

厚生労働科学研究費補助金（化学物質リスク研究事業）  
分担研究報告書

化学物質の有害性評価手法の迅速化、高度化に関する研究  
- 網羅的定量的大規模トキシコゲノミクスデータベースの維持・拡充と  
毒性予測評価システムの実用化のためのインフォマティクス技術開発 -

分担研究課題：「Percellome 3次元データ等のための専用解析ソフトウェアの開発研究」

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・第一室長

研究要旨

先行研究に於いて遺伝子発現データの絶対量化技術である Percellome 手法を実用化し、既に 100 余の化学物質による遺伝子発現変動情報を収録して現在もさらに拡張しつつある網羅的遺伝子発現情報データベース（Percellome データベース）を構築している。この高精度大規模データベースの有効活用には、時間×暴露用量×遺伝子発現量からなる遺伝子発現変動情報を生物学者によるデータ把握が容易な 3次元波動面（Surface グラフ）に変換して解析する直感的で効率のよい手法（”Millefeuille”（ミルフィーユ、MF）データ処理）を採用している。本分担研究では、Percellome 手法や Surface グラフの特徴を活かし、遺伝子発現反応の網羅的解析を効率良く行うためのプログラムの開発・改良を行い、トキシコゲノミクス技術を利用した毒性予測・評価システムの実用化を推進する。

平成 24 年度は、Percellome 法の適用を介して、医薬基盤研・トキシコゲノミクスプロジェクト（TGP）のラットデータを、マウスデータからなる Percellome データベースに結合し、統合トキシコゲノミクスデータベースの構築を進めた。また非 Percellome データの絶対量推定ソフトウェア SnCalc.exe を開発した。さらに Percellome データベースシステムの WebAPI を開発し、ライフサイエンス研究用ソフトウェアの国際共通プラットフォーム GARUDA プロジェクト等、外部の学術・行政システムからのオンラインアクセスを可能にする本格的な一般公開の準備を整えた。

平成 25 年度は各班員のデータ解析の効率と精度の向上を優先して平成 26 年度予定の計画を実施し、候補遺伝子抽出プログラム RSort に新規フィルタールーチンを追加して自動抽出精度を向上させたほか、経時変化データ等の線グラフ表示データを対象とした発現変動起点の自動抽出アルゴリズムの新規開発や、前年度に開発

した SnCalc.exe のラットデータ対応のためのリファレンスデータベースの構築、及び Percellome データベース一般公開システムの WebAPI の機能拡張を実行した。

#### A. 研究目的

本研究班が化学物質による生体影響の分子メカニズムに依拠した毒性評価手法の実用化の為にインフォマティクス開発を目指す中において特に、100余の化学物質による遺伝子発現変動情報を絶対量化して収録している高精度大規模遺伝子発現情報データベース（Percellome データベース）を活用し、遺伝子発現反応の網羅的解析を効率良く行うためのプログラムの開発・改良を行って、トキシコゲノミクス技術を利用した毒性予測・評価システムの実用化を推進する。

#### B. 研究方法

##### B-1: 参照データ

a) 遺伝子発現データ：Percellome プロジェクトで生成したマウスの遺伝子発現データ、および TGP プロジェクトで生成したラットの遺伝子発現データを用いた。データ標準化には Percellome 法を利用した。

b) 遺伝子のアノテーション情報：米 Affymetrix 社および Gene Ontology Consortium、独 BIOBASE 等が提供している情報を参照した。

c) 原著論文（J Toxicol Sci. 38(4):643-54）等のヒューマンキュレーションを経た候補遺伝子リスト。

##### B-2: ソフトウェア作成に用いた開発言語及びコンポーネント

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 開発は RAD（Rapid Application Development）対応の

Delphi ver.7 もしくは XE3（Object Pascal 言語、USA, Embarcadero Technologies, Inc.）を用いた。Web アプリケーション開発には Delphi もしくは PHP、JavaScript を用いた。1 億件以下の小～中規模のデータベース処理に際しては組み込み型リレーショナルデータベースコンポーネントの 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。また WebAPI の構築には DataSnap（USA, Embarcadero Technologies, Inc.）を用いた。

##### B-3: データ解析計算

主たる計算は独自に開発したプログラムにより実施した。検証は必要に応じて Excel（USA Microsoft Corporation）や R 言語（オープンソース R Development Core Team）で実施し、浮動小数点誤差以上の乖離がないことを確認した。

大規模計算が必要な場合は、高速データベースエンジンである Teradata データベース（日本 Teradata 社）を装備する研究計算用サーバーシステム（MF サーバーシステム）にアプリケーションソフトウェアを移植して実施した。

#### C. 研究結果

平成24年度は、化学物質評価用の遺伝子発現データベースの有用性を高めるために、医薬基盤研トキシコゲノミクスプロジェクト（TGP）のラットデータをマウスの遺伝子発現データからなる

Percellomeデータベースに統合すべく、マウスとラットのデータをシームレスに一括解析する手法（マウス・ラットの相同遺伝子対を一意に示す統合IDとその効率的な処理アルゴリズム）を生成し、TGPデータをPercellome変換（絶対量変換）した上で、Percellomeデータベースとの統合を進めた。

またデータベース及び解析ツールの公開に向け、Percellome化せず取得した外部のマイクロアレイデータとPercellome絶対量化データとの比較を可能にする絶対量推定ソフトウェアSnCalc.exeを開発したほか、外部研究システムから直接、Percellomeデータベースを利用するためのWebAPIを作成し、運用を開始した。

平成25年度は、各班員のデータ解析の効率と精度の向上を優先して平成26年度予定であった計画を実施し、候補遺伝子抽出プログラムRSortに新規フィルター機能を追加して自動抽出精度向上の改良をお加えたほか、微分解析による遺伝子変動起点の自動抽出アルゴリズムの新規作成を中心に、前年度に作成したSnCalc.exeのラットデータ対応のためのリファレンスデータベースの構築やPercellomeデータベース一般公開システムのWebAPIの機能拡張を進めた。詳細は以下の通り。

#### C-1：RSortの改良

RSortは、2条件（主に用量、時間）の組み合わせにおける遺伝子発現量を3次元平面で表現した遺伝子発現変動パターンの凹凸を評価し、全遺伝子に対応する総数4~5万個の発現データから、ノイズや解析不能

パターンを除去して、候補遺伝子リスト（通常、総数の1/10程度）を短時間（1~2分）で自動抽出し、データ解析の作業効率を大幅に高めている。ただし、毒性評価上の必要から偽陰性を極力排除するよう最適化されていることもあり、従来版（ver.0.17）では偽陽性が多発し、自動抽出結果の2/3~3/4を占めていたため、一層の改良が必要であった。

そこで偽陽性データの混入を減らすべく、既存のroughnessフィルタと独立した追加フィルタの作成を検討し、3次元の遺伝子発現変動パターンを構成する各2次元切片を対象として、増減パターン組み合わせ数や変曲点数、連続性指標、グラフ密度などを検討して、最適条件を求めた。

その結果、偽陰性データの発生数を数個~200個程度に抑えつつ、従来版RSortの抽出結果から数百~2000個程度の偽陽性データを削減することに成功した。

#### C-2：発現変動起点の自動抽出アルゴリズムの開発

RSortは、時間と用量と発現量の3次元表示データを効率的に処理すべく開発されてきたものなので、経時変化データ等の線グラフ表示データの処理については精度が低下する傾向があった。そこで、分担研究(2)が対象とする個体発生の経時変動データをサンプルとし、線グラフ表示データからの効率的な候補遺伝子抽出アルゴリズムを作成した。

特に、データ点数の多い線グラフデータを解析する場合、遺伝子誘導の結果であるピークに着目して線グラフの凹凸を評価するより、遺伝子発現の起点（発現変動起点）

に着目したほうが、シグナルネットワーク描出のために有用である。

そこで、各群の平均値をなめらかに繋ぐ 3 次スプライン曲線を求め、その一階微分曲線と二階微分曲線、三階微分曲線を算出して、これらを元に発現変動起点を推定するアルゴリズムを作成した。

実証のためのプログラムを作成し、性能試験を行ったところ、分担研究(2)の成果で詳細を記載したとおり、既知のシグナルネットワークに対応する候補遺伝子リストを適切に抽出する事に成功した。

#### C-3:SnCalc.exe のラットデータ対応のためのリファレンスデータベースの構築

平成 24 年度に開発した非 Percellome データの絶対量推計ソフトウェア SnCalc.exe は、当初、マウスデータを対象としていたが、ラットやヒトなど、他の動物種由来のデータにも対応できるよう改良を加えた。実際に絶対量推定計算を行うためには、リファレンスとなる、各動物種の Percellome データが一定量以上必要だが、ここでは医薬基盤研・トキシコゲノミクスプロジェクト(TGP)のラットデータを利用して、リファレンスデータベースを構築した。

#### C-4:Percellome データベース一般公開用 WebAPI の機能拡張

初年度に Percellome 公開用 Web サーバーに実装した REST (現時点で最も普及しているソフトウェアアーキテクチャの 1 つ) インターフェイスについては、当該システムを利用したオンライン解析能力を向上させるべく、機能を拡張し、RSort で抽出した候補遺伝子リストをインターネット経由

で提供するなどの機能強化を行った。

#### D. 考察

Percellome 3 次元データ等の為の専用解析ソフトウェアの開発研究においては、本年度、自動的にノイズデータを除外し、研究者が操作する候補遺伝子数を減らすことで解析効率の向上に大きな役割を果たしている RSort の性能を一層向上させた。今までは、生物学的な意味のある遺伝子発現データを除外してしまわないように、若干の余裕をもってノイズデータを除去していたことから、研究者が操作する段階で、若干のノイズデータが残存していた。平成 25 年度は、この残存ノイズデータを更に高精度に除去する改良を加えた。この改良は、個々のデータの網羅的遺伝子解析作業の効率を高めるのみならず、複数の候補遺伝子リストの和集合計算を行う必要性が生じる大規模解析、例えば複数の化学物質毎に特異的に発現する遺伝子群の比較解析時に、その効果が明瞭に現れる重要な技術である。

また今までの時間と用量と発現量の 3 次元表示データを中心に開発してきた解析ソフトウェアに加え、経時変化データ等の線グラフ表示データに特化した候補遺伝子抽出アルゴリズムによるソフトウェアを用意することが出来た。これにより 3 次元表示データに匹敵する解析精度や作業効率の向上が見込まれ、この一部の技術は 3 次元表示データ解析にも還元可能であり、相互に統合的解析が可能となる。

#### E. 結論

Percellome 3 次元データ等の為の専用解析ソフトウェアの開発研究では、平成 24 年

度成果の異種動物由来のデータ統合技術や絶対量化されていない非 Percellome データの絶対量推定技術、及び今年度成果の RSort 技術の改良により、網羅的遺伝子発現解析の効率向上を押し進めている。

また平成 24 年度から公開している Percellome データベース一般公開システムの WebAPI の改良により、外部のオンライン解析ソフトウェアからも、RSort による候補遺伝子情報にアクセス出来るようになり、解析効率の向上が期待される。

## F. 研究発表

### (1) 論文発表

Kanno J, Aisaki K, Igarashi K, Kitajima S, Matsuda N, Morita K, Tsuji M, Moriyama N, Furukawa Y, Otsuka M, Tachihara E, Nakatsu N, Kodama Y. (2013) Oral administration of pentachlorophenol induces interferon signaling mRNAs in C57BL/6 male mouse liver. *J Toxicol Sci.* 38(4):643-54.

Abe S, Kurata M, Suzuki S, Yamamoto K, Aisaki K, Kanno J, Kitagawa M.(2012) Minichromosome maintenance 2 bound with retroviral Gp70 is localized to cytoplasm and enhances DNA-damage-induced apoptosis. *PLoS One.* 7(6):e40129.

### (2) 学会発表

北嶋 聡,小川幸男, 大西 誠 ,相磯成敏,相崎健一,五十嵐勝秀, 高橋祐次, 菅野 純  
シックハウス症候群レベルの極低濃度暴露の際の海馬における Percellome 法に

よる吸入トキシコゲノミクス 第 40 回  
日本毒性学会学術年会(2013.6.18)

Kitajima S, Aisaki K, Igarashi K, Kanno J. Application of Percellome Toxicogenomics approach to food safety: A flavor, estragole appears to be a PPAR-alpha agonist. The XIII International Congress of Toxicology 2013 (ICT 2013)] (2013.7.3.), Seoul, Korea

Tanemura K, Igarashi K, Furukawa Y, Otsuka M, Aisaki K, Kitajima S, Sato E, Kanno J. Delayed Effects on CNS Induced by Disturbance of Neural Activity during Development – Behavioral Impairment in Male Adult Mice Induced by Postnatal Oral Intake of Acephate. The XIII International Congress of Toxicology 2013 (ICT 2013)] (2013.7.3.), Seoul, Korea

## G. 知的財産所有権の出願・登録状況(予定も含む)

- 1 . 特許取得  
なし
- 2 . 実用新案登録  
なし
- 3 . その他  
なし