

2024年2月5日

時系列 scRNA-seq データから細胞分化ダイナミクスを推定 - 包括的軌跡推定フレームワーク scEGOT の開発 -

概要

ヒトの発生過程において、細胞がどのように特殊化された細胞型に変化するかを理解することは、生物学における最重要課題の一つです。この複雑なプロセスは「細胞分化」と呼ばれ、発生生物学や再生医療の進歩において重要な鍵を握っています。近年大きく発展した1細胞遺伝子発現解析技術（scRNA-seq）は、1つの細胞が持つすべての遺伝子発現情報を取得する手法で、発生生物学の分野でも広く普及しています。しかしながら、一つの実験から得られる膨大なデータから、目的とする動的現象に関連する情報を正確に取得することは、大きな課題の一つでした。

京都大学高等研究院ヒト生物学高等研究拠点（WPI-ASHBi）谷地村敏明 特定研究員（研究当時、現：東北大学 数理科学共創社会センター 助教）、平岡裕章 副拠点長/主任研究者らの研究チームは、正規化混合ガウス最適輸送理論に基づく、新しい包括的軌跡推定フレームワーク「single-cell trajectory inference framework based on entropic Gaussian mixture optimal transport (scEGOT)」を開発しました。scEGOT は従来の軌跡予想の手法の限界を克服し、単一細胞が異なる細胞に分化する動的な過程を解析する新たな方法を提供します。

本成果は、2024年12月23日に国際学術誌「BMC Bioinformatics」に掲載されました。

**正規化混合ガウス最適輸送理論 (EGOT) に基づいて、
時系列 scRNA-seq データから細胞分化ダイナミクスを推定する
包括的軌跡推定フレームワーク scEGOT を開発**

従来方法と比較して、高い生物学的解釈性を保ったまま非常に低い計算コストで細胞分化ダイナミクスを推定する生成モデルを構築可能
データ駆動型の手法であるため、scATAC-seq データなどの他の単一細胞データにも適用可能

① Cell
② RNA extraction
③ Library prep
④ Sequence
⑤ Time series scRNA-seq data

✓ Cell state graph
✓ Velocity
✓ Interpolation
✓ Animation (dynamics)
✓ Gene regulatory network
✓ Landscape

<https://github.com/yachimura-lab/scEGOT>

©Kanon Tanaka

本研究では、ヒト iPS 細胞から始原生殖細胞様細胞 (hPGCLCs) ^{<注1>}を誘導するシステムに焦点を当て、各誘導段階での scRNA-seq データを用いて、包括的に軌跡推定を行う新規手法の開発を試みました。従来の手法では、「観測時点間の細胞の微細な動的変化を捉えることが難しい」「複雑なアルゴリズムにより膨大な計算資源を必要とする」「ニューラルネットワークに代表されるようなブラックボックスアプローチでは軌跡推定の生物学的解釈が理解しづらい」などの課題がありました。この課題を克服するために、混合ガウスモデル^{<注2>}と最適輸送理論^{<注3>}に基づく包括的軌跡推定フレームワーク scEGOT を開発しました。scEGOT は時系列 scRNA-seq データから、細胞分化に関する状態グラフ、速度場、時間補間 (scRNA-seq 間の中間状態生成や遺伝子発現アニメーション)、動的遺伝子制御ネットワークの推定、ワディントンランドスケープの再構築など、軌跡推定のための多目的なツールを提供し、高い生物学的解釈可能性と計算効率を兼ね備えています。研究チームは、hPGCLCs 誘導システムから得られた時系列 scRNA-seq データに対して scEGOT を用いて解析することにより、生殖細胞の前駆細胞集団を特定しました。また、これらの細胞が体細胞系列 (組織や臓器を構成する細胞) と分岐するタイミングを明らかにしました。さらに、*TFAP2A* や *NKX1-2* といった遺伝子が生殖細胞の分化に重要な役割を果たしていること、*MESP1* や *GATA6* が体細胞系列の分化で重要であることを見出しました。さらに、一般的なグラフや速度場の表記だけでなく、遺伝子発現アニメーションやワディントンランドスケープの様なデータ表現にも対応しており、視覚的かつ直感的に細胞分化の動的過程を理解することが可能とする包括的な軌跡推定フレームワークとなっています。scEGOT の開発により、今後はこれまで難しかった細胞集団の分化動態の推定が進展することが期待されます。

また、この研究は数学と生物学が学問の垣根を超えて協働することで、科学的課題に取り組む重要性を示しています。

<研究者のコメント>

scEGOT は多用途なデータ駆動型の手法であり、scRNA-seq に限定されることなく、scATAC-seq などの他の単一細胞データにも適用可能です。エピジェネティックなメカニズムを解明し、細胞分化や系統決定の理解を深めることで、発生生物学の研究がさらに躍進することを期待しています。(谷地村敏明)

<用語解説>

1. 始原生殖細胞様細胞 (Primordial Germ Cells Like Cells, PGCs) : iPS 細胞や ES 細胞などの多能性幹細胞を起点として、成長因子や小分子化合物を用いて培養ディッシュ上で誘導することができる始原生殖細胞 (PGC) によく似た性質を持つ細胞のこと。PGC はヒト胚では受精後 2 週間目に形成される、最も未分化な生殖細胞。
2. 混合ガウスモデル : 複数の正規分布 (ガウス分布) の混合によってデータが生成されると仮定する確率モデルである。各正規分布はデータのサブグループ (クラスター) を表し、それぞれの平均、分散、混合比率というパラメータによって特徴付けられる。クラスターリングや密度推定などに広く用いられる。
3. 最適輸送理論 : 確率測度に関する変分問題で、確率測度間における距離や最適なマッチングを与える数学理論。

< 論文書誌情報 >

Yachimura, T., Wang, H., Imoto, Y., Yoshida, M., Tasaki, S., Kojima, Y., Yabuta, Y., Saitou, M., & Hiraoka, Y. (2024). scEGOT: single-cell trajectory inference framework based on entropic Gaussian mixture optimal transport. *BMC Bioinformatics*. <https://doi.org/10.1186/s12859-024-05988-z>