

1つの化合物のデータは約45,000 probesetの曲面からなる
(MOE430 v 2 GeneChipの場合)

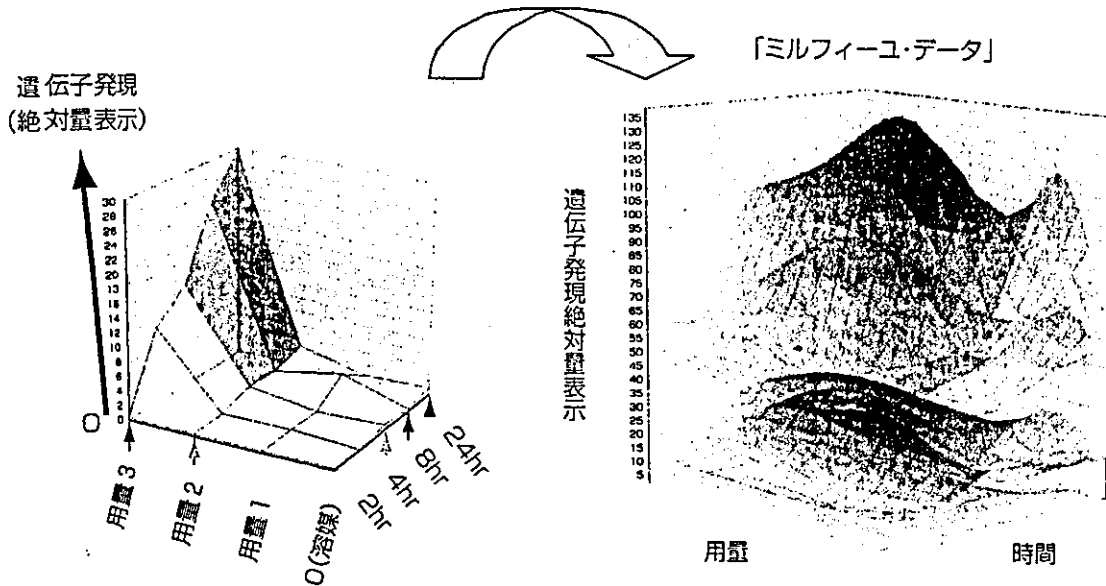


図6 トキシコゲノミクス・プロジェクトにおける単回投与実験の基本構成とミルフィーユ・データ

時間と用量の組合せからなる4×4のマトリックス構造のプロトコールを示す。各群3匹、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）をプロットすることにより、1つのプローブセットごとに1枚の発現局面を描くことができる。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、1つの化合物のトランスクリプトーム情報は45,000枚の局面の集合体（ミルフィーユ・データ）であらわされる。→巻頭カラー図参照

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、クラスター数を指定せず、45,000プローブセット（MOE430v2）を小さいクラスターから数百のクラスターに分類する。複数の化学物質からのクラスターデータの解析と、適切な遺伝子欠失マウスによるMF dataにより、客観的な遺伝子カスケードの描出を試みている。これと既知の情報との比較を行い、必要に応じて不足部分の確認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、これにもと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。

〈謝辞〉

本システムの開発とプロジェクトの遂行に当たり、当毒性部の諸先生方および安東朋子、森山紀子、近藤優子、中村祐子、安部麻紀、吉木健太、松田菜恵、森田祐一、今井あや子、青柳千百合、相原妃佐子の各氏に深く感謝する。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およびH15-化学-002による。

参考文献

- 1) Ema, M. et al. : J. Biol. Chem., 269 : 27337-27343, 1994
- 2) Jacobson, J. L. & Jacobson, S. W. : Neurotoxicology, 18 : 415-424, 1997
- 3) Jacobson, J. L. & Jacobson, S. W. : Obstet Gynecol Surv., 59 : 412-413, 2004
- 4) Waters, M. D. et al. : Mutat. Res., 544 : 415-424, 2003
- 5) Hill, A. A. et al. : Genome Biol., 2 : RESEARCH0055, 2001
- 6) van de Peppel, J. et al. : EMBO Rep., 4 : 387-393, 2003
- 7) Hekstra, D. et al. : Nucleic Acids Res., 31 : 1962-1968, 2003
- 8) Sterrenburg, E. et al. : Nucleic Acids Res., 30 : e116, 2002
- 9) Talaat, A. M. et al. : Nucleic Acids Res., 30 : e104, 2002
- 10) Bolstad, B. M. et al. : Bioinformatics, 19 : 185-193, 2003
- 11) Lee, P. D. et al. : Genome Res., 12 : 292-297, 2002
- 12) Wilson, M., et al. : Proc. Natl. Acad. Sci. USA, 96 : 12833-12838, 1999

7
ゲノム創薬と
薬理ゲノミクス

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

形質非依存型トキシコゲノミクスに適用するため、マイクロアレイから細胞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で表示し直接比較することができるようになった。そのため、今まで用いられてきたコントロールに対する比率表示と違って、割り算をする必要がなく、発現値ゼロの表示が自由に行え、コントロール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原則的に不必要なため、測定したすべての遺伝子についてマイクロアレイ間はもとより、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やすいよう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

① 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 E-mail: kanno@nihs.go.jp

1985年東京医科歯科大学大学院医学研究科博士課程修了, 医学博士(人体病理学, 実験病理学専攻), 国立医薬品食品衛生研究所毒性部室長を経て, 2002年より同部長。内分泌かく乱関連などの分子毒性学研究, トキシコゲノミクスプロジェクトなどを所掌業務との有機的連携のもとに推進。

相崎健一, 五十嵐勝秀, 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)との相互作用の中で存在している。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、薬草を見いだすなど太古の昔から存在しているが、近代生活にお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、生活の利便性のために開発・利用され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、身体に入らなまでの“物質の分類”はもはや重要ではなく、むしろ身体に入った後にどのような反応がいかにか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、目的に合った方法でスクリーニングする方策がとられる。“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。いずれにせよ、“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なおさらである。

本稿では、cost-benefitの対象を人(ヒト)に限定して話を進める。ヒトにおける毒性を検討するためには、ヒトからの情報が一番正確なことは言うまでもない。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、ヒトに使っても薬効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、厳重な管理体制のもとで、かつ本人の了解を得たうえで“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生死に関わるような大量投与を行うことはない。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、例えば、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、どのくらいの量までなら安全と見な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でき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できれば、それは可能かもしれないが、発癌性が疑われたり、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”が事実上できない対象がある。それは、胎児と子どもである。いずれの場合も、現在のところ、ヒトの身代わりとしてモデル動物を用いることになる。

I. 毒性における量と質の問題

では、どのくらいの量までなら安全と見なせるか。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）のもとでは、“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などの実験動物で求め、種差や個体差を勘案した係数（不確定係数あるいは安全係数と呼ぶ）で割って、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ヒトでの中毒事例、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触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、ラットなど単回曝露時の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ように取り扱われてきたか。生物学が現象の記述学に基づ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“不確定係数”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、ヒト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。このような関連付けを“phenotypic anchoring”と呼ぶことがある¹⁾。

II. 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、トランスクリプトーム（transcriptome）の場合にはすべての遺伝子の情報をもとにした遺伝子カスケードの全容解明である（図1）。これがわかれば、膨大な時間と費用の掛かる長期毒性試験（ラットなどを用いる）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ことが強く期待される。特に胎児、新生児、小児、成人、老人の各発達段階にお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ようになると考えられる。すなわち、実験動物で得た所見をヒトに外挿する際に、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“不確定係数”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ことになる（図2）。これを実現させるためには、例えば、マウスにお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、これら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かかわらずすべ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えない。

さらに、創薬における毒性分野への要求の1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。すなわち、“動物実験では毒性が

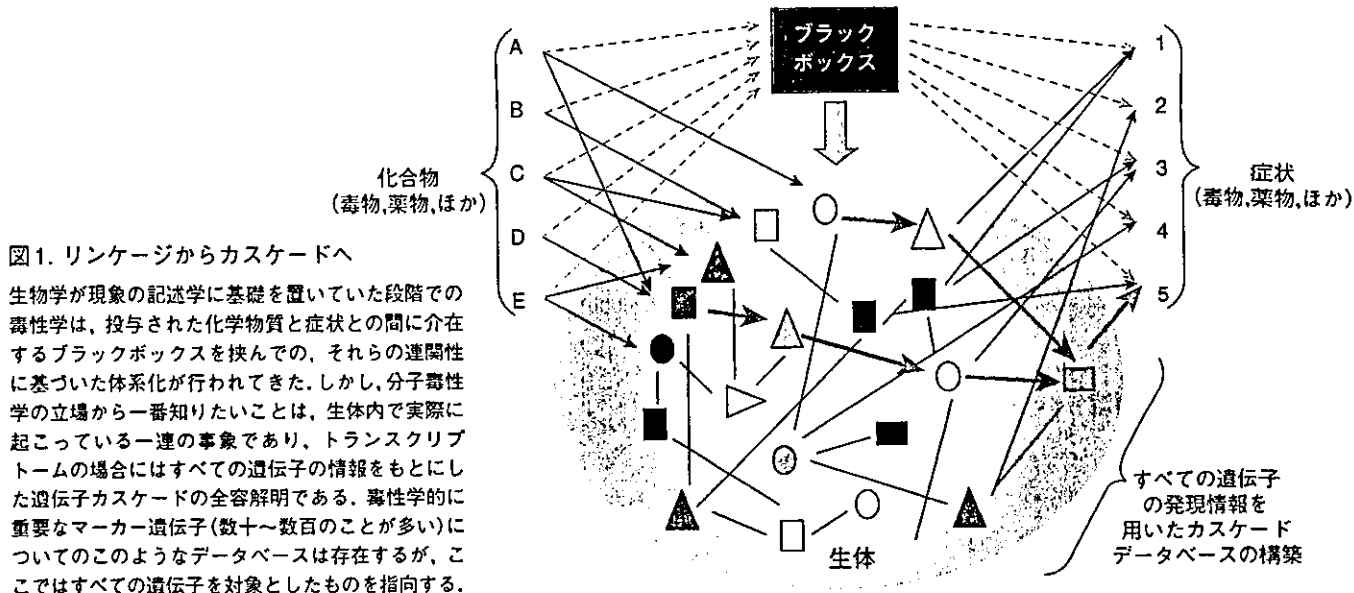


図1. リンケージからカスケードへ

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ブラックボックスを挟んで、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、トランスクリプトームの場合にはすべての遺伝子の情報をもと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)についてのこのようなデータベースは存在するが、ここではすべての遺伝子を対象としたものを指向する。

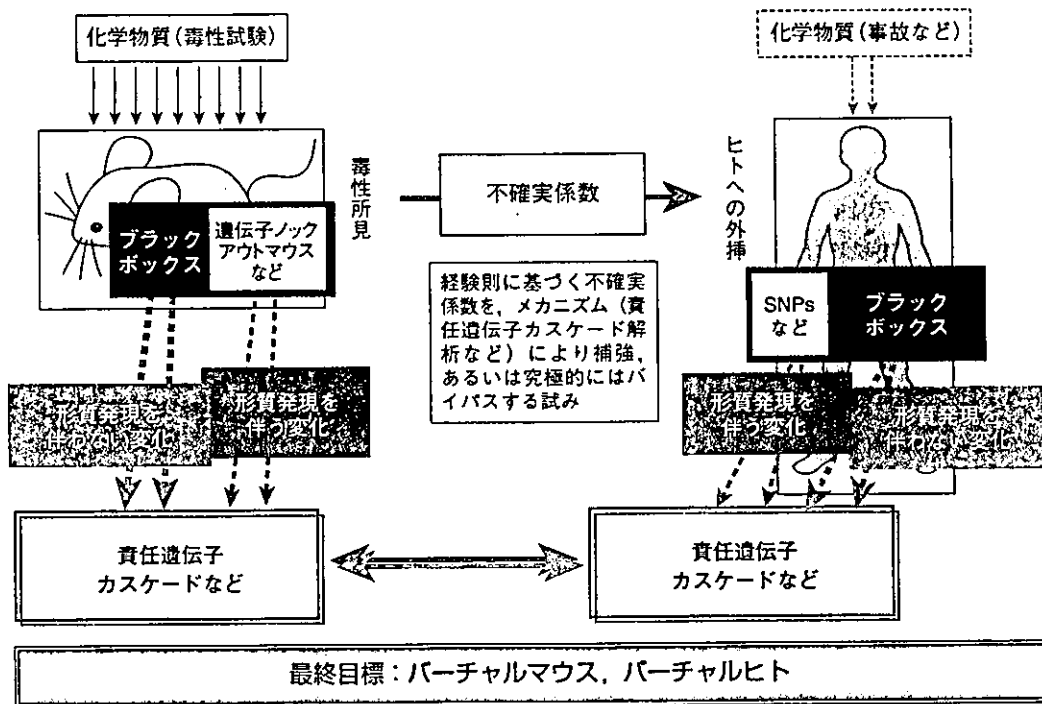


図2. 毒性評価法への分子毒性学の導入

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、ブラックボックス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おけるSNPs情報が活用できるが、いずれも形質発現を伴う場合にのみ有効に働く。形質発現と結び付かない部分については、網羅的な情報収集を行わざるをえないと考える。

な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。そのために何億円もの経費が無駄になった。このような事態を回避せよ、言い換えると、動物で所見がなくともヒトでの毒性を予測することが要求されているわけである。サリドマイドはマウスには目に見え

る変化を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良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、形質発現の有無にかかわらず mRNA やタンパク質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。

このように、今後の毒性学におけるトランスクリプトーム解析、すなわ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。

Ⅲ. 形質非依存型トキシコゲノミクスの条件

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ことから始まる。これに対して、形質非依存型トキシコゲノミクスの特徴は、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。すなわち、ある毒性所見にリンクしたマーカー遺伝子を認定できないので、測定するすべ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、そのすべ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。さらに、幾多の化学物質を検討した結果初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わたり集積し、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イ手法には問題があった。まず、マイクロアレイの性能として、mRNA の測定可能な範囲が比較的狭いために1枚当たりに用いる総 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いためデータが得られず、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。このよう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もとにした変動遺伝子の抽出が行われる。このような計算に際しては、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”と位置付けられることとなる。また、変動の大きさを表現するためにコントロール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、コントロール群のサンプルでほとん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、コントロール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。

Ⅳ. Percellome とミルフィーユ・データ (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、筆者らは、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つの要素から成っている。第1に RNA 用に準備したサンプル破砕液のごく一部からその DNA 濃度を簡便に測定する方法、第2に用量関係を考慮し工夫されたスパイク RNA 液の調製と、その破砕液への添加法、第3に Hill 式に基づいた絶対化アルゴリズム、そして第4に、マイクロアレイ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イ間のデータ変換、ひいては、異なったメーカーのマイクロアレイ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コントロールに対する比率表示と違って、割り算をする必要がないため、発現値がゼロの場合の表示が自由に行えるうえ、何よりも、コントロール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必要であるため、測定したすべての遺伝子について、マイクロアレイ間はもとより、実験間の直接比較が可能となった。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やすくなり、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わた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る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。

1. 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こと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 調製用の組織破砕液のごく一部 (通常、10 μ l) を DNA 測定に用いるプロ

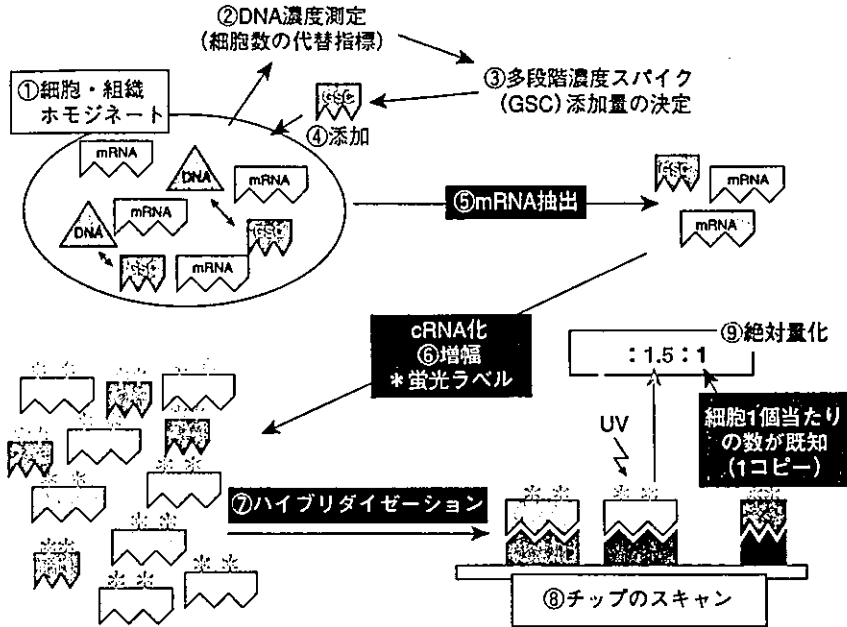


図3. 絶対量化の概略

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量としてとらえ、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 (dose-graded spike cocktail ; 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求めることができる (⑨)。GeneChipにおいては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、それをを用いた変換式により測定されたすべ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：アフィメトリクス社のプロトコール手順。

トコールを確立した。

(2) 多段階濃度スパイクカクテル (dose-graded spike cocktail ; GSC)

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が使用者のために用意していた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2,000塩基の長さに合成し、5段階の用量に配合したカクテルを作製した。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すべてのサンプルに導入することが可能となった。

(3) 絶対量化プログラム

アフィメトリクスGeneChipにより、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。

(4) GeneChipの用量相関性確認およびバージョン間・プラットフォーム間データ変換対応のためのLBM (liver-brain mix) 標準サンプルおよ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

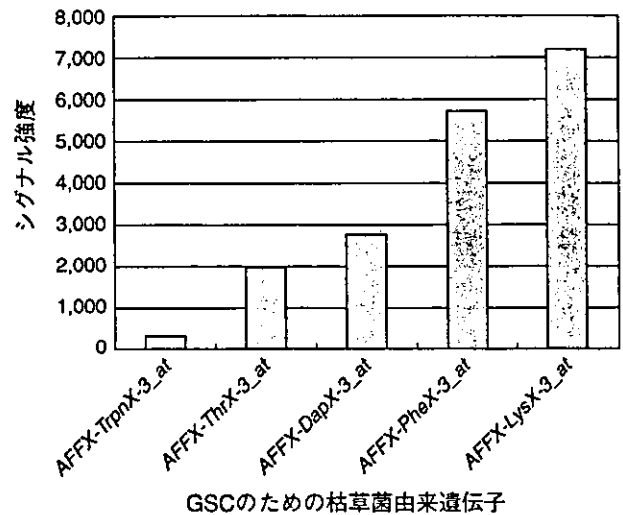


図4. 多段階濃度スパイクRNAカクテル (GSC)

GSCはグラフに示すように、低い値から高い値まで幅広い領域をカバーするように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すべてのサンプル中に細胞1個当たりの指標とmRNA検査線を導入している。その結果GeneChip1枚ごと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。

ルから成るセットを用意した。

2. 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を添加することで、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ことが特徴である(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ことになり、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、すべての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すべての測定値は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。

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

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、 SC_{50} は50%反応濃度、 γ はHill係数を示す。なお、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。

1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。

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のプロープセットの設計に依存する。アフィニティメトリクスはプロープ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。

V. LBM

1. 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よう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、理想的にはすべての遺伝子が50:50のところまで1の値をとり、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、マイクロアレイの定量性が確認される。

2. GeneChipの新旧バージョン間のデータ変換

さらに、このような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

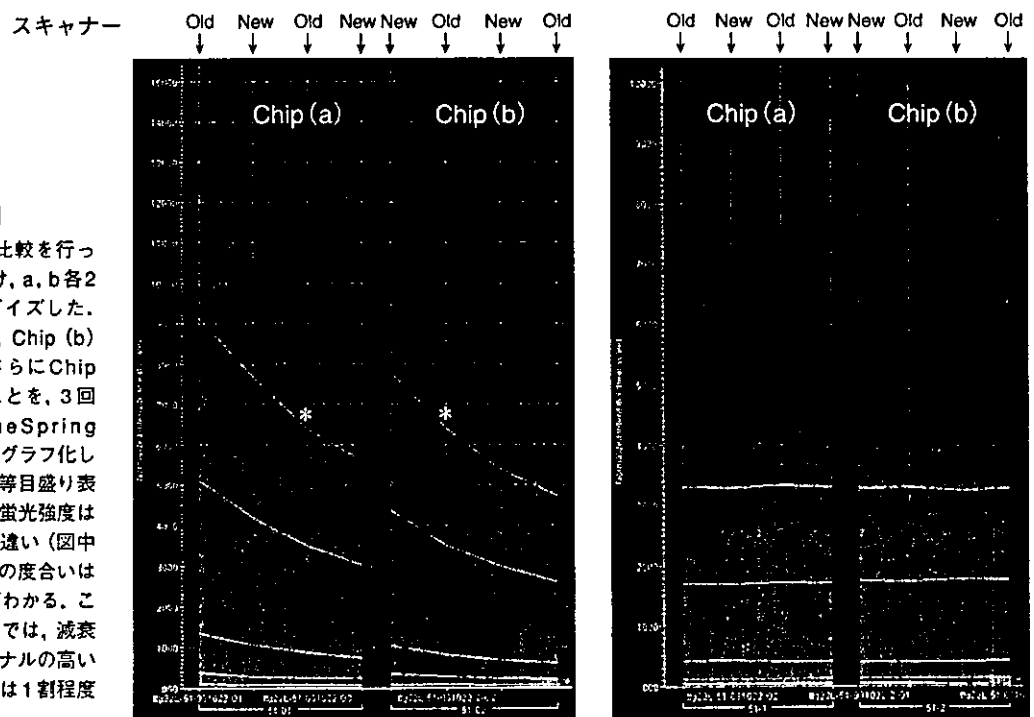


図5. 絶対量化の効果の1例

新旧2台のスキャナーの特性比較を行った。1つのサンプルを2つに分け、a, b各2枚のGeneChipにハイブリダイズした。Chip (a)を新スキャナーにて、Chip (b)を旧スキャナーにて測定し、さらにChipを取り替えて再スキャンすることを、3回繰り返した。その結果をGeneSpring (Silicon Genetics社)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)。スキャ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ャナーで若干強いことがわかる。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。

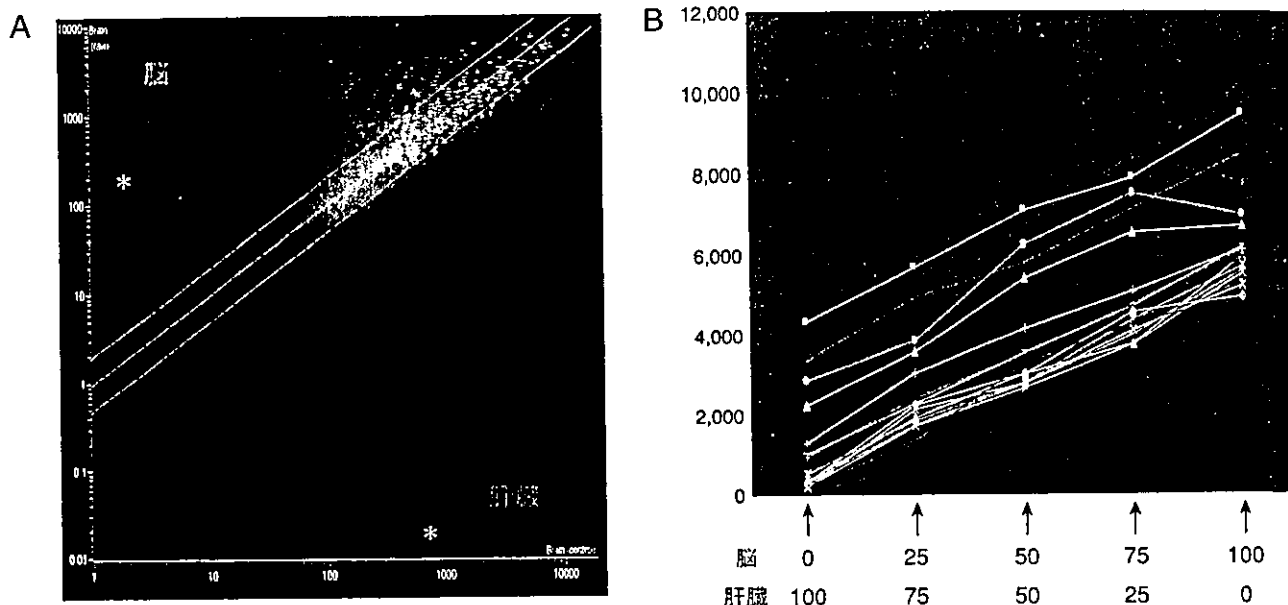


図6. LBM (liver-brain mix) 標準サンプルセットによるシステムの定量性の検定

Aのスカッターグラフが示すように脳と肝臓では発現遺伝子のレポーターが大きく異なり、X軸およびY軸に沿って遺伝子が線状に並ぶ(*)こととわかるとおり、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製し、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のように、ほぼ直線を得る。すべての遺伝子について50:50の値に対する比を求め、同様にグラフを描くと、理想的な性能を有するマイクロアレイではすべての遺伝子が50:50のところを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。

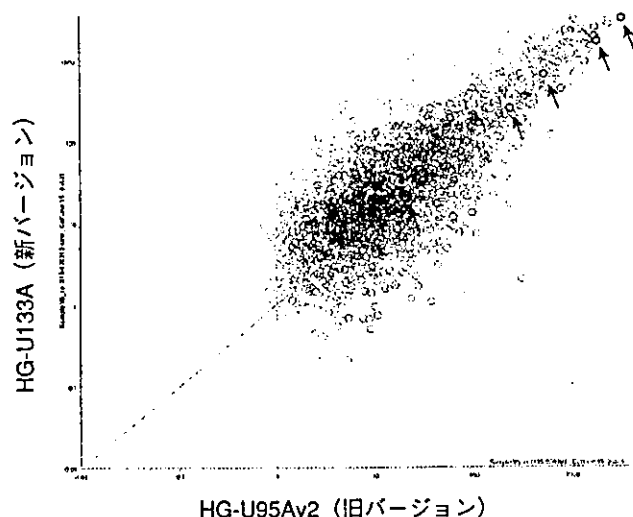


図7. GeneChipの新旧バージョン間のデータの関係

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、そのような点が5組のデータから5点得られることから、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イプラットフォームにも原理的に拡張可能である。

ておくことにより、LBMに含まれるすべての遺伝子について、5点から成る新旧のチップにおける用量相関関数を求めることができる。LBMに他の臓器の組み合わせを用いることで取り扱える遺伝子数を増やすことが可能である(図7)。

3. データ互換性の他システムへの拡張

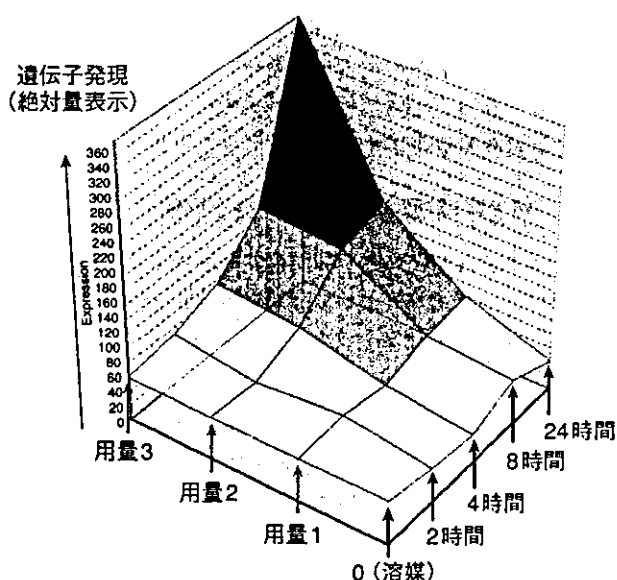
本システムのGSCを添加したサンプルはスパイクRNAを検出するプライマーセットを用意することでPCRにおいても容易に絶対量データを得ることができる。詳細は他に譲るが、プライマーペアの増幅効率のばらつきを勘案した絶対化アルゴリズムとともにPerccellome定量PCRシステムを構築中である。アフィメトリクスGeneChip以外のプラットフォームとのデータ互換も原理的に可能である。本システムが適応可能なプラットフォームの条件としては、GSCを受け付けるプローブセットが用意されていること、および用量相関性が確保されていることの2点を満たしている必要がある(現在、2社検討開発中)。

VI. 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、Perccellomeシステムのもとで進行中である(表1)。両プロジェクトではともに、4~5段階の用量について4時点

表1. 厚生労働省におけるトキシコゲノミクス研究

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、萌芽の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る、臨床段階での開発中止の回避、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	日常生活にお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ことによる国民の安全・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



での遺伝子発現を観測する16～20群 (1群3匹) の構成から成るプロトコルを採用した。1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Perclomene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軸に時間、Y軸に用量、Z軸に発現量(ゼロからの均等目盛り表示)をプロットすることにより、1つの遺伝子につき16～20格子点(48～60枚の

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成る4×4～4×5のマトリックス構造のプロトコ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軸に時間、Y軸に用量、Z軸に発現量(ゼロからの均等目盛り表示)をプロットすることにより、1つの遺伝子につき1枚の局面を描くことができる。現在使用中のMOE430v2は約45,000のプロープセット情報を生成するため、1つの化合物のトランスクリプトーム情報は45,000枚の局面の集合体(ミルフィーユ・データ)で表される。

GeneChipからのデータ) から成る1枚の局面を描くことができる(図8)。1つのGeneChipが45,000のプロープセットから成る場合、1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(ミルフィーユ・データと名付けた)。このミルフィーユ・データは各格子点が3匹の動物に由来する3つのデータをもとにしており、格子点のデータの信頼性の評価を含めて、アーチファクトの除去や、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うえに、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。

VII. 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

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、すで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)日立製作所とともに進めている。他方、筆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おいては、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、完全な教師なしクラスタリングを実施する。ミルフィーユ・データを基礎に、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)と共同開発し、Teradata〔日本NCR(株)〕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、クラスター数およびクラスター径を指定せず、通常45,000プローブセット(MOE430v2)を小さいものから順に数百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ミルフィーユ・データ生成により、客観的な遺伝子カスケード構築を進める。そのうえで、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。

おわりに

毒性学は毒性という形質発現をもと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動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、筆者ら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イから細胞1個当たりのmRNA絶対量情報を生成する

Perccellome システムを開発した。

絶対量化されたPerccellomeデータは、そのすべてを生物学者にとってわかりやすい3次元ミルフィーユ・データとして表示することが可能であり、その結果、クラスタ化など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ことができるようになった。本システムは4×4~4×5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。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ような統計手法を用いる必要がなく、個々の遺伝子について逐一比較ができるためである。

絶対量化された発現値はコントロール群を含めて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そのままミルフィーユ・データ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、さらにはアフィメトリクス以外のPerccellome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、Perccellomeシステムはミルフィーユ・データと相まってトランスクリプトームの精度と相互互換性を有意に高めること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4×4ないし4×5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Perccellomeコンソーシアム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および安東朋子、森山紀子、近藤優子、中村祐子、安部麻紀、吉木健太、松田菜穂、森田紘一、今井あや子の各氏の強力なサポート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およびH15-化学-002による。

文献

- 1) Waters MD, et al: *Mutat Res* (2003) 544: 415-424
- 2) Hill AA, et al: *Genome Biol* (2001) 2: RESEARCH0055
- 3) van de Peppel J, et al: *EMBO Rep* (2003) 4: 387-393
- 4) Hekstra D, et al: *Nucleic Acids Res* (2003) 31: 1962-1968
- 5) Sterrenburg E, et al: *Nucleic Acids Res* (2002) 30: e116
- 6) Talaat AM, et al: *Nucleic Acids Res* (2002) 30: e104
- 7) Bolstad BM, et al: *Bioinformatics* (2003) 19: 185-193
- 8) Lee PD, et al: *Genome Res* (2002) 12: 292-297
- 9) Wilson M, et al: *Proc Natl Acad Sci USA* (1999) 96: 12833-12838
- 10) Heller RA, et al: *Proc Natl Acad Sci USA* (1997) 94: 2150-2155