

SOKOSを用いたヒトゲノム中の機能性RNA部位を予測

# 開発 ヒトゲノムの核小体低分子RNA部位予測手法の

近年の実験機器・技術の向上により、これまでに検出が困難であった細胞内の小さな分子(低分子)の存在やその構造・機能が明らかにされてきている。これらの低分子は機能性RNA (fRNA: functional RNA)と呼ばれ、 mRNA や rRNA の成熟に関与しており、その機能や発現時期、部位、また遺伝子を特定することはポストゲノム研究の一つとして注目されている。

その一つである核小体低分子RNA (snoRNA: small nucleolar RNA)は、rRNA の成熟に関与する低分子RNAであり、リボースのメチル化(box C/D 型)やウリジン残基の擬ウリジル化(box H/ACA 型)を手助けする役割を果たしている(図1)。また、酵母を除く真核生物の snoRNA の遺伝子は他の遺伝子(宿主遺伝子)のイントロンに内在し、また酵母では独立に存在することから、生物の進化を考える上でも非常に興味のある物質でもある。snoRNA は塩基数が数十~数百塩基と非常に短く、また保存されているモチーフも非常に少ないため、従来のBLASTなど配列の類似性を用いた手法では検出が困難である。

また、酵母を除く真核生物においては、ゲノム中のイントロン(介入配列)に内在しており、従来の遺伝子予測手法で見つけることも困難である。

そこで我々は、GeneDecoder により予測されたヒトゲノムのイントロン配列と、確率文脈自由文法(SCFG: Stochastic Context Free Grammars)を用いた手法(SOKOS: a software for kernel computation over scfg)により snoRNA 部位の予測を行っている。SOKOS は、与えられた配列から予測できるあらゆる高次構造を考慮しながら各配列の類似度を比較する手法である。この手法は snoRNA のような二次構造の保存性は高いが、共通モチーフが少ないというような配列を分類することに適しており、従来の配列の類似性を用いた手法では検出が困難であった配列を検出することができると思われる(図2)。現在、我々は予測されたヒトゲノムのイントロン配列から snoRNA と予測される候補配列を抽出しており、今後、このようなタイプの fRNA 部位の予測を行うことができると期待される。

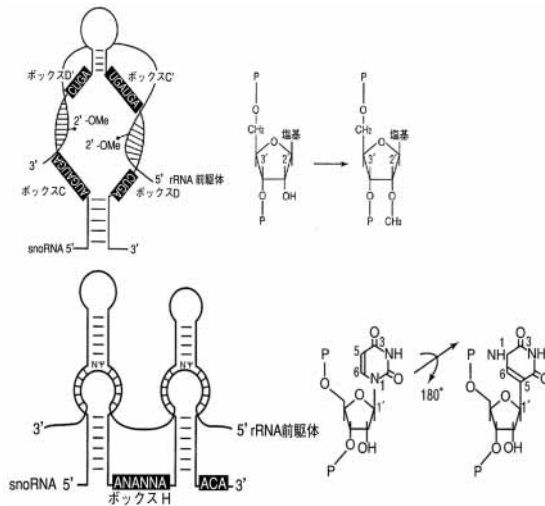


図1 (上) Box C/D型 snoRNA (下) Box H/ACA型 snoRNAの二次構造

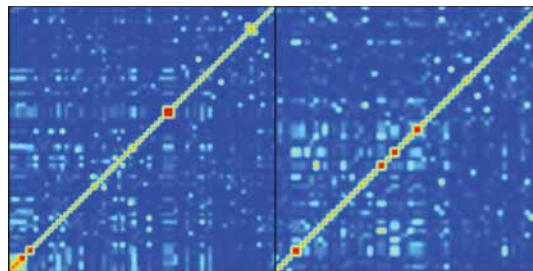


図2 SOKOSを用いたBox C/D型 snoRNAの分類テスト

左: Box C/D型 snoRNA の配列39本によるトレーニングデータを用いた分類。右: Box C/D型 snoRNA の配列50本によるトレーニングデータを用いた分類。トレーニングデータの本数を増やすことにより判定の精度が向上する。



さがらじゅんいち  
相良純一  
jun@neptune.kanazawa-it.ac.jp  
生命情報科学研究センター\*

関連情報

- 共同研究者: 中村周吾(東京大学), 剣持直哉(宮崎大学)。
- J. Sagara, S. Nakamura, N. Kenmochi, K. Asai: RNA, Vol. 1, 2003, 165 (2003).
- T. Kin, K. Tsuda, K. Asai: Genome Informatics, Vol. 13, 112-122 (2002).
- ※ (現) 金沢工業大学ゲノム生物学研究所