

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

杉原 礼一¹ 加藤 有己^{1,2,a)} 森 智弥³ 河原 行郎^{1,2}

概要: シングルセル遺伝子発現解析が盛んに行われている。ある組織の時系列シングルセルデータの活用例として、細胞状態の遷移過程を捉えることを目指す疑似時間解析が挙げられる。特に、実験条件の異なるデータから得られる、細胞集団の疑似時間経路同士を比較することにより、分化などの動的な過程を制御する遺伝子の同定につながる事が期待される。しかしながら、既存手法では線形構造の比較しか行えないため、経路に分岐が含まれる場合、どの経路同士を比較すべきかは解析者の判断に委ねられている。本稿では、経路に分岐が含まれるような複雑な場合にも、客観的に比較可能な手法 CAPITAL を開発したことを報告する。

1. はじめに

近年のシングルセル RNA-seq 技術により、細胞集団の遺伝子発現を 1 細胞単位でトランスクリプトーム網羅的に解析することが可能になった [1]。これに伴い、細胞の分化などの時間経過により変化する生命現象を理解するため、経過時間がそれぞれ異なる細胞集団の遺伝子発現を 1 細胞レベルで調べ、それらの疑似的な時間経過変化における発現ダイナミクスを追うための数理モデルが盛んに開発されている。ここでは、各細胞を他の細胞との発現レベルの差で重み付けした辺で接続するグラフ構造を考え、発現変化が最小となるように細胞間を最短で結ぶ経路を疑似時間経路 (pseudotime trajectory) と呼んでいる。疾患モデルなどの刺激を加えた実験系と、同一細胞集団のコントロールを比較する際、数理モデルで予測した各実験系での疑似時間経路同士を比較することで、細胞の運命を決定する遺伝子を同定できることが期待されている。ただし、現在の経路比較手法では、線形状の経路の比較しか行えず、疑似時間経路に分岐が存在する場合、経路選択に関する客観性が担保できず、以降の下流解析の信頼性を揺るがしている。

本稿では、各実験系の分岐を含むような疑似時間経路をグラフ理論的手法で構築した後、両者のアラインメント計

算を実現した手法 CAPITAL [2] の開発を報告する。これにより、グローバルな木構造間の比較を実現し、解析者によるバイアスがかからない客観的な細胞集団の時系列比較が行えるようになった。互いに関連性があり、条件の異なる 2 つの実験系から得られたシングルセル RNA-seq データに対して提案手法を適用した結果、適切な細胞間アラインメントが得られ、かつ発現ダイナミクスの変化を捉えることに成功した。

2. 手法

2 つの実験系それぞれについて、シングルセル RNA-seq から得られた遺伝子発現プロファイル行列が与えられた時、CAPITAL では対応する疑似時間経路を計算し、それらの対応関係を求める。ここで、計算量を削減し、かつシングルセルデータに特有のノイズの影響を最小限に抑えるため、CAPITAL が扱う疑似時間経路グラフの頂点は、個々の細胞ではなく類似の細胞集団を表すクラスターとする。また、疑似時間経路に分岐がある場合は木構造とみなし、木構造同士のアラインメントを動的計画法により計算する [3]。アラインメント結果は木構造で表現されるが、CAPITAL ではアラインメントの開始点 (根) から全ての葉に至る経路について、対応する細胞集団を復元し、それらの疑似時間順序を求める [4]。さらに、特定の遺伝子に関する dynamic time warping [5] を計算することで、遺伝子発現のダイナミクスを追うことが可能である。なお、CAPITAL アルゴリズムの詳細については文献 [2] を参照されたい。

¹ 大阪大学 医学部
Faculty of Medicine, Osaka University, 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan

² 大阪大学 大学院医学系研究科
Graduate School of Medicine, Osaka University, 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan

³ 京都大学 化学研究所
Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan

a) ykato@rna.med.osaka-u.ac.jp

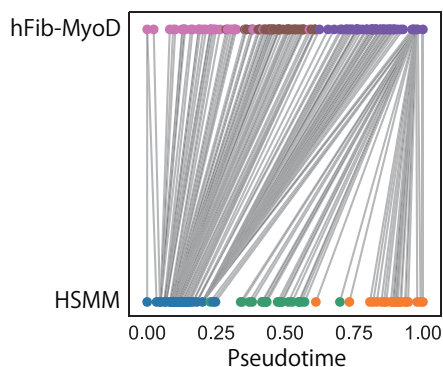


図 1 hFib-MyoD と HSMM のアラインメントから得られる経路に対し、遺伝子 CDK1 に関する 1 細胞間の dynamic time warping の結果. 各細胞の色は文献 [2] の Fig. 2 に対応する.

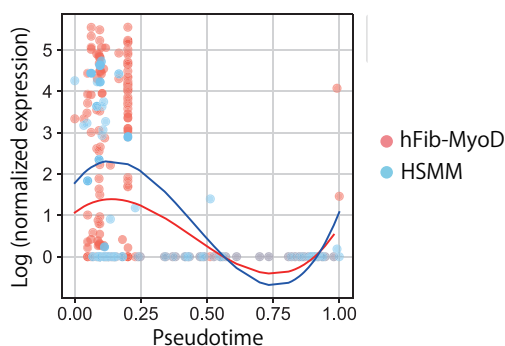


図 2 遺伝子 CDK1 に関する dynamic time warping から導出される発現ダイナミクス.

3. 結果

3.1 筋細胞データ

本研究では、ヒトの線維芽細胞の初期化データ (hFib-MyoD) (GSE105211)[6] およびヒトの筋芽細胞の分化データ (HSMM) (GSE52529)[7] を比較に用いた. なお、これらのデータは疑似時間経路に分岐を含むことが報告されている.

3.2 アラインメント

各データセットにおける CAPITAL を用いた細胞クラスタリング、および疑似時間経路推定とアラインメント結果はそれぞれ文献 [2] の Fig. 2 および Fig. 3 に示す通りである. これは、CAPITAL が疑似時間構造を保存したまま、細胞クラスター間のマッチングを客観的に計算可能であることを示す結果である.

次に、マッチングクラスター数が最大であるアラインメント経路ペアに対して、遺伝子 CDK1 に関する dynamic time warping を計算した結果を図 1 に、対応する発現ダイナミクスを図 2 に示す. ここで、CDK1 が疑似時間初期に両実験系で高発現であることが観察できるが、当該遺伝子

が活発な増殖に関係するという既知の結果に合致する. また、本研究の結果 [2] は文献 [6] で報告されている結果と類似しており、提案手法の妥当性がある程度示されていると考える.

4. おわりに

CAPITAL では、比較すべき時系列細胞集団の疑似時間経路に分岐が含まれる場合でも、可能なアラインメントを全て一括で計算することが可能となった. これにより、例えば分化が成功した細胞集団の主要な疑似時間経路の比較のみならず、分化が失敗した経路についても同時に比較することができるため、細胞状態の遷移過程に関する分子レベルでの網羅的な観察が可能である. 今後シングルセルデータ解析がますます普及していくことが予想され、CAPITAL による細胞集団の時系列解析への貢献が期待される.

謝辞 本研究成果の一部は、大阪大学サイバーメディアセンターのスーパーコンピュータシステム、および情報・システム研究機構国立遺伝学研究所が有する遺伝研スーパーコンピュータシステムを利用して得られたものである.

参考文献

- [1] Luecken, M. D. and Theis, F. J.: Current best practices in single-cell RNA-seq analysis: a tutorial, *Mol. Syst. Biol.*, Vol. 15, p. e8746 (2019).
- [2] Sugihara, R., Kato, Y., Mori, T. and Kawahara, Y.: Alignment of time-course single-cell RNA-seq data with CAPITAL, *Preprint bioRxiv*, (online), available from (<https://doi.org/10.1101/859751>) (2019).
- [3] Jiang, T., Wang, L. and Zhang, K.: Alignment of trees – an alternative to tree edit, *Theor. Comput. Sci.*, Vol. 143, pp. 137–148 (1995).
- [4] Haghverdi, L., Büttner, M., Wolf, F. A., Büttner, F. and Theis, F. J.: Diffusion pseudotime robustly reconstructs lineage branching, *Nat. Methods*, Vol. 13, pp. 845–848 (2016).
- [5] Giorgino, T.: Computing and visualizing dynamic time warping alignments in R: the dtw package, *J. Stat. Softw.*, Vol. 31, pp. 1–24 (2009).
- [6] Cacchiarelli, D., Qiu, X., Srivatsan, S., Manfredi, A., Ziller, M., Overbey, E., Grimaldi, A., Grimsby, J., Pokharel, P., Livak, K. J. et al.: Aligning single-cell developmental and reprogramming trajectories identifies molecular determinants of myogenic reprogramming outcome, *Cell Syst.*, Vol. 7, pp. 258–268 (2018).
- [7] Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N. J., Livak, K., Mikkelsen, T. S. and Rinn, J. L.: The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells, *Nat. Biotechnol.*, Vol. 32, pp. 381–386 (2014).