

Title	グラフ構造に基づいた遺伝子相互作用の推定に関する研究
Author(s)	大谷, 俊朗
Citation	
Issue Date	2007-03
Type	Thesis or Dissertation
Text version	author
URL	<a href="http://hdl.handle.net/10119/3609">http://hdl.handle.net/10119/3609</a>
Rights	
Description	Supervisor:平石 邦彦, 情報科学研究科, 修士

# グラフ構造に基づいた遺伝子相互作用の推定に関する研究

大谷 俊朗 (510020)

北陸先端科学技術大学院大学 情報科学研究科

2007年2月8日

キーワード: 遺伝子相互作用, マイクロアレイ, バインディングサイト.

遺伝子とは DNA 中の機能的な役割を持つ領域を指し, その発現によって生体機能にかかわるタンパク質が生成される. タンパク質の中で, 自身をコードしていた遺伝子や他の遺伝子の転写を促進, あるいは抑制する機能を持つものを転写制御因子と呼ぶ. 各遺伝子から生成されたタンパク質により互いの発現を制御し合う関係を遺伝子相互作用と呼ぶ. 遺伝子間の相互作用を解明することで, 細胞内の様々な反応を一つのシステムとしてモデル化することが可能になり, 新薬の開発などに繋がると考えられる. したがって, 遺伝子相互作用の推定精度のさらなる向上が望まれている.

遺伝子間の相互関係推定に関する従来研究としては, 生物学的な性質を利用した方法と, ある種の数理モデルを応用した解析の 2 種類がある. 生物学的性質を用いた推定では, 転写制御因子は特異的な塩基配列を認識, 結合することで遺伝子の発現に影響を与えるという生物学的知見に基づき, 多重配列アラインメントを用いて依存関係が既知の遺伝子の転写制御領域の間に類似した塩基配列を発見する. そして依存関係が未知の遺伝子群の中から類似した塩基配列を持つ遺伝子を探索する方法である. これに対し, マイクロアレイ実験により得られた遺伝子発現変化のパターンの情報を用いて, 遺伝子間の相互作用や調節機構を解釈するための遺伝子発現パターンのモデルを構築しようとする試みが, 数理モデルを応用した方法である. 代表的なものに, 遺伝子の発現を 2 値化して遺伝子間の依存関係を論理関数として表現するブーリアンネットワークを用いた方法や, 遺伝子の発現を静的で確率的な事象としてモデル化し, 確率変数間の依存関係を非循環な有効グラフで表現するベイジアンネットワークによる推定がある. しかし, ブーリアンネットワークによる推定は, 測定が困難な発現状況の時系列データが必要であるうえ, 推定する遺伝子数に対し多数の DNA マイクロアレイデータを必要とする. また, ベイジアンネットワークは現在この分野において広く用いられている手法であるが, 同じデータを説明可能な複数のモデルが存在するなどの問題がある. これらの手法はマイクロアレイデータの精度に強く依存しているが, 現在の実験技術ではマイクロアレイデータには 30 ~ 40 % 程度の誤差が存在しており, マイクロアレイデータのみから正確なモデルを構築することは難しいと考えられる.

本研究では、平石・土居らの提案による DNA マイクロアレイと DNA 配列比較の組合せによる遺伝子間相互作用の推定方法の改良を行う。既存手法の推定手順は、まず遺伝子  $a$  と比較して発現強度の相関の高い上位遺伝子を抽出し、転写制御領域に  $a$  と対象遺伝子間で全ての位置のウインドウ（長さ  $l$  の塩基配列）間の類似度を計算し、 $a$  の各位置に対する最大ウインドウ類似度を計算する。最大ウインドウ類似度が閾値以上であり、かつ、極大値を持つような位置（ピーク位置）を求める。ピーク位置がある程度連続した領域があれば、その領域と高いウインドウ類似度を持つ遺伝子群は同一の転写制御因子の被制御遺伝子の可能性が高いと推定する。マイクロアレイにより得られた遺伝子発現傾向の相関の高さという統計的性質と、転写制御領域における塩基配列の類似性という生物学的性質の異なる 2 つの性質を用いることで、一つの情報に依存しすぎず、また、転写制御因子の結合サイトを同時に推定可能である点を特徴としている。しかしながら、既存手法では以下の点が未解決である。1. 既知の転写制御因子の結合サイトがピーク位置に選択されない場合がある。2. 転写制御因子の結合サイトを特定可能なほど十分に絞り込むことができない。3. 推定に用いる各種パラメータについて十分な調査に基づいて決定していない。

本研究ではこれらの問題点に対し、まず、結合サイトを含むウインドウ部分が、ピーク位置として選択されないケースについて調査を行い、ウインドウ位置と最大ウインドウ類似度をとる位置の関係を利用した決定方法に改良した。これにより既存のピーク位置決定方法では、ピーク位置として選択されなかった結合サイトが選択可能になることを確認した。次に、結合サイト候補の数を絞り込むため、遺伝子間の類似性の関係を使用して被制御遺伝子の数を絞り込む方法を提案した。具体的には、既存手法により得られた遺伝子間の類似性をグラフとして表したとき、同一転写因子の被制御遺伝子ならばそれらの遺伝子はクリークを構成する可能性が高いことから、3 頂点以上の頂点からなるクリークを構成するピーク位置を結合サイトの候補が高い部分として判別した。これにより結合サイト候補の数を既存手法に比べ  $1/3$  にまで絞り込むことに成功した。また、可能性が高い位置を特定する方法として、ランダムに生成した配列との比較により特異的に高いウインドウ類似度を持つ結合サイト候補を判別する方法を提案した。この方法では結合サイトの長い転写制御因子の被制御遺伝子の場合には高い確率で発見できることが確認できた。また、各種パラメータについて全体の推定結果に対する最適な設定値の調査を行った。