

厚生労働行政推進調査事業費（化学物質リスク研究事業）
令和2年度 分担研究報告書

新型毒性試験法とシステムバイオロジーとの融合による有害性予測体系の構築
(H30-化学-指定-001)

システム毒性解析の人工知能化

研究分担者： 北野宏明

特定非営利活動法人システム・バイオロジー研究機構 会長

研究協力者：長谷武志（特定非営利活動法人 システム・バイオロジー研究機構）

Natalia Polouliakh （株式会社ソニーコンピュータサイエンス研究所）

研究要旨

システム毒性では、一連の解析手順の高度な連携と同時に、大規模データベースから多くの情報を抽出し、それを解析へと結びつける必要がある。本研究では、従来、解析工程の律速段階となっていた、Percellome データから有意に変動した遺伝子を専門的判断に基づき逐一抽出する作業を完全自動化すべく、深層学習(Deep Learning)を用いて膨大な遺伝子変動データから有意に変動した遺伝子を高精度で自動同定させる技術、ならびに転写制御領域の解析ソフトウェア SHOE と GarudaPlatform の連携を中心に解析パイプラインの機能強化を実施している。

本年度も新たな学習用データを追加して予測能力のより一層の向上を実現した。また予測モデルの判断基準を可視化する explainability model (grad cam)を参考にモデルの最適化を進めた上で、これらを組み込んだユーザーフレンドリな GUI を実装した解析ソフトウェアを開発した。さらに、遺伝子発現パターンによるクラスタリングソフトウェア AGCT と転写制御領域の解析ソフトウェア SHOE の改良を重ね、Garuda Platform のソフトウェア等、複数の解析ツールを使った統合パイプラインを構築・強化した。

A. 研究目的

システム・レベルで毒性を理解するには、膨大な実験データを格納したデータベース、文献、数値モデルなどを統合的に解析する必要があり、大規模かつ複雑なデータを意味のある形で解析するには、深層学習やテキストマイニングなどを含めた一連の人工知能 (AI) アルゴリズム群の連携が有効である。さらに、複数の解析ツールをスムーズに連動させる必要

がある。本分担研究では、一連の解析過程の AI 化を実施し、ツール間連動を強化することで、高度な AI 駆動型システム毒性学基盤の構築を推進する。

B. 研究方法

●深層学習を用いた大規模遺伝子発現データベースからの重要遺伝子群の判別

平成 30 年度に深層学習を用いた 3 次元グラフの画像解析システム DTOX を構築した。この深層学習モデルでは、有意な遺伝子のうち、化学物質曝露により発現が上昇する(up-regulate)ものを positive 群、低下する(down-regulate)ものを negative 群、さらに有意で無い遺伝子は、non-significant 群として予測する。平成 31/令和元年度には多様な描画角度から 3 次元グラフを作成してトレーニング画像数を 10 倍に増加させることで、深層学習による分類器の精度を大幅に向上させた。令和 2 年度はさらに特殊パターン（飽和パターンや溶媒群に変動のあるパターンなど）を追加してトレーニングを実施した。

また、DTOX システムに対して、ユーザーが使いやすいグラフィカルユーザーインターフェース(GUI)の実装を進めた。GUI は、python の代表的な GUI 作成用のモジュールである、PYQT5 と、Qt designer を用いて作成を進めた。

DTOX の性能比較のためには、一般的な変動遺伝子解析(DEG: differential expression gene analysis)を利用した。DEG では、投与量が 0 で時間が 0 の場合のデータをコントロール群とし、その他の条件下においてコントロール群と比較して有意に遺伝子発現が大きい若しくは小さい遺伝子群を、その条件下において有意であると見なした。有意性の検定では、t-test で求めた p 値に対して多重検定補正を実行して FDR を求め、FDR < 0.05 である場合に有意であるとした。コントロール以外に 15 条件あるが、15 条件の内、1 条件以上において FDR < 0.05 で log₂ Fold change が 1 以上の遺伝子（ただし、log₂ Fold change が -1 以下となる場合が無い場合）を positive 群とした。一方、15 条件の内、1 条件以上において FDR < 0.05 で log₂ Fold change が -1 以下の遺伝子（ただし、log₂ Fold

change が 1 以上となる場合が無い場合）を negative 群とした。positive 群でも negative 群でも無い遺伝子を、non-significant 群とした。

●Garuda Platform 上での Percellome と SHOE の連動強化

転写領域解析ソフトウェア SHOE 及びクラスタリングソフトウェア AGCT の開発は、Java 言語 (USA, Oracle Inc.) で行った。Garuda Platform 用ソフトウェア (Garuda ガジェット) の開発や他の Garuda ガジェットとの連動については、GarudaDevPack を使用した。性能評価や試験運用には、Percellome データベースより実際の化学物質曝露による遺伝子発現時系列データを用いた。生成クラスタのパスウェイ解析には Ingenuity Pathway Analysis (Germany, QIAGEN)を使用した。

C. 研究結果

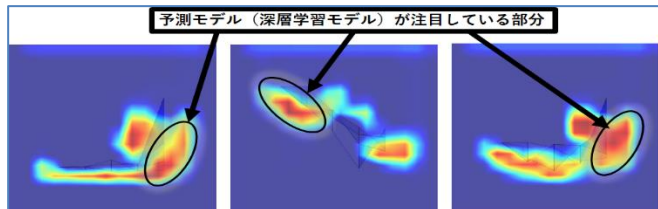
当初計画に沿って研究を行い、下記の成果を得た。

●深層学習を用いた大規模遺伝子発現データベースからの重要遺伝子群の判別

平成 30 年度に深層学習を用いて開発した 3 次元グラフの画像解析システム DTOX は、専門家の判断を訓練データとして用いた深層学習モデルによって画像を解析し、各遺伝子を発現上昇、発現低下、それ以外に分類する。平成 31/令和元年度も多様な描画角度から 3 次元グラフを作成してトレーニング画像数を 10 倍に増加させることで、DTOX による分類精度を大幅に向上させ、令和 2 年度も特殊パターン（飽和パターンや溶媒群に変動のあるパターンなど）を追加してトレーニングすることで、分類器の精度と網羅性を向上させた。

また機械学習の予測モデルの判断基準を明確にする explainability model (grad cam : R. R. Selvaraju, M. et al. (2019) Grad-CAM: Visual Explanations from Deep

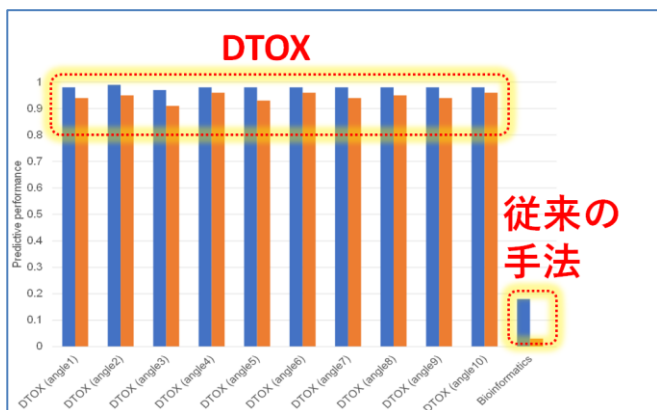
Networks via Gradient-Based Localization. Int J Comput Vis doi:10.1007/s11263-019-01228-7) を導入してモデルの最適化を進めた。



【Explanatory model (Grad cam) による深層学習モデルの解析の結果：深層学習モデルは、赤く強調されている所に着目して、判別を行っている。】

深層学習モデルによる DTOX と、一般的なバイオインフォマティクス解析パイプラインとの予測精度の比較を行った結果、下図に示すように、深層学習モデルは 97%以上の高い正解率を示したが、バイオインフォマティクス解析パイプラインは 18%と非常に低い正解率であった。このことから、我々の深層学習モデル DTOX は、一般的に使用されてきた手法と比べて、より高精度に化合物投与下で影響を受ける遺伝子群を探索できることが示された。

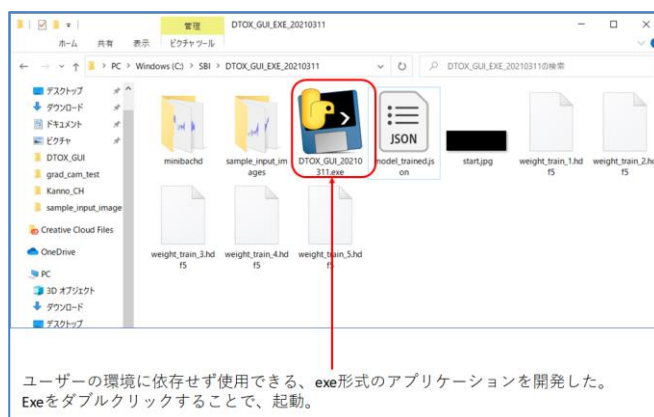
以上の結果から、DTOX を解析パイプラインに組み込むことにより、遺伝子群の探索の精度を大幅に改善することが出来るので、その下流のメカニズムに関する解析 (パスウェイ解析等) をより高精度に行うことが出来るようになることが期待される。



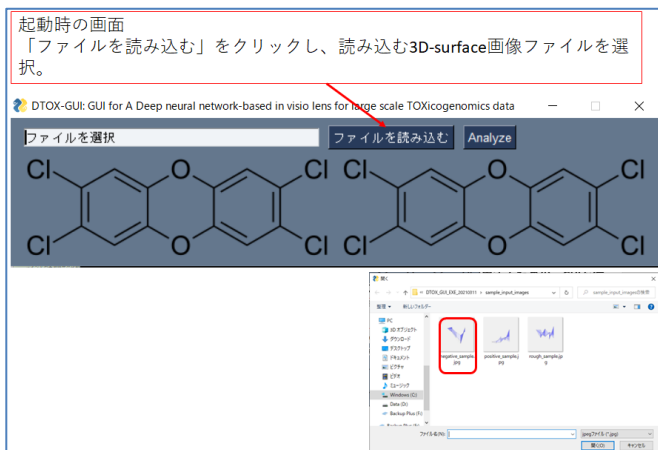
深層学習モデル DTOX の予測精度と一般的なバイオインフォマティクス解析パイプライン(右端)の予測精度の比較

ところで、H3 1年度の段階では、深層学習モデル DTOX には、グラフィカルユーザーインターフェース (GUI) は実装されておらず、また、DTOX を使用した解析には、深層学習計算を実施するための特殊な計算環境を整える必要があった。そのため、深層学習モデル DTOX を、ユーザーの計算環境にインストールし実行するためには、コンピューターサイエンスと深層学習の専門知識を備えた専門家のサポートが必要不可欠であった。この問題点は、DTOX を研究者等のユーザーが使用するための、大きな障害となっていた。

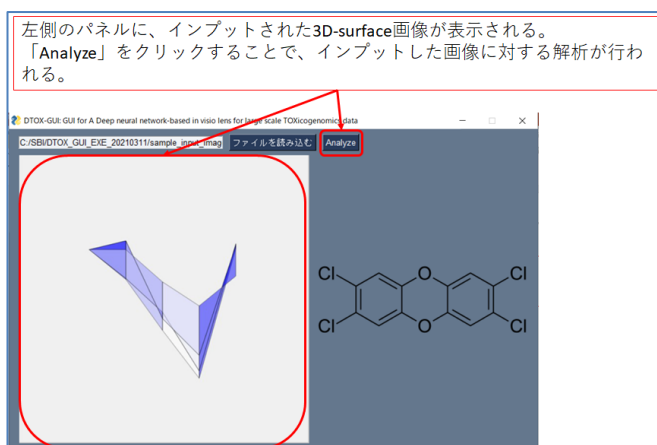
そこで、R2年度は、研究者等のユーザーが使いやすい GUI の実装を進めた。特に、DTOX で必要なプログラム群とトレーニング済みの予測モデルを圧縮し組み込むことにより、ユーザーの計算環境に依存せず実行可能な、GUI ソフトウェアとして実装を行った。具体的には、研究者等のユーザーは、exe 形式のファイルをクリックするのみで、深層学習モデル DTOX を使用した GUI を利用できる。



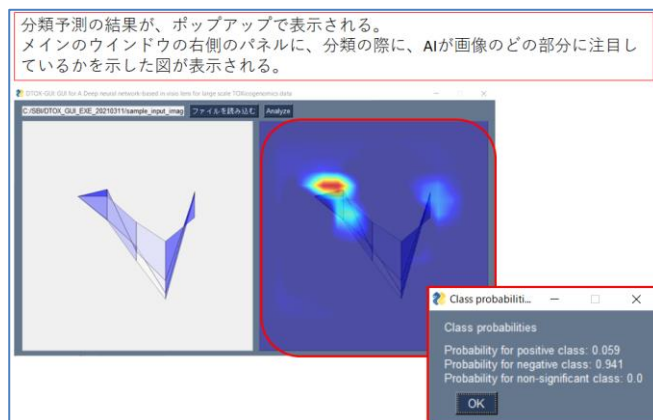
DTOX の GUI では、まず“ファイルを読み込む”ボタンをクリックして、遺伝子の三次元画像ファイルを選択し読み込む。



読み込んだ画像は、の左側のサブウィンドウに表示される。その後、“Analyze” ボタンをクリックすることで、読み込んだ画像に対して、深層学習を用いた解析が実行され予測が行われる。



予測結果はポップアップウィンドウに表示され、一方、深層学習モデルの判断根拠に関する解析の結果は右側のサブウィンドウに表示される。

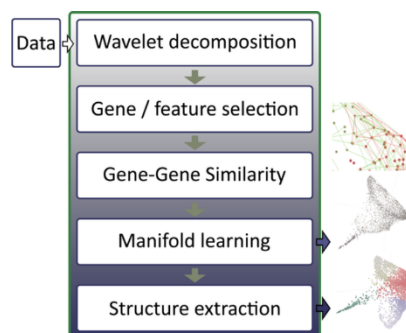


この画像の例では、この遺伝子は化学物質により down-regulation されている確率(negative class の class probability)が 0.941 であると予測されている。また、判断根拠とされた画像領域は赤色でハイライトされている。

●Garuda Platform 上での Percellome と SHOE の連動強化

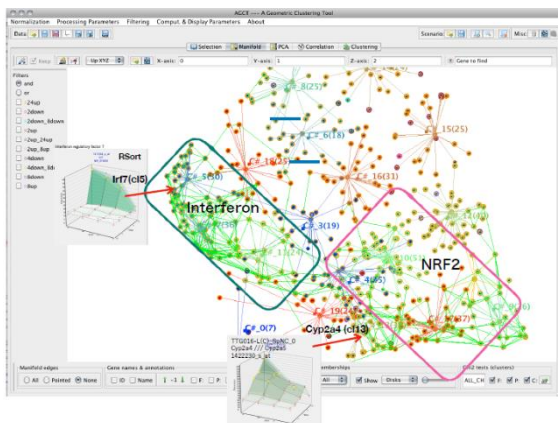
今年度は、先行研究で開発を進めていた転写領域解析ソフトウェア SHOE の改良に加え、別途開発していたクラスタリングソフトウェア AGCT の整備を完了し、性能評価の上、論文発表を行った。

AGCT は時系列データとしてトランスクリプトームデータを評価してクラスタリング計算を行う。解析ワークフローは次の通り。



性能評価のために Percellome データベースに収録されている Pentachlorophenol(PCP)の単回投与データセットから有意な変動を示した 513 プローブセットの時系列データを抽出し、これを用いて教師なしクラスタリング

を実行した。



この結果、Affinity Propagation method では 21 クラスターが生成され、特に Interferon signaling 関係の遺伝子が集まった 3 クラスターと NRF2 関係の遺伝子が集まった 4 クラスターを得た。これは菅野らが 2013 年に報告した論文 (DOI: 10.2131/jts.38.643) と一致するものであり、解析精度を維持しつつ解析工程の効率化に成功したものである。

また Garuda プラットフォームをハブにして、SHOE や AGCT、その他の Garuda ガジェットの連携を整備し、動作を確認した。

A new Toxicology gadget to connect SHOE, Target Mine and Reactome



D. 考察

本研究で開発した深層学習画像解析システム DTOX の予測精度は、一般的なバイオインフォマティクス解析パイプラインの精度を大きく上回ることが判った。これは、DTOX の深層学習モデルは、研究者が Visual Selection の際に評価している領域を着目して判定していることから、トレーニングにより人の判断のパターンを上手く捉えることで、高い精度を実現している可能性を示唆している。深層学習モデルの精度はデータの数に依存するため、更なる精度の向上を目指して、トレーニングデータの追加増強を進め、これをもとに、深層学習モデルの再チューニングを進める。

E. 結論

本研究は、ほぼ計画通りに進捗した。

一般の毒性学研究者にも使いやすい GUI を備えた解析ソフトウェアの導入により、遺伝子発現解析工程の効率と精度の向上が見込まれ、従来は困難であった規模の毒性シグナル連動解析が可能になると期待される。また、SHOE と Percellome の連動解析においても AGCT によるクラスタリング工程を前処理に加えることにより分解能と効率が向上し、よりスムーズな解析が可能となった。今後は、これらの成果をさらに多くの解析プロセスに展開することを目指す。

F. 研究発表 (3 年間分)

1. 論文発表

●令和 2 年度

(1) Nock R, Poluliakh N, Nielsen F, Oka K, Connell CR, Heimhofer C, Shibani K, Ghosh S, Aisaki K, Kitajima S, Kanno J, Akama K, Kitano H, A Geometric Clustering Tool (AGCT) to robustly unravel the inner cluster structures of time-series gene expressions. PLOS ONE 15(7): e0233755. 2020

[doi: 10.1371/journal.pone.0233755]

●令和元年度

(2) Gupta V, Crudu A, Matsuoka Y, Ghosh S, Rozot R, Marat X, Jäger S, Kitano H, Breton L. Multi-dimensional computational pipeline for large-scale deep screening of compound effect assessment: an in silico case study on ageing-related compounds. *npj Syst Biol Appl* 5, 42 2019.

(3) Uchida S, Asai Y, Kariya Y, Tsumoto K, Hibino H, Honma M, Abe T, Nin F, Kurata Y, Furutani K, Suzuki H, Kitano H, Inoue R, Kurachi Y. Integrative and theoretical research on the architecture of a biological system and its disorder *J Physiol Sci* 69 433 2019

(4) Kato Y, Tabata K, Kimura T, Yachie-Kinoshita A, Ozawa Y, Yamada K, Ito J, Tachino S, Hori Y, Matsuki M, Matsuoka Y, Ghosh S, Kitano H, Funahashi Y. Lenvatinib plus anti-PD-1 antibody combination treatment activates CD8+ T cells through reduction of tumor-associated macrophage and activation of the interferon pathway *PLOS ONE*
doi.org/10.1371/journal.pone.0212513

(5) Polouliakh N. In Silico Transcription Factor Discovery via Bioinformatics Approach: Application on iPSC Reprogramming Resistant Genes. *Leveraging Biomedical and Healthcare Data, 2019 - Elsevier*

●平成 30 年度

(6) Natalia Polouliakh, Paul Horton, Kazuhiro Shibana, Kodai Takata, Vanessa Ludwig, Samik Ghosh and Hiroaki Kitano Sequence homology in eukaryotes (SHOE): interactive visual tool for promoter analysis; *BMC Genomics* September 2018, 19: 715;

2. 学会発表

●令和 2 年度

(1) Samik GHOSH, Takeshi HASE, Ayako YACHIE, Sucheendra Kumar PALANIAPPAN (2020) Towards an optimal dose of AI in biomedicine and healthcare. 第47回日本毒性学会学術年会(2020.6.30.) オンライン

●令和元年度

(2) Natsume-Kitatani Y, Mizuguchi K, Aisaki K, Kitajima S, Ghosh S, Kitano H, Kanno J. Pentachlorophenol affects RIG-1 antiviral pathway that produces type 1 interferon at the transcriptional level *ISMB/ECCB 2019 バーゼル (スイス)*, 2019/07/24

(3) Natsume-Kitatani Y, Aisaki K, Kitajima S, Ghosh S, Kitano H, Mizuguchi K, Kanno J. Cross Talks among PPARα, SREBP, and ER Signaling Pathways in the Side Effect of Valproic Acid *IUTOX2019 ホノルル (ハワイ)*, 2019/07/16

(4) 夏目 やよい, 相崎 健一, 北嶋 聡, Gosh Samik, 北野 宏明, 水口 賢司, 菅野 純 *Garudaプラットフォームによる多角的毒性予測* 第46回日本毒性学会学術年会 徳島, 2019/06/28

●平成 30 年度

(5) Yayoi Natsume-Kitatani, Ken-ichi Aisaki, Satoshi Kitajima, Samik Ghosh, Hiroaki Kitano, Kenji Mizuguchi, Jun Kanno, Inferred role of crosstalk between PPARα and ER signaling pathways in the toxicity of valproic acid: systems toxicology approach, *International Society for Computational Biology (ISMB) 2018, (2018.7.6-10) Chicago, USA*

⑥ Takeshi Hase, Samik Ghosh, Ken-ichi Aisaki, Satoshi Kitajima, Jun Kanno, and Hiroaki Kitano. DTOX: Deep neural network-based computational framework to analyze omics data in Toxicology. OPENTOX ASIA 2018, Asahi Seimei Otemachi Building, Tokyo, May 25, 2018.

⑦ Kitano, H. AI-driven systems toxicology. Toxicogenomics for Accelerated and Refined Hazard Identification of Chemicals (Joint Symposium between IUTOX and IUPHAR), 18th World Congress of Basic and Clinical Pharmacology, Kyoto International Conference Center, Kyoto, July 6, 2018. (invited)

(8) 北野 宏明. AI駆動型システム医科学の展望. 山口大学第2回シンポジウム: 人工知能・システム医学による難治性疾患への新たな挑戦, KKR山口あさくら, 山口, Aug. 4, 2018. (invited)

(9) Kitano, H. AI Grand Challenges. 2018年日台科学技術フォーラム, リージェント台北, 台湾, Sep. 10, 2018. (invited)

⑩ 北野 宏明. 人工知能駆動型システム毒性学の展望. CBI学会2018年大会, タワーホール船堀, 東京, Oct. 9, 2018. (invited)

(11) 北野 宏明. AI戦略とムーンショット. データサイエンティスト協会シンポジウム, JPタワーホール&カンファレンス東京, 東京, Oct. 19, 2018. (invited, keynote)

⑫ Natsume-Kitatani Y., Aisaki K., Kitajima S., Ghosh S., Kitano H., Mizuguchi K., Kanno J. “Inferred role of crosstalk between PPAR α and ER signaling pathways in the toxicity of valproic acid: systems toxicology approach” ISMB 2018, (2018.7.7, USA), (Poster)

G. 知的所有権の取得状況

1. 特許取得

なし

2. 実用新案登録

なし

3. その他

なし