

細胞分化クロストークのモデル化と 細胞分化クロストーク遺伝子の推定手法

吉澤 陽 志^{†1} 瀬尾 茂 人^{†2}
竹中 要 一^{†2} 松田 秀 雄^{†2}

本研究では細胞分化クロストークの数理モデル化を提案する。細胞分化は ES 細胞や iPS 細胞を代表に近年注目を浴びている研究対象である。細胞分化における重要な遺伝子制御ネットワークとして細胞分化クロストークがある。しかし細胞分化クロストークに対して生物学的実験による研究は行われているが、情報科学的アプローチによる推定手法は確立していない。そこで我々は細胞分化クロストークの数理モデル化を提案した。提案手法をマウスの細胞分化の過程を観測した遺伝子発現量データに対して適用し、有効性を検証した。

A Crosstalk Model and an Estimation Method to Find Crosstalk Genes on Cell Differentiation

KIYOSHI YOSHIZAWA,^{†1} SHIGETO SENO,^{†2}
YOICHI TAKENAKA^{†2} and HIDEO MATSUDA^{†2}

We propose a mathematical model of crosstalks on cell differentiation. Cell differentiation is the process that a stem cell (such as an ES or iPS cell) changes to a specific type of cell, and attracts lots of attention from researchers of medicine and biology. Crosstalks on cell differentiation are a gene regulatory network that plays a role to determine the specialized cell type. To analyze the crosstalk, biological approaches have been adopted but information science approaches have not been established. In this study, we demonstrated the effectiveness of our model by applying it to a differentiation process to bone cell.

^{†1} 富士通株式会社

Fujitsu Limited

^{†2} 大阪大学大学院情報科学研究科バイオ情報工学専攻

Department of Bioinformatic Engineering, Graduate School of Information Science and Technology, Osaka University

1. はじめに

世の中にはネットワークと呼ばれるものが数多く存在する。身近な例ではインターネットや鉄道の交通網、人間の社会的ネットワーク等があげられる。これらを抽象化し、グラフとして表現することによる数理モデル化が行われてきた。また生物の細胞内にも代謝反応ネットワーク、シグナル伝達系ネットワーク、遺伝子制御ネットワークといったネットワークが存在している。そのうち遺伝子制御ネットワークは、遺伝子をノードとした有向グラフであり、辺は遺伝子間の制御関係を表現している^{1),2)}。制御関係とは有向辺の始点ノードに対応する遺伝子の発現量が細胞内で変化することに起因して、終点ノードに対応する遺伝子の発現量が変化することを意味している。遺伝子制御ネットワークの構造を解明するための研究の1つに、遺伝子発現量のデータから数理モデルに基づいてネットワーク構造の推定を行う研究があげられる。このような数理モデルを利用した推定手法の例としてはグラフィカルガウシアンモデリング³⁾やベイジアンネットワーク⁴⁾、微分方程式モデル⁵⁾等が提案されている⁶⁾。

多細胞生物は一般に特定の役割を有する複数の種類の細胞から構成されている。以下、細胞の種類のことを細胞種と呼ぶ。細胞の中には、別の細胞種に変化する能力を有するものがあり、このような能力を持つ細胞を幹細胞と呼ぶ。またこの変化の過程を細胞分化と呼ぶ¹⁾。細胞分化は近年注目を浴びている生命現象の1つであり、ES細胞⁷⁾やiPS細胞⁸⁾を代表に医療分野への応用が期待されている。それ以外にも、糖尿病に関わる脂肪細胞や骨粗しょう症に関わる骨芽細胞等に関する多くの研究が報告されている^{9)–11)}。これらの細胞内に存在する遺伝子制御ネットワークは全細胞種に共通するもの、各細胞種に固有のもの、複数の細胞種に共通するものがある。その中で複数の細胞種に共通する例としてあげられるものが細胞分化に関連する遺伝子制御ネットワークである。細胞分化に関連する遺伝子制御ネットワークの研究は生物学的実験によってなされている¹²⁾。計算機を用いた研究としては複数の遺伝子発現量データを利用した研究は行われている^{13),14)}が、複数の細胞種間での変化を制御する遺伝子制御ネットワークを推定する研究はほとんど行われていない。生物学的実験にはコスト等の問題があり、情報科学的アプローチが求められている。

細胞分化に関連する遺伝子制御ネットワーク上で、細胞分化における重要な働きをしていると考えられるものが細胞分化クロストークである^{12),15)}。細胞分化クロストークは、細胞分化時に目的とする細胞種への分化を決定づける働きとその他の細胞種への分化を抑制する働きを行っている。細胞分化クロストークにはその遺伝子制御ネットワークの働きを左

右する核となる遺伝子，すなわち細胞分化クロストーク遺伝子が存在すると考えられている^{12),15)}．細胞分化は時間の経過ともなう発現量の変化が重要となる現象であるため，本研究では，時系列データから細胞分化クロストーク遺伝子を推定することを目的とする．

細胞分化は複数の細胞種にまたがる現象であるため，複数の細胞種からの遺伝子発現量のデータを各細胞種の違いを反映して結合させたいうで推定を行う必要がある．そのため既存の数値モデルを用いて遺伝子制御ネットワークを推定するのみでは細胞分化クロストークを推定することは難しく，細胞分化クロストークの特殊性をふまえた新たな数値モデルが必要とされる．

そこで本研究では細胞分化クロストークのモデルを構築し，そのモデルを用いて細胞分化クロストーク遺伝子を求めるための解法を提案する．なおこの細胞分化クロストークの問題は最短経路探索問題に帰着することにより解くことが可能である．そのため最短経路探索法の Dijkstra のアルゴリズム^{16),17)} と Yen のアルゴリズム^{18),19)} を組み合わせたアルゴリズムを基にしている．

2. 細胞分化クロストーク

2.1 細胞分化と細胞分化クロストーク

多細胞生物の生命は受精卵という 1 つの細胞から始まる．受精卵は活発に細胞分裂を繰り返し細胞の数を増やしていく．ここで行われることが細胞分裂のみであれば全細胞が受精卵と同じ種類の細胞となる．しかし，人間の体では骨，筋肉，脂肪というように部位，機能によって細胞の種類は異なっている．つまり，細胞が特殊な構造や機能を獲得する過程が存在しており，このことを細胞分化という¹⁾．細胞分化を行う代表的な細胞の例としてはすべての血液細胞に分化する造血幹細胞や図 1 の骨芽細胞，軟骨細胞，脂肪細胞等に分化する間葉系幹細胞が知られている¹⁾．また様々な細胞へと分化する機能を持つ細胞として ES 細胞⁷⁾ や iPS 細胞⁸⁾ がある．基本的にすべての細胞は同じ遺伝子セットを持つため，細胞が多様化するには各細胞種ごとに遺伝子が差異的に発現する必要がある．よって細胞分化は細胞分化の際に個々の細胞種で特徴的である遺伝子制御ネットワークによって制御されているといえる．

細胞分化を制御するには目的とする細胞種への分化を促進すること以外に他の細胞種への分化を抑制することが重要となる．このように，ある細胞種への分化の促進と別の細胞種への分化の抑制が，遺伝子制御ネットワーク上で同時に起こることを細胞分化クロストークと呼び，目的とする細胞種への分化を決定づける働きをしていると考えられている^{12),15)}．事

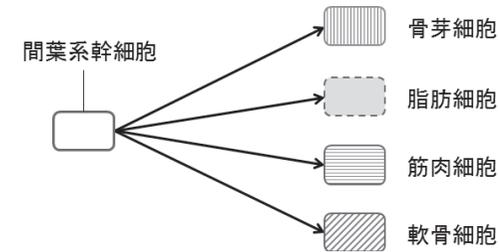


図 1 間葉系幹細胞の細胞分化

Fig. 1 Cell differentiations of the mesenchymal stem cell.

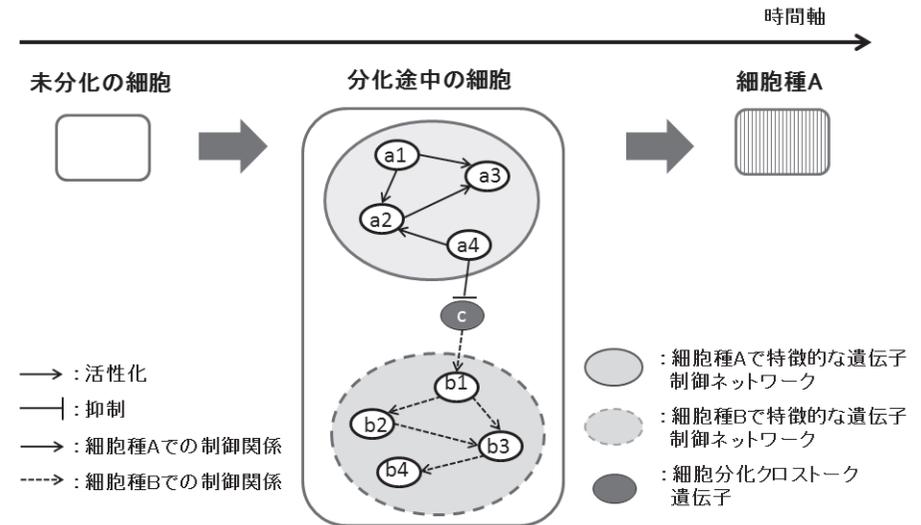


図 2 細胞種 A へ分化する際のクロストーク概念図

Fig. 2 Diagram of a crosstalk of cell differentiation.

実，細胞分化の過程において一方では目的とする細胞種への分化に促進的に働き，他の細胞種への分化に対し抑制的に働く現象が存在することが確認されている^{12),15)}．例として図 2 を用いて細胞分化クロストークの説明を行う．図 2 は未分化である細胞が細胞種 A と細胞種 B に分化可能であるという条件下で，細胞種 A へ分化する場合の細胞分化クロストークの構造を表している．特徴的な遺伝子とは細胞分化の際に重要な働きをする遺伝子のことを

指し、生物学的実験によりいくつかの遺伝子が知られている^{9),10)}。図2中の分化途中の細胞では、遺伝子 a4 が遺伝子 c を経由して遺伝子 b1 を抑制することにより細胞種 A での特徴的な遺伝子制御ネットワークから細胞種 B での特徴的な遺伝子制御ネットワークに対し抑制する制御関係ができています。つまり、細胞種 A で特徴的な遺伝子は機能し続け細胞種 B で特徴的な遺伝子が機能しなくなることにより、細胞種 A への分化が促進され細胞種 B への分化が抑制されることとなる。この構造が細胞分化クロストークである。

また細胞分化クロストークにはその働きを左右する核となる遺伝子が存在すると考えられている^{12),15)}。核となる遺伝子によりどの細胞種へ分化するかを選択が行われる。この核となる遺伝子を細胞分化クロストーク遺伝子と呼ぶ。細胞分化クロストーク遺伝子を操作することにより細胞分化を人工的に制御することが可能となるため発見が期待されている。また細胞分化クロストーク遺伝子を発見することは細胞分化が関係する病気の治療の出発点となりうる。細胞分化が関係する病気の例としては骨粗しょう症が知られている²⁰⁾。骨粗しょう症は骨芽細胞の働きの低下により骨がもろくなる病気である。人工的に脂肪細胞への分化の抑制と骨芽細胞への分化を促進することが可能となれば治療法につながると考えられるため、細胞分化クロストーク遺伝子の発見が重要となる。本研究では、以下の条件をすべて満たす遺伝子 x を細胞分化クロストーク遺伝子の定義とする^{12),15)}。

前提 細胞種 A と細胞種 B へ分化可能な細胞が細胞種 A へ分化する。

- (1) 遺伝子 x は細胞分化クロストークに関与する遺伝子である。
- (2) 遺伝子 x は細胞種 A において細胞種 A で特徴的な遺伝子から制御を受けている。
- (3) 遺伝子 x は細胞種 B において細胞種 B で特徴的な遺伝子を制御している。
- (4) (2) と (3) における制御関係が、(2) が活性化かつ (3) が抑制、または、(2) が抑制かつ (3) が活性化となる。

例として図2では遺伝子 c が細胞分化クロストーク遺伝子となる。

2.2 既存手法による細胞分化クロストーク推定の困難さ

生物学的実験により細胞分化クロストーク、また細胞分化クロストーク遺伝子を解明するための試みの1つとして、標的遺伝子の不活性化(ノックダウン)方法の一種である RNA interference (RNAi) 法²¹⁾ 等による遺伝子の発現をコントロールする実験があげられる。以下では RNAi 法に基づく制御関係推定の実験に関して述べる。遺伝子 g の発現を抑制したとき、発現量が低下する遺伝子は遺伝子 g から正の制御を、発現量が上昇する遺伝子は負の制御を受けていると考えられる。

分化過程においてこの実験を行うことで遺伝子の制御関係を確認でき、細胞分化クロス

トーク遺伝子の発見につなげることが可能である。ただし遺伝子をノックダウンすることで明らかにできる制御関係はノックダウンした遺伝子の制御を受ける下流側に制限され、ノックダウンした遺伝子を制御する上流側の制御関係は明らかにできない。それに加え RNAi 法の実験は1つ1つに大きな手間がかかるため、これらの実験には時間的・金銭的に大きなコストが発生する。よって情報科学的アプローチにより細胞分化クロストーク遺伝子を推定し、あらかじめ実験対象を絞ることが必要となってくる。

情報科学的アプローチとして、遺伝子発現量のデータから計算機を用いて遺伝子制御ネットワークを推定する手法は数多く提案されている。しかし、その多くは1種類の細胞から取得された遺伝子発現量データ上の遺伝子制御ネットワークを推定するものである。複数の独立した実験から得られた遺伝子発現量データを利用した遺伝子制御ネットワークの推定も研究されている^{13),14)} が、実験の途中で分岐する細胞分化クロストークを対象とする推定手法は確立していない。細胞分化クロストーク遺伝子は目的とする細胞種への分化の際の遺伝子制御ネットワークとその他の細胞種への分化の際の遺伝子制御ネットワークとを比較することで、初めて発見が可能となる。つまり異なる細胞種へ分化する際の複数の遺伝子発現量のデータを各細胞種の違いを反映して結合させ、複数の遺伝子発現量のデータにまたがった制御関係を推定することが必要となる。そのため既存の数理モデルに基づき遺伝子制御ネットワークを推定するのみでは細胞分化クロストークを推定することは不可能である。よって細胞分化クロストークを計算機で推定するためには新たに細胞分化クロストークのモデルが必要となる。

3. 提案モデルと推定手法

本章では細胞分化クロストーク推定問題を解くための数理モデル化について述べる。なお以下では細胞種 A と細胞種 B に分化可能である細胞が、細胞種 A に分化するという前提のもとで記述する。

3.1 細胞分化上の遺伝子制御ネットワークのグラフ表現

生物学の問題を情報科学的アプローチにより解くためには、まず情報科学的な定義を行う必要がある。本節では遺伝子制御ネットワークとグラフ表現の対応付けを行う。

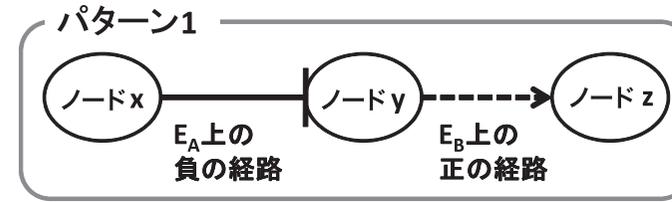
遺伝子制御ネットワークは有向グラフ $G = (V, E)$ として表すことができる。遺伝子の集合はノード集合 $V = \{v_1, v_2, \dots, v_n\}$ に対応し、遺伝子間の制御関係の集合が辺集合 $E = E_A \cup E_B$ に対応する。ここで辺 $e \in E_A$ が細胞種 A 上での遺伝子間の制御関係を表し、辺 $e \in E_B$ が細胞種 B 上での制御関係を表す。そのためグラフ G は、頂点 $v_i, v_j \in V$

($1 \leq i \leq n, 1 \leq j \leq n, i \neq j$) 間に E_A と E_B に属する複数の辺を持つ多重グラフとなる．辺 $e \in E$ は実数の重み $w(e)$ を持ち、遺伝子間の制御関係の強さを表す ($w(e)$ の詳細については 4.1 節で述べる)．なお辺の重みの絶対値 $|w(e)|$ が小さいほど制御関係が強いことを意味する．辺 $e = (v_i, v_j)$ の重み $w(e)$ は v_i が v_j を活性化させるときは正の実数となり、抑制するときには負の実数となる．

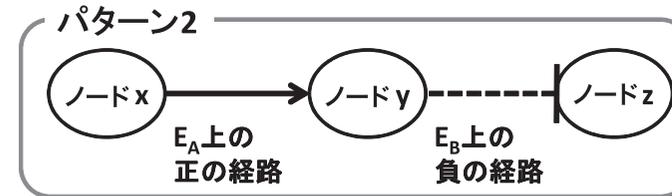
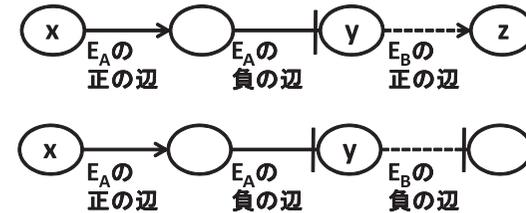
細胞種 A の特徴的な遺伝子集合を出発地の集合 S 、細胞種 B の特徴的な遺伝子集合を目的地の集合 D とする．2 遺伝子 v_i, v_j 間の間接的な制御関係、すなわち 2 遺伝子間に複数の遺伝子を介した制御関係は経路 $path(v_i, v_j)$ として表す．経路 $path(v_i, v_j) = (e_1, e_2, \dots, e_n)$ の距離は $dist(path(v_i, v_j)) = \sum_{e \in path(v_i, v_j)} |w(e)|$ により定義され、2 遺伝子間の間接的な制御の強さを表現する．また特にあるノード v_k を経由する経路を表現する場合は $path(v_i, v_k, v_j)$ とする．

3.2 細胞分化クロストークの提案モデル

2.1 節で述べたように、細胞分化クロストークは細胞種 A の特徴的な遺伝子から細胞分化クロストーク遺伝子を仲介し、細胞種 B の特徴的な遺伝子への抑制する制御関係を指す．また細胞種 A で特徴的な遺伝子から細胞分化クロストーク遺伝子への制御は細胞種 A 上で行われ、細胞分化クロストーク遺伝子から細胞種 B の特徴的な遺伝子への制御は細胞種 B 上で行われる．その定義に従い細胞分化クロストークをモデルとして表したものが図 3 である．細胞分化クロストークを表現する経路は、細胞分化クロストーク遺伝子が細胞種 A で活性化されるか抑制されるかによって 2 つのパターンに分類される．この 2 つのパターンを図 3 に示す．図 3 のノード x は細胞種 A の特徴的な遺伝子、ノード y は細胞分化クロストーク遺伝子、ノード z は細胞種 B の特徴的な遺伝子である．また、無印のノードは x から y を経由し z に至る経路上にあり、細胞種 A と細胞種 B のいずれにも特徴的であることが知られていない遺伝子である．各ノードを結ぶ経路は複数のノードを経由した経路を意味する．パターン 1 ではノード x からノード y へ E_A の辺を通る負の経路で結ばれ、ノード y からノード z へは E_B の辺を通る正の経路となる．ここで正の経路とは、含まれる負の辺が偶数個となることで活性化が行われることを意味する．同様に負の経路とは、含まれる負の辺が奇数個となることで抑制が行われることを意味する．経路上の辺の数が 1 であれば直接的な活性化または抑制を表し、辺の数が 2 以上であれば間接的な活性化または抑制を表す．パターン 2 ではノード x からノード y へは E_A の辺からなる正の経路をとり、ノード y からノード z へ E_B の辺からなる負の経路をとる．ともにノード x からノード z への総合的な経路は負となり、ノード y で通る辺が $e \in E_A$ から $e \in E_B$ へ変わる．つま



パターン1の例



パターン2の例

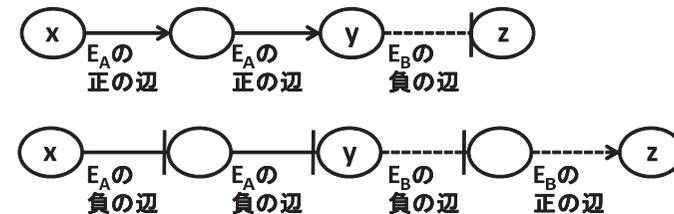


図 3 細胞分化クロストークのグラフ上でのモデル

Fig. 3 Scheme of cell differentiation crosstalk with graph theory.

り, 図 3 の細胞分化クロストークモデルは以下の条件を満たす経路である.

制約条件 1 経路中に含む負の辺が奇数個.

$$\prod_{e \in \text{path}(s,d)} w(e) < 0, s \in S, d \in D$$

制約条件 2 経路中の辺は初め $e \in E_A$ を通り後半 E_B を通る. E_A から E_B への遷移は 1 度だけ行われる.

$$1 \leq i < j \leq n, e_i \in E_A, e_j \in E_B$$

辺の重みの絶対値 $|w(e)|$ が小さいほど制御関係が強いことを意味していることから, 制約条件を満たす経路のうち出発地 $s \in S$ から目的地 $d \in D$ への経路の距離 $\text{dist}(\text{path}(s,d))$ が小さい経路ほど制御関係が強い細胞分化クロストークであるといえる.

3.3 提案モデルを用いた細胞分化クロストーク遺伝子推定手法

3.2 節で定義した制約条件に従い, 最短経路探索法を基にしたアルゴリズムを解くことにより制御関係が強い細胞分化クロストークの推定が可能となる. なお, 最短経路が必ずしも生物学上の真の細胞分化クロストークとは限らないため, 複数の経路を求めることが望まれる. そこで, 手法に次の条件を加える.

制約条件 3 K 個の最短経路を求める (K は自然数)

提案解法では以下の 4 ステップで制約条件を満たす経路を求める.

出発地 $s \in S$ から各ノードへの, 負の辺を偶数個含む正制御となる経路と負の辺を奇数個含む負制御となる経路の両者を持つノードを考え, 正制御と負制御のそれぞれにおいて最短経路を求める. Step1 ではそれぞれに対応するノードを作成する. これにより選択的に負制御の経路の探索を可能とする. Step2 と 3 において出発地から目的地へ複数のグラフをまたいだ経路探索を行う. 経路探索手法は Dijkstra のアルゴリズム^{16),17)} と Yen のアルゴリズム^{18),19)} を組み合わせたアルゴリズムを基にしている. Step4 では求めた経路から細胞分化クロストーク遺伝子と見なすノードを取り出す.

Step3 で利用する Yen のアルゴリズムは論文 18) で提案された loop を含まない k 番目の最短経路を求めるアルゴリズムである. このアルゴリズムの概要は以下のとおりである. すでに求めてある 1 番目から $k-1$ 番目の各最短経路に対して, 1) 経路中の各辺を順番に除去し, 2) 除去辺以前の経路中の各ノードから目的地までの最短経路を探索し, 3) 出発地から各ノードまでの最短経路と合成することにより k 番目の最短経路候補を導出する. 4) 最も距離の短い候補を k 番目の最短経路として出力する.

各ステップの具体的な手順を以下に示す.

入力 グラフ $G = \{V, E\}$, 出発地 s とする遺伝子, 目的地 d とする遺伝子

出力 順位付けされた細胞分化クロストーク遺伝子に相当するノードのリスト

Step1 選択的に負制御の経路を探索可能とするためにグラフ $G_{new} = (V_{new}, E_{new})$ を作成する. Step2 以降ではグラフ G_{new} を入力として用いる.

Step1.1 ノード集合 V を複製し, 2 セットのノード集合 $V_{posi} = \{v_{1posi}, v_{2posi}, \dots, v_{nposi}\}$, $V_{nega} = \{v_{1nega}, v_{2nega}, \dots, v_{nnega}\}$ を作成し, $V_{new} = V_{posi} \cup V_{nega}$ とする. それぞれのノード集合中の各ノード $v_{iposi} \in V_{posi}$, $v_{inega} \in V_{nega}$ はノード $v_i \in V$ と対応する. つまり元の遺伝子との対応関係は保存される.

Step1.2 ノード集合 V と辺集合 E の関係性を基に新たに辺集合 $E_{new} = E_{new,A} \cup E_{new,B}$ を作成する. $E_{new,A}$ は次のように定義される.

$$E_{new,A} = E_{posi,A} \cup E_{nega,A}$$

where

$$E_{posi,A} = \{(v_{iposi}, v_{jposi}) \mid v_{iposi}, v_{jposi} \in V_{posi} \wedge (v_i, v_j) \in E_A \wedge w((v_i, v_j)) > 0\}$$

$$\cup \{(v_{inega}, v_{jnega}) \mid v_{inega}, v_{jnega} \in V_{nega} \wedge (v_i, v_j) \in E_A \wedge w((v_i, v_j)) > 0\}$$

$$E_{nega,A} = \{(v_{iposi}, v_{jnega}) \mid v_{iposi} \in V_{posi} \wedge v_{jnega} \in V_{nega} \wedge (v_i, v_j) \in E_A \wedge w((v_i, v_j)) < 0\}$$

$$\cup \{(v_{inega}, v_{jposi}) \mid v_{inega} \in V_{nega} \wedge v_{jposi} \in V_{posi} \wedge (v_i, v_j) \in E_A \wedge w((v_i, v_j)) < 0\}$$

$E_{new,B}$ も同様に定義される.

Step2 s に対応する V_{posi} のノードを出発地 s_{posi} に, d に対応する V_{nega} のノードを目的地 d_{nega} とし, 制約条件 1, 2 を満たす最短経路を探索する.

Step2.1 G_{new} の部分グラフ $G_{new,A} = (V_{new}, E_{new,A})$ において出発地 s_{posi} からすべてのノード $v \in V_{new}$ への最短経路 $\text{path}(s_{posi}, v)$ を Dijkstra のアルゴリズムを用いて探索する.

Step2.2 G_{new} の部分グラフ $G_{new,B} = (V_{new}, E_{new,B})$ 上ですべてのノード v から目的地 d_{nega} への最短経路 $\text{path}(v, d_{nega})$ を Dijkstra のアルゴリズムにより探索する.

Step2.3 各ノード v ごとに Step2.1 で求めた経路 $\text{path}(s_{posi}, v)$ と Step2.2 で求めた経路 $\text{path}(v, d_{nega})$ を統合し, ノード v を経由する 2 ノード間 s_{posi}, d_{nega} の経路 $\text{path}(s_{posi}, v, d_{nega})$ を導出する. 経路 $\text{path}(s_{posi}, v, d_{nega})$ の距離は Step2.1 の経

路 $path(s_{posi}, v)$ の距離と Step2.2 の経路 $path(v, d_{nega})$ の加算で求められる．

Step2.4 Step2.3 で算出した中で距離が最も小さい経路を最短経路とする．

Step3 Yen のアルゴリズムに基づき G_{new} を変更し，Step2 により k 番目の最短経路を探索する ($2 \leq k \leq K$)．

Step4 各最短経路から細胞分化クロストーク遺伝子を取り出す．経路中で辺の種類の切替え ($e_A \in E_{new,A}$ から $e_B \in E_{new,B}$ への遷移) が行われているノードが細胞分化クロストーク遺伝子に対応するノードである．

提案手法の各ステップの動作の概念図を図4に示す．今，5つの頂点からなるグラフ G に対して，出発地 v_1 から目的地 v_5 に至る細胞分化クロストーク候補経路を K 個を求めることを考える．グラフ G より Step1 を経て新しく作成されたグラフが右上のグラフ G_{new} である．各ノードごとに v_{posi} と v_{nega} の2種類のノードが作成され，辺も新たに引かれる．また新しい出発地として v_{1posi} ，目的地として v_{5nega} が設定される．次にグラフ G_{new} から Step2 を用いて1番目の最短経路を求める．左側中段にある経路 $path(v_{1posi}, v_{2posi}, v_{5nega})$ が1番目の最短経路として求められた経路である．2番目以降の最短経路は Step3 を用いて求める．2番目に得られた最短経路が右下にある $path(v_{1posi}, v_{4nega}, v_{5nega})$ である．引き続き Step3 により K 番目までの最短経路を求める．得られた K 個の最短経路から Step4 により細胞分化クロストーク候補遺伝子に相当するノードを取り出す．図の左下のように1番目の最短経路からは v_{2posi} を2番目の最短経路からは v_{4nega} が得られる．以降の3番目から K 番目までの最短経路からも同様にしてノードを取り出し，結果， K 個の細胞分化クロストーク遺伝子が得られる．

4. 検証実験

提案した細胞分化クロストークモデルの妥当性を検証するために実験を行う．本実験では細胞分化クロストーク遺伝子であると推定された遺伝子を細胞分化クロストーク候補遺伝子と呼ぶ．検証実験では生物学的実験によって得られた細胞分化クロストーク候補遺伝子を用いて提案手法によって得られる細胞分化クロストーク遺伝子の評価を行い，提案モデルの有用性を示す．

4.1 実験条件

遺伝子発現量データからモデルに従い細胞分化クロストーク遺伝子を推定する実験を以下の条件で行った．

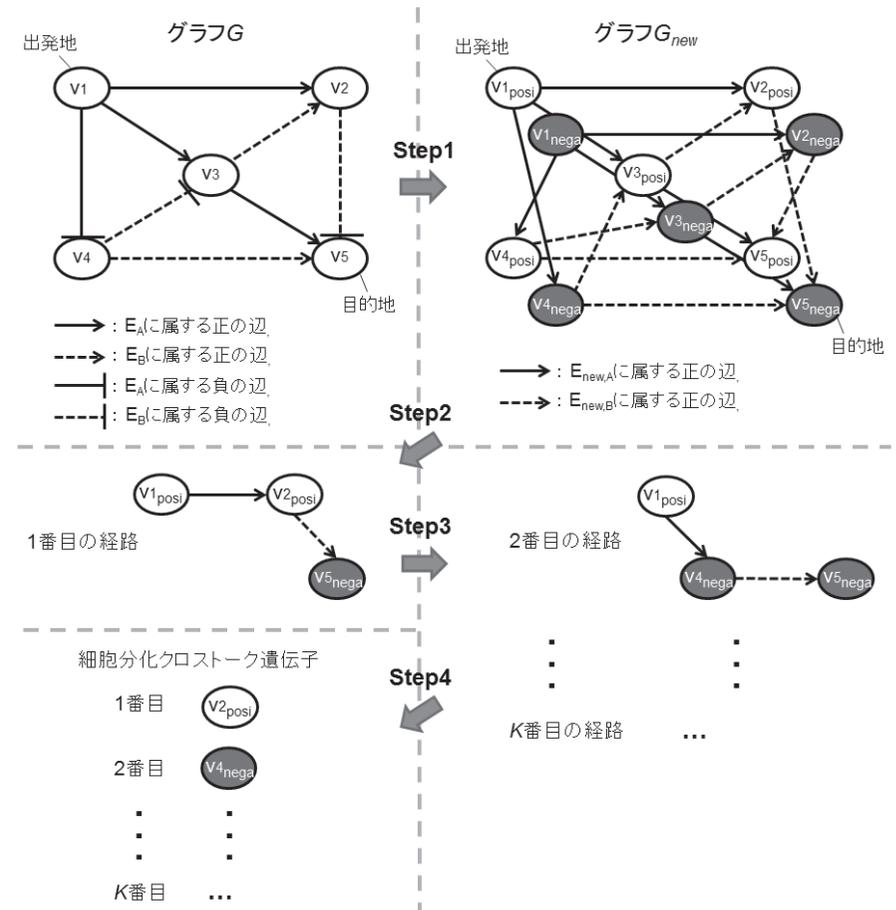


図4 提案手法の概要図
 Fig. 4 Scheme of proposed method.

利用した生物実験情報

● 遺伝子発現量データ

本研究の一連の実験においては遺伝子発現量データとして、マウスの脂肪細胞分化および骨芽細胞分化の時系列遺伝子発現量データを使用する。測定時点数は 0 から 192 時点まで 6 時間間隔の 33 時点である。実験データにおける遺伝子数は 557 個である^{*1}。

● 発現変動遺伝子

2.2 節で述べた RNAi 法により骨芽細胞で特徴的な遺伝子 Sp7⁹⁾, Runx2¹¹⁾ をノックダウンした結果、いずれかの実験において発現量分布の標準偏差 $\times 2$ 以上発現量を変化した遺伝子が 65 個検出された。この 65 個の遺伝子を発現変動遺伝子と定義する。

入力

● グラフ $G = \{V, E\}$

ノード集合 V は遺伝子集合に対応している。よって前述した遺伝子発現量データから本実験でのノードの数は 557 個となる。

遺伝子間の制御関係推定には時間遅延を考慮したピアソン相関係数²²⁾ を利用する。時間遅延を τ としたとき、遺伝子 i から遺伝子 j への制御関係はピアソン相関係数 $R = (r_{i,j})$ を変形した次の式により定義される。遺伝子 i の時点 t での発現量を $x_i(t)$ 、遺伝子 j の時点 t での発現量を $x_j(t)$ 、データの測定時点数を n とおく。

$$S_{i,j}(\tau) = \sum_{t=1}^{n-\tau} (x_i(t) - \bar{x}_i)(x_j(t+\tau) - \bar{x}_j) \quad (1)$$

$$r_{i,j}(\tau) = \frac{S_{i,j}(\tau)}{\sqrt{S_{i,i}(\tau)S_{j,j}(\tau)}} \quad (2)$$

遺伝子 i, j がノード v_i, v_j に対応するとき、辺 $e = (v_i, v_j)$ の重みは下記で定義する²³⁾。

$$w(e) = w((v_i, v_j)) = \begin{cases} +\varepsilon & (r_{i,j} = 1) \\ 1 - r_{i,j}(\tau) & (r_{i,j}(\tau) \geq 0) \\ -1 - r_{i,j}(\tau) & (r_{i,j}(\tau) < 0) \\ -\varepsilon & (r_{i,j} = -1) \end{cases}$$

ここで ε は十分に小さい数である。

*1 なお元のデータには遺伝子のプローブが 45,101 個含まれている。スクリーニングとして転写因子かつ脂肪細胞分化時に Msx2 遺伝子以上の、骨芽細胞分化時に Runx2 遺伝子以上の発現量変化を示すものを抽出する作業を行った結果、557 個の遺伝子が得られた。なお Msx2, Runx2 はそれぞれ脂肪細胞、骨芽細胞において中心的な役割をするネットワークに含まれる遺伝子である。

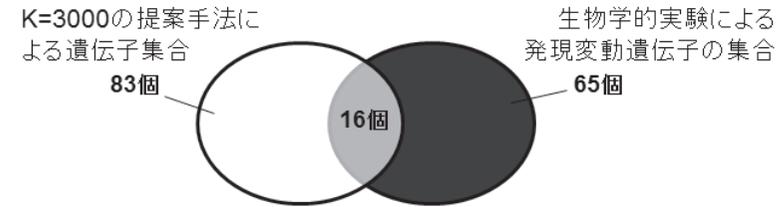


図5 提案手法により導出した遺伝子と発現変動遺伝子のベン図
Fig.5 Venn diagram of genes derived by our method and genes whose expression values changed.

本実験では $\tau = 1$ として推定を行う。なお G は完全グラフを想定しており、各遺伝子間で双方向の制御関係を推定する。つまり遺伝子集合から $(1 \leq i, j \leq 557, i \neq j)$ の範囲で遺伝子 i, j を選択して、上記の式により辺 $e = (v_i, v_j)$ を導く。遺伝子を選択する順番に意味があるため辺の本数は $557P_2$ となる。骨芽細胞から得られる辺集合を E_A 、脂肪細胞から得られる辺集合を E_B とする。 E_A, E_B それぞれの辺の本数は $557P_2 = 309,692$ 本となる。

● 出発地 s , 目的地 d

出発地 s を骨芽細胞で特徴的な遺伝子 Sp7, 目的地 d を脂肪細胞で特徴的な遺伝子 Pparg とする。生物学的実験により Sp7 は骨芽細胞への分化, Pparg は脂肪細胞への分化が行われる際に重要な遺伝子として知られている^{9),10)}。

パラメータ

● 最短経路数 K

本実験では $K = 3,000$ とする。1 つの経路につき 1 つの細胞分化クロストーク候補遺伝子が得られる。しかし細胞分化クロストーク遺伝子が同一であってもそれ以外の経路構成に差異がある場合、別々の経路として導出される。そのため経路の個数と得られる細胞分化クロストーク候補遺伝子の種類数が異なる。そこで最短経路数 K を遺伝子数 575 個より多く設定している。

4.2 実験結果

実験条件に従い 3,000 位までの最短経路を求めることにより、83 個の細胞分化クロストーク候補遺伝子が得られた。また図 5 に示すように 83 遺伝子中 16 遺伝子が発現変動遺伝子と一致した。得られる遺伝子数は求める最短経路数 K の値に依存する。そこで細胞分化クロストーク候補遺伝子数と最短経路順位の関係を図 6 に示す。また細胞分化クロストーク候補遺伝子に含まれる発現変動遺伝子数の関係も図 6 中に示している。この図より、最短

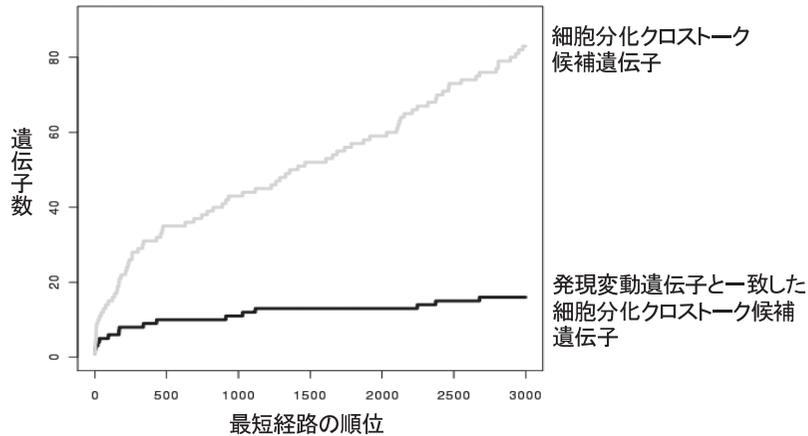


図 6 最短経路の順位に対する細胞分化クロストーク候補遺伝子の累計の推移
Fig. 6 The number of ranks of paths and the number of candidates.

経路順位の初期に多くの候補遺伝子が導出されたことおよび、その後も最短経路数の増加にともない候補遺伝子が得られていることが分かる。

次に得られる細胞分化クロストーク候補遺伝子数と発現変動遺伝子数の関係を図 7 に示す。図には今回の実験対象である遺伝子 557 個から等確率でランダムに抽出した際に含まれる発現変動遺伝子数の期待値と、期待値の標準偏差の幅も表示している。図 7 より等確率のランダム抽出よりも提案モデルを用いた手法が上回っていることが分かる。

4.3 考 察

今回の実験では最短経路数 $K = 3,000$ とすることによって 83 個の細胞分化クロストーク候補遺伝子を制御関係の強さに基づく順位とともに得ることができた。得られる細胞分化クロストーク候補遺伝子数は最短経路数 K を大きくすることにより増やすことができるため、今回導出していない遺伝子にも順位を付与することが可能である。一方、発現変動遺伝子を得るための生物学実験では順位付けを行うことはできない。これは提案モデルに基づく本手法が優位である部分である。なぜなら本手法や 2.2 節で述べた生物学実験により導出される細胞分化クロストーク候補遺伝子はあくまで推定結果にとどまるため、さらに細胞分化クロストーク遺伝子であると確定させるための実験が必要とされる。そのため確定させるための実験として、細胞分化クロストーク候補遺伝子の 1 つ 1 つに対してコストの大きい個別の生物学実験が行われるからである。この実験を行う際に遺伝子に優先順位が付与

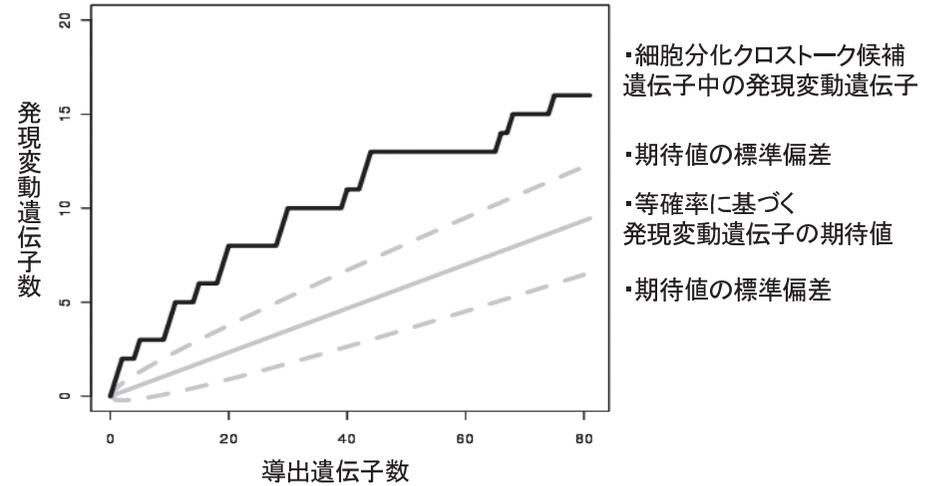


図 7 導出された発現変動遺伝子数の推移
Fig. 7 The number of candidates and the number of candidates matched with wet candidates.

されていることは実験者にとって大きなメリットとなる。

5. おわりに

本研究では細胞分化クロストークの数理モデル化の提案を行った。これまで細胞分化クロストークに対する情報科学的アプローチは行われていなかったが、提案したモデルにより細胞分化クロストークを情報科学的に扱うことを可能とした。このモデルに基づき細胞分化クロストーク遺伝子を推定したところ有効性を示す結果が得られた。さらに既存の生物学実験による細胞分化クロストーク遺伝子の推定とは異なる点として遺伝子を順位付けできることがあげられる。これにより細胞分化クロストーク遺伝子を発見するために必要な生物学実験の効率化に貢献できると考えている。

参 考 文 献

- 1) Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P.: *Molecular Biology of the Cell*, 5th edition, Garland Science (2008).

- 2) De Smet, R. and Marchal, K.: Advantages and Limitations of Current Network Inference Methods, *Nature reviews, Microbiology*, Vol.8, No.10, pp.717–729 (2010).
- 3) Toh, H. and Horimoto, K.: Inference of a Genetic Network by a Combined Approach of Cluster Analysis and Graphical Gaussian Modeling, *Bioinformatics*, Vol.18, No.2, pp.287–297 (2002).
- 4) Friedman, N., Linial, M., Nachman, I. and Pe'er, D.: Using Bayesian Networks to Analyze Expression Data, *Journal of Computational Biology*, Vol.7, No.3-4, pp.601–620 (2000).
- 5) Chen, T., He, H.L. and Church, G.M.: Modeling Gene Expression with Differential Equations, *Pacific Symposium on Biocomputing (PSB'99)*, Vol.4, pp.29–40 (1999).
- 6) Hecker, M., Lambeck, S., Toepfer, S., van Someren, E. and Guthke, R.: Gene Regulatory Network Inference: Data Integration in Dynamic Models—A Review, *Bio Systems*, Vol.96, No.1, pp.86–103 (2009).
- 7) Evans, M. and Kaufman, M.: Establishment in Culture of Pluripotential Cells from Mouse Embryos, *Nature*, Vol.292, pp.154–156 (1981).
- 8) Takahashi, K. and Yamanaka, S.: Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors, *Cell*, Vol.126, No.4, pp.663–676 (2006).
- 9) Nakashima, K., Zhou, X., Kunkel, G., Deng, J.M., Behringer, R.R. and de Crombrughe, B.: The Novel Zinc Finger-Containing Transcription Factor Osterix is Required for Osteoblast Differentiation and Bone Formation, *Cell*, Vol.108, pp.17–29 (2002).
- 10) Tamori, Y., Masugi, J., Nishino, N. and Kasuga, M.: Role of Peroxisome Proliferator-Activated Receptor-gamma in Maintenance of the Characteristics of Mature 3T3-L1 Adipocytes, *Diabetes*, Vol.51, No.7, pp.2045–2055 (2002).
- 11) Komori, T., Yagi, H., Nomura, S., Yamaguchi, A., Sasaki, K., Deguchi, K., Shimizu, Y., Bronson, R.T., Gao, Y.H., Inada, M., Sato, M., Okamoto, R., Kitamura, Y., Yoshiki, S. and Kishimoto, T.: Targeted Disruption of Cbfa1 Results in a Complete Lack of Bone Formation Owing to Maturational Arrest of Osteoblasts, *Cell*, Vol.89, pp.755–764 (1997).
- 12) 坊農秀雅, 八木 研, 仲地 豊, 二階堂愛, 岡崎康司: 脂肪・骨芽細胞分化ネットワークのクロストークと冗長性の解明, 蛋白質・核酸・酵素, Vol.49, No.17, pp.2965–2969 (2004).
- 13) Luscombe, N.M., Babu, M.M., Yu, H., Snyder, M., Teichmann, S.A. and Gerstein, M.: Genomic Analysis of Regulatory Network Dynamics Reveals Large Topological Changes, *Nature*, Vol.431, No.7006, pp.308–312 (2004).
- 14) Steele, E. and Tucker, A.: Consensus and Meta-analysis Regulatory Networks for Combining Multiple Microarray Gene Expression Datasets, *Journal of Biomedical Informatics*, Vol.41, No.6, pp.914–926 (2008).
- 15) Tokuzawa, Y., Yagi, K., Yamashita, Y., Nakachi, Y., Nikaido, I., Bono, H., Ninomiya, Y., Kanasaki-Yatsuka, Y., Akita, M., Motegi, H., Wakana, S., Noda, T., Sablitzky, F., Arai, S., Kurokawa, R., Fukuda, T., Katagiri, T., Schönbach, C., Suda, T., Mizuno, Y. and Okazaki, Y.: Id4, a New Candidate Gene for Senile Osteoporosis, Acts as a Molecular Switch Promoting Osteoblast Differentiation, *PLoS genetics*, Vol.6, No.7, pp.1–15 (2010).
- 16) Dijkstra, E.W.: A Note on Two Problems in Connexion with Graphs, *Numerische Mathematik*, Vol.1, pp.269–271 (1959).
- 17) Sniedovich, M.: Dijkstra's Algorithm Revisited: The Dynamic Programming Connexion, *Control and Cybernetics*, Vol.35, No.3, pp.599–620 (2006).
- 18) Yen, J.Y.: Finding the K Shortest Loopless Paths in a Network, *Management Science*, Vol.17, No.11, pp.712–716 (1971).
- 19) Hershberger, J., Maxel, M. and Suri, S.: Finding the K Shortest Simple Paths: A New Algorithm and its Implementation, *ACM Trans. Algorithms*, Vol.3, No.4, p.45 (2007).
- 20) Burkhardt, R., Kettner, G., Böhm, W., Schmidmeier, M., Schlag, R., Frisch, B., Mallmann, B., Eisenmenger, W. and Gilg, T.: Changes in Trabecular Bone, Hematopoiesis and Bone Marrow Vessels in Aplastic Anemia, Primary Osteoporosis, and Old Age: A Comparative Histomorphometric Study, *Bone*, Vol.8, pp.157–164 (1987).
- 21) Hannon, G.J.: RNA-interference, *Nature*, Vol.418, pp.244–251 (2002).
- 22) Schmitt, W.A.J., Raab, R.M. and Stephanopoulos, G.: Elucidation of Gene Interaction Networks through Time-Lagged Correlation Analysis of Transcriptional Data, *Genome Research*, Vol.14, No.8, pp.1654–63 (2004).
- 23) Ruan, J., Dean, A.K. and Zhang, W.: A General Co-expression Network-based Approach to Gene Expression Analysis: Comparison and Applications, *BMC Systems Biology*, Vol.4, No.1, p.8 (2010).

(平成 23 年 3 月 25 日受付)

(平成 23 年 6 月 1 日再受付)

(平成 23 年 6 月 21 日採録)



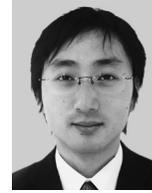
吉澤 陽志 (正会員)

2009年大阪大学基礎工学部情報科学科卒業。2011年大阪大学大学院情報科学研究科バイオ情報工学専攻博士前期課程修了。同年富士通株式会社入社。



瀬尾 茂人 (正会員)

大阪大学大学院情報科学研究科助教。2006年大阪大学大学院情報科学研究科修了。同年同研究科助手。2007年より現職。バイオインフォマティクスの研究に従事。日本バイオインフォマティクス学会、日本分子生物学会各会員。博士(情報科学)。



竹中 要一 (正会員)

大阪大学大学院情報科学研究科准教授。1995年大阪大学基礎工学部情報工学科退学(飛び級)。同年大阪大学大学院基礎工学研究科物理系専攻博士前期課程入学。1997年同修了。同年大阪大学大学院基礎工学研究科博士後期課程入学。2000年同修了。同年博士(工学)の学位取得。同年大阪大学大学院基礎工学研究科助手。2002年大阪大学大学院情報科学研究科助教。2007年より現職。生物情報学の研究に従事。電子情報通信学会、日本バイオインフォマティクス学会、言語処理学会、ISCB、IEEE 各会員



松田 秀雄 (正会員)

大阪大学大学院情報科学研究科教授。1987年神戸大学大学院自然科学研究科修了(学術博士)。同年同大学工学部助手。1994年大阪大学基礎工学部助教授。2002年より現職。バイオインフォマティクスの研究に従事。日本バイオインフォマティクス学会、ISCB、IEEE CS、ACM 各会員。