

「生命活動のプログラム」  
平成7年度採択研究代表者

小原 雄治

(国立遺伝学研究所 教授)

## 「線虫全発生過程の遺伝子発現プログラム」

### 1. 研究実施の概要

ゲノムDNAにはその生物を規定するすべての遺伝情報のセット(これをゲノムとよぶ)が書き込まれている。生物が卵から発生していくいわば「台本」が書き込まれているのである。ゲノムDNAの塩基配列の決定は多くの生物で決定あるいは進行中であるが、われわれはまだ塩基配列から「台本」を読み取るルールを知らない。本研究のねらいは、シンプルだが優れたモデル系である線虫を使って、その「台本」を明かにしようとするものである。このために「役者」である遺伝子のすべてを見つけだし、それらが発生の「いつ、どこで、何を」しているかを網羅的に調べる。その結果から、情報学的手法を駆使して、発生のプログラム(すなわち台本)を逆にあぶりだそうとするものである。線虫でのプログラムは高等生物でも共通なことが多いので、ヒトゲノムなどの機能解析への利用価値は測りしれない。

線虫のゲノムDNAの塩基配列は99%が決定されたが、われわれは相捕的なアプローチとして、発現遺伝子の総ざらえをおこなってきた。すなわちmRNAをDNAの形にかえたcDNAのセットを作成し、塩基配列をもとに分類し、約10,000遺伝子を見いだした。これは予想される全遺伝子の1/2以上である。これまでに7,000遺伝子について発生のどの時期のどの細胞/組織で発現しているか解析を終え、残りも今年度中に終了予定である。その発現パターンを情報学的手法で似たものに分類するクラスタリング解析を進め、共通パターンの遺伝子群については、発現調節領域や遺伝子機能の共通性について実験的検証をおこなっている。遺伝子機能については、RNAi(2本鎖RNA注入による遺伝子機能障害)を体系的に実施した。現在はまず初期胚に焦点をあて、この局面に関わる全遺伝子の働きと関係付けを行なっている。今後、対象の発生局面を少しずつ後期にずらせて、発生の「台本」読み取り作業を進めて行く予定である。

### 2. 研究実施内容

#### (1) cDNAクローンのタグ配列決定、分類、マッピングによる遺伝子構造解析

これまでに、約65,000クローンを両端タグ配列を決定し公開した。さらに菅野純夫(東大医科研)との共同研究で、オリゴキャッピング法での完全長cDNAラ

イブラリーを試作し、約4000クローンの解析を追加した。この結果、約10000種のcDNAに分類した。これは19,000と推定される全遺伝子の約半分以上にあたる。新たなトランススプライスリーダー配列など多くの情報が得られた。

昨年に引き続き、ゲノムシーケンシス(99Mb以上)へのin silicoマッピングをおこない、エキソン/イントロンの対応付けを行うことにより、遺伝子構造の確定を進めた。このためにABIシーケンサーのクロマトファイルをUNIXワークステーションに読み込み、ゲノム配列と比較し、エキソンを次々同定していくプログラムをJean Thierry-Mieg (CNRS, Montpellier, France)と共同で開発した。その結果、ゲノムシーケンシングチームが予測した遺伝子のセットであるWormPepは、その半分は何らかの修正が必要であることがわかった。また、多数のalternative splicingパターンやイントロン中の逆向き遺伝子の存在などを見だし、線虫のトランスクリプトームが予想以上に複雑であることが判明した。

分類済み約10,000cDNA種の代表クローンについてはベクタープライマーを用いてcDNA部分を増幅し、マイクロアレイを作成し、内外の共同研究に供している。

## (2) 発生における遺伝子発現パターンの解析

本プロジェクトで開発したwhole mount embryoのマルチウェルin situハイブリダイゼーション法を用いて、7000遺伝子のmRNA発現パターンを解析した。胚発生期については10ステージ(2細胞期、4細胞期、8-12細胞期、原腸陥入開始期、同中期、同後期、コンマ期、1.5折れ期、2折れ期、3折れ期)、後胚発生期については4ステージ(L1-L2、L2-L3、L3-L4、L4-成虫)に分け、それぞれ典型的な試料を含む画像を保存してデータベース化すると共に、各ステージでの発現パターンのアノテーションをつけた。発現パターン情報はNEXTDB (Nematode Expression Pattern Database) に集約し、WWW上で公開を始めた(<http://watson.genes.nig.ac.jp/db/index.html>)。発現の類似性で検索するなど、様々な検索が可能な全く新しいゲノムデータベースになっている。研究コミュニティからのフィードバック情報の受け皿の仕組みも構築し、発現情報のセンターとして機能している。

クラスタリング解析の最初の試みとして、これまでにアノテーションを終えた第3染色体とX染色体を中心とする約5,000遺伝子について発現パターンのクラスタ分析を行った。クラスタリングの方法として、非類似度はユークリッド距離と標準ユークリッド距離の2種類、アルゴリズムは最短距離法、最長距離法、群平均法、メジアン法、k-means法の5種類を用いて、計10種類のクラスタ分析をした。一連のクラスタ分析結果は、中身は多少異なるが、1)ひとつの遺伝子のみからなるシングルクラスタが多数存在し、2)逆に100個以上の遺伝子からなる

巨大クラスターの存在も観察された。現在、クラスター分析結果を詳細に検討し、最適のクラスタリングの方法を設定を試みている。

(3) 母性遺伝子のRNAiによる系統的機能解析とタンパク発現パターン解析

C.elegansの初期発生では受精後4回の卵割の間に各割球の運命決定がおこる。この過程は母性遺伝子に支配されている。これまでの経験や予備的実験から、(1)卵母細胞に存在し、(2)受精後速やかに消失するかあるいは特定細胞系譜に局在化する、という遺伝子が初期発生の運命決定に関与する可能性が高いと考えられた。このような基準でNEXTDBをサーチすると約200遺伝子(約5%)見つかった。この遺伝子群に対しRNAi(RNA mediated interference)による機能解析実験をおこなった。具体的には、標準N2株に各遺伝子の2本鎖RNAを顕微注入し、注入した親虫、F1胚、孵化した幼虫(escaper)の成長、生殖能、形態等の観察を行った。現在のところ、24%にF1胚致死またはF1数の減少、36%にescaperのみ表現型が見られ、40%には全く表現型が見られなかった。内100遺伝子についてタンパク発現パターンを明かにするために、大腸菌で調製した部分蛋白でラット1匹を免疫し、現在までに63種に対する抗血清を得、これによる抗体染色の結果を明らかにした。染色結果の正当性はRNAiをした胚で染色が消えることで確認している。これらの情報はNEXTDBに統合化し、その関係付けを進めている。

(4) 線虫C. elegans初期胚の4次元遺伝子発現データベースの構築

発生過程のコンピューターシミュレーションをめざして、胚発生過程のコンピューターグラフィックス(CG)化、その上に遺伝子発現パターンの重ね合わせの試みをおこなった。いわゆる4D画像(発生の様子を一定の時間間隔でノマルスキー微分干渉顕微鏡で焦点の段階的变化による光学的切片像を多数とったもの)の元画像を用いて細胞と核の輪郭をトレースし3次元再構成をおこない、さらに次の像との間で補間をおこない100細胞期までCG化した。この上に母性遺伝子や極初期の接合体型遺伝子の発現パターン(共焦点顕微鏡による3DのmRNAの分布、蛋白の分布)を重ね合わせる仕組みを開発し、多数の母性遺伝子の発現パターンをCGに重ね合わせ、データベース化した。これはインターネット上での利用が可能であり(研究室内で実行済み)、近いうちに公開予定である。

(5) 線虫 C.elegans発生における遺伝子発現制御の解析

(a) 母性 mRNAの翻訳調節

線虫初期胚の各割球の運命決定は母性遺伝子に支配されているが、この母性遺伝子の発現制御には翻訳制御によるものが多い。これまでに生殖系列の確立に必須なpos-1遺伝子(TIS11型のジンクフィンガータンパク)を見つけてきた。これと相互作用する新規のRNA結合タンパク質遺伝子pip-1を見つけ、これらの機能解析をおこなった。両者のRNAiの表現型は同じであり、いくつかの母性遺

伝子発現を調べたところ、前割球AB由来のみでタンパクが観察されるGLP-1 (Notchホモログ)の翻訳がpos-1, pip-1の制御を受けていることが判明した。すでにAPX-1 (Deltaホモログ)の翻訳制御への関与も見いだしているため、今後関与する遺伝子群を絞りこむことにより、母性遺伝子の翻訳制御メカニズムが明らかになる可能性が出てきた。

(b) 線虫生殖顆粒 (P-granules) 構成因子の機能解析

P granulesは、成熟精子を除く全ての生殖系列細胞に、その発生の全段階を通して観察される。未同定の (poly-A) RNA成分と複数のタンパク質成分から構成されると推定されているが、その成分の多くは未同定のままであった。我々はこれまでにその主要な構成成分のひとつとしてPGL-1タンパク質を同定し、さらにPGL-1と相互作用するPGL-1に相同な二つのタンパク質、PGL-2とPGL-3を同定した。RNAi法を用いて、*C. elegans*の生殖系列の発生において、少なくともPGL-3がPGL-1と重複して機能していることを明らかにしたが、*pgl-3*の欠失変異を逆遺伝学的に作製し、現在その表現型の解析を進めている。

3 . 主な研究成果の発表 (論文発表)

Tabara, H., Hill, R.J., Mello, C., Priess, J. and Kohara, Y.: Pos-1 encodes a cytoplasmic zinc-finger protein essential for germline specification in *C. elegans*. *Development* 126, 1-11 (1999).

Fields, S., Kohara, Y. and Lockhart, D.J.: Functional Genomics. *Proc. Natl. Acad. Sci. USA* 96, 8825-8826 (1999).

Motii, M., Yoshida, S., Morita, K., Kohara, Y. and Ueno, N.: Identification of TGF- $\beta$ -regulated genes in *Caenorhabditis elegans* by differential hybridization of arrayed cDNAs. *Proc. Natl. Acad. Sci. USA*, 96, 15020-15025 (1999).

Cassata, C., Kagoshima, H., Andachi, Y., Kohara, Y., Durenberger, M.B., Hall, D.H., and Burglin, T.R.: The Lim Homeobox Gene *ceh-14* Confers Thermosensory Function to the AFD Neurons in *Caenorhabditis elegans*. *Neuron* 25, 587-597 (2000).