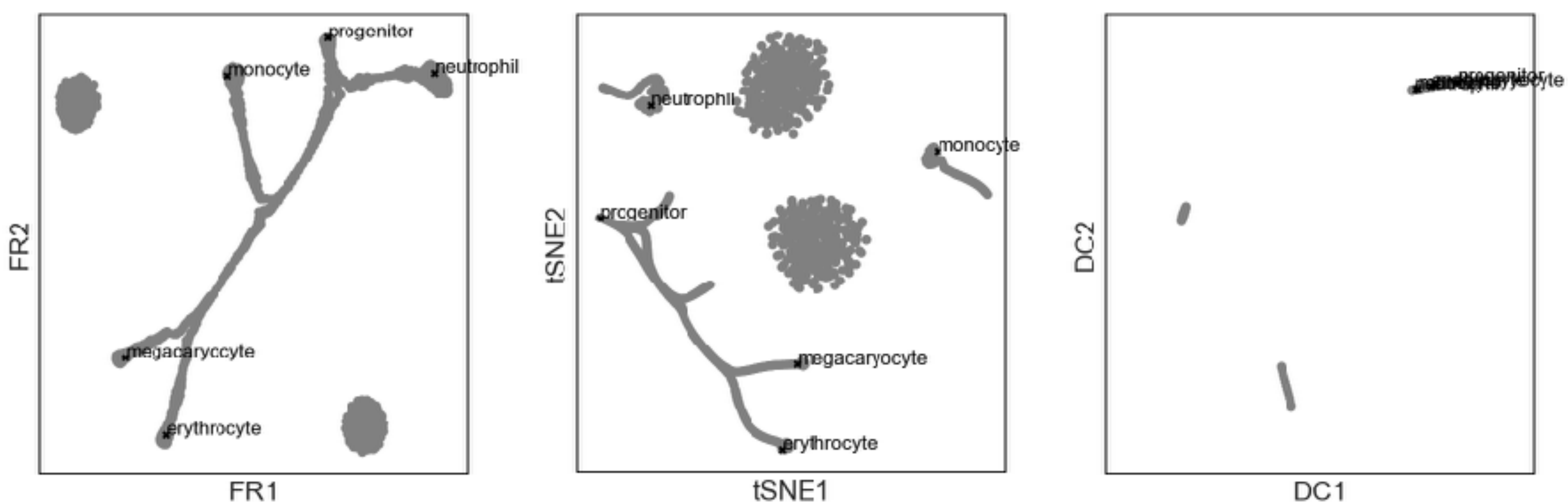
For illustration, model myeloid differentiation…



… and add clusters to the data to model imperfect sampling.

Graph-drawing <u>often</u> conserves topology of single-cell graph.

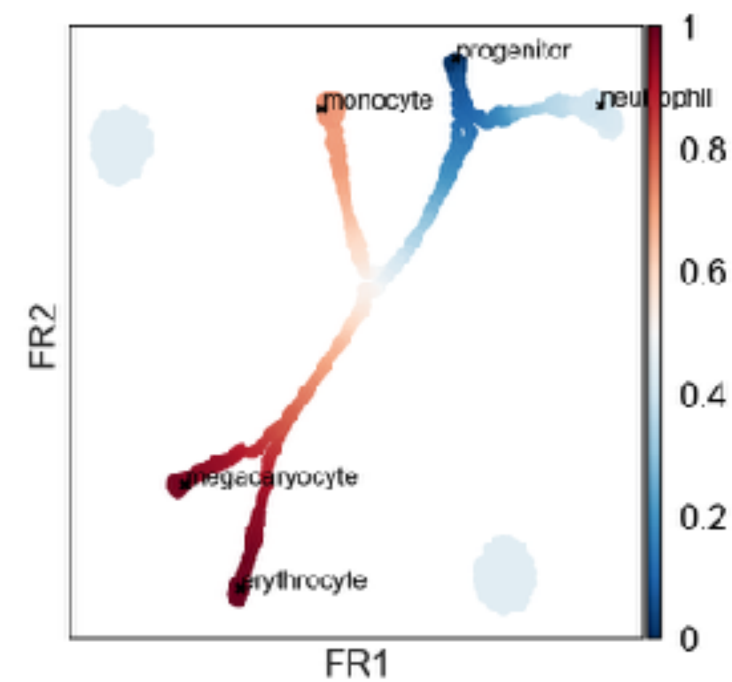
Weinreb et al., bioRxiv (2017)



Graph-based coordinates

Continuous coordinate: random-walk based distance on graph

Generalize scale-free random-walk based distance measures to disconnected graphs



$$\text{mean commute time}(\iota_1, \iota_2) = 2n_{\text{edges}} \sum_{r=1+n_{\text{comps}}}^{n_{\text{nodes}}} \left(\frac{1}{1-\lambda_i}\right)^2 (v_{r\iota_1} - v_{r\iota_2})^2,$$

extends Lovász, Combinatorics (1993)

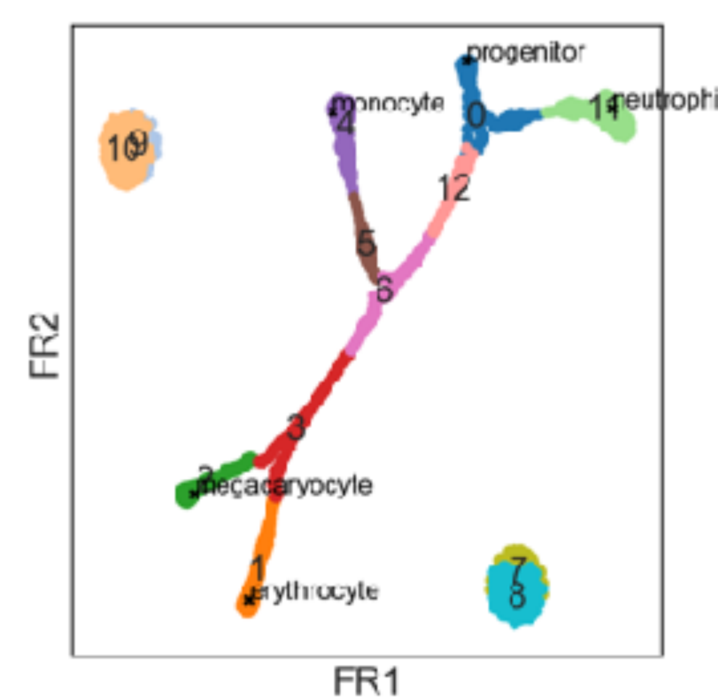
$$\text{dpt}(\iota_1, \iota_2) = \sum_{r=1+n_{\text{comps}}}^{n_{\text{nodes}}} \left(\frac{\lambda_i}{1-\lambda_i}\right)^2 (v_{r\iota_1} - v_{r\iota_2})^2,$$

extends Haghverdi *et al.*, Nat. Meth. (2016)

Categorical coordinate: cluster index

Optimizing <u>graph modularity</u> is sensitive to changes in topology

Newman, Phys. Rev. E (2004)
Blondel *et al.*, J. Stat. Mech. (2008)
Levine *et al.*, Cell (2015)

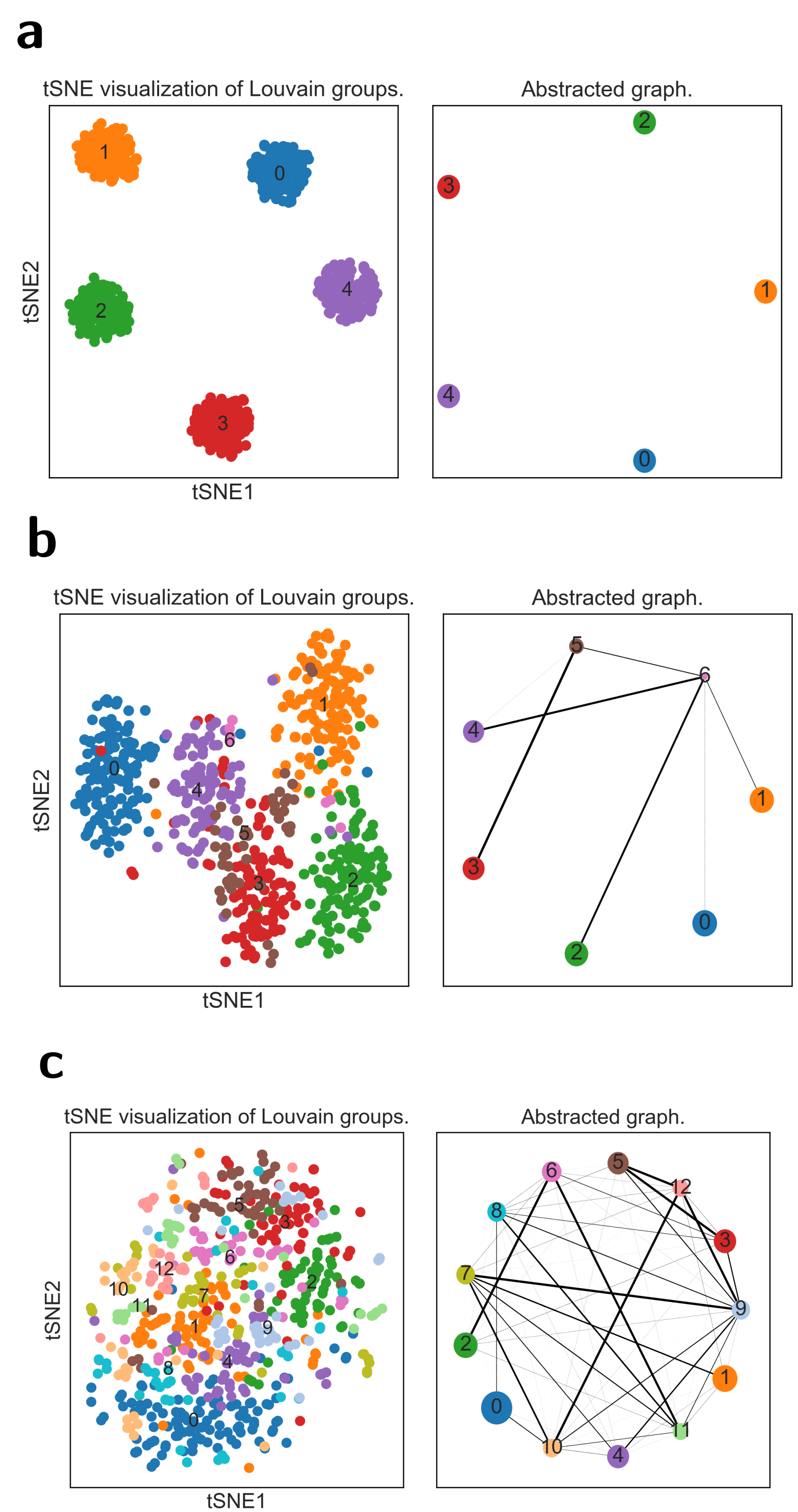


Relating clusters with each other

Develop statistical test of <u>connectedness of clusters</u>

Consider clusters as connected if they have more inter-edges than expected under random connections

$$M_{ij} = K_{ij}/n_{\text{edges}} - \theta_i\theta_j$$
$$\mathrm{E}\left[M_{ij}\right] = 0$$
$$\mathrm{var}\left[M_{ij}\right] = \theta_i\theta_j(1-\theta_i\theta_j)/n_{\text{edges}}$$

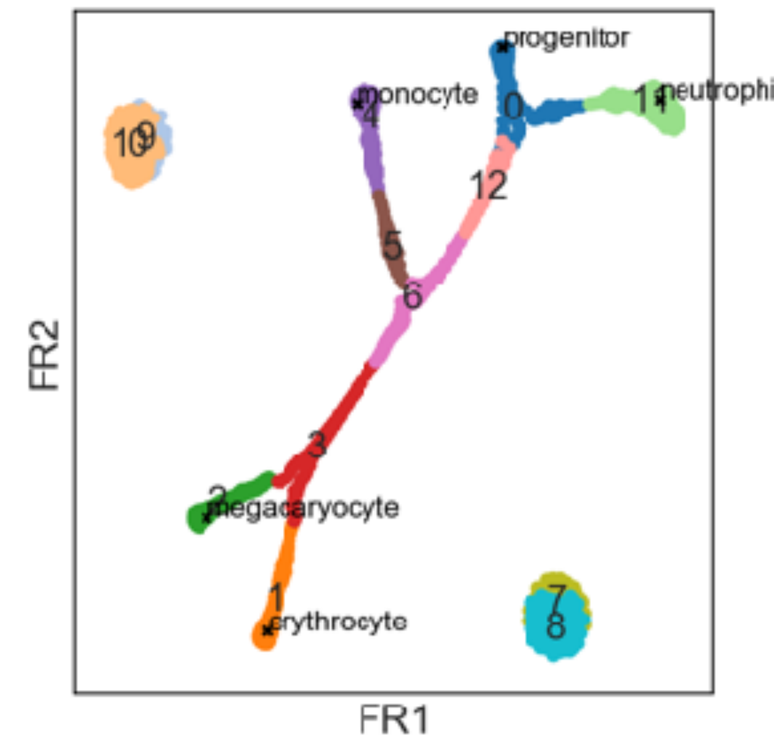


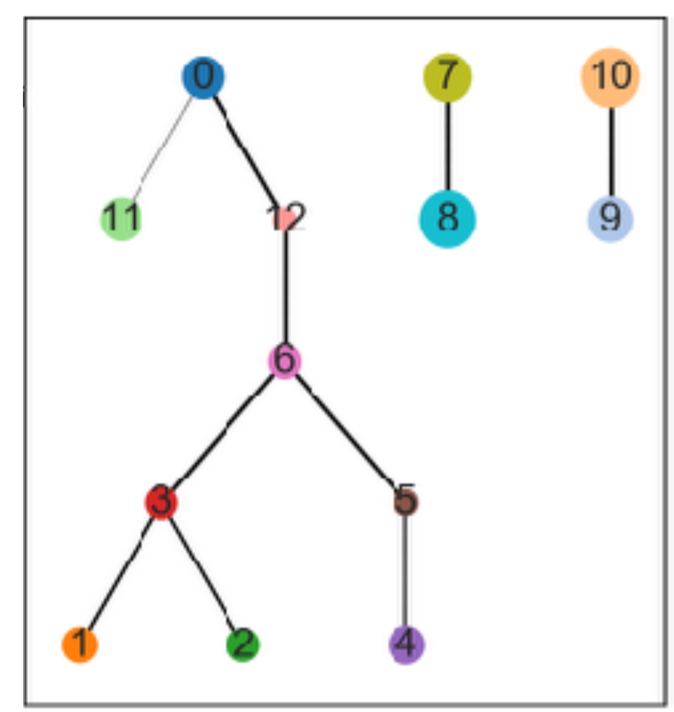
Relating clusters with each other

Develop statistical test of connectedness of clusters

Consider clusters as connected if they have more inter-edges than expected under random connections

$$
\begin{aligned}
M_{ij} &= K_{ij}/n_{\text{edges}} - \theta_i\theta_j \\
\mathrm{E}\left[M_{ij}\right] &= 0 \\
\operatorname{var}\left[M_{ij}\right] &= \theta_i\theta_j(1-\theta_i\theta_j)/n_{\text{edges}}
\end{aligned}
$$



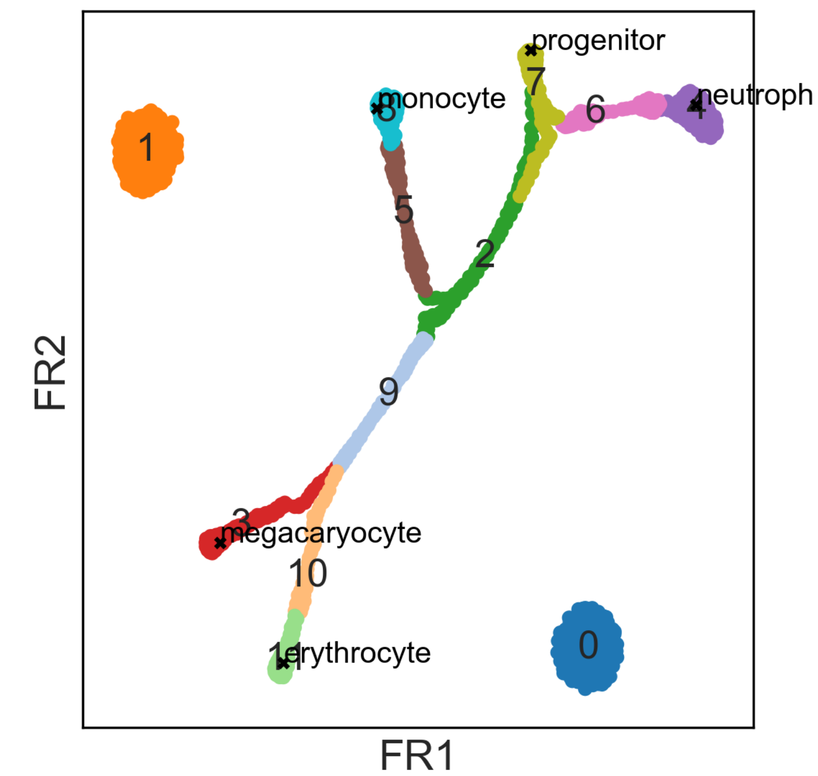


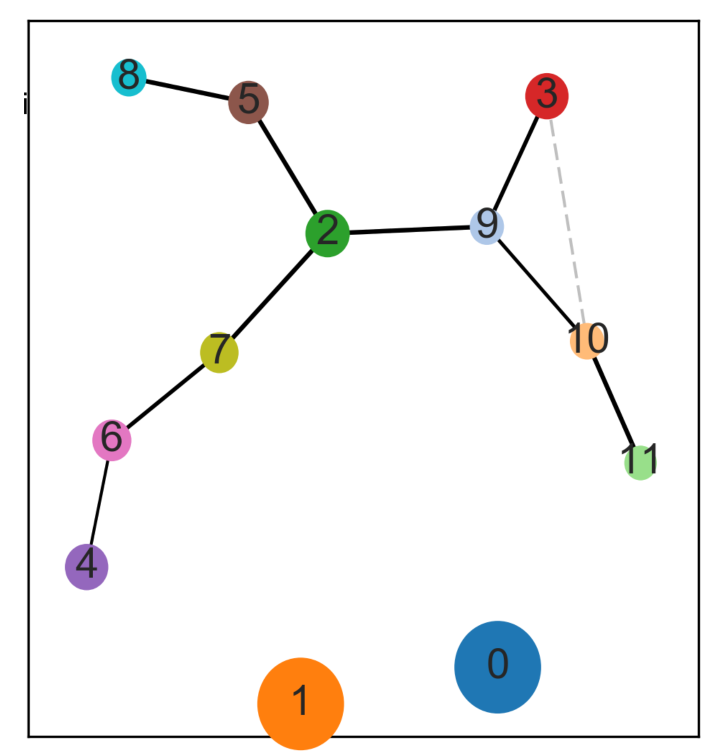
Relating clusters with each other

Develop statistical test of <u>connectedness of clusters</u>

Consider clusters as connected if they have more inter-edges than expected under random connections

$$M_{ij} = K_{ij}/n_{\text{edges}} - \theta_i\theta_j$$
$$\mathrm{E}\left[M_{ij}\right] = 0$$
$$\mathrm{var}\left[M_{ij}\right] = \theta_i\theta_j(1-\theta_i\theta_j)/n_{\text{edges}}$$



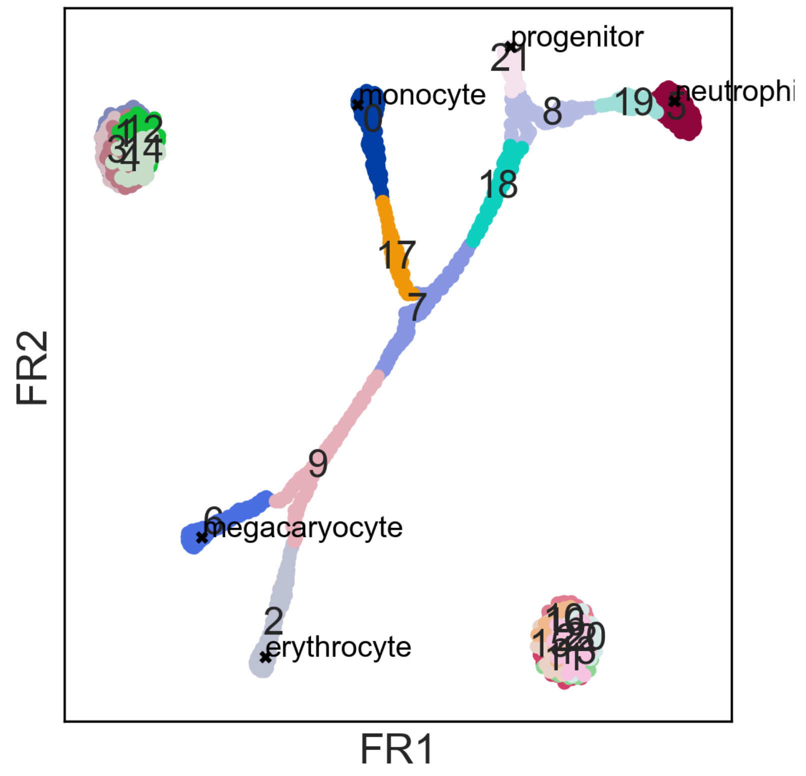


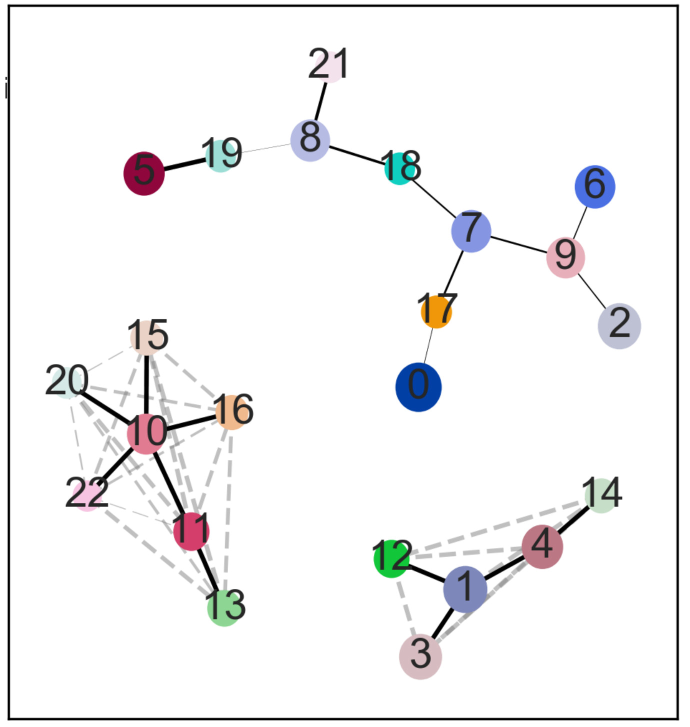
Relating clusters with each other

Develop statistical test of
<u>connectedness of clusters</u>

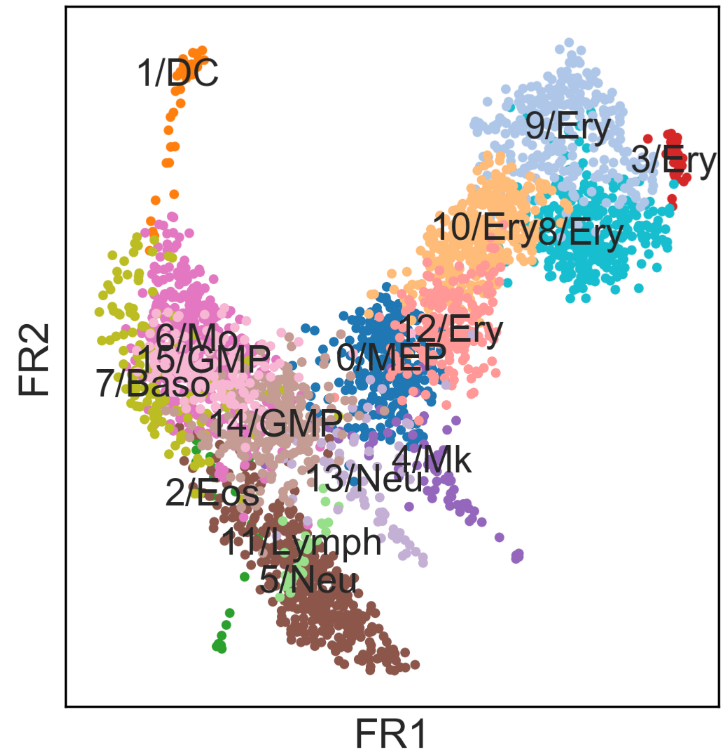
Consider clusters as connected if they have more inter-edges than expected under random connections

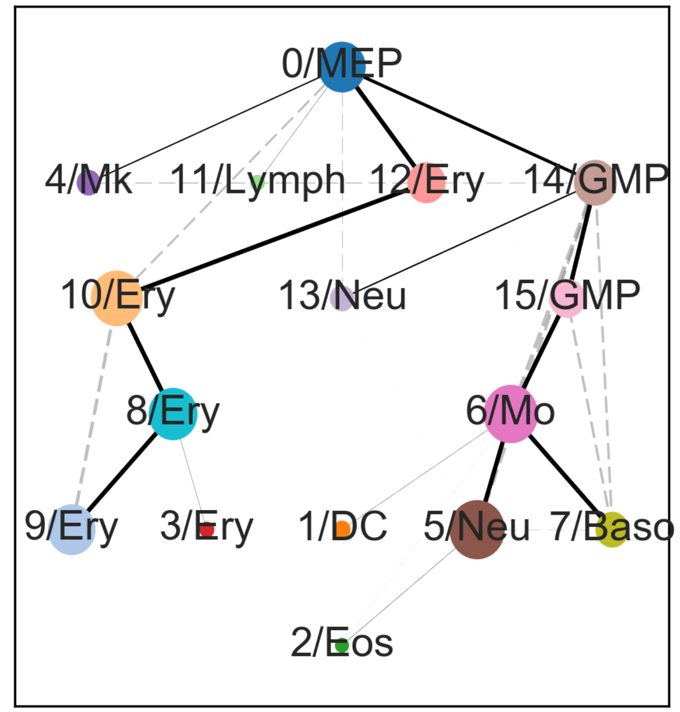
$$M_{ij} = K_{ij}/n_{\text{edges}} - \theta_i\theta_j$$
$$\mathrm{E}\left[M_{ij}\right] = 0$$
$$\mathrm{var}\left[M_{ij}\right] = \theta_i\theta_j(1-\theta_i\theta_j)/n_{\text{edges}}$$



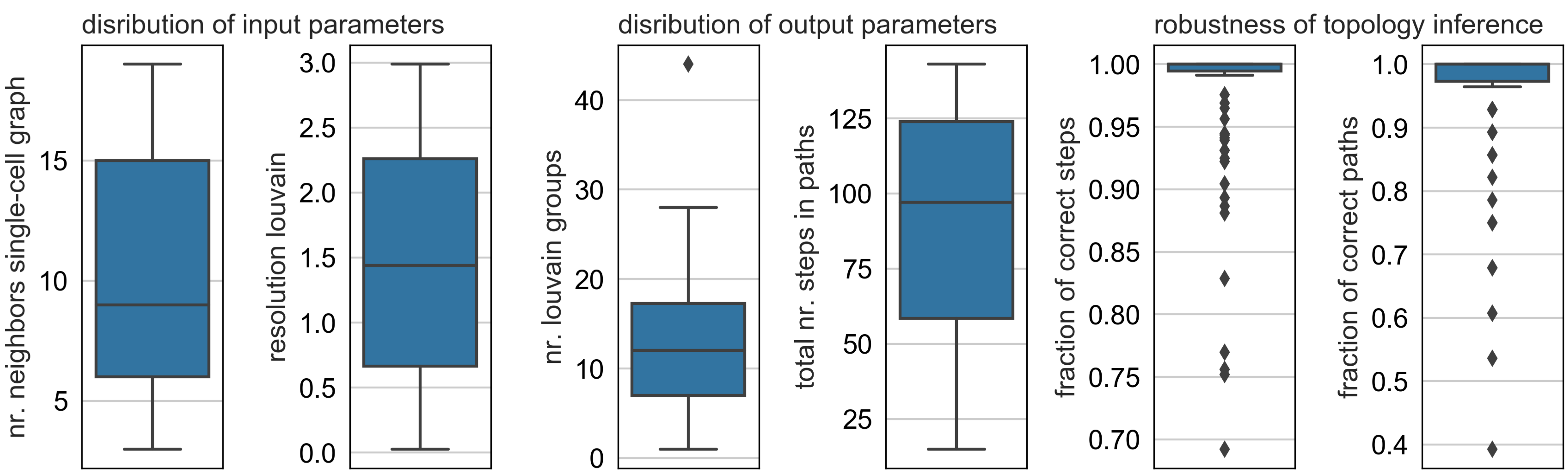


Abstracted topology is robust

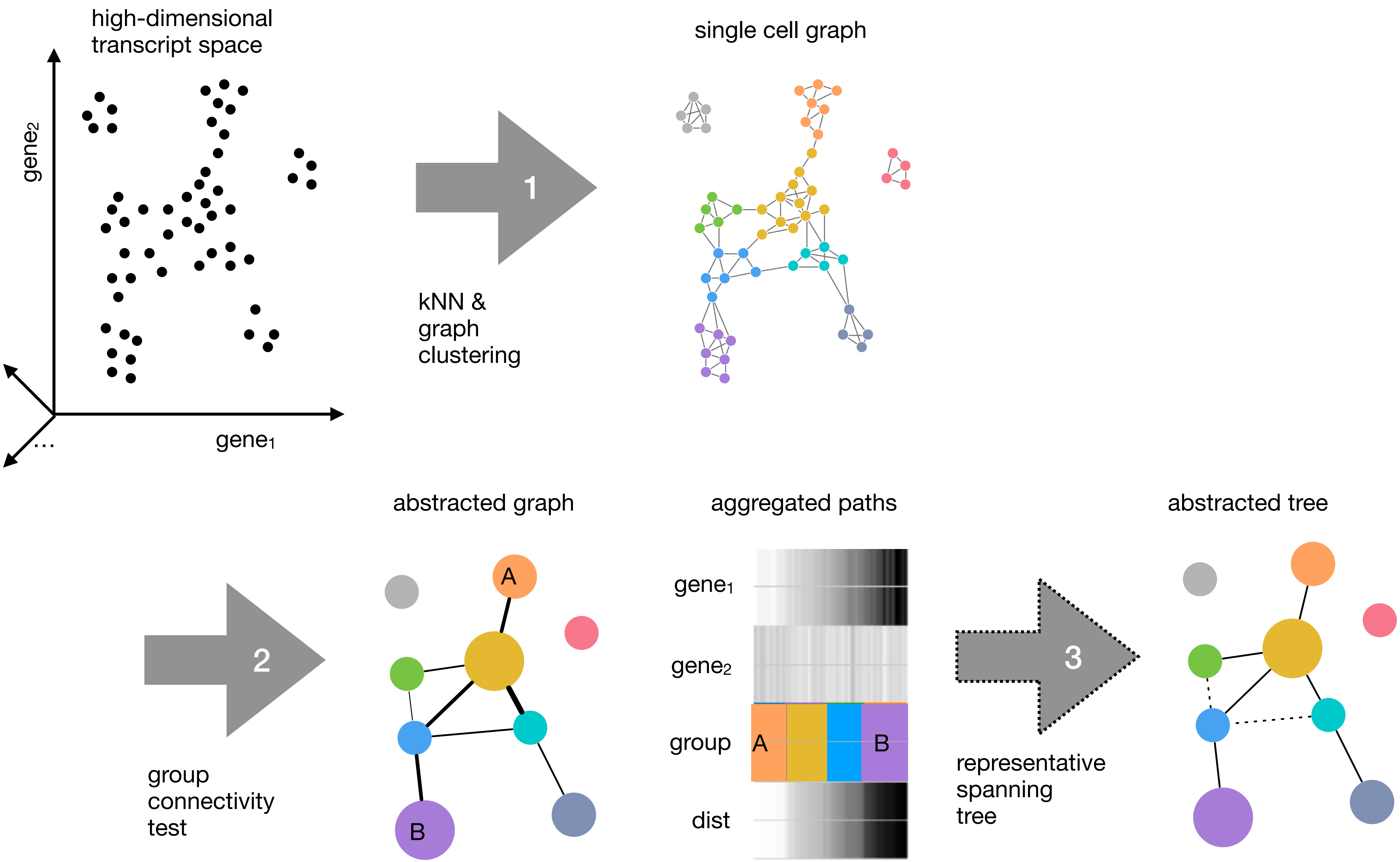




data of
Paul *et al.*, Cell (2015)

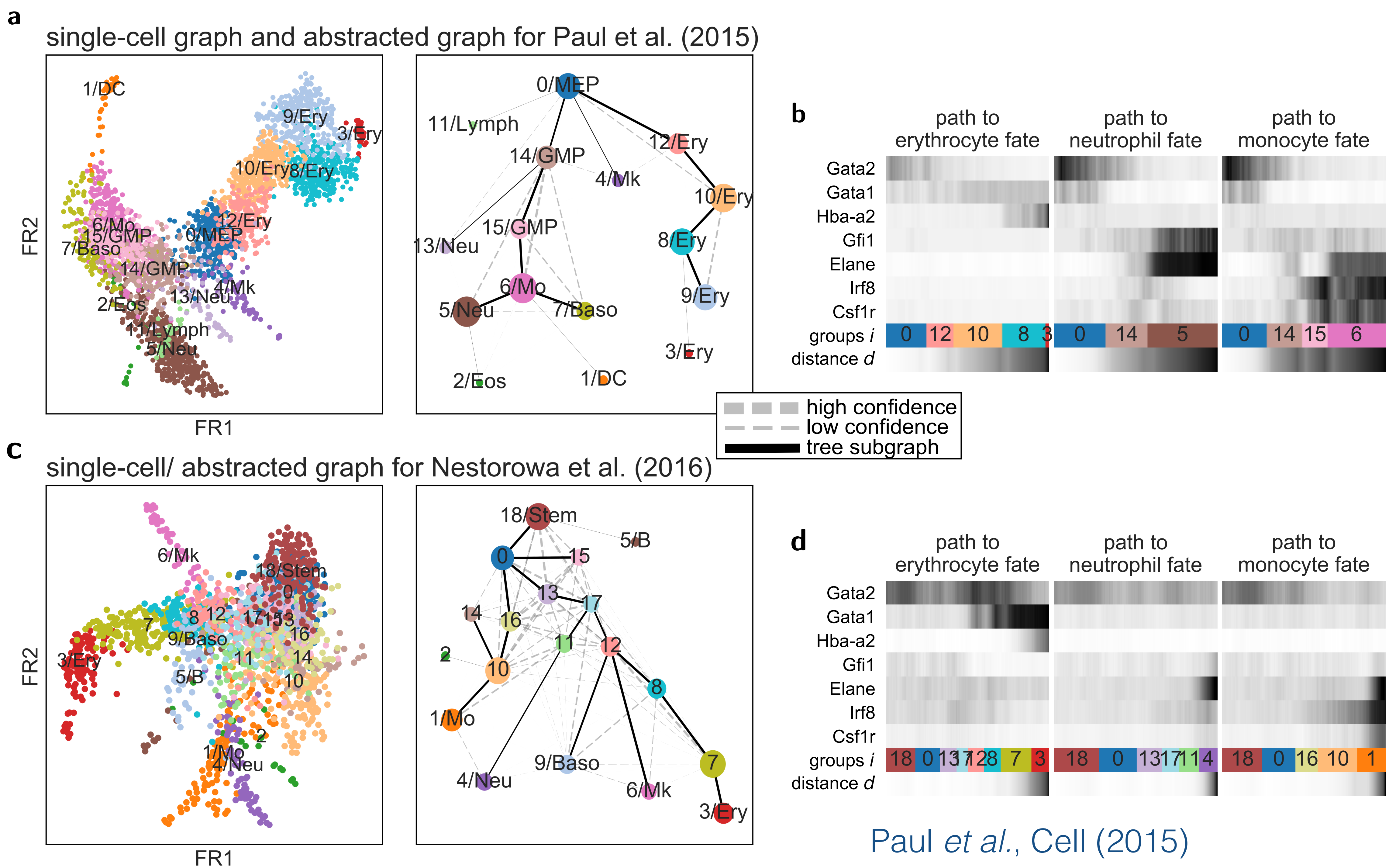


Graph abstraction: overview



Consistent continuous gene changes

… across datasets from different labs.



Paul *et al.*, Cell (2015)

Nestorowa *et al.*, Blood (2016)

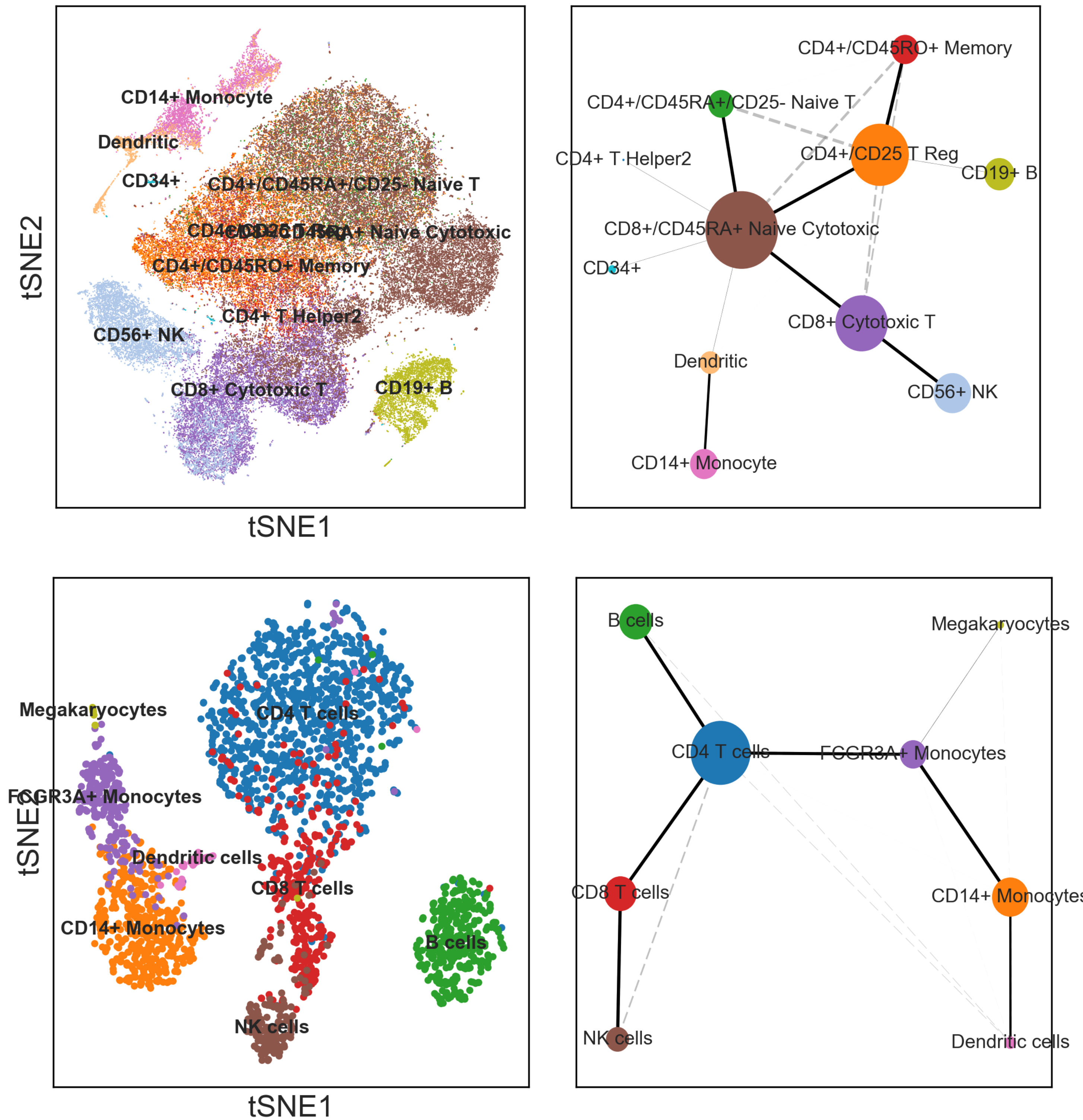
Learn where data is connected

Partial reconstruction of the PBMC lineage tree.

Only motifs can be recovered as data mostly consists of differentiated cells.

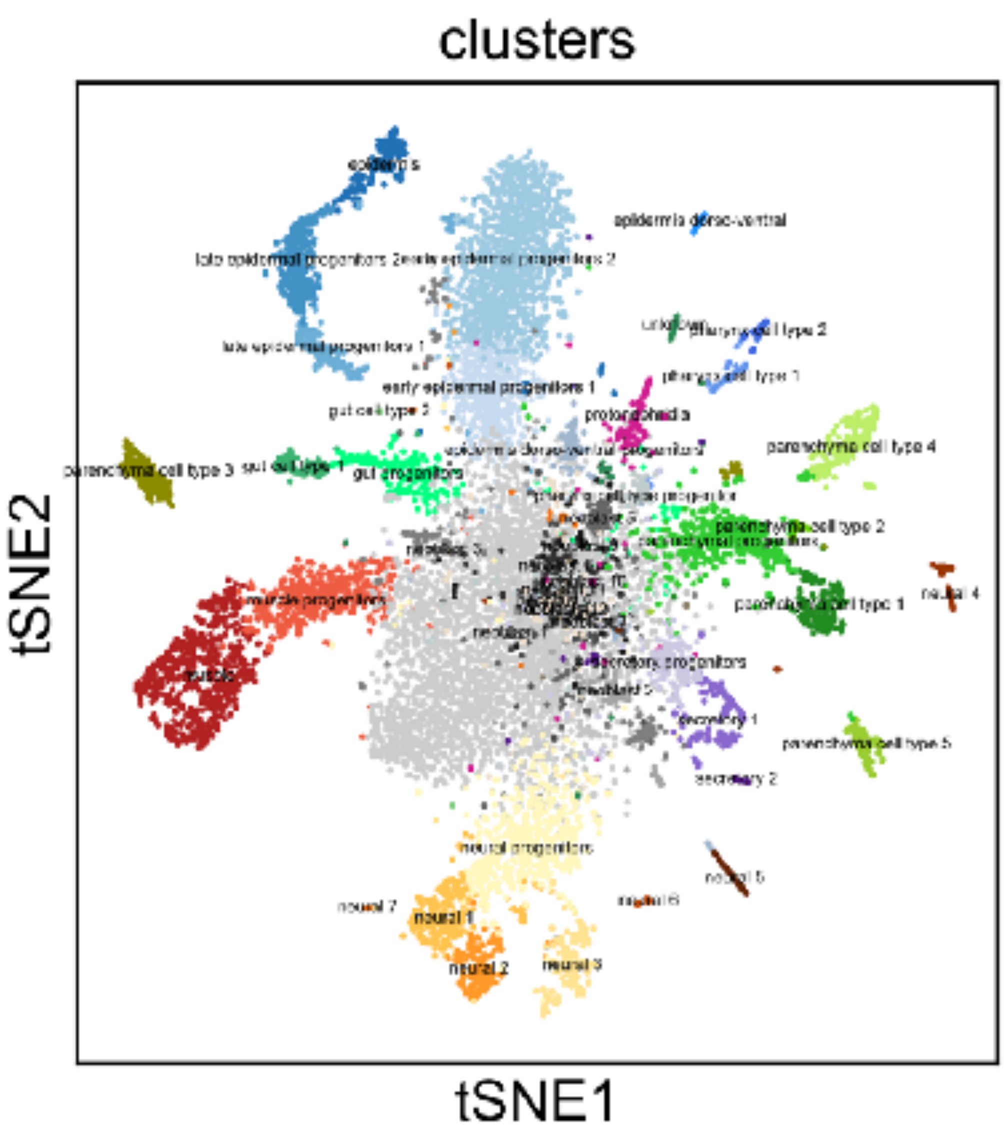
data for 68k cells from Zheng *et al.*, Nat. Comms. (2017)

data for 3.6k cells from 10X Genomics

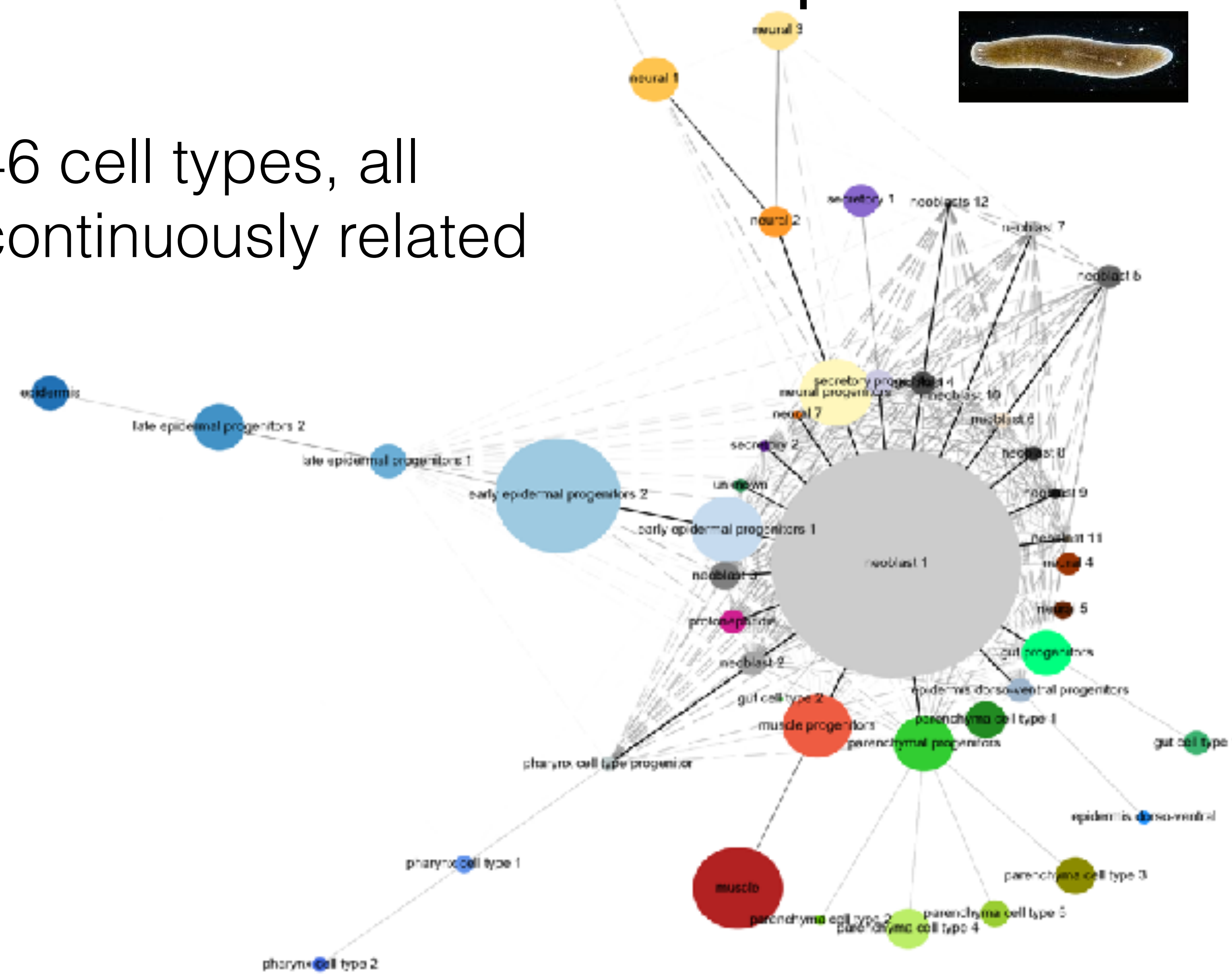


Plass, Solana, …, Rajewski, unpublished (2017)

Lineage tree of an adult animal: planaria



46 cell types, all continuously related



- Likely candidate for differentiation tree within abstracted graph
- Key genes during differentiation by following paths in abstracted coordinate system

Thanks to

Helmholtz Munich, Theis Lab
Sophie Tritschler
Lukas Simon
Fabian Theis

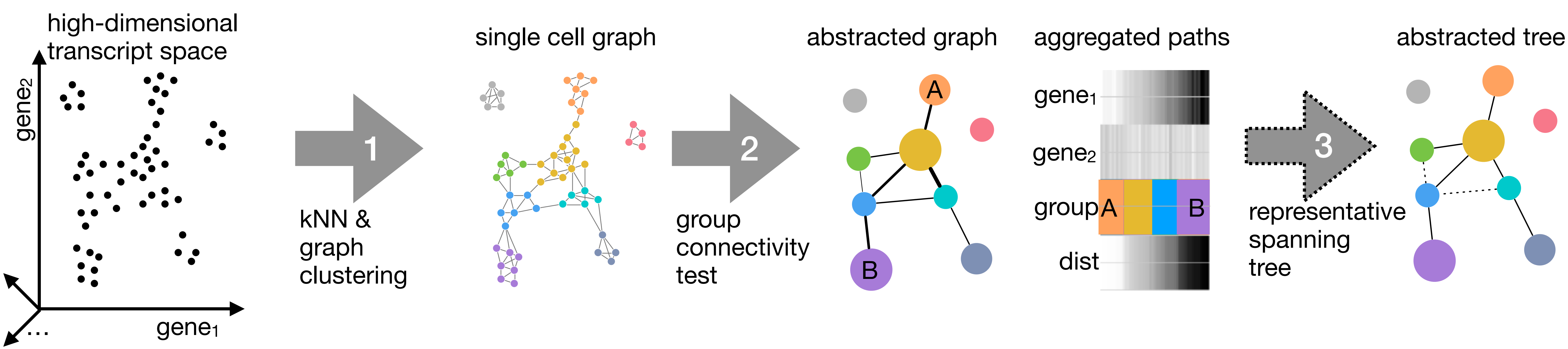
HelmholtzZentrum münchen
Deutsches Forschungszentrum für Gesundheit und Umwelt

Cambridge U, Göttgens Lab
Fiona Hamey

MDC Berlin, Rajewski Lab
Mireya Plass
Jordi Solana
Nikolaus Rajewski

HELMHOLTZ ASSOCIATION

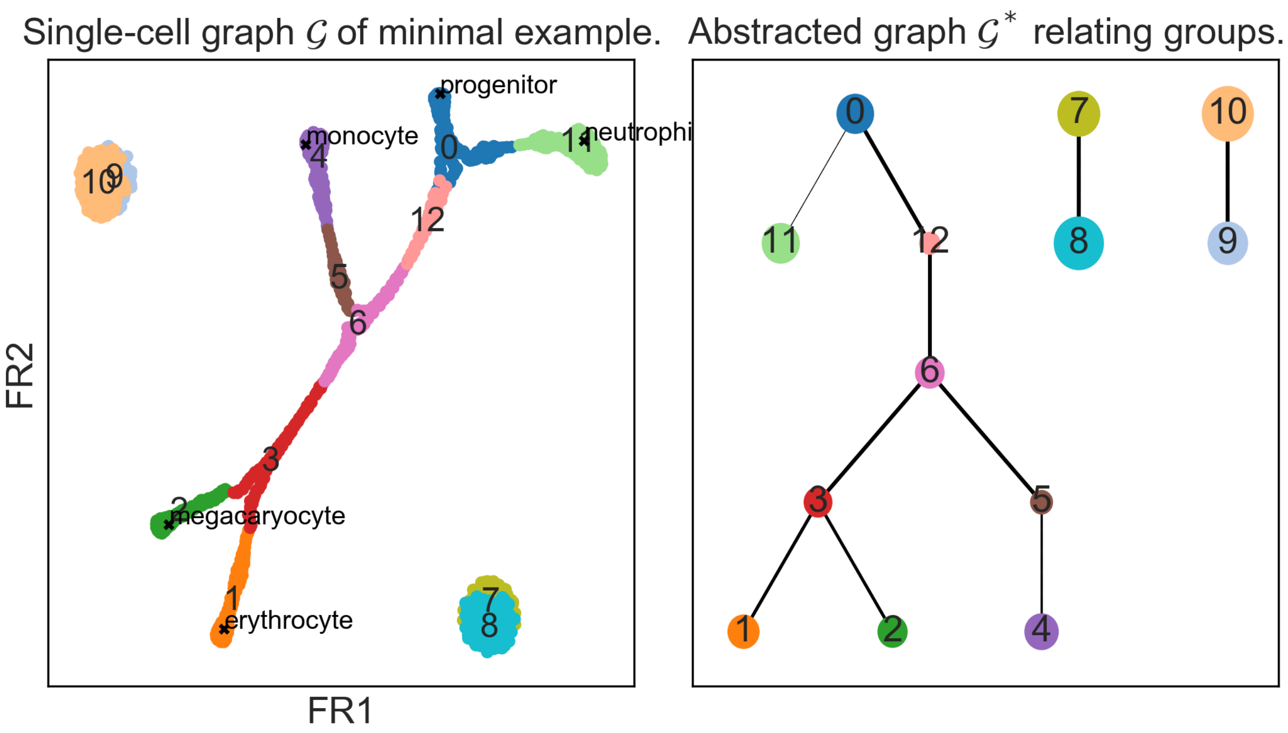
Thank you for your attention!

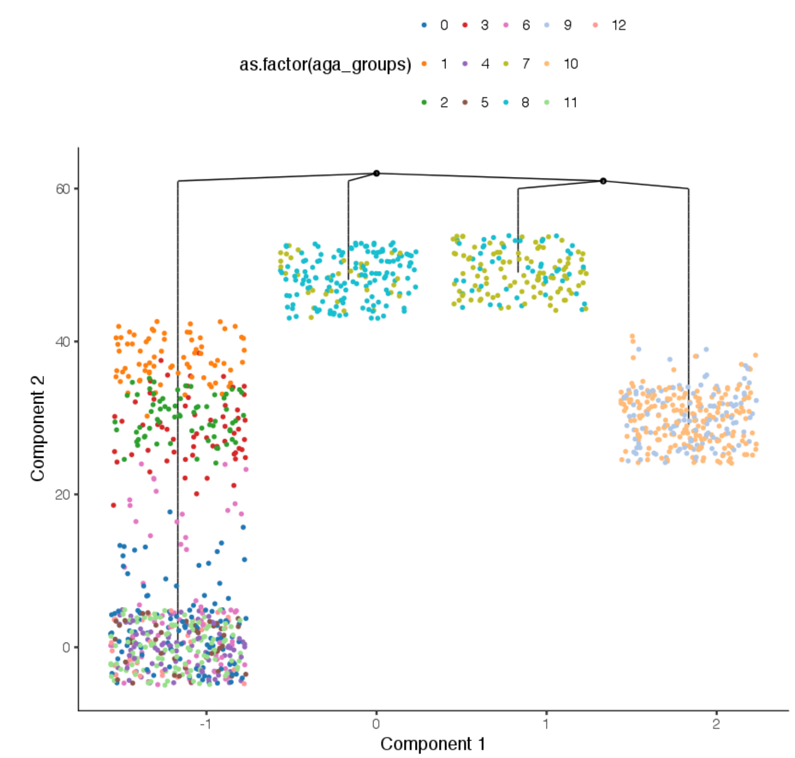


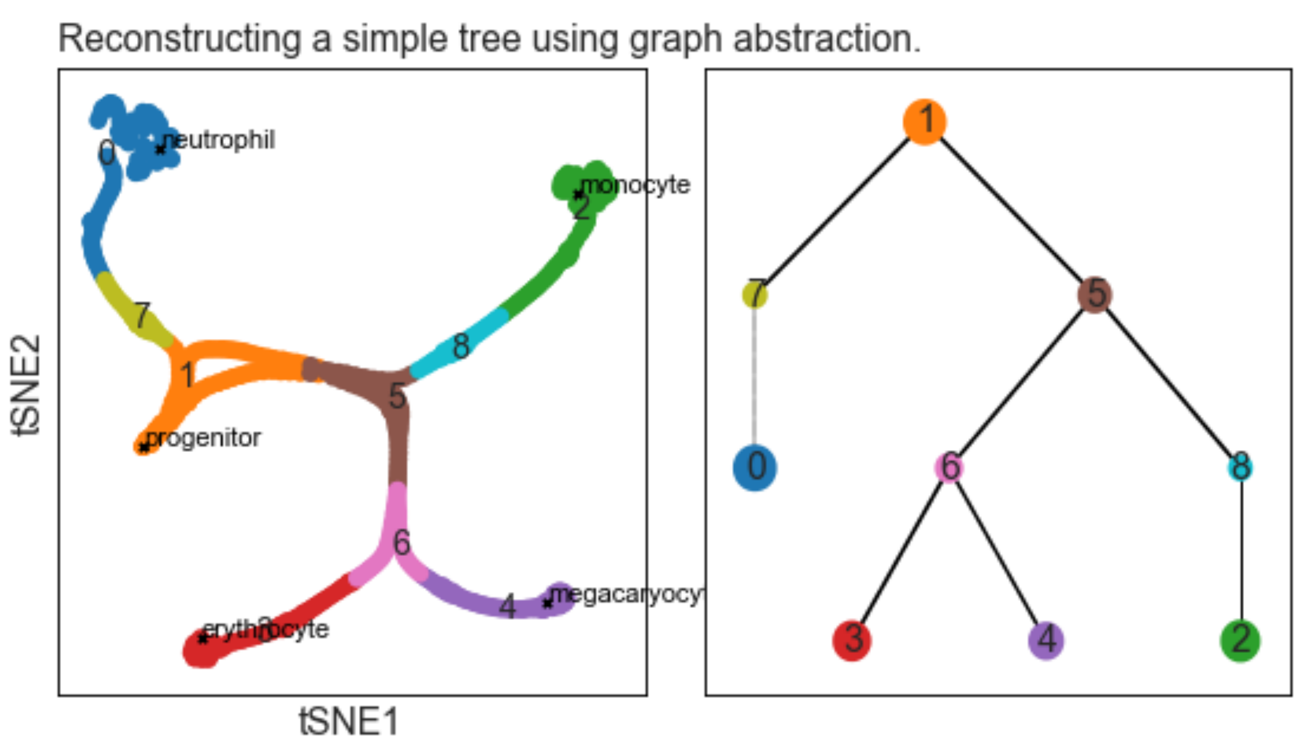
Code and documentation: https://github.com/theislab/graph_abstraction
On bioRxiv within the next days

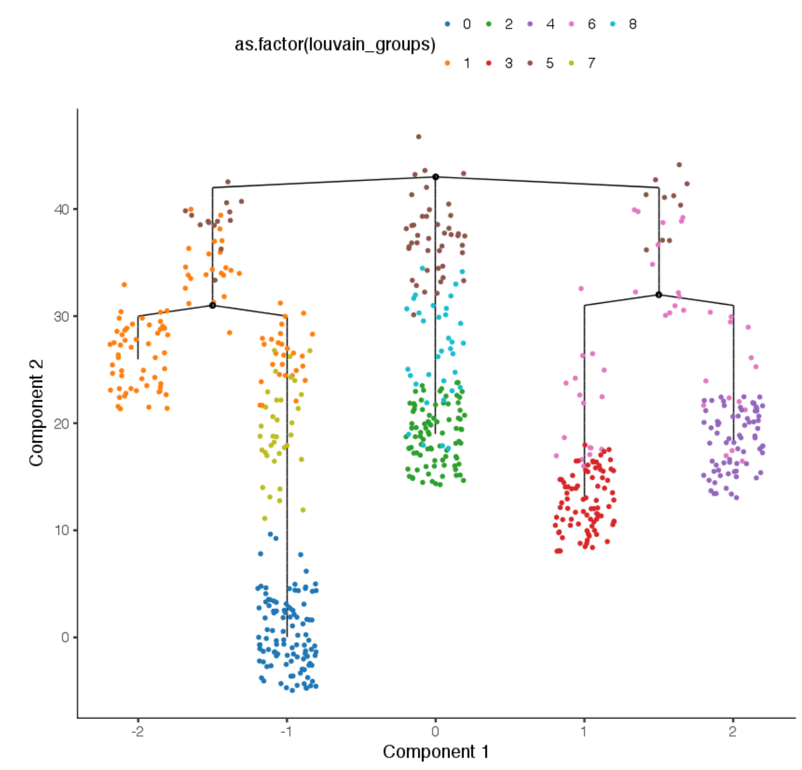
Trapnell et al., Nat. Biotech. (2014)
Qiu et al., Nat. Meth. (2017)

Comparison with Monocle









Comparison with stemID

Grün et al., Nature (2016)
Grün et al., Cell Stem Cell (2017)