

## 転写制御ネットワークと ncRNA-タンパク質相互作用 の構造と機能の関係

荒木直央<sup>†1</sup> Jose Nacher<sup>†1</sup>

本研究は、転写制御ネットワークと non-coding RNA が関与する相互作用ネットワークの構造と頂点を持つ分子機能の分布の関係について分析を行う。分析には2つのパラメータを用い、ネットワークが2つの異なる機能クラスに属するときの頂点の分布を特定した。この結果、転写制御ネットワークと ncRNA 相互作用ネットワークの双方において、ネットワーク構造と機能の関係がランダムな分布とは大きく異なることが示された。このことから、ncRNA 分子が転写制御機能に関係している可能性が考えられる。

### On the relation between structure and biological function in transcriptional networks and ncRNA-mediated interactions

NAO ARAKI <sup>†1</sup> and JOSE NACHER<sup>†1</sup>

We investigate the distribution of molecular functionalities in transcriptional networks and non coding RNA-mediated interactions. When network nodes belong to two distinct functional classes, the networked structure is analyzed by means of two parameters. The results show that both transcriptional and ncRNA-interactions exhibit far from random couplings between network and functionality, supporting growing evidences for ncRNA regulatory capability.

#### 1. はじめに

遺伝情報は「DNA→mRNA→タンパク質」という経路で伝達されるという考えは一般的な概念である。この考えより、遺伝子とタンパク質を一对一対応させて考えることが可能と

なる。単純な原核生物においては、この対応の正当性が確認されている。しかし近年の研究により、タンパク質をコードしていない遺伝子 (non protein coding gene) の割合が、複雑な生体機能をもつ高等生物になるに従い高くなることが分かってきた<sup>1)</sup>。高等生物の代表ともいえるヒトでは、その比率は98.5%となる。高等生物では non coding RNA (ncRNA) への転写が多いというこの発見は、現在一般的とされている遺伝情報の伝達過程では説明ができない。このことから、遺伝情報の伝達に係わる、新たな伝達過程の存在が示唆される<sup>1)-3)</sup>。この過程は RNA 分子が高等生物の転写に適応し、進化したものと考えられる。

ncRNA とはタンパク質に翻訳されない RNA 分子のことを指す。ncRNA はタンパク質に翻訳されない他の転写物と同じく、遺伝子のイントロン領域にコードされている。一部の機能が知られている ncRNA 分子には転写 RNA (tRNA) やリボソーム RNA (rRNA), snoRNAs, microRNAs, siRNAs, piRNAs などが存在する<sup>3)</sup>。

本研究では、転写制御ネットワークと ncRNA 相互作用において、構造と頂点を持つ分子機能の関係について分析を行う。分析の対象は有機体 *S. cerevisiae* の転写制御ネットワーク、および同有機体の ncRNA 相互作用ネットワークとする。対象のネットワークはインターネット上に公開されているデータベースより作成する。転写制御ネットワークは遺伝子発現の制御を行う転写因子 (TFs) と制御される標的遺伝子によって定義される転写制御のデータから作成した。ncRNA 相互作用ネットワークは、ncRNA とタンパク質の機能的相互作用データを利用している<sup>4),5)</sup>。

分析を行う頂点機能については、各頂点 (分子) が持つ生体機能の情報を用いた。この情報に従う分類によって、各頂点はいくつかの機能クラスに属することになる。機能クラスに属する場合、その頂点状態を1で表し、属さない場合は0とする。本研究では参考文献6), 7)の分析手法に従い  $D$ ,  $H$  の2つのパラメータを用いて、各機能クラスに属する頂点の分布について調べる。この手法は過去に携帯電話サービスとタンパク質の相互作用の機能的な特徴についての研究に用いられおり、機能的特徴とネットワーク構造の関係について特定することに成功している<sup>6)</sup>。本研究では上記の手法を用いて「ncRNA が類似した機能を持つタンパク質との間に辺を持つ可能性は、ランダムな場合と比べ高くなるのか」という問いに対する答えを提示する。また、同様の分析を転写制御ネットワークに対しても行い、転写制御ネットワークと ncRNA 相互作用ネットワークの構造と機能の関係を比較し、ncRNA が制御機能を持つ可能性について考察する。

<sup>†1</sup> 公立はこだて未来大学 Future University-Hakodate

ID	機能クラス (ncRNA, タンパク質の頂点数)	$D$	$H$
1	DNA stability (1, 0)	...	3.0311
2	DNA transcription (0, 10)	...	0.7965
3	RNA modification (15, 7)	0.0000	0.6034
4	RNA processing (10, 81)	1.5278	0.3672
5	RNA translation (0, 6)	...	0.4670
6	RNA translocation (0, 2)	...	0.3394
7	Protein transport (1, 2)	28.9637	0.2281

表 1 ncRNA-タンパク質ネットワークの各機能クラス  
Table 1 functional class and  $D$ ,  $H$  of ncRNA-protein network

ID	機能クラス (TF, 標的遺伝子の頂点数)	$D$	$H$
1	DNA stability (0, 0)	...	...
2	DNA transcription (120, 323)	2.0032	1.2581
3	RNA modification (0, 27)	...	0.4416
4	RNA processing (4, 192)	0.1425	0.7975
5	RNA translation (1, 295)	0.5546	0.7786
6	RNA translocation (1, 130)	0.0000	0.6050
7	Protein transport (0, 67)	...	0.5278

表 2 転写制御ネットワークの各機能クラス  
Table 2 functional class and  $D$ ,  $H$  of TFs-gene network

## 2. ネットワーク構造と機能クラス

分析対象の転写制御ネットワークと ncRNA-タンパク質相互作用ネットワークはデータベースの情報をもとに作成した<sup>5),8)</sup>。作成したネットワークは、2つの頂点集合からなる二部グラフと定義する。ネットワークの辺は TFs から標的遺伝子（または、ncRNA からタンパク質）への機能的な相互作用を意味する。利用したデータは、転写制御ネットワークでは TFs 121 個と遺伝子 896 個を含み、ncRNA 相互作用では 25 個の ncRNA 分子と 105 個のタンパク質から成る。また、各頂点の機能クラスに関する情報は NONCODE 9) と MIPS 10) から取得した。転写制御ネットワークと ncRNA-タンパク質相互作用ネットワークの各機能クラスへの分類を表 1, 2 の 1, 2 列目に示す。

## 3. ネットワークの分析方法

### 3.1 頂点と辺の分類

ネットワークの各頂点に対して、特定の機能クラスに属する場合は 1, 属さない場合は 0

の値をとるとする。値 1 の頂点の個数は  $n_1$ , 値 0 の頂点個数は  $n_0$  で表す。よって、頂点の総数は  $N = n_1 + n_0$  となる。両端の頂点が 1-1 (1-0) の値をとる辺の数は  $m_{11}$  ( $m_{10}$ ) によって表す。値 1 の頂点がネットワーク内にランダムに分布する場合  $m_{11}$  と  $m_{10}$  の期待値は次のようになる。

$$\overline{m_{11}} = \binom{n_1}{2} \times p = \frac{n_1(n_1-1)}{2} p \quad (1)$$

$$\overline{m_{10}} = \binom{n_1}{1} \binom{n_0}{1} \times p = n_1(N-n_1)p \quad (2)$$

ここでは、 $p = 2M/N(N-1)$ ,  $M$  は辺の総数  $M = m_{11} + m_{10} + m_{00}$  を示す<sup>11)</sup>。実際のネットワークの  $m_{11}$  と  $m_{10}$  の値と、ランダムな場合の  $m_{11}$  と  $m_{10}$  の期待値を比べることにより、実際ネットワークの特徴を示すパラメータ  $D$ ,  $H$  が求まる。

### 3.2 頂点間の関係を表すパラメータ $D$ , $H$

$D$  dyadicity と  $H$  heterophilicity の 2 つのパラメータを求めることにより、実際ネットワークの  $m_{11}$ ,  $m_{10}$  の値とランダムなネットワークの場合の期待値との偏差を調べることができる<sup>6)</sup>。

$$D = \frac{m_{11}}{\overline{m_{11}}}, \quad H = \frac{m_{10}}{\overline{m_{10}}} \quad (3)$$

パラメータ  $D$  の値が高い場合 ( $D > 1$ )、ネットワークはランダムな場合と比べて、対象としている機能クラスに属する頂点  $n_1$  同士の結びつきが強い事を表し、そのネットワークは dyadic な特徴を持つと言える。 $D < 1$  となる anti-dyadic なネットワークは、同じ機能クラスに属する頂点同士は辺で結ばれない傾向にある事を示す。パラメータ  $H$  は機能クラスに属する頂点  $n_1$  と属さない頂点  $n_0$  の辺の繋がりについて表している。 $H > 1$  は異なる機能に属する頂点同士が繋がる傾向にあり (heterophilic),  $H < 1$  は繋がり避ける傾向にあ

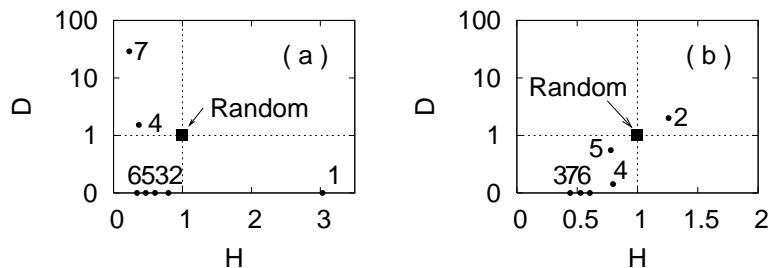


図1 ncRNA-タンパク質ネットワーク (a) と転写制御ネットワーク (b) の  $D, H$   
Fig.1 Distribution  $D, H$  for ncRNA-protein network(a) and transcriptional regulatory network(b)

ることを示す (heterophobic).  $D$  と  $H$  の取り得る値の範囲は、機能に属する頂点の個数  $n_1$  とネットワーク構造によって制限を受ける. この  $D, H$  の値が取り得る範囲を二次元上で表したものが境界面図である. 境界面と実際の  $D, H$  の値を比較することにより, 対象ネットワークの構造と  $n_1$  の配置の関係について確認することができる.

#### 4. 結果

ネットワーク構造と機能の関係では, データベースから集めた生体機能の情報と分子間の相互作用の情報を用いて, 各頂点 (分子) がもつ機能と構造の関係について分析を行った. 分析の対象は, 有機体 *S. cerevisiae* の転写制御ネットワークと同じ有機体の ncRNA-タンパク質相互作用ネットワークとした. 分析では, ランダムな場合と実際ネットワークとの偏差を表す  $D$  と  $H$  の値を求め, 2つのネットワークの比較を行った.

##### 4.1 転写制御ネットワーク

転写制御ネットワークでは, 分析を行った大半の機能クラスの結果 ( $D, H$ ) が heterophobic な状態を示した. また, 機能クラス 2 では dyadic で heterophilic な特徴を持つことが分かった (表 2, 図 1 (b)).

##### 4.2 ncRNA 相互作用ネットワーク

ncRNA 相互作用ネットワークの分析では, 大半の機能クラスが heterophobic な特徴を持つことが分かる (表 1). この結果は, 転写制御ネットワークの結果と類似している. heterophobic を示した機能は, 機能クラス番号 2 の DNA transcription, 3.RNA modification,

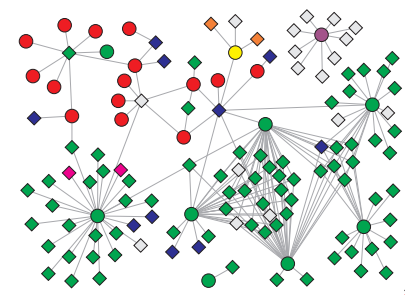


図2 *S. cerevisiae* の ncRNA-タンパク質ネットワーク  
Fig.2 ncRNA-protein network of *S. cerevisiae*

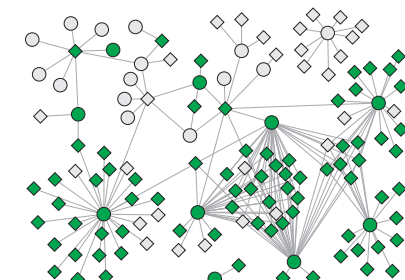


図3 機能クラス RNA processing のネットワーク  
Fig.3 RNA processing functional class

4.RNA processing, 5.RNA translation, 6.RNA translocation, 7.Protein transport である. さらに RNA processing と Protein transport は dyadic な分布も示している. ncRNA-タンパク質ネットワークの大半の機能クラスが heterophobic な特徴を持つという結果と, 転写制御ネットワークの特徴の結果が一致していることは注目すべき点である. 関連研究 6) のタンパク質相互作用ネットワークの結果では 1つ機能クラスのみが anti-dyadic な分布をとり, 残りの機能クラスは全て dyadic と heterophobic な特徴を示している. ncRNA-タンパク質相互作用ネットワークの各機能クラスの  $D, H$  の値を図 1 (a) に示す. この図より, ncRNA 相互作用ネットワークの各機能クラスは, ランダムな分布とは異なる値をとることが分かる. 図 2, 3 は ncRNA-タンパク質相互作用ネットワークと機能クラス 4.RNA processing のネットワークである.

##### 4.3 ネットワークの境界面

機能クラスに属する頂点  $n_1$  と属さない頂点  $n_0$  をネットワーク上に配置し, その配置の組合せを全て計算することによってパラメータ  $D, H$  の境界面図を作成することができる. この方法は小さなネットワークに対しては有効であるが, 巨大なネットワークの場合は計算量が膨大になる. よって, ヒューリスティックアルゴリズムを利用し, 最適な  $D, H$  の値を特定するほうが効率的である. 本研究では Ant Colony Optimization (ACO) と Fiduccia-Mattheyses (FM) アルゴリズムを利用し  $D, H$  の境界面図を計算した (図 4). 図は実線が ACO, 破線が FM の境界面, 白い四角は実際の  $D, H$  の値を示す. 図 4 (a) は ncRNA-タンパク質ネットワークの機能クラス 3, (b) は機能クラス 4, (c) は 5, (b) は

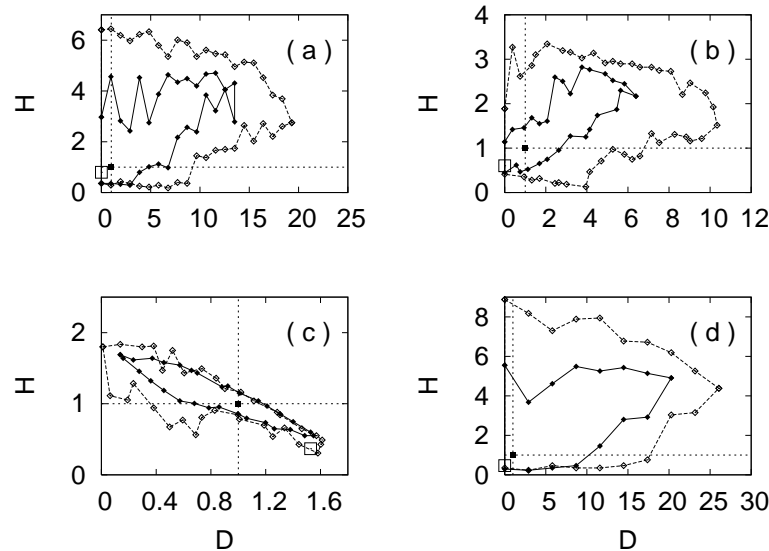


図4 ACO と FM による ncRNA-タンパク質ネットワークの境界面

Fig. 4 Phase boundaries for the ncRNA-protein network. Results are shown for ACO and FM algorithms.

6の境界面である。この結果から境界面の形、大きさは2つのアルゴリズムで類似したものとなる事が分かる。しかし、FMアルゴリズムのほうが早い計算時間で、より最適な境界面（境界内の面積積を最大とするための $D$ ,  $H$ の最大値、最小値）を見つけることができた。ACO, FMのアルゴリズムの概略を次に記す。

ACOによる計算方法：

1. ネットワークを隣接行列によって表現し、全ての頂点を $n_0$ に初期化する。辺に対応するフェロモン行列は値0.001の完全行列として初期化する。
2. 各蟻はランダムに1つの頂点を選択する。選択された頂点は機能クラスに属する頂点 $n_1$ となる。次に蟻は $n_0$ の頂点を辺のフェロモン値に従い優先的選択によって選択する。ステップ2を頂点 $n_1$ が決められた個数に達するまで繰り返す。この過程を全ての蟻において行う。

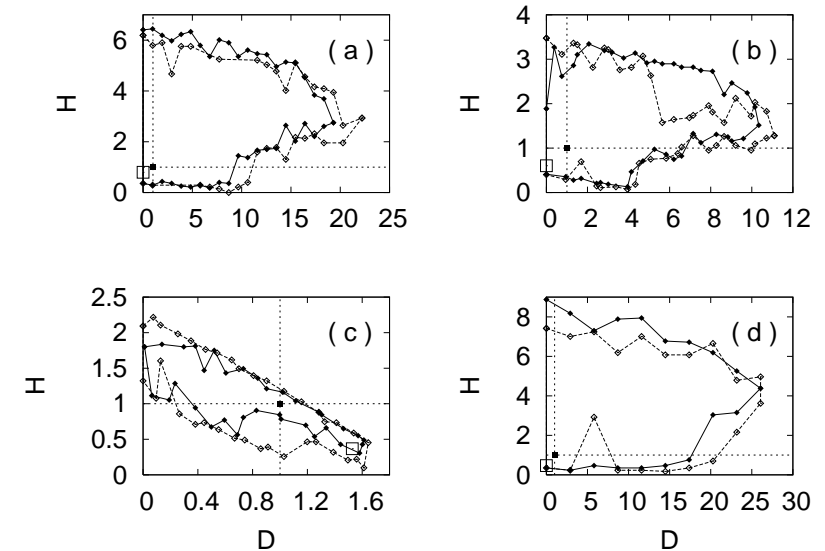


図5 FMアルゴリズムによる ncRNA-タンパク質ネットワークの境界面

Fig. 5 Phase boundaries for the ncRNA-protein network. Results are shown for FM algorithm.

3. 全ての蟻の $m_{11}$ と $m_{10}$ を計算し、最適な値を保存する。
4. フェロモンの蒸発率と最適な蟻の $m_{11}$ ,  $m_{10}$ の値を用いて、フェロモン行列を更新する。
5. ステップ1から4を100回繰り返す。

FMによる計算方法：

1. ネットワーク上に、機能クラスに属する頂点 $n_1$ をランダムに配置する。
2. 1つ $n_1$ 頂点と、交換する候補 $n_0$ 頂点（複数個）を選択する。
3. 選択した頂点 $n_1$ と $n_0$ の機能クラスを交換した場合の $m_{11}$ ,  $m_{10}$ の増加量を計算する。
4.  $m_{11}$ ,  $m_{10}$ の増加量が最適となる $n_1$ ,  $n_0$ の組合せを選び、機能クラスを交換する。
5. 交換後のネットワーク全体の $m_{11}$ と $m_{10}$ の値を計算し、最小値（最大値）の更新を行う。
6. ステップ1から5を、まだ交換されていない $n_1$ に対して繰り返す（ $n_1 \times n_0$ 回）。
7. ステップ1から5を100回繰り返す。

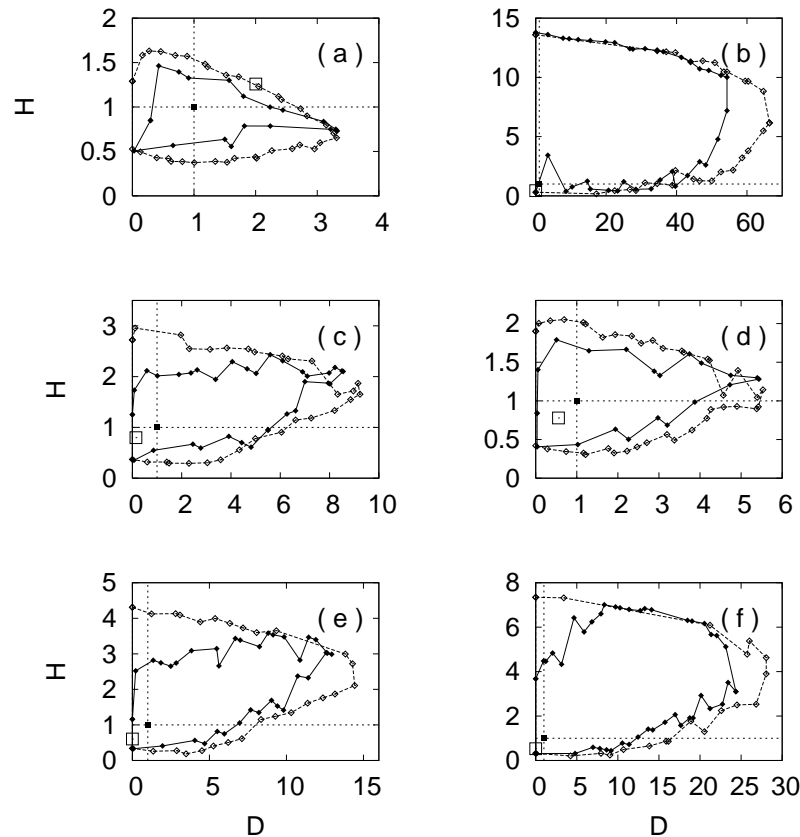


図 6 FM アルゴリズムによる転写制御ネットワークの境界面

Fig. 6 Phase boundary for the transcriptional regulatory network. Results are shown for FM algorithm.

FM については、上記の計算方法と参考文献 (6), (13) の方法による境界面を求め結果の比較を行った (図 5, 6) . 図の実線は上記の方法による境界面、破線は参考文献の結果である。白い四角は実際ネットワークの  $D$ ,  $H$  の値を表す。図 5 の (a-d) は機能クラス 3, 4, 5, 6 と対応している。図 6 の (a-f) は機能クラス 2, 3, 4, 5, 6, 7 の境界面と対応している。これらの図より、計算方法が異なる境界面の結果は大きさは異なるが、形は類似していることが分かる。

## 5. 考察と結論

本研究では、ネットワーク構造と機能の関係を特定するために 2 つの分析を行った。1 つ目は、転写制御と ncRNA-タンパク質ネットワークのネットワーク構造と機能の関係と、その結果の比較である。ncRNA-タンパク質ネットワークにおいて、分析した機能クラスは heterophobic な分布であることが分かった。このことは、大部分の ncRNA が、同じ機能クラスに属するタンパク質との間に辺を持つ傾向があることを示している。この heterophobic な分布は、転写制御ネットワークにおいても確認できた。これは、いくつかの機能において、TFs も同じ機能に属する標的遺伝子との間に辺を持つ傾向があることを示している。この結果は、転写制御ネットワークと ncRNA 相互作用においてネットワーク構造と機能の関係が類似していることを示している。

2 つ目の分析では、転写制御ネットワークと ncRNA 相互作用ネットワークの各機能クラスの  $D$ ,  $H$  と境界面図との比較を行った。この結果、 $D$ ,  $H$  の分布は、境界面上にランダムに分布するのではなく、より heterophobic な部分に位置していることが分かった。ncRNA 相互作用の構造と機能の分布がランダムでないという結果から、ncRNA 分子が特異的な生体機能を持つ可能性が考えられる<sup>1),14)</sup>。近年の研究により、ncRNA 分子が細胞分裂や腫瘍形成の制御過程に関係することが分かってきた。その他にも、ncRNA 分子が冠状動脈障害や糖尿病のような複雑な病気と関係していることも解明されつつある<sup>2)</sup>。遺伝子発現データの分析からは、microRNA の制御とタンパク質の相互作用間に関係がある事が報告されている<sup>15)</sup>。本研究の機能クラスと頂点の分布についての分析結果は、上記した最近の研究結果が示唆している仮説「non coding RNA が制御機能を持つため、高等生物になるに従いタンパク質をコードしていない non coding DNA の割合が高くなる」を支持し、non coding 分子の相互作用に対する新たな考えを支援する構造と機能の関係を提供する。

謝辞 本研究は公立ほこだて未来大学特別研究費重要区分（平成 22 年度）の助成をうけたものである。

15) Liang, H. and Li, W.-H.:MicroRNA regulation of human protein-protein interaction network,*RNA* , Vol.13, pp. 1402-1408 (2007).

### 参 考 文 献

- 1) Mattick, J.S.:RNA regulation: a new genetics, *Nature Review Genetics*, Vol.5, pp. 316-323 (2004).
- 2) Amaral, P.P., Dinger, M.E., Mercer, T.R. and Mattick, J.S.:The eukaryotic genome as an RNA machine, *Science*, Vol.319, pp. 1787-1789 (2008).
- 3) Mattick, J.S. and Makunin, I.V.:Small regulatory RNAs in mammals, *Human Molecular Genetics*, Vol.14, pp. R121-R132 (2005).
- 4) Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., et al.:Transcriptional regulatory networks in *Saccharomyces cerevisiae*, *Science*, Vol.298, pp. 799-804 (2002).
- 5) Wu, T.,Wang, J., Liu, C., Zhang, Y. and Shi, B.:NPInter: the noncoding RNAs and protein related biomacromolecules interaction database,*Nucleic Acids Research*, Vol.34, pp. D150-D152 (2006).
- 6) Park, J. and Barabási, A.-L.:Distribution of node characteristics in complex networks, *Proc. Natl. Acad. Sci. USA*, Vol.99, pp. 17916-17920 (2007).
- 7) Wasserman, S. and Faust, K.:*Social network analysis*, Cambridge University Press: New York (2007).
- 8) Luscombe, N.M., Babu, M.M., Yu, H., Snyder, M., Teichmann, S.A. et al.:Genomic Analysis of regulatory network dynamic reveals large topological changes,*Nature* , Vol.431, pp. 308-312 (2004).
- 9) Changning, L., Baoyan, B., Geir, S., Lun, C. et al.:NONCODE: an integrated knowledge database of non-coding RNAs, *Nucleic Acids Research*, Vol.33, pp. D112-D115 (2005).
- 10) MIPS Classification: <http://mips.helmholtz-muenchen.de/funecatDB/>
- 11) Alba, R.D.:Graph-theoretic definition of a sociometric clique, *J. Math. Soc.*, Vol.3, pp. 113-126 (1973).
- 12) Bonabeau, E., Dorigo, M., Theraulaz, G.:*Swarm Intelligence: From natural to artificial systems*, Oxford university Press: New York (1999).
- 13) Fiduccia, C.M. and Mattheyses, R.M.:A linear-time heuristic for improving network partitions, Proc. of the 19th Conference on Design Automation (IEEE, Los Alamitos, CA), pp. 171-181, doi:0420-0098/82 (1982).
- 14) Shimoni, Y., Friedlander, G., Hetzroni, G., Niv, G. et al.:Regulation of gene expression by small non-coding RNAs: a quantitative view, *Molecular Systems Biology*, Vol.3, pp.1-9 (2007)