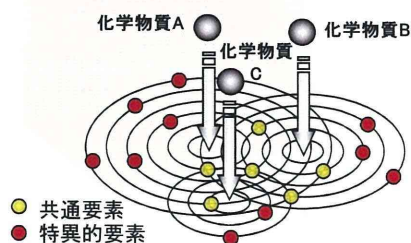




RSortの改良(要点)

- 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。



9



RSort : false positive対策

Materials:

TTG016-L(C) Surface DataSet

Matrix: 4x5

全プローブ数 : 45101

RSort(Expand_G2_AP) : 4420

目視による選別 : 1255

TTG020-L Surface DataSet

Matrix: 5x5

全プローブ数 : 45101

RSort(Expand_G2_AP) : 2713

目視による選別 : 302

10



RSort : false positive対策

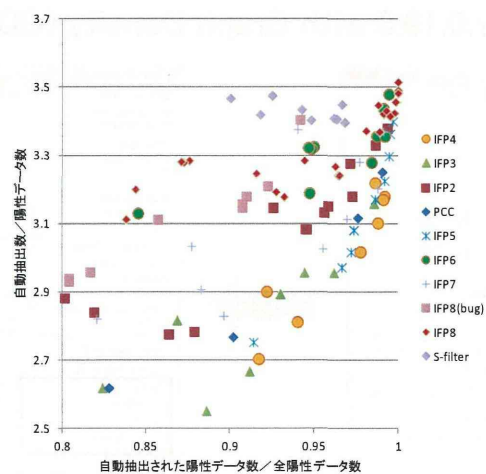
検討したアルゴリズム

- 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)

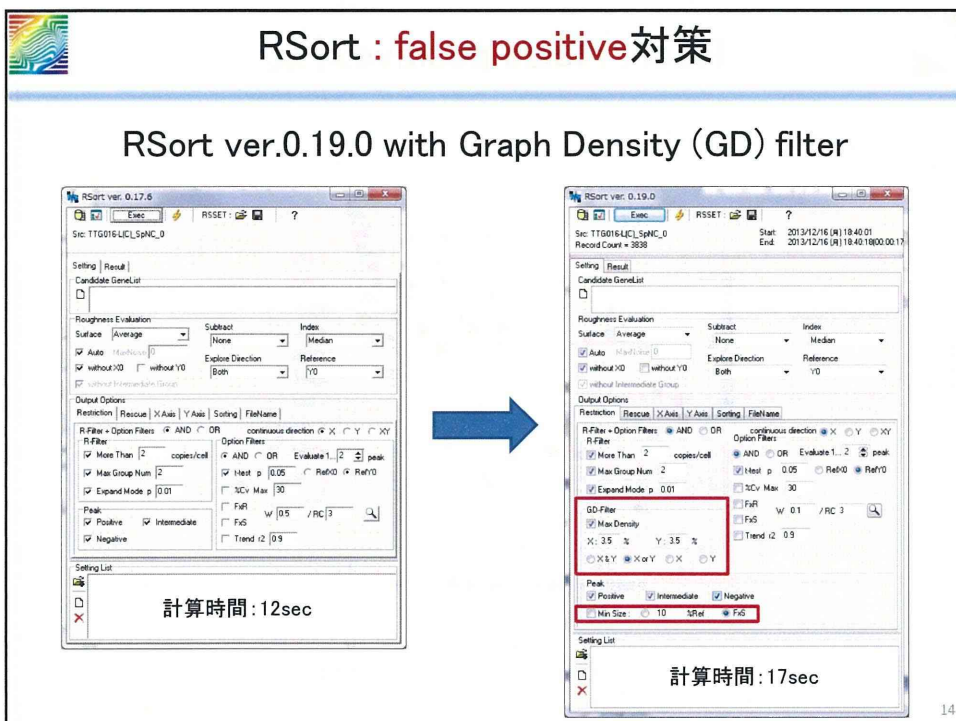
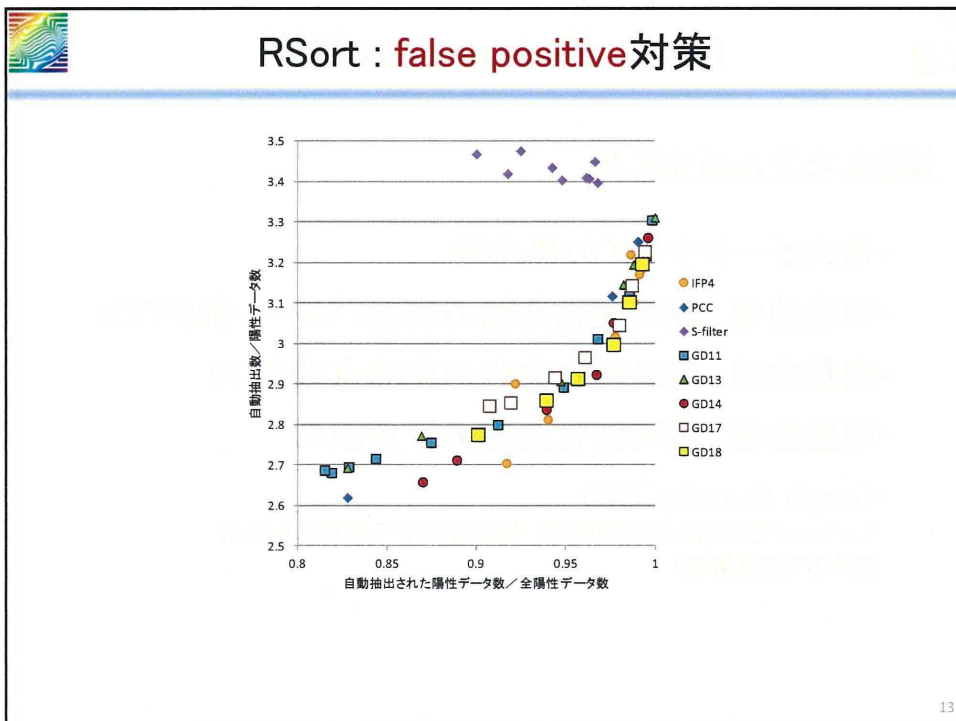
11



RSort : false positive対策



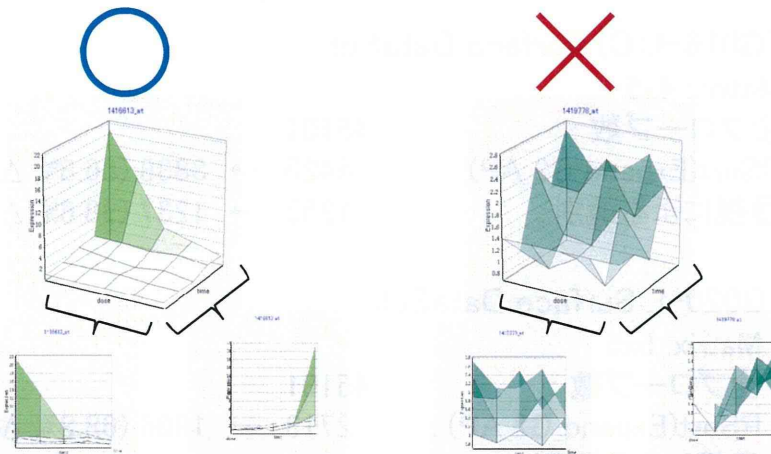
12





RSort : false positive対策

Graph Density (GD) filter

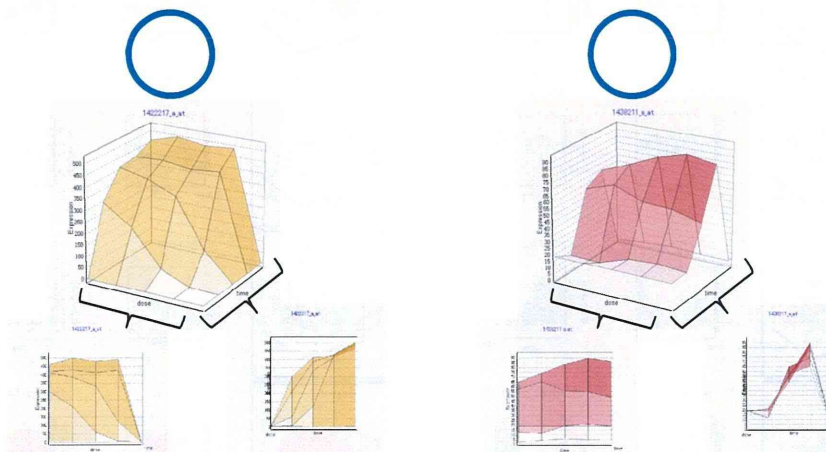


15



RSort : false positive対策

Graph Density (GD) filter



16



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

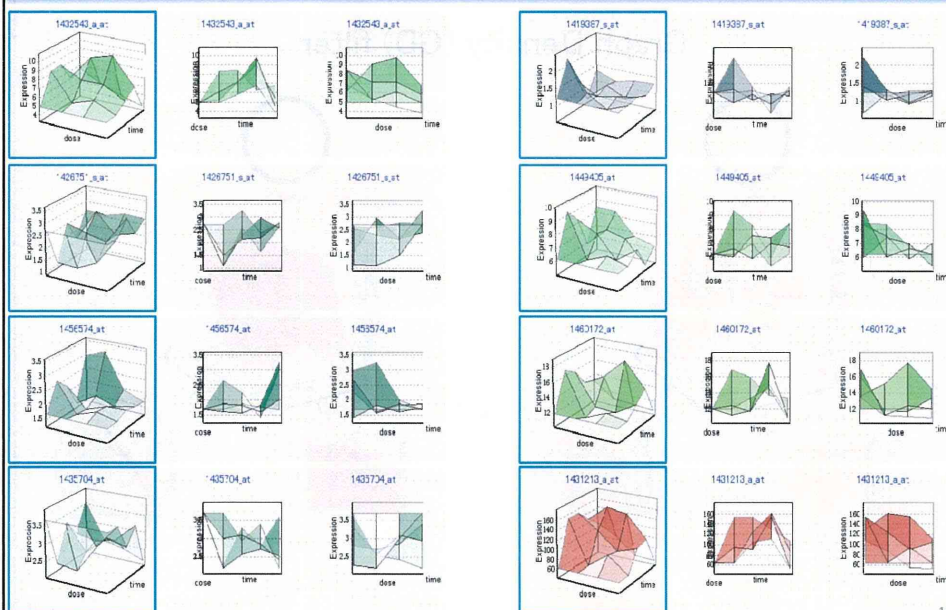
RSort(Expand_G2_AP) : 2713 → 1866 (68.8%, Δ847)

目視による選別 : 302 → 292 (96.7%, Δ10)

17



TTG016-L(C)におけるGD filterの過剰カット

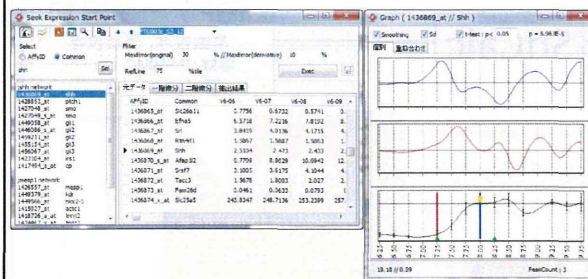


18

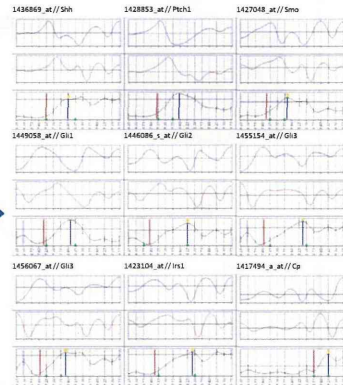


微分解析による発現変動起点の検出

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



19



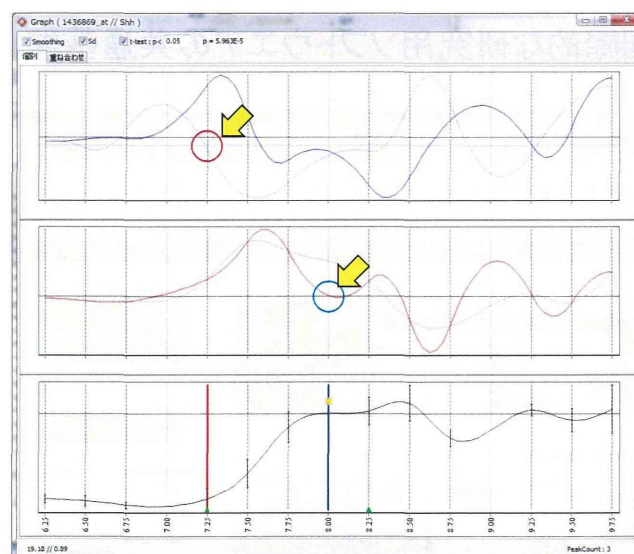
微分解析による発現変動起点の検出

二階微分

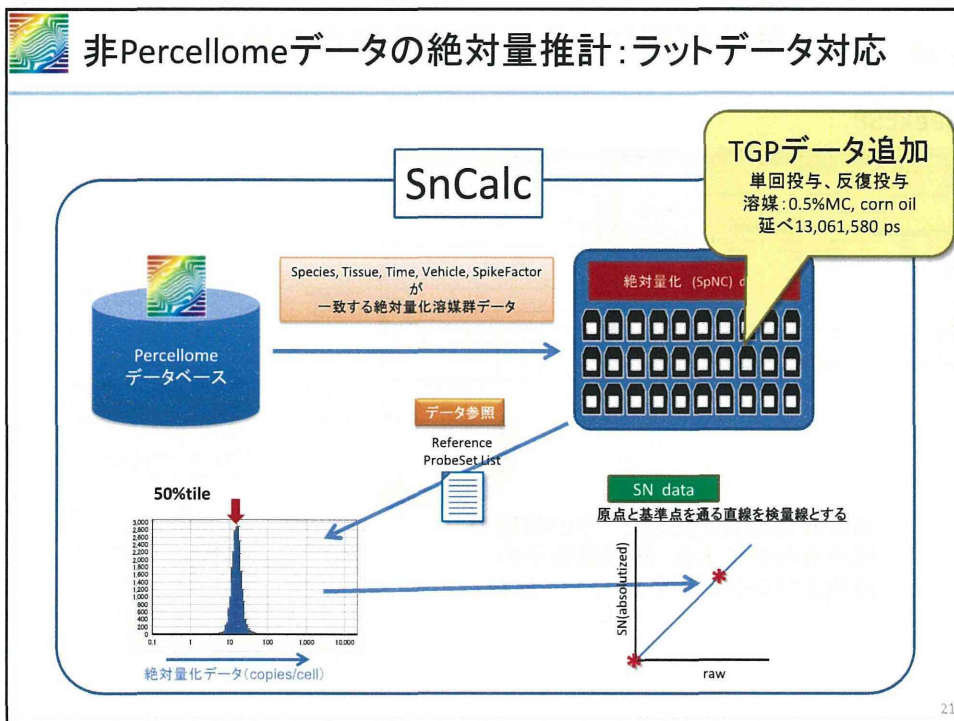
三階微分

一階微分

元データ



20



SBI The Systems Biology Institute

Garuda Alliance

国際的な研究用ソフトウェアの共通基盤

Garuda Alliance
THE WAY BIOLOGY CONNECTS

Home Overview Garuda Alliance Members Garuda Resources User Stories News/Topics Events Developer Center

About Garuda Alliance

User Stories

News/Topics

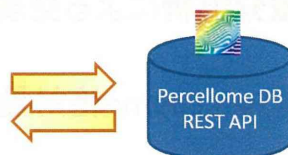
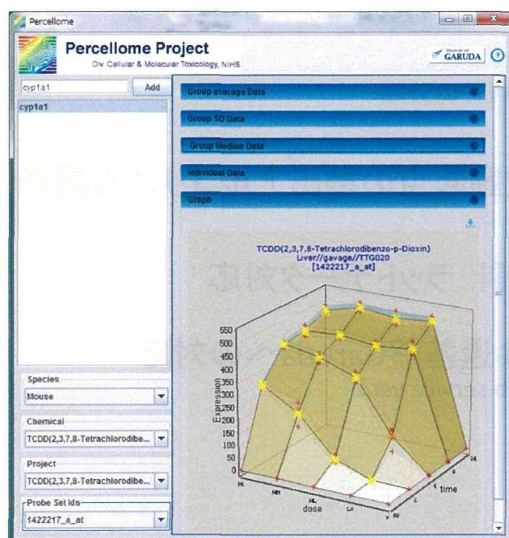
22



Percellome Gadget with 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.nihs.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



23



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>

24



H25年度の成果

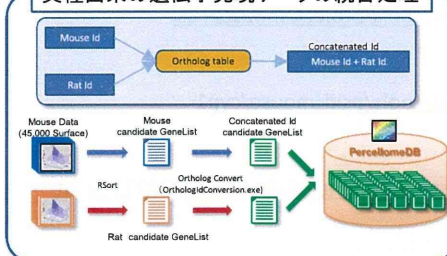
- RSortの改良
自動抽出の性能向上
溶媒影響(概日リズムを含む)の分離
- 微分解析による発現変動起点の検出アルゴリズム開発
- 非Percellome絶対量推計:ラットデータ対応
- 国際的なソフトウェア共通基盤Garudaへの対応
REST API of PercellomeWeb service ver.2

25

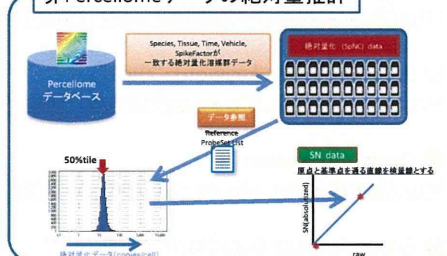


成果一覧 (H24~25)

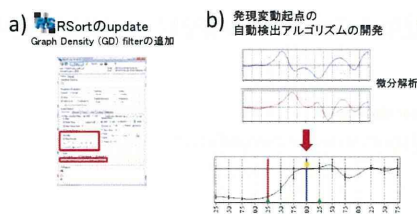
異種由来の遺伝子発現データの統合処理



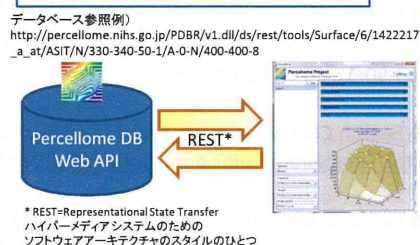
非Percellomeデータの絶対量推計



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



26



H26年度計画

- Garuda Platformの活用

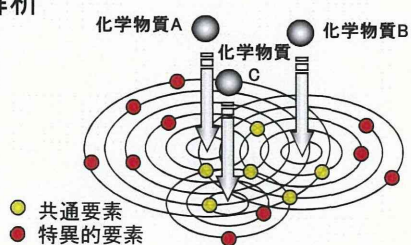
REST API of PercellomeWeb service の機能拡張

Percellome関係Gadgetの作成

→ 解析ルーチンにおける実用化

- PercellomeExplorerの機能拡張

非共通要素(特異的要素)解析



27

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
ー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ー

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する。

今年度は、遺伝子制御関係推定アルゴリズムとして、複数のアルゴリズムの併用を行う方式を開発し、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の開発を行った。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。また 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 の開発者向けバージョンとして国際的に配布を開始した。正式版は、2014 年春を目処に公開される予定である。

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する。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

ウェアとして実現し、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する。

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

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う。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。本研究では、この手法をさらに改良し、より精度の高いものに改良する。同時に、発現の位相差、時間差、ゲノム情報などを統合し、相関関係ではなく因果関係に関するネットワークを高精度で構築可能な手法の開発を行う。

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う。

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。同時に、発病した状態に対応するネットワークを、正常な状態にもどす制御を可能とする別の数個の遺伝子を同定する手法も開発している。現在のところ

ろ、この手法は研究の初期段階であり、その計算手法や予測精度、応用範囲など、多くの研究の余地が残されている。本研究では、この手法に関する研究を進めることで、システム毒性研究における重要な手法とすることを目指す。この研究には、遺伝子制御ネットワークの変化だけでは無く、影響を及ぼすドライバー遺伝子の同定とそれに関わる転写因子や受容体の同定なども含める。また、SNPs などの違いがこのネットワークの挙動に影響を及ぼすことが分かった場合には、これらの影響を組み込む手法の開発も試みる。

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、これらに関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。この研究項目は、このプロセスに関する研究とそのコンセプトを実証するソフトウェアの試作を行っている。

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

コロジーに関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・啓蒙を行う。

従来の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」解析基盤を開発する。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する。

H25 年度は、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して、状態制御遺伝子群 (ネットワークを制御することができる遺伝子群) を大量データから同定する手法を開発する。

H26 年度において、大規模データからの遺伝子制御関係抽出及び状態制御遺伝子群推定により、Percellome トキシコゲノミクスデータベース等の大規模データベースを対象としたネットワーク情報の抽出ならびこれらデータの応用の手法を検討する。

C. 研究成果

遺伝子制御関係推定アルゴリズムでは、複数のアルゴリズムの併用を行う方式を開発し、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ている。さらに、より一層の大規模化に備え、この性能を維持しつつ、情報エントロピーを利用しながら計算量を減少させる手法

の開発に成功した。また、これらを 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1)。

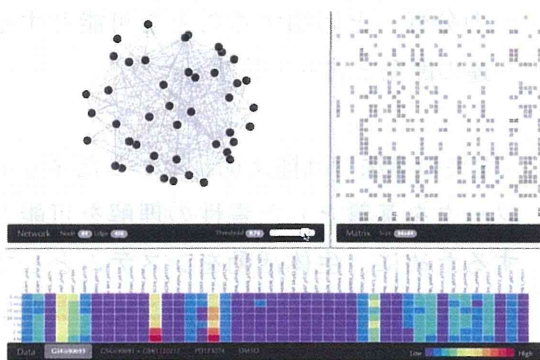


図1： Inference Cloud の画面例

状態制御遺伝子群推定アルゴリズムに関しては、以前から開発している A Gene geometric Clustering Tool (AGCT) で Percellome データの特徴を考慮に入れた解析を容易に実行できるようにするための開発を行った。具体的には、全データからのドーズ選択（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追加した。これらの機能では Percellome データのニーズにそくし、事前データ削除を行わずに全データを計算した後、必要な部分だけをユーザが取り出せるようになっている。更に大規模データをより短時間で計算できるようにした。

制御遺伝子群の同定についての予備検討では、幾つかの事例において、実験的に確認されている遺伝子群と合致することなどを確認した。

さらに、Garuda Platform の開発者向けパッケージβ2版の配布開始（2013年12月）にあわせ、ネットワークの推定や

AGCT など一連のアプリケーションを Garuda に準拠し、毒性学に利用できるように対応したほか、オルソログ制御領域分析ソフトウェア SHOE（Sequence Homology in Higher Eukaryotes）に関してはオンラインで Garuda 経由でサービス提供できるように準備を進めた。また Percellome DB の Garuda 上への実装に協力し、Percellome DB の国際的普及への準備を行った。

制御関連遺伝子の推定技術開発では実用レベルのプログラム開発を行い、解析研究を推進した。また遺伝子制御関係推定では、実際の網羅的トランスクリプトームデータを用いて最適化を進めた。さらに Percellome DB 対応の Garuda ガジェットプログラムを作成・改良し、配布を開始するなど、その国際的普及活動を進めた。また、遺伝子クラスター解析手法の開発に関しても、開発中の AGCT (A Geometric Clustering Tool) の Garuda 対応が完了している。

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年度からの導入を決めるなど、普及が加速している。

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。

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たい。

F. 研究発表

(1) 論文発表

Hase T, Ghosh S, Yamanaka R, Kitano H. (2013) Harnessing Diversity towards the Reconstructing of Large Scale Gene Regulatory Networks. *PLOS Computational Biology*. 9(11): e1003361, doi:10.1371/journal.pcbi.1003361.

Matsuoka Y, Matsumae H, Katoh M, Einfeld AJ, Neumann G, Hase T, Ghosh S, Shoemaker JE, Tiago Lopes, Watanabe T, Watanabe S, Fukuyama S, Kitano H, Kawaoka Y. (2013) A comprehensive map of the influenza A virus replication cycle. *BMC Systems Biology*. 7, 97, doi:10.1186/1752-0509-7-97.

Yamashita F, Sasa Y, Yoshida S, Hisaka A, Asai Y, Kitano H, Hashida M, Suzuki H. (2013) Modeling of Rifampicin-Induced CYP3A4 Activation Dynamics for the Prediction of Clinical Drug-Drug Interactions from In Vitro Data. *PLOS ONE*. 8, 9, e70330.

Fujita K, Ostaszewski M, Matsuoka Y, Ghosh S, Glaab E, Trefois C, Crespo I, Perumal TM, Jurkowski W, Antony PMA, Diederich N, Buttini M, Kodama A, Satagopam VP, Eifes S, Del Sol A, Schneider R, Kitano H, Balling R. (2014) Integrating Pathways of Parkinson's Disease in a Molecular Interaction Map. *Molecular Neurobiology*. 49:88-102, DOI 10.1007/s12035-013-8489-4.

Naito T, Yatsuhashi A, Kaji N, Ando T, Sato, K, Moriya H, Kitano H, Yasui T, Tokeshi M, Baba Y. (2013) Parallel Real-Time PCR on a Chip for Genetic Tug-of-War (gTOW) Method. *Analytical Sciences*. 29(3): 367-71.

Makanae K, Kintaka R, Makino T, Kitano H, Moriya H. (2013) Identification of dosage-sensitive genes in *Saccharomyces cerevisiae* using the genetic tug-of-war method. *Genome Research*. 23, 300-311.

Schaefer MH, Lopes T, Mah N, Shoemaker JE, Matsuoka Y, Fontaine JF, Louis-Jeune C, Einfeld AJ, Neumann G, Perez-Iratxeta C,

- Kawaoka Y, Kitano H, Andrade-Navarro MA. (2013) Adding Protein Context to the Human Protein-Protein Interaction Network to Reveal Meaningful Interactions. *PLOS Computational Biology*. 9, 1. doi:10.1371/journal.pcbi.1002401
- Polouliakh N (2013) Reprogramming resistant genes: in-depth comparison of gene expressions among iPS, ES and somatic cells. *Frontier of Physiology* doi:10.3389/fphys.2013.00007.
- (2) 学会発表
- Kitano, H. Systems drug discovery and Neuro-degenerative diseases. CHDI's 8th Annual Huntington's Disease therapeutics Conference: A Forum for Drug Discovery & Development, Molino Stucky Hilton Hotel, Venice, Italy, Apr. 9, 2013. (invited)
- Kitano, H. Systems Biomedicine for Drug Discovery and Personalized Healthcare. Seminar at Institute for Infocomm Research(I2R), Institute for Infocomm Research(I2R), Singapore, Apr. 12, 2013. (invited)
- Kitano, H. Systems Biology: past, present, future. Talk at AIBN at the University of Queensland, Australian Institute for Bioengineering and Nanotechnology, University of Queensland, Brisbane, Australia, May 9, 2013. (invited)
- Kitano, H. Systems biology in the context of systems and precision engineering. Talk at Australian Institute of Marine Science(AIMS), AIMS, Townsville, Australia, May 10, 2013. (invited)
- 北野宏明. システムトキシコロジープラットフォーム. 第40回日本毒性学会学術年会 シンポジウム7「毒性オミクス」, 幕張メッセ, 千葉, June 18, 2013. (invited)
- 北野宏明. 低レベル被曝リスクに関する議論における問題点の検証とリスク評価の困難さ. 第40回日本毒性学会学術年会 シンポジウム10「放射線毒性学における課題」, 幕張メッセ, 千葉, June 19, 2013. (invited)
- Kitano, H. Systems biology and systems biomedicine: integrative systems sciences and biomedical sciences. 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Osaka International Convention Center, Osaka, July 7, 2013. (invited keynote)
- Kitano, H. Systems Biology Platform for Drug Discovery. Talk at Imperial College, London, UK, July 30, 2013. (invited)
- Kitano, H. Systems Biology and Software Platform. Nature Publishing Group Asian Academic Publishing team meeting, Sheraton Miyako Hotel, Sep. 6, 2013. (invited)
- Kitano, H. Garuda Platform: An integrated inter-operability for biomedical software and data resources. COMBINE 2013, Institut Curie, Paris, France, Sep. 16, 2013. (invited)
- 北野宏明. システムバイオロジーを活用

したバイオマーカー評価. 日経バイオテク/日経バイオテク ONLINE プロフェッショナルセミナー「実例から学ぶクリニカルバイオマーカー時代の到来」, コクヨホール, Oct. 2, 2013. (invited)

2. 実用新案登録
なし

3. その他
なし

Kitano, H. Systems Drug Discovery and Systems Biomedicine. Systems Biology Seminar co-hosted by Garvan Institute and UNSW, Garvan Institute, Sydney, Oct. 11, 2013. (invited)

北野宏明. システムバイオロジーを活用したバイオマーカー評価. 講演@東北メディカルメガバンク機構. Nov. 1, 2013. (invited)

北野宏明. Systems drug discovery and its software platform システム創薬とソフトウェアプラットフォーム. セミナー@武田薬品工業株式会社 湘南研究所, Dec. 3, 2013. (invited)

北野宏明. 多階層生体機能学: モデル創薬から臨床推論サポートまで. 金沢医科大学 第15回KMU研究推進セミナー, 金沢医科大学病院, Dec.4, 2013. (invited)

北野宏明. Current Status and Perspectives on Systems Drug Discovery. システム創薬の現状と今後の展望. アスビオファーマ社内セミナー, アスビオファーマ株式会社, Dec. 24, 2013. (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



GARUDA

THE WAY BIOLOGY CONNECTS

1

Garuda Alliance (snapshot May 2013)



Percellome Project
Div. Cellular & Molecular Toxicology, NIHS



PhysioDesigner

CellDesigner.org

2

Software problems

- Software for biomedical research is the critical components for success of research
- Nobody can develop entire software systems alone
- However
 - Tools are developed independently
 - Different GUI, different operating procedure, different APIs, etc.
 - Need to launch tools independently
 - No direct data sharing, etc
- Inter-operability is missing!!!!
- Extra work needed for users and developers

3



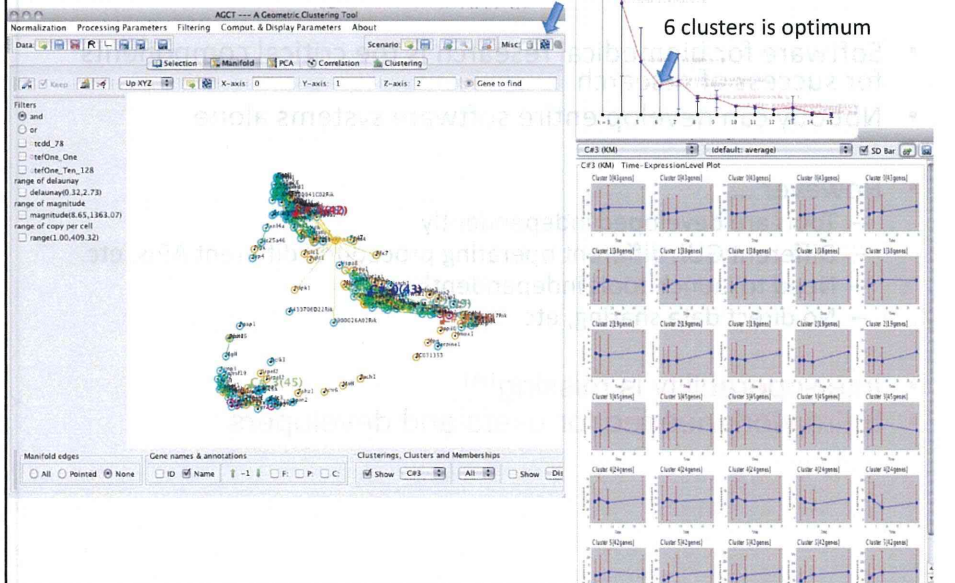
102

Percellome data clustered

Clustering 212 selected probes, which are three types of labels.
TCDD only (78). TCDD-TCDF (TEF=1.0, 6), TCDD-TCDF (TEF=0.1, 128)

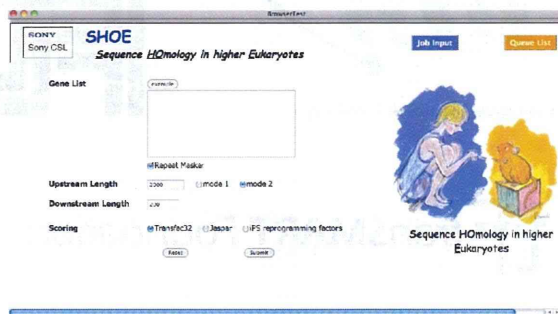
Save clusters on Desktop

6 clusters is optimum



SHOE

Interface



Sony CSL Server

