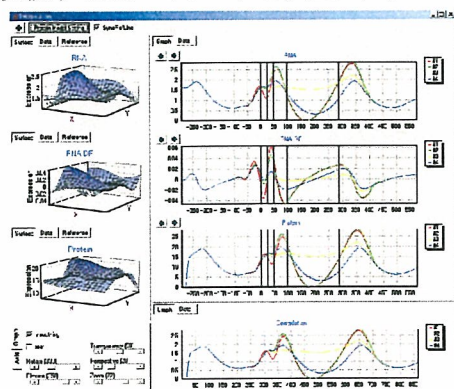


時系列ソーティングのための基本技術として、測定点間のデータ補間および日内変動リズムを考慮した測定区間外へのデータ外挿を検討した。実際の測定値を正確に反映し、なおかつ測定点間を滑らかに接続すること、および補間関数自体を微分・積分できることから補間関数としてスプライン関数を採用した。測定区間外へのデータ外挿に際しては測定区間に於ける溶媒コントロールの補間関数をベースとして化合物投与による変動曲線に滑らかに接続するようパラメータを調整した。



D. 考察

1. 発現同期遺伝子の抽出・解析技術の開発・改良

今回開発した発現同期率計算結果の閲覧ソフトウェアは、同期率計算結果に応じて遺伝子情報や各化合物投与実験での発現パターン、さらには同じクラスタに所属する他の遺伝子の情報などを簡単に検索・表示できるようになっているが、基本的にまだキャラクターベースの表示であり、同期率計算結果全体をより直感的に捉えられるようなグラフィカル表示の開発が課題である。

2. エクソンレベルでの遺伝子発現データの解析技術開発

数値化アルゴリズム開発にはa)プローブ毎

のバックグラウンド値の取り扱い方法とb)同一エクソンに属する複数プローブの値をまとめる方法の2点を解決する必要がある。aについては標準サンプルを準備して、その実測値からアルゴリズムの考案・開発を目指す。bについては同一遺伝子(正確には同一メッセンジャーRNA)に属するエクソン間でデータ値レベルが同等になるようなパラメータ設定方法を模索する。

3. プラットホーム間データ相互変換技術の開発

用量相関性の悪いプローブ(セット)についてはマイクロアレイのプローブ設計を変更しなければならないため、考察の対象外とする。各プラットホームにおいて用量相関性が良好なプローブ(セット)であるにも関わらずデータ相互変換の信頼性が悪かったものは標準データにおいて発現値が大きすぎて飽和しているか、小さすぎて測定値がきちんと得られていないかのいずれかが多かった。これらの問題は標準データを増やすことである程度解消されると予測された。

4. 時系列ソーティング技術の基礎開発

生成した変動曲線方程式は微積分可能であるため、発現誘導やタンパク質合成のパターンも推測可能であり、今後はアルゴリズムの改良を進めつつ、得られた推測式を利用して時系列シフトした発現パターンや推測された発現誘導因子パターン・タンパク質合成パターンを鋳型としてシグナルへの関与の可能性のある遺伝子の抽出を検討する。

E. 結論

今回、トキシコゲノミクスのためのデータ変

換技術や高次解析アルゴリズムなどの開発を行った。発現同期率計算・評価やプラットフォーム間データ変換技術、時系列ソーティングのための補間・外挿アルゴリズム開発については一定の成果を得た。今後はこれら技術を基礎とした応用研究を進める。

一方、エクソンアレイについては既存の数値化アルゴリズムや基本性能に問題が見つかったため、今後、研究計画よりもさらに一步踏み込んだ性能評価と技術開発を行うこととする。

F. 健康危惧情報

特になし

G. 研究発表

1. 論文発表

1) 書籍

なし

2) 雑誌

なし

2. 学会発表

1) Percellome データ解析システム (Millefeuille システム)の開発[第 33 回日本トキシコロジー学会学術年会] 2006 年 7 月 (相崎 健一、中津 則之、北嶋 聡、五十嵐 勝秀、菅野 純)

2) Millefeuille システム～Percellome プロジェクトの解析システム[第 1 回毒性オミクスフォーラム(第 33 回日本トキシコロジー学会学術年会との共催)] 2006 年 7 月 (相崎 健一)

H. 知的財産所有権の出願・登録情報

1. 特許取得

なし

2. 実用新案登録

なし

3. その他

なし






















「非形質依存的」の真意

- 本研究の最終目標は、形質である毒性症状に結びついた解析結果を求めること、である。
- しかしながら、形質が明らかに誘導される条件で遺伝子発現を解析(=形質依存的な解析)しても、形質を誘導する引き金となった分子標的や分子機序を見いだすことは困難である。
- そこで敢えて、形質が顕在化しないギリギリの条件で初期～中期における遺伝子発現を解析(=非形質依存的な解析)し、始原反応の捕捉・解析を試みる。

背景技術

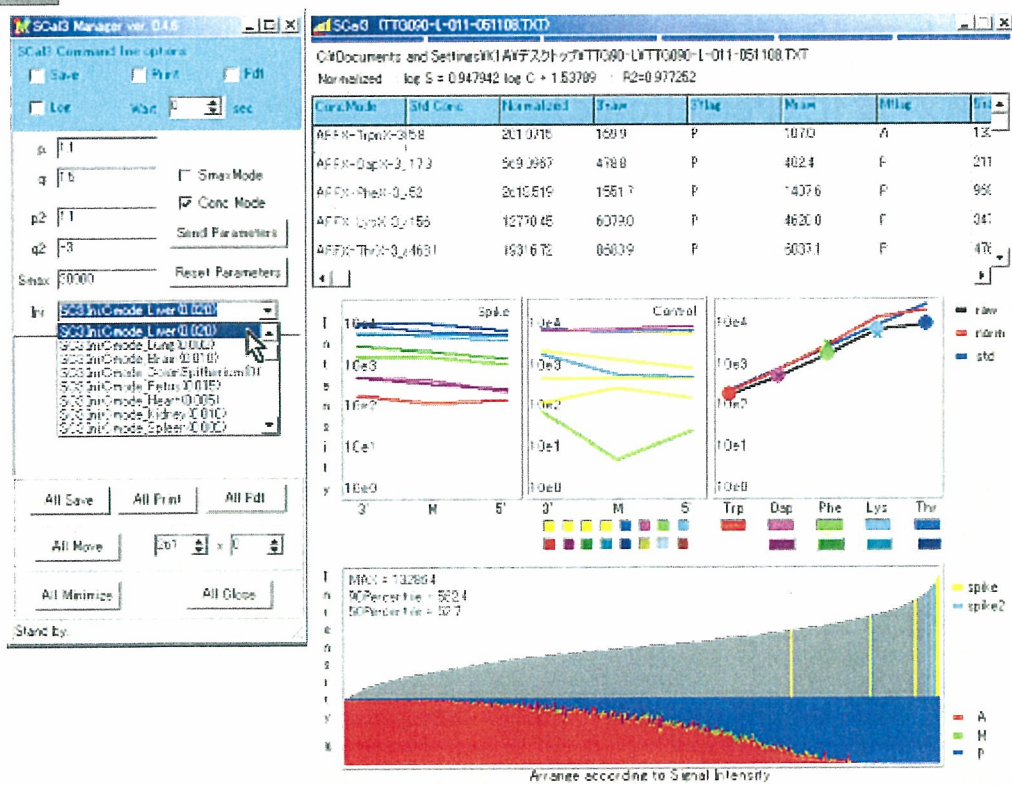
- データ変換技術
 - 絶対量化(PerCellome)
 - 亜飽和データの補正
- 遺伝子発現の可視化技術
 - Surface
- 候補遺伝子の抽出技術
 - 発現パターン類似度評価(tmf)
 - アンスーパーバイズドクラスタリング
 - 発現パターンの整・不整評価(RSort)
- データマイニング技術
 - GoPlot

MF tools

Percellome計算、QC	 SCal,  MFScatter,  MFSurface
データ可視化	 MFSurface,  MFScatter,  MFLine,  MFHist
発現パターン類似度評価	 MFMatch,  MFCMF2
実験間比較	 MSV2,  MFCompare,  MFHist
クラスタリング	 MFCV,  MFGalaxy,  MFSyncCluster
遺伝子リスト操作	 MFFilter,  GLM,  GLCalc,  GE
データ解析	 MFAnalyzer,  MFGoPlot

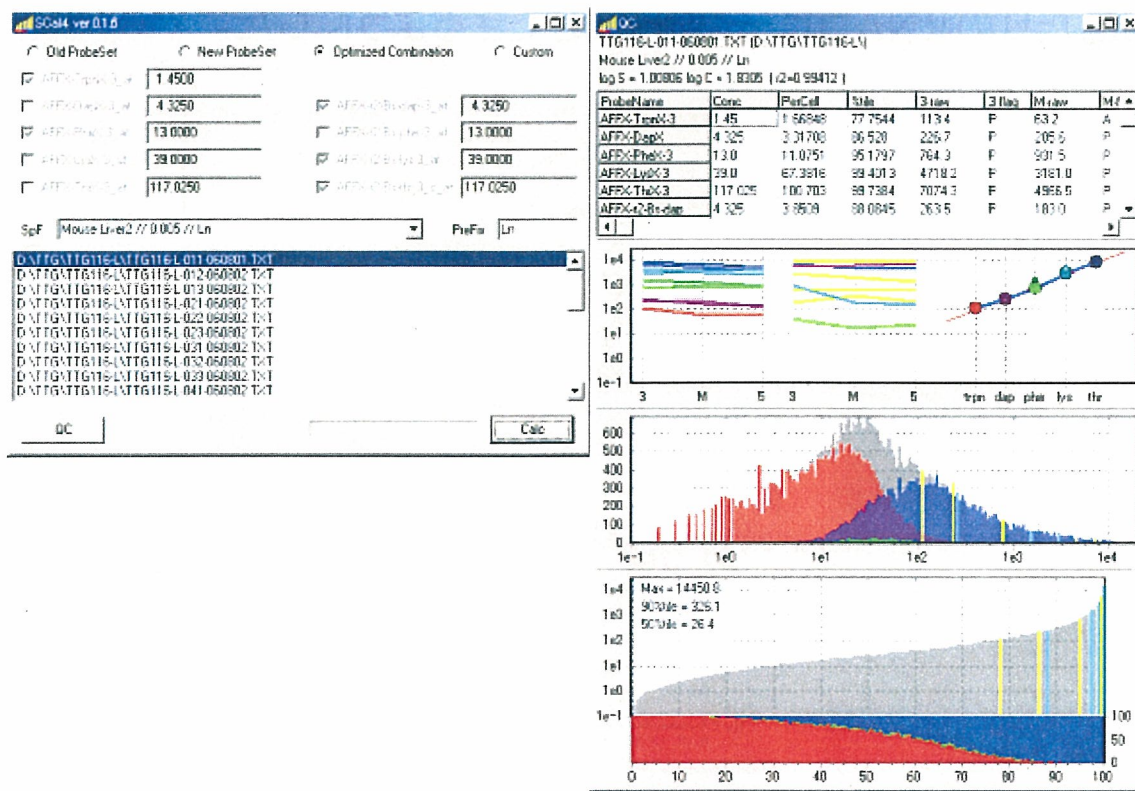
Percellome Normalization (細胞1個当たりのmRNAコピー数換算)

SCal3

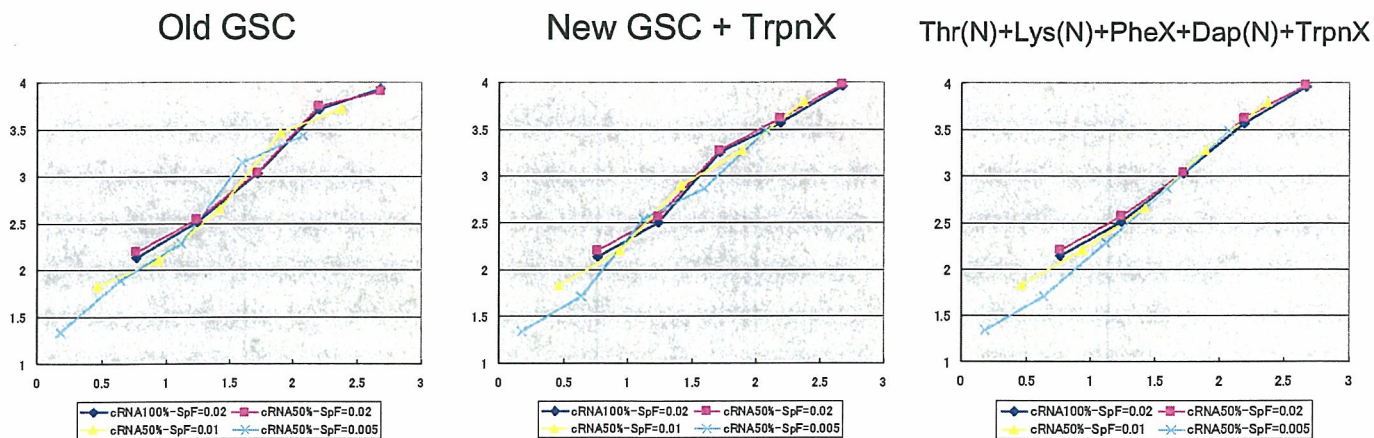


Percellome Normalization (細胞1個当たりのmRNAコピー数換算)

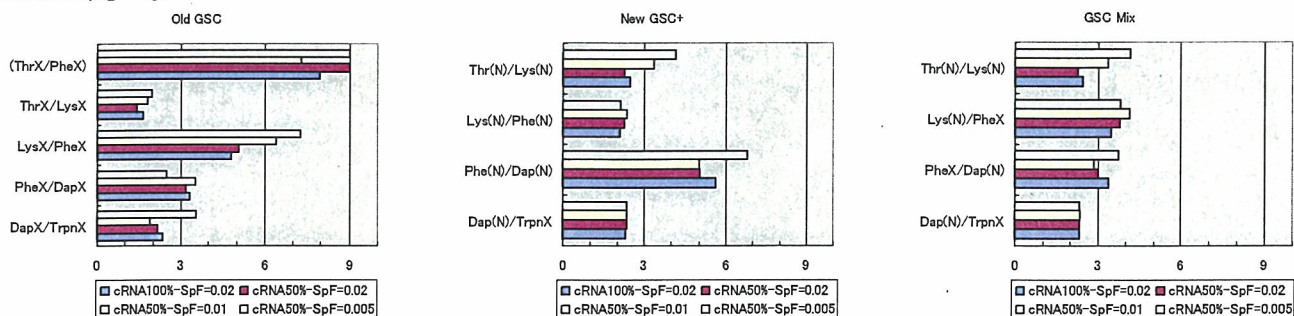
SCa14



GSCプローブの組み合わせ最適化によるデータ精度の改善

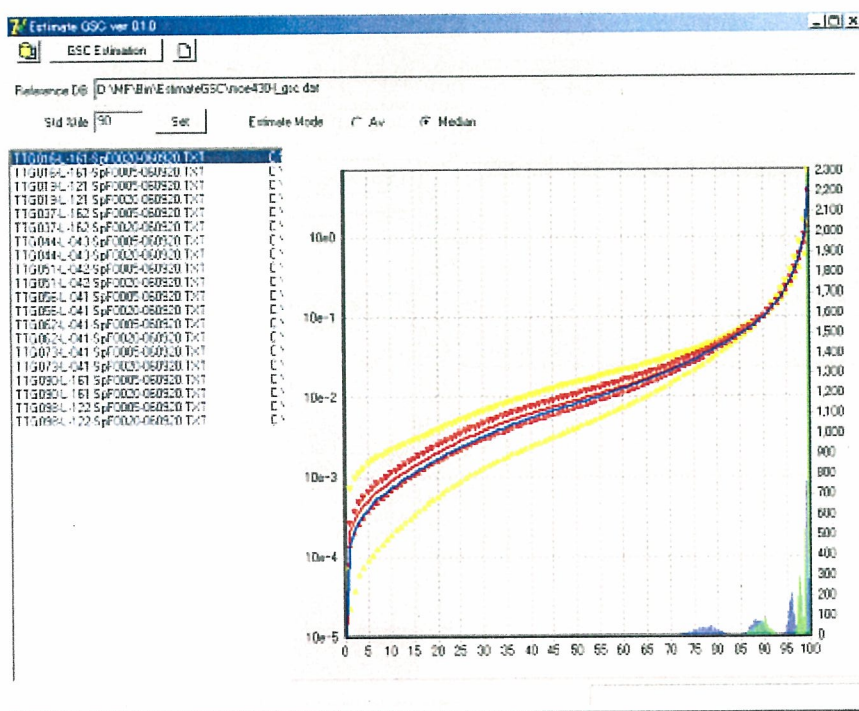


公比分布



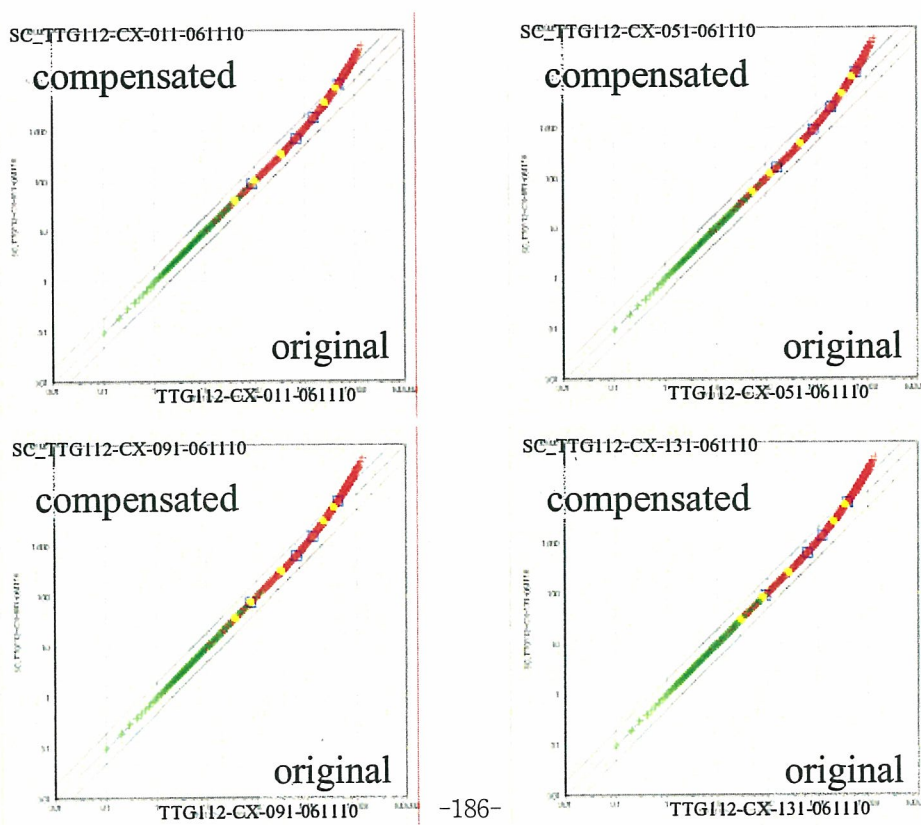
EstimateGSC

GSCを添加していなかった過去のサンプルについても条件付きながら、細胞1個あたりのRNAコピー数を推定できるようにした。



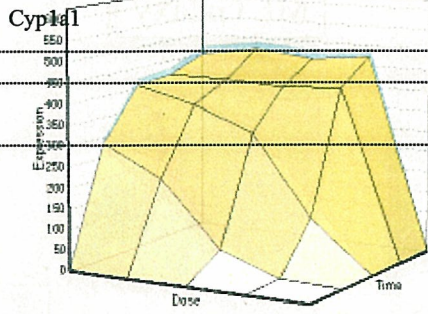
SCancel ver.0.2.0

～ GeneChipの亜飽和データを補正する ～



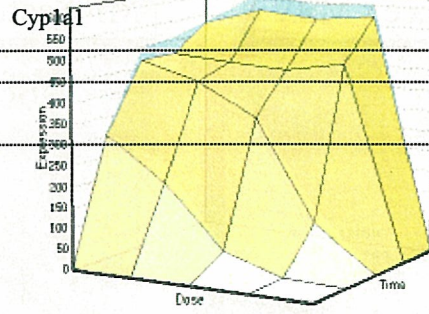
TTG020_L_SpNC_0_2217_3.st

standard



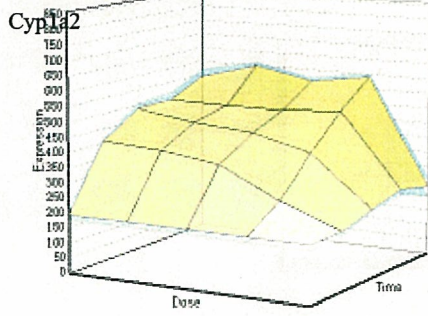
TTG020_L_SpNC_0_2217_3.st

PScombination + Saturation-Canceled



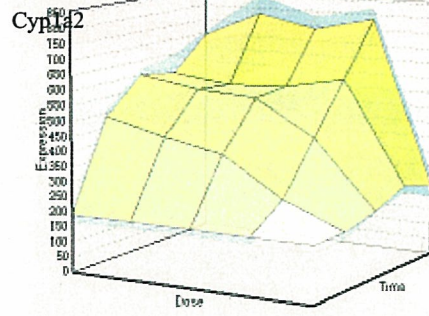
TTG020_L_SpNC_0_20715_3.st

standard



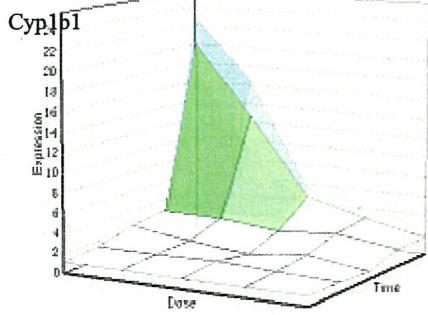
TTG020_L_SpNC_0_20715_3.st

PScombination + Saturation-Canceled



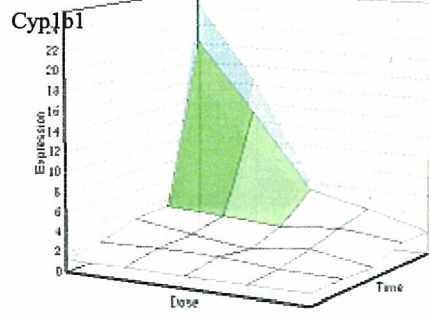
TTG020_L_SpNC_0_19513_3.st

standard



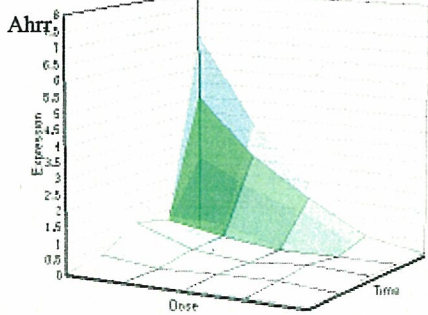
TTG020_L_SpNC_0_19513_3.st

PScombination + Saturation-Canceled



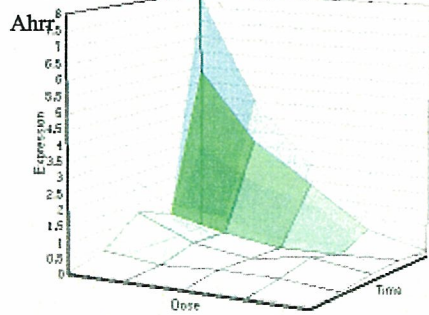
TTG020_L_SpNC_0_20795_3.st

standard



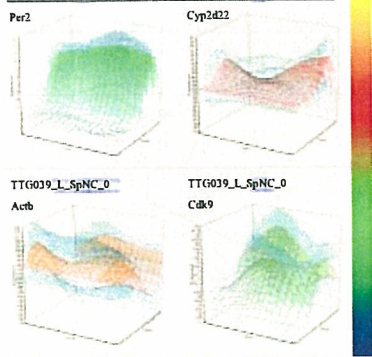
TTG020_L_SpNC_0_20795_3.st

PScombination + Saturation-Canceled

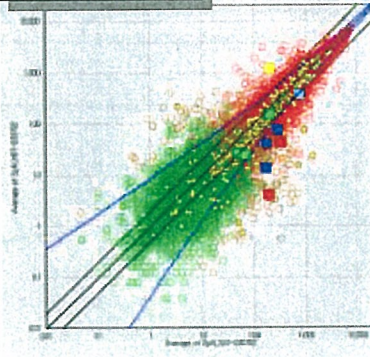


遺伝子発現パターンの可視化

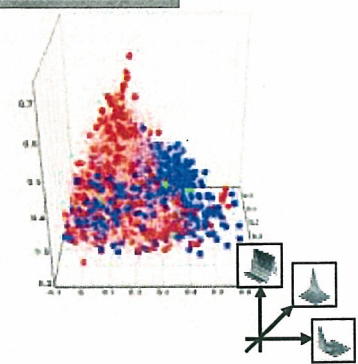
MF Surface / MSV



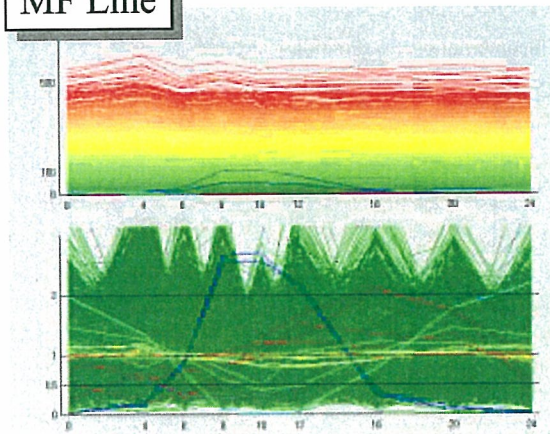
MF Scatter



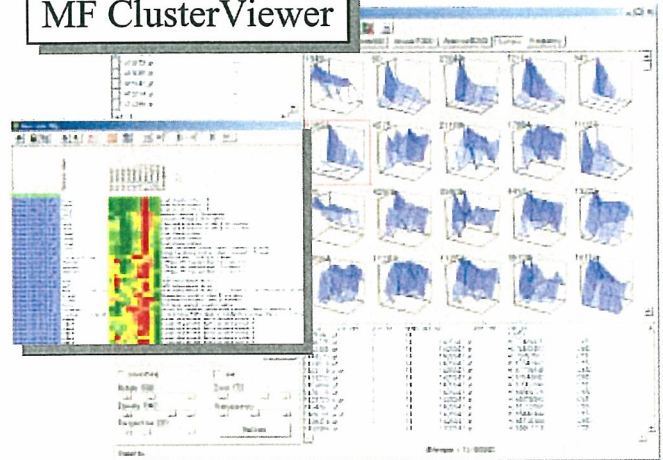
MF Galaxy



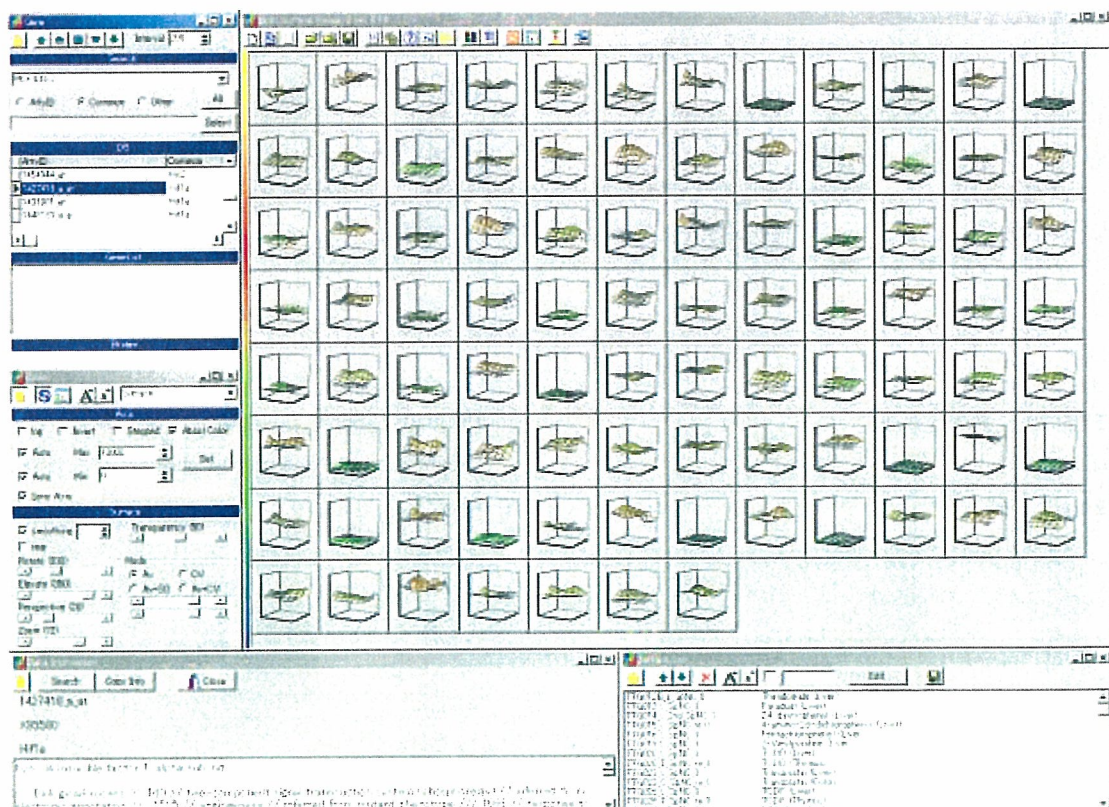
MF Line

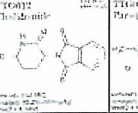
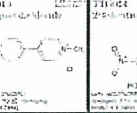
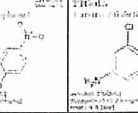
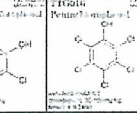
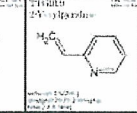
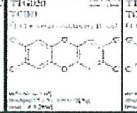
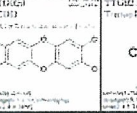
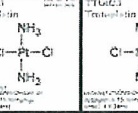
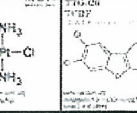
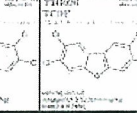
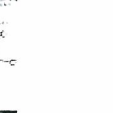


MF ClusterViewer

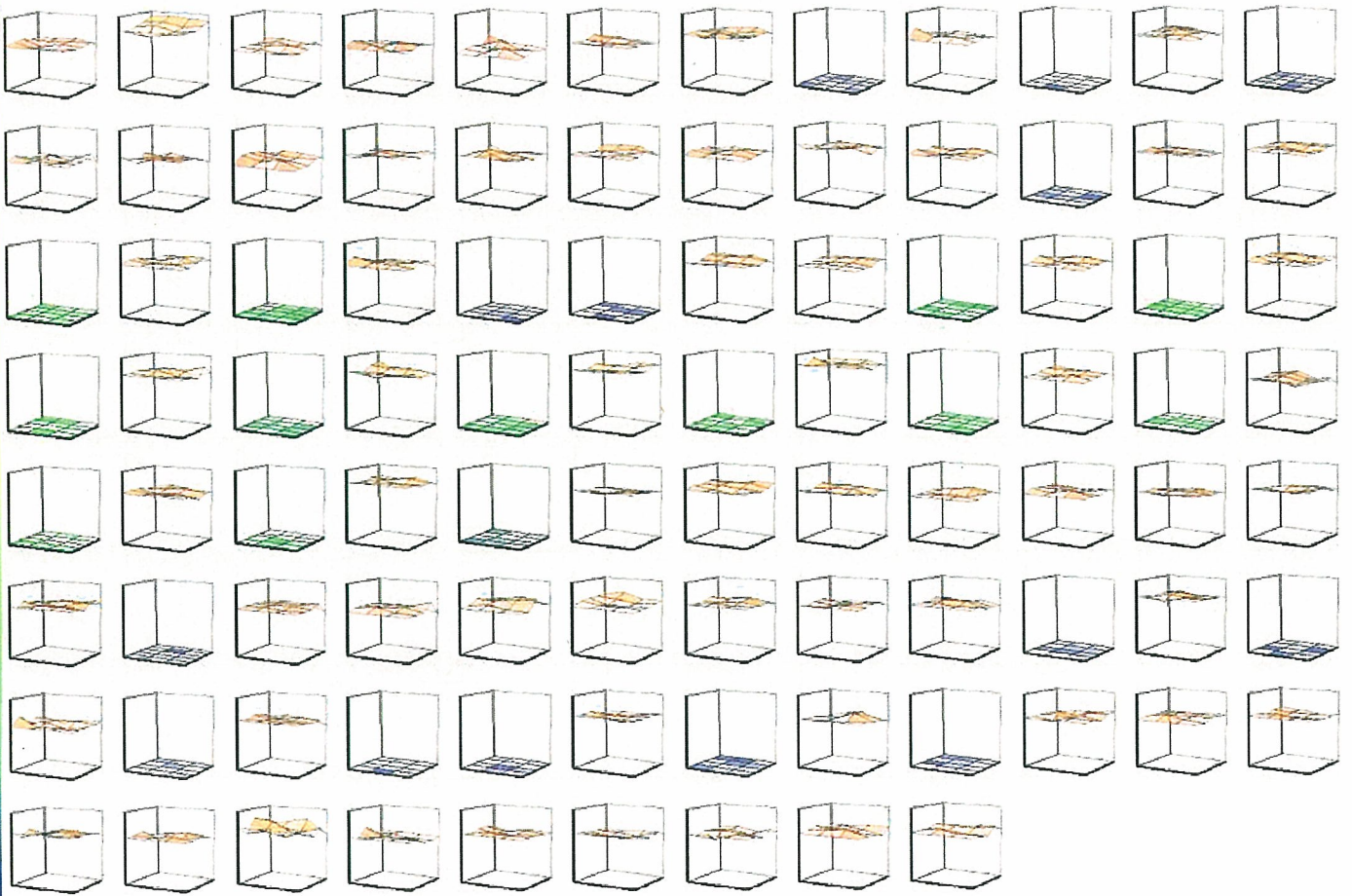


発現パターンの化合物間評価 (MSV2)

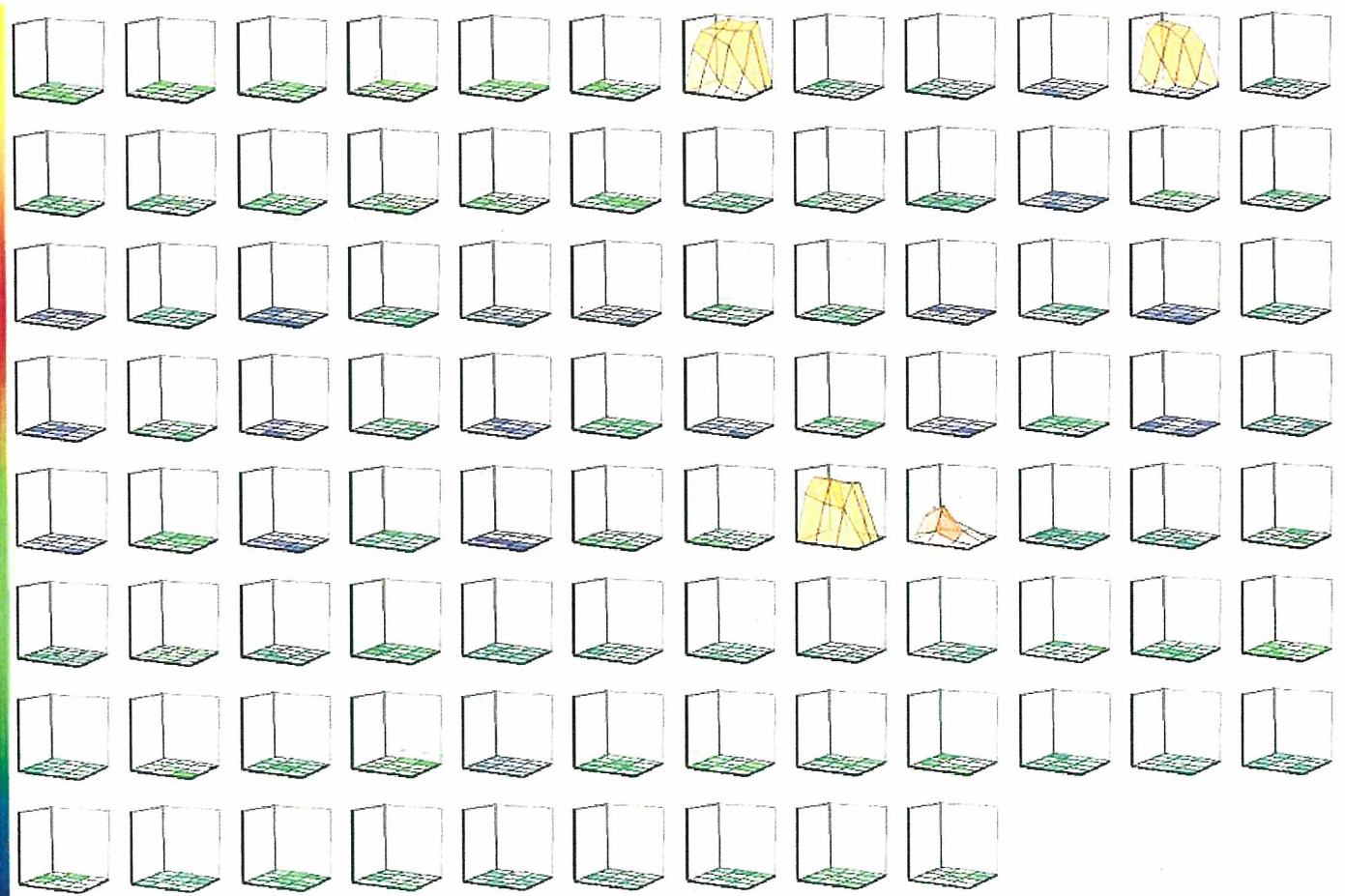


<p>TH002 Thalidomide</p> 	<p>TH003 Paraquat dichloride</p> 	<p>TH004 2,4-Dinitrophenol</p> 	<p>TH005 4-amino-2,6-dichlorophenol</p> 	<p>TH006 Pentachlorophenol</p> 	<p>TH007 2-Vinylpyridine</p> 	<p>TH008 TCDD</p> 	<p>TH009 TCDD</p> 	<p>TH010 Transplatin</p> 	<p>TH011 Transplatin</p> 	<p>TH012 TCDF</p> 	<p>TH013 TCDF</p>
---	--	--	---	--	--	--	---	--	--	---	---

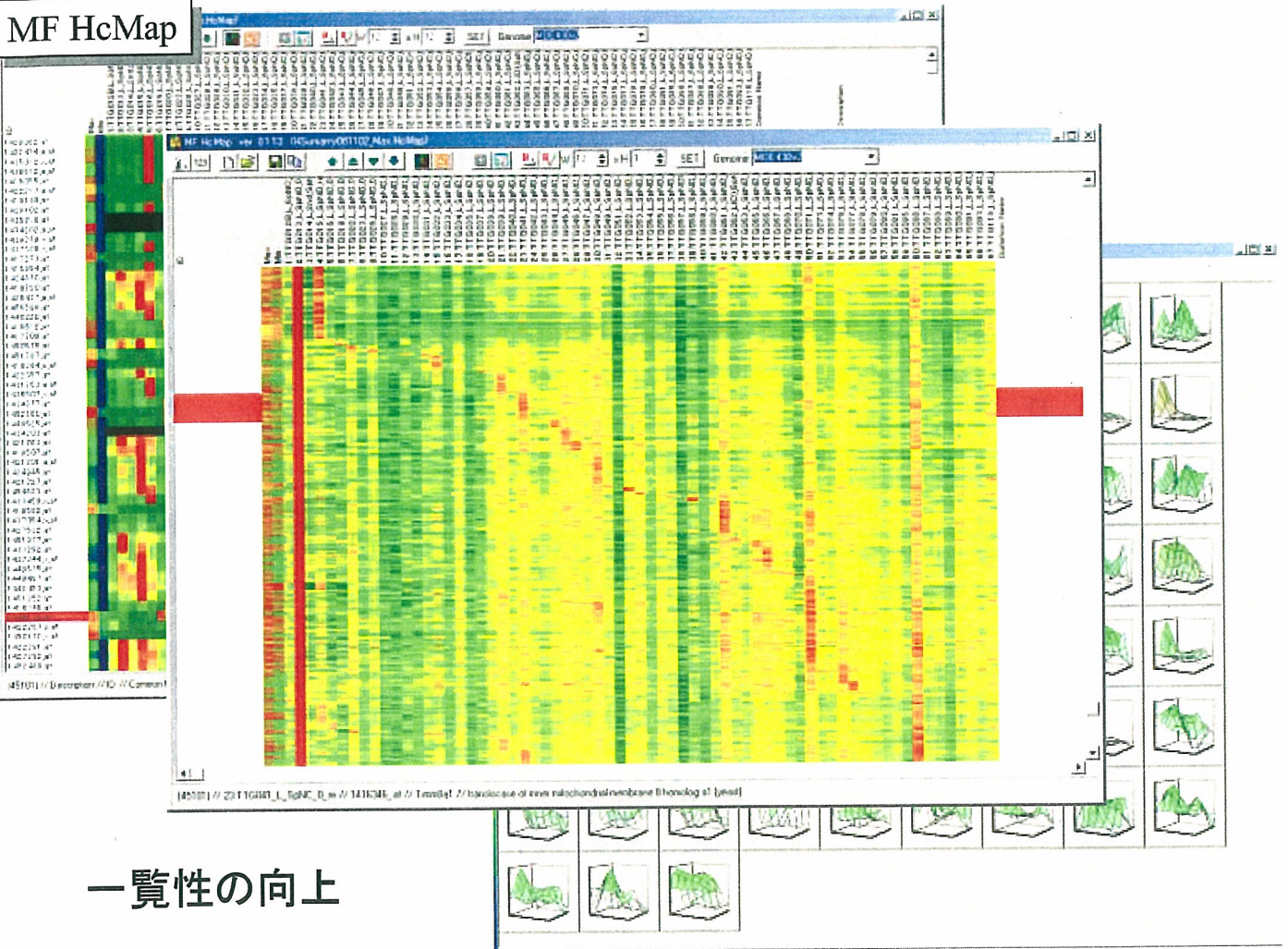
Thalidomide (Liver)	Paraquat (Liver)	2,4-dinitrophenol (Liver)	4-amino-2,6-dichlorophenol (Liver)	Pentachlorophenol (Liver)	2-Vinylpyridine (Liver)	TCDD (Liver)	TCDD (Thymus)	Transplatin (Liver)	Transplatin (Colon)	TCDF (Liver)	TCDF (Thymus)
1,2,3-Triazole (Liver)	1,2,4-Triazole (Liver)	2-Aminomethylpyridine (Liver)	N-Methylaniline (Liver)	2-Chloro-4,6-dimethylaniline (Liver)	3-Amino-1H-1,2,4-triazole (Liver)	1,2-Dichloro-3-nitrobenzene (Liver)	4-Ethynitrobenzene (Liver)	Acetaminophen [2hd] (Liver)	Acetaminophen [2hd] (Kidney)	Phenobarbital [2nd] (Liver)	Hydroxyctic Acid (Liver)
Hydroxyctic Acid (Testis)	Citric acid-calcium salt (Liver)	Citric acid-calcium salt (Testis)	Cisplatin [2nd] (Liver)	Cisplatin [2nd] (Colon)	Cisplatin [2hd] (Kidney)	Valproic Acid [2nd] (Liver)	Ethynyl estradiol (Liver)	Ethynyl estradiol (Testis)	Testosterone propionate (Liver)	Testosterone propionate (Testis)	Clofibrate (Liver)
Clofibrate (Testis)	Troglitazone (Liver)	Troglitazone (Testis)	Levothyroxine (Liver)	Levothyroxine (Testis)	Bisphenol A (Liver)	Bisphenol A (Testis)	Genistein (Liver)	Genistein (Testis)	Tributyltin (Liver)	Tributyltin (Testis)	9-cis retinoic acid (Liver)
9-cis retinoic acid (Testis)	All trans retinoic acid (Liver)	All trans retinoic acid (Testis)	Ethynyl estradiol (Liver)	Ethynyl estradiol (Testis)	Diethylnitrosamine (C57BL/6) (Liver)	N-ethyl-N-nitrosourea (Liver)	3-methylcholanthrene (Liver)	Indigo (Liver)	Diethylnitrosamine (C3H) (Liver)	Caffeine (Liver)	Forskolin (Liver)
Paraquat [2nd] (Liver)	Paraquat [2nd] (Lung)	Dexamethasone (Liver)	Sodium Arsenite (Liver)	Methoprene acid (Liver)	Methoprene (Liver)	Acephate (Liver)	Carbaryl (Liver)	Formalin (Liver)	Formalin (Lung)	Acetaldehyde (Liver)	Acetaldehyde (Lung)
Toluene (Liver)	Toluene (Lung)	Bromobenzene (Liver)	Bromobenzene (Lung)	Bromobenzene (Kidney)	Carbon tetrachloride (Liver)	Carbon tetrachloride (Lung)	Monocrotaline (Liver)	Monocrotaline (Lung)	Ethanol (Liver)	Methanol (Liver)	DMSO (Liver)
Warfarin (Liver)	Tributyltin (Liver)	Coenzyme Q10 (Liver)	Pyriproxyfen (Liver)	Tebufenozide (Liver)	Rifampicin (Liver)	Pregnenolone Carbonitrile (Liver)	Azacytidine (Liver)	AraC (Liver)			



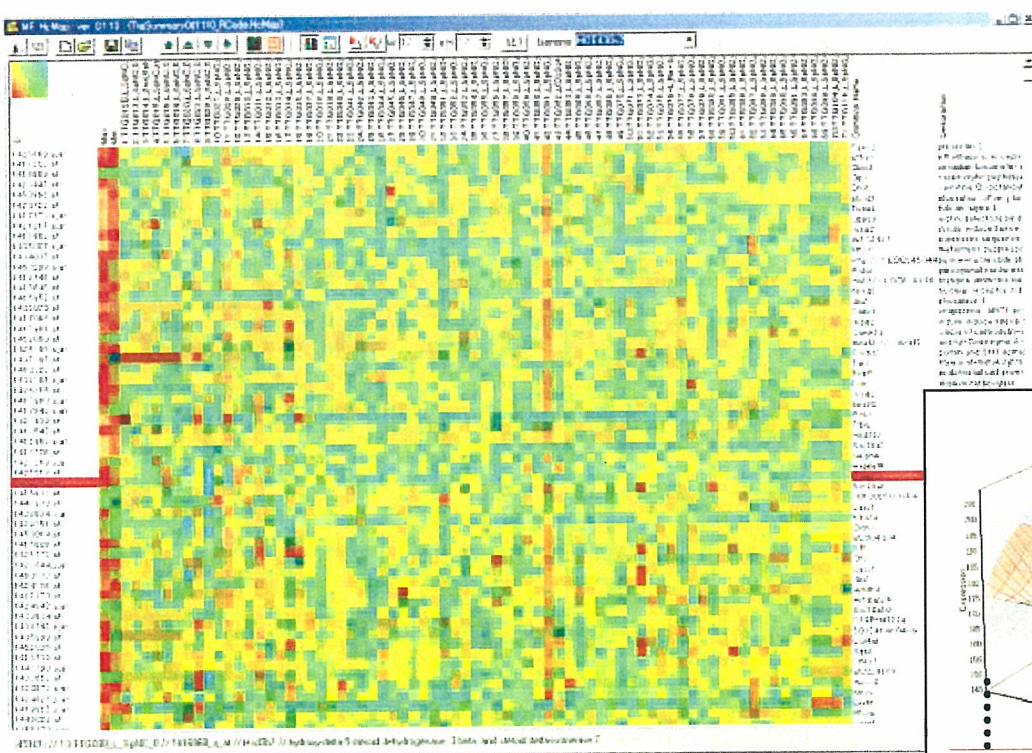
Cyp2a12



Cyp1a1

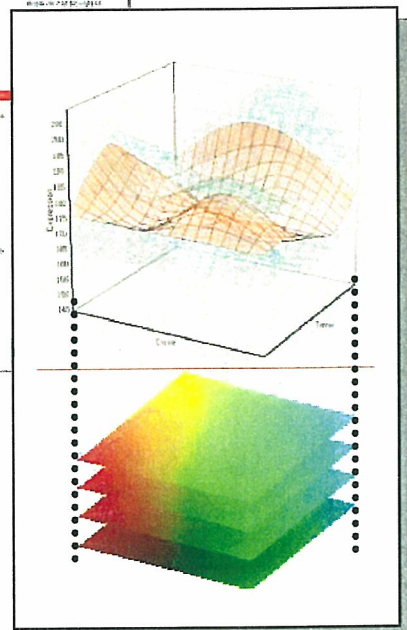


一覧性の向上

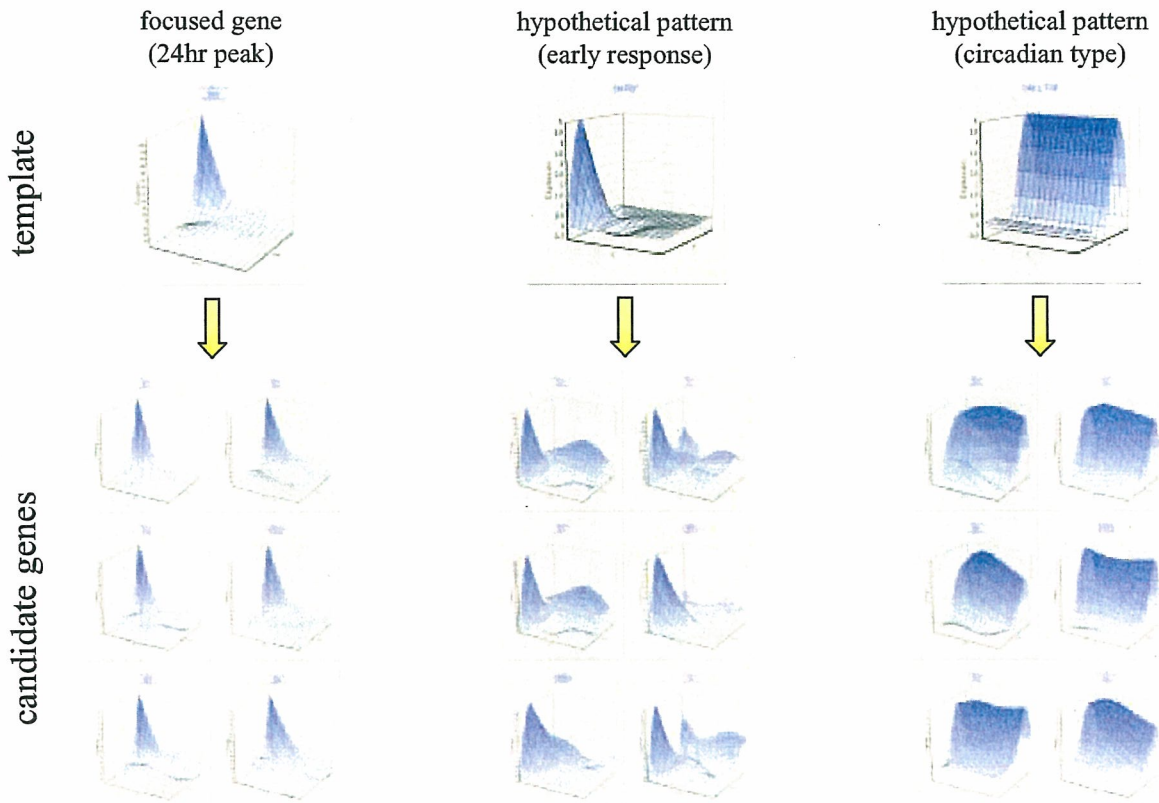


形状特徴抽出技術 (RSort) を応用した
データ可視化例

最大ピーク位置(条件) = 色相
ピーク数 = 彩度

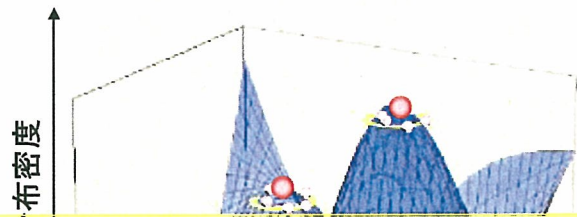
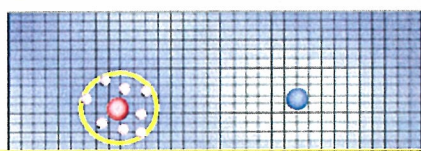


類似した発現パターンをとる遺伝子の抽出 (TMF)



クラスタリングのコンセプト

発現パターン類似度空間の密度分布



- 5万プローブセットを直接、約1000個のクラスタに分ける。
- 同時に、独特の発現パターンを呈するプローブセットも数百個抽出される。

- Rigorous (クラスタの代表遺伝子)
- Shell (クラスタのメンバー遺伝子)
- Isolate (2遺伝子からなるクラスタ)
- Unique (1遺伝子からなるクラスタ)

Matsumono, S., Aisaki, KI., Kanno, J. Mass Distributed Clustering: A New Algorithm for Repeated Measurements in Gene Expression Data [Genome Informatics 16 183 2005]

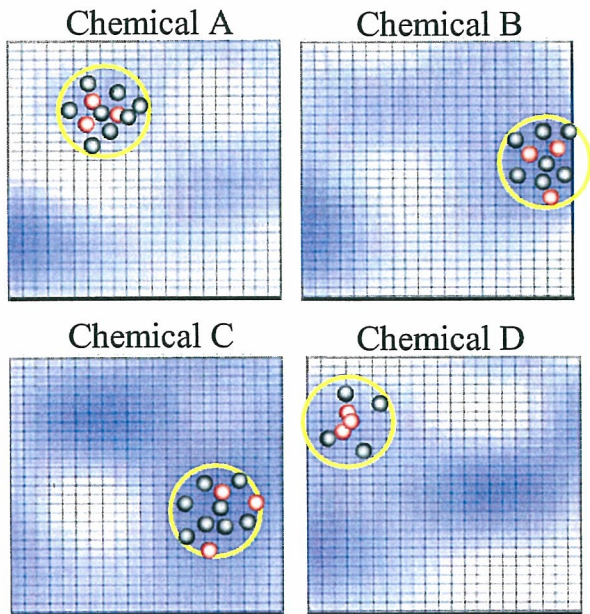
研究計画

- データ(ベース)の量的・質的な拡張に関わる技術
 - プラットホーム間データ変換技術
 - Exonアレイデータによる遺伝子発現の詳細解析技術
- データ解析に関わる技術
 - 発現同期遺伝子の抽出・解析技術
 - 時系列ソーティング技術
 - 化合物反応(毒性反応)の類似性評価技術
- データそのものの取り扱い技術
 - 遺伝子リストの大量・高速ハンドリング方法

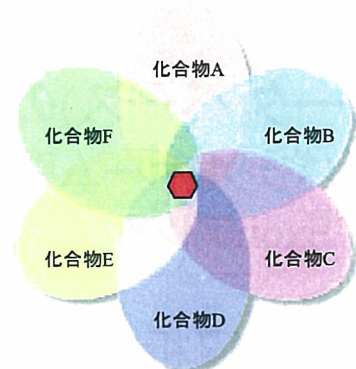
H18年度計画

- 発現同期遺伝子の抽出・解析技術の開発・改良
- Exonレベルでの遺伝子発現データの解析技術開発(1)
- プラットフォーム間データ変換の検証

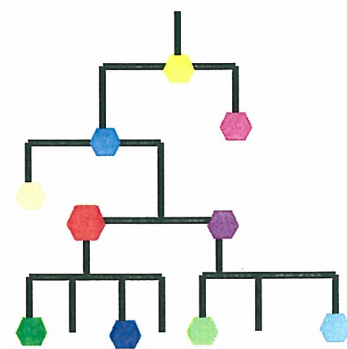
発現同期遺伝子の抽出・解析技術の開発・改良



- 常に同期発現しているクラスタ要素
- その他のクラスタ要素
- 発現パターンの類似領域



シグナルカスケード

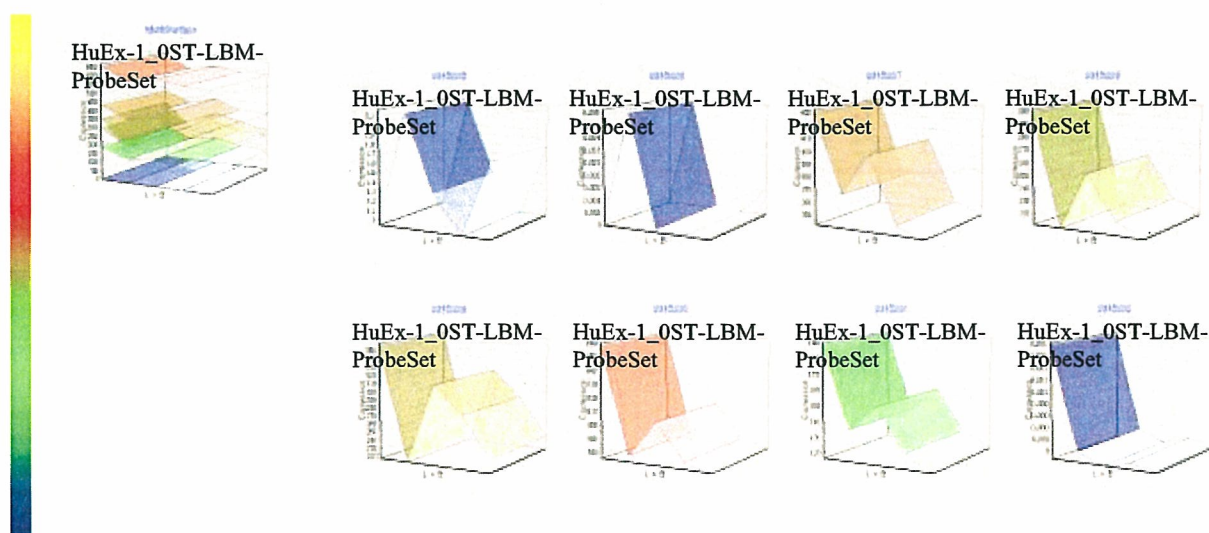


同期率計算結果の効率的な閲覧方法の模索

Exonレベルでの遺伝子発現データの解析技術開発(1)

- LBMサンプルによってプローブセット毎の定量性・再現性等の基本性能を評価。

→ 同一遺伝子のエクソン間でのデータ乖離が顕著であることが判明。まず数値化アルゴリズムの改良による対処を試みる。



プラットフォーム間データ変換の検証

Sample: TCDD-administrated mice liver

Dosage: Vehicle, Low, Middle-low, Middle-High, High

Sampling Time: 2, 4, 8, 24hr after TCDD administration

→ 5 x 4 x triplicate = 60 samples

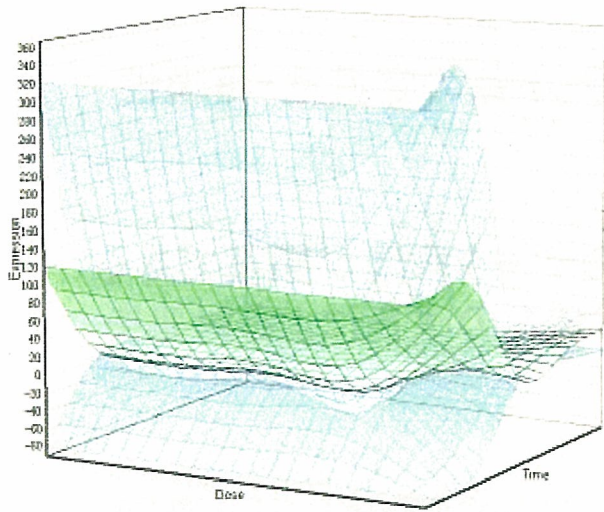
Platforms: Affymetrix MOE430v2

Agilent NIHS44K

Q-PCR

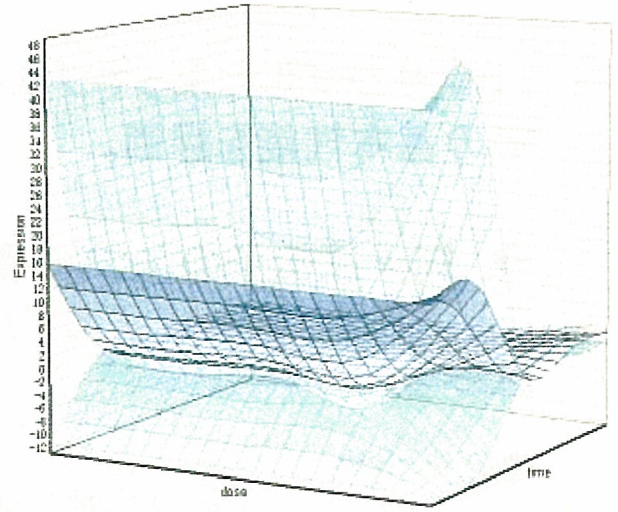
Affymetrix MOE430v2

1427747_a_at



Agilent PercellomeArray

16769

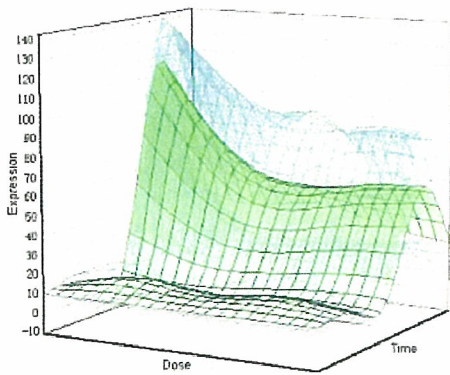


Lcn2

- [-] A2A_ByUnigene@
 - [+] MOE430v2
 - 1427747_a_at
 - [+] AgilentID
 - A_51_P510156
 - [+] AgilentNIHS44K
 - 16769

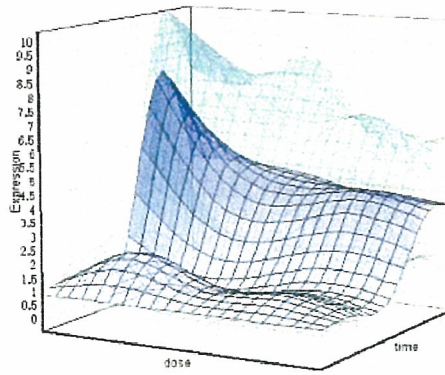
Affymetrix MOE430v2

1438743_at



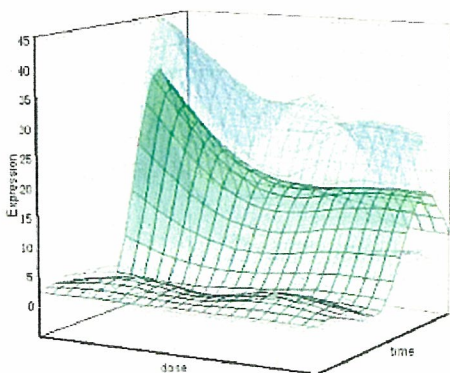
Agilent PercellomeArray

20078



Agilent PercellomeArray

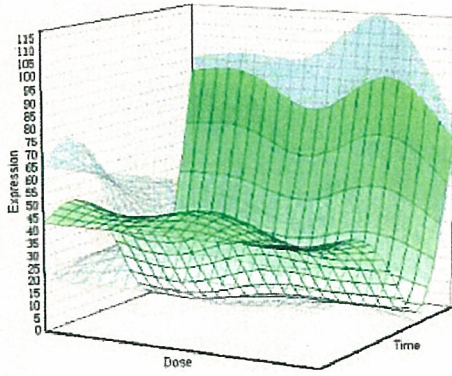
40756



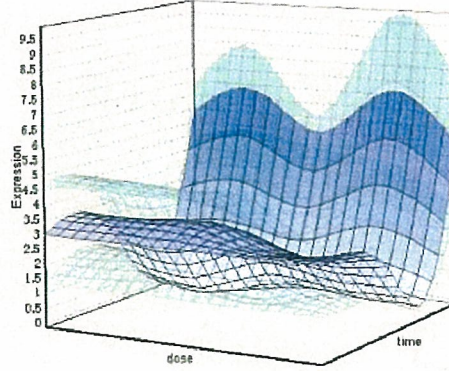
- [-] A2A_ByUnigene@
 - [+] MOE430v2
 - 1438743_at
 - [+] AgilentID
 - A_51_P290981
 - A_52_P84027
 - [+] AgilentNIHS44K
 - 20078
 - 40756

Cyp7a1

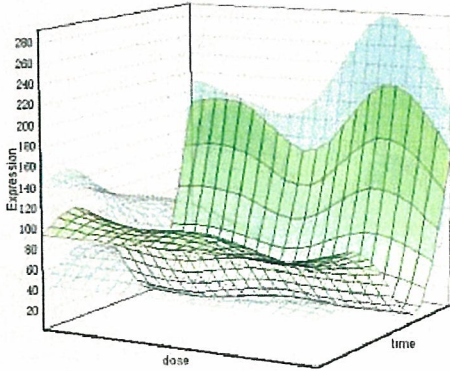
Affymetrix MOE430v2 1427347_s_at



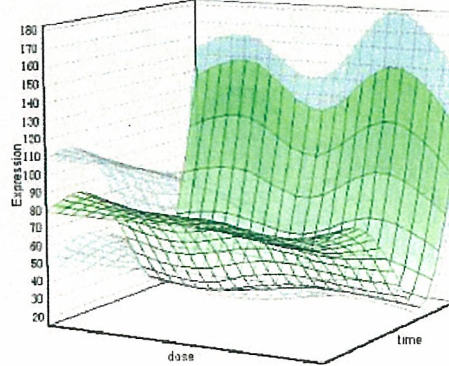
Agilent PercellomeArray 17388



Agilent PercellomeArray 25843



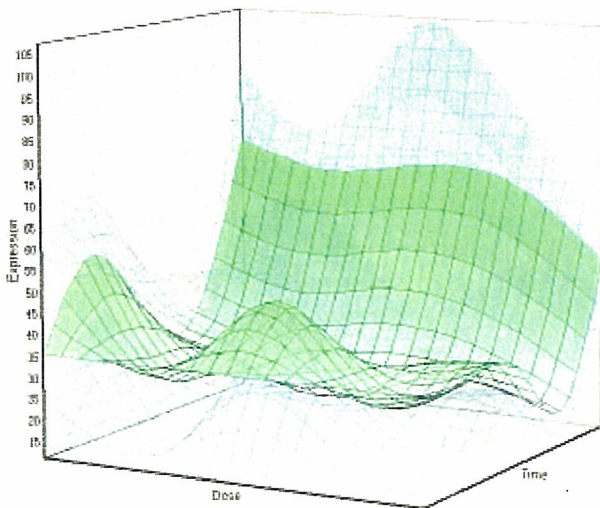
Agilent PercellomeArray 32223



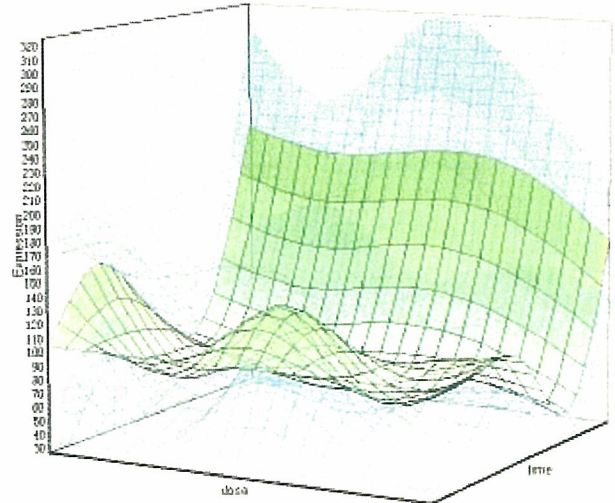
- [-] A2A_ByUnigene@
 - [+] MOE430v2
 - [+] 1427347_s_at
 - [+] AgilentID
 - [+] A_51_P103659
 - [+] A_51_P490023
 - [+] A_52_P621603
 - [+] AgilentNIHS44K
 - [+] 17388
 - [+] 25843
 - [+] 32223

Tubb2 /// 2410129E14Rik /// LOC544939

Affymetrix MOE430v2 1-117828_at

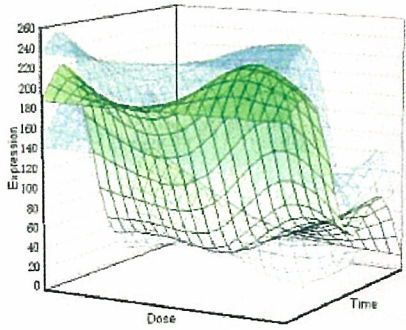


Agilent PercellomeArray 38234

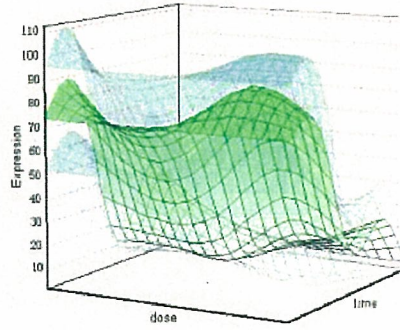


- [-] A2A_ByUnigene@
 - [+] MOE430v2
 - [+] 1417828_at
 - [+] AgilentID
 - [+] A_51_P183261
 - [+] AgilentNIHS44K
 - [+] 38234

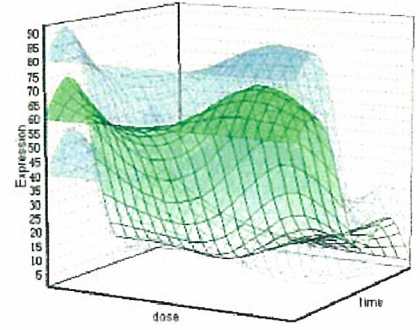
Affymetrix MOE430v2 1426516_a_at



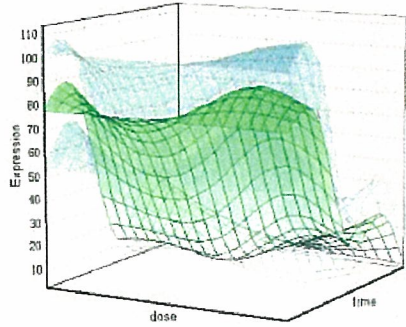
Agilent PercellomeArray 18916



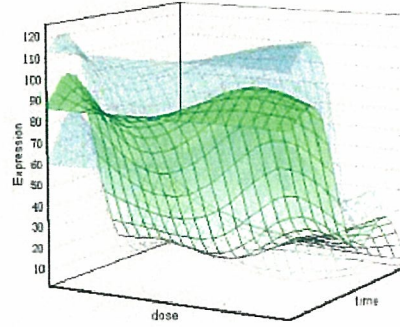
Agilent PercellomeArray 28341



Agilent PercellomeArray 6632



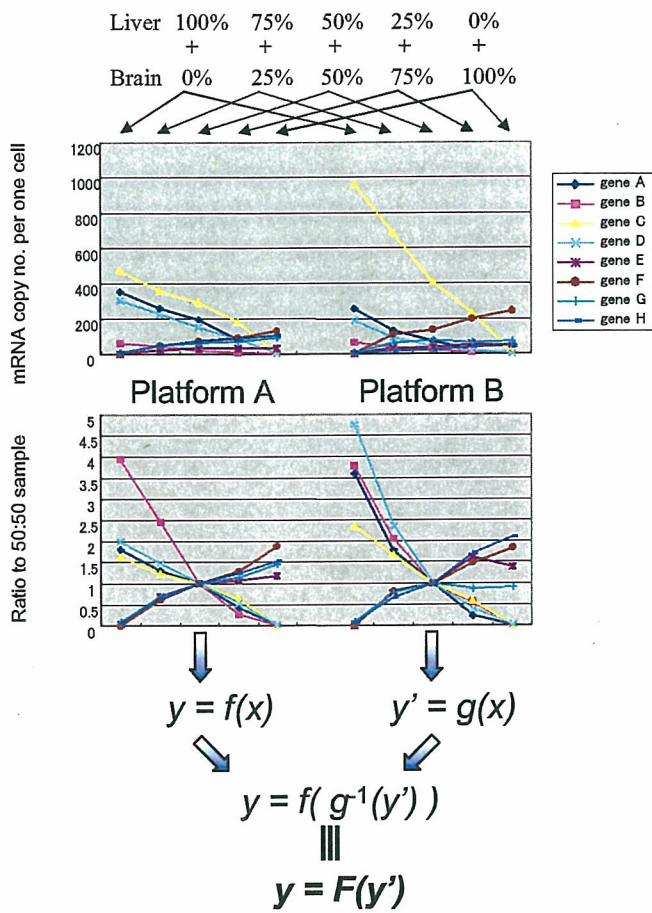
Agilent PercellomeArray 7001



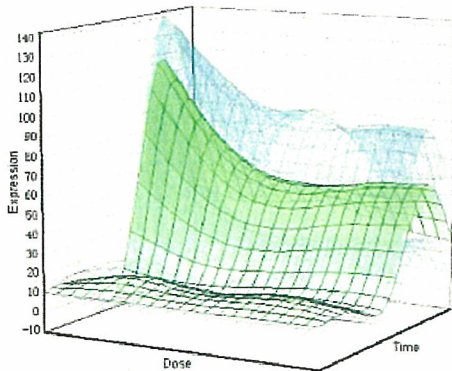
- [-] A2A_ByUnigene@
 - [+] MOE430v2
 - ▷ 1426516_a_at
 - [+] AgilentID
 - ▷ A_51_P394984
 - ▷ A_52_P483101
 - ▷ A_52_P483104
 - ▷ A_52_P483105
 - [+] AgilentNIHS44K
 - ▷ 18916
 - ▷ 28341
 - ▷ 6632
 - ▷ 7001

Lpin1

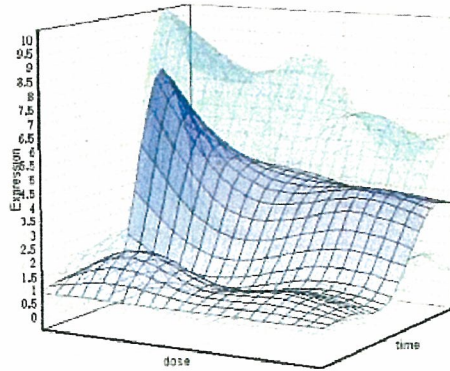
発現値の補正



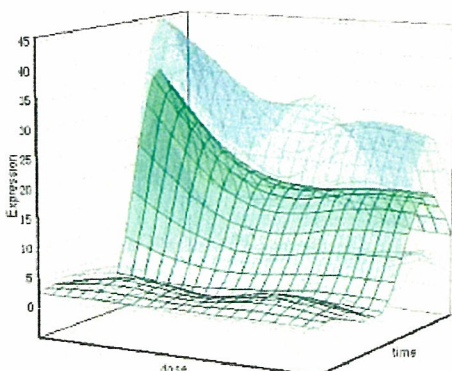
Affymetrix MOE430v2 1438743_at



Agilent PercellomeArray 20078



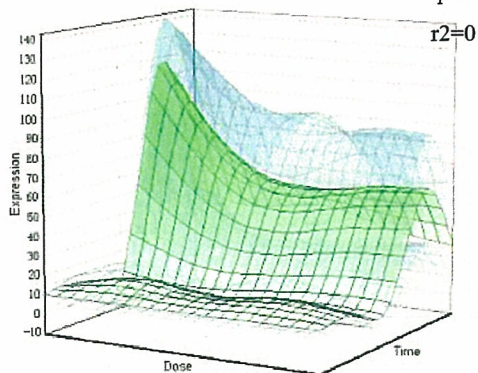
Agilent PercellomeArray 40756



- [-] A2A_ByUnigene@
 - [+] MOE430v2
 - [+] 1438743_at
 - [+] AgilentID
 - [+] A_51_P290981
 - [+] A_52_P84027
 - [+] AgilentNIHS44K
 - [+] 20078
 - [+] 40756

Affymetrix MOE430v2 1438743_at

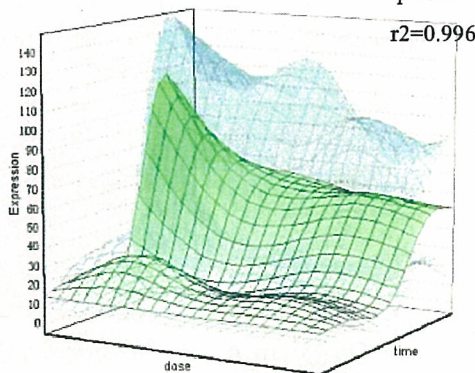
p=1
q=0



r2=0

Agilent PercellomeArray 20078

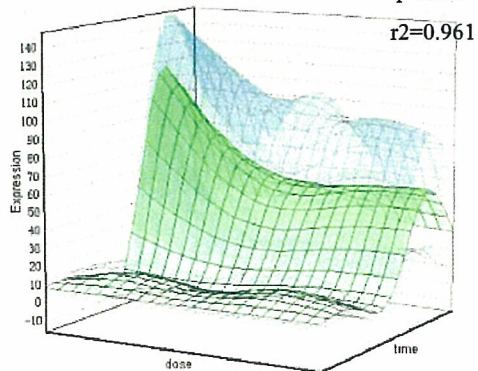
p=14.34
q=0.532



r2=0.996

Agilent PercellomeArray 40756

p=3.228
q=0.181

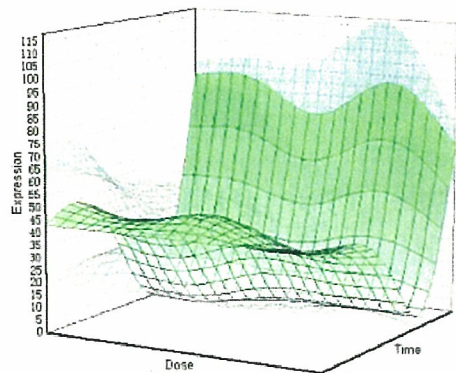


r2=0.961

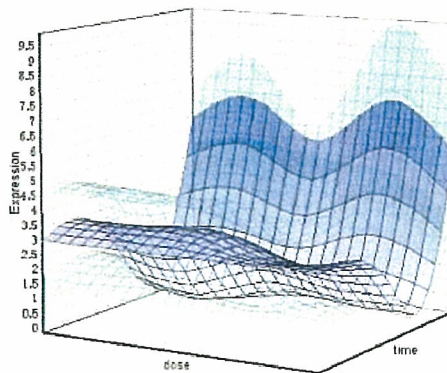
Cyp7a1

- [-] A2A_ByUnigene@
 - [-] MOE430v2
 - 1438743_at
 - [-] AgilentID
 - A_51_P290981
 - ▸ A_52_P84027
 - [-] AgilentNIHS44K
 - 20078
 - 40756

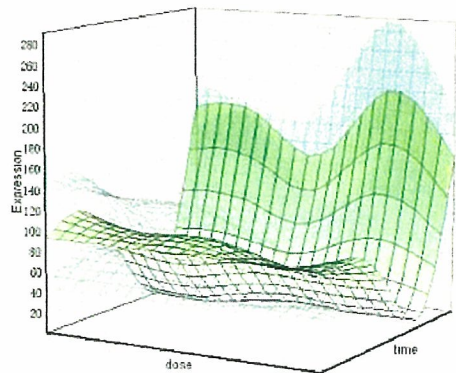
Affymetrix MOE430v2 1427347_s_at



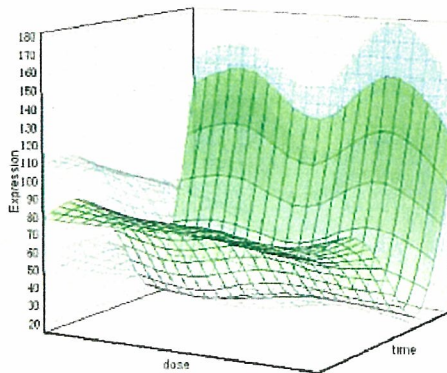
Agilent PercellomeArray 17388



Agilent PercellomeArray 25843



Agilent PercellomeArray 32223



- [-] A2A_ByUnigene@
 - [-] MOE430v2
 - 1427347_s_at
 - [-] AgilentID
 - A_51_P103659
 - A_51_P490023
 - A_52_P621603
 - [-] AgilentNIHS44K
 - 17388
 - 25843
 - 32223