

# 遺伝子発現プロファイルを用いた遺伝子制御ネットワーク推定のためのバイクラスタリングの利用

灌 浩平<sup>†</sup> 竹中 要一<sup>†</sup> 松田 秀雄<sup>†</sup>

遺伝子発現プロファイルの蓄積にともない、遺伝子制御ネットワークの推定に、より多くの実験条件を含む発現プロファイルを用いることが可能になった。しかし、モジュールネットワークのような推定手法を、そのような発現プロファイルに対して適用すると、推定精度の低下を招く恐れがある。モジュールネットワークは、発現プロファイルの実験条件の大半で類似した発現を示す遺伝子が多数存在することを前提とするが、多くの実験条件を含む発現プロファイルほどそのような遺伝子は少ない。そこで本研究では、発現プロファイルのバイクラスタリングの結果を利用する。発現プロファイルの実験条件がバイクラスタに含まれる部分のみを用いて推定を行うことで、推定精度の低下の軽減を図る。本手法を出芽酵母の複数の発現プロファイルに対して適用し、有効性を検証した。

## Inference of Gene Regulatory Networks from Gene-expression Profiles with Utilization of Biclustering Results

KOHEI TAKI,<sup>†</sup> YOICHI TAKENAKA<sup>†</sup> and HIDEO MATSUDA<sup>†</sup>

The accumulation of gene-expression profiles can allow an inference of a gene regulatory network by using a profile measured under a number of experimental conditions. However, in case of applying to such profile, the conventional methods such as module network model may not perform an inference accurately enough, because of following two facts. 1) Module network can accurately perform an inference only for regulated genes that show similar gene-expression patterns under almost all experimental conditions. 2) In a gene-expression profile that includes more conditions, fewer genes show similar gene-expression patterns. To alleviate the accuracy loss, we utilized a biclustering result for an inference. We performed an inference for regulated genes that were included in a detected bicluster by using gene-expression patterns only under experimental conditions included in the bicluster. We demonstrate the effectiveness of our method by applying to inferences of gene regulatory networks by using various gene-expression profiles of budding yeast.

### 1. はじめに

ヒトをはじめとする様々な生物種で展開されたゲノムプロジェクトの成果により、300種類を超えるゲノムの全 DNA 塩基配列が完全に解読された。このような遺伝子機能解析の基礎的データの整備にともない、個々の遺伝子の機能解析から、複数の遺伝子が協調して働くことで果たされる機能の解析へと焦点が移りつつある。このような複雑な機能解析のためには、遺伝子相互の機能的関連により形成される遺伝子ネットワークの構造を解明することが求められる。そのためには、従来のゲノム解析よりもさらに複雑かつ大量の

データを取り扱わなければならない、新たな情報科学的解析技術の開発が必要不可欠である。

遺伝子ネットワークの構造を解明するための研究の1つとして、遺伝子発現プロファイルを用いて遺伝子制御ネットワークを推定する試みがなされている。遺伝子発現プロファイルとは、細胞を様々な条件下に置いた場合に各遺伝子が働いた量（遺伝子発現量）を示す、遺伝子×実験条件の行列データである。遺伝子制御ネットワークとは、遺伝子の間の発現量の制御関係を表したグラフである。

遺伝子制御ネットワークの推定は、ベイジアンネットワーク<sup>1)</sup>、微分方程式モデル<sup>2)</sup>、ガウシアングラフィカルモデリング<sup>3)</sup>などを用いて研究が進められてきた。また、推定における組合せ爆発の問題を軽減させるために、推定の前段階として同じ制御関係に従う遺伝子をまとめる、遺伝子のクラスタリング手法<sup>4)</sup>の研究

<sup>†</sup> 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

が進められてきた．そして，クラスタリングとネットワーク推定を同時に行う手法として，モジュールネットワークモデル<sup>5)</sup>に基づいた推定の試みがなされている．本研究ではモジュールネットワークモデルに基づいて，遺伝子制御ネットワークを推定する．

遺伝子制御ネットワークの推定では，より多くの実験条件を含む発現プロファイルを用いた方が，精度の高い結果が得られることが期待される．近年の発現プロファイルの蓄積と遺伝子発現データベース<sup>6)~8)</sup>の整備により，より多くの実験条件を含む発現プロファイルの取得が可能になってきた．このため，その利用によってより精度の高い推定が可能になると期待されている．本研究ではこのような多くの実験条件を含む発現プロファイルを用いて，遺伝子制御ネットワークの推定を行う．

しかし，モジュールネットワークモデルによる推定では，このような発現プロファイルに対しては十分な推定精度を得られない可能性がある．モジュールネットワークモデルは，大半の実験条件で類似した発現量を示す遺伝子の集合（モジュール）が，多数存在することを前提として制御関係の推定を行う．より多くの実験条件を含む発現プロファイルでは，大半の実験条件で類似した発現量を示す遺伝子は多くない．このため，同じモジュールにまとめられた遺伝子の間でも，多数の実験条件で発現量が類似しないことが多くなり，推定精度が低下することが予測される．

そこで本研究では，発現プロファイルから推定に用いる実験条件を選別し，選別された実験条件のみを用いてモジュールネットワークモデルによる推定を行う手法を提案する．実験条件の選別には，遺伝子と実験条件のバイクラスタリング<sup>9)~12)</sup>を利用する．バイクラスタリングによって，遺伝子発現パターンが部分的に類似した遺伝子群と実験条件群の組が得られる．これを利用することで，発現パターンが十分に類似した実験条件のみを用いて制御関係を推定することが可能になり，より多くの実験条件を含む発現プロファイルへの適用が可能になると考えられる．

以下ではまず，2章で遺伝子発現プロファイルとそれを用いた遺伝子制御ネットワークの推定について説明する．3章でバイクラスタリングについて説明し，4章でバイクラスタリングの結果をモジュールネットワークの推定に利用する方法を提案する．5章では，本手法の有効性を評価するため，遺伝子発現データベースに登録された出芽酵母の発現プロファイルに対して本手法を適用した結果を示し，6章でその考察を行う．

## 2. 遺伝子の制御関係の推定

### 2.1 遺伝子発現プロファイルと制御関係の推定

遺伝子とは DNA 配列上の領域であり，生命活動に不可欠とされるタンパク質の設計図である．タンパク質はそれぞれ，遺伝子がコードされた DNA 配列上の領域が，まず mRNA に転写され，次にそれを鋳型としてアミノ酸の配列に翻訳されるという 2 つの段階を通じて作られており，これを遺伝子の発現と呼ぶ．遺伝子が mRNA に転写された量は転写レベルでの遺伝子発現量と呼ばれ，DNA マイクロアレイ<sup>13)</sup>などを用いた実験によってその量を測定することができる．以下，本論文ではこの転写レベルでの遺伝子発現量を，単に遺伝子の発現量と呼ぶ．

遺伝子発現プロファイルとは，遺伝子  $\times$  実験条件の発現量の行列データである．発現量の測定対象となる細胞を様々な条件下に置いて実験を行い，各条件下における各遺伝子の発現量を測定することで得られる．本論文では，遺伝子  $i$  を記号  $g_i \in G$  で，実験条件  $j$  を記号  $c_j \in C$  で表し，発現プロファイルを図 1 に示すように記号  $X$  で表す． $G = \{g_1, \dots, g_m\}$  および  $C = \{c_1, \dots, c_n\}$  は，それぞれ発現プロファイルに現れる遺伝子と実験条件の全体を表す集合とする．発現プロファイル  $X = (x_{(i,j)})_{m \times n}$  の  $i$  行目の行ベクトルを遺伝子  $g_i$  の発現パターンと呼び，記号  $\mathbf{x}_{g_i} = (x_{(i,1)}, \dots, x_{(i,n)})$  で表す．同様に実験条件  $c_j$  の列ベクトルを記号  $\mathbf{x}_{c_j} = (x_{(1,j)}, \dots, x_{(m,j)})^T$  で表す．

発現量には遺伝子間で依存関係があることが観測されている．遺伝子  $g_1$  の発現量の増加が遺伝子  $g_2$  の発現量の増加をもたらすとき， $g_1$  は  $g_2$  を活性化するといひ，逆に減少をもたらすとき，不活性化するという．遺伝子  $g_1$  の発現によって作られる転写因子と呼ばれるタンパク質が，遺伝子  $g_2$  がコードされた DNA 配列の近傍に存在する領域に結合して転写を促進・抑制することで， $g_2$  を活性化・不活性化する．

本論文では活性化と不活性化をまとめて遺伝子の

		実験条件 C			
		$c_1$	$c_2$	$c_3$	$c_4$
遺伝子 G	$g_1$	$x_{(1,1)}$	$x_{(1,2)}$	$x_{(1,3)}$	$x_{(1,4)}$
	$g_2$	$x_{(2,1)}$	$x_{(2,2)}$	$x_{(2,3)}$	$x_{(2,4)}$
	$g_3$	$x_{(3,1)}$	$x_{(3,2)}$	$x_{(3,3)}$	$x_{(3,4)}$

発現パターン  $\mathbf{x}_{g_3}$

図 1 遺伝子発現プロファイル  $X$   
Fig. 1 Gene expression profile  $X$ .

制御関係と呼ぶ。遺伝子の制御関係は連なって遺伝子制御ネットワークを形成している。遺伝子制御ネットワークは有向グラフによって表現され、遺伝子をノードで、その間の制御関係を有向辺で表す。有向辺の向きは制御関係の向きを表しており、遺伝子  $g_1$  が遺伝子  $g_2$  を制御することは、 $g_1$  に対応するノードから  $g_2$  に対応するノードへの有向辺によって表される。

発現プロファイルによる遺伝子制御ネットワークの推定はこれまで、ベイジアンネットワーク<sup>1)</sup>、微分方程式モデル<sup>2)</sup>、ガウシアングラフィカルモデリング<sup>3)</sup>などを用いて研究が進められてきた。本研究ではベイジアンネットワークから派生した手法を利用する。ベイジアンネットワークの推定は、可能なネットワークのモデルの中から、評価値が最大となるモデルを探索する組合せ最適化問題となる。ネットワーク推定問題は各遺伝子を制御する遺伝子を推定する問題に分割され、それぞれが制御遺伝子の組合せ最適化問題となる。制御関係の推定では、発現プロファイル  $X$  に観測された、遺伝子間の発現パターンの依存関係をベイズ統計に基づいて評価する。遺伝子  $g_i$  の発現量の確率変数  $X_{g_i}$  を定義し、 $g_1$  が  $g_2$  に制御されることを、制御遺伝子  $g_2$  の発現量  $X_{g_2}$  が与えられたときの  $X_{g_1}$  の条件付き確率  $P(X_{g_1}|X_{g_2})$  によって表す。 $g_1$  が  $g_2$  に制御されるモデルに対しては、 $P(X_{g_1}|X_{g_2})$  から発現パターン  $x_{g_1}$  の尤度関数の値が求められ、 $g_1$  の制御関係に対するこのモデルの評価値となる。

制御関係の向きの推定は、上記のような 1 対 1 の関係を対象とする場合は難しいが、遺伝子  $g_1$  と  $g_2$  が  $g_3$  を制御するような多対 1 の関係を対象とする場合には可能である。尤度関数の値は、 $g_1$  と  $g_2$  が  $g_3$  を制御するモデルでは  $P(X_{g_3}|X_{g_1}, X_{g_2}) \cdot P(X_{g_1}) \cdot P(X_{g_2})$  から、 $g_1$  と  $g_2$  が  $g_3$  に制御されるモデルでは  $P(X_{g_1}, X_{g_2}|X_{g_3}) \cdot P(X_{g_3})$  から求められ、尤度が高い方のモデルが解となる。 $x_{g_1}$  と  $x_{g_2}$  が互いに独立した値を示す場合には、2 つ目のモデルは 1 つ目のモデルに対して尤度が低くなるため、制御関係の向きが推定できる。

上記のようにネットワーク推定は制御遺伝子の組合せ最適化問題である。ヒトでは約 3 万個の遺伝子があると考えられており、制御遺伝子の組合せの数は膨大となるため、遺伝子数に対する組合せ爆発が問題となる。このため推定の前段階として同じ制御関係に従う遺伝子をまとめるために、階層型や分割型のクラスタリングの手法を用いた研究が進められてきた<sup>4)</sup>。そして、クラスタリングとネットワーク推定を同時に行う手法として、分割型クラスタリングとベイジアンネッ

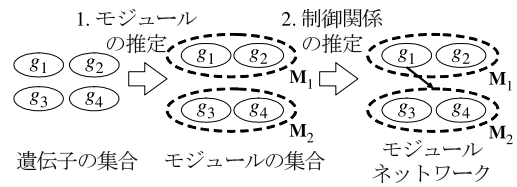


図 2 モジュールネットワークの推定  
Fig.2 Inference of a module network.

トワークを組み合わせた、モジュールネットワークモデルに基づく推定手法が提案されている<sup>5)</sup>。モジュールネットワークモデルでは、推定精度の低下を抑えながら制御遺伝子の組合せ爆発を緩和できることが報告されている。本研究ではモジュールネットワークモデルによって、遺伝子制御ネットワークの推定を行う。

## 2.2 モジュールネットワーク

これまでに解明された遺伝子間の制御関係には、1 つの遺伝子が複数の遺伝子を制御する例が多数存在する。さらに、同じ遺伝子に制御される遺伝子は類似した発現パターンを示すことが、多くの遺伝子発現プロファイルで確認された。モジュールネットワークモデルは、同じ遺伝子に制御され類似した発現パターンを示す遺伝子の集合を、モジュールとしてまとめて扱うモデルである。同じモジュールに含まれるすべての遺伝子が、同じ遺伝子に制御されるようにモデルを制限することで、推定における組合せ爆発を緩和している。

モジュールネットワークはモジュールの集合とその間を結ぶ有向辺によって表される。モジュール  $M_k \subseteq G$  は同じ遺伝子に制御される遺伝子の集合として定義される。有向辺は遺伝子からモジュールへ結ばれ、その遺伝子がモジュールに含まれるすべての遺伝子を制御することを表す。図 2 は、遺伝子  $g_3, g_4$  が同じモジュール  $M_2 = \{g_3, g_4\}$  に含まれ、遺伝子  $g_1$  に制御されることを表すモジュールネットワークの例である。

モジュールネットワークの推定は図 2 に示すように、1) 各遺伝子が含まれるモジュールの推定、2) 各モジュールに対する制御関係の推定、の 2 つの段階に分かれる。ネットワークの評価値が収束するまで 2 つの段階を交互に繰り返すことで、モジュールネットワークは推定される。評価値はベイジアンネットワークの尤度関数を用いて求められる。尤度関数はモジュールごとに分割して計算できることが仮定され、これによりモジュールごとに独立に評価ができる。本論文では各モジュール  $M_k$  の尤度関数を以下のように記述する。

$$score_k = L(M_k, \mathbf{Pa}_{M_k} : X)$$

ただし、 $\mathbf{Pa}_{M_k}$  はモジュール  $M_k$  を制御する遺伝子

の集合を表す．ネットワーク全体の尤度は，上記を用いて  $\prod_k score_k$  として求められる．

推定の入力として，制御関係の推定対象となる遺伝子からなる発現プロファイル  $X$  と，制御遺伝子の候補の集合  $G_e \subseteq G$  が与えられる． $G_e$  は，各モジュールを制御する遺伝子を， $G_e$  に含まれる遺伝子に限定する ( $Pa_{M_k} \subseteq G_e$ ) ために用いられる．これによって制御遺伝子の組合せ爆発を緩和できる．

### 2.3 遺伝子発現プロファイルの蓄積とデータベースの整備

制御関係の推定では，推定に用いた遺伝子発現プロファイルに含まれる実験条件の数が多いほど，推定結果からより具体的な制御関係の仮説を導くことができる場合が多い．現在，発現プロファイルは，DNA マイクロアレイの技術発展にともなって膨大な数が蓄積されつつあり，より多くの実験条件を含む発現プロファイルを用いた推定が可能になりつつある．

さらに，公共遺伝子発現データベースの整備によって，多数の発現プロファイルを集めることが容易になりつつある．日米欧それぞれで公共遺伝子発現データベース (CIBEX<sup>6)</sup>，GEO<sup>7)</sup>，ArrayExpress<sup>8)</sup>) が整備されており，DNA マイクロアレイを用いた研究報告に対して，公共データベースへの登録を義務付ける論文誌が多くなってきた．論文投稿時に解析を行った一連の実験データを登録・公開することで，自由に検証や解析を行うことができる環境が整いつつある．

このような環境の整備にともなって，複数の発現プロファイルを用いた解析が試みられるようになってきた．たとえば文献 14) では，異なる文献で公表されている複数の発現プロファイルを，実験条件について連結することで，多くの実験条件を含む 1 つの発現プロファイルとして解析している．この方法を遺伝子発現データベースに適用することによって，より多くの実験条件を含む発現プロファイルの解析が可能になる．この連結した発現プロファイルを用いることで，より具体的な遺伝子制御ネットワークの推定を試みるのが，本研究の目的である．

連結した発現プロファイルは，それぞれが行列である  $L$  個の発現プロファイル  $X^{(1)}, \dots, X^{(L)}$  を， $(X^{(1)} \dots X^{(L)})$  のように列方向に並べた行列とする． $X^{(l)}$  の遺伝子の集合  $G^{(l)}$  は完全に同じではないが，連結した発現プロファイルは  $G^{(l)}$  の遺伝子をすべて含むようにする． $G^{(l)}$  に遺伝子  $g_i$  が含まれず， $X^{(l)}$  に値が存在しない  $g_i$  の要素は欠損値として扱う．また 2.1 節で述べたように，本研究で扱う推定手法は各遺伝子の発現量の分布からモデルを評価するため，連

結する順番は推定結果に影響しない．

連結した発現プロファイルを以降で記号により表すため， $X, G, C$  を次のように再定義する． $l$  番目に連結する発現プロファイルを  $X^{(l)} = (x_{(i,j)}^{(l)})_{m^{(l)} \times n^{(l)}}$  で表し，その実験条件の集合を  $C^{(l)} = \{c_1^{(l)}, \dots, c_{n^{(l)}}^{(l)}\}$  で，遺伝子の集合を  $G^{(l)} = \{g_1^{(l)}, \dots, g_{m^{(l)}}^{(l)}\}$  で表す．ただし，任意の  $l_1, l_2$  について  $C^{(l_1)} \cap C^{(l_2)} = \phi$  とする．連結した発現プロファイルの遺伝子および実験条件の集合として， $G \equiv \bigcup_l G^{(l)}$ ， $C \equiv \bigcup_l C^{(l)}$  のように再定義する．集合の要素を  $G = \{g_1, \dots, g_m\}$ ， $C = \{c_1, \dots, c_n\}$  で表す．連結した発現プロファイルを表す記号を  $X \equiv (x_{(i,j)})_{m \times n}$  として再定義する．ただし，

$$x_{(i,j)} = \begin{cases} x_{(i',j')}^{(l')} & : g_i \in G^{(l')} \\ null & : g_i \notin G^{(l')} \end{cases}$$

とする． $i', j', l'$  は  $x_{(i,j)}$  の添え字  $i, j$  に対して  $g_i = g_{i'}^{(l')}$  および  $c_j = c_{j'}^{(l')}$  から決まる添え字とし， $null$  は欠損値を表す記号とする．

上で述べたように連結した発現プロファイルは多くの実験条件を含むが，モジュールネットワークモデルのような従来の推定手法を適用する場合には，次の 2 つの理由から十分な推定精度を得られない可能性がある．1 つ目の理由は，モジュールネットワークモデルに基づいた従来の推定手法では，大半の実験条件で遺伝子発現パターンが類似した遺伝子からなるモジュールが多数存在することを前提として，制御関係の推定を行うことである．モジュールネットワークモデルでは，同じモジュールに含まれる複数の遺伝子の発現パターンを，共通の条件付き確率分布でモデル化する．したがって，発現パターンが類似しない実験条件では，どのような制御関係のモデルを選択しても，発現パターンとモデルが表す分布の間に一定以上のずれが生じる．このため，多くの実験条件で発現パターンが類似しないモジュールに対して，制御する遺伝子を推定する場合には，その精度が低下することが予測される．

2 つ目の理由は，本研究で解析の対象とする多数の実験条件を含む発現プロファイルを用いた場合には，同じモジュールに含まれる遺伝子の発現パターンが，あまり類似しなくなることが予測されることである．遺伝子発現データベースに含まれる複数の発現プロファイルを連結して用いるため，多数であると同時に多様な実験条件を含んでいる．発現プロファイルに含まれる実験条件が多様であるほど，多様な変化を示す発現パターンが観測されるため，任意の 2 つの遺伝子の発現パターンが高い類似性を示すことは少なくなる．

このため、多くの実験条件で発現パターンが類似しない遺伝子が、同じモジュールに分類されることが多くなると考えられる。以上の2つの理由から、モジュールネットワークモデルを適用した場合には、十分な推定精度を得られない可能性がある。

2つ目の理由で示した傾向が見られるか確認するため、遺伝子発現データベース GEO に登録された発現プロファイルに分割型クラスタリングを適用して、細胞周期で同じ時期に制御されることが既知の遺伝子104個をモジュールの集合へ分割した。細胞周期に関する実験条件の発現プロファイルでは相関係数<sup>15)</sup>の中央値が0.71と類似した発現パターンを示したが、連結した発現プロファイルでは0.35と類似した発現パターンは示さなかった。このように連結した発現プロファイルでは類似した発現パターンを示し難いことが確認できるため、これらの遺伝子に対しては1つ目の理由から十分な推定精度が得られなくなると考えられる。

この問題の解決のため本研究では、遺伝子と実験条件のバイクラスタリングの結果を、モジュールネットワークの推定に利用することを提案する。以下では、バイクラスタリングの概念について説明した後、バイクラスタを制御関係の推定に利用する方法を提案する。

### 3. 遺伝子と実験条件のバイクラスタリング

これまでさかんに研究されてきた遺伝子のクラスタリングは、遺伝子発現パターンの類似性から遺伝子を複数のグループに分類するための方法である。しかし、遺伝子の集合が類似した発現パターンを示すのは一部の実験条件に限られ、その他の実験条件ではほとんど独立に発現するか、ほとんど発現しないことが観測されている。このような局所的にしか類似しない発現パターンを発見するために開発されたのがバイクラスタリング<sup>9),10)</sup>であり、遺伝子発現プロファイルを遺伝子と実験条件の2つの次元で同時にクラスタリングする。

クラスタが遺伝子全体に対する部分集合  $G_k \subseteq G$  であるのに対して、バイクラスタは遺伝子の部分集合と実験条件の部分集合の組  $(G_k \subseteq G, C_k \subseteq C)$  として表される。4章の図3の例のようにバイクラスタ  $B_k = (G_k, C_k)$  によって、行列データである発現プロファイルの対応する列と行からなる部分行列  $X_{(G_k, C_k)}$  を表すことができる。以降、遺伝子の集合  $G_k = \{g_{i_1}, g_{i_2}, \dots, g_{i_{m_k}}\}$  と実験条件の集合  $C_k = \{c_{j_1}, c_{j_2}, \dots, c_{j_{n_k}}\}$  に対する、発現プロファイル  $X$  の部分行列を  $X_{(G_k, C_k)}$  で表し、その行ベクト

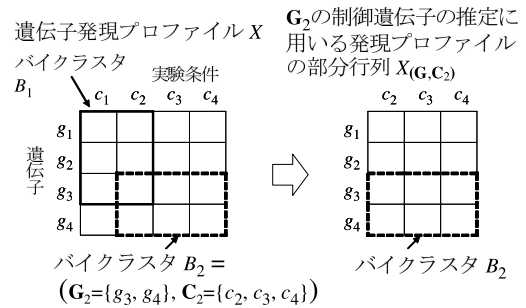


図3 バイクラスタに含まれる実験条件の集合に対して、遺伝子発現プロファイルの対応する列の部分行列を取り出す  
Fig. 3 Extracting column vectors corresponding to experimental conditions included in the bicluster  $B_2$ .

ルを  $x_{(g_i, C_k)}$  で、列ベクトルを  $x_{(G_k, c_j)}$  で表す。

$$X_{(G_k, C_k)} = X[i_1, i_2, \dots, i_{m_k}; j_1, j_2, \dots, j_{n_k}]$$

$$x_{(g_i, C_k)} = (x_{(i, j_1)}, x_{(i, j_2)}, \dots, x_{(i, j_{n_k})})$$

$$x_{(G_k, c_j)} = (x_{(i_1, j)}, x_{(i_2, j)}, \dots, x_{(i_{m_k}, j)})^T$$

ただし、 $i_1 < i_2 < \dots < i_{m_k}$ ,  $j_1 < j_2 < \dots < j_{n_k}$  とする。

クラスタリングではクラスタ内の遺伝子間の類似性を評価するのに、すべての実験条件からなる発現パターン  $x_{g_i}$  を用いるのに対して、バイクラスタリングではバイクラスタに含まれる実験条件  $C_k$  のみからなる発現パターン  $x_{(g_i, C_k)}$  だけを用いる。バイクラスタリングの目的は、特定の実験条件でのみ類似した発現パターンを示す遺伝子の集合を見つけることである。バイクラスタリングの入力は発現プロファイル  $X$ , 出力はバイクラスタの集合  $\{B_1, \dots, B_K\}$  となり、 $B_k = (G_k, C_k)$  はそれが表す部分行列  $X_{(G_k, C_k)}$  内で各遺伝子が類似した発現パターンを示すバイクラスタとする。

バイクラスタリングを用いた発現プロファイルの解析の研究としては、Plaid モデル<sup>11)</sup> や SAMBA<sup>12)</sup> を用いた事例があげられる。Plaid モデルでは、発現プロファイルの発現量  $x_{(i, j)}$  を、バイクラスタの集合  $\{B_1, \dots, B_K\}$  によって以下のモデル  $\hat{x}_{(i, j)}$  で表す。

$$\hat{x}_{(i, j)} = \sum_{k | g_i \in G_k \cap c_j \in C_k} \hat{x}_{(i, j, k)}$$

$$\hat{x}_{(i, j, k)} = \bar{x}_{(g_i, C_k)} + \bar{x}_{(G_k, c_j)} - \bar{x}_{(G_k, C_k)}$$

ただし、 $\bar{x}_{(g_i, C_k)}$ ,  $\bar{x}_{(G_k, c_j)}$ ,  $\bar{x}_{(G_k, C_k)}$  はそれぞれ、 $x_{(g_i, C_k)}$ ,  $x_{(G_k, c_j)}$ ,  $X_{(G_k, C_k)}$  の要素の平均を表す。モデルが表す発現量の値  $\hat{x}_{(i, j)}$  と実際の値  $x_{(i, j)}$  の間の偏差の二乗和  $\sum_{i, j} (\hat{x}_{(i, j)} - x_{(i, j)})^2$  を発現プロファイル全体について求め、この偏差の二乗和によってモデルを評価する。偏差の二乗和ができるだけ小さ

くなるモデルを表すバイクラスタの集合を探索する．この手法では、複数のバイクラスタの重複を考慮した柔軟なモデルの構築ができる．

SAMBA はグラフアルゴリズムを利用してバイクラスタを探索する手法であり、二部グラフによって発現プロファイルを表す．遺伝子と実験条件それぞれを二部グラフのノードと見なし、大きな発現量が観測された遺伝子と実験条件のノードの間に辺を引いた二部グラフを作る．二部グラフからできるだけ大きく密な部分グラフを見つけることで、同じ実験条件で同時に大きな発現量を示した遺伝子の集合が見つかる．その遺伝子と実験条件の集合を、バイクラスタとして検出する．

#### 4. バイクラスタリングを利用した遺伝子制御ネットワークの推定

2.3 節で述べたように、多くの実験条件を連結した遺伝子発現プロファイルにモジュールネットワークモデルを適用すると、同じモジュールに含まれる遺伝子の間で遺伝子発現パターンが類似しない実験条件が多くなる問題点がある．本研究ではこの問題点を解決するために、モジュールネットワークモデルに基づく制御関係の推定に、バイクラスタリングの結果を利用することを提案する．提案手法では、バイクラスタリングの結果を用いて、モジュールに含まれる遺伝子の間で発現パターンが類似した実験条件を選別する．そして、選別した実験条件のみを用いて、各モジュールを制御する遺伝子の集合を推定する．これによって、それぞれのモジュール内で類似する発現パターンだけを用いて、制御関係を推定することを実現する．

図 2 で示したように、モジュールネットワークの推定は、モジュール推定と制御関係の推定の、2 つの段階からなる．提案手法では、モジュール推定の段階をバイクラスタリングで置き換えることで、モジュールの推定と実験条件の選別を同時に行う．制御関係の推定の段階では、バイクラスタリングによって選別された実験条件のみを用いて推定を行う．以下では、提案手法におけるこの 2 つの段階の処理の内容と評価方法に関して順に説明する．

提案手法の入力は従来手法と同様に、発現プロファイル  $X$  と、制御遺伝子候補の集合  $G_e$  となる．まずモジュール推定の段階の代わりに、発現プロファイル  $X$  に対してバイクラスタリングを行う．検出されたバイクラスタの集合  $\{B_1, \dots, B_K\}$  に含まれる各バイクラスタ  $B_k = (G_k, C_k)$  によって、発現パターンが類似した実験条件の集合  $C_k$  が、遺伝子の集合  $G_k$  ごとに得られる．これらをそれぞれ、実験条件の選別とモジュー

ル推定の結果と見なして利用する． $B_k$  の遺伝子の集合  $G_k$  をモジュール  $M_k$  とし、 $B_k$  の実験条件の集合  $C_k$  を、 $G_k$  の制御関係の推定のために選別された実験条件として用いる．たとえば図 3 に示したバイクラスタ  $B_1$  に対しては、その遺伝子の集合  $G_1 = \{g_1, g_2, g_3\}$  をモジュール  $M_1$  とし、 $M_1$  の制御関係の推定のために選別された実験条件の集合を  $C_1 = \{c_1, c_2\}$  とする．同様に  $B_2$  に対しては、 $G_2 = \{g_3, g_4\}$  をモジュール  $M_2$  とし、 $C_2 = \{c_2, c_3, c_4\}$  を選別する．

次に制御関係の推定の段階として、 $G_k$  のために選別された実験条件  $C_k$  のみからなる、遺伝子  $g_i \in G_k$  の発現パターン  $x_{(g_i, C_k)}$  を用いて、 $g_i$  を制御する遺伝子を推定する．このために図 3 に示すように、行列である発現プロファイル  $X$  から、実験条件  $c_j \in C_k$  に対応する列のみからなる部分行列  $X_{(G, C_k)}$  を取り出す．ただし、 $G$  は  $X$  に含まれるすべての遺伝子を表す集合である． $X$  の代わりにこの部分行列  $X_{(G, C_k)}$  を用いて  $G_k$  の制御関係の評価を行い、これに基づいて推定を行う．

制御関係の推定は、検出された  $\{B_1, \dots, B_K\}$  のそれぞれのバイクラスタについて行われる．バイクラスタ  $B_k$  ごとに含まれる実験条件の集合  $C_k$  が異なるため、 $B_k$  ごとに  $X$  の異なる列からなる部分行列  $X_{(G, C_k)}$  を用いて、 $G_k$  の制御関係の評価を行う．図 3 の例では、バイクラスタ  $B_1$  に含まれる遺伝子  $g_1, g_2, g_3$  を制御する遺伝子の推定には、 $C_1$  に含まれる実験条件  $c_1, c_2$  の列のみからなる部分行列  $X_{(G, \{c_1, c_2\})}$  を用いる．これに対して、バイクラスタ  $B_2$  に含まれる遺伝子  $g_3, g_4$  を制御する遺伝子の推定には、 $C_2$  に含まれる実験条件  $c_2, c_3, c_4$  の列のみからなる部分行列  $X_{(G, \{c_2, c_3, c_4\})}$  を用いる．

提案手法では、モジュール  $M_k$  を制御遺伝子の集合  $\text{Pa}_{M_k}$  が制御するモデルの評価に、2.2 節であげたモジュールネットワークで用いる尤度関数  $L(M_k, \text{Pa}_{M_k} : X)$  を利用する．各バイクラスタ  $B_k = (G_k, C_k)$  に対して、 $G_k$  を制御遺伝子の集合  $\text{Pa}_{G_k}$  が制御するモデルを、尤度関数を以下のように適用して評価する．

$$\text{score}_k = L(G_k, \text{Pa}_{G_k} : X_{(G, C_k)})$$

ただし、上の式では  $B_1, \dots, B_K$  の間で異なった部分行列  $X_{(G, C_k)}$  が混在するため、ネットワーク全体の尤度は求められない．そこで、2.2 節で示したように尤度関数をモジュールごとに独立に求められることを利用して、モジュールごとに独立に制御関係の推定を行う．また、同じ理由からモジュール推定の段階と制御関係の推定の段階を繰り返すことができない．この

ため、2つの段階を1回だけ行った後に、最も評価が高いモデルを解として出力する。

## 5. 評価実験

遺伝子発現プロファイルから遺伝子制御ネットワークを推定する実験を、以下の条件で行った。発現プロファイルは、遺伝子発現データベース GEO<sup>7)</sup> から得られる出芽酵母の発現プロファイルを用いた。出芽酵母は多細胞生物に比べて遺伝子間の制御関係が単純であり、さらに制御関係の研究が比較的進んでいるため、制御関係の推定精度の評価に適していると考えられる。複数の発現プロファイルの実験条件を連結して、448個の実験条件を含む発現プロファイル  $X$  を作成した。ただし、名称に dye swap を含む発現プロファイルや、実験条件数が5個未満の発現プロファイルは利用しなかった。また、2色蛍光のマイクロアレイによる発現プロファイルのみを利用した。遺伝子間の転写制御関係のデータベースである TRANSFAC<sup>16)</sup> に記載された制御関係に登場する、91個の制御遺伝子と、これらの制御遺伝子に制御される182個の遺伝子(被制御遺伝子)の間の制御関係を、推定の対象とした。この91個の制御遺伝子を、制御遺伝子候補の集合  $G_c$  として推定で用いた。

TRANSFACに記載された91個の制御遺伝子と182個の被制御遺伝子の間の276個の制御関係を、既知の制御関係として推定結果の評価に用いた。TRANSFACに記載された制御関係は、DNAマイクロアレイ実験の結果に基づいて推定を行う本研究のような間接的な方法ではなく、直接的により正確な制御関係を検出できるフットプリント法のような実験の結果に基づいている。フットプリント法<sup>17)</sup> はDNA配列とタンパク質の結合を調べるための実験である。制御遺伝子がコードする転写因子タンパク質と、その被制御遺伝子がコードされたDNA配列の近傍に位置する領域の結合を調べることによって、これらの遺伝子の間の制御関係を調べることができる。ただし、このような実験では制御関係を網羅的に調べることはできない。

実験条件の選別を行わずに、モジュールネットワークモデルに基づいた推定を行う手法を従来手法とした。モデルの評価には、ベイズ統計の下で最小記述長基準(MDL基準)<sup>18)</sup> に基づいて定義された尤度関数を用

いた。発現プロファイルで欠損値を示す要素は全体の平均値で置き換えた。2色蛍光のマイクロアレイによる測定値は対数正規分布を示し、平均値付近が最頻値となるため、平均値で置き換えた場合に結果に与える影響が最も小さくなる。本研究で用いた発現プロファイルでは、欠損値は全体の3%程度であったため、この扱いが結果に与える影響は小さいと考えられる。

以下では、まず5.1節で、本研究で提案する発現プロファイルの実験条件の選別に関して、妥当性の検証を行った結果を示す。次に5.2節で、提案手法と従来手法によって遺伝子制御ネットワークを推定し、その推定結果の評価結果について比較した結果を示す。

### 5.1 実験条件選別の効果の検証

本論文で提案したように、遺伝子発現パターンの類似性から選別した実験条件のみを利用することが、制御関係の推定にどのように寄与するか検証を行った。

まずはじめに、実験条件の選別によって、制御関係の推定精度が向上するか、例を示して個別に検証した。遺伝子 EUG1 と PDI1 は、遺伝子 HAC1 に制御されることが知られている。これらの遺伝子 EUG1 と PDI1 からなるモジュールに対して、制御遺伝子を推定した。

その結果、すべての実験条件を用いて EUG1 と PDI1 を制御する遺伝子を推定した場合には、既知の制御遺伝子 HAC1 とはまったく異なる発現パターンを示す制御遺伝子 REB1 が推定された。一方で EUG1 と PDI1 が類似した発現パターンを示す実験条件のみを用いて、この2つの遺伝子を制御する遺伝子を推定すると、HAC1 と類似した発現パターンを示す制御遺伝子 CDC10 が推定された。

本手法のように遺伝子発現プロファイルのみから推定を行う場合には、発現パターンが類似した遺伝子どうしを区別することは困難であると考えられる。このため、既知の制御遺伝子と発現パターンの類似した遺伝子が、制御遺伝子として推定される傾向がある。そこで本研究では、推定された制御遺伝子だけでなく、それと発現パターンの類似した遺伝子も推定された制御遺伝子の候補として提示する。ここでは遺伝子の発現パターンの類似性を測る指標として相関係数を用いた<sup>15)</sup>。REB1 と HAC1 の間の相関係数が0.23と無相関に近かったのに対して、CDC10 と HAC1 の間の相関係数は0.75と比較的強い相関があった。

以上から、推定された制御遺伝子の発現パターンが、既知の制御遺伝子の発現パターンとどれだけ類似しているかで、推定結果の評価を行うことにする。しかし、相関係数は発現パターンに含まれている実験条件の数

dye swap は同じ実験を繰り返すことによって、測定誤差を評価するために行った実験を指す。連結した発現プロファイルに含まれる実験条件が重複するのを防ぐため、dye swap 実験に相当する発現プロファイルは利用しなかった。

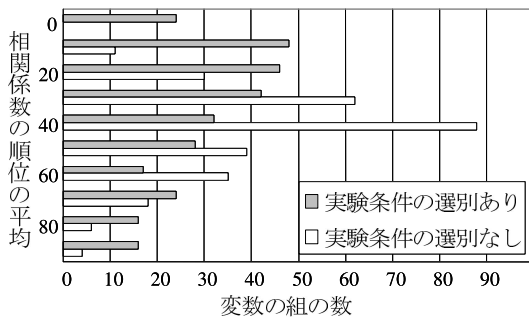


図 4 推定に利用する実験条件を選別した場合とすべての実験条件を用いた場合の間の相関係数の順位と比較

Fig. 4 Comparing the ranking of correlation coefficients in inferences with selections of experimental conditions to the ranking of ones without selections.

が多いほど低下する傾向にあるため、相関係数の値そのもので評価するのは適切ではないと考えられる。そこで、相関係数の値の大きさではなく順位から評価を行った。まず、推定された遺伝子と制御遺伝子候補に含まれる 91 個の遺伝子の中で相関係数を求めた。そして、既知の制御遺伝子との相関係数が 91 個中の上位何番目に入っているかによって、推定された制御関係の妥当性を評価した。利用する実験条件を選別した場合に制御遺伝子として推定された CDC10 の順位は 3 位だったのに対して、すべての実験条件を利用した場合に制御遺伝子として推定された REB1 の順位は 36 位だった。

このように、被制御遺伝子 EUG1, PDI1 が類似した発現パターンを示す実験条件のみを用いて制御関係を推定することによって、既知の制御遺伝子 HAC1 そのものは推定できなかったが、既知の制御遺伝子と発現パターンの類似した遺伝子 CDC10 を制御遺伝子として推定することができた。以上から、既存の手法よりも精度の高い推定が可能になることが期待される。

そこで次に、上の制御関係の推定の例と同様の方法を用いて、全体の傾向について調べた。同じ遺伝子に制御されることが既知の任意の 2 つの遺伝子からなるモジュールを作り、それぞれのモジュールを制御する遺伝子を推定した。推定された制御遺伝子と既知の制御遺伝子の発現パターンが類似しているか調べた。推定された制御関係の評価は、上で示した例と同様の方法で行った。

2 つの遺伝子が類似した発現パターンを示す実験条件のみを選別して用いた場合の推定結果と、すべての実験条件を用いた場合の推定結果を比較した結果を、図 4 に示した。図 4 は、それぞれの結果の相関係数の順位をヒストグラムで示している。推定に利用する

表 1 提案手法と従来手法の正解率の比較

Table 1 Comparing accuracy rates of proposed method with ones of the conventional method.

推定手法	正解率 (一致)	正解率 (類似)
提案手法 (Plaid)	2.0%	11.2%
提案手法 (SAMBA)	2.9%	9.0%
モジュールネットワーク	2.2%	6.3%

実験条件を選別した場合の相関係数の順位分布は、すべての実験条件を用いた場合に比べて、順位の高いグラフの上側に偏っていることが分かる。したがって図 4 が示す比較結果は、発現パターンの類似性から利用する実験条件を選別した場合の方が、多くの制御関係について良い推定結果を示す傾向があることを示していると考えられる。

## 5.2 遺伝子制御ネットワークの推定

既知の制御関係における被制御遺伝子 182 個に対して、これらの遺伝子を制御する遺伝子を推定した。提案手法と従来手法それぞれを用いて、制御関係の推定を行った。提案手法では、バイクラスタリングの手法として Plaid モデルおよび SAMBA を利用した。提案手法の性能評価のため、推定された制御関係それぞれを既知の制御関係を用いて評価し、各手法の推定結果の正解率を比較した。

まずはじめに、推定された制御関係が既知の制御関係と一致する場合のみを正解として、推定結果の正解率を評価した。表 1 の 2 列目 (正解率 (一致)) にその結果をまとめた。この結果から、提案手法は従来手法とほとんど変わらない正解率しか示さなかったことが分かる。SAMBA を用いた場合は従来手法よりも精度が高いが有意な差があるとはいえず、Plaid モデルを用いた場合では従来手法よりも正解率が低い。

5.1 節で述べたように、本手法では類似した遺伝子発現パターンを示す遺伝子どうしを区別することは困難であるため、推定された制御関係が既知の制御関係と正確に一致することはあまり期待できない。そこで次に 5.1 節と同様に、推定された制御遺伝子と既知の制御遺伝子との相関係数が全体の上位何番目に入るかによって、推定された制御関係の正否を評価した。既知の制御遺伝子との相関係数が上位 5 番目までに入っている制御関係を正解とする場合について、正解率をまとめたのが表 1 の 3 列目 (正解率 (類似)) である。これらの正解率から、提案手法を用いた結果は、従来手法よりも高い正解率を示したことが分かる。提案手法どうしを比べた場合には、Plaid モデルの方が SAMBA を用いた場合よりも、高い正解率を示したこ



とが分かる。

推定された制御関係を正解とする順位の閾値として、5以外の値を選んだ場合について正解率を比較した場合にも、提案手法は従来手法よりも高い正解率を示した。また、閾値の値によっては、Plaid モデルを用いた結果と SAMBA を用いた結果の間で、正解率の大小の関係が入れ替わる場合があり、Plaid モデルが SAMBA よりつねに高い正解率を示したわけではなかった。

## 6. 考 察

表 1 の推定精度に関して考察を行う。まずはじめに、提案手法と従来手法の間の正解率の差について考察し、次に、提案手法の正解率の大きさについて考察する。

表 1 の正解率の差について考察する。表 1 で示したように、推定された制御関係と既知の制御関係が一致する場合のみを正解とした場合には、提案手法と従来手法の正解率はほとんど同じだった。遺伝子の制御関係はネットワークを形成しているため、制御遺伝子自体も他の制御遺伝子によって制御されることが多い。このため制御遺伝子と同じ遺伝子によって制御される遺伝子群が存在し、それらの遺伝子群は制御遺伝子と類似した遺伝子発現パターンを示す。推定の際には、このような類似した発現パターンを示す遺伝子が制御遺伝子を最適解から覆い隠したため、最適解に偶然による選択が占める割合が大きくなったと考えられる。表 1 で示した正解率（一致）が提案手法と従来手法で変わらない要因として、以上のような説明が考えられる。

表 1 の正解率（一致）は提案手法と従来手法の間でほとんど差が見られないため、ほぼ同じ制御関係が推定されたとも考えられる。推定された制御関係が同じならば、正解率（類似）の向上は推定精度の改善を意味しないため、それぞれの手法で推定された制御関係がどの程度異なるかを調べた。その結果として、提案手法（Plaid）によって推定された制御関係の 34.1% は、従来手法で推定された制御関係と重複していた。同様に提案手法（SAMBA）の 31.8% が従来手法と重複していた。提案手法と従来手法では推定された制御関係は十分に異なっており、正解率（類似）の向上には推定精度の改善が寄与していると考えられる。

次に、提案手法の正解率の大きさは、十分な水準に達しているか考察する。計算機上で推定された制御関係は実用するには信頼性が不十分のため、5 章のはじめに TRANSFAC についての説明で述べたような

実験を行って確認する必要がある。しかし、表 1 で示した正解率は、確認実験を行うには十分とはいえない。確認実験の費用は一般に高額なため、失敗を前提として確認実験を計画することはできない。計画した確認実験の 7~8 割程度で制御関係の検出が成功する必要があると考えられる。ただし、表 1 の正解率は、TRANSFAC にまだ記載されていない、未知の制御関係が多数存在することを考慮していない。そこで仮に全体の半数程度の制御関係が判明していると仮定しても、表 1 の正解率では、制御関係の検出の成功が期待できるのは計画した確認実験の 2 割程度のみとなる。計画した確認実験の 7~8 割程度が成功することを期待できるようにするには、本研究の評価実験の条件でも、少なくとも 30% 以上の正解率が必要となると考えられる。したがって、表 1 の正解率はまだ十分な水準とはいえない。

本研究で十分な正解率が達成できない理由として、以下の 3 つの要因があげられる。

- (1) 発現プロファイルの測定精度がまだ十分ではない<sup>19)</sup>。
- (2) 細胞が示す状態の多くは発現プロファイルに観測されていない<sup>19)</sup>。
- (3) 発現プロファイルから推定することが困難な制御関係も TRANSFAC に記載されている。

(1) と (2) による影響は、DNA マイクロアレイなどの測定技術の精度が現在も向上しつつあることと、遺伝子発現データベースへの登録数が加速度的な増加傾向を示していることから、今後緩和される傾向にあると期待できる。(3) はたとえば、制御遺伝子の発現量とその遺伝子がコードする転写因子の活性の強さが連動しない場合には、発現プロファイルから制御関係を推定することは難しいことがあげられる。迅速な制御が求められるような環境変化への応答においては、制御遺伝子の発現を介さずに転写因子を直接的に活性化して、必要な制御を行うことが知られている。このような制御関係に関しては、被制御遺伝子の発現量が変化している場合でも、制御遺伝子の発現量には変化が観測されないことがあると予測される。このような問題は、発現プロファイルのみから推定を行う限りは、本質的には解決できないと考えられる。

## 7. おわりに

遺伝子発現プロファイルに基づく遺伝子制御ネットワークの推定に、バイクラスタリングを利用する方法を提案した。発現プロファイルの蓄積によって得られるようになった、多くの実験条件を含む発現プロファ

イルを用いて遺伝子制御ネットワークを推定する場合に、この手法は有効である。遺伝子発現パターンが類似した実験条件を選別することによって、制御関係の推定の正解率が向上することを検証し、出芽酵母の遺伝子制御ネットワークの推定結果の正解率を評価した。

今後の研究課題として、制御関係の推定結果をバイクラスタリングの段階にフィードバックする方法を検討することがあげられる。

謝辞 本研究は一部、服部報公会工学研究奨励援助金によっている。

### 参 考 文 献

- 1) 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).
- 2) Chen, T., He, H.L. and Church, G.M.: Modeling Gene Expression with Differential Equations, *Proc. 4th Pacific Symposium on Biocomputing (PSB'99)*, pp.29-40 (1999).
- 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) Hartigan, J.A.: *Clustering Algorithms*, John Wiley & Sons, Inc., New York, NY, USA (1975).
- 5) Segal, E., Pe'er, D., Regev, A., Koller, D. and Friedman, N.: Learning Module Networks, *Proc. 19th Annual Conference on Uncertainty in Artificial Intelligence (UAI-03)*, Acapulco, Mexico, pp.525-534, Morgan Kaufmann Publishers (2003).
- 6) Ikeo, K., Ishi-i, J., Tamura, T., Gojobori, T. and Tatenoy, Y.: CIBEX: Center for Information Biology gene EXpression database, *Comptes rendus biologiques*, Vol.326, No.10, pp.1079-1082 (2003).
- 7) Barrett, T., Suzek, T.O., Troup, D.B., Wilhite, S.E., Ngau, W.C., Ledoux, P., Rudnev, D., Lash, A.E., Fujibuchi, W. and Edgar, R.: NCBI GEO: Mining millions of expression profiles—database and tools, *Nucleic Acids Research*, Vol.33, Database Issue, pp.D562-D566 (2005).
- 8) Brazma, A., Parkinson, H., Sarkans, U., Shojatalab, M., Vilo, J., Abeygunawardena, N., Holloway, E., Kapushesky, M., Kemmeren, P., Lara, G.G., Oezcimen, A., Rocca-Serra, P. and Sansone, S.A.: ArrayExpress — a public repository for microarray gene expression data at the EBI, *Nucleic Acids Research*, Vol.31, No.1, pp.68-71 (2003).
- 9) Cheng, Y. and Church, G.M.: Biclustering of Expression Data, *Proc. 8th International Conference on Intelligent Systems for Molecular Biology (ISMB 2000)*, pp.93-103, AAAI Press (2000).
- 10) Madeira, S.C. and Oliveira, A.L.: Biclustering Algorithms for Biological Data Analysis: A Survey, *IEEE/ACM Trans. Computational Biology and Bioinformatics*, Vol.1, No.1, pp.24-45 (2004).
- 11) Lazzeroni, L. and Owen, A.: Plaid Models for Gene Expression Data, *Statistica Sinica*, Vol.12, No.1, pp.61-86 (2002).
- 12) Tanay, A., Sharan, R. and Shamir, R.: Discovering statistically significant biclusters in gene expression data, *Bioinformatics*, Vol.18, Suppl.1, pp.S136-S144 (2002).
- 13) DeRisi, J.L., Iyer, V.R. and Brown, P.O.: Exploring the Metabolic and Genetic Control of Gene Expression on a Genomic Scale, *Science*, Vol.278, No.5338, pp.680-686 (1997).
- 14) 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, pp.308-312 (2004).
- 15) Spellman, P.T., Sherlock, G., Zhang, M.Q., Iyer, V.R., Anders, K., Eisen, M.B., Brown, P.O., Botstein, D. and Futcher, B.: Comprehensive Identification of Cell Cycle-regulated Genes of the Yeast *Saccharomyces cerevisiae* by Microarray Hybridization, *Molecular Biology of the Cell*, Vol.9, No.12, pp.3273-3297 (1998).
- 16) Matys, V., Fricke, E., Geffers, R., Gößling, E., Haubrock, M., Hehl, R., Hornischer, K., Karas, D., Kel, A.E., Kel-Margoulis, O.V., Kloos, D.U., Land, S., Lewicki-Potapov, B., Michael, H., Münch, R., Reuter, I., Rotert, S., Saxel, H., Scheer, M., Thiele, S. and Wingender, E.: TRANSFAC<sup>®</sup>: Transcriptional regulation, from patterns to profiles, *Nucleic Acids Research*, Vol.31, No.1, pp.374-378 (2003).
- 17) 大藤道衛, 高井貴子, 高木利久: これからのバイオインフォマティクスのためのバイオ実験入門, 羊土社 (2002).
- 18) Lam, W. and Bacchus, F.: Learning Bayesian Belief Networks: An Approach Based on the MDL Principle, *Computational Intelligence*, Vol.10, pp.269-293 (1994).
- 19) Kohane, I.S., Kho, A. and Butte, A.J.: *Microarrays for an Integrative Genomics*, MIT Press, Cambridge, MA, USA (2002). 星田有人

(訳): 統合ゲノミクスのためのマイクロアレイデータアナリシス, シュプリング・フェアラーウ東京 (2004).

(平成 17 年 11 月 24 日受付)  
(平成 18 年 1 月 11 日再受付)  
(平成 18 年 2 月 1 日採録)



瀧 浩平 (学生会員)

昭和 54 年生。平成 14 年大阪大学基礎工学部情報科学科卒業。平成 16 年同大学院情報科学研究科バイオ情報工学専攻 (修士課程) 修了。現在同研究科博士後期課程在学中。



竹中 要一 (正会員)

昭和 48 年生。平成 7 年大阪大学基礎工学部情報工学科中退。平成 9 年同大学院基礎工学研究科物理系専攻情報工学分野 (博士前期課程) 修了。平成 12 年同研究科情報数理系計算機科学分野 (博士後期課程) 修了。同年同大学院基礎工学研究科助手となり、平成 14 年より同大学院情報科学研究科助教授、現在に至る。工学博士。遺伝子情報処理の研究に従事。電子情報通信学会, IEEE, 日本バイオインフォマティクス学会, 日本知財学会各会員。



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)。同年同大学助手。1994 年大阪大学助教授。2002 年より現職。バイオインフォマティクス, データグリッドの研究に従事。日本バイオインフォマティクス学会, 日本分子生物学会, ISCB 各会員。