

赤池情報量規準を用いた遺伝子発現解析

組織特異的遺伝子とは、ある組織で特異的に発現し他の組織では(少量しか)発現しない遺伝子を指す。それらは、我々の体を形成している様々な器官(臓器)を特徴づける鍵となる役割を果たしていること、また腫瘍マーカーとしての役割を果たす可能性から様々な手法による同定が試みられている。

何千もの遺伝子発現データを一度の実験で観測できるDNAチップの利用は、多くの正常臓器組織に対して蓄積されたデータをもとに、計算機的手法によって様々な目的組織特異的発現パターンを示す遺伝子をスクリーニングする試みを可能にした。しかしながら従来法では、信頼度を変えることによる結果(遺伝子数)の違いや、空間的に近い別組織での高発現遺伝子の過剰検出などの問題が残されている。当研究センターでは、赤池情報量規準(AIC)を適用した、ある遺伝子の発現プロファイル中で特異的に高(低)発現している値を“外れ値”として検出する新しい方法¹⁾を見出した。

AICとは、統計モデルの良さを評価するための手法の一つである。AICはモデルの悪さの指標であり、様々なモデルの中から最小AICのものを最適モデルとする。ここでは、ある遺伝子の様々な組織での発現データ(観測

値)に対して、他の測定値から飛離れた組織の観測値を上位何組織までの値を外れ値とするかという様々なモデルに対するAICを計算する。そしてそれらを比較することで、最小AIC値の組合せモデルから得られる“外れ値”に対応する組織を“特異的発現”組織として検出することに応用している。

本手法の主な長所としては、外れ値として検出されるほど飛離れた観測値の組織だけが検出されるため、空間的および発現プロファイル的に近い別の組織の存在²⁾による影響を受けにくい(図)ことや、最小AIC値に対応する外れ値の組合せが結果として返されるので、一つの遺伝子の様々な組織における発現プロファイルに対して、特異的発現がない(外れ値がない)場合を含んだ複数外れ値の場合にも同時対応可能であることが挙げられる。

我々は、本手法を癌と正常サンプルで発現が異なる疾患感受性遺伝子の同定に悪影響を与える“外れサンプル”の検出にも応用し、検出された外れサンプルの有無による候補遺伝子群や、分類器として用いた場合の分類精度に有意な違いが見られた³⁾。このことから、本手法は今後も様々な生命情報科学研究分野での応用が期待される。



図 大脳と眼球で特異的に高発現している遺伝子の抽出例

AIC(上)は目的組織でのみ高発現している4つの遺伝子を検出しているが、既存の方法(下、パターンマッチング法)では、検出された上位4つの遺伝子のうち3つが目的組織と空間的に近い小脳でも高発現している遺伝子抽出していることが分かる。

関連情報

- 共著者: 高橋勝利(生命情報科学研究センター)
- 1) K. Kadota, S. Nishimura, H. Bono, S. Nakamura, Y. Hayashizaki, Y. Okazaki, and K. Takahashi : *Physiol. Genomics*, 12(3), 251-259 (2003).
- 2) K. Kadota, R. Miki, H. Bono, K. Shimizu, Y. Okazaki and Y. Hayashizaki : *Physiol. genomics*, (3), 183-188 (2001).
- 3) K. Kadota, D. Tominaga, Y. Akiyama and K. Takahashi : *Chem-Bio Informatics J.*, 3(1), 30-45 (2003).



かどたこうじ
門田幸二
koji-kadota@aist.go.jp
生命情報科学研究センター