

令和元年6月14日現在

機関番号：82626

研究種目：基盤研究(C) (一般)

研究期間：2016～2018

課題番号：16K07225

研究課題名(和文) システム構造化手法を用いた細胞分化メカニズムの解明

研究課題名(英文) Elucidation of Cell Differentiation system by Structural Method

研究代表者

油谷 幸代 (Aburatani, Sachiyo)

国立研究開発法人産業技術総合研究所・生命工学領域・副ラボ長

研究者番号：10361627

交付決定額(研究期間全体)：(直接経費) 3,500,000円

研究成果の概要(和文)：本研究では、システム構造化手法を適用したネットワークモデル推定手法の開発を行った。さらにネットワークグラフの構造に制約条件を入れたネットワークモデルの開発を実施した。特に、因果関係の有無が確定している遺伝子間において、その知見を包含したネットワークモデルを構築するため、最適化アルゴリズムを実施する時、共分散行列上で特定の要素に値を持たせたままで数値のみ最適化する方法を開発し適用した。構築したネットワークモデル構造において、一部を固定したうえで、GAの点変異を利用して改良した最適化アルゴリズムのを実施することで、少数の特定条件下で活性化するサブグラフ構造の抽出方法を開発した。

研究成果の学術的意義や社会的意義

本研究の学術的意義としては、従来法では困難であった少数データからのネットワーク構造推定が可能になる点が挙げられる。

具体的には、1. 特定の条件下で活性化しているサブグラフの構造が明らかになることで、実際に生体内で機能している制御プロセスの解明に有用な情報になる。2. 細胞内における異なった細胞内分子間の関係性が明らかになり、メカニズム解明に有用な情報になる。3. サブグラフ推定によって、細胞内のフィードバック制御やサイクリック制御のモデル化が可能になる。4. 活性化部分の推定には多変量を必要としないことから、今後増加すると予想される1細胞計測データに適用可能である。

研究成果の概要(英文)：In this study, I developed a new method for network modeling based on structural method. Furthermore, I improved the network modeling method for addition of structural restriction. For example, if there are known causality between the genes, I can estimate the network structure including the known causalities. I developed the numerical method to keep the elements which were known causalities within the co-variance matrices. I also improved my network optimization algorithm combined with GA algorithm to avoid the local minimum. I applied my developed and improved methods to real expression data, and I could select the activated sub-graph structure at the specific conditions.

研究分野：システム生物学

キーワード：システム構造化 構造方程式モデリング ネットワークモデリング 遺伝子発現制御 細胞分化メカニズム

## 様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

### 1. 研究開始当初の背景

近年、次世代シーケンサーの普及により、細胞内における転写産物量を測定した実験研究が数多く行われ、各種細胞・器官で測定されたデータも加速度的に蓄積されている。実験的に得られたデータの蓄積に伴い、遺伝子発現データから遺伝子間ネットワークを推定する数理モデルも数多く発表されており(Akutsu et al., 2000, Friedman et al., 2000, Horimoto & Toh, 2003)、遺伝子発現データからその生物種の遺伝子間相互作用を推定するのに有用であることが明らかになっている。最近では、NCBI によって GEO (Gene Expression Omnibus) という遺伝子発現データベースが整備され、データ集積が体系的に行われている。このように充実してきた発現データを用いて遺伝子ネットワークを推定することは、統計的に優位かつ実際に生体細胞内で起こりうる制御関係を包括したモデルを構築できることが知られている。

一方、多能性幹細胞における選択的分化や、発生過程における細胞分化制御では、初期胚におけるこの遺伝子ネットワークの違いに起因した細胞内バランスの違いに依るものとも考えられている。そのため、細胞分化の制御構造を明らかにするためには、包括的な遺伝子ネットワークモデルの中で、どの部分が活性化しているかを明らかにしていく必要がある。さらに、細胞分化の過程では細胞が時間的・空間的に変化することから、従来の遺伝子間の相互関係だけでは現象のメカニズム解明は困難である。事実、モデル生物では、発生の初期において、細胞質因子や中心体など遺伝子やタンパク質以外の物質による遺伝子発現制御が確認されている (Seydoux et al., 1996, Goldstein et al., 1996)。さらに、近年では ncRNA の一部によって遺伝子の発現制御が行われていることも示唆されている。このように、細胞内における様々な要因が遺伝子発現制御を行っている可能性が高いにも関わらず、これら別因子は遺伝子発現プロファイルデータ上にはデータとして存在しなかった。よって、細胞分化制御のメカニズム解明には、従来の遺伝子間の関連構造推定だけではなく、遺伝子と他の細胞内因子の関連性を網羅した細胞内因子間ネットワーク推定が必要であると考えられた。

### 2. 研究の目的

本研究では、公開されている大量の発現データの情報を利用することで、少数サンプルや一点計測由来のデータから活性化しているサブグラフを推定する手法を開発することを目的とした。従来のネットワーク構造推定手法は多変量データ解析を基盤としていることが多く、少数サンプルや一点計測データに適用が困難であった。本研究では、発生過程における発現データが充実している線虫・ショウジョウバエ・マウス ES 細胞において、発現データから包括的ネットワークモデルを構築する手法を開発した。推定された包括的ネットワーク構造をグラフ構造の制約条件とし、細胞分化過程やある細胞系譜上で測定されたデータにもっとも近くなるようにグラフ上の各エッジの係数を推定した。特定の条件下で測定されたデータが複数ある場合には、グラフ構造上の制約条件はそのまま構造方程式モデリングを適用し、より正確な各係数の推定を行った。

### 3. 研究の方法

本研究における全体的計画では、第一に ISM 法や DEMATEL 法によって推定された包括的遺伝子ネットワーク構造から、実測された特定の発現プロファイルパターンを再現するために最適な部分構造を推定する手法の開発を行った。次に、開発した手法をモデル生物の細胞分化過程で測定された発現プロファイルへ適用し、細胞分化や特定の器官で測定された発現データから、それぞれの細胞内で活性化している遺伝子発現制御ネットワークモデルの部分構造をサブグラフとして推定した。最後に、様々な条件下で推定した活性化サブグラフの構造、および各パラメータの値から細胞分化や器官形成プロセスのメカニズムを推定した。

研究項目としては、以下の項目を実施した。

- 1) 線虫・ショウジョウバエ・マウス ES 細胞などにおける遺伝子発現データの網羅的収集。
- 2) ISM 法、DEMATEL 法による遺伝子ネットワークモデル推定手法の開発。
- 3) 2) で開発した手法を 1) に適用した各モデル生物における遺伝子ネットワークモデルの構築。
- 4) 3) で構築したモデル構造を制約条件とした、構造方程式モデリングを基盤とする新規活性化部位推定予測手法の開発。
- 5) 開発した手法を、実データである発現プロファイルへ適用し、活性化している遺伝子ネットワークのサブグラフを推定。
- 6) 推定したサブグラフ上における各パラメータ値・潜在変数等の検証。

### 4. 研究成果

研究項目 1) 各モデル生物における発現プロファイルデータの網羅的収集：

GEO (Gene Expression Omnibus) から、線虫の各細胞系譜、ショウジョウバエ初期胚における細胞分化過程での遺伝子発現データ、さらにマウス ES 細胞における多能性に関する遺伝子発現データを網羅的に収集した。収集したデータを実験条件や細胞分化の段階によって分類し、解析対象とする現象に適応した実験条件下で測定された発現プロファイル群を選択した。選択した発現プロファイル群のデータフォーマットを整備した後、有意差解析・相関分析・ANOVA・クラスタリングなどの多変量解析を行い、特徴的な発現をしている遺伝子群を抽出した。

研究項目 2 ) ISM 法、DEMATEL 法による遺伝子ネットワークモデル推定手法の開発：  
システム構造化手法の遺伝子発現データへの適用のための実装プログラムを開発した。

研究項目 3 ) 各モデル生物における遺伝子ネットワークモデルの構築：

2 ) で開発した手法を、1 ) で選択した遺伝子群・収集した各モデル生物の全データに対し適用し、包括的な遺伝子ネットワーク構造推定を行った。基本的には、遺伝子対遺伝子でネットワークモデルを構築したが、選択された遺伝子数が 1000 個以上になる場合には、クラスタリングによるグループ化を行い、グループ間の階層構造モデル推定 各グループ内での制御構造モデル推定 と 2 段階で推定した。この段階ではなるべく多くの制御構造を包含したモデルを構築するようにパラメータを調整した。

研究項目 4 ) 3 ) で構築したモデル構造を制約条件とした、構造方程式モデリングを基盤とする新規活性化部位推定予測手法の開発：

これまでに開発したネットワークモデル推定手法を各モデル生物における遺伝子発現プロファイルに適用し、遺伝子ネットワークモデルを構築を行った。さらに、構築したモデル構造を制約条件とした、構造方程式モデリングを基盤とする新規活性化部位推定予測手法を開発した。具体的には、第一に初期モデルの構築手法を新規に開発した。従来、既知の知見から SEM の第一段階である初期モデルを構築してきたが、ベイジアンモデルや ブーリアンモデルを遺伝子発現データに適用し、複数のモデル推定手法でロバストな構造を抽出することで、初期モデルを自動的に生成することを可能にした。

研究項目 5 ) 開発した手法を、実データである発現プロファイルへ適用し、活性化している遺伝子ネットワークのサブグラフを推定：

特定の条件下で活性化している部位を推定するため、疑似逆行列をつかった重み行列を計算し、SEM で最適化された重み行列からの差異を計算した。特に、3 ) で構築した包括的遺伝子ネットワークモデルのグラフ構造を制約条件とした最適化では従来の MI スコアによる最適化では、精度が保証されなかったことから、エッジの削除や追加は行わず、エッジの持つ係数を調整することで特定条件下にて測定された遺伝子発現プロファイルと同じパターンの発現プロファイルになるような手法の開発を行った。

研究項目 6 ) 推定したサブグラフ上における各パラメータ値・潜在変数等の検証：

構築した包括的遺伝子ネットワークモデルで推定された各エッジの weight や p 値を元にランダムに数値を与えることで、ネットワークを構成する遺伝子群の値を計算し、実測値との差を検定した。

最終的に、ネットワークグラフの構造に制約条件を入れたネットワークモデルの開発を実施した。特に、生物学的知見等から因果関係の有無が確定している遺伝子間において、その知見を包含したネットワークモデルを構築するため、最適化アルゴリズムを実施する時、共分散行列上で特定の要素に値を持たせたまま まで数値のみ最適化する方法を開発し、適用した。システム構造化手法を適用したネットワークモデル推定手法の開発を行ってきたが、開発していく中で、最適化したネットワークモデルが局所解になっている可能性が考えられたので、GA の点変異を利用して最適化アルゴリズムの改良を行った。そのうえで、構築したネットワークモデル構造において、一部を固定したうえでこの最適化アルゴリズムを実施することで、少数の特定条件下で活性化するサブグラフ構造の抽出方法を開発した。具体的には、従来モデル最適化アルゴリズムで使用してきた Modification Index (MI) スコアについて、閾値を設定し 2 段階でモデル構築を行うアルゴリズムを開発した。次に、これまでに開発した疑似逆行列をつかった重み行列からの活性化部位の推定手法について、疑似逆行列と SEM で最適化された重み行列の構造類似性について、ハミング距離、相関係数等で比較した。SEM で最適化された重み行列からの差異を計算した。これらの手法をまとめた結果を、国際学会で発表するとともに、開発した手法を実データに適用した結果の論文を作成中である。また、個別の手法については、随時各生物のデータに適用し論文発表を行った。

## 5 . 主な発表論文等

〔雑誌論文〕(計 4 件)

1. PS Wong, K. Tashiro, K. Kuhara, S. Aburatani, Elucidation of the sequential transcriptional activity in Escherichia coli using time-series RNA-seq data, *Bioinformatics*, *Bioinformatics*, 2017, 13(1), pp.25-30
2. Itaru Tkaeda, Hiroya Itoh, Makoto Matsui, Takashi Shibata, Masayuki Machida, Sachiyo Aburatani, Exploration of Activated Pathways for Improving Antifungal Agent FR901469 Productivity in Fungal Species No. 11243 Using Comprehensive Pathway Model, *Journal of Biosciences and Medicines*, 2017, vol.5, pp.16-31
3. Koji Ishiya, Sachiyo Aburatani, Outlier Detection for Minor Compositional Variations in Taxonomic Abundance Data, *Applied Sciences*, 2019, 9(7),
4. S Aburatani, T Tamura, Development of a new classification approach for detecting the heterogeneity in cell population, *FEBS OPEN BIO*, 2018, vol8

〔学会発表〕(計 10 件)

1. 井手圭吾, 油谷幸代, 竹山春子, Maximal Information Cross Correlation を用いた生物ネットワークの構築, 第6回日本生物物理学会関東支部会, 2017, 13<sup>th</sup>, Mar.-14<sup>th</sup>, Mar, 早稲田大学先端生命医科学センター
2. Sachiyo Aburatani, Network Inference by Structural Equation Modeling for systems biology, Systems and Synthetic Bacterial biology (国際学会), 2018,
3. Keigo Ide, Toru Maruyama, Michihiro Ito, Hiroyuki Fujimura, Yoshikatu nakano, Soichiro Suda, Sachiyo Aburatani, Haruko Takeyama, Network Analysis to Reveal Microbial Community Dynamics in the Coral Reef Ocean, 19<sup>th</sup> International Conference on Systems Biology (国際学会), 2017
4. Haruka Ito, Toru Maruyama, Michihiro Ito, Hiroyuki Fujimura, Yoshikatu nakano, Soichiro Suda, Sachiyo Aburatani, Haruko Takeyama, Gene Expression Analysis for Corals / Zooxanthellae under High Seawater Temperature Stress, 19<sup>th</sup> International Conference of Systems Biology (国際学会), 2017
5. Sachiyo Aburatani, Network inference of AP pattern formation system in *D. melanogaster* by structural equation modeling, 4<sup>th</sup> Core-to-Core International Symposium (国際学会), 2017年
6. Sachiyo Aburatani, Inference of Pluripotency Control in mouse Embryonic Stem Cells, 9<sup>th</sup> International Conference and Exhibition on Metabolomics & Systems Biology (国際学会), 2017年
7. Sachiyo Aburatani, Tomohiro Tamura, Statistical Inference of Gene Regulatory Network from Gene Expression Profiles, III International Symposium on Profiling (国際学会), 2017
8. Sachiyo Aburatani, Statistical Approach for Network Inference by Structural Equation Modelling, International Congress on Biological and Medical Sciences 2018 (招待講演)(国際学会), 2018
9. Sachiyo Aburatani, Tomohiro Tamura, Development of a New Classification Approach for detecting the heterogeneity in cell population, FEBS Congress 2018 (国際学会), 2018
10. Koji Ishiya, Sachiyo Aburatani, Robust age prediction model based on massive methylation data among multi-tissue, 19<sup>th</sup> International Conference on Systems Biology (国際学会), 2018

〔図書〕(計0件)

〔産業財産権〕

出願状況(計0件)

名称：  
 発明者：  
 権利者：  
 種類：  
 番号：  
 出願年：  
 国内外の別：

取得状況(計0件)

名称：  
 発明者：  
 権利者：  
 種類：  
 番号：  
 取得年：  
 国内外の別：

〔その他〕

ホームページ等

## 6. 研究組織

(1)研究分担者

研究分担者氏名：

ローマ字氏名：

所属研究機関名：

部局名：

職名：

研究者番号（8桁）：

(2)研究協力者

研究協力者氏名：

ローマ字氏名：

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。