
RNA 構造のライブラリ化を通じて RNA 構造ごとにおける RNA-タンパク質相互作用を 大規模に解析するシステム “FOREST” の開発 —RNA を標的とする創薬に道—

ポイント

- ゲノムに潜む RNA^{注1)}構造データを網羅的に収集した上で、数千-数万種類の RNA 構造ライブラリを作成し、タンパク質との相互作用を生化学的に大規模に解析するシステム・FOREST (Folded RNA Element Profiling with Structure library)を開発した。
- FOREST は、タンパク質が RNA と相互作用する時に、「どのような RNA の構造を好んで結合するのか？」というこれまで取得が困難であった情報を定量的かつ網羅的に取得できた。
- これまで大規模解析が困難であった RNA 高次構造である RG4 構造^{注2)}の定量に成功し、それらと相互作用する3つのタンパク質の結合特異性と強度を明らかにした。
- 得られた結合データを分析し、RG4 構造をヒトのマイクロ RNA 前駆体から新たに見出すことに成功した
- FOREST は RNA 構造と様々な相互作用を大規模に生化学解析可能な幅広いプラットフォームとなり、今後 RNA に関連する基礎研究や創薬開発など、さまざまな領域で活用が期待できる。

1. 要旨

RNA は生体内で様々な構造をつくるのが近年わかってきましたが、それらの構造がどのような機能を持つかを調べるためには、RNA1 つずつを生化学的に検証するしかありませんでした。このため、RNA 構造の効率的な研究を行うために、RNA 標的タンパク質との相互作用を、一度の実験で数千-数万種類の組み合わせについて定量する「RNA 構造が有する機能の大規模解析」の確立が待たれていました。

今回、小松リチャード馨 (CiRA 未来生命科学開拓部門 大学院生, xFOREST Therapeutics 代表取締役 CTO)、樫田俊一 (同社 代表取締役社長 CEO, 元 CiRA 特定研究員) および齊藤博英教授 (CiRA 同部門)らの研究グループは、数千-数万種類の RNA 構造が行う RNA-タンパク質相互作用を1度に大規模に解析可能なシステム、FOREST (Folded RNA Element Profiling with Structure Library) を開発しました (図 1)。FOREST では通常の実験で使用する RNA1 種類の代わりに、多様な生物種、細胞種、ウイルスゲノム等から抽出した RNA のモチーフ^{注3)}群である RNA 構造ライブラリを実験に使用します。ひいては、RNA を標的として結合するタンパク質が、RNA 構造ライブラリ中のどのような RNA 構造にどの程度結合するのかという情報を一度の実験で大量に取得することができます。

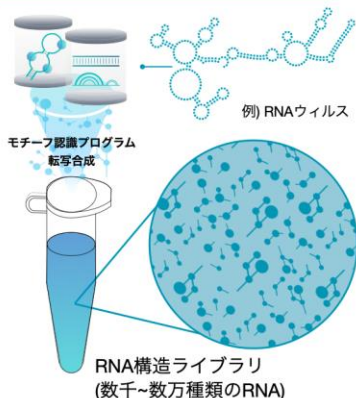
本研究では、実証実験として、複数のタンパク質を用いて FOREST の実用性の証明と、新しい相互作用の発見を試みました。結果として、RNA 構造ライブラリとタンパク質の相互作用を定量的スコアによるランクとして描出できることを確認し、機能をもつ RNA 構造の発見に至りました。さらに、FOREST は、従来法では解析が難しかった RG4 構造の大規模定量に成功しました。そして、複数の RG4 結合タンパク質に対する RNA 結合強度の定量化が可能になったため、新しく RG4 結合タンパク質の結合特性を明らかにできました。応用例として相互作用データをもとにヒトマイクロ RNA 前駆体から RG4 構造を新たに見出すことに成功しました。FOREST は概して、RNA 構造の相互作用を大規模に解析する幅広いプラットフォームとなり、今後 RNA に関連する基礎研究や創薬開発など、さまざまな領域で活用が期待されます。

この研究成果は2020年12月8日午後19時（日本時間）に英国科学誌「*Nature Communications*」でオンライン公開されました。

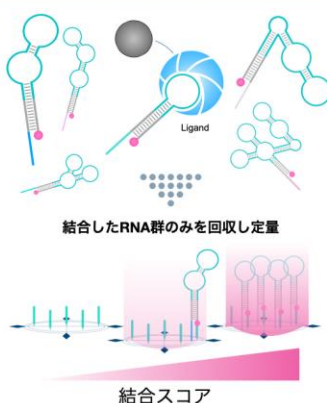
FOREST

Folded RNA Element Profiling with Structure Library

データベースからRNA構造を集めてライブラリ化する



任意の分子と構造ライブラリを反応させて、それぞれ分析する



任意の分子との結合に重要なRNA構造を特定する

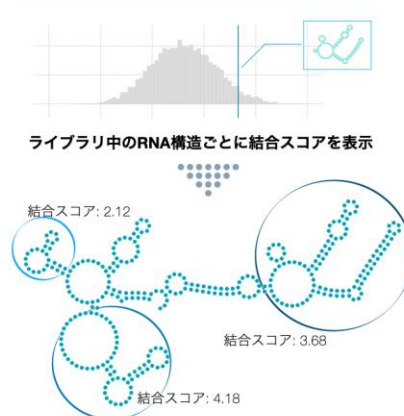


図1: FOREST の概要

FOREST は RNA 構造ライブラリを用いて一度に数千~数万種類の組み合わせの相互作用を定量する。

2. 研究の背景

RNA は配列をもとに複雑な高次構造を形成し、構造単位でユニークな機能を有することが知られています。それゆえ、多種多様な RNA の構造を特定して機能の検証を行うことは、生命を司る特別な RNA を見つけることに繋がります。これまで、並列シーケンサーと RNA 生化学実験における技術の進展によって、自然界に存在する多くの RNA 構造が明らかになってきました。しかし、RNA 構造の種類は膨大であるために、どのような特徴や機能を有しているのかわかるところを一つ一つ検証していく手法は費用と時間がかかりました。一方で、RNA 構造の機能解析に特化した大規模解析手法は未だ確立されておらず、一度にたくさんの RNA 構造を用いて実験することはできませんでした。ひいては、RNA の構造が果たす細胞の役割は多くがわかっていませんでした。

この課題を克服すべく、本研究グループは新たな解析システム・FOREST (Folded RNA Element Profiling with Structure Library) を開発しました。FOREST は、任意の分子を利用した多重化結合アッセイへと展開可能で、1 つのチューブで短時間のうちに数千~数万種類の相互作用を定量することができます。

3. 研究結果

1) FOREST (Folded RNA Element Profiling with Structure Library) の開発

FOREST による大規模解析を実現するために、まず RNA 構造の機能的な単位を大量に収集する必要があります。そこで、解析したいモチーフを定義し、抽出するためのプログラムを作成しました。その後プログラムによってコンピュータ内で天然に存在する RNA 構造のデータセットからモチーフ抽出が自動で行われ、任意のデータから RNA 構造を大量に収集できるようになりました。そして生化学実験に使用するために、試験管内転写反応を行い、一本のチューブ内で数千から数万種の RNA 構造を含む「RNA 構造ライブラリ」を生合成させました(図 2)。本研究では実証のためにヒトのマイクロ RNA 前駆体、メッセンジャー RNA、HIV-1 ウィルスの RNA 構造を収集し、幅広い種やゲノムの RNA に適用可能であることを示しました。

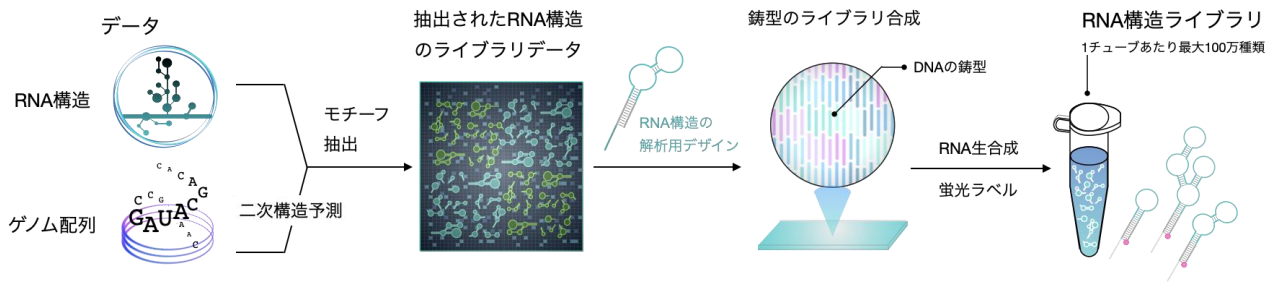


図 2: RNA 構造ライブラリの開発

生命が持ち得る RNA 構造群のデータセットから各構造のモチーフ単位をコンピュータ上で収集し、それらをライブラリとして生合成する。

あわせて、RNA 構造の定量に特化した DNA バーコードマイクロアレイ^{注4)}を開発し、RNA 構造ライブラリ中の各 RNA 構造の量をそれぞれ定量することに成功しました(図 3)。ここで、1 本のチューブで実験しても 1 回の実験で数千、数万種類の RNA 構造の機能を評価することができる大規模解析を実現できました。くわえて、並列シーケンサー解析に必要な PCR 反応や逆転写反応を阻害するため従来の方法では検出が困難であった RNA 高次構造 (RG4 構造) や疾患に関連するトリプレットリピート^{注5)}を通常の RNA 構造と同様に検出することができ、多様な種類の RNA 構造の機能を解き明かす基盤が整いました(図 4)。

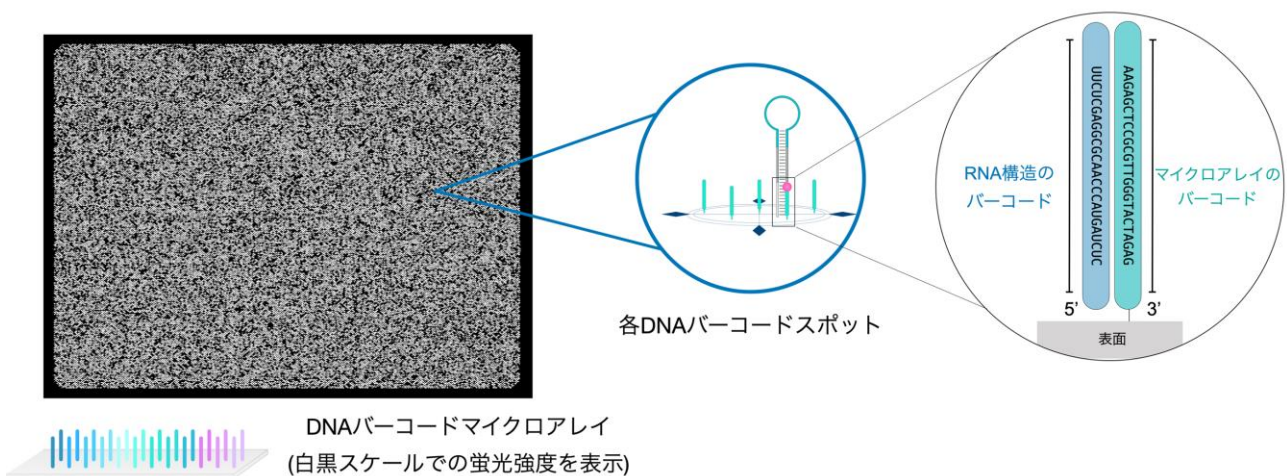
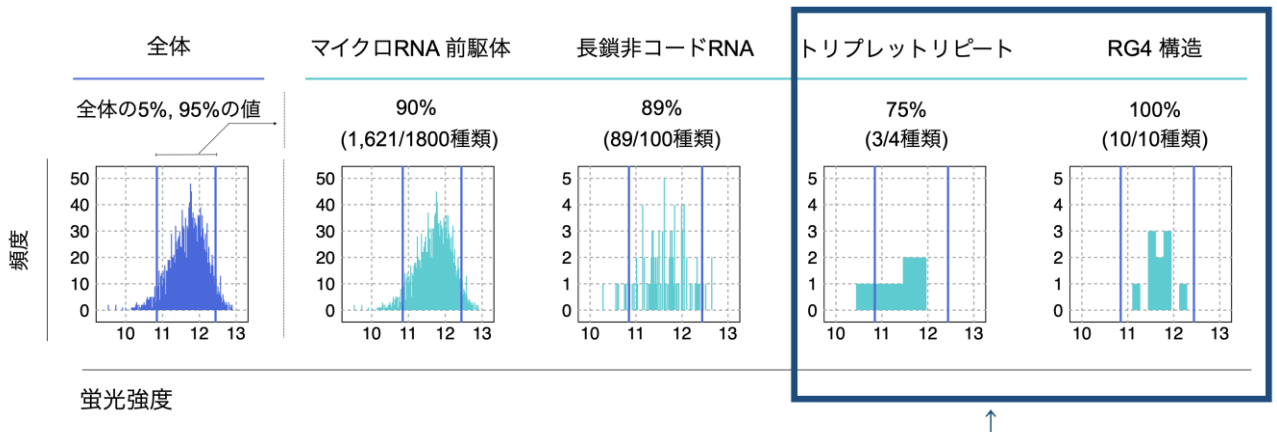


図 3: RNA 構造ライブラリの大規模定量を実現した DNA バーコードマイクロアレイの開発

RNA バーコードは相補的な DNA バーコードとしか結合しないことを利用し、各 RNA 構造を指定箇所配置、その後 RNA に付与した蛍光色素を通じて、ライブラリ中の全 RNA 構造の量を大規模に測定する。



↑
従来では逆転写・PCRが阻害され通常検出が困難

図 4: DNA バーコードマイクロアレイによる RNA 構造ライブラリの検出・定量

横軸: 蛍光強度 (RNA の量を反映)、縦軸: 頻度を表す。紫の 2 つの線は全体の 5%–95% の範囲を表す。
検出が困難な RG4 構造は、他のクラスの RNA と同等に検出可能であった。

2) RNA-タンパク質相互作用解析への展開

RNA の代表的な機能的な役割として、タンパク質と結合し、両者で細胞の運命を制御することがあげられます。しかしその結合に必要な RNA 構造が不明であるケースや、RNA 構造との結合特異性^{注6)}が不明である種類のタンパク質もありました。そこで、本研究グループはタンパク質に結合したライブラリ中の各 RNA 構造を検出し、かつその量を測定することで、FOREST でタンパク質と RNA 構造の結合を測定、結合特異性の違いを評価できると考えました。まず、このコンセプトを実証するために、ヒトのタンパク質を複数種類用いて、既知の文献や実験データと一致することを調べ、FOREST は RNA 構造-タンパク質の相互作用に適用可能ということを実証しました。例えば、U1A タンパク質は、既知のコントロールである snRNA 由来のステムループがもっともよく結合すること、そしてその変異体が結合しないこと、GCAC という配列が重要であることを確かめました。また LIN28A タンパク質においても既知の結合群である let-7 ファミリーに属するマイクロ RNA 前駆体が結合することを確認しました。同時に EMSA 法を用いて U1A と LIN28A の結合スコアと実際の結合親和性が相関することを実験的に示しました。他にも細胞由来の Roquin や EIF3 タンパク質複合体に適用し、様々な実験条件や分子で FOREST が機能したことを確認し、FOREST の有用性と汎用性を確認しました。

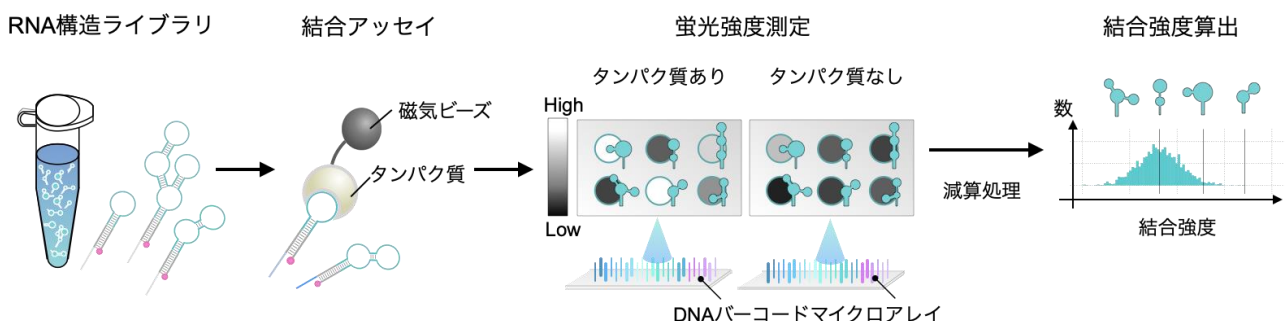


図 5: ライブラリ規模での RNA-タンパク質結合アッセイ

タンパク質は磁気を帯びたビーズに結合しているため、RNA はそれらを介して磁石に引き寄せられる。
その後、磁石に引き寄せられたタンパク質結合群を DNA バーコードマイクロアレイで定量する。
任意のタンパク質 1 つにつき、数千–数万種類の RNA に対する結合強度の数値が算出される。

最後に、RG4 の探索に FOREST のデータを応用しました。RG4 を形成できる領域の探索手法には長らく課題があり、RG4 が存在するゲノムの領域はその多くがはっきりとわかっていませんでした。さらに、どこからどこまでが RG4 なのかどうかという明確な RG4 の構造的境界の情報を伴った大規模解析は未だ実現できていません。しかし、FOREST による解析で CIRBP 結合スコアが RG4 の指標と関連したことから DHX36 が RG4 に対して高い特異性を持つと判明したことから、本研究グループは FOREST の特異的な結合データを統合することで RNA 構造ライブラリから RG4 構造を見つけられると着想しました。この仮説を実証するために、ヒトのマイクロ RNA 前駆体のループ構造のうち、RG4 を形成できるものを CIRBP と DHX36 の結合スコアを頼りに探索しました。結果的に、FOREST で見出した RG4 候補である 4 つのマイクロ RNA 前駆体ループ構造は、NMM という RG4 の指示薬で高い蛍光を示しました (図 7)。この結果はこれら 4 つの RNA は RG4 構造を有することを意味しています。これにより、FOREST は通常では判別できない特定の構造に対して高い特異性を持つ分子を明らかにし、それらデータの解析で特定の RNA 構造を新しく発見できることを証明しました。

4. 本研究の意義と今後の課題

本研究では、RNA 構造-タンパク質相互作用の解析を大規模に実行するためのシステム・FOREST を新たに開発しました。いくつかのタンパク質をモデルとして用いて、FOREST は相互作用を定量化できること、結合を行う新しい RNA 構造を発見すること、RNA 構造への結合特異性の違いを調べられること、そして RG4 の定量と発見をできることを示しました。今後、FOREST は、タンパク質の結合に重要な RNA 構造をゲノムの中から自在に探索できる幅広いプラットフォームとなり、RNA に関連する基礎研究や創薬開発の促進など、さまざまな領域で活用が期待されます。

5. 論文名と著者

○ 論文名

“RNA structure-wide discovery of functional interactions with multiplexed RNA motif library.”

○ ジャーナル名

Nature Communications

○ 著者

Kaoru R. Komatsu^{1*}, Toshiki Taya², Sora Matsumoto¹, Emi Miyashita¹, Shunnichi Kashida^{1**} and Hirohide Saito^{1**}

○ DOI

<https://doi.org/10.1038/s41467-020-19699-5>

○ 著者の所属機関

1. 京都大学 iPS 細胞研究所 (CiRA)
2. ツイストバイオサイエンス社 (サンフランシスコ、米国)

*:筆頭著者、**:責任著者

6. 本研究への支援

本研究は、下記機関より支援を受けて実施されました。

- 日本学術振興会 (JSPS) 科研費 [15H05722] [17H05601] [17J09014] [19K22387]
- 公益財団法人 三菱財団
- 公益財団法人 内藤記念科学振興財団

7. 用語説明

注 1) RNA

リボ核酸のこと。細胞内で様々な役割を担っており、DNA の塩基情報が RNA (メッセンジャーRNA) に転写され、その情報をもとにタンパク質が合成される。

注 2) RG4 (RNA G-quadruplex, グアニン四重鎖)

塩基の1つであるグアニンは隣接した位置に連続して存在するとき、フーグスティーン塩基対を形成し、四重鎖構造を形成できる。この構造では、グアニン塩基が水素結合により平面構造(G-カルテット)をなし、それが積み重なっている形をした状態で安定している。RG4 は遺伝子の発現や細胞の機能に大きく関わっていることが知られている。

注 3) モチーフ

偶然起こったとは考えにくい確率で自然界、生物の中で頻出する単位のこと。本研究では RNA の配列と構造、またはその両方から定義されるものを指す。モチーフの発見をして、その役割を調べることは RNA が行う細胞内の現象の発見に繋がる。

注 4) DNA バーコードマイクロアレイ

スライドガラスの一面にわたって多種類の DNA を極小のスポット状に無数に生やしたものを通常 DNA マイクロアレイと呼ぶ。その DNA を RNA 構造ライブラリ中の RNA が有するバーコード配列と相補的になるように本研究グループが開発・設計したもの。これにより、RNA 構造ライブラリ中の各 RNA 構造を指定したスポットに配置し、定量する大規模解析が可能になる。

注 5) トリプレットリピート

特定の 3 つの塩基を一つの単位とした繰り返し配列。一部の疾患においてこれらの繰り返し回数の伸長が認められる。本研究では GAA, CUG, CGG, CAG リピートを実際に解析に使用した。

注 6) 特異性

ここでは複数の種類の物質が混在する中で、ある特徴をもつ標的分子だけに結合できるかどうかの指標を意味する。