

令和 4 年 6 月 26 日現在

機関番号：32689

研究種目：若手研究(A)

研究期間：2016～2019

課題番号：16H05879

研究課題名（和文）機能エレメントと深層学習に基づく長鎖ノンコーディングRNAの機能分類

研究課題名（英文）Functional classification of long noncoding RNAs based on functional elements and deep learning

研究代表者

浜田 道昭 (Hamada, Michiaki)

早稲田大学・理工学術院・教授

研究者番号：00596538

交付決定額（研究期間全体）：（直接経費） 18,000,000円

研究成果の概要（和文）：タンパク質に翻訳されずにRNA自体が機能を有する長鎖ノンコーディングRNA（lncRNA）の機能を明らかにするために、RNAの配列・構造・修飾・生体高分子との相互作用などの「機能エレメント」に着目し情報学的な観点から多くの研究を行った。例えば、ジャンクだと考えられていたリピート配列がlncRNAの組織特異的な発現に寄与していたり、タンパク質やDNAとの相互作用に寄与していることを明らかにすることによりlncRNAの機能の解明を試みた。これらを通してlncRNAの機能の分類に寄与する情報基盤技術確立し、広く公開を行った。

研究成果の学術的意義や社会的意義

現在機能のわかっていないlncRNAの中には、がんや神経変性疾患などの重篤な疾患と関連するものが存在すると考えられている。今回開発した技術や解析結果により、今後lncRNAと疾患との関連性が明らかになった暁には、RNAをターゲットにした医学・薬学研究などに発展する可能性が高い。

研究成果の概要（英文）：To elucidate the function of long non-coding RNAs (lncRNAs), which are not translated into proteins but have their own functions, we have conducted many studies from an informatics perspective, focusing on "functional elements" such as RNA sequences, structures, modifications, and interactions with biological macromolecules. For example, we attempted to elucidate the functions of lncRNAs by clarifying that repeat sequences, which were thought to be junk, contribute to the tissue-specific expression of lncRNAs and to their interactions with proteins and DNA. Through these efforts, we have established information infrastructure technology that contributes to the classification of lncRNA functions, and have made it widely available to the public.

研究分野：バイオインフォマティクス

キーワード：lncRNA ノンコーディングRNA

## 様式 C - 19 , F - 19 - 1 , Z - 19 ( 共通 )

### 1 . 研究開始当初の背景

近年の研究により、ヒトで数万種類以上の、タンパク質に翻訳されない「長鎖ノンコーディング RNA (lncRNA)」が存在し、細胞内における様々な機能を担っていることが示唆されているが、その機能の詳細が解明されているものは 1%程度であると考えられている。一方で lncRNA の一部はがんや神経変性疾患などの疾患に寄与しているものも見つかってきている。残された lncRNA の機能の全貌を解明することは現代分子生物学における喫緊の課題である。

### 2 . 研究の目的

本研究では、RNA バイオインフォマティクスに関する多くの研究実績を有する申請者が、lncRNA を『機能エレメント』に基づき分類しその機能を推定するために、深層学習を利用した情報技術の開発を行う。開発した情報技術はソフトウェアやデータベースなどのリソースとして広く公開し lncRNA 研究に役立ててもらう。

### 3 . 研究の方法

機能エレメントに基づいた lncRNA の機能解明を行う。機能エレメントとしては、RNA の配列(トランスポゾンなどのリピート要素を含む)・構造、RNA 修飾、他の生体高分子との相互作用を対象とする。

### 4 . 研究成果

#### 4.1 RNA の配列に関する研究

lncRNA は mRNA に比べて組織特異的に発現するものが多いことが知られている。一方で、lncRNA の配列には従来はジャンク(ごみ)だと考えられていたトランスポゾンが多く含まれることも知られていた(mRNA にはほとんど含まれていない)。本研究では、lncRNA の組織特異的発現にトランスポゾンが関連しているという仮設のもので、lncRNA とトランスポゾンのアノテーションと発現データを用いて、統計解析により組織特異的発現に関与するトランスポゾンを網羅的に同定した。その結果、例えば ERV1 / Alu は精巣での特異的発現に関与していることがわかった。さらに、L1PA2 は胎盤特異的発現に関与し、さらにプロモータとして働いている可能性を示唆した。

細胞質および核局在 RNA-seq データ解析によって、スプライシングが RNA の細胞質局在を促進することが観測された。更に選択的スプライシングによるイントロンリテンションが起こると、RNA が細胞核に局在しやすくなることが発見した。リテンしたイントロ配列上の RNA 結合タンパク質の解析によって、CUG 反復配列に結合し、核局在化シグナルを持つタンパク質 MBNL1 が核局在因子の役割があると考察した。本研究は選択的スプライシング異常を伴う疾患の解明に貢献することが期待される。

#### 4.2. RNA 修飾に関する研究

lncRNA の多くに見られる修飾の一つである N6-メチルアデニン(m6A)に関して、情報解析技術を作った。

第一に、miCLIP-Seq のデータから m6A を含む配列を予測し、その周辺の生物学的特徴を明らかにするために、DeepM6Aseq と名付けた深層学習フレームワークを実装した。DeepM6Aseq は、他の機械学習分類法と比較して、優れた性能を示した。さらに、m6A を含むゲノム領域を特定する m6A-Seq データを用いた独自のテストにより、本モデルが m6A を含む配列の予測に優位性を持つことが明らかになった。DeepM6Aseq で学習されたモチーフは、既知の m6A リーダーに対応していることが示された。さらに、深層学習モデルの顕著性マップを活用することで、m6A 部位の位置を可視化できることがわかった。

第二に、MeRIP-Seq の濃縮領域を検出し、混合負二項モデルに基づいてシグナル比率を推定するソフトウェア「model-based analysis and inference of MeRIP-Seq (MoAIMS)」を開発した。MoAIMS はトランスクリプトーム免疫沈降シーケンス実験のために設計されているため、異なる RNA シーケンスプロトコルと互換性がある。MoAIMS は、他のツールと比較して、優れた処理速度と競争力のある性能を達成した。MoAIMS を m6A の研究に適用した場合、検出された濃縮領域は m6A の既知の生物学的特徴を含んでいる。さらに、m6A 処理(m6A メチルトランスフェラーゼの損傷)データセットに対して MoAIMS から推定されるシグナル比率は、実験観察と一致する減少傾向を示し、シグナル比率が処理効果の直感的な指標として使用できることが示唆された。

第三に、統合的な計算フレームワークを用いて、m6A に関連する RBP の候補を同定することを

試みた。本研究で新しく提案したフレームワークは、エンリッチメント解析と分類モデルで構成されている。このフレームワークを用いて、RBP の結合データを活用し、独立した研究による再現性の高い m6A 領域を解析した。エンリッチメント解析により、YTH ドメインを含む既知の m6A 関連 RBP が同定され、さらに RBM3 がマウスの m6A 関連 RBP の候補として同定された。さらに、同定された m6A 関連 RBP には、遺伝子発現レベルではなく、タンパク質発現レベルで有意な相関が観察された。一方、RBP の結合データを用いて、再現性のある m6A 領域に対して Random Forest 分類モデルを構築した。この結果、RBP の結合データを利用することで、配列レベルを超えて m6A と m6A に関連する RBP の相互作用を推定できることが示唆された。以上のように、我々は、既知および潜在的な m6A 関連 RBP を同定するための統合的な計算機フレームワークを設計した。本解析が、m6A と RNA 修飾の研究において、より多くの知見をもたらすことを期待する。

#### 4.3. 他の生体分子との相互作用

**DNA との相互作用：**R ループ (R-loop) とは、2 本鎖 DNA のうちの片方の鎖が 1 本鎖 RNA と相補的な結合によってハイブリッドを形成し、もう片方の DNA 鎖が 1 本鎖である状態が蓄積するという構造である。近年、R ループは真核生物のゲノムに広く分布しており、遺伝子制御やゲノムの完全性に関与していることが示唆されている。しかし、R ループ構造の形成についての理解はまだ限られている。我々は今までジャンクだと考えられていたゲノム反復配列が、R ループ構造の形成に寄与していることを明らかにした。ヒト、ミミバエとシロイヌナズナでは、satellite、LINE と DNA 反復要素がそれぞれ R ループに特異的に富むことが観測された。また、種間における R ループが、low-complexity や simple repeat の領域に形成しやすい傾向を示している。R ループの形成に関わる反復要素は、発生段階によって異なることも明らかにした。

**RNA 結合タンパク質 (RBP) との相互作用：**近年、イントロンやロングノンコーディング RNA (lncRNA) に埋め込まれた繰り返し配列が RNA 結合タンパク質 (RBP) の標的となり、RNA スプライシングや転写調節などの生命現象に寄与することが明らかにされている。これらのことから、繰り返し配列由来の RNA は RBP の足場や機能素子として重要であることが示唆されている。しかし、繰り返し配列由来の RNA の全体的な機能配列は十分に解明されていない。ここでは、ENCODE eCLIP データをもとに、RBP との結合パターンを解析することで、機能的と考えられるリピート由来 RNA を示す eCLIP の全リードをリピート配列にマッピングしたところ、K562 細胞では平均 10.75 %、HepG2 細胞では平均 7.04 % のリードがリピートに濃縮 (コントロールに比べて少なくとも 2 倍) されていることが確認された。これらのデータを用いて、long interspersed element 1 (LINE1) 配列のセンス鎖およびアンチセンス鎖上に機能的な RNA エlement を予測した。さらに、他のトランスポーザブルエレメント (TE) ファミリーに由来する断片上に、いくつかの新しい RBP セットを見出した。これらの断片の中には、特異的で安定した二次構造を示し、遺伝子や lncRNA のイントロンに挿入されるものも見いだされた。これらの結果は、リピート由来の RNA 配列が内在性ノンコーディング RNA の機能的 RNA エlement の有力な候補であることを示唆している。

## 5. 主な発表論文等

〔雑誌論文〕 計15件（うち査読付論文 15件 / うち国際共著 0件 / うちオープンアクセス 6件）

1. 著者名 Tani Hidenori, Matsutani Taro, Aoki Hiroshi, Nakamura Kaoru, Hamaguchi Yu, Nakazato Tetsuya, Hamada Michiaki	4. 巻 512
2. 論文標題 Identification of RNA biomarkers for chemical safety screening in neural cells derived from mouse embryonic stem cells using RNA deep sequencing analysis	5. 発行年 2019年
3. 雑誌名 Biochemical and Biophysical Research Communications	6. 最初と最後の頁 641 ~ 646
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.bbrc.2018.11.141	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Fukunaga Tsukasa, Iwakiri Junichi, Ono Yukiteru, Hamada Michiaki	4. 巻 10
2. 論文標題 LncRRsearch: A Web Server for lncRNA-RNA Interaction Prediction Integrated With Tissue-Specific Expression and Subcellular Localization Data	5. 発行年 2019年
3. 雑誌名 Frontiers in Genetics	6. 最初と最後の頁 eCollection
掲載論文のDOI (デジタルオブジェクト識別子) 10.3389/fgene.2019.00462	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Matsumura Yuko, Ito Yasuhiko, et al. Hamada Michiaki, Orimo Akira	4. 巻 2
2. 論文標題 Stromal fibroblasts induce metastatic tumor cell clusters via epithelial-mesenchymal plasticity	5. 発行年 2019年
3. 雑誌名 Life Science Alliance	6. 最初と最後の頁 eCollection
掲載論文のDOI (デジタルオブジェクト識別子) 10.26508/lisa.201900425	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Matsutani Taro, Ueno Yuki, Fukunaga Tsukasa, Hamada Michiaki	4. 巻 35
2. 論文標題 Discovering novel mutation signatures by latent Dirichlet allocation with variational Bayes inference	5. 発行年 2019年
3. 雑誌名 Bioinformatics	6. 最初と最後の頁 4543 ~ 4552
掲載論文のDOI (デジタルオブジェクト識別子) 10.1093/bioinformatics/btz266	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Shi Hangchuan, Sun Yin, He Miao, Yang Xiong, Hamada Michiaki, Fukunaga Tsukasa, Zhang Xiaoping, Chang Chawnshang	4. 巻 39
2. 論文標題 Targeting the TR4 nuclear receptor-mediated lncTASR/AXL signaling with tretinoin increases the sunitinib sensitivity to better suppress the RCC progression	5. 発行年 2019年
3. 雑誌名 Oncogene	6. 最初と最後の頁 530 ~ 545
掲載論文のDOI (デジタルオブジェクト識別子) 10.1038/s41388-019-0962-8	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Fukunaga Tsukasa, Hamada Michiaki	4. 巻 -
2. 論文標題 Computational approaches for alternative and transient secondary structures of ribonucleic acids	5. 発行年 2019年
3. 雑誌名 Briefings in Functional Genomics	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) 10.1093/bfgp/ely042	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Zeng Chao, Hamada Michiaki	4. 巻 19
2. 論文標題 Identifying sequence features that drive ribosomal association for lncRNA	5. 発行年 2018年
3. 雑誌名 BMC Genomics	6. 最初と最後の頁 906
掲載論文のDOI (デジタルオブジェクト識別子) 10.1186/s12864-018-5275-8	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Zhang Yiqian, Hamada Michiaki	4. 巻 19
2. 論文標題 DeepM6ASeq: prediction and characterization of m6A-containing sequences using deep learning	5. 発行年 2018年
3. 雑誌名 BMC Bioinformatics	6. 最初と最後の頁 524
掲載論文のDOI (デジタルオブジェクト識別子) 10.1186/s12859-018-2516-4	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Nishida Shimpei, Sakuraba Shun, Asai Kiyoshi, Hamada Michiaki	4. 巻 -
2. 論文標題 Estimating energy parameters for RNA secondary structure predictions using both experimental and computational data	5. 発行年 2018年
3. 雑誌名 IEEE/ACM Transactions on Computational Biology and Bioinformatics	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/TCBB.2018.2813388	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Fukunaga Tsukasa, Hamada Michiaki	4. 巻 25
2. 論文標題 A Novel Method for Assessing the Statistical Significance of RNA-RNA Interactions Between Two Long RNAs	5. 発行年 2018年
3. 雑誌名 Journal of Computational Biology	6. 最初と最後の頁 976 ~ 986
掲載論文のDOI (デジタルオブジェクト識別子) 10.1089/cmb.2017.0260	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Chishima Takafumi, Iwakiri Junichi, Hamada Michiaki	4. 巻 9
2. 論文標題 Identification of Transposable Elements Contributing to Tissue-Specific Expression of Long Non-Coding RNAs	5. 発行年 2018年
3. 雑誌名 Genes	6. 最初と最後の頁 23 ~ 23
掲載論文のDOI (デジタルオブジェクト識別子) 10.3390/genes9010023	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Chishima Takafumi, Iwakiri Junichi, Hamada Michiaki	4. 巻 9
2. 論文標題 Identification of Transposable Elements Contributing to Tissue-Specific Expression of Long Non-Coding RNAs	5. 発行年 2018年
3. 雑誌名 Genes	6. 最初と最後の頁 23 ~ 23
掲載論文のDOI (デジタルオブジェクト識別子) 10.3390/genes9010023	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Takeda Taikai, Hamada Michiaki	4. 巻 34
2. 論文標題 Beyond similarity assessment: selecting the optimal model for sequence alignment via the Factorized Asymptotic Bayesian algorithm	5. 発行年 2017年
3. 雑誌名 Bioinformatics	6. 最初と最後の頁 576 ~ 584
掲載論文のDOI (デジタルオブジェクト識別子) 10.1093/bioinformatics/btx643	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Iwakiri Junichi, Terai Goro, Hamada Michiaki	4. 巻 12
2. 論文標題 Computational prediction of lncRNA-mRNA interactions by integrating tissue specificity in human transcriptome	5. 発行年 2017年
3. 雑誌名 Biology Direct	6. 最初と最後の頁 15
掲載論文のDOI (デジタルオブジェクト識別子) 10.1186/s13062-017-0183-4	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Fukunaga Tsukasa, Hamada Michiaki	4. 巻 33
2. 論文標題 Rlblast: an ultrafast RNA-RNA interaction prediction system based on a seed-and-extension approach	5. 発行年 2017年
3. 雑誌名 Bioinformatics	6. 最初と最後の頁 2666 ~ 2674
掲載論文のDOI (デジタルオブジェクト識別子) 10.1093/bioinformatics/btx287	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

〔学会発表〕 計4件 (うち招待講演 2件 / うち国際学会 1件)

1. 発表者名 Takafumi Chishima and Michiaki Hamada
2. 発表標題 Subcellular localization of lncRNAs is affected by Transposable Element (TE) derived sequences inside lncRNAs
3. 学会等名 The Sixteenth Asia Pacific Bioinformatics Conference (APBC 2018) (国際学会)
4. 発表年 2018年

1. 発表者名 浜田道昭
2. 発表標題 長鎖ノンコーディングRNA の機能の解明に向けたバイオインフォマティクス技術
3. 学会等名 EWE 三月会 11 月例会 (招待講演)
4. 発表年 2017年

1. 発表者名 浜田道昭
2. 発表標題 生命情報科学と私
3. 学会等名 第9回生命情報科学若手の会 (招待講演)
4. 発表年 2017年

1. 発表者名 Takafumi Chishima, Junichi Iwakiri and Michiaki Hamada
2. 発表標題 Identification of Transposable Elements which contribute to tissue specific expression of lncRNAs
3. 学会等名 第6回生命医薬情報学連合大会 (IIBMP 2017)
4. 発表年 2017年

〔図書〕 計1件

1. 著者名 藤 博幸、岩部 直之、川端 猛、浜田 道昭、門田 幸二、須山 幹太、光山 統泰、黒川 顕、森 宙史、東 光一、吉沢 明康、片山 俊明	4. 発行年 2018年
2. 出版社 講談社	5. 総ページ数 208
3. 書名 よくわかるバイオインフォマティクス入門	

〔産業財産権〕

〔その他〕

-



6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------