

科学研究費助成事業 研究成果報告書

令和元年6月13日現在

機関番号：17104

研究種目：挑戦的研究(萌芽)

研究期間：2017～2018

課題番号：17K19565

研究課題名(和文) ゲノムアノテーションで見過された短鎖ペプチドコードORFの発見と疾患マーカー開発

研究課題名(英文) Systematic discovery of short peptide coding ORFs in the human genomes

研究代表者

矢田 哲士(Yada, Tetsushi)

九州工業大学・大学院情報工学研究院・教授

研究者番号：10322728

交付決定額(研究期間全体)：(直接経費) 5,000,000円

研究成果の概要(和文)：ヒトゲノムには数多くの長鎖ノンコーディングRNA(lncRNA)遺伝子が同定されているが、その多くの機能は分かっていない。一方、lncRNAだと考えられていた遺伝子が短いペプチドをコードしている例が報告されている。ここでは、分子生物学と生化学を使った実験的アプローチと情報生物学に基づく計算的アプローチを連携させて、lncRNAとして同定されていたヒトのRNAの中から、短いペプチドをコードしているRNAを体系的に発見した。さらに、それらのペプチドのヒト細胞における生理機能を明らかにした。

研究成果の学術的意義や社会的意義

ヒトlncRNAの多くの機能は分かっていないが、それらの幾つかは、特定の疾患組織で過剰に発現することから、疾患マーカーとして利用されている。しかし、それらは、疾患マーカーとしての血中での安定性の低さが指摘されている。一方、lncRNAだと考えられていた遺伝子が短いペプチドをコードしている例が報告されている。ペプチドマーカーは、RNAマーカーに比べて血中での安定性が高く、また、臨床現場への導入も容易であることから、その開発が渴望されている。ここでの研究成果は、ヒトゲノムから網羅的にペプチド疾患マーカーを発見することに寄与すると期待される。

研究成果の概要(英文)：Although a large number of long non-coding RNA (lncRNA) genes have been identified in the human genome, a large part of them have not been functionally characterized, yet. On the other hand, it has been reported that genes annotated as lncRNA have actually coded short peptides. In this research, we have adopted an integrative approach of experimental and computational methods: the former is based on molecular biology and biochemistry, and the latter is based on bioinformatics. Our approach has systematically discovered human short peptide coding RNAs which have been annotated as lncRNAs, and has successfully revealed their physiological functions.

研究分野：バイオインフォマティクス

キーワード：ショートORF(sORF) 短鎖ペプチド ノンコーディングRNA(ncRNA) ヒトゲノム 疾患マーカー

様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

1. 研究開始当初の背景

20 世紀後半からはじまったゲノムプロジェクトやトランスクリプトームプロジェクトの進展にとともに、高等動物のゲノムからは、さまざまなノンコーディング RNA が発現することが明らかになった。ノンコーディング RNA は、メッセンジャー RNA とは異なり、タンパク質のアミノ酸配列の情報をもたず(つまり、翻訳の鋳型にならず) それ自体で生理機能を発現する機能的 RNA の総称である。大量のノンコーディング RNA の発見から約 15 年が経過し、この間、数多くの発見がなされ、今や、ノンコーディング RNA が重要な生体分子であることは論を俟たない。ノンコーディング RNA は、そのサイズに応じて、大きく 2 つのグループに分けられる。ひとつは、miRNA に代表される全長が二十数塩基長程度の小分子ノンコーディング RNA である。もう一つは、全長が数百から数万塩基長になる長鎖ノンコーディング RNA (long non-coding RNA, lncRNA) である。

これまで、機能未知の RNA が見つかり、その一次配列の中に、タンパク質をコードすることができる「開いた」読み枠(open reading frame, ORF)が探索され、300 塩基以上の ORF が見つからなければ、その RNA はノンコーディング RNA に分類されてきた。これは、「タンパク質が機能を持つための必要なサイズとして 100 アミノ酸程度を要求する」という一般的な知見に基づいた分類である。しかし、この知見には例外が多く、例えばペプチドホルモンなど、100 アミノ酸以下の機能的タンパク質は数多く存在している。

このような背景のもと、分担研究者の秋光の予備的な研究により、これまで lncRNA に分類されていた RNA がリボソームと相互作用していることが明らかになった。このことは、lncRNA として分類されていた RNA が翻訳されていることを示唆している。実際、リボソームと相互作用する lncRNA には、機能的な短い ORF(short ORF, sORF)が予測され、短いペプチドをコードしている可能性が高まった。

2. 研究の目的

ここでは、分子生物学と生化学に基づいた実験的アプローチと情報生物学に基づいた計算的アプローチを有機的に連携させて、これまでは lncRNA として分類されてきたヒト RNA の中から、ペプチドをコードする RNA を同定し、その RNA にコードされているペプチドのヒト細胞における生理機能の解明を目指す。すなわち、ゲノムアノテーションの初期解析で lncRNA として注釈付けられたため、これまでの機能解析では見過がれていた短鎖ペプチドをコードする sORF をもつ RNA を同定する。さらに、同定した RNA と疾患との生理学的な関係を明らかにすることに挑む。

3. 研究の方法

(1) 実験的アプローチでは、次世代シーケンス技術を活用し、mRNA の翻訳活性を測定する ribosome profiling 法と polysome profiling 法のデータから、リボソームによって翻訳される「これまでノンコーディング RNA として分類されていた RNA」が存在するかを検討した。さらに、同定した RNA を TUF (Transcripts with unknown function)と呼び、それらの機能を分子生物学的に検証した。

(2) 計算的アプローチでは、短鎖ペプチドをコードする sORF をもつ RNA を網羅的に同定するために、公共のデータベースに登録されている sORF の塩基配列から統計的な特徴を抽出し、その特徴に基づいて RNA の塩基配列から短鎖ペプチドをコードする sORF を発見する計算プログラムを開発した。そして、このプログラムを使い、TUF やこれまで lncRNA に分類されていたヒト RNA の短鎖ペプチドコーディング性を評価した。

4. 研究成果

(1) 実験的アプローチ

NCBI のデータベースに長鎖ノンコーディング RNA としてアノテーションされていた 7,841 種類の RNA について、ribosome profiling 法による翻訳活性と NMD 耐性を指標に翻訳される可能性の高い RNA をスクリーニングした。その結果、翻訳されている可能性の高い RNA として 61 種類の RNA を同定した(Figure1)。以下の実験では、これらの RNA を TUF と呼ぶ。

細胞内には、不要な RNA を分解排除する仕組みとして、nonsense-mediated mRNA decay (NMD)が存在する。そこで、新たに同定した 61 個の TUF が NMD で分解されるか否かを指標に、これらが細胞にとって必要な RNA としてのポテンシャルがあるか否かを検討した。NMD 経路の必須因子である UPF1 をノックダウンした細胞を使った検討から、61 個中 45 個が NMD で分解されない、すなわち、細胞にとって必要な RNA としてのポテンシャルを持つことがわかった(Figure2)。

NMD 抵抗性の TUF のなかで、HeLa 細胞で発現量が高かった LINC00493 と LINC00998 について、計算的に予想された ORF の翻訳活性をルシフェラーゼレポーターアッセイで調べた。その結果、ORF が翻訳活性を持つことを確認した (data not shown)。さらに、これら 2 つの TUF 中の ORF に FLAG タグを付けたタンパク質発現プラスミドを構築し、実際に細胞内発現をするかを検証した。その結果、LINC00493 からのペプチド発現をウエスタンブロットで確認できた(data not shown)。そこで、以下の実験では LINC00493 にフォーカスして研究を進めた。

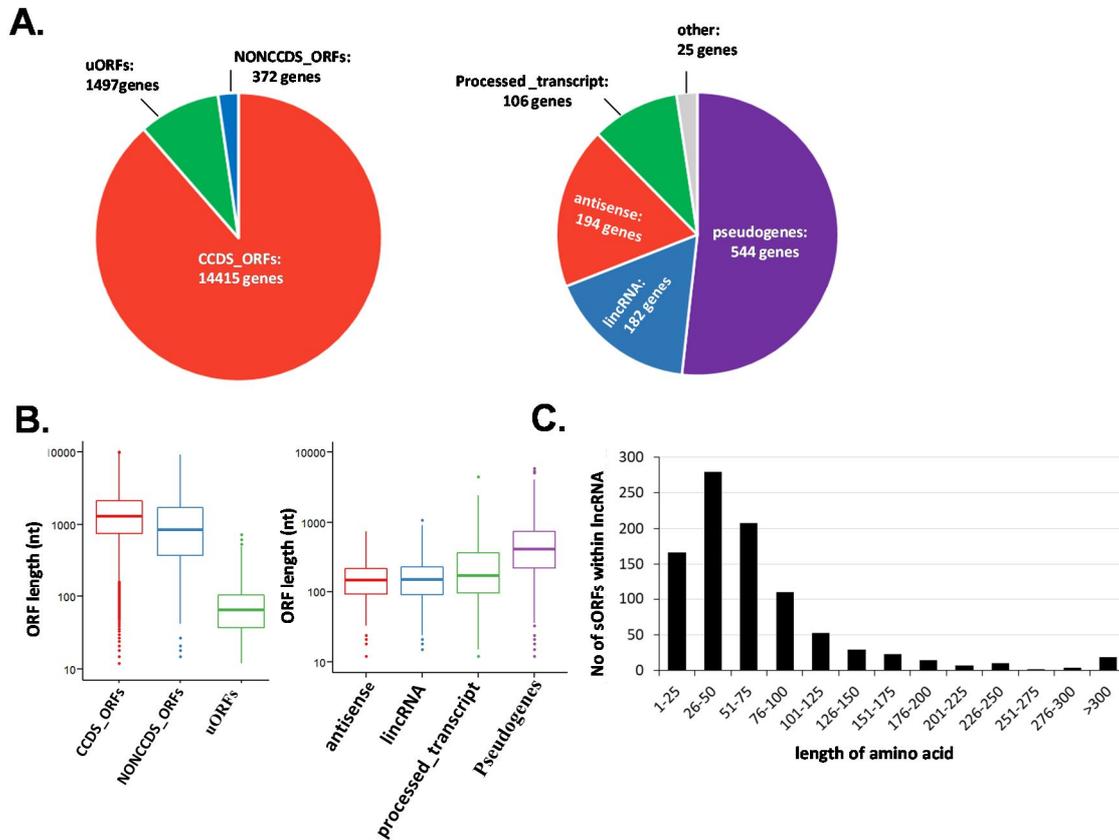


Figure 1: ノンコーディングRNAとして登録されているRNAにおけるORFの予測
 A. RiboTaperで予測した、ORFを有するmRNA (CCDS_ORFs)、ノンコーディングRNAとして登録されているRNAで予測されたORF (NONCCDS_ORFs)、upstream ORFs (uORFs) のパイチャート。本研究の解析に用いたデータにおける、RiboTaperで予測したlincRNAs, antisense RNAs, processed transcripts, pseudogenes with the putative ORFs の数の割合。B. canonical ORFs (CCDS_ORFs)、non-canonical ORFs (NONCCDS_ORFs)、upstream ORFs (uORFs) within protein-coding RNAs、について、ORF長の比較。antisense RNAs, lincRNAs, processed transcripts、pseudogenes with the putative ORFs についても、同様にORF長の比較をおこなった。C. 長鎖ノンコーディングRNAにおいて予想されるORFの長さの分布。

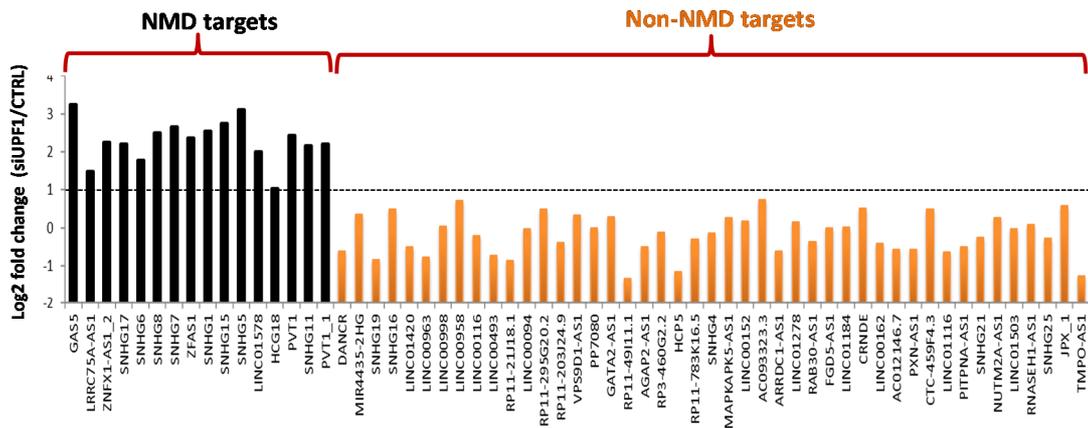
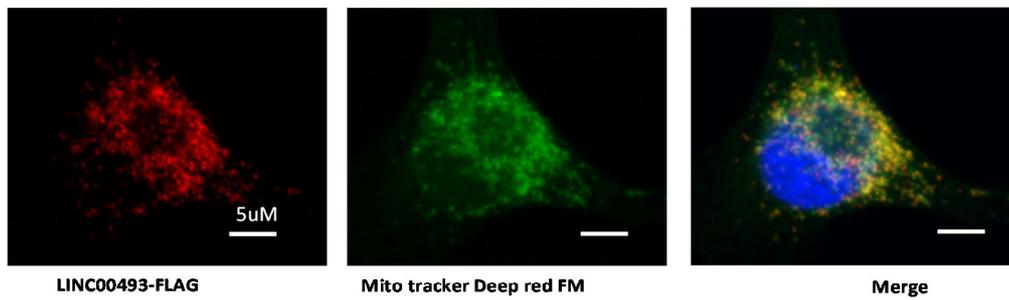


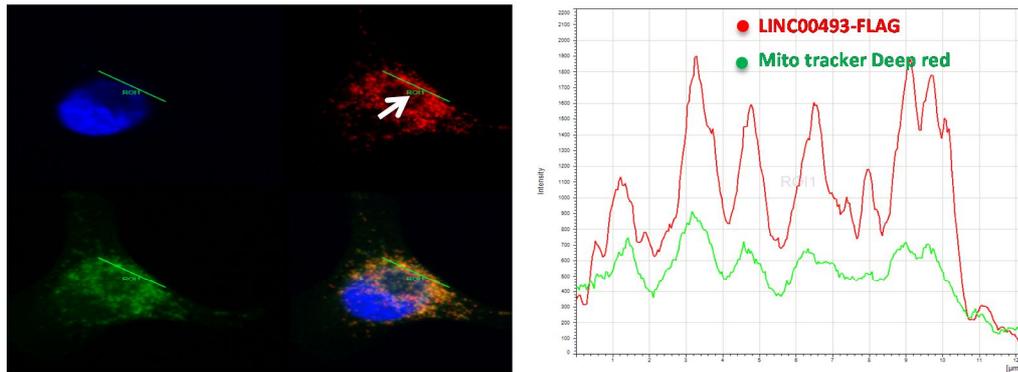
Figure 2: TUFがNMD標的となっているかの検証
 NMD機構の必須因子であるUPF1をノックダウンすることによって、各TUFの発現量が増加するか否かを定量的RT-PCR法で調べた。UPF1ノックダウンで発現量が2倍以上増加したTUFをNMD標的と判断した。

機能未知分子の機能を予測するためには、その相互作用分子を同定することが有効である。そこで、LINC00493 ペプチドの生理機能を予測するため、LINC00493 ペプチドを免疫沈降法で回収し、質量分析法で相互作用するタンパク質を特定した。その結果、LINC00493 ペプチドが多くミトコンドリアタンパク質と相互作用するとの結果を得た。

A.



B.



C.

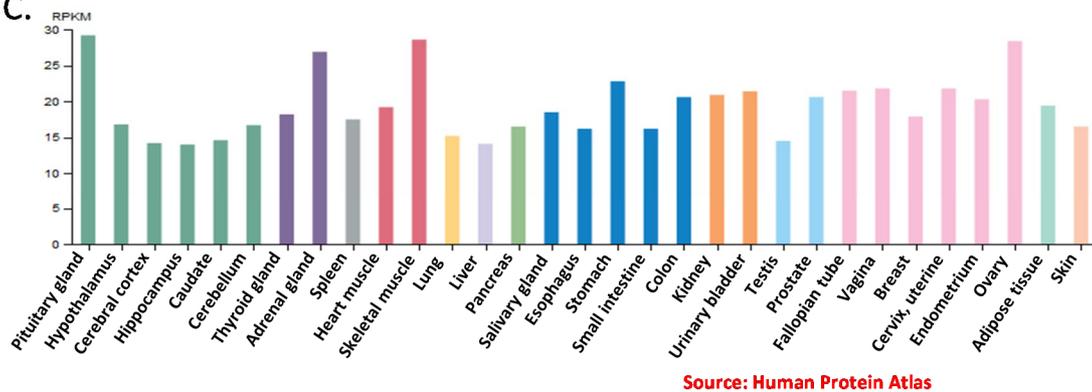


Figure 3: LINC00493 の細胞内局在

A.HeLa細胞にFLAG-LINC00493を発現し、固定後、FLAGに対する特異的抗体で免疫染色を行った(赤色)。ミトコンドリアはMitotrackerで染色した(緑色)。核をDAPIで染色した(青色)。B.Human Protein Atlasのデータから、LINC00493の各組織における発現を調べた。

次に、LINC00493 ペプチドがミトコンドリアに局在するか否かを検証した。FLAG タグペプチドタグを付加した LINC00493 ペプチドを HeLa 細胞に発現させ、FLAG タグに対する抗体で免疫染色実験を行った。ミトコンドリアは mitotracker で同時染色した。その結果、FLAG-LINC00493 が確かにミトコンドリアに局在化することを確認できた(Figure3)。さらに、ミトコンドリアにいて機能するならば、全身性に発現していることが予想された。Human protein atlas のデータを解析した結果、予想通り、LINC00493 は全身性に発現していた。

さらに、GO ターム解析を行った結果、LINC00493 ペプチドと相互作用するタンパク質がアポトーシス関連分子でもあることが判明した。このことから、LINC00493 ペプチドがアポトーシスに関連した分子である可能性が考えられる。

以上に述べたように、本研究では多数の TUF (ペプチドをコードすることが予想される RNA 分子) を発見した。そのうちのひとつである LINC00493 が確かにペプチドをコードしていることを実験的に示した。さらに、LINC00493 ペプチドがミトコンドリアに局在化することを示した。また、GO ターム解析から、LINC00493 がアポトーシスと関連する可能性を提唱するに至った。以上の結果は、これまで見過ごされてきたノンコーディング RNA の翻訳ポテンシャルの存在を改めに示すものである。そのため、この研究は、ヒトゲノムの機能解明に大きく貢献する内容である。さらに、ペプチドは中分子創薬のシーズともなるため、新たなペプチドの発見は創薬研究に有用な情報を提供する。

(2) 計算的アプローチ

ヒト sORF の塩基配列から統計的な特徴を抽出するために、まず、ribosome profiling 法によって同定された sORF の情報をまとめた SORFS.ORG (<http://sorfs.org/>) からヒト sORF 1,086 配列を取得した。そして、それらのアミノ酸組成とコドンの出現頻度を調べ、それらの統計量をヒトゲノムに注釈付けされている全ての CDS の統計量と比較した(Figure 4 (a) (b))。すると、両者の間には、アミノ酸組成より、コドンの出現頻度に大きな違いが観察された。具体的には、CDS で高い頻度で現われるコドンは sORF ではより高い頻度で、CDS で低い頻度で現われるコドンは sORF ではより低い頻度で現われる傾向が観察された。このことは、RNA の塩基配列から短鎖ペプチドをコードする sORF を発見する計算プログラムを開発する場合、全ての CDS の統計量ではなく、sORF の統計量を用いるべきであることを示している。

ここでは、上記の統計的な特徴を際立たせるために、ヒト sORF の配列データを 3 つのグループに分類し、各々のグループから統計量を得た。ひとつは高い頻度で現われるコドンの特徴を際立たせたグループ(Figure 4 (c)左)(427 配列)、もうひとつは低い頻度で現われるコドンの特徴を際立たせたグループ(Figure 4 (c)中)(460 配列)、そして残りのグループ(Figure 4 (c)右)(199 配列)である。Figure 4 (c)右では、sORF におけるコドンの出現頻度のばらつきが小さくなっている。このことは、このグループの sORF の塩基配列には、コドンの出現頻度に際立った特徴がないことを示している。

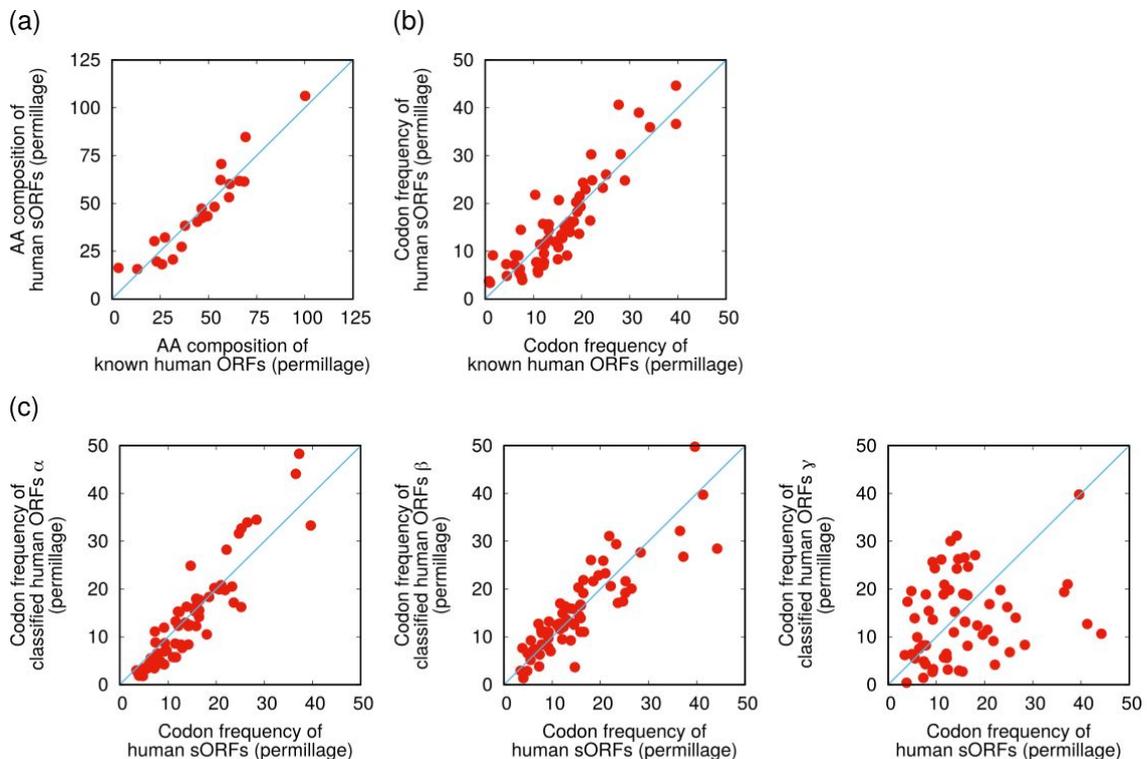


Figure 4: ヒトsORFの一次配列に観察される統計量。(a)はアミノ酸の組成(縦軸)をヒトの全CDSのもの(横軸)と比較した。(b)はコドンの出現頻度(縦軸)をヒトの全CDSのもの(横軸)と比較した。(c)は、ヒトsORFを3つのグループに分け、各々のグループのコドンの出現頻度(縦軸)を全体のもの(横軸)と比較した。

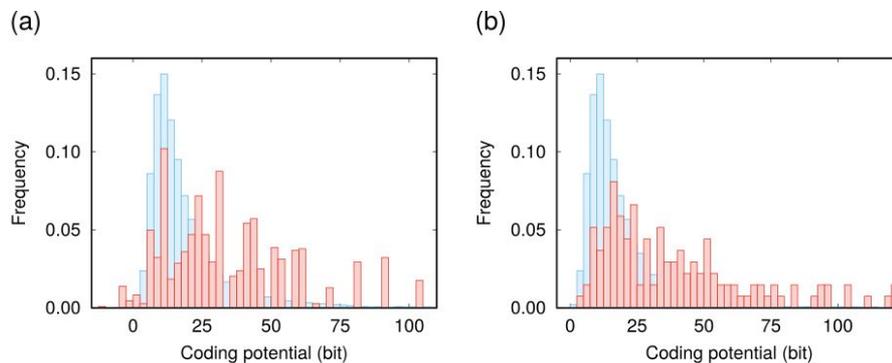


Figure 5: RNAの一次配列から短鎖ペプチドをコードするsORFを発見するHMMを使った各配列データの評価。(a)はヒトsORF(赤)とヒトncRNA(青)、(b)はTUF(赤)とヒトncRNA(青)の短鎖ペプチドコーディング性を比較した。

ここでは、Figure 4 (c)の統計量を隠れマルコフモデル(hidden Markov model, HMM)でモデル化し、RNA の塩基配列から短鎖ペプチドをコードする sORF を発見する計算プログラムを開発した。そして、このプログラムを用い、ヒト sORF、ヒト ncRNA、上記の TUF の各配列データの短鎖ペプチドコーディング性を評価した(Figure 5)。ヒト ncRNA の配列データは、GENCODE (<https://www.genencodegenes.org/>)から取得した(13,299 配列)。Figure 5 より、ヒト sORF と TUF の配列データは、ヒト ncRNA の配列データに比べ、このプログラムによって与えられるスコア(coding potential)の高い配列に富むことが分かる。しかし、ヒト sORF と TUF の一部の配列は、このスコアでは ncRNA の配列と見分けることができない。この sORF と TUF の配列の多くは、Figure 4 (c)右の統計量で評価されていた。これらの配列をどのようにして ncRNA と見分けるかは、計算的アプローチにおける今後の研究課題のひとつである。一方、ncRNA の配列の一部は、sORF と TUF の配列の多くに匹敵するスコアを示している。このことは、ncRNA に分類されている RNA の中には、短鎖ペプチドをコードするものが含まれている可能性を示している。

5 . 主な発表論文等

〔雑誌論文〕(計 2 件)

1 Genome sequence alignment, [Yada T](#), Encyclopedia of Bioinformatics and Computational Biology (Gaeta B, Nakai K, ed.), Elsevier, 268-283 (2019). 査読あり

2 Micropeptides encoded in transcripts previously identified as long noncoding RNAs: A new chapter in transcriptomics and proteomics, Yeasmin F, [Yada T](#), [Akimitsu N](#), Front Genet, 25 April (2018). doi: 10.3389/fgene.2018.00144 査読あり

〔学会発表〕(計 1 件)

1 Yeasmin F, Yada T, Akimitsu N, Functional analysis of TUFs. 2017 年 12 月 3 日から 6 日、横浜市、第 40 回分子生物学会年会 (2017 年・ConBio2017)

〔図書〕(計 0 件)

〔産業財産権〕

○出願状況 (計 0 件)

○取得状況 (計 0 件)

〔その他〕

ホームページ等

6 . 研究組織

(1)研究分担者

研究分担者氏名：秋光 信佳

ローマ字氏名：AKIMITSU, Nobuyoshi

所属研究機関名：東京大学

部局名：アイソトープ総合センター

職名：教授

研究者番号 (8 桁): 40294962

※ 科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。