

「ゲノムの構造と機能」  
平成10年度採択研究代表者

松原 謙一

(財)国際高等研究所 副所長)

## 「器官形成に関するゲノム情報の解読」

### 1. 研究実施の概要

高等動物の器官形成は、器官の"芽"となる細胞がまず生じ、それが周辺の細胞とコミュニケーションしながら分化・増殖をくり返して複雑な細胞集団へと発展してゆく過程である。コミュニケーションにさいしては、細胞外分泌物質やそのリセプターをはじめ、各種の誘導・抑制シグナル伝達のカスケードが次々と関与し、整然と働く筈である。

我々はこの過程で動員される数万に及ぶと考えられる遺伝子をゲノム解析手法によって同定・定量し、各遺伝子の経時的発現プロファイルのデータベース作成を進める。このために発生途上のマウスの小脳をモデル系として選び、そこで作動する約1000種の遺伝子の活動度についてデータベースを作成した。昨年に引き続きデータベースを続けて充実させてきたが、それと同時に大量のデータを解析する統計学的手法を開発した。研究手法は、器官形成、器官再生、器官に対する薬理作用、疾病要因の問題等を解析する新しい方法論として展開できる。

### 2. 研究実施内容

マウスの小脳は胎生10日目頃に中脳と後脳の間を生じた小突起から発生して来るが、組織の形成は出生後に活発に起こり、そのさい経時的に反応が同調して起こるのが特徴である。すなわち、生後4日目頃は顆粒細胞のさかんな増殖が特徴的で、12日目頃になると、顆粒細胞の増殖は完了し、かわって、いわゆる分子層における神経細胞のさかんな樹状突起の伸長と、シナプス形成が主となる。これらの反応は生後3週には終了し、成熟した小脳となる。これら一連の反応はゲノム情報の逐次的な発現によって整然と進められるものと考えられている。昨年度は比較的少数の遺伝子に関して解析を行ったが、今年度は測定技術、データ解析法の細部を吟味するとともにより大量のデータを収集した。

#### (1) ATAC-PCRのABI 3700 DNA analyzer への最適化。

ABI 3700 DNA analyzerは、これまでのautosequencerとは異なり、96キャピラリを使っているため、圧倒的な処理速度を持っている。しかしながら、まだ完成されたハードウェア、ソフトウェアではないためにいくつかの問題点がある。例え

ば、キャピラリー毎に感度が一定でないため、至適なサンプル量の幅が狭い。また、断片のサイズの計算値が実際の値よりも5、6塩基頻繁にずれる。これらの問題を、反応と電気泳動の至適化、及びピーク自動抽出のアルゴリズムの改良により解決し結果的に、1000反応/日の速度を維持できるようになった。

また、大量のデータを扱うに際し、新しく浮上してきたのは、データの品質管理の問題である。6アダプターを使用するATAC-PCRでは、その3つに量の異なる内部標準を使って、標準曲線を作製する。この標準曲線が妥当であるかどうかにより、測定結果の品質を判定することができることがわかった。一定の基準に満たないデータを除去することにより、正確なデータセットをつくることができる。この品質管理は非常に重要で、後述するように正確なクラスター分析ができなくなる。

(2) クラスター分析の吟味。

クラスター分析には、階層的クラスター分析の他、自己組織化マップやk平均法などいくつかのアルゴリズムが考案されているが、階層的クラスター分析の一部のものはデータの類似性の距離尺度に関して、いろいろな係数を選べるというメリットがある。しかしながら初期に誤って融合された発現パターンが、そのまま最後のグループ分けまで残ってしまう可能性がある。我々のデータセットに関して、階層的クラスター分析（UPGMA法とWard法）と自己組織化マップを比較検討したが、特に階層的クラスター分析で誤ったクラスタリングが行われたケースは見あたらなかった。自己組織化マップは、あらかじめ分類するグループの数を決定するため、後述する例のように発現パターンの数を類推する目的にはあまり適しない。また、現時点ではmicroarrayやATAC-PCRともに測定値の分析に対する精度が不明確なため、距離尺度を選択できる手法のほうが有利だと考えた。従って、階層的クラスター分析を標準的な手法として選択した。

(3) 小脳皮質発生過程の遺伝子発現パターンのクラスター分析

小脳皮質の生後2日、4日、8日、12日、3週、6週の6点に関して2000個の遺伝子について、成体の大脳を内部標準として相対的発現量を測定した。この発現パターンを、距離尺度としてはピアソンの積率相関係数、アルゴリズムとしてはUPGMAを用いてクラスター分析を行った。その結果約9割の遺伝子は図1に示すように6つの発現パターンに収斂した。3つは、発生の初期に発現量が多く、後の3つは発生の後期ないしは成体（6週）で発現量が多い。発生の後期に発現量の高い遺伝子の中には、発生過程を通じてあまり変動がないが、リボソーム蛋白など発現量の絶対値が高いものの発現量が減少したために、相対的に発現量が上昇しているものが多数含まれている。そのため他のグループよりも属する遺伝子が多い。

クラスター分析に必要なデータの精度を決めるために、内部標準を使わずにエ

レクトロフェログラムのピークの高さから直接相対的発現量を推定した場合は、上記のように6つのクラスターには分かれず、いろいろな複雑なパターンが現れた。また、後述するような機能との連関もわからなくなってしまった。従って少なくとも哺乳類の組織についての時系列データの場合は、マイクロアレイや内部標準を使わないIATAC-PCRのような不正確な測定ではクラスター分析に耐えるデータを収集できないことがわかった。

#### (4) 遺伝子機能と発現パターンを関連づける一般的方法

遺伝子の機能と発現パターンを関連づけるためには、エキスパートによるデータの観察が主であったが、大量のデータを取り扱う場合煩雑であり、また研究者により判定結果がことなる可能性がある。そのため、より客観的な方法が望ましい。出芽酵母の場合には、MIPSによる全遺伝子の機能分類があり、それをもとにして統計的に有意な機能と発現パターンの連関を同定する方法が試みられている。このMIPS分類は単細胞機能に関するものであり、細胞間情報伝達の占める比重が大きい神経組織には使えない。従って、我々が作製したマウス小脳及び海馬の遺伝子に関して機能と発現パターンを結びつけるツールを開発した。ただ、機能と一口にいても酵素活性、生物学的現象、発現する細胞種など多岐にわたるため、一つの遺伝子を一つの機能分類に入れることはせず、遺伝子ごとにその機能を代表するキーワードをつけることにした。キーワードの数は一つの遺伝子あたり4個まで、全部で90種類となった。我々のESTデータベースの既知遺伝子の数は、1600個あまりであり、現在のマウスUniGeneの既知遺伝子約4800個の約3分の1をカバーしている。今回測定した2000個の遺伝子のうち、約1100個が既知遺伝子であった。

キーワードの出現頻度が小脳発生過程の発現パターンによって統計的に有意に異なるものを選択した。11個のキーワードが選択されたが、リボゾーム蛋白、癌関連遺伝子など細胞増殖に関係したものは発生の初期に発現量の多いものに多く、トランスポーター、神経伝達物質受容体など神経細胞の機能に直接関わるものは分化した時期に高発現するグループに頻度が高いことがわかった。これは、それぞれのステージの小脳皮質の状況に見事に一致した。

### 3. 主な研究成果の発表（論文発表）

Matoba, R., Kato, K., Sakakibara, Y., Maruyama, C., and Matsubara, K. Expression profiling of active genes, In Current Options for the Human Genome Project, edited by Fundacion BBV pp.113-124. Fundacion BBV, Bilbao, Spain, 2000.

Matoba, R., Kato, K., Saito, S., Kurooka, C., Maruyama, C., Sakakibara, Y. and Matsubara, K., Gene expression in mouse cerebellum during its development. Gene, 241, 125-131, 2000.