

平成21年6月22日現在

研究種目：若手研究（A）

研究期間：2006～2008

課題番号：18681031

研究課題名（和文） 遺伝子の発現情報に基づく生命現象の因果性に関する統計解析

研究課題名（英文） Statistical Analysis of causal relationship between the genes from gene expression data.

研究代表者 油谷 幸代（ABURATANI SACHIYO）

独立行政法人産業技術総合研究所・生命情報工学研究センター・研究員

研究者番号：10361627

## 研究成果の概要：

本研究では、発現プロファイルデータに既知の生物学的知見を加えて遺伝子間の因果性を推定する統計的手法を開発し、関連の有無や強弱に加え遺伝子間の因果性の解明を行った。また、開発した手法をウェブツールとして広く公開した。本研究で開発した手法は、数値データに既知の生物学的知見を加えた統計的手法であり、この手法を用いることによって、遺伝子間の発現相互作用について関連性の有無や強弱に加え遺伝子間の因果性の解明までを行うことが可能になった。

## 交付額

（金額単位：円）

	直接経費	間接経費	合計
平成18年度	9,400,000	2,820,000	12,220,000
平成19年度	6,400,000	1,920,000	8,320,000
平成20年度	6,500,000	1,950,000	8,450,000
年度			
年度			
総計	22,300,000	6,690,000	28,990,000

研究分野：複合新領域

科研費の分科・細目：ゲノム科学・基礎ゲノム科学

キーワード：生体生命情報学・マイクロアレイ・発現制御・遺伝子ネットワーク・グラフィカル連鎖モデル

## 1. 研究開始当初の背景

研究開始当初、ポストシーケンス研究としてトランスクリプトーム解析が盛んであった。実験的にはマイクロアレイ法が確立されたことによって、*in vivo*での遺伝子発現プロファイルを数値データとして得ることが可能になっていた。当時は、この実験的手法を用いた数多くの遺伝子発現データが収集され、データベース上に情報として蓄積さ

れつつあった。

これらの遺伝子発現プロファイルデータに対して様々な数理モデルを適用し、細胞内における遺伝子間相互作用の推定や、遺伝子の機能分類の研究が数多く行われてきたが、中でも、統計的手法は大量の発現プロファイルデータからゲノムレベルでの遺伝子間相互作用を推定するために有用であると考えられていた。しかしながら、ゲノムレベルでの遺伝子間の因果関係を推定する統計的

法は当時確立されていなかった。そこで、本研究でゲノムレベルでの遺伝子間の因果性を推定する統計的手法を開発し、細胞周期の各段階や発生のステージ間の遺伝子制御など、異なる細胞状態の遺伝子群間の因果関係を解明することを目標とした。

## 2. 研究の目的

本研究の目的は、発現プロファイルデータに既知の生物学的知見を加えて遺伝子間の因果性を推定する統計的手法を開発し、実際に開発した手法によってこれまで困難であった経時的な遺伝子発現の制御システムの解明である。そこで、具体的に以下の8項目を目標とした。

- (1) 生物学的知見を加えた、遺伝子間の因果性を推定する統計的手法の開発。
- (2) 開発した統計的手法を発現プロファイルデータへ適用し、遺伝子間の因果性推定。
- (3) 推定された因果関係の強弱について統計検定を行い、統計的に有意な因果関係を抽出。
- (4) 統計的に有意な因果関係を持つ遺伝子群と既知の制御関係との比較検証による、開発した手法の評価
- (5) 有意な因果関係を持つ遺伝子群の生化学的機能から、推定手法の正当性を評価。
- (6) 細胞周期や発生段階における既知の知見から、推定した因果関係の評価。
- (7) 推定した因果関係から、細胞周期や発生段階の遺伝子レベルでの制御システムの解明。
- (8) (1)～(7)で、新規知見を得るために有用と判断された統計的手法のツール化と、ウェブ公開。

## 3. 研究の方法

研究目的で記載した8項目を遂行するために、具体的な研究方法として、以下の方法を用いた。

- (1) 収集した発現プロファイルデータの統合  
複数の実験から得られた発現プロファイルデータを扱う上で、データフォーマットが異なる複数のデータの統合を行った。データベ

ースから収集した発現プロファイルデータについて、データフォーマットの整備を行い、さらに解析に必要な部分を抽出し、データ統合を行った。

### (2) 遺伝子選択を行う統計的手法の開発

ゲノムレベルで測定された発現プロファイルデータから、細胞周期の各段階において、特異的に発現変動している遺伝子群の選択を行った。

### (3) 遺伝子発現の因果性を解析する統計的手法の開発

本研究の準備としてグラフィカル・ガウシアン・モデルは、遺伝子間相互作用の推定に有用であると考えられた。そこで、このモデルをさらに発展させ、因果性までを推定することができる統計的手法を開発した。

### (4) 開発した統計的手法の発現プロファイルデータへの適用

開発した統計的手法を、出芽酵母・分裂酵母の細胞周期に関連した発現プロファイルデータに適用し、細胞周期における遺伝子発現の因果性推定を行った。

### (5) 推定した因果性の強弱を評価する統計検定

数値データとして得られた因果性について統計検定を行い、有意に強い因果性の抽出を行った。

### (6) 評価基準となる既知の制御関係データの整備

収集した既知の制御関係データを、①細胞分裂に関与する制御関係 ②恒常的に行われる制御関係 ③特定の条件下で起こる制御関係の3種類に分類した。

### (7) 既知の制御関係からの推定した遺伝子間の因果性の評価

既知の制御関係と、本研究で開発した統計的手法にて推定された遺伝子発現の因果性を比較し、開発した統計的手法によって得られる生物学的知見の正当性評価を行った。

### (8) 既知の生物データとの照合による、推定した遺伝子間の因果性評価

既知の生物学データから、該当遺伝子に関する情報の抽出を行った。その後、推定され

た因果性について、生命維持における生化学的機能の関係性や転写因子の結合サイトの有無などを調べ、推定した因果性の生物学的な正当性の評価を行った。

(10) ガン進行過程における発現プロファイルデータの収集と整備

肝臓ガンの進行過程について、ガン進行における各段階の遺伝子発現を測定した発現プロファイルデータを実験研究者より提供して頂いた。提供された肝臓ガンの発現プロファイルデータに対し、本研究で開発した手法を適用できるようにデータ整備を行った。

(11) ガン進行過程における各フェーズごとに発現している遺伝子の選択

先に確立した遺伝子選択の統計的手法に基づき、フェーズ特異的に発現している遺伝子群の選択を行った。

(12) 推定した遺伝子間の関連性の評価

既知のガン細胞に関する既知の知見と、推定した遺伝子間の関連性との整合性を調べた。次に、比較した既知の生物データと推定した因果性の中で、実験的には確認されていないものの生物学的に有り得ると判断された関連構造の抽出を行った。さらに、これらの因果性の統計検定を行い、その有意性の評価を行った。

(13) これまでに確立した因果性推定のための統計的手法のパッケージ化

本研究によって、確立した遺伝子発現の因果性推定のための統計的手法、および推定された因果性の評価を行う統計検定すべてを包含したプログラムパッケージの開発を行った。

(14) 開発したツールのウェブ上での公開

開発したプログラムパッケージを統合的な遺伝子相互作用推定サーバに組み込み、ウェブ上での公開を行った。

#### 4. 研究成果

研究の目的にて記載した8項目について、各年度毎に以下の成果を挙げた。

(1) 平成18年度

平成18年度は、研究目的の(1), (2), (3)について主に行った。具体的な成果は以下の3項目である。

① グラフィカル連鎖モデルによる遺伝子間の因果性推定手法の開発

グラフィカルモデルの一つであるグラフィカル連鎖モデルを改良することで、時系列に沿って明確な区分のできる時期に発現する遺伝子間の因果的関連性を検出する方法の開発を行った。

② グラフィカル連鎖モデルの遺伝子発現プロファイルデータへの適用

SMDにて公開されていた酵母細胞周期の各段階で特異的に遺伝子発現する遺伝子群の発現プロファイルデータに、開発したグラフィカル連鎖モデルを適用し、遺伝子間の因果性推定を行った。

③ 推定した因果関係への統計的検定の適用

本研究で開発した手法では、遺伝子間の因果関係の強弱が偏相関係数の絶対値として評価可能であった。そこで、推定された遺伝子間の因果関係の中から、統計的に有意に強い因果関係を抽出するため、偏相関係数のt検定を行った。

(2) 平成19年度

平成19年度は、研究目的の(4), (5), (6)について主に行った。

① 既知の制御関係との比較検証

これまでに開発した手法によって、統計的に有意な因果関係を持つと推定された遺伝子群について、既知の制御関係との比較検証を行った。具体的には、収集した既知の制御関係データを、a)細胞分裂に関与する制御関係 b)恒常的に行われる制御関係 c)特定の条件下で起こる制御関係の3種類に分類し、分類した既知の制御関係それぞれと、本研究で開発した統計的手法にて推定された遺伝子発現の因果性を比較した。

② 開発した統計的手法の正当性評価

さらに、他のデータベースから得られた様々な生物学的知見を元に、推定された遺伝子間の因果関係の正当性の評価を行った。具体的には、Yeast Proteome Database (YPD) や mips, TRANSFAC から、出芽酵母・分裂酵母におけるコードタンパク質の生化学的機能に関するデータの収集を行い、推定された因果性について、生命維持における生化学的機能の関係性や転写因子の結合サイトの有無などを調べた。特に、細胞周期における既知の知見から、推定した因果関係の評価を行った。

③ ヒト肝臓ガン細胞における遺伝子間の関連性推定

開発してきた手法は、生物学的な情報を組み込んだ遺伝子ネットワーク推定手法であ

ることから、ガン進行過程などの不可逆な細胞変化における遺伝子間の因果性推定も可能であると考えられた。そこで、実験研究者から、肝臓ガンの進行過程における遺伝子発現プロファイルデータの供与を受け、肝臓ガンにおける遺伝子間の関連構造解析を行った。

### (3) 平成 20 年度

平成 20 年度は、研究目的(7)、(8)について主に行った。

#### ① 遺伝子選択手法の確立

各段階における遺伝子発現を測定した発現プロファイルデータから先に確立した遺伝子選択の統計的手法に基づき、ステージ特異的に発現している遺伝子群の選択を行った。因果性推定研究において、結果として発現変動を起こした遺伝子群に注目する場合と、原因となる発現変動を起こした遺伝子群に注目する場合で異なった遺伝子選択手法を用いることにより、解析目的に沿った遺伝子選択が可能となった。

#### ② 細胞周期の制御システムの解明

経時的に変化する細胞間における遺伝子間の制御ネットワークを推定するために、これまでに開発した統計的手法をより網羅的な細胞周期の発現プロファイルデータに適用し、遺伝子発現の因果性を推定した。さらに、整備した既知の発生制御に関する既知の知見と、推定した遺伝子間の因果性との整合性を調べた。最終的には、比較した既知の生物データと推定した因果性の中で、実験的には確認されていないものの生物学的に有り得ると判断された因果性の抽出を行い、これらの因果性の有意性の評価を行った

#### ③ 開発した手法のウェブツール化

これまでの研究によって、確立した遺伝子発現の因果性推定のための統計的手法をプログラムパッケージとして開発した。さらに、開発したプログラムパッケージを統合的な遺伝子相互作用推定サーバに組み込み、ウェブ上での公開を行った。特に、大量データの解析に耐えるようにハイパフォーマンスのサーバを完備し、様々な分野の研究者がストレスなく解析を行うことができる環境整備を行った。

#### ④ 開発した手法の紹介

開発公開したウェブツールについて外国および国内書籍にて広く紹介した。

本研究で開発された統計的手法をウェブ上で公開したことによって、実験研究者による遺伝子間の因果性解析が推進され、生命シス

テムの解明に貢献できたと考える。

### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 9 件)

① Aburatani, S. and Horimoto, K., "Statistical Analysis of the Relationship between LexA-Regulated Genes from Expression Profiles", Res. Commun. Biochem. Cell Mol. Biol., Vol.12, NOS.1&2, pp.119-135, 2008, 査読有

② Hayashida, M., Sun, F., Aburatani, S., Horimoto, K. and Akutsu, T., "Integer Programming-based Approach to Allocation of Reporter Genes for Cell Array Analysis", Int. J. Bioinformatics Research and Applications, Vol.4(4), pp.385-399, 2008, 査読有

③ Saito, S., Aburatani, S. and Horimoto, K., "Network evaluation from the consistency of the graph structure with the measured data", BMC Systems Biology, Vol.2:84, pp.1-14, 2008, 査読有

④ Aburatani, S., Sun, F., Saito, S., Honda, M., Kaneko, S. and Horimoto, K., "Gene systems network inferred from expression profiles in hepatocellular carcinogenesis by graphical Gaussian model", EURASIP J. Bioinfo. Systems Biol., Vol.47214, pp.1-11, 2007, 査読有

⑤ Hayashida, M., Sun, F., Aburatani, S., Horimoto, K. and Akutsu, T., "Integer Programming-based Approach to Allocation of Reporter Genes for Cell Array Analysis", Proceedings of OSB2007, pp.21-28, 2007, 査読有

⑥ Aburatani, S., "ASIAN: A Network Inference Web Server for Biologists", Proceedings of the 11th World Multiconference on Systemics, Cybernetics and Informatics, Vol.4 pp.1-6, 2007, 査読有

⑦ Aburatani, S., "Inference of Complex Regulatory Network for the Cell Cycle System in *Saccharomyces cerevisiae*", Algebraic Biology 2007 Lecture Notes in Computer Science, pp.350-364, 2007, 査読無

⑧ Aburatani, S., Saito, S., Toh, H., Horimoto, K., "A Graphical Chain Modeling Approach for Analyzing Gene Expression Profiles", *Statist. Method.*, Vol. 3, pp. 17-28, 2006, 査読有

⑨ Aburatani, S. and Horimoto, K., "Serial Network Inference in Cell Cycle regulation on Yeast", *Proceedings of the 10th World Multiconference on Systemics, Cybernetics and Informatics*, Vol. 4, pp. 1-6, 2006, 査読有

[学会発表] (計 4 件)

① Aburatani, S., "Inference of Complex Regulatory Network for the Cell Cycle System in *Saccharomyces cerevisiae*", *Algebraic Biology 2007*, RISC (Austria), 2007. 7

② Aburatani, S., "ASIAN: A Network Inference Web Server for Biologists", *WMSCI2007*, Orland (USA), 2007. 7

③ Aburatani, S., "Serial Network Inference in Cell Cycle Regulation on Yeast", Horimoto, K., *WMSCI2006*, Orland (USA), 2006. 7

④ Aburatani, S., Saito, S., Horimoto, K., "Inference of Cell Cycle Regulatory Network from Expression Profiles by Graphical Chain Modeling", 20th IUBMB International Congress of Biochemistry and Molecular Biology, Kyoto (Japan), 2006. 6

[図書] (計 3 件)

① "Bioinformatics for Systems Biology : Chapter 29 :ASIAN: Network Inference Web Server", Aburatani, S., Saito, S., Horimoto, K., *Humana Press*, pp. 563-577, Stephen Krawetz 編集, 全 740 ページ, 2009. 2

② "マイクロアレイデータ統計解析プロトコール: 偏相関係数によるネットワーク推定グラフィカル・ガウシアン・モデル法", 油谷幸代, 羊土社, pp. 224-233, 藤渕航・堀本勝久 / 編, 全 255 ページ, 2008. 4

③ "はじめてのバイオインフォマティクス: トランスクリプトームとプロテオーム", 油谷幸代, 講談社サイエンティフィック, pp. 97-123, 籾博幸 編集, 全 166 ページ,

2006. 12

[その他]

ASIAN : 階層クラスタリングとグラフィカル・ガウシアン・モデリング (GGM) の組み合わせによるネットワーク推定ツール  
<http://eureka.cbrc.jp/asian/>

## 6. 研究組織

### (1) 研究代表者

油谷 幸代 (ABURATANI SACHIYO)  
産業技術総合研究所・  
生命情報工学研究センター・研究員  
研究者番号 : 10361627

### (2) 研究分担者

### (3) 連携研究者