

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4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、毒性の分子機序に基づいて、現行の不確実係数（安全係数）を利用する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して、マイクロアレイ（GeneChip）と次世代シーケンサを用いて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、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る。

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**の蓄積によりプロトタイプを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、反映することにより、解析精度を向上させる。単回曝露のデータベースと共にこれを利用することで、現在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・効率化を検討する。

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る。

令和4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、遺伝子発現データとエピゲノムデータの連携解析ユニットの開発、ゲノムデータ等の表示・解析モジュールのデザインを進めつつ、並行して変動遺伝子の自動抽出パイプラインの改良を実施した。

(*) 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に行う。

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る。

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となる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る。

B. 研究方法

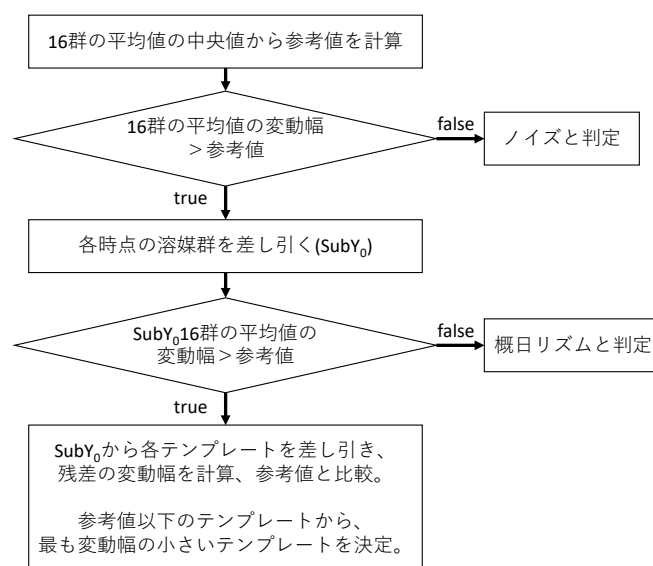
ソフトウェアの in house 開発に際しては、先行研究で開発したソフトウェアの改良の際は開発効率と生成する実行バイナリの実行速度を重視して、Win32/64 開発及び Web アプリケーション開発は RAD (Rapid Application Development) 対応の Delphi (Object Pascal 言語、USA, Embarcadero Technologies, Inc.) を用いた。データベースエンジンには組込型の DBISAM (USA, Elevate Software, Inc.) を、一般的なグラフ描画には TeeChart (Spain, Steema 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 (ver.3.9.1)を用いた。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.3.2.2) を用いた。

C. 研究結果

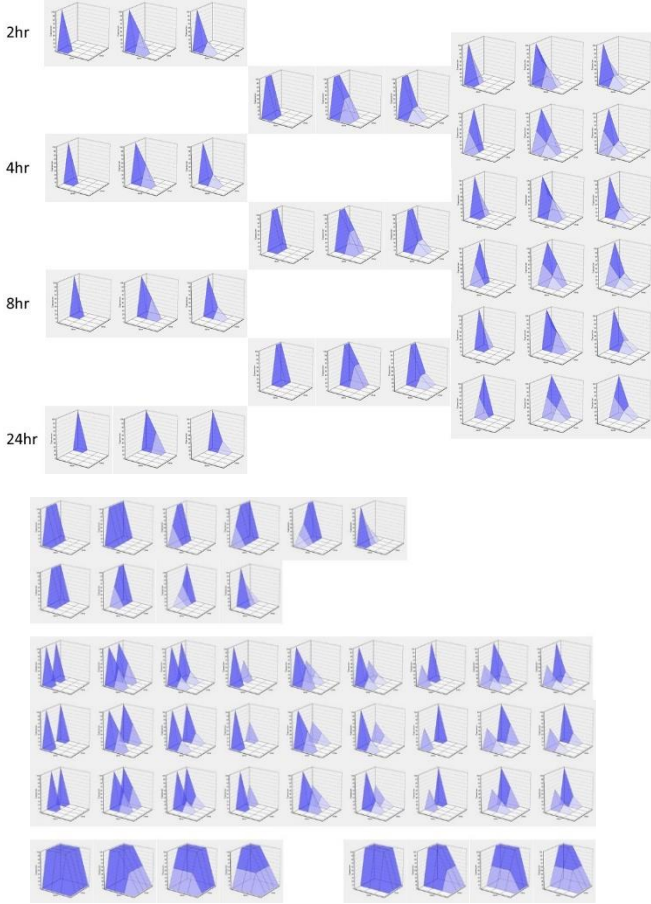
令和 4 年度も最終目標である新型化学物質有害性評価システムの解析コアの主要ユニットとしての、遺伝子発現とエピゲノムのデータ解析ユニットの開発を進めた。研究計画初期のデザインのままでは、解析工程のボトルネック、即ち、化学物質曝露による発現変動遺伝子の抽出作業の効率化が不十分である可能性が浮上したため、新たに遺伝子発現の 3 次元パターン (Surface) の特徴分類アルゴリズムの開発を進めた。

具体的な効率化の内容としては、ノイズパターン除去の強化と、発現パターンからの自動時系列分類である。基本的なアルゴリズムは、①変動幅が参考値以下の Surface のノイズ判定、②溶媒群と投与群の比較による概日リズムパターンの自動判別、及び③時系列パターンや飽和パターンなどの「テンプレート」との比較による自動 (時系列) 分類、から構成される (下図)。



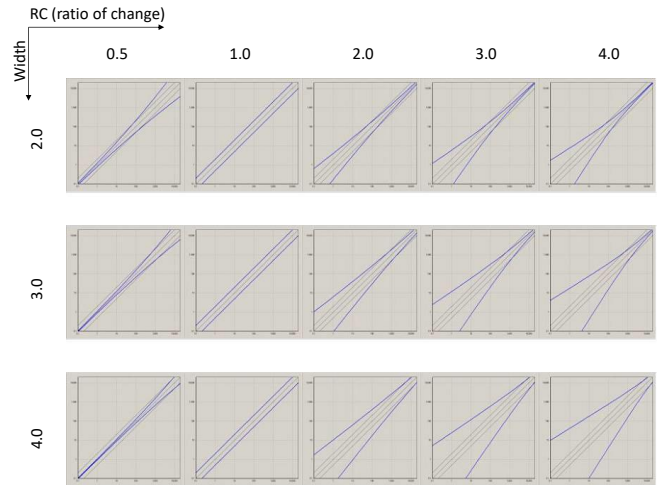
4 用量×4 時点の Surface の基本解析フロー

テンプレートパターン



想定されるテンプレートを用意することにより、発現パターンの自動分類を実現する。ピーク数 1~2、ピーク時点、ピーク幅、立ち上がり方の組み合わせと、飽和パターンを合わせたテンプレート基本セットを示す。

このとき参考値の算出は、単純な fold change ではなく、先行研究で考案した empirical な曲線（実際の Percellome 法により絶対量化した発現値の分散からパラメータを決定する指数関数（下図、Fx 曲線と仮称）を用いた。



対数グラフ上で、パラメータによって曲率に変化する Fx 曲線

例えば、下図のような 8 時間ピークの Surface (赤色で表示)については、概日リズムパターンではない、有意な変動幅を呈するピークであり、8 時間高用量ピークに分類された。

有意な変動を示すパターンの判定例

| ProbeID | Au-Av | Au-SGP | Au-Max | Au-Min | Au-MedA | Au-Dif | TC | S | TTC | QS |
|----------------|----------|----------|----------|----------|----------|----------|-------|---------|-----|----|
| 1483337 | 45.6128 | 13.5463 | 134.8516 | 53.3056 | 58.3227 | 161.5497 | TRUE | 51.2577 | | |
| vehicle | 39.0062 | 4.07623 | 48.0762 | 36.5169 | 36.8623 | 9.50117 | | | | |
| SubY0 | 6.52177 | 21.5063 | 68.7758 | -3.26221 | 0 | 62.0390 | TRUE | | | |
| SubY0-Temporal | Au-Av | Au-SGP | Au-Max | Au-Min | Au-MedA | Au-Dif | | | | |
| SubY0-T01 | 0.03881 | 17.7116 | 61.8542 | -62.223 | 3.07616 | 184.0778 | TRUE | | | |
| SubY0-T02 | 0.07766 | 34.4238 | 61.8542 | -62.223 | 3.07616 | 184.0778 | TRUE | | | |
| SubY0-T03 | 2.42773 | 131.2424 | 61.8542 | -62.223 | 3.07616 | 184.0778 | TRUE | | | |
| SubY0-T04 | 0.053881 | 16.02293 | 61.8542 | -60.4879 | 3.07616 | 172.3427 | TRUE | | | |
| SubY0-T05 | 0.07766 | 11.82527 | 61.8542 | -60.4879 | 3.07616 | 172.3427 | TRUE | | | |
| SubY0-T06 | 2.42773 | 137.3053 | 61.8542 | -60.4879 | 3.07616 | 172.3427 | TRUE | | | |
| SubY0-T07 | 0.053881 | 6.53002 | 12.10586 | -0.18408 | 3.07616 | 12.0969 | FALSE | | | |
| SubY0-T08 | 0.07766 | 11.69423 | 12.10586 | -42.1722 | 3.07616 | 54.32061 | TRUE | | | |
| SubY0-T09 | 2.42773 | 140.0439 | 12.10586 | -18.1624 | 3.07616 | 91.2026 | FALSE | | | |
| SubY0-T10 | 0.053881 | 12.48197 | 61.8542 | -60.7934 | 3.07616 | 182.5483 | TRUE | | | |
| SubY0-T11 | 0.07766 | 34.453 | 61.8542 | -60.7934 | 3.07616 | 182.5483 | TRUE | | | |
| SubY0-T12 | 2.42773 | 130.247 | 61.8542 | -60.7934 | 3.07616 | 182.5483 | TRUE | | | |
| SubY0-T13 | -1.68859 | 38.6203 | 61.8542 | -62.223 | 3.07616 | 184.0778 | TRUE | | | |
| SubY0-T14 | -7.60208 | 39.8129 | 61.8542 | -62.223 | 2.78491 | 184.0778 | TRUE | | | |
| SubY0-T15 | -4.71477 | 38.39627 | 61.8542 | -62.223 | 2.78491 | 184.0778 | TRUE | | | |
| SubY0-T16 | -1.68859 | 13.98785 | 12.10586 | -62.223 | 3.07616 | 104.3798 | TRUE | | | |
| SubY0-T17 | -7.60208 | 35.64962 | 12.10586 | -62.223 | 2.78491 | 184.3798 | TRUE | | | |
| SubY0-T18 | -4.71477 | 23.96426 | 12.10586 | -62.223 | 2.78491 | 104.3798 | TRUE | | | |
| SubY0-T19 | -1.68859 | 18.07383 | 61.8542 | -62.223 | 3.07616 | 184.0778 | TRUE | | | |
| SubY0-T20 | -7.60208 | 11.84278 | 61.8542 | -62.223 | 3.07616 | 184.0778 | TRUE | | | |
| SubY0-T21 | -4.71477 | 44.29712 | 61.8542 | -62.223 | 3.07616 | 184.0778 | TRUE | | | |
| SubY0-T22 | -1.68859 | 20.56447 | 12.10586 | -60.4879 | 3.07616 | 60.6378 | TRUE | | | |
| SubY0-T23 | -7.60208 | 21.98422 | 6.00956 | -60.4879 | 2.104624 | 87.00380 | TRUE | | | |
| SubY0-T24 | -4.71477 | 10.55783 | 6.00956 | -60.4879 | 2.104624 | 87.00380 | TRUE | | | |
| SubY0-T25 | -1.68859 | 18.24763 | 61.8542 | -60.7934 | 3.07616 | 182.5483 | TRUE | | | |
| SubY0-T26 | -7.60208 | 18.85887 | 61.8542 | -60.7934 | 3.07616 | 182.5483 | TRUE | | | |
| SubY0-T27 | -4.71477 | 38.81112 | 61.8542 | -60.7934 | 3.07616 | 182.5483 | TRUE | | | |
| SubY0-T28 | -1.68859 | 22.243 | 12.10586 | -60.7934 | 3.07616 | 182.5483 | TRUE | | | |
| SubY0-T29 | -7.60208 | 24.86488 | 12.10586 | -60.7934 | 3.07616 | 182.5483 | TRUE | | | |
| SubY0-T30 | -4.71477 | 13.82759 | 12.10586 | -60.7934 | 3.07616 | 182.5483 | TRUE | | | |
| SubY0-T31 | 0.07766 | 13.91136 | 61.8542 | -62.223 | 3.07616 | 184.0778 | TRUE | | | |
| SubY0-T32 | -3.53966 | 31.7719 | 61.8542 | -62.223 | 2.78491 | 184.0778 | TRUE | | | |
| SubY0-T33 | -1.40059 | 24.02425 | 61.8542 | -62.223 | 2.78491 | 184.0778 | TRUE | | | |

D. 考察

統合ツール“Percellome Integrator”の開発準備はほぼ計画通りに進行している。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、ライブラリの組み込みテストを進めている。

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開発については、期待されたノイズ除去と自動(時系列)分類の性能を満たしつつある。ノイズ除去が強すぎる傾向があるため、参考値の算出に利用したFx曲線の改良を進めており、現在のところ、パラメータを1つ追加することで、曲率調整の柔軟性を向上させ、過度の足切りが起こらない様にする事に成功している。

現在使用しているRSort(Surfaceの凹凸を解析し、有意な変動を示す候補遺伝子を数的に抽出する)とDTox(北野分担研究者らによるAI。熟練研究者の判定を学習、再現する)で処理した後、研究者による最終的なヒューマンキュレーションの前に、本アルゴリズムの実装フィルターを挿入することで、データ解析工程の大幅な効率化が期待できる。

E. 結論

遺伝子発現データとエピゲノムデータの連携解析ユニットの開発、ゲノムデータ等の表示・解析モジュールのデザインを進めつつ、並行して変動遺伝子の自動抽出パイプラインの改良を行い、統合ツールがより機能的・実用的なものになることが期待される。

F. 研究発表

1. 論文発表

① Aisaki KI, Ono R, Kanno J, Kitajima S. [Percellome Project: research on molecular mechanisms of toxicological responses based on transcriptomics and epigenetics]. Nihon Yakurigaku Zasshi. 2022;157(3):200-206. [doi.org/10.1254/fpj.21122] (Japanese).

(2) Yuhji Taquahashi, Shuji Tsuruoka, Koichi Morita, Masaki Tsuji, Kousuke Suga, Ken-ich Aisaki, Satoshi Kitajima: A novel high-purity carbon-nanotube yarn electrode used to obtain biopotential measurements in small animals: flexible, wearable, less invasive, and gel-free operation. Fundam. Toxicol. Sci. 2022; 9: 17-21. [doi.org/10.2131/fts.9.17]

2. 学会発表

(1) 相崎健一、小野竜一、菅野純、北嶋聡: Percellome プロジェクト ～トランスクリプトミクスとエピジェネティクス、インフォマティクスによる毒性分子機序の探求～、第96回日本薬理学会年会、(2022.12.2)、横浜

② J. Kanno, K.-I. Aisaki, R. Ono, S. Kitajima: Histone Modification, DNA Methylation, and mRNA Expression Analysis of Murine Liver Repeatedly Exposure to a Chemical. The XVITH International Congress of Toxicology (ICT2022), (2022.9.19), Maastricht, The Netherlands Oral.

③ 菅野純、相崎健一、小野竜一、北嶋聡 Percellome project からみた毒性 AI の展望. 第49回日本毒性学会学術年会、(2022.7.2)、札幌コンベンションセンター、シンポジウム、口演。

④ 菅野純、相崎健一、小野竜一、北嶋 聡 新型反復曝露トランスクリプトミクスから見た発癌エピソードジェネティクスの考察. 第 49 回日本毒性学会学術年会、(2022.6.30)、札幌コンベンションセンター、シンポジウム、口演.

(5) 高橋祐次、鶴岡秀志、大久保佑亮、相崎健一、北嶋 聡：バイタルサインの統合的評価による急性毒性試験の致死性予測 第 49 回日本毒性学会学術年会 (2022.6.30)

(6) 小野竜一、山本 雄介、成瀬 美衣、田邊 思帆里、吉岡 祐亮、相崎 健一、広瀬 明彦、落谷 孝広、平林 容子、北嶋 聡 cfDNA による毒性評価 第 49 回日本毒性学会学術年会 (2022.6.29.) 札幌コンベンションセンター、シンポジウム、口演.

⑦ J. Kanno, K. Aisaki, R. Ono, and S. Kitajima: Comprehensive Histone, DNA Methylation, and mRNA Expression Analysis of Murine Liver Repeatedly Exposed to Chemicals: Percellome Project 2022 Update. Society of Toxicology (SOT) 61st Annual Meeting (SOT2022), (2022.3.30), San Diego, USA, Poster.

⑧ Natsume-Kitatani Y; Aisaki K; Kitajima S; Kanno J「Comparative study of dynamic changes in gene expression profiles induced by PPARα ligands」 ECCB2022 2022/9/18 スペイン

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