<ホットトピックス (3)-5>

機能性 RNA の世界

岡田 浩美1) 林崎 良英1)

要旨: 林崎らは大規模な包括的トランスクリプトーム解析に向けて、完全長 cDNA ライブラリーの構築ならびに、それらをハイスループットシーケンシング解析に応用する新たな技術を開発してきた。彼らが組織する FANTOM データベースの解析により、ゲノム DNA の約2%だけが機能を持つという常識に反し、ゲノム DNA の 70%以上が転写され、大多数の RNA が非タンパクコード RNA(non-codingRNA;ncRNA)であることを明らかにした。これらの一連の知見により新たに「RNA 大陸」という広大な研究分野が創出された。近年の RNA 研究の進展により、生命活動における ncRNA の多様な機能がつぎつぎと報告されている。

(臨床神経 2013;53:957-961)

Key words: 非コード RNA, 長鎖非コード RNA, センス-アンチセンス RNA. トランスクリプトーム解析

はじめに

「ゲノム配列には何が書かれているのか?」染色体の DNA に4種類の塩基の配列で記された「生命の設計図」であるヒ トゲノムは約30億個の塩基配列からなる.数年前までは、 ヒトゲノムではタンパク質の設計情報が書かれている領域は 全体の2%程度とされ、生物にとって重要なゲノム上の情報 は、タンパク質のアミノ酸配列情報とプロモーターなどの発 現調節領域情報であり、それ以外の部分は意味をもたない塩 基配列の連続であると信じられていた. イメージするなら. ゲノムとは見渡すかぎり広がるジャンク DNA の砂漠に重要 な領域がオアシスのように散在するだけの世界だと考えられ ていたのである.しかしながら、林崎らの研究グループによっ て、タンパク質をコードしていない RNA (non-coding RNA; ncRNA)が大量に存在していることが明らかとなり、さらに、 近年の飛躍的な RNA 研究の進展によって、かつては意味を もたない転写産物または転写の際に生じたジャンクと考えら れてきた多くの ncRNA は、実際には、ヒトの生命活動にお いて重要かつ多様な機能をもち、ゲノムは総体として働いて いるという、旧来のゲノム観を大きく覆す新たなゲノム観が 誕生した.機能性 RNA の存在とともにその機能の破綻は, 他の分子の破綻と同様に、様々な疾患をひきおこす可能性を 容易に推察できる. 実際, がん領域など多くの疾患において, 機能性 RNA として注目されている micro RNA (miRNA) と の関連性を解析する研究は盛んにおこなわれており、アルツ ハイマー病やパーキンソン病、筋萎縮性側索硬化症(ALS) などの神経変性疾患領域においても、ncRNA の機能と発症 メカニズムの関連性について研究が進められている. 本稿で は、林崎らのグループが開発した完全長 cDNA 技術と国際研究コンソーシアム「Functional Annotation of Mammalian Genome (FANTOM)」の活動の歴史、そして、その活動から明らかとなった ncRNA の多様な機能の一部を紹介したい.

FANTOM の歴史と RNA 新大陸の発見

完全長 cDNA プロジェクト

1990年に国際ヒトゲノムコンソーシアムによって開始された「ヒトゲノム計画」は2003年をもって完了した。ゲノム上の機能領域を知るにはゲノムシーケンスだけではなく転写産物の大規模解析(トランスクリプトーム解析)が不可欠であり、1995年より林崎らは、成熟化したRNAの全長塩基配列を完全な形で解析するための完全長cDNA 技術を開発し、その技術をもちいてトランスクリプトーム(RNA)解読をおこなうことを進めてきた。

FANTOM の歴史

収集した完全長 cDNA データの機能注釈付けを目的とした共同研究を呼び掛け、2000 年に 11 ヵ国 45 機関によって結成された FANTOM コンソーシアムが結成された。これまでに 5 段階のプロジェクトのほぼ完了を迎え、続く、第 6 段階へとプロジェクトは現在も進行中である。FANTOM1 と 2 では、総計 60,770 個の完全長 cDNA クローンを対象に全長配列決定と機能注釈付けを実施した $^{1)\sim3}$)。続く FANTOM3では、総計約 103,700 個の完全長 cDNA の機能注釈付けが行われた 4

 $^{^{1)}}$ 理化学研究所社会知創成事業予防医療・診断技術開発プログラム〔〒 351-0198 埼玉県和光市広沢 2-1〕(受付日:2013 年 5 月 31 日)

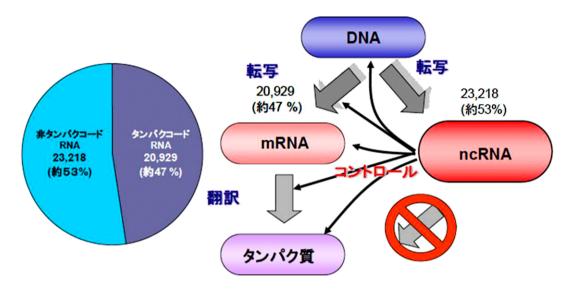


Fig. 1 タンパク質をコードしないで機能する RNA. ncRNA は DNA の転写レベルとタンパク質の翻訳レベルを制御する機能をもっている.

CAGE 法(cap analysis gene expression) と次世代シーケンサー いろいろな階層で制御され、細胞レベルで特異的に発現す る何万もの遺伝子に由来するトランスクリプトームを包括的 にその特性を解明するには、個々のトランスクリプトームの 複雑さを明らかにしてくれる新しいアプローチ方法で取り組 む必要がある. 新たな解析手法として. 転写開始点や転写終 了点を網羅的かつハイスループットに同定できる「Cap Analysis of GeneExpression (CAGE) 法」と「Gene Signature Cloning (GSC) 法」が開発され、完全長 cDNA 技術だけで はえることができない新たな知見が数多くえられた. CAGE 法は、完全長 cDNA の5'末端にアダプターをライゲーション させる完全長 cDNA ライブラリー作製技術をベースとして 開発された. このアダプターライゲーションにより. cDNA の5'末端に隣接した部分にクラス Ⅲ 制限酵素の認識サイトが 付加され、そしてクラス III 制限酵素(EcoP15I) による処 理で、転写産物の5、末端由来の短いタグがクローニングされ、 次世代高速シーケンス解析をおこなうことが可能となる. さ らに、CAGE タグをゲノム配列上でマッピングすることによ り、転写開始点の特定ができる. これらの技術と、シンガポー ルゲノム研究所が開発した Gene identification Signature (GIS) をもちいて, 転写開始点と終了点の解析情報を, マウスで 1156万7973個、ヒトで1370万6472個収集した、これらの 解析の結果、従来転写はゲノムの2%しかされていないと考 えられていた常識が覆され、ゲノムの70%以上がRNAとし て転写されていることが明らかになった。23,000 個以上の ncRNA が存在し、転写産物の半分以上が様々な機能を有す る ncRNA であることが明らかとなり⁴⁾. これらは「RNA 新 大陸」の発見として世界に大きなインパクトを与えるものと なった (Fig. 1).

多彩な ncRNA の種類と機能

ncRNA の種類

ncRNA はそのサイズにより、miRNA に代表される約20~ 30 塩基程度の小分子 RNA (small RNA) と、数百から数十 kbp におよぶ長鎖 ncRNA(lncRNA)に分けられる. Small RNAに は、rRNA の形成過程でのプロセシングや塩基の修飾のため に必要な核小体 RNA である snoRNA (small nucleolar RNA) や snRNA (small nuclear RNA) が、miRNA の研究が盛んに なる前から存在が知られており、レトロトランスポゾンの発 現抑制にかかわる piRNA (PIWI-interacting RNA) や esiRNA (endogenous small interfering RNA) などもふくまれる. 一方, lncRNA は、1990 年代には性染色体の活性化や不活性化にか かわる Xist などが報告されていたが、近年になって、ヒス トン修飾や DNA メチル化によるクロマチン修飾の機能が報 告されるなど注目を集めている。 さらに ncRNA の機能はパ ラスペックルなど核内構造体の構築⁵⁾ やエピジェネティッ クな遺伝子発現制御、そして、細胞質での標的 mRNA の分 解や翻訳制御まで多彩であることが知られている 6)7).

センス – アンチセンス RNA 転写

さらに、網羅的トランスクリプトーム解析結果では、多くの DNA 領域において両方向から転写がおこっており、73% の転写産物がセンス-アンチセンス転写をおこなっていることが明らかとなった。その中には、coding RNA と ncRNA のペアだけではなく、coding RNA のみのペアや、ncRNA のみのペアもふくまれており、これらのセンス-アンチセンス RNA ペアには転写レベルで遺伝子の発現制御にかかわるものも多く存在することが示された®、また、ncRNA のうち

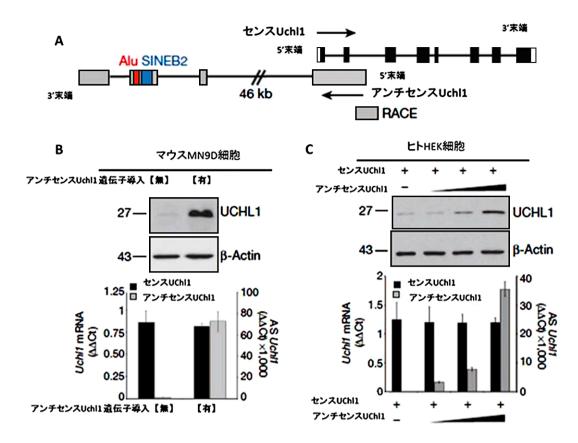


Fig. 2 タンパク質合成を促進するアンチセンス RNA.

(A) センス Uchl1 とアンチセンス Uchl1 のゲノム上の配列構造。センス Uchl1 の黒い部分はエクソン,白い部分は非翻訳領域を表す。アンチセンス Uchl1 の灰色の部分はエクソン,Alu,SINEB2 は反復配列領域である。5 末端が重要であることが示された。(B) マウス由来細胞株(MN9D)にアンチセンス Uchl1 を導入し,UCHL1 タンパク質が発現した。(C) とヒト由来細胞株(HEK)にセンス Uchl1 およびアンチセンス Uchl1 を導入し,UCHL1 タンパク質が発現した。

塩基配列が $100\sim200$ bp 以上のアンチセンス lncRNA は、これと相補的な配列を持つ mRNA(センス RNA)と結合して、その翻訳を阻害することが知られてきた。最近では、細胞質において lncRNA の TINCR が STAU1 タンパク質とともに標的 mRNA の安定性を制御する可能性なども報告されている 90

翻訳レベルで遺伝子発現を制御するアンチセンス RNA

Carrieri らは、脳機能や神経変性疾患に関与して脳内でセンス RNA-アンチセンス RNA のペアが発現している遺伝子 Uchl1 に着目し、その転写翻訳機構の解析をおこなった $^{10)}$. マウスの細胞から Uchl1 のアンチセンス IncRNA (アンチセンス Uchl1) の配列を解析したところ、Uchl1 のセンス RNA (センス Uchl1) と相補しない領域に 2 つの SINE (SINEB2, Alu) を同定した。アンチセンス Uchl1 の 5 末端には、センス Uchl1 の 5 末端と結合する配列を有しており、このアンチセンス Uchl1 の 5 大端と結合する配列が哺乳類の遺伝子に共通に存在することが明らかとなった。

センス Uchl1 の発現がみられるマウスの細胞では、アンチ

センス Uchl1 を発現させると、センス Uchl1 量に変化はなかっ たものの、UCH11 タンパク質の合成量が増加した. 本来 Uchl1 を発現しないヒトの細胞においても、センス Uchl1 と アンチセンス Uchl1 を強制発現させたばあいには UCHl1 タ ンパク質の合成量が増加した (Fig. 2). また、タンパク質の 合成量増加には SINEB2 が必要であり、アンチセンス Uchl1 内に存在する位置が重要であることが明らかとなった。さら に、アンチセンス Uchl1 のタンパク質合成を促進する経路を 明らかにするため、一般的な mRNA の翻訳開始機構である CAP 構造を抑制して翻訳機能を阻害したところ、アンチセ ンス Uchl1 の SINEB2 が、通常の翻訳とはことなる機能で mRNAのリボソームへの移行し、タンパク質合成を促進した. また、タンパク質の生合成を制御する酵素 mTORC1 の阻害 を抑制すると、通常は核に多く存在するアンチセンス Uchl1 を細胞質への移行がみとめられた. このようなアンチセンス Uchl1 にみられる機能は、ストレスなどの外的要因により通 常の翻訳開始機構を担う遺伝子発現が抑制された時に、必要 なタンパク質を合成して生き残るための保存的機能である可 能性が示唆された。

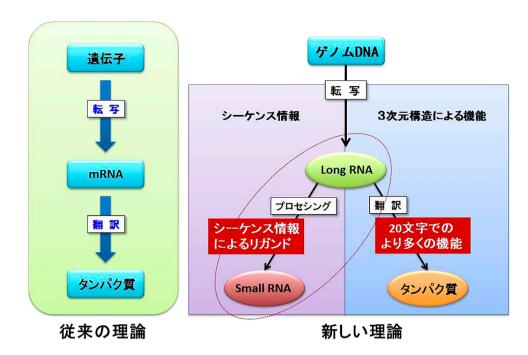


Fig. 3 新しいセントラルドグマの概念. Long RNA と Small RNA は遺伝子発現制御の中心的な役割を果たす.

おわりに(新しいセントラルドグマの概念と今後の展望)

生命科学では、DNAにコードされている遺伝子情報はRNAへ書き換えられ(転写)、それをもとにタンパク質へと変換される(翻訳)という生命活動の流れは揺るぎない柱「セントラルドグマ」とされてきた。しかし、RNA研究が進むにつれて、ncRNAは生命を理解するためには必須の存在であり、DNA・RNA・タンパク質と続く従来のセントラルドグマの中に組み込んで考える必要があるだろう(Fig. 3).

ncRNA の機能を解明する研究領域が進展することにより、今までタンパク質のアミノ酸配列を有する遺伝子の違いでは説明できなかった生物種の違いや、生命の発生や、分化、これまで発症原因が不明とされている多くの疾患について、その発症メカニズムを包括的に解明される可能性が考えられる。また、機能性 RNA の研究が、早期診断のためのバイオマーカーへの応用や、その多様な機能を利用した新しい医薬品開発や再生医療技術をふくめた医療領域へ応用されることが今後さらに期待される。

※本論文に関連し、開示すべき COI 状態にある企業、組織、団体はいずれも有りません。

文 献

1) Kawai J, Shinagawa A, Shibata K, et al. Functional annotation of

- a full-length mouse cDNA collection. Nature 2001;409:685-690.
- 2) Lander ES, Linton LM, Birren B et al. Initial sequencing and analysis of the human genome. Nature 2001;409:860-921.
- Okazaki Y, Furuno M, Kasukawa T, et al. Analysis of the mouse transcriptome based on functional annotation of 60,770 fulllength cDNAs. Nature 2002;420:563-573.
- Carninci P, Kasukawa T, Katayama S et al. The transcriptional landscape of the mammalian genome. Science 2005;309:1559-1563
- Naganuma T, Nakagawa S, Tanigawa A, et al. Alternative 3'-end processing of long noncoding RNA initiates construction of nuclear paraspeckles. EMBO J 2012;31:4020-4034.
- 6) Gong C, Maquat LE. IncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3'UTRs via Alu elements. Nature 2011;470:284-288.
- Yoon JH, Abdelmohsen K, Srikantan S. et al. LincRNA-p21 suppresses target mRNA translation. Mol Cell 2012;47:648-655.
- 8) Katayama S, Tomaru Y, Kasukawa T, et al. Antisense transcription in the mammalian transcriptome. Science 2005; 309:1564-1566.
- Kretz M, Siprashvili Z, Chu C, et al. Control of somatic tissue differentiation by the long non-coding RNA TINCR. Nature 2013;493:231-235.
- 10) Carrieri C, Cimatti L, Biagioli M, et al. Long non-coding antisense RNA controls *Uchl1* translation through an embedded SINEB2 repeat. Nature 2012;491:454-457.

Abstract

The world of functional RNA

Hiromi Okada, Ph.D.¹⁾ and Yoshihide Hayashizaki, M.D., Ph.D.¹⁾

¹⁾Preventive Medicine and Diagnosis Innovation Program, RIKEN Research Cluster for Innovation

The development of next-generation sequences has brought not only high-throughput sequencing but also new possibilities for various kinds of analysis methods of genetic information. Dr. Hayashizaki et al. developed new technologies to construct the full-length cDNA library and applied them to high-throughput sequencing technologies for large-scale transcriptome analysis. These analysis results overturned the conventional assumption the 2% of the genome is transcribed by showing that 70% or more of the genome is transcribed as RNA through FANTOM activities which was founded in 2000 on their initiative. Further, the existence of 23,000 non-protein coding RNAs was confirmed. These new findings redefine the central dogma into a new picture containing new interaction cascade and the unexpected complexity of combined omics. The neo central dogma shows that there are three types of final products derived from genes; long ncRNA, small ncRNA, and protein. They play essential roles by forming complexes with each other to maintain life. Long ncRNA and small ncRNA play a role as a ligand with sequence information. Long ncRNA and protein play a role as a functional molecule. Here, I would like to introduce the neo central dogma concept and some of the mechanisms of ncRNAs.

(Clin Neurol 2013;53:957-961)

Key words: non-coding RNA, long non-coding RNA, sense-antisense RNA, transcriptome analysis