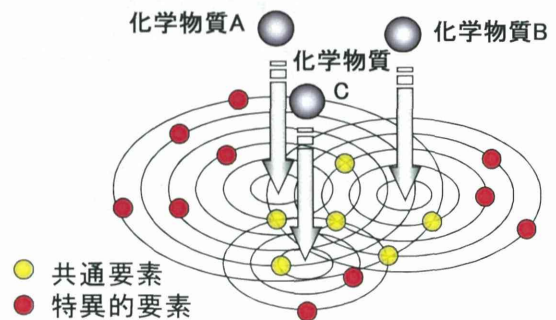




RSortの改良(要点)

- 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。



31



RSort : false positive対策

検討したアルゴリズム

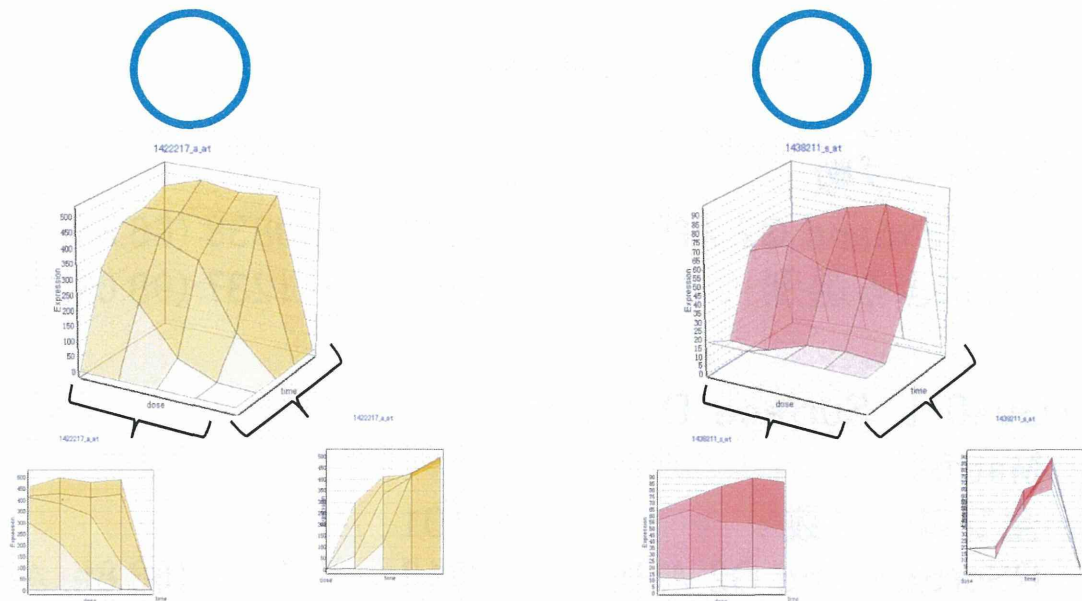
- 最大ピークサイズの最小値
- 用量曲線、経時変化曲線の変動パターン数(PCC)
- 用量曲線、経時変化曲線の変曲点数(IFP)
- 用量曲線、経時変化曲線の平滑さ(S)
- Graph Density(GD)
Surfaceグラフ曲面を用量軸側、時間軸側から観察した時の描画密度(重複度)

32



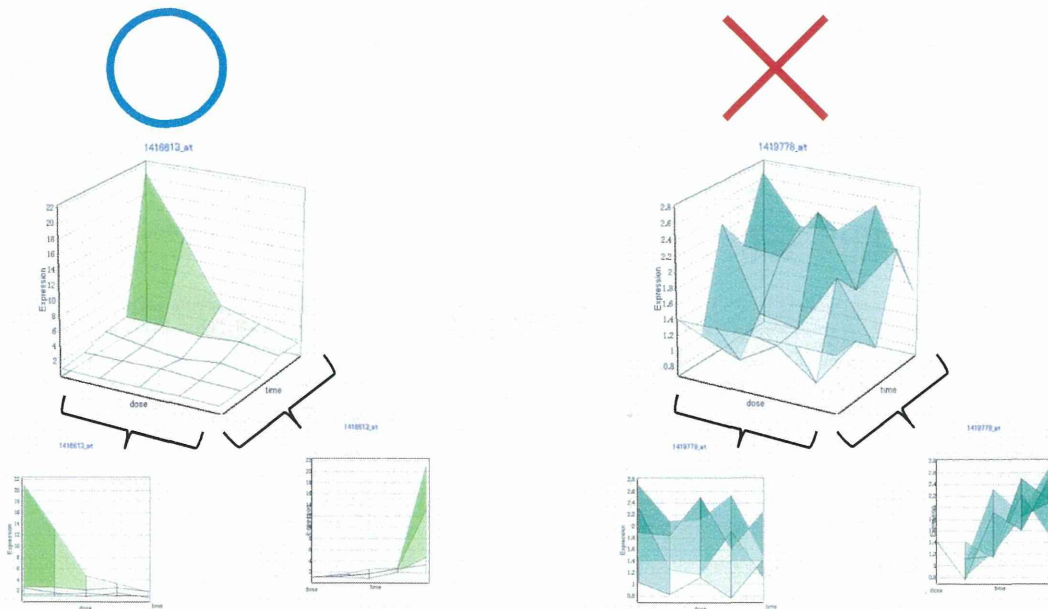
RSort : false positive 対策

Graph Density (GD) filter



RSort : false positive 対策

Graph Density (GD) filter





RSort :Graph Density (GD) filter

設定) GD ver.18 // Max DensityX = 3.5, Y=3.5, X or Y

TTG016-L(C) Surface DataSet

Matrix: 4x5

全プローブ数 : 45101

RSort(Expand_G2_AP) : 4420 → 3838 (86.8%, Δ582)

目視による選別 : 1255 → 1237 (98.6%, Δ18)

TTG020-L Surface DataSet

Matrix: 5x5

全プローブ数 : 45101

RSort(Expand_G2_AP) : 2713 → 1866 (68.8%, Δ847)

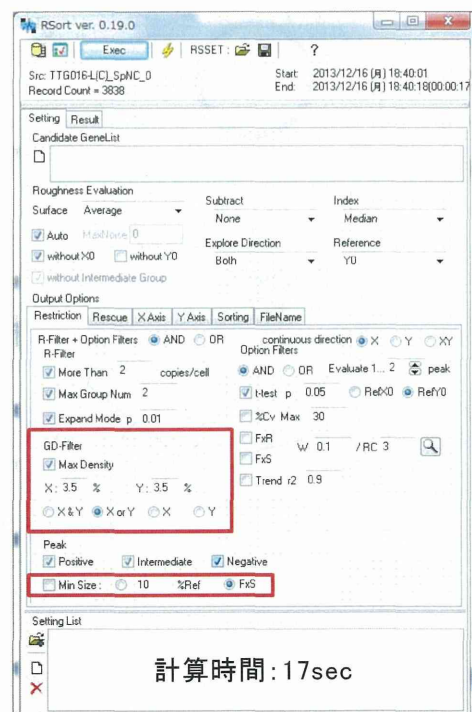
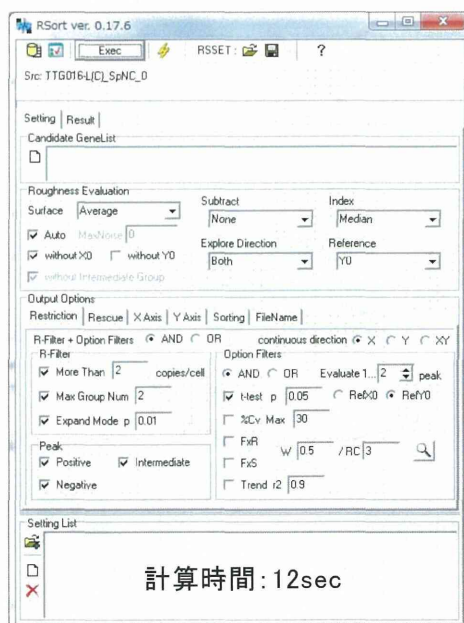
目視による選別 : 302 → 292 (96.7%, Δ10)

35



RSort : false positive対策

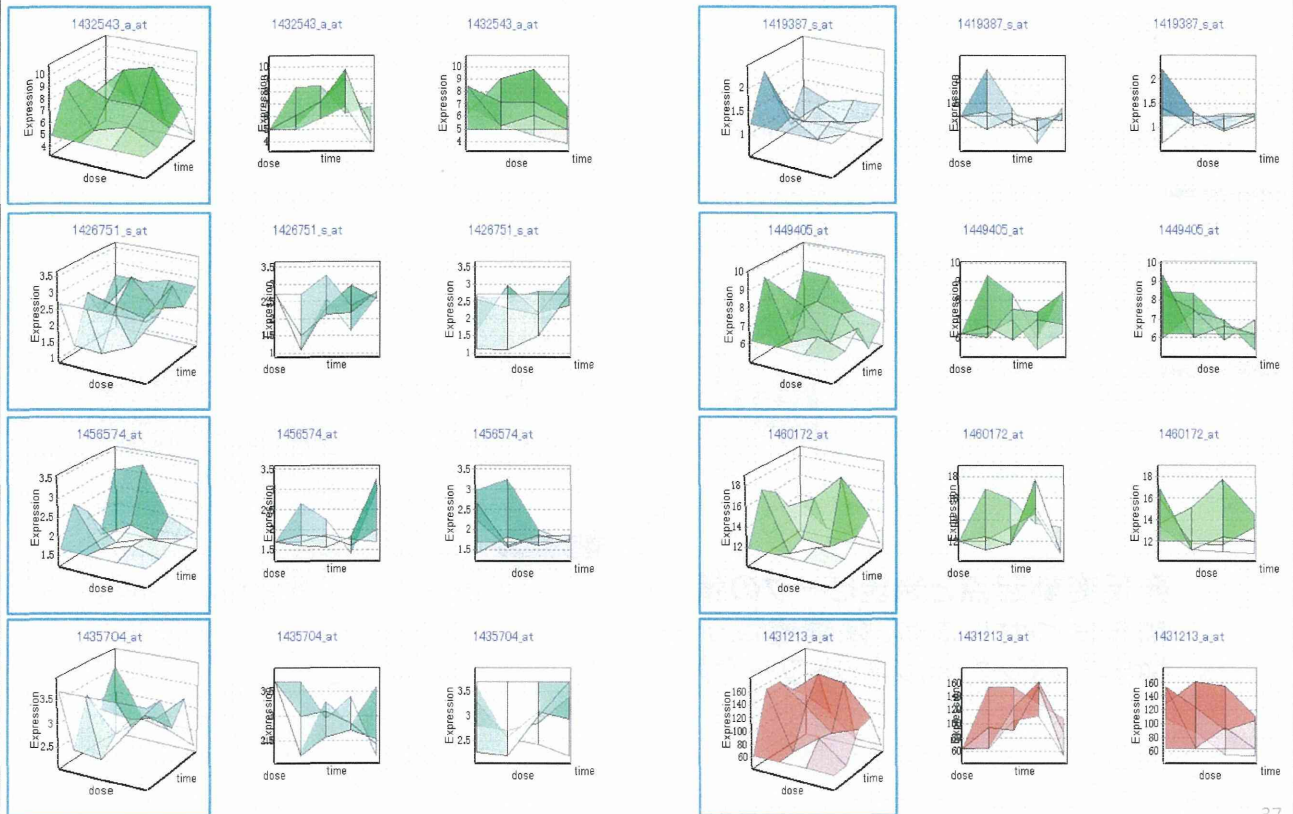
RSort ver.0.19.0 with Graph Density (GD) filter



36



TTG016-L(C)におけるGD filterの過剰カット



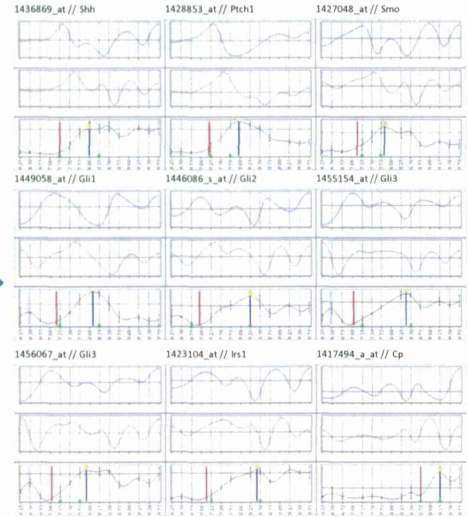
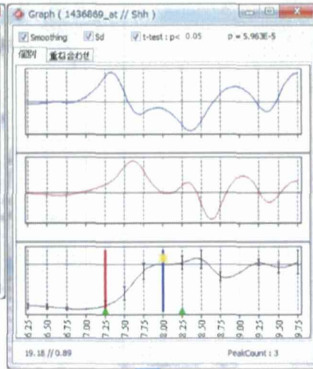
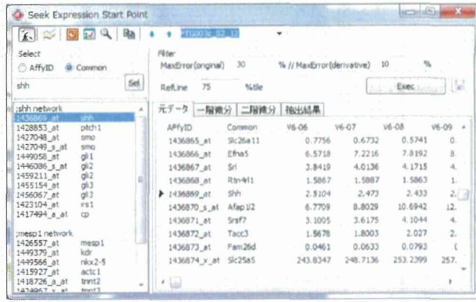
発現変動起点の検出

false positive削減の試み



微分解析による発現変動起点の検出

SeekESP



発現変動起点と発現ピークの情報の組み合わせによる、候補遺伝子の抽出及びシグナルネットワーク解析

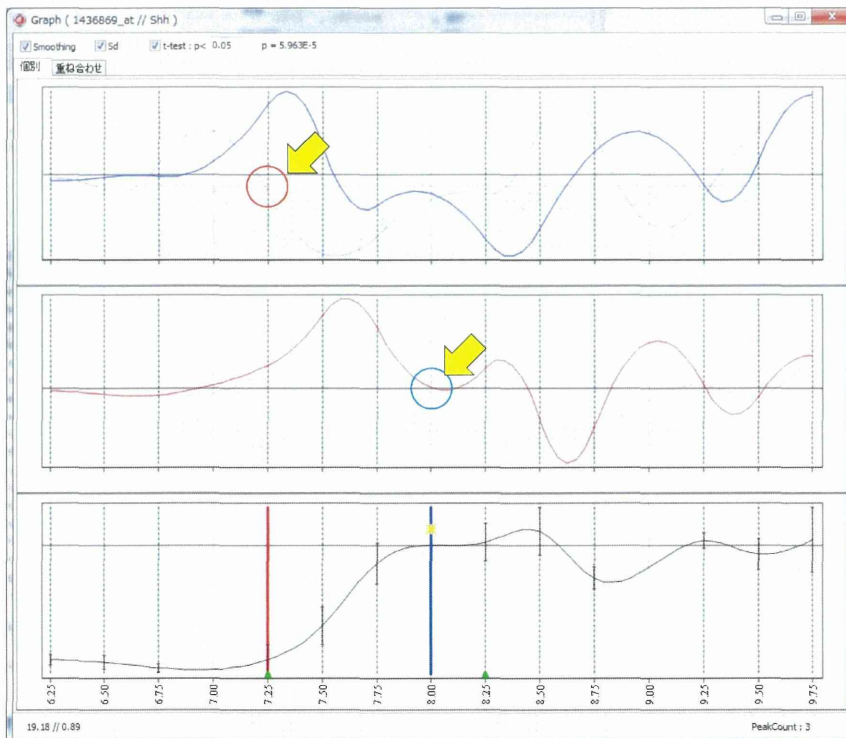


微分解析による発現変動起点の検出

二階微分
三階微分

一階微分

元データ





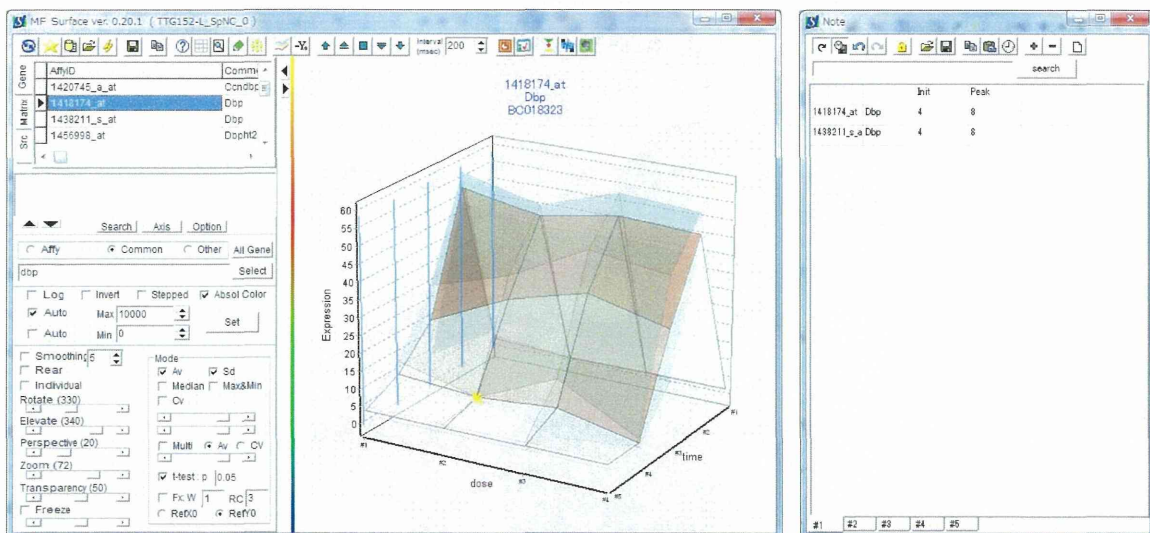
MFSurfaceの機能拡張

評価・解析業務の効率化

41



Note機能の追加など



42

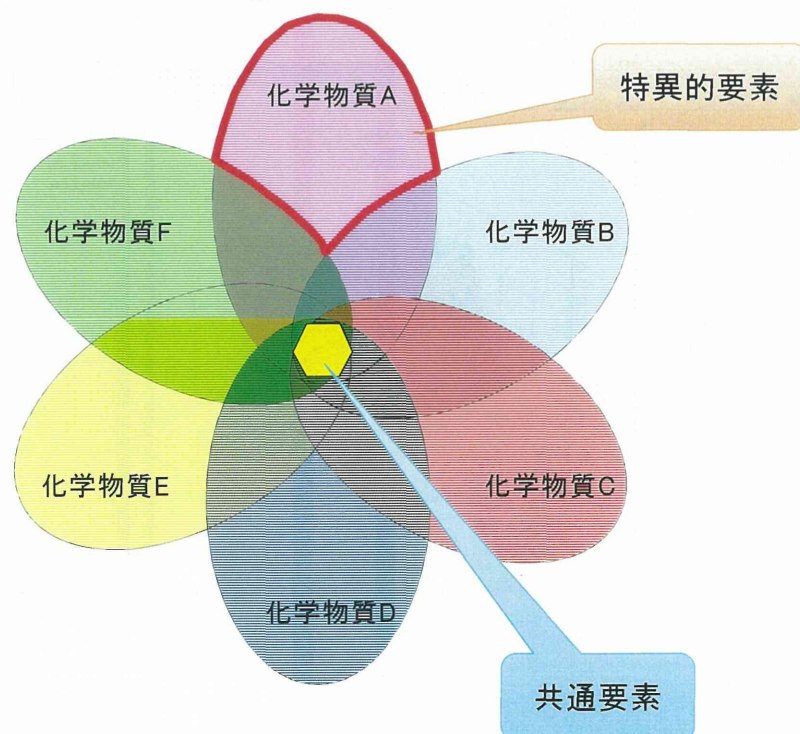


特異的要素解析

43



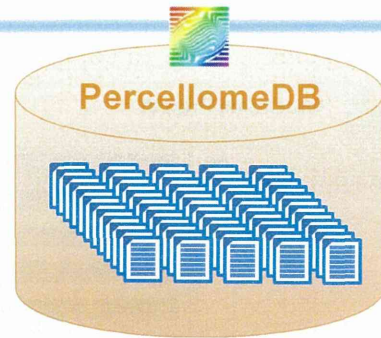
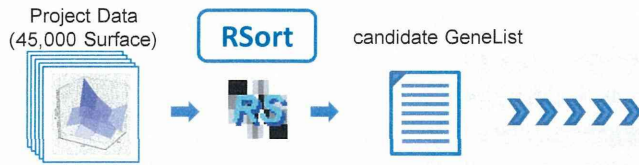
共通要素と特異的要素



44



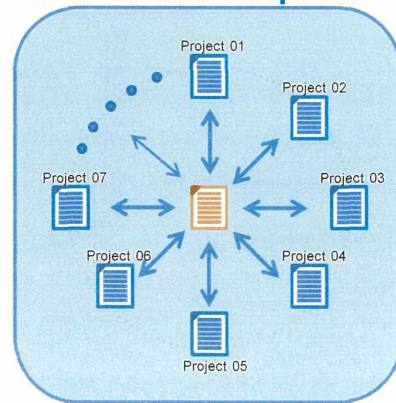
Percellome Explorer



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の高速抽出・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化学物質[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。

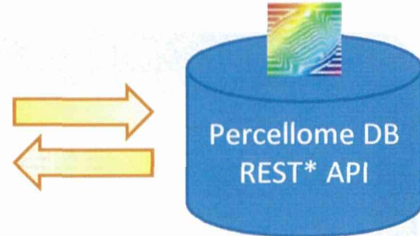
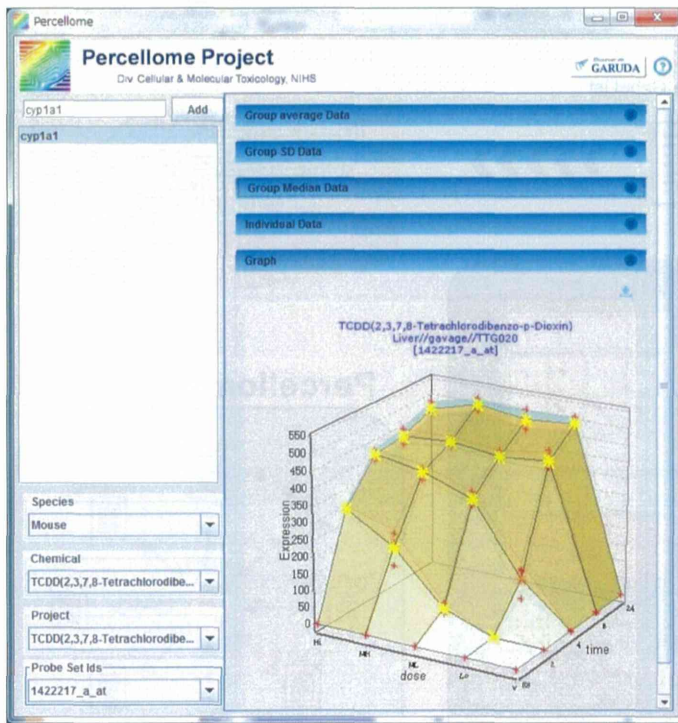
Percellome Explorer



Percellomeプロジェクト成果の 一般公開状況



Percellome Gadget



*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

API
(Application Programming
Interface)
アプリケーション(ユーザのプログラム)がOS等
の公開機能を使うためのインターフェース。

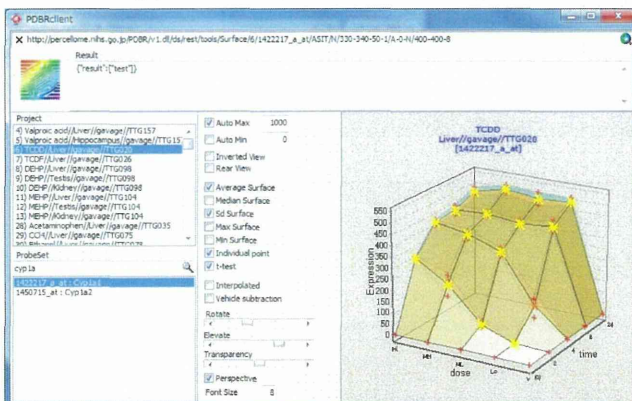


REST API of PercellomeWeb

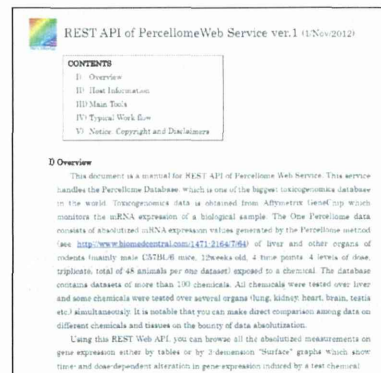
REST (Representational State Transfer)
ハイパーメディアシステムのためのソフトウェア
アーキテクチャのスタイルのひとつ

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



Percellome Gadget with REST API of PercellomeWeb

- Project (ex. keyword="TCD")

<http://percellome.nihs.go.jp/PDBR/v1.dll/ds/rest/tools/project/mouse/TCD>

- Probe Set ID (ex. keyword="cyp1")

<http://percellome.nihs.go.jp/PDBR/v1.dll/ds/rest/tools/psid/mouse/cyp1>

- Gene Information

http://percellome.nihs.go.jp/PDBR/v1.dll/ds/rest/tools/geneinfo/mouse/1422217_a_at

- Group Average Data (formatted)

http://percellome.nihs.go.jp/PDBR/v1.dll/ds/rest/tools/AvData/4/1422217_a_at

- Surface Graph

http://percellome.nihs.go.jp/PDBR/v1.dll/ds/rest/Tools/Surface/6/1422217_a_at/AS/N/330-340-50-1/A-0-N/300-300-8

- New ● Search Chemicals by keyword (ex. keyword="oxidative")

<http://percellome.nihs.go.jp/pdbr/v1.dll/ds/rest/tools/chemical/oxidative>

- New ● Retrieve RSort Genelist (ex. ProjectId=4)

<http://percellome.nihs.go.jp/pdbr/v1.dll/ds/rest/tools/genelist/4>

49



H24年度の成果

- TGPラット・トキシコゲノミクスデータの統合
～シームレスなOrtholog発現解析技術の開発
- 一般的な(外部Spikeが追加されていない)測定済み
マイクロアレイデータの絶対量推定技術の開発
- 国際的なソフトウェア共通基盤Garudaへの対応
～REST API of PercellomeWeb service ver.1

50



H25年度の成果

- RSortの改良
自動抽出の性能向上
溶媒影響(概日リズムを含む)の分離
- 微分解析による発現変動起点の検出アルゴリズム
開発
- 国際的なソフトウェア共通基盤Garudaへの対応
REST API of PercellomeWeb service ver.2

51



H26年度の成果

- 非Percellome絶対量推計:ラットデータ拡充
- ideal controlデータの生成
~新型反復暴露実験の基線反応評価の精度向上
- MFSurfaceなどの、解析作業を効率化するための
ユーザーインターフェイス拡張
- PercellomeExplorerを利用した特異的要素解析の
検討

52

厚生労働科学研究費補助金（化学物質リスク研究事業）
分担研究報告書

「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－」
(H24-化学-指定-006)

分担研究課題：「システムトキシコロジー解析基盤の研究開発」

研究分担者： 北野 宏明

特定非営利活動法人システム・バイオロジー研究機構 会長

研究要旨

本研究では、システムバイオロジーをトキシコロジーに応用し、大規模かつ網羅的なデータを有効に利用し、化学物質の毒性の機序の理解を推進すると同時に、遺伝子発現データに反映される生理学的影響から、その原因物質が、遺伝子制御ネットワークのどの部分に影響を及ぼしているのかなどを同定する手法を開発した。これは、昨年度に開発された複数のアルゴリズムの併用を行う方式の遺伝子制御関係推定アルゴリズムであり、現状での最も優れた方式と同等の結果を安定的に維持することに成功したほか、情報エントロピーの計算より最適なアルゴリズムの組み合わせを探索し、動的に推論エンジンを構築する技術を実現した。さらに、このシステムを Inference cloud として Web-service として提供できるように開発を行った。また、クラスタリングソフトウェア（AGCT）の Percellome データへの最適化や計算規模及び速度の向上を実施し、高速化を達成した。さらに、Percellome データベースの、国際的なソフトウェアの共通基盤 Garuda Platform 上への実装を進め、そのソフトウェアを Garuda platform の一般公開バージョンとして配布を開始した。

A. 研究目的

本研究では、大規模かつ網羅的なデータを有効に利用し、化学物質の毒性の機序の理解を推進すると同時に、遺伝子発現データに反映される生理学的影響から、その原因物質が、遺伝子制御ネットワークのどの部分に影響を及ぼしているのかなどを同定する手法を開発した。この手法によって、大量の文献情報などから集約したボトムアップな知識の集積によって毒性の機序が理

解できると同時に、大規模データから毒性に関連して変動する遺伝子群の同定とそれらの相互作用関係、さらには、各々の生理学的影響または毒性に相関すると思われる遺伝子の変動に対して、それらの上流に存在するどの遺伝子群やどの遺伝子群を制御する転写因子や受容体が重要であるかを判別することが出来る。

同時に、これら一連の手法は、ソフトウ

エアとして実現し、広く利用されることが望ましい。そこで、我々が、提唱し、開発が進んでいる、Garuda Common Platformに準拠した一連のソフトウェアツールを開発した。

B. 研究方法

本研究は、システムバイオロジーをトキシコロジーに応用する研究であり、従来の手法ではとらえきれなかったネットワークを基盤とした毒性の理解を可能とする。このため、大規模データを利用する手法と大規模文献情報などをいかに集約するかに関する研究、そしてそれらを統合的なプラットフォームに実装し広く利用できるようにする研究によって構成される。よって、この研究は、以下の項目に区分できる：

- (1) 大規模データからの遺伝子制御関係推定アルゴリズムの開発と改良
- (2) 大規模データからの状態制御遺伝子群推定アルゴリズムの開発と改良
- (3) 大規模文献情報などの集約手法と伝達手法の研究
- (4) Garuda Platform とその関連ソフトウェアの研究開発

C. 研究成果

1: 大規模データからの遺伝子制御関係推定アルゴリズムの開発と改良

Percellome や多くの研究で利用されている遺伝子発現データを利用し、遺伝子間の発現変動の相互関係と因果関係ネットワークの構築を可能とする手法の研究を行っ

た。現在、遺伝子発現データからのネットワーク抽出の研究は多く行われているが、その精度は高くない。例えば、広く使われているベイズ統計を利用した方法は、比較的少数の遺伝子に対して大量のデータが存在するときには適切な方法とは言えない。このような状況では、相互情報量(Mutual Information)を利用する手法の方が適しているが、その計算の具体的な手法により、推定結果の精度などは大きく違ってくる。我々では、既に、複数の計算手法の組み合わせを利用して、これらの現存手法より、精度の高い手法を開発した。

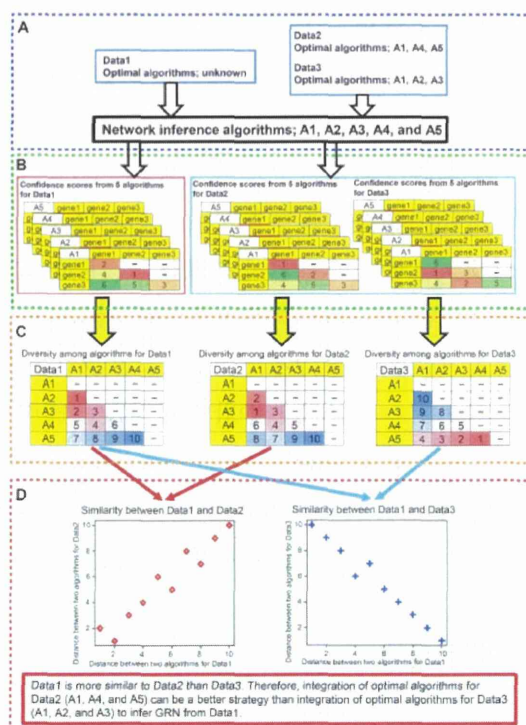


図1：データの特徴分析から最適アルゴリズム群を決定

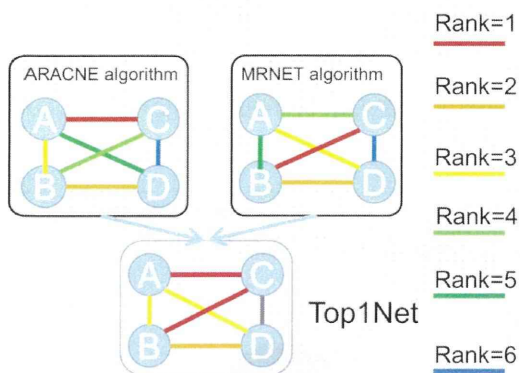


図 2 : 最適アルゴリズム群内の複数の手法からの遺伝子制御ネットワーク推定方法

推定アルゴリズムの手法の違いにより、どのようなデータにはどのアルゴリズムが最も適切かが違う。そこで、我々の手法では、データの全体的な特徴で利用するアルゴリズム群を決定し、個別の因果関係に関して各々の手法からの推定信頼度を参照して、最終的な推定ネットワークを構築する方法を開発した。

すなわち、我々の手法は、データの特徴分析から、そのデータの解析に最適なアルゴリズム群を決定する段階（図 1）と、そこで選択されたアルゴリズム群の推論結果から、同じデータにおいても推定信頼度が変わる側面を捉え、最も信頼度の高い相互作用の推定結果を集約する段階（図 2）の 2 段階を経る。

実際に、図 3 に示したように、アルゴリズムごとにどの相互作用の推定信頼度が高くなるかは、顕著に違いが現れる。

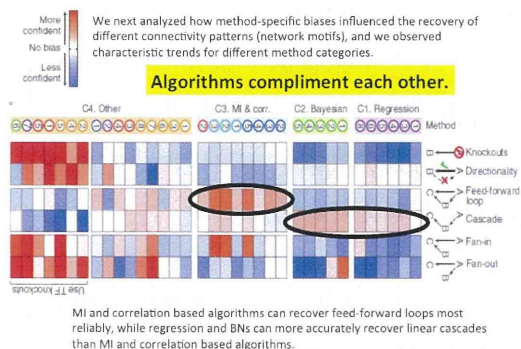


図 3 : 推定された相互作用の信頼度マップ

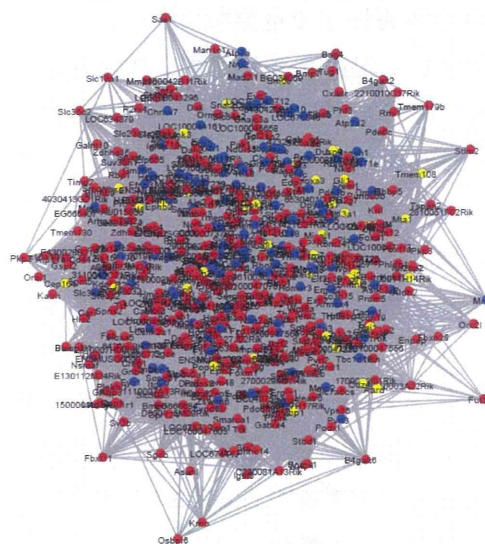


図 4 : 推定された遺伝子制御ネットワーク

また、これらの成果は、Inference cloud という Web-Service の形で実装し、オンライン上で利用可能な形態にした（図 5）。

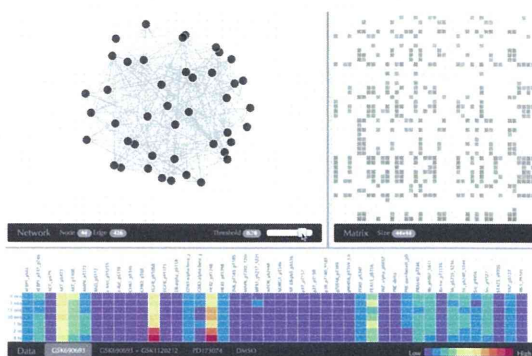


図 5 : Inference Cloud の画面例

このような手法開発は、継続的に改良を重ねていくことが重要であり、今後とも研究は継続される。しかし、ここまでの成果で、この方式の利点と実用性に一定の目処が立ったと考えている。

2: 大規模データからの状態制御遺伝子群推定アルゴリズムの開発と改良

システム毒性学で重要なことは、ネットワークとして記述された生体システムが、毒性物質の影響でどのように「制御」されてしまうかを理解すること（フォワードアナリシス）であり、また、ネットワークを特定の状態に遷移させた原因は、どの遺伝子群への影響によるものであるかを特定できるということ（バックワードアナリシス）である。フォワードアナリシスは、研究項目1と3で実現する。この研究項目では、ネットワークを制御することができる遺伝子群の同定を大規模データから行う手法の研究を行った。

我々は、すでにプリオン病のデータを利用して、遺伝子相互作用ネットワークの分析から、プリオン病と同等の遺伝子発現変化を生じさせる少数のドライバー遺伝子を同定し、その効果をシミュレーションで確認している。ここでは、前項で開発された遺伝子制御推定システムの推論結果をもとに、ネットワーク制御推定アルゴリズムを適応してその効果を確認するとともに手法の改良を行った。

特に大規模なネットワークを扱うことを考慮して、個別遺伝子の制御可能性の推定に加え、多くの遺伝子を群として扱い、その遺伝子群の関係において制御可能性を

推定するという手法を開発した（図6）。これにより、従来より大規模なネットワークの制御可能性の解析が可能となった。

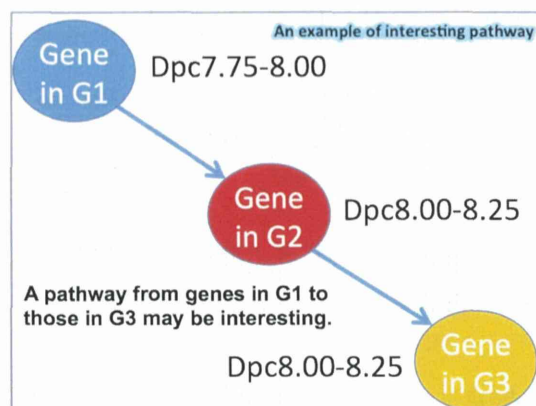


図6：遺伝子群レベルでの制御可能性推定

さらにプロモーター領域の解析も付加するシステムSHOEを開発し、ゲノムデータの裏付けを基に制御ネットワークの推定を行うシステムも開発した。図7には、そのシステムの解析結果の可視化として、特定の遺伝子群から別の遺伝子への制御可能性分布を、図8には、転写因子から遺伝子群への制御可能性分布を可視化したものを示している。この過程において、全データからの用量選択、細胞選択、時計遺伝子効果削除、スペクトラルクラスタリングや主成分分析の表示の上で6つのアルゴリズムによるクラスタリング等の機能を加え、より正確な状態制御推定を可能とした。

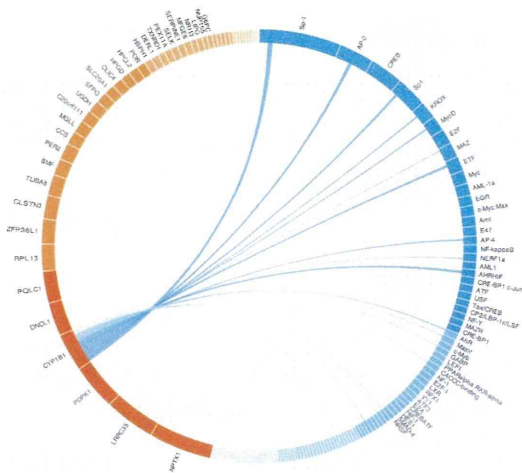


図 7：特定の遺伝子群から別の遺伝子群への制御可能性分布の可視化

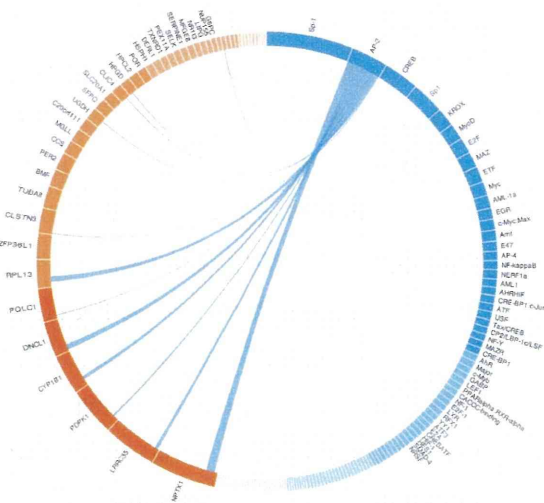


図 8：転写因子から遺伝子群への制御可能性分布の可視化

3: 大規模文献情報などの集約手法と伝達手法の研究

フォワードアナリシスでは、大規模データのみならず、大量の文献やデータベースなどを総動員して精度の高い分子相互作用ネットワークと遺伝子制御ネットワークを構築する必要がある。すでに我々は、これらに關係するソフトウェアである

CellDesigner や Payao などを構築して、このプロセスの効率化と高精度モデルの構築ノウハウを蓄えてきた。この蓄積から、次のステップとして大量に存在する文献などの情報を、知的に集約し、モデル構築を行う研究手法などを提示し、可能な範囲で自動的にモデルを構築する情報を抽出する手法の開発が必要であると結論している。これに関しては、文献情報からパスウェイ構築を効率化するために、パスウェイエディタ CellDesigner を改良し、パスウェイの各パスをモジュール化して扱えるように、グラフィックデータベースにモデルを格納する Ver. 5 の試作を行ない、基本的なソフトウェア構成を確認し、大量データを扱う実行速度や安定性などが確認された。

4: Garuda Platform に準拠する毒性解析ソフトウェアの開発

本研究で開発される手法を実装するソフトウェアや Percellome などは、国内外の幅広い研究者に利用されることで、トキシコロジーに關連する研究の発展とその成果の利用を促進する必要がある。我々は、国際的なソフトウェアの共通基盤として Garuda Platform を開発している。ここでは、この Garuda Platform をトキシコロジーの分野へと応用することを可能とする一連の開発と普及・啓蒙を行った。

従来手法では捕え切れなかったネットワークを基盤とした毒性の理解を可能とするために、生体活動をシステムとして把握するシステムバイオロジーをトキシコロジーに応用した「システムトキシコロジー」

解析基盤を開発し、その成果をライフサイエンスソフトウェアの国際的共通基盤 (Garuda Platform) を介して国内外の研究者に提供し、国際的な研究活動を促進した。Garuda Platform は、国際的に非常に高い評価を得ており、普及は予想以上に順調に進んでいる。実際に、武田薬品工業株式会社が、2014年度から導入するなど、普及が加速している。また、日本や欧州の製薬会社、化粧品会社、食品会社が大きな興味を示しており、本研究のインパクトは大きなものがあると感じている。

D. 考察

毒性学にシステムバイオロジーを適用する一連の手法が具体的な形となっており、複数のソフトウェアの連動が実現されつつある。特に、Garuda は、今後極めて重要なプロジェクトに成長すると思われる。ソフトウェアを開発する側の研究者のみならず、ユーザとなる製薬企業や生命科学の基礎研究に携わるものにとっても、重要な Platform であるとの認識が広まっている。この Garuda Platform にシステム毒性学の研究を可能とするソフトウェア群を実現する事は、今後の研究の展開にとって非常に重要であると考えられる。

E. 結論

本研究は、システムバイオロジーを毒性学に応用しようとするものであり、そのために一連の方法論とソフトウェアを開発し、応用するものである。現在、この目標に対して、順調に展開している。すでに幾つかのソフトウェアが、Garuda Platform 上に実装され連動するなどの成果をあげている。

今後、これらのソフトウェアを利用した具体的適用例を作る事や普及などの展開にも力を入れ、実用レベルでのシステム開発と世界的な普及に力を入れたい。

F. 研究発表

(1) 論文発表

1. Maria Marti-Solano, Ewan Birney, Antoine Bril, Oscar Della Pasqua, Hiroaki Kitano, Barend Mons, Ioannis Xenarios and Ferran Sanz. Integrative knowledge management to enhance pharmaceutical R&D. *Nature Reviews Drug Discovery*. 13, 4, 239-240, Apr. 1, 2014.
2. Yoshiyuki ASAI; Takeshi ABE, Hideki OKA, Masao OKITA, Ken-ichi HAGIHARA, Samik GHOSH, Yukiko MATSUOKA, Yoshihisa KURACHI, Taishin NOMURA, Hiroaki KITANO. A Versatile Platform for Multilevel Modeling of Physiological Systems: SBML-PHML Hybrid Modeling and Simulation. *ABE Advanced Biomedical Engineering*. 3, 50-58, May 17, 2014.
3. Tokiko Watanabe, Eiryō Kawakami, Jason E. Shoemaker, Tiago J.S. Lopes, Yukiko Matsuoka, Yuriko Tomita, Hiroko Kozuka-Hata, Takeo Gorai, Tomoko Kuwahara, Eiji Takeda, Atsushi Nagata, Ryo Takano, Maki Kiso, Makoto Yamashita, Yuko Sakai-Tagawa, Hiroaki Katsura, Naoki Nonaka, Hiroko Fujii, Ken Fujii, Yukihiro Sugita, Takeshi Noda, Hideo Goto, Satoshi Fukuyama, Shinji

Watanabe, Gabriele Neumann, Masaaki Oyama, Hiroaki Kitano, and Yoshihiro Kawaoka. Influenza Virus-Host Interactome Screen as a Platform for Antiviral Drug Development. Cell Host & Microbe, 16, 6, 795-805, Dec.10, 2014.

4. Jablonska Agnieszka and Natalia Polouliakh In silico discovery of novel transcription factors regulated by mTOR pathway activities. Frontier in Cell and Developmental Biology ISSN: 2296-634X, DOI:10.3389/fcell.2014.00023 Vol:2 No.23 June 3, 2014.

(2) 学会発表

1. Kitano, H. Frontiers of Systems Biology and Software Platform. Presentation at L'ORÉAL, L'ORÉAL Research & Innovation, Aulnay-sous-Bois, France, Apr. 22, 2014. (invited)
2. Kitano, H. Multiscale disease systems modelling. 7th Noordwijkerhout Symposium on Pharmacokinetics, Pharmacodynamics and Systems Pharmacology: "SYSTEMS PHARMACOLOGY IN DRUG DISCOVERY AND DEVELOPMENT", NH Conference Centre Leeuwenhorst, Noordwijkerhout, the Netherlands, Apr. 24, 2014. (invited)
3. 北野宏明. システム毒性学の戦略とプラットフォーム技術. 第41回日本毒性学会学術年会 シンポジウム: 毒性オミクス-遺伝子発現ネットワークを

標的とした、治療、毒性、及びそれらの評価の新動向-、神戸国際会議場, July 2, 2014. (invited)

4. Ghosh, S.; Kitano, H. Garuda: Fly to the future of biology. ISMB 2014, John B. Hynes Memorial Convention Center, Boston, USA, July 14, 2014.
5. Kitano, H. Systems Drug Discovery and Software Platform. Unlocking unique clinical research roadmaps using a systems approach, ICSB 2014, Melbourne Convention and Exhibition Centre, Sep. 15, 2014.
6. Kitano, H. Systems Biology and Applications. ICSB 2014, Melbourne Convention and Exhibition Centre, Sep. 16, 2014. (invited)
7. Kitano, H. Garuda Platform: An integrated software solution for data-driven medical sciences. World Health Summit 2014, Federal Foreign Office, Berlin, Germany, Oct. 20, 2014. (invited)

G. 知的財産所有権の出願・登録状況 (予定も含む)

1. 特許取得
なし
2. 実用新案登録
なし
3. その他
なし

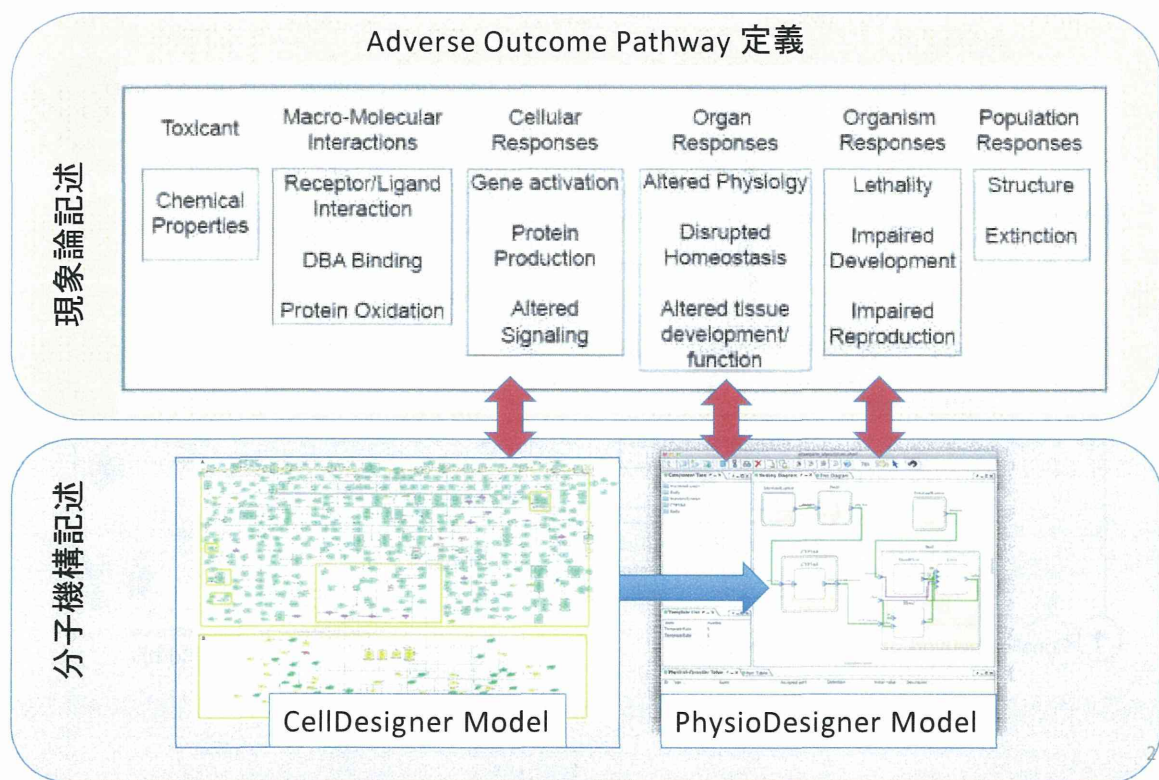
Next Steps

The Systems Biology Institute

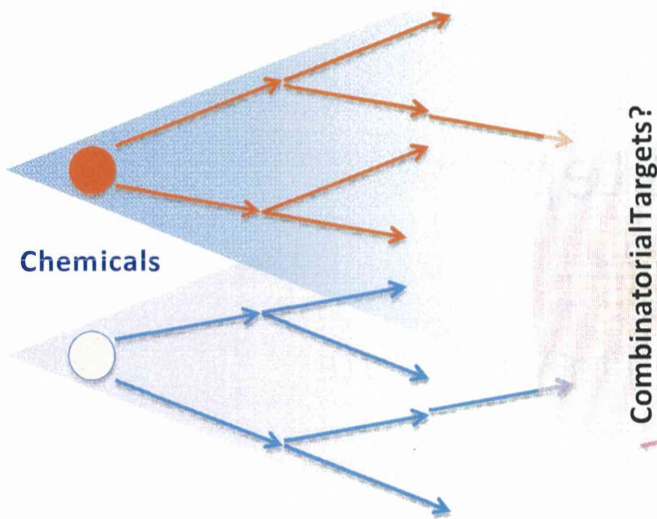
13 Feb. 2015

1

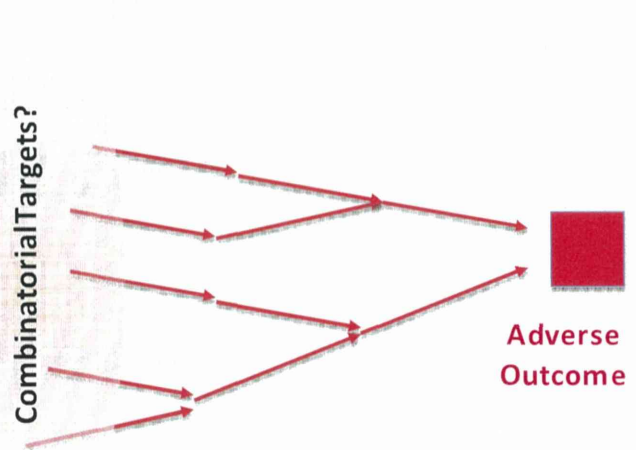
Integration with APO



Forward Effect Analysis

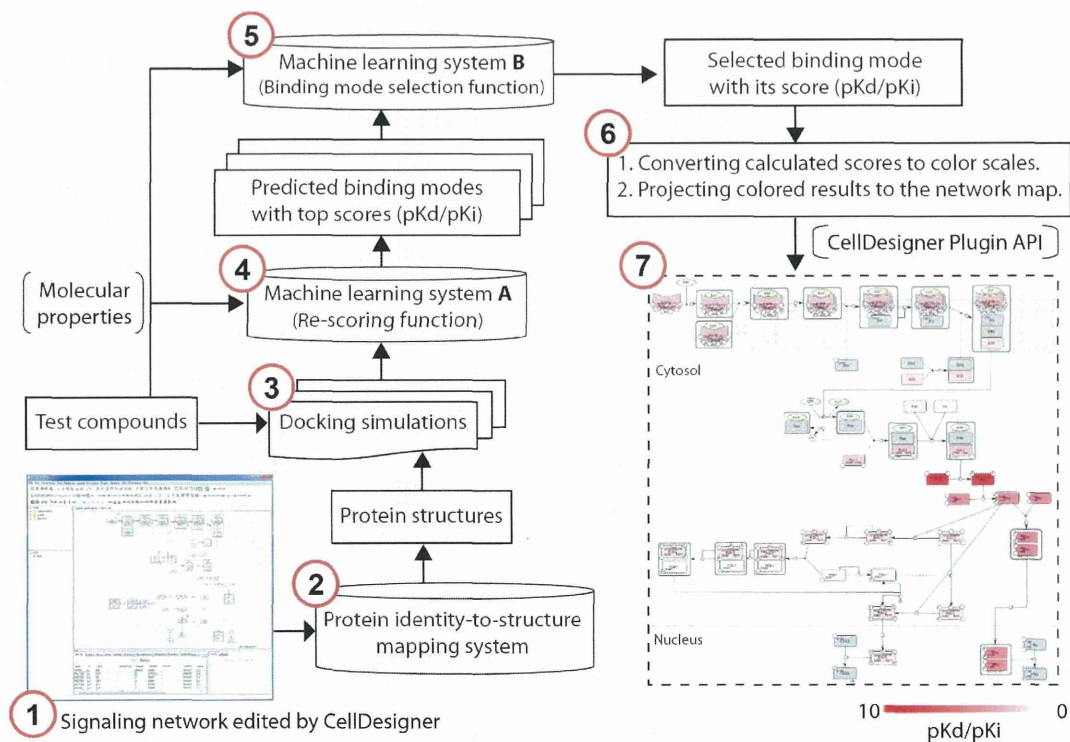


Backward Effect Analysis



3

Network-based screening pipeline



4