

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
2015年度採択研究代表者

2019年度 実績報告書

大浪 修一

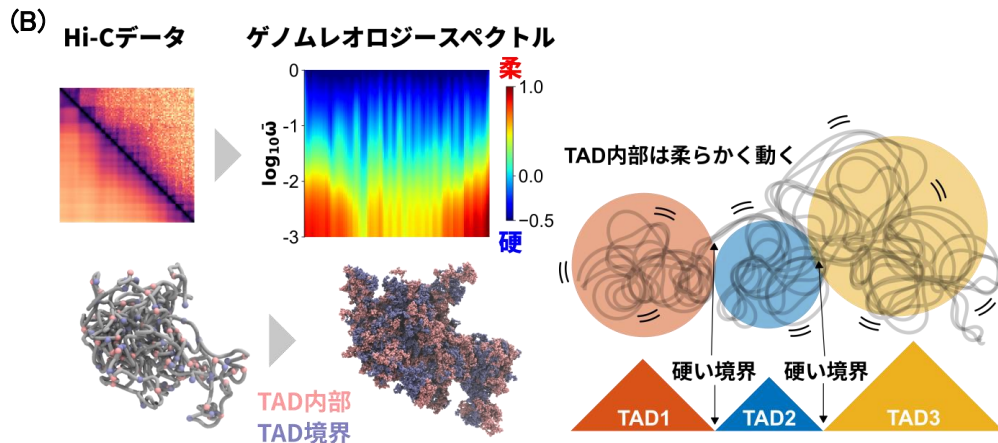
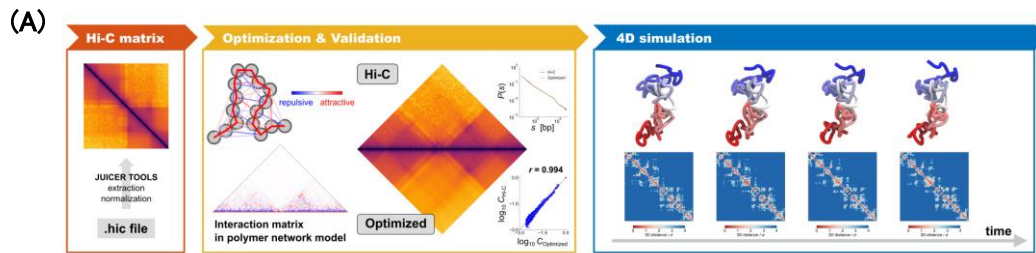
理化学研究所生命機能科学研究センター
チームリーダー

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

§ 1. 研究成果の概要

本研究は、因果推論を用いたビッグデータの解析(川上グループ)とビッグデータ解析における可視化技術開発(小山田グループ)にそれぞれ高い実績を持つ2つの研究グループと連携し、大浪グループがこれまで進めてきた、生命動態の時空間定量計測データを中核データとして使用して多細胞生物の発生のメカニズムを解明するデータ駆動型の研究手法の開発を強力に推進することを目的とする。第5年次である令和元年度は、前年に引き続き、3グループが共同して、計算表現型解析、クラスタ解析、表現型特徴の因果推論の手法の改良と、これらの解析結果からの知識発見を促進する可視化手法の開発を行った。加えて、大浪グループは発生のデータ駆動解析に適用する新たな大規模データの取得と先端バイオイメージデータの世界規模での共有化に向けての取り組みを、小山田グループは論文情報から因果関係を抽出するプログラム・ツールの開発を前年に引き続き行った。また、大浪グループはゲノム3次元構造データを用いて染色体の4次元ダイナミクスの特性を解析する新たな手法の開発を行った。

具体的には、大浪グループは、GO enrichment 解析の方法を改良し、胚発生必須遺伝子(263種)を分類した24クラスタ各々に対応する生物学的プロセスを推定し、これを用いて22種の未解明遺伝子の機能を推定した。また、これまで開発してきた表現型特徴発現の因果関係ネットワークを推定する手法を、前年度までに取得した線虫胚の1細胞期から90細胞期までの細胞核および細胞膜の動態の計測データ(50個体分)に適用し、50細胞期までの表現型特徴発現因果関係ネットワークを推定した。また、昨年度に引き続き、ゲノム編集を用いた蛍光タンパク質標識株の作成を行い、通算で127遺伝子に対する蛍光標識株を樹立した。更に、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法を開発し、染色体機能構造ドメインに関する新しいモデルを提唱した。加えて、生命動態データとバイオイメージデータの統合データベースSSBD:databaseを更新し、3076種の画像データと201種の生命動態定量データを新たに公開した。加えて、Global BioImaging および BioImage Archive の会議に参加し(いずれも招待)、先端生命科学画像データベースの世界連携を進めた。川上グループは、これまでに抽出された特徴量間の因果推論手法の問題点を検討し LiNGAM をはじめとした統計的因果推論手法についての検討と実装を行った。大浪グループから線虫胚の細胞核の4次元分裂動態の計測データの提供を受け、LiNGAM による統計因果推論およびバリデーションを行った。また、異なる特徴量データ間を連結し、多次元の相関・因果推論結果を提示する手法を開発した。小山田グループは表現型ネットワークに含まれるグループ構造を利用したネットワーク可視化を高度化し、加えて、表現型—遺伝子ネットワーク探索のための視覚的分析システムのプロトタイプを構築し、チーム内外のドメインエキスパートが利用可能な環境構築の準備を進めた。また、大規模化する表現型特徴データや遺伝子データを効率よく可視化および分析するために大規模グラフ粗視化技術を活用した可視化技術の性能のユーザー実験による評価を行い、その成果を国際会議および学術論文にて公表した。さらに、学術テキストデータからの因果関係抽出については、文書の真偽判定を向上させるとともに、多忙な専門家が短時間でテキストを読み作業が可能な、操作性を備えた文書分類のためのアノテーションツールの開発を行った。



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 法の概要(A)と、PHi-C 法を拡張したゲノムレオロジー解析により提唱された染色体機能構造ドメインに関する新しいモデル(B)

【代表的な原著論文】

Soya Shinkai, Masaki Nakagawa, Takeshi Sugawara, Yuichi Togashi, Hiroshi Ochiai, Ryuichiro Nakato, Yuichi Taniguchi, and Shuichi Onami, “PHi-C: deciphering Hi-C data into polymer dynamics”, NAR Genomics and Bioinformatics, vol. 2, No. 2, lqaa020, 2020

Soya Shinkai, Takeshi Sugawara, Hisashi Miutra, Ichiro Hiratani, and Shuichi Onami, “Michrorheology for Hi-C data reveals the spectrum of the dynamic 3D genome organization”, Biophysical Journal, vol. 118, No. 9, pp.2220–2228, 2020

Nozomi Aoyama, Yosuke Onoue, Yuki Ueno, Hiroaki Natsukawa, and Koji Koyamada. “User Evaluation of Group-in-a-box Variants”, Proceedings of IEEE Pacific Visualization Symposium (PacificVis), pp.127–136, 2019

§ 2. 研究実施体制

(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