

# 構造モデルに基づく塩基配列からの boxC/D型 snoRNA 遺伝子検出法

山森 一人<sup>1,a)</sup> 薛 晨<sup>2</sup> 岩切 淳一<sup>3</sup> 剣持 直哉<sup>4</sup> 吉原 郁夫<sup>5</sup>

受付日 2012年8月20日, 再受付日 2012年10月9日,  
採録日 2012年10月24日

**概要:** 本論文では, boxC/D 型 snoRNA の一次構造, および二次構造を考慮した構造モデルに基づき, 同遺伝子を高精度に検出する手法を提案する. 機能性 RNA の 1 つである boxC/D 型 snoRNA は真核生物の核小体内に存在し, リボソーム RNA などのメチル化などに関与すると考えられている. RNA 修飾の意義はよく分かっておらず, 機能性 RNA の役割を明らかにするためにも, より多くの機能性 RNA 遺伝子を発見することが望まれている. 提案手法では, boxC/D 型 snoRNA が持つ特徴的な塩基配列であるボックスと, ステム構造と呼ばれる相補塩基対で構成された二次構造の双方に着目し, ボックス存在位置などにモデルに基づく制約を課することで高精度に snoRNA 遺伝子を検出する. 実験により, 提案手法は 97.5% の精度で boxC/D 型 snoRNA 遺伝子を検出できることを示した.

**キーワード:** boxC/D 型 snoRNA, 構造モデル, ボックス構造, ステム構造

## A Structure Model Based Method to Detect boxC/D Type snoRNA Genes from DNA Sequences

KUNIHITO YAMAMORI<sup>1,a)</sup> XUE CHEN<sup>2</sup> JUNICHI IWAKIRI<sup>3</sup> NAOYA KENMOCHI<sup>4</sup>  
IKUO YOSHIHARA<sup>5</sup>

Received: August 20, 2012, Revised: October 9, 2012,  
Accepted: October 24, 2012

**Abstract:** This paper proposes a method to detect boxC/D type snoRNA genes using structure model that consisting with characteristic base sequences and secondary structure. The boxC/D type snoRNAs exist in nucleolus of eukaryote, and they may work methylation of rRNA, but the function of such RNA modification remains unknown. Our method utilizes both characteristic sequences called as boxes and secondary structure called as stems. We develop a structure model of boxC/D type snoRNA from actual sequences, and also determine a sequence include a boxC/D type snoRNA gene or not based on the matching ratio of the sequence and the model. Our method achieves 97.5% detection ratio.

**Keywords:** boxC/D type snoRNA, structure model, box structure, stem structure

<sup>1</sup> 宮崎大学工学教育研究部  
Faculty of Engineering, University of Miyazaki, Miyazaki  
889-2192, Japan  
<sup>2</sup> ソフトバンクモバイル株式会社  
SOFTBANK MOBILE CORP., Minato, Tokyo 105-7317,  
Japan  
<sup>3</sup> 東京大学大学院新領域創成科学研究科情報生命科学専攻  
Department of Computational Biology, Graduate School of  
Frontier Science, University of Tokyo, Kashiwa, Chiba 277-  
8561, Japan  
<sup>4</sup> 宮崎大学フロンティア科学実験総合センター  
Frontier Science Research Center, University of Miyazaki,  
Miyazaki 889-2192, Japan

### 1. はじめに

ヒト DNA 上の塩基配列をすべて読み取るヒトゲノム計画 [1], [2] が 2003 年に完了し, 得られた塩基配列と生命活動の関わり, すなわち, 遺伝子がどのように発現し, 生命活動をどのように制御しているのか, 多くの科学者が研究を行っている. ゲノムはたんぱく質のアミノ酸配列をコー

<sup>5</sup> 宮崎大学  
University of Miyazaki, Miyazaki 889-2192, Japan  
a) yamamori@cs.miyazaki-u.ac.jp

ドするコーディング領域と、それ以外のいわゆるノンコーディング領域に大別される。ゲノム配列解読当初、ノンコーディング領域についてはその一部が遺伝子発現調節などに関与することが知られていたものの、大部分は意味を持たないものと考えられ、ジャンク DNA とも呼ばれていた。現在では、遺伝子発現調節のほか、たんぱく質の翻訳情報を持たないノンコーディング RNA 遺伝子など、生体機能に必須な情報がこの領域に多く含まれることが明らかにされつつある [3], [4]。

21 世紀に入り、ノンコーディング RNA の 1 つである機能性 RNA を発見し、生命活動における役割を明らかにすることは、分子細胞生物学やバイオフィーマティクス双方において最も重要な研究課題の 1 つになっている [5], [6], [7]。機能性 RNA と疾患との関わりに関する研究成果も次々に報告されており [8], [9]、創薬や再生医療分野などで大きな進展をもたらすことが期待されている。しかし、塩基配列の量は膨大で、医学的な検出方法では多くの手間と時間がかかる。そのため、コンピュータによる機能性 RNA 遺伝子の自動検出法の開発が必要とされている。

snoRNA は small nucleolar RNA の略称で日本語では核小体低分子 RNA と呼ばれ、酵母からヒトまで広く存在している。これまでに、ヒトで 236 種類、酵母で 75 種類の snoRNA が見つかり [9]、リボソーム RNA (rRNA) 前駆体の修飾やプロセッシングなどに関与すると考えられている。しかし、RNA 修飾の意義などは十分解明されておらず、ゲノム研究のいっそうの推進のためにも、より多くの snoRNA 遺伝子の検出が期待されている。

本論文では、機能性 RNA の 1 つである核小体低分子 RNA (snoRNA) [10] を対象とし、構造モデルとの適合度に基づき snoRNA 遺伝子を自動検出する方法を提案する。提案手法は、snoRNA に含まれる特徴的な塩基配列とステム構造の存在確率を求め、構造モデルとの一致度を評価することで、boxC/D 型 snoRNA 遺伝子を高精度に検出する。

本論文の構成は以下のとおりである。2 章で snoRNA について紹介する。3 章では、boxC/D 型 snoRNA 遺伝子の構造モデル、およびボックス存在確率とステム存在確率の導出方法、さらに boxC/D 型 snoRNA 遺伝子有無の判定に用いるしきい値の決定法について述べる。4 章では、提案手法による boxC/D 型 snoRNA 遺伝子の検出精度について議論する。5 章は本論文のまとめである。

## 2. snoRNA とその従来検出法

### 2.1 snoRNA の構造と機能

snoRNA はその構造上の特徴から boxC/D 型と boxH/ACA 型の 2 つに分類される。図 1 と図 2 に boxC/D 型、boxH/ACA 型 snoRNA の二次構造をそれぞれ示す。rRNA、またはその他の RNA に対し、boxC/D 型 snoRNA はメチル化、boxH/ACA 型 snoRNA はシュドウリジン化

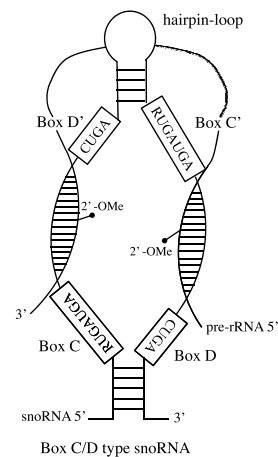


図 1 boxC/D 型 snoRNA の二次構造

Fig. 1 Secondary structure of boxC/D type snoRNA.

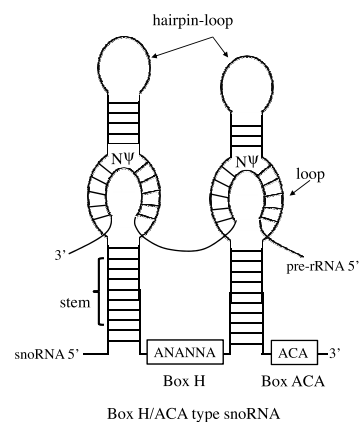


図 2 boxH/ACA 型 snoRNA の二次構造

Fig. 2 Secondary structure of boxH/ACA type snoRNA.

などに関与している。本論文では、これら 2 種類の snoRNA 遺伝子のうち、boxC/D 型を対象とする。

boxC/D 型 snoRNA はおおむね 60~100 塩基の長さで、両末端の塩基がワトソン-クリックの相補塩基対 (A-U, C-G) で構成されるステム構造を持ち、これに隣接して、特徴的な配列 boxC と boxD がある。また、ヘアピンループ直下にもステム構造を持ち、このステム構造に隣接してもう 1 セット特徴的な配列 (boxC' と boxD') を持つ場合もある。boxC の 3' 側と boxD の 5' 側に、標的となる rRNA と相補的な配列があり、rRNA のメチル化はこの領域で起こる。

snoRNA による rRNA 修飾の意義は今のところ不明であるが、修飾の異常、または snoRNA そのものが関わっている可能性のある疾患として、先天性角化不全症、B 細胞悪性リンパ腫、プラダー・ウィリー症候群などが報告されている [11]。

### 2.2 snoRNA 検出の従来法

塩基配列が与えられたとき、転移 RNA (tRNA) 遺伝子や rRNA 遺伝子、snoRNA 遺伝子などが含まれるか否かを

判別する様々な手法が考案されている。その手法の1つとして、サポートベクタマシン (Support Vector Machine, SVM) [12] とカーネル関数 [13] を組み合わせた手法が近年注目されている。

SVM は入力サンプルを2つのクラスに分類する手法であり、学習サンプルから最適な識別境界を決定する。与えられたサンプル集合が線形分離可能な場合と不可能な場合の両方に適応可能であり、線形分離不可能な場合は識別関数にカーネル関数を適用して特徴空間を写像し、識別関数を線形化することで適切な分類結果を得ることができる。

カーネル関数は配列や木構造、グラフ、画像などといった非数値的なデータも扱うことができ、バイオインフォマティクス解析にも応用されている [13]。本論文で比較対象としたカーネル関数は、文字列部分列カーネル (String Subsequence Kernel, SSK) [14]、ステムカーネル (Stem Kernel, StK) [15] とその改良法である DAG (Directed Acyclic Graphs) based stem kernel (DAG-based StK) [16]、BPLA (Base-pairing Profile Local Alignment) kernel [17] である。

文字列部分列カーネル (SSK) は文字列に対するカーネルであり、文字配列の類似度を求めるものである。SSK では、文字列  $x$  と  $y$  の類似度について、連続、または不連続の部分配列  $k$  の出現頻度をカウントして評価する。SSK での snoRNA 遺伝子検出精度は 65% 程度 [14] であり、十分な精度とはいえない。

ステムカーネル (StK) は RNA 配列に特化したカーネルであり、RNA の二次構造を考慮した特徴空間をとり、任意の長さの連続、または不連続のステム構造候補の出現頻度を数えて特徴ベクトルとする。すなわち、ステム構造の間にギャップが挿入される場合も含めて、すべての可能なステム構造の候補をカウントする。boxC/D 型 snoRNA 遺伝子の StK での検出精度は 64% 程度 [15] であるが、処理速度と特異度を大幅に向上させた DAG-based StK、さらなる性能向上を果たした BPLA kernel と改良が続けられており、BPLA kernel を用いた SVM で boxC/D 型 snoRNA 遺伝子を含まないサンプルを正しく判別できる確率である特異度が 0.92~1.0、boxC/D 型 snoRNA 遺伝子を含むサンプルを正しく判別できる確率である感度が 0.71 と報告されており [17]、感度の点で改良の余地がある。

Yang ら [18] は boxC/D 型、boxH/ACA 型 snoRNA 遺伝子を検出する一連のソフトウェア snoSeeker を開発している。このうち、boxC/D 型 snoRNA 遺伝子を検出する CDseeker によりヒトのゲノムから 212 個の boxC/D 型遺伝子候補を検出しており、そのうち 191 個が既知の boxC/D 型 snoRNA 遺伝子であり、標的不明の boxC/D 型 snoRNA 遺伝子 5 個、既知の snoRNA 遺伝子のアイソフォーム型 2 個を含む残る 21 個が新規の snoRNA 遺伝子候補と報告している。CDseeker は隠れマルコフモデルを用いたボックス、および標的 RNA との相補領域の検出と、確率文脈

自由文法を用いた末端のステム構造の検出を組み合わせるスコアを導出し、このスコアが一定値以上を示した場合に boxC/D 型 snoRNA 遺伝子と判定する。

Hertel ら [19] は、最小自由エネルギー値やステムを構成する相補塩基対数などを入力とする SVM ベースの snoRNA 遺伝子検出法を提案し、これを SnoReport として公開している。

### 3. 構造モデルに基づく boxC/D 型 snoRNA 検出法

#### 3.1 構造モデルの設計

提案手法では、ボックスと呼ばれる特徴的塩基配列と、ステムと呼ばれる二次構造の両方に着目し、boxC/D 型 snoRNA 遺伝子の検出を行う。そこで、既知の boxC/D 型 snoRNA 遺伝子を調べ、どの位置にどのような塩基が表れるかを定めた構造モデルを作成し、このモデルと一致する箇所に snoRNA 遺伝子が存在すると判定する。そのため、文献 [9]、[20]、および snoRNA 遺伝子データベース “snOPY” [21] から塩基配列を調べる。

boxC/C' は “RUGAUGA” (R は A, または G) の 7 塩基で構成されており、boxD/D' は “CUGA” の 4 塩基で構成されている。boxC' と boxD' は、boxC/D 型 snoRNA 遺伝子内に存在しない場合もある。snoRNA の二次構造を平面上に展開すると、図 3 のようになる。図 3 に示すとおり、boxC と boxD の両末端に stem-1 が構成され、boxC' と boxD' 間に stem-2 が構成される。stem-1 と stem-2 は最短 1 ペアの相補塩基対を構成し、相補塩基対が多いほど snoRNA 遺伝子の構造が安定する。図 3 の黒い長方形部分には標的となる rRNA と相補的な配列があり、9~20 塩基からなる。ヘアピンループに関しては、RNA 二次構造の予測ツール [CentroidFold] [22] を使って、“snOPY” から取り出した 1,082 個の boxC/D 型 snoRNA データに対する二次構造を予測すると、その塩基数は最小 4 塩基であることが分かった。

これらの結果から、図 4 に示す構造モデルを作成した。図 4 の  $\frac{n}{m}$  は  $m$  塩基中  $n$  塩基以上が一致する必要があることを表し、 $n \sim$  は当該箇所が  $n$  塩基以上で構成されている

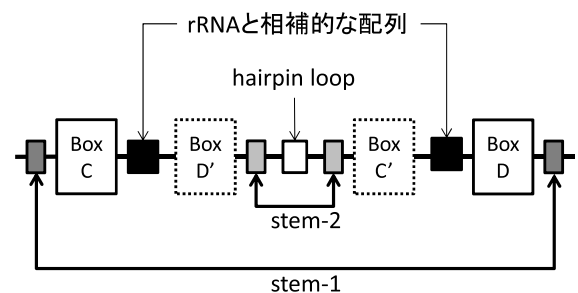


図 3 boxC/D 型 snoRNA の二次構造の展開図  
Fig. 3 Development of the secondary structure of boxC/D type snoRNA.

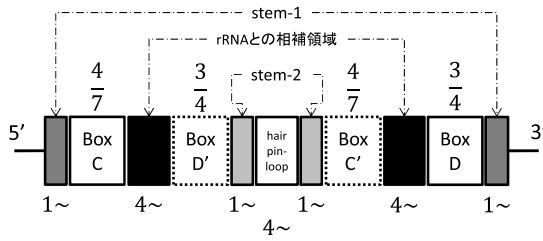


図4 boxC/D型 snoRNA 遺伝子の構造モデル

Fig. 4 Proposed structure model for boxC/D type snoRNA.

ことを表す。

まず、5' 端から説明する。5' 端には 3' 端とペアになる stem-1 が存在する。したがって、少なくとも 1 つの相補対があるものとし、boxC は 5' 端から 1 塩基以上離れているものとする。次に boxC であるが、5' 端に最も近く、かつ boxC と最も多く塩基が一致する箇所に定める。このとき、少なくとも boxC を構成する 7 塩基の半分以上、すなわち  $\frac{4}{7}$  が一致する必要があるとする。boxC に続き、標的 rRNA との相補的な塩基配列がある。先述のとおり、この領域は 9~20 塩基、すなわち片側で平均 4.5~10 塩基からなる。そこで、標的 rRNA と相補的な配列は片側 4 塩基以上からなるものとした。次の boxD' は、boxC 端から標的 rRNA との相補配列分の 4 塩基以上離れ、かつ boxD' を構成する 4 塩基の半分以上、すなわち  $\frac{3}{4}$  の塩基が一致する箇所に定める。boxD' 候補が複数ある場合については後に述べる。boxD' の後には stem-2 がある。stem-1 と同じく、ここにも 1 つ以上の相補対があるものとする。stem-2 の次に、構造の折り返しとなるヘアピンループがある。「CentroidFold」による予測結果から、この領域は 4 塩基以上からなるものとした。以下、1 塩基以上からなる stem-2、7 塩基中 4 塩基以上が一致している boxC' 領域、標的 rRNA との相補的な 4 塩基以上の領域、4 塩基中 3 塩基以上が一致する boxD' 領域、1 塩基以上の stem-1、と続く。

以上が、本研究で用いる boxC/D 型 snoRNA 遺伝子構造モデルである。

### 3.2 boxC/D 型 snoRNA の存在判定法

snoRNA 遺伝子の有無を判断する流れについて説明する。まず、対象とする塩基配列に各ボックスが含まれる平均確率  $P_B$  を求める。次に、各ボックスが一定以上の確率で含まれる塩基配列において、相補対により形成されるステム構造が存在する確率  $P_S$  を計算する。式 (1) に示すとおり、 $P_B$  と  $P_S$  の重み付き和がしきい値  $P_{th}$  を超える場合、対象とする塩基配列に snoRNA 遺伝子が含まれると判断する。

$$P_{th} = \alpha P_S + \beta P_B, \quad (1)$$

$$\alpha + \beta = 1.0.$$

ここで、 $\alpha, \beta (\geq 0)$  はそれぞれ、 $P_S, P_B$  の寄与を調整

するパラメータである。

3.3 節と 3.4 節では、ボックス存在確率  $P_B$  とステム存在確率  $P_S$  の導出方法について説明する。

### 3.3 ボックス存在確率の導出

#### 3.3.1 ボックス存在確率の定義

boxC/D 型 snoRNA 遺伝子におけるボックス存在確率の導出法について説明する。boxC/D 型 snoRNA は boxC/D を持つが、boxC'/D' は持たない場合もある。そこで、各ボックスの存在確率をそれぞれ計算し、その平均をボックス存在確率  $P_B$  とし、式 (2) で定義する。

$$P_B = \frac{\sum \text{各ボックスの存在確率}}{\text{検出されたボックスの数}} \quad (2)$$

以下、各ボックスの存在確率の計算方法について詳しく述べる。

#### 3.3.2 boxC の探索

boxC の探索では、7 塩基からなる比較領域を、5' 端から 3' 端に向かって 1 塩基ずつずらしつつ、boxC 存在確率  $P_B^C$  を計算する。boxC 存在確率とは、ボックスを構成する塩基配列と対象配列が一致する割合であり、式 (3) で定義する。

$$P_B^C = \frac{1}{N} \sum_{i=1}^N d(C_i, T_i), \quad (3)$$

$$d(C_i, T_i) = \begin{cases} 1 & C_i = T_i, \\ 0 & C_i \neq T_i. \end{cases}$$

式 (3) において、 $N$  は比較する塩基数、 $C_i, T_i$  はそれぞれ対象ボックスと比較領域の  $i$  番目の塩基を表す。各比較領域での  $P_B^C$  を比較し、最も大きい  $P_B^C$  を持つ比較領域を boxC と判断する。同じ  $P_B^C$  を持つ領域が複数あった場合、塩基配列の 5' 側に最も近い領域を採用する。これは、1 塩基ずつずらしながら  $P_B^C$  を求めていくことから、現在 5' 端から遠い boxC 候補領域もいずれ最も 5' 端に近い領域として選択されるからである。

boxC 領域は boxC/D 型 snoRNA 遺伝子に必ず含まれている。そこで、boxC を構成する 7 塩基のうち、boxC 候補領域は少なくともこの半分以上の塩基が一致する必要があるとする。すなわち、boxC 存在確率が  $\frac{4}{7}$  未満の場合は、当該配列は boxC/D 型 snoRNA 遺伝子を含まないと判断する。

#### 3.3.3 boxD の探索

boxD の探索では、先に決定した boxC の末端から 3' 端まで順に、4 塩基からなる比較領域を 1 塩基ずつずらしつつ、boxD の存在確率  $P_B^D$  を計算する。 $P_B^D$  も式 (3) により boxC と同様求める。なお、boxD は 4 塩基からなるので、式 (3) 中の  $N$  は 4 である。各比較領域での  $P_B^D$  を比較し、最も大きい  $P_B^D$  を持つ比較領域を boxD 領域とする。このとき、同じ  $P_B^D$  を持つ領域が複数ある場合が考えられる。

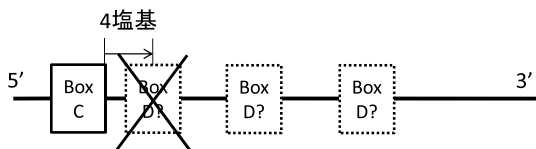


図 5 標的 rRNA との相補的塩基配列制約による boxD 候補の削除  
 Fig. 5 Elimination of boxD candidates by the rule of complementary base pairs for target rRNAs.

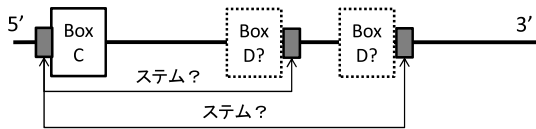


図 6 boxD を決定するための boxC/D 両末端ステム構成のチェック  
 Fig. 6 Stem structure checking to decide the position of boxD.

この場合、まず図 4 の「rRNA と相補的な配列」の塩基数制限を用いて、図 5 のように BoxC から 3' 末端 4 塩基以内にある boxD 候補を除外する。それでもなお複数の boxD 候補が残った場合、図 6 に示すように boxC/D 両末端には stem-1 が存在することから、各 boxD 候補について計算した stem-1 存在確率が最大となる boxD 候補を boxD として採用する。stem-1 存在確率については 3.4.2 項で詳しく説明する。それでもなお複数の boxD 候補が残った場合、3' 端に最も近い boxD 候補を採用する。これは、boxC/D 間の塩基数が多いほど、次に探索する boxC', および boxD' が存在する可能性が高くなり、中央部の stem-2 やヘアピンループを構成できる可能性が高くなるためである。

boxD は“CUGA”の 4 塩基からなる。boxC と同じく、対象配列の過半数、すなわち、boxD 存在確率が  $\frac{3}{4}$  未満の場合、対象配列は snoRNA 遺伝子を含まないものとして扱う。

### 3.3.4 boxC' と boxD' の探索

boxC と boxD の位置を決定後、boxC の 3' 端から boxD の 5' 端まで順に 1 塩基ずつ比較領域をずらしつつ、boxC'/D' の存在確率  $P_B^{C'}$ ,  $P_B^{D'}$  を計算する。存在確率の計算には、boxC/D と同じく式 (3) を用いる。なお、boxC, boxD は各ボックスと最もよく一致する領域に定めているので、その中間である探索範囲に boxC や boxD とよりよく一致する領域は存在しない。

求めた  $P_B^{C'}$ ,  $P_B^{D'}$  を比較し、最も  $P_B^{C'}$ ,  $P_B^{D'}$  が大きい比較領域を boxC'/D' と判断する。同じ  $P_B^{C'}$ ,  $P_B^{D'}$  を持つ boxC'/D' 候補が複数ある場合、以下の条件にあてはまる候補を除外する。

- 図 5 と同様に「rRNA と相補的な配列」の塩基数制限を用いて、boxC や boxD に近すぎる候補を除外する。
- boxC' は boxC と boxD の中間より 3' 側、boxD' は 5' 側に存在する事実に基づき、図 7 に示すように、本来存在しえない位置にある boxC'/D' 候補を除外する。
- 図 8 に示すように、boxC'/D' の間に最小 4 塩基から

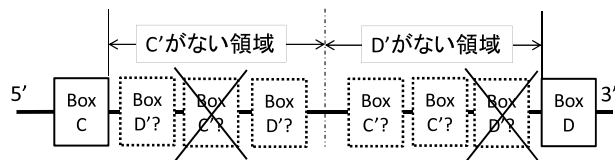


図 7 boxC'/D' の出現順制約による boxC'/D' 候補の除外  
 Fig. 7 Elimination of boxC'/D' candidates by appearance order rule.

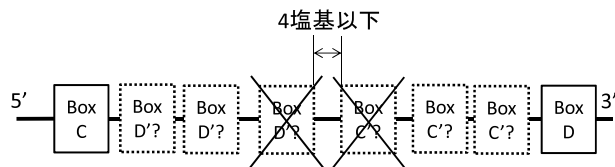


図 8 ヘアピンループ存在制約による boxC'/D' 候補の除外  
 Fig. 8 Elimination of boxC'/D' candidates by hairpin-loop existence rule.

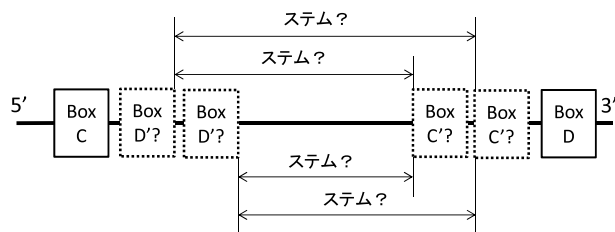


図 9 複数の boxC'/D' 候補が存在する場合のステム存在確率計算パターン  
 Fig. 9 Stem existence probability calculation on a sequence including multiple boxC'/D' candidates.

なるヘアピンループ領域を構成しえない候補を除外する。

上記の条件にあてはまる boxC'/D' 候補領域を除外したのち、boxC'/D' 候補間には stem-2 が存在することを考え、図 9 のように残っている候補間で、3.4.3 項で述べる stem-2 存在確率を計算し、stem-2 存在確率が最大の boxC'/D' 候補を boxC'/D' と決定する。boxC'/D' 間の stem-2 存在確率を評価したあとでもなお複数の boxC'/D' 候補が存在する場合、boxC' 候補と boxD' 候補の間が一番離れている組合せを採用し、boxC'/D' に決定する。これは、boxC' と boxD' が離れるほど、boxC'/D' 間に stem-2 を構成できる塩基数が増えるためである。

boxC/D と異なり、 $P_B^{C'}$  が  $\frac{4}{7}$  未満、 $P_B^{D'}$  が  $\frac{3}{4}$  未満の場合、当該配列は snoRNA 遺伝子を含まないのではなく、boxC' や boxD' を含まないものとして扱う。

## 3.4 ステム存在確率の導出

### 3.4.1 ステム存在確率の定義

ステム存在確率の定義について説明する。2.1 節で説明したように、ステムはワトソン-クリックの相補塩基対により構成される。なお、RNA ではワトソン-クリックの相

補塩基対だけでなく、G-U 塩基対や G-A 塩基対など様々な塩基対が形成されることが知られているが [9], 本手法では最も出現頻度の高いワトソン-クリックの相補塩基対に着目する. 図 3 に示したとおり, boxC/D 型 snoRNA は boxC と boxD の両端にステム構造 stem-1 を持ち, boxC' と boxD' の間にもまた別のステム構造 stem-2 を持つ. boxC' と boxD' を含まない場合, stem-2 は boxC と boxD の間に構成する. したがって, 前節までに確定した boxC/D や boxC'/D' の間に存在するはずの 2 つのステム構造それぞれに対して stem 存在確率を計算し, その平均をステム存在確率  $P_S$  とする. ステム存在確率は式 (4) で定義する.

$$P_S = \frac{\text{stem-1 存在確率} + \text{stem-2 存在確率}}{2} \quad (4)$$

### 3.4.2 stem-1 存在確率の導出

boxC/D 端に存在する stem-1 の存在確率を求めるため, boxC 端から 5' 側の 10 塩基, boxD 端から 3' 側の 10 塩基に着目する. 両端の塩基数が 10 未満の場合は, どちらか短い方に合わせる. 着目する塩基数を 10 とした理由は, 「CentroidFold」を用いた予備実験で stem-1 を構成する相補塩基対数の最大値が 7 と求められたため, 余裕を含んで設定したものである.

stem-1 存在確率は, ボックス端に塩基の挿入や欠損が発生している場合を考慮して図 10 に示す 3 つのパターンで計算し, ステムを構成できる相補塩基対数が最も多い場合を採用する. 例として, ある塩基配列 X 「...AAAAGUCA (Box C) ... (BoxD) GACUUUUU...」における stem-1 出現確率を求める場合について説明する.

boxC の 5' 側端 8 塩基, boxD の 3' 側端 8 塩基すべてがステムを構成すると仮定した場合, 相補塩基対数は 8 であり, これを分母として用いる. 図 10(a) に示すように, boxC 端と boxD 端に塩基挿入がない場合は (A-U) が 4 つ含まれることから, stem-1 存在確率は  $\frac{4}{8} = 0.5$  となる. 次に, 図 10(b) に示すように boxC 端に塩基挿入があった場合を考える. 図 10(b) の例では, (G-C) が 1 つ, (A-U) が 1 つ, (C-G) が 1 つ, (U-A) が 4 つの計 7 となり, この場合の stem-1 存在確率は  $\frac{7}{8} = 0.875$  となる. 最後に, 図 10(c) に示すように boxD 端に塩基挿入があった場合を考える. 図 10(c) の例では, (A-U) が 3 つが含まれるので, stem-1 存在確率は  $\frac{3}{8} = 0.375$  となる. したがって, この例では boxC 端に塩基の挿入があった場合の図 10(b), 0.875 が stem-1 存在確率となる.

### 3.4.3 stem-2 存在確率の計算

図 3 に示したとおり, snoRNA は boxC' と boxD' の間にステム構造 stem-2 を構成する. boxC' や boxD' が含まれない場合は boxC/boxD 間で stem-2 を構成する. stem-2 存在確率を求めるとき, 提案手法ではヘアピンループを考慮せずに行う. これは, ヘアピンループ部分の位置が不確定のためである.

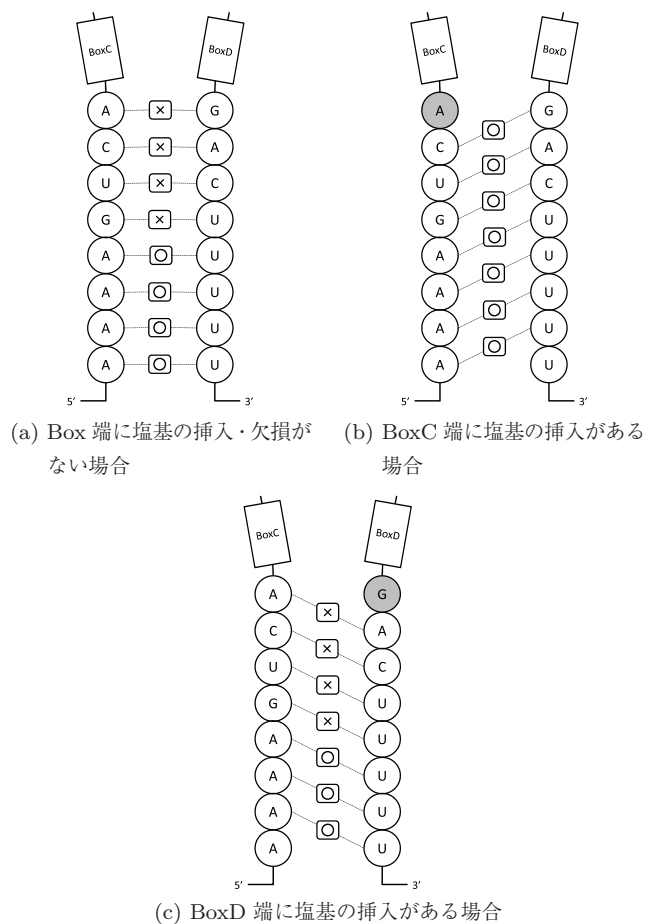


図 10 stem-1 存在確率を求めるときの相補塩基対構成パターン  
**Fig. 10** Patterns of complementary base pairs to calculate stem-1 existence probability.

stem-2 存在確率は, 3.4.2 項で示した stem-1 存在確率の導出時と同じく, boxC' と boxD' 端に塩基挿入がない場合, boxC' 側に 1 塩基が挿入されている場合, boxD' 側に 1 塩基が挿入されている場合の 3 つのパターンで計算する. このうち, ステムを構成できる相補塩基対数が最も多い場合を採用し, stem-2 存在確率とする.

例として, ある塩基配列 Y 「... (box D') CGGGCGGU-CUGAAA (box C') ...」における stem-2 存在確率を求める場合について考える. 14 塩基からなる塩基配列 Y がすべてステムを構成すると仮定した場合, 相補塩基対数は 7 となる.

まず, boxC' と boxD' 間に塩基の挿入がない場合について相補塩基対数を求める. この場合, (G-C) が 1 つが含まれることから, stem-2 存在確率は  $\frac{1}{7} = 0.143$  となる. 次に, boxC' 側に 1 塩基の挿入があった場合について, 相補塩基対数を求める. この場合ではステムを構成できる相補塩基対は存在せず, stem-2 存在確率は  $\frac{0}{7} = 0.0$  となる. 最後に, boxD' 側に 1 塩基の挿入があった場合について相補塩基対数を求める. この場合では, (G-C) が 1 つ, (C-G) が 1 つの計 2 つの相補塩基対が含まれることから, stem-2 存在確率は  $\frac{2}{7} = 0.286$  となる. これら 3 つの stem-2 存在

確率のうち最も大きい場合、すなわち boxD' 側に 1 塩基の挿入があったとした場合の 0.286 が最終的な stem-2 存在確率となる。

## 4. 実験と考察

### 4.1 実験に使用する塩基配列データ

評価実験を行うにあたり、boxC/D 型 snoRNA 遺伝子を含む正例データと、それを含まない負例データを作成した。これら評価実験で用いるデータは、データベース“snOPY” [21] から取得した。実験に使用する boxC/D 型 snoRNA 遺伝子の鎖長分布を図 11 に示す。

評価実験で用いる正例データは、snoRNA 遺伝子を中央に配置するよう長さ 120 塩基で切り出した配列で、1,100 個作成した。負例データは、正例データと同じ生物種の塩基配列から snoRNA 遺伝子をできる限り含まないように、snoRNA 遺伝子と離れた位置から長さ 120 塩基で 1,100 個作成した。作成したデータのうち、正例 100 個と負例 100 個は予備実験に使用する。作成した負例データは、塩基配列内に snoRNA 遺伝子が含まれていないか「CLUSTALW」 [23] を用いて検証を行った。「CLUSTALW」は広く用いられている多重整列プログラムであり、これによりペアワイズアラインメントを行った後、正例・負例間に重複がないことを確認した。

### 4.2 ボックス存在確率係数、ステム存在確率係数、およびしきい値の決定

予備実験として、3.2 節で述べた boxC/D 型 snoRNA 遺伝子の有無を判断するしきい値  $P_{th}$ 、および判定に係るボックス存在確率とステム存在確率の寄与を調整するパラメータ  $\alpha$ 、 $\beta$  を決定する。ここでは先にとりおいた正例データ 100 個と負例データ 100 個を用いて、 $\alpha$  と  $\beta$  を 0.0 から 1.0 まで 0.1 刻みで変更し、各  $\alpha$ 、 $\beta$  の組合せについて  $P_{th}$  を 0.1 から 0.9 まで 0.1 刻みで変更して最も高い検出精度 (Accuracy) を示した  $\alpha$ 、 $\beta$ 、 $P_{th}$  の組合せを決定する。

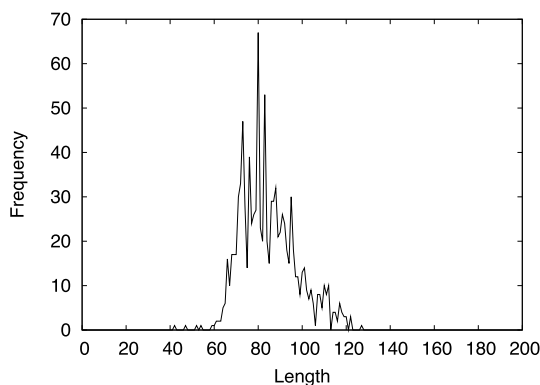


図 11 実験に使用した正例データの鎖長分布

Fig. 11 Length distribution of boxC/D type snoRNA genes for experiments.

ここで検出精度とは、正例、負例とその判別結果を表 1 のように分類したとき、式 (5) で定義される。

$$Accuracy = \frac{A + D}{A + B + C + D}. \quad (5)$$

予備実験における各  $\alpha$ 、 $\beta$ 、 $P_{th}$  での検出精度を表 2 に示す。表 2 から、最も精度良く snoRNA 遺伝子を検出した  $(\alpha, \beta, P_{th})$  は、(0.5, 0.5, 0.5), (0.3, 0.7, 0.6), (0.1, 0.9, 0.7) であり、表 2 中に太字で表す。評価実験では、この 3 つの組合せ、およびボックスとステムを均等に考慮 ( $\alpha = \beta = 0.5$ ) したときの  $P_{th}$  の影響を調べるため、(0.5, 0.5, 0.1), (0.5, 0.5, 0.9) の 2 つを加えた 5 つの組合せで、それぞれ boxC/D 型 snoRNA 遺伝子の検出を行う。

### 4.3 boxC/D 型 snoRNA 遺伝子検出実験結果

予備実験で決定した  $(\alpha, \beta, P_{th})$  を用いて、boxC/D 型 snoRNA 遺伝子の検出実験を行う。正例データと負例データをそれぞれ 1,000 個を使用して実験を行い、感度 (Sensitivity)、特異度 (Specificity)、検出精度を比較する。感度とは正例を正しく判定できる確率であり、検証用データと実験結果を表 1 のとおり分類すると式 (6) で定義される。特異度とは負例を正しく判定できる確率であり、式 (7) で定義される。

$$Sensitivity = \frac{A}{A + B}, \quad (6)$$

$$Specificity = \frac{D}{C + D}. \quad (7)$$

5 種類の  $(\alpha, \beta, P_{th})$  の組合せについて、snoRNA 遺伝子検出実験を行った結果を表 3 に示す。

表 3 に示したとおり、(0.5, 0.5, 0.5), (0.3, 0.7, 0.6), (0.1, 0.9, 0.7) の組合せにおいて、検出精度 0.97 以上を達成している。表 3 から、 $\beta$  を増加すると特異度が向上する傾向がみられる。また、(0.5, 0.5, 0.1), (0.5, 0.5, 0.5), (0.5, 0.5, 0.9) での結果を比較すると、 $P_{th}$  を小さくしてもあまり感度は変化しないことが分かる。これは、提案手法では boxC は 7 塩基中 4 塩基、boxD は 4 塩基中 3 塩基以上の一致を要求しており、これを満たすサンプルの  $P_B$  は約 0.66、 $\beta = 0.5$  によりこれが 2 分の 1 となっても約 0.33 で、仮にまったくステムを構成できないサンプルであっても ( $P_S = 0.0$ )、 $P_{th} = 0.1$  を超えるからと考えられる。また、 $P_{th}$  が小さくなることで一部の負例サンプルが正例として検出されてしまい、特異度を低下させていることが分かる。一方、 $P_{th}$  を大きくすると、snoRNA 遺伝子とし

表 1 判別結果の分類

Table 1 Classification of samples and detection results.

		判別結果	
		正	負
検証用 データ	正例	A	B
	負例	C	D

表 2  $\alpha, \beta, P_{th}$  の組合せによる検出精度  
Table 2 Accuracy on each combination of  $\alpha, \beta$  and  $P_{th}$ .

$\alpha$	$\beta$	$P_{th}$								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1.0	0.0	0.940	0.945	0.925	0.850	0.755	0.640	0.550	0.515	0.500
0.9	0.1	0.925	0.960	0.935	0.895	0.790	0.680	0.565	0.515	0.500
0.8	0.2	0.920	0.930	0.970	0.940	0.855	0.730	0.590	0.515	0.500
0.7	0.3	0.920	0.920	0.945	0.975	0.890	0.780	0.630	0.515	0.500
0.6	0.4	0.920	0.920	0.925	0.970	0.960	0.855	0.715	0.515	0.500
0.5	0.5	0.920	0.920	0.920	0.945	<b>0.985</b>	0.920	0.785	0.530	0.500
0.4	0.6	0.920	0.920	0.920	0.920	0.970	0.975	0.830	0.605	0.500
0.3	0.7	0.920	0.920	0.920	0.920	0.945	<b>0.985</b>	0.955	0.725	0.500
0.2	0.8	0.920	0.920	0.920	0.920	0.920	0.970	0.975	0.815	0.565
0.1	0.9	0.920	0.920	0.920	0.920	0.920	0.925	<b>0.985</b>	0.920	0.670
0.0	1.0	0.920	0.920	0.920	0.920	0.920	0.920	0.975	0.955	0.745

表 3 boxC/D 型 snoRNA 遺伝子の検出精度

Table 3 Sensitivity, specificity and accuracy on boxC/D type snoRNA gene detection experiments.

$\alpha$	$\beta$	$P_{th}$	感度	特異度	検出精度
0.5	0.5	0.5	0.951	0.990	0.971
0.3	0.7	0.6	0.955	0.994	0.975
0.1	0.9	0.7	0.953	0.999	0.976
0.5	0.5	0.1	0.958	0.825	0.892
0.5	0.5	0.9	0.413	0.961	0.687

て検出されるためには、より多くの塩基が boxC/D を構成する配列と一致したうえ、ステムを構成する相補塩基対数も多くなければならず、感度は大きく低下する。この感度低下により、検出精度も低下することが表 3 から分かる。

表 3 の最初に示した 3 種類の ( $\alpha, \beta, P_{th}$ ) の組合せのうち、最も感度が高い (0.3, 0.7, 0.6) について、従来研究との比較を行った結果を表 4 に示す。表 4 のうち、SSK, StK, DAG-based Stk, BPLA kernel, SnoReport については文献記載の値である。CDseeker [24] の結果は正例として提案手法と同じ boxC/D 型 snoRNA 遺伝子そのものの配列データを用いたときの値であり、サンプルの長さを 120 塩基に揃えていないことに注意する必要がある。これらのうち、感度、特異度、検出精度すべてが比較できる SSK, StK, CDseeker と提案手法との比較を図 12 に示す。

表 4, および図 12 から、提案手法は検出精度で従来研究より大幅に向上し、感度について SnoReport と同等、特異度について BPLA kernel, CDseeker と同等となっている。感度を比較すると、SSK は 0.440, StK は 0.800, CDseeker は 0.332 であるのに対し、提案手法は 0.955 と大きく向上している。特異度については、BPLA kernel と CDseeker が 1.00 と最も高く、次いで提案手法の 0.994 と続く。検出精度は提案手法が 0.975 と最も高く、続いて CDseeker の 0.666, SSK の 0.655, StK の 0.640 となっており、提案手法での感度の向上が検出精度の向上に大きく貢献していることが分かる。

#### 4.4 考察

予備実験における  $\alpha, \beta, P_{th}$  について考える。表 3 に示すように、 $\beta$  を大きくすると特異度が向上する傾向が見られた。その理由として、boxC/D 型 snoRNA 遺伝子における「ボックス」という特徴的塩基配列が重要であることが考えられる。boxC は 7 塩基で構成されており、boxC に完全一致する塩基配列が偶然出現する確率は小さい。そのため、snoRNA 遺伝子内に高い確率で含まれる boxC の存在は、boxC/D 型 snoRNA 遺伝子の判別において非常に重要な意味を持つといえる。

感度について考える。図 12 に示すように、CDseeker の感度が最も低く、次いで SSK の感度が低い。CDseeker では、snoRNA 遺伝子であると判断した場合のみボックスやステムの位置などの情報を出力し、snoRNA 遺伝子でないと判断された場合にも出力しないので、どのような理由で正例が snoRNA 遺伝子でないと判断されたのかは不明である。CDseeker では様々なパラメータを調整できることから、これらを適切に設定することで感度は向上する可能性もある。SSK については、ボックスの情報だけで snoRNA 遺伝子の有無を判定するため、boxC に比べ少ない塩基数で構成される boxD と同じ塩基配列が偶然表れる可能性が高く、多くの配列を boxD と誤認する可能性があるため、感度が低くなったと考えられる。

特異度について考える。図 12 から、StK の特異度が最も低い。StK は二次構造であるステムの存在確率から snoRNA 遺伝子の有無を予測するが、その際、連続、または不連続のステム構造候補を詳細に調べてすべてカウントする。このため、ステム構造候補が多く出現して snoRNA 遺伝子ありと判断するケースが多くなり、感度が高く、特異度が低くなったと考えられる。一方、StK を改良した DAG-based StK, BPLA kernel では、表 4 から分かるように特異度の大幅な向上を果たしていることが分かる。

提案手法では、boxC/D 型 snoRNA の構造モデルに基づき、図 4 に示したように、ボックス候補の位置にいくつ



表 4 従来研究との比較

Table 4 Comparison of proposed method with previous works.

手法	$\alpha$	$\beta$	$P_{th}$	感度	特異度	検出精度
SSK [15]				0.440	0.772	0.655
StK [15]				0.800	0.606	0.640
DAG-based StK [16]				0.71	0.91	-
BPLA kernel [17]				0.71	1.00	-
CDseeker v1.1				0.332	1.00	0.666
SnoReport [17]				0.96	0.91	-
提案手法	0.3	0.7	0.6	0.955	0.994	0.975

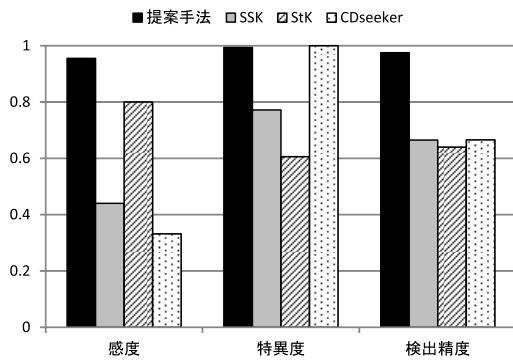


図 12 従来研究との比較

Fig. 12 Comparison of proposed method with previous works.

かの制限を課している。そのため、ボックスの探索時に候補が多数出現しても、その制限により存在し得ないボックス候補を除外し、残る候補についてステム構造を構成しうるかを考慮し、snoRNA 遺伝子か否かを判別する。すなわち、ボックスとステム双方を総合的に考慮したことから、感度、特異度、検出精度のすべてで性能が向上したと考えられる。

#### 4.5 長大配列からの snoRNA 遺伝子検出実験

4.3 節では、120 塩基で切り出した配列に対し、snoRNA 遺伝子が含まれるか否かを判定し、良好な結果を得た。そこで本節では、提案手法により長い塩基配列上の snoRNA 遺伝子を正しく検出できるか検証する。データベース “snOPY” から、snoRNA 遺伝子を含む「host gene」からサンプル 100 個を取り出す。生物種はヒト、線虫、ネズミとシロイヌナズナの 4 種類である。長さは生物種ごとに異なっているが、少なくとも 2,000 塩基以上とした。

実験方法は以下のとおりである。まず、snoRNA 遺伝子が何塩基目から何塩基目までを占めるか記録しておく。次に、配列の先頭から 1 塩基ずつずらしながら、120 塩基長からなるサンプルを抜き出す。抜き出した各サンプルに対し提案手法で snoRNA 遺伝子の有無を調べ、snoRNA 遺伝子が含まれていた場合は、配列の先頭からの位置を求める。記録しておいた位置と検出された snoRNA 遺伝子の位置を比較し、これらが一致するかどうかを評価する。

両端の stem-1 から  $\pm 5$  塩基の検出誤差を許して実験を

行ったところ、100 サンプル中 94 サンプルで snoRNA 遺伝子を正しく検出することができた。正しく検出することができなかったサンプル 6 個のうち 4 つがヒトのデータであった。例として、“SNORD74L2” というヒト遺伝子について説明する。“SNORD74L2” では、4 塩基中 3 塩基以上が一致する boxD 候補が 2 つ見つかる。図 4 に示したモデル構造に基づく、boxC の末端から boxD の先頭までは 26 塩基以上離れていなければならない。しかし、“SNORD74L2” で見つかる 2 つの BoxD 候補はこの制限を満たしておらず、snoRNA 遺伝子として検出することができなかった。

#### 5. おわりに

近年、ノンコーディング RNA の 1 つである機能性 RNA を発見し、その役割を解析することは、分子細胞生物学やバイオインフォマティクス双方において、最も重要な研究課題の 1 つになっている。現在では、基本的な代謝から個体発生、細胞分化までの様々な生命現象に関与する機能性 RNA が多く見いだされている。機能性 RNA と疾患との関わりに関する研究成果も次々に報告されており、創薬や再生医療分野などで大きな進展をもたらすことが期待されている。本論文では、機能性 RNA の 1 つである核小体低分子 RNA (snoRNA) を対象とし、コンピュータによって boxC/D 型 snoRNA 遺伝子を高精度に自動検出することを目的とした。

本論文では、boxC/D 型 snoRNA の特徴であるボックスとステムの 2 つの構造両方に着目し、構造モデルとの一致確率に基づく塩基配列からの snoRNA 遺伝子検出法を提案した。構造モデルは、1 つ以上の相補塩基対からなるステム領域 2 カ所、7 塩基中 4 塩基以上が一致している boxC/C' 領域、4 塩基中 3 塩基以上が一致する boxD/D' 領域、4 塩基以上の標的 rRNA との相補領域、4 塩基以上のヘアピンループ領域からなる。このモデル構造に基づき、各ボックスの存在可能位置などいくつかの制限を設定している。そのため、ボックス候補が多数存在しても、モデル構造によって存在しえない候補を除外することができる。

実際の塩基配列を用いて検出精度を検証した結果、boxC/D 型 snoRNA について 97.5% の検出精度を得ること

ができた。さらに、長さ 2,000 塩基以上の、ヒトや線虫、ネズミ、シロイヌナズナの塩基配列を使い、正しく snoRNA 遺伝子の位置を特定できるかどうかの実験を行った。その結果、100 サンプル中 94 サンプルで snoRNA 遺伝子を正しく検出することができた。このことより、提案手法の有効性が示された。

今後の課題としては、長大塩基配列からの snoRNA 遺伝子検出実験で一部位置の特定ができないサンプルがあったことから、構造モデルを改良しての精度向上や、G-U 塩基対など RNA 中に見られる塩基対の考慮、任意位置での塩基の挿入・欠損を考慮したステム存在確率の計算などがあげられる。

### 参考文献

[1] Collins, F.S., Patrinos, A., Jordan, E., Chakravarti, A., Gesteland, R. and Walters, L.: New Goals for the U.S. Human Genome Project: 1998-2003, *Science*, Vol.282, No.5389, pp.682-689 (1998).

[2] ブラウン, T.A.: ゲノム新しい生命情報システムへのアプローチ, *メディカルサイエンスインターナショナル* (2007).

[3] Pennisi, E.: ENCODE Project Writes Eulogy for Junk DNA, *SCIENCE*, Vol.337, No.6099, pp.1159-1161 (2012).

[4] Encyclopedia of DNA Elements Consortium: Encyclopedia of DNA Elements, National Human Genome Research Institute (online), available from <http://encodeproject.org/ENCODE/> (accessed 2012-9-25).

[5] 浅井 潔: ポストゲノム時代の配列情報解析, *情報処理*, Vol.46, No.3, pp.271-276 (2005).

[6] 鈴木 勉: リボヌクレオーム: 広がる RNA 転写後プロセスの世界, *Cell technology*, Vol.22, No.9, pp.936-938 (2003).

[7] Eddy, S.R.: Non-coding RNA genes and the modern RNA world, *Nat. Rev. Genet.*, Vol.2, No.12, pp.919-929 (2001).

[8] Peters, J.: Prade-Willi and snoRNAs, *Nat. Genet.*, Vol.40, No.6, pp.688-689 (2008).

[9] 河合剛太, 清澤秀孔: 機能性 RNA の分子生物学, *クバプロ* (2010).

[10] Smith, C.M. and Steitz, J.A.: Sno Storm in the Nucleolus: New roles for Myriad small RNPs, *Cell*, Vol.89, pp.669-672 (1997).

[11] 河合剛太, 金井昭夫: 機能性 Non-coding RNA, *クバプロ* (2010).

[12] 澤田石翔太, 三功浩嗣, 土井晃一郎, 山本章博: 内包カーネルと配列分割法を用いた RNA 識別, *情報処理学会研究報告 (バイオ情報学)*, Vol.2008, No.15, pp.71-78 (2008).

[13] Schölkopf, B., Tsuda, K. and Vert, J.P.: *Kernel Methods in Computational Biology*, The MIT Press (2004).

[14] Lodhi, H., Saunders, C., Shawe-Taylor, J., Cristianini, N. and Watkins, C.: Text Classification using String Kernels, *J. Mach. Learn. Res.*, Vol.2, pp.419-444 (2002).

[15] Sakakibara, Y., Asai, K. and Sato, K.: Stem Kernels for RNA Sequence Analyses, *Bioinformatics Research and Development* (Hochreiter, S. and Wagner, R., eds.), Lecture Notes in Computer Science, Vol.4414, pp.278-291, Springer Berlin/Heidelberg (2007).

[16] Sato, K., Mituyama, T., Asai, K. and Sakakibara, Y.: Directed acyclic graph kernels for structural RNA anal-

ysis, *BMC Bioinformatics*, Vol.9, No.1, p.318 (2008).

[17] Morita, K., Saito, Y., Sato, K., Oka, K., Hotta, K. and Sakakibara, Y.: Genome-wide searching with base-pairing kernel functions for noncoding RNAs: computational and expression analysis of snoRNA families in *Caenorhabditis elegans*, *Nucleic Acids Research*, Vol.37, No.3, pp.999-1009 (2009).

[18] Yang, J.-H., Zhang, X.-C., Huang, Z.-P., Zhou, H., Huang, M.-B., Zhang, S., Chen, Y.-Q. and Qu, L.-H.: snoSeeker: an advanced computational package for screening of guide and orphan snoRNA genes in the human genome, *Nucleic Acids Research*, Vol.34, No.18, pp.5112-5123 (2006).

[19] Hertel, J., Hofacker, I.L. and Stadler, P.F.: SnoReport: computational identification of snoRNAs with unknown targets, *Bioinformatics*, Vol.24, No.2, pp.158-164 (2008).

[20] 長沼孝雄, 廣瀬哲郎: noncoding RNA の機能単位, *実験医学*, Vol.28, No.10 (増刊), pp.30-36 (2010).

[21] RI Laboratory Frontier Science Research Center: snOPY, University of Miyazaki (online), available from [http://snoopy.med.miyazaki-u.ac.jp/snorna\\_db.cgi](http://snoopy.med.miyazaki-u.ac.jp/snorna_db.cgi) (accessed 2012-7-20).

[22] Sato, K., Hamada, M., Asai, K. and Mituyama, T.: CENTROIDFOLD: a web server for RNA secondary structure prediction, *Nucleic Acid Research*, Vol.37, No. suppl 2, pp.W277-W280 (2009).

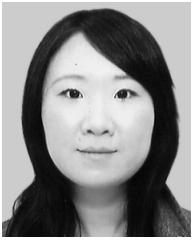
[23] Thompson, J.D., Higgins, D.G. and Gibson, T.J.: CLASTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acid Research*, Vol.22, No.22, pp.4673-4680 (1994).

[24] Yang, J.-H.: snoSeeker, SUN YAT-SEN UNIVERSITY (online), available from <http://genelab.sysu.edu.cn/snoseeker/> (accessed 2012-10-4).



山森 一人 (正会員)

1997 年北陸先端科学技術大学院大学情報科学研究科博士後期課程修了。同年北陸先端科学技術大学院大学情報科学研究科助手。2001 年宮崎大学工学部助教授。2007 年同大学准教授。並列処理, および神経回路網応用, 進化型計算に関する研究に従事。博士 (情報科学)。IEEE-CS, 電子情報通信学会, 日本バイオインフォマティクス学会各会員。



**薛 晨**

2007年天津冶金職業技術大学工学部電気電子工学科卒業。2009年宮崎大学工学部情報システム工学科研究生。2010年宮崎大学大学院工学研究科情報システム工学科入学，2012年同専攻修了。機能性RNA等の遺伝子解析に関する研究に従事。工学修士。情報処理学会第74回全国大会にて推奨修士論文認定。



**岩切 淳一**

2012年宮崎大学大学院医学系研究科博士課程修了。同年東京大学大学院新領域創成科学研究科特任研究員。バイオインフォマティクス，特に生体高分子間相互作用に関する研究に従事。博士(医学)。日本生物物理学会，日本分子生物学会各会員。



**劍持 直哉**

1978年東北大学工学部卒業。同年新潟大学医学部教務職員。1990年琉球大学医学部助手。1993～1995年マサチューセッツ工科大学ホワイトヘッド研究所研究員。2001年宮崎医科大学実験実習機器センター助教授。2003年宮崎大学フロンティア科学実験総合センター助教授。2008年同教授。分子遺伝学およびゲノム科学の手法を用いた機能性RNAの解析に従事。博士(医学)。日本分子生物学会，日本生化学会，日本RNA学会，米国人類遺伝学会各会員。



**吉原 郁夫** (正会員)

1969年東京工業大学理工学部物理学科卒業。1971年東京教育大学大学院理学研究科物理学専攻修士課程修了。1971～1999年(株)日立製作所中央研究所，システム開発研究所にて，ビル防災，環境制御，数値計算，ニューラルネットワーク，遺伝的アルゴリズムの研究に従事。その間，1997～1999年東北大学客員助教授，1999年宮崎大学客員教授を兼務。1999年より宮崎大学工学部教授。2006～2008年同大学総合情報処理センター長。進化型計算，ゲノム情報処理の研究に従事。2012年同大学名誉教授。著書『知的情報処理』(共著)ほか。2003年FPGA/PLDデザイン会議，2004年SEAL'04，2008年ICES2008にてそれぞれ最優秀論文賞受賞，ほか。工学博士。IEEE，電子情報通信学会，電気学会，計測自動制御学会，日本応用数理学会，進化計算学会各会員。