

## 反復(新型慢性)暴露マウス肝トキシコゲノミクス

### 実験計画

物質A(14日間)	物質(単回)	
Clofibrate	Clofibrate	済
Clofibrate	PCN	済
Clofibrate	ATRA	済
CCl <sub>4</sub>	CCl <sub>4</sub>	済
CCl <sub>4</sub>	Clofibrate	済
CCl <sub>4</sub>	Phenobarbital	済
TBT	TBT	済
TBT	Clofibrate	済
TBT	Phenobarbital	済

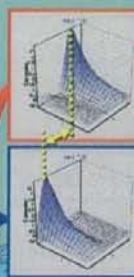
先行トキシコゲノミクス研究の成果の継承  
化合物リストの継承

物質Aの反復暴露群  
(遺伝子改変動物)

非暴露対照群  
(野生型動物)

物質Bの  
単回暴露

データ取得見込み  
① A3 × B3 = 9  
② 3 × 多臓器



●物質Bに対する2, 4, 8及び、24時間目の反応性の比較解析

●多臓器の比較

①物質Aの慢性影響の解析(遺伝子改変の影響の解析)  
②多臓器関連性の解析

## 反復(新型慢性)暴露マウス肝トキシコゲノミクス

### 実験計画

物質A(14日間)	物質(単回)	
VPA	VPA	済
VPA	Aspirin	済
VPA	Thalidomide	済
Deet	Deet	済
Deet	Permethrin	3月予定
Deet	Phenobarbital	予定

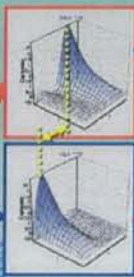
先行トキシコゲノミクス研究の成果の継承  
化合物リストの継承

物質Aの反復暴露群  
(遺伝子改変動物)

非暴露対照群  
(野生型動物)

物質Bの  
単回暴露

データ取得見込み  
① A3 × B3 = 9  
② 3 × 多臓器



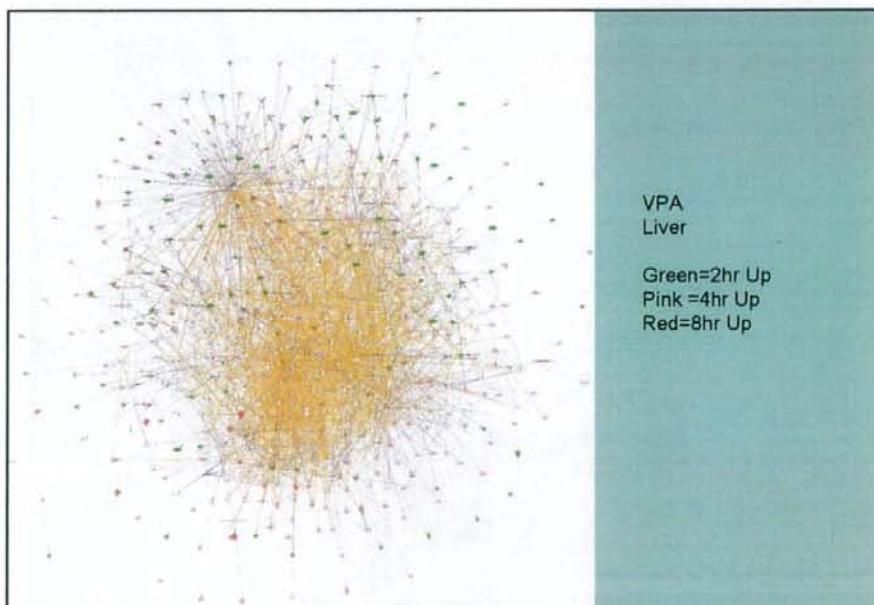
●物質Bに対する2, 4, 8及び、24時間目の反応性の比較解析

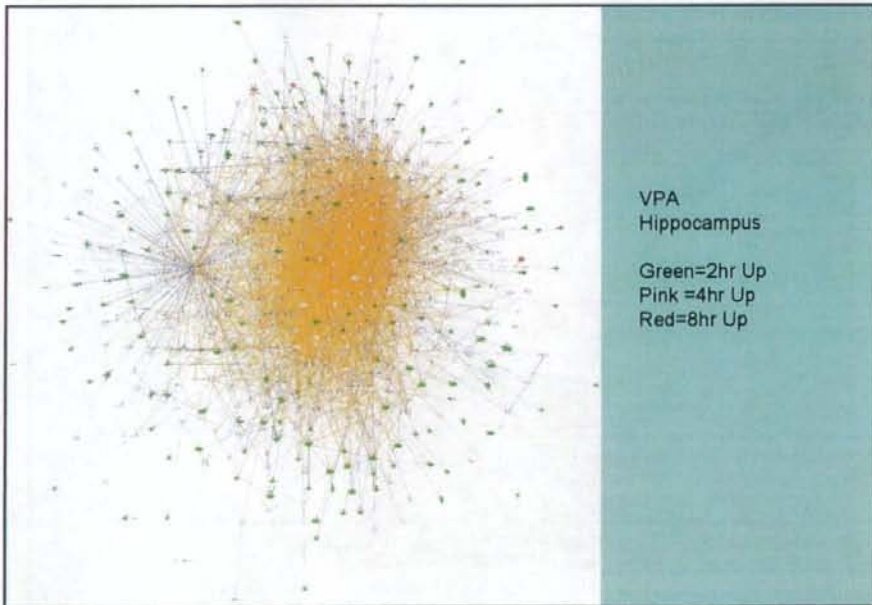
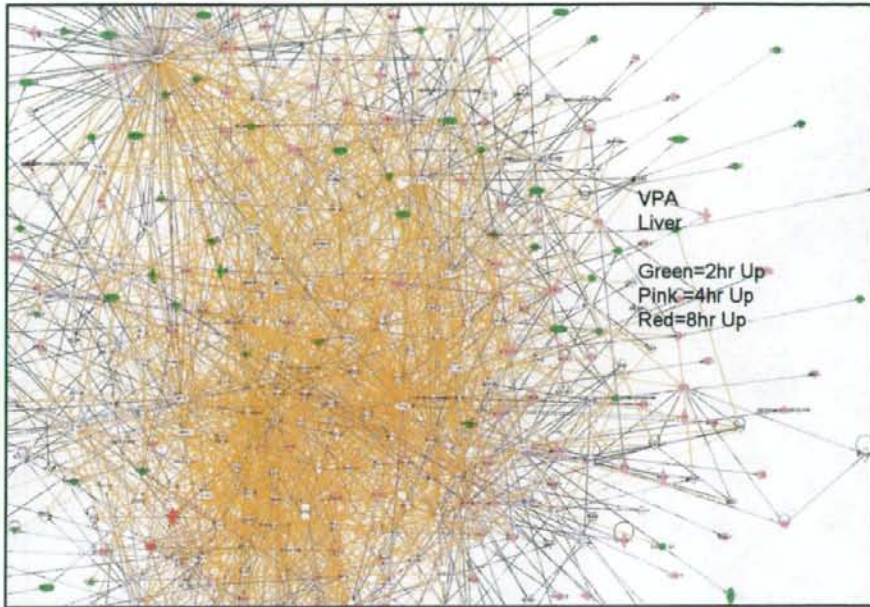
●多臓器の比較

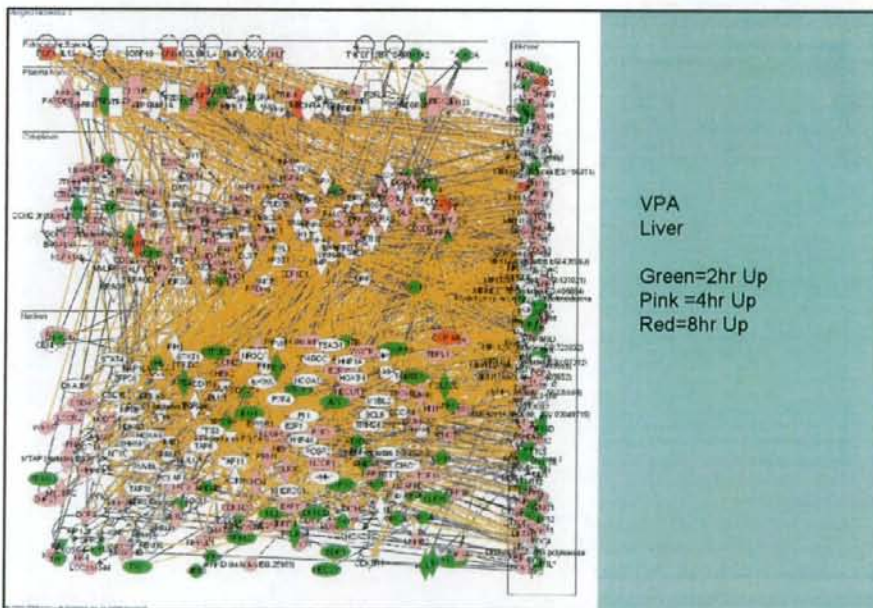
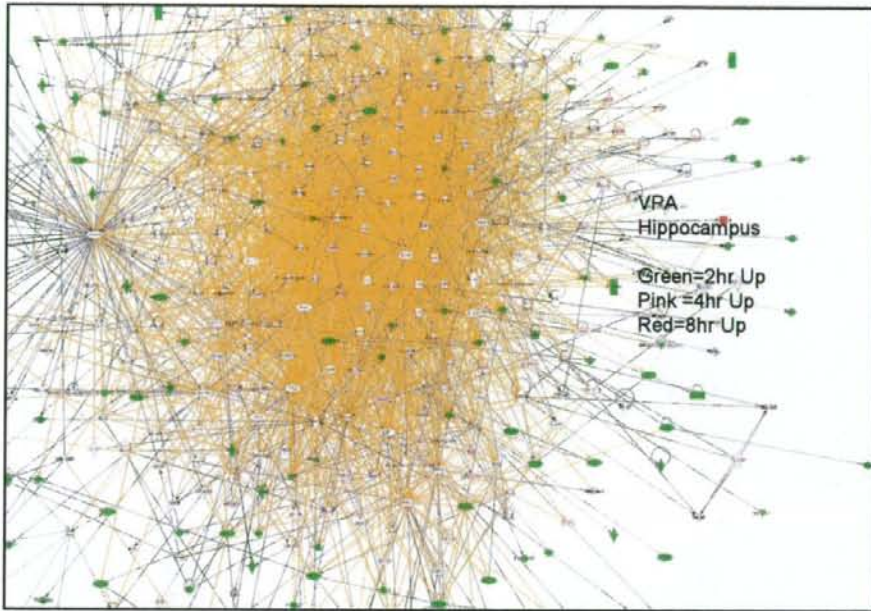
①物質Aの慢性影響の解析(遺伝子改変の影響の解析)  
②多臓器関連性の解析

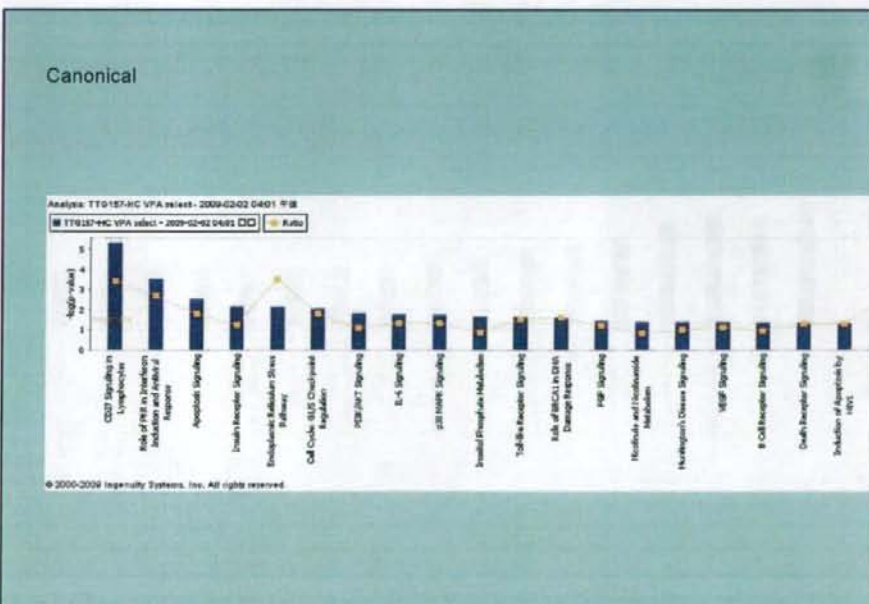
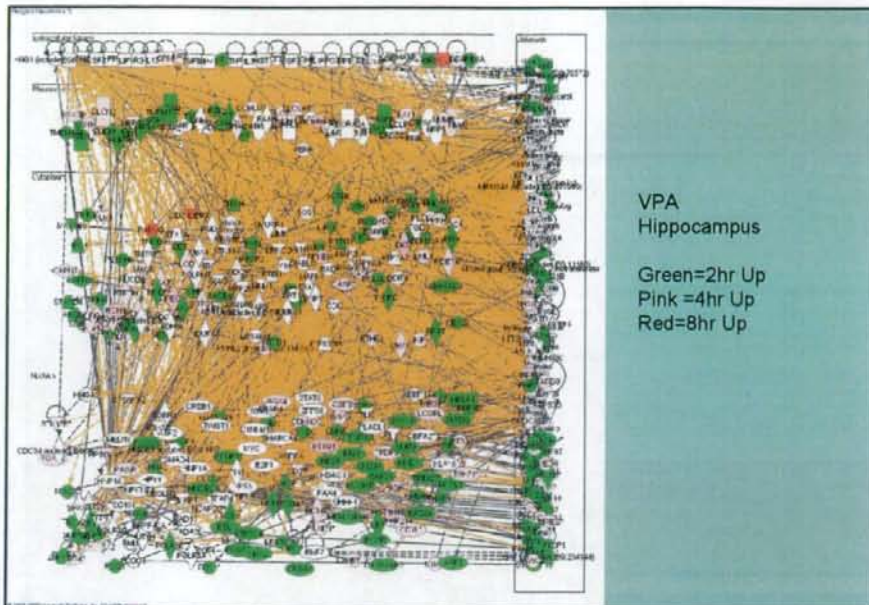
## 多臓器連関

Valproic acid  
Multi-organ



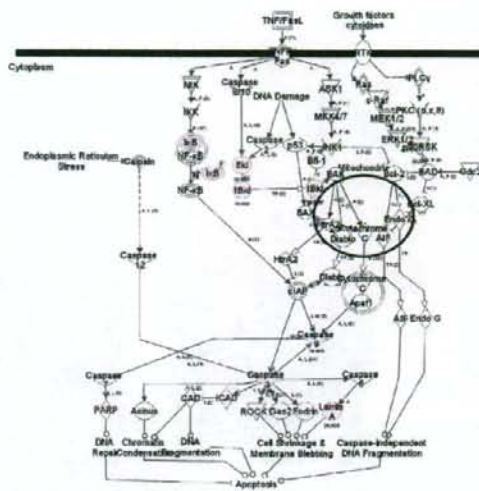






## Canonical Apoptosis

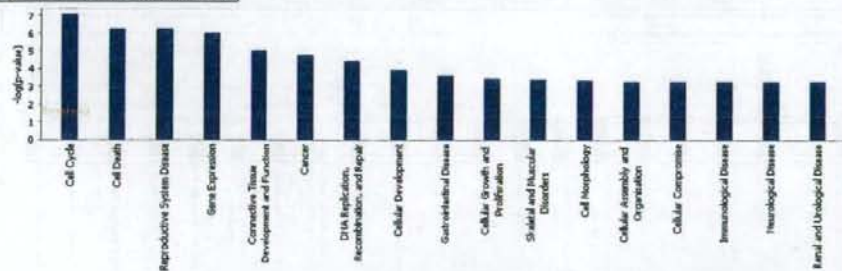
Apoptosis Signaling



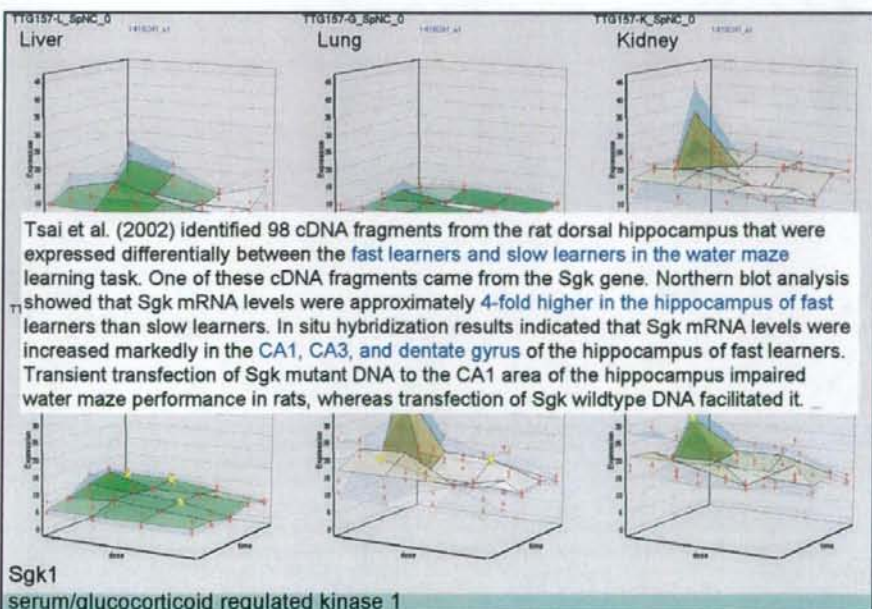
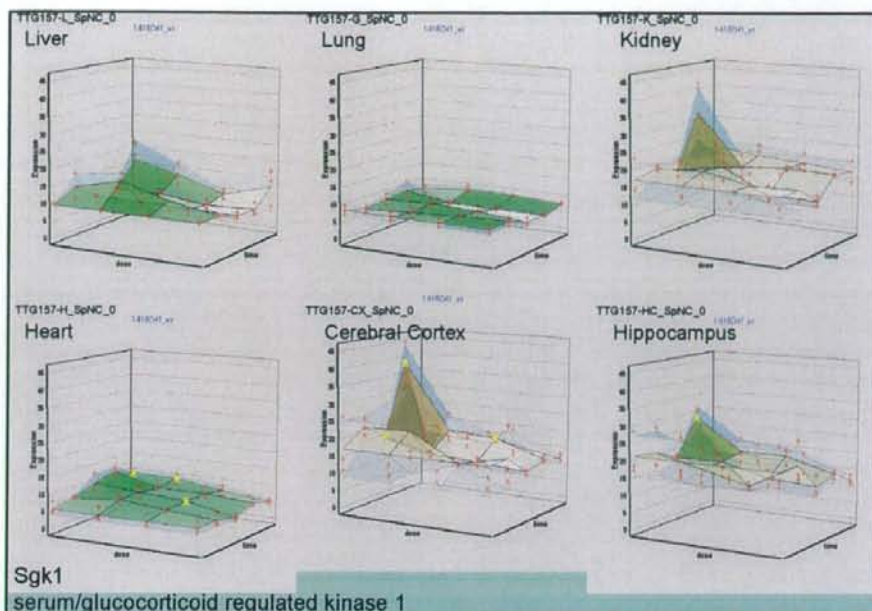
## Bio Function

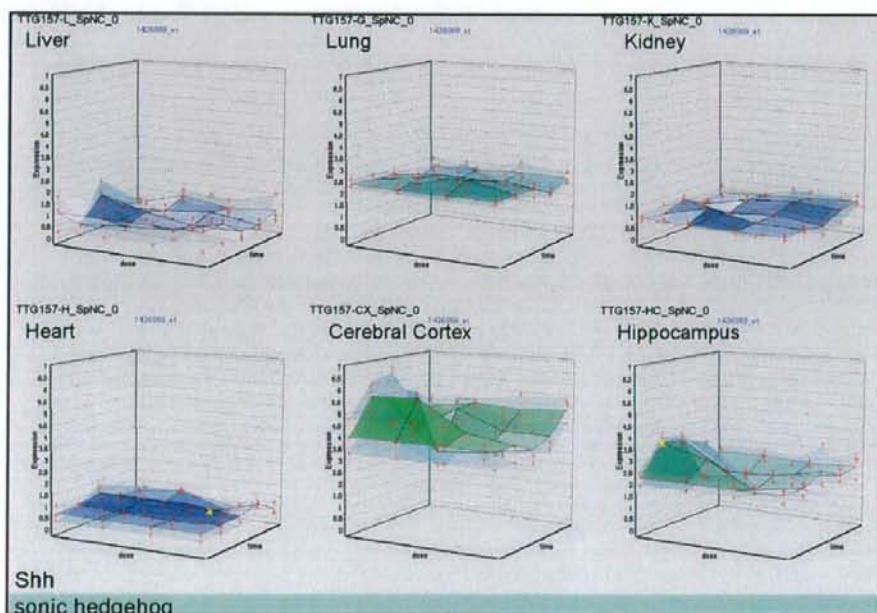
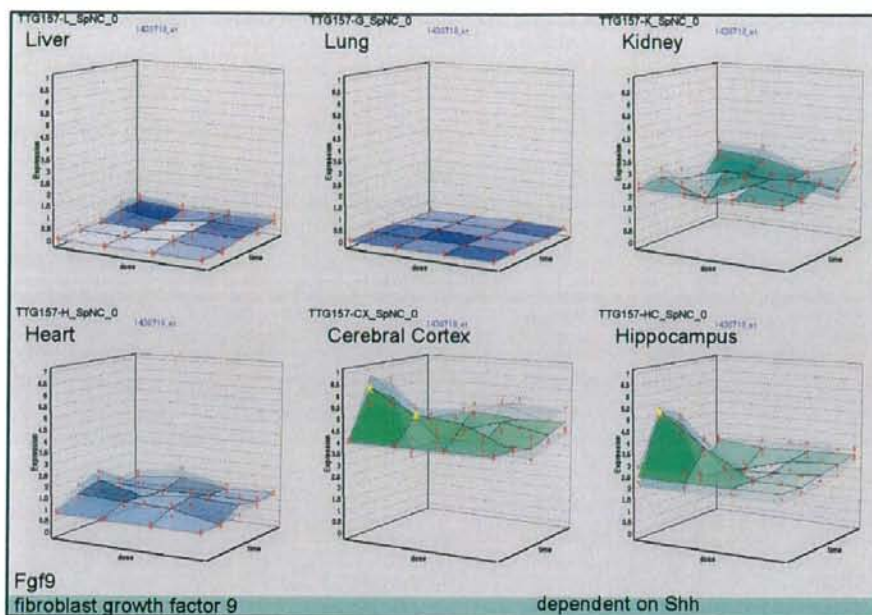
Analysis: TTG157-HC VPA select - 2009-02-02 04:01 48

■ TTG157-HC VPA select - 2009-02-02 04:01 □

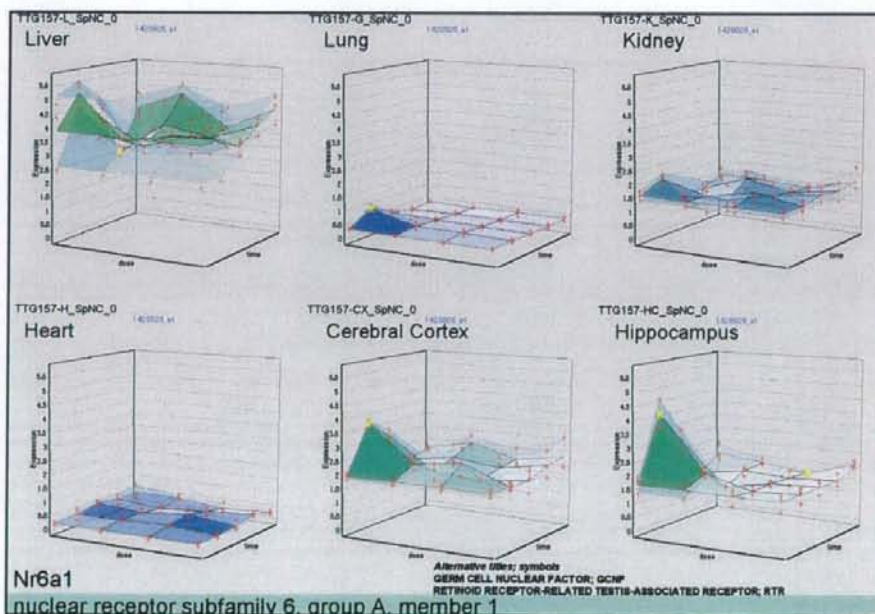
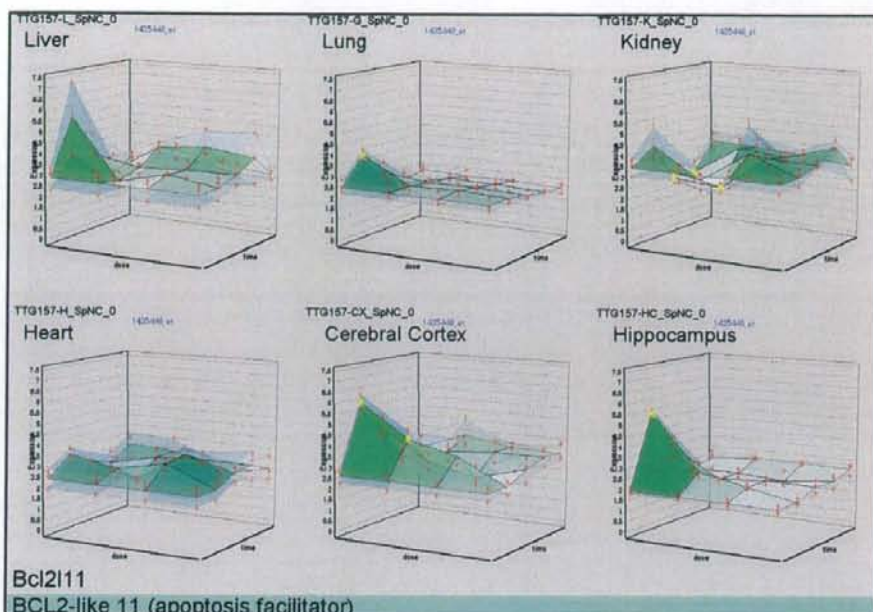


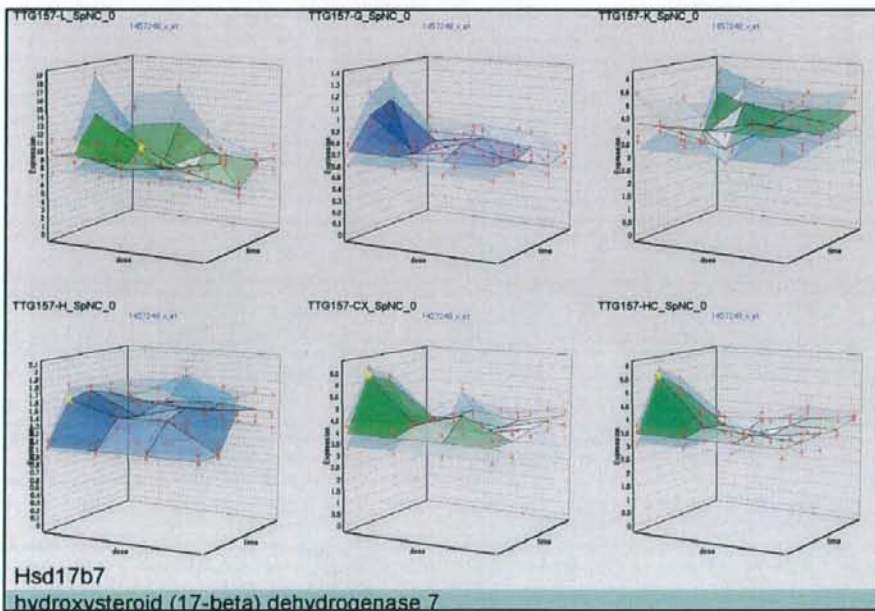
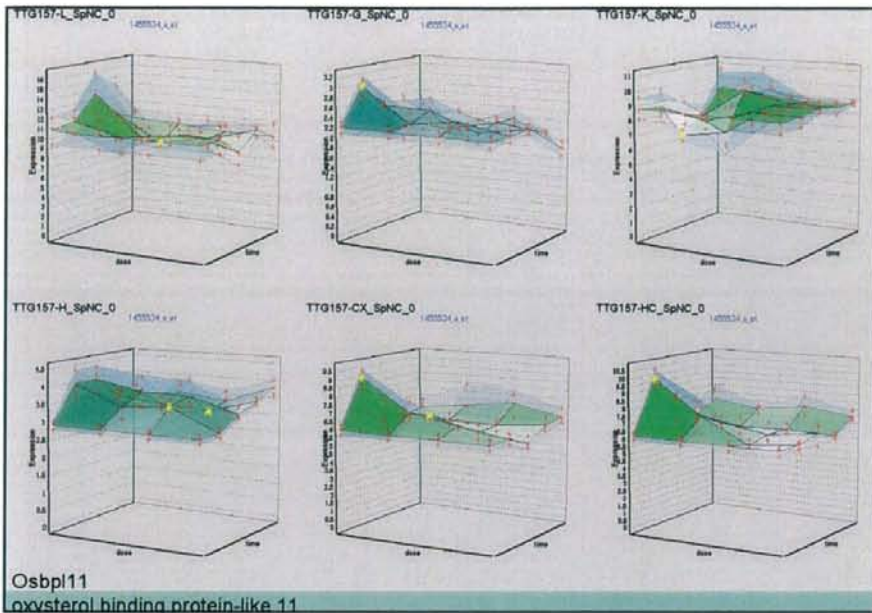
© 2000-2009 Ingenuity Systems, Inc. All rights reserved.











- NTT data / Teradata との共同研究

## Langmuirの方程式

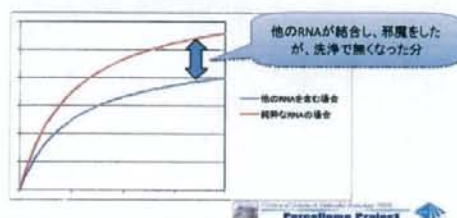
- Irving Langmuirによって1918年に導出された理論的な吸着等温式である。以下のような仮定を持っている。
  - 吸着媒には有限な数Nの吸着サイトがあり、そこだけで吸着質分子と結合する。
  - すべての吸着サイトは等価である。
  - 1つの吸着サイトは1つの吸着質分子としか結合しない。
  - 空の吸着サイトM、気相中の吸着質S、吸着サイトに結合した吸着質M-Sの間に  $M + S \rightleftharpoons M-S$  の化学平衡が成立する。
- 各プローブが飽和している場合には、Langmuirの方程式に従うと仮定する

$$\log(PM_{pjl}) = \log \left( I_p \frac{k_p c_j}{1 + k_p c_j} + bg_p \right) + \varepsilon_{pjl}$$

$p = 1, \dots, P$ : probe\_index  
 $j$ : concentration\_index  
 $c$ : concentration  
 $l = 1, \dots, L$ : replicate\_index  
 $I$ : Saturation\_Intensity  
 $k$ : equilibrium\_constant  
 $bg$ : background\_component

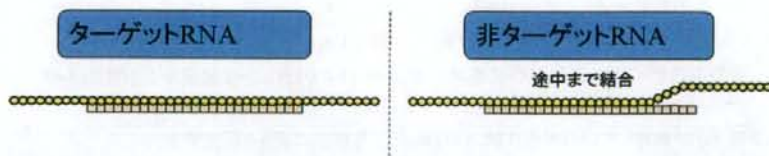
## 仮説: RNAの競合吸着

- Langmuir吸着等温方程式を各RNAごとに飽和補正を実施すると、大きめに推定されているのではないか？
- 仮説
  - 一つの吸着サイトに、ターゲット以外のRNAが結合し、ターゲットの結合ができなくなり、見掛け上最大飽和値が小さく推定される。
  - 特定のRNAが大きな発現を示す場合には、他のRNAが結合する間もなくターゲットで埋まってしまうため、見掛け上の最大飽和値がターゲットのみの場合に近くなるのではないだろうか？



補正係数算出のためのLBMは、何らかの工夫が必要である。

## Cross-HybridizationとWashingによる見かけ上の最大値の変化の考察



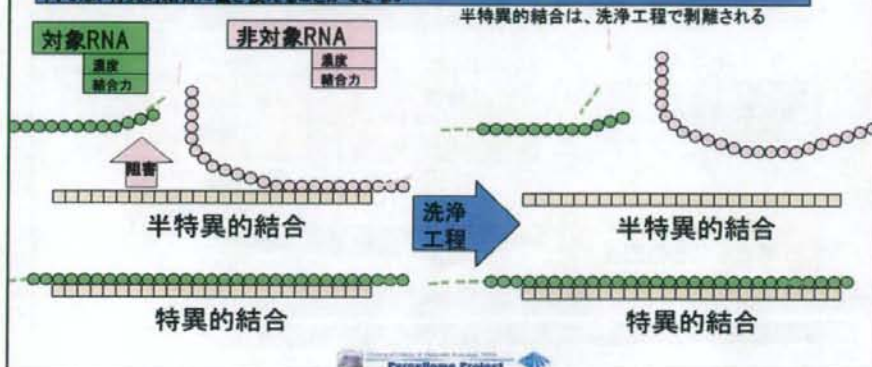
### 仮説

非ターゲットRNAは、洗浄工程で落とされ発現値として計測されないが、見掛け上の最大値を引き下げる。また、ターゲットRNAが極端に増えた場合には、通常の見掛け上の最大値を大幅に上回る

## Langmuir吸着等温式の多溶質への拡張

半特異的結合 (Semi-specific Hybridization) という概念を導入し、モデルを構築する

半特異的結合は、プローブと間にある程度の大きさの結合力が働き外れにくく、Hybridization時間内では、特異的結合に置き換えることができない



## Langmuir吸着等温式の多溶質への拡張

多溶質による吸着のモデル化、Langmuir吸着等温式と同じ仮定

吸着速度は、吸着体の残り個数と溶質の濃度に比例すると仮定する

$$V' = K \cdot \left( N - \sum_{i=1}^n N_i \right) \left( \sum_{i=1}^n C_i \right)$$

個々の溶質の吸着速度は、溶質の濃度の比に比例する

$$V'_i = \frac{C_i}{\sum_{j=1}^n C_j} V'$$

個々の溶質の離脱速度は、吸着した個数に比例する

$$V_i = \alpha_i \cdot N_i$$

式変形

プローブに吸着したターゲットの個数

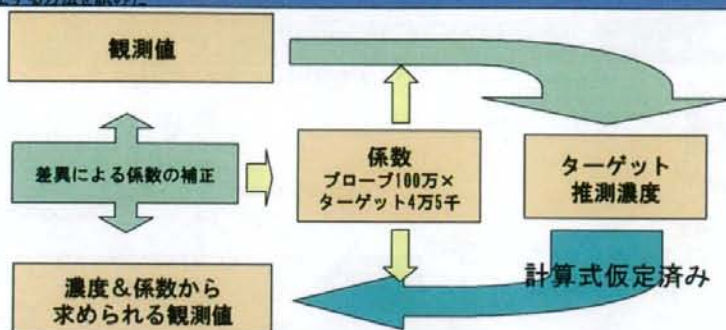
$$N_y = \frac{C_y}{\alpha_y} \cdot K \cdot \left( N - \sum_{j=1}^n N_j \right) + \sigma_y$$

プローブに吸着したターゲットの離脱速度係数  
吸着速度係数は1とし、 $\alpha$ で調整する

$$\alpha_y \geq 0$$

## 係数補正による係数の同定

プローブとターゲットのペアの塩基配列に基づいた結合力の同定は、ターゲット範囲外での結合など、未知要素が多く、困難なので、観測値から推測値を求め、更に観測値を推定し、その誤差で、係数を補正する方法を試みた。



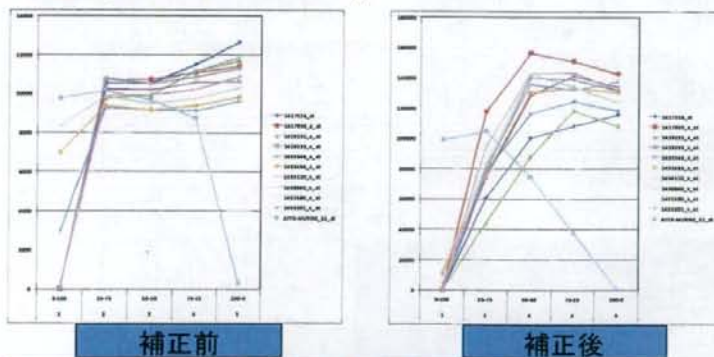
複数のチップ、実験条件で、繰り返し実施することで本来あるべき係数に近づいていく

ただし、計算時間がかかる。また、収束が保証できない

PerceHomo Project

## 係数補正による係数の同定(試行1)

- 飽和補正結果(LBM: 単Langmuir(昨年度実施))
- 高発現域(50%:50%で、10000以上)

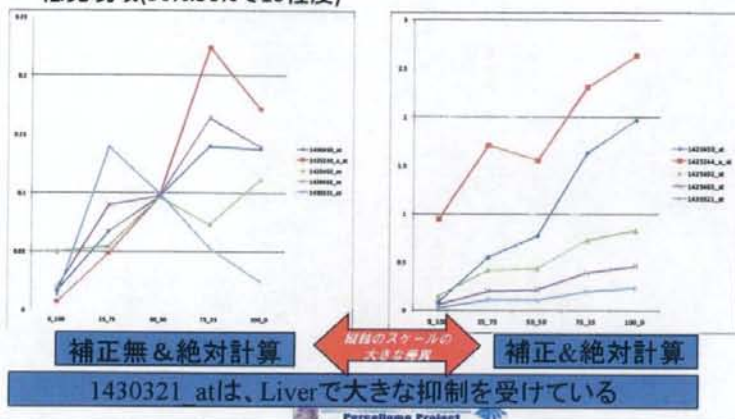


50%までは直線に引き延ばされている

PerceHomo Project

## 係数補正による係数の同定(試行1)

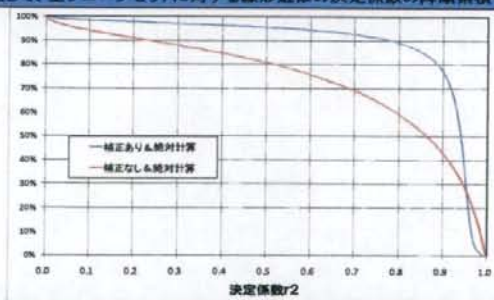
- 飽和補正結果(LBM: 多Langmuir&絶対化)
- 低発現域(50%:50%で10程度)



## 係数補正による係数の同定(試行1)

- LBM(補正&絶対計算後の直線性)

LBMは、配合割合を変更しているだけなので、推定精度は直線になるはずである。直線に近づいたかの指標として、全プローブセットに対する線形近似の決定係数の降順累積グラフを示す



Percellome絶対計算の直線性と比して格段に改善していると考えられる  
決定係数1付近が少ないのは、ControlRNAを用いた補正の結果と思われる

## 委託研究報告書(STEP5~7)

### 目次

1. Exon Chip実験結果のAICを用いた推定
2. MADICパラメータ検討
3. 局所変動性指標の作成
4. 化合物間クラスタリング
5. 飽和プローブセットの調査と対処方法の検討
6. プローブ配列に基づくマイクロアレイデータ補正



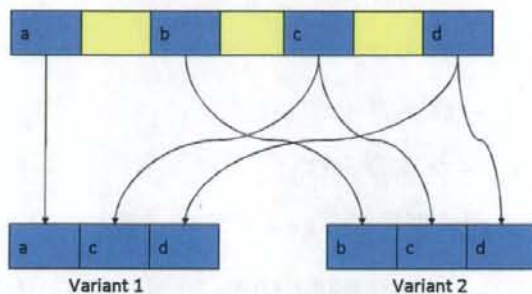
# 1. EXON CHIP実験結果のAICを用いた推定

## ExonとSplicing Variantの基本的な考え方



## 計算手順概要

Splicing Variantが存在する場合、個々のExon単位で発現量調整が行われている。



## 計算手順概要

Splicing Variantの種類によって、連立方程式を立てることが可能であると考えられる

The diagram shows a gene with four exons: a, b, c, and d. Exons a, b, and c are blue, and exons b and c are highlighted in yellow.

$$\begin{array}{l} \text{Variant 1} \quad \alpha \quad + \alpha + \alpha \\ \text{Variant 2} \quad \beta + \beta + \beta \end{array}$$

Splicing Variantごとに各Exonを使う/使わないの情報とSplicing Variantの存在量が既知ならば、各Exonの存在量も分かる。逆に、各Exonの量が与えられた時に、使う/使わないの情報と、Splicing Variantの存在量も計算可能である。

## 計算手順概要

観測結果は、モデルと誤差項の組み合わせであると考える

Exon量の観測値

観測誤差項

$$e_a = \alpha + \sigma_a$$

$$e_b = \beta + \sigma_b$$

$$e_c = \alpha + \beta + \sigma_c$$

$$e_d = \alpha + \beta + \sigma_d$$

$$AIC = -2(\text{最大対数尤度}) + 2(\text{パラメータ数})$$

組合せの全パターンに対して、AIC(赤池情報量規準)を計算し、最も小さいAICとなるパターンを採用する。全パターンを計算しつくすのは、膨大な計算時間が必要となるので、バリエーション数を制限する。

## 計算手順概要

n種類のバリエーションが存在する遺伝子であると仮定する。LiverとBrainにおいて各々違う比率で発生していると仮定する。エクソン数がm個であるとして数式をたてる。

$$\begin{pmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1n} \\ \beta_{21} & \beta_{22} & & \beta_{2n} \\ \vdots & & \ddots & \vdots \\ \beta_{m1} & \beta_{m2} & \cdots & \beta_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

$\beta_{ij}$ : 発現パターン係数: 0か1の値

$x_j$ : 発現量

$y_i$ : 発現観測量

$\varepsilon_i$ : 発現観測誤差量

## 計算手順概要

LiverとBrainを同時に式に与え、複雑な式になるものを行列として表現する。

$$\begin{pmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1n} \\ \beta_{21} & \beta_{22} & & \beta_{2n} \\ \vdots & & \ddots & \vdots \\ \beta_{m1} & \beta_{m2} & \cdots & \beta_{mn} \end{pmatrix} \begin{pmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1n} \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{1n} \end{pmatrix} = \begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n} \end{pmatrix}$$

$$\begin{pmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1n} \\ \beta_{21} & \beta_{22} & & \beta_{2n} \\ \vdots & & \ddots & \vdots \\ \beta_{m1} & \beta_{m2} & \cdots & \beta_{mn} \end{pmatrix} \begin{pmatrix} x_{21} \\ x_{22} \\ \vdots \\ x_{2n} \end{pmatrix} + \begin{pmatrix} \varepsilon_{21} \\ \varepsilon_{22} \\ \vdots \\ \varepsilon_{2n} \end{pmatrix} = \begin{pmatrix} y_{21} \\ y_{22} \\ \vdots \\ y_{2n} \end{pmatrix}$$

$\mathbf{B} \cdot \mathbf{x}_L + \varepsilon_L = \mathbf{y}_L$ 
 $\mathbf{B} \cdot \mathbf{x}_B + \varepsilon_B = \mathbf{y}_B$

$\begin{pmatrix} \mathbf{B} & \\ & \mathbf{B} \end{pmatrix} \begin{pmatrix} \mathbf{x}_L \\ \mathbf{x}_B \end{pmatrix} + \begin{pmatrix} \varepsilon_L \\ \varepsilon_B \end{pmatrix} = \begin{pmatrix} \mathbf{y}_L \\ \mathbf{y}_B \end{pmatrix}$

## 計算手順概要

LiverとBrain、更にMixtureも考慮した式を作成する。

$$\begin{pmatrix} \mathbf{I} & \mathbf{0} \\ 0.8 \cdot \mathbf{I} & 0.2 \cdot \mathbf{I} \\ 0.6 \cdot \mathbf{I} & 0.4 \cdot \mathbf{I} \\ 0.4 \cdot \mathbf{I} & 0.6 \cdot \mathbf{I} \\ 0.2 \cdot \mathbf{I} & 0.8 \cdot \mathbf{I} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{0} \\ \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_{1202} \\ \varepsilon_{1204} \\ \varepsilon_{1206} \\ \varepsilon_{1208} \\ \varepsilon_p \end{pmatrix} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_{1202} \\ \mathbf{y}_{1204} \\ \mathbf{y}_{1206} \\ \mathbf{y}_{1208} \\ \mathbf{y}_p \end{pmatrix}$$

誤差項の二乗和を最小化する発現推定量が、その発現パターンに対する推定量であるとみなすことができる。

$$g(\mathbf{x}) = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_{1202} \\ \varepsilon_{1204} \\ \varepsilon_{1206} \\ \varepsilon_{1208} \\ \varepsilon_p \end{pmatrix} \begin{pmatrix} \varepsilon_1 \\ \varepsilon_{1202} \\ \varepsilon_{1204} \\ \varepsilon_{1206} \\ \varepsilon_{1208} \\ \varepsilon_p \end{pmatrix}$$