

バイオインフォマティックスの手法を用いたデータマイニング

千葉県産業振興センター (研究員) 三沢 計治

【目的と概要】

ヒトおよびマウスの脳を中心とした領域で発現する mRNA の配列を比較解析し、配列と発現量の関連を詳細に調べ、さらに進んで配列から機能を解明する手がかりとすることを目指す。そのために、まず、mRNA の配列を比較する方法の理論的研究を行い、もっともふさわしい比較解析法を開発した。その後、ヒトおよびマウスの mRNA の配列解析を行い、配列上のどこにどのような違いがあるのか調べた。その結果を発現部位や発現量と比較し、機能が配列へ与える影響を調べることを目指す。

【研究成果の概要と今後の取り組み】

(1) UTR における発現調節因子の探索

ヒトおよびマウスの脳を中心とした領域で発現する mRNA の配列を比較解析した。材料としてはかずさ DNA 研究所との共同研究で得られた KIAA/mKIAA を用いた。解析にあたっては mRNA を2つの領域に分けた。たんぱく質をコードしている領域 (以下 CDS) としていない領域 (UTR) である。まず、UTR について、配列と発現量の関連を詳細に調べた。すると、3'UTR 領域の長さや発現量に相関があることがわかった。この結果は DNA research に発表された (Okazaki et al. 2003)。MEME 法で 3'UTR の中に発現量をコントロールしている配列があるかどうか探索したところ、いくつか候補が見つかった。これからは、これらの配列が本当に発現量に影響を与えているのか、遺伝学的・分子生物学的解析が行われる予定である。

(2) CDS と発現量の関係

CDS について、ヒトおよびマウスの mRNA の発現量との比較を行った。ひとつのアミノ酸を code する DNA の3つ組 (codon) は多くの場合複数あり、同じアミノ酸配列を作る遺伝子でも DNA 配列にわずかな違いが許されていて、どの codon を使うか、頻度が遺伝子ごとに違っている。先行研究から、大腸菌やショウジョウバエなどの実験用モデル生物では、発現量と codon の使用頻度に相関があることが知られている。なぜならば、使用される codon によって、RNA 配列からアミノ酸配列へと翻訳される際の効率が違うためである。大腸菌やショウジョウバエなどでは、このことを逆に利用し、配列から発現量を推定することが行われることもある。しかしながら、発現と codon 頻度の相関は、ヒトやマウスでは見つかっていなかった。そこでわれわれは、codon 頻度と発現量の相関に注目して解析を行った。しかしながら、先行研究と同じく、相関は見られなかった。

相関が出てこない理由のひとつとして、突然変異の偏りがあると考えられる。ヒトやマウスの codon 頻度は突然変異の偏りに影響を受けている。われわれは、特に CpG hypermutability に注目した。CpG hypermutability とは、DNA 配列中の CG の C が特異的にメチル化され、その methyl-C が脱アミノ化して T へと変化することによって、C から T への突然変異が高い率で生じる現象である。このタイプの突然変異は、隣の塩基の影響を受け、時間軸に対称でないという特徴がある。我々は、ヒト・チンパンジー・マウスのコード領域を比較し、コドン置換速度を調べた所、CpG hypermutability が引き起こす置換速度は他のものよりも約 10 倍大きいことがわかった。また、全体のコドン置換のうち、およそ2割が CpG hypermutability に引き起こされていた。CpG hypermutability のために、codon の中でも C と G が並んでいるようなものは数が少ない。CpG hypermutability を考慮に入れた上で、codon 頻度と発現量の相関を調べたところ、優位な相関が見られた。このことから2つのことがわかった。(1) ヒトやマウスでも、codon によって発現の効率が違う。(2) しかしその影響は、突然変異率の影響よりも小さい。先行研究で相関が見つかっていなかったのは突然変異率を考慮に入れていなかったためらしい。このことを利用すると、突然変異率の大きさと比較することで、codon 頻度が翻訳効率に及ぼす影響を定量化することができる。

(3) ヒトとマウスの系統で観測されるアミノ酸変化のモデル化

この研究の副産物として、たんぱく質を構成するアミノ酸が、長い時間の中で少しずつ置換されていく傾向を平均化し、モデル化することができた。これもヒトとマウスの間での突然変異率を考慮に入れることで可能となった。このモデルを使うと、ヒトまたはマウスに至る系列での変化が、平均から外れている遺伝子を統計処理で探し出すことができる。ヒトが進化した過程で変化した遺伝子は、ヒトの特徴を作り出している遺伝子の候補である。このような遺伝子は幾つか見つかったが、残念ながら、そのうちの一部

は、最初からたんぱく質をコードしていない、または途中でたんぱく質をコードしなくなった遺伝子（偽遺伝子）である。とはいえ、この結果は、従来の遺伝子推定法に誤りが含まれ、今回開発した方法が、たんぱく質をコードしている遺伝子だけを探し出すことに応用できることを示している。