様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　５ 年　 ５ 月　 ８ 日現在

機関番号：１４４０１
研究種目：基盤研究(C)（一般）
研究期間：2018～2022
課題番号：１８Ｋ１１５２６

研究課題名（和文）種間で保存されたRNAグアニン４重鎖構造のゲノムワイド解析法の確立

研究課題名（英文）Establishing genome-wide analysis of RNA guanine quadruplexes conserved among species

研究代表者
加藤　有己（Kato, Yuki）

大阪大学・大学院医学系研究科・准教授

研究者番号：１０５１１２８０

交付決定額（研究期間全体）：（直接経費）　　3,300,000円

研究成果の概要（和文）：近年の網羅的解析により、細胞内に発現するRNAの大多数がタンパク質をコードしない非コードRNA（ncRNA）であることが明らかとなった。Ｇ４重鎖は核酸中のグアニンに富んだ配列に形成される特殊な高次構造であり、ncRNA中に豊富に存在している可能性が近年指摘されている。しかし、生物学的意義や異なる種間での保存性、そのゲノム全体での規模など未解明な部分が多い。本研究では、比較構造情報解析を行うことで、従来見過ごされてきたncRNA中の種間で保存されたＧ４重鎖領域をゲノムワイドに同定することを最終目標とし、そのためＧ４重鎖を配列から分類するモデルを開発した。

研究成果の学術的意義や社会的意義
本研究を発展させることによって保存されたＧ４重鎖領域を網羅的に同定できれば、これを契機に作動装置を特定し、ひいてはncRNA中に埋め込まれたＧ４重鎖の生物学的意義を明らかにすることへとつながることが期待できる。Ｇ４重鎖はALSなどの神経変性疾患にも関与していることが明らかになりつつあるため、ncRNAの機能を規定する要素として保存されたＧ４重鎖領域を同定する本研究は、疾患病態を解明するための一里塚としても社会的意義に富んだ内容だと考える。

研究成果の概要（英文）：Recent exhaustive analyses have revealed that a large majority of RNAs expressed in cells are non-coding RNAs (ncRNAs), which do not code proteins. G-quadruplexes (G4s) are special substructures that are formed in G-rich sequences, and recent reports say that ncRNAs may contain abundant G4s. However, it is almost unclear what a biological meaning of G4s is, how they are conserved among different species, and to what extent they are occupied in the whole genome. In this study, we developed a classification model for G4s, which will lead to identifying G4 regions in ncRNAs genome-wide among different species.

研究分野：バイオインフォマティクス

キーワード：RNA　G4重鎖　トランスクリプトーム

１．研究開始当初の背景

近年、ゲノムの 80%以上の領域から RNA が転写され、そのほとんどがタンパク質をコードしない非コード RNA (ncRNA) であることが明らかとなった (Dunham et al. *Nature* 2012)。これには、microRNA のような 20～30 塩基長程度の small RNA から、1 キロ以上もある long ncRNA (lncRNA) など多種多様な RNA が含まれるが、中でも lncRNA の大多数は、その１次配列だけ見れば種間の保存性も低いことから、機能が解明されたものはごく一部に留まる。この種間の保存性が低いことについては、実際には lncRNA 内に共通の高次構造が埋め込まれていて、その構造保存性を介して共通の機能を果たしている可能性が考えられている。しかし、高次構造の種間保存性に着目した解析は、予測手法の困難さなどによりこれまでほとんど行われてこなかった。

一方、RNA 中のグアニン (G) が豊富な領域には、G４重鎖と呼ばれる特殊な折り畳み構造が形成されることが知られている (**図１**)。近年のトランスクリプトーム解析によって、これらＧ４重鎖が RNA 中に豊富に存在していることが予測されており、個別のケースについては転写やスプライシングなどに関与することが報告されている (Agarwala et al. *Org Biomol Chem* 2015)。しかし、ncRNA 中のＧ４重鎖の生物学的意義は概ね不明であり、これを明らかにする第一歩として、従来行われてきた個々の RNA での領域予測を越えて、種間でのゲノムワイドな比較によって、Ｇ４重鎖構造保存領域を明らかにする必要がある。実際、RGG (arginine-glycine-rich) ドメインを持った複数の RNA 結合タンパク質 (FMRP, Ncleolin など) はＧ４重鎖に結合することが知られていることから (Vasilyev et al. *PNAS* 2015)、種間で保存されたＧ４重鎖は、ncRNA に共通した要素として機能していることが予測できる。

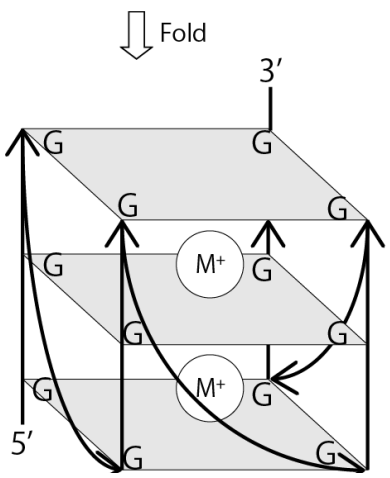

図１　RNA のＧ4 重鎖（M+は１価の陽イオンを表す）

２．研究の目的

本研究では、「種間で保存されたＧ４重鎖はどのような生物学的意義を持つか？」に答えるべく、情報科学の立場から高度な数理モデルやアルゴリズムを開発することで、種間で保存性高く埋め込まれた Ｇ４重鎖と呼ばれる高次構造をゲノムワイドに同定することを目的としている。これは単に ncRNA の機能解析への応用につながるのみならず、Ｇ４重鎖モチーフを ncRNA の機能を規定する要素として捉え、その有無によって ncRNA を新たに分類できる可能性を秘めている。

３．研究の方法

(1) 深層学習によるＧ４重鎖配列の予測

ゲノム中の Ｇ４重鎖候補領域をその配列モチーフを手掛かりに発見するため、畳み込みニューラルネットワーク (convolutional neural network, CNN) を応用する。ここで、Ｇ４重鎖 RNA 検出に特化した次世代シークエンシング技術 (Kwok et al. *Nat Methods* 2016) により得られたデータを利用して CNN を学習する。CNN への入力を配列情報から作る際、RNA 構造情報の１つである塩基対確率を ViennaRNA パッケージ (Lorenz et al. *Algorithms Mol Biol* 2011) を用いて追加することで、RNA 配列特有のモチーフに対応させることを検討する。また、既知のＧ４重鎖 RNA 配列を格納したデータベースである G4RNA (Garant et al. *Database* 2015) を利用して、モデルの識別能力を評価する。

(2) Ｇ４重鎖構造情報の粗視化による高速RNA配列比較法の開発

２つの RNA 配列が与えられたときに、Ｇ４重鎖の存在を考慮した構造類似度スコアを高速に計算するアルゴリズムを開発する。ここで、Ｇ４重鎖構造比較の計算量の削減のため、各配列において２次元の構造情報を１次元の２進列に圧縮する。これにより、例えば２本の文字列の内積を計算することで、高速に類似度スコアを導出することが可能となる。構造情報を表すＧ４重鎖エネルギー行列の計算は ViennaRNA パッケージを用いることで対応可能である。

(3) ２種の生物間におけるＧ４重鎖を考慮したゲノム網羅的配列比較

長さ固定のスライディング・ウインドウを用いて、ヒトとマウスのゲノム配列に対して部分配列を連続的に取り出し、(1)で開発した CNN モデルに入力することで、Ｇ４重鎖候補領域の集合

を得る。次に、得られたG４重鎖候補配列群を(2)で開発した構造比較アルゴリズムへ入力し、その出力結果である類似度スコアの大小をもとに、各領域が G４重鎖を形成するか否かを判定する。その際に用いる閾値を、既知のG４重鎖RNA配列から得られた情報を利用して推定することで、開発手法の実用性を担保する。

(4) ゲノムワイド多重比較への拡張とncRNA分類への適用

開発したペアワイズゲノム配列比較技術を、３種以上のモデル生物に適用し、G４重鎖 RNA分類のための絞り込みを検討する。ここで、累進法によるマルチプルアラインメントと同様の手法で、ペアワイズ比較時に得られた類似度スコアを組み合わせて系統樹を計算することで、多重比較解析へ拡張する。多重比較により得られたG４重鎖遺伝子情報を、アノテーション済み遺伝子とゲノムワイドに比較し、種間で保存されたG４重鎖モチーフや遺伝子数などの規模、位置などを解析することで、ncRNA を新たに分類することを試みる。

４．研究成果

本研究では RNA の G４重鎖を主な研究対象としており、DNA のそれと区別するため、以下rG4 と呼ぶことにする。先述の通り、rG4 配列の特徴を自動的に抽出するための方策として、CNNをベースとした教師あり機械学習モデル D-Quartet (deep learning-based quadruplex RNA detector)を開発した。具体的に、主として 1 層の畳み込み層と 1 層の全結合層からなり、2 クラス (rG4か否か) を予測する比較的シンプルなモデルを設計した。

使用するデータセットとして、rG4-seq と呼ばれる rG4 に特化した次世代シークエンシング技術 (Kwok et al. *Nat Methods* 2016) により得られる、ヒトの HeLa 細胞由来の短い RNA 配列断片を入手した。これらは rG4 であることが実験的に確認された配列であり、機械学習における正例として扱うことができるものである。ただし、CNN といった深層学習に用いるためには配列の本数が著しく不足しているため、データ拡張を行って合計 70,000 本の正例の配列を生成した。一方、rG4 でない負例の配列として、ヒトの UTR 配列から正例に重複しない領域をランダムに選択し、70,000 本の配列を生成した。

教師あり学習モデル D-Quartet を評価するために、上で生成した合計 140,000 本の配列を、クラスの比率が均等になるように、ランダムに 126,000 本の訓練データセットと 14,000 本のテストセットに分割した。CNN モデルを訓練データセットにより学習させるため、層化 10 分割交差検証を行い、モデルのハイパーパラメータをベイズ最適化により決定した。

次に、テストデータセットを用いて D-Quartet の性能評価を行った。ここで、既存の手法として、機械学習に基づく G4NN (Garant et al. *Bioinformatics* 2017)、組成塩基のスコア付け体系に基づく G4Hunter (Bedrat et al. *Nucleic Acids Res* 2016)、pqsfinder (Hon et al. *Bioinformatics* 2017) との比較を行った。**図２**に示す通り、D-Quartet の受信者操作特性曲線下面積 (area under the ROC curve, AUC) の値は他手法と比べて最も高く、rG4 か否かを分類する上で高い性能を上げることに成功した。

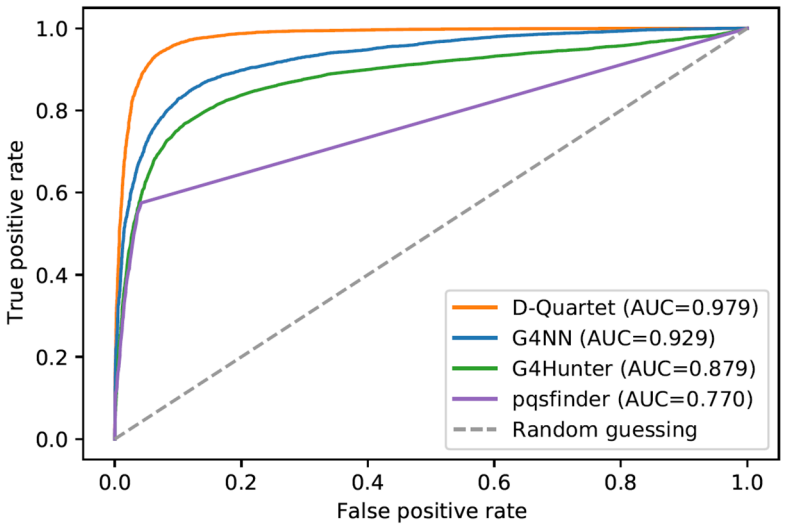

図２　開発手法のG4重鎖識別精度（AUC: area under the ROC curve）

しかしながら、先述の G4RNA データベースから取得した配列群をテストデータセットとして用いた時、分類性能の著しい低下が見られた。これは、訓練データセットとテストデータセットの塩基組成などの違いから生じるものと考えている。本研究期間中に対応策を考えていたが、新しいデータを生成して学習の規模を大きくする以外に、良い打開策を見つけられなかった。データが少ない、あるいはデータの分布が異なる問題は、データ駆動型アプローチを採用する上で不

可避の問題であり、今後の解決が待たれる状況である。
　データセットの本質的な違いによる性能低下問題をクリアした暁には、例えばヒトで学習したモデルを使ってマウスの RNA 配列の分類を行い、rG4 であることを実験的に確証したマウスのデータと比較することで、種間の保存性を検討することが可能である。今後ウェット研究者の協力を仰ぎ、引き続き研究を実施したいと考えている。

５．主な発表論文等

〔雑誌論文〕　計12件（うち査読付論文　12件／うち国際共著　1件／うちオープンアクセス　4件）

１．著者名 Sugihara Reiichi、Kato Yuki、Mori Tomoya、Kawahara Yukio	４．巻 13
２．論文標題 Alignment of single-cell trajectory trees with CAPITAL	５．発行年 2022年
３．雑誌名 Nature Communications	６．最初と最後の頁 5972
掲載論文のDOI（デジタルオブジェクト識別子） 10.1038/s41467-022-33681-3	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 Sato Kengo、Kato Yuki	４．巻 23
２．論文標題 Prediction of RNA secondary structure including pseudoknots for long sequences	５．発行年 2021年
３．雑誌名 Briefings in Bioinformatics	６．最初と最後の頁 1～9
掲載論文のDOI（デジタルオブジェクト識別子） 10.1093/bib/bbab395	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 Kim Jung In、Nakahama Taisuke、Yamasaki Ryuichiro、Costa Cruz Pedro Henrique、Vongpipatana Tuangtong、Inoue Maal、Kanou Nao、Xing Yanfang、Todo Hiroyuki、Shibuya Toshiharu、Kato Yuki、Kawahara Yukio	４．巻 17
２．論文標題 RNA editing at a limited number of sites is sufficient to prevent MDA5 activation in the mouse brain	５．発行年 2021年
３．雑誌名 PLOS Genetics	６．最初と最後の頁 e1009516
掲載論文のDOI（デジタルオブジェクト識別子） 10.1371/journal.pgen.1009516	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 Nakahama Taisuke、Kato Yuki、Shibuya Toshiharu、Inoue Maal、Kim Jung In、Vongpipatana Tuangtong、Todo Hiroyuki、Xing Yanfang、Kawahara Yukio	４．巻 54
２．論文標題 Mutations in the adenosine deaminase ADAR1 that prevent endogenous Z-RNA editing induce Aicardi-Goutieres syndrome-like encephalopathy	５．発行年 2021年
３．雑誌名 Immunity	６．最初と最後の頁 1976～1988
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.immuni.2021.08.022	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 Uyeda Akiko、Quan Lili、Kato Yuki、Muramatsu Nagaaki、Tanabe Shogo、Sakai Kazuhisa、Ichinohe Noritaka、Kawahara Yukio、Suzuki Tatsunori、Muramatsu Rieko	4．巻 69
2．論文標題 Dimethylarginine dimethylaminohydrolase 1 as a novel regulator of oligodendrocyte differentiation in the central nervous system remyelination	5．発行年 2021年
3．雑誌名 Glia	6．最初と最後の頁 2591～2604
掲載論文のDOI（デジタルオブジェクト識別子） 10.1002/glia.24060	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 Inoue Maal、Nakahama Taisuke、Yamasaki Ryuichiro、Shibuya Toshiharu、Kim Jung In、Todo Hiroyuki、Xing Yanfang、Kato Yuki、Morii Eiichi、Kawahara Yukio	4．巻 207
2．論文標題 An Aicardi-Goutieres syndrome-causative point mutation in Adar1 gene invokes multi-organ inflammation and late-onset encephalopathy in mice	5．発行年 2021年
3．雑誌名 The Journal of Immunology	6．最初と最後の頁 3016～3027
掲載論文のDOI（デジタルオブジェクト識別子） 10.4049/jimmunol.2100526	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 Costa Cruz Pedro Henrique、Kato Yuki、Nakahama Taisuke、Shibuya Toshiharu、Kawahara Yukio	4．巻 26
2．論文標題 A comparative analysis of ADAR mutant mice reveals site-specific regulation of RNA editing	5．発行年 2020年
3．雑誌名 RNA	6．最初と最後の頁 454～469
掲載論文のDOI（デジタルオブジェクト識別子） 10.1261/rna.072728.119	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 Vongpipatana Tuangtong、Nakahama Taisuke、Shibuya Toshiharu、Kato Yuki、Kawahara Yukio	4．巻 204
2．論文標題 ADAR1 Regulates Early T Cell Development via MDA5-Dependent and -Independent Pathways	5．発行年 2020年
3．雑誌名 The Journal of Immunology	6．最初と最後の頁 2156～2168
掲載論文のDOI（デジタルオブジェクト識別子） 10.4049/jimmunol.1900929	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 Shimmura Keisuke、Kato Yuki、Kawahara Yukio	4．巻 36
2．論文標題 Bivartect: accurate and memory-saving breakpoint detection by direct read comparison	5．発行年 2020年
3．雑誌名 Bioinformatics	6．最初と最後の頁 2725～2730
掲載論文のDOI（デジタルオブジェクト識別子） 10.1093/bioinformatics/btaa059	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 Ito Masumi、Muramatsu Rieko、Kato Yuki、Sharma Bikram、Uyeda Akiko、Tanabe Shogo、Fujimura Harutoshi、Kidoya Hiroyasu、Takakura Nobuyuki、Kawahara Yukio、Takao Masaki、Mochizuki Hideki、Fukamizu Akiyoshi、Yamashita Toshihide	4．巻 1
2．論文標題 Age-dependent decline in remyelination capacity is mediated by apelin-APJ signaling	5．発行年 2021年
3．雑誌名 Nature Aging	6．最初と最後の頁 284～294
掲載論文のDOI（デジタルオブジェクト識別子） 10.1038/s43587-021-00041-7	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 Nakasone Akino、Muramatsu Rieko、Kato Yuki、Kawahara Yukio、Yamashita Toshihide	4．巻 500
2．論文標題 Myotube-derived factor promotes oligodendrocyte precursor cell proliferation	5．発行年 2018年
3．雑誌名 Biochemical and Biophysical Research Communications	6．最初と最後の頁 609～613
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.bbrc.2018.04.118	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 Nakahama Taisuke、Kato Yuki、Kim Jung In、Vongpipatana Tuangtong、Suzuki Yutaka、Walkley Carl R、Kawahara Yukio	4．巻 19
2．論文標題 ADAR1-mediated RNA editing is required for thymic self-tolerance and prevention of autoimmunity	5．発行年 2018年
3．雑誌名 EMBO reports	6．最初と最後の頁 e46303
掲載論文のDOI（デジタルオブジェクト識別子） 10.15252/embr.201846303	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 該当する

〔学会発表〕　計13件（うち招待講演　1件／うち国際学会　2件）

1．発表者名
加藤 有己

2．発表標題
シングルセルデータを用いた細胞動態解析アルゴリズムの開発

3．学会等名
NGS EXPO 2022（招待講演）

4．発表年
2022年

1．発表者名
杉原 礼一，加藤 有己，森 智弥，河原 行郎

2．発表標題
1細胞RNA-seqデータを用いた細胞状態遷移経路アラインメント法の開発

3．学会等名
RNA Frontier Meeting 2022

4．発表年
2022年

1．発表者名
Yuki Kato, Kengo Sato, Jakob Hull Havgaard and Yukio Kawahara

2．発表標題
Deep learning-based prediction of potential RNA G-quadruplexes with D-Quartet

3．学会等名
29th Conference on Intelligent Systems for Molecular Biology and the 20th European Conference on Computational Biology (ISMB/ECCB2021)（国際学会）

4．発表年
2021年

1．発表者名
Kengo Sato and Yuki Kato

2．発表標題
Prediction of RNA secondary structure including pseudoknots for long sequences

3．学会等名
第68回情報処理学会バイオ情報学研究会

4．発表年
2021年

１．発表者名
Yuki Kato, Reiichi Sugihara, Tomoya Mori and Yukio Kawahara

２．発表標題
Alignment of complex single-cell trajectories with CAPITAL

３．学会等名
28th Conference on Intelligent Systems for Molecular Biology (ISMB2020)（国際学会）

４．発表年
2020年

１．発表者名
加藤 有己, 杉原 礼一, 森 智弥, 河原 行郎

２．発表標題
CAPITALによるシングルセルデータの疑似時系列比較解析

３．学会等名
第43回日本分子生物学会年会

４．発表年
2020年

１．発表者名
Taisuke Nakahama, Tuangtong Vogpipatana, Toshiharu Shibuya, Yuki Kato and Yukio Kawahara

２．発表標題
ADAR1 regulates early T cell development via MDA5-dependent and -independent pathways

３．学会等名
第21回日本RNA学会年会

４．発表年
2019年

１．発表者名
Pedro Henrique Costa Cruz, Yuki Kato, Taisuke Nakahama, Toshiharu Shibuya and Yukio Kawahara

２．発表標題
A comparative analysis among ADAR mutant mice reveals site-specific regulation of RNA editing

３．学会等名
第21回日本RNA学会年会

４．発表年
2019年

１．発表者名
加藤 有己，佐藤 健吾，Jakob Hull Havgaard，河原 行郎

２．発表標題
畳み込みニューラルネットワークによるRNAグアニン４重鎖領域予測

３．学会等名
第21回日本RNA学会年会

４．発表年
2019年

１．発表者名
加藤 有己，杉原 礼一，森 智弥，河原 行郎

２．発表標題
複雑な細胞運命比較のための時系列１細胞RNA-seqデータのアラインメント

３．学会等名
第42回日本分子生物学会年会

４．発表年
2019年

１．発表者名
杉原 礼一，加藤 有己，森 智弥，河原 行郎

２．発表標題
分岐を考慮した時系列シングルセルデータのアラインメント

３．学会等名
第61回情報処理学会バイオ情報学研究会

４．発表年
2020年

１．発表者名
加藤 有己，佐藤 健吾，Jakob Hull Havgaard，河原 行郎

２．発表標題
深層学習に基づくRNAグアニン４重鎖構造識別法の検討

３．学会等名
第20回日本RNA学会年会

４．発表年
2018年

１．発表者名 加藤 有己，新村 啓介，河原 行郎
２．発表標題 配列リード直接比較による高精度省メモリゲノム構造変異解析
３．学会等名 第41回日本分子生物学会年会
４．発表年 2018年

〔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			
デンマーク	コペンハーゲン大学			