

令和 4 年 6 月 1 日現在

機関番号：14401

研究種目：挑戦的研究（萌芽）

研究期間：2019～2021

課題番号：19K22894

研究課題名（和文）情報量基準に基づく細胞集団の多重分類・比較方式の開発と細胞アトラスへの応用

研究課題名（英文）Development of multiple classification and comparison methods of cells based on information criteria and their applications to Cell Atlas

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） 5,000,000円

研究成果の概要（和文）：1細胞RNAシーケンシングにより取得した1細胞発現プロファイルを、細胞アトラスと細胞集団間の類似性を比較することを目的として、比較を妨げる要因であるバッチ効果を補正して細胞集団をクラスタリングする手法を開発した。これを発展させて、細胞集団の細胞系譜を推定し、2群で異なる発現変動を示すマーカー遺伝子を検出する手法を開発した。実際に、ヒトサンプルから得られた疾患型と健常型の2種類の細胞のデータに本手法を適用したところ、T細胞の分化過程を反映した細胞系譜が得られ、さらにその系譜上で疾患型の細胞集団に特徴的な発現変動を示す新規のマーカー遺伝子を検出でき、本手法の有効性が確認できた。

研究成果の学術的意義や社会的意義

研究者は、自身の取得した細胞サンプルの遺伝子発現プロファイルを問合せとして細胞アトラスを検索し、類似したプロファイルを持つ細胞の情報を得ることで、細胞の機能についての手がかりをつかむことが期待される。しかし、異なるサンプル間でのデータの偏りにより生じるバッチ効果のため、細胞サンプル間の比較は容易ではなかった。本研究は、複数の細胞集団で、バッチ効果を補正して相互の類似性を基にクラスタリングする手法を開発した。さらに、この手法を応用して、疾患型と健常型など生物学的背景の異なる細胞集団間で分化系譜を推定し、対応する系譜間で疾患型のみで強く発現するマーカー遺伝子を検出することができるようになった。

研究成果の概要（英文）：Aiming to compare the similarity between cell atlas and cell populations by comparing expression profiles obtained by single-cell RNA sequencing, we developed a method for clustering cell populations by removing batch effects, a factor that hinders comparison. Expanding on this, we developed a method to infer the cell lineages of cell populations and to detect biomarker genes that are differentially expressed in the two groups. In fact, when this method was applied to the data of two types of T cells obtained from human samples, a cell lineage reflecting the differentiation process of T cells was obtained, and a new marker gene that shows characteristic expression variation in the cell population of the disease type was detected on the lineage, confirming the validity of this method. The effectiveness of this method was demonstrated by this analysis results.

研究分野：バイオインフォマティクス

キーワード：クラスタリング 細胞系譜推定 細胞アトラス 1細胞RNAシーケンシング解析 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) 1細胞 RNA シーケンシング技術の発展により、ヒトやマウスなどの生物の様々な臓器・組織を構成している細胞集団の持つ遺伝子発現プロファイルが得られつつある。遺伝子発現プロファイルとは、細胞ごとに個々の遺伝子の発現量を成分とするベクトルであり、遺伝子数を次元とするためヒトやマウスでは数万次元という高次元ベクトルとなる。得られた遺伝子発現プロファイルは、採取した個体・臓器・組織や、それらの環境条件、細胞の種類に関する情報などの付随情報と共に、細胞アトラスという網羅的な細胞情報データベースにまとめられている。

(2) 研究者は、自身の取得した細胞サンプルの遺伝子発現プロファイルを問合せとして細胞アトラスを検索し、類似したプロファイルを持つ細胞の情報を得ることで、細胞サンプルの機能についての手がかりをつかむことが期待されるが、別々の 1細胞 RNA シーケンシングにより得られた遺伝子発現プロファイルでは、単純にデータを併合して次元削減するとバッチ効果のため、同一組織・同一細胞種であっても分離してしまうことが多かった。

### 2. 研究の目的

(1) 遺伝子発現プロファイルを次元削減による情報損失なしに細胞アトラスと比較できるようにすることを研究の目的とした。このためには、別々の 1細胞 RNA シーケンシングで得られた遺伝子発現プロファイルであっても、バッチ効果で細胞集団が分離してしまうことなしに、細胞単位で細胞アトラスと比較する手法を開発することを目指した。

(2) (1)で開発した手法を、実際に細胞分化や細胞の刺激応答での 1細胞 RNA シーケンシングで得られた複数の遺伝子発現プロファイルに適用して、クラスタリングや細胞系譜解析を行うことで、手法の有効性を評価することを目指した。

### 3. 研究の方法

(1) バッチ効果を補正したクラスタリング手法を開発し、同一組織由来の細胞集団について、異なる実験条件で個別に 1細胞 RNA シーケンシングを行うことにより得られる複数の遺伝子発現プロファイルに適用することで性能を評価した。

(2) (1)のクラスタリング手法を、細胞分化や細胞の刺激応答の進行過程に応じて発現が変化するような細胞集団に適用することで、クラスタリング結果から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、この手法により、実際に細胞分化の 1細胞遺伝子発現プロファイルから細胞分化の系譜を推定することで、手法の有効性を評価した。

### 4. 研究成果

(1) 相互近傍法に基づいてバッチ効果を補正し、細胞集団のクラスタリングを行う手法を開発した。マウスの胚性幹細胞の野生型と特定の遺伝子をノックアウトした変異体について、それぞれ神経細胞への分化誘導をかけて、分化誘導前と分化誘導後 5 日後のサンプルについて 1細胞 RNA シーケンシングを行うことで取得した 1細胞遺伝子発現プロファイルに、本手法を適用した結果を図 1 に示す。図 1 で青緑色の点が野生型、赤色の点の変異体を表す。分化誘導前の細胞集団は左側の丸い部分であり、野生型と変異体が共に重なってまとまっていることから、バッチ効果が補正できていることがわかる。また、分化誘導後 5 日後の

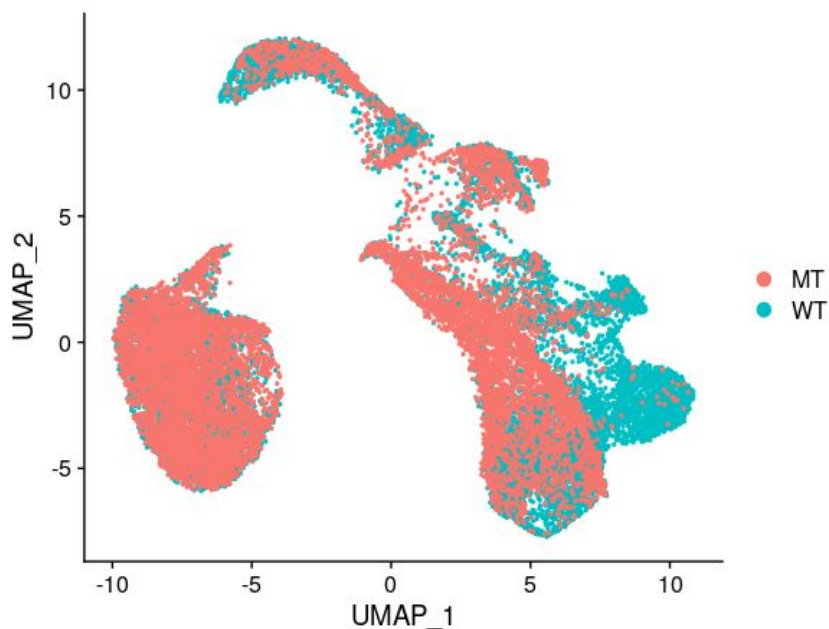
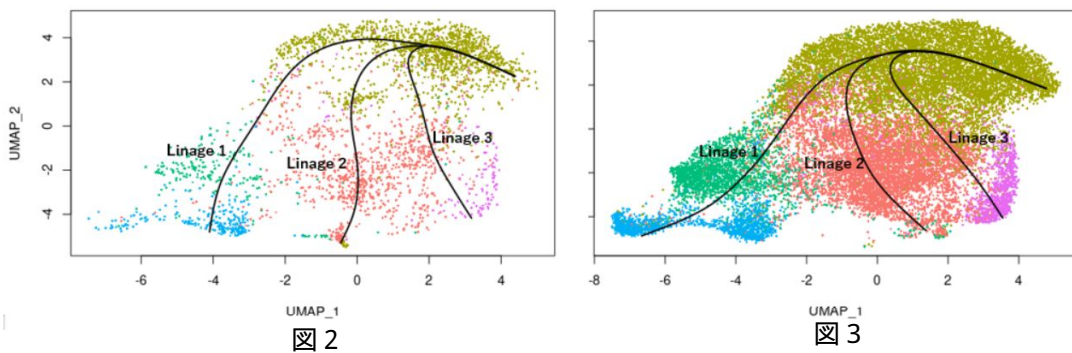


図 1

サンプルは、図1の右側に広範囲に広がっており、特に右下の部分において変異体は野生型と比べてより分化誘導前の細胞集団に近い位置を占めている。実際に、この変異体は分化抵抗性の表現型を示しており、図1のクラスタリング結果と整合している。以上のことから、本研究のクラスタリング手法は、バッチ効果を正しく補正してクラスタリングできていることがわかる。

(2) 生物学的背景の異なる2群(例えば、疾患型と健常型)から取得した1細胞RNAシーケンシングデータを入力として、細胞集団の応答過程に基づく経時的な細胞系譜を推定する手法を開発した。従来の細胞系譜推定手法は、細胞集団を遺伝子発現量の差異によりクラスタリングで分類し、クラスタ間の距離を基にした最小全域木により細胞系譜を求めていた。これに対して本研究では、遺伝子発現量だけでなく各クラスタの細胞集団中に占める細胞種の割合をクラスタ間の距離計算で考慮することにより、より生物学的な変化過程を反映した細胞系譜を推定することを目指した。実際に、ヒトサンプルから得られた疾患型と健常型の2種類のT細胞について、ナイーブT細胞から始まる分化過程で得られた1細胞RNAシーケンシングのデータに本手法を適用したところ、T細胞の分化過程を反映した細胞系譜が得られた(図2、図3)。図2と図3はそれぞれ、健常型とサルコイドーシスという疾患でのT細胞の細胞系譜を示しており、左から順にガンマ・デルタT細胞、メモリT細胞、制御性T細胞への分化を表している。この図に示されているように、異なる条件の細胞集団であっても、バッチ効果を補正することで、同じ種類の細胞が対応する場所に配置されることが示された。



(3) 図2と図3の細胞集団において、推定された細胞系譜上で健常型と疾患型で異なる発現変動を示すマーカー遺伝子を検出する手法を開発した。従来は、個々の細胞集団ごとに細胞系譜上での発現変動遺伝子を個別に検出し、遺伝子セット間の差異を見ていたが、この手法であると群間での発現変動の量的な違いを考慮できていなかった。本手法では、発現変動遺伝子の検出過程で、情報量基準に基づいて2群間の統計量の差異を比較解析することで、群間での発現変動の大きさを考慮した特徴的な遺伝子の検出ができるようになった。実際に、図2、図3の系譜上で発現変動の異なる遺伝子を検出した結果、健常型と比べて疾患型の細胞集団に特徴的な発現変動を示す新規のマーカー遺伝子が検出でき、本手法の有効性が確認できた。検出されたマーカー遺伝子を図4に示す。

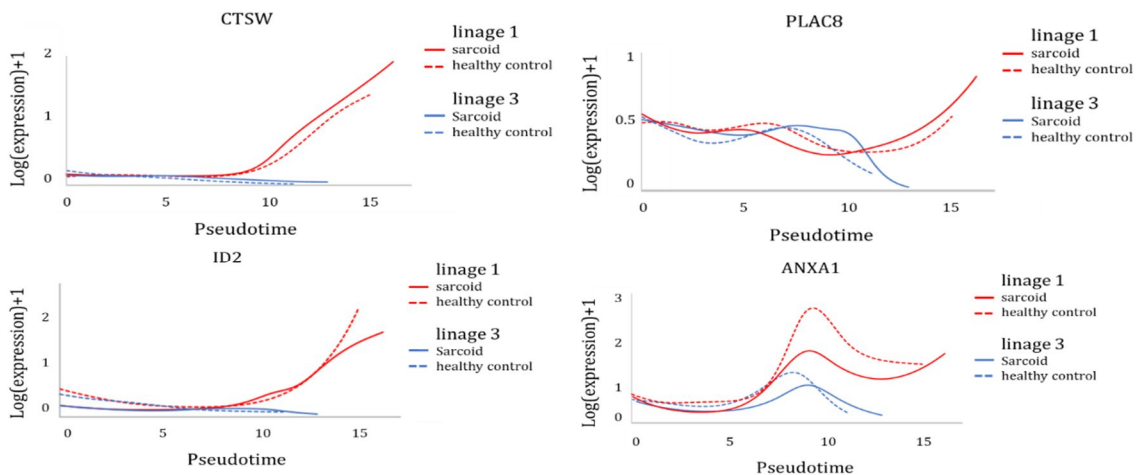


図4

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2件/うち国際共著 0件/うちオープンアクセス 2件）

1. 著者名 Akihiro Nomura, Hideo Matsuda	4. 巻 12
2. 論文標題 Identification of Lineage Markers for T Cell Immune Dysregulation in Sarcoidosis using Single-Cell RNA-seq	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 -
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 -

〔学会発表〕 計2件（うち招待講演 0件/うち国際学会 2件）

1. 発表者名 Takumi Adachi, Junko Yoshida, Shigeto Seno, Kyoji Horie, Hideo Matsuda
2. 発表標題 Cell trajectory inference for revealing differentiation process of mouse stem cells using single cell RNA-seq data
3. 学会等名 28th Conference on Intelligent Systems for Molecular Biology (ISMB) (国際学会)
4. 発表年 2020年

1. 発表者名 Shuhei Yao, Hironori Shigeta, Shigeto Seno, Hideo Matsuda
2. 発表標題 A study on novel estimation method of cell differentiation lineage by single cell trajectory inference
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

〔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	石井 優  (ISHII Masaru)  (10324758)	大阪大学・生命機能研究科・教授    (14401)	

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