



Title	遺伝子発現情報に基づくデータの集約化と視覚化
Author(s)	峯田, 克彦; Mineta, Katsuhiko; 池尾, 一穂 他
Citation	情報知識学会誌, 16(3), 50-55
Issue Date	2006-08-08
Doc URL	https://hdl.handle.net/2115/44323
Rights	本文データは学協会の許諾に基づきCiNiiから複製したものである
Type	journal article
File Information	mineta.pdf



遺伝子発現情報に基づくデータの集約化と視覚化

Integration of Gene Expression Profiles into the Visualized Data Management System

峯田 克彦* 池尾 一穂[†] 田中 譲* 五條堀 孝^{†,‡}

Katsuhiko MINETA Kazuho IKEO
Yuzuru TANAKA and Takashi GOJOBORI

近年の遺伝子発現情報の爆発的な増大は、その基盤となる技術の革新とともに研究者に遺伝子がどのように使われているかという情報を網羅的に入手する機会を提供している。しかしながら、データの増加に伴い、データの取り扱いや検索などの解析において、これまでの生命研究では直面しなかった新たな課題が生じている。本稿では、このような課題に対して我々が取り組んでいる高次元生物情報可視化プラットフォームを紹介するとともに、今後の遺伝子発現情報を用いた研究に関して議論する。

An explosive increase of the gene expression data provides us a chance to analyze fundamental biological phenomena such as how the genes are being used in the organisms. As the data increases, however, we faced new problems how we deal with and analyze an enormously large amount of data. Moreover, the gene expression data includes information of the time, location and network of the genes that cannot be managed easily. Thus, the establishment of a novel management system for the gene expression data is one of the most essential tasks in the biological data analyses. Here, we introduce our visualizing platform of higher-dimensional biological data for the management of the gene expression data.

キーワード：遺伝子発現情報，可視化，ネットワーク，データベース
Gene Expression, Visualization, Network, Database

1 背景

1.1 遺伝子発現情報

遺伝子発現解析は、生物個体中に存在するゲノム配列に記されている遺伝情報がいつ、どこで、どのようなタイミングで利用されているかという情報を提供する。この遺伝子発現情報を解明することは、例えば疾患の発生機序の解明、新薬の開発、食品工学などの応用に結びつくことから生命科学における最も重要な課題の1つとなっている。遺伝子発現情

報は、特定の生命現象に注目しても多くの遺伝子が関与し、かつ各遺伝子間で複雑なネットワークを構成していることが知られている。このネットワークは、局所的な同一部位中での反応だけではなく、隣接する部位への情報伝達によって3次元空間上に経時的に構成されるものである。

1.2 遺伝子発現情報の実験的取得方法と問題点

このような遺伝子発現情報は、実験的にはある時系列のある時点における、さらに各組織の各細胞における遺伝子発現頻度(発現している分子数)の集積という形で収集されている。近年の遺伝子発現測定技術の進歩により同時に観察できる遺伝子発現頻度の数は爆発

* 北海道大学大学院情報科学研究科
Graduate School of Information Science and Technology, Hokkaido University

[†] 国立遺伝学研究所 生命情報・DDBJ 研究センター
CIB-DDBJ, National Institute of Genetics

[‡] tgojobor@genes.nig.ac.jp

的に増加しており、遺伝子発現情報自体の蓄積もそれに伴い増加している。そのため、解析手法においても、これまでの単独～複数個の遺伝子発現情報を取り扱うための方法では対処することができない状況が生まれてきた。現在では、例えば多変量解析や統計的な手法をこれら大規模データに用いることで、共通に発現する遺伝子群の同定や、ある機能の特徴づけるような遺伝子発現パターンの解明などが行われている。

このような解析においては、一般に個々に分離した細胞や組織のサンプルにおける発現頻度を計測し、それを単体のデータや平面的な表形式データとしてだけ取り扱っており、それに対する解析は、ある組織でのある遺伝子の時系列の変化をそれぞれ単独でグラフ化することや比較することによって行われている。

しかしながら、前述した解析方法においては、各細胞や各組織に関する物理的な距離に関する情報や三次元的な空間情報が加味することができず、実際の生物体内で生じる細胞、組織間でのより高次の遺伝子発現情報・ネットワークの解明においては、十分な情報・解析を提供できるとは言い難い。加えて、観察者である研究者の発想は、二次元に投射されたデータを俯瞰し、もしくはそれに対する種々の解析を行ったとしても、生物体自体からの連想を喚起するためには貧弱な情報であり、新規知識発見を行うためにも上記のような問題点を克服できるような手法、考え方が必要である。

1.3 先行研究の例

これまでにこのような問題意識の下、国外のグループにおいて、マウスの脳に注目したデータベースの開発が報告されている^[1]。彼らは、*in situ* ハイブリダイゼーションという遺伝子発現の観察手法に基づき、遺伝子発現情報を収集するとともに、マウス脳の三次元データを構築し、その発現パターンを3次元空間上にマップしている。*in situ* ハイブリダイゼーションという手法は、特定の遺伝子の

発現パターンを特定の部位(ここでは脳の切片画像)で観察する手法であり、原則として、調べたい遺伝子の数 \times 部位の断面数の画像データが必要となる。彼らは、この大量画像データの処理を自動化しており、今後の研究の進捗にともしない有力なデータベースとなることが期待できる。

1.4 今後の展望

一方で、遺伝子の機能情報や発現情報は、ヒトだけに限らず、むしろ実験の容易なモデル生物(マウス、ラットなど)において集積されているという状況がある。一般に生物が持つ遺伝情報は生物進化の過程で保存されてきているため、ヒトでの特定遺伝子の機能を調べるためにマウスやラットで実験を行ってその機能を推測するという手法は、非常に良く用いられている。つまり、生物の機能をより詳細に理解していくためには、特定の生物種の情報だけに依らず、むしろ複数の生物種の情報から互いに補完・推測することでより付加価値の高い解析を行うことが可能となる。

前述したようにデータ解析の欠点を補い、さらに生物種の比較までも含めた遺伝子発現情報解析を行う仕組みというのは、これまでに存在していない。そのため、ここで問題としているようなすべての情報を内包し、かつ解析可能とするためのプラットフォーム、すなわち生物の空間情報、時間情報、進化情報という異なる軸の情報を3次元形状データを核として統合した新規解析プラットフォームの開発は世界的にも緊急の課題であるととも非常に興味深い研究対象となっている。また、このようなプラットフォームを開発することは、新たな研究分野の創出や遺伝子によるネットワークの解明などにつながる遺伝子発現情報解析への支援といった視点からも大変意義深いものと考えられる。加えて、上記アプローチを実装することで、単独に存在する種々の遺伝子発現情報を生物の形状データを元に再構成することが可能となり、今まで数日～数ヶ月かかっていた実験を既知の情報から再構築し、

コンピュータ上で実験結果をシミュレートすることも可能となる。以下の項では、我々の開発しているプラットフォームについて紹介する。

2 高次元生物情報可視化プラットフォーム

我々は以上のような背景の下、高次元生物情報可視化プラットフォームを開発している。このプラットフォームは、上記のような空間的・時間的な遺伝子発現情報を統合し、利用者の研究を活性化することを目指している。そのためには、生物の細胞単位や部位単位にその形状データと遺伝子発現情報を備えること、そして、それらに対する解析の手法を実装することが必要である。我々は、コンピュータ上に構築した3次元空間内で生物の形状データの表示と遺伝子発現情報の可視化、およびそれらの解析を行えるシステムの開発を行っている。詳細を以下に説明する。

2.1 立体形状データ

生物の3次元デジタル情報として、核磁気共鳴画像法などによる実測に基づく方法と擬似的なコンピュータグラフィクス(CG)を用いる方法がある。実測から得られる情報は、非常に正確な形状を提供することができる反面、個人の識別をも可能になるという危険を内包している。そのため、ヒトの情報を取り扱うことを念頭におくと、現在我々を取り巻く環境下では難しい状況にある。一方、CGにより作成された形状データは、厳密性という意味で実測データには及ばない点もあるが、不特定多数の標準的な形状をモデル化して提供することが可能となる。CGによる形状作成の場合も、実測データを参考にすることで非常に精度の高い形状データを入手することが可能である。そこで我々はCGによる形状データを採用してシステムの基盤としている。

2.2 遺伝子発現情報の取り扱い

遺伝子の発現情報を観察する手法には、一般に網羅的に収集する方法と個別に収集する手法の2種類の手法が存在する^[2]。詳細な説明は他の書籍に譲るが、前者にはcDNAマイクロアレイ、DNAチップ、SAGEなどといった手法が、後者にはノーザンハイブリダイゼーション、*in situ* ハイブリダイゼーションなどといった手法が含まれる。特に前者の手法により産み出されるデータは、大量の発現情報が収集されるという意味において、我々のプラットフォームでぜひとも使用したい情報である。ここで例に挙げたcDNAマイクロアレイやDNAチップといった手法による遺伝子発現情報は、実験手法や使用した材料などの差異により、異なった研究グループのデータ間での比較ができないことが長年の問題であった。近年、このような遺伝子発現情報を有効利用していくためにMicroarray Gene Expression Database 会議^[3]が国際に統一した実験手法、用語を提案しており、徐々に共通化の基盤が整備されつつある。一方、後者の個別に収集される遺伝子発現情報は、網羅的な収集で見落とされる可能性のある詳細な遺伝子発現情報の入手が可能であり、どちらのタイプのデータも我々のプラットフォームには不可欠な情報と考えている。前出の通り、遺伝子発現情報は、上記のいずれの方法であっても、ある遺伝子がどこでどの程度発現しているかという表で表現されるものであり、我々のプラットフォームでは、この形式のデータは全て取り扱いが可能である。

2.3 遺伝子発現情報の可視化

我々のプラットフォームにおいて、もっとも主要な特色として遺伝子発現情報を3次元で表現した形状データ上に表示する機能がある(図1)。このことで、どの遺伝子がいつ、どこで発現しているのかということを利用者は直感的に理解することが可能となる。また、3次元形状データはコンピュータ上で任意に移動、

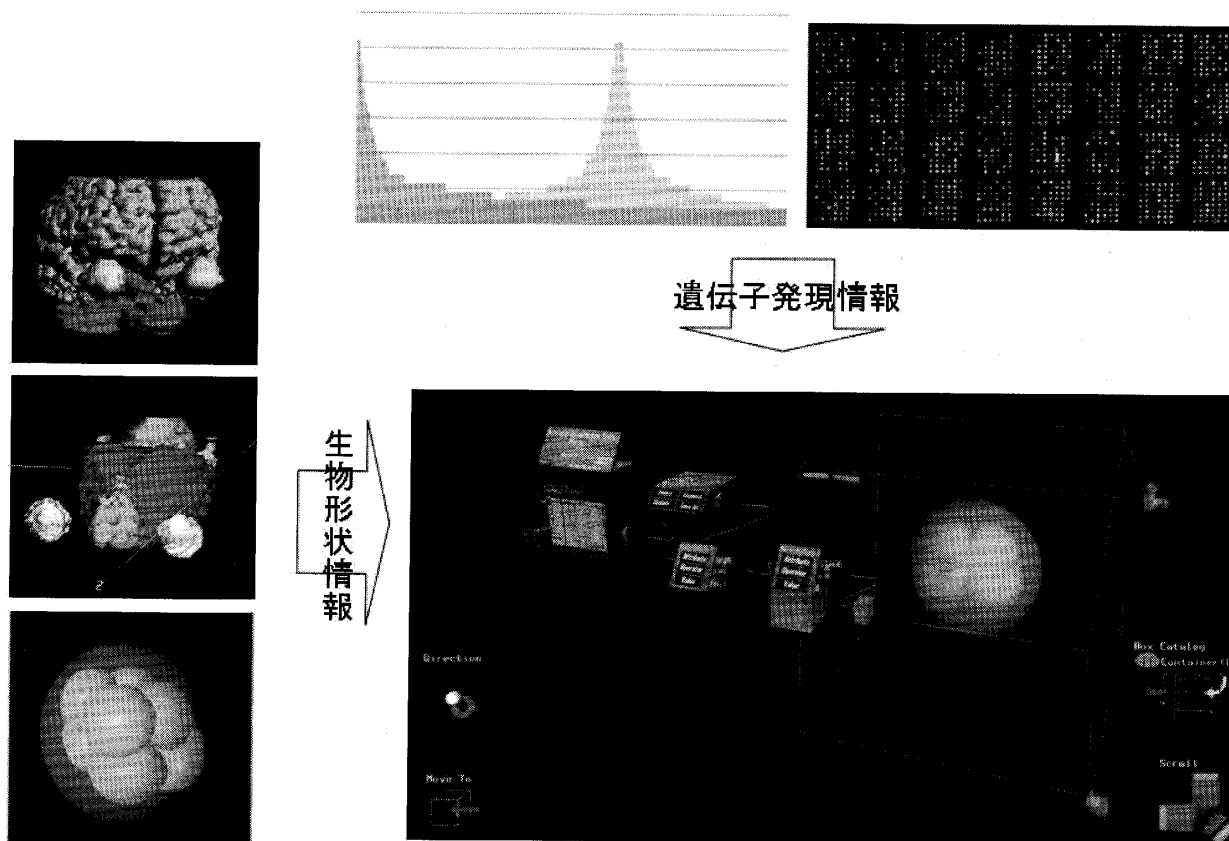


図1 多次元生物情報可視化プラットフォームの表示例

左側の図は、生物形状のCGモデルの例であり、上からヒト脳、マウス脳、ホヤの細胞(8割球期)を表す。上側の図は、遺伝子発現情報の例であり、右側はcDNAマイクロアレイの実験結果、左側は遺伝子発現情報をグラフ化したものである(縦軸:遺伝子の発現頻度、横軸:遺伝子)。中央の図は、多次元生物情報可視化プラットフォームの表示例である。ここでは、ホヤの細胞を表示させ、データベースからの遺伝子発現情報を視覚的に操作している。

切断することが可能であるので、実際の生物を実験的に扱っている際には見ることでできない、内側や裏側といった部分での発現情報も容易に観察することができる。また、時間的にどの段階から遺伝子が発現するようになるのかといった情報の視覚化も可能である。

2.4 遺伝子発現情報の空間的比較・検索

遺伝子発現情報は、データベースとして内部に格納されているので、通常の検索を行うことはもちろん可能であるが、さらに特定の3次元形状に注目して、ある細胞で発現している遺伝子を表示させたり、また細胞a, b, cに共通に発現するものを選ぶといった特殊

な検索を行ったりすることも可能である。また、検索結果を可視化して3次元空間上に表示すること(図2)で、遺伝子間の相互作用や、on/offといったネットワーク解明へとつながる情報を解析者に視覚的に認識させることができる。

2.5 遺伝子発現情報の時間的比較・検索

上記に加えて、生物の発生過程などといった経時的な遺伝子発現情報の変化を検索によって、見つけ出すことも可能である。例えば、発生初期のみに発現する遺伝子群を表示させたり、成体中で発現している遺伝子が発生過程のある時期にどこで使われているのかといっ

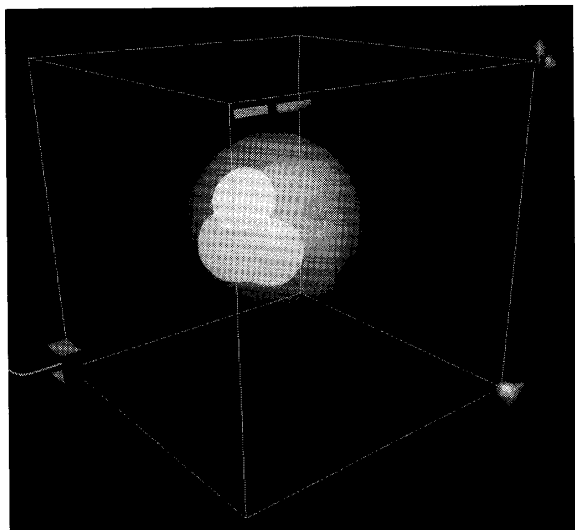


図2 遺伝子発現情報の可視化例
ホヤの初期発生での各割球において、発現している遺伝子を異なる色で表示させている。また、ここでは発現の程度は各色の濃淡で示してある。このように3次元空間中の生物形状データ上で実際に遺伝子がどこで、どの程度発現しているのかを利用者は認識することができる。

た情報も直感的に認識できる形で検索を行うことができる(図2)。

2.6 遺伝子発現情報の進化的比較・検索

前項でも述べた通り、単独の生物種に注目した情報だけでは限界がある。そこで、本プラットフォームにおいては、複数の生物種の情報も横断的に視覚化するとともに、発現情報を検索することができる(図1)。その結果、ヒトで情報のない(少ない)ある遺伝子の発現情報をマウスの結果から検索し、ヒトでの機能を推測するといった生物の進化的時間軸に注目した比較解析を行うことが可能となる。また、進化的に遺伝子の機能は必ずしも同一でない場合も考えられる。このような同一性と相違性に注目することで、配列の相同性と機能の相同性の比較、ひいては進化的なネットワークの比較といった研究を支援していくことが可能である。

2.7 関連する生命情報との統合

遺伝子発現情報は、遺伝子がいつ、どこで、どのように使われているかという情報であるが、それに加えて、例えばゲノム中での位置関係、翻訳されるアミノ酸配列やタンパク質の立体構造情報、実験的に明らかになっている機能などといった付加的な情報も多くはデータベースや文献情報として公開されている。また、生物学的な基本情報として、例えば細胞の発生系譜、細胞・組織・器官の階層性などといった情報も同じく遺伝子発現情報以外の情報として重要である。遺伝子発現情報を元にした解析を行う際にもこのような付加的な情報は不可欠であり、我々のプラットフォームにおいても有機的に各種データを結びつけて表示する機能を備えることが今後、重要になってくると考えている。

3 結び

遺伝子発現情報は、生命科学の基盤となる多くの情報を有しており、そこから如何にして有用な知識を発見していくかということが、今後の研究の進展に不可欠な課題である。我々は、その課題を解決するために、上記のような遺伝子発現情報を空間・時間・進化という複数の軸に基づき、3次元形状データを基盤としたプラットフォームを開発している。本プラットフォームは、研究支援を行うためのシステムとして現状開発が進んでいるが、我々は最終的には遺伝子発現情報のみならず、すべての生物学的情報をこのプラットフォームに統合することを目指している。本プラットフォーム上で生命体自体を再構成し、さらには生命現象をシミュレートすることが可能になるためには、まだまだ多くの生物研究の進展と情報の蓄積が必要ではあるが、このような研究を進めていくことで、生物学研究の根源的な命題である「生命とは何か」という問いに答えることが可能となるよう、引き続き研究を進めていきたい。

参考文献

[1] Carson J.P., Ju T., Lu H.C., Thallar C., Xu M., et al.: "A digital atlas to characterize the mouse brain transcriptome.", PLoS Comput. Biol. Vol.1, No.4:e41, pp.289-296, 2005.

[2] 五條堀孝編：「生命情報学」，シュプリンガー・フェアラーク東京，238p.，2003.

[3] <http://www.mged.org/>

(2005年12月1日受付)

(2006年4月14日採択)