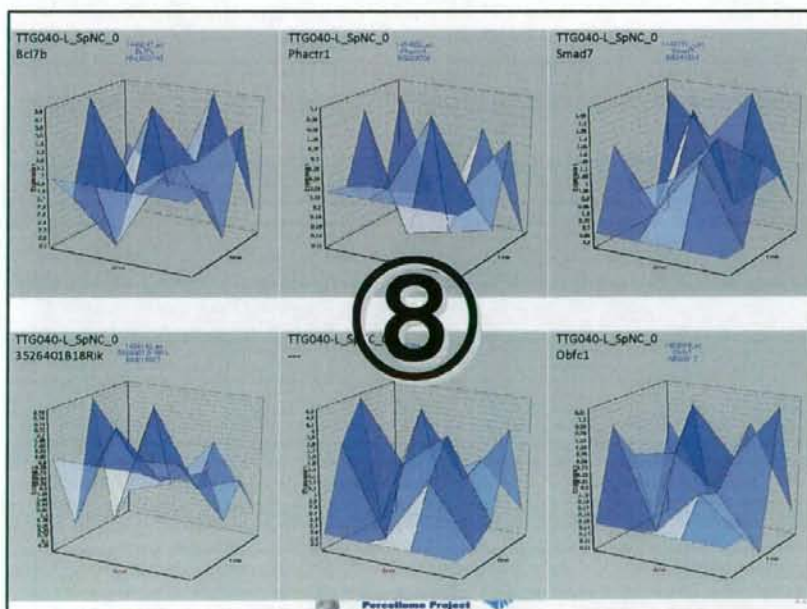


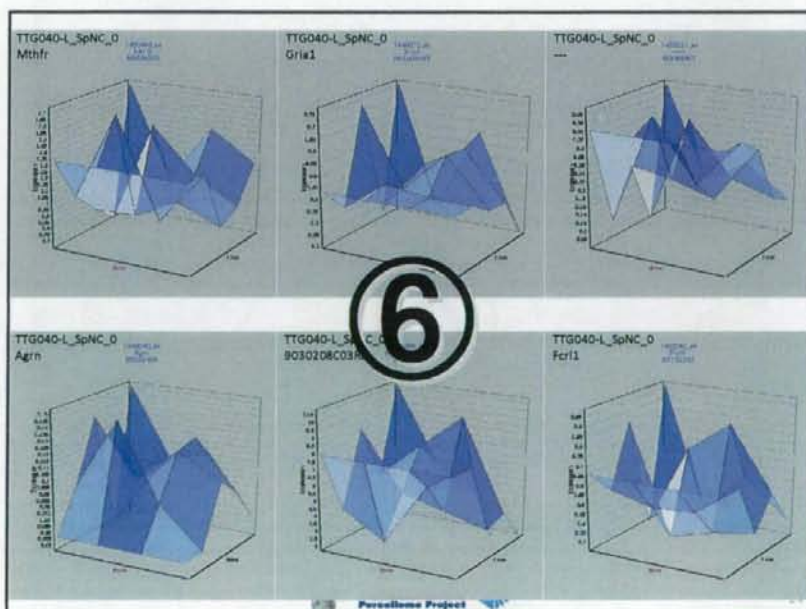
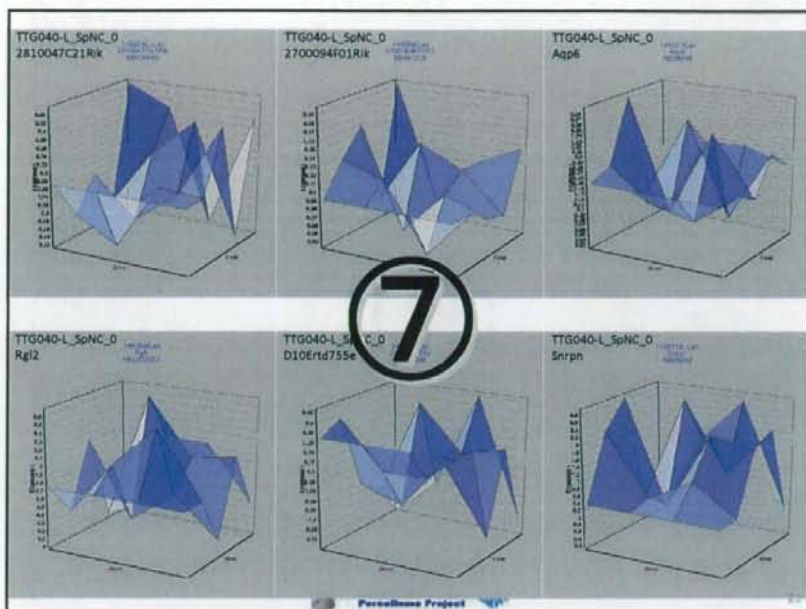
## RSort

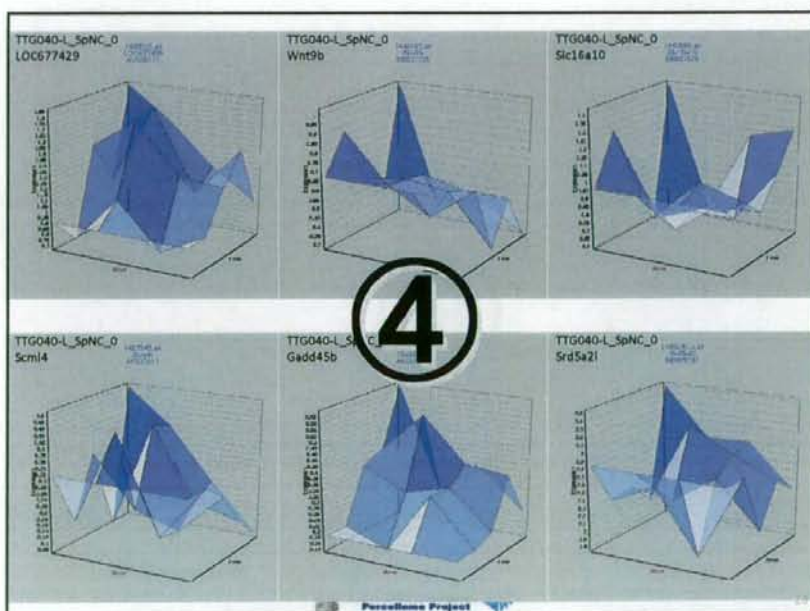
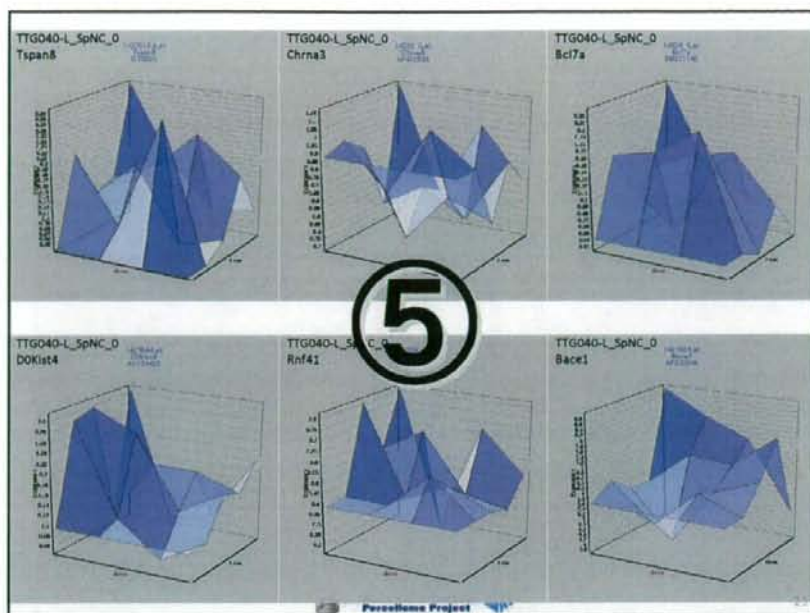


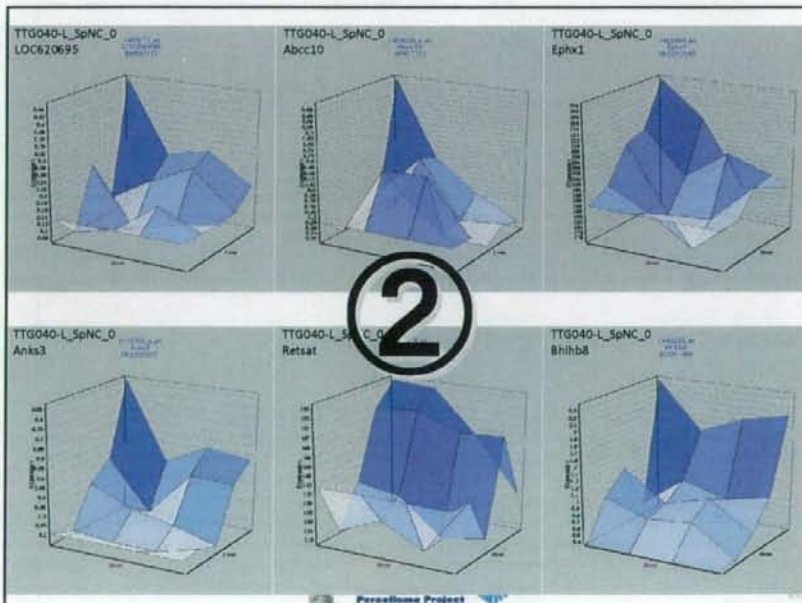
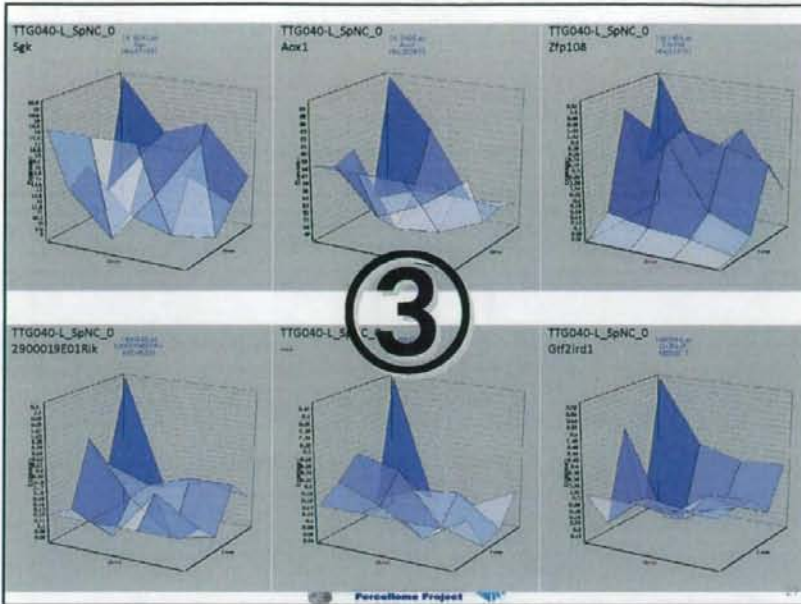
- 遺伝子発現を3次元曲面として表示した「Surface」グラフを元に、その凸凹の程度(Roughness)とピークの位置などを評価して、候補遺伝子リストの並べ替えを行うアルゴリズム

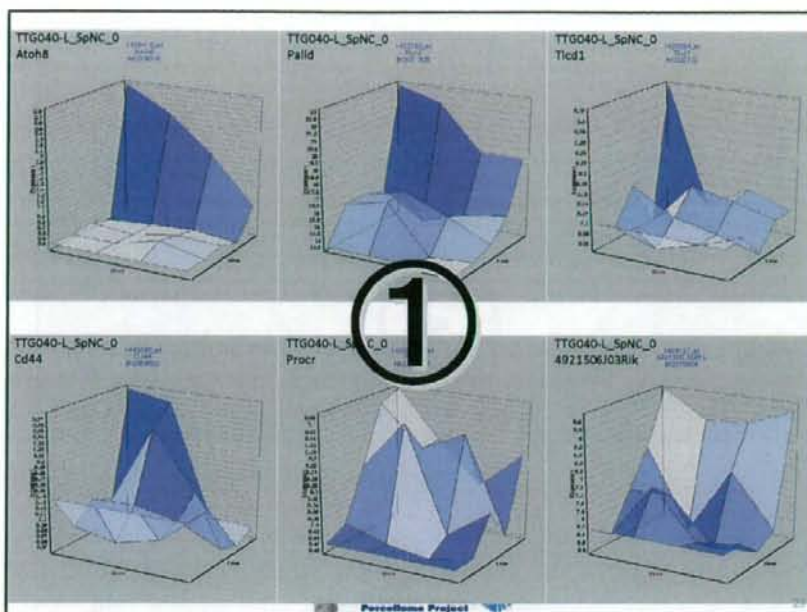
- 誘導抑制の検出に強い。
- 計算処理が軽い。











## RSortの改良点(1)

- 外れ値の処理
- dose軸方向では、隣り合う点との連続性が重視される
- time軸方向では、連続性だけでなく独立性も評価される
- Vehicle Effectの処理



Persephone Project

## RSortの改良点(1)

- 外れ値の処理  
>>>> t-test等、統計処理との併用
- dose軸方向では、隣り合う点との連続性が重視される  
>>>> Trend評価を利用
- time軸方向では、連続性だけでなく独立性も評価される  
>>>> 最大ピークだけでなく、第2、第3ピークの抽出オプションを追加
- Vehicle Effectの処理  
>>>> 再現性を評価した上で、Vehicleデータをsubtractionする

## RSortの改良点(2)

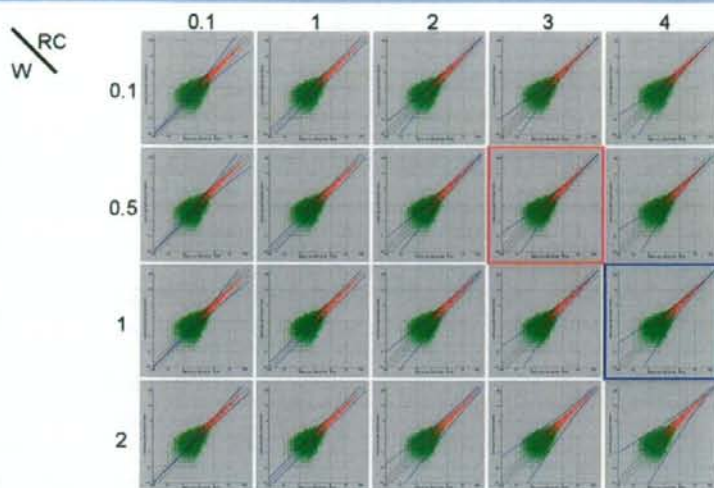
ピーク数3個以下、t検定で $p < 0.1$ のみ  
>>>> 4用量x4時点のデータにおいて、  
約4000~8000個の候補遺伝子を自動抽出する



うち半数以上が研究者の手作業により候補から外される

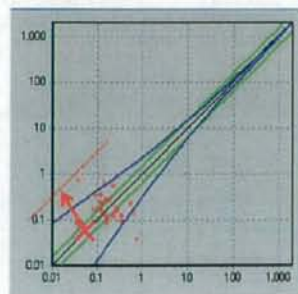
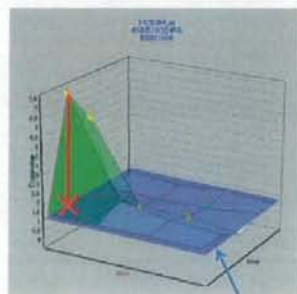
- 再現性は比較的良いが、変動量が系統誤差以下の候補遺伝子を除外  
>>>> Fx Curveの併用

## Fx曲線

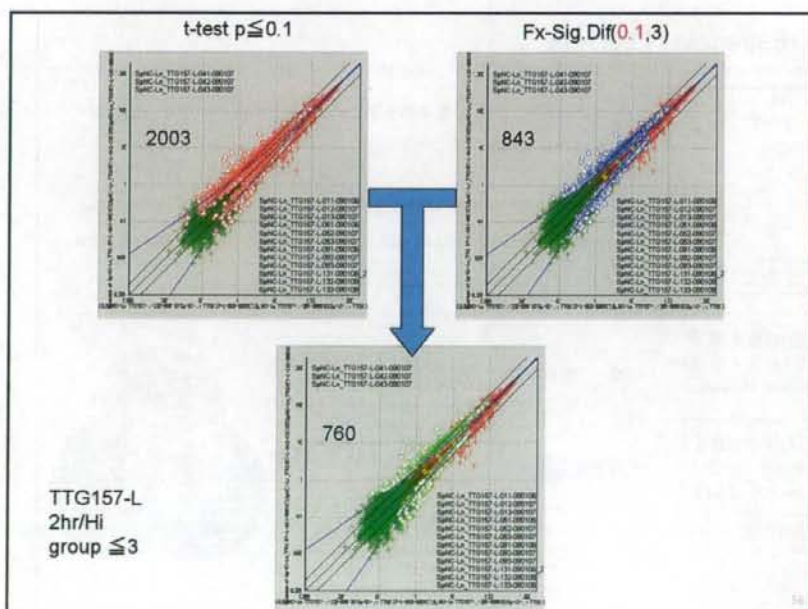
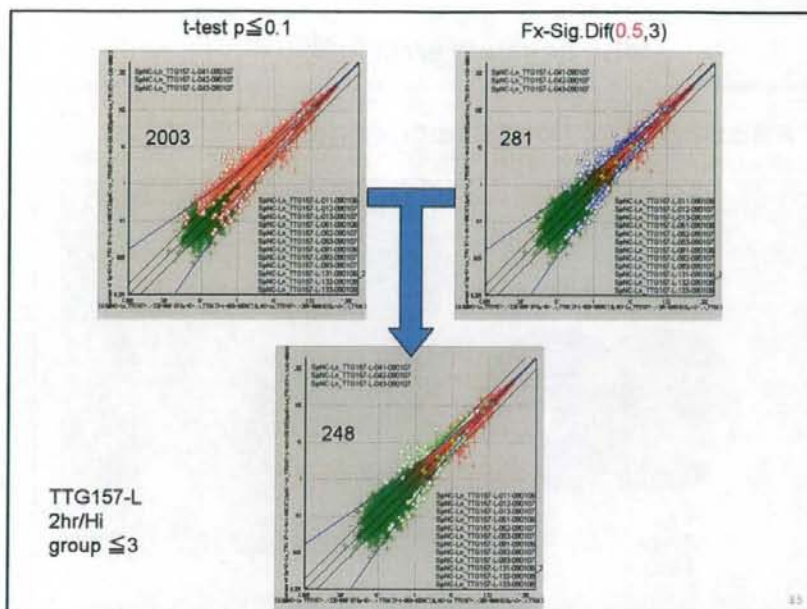


## RSortへの適用

最大変動量がFx曲線の外側であれば、有意と判定する



基準平面(通常、溶媒対照群の平均値)



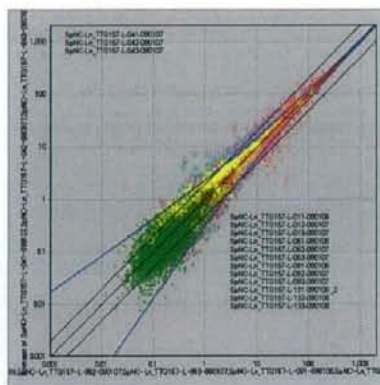


## false negative errorを減らす

再現性が極めて良好なもの(t-test  $p < 0.01$ )を拾い直す

t-test( $p < 0.1$ ) and Fx-Sig.Dif(0.5,3)  
 >>>> 248 probe-set

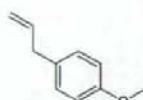
t-test( $p < 0.01$ )  
 >>>> 649 probe-set



## 例) Estragole (TTG147)

TTG147  
Estragole

CAS 149-07-0  
MW=148.20



vehicle(corn oil)  
dosage(0, 10, 30, 100mg/kg)  
time(2, 4, 8, 24hr)

・香料イランイランや香辛料タラゴン(エストラゴン)やバジル、八角(ダイウイキョウの実)に含まれる。

・mouseのTD50は51.8mg/kg/day[Carcinogenic Potency Database].  
HepG2ではDNA adduct形成が報告されている[Environ Mol Mutagen 38, 715, 2007]

### RSort基本設定

time: 2, 4, 8, 24hr } にピークがあるもの  
dose: Hi, Mid }

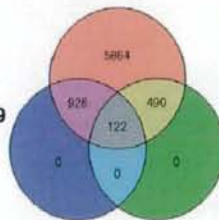
Max group: 3  
t-test:  $p < 0.1$   
>>>> 7402

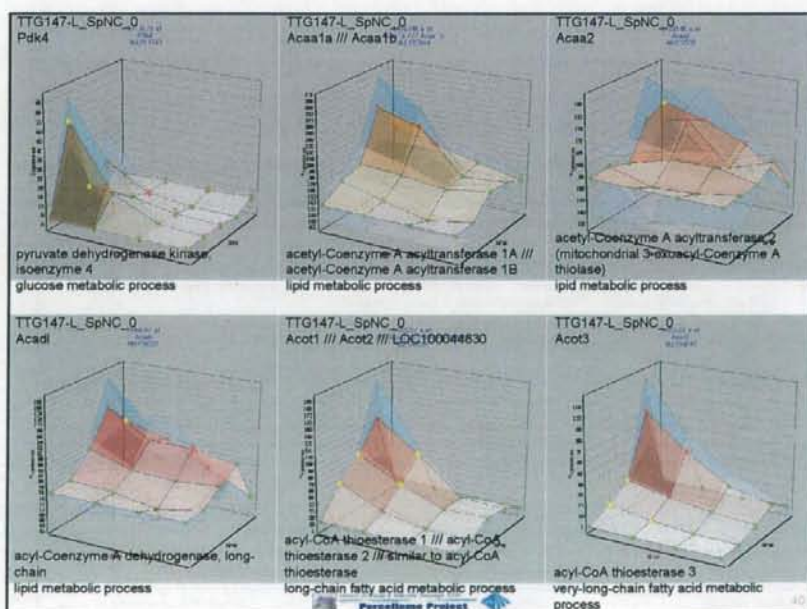
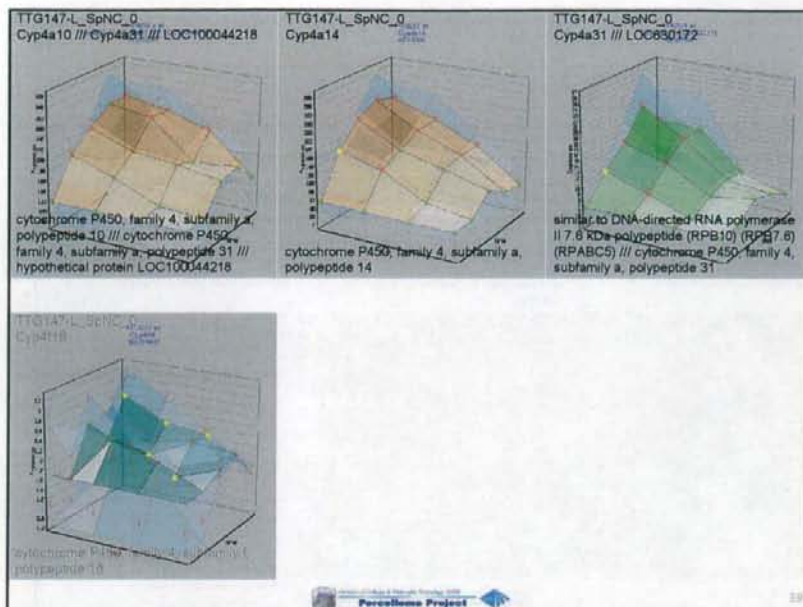


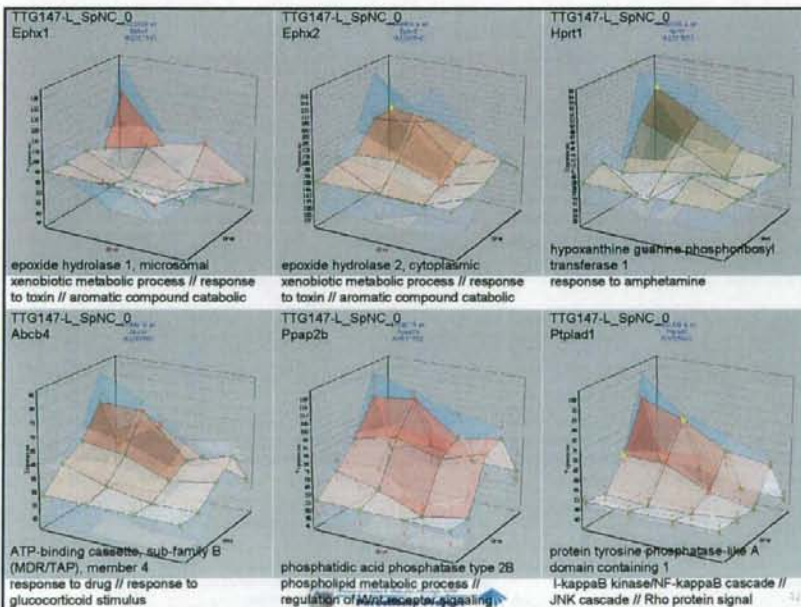
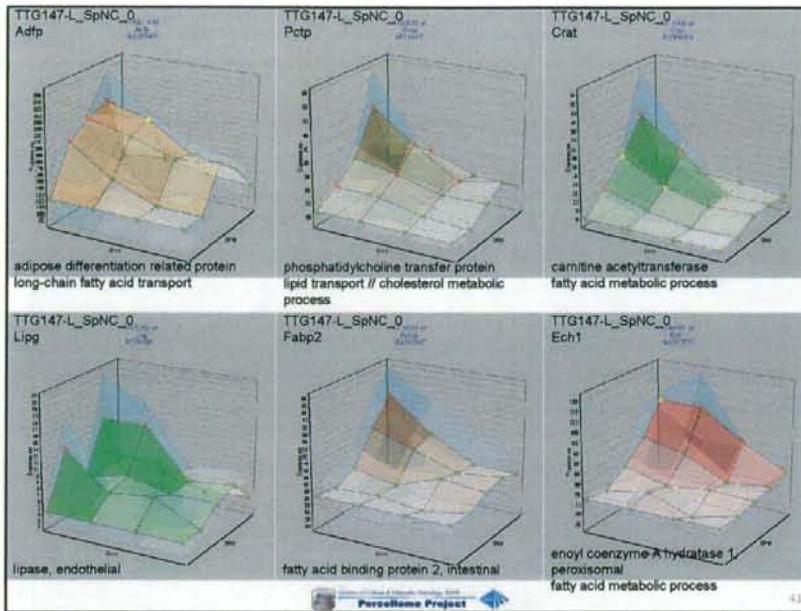
Max group: 3  
t-test:  $p < 0.01$   
>>>> 1048

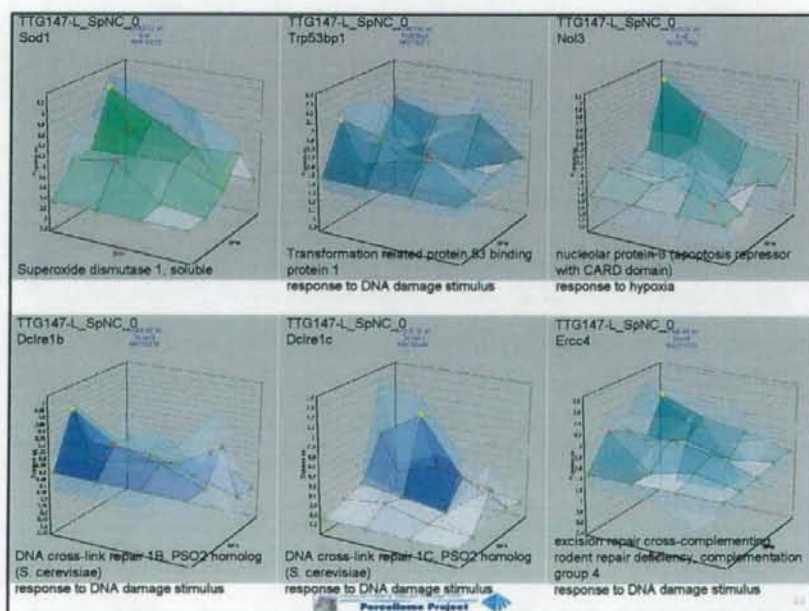
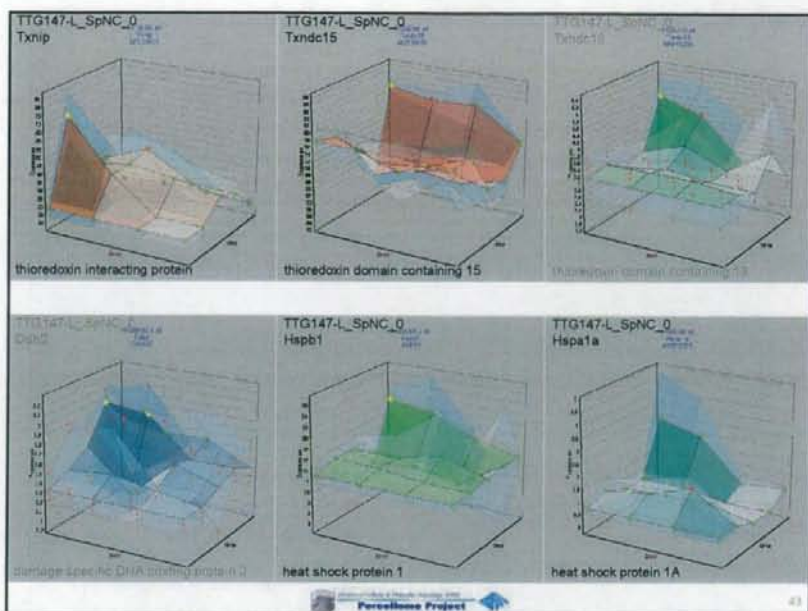
+ >>>> 1539

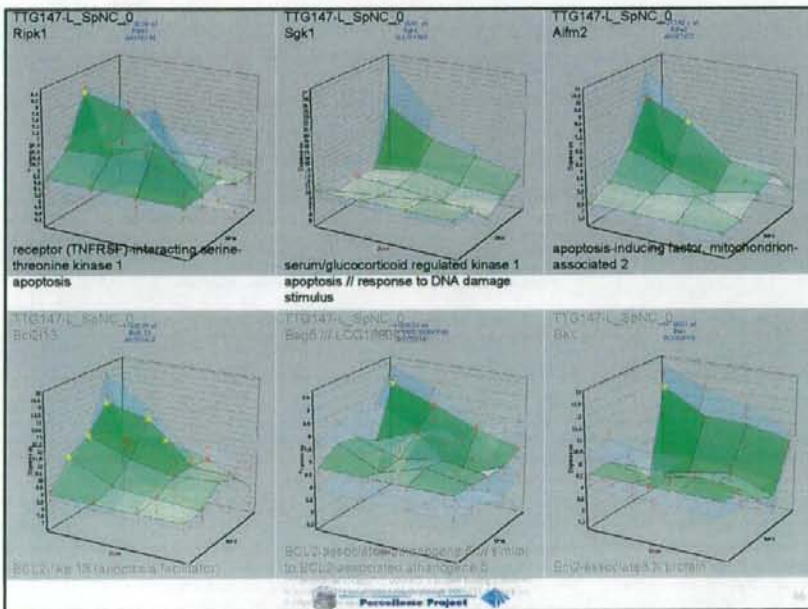
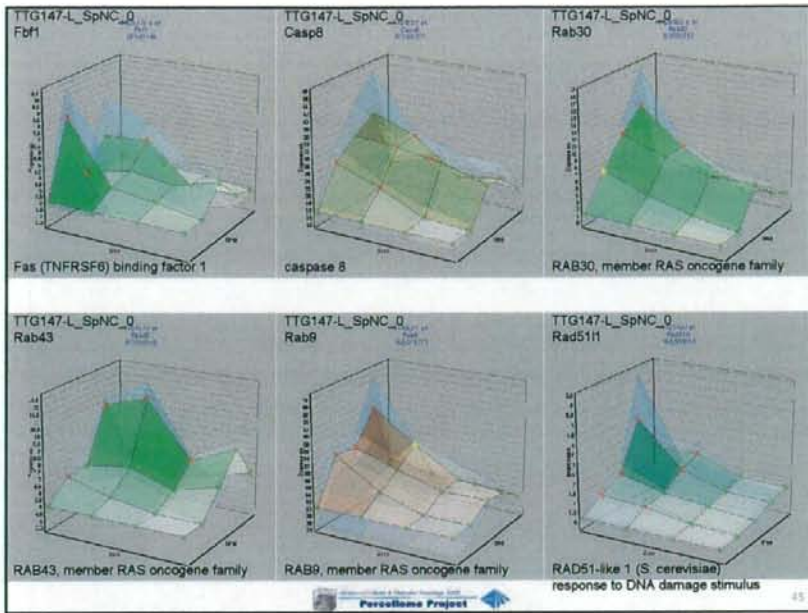
Max Group: 3  
t-test:  $p < 0.1$   
Fx-Sig.Dif(0.5,3)  
>>>> 612





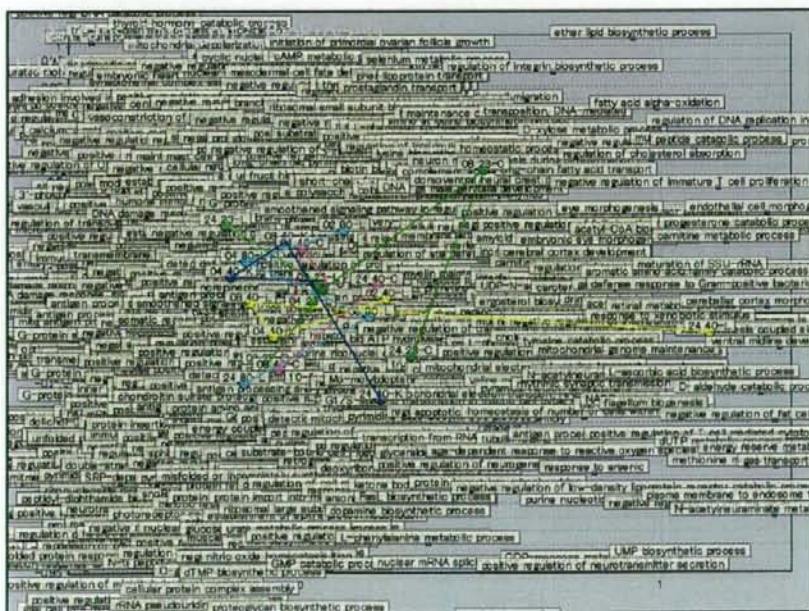
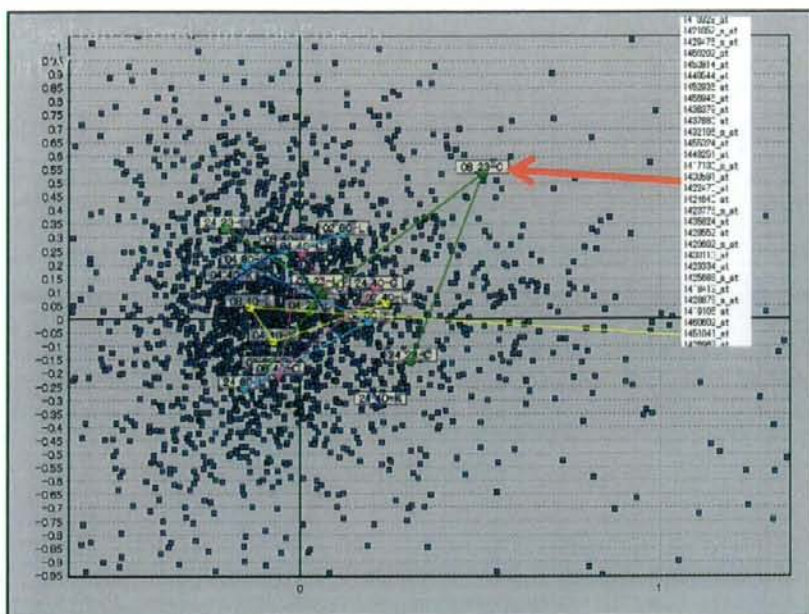




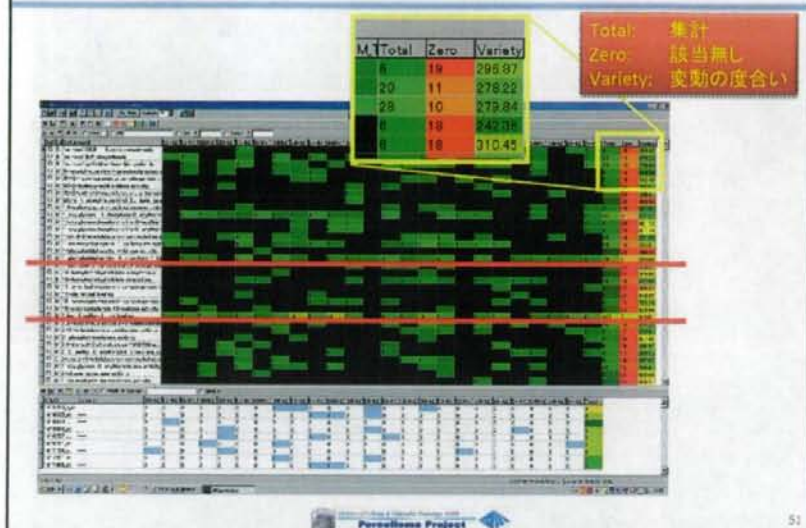


## 遺伝子クラスタの意味抽出

## GoPlot



## 候補遺伝子リスト群のGO解析 (MFGoViewer)

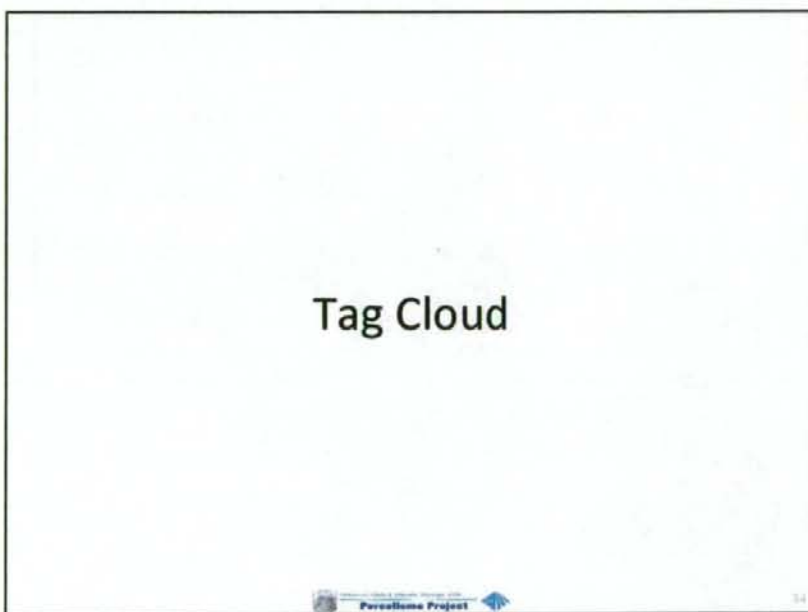
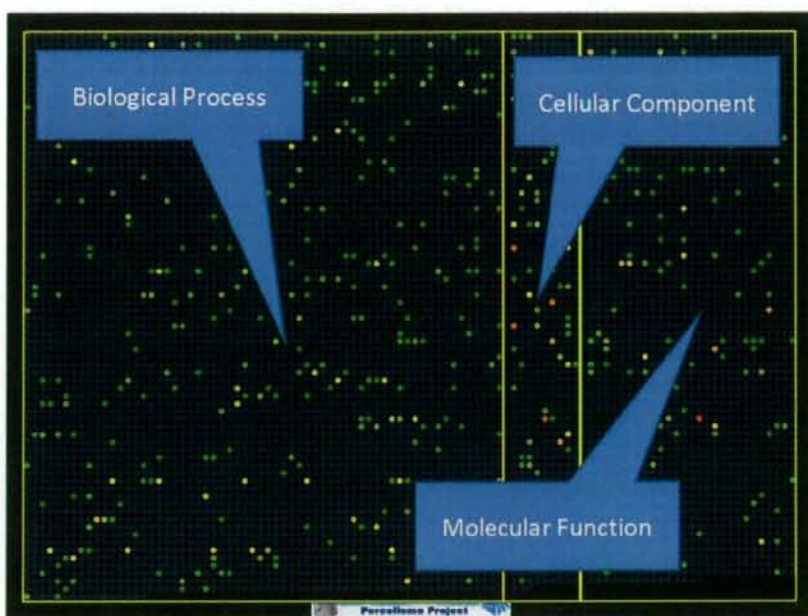


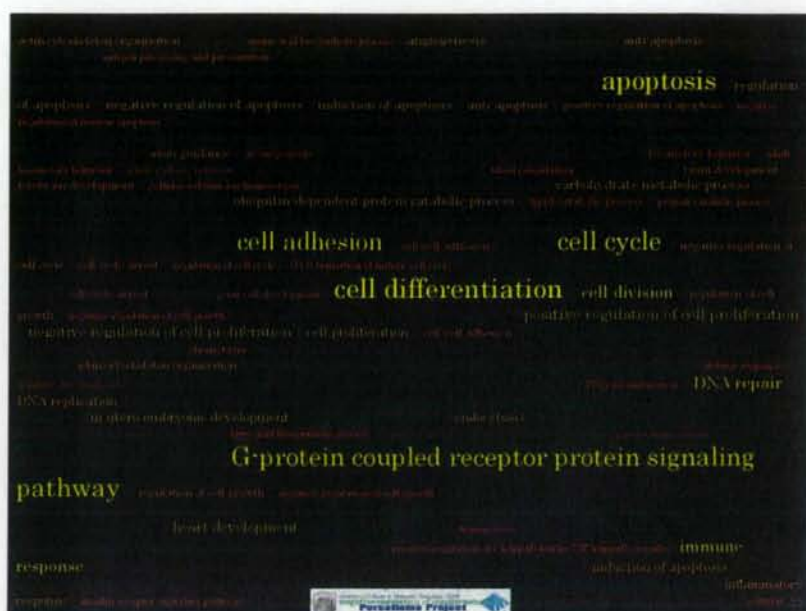
51

GoSheet


52



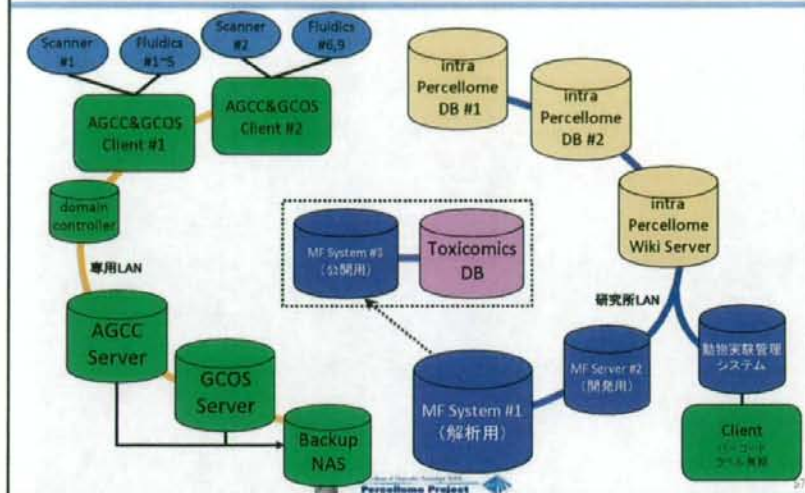




Percellomeデータベースの最適化と  
ナレッジデータ強化



## Percellome Projectの データ解析・公開システム (MF System) の構成



## MF Serverハードウェア



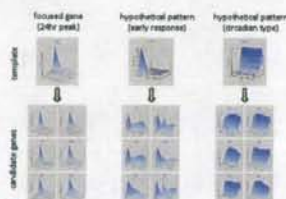
## 遺伝子発現変動パターンを全て網羅する方策

- RSort
  - 発現パターンの単純さ(複雑さ)・特徴点の位置(実験条件)を評価し、候補遺伝子リストを並べ替える
  - 解釈が容易な(単純な)発現パターンを呈する候補遺伝子を抽出する



### tmf Clustering = tmf + 教師なしクラスタリング(MADIC)

- tmf: 2つの発現パターンを比較し、類似度を算出するアルゴリズム
- 想定外の発現パターンも含め、代表的なパターンを網羅的に自動抽出する
- 正立に加え倒立での評価計算を加え、網羅性を高めた



## Tmf Clusteringによって抽出された 代表的な発現変動パターン

