

Abstract: Decoding temporal order and lineage choice of asynchronously differentiating cells from single cell gene expression data

Single-cell technologies have recently gained popularity in developmental biology because they allow resolving potential heterogeneities due to asynchronicity of differentiating cells. Popular multivariate approaches for analyzing such data are based on data normalization, followed by dimension reduction and clustering to identify subgroups. However, in the case of cellular differentiation, we cannot expect clear clusters to be present - instead cells tend to follow continuous branching lineages.

We show that modeling the high-dimensional state space as a diffusion process, where cells move to close-by cells with a distance-dependent probability well reflects the differentiating characteristics. Based on the underlying diffusion map transition kernel, we then propose to order cells according to a diffusion pseudo time, which measures transitions between cells using random walks of arbitrary length. This allows for a robust identification of branching decisions and corresponding trajectories of single cells. We demonstrate the method on single-cell qPCR data of differentiating mouse haematopoietic stem cells as well as on RNA sequencing profiles of embryonic stem cells.