

遺伝子発現制御配列と反復配列の進化的変動の分析

●大島 一彦

長浜バイオ大学 バイオサイエンス学部

〈研究の目的と進め方〉

アレイデータの蓄積などにより、遺伝子発現ネットワークの大局的把握が可能になりつつある。遺伝子発現ネットワーク解明のアプローチの一つとして、遺伝子発現制御配列を特定の転写制御因子結合配列の統計的傾向から推定する等、様々な試みがなされている。この際、従来見逃されがちであったのが反復配列の存在である。ゲノムにはマイクロサテライトや転移因子由来の散在性配列など、多様な反復配列が存在しており、ヒトではゲノムの40%以上を占めていることは、よく知られている。従来の分子生物学的研究では、これらの配列の存在意義は不明であり、全く無意味な配列（ジャンクDNA）であるとの推測すらなされていた。ところが近年、これら反復配列の一部が遺伝子発現制御に積極的に関与している事例が続々と報告され、さらに散在性配列（Alu配列）がゲノム大域的に何らかの貢献をしている可能性も示唆されている（International Human Genome Sequencing Consortium 2001）。本研究では、転写制御と反復配列の相関について次のアプローチで取り組むことを目的とした。

- 1) 発現情報と配列情報の相関分析
- 2) 個別の反復配列に着目した種間比較

〈研究開始時の研究計画〉

1) 発現情報と配列情報の相関分析

ヒトを中心に、発現情報と対応遺伝子制御領域（主に5'上流配列）の構造相関を分析する。共通の発現パターンを示す遺伝子組に共通の非コード配列の抽出を行う。その際、反復配列を完全にはマスクせず、反復配列が転写制御に関与する可能性に注意して解析する。このような分析をおこなう際に、国立遺伝学研究所・池村博士のグループが開発した「自己組織化地図(SOM)を用いたゲノム解析法」が適用可能であるか検討をおこないたい。

また、反復配列が転写制御に関与する事例を収集し、共通配列からの変異傾向や発現パターンとの相関を解析する。

2) 個別の反復配列に着目した種間比較、および偽遺伝子群の転写活性喪失の解析

ゲノムには2種類の偽遺伝子（プロセシング済み偽遺伝子とDNA重複型偽遺伝子）が存在するが、DNA重複型では、その生成初期には転写制御配列が元の遺伝子と完全に同一であり、やがて転写制御配列が変異し転写活性を喪失する。そこで、DNA重複型偽遺伝子の転写制御領域の変異傾向を分析し、転写制御配列の抽出の資料とする。

〈研究期間の成果〉

- 1) 2003年度に、パイロット研究として遺伝子の発現と発現制御配列の相関を、反復配列の存在に注目して分析する研究に取り組んだ。公共データベースより、ヒトとマウスの海馬（脳の一部）で特異的に発現していると思われる遺伝子を選別し、各々の転写開始

点上流域の配列を調べた。その結果、上流500～1000塩基の範囲内に存在する反復配列の分布に、ヒトとマウスで異なる傾向が認められた。この結果の有意性や一般性、及びに他組織の状況を検討するため、2004年度にさらに以下の解析を行った。

選別したヒト公開アレイデータを用いて、クラスタリングを行ない、絞り込んだ遺伝子セットに関して、発現情報と遺伝子制御領域構造（5'上流配列）の相関を分析した。その際、反復配列が転写制御に関与する可能性に注意して解析を行なった。その結果、次のような特徴が明らかになった。①組織特異的発現を示すヒト遺伝子群の5'上流500bpにおいて、Low complexity repeats及びSimple repeatsの頻度が顕著に増加している。②発現組織の違いにより、5'上流域における反復配列の分布パターンに差異が認められる。また転写開始点のマッピング・アノテーション精度に留意した解析を再度行ない、上記傾向の再現を確認した。今後、個別の反復配列の種間差異が、遺伝子発現に及ぼす影響を評価するために、マウスの発現情報も統合して解析したい。

- 2) 哺乳類ゲノムには多量の偽遺伝子が存在する。DNA重複型偽遺伝子では、その生成初期には転写制御配列が元の遺伝子と完全に同一であり、やがて転写制御配列が変異し転写活性を喪失する。このため、DNA重複型偽遺伝子の転写制御領域の変異傾向は、機能遺伝子制御配列を抽出する上での重要な資料となる可能性がある。また、プロセシング済み偽遺伝子は一般には転写活性を示さないが、一部に転写産物を生じる（偽）遺伝子も知られている。このような観点も込めて、偽遺伝子に関する研究を進めている。

以前より進めていた、ヒト・プロセシング済み偽遺伝子の全ゲノム解析の成果を2003年末に公表した（公表リスト1）。「プロセシング済み偽遺伝子」は、細胞のmRNAが逆転写反応によりcDNA化し、ゲノムに再挿入されて生じたものである。これらをゲノム全体から網羅的に抽出し生成年代を推定したところ、その分布は一つの頂点を持つ山型を示した。さらにその平均年代は、やはり逆転写反応により増幅するAlu SINE（ヒトゲノムの10%を占める）とほぼ等しいことが明らかになった。この事実は、霊長類進化のある時期ヒトの祖先のゲノムで、プロセシング済み偽遺伝子とAlu SINEが爆発的に増幅した可能性を示唆している。その時期は4000～5000万年前と推定された。

個別の反復配列の分析に関連して、RNA転移因子（レトロポゾン）に関する研究代表者の従来の成果を総括する総説を出版した（公表リスト2）。現在までに蓄積した大量情報の整理を行ない、依然未解明な問題を明確にし、新たな課題を提示した。例えば、完全長データの入手が可能ならば全てのLINE配列を用いた大規模な系統樹を作成し、3'末端配列がSINEと共通なLINEの分布に、偏りがあることを示した。特定のLINEが、頻繁に新奇な

SINEを生成した結果だと思われるが、更なる検証が必要であると思われる。多くの著名なレトロトランスポゾン研究者が寄稿した特集号“Retrotransposable Elements and Genome Evolution”の中の一編である。

〈国内外での成果の位置づけ〉

ヒト・プロセシング済み偽遺伝子の全ゲノム規模の探索に関する成果の公表は、際どいタイミングとなった(公表リスト1)。複数の研究グループが同時期に成果を報告したのである (Ohshima et al. 2003 Oct, Zhang et al. 2003 Dec, Torrents et al. 2003 Dec)。以下は、他グループが翌年出版した総説からの引用である。

“2003 proved to be an exciting year, as three groups (Tokyo, Yale, and EMBL) independently published comprehensive surveys of pseudogenes in the entire human genome” (Zhang & Gerstein 2004)。

また同年 (2003)、偽遺伝子RNAがタンパク質には翻訳されずに、起源遺伝子のmRNAに作用してその安定性を調節する例が発見され (Hirotsume et al., Nature 423, 91-96, 2003)、また

Marques et al.は若い偽遺伝子のかなりの数のものに機能がある可能性を指摘している (Marques et al. PLoS Biology 3, e357, 2005)。このように偽遺伝子解析が急ピッチで進む中、国際的にも一定の評価が得られたと考えている。

〈達成できなかったこと、予想外の困難、その理由〉

「1) 発現情報と配列情報の相関分析」は、反復配列に着目した解析を行う点で、国内外を通じて極めて新奇な取り組みであり、上記のように一定の成果も得られたが、2年の期間内では成果の出版に至らなかった。

〈今後の課題〉

1) 発現情報と配列情報の相関分析：結果の種間比較や、成果の出版を目指したい。2) 個別の反復配列に着目した種間比較：プロセシング済み偽遺伝子に関する全ゲノム・システム研究、および局所的機能解析の双方を推進する。

〈研究期間の全成果公表リスト〉

1. 0404091739

Ohshima, K., Hattori, M., Yada, T., Gojobori, T., Sakaki, Y. and Okada, N. Whole-genome screening indicates a possible burst of formation of processed pseudogenes and Alu repeats by particular L1 subfamilies in ancestral primates. *Genome Biol.* 4:R74 (2003).

2. Ohshima, K., Okada, N. SINEs and LINEs: symbionts of eukaryotic genomes with a common tail. *Cytogenet. Genome Res.* 110:475-490 (2005).