

News Release



2024 年 1 月 18 日

京都大学 iPS 細胞研究所 (CiRA)

早稲田大学

科学技術振興機構 (JST)

機能性 RNA の配列設計を支援する深層生成モデル“RfamGen”の開発

ポイント

- 特定の機能をもつ RNA 配列を学習し、同等の機能を発揮する新規の配列を生成する深層生成モデル^{注1)} “RfamGen^{注2)}”を構築した。
アルファジェン
- 変分オートエンコーダ (VAE)^{注3)} に RNA の数理モデルである共分散モデル^{注4)} を統合することで、新規 RNA 生成の性能を高め、少数データでも安定的な性能を実現した。
- RfamGen は入力データの特徴を捉えながら情報を集約しており、RNA 設計のカスタマイズに有用である。
- RfamGen による人工 RNA は、学習 RNA 群と同等の構造と機能を保持し、天然 RNA よりも高い機能活性をもつ可能性がある。
- 創薬や基礎研究における RNA 設計のコスト削減と高速化につながることを期待される。

1. 要旨

角俊輔 氏 (京都大学 iPS 細胞研究所 (CiRA) 未来生命科学開拓部門 大学院生、早稲田大学理工学術院研究室受け入れ)、浜田道昭 教授 (早稲田大学理工学術院)、齊藤博英 教授 (CiRA 同部門) は、目的の機能と構造をもつ人工 RNA 設計を支援する世界初の深層生成モデル“RfamGen”を開発しました。

RfamGen は、深層生成モデルで広く用いられている手法の一つである変分オートエンコーダ (VAE) と、RNA 配列と二次構造^{注5)} の情報から機能性 RNA を分類することのできる共分散モデルを組み合わせたもので、特定の機能と構造の特徴をもつ RNA 群の特徴を学習し、人工配列を生成することができます。

研究グループは、RfamGen が学習した RNA 群と相同な構造と機能をもつ RNA 配列が安定的に生成できることをコンピュータ上の解析と生化学実験の両方で確認しました。また、この RfamGen の性能は、深層生成モデルに共分散モデルを適用した結果であることがわかりました。さらに、RfamGen による生成配列の RNA を大規模に合成し、網羅的にその活性を検証したところ、生成配列の RNA は天然の RNA よりも高い活性を示す傾向もみられました。

RfamGen による学習結果を調べたところ、入力データの RNA 群の二次構造や機能性のモチーフなどのバリエーションを、入力データの特徴の分布として効果的に集約していました。これにより、研究者が利用したい RNA の特徴をカスタマイズして、配列を生成することが容易になります。

RfamGen により人工知能支援型の RNA 設計が可能となることで、従来の RNA 設計と比較し、開発コスト削減と高速化が実現し、核酸医薬や遺伝子治療などの RNA 創薬の研究開発に貢献することが期待されます。

この研究成果は、2024 年 1 月 18 日に英科学誌「Nature Methods」で公開される予定です。

2. 研究の背景

RNA 分子は、遺伝子の転写調節や酵素活性など、その配列に応じてさまざまな機能を発揮し、基礎研究から医療まで幅広い場面で利用されています。しかし、利用目的に適した機能をもつ RNA の塩基配列を設計することは高度な専門性と労力を要するため、RNA 配列の特徴を適切に捉えて、機能性 RNA を効率的よく設計できる、コンピュータを活用した手法の開発が期待されています。

これまでに機能性 RNA の設計法として、RNA 逆フォールディング^{注6)}が主に研究されています。しかし、この手法には、手法の性質上、その正確性や汎用性にいくつか課題がありました。今回、研究グループは、RNA の配列と二次構造を数理的に記述できる共分散モデルと VAE を統合した機能性 RNA 生成のための世界初の深層生成モデル RfamGen を開発しました。

共分散モデルは、配列から特定の二次構造を検出し、幅広い種類の RNA を RNA ファミリーとして分類でき、ゲノム配列から多くの機能性 RNA を発見することに長年使われてきました。これまでに人工 RNA の設計に共分散モデルが活用された例はありませんでしたが、研究グループは RNA 分類に有用な共分散モデルを利用することで、従来の技術的課題を解決することができるのではないかと考えました。

3. 研究結果

1) RNA ファミリー配列を設計する深層生成モデル RfamGen

RfamGen は、機能性 RNA の生成性能を高めるため、VAE に RNA の分類に用いられる共分散モデルを統合しています (図 1)。

共分散モデルは、RNA の配列と二次構造に基づき、複数の RNA 配列どうしを互いに揃うように並べること (マルチプルアライメント) ができます。RfamGen では、はじめに目的の機能をもつ既存の RNA 配列群を用意し、これを一つの RNA ファミリーと見立てて、それぞれの配列に対して共分散モデルによるアライメントを行います (図 1 左)。

RNA 群のアライメント結果を、VAE の入力データとして使用します。VAE では、入力したデータ群の特徴を学習し、入力データの特徴を確率分布として表現する「潜在空間」を構築します。RfamGen では、RNA ファミリーとみなした RNA 群の特徴を確率分布として表現する潜在空間が構築されます (図 1 中央)。

この潜在空間から出力されるデータは、入力に用いた RNA 群の共分散モデルによる特徴を示すように生成されます。出力データを共分散モデルを介して配列に再構築することで、最終的に目的の機能を獲得した人工 RNA 配列を得ることができます (図 1 右)。

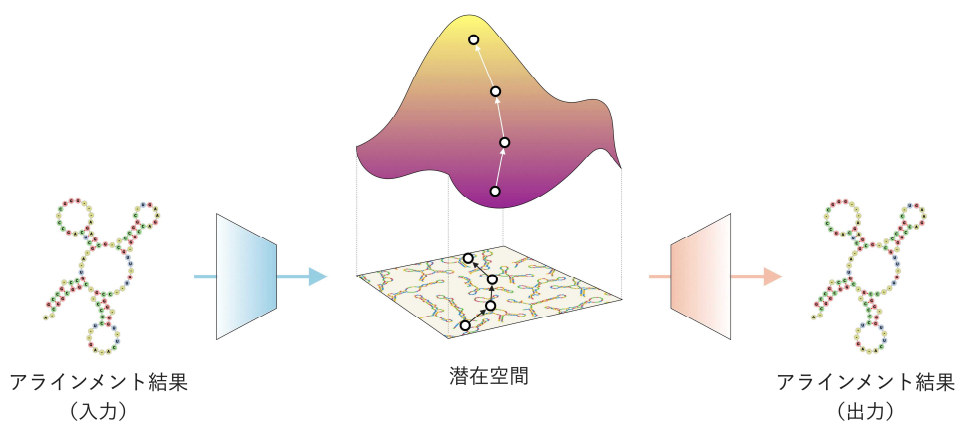


図 1 RfamGen の概要

2) 共分散モデルは深層生成モデルの性能と安定性を高める

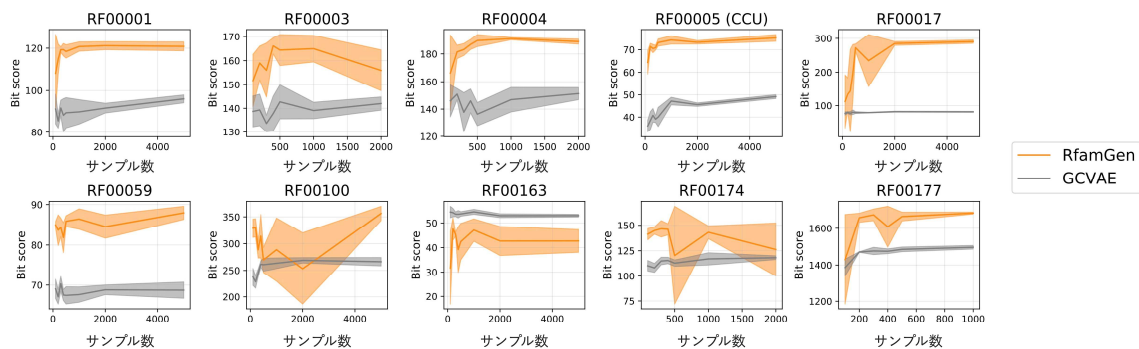
RfamGen は共分散モデルを利用することで、アライメントと二次構造を学習します。それぞれの要素が生成能力にどのように関わるかを検証するため、比較対象として次の3種のモデルを用意しました。

- ① 共分散モデルによる入力データに含まれるアライメントの情報を利用する深層生成モデル (GCVAE)
- ② 共分散モデルによる入力データに含まれる二次構造の情報を利用する深層生成モデル
- ③ 二次構造とアライメントの情報をいずれも利用しない深層生成モデル

これら3種の深層生成モデルと RfamGen に、既知の RNA ファミリー配列を学習させ、ランダムに 1,000 の配列を生成させました。そのうえで、学習に用いた RNA ファミリーに共通の構造とどの程度、相同性をもつかをコンピュータ上で計算し比較しました。その結果、RfamGen が最も良いスコアを示すことがわかりました。

次に、学習に用いる RNA 配列群のサンプル数の増減によりスコアがどのように変動するかについて研究グループは検証しました。まず、RfamGen に次いで良いスコアを出したアライメント情報のみを利用する深層生成モデル (GCVAE) を比較対象として RfamGen を検討した結果、多くの場合で RfamGen が GCVAE よりも高スコアを取ること、そして、サンプル数が少ない場合でも RfamGen の生成能力が安定して発揮されることが示唆されました (図 2)。

これらの結果から、二次構造とアライメント両方の学習が重要であり、共分散モデルを深層生成モデルと組み合わせることで、品質の高い人工 RNA 配列を安定的に生成できることが示唆されました。図 2



サンプル数による生成性能の変動

表の縦軸 Bit Score : RNA ファミリー配列らしさ

3) RfamGen による潜在空間の学習

次に、VAE の潜在空間が、学習した RNA 配列の特徴を的確に反映したものとなっているかを調べました。その結果、潜在空間を二次元に可視化したところ、RNA の二次構造にみられる多型領域や、標的タンパク質に結合する配列（モチーフ）など、入力データである RNA 配列の特徴の分布が、潜在空間に効果的に集約されていることがわかりました（図3）。このことから、RfamGen は、人工 RNA 設計支援ツールとして有用な、研究者が目的の機能と構造をもつ配列をより詳細にカスタマイズできる性能ももつことが示されました。

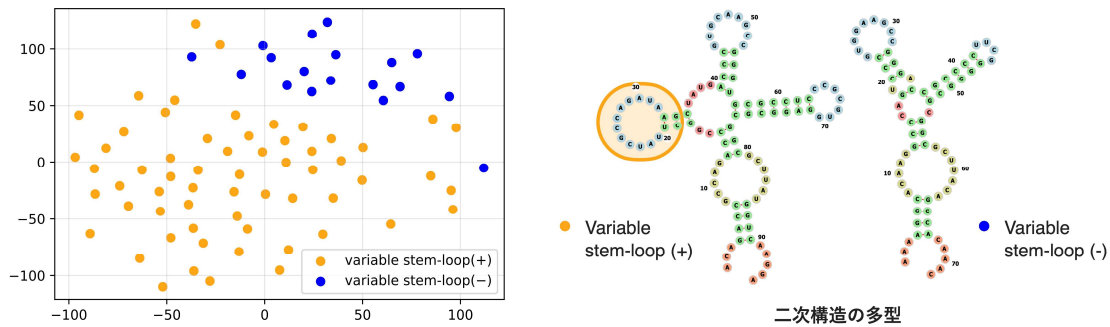


図3 潜在空間における配列情報の効果的な集約

4) RfamGen は高確率に活性配列を生成し、高活性を示す傾向がある

さらに、RfamGen を用いた人工 RNA の性能を大規模な生化学実験によって評価しました。研究グループは、RNA 分子のうち自己切断という酵素活性をもつ数百の RNA 酵素（リボザイム）の配列情報を既存のデータベースから取得し、RfamGen の学習に使用しました。その結果、RfamGen から生成された配列が、実際に天然の RNA と相同な構造（図4左）と酵素活性をもつことがわかりました。この結果から、少数のデータで学習した場合も、RfamGen が期待した配列を生成できることを実験によって確認しました。

また、低分子に結合することで自己の RNA 配列を切断する活性をもつ RNA 酵素である glmS リボザイムを例に、RfamGen により新規に 1,000 の配列を生成し、大規模に生成 RNA 配列の網羅的解析を行いました。その結果興味深いことに、RfamGen は酵素活性の高い配列を高確率に生成できることがわかりました(図4右)。

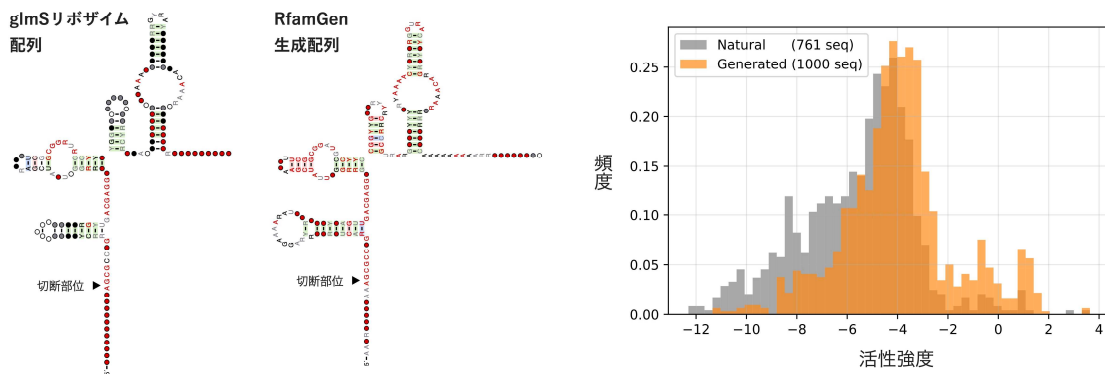


図4 RfamGen 生成配列の構造と機能の評価

左：二次構造が相同な生成配列と天然配列

右：酵素活性の性能分布の比較（オレンジ：生成配列群、灰色：天然配列群）

4. まとめと展望

本研究では、RNA 分類に利用される共分散モデルと深層生成モデルを統合し、人工 RNA 配列設計支援に用いることのできる RfamGen を構築しました。さらに、コンピュータ上と実験による性能評価によって、少数の入力データで学習した場合でも十分な性能が期待できることや、研究者が生成配列を詳細にカスタマイズ可能であること、入力データよりも高性能な人工 RNA 配列も生成しうることなど、RfamGen の有用性を示しました。今後、RfamGen を活用して、人工 RNA 設計を低コスト化、高速化し、生物学や医学など幅広い領域で RNA が活用されることに貢献することが期待されます。

5. 論文名と著者

○ 論文名

“Deep generative design of RNA family sequences”

DOI: 10.1038/s41592-023-02148-8

○ ジャーナル名

Nature Methods

○ 著者

Shunsuke Sumi^{1,2,3}, Michiaki Hamada^{3,4,5,*}, Hirohide Saito^{1,*}

*: 責任著者

○ 著者の所属機関

1. 京都大学 iPS 細胞研究所 (CiRA)
2. 京都大学大学院医学研究科
3. 早稲田大学理工学術院
4. 産業技術総合研究所生体システムビッグデータ解析オープンイノベーションラボラトリ (AIST CBBB-OIL)
5. 日本医科大学大学院医学研究科

6. 本研究への支援

本研究は、以下の支援を受けて実施されました。

- 科学技術振興機構 (JST) 戦略的創造研究推進事業 CREST 「イノベーション創発に資する人工知能基盤技術の創出と統合化」研究領域 (研究総括: 栄藤稔) 「AI アプタマー創薬プロジェクト」 (研究代表者: 浜田道昭、主たる共同研究者: 齊藤博英、 Grant 番号: JPMJCR21F1)
- 科学技術振興機構 (JST) 戦略的創造研究推進事業 CREST 「細胞操作」研究領域 (研究総括: 宮脇敦史) 「機能性 RNA・RNP 進化プラットフォームの構築と細胞制御技術の開発」 (研究代表者: 齊藤博英、主たる共同研究者: 足立俊吾、浜田道昭、 Grant 番号: JPMJCR23B3)
- 日本学術振興会 (JSPS) 科学研究費補助金 特別推進研究

7. 用語説明

注 1) 深層生成モデル

コンピュータ上の多層化したニューラルネットワークにより情報の処理を行い、学習したデータの特徴をもったデータを新たに生成するモデルのこと。

注 2) RfamGen

開発した深層生成モデルの名称。RNA ファミリー (RNA family) 配列の生成モデル (generator) であることから“RfamGen”と名付けた。

注 3) 変分オートエンコーダ (VAE)

深層生成モデルの手法の一つ。入力データを元にその特徴を確率分布として潜在空間にとらえ、入力データと似たデータを新たに生成 (出力) することができる。VAE は Variational Autoencoder の略。

注 4) 共分散モデル

RNA 配列の相同性を評価するアライメントに用いるモデル。ゲノム中の機能性 RNA の探索に長年用いられている。共分散モデルにより、RNA 配列は数千の RNA ファミリーに分類されている。

注 5) 二次構造

1 本鎖 RNA の配列に応じて局所的に形成される塩基対構造。

注 6) RNA 逆フォールディング

RNA の構造から配列を計算する方法。配列から構造を計算するフォールディングの逆の流れのため、逆フォールディングと呼ばれる。

本件担当： 京都大学 iPS 細胞研究所(CiRA)
国際広報室 三澤
TEL: 075-366-7005
Email: media[at]cira.kyoto-u.ac.jp

早稲田大学 広報室広報課
TEL: 03-3202-5454
Email: koho[at]list.waseda.jp

科学技術振興機構 広報課
TEL: 03-5214-8404
Email: jstkocho[at]jst.go.jp

(JST 事業に関すること)
科学技術振興機構 戦略研究推進部
ICT グループ
前田さち子

TEL: 03-3512-3526

Email: crest[at]jst.go.jp