Alignment of single-cell trajectory trees with CAPITAL

Nat Commun 13, 5972 (2022). https://doi.org/10.1038/s41467-022-33681-3

Reiichi Sugihara, Yuki Kato, Tomoya Mori & Yukio Kawahara

Single-cell RNA sequencing (scRNA-seq) provides large-scale, complex gene expression data at the single-cell level. To better understand cellular and tissue characteristics, methods that can accurately and efficiently analyze such data are increasingly needed. Pseudotime analysis, which reconstructs cellular state transitions, has become an important approach for studying gene expression changes and fate decisions during cell differentiation. Comparing differentiation processes across multiple datasets can reveal how different conditions or diseases affect gene expression and may lead to new insights into disease mechanisms and therapies.

However, most conventional methods can only handle simple, linear differentiation trajectories. In practice, scRNA-seq data often show complex branching trajectories, which are difficult to analyze with existing approaches.

In this study, we developed a method to compare differentiation trajectories obtained from multiple scRNA-seq datasets. First, we estimate the differentiation trajectory for each dataset using clustering and pseudotime analysis. Then, we apply a tree alignment algorithm to align the branching trajectories between datasets. This allows detailed comparison of cell differentiation processes and gene expression changes under different conditions.

We evaluated our method using synthetic scRNA-seq datasets and showed that it achieved higher accuracy and robustness than previous methods. We also applied it to public scRNA-seq datasets of human and mouse bone marrow cells, which include hematopoietic stem cells, progenitor cells, and cells of the lymphoid and myeloid lineages. Our method successfully aligned the complex differentiation trajectories and identified gene sets with similar or different expression patterns between species.

This method enables the comparison of complex branching differentiation trajectories across datasets, which has been difficult with conventional approaches. It provides a useful tool for studying cellular differentiation and gene expression dynamics and is expected to contribute to many fields using scRNA-seq data.