

ゲノム情報工学講座

～様々な生命科学データの情報処理



研究室紹介

瀬尾 茂人*

Genome Information Engineering Laboratory:
Information processing of multi-modal life science data
Key Words: Bioinformatics, Machine Learning, Image Processing

はじめに

生命科学の進展により、単一細胞レベルという高解像度の遺伝子発現データや、細胞・組織の画像・動画データなど、多様で大規模な生命データが日常的に得られるようになってきています。これらのデータは、生命現象を深く理解するための豊かな情報を含んでいる一方で、その規模と複雑さゆえに、人の直感だけで十分に読み解くことは難しくなっています。このような背景のもと、情報科学の手法を用いて生命データを解析し、新たな知見を引き出すことの重要性はますます高まっています。

私たちの研究室では、バイオインフォマティクスおよびバイオイメージングインフォマティクスを基盤として、生命科学における多様なデータ解析に取り組んできました。遺伝子発現データから細胞状態や制御機構を探る研究と、画像や動画から細胞や組織の形態・動態を定量化する研究は、一見すると異なる対象を扱っているように見えます。しかし実際には、いずれも大量で複雑なデータから意味のある構造を抽出するという共通の課題を持っています。

筆者は2026年8月に、ゲノム情報工学講座¹⁾を引き継ぐ形で教授に着任しました。前任の松田秀雄教授は、筆者の学生時代の指導教員であり、その後も四半世紀にわたりご教示をいただいできました。松田教授のもとで培われた「生命現象をデータから“読み解く”」という視点は、本講座の重要な基盤であり、

現在の研究にも受け継がれています。本稿では、その流れを踏まえつつ、これまでの研究の展開と、今後のマルチモーダルなデータ解析への展望について紹介します。

発現解析と画像処理

生命科学におけるデータ解析は、近年ますます機械学習との結びつきを強めています。遺伝子発現データや画像データに代表される生命科学データは、高次元でノイズが多く、かつ複雑な構造を持つため、人手による解析には限界があります。このようなデータから有意なパターンや構造を抽出するには、統計的手法や機械学習に基づくモデル化が不可欠です。特に、次元削減やクラスタリング、分類といった基本的手法に加え、近年では深層学習などの高度な手法の導入により、従来は困難であった複雑な生命現象の解析が可能になってきています。

このような背景のもと、シングルセルRNAシーケンス (scRNA-seq) データを用いた遺伝子制御ネットワークの推定は重要な研究課題となっています。シングルセルデータでは、個々の細胞の遺伝子発現プロファイルが得られるため、細胞間の不均一性や状態遷移を詳細に解析することができます。しかし、観測されるデータは静的なスナップショットであり、時間的変化は直接観測されません。この課題に対して、細胞間の発現パターンの類似性に基づいて疑似時間を推定し、細胞変化を仮想的な時系列として再構成する手法が提案されています。さらに、この時系列データに深層学習モデルを適用することで、遺伝子間の因果関係や制御ネットワークの推定²⁾が可能となります。

また筆者らは、シングルセルデータ解析における前処理段階、特に遺伝子発現量の定量化手法の違いが解析結果に大きな影響を与える点に着目し、複数



* Shigeto SENO

1979年2月生まれ
大阪大学大学院 情報科学研究科 博士後期課程 修了 (博士 (情報科学)、2006年)
現在、大阪大学大学院 情報科学研究科 バイオ情報工学専攻 教授
TEL : 06-6879-4390
E-mail: senoo@ist.osaka-u.ac.jp
E-mail: doi@civil.eng.osaka-u.ac.jp

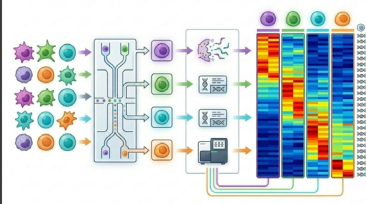
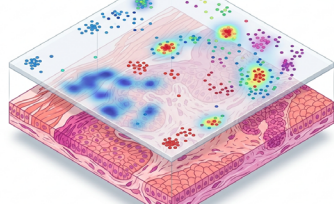
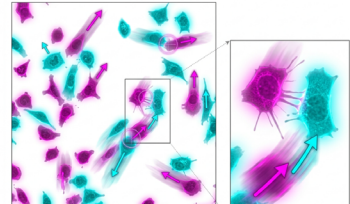
| | シングルセルオミクス | 空間オミクス | ライブイメージング |
|-----|---|--|---|
| | <p>シングルセル（単一細胞）を単位として測定されたオミクスデータ</p>  <p>各細胞での遺伝子発現量を示す「遺伝子×細胞」のテーブルデータ</p> | <p>空間情報を保持したまま測定されたオミクスデータ</p>  <p>様々な種類の分子の位置情報と、組織の画像データのセット</p> | <p>高解像度・長時間観察された生体内の細胞動画像</p>  <p>マルチチャネルの動画像（生きた細胞の機能や動態）</p> |
| 遺伝子 | ○ ほぼ全ての遺伝子 | △ 数百種類 | × 数種類 |
| 細胞 | ○ 数万個 | ◎ 数万個～ | △ 数十個 |
| 空間 | × 位置情報は失われる | ○ 広い（5mm～） | △ 狭い（数百μm～） |
| 時間 | △ 数点の経時観察 | × ほぼ1点（高コスト） | ◎ 連続した動態 |

図1 マルチモーダルな生命科学データとその性質

の定量化手法から得られる発現プロファイルを統合する解析手法を提案しています³⁾。この手法では、異なる定量化方法に由来するデータをジョイント非負値行列分解 (joint-NMF) により同時に分解し、各データに固有の成分と、細胞に共通する潜在表現を分離して抽出します。これにより、定量化手法の違いに起因するばらつきを抑えつつ、より本質的な細胞構造を反映したクラスタリングや特徴抽出を行うことが可能となります。さらに、この枠組みにより、細胞集団に特徴的な遺伝子群の同定も可能であることを示しており、シングルセル解析におけるロバストな情報抽出手法として有効であると考えています。

一方、画像データ解析においても、機械学習の導入により大きな進展が見られます。特に組織画像解析では、細胞のセグメンテーションや分類を通じて、組織状態を定量的に評価することが求められます。近年では、深層学習と画像処理を組み合わせることで、筋組織の回復過程を定量的に評価する手法を開発しました。この手法では、ヘマトキシリン・エオジン染色画像から細胞領域を抽出し、それらを分類することで回復段階を推定します。特に、ラベル比率学習などの弱教師あり学習を用いることで、詳細なアノテーションを必要とせず高精度な解析を実現している点が特徴⁴⁾です。

さらに、画像解析は静止画像にとどまらず、ライブイメージングによる動画データへと応用しています。細胞の移動や相互作用といった動的現象を捉え

るためには、細胞の検出・追跡・状態識別を統合的に行う必要があります。これにより、免疫細胞の遊走ダイナミクスのような複雑な生体内挙動を定量化し、その背後にある生物学的メカニズムの理解へとつなげることが可能となります。

このように、発現解析と画像処理は、いずれも機械学習に基づくデータ駆動型アプローチによって発展してきた分野であり、異なるデータ形式を扱いながらも共通の方法論に支えられています。これらを統合的に発展させることが、生命現象のより深い理解に向けた重要な鍵となります。

マルチモーダルな生命科学データの解析

今後、生命データ解析においてますます重要になるのは、異なる種類のデータを統合して扱うマルチモーダル解析です。近年では、シングルセルRNAシーケンスにより細胞ごとの遺伝子発現を網羅的に測定できるようになり、さらに空間トランスクリプトーム技術によって、組織内での分子の位置情報を保持したまま遺伝子発現を観測することが可能となってきました。また、ライブイメージング技術の進展により、生体内で細胞がどのように移動し、相互作用するかを時間的に連続して観察できるようになっています(図1)。

これらのデータは、それぞれ異なる特徴と制約を持っています。シングルセルオミクスは網羅性が高い一方で空間情報を失いやすく、ライブイメージン

グは動態を捉えられるものの観測できる分子数は限られています。その中間に位置する空間オミクスは、分子情報と空間情報を同時に扱うことができますが、時間情報は乏しいという特徴があります。このように、各モダリティは互いに補完的な関係にあります。

しかし、細胞の分子状態だけを見ても、その細胞がどのように振る舞うかを直接理解することは難しいです。逆に、画像から細胞の動態を観察できても、その背後にある分子機構を明らかにすることは容易ではありません。したがって、これらのデータを統合し、細胞の「状態(分子)」と「振る舞い(動態)」、さらには「位置(空間)」を同時に理解する枠組みが必要となります。

私たちの研究室では、これまでに蓄積してきた発現解析と画像解析の手法を基盤として、こうしたマルチモーダル統合へと研究を展開しています。異種データ間の対応付け、時空間情報の統合、さらには解釈可能な数理モデルの構築といった課題に取り組むことで、細胞間相互作用や組織レベルでの秩序形成といった現象の理解を目指しています。このようなアプローチは、生命現象を静的な断面としてではなく、時空間的に変化するシステムとして捉えるための基盤となるものです。

おわりに

これまで進めてきた発現解析と画像処理の研究は、それぞれ独立したものではなく、生命データをどのように理解するかという共通の問いに支えられています。そしてその蓄積は、異なる種類のデータを統合して生命現象を捉えるマルチモーダル解析へと自然につながっていきます。情報科学の手法を用いて生命科学データを横断的に扱うことは、今後の生命科学や医療研究において、ますます重要になると考えられます。

筆者が愛読してきたSF作品『月は無慈悲な夜の女

王』⁵⁾に描かれた、人が人工知能の相談相手となる世界は、かつては遠い未来像のように思われました。しかし現在では、AIエージェントと対話しながらデータ解析を進めることが現実の研究環境となりつつあり、その有用性と新しさを日々実感しています。一方で、こうした技術の進展が大学や研究活動の在り方をどのように変えていくのかについては、慎重に考えるべき側面もあります。

今後も、新しい技術を積極的に取り入れつつ、生命現象の理解に本質的に迫る研究を進めていきたいと考えています。そのためには、データの統合だけでなく、それを解釈するための理論やモデルの深化が不可欠であり、情報科学と生命科学のさらなる融合が求められています。

参考文献

- 1) 大阪大学 大学院情報科学研究科 ゲノム情報工学講座 <https://www-gie.ist.osaka-u.ac.jp/>
- 2) Yao S, Uemura K, Seno S, Matsuda H, “A Novel Method for Gene Regulatory Network Inference with Pseudotime Data Using Information Criterion”, *International Journal of Bioscience, Biochemistry and Bioinformatics*, 12(3), pp.43-52 (2022.6)
- 3) Shiga M, Seno S, Onizuka M, Matsuda H, “SC-JNMF: Single-cell clustering integrating multiple quantification methods based on joint non-negative matrix factorization”, *PeerJ*, 9:e12087 (2021.9).
- 4) Yamaoka Y, Seno S, Chan WY, Iwamori K, Fukada S, Matsuda H, “Quantifying the recovery process of skeletal muscle on hematoxylin and eosin stained images via learning from label proportion”, *Scientific Reports*, Vol. 14, 27044, (2024).
- 5) ロバート・A・ハインライン, 矢野徹(訳), “月は無慈悲な夜の女王”, 早川書房 (1976).