

厚生労働行政推進調査事業費（化学物質リスク研究事業）  
トキシコゲノミクスとシステムバイオロジーとの融合による  
新型化学物質有害性評価系の実装研究  
（21KD2001）

令和5年度 分担研究報告書

統合ツール“Percellome Integrator”の開発

研究分担者 相崎 健一

国立医薬品食品衛生研究所

安全性生物試験研究センター 毒性部

第一室 室長

研究要旨

本研究は、毒性の分子生物学的機序に基づいて、現行の不確実係数（安全係数）を利用する数量的な有害性評価手法を補強し、より迅速で、高精度且つ省動物を具現化した新たな有害性評価系の開発を目標とする。方法として、マイクロアレイ（GeneChip）と次世代シーケンサを用いて分子生物学的な情報の基盤となる遺伝子発現及びエピゲノムの網羅的データを得つつ、独自開発のソフトウェア群による化学物質の生体影響の網羅的分析法の体系化を行い、これに、毒性学・分子生物学に精通したデータサイエンス専門家を擁してのシステムバイオロジー及び人工知能（AI）技術を融合した新たな有害性評価系の開発を進めた。

特に先行研究において、Percellome 法\*を基盤とする「新型」反復曝露実験\*\*の蓄積により構築した化学物質の反復曝露による生体影響のデータベースについては、溶媒の反復曝露影響や、反復曝露影響の可逆性・非可逆性を遺伝子単位で取得した網羅的情報をデータベースに反映することにより、評価系の解析精度を向上させる。既に構築済みの単回曝露実験のデータベースと共にこれを利用することで、現段階では長い時間と多額の費用を要している長期反復曝露の毒性評価の期間短縮と効率化を検証した。

令和5年度は、新型化学物質有害性評価システムの解析コアの主要ユニットとしての解析モジュールのデザインを終え、コンセプトモデルを作成した。また変動遺伝子の自動抽出パイプラインの改良を継続し、テンプレートコレクションの充実を図った。

-----  
(\*) mRNA発現値を細胞1個当たりのコピー数として絶対定量する方法。

(\*\*) 全動物に同量の検体を反復曝露し、遺伝子発現測定直前の曝露時に、溶媒群、低用量群、中用量群、高用量群に分けて最終曝露を一回行う。実験の反復曝露と単回曝露の回数をもとに[14+1]、[4+1]、[0+1]等と表記することとした。

## A. 研究目的

本研究は、独自構築したトキシコゲノミクス・データベース (DB) にインフォマティクス、及び、人工知能 (AI) を拡大適用し、化学物質が実験動物に惹起する遺伝子発現変動等の分子毒性学情報から、科学的根拠に基づく有害性予測評価手法を確立する。これにより「安全係数」を用いる従来の有害性評価手法を補強するとともに、迅速、高精度、省動物を具現化する新たな評価システムを構築することを目的とする。

即ち、先行研究にて構築済みの延べ 8 億 5 千万遺伝子発現情報からなる高精度トキシコゲノミクスデータベースと単回曝露及び反復曝露の毒性ネットワーク解析技術を基盤に、これらを維持・拡充しつつ、さらに臓器別のゲノム DNA メチル化及び代表的物質の反復曝露によるヒストン修飾情報を加えて、毒性ネットワーク解析による、短期間試験での反復曝露毒性の予測評価技術を開発する。この際、インフォマティクス専門家によりシステムトキシコロジーや人工知能の技術を融合し、反復曝露にも対応する新型化学物質有害性評価系の実装を進めた。

本分担研究では、新型化学物質有害性評価系の実装に必要な全ゲノム配列を参照する Percellome 解析に最適化された統合ソフトウェアの開発を進めた。

## B. 研究方法

ソフトウェアの in house 開発に際しては、先行研究で開発したソフトウェアの改良の際は開発効率と生成する実行バイナリの実行速度を重視して、Win32/64 開発及び Web アプリケーション開発は RAD (Rapid Application Development) 対応の Delphi (Object Pascal 言語、USA, Embarcadero Technologies, Inc.) を用いた。データベースエンジンには組込型の DBISAM (USA, Elevate Software, Inc.) を、一般的な

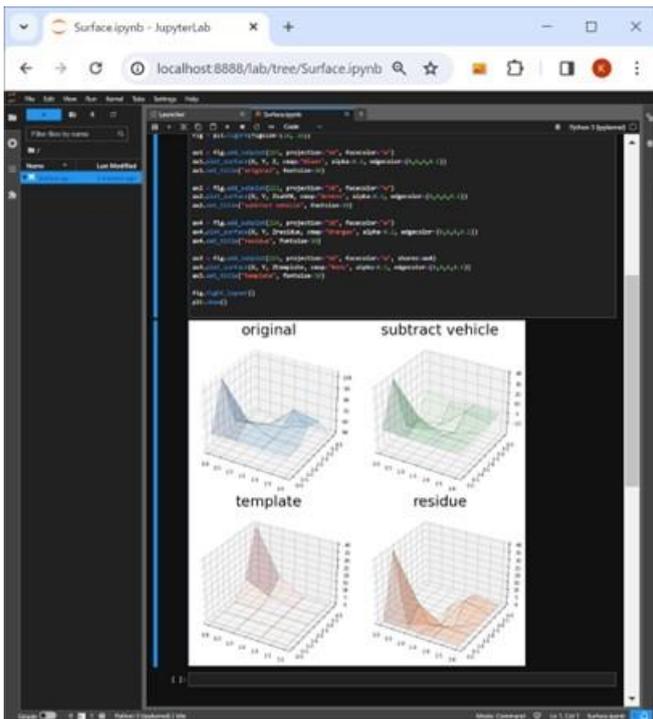
グラフ描画には TeeChart (Spain, Steema Software SL) を利用した。新たに開発するソフトウェアについては、ライブラリが充実している Python 言語 (ver.3.9.1) を使用した。主な解析ライブラリとしては numpy (ver.1.19.5)、pandas (ver.1.2.1)、scikit-learn (ver.1.2.0)を、グラフ描画ライブラリとして matplotlib (ver.3.3.4) を用いた。

## C. 研究結果

令和5年度も最終目標である新型化学物質有害性評価システムの解析コアの主要ユニットとしての、データ解析ユニットの開発を進めた。

今回、コンセプトモデルのプラットフォームとしては、実行した Python コードと出力結果を記録しつつ、容易に様々な試行を実行可能、つまり探索的解析に適した Jupyter Notebook/Lab を採用した。

ある程度定型化した処理については、雛形となるソースコード(スクリプト)を集めたりポジトリを整備する。プログラミングできないユーザーでも、これらを Jupyter Notebook/Lab に読み込み実行すれば、解析処理が実行され、計算結果の数値が出力され、数値やグラフ・画像がソースコードに続けて出力、表示される。



ここに考察などのメモを追加することも可能であり、電子的な実験ノートとしての活用が期待される。これら、ソースコード(プログラム/スクリプト)とその実行結果、実験ノートを一括して、専用ファイル形式(.ipynb)で保存できる。このファイル形式は基本的にはテキストデータで、現在は勿論、将来的にも可用性が高く、ファイルが開けないなどといった事態に陥ることもなく、記録データ形式としては理想的なものである。

また昨年度に引き続き、遺伝子発現の3次元パターン(Surface)の特徴分類アルゴリズムの開発については、分類指標となるテンプレートのコレクションの網羅的な作成を進めた。具体的には、クラスタリングを行い、各クラスタのセントロイドを代表パターンとして採用した。今回は、密度準拠クラスタリングアルゴリズムであるDBSCANを用い、主要なパターンを収集した。

## D. 考察

開発はほぼ計画通りに進行した。新型化学物質有害性評価システムの解析コア開発については、コンセプトモデルの作成を終えた。システムのメンテナンス性、拡張性、ユーザー利便性、及び、完全再現性を実現するための記録保持、即ち使用プログラムのバージョン管理、解析パラメータの保存、解析計算過程の記録、解析結果の保存、実験ノートの保存、の全てに対応しうる環境として、Jupyter Notebook/Labを選択するのは、オープンソース及びオープンサイエンスの流れに沿った適切なものと考えられる。また専用ファイル形式.ipynbが基本的にテキストデータであることは、将来的な技術変遷にも対応しうる利点である。GUIに慣れたエンドユーザーにとって、ソースコードとの距離が近い事は当初、戸惑うかもしれないが、必要なければ書き換える必要が無くブラックボックスとして扱えば良い事が分かれば、直ぐに対応が可能であると思われる。また実験・解析記録として、実行したソースコードと計算結果(グラフや画像を含む)、考察などのメモを残すことの出来るMarkdownデータがオールインワンで揃っている.ipynbファイルは、ほぼ完全なものとなりうる。

今後は解析用プログラム(スクリプト)のリポジトリを充実させ、充分なセキュリティを考慮した上で、GitHubなどオンラインリポジトリとの連携も考える。また研究が進み、完全に定型化した工程については、ユーザー利便性が増すなら、そのソースコードを基にElectron、Elmなどを活用して独立起動可能なアプリケーションソフトウェアに纏めることとする。

Garuda Platformとの連携についてはまだ実現できていないが、アダプタとなるGadgetを作成することで実現可能であり、優先的に開発を進める。

遺伝子発現の3次元パターン(Surface)の特徴分類については、テンプレートライブラリの網羅的な拡充にも目処が立ち、早期の実用化が可能である。当初はPercellome Integratorのスクリプトリポジトリに

加えることで最適化を兼ねた早期稼働を目指す。

なお今回実施したクラスタリングによる代表パターンの抽出方法では、孤立パターン（クラスタを形成しないもの）を抽出する事が出来ないため、クラスタを形成しなかったパターンから何らかの方法でノイズを除外し、残ったパターンから生物学的に有意な変動パターンを取得する方法の開発も進める。

## E. 結論

新型化学物質有害性評価システムの解析コア主要ユニットとしての、データ解析ユニットのコンセプトモデルの作成を終えた。現在のワークフローに適応した、探索的解析に最も適した設計となった。多方面で実績のある Jupyter Notebook/Lab ベースとすることで、解析履歴の保存が難しかった *in silico* の解析操作について、より完全性の高い記録が可能で、尚且つユーザービリティの高いシステムの実現が期待される。

## F. 研究発表

### 1. 論文発表

- (1) Takeshi Hase, Samik Ghosh, Ken-ichi Aisaki, Satoshi Kitajima, Jun Kanno, Hiroaki Kitano, Ayako Yachie. DTox: A Deep neural network-based in visio lens for large scale Toxicogenomics data. J. Toxicol. Sci. 2024; 49(3): 105-115.[doi.org/10.2131/jts.49.105].
- (2) 菅野純、相崎健一、北嶋 聡: 遺伝子発現を指標とした毒性評価・予測, 単行本「化学物質の複合影響と健康リスク評価」, 2024; 第2章 複合曝露による毒性の評価手法 第1節, 67p, 医歯薬出版(東京) [ISBN: 978-4-263-73220-5]

### 2. 学会発表

- (1) Jun Kanno, Keynote Lecture: Molecular toxicology in health and diseases, The 13 th National Conference in Toxicology (NCT13), Thai Society of Toxicology, (2023.08.06), Bangkok, Oral.
- (2) 菅野 純、相崎 健一、小野 竜一、北嶋 聡:

新型反復曝露実験による PFOA の毒性発現分析 - Clofibrate の網羅的エピジェネティック情報を参照して - . 第50回日本毒性学会学術年会、(2023.6.19)、パシフィコ横浜 会議センター、シンポジウム、口演

- (3) 夏目 やよい、相崎 健一、北嶋 聡、菅野 純: PPAR alphaリガンドが惹起する遺伝子発現プロファイルの比較解析. 第50回日本毒性学会学術年会、(2023.6.19)、パシフィコ横浜 会議センター、イーブンニングフォーラム、口演
- (4) Yayoi Natsume-Kitatani、Ken-Ichi Aisaki、Satoshi Kitajima、Jun Kanno : Comparative analysis of gene expression profiles induced by chemicals with the same target molecule、ISMB/ECCB 2023

## 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