

Title	転写制御領域の解析と破壊株データからの遺伝子の依存関係推定に関する研究
Author(s)	小倉, 亨
Citation	
Issue Date	2003-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1697
Rights	
Description	Supervisor:平石 邦彦, 情報科学研究科, 修士

転写制御領域の解析と破壊株データからの 遺伝子間の依存関係推定に関する研究

小倉 亨 (110028)

北陸先端科学技術大学院大学 情報科学研究科

2003 年 2 月 14 日

キーワード: 遺伝子, 遺伝子間の依存関係, DNA, 破壊株データ, 統計的手法.

近年、ヒトを始めとする様々な生物のデオキシリボ核酸 (DNA) 上の塩基配列が決定した。そのため、研究の焦点は遺伝子や生体活動に関わる蛋白質などを明らかにしようとする動きに変わってきている。

DNA は遺伝情報を担っており、細胞内の核に存在する物質である。遺伝子は DNA 上ので生体機能に関わる蛋白質を生成する発現を行う部分の事を指す。遺伝子の発現は転写、翻訳の段階を経て蛋白質を生成する現象である。発現は遺伝子上流部分にある転写制御領域と呼ばれる部位に転写活性化因子と呼ばれる蛋白質が結合することにより開始する。転写段階では遺伝子上の塩基配列と対になる配列を持つメッセンジャーリボ核酸 (mRNA) が生成される。転写段階で生成された mRNA は核の外に移動し翻訳の段階で蛋白質を生成する元になる。翻訳段階では mRNA の塩基配列から対応したアミノ酸を結合させて蛋白質を生成する。また、上流部分にある別の転写制御領域に転写抑制因子と呼ばれる蛋白質が結合する事により転写は抑制される。

転写を活性化させる蛋白質や転写を抑制する蛋白質の事を転写因子と呼ぶ。また、転写因子を発現によって生成する遺伝子は調節遺伝子と呼ばれる。調節遺伝子が発現により生成する転写因子によって発現に影響を受ける遺伝子は調節遺伝子と依存関係を持つ。

調節遺伝子と依存関係を持つ遺伝子上流部分には調節遺伝子が発現して生成した転写因子が結合する転写制御領域が存在し、その領域内の塩基配列は特異的である。本論文では DNA 配列を文字列として扱うため、転写制御領域内の塩基配列は転写制御領域の部分文字列として扱う。この様な部分文字列の事を本論文では文字列パターンと呼ぶ。ゆえに調節遺伝子と依存関係を持つ遺伝子は特定の文字列パターンを持つ。文字列パターンを持つ遺伝子とは転写制御領域にその文字列パターンを含む事を意味している。本論文では同一の文字列パターンを持つ遺伝子群は共通の転写因子によって影響を受けているかどうかを破壊株データを用いて検証する。この方法により遺伝子間の依存関係推定における転写制御領域の解析の有意性を示す事を目的とする。破壊株データとは標的となる遺伝子を破

壊し各遺伝子の発現量を観測した DNA マイクロアレイデータの事である。本論文の目的は遺伝子間の依存関係推定に対して転写制御領域を考慮した推定方法の基礎的部分を担うものである。推定された遺伝子間の依存関係は細胞内の状態変化を定性的な解析に利用する事が出来る。また、細胞内の状態変化を解析する事は新薬開発など様々な利用法が考えられる。本論文では対象となる生物として DNA 上の塩基配列や各遺伝子の場所などが知られている枯草菌を用いる。

本論文の手法としては枯草菌の DNA 塩基配列を文字列として表現したデータから調節遺伝子が発現して生成する転写因子が結合する事が既知である文字列パターンを持つ遺伝子をパターンマッチングを用いて調査する。DNA は二つのストランドと呼ばれる塩基配列が互いに結合して二重螺旋構造を形成している。今回は各遺伝子の上流部分を両方のストランドから抜き出し調査を行った。遺伝子の上流部分を特定するため、各遺伝子の場所、転写方向等のデータを利用して上流部分を範囲を決めて抜き出した。各遺伝子の上流部分を文字列として抜き出して、調べる対象となる文字列パターンについて全遺伝子の上流部分を調査した。

文字列パターンを持つ遺伝子群が調節遺伝子に影響を受けている事を調べるために破壊株データを用いて統計学的手法を行った。調節遺伝子と他の遺伝子との相関係数を破壊株データから求め、調節遺伝子と文字列パターンを持つ遺伝子との相関係数が高い場合にはその二つの遺伝子間には依存関係を持つ可能性がある。文字列パターンを持つ遺伝子群を標本とし、全遺伝子を母集団とした時に調節遺伝子との相関係数を求め、標本と母集団の相関係数の平均値に差が有意であるか調査した。

今回は文字列パターンを持つ依存関係既知の遺伝子群、依存関係既知の遺伝子と一律に発現に対して影響を受ける遺伝子群（オペロン）、文字列パターンを持つ全ての遺伝子群、文字列パターンを持つ依存関係未知の遺伝子群を標本とし調査を行った。結果としては文字列パターンを持つ依存関係既知の遺伝子群やそのオペロン群に対しては本研究の統計的手法において半数以上の調節遺伝子に対して標本と依存関係を持つ可能性があると判断する事が検証できた。これは DNA マイクロアレイデータを用いた本論文の解析手法で半数以上の調節遺伝子に対して既知の事実と一致する事を示している。しかし、文字列パターンを持つ全ての遺伝子群では依存関係を持つ可能性がある調節遺伝子は半数以下になり、文字列パターンを持つ依存関係未知の遺伝子群では1つの調節遺伝子のみに関して依存関係がある可能性を持つという結果になった。

文字列パターンを持つ依存関係未知の遺伝子群の調節遺伝子との相関係数を調べると文字列パターンを持つ依存関係既知の遺伝子群に近い相関係数を持つ遺伝子は存在するが、相関係数が低い遺伝子を多数含むため、依存関係未知の遺伝子群全体の平均値が低くなる事が分かった。

文字列パターンを持つ依存関係未知の遺伝子群の中に低い相関係数を含む理由としては、未知の転写因子によって遺伝子が影響を受けることと、依存関係を持たない遺伝子が存在することが挙げられる。

未知の転写因子によって影響を受ける場合にはその転写因子を生成する調節遺伝子の特定が必要になってくる。また、依存関係を調べる方法として調節遺伝子との相関係数ではなく、偏相関係数を用いる方法でこの様な問題を解決することが出来ると考えられる。

標本内に依存関係を持たない遺伝子を含む事に対しては、調節遺伝子との相関係数で調査するのではなく、既知に影響を受ける遺伝子との相関係数を求める必要がある。このような相関係数を疑似相関係数と呼び、疑似相関係数が正に高いほど既知の遺伝子とデータ上での値の傾向が類似する事を意味する。この疑似相関係数に閾値を設けて依存関係の無い遺伝子を削除していく方法が有効であると考えられる。