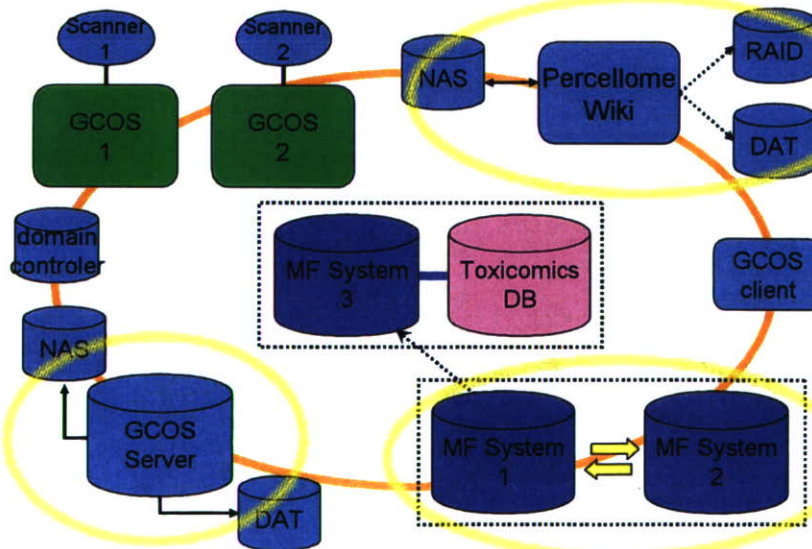


進捗状況

- アルゴリズム
 - 新規開発:
 - 飽和対策
 - 改良:
 - RSort
 - GoPlot
- クライアントツール (MFtool)
 - 新規開発:
 - FxFilter
 - MFGoViewer
 - CreateDfSurface
 - 改良
 - MFHcMap
 - MFSurface, MSV2
 - RSort
- MFサーバーシステム
 - 専用Teradataサーバー導入
 - 処理能力強化のため
- 公開DB: 稼働中
 - ユーザー数 = 延べ82名
 - データ不足・機能不足の解消が課題 (機能については予算獲得次第で拡張可能)。
- PercellomeWiki
 - データの一元管理
 - 解析結果の共有も視野

MF Serverのシステム構成





Wikiサーバーの構築

- DualCore Xeon 2.66GHz
- RAM 4GB
- HDD 1.5TB

- Windows 2003 Server
- Apache2
- PHP5
- PukiWiki1.4.7



PercellomeWiki

The screenshot shows a web browser displaying the PercellomeWiki interface. The page lists data files for three cell lines: Ljiver, Kidney, and Testis. Each cell line has a list of files with their respective formats and download links. The browser's address bar shows the URL: http://10.12.1.158/percellomewiki/index.php/TTG099-4. The page content includes a sidebar with navigation links and a main content area with file lists. At the bottom of the page, there is a search bar and a 'コメントの挿入' (Insert Comment) button.

Ljiver

- SPF=0.02
- cDNA=50%
- GSC combination=optimized
- QC:http://10.12.1.158/MFDB/TTG099-4_SC4QC.zip
- raw (PrivotAnalysis Format):http://10.12.1.158/MFDB/TTG099-4_raw.zip
- SpNC (Individual Files):http://10.12.1.158/MFDB/TTG099-4_SpNC.zip
- Analyzer Data Set:http://10.12.1.158/MFDB/TTG099-4_SpNC_Analyzer.zip
- Surface Data Set:http://10.12.1.158/MFDB/TTG099-4_SpNC_Surface.zip
 - AvS

Kidney

- SPF=0.002
- cDNA=50%
- GSC combination=optimized
- QC:http://10.12.1.158/MFDB/TTG099-K_QC.zip
- raw (PrivotAnalysis Format):http://10.12.1.158/MFDB/TTG099-K_raw.zip
- SpNC (Individual Files):http://10.12.1.158/MFDB/TTG099-K_SpNC.zip
- Analyzer Data Set:http://10.12.1.158/MFDB/TTG099-K_SpNC_Analyzer.zip
- Surface Data Set:http://10.12.1.158/MFDB/TTG099-K_SpNC_Surface.zip
 - AvS

Testis

- SPF=0.015
- cDNA=50%
- GSC combination=optimized
- QC:http://10.12.1.158/MFDB/TTG099-Y_SC4QC.zip
- raw (PrivotAnalysis Format):http://10.12.1.158/MFDB/TTG099-Y_raw.zip
- SpNC (Individual Files):http://10.12.1.158/MFDB/TTG099-Y_SpNC.zip
- Analyzer Data Set:http://10.12.1.158/MFDB/TTG099-Y_SpNC_Analyzer.zip
- Surface Data Set:http://10.12.1.158/MFDB/TTG099-Y_SpNC_Surface.zip
 - AvS

Comments

お名前: コメントの挿入

Probe飽和の補正

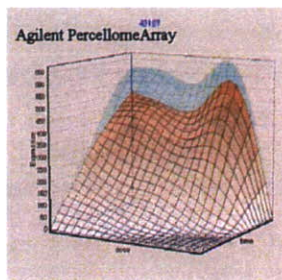
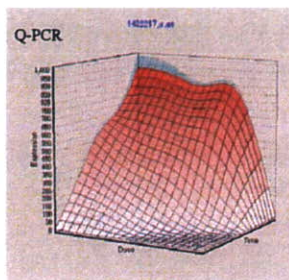
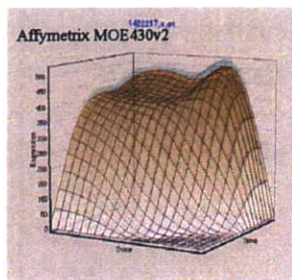
データ解析の妨げになっている原因の
1つ、飽和現象の補正を試みた

Probe飽和の例

Affy

Q-PCR

Agilent



AZA_ByUriagene®
 ↳ MOE430v2
 ↳ 1422217_a_at
 ↳ AgilentID
 ↳ A_51_P279693
 ↳ AgilentNHIS44K
 ↳ 43187

Cyplal



7

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



8

AIC(Akaike Information Criteria)

- すべてのプローブが飽和しているとは考えられない。また、飽和していたとしても、一定の値を示しているため、飽和しているかを判別できないことが考えられる。
>>>> Langmuirモデル、線形モデル、定数モデルを選択するために使用する
- 元統計数理研究所長の赤池弘次によって1971年に考案された統計モデルの良さを示す指標
 - Akaike, Hirotugu (1974). "A new look at the statistical model identification". IEEE Transactions on Automatic Control 19 (6): 716-723.

$$AIC = -2 \log(L) + 2k$$

L: 尤度

k: パラメータ数

$$\text{確率密度関数: } P_{\mu, \sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\text{尤度関数: } L(\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$L_{\mu}(\mu, \sigma) = \prod_{i=1}^n L_i(\mu, \sigma) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$\log L_{\mu}(\mu, \sigma) = \log \left(\prod_{i=1}^n L_i(\mu, \sigma) \right) = \log \left(\prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \right)$$

$$= \sum_{i=1}^n \log \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \right)$$

$$= \sum_{i=1}^n \log \left(\frac{1}{\sqrt{2\pi}\sigma} \right) + \sum_{i=1}^n \log \left(e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \right)$$

$$= -n \log \sqrt{2\pi}\sigma - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

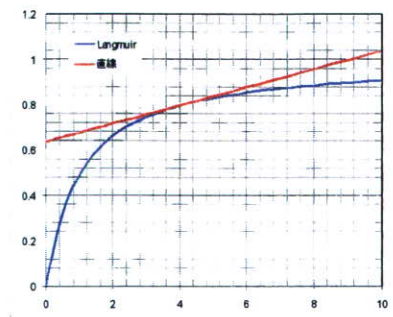
数値計算上の技巧: Langmuir方程式と直線を組み合わせる

- 数値計算を行う上で、Langmuirの方程式は上に有界な数式である。次の2点で数値計算上問題がある。
 - 濃度に対する係数に鈍感な関数で、完全飽和に近い状態で適切な収束が得られない。
 - 飽和していない場合には、AICにより、排除される
 - 今回は逆関数にするので、逆関数の際に右に有界で定義域が制限される。
- 高濃度において、直線にフィット*するものとみなす。
 - この領域に入った場合には詳細な検討が必要である。

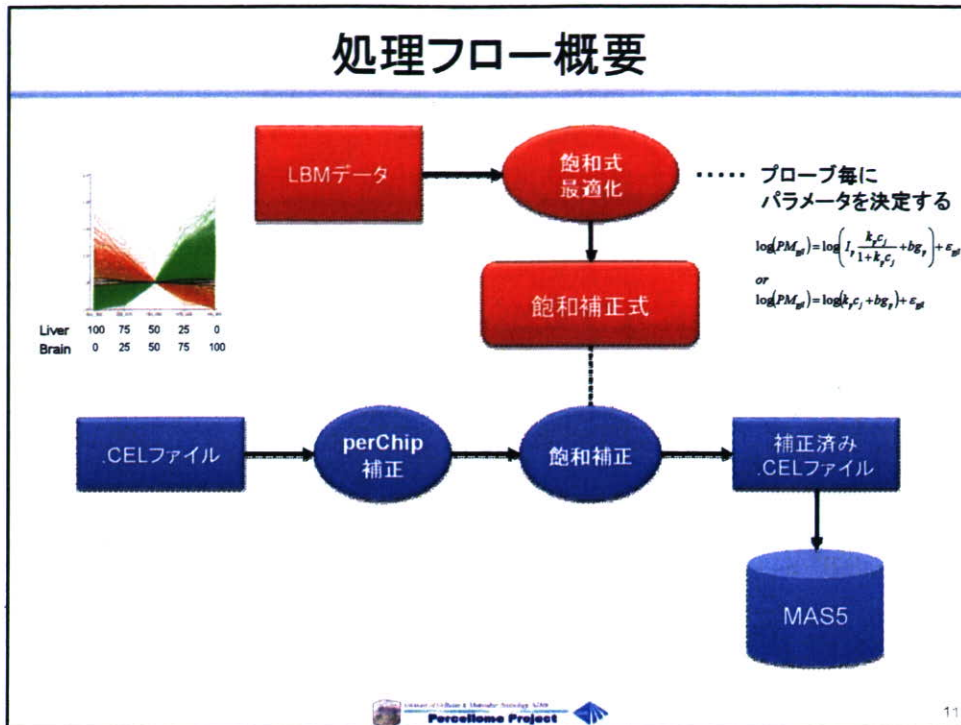
* 高濃度域での放物線フィットも試行したが、直線フィットの方が安定性が高かった

$$f(x) = \begin{cases} \frac{x}{x+1} & \text{for } x \leq 4 \\ 0.04x + 0.64 & \text{for } x \geq 4 \end{cases}$$

$$I_H = \begin{cases} I_p \frac{k_p c_j}{k_p c_j + 1} + b g_p & \text{for } k_p c_j \leq 4 \\ I_p (0.04 k_p c_j + 0.64) + b g_p & \text{for } k_p c_j \geq 4 \end{cases}$$



処理フロー概要






飽和補正式の作成手順

1. LBMデータをデータベースへ読み込み
2. チップごとの標準化を実施
3. データベース内で振り分け用単純モデル作成
4. 振り分け情報に基づき、データ出力及び非線形最適化のR用スクリプトの作成
5. Rによる非線形最適化実施し、飽和方程式の決定

RSort

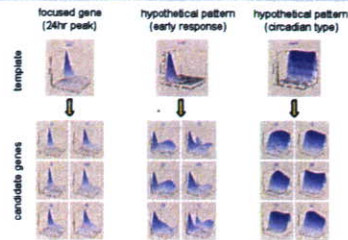
簡便・効率的な候補遺伝子の抽出法

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,  MFCMF,  MFHeMap
特徴パターン抽出	 RSort,  MFHeMap,  MFkFilter
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	 MFAnalyze,  MFGoViewer,  MFGoPlot

現れている変動パターンを網羅するために

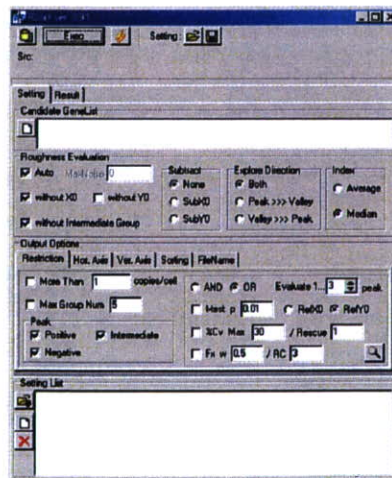
- tmf
 - 予め想定した発現パターンを元に検索
- MADIC
 - 代表的なパターンを抽出する
 - 想定外の発現パターンも網羅的に抽出可能
 - 想定した発現パターンを呈する集団の存在様式の判定 (類似遺伝子が多数あるか、派生パターンの集団はあるか、など)
- RSort
 - 発現パターンの単純さ(複雑さ)・特徴点の位置を評価し、候補遺伝子リストを並べ替える
 - 解釈が容易な(単純な)発現パターンを呈する候補遺伝子を抽出する



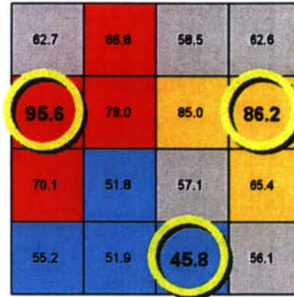
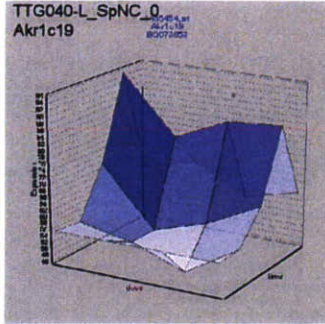
RSort



- 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。



RSortの原理



mean 63.6
SD 13.1
intermediate 57.1~70.2
凹凸判定 = 3

データ

62.7	66.8	58.5	62.6
95.6	78.0	85.0	86.2
70.1	51.8	57.1	65.4
55.2	51.9	45.8	56.1



データ平均

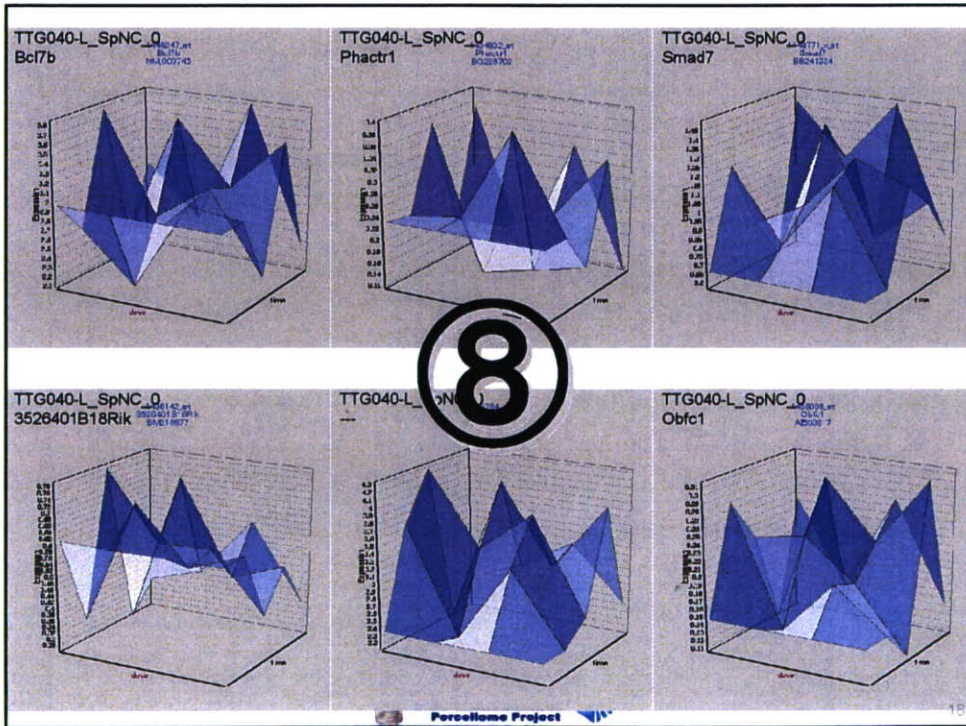
1.0	3.2	5.1	1.1
32.0	14.3	21.4	22.5
6.5	11.9	6.6	1.7
8.4	11.8	17.9	7.6



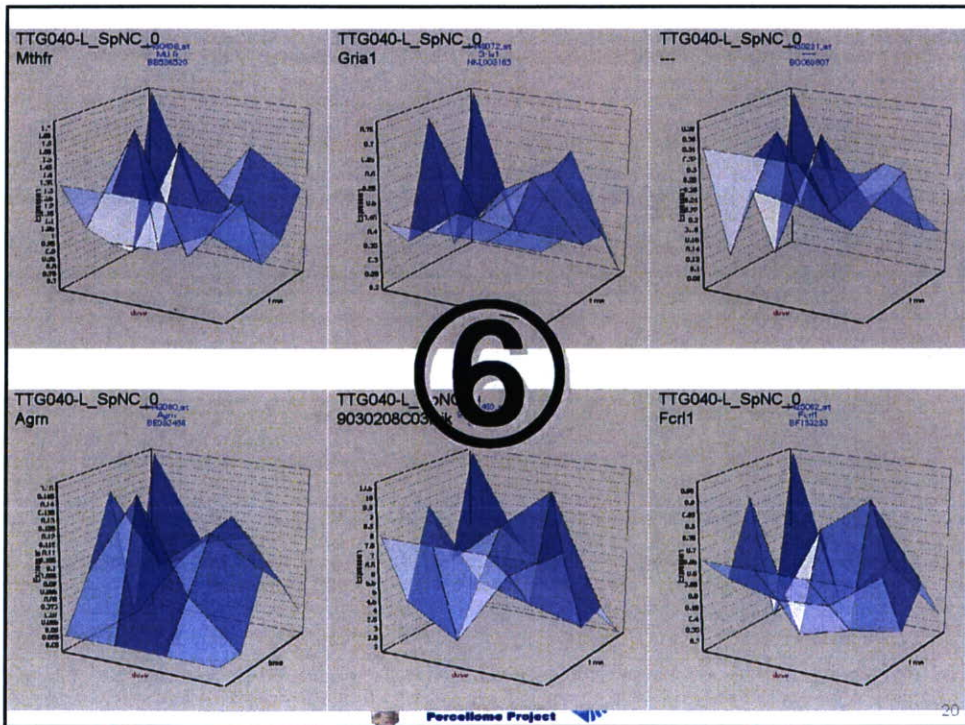
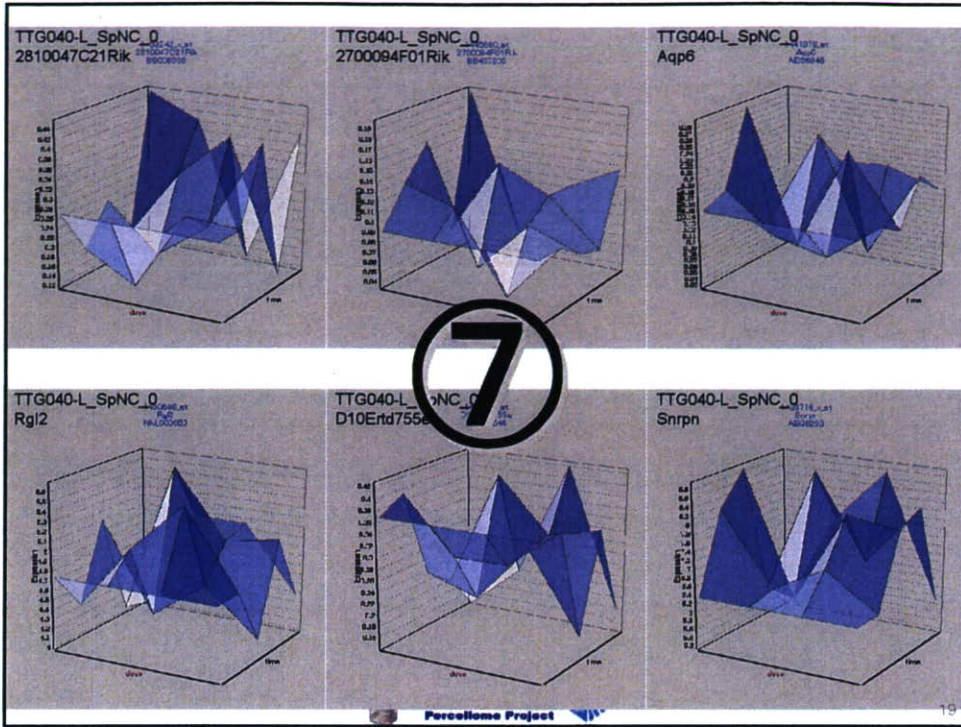
順位

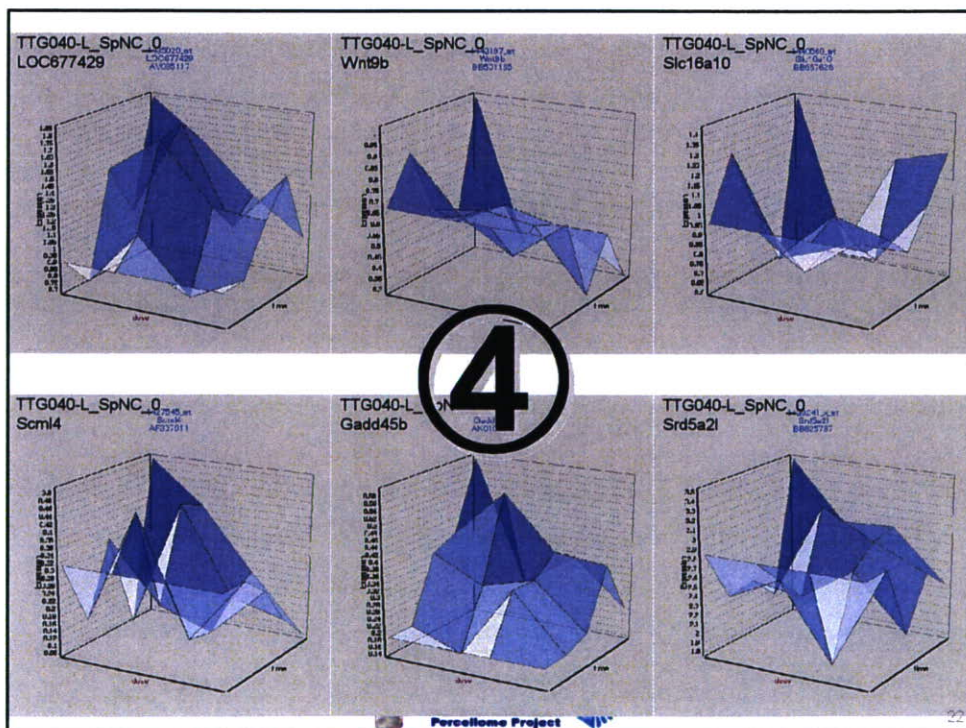
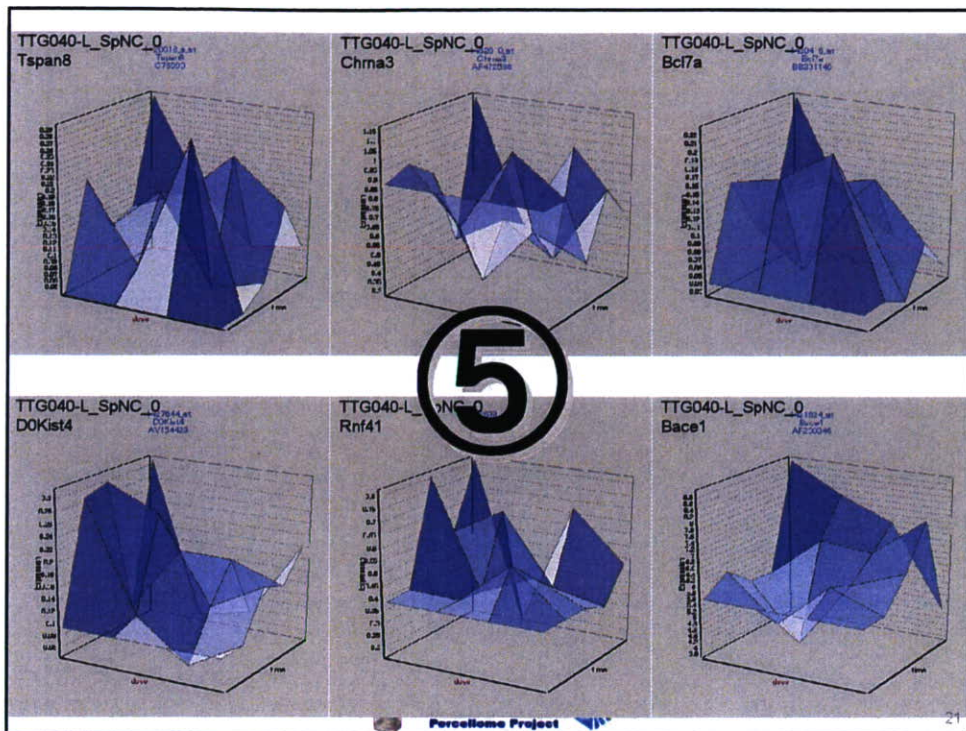
16	15	12	15
1	5	3	2
11	6	10	14
8	7	4	9

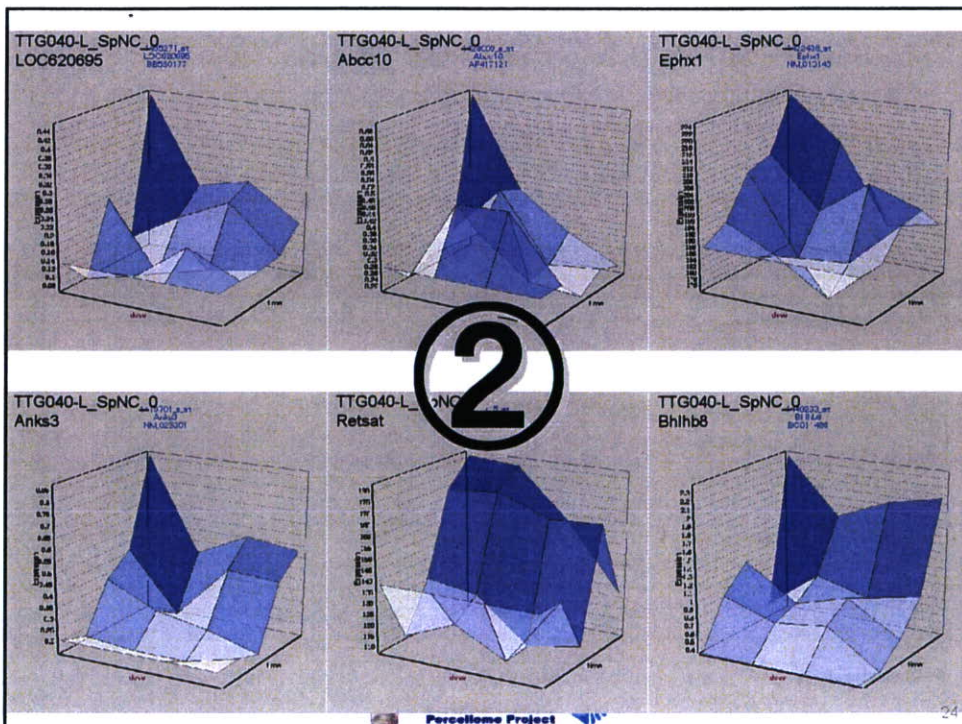
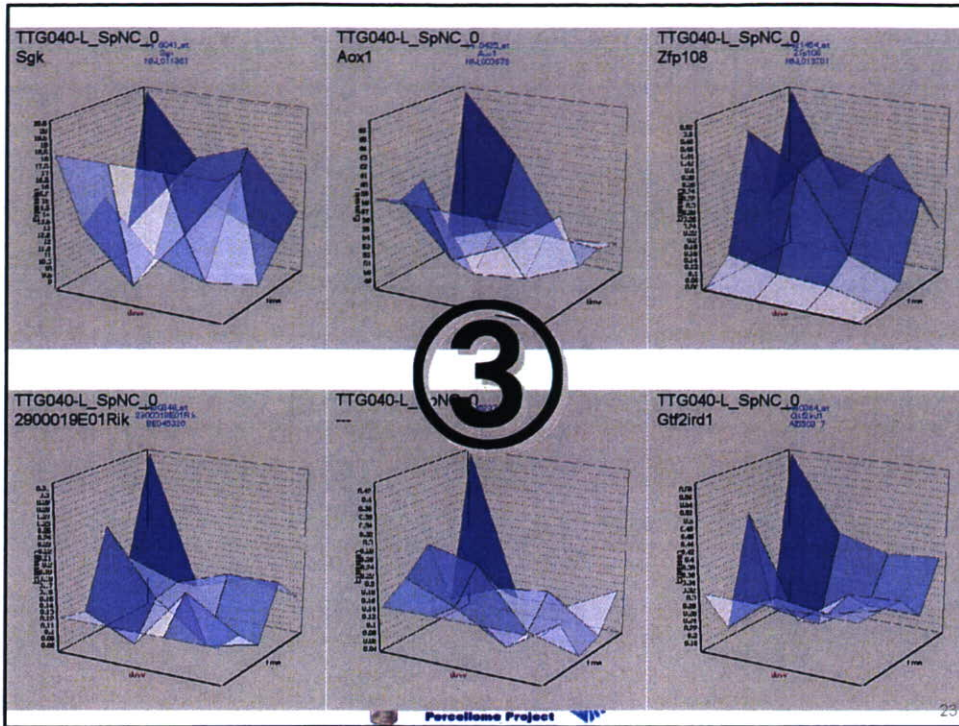
17

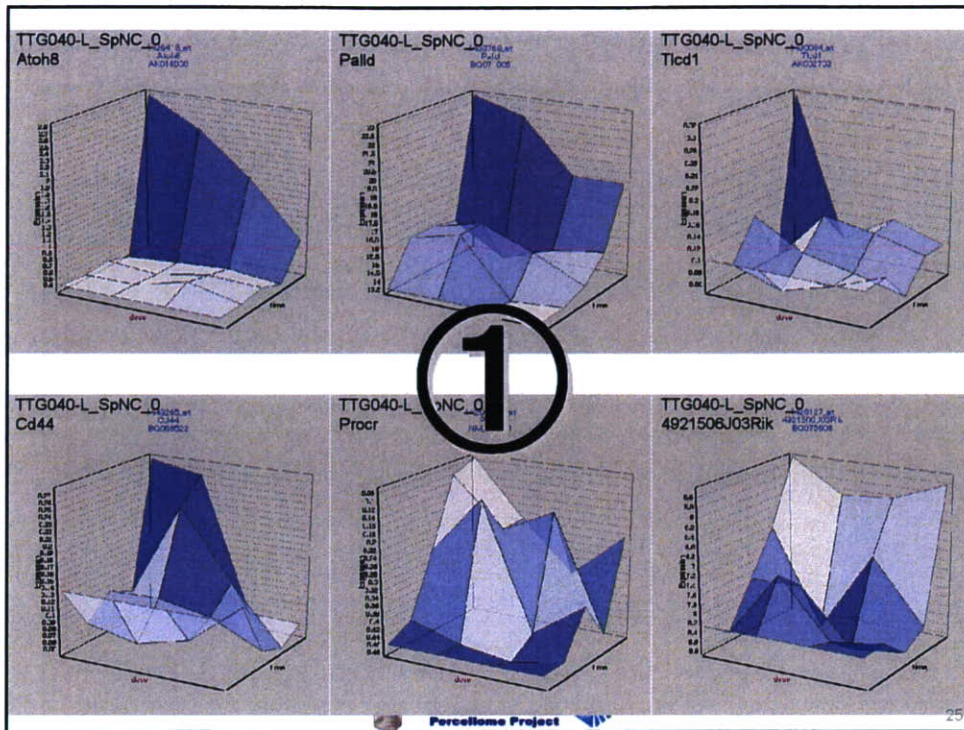


18

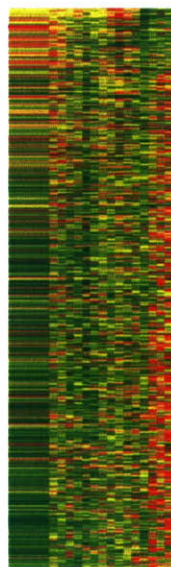




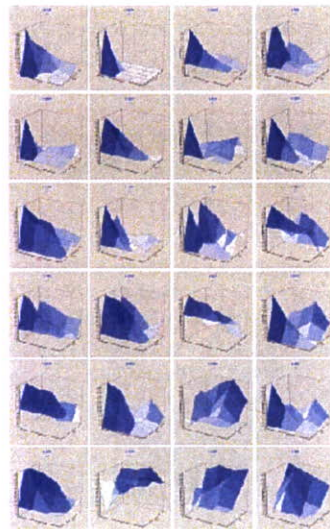
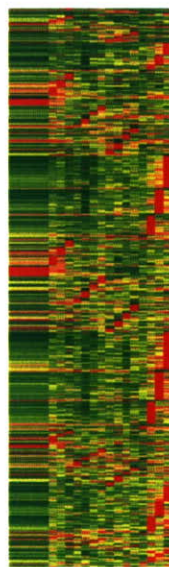




発現パターンの整列



並べ替え



RSortによる候補遺伝子の抽出

・凹凸

少ないもの ⇔ 候補遺伝子
比較的単純な発現誘導(解釈可能)

多いもの
複雑な発現誘導(解釈困難)

or
ノイズ

・最大変動を観測した時間と用量
容量相関、経時変化

The screenshot shows the RSort software interface with several callout boxes:

- 実**: A large character in the top right corner of the main window.
- Fx, Cviによる絞り込み**: A callout pointing to the 'Candidate Genes List' table, which lists genes like 1415797_Ddr1, 1415798_Ddr1, 1415799_Wbp11, and 1415800_Gja1.
- 複数設定での自動実行**: A callout pointing to the 'Roughness Evaluation' and 'Output Options' sections.
- dose指定**: A callout pointing to the 'Dose' field in the 'Roughness Evaluation' section.
- 複数Surfaceデータセットでの自動実行**: A callout pointing to the 'Setting List' section, which contains multiple entries for different surface data sets.
- time指定**: A callout pointing to the 'Time' field in the 'Roughness Evaluation' section.

The interface includes various control panels for 'Roughness Evaluation', 'Output Options', and 'Setting List', along with a 'Candidate Genes List' table and a 'Threshold Curve' graph.

候補遺伝子群が示す意味

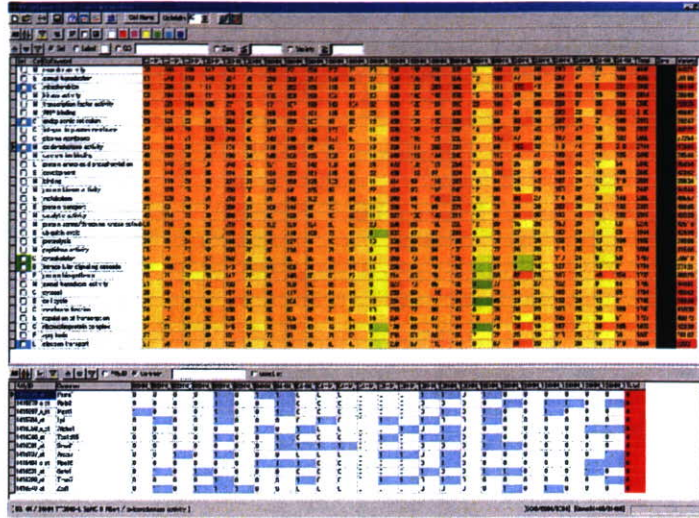
候補遺伝子の
抽出

GOを絞り込み

GoPlotで可視化

候補遺伝子リスト群のGO解析

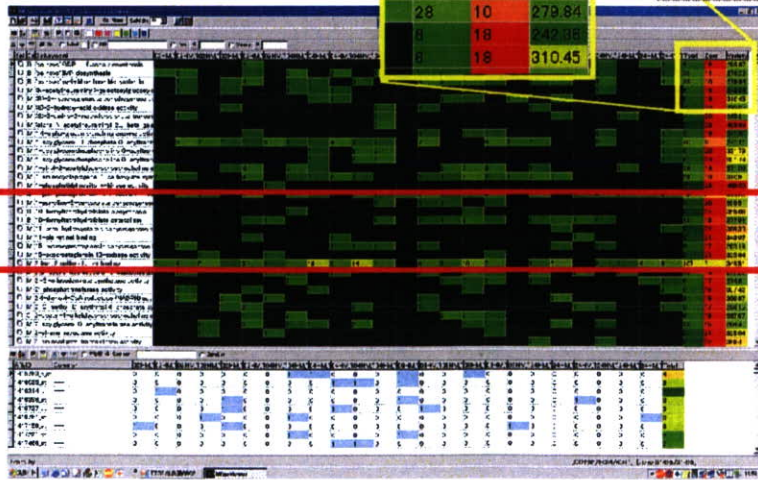
MF GoViewer



候補遺伝子リスト群のGO解析

M	Total	Zero	Variety
6	19	296.87	
20	11	279.22	
28	10	279.84	
6	18	342.38	
6	18	310.45	

Total: 集計
Zero: 該当無し
Variety: 変動の度合い



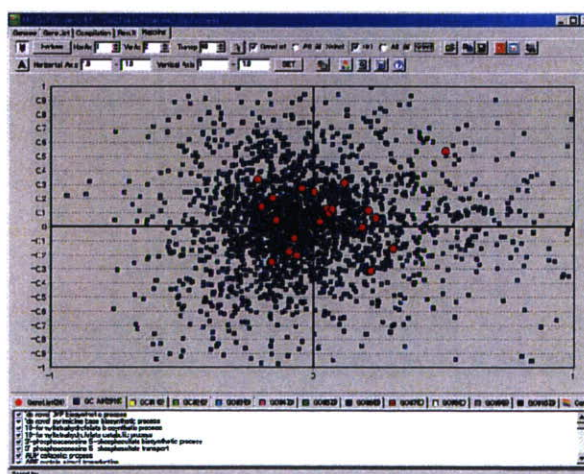
頻度分布による注目すべきGOの抽出



33

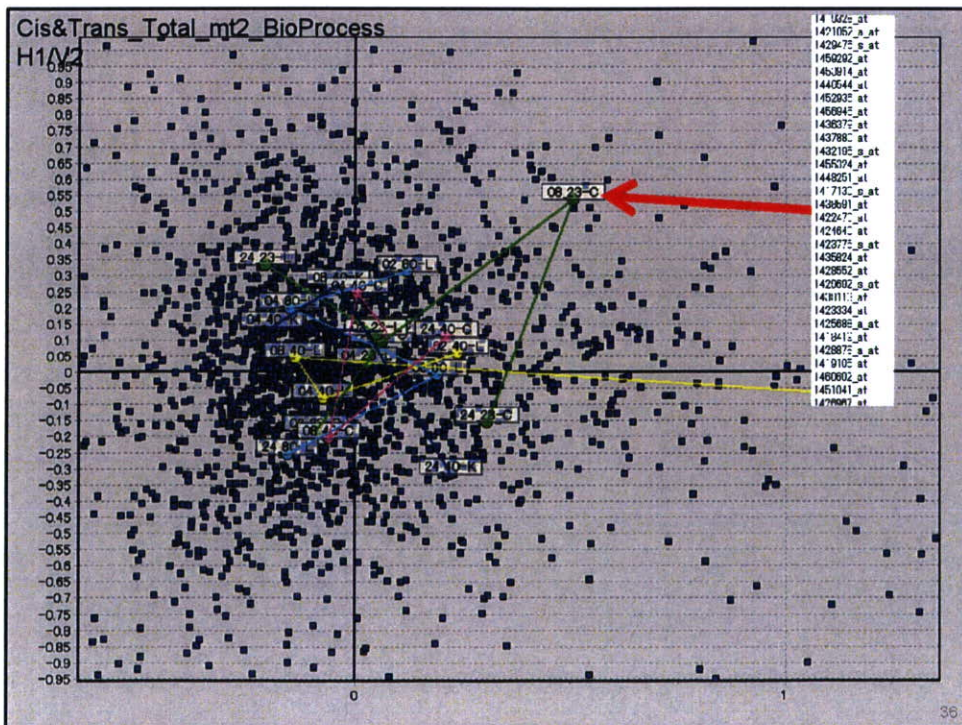
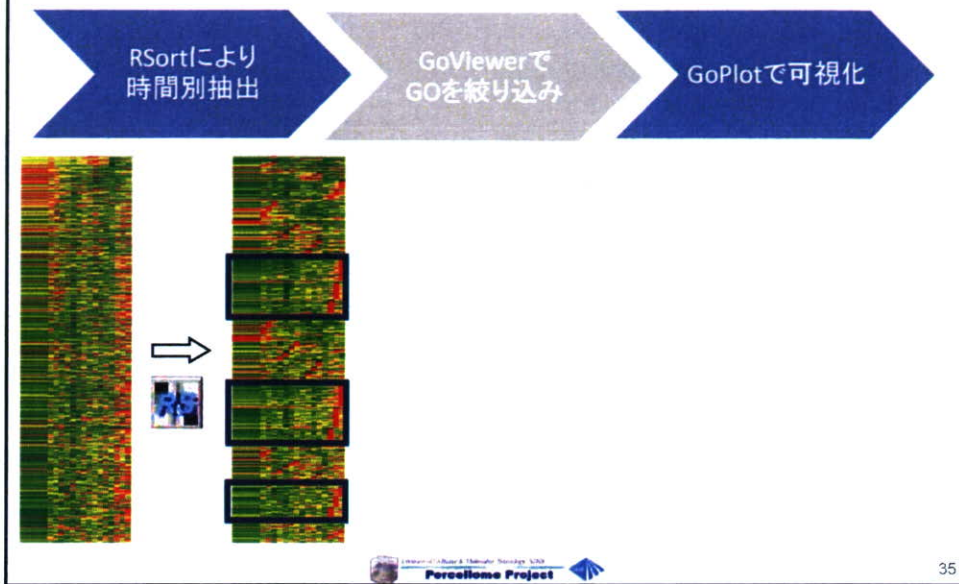
GoPlot

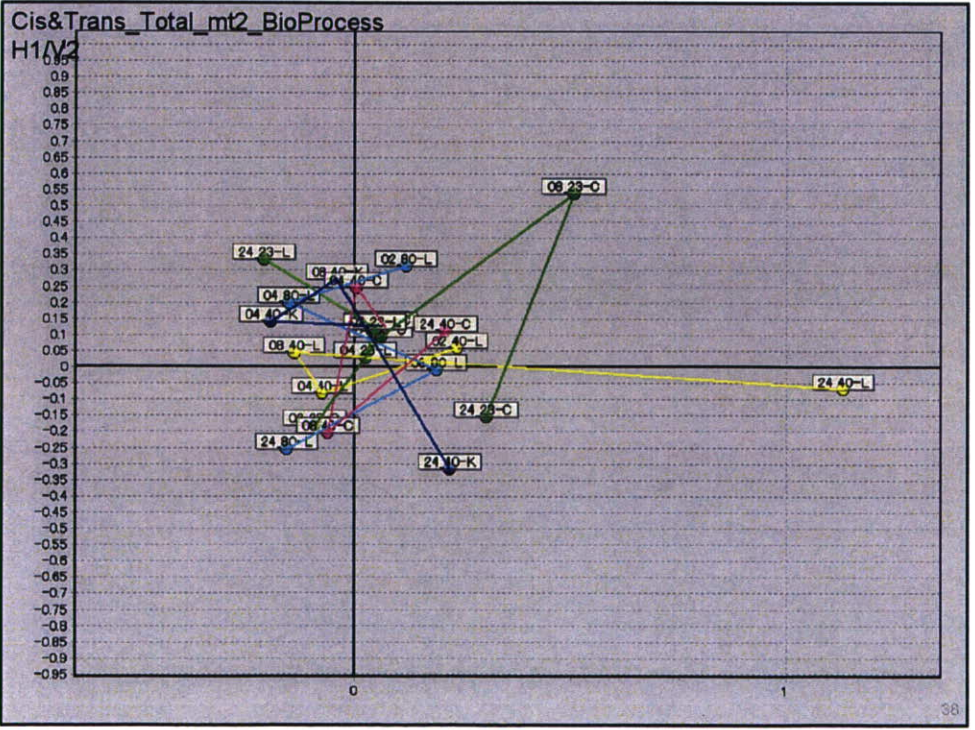
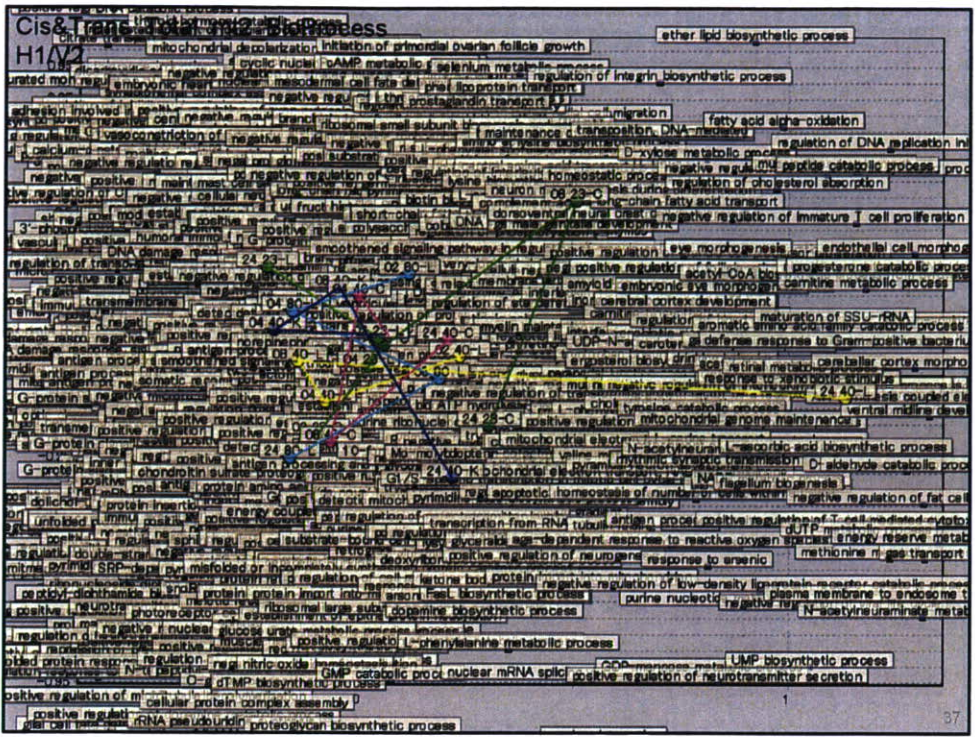
➤ 候補遺伝子群とGOキーワードの関連性を可視化する



34

例1:トランスクリプトーム経時変化の化合物間・臓器間比較





例1:トランスクリプトーム経時変化の化合物間・臓器間比較

