

真核生物ゲノム間での制御領域の比較解析支援ツール SHOE の開発

ナターリヤ・ポリュリヤーフ
(配列解析チーム 産総研特別研究員)



地球上には数千万種とも言われる多種多様な生命が存在しています。それらは種として近縁種であるほど共通の遺伝子をたくさん(ヒトとマウス間ではほぼ半分)持っているのにも関わらず、なぜわれわれの体(表現型)がこんなに違うのかは大きな謎です。その秘密は特定の遺伝子群の発現パターンの違いに隠されていると思われる。高等生物間にもっとも大きな制御の変化を与えているのは脳の働きに関わる遺伝子とその転写因子です。ヒトのタンパク質遺伝子の1割を占めている転写因子が、生命の謎を解くためのカギを握っていると言えるでしょう。

ゲノムの制御領域を比較すると、進化の過程で同じ役割を保ってきた遺伝子群と役割が変化した遺伝子がわかります。しかしこのとき、役割の変化は、遺伝子そのものの機能変化より、転写の制御により役割が変わることが多いことが最近わかってきました。

我々は 7,500 万年の進化距離を持つヒト・マウス・ラットの制御領域の比較解析ツール "SHOE - Sequence HOmology in higher Eukaryotes"(<http://shoe.cbrc.jp>) (図 1) を開発しています。真核生物の転写制御要素は転写開始点付近にあるプロモーター以外にも数千塩基に離れているエンハンサーがあって、それに結合するアクティベーターやインヒビター(図 2) による制御の調節が行われています。これを考慮し、SHOE では 10 万塩基長までの領域を扱えるようにしました。今までのツールと比べて、転写因子の認識するパターンを表す "PSSM (位置依存スコア行列)" のみならず、共発現と進化保存度の情報をあわせて精度の高い予測ができることが特徴として挙げられます。

共発現頻度の高いヒトの遺伝子ペアについては、ヒト・マウス・ラット間によく保存されている領域を見つけ、共通の転写因子の結合部位(モチーフ)の検索を行います。更に保存度の高い共通のモチーフを持っている遺伝子と、モチーフに結合すると思われる転写因子との共発現を確認します。

発現データは CBRC が公開しているデータベース Cell Montage、モチーフ検出には TRANSFAC や JASPAR の PSSM を使っていますが、将来的に新規モチーフ発見機能を付け加える予定です。

参考文献:

- Polouliakh, N., Natsume, T., Harada, H., Fujibuchi, W., Horton, P.: "Comparative genomic analysis of transcription regulation elements involved in the human MAP-kinase G-protein coupling pathway", *Journal of Bioinformatics and Computational Biology*, **4** (2) (2006).
- Suzuki, Y., Yamashita, R., Shirota, M., Sakakibara, Y., Chiba, J., Mizushima-Sugano, J., Nakai, K., Sugano, S.: "Sequence comparison of human and mouse genes reveals a homologous block structure in the promoter regions", *Genome Res.*, **14**, pp.1711-1718 (2004).

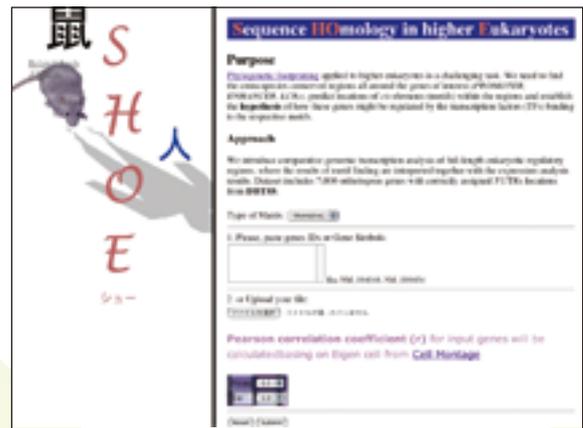


図 1: SHOE のトップページ

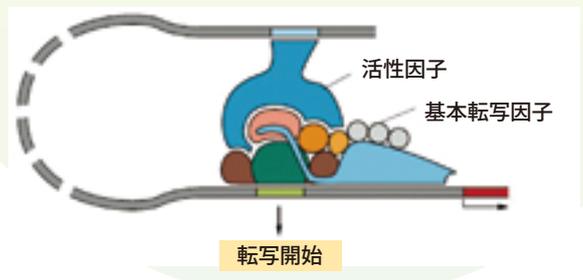


図 2: 真核生物の転写制御メカニズム
(<http://www.csu.edu.au/> から改変)