

Special Review

ゲノム毒性学：形質非依存型トキシコゲノミクスの導入

Toxicology in Genome Age: Introduction of Phenotype-Independent Toxicogenomics.

菅野 純、相崎健一、五十嵐勝秀、小野 敦、中津則之

Jun Kanno, Ken-ichi Aisaki, Katsuhide Igarashi, Atsushi Ono, Noriyuki Nakatsu

abstract :

形質非依存型トキシコゲノミクスに適用するため、マイクロアレーから細胞1個当たりの mRNA 絶対量を得る方法 (Percellome) を開発した。これにより遺伝子発現量を、ゼロを起点とする均等目盛りで表示し直接比較することが出来るようになった。そのため、今まで用いられてきた対照に対する比率表示と違って、割り算をする必要が無く、発現値ゼロの表示が自由に行え、対照群も処置群も同列に表示することが可能となった。また、更なる標準化操作が原則的に不必要なため、測定した全ての遺伝子についてマイクロアレー間はもとより、実験間での直接比較が行える。この特長は、生物学者が内容を直感的に把握し易い様なデータの可視化にも役立ち、その後のデータ解析とインフォマティクス形成を促進することが示されつつある。異なったプラットフォーム間でのデータ互換にも拡張可能であり、トキシコゲノミクスに必要な大型データベースやコンソーシアム構築にも貢献する可能性が高い。

Key Words :

マイクロアレイ技術, トキシコゲノミクス, 分子毒性学, 創薬支援, 化学物質安全性評価

著者略歴

かんの じゅん

菅野 純 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・毒性部長)

e-mail: kanno@nihs.go.jp

東京医科歯科大学大学院医学研究科博士課程終了、医学博士

1985 年東京医科歯科大学大学院医学研究科博士課程修了。人体病理学、実験病理学専攻。国立衛研毒性部室長を経て 2002 年より現職。内分泌かく乱関連等の分子毒性学研究、トキシコゲノミクスプロジェクト等を厚生労働所掌業務との有機的連携のもとに推進。

相崎健一、五十嵐健一、小野 敦、中津則之

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

はじめに：

生物界 (Biosphere) は化学物質界 (Chemosphere) との相互作用のなかで存在している。食べ物も薬も毒も、経口、吸入、あるいは皮下や血管内へ進入してきて生体分子と相互作用を起こす。この関係は、食物を選ぶ過程で例えば毒のあるものを避け、薬草を見出すなど太古の昔から存在しているが、近代生活に於いてその複雑さが急速に増してきた。体内への直接的な摂取を目的としたものに加え、生活の利便性ために開発・利用する物質の増加が国民の安全と安心

に係る問題として注目されている。しかし、それらの規制決定に係る毒性評価を生体側から見ると、身体に入るまでの「物質の分類」はもはや重要ではなく、むしろ身体に入ったあとにどのような反応が如何に惹起されるかが問題となる。

創薬の世界では、薬効のある物質を見つけ出すことが重要であり、次いで、その毒性が検討される。つい最近までは、「アマゾンに新しい植物を探しに行く」ことが主流であったが、近年は「何十万ものリード化合物のライブラリー」を充実させ、目的とする薬効を発揮する物質を、目的にあった方法でスクリーニングする方策が採られる。「薬」の毒性は、概念上、薬効の延長線上の毒性（例えば過剰インシュリンによる低血糖）と予期せぬ副作用（まさしく“side effect”）とがある。何れにせよ、「薬」については cost-benefit（費用便益・費用対効果）の概念が強く働き、多少の毒性があっても使用することが少なくないし、患者側からの要望があれば尚更である。

ここでは、cost-benefit の対象を人（ヒト）に限定して話を進める。人に於ける毒性を検討するためには、人からの情報が一番正確なことは言うまでも無い。薬の開発の過程では、「臨床試験」なる「人体実験」が可能である。もちろん、人に使っても薬効のもたらす利益よりも副作用たる毒性が十分に小さいであろうことを、各種の動物実験によって確かめてから、厳重な管理体制の元で、且つ本人の了解を得た上で「人体実験」に入るわけである。この場合の毒性には、用量作用関係の概念が乏しい。すなわち、実際に薬として投与するときの薬用量において、どのような毒性（副作用）が現れるかが、最大の焦点なのである。この貴重な人体実験でわざわざ「自殺目的」の大量投与を行うことは無い。また、薬効が期待できないような微量の投与も当然行わないわけである。

これに対して、いわゆる化学物質、たとえば、家庭用品、工業製品、食品添加物などの現代生活の利便性に欠かせない物質に由来する化学成分の体内への侵入に対しては、一般的に cost-benefit の概念が弱く働き、可能ならばゼロにしたいという傾向がある。しかし、「完全ゼロ」は使用する限り基本的には不可能であるので、どの位の量までなら安全と見做せるかを検討することが行われてきている。

これらの物質については、人体実験が倫理的にも現実的にも出来ないと考えるのが通常である（ボランティアを募ることが出来れば、それは可能かもしれないが、発がん性が疑われたり、蓄積性が高いもの、例えばダイオキシンや PCB のようなものは、いくらボランティアが名乗り出てくれても投与させてもらう気にはならないものである）。なお、薬でも「人体実験」が事実上出来ない対象がある。それは、胎児と子供である。いずれの場合も、現在のところ、人の身代わりとしてモデル動物を用いることになる。

毒性における量と質の問題：

では、どの位の量までなら安全と見做せるか。多量に摂取すれば毒性は強く、少量になれば毒性は弱まるという大原則（毒性は用量に関して単調増加する）の下では、「毒性に閾値がある」と考えられる場合と、「閾値が存在しない」と考えられる場合とで、扱いを分けている。前者の場合は無毒性量あるいは無作用量をラット等の実験動物で求め、種差や個体差を勘案した係数（不確実係数あるいは安全係数と呼ぶ）で除して、安全の目安となる基準値とする。後者の場合は、無毒性量の代わりに、俗に「運悪く雷に打たれて死ぬ確率」を目安とする実質安全量（virtually safe dose、通常 10^{-5} ないし 10^{-6} の危険率を適用）を採用し、同様の手続きを経て

ヒトへの外挿を行っている。これらの判断が正しいか否かを検討する材料としては人での中毒事例、自殺事例、事故事例やそれらに関する疫学調査が活用され、それに基づく基準設定法の修正が折にふれて加えられてきた歴史がある。(他方、化学物質の輸送や取り扱いに際した注意度を定めるために、毒物・劇物の指定が行われているが、これは、ラットなど単回暴露時のLD₅₀(半数致死量：動物の半数が14日以内に死亡する量)が低値のものを「危険度が高い」として規制するものである。)

それでは、毒性の質的な問題はどの様に取り扱われてきたか。生物学が現象の記述学に基礎を置いていた段階での毒性学は、創薬の場にしろ、一般的な化学物質の毒性評価の場にしろ、その要求される役割を果たすために、投与された化学物質と症状との関連性に基づいた化学物質の体系化を基盤として発達してきた。その過程での様々な経験を取り入れる形で、前述の「不確実係数」や「LD₅₀」の概念が利用され、現在まで、非常に有効に機能してきている。ここまでの毒性学は、化学物質の投与とそれによる症状発現(毒性)の関連性をブラックボックスを介して分類し体系化するものであり、回帰モデル(Regression model)の概念に根差した後向きの検討が行われることが多かった。しかし、サリドマイド禍(奇形発生)に代表されるようにげっ歯類の実験動物では毒性が確認されず、人に使用して初めて催奇形性が明らかになった事例の存在は、この方法の限界を示している。

近年、科学の進歩により、毒性学は生体内で引き起こされる反応の分子レベルから形態レベルまでメカニズム記述を基礎とするものへと変貌しつつある。ここで、活躍するのがハイスループット性の高いマイクロアレイ技術である。しかし、マイクロアレイから得られた遺伝子発現プロファイルによる検討も、依然としてブラックボックスが介在している場合は、その時に観測される毒性形質と関連付け、いわゆる化学物質のフィンガープリント(指紋)として毒性反応の類型化を行うことが多い。この様な関連付けを「phenotypic anchoring」と呼ぶことがある(1)。

形質非依存型トキシコゲノミクス (phenotype-independent toxicogenomics) :

これに対して、分子毒性学の立場から一番知りたいことは、生体内で実際に起こっている一連の事象であり、transcriptome の場合には全ての遺伝子の情報を元にした遺伝子カスケードの全容解明である(図1)。これが分かれば、膨大な時間と費用の掛かる長期毒性試験(ラット等を用いる)の代替として、より早く、安く、正確な評価、種差や個人差を勘案した正確なヒト毒性予測が可能となる事が強く期待される。特に胎児、新生児、小児、成人、老人の各発達段階に於ける生体側の反応様式・感受性の変化や、複数の物質の進入による複合作用なども包括的に扱える様になると考えられる。即ち、実験動物で得た所見を人に外挿する際に、実験動物のブラックボックスとヒトのブラックボックスを繋げる経験則が「不確実係数」であるが、これを責任遺伝子カスケードの解明によってバイパスする方策を得る事になる(図2)。これを実現させるためには、例えば、マウスに於いては遺伝子ノックアウト手法により遺伝子ごとの機能解析が可能であり、ヒトではSNPs解析が同様に利用できる。しかし、形質発現が伴わない場合には解析が行き詰まることが多い点で、これは形質発現に依存的な手法である。全ゲノムが明らかになった現在、この目的のためには形質発現の有無に関らず全ての遺伝子の発現をモニターすることを目的としたアプローチを考慮せざるを得ない。

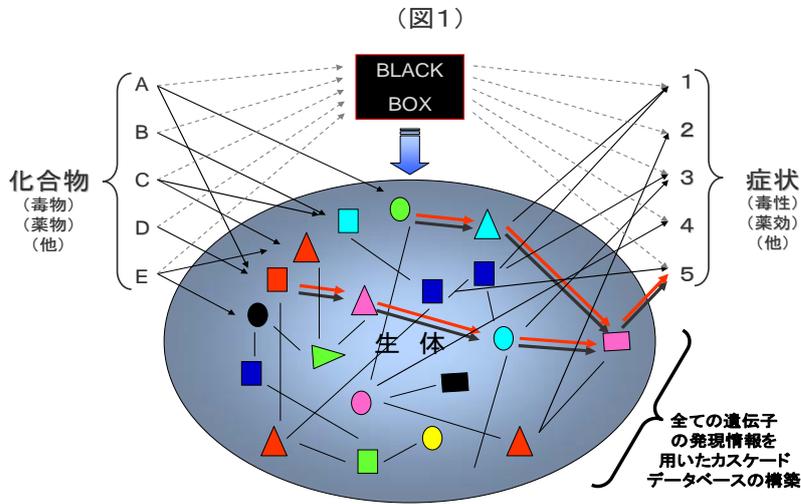


図1. リンケージからカスケードへ

生物学が現象の記述学に基礎を置いていた段階での毒性学は、投与された化学物質と症状との間に介在するBlackbox を挟んでの、それらの関連性に基づいた体系化が行われてきた。しかし、分子毒性学の立場から一番知りたいことは、生体内で実際に起こっている一連の事象であり、transcriptome の場合には全ての遺伝子の情報を元にした遺伝子カスケードの全容解明である。毒性学的に重要なマーカー遺伝子(数十～数百のことが多い)についてのこの様なデータベースは存在するが、ここでは全ての遺伝子を対象としたものを指向する。

(図2)

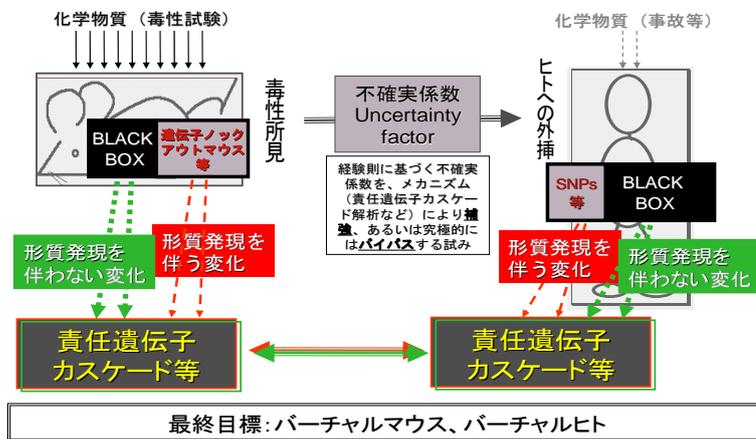


図2. 毒性評価法への分子毒性学の導入

毒性学の近代化のための分子毒性学の導入は、実験動物のブラックボックスとヒトのブラックボックスを繋げるために採用されている経験則ベースの「不確実係数」を、何らかのメカニズム解析により補強、あるいは究極的にバイパスすることを目的とする。その際、Blackbox の解明手段には例えば遺伝子ノックアウトマウスやヒトに於ける SNPs 情報が活用できるが、何れも形質発現を伴う場合にのみ有効に働く。形質発現と結びつかない部分については、網羅的な情報収集を行わざるを得ないと考える。

さらに、創薬における毒性分野への要求のひとつに、副作用による臨床段階での開発中止例、あるいは市販後の販売中止例の減少が挙げられる。即ち、「動物実験では毒性が無かったことから臨床試験に進んだところ、ヒトで毒性が現れ、開発中止となった。その為に何億円もの経費が無駄になった。この様な事態を回避せよ」、即ち、動物で所見が無くともヒトでの毒性を予測する事が要求されている訳である。ここでも形質発現に依存しない方策が要求されている。サリドマイドはマウスには目に見える変化は起こさないかもしれないが、血管新生や免疫修飾など、種々の作用が誘導されることが報告されている。これは、マウスでは形質発現が明らかでなくとも、ヒトへの影響を予測する方策の存在の可能性を示している。

また、恒常性維持機構に深く関わる内分泌かく乱化学物質の問題など、外界からの影響が効率よく中和されてしまい、形質変化がモニターしにくい対象を扱う場合にも、形質発現の有無に関らず mRNA や蛋白の発現修飾を観測することが有効な影響解析手段となることが考えられる。

この様に、今後の毒性学における transcriptome 解析、即ちトキシコゲノミクス (toxicogenomics) は、従来の「形質依存型」のものから「形質非依存型 (phenotype-independent)」に発想を転換する時期に来ていると言えよう。

形質非依存型トキシコゲノミクス (phenotype-independent toxicogenomics) の条件：

形質依存型では、ある特定の毒性所見にリンクした遺伝子をマーカーとして選択し、それが毒性発現に重要であると認定する事から始まる。これに対して、形質非依存型トキシコゲノミクス (phenotype-independent toxicogenomics) は、まずは形質発現情報などの情報を用いずに、自らの遺伝子発現プロファイル情報のみを頼りに遺伝子発現変化の解析を開始しようとする点にある。即ち、ある毒性所見にリンクしたマーカー遺伝子を認定出来ないの、測定する全ての遺伝子はどれも平等に重要であると仮定する必要がある。そして、その全てがどれだけ発現増加したか、減少したか、あるいは不変であったかを正確に観測する必要がある。更に、幾多の化学物質を検討した結果はじめて全体像が明らかになるため、複数の実験の結果を長きに亘り集積し、それらのデータを縦横に解析する必要がある。

この条件を満たすためには、今までのマイクロアレー手法には問題があった。まず、マイクロアレーの性能として、mRNA の測定可能な範囲が比較的狭いために 1 枚当たり用いる総 mRNA 量を一定量に揃える必要があった点である。これは mRNA が少な過ぎると蛍光シグナルが弱くてデータが得られず、多過ぎると蛍光シグナルが飽和してしまって定量性の良いデータが得られない事態を回避するための措置である。この場合、サンプル中の細胞 1 個当たりの mRNA の絶対的な多寡に関する情報は消失してしまう。この様な相対的な情報でのサンプル間の mRNA 発現の比較のために、種々の標準化手法が編み出されている (2-10)。原則的には、統計学的な有意差検定を基にした変動遺伝子の抽出が行われる。この様な計算に際しては、大半の遺伝子はサンプル間で不変であるとの前提が必要であり、その結果、多数の遺伝子が「変動したとは言えない」と位置付けられる事となる。また、変動の大きさを表現するために対照群のサンプルに対して何倍変化したかを比率表示することが多い。この場合、対照群のサンプルで殆ど発現していない遺伝子は表示が困難となるばかりでなく、異なる時期に実施した複数の実験を比較する際に、対照群の実験間変動を吟味する情報が消失してしまうという問題が加わる。

Percellome と millefeuille data :

このような問題を解決し、形質非依存型トキシコゲノミクスに適用するため、我々は、細胞1個当たりの mRNA 絶対量を得る方法 (Percellome) を、当時それに必要な条件を満たしていたアフィメトリクス社の GeneChip を対象に開発した (特許出願中、投稿中)。このシステムは大きく4つの要素からなっている。第一に RNA 用に準備したサンプル破砕液の極一部からその DNA 濃度を簡便に測定する方法、第二に用量関係を考慮し工夫されたスパイク RNA 液の調整と、その破砕液への添加法、第三に Hill 式に基づいた絶対化アルゴリズム、そして、マイクロアレーの用量相関性能の検証や、バージョンが異なるマイクロアレー間のデータ変換、曳いては、異なったメーカーのマイクロアレー間のデータ変換に用いる標準サンプルセットとデータ変換アルゴリズムである。Percellome データは細胞1個当たりの絶対量であるので、各遺伝子の発現量をゼロを起点とする均等目盛りで表示し直接比較することが可能である。今まで用いられてきた対照に対する比率表示と違って、割り算をする必要が無いため、発現値がゼロの場合の表示が自由に行える上、何よりも、対照群も処置群も同列に表示することが可能となった。また、さらなる標準化操作が不要であるため、測定した全ての遺伝子について、マイクロアレー間はもとより、実験間の直接比較が可能となった。さらに、GeneChip の新旧バージョン間のデータ変換も可能となった。また、データを可視化することが非常に容易になったため、生物学者がその内容を直感的に把握し易くなり、その後のデータ解析とインフォマティクス形成に大きく貢献することが示されつつある。これらの機能は複数の実験からの結果を長きに亘り蓄積する必要があるトキシコゲノミクス研究には重要なことである。また、後述するように異なったプラットフォーム間でのデータ互換にも拡張可能であり、共通の大型データベースやコンソーシアム構築にも貢献する可能性が高い。

方法の概略:

1. DNA 測定: 細胞1個当たりの mRNA 情報を得るために、サンプルを構成する総細胞数を測定する。実際に細胞数を計測する事は特に実質臓器の場合には困難なため、その代替指標として、細胞核内のゲノム DNA 量を用いる。サンプルを DNA 測定専用消費することを避けるため、RNA 調整用の組織破砕液の極一部 (通常、 $10\mu\text{l}$) を DNA 測定に用いるプロトコルを確立した。
2. 多段階濃度スパイクカクテル (GSC : dose-graded spike cocktail) : 細胞1個当たりの mRNA の標準として、組織破砕液に添加するスパイク RNA には、アフィメトリクス社の GeneChip が使用者の為に用意していた5種類の枯草菌由来遺伝子の RNA を用いた。5種類の枯草菌 RNA を各々約 2000 塩基の長さ合成し、5段階の用量に配合したカクテルを作成した。これにより、広い濃度範囲をカバーする標準用量作用曲線を全てのサンプルに導入する事が可能となった。
3. 絶対量化プログラム: アフィメトリクス GeneChip は、蛍光シグナルと mRNA 量との間に Hill 式に従う関係が成立することを後述の LBM 標準サンプルなどにより確認した。その結果から、Hill 式の直線化式により GSC を直線化して絶対量化を行う変換アルゴリズムを開発し、それを自動実行するプログラムを独自に開発した (相崎)。
4. GeneChip の用量相関性確認及びバージョン間・プラットフォーム間データ変換対応のための LBM (liver-brain mix) 標準サンプル及びデータ変換アルゴリズム: 遺伝子発現プロファイルが大きく異なる一対の組織を一定の比率で相互に希釈し合ったサンプルセットを表記の目的のために用意した。具体的には、肝と脳を用い、100:0、75:25、50:50、25:75、および 0:100 の混合比の5サンプルからなるセットを用意した。

絶対量化の原理：

基本的原理は、サンプルの細胞数(ゲノム DNA 濃度で代替)に比例した分子数のスパイク RNA を添加する事で、サンプルの細胞当たりの mRNA 絶対量(コピー数)の指標をサンプル中に導入するものである(図3)。ただし、スパイク RNA は1点を規定するものではなく、5種類の枯草菌遺伝子に対する RNA(哺乳類の配列と交叉しない)を適切な公比を持たせて5段階の濃度に割り振ったカクテルとして用いる事が特徴である(図4)。これにより、絶対コピー数の指標となると同時に、広い用量範囲について検量線を各サンプルに導入した事になり、mRNA 抽出から GeneChip の蛍光測光までの過程で生じるデータ全体の歪みを補正する際に威力を発揮するとともに、全ての GeneChip の発現値を統一基準下で安定的に絶対量化する効果を有している。サンプルに由来する全ての測定値は Hill 関数の直線化式により直線化された GSC 検量線に基づいて絶対量に変換される。

$$\text{Log}(S/S_{\text{MAX}} \cdot S) = \gamma \log C - \gamma \log S C_{50}$$

(式中、Sは測定値、S_{MAX}は最大測定値、Cはスパイク RNA の濃度、S C₅₀は50%反応濃度、γは Hill 係数を示す。尚、高発現側の歪みを気にしない場合には、S_{MAX}を無限大に置いた近似式での代用が可能である)。

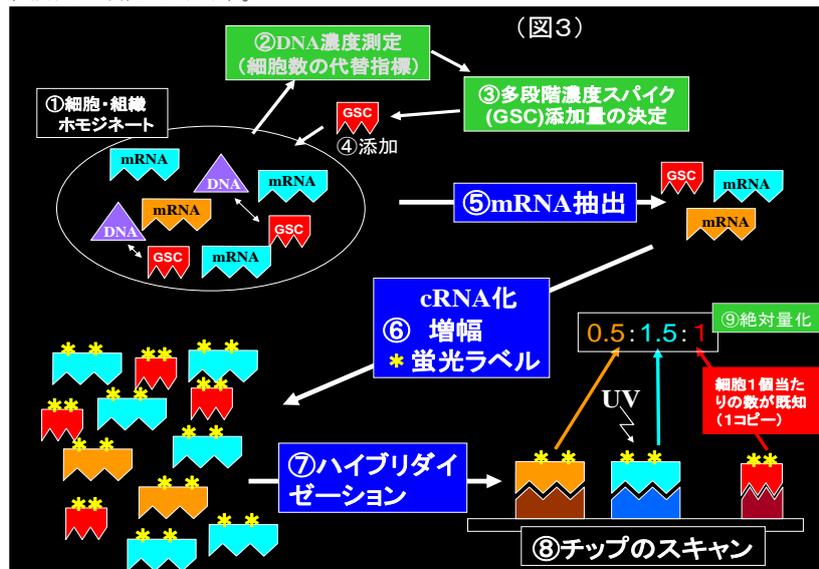


図3. 絶対量化の概略

絶対量化の原理は、①のサンプル・ホモジネートの細胞数を DNA 量として捕らえ、細胞個数に比例した量の多段階濃度スパイク RNA カクテル(GSC)を添加する(②~④)。その後の⑤~⑧は通常の手順を踏む。⑧における GSC のシグナルが細胞1個当たりの既知コピー数を示している。測定したいサンプル中の mRNA の細胞1個当たりのコピー数は GSC のシグナル強度との比較からもとめることが出来る(⑨)。GeneChip に於いては④の RNA 濃度と⑧の蛍光強度の関係が Hill 式で記述できることを確認しており、それを用いた変換式により測定された全ての遺伝子についてサンプルの細胞1個当たりの絶対量が導き出される。緑の囲み:本法で追加された手順。青の囲み:アフィニトリクスのプロトコール手順。

(図4)

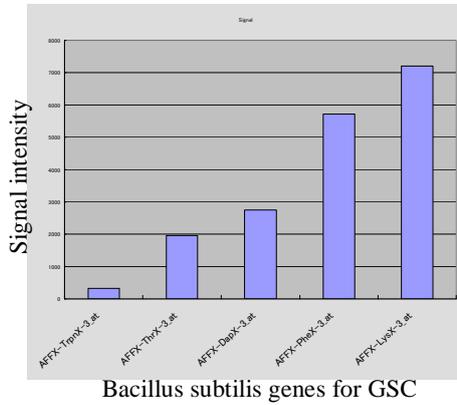


図4. 多段階濃度スパイク RNA カクテル (GSC)

GSC はグラフに示す如く低い値から高い値まで幅広いよう領域をカバーする様に 5 種類の枯草菌 mRNA を合成して 5 段階の濃度に混合したものである。これを適切に添加することにより全てのサンプル中に細胞 1 個当たりの指標と mRNA 検量線を導入している。その結果 GeneChip 一枚毎のデータの歪みを検出することが可能となり絶対量化の精度を格段に高める結果となっている。また、同様の理由で新旧のバージョン間や異なったプラットフォーム間のデータ変換の際にも、標準曲線として有効に機能する。

一例として、スキャナーを取り替えた際のデータの歪みを矯正した事例を紹介する(図5)。複数のサンプルの間での、あるいは複数の実験間でのある遺伝子の発現変動の比較は本システムにより飛躍的に向上することが示されている。例えば、日内変動遺伝子の日内変動が絶対表示により直読可能であり、その発現様態は発現値をも含めて実験間で再現されている。

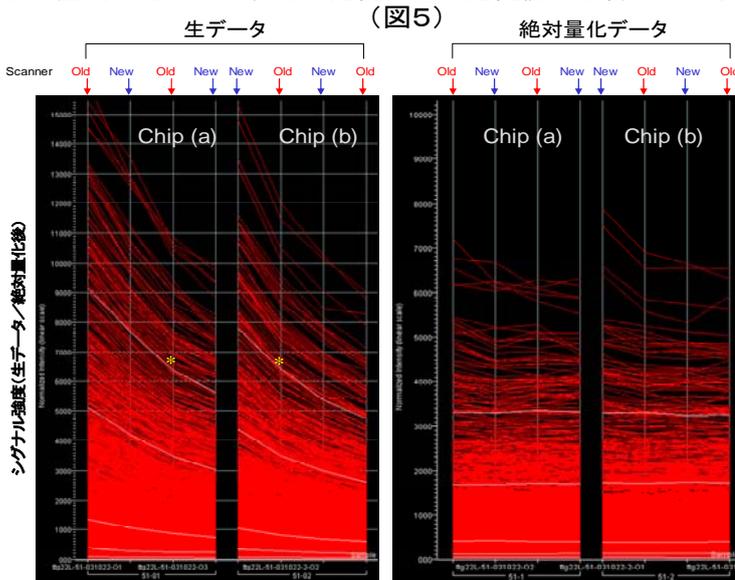


図 5: 絶対量化の効果の一例

新旧2台のスキヤナーの特性比較を行った。ひとつのサンプルを二つに分け、a、b2枚の GeneChip にハイブリダイズした。Chip(a)を新スキヤナーにて、Chip(b)を旧スキヤナーにて測定し、さらに Chip を取り替えて再スキヤンすることを、三回繰り返した。その結果を GeneSpring(Silicon Genetics)を用いてグラフ化した(縦軸はゼロを起点とする均等目盛り表示)。スキヤンを繰り返すたびに蛍光強度は減衰してしまう。傾斜の角度の違い(図中*の部分の角度参照)から減衰の度合いは旧スキヤナーで若干強いことが分かる。このデータを絶対量化したデータでは、減衰がほぼ完全に補正される(シグナルの高いところでも、値の相対的な振れは1割程度に収まっている)。白線は GSC。

他方、チップ内での異なる遺伝子の発現量の正確さに関しては、GeneChip のプローブセットの設計に依存する。アフィメトリクスはプローブの設計に際してそれらの t_m 値を一定に保つアルゴリズムを用いている。これについては、利用者として個々に定量的 PCR などにより検証する必要がある。

LBM :

システムの定量性の検定 :

LBM は当方の便宜上、肝と脳の組み合わせを用いたが、遺伝子発現プロファイルの異なるペアであればどの様な組み合わせでも利用可能である。複数のペアを併用すればさらに精度の良い検定が可能となる。GSC を DNA 濃度に応じて添加した LBM セットを測定し、絶対量化した結果は、グラフ化すると直線を描くはずであり(図 6)、さらに 50:50 のサンプルで除した場合、理想的には全ての遺伝子が 50:50 のところで 1 の値を取り、100:0 あるいは 0:100 では 0 から 2 の間の値をとるところの直線を描くはずである。この結果から、マイクロアレーの定量性が確認される。

(図6)

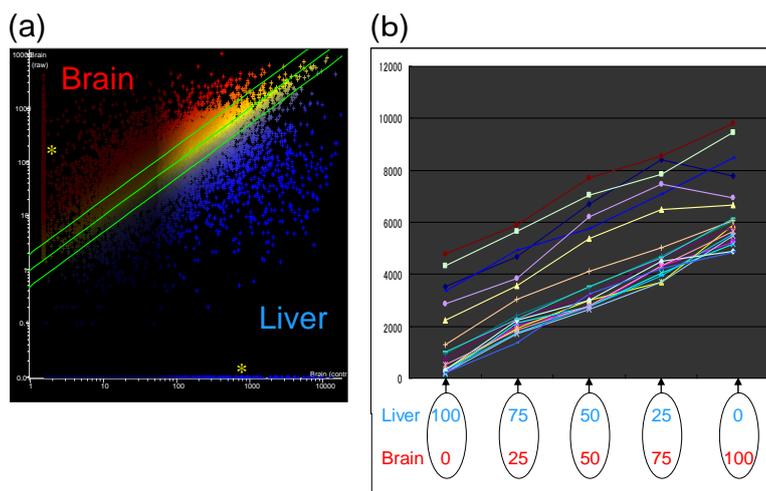


図 6: LBM (Liver-Brain Mix) 標準サンプルセットによるシステムの定量性の検定

(a)のスカッターグラフが示す如く脳と肝臓では発現遺伝子のレパートリーが大きく異なり、X軸およびY軸に沿って遺伝子が線状に並ぶ(*)ことで判るとおり、片方の臓器にのみ発現する遺伝子が多い。このような臓器の対からサンプルを調整し(b) 100:0、75:25、50:50、25:75、および 0:100 の混合比の 5 サンプルを GeneChip (MG-U74v2A) にて測定すると(b)の如く、ほぼ直線を得る。全ての遺伝子について 50:50 の値に対する比を求め、同様グラフを描くと理想的な性能を有するマイクロアレーでは、全ての遺伝子が 50:50 のところで 1 を通り 100:0 における y 切片が 0~2 の範囲に収まる直線を描く。

GeneChip の新旧バージョン間のデータ変換 :

さらに、この様な LBM サンプルをバージョンアップ前の古い GeneChip と新しいバージョンの GeneChip で測定しておくことにより、LBM に含まれる全ての遺伝子について、5 点からなる新旧のチップに於ける用量相関関数を求めることが出来る。LBM に他の臓器の組み合わせを用いる事で取り扱える遺伝子数を増やすことが可能である(図 7)。

(図 7)

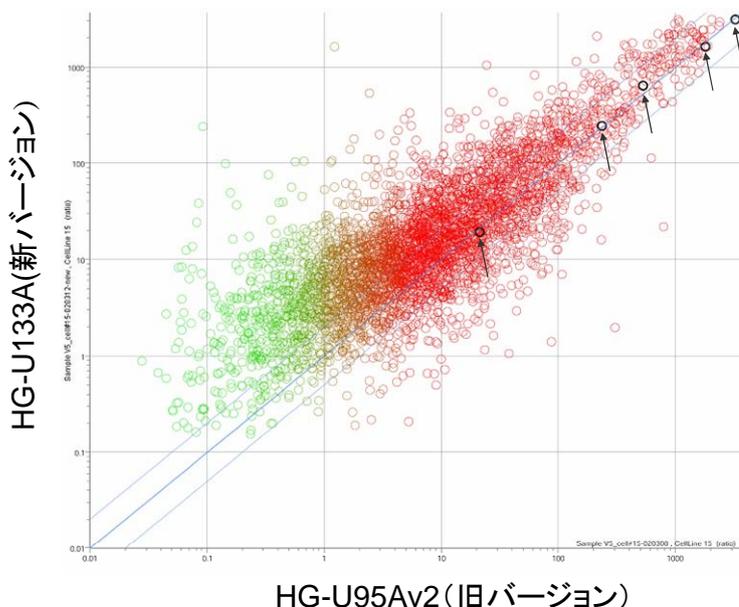


図 7. GeneChip の新旧バージョン間のデータの関係

LBM サンプルセットを新旧のバージョンの GeneChip において測定する。するとここにスカッターグラフで示すような関係が5組得られる。矢印で示す黒丸が GSC である。GSC を基準に新旧の Chip での発現値が標準化され、その様な点が5組のデータから5点得られる事から、ここにプロットされた遺伝子(両バージョンに同一または対応するアノテーションが得られ、かつ、LBM サンプルに発現されているもの)については個々について直接変換式が得られる。これは、条件により定量PCRやアフィメトリクス以外のマイクロアレープラットフォームにも原理的に拡張可能である。

データ互換性の他システムへの拡張 :

本システムの GSC を添加したサンプルはスパイク RNA を検出するプライマーセットを用意す

る事で PCR に於いても容易に絶対量化データを得る事が出来る。詳細は他に譲るが、プライマーペアの増幅効率のばらつきを勘案した絶対化アルゴリズムと共に Percellome 定量 PCR システムを構築中である。アフィメトリクス GeneChip 以外のプラットフォームとのデータ互換も可能である。本システムが適応可能なプラットフォームの条件としては、GSC を受け付けるプローブセットが用意されていること、及び用量相関性が確保されてることの 2 点を満たしている必要がある（現在、2 社検討開発中）。

形質非依存型トキシコゲノミクスへの適用：

創薬開発促進のためのトキシコゲノミクス・プロジェクトと、化学物質の安全性評価のためのプロジェクトの 2 つが現在、Percellome システムの下で進行中である(表 1)。両プロジェクトでは共に、4~5 段階の用量について 4 時点での遺伝子発現を観測する 16~20 群(一群 3 匹)の構成から成るプロトコルを採用した(図 8)。ひとつの化合物について 48~60 匹の動物からのサンプルを解析し Percellome データを生成する。遺伝子の発現値を 3 次元表示することでその用量・時間依存性が視覚化できる。X 軸に用量、Y 軸に時間、Z 軸に発現量（ゼロからの均等目盛り表示）をプロットする事により、ひとつの遺伝子につき 16~20 格子点（48~60 枚の GeneChip からのデータ）から成る 1 枚の局面を描くことが出来る(図 9)。ひとつの GeneChip

プロジェクト名	「トキシコゲノミクス手法を用いた医薬品安全性評価予測システムの構築とその基盤に関する研究」 (厚生労働科学研究費補助金、萌芽的先端医療技術推進研究事業)H14年度~5年計画	「化学物質リスク評価の基盤整備としてのトキシコゲノミクスに関する研究」 (厚生労働科学研究費補助金、化学物質リスク研究事業)H15年度~3年計画
組織	創薬関連企業17社参加「産学官連携」プロジェクト(プロジェクトリーダー:長尾拓(国立衛研所長))	研究班体制の研究プロジェクト 国立衛生研・安全性生物試験研究センター内(主任研究者:菅野 純)
目標	創薬過程における安全性の早期予測システムの構築	国が行う既存化学物質の点検を、より迅速、安価かつ正確に実施する毒性予測システムの構築
期待される効果	医薬品による副作用の早期予測により、臨床段階での開発中止を回避し、以って創薬の経費削減と効率化の促進 予期せぬ副作用の低減による国民被害の減少、より安全性の高い医薬品の創製による製薬企業の活性化と国民の健康増進への寄与	日常生活に於いて使用される数万種の化学物質の毒性を、従来の毒性試験よりも、迅速、安価かつ網羅的に予測する事による国民の安全・安心の向上 毒性発現メカニズムに支えられた包括的な毒性評価の体制の整備
検討物質	開発中止、あるいは販売中止となった医薬品・医薬品候補物質を中心としたDrug-likeな化合物(150物質/5年)	日常使用される数万種に及ぶ化学物質を中心とした各種の物質(約90物質/3年)
モデル動物	ラット(医薬品の審査に使用されるSD/IGSラット)	マウス(遺伝子改変マウスの活用を見越しC57BL/6マウス)
検索臓器	肝および腎。ヒトおよびラット肝細胞由来培養細胞	肝および化学物質固有の標的臓器

表1:厚生労働省におけるトキシコゲノミクス研究

が 45,000 のプローブセットから成る場合、ひとつの化合物の用量・時間依存的データ 3 次元表示では 45,000 枚の局面の層状集合体から成る (millefeuille data と名付けた)。この millefeuille data は各格子点が 3 匹の動物に由来する 3 つのデータを基にしており、格子点のデータの信頼性の評価を含めて、artifact の除去や、生物学的な蓋然性のある変化であるか否かの判別に適している上に、類似の用量・時間反応を示す遺伝子の選別に威力を発揮する。

(図8)

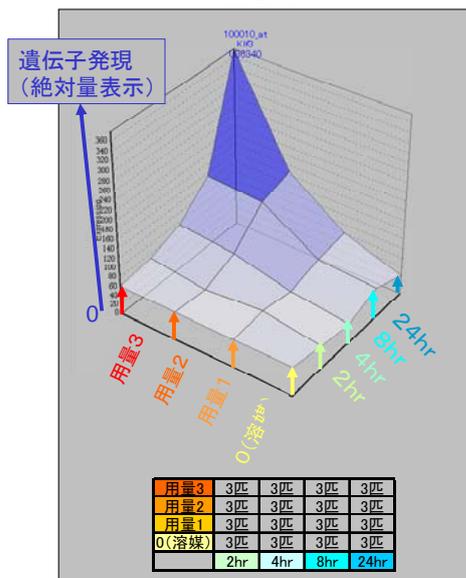


図8:トキシコゲノミクス・プロジェクトにおける単回投与実験の基本構成

時間と用量の組み合わせからなる4X4~4X5 のマトリクス構造のプロトコールにてデータを生成中である。各群 3 匹とし、サンプルはプールせず個別に GeneChip 解析を実施している。X 軸に用量、Y 軸に時間、Z 軸に発現量(ゼロからの均等目盛り表示)をプロットする事により、ひとつの遺伝子 1 枚の局面を描くことが出来る。現在使用中の MOE430v2 は約 45,000 のプローブセット情報を生成するため、ひとつの化合物の transcriptome 情報は 45,000 枚の局面の集合体(ミルフィーユ・データ)で表される。

形質非依存型トキシコゲノミクスにおけるデータ解析法：

創薬開発促進のためのトキシコゲノミクス・プロジェクトでは、特に開発中止となった化合物を含む薬剤関連の化学物質を中心としてラットを用いた実験を進めており、既に蓄積されている膨大なラット毒性情報との対比に重点を置いた解析を(株)日立製作所と共に推し進めている。他方、著者らが進めている化学物質の安全性評価のためのプロジェクトに於いては、一般的な化学物質が対象であるため毒性データが必ずしも豊富でないこともあり、生体反応の分子メカニズム解析(カスケード解析)に重点を置き、遺伝子欠失動物の活用を見込んで、マウスを用いた実験を重ねている。こちらでは遺伝子発現プロファイルを体系化するために形質発現情報に頼らず、完全な教師無しクラスタリングを実施する。Millefeuille data を基礎に、生物学者が視覚的に確認できる変数を利用する方法を NTT コムウェア株式会社と共同開発し、Teradata (日本 NCR 株式会社) による解析・データベース上に搭載した。このクラスタリング手法は、クラスター数及びクラスター径を指定せず、通常 45,000 プローブセット (MOE430v2) を小

さいクラスターから数百万クラスターに分類する。今後この方法と、適切な遺伝子欠失マウスによる *millefeuille data* 生成により、客観的な遺伝子カスケード構築を目指している。その上で、既知の情報との比較を行い、必要に応じて確認のための小実験を別途追加して実施し、最終的に信頼性の高い遺伝子カスケードデータベースの構築と、これに基づいた効率的で正確な毒性評価・予測技術の開発を目指している。

まとめ：

毒性学は毒性という形質発現を基に成り立ってきているが、その効率化と正確性向上のために分子毒性学的なメカニズム解析の導入を図るに当たり、形質が発現する以前の段階、あるいはフィードバック機構が働くために形質発現が乏しい状態など、形質発現を直接には伴わないところでの遺伝子発現変動を網羅的に捕らえて毒性に関わる遺伝子発現カスケードの全容を解明する必要に迫られている。これに応えるために、我々は形質非依存型トキシコゲノミクスの概念の導入と、それに必要な技術であるマイクロアレーから細胞 1 個当たりの mRNA 絶対量情報を生成する *Percellome* システムを開発した。

絶対量化された *Percellome* データは、その全てを生物学者にとって分かりやすい 3 次元 *millefeuille data* として表示する事が可能であり、その結果、クラスター化等のデータ解析過程を生物学的蓋然性に基づいて比較的容易に検証する事が出来る様になった。本システムは 4x4 ~ 4x5 マトリックス方式の大型プロジェクトを対象として開発したものであるが、実際には小規模の実験サンプルに対しても有用性が高いことが実証されている。また、特に変動遺伝子リストの遺伝子数が飛躍的に増大することが多い。それは、変動比率による足切りやハズレ値計算の様な統計手法を用いる必要が無く、個々の遺伝子について逐一比較が出来るためである。

絶対量化された発現値は対照群を含めて *GeneChip* 間あるいは実験間でそのままの形で相互に直接比較が可能であり、例えば日内変動遺伝子があるまま *millefeuille data* として表示されている。さらに、異なったバージョン間、定量 PCR との間、更にはアフィメトリクス以外の *Percellome* に対応したプラットフォームとの間でのデータ変換のための変換関数を導き出すことが可能であり、この特徴は複数の研究者や組織がデータを持ち寄るデータベースの構築に貢献する可能性が期待される。

以上、*Percellome* システムは *millefeuille data* と相まって *transcriptome* の精度と相互互換性を有意に高める事が示唆されつつある。現在まで創薬プロジェクトも当方の化学物質プロジェクトも 4x4 ないし 4x5 マトリックス規模の実験を 30~40 化合物程度実施し、膨大なデータの蓄積を開始した。今後、この形態のデータの有用性を客観的に評価頂くための *Percellome consortium* の構築を考えていきたい。

謝辞：

本システムの開発とプロジェクトの遂行は、当毒性部の諸先生方及び安東朋子、森山紀子、近藤優子、中村祐子、安部麻紀、吉木健太、松田菜恵、森田紘一、今井あや子の各氏の献身的努力に負っておりここに深く感謝する。本研究は厚生労働科学研究費補助金 H13-生活-012、H13-生活-013、H14-トキシコ-001 及び H15-化学-002 による。

参考文献：

1. Waters, M.D., Olden, K. and Tennant, R.W. (2003) *Mutat Res*, 544, 415-424.
2. Hill, A.A., Brown, E.L., Whitley, M.Z., Tucker-Kellogg, G., Hunter, C.P. and Slonim, D.K. (2001) *Genome Biol*, 2, RESEARCH0055.
3. van de Peppel, J., Kemmeren, P., van Bakel, H., Radonjic, M., van Leenen, D. and Holstege, F.C. (2003) *EMBO Rep*, 4, 387-393.
4. Hekstra, D., Taussig, A.R., Magnasco, M. and Naef, F. (2003) *Nucleic Acids Res*, 31, 1962-1968.
5. Sterrenburg, E., Turk, R., Boer, J.M., van Ommen, G.B. and den Dunnen, J.T. (2002) *Nucleic Acids Res*, 30, e116.
6. Talaat, A.M., Howard, S.T., Hale, W.t., Lyons, R., Garner, H. and Johnston, S.A. (2002) *Nucleic Acids Res*, 30, e104.
7. Bolstad, B.M., Irizarry, R.A., Astrand, M. and Speed, T.P. (2003) *Bioinformatics*, 19, 185-193.
8. Lee, P.D., Sladek, R., Greenwood, C.M. and Hudson, T.J. (2002) *Genome Res*, 12, 292-297.
9. Wilson, M., DeRisi, J., Kristensen, H.H., Imboden, P., Rane, S., Brown, P.O. and Schoolnik, G.K. (1999) *Proc Natl Acad Sci U S A*, 96, 12833-12838.
10. Heller, R.A., Schena, M., Chai, A., Shalon, D., Bedilion, T., Gilmore, J., Woolley, D.E. and Davis, R.W. (1997) *Proc Natl Acad Sci U S A*, 94, 2150-2155.