

令和 4 年 6 月 1 日現在

機関番号：14401

研究種目：基盤研究(A) (一般)

研究期間：2018～2021

課題番号：18H04124

研究課題名(和文) ベイジアンネットワークによる遺伝子制御予測に基づく細胞多様性の解析手法の開発

研究課題名(英文) Development of cell diversity analysis method based on gene regulatory prediction by Bayesian network

研究代表者

松田 秀雄 (MATSUDA, Hideo)

大阪大学・情報科学研究科・教授

研究者番号：50183950

交付決定額(研究期間全体)：(直接経費) 27,200,000円

研究成果の概要(和文)：細胞分化や細胞の刺激応答など、細胞内の遺伝子制御が経時的に変化する生体現象について、1細胞RNAシーケンシングのデータを取得し、各細胞を疑似的な時間軸上で整列させた細胞系譜の推定手法を開発した。この細胞系譜推定により得られる系譜上の各細胞の遺伝子発現量を時系列発現プロファイルとみなして、細胞ごとの個々の遺伝子間の制御関係を新たに考案したエッジゲインと呼ぶスコアで定量化した。このスコアを基に、細胞系譜に沿って得られた時系列遺伝子発現プロファイルから動的ベイジアンネットワークモデルにより遺伝子制御ネットワークを推定することで、従来の手法よりも高い精度で推定できることを示した。

研究成果の学術的意義や社会的意義

本研究では、1細胞RNAシーケンシング技術により得られる細胞ごとの遺伝子発現プロファイルから、細胞分化や細胞の刺激応答などの進行過程を表す細胞系譜を推定する手法を開発した。さらに、細胞系譜上で各細胞の遺伝子発現量を抽出することで、非常に細かい時間間隔で時系列遺伝子発現プロファイルを構成して、動的ベイジアンネットワークモデルにより遺伝子制御ネットワークを推定する手法を開発した。実際に、造血幹細胞からの細胞分化や免疫細胞の刺激応答に本手法を適用することで、細胞系譜と遺伝子制御ネットワークを高い精度で推定することが示され、本手法が多様な生命現象に適用可能であることが示唆された。

研究成果の概要(英文)：For biological phenomena in which intracellular gene regulation changes over time, such as cell differentiation and cellular stimulus response, we have developed a cell lineage inference method that obtains data from single-cell RNA sequencing and maps each cell on a pseudo-temporal time. The cells were sorted by this cell lineage inference, and the gene expression levels of each cell were considered as a time-series expression profile, and the regulatory relationship between individual genes in each cell was quantified by a newly developed score called "edge gain". Based on this score, gene regulatory networks can be inferred by a dynamic Bayesian network model from the time-series gene expression profiles obtained along the cell lineage, and it was shown that the inference accuracy was higher than that of existing methods.

研究分野：バイオインフォマティクス

キーワード：遺伝子制御ネットワーク推定 動的ベイジアンネットワークモデル 1細胞RNAシーケンシング 細胞系譜推定 バイオインフォマティクス

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

(1) 数千個の細胞集団に対して一度に全細胞の遺伝子発現量を1細胞単位で計測できる1細胞RNAシーケンシング技術が開発されている。その結果、従来は均質だと思われていた細胞集団(例えば、同じ生体内の臓器から採取された細胞群)において、細胞ごとに個々の遺伝子の発現量がばらついていることが明らかになっていた。

(2) 1細胞RNAシーケンシングにより得られる遺伝子発現量のデータ(以下、1細胞発現データと呼ぶ)は、細胞と遺伝子で構成される2次元のマトリックスで表される。従来の解析では、各細胞を、遺伝子数分の次元を持つ高次元のベクトルと考え、主成分分析等による次元削減で低次元空間に投影したとき、近い位置に集まる細胞群をまとめることで、細胞のサブタイプを探索していた。

(3) 1細胞発現データには欠損値が存在することが多く、単純に主成分分析を行うと欠損値の影響で細胞がうまく分類できない場合があった。また、経時的に内部の状態を変化させる過程(例えば、細胞分化や、生体組織に対する刺激応答など)にある細胞集団では、単に個々の細胞を分類するだけでなく、その過程でどのような遺伝子制御ネットワークが働いていて、そのネットワークを構成している各遺伝子に対する制御が細胞ごとにどれだけ異なっているかを解き明かしたいという要求があった。

2. 研究の目的

(1) 数千個の細胞集団に対して一度に全細胞の遺伝子発現量を1細胞単位で計測できる1細胞RNAシーケンシング技術により、個々の細胞単位で遺伝子発現量を計測し、そこから遺伝子制御ネットワークを推定することで、細胞ごとにどのような遺伝子制御が存在するかを定量的に明らかにするための解析手法を開発することを目的とした。

(2) 細胞分化や細胞の刺激応答など、細胞内の遺伝子制御が経時的に変化する生体現象について、1細胞RNAシーケンシングのデータを取得し、細胞系譜推定により各細胞を時間軸上で整列させ、時間軸上での各細胞の遺伝子発現量を時系列発現プロファイルとみなして、ベイズ情報量基準に基づく動的ベイジアンネットワークモデルにより遺伝子制御ネットワークを推定することで、細胞ごとの個々の遺伝子間の制御の「強さ」を統一的な指標で定量化して比較・評価することを目指した。

3. 研究の方法

(1) 1細胞RNAシーケンシングにより得られた1細胞発現データから、細胞系譜解析により各細胞の時系列上で占める位置を推定する手法を開発した。具体的には、細胞分化や刺激応答などの細胞の状態遷移過程において、各細胞の状態に応じてクラスタリングにより細胞集団を分類した後、各細胞を順序を付けてソーティングする手法を開発した。従来、このような細胞の順序付けは、個々の生体現象ごとに、細胞の状態を反映していると考えられるマーカー遺伝子を使って行われてきた。しかし、マーカー遺伝子のみで細胞の状態を表すと、他の大多数の遺伝子の発現が反映されないことになり、細胞ごとの多様性の情報が失われる。そこで、まず大まかに細胞集団単位での系譜を求め、各細胞がいずれかの系譜に割り当たるように系譜を更新していくことにした。これにより、すべての細胞について系譜上の位置を表す疑似時刻を付与することができるようになった。

(2) 疑似時刻で並べられた1細胞発現データを時系列発現プロファイルとみなして、動的ベイジアンネットワークによる遺伝子制御ネットワークの推定を行った。時系列上の各時点で、ベイズ情報量基準を用いてネットワーク中の遺伝子間の制御関係の強さを時点ごとに求める手法を開発し、分岐を含む細胞系譜の始点から末端の端点までの直線的な経路について時系列発現プロファイルを求めて遺伝子制御ネットワークの推定に利用した。

4. 研究成果

(1) RNAシーケンシングにより、細胞の時系列での変化に伴う遺伝子配列に生じる変異の蓄積を評価する研究を行った。具体的には、がんの進行過程でがん細胞に生じる遺伝子変異の蓄積過程を表す融合遺伝子を探索する手法を開発した。遺伝子のRNAシーケンシングにより得られるリード配列(読み取り断片配列)を高速にソーティングするSlideSortを応用し、部分的に重複するリード配列を重ね合わせて伸長することによりmRNA配列を再構築した。再構築したmRNA配列をゲノム上にマッピングすることで、リード配列の部分配列がゲノム上で離れた位置にマッピングされる場合に融合遺伝子の候補として提示する手法を、乳がん培養細胞から得たトランスクリプトームデータに適用し、一部、PCR実験結果等との比較も行い、新規の融合遺伝子を探

索できることを示した。

(2) 1細胞 RNA シーケンシングにより得られた1細胞発現データについて、時系列上で占める位置を推定する手法を開発した。マウスの骨髄や末梢血由来の好中球や単球等の免疫細胞の1細胞トランスクリプトームの RNA シーケンシングデータを取得し、マクロファージや樹状細胞に分化することができる単球と、分化が完了している好中球など、複数種類の免疫細胞の状態に応じた順序を付けて、細胞分化の疑似的な経過時間を基に細胞系譜として推定することができた(図1)。図1で青色、緑色、黄色、オレンジ色、赤色、ピンク色、水色の点は、それぞれ骨髄由来、末梢血由来、脾臓由来、PMA 刺激6時間後、PMA 刺激24時間後、GMCSF 刺激、LPS 刺激の好中球細胞を指す。

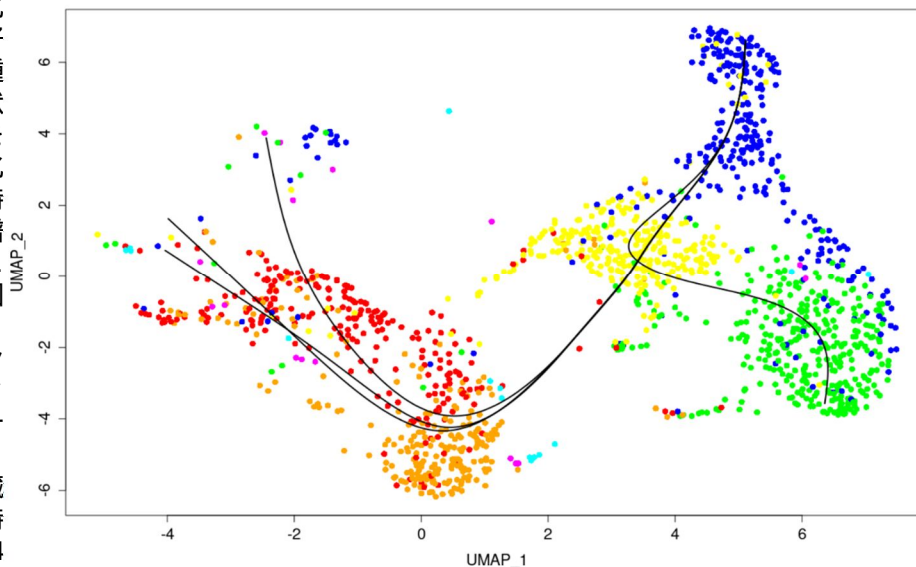


図1

(3) マウス脂肪組織のトランスクリプトームデータに対して遺伝子発現解析を行うことで、特定の組織由来の細胞集団でのみ特異的に発現する新規のノンコーディング RNA を発見した。このノンコーディング RNA は、寒冷刺激の初期に生じる Ucp1 の発現上昇に伴う熱産生を受けて、寒冷刺激時間に応じて継続的に発現が上昇することで、Ucp1 の発現を段階的に抑制し熱産生を抑えることが示された。これにより、寒冷刺激の時間経過に応じて熱産生を段階的に抑えていくというユニークな調節機構を持つことが示唆され、本研究での解析の有効性が示された。

(4) (1)で開発した細胞集団のクラスタリング手法を、公開されている口腔扁平上皮癌の患者由来の培養細胞株から取得された1細胞 RNA シーケンシングデータに適用したところ、抗がん剤投与前の薬剤感受性状態から、長期間投与後の薬剤抵抗性状態、さらには、薬剤抵抗性状態から薬剤投与を停止した薬剤休止状態のそれぞれの状態にある細胞群を、細胞株の違いによるバッチ効果の影響を除去して、それぞれの状態を表すクラスタにまとめることができた。また、(1)の細胞系譜推定手法により、薬剤感受性状態から、薬剤抵抗性状態を経て、薬剤休止状態に向かう一連の細胞系譜を検出することに成功した(図2)。次に、検出された細胞系譜において、発現変動を示す遺伝子を探索したところ、薬剤感受性状態から薬剤抵抗性状態に向けて、卵巣がんなどで薬剤抵抗性に関与することが報告されている遺伝子群が発現上昇していることが示された。

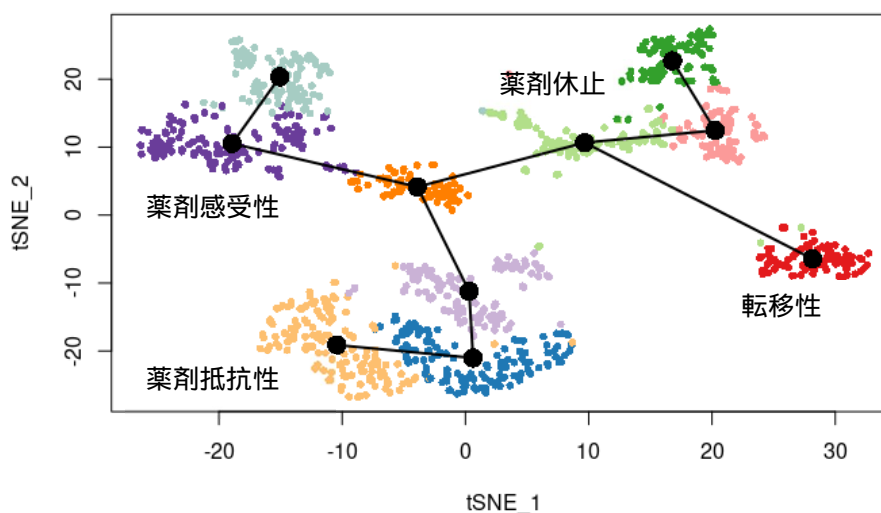


図2

(5) 1細胞 RNA シーケンシングデータから時系列遺伝子発現プロファイルを構成するため、細

胞系譜上に細胞を整列させて、遺伝子発現量が連続的に変化するようにフィッティングすることで、系譜に沿った連続的な時系列遺伝子発現プロファイルが得る方法を開発した。得られた時系列遺伝子発現プロファイルから、動的ベイジアンネットワークモデルにより遺伝子制御ネットワークを推定する手法を開発した。この方法では、エッジゲインと呼ばれる、ネットワークの制御辺ごとの遺伝子制御関係に与える寄与度をネットワークのスコアとするのが従来の手法にない特徴である。細胞系譜上でブートストラップサンプリングした細胞の遺伝子発現量から求めたエッジゲインをスコアとして動的ベイジアンネットワークモデルで遺伝子制御ネットワークを推定することで、細胞分化等の長時間にわたる生命現象での遺伝子制御ネットワークを高精度で求めることができた（図3）。

図3で、赤、緑、青の線はそれぞれ、エッジゲインによる動的ベイジアンネットワーク（本手法）、ブートストラップ確率による動的ベイジアンネットワーク、SINCERITIESの手法で推定を行った時のROC曲線を示す。本手法が最も高い精度を達成していることを示している。実際に、造血幹細胞から顆粒球に分化する過程での1細胞 RNA シーケンシングデータから、本手法により、細胞系譜の推定と時系列遺伝子発現プロファイルの構成および遺伝子制御ネットワークの推定を行ったところ、既知の分化マーカーの発現変動の比較から、従来の手法よりも高精度で細胞系譜と遺伝子制御ネットワークを推定できることが示された。

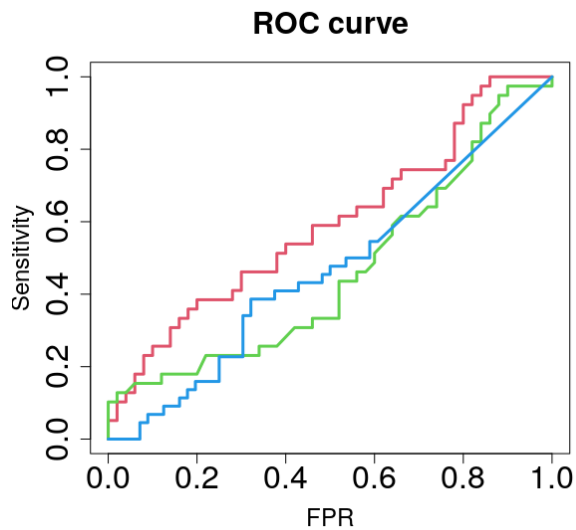


図3

5. 主な発表論文等

〔雑誌論文〕 計6件（うち査読付論文 6件/うち国際共著 1件/うちオープンアクセス 6件）

1. 著者名 Shuhei Yao, Kaito Uemura, Shigeto Seno, Hideo Matsuda	4. 巻 12
2. 論文標題 A Novel Method for Gene Regulatory Network Inference with Pseudotime Data Using Information Criterion	5. 発行年 2022年
3. 雑誌名 International Journal of Bioscience, Biochemistry and Bioinformatics	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
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 -

1. 著者名 Kosho Murayama, Hideo Matsuda	4. 巻 27
2. 論文標題 Detection of Biomarkers for Epithelial-Mesenchymal Transition with Single-Cell Trajectory Inference	5. 発行年 2022年
3. 雑誌名 Frontiers in Bioscience-Landmark	6. 最初と最後の頁 127 ~ 127
掲載論文のDOI（デジタルオブジェクト識別子） 10.31083/j.fbl2704127	査読の有無 有
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 -

1. 著者名 Kosho Murayama, Hideo Matsuda	4. 巻 -
2. 論文標題 A Method for Detection of Markers for Epithelial-Mesenchymal Transition based on Single Cell Transcriptomic Data	5. 発行年 2022年
3. 雑誌名 Proceedings of 12th International Conference on Bioscience, Biochemistry and Bioinformatics	6. 最初と最後の頁 57 ~ 62
掲載論文のDOI（デジタルオブジェクト識別子） 10.1145/3510427.3510436	査読の有無 有
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 -

1. 著者名 Kosho Murayama, Hideo Matsuda	4. 巻 11
2. 論文標題 Detecting Lineage-Specific Marker Genes for Tumor Evolution Based on Single Cell Transcriptome	5. 発行年 2021年
3. 雑誌名 International Journal of Bioscience, Biochemistry and Bioinformatics	6. 最初と最後の頁 50 ~ 57
掲載論文のDOI（デジタルオブジェクト識別子） 10.17706/ijbbb.2021.11.3.50-57	査読の有無 有
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 -

1. 著者名 Mari Iwase, Shoko Sakai, Shigeto Seno, Yu-Sheng Yeh, Tony Kuo, Haruya Takahashi, Wataru Nomura, Huei-Fen Jheng, Paul Horton, Naoki Osato, Hideo Matsuda, Kazuo Inoue, Teruo Kawada, Tsuyoshi Goto	4. 巻 84
2. 論文標題 Long non-coding RNA 2310069B03Rik functions as a suppressor of Ucp1 expression under prolonged cold exposure in murine beige adipocytes	5. 発行年 2020年
3. 雑誌名 Bioscience, Biotechnology, and Biochemistry	6. 最初と最後の頁 305 ~ 313
掲載論文のDOI (デジタルオブジェクト識別子) 10.1080/09168451.2019.1677451	査読の有無 有
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 該当する

1. 著者名 Yoshiaki Sota, Shigeto Seno, Hironori Shigeta, Naoki Osato, Masafumi Shimoda, Shinzaburo Noguchi, Hideo Matsuda	4. 巻 17
2. 論文標題 Improvement of detection performance of fusion genes from RNA-seq data by clustering short reads	5. 発行年 2019年
3. 雑誌名 Journal of Bioinformatics and Computational Biology	6. 最初と最後の頁 1940008 ~ 1940008
掲載論文のDOI (デジタルオブジェクト識別子) 10.1142/S0219720019400080	査読の有無 有
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3件 (うち招待講演 0件 / うち国際学会 3件)

1. 発表者名 Kaito Uemura, Naoki Osato, Hironori Shigeta, Shigeto Seno, Hideo Matsuda
2. 発表標題 A method for inferring gene regulatory networks based on pseudo time-series gene expression profiles from single-cell RNA-seq data
3. 学会等名 27th Conference on Intelligent Systems for Molecular Biology and 18th European Conference on Computational Biology (ISMB/ECCB) (国際学会)
4. 発表年 2019年

1. 発表者名 Naoki Osato, Hironori Shigeta, Shigeto Seno, Yutaka Uchida, Masaru Ishii, Hideo Matsuda
2. 発表標題 Single-cell transcriptome analysis for elucidating cell dynamics
3. 学会等名 Single cell biology meets diagnostics - 12th International workshop on approaches to single cell analysis (国際学会)
4. 発表年 2019年

1. 発表者名 Yoshiaki Sota, Shigeto Seno, Hironori Shigeta, Naoki Osato, Masafumi Shimoda, Shinzaburo Noguchi, Hideo Matsuda
2. 発表標題 Detection of fusion genes from human breast cancer cell-line RNA-Seq data using shifted short read clustering
3. 学会等名 IEEE 18th International Conference on Bioinformatics and Bioengineering (BIBE 2018) (国際学会)
4. 発表年 2018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
研究分担者	瀬尾 茂人 (SENO Shigeto) (30432462)	大阪大学・情報科学研究科・准教授 (14401)	

	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
研究協力者	河田 照雄 (KAWADA Teruo) (10177701)	京都大学・農学研究科・教授 (14301)	

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