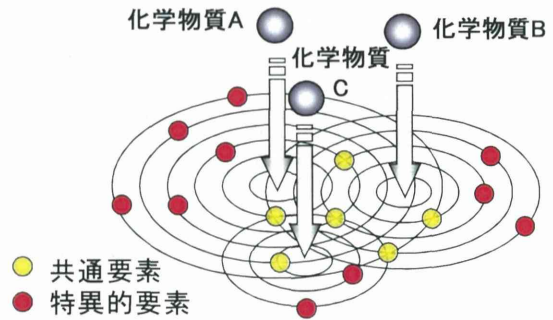


RSortの改良(要点)

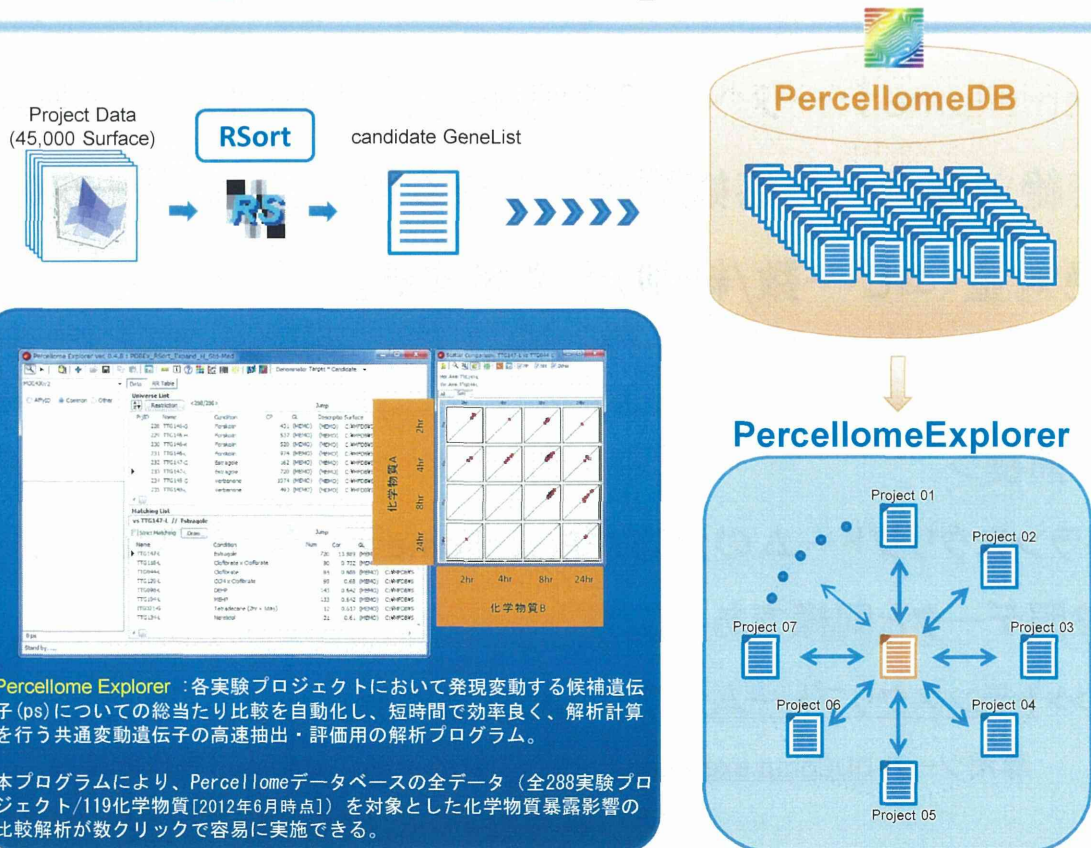
- RSortは見落としを許さない毒性学特有の前提条件に準拠しているため、元々、false positiveが出易い。
 - PercellomeExplorerの機能拡張(非共通要素＝特異的要素の解析機能)に際して、RSortの自動抽出精度を向上させる。
 - 計算処理時間を極力増大させない。
- RSortのroughness filterとの併用を前提に開発。



TTG4 菅野班 H26年度 総括 2012 02-13 version

43

PercellomeExplorer



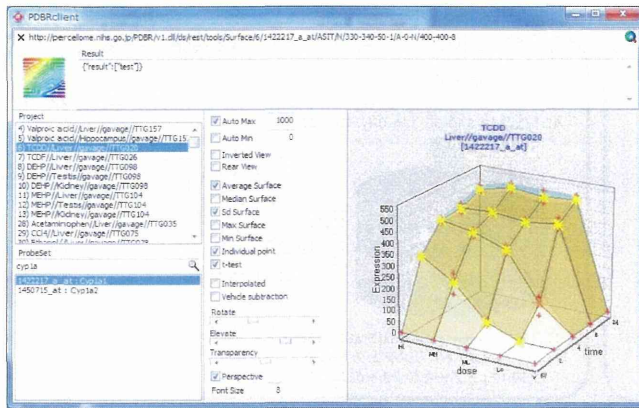
44

REST API of PercellomeWeb

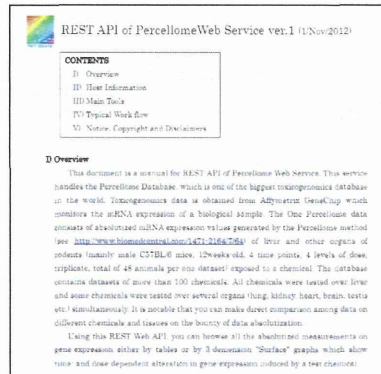
REST (Representational State Transfer)
ハイパーメディア システムのためのソフトウェア
アーキテクチャのスタイルのひとつ

http://percellome.nih.go.jp/PDBR/v1.dll/ds/rest/
tools/Surface/6/1422217_a_at/ASIT/N/330-340-
50-1/A-0-N/400-400-8

REST client 例



User manual



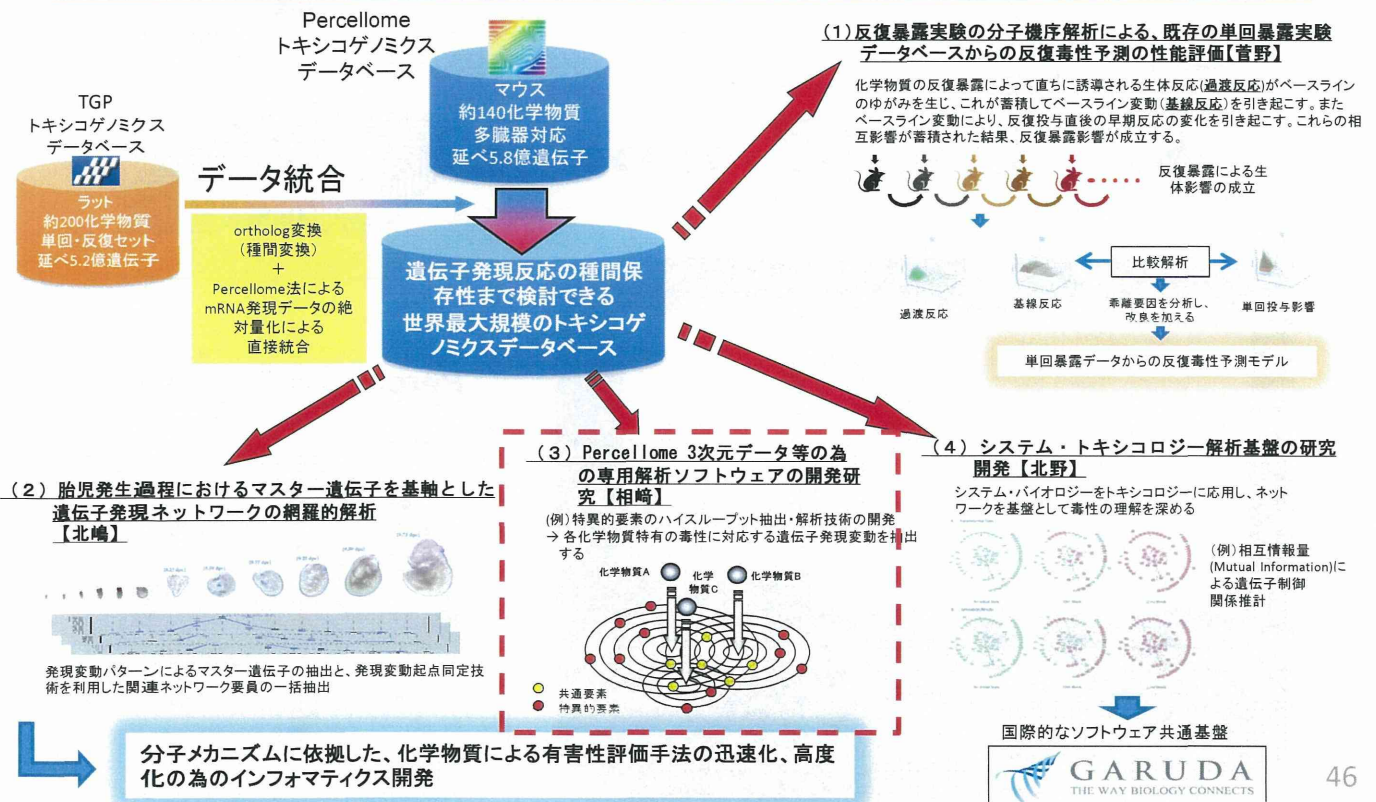
Functions

- Announcement
- PercellomeLogo
- Project
- ProbeSetId
- GeneInformation
- GroupAverageData(formatted)
- GroupSdData(formatted)
- GroupMedianData(formatted)
- GroupMaximumData(formatted)
- GroupMinimumData(formatted)
- IndividualData(formatted)
- SurfaceGraph

45

化学物質の有害性評価手法の迅速化、高度化に関する研究

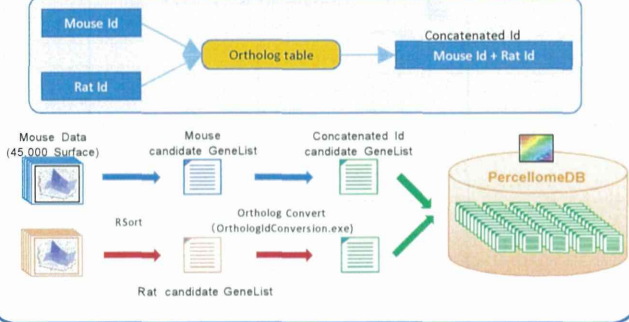
一網羅的定量的大規模トキシコゲノミクスデータベースの維持・拡充と毒性予測評価システムの実用化の為のインフォマティクス技術開発一



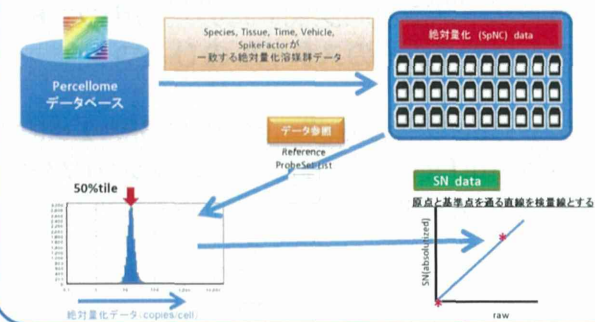
46

【分担研究】Percellome 3次元データ等の為の専用解析ソフトウェアの開発研究

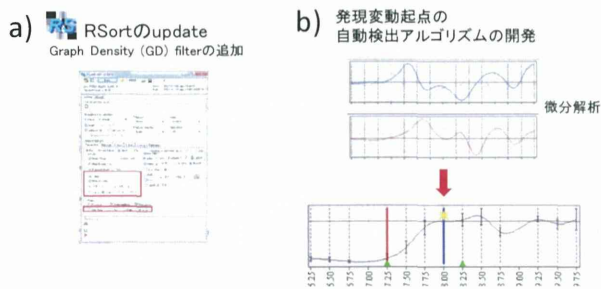
異種由来の遺伝子発現データの統合処理



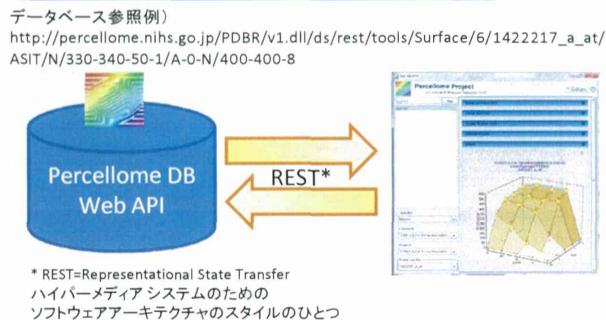
非Percellomeデータの絶対量推計



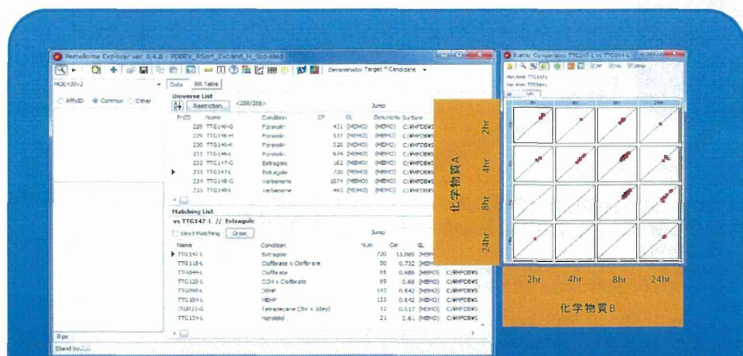
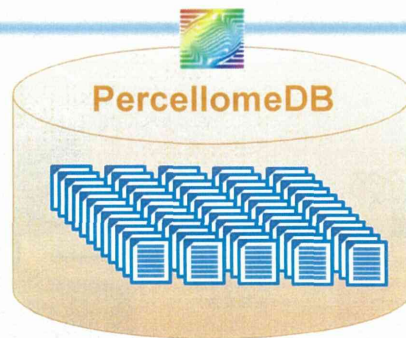
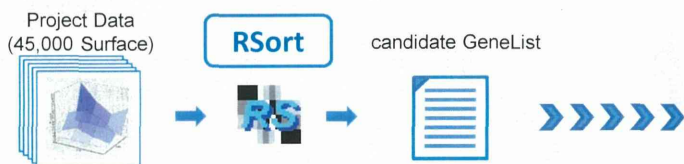
候補遺伝子自動抽出プログラムの改良



Percellome公開サーバーの機能拡張



Percellome Explorer

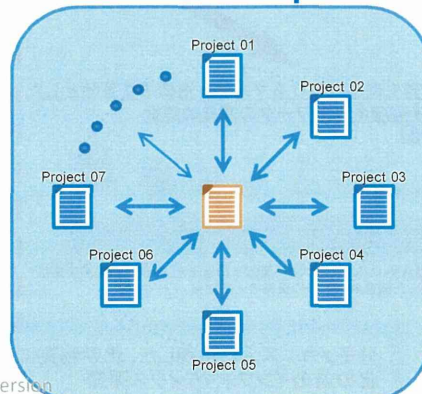


Percellome Explorer : 各実験プロジェクトにおいて発現変動する候補遺伝子 (ps) についての総当たり比較を自動化し、短時間で効率良く、解析計算を行う共通変動遺伝子的高速抽出・評価用の解析プログラム。

本プログラムにより、Percellomeデータベースの全データ (全288実験プロジェクト/119化合物 [2012年6月時点]) を対象とした化学物質暴露影響の比較解析が数クリックで容易に実施できる。

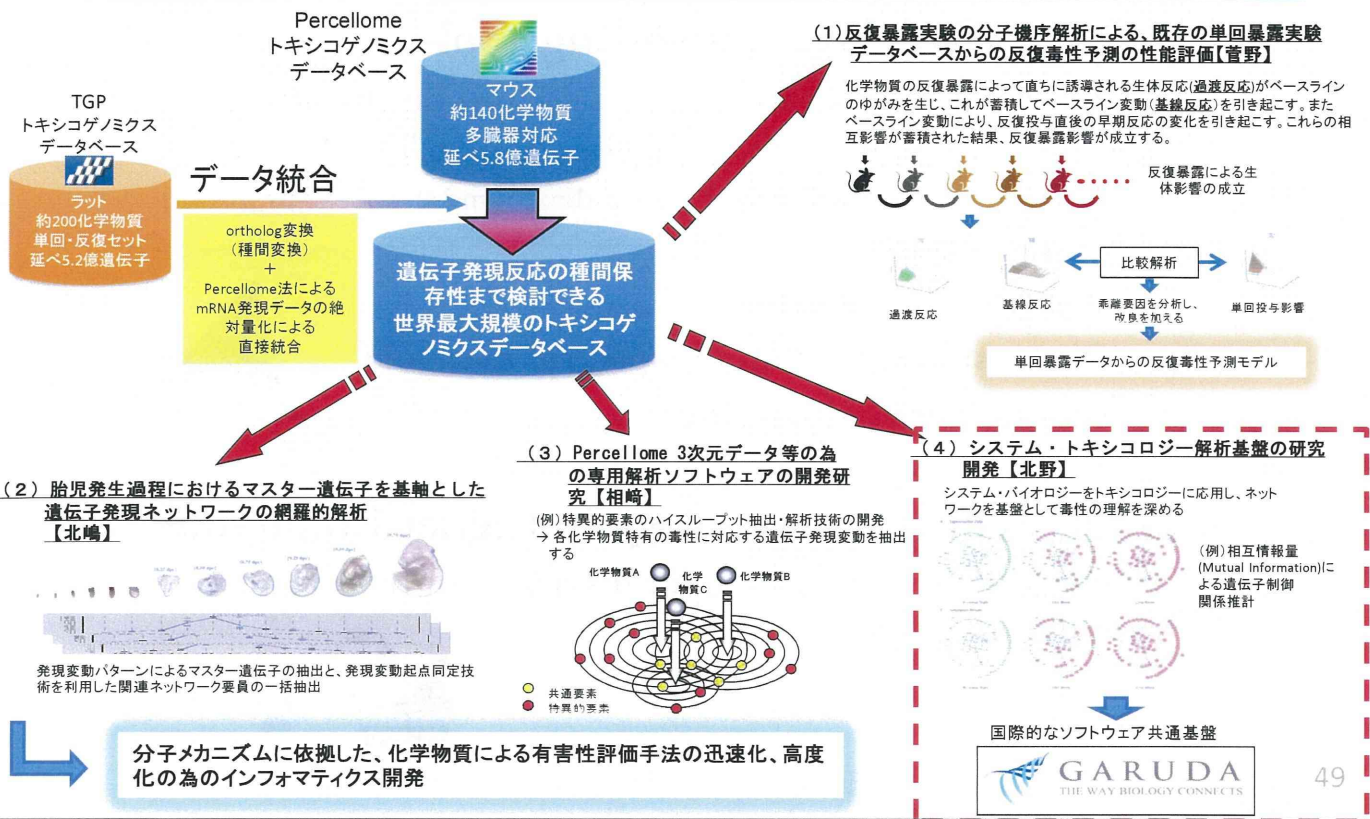
TTG4 菅野班 H26年度 総括 2012-02-13 version

Percellome Explorer

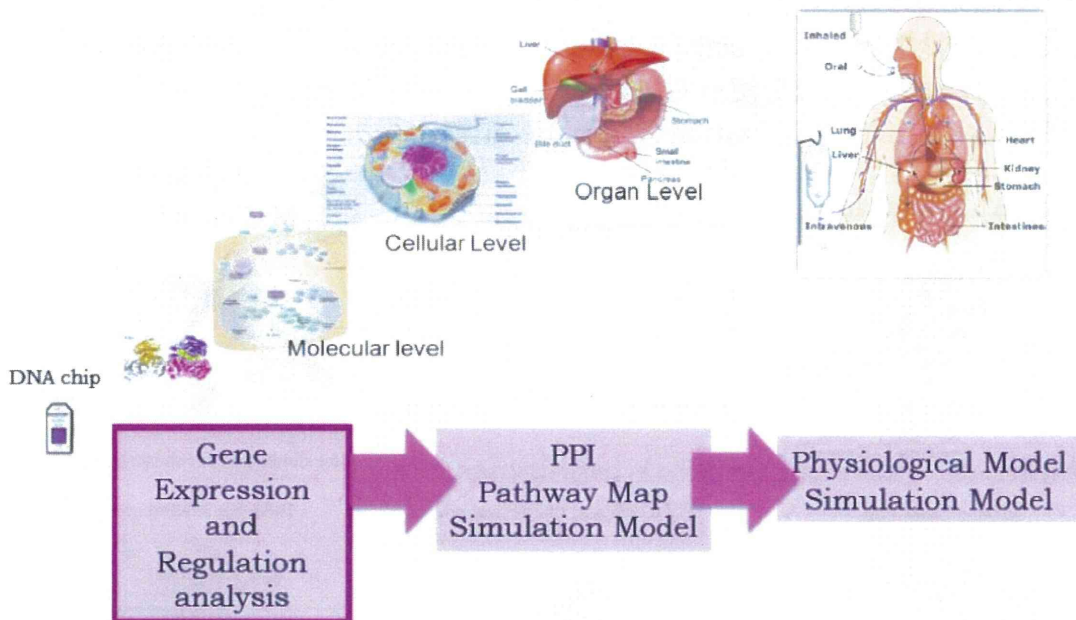


化学物質の有害性評価手法の迅速化、高度化に関する研究

— 網羅的定量的大規模トキシゲノミクスデータベースの維持・拡充と毒性予測評価システムの実用化の為にインフォマティクス技術開発 —



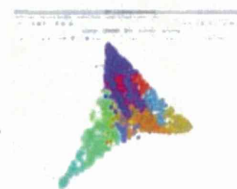
Data-driven analytic pipeline



Research Agenda

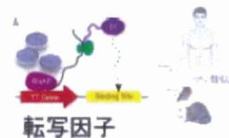
▶ **AGCT** A Geometric Clustering Tool

- ▶ Clustering Percellome data based on similarity of gene expression profile. Application to TCDD and TCDF (2,3,7,8-Tetrachlorodibenzo-p-dioxin and 2,3,7,8-Tetrafuran) chemicals.
- ▶ Sample normalization
- ▶ Dimension reduction on spectral manifold
- ▶ Unsupervised clustering



▶ **SHOE** Sequence Homology in Higher Eukaryote

- ▶ Phylogenetic footprinting for discovery of transcription regulation network



▶ 3

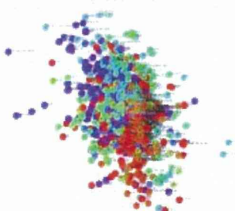
TTG4 菅野班 H26年度 総括 2012-02-13 version

51

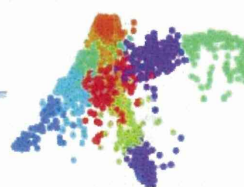
Processing data on AGCT

1. 時系列データ前処理:線形回帰/ウェーブレット変換
2. 遺伝子間の類似度マトリックス
3. 低次元に落とすためにSpectral clusteringを行う。通常の主成分分析も行う。
4. 発見的なClustering法を使って構造上でデータの分割を行う。
5. 結果のinteractive visualizationやscenario 記録を行う。

PCA : $M \times N$ matrix



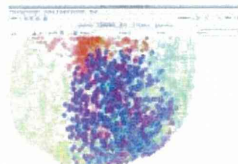
Spectral clustering:
 $M \times M$ matrix



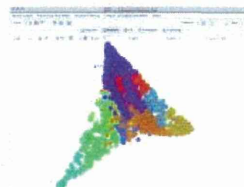
Orthogonal Laplacian matrix to compute one dimension per cluster/gene

Examples of different network topologies

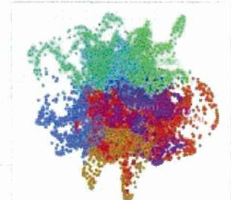
Mouse Stem cell



TCDD affected mouse liver cell



Influenza affected mouse bronchi cell

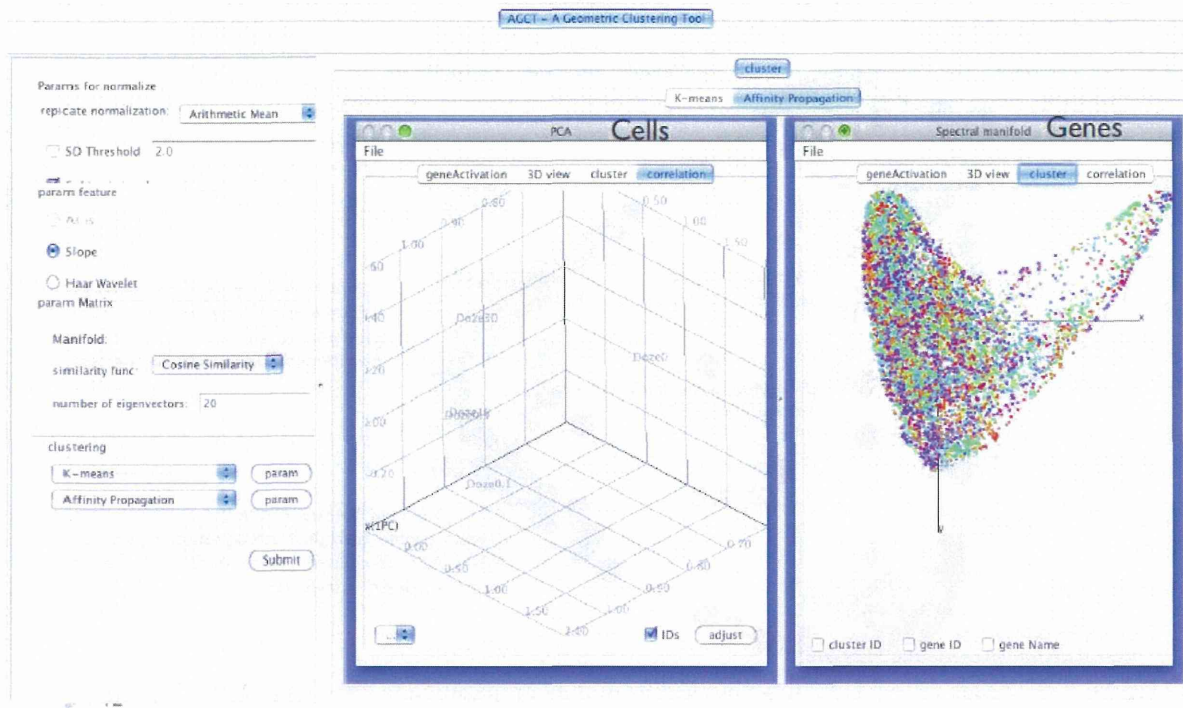


▶ 6

TTG4 菅野班 H26年度 総括 2012-02-13 version

52

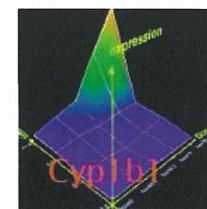
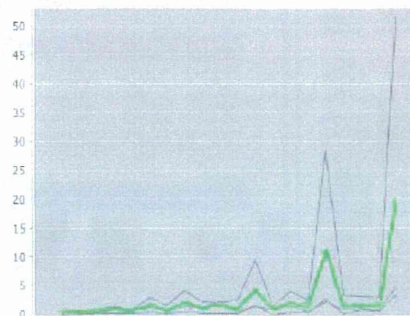
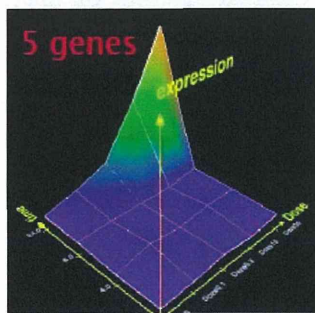
TCDD spectral clustering view



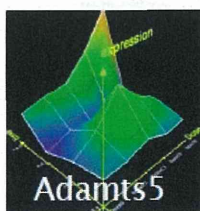
TTG4 菅野班 H26年度 総括 2012-02-13 version

53

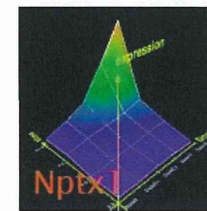
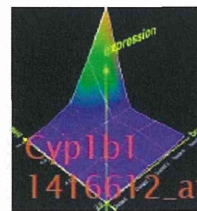
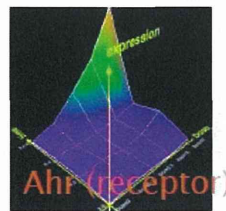
TCDD Ahr cluster



Metabolic enzyme



metalloproteinase



binding protein for the snake venom toxin taipoxin

▶ 16

TTG4 菅野班 H26年度 総括 2012-02-13 version

54

Gene ontology for 8 TCDD clusters

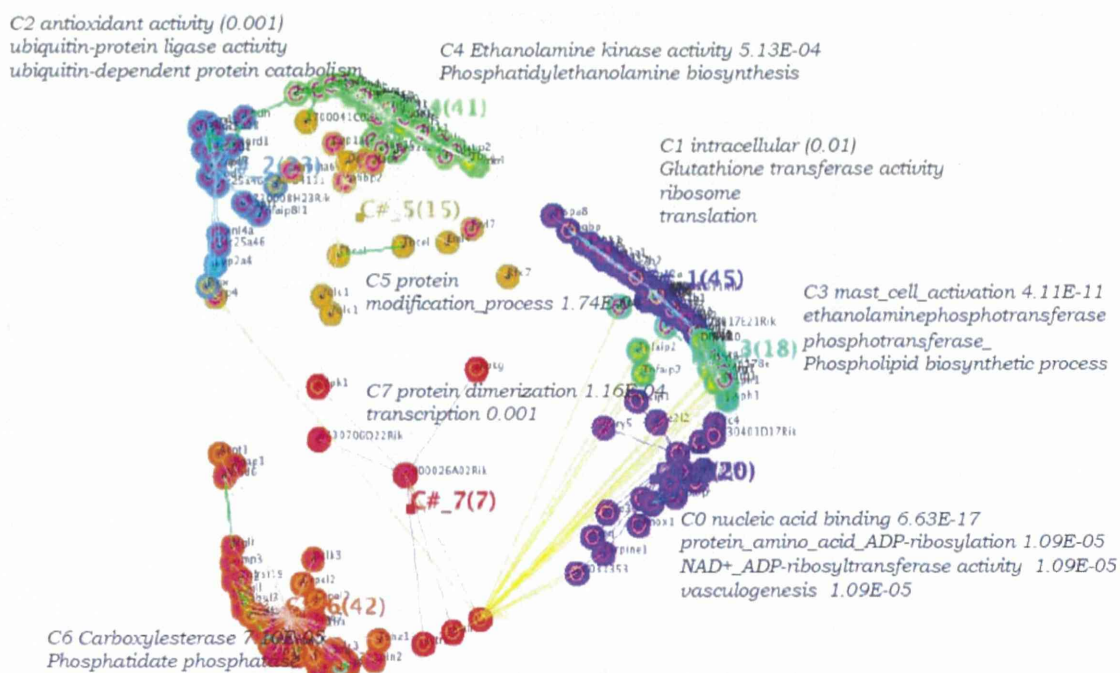
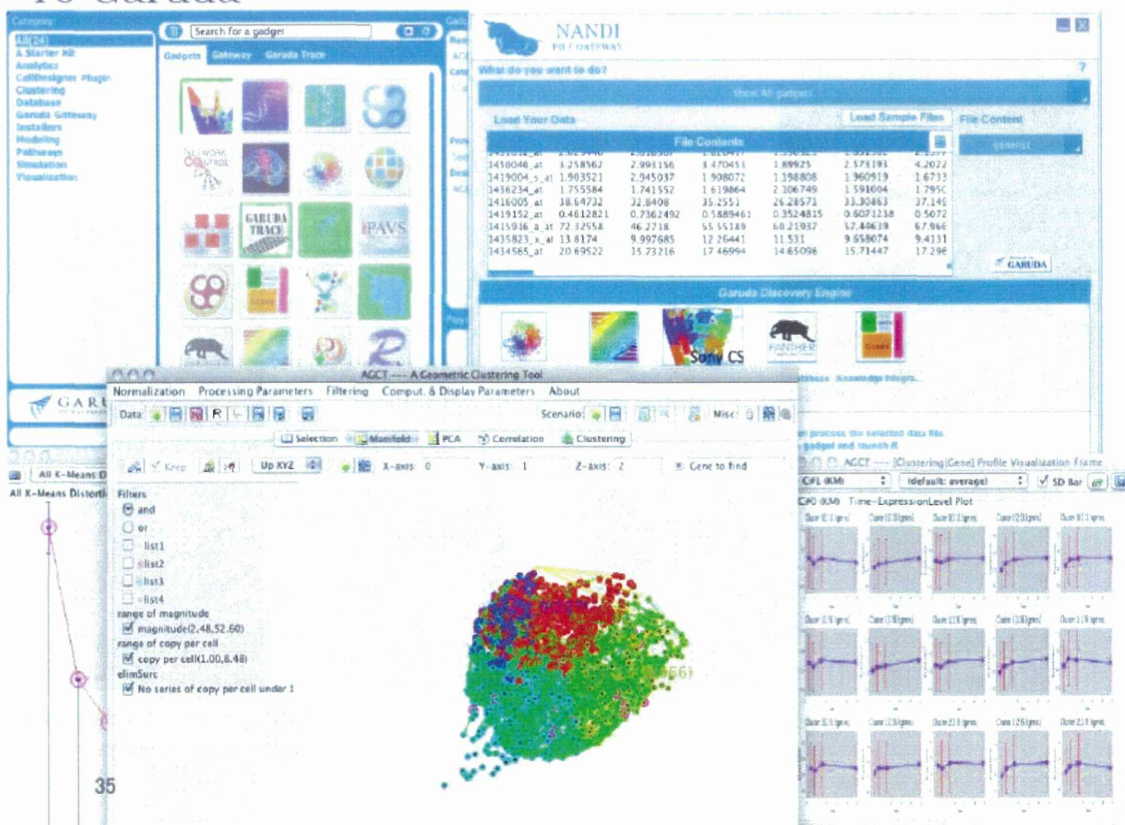


Figure 2: GO terms for eight clusters obtained by Affinity propagation method on 211 TCDD probes.
28

To Garuda



BrowserTest

SONY Sony CSL **SHOE** Sequence HOmology in higher Eukaryotes Job Input Queue List

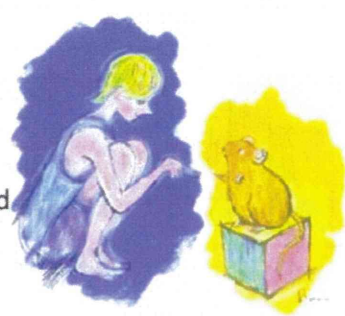
Gene List
 NM_001287
 NM_004068
 NM_001654
 NM_006136
 NM_004930
 NM_012120

Repeat Masker

Upstream Length 2000 mode 1 mode 2

Downstream Length 200

Scoring Transfac32 Jaspas IPS reprogramming factors



Sequence HOmology in higher Eukaryotes

1. Activated

BrowserTest

SONY Sony CSL **SHOE** Sequence HOmology in higher Eukaryotes Job Input Queue List

146 126 Compare

No	Status	Start Date	End Date	Gene List	Repeat Masker	Upstream Length	Mode	Downstream Length	Scoring	Delete
219	finished Result Relations	2014/09/09 16:23:13	2014/09/09 16:28:17	NM_009993 NM_009992 (2 genes)	Checked	2000	mode2	200	transfac32	Delete
216	finished Result	2014/09/09 15:56:26	2014/09/09 16:01:03	NM_009993 NM_009992		5000	mode2	200	transfac32	Delete

2. Pairwise job comparison

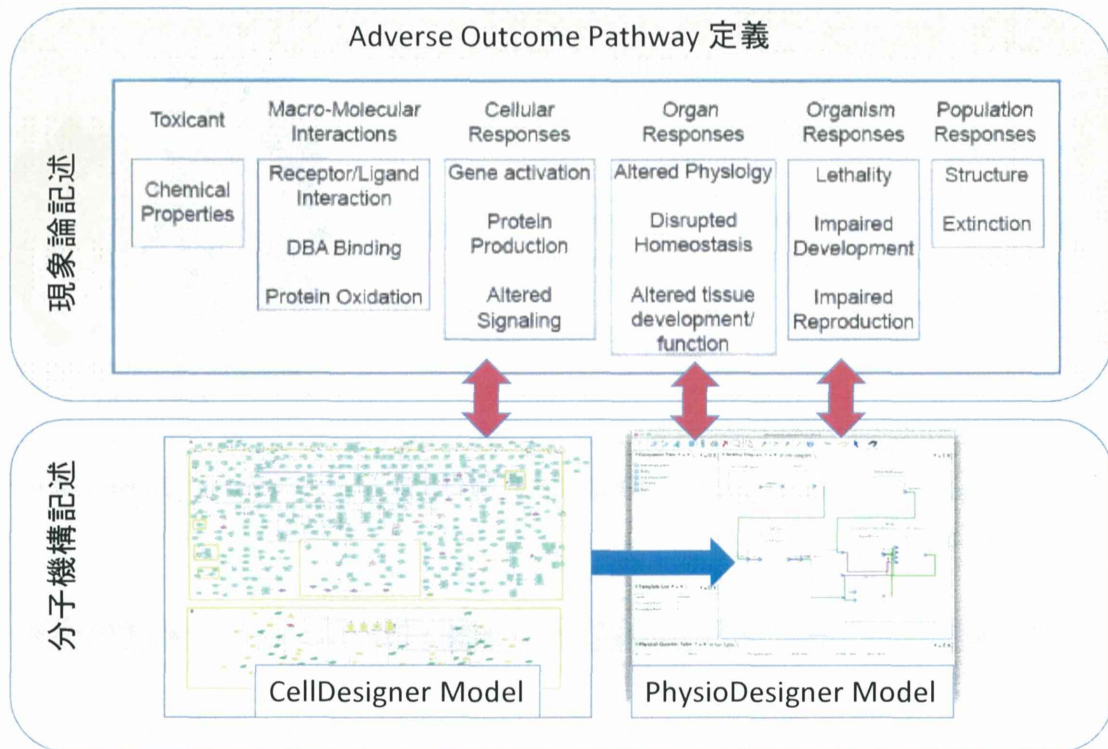
TTG4 菅野班 H26年度 総括 2012-02-13 version

57

Next Steps

The Systems Biology Institute

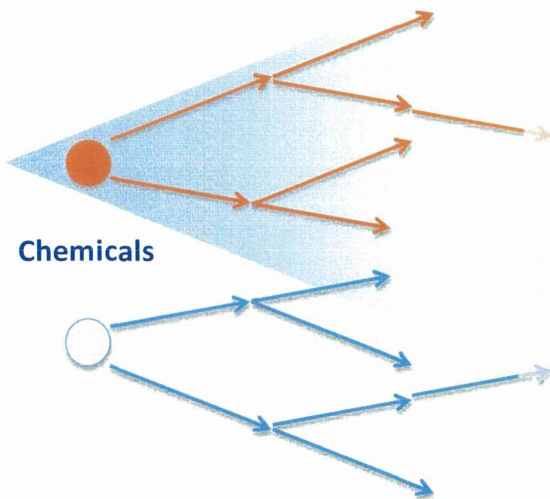
Integration with APO



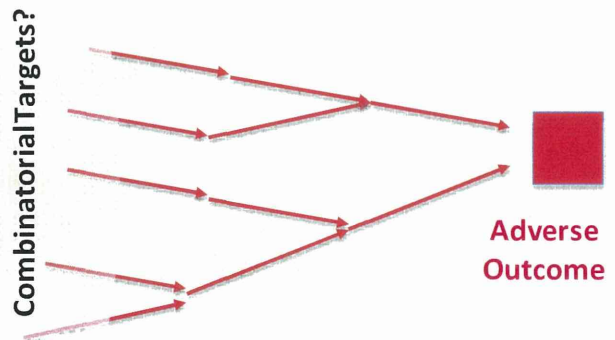
TTG4 菅野班 H26年度 総括 2012-02-13 version

59

Forward Effect Analysis



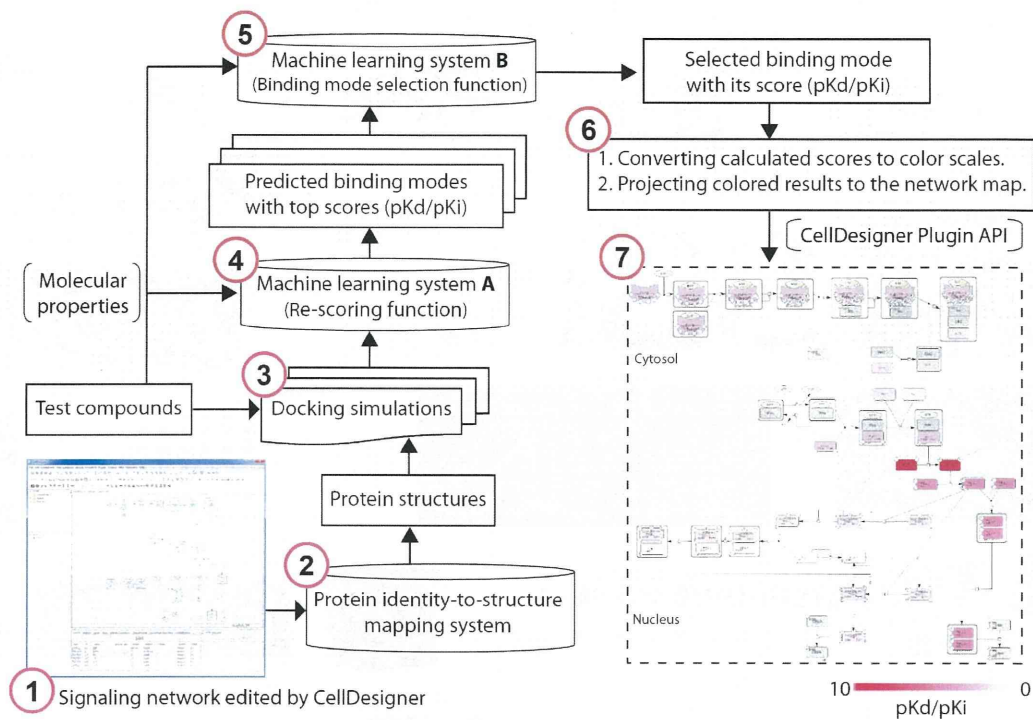
Backward Effect Analysis



TTG4 菅野班 H26年度 総括 2012-02-13 version

60

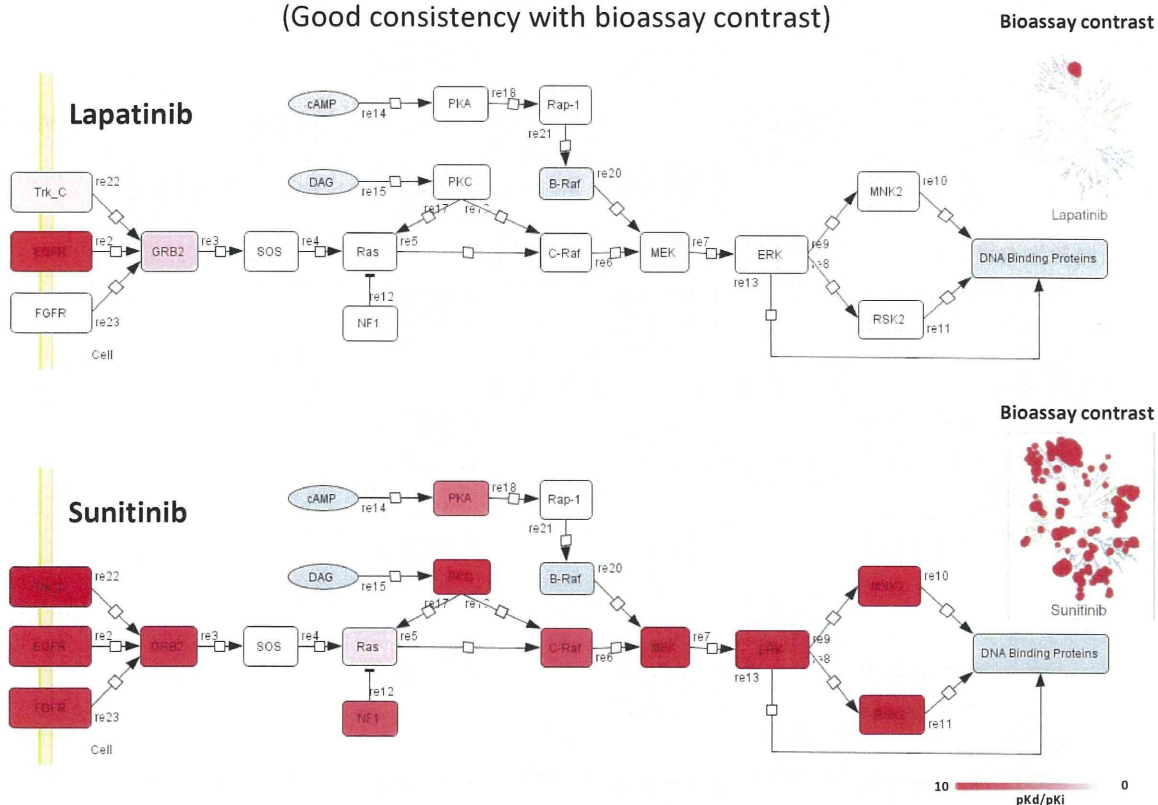
Network-based screening pipeline



TTG4 菅野班 H26年度 総括 2012-02-13 version

61

Case Study Using Our Method over MAPK Pathway (Good consistency with bioassay contrast)



TTG4 菅野班 H26年度 総括 2012-02-13 version

62