

バイオイメージ・インフォマティクスと可視化技術を統合した ユーザー主導型データ探索*

夏川 浩明**, 尾上 洋介***, 小山田 耕二****

User-driven Data Exploration Integrating Bioimage Informatics and Visualization Technology

Hiroaki NATSUKAWA, Yosuke ONOUE and Koji KOYAMADA

1. 緒 論

生命は生化学現象を基盤とし、遺伝子や mRNA、タンパク質、表現型特徴などの異なるスケールで制御される複雑系からなる。この生命現象を理解するためには、現象のダイナミクスや異なるスケール間の相互作用を具に観察する必要がある。ここ数十年で顕微鏡により計測された細胞の時空間動態を自動的に計測するバイオイメージ・インフォマティクス技術が発達し、大量の生命科学データが取得されている。また、ImageJ¹⁾に代表されるように取得された画像データを定量化し可視化する汎用的なツールも盛んに開発されている²⁾³⁾⁴⁾。増え続ける生物学データを有効に利用し、新たな生物学的知見を目指すデータ駆動型科学の存在が重要性を増す中、それを支援するためには、異種のデータを如何に見るかという点を考える必要がある。ここでデータ駆動型科学とは、事前の仮説無しにデータから何が言えるのかを考え、科学的手法における仮説生成をデータに基づき行うアプローチのことを指す。生物学データとその解析に可視化技術を統合させるアプローチが、データ駆動型科学の基礎となるユーザー主導型のデータ探索を促進させると著者らは考えている。

本稿ではバイオイメージ・インフォマティクスにより取得された線虫 (C. elegans) の発生ダイナミクスのデータに対して、可視化技術によりデータ駆動型アプローチの促進を目指した我々の取り組みを紹介する。次節では、バイオイメージ・インフォマティクスの方法で定量化された表現型特徴ネットワークから、生物学的な因果関係を示唆する関係性抽出を目指した可視化システム Causal Net⁵⁾について解説する。続けて、遺伝子と表現型という異なる生物学ネットワークの探索を支援す

る可視化システム PheGeNet の取り組み⁶⁾を紹介する。最後に生物の発生の根源になる卵母細胞のカルシウム波現象に注目した視覚的分析⁷⁾について紹介した後に、バイオイメージ・インフォマティクスと可視化についての我々の取り組みを総括する。

2. 表現型特徴ネットワーク探索のための統合的可視化システム

動画画像に画像認識技術を適用して生物特徴の動態を計測する技術であるバイオイメージ・インフォマティクスにより高スループットに生命動態を定量化できるようになり、大規模なデータベースが構築されている⁸⁾。本研究ではモデル系として研究の進んだ多細胞生物である線虫 C. elegans を対象とし、発生中の正常な胚の細胞の動態を 4 次元的に計測した顕微鏡画像を取得した。取得した動画画像データに対して画像認識プログラムを適用し、核の大きさ等の胚の表現型特徴の動態を計測する。次に計測された表現型特徴間の相関を計算することで、表現型特徴間の関係性から Fig. 1 のように表現型特徴ネットワークが構築される。ここでノードは表現型特徴を表し、リンクは表現型特徴相関を表す。しかしながら表現型特徴ネットワークは密なネットワークになるため、辺交差が多く特徴点を視認して抽出するのは困難である。さらに、データ探索はユーザーの関心事や専門知識を基に行われるため、ウェットな実験データと相補関係にあ

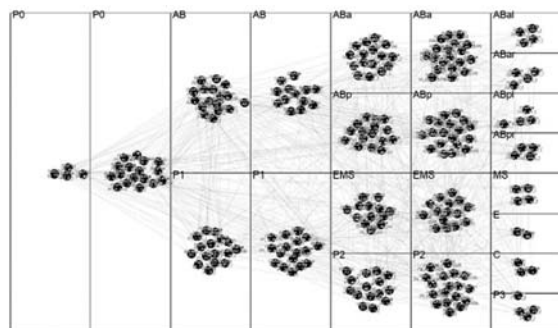


Fig. 1 Example of phenotypic character network.

* 原稿受付 2019 年 10 月 30 日

** 正会員 京都大学 学術情報メディアセンター
(〒 606-8501 京都市左京区吉田本町,
E-mail: natsukawa.hiroaki.3u@kyoto-u.ac.jp)

*** 正会員 日本大学 文理学部 情報科学科

**** 正会員 京都大学 学術情報メディアセンター

る論文文献情報（ここではドライなデータと呼ぶ）といった専門知識を補うドライデータの利活用が求められている。そのため、大規模表現型特徴ネットワークから、科学的発見を促進させるためには、大規模ネットワークの視認性を向上し探索を支援する可視化技術と、過去の論文情報や現在の生命科学者の専門知識を統合した分析環境が必要である。そこで我々は大規模な表現型ネットワークに対して粗視化技術やフィルタリングを可能とし且つ論文情報を紐づけることでインタラクティブなグラフ探索を可能とする可視化システム CausalNet を開発した⁵⁾。本システムは生物学者との意見交換で明らかにした機能要件に基づき、①大規模ネットワークに対して対話的にバイクラスタ分析を適用することで部分グラフの抽出を可能にし、②抽出された部分グラフについて論文情報データベースや生命科学者の知見に基づいて効果的に絞り込むことを可能にするシステムである⁵⁾。そして③本システムから抽出された部分グラフに対して、潜在因子を挿入したモデルの確からしさを解析するため、構造方程式モデリング (SEM) 解析が行えるツール iSEM⁹⁾ と連携するためのデータを出力可能とした。このようにウェットのみならずドライ環境において取得されたビッグデータから抽出された因果グラフにおいて、潜在変数発見を促進する視覚的分析環境の構築し、使用事例を通じて有効性を検証した。本稿では使用事例の一つを紹介する（本システムの詳細は文献⁵⁾を参照された）。使用事例では1細胞期から8細胞期までに定義された線虫の272の表現型特徴を対象とし、文献情報には線虫の研究データベースである WormBase から得られた19,888件の抄録のうち244,920件の文章を抽出した。抽出した文章に自然言語処理を用いた前処理をすること

で生物発生のダイナミクスの因果関係を示唆する12,470文を得た。Fig. 2にシステムのGUIを示す。論文より抽出した文章と文献リストは画面下部に表示されノード選択と連携して動作する。また、画面左のクエリインターフェースで設定したグラフが Sugiyama レイアウト¹⁰⁾にて描画されている。また辺交差を低減し可視性を高めると共に完全二部グラフや準バイクリークとしてのつながりを可視化するために、バイクラスタリングを用いた辺集中化¹¹⁾を行っている。本事例では9つの表現型特徴を含むバイクラスタを選択し、選択した表現型特徴の定量データから、7つの観測変数と3つの潜在変数を持つ統計モデルを構築し SEM 解析を行ったところ、ある AB 細胞の分裂軸がほかの後期の細胞期の表現型特徴に対して負の相関があることがわかり、対応する関係性については文献の記述も確認された。表現型特徴のデータ駆動型アプローチとしての CausalNet の取り組みは新たな知見の創出に資する可視化システムとして専門家からも評価されている⁵⁾。

3. 表現型特徴ネットワークと遺伝子ネットワークの探索支援システム

CausalNet は表現型特徴ネットワークの探索を支援するための可視化システムであったが、表現型特徴の情報のみから背景の生物学的なプロセスを推し量るのは困難である。生物学において遺伝子やタンパク質、表現型特徴など異なる階層の情報をまるごと理解することが生物の理解につながると期待されるため、我々は表現型特徴ネットワークに加えて遺伝子ネットワークを分析対象とした⁸⁾¹²⁾。理由としては遺伝子と表現型特徴の関係性は次のように実験的に対応付けられるためである。まずは

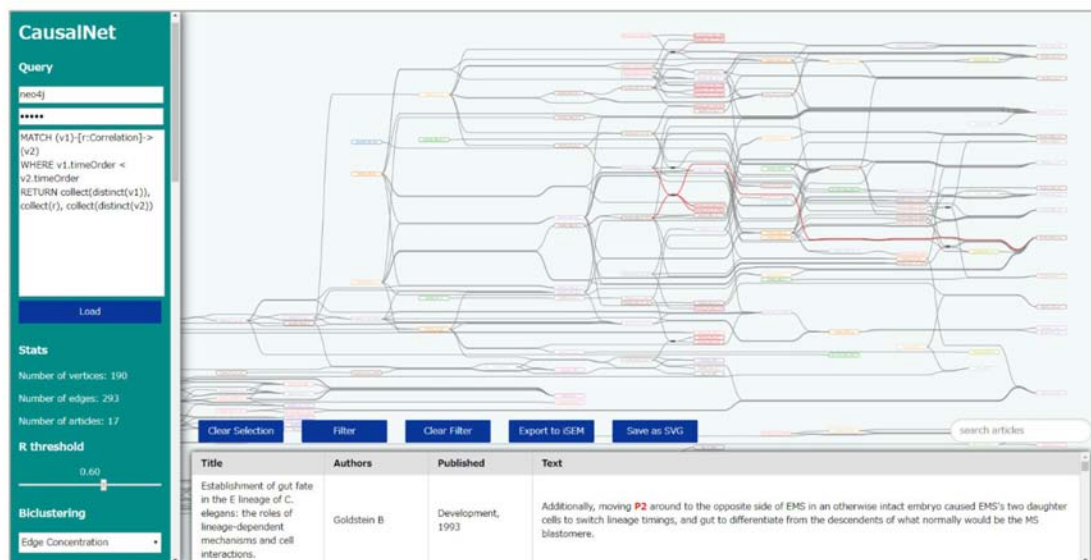


Fig. 2 CausalNet. A screenshot of the integrated visualization system for phenotypic character networks. The main view shows the phenotypic character network. The left column and bottom of the screen display the filtering interface for the network and the related text information, respectively.

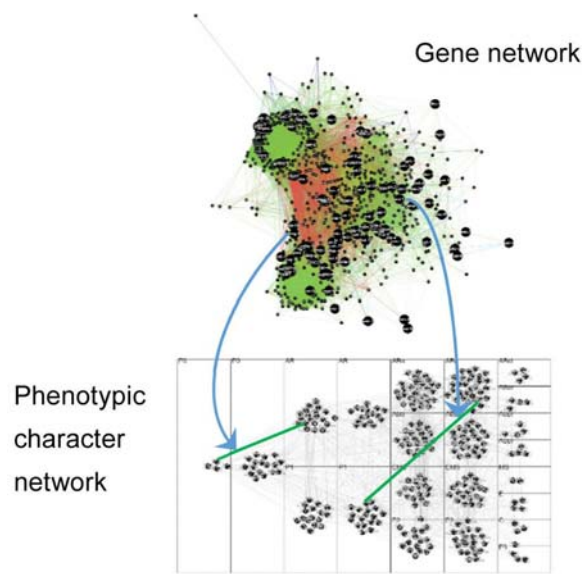


Fig. 3 Integration between gene network and phenotypic character network.

野生胚において表現型特徴を計測したのちに、胚発生に関わる 351 種の遺伝子を不活化した胚に対して同様の表現型特徴を定量化し、正常な胚で観察される表現型特徴相関関係が不活化により変化した場合は、その遺伝子は特徴の発現に関与していると考えられる。このように実験的に表現型特徴関係と遺伝子に対応付けることができる。加えて、遺伝子についても遺伝子間の共発現特徴やタンパク質間相互作用により定められたネットワークが既に構築されており¹²⁾、先行研究の遺伝子ネットワークの情報を利用可能である。Fig. 3 にこれらの表現型特徴

ネットワークと遺伝子ネットワーク、および異種のネットワーク間の関係性を図示した。先述した実験より遺伝子ネットワークにおけるノード（遺伝子）が表現型特徴ネットワークにおけるリンク（表現型特徴相関）に対応付けられる。今、この表現型特徴—遺伝子ネットワークを眺めた時にあらゆる可能性がありうる。このあらゆる可能性の中から専門家の知識や種々の情報、ユーザー操作をもとにネットワークを絞り込み、生物学者が興味を持つ「検証可能な仮説」を導出するには、インタラクティブな可視化システムを用いたユーザー主導型のデータ探索を可能とする視覚的分析が有用である。そこで我々は、生物学における科学的発見を促進するために表現型特徴—遺伝子ネットワークの横断的探索を支援する視覚的分析システムを開発している。

ネットワークの横断的探索を可能にし、新たな仮説構築を支援する視覚的分析システムを構築するにあたり、共著者である分子生物学者との議論をもとに以下の 5 つのシステムの要件を導出した。①両ネットワークが Multiview で閲覧でき、実験的な対応関係をもとに連携して可視化されること。②関心のある遺伝子を直接可能であり、遺伝子オントロジーのような遺伝子情報をもとにフィルタリング可能なこと。③表現型特徴も同様にフィルタリングが可能なこと。④ネットワーク上でのパス探索の支援。⑤グラフ探索のプロセスを容易に把握できること。我々はこれらを満たす Fig. 4 のような視覚的分析システムのプロトタイプを構築した。本システムは 2 つのネットワークを同時に俯瞰でき、ユーザー操作によるグラフハイライトにより連携可視化を実現した (①)。また遺伝子情報や表現型特徴情報に基づきネット

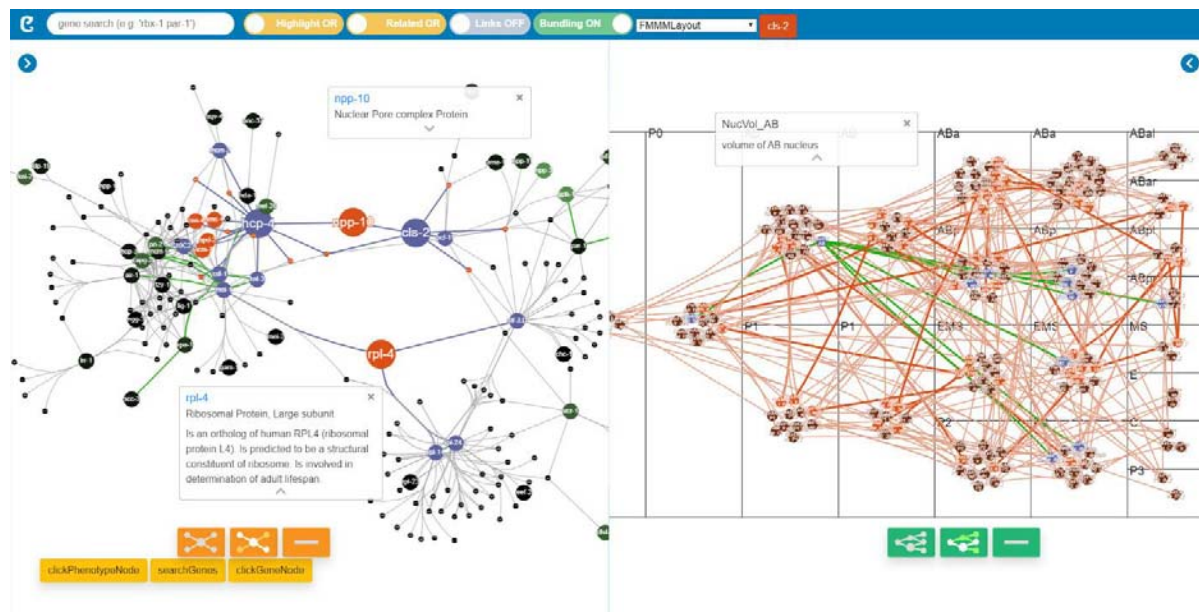


Fig. 4 PheGeNet. A graphical user interface of our proposed explorative visualization system for phenotypic character network (right) and gene network (left). Gene network is visualized using FM3 layout¹³⁾ and phenotypic character network is visualized using group-in-a-box layout¹⁴⁾¹⁵⁾.

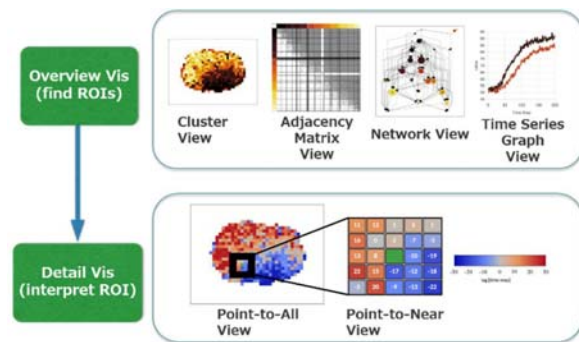


Fig. 7 Overview and detail analysis pipeline in the calcium wave visualization system.

ルシウム波の実データ分析を行った。その結果、提案システムを用いた場合の方が特異点発見の正解率が増加し、さらに専門家が今までに着目していなかった新たな ROI の発見を可能にし、分析支援ができることがわかった。

5. おわりに

本稿では、バイオイメージ・インフォマティクスにより取得された線虫の発生ダイナミクスのデータに対して、可視化技術によりデータ駆動型アプローチの促進を目指したいくつかのアプローチを紹介した。とするとこれまでの研究プロセスの中で研究者の頭の中で閉じていた思考や、研究グループの言論空間に埋め込まれていたアイデアをこのような可視化技術を用いて表出させることは、生物学の研究プロセスを促進させる新たなアプローチになりうると考えている。

また、異なる階層のネットワーク（マルチレイヤーネットワーク）に代表されるように、生物学のドメインにおいては大規模・複数・異種のデータから新たな知見を得るためのデータ駆動型科学的アプローチを目指した取り組みが盛んであるが、可視化の基盤技術や可視化の文脈でのタスク分類はあれども、生物学ドメインの研究者と密に連携して現場で必要とされている要件を洗い出し、システムを設計し、応用まで昇華した例はまだ少ない。これからは可視化と生物学ドメインのギャップを埋めるために、両学問分野が連携し、知の創出を促進させることが求められるであろう。

謝 辞

本研究を共同で遂行していただいた理化学研究所の大浪修一先生、京田耕司研究員、ならびに大浪グループの方々に感謝を表す。また、本研究は JST, CREST, JPMJCR1511 の支援を受けたものである。

参 考 文 献

- 1) Schneider, C.A., Rasband, W.S., Eliceiri, K.W.: NIH Image to ImageJ: 25 years of image analysis, Nat. Methods, Vol. 9, No. 7 (2012) pp. 671-675.
- 2) Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.Y., White, D.J., Hartenstein, V., Eliceiri, K., Tomancak, P., Cardona, A.: Fiji: an open-source platform for biological-image analysis, Nat. Methods, Vol. 9, No. 7 (2012) pp. 676-682.
- 3) Carpenter, A.E., Jones, T.R., Lamprecht, M.R., Clarke, C., Kang, I.H., Friman, O., Guertin, D.A., Chang, J.H., Lindquist, R.A., Moffat, J., Golland, P., Sabatini, D.M.: CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol., Vol. 7, No. 10 (2006) pp. R100.
- 4) Chen, J., Ding, L., Viana, M.P., Hendershott, M.C., Yang, R., Mueller, I. A., Rafelski, S. M.: The Allen Cell Structure Segmenter: a new open source toolkit for segmenting 3D intracellular structures in fluorescence microscopy images, bioRxiv, (2018) doi: <https://doi.org/10.1101/491035>
- 5) Onoue, Y., Kyoda, K., Kioka, M., Baba, K., Onami, S., Koyamada, K.: Development of an integrated visualization system for phenotype character networks, In Proc. of IEEE Pacific Visualization 2018 (VisNotes), (2018) pp. 21-24.
- 6) Natsukawa, H., Ishino, S., Onoue, Y., Kyoda, K., Onami, S., Koyamada, K.: Explorative visualization for phenotypic-genotypic network, IEEE Visualization 2019 (Poster), (2019) pp. 1-2.
- 7) Umezawa, K., Natsukawa, H., Onoue, Y., Koyamada, K.: A visual analytics system to support the formation of a hypothesis from calcium wave data, Visual Informatics, Vol. 2, No. 1 (2018) pp. 2-13.
- 8) Kyoda, K., Adachi, E., Masuda, E., Nagai, Y., Suzuki, Y., Oguro, T., Urai, M., Arai, R., Furukawa, M., Shimada, K., Kuramochi, J., Nagai, E., Onami, S.: WDDD: Worm Developmental Dynamics Database, Nucleic Acids Res., Vol. 41 (Database issue), (2013) pp. D732-D737.
- 9) Ashida, Y., Umezawa, K., Onoue, Y., Koyamada, K.: Development of a visual analytics system for structural equation modeling, In Proc. of IEEE Pacific Visualization 2017 (Poster), (2017) pp. 338-339.
- 10) Sugiyama, K., Tagawa, S., Toda, M.: Methods for visual understanding of hierarchical system structures, IEEE Transactions on Systems, Man, and Cybernetics, Vol. 11, No. 2 (1981) pp. 109-125.
- 11) Onoue, Y., Kukimoto, N., Sakamoto, N., Koyamada, K.: Minimizing the number of edges via edge concentration in dense layered graphs, IEEE TVCG., Vol. 22, No. 6 (2016) pp. 1652-1661.
- 12) Gansalus K. C., et al., Predictive models of molecular machines involved in C. elegans early embryogenesis, Nature, Vol. 436, No. 7052 (2005) pp. 861-865.
- 13) Giot, R., Bourqui, R.: Fast graph drawing algorithm revealing networks cores, In Proc of 19th International Conference on Information Visualisation (2015) pp. 259-264.
- 14) Aoyama, N., Onoue, Y., Ueno, Y., Natsukawa, H., Koyamada, K.: User Evaluation of Group-in-a-Box Variants, In Proc. of IEEE Pacific Visualization Symposium (PacificVis) 2019, (2019), pp. 127-136.
- 15) Ueno, Y., Natsukawa, H., Aoyama, N., Koyamada, K.: Exploration behavior of group-in-a-box layouts, Visual Informatics, Vol. 3, No. 1 (2019) pp. 38-47.