作成／更新日:令和８年４月２８日
課題管理番号: 25ae0121049h0005

日本医療研究開発機構
次世代治療・診断実現のための創薬基盤技術開発事業
(RNA 標的創薬技術開発)
事後成果報告書

公開

I 基本情報

研究開発課題名： (日本語) 機能解析に基づく RNA 標的創薬のための統合 DB と AI システムの構築
(英　語) Construction of an integrated DB and AI systems for RNA-targeted drug discovery based on functional analysis

研究開発実施期間：令和 3 年 12 月 1 日～令和 8 年 3 月 31 日

研究開発代表者　氏名：(日本語) 中谷　和彦
(英　語) Kazuhiko Nakatani

研究開発代表者　所属機関・部署・役職：
(日本語) 国立大学法人大阪大学・産業科学研究所・特任教授
(英　語) The University of Osaka・SANKEN・Specially Appointed Professor

II 研究開発の概要

１．概要

本課題は、疾患関連 RNA を創薬標的として活用するために、

1) RNA と化合物・タンパク質との相互作用、
2) lncRNA の機能、
3) 病理組織・オルガノイドでの RNA 発現

といった多様なデータを統合した「RNA 標的創薬統合データベース」と、オフターゲットを考慮した「標的 RNA 予測 AI」と「化合物選別 AI」を構築することで、我が国独自の RNA 標的創薬基盤を整備することを目指した。

２．背景と目的

ENCODE プロジェクトにより、ヒトゲノムの約 80%が何らかの機能をもち、その多くを non-coding RNA (ncRNA) が担うことが明らかとなり、ncRNA は新たな創薬標的として注目されている。一方、アンチセンス核酸医薬（スピンラザなど）や低分子薬 (Risdiplam) の成功により、「核酸を標的とする創薬」が現実味を帯びてきたが、以下の課題が残る。

- どの RNA、どの配列・構造を標的とすべきか
- オフターゲットをどう回避するか
- どのような低分子が特定の RNA 構造に選択的に結合するか

欧米ではベンチャーやメガファーマが RNA 標的創薬に参入している一方、我が国の取り組みは遅れており、**日本発の RNA 標的創薬基盤技術を整備し、国際競争力を確保すること**が喫緊の課題であった。

本研究では、

- 疾患関連 RNA・構造エレメント・モダリティ情報を統合した「統合 DB」
- オフターゲットを考慮して標的 RNA・標的部位を提示する「標的予測 AI」
- 大規模化合物ライブラリから標的 RNA 構造に結合しやすい化合物を選別する「化合物選別 AI」

を構築し、製薬企業に提供することで、シード探索からリード創出までの期間短縮とコスト削減を狙った。

３．全体構想とデータフロー

研究開発は６つの項目 A〜F で構成した

1. **A：RNA と化合物や蛋白質等との相互作用解析**
2. **B：ノックダウン等による lncRNA 機能解析**
3. **C：病理組織・オルガノイドにおける RNA 発現解析**
4. **D：RNA 標的創薬統合データベースの構築**
5. **E：RNA 創薬標的予測 AI システムの開発**
6. **F：疾患関連 lncRNA に対する DB・予測システムの検証・実証**

A〜C で取得した独自データと、外部オミクスデータベースを D で統合し「統合 DB」を構築し、これを基盤として E で標的予測 AI を構築・高度化し、F で実際の疾患関連 lncRNA・化合物を用いてシステムの有効性を検証した。同時に、A で得られる RNA-化合物相互作用データを用いて化合物選別 AI を構築した。

４．研究項目 A〜C：データ取得と基礎情報の整備

A)　RNA と化合物や蛋白質等との相互作用解析

A-1：RNA と化合物の相互作用解析

- 参画企業と連携し、構造多様性をカバーする標準化合物を選定
- lncRNA に頻出するステムループ／ヘアピンループなどの標準 RNA 構造エレメントを設定
- BIAcore 8K+等の SPR 装置を用いて、標準 RNA×標準化合物の相互作用を体系的に取得
- 標準 RNA の一部をランダム化した「部分ランダム RNA」と鍵化合物固定化樹脂を用いて、プルダウン＋NGS により「鍵化合物が好む RNA 配列モチーフ」を網羅的に取得

A-2：NMR による相互作用解析と結合サイト同定

- 47 種類のテスト用 RNA と 5 種類の化合物について NMR 法による相互作用解析を実施し、イミノプロトンの化学シフト変化・線幅変化等を定量化し、データベースに統合
- A-1 で相互作用が確認された RNA-化合物ペアについて、標準化した条件下で NMR スペクトルを取得し、相互作用を検証
- 残基特異的安定同位体標識 RNA を用い、効率的に結合残基を推定する手法を開発

B) ノックダウン等による lncRNA の機能解析

- 難溶性 RNA-seq により、細胞核内でハブ構造を形成する多数の arcRNA 候補を各種がん細胞株・臨床検体から網羅的に抽出
- CRISPRi を用いた機能スクリーニング系を整備し、lncRNA を網羅的にノックダウン解析により、癌細胞株の細胞増殖に必要な arcRNA を含む機能性 lncRNA を選別
- 核内でハブ構造体を形成する arcRNA を同定し、その相互作用タンパク質と標的遺伝子候補を取得。
- arcRNA の発現やハブ構造体形成に関わる RNA 領域の同定

C) 病理組織の RNA 発現データ取得

- 慢性間質性肺炎病理組織における再生領域と気管支上皮化生領域における lncRNA 発現を網羅解析
- in situ hybridization で病理検体における発現検証
- 肺由来オルガノイド（肺胞・気道）を樹立し、炎症誘導などにより肺胞→気管支上皮への化生を in vitro で再現、その過程での lncRNA 発現変化を解析

- 難溶性 RNA-seq により、病理組織・オルガノイドにおける arcRNA 発現情報の取得
- 候補 lncRNA をノックダウンおよび過剰発現しオルガノイドでの化生変化を解析し、機能性 lncRNA を同定
- 機能性 lncRNA と推定されるものについてそのターゲットとの関係を免疫染色＋in situ hybridization で検証

５．研究項目 D〜F：統合 DB・AI システムの構築と実証

D）RNA 標的創薬統合データベースの構築

- 公共 DB（TCGA、ENCODE 等）からの RNA 発現・オミクスデータに加え、A〜C で得られる独自データを統合し、疾患関連 RNA・創薬標的候補 RNA・構造エレメントの情報を収載した**統合 DB** を構築
- 疾患関連 RNA の網羅的選定、RNA 修飾、RNA 結合タンパク質結合、RNA 構造などのオミクス情報に加え、RNA 構造安定性・保存性指標の導出・登録を行なった
- ユーザがシナリオに沿って絞り込みを行うインターフェースを開発
- ユーザ管理などのセキュリティ技術を実装

E）RNA 創薬標的予測 AI システムの開発

- 任意の RNA 構造の検索を行うソフトウェアを開発し、オフターゲット、オンターゲット予測に寄与するシステムを構築
- AI 技術を用いて保存されている RNA 構造・領域を同定する新規手法を開発
- Cryptic Binding site を予測する AI 技術を開発
- オミクス情報を統合し、疾患特異的遺伝子発現変化の予測と解釈する AI 手法を開発

F）疾患関連 lncRNA に対する有効性の検証・実証

- **慢性間質性肺炎と肺癌**に対象疾患を絞り、統合 DB・標的予測 AI・化合物選別 AI の有効性を検証
- 慢性間質性肺炎・肺癌関連 lncRNA を探索・機能解析し、
    - それらに結合する化合物の探索と相互作用検証、
    - オルガノイド等疾患モデルでの機能影響評価、
    - 得られた結果をもとに化合物選別 AI・標的予測 AI を至適化

６．期待される成果と社会的インパクト

- RNA と化合物・タンパク質の網羅的高品質データ、lncRNA 機能、病理組織発現データを統合した**日本独自の RNA 標的創薬基盤 DB** を構築。
- オフターゲットを考慮した標的 RNA・標的エレメントの予測 AI、および標的 RNA 構造に結合する化合物を効率選別する AI により、
    - シード／ヒット化合物のヒット率向上、
    - リード化合物取得までの期間・コストの削減、

    を実現し、国内製薬企業の国際競争力強化に貢献する。
- arcRNA や異常伸長リピート RNA、などを標的とすることで、従来治療介入が難しかった遺伝性神経変性疾患、がんのエピジェネティック制御等に対する新規治療薬開発の可能性を拓く。
- 純国産技術による RNA 標的医薬の開発は、高額医薬品の輸入依存を軽減し、医療費の国内循環・外貨獲得に資するエコシステム構築へとつながる。

II. Overview of the Research and Development

1. Summary

This project aimed to establish a Japan-origin platform for RNA-targeted drug discovery by developing:

1. a “Comprehensive Database for RNA-Targeted Drug Discovery” that integrates diverse data on
   - RNA-small molecule / RNA-protein interactions,
   - lncRNA functions, and
   - RNA expression in pathological tissues and organoids; and
2. an “RNA Target Prediction AI” and a “Compound Prioritization AI” that explicitly take off-target effects into account.

2. Background and Objectives

The ENCODE project has revealed that approximately 80% of the human genome is functionally relevant, and that a large fraction of this functionality is mediated by non-coding RNAs (ncRNAs), which are now recognized as promising new drug targets. In parallel, the clinical success of antisense oligonucleotide drugs (e.g., nusinersen) and small-molecule drugs (e.g., risdiplam) has made “drug discovery targeting nucleic acids” a realistic strategy. However, several key challenges remain:

- Which RNA species, and which sequences or structural elements, should be targeted?
- How can off-target effects be avoided?
- What kinds of small molecules can selectively bind to specific RNA structures?

While venture companies and major pharmaceutical firms in Europe and the United States are actively entering the RNA-targeted drug discovery space, efforts in Japan have lagged behind. Establishing a Japan-origin technological foundation for RNA-targeted drug discovery and securing international competitiveness has therefore been an urgent issue.

In this study, we set out to build and provide to pharmaceutical companies:

- a “Unified Database” that integrates information on disease-associated RNAs, structural elements, and therapeutic modalities,
- an “RNA Target Prediction AI” that proposes target RNAs and target sites while considering off-target effects, and
- a “Compound Prioritization AI” that selects compounds likely to bind target RNA structures from large-scale compound libraries,

with the goal of shortening the time and reducing the cost required to progress from seed identification to lead generation.

3. Overall Concept and Data Flow

The research and development program consisted of six components, A-F:

1. A: Analysis of interactions between RNA and small molecules
2. B: Functional analysis of lncRNAs by knockdown and related approaches
3. C: Analysis of RNA expression in pathological tissues and organoids
4. D: Construction of a unified database for RNA-targeted drug discovery
5. E: Development of an AI system for predicting RNA drug targets
6. F: Validation and demonstration of the database and prediction system for disease-associated lncRNAs

Proprietary data obtained in A-C, together with external omics databases, were integrated in D to construct the “Unified Database.” Based on this database, we developed and refined the RNA Target Prediction AI in E, and then evaluated the performance of the system in F using actual disease-

associated lncRNAs and compounds. In parallel, we constructed the Compound Prioritization AI using RNA-compound interaction data obtained in A.

4. Research Components A-C: Data Acquisition and Preparation of Fundamental Information

A) Analysis of interactions between RNA and small molecules / proteins

A-1: Analysis of RNA-small molecule interactions

- In collaboration with participating companies, standard compounds covering structural diversity were selected.
- Standard RNA structural elements, such as stem-loops and hairpin loops frequently found in lncRNAs, were defined.
- Using SPR instruments such as Biacore 8K+, interaction data for combinations of standard RNAs and standard compounds were systematically acquired.
- "Partially randomized RNAs," generated by randomizing parts of the standard RNAs, and resins immobilized with key compounds were used for pull-down followed by NGS to comprehensively identify "RNA sequence motifs preferred by key compounds."

A-2: Interaction analysis and binding-site identification by NMR

- Interaction analyses by NMR were performed for 47 test RNAs and 5 compounds. Changes in imino proton chemical shifts, line broadening, and related parameters were quantified and integrated into the database.
- For RNA-compound pairs whose interaction had been confirmed in A-1, NMR spectra were acquired under standardized conditions to validate the interactions.
- A method was developed to efficiently identify binding residues using residue-specifically isotope-labeled RNAs.

B) Functional analysis of lncRNAs by knockdown and related approaches

- Using insoluble RNA-seq, numerous candidate arcRNAs forming nuclear hub structures were comprehensively extracted from various cancer cell lines and clinical specimens.
- A CRISPRi-based functional screening system was established, and genome-wide knockdown analysis of lncRNAs was performed to identify functional lncRNAs, including arcRNAs required for the proliferation of cancer cell lines.
- arcRNAs that form nuclear hub structures were identified, and their interacting proteins and candidate target genes were obtained.
- RNA regions involved in arcRNA expression and hub formation were determined.

C) Acquisition of RNA expression data from pathological tissues

- Genome-wide analysis of lncRNA expression was carried out in regenerative regions and bronchiolar metaplastic regions within lung biopsy specimens from patients with chronic interstitial pneumonia.
- Expression in pathological specimens was validated by in situ hybridization.
- Lung-derived organoids (alveolar and airway) were established, and transdifferentiation from alveolar to bronchiolar epithelium was reproduced in vitro by inflammatory stimulation; changes in lncRNA expression during this process were analyzed.
- Insoluble RNA-seq was used to obtain information on arcRNA expression in pathological tissues and organoids.
- Candidate lncRNAs were knocked down or overexpressed in organoids, and changes in metaplastic transformation were analyzed to identify functional lncRNAs.

- For lncRNAs presumed to be functional, their relationship with target molecules was examined by immunostaining combined with in situ hybridization.

5. Research Components D-F: Construction and Demonstration of the Unified DB and AI Systems

D) Construction of a unified database for RNA-targeted drug discovery

- A unified database was constructed that integrates RNA expression and omics data from public databases (TCGA, ENCODE, etc.) with proprietary data obtained in A-C, and stores information on disease-associated RNAs, candidate drug target RNAs, and structural elements.
- In addition to comprehensive selection of disease-associated RNAs, omics information on RNA modifications, RNA-binding protein interactions, RNA structure, etc. was incorporated, and indices of RNA structural stability and conservation were derived and registered.
- A user interface was developed that enables stepwise narrowing down of candidates along user-defined scenarios.
- Security functions, including user management, were implemented.

E) Development of an AI system for predicting RNA drug targets

- Software was developed to search arbitrary RNA structures, and a system was constructed to support on-target and off-target prediction.
- A novel AI-based method was developed to identify conserved RNA structures and regions.
- AI techniques were developed to predict cryptic binding sites.
- By integrating omics information, AI methods were developed to predict and interpret disease-specific changes in gene expression.

F) Validation and demonstration for disease-associated lncRNAs

- The target diseases were focused on chronic interstitial pneumonia and lung cancer, and the effectiveness of the Unified DB, Target Prediction AI, and Compound Prioritization AI was evaluated.
- Disease-associated lncRNAs in chronic interstitial pneumonia and lung cancer were identified and functionally analyzed, and:
    - compounds binding to these lncRNAs were searched and their interactions validated,
    - functional effects were evaluated in organoid and other disease models, and
    - based on the obtained results, the Compound Prioritization AI and Target Prediction AI were further optimized.

6. Expected Outcomes and Societal Impact

- A Japan-origin foundational database for RNA-targeted drug discovery was constructed by integrating high-quality comprehensive data on RNA-small molecule / RNA-protein interactions, lncRNA functions, and RNA expression in pathological tissues.
- Through an AI system that predicts target RNAs and target structural elements while accounting for off-targets, and an AI system that efficiently prioritizes compounds binding to target RNA structures, we aim to:
    - increase the hit rate for seed and hit compounds, and
    - shorten the time and reduce the cost required to obtain lead compounds,

thereby strengthening the international competitiveness of domestic pharmaceutical companies.

- By targeting arcRNAs, aberrantly expanded repeat RNAs, and related species, this platform opens the possibility of developing novel therapeutics for diseases that have been difficult to treat, such as hereditary neurodegenerative diseases and cancers driven by epigenetic

dysregulation.

- The development of RNA-targeted therapeutics based on purely domestic technologies will reduce dependence on imported high-cost medicines and contribute to building an ecosystem that promotes domestic circulation of medical expenditures and acquisition of foreign currency.