

2001年2月8日  
独立行政法人 理化学研究所

## マウス cDNA アノテーション情報の公開

### - 約 12, 400 種のマウス新規遺伝子を発見 -

理化学研究所（小林俊一理事長）は、機能アノテーション（機能注釈）情報を付与したマウスの完全長 cDNA 情報を公開します。理研横浜研究所のゲノム科学総合研究センター（和田昭允センター所長）の林崎良英プロジェクトディレクター（遺伝子構造・機能研究グループ）は、世界に先駆けて「マウス遺伝子エンサイクロペディア計画」としてマウス完全長 cDNA<sup>\*1</sup>（約 21, 000 クローン）の解析を進めてきました。さらに昨年、「マウス cDNA 機能アノテーション会議（Functional Annotation of Mouse [FANTOM] meeting）<sup>\*2</sup>」を開き、遺伝子の機能注釈のルールや方法について取り決めを行い、遺伝子の機能注釈を効率的に行なうシステムを開発し、それらを用いて世界標準となる機能注釈をつけました。

本研究の大きな成果として下記の 2 点があげられます。

1. 約 12, 400 種のマウス新規遺伝子を発見し、中でも約 8, 200 種は哺乳類において全く新しい遺伝子であることがわかりました。
2. 新規遺伝子中には、発生の段階で遺伝子の転写調節を行なっていると考えられる新規遺伝子や、がんの増殖や抑制に関する遺伝子など病気に関連すると思われる新規遺伝子も見つかりました。

なお、当研究所は本年 5 月をめどに、機能アノテーションを付与した完全長 cDNA クローンを国内外の希望者が利用できる体制を整えていく計画です。

本研究成果は、英国科学雑誌「nature」の 2 月 8 日号に掲載されるとともに、理化学研究所(<http://www.gsc.riken.go.jp/e/FANTOM/>)などでも公開されます。

## 1. 背景

マウスは、遺伝子が約 4 万～10 万種あると予測されており、“ヒトの遺伝子とほとんど同じであること”、“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”、“実験的にも扱いやすいこと”などから、ヒトのモデル動物とされています。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きており、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遺伝子構造・機能研究グループでは、これまで、約 127, 000 クローンのマウス完全長 cDNA の末端塩基配列データをホームページ (<http://genome.gsc.riken.go.jp/>) 上で公開してきました。

さらに昨年 8 月 28 日から 9 月 8 日まで、国内外のバイオインフォマティクスおよび、ゲノム科学などの専門家が集い、上記マウス完全長 cDNA クローンのうち、

全長塩基配列の解析を完了した約 21, 000 クロンの完全長 cDNA について、機能アノテーションを行う国際会議 (FANTOM) が理研筑波研究所で開かれました。その結果の一部は、成田市で開かれた「第 14 回 国際マウスゲノム学会 (14th International Mouse Genome Conference : IMG, オーガナイザー : 林崎良英)」で報告されています。

## 2. マウス cDNA の機能アノテーション

### 1) 意義・目的

ヒトゲノムなど莫大なゲノム情報 (DNA 塩基配列) が全世界で急激に解読されつつあるなか、それらの情報をもとに、革新的な医薬品の開発等につなげるための、いわゆるポストゲノム研究の推進が重要となっています。近年、医薬品開発に直結するポストゲノム研究の主要な研究課題として、遺伝子から合成されるタンパク質 (生体反応を担う主役) に関する研究が重要視されており、激しい国際競争が展開されつつあります。本研究の対象である完全長 cDNA (遺伝子の本体) は、この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出来ることから、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。

今回、約 21, 000 クロンもの完全長 cDNA について、全塩基配列情報の解読に加え、それらが既知の遺伝子なのか、新規の遺伝子なのか、既知の遺伝子とはどのような類似性を有するのか、どのような性質の遺伝子なのかなどの機能情報を注釈付け (機能アノテーション) を行いました。このような膨大な量の完全長 cDNA について、単なる塩基配列情報だけでなく、機能情報をも付加することは、「基盤ツール」としての価値を飛躍的に高めるとともに、今後のポストゲノム研究の効率的推進に大きく貢献するものと期待されています。なお、昨年 8 月に行われた cDNA に関する機能アノテーション会議は、高等生物では世界初の試みであったことから、高等生物の cDNA についての機能アノテーションの世界標準を確立することにも貢献しました。

### 2) 機能アノテーションを効率的に行う新システム

cDNA に対する機能アノテーションは、既知遺伝子塩基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 (2 次構造あるいはそれらの組み合わせの構造情報) からある程度、機械的に推測することは可能です。しかし、高品質なアノテーションを付与するためには、遺伝学、免疫学、がん、神経系などの専門家がコンピュータ処理の結果を検証する作業が不可欠となります。ゲノム科学総合研究センターでは、NTT ソフトウェアとの共同研究により、高品質な機能アノテーションを決定するための支援システムを開発しました。本システムの特徴は、以下の 2 点です。

1. 機能アノテーションを素早く決定できるように、各種のコンピュータ処理の結果を図を用いて見やすく表示します。
2. 基本的な機能アノテーションの入力は、手入力ではなく、マウス操作による選択を行うだけで実現できるようにするなど、入力の間違いを極力減らすような設計がされています。

### 3) 得られた成果と情報の公開

本研究で解析した 21, 076 のマウス完全長 cDNA クローンには、選択的スプライシング\*2などによる重複を除いた結果、約 15, 200 種の遺伝子が含まれていることがわかりました。この約 15, 200 種の遺伝子のうち、約 2, 800 種の遺伝子はマウス既知遺伝子でしたが、残りの約 12, 400 種の遺伝子はマウス新規遺伝子です。

さらに、約 12, 400 種のマウス新規遺伝子の中には、哺乳類において全く新しい遺伝子が約 8, 200 種見つかりました。これらの中には、発生の段階で遺伝子の転写調節を行なっていると考えられる新規遺伝子や、がんの増殖や抑制に関連すると思われる新規遺伝子も見つかりました。

これらの研究成果は、2月8日より理研ゲノム科学総合研究センターのホームページ(<http://www.gsc.riken.go.jp/e/FANTOM/>)及び「nature」のウェブサイト(<http://www.nature.com/>)から公開される予定であり、また、日本 DNA データバンク (DDBJ) 及び、米国のジャクソン研究所

(<http://www.informatics.jax.org/>) のマウスゲノムデータベース中にも登録され、アクセスが可能になる予定です。

このようなヒトモデル動物であるマウスの新規遺伝子の大量な発見とそれら遺伝子の全塩基配列情報及び機能アノテーション情報の公開により、すでに公開されているヒトゲノムドラフト情報などと相まって、ヒト遺伝子の機能解明などの研究が飛躍的に促進することが期待されます。

## 3. 今後の予定

### 1) 今後の研究展開

本研究は、今後も「マウス遺伝子エンサイクロペディア計画」の一環として、マウス完全長 cDNA の取得と、その全長塩基配列の解析を進めるとともに、機能アノテーションを継続して実施し、順次、その研究成果を公開していきます。本年 9 月には第 2 回マウス cDNA 機能アノテーション会議を予定しており、世界各国から研究者が集い新たに 2 万個以上の完全長 cDNA クローンに注釈付けを行い、遺伝子の機能解析を推進する予定です。

また、マウス完全長 cDNA を活用した遺伝子機能解析等に関する共同研究についても国内外の研究機関と積極的に推進していきたいと考えております。

### 2) マウス cDNA クローンの利用

当研究所では、全長シーケンスを完了し、機能アノテーションを付与した今回の約 2 万クローンについて、その塩基配列及び機能アノテーション情報の公開だけでなく、完全長 cDNA クローンそのものについても本年 5 月をめどに、国内外の希望者が利用できるように体制を整えていきます。完全長 cDNA クローンの利用に関しては、学術利用を促進するため、“学術研究利用”と“産業利用”とを区別した利用料金設定をすることとしており、本年 4 月には利用方法や利用条件などを含め、その詳細について発表する予定です。

当研究所では今後さらに、全長シーケンスを完了し、機能アノテーションを付与した完全長 cDNA クローンについても、随時、国内外の希望者が利用できる

ようにしていきます。

(問い合わせ先)

<研究成果>

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

プロジェクトディレクター

林崎 良英

チームリーダー

岡崎 康司

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

<マウス完全長 cDNA クローン利用>

研究業務部

前川 治彦

Tel : 048-467-9762 / Fax : 048-462-4609

(報道担当)

独立行政法人理化学研究所 広報室

嶋田 庸嗣

Tel : 048-467-9272 / 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。

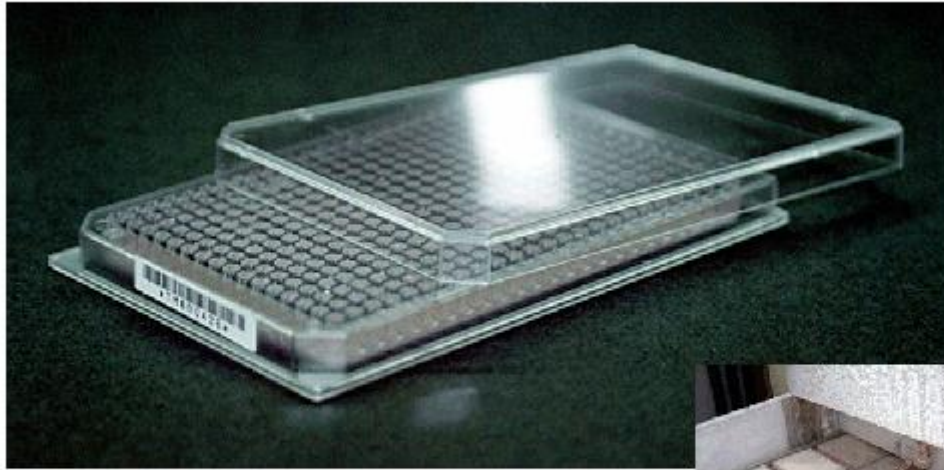
### ※2 マウス cDNA 機能アノテーション会議

機能アノテーションとは、解析対象である遺伝子について機能の注釈付けを行うこと。例えば、既知遺伝子塩基配列情報とのホモロジー (類似性) 解析、遺伝子産物であるタンパク質のモチーフ (2 次構造あるいはそれらの組み合わせの構造) 解析、遺伝子の属性分類等の方法が用いられる。

本会議では、海外の EMBL-EBI (European Bioinformatics Institute)、NCBI (National Center for Biotechnology Information)、TIGR (The Institute for Genomic Research)、The Jackson laboratory、UCB (University of California, Berkeley)、国内の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、国内外のバイオフィニクスおよび、ゲノム科学などの研究機関から国外 39 人、国内 16 人の専門家が参加し、共同解析を行った。

### ※3 選択的スプライシング

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）。



384 穴遺伝子（完全長 cDNA）凍結保存プレート



## 機能アノテーションの流れ



# FANTOM Viewer

The screenshot shows a Netscape browser window displaying the FANTOM website. The browser's address bar shows the URL <http://www.gsc.riken.go.jp/e/FANTOM/>. The browser's toolbar includes icons for back, forward, stop, refresh, home, auto-fill, print, and mail. The website content is as follows:

**Genome Exploration Research Group**  
RIKEN Genomic Sciences Center  
**Genome Science Lab**  
RIKEN Tsukuba Life Science Center


## FANTOM

### FUNCTIONAL ANNOTATION OF MOUSE

The FANTOM is a collection of functional annotations for the RIKEN full-length cDNA clones and provides not simply the gene functions themselves, but also many other informative and useful data describing functional information of the clones. FANTOM annotation data are available through [FANTOM annotation viewer](#).

- [About FANTOM](#)
- [Viewer Help](#)
- [Annotation viewer](#)
- [Reference](#)
- [FANTOM meeting](#)
- [Link](#)
- [FAQ](#)

[Contact Information](#)  
Last modified : 13-Feb-2001



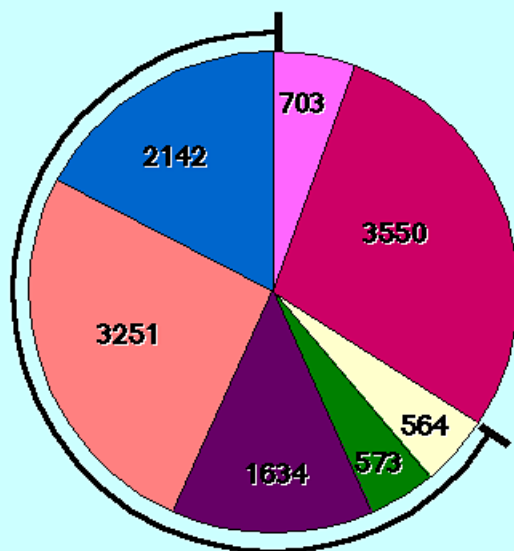
At the bottom of the browser window, the status bar reads "インターネットゾーン".

# Web interface

- \* 図を用いて各種の解析結果が見やすく表示される。
- \* 入力の間違いを極力減らすような設計がなされている。
- \* 遺伝子に付随する様々な機能情報を入力することが可能。

The screenshot shows a web browser displaying a gene analysis page. The page title is 'Transcription elongation factor 7 (by FASTA [protein]) SPTF-7cdJ'. The main content area shows a BLAST search result for 'Transcription elongation factor 7 (by FASTA [protein]) SPTF-7cdJ'. Below the BLAST result, there is a table with columns for 'Qualifier', 'Annotation text', 'Data source', and 'Accession'. The table contains several rows of data, including 'Transcription elongation factor 4 (SPTF-4)' and 'Transcription elongation factor 5-related protein 4 (SPTF-5)'. Callouts point to different parts of the interface: '情報表示フレーム' points to the top navigation bar, '解析結果表示フレーム' points to the BLAST result area, '入力支援画面' points to a 'Gene def' form, and 'アノテーション入力フレーム' points to the table below.

## マウス新規遺伝子の分類



- マウス既知タンパク質類似遺伝子
- ほ乳類既知タンパク質類似遺伝子
- 既知タンパク質類似遺伝子
- 既知モチーフを含む遺伝子
- 推定タンパク質をコードする遺伝子
- 既知転写産物類似遺伝子
- 分類不能遺伝子

ほ乳類新規遺伝子

ほ乳類新規遺伝子