

研究終了報告書

「移流拡散過程に基づく環境依存的細胞状態ダイナミクスの推定」

研究期間：2020年12月～2023年3月

研究者：小嶋 泰弘

加速フェーズ期間：2023年4月～2024年3月

1. 研究のねらい

数理モデルは、科学の幅広い諸分野で様々な現象を記述し、その定量的な理解をするにあたり重要な役割を担ってきた。しかしながら、特に生命現象を解析する場合には、モデルの構造を定める過程のみならず、数理モデルを構成するパラメータの値も多くの場合で直接の観測が難しい場合が多い。このような問題に対して、研究者は、大量観測データの確率的モデリングと逆推定により得られた数理モデルの解析というアプローチを、様々な種類の細胞がダイナミックな組織形態の変形とともに誕生する胚発生の過程に応用することで、観測することが困難な遺伝子発現の連続的な時空間パターンを推定する手法の開発をおこなった。この研究では、多様なデータに対する統合的確率モデリングと機械学習技術を用いた推定を行うことにより、それぞれの細胞が変化していくその時空間ダイナミクスを解析することが可能となり、特定の細胞状態への遷移がいつどこから始まるのか、といったこれまでにない解析を行うことが可能となった。一方で、胚発生の初期を除く多くの生命現象では、ほとんど同じ時空間プロセスが異なる個体で再現されることは少なく、ある程度の決まった構造を伴いながらも多様な細胞種が空間的にランダムな混在を呈する場合がある。その一方で、このようなある種のランダムネスは、状況に応じて画一的に変動していく平均的な動きのみからは生まれ得ないものであり、細胞自身が示す確率的なゆらぎを伴う挙動の存在が示唆される。それゆえ、このような多様な細胞種の混在が見られ、これまでに研究者が用いてきた理論的枠組みでは扱いが難しい場合について、適切な数理モデルを確立し、それを時空間的な遺伝子発現観測から逆推定することができれば、細胞の運命決定の中に内包される環境依存的な平均的動態とゆらぎを同時に抽出可能となることが期待される。

本研究では、近年発達の著しい AI 技術と研究者が開発した移流拡散モデルの最適化手法を複合的に運用することで、環境要因依存的な細胞状態ダイナミクスの推定手法を開発し、細胞状態を変更する環境要因と細胞のゆらぎのデータ駆動的推定を行う。さらに、推定されたモデルの前向き解析から、それぞれの要因が全体に及ぼす影響を予測する技術の開発も目的とする。これらの要因を外的介入により調整することで生体組織そのものの運命を定量的に制御可能な社会の到来を手繰り寄せる。

2. 研究成果

(1) 概要

本研究では、環境依存的な細胞状態のダイナミクスを移流拡散過程の形で推定する。これにより、環境因子により説明可能な細胞状態の運命分岐と細胞の内的なゆらぎにより生まれる分岐とを区別することを目指し、研究開発をおこなった。まず、第一段階として一細胞トランスクリプトームの観測データから、深層生成モデル技術の一つである変分自己符号化器とスプライシング数理モデルとを融合することで、潜在的な細胞状態の確率的なダイナミ



クスを推定するための枠組みを世界に先駆けて開発した。これにより、細胞状態の多様化が起きる段階では、細胞状態の時間変化の不確実性が大きく向上することが確認された。一方で、細胞トランスクリプトーム観測では、その細胞の環境情報は欠落してしまう欠点がある。そのため、本手法で明らかになった細胞状態の時間変化の不確実性が、内因的なゆらぎに基づくものであるのか、それとも細胞の置かれた環境のばらつきに由来するものであるのか、区別することは依然として困難であった。このような細胞の環境情報を補う上で有用と考えたのが、遺伝子の発現を空間的な座標情報を保持したまま観測することを可能にする空間トランスクリプトーム観測の技術である。ただ、現状普及が進む空間トランスクリプトーム技術では、一細胞の解像度がなく細胞状態の時間変化と紐付けた解析を行うことが困難であった。そこで、本研究では変分自己符号化器により得られる細胞状態に基づいて、細胞ごとの空間分布を復元、細胞の共局在関係を明らかにするための方法論の開発をおこなった。これらの結果は、本研究の成果が、一細胞トランスクリプトーム解析により明らかにされつつある多様な細胞状態が生成する過程をその環境要因や内因的なゆらぎに基づき理解するための基盤技術として有用であることを示している。

加速フェーズについて

加速フェーズにおいては、本研究開発で培った深層生成モデルに基づく細胞状態の解析技術をさらに発展させ、一細胞・空間トランスクリプトームの情報が解析技術を新たに開発した。具体的には、また、近年開発されている Subcellular 解像度のトランスクリプトームに対して、3次元の細胞内トランスクリプトームの推定を可能とする深層生成モデルの開発を行なった。最後に、これまでの細胞状態を中心とした情報解析ツールを発展させることで、細胞状態の分布によって規定される微小環境の潜在状態を解析可能な深層生成モデルを開発した。これらの研究開発により、加速フェーズでは、これまで開発を行ってきた細胞状態の解析基盤をマイクロ、マクロな方向へと発展させることで、微小環境から細胞内に至るまで様々な現象のオミクス解析を行うための方法を構築した。

(2) 詳細

1. VAE に基づくゆらぎを伴う細胞状態ダイナミクスの推定

本研究においては細胞状態のダイナミクスを移流拡散過程としてモデル化をし、それを変分推定の枠組みにより、spliced と unspliced のトランスクリプトーム情報に基づき推定する。具体的には、spliced のトランスクリプトーム観測に対し VAE を適用することにより、spliced の transcriptome 情報に基づく細胞状態 z への Encoder と Decoder を学習する。ここで、 z における移流と拡散を二つの Neural network によりモデル化することで、微小時間後の細胞状態の変化 d をガウス分布に従う確率変数と考えることができる。この微小

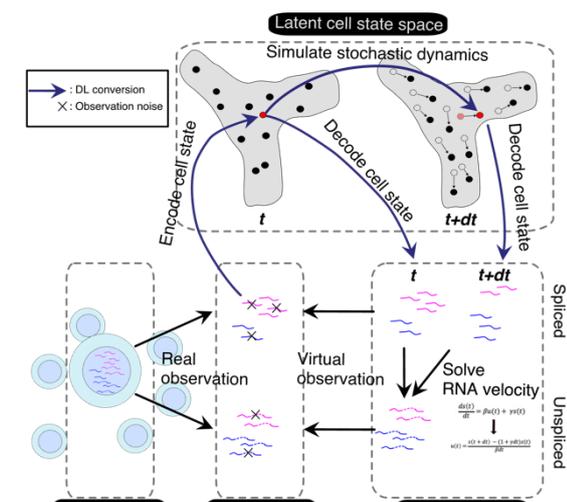


図 1 ゆらぎを伴うダイナミクスの推定手法

時間後の細胞状態の変化と実際の発現変化への Decoder を用いることにより、実際の遺伝子発現空間における時間変化を Decode することが可能となる。ここで、unspliced と spliced、spliced の時間変化を結びつける数理モデルを用いることにより、spliced の観測のみならず、unspliced の発現プロファイルを再構築することが可能となる(図1)。これにより、End to End で spliced の transcriptome の VAE と細胞状態変化の移流項と拡散項の同時最適化が可能となる手法を構築した。シミュレーションデータに対する適用では、シミュレーション時に仮定した分化方向を再構築することに成功した。また、この手法を公開されている Pancreas endocrine development のデータに対して適用を行なったところ、期待される Progenitor 集団から Alpha, Beta, Delta, Epsilon 細胞への分化ダイナミクスが推定されることを確認した。

2. 細胞運命分岐において増大する細胞状態ダイナミクスのゆらぎ

本手法では、細胞状態のダイナミクスを移流拡散過程に基づきモデル化し推定を行なっている、そのため微小時間後の変化 d について確率的に繰り返しサンプルすることが可能となる。繰り返しサンプルをした発現変動を Decoder により実際の発現変動に変換することにより各遺伝子の発現のばらつきを評価することが可能となる。これに基づき Pancreas endocrine development のデータについて各細胞におけるゆらぎの高さを評価した。その結果、既存の研究から様々な細胞種への分化が期待されている Ngn3 陽性の Endocrine progenitor の分画においてゆらぎが高まっていることが確認された(図 2)。このことは、細胞の運命が分岐する際に、遺伝子発現変動のダイナミクスの不確実性が増大することを示唆する結果となる。

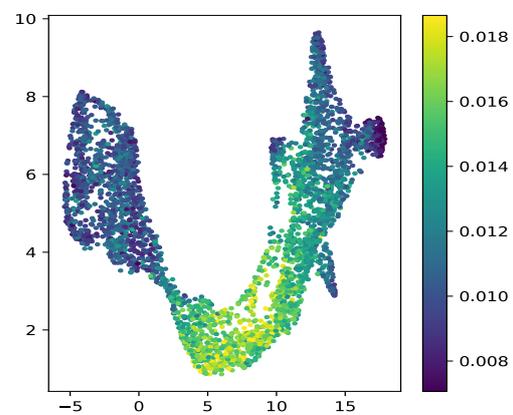


図 2 発現変動ゆらぎの大きさ

3. ゆらぎを実現する遺伝子の抽出

上述の方法により推定した細胞運命の分岐において高まるゆらぎに貢献している遺伝子の抽出を試みた。上位 20%となる細胞について、そのゆらぎのプロファイルに基づきそれらの細胞のクラスタ化を行ない、そのうちの Alpha、Beta 細胞と類似するクラスタについて、各遺伝子が Alpha 細胞、Beta 細胞のどちらがわのゆらぎに貢献しているかどうかを評価した。その結果、Arx 遺伝子が Alpha 細胞側のゆらぎの上位 8 位に、Pdx1 が Beta 細胞側の上位 6 位に位置していることが確認された。Arx と Pdx1 は、Endocrine progenitor において強制発現をさせるとそれぞれ Alpha 細胞、Beta 細胞に分化が偏ることが既存の研究により報告されており、この結果は、本研究が開発した手法が細胞の運命分岐を捉え、そこで重要となる因子の抽出に有効な手段となることを示している。

4. 空間・一細胞トランスクリプトームデータの統合による共局在解析

本研究では、まず一細胞トランスクリプトームデータに対して変文事後符号化器の適用により、遺伝子発現プロファイルから細胞状態への変換と細胞状態から遺伝子発現プロファイルの変換を学習する。それぞれの空間トランスクリプトームのスポットにどの細胞が入っているかを潜在変数依存的な関数として学習を行う。これにより、それぞれの細胞についてどのスポットに存在するかという空間的な情報が復元される。本手法では、この空間分布の重なりを共局在性の強さのスコアとして用いることにより、共局在する細胞集団の抽出を行う。これにより、共局在をする細胞集団の特徴づけや、細胞間コミュニケーションの推定を行うことが可能となる。実際に、扁平上皮癌のデータセットへの適用においては、Fibroblast と浸潤傾向のあるがん細胞の間に共局在により規定される集団を同定することができた(図 3)。これらの成果は、Preprint を公開し、国際学術誌 Cell Systems に現在投稿中である。

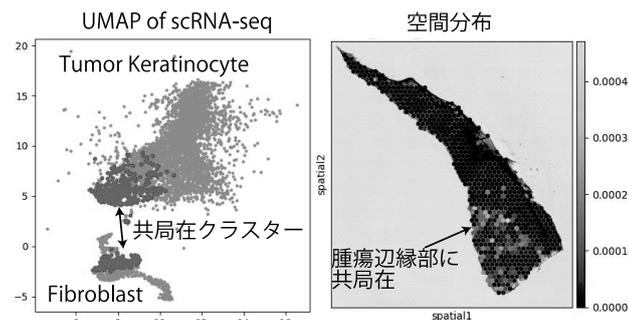


図 3. 潜在状態と空間における共局在集団の分布

5. 条件依存的なダイナミクスの推定

前年度までに開発をおこなった深層生成モデルに基づく細胞状態ダイナミクス推定の手法を、共変量に依存可能な形に実装を変更した。これにより、同じ細胞であっても、異なる共変量を持つ場合に、ダイナミクスがどのように変化をするか解析が可能となっている。

6. 共局在依存的細胞状態ダイナミクスの推定

上述の条件依存的な細胞状態ダイナミクスの推定手法、並びに一細胞共局在解析手法を組み合わせることにより、周辺に存在する細胞の発現プロファイルに依存する形で細胞状態のダイナミクスを推定する方法論を構築した。これにより、特定の細胞種と共局在すると特定の細胞状態への遷移が促進されるというような環境的な要因による遺伝子発現変動の変化を網羅的に捉えることを可能にした。

7. クロマチン状態依存的細胞状態ダイナミクスの推定

近年、トランスクリプトームとクロマチンアクセシビリティの同時プロファイリングが、様々な生物系に広く応用されている。また、クロマチンアクセシビリティは遺伝子発現の変化に先行して変化する傾向があることが、いくつかの研究で示されている。本研究では、条件付きダイナミクス推定の枠組みを拡張することにより、造血幹細胞分化の系において、クロマチンアクセシビリティから算出した転写因子(TF)活性により条件づけられたダイナミクスの推定を行った。この際には、TF activity への摂動に対するダイナミクスの応答を Jacobian 行列として細胞ごとに計算した上で、それらを連結したものに対して、Singular value decomposition を行うことで、ダイナミクスの変化をもたらす TF 活性の主要なパターンの抽出を行った。その結果、T 細胞と B 細胞への分化を促進する 2 パターンの TF activity の変動では、GATA3 と

SPIB というそれぞれの細胞系譜の発達に重要な因子が上位に登場することが確認された。これらの結果は、本方法論の応用範囲の広さと細胞状態の多様化プロセスを解明する上での有用性を示す結果となっている。

以降加速フェーズ

8. 深層生成モデルによる一細胞レベルのスプライシング、分解の推定

細胞の遺伝子発現制御、特に転写産物の発現レベルの制御には、転写自体の活性の他にも、転写産物のスプライシング、分解の速度が深く関わることが知られている。それゆえ、これらの転写産物の制御パラメータを細胞ごとに推定することができれば、細胞状態ごとの遺伝子発現の多様性を生み出す制御機構を抽出可能なことが期待される。その一方で、本研究で提案した枠組みを含めて、既存の RNA velocity の推定手法においては、スプライシング、分解の速度は一定と仮定されていた。本研究では、これまでに開発していた細胞状態のダイナミクスをスプライシング数理モデルにより推定するアルゴリズムを拡張し、細胞状態ごとに異なるスプライシング、分解速度を取れるようにモデルの拡張を行なった。

9. 細胞内トランスクリプトームの立体再構築

空間トランスクリプトームの発展により、その空間解像度は今や細胞内のオルガネラ等、微細な構造を描出可能なまでに高精細化が進んでいる。その一方で、観測されるプロファイルは非常にスパースであり、空間パターンを細胞間に対応づけて解析することも困難であった。

本研究においては、これまでの研究で用いてきた一細胞レベルの発現プロファイルから細胞状態を抽出し様々な生命現象と紐付ける枠組みを発展させることで、オミクス観測へと拡張することにより、細胞状態の推定とそれに紐づく立体的なトランスクリプトームを End-to-End で学習可能な方法論を構築した。

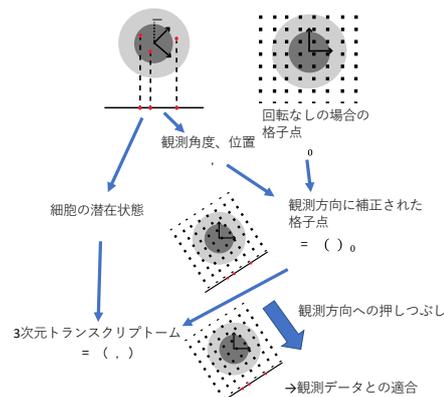
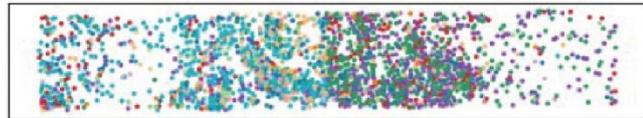


図 4. 立体再構築アルゴリズム

10. Transformer による微小環境状態の推定

これまで、本研究を含めオミクス解析の発展は、細胞を中心とした生命現象の理解を大きく進めてきた。その上で、本研究で活用してきた深層生成モデルに基づく、細胞状態の概念は大きな役割を果たしてきた。その一方で、実際の組織内では、これらの多様な細胞は、微小環境において相互作用をしながら組織全体の運命を大きく変えていく。そのため、このような微小環境の多様性をその内部に存在する細胞状態の多様性に基づいて解析するための、新たな情報解析基盤の創出が求められていた。そこで、本研究では、Transformer を用いることにより、周囲の細胞状態から Niche の潜在状態を encode する Nicheformer の開発を行った。Nicheformer では微小環境状態は、細胞状態の分布として decode されるため、その微小環境内にどのような細胞状態が集積するかに基づいて解析を行うことが可能となる。これにより、微小環境の変遷による細胞状態の変遷や、増減を体系的に扱うことを可能とする。

細胞状態のクラスタリング結果



微小環境状態のクラスタリング結果

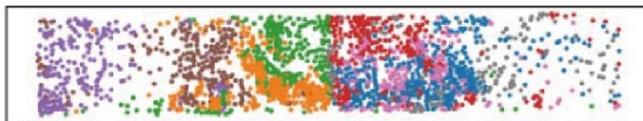


図 5. 細胞と微小環境状態の空間分布

3. 今後の展開

本研究においては、環境依存的な細胞状態のダイナミクスを内的なゆらぎを包含する形で推定するフレームワークの確立に成功し、実際に細胞状態の多様化において状態変化のゆらぎが重要な役割を示すことや、特定の細胞状態への遷移に重要となる細胞間の相互作用を同定することができた。今後の方向性としては、主に二つの方向性で研究を発展させていくことを考えている。一つは、高精細化が大きく進み Subcellular 解像度にも達しつつある空間トランスクリプトーム観測のデータに対する新規手法を開発することにより、これらのゆらぎを伴う発現変動や、細胞間相互作用による発現制御により細胞内のトランスクリプトーム空間分布がどのように変化していくか、そのホットスポットを細胞内の空間構造として探求することにある。これにより、ホットスポットを共有する転写産物からの配列的特徴の抽出等を通して、細胞内の各遺伝子産物の空間分布を調整する分子機構の解明につながることを期待される。

また、もう一つの方向性として考えられるのが、今回の研究により推定される遺伝子発現変動のゆらぎや環境依存的な発現変動が、ゲノム上のどのような制御配列により規定されるか、を体系的に究明することにある。具体的には、確率的な不確実性を記述可能な数理モデルのパラメータを、細胞状態、近傍細胞の状態、遺伝子周辺の制御配列により推定する枠組みへと、本研究で開発した手法を拡張することを考えている。これにより、Explainable AI の技術を応用することで、環境依存性やゆらぎに関わる配列的特徴の解明を行う(2年後)。これにより、制御配列領域における疾患関連遺伝子変異の細胞間相互作用等への影響を推定し、新たな治療標的の探索につながることを期待される。

4. 自己評価

2022 年度における研究目的の達成状況としては、①「環境依存的細胞状態ダイナミクスの推定手法の確立」、②「細胞ダイナミクスの変化を実現する分子機構の探索」については、すでに研究開発を完了しており想定以上のペースで研究が進行しており、関連する二つの論文が公開されており(成果1、2)、もう一報の論文も現在準備中となっている。③「細胞状態ダイナミクスにおける環境要因とゆらぎの生体組織全体への影響探索」については、すでに基礎的な技術の開発を終えており、実データでの評価を開始しており想定通りの進捗状況と言える。④「抗腫瘍免疫応答における細胞の環境依存的ダイナミクスの解明」については、研究環境の変化により当初予定していた前向き実験での解析が困難となったが、公開データを元にした解析で同様の結果を得ることに成功している。

研究の進め方については、研究者本人が主体となって研究を進めている。本研究において開発を行った手法により、7件もの共同研究が進行しており、分野の進展を大きく助ける成果となっていると言える。研究費の執行状況としては、開発用 PC、ワークステーションの導入により大きく研究を加速させることが可能となった。

本研究で開発を行った新技術の波及効果としては、特に組織内の細胞間相互作用に基づく疾患メカニズムを解明することにより、新たな治療標的薬の発見につながることを期待される。本研究では、これまでの疾患の結果となる細胞集団に焦点を当てた解析から、その集団に至るまでのメカニズムを細胞間の相互作用による発現変動や細胞の内因的なゆらぎに求めるアプローチに転換することにより、疾患のアキレス腱を効率的に探索することを可能にすることが期待される。

5. 主な研究成果リスト

(1) 代表的な論文(原著論文)発表

研究期間累積件数: 4件

1. Kojima Y. (co-corresponding) et al. Single-cell colocalization analysis using a deep generative model. Cell Syst. 2024 Feb 21;15(2):180-192.e7. (査読あり).

本研究においては、深層生成モデルに基づいて、空間トランスクリプトームと一細胞トランスクリプトームの統合により、細胞間の共局在関係を推定する手法 DeepCOLOR の開発を行った。DeepCOLOR では、一細胞トランスクリプトームの空間情報を復元することで、共局在関係を推定し、共局在により細胞間の相互作用により影響を受けた集団を効率的に同定、その分子的な特徴づけを行うことを可能にした。

2. Nagaharu K., Kojima Y. (equally co-first) et al. A bifurcation concept for B-lymphoid/plasmacytoid dendritic cells with largely fluctuating transcriptome dynamics. Cell reports. 2022 Aug 30; 40.9 (2022): 111260. (査読あり)

本研究においては、近年物理学での研究が盛んに行われている Physics Informed Neural Network の考え方を数理生物学と深層学習の融合へ応用し、潜在的な細胞状態の確率的な変化をスプライシング現象が記述するダイナミクスと対応をさせることで、データ駆動的に不確実性を伴う遺伝子発現変動を推定可能な手法 VICDYF を開発した。本研究では、細胞状態ダイナミクスの不確実性が細胞状態の多様化と関連づけられることを示した。



3. Ozato Y. Kojima Y. (equally co-first) et al. Spatial and single-cell transcriptomics decipher the cellular environment containing HLA-G+ cancer cells and SPP1+ macrophages in colorectal cancer, Cell Reports, Vol. 42, Issue 1, 111929, 2023.

本研究では、ベイズモデルに基づく細胞種存在量の推定を行う解析手法を用いて、細胞種間の共局在関係の解析を行なった。これにより、Wet 研究者による検証も併せて、がん細胞の発現する HLA-G により免疫抑制的な SPP1+Macrophage を誘導することを明らかにした。

(2)特許出願

該当なし

(3)その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

(招待講演)

1. 小嶋 泰弘 “深層生成モデルに基づく一細胞・空間トランスクリプトームの解析技術” NGS Expo 2023, Japan, Nov 2023
2. Yasuhiro Kojima , “Inferring cell state dynamics dependent on tumor microenvironments”, 2nd JCA-AACR Precision Cancer Medicine International Conference, Japan, Jun 2023
3. 小嶋泰弘 “変分オートエンコーダを用いた細胞状態ダイナミクスの推定” 2022 年度統計関連学会連合大会 Japan, Sep 2022
4. 小嶋泰弘 “深層生成モデルによる一細胞・空間トランスクリプトームの解析” 第二回分子病理学教室セミナー Japan, Aug 2022
5. Yasuhiro Kojima, Teppei Shimamura, “Dynamics and colocalization of deep learning-based cell states behind single cell and spatial transcriptome observation”, The 40th Sapporo International Cancer Symposium, Japan, Jul 2022
6. 小嶋泰弘 “空間トランスクリプトームの一細胞分解に基づく細胞間共局在関係の推定” 2021 年度日本数理生物学会年会 Japan, Sep 2021