

文献情報からの多体相互作用パスウェイ・ネットワーク抽出および協調作用に関する研究

●井原 茂男

東京大学・先端科学技術研究センター

〈研究の目的と進め方〉

現在、生命情報を網羅的に解析する技術が進展し、そのデータも急増している。これらの情報から生命活動を理解し、創薬、診断、治療などに応用していくには、各因子の相互作用、パスウェイの理解が必要不可欠である。パスウェイの生物学・医学的意味を明確化するためには、DNA、蛋白質、あるいは生物物質など、いくつかの作用が時間的に空間的に連動する協調現象の理解が必要である。

本研究の目的は、文献情報から集めた蛋白質間の2項関係をもつ情報のなかから、多数の蛋白質の間に特有な多体相互作用に起因するパスウェイ情報を抽出するための技術開発にある。文献検索から対象のもつ多体関係を具体的にひろいだし、蛋白質ネットワークの詳細な意味付けを行うことは、新規の試みである。

〈研究開始時の研究計画〉

- 1) 転写制御を行うための関連DNA、蛋白質名、さらに転写制御を抽出するのに必要な動詞、名詞句などを多体相互作用の事例の収集と分析し、選定し、解析するための環境を構築し、転写制御領域に対して解析を実行する。
- 2) さらにシミュレーション、マイクロアレイ実験を併用して、多体相互作用に起因する協調作用、例えば転写開始、複合蛋白質、メチレーションなどを解析する環境を構築し、いくつかについて解析する。
- 3) 複合蛋白質解析のための課題の抽出

〈研究期間の成果〉

転写制御を行うための関連DNA、蛋白質名、さらに転写制御を抽出するのに必要な動詞、名詞句などを多体相互作用の事例の収集、分析、選定、解析を行うための環境を構築し、転写制御領域に対して解析を実行した。

また、複合蛋白質解析のための課題の抽出を行った。具体的には以下の通り：

- 1) 転写制御に特化し、典型的な文型を選定して用語整備を進め、転写制御で最も重要な相互作用であるbindに対して、1970年から2003年までのMEDLINEに対して検索を行い、解析を試みた。様々な転写因子に関係する3体の関係性の情報抽出を行ったところ、403件ヒットした。専門家に検証を依頼した結果、そのうち正しい転写関連の関係性の文章は280件であり、正解率は70%であった。
- 2) 不正解の要因を分析した結果、修飾が複雑な文体の場合での多体関係の同定の誤り、細胞名と蛋白質名が同一のシノニムによる誤認定、多体の関係ではあるが転写因子ではない蛋白質を転写因子と誤同定したことが判明した。
- 3) 文献検索で用いるSVM法の改良の試みとして、マッチングを容易にし、アウトライヤーを容易に抽出するためのSVMの手法を改良した。いくつかの例では従来法で見出しにくかったアウトライヤーを見出すことがで

きるようになった。

〈国内外での成果の位置づけ〉

本研究を遂行していた期間、欧米のいくつかの研究機関から相互作用は明示しないが、ネットワーク上でのトポロジーからモチーフを抽出する試みが行われた。本研究は、さらに進んでトポロジーに相互作用上の意味を与えることができ、現在でも、解析水準は進んでいると思われる。

〈達成できなかったこと、予想外の困難、その理由〉

初年度において、多体相互作用の事例の収集と分析に時間がかかり、シミュレーション、マイクロアレイ実験との相乗効果の結果をだすまでにはいたらなかった。次年度において、多体相互作用の事例の収集と分析をさらに進め、転写制御に対する結果をデータベースCOOPERIONとして公開予定であったが、単年度のみの研究期間となったためデータ収集ができず、データベース化は実現できなかった。

〈今後の課題〉

より多くの関係性を集めるべく文型を収集し、本研究で一部分構想したコンセプトが実際の問題でも有効であることを示す。

〈研究期間の全成果公表リスト〉

1. 0404071441
Ihara, S., Nishikawa, N, S., Ohta, Y., Abe, K., Tsuji, S., Tsutsumi, S., Yamamoto, S., and Aburatani, H., Response Analysis in Pharmacogenomics by Literature Mining, Joint Cold Spring Harbor Laboratory/Wellcome Trust Conference, Pharmacogenomics, 1st meeting on pharmacogenomics and related applications of genomics, 25 (2003).
2. 0404071449
Komura, D., Nakamura, H, Tsutsumi, S., Aburatani, H., and Ihara, S., Multidimensional SupportVector Machines for visualization of gene expression data, ACM Symposium on Applied Computing, Nicosia, Cyprus, 175-179 (2004).
3. 0404071455
Komura, D., H. Nakamura, H., Tsutsumi, S., Aburatani, H., and Ihara, S., Features of Gene: Extraction by Nonlinear Support Vector Machines in Gene Expression Analysis, 14th International Conference on Genome Informatics, 14, 322-323 (2003).
4. 0404071501
Tsutsumi, S., Ihara, S., and Aburatani, H., Two Distinct Gene Expression Signatures in Predicted Acute Lymphoblastic Leukemia with MLL Rearrangements, Cold Spring Harbor Laboratory Meetings, System Biology:Genomic Approaches to Transcriptional

Regulation, 102 (2004).

5. 0404071510

堤修一、井原茂男、油谷浩幸, アレイ技術とがん研究,
血液、腫瘍科 (科学評論社), 48, 182-189 (2004).