

令和 6 年 6 月 14 日現在

機関番号：32689

研究種目：基盤研究(A)（一般）

研究期間：2020～2022

課題番号：20H00624

研究課題名（和文）リピート要素のde novo発見に基づく長鎖ノンコーディングRNAの機能の解明

研究課題名（英文）Elucidation of the function of long non-coding RNAs based on de novo discovery of repeat elements

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） 34,700,000円

研究成果の概要（和文）：本研究では、長鎖ノンコーディングRNA(lncRNA)の機能をリピート要素の観点から解明することを目的とした。手法開発として、新規リピート同定ツールREPriseを開発した。またリピートの機能の解明の研究として、脳細胞特異的な発現に關与するリピート要素を発見した。さらに、リピート要素とRNA結合タンパク質の相互作用、R-loop形成への關与を明らかにした。以上より、リピート要素がlncRNAの発現制御や機能に重要な役割を果たしていることが示唆された。

研究成果の学術的意義や社会的意義

本研究の学術的意義は、lncRNAの機能解明に向けて、リピート要素という新たな切り口から研究を進めた点にある。新規リピート同定ツールREPriseの開発、脳細胞の特異性關与するリピート要素の同定、リピート要素とRNA結合タンパク質の相互作用の解明、R-loop形成へのリピート要素の關与の発見は、分子制御機構の理解を大きく前進させた。

社会的意義としては、lncRNAの機能解明が、がんや神経変性疾患などの疾患メカニズムの理解や治療法の開発に繋がる可能性がある。リピート要素の観点からlncRNAの発現制御機構を明らかにすることで、疾患關連lncRNAの同定や創薬ターゲットの発見が期待される。

研究成果の概要（英文）：In this study, we aimed to elucidate the functions of long non-coding RNAs (lncRNAs) from the perspective of repeat elements. As a methodological development, we developed a novel repeat identification tool called REPrise. In our research on uncovering the functions of repeats, we discovered repeat elements involved in brain cell-specific expression. Furthermore, we clarified the interactions between repeat elements and RNA-binding proteins, as well as their involvement in R-loop formation. These findings suggest that repeat elements play crucial roles in the regulation of lncRNA expression and function.

研究分野：バイオインフォマティクス

キーワード：リピート要素 トランスポゾン アルゴリズム lncRNA

## 様式 C - 19 , F - 19 - 1 (共通)

### 1. 研究開始当初の背景

長鎖ノンコーディング RNA (lncRNA) は、近年ヒトなどの高等真核生物において多数発見されている。タンパク質に翻訳されずに細胞内で機能を有する長い RNA である。lncRNA は転写・翻訳・エピジェネティクスなどの細胞内制御に深く関わっており、その一部は、がんや神経変性疾患などの重篤な疾患の原因となっていることが知られている。近年のシーケンサーを用いた大規模な発現解析により、ヒトでは、タンパク質コード遺伝子より多い数万種類の lncRNA が存在することがわかってきた。しかしながら、ヒトの lncRNA の大部分の機能は未知であり、これらの機能を解明することが現在の生物学における喫緊の課題となっている。

一方、リピート要素(反復配列)は、ゲノム DNA の配列の中で、複数回出現する配列の総称である。ヒトゲノムの場合、反復配列は全体の 60% 近くを占め、一昔前まではジャンク(ゴミ)であると考えられてきた。リピート要素には、いくつかの種類が存在するが、本研究で着目するのは、散在反復配列と呼ばれる種類の、ゲノム中の様々な場所に出現するリピート要素であり、トランスポゾンと呼ばれるゲノム中を移動(転移)する配列に由来すると考えられている。興味深いことに、ヒトの lncRNA の約 75% がトランスポゾン由来の配列を含むことが知られている(mRNA では 0.2%)。そのため、lncRNA の機能にリピート要素が関連しているのではないかという仮説も存在している。ゲノム配列中に一度しか現れない配列に比べて、リピート要素は情報科学的にも取り扱いにくく、これがリピート要素に関する研究が遅れている一因となっている。

このような中、申請者のグループは lncRNA のリピート要素と機能の関連性にいち早く注目し研究を行ってきた。1 例として、最近、lncRNA の重要な特徴の一つである「組織特異的発現」にリピート要素が関連しているのではないかと考え、リピート要素と発現量の関連性の網羅的な研究を行い、様々なリピート要素(トランスポゾン)が特定の組織における特異的な発現上昇/減少に関与していることを明らかにした。さらに、申請者は、これ以外にも lncRNA の機能の解明に向けたバイオインフォマティクス研究を数多く行ってきた。しかしながら、申請者も含め多くの研究者が研究を行っているにもかかわらず、lncRNA の機能の全貌は明らかになっていない。中でも、その情報解析の困難さにも起因して、リピート要素の lncRNA の機能における役割に関する研究は進んでおらず、次のような「問い」に答える必要がある。「どのような種類のリピート要素が lncRNA のどのような機能(局在・発現や疾患)に関連しているか?」また、「リピート要素がどのようなメカニズム(作動原理)で機能を実現しているか?」。さらに、この問いに答える際に根幹となるリピート要素に関しても、特に短いリピートや変異が蓄積したものを中心に未定義・未発見のものが多数あると考えられ、根本的にこの問題を解決する必要もある。

### 2. 研究の目的

lncRNA の機能をリピート要素の観点から解明する。そのために、研究の根幹であるリピート要素の再定義を行う。その後、網羅的実験データと情報技術を駆使することにより、リピート要素の作動原理の解明、リピート要素を含む機能エレメントの組み合わせである機能モジュールを導出し、これらに基づき lncRNA の局在・発現、疾患などの機能との関連性を明らかにする。

### 3. 研究の方法

上記の目的を達成するために主に以下の研究を実施した。

- (1) 新規のリピート要素をゲノム配列情報のみから同定する新規の手法である REPrise を開発
- (2) リピート要素(トランスポゾン)が脳細胞の特異性にどのように関与しているかを明らかにする研究
- (3) リピート要素と RNA 結合タンパク質の相互作用を網羅的に明らかにする研究
- (4) RNA と DNA の相互作用の一つである R-loop の形成にリピート要素がどのように関与しているかを明らかにする研究

### 4. 研究成果

#### (1) *de novo* リピート発見アルゴリズム・ツール REPrise の開発

既知のリピート配列データベースを使わずにゲノム中の反復配列(interspersed repeats)を高感度に検出できる新しいソフトウェア REPrise を開発した。REPrise の主な特徴は以下の通りである。

1. seed と呼ばれる k-mer を用いた seed-and-extension アルゴリズムを採用。seed の探索では

- 完全一致でなく、d 塩基のミスマッチを許容する inexact seeding を導入し、感度を向上。
2. extension ステップでのアラインメントスコアに affine gap ペナルティを導入。
  3. seed のマスクングを既存手法より緩和することで、感度低下を防いだ。
  4. ヒトゲノム全長配列にも適用可能に実装し、計算時間がゲノムサイズに対して線形時間で済む。

イネやシミュレーションデータを用いた検証実験の結果、REPrise は既存の代表的手法 RepeatScout よりも高感度であることが示された。特に反復配列間の塩基置換や indel が多い場合に高感度だった。さらに REPrise をヒトゲノム全長配列に適用したところ、17 個の新規反復配列ファミリー候補が同定された。そのうちの 1 つは、ある遺伝子ファミリーのプロモーター領域に由来すると推測された。また、2 つの候補は既知の LTR ファミリーの新規サブファミリーである可能性が示唆された。REPrise は反復配列の検出感度とゲノム全長への適用性に優れており、特に反復配列のデータベースが整備されていない生物種のゲノム解析に有用と期待される。

## (2) 脳細胞の細胞タイプとリピート要素の関連性を明らかにする研究

この研究では、マウス脳の単一細胞レベルのオープンクロマチン解析により、特定の細胞タイプに重要なトランスポゾン(TE)由来のシスエレメントを発見した。主な結果は以下の通りである。

1. MER130 と MamRep434 という TE ファミリーが、それぞれ Neurod2 および Lhx2 転写因子の結合モチーフを内部に持ち、グルタミン酸性神経前駆細胞においてシスエレメントとして機能していることが示唆された。
2. MER130 由来のシス配列はアムニオタ(羊膜類)の祖先で獲得され、MamRep434 由来の配列は真獣類の祖先で増幅されたと推定された。
3. これらの TE は異なる時期に獲得され、Neurod2 や Lhx2 を介して神経発生に関わる遺伝子の発現制御に寄与することで、脳の機能や形態の進化に貢献した可能性があることが示唆された。
4. MER130 は脳のシナプス膜の調節や形態形成など基本的な機能に、MamRep434 は真獣類で発達した大脳皮質の層構造の調節に関わる神経新生の制御に関与すると示唆された。
5. このように進化の異なるステージで TE を介したシス配列の獲得が起こり、脳の高次機能の獲得に段階的に寄与した可能性が示唆された。

以上の結果から、トランスポゾンが脳の特定細胞における遺伝子発現制御に重要な役割を果たし、脳の進化に貢献してきたことが明らかになった。

## (3) リピート要素と RNA 結合タンパク質 (RBP) の相互作用を網羅的に明らかにする研究

この研究では、ENCODE eCLIP データを用いて、ヒトのリピート配列に RNA 結合タンパク質(RBP)がどのように結合しているかを網羅的に解析した。主な結果は以下の通りである。

1. LINE1 の sense 鎖と antisense 鎖の両方で、特定の領域に RBP が集中して結合していることがわかった。これらの領域は安定した二次構造を形成することが予測された。
2. LINE1 のサブファミリー間で RBP 結合パターンが異なることがわかった。特に L1PA4-8 の sense 鎖の 3'UTR 領域には、SAFB、SAFB2、HNRNPL などの RBP が結合する特徴的なモチーフがあり、これらの RBP はヘテロクロマチン形成に関与することが示唆された。
3. HERV のサブファミリーでも、RBP が集中的に結合する領域が見つかった。例えば HERV39 の sense 鎖には IGF2BP1、SRSF7、TARDBP、EIF4G2 などが結合する UG リピート配列があり、選択的スプライシングに関与すると推測された。
4. これらの RBP 結合領域を含む LINE1 や HERV 由来の配列は、神経発生やシナプス機能に関連する遺伝子のイントロンに挿入されている傾向があった。
5. 以上の結果から、LINE1 や HERV 由来の配列の中には機能的な RNA 配列が含まれており、RBP の結合を介してスプライシングやクロマチン制御などに関与している可能性が示唆された。
6. トランスポゾン由来配列は宿主ゲノムにとって有用な制御配列のアーキタイプとなっており、進化の過程で宿主ゲノムに取り込まれ共進化してきたと考えられる。

## (4) RNA と DNA の相互作用の一つである R-loop の形成にリピート要素がどのように関与しているかを明らかにする研究

この研究では、DRIP-seq データの再解析により、ヒト、ショウジョウバエ、シロイヌナズナに

において、リピート配列が R-loop の形成にどのように関与しているかを調べた。主な結果は以下の通りである。

1. ヒト、ショウジョウバエ、シロイヌナズナにおいて、それぞれ satellite, LINE, DNA トランスポゾンが R-loop に濃縮されていた。また、低複雑度領域や simple repeat も R-loop 形成に関与していた。
2. R-loop 形成に関与するリピート配列は、生物種によって異なっていた。例えば、ショウジョウバエの胚では LINE や LTR が R-loop を形成しやすいのに対し、S2 細胞では低複雑度領域や simple repeat に R-loop が集中していた。
3. リピート配列の R-loop 形成への寄与は、胚よりも S2 細胞で小さくなっており、発生段階によって変化することが示唆された。
4. R-loop 形成は、もともとトランスポゾンの制御機構として進化した可能性がある。R-loop を介したトランスポゾンのサイレンシングは、宿主ゲノムにとって有利な場合には共進化によって保存されたのかもしれないことが示唆された。
5. 以上より、リピート配列は R-loop 形成を介して、生物種特異的あるいは発生段階特異的な遺伝子発現制御に関与している可能性が示唆された。

本研究は、リピート配列と R-loop 形成の関連性を生物種や発生段階横断的に解析した初めての例であり、リピート配列の新しい機能の可能性を示唆するものである。

## 5. 主な発表論文等

〔雑誌論文〕 計5件（うち査読付論文 2件/うち国際共著 0件/うちオープンアクセス 3件）

1. 著者名 Yamada Keisuke, Hamada Michiaki	4. 巻 2
2. 論文標題 Prediction of RNA-protein interactions using a nucleotide language model	5. 発行年 2022年
3. 雑誌名 Bioinformatics Advances	6. 最初と最後の頁 online
掲載論文のDOI（デジタルオブジェクト識別子） 10.1093/bioadv/vbac023	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 Atsushi, Nonaka Daisuke, Imazu Yuta, Fukunaga Tsukasa, Hamada Michiaki	4. 巻 na
2. 論文標題 REPrise: <i>de novo</i> interspersed repeat detection using inexact seeding	5. 発行年 2024年
3. 雑誌名 BioRxiv	6. 最初と最後の頁 na
掲載論文のDOI（デジタルオブジェクト識別子） 10.1101/2024.01.21.576581	査読の有無 無
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 Onoguchi Masahiro, Zeng Chao, Matsumaru Ayako, Hamada Michiaki	4. 巻 3
2. 論文標題 Binding patterns of RNA-binding proteins to repeat-derived RNA sequences reveal putative functional RNA elements	5. 発行年 2021年
3. 雑誌名 NAR Genomics and Bioinformatics	6. 最初と最後の頁 na
掲載論文のDOI（デジタルオブジェクト識別子） 10.1093/nargab/lqab055	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 Zeng Chao, Onoguchi Masahiro, Hamada Michiaki	4. 巻 12
2. 論文標題 Association analysis of repetitive elements and R-loop formation across species	5. 発行年 2021年
3. 雑誌名 Mobile DNA	6. 最初と最後の頁 na
掲載論文のDOI（デジタルオブジェクト識別子） 10.1186/s13100-021-00231-5	査読の有無 無
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 Sekine Kotaro, Onoguchi Masahiro, Hamada Michiaki	4. 巻 6
2. 論文標題 Transposons contribute to the acquisition of cell type-specific cis-elements in the brain	5. 発行年 2023年
3. 雑誌名 Communications Biology	6. 最初と最後の頁 na
掲載論文のDOI (デジタルオブジェクト識別子) 10.1038/s42003-023-04989-7	査読の有無 無
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 -

〔学会発表〕 計0件

〔図書〕 計0件

〔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
研究分担者	福永 津嵩  (Fukunaga Tsukasa)  (80791433)	早稲田大学・高等研究所・講師(任期付)    (32689)	
研究分担者	足達 俊吾  (Adachi Shugo)  (90783803)	国立研究開発法人産業技術総合研究所・生命工学領域・主任 研究員   (82626)	
研究分担者	小野口 真広  (Onoguchi Masahiro)  (30645297)	早稲田大学・理工学術院総合研究所(理工学研究所)・次席 研究員(研究院講師)   (32689)	

#### 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
---------	---------