

## 計算手順概要

誤差項の2乗和は、2次式なので微分して0となる点が発現量推定値であると考えることができる。

$$\frac{\partial}{\partial x_j} g(\mathbf{x}) = 0$$

$$\begin{pmatrix} \mathbf{y}_L \\ y_{L75025} \\ y_{L50050} \\ y_{L25075} \\ \mathbf{y}_B \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ 0.75 \cdot \mathbf{I} & 0.25 \cdot \mathbf{I} \\ 0.50 \cdot \mathbf{I} & 0.50 \cdot \mathbf{I} \\ 0.25 \cdot \mathbf{I} & 0.75 \cdot \mathbf{I} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{pmatrix} \frac{\partial}{\partial x_j} \begin{pmatrix} \mathbf{x}_L \\ \mathbf{x}_B \end{pmatrix} = \mathbf{0}$$

推定量から誤差値を求めることができ、尤度を計算してAICを計算する。

## 計算手順概要

本数式は、 $2n$ 個（バリエーション数 $\times$ L B）の連立一次方程式を作成する。したがって、この連立方程式を解けば、発現パターンごとの発現推定量を求めることができる。この発現推定量に対する、尤度を次のように仮定して求める。

$$L_x(\mu, \sigma) = p_{\mu, \sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

この尤度を用いてAICを計算し、最適な発現パターンを計算する。しかしながら、発現パターンは、 $2^n$ 個存在する。 $nm$ が大きくなると、計算時間がかかりすぎて、実用的ではなくなる。対処が必要である。

## 推定用計算式作成

細かく展開した式を作成する

Mixtureも考慮した式にする

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。

$$\begin{pmatrix}
 \mathbf{I} & \mathbf{0} \\
 0.75\mathbf{I} & 0.25\mathbf{I} \\
 0.50\mathbf{I} & 0.50\mathbf{I} \\
 0.25\mathbf{I} & 0.75\mathbf{I} \\
 \mathbf{0} & \mathbf{I} \\
 \mathbf{I} & \mathbf{0} \\
 0.75\mathbf{I} & 0.25\mathbf{I} \\
 0.50\mathbf{I} & 0.50\mathbf{I} \\
 0.25\mathbf{I} & 0.75\mathbf{I} \\
 \mathbf{0} & \mathbf{I} \\
 \mathbf{I} & \mathbf{0} \\
 0.75\mathbf{I} & 0.25\mathbf{I} \\
 0.50\mathbf{I} & 0.50\mathbf{I} \\
 0.25\mathbf{I} & 0.75\mathbf{I} \\
 \mathbf{0} & \mathbf{I}
 \end{pmatrix}
 \begin{pmatrix}
 \mathbf{0} \\
 \mathbf{0} \\
 \mathbf{0} \\
 \mathbf{0} \\
 \mathbf{0} \\
 \mathbf{0} \\
 \mathbf{0} \\
 \mathbf{0} \\
 \mathbf{0} \\
 \mathbf{0} \\
 \mathbf{0} \\
 \mathbf{0} \\
 \mathbf{0} \\
 \mathbf{0} \\
 \mathbf{0}
 \end{pmatrix}
 +
 \begin{pmatrix}
 \mathbf{0} \\
 \mathbf{0} \\
 \mathbf{0} \\
 \mathbf{0} \\
 \mathbf{0} \\
 \mathbf{0} \\
 \mathbf{0} \\
 \mathbf{0} \\
 \mathbf{0} \\
 \mathbf{0} \\
 \mathbf{0} \\
 \mathbf{0} \\
 \mathbf{0} \\
 \mathbf{0} \\
 \mathbf{0}
 \end{pmatrix}
 \begin{pmatrix}
 \mathbf{x}_L \\
 \mathbf{x}_B \\
 \mathbf{x}_L \\
 \mathbf{x}_B \\
 \mathbf{x}_L \\
 \mathbf{x}_B \\
 \mathbf{x}_L \\
 \mathbf{x}_B \\
 \mathbf{x}_L \\
 \mathbf{x}_B \\
 \mathbf{x}_L \\
 \mathbf{x}_B \\
 \mathbf{x}_L \\
 \mathbf{x}_B \\
 \mathbf{x}_L \\
 \mathbf{x}_B
 \end{pmatrix}
 =
 \begin{pmatrix}
 \mathbf{y}_L \\
 \mathbf{y}_B \\
 \mathbf{y}_L \\
 \mathbf{y}_B \\
 \mathbf{y}_L \\
 \mathbf{y}_B \\
 \mathbf{y}_L \\
 \mathbf{y}_B \\
 \mathbf{y}_L \\
 \mathbf{y}_B \\
 \mathbf{y}_L \\
 \mathbf{y}_B \\
 \mathbf{y}_L \\
 \mathbf{y}_B \\
 \mathbf{y}_L \\
 \mathbf{y}_B
 \end{pmatrix}$$

$$g(\mathbf{x}) = \begin{pmatrix} e_{L1} \\ e_{L2} \\ e_{L3} \\ e_{L4} \\ e_{L5} \\ e_{L6} \\ e_{L7} \\ e_{L8} \\ e_{L9} \\ e_{L10} \\ e_{L11} \\ e_{L12} \\ e_{L13} \\ e_{L14} \\ e_{L15} \\ e_{L16} \end{pmatrix} \begin{pmatrix} e_{B1} \\ e_{B2} \\ e_{B3} \\ e_{B4} \\ e_{B5} \\ e_{B6} \\ e_{B7} \\ e_{B8} \\ e_{B9} \\ e_{B10} \\ e_{B11} \\ e_{B12} \\ e_{B13} \\ e_{B14} \\ e_{B15} \\ e_{B16} \end{pmatrix}$$

このような方程式を過剰条件の連立方程式と呼ぶ

## 推定用計算式作成

- 過剰条件の連立一次方程式には幾つかの方法が提唱されている

- 正規方程式に解法

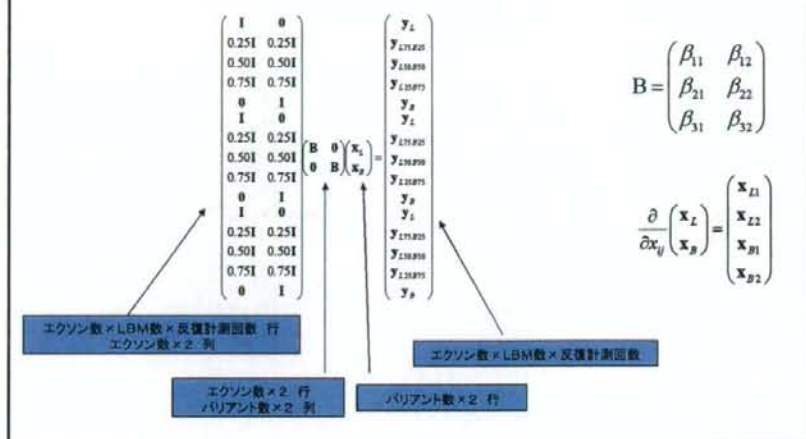
$${}^t\mathbf{A}\mathbf{A}\mathbf{x} = {}^t\mathbf{A}\mathbf{y}$$

- QR分解法

などの方法がある

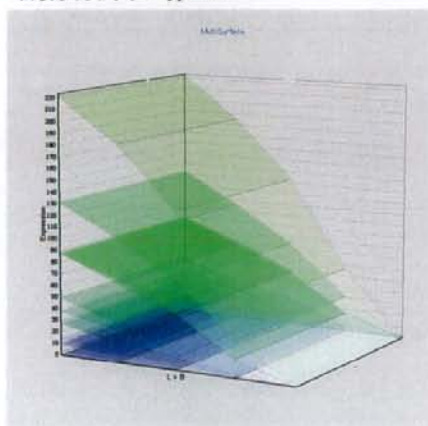
## 推定用計算式作成

3個のエクソン、2種類のバリエントで、数式を組み立ててみる



## 推定用計算試行

エクソンチップの実験結果を用いて、計算を行ってみる  
 計算実験対象: Cyp2a12



バリエント数分の直線で近似する  
 同じタンパクを生成するExonのはず(?)

タンパク1個で、Exonが重複する?  
 行列  $\beta$  の作り方の問題

計測誤差?

## 結論

- 現段階において、Exon Chipの定量性に問題があり、Splicing Variantの特定が困難である。

## 2. MADICパラメータ検討

## 背景と目的

- 背景
  - MADIC(Mass Distributed Clustering : A New Clustering Algorithm for Repeated Measurements in Gene Expression Data, Genome Informatics 16(2): 183-194 (2005), PMID: 16901101、特許番号3995099 (平19.8.10))は論文作成時に、Moe430Aを対象としており、現行チップ、Moe430 2.0ではプローブセットが増えている。
- 目的
  - 増加したプローブセット数に対して、適切なパラメータを導き出す。

## MADICクラスタ分類

### 密度が高い領域

- RIGOROUS
  - クラスタの中心
- SHELL
  - RIGOROUSから一定の距離以内に存在する
- ADHESIVE
  - RIGOROUSから少し離れている

### 密度が低い領域

- UNIQUE
  - 単一で密度が高いとみなせる(SDが小さい)
- ISOLATE
  - 単一で密度が高いとみなせる
- UNSTABLE
  - 単一で密度が高いとは考えられない

## 検討パターン

- 第1パターン
  - 論文時と同様のパラメータを用いる
  - $\rho = 1.4$
- 第2パターン
  - 密度閾値を若干厳しくし、クラスタとして現れるのを少し遅らせる
  - $\rho = 1.7$
- クラスタの最後の閾値は同じものを使用する

## 結果

2パターンの結果分類の対比

		パターン2						
パターン1	標準	RIGOROUS	SHELL	ADHESIVE	UNIQUE	ISOLATE	UNSTABLE	総計
	RIGOROUS	480	30	8				518
	SHELL	13	13,513	3,045				16,571
	ADHESIVE	1	1,174	22,020				23,195
	UNIQUE				4,322			4,322
	ISOLATE					8		8
	UNSTABLE						487	487
	総計	494	14,717	25,073	4,322	8	487	45,101

厳しい基準にすることにより、RIGOROUSがSHELLやADHESIVEへ移っていった。しかし、大きな違いは存在しない

## 結果

2パターンのうち片方だけでRIGOROUSとなったブローブ

品名	パターン1(標準)				パターン2			
	品名	規格	検査	結果	品名	規格	検査	結果
MOE430A	MOE430A	MOE430A	MOE430A	MOE430A	MOE430A	MOE430A	MOE430A	MOE430A
MOE430B	MOE430B	MOE430B	MOE430B	MOE430B	MOE430B	MOE430B	MOE430B	MOE430B
MOE430C	MOE430C	MOE430C	MOE430C	MOE430C	MOE430C	MOE430C	MOE430C	MOE430C
MOE430D	MOE430D	MOE430D	MOE430D	MOE430D	MOE430D	MOE430D	MOE430D	MOE430D
MOE430E	MOE430E	MOE430E	MOE430E	MOE430E	MOE430E	MOE430E	MOE430E	MOE430E
MOE430F	MOE430F	MOE430F	MOE430F	MOE430F	MOE430F	MOE430F	MOE430F	MOE430F
MOE430G	MOE430G	MOE430G	MOE430G	MOE430G	MOE430G	MOE430G	MOE430G	MOE430G
MOE430H	MOE430H	MOE430H	MOE430H	MOE430H	MOE430H	MOE430H	MOE430H	MOE430H
MOE430I	MOE430I	MOE430I	MOE430I	MOE430I	MOE430I	MOE430I	MOE430I	MOE430I
MOE430J	MOE430J	MOE430J	MOE430J	MOE430J	MOE430J	MOE430J	MOE430J	MOE430J
MOE430K	MOE430K	MOE430K	MOE430K	MOE430K	MOE430K	MOE430K	MOE430K	MOE430K
MOE430L	MOE430L	MOE430L	MOE430L	MOE430L	MOE430L	MOE430L	MOE430L	MOE430L
MOE430M	MOE430M	MOE430M	MOE430M	MOE430M	MOE430M	MOE430M	MOE430M	MOE430M
MOE430N	MOE430N	MOE430N	MOE430N	MOE430N	MOE430N	MOE430N	MOE430N	MOE430N
MOE430O	MOE430O	MOE430O	MOE430O	MOE430O	MOE430O	MOE430O	MOE430O	MOE430O
MOE430P	MOE430P	MOE430P	MOE430P	MOE430P	MOE430P	MOE430P	MOE430P	MOE430P
MOE430Q	MOE430Q	MOE430Q	MOE430Q	MOE430Q	MOE430Q	MOE430Q	MOE430Q	MOE430Q
MOE430R	MOE430R	MOE430R	MOE430R	MOE430R	MOE430R	MOE430R	MOE430R	MOE430R
MOE430S	MOE430S	MOE430S	MOE430S	MOE430S	MOE430S	MOE430S	MOE430S	MOE430S
MOE430T	MOE430T	MOE430T	MOE430T	MOE430T	MOE430T	MOE430T	MOE430T	MOE430T
MOE430U	MOE430U	MOE430U	MOE430U	MOE430U	MOE430U	MOE430U	MOE430U	MOE430U
MOE430V	MOE430V	MOE430V	MOE430V	MOE430V	MOE430V	MOE430V	MOE430V	MOE430V
MOE430W	MOE430W	MOE430W	MOE430W	MOE430W	MOE430W	MOE430W	MOE430W	MOE430W
MOE430X	MOE430X	MOE430X	MOE430X	MOE430X	MOE430X	MOE430X	MOE430X	MOE430X
MOE430Y	MOE430Y	MOE430Y	MOE430Y	MOE430Y	MOE430Y	MOE430Y	MOE430Y	MOE430Y
MOE430Z	MOE430Z	MOE430Z	MOE430Z	MOE430Z	MOE430Z	MOE430Z	MOE430Z	MOE430Z

## 結論

- RIGOROUSになるかどうかで、一部違いはみられたものの、両パターン間で有意な差異はなかった。
- MOE430Aで使用したパラメータをMOE430 2.0でもそのまま使用可能と考えられた。

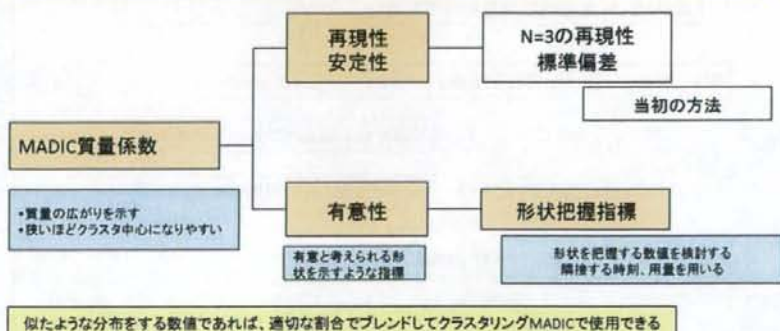
### 3. 局所変動性指標\*の作成

\* 時間、用量の実験条件に対して、局所的に誘導がかかっていることを示す指標

25

### 局所変動性指標の検討

クラスターリング手法 (MADIC) では、再現性の指標を用いて、クラスターの発見を行っているが、一昨年の研究成果により、高濃度のRNAでは、再現性が高いにもかかわらず、実験時期などの影響を大きく受けることが示された。  
誘導がかかっていることを示す指標を用いて、クラスターリングを実施すれば、適切なクラスターの発見につながるはずである。





## 再現性指標と局所変動性指標

### 再現性指標

MADIC当初の質量分布を規定する指標

平均サーフェースから全観測の標準偏差

再現性の高い場合に、小さな値を示す

### 発現局所変動性指標

今回検討した指標

平均値の時刻差分と用量差分の絶対値の中央値

単純な誘導がかかっている場合に、小さな値を示す

## 形状把握指標 アイデア

変化が特定の時刻、用量に集中している形状は価値がある

サーカディアンは、時間方向変化はあるが、用量方向変化は小さい

単峰は、特定の時間用量だけで変化

MAD: the median for absolute deviation

$$SD = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

標準偏差 = 偏差の平方の平均の平方根

$$MAD = \mathit{median}_{i=1}^n (abs(x_i - \bar{x}))$$

MAD = 偏差の絶対値の中央値

$$MAD1 = \mathit{median}_{i=1}^n (abs(\Delta Q_i))$$

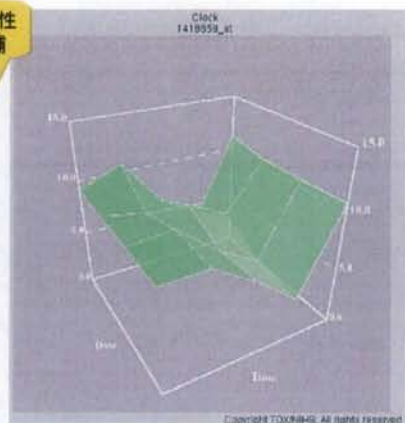
$$MAD2 = \mathit{median}_{i=1}^n (abs(\Delta Q_i - \overline{\Delta Q}))$$

## 試行結果(Ethanol)

局所変動性  
指標候補

	正規化MAD
実	0.8678
1階差分	0.8358
1階差分(平均差)	0.8508
時間1階差分	1.3930
用量1階差分	0.3448
平均	0.8689
幾何平均	0.8636
最小	0.3448

Clock遺伝子は、サーカディアン遺伝子として、用量の影響を受けず、時間1階差分は大きく、用量1階差分は大きくなっている。

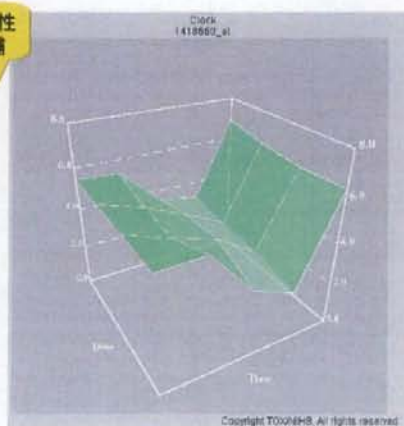


## 試行結果(Ethanol)

局所変動性  
指標候補

	正規化MAD
実	0.9187
1階差分	0.4968
1階差分(平均差)	0.5831
時間1階差分	1.3941
用量1階差分	0.1699
平均	0.7820
幾何平均	0.7497
最小	0.1699

Clock遺伝子は、サーカディアン遺伝子として、用量の影響を受けず、時間1階差分は大きく、用量1階差分は大きくなっている。

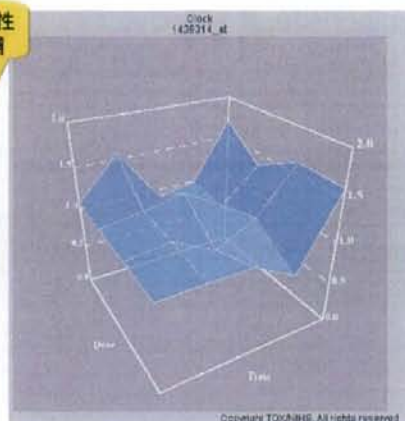


## 試行結果(Ethanol)

局所変動性  
指標候補

	正規化MAD
実	0.7848
1階差分	0.9622
1階差分(平均差)	0.9941
時間1階差分	1.3420
用量1階差分	0.9002
平均	1.1211
幾何平均	1.0385
最小	0.9002

Clock遺伝子のこのプローブセットでは発現量が低く、偶然影響と考えられるように、面が荒れていると思われる。他のclockプローブセットより1階差分の値が大きい。

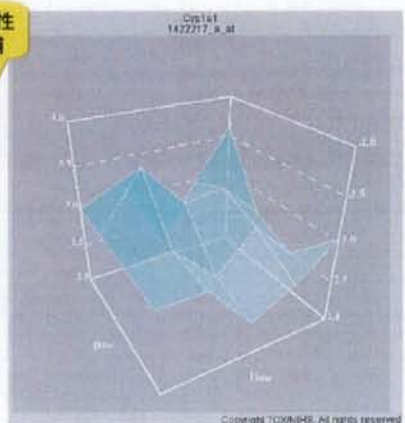


## 試行結果(Ethanol)

局所変動性  
指標候補

	正規化MAD
実	0.6505
1階差分	1.2033
1階差分(平均差)	1.2033
時間1階差分	1.1620
用量1階差分	0.9769
平均	1.0694
幾何平均	1.0257
最小	0.9769

Cyp1a1遺伝子は、偶然影響と考えら、面が荒れていると思われる。1階差分の値が大きい。

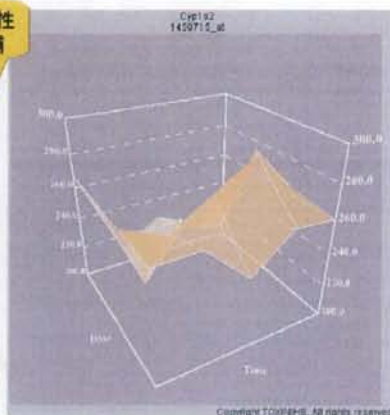


## 試行結果(Ethanol)

局所変動性  
指標候補

	正規化MAD
実	0.7954
1階差分	0.6103
1階差分(平均差)	0.7575
時間1階差分	0.5337
用量1階差分	0.8979
平均	0.7158
幾何平均	0.8632
最小	0.5337

Cyp1a2遺伝子は、用量の影響が大きな遺伝子であり、1階差分の値が小さい。

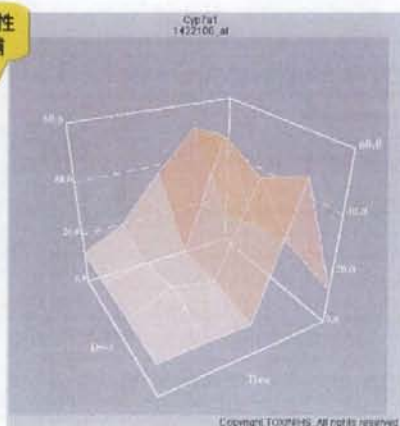


## 試行結果(Ethanol)

局所変動性  
指標候補

	正規化MAD
実	0.7424
1階差分	0.4077
1階差分(平均差)	0.4077
時間1階差分	1.8685
用量1階差分	0.2720
平均	1.0703
幾何平均	0.8734
最小	0.2720

Cyp7a1遺伝子では、8hの全用量でピークを迎えており、単純な誘導がかかっていると考えられ、1階差分でのMADが小さな値となっている。

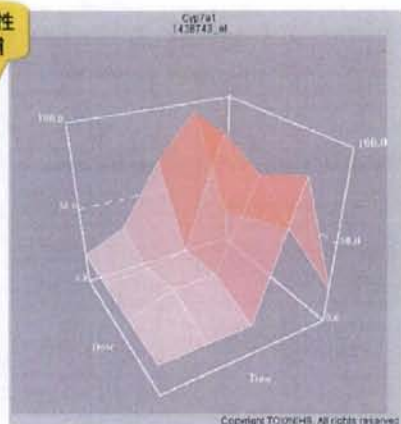


## 試行結果(Ethanol)

局所変動性  
指標候補

	正規化MAD
実	0.8359
1階差分	0.4025
1階差分(平均差)	0.5019
時間1階差分	1.6014
用量1階差分	0.3769
平均	0.9891
幾何平均	0.9039
最小	0.3769

Cyp7a1遺伝子では、8hの全用量でピークを迎えており、単純な誘導がかかっていると考えられ、1階差分でのMADが小さな値となっている。

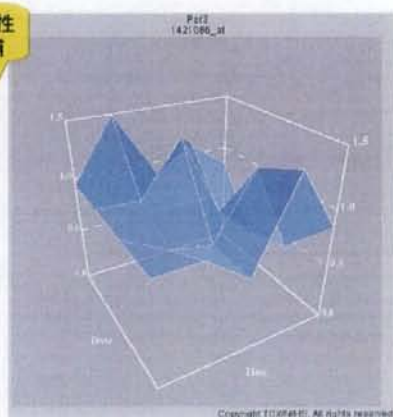


## 試行結果(Ethanol)

局所変動性  
指標候補

	正規化MAD
実	0.7055
1階差分	1.3817
1階差分(平均差)	1.3816
時間1階差分	1.2753
用量1階差分	1.3649
平均	1.3201
幾何平均	1.1172
最小	1.2753

Per3遺伝子のこのプローブセットでは、偶然の影響と思われる荒れ方をしており、1階差分での値が大きくなっている。

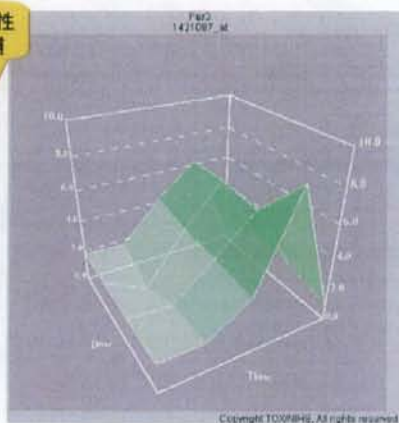


## 試行結果(Ethanol)

局所変動性  
指標候補

	正規化MAD
実	0.7247
1階差分	0.7649
1階差分(平均差)	0.8055
時間1階差分	1.2594
用量1階差分	0.2554
平均	0.7574
幾何平均	0.7971
最小	0.2554

Per3のこのブロープセットでは、8hの全用量でピークを迎えており、単純な誘導がかかっていると考えられ、1階差分でのMADが小さな値となっている。



## 試行結果(Ethanol)

局所変動性  
指標候補

	正規化MAD
実	0.8238
1階差分	1.1517
1階差分(平均差)	1.1517
時間1階差分	1.2655
用量1階差分	0.8342
平均	1.0498
幾何平均	1.0109
最小	0.8342

Per3のこの遺伝子では、ピークがわかりにくく、1階差分も大きな値となっている。

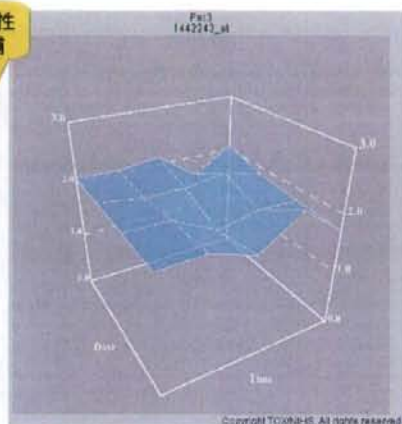


## 試行結果(Ethanol)

局所変動性  
指標候補

	正規化MAD
実	0.7825
1階差分	1.0701
1階差分(平均差)	1.0960
時間1階差分	1.1969
用量1階差分	0.9101
平均	1.0535
幾何平均	1.0173
最小	0.9101

Per3のこのプローブセットでは、ピークがわかりにくく、若干大きな値となっている。

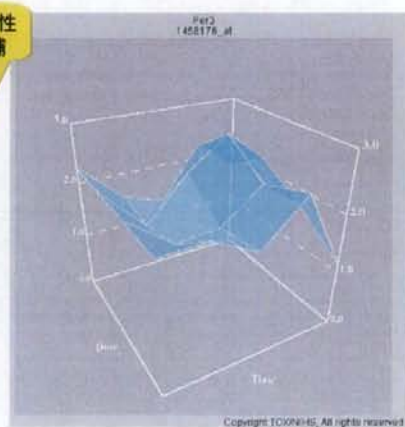


## 試行結果(Ethanol)

局所変動性  
指標候補

	正規化MAD
実	0.7677
1階差分	1.2032
1階差分(平均差)	1.1674
時間1階差分	1.2703
用量1階差分	1.0758
平均	1.1731
幾何平均	1.0645
最小	1.0758

Per3のこのプローブセットでは時間変化と用量変化があり、1階差分での値が大きくなっている。

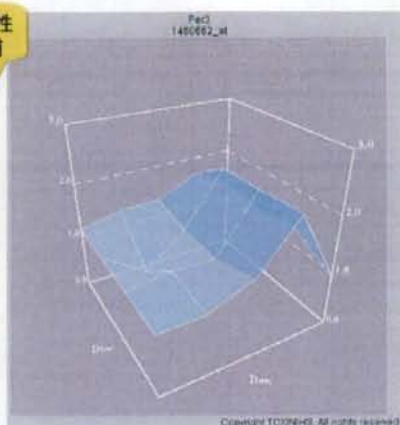


## 試行結果(Ethanol)

局所変動性  
指標候補

	正規化MAD
実	0.4851
1階差分	0.8281
1階差分(平均差)	0.8202
時間1階差分	1.5005
用量1階差分	0.2900
平均	0.8952
幾何平均	0.8467
最小	0.2900

Per3のこのプローブセットでは、8hでピークがみられ、若干小さな値となっている。

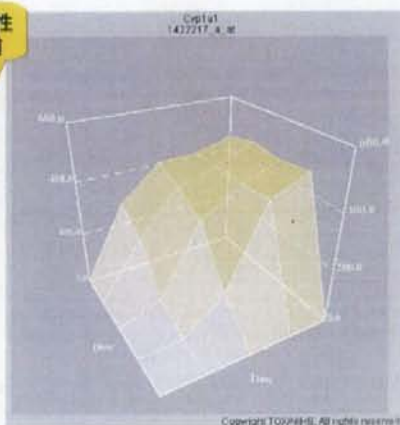


## 試行結果(TCDD)

局所変動性  
指標候補

	正規化MAD
実	0.9053
1階差分	0.2799
1階差分(平均差)	0.3959
時間1階差分	0.2697
用量1階差分	0.4910
平均	0.3803
幾何平均	0.6674
最小	0.2697

TCDDにより大きな誘導のかかる遺伝子であり、1階差分の値も小さくなっている



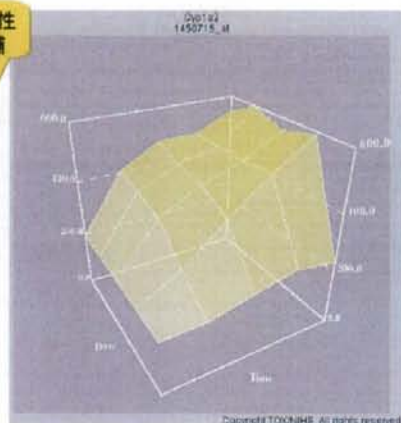


## 試行結果(TCDD)

局所変動性  
指標候補

	正規化MAD
実	0.6097
1階差分	0.4056
1階差分(平均差)	0.2829
時間1階差分	0.2361
用量1階差分	0.4564
平均	0.3462
幾何平均	0.6404
最小	0.2361

TCDDにより大きな誘導のかかる遺伝子であり、1階差分の値も小さくなっている

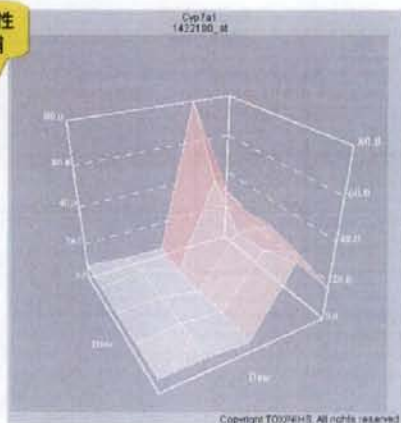


## 試行結果(TCDD)

局所変動性  
指標候補

	正規化MAD
実	0.6699
1階差分	0.2460
1階差分(平均差)	0.3787
時間1階差分	1.0470
用量1階差分	0.2190
平均	0.6330
幾何平均	0.7449
最小	0.2190

8hで大きく誘導がかかっており、1階差分の値も小さくなっている

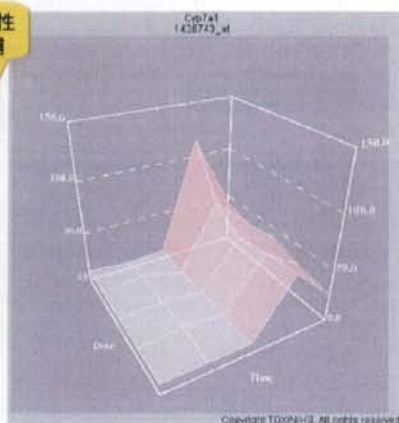


## 試行結果(TCDD)

局所変動性  
指標候補

	正規化MAD
実	0.7587
1階差分	0.2792
1階差分(平均差)	0.3882
時間1階差分	1.3151
用量1階差分	0.2428
平均	0.7790
幾何平均	0.7959
最小	0.2428

8hで大きく誘導がかかっており、1階差分の値も小さくなっている



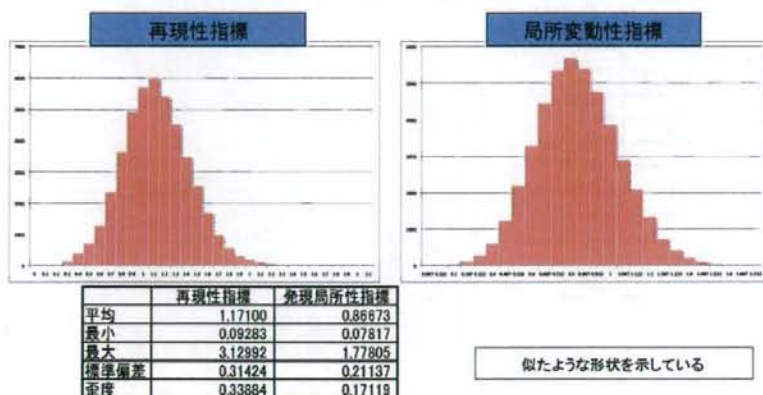
## 試行結果まとめ

数値と形状から次のような関係があるのではないかと想像される

範囲	状況
0.4以下	特殊な形状(サーカディアン、TCDDのCyp1a1)
0.4~0.8	意味のありそうな形状
0.8~1.0	判断がつきにくい形状
1.0以上	あまり、意味があるとは思えない形状

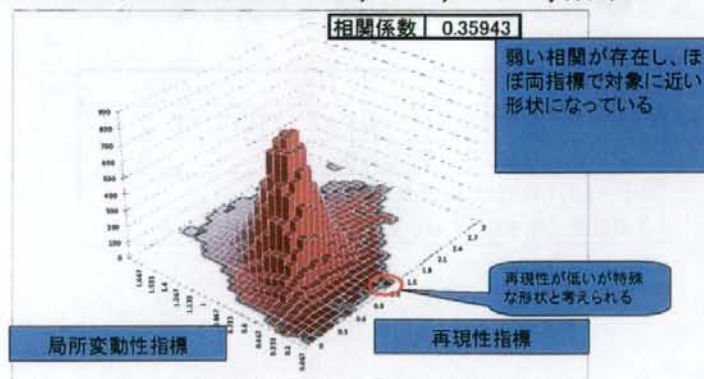
## 局所変動性指標 試行計算結果

- 全プローブセットに対して計算を実施し、再現性指標と比較した  
- TTG020(Chemical=TCDD, PJ=1, SFC=3)結果



## 局所変動性指標 試行計算結果

- TTG020(Chemical=TCDD, PJ=1, SFC=3)結果



## 局所変動性指標 試行計算結果

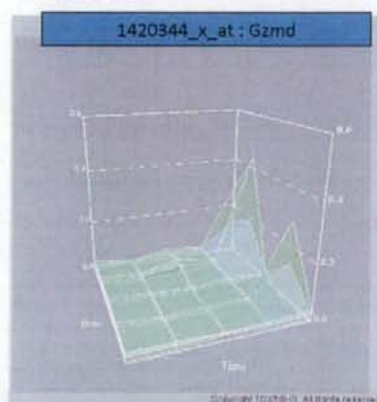
- 再現性が低い特徴的な形状と考えられる遺伝子

prb_no	sd_sgnl	mad	ant_val	Prb_ID
4650	1.3173	0.1460	Gzmd	1420344_x_at
4744	1.4908	0.0919	Orm2	1420438_at
10461	1.0746	0.1613	Osr2	1426155_a_at
12053	1.4731	0.1066	Lcn2	1427747_a_at
33626	1.2750	0.1956	Saa2	1449326_x_at

特定時刻、用量で再現性が悪い。

## 局所変動性指標 試行計算結果

再現性が低い、特殊な形状



24時間後、Low及びMidHighで、再現性が悪い