

2002年12月5日
独立行政法人 理化学研究所

遺伝子の機能解析を飛躍的に進める世界最大規模の遺伝子情報を公開

- 遺伝子として認知されていなかった部分が転写されていることを実証 -

理化学研究所（小林俊一理事長）は、マウスの完全長 cDNA^{*160}、770 クローンの塩基配列および機能アノテーション（機能注釈）情報を公開します。これは、現在までに人類が取得している遺伝子の約 9 割に当り、1 つの生物種のトランスクリプトーム^{*2}データとして世界最大、最高品質のものです。さらに、遺伝子としては認知されていなかった部分が、実際には遺伝子として転写されていることを初めて科学的に実証しました。理研横浜研究所ゲノム科学総合研究センター（和田昭允センター所長）遺伝子構造・機能研究グループ（林崎良英プロジェクトディレクター）および FANTOM コンソーシアム^{*3}によって成し遂げられた研究成果です。

ゲノムの塩基配列だけからは、遺伝子を完全に予測することは不可能であり、遺伝子の予測、さらには実在を証明するためにはトランスクリプトームによる裏づけが必要不可欠です。さらに、最終的な生物機能活性を有するタンパク質を得るためには、完全長 cDNA の解析なしでは得られません。今回、得られた研究成果の核心は、以下の通りです。

- ①高品質の完全長 cDNA クローンの最大規模の収集と塩基配列
- ②高精度の機能アノテーション情報
- ③マウスゲノムとトランスクリプトームとが一体となったデータ

マウスは、ヒトのモデル動物として医学的に重要な役割を果たします。本研究成果であるマウスのトランスクリプトームと、マウスの全ゲノムシーケンスとが併せて解析されることにより、遺伝子の機能研究、タンパク質の構造解析研究、がん遺伝子の研究、および新しい薬の開発を劇的に進展させるものと期待されます。

本研究成果は、英国の科学雑誌『nature』12月5日号に掲載されるとともに、新たな機能アノテーションを付与した 60, 770 クローンの塩基配列と、機能アノテーション情報は、理研内のホームページ(<http://fantom2.gsc.riken.go.jp/>) および DDBJ(<http://www.ddbj.nig.ac.jp/>)にて公開されます。また、FANTOM2 クローンセットは、「RIKEN-FANTOM2 クローンセット」の名称で頒布する予定です。

1. 背景

マウスは、“ヒトとほとんど同じ遺伝子を持っていること”、“ヒトでは取れないライフサイクルすべてのステージの遺伝子が取れること”、“遺伝学的実験が行いやすいこと”“発生生物学的手法による、遺伝子破壊マウスなどが作成しやすいこと”などから、ヒトのモデル動物として医学生物学的研究の中心となっています。マウス遺伝子（完全長 cDNA）のすべてを取り出し、その塩基配列情報などを体系的に整理した辞書となる「マウス遺伝子エンサイクロペディア」の作成は、ヒト遺伝子の機能解明の促進など、医学・生物学といったさまざまな分野での貢献が期待されます。また、マウスのゲノムプロジェクトは、ヒトゲノムプロジェクトと並行して進めら

れており、世界的にも重要課題とされています。それらの中で、非常に高い完全長 cDNA のカバー率を誇るマウスエンサイクロペディア計画は極めて重要です。

ゲノム科学総合研究センター (GSC) 遺伝子構造・機能研究グループでは、これまでマウス完全長 cDNA100 万クローン以上の末端塩基配列データを、日本 DNA データバンク (DDBJ) および理研内のホームページ

(<http://genome.gsc.riken.go.jp/>) で公開してきました。また、2000 年 8 月から 9 月にかけて行われた FANTOM (Functional Annotation of Mouse)¹ マウス cDNA 機能アノテーション国際会議において、約 2 万クロンの完全長 cDNA 配列に対する機能情報を付加し、その成果を『nature』(2001 年 2 月 8 日号)に発表しました。さらに、2002 年 2 月より 4 月にかけて、「機能アノテーション国際遠隔会議

(MATRICS : Mouse Annotation Teleconference for RIKEN cDNA sequence)」を開催し、国内外のゲノム科学、生物学などの専門家と共同で FANTOM1 の 2 万クローンに加え、今回新たに決定した 4 万クローンを加えた約 6 万クロンのマウス cDNA の完全長配列について、機能アノテーション情報を付与し、マウス百科事典の世界標準化を目指した機能注釈などについて議論してきました。4 月 29 日から 5 月 5 日にかけては、「FANTOM2 Cherry Blossom 国際会議」を開催し、新規遺伝子の発見などについて活発な報告と議論を行いました。

一方、理研のマウスエンサイクロペディア計画から得られたマウス完全長 cDNA の塩基配列と、マウスゲノムの塩基配列決定を行う国際組織である MGSC (Mouse Genome Sequencing Consortium) が決定したゲノムの塩基配列とを相互に交換し解析することにより、互いの成果を相補的かつ飛躍的に発展させました。

2. 研究手法と成果

本研究の対象である完全長 cDNA (遺伝子の本体) は、タンパク質を合成するための設計情報をすべて有し、タンパク質そのものを合成することができることから、ポストゲノム研究における非常に重要な「基盤ツール」となります。研究グループでは今回、約 6 万クローンもの完全長 cDNA について、“それらの配列はどのような性質の遺伝子なのか”、“どのようなタンパク質をコードしているのか”などの機能情報の注釈付け (機能アノテーション) を行いました。cDNA に対する機能アノテーションを行うためには、コンピューターによる解析と、それを専門家により確認するための支援システムが必要不可欠です。研究グループでは、「機能アノテーション自動システム^{*4}」、「機能アノテーション入力システム^{*5}」、「cDNA マッピングシステム^{*6}」、「cDNA クラスタリングシステム^{*7}」を開発し、解析に取り組みました。その結果、以下の研究成果が得られました。

1.FANTOM2 クローンセット (約 6 万クローン) のマウス完全長 cDNA について、配列相同性解析、機能モチーフ解析、タンパク質コード領域解析に加えて、マウスゲノムへのマッピング情報や発現情報を用いた総合的な機能アノテーションを行いました。

2.マウスゲノム配列との比較から、FANTOM2 クローンセットの中には約 33, 000 種類の遺伝子が含まれていることが分かりました。この FANTOM2 クローンセットは、現在、公共データベースに登録されたマウスの遺伝子を合わせた全遺伝子の 90% をカバーしています。

3. FANTOM2 クローンセットに含まれる約 33, 000 種類の遺伝子のうち、約 16, 000 種類は、今回新たに発見された遺伝子です。また、FANTOM2 クローンセットの中には、ゲノム上において互いに逆向きに転写されている遺伝子ペアが約 2, 000 あること分かりました。
4. マウスゲノムとトランスクリプトームを一体で解析することにより、ゲノム解析だけでは遺伝子としては認知されていない部分が、実際には遺伝子として転写されていることを世界で初めて科学的に実証しました。このことは、ゲノムの塩基配列をもとにしたコンピューターによる遺伝子予測では、正確な遺伝子数を確定できないことを物語っています。
5. 確認された全遺伝子のうち、約 40% が選択的スプライシング（オルタネイティブ・スプライシング）^{*8} され、そのうちの約 80% が、異なる構造を持つタンパク質をコードすることが分かりました。このことにより、タンパク質の数は、遺伝子の数よりはるかに多いということが判明しました。
6. 遺伝子の中にはタンパク質をコードする「coding RNA」のみならず、多数の「non-coding RNA^{*9}（タンパク質情報をコードしていない RNA）」があることが分かりました。これらの non-coding RNA は、ゲノムの塩基配列だけでは予測ができず、トランスクリプトームのデータを用いてのみ予測されます。このことは、タンパク質と同じく、non-coding RNA も病気などの発症に関わっていることが考えられ、新しい治療法の開発や創薬などに影響をおよぼす、非常に重要な発見と考えられます。

3. 今後の展開

本研究で得られた 60, 770 クローンの配列は、完全長 cDNA としての高いカバー率を誇ります。これらのクローンセット、塩基配列、機能アノテーション情報は、MGSC により配列決定されたマウス全ゲノムドラフト配列とともにヒト、マウスを始めとするさまざまな生物についての基礎研究及び医学または創薬などへの応用研究のための世界共通のプラットフォームとなるものです。

理研では今後、全長シーケンスを完了し、新たな機能アノテーションを付与した 60, 770 クローンの塩基配列と機能アノテーション情報を理研内のホームページ (<http://fantom2.gsc.riken.go.jp/>) および DDBJ (<http://www.ddbj.nig.ac.jp/>) より公開します。また、マウス完全長 cDNA 60, 770 クローンからなる FANTOM2 クローンセットについては、「世界標準」として国内外の分譲を希望する科学研究者の利用に供するため、頒布体制を整えていきます。

(問い合わせ先)

独立行政法人理化学研究所 横浜研究所

ゲノム科学総合研究センター

遺伝子構造・機能研究グループ

グループディレクター

林崎 良英

チームリーダー

岡崎 康司

Tel : 045-503-9222 / Fax : 045-503-9216

横浜研究所 研究推進部

木村 優

Tel : 045-503-9117 / Fax : 045-503-9113

(報道担当)

独立行政法人理化学研究所 広報室

嶋田 庸嗣

Tel : 048-467-9271 / Fax : 048-462-4715

<補足説明>

※1 完全長 cDNA

cDNA とは、ゲノム DNA の中から不要な配列を除き、タンパク質をコードする配列のみに整理された遺伝情報物質である mRNA (メッセンジャーRNA) を鋳型にして作られた DNA のこと。完全長 cDNA は、断片 cDNA と異なり、タンパク質を合成するための設計情報をすべて有しているため、タンパク質を合成することができる。この完全長 cDNA を効率的に合成するためには、非常に高い技術を必要とし、わが国が世界に先んじている。

※2 トランスクリプトーム

生物のゲノムから転写された遺伝子全体を指す用語。トランスクリプトームの解析を行うことにより、それぞれの遺伝子が転写されている組織や、発生段階などの情報が得られ、ゲノムの塩基配列から予測されるもののみではなく、コンピューターでは予測できない遺伝子の存在も実証することができる。

※3 FANTOM コンソーシアム

理研ゲノム科学総合研究センター遺伝子構造・機能研究グループの主導により結成された、マウス遺伝子の機能解析を行う組織。国内外のゲノム科学、生物学などの専門家と共同で、マウス cDNA に機能アノテーション情報を付与した。2002年4月29日から5月5日にかけて FANTOM2 (Functional annotation of Mouse) Cherry Blossom 国際会議では理研のほか、海外から The Jackson laboratory (米国)、TIGR (The Institute for Genomic Research : 米国)、NCBI (National Center for Biotechnology Information : 米国)、University of Queensland (オーストラリア)、EBI (European Bioinformatics Institute : ヨーロッパ)、国内から国立遺伝学研究所生命情報研究センター、大阪大学、慶應義塾大学など国内外のバイオインフォティクス、ゲノム科学、および生物学の研究機関から国外 38 人、国内 12 人の専門家が参加し、共同解析を行った。

※4 機能アノテーション自動システム

コンピューターにより、既知遺伝子配列情報とのホモロジー (類似性) 情報、遺伝子産物であるタンパク質のモチーフ情報 (類似機能を有するタンパク質群中で保存されている特徴的な構造) などから機械的に機能アノテーションを決定するための

システム。理化学研究所および NTT ソフトウェアとの共同研究により開発された。

※5 機能アノテーション入力システム

ゲノム科学や生物学の専門家がコンピューターによるさまざまな解析結果を参照しながら、必要に応じて機械的に決められた機能アノテーション情報の補正や、新たな機能情報を追加入力するためのシステム。理化学研究所および NTT ソフトウェアとの共同研究により開発された。

※6 cDNA マッピングシステム

cDNA が転写されたマウスゲノム上の位置を推定するための計算を行うシステムと、計算結果を専門家が確認するための表示システム。

※7 cDNA クラスタリングシステム

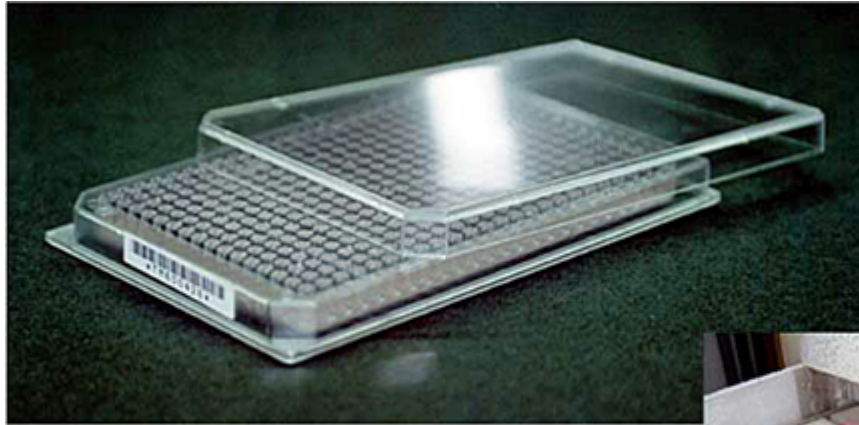
マウスゲノム上の同一位置から転写された cDNA をまとめ、選択的スプライシングが行われている mRNA を専門家が確認できるように表示するシステム。

※8 選択的スプライシング(オルタネイティブ・スプライシング)

選択的スプライシングとは、1つの遺伝子から複数の種類の mRNA が合成される現象のこと (cDNA は、mRNA を鋳型にして作られた DNA であるので、1つの遺伝子から複数の種類の cDNA が生じることがある)。

※9 non-coding RNA

タンパク質情報をコードしていない RNA。最近になってこのような non-coding RNA が生体内で実際に機能をもって遺伝子と同じような振る舞いをする例が報告され、注目されている。今回の研究成果の注目すべき点として、このような例が少なからずあることを証明した。

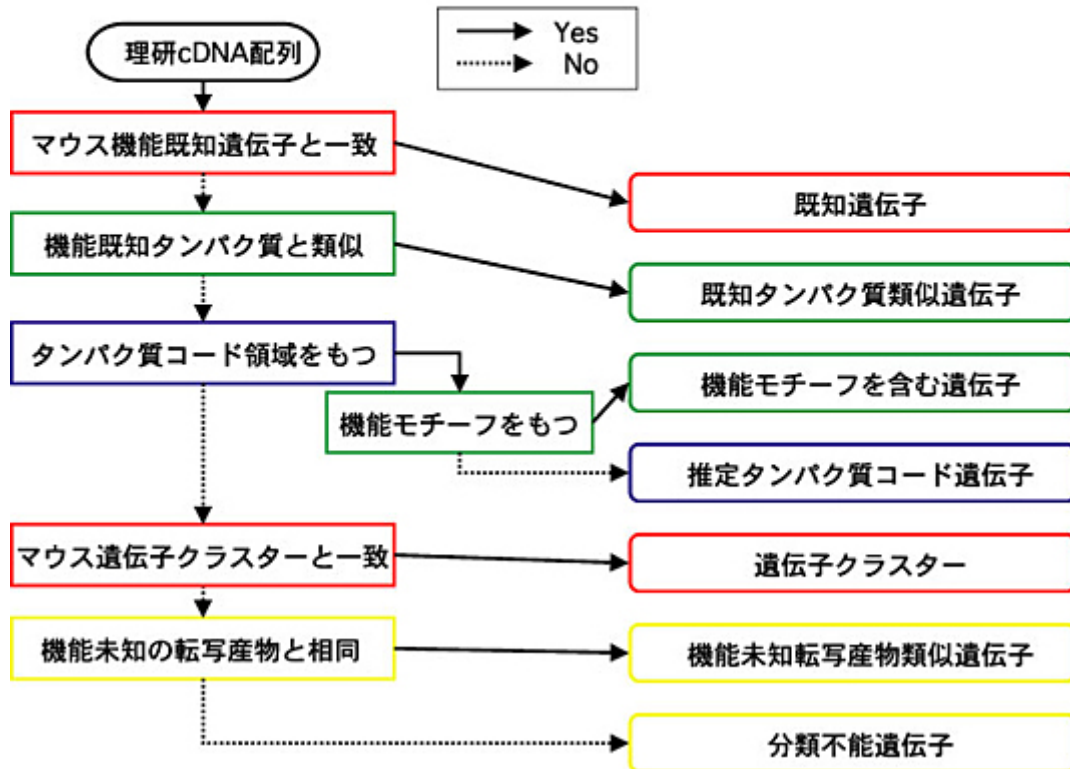


384穴遺伝子（完全長cDNA）凍結保存プレート

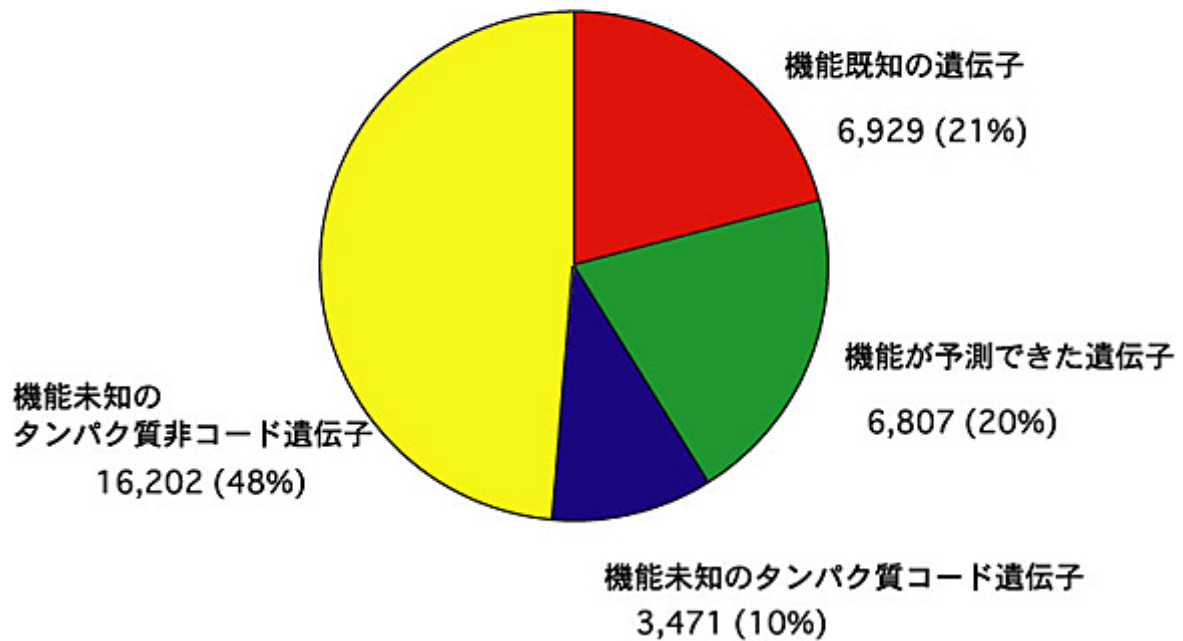


RIKEN-FANTOM2 クローンセット

機能アノテーションの方法



FANTOM2クローンの 機能アノテーションによる分類



FANTOM2 Cherry Blossom Meeting
(2002.4.29~5.5 於：理研横浜研究所)