

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 6 日現在

機関番号：12601

研究種目：基盤研究(A)

研究期間：2010～2013

課題番号：22240031

研究課題名(和文)マイクロRNAおよびsnoRNAとその標的の網羅的予測

研究課題名(英文)Exhaustive prediction of microRNAs, snoRNAs and their targets

研究代表者

浅井 潔 (Asai, Kiyoshi)

東京大学・新領域創成科学研究科・教授

研究者番号：30356357

交付決定額(研究期間全体)：(直接経費) 35,700,000円、(間接経費) 10,710,000円

研究成果の概要(和文)：本研究では、主として以下の成果を得た。(a)新しいマイクロRNA予測ツールを開発し、既存のツールよりも、誤検出が少ないことを示した。(b)二次構造的なアクセシビリティを検出するツールを開発し、ゲノム上のマイクロRNA結合サイト周辺に働く二次構造的な制約を検出した。(c)変異による構造エネルギーや揺らぎの変化を計算するツールを開発した。これを用いて、ミトコンドリアtRNAの解析を行い、構造揺らぎの増加が疾患と関連することを示した。(d)RNA配列の各位置の構造的文脈を予測するツールを開発し、RNA結合たんぱく質の結合領域周辺の構造的な特異性を検出した。

研究成果の概要(英文)：The following is the main results of this project. (a) We developed a microRNA finding tool, which combines structural and evolutionary features of microRNAs. We showed that our tool outperforms previous methods with less false positives. (b) We developed an algorithm for computing secondary structural accessibilities of RNA sequences. We were able to detect the evolutionary constraints around the target sites of microRNAs. (c) We developed a tool to calculate the changes of free energies and entropies of RNA secondary structures in response to mutations. We applied our method to mitochondrial tRNAs and showed that the increase of structural fluctuation caused by mutations correlated with disease susceptibility. (d) We developed a tool to predict potential secondary structural contexts for each base of an RNA sequence. We applied our method to binding sites of RNA-binding proteins and detected the structural specificities of the proteins.

研究分野：総合領域

科研費の分科・細目：情報学・生体情報生命学

キーワード：バイオインフォマティクス 機能性RNA RNA snoRNA マイクロRNA 確率モデル

1. 研究開始当初の背景

(1) 2000 年前後に最初のヒトゲノム配列が公開されてから、ゲノム配列機能の網羅的解析がおこなわれてきた。中でも ACGT の 4 文字からなる文字列で表現されたゲノム配列は、計算機を用いた情報解析と親和性が高く、遺伝子の構造や機能のバイオインフォマティクス解析が広く行われている。

(2) それと前後して、ゲノムからの転写物を網羅的に計測する大規模な研究もおこなわれていた。これらの研究により、ヒトゲノムの大半の領域が、何らかの条件下、あるいはタイミングで、一度は転写されることが明らかにされた。また、マイクロ RNA や snoRNA などの、機能性構造 RNA も多数発見された。しかし、発見された RNA の大部分は、機能未知であり、それを明らかにすることが重要課題となっている。

2. 研究の目的

(1) 様々な非コード RNA の中でも、マイクロ RNA や snoRNA は、比較的安定した二次構造を持ち、ループ部分がターゲット RNA と相補的に結合するという点で、共通点がある。このような特徴は、二次構造解析や配列アライメントを行うバイオインフォマティクス解析と親和性が高い。そこで、計算機解析をフル活用して、ゲノムからマイクロ RNA と snoRNA を網羅的に発見するツールを開発する。

(2) マイクロ RNA や snoRNA が見つかったとしても、その機能がわからなければ生命の理解につながらない。このため、ターゲット候補領域周辺の二次構造的アクセシビリティ計算などを行い、これらの RNA のターゲット領域を精度よく網羅的に予測する。予測の正しさは次世代シーケンシング解析などにより検証する。

3. 研究の方法

(1) マイクロ RNA ・ snoRNA の発見には、細胞内の転写物を次世代シーケンシング解析により計測する実験的方法と、マイクロ RNA ・ snoRNA の二次構造に注目し、ゲノム配列からマイクロ RNA ・ snoRNA と同じ二次構造を取りうる領域を予測する計算機的方法があるが、本研究では、計算機的方法でマイクロ RNA ・ snoRNA を高い精度で予測する方法を開発する。そのために、RNA のエネルギーパラメータに基づく塩基対確率と、マイクロ RNA ・ snoRNA の進化的保存度を両方考慮したモデルにより、マイクロ RNA ・ snoRNA 遺伝子領域を発見する手法を開発する。

(2) マイクロ RNA ・ snoRNA の結合領域は、マイクロ RNA ・ snoRNA と相補鎖をもたねばならないため、転写物の相補領域が、候補となるが、単に相補領域を探すのみでは、特異性

が低すぎて、擬陽性がほとんどになってしまう。本研究においては、結合領域が二次構造的にアクセシブルでなければ、マイクロ RNA ・ snoRNA は効率的に相互作用できない、という特徴を利用し、アクセシビリティを網羅的に計算するツールの開発を行い、結合領域予測における擬陽性を減らすことを行う。

4. 研究成果

本研究課題はマイクロ RNA と snoRNA とそれらのターゲットサイトを網羅的に予測し、それを次世代シーケンシング解析などにより実験的に検証することであった。研究期間の半ばでメンバーの異動に伴う変更にもない、多少研究方針が修正され、RNA の二次構造解析手法の開発に、より重点をおいた研究を行うことになった。

結果として、二次構造アルゴリズムのゲノム解析への応用に関しては、国際的にみてもトップクラスのレベルの高い研究成果が続々と生まれた。このような研究を継続していくことができれば、今後、RNA のバイオインフォマティクス研究の国際的な拠点が形成されるだろうと考えられる。

研究期間内で得られた主な研究成果は以下のとおりである。

(1) ゲノムからのマイクロ RNA 遺伝子の発見

マイクロ RNA の前駆体はヘアピン型の二次構造を組むことが知られている。また、多くのマイクロ RNA のファミリーでは、ヘアピン構造のステム領域は非常によく保存していることが知られている (図 1)。

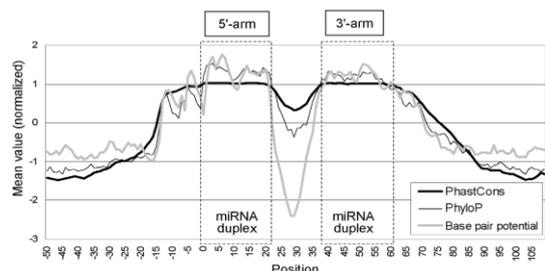


図 1. マイクロ RNA 遺伝子領域の進化的保存性 (実線) とステム形成確率 (灰色)。

これらの特徴をベクトル量として、条件付き確率場 (Conditional Random Field) の方法により、マイクロ RNA 遺伝子をモデル化し、既知のマイクロ RNA 遺伝子を用いて、パラメータを最適化した。この最適化したモデル (miRRim2) により、ゲノムから網羅的にマイクロ RNA を予測した所、既存の手法より、精度が向上することがわかった。

(2) RNA のアクセシビリティ計算法の開発

siRNA やマイクロ RNA が効率的にメッセンジャー RNA に結合するためには、ターゲット領域周辺が二次構造的に開かれた状態である必要がある。従って、一次配列的特徴から

予測された、各ターゲット領域候補について二次構造的なアクセシビリティを計算することにより、siRNA やマイクロ RNA が効率的に相互作用するかどうかを調べることができる。そこで、RNA 配列から、二次構造アルゴリズムを用いて、網羅的に配列の各領域のアクセシビリティを計算するアルゴリズムを開発し実装した (Raccess)。このプログラムを用いて、siRNA の抑制効率とアクセシビリティの相関をとったところ、siRNA のターゲットサイトの 3' 末端がアクセシブルであることが、抑制効率を最も高めることが明らかになった (図 2)。

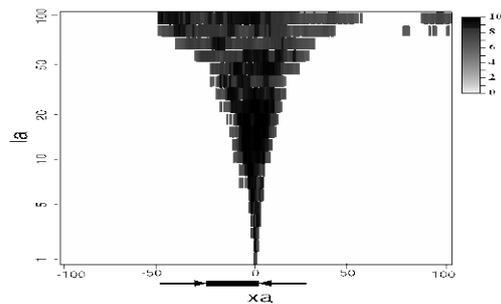


図 2. 抑制効率とアクセシビリティの相関の密度グラフ。黒いほど相関が高い。(横軸：ターゲット領域の位置。縦軸：アクセシビリティを計算する幅)

また、siRNA が、意図せずマイクロ RNA の作用パスウェイに入り、様々な遺伝子を抑制してしまうオフターゲット効果も検出することができた。

(3) 塩基変異のもとでの二次構造変化の計算
塩基変異のもとで、RNA の構造や機能がどう変化するかを調べるのは、RNA の進化解析や機能解析で重要である。この研究では、塩基変異のもとでの、RNA 二次構造の自由エネルギーやエントロピーの変化量を網羅的に計算する手法を開発した (Rchange) (図 3)。

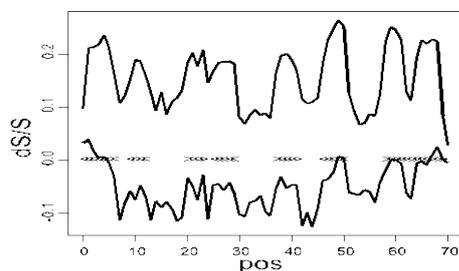


図 3. tRNA の各位置(横軸)に入った変異により、構造のエントロピー(縦軸)が増減する。

このプログラムを用いて、ヒトのミトコンドリア tRNA 遺伝子の変異と疾患の関係を調べたところ、変異による二次構造揺らぎの増加が、さまざまな疾患と関連していることが示された。

(4) RNA 結合たんぱく質の二次構造特異性
ヒトにおいて、たんぱく質コード遺伝子の 90% 以上は、何らかの条件・タイミングで、選択的スプライシングを受けることが知られている。スプライシングパターンは多くの場合、RNA 結合たんぱく質が、スプライシング領域周辺に結合することにより制御される。RNA 結合たんぱく質は、メッセンジャー RNA の一次配列的特徴を認識するが、DNA 結合性の転写因子などと異なり、二次構造的な特徴にも特異性をもつと予想される。そこで、RNA 二次構造アルゴリズムを用いて、RNA 配列の各塩基の二次構造的なコンテキストを網羅的に計算するアルゴリズムを開発し、実装した (CapR)。

このプログラムを用いて、CLIP-seq の手法により計測された、さまざまな RNA 結合たんぱく質の結合領域について、結合領域周辺の二次構造的な特徴を調べたところ、RNA 結合たんぱく質の多くは、強い二次構造的な特異性を持っていることを示すことができた。

例えば、Nova たんぱく質は、脳で特異的に発現し、スプライシングを制御する RNA 結合たんぱく質だが、このたんぱく質の結合領域は、結合モチーフ領域の直前が、非構造領域であることが多いことを示すことができた (図 4)。

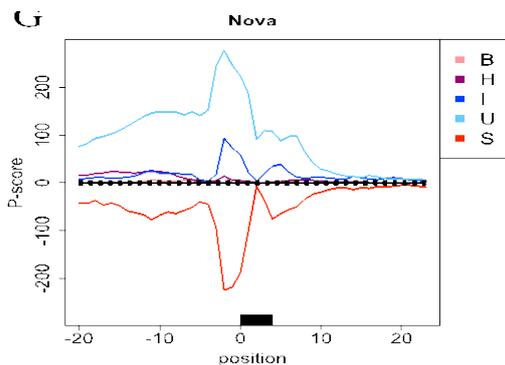


図 4. Nova 結合領域の二次構造特異性。モチーフ(黒棒)の直前で、非構造領域(U)が多く、ステム(S)領域は少ないことを示す。

5. 主な発表論文等

[雑誌論文] (計 31 件)

- ① 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) 査読有り
doi:10.1186/gb-2014-15-1-r16
- ② Hirotaka Matsumoto, and Hisanori Kiryu, “MixSIH: a mixture model for single individual haplotyping” *BMC*

- genomics 14:S5 (2013) 査読有り
doi: 10.1186/1471-2164-14-S2-S5
- ③ Junichi Iwakiri, Tomoshi Kameda, Kiyoshi Asai, Michiaki Hamada, “Analysis of base-pairing probabilities of RNA molecules involved in protein-RNA interactions.” *Bioinformatics* 29 2524-2528 (2013) 査読有り
doi: 10.1093/bioinformatics/btt453
- ④ Haruka Yonemoto, Kiyoshi Asai, Michiaki Hamada “CentroidAlign-Web: A Fast and Accurate Multiple Aligner for Long Non-Coding RNAs.” *International journal of molecular sciences* 14 6144-6156 (2013) 査読有り
doi: 10.3390/ijms14036144
- ⑤ Hisanori Kiryu, and Kiyoshi Asai, “Rchange: Algorithms for computing the energy changes of RNA secondary structures in response to base mutations.” *Bioinformatics* 28(8) 1093-1101 (2012) 査読有り
doi: 10.1093/bioinformatics/bts097
- ⑥ Kengo Sato, Yuki Kato, Tatsuya Akutsu, Kiyoshi Asai, Yasubumi Sakakibara “DAFS: simultaneous aligning and folding of RNA sequences via dual decomposition.” *Bioinformatics* 28: 24. 3218-3224 (2012) 査読有り
doi: 10.1093/bioinformatics/bts612
- ⑦ Goro Terai, Hiroaki Okida, Kiyoshi Asai, Toutai Mituyama “Prediction of Conserved Precursors of miRNAs and Their Mature Forms by Integrating Position-Specific Structural Features.” *PLoS One* 7 (2012) 査読有り
doi: 10.1371/journal.pone.0044314
- ⑧ Yuki Kato, Kengo Sato, Kiyoshi Asai, Tatsuya Akutsu, “Rtips: fast and accurate tools for RNA 2D structure prediction using integer programming.” *Nucleic Acids Res* 40: W29-W34 (2012) 査読有り
doi: 10.1093/nar/gks412
- ⑨ Michiaki Hamada, Kiyoshi Asai, “A classification of bioinformatics algorithms from the viewpoint of maximizing expected accuracy (MEA).” *J Comput Biol* 19 532-549 (2012) 査読有り
doi: 10.1089/cmb.2011.0197
- ⑩ Hisanori Kiryu, “Sufficient statistics and expectation maximization algorithms in phylogenetic tree models”, *Bioinformatics* 27, 2346-2353 (2011) 査読有り
doi:10.1093/bioinformatics/btr420
- ⑪ Michiaki Hamada, Koichiro Yamada, Kengo Sato, Martin C Frith, Kiyoshi Asai, “CentroidHomfold-LAST: accurate prediction of RNA secondary structure using automatically collected homologous sequences.” *Nucleic Acids Res* 39 W100-W106 (2011) 査読有り
doi: 10.1093/nar/gkr290
- ⑫ Hisanori Kiryu, Goro Terai, Osamu Imamura, Hiroyuki Yoneyama, Kenji Suzuki, and Kiyoshi Asai, “A detailed investigation of accessibilities around target sites of siRNAs and miRNAs”, *Bioinformatics* 27, 1788-1797 (2011) 査読有り
doi:10.1093/bioinformatics/btr27
- ⑬ Michiaki Hamada, Hisanori Kiryu, Wataru Iwasaki, and Kiyoshi Asai, “Generalized Centroid Estimators in Bioinformatics” *PLoS ONE* 6: e16450, (2011) 査読有り
doi:10.1371/journal.pone.0016450
- ⑭ Kengo Sato, Yuki Kato, Michiaki Hamada, Tatsuya Akutsu, Kiyoshi Asai “IPknot: fast and accurate prediction of RNA secondary structures with pseudoknots using integer programming.” *Bioinformatics* 27 i85-i93 (2011) 査読有り
doi: 10.1093/bioinformatics/btr215
- ⑮ Hironori Adachi, Akira Ishiguro, Michiaki Hamada, Eri Sakota, Kiyoshi Asai, Yoshikazu Nakamura “Antagonistic RNA aptamer specific to a heterodimeric form of human interleukin-17A/F.” *Biochimie* 93 1081-1088 (2011) 査読有り
doi: 10.1016/j.biochi.2011.04.003
- ⑯ Michiaki Hamada, Kengo Sato, Kiyoshi Asai “Improving the accuracy of predicting secondary structure for aligned RNA sequences.” *Nucleic Acids Res* 39 393-402 (2011) 査読有り
doi: 10.1093/nar/gkq792
- ⑰ Michiaki Hamada, Kengo Sato, Kiyoshi Asai, “Prediction of RNA secondary structure by maximizing pseudo-expected accuracy.” *BMC Bioinformatics* 11: 586 (2010) 査読有り
doi: 10.1186/1471-2105-11-586
- ⑱ Yuki Kato, Kengo Sato, Michiaki Hamada, Yoshihide Watanabe, Kiyoshi Asai, Tatsuya Akutsu, “RactIP: fast and accurate prediction of RNA-RNA interaction using integer

programming.” *Bioinformatics* 26
i460-i466 (2010) 査読有り

doi: 10.1093/bioinformatics/btq372

- ⑱ Kengo Sato, Michiaki Hamada, Toutai Mituyama, Kiyoshi Asai, Yasufumi Sakakibara, “A non-parametric Bayesian approach for predicting RNA secondary structures” *Journal of Bioinformatics and Computational Biology* 8 727-742 (2010) 査読有り
doi: 10.1142/S0219720010004926

- ⑳ Goro Terai, Aya Yoshizawa, Hiroaki Okida, Kiyoshi Asai, Toutai Mituyama, “Discovery of short pseudogenes derived from messenger RNAs. *Nucleic Acids Res* 38 1163-1171 (2010) 査読有り
doi: 10.1093/nar/gkp1098

[学会発表] (計1件)

- ① Hisanori Kiryu and Kiyoshi Asai, “On algorithms for computing thermodynamic entropy of RNA secondary structures” 第34回日本分子生物学会年会、2011年12月14日

[その他]

<http://www.ncrna.org/>

6. 研究組織

(1) 研究代表者

浅井 潔 (ASAI, Kiyoshi)
東京大学・大学院新領域創成科学研究科・教授
研究者番号：30356357

(2) 研究分担者

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

(3) 研究分担者

廣瀬 哲郎 (HIROSE, Tetsurou)
北海道大学・遺伝子病制御研究所・教授
研究者番号：30273220

(3) 連携研究者

佐藤 健吾 (SATO, Kengo)
慶應義塾大学・理工学部・講師
研究者番号：20365472