

non-coding RNA-タンパク質相互作用ネットワークとその制御性の特徴分析

香々見春菜^{†1} 阿久津達也^{†2} ナチエル・ホセ^{†1}

ネットワーク制御性は複雑ネットワークの分野で最近研究が開始され、またネットワークの中でドライバー・ノード(あるいはコントローラー)を決定するための様々なアプローチが提案された。しかし、ほとんどのアプローチは一種類の頂点からなるネットワークを対象としていた。一方、最近の研究において、いくつかのRNA分子はタンパク質(非コードのRNA)に翻訳されないが、それらが細胞の重要な生物学的機能を持ち、また疾患に関連する可能性があることが示された。この複雑な生物システムを明確にするために、我々は、non-coding RNA-タンパク質相互作用の二部ネットワークの構造と制御性の特徴を解析した。

Analysis of non-coding RNA-protein interaction network and its controllability features

HARUNA KAGAMI^{†1} TATSUYA AKUTSU^{†2} JOSE C. NACHER^{†1}

Network controllability has recently emerged in complex network field and various approaches to determine driver nodes (or controllers) in networks have been proposed. However, most of the approaches have focused on unipartite networks. On the other hand, there are growing evidences showing that although some RNA molecules are not translated into a protein (non-coding RNA), they can still play key biological functions in a cell and are potentially linked to diseases. To shed light on this complex biological system, we investigate structure and controllability features of the bipartite non-coding RNA-protein interaction network.

1. はじめに

最近の研究により、哺乳類など高等生物のゲノム中のコード化情報の多くは、タンパク質に翻訳されない RNA (ncRNA) へ転写されることが判明してきた。さらに、このことに対して原核生物および真核生物の間の不均衡な複雑さスケールによる説明がなされている[1]。興味深いことに、これらのRNA分子はタンパク質に翻訳されないが、細胞において重要な生物学的機能を持つ。リボソーム RNA や転移リボ核酸はタンパク質合成と関係する機能を持つが、小さなRNA分子(snoRNAs, miRNAs および siRNAs)は、タンパク質発現量を変化させる機能を持つことが示唆されている。さらに、ウイルス感染と腫瘍形成なども含め、ncRNA と複雑な疾患との関連性が多数、報告されている。また、miRNAs の変化および調節異常が、様々な疾患に帰着することを示唆する結果も得られつつある。これらの生物学的機能および疾患関連と関係する実験結果などから、高等生物中の遺伝情報の発現機構には多くの ncRNAs からなる未開拓の層があることが示唆されている[2]。

一方、近年、ネットワーク科学において、大規模な有向ネットワークや、無向ネットワーク中の可制御性を解析するための様々な方法が開発されてきた[3][4][5]。それらの研究は一種類の頂点のみからなるネットワーク(一部ネットワーク)を対象としていたが、最近、二部ネットワーク構造の可制御性に関して、グラフ理論における最小支配集合(Minimum Dominating Set (MDS))に基づいた理論および解析手法が開発された[6]。この MDS のアプローチの使用に

より、ネットワーク全体を制御するのに必要なコントローラーの最小数を決定することが可能である。このように理論展開は急速に進歩したが、ここでは実際の生物学におけるネットワーク中のコントローラーの特徴づけに関する研究はほとんどなされていなかった。Wuchty は、タンパク質相互作用(PPI)ネットワーク(一部ネットワーク)中の可制御性を研究し、MDS 中の頂点の多くは、重要な遺伝子、癌関連の遺伝子およびウイルスにターゲットとされた遺伝子であることを発見した[7]。本研究では、この複雑な生物システムについて理解を深めるために、二部グラフ構造を持つ ncRNA-タンパク質相互作用ネットワークの構造と可制御性の特徴について解析を行った。

2. 研究方法

(1) データセット

ncRNA とタンパク質間の相互作用は NPInter データベースから検索することにより取得した。2005年にリリースされた NPInter の最初のバージョンでは、6種類のモデル生物からのわずか 700 個の相互作用しか含まれていなかったが、最新の NPInter v2.0 では 18 種類の生物種からの ncRNA と他の分子との間の 200,000 を超える相互作用が含まれている。したがって、現在のバージョンを利用することにより、より詳細なネットワーク分析が可能となる。そこで、この最近の文献および関連するデータベースからのデータを蓄積する NPInter v2.0 を使用し、ヒトを対象に ncRNA とタンパク質との相互作用を抽出した結果、2,453 個の ncRNA、4,872 個のタンパク質、および、35,816 個の相互作用データを得た。これらの分子とそれらの相互作用は、二部グラフ構造を持つ相互作用ネットワークとして表現することができる。

^{†1} 東邦大学
Toho University
^{†2} 京都大学
Kyoto University

(2) MDS の計算方法

二部グラフ $G(V_T, V_B; E)$ は、2種類の頂点集合 V_T と V_B 、および、それらの間の辺集合 $E \subseteq V_T \times V_B$ から構成される。我々の問題では、 V_T の要素は ncRNA 分子、 V_B の要素はタンパク質に対応する (図 1 参照)。このような二部グラフ構造を持つ ncRNA-タンパク質相互作用ネットワークの構造を次数分布および累積分布を利用して分析した。なお、このネットワークにおいてはすべての辺は、ncRNA からタンパク質へという方向に向きづけられている。さらに、ncRNA-タンパク質の二部からなるネットワークの可制御性に関する特徴も分析した。ncRNA コントローラーの最小数は MDS の計算することにより得ることができる (図 1 参照)。二部グラフ構造を持つネットワークにおいては、MDS の頂点がすべて V_T に属するという制約のもとで、MDS は最小セットカバーと等価となる。MDS の計算、および、最小セットカバーの計算は NP 困難な問題である [8][9][10]。しかしながら、二部ネットワーク中における MDS の最適解は、 10^5 個までの頂点からなる大規模なネットワークに対してもわずかな数秒で計算することができる [6]。本研究では既存のアルゴリズムを使用する代わりに、二部グラフにおける MDS 計算問題を整数計画問題 (ILP) として以下のように定式化することにより最適解を計算した:

$$\begin{aligned} & \text{minimize} && \sum_{v \in V_T} z_v \\ & \text{subject to} && \sum_{\{v,u\} \in E} z_v \geq 1 \quad \forall u \in V_B \\ & && z_v \in \{0,1\} \quad \forall v \in V_T \end{aligned} \quad (\text{Eq. 1})$$

なお、この定式化において、MDS は $\{z | z_v = 1\}$ より得られる。

3. 結果

2 節で述べた手法を用いて二部グラフ構造を持つ ncRNA-タンパク質相互作用ネットワーク中の MDS を計算することにより、ネットワーク全体を制御するのに必要なコントローラーの最小数を求めた。本研究で用いた ncRNA の総数は 2,453 個であり、その中で MDS に含まれる ncRNA の個数、つまり、4,872 個のタンパク質を制御するために必要な ncRNA の個数は 469 個であった。それは ncRNA 全体の 19.1% を占めるだけである。この結果は、すべてのプロテオームを制御する際に、ncRNA 全体の 5 分の 1 だけが重大な役割を果たすことを示している。MDS の比率が約 19% という結果は、タンパク質ネットワーク中の制御性に関する従来の研究と同様の値となっており、注目に値する [5][7]。

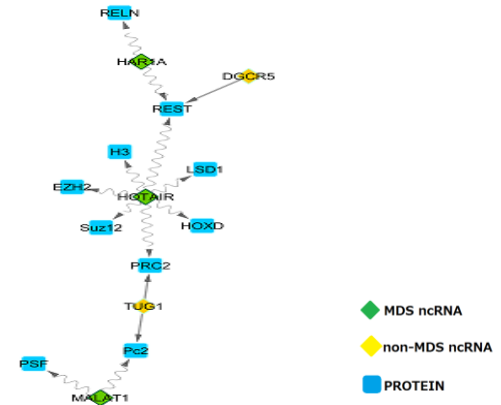


図 1 ncRNA-タンパク質相互作用ネットワークの一部分

Figure 1. A subgraph of ncRNA-protein interaction network.

4. 今後の課題

ncRNA ネットワークの位相的な特徴と疾患情報を組み合わせることで解析することにより、ncRNA と疾患の統計関連性を見出すことが可能であると考えられる。そのため、現在、本研究について提案した手法の拡張について様々な検討を行っている。

参考文献

- 1) Frith, M.C., Pheasant, M. and Mattick, J.S.: The amazing complexity of the human transcriptome. *Eur. J. Hum. Genet.*, **13**, pp. 894–897 (2005).
- 2) Amaral, P.P., Dinger, M.E., Mercer, T.R., Mattick J.S.: The eukaryotic genome as an RNA machine. *Science* **319**, pp.1787–9 (2008).
- 3) Nacher, J. C. and Akutsu, T.: Analysis of critical and redundant nodes in controlling directed and undirected complex networks using dominating sets. *Journal of Complex Networks*, in press (2014).
- 4) Liu, Y.-Y., Slotine, J.-J. & Barabási, A.-L.: Controllability of complex networks. *Nature* **473**, pp.167–173 (2011).
- 5) Nacher, J. C. and Akutsu, T.: Dominating scale-free networks with variable scaling exponent: heterogeneous networks are not difficult to control. *New Journal of Physics* **14**, 073005 (2012).
- 6) Nacher J.C. and Akutsu T.: Structural controllability of unidirectional bipartite networks, *Scientific Reports* **3**, 1647 doi:10.1038/srep01647 (2013).
- 7) Wuchty, S.: Controllability in protein interaction networks. *Proc. Natl. Acad. Sci.* **111**, pp.7156–7160 (2014).
- 8) Chvátal, V.: A greedy heuristic for the set-covering problem. *Mathematics of Operations Research* **4**, pp.233–235 (1979).
- 9) Lovász, L.: On the ratio of optimal integral and fractional covers. *Discrete Mathematics* **13**, pp.383–390 (1975).
- 10) Johnson, D. S.: Approximation algorithms for combinatorial problems. *Journal of Computer and System Sciences* **9**, pp.256–278 (1974).