



2009年10月30日  
独立行政法人理化学研究所

## 世界の研究者が協調し、マウス表現型データ共有化に着手

—マウス情報の次世代標準規格確立に向けた国際協力がスタート—

### 本研究成果のポイント

- 主要国のマウス表現型データを集約した、国際ポータルサイトの共同構築へ
- 国際間のデータベース統合に向け、日本のデータ集積拠点を理研サイネスが担う
- クラウドコンピューティング技術を活用し、大規模なデータの編纂を効率的に実施

EU・米国・カナダとわが国の研究機関は共同で、世界中のマウス表現型<sup>\*1</sup>データの共有化に向けた取り組みを開始し、独立行政法人理化学研究所（野依良治理事長）が、理研サイネス<sup>\*2</sup>を活用して、世界的なデータベース連携網における日本のハブ機能を担います。理研生命情報基盤研究部門豊田哲郎部門長、理研バイオリソースセンター（理研 BRC、小幡裕一センター長）マウス表現型知識化研究開発ユニット榎屋啓志ユニットリーダーらの研究グループを中心とする国際コンソーシアム「インターフェノーム (InterPhenome) <sup>\*3</sup>」による取り組みです。

ライフサイエンスの発展に伴い、遺伝子の働きが形質として現れた「表現型」の解析が重要視されるようになり、遺伝子に変異を持つマウスが多数作製され、形態から行動までの網羅的な解析が主流になりつつあります。表現型データは、多面的な観察結果を記録した複合データで、データ量も膨大です。加えて、調べるべき遺伝子の数も多く、単独の研究機関では対処しきれない事態が世界中で生じるようになりました。「役割分担」と「データ共有」に向けた国際的な連携が、その問題解決のための鍵とされていますが、用語や定義、データフォーマットの不統一などが、そうした連携を阻んでいました。

研究グループと EU・米国・カナダの研究機関は、2009年7月12、13の両日に滋賀県で開催した「国際マウス表現型情報統合シンポジウム (International Phenome Integration Meeting)」で、国際的な研究の分担やデータ共有に向けて共同で取り組むことに合意しました。具体策として、国際ポータルサイトの共同構築、語彙（ごい）の相互運用性確保、データライセンス付与システムの構築などを行うことになりました。その後、分担や方針、ライセンスなどの議論を経て、理研は、理研が得意とする次世代情報技術「セマンティックウェブ<sup>\*4</sup>」に準拠した、マウス表現型データの標準化作業を行うこととなりました。特に、国際間のデータベース統合に向けて、理研サイネスを活用し、日本のデータ集積拠点を担います。さらに、理研サイネスと理研のスーパーコンピュータをクラウドコンピューティング<sup>\*5</sup>技術で接続し、理研 BRC で生産している大規模な表現型データを効率的に取り扱います。

国際コンソーシアムは、すでにデータ提供用ポータルサイトの構築に着手しており、この取り組みについて、11月1日（日）～4日（水）に米国カリフォルニア州で開催される国際マウス哺乳類学会 (IMGC) において発表します。

## 1. 背景

病気の原因解明とその治療や予防方法などを研究する「疾患研究」では、マウスなどの実験動物を使って、遺伝子に変異した場合に病気などとしてどのように現れるか、すなわち「表現型」を詳細に観察することが重要視されています。このため、近年、大規模な突然変異マウス開発プロジェクトや、ノックアウトマウスを作製する実験が盛んになり、数多くの変異マウスが作製され、その表現型解析が世界中で精力的に行われています。これまでは、各研究機関が独自に表現型解析を行い、自らデータを管理するのが一般的でしたが、現在では、表現型を網羅的に解析してデータを集め、バイオインフォマティクス<sup>※6</sup>の手法を用いて大規模解析を行うことが主流となっています。表現型データは、病院の電子カルテに似て、多面的な観察結果を記録した複合データで、データ量も膨大です。このため、変異マウスの数と解析項目数が増えるのに伴い、指数関数的にデータ量が増大し、単独の研究機関では解析が追いつかない状態となっています。こうした問題を受け、国際的に研究を分担し、データを共有することが緊急の課題となってきました。

## 2. データ統合と共有化

文部科学省統合データベースプロジェクトの一環として、理研は2009年7月12、13日の両日、データ連携のための国際会議「国際マウス表現型情報統合シンポジウム (International Phenome Integration Meeting)」を滋賀県のホテルラフォーレ琵琶湖で開催しました(図1)。この国際会議では、マウス表現型における国際統合データベースの構築を最終目標として、世界中でマウスを用いた研究を効率的に推進するために、理研、国立遺伝学研究所(日本)、ユーロフェノーム (EuroPhenome : 欧州の表現型に関する統一機関) を代表して英国のハウエル研究所、米国のジャクソン研究所、カナダのヒト疾患モデル研究センター、カシミア (CASIMIR : 国際マウス情報資源維持連携コンソーシアム) の研究者が集まり、データの形式統一化などについて議論しました。その結果、(1) 国際ポータルサイトの共同構築 (2) 語彙の相互運用性確保 (3) データライセンス付与システム構築などを行うことで合意しました。

合意後、分担や具体的な方策について議論を行った結果、以下の方針が決定し、必要な開発に着手しました。

### (1) 国際ポータルサイトの共同構築

各機関は表現型データを収集して管理するとともに、そのデータを互いに共有化することで、インターフェノームでポータルサイトを共同構築する。一般利用者は、ウェブサイトを通じて、データベースからデータを自動的に抽出できるように、データはセマンティックウェブ形式で提供する(図2)。理研は、理研サイネスを活用することで、ポータルデータのセマンティックウェブ形式への変換を担当するとともに、国際ポータルサイトの共同構築においても、日本のデータ集積拠点として機能する。

### (2) 語彙の相互運用性確保

管理データなどの詳細情報、さらに疾患やゲノムなどの幅広い情報を表現型デー

タと統合して新発見を促すために、セマンティックウェブ技術への期待が高まっている。将来的に語彙の相互運用性を確保するために、セマンティックウェブの中核技術である RDF/OWL 技術<sup>\*4</sup> を利用した高度な統合技術の開発を行う。オントロジー<sup>\*7</sup>には、MA (Mouse Anatomy: マウス解剖学オントロジー)、MP (Mammalian Phenotype: マウス表現型オントロジー)、PATO (Phenotypic Quality Ontology: 表現型・性質オントロジー) を共通利用する。これらと理研が採用するオントロジー「YAMATO (Yet Another More Advanced Top-level Ontology)」に基づき、データ形式の運用を包括的に行う (図 3)。この先進的な情報統合の試行は、理研が基盤システムとして理研サイネスを用いて主導する。

### (3) データライセンス付与システム構築

データへの自由なアクセスを奨励するため、データライセンス付与システムを構築する。データライセンスについては、公式に Creative Commons (クリエイティブ・コモンズ)<sup>\*8</sup> のライセンス (パブリックドメイン<sup>\*9</sup> の宣言も含む) の採用を決定した。理研はこのシステムの構築・運用を担当し、必要な情報技術の開発を行う。なお、すでに理研サイネスでは、データを公開したい各ユーザーが好きなライセンスタイプを選択して公開している。

このような国際連携をリードするためには、優れた情報基盤をいち早く構築できるかどうか鍵になります。理研は、これまでライフサイエンス分野で培ってきた情報基盤をもとに、理研サイネスと、理研のスーパーコンピュータをクラウドコンピューティング技術で接続し、理研 BRC で生産している大規模な表現型データを効率的に取扱う情報基盤を整備し、世界をリードします (図 4)。

## 3. 今後の期待

マウス表現型データを大量に所有する世界トップ機関による合意を経て、これらの機関が具体的な取り組みが開始したことから、マウス表現型データの世界標準が確立するとともに、マウスを利用する世界中の研究者が、より正確で容易に研究を行うことができる環境が整います。特に、マウスの利用が不可欠な創薬や治療法開発などで、大きく貢献するものと期待できます。

<報道担当・問い合わせ先>

(問い合わせ先)

独立行政法人理化学研究所

生命情報基盤研究部門 部門長

豊田 哲郎 (とよだ てつろう)

TEL : 045-503-9610 FAX : 045-503-9553

バイオリソースセンター

マウス表現型知識化研究開発ユニット

ユニットリーダー

榎屋 啓志 (ますや ひろし)

TEL : 029-836-9013 FAX : 029-836-9017

横浜研究推進部 企画課

TEL : 045-503-9117 FAX : 045-503-9113

筑波研究推進部 企画課

TEL : 029-836-9136 FAX : 029-836-9100

(報道担当)

独立行政法人理化学研究所 広報室 報道担当

TEL : 048-467-9272 FAX : 048-462-4715

<補足説明>

※1 表現型(ひょうげんけい・ひょうげんがた)

個体の持つ遺伝子の変異や、ゲノムの違いなどの結果として、個体差となって現れる形質のこと。遺伝子にある変異から生じる個体の差は、遺伝子発現の違い、タンパク質の違い、代謝産物の違いという段階を経て到達する。疾患モデル動物であるマウスでは、病的な症状や行動特性の違いなどで現れるが、ほかの段階と違い形質の定量化が難しいため、研究室間での比較やデータベースの統一化が難しく、課題となっていた。

※2 理研サイネス

理研サイネス (RIKEN SciNeS : RIKEN Life Science Networking System) は、クラウドコンピューティングによる大規模なデータベースの構築基盤システムである。理研サイネスを利用することにより、個々の研究者は自らウェブサーバーを維持することなく、各自のデータベースを研究成果物としてスムーズに外部へ向け発信・発表することができる。このため、日本の研究者が国際連携研究を進めるための情報基盤として役立つ。

参考資料 <http://omicspace.riken.jp/publications/toyoda2009ja.pdf>

### ※3 インターフェノーム (InterPhenome)

2006 年発足。マウス研究者が容易にマウス表現型データにアクセスできることを目指して、世界中のマウスセンターにより結成された国際コンソーシアム。

### ※4 セマンティックウェブ、RDF/OWL 技術

セマンティックウェブは、従来の World Wide Web (WWW)のウェブページの閲覧という行為に加え、コンピュータが意味を理解できるようにすることで、自動的な情報収集を可能にするための標準化された技術を指す。RDF (Resource Description Framework) は、セマンティックウェブの中核技術であり、相互運用可能な形でデータ共有や再利用の枠組みを提供する世界標準のデータ形式。主語、述語、目的語の 3 者関係によって、関係の連鎖をたどることができるようなデータモデルを構築する。データ共有のための標準形式として、アプリケーション、企業、コミュニティといった境界を越えて利用できる。OWL (Web Ontology Language) は、「用語・語彙とそこに含まれる各要素の関連の明確な表現」を目的として RDF を拡張し定義されたデータ形式である。

### ※5 クラウドコンピューティング

多数の計算サーバーやデータサーバー群の固まりを集合的に利用することで大規模な情報処理ニーズを迅速かつ効率的に満たす技術。理研サイネスではデータ保存の分散化と、それらの定期的な自動編さん処理の分散化を大規模に行っている。

### ※6 バイオインフォマティクス

応用数学、情報学、統計学、計算機科学などの技術応用によって生物学の問題を解こうとする学問。「生命情報学」「生物情報学」などと訳される。近年、多くの生物を対象に実施されているゲノムプロジェクトや構造ゲノムプロジェクトによって、大量のバイオ関連情報が得られるようになり、それらの情報をタンパク質の系統解析、構造予測、相互作用予測など有用なバイオインフォマティクス技術につなげることが求められている。

### ※7 オントロジー

WWW の驚異的な成長を可能にしている構成原理を解明する「ウェブサイエンス」において、「分類体系」や「推論ルール集」の意味で用いられている。対象とする世界に存在するものごとを体系的に分類し、その関係を記述するものとして、言語学や人工知能研究でも用いられ、セマンティックウェブを誰もが利用可能なものにするための要となっている。MA, MP, PATO は、生物学分野で国際的に提案されているオントロジーであり、YAMATO は大阪大学で開発されている国産技術としての包括的オントロジーである。

### ※8 Creative Commons (クリエイティブ・コモンズ)

創造的な作品に柔軟な著作権を定義するライセンスシステム。ウェブなどのコンテンツに対して、著作権を保持しながら一定の自由を事前に許諾していることを分かりやすく表示することで、より自由な著作権ルールを実現し、より豊かな情報流通と文化・科学技術の発展を目指している。

## ※9 パブリックドメイン

著作権保護期間切れや著作者の権利放棄などにより、知的財産権が誰にも帰属しない状態。日本では、権利を消滅させて公有することを指す場合もある。

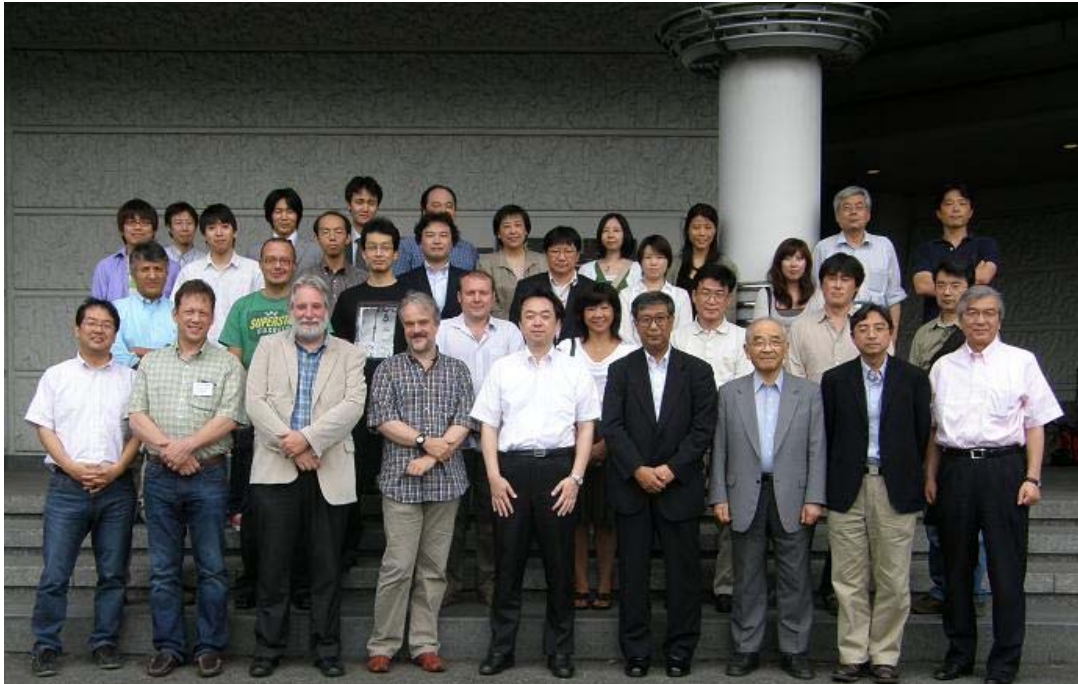


図1 国際マウス表現型情報統合シンポジウム コアメンバー  
(2009年7月13日 ホテルラフォーレ琵琶湖にて)

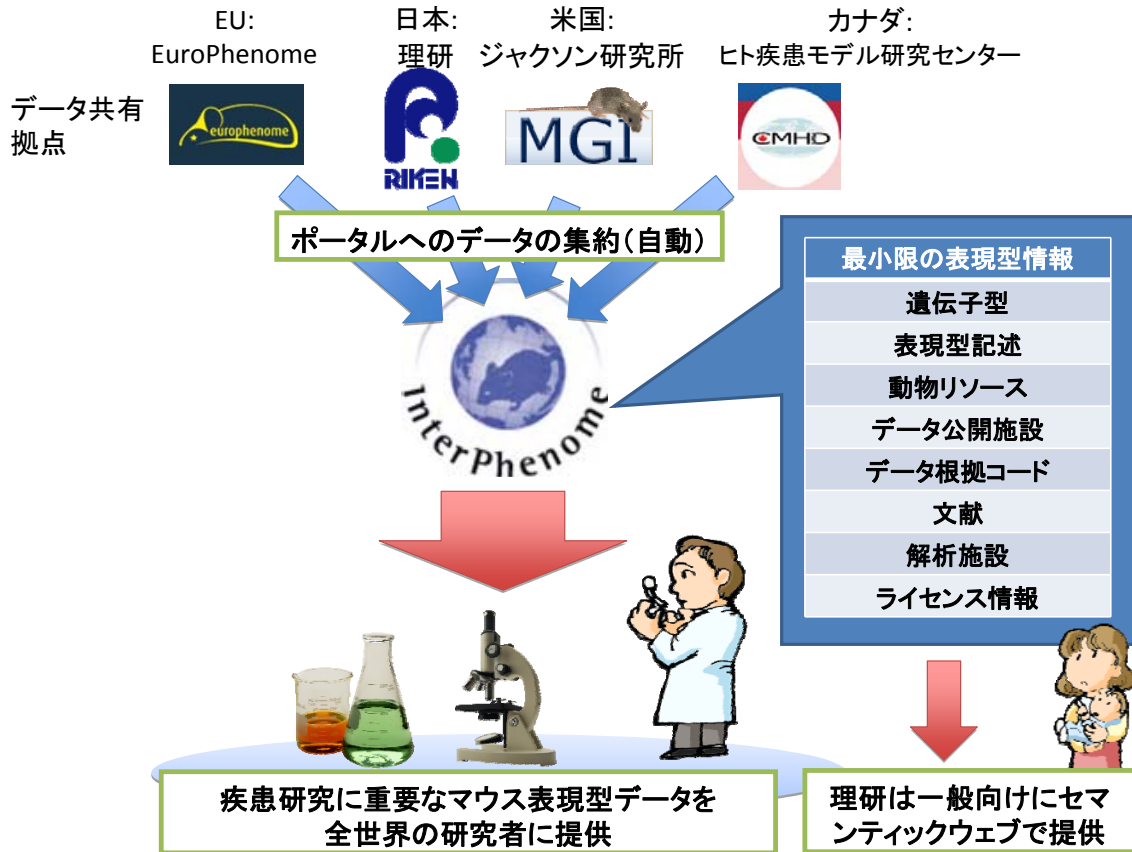


図2 国際マウス表現型ポータルサイトの概要

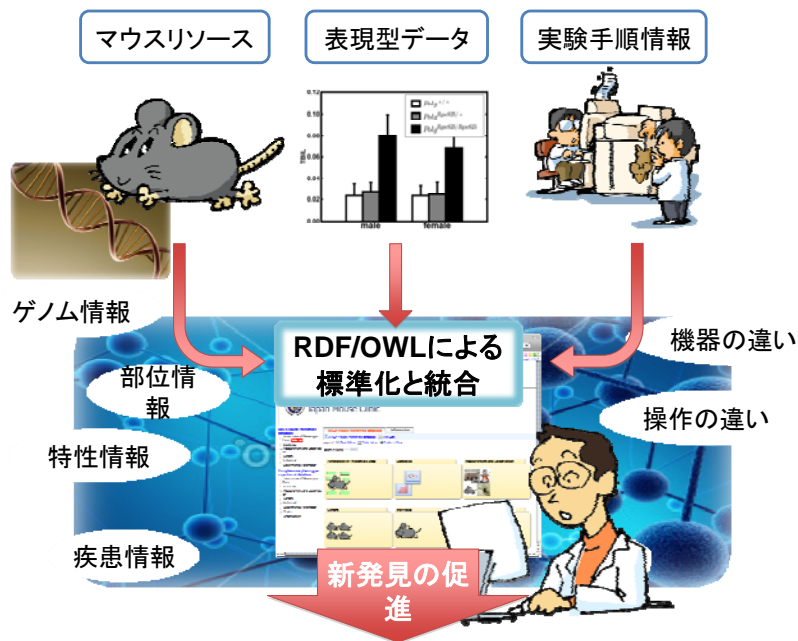


図3 より高度な表現型データ統合の概要



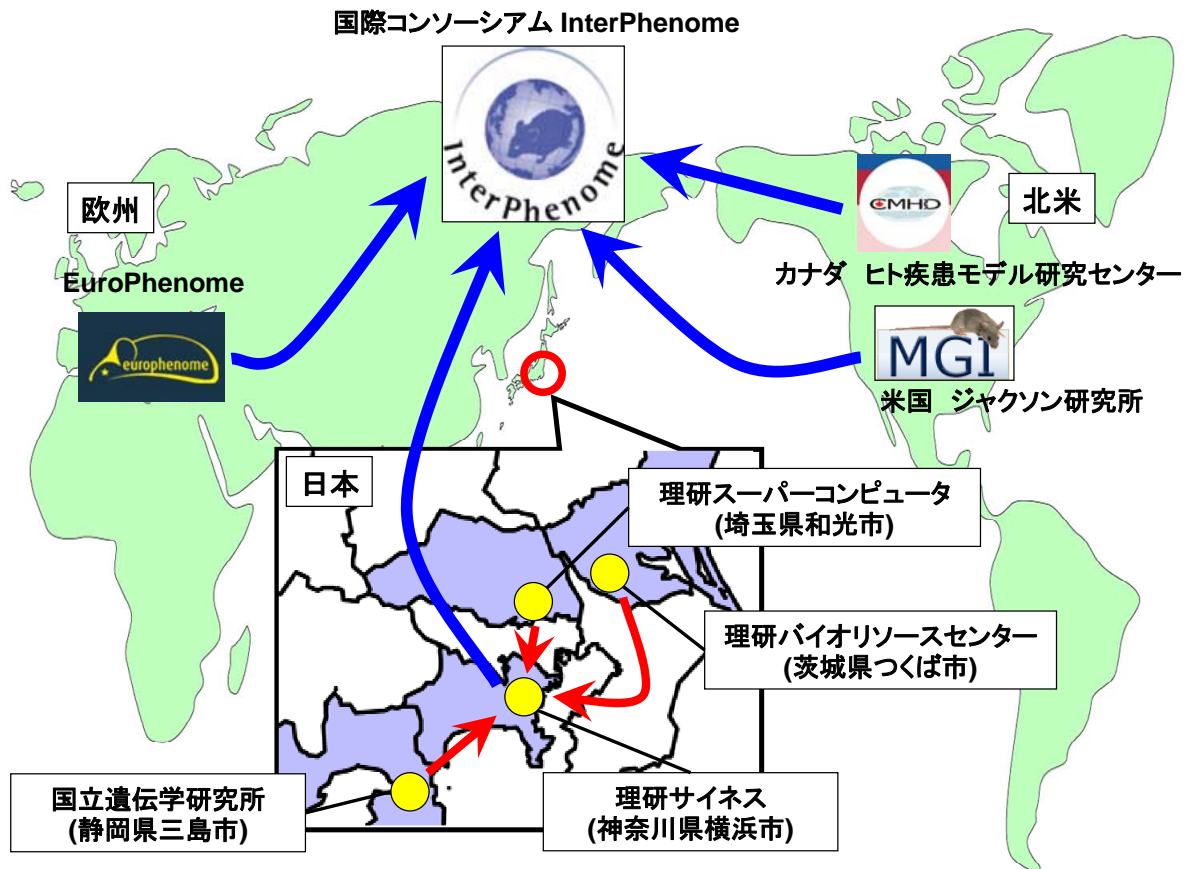


図 4 国際連携の体制