Garvan Institute of Medical Research

Trajectory analysis in single cells

Associate Professor Joseph Powell
Director, Garvan-Weizmann Centre for Cellular Genomics
Deputy Director, UNSW Cellular Genomics Futures Institute

SISG - 2019

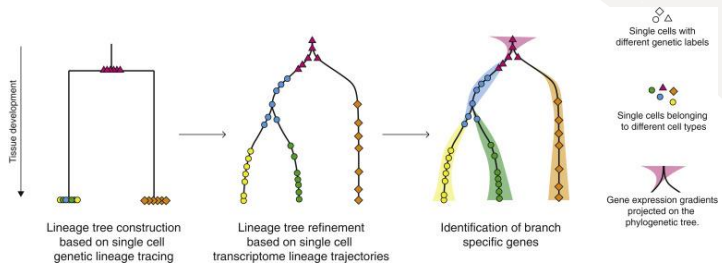
What is a trajectory?

Garvan Institute of Medical Research

- Reconstructing lineage relationships between cells within a tissue or organism is a long-standing aim in biology.
- Traditionally, lineage tracing has been achieved through the (genetic) labelling of a cell followed by the tracking of its offspring.
- Currently, lineage trajectories can also be predicted using single-cell transcriptomics. Although single-cell transcriptomics provides detailed phenotypic information, the predicted lineage trajectories do not necessarily reflect genetic relationships

What is a trajectory?

- Reconstructing lineage relationships between cells within a tissue or organism is a long-standing aim in biology.
- Traditionally, lineage tracing has been achieved through the (genetic) labelling of a cell followed by the tracking of its offspring.
- Currently, lineage trajectories can also be predicted using single-cell transcriptomics. Although single-cell transcriptomics provides detailed phenotypic information, the predicted lineage trajectories do not necessarily reflect genetic relationships



This comes back to the concepts of cell type vs cell state..

Cell types

- Function
- Morphology
- Set of molecular constituents
- End or start point of differentiation
- No interconversions possible

Examples:

- T-cell, hepatocyte, epithelial cell

Cell states

- Related but distinct functions
- Partially overlapping sets of molecular constituents
- Interconversions possible
- Transitional versus cell states

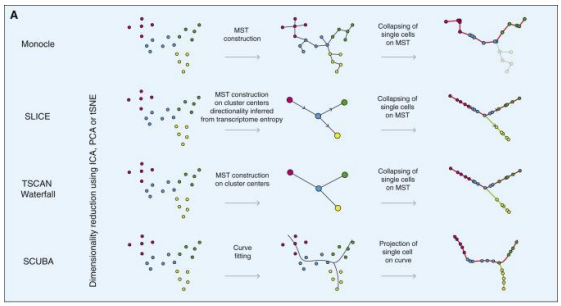
Examples:

- Cell cycle

4

Lineage reconstruction

Garvan Institute of Medical Research



Lineage reconstruction algorithms based on dimensional reduction.

Monocle uses independent-component analysis, followed by the construction of a minimum spanning tree (MST) connecting all cells. Large side branches are excluded, and remaining cells are projected onto the pseudotime backbone.

SLICE constructs a MST of cluster centres, and directionality is inferred from transcriptome entropy. Single cells are projected on the edges connecting the cluster centers.

TSCAN and Waterfall also constructs a MST based on the cluster centers, followed by projection of the single cells onto the edges to align cells in pseudotime.

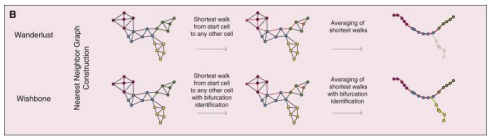
SCUBA uses tSNE for dimensionality reduction followed by the fitting of a smooth curve. Single cells are projected on the smooth curve to order them in pseudotime.

All require user input to infer directionality

5

Lineage reconstruction

Garvan Institute of Medical Research



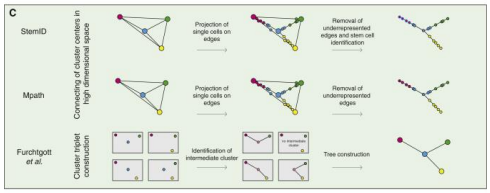
Lineage reconstruction algorithms based on NNGs.

Both Wanderlust and Wishbone start with the construction of a NNG. A collection of shortest walks, from a user-defined root cell to all other cells in the graph, is then used to construct the lineage trajectory. Wishbone has the added benefit that it can identify bifurcations in the lineage trajectory.

6

Lineage reconstruction

Garvan Institute of Medical Research



Lineage reconstruction algorithms based on cluster networks.

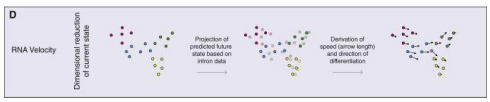
Both StemID and Mpath start by connecting all cluster centers in a high dimensional space. Single cells are then projected on the edges between the clusters, and underrepresented edges are removed from the graph.
StemID identifies a potential stem cell population based on transcriptome entropy.

The Furchgott method infers the intermediate cluster (if possible) from each triplet of clusters in the data, followed by tree construction based on the triplet relations.

7

Lineage reconstruction

Garvan Institute of Medical Research



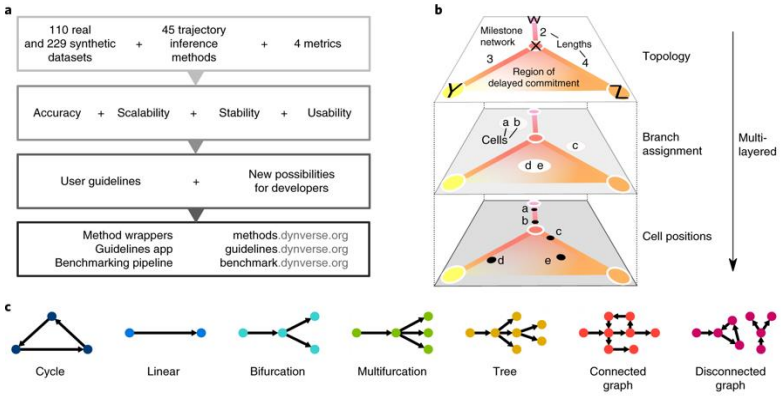
Graphical representation of RNA velocity.

Current and future state of the cell are computed based on exon and intron data, respectively.
Difference between current and future state determines the direction and speed of differentiation.
No user input is required to infer these parameters

8 Manno et al. Nature 2018

Comparing trajectory methods

Garvan Institute of Medical Research



a schematic overview of our evaluation pipeline.
b To make the trajectories comparable to each other, a common trajectory model was used to represent reference trajectories from the real and synthetic datasets, as well as any predictions of TI methods.
c Trajectories are automatically classified into one of seven trajectory types, with increasing complexity

9

Methods comparison

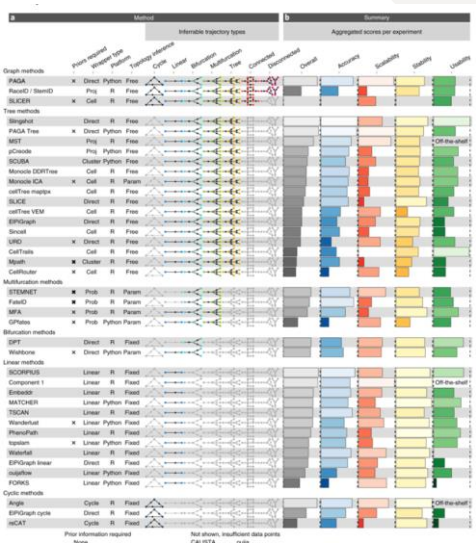
Garvan Institute of Medical Research

Method	SCUBA pseudotime	Wanderlust	Wishbone	SLICER	SCOUP	Waterfall	Mpath	TSCAN	Monocle	SCUBA
Visual abstract										
Structure	Linear	Linear	Single bifurcation	Branching	Branching	Linear	Branching	Linear	Branching	Branching
Robustness strategy	Principal curves	Ensemble, starting cell	Ensemble, starting cell	Starting cell	Starting population	Clustering of cells	Clustering of cells using external labelling	Clustering of cells	Differential expression	Simple model
Extra input requirements	None	Starting cell	Starting cell	Starting cell	Starting population	None	Time points	None	Time points	Time points
Unbiased	+	±	±	±	±	+	−	+	−	−
Scalability w.r.t. cells	−	−	±	±	−	±	+	+	−	±
Scalability w.r.t. genes	+	+	+	+	−	+	±	±	±	+
Code and documentation	−	±	+	±	+	±	+	+	+	±
Parameter ease-of-use	+	+	+	+	−	±	−	+	+	+

10 Cannoodt et al. Eur J Immunology 2016

Methods comparison

Garvan Institute of Medical Research

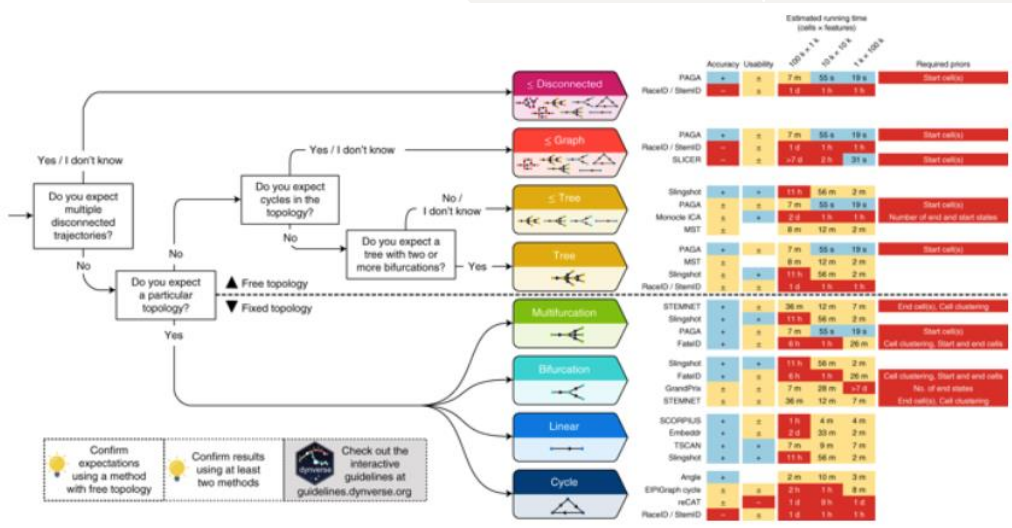


Take home message: most methods are situation specific

11 Saelens et al. Nat Biotech 2019

How to decide on the best method?

Garvan Institute of Medical Research



12 Saelens et al. Nat Biotech 2019