

厚生労働科学研究費補助金（地球規模保健課題推進研究事業）
化粧品等の QSAR/in silico/インフォマテクス技術等の安全性評価応用
に関する調査研究
平成 25 年度分担研究報告

- 毒性評価データベースに関する動向調査 -

研究分担者：水口 賢司（(独)医薬基盤研究所 バイオインフォマテクスプロジェクト
プロジェクトリーダー）

研究要旨：インシリコでの毒性予測や毒性発現メカニズムの解明には、化学構造に加えて化合物暴露に対する遺伝子発現情報などの利用が有効であると考えられている。そのようなトキシコゲノミクス研究の進展には、大規模な遺伝子発現情報をコンピュータ解析可能な形で如何に整理して、他のデータと統合するかが鍵になる。本研究では、毒性評価の基礎となるデータベースの現状を調査し、毒性メカニズムのモデリングに向けた将来への展望を議論した。

キーワード：トキシコゲノミクス、遺伝子発現情報、データ統合、モデリング

A. 研究目的

化学構造と活性との関係をモデル化する定量的構造活性相関 (quantitative structure-activity relationship; QSAR) は、結合親和性などの比較的単純なエンドポイントに対しては有効で、幅広く用いられているが、肝毒性などへの適用には限界がある。実際、Low らは (Low *et al.*, Chem. Res. Toxicol. 24:1251-1262, 2011)、後述のトキシコゲノミクスプロジェクトによるデータを解析し、QSAR による化合物の肝毒性予測では、限られた精度しか達成できないことを報告している。そのため、化学構造以外の情報、特に実験的に取得した遺伝子発現情報を利用して毒性発現メカニズムの解明や毒性予測を目指すトキシコゲノミクスに期待が持たれている。遺伝子発現情報などの大規模データから経験則を抽出して現象をモデリングする試みは、工学、

医学、薬学を含む幅広い分野でさかんに研究が進められている。そのような研究の前提として、現象に関連するデータが、コンピュータ解析可能な形で整理されている、すなわちデータベースが整備されていることが重要である。そこで本研究では、毒性評価の基礎となるデータベースの現状を調査し、その展望と課題を明確化することを目的にした。

B. 研究方法

関連データベースは、文献検索やインターネット上の検索により調査した。我々自身が開発した、トキシコゲノミクス統合解析プラットフォーム Toxygates は、後述の Open TG-GATEs による公開データを元にして、セマンティックウェブ技術と key-value ストアと呼ばれるデータベース技術を用いて構築した（詳しくは、

Nystrom-Persson *et al.*, *Bioinformatics*, 29:3080-3086, 2013)。これらのデータベースをウェブ上の操作により比較、検討した。

C. 研究結果

(1) 毒性評価に関連する既存データベース

官民共同研究としての日本のトキシコゲノミクスプロジェクト(以下、TGP と呼ぶ。 <http://www.tgp.nibio.go.jp/index.html>) は、医薬品などの化合物をラット個体や細胞に暴露した際の毒性情報と遺伝子発現情報を網羅的に収集することで、創薬研究早期での毒性発現メカニズムの解明や毒性予測を目指したものである。Open TG-GATEs というデータベース名で、遺伝子発現データ、病理所見と高解像度病理画像、生化学データが公開されている (<http://toxico.nibio.go.jp>)。

一方、海外の関連するデータベースとしてまず、Comparative Toxicogenomics Database をあげることができる (<http://ctdbase.org>; Davis *et al.*, *Nucleic Acids Res.*, 41, D1104–D1114, 2013)。このデータベースは、文献情報のキュレーションにより、医薬品と遺伝子や疾患との関係性をまとめたものだが、実際の遺伝子発現データについては提供されていない。遺伝子発現データの大規模データベースとしては、Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>; Barrett *et al.*, *Nucleic Acids Res.*, 39, D1005–D1010, 2011) や ArrayExpress (<https://www.ebi.ac.uk/arrayexpress/>; Rustici *et al.*, *Nucleic Acids Res.*, 41, D987–D990, 2013) が存在する。これらには、化合物暴露に対する遺伝子発現の変

化についての実験データが多数格納されており、トキシコゲノミクス研究に大いに関連するデータベースと言える。但し、これらのデータベースは毒性学に特化したものではなく、遺伝子発現データ一般についてのレポジトリであるため、測定プラットフォーム、対象となる生物種、化合物の種類や投与方法、その他実験プロトコルについて、多種多様なデータが混在している。従って、毒性評価のためのデータ解析には、やや使いづらい面もある。DrugMatrix (<https://ntp.niehs.nih.gov/drugmatrix/index.html>; Ganter, *et al.*, *J. Biotechnol.*, 119, 219–244, 2005) は、600 以上の化合物をラット個体に投与した際の遺伝子発現データをマイクロアレイにより収集したものである。上記の TGP データは、化合物数としては DrugMatrix に及ばないが、統一された実験デザインに基づいて、より多くの投与量や時点でのデータがあり、系統的なトキシコゲノミクス解析に有利な点を持っている。

The Connectivity Map(以下、cmap と呼ぶ。 <http://www.broadinstitute.org/cmap/>; Lamb *et al.*, *Science*, 313, 1929–1935, 2006) は、生物活性を持つ化合物をヒト培養細胞に暴露した際の遺伝子発現情報を網羅的に収集して公開している。後述する、我々の開発した Toxygates と同様、遺伝子発現パターンの類似度に応じて化合物をランキングするシステムを提供している。但し、cmap では、複数のプラットフォームを用いた測定がなされているため、コルモゴロフスミルノフ検定によって化合物の順位付けを行なっているが、Toxygates の場合は、元になる TGP データが単一プラットフォームを採用しているため、より直接的な発現データ間の相関係数を用いたランキング

が可能になっている。

ToxBank (<http://toxbank.net> ; Kohonen *et al.*, *Mol. Inform.*, 32, 47–63, 2013)は、毒性評価のための、より一般的なデータ統合プラットフォームを目指している。遺伝子発現データ解析に限らず、毒性学研究一般についてのデータが提供されている。

(2)トキシコゲノミクス統合データ解析プラットフォームとしての Toxygates

上で述べたように、TGP によるトキシコゲノミクスデータは、統一したプラットフォームとプロトコルを特徴とし、毒性評価の基礎データとして極めて重要なものである。但し、Open TG-GATEs データベースでは、マイクロアレイの生データのダウンロードを可能にしているだけで、データの統合や解析という機能は提供されていない。そこで我々は、セマンティックウェブと呼ばれる技術を用いて、Open TG-GATEs と KEGG データベースによるパスウェイ情報 (<http://www.genome.jp/kegg/pathway.html>)、Gene Ontology 機能注釈情報 (<http://www.geneontology.org/>)、ChEMBL データベースからの化合物—ターゲット情報 (<https://www.ebi.ac.uk/chembl/>) 等の外部データとを統合し、化合物投与に反応する遺伝子の同定と絞り込みが可能なシステムを構築した。Toxygates と名付けたこのシステムは、トキシコゲノミクスデータ統合解析プラットフォームと位置づけることができる (<http://toxygates.nibio.go.jp>; Nystrom-Persson *et al.*, *Bioinformatics*, 29:3080-3086, 2013)。Toxygates を用いることで、特定の化合物投与後にどのよう

な遺伝子が発現変動したかを調べ、それと似たような反応を示す他の化合物をランキングすることが可能になった(図 1)。現在、このデータベースシステムを拡張して、アジュバント (免疫賦活剤) の有効性と安全性の指標となるバイオマーカー探索の基礎となるデータベースの構築を進めている (<http://adjuvantdb.nibio.go.jp>)。

D. 考察

本研究で調査したデータベースの中でも、TGP データ (Open TG-GATEs および Toxygates) は、統一したプラットフォームとプロトコルに基づく動物個体に対する大規模なデータとして貴重なものと言える。また、同じ化合物を細胞に暴露した際の遺伝子発現情報についても収集されているので、個体レベルと細胞レベルとの架け橋となる可能性も有している。但し、単純に発現量変動遺伝子を比較するという解析では、ラット個体と細胞でのデータには大きな差があり、両者を結びつけることは難しく思われる。化合物作用と遺伝子発現というエンドポイントの間には、シグナル伝達や転写制御など様々なプロセスが関与しており、それらに関して何らかのモデル化を試みることで、より抽象的なレベルで個体レベルと細胞レベルのデータを関連付けることが必要であろう。複雑なシグナル伝達プロセスのモデル化は容易ではないが、そのネットワーク構成要素に関しては、タンパク質間相互作用、化合物—タンパク質相互作用、転写因子—標的遺伝子相互作用などについて、多くのデータが公共データベース上に蓄積されつつある。Open TG-GATEs による遺伝子発現情報と公共データベース上のパスウェイや化合物情

報を統合する Toxygates は、そのようなデータ統合に向けた試みの最初のステップと位置づけることができる。また、170 程度という化合物の数は、化学構造に基づく毒性評価という目的には極めて少なく、今後大規模に動物個体による実験データを追加していくことは困難であろうから、上記のインシリコ解析により個体と細胞とを関連付けるモデル化を行ない、それに基づいて細胞レベルでの実験をデザインして遂行するという戦略が有効ではないかと考えられる。

E. 結論

本研究では、毒性評価の基礎となるデータベースの現状を調査することで、今後のインシリコ毒性予測への期待とその実現に向けた課題を明らかにした。考察で述べたように、ドライの解析が主導する形で実験をデザインし、より一層のデータの蓄積と統合を実現できるかが、今後のこの分野の進展の鍵になると考えられる。そのためには、ウェットとドライ研究の緊密な連携が必須であろう。

F. 研究発表 学会発表

1. 水口 賢司：データ統合とネットワーク解析による創薬初期研究の支援、第3回シスメックスプロテインカンファレンス(2013-10-18、品川プリンスホテル)
2. 水口 賢司：データ統合とネットワーク解析による創薬支援、第9回霊長類医科学フォーラム(2013-11-14、文部科学省研究交流センター)
3. 水口 賢司：創薬の初期研究におけるデータ統合：ターゲットと安全性

の評価、第345回CBI学会研究講演会(2014-1-9、東京大学山上会館大会議室)

4. 水口 賢司：‘アジュバントゲノミクス’に向けた統合データベースの現状、第7回次世代アジュバント研究会(2014-1-21、千里ライフサイエンスセンター)
5. 水口 賢司：データベースは、創薬初期でのターゲット評価と安全性の予測に役立つか？、MEDALS 第三回データベース講習会(2014-1-24、産総研・関西センター)

G. 知的所有権の取得状況

無し

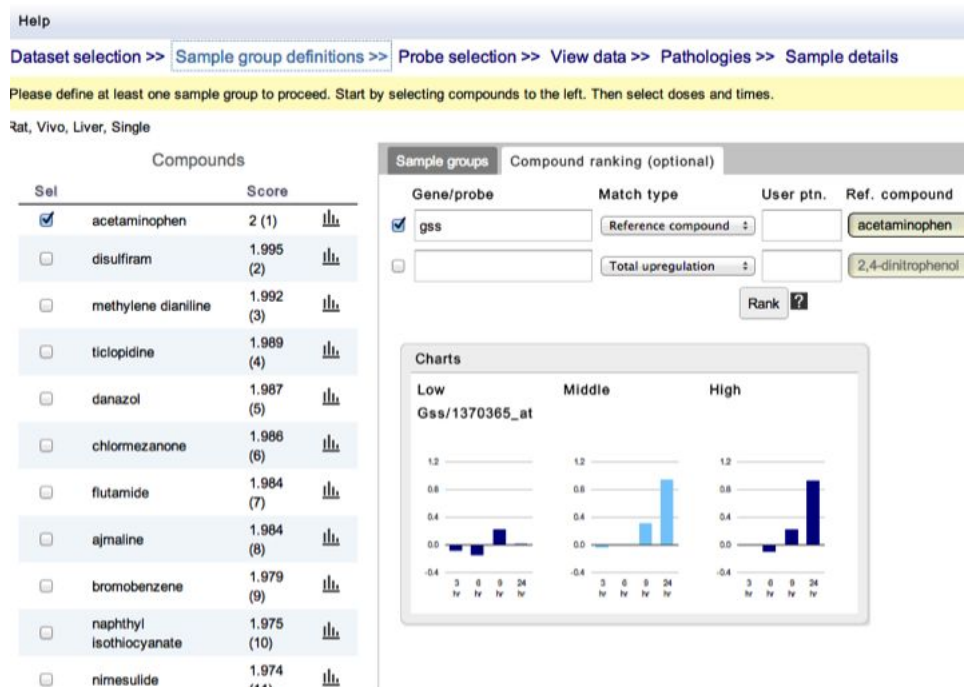


図1 トキシコゲノミクスデータ統合解析プラットフォーム Toxygates による、遺伝子発現パターンを用いた化合物ランキング