タンパク質相互作用ネットワークからの密結合モジュールの列挙法

津田　宏治
RNA情報工学チーム
研究員

タンパク質が細胞内において単独で機能をもつことは比較的稀であり、多くの場合、他のタンパク質と複合体を構成して機能しています。機能が既知の複合体に関しては、MIPSなどにデータベースとしてまとめられていますが、とくにヒトに関しては未だ完全とはいえず、未知の複合体が存在することに疑いはありません。また、複合体によっては、組織特異的に、新たなタンパク質が加わったり、除かれたりして組成が変わり、その機能を変化させることが知られています。

タンパク質の物理的相互作用を表すネットワークが与えられたとき、複合体は、そのなかの密結合モジュールに対応していると考えられます。密結合モジュールとは、エッジによって密に結ばれているノードの部分集合のことを指します。複合体であっても、すべてのタンパク質が他のすべてに結合するわけではないため、必ずしも完全グラフにはなりません。ネットワークから複合体を発見するため、さまざまなクラスタリング法が提案されていますが、多くの場合、それらはネットワークの分割に基づいています（図１A左）。このような分割法では、ひとつのタンパク質が複数の複合体に割りつけられることはありません。しかし、実際の複合体を見てみると、多くの場合、ひとつのタンパク質が複数の複合体に属することが知られています。分割法ではこのような場合を正しく扱うことができません。本研究では、エッジ密度が、ある閾値以上のモジュールを、重なりを許してすべて列挙するというアプローチをとっています（図１A右）。これにより、複合体の組成の可能性をすべて数え上げることが可能になり、分割法に比べて、とりこぼしを少なくすることができます。

エッジ密度が高い複合体であっても必ずしも機能を持つとは限りません。そこで本研究では、遺伝子発現や進化系統樹のプロファイルデータも考慮して、より機能を持つ可能性の高いモジュールを発見することができます（図１B）。たとえば、各組織における発現プロファイルが与えられた際には、k 個以上の組織ですべてのタンパク質が発現しているモジュールだけを列挙するというような操作ができます。

論文[1]に示した通り、我々の列挙法(Dense Module Enumeration)は、これまでの手法(Markov Clustering, Clique Percolation Method)よりも、MIPSに登録されている複合体を精度よく予測することができました。また、SCF E3 Ubiquitin Ligase Complexの組織特異的な再構成の様子もとらえることができました（図２）。この図では、５個のモジュールと、各々が発現している組織のリストが表わされています。中央の三つのタンパク質は共通で、各モジュールは組織によって、それぞれ一つの追加タンパク質を持っています。Q86VP6以外は、F-boxタンパク質であり、SCF複合体がユビキチンを付加する対象を認識する機能を持っています。この結果からは、SCF複合体が組織によってユビキチンを付加する対象を選択する様子がよくわかります。

本手法は、ネットワークとプロファイルさえあれば、様々な目的に利用可能です。今後CBRCで行われているプロジェクト等に生かされれば素晴らしいことであると思っています。

References

[1]E. Georgii, S. Dietmann, T. Uno, P. Pagel and K. Tsuda: Enumeration of condition-dependent dense modules in protein interaction networks, Bioinformatics, 25:7, pages 933-940, 2009.

図１

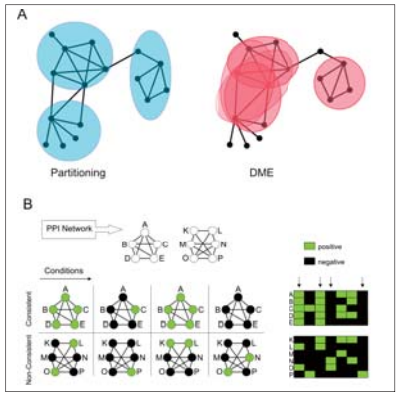
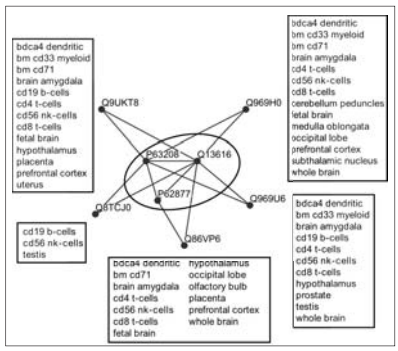
図２