

戦略的創造研究推進事業 CREST
研究領域「科学的発見・社会的課題解決に向けた
各分野のビッグデータ利活用推進のための次世代
アプリケーション技術の創出・高度化」
研究課題「データ駆動型解析による多細胞生物の
発生メカニズムの解明」

研究終了報告書

研究期間 2015年10月～2021年3月

研究代表者：大浪 修一
(理化学研究所生命機能科学研究センター
発生動態研究チーム、チームリーダー)

§ 1 研究実施の概要

(1) 実施概要

細胞や個体を生きたまま顕微鏡観察するライブイメージング技術と、ライブイメージング画像から細胞や個体の時空間動態を計測する画像処理技術の急速な発展により、現在の生命科学分野では、遺伝子不活化等の様々な条件下で細胞や個体の時空間動態を計測したデータの蓄積が急速に進んでいる。研究代表者の大浪は、発生過程で発現する様々な形態的特徴間の因果関係の推定手法や、これらの因果関係を生み出す分子メカニズムに関与する遺伝子の推定手法など、これらのデータからデータ駆動的に多細胞生物の発生メカニズムを解明する手法を世界に先駆けて開発してきた。このようなデータ駆動型の研究手法は、21世紀の生命科学の主要な推進力となると期待されている。本研究では、個体発生の時空間定量計測データが最も蓄積している線虫 *C. elegans* をモデル系として使用し、多細胞生物の発生メカニズムを解明するデータ駆動型の研究手法の開発と、それを利用した線虫の発生メカニズムの解明を目標とした。

上記の目標を達成するため、本研究では、システム生物学・計算生物学を専門とする大浪と、生物医学データの統計処理による因果推論を専門とする川上、ビッグデータの可視化情報解析技術の開発を専門とする小山田の3グループの共同研究体制を構築した。研究期間全般にわたり、各グループで要素技術を開発し、3グループ共同で、これら要素技術の統合を行うというアプローチを採用した。

本研究課題の最大の成果は、線虫胚発生の 64 細胞期までのデータ駆動モデルである。当モデルの構築の為に、大浪グループは、当研究課題の開始前に開発した 4 次元微分干渉顕微鏡画像の画像認識技術を適用し、胚発生に必須な全ての遺伝子 (350 遺伝子) に対する遺伝子ノックダウン胚の細胞核分裂動態の 4 次元計測データの多重複度コレクションを構築した。また、細胞核および細胞膜を蛍光標識した線虫胚の 4 次元高速共焦点顕微鏡画像から細胞核と細胞膜の動態を自動計測する装置を開発し、1-90 細胞期までの細胞核/細胞膜動態の 4 次元計測データの多重複度コレクションを構築した。大浪が以前より開発を進めてきた手法を川上グループと共同で改良し、これらの多重複度コレクションから 1-8 細胞期の細胞核動態に関する表現型発現因果関係ネットワーク、1-8 細胞期で推定された因果関係を生み出す遺伝子群、1-64 細胞期までの細胞核・細胞膜動態に関する表現型発現因果関係ネットワークの推定にそれぞれ成功した。川上グループは更に、異なるデータコレクションに由来する表現型発現因果関係ネットワークを統合する統計的手法の開発を行った。小山田グループは密グラフの粗視化方法、背景知識に基づくグラフのレイアウト方法、文献情報からの遺伝子・細胞間の因果関係推定手法を開発し、また、表現型発現因果関係ネットワークと遺伝子ネットワークをインタラクティブに表示し科学的発見を支援する可視化システムを開発した。これら各グループで開発した要素技術や推定結果を統合し、表現型発現因果関係ネットワークと遺伝子ネットワークを横断的に探索し、生物学における仮説構築を支援する可視化システム PheGeNet を開発した。PheGeNet には、当 CREST 研究で導出した 1-8 細胞期の細胞核動態に関する表現型発現因果関係ネットワーク、1-8 細胞期の細胞核動態表現型間で推定された因果関係を生み出す遺伝子群、1-64 細胞期までの細胞核・細胞膜動態に関する表現型発現因果関係ネットワークと、GeneMania より取得した最新の遺伝子ネットワークの統合ネットワークが表示可能で、グラフの粗視化率やレイアウトをインタラクティブに変更しながら知識の探索が可能である。また、線虫の統合データベース Wormbase より抽出した文献情報や、文献情報から推定した遺伝子間、細胞間の因果関係が同時に表示される。当 PheGeNet は実験生物学分野の共同研究者へのインターネット上での限定公開を既に開始している。今後、実地開催される国際的な学会での発表と同時に一般公開する予定である。

上記の PheGeNet に統合された成果以外にも、当 CREST 研究ではデータ駆動型の生命科学研究の未来を開拓する多くの成果の創出に成功した。生命動態定量データと生命科学画像データの統合データベース SSBD の構築は、近年のオープンサイエンス推進の世界的な潮流

に呼応し、生命科学画像データの公開と共有に関する国際連携プロジェクトへと発展した。大浪は中核メンバーの一人としてこの国際連携プロジェクトを牽引している。本研究で構築した胚発生に必須な全ての遺伝子に対する遺伝子ノックダウン胚の細胞核分裂4次元計測データの多重複度コレクションに多重検定やクラスタリング法を適用し、27,174種のRNAi表現型や18種の機能未知遺伝子の遺伝子機能の予測を行った。遺伝子ノックアウト胚の多重複度コレクションでのみ可能な本手法の成功は、今後の発生生物学分野の表現型解析の枠組みに大きな影響を与えるものと考えられる。本研究で開発したデータ駆動型的手法を遺伝子発現の時空間制御に適用するためのデータリソースの開発も行った。胚発生必須遺伝子130遺伝子に対して蛍光タンパク質標識した線虫株を作成し、これらの遺伝子発現の時空間パターンを4次元高速共焦点顕微鏡で記録した画像データの大規模なコレクションを作成した。当コレクションを本研究で開発したデータ駆動型手法に適用することにより、遺伝子発現の時空間発現動態と表現型発現動態の間の因果関係の推定が実現し、より予測精度の高いデータ駆動型発生モデルの構築が可能になると期待される。本研究で開発したデータ駆動型手法を線虫の一生のライフコースの研究に適用するための技術開発も行った。線虫の一生を組織レベルの高解像度でライブ撮影するマイクロデバイスを開発した。線虫は寿命が2-3週間と短く、老化研究が最も発達したモデル生物である。当デバイスを活用し、今後、老化研究にデータ駆動型の新たな研究を展開していきたいと考えている。本研究で開発したデータ駆動型的手法をマウス等の高等生物やオルガノイド等に展開するための技術開発も行った。本研究で開発したデータ駆動型的手法をマウス胚やオルガノイドに適用する上での最大の障壁はライブイメージングのz軸方向の解像度であった。そこで、本CREST研究の期間にわたり、マウス胚・オルガノイドに最適化した二軸の光シート顕微鏡システムの開発を行った。当システムの完成により、マウス胚・オルガノイドを用いたデータ駆動型解析の展開が期待できる。動的システムに潜む因果関係の推定手法として、本CREST研究の予算で2017年に招聘したUCSDのGeorge Sugihara教授とのEDM(Empirical Dynamic Modeling)の可視化システムの開発も行った。生体内の分子動態の解析や、細胞間シグナル伝達等の解析において、当システムの今後の活用が期待される。近年注目を集めている細胞核の4次元構造の生理機能の解明に関する開発も行った。次世代シーケンサーを使って取得した染色体の3次元構造に関するHi-Cデータから、4次元の染色体動態シミュレーションを可能にする理論とソフトウェアを開発した。当手法により、細胞核の構造と生理機能の解明が飛躍的に前進することが期待される。

以上のように、本CREST研究では、3グループの密接な協力の下、今後の生命科学研究に新たな可能性を提供するデータ駆動型研究に関する多くの要素技術、研究手法の開発とその実証が行われた。今後は、これらの成果を更に発展させ、基礎生物学の発展のみならず、がん治療や再生医療、生殖医療等を含めた生命医科学分野の発展に広く貢献していきたいと考えている。

(2) 顕著な成果

<優れた基礎研究としての成果>

1. 生命科学イメージングデータと生命動態定量データの統合データベースSSBDの構築

概要:

バイオイメージング技術と画像処理技術の発展により、様々な生命現象に関して、生命分子や細胞、組織などの時空間動態の大規模な計測が可能になったが、これらの計測データは研究室や雑誌のウェブサイトから個々に公開されており、第三者からのデータの再利用性が低かった。本研究では、国内外の生命動態の定量計測データと計測に利用した画像データ、および生命動態のシミュレーションデータを集積し共有する世界初の統合データベースを構築した。当データベースの構築は、今後の生命科学分野の画像データのオープンサイエンスのエコシステムの構築に大きな影響を与える先導的な研究成果であり、Nature誌に取材記事が掲載されるなど国際的に高い評価を得ている。

2. 表現型特徴と遺伝子ネットワークの横断的探索を可能にする Web ベース可視化システム (PheGeNet) の構築

概要:

本課題で初めて導出に成功した線虫の表現型特徴ネットワークと既存の遺伝子ネットワーク、および RNAi により対応付けた表現型特徴と遺伝子の関係性のデータをもとに、表現型特徴ネットワークと遺伝子ネットワークを横断的に探索し、生物学における仮説構築を支援する可視化システム PheGeNet を開発し、幅広く利用できるウェブサイトを構築した。表現型特徴と遺伝子の関係性を探ることは根源的なアプローチであるが、インタラクティブな可視化システムを通じてデータ駆動的に実現した初めての例である。また、線虫のデータベースである Wormbase の情報や論文情報をシステム内で結び付けることで、機能の関連が知られていなかった未知な遺伝子や生物学的プロセスへの推論を支援し、研究サイクルを加速させることが期待される。当システムは画像データを基盤とした生命科学分野のデータ駆動型研究に新しい方向性を与える先導的な研究成果である。

3. ゲノム3次元構造データを4次元高分子ダイナミクスに変換する PHi-C 法の開発

概要:

近年のゲノム解析技術の向上により、次世代シーケンサーを使って細胞核内の染色体3次元構造を解析する Hi-C 法が進展してきている。しかし、染色体領域間の近接頻度の情報である Hi-C データを染色体のダイナミクスに関連づける方法が存在せず、Hi-C データの解釈は進んでいなかった。本研究では、Hi-C データを4次元高分子ダイナミクスに変換する理論とシミュレーション手法である PHi-C 法を開発した。本手法により近年の生命科学分野において最も注目されているトピックの一つである、染色体ダイナミクスの機能と機構の解明が飛躍的に前進することが期待される。本手法については、既に多くの国際学会で招待講演の依頼を受けるなど、国際的な評価が高い。

< 科学技術イノベーションに大きく寄与する成果 >

1. バイオイメージング・データのデータ形式の標準化とデータ共有システムの整備に向けた国際提言

概要:

近年、超解像度顕微鏡や光シート顕微鏡など、新たなバイオイメージング技術の開発が加速している。しかし、これらの最新技術の多くは高度な光学技術・知識をもつ技術者の自作品であり、製品化された装置はきわめて高額である等の理由から、ごく限られた大学・研究所のみがこれらの最新技術を導入し活用しているのが実態であった。この問題の解決のため、近年、バイオイメージング・データの共有のしくみづくりが国際的な枠組みで進んでいる。本研究では、国際連携組織 Global BioImaging でのこれまでの議論をとりまとめ、バイオイメージング・データのデータ形式の標準化とデータ共有システムの整備についての提言を発表した。

2. 英文論文誌への投稿論文の採録・不採録の可視化分析

概要:

本研究は、英文論文誌 Journal of visualization (JOV) に提出された投稿 591 件の要約データを基に、要約データの採録・不採録の違いを探ることを目的に、各要約データの分散表現を計算した。多次元データとして表現される分散表現データを次元圧縮して、二次元平面に写像することにより、要約の構造に大きな違いがあることを示した。また、機械学習技術を用いた採録判定モデルの開発にも成功しているため、レビューアの負担を軽減し、レビュー品質を向上させる半自動レビューシステムの開発に先鞭をつけることができた。

3. 無染色画像データを用いた老化に伴う卵母細胞の品質低下の定量化

概要:

老化に伴い卵母細胞の品質は低下し、その結果として受精や受精後の胚の細胞分裂の失敗などが誘発される。本研究では、統計的画像処理法を用いて、線虫 *C. elegans* の老化に伴う胚の品質低下を定量化する手法を開発した。1 次統計量である Max-min 値および 2 次統計量である gray-level co-occurrence matrix の Correlation 値を用いると、老化に伴う卵母細胞の変化を無染色の画像から有意に分類できることが示された。316 種の胚発生必須遺伝子に対して遺伝子ノックダウン処理をした 1728 セットの線虫胚の画像に対して、当画像処理システムを適用することにより、卵母細胞の老化メカニズムに関連することが予想される 6 種の候補遺伝子が導出された。今後の本手法のヒト卵母細胞への応用発展が期待される。

< 代表的な論文 >

1. Tohsato, Y., Ho, K. H. L., Kyoda K., and Onami, S. "SSBD: a database of quantitative data of spatiotemporal dynamics of biological phenomena" *Bioinformatics*, vol. 32, no. 22, pp.3471-3479, 2016.

概要:

バイオイメージング技術と画像処理技術の発展により、様々な生命現象に関して、生命分子や細胞、組織などの時空間動態の大規模な計測が可能になったが、これらの計測データは研究室や雑誌のウェブサイトから個々に公開されており、第三者からのデータの利用率が低かった。本研究では、国内外の生命動態の定量計測データと計測に利用した画像データ、および生命動態のシミュレーションデータを集積し共有する世界初の統合データベースを構築した。

2. Onoue, Y., Kukimoto, N., Sakamoto, N., and Koyamada, K. "Minimizing the number of edges via edge concentration in dense layered graphs" *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 6, pp.1652-1661, 2016.

概要:

表現型特徴因果関係ネットワークは辺数が多い密グラフであるため、これを階層グラフとして可視化した場合に多くの辺交差が現れる。辺交差の発生はグラフの視認性に悪影響を与えることが知られており、これを解消することが課題であった。本論文では、辺交差を削減するために新たな辺集中化アルゴリズムを開発し、表現型特徴因果関係ネットワークへの適用を行った。その結果、53 パーセント以上の辺交差が削減され、大幅な視認性の向上が達成された。

3. Swedlow, J. R., Kankaanpää, P., Sarkans, U., Goscinski, W., Galloway, G., Malacrida, L., Sullivan, R. P., Härtel, S., Brown, C. M., Wood, C., Keppler, A., Paina, F., Loos, B., Zullino, S., Longo, D. L., Aime, S., and Onami, S. "A global view of standards for open image data format and repositories", *Nature Methods*, 2021. <https://doi.org/10.1038/s41592-021-01113-7>

概要:

近年、超解像度顕微鏡や光シート顕微鏡など、新たなバイオイメージング技術の開発が加速しているが、ごく限られた大学・研究所のみがこれらの最新技術を導入し活用しているのが実態である。この問題の解決のため、近年、バイオイメージング・データの共有のしくみづくりが国際的な枠組みで進んでいる。本研究では、国際連携組織 Global BioImaging でのこれまでの議論をとりまとめ、バイオイメージング・データのデータ形式の標準化とデータ共有システムの整備についての提言を発表した。

§ 2 研究実施体制

(1) 研究チームの体制について

(1) 研究代表者グループ

① 研究代表者: 大浪 修一 (理化学研究所生命機能科学研究センター、チームリーダー)

② 研究項目

- ・ 生命動態データとオミクス計測データを統合して多細胞生物の発生メカニズムを解明するデータ駆動型の研究手法の開発と実証
- 開発済みの手法の改良
- 最新データ及び改良手法を用いた発生メカニズムの推定
- 推定した発生メカニズムの検証
- 発生メカニズムを解明するデータ駆動型の新規手法の開発

(2) 共同研究グループ(1)

① 主たる共同研究者: 川上 浩司 (京都大学大学院医学研究科、教授)

② 研究項目

- ・ 計算表現型解析における新規統計手法の開発と実装
- 表現型間の因果推論のための統計手法の開発
- 因果推論により推定された発生モデルの適合性指標の開発
- 発生メカニズムを解明する新規統計手法の開発
- 新規統計手法により推定されたモデルの適合性指標の開発

(2) 共同研究グループ(2)

① 主たる共同研究者: 小山田 耕二 (京都大学学術情報メディアセンター、教授)

② 研究項目

- ・ 基礎生命科学の発見を促進するビッグデータ可視化技術の開発
- 生命科学者のニーズ分析
- 大規模グラフの粗視化技術の開発
- 潜在変数探索支援技術の開発
- 時系列複数密度データからの因果発見支援技術の開発

(2) 国内外の研究者や産業界等との連携によるネットワーク形成の状況について

① 大浪グループ

理化学研究所生命機能科学研究センターの渡邊朋信チームリーダーと連携して、当 CREST 課題で線虫を対象に開発している技術を、マウス胚やオルガノイドに適用するために必須の 4 次元顕微鏡撮影装置の開発を行っている。

理化学研究所生命機能科学研究センターの清末優子チームリーダーと神戸大学大学院システム情報学研究科助教/情報通信研究機構脳情報通信融合研究センター特別研究員の松原索博士と米国 HHMI Janelia Research Campus の Eric Betzig 博士が開発した Lattice Light Sheet 顕微鏡の細胞の 4 次元動態データからの細胞情報抽出の共同研究を行っている。清末チームリーダーは本連携研究を含む研究課題で平成 30 年度 CREST「情報計測」領域に採択された。(当課題に大浪は研究グループのメンバーの一人として参加している)

理化学研究所バイオリソース研究センターの榎屋啓志室長、同生命医科学研究センターの粕川雄也チームリーダー、同情報システム本部小林紀郎開発ユニットリーダー、同環境資源科学研究センターの有田正規チームリーダー、同脳神経科学研究センターの田中啓治チームリーダー、同光量子工学研究センターの横田秀夫チームリーダーらと、理化学研究所の

生命科学系の研究者が生産するデータを統合するプラットフォームを構築し、データ駆動型研究を展開するオープンライフサイエンスプロジェクトを行っている（大浪はプロジェクトの代表者）。当プロジェクトは令和元年度に理化学研究所のオープンサイエンス事業の一部として資金獲得に成功し、令和2年9月には当プロジェクトを推進するための生命科学データ共有開発ユニットが同研究所の情報システム本部に設置された（大浪は同開発ユニットの開発ユニットリーダーに就任（兼任））。

大阪大学の永井健治教授、徳島大学の堀川一樹教授、東京大学の岡崎拓教授らと、バイオイメージインフォマティクスを活用したビッグデータ解析による細胞集団の状態変化に関わる共同研究を行っている。当共同研究は、平成30年度に新学術領域研究「シンギュラリティ生物学」として採択された（永井教授が領域代表）。大浪は同領域研究の計画班の研究代表を務めている。

東北大学の杉本亜砂子教授と、当 CREST プロジェクトの研究成果である遺伝子ネットワークと表現型発現ネットワークの情報を活用した実験発生生物学研究の共同研究を開始している。

大阪大学の佐々木洋教授と、当 CREST プロジェクトの研究成果である細胞の画像認識および4次元追跡技術のマウス胚発生への応用に関する共同研究を行っている。

国立遺伝学研究所の前島一博教授と、当 CREST プロジェクトの研究成果である PHi-C 法を用いた細胞核の4次元構造とその生理的機能の解明に関する共同研究を行っている。

がん研究所の八尾良司部長と、当 CREST 研究で開発したデータ駆動型の解析手法を、ヒト大腸がんオルガノイドを用いたがん研究に応用することを目指す共同研究を開始した。

CREST「多細胞」の松田道行総括より、同 CREST プロジェクトの画像データを中心とした研究データの管理とオープン化の支援を依頼され、CREST「多細胞」領域研究者全体との連携を進めている。大浪は、上記の役割を遂行するため、理化学研究所生命機能科学研究センターの藤原裕展チームリーダーが研究代表を務める研究課題に主たる共同研究者の立場で参画することとなった。

The University of British Columbia の杉岡賢史助教授（Assistant Professor）と、当 CREST プロジェクトの研究成果である遺伝子ネットワークと表現型発現ネットワークの情報を活用した実験発生生物学研究の共同研究を開始している。

OMERO および IDR の開発中心者である、英国 University of Dundee の Jason Swedlow 教授と画像データベースのデータ共有技術の開発について連携している。また、OMERO システムを基盤とする生命科学画像データベースの世界連携機構の構築を進めている。

欧州の Euro-BioImaging が中心となり構築した、全世界の最先端イメージング関連技術の共有化の連携組織である、Global BioImaging に、生物画像の公共データベースの日本代表として参画している。また、University of Dundee の Jason Swedlow 教授とともに、同組織の Image Data Management Working Group の座長を勤め、当 Working Group の議論を牽引している。

独 University of Frelburg の Roland Nitschke 博士を中心に光学顕微鏡の機器とイメージングの品質と再現性の向上のために世界規模で組織された QUAREP (Quality Assessment and Reproducibility for Instruments & Images in Light Microscopy) に招待され、世界各国の専門家と共に、顕微鏡画像撮影の品質の向上のために協働作業を行っている。

仙台ソレイユ母子クリニック（村川晴生院長）と、DIC 画像の画像認識を用いた受精卵の品質判定に関する共同研究を行っている。

顕微鏡画像データの解析フローのオープン化に対応するため、株式会社ニコンと ROI データのデータフォーマットに関する共同研究を行っている。2020 年には株式会社ニコンの顕微鏡画像解析ソフトウェア NIS-Elements から、同ソフトウェアを用いて取得した ROI データを我々が開発した BDML フォーマットで出力することを可能にするソフトウェアアップデートが行われた。

②川上グループ

滋賀大学データサイエンス学部教授、理化学研究所革新知能統合研究センター汎用基盤技術研究グループ因果推論チームチームリーダーの清水昌平氏と線虫胚発生時の特徴量データに対する LiNGAM の適応について研究打合せを行い、LiNGAM 法の活用について継続してアドバイスを頂いている。

③小山田グループ

「ビッグデータ統合利活用のための次世代基盤技術の創出・体系化」領域、研究代表者黒橋禎夫グループと生命科学論文ビッグデータからの情報抽出とその利用を目指した協業を行っている。協業の成果は以下のとおりである。

1. WormBase に収録されている線虫に関する論文のうち、19888 件のアブストラクト掲載ページにアクセスし、題名・著者・掲載誌・説明文等の項目別にテキストを取得、定型化されたリストを自動的に生成するプログラムを開発した。
2. 1 で取得したテキストデータについて、自然言語処理技術および機械学習を用い、記載されている細胞名別にグループ分けを行った。
3. 2 で抽出した語に加え、認知心理学の研究で用いられる因果関係を示す動詞リスト、およびドメインエキスパートから提供された表現型特徴リストをもとに、該当する記述を含むセンテンスを検索、構文解析結果をもとに、因果表現の候補となる主語-動詞-目的語のセットのリスト 295 件を抽出するプログラムを開発した。
4. 実験室で得られたウェットデータをもとに作成された表現型ネットワーク図に、3 で得られた因果関係情報を統合、図中のアークを指定すると当該テキストが出力される、GUI インターフェイスのプロトタイプを開発した。

時系列複数密度データからの因果発見支援技術の開発に際して、時系列因果検出において顕著な業績のある UC San Diego の George Sugihara 教授らと共同研究を行い、動的ネットワークからシステムの状態を特定しその遷移を可視化する視覚的分析システムを開発した。この成果は国際共著論文として IEEE TVCG に採録が決定した。また、時系列因果検出においてソーク研究所の Gerald Pao 博士、Boston 大学の Ethan Deyle 博士 (R1 年度まで UCSD 所属)、SFNRC (南フロリダ自然資源センター) の Joseph Park 教授とも密に連携し研究を進めている。

「ビッグデータ統合利活用のための次世代基盤技術の創出・体系化」領域、研究代表者三好建正グループ前島研究員と気候シミュレーション結果と河川水位データとの因果関係を探索するための可視化技術の研究開発に着手した。具体的には、京コンピュータでの計算結果と神戸市都賀川水位データを使った因果ボリュームデータの作成とその可視化を行っている。

本グループ研究に関係する生命科学論文 DB から知見を得るための視覚的分析技術を開発するにあたって、CREST ビッグデータ応用「構造理解に基づく大規模文献情報からの知識発見」領域、研究代表者の松本裕治先生と意見交換を行った。

新型コロナウイルス感染に関して、無症状者の喫煙者からの伝播にあたって、呼出する煙で遠くにウイルス粒子が運ばれるかどうかについて、CREST ビッグデータ応用「大規模生物情報を活用したパンデミックの予兆、予測と流行対策策定」領域、研究代表者の西浦博先生と意見交換を行った。