

令和 4 年 6 月 9 日現在

機関番号：63801

研究種目：新学術領域研究（研究領域提案型）

研究期間：2017～2021

課題番号：17H06399

研究課題名（和文）単一細胞シーケンスデータに基づく細胞社会学のための情報手法の開発とデータ解析

研究課題名（英文）Study of Information analysis and modelinig

研究代表者

池尾 一穂 (IKEO, KAZUHO)

国立遺伝学研究所・ゲノム・進化研究系・准教授

研究者番号：20249949

交付決定額（研究期間全体）：（直接経費） 105,400,000円

研究成果の概要（和文）：単一細胞データポータルを構築し、領域内研究グループから収集したデータをの解析を行い、その全データをダウンロード・表示できる形でまとめ公開した。細胞集団の動態モデルの構築に向けて、1細胞レベルの体細胞変異パターンを用いて、既知の知識と矛盾のない細胞系譜を推定できることを示した。また、従来の機械学習方法を見直し、より少ない細胞の発現データから、遺伝子の相互作用や共発現に関する有用な情報を抽出するための数学的手法の開発に成功した。機械学習モデルを用いて、APA破綻遺伝子リストのエンリッチ解析から疾患関連遺伝子抽出を可能とした。

研究成果の学術的意義や社会的意義

今後ますます増加していくシングルセルデータの解析の発展に大きく貢献することが期待される解析フローを構築した。これは、今後、医学領域に加え、基礎生物学領域での利用も期待される。また、少数細胞のデータから既知の細胞種を選別することばかりでなく、新規の細胞種の定義を定量的に行うことが可能となった。細胞系譜再構築手法を加えて、シングルセルレベルのデータを大規模かつ効率的に解析する手法は今後、様々な単一細胞研究に貢献することが期待される。炎症細胞社会データベースとして公開した領域全体のデータは、様々な疾患に関わる炎症の分子レベルでの理解の助けとなることが期待される。

研究成果の概要（英文）：We constructed a single-cell data portal, analyzed the data collected from the research groups in the area, and released all the data in a form that can be downloaded and displayed.

It was shown that a cell lineage consistent with known knowledge can be estimated using a somatic mutation pattern at the 1-cell level for the construction of a dynamic model of a cell population. We also reviewed conventional machine learning methods and succeeded in developing a mathematical method for extracting useful information on gene interactions and co-expressions from less cell expression data. Using a machine learning model, we made it possible to extract disease-related genes from the enrichment analysis of the APA disrupted gene list.

研究分野：比較ゲノム、バイオインフォマティクス

キーワード：単一細胞トランスクリプトーム 炎症細胞の遷移 NGSデータ解析 細胞系譜再構築 データベース

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

急速に高齢化が進行する我が国の社会を持続可能なものとするためには、慢性炎症を基盤とする生活習慣病・線維症・がんなどを予防し、また一度発症しても早期に診断・介入する健康維持システムの構築が喫緊の課題である。従来の予防医学では、疫学および生化学・細胞生物学に基づき危険因子を推測し、健康リスクを低減するアプローチに主眼が置かれてきた。古典的な疫学は、疾患の発症における多数の危険因子の存在、例えば自然環境、社会環境、労働環境などの外的環境因子や、生活習慣などで変化する内的環境因子を詳らかにし、時として疾患の克服をもたらした。また、ヒトゲノムプロジェクトに端を発したゲノム疫学の進歩は、ゲノム点変異などの遺伝要因の解明に加えて、多くの難治性疾患の発症には遺伝要因以外の外的・内的環境因子が大きな影響を与えること、すなわち可塑性があるが故に予防可能であることも明らかにした。しかしながら、日常生活から全ての危険環境因子を排除することは至難であることから、予防医学における現実的な課題は、疾患の生物学的な発症機序に基づくリスクの重み付けと、バックアップとしての疾患の早期診断・介入手段の確立といえる。生化学・細胞生物学的な疾患へのアプローチは、慢性炎症をもたらす個々の環境構成物質、例えば活性酸素種や毒性物質などの化学的因子や微生物成分などに対する細胞・分子レベルのストレス応答を解明してきた。一方、臓器・個体レベルにおいて、**細胞・組織・個体という各階層で働く恒常性維持機構**(例えば免疫システムによる非自己・異常自己の排除や、傷害部位における治癒応答など)が、ストレス応答によりいかに破綻し、慢性炎症性疾患の病像に帰結するのかは、未だブラックボックスの中にあり、それらの機序に対する理解は、予防医学に求められる要求を満たすに至っていない。

2. 研究の目的

炎症記憶の形成過程を細胞や分子レベルでの細胞間相互作用(ここで「炎症細胞社会」とよぶ)として解明するために、数千個の細胞からなる炎症組織の構成細胞それぞれについて、以下の定性的・定量的情報を「細胞状態変数」として収集蓄積し、統合する。「細胞状態変数」の経時的变化(経時プロファイル)、空間的变化(空間プロファイル)を(隣接)細胞間で比較し、相関を検出する。特に転写産物相対量のプロファイルが異なる細胞間で相関する場合、その遺伝子同士は細胞間で機能的に相互作用していると仮定できる。そこで、そのような相関する遺伝子について、既知機能情報に基づき、相互作用によりもたらされる生化学・分子生物学・細胞生物学的帰結を推定しモデル化を試みる。さらに、全ての相互作用の総体を、炎症組織において観察される様々な状態変化と関連づけ「場の記憶」を細胞ならびに細胞集団の状態変化として記述する。それは一種の仮説であるので、この過程を経て、「場の記憶」を理解することを目指す。

3. 研究の方法

炎症記憶の形成過程を細胞や分子レベルでの細胞間相互作用として解明を目指し、数千個の細胞からなる炎症組織の構成細胞について、定性的・定量的情報を「細胞状態変数」として収集蓄積、統合するために必要な手法の開発とそれを用いた実際のデータ解析を行うために、単一細胞プロファイルに基づく、「炎症の場」としての細胞社会の記述を可能とすることを旨とし以下のように研究を進めた。1)単一細胞トランスクリプトーム動態の解析・モデリングのための単一細胞遺伝子発現プロファイルデータ解析手法の確立と解析(池尾)では、各種解析フローの整備、改良と一時データの解析を進める2)細胞集団の動態モデルの構築に向けて、様々な情報学的手法とシミュレーションを用いたモデル構築のための手法のモデル構築へ応用(太田)では、突然変異情報を抽出し遺伝子変異に基づく細胞系譜作成を行うアルゴリズムの開発を進める。1)細胞トランスクリプトーム配列データを用いて1)細胞レベルで体細胞変異の検出を行い、その突然変異パターンを用いて細胞の系統樹(すなわち、細胞系譜樹)を推定するための解析パイプラインの開発を行う。3)各細胞状態に対

応するマーカー遺伝子候補の抽出とそれらの機能情報に基づく炎症記憶形成における役割の推定と検証(渡邊)では、細胞分類、新規細胞の特徴づけを行うため、機械学習や統計解析を含むこれまでの単一細胞遺伝子発現研究の問題である細胞数不足の問題を回避する手法を開発する4)機械学習モデルから炎症細胞へ遷移する細胞の予測(小倉)。また、全員で協力して計画班を中心にA01,A02の各グループに対して、データ解析の支援をおこなう。

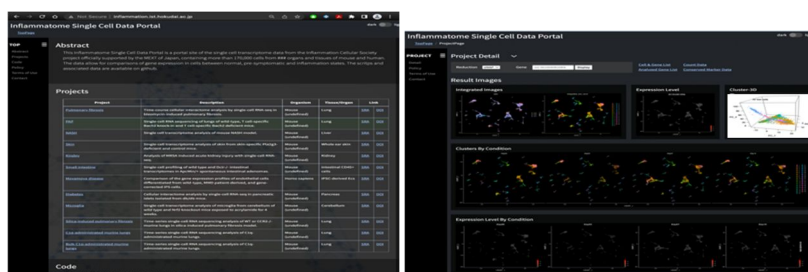
4. 研究成果

炎症記憶の形成過程を細胞や分子レベルでの細胞間相互作用として解明を目指し、数千個の細胞からなる炎症組織の構成細胞について、定性的・定量的情報を「細胞状態変数」として収集蓄積、統合するために必要な手法の開発とそれを用いた実際のデータ解析を進めた。単一細胞プロファイルに基づく、「炎症の場」としての細胞社会の記述を可能とすることを目指し以下のように研究を進めた。1)単一細胞トランスクリプトーム動態の解析・モデリングのための単一細胞遺伝子発現プロファイルデータ解析手法の確立と解析(池尾)では、コンテナ技術を導入する事により各種シングルセル解析パイプラインを遺伝研スパコン上で利用可能とし従来から10倍以上の性能改善をおこなった。また、この性能改善を受けて、従来のマッピング手法を改良する事により、全ゲノムリファレンス配列を用いたシングルセルデータマッピングを実用に耐える性能で可能とし、ノンコーディングRNAなども対象とした解析を可能とした。2)細胞集団の動態モデルの構築に向けて、様々な情報的手法とシミュレーションを用いたモデル構築のための手法のモデル構築へ応用(太田)では、突然変異情報を抽出し遺伝子変異に基づく細胞系譜作成を行うアルゴリズムを開発し実データの解析を行なった。1細胞トランスクリプトーム配列データを用いて1細胞レベルで体細胞変異の検出を行い、その突然変異パターンを用いて細胞の系統樹(すなわち、細胞系譜樹)を推定するための解析パイプラインを開発し、領域内で得られたデータを解析することが当初の目的であった。このうち、解析パイプラインの開発は終了しており、細胞系譜が既知のパブリックデータ(ヒトの胎盤組織)を用いて検証を行なっている。また、島野班の1細胞トランスクリプトーム配列データを用いて、実際の解析を行なった。その意味で、当初の目的はほぼ達成できたと考えている。その一方で、計算時間の短縮や情報量の少ないデータをどう扱うかは継続して改良すべき課題である。3)各細胞状態に対応するマーカー遺伝子候補の抽出とそれらの機能情報に基づく炎症記憶形成における役割の推定と検証(渡邊)では、実際のデータに対して細胞分類、新規細胞の特徴づけを進めた。機械学習や統計解析を含むこれまでの単一細胞遺伝子発現研究では、個々の遺伝子の細胞状態への寄与のみが論じられてきた。しかし、実際の細胞内での遺伝子は、単体で機能するばかりでなく、他の遺伝子や翻訳産物との相互作用を通して機能するため、遺伝子間相互作用、すなわち、2次以上の項も考慮した機械学習や統計学的解析が必要である。従来、その必要な解析が行われていないのは、たとえば2次の項を加える場合、遺伝子数の二乗の数の項の重みを計算する必要が生じ、細胞数の不足がより顕著となるからである。この問題に対して、研究項目A03において、各遺伝子は他の全遺伝子と一様に相互作用するわけではない、という生物学的知識に基づく独自の数学的手法を開発し、細胞数不足の問題を回避することで、2次の項(相互作用項)についても重み(寄与度)を計算することを世界で初めて可能とした。実際、実験によって明らかにされた遺伝子間制御関係などを、単一細胞遺伝子発現データの解析によって示すことができるということを確認した。この手法は、単一細胞遺伝子発現データに特化したものではなく、多くの生物学的データに応用が可能であり、革新的・創造的学術研究の発展を促す可能性が十分にある。4)機械学習モデルから炎症細胞へ遷移する細胞の予測(小倉)。機械学習を用いて、細胞の分類を行い細胞の遷移を遺伝子発現による細胞系

列の分類手法の試行を進め、実データへの適用を進めた。また、全員で協力して計画班を中心に A01, A02 の各グループに対して、データ解析の支援をおこない、10 機関、総リード数にして 200 億リード以上の解析処理を行うとともに、特徴遺伝子抽出、クラスタリングなどの基本的な解析に加え、細胞腫の推定、細胞系譜構築、細胞遷移、スプライスバリエーション推定などの個別の解析を行い提供した。

すでに述べたように、単一細胞トランスクリプトーム動態の解析・モデリングのための単一細胞遺伝子発現プロファイルデータ解析手法の確立では、プロセスを大幅に見直し改良を加える事により、大規模なシングルセルデータの処理能力を大幅に改善した。これは、今後のデータの増大に大いに役立つとともに将来の全ゲノム配列に対する解析をも可能とする大事な進歩である。また、これらの解析はパイプライン化され、今後、さまざまなプロジェクトで簡便に使用することが可能である。

研究項目 A03 において、単一細胞データポータル（データ収集・解析・公開用サーバー）を構築し、異なる組織から得た正常・炎症過程の細胞の単一細胞遺伝子発現データを領域内研究グループから収集し、統一したデータへの変換と統一パラメータ・手法による解析を行い、その全てのデータをダウンロード・表示できる形でまとめている。



細胞集団の動態モデルの構築に向けて、様々な情報的手法とシミュレーションを用いたモデル構築のための手法のモデル構築では、1細胞トランスクリプトーム配列データから検出した1細胞レベルで体細胞変異パターンを用いて、既知の知識と矛盾のない細胞系譜を推定できることがわかった。この結果は擬似時間解析の結果と整合性があるだけでなく、擬似時間解析の結果のみからは推定できない細胞軌道の方向性についての情報も含んでいる。私たちはこの方法を Realtime Course Analysis と名付け、特許出願準備中である。

各細胞状態に対応するマーカー遺伝子候補の抽出とそれらの機能情報に基づく炎症記憶形成における役割の推定では、従来の細胞種推定のための機械学習アプローチを見直す事により、少ない束縛条件（少数細胞の発現データから）でも、遺伝子の相互作用や共発現に関する有用な情報を抽出するための数学的手法の開発を行い成功した。

機械学習モデルから炎症細胞へ遷移する細胞の予測では、選択的ポリアデニル化 (APA) とは、細胞が状況に応じて RNA の様々な部位に PolyA-tail を付加する現象で、ポリアデニル化破綻を推定する手法を開発し、APA 破綻遺伝子リストのエンリッチ解析から疾患関連遺伝子抽出を可能とした。

5. 主な発表論文等

〔雑誌論文〕 計16件（うち査読付論文 14件／うち国際共著 3件／うちオープンアクセス 13件）

1. 著者名 Satoshi Oota, Kuniya Abe, Hideo Yokota, *Kazuho Ikeo.	4. 巻 -
2. 論文標題 Real-Time Course: Reconstruction of cellular diversity and lineage trajectory based on somatic mutational patterns detected from low-pass single-cell transcriptome data	5. 発行年 2022年
3. 雑誌名 Research Square	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) 10.21203/rs.3.rs-1558242/v1	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 Iwata Y, Satou K, Furuichi K, Yoneda I, Matsumura T, Yutani M, Fujinaga Y, Hase A, Morita H, Ohta T, Senda Y, Sakai-Takemori Y, Wada T, Fujita S, Miyake T, Yasuda H, Sakai N, Kitajima S, Toyama T, Shinozaki Y, Sagara A, Miyagawa T, Hara A, Shimizu M, Kamikawa Y, Ikeo K, Shichino S, Ueha S, Nakajima T, Matsushima K, Kaneko S, Wada T.	4. 巻 91
2. 論文標題 Collagen adhesion gene is associated with bloodstream infections caused by methicillin-resistant Staphylococcus aureus.	5. 発行年 2020年
3. 雑誌名 Int J Infect Dis	6. 最初と最後の頁 22-31
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.ijid.2019.11.003	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 Iwata Y, Sakai N, Yoneda I, Satou K, Furuichi K, Senda Y, Sakai-Takemori Y, Wada T, Fujita S, Ogura H, Sato K, Minami T, Yamaguchi K, Kitajima S, Toyama T, Yamamura Y, Miyagawa T, Hara A, Shimizu M, Sakai Y, Ikeo K, Shichino S, Ueha S, Nakajima T, Matsushima K, Wada T.	4. 巻 26(6)
2. 論文標題 The increased frequency of methicillin-resistant Staphylococcus aureus with low MIC of beta-lactam antibiotics isolated from hospitalized patients.	5. 発行年 2020年
3. 雑誌名 J Infect Chemother	6. 最初と最後の頁 604-610
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.jiac.2020.01.016	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 Hashimoto S, Shirasaki T, Yamashita T, Iwabuchi S, Suzuki Y, Takamura Y, Ukita Y, Deshimaru S, Okayama T, Ikeo K, Kuroki K, Kawaguchi K, Mizukoshi E, Matsushima K, Honda M, Kaneko S	4. 巻 16(2)
2. 論文標題 DOCK11 and DENND2A play pivotal roles in the maintenance of hepatitis B virus in host cells.	5. 発行年 2020年
3. 雑誌名 PLoS One	6. 最初と最後の頁 e0246313
掲載論文のDOI (デジタルオブジェクト識別子) 10.1371/journal.pone.0246313	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 Iwata Y, Sakai N, Yoneda I, Satou K, Furuichi K, Senda Y, Sakai-Takemori Y, Wada T, Fujita S, Ogura H, Sato K, Minami T, Yamaguchi K, Kitajima S, Toyama T, Yamamura Y, Miyagawa T, Hara A, Shimizu M, Sakai Y, Ikeo K, Shichino S, Ueha S, Nakajima T, Matsushima K, Wada T.	4. 巻 26(6)
2. 論文標題 The Increased Frequency of Methicillin-Resistant Staphylococcus Aureus With Low MIC of Beta-Lactam Antibiotics Isolated From Hospitalized Patients	5. 発行年 2020年
3. 雑誌名 Journal of Infection and Chemotherapy	6. 最初と最後の頁 604-610
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.jiac.2020.01.016	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 Iwata Y, Satou K, Furuichi K, Yoneda I, Matsumura T, Yutani M, Fujinaga Y, Hase A, Morita H, Ohta T, Senda Y, Sakai-Takemori Y, Wada T, Fujita S, Miyake T, Yasuda H, Sakai N, Kitajima S, Toyama T, Shinozaki Y, Sagara A, Miyagawa T, Hara A, Shimizu M, Kamikawa Y, Ikeo K, Shichino S, Ueha S, Nakajima T, Matsushima K, Kaneko S, Wada T	4. 巻 91
2. 論文標題 Collagen adhesion gene is associated with bloodstream infections caused by methicillin-resistant Staphylococcus aureus	5. 発行年 2020年
3. 雑誌名 International Journal of Infectious Diseases	6. 最初と最後の頁 22-31
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.ijid.2019.11.003	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 Tabuchi Y, Hirohashi Y, Hashimoto S, Mariya T, Asano T, Ikeo K, Kuroda T, Mizuuchi M, Murai A, Uno S, Kawai N, Kubo T, Nakatsugawa M, Kanaseki T, Tsukahara T, Saito T, Torigoe T.	4. 巻 106
2. 論文標題 Clonal analysis revealed functional heterogeneity in cancer stem-like cell phenotypes in uterine endometrioid adenocarcinoma.	5. 発行年 2019年
3. 雑誌名 Exp Mol Pathol.	6. 最初と最後の頁 78-88
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.yexmp.2018.11.013.	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Kinjo S, Monma N, Misu S, Kitamura N, Imoto J, Yoshitake K, Gojobori T, Ikeo K.	4. 巻 2018
2. 論文標題 Maser: one-stop platform for NGS big data from analysis to visualization.	5. 発行年 2018年
3. 雑誌名 Database (Oxford).	6. 最初と最後の頁 1-12
掲載論文のDOI (デジタルオブジェクト識別子) 10.1093/database/bay027.	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7件（うち招待講演 4件 / うち国際学会 6件）

1. 発表者名 Kazuho Ikeo
2. 発表標題 Reconstruction of cell lineage trees by using single-cell RNA sequence data
3. 学会等名 Genome Concept Centennial Conference (国立遺伝学研究所国際研究集会) (招待講演) (国際学会)
4. 発表年 2021年

1. 発表者名 Oota S, Ikeo K
2. 発表標題 Reconstruction of a phylogenetic tree of cells by using single-cell human transcriptome data
3. 学会等名 SMBE2021 (国際学会)
4. 発表年 2021年

1. 発表者名 Hidemi Watanabe
2. 発表標題 Single-cell transcriptomics as big data biology
3. 学会等名 International Symposium on Big-Data, Cybersecurity and IoT, Zoom conference (招待講演) (国際学会)
4. 発表年 2021年

1. 発表者名 柳澤諒, 渡邊日出海
2. 発表標題 単一細胞解析における正規化及び特徴選択に関する研究
3. 学会等名 日本分子生物学会年会
4. 発表年 2020年

1. 発表者名 Oota S, Ikeo K
2. 発表標題 Reconstruction of a cell lineage tree by using single-cell level somatic mutations with a distance matrix-based approach.
3. 学会等名 SMBE2019 (国際学会)
4. 発表年 2019年

1. 発表者名 Hidemi Watanabe
2. 発表標題 Machine-learning-based Cell Type Identification
3. 学会等名 1st International Symposium on Inflammation Cellular Sociology (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 Kazuho Ikeo
2. 発表標題 Cell type and marker gene prediction for single-cell transcriptome analysis
3. 学会等名 1st International Symposium on Inflammation Cellular Sociology (招待講演) (国際学会)
4. 発表年 2019年

〔図書〕 計0件

〔出願〕 計1件

産業財産権の名称 細胞の系統解析を行う方法	発明者 太田聡史	権利者 同左
産業財産権の種類、番号 特許、2022-23985	出願年 2022年	国内・外国の別 国内

〔取得〕 計0件

〔その他〕

単一細胞データポータル(データ収集・解析・公開用サーバー)
 Inflammatome Single Cell Data Portalサーバー, <https://inflammation.ist.hokudai.ac.jp/>

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	渡邊 日出海 (watanabe hidemi) (30322754)	北海道大学・情報科学研究院・教授 (10101)	
研究分担者	太田 聡史 (oota satoshi) (30391890)	国立研究開発法人理化学研究所・光量子工学研究センター・専任研究員 (82401)	
研究分担者	小倉 淳 (ogura atsushi) (60465929)	長浜バイオ大学・バイオサイエンス学部・教授 (34204)	

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