

解読されたゲノム・トランスクリプトーム情報と 新しい医学研究の展望

林崎 良英^{*,**} 金森 睦^{*}

要 旨: マウスエンサイクロペディア計画について、このプロジェクトを開始するに至った経緯、戦略、得られた結果、今後の展望について詳述したい。これにより、予想を上回る数のncRNAやアンチセンスRNAの存在が明らかになった。さらに、予想をはるかに上回るバリエーションをもつRNAが存在することも証明された。ゲノム情報が次々に解明される中、効率の良い情報提供手段の開発(DNABookTMなど)についても紹介する。(J. Jpn. Coll. Angiol., 2004, **44**: 53-57)

Key words: ncRNA, Mouse encyclopedia project, DNABookTM

ゲノムが「生物が生活環を営む上で必要な遺伝情報のすべて」と定義される基本設計図であることから、ヒトを含めたゲノム計画の必要性が議論され、急速な展開を遂げてきた。ヒトゲノムのドラフト配列が発表された翌々年には29億塩基対のヒトゲノム解読完了が宣言された^{1,2)}。しかしこの情報は、AGCTの4文字で記載された暗号情報であり、このままでは何を意味しているのかわからない。ゲノムに収められた遺伝情報を理解するためには、その最終的機能産物であるタンパク質などの構造遺伝子部分の情報とその発現情報を担う発現調節部位の情報を俯瞰することが必須である。この遺伝情報を担う大きな機能のひとつが、RNAに転写される領域であり、生理活性を有するタンパク質やRNAの情報、最終産物の構造情報を備えて生命の営みを支え、我々が生命を理解することに直結するものである。このRNA全体をトランスクリプトームと呼ぶが、マウスはゲノム及びトランスクリプトームの双方が解読された科学史上初めての生物である。

本稿では、マウスのゲノム解読の持つ意味と広範なトランスクリプトーム解析を紹介するとともに、マウスの一生の間に各組織でゲノムから転写されるmRNAを完全長cDNAのかたちで網羅的に収集し、それらの発現プロフィール、タンパク質間相互作用などの情報を体系的に整理するマウスエンサイクロペディア計画

について、このプロジェクトを開始するに至った経緯、戦略、得られた結果、さらに今後の展望について、詳述したい。

マウスゲノム解読の意義 —トランスクリプトーム解析

1990年にヒトゲノムプロジェクトが始まった時、ゲノム配列解析を実施する5種類の主要なモデル生物の1つにマウスが含まれていた。ヒトゲノム配列の解析が相当に進んだ1999年になると、ゲノム配列解析機能を備えた大規模施設が協力し、マウスゲノム解読コンソーシアムMGSC(Mouse Genome Sequencing Consortium)の名の下でマウスゲノムの配列解析を始めた。ゲノム配列には遺伝子の塩基配列が含まれており、種々のコンピュータソフトにより予測可能であるが、実際にコードする領域を予測することはまだまだ容易ではない。そこでゲノムDNAから転写された産物(transcripts, 総体をtranscriptomeという)であるRNAを同定し、遺伝子産物の機能を同定するトランスクリプトーム解析が必須である。このトランスクリプトーム解析は、理化学研究所の筆者らのグループを中心に、世界中から100人以上の一流の研究者を集め、国際FANTOM(Functional Annotation of Mouse cDNA)コンソーシアムにより、大規模な発現遺伝子解析結果(トランスクリプトームデータとアノテーションデータベース)としてまとめられた。これはMGSCが行ったマウス

* 独立行政法人理化学研究所ゲノム科学総合研究センター(GSC)遺伝子構造・機能グループ

** 同生体分子機能研究室(和光)

2004年2月5日受理

ゲノムの全解析結果とあわせて、2002年12月にマウスゲノム特集号としてNatureに報告された^{3,4)}。マウスゲノムについては、既にセラ・ジェノミクス社(Celera Genomics)が、ゲノム全体の塩基配列を一度に決定できる全ゲノムショットガン配列決定法により、マウスのドラフト配列を発売していた。ゲノムカバー率5倍のマウスゲノムが民間の顧客に提供されているが、有償であり解析されたデータには制限が設けられていたもので、これらの報告により広く一般の研究者にマウスゲノム情報が提供されることになった。

大規模完全長cDNA発現解析

ヒトゲノムドラフト配列の完成で得られるゲノム情報は、疾患を解明し創薬につなげるなどの研究に非常に有用である。しかし同時に、疾患メカニズムの解析などにおいては、ゲノム配列だけでは完全にわからない部分に焦点をあてる必要がある。先にも述べたが、ゲノム情報で得られるのは染色体のDNA配列であり、個々の遺伝子についての正確な構造予測は現在では不可能である。また遺伝子構造がわかったとしても、タンパク質の機能や高次構造などを完全に予測することも不可能である。この意味で、遺伝子の構造配列である完全長cDNAの研究は、ゲノム計画と並び称されるきわめて重要なもう一つのゲノムプロジェクトである。

筆者らのグループでは、マウスを題材としてすべての遺伝子の完全長cDNAを単離するという「マウスエンサイクロペディアプロジェクト(マウス百科事典計画)」を行い発表してきた。効率的に遺伝子を収集するため、完全長cDNAライブラリー作製技術の開発を行った。これにより全クローンの90~95%が全長を含有するような完全長cDNAライブラリーを出発点とすることに成功した。これは可能な限り多様な組織を用いた総数246の異なる成長段階・組織に由来するライブラリーを用い、受精卵から成体までのマウスのほとんど全ての成長段階を網羅している。この理研完全長cDNAの多くは、既知遺伝子、EST発現遺伝子情報、マーカー遺伝子などとともにマウスゲノム上にはもちろんのこと、ヒトゲノム配列上にもマップされている。よって疾患を遺伝学的にヒトゲノム配列上にマップすれば、疾患原因遺伝子を効率的に選り出すことが出来るように整理されている。解析から得られた

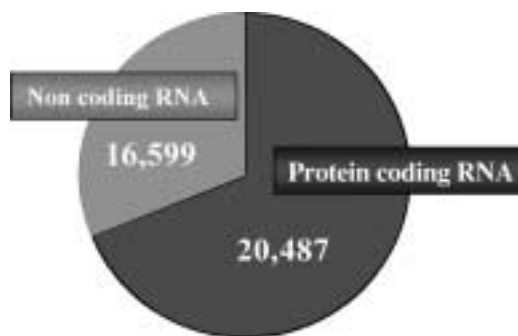


Figure 1 Results of FANTOM.

60,770個の配列情報には機能情報を注釈(annotation)し、エンサイクロペディアとしての機能を充実させた結果、クラスターごとに分類された遺伝子の種類数は37,086個であった。おもしろいことに、この全クラスターに対して機能解析を行った結果、タンパク質を発現するものは全体の53%であり、47%はタンパク質合成には直接関係していないncRNA(non-coding RNA)であることが判明した(Fig. 1)。

ncRNA

ncRNAには既知である、タンパク質合成に関与するリボソームRNA(rRNA)や転移RNA(tRNA)、スプライシングに関与する核内低分子RNA(small nuclear RNA: snRNA)のようによく知られているもの以外に、多様な機能性RNAが含まれている(Table)。マウスRNAの47%のncRNAのうち、71%がスプライシングを受けておらず単一のエキソンのみで構成されていた。これはmRNAでは80%以上がスプライシングを受けていることと対照的である。さらにスプライシングされていないncRNAのうち26%はマウスゲノムと比較してスプライシングされている構造をなしていた。タンパク質をコードしていないにもかかわらずスプライシングされているということは、何らかの生物学的機能を有していることを強く示唆している。ncRNAは細菌から哺乳動物まで広い生物種に多数存在しており、それ自身で機能を持ち、RNAやタンパク質、あるいはゲノムに直接作用して遺伝子発現における様々な働きを担うことが明らかになりつつある。例えば哺乳動物に先行した線虫のゲノムトランスクリプトーム解析におい

Table Classification of functional ncRNAs

RNAの種類	特 徴
(1) tRNA	タンパク質合成(翻訳)過程で働く。
(2) catalytic RNA	rRNAに代表される酵素活性を持つRNA分子。 自己スプライシング能などに働くリボザイムRNAなど。
(3) stRNA (small temporal RNA)	発生に関与する特異的な遺伝子のmRNAの5'-非翻訳領域に結合し、 その翻訳を抑制することにより発生のタイミングを制御すると推察されている。
(4) snRNA (small nuclear RNA)	スプライシングを司るタンパク質とRNAの複合体であるスプライソソームの実体であるRNA。 U1, U2, U4, U5, U6はmRNAのスプライシングに関与。
(5) snoRNA (small nucleolar RNA)	snRNAのうち細胞内の核小体にあるものを呼び、主にrRNAの生成に関連している。 RNA修飾に機能していると推察されている。 U3, U8, U13, U14, U15など、リボソームRNAのプロセッシングに関与するものが多い。
(6) miRNA (micro RNA)	ほとんどのmiRNAが特異的なmRNAの5'-非翻訳領域に結合してその翻訳を 抑制すると考えられている。
(7) sense/antisense RNA	センス遺伝子エキソン中の配列でペアを形成する、少なくとも20塩基以上の センス鎖-アンチセンス鎖。siRNA (small interfering RNA)が介在するとされる、 RNA interference(RNA干渉: RNAi)のような特異的RNA分解にも関与していると考えられる。

て、時期特異的な発現を示して発生のタイミングを制御している、翻訳されない低分子RNAが数多く発見されている。哺乳動物では正常胚発生に重要であるX染色体の不活性化において、従来知られていたDNAやヒストンのエピジェネティックな修飾に加えて、数種のncRNAがきわめて重要な役割を果たしていることが判明した。これらはncRNAが遺伝子発現を制御する未知のエピジェネティックな修飾構築などに深く関与していることを容易に推察させる。

ヒト21番染色体の配列全体をマウスゲノム中の相同領域と比較したところ、機能未知の保存されたブロックが多数同定され、ほとんどがncRNAであることが示された。

機能性RNAの重要性は知られていても従来の研究方法ではなかなか単離が難しく、マウストランスクリプトーム解析から明らかになった多数のncRNAは網羅的トランスクリプトームの最大の功績であると言っても過言ではない。この「RNA大陸」の発見と呼べるほどの成果に関してはまだまだ未知数であるが、今後の研究が大いに期待される。

また選択的スプライシングが組織特異的に起こることを示す例は非常に多く、mRNAそしてタンパク質の

多様性を作り出す最大の要因といえる。ヒトの疾患に関連した遺伝子変異のうち、約15%は選択的スプライシング制御がうまくいかなかったり、mRNAプロセッシングの誤りが関係していると言われている。このような疾患の解析にもマウストランスクリプトーム解析の担う重要性は明快であるといえる。

DNABook™

このように、膨大なゲノム情報が次々に解明される中、これらを効率よく活用することが個々の医学研究者に求められるのは明らかである。また情報発信側にも効率の良い情報提供手段の開発が必要である。

マウスゲノムエンサイクロペディアプロジェクトの目的である、遺伝子クローンと遺伝子情報の両方を、研究者など利用者に提供する方法を実現したのがDNABook™である。**Fig. 2**は理化学研究所で試作したマウスDNABook™である(**Fig. 2**, マウスDNABook™第1版)。ここには**Fig. 2**と同じ60,770個の遺伝子クローンが一冊のDNABook™に非常にコンパクトに収録されている。DNABook™の画期的な特徴は、これまで情報の媒体であった書籍を、物質を運搬する媒体として利用する点にある。利用者は書籍を購入する手軽さで

DNABook™を発注し、クローンを頒布する機関は在庫のDNABook™を宅配便で発送し、利用者は受け取ったDNABook™から遺伝子情報だけでなく、遺伝子クローンそのものを容易に取り出し、即座に自分の実験を始めることができる(**Fig. 3**, 概念図)。あらかじめDNABook™を一冊購入して実験机に常備しておけば、実験のアイデアを思いついたその日に、実験を始めることも可能になる。DNABook™は、遺伝子クローンが整備されたポストゲノム時代の遺伝子研究の新しいスタイルを可能にする。

ゲノムネットワーク

DNABook™は、非常に多くの情報を運ぶが、この情報は、少なくとも生命科学が目指す最終駅ではない。「生命戦略」を解明するための一里塚は、生命戦略を「分子システム」として理解することである。そのためには、生命の営む現象または、外的刺激に対する生命体の反応などを分子のレベルで明らかにする必要がある。たとえば、疾患の原因遺伝子と疾患の症状、薬の標的分子と薬効、改変遺伝子と遺伝子改変作物・家畜などなどを分子のレベルで結びつけることである。そのためには、これらの間を分子レベルで説明するいわゆる分子ネットワーク(分子回路)の解明が必要であ

る。これらの生命現象を説明する網羅的な分子ネットワークを「ゲノムネットワーク」と呼んでいる。ゲノムネットワークの解明こそ分子生物学が目指す最初のゴールであり、間違いなく、これからの生命科学のカギになるであろう。

謝 辞

一連の研究は、文部科学省の「マウスcDNAエンサイクロペディア計画」の一環として行われたものである。この原稿を作成するにあたって西川実希氏の協力を得た。

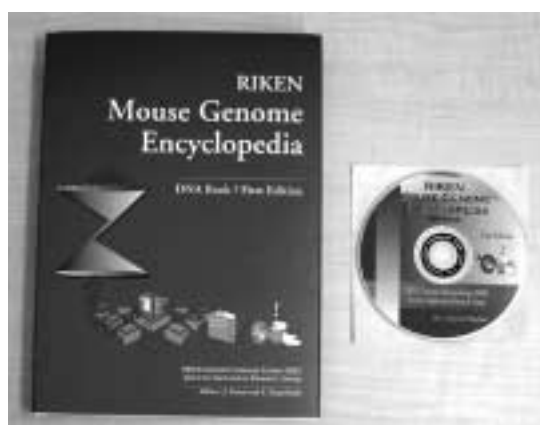


Figure 2 'DNABook™'.

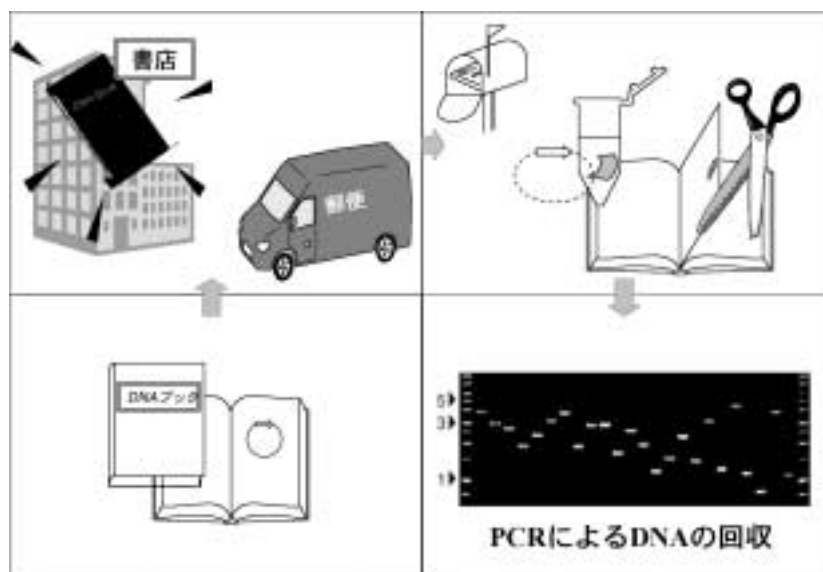


Figure 3 Concept of 'DNABook™'.

文 献

- 1) International Human Genome Sequencing Consortium: Initial sequencing and analysis of the human genome. *Nature*, 2001, **409**: 860-921.
- 2) Venter JC, Adams MD, Myers EW et al: The sequence of the human genome. *Science*, 2001, **291**: 1304-1351.
- 3) Mouse Genome Sequencing Consortium: Initial sequencing and comparative analysis of the mouse genome. *Nature*, 2001, **420**: 520-562.
- 4) The FANTOM Consortium & the RIKEN Genome Exploration Research Group Phase I & II Team: Functional annotation of a full-length mouse cDNA collection. *Nature*, 2002, **420**: 563-573.

Outlook for Dynamic Transcriptome Analysis and Future Life Science

Yoshihide Hayashizaki^{*,**} and Mutsumi Kanamori^{*}

^{*} Genome Exploration Research Group, Genome Science Centre (GSC), RIKEN

^{**} Genome Science Laboratory (Wako), RIKEN

Key words: ncRNA, Mouse encyclopedia project, DNABookTM

One of the goals of life science is to understand all of the biological phenomena in a system of biomolecules, named “genome network”.

As the first step for the analysis of “transcriptomics”, we have been working to establish the mouse’s comprehensive full-length cDNA collection, named Riken mouse genome encyclopedia. The findings revealed that these sequences contained 33,409 unique sequences with more than 20,487 clear protein-encoding genes. Using these results, the functional analysis of specific proteins gave a new discovery to proteome research.

In this analysis it was turned out that the significant population of non-coding RNA as well as protein, functions directly by various mechanisms. Non-coding RNA will be an essential component in genome network analysis.

One of the challenges in genome network analysis is the delivery of full-length cDNA clone distribution. “DNA BookTM” is a powerful tool to distribute the full length cDNA clones essential for genome network research. In this technology, DNA clones or PCR products are printed onto the pages of books. (*J. Jpn. Coll. Angiol.*, 2004, **44**: 53-57)