

令和 4 年 6 月 1 日現在

機関番号：14301

研究種目：若手研究

研究期間：2019～2021

課題番号：19K20399

研究課題名（和文）単一細胞オミックスデータに基づく細胞系譜推定および比較アルゴリズムの開発

研究課題名（英文）Development of cell trajectory inference and comparison algorithm based on single-cell omics data

研究代表者

森 智弥（Mori, Tomoya）

京都大学・化学研究所・助教

研究者番号：50795333

交付決定額（研究期間全体）：（直接経費） 2,000,000円

研究成果の概要（和文）：本研究では、単一細胞遺伝子発現データに基づいて細胞系譜推定から比較までを行う統合解析アルゴリズムを開発した。本手法をヒト繊維芽細胞およびヒト骨格筋芽細胞のデータに適用した計算機実験では、2つのデータセット間の細胞群の対応関係を求め、木のアラインメント上のパスに対して動的時間伸縮法を用いて擬似時間の比較を行ったところ、既存の報告と同様の結果が得られることを確認した。また、人工データやヒト骨髄の細胞データを用いた計算機実験でも本手法の有効性を確認した。さらに、他手法との比較および機能遺伝子群解析に基づいた本手法の有用性に対する考察と得られた結果の生物学的解釈も合わせて論文投稿を行った。

研究成果の学術的意義や社会的意義

本研究課題の目的は、単一細胞遺伝子発現データに基づいて細胞系譜を推定し、それらの比較を通じて生物学的および医学的知見を獲得するための高速かつ高精度な統合解析アルゴリズムを開発することである。本研究課題で開発した手法をさらに発展させ、細胞間の系統関係をより詳細なレベルで明らかにすることができれば、正常組織の発達、恒常性の維持機能、発達障害、そして癌などの病理について重要な情報を提供することができるだけでなく、分化性能の良いiPS細胞を選別することや細胞が正常に分化しない場合の原因究明などに役立つと考えられるため、再生医療研究の一助となることが期待される。

研究成果の概要（英文）：In this study, we have developed an algorithm for cell trajectory inference and comparison with high accuracy based on single-cell gene expression data. We applied this algorithm to single-cell human fibroblast and human skeletal muscle myoblast data, and after finding the correspondence of cell clusters between the two datasets using tree alignment, a pseudotime comparison was performed based on the dynamic time warping for the paths on the alignment. As a result, it was confirmed that the analysis results were consistent with existing reports. Its effectiveness was also confirmed in the experiments using synthetic data and real data of human bone marrow. In addition, we compared it with other methods and performed functional gene set analysis, and then submitted a paper together with consideration of the usefulness of this method and biological interpretation of the results.

研究分野：バイオインフォマティクス

キーワード：細胞系譜比較 細胞系譜推定 シングルセル解析 木のアラインメント 木の編集距離 機能遺伝子群解析

様式 C-19、F-19-1、Z-19 (共通)

1. 研究開始当初の背景

生体を構成する細胞間の系譜を再構築することは生物学における長年にわたる最重要目的の一つである。これらの細胞間の系統関係を明らかにすることができれば、正常組織の発達および恒常性の維持機能だけでなく、発達障害や癌などの病理についても重要な情報を提供することができる。伝統的に系譜推定は細胞を遺伝的に標識化し、その子孫細胞を追跡していくことで達成されてきた。子孫細胞に見出される細胞種の多様性が創始細胞の分化可能性を示しているため、系譜追跡には正確な細胞種同定が必要となる。理想的にはできるだけ多くの標識を用いる必要があるが、一般的に細胞種の同定は限られた数の標識に基づいて行われるため、細胞亜集団の多様性を明らかにすることが難しく、複雑な組織の場合にはバイアスのかかった結果が得られてしまう可能性がある。一方で、近年では単一細胞トランスクリプトーム（細胞中に存在する mRNA の総体）解析の技術進歩により、数千もの単一細胞トランスクリプトームのプロファイリングが可能になっているため、細胞種同定はこれまでにない発展を遂げている。このことから組織や臓器を構成する個々の細胞の不均一性に対する理解も深まっており、単一細胞トランスクリプトームに基づいた系譜推定法がいくつか提案されてきた。しかしながら、再構築した系譜の明確な評価基準や推定された系譜の応用可能性についての議論が十分にされておらず、実際に医療応用を目指す際の大きな課題となっている。

2. 研究の目的

本研究の目的は、単一細胞データに基づいて細胞系譜推定から比較までを高精度に行う統合解析アルゴリズムを開発することである。これまで単一細胞トランスクリプトームデータに基づく系譜推定アルゴリズムはいくつか提案されているが、いずれの手法においても遺伝子発現プロファイルから細胞の分化進行度を示す擬似時間を推定する際に情報の縮退が見られる。また、得られた推定結果を定量的に判断する基準が存在しないため、どの手法が優位であるのかが明らかになっていない。本研究の目的が達成されることにより、高精度な系譜推定・比較が可能となれば、例えば正常細胞の細胞系譜と iPS 細胞由来の細胞系譜を比較することにより、分化性能の良い iPS 細胞を選別することや細胞が正常に分化しない場合の原因究明などに役立つことが期待される。

3. 研究の方法

本研究では目的達成のために下記2つの研究項目を実施した。

- (1) 情報の損失をできるだけ抑えた細胞系譜推定手法の開発
- (2) 木構造データ比較に基づく新たな細胞系譜比較手法の開発

(1) 情報の損失をできるだけ抑えた細胞系譜推定手法の開発

単一細胞解析においてしばしば問題となる個々の細胞間の実験バイアスを取り除くために細胞のクラスタリングを行った。細胞群を複数のグループに分類し、そしてクラスターの重心を頂点とする最小全域木を計算することで木構造を構築し、細胞系譜を推定する。細胞系譜の擬似時間を計算する既存手法は情報を一次元に圧縮して行うことが多いが、(2)で説明する木のアラインメントを用いることで、木構造情報を保持したまま擬似時間を計算する手法を提案した。

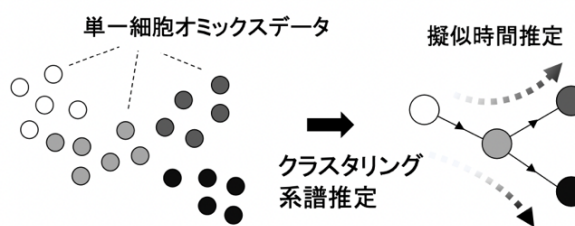


図1: 系譜推定手法の概念図

(2) 木構造データ比較に基づく新たな細胞系譜比較手法の開発

細胞系譜の比較はこれまで細胞を一次元に圧縮して動的時間伸縮法を適用する手法が主流であったが、圧縮させることなく系譜比較を行う手法を新たに開発した。具体的には、上記(1)の手順で生成した木構造データを、木のアラインメントと呼ばれる計量を用いて比較した。ここで、一般の場合の木のアラインメントを計算する問題は NP 困難と呼ばれ、多項式時間で解くことができる効率的なアルゴリズムが存在しない。しかしながら、木の最大次数が定数で抑えられている場合、多項式時間で計算できることが知られている。これを適用することで細胞系譜を一次元に圧縮することなく直接比較することができる。細胞系譜の比較において、このような手法は未だ提案された事がなく、全く新しいアプローチであると言える。

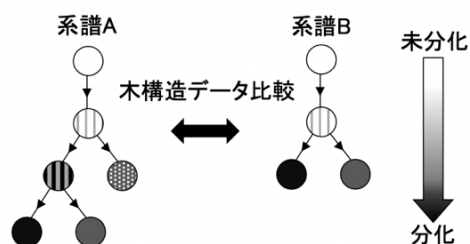


図2: 系譜比較手法の概念図

4. 研究成果

上記方法に基づいて、大阪大学大学院医学系研究科の加藤有己准教授らと単一細胞データを用いた細胞分類・細胞状態遷移経路推定アルゴリズム CAPITAL (comparative analysis of pseudotime trajectory inference with tree alignment) を開発し、プレプリントリポジトリ「bioRxiv」にて公開した [1]. Gene Expression Omnibus (GEO) データベースからダウンロードした MyoD を介したヒトの線維芽細胞 (hFib-MyoD) のリプログラミングのデータ (GSE105211) とヒト骨格筋芽細胞 (HSMM) の分化データ (GSE52529) を木のアラインメントを用いて比較した計算機実験では CAPITAL が実験条件の異なる 2 つのデータセットの擬似時間構造を保ちながらできるだけ多くのクラスターをマッチングさせることが示された (図 3).

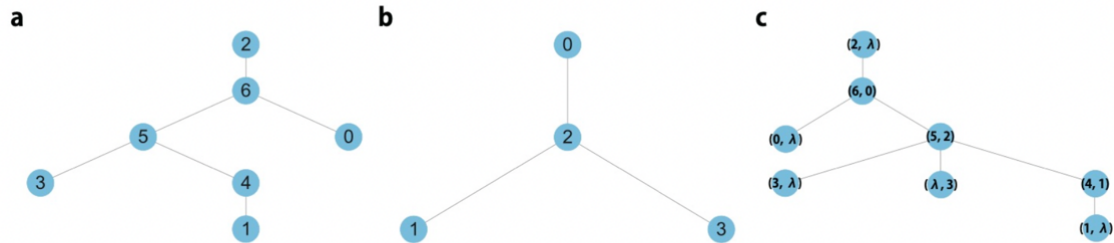


図3: 木のアラインメントによるヒト線維芽細胞データとヒト骨格筋芽細胞の細胞系譜比較. (a)ヒト線維芽細胞データ, (b)ヒト骨格筋芽細胞データ, (c)木のアラインメント. 頂点に付随する番号は入力データのクラスター番号を示す. 文献[1]より引用.

また、木のアラインメントから得られた中心パス (頂点 (6, 0) から頂点 (4, 1) に至るパス) について、擬似時間に沿って並べた線維芽細胞と骨格筋芽細胞のデータに動的時間伸縮法を適用したところ、細胞増殖に関連するマーカー遺伝子 CDK1 と骨格筋の転写に関連するマーカー遺伝子 MEF2C の発現レベルに対する 2 つのデータセット間の擬似時間のずれと、同じく骨格筋の転写に関連するマーカー遺伝子 MYOG の発現レベルに対する擬似時間のずれには異なる傾向が見られた (図 4).

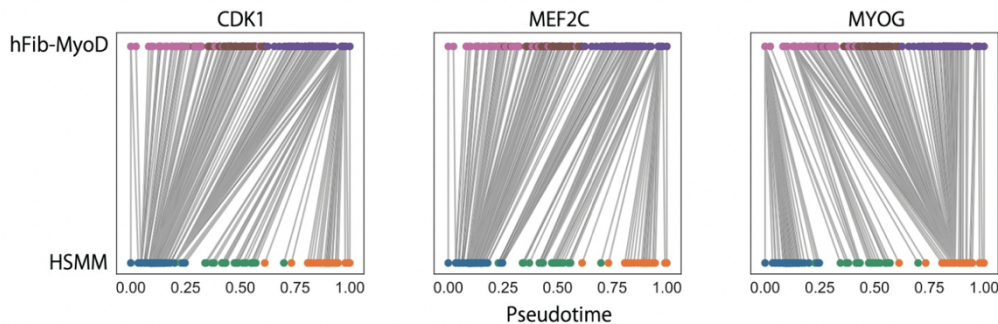


図4: 動的時間伸縮法による細胞データのアラインメント(文献[1]より引用).

これらのマーカー遺伝子について擬似時間に対する発現レベルのダイナミクスを調べたところ、CDK1 と MEF2C の発現レベルは擬似時間経過と共に同様の振る舞いを示した. 一方で、MYOG は hFib-MyoD データでは擬似時間経過に沿った発現レベルの大きな変化はなかったものの HSMM データでは後半に大きな発現レベルを示していた. 同様の結果は文献[2]でも示されており、これらの結果から CAPITAL の有効性が確認された.

さらに、ヒト線維芽細胞および骨格筋芽細胞以外にも人工データやヒト骨髄の細胞データへと提案手法の適用範囲を拡大し、本手法の有効性を確認する実験を追加で行った. 他手法との比較および機能遺伝子群解析なども行い、手法の有用性に対する考察と本手法によって得られた結果の生物学的解釈をさらに充実させたうえで国際学術論文誌に投稿中である.

<参考文献>

- [1] R Sugihara et al. Alignment of time-course single-cell RNA-seq data with CAPITAL. *bioRxiv*. 2019. doi: <https://doi.org/10.1101/859751>.
- [2] D Cacchiarelli et al. Aligning single-cell developmental and reprogramming trajectories identifies molecular determinants of myogenic reprogramming outcome. *Cell Syst.*, 7:258-268, 2018.

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 0件／うち国際共著 0件／うちオープンアクセス 1件）

1. 著者名 Reiichi Sugihara, Yuki Kato, Tomoya Mori, Yukio Kawahara	4. 巻 -
2. 論文標題 Alignment of time-course single-cell RNA-seq data with CAPITAL	5. 発行年 2019年
3. 雑誌名 bioRxiv	6. 最初と最後の頁 1-11
掲載論文のDOI（デジタルオブジェクト識別子） 10.1101/859751	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計4件（うち招待講演 0件／うち国際学会 1件）

1. 発表者名 Yuki Kato, Reiichi Sugihara, Tomoya Mori, Yukio Kawahara
2. 発表標題 Alignment of complex single-cell trajectories with CAPITAL
3. 学会等名 28th Conference on Intelligent Systems for Molecular Biology (ISMB2020) (国際学会)
4. 発表年 2020年

1. 発表者名 加藤 有己, 杉原 礼一, 森 智弥, 河原 行郎
2. 発表標題 CAPITALによるシングルセルデータの疑似時系列比較解析
3. 学会等名 第43回日本分子生物学会年会
4. 発表年 2020年

1. 発表者名 加藤 有己, 杉原 礼一, 森 智弥, 河原 行郎
2. 発表標題 複雑な細胞運命比較のための時系列 1 細胞RNA-seqデータのアラインメント
3. 学会等名 第42回日本分子生物学会年会
4. 発表年 2019年

1. 発表者名 杉原 礼一, 加藤 有己, 森 智弥, 河原 行郎
2. 発表標題 分岐を考慮した時系列シングルセルデータのアラインメント
3. 学会等名 情報処理学会 第61回BIO研究発表会
4. 発表年 2020年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関