

# 相互作用ネットワーク検索によるパスウェイ情報補完

関 美緒 (指導教員：瀬々 潤)

## 1 はじめに

現在までに様々な種のゲノム配列が読まれ、遺伝子の情報が調べられてきたが、細胞内にどのようなシグナル伝達経路が存在しているかは依然としてほとんど明らかになっていない。シグナルの入力、及び、出力となる現象は知られていても、入力から出力までの間にどのような遺伝子が関わっているかがわからない、ということがほとんどである。既存のパスウェイ [1] は、人手によって整理されており、生体内のパスウェイを網羅するには膨大な労力を要する。また、文献等の整理では未知のパスウェイを推定することは難しい。そこで、シグナルの入力、及び、出力から、あまり労力を要さず、そのシグナルのパスウェイを推定することを考えた。本研究では、たんぱく質-たんぱく質相互作用ネットワーク情報を利用することで、新たなパスウェイの発見を目指す。具体的には、膨大な相互作用ネットワークから、着目の遺伝子群を繋がるべくエッジが少ない部分グラフの抽出をすることで、新しいパスウェイの推定を行う。シグナルの入力、及び、出力が1遺伝子ずつとは限らないので、3遺伝子以上の着目遺伝子群にも対応できる手法を提案する。我々のアルゴリズムにDNA修復複製制御に関与する3遺伝子(RFC1, RAD17, RAD53)を入力として、これらを結ぶ部分グラフを抽出したところ、KEGG [1]に登録されているDNA傷害チェックポイントのパスウェイとは異なるものの、大きさが同一のネットワークを1秒以内に求めることができた。大きさが同一であることから、生体内で利用されている未発見のパスウェイの可能性はある。

## 2 関連研究

たんぱく質相互作用ネットワークから、パスウェイを推定する試みとして、着目する2つの遺伝子を結ぶ最短のパスをダイクストラ法によって見つける提案が為されている [2] が、パスウェイに対するシグナルの入力、及び、出力は複数考えられるので、3遺伝子以上でも発見する必要がある。3つ以上の与えられた点を全て通過する経路を求める問題としては、辺の重みの総和が最小となる最短経路をグラフから求める問題が知られているが、本研究で発見したいパスウェイは分岐構造も現れるため、単純な最短経路探索問題とは異なる。

## 3 研究内容

シグナルの入力となる遺伝子群、及び、出力となる遺伝子群から、それらを結ぶ経路を推定する方法として、たんぱく質相互作用ネットワーク(ノードが遺伝子、エッジが相互作用に相当)上で、入力、出力遺伝子群を結ぶ部分グラフの抽出が考えられる。手法の概要を図1に示した。また、エッジの重さは全て1で均一であると仮定し、エッジの数をグラフの大きさとする。また、エッジは相互作用を表しているの、求めた部分グラフのエッジ数が少ないグラフ程、実際にシ

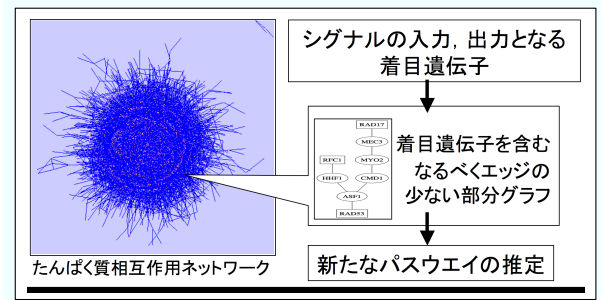


図1: 提案手法概要

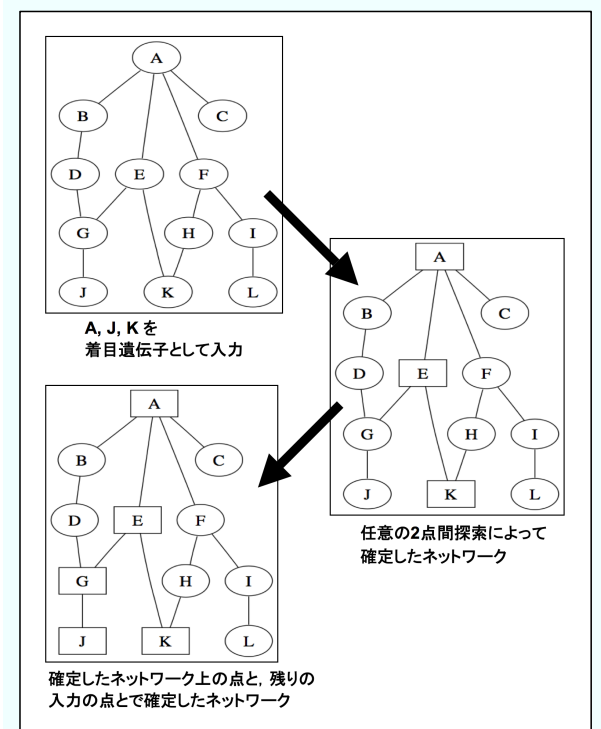


図2: 3点を入力した場合の探索

グナル伝達として利用されている可能性があると考えるのが妥当であろう。

しかし、一般に、与えられたグラフの部分グラフの内、指定した全てのノードを含む最も小さな部分グラフを見つけることは計算時間などの問題により容易ではない。本研究では、着目した遺伝子群を含みなるべくエッジの本数が少ない部分グラフを現実的な時間で発見する方法を提案する。具体的には、部分グラフの探索を次の様な手順で2点間の最短パスを求める問題に変換する。全てのエッジの重みが均一であると仮定しているの、2点間の最短パスの発見には幅優先探索を用いた。

### 手順

1. 与えられた相互作用ネットワークグラフを  $G$  とする。初期状態としてグラフ  $T$  を空のグラフとする。  $T$  を確定ネットワークと呼び、最終的に  $T$  に求めるグラフが入る。

2. 着目遺伝子群 (遺伝子数 2 以上)  $V$  の任意の 2 点間で最短パス  $P$  を求める.
3.  $T = P$  とする.
4.  $V \cap (G - T) = \phi$  となるまで 5, 6 を繰り返す.
5. ノード  $x \in T, y \in V \cap (G - T)$  のうち, 2 点間の距離が最短となる  $x, y$  及び最短パス  $P'$  を求める.
6.  $T = T \cup P'$  とする.

図 2 にアルゴリズムの動作例を示す. 四角のノードが確定ネットワーク  $T$  上のノードである.

### 相互作用ネットワークデータ

相互作用ネットワークとして Yeast two-hybrid(Y2H) データ (Ito らの Full data [3, 4] と Uetz ら [5]) 及び TAP データ (Krogan ら [6] のデータの内 Confidence Score が 0.8 以上のもの) を統合して使用した. ノード数 4,123, エッジ数 7,972 のネットワークである. また, これらの実験において bait,prey 間の関係と既知のシグナルが伝達される向きに有意な関係が見られなかったことから, ネットワークは無向グラフとして扱った.

## 4 実行結果

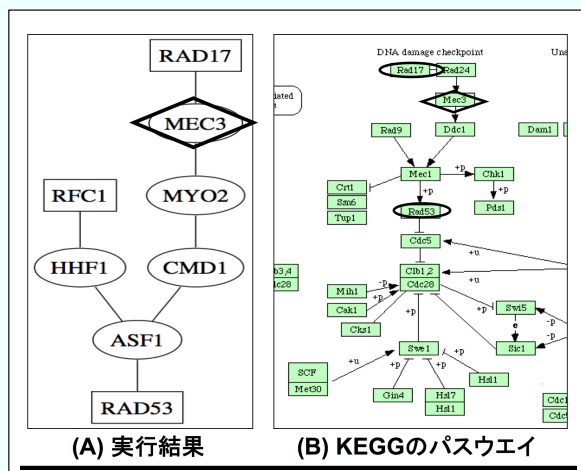


図 3: 実行結果と KEGG パスウェイの比較

実行結果の例を図 3(A) に示す. 長方形のノードが入力した着目遺伝子, 楕円のノードは探索により求められた部分グラフ上に現れた遺伝子である. 例では, DNA 修復複製制御に関連する RFC1, RAD17, RAD53 の 3 遺伝子を着目遺伝子として入力した. 結果では, RFC1 が関係する DNA 複製チェックポイントのシグナルと, RAD17 が関係する DNA 傷害チェックポイントのシグナルはともに合流し, RAD53 がさらに下流へ伝えていく予測となった. これは, 既知の知見とも一致する.

図 3(B) は, KEGG 上の DNA 傷害チェックポイントのパスウェイの一部である. 楕円で囲んだ遺伝子は, 上で述べた着目遺伝子 RAD17 と RAD53 である. ひし形で囲んだ遺伝子は図 3(A) のひし形で囲んだ遺伝子と対応する. 今回の実行によって得られた経路のうち右側の経路には KEGG 上のパスウェイとの類似が見られた. よって, 実行結果上に現れた, KEGG のパスウェイと一致していない他の遺伝子も DNA 傷害チェックポイントに関わっている可能性があり, 実行結果は KEGG とは一部の経路が異なる新たなパスウ

エイの可能性もある. また, 実行結果上の RAD17 と RAD53 間のパスウェイの長さは KEGG 上の同一遺伝子間の経路とほぼ等しく, 推定したパスウェイも十分存在する可能性がある.

例で示した 3 遺伝子間での探索に要した時間は, 平均 0.57 秒であり (1000 回実行の平均, CPU は Core 2 Duo, メモリ 2GB, 言語は Java, 遺伝子間の関係は PostgreSQL を用いデータベースに格納), 現実的な時間でエッジの少ない部分グラフを求めることができた.

しかしながら, Y2H, TAP のみのデータ上での探索であるため, 実際のシグナル伝達経路に多く存在するリン酸化や化合物などは含んでおらず, 着目遺伝子や着目シグナルによっては, 本手法では, とても大きなネットワークが求められてきてしまい, 新たなパスウェイであると推測するのは難しい結果となった.

## 5 まとめと今後の課題

本研究では, たんぱく質-たんぱく質相互作用ネットワーク情報を利用して, 膨大なネットワーク上から, 着目遺伝子を含みなるべくエッジの本数が少ない部分グラフを探索する, 3 遺伝子以上にも対応する手法を述べた. 本手法により, 着目シグナルに関与する, 存在する可能性のある新たなパスウェイを現実的な時間で発見できることがわかった.

今後の課題としては, 得られた実行結果が生物学的に有意であるか検証していくことと, 求めた部分グラフは最短のグラフではないので, なるべくエッジの少ないグラフ探索手法の改良を行う. また, 検索するネットワーク情報に, リン酸化ネットワークの実験結果 [7] などを含めることで, 実行結果として求めたネットワークやその探索自体の有意性を高め, パスウェイ情報の推定に, より有用なものとした.

## 参考文献

- [1] M. Kanehisa, and S. Goto. *KEGG: Kyoto Encyclopedia of Genes and Genomes*. Nucleic Acids Res, 28: 27-30, 2000.
- [2] H. Hermjakob, et al. *IntAct: an open source molecular interaction database*. Nucleic Acids Res, 32: D452-455, 2004.
- [3] T. Ito, et al. *Toward a protein-protein interaction map of the budding yeast: a comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins*. Proc. Natl. Acad. Sci. 97, 1143-1147, 2000.
- [4] T. Ito, et al. *A comprehensive two-hybrid analysis to explore the yeast protein interactome*. Proc. Natl. Acad. Sci. 98, 4569-4574, 2001.
- [5] P. Uetz, et al. *A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae*. Nature 403, 601-603, 2000.
- [6] N. J. Krogan, et al. *Global landscape of protein complexes in the yeast Saccharomyces cerevisiae*. Nature 440, 637-643, 2006.
- [7] J. Ptacek, et al. *Global analysis of protein phosphorylation in yeast*. Nature 438. 679-684, 2005.