

令和 5 年 5 月 19 日現在

機関番号：12601

研究種目：若手研究

研究期間：2020～2022

課題番号：20K19916

研究課題名（和文）細胞分化における時変遺伝子ネットワーク推定手法の開発

研究課題名（英文）Inference of time-varying genetic networks for cell differentiation

研究代表者

仲嶋 なつ（Nakajima, Natsu）

東京大学・医科学研究所・特任研究員

研究者番号：60848373

交付決定額（研究期間全体）：（直接経費） 3,200,000円

研究成果の概要（和文）：細胞分化や疾患発症における各細胞の状態は、多くの遺伝子を含む相互作用により制御されて異なっているため、各細胞種での遺伝子間相互作用を推定することは重要となる。本研究では、1細胞の遺伝子発現データから、細胞の擬似時系列を推定し、各細胞種での遺伝子ネットワークを推定する手法を開発した。造血幹細胞分化の発現量データに適用し計算機実験を行なったところ、数種の細胞種については分化点を推定することが可能であり、また細胞数が多いデータに対しても、全計算過程の並列化を行うことによって、計算速度が大幅に向上することを確認した。

研究成果の学術的意義や社会的意義

本研究では、細胞状態が動的に変化する仕組みに着目し、複数の細胞種への分化に対応した時変遺伝子ネットワーク推定手法を開発する。1細胞の遺伝子発現データを適用することで、細胞分化や疾患発症において、細胞状態の遷移を引き起こす遺伝子間相互作用を解明することが期待される。

研究成果の概要（英文）：In the cell differentiation or disease develop, since each of the cell state is different due to the interaction involving many genes, it is crucial to infer the genetic interaction for each cell type. In this study, we develop a method to infer the time-varying genetic networks for cell differentiation with pseudotime analysis from single-cell RNA-seq data. We applied this method to the Hematopoietic stem cell scRNA-seq data and performed the computational experiments. The results indicate that this method can infer differentiation time points of several cell types and as for the datasets which consist of a large number of cells, the run time can be improved through parallel computing for all computation.

研究分野：バイオインフォマティクス

キーワード：時変遺伝子ネットワーク 1細胞解析 細胞分化 LASSO回帰 動的計画法 並列化 マルチプロセス

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

### 1. 研究開始当初の背景

造血系や心血管系の発生分化や個体発生、疾患発症の機序を制御している遺伝子発現は、多様な機能を持つ細胞種から組織を構成し、多数の遺伝子を含むそれぞれに異なる相互作用により特徴づけられる。ゆえに、異なる細胞状態での遺伝子間相互作用を明らかにすることは、発生や疾患発症機序を理解する上でも重要である。

バイオインフォマティクス分野においては、このような遺伝子間相互作用を、遺伝子のネットワーク (図 1) として捉え、次世代シーケンシングやマイクロアレイなどにより得られた遺伝子発現データをもとに、数的手法を用いてモデル化する研究が数多く行われている。このような推定は、ある時点での静的な遺伝子発現を再現しているが、実際の遺伝子発現では、細胞の状態遷移により動的に変化する仕組みを持つ。しかしながら、それらのネットワーク推定では、モデル化が複雑であることや十分なサンプル数のデータが必要となるため、あまり研究がなされていないのが現状である。

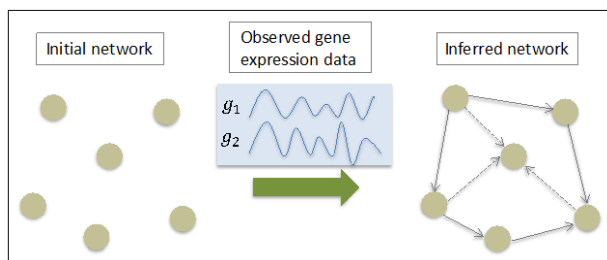


図 1 遺伝子ネットワーク推定の概要。頂点が遺伝子、辺が遺伝子間の相互作用を表す。

### 2. 研究の目的

本研究では、これらの課題をふまえ、細胞分化などの細胞状態が遷移する系において、複数の細胞種への分化に対応した時変遺伝子ネットワーク推定手法 (図 2) を開発する。1 細胞での mRNA 発現量 (scRNA-seq) データに適用することで、どの時点で細胞状態が変化し、時間経過に伴って遺伝子間相互作用がどのように変化するかを推定する。擬似時系列推定による二次元的な細胞種分類をもとに、各細胞種での遺伝子間相互作用を推定することで、発生や分化、疾患の機序への理解を深めることが期待される。

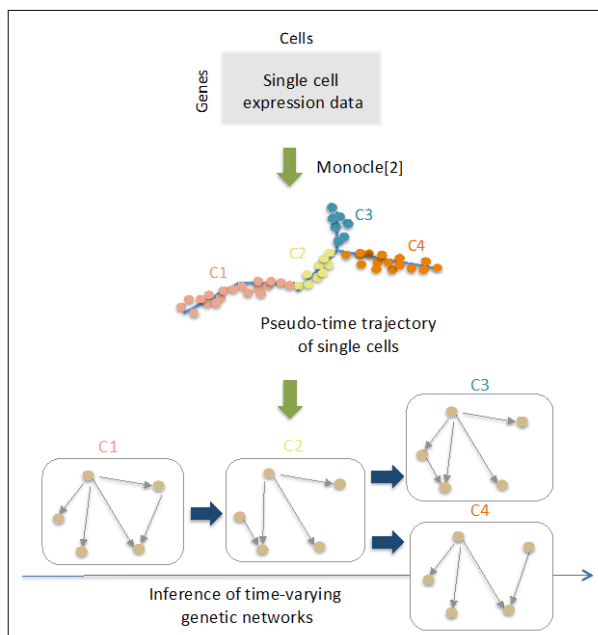


図 2 細胞分化における時変遺伝子ネットワーク推定の概要。シングルセル遺伝子発現量データから、分化軌道を推定し、時間軸に沿ったデータを作成して、それをもとに分化点と各細胞種での遺伝子ネットワークを推定する。

### 3. 研究の方法

本研究では、scRNA-seq から得られる分化の擬似時系列をもとに、以下の推定を行う手法を開発した。

#### (1) scRNA-seq データからの擬似時系列データの作成

細胞周期や細胞分化、細胞活性などの細胞状態の遷移を捉えるために、scRNA-seq データを用いた擬似時系列解析により、分化軌道を推定する。軌道推定とは、遺伝子発現パターンによ

る細胞間類似度をもとに、擬似時間軸に沿って細胞を並び替えることで、数多くの推定手法が開発されている [1]。細胞分化においては、細胞の状態が直線状や環状、木構造状などの様々な形状の分化軌道となる。また、scRNA-seqデータは非時系列であるため、分化軌道については、細胞を擬似時系列に並び替える手法 [2, 3]に適用することで、細胞状態の遷移を推定することができる。そこで本研究では、推定された擬似時系列から計算される各細胞の二次元座標をもとに、十分なサンプル数の発現量データを作成する手法を開発した。これらのデータは、数百から数千細胞から構成されるため、複数の細胞種へ分化する系や遺伝子数が多いネットワークの推定には有用である。

#### (2) LASSO正則化と二重動的計画法を用いた時変遺伝子ネットワークの推定

これまでに、バルクの時系列データから時変遺伝子ネットワークを推定する手法 [4, 5]やシングルセルデータから推定する手法 [6, 7]が開発されている。バルクデータからの推定については、観測値とモデル化された予測値の二乗誤差が最小となるような分化の変化点と、各細胞種におけるネットワーク構造を推定する最小化問題として定式化を行い、この問題に対して、最小二乗法と動的計画法を組み合わせた計算手法 [5]が開発されている。直線状に遷移する系や小規模ネットワークに対しては適用が可能である。

そこで提案手法では、一細胞解析におけるネットワーク推定問題として、大規模なネットワークや、一遺伝子あたりの検出感度が低いscRNA-seqデータに対しても適用可能となるように、LASSO正則化によるスパースモデリングを用いて、有効な変数選択を行う手法へと拡張した。各遺伝子ごとにLASSO回帰により二乗誤差を計算し、回帰係数の大きな遺伝子を特徴量として抽出した。また、分化点と各細胞種における遺伝子ネットワーク構造を同時に推定するために、二重動的計画法 [5]に基づいた計算を行うが、計算量が増大するため、並列計算が可能となるように実装の改良を行なった。

## 4. 研究成果

本研究では、提案手法を造血幹細胞分化の scRNA-seq データに適用し有効性を検証した。擬似時系列データの作成については、軌道推定手法である Monocle [2]に適用し、擬似時系列推定による二次元座標をもとに、各細胞種において、中央値を示す細胞から数十細胞を抽出し、推定に用いる十分なサンプル数の発現量データを作成した。ネットワーク構造の推定に関しては、各遺伝子ごとに、LASSO 回帰によって二乗誤差を計算し、動的計画法を用いて二乗誤差の総和が最小となるように推定を行った。

遺伝子数が膨大なため、分散の大きな遺伝子を抽出した造血幹細胞分化のデータを用いて、遺伝子ネットワーク推定の計算機実験を行い、手法の有効性について検証したところ、数種の細胞種については、分化点を推定することが可能であり、疎な発現量データからも特徴量の推定を行えることが分かった。また、バルクのショウジョウバエ発生の時系列データに適用し、最小二乗法を用いた場合と同様の辺を推定できることを示した。

さらに、遺伝子数や細胞数が多いデータについては、計算量が増大するため、並列計算が可能となるように実装を行なった。特に、二乗誤差の計算や動的計画法を用いたネットワーク構造推定などの高速化を可能とする全計算過程に対して、マルチプロセスと共有メモリを用いたプロセス間通信を行う並列化を実装した。細胞数が多いデータに適用して計算機実験を行い、手法の有効性を検証したところ、シングルコアでの計算速度よりも、大幅に計算速度が向上することを確認した。

<参考文献>

- [1] N. Nakajima, Trends and outlook of single-cell analysis, JSBi Bioinformatics Review., 3:61-74, 2022.
- [2] J. Cao, M. Spielmann, X. Qiu et al., The single cell transcriptional landscape of mammalian organogenesis, Nature., 566:496-502, 2019.
- [3] G. L. Manno, R. Soldatov, A. Zeisel et al., RNA velocity in single cells, Nature., 560:494-498, 2018.
- [4] S. Lèbre, J. Becq, F. Devaux et al., Statistical inference of the time-varying structure of gene-regulation networks, BMC Biol., 4:130, 2010.
- [5] N. Nakajima and T. Akutsu, Exact and heuristic methods for network completion for time-varying genetic networks, Biomed Res Int., 2014:684014, 2014.
- [6] Z. Wu, S. Gao, C. Diamond et al., Sequencing of RNA in single cells reveals a distinct transcriptome signature of hematopoiesis in GATA2 deficiency, Blood Adv., 4:2656-2670, 2020.
- [7] S. Gao, Y. Chen, Z. Wu et al., Time-varying gene expression network analysis reveals conserved transition states in hematopoietic differentiation between human and mouse, Genes., 13:1890, 2022.

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計0件

〔図書〕 計0件

〔産業財産権〕

〔その他〕

仲嶋なつ 「シングルセルデータを用いた細胞分化における時変遺伝子ネットワーク推定手法」細胞, 55(7), 43-45, 2023.

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
---------	---------