

活性化情報に基づくパスウェイ横断ネットワークの推定

●瀬々 潤

お茶の水女子大学 大学院人間文化創成科学研究科

<研究の目的と進め方>

パスウェイ情報の蓄積が進んでいるが、現在データベースに登録されているネットワークは生体全パスウェイに比べ極一部であると考えられる。パスウェイの理解を深めることは、生体のシステムの挙動の理解及び創薬ターゲットの発見において重要な役割を果たす。この目的のため新規パスウェイの予測が行われてきたが、生体から観測されたデータは揺らぎやノイズが大きいため、成功を見てはいない。

本研究では、既知のパスウェイ間には協調して働く関係にあるパスウェイがあるであろうという既知のパスウェイから見られる仮説から、新規ネットワークの予測を行う。特に(1)既知のパスウェイやたんぱく質相互作用のネットワークとマイクロアレイ等で得られた遺伝子発現量の情報を照らし合わせることで、どのネットワークがいつ利用されているかのネットワーク活性化情報を抽出し、(2)得られた活性化情報を基に、接続されるべきだが、現在はネットワークが描かれていないパスウェイを求める。

本研究により、パスウェイや相互作用ネットワークのどの部分がいつ働いているかの注釈付けが行え、現在のパスウェイ情報を超えて、副作用が起きているネットワークの推定が期待できる。

<2008年度の研究の当初計画>

本年度は主に、既知のパスウェイやたんぱく質相互作用ネットワーク情報と、遺伝子発現量の情報を照らし合わせることで、どのネットワーク(サブネットワーク)が、どのような環境下で活性化しているかの注釈付けを行う。特に、次の3点を行う(1)ネットワークの注釈付けに必要なアルゴリズムの構築、(2)発現量データの収集と選別、(3)遺伝子発現やネットワークは、ほ乳類などの多細胞で高等な生物において、複雑な調整が成されていると推定できるため、本研究では酵母などの単細胞生物から研究をスタートし、より高等な生物への適用をする。

<2008年度の成果>

2008年度計画に沿い、活性化ネットワーク同定方法の開発及び、酵母データへの適用を行った。

活性化ネットワークの同定方法として、(1)様々な条件下で観測した遺伝子の発現量情報を有意に発現変化した条件集合へと変換した後、(2)ネットワーク中で連結しているかつ、発現変動した条件を共有するネットワークを同定する手法 COmmon Patteern Itemset NEtetwork (COPINE)を開発した。発現が変化した条件が遺伝子毎にランダム

に現れると仮定した場合、ネットワークで連結している2遺伝子で発現変化条件を共有している確率は低い。更に、より大きなネットワーク中全ての遺伝子で条件が共有されている確率は低くなる。よって、一定以上の大きく、条件を共有するネットワークを抽出することで、生物学的に活性化している可能性の高いネットワークの抽出が可能となる。また、ネットワークを探索する際、ネットワークの大きさが大きくなるに従い爆発的に探索すべきネットワーク数が大きくなる。近年では、KEGGやReactome.orgで公開されているパスウェイ、相互作用情報は50万以上の相互作用を有しており、これらの解析には上記で示した条件共有ネットワークを効率よく、高速に探索できる必要がある。我々は、このようなネットワークの探索を高速に行う手法を開発し、100万以上の相互作用、100以上の実験条件がある環境下でも、解を得られるアルゴリズムを開発した。

更に、このアルゴリズムを酵母の相互作用ネットワーク(Ito et al. 2000, Uetz et al. 2000, Nevan et al. 2006)と様々なストレス環境下における遺伝子発現量情報(Gasch et al, 2000)を組み合わせることで、解析を行った。この結果を図1に示す。図1には、大きなネットワーク、つまり、より同時に活性化している可能性の高いネットワーク上位5つを示した。特筆すべき点は3点有る。1点目は環境特異的に活性化しているネットワークが自動的に発見できていた点である。図の中で(A)はジアミド条件下で、(B)、(E)は過酸化水素過剰な条件下で、(C)は窒素枯渇の条件下で、(D)はヒートショック条件下で活性化しているネットワークであり、膨

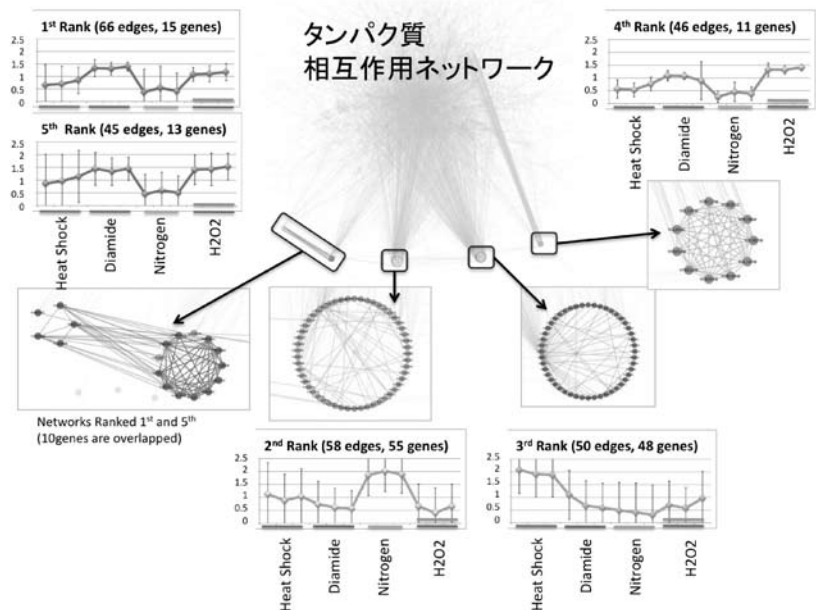


図1. タンパク質相互作用と遺伝子発現量の複合解析例。条件特異的に働く相互作用ネットワークが抽出できている。

大な相互作用ネットワーク、多様な条件から、条件特異的に働いているネットワークとその条件が自動的に抽出できることが示唆される。第2点目は既知の情報と強く関連したネットワークが抽出されていることである。(A), (B), (E)はプロテアソーム複合体の一部であり、プロテアソーム複合体は20Sと19Sの2つの大きなサブユニットを持つが、(A), (B)は20Sの一部、(E)は19Sの一部である。(A)と(B)は共通する遺伝子が多いが、(A)に含まれ(B)に含まれない遺伝子には、環境が変化すると局在箇所が変化するタンパク質が含まれており、本手法により環境によるタンパク質の特性の変化を調査できる可能性が示唆されている。(C), (D)はそれぞれオートファジー関連、サイトスケルトンの立体構造変化に関連した遺伝子群が多く含まれている。第3点目の特筆すべき点は、疎な結合のネットワークの機能を明らかに出来る可能性である。今までグラフを用いた注釈には主に密な結合を持つ部分に特化しているものが多かった。一方、パスウェイは結合が疎であっても重要な情報伝達を担う遺伝子が存在しており、このようなネットワークは困難であった。我々のアルゴリズムで抽出されているネットワークには、(A), (B), (E)の様に密なネットワークも含まれるが、(D)の様にハブを有しているネットワークも抽出可能となっている。更に、一般にハブのネットワークは機能的に重要であることが叫ばれつつ、どのような環境で、どのような機能を果たしているか細かい機能が議論されることが少なかった。本実験結果により、ハブの遺伝子がどのような環境で、どの遺伝子群と協調しているかが明らかになり、遺伝子機能の理解が進む物と考えられる。

また、上記アルゴリズムは既知のネットワークと実験条件の組み合わせを詳細に把握する事に適しているが、パスウェイを横断するネットワークの検出には、既知のネットワークでは捕えられていない、より大域的なネットワークの検出も必要である。しかし、遺伝子発現量を含む生物学データは観測ノイズを多く含むだけでなく、全遺伝子が一齐に変化するのではなく、一部の遺伝子の状態が変化することで適用する事が多い。よって、弱い相関を発見する、逆に言うと無相関であることを発見する技術が必要であった。この問題に対し、密度比を利用することで、事前に分布を与える必要が無く、モデル選択可能な相互情報量推定手法を開発した。また、本手法を上記同様ストレス環境下における酵母発現量情報に適用することで、ストレス刺激に対しての酵母の大

域的な応答を把握する事が可能となったことを確認した。図2は、その相関を表した物で、DNAやRNAにダメージを起こすストレスは、細胞活動全体に刺激を与え、ビタミンKによる刺激はタンパク質の局在にのみ影響を与えることが確認できた。

<国内外での成果の位置づけ>

本研究はデータマイニング分野において研究されている頻出グラフ列挙、グラフクラスタリング、制約付きクラスタリングの分野を跨ぐ研究成果である。頻出部分グラフでは、多数のグラフを含むデータベースから共通のグラフを発見するが、遺伝子の発現条件を扱うことができない。グラフクラスタリングでは密なネットワークを見つけることが出来るが、パスウェイの様に疎な結合でも重要なネットワークを把握する事ができない。制約付きクラスタリングでも、密なネットワークを要求するアルゴリズムが多いこと、また、一般に全ての実験条件で協調していないとクラスターが構成できない。よって、我々の手法は、他の手法では発見できず、かつ、生物学的な解釈が行いやすい結果を得られる点で優れている。

<達成できなかったこと、予想外の困難、その理由>

酵母は工業的にも有用なため、調査価値のある種であるが、酵母で利用可能な手法が、より高等生物で利用可能であるとは限らない。本年、この多細胞生物のデータ収集まで行ったが、解析結果まで出せなかった。これは、次年度の研究内容にもなるが、多細胞生物はネットワークを巧みに利用して活動しており、酵母に比べると非常に複雑な機構を解析する必要が生じるためである。当初から予想されていたことではあるが、解析を進めることで改めて複雑なネットワーク利用が散見され、解析が予想以上に困難な様子が見えてきている。

<今後の課題>

課題は2つ有る。1つは、特に、マウスやヒトは酵母の5倍程度の遺伝子数で多細胞生物として様々な器官形成を行っており、細胞内ネットワークを巧みに利用していることが考えられる。この複雑なネットワークを解析していくこと。もう1つは、パスウェイを跨いで働いているネットワークの解析手法を確立することである。本年度は連結したネットワークの発見に着目した。これによりプロテアソーム複合体の例に見られる様に、本来複合体として働く複合体が正しく抽出できることが確認できた。今後は、このように協調して働くネットワークを意図的に見つけ出すアルゴリズム開発を行い、現在は繋がっていないにも関わらず、協調して働く、つまり副作用を起こす可能性のあるネットワークの発見を行っていく。

<成果公表リスト>

- 0901120751
Mio Seki and Jun Sese. Identification of active biological networks and common expression conditions. 8th IEEE International Conference on Bioinformatics and BioEngineering (BIBE 2008), Athens, Greece, Oct. 8-10, 2008
- 0901120757
Taiji Suzuki, Masashi Sugiyama, Takafumi Kanamori, and Jun Sese.. Mutual information estimation reveals global associations between stimuli and biological process. BMC Bioinformatics 2009, 10 (Suppl 1):S52.

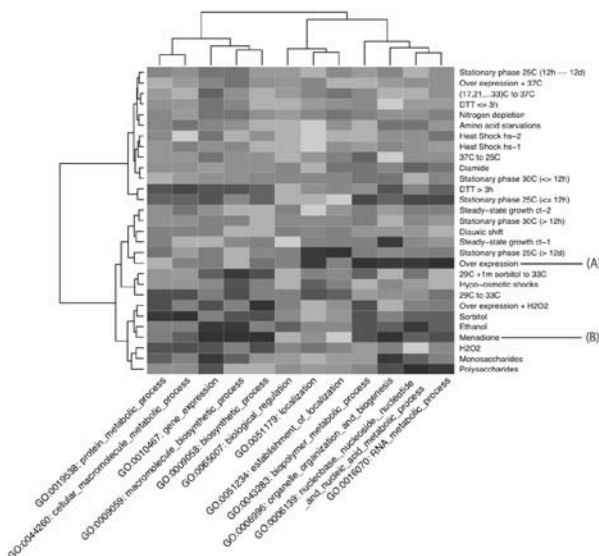


図 2. ストレス刺激と遺伝子機能の相関関係