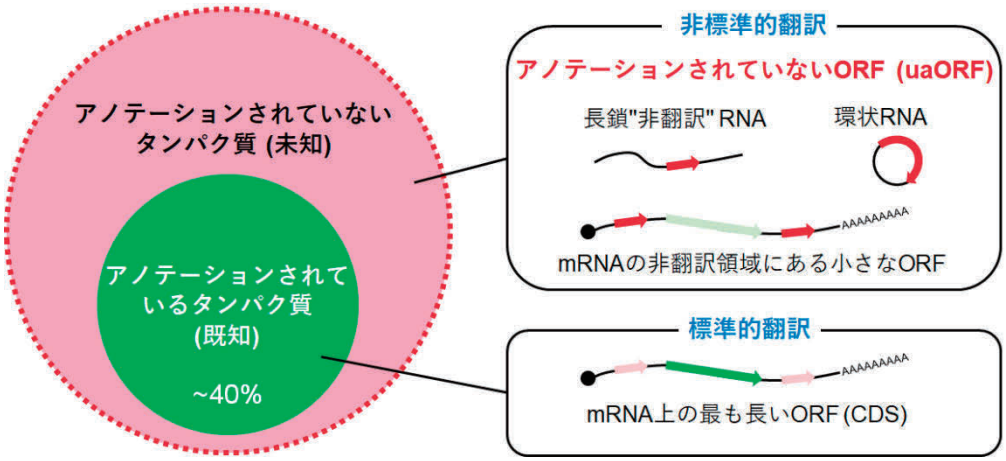
73　完全な翻訳地図と翻訳予測モデルによる細胞機能の理解	高橋　和利

【目的】従来、真核生物において、mRNA 中の一番長い open reading frame（ORF）配列である coding sequence（CDS）のみがタンパク質として翻訳されると考えられてきた。しかし、リボソームプロファイリング法により翻訳されている RNA の領域を調べたところ、CDS 以外の非翻訳と考えられてきた領域も翻訳されていることが示唆されている。質量分析により細胞で発現するタンパク質の解析を行うと、CDS 由来の既知タンパク質と一致するスペクトルは約 40%である。これまで、残りの 60%はデータとして高品質で信頼性が高いものであっても、CDS がコードする既知タンパク質と一致しないという理由で無視されていた。本研究では、これら既知のタンパク質をコードしない小さな ORF を一括りにして unannotated ORF（uaORF）と定義し、研究対象とする。uaORF がコードするタンパク質はほとんどが未同定・未解析であり、細胞や生物を説明する上で「足りないもの」の極めて有力な候補である。現在までに同定されている既知タンパク質に uaORF がコードする新規タンパク質を加えれば、現行のプロテオームデータが大幅に拡張でき、細胞機能の実行役を担うプレイヤーが出そろうだろう。そこで本研究では、ヒト多能性幹細胞における uaORF の同定とマッピングを目的とする。

【方法】二つの独立したヒト多能性幹細胞を用いて、直鎖 RNA（mRNA および長鎖非翻訳 RNA）のマッピング、環状 RNA のマッピングおよびリボソーマルプロファイリングによる翻訳領域の同定を行った。

【結果】直鎖 RNA（mRNA および長鎖非翻訳 RNA）は通常の RNA-seq を実施することでデータを得た。次に環状 RNA を濃縮する目的で、直鎖 RNA を特異的に消化する RNase R でトータル RNA を処理し、環状 RNA を RNA-seq により効率的に検出した。結果として、二つの独立したヒト多能性幹細胞株でそれぞれ 10,000 種類の環状 RNA が検出され、共通して発現していた環状 RNA は 6,788 個であった。さらに、リボソーマルプロファイリングを実施し、データベース上に存在しない新規翻訳領域を塩基レベルで見出した。uaORF マッピングデータとリボソーマルプロファイリングデータと組み合わせることにより、uaORF のうち、実際に翻訳されている領域を一塩基レベルで正確に決定した。その結果、検出した全 ORF のうち 62%が既知の CDS であった。一方で、新規の uaORF は 4,500 個以上同定され、その割合は全 ORF の 19%であった。また、報告されている翻訳開始点とは異なる位置から翻訳が開始されている CDS のバリアントは全 ORF の 19%であった。

これまで見過ごされてきたアノテーションされていないタンパク質



73. 完全な翻訳地図と翻訳予測モデルによる細胞機能の理解

高橋　和利

京都大学 iPS 細胞研究所 未来生命科学開拓部門

Key words：非標準的翻訳，多能性幹細胞，ORF

緒　言

1. 今見直され始めた翻訳の定義：見逃されてきたタンパク質を探す

従来、真核生物において、mRNA 中の一番長い open reading frame（ORF）配列である coding sequence（CDS）のみがタンパク質として翻訳されるとされてきた。そして、CDS の定義は「100 アミノ酸以上のタンパク質をコードするもの」とされたため、小さな ORF のみを持つ RNA は「ノンコーディング（非翻訳）」とされた。しかし、リボソーマルプロファイリング法により翻訳されている RNA の領域を調べたところ、CDS 以外の非翻訳と考えられてきた領域も翻訳されていることが示唆されている［1］。また、非翻訳と考えられてきた長鎖ノンコーディング RNA や環状 RNA も翻訳を受けている可能性が示唆されている［2］。CDS の上流に存在する小さな ORF はウイルスにおいて見出され、研究の歴史は古く 1980 年代から始まっている。真核生物において、CDS の上流に存在する小さな ORF は下流にある CDS の翻訳効率を調節する役割を担っているとの考え方が主流である。環状 RNA に関する研究は 1970 年代から行われているが［3］、現在に至るまでそのほとんどがウイルスに関するものである。真核生物における環状 RNA の翻訳に関する報告は少ないが、環状 RNA がコードする小さなタンパク質が筋組織で重要な役割を果たしていることが報告された［4］。近年の技術革新によって新たな翻訳領域の同定が可能になり、今後当該分野が急速に発展すると考えられている。質量分析により細胞で発現するタンパク質の解析を行うと、CDS 由来の既知タンパク質と一致するスペクトルは約 40%である（図 1）。これまで、残りの 60%はデータとして高品質で信頼性が高いものであっても、CDS がコードする既知タンパク質と一致しないという理由で無視されていた。本研究では、これら既知のタンパク質をコードしない小さな ORF を一括りにして unannotated ORF（uaORF）と定義し、研究対象とする（図 1）。uaORF がコードするタンパク質はほとんどが未同定・未解析であり、細胞や生物を説明する上で「足りないもの」の極めて有力な候補である。現在までに同定されている既知タンパク質に uaORF がコードする新規タンパク質を加えれば、現行のプロテオームデータが大幅に拡張でき、細胞機能の実行役を担うプレイヤーが出そろうだろう。

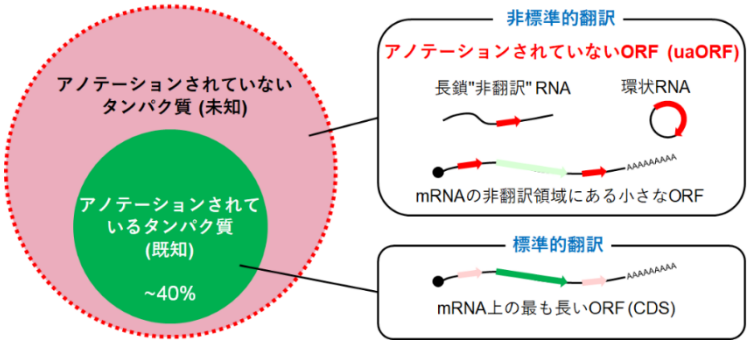


図 1.　これまで見過ごされてきた uaORF
これまで CDS と一致しないという理由で廃棄されてきたタンパク質は、アノテーションされていない ORF（uaORF）によりコードされている可能性がある。

方法および結果

1. ヒト多能性幹細胞の調製

本研究では対象をヒト多能性幹細胞とし、基礎研究目的に広く利用されている H9 ヒト胚性幹（Embryonic Stem：ES）細胞株および WTB6 ヒト人工多能性幹（induced Pluripotent Stem：iPS）細胞株を用いた［5, 6］。これらのヒト多能性幹細胞株は iMatrix-511 silk でコートした培養皿に播種し、StemFiT AK02 培地中で培養を行った。3 日に一回程度の頻度で継代を行い、定期的にマイコプラズマが陰性であることを確認した。また、正常な核型を保持し、多能性幹細胞マーカーが均一に発現していることを確認した。それぞれの細胞株からトータル RNA を回収し、後述する uaORF のマッピングである「直鎖 RNA（mRNA および長鎖ノンコーディング RNA）マッピング」および「環状 RNA マッピング」に用いた。さらに「リボソーマルプロファイリングによる翻訳領域の決定」を行うための試料調製を行った。実験結果の再現性を確認するために、3 つの異なる継代数において試料の回収を行った。

2. uaORF の検出とマッピング

1）直鎖 RNA（mRNA および長鎖ノンコーディング RNA）マッピング

上述のヒト多能性幹細胞由来トータル RNA を用いて、ライブラリーを作製し、標準的な RNA-seq 法を実施することで直鎖 mRNA と長鎖非翻訳 RNA の発現および配列情報を取得した。uaORF をマッピングする上で RNA の正確な配列を得ることは重要であるため、理化学研究所の CAGE（Cap Analysis of Gene Expression）データベースを使用し、正確な転写開始点情報を得た。

2）環状 RNA マッピング

環状 RNA の性質上、環状 RNA 本体のほとんどの部分の塩基配列は同遺伝子座にある mRNA の配列と重複するため通常の RNA-seq では区別が困難であった。この問題点については、直鎖 RNA を特異的に消化する RNase R でトータル RNA を処理し、環状 RNA を濃縮することで克服した。環状 RNA と mRNA を区別するためには、環状 RNA のみが持つ接合部位配列を頼りにマッピングを行う必要がある。接合部位を効率よく検出するために、ロングリードの RNA-seq を行った。環状 RNA の接合部位配列はゲノム配列には存在しないため、専用のアルゴリズムを用いて解析する必要があった。そこで、得られた RNA-seq データについて circExplorer［7］などの公開プログラムを独自に改変したものを利用し、ヒト ES 細胞およびヒト iPS 細胞において 10,000 種以上の環状 RNA を検出した。

3）リボソーマルプロファイリングによる翻訳領域の決定

リボソーマルプロファイリングは、リボソームが結合し実際に翻訳を受けている RNA の領域を実験的に決定する方法である［1］。翻訳阻害剤である Cycloheximide（CHX）によりリボソーム-tRNA 結合状態が固定された RNA を RNase で処理し、リボソーム-tRNA が結合していない RNA を RNase 処理により分解した。リボソーム-tRNA が結合した未消化の RNA 配列を精製しシークエンス用のライブラリーを作製した。この方法により、実際に RNA にリボソームが結合していた場所＝翻訳を受けている箇所を実験的に決定し、データベース上に存在しない新規翻訳領域を塩基レベルで見出すことを試みた。上述の uaORF マッピングデータとリボソーマルプロファイリングデータと組み合わせることにより、uaORF のうち、実際に翻訳されている領域を一塩基レベルで正確に決定した。その結果、検出した全 ORF のうち 62%が既知のタンパク質をコードする CDS であった（図 2）。一方で、新規の uaORF は検出された全 ORF の 19%であった。また、報告されている翻訳開始点とは異なる位置から翻訳が開始されている CDS のバリアントは全 ORF の 19%であった。

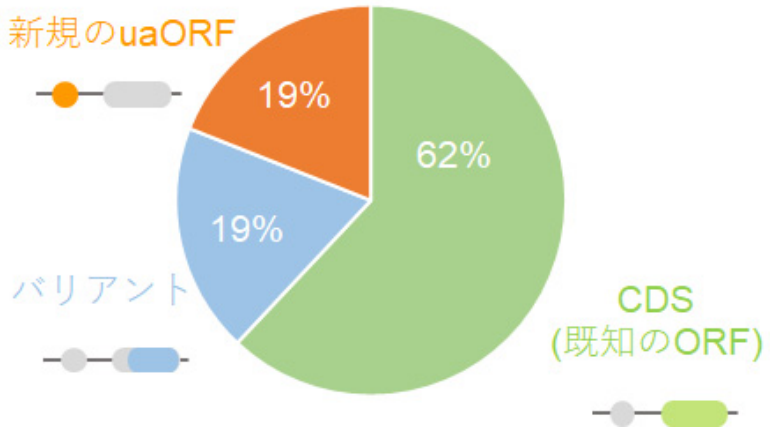


図 2. ヒト多能性幹細胞で発現する uaORF

ヒト多能性幹細胞株（ES 細胞および iPS 細胞）を用いて行った RNA-seq およびリボソームプロファイリングの結果により、新規の uaORF を同定した。加えて既知の翻訳開始点とは異なる一から翻訳が開始されているバリアントも多数同定された。

考　察

本研究の成果により、ヒト多能性幹細胞において、既知のタンパク質とアノテーションされていない新規の ORF を 4,500 種類以上同定した。これまで細胞の性質を理解する目的で、mRNA や mRNA の CDS がコードするタンパク質に焦点を当てた研究が数多く行われてきた。本研究の成果は、これら既知のタンパク質に加えて、uaORF がコードする未同定のタンパク質が細胞機能を担っている可能性を示唆するものである。さらに本研究では、二株のヒト多能性幹細胞で発現する環状 RNA を 10,000 種類以上同定した。環状 RNA は自身が機能性 RNA として働く可能性に加えて、タンパク質をコードする可能性も持っている。環状 RNA がコードするタンパク質も未同定のものがほとんどであり、これまで理解されなかった細胞機能の分子機構を解明する一助となることが期待される。

共同研究者・謝辞

本研究の共同研究者は、理化学研究所岩崎 RNA システム生化学研究室の岩崎慎太郎博士、京都大学 iPS 細胞研究所の岩崎未央講師である。

文　献

1） Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science. 2009 Apr 10;324(5924):218-23. Epub 2009 Feb 12. PMID: 19213877 DOI: 10.1126/science.1168978.

2） Fields AP, Rodriguez EH, Jovanovic M, Stern-Ginossar N, Haas BJ, Mertins P, Raychowdhury R, Hacohen N, Carr SA, Ingolia NT, Regev A, Weissman JS. A Regression-Based Analysis of Ribosome-Profiling Data Reveals a Conserved Complexity to Mammalian Translation. Molecular Cell. 2015 Dec 3;60(5):816-827. PMID: 26638175 DOI: 10.1016/j.molcel.2015.11.013.

3） Diener TO. Potato spindle tuber "virus". IV. A replicating, low molecular weight RNA. Virology. 1971 Aug;45(2):411-28. PMID: 5095900 DOI: 10.1016/0042-6822(71)90342-4.

4) Pamudurti NR, Bartok O, Jens M, Ashwal-Fluss R, Stottmeister C, Ruhe L, Hanan M, Wyler E, Perez-Hernandez D, Ramberger E, Shenzis S, Samson M, Dittmar G, Landthaler M, Chekulaeva M, Rajewsky N, Kadener S. Translation of CircRNAs. Molecular Cell. 2017 Apr 6;66(1):9-21.e7. Epub 2017 Mar 23. PMID: 28344080 DOI: 10.1016/j.molcel.2017.02.021.

5) Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, Jones JM. Embryonic stem cell lines derived from human blastocysts. Science. 1998 Nov 6;282(5391):1145-7. PMID: 9804556 DOI: 10.1126/science.282.5391.1145.

6) Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell. 2007 Nov 30;131(5):861-72. PMID: 18035408 DOI: 10.1016/j.cell.2007.11.019.

7) Hansen TB, Venø MT, Damgaard CK, Kjems J. Comparison of circular RNA prediction tools. Nucleic Acids Research. 2016 Apr 7; 44(6): e58. Epub 2015 Dec 10. PMID: 26657634 DOI: 10.1093/nar/gkv1458