

EST クラスタリングと遺伝子発現データベース

理論構築グループ研究員 長谷川 亜樹

マウスの初期胚における cDNA ライブラリから得られた 25,438 個の EST (Expressed Sequence Tag) から、各発生段階において発現している遺伝子を独自の手法を用いて予測した。その結果、9,718 個の遺伝子が予測され、それらの各遺伝子が発現している発生段階をも特定した。以上の結果を遺伝子発現パターンデータベースとしてまとめ、インターネット上で公開した。

1. はじめに

個体の発生に伴って起こる細胞分裂は、二本鎖 DNA や中心粒における新旧の非対称性を論じる上での重要な鍵を握っている。そこは染色体 DNA や中心粒の非対称な分配が行われる場であり、そのような生命現象を理解するためには、各細胞分裂毎にどのような遺伝子が発現しているかを網羅的に把握することが重要である。本研究はマウス発生初期胚における発現遺伝子を発生段階ごとに特定する解析の一環を担うものとして位置付けられる [1]。理論構築グループでは、ゲノム非対称性グループによって調製されたマウス初期胚 cDNA クローンライブラリより得られた EST 集合を用い、計算機による解析を行った。この解析には発現遺伝子の予測、新規遺伝子の特定、既知の遺伝子に対する注釈づけ、個々の遺伝子の発現している発生段階の特定、これらデータの共有利用を図るためのデータベース構築が含まれる。

cDNA は、遺伝子の転写産物である mRNA に対する相補的な DNA 配列であり、EST は cDNA を部分的に配列決定したものである。そのため、EST は発現している遺伝子を特定するための重要な手がかりであるといえる。ある一つの遺伝子からはその発現量に応じた量の mRNA が転写されており、発現量を反映するように調製された cDNA クローンライブラリからは、その遺伝子に由来する EST がほぼ発現量を反映したかたちで得られるようになってきている。したがって EST 集合には同一遺伝子に由来する EST が複数含まれている場合があり、与えられた EST 集合から発現している遺伝子群を予測するためには、それらを一つにまとめ、他の遺伝子に由来する EST と区別することが重要となる。

ライブラリからは、7つの発生段階 (未受精卵, 受精卵 1細胞期, 2細胞期, 4細胞期, 8細胞期, 16細胞期, 胚盤胞期 (着床直前)) について約 3,000~5,700 個ずつ、合計 25,438 個の 3' 末端 EST が配列決定されており、まずこれらをクラスタリングして発現遺伝子を同定する必要があった。クラスタリング手法は現在までにいくつか提案されているが、当時適切に利用できるものは少なく困難であり、そのため我々は相同性検索プログラム NCBI BLAST2 を利用した独自の手法を考案してクラスタリングを行った。

その結果、同ライブラリに含まれると考えられる遺伝子が 9,718 個存在すると予測され、それらの遺伝子がマウス初期発生のどの段階にどの程度発現しているかという発現パターンが明らかとなった。また、それらの内 4,351 個の遺伝子が全く新規のものであると判明し、1,065 個の遺伝子について注釈づけを行った (1999 年 2 月 8 日時点)。さらにこれらの結果をまとめたデータベースを構築し、インターネット上で公開した。

2. EST クラスタリング

基本的に EST は発現した遺伝子に対応する配列を示す指標配列であり、一般に長く読み誤りの無いものであればクラスタリングを行う上で正確な遺伝子予測が可能となると考えられる。しかしながら、技術的・時間的制約、その他のさまざまな制約から通常はそのようなものは期待できない。そこで、クラスタリングを行う際には、以下のような幾つかの点を考慮して行う必要がある。まず、配列決定の際に一度の読み取りで可能な長さには限界があり、平均して 400~500bp 長くても 700bp 程度であること；読み取り精度の限界から EST には誤って読み取られた塩基が少数ではあるが不特定個含まれる場合があること；また、同一遺伝子由来の EST であってもオルタナティブ・スプライシングを反映した結果、異なるものになっている場合や、反復配列を含むため異なる遺伝子由来の EST の間に高い相同性が見られる場合があることなどである。これらの問題を考慮しつつ相同性検索プログラム BLAST2 を利用した独自の分類評価基準を考案し、これを用いて各発生段階毎に同一遺伝子に由来する EST 集合を予測した。

3. 遺伝子発現データベース

上記の EST 解析の結果を集約し、インターネット上で遺伝子発現データベースとして公開した (<http://www.bioa.jst.go.jp/>)。ここでは、EST クラスタリングに用いられた全ての EST を検索できるだけでなく、染色体上のマーカーに対応付けされた予測遺伝子とその発生段階ごとの発現の程度が検索できるようになっている。さらに、その遺伝子に由来するクローンの ID や EST 配列、DDBJ アクセション番号も含まれており、DDBJ の提供する検索サービスとリンクすることができるようになっている。また、発現パターンから遺伝子を検索する機能の提供も予定している。例えば、7つの発生段階においてとびとびに発現している可能性があるもの(未受精卵, 2細胞期, 8細胞期, 胚盤胞期に発現しており, 他の段階では発現していないもの)などを検索することができる。

4. おわりに

本遺伝子発現データベースの構築は、染色体 DNA 及び中心粒の細胞分裂による分配のたびに生じる細胞分裂というイベントにおいて、遺伝子の発現パターンがどのように変化するかを探る上での貴重な情報源となることを期待するものである。また、各卵割球を識別して染色体 DNA や中心粒の分配を追跡することが可能となれば、さらに卵割球レベルでの非対称な遺伝子発現を明らかにすることが可能となり、土居総括責任者の提唱する生物現象における非対称性に迫る結果が得られるであろうと期待される。

参考文献

- [1] Minoru S. H. Ko, John R. Kitchen, Xiaohong Wang, Tracy A. Threat, Xueqian Wang, Aki Hasegawa, Tong Sun, Marija J. Grahovac, George J. Kargul, Meng K. Lim, YuShun Cui, Yuri Sano, Tetsuya Tanaka, Yuling Liang, Scott Mason, Paul D. Paonessa, Althea D. Sauls, Grace E. DePalma, Rana Sharara, Lucy B. Rowe, Janan Eppig, Chris Morrell, and Hirofumi Doi. Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development. *Development*, Vol. 127, No. 8, pp. 1737-1749, 2000.