

## 科学研究費助成事業 研究成果報告書

平成 29 年 5 月 15 日現在

機関番号：12601

研究種目：若手研究(B)

研究期間：2013～2016

課題番号：25870190

研究課題名(和文) RNA二次構造の大域的性質の集団遺伝解析

研究課題名(英文) A population genetics analysis of RNA secondary structures

研究代表者

木立 尚孝(Kiryu, Hisanori)

東京大学・大学院新領域創成科学研究科・准教授

研究者番号：80415778

交付決定額(研究期間全体)：(直接経費) 3,200,000円

研究成果の概要(和文)：細胞内のRNA分子は、遺伝子に書き込まれた情報がタンパク質になり生理的な機能を発揮する上で非常に重要な役割を持つ。RNA分子の立体構造は、ステムと呼ばれる局所的二重らせん構造(二次構造)の3次元的な配置としてよく理解されることが知られており、二次構造の性質を解明することは、RNAの機能を理解する上で重要である。本研究において、我々は、メッセンジャーRNAや長鎖非コードRNAと呼ばれる長大なRNA分子に対して、その二次構造的性質を正確に計算することができるアルゴリズムの開発に世界で初めて成功した。また、RNA結合タンパク質の結合領域周辺がどのような二次構造的特徴をもつかを計算する手法も開発した。

研究成果の概要(英文)：RNA molecules in a cell play very important roles so that genetic information encoded in the genome is instantiated as proteins and exerts actual functions. Three dimensional structure of an RNA is understood by its arrangement of stem structures, and it is important to investigate the properties of this secondary structure to understand the functions of RNAs. In this study, we have succeeded in developing an algorithm (ParasoR) for computing several properties of RNA secondary structures of very long RNAs such as messenger RNAs and long non-coding RNAs for the first time. We have also succeeded in developing an algorithm (CapR) for computing secondary structural context around the binding regions of RNA binding proteins.

研究分野：バイオインフォマティクス

キーワード：バイオインフォマティクス RNA RNA二次構造 アルゴリズム RNA結合タンパク質 文脈自由文法 RNA進化

### 1. 研究開始当初の背景

(1) DNA シーケンシング技術の劇的な低コスト化により、大規模な個人ゲノムのシーケンシングが行われ、全ゲノム関連解析などにより、疾患に関連する一塩基多型 (SNP) の探索が広く行われた。その結果として分かったのは、多くの疾患関連 SNP は、アミノ酸をコードしない非コード領域にあるということだった。例えば、全体の疾患関連 SNP の 40% 以上は、イントロン領域に、存在することが示唆された。タンパク質をコードしないゲノム領域上で、これらの SNP がどのような機序で疾患感受性に影響を与えるのかを解明することが重要な課題となっていた。

(2) 研究代表者は、長年 RNA 情報解析ツールの開発を行っている。RNA 解析の中でも、RNA の二次構造は、確率文脈自由文法と呼ばれる情報科学的な手法によりうまくモデル化できることが知られており、この技術を活用して、疾患関連 SNP の機能を解明する研究に取り組むことにした。

### 2. 研究の目的

非コード転写領域の SNP は、転写された非コード RNA や、メッセンジャーRNA の構造変化を通して、生命活動に影響を与える可能性が考えられる。この仮説のもとに、我々は、これらの RNA の二次構造が、塩基置換などによりどのような影響を受けるかを網羅的に計算するツールの開発を行い、ゲノム上の非コード領域に働く、二次構造的な制約を明らかにすることを目標とした。

### 3. 研究の方法

(1) RNA 二次構造アルゴリズムを改変し、メッセンジャーRNA、長鎖非コード RNA などの長大な RNA 分子に対して、二次構造計算を正確に行うことができる新規の並列計算アルゴリズム (ParasoR) を開発する。

(2) ParasoR を用いて、長い配列に対する二次構造予測の性質を調べ、既存手法に対する有意性を示す。

(3) ParasoR をヒトゲノム全体に適用し、ゲノムの各機能領域でどのような二次構造的な拘束が働いているのかを調べる。

(4) RNA 二次構造アルゴリズムを応用し、RNA 配列の各位置がどのような二次構造的な文脈にあるかを網羅的に計算する新規のアルゴリズム (CapR) を開発する。

(5) CapR を、CLIP-seq データにより測定される、RNA 結合タンパク質結合領域周辺の RNA 配列に適用し、RNA 結合タンパク質がどのような二次構造的な特性をもって結合するターゲットを認識するのかを調べる。

### 4. 研究成果

(1) ParasoR の精度を調べたところ、従来のソフトウェア・ツールと同等程度またはそれ以上の正確さを持つことが分かった。

(2) ParasoR のスケーラビリティを調べたところ、メッセンジャーRNA、非コード RNA など、従来計算が不可能だった長大な RNA 配列でも数値精度の誤りなく二次構造計算をすることができた。試しにヒトゲノムサイズの人工配列を与えても二次構造を計算することができたことが分かった。これにより、二次構造計算アルゴリズムのスケーラビリティの問題を世界で初めて解決することができたと考えている。

(3) ParasoR をヒトゲノムの各領域に適用して、各領域のステム確率を計算したところ、ヒトゲノムの機能領域には強い二次構造的な制約が働いていることが明らかになった。例えば、ヒトのイントロン領域では、全体に渡って、ステムを作ろうとする強い二次構造的な制約が働いていることが分かった。(図1) また、スプライシングサイト周辺では、スプライシングにより構造を取りにくい配列構造へとへんかすることが分かり、アミノ酸翻訳速度を落とさないようにする進化的な制約が働いていることが示唆された。

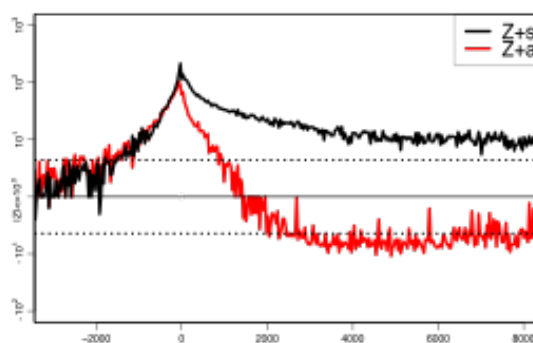


図1：ヒトゲノムのイントロン領域に働く二次構造的な制約。：横軸は、スプライシングサイトからの遺伝子上の位置を表す。0より左がコード領域、右がイントロン領域。縦軸は、遺伝子間領域に比べ、ステムを組む傾向が強いか弱いかを表している。黒線から、イントロンの全域に渡り、二次構造を組もうとする強い傾向があることが分かる。赤線はイントロン配列の相補鎖を表し、こちらは、イントロンとは逆の二次構造的傾向を持つことを示している。

(4) 既知 RNA 構造のデータベース Rfam のデータを用いて、CapR の性能を調べたところ、ヘアピン、ステム、内部ループ、外部ループの各二次構造的な文脈を非常に高い精度で予測できることが分かった。

(5) 公開されている CLIP-seq データと CapR を用いて、RNA 結合タンパク質の結合領域周辺がどのような二次構造的な特性をとることが多いかを調べたところ、RNA 結合タンパク質

の種類に依存して、ステムに結合しやすいもの、内部ループに結合しやすいもの、ヘアピンループに結合しやすいものなど、様々な二次構造的な特異性を持ちうることが予測された。

(6) RNA と結合した X 線結晶構造が知られている幾つかのタンパク質 (例えば、脳で特異的に発現し、RNA のスプライシング制御に関与する Nova タンパク質) などと比較すると、CapR による予測と X 線結晶構造は、互いに矛盾のない結論を導くことが分かった。(図 2)

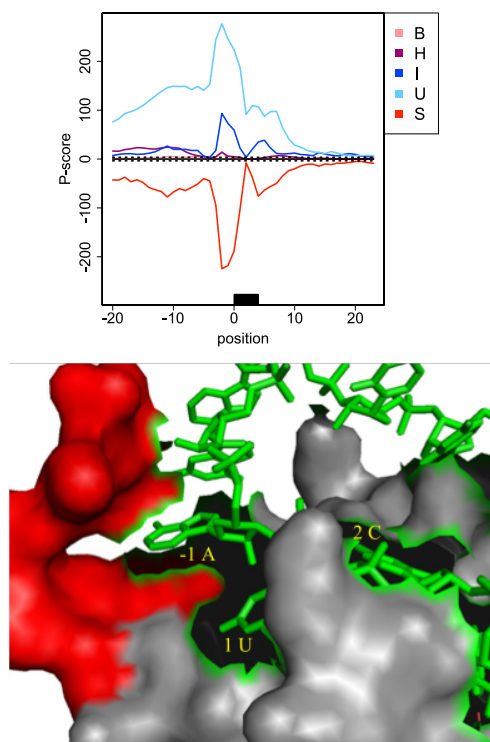


図 2 : (上) : CapR による Nova タンパク質結合領域周辺の二次構造的な特徴。横軸は配列上の位置で、縦軸は各二次構造的な文脈 (B : バルジ、H : ヘアピン、I:内部ループ、U:構造なし、S:ステム) の取りやすさを示す。黒の横線で示した、結合配列モチーフの直前で特に、二次構造を組みにくいことを示している。

(下) X 線構造解析による、Nova と RNA の結合構造。結合配列モチーフの直前は、丁度タンパク質のポケット領域に入っており、この部分が構造を組まないことは、Nova の RNA 認識における重要性を示唆している。

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 6 件)

① Hiroataka Matsumoto, Hisanori Kiryu, Chikara Furusawa, Minoru S. H. Ko, Shigeru B. H. Ko, Norio Gouda, Tetsutaro Hayashi, Itoshi Nikaido  
SCODE: An efficient regulatory network inference algorithm from single-cell RNA-

Seq during differentiation.

Bioinformatics; btx194 (2017)

doi: 10.1093/bioinformatics/btx194 査読有

② Hiroataka Matsumoto and Hisanori Kiryu

SCOUP: a probabilistic model based on the Ornstein–Uhlenbeck process to analyze single-cell expression data during differentiation

BMC Bioinformatics 17:232 (2016)

doi: 10.1186/s12859-016-1109-3 査読有

③ Risa Kawaguchi and Hisanori Kiryu

Parallel computation of genome-scale RNA secondary structure to detect structural constraints on human genome

BMC Bioinformatics 17:203 (2016)

doi:10.1186/s12859-016-1067-9 査読有

④ Michiaki Hamada, Yukiteru Ono, Hisanori Kiryu, Kengo Sato, Yuki Kato, Tsukasa Fukunaga, Ryota Mori, and Kiyoshi Asai

Rtools: a web server for various secondary structural analyses on single RNA sequences

Nucleic Acids Res. gkw337 (2016)

doi: 10.1093/nar/gkw337 査読有

⑤ Hiroataka Matsumoto, and Hisanori Kiryu

Integrating dilution-based sequencing and population genotypes for single individual haplotyping

BMC Genomics 15:733 (2014)

doi:10.1186/1471-2164-15-733 査読有

⑥ Tsukasa Fukunaga, Haruka Ozaki, Goro Terai, Kiyoshi Asai, Wataru Iwasaki and Hisanori Kiryu

CapR: revealing structural specificities of RNA-binding protein target recognition using CLIP-seq data

Genome Biology 15:R16 (2014)

doi:10.1186/gb-2014-15-1-r16 査読有

[学会発表] (計 6 件)

① Hiroataka Matsumoto, Inference of gene regulatory network and drivers of differentiation from time-course single-cell RNA-Seq, 第 5 回生命医薬情報連合大会 (IIBMP2016), 東京国際交流館プラザ平成 (東京都・江東区) 2016 年 9 月 30 日

② Risa Kawaguchi, Hisanori Kiryu, “Parallel computation of genome-scale RNA secondary structure to detect structural constraints on human genome”. 第 5 回生命医薬情報連合大会 (IIBMP2016) 東京国際交

流館プラザ平成（東京都・江東区）2016年9月30日

③ 松本 拓高 1細胞発現データと確率過程から読み解く細胞分化機序, 生命情報科学若手の会第7回研究会、慶應義塾大学鶴岡タウンキャンパス・TTCK レクチャーホール（山形県・鶴岡市）2015年10月1日

④ Risa Kawaguchi, “Genome-wide computational analysis of local RNA secondary structure ” Computational Analysis of RNA Structure and Function, Centro de ciencias de Benasque Pedro Pascual (Benasque, Spain), 2015年7月30日

⑤ Risa Kawaguchi, Hisanori Kiryu. “A comprehensive analysis of RNA secondary structure with deletion, insertion, and mutation” RiboClub meeting, Hotel et Villégiature Chèribourg (Québec, Canada), 2013年9月25日

⑥ Risa Kawaguchi, Hisanori Kiryu. “An algorithm for evolutionary analysis about the influence of single nucleotide variants on RNA secondary structure landscape”. Asian Young Researchers Conference on Computational and Omics Biology (AYRCOB), 7th meeting. 産業技術総合研究所 生命情報工学研究センター（東京都・江東区）2013年9月10日

〔図書〕（計0件）  
該当なし

〔産業財産権〕  
該当なし

○出願状況（計0件）  
該当なし

○取得状況（計0件）  
該当なし

〔その他〕  
ホームページ等

①SCODE: 一細胞シーケンシングデータから、遺伝子の発現相関を検出するソフトウェアの配布サイト:  
<https://github.com/hmatsu1226/SCODE>

②Branco: 系統樹上での塩基保存度の期待値を計算するソフトウェアの配布サイト:  
<https://github.com/izziyt/phylozzii>

③SCOUP: 一細胞シーケンシングデータから、各細胞の分化進行度や、遺伝子ごとの発現揺らぎの強さを推定するソフトウェアの配布サ

イト:  
<https://github.com/hmatsu1226/SCOUP>

④Rtools: RNA配列から、RNA二次構造に関する様々な統計量を計算するウェブサーバー:  
<http://rtools.cbrc.jp>

⑤ParasoR: 長い配列に対して、並列計算で二次構造を計算するソフトウェアの配布サイト:  
<https://github.com/carushi/ParasoR>

⑥CSP: 次世代シーケンシングデータから、SNPフラグメントのエラーを除去するソフトウェアの配布サイト:  
<https://sites.google.com/site/hmatsu1226/software/csp>

⑦CapR: RNA配列の各塩基について、その二次構造的なコンテキストを出力するソフトウェアの配布サイト:  
<https://github.com/fukunagatsu/CapR>

## 6. 研究組織

### (1) 研究代表者

木立 尚孝 (KIRYU Hisanori)  
東京大学・大学院新領域創成科学研究科・  
准教授  
研究者番号: 80415778

(2) 研究分担者  
該当なし

(3) 連携研究者  
該当なし

### (4) 研究協力者

浅井 潔 (ASAI, Kiyoshi)  
浜田 道昭 (HAMADA, Michiaki)  
佐藤 健吾 (SATO, Kengo)  
加藤 有己 (KATO, Yuki)  
岩崎 渉 (IWASAKI, Wataru)  
小野 幸輝 (ONO, Yukiteru)  
寺井 悟朗 (TERAI, Goro)  
尾崎 遼 (OZAKI, Haruka)  
松本 拓高 (MATSUMOTO, Hirotaka)  
福永 津嵩 (FUKUNAGA, Tsukasa)  
森 遼太 (MORI, Ryota)  
柏原 裕樹 (KASHIHARA, Yuki)  
河口 理紗 (KAWAGUCHI, Risa)