

令和 5 年 6 月 1 日現在

機関番号：12601

研究種目：基盤研究(C) (一般)

研究期間：2020～2022

課題番号：20K12041

研究課題名(和文) アーカイブデータの大規模解析による長鎖非コードRNA機能解析のための情報基盤構築

研究課題名(英文) Large-scale analysis of archived sequencing data in public database for functional analysis of long non-coding RNAs

研究代表者

岩切 淳一 (Iwakiri, Junichi)

東京大学・大学院新領域創成科学研究科・助教

研究者番号：40770160

交付決定額(研究期間全体)：(直接経費) 3,400,000円

研究成果の概要(和文)：長鎖非コードRNA (lncRNA) の1種であるリードスルー転写産物がヒトのどのような組織・細胞内小器官で発現しているのかを調べるため、ENCODEプロジェクトが公開している様々なRNA-seqデータの再解析を行った。その結果、リードスルー転写産物の多くは、組織特異的な発現パターンを有し、mRNAとは異なり、転写後は3'末端のpoly-A化を受けず、核内に留まっているものが非常に多いことが明らかとなった。またリードスルー転写産物の多くは、様々な細胞ストレスによって発現が誘導され、難抽出性という特徴があり、核内に存在する様々な非膜性構造体の骨格として働く可能性があ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

公共データベースにアーカイブされているRNA-seqなどの大規模な実験データは、世界中からの学術論文発表に伴って飛躍的に増大しているが、本研究で着目したlncRNAのように、それら大規模なデータの中に埋もれている生命現象がまだ多く残されている。本研究によって、lncRNAの1種である多様なリードスルー転写産物が、組織特異的な発現パターンを示すこと、様々な細胞ストレスによって発現が誘導されていること、難抽出性という特徴有すること等、徐々にその細胞内での役割の一端が明らかにされつつある。このようなアーカイブデータ再活用は、近年の深層学習等の発展に伴って、今後さらに重要になっていくものと考えられる。

研究成果の概要(英文)：To investigate the expression of read-through transcripts, a type of long non-coding RNA (lncRNA) that originated from the transcriptional elongation beyond the normal transcription termination site of protein-coding genes into the downstream intergenic regions, we reanalyzed various RNA-seq data published by the ENCODE project. The results revealed following the three characteristics of the read-through transcripts: tissue-specific expression patterns similar to the other lncRNAs, no polyadenylation at the 3' end, and retained in the nucleus after the transcription. In addition, most of the read-through transcription is induced by various cellular stresses and the transcripts exhibit the semi-extractable feature. Furthermore, these transcripts are found to be localized in membrane-less nuclear body.

研究分野：計算生物学

キーワード：長鎖非コードRNA lncRNA リードスルー転写産物

## 様式 C - 19、F - 19 - 1、Z - 19 (共通)

### 1. 研究開始当初の背景

現在、ヒトゲノム上には約 20,000 個のタンパク質コード遺伝子以外に、タンパク質に翻訳されない非コード RNA 遺伝子が多く存在しており、これら非コード RNA 遺伝子から転写される RNA 分子 (非コード RNA : ncRNA) が細胞内の様々な制御に関与していることが知られている。近年の次世代シーケンサーを用いた RNA-seq 等の網羅的な実験手法の普及や、MiTranscriptome・FANTOM CAT 等の国際コンソーシアムによる大規模な実験により、ヒト生体内には約 60,000 個にも及ぶ長鎖非コード RNA (lncRNA) 遺伝子が存在することが明らかになったが、これら lncRNA の生体内機能はほとんど明らかにされていない。一方、RNA-seq などの大規模な実験データは、学术论文投稿の際に公共データベースに登録することが義務付けられているため、世界中からの様々な論文発表に伴って膨大な量の実験データがデータベース上にアーカイブされ続けている。しかしながら、多くの研究ではそれらデータに含まれる mRNA だけに着目しており、lncRNA についての解析はほとんど行われていない。

### 2. 研究の目的

ヒトの lncRNA の生体内での機能解析のための基礎的な情報基盤の構築を目的として、公共データベースに蓄積された膨大な量の RNA-seq およびその応用手法の実験データの再解析を通じて、公共データベース中に埋もれている lncRNA の発現量や細胞内局在、分子間相互作用などの lncRNA の機能情報の網羅的な抽出を行う。

### 3. 研究の方法

本研究では、公共データベースに登録されている RNA-seq データ (lncRNA の発現条件) CLIP-seq データ (lncRNA とタンパク質間の相互作用) PARIS・LIGR-seq 等の近接ライゲーション法を用いた実験データ (lncRNA とその他の RNA 間の相互作用) の再解析を通じて、lncRNA の機能解析の情報基盤構築を行う。

### 4. 研究成果

#### 4 - 1. リードスルー転写産物の解析と難溶性 RNA-seq

近年、lncRNA の 1 種で、タンパク質コード遺伝子の転写が通常的位置で終結せず非常に長く伸長されることで生成されるリードスルー転写産物が発見され、それら lncRNA の細胞内での役割について注目されている。このようなリードスルー転写産物が公共データベースにアーカイブされている RNA-seq データからどの程度検出されるのか、どのような組織、実験条件で検出されるのかを調べるために、リードスルー転写産物が生じている遺伝子の検出パイプラインを開発し、ENCODE 国際コンソーシアムが公開している RNA-seq データに適用することで、網羅的な検出を行った。リードスルー転写産物の発現条件を調べるため、ENCODE が公開している RNA-seq データのうち、46 組織由来サンプルの Total RNA-seq、培養細胞内の各細胞画分 (核、核小体、核質、クロマチン、細胞質可溶性、細胞質不溶性、細胞膜) での Total RNA-seq (rRNA のみ除去) poly-A 濃縮 RNA-seq (poly-A plus: 一般的な mRNA の RNA-seq) poly-A 除去 RNA-seq (poly-A minus: rRNA と poly-A RNA を除去) のデータについて解析を行った。その結果、組織由来サンプルでは、リードスルー転写産物が生じている遺伝子の検出数が最小の組織では約 100 個、最大の組織では約 3,200 個検出され、組織間の差が非常に顕著であった。このことは、ヒトで既に発見されている lncRNA の多くが組織特異的な発現パターンを示すことと共通しており、リードスルー転写産物も多くの lncRNA と同様の特徴を有していることが示唆された。

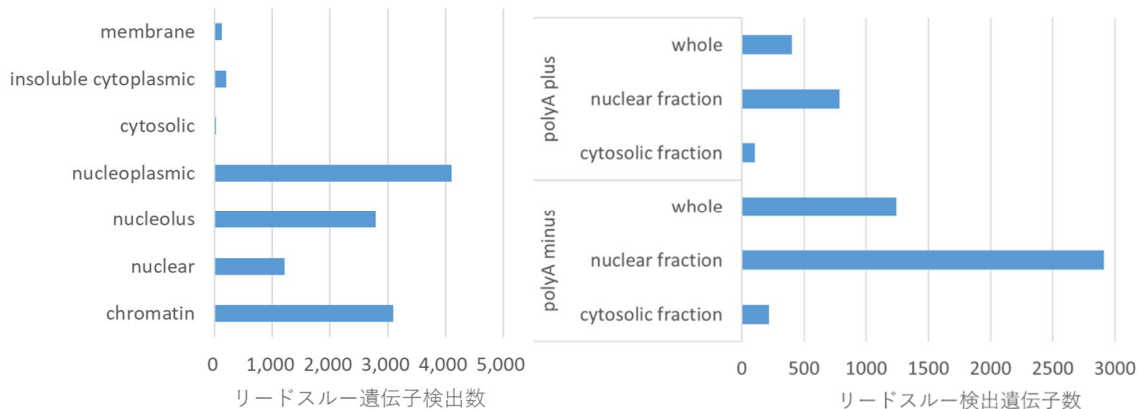


図1 ENCODE プロジェクトから公開されている細胞画分ごとの RNA-seq データにおけるリードスルー転写産物が検出された遺伝子数

また、培養細胞内の細胞内画分ごとの Total RNA-seq データの解析では、リードスルー転写産物は核画分には平均 2,800 個検出される一方、細胞質画分では平均 100 個程度しか検出されなため、リードスルー転写産物の大部分は転写された後に、核内に留まっていることが示唆された。さらに、核画分由来 RNA について、poly-A plus と poly-A minus の RNA-seq データを比較したところ、リードスルー転写産物は、poly-A minus のサンプルでより多く検出された。このことから多くのリードスルー転写産物の 3' 末端は poly-A 化されていないことが示唆された(図 1)

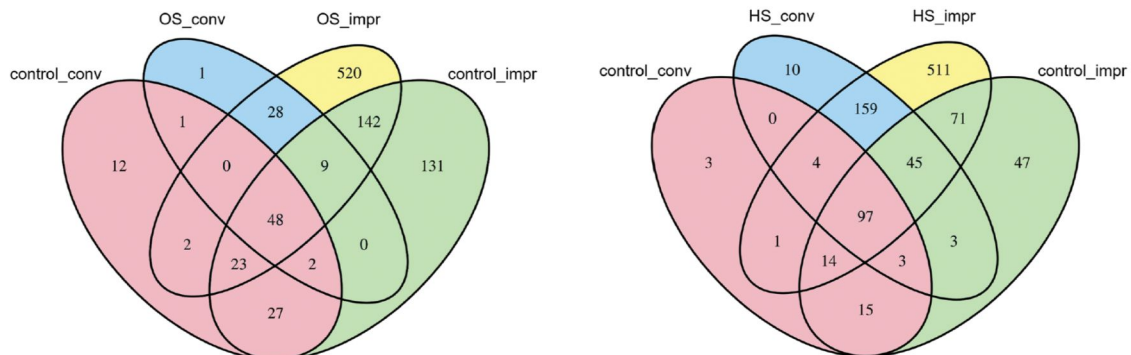


図2 難抽出性 RNA-seq と通常 RNA-seq によりリードスルー転写産物が検出された遺伝子数 (OS : 浸透圧ストレス、HS : 熱ストレス、conv : 一般的な RNA 抽出法、impr : 改善 RNA 抽出法)

上記のような公共データベースにアーカイブされた RNA-seq データからリードスルー転写産物を検出する研究と並行して、通常の RNA-seq で用いられる一般的な RNA 抽出法では抽出されづらい RNA を効率的に抽出・シーケンシングを行う新しい実験手法である難抽出性 RNA-seq のデータ解析を共同研究として実施した。この研究では、一般的な RNA 抽出と改善抽出法、細胞ストレスの有無(熱・浸透圧)の組み合わせでシーケンシングを実施後、リードスルー転写産物の検出・比較を実施した。その結果、多くのリードスルー転写産物は難抽出性の性質を持っており、さらにそのような RNA は細胞ストレス条件下で非常に多く発生することが明らかとなった(図2)。このことから難抽出性 RNA-seq は、リードスルー転写産物を効率的に検出できる強力な実験手法であることが示唆された。また、ストレス条件下で検出されたリードスルー転写産物の一部は、核内で RNA を骨格とした非膜性の構造体を形成することも示され、このようなストレス誘導性転写産物の細胞内での機能研究のきっかけとなることが期待される。

#### 4 - 2 . RNA - タンパク質相互作用の解析

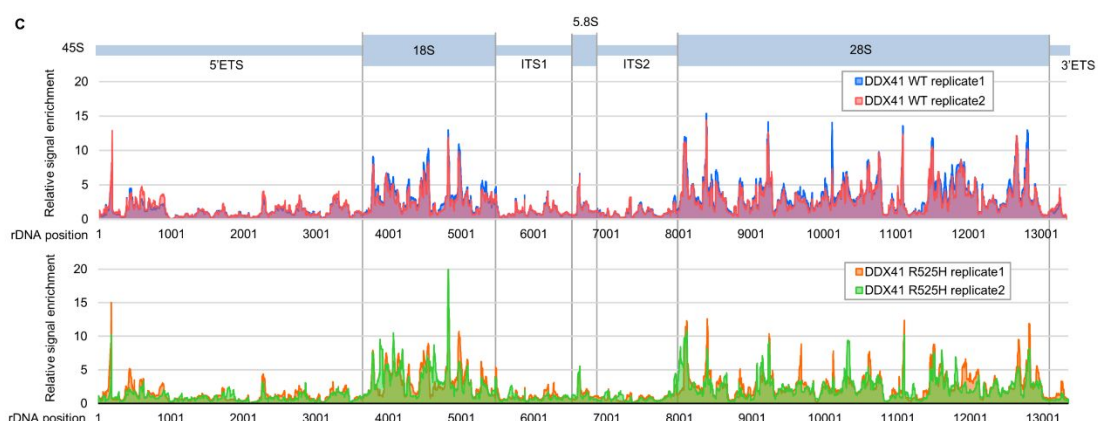


図3 CLIP-seq で見られる rRNA と DDX41 タンパク質との結合

近年、RNA - タンパク質相互作用を大規模にスクリーニングする実験手法として CLIP-seq 法が広く普及しており、ncRNA やそれに結合するタンパク質の機能研究において強力な実験手法となっている。RNA 結合タンパク質の1つである DDX41 は、血液悪性腫瘍で変異が見られるタンパク質であり、このタンパク質が結合するターゲット RNA を探索する共同研究に参画し、CLIP-seq データの解析を実施した。その結果、DDX41 の主要な結合ターゲット RNA は rRNA であり、特に成熟型の 18S と 28S rRNA との結合が確認された他(図3)、pre-mRNA のイントロンスプライス部位(特に5' SS)に選択的に結合していることが確認された。

## 5. 主な発表論文等

〔雑誌論文〕 計9件（うち査読付論文 9件/うち国際共著 1件/うちオープンアクセス 5件）

1. 著者名 Sakuraba Shun, Xie Qilin, Kasahara Kota, Iwakiri Junichi, Kono Hidetoshi	4. 巻 18
2. 論文標題 Extended ensemble simulations of a SARS-CoV-2 nsp1 5' -UTR complex	5. 発行年 2022年
3. 雑誌名 PLOS Computational Biology	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) 10.1371/journal.pcbi.1009804	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 Ninomiya Kensuke, Iwakiri Junichi, Aly Mahmoud Khamis, Sakaguchi Yuriko, Adachi Shungo, Natsume Tohru, Terai Goro, Asai Kiyoshi, Suzuki Tsutomu, Hirose Tetsuro	4. 巻 40
2. 論文標題 m6A modification of HSATIII lncRNAs regulates temperature dependent splicing	5. 発行年 2021年
3. 雑誌名 The EMBO Journal	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) 10.15252/embj.2021107976	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 Ito-Ishida Aya, Baker Steven A., Sillitoe Roy V., Sun Yaling, Zhou Jian, Ono Yukiteru, Iwakiri Junichi, Yuzaki Michisuke, Zoghbi Huda Y.	4. 巻 40
2. 論文標題 MeCP2 Levels Regulate the 3D Structure of Heterochromatic Foci in Mouse Neurons	5. 発行年 2020年
3. 雑誌名 The Journal of Neuroscience	6. 最初と最後の頁 8746 ~ 8766
掲載論文のDOI (デジタルオブジェクト識別子) 10.1523/jneurosci.1281-19.2020	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 該当する
1. 著者名 Sakuraba Shun, Iwakiri Junichi, Hamada Michiaki, Kameda Tomoshi, Tsuji Genichiro, Kimura Yasuaki, Abe Hiroshi, Asai Kiyoshi	4. 巻 16
2. 論文標題 Free-Energy Calculation of Ribonucleic Inosines and Its Application to Nearest-Neighbor Parameters	5. 発行年 2020年
3. 雑誌名 Journal of Chemical Theory and Computation	6. 最初と最後の頁 5923 ~ 5935
掲載論文のDOI (デジタルオブジェクト識別子) 10.1021/acs.jctc.0c00270	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 Takizawa Hiroki、Iwakiri Junichi、Asai Kiyoshi	4. 巻 21
2. 論文標題 RintC: fast and accuracy-aware decomposition of distributions of RNA secondary structures with extended logsumexp	5. 発行年 2020年
3. 雑誌名 BMC Bioinformatics	6. 最初と最後の頁 210
掲載論文のDOI (デジタルオブジェクト識別子) 10.1186/s12859-020-3535-5	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 Takizawa Hiroki、Iwakiri Junichi、Terai Goro、Asai Kiyoshi	4. 巻 36
2. 論文標題 Finding the direct optimal RNA barrier energy and improving pathways with an arbitrary energy model	5. 発行年 2020年
3. 雑誌名 Bioinformatics	6. 最初と最後の頁 i227 ~ i235
掲載論文のDOI (デジタルオブジェクト識別子) 10.1093/bioinformatics/btaa469	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 Iwakiri Junichi、Tanaka Kumiko、Chujo Takeshi、Takakuwa Hiro、Yamazaki Tomohiro、Terai Goro、Asai Kiyoshi、Hirose Tetsuro	4. 巻 29
2. 論文標題 Remarkable improvement in detection of readthrough downstream-of-gene transcripts by semi-extractable RNA-sequencing	5. 発行年 2022年
3. 雑誌名 RNA	6. 最初と最後の頁 170 ~ 177
掲載論文のDOI (デジタルオブジェクト識別子) 10.1261/rna.079469.122	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 Shinriki Satoru、et. al	4. 巻 36
2. 論文標題 DDX41 coordinates RNA splicing and transcriptional elongation to prevent DNA replication stress in hematopoietic cells	5. 発行年 2022年
3. 雑誌名 Leukemia	6. 最初と最後の頁 2605 ~ 2620
掲載論文のDOI (デジタルオブジェクト識別子) 10.1038/s41375-022-01708-9	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 岩切 淳一	4. 巻 3
2. 論文標題 大規模配列データにより加速するノンコーディングRNA研究	5. 発行年 2022年
3. 雑誌名 JSBi Bioinformatics Review	6. 最初と最後の頁 1~10
掲載論文のDOI (デジタルオブジェクト識別子) 10.11234/jsbibr.2022.primr1	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 -

〔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
---------------------------	-----------------------	----

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