

● 第2章 創薬分野への応用

第2節 毒性評価 トキシコゲノミクス

1. 創薬における 毒性試験とゲノム解析

図1に創薬過程を模式的に示した。旧来の創薬は、図の上のような経過をたどっていた。まず、創薬ターゲットを絞るために、対象疾患を決定する。次に、動物で病態モデルを作成する。このモデルを用いて出発化合物を発見するのであるが、このスクリーニング効率が最も重要であって、候補化合物の合成が律速になることはあまりなかった。通常製薬会社には多くの化合物ライブラリーがあり、また天然物には無限といえる多種多様な成分が存在しているため、スクリーニング効率が律速となっていたのである。

モデル動物による出発化合物の最適化と並行して、動物を用いた候補化合物の毒性評価がなされてきた。問題は、これら非臨床試験の臨床への外挿性である。病態モデルは必ずしも臨床の病態を反映したものとは限らなかった。たとえば現在、消化性潰瘍の主因は *Helicobacter pylori* であることが定説であるが、現在繁用されている潰瘍治療薬は(除菌薬

を除き)感染症モデルを用いて開発されたものではない。幸いなことに、潰瘍治療薬の臨床での有効性は抜群であったが、「非常に有効な、ラットに限定された治療薬」なるものが生み出されてしまい、臨床開発を断念した例は多い。

このような状況にパラダイムシフトをもたらしたのがヒトゲノムプロジェクトである。図1の下は、ポストゲノム時代の創薬スキームを示す。創薬ターゲットとしてヒト型のタンパク質を設定することが可能になり、標的タンパク質が病因に関係している限り、その有効性は保証されている。問題は薬物動態の種差であったが、これも、ヒト型薬物代謝酵素やトランスポーターが容易に入手できるようになり、急速に予測性が改善されてきた。1991年の米国の調査によれば、過去10年間に開発中止となった薬物の中止理由のうち、最大のものは薬物動態(40%)であり、毒性関係は20%程度であったが、2000年の調査では、薬物動態を理由とするものが10%を割り、相対的に毒性関係が40%近くまで達している¹⁾。2000年時点ではヒトゲノムプロジェクトの直接の成果はまだ反映されていないが、網羅的ゲノム解析に先行して、代謝酵素やトランスポーターの解析がなされたことが大きく寄与したのは明らかである。したがって、ヒトゲノムプロジェクトが完了した後は、創薬効率が飛躍的に伸びると誰もが予想していた。

しかしながら、この予想は二つの面で外れることになる。一つは、ヒトゲノムは決して創薬ターゲットの宝庫ではなかったことである。重要な疾患のターゲットの多くは、旧来の手法で既に研究し尽くされており、新たに疾患関与性の濃い因子が見出されても、非常に特殊な疾患のターゲットにしかならないことが多かった。もう一つの期待外れは、副作用に関するものであった。重大な副作用により、市場から撤退する薬物はあとを絶たず、最近ではかえって

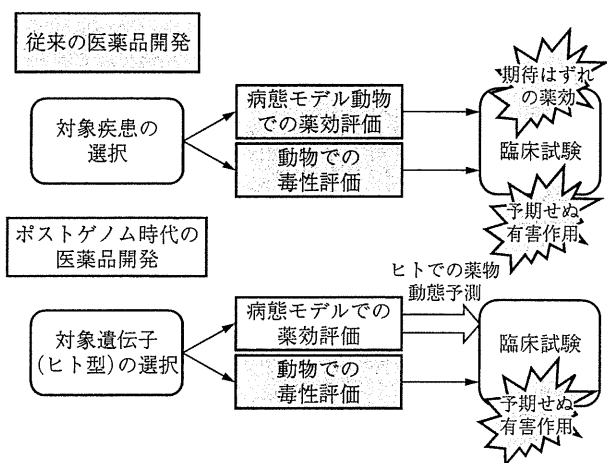


図1 創薬過程の変遷

数が増えている。これは明らかに、薬物作用の特異性・有効性と、使用上の安全性は異なった次元の問題であることを示している。創薬ターゲットは現在ほぼ100% ヒト型が利用できるが、安全性研究はこの進歩を享受できず、相変わらず動物を用いた旧態然の安全性試験を行わざるを得ない。これでは、臨床における予期しない副作用発現の頻度は低下するはずがない。

臨床での医薬品安全性における重要な要素として、いわゆる idiosyncrasy (特異体質) がある。近年問題になった、大型新薬の市場撤退の例として、トログリタゾン²⁾やセリバスタチン³⁾があるが、これらの有害作用の発生率は、数十万例に1例でしかない。数千例の動物実験ですら非現実的であり、種差が全くないとしても、動物試験でこの副作用を予測することは事実上不可能である。それでも安全性予測が可能とするなら、図2のような仮定が必要である。ここで横軸を用量、縦軸を毒性の発生率とすると、臨床用量域Aでは発生率が低すぎて検出できない毒性Xも、毒性発生率が用量依存的に増加するならば、いずれ検出が可能となると期待できる(曲線右)。しかしながら多くの場合、注目している毒性以外にも、多くの臓器において種々の毒性が発現する。もし、その中に動物にとって致死的な毒性があり、その発生率が左側の曲線のようなであれば、毒性Xの検出はきわめて困難となる。用量域Cでは理論的に検出不能であり、用量域Bでも検出はきわめて困難であろう。実際の例としては、非ステロイド系抗炎症薬による肝障害や腎障害がある。臨床の場でこれらの発生は無視できないが、ラットに

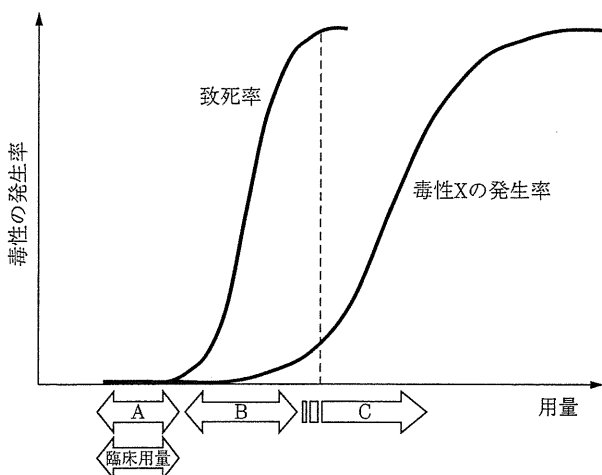


図2 仮想的な用量・毒性曲線

連続投与すると、低用量でも消化管穿孔で死亡するため、典型的な肝障害や腎障害を観察することは難しい。さらに、この議論は、個体差を考慮していない。臨床では、ある特定の遺伝子型や病態をもつ集団に発生リスクが高いことが idiosyncrasy の原因だとすれば⁴⁾、これを反映しないような、遺伝的に均一な動物の実験から予測が不可能なことは自明であろう。

環境中物質や、食品添加物などの場合と異なり、医薬品は生体に働きかけることを前提として積極的に投与するものである。したがって、NOEL (No Observable Effect Level) は原理的にあり得ず、リスクとベネフィットを秤にかけたうえで、NOAEL (No Observable Adverse Effect Level) を超えた量でさえも投与される。薬の専門家でない人々がよく発する言葉に、「副作用のない薬を目指す」というものがある。これは原理的にはあり得ないのであるが、「有害作用を最小とするために、起こりうる有害作用の内容と、起こりやすい集団の性質を前もって把握し、備えておくこと」は重要である。そのためには、毒性発現のメカニズム解析が最大の武器になる。しかし、ターゲットを絞って開発された薬物の作用機序の解析とは異なり、毒性発現の可能性は無限にある。これらすべてを事前にチェックすることは原理的に不可能であると、誰もが考えていた。極端な言い方をすれば、古典的な毒性試験は、「やっても効果はないが、やることに決まっているから義務的に行うもの」という一面をもっていた。しかし、オミクステクノロジーの登場によって、膨大な数のパラメータを網羅的に測定することが可能になり、これこそが、トキシコロジーに最も有用な技術であるとの期待が集まった。

2. トキシコゲノミクスとは

「全体を測定する学問・技術」のオミクステクノロジーのうち、遺伝子全体=ゲノムを扱うゲノミクスを、毒性学=トキシコロジーに応用したものがトキシコゲノミクスである。ゲノミクスの解析手法に関しては第1編4・5章で詳細な解説がなされており、トキシコゲノミクス特異的な方法論があるわけではないので、ここでは創薬における安全性研究に応用

する場合の問題点に絞って論ずる。

トキシコゲノミクスという用語は通常2種類の意味に使われる。一つは「遺伝子型によって毒性発現が異なるという現象」を網羅的に解析する、というものである。薬理学=ファーマコロジーにゲノミクスを組み合わせたファーマコゲノミクス、すなわち、遺伝子型による薬効解析では、臨床における薬物代謝に関連づけた解析の成功例が多く、ファーマコゲノミクスといえ、薬物代謝酵素のSNPsに起因する薬物動態の個体差の解析を意味するかのような勢いを示している。他方、網羅的遺伝子発現解析による毒性研究を指す場合がある。網羅的遺伝子発現=トランスクリプトームであるから、トキシコトランスクリプトミクスと呼ぶべきであろうが、これもトキシコゲノミクスと呼び習わす。

前者のトキシコゲノミクスの場合、これはかなり有望な戦略といえよう。現在各方面で、患者のジェノタイプと病気についてのデータベース構築が進行している。2007年のScience誌においてHuman Genetic Variationがbreakthrough of the yearに選ばれたことでも分かるとおり(<http://www.sciencedaily.com/releases/2007/12/071220140817.htm>)、疾患と遺伝子型の関連は、現在最も注目されている領域の一つである。たとえばある疾患について、疾患関連遺伝子の遺伝子型(主にSNPs)についてロジスティック解析を行えば、数学的に厳密な予測モデルが構築可能である。現在は各病態に対する寄与度の大きい遺伝子が解明途上であるため、個々の疾患例についての将来予測は困難であるものの、ある集団における予測性は向上している。生活習慣病のような多因子疾患の場合には、旧来の診断法と大差ないという批判⁵⁾もあるが、遺伝子の選択とデータの蓄積により、大幅な改良が望める。データ集積は、SNPsアレイの使用により加速されると思われる。

この方法の最大の強みは、遺伝子型には生物学的なばらつきが全くないことである。この方式は、発病率ばかりでなく、医薬品の副作用予測にも応用できよう。しかしながら、動物を用いた非臨床試験には、即座に応用することができない。

3. 網羅的遺伝子発現解析による毒性予測

通常トキシコゲノミクスというと、後者のトランスクリプトミクスを安全性試験に応用する場合を指すことが多い。新たに開発しようとする薬物の有害作用は、薬理作用と異なって、どこにどのようなものが現れるのか全く予想がつかないので、すべてを測定する必要がある。これは、たとえばGeneChipを用いれば、ある組織で発現しているすべての遺伝子を定量することができる。

さて、ある薬物によってある組織のすべての遺伝子発現変化を定量できたとしよう。そのデータを用いて、どのようにして臨床につながる安全性予測が可能となるのであろうか。ここには、以下に示す三つの問題がある。

(1) データ処理の問題

遺伝子型の場合と異なり、網羅的発現解析の場合には、測定値に生物学的なばらつきが伴う。以前は、マイクロアレイを用いた定量には再現性が乏しく、結果は必ず定量PCRなどで確認すべきであるとの意見が多かった。しかし、たとえばGeneChipのような、高度な定量化を達成したシステムの場合、定量性の問題はアレイ側にあるというより、生物側の影響が大きいと思われる。毒性試験は、その性質上第2種の過誤を小さくする必要があるので、あまり第1種の過誤を小さくできない。たとえば数十項目が一斉に測定される血液生化学値のような古典的パラメータは、何ら毒性影響が無くても「統計的有意差」が認められる項目が複数個出現するのが常である。しかしその多くは、統計学的に有意であっても「毒性学的には」有意でないと判断されてきた。しかしながら、網羅的発現解析の場合は、この方法が使えないのである。

ここで、薬物投与による肝臓の遺伝子発現変化を、対照群と処置群それぞれ5例のラットを用いて検出したとしよう。Affymetrix社のGeneChipを用いると、1枚のチップで32000プローブセット(遺伝子数)の測定が可能であるから、32,000組の有意差検定を行うことになる。危険率 $\alpha < 5\%$ で補正なしのt検定を行ったとすると、有意差ありと判定されるペアのほとんどが、全く意味をもたない、偶然の産物である。統計学には多重比較の補正という方

法があるが、たとえば Bonferroni 補正を行って、危険率を $5/32,000=0.00016\%$ としても問題の解決にはならない。「有意差」を示す遺伝子数は激減するであろうが、「毒性学的に有意義な遺伝子の発現変化」の多くは排除されてしまう。このような問題に対して、FDR 法(擬陽性の割合を一定値以下に制御する方法)などが開発されているが、本質的な解決にはならない。

そもそも、いかなる解析を行うにも、解析対象となる項目数とデータ数は同じオーダーでなければならないというのが統計学の基本である。しかし現実には可能な動物実験は各群 10 例以下であり、マイクロアレイのコストを考えると $N=5$ 以下が通常である。数万の項目(遺伝子)に対して、数例のデータでは所詮合理的な解析をすることはできず、解析の前に対象となる遺伝子の絞り込み(フィルタリング)を行う必要がある。しかしながら、どのようにして遺伝子を絞り込むかは、次項の問題とも関連し、この手法の最大の問題点であると言っても過言でない。

(2) 遺伝子発現データと毒性の関連づけ

医薬品の場合、NOEL を投与することはあり得ない。薬効が期待できる用量を投与したとき、有害作用が出るか否かを検討するのが安全性試験であるが、遺伝子発現変化を観察するのみでは、それが毒性か否かを判断することが困難である。

遺伝子発現変化から毒性の評価・予測をする場合、発現変化と病理学的・病態生理学的な表現型(フェノタイプ)を関連づける必要がある。これをフェノタイプアンカーリングというが、トキシコゲノミクス手法における最も重要な課題である。

たとえば薬物 A の肝毒性を評価する目的で、A を投与した動物の肝臓での網羅的遺伝子発現解析を行い、膨大な数の遺伝子の、「生物学的に意味のある」発現変化を観測できたとしよう。このとき、どうすれば遺伝子発現変化から肝毒性の程度を評価できるだろうか。一つの方法は、既知の毒性学的フェノタイプが観測された時点・用量において、これと関連する変化を示す遺伝子を抽出することである。しかしこの戦略では、古典的毒性学を凌駕できない。既知のフェノタイプと完全に関連する遺伝子発現を測定するのであれば、そもそもそのフェノタイプを測定すればよく、わざわざコストのかかる実験をする必要がない。トキシコゲノミクスのメリットを生かすためには、通常の毒性試験より高感度であ

るか、早期に予測できることが必要である。

この2点をクリアする戦略として、あるフェノタイプが現れる用量より低い水準で起こる遺伝子変化、あるいはフェノタイプが現れる時点より以前で起こる遺伝子変化をとらえることにより、高感度・短期間で安全性評価・予測をしようとする試みがある。このことにより、図2で指摘したような問題は回避でき、また、特に発がん実験のように、長期間の投与を必要とする膨大なコストを削減できる可能性がある。しかしこの方式は、遺伝子発現変化と毒性発現の因果関係が保証されないと、意味をなさない。たとえば、Heme oxygenase I(HmoxI)という酵素は、酸化ストレスを引き起こす多くの肝毒性物質によって著明に誘導され、ストレスマーカーとよいてよい。それでは、HmoxI を誘導する物質は、肝毒性物質と分類してよいだろうか。実際この誘導は、多くの医薬品の薬理量で認められる。医薬品は生体に積極的に働きかけるものであり、何らかの応答は避けられない。問題は、実際の治療において、この酸化ストレスが容認できるか否かにある。また、HmoxI はストレスに対する防御反応と考えられ、ストレス下においては、生体にとって有益な応答である。ここに酸化ストレスを起こす薬物 A、B があって両者とも HmoxI を誘導するが、B は別経路を介してその誘導を抑制するとしよう。表面的には薬物 A の方が危険に思えるが、実際は防御機能の働かない B の方が危険であろう。

フェノタイプアンカーリングを低用量・早期時点に拡張したとき、どのようにして診断・予測の精度を保証するかが大きな重要な課題である。ここで最も望ましいのは、判断の基になる遺伝子群が、毒性学的パスウェイ上に適切に配置され、かつ毒性発現機序が説明されることである。しかしながら現時点では二つの大きな障害がある。一つは、遺伝子のアノテーションの不足である。たとえば GeneChip を用い、有意な変動が観測された遺伝子を抽出したとしても、通常そこには実体の不明な遺伝子が多く含まれている。たとえ命名されている遺伝子であっても、生理機能、特に毒性にどう関与しているのかが分かっている場合はほとんどないのが現実である。抽出された多くの遺伝子を機能分類するには、Gene Ontology Analysis(たとえば、DAVID <http://david.abcc.ncifcrf.gov/>)が使用されるが、これで生理機能を推定するのは困難であり、参考程度の情報

しか得られない。ここでの最大の障害は、毒性学的パスウェイに関する決定的な情報不足である。むしろ、トキシコゲノミクス手法により、毒性学的パスウェイが詳細に明らかにされ、これが毒性診断・予測の精度を上げていく、という正のスパイラルが望まれている。

(3) 種差のブリッジング

最近マイクロドージングの導入により、創薬の早期から候補化合物を人体に投与することが可能となった。しかしながら、これで得られる情報は薬物動態学的性質が主であり、直接的な毒性に関しては人体実験が不可能であることから、何らかの代替法を考えねばならない。トキシコゲノミクス手法を用いた有望な方法としては次の①～④があるが、いずれも決定的なものではない。

① 培養細胞を用いる方法

ヒト由来であっても、培養細胞を用いた実験では、*in vivo* と *in vitro* の差が大きすぎ、現在の技術ではいまだ使用に耐えない。直接の細胞毒性であれば、外挿は可能であろうが、臨床における有害作用は、血流や神経を介したものが多い。また、薬物それ自身ではなく代謝活性化を受けたものが毒性の本体である場合が多いが、一次培養肝細胞であっても、生体のような薬物代謝は再現できない。*in vitro* で再現可能な毒性学的パスウェイをなるべく多く見いだして、これを参照するという使い方が現状ではベストであろう。

② ヒト型臓器をもった動物を用いる方法

これにはマウスなどを利用した二つの方法がある。一つは、ヒト型遺伝子を組み込む(相同遺伝子と置き換えるのが望ましい)方法であり、他方は、ヒトの臓器を移植・定着させる方法である。前者は、すべての関連遺伝子をヒト化することが困難なこと、プロモーター領域込みでヒト化しないとモデルとして成立しないことなどの欠点があり、特定の遺伝子の関与を検討する場合にはよいが、一般的なスクリーニングにはまだ程遠い。後者は、免疫不全マウスの肝臓をヒト肝細胞で置き換えたマウスが作られているが⁶⁾、高価であること、免疫不全状態という特殊な環境下での毒性発現が臨床に外挿できるか、など問題が多く、これも一般化にはまだ遠い。

③ リンパ球を用いたトランスクリプトーム

たとえばラット肝臓の遺伝子発現解析によって、ある医薬品の毒性が良好に予測できたとしても、よ

ほどのことがない限り、患者の肝臓の生検を行って比較することはできない。非臨床データを臨床につなげるには、非侵襲的サンプルで判断することが望まれる。これには、非臨床データから予測される血中あるいは尿中の代替マーカーを提案するか、生検しなくても得られるサンプルの遺伝子発現を測定するしかない。前者の場合、薬物の毒性発現に関連して或る遺伝子の発現が変化したとしても、その遺伝子産物が血中に出てこない限り検出できない。それが分泌されるものなら、毒性に関連した発現変化なのか分泌自体の変化なのか、判断がつかない場合がでてこよう。非分泌型のものであれば、血中で検出されるということは細胞の破壊を意味するから、古典的な逸脱酵素の検出と変わらないことになる。ただし、逸脱酵素よりはるかに感度と特異性が高ければ、血中への臓器特異的 mRNA の漏出を測定する意味はある⁷⁾。

後者の候補としては、血液中のリンパ球がほとんど唯一のものである。薬物投与時に全身で起こる有害作用が、直接あるいは間接的にリンパ球の遺伝子発現変化として反映されることは十分考えられる。実際、ラットにおいて、アセトアミノフェン投与により薬物特異的な遺伝子変化が検出できたとの報告があり⁸⁾、また、臨床研究も進められている。

④ 毒性学的パスウェイを介したブリッジング

これまでの経験では、動物細胞において薬物の毒性に特徴的な遺伝子群の変化が認められたとしても、これらを単純にオルソログ変換したものは、ヒト細胞における発現変化を示す遺伝子群との一致率は高くない。しかし、個々の遺伝子どうしの対応は低いにしても、毒性物質に対する生理・病理的応答は共通性が見いだせる場合が多いようである。そこで、動物実験(*in vivo*, *in vitro*)で得られた結果を、毒性学的パスウェイ上にマッピングして、「ヒトにおいても、このようなメカニズムで毒性の発現する危険がある」ということを提唱することが、最も確実な戦略であろう。ただしこれを実現するには、いわゆるシステムズバイオロジーの飛躍的な進歩が必要である。

(4) 世界の現状とトキシコゲノミクスプロジェクト

遺伝子全体を扱うと、その数が膨大であるために、データ数も大きなものが必要となってくる。必然的に、小さな研究室単位での研究は困難で、巨大なデータベースの構築が必要であることから、コン

ソーシウム形式で、疾患関連遺伝子や薬物代謝関連遺伝子の SNPs データベース構築が世界中で推進されている。

毒性学的トランスクリプトミクスの場合、プラットフォームやプロトコルの違いによりデータの互換性が保証されない場合が多く、多施設で得られたデータを *in silico* で統合しても利用価値の高いデータベースとなるかどうかは疑問である。しかし、コストを考えると、アカデミアでは達成不可能で、欧米の巨大製薬会社でようやく可能となる程度のものである。トキシコゲノミクス技術が開発された当時、わが国の創薬の国際競争力を担保するためには、欧米に伍するデータベース構築が必要であるとの考えから、国衛研と製薬 17 社による官民共同プロジェクトであるトキシコゲノミクスプロジェクトが 2002 年に発足した。これは、安全性試験に使用されるラットに 150 種の医薬品を中心とする化学物質を投与し、肝臓(一部腎臓)の遺伝子発現変化を GeneChip で測定してデータベースとするものであった。このデータベースの特長は、同一プラットフォームで質の高いデータを取得したことと、プロトコルとして用量水準(溶媒対照, 低, 中, 高用量, $N=3$)と時点(単回投与後 3, 6, 9, 24 時間, 連続投与 3, 7, 14, 28 日, $N=3$)を豊富にもつことが挙げられる。用量・時点が豊富であることのメリットは、既報を参照されたい⁹⁾。

プロジェクトではこれ以外に、ヒト凍結肝細胞、ラット一次培養肝細胞でのデータ(溶媒対照, 低, 高用量, 処置後 8, 24 時間, $N=2$)での暴露実験を

行い、一化合物あたり総計 120 枚の GeneChip(腎臓も解析した化合物の場合は 216 枚)を費やしている。これら膨大なデータは古典的毒性学パラメータや病理とひもづけた形でデータベース化し、解析システム、予測システムを組みこんだ統合システム TG-GATEs (ToxicoGenomics project: Genomics Assisted Toxicity Evaluation System) として結実した⁹⁾。2007 年からは、後継プロジェクトにおいて、TG-GATEs を活用することにより、安全性バイオマーカー創出の努力が続けられている (<http://www.tgpnibio.go.jp/>)。現在はバイオマーカー候補の提案の段階であるが、実例の詳細は既報を参照されたい^{10~12)}。

4. おわりに

わが国におけるトランスクリプトームのデータベースとしては、国衛研毒性部によるマウスの各臓器における化学物質の影響、NEDO によるラットの肝発ガンデータベース(一部 GEO に登録)などがある。表 1 に、現在アクセスできるデータベースを示すが、米国においては現在、製薬企業が中心になったトキシコゲノミクスコンソーシアム、欧州では、オランダが中心となったコンソーシアムが動いている。

2005 年、FDA が新薬の申請時にゲノミクスデータのボランティアサブミッションを開始したことか

表 1 各種データベースと URL

データベースの名称	URL
Array Track	http://www.fda.gov/nctr/science/centers/toxicoinformatics/ArrayTrack/
ArrayExpress	http://www.ebi.ac.uk/microarray-as/ae/
Chemical Effects in Biological Systems (CEBS)	http://cebs.niehs.nih.gov/cebs-browser/cebsHome.do
Center for Information Biology gene EXpression database (CIBEX)	http://cibex.nig.ac.jp/index.jsp
Comparative Toxicogenomics Database (CTD)	http://ctd.mdibl.org/
dbZach	http://dbzach.fst.msu.edu/
Gene Expression Omnibus (GEO)	http://ncbi.nlm.nih.gov/geo/
Percellome Project	http://www.nihs.go.jp/tox/TTG_Archive.htm
PharmGKB	http://www.pharmgkb.org/
SYMATLAS	http://symatlas.gnf.org/SymAtlas/
Tox-MIAMExpress	http://www.ebi.ac.uk/miamexpress/

ら、この分野に注目が集まった(<http://www.fda.gov/Cder/genomics/default.htm>)。またFDAは臨床研究におけるバイオマーカーの重要性を発信しており、今後ますますこの分野における競争が激化すると思われる(<http://www.fda.gov/oc/initiatives/criticalpath/>)。創薬における安全性研究は、トキシコゲノミクス手法に限らず、最先端の科学知識・技術を可能な限り広範囲から取り入れ、統合していく必要があり、一層の進歩が期待される。

【引用・参考文献】

- 1) I. Kola and J. Landis: *Nature Rev. Drug Discov.*, **3**, 711(2004).
- 2) M. M. K. Ong, C. Latchoumycandane and U. A. Boelsterli: *Toxicol. Sci.*, **97**, 205(2007).
- 3) 斉藤充生, 平田睦子, 三宅真二, 長谷川隆一: *Bull. Natl. Inst. Health Sci.*, **123**, 41(2005).
- 4) J. B. Meigs, P. Shrader, L. M. Sullivan, J. B. McAteer, C. S. Fox, J. Dupuis, A. K. Manning, J. C. Florez, P. W. F. Wilson, R. B. D'Agostino and L. A. Cupples: *N Engl J Med.*, **359**, 2208(2008).
- 5) R. A. Roth, J. P. Luyendyk, J. F. Maddox and P. E. Ganey: *J. Pharmacol. Exp. Ther.*, **307**, 1(2003).
- 6) K. Ohashi, P. L. Marion, H. Nakai, L. Meuse, J. M. Culen, B. B. Bordier, R. Schwall, H. B. Greenberg, J. S. Glenn and M. A. Kay: *Nature Med.*, **6**, 327(2000).
- 7) Y. Kudo, T. Ochi, H. Shimada and K. Shinjo: *Vet. Med. Sci.*, **70**, 993(2008).
- 8) P. R. Bushel, A. N. Heinloth, J. Li, L. Huang, J. W. Chou, G. A. Boorman, D. E. Malarkey, C. D. Houle, S. M. Ward, R. E. Wilson, R. D. Fannin, M. W. Russo, P. B. Watkins, R. W. Tennant and R. S. Paules: *Proc Natl Acad Sci USA*, **104**, 18211(2007).
- 9) T. Urushidani: Hepatotoxicity from genomics to in vitro and in vivo models. ed by S. C. Sahu, Wiley, pp. 507-529(2007).
- 10) T. Uehara, N. Kiyosawa, T. Shimizu, K. Omura, M. Hirode, T. Imazawa, Y. Mizukawa, A. Ono, T. Miyagishima, T. Nagao and T. Urushidani: *Human Exp. Toxicol.*, **27**, 23(2008).
- 11) M. Hirode, A. Ono, T. Shimizu, T. Nagao, Y. Ohno and T. Urushidani: *Toxicol Appl Pharmacol.*, **229**, 290(2008).
- 12) T. Uehara, A. Ono, M. Hirode, N. Kiyosawa, K. Omura, T. Shimizu, Y. Mizukawa, T. Miyagishima, T. Nagao and T. Urushidani: *Toxicology*, **250**, 15(2008).

〈漆谷 徹郎〉

20

Prediction of Hepatotoxicity Based on the Toxicogenomics Database

Tetsuro Urushidani

20.1 Introduction

Today, in the post-genomic era, there have been remarkable advances in the technology of searching for seeds of new drugs. However, the success rate of drug development is nevertheless decreasing not only in Japan but also worldwide. According to the investigation performed by the Japan Pharmaceutical Manufacturers Association (<http://www.jpma.or.jp/12english/publications/index.html>), 422 653 seeds for new drugs were synthesized or extracted from 1996 to 2000 in Japan, whereas only 63 were approved. Moreover, half of those approved were chemicals created by and purchased from foreign countries. It is obvious that the opportunity to create an original drug is now extremely low.

Drug development in the previous century was usually based on screening by measuring effects of the chemicals in model animals with artificially created diseases, and subsequently it sometimes happened that a quite excellent drug was produced not for humans but for rats. In recent years, however, it has been possible to start the development by targeting disease-related genes whose molecular functions are well elucidated, and indeed, human-type genes are always available. Therefore, it is now easy to select a chemical, which is effective on the human-type molecule for at least the *in vitro* level. Even with this advantage, many candidate drugs have been eliminated largely because of toxicity, which could not be found in pre-clinical tests but appeared in clinical trials (Ismail and Landis, 2003). This not only brings about an increment in the cost for drug development that will be shifted to the medical expense, but also causes a serious ethical problem that the toxicity of the chemical is proven by using humans. From another

point of view, the low predictivity of present toxicological tests for clinical toxicity also means that many drug candidates might have been abandoned although they were in fact excellent.

The current toxicological tests are much improved and controlled to assure reliability of the data in comparison to past adverse cases, such as thalidomide. However, reliability of the pre-clinical data does not necessarily lead to clinical safety. When the drug possesses an obvious organ or cellular toxicity as part of its pharmacological effects, its toxicity can be easily predicted from its dose-response relationship and thus controlled. However, many of the practical side effects in the clinical field are scarcely related to the pharmacological target and the origin of the toxicity is usually not the drug itself but its metabolite(s). If a serious adverse effect happened in the clinical dose range at a rate of 1/1000 (an impractically large number), such a drug should be withdrawn from the market at once. However, it is theoretically impossible to detect the effect with such a low incidence using 10 000 rats or more, even though there is no species difference, and this is, of course, still an impractical number for safety tests. Needless to say, the barrier of species difference is quite high and we sometimes encounter a case that there is a difficulty to find out an animal species accurately reproducing a clinically evident toxicological profile. In practical toxicological tests, we inevitably employ the hopeless assumption that any pathophysiological changes in the experimental animals associated with the over-dosage of the test drug can be extrapolated to the corresponding adverse effects with low incidence in humans at low dose. Many conscientious toxicologists feel that such a preclinical safety evaluation is not a certification but an excuse.

In order to 'predict the unpredictable toxicity', the most realistic strategy would be 'toxicogenomics' which enables us to analyze all of the gene expression changes caused by an external stimulus in an organism. Even when an 'unexpected' hazard is elicited by drug administration at a very low incidence, its causal events should have occurred and some of them should be detectable as a subtle change. If all of the events were observed and their relationships were analyzed, a perfect prediction would be realized. Observation of all of the physiological processes had been thought to be practically impossible, but advances in biotechnology have made it possible. Since the most sensitive change among biological processes is that of mRNA, and many of the physiological responses are expected to be associated with changes in the expression level of mRNA, the most promising measure is the amount of mRNA. The development of microarray technology has made it possible to substantially quantify the expression level of all of the genes at once.

Under such a circumstance, the Ministry of Health, Labour and Welfare, National Institute of Health Sciences (NIHS) and the working group of the Japan Pharmaceutical Manufacturers Association planned 'The Toxicogenomics Project (TGP)', a collaborative project of the government and private companies (Urushidani and Nagao, 2005). This is a 5-year project from 2002 to 2007, and after the rearrangements of the organization and mergers of the companies, the members are the NIHS, 15 pharmaceutical companies (Astellas, Chugai, Daiichi, Dainippon-Sumitomo, Eisai, Kissei, Mitsubishi, Mochida, Ono, Otsuka, Sankyo, Sanwa, Shionogi, Takeda and Tanabe) and the National Institute for Biomedical Innovation (NIBIO) as the core institute where the actual work is being performed. About half of the budget is from a grant of the Ministry of Health, Labour and Welfare, and the remaining is from the companies.

20.2 The Features of The Toxicogenomics Project (TGP)

The plan of the project at the beginning was as follows. The main goal is to create a gene expression database of 150 chemicals, mainly medical drugs. The target organ is mainly the liver. The reason was that most of the clinically serious adverse effects have occurred in this organ, and that the composition of the cell type is relatively homogenous in this organ and thus the expected variation based on differences of sampling would be minimal. Subsequently, the liver was considered to be favorable to accumulate the know-how of the toxicogenomics technique. Nephrotoxicity was of course considered to be important and the sampling and pathological examination of the kidney in addition to liver was to be performed in all of the animals. Among the animals showing nephrotoxic phenotypes, up to 20 % of the total compounds were subjected to transcriptome analysis of the kidney.

For the test animal, 6 week- old male Spraque–Dawley rats were employed. At the starting point, the rat genome was not fully analyzed and there was a firm opinion that the mouse should be used because of the enrichment of the genome information. However, all of the classical toxicological tests had been done with rats and there was a vast amount of knowledge accumulated. When we were venturing into a totally new field, we believed that the accumulated knowledge should be used as much as possible. Another advantage was that more data, such as biochemistry and hematology, could be obtained from the rat rather than from the mouse.

The biggest problem at the start of the TGP was the fact that we were behind the large pharmaceutical companies and bioventures in Europe and America who had already started similar projects to create databases of toxicogenomics (Porter *et al.*, 2003; Boverhof and Zacharewski, 2006). Therefore, the following strategy was employed to catch up with the preceding projects. (1) Establishment of quantitativity and reproducibility. Data with excellent quantitativity are acquired using the ‘Affymetrix GeneChip’ and a new normalization method based on the externally added spike RNAs proportional to the sample DNA contents (Kanno *et al.*, 2006) was employed. (2) Selection of test compounds. The chemicals, mainly medicinal, contain drug candidates withdrawn from development because of their hepato- or nephrotoxicity, that are supplied from the member companies. (3) Bridging between species. In addition to the *in vivo* experiments, exposure to the primary culture of rat liver and to human hepatocyte culture was performed. (4) Enrichment of protocol. Gene expression data possess multi-dose, multi-time points that link to various toxicological measures. These points are discussed in this order below.

20.2.1 Establishment of Quantitativity and Reproducibility

The ‘Affymetrix GeneChip’, an oligonucleotide array, is known to be superior to other arrays, such as the ‘Stanford’ type (Wildsmith and Spence, 2003). In the TGP, data with excellent quantitativity are being accumulated under thoroughly optimized and controlled conditions in order to draw their full ability. During the first two years (up to ca. 35 chemicals), the ‘Rat Expression Array 230A’ containing 15 923 probe sets was used while it was then shifted to the ‘Rat Genome 230 2.0 Array’ containing 31 099 probe sets after the version upgraded.

The TGP started with employing a new normalization method, ‘percellome’, using an externally added spike RNA (Kanno *et al.*, 2006). This method makes it possible to express

each gene expression as copy numbers per cell (per DNA) by adding external *Bacillus subtilis* mRNA proportional to the DNA contents in the homogenate. In our database, raw data, per-chip normalized data as well as spike-normalized data, are usable.

Per-chip normalization (global normalization) is fundamentally based on the assumption that the total amount of mRNA is constant, and thus the change of each mRNA cannot be precisely estimated when the total transcription is drastically changed (Kanno *et al.*, 2006). In the TGP, the version of 'GeneChip' was changed and this made it impossible to make a comparison between different chips based on values with global normalization. Moreover, in the *in vitro* experiments where extremely high concentrations of the chemicals were applied, drastic changes of total mRNA were often observed. Especially, when a chemical having direct cytotoxicity to hepatocyte was applied, total mRNA was obviously decreased, and many genes, which were actually down-regulated, were estimated as apparently up-regulated because of normalization by the reduced value. In analysis of the kidney, the cell composition varies with the particular part of the organ. It was found that not only the members of expressed genes but also the total mRNA amounts differed among cortex, medulla and papilla. When region-specific genes were extracted, it was revealed that global normalization led to a biased conclusion because of the large difference in the total mRNA (Tamura *et al.*, 2006a).

Contrary to the above facts, we concluded that there is no problem in the usual analysis based on global normalization, at least *in vivo* liver data, since total RNA contents were almost unaffected, even at the toxic dose. When the ratio to control value is employed, global normalization is rather superior to spike normalization because the extra procedure introducing an additional error is avoided. In this present paper, analyses are based on global normalization.

20.2.2 Selection of Test Compounds

The TGP started with five representative hepatotoxicants, i.e. acetaminophen, isoniazid, carbon tetrachloride, phenobarbital and valproic acid. The 150 chemicals were selected and their exposure to rats have been completed – these are categorized in Figure 20.1. Although they are somewhat biased, they cover most of the therapeutic categories. At present, the drugs in the Japanese market number about 1500, including those not suitable for transcriptome analysis of the liver or kidney, e.g. drugs for dermatology or ophthalmology, drugs with almost identical structures and anti-cancer drugs whose toxicity to these organs is not a primary problem. Our aim is to create a database of representative medical compounds. In this sense, the number of 150, 10% of total drugs, is considered to be enough for a 'textbook of toxicology'. It would be interesting to extend the project toward a structure-toxicity relationship based on the present database.

One unique feature of our project is the supply of chemicals from the member companies. These are drug candidates that were withdrawn from the various stages of the development process because of the emergence of toxicity in the liver or kidney. These are usually impossible to obtain, and are quite valuable and interesting samples. The candidates that were withdrawn mean that they were once considered to be hopeful candidates; in other words, their potential toxicity was underestimated in the early stage. Thus, they could be useful for the evaluation of our database as good model cases after our system is established.

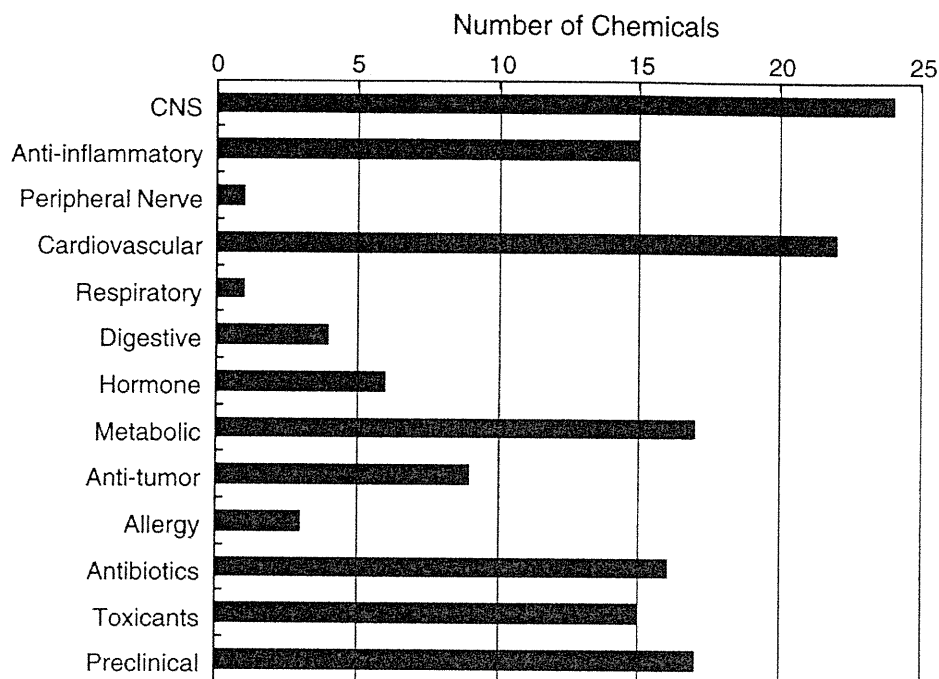


Figure 20.1 Chemicals selected (150 total). The chemicals selected for the Toxicogenomics Project are classified by their category: CNS, drugs for central nervous system; toxicants, not medical drugs but representative chemicals with hepato- or nephrotoxicity, such as carbon tetrachloride, hexachlorobenzene, aryl alcohol, etc; preclinical, drug candidates supplied from the member companies, which were withdrawn in various stages of drug development because of hepato- or nephrotoxicity

20.2.3 Bridging between Species

Even if the perfect prediction of hepatotoxicity is attained in the rat, this does not necessarily mean the improvement of prediction of toxicity in humans. The final goal should be the prediction in humans for drug development. There are too many problems in extrapolation of phenotype from rodent to human. However, if general toxic mechanisms or toxicological pathways are conserved over species, they would obviously be useful for bridging between animal models and clinical events. In the preliminary studies performed before the project, it was found that the gap between hepatocyte cell lines and normal liver were too wide to bridge even if they were both human, and thus normal human hepatocyte culture was employed as the only choice. At present, data collection has been completed for rat primary hepatocytes, whereas that of human hepatocytes has not been done yet, and their comparative analysis is a future subject.

In the system of analysis for the TGP database described later, the object is to overview the responsiveness of the test compounds to the biomarker gene lists and make it easy to compare them between rat and human by an automatic ortholog conversion of the marker genes. However, there have been many difficult problems pointed out in the bridging between *in vivo* and *in vitro* experiments before solving the species difference (Boess *et al.*, 2003). We also recognized this problem as the *in vitro* data accumulated. Although our final goal should be the prediction of clinical toxicity, we are now focusing on the increment of the efficacy of preclinical study as a more realistic goal. Namely, the increment of

predictivity of hepatotoxicity in rats is the first important step and thereafter the bridging between species difference is envisioned as the next step. Therefore, this topic is omitted from this present paper.

20.2.4 Enrichment of the Protocol

This point is the largest merit of the TGP and is thus discussed in detail.

20.2.4.1 Standard Protocol and Problems in Statistical Analysis of Gene Expression Data

The protocol employed in the TGP is summarized in Table 20.1. *In vivo* experiments consist of single and repeated oral administrations to rats ($N = 5$ for each group; 3 doses + vehicle control) and autopsy is done 3, 6, 9 and 24 h after the single dose and 24 h after repeated dose for 3, 7, 14 and 28 days. Data of blood biochemistry, hematology and histopathology (both liver and kidney) are obtained from all animals and the gene expression analysis in liver (also kidney in some cases) is performed in 3 out of 5 rats. *In vitro* experiments consist of rat and human hepatocytes ($N = 2$ for each group; 3 concentrations + vehicle control) and cell harvest is done at 2, 8 and 24 h after exposure.

Table 20.1 Standard protocols employed in the Toxicogenomics Project

<i>In vivo</i>	
Animal	Sprague–Dawley rat 6 week old $N = 5$ for each group
Vehicle	0.5% methylcellulose or corn oil
Dose	Low, middle, high (1:3:10)
Route	Oral (intravenous in a few cases)
Sacrifice	3, 6, 9 and 24 h after a single administration 24 h after the last dose of repeated administration for 3, 7, 14 and 28 days
Sampling	Liver, kidney, plasma
GeneChip analysis	$N = 3$
Items examined	Histopathology: liver and kidney Body weight, organ weight (liver and kidney), food consumption Biochemistry, hematology: 37 standard items
<i>In vitro: rat</i>	
Animal	Sprague–Dawley rat 6 week old
Cell	Hepatocyte isolated by collagenase digestion
Vehicle	Culture medium or DMSO
Concentration	Low, middle, high (1:5:25)
Treatment	2, 8 and 24 h
GeneChip analysis	Duplicate
Items examined	Cell biability (LDH release and DNA contents)
<i>In vitro: human</i>	
Cell	Human frozen hepatocytes
Vehicle	Culture medium or DMSO
Concentration	Low, middle, high (1:5:25, low is omitted in some cases)
Treatment	2, 8 and 24 h
GeneChip analysis	Duplicate
Items examined	Cell biability (LDH release and DNA contents)

We believe that no other database contains data with such an enriched protocol in the world. Especially the fact that dose-dependency with $N = 3$ for each time point can be estimated is very powerful in erasing 'noises' inevitably associated with the statistical analysis of vast numbers of measurements. Although the price of 'GeneChip' is decreasing as it becomes more popular, the cost of the experiments using this device is still so high that either the sample numbers in a group, time points, or dose levels need to be reduced. However, the reduction of the dose levels ruins the analysis. The analysis of microarray data is usually performed by multivariate statistical techniques such as hierarchical clustering, k -means clustering, self-organizing map or principal component analysis (Kaminski and Friedman, 2002; Draghici, 2003). When the data sets are divided into 'positive' and 'negative' by any definition, discriminant analysis such as PAM (Tibshirani *et al.*, 2002) or SVM (Brown *et al.*, 2000) can be used. However, in any case, no confident results are obtained unless the size of the gene list is reduced to a reasonable level.

Suppose only one 'toxic' dose was tested and compared with its control. When using the A chip, 15 923 pairs should be made and 31 099 pairs for the v.2.0 chip. Even when the p value is set to a very low value, too many 'significant' differences without biological significance are obtained. On the other hand, when one does not want to miss a certain gene with high significance in terms of biology or toxicology, the p value cannot be lowered in the experiments with such small N values. It has been pointed out that application of the usual biostatistics is difficult when the evaluation is done with very small N values against vast numbers of measurements such as whole genes. Especially when the data contain biologically based variations, no improvement is expected by the use of any sophisticated statistical technique. When data with multiple doses or time point are available, however, the extraction of biologically meaningful changes is easier.

20.2.4.2 Example 1: The Case of Omeprazole

Let's refer to the actual data. The first one is omeprazole, which was administered to rat at 100, 300, and 1000mg/kg and analyzed with the A chip (15 923 probe sets). In the TGP database, there are three dose levels with vehicle control for each of four time points in single and repeated doses, respectively. When comparison between control and treated rats is made in each point, it will be 380 740 pairs. When an uncorrected Student's t -test is applied, 36 883 pairs of 'significant at $p < 0.05$ ' are obtained. No one would like to perform a precise analysis of these probe sets and it is practically impossible. It is common sense in statistics that this number is overestimated and some correction is needed. It is necessary to erase the accidental difference not related to the drug effect, but the usual statistics only tells you to reduce the p value based on some criteria. A simple reduction of the p value to 0.001 results in the reduction of the 'significant' pair numbers to 1639, which is still a quite large number, but the main question here is whether this has elucidated the really biological meaningful differences. Regrettably, the answer is no, in most cases.

Suppose 'genes significantly changed by 100 mg/kg omeprazole at any time point' were extracted by the t -test without correction. Figure 20.2(a) shows an arbitrarily selected probe set, 13 938 56 \times . At $p < 0.05$, this gene is considered to be up-regulated at 3 h whereas it is down-regulated at 9 h after administration. However, when all of the data are reviewed (Figure 20.2(b)), it is quite easy to conclude that no biologically meaningful change is caused by the drug, since there is no dose- or time-dependency. This type of extraction only creates a 'heap' of 'junky' genes.

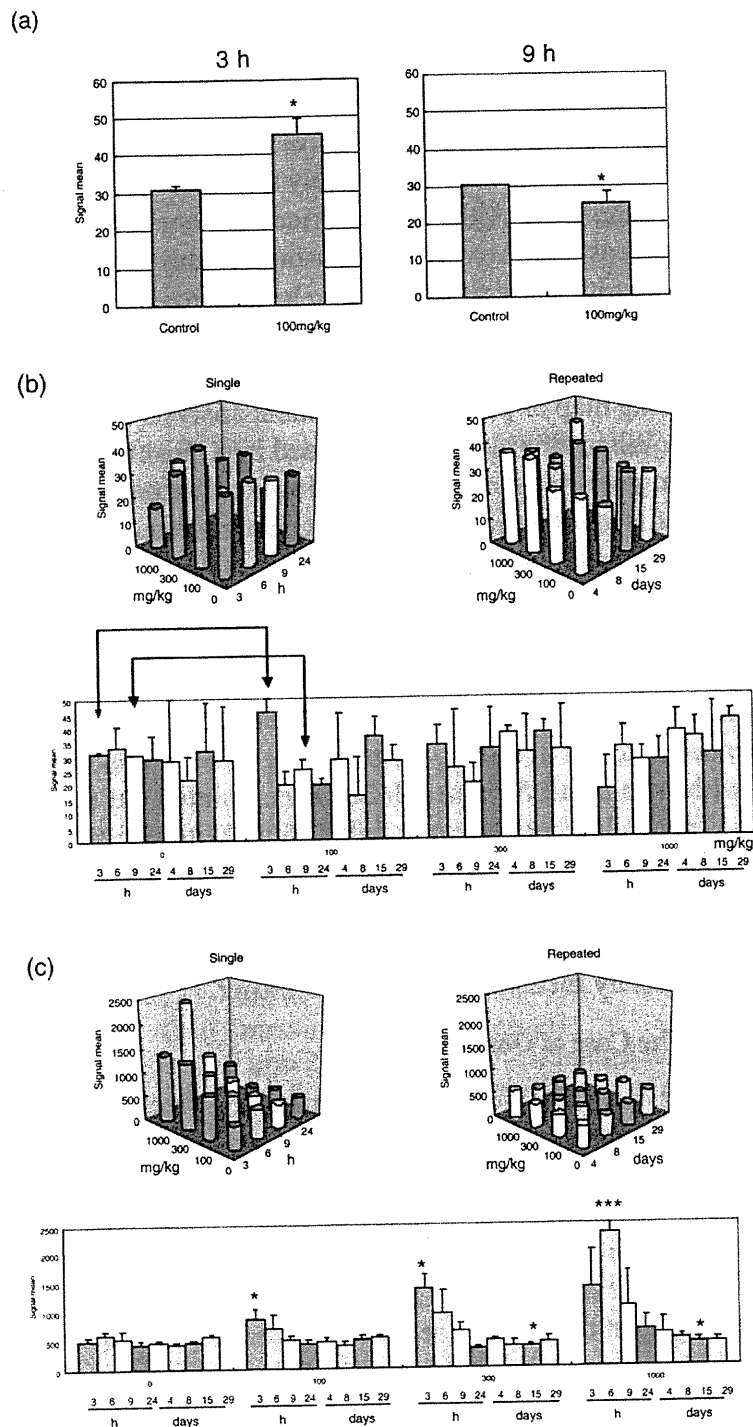


Figure 20.2 Effects of omeprazole on gene expression. (a) Effects of a single oral dose of omeprazole (100 mg/kg) on the expression of an arbitrarily selected probe set, 1393856x at 3 and 9 h after dosing. (b) Effects of single and repeated oral doses of omeprazole (100, 300, 1000 mg/kg) on the expression of 1393856x. The upper panels show the 3-dimensional graphs of single (left) and repeated (right) dosing, while the lower panel shows the 2-dimensional graph with error bars. The arrows indicate where 'statistical significance' was noted in Figure 20.2(a). (c) Effects of single and repeated oral dose of omeprazole (100, 300, 1000 mg/kg) on the expression of heme oxygenase-1 (1370080.at). The upper panels show the 3-dimensional graphs of single (left) and repeated (right) dosing. The lower panel shows the 2-dimensional graph with error bars. Expression data were normalized by mean value and multiplied by 500, and expressed as the mean of $N = 3$ with SD where indicated. The asterisks indicate 'statistical significance' by the uncorrected Student's t-test at $*p < 0.05$ and $***p < 0.001$

Next, attention is paid to heme oxygenase-1, which is known to play important roles in cellular stress response (Figure 20.2(c)). These data obviously indicate that omeprazole potently and dose-dependently induces this gene at 3 h after dosing. Here, if gene extraction is performed by using only the highest dose (1000 mg/kg), only the change at 6 h should be noted. An even worse point is that one would conclude that this important gene is not responding to 1000 mg/kg omeprazole if extraction is performed without using the data of 6 h. Looking at Figure 20.2(c), most researchers would agree that omeprazole induces the expression of heme oxygenase-1. However, it is highly possible that omeprazole is regarded as a negative inducer of this gene by experiments with reduced data points and restricted statistics.

20.2.4.3 Example 2: Age-Related Difference in Toxicity

Before the start of the TGP, there was an argument regarding the age of rats. One asserted that 6-week old rats should be used because this age is recommended in the standard toxicity tests, whereas the other claimed that confident data with small variations would be only obtained from mature animals not younger than 10 weeks. Before reaching a conclusion to use 6-week old rats, a study to compare 6- and 12-week rats was performed regarding the sensitivity of hepatotoxicity.

Acetaminophen, isoniazid and carbon tetrachloride were selected as representative hepatotoxicants and the sensitivity to those were compared between 6- and 12-week old rats by a single administration protocol. Although the latter two showed no age-related difference, acetaminophen was found to be more toxic in 12-week old rats than in 6-week old rats. The causal factors were suggested to be the higher expression of CYP3A13 that produces an active metabolite and/or the lower expression of a subtype of glutathione transferase (Morishita *et al.*, 2006). The most interesting findings in this study were the following points. In order to compare gene expression changes at 24 h after dosing where pathological changes emerged, differentially expressed genes between 6- and 12-week old rats were extracted by statistics. In the usual course of investigation, one would try to attribute the difference in toxicity to the expression level of these genes by using correlation or discriminant analysis. However, a precise review of these genes revealed other features. Among the stress-responsive genes, many genes showed age-related difference not in the extent but in the time course.

Figure 20.3 shows again a representative stress responsive gene, heme oxygenase-1. In 6-week old rats, the peak of induction by acetaminophen appeared at 9 h or earlier and the expression returned to basal level at 24 h, whereas the peak was later than 9 h in 12-week old rats. If the observation were done at 24 h only, the expression of heme oxygenase-1 would have been judged as 'yes' in 12-week old rats and 'no' in 6-week old rats by an 'all-or-none' manner. However, the actual response was 'yes' in both cases and the difference was present in the response time. Other information from Figure 20.3 is that a threshold dose of acetaminophen exists in the induction of at least this gene. This is reasonable, considering the widely accepted toxic mechanism of this drug, i.e. the hepatocyte damage does not occur when glutathione is not depleted and the detoxification system for active metabolite is active (James *et al.*, 2003). In the present case, the induction of this gene would be undetectable unless the test with the highest dose was performed.

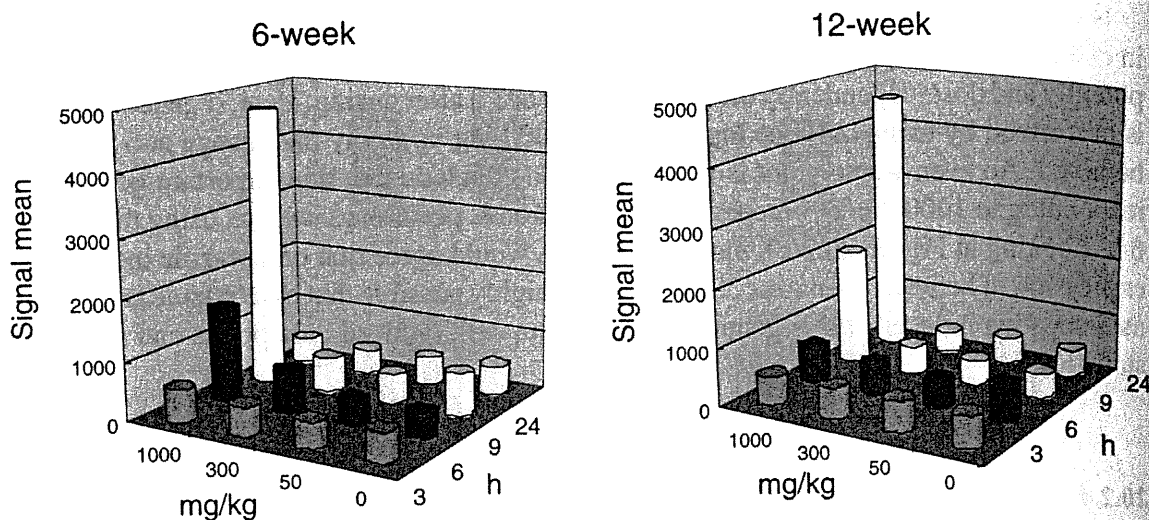


Figure 20.3 Effects of a single oral dose of acetaminophen (50, 300, 1000 mg/kg) on the expression of heme oxygenase-1 (1370080_at). The left and right panels show the data of 6-week old and 12-week old rats, respectively. Expression data were normalized by mean value and multiplied by 500, and expressed as the mean of $N = 3$

20.2.4.4 Advantage of Multi-Time, Multi-Dose Protocol

One would expect that the sensitivity of transcriptome analysis should be always higher than any traditional toxicological technique. Who wants to do additional, quite expensive and tedious tests when obvious toxicological phenotypes are already obtained? However, the previous example tells us that this hope is too optimistic, especially when a toxicological threshold may exist in a certain step of a sequence of gene expression, as in phenotypes. There would be two alternative answers. One is that prediction of toxicity by toxicogenomics technology is only possible in a chronological manner where the administration of toxic doses is essential. In this case, the prediction would be 'Keeping this dosage will cause that phenotype, etc'. The other is that the toxicity of overdose with apparent threshold can be indirectly predicted even at the low dose by gene expression changes somewhere in the toxicological pathway if threshold does not exist in that step. This point will be discussed later.

Figure 20.4 depicts a schematic expression of time- and dose-dependencies of gene expression changes. Needless to say, the conclusions drawn from using multiple components with different time- and dose-dependencies must differ among the cases, where a single observation each is made at time point A or B, or at dose level X or Y. In the case of the usual biology test focusing on a particular target, it would be possible to set an appropriate time and dose in a preliminary study. However, toxicogenomics is designed to observe any gene expression changes reflecting 'any toxic phenomena in the future' one chip at a time, and so it is practically impossible. In order to 'mine' the data with biological significance (this does not always coincide with statistical significance), data with multiple time and dose are considered to be essential. Any sophisticated statistical procedure cannot create anything from no data.

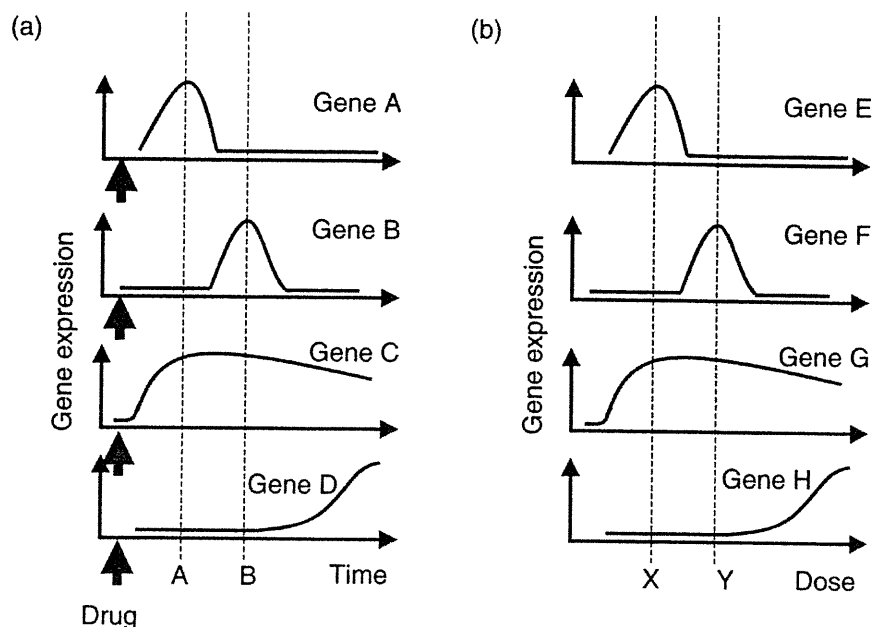


Figure 20.4 Schematic expressions of (a) time- and (b) dose-dependencies of gene expression changes. (a) Some of the genes might be up-regulated by the administration of a drug (indicated by arrows) either transiently at an early time point (A) or at a later time point (B), or continuously with early (C) or later onsets (D). If observations are made either at the time point A or B, the results would be quite different, i.e. up-regulation is observed in A and C at time point A, whereas up-regulation occurs in B and C at point B and nothing is detectable in D. (b) Similar to the above, some of the genes might be up-regulated by the administration of a drug with either bell-shaped (E, F) or sigmoidal (G, H) dose-dependency. If observations are made at the concentrations of either X or Y, the results would be quite different, i.e. up-regulation in E and G at concentration X, whereas up-regulation in F and G at Y with nothing detectable in H

Even with such enriched data stored in our database, the procedure as to how to efficiently withdraw significant genes is immature. A statistical test covering dose-dependency, Williams test, for example, is available, but it does not always work efficiently. In general, three for each group is too small for any statistical analysis. However, the cost effectiveness is still questionable when the number is increased to, say, five. In order to make the statistical analysis applied to over 10 000 measures meaningful, N should be increased to a similar order, which is impractical in the biological data. In our experience, too strict statistics should not be applied for extracting significantly mobilized genes, in order not to overlook important genes with biological variance. We use various properly applied approaches case-by-case, e.g. ANOVA with a relatively large p , followed by proper filtering (elimination of the genes with significance in low dose only, or whose expression were inversely correlated with dose, etc.) or selection without statistics, based on the value of ratio to control. In any case, all of the data of extracted genes down to a reasonable number are stored in the database and so at anytime can refer their dose- or time-dependencies.

20.2.4.5 New Knowledge from Accumulated Data

When vast amounts of data are accumulated in the database, an interesting thing emerges by simple alignment of the data. Figure 20. 5(a) presents a summary of the vehicle control

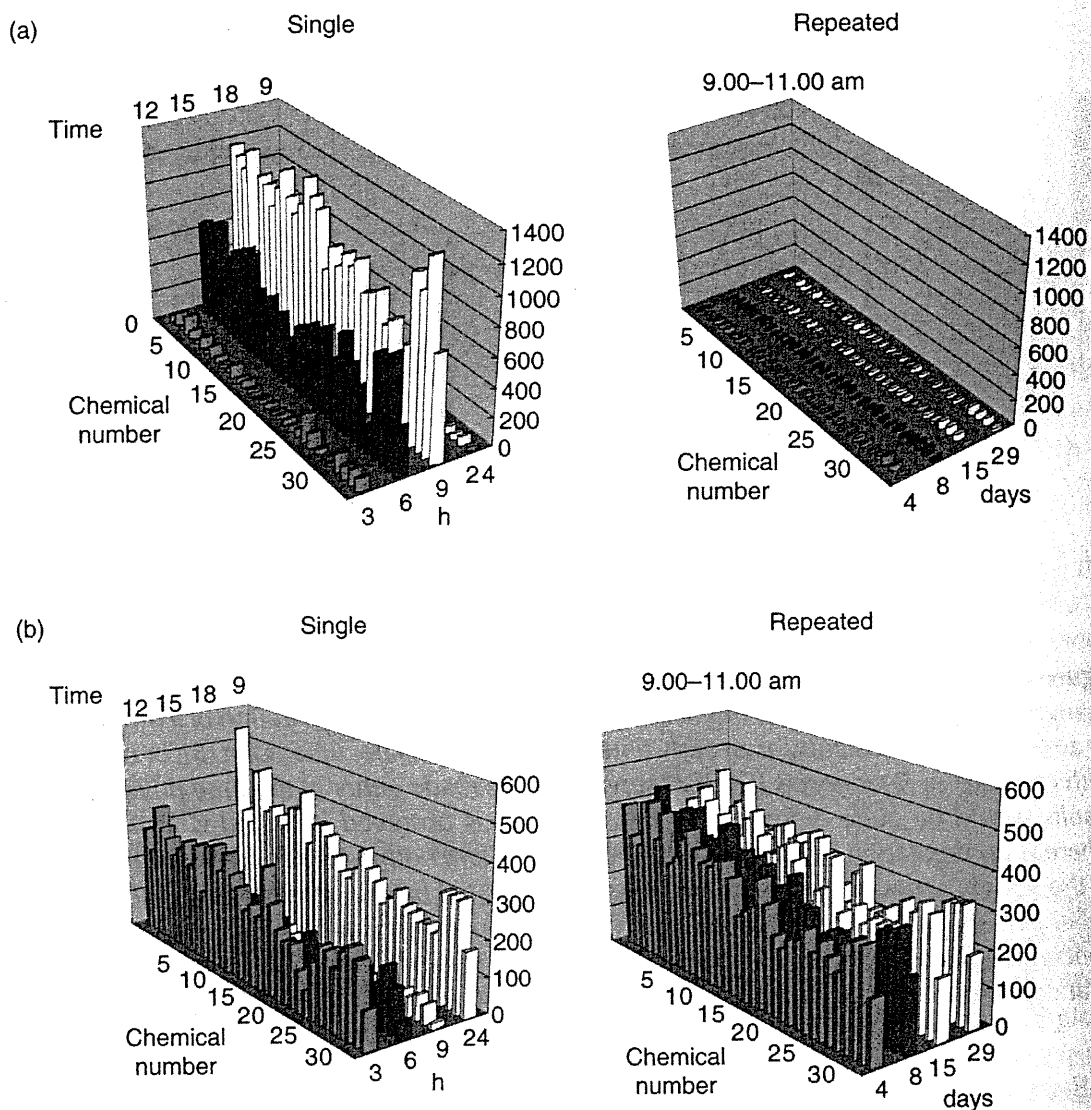


Figure 20.5 Circadian changes in the expression of (a) *D* site albumin promoter binding protein, 1387874 and (b) the aryl hydrocarbon receptor nuclear translocator-like, 1370510 in vehicle controls for the early 35 chemicals in the database tested by using the RAE 230A chip. Left panels – circadian changes of the expression within a day. Sacrifice of the animals in single-dose experiments is done 3, 6, 9 and 24 h after oral administrations in the morning. They thus correspond to about 12.00 midday, 15.00 pm, 18.00 pm and 9.00 am the next day, respectively. Right panels – in repeated administration, sacrifice is done in the morning at around 9:00 am to 11:00 am, and thus the expression level is constant to the value in the morning through all of the chemicals and days. It is also obvious that expression does not change with age

group of single- and repeated-dose experiments. The left panel shows the expression level of a representative circadian gene, *D* site albumin promoter binding protein, for 3, 6, 9 and 24 h data of the vehicle controls from the first 35 chemicals tested with an old A chip. As the administration is done in the morning, this pattern shows the reproducible increase of expression in the afternoon. The right panel shows its expression level in the repeated dosing for 3, 7, 14 and 28 days. As the autopsy was done 24 h after the last dose, i.e. in the morning, the expression level appears uniformly low. On the other hand, there are genes

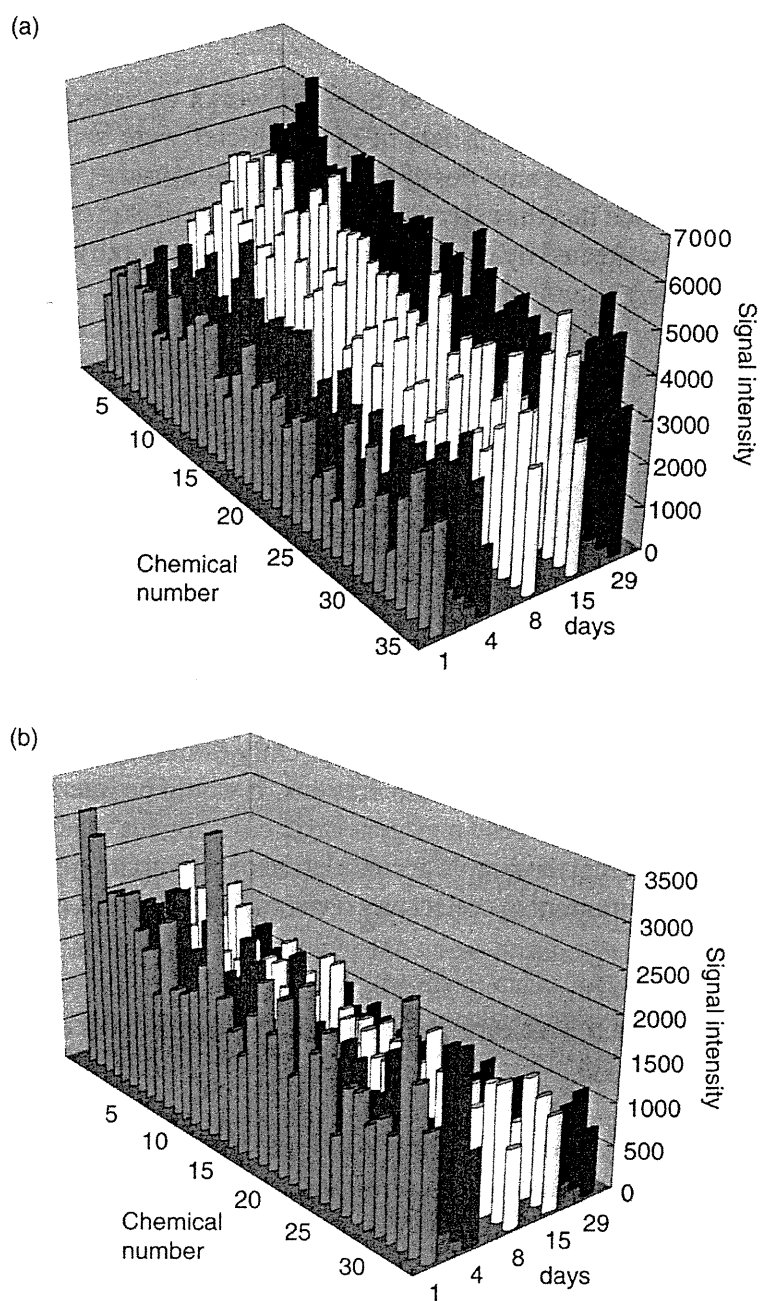


Figure 20.6 Expression levels of (a) steroid delta-isomerase 3 and (b) the hemoglobin beta chain complex in vehicle controls for the early 35 chemicals in the database tested by using the RAE 230A chip

with reversed circadian rhythm such as ARNT-like (Figure 20.5(b)), i.e. high expression level at noon and it goes down towards the night. The field of circadian gene expression has recently attracted attention (Ueda *et al.*, 2004). Using our TGP database, many genes with circadian rhythm, as well as potential drugs that affect the rhythm, can be easily extracted.

Reviewing the vehicle control data of repeated dosing experiments, we noticed a group of genes whose expression level changed with age. Figures 20.6(a) and 20.6(b) show the expression levels of steroid delta-isomerase 3 and hemoglobin beta chain complex, respectively. The former increased, while the latter decreased with age, possibly reflecting