

科学研究費助成事業 研究成果報告書

令和 5 年 6 月 8 日現在

機関番号：82401

研究種目：若手研究

研究期間：2019～2022

課題番号：19K20406

研究課題名（和文）非負値テンソル分解を用いたオーファン受容体結合リガンドの同定及び機能解明

研究課題名（英文）Identification and functional elucidation of orphan receptor binding ligands using non-negative tensor decomposition

研究代表者

露崎 弘毅 (Tsuyuzaki, Koki)

国立研究開発法人理化学研究所・生命機能科学研究センター・基礎科学特別研究員

研究者番号：70769520

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）：本研究の基盤技術である、非負値テンソル分解を使用して、1細胞RNA-Seqデータに含まれる細胞間相互作用（Cell-Cell Interaction: CCI）を検出し、受容体とリガンドのペアを同定するための解析手法scTensorについては、類似手法との網羅的な比較及び論文投稿まで、データの準備に関してはオーファン受容体のリストの作成までが進められた。今後は、これらを利用して、オーファン受容体ペアに結合する候補リガンドを列挙するシステムを構築していく予定である。

研究成果の学術的意義や社会的意義

Gタンパク質共役型受容体は、既存薬の多くが関与する、重要な創薬ターゲットであるが、リガンド未知のオーファン受容体が未だ多く存在する。本研究で得られる、オーファン受容体-リガンド-細胞機能の三項関係は、新規創薬ターゲットとして期待される。本研究で利用するアルゴリズム、非負値テンソル分解は、データを $A \times B \times C$ のように、三つ組の情報、テンソルとして表現し、その中に含まれる情報を抽出する。これにより、従来の行列分解ベースのアルゴリズムでは、「AとBの関係性」についてのパターンしか抽出できなかったのに対して、「AとBがCによって関係している」という、より高解像度なパターンの抽出が実現できる。

研究成果の概要（英文）：The basic technology of this study, scTensor, a method for detecting cell-cell interactions (CCI) and identifying receptor-ligand pairs in one-cell RNA-Seq data using non-negative tensor decomposition, was subjected to comprehensive comparison with similar methods and submission for publication. The data preparation was also completed to create a list of orphan receptors. In the future, we plan to use these methods to construct a system that enumerates candidate ligands that bind to orphan receptor pairs.

研究分野：バイオインフォマティクス

キーワード：バイオインフォマティクス 1細胞RNA-Seq 機械学習 オミックス 次元圧縮 テンソル分解 脱オーファン化 逆薬理学

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。

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

Gタンパク質共役型受容体は、既存薬の多くが関与する、重要な創薬ターゲットであるが、リガンド未知のオーファン受容体が未だ多く存在する。

本研究では、非負値テンソル分解というアルゴリズムを利用して、1細胞 RNA-Seq データに含まれる細胞間相互作用 (Cell-Cell Interaction: CCI) を網羅的に検出し、CCI に特異的に共発現するオーファン受容体とリガンドのペアを特定することで、オーファン受容体に結合するリガンド、及びそれらの結合が関与する細胞機能を同時に推定する。本研究で得られる、オーファン受容体-リガンド-細胞機能の三項関係は、新規創薬ターゲットとして期待される。

2. 研究の目的

薬は生体内の分子に作用し、その薬効を示す。その中でも Gタンパク質共役型受容体 (G protein-coupled receptor: GPCR) は、既存薬の多くが関与する、重要な創薬ターゲットであるが、リガンド未知のオーファン受容体が未だ多く存在する。

本研究では、非負値テンソル分解というアルゴリズムを利用して、1細胞 RNA-Seq データに含まれる細胞間相互作用 (Cell-Cell Interaction: CCI) を網羅的に検出し、CCI に特異的に共発現するオーファン受容体とリガンドのペアを特定することで、オーファン受容体に結合するリガンド、及びそれらの結合が関与する細胞機能を同時に推定する。本研究で得られる、オーファン受容体-リガンド-細胞機能の三項関係は、新規創薬ターゲットとして期待される。

3. 研究の方法

本研究で利用するアルゴリズム、非負値テンソル分解は、データを $A \times B \times C$ のように、三つ組の情報、テンソルとして表現し、その中に含まれる情報を少数のベクトルの外積として近似する。これにより、従来の行列分解ベースのアルゴリズムでは、「A と B の関係性」についてのパターンしか抽出できなかった場合と比較して、「A と B が C によって関係している」という、より高解像度なパターンを抽出が実現できる。非負値テンソル分解はすでに R のパッケージ nnTensor として実装してある。

4. 研究成果

本プロジェクトの基盤となっている scTensor については、CCI を検出する上での根拠となっている、既知の受容体-リガンドペアの適切なペアリング方法、現在従来の CCI 検出手法との比較、実データに対する有効性などを検証した論文 (雑誌論文 12 番目) として bioRxiv から公開しており、現在査読付き雑誌にて査読中である。オーファン GPCR のリスト、及びあらゆる組織・細胞における CCI の予測結果のデータベース (CellCellDB) は手元に作成してあるため、論文作業と引き続き、推定したオーファン GPCR の検証を進めていく予定である。

また、この研究テーマと関連して、様々な研究活動を進めることができたため、以下に記す。

【解析ツール論文】

scTensor の内部でテンソル分解を実行しているツール自体は、R パッケージ nnTensor として実装されており、本研究テーマで利用したデータに限らず、様々な問題に適用可能である。この nnTensor の実装自体に対するソフトウェア論文が Journal of Open Source Software に受理された (雑誌論文 9 番目)。

【ODEGR】

1細胞 RNA-Seq データ解析では、既知のゲノム配列上の遺伝子領域に由来するリード配列 (次世代シーケンサーが計測する RNA 分子量を反映した断片 DNA) の数を根拠として、遺伝子発現量が見積もられる。一方、遺伝子領域自体がまだ理解が進んでいない部分もあることから、本研究では既存の遺伝子領域を細かいピンで区切り、ピンごとの遺伝子発現量データに対して、非負値行列分解を適用することで、新しい遺伝子領域候補を発見することに成功した (雑誌論文 1 番目)。

【OnlinePCA.jl】

1細胞 RNA-Seq データでは、RNA × 細胞という行列形式で表現され、様々な次元圧縮法が適用される。しかしながら、近年の細胞方向の高出力化に伴い、データが計算機のメモリに載らない 100 万細胞レベルにまで達しており、計算自体ができないことが問題となっている。本研究では、そのような大規模データに対しても精度を落とすことなく適用できる主成分分析のアルゴリズムや実装を評価したベンチマーク実験であり、性能が良かった Randomized SVD を Out-of-core かつスパース形式で計算できるように実装した Julia パッケージも公開した (雑誌論文 2 番目)。

【WormTensor】

線虫の神経活動データに対して、テンソル分解を適用しており、複数個体間で共通するような機

能モジュールの検出を行った。各個体のクラスタリング結果を単純にマージするコンセンサスクラスタリングと比較して、この手法は各個体ごとに信頼度を意味する重みを推定する方法であり、これによりクラスタリング性能の向上や、低クオリティな個体のフィルタリングといった解析へ寄与することを示した。また、この解析手法を実装した R パッケージを WormTensor という名前で CRAN レポジトリから公開した（雑誌論文 10 番目）。

【COVID-19 への貢献】

新型コロナウイルスワクチンの副作用に関するアンケートデータにおいても、テンソル分解を適用しており、副反応の出方には大きく 4 パターンあることを突き詰めている（雑誌論文 7 番目）。この論文で実行した解析の手順は、TensorLyCV という名前のワークフローとして独自開発したもので、この実装自体も、プロトコール論文として公開した（雑誌論文 11 番目）。

【行列・テンソル分解の啓蒙活動】

行列・テンソル分解自体は、様々な問題に適用可能な汎用的な枠組みであることから、日本語の総説論文として、JSBi Bioinformatics Review で連載を進めており、現在までのところ、第 4 回まで公開されている（雑誌論文 3, 4, 6, 8 番目）。

5. 主な発表論文等

〔雑誌論文〕 計14件（うち査読付論文 13件 / うち国際共著 1件 / うちオープンアクセス 14件）

1. 著者名 露崎 弘毅	4. 巻 2
2. 論文標題 行列・テンソル分解によるヘテロバイオデータ統合解析の数理	5. 発行年 2021年
3. 雑誌名 JSBi Bioinformatics Review	6. 最初と最後の頁 15～29
掲載論文のDOI（デジタルオブジェクト識別子） 10.11234/jsbibr.2021.6	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 露崎 弘毅	4. 巻 1
2. 論文標題 行列・テンソル分解によるヘテロバイオデータ統合解析の数理	5. 発行年 2021年
3. 雑誌名 JSBi Bioinformatics Review	6. 最初と最後の頁 18～25
掲載論文のDOI（デジタルオブジェクト識別子） 10.11234/jsbibr.2021.1	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 Amorim Sabrina T., Tsuyuzaki Koki, Nikaido Itoshi, Morota Gota	4. 巻 53
2. 論文標題 Improved MeSH analysis software tools for farm animals	5. 発行年 2021年
3. 雑誌名 Animal Genetics	6. 最初と最後の頁 171～172
掲載論文のDOI（デジタルオブジェクト識別子） 10.1111/age.13159	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 露崎 弘毅	4. 巻 1
2. 論文標題 行列・テンソル分解によるヘテロバイオデータ統合解析の数理	5. 発行年 2021年
3. 雑誌名 JSBi Bioinformatics Review	6. 最初と最後の頁 18～25
掲載論文のDOI（デジタルオブジェクト識別子） 10.11234/jsbibr.2021.1	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 Koki Tsuyuzaki, Hiroyuki Sato, Kenta Sato and Itoshi Nikaido	4. 巻 21 (12)
2. 論文標題 OnlinePCA.jl: an ultrafast and memory-efficient incremental principal component analysis for million-level single cell RNA-sequencing	5. 発行年 2020年
3. 雑誌名 BMC Genome Biology	6. 最初と最後の頁 0-0
掲載論文のDOI (デジタルオブジェクト識別子) 10.1186/s13059-019-1900-3	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 Hirotaka Matsumoto, Tetsutaro Hayashi, Haruka Ozaki, Koki Tsuyuzaki, Mana Umeda, Tsuyoshi Iida, Masaya Nakamura, Hideyuki Okano, and Itoshi Nikaido	4. 巻 2 (1)
2. 論文標題 A NMF based approach to discover overlooked differentially expressed gene regions from single-cell RNA-seq	5. 発行年 2019年
3. 雑誌名 NAR Genomics and Bioinformatics	6. 最初と最後の頁 lqz020
掲載論文のDOI (デジタルオブジェクト識別子) 10.1093/nargab/lqz020	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 Kenta Sato, Koki Tsuyuzaki, Kentaro Shimizu and Itoshi Nikaido	4. 巻 20 (31)
2. 論文標題 CellFishing.jl: an ultrafast and scalable cell search method for single-cell RNA-sequencing	5. 発行年 2019年
3. 雑誌名 BMC Genome Biology	6. 最初と最後の頁 0-0
掲載論文のDOI (デジタルオブジェクト識別子) 10.1186/s13059-019-1639-x	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 露崎 弘毅	4. 巻 3
2. 論文標題 行列・テンソル分解によるヘテロバイオデータ統合解析の数理	5. 発行年 2022年
3. 雑誌名 JSBi Bioinformatics Review	6. 最初と最後の頁 20 ~ 33
掲載論文のDOI (デジタルオブジェクト識別子) 10.11234/jsbibr.2022.1	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 Ikeda Kei, Nakada Taka-Aki, Kageyama Takahiro, Tanaka Shigeru, Yoshida Naoki, Ishikawa Tetsuo, Goshima Yuki, Otaki Natsuko, Iwami Shingo, Shimamura Teppei, Taniguchi Toshibumi, Igari Hidetoshi, Hanaoka Hideki, Yokote Koutaro, Tsuyuzaki Koki, Nakajima Hiroshi, Kawakami Eiryō	4. 巻 25
2. 論文標題 Detecting time-evolving phenotypic components of adverse reactions against BNT162b2 SARS-CoV-2 vaccine via non-negative tensor factorization	5. 発行年 2022年
3. 雑誌名 iScience	6. 最初と最後の頁 105237 ~ 105237
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.isci.2022.105237	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 露崎 弘毅	4. 巻 3
2. 論文標題 行列・テンソル分解によるヘテロバイオデータ統合解析の数理	5. 発行年 2022年
3. 雑誌名 JSBi Bioinformatics Review	6. 最初と最後の頁 33 ~ 46
掲載論文のDOI (デジタルオブジェクト識別子) 10.11234/jsbibr.2022.2	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 Tsuyuzaki Koki, Nikaido Itoshi	4. 巻 8
2. 論文標題 nnTensor: An R package for non-negative matrix/tensor decomposition	5. 発行年 2023年
3. 雑誌名 Journal of Open Source Software	6. 最初と最後の頁 5015 ~ 5015
掲載論文のDOI (デジタルオブジェクト識別子) 10.21105/joss.05015	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 Tsuyuzaki Koki, Ishii Manabu, Nikaido Itoshi	4. 巻 0
2. 論文標題 scTensor detects many-to-many cell-cell interactions from single cell RNA-sequencing data	5. 発行年 2022年
3. 雑誌名 bioRxiv	6. 最初と最後の頁 0
掲載論文のDOI (デジタルオブジェクト識別子) 10.1101/2022.12.07.519225	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 -

1. 著者名 Koki Tsuyuzaki, Kentaro Yamamoto, Yu Toyoshima, Hirofumi Sato, Manami Kanamori, Takayuki Teramoto, Takeshi Ishihara, Yuichi Iino and Itoshi Nikaido	4. 巻 0
2. 論文標題 WormTensor: a clustering method for time-series whole-brain activity data from <i>C. elegans</i>	5. 発行年 2023年
3. 雑誌名 BMC Bioinformatics	6. 最初と最後の頁 0
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 Koki Tsuyuzaki, Naoki Yoshida, Tetsuo Ishikawa, Yuki Goshima, and Eiryō Kawakami	4. 巻 0
2. 論文標題 Non-negative tensor factorization workflow for time-series biomedical data	5. 発行年 2023年
3. 雑誌名 STAR Protocol	6. 最初と最後の頁 0
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 -

〔学会発表〕 計12件 (うち招待講演 10件 / うち国際学会 3件)

1. 発表者名 露崎弘毅
2. 発表標題 生命科学にパラダイムシフトを起こすテンソル分解技術
3. 学会等名 IP SJ-ONE (招待講演)
4. 発表年 2022年

1. 発表者名 露崎弘毅
2. 発表標題 多様な多次元データ構造を扱うテンソルパッケージの紹介
3. 学会等名 第44回日本分子生物学会年会 (招待講演)
4. 発表年 2021年

1. 発表者名 露崎弘毅
2. 発表標題 細胞間相互作用解析ワークショップ
3. 学会等名 BioC Asia 2021 (招待講演) (国際学会)
4. 発表年 2021年

1. 発表者名 露崎弘毅
2. 発表標題 Guidelines to handle large-scale and complex tensor data in R
3. 学会等名 BioC Asia 2021 (招待講演) (国際学会)
4. 発表年 2021年

1. 発表者名 露崎弘毅
2. 発表標題 Opening Remark
3. 学会等名 BioC Asia 2021 (招待講演) (国際学会)
4. 発表年 2021年

1. 発表者名 露崎弘毅
2. 発表標題 事例紹介: Snakemake
3. 学会等名 2021年日本バイオインフォマティクス学会年会・第10回生命医薬情報学連合大会 (IBMP2021) (招待講演)
4. 発表年 2021年

1. 発表者名 Koki Tsuyuzaki, Hiroyuki Sato, Kenta Sato, Itoshi Nikaido
2. 発表標題 Benchmarking principal component analysis for large-scale single-cell RNA-sequencing
3. 学会等名 IIBMP2020 2020年9月1日
4. 発表年 2020年

1. 発表者名 Koki Tsuyuzaki
2. 発表標題 Tensor Decomposition: A Versatile Method for Heterogeneous Biological Data Fusion
3. 学会等名 BioC Asia 2020 (招待講演)
4. 発表年 2020年

1. 発表者名 露崎弘毅
2. 発表標題 行列・テンソル分解を用いた 異種バイオデータ統合解析の数理
3. 学会等名 情報処理学会第64回バイオ情報学研究会/東北大学電気通信研究所第35回生体生命工学研究会 (招待講演)
4. 発表年 2020年

1. 発表者名 露崎弘毅
2. 発表標題 大規模オミックスデータのためのテンソル分解の高速化
3. 学会等名 IPR Seminar (招待講演)
4. 発表年 2021年

1. 発表者名 露崎弘毅
2. 発表標題 OSSを用いた1細胞オミックスデータ解析の現状と課題
3. 学会等名 第420回CBI学会講演会（招待講演）
4. 発表年 2021年

1. 発表者名 露崎弘毅
2. 発表標題 イントロダクション: テンソルとしてのバイオデータ
3. 学会等名 第42回分子生物学会年会
4. 発表年 2019年

〔図書〕 計2件

1. 著者名 露崎弘毅	4. 発行年 2019年
2. 出版社 羊土社	5. 総ページ数 255
3. 書名 RNA-Seqデータ解析 WETラボのための鉄板レシピ（実験医学別冊）, 「リファレンスゲノムにマッピングしない方法～salmon, kallisto, tximport & RNA-Seq定量にまつわるFAQ」	

1. 著者名 小野 浩雅	4. 発行年 2022年
2. 出版社 羊土社	5. 総ページ数 248
3. 書名 バイオDBとウェブツール ラボで使える最新70選	

〔産業財産権〕

〔その他〕

大規模データに対する主成分分析の性能を評価
https://www.riken.jp/press/2020/20200120_2/index.html
高速検索エンジン「CellFishing.jl」を開発
https://www.riken.jp/press/2019/20190225_1/index.html

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