

平成 21 年 5 月 29 日現在

研究種目：基盤研究（B）  
 研究期間：2006 年度～2008 年度  
 課題番号：18300097  
 研究課題名（和文） 遺伝子ネットワークによる創薬ターゲットパスウェイのイン・シリコ探索技術  
 研究課題名（英文） In Silico Search for Drug Target Pathways by Gene Networks

研究代表者  
 宮野 悟（MIYANO SATORU）  
 東京大学・医科学研究所・教授  
 研究者番号：50128104

研究成果の概要：遺伝子ノックダウンや薬剤応答などに基づいた遺伝子発現データから大規模遺伝子ネットワークを推定する方法と統計的解析とシミュレーションによる遺伝子ネットワークの解析方法を用いて、遺伝子ネットワークに基づく、創薬ターゲットパスウェイを探索するためのバイオインフォマティクス技術を開発した。この技術を使って、薬剤 Fenofibrate, Gefitinib 並びに漢方薬が影響を与えるパスウェイの解析を実現した。

## 交付額

（金額単位：円）

	直接経費	間接経費	合計
2006 年度	4,700,000	1,410,000	6,110,000
2007 年度	5,300,000	1,590,000	6,890,000
2008 年度	4,700,000	1,410,000	6,110,000
年度			
年度			
総計	14,700,000	4,410,000	19,110,000

## 研究分野：総合領域

科研費の分科・細目：情報学・生体生命情報学

キーワード：遺伝子ネットワーク，システム生物学，シミュレーション，遺伝子発現マイクロアレイデータ，創薬ターゲット遺伝子，Fenofibrate，Gefitinib，漢方薬

## 1. 研究開始当初の背景

ドラッグディスカバリーは、ドラッグターゲットの選択とドラッグ自身の設計の2つに大きく分けることができる。ドラッグの設計では、既に解明されているターゲットに対して最適な低分子薬物候補化合物を探索することに主力が置かれ、計算化学やコンビナトリアルケミストリーなどの方法により、合理的にリード分子の設計・選択が行われている。一方、ターゲットの選択では、ゲノムワイドな遺伝子情報などが利用できる状況になっているが、そうした情報を積極的に利用し、

システムティックにターゲット選択する合理的・効率的な方法は未開拓の状態であった。このターゲット探索では、高コストの伝統的な分子生物学研究の手法が主流である。しかし、ゲノム研究からシステムバイオロジーへとつながる世界的な研究の流れと、ヒトゲノム解読後の NIH Biomedical Research Roadmap に基づいた米国の医学・生命科学への投資戦略をみれば、近い将来に、創薬のパラダイムシフトが起こることが想定された。こうした動向のなかで、マイクロアレイ解析による遺伝子発現データを初め

とするゲノムワイドなデータや Gene Ontology などの生命情報の概念整理, 電子化された生命情報など, 様々の情報から, 創薬のターゲットの可能性をもった遺伝子及びパスウェイを探索するためのバイオインフォマティクス技術が重要になっていた.

## 2. 研究の目的

遺伝子ネットワーク情報を用いて, 創薬のターゲットになるパスウェイを探索するためのバイオインフォマティクスの技術を開発することを目的とした.

## 3. 研究の方法

(1) 遺伝子ノックダウンによる遺伝子発現データ及び薬剤応答による遺伝子発現データから, 薬剤が経時的に影響を及ぼしていくパスウェイを推定するバイオインフォマティクス技術を開発する.

(2) 動的遺伝子ネットワークのシミュレーションとそのネットワーク構造等の情報から, 創薬ターゲット遺伝子及びパスウェイを絞り込むためのバイオインフォマティクス技術を開発する.

## 4. 研究成果

(1) ベイジアンネットワークは, 複数の確率変数間の依存関係を記述することが可能な多変量統計モデルの一種である. 遺伝子発現を確率変数とみなし, ベイジアンネットワークを用いてその依存関係を推定することにより, 遺伝子間の発現の依存関係をネットワークとしてモデル化することができる. ベイジアンネットワークは, 確率変数をノード, 確率変数間の依存関係を有向エッジで表した非閉路有向グラフとして表現される. 数学的には, 確率変数間の矢印の向きに従うマルコフ連鎖律を仮定した非閉路有向グラフにより確率変数の同時確率が各確率変数のグラフ上の親確率変数所与の下での条件付き確率の積に分解されることがベイジアンネットワークにおいて本質的である. ある非閉路有向グラフを仮定することでこの分解は一意に定まる. したがって, 確率変数間の依存関係が未知の状況下において, 最適な非閉路有向グラフは, その構造により得られる分解に基づく統計モデルの汎化能力を最大にするものとして決定される. この最適化問題を, ベイジアンネットワークの構造推定問題とよんでいる. ベイジアンネットワークの構造推定問題を, 遺伝子発現データに基づき解くことにより, 遺伝子間の転写制御ネットワーク構造推定を行う方法が本研究代表者等のグループにより開発されており, 本研究ではこれを利用している.

ダイナミックベイジアンネットワークは,

時系列観測データに基づき確率変数間の依存関係を記述するためのベイジアンネットワークの一つの拡張である. 統計モデルとしては, 統計的時系列データ解析手法の一つである一次の自己回帰モデルを用いてベイジアンネットワークの条件付き確率を表現したモデルに相当する. 通常のベイジアンネットワークが  $v$  構造という限定された状況下でしか確率変数間の因果関係を推定できないことに比べ, 時間の情報を考慮に入れることで Granger の意味での確率変数間の因果関係を捉えることができるモデルである.

本研究では, まず, このダイナミックベイジアンネットワークにノンパラメトリック回帰を融合させた方法により, 転写制御ネットワーク構造を時系列データから抽出し, 創薬ターゲット探索とバリデーションを行う統計的解析方法及び計算方法を検討した [7,9].

そして, ヒト血管内皮細胞(HUVEC)においてアポトーシスを誘導させた時系列マイクロアレイデータからダイナミックベイジアンネットワークにノンパラメトリック回帰を組み合わせた方法で, 遺伝子ネットワークを推定することに成功した. また, siRNA による遺伝子ノックダウンマイクロアレイデータを用いて, このネットワークを比較解析した. その結果, この遺伝子ネットワーク推定による方法が薬剤ターゲットの探索に有効であることが示唆された [5].

さらに, HUVEC を用いた, 400 の遺伝子に対するノックダウンマイクロアレイデータ, 薬剤 Fenofibrate を投与した HUVEC 時系列遺伝子発現データ, 及びタンパク質相互作用データから, 動的に変化する転写ネットワークとシグナル伝達パスウェイを結びつけたネットワークを抽出する方法を開発し, この薬剤に反応した自己分泌シグナル伝達パスウェイを推定することに成功した [1]. 図 1 はその方式の概略である.

また, 新たな方式として, 状態空間モデルに基づいたネットワーク解析技術を開発した. 状態空間モデルは, 統計的時系列モデルの一種で, 気象現象などの物理学の分野において予測モデルを構築するために用いられている. 通常, 多時点数データを対象とし, オブジェクト対オブジェクトの関係を直接的に表現した一次のベクトル自己回帰モデルを考える. 状態空間モデルにおけるオブジェクトに遺伝子を対応させ, そのオブジェクトの量的観測データを遺伝子発現データとしてとらえる. 本研究で用いる状態空間モデルでは, 時刻  $t$  において観測された遺伝子発現データを  $y_t$  と表し, まとめて  $Y = \{y_t\}_{t=1, \dots, T}$  とおく. ベクトル  $y_t$  の次元を  $m$  と表す. つまり,  $m$  個の遺伝子の時刻  $t$  における発現値が  $y_t$  には並んでいることになる. 状

態空間モデルは、次の2式で定義される：

$$y_t = Hx_t + u_t \quad (a)$$

$$x_t = Fx_{t-1} + w_t \quad (b)$$

ここで、 $x_t$  は  $k$  次元状態ベクトル ( $k < m$ )、 $H$  は  $m \times k$  次観測行列、 $F$  は  $k \times k$  次システム行列、 $u_t$  と  $w_t$  はそれぞれ観測ノイズ、システムノイズであり、多変量正規分布  $N(0, R)$ 、 $N(0, \Lambda)$  に従う。ここで、状態ベクトル  $\{x_t\}_{t=1, \dots, T}$  は観測されない変数であり、パラメータ  $H, F, R$  と共にデータから推定する必要がある。また、観測ノイズの分散共分散行列  $R$  は、対角行列であるとする。(a)式は、観測モデル、(b)式はシステムモデルとよばれる。状態空間モデルによる細胞内での遺伝子発現のシステムのモデリングでは、例えば、転写因子が多くの遺伝子を制御していることから、状態空間モデルの状態ベクトルは、リン酸化等により活性化された転写因子の濃度プロファイルのように解釈をすることもできる。

本研究が対象とする時系列量的データは、薬剤反応などによる時系列遺伝子発現データであり、シミュレーションモデルに取り入れる遺伝子の数（数百から数千）に比べて、計測時点数（10 時点程度）が非常に少ない超高次元短時系列という特性がある。そのため、通常の自己回帰モデルに含まれるパラメータを推定することは技術的・理論的に困難である。本研究代表者等のグループは、状態空間モデルにおいて、多数の遺伝子の発現状態を、少数の潜在変数の発現状態へ縮約（次元圧縮）することにより、パラメータ推定の困難を克服する一連の技術を既に開発しており、この技術を本研究に利用している。

時系列マイクロアレイデータ解析に適應した状態空間モデルを用いて、異なる細胞株間や、薬剤投与条件に差のある細胞間の遺伝子制御システムの差異に関わる遺伝子群を、シミュレーションモデルからの予測と観測結果の差異から抽出する方法を開発した。この方法を、ヒト小気道上皮細胞(SAEC)をEGFで刺激した場合と、EGF刺激及びGefitinib投与の場合の、2種類の時系列マイクロアレイデータに適用することにより、薬剤投与により制御関係に変異が生じた遺伝子群を抽出することに成功した[4]。図2は、EGF+GefitinibのシステムとEGFのみのシステムの違いを示す遺伝子を探索した一例である。

(2) 遺伝子オントロジーを使って遺伝子発現データを解析する統計的方法を構築した[6]。この方式を、MetaGP (MetaGeneProfiler)と

いうソフトウェアとして実現した。この遺伝子オントロジーを利用する方法は、マイクロアレイ遺伝子発現データを用いて、その発現データから、その背景にある細胞内メカニズムに関係深い遺伝子機能を予測するものである。まず  $p$  値などを用いてランク付けする既存の方法は、偽陽性の予測など、様々な深刻な問題を抱えていることを指摘した。特に  $p$  値によるランク法は、背景となっている細胞内メカニズムとは関係のない遺伝子群を統計的に有意差があるように判定することが多いことがわかった。本研究では、まずこうした問題を解決し、統計的理論に裏打ちされた、オントロジーという知識の分類と関係付けを用いた新しい方式を開発した。そして Monte Carlo 実験を行うことで、既存法の問題点が、本方式では解決されていることを実証し、その有効性を示した。

この方法に基づいて、ケース・コントロールのサンプル間で活動度の変化する転写因子を、マイクロアレイで取得された遺伝子発現データから、仮説検定により発見するための手法を開発した。その手法を適用することにより、転写因子の活動度が、各実験間で変動する様子を、網羅的に俯瞰するマップを得ることが可能となった。我々は、この手法を、中枢神経系において脱髄したミエリン鞘の再髄鞘化を促す働きをもつ漢方薬を、マウスに投与したケース・コントロール実験から得られたデータに対して適用し、そのターゲットパスウェイ群の推定を行った[8]。一般に漢方薬は、単一の化合物ではなく、混合した生薬の抽出物からなるため、作用領域は複数あると考えられる。そのため、通常行われるように単に少数の高発現差のある遺伝子群を列挙しただけでは、その作用機序や生体システムの働きを推察することは困難である。それに対し、本研究で開発した方法では、転写因子活動度を含め、gene ontology 等、生体システムの働きの特徴を要約する、さまざまな機能的アノテーションと共に適用可能であり、システムの機能を、より総体的に理解するための情報を抽出することができる。さらに、漢方薬の黄連解毒湯をマウスに投与して解析したマイクロアレイデータに対して、Cell Illustrator, MetaGP, 及びパスウェイデータベース TRANSPATH を統合して用いるバイオインフォマティクス技術を構築し、パスウェイ解析を行い、この漢方薬の作用しているパスウェイの抽出を行った[2]。図3は、その解析法を[8]のデータで行ったときのようなすである。併せて、病気及び薬剤応答に関するパスウェイ情報を文献から抽出する技術を開発した[3]。

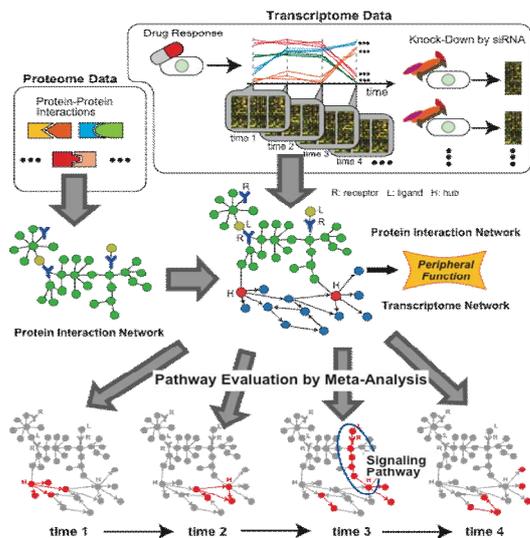


図1：薬剤応答ネットワークの抽出スキーム。Fenofibrate を投与した時系列遺伝子発現データからダイナミックベイジアンネットワークにより推定したネットワークと PPI ネットワークのつながりを推定している。

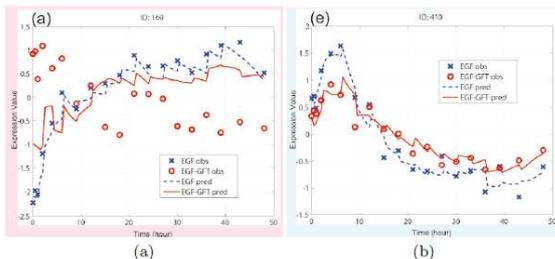


図2：(a) は、ID169 遺伝子の時系列プロファイル。赤は EGF+Gefitinib、青×は EGF のみを投与したときのもの。赤実線と青破線は状態空間モデルによるシミュレーション結果。観測と予測が一致していない。一方(b)の遺伝子 ID160 では、状態空間モデルによる予測と観測結果が合致している。このことから、遺伝子 ID169 は、Gefitinib からの影響により、ネットワークが変化していると考えられる。一方、遺伝子 ID410 には影響がないと考えられる。こうして、薬剤投与によってシステムとしての変化をしてしまう遺伝子群を抽出している。

### 5. 主な発表論文等

[雑誌論文](計9件)

[1] Tamada, Y., Araki, H., Imoto, S., Nagasaki, M., Doi, A., Nakanishi, Y., Tomiyasu, Y., Yasuda, K., Dunmore, B., Sanders, D., Humphreys, S., Print, C., Charnock-Jones, D.S., Tashiro, K., Kuhara, S., Miyano, S. Unraveling dynamic activities of autocrine

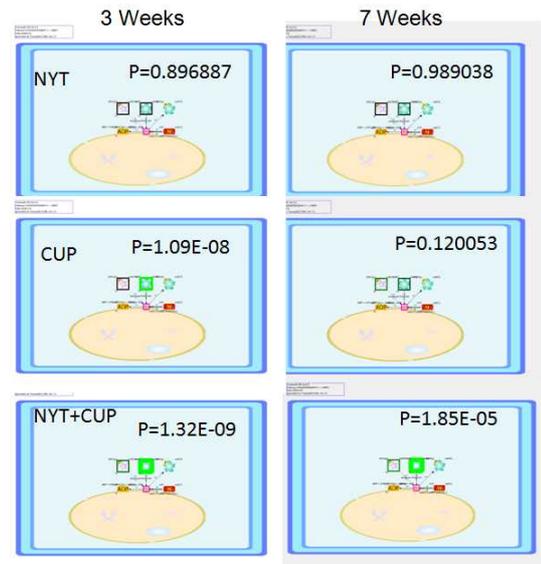


図3：Cuprizone (CUP)を使ってマウスのミエリン鞘を脱髄鞘させたもの、人参養栄湯(NYT)を投与したもの、CUP で脱髄鞘してNYTを投与した3つの場合のマウスを使っている。3週後、7週後にマイクロアレイ解析を行い、どのようなパスウェイが有意に表れているかを解析し、Cell Illustrator を使って解析結果を表示している。各条件間でパスウェイの有意差とどの遺伝子が効いているかを一目で比較でき、CUPで脱髄鞘したものが、NYTの投与で、再髄鞘化していくプロセスに現れるパスウェイが見える。

pathways that control drug-response transcriptome networks. Pacific Symposium on Biocomputing. 14: 251-263 (2009). (査読有)

[2] Watanabe-Fukuda, Y., Yamamoto, M., Miura, N., Fukutake, M., Ishige, A., Yamaguchi, R., Nagasaki, M., Saito, A., Imoto, S., Miyano, S., Takeda, J., Watanabe, K. Orengedokuto and berberine improve indomethacin-induced small intestinal injury via adenosine. J. Gastroenterology. 44(5): 380-389 (2009). (査読有)

[3] Hishiki, T., Tamada, I., Okubo, K. Gene-L'EXPO: a tool to extract knowledge from transcriptomes and find 'literature-sparse' relationships between genes and tissues. Proc. AMIA Annual Symposium 2008. 313-317 (2008). (査読有)

[4] Yamaguchi, R., Imoto, S., Yamauchi, M., Nagasaki, M., Yoshida, R., Shimamura, T., Hatanaka, Y., Ueno, K., Higuchi, T., Gotoh, N., Miyano, S. Predicting differences in gene regulatory systems

- by state space models. *Genome Informatics*. 21:101-113 (2008). ( 査読有 )
- [5] Affara, M., Dunmore, B., Savoie, C.J., Imoto, S., Tamada, Y., Araki, H., Charnock-Jones, D.S., Miyano, S., Print, C. Understanding endothelial cell apoptosis: What can the transcriptome glycome and proteome reveal? *Philosophical Transactions of Royal Society*. 362(1484): 1469-1487 (2007). ( 査読有 )
- [6] Gupta, P.K., Yoshida, R., Imoto, S., Yamaguchi, R., Miyano, S. Statistical absolute evaluation of gene ontology terms with gene expression data. *Lecture Notes in Bioinformatics*. 4463:146-157 (2007). ( 査読有 )
- [7] Imoto, S., Tamada, Y., Savoie, C.J., Miyano, S., Analysis of gene networks for drug target discovery and validation. *Methods in Molecular Biology*. 360:33-56 (2007). ( 査読有 )
- [8] Yamaguchi, R., Yamamoto, M., Imoto, S., Nagasaki, M., Yoshida, R., Tsujii, K., Ishiga, A., Asou, H., Watanabe, K., Miyano, S. Identification of activated transcription factors from microarray gene expression data of Kampo-medicine treated mice. *Genome Informatics*. 18: 119-129 (2007). ( 査読有 )
- [9] Tamada, Y., Imoto, S., Miyano, S. Estimating gene networks from gene expression data utilizing biological information. *Proceedings of The Institute of Statistical Mathematics*. 54(2): 333-356 (2006). ( 査読有 )

[学会発表](計7件)

1. 宮野 悟 , Unraveling dynamic activities of autocrine pathways that control drug-response transcriptome networks, The 14<sup>th</sup> Pacific Symposium on Biocomputing, 2009年1月6日, Fairmont Orchid, Big Island of Hawaii, USA.
2. 宮野 悟 , Predicting differences in gene regulatory systems by state space models, The 19<sup>th</sup> International Conference on Genome Informatics, 2008年12月2日, Gold Coast, Australia.
3. 日紫喜光良 , Gene-L'EXPO: a Tool to Extract Knowledge From Transcriptomes and Find 'Literature-Sparse' Relationships Between Genes and Tissues, AMIA Annual Symposium 2008, 2008年11月10日, Washington D.C., USA.
4. 宮野 悟 , Computational Systems Biology, 2007 INDY Bioinformatics Conference, 2007年6月1日, Indiana University Medical School, Indianapolis, USA.
5. 宮野 悟 , Computational Challenges for Top-Down Modeling and Simulation of Biological Pathways, The 3<sup>rd</sup> International *E. Coli* Alliance Conference on Systems Biology, 2006年11月1日, International Convention Center Jeju, Republic of Korea.
6. 宮野 悟 , Computational Systems Biology and Its Applications to Drug Target Gene Discovery, INTERNATIONAL CONGRESS OF IMMUNOGENOMICS AND IMMUNOMICS 2006, 2006年10月11日, Budapest Congress & World Trade Center, Budapest, Hungary.
7. 宮野 悟 , Gene Networks for Drug Target Pathway Discover, The First Toxicomics Forum, 2006年7月5日, 名古屋国際会議場 センチュリーホール .

[その他]

MetaGeneProfiler による解析サービス  
サイト : <http://metagp.ism.ac.jp/>

## 6. 研究組織

### (1) 研究代表者

宮野 悟 (MIYANO SATORU)  
東京大学・医科学研究所・教授  
研究者番号 : 50128104

### (2) 研究分担者

井元 清哉 (IMOTO SEIYA)  
東京大学・医科学研究所・准教授  
研究者番号 : 10345027  
長崎 正朗 (NAGASAKI MASAO)  
東京大学・医科学研究所・助教  
研究者番号 : 90396862  
日紫喜 光良 (HISHIKI TERUYOSHI)  
東邦大学・理学部・准教授  
研究者番号 : 30324271  
山口 類 (YAMAGUCHI RUII)  
東京大学・医科学研究所・講師  
研究者番号 : 90380675  
吉田 亮 (YOSHIDA RYO)  
情報・システム研究機構・統計数理研究所・助教  
研究者番号 : 70410126