

ゲノムワイドの遺伝子発現データマイニング

— 酵母菌老化研究への応用

理論構築グループ 研究員

上村 泰央

DNAマイクロアレイなどの大規模遺伝子発現解析実験の結果得られる大量の遺伝子発現データを解析するための一連のソフトウェアシステムを構築し、これを酵母菌の老化に伴う遺伝子発現解析実験データへ適用した。酵母菌の遺伝子・ゲノムに関する知識を活用し、とくに染色体地図、遺伝子発現調節ネットワーク、タンパク質間相互作用ネットワーク上での各遺伝子の位置と発現量の変化に注目し解析を行なった。このようなゲノムワイドの遺伝子発現データマイニングは、生体内で起きている変化の全体像を捉えるうえで有効な手段である。

1. はじめに

近年のDNAマイクロアレイ、DNAチップ技術とゲノム配列決定プロジェクトの急速な発展により、生物が保持するすべての遺伝子の発現状態を1度に解析することが可能になってきている^{1,2)}。こうしたゲノム規模での遺伝子発現解析は、ポストゲノムプロジェクトとして、ゲノム配列が決定されたあとに当然主流となるべき研究のひとつであり、現時点においてもこれらの技術を用いた新しい結果が数多く出てきている。一方で、この解析の結果、必然的に産み出されてくる大量のデータを扱うことは困難であり、そのためにコンピュータを用いた体系的なデータの蓄積、管理を行なう必要性が生じている³⁾。

こうした背景のなか、当プロジェクトにおいても出芽酵母(*Saccharomyces cerevisiae*)の老化によって起こる遺伝子の発現量の変化をゲノム規模で調べるためにDNAマイクロアレイを用いた解析を進めてきた。出芽酵母は生物学者たちによって最もよく研究されているモデル生物のひとつであり、全ゲノム配列を含め、数多くの遺伝子の機能に関する豊富な知識が利用できる生物であ

る^{4,5,6)}。そこで本研究では、まずDNAマイクロアレイを用いた遺伝子発現解析データを扱ううえで必要となるソフトウェアシステム環境を構築した。とくに酵母菌ゲノム・遺伝子に関する知識を最大限に活用し、発現データと統合することにより、個々の遺伝子レベルでの発現量の変動を調べただけに留まらず、よりマクロなレベルで細胞内状態変化の全体像を捉えることを念頭においたシステムの構築を行なった。さらにこのシステムを実際に用いて酵母菌の若い細胞と老化細胞との間での各遺伝子の発現量の変化を、染色体地図上、遺伝子発現調節ネットワーク上、タンパク質間相互作用ネットワーク上での各遺伝子の位置と関連づけて解析を行なった。

2. 遺伝子発現データ解析システムの構築

DNAマイクロアレイの実験結果として得られる画像データから、各スポットにおけるシグナル強度を測定し、各遺伝子の発現量を定量化するソフトウェアSpotEditor、研究者による発現プロファイルの解釈を支援するためのデータベースシステムSpotDB、遺伝子発現プロファイルの染色

体上での解析を可能にする染色体表示解析ソフトウェア Chromosome Viewer などを開発した。

一般に大規模遺伝子発現解析において重要なのは、各遺伝子の発現状況からその生体内でどのような現象が起きているかを総合的に理解することである。そのためには、個々の遺伝子の機能や実験対象の生物についての専門的な知識とあわせて発現データを解釈することが必要である。SpotDBは、このような知識と発現プロファイルを統合したカード型データベースであり、研究者が実験結果をより高度な視点から解釈するための環境を提供する目的で開発された。本データベースシステムは、MIPS⁹⁾のYeastデータベースに蓄積されている6000以上にも及ぶ酵母の全遺伝子に関する知識をベースに作成された。具体的には、染色体についての情報、各遺伝子の機能についての情報のほかに、全遺伝子を機能によって分類したFunctional Catalogや、複合体によって分類したComplexes Catalogなどを含むさまざまなカタログ情報も、データベースに取り込まれている。このような情報を統合することにより、研究者は任意のカテゴリ別にそれらの発現情報を容易に調べることができる。

3. 酵母菌の老化に伴う遺伝子発現データへの適用

3.1. 染色体上での遺伝子の位置と発現量の解析

染色体地図上での各遺伝子の位置関係と、老化に伴う発現量の変化をChromosomeViewerを用いて解析した。計6144個の遺伝子の発現プロファイルのうち、若い細胞に多く発現している遺伝子278個と老化細胞に多く発現している遺伝子161個を選び、それぞれ染色体地図上にマップした。その結果、老化に伴い発現量が変化する遺伝子は染色体上に均一に配置されているのではな

く偏って配置されているという傾向がみられた。いくつかの染色体では、若い細胞に多く発現する遺伝子が集中する領域と老化した細胞に多く発現する遺伝子が集中する領域とが顕著に現れた。この結果から、老化によって染色体上の位置に依存した遺伝子発現量の変化が起きていると考えられる。

3.2. 遺伝子発現調節ネットワークの解析

遺伝子発現調節ネットワーク上で各遺伝子の発現量を若い細胞と老化細胞で比較することにより、若い細胞で活発に機能している発現調節ネットワークの系と老化した細胞で活発に機能しているネットワークの系を比較することができる。そこで、YPD⁹⁾から既知の転写調節情報を集め、計631個の遺伝子と1218の転写調節関係からなる酵母遺伝子発現調節ネットワークを得た。これに発現データを重ね合わせて解析した結果、若い細胞、老化細胞それぞれで活性化されているとみられる部分的なネットワークを見つけることができた。

3.3. タンパク質間相互作用ネットワークの解析

次に、タンパク質間相互作用ネットワーク上で遺伝子発現データの解析をおこなった。そのためにYPDなど⁹⁾から既知のタンパク質間相互作用情報を集め、計1855個のタンパク質と2760の相互作用からなる酵母タンパク質間相互作用ネットワークを得た。これに発現データをあわせてみることによって、若い細胞と老化した細胞のそれぞれに多く発現している遺伝子のクラスターを相互作用ネットワーク内に見つけることができた。これにより老化に伴い生じた細胞内の変化について、タンパク質間相互作用、すなわちタンパク質どうしがどのように機能しているか、という観点

から考察することができた。

4. おわりに

本研究において開発されたシステムは、大規模遺伝子発現解析データをもとに、(1)生物学者が経験と専門知識を用いて発現データを解釈するため、(2)情報学者がコンピュータを用いて体系的に発現データを解析するため、の両方面からの解析を行なうために共有されるべきデータおよびソフトウェア環境の基盤を提供するものである。我々の本来の目的は、コンピュータによって処理、蓄積されたデータから、その裏に潜む重要な法則や事実を発見するデータマイニングにあり、あるいは実験にフィードバックされるような重要な知見を実験者に還元することにある。

本システムを用いて遺伝子発現プロファイルを作成することにより、計算機上に実験データを蓄積することが可能になり、そのデータを利用して染色体、遺伝子発現調節ネットワーク、タンパク質間相互作用ネットワークの上で体系的に解析をすることができた。その結果いくつかの知見を実際に実験者に還元できたと考えている。

染色体地図上での遺伝子の位置関係と発現情報をあわせてみることは、遺伝子が染色体上の連続鎖と不連続鎖のどちらに配置するかという染色体上の遺伝子の方向性と遺伝子発現の関係をさぐるためにも意味のある解析である。老化によって発現量に違いが生じる遺伝子の染色体上での偏った局在は、老化という現象を非対称なDNA複製の積み重ねによって起きた細胞の変化と考えるならば、その結果生じた、染色体上の位置に依存した遺伝子発現の変化と非対称なDNA複製機構との相関を示唆する興味深い結果である。

謝辞

本研究におけるソフトウェアシステムの開発は非常勤技術員のチャン・ナム・トアン氏、アルバイトスタッフの張巍氏、付欣氏、周毅氏の協力によるところが大きかった。ここに深く感謝の意を表したい。

参考文献

- 1) Duggan, D.J. *et al.* Expression profiling using cDNA microarrays. *Nature Genet.* **21**, 10-14 (1999).
- 2) Brown, P.O. & Botstein, D. Exploring the new world of the genome with DNA microarrays. *Nature Genet.* **21**, 33-37 (1999).
- 3) Ermolaeva, O. *et al.* Data management and analysis for gene expression arrays. *Nature Genet.* **20**, 19-23 (1998).
- 4) Mewes, H.W. *et al.* MIPS: a database for protein sequences, homology data and yeast genome information. *Nucleic Acids Res.* **25**, 28-30 (1997).
- 5) Costanzo, M.C. *et al.* The yeast proteome database (YPD) and *Caenorhabditis elegans* proteome database (WormPD): comprehensive resources for the organization and comparison of model organism protein information. *Nucleic Acids Res.* **28**, 73-76 (2000).
- 6) Aach, J. *et al.* Systematic Management and Analysis of Yeast Gene Expression Data. *Genome Res.* **10**, 431-445 (2000).
- 7) Uetz, P. *et al.* A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623-627 (2000).