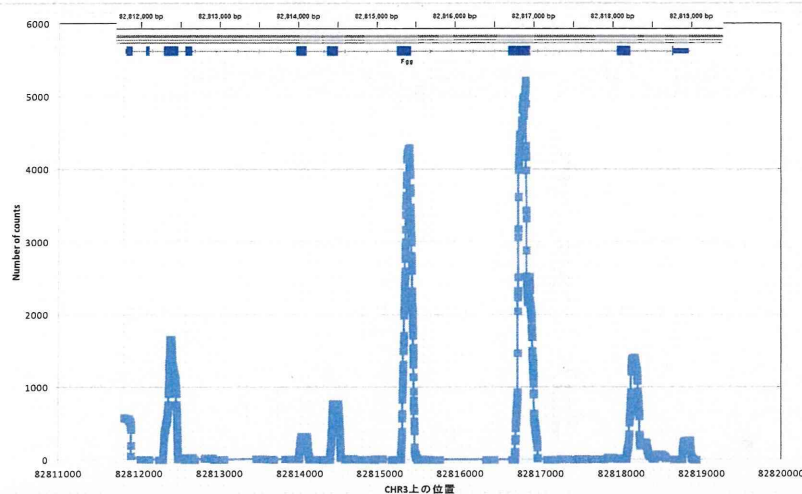
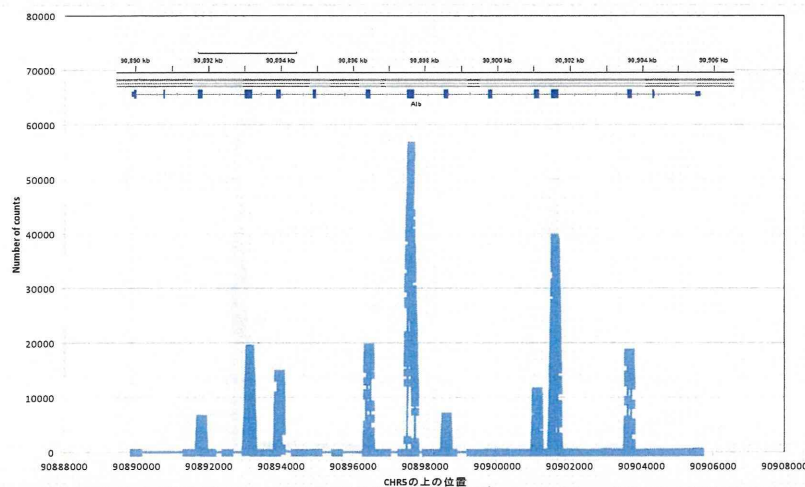


4.5.7.Fggの割り付け結果



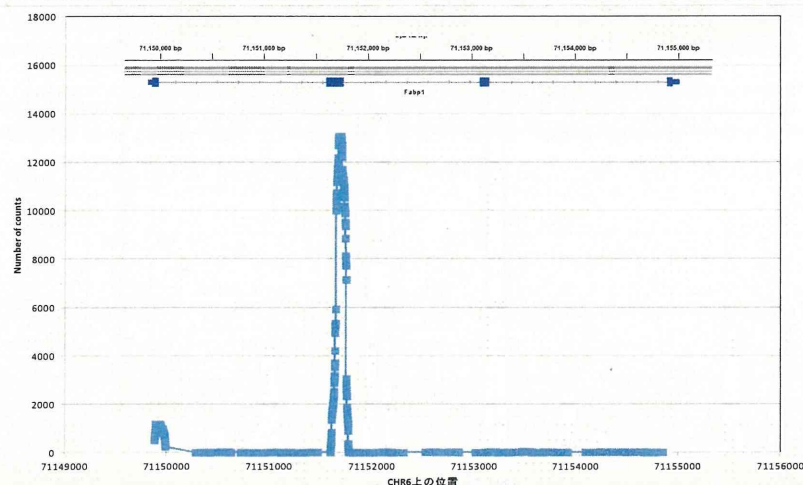
長いエクソンほどピークが高くなっている

4.5.8.Albの割り付け結果



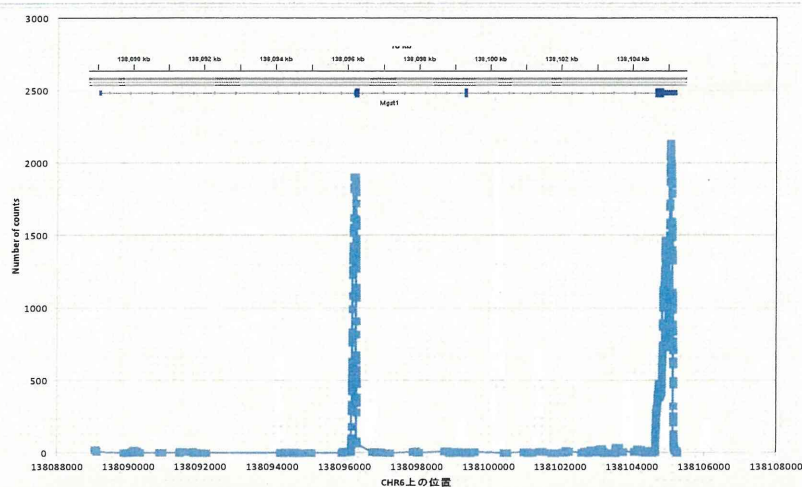
他の遺伝子より桁違いに多い。これが現実であり、マイクロアレイでは飽和が発生し、濃度を推定しにくい遺伝子と思われる

4.5.9.Fabp1の割り付け結果



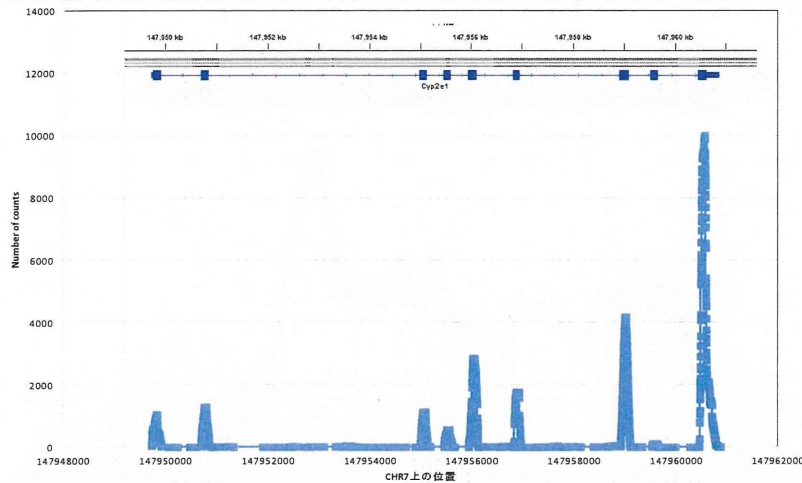
中央の飛びぬけて大きな値を示すエクソンは、長さがあり、割り付けが成功したと思われる

4.5.10.Mgst1の割り付け結果



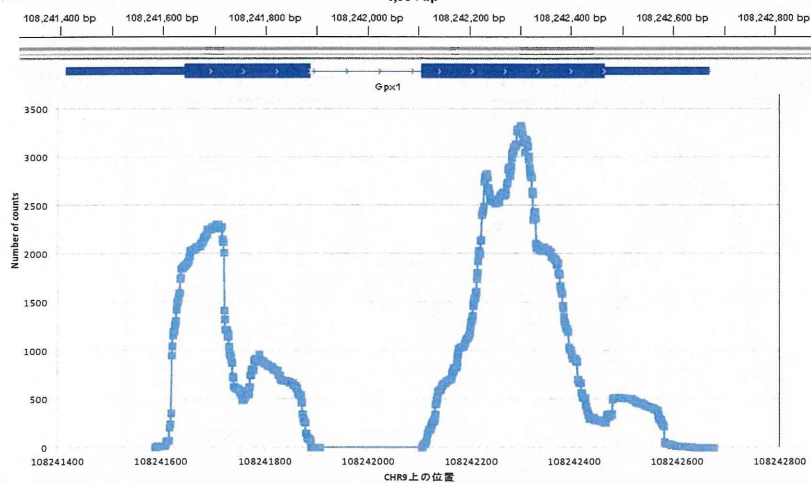
二つのピークとなっているエクソンは、長さがあり、割り付けが成功したと思われる

4.5.11.Cyp2e1の割り付け結果



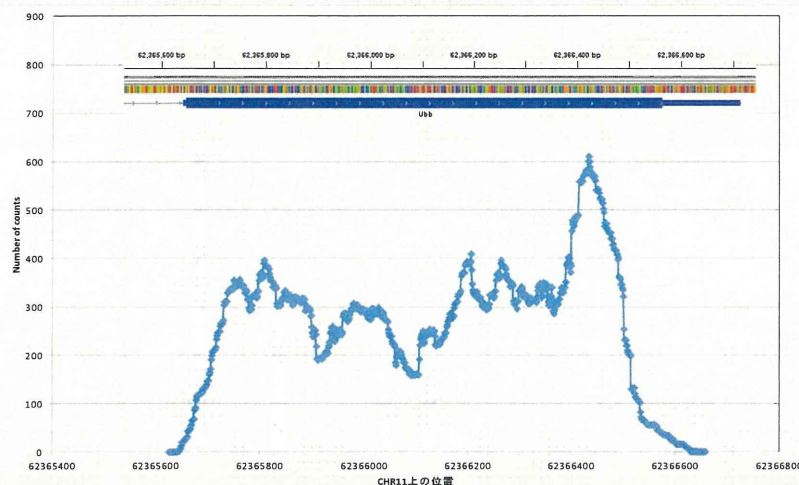
右側のエクソンは、長さがあり、割り付けが成功したと思われる

4.5.12.Gpx1の割り付け結果



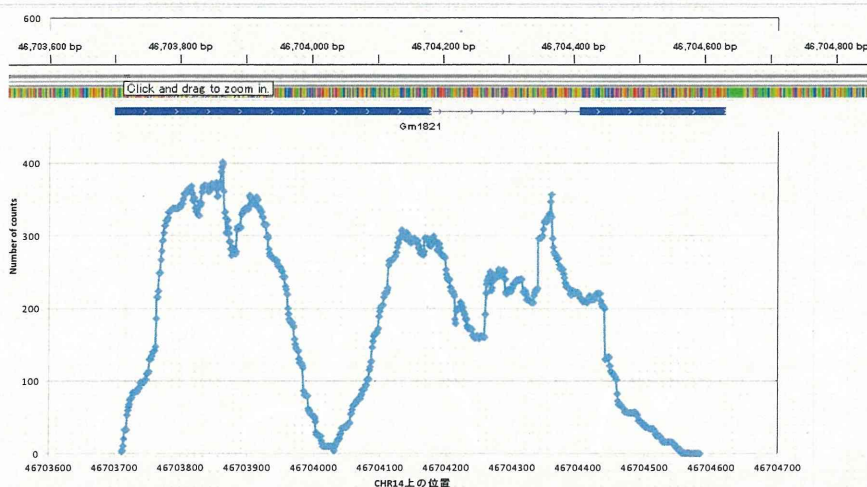
エクソン中に左側にピークがあるのは、ペアエンド法を実施しているためと考えられる。割り付けアルゴリズムも対応させるべきである

4.5.13.Ubbの割り付け結果



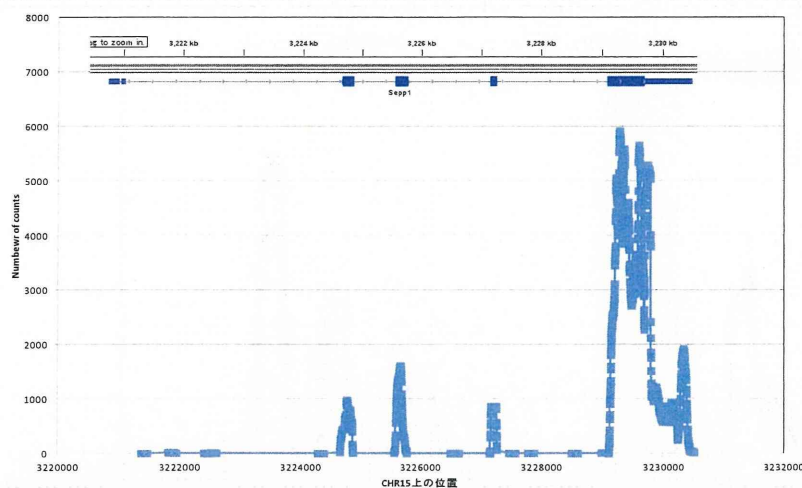
単一のエクソンで構成される遺伝子で、エクソンが長いため、うまく割り付けできていると思われる。

4.5.14.Gm1821の割り付け結果



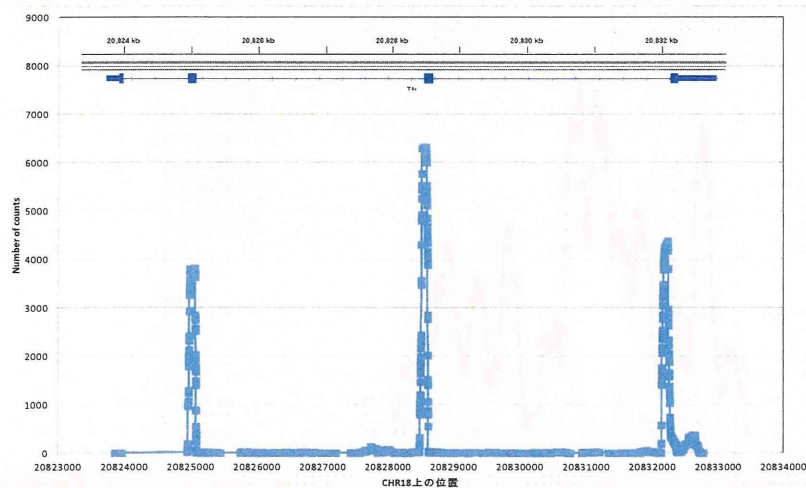
イントロン部分での割り付け量が多い。複数箇所へ割り付け可能としていることの影響かもしれない。

4.5.14. Sepp1 の割り付け結果



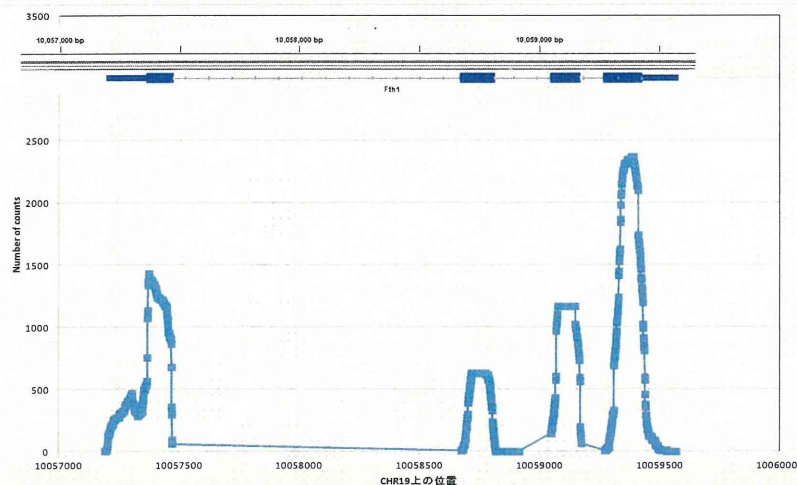
右側の長いエクソン部分はうまく割り付けできていると思われる。

4.5.15. Ttr の割り付け結果



各エクソンが適度な長さがあり、割り付けがうまくいっていると思われる。

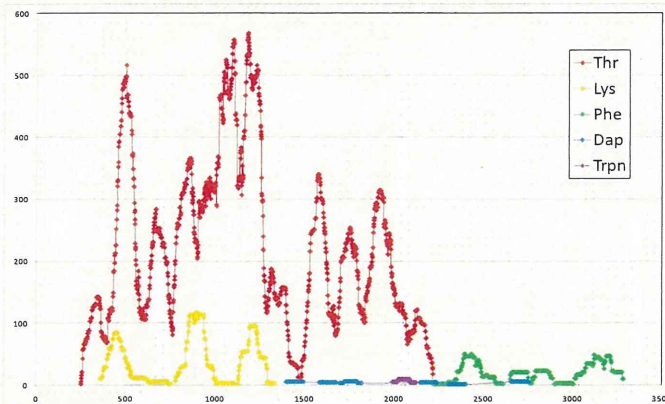
4.5.16.Fth1の割り付け結果



各エクソンが適度な長さがあり、割り付けがうまくいっていると思われる。

4.5.17.GSC-RNAの発現量

GSCの5種類のRNAの割り当てを確認した



同一RNAの中で、配列位置により、計測されたタグの数が異なっている。分配係数で補正可能なものか？何らかのバイアスが発生していないか？ → 今後検討する

4.6.アライメント試行のまとめ

以下にあげる現象が見つかり、本格的なアライメント用アルゴリズムのためには対処が必要である

- 1 長いエクソンは割当量が多くなり、短いエクソンは割当量が少なくなる

読み込まれたタグの100塩基完全一致を行っている

スプライシングの影響が考えられるので、スプライシングに対応したアルゴリズムが必要

- 2 3'末端側の割り当て量が多くなる場合がある。

計測はペアエンド法を用いている

ペアエンド法に対応したアルゴリズムが必要

- 3 インロン部分で大量の割り当てが存在した

単一のタグが複数個所に割り当てるようにした

複数個所の割り当てを適切に分配させるアルゴリズムが必要

厚生労働科学研究補助金（化学物質リスク研究事業）
「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
ムの実用化の為にインフォマティクス技術開発-」
(H21-化学-一般-001)
分担研究報告書

インフォマティクス開発研究

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

本研究は毒性分子メカニズムに依拠した毒性予測評価システムの実用化の最終段階としてのインフォマティクス技術開発を目的とし、超多次元大規模データから初期反応に関わる遺伝子発現ネットワークを同定する手法の開発に向け、実データを用いた基本推計手法、とりわけ独自のクラスタリングプログラムAGCTの開発と最適化を進めた。さらに共発現する遺伝子が同じ転写因子による制御を受けている可能性を検討するために、オルソログ制御領域の解析を行うプログラムSHOEを開発した。これらの成果により、網羅性を維持しつつアーティファクトを回避し、生物学的意味を持つ遺伝子発現変動のハイスループット抽出を中心とした解析技術の開発に目処がついた。一方、より精密に個々の遺伝子間の相互作用関係の推定を行う手法として、複数の異なる推定方式から仮説群を求め、最も可能性の高い仮説の抽出を繰り返し実施して、一連の仮説を最も確からしい相互作用ネットワークとするアルゴリズムを考案した。試計算に拠れば、現在知られている方法の中で、本法が最も正確に相互作用を推計できることを確認した。

A. 研究目的

本研究は、先行実施された化学物質リスク評価の基盤整備としてのトキシコゲノミクス研究の成果を受け継ぎ拡充しつつ、毒性分子メカニズムに依拠した毒性予測評価システムの実用化の最終段階としてのインフォマティクス技術を開発することを目的とする。

ベイジアンネットワーク等のアルゴリズムの複合適応による初期反応カスケードの描出を行う。大規模データから擾乱初期に

関わる分子標的を同定する手法の開発を行い、実際のデータを基にそのカスケードを抽出する。

B. 研究方法

大規模データから初期反応に関わる発現ネットワークを同定する手法の開発に向け、実データを用いて基本推定手法の最適化を行う。

プログラム開発は、成果としての実行ファイルの柔軟性（特定のハードウェア、オ

ペレーションシステムに束縛されない) や開発効率に優れた Java 言語を用い、研究開発の効率化・高度化を実現した。

性能評価や最適化に用いる実データとしては Percellome データベースの高精度・絶対量化されたマイクロアレイ測定データを用い、常に実用化を意識して開発を実施した。

研究の最終目標を達成するために、具体的には (1) 超多次元大規模データの解析に関する新たな方式を考案し、実データを用いた試行解析からプログラムの最適化を行うと同時に、(2) 遺伝子発現データからのネットワークの推定方法を新たに考案した。

C. 研究成果

(1) 大規模データクラスタリングからの遺伝子群間の相互作用同定

超多次元大規模データの解析に関しては、計算幾何学の手法を導入し開発した。具体的には、大規模データから特徴的なシグナル伝達系や代謝系などの活動の発見のために独自のクラスタリングプログラム AGCT (A Geometric Clustering Tool)を開発し、最終的な最適化を行った。

AGCT の処理は、4 段階で構成されている: ①時系列データの前処理、②構造計算、③クラスタリング、④出力結果フィルタリング。

①の前処理としては、多次元データの時間要素の特徴をより細かく捉えるために幾何学手法の Wavelet transform を適用する。その後、②として主成分分析 (PCA) 法やスペクトラルクラスタリング法によってデータ構造の構築計算を行う。③としては、得られた構造に対して幾つかのクラスタリング手法 (Affinity Propagation, K-means, ect) を利用して、遺伝子の発現類似度によ

ってグルーピングを行う。これらの処理により、大規模データから、遺伝子群間の相互作用関係の抽出が可能となった。クラスタリングを行った後ユーザにとって重要となるのは、全体構造の形成しているクラスター群から刺激反応を示すクラスターを選択することである。そのために今年度は、④出力結果のフィルタリングの手続きを幾つか加えた。無関係と思われる遺伝子情報を最初から取り除くのではなく、全ての情報の中から必要な情報がユーザによって決められていくコンセプトになっている。それによってデータの漏れを防ぐことができるのであろう。

また、共発現する遺伝子は同じ転写因子による制御を受けている可能性が高いので、得られたクラスター情報をもとに、今年度から開発を行っている SHOE (Sequence Homology in Higher Eukaryotes) ソフトウェアによってオルソログ制御領域の分析を行った。これにより、推定された相互作用関係の信頼度が向上する。

さらに、上記の解析手法により効率的に得られた三次元ネットワーク/クラスター構造をもとに、さらに仔細なアノテーションやネットワーク解析、シミュレーションを行うために SBML (システム・バイオロジー記述言語 sbml.org) 形式のモデルに変換する CellDesigner (パスウェイモデリングエディタ celldesigner.org) の Plugin を開発した。さらに、CellDesigner で作成された SBML モデルをオンライン上で共有、キュレーションするためのプラットフォーム Payao (payaologue.org) にもネットワークモデルを搭載、そのモデルから Percellome データベースの実データへのリンクも試験中である。

(2) 発現データからの遺伝子間相互作用の

同定

前項での解析は、大規模データから遺伝子群の間での相互作用の推定を行ったのに対して、ここでは、より精密に個々の遺伝子間の相互作用関係の推定を行う方式を開発した。従来から、遺伝子発現データからの相互作用関係の推定には、ベイズ推定や相互情報量などの数理統計手法が導入されてきた。しかし、これらの方法には、得意不得意があると同時に、その推定精度も高くは無い。今回、開発した方式は、複数の違った推定方式を一定の方法で組み合わせ、その仮説群から最も可能性の高いと思われる一連の仮説を最終的な相互作用ネットワークとして推定するものである。現在までのベンチマークでは、この方法が、現在知られている方法の中で最も正確に相互作用を推定することが出来た。

D. 考察

大規模データから初期反応に関わる発現ネットワークを同定する手法の開発に向け、H21年度より開発を継続した独自のクラスタリング技術 (AGCT) など基本推定手法及びこれを実装した解析ソフトウェアの最終的な拡張と最適化を行った。

Percellome データベースの大量の実データを用いて、性能を検証したが、例えばダイオキシン類の単回暴露実験データからは、TCDD 誘導要員として 3117 プローブ, 498 クラスタ、TCDF 誘導要員として 3711 プローブ, 369 クラスタが検出された。その中で、既知の Aryl hydrocarbon receptor pathway に関わる 350 プローブが入っているが、新しい候補遺伝子も多数上がっており、研究計画で期待した性能を実現している。

また今年度から開発を開始した、比較ゲ

ノム解析ツール SHOE (Sequence Homology in Higher Eukaryotes) や複数の手法を組み合わせた遺伝子間相互作用推定を導入することで、より一層、解析結果の精度向上見込まれており、今後さらにこの研究を継続するところで、極めて有用な手法の実用化と、毒性分野への効用が期待できる。

なお、本研究では多数の独自技術を生成しているが、それに固執せず共同研究を展開する柔軟性を維持している。現在、コミュニティによるパスウェイデータ構築を企図する研究 (PAYAO) を含む国際的なバイオインフォマティクスプロジェクト (GARUDA) との連携を検討しており、研究進捗の飛躍的な加速が期待される。

E. 結論

本研究は、分子メカニズムに依拠した毒性予測評価システムの実用化の最終段階として、毒性ネットワーク描出の為にインフォマティクス技術を開発することである。単純なバイオマーカー抽出に比し、毒性ネットワーク描出には格段に高度な解析技術が必要であって、その実現には大きな困難が伴うが、独自開発による教師無しクラスタリングや、複数の特徴抽出アルゴリズムによる解析技術の併用といった今年度までの研究成果によって、アーティファクト回避と網羅性維持を確保しつつ、生物学的意味を持つ遺伝子発現変動のハイスループット抽出を中心とした解析技術の開発に目処がついた。

これら技術の利用により、既にネットワークの要員遺伝子の抽出が時間的要素を維持したままで可能となっており、このアプローチによるネットワーク描出の糸口が次々に得られている。現在、これらの要員遺伝子の連結作業を加速する手段としての

インフォマティクス研究を本格的に開始しており、最終的には網羅的な毒性ネットワークの生成技術が開発可能と見込まれる。

F. 研究発表

1. 論文発表

1. Kitano, H.; Ghosh, S.; Matsuoka, Y. Social engineering for virtual "big science" in systems biology. *Nature Chemical Biology*. 7(6), 323-326, published online May 17, June 2011.
2. Tiago J.S. Lopes; Martin Schaefer; Jason Shoemaker; Yukiko Matsuoka; Jean-Fred Fontaine; Gabriele Neumann; Miguel A. Andrade-Navarro; Yoshihiro Kawaoka and Hiroaki Kitano. Tissue Specific subnetworks and characteristics of publicly available human protein interaction databases. *Bioinformatics* (2011) doi: 10.1093/bioinformatics/btr414, published online July 28, 2011. 27(17): 2414-21
3. Ghosh, S.; Matsuoka, Y.; Asai, Y.; Hsin, K-Y.; Kitano, H. Software for systems biology: from tools to integrated platforms. *Nature Reviews Genetics*. Published online Nov. 3, 2011.

2. 学会発表

1. 北野宏明. シグナルネットワーク創薬の最先端. バイオフィナンスギルド第9期第8回/9回セミナー「次世代のシグナル伝達医薬・抗がん剤」, バイオフロンティアパートナーズ特別講堂, Apr. 8, 2011. (invited)
2. 北野宏明. Systems medicine. LO皮膚科学研究会夏季シンポジウム, チサンホテル浜松町, June 11, 2011. (invited)
3. 北野宏明. システムバイオロジーの起源と展開. 電子情報通信学会ニューロコンピューティング研究会/情報処理学会バイオ情報学研究会 合同研究会, 琉球大学, June 23, 2011. (invited)
4. 北野宏明. システムトキシコロジープラットフォーム. 第38回日本トキシコロジー学会学術年会, パシフィコ横浜, July 12, 2011. (invited)
5. 北野宏明. システム創薬と新たなグローバルプロジェクトの可能性. 東京大学医科学研究所 G-COE 特別セミナー, 東京大学医科学研究所 (医科学教育セミナー), July 13, 2011. (invited)
6. Polouliakh, N. 時系列データから遺伝子ネットワーク推定. Japanese Society of Bioinformatics (JSBi) 第3回 応用システムズバイオロジー研究会, 九州大学, 福岡, July 27, 2011. (invited)
7. Kitano, H. Robustness and Fragility of Biology Systems. Karles' Invitational Conference on Microbial Systems and Synthetic Biology, Naval Research Laboratory, USA, Aug. 15, 2011. (invited)
8. Kang, H.; Ghosh, S.; Suzuki, K.; Matsuoka, Y.; Kitano, H. Integrated experimental and computational analysis for the dynamics of Tamoxifen resistance in MCF-7 cell lines. ICSB 2011, Rosengarten

- Conference Center, Mannheim, Germany, Aug. 29, 2011. (poster presentation)
9. Ghosh, S.; Matsuoka, Y.; Kitano, H. iPathways- biological pathways on your palm! ICSB 2011, Rosengarten Conference Center, Mannheim, Germany, Aug. 30-31, 2011. (poster presentation)
 10. S. Ghosh, Y. Matsuoka, A. Funahashi, H. Takizawa, N. Hiroi, Y. Asai, T. Abe, T. Okamoto, Y. Kido, H. Oka, T. Nomura, R. Adams, N. Hanlon, N. Tsorman, A. Clark, S. Gilmore, A. Yamaguchi, H. Mi, A. Muruganujan, P. D. Thomas, V. Satagopam, R. Schneider, T. Ohta, M. Miwa, S. Ananiadou, J. Tsujii, M. Wybrow, K. Marriott, I. Goryanin, H. Kitano. Garuda™ - Towards an integrated platform for systems biomedicine. ICSB 2011, Rosengarten Conference Center, Mannheim, Germany, Aug. 31, 2011. (poster presentation)
 11. Kitano, H. How networks can be used to find diagnostic and therapeutic candidates. ICSB 2011 Workshop 4: Systems biology to personalize medication and to find novel drug candidates, University of Heidelberg, Heidelberg, Germany, Sep. 1, 2011. (invited)
 12. S. Ghosh, Y. Matsuoka, A. Funahashi, H. Takizawa, N. Hiroi, Y. Asai, T. Abe, T. Okamoto, Y. Kido, H. Oka, T. Nomura, R. Adams, N. Hanlon, N. Tsorman, A. Clark, S. Gilmore, A. Yamaguchi, H. Mi, A. Muruganujan, P. D. Thomas, V. Satagopam, R. Schneider, T. Ohta, M. Miwa, S. Ananiadou, J. Tsujii, M. Wybrow, K. Marriott, I. Goryanin, H. Kitano. Garuda™ - Towards an integrated platform for systems biomedicine. Garuda 5 Workshop, University of Heidelberg, Heidelberg, Germany, Sep. 1, 2011. (poster presentation)
 13. Kitano, H. Garuda Vision and Goals. Garuda Five Workshop, University of Heidelberg, Heidelberg, Germany, Sep. 1, 2011.
 14. Kitano H. Robustness and Fragility of Disease Networks. 1st International Systems Medicine Symposium. Luxembourg Centre for Systems Biomedicine, Luxembourg, Sep. 27, 2011. (invited)
 15. Kitano, H. Garuda Goals. Garuda Six Workshop, Pavillon Skip, Luxembourg, Sep. 28, 2011.
 16. Kitano H. Novel scientific approaches in chronic diseases: Systems biology. India-Japan Symposium on Global Challenges in Health and Environment, Indian Embassy Auditorium, Tokyo, Oct. 7, 2011. (invited)
 17. Kitano, H. Systems Drug Design. Invited presentation at Monash University, Monash University, Melbourne, Australia, Oct. 19, 2011. (invited)
 18. Kitano, H. Systems biology and its

medical and industrial implications. Systems Biology Forum, Monash University, Australia, Oct. 21, 2011. (invited)

19. 北野宏明. システムバイオロジーと創薬パイプライン. 神奈川科学技術アカデミー(KAST) 教育講座 システムバイオロジー・合成バイオロジーの最新動向コース, Nov. 7, 2011. (invited)
20. Kitano, H. Unmet medical needs in the aging society. Economic Opportunities from the Ageing Society: Policies and Challenges, Asian Development Bank Institute, Tokyo, Nov. 18, 2011. (invited)
21. Kitano, H. HD-Physiology Project and the Garuda Alliance Software and Resource Platform. The 4th Global COE International Symposium on Physiome and Systems Biology for Integrated Life Sciences and Predictive Medicine, Senri-Hankyu Hotel, Osaka, Nov. 21, 2011. (invited)
22. 北野宏明. システム創薬の展開. 第3回三井情報バイオサイエンスセミナー, 三井情報株式会社, 東京, Nov. 28, 2011. (invited)
23. Kitano, H. Introduction to Systems biology and to multi-level modeling with CellDesigner. Workshop “Can Systems Biology Aid Personalized Medication?” Linköping University, Sweden, Dec. 5, 2011. (invited)
24. Kitano, H. Overview of HD-physiology network. Workshop “Can Systems Biology Aid Personalized

Medication?” Linköping University, Sweden, Dec. 6, 2011. (invited)

G. 知的財産権の出願・登録状況
特になし

Systems Toxicology

1

- Garuda Alliance : Open Common Software Platform
- Network Reconstruction
- AGCT Clustering data analysis

2



GARUDA
THE WAY BIOLOGY CONNECTS

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**

(C) Hiroaki Kitano, 2010 ***
LIMITED CIRCULATION ***

Data and Knowledge base Problems

- Too many fragmented DBs and KBs.
- Inconsistency/maintenance/error-correction
- Users are forced to integrate by them self.
- Poor feedback mechanism exists that prevents DB/KB to improve their quality

(C) Hiroaki Kitano, 2010 ***
LIMITED CIRCULATION ***

The Garuda Alliance

- To develop an open & common software and data platform for biological and medical research
- Consistent GUI, APIs, and other development framework
- Provide a consistent user experience
- Enables efficient and quality software development
- Effective dissemination of tools and resources

The Garuda Alliance

- The global alliance for research, development, and sustainable deployment of common platform for systems biology and drug discovery.
- Launched at Garuda One Workshop at Okinawa in Feb 2010.

(C) Hiroaki Kitano, 2010 ***
LIMITED CIRCULATION ***

The Garuda Alliance

- One stop service
- Consistent and unified user experiences
- Quality standard
- Additional benefits

- Other points to be explored...

(C) Hiroaki Kitano, 2010 ***
LIMITED CIRCULATION ***

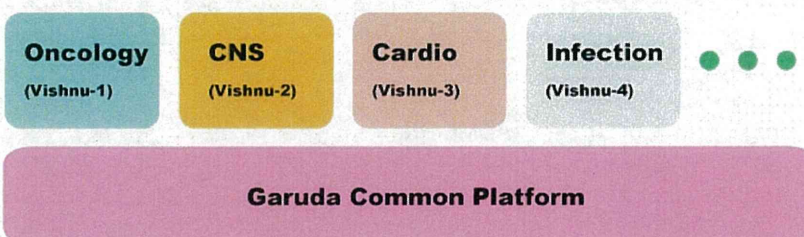
What is Garuda?

Garuda is the bird in Hindu mythology that Vishnu rides. Vishnu is one of three Hindu Gods that governs maintenance of the universe and commanded the Churning of the Sea of Milk.

Churning of the *Sea of Milk*: A *Hindu* myth in which gods and demons cooperate to churn the primordial ocean, in order to produce amrita, the elixir of immortality.

Garuda Vision

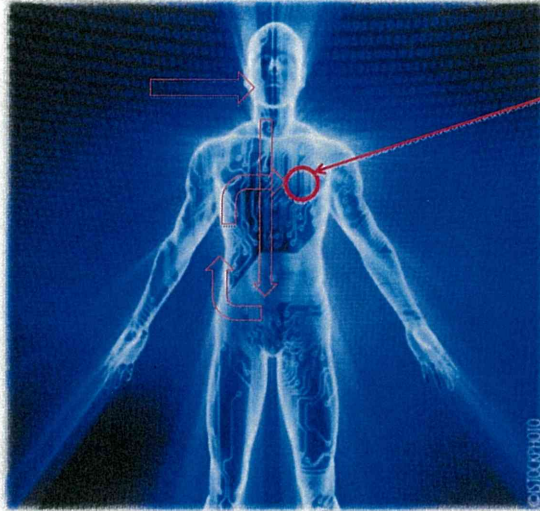
A common platform of tools that supports applications



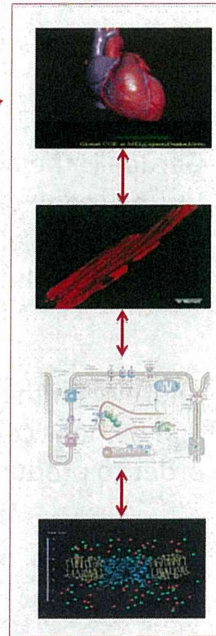
Garuda modules can be tailored to leverage functions across disparate tools which otherwise do not inter-operate, while integrating public domain knowledge spread across isolated databases



HD-Physiology Project



ADME/PK model



Heart model

Inter-layer interactions

