

# 正則化混合ガウス最適輸送理論による 細胞分化の軌跡推定手法の構築

谷地村 敏明\*

Development of a Trajectory Inference Framework  
by Entropic Gaussian Mixture Optimal Transport

Toshiaki YACHIMURA\*

This study proposes a novel mathematical model and analysis method called “scEGOT” based on optimal transport theory to quantitatively reconstruct the dynamics of the cell differentiation process. By extending the entropic Gaussian mixture optimal transport theory, we captured dynamic information such as the speed and intermediate states of gene expression, which were difficult to extract using conventional cell lineage inference methods based on scRNA-seq data. Moreover, it can also reconstruct Waddington’s epigenetic landscape and infer the underlying gene regulatory networks. When applied to time-series data from iPS cells induced to form primordial germ cell-like cells (PGCLCs), our approach identified not only known marker genes but also novel genes including *NKX1-2*, as well as groups of genes involved in inhibiting differentiation into PGCLCs.

## 1. 背景・目的

幹細胞（例えばES細胞やiPS細胞）が特定の機能をもつ細胞（神経、筋肉、血液など）に変化する過程のことを細胞分化という。全ての細胞は遺伝子の発現量によって特徴づけられ、遺伝子間の相互作用（遺伝子制御ネットワーク）により遺伝子の発現量が互いに制御されることで分化が進行する。この細胞分化のダイナミクスに関する概念的モデルとして、1957年にC. H. Waddingtonによってエピジェネティックランドスケープが提唱された<sup>1)</sup>。この細胞分化の概念モデルにおける各々の谷は細胞種を表しており、多能性を持つ細胞が谷を下っていくことでより特化した細胞種へと分化していく。このランドスケープの形状は遺伝子制御ネットワークによって形成される（図1参照）。細胞分化過程においてこのランドスケープを詳細に知ることができれば、細胞分化を予測・制御することが可能となり、生命の発生過程における原理解明だけでなく再生医療等への臨床的应用も期待できる。一方で、このランドスケープ描像は単なる比喻であり、実際の細胞分化においてどのような形状を成しているか、及びその形状を形成している背景の遺伝子制御ネットワークは自明ではない。

近年、計測技術の発展に伴い、単一細胞の遺伝子発現を網羅的に計測する単一細胞RNAシーケンス（scRNA-seq）技術が注目されている。

この技術により組織における一つ一つの細胞が持つ遺伝子発現量を得ることができ、出力としては細胞数×遺伝子数の行列データが得られる。一般に遺伝子数は数万程度（ヒトであれば約2万個）であるためscRNA-seqデータは超高次元のビッグデータであり、細胞集団はこの超高次元データの空間（遺伝子発現空間）における点群とみなすことができる。scRNA-seqデータを用いることにより、遺伝子発現空間においてこのWaddingtonのランドスケープをデータ駆動的に再構成できる土壌が整いつつあり、細胞系譜推定問題として現在の発生生物学やシステム生物学における中心的話題の一つとして盛んに研究されている。

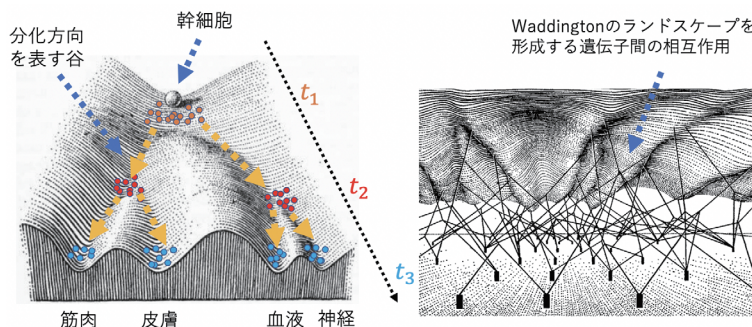


図1 Waddingtonのエピジェネティックランドスケープ（左）と遺伝子制御ネットワーク（右）の概念図。

2025年3月5日 受理

\* 豊田理研スカラー

東北大学数理学共創社会センター

これまで知られている細胞系譜推定の手法の多くは単一の scRNA-seq データに対して設計されたものであり、非線形次元削減とグラフ理論に基づき遺伝子発現空間における個々の細胞に対して遺伝子発現の類似性から分化の順序を推定するものであった。一方で、単一の scRNA-seq データに基づく細胞系譜推定では細胞分化の過程全体におけるダイナミクス（例えば分化の速度や中間状態）を推定することは困難であった。

そこで本研究では、確率分布間における距離や最適マッチングを提供する数学理論である最適輸送理論を基として、時系列 scRNA-seq データに対して細胞分化のダイナミクスを推定し、遺伝子発現空間における Waddington のエビジェネティックランドスケープを再構成できる系譜推定の数理基盤の構築を目的とした。また、構築された数理に基づき細胞分化のダイナミクスを推定するソフトウェアを開発し、iPS 細胞から始原生殖細胞への誘導系等の実際の時系列 scRNA-seq データに適用することで、それらの細胞分化機構を司る重要遺伝子や遺伝子制御関係の解明について取り組んだ。

## 2. 結果

正則化混合ガウス最適輸送に関する理論的側面を進展させ、その理論に基づく時系列 scRNA-seq データに対する細胞分化ダイナミクス推定手法の構築を行った。また、この理論に基づく軌跡推定ソフトウェア scEGOT (single-cell trajectory inference framework based on Entropic Gaussian mixture Optimal Transport) を開発した<sup>2)</sup> (ソフトウェアについては <https://github.com/yachimura-lab/scEGOT> 参照)。scEGOT を用いることにより、従来の軌跡推定手法が構築してきた細胞分化の状態グラフだけでなく、各細胞の細胞分化に伴う遺伝子発現の速度 (cell velocity) や動態 (アニメーション) を推定することが可能となる。また、推定された速度場に基づいて遺伝子発現空間における分化ポテンシャル推定 (Waddington ランドスケープの再構成) や Waddington ランドスケープを形成する背景の遺伝子制御ネットワークの推定をも行うことができる (図2参照)。

iPS 細胞から始原生殖細胞を誘導する時系列 scRNA-seq データに対して scEGOT を適用することにより、*TFAP2A* や *NKX1-2* といった遺伝子が始原生殖細胞の前駆細胞におけるマーカー遺伝子であることを発見した。*TFAP2A* は近年始原生殖細胞の前駆細胞におけるマーカー遺伝子であることが報告されており注目されつつある遺伝子の一つだが、*NKX1-2* と始原生殖細胞との関連はこれまで全く報告されておらず、本データ解析によって知り得た新規遺伝子である。さらに、*GATA6* や *MESPI* といった遺伝子が始原生殖細胞系列への分化を阻害していることも発見した。

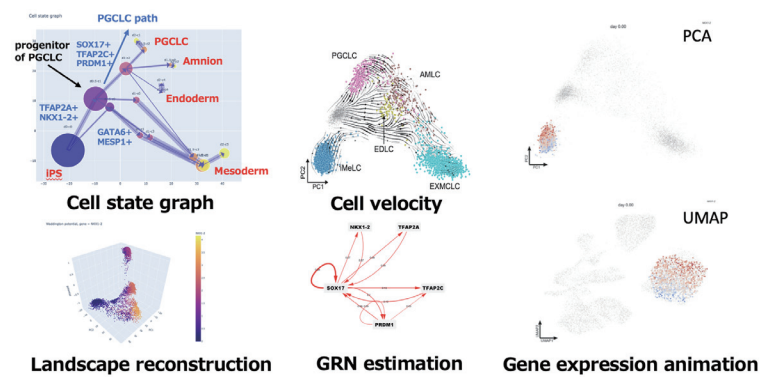


図2 scEGOT の時系列 scRNA-seq データへの適用。

## 3. 今後の展望

scEGOT が scRNA-seq データに限らないデータ駆動型手法である利点を活かして、本手法の適用範囲を広げていく予定である。特に単一細胞のクロマチン情報を網羅的に計測する scATAC-seq データのような他の単一細胞データへの適用について考えていく。また、scRNA-seq データのみを用いて細胞の空間位置を再構成する novoSpaRc などの遺伝子発現カートグラフィー手法と本手法を組み合わせることで、3次元における組織や器官の形成過程における遺伝子の時空間発現パターンのダイナミクスを推定するアルゴリズムの構築を考えていきたい。

## REFERENCES

- 1) C. H. Waddington, *London: George Allen & Unwin, Ltd.*, 1957.
- 2) T. Yachimura, *et al.*, *BMC Bioinformatics*, **25** (2024) 388.