

ネットワーク分析によるタンパク質相互作用データベースからの知識発見

Knowledge Discovery from Protein-Protein-Interaction Data through Network Analysis

鈴木 準二^{*1}
Junji Suzuki

松山 科子^{*2}
Shinako Matsuyama

白石 哲也^{*3}
Tetsuya Shirashi

寺野 隆雄^{*1}
Takao Terano

^{*1} 東京工業大学
Tokyo Institute of Technology

^{*2} ソニー株式会社
Sony Corporation

^{*3} 株式会社 ソニーコンピュータサイエンス研究所
Sony Computer Science Laboratories, Inc.

“Molecular switches” are molecules related to the signal transduction which switch on / off to the cell function modules. Analyzing the molecular switches is expected to contribute the presumption of the carcinogenesis process and discovery the target of treatments. In this paper we have extracted the “molecular switches” from the protein-protein-interaction network and analyzed the characteristics of them.

1. はじめに

従来の癌治療方法は、網羅的解析による癌の理解と単一分子に着目することが主流であった。しかし、この手法には限界があり、システムの捉え方で研究を進める必要がある([白石06])。著者らは生体を複雑ネットワークであるシステムとしてとらえ、そのネットワークの特性を計測することで生命システムを解析し、癌の早期発見や治療への貢献を目指している。

本論文では、知識発見の適用の観点から、タンパク質相互作用データからネットワークを生成し、分子スイッチ(細胞機能モジュールに対して情報の流れをスイッチ制御するタンパク質)の抽出を行いその特徴を解析する。

2. タンパク質相互作用ネットワークの生成

本章では、タンパク質相互作用ネットワークの生成手順を述べる。

2.1 データ

まず、Ariadne Genomics 社の PathwayStudio ResNet データベース(version 3.0)からタンパク質間相互作用の中で発現制御や機能調整に関わるリンク情報を取得した。PathwayStudio ResNet データベースは、膨大な科学文献からタンパク質相互作用情報を自然言語処理技術により抽出したものであり、50 万以上の相互作用データが存在する。本データベースを用いた理由は、医学・バイオ関連の論文データベースとしてタンパク質の相互関係データに関し最もデータ数が多いためである。

本データベースから、リンクの特性として、機能を促進・作用を増強するもの、抑制・減少するものの種別が明記されているもの、さらにデータの元となる掲載された論文数が3以上のものを解析対象とした。結果、ノード(タンパク質)数 3682、エッジ(タンパク質間の相互作用)数 19,178 のネットワークが生成された。エッジの属性として機能を促進・作用を増強するものを Positive、その逆を Negative とした。

2.2 ネットワークの特徴

2.1で生成したネットワークに対し、ネットワークの特徴を示す主要な指標である度数分布とクラスタリング係数を計測した。度数分布はノード毎のリンク数の分布、クラスタリング係数は密度を測る指標であり、ノードが持つエッジ群が互いに接続されてい

る確率である。

図 1に度数分布を示す。X軸が度数 k 、Y軸は k 以上の度数を持つユーザ数の両対数表示である。度数分布はランダムよりは冪則に近い形態を持つ。

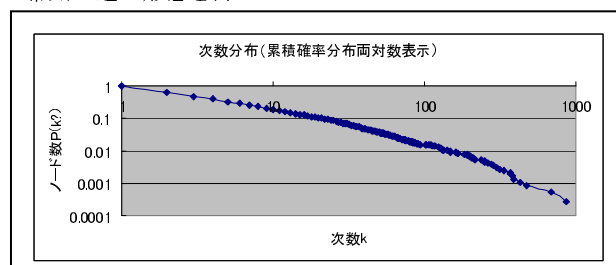


図 1 タンパク質相互作用ネットワークの性質

次に対象ネットワークのクラスタリング係数を、対象ネットワークの平均度数(10)を持つ正則ランダムグラフでの値と比較したところ約 100 倍であった(それぞれ 0.003, 0.25)。また、2 ノード間の平均最短パス長はそれぞれ、5.37, 3.41 であった。正則ランダムグラフでの値と比べた際に最短パス長に差がなく、クラスタリング係数が大きい場合、そのネットワークは「スモールワールドネットワーク」と定義される[Watts 98]。この値より、本ネットワークはスモールワールドであると言える。

3. 分子スイッチの検出

本章では、分子スイッチの検出について述べる。

3.1 分子スイッチの定義

まず、既存研究[Gardner 2000]で定義されている大腸菌におけるスイッチパターンを type1 とした。次に type1 を拡張し、他ノード(タンパク質)との相互作用リンクをも考慮したものを type2 とした。続いて、[Kobayashi 03]で述べられている、安定均衡のスイッチネットワークの要件であるポジティブフィードバックのうち、1ノードを介したリンクを対象とし、type3 とした。著者が定義した分子スイッチパターンを図 2に示す。塗りつぶした丸印はスイッチ候補のノード組、ドット印は関連ノードである。また、→線は Positive リンク、-リンクは Negative リンクである。Type2 のパターンは、スイッチ候補ノードと関連ノード間のリンクが双方向 Positive、双方向 Negative 全パターンがありうるため記載パターンを含めて 3 パターン存在する。Type3 も同様に考え記載

パターンを含めて6パターン存在する.

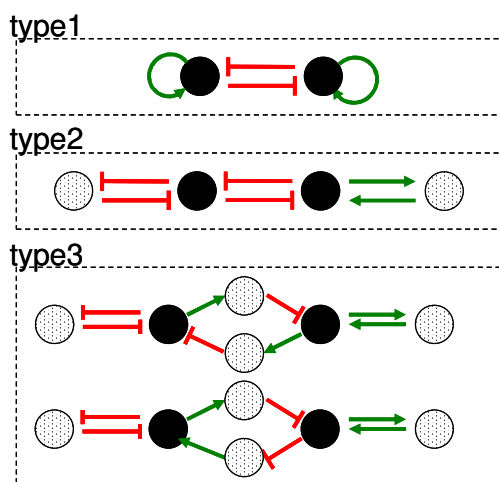


図2 分子スイッチの定義

3.2 分子スイッチの検出

生成したネットワークから、図2で定義した部分構造を抽出した。結果、Type1が40、Type2が49、Type3が2414パターン検出された。検出されたパターンの該当ノード(タンパク質)を「スイッチノード」と呼び、対象2ノードを「スイッチペア」と呼ぶ。スイッチノードは、439個であった。図3に検出したスイッチの一例を示す。このスイッチは良く知られたもので、FGF2とPTK2Bの発現量で情報流通の向きが入れ替わる。例えば、PTK2Bの相対的発現量がFGF2よりも多い場合、apoptosisが抑制され、angiogenesisが促進されることで細胞の異常増殖につながる。

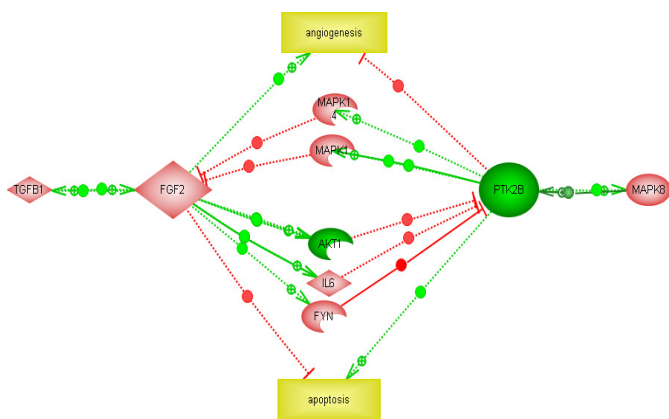


図3 検出されたスイッチの例

4. 分子スイッチのネットワーク内での特徴

本章では、検出された439のスイッチノードのネットワーク内での特性を見る。

全ノードに対する次数(リンク数)、クラスタリング係数、および他のノードを仲介している割合を示す媒介中心性を計算し、全ノードの平均値とスイッチノード間での平均値を比較した。表1に数値を掲載する。全ての項目に対しスイッチノード群は値が高い。従ってこれらのノードはネットワークの中心であると言える。

表1 スイッチノードのネットワーク特性

	全ノード平均	スイッチノード平均
ノード数	3,682	439
次数	10.15	45.59
クラスタリング係数	0.25	0.41
媒介中心性	0.0002	0.0012

5. 考察

既存研究にて、タンパク質相互作用ネットワークは次数分布が冪則に従うスケールフリー性を持ち、ハブになるタンパク質は生存に重要な機能を持つことが多いとされている([五斗 05])。本論文で、今回抽出したネットワークでも同一傾向が判明した。

分子スイッチはその定義から次数は最多で4持つことが条件となるが、実際のスイッチノードの平均は約46であり、全ノードの平均の約4.5倍持つ。これは、検出されたスイッチノードのほとんどが複数相手とのスイッチペアを持つ(スイッチノードの平均スイッチ個数は223、88%のスイッチノードが2つ以上のスイッチを持つ)ことが影響している。媒介中心性やクラスタリング係数が高いのも一つは次数が高い事が影響しているためであるが、その他の要因として、スイッチノードは全て3モチーフのノード(2タンパク質間と相互の接続: 三角形の関係を持つ)にもなっているためと推測される。

6. まとめ

本研究では、構造的な特徴解析技術を用いてタンパク質相互作用ネットワークから実際にいくつかの分子スイッチを検出した。また、分子スイッチのタンパク質群はネットワーク内で次数、密度、媒介性の観点からネットワークの中心になっていることを示した。

生体内では、環境や時間条件により、今回検出したスイッチペアのうちいくつかは分子スイッチとして働くと考えられる。条件毎に機能するスイッチの詳細解析を行うことが課題である。またモチーフも生体ネットワークにおいては重要なタンパク質構造として着目されており、スイッチとモチーフの関連も分析したい。

本研究の接近法は、ネットワーク分析手法を適用してデータベースからの知識発見を行うものであり、生物学的な知見は利用していない。それにもかかわらず既存の分子スイッチならびにその候補が数多く見出されたことは興味深い。本アプローチにより、生化学的な実験の妥当性・冗長性・不備などを探る客観的な手がかりも得られるものと期待される。

参考文献

[Gardner 2000] Gardner, T.S., Cantor, R.C., and Collins, J.J.: Construction of a genetic toggle switch in Escherichia coli, Nature, Vol 403, pp. 339-342, 2000.
 [Kobayashi 03] Kobayashi, T., Chen, L. and Aihara, K.: Modeling Genetic Switches with Positive Feedback Loops, J. theor. Biol, 221, pp.379-399, 2003.
 [Watts 98] Watts, D.J. and Strogatz, S.H.: Collective dynamics of 'small-world' networks, Nature, Vol. 393, No.4, pp.440-442, 1998.
 [五斗 05] 五斗進: 生体分子の相互作用ネットワークにおけるスケールフリー性, 人工知能学会誌 20 巻 3 号, pp.280-283, 2005.
 [白石 06] 白石哲也: 癌のシステム生物学とグリオーマ, グリオーマ: 病態と治療, pp.77-92, シュプリンガー・フェアラーク東京, 2006.