Inference of multiple trajectories in single-cell RNA-Seq data from RNA-velocity

Ziqi Zhang[1], Xiuwei Zhang[1,*]
[1]Georgia Institute of Technology, Computational Science and Engineering, Atlanta, GA

Introduction

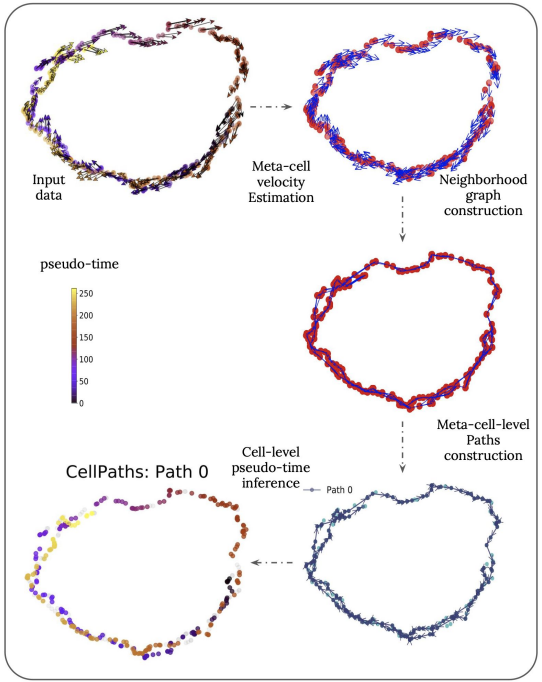
The availability of large scale single-cell RNA-Sequencing (scRNA-Seq) data allows researchers to study the underlying mechanisms that drive the change of cells within dynamic processes such as stem cell differentiation and cancer cell development. Trajectory inference (TI) methods are often used to infer the trajectory of this dynamic process, namely, assign developmental lineages and pseudo-time for every cell.

Most of the current TI methods infer cell developmental trajectories based on the transcriptome similarity between cells, using only scRNA-Seq data. The disadvantages of these methods are: 1) a method is often restricted to certain trajectory structures like trees or simple cycles; 2) directions of the trajectory cannot be inferred and the root cell is often required as a prior.

The recent surge of RNA-velocity estimation methods has opened up a new perspective for trajectory inference of cells. RNA-velocity provides a short-term prediction of gene expression profile in each cell by incorporating unspliced mRNA counts, the prediction can be used as directional information for trajectory inference.

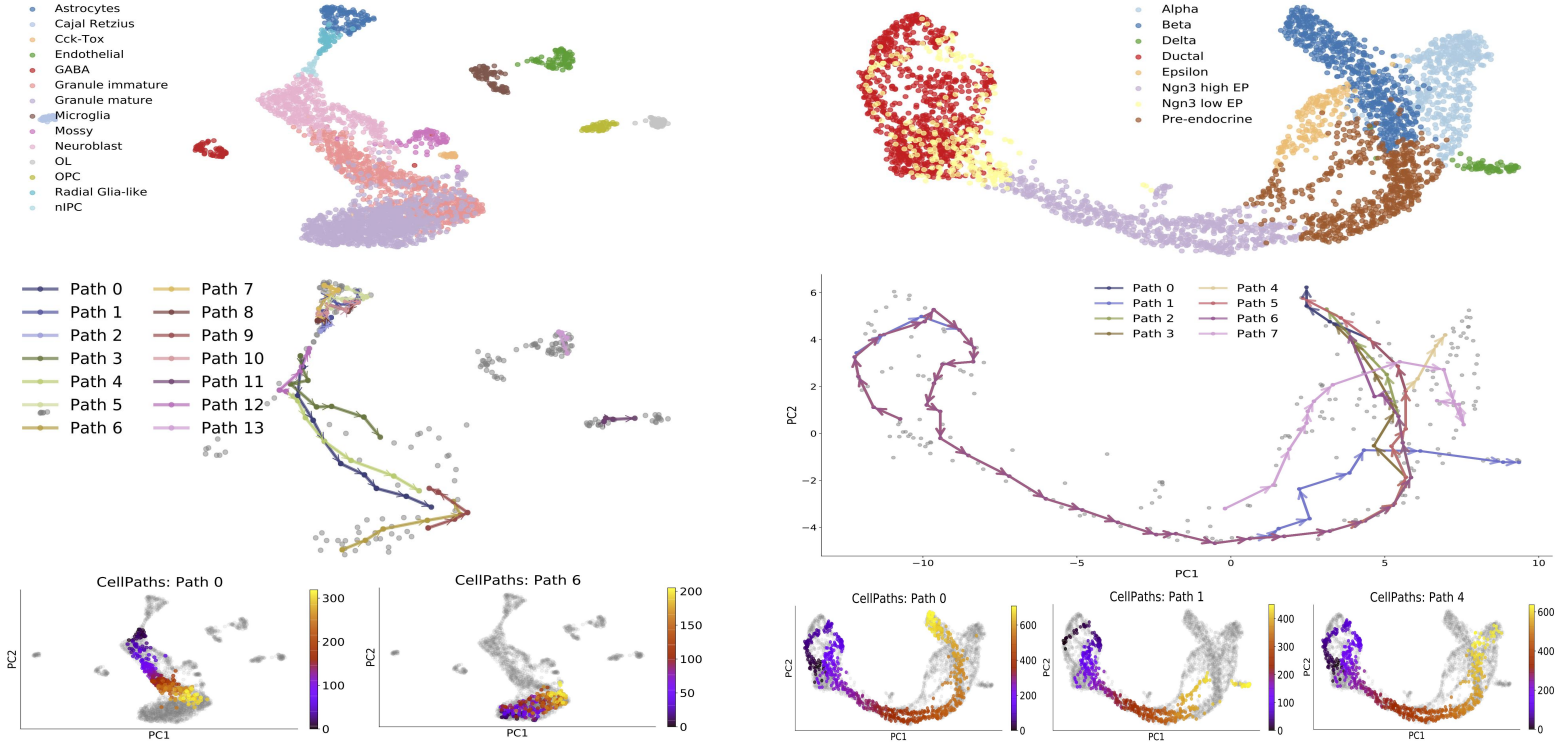
We present CellPath, a single cell TI method that infers multiple high-resolution developmental trajectories by integrating RNA velocity information.

Schematics of CellPath



1. CellPath first constructs meta-cells and infers meta-cell gene expression and RNA-velocity values to accommodate the noise in the original data.
2. CellPath then builds a neighborhood graph of the meta-cells
3. CellPath adopts shortest path algorithms to find possible paths in the graph and uses a greedy algorithm to select the final paths as the meta-cell level paths on the neighborhood graph.
4. CellPath assigns cell level pseudo-time.

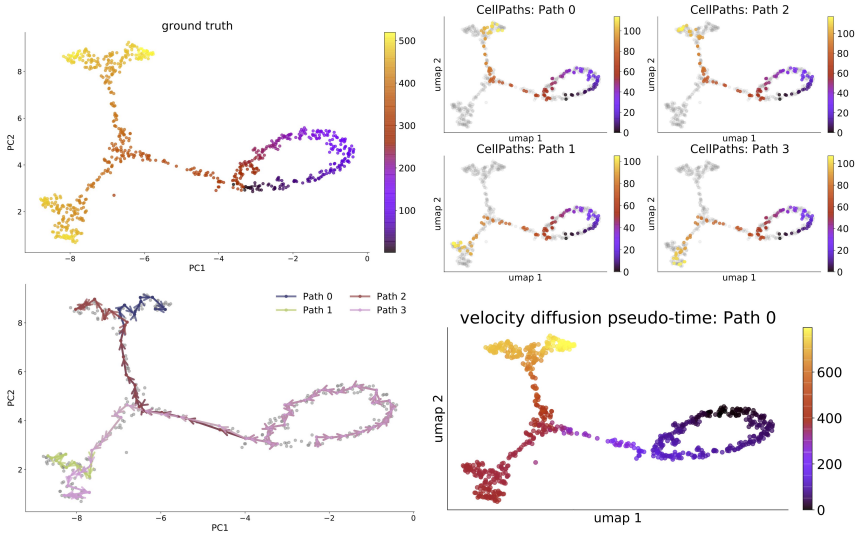
CellPath on real dataset



We test CellPath on two real datasets: dentate-gyurs dataset(left) and pancreatic-endocrinogenesis dataset (right), The result shows that CellPath robustly infer the accurate branching structure for data of different trajectory topologies.

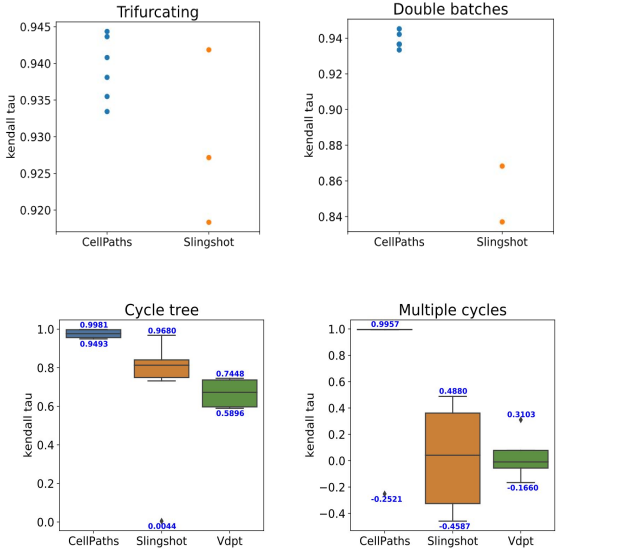
CellPath accurately infers multiple branching lineages starting with different root cells in dentate-gyrus dataset. CellPath discovers multiple lineages that correspond to different endocrine cell sub-types genesis processes in pancreatic-endocrinogenesis dataset.

CellPath on simulated dataset



We test CellPath on a cycle-tree structured simulated data. CellPath accurately find all four sub-branches and the cell-cycle structure within the dataset, Whereas Slingshot wrongly infers it as a trifurcating structure and velocity diffusion pseudo-time cannot discover the cell-cycle within the structure.

CellPath is available as a python package at:
https://github.com/PeterZZQ/CellPath



	Trifurcating	Double batches
CellPath	**1.514**	**1.745**
Slingshot	1.568	2.014

We benchmark CellPath on multiple simulated datasets of different trajectory topologies, and we measure

- The pseudo-time assignment accuracy using Kendall rank correlation coefficient (boxplot on above)
- Trajectory assignment by average entropy (table above, lower values are better).

The results show that CellPath perform better in both pseudo-time and branch assignment.