

# 平成 27 年度 研究開発実施報告書

ライフサイエンスデータベース統合推進事業（統合化推進プログラム）

研究代表者

榎屋 啓志

理化学研究所 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

「生命と環境のフェノーム統合データベース」



©2016 榎屋 啓志(理化学研究所) licensed under CC表示2.1 日本

## §1. 研究実施体制

### (1) 代表者グループ

- ① 研究代表者：榎屋 啓志（理化学研究所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）
- ② 研究項目
  - ・ 【提供データ拡充】提供するフェノーム情報の質と量の拡大
  - ・ 【データ収集技術】生物学者向けの表現型データ入力システム開発
  - ・ 【データ活用技術】モデル生物表現型と疾患との関連づけと可視化

## §2. 研究実施内容

### (1) 通期及び本年度の研究開発計画と達成目標

#### **【提供データ拡充】提供するフェノーム情報の質と量の拡大**

＜通期目標＞

各コミュニティの表現型データベースから「分子表現型」600 系統、「単純表現型」1600 系統、をリンク作成用参考データとしての「複合表現型」17000 疾患のデータを収集し、NBDC の各プロジェクトと連携できるような RDF とした上で、標準化・統合化・体系化して公開する。

＜本年度達成目標＞

1) データ導入ワークフロー、オントロジーマッピングデータの作成:

・メダカ表現型の新規アノテーション、・ゼブラフィッシュジーントラップのゲノム挿入部位、・国際マウス表現型解析コンソーシアム (International Mouse Phenotyping Consortium: IMPC) の網羅的表現型データ

2) データ RDF 化と公開: H26 年度に収集した下記のデータを RDF として公開する

・マウス Cre ドライバー系統、・ゼブラフィッシュの Gene Trap 系統遺伝子発現情報、・マウス網羅的表現型解析データ、・ラット表現型解析データ、・メダカ表現型、・第 1 期マウス系統表現型データ、・第 1 期細胞株表現型データ、・第 1 期微生物株データ

#### **【データ収集技術】生物学者向けの表現型データ入力システム開発**

＜通期目標＞

一般研究者から直接表現型データ登録を受け付けるシステムとして、今後爆発的に増加すると見込まれる「ゲノム編集技術」による多様な生物の変異体の表現型を主な収集するシステムを開発する。

＜本年度達成目標＞プロトタイプを作成

#### **【データ活用技術】モデル生物表現型と疾患との関連づけと可視化**

＜通期目標＞

器官組織、形質、遺伝子の相同性を通じて、ヒト疾患（複合表現型）と、モデル生物の表現型（単純表現型の組み合わせ）の生物学的対応が、より高い信頼性をもって評価できるような関連化および可視化を行う。

＜本年度達成目標＞プロトタイプを作成

### (2) 本年度の研究開発進捗状況

#### **【提供データ拡充】提供するフェノーム情報の質と量の拡大**

平成 27 年度は下記のデータの RDF 化および公開を行った。RDF データ公開の基盤には、理研メタデータベース (<http://metadb.riken.jp/metadb/front>) を用いた。本データベースを用いることで、RDF ファイル公開、SPARQL エンドポイントの公開、Web によるデータ閲覧が可能となった。当初の計画になかった国際コンソーシアムのマウス網羅的表現型データの提供を受けたことで、公開データは、表現型データとして 120 万を超え、「表現型ビッグデータ」と呼べる規模となっている。一方で、データ提供者との協議調整等で一部のデータが未公開となっている。

<公開済み>

カテゴリ	データベース	URL	表現型データ ポイント数	総トリプル数
単純表現型	NBRP メダカ表現型メタデータ	<a href="http://metadb.riken.jp/metadb/db/NBRP_medaka">http://metadb.riken.jp/metadb/db/NBRP_medaka</a>	581	12,268
単純表現型	JCGGDB: ノックアウトマウスを用いた機能糖鎖科学データベースのメタデータ	<a href="http://metadb.riken.jp/metadb/db/Glycomics_mouse">http://metadb.riken.jp/metadb/db/Glycomics_mouse</a>	91	1,462
単純表現型	国際マウス表現型解析コンソーシアム(IMPC)の網羅的表現型データ	<a href="http://metadb.riken.jp/metadb/db/IMPC_RDF">http://metadb.riken.jp/metadb/db/IMPC_RDF</a>	1,225,386	53,907,630
第1期データ	BRC マウスリソース表現型メタデータ	<a href="http://metadb.riken.jp/metadb/db/rikenbrc_mouse">http://metadb.riken.jp/metadb/db/rikenbrc_mouse</a>	13,066	518,061
第1期データ	BRC 細胞リソースメタデータ	<a href="http://metadb.riken.jp/metadb/db/rikenbrc_cell">http://metadb.riken.jp/metadb/db/rikenbrc_cell</a>	6,170	195,055
第1期データ	JCM 微生物リソースメタデータ	<a href="http://metadb.riken.jp/metadb/db/rikenbrc_jcm_microbe">http://metadb.riken.jp/metadb/db/rikenbrc_jcm_microbe</a>	-	483,057
小計			1,245,294	55,117,533

<未公開>

カテゴリ	データベース	ステータス	表現型データ ポイント数	総トリプル数
単純表現型	NBRP ラット表現型メタデータ	データ公開のライセンスに関して最終調整中(2016年5月に調整終了)	25,965	289,606
単純表現型	遺伝研 マウス表現型データベースメタデータ	スキーマの最終調整中(2016年5月に調整終了)	2,130	23,936
分子表現型	遺伝研 ゼブラフィッシュ表現型データ	データスキーマに関して遺伝研と協議中	1,383	39,717
分子表現型	マウス Cre ドライバー系統マーカー遺伝子の発現情報	データスキーマに関して熊本大学と協議中	3,279	43,680
複合表現型	厚生労働省指定難病と難病オントロジーORDO との関連性データ(rear-disease vocabulary in Japanese)	訳語の調整中	684	2,376
小計(見込み)			31,311	375,379

**【データ収集技術】生物学者向けの表現型データ入力システム開発**

プロトタイプの開発を行った。予め用意しておいた語彙(オントロジー)を画面遷移に沿って選んで行くことで、対象のゲノム改変生物、およびその表現型について、本プロジェクトで設計したデータスキーマに準拠し、かつ適切なオントロジーでアノテーションされた RDF が生成される。昨年度計画では、早期に開発を進行させ、プロトタイプをゲノム編集コンソーシアムに限定公開する予定であったが、プロトタイプ開発の開発終了にとどまった。平成 28 年度の開発で評価と拡張を行い、最終バージョンとする予定である。

**【データ活用技術】モデル生物表現型と疾患との関連づけと可視化**

平成 27 年度は、各種データの連結の検証作業と、Web アプリケーションのプロトタイプ開発を行

った。本アプリケーションでは、UberPheno オントロジーのリンクを通じて、疾患語彙と関連する表現型を示すモデル生物(マウス、ラット、メダカ等)を検索することができる。下記の機能を備える。

1. ヒト疾患名を記入して、疾患モデル動物の候補を検索する機能

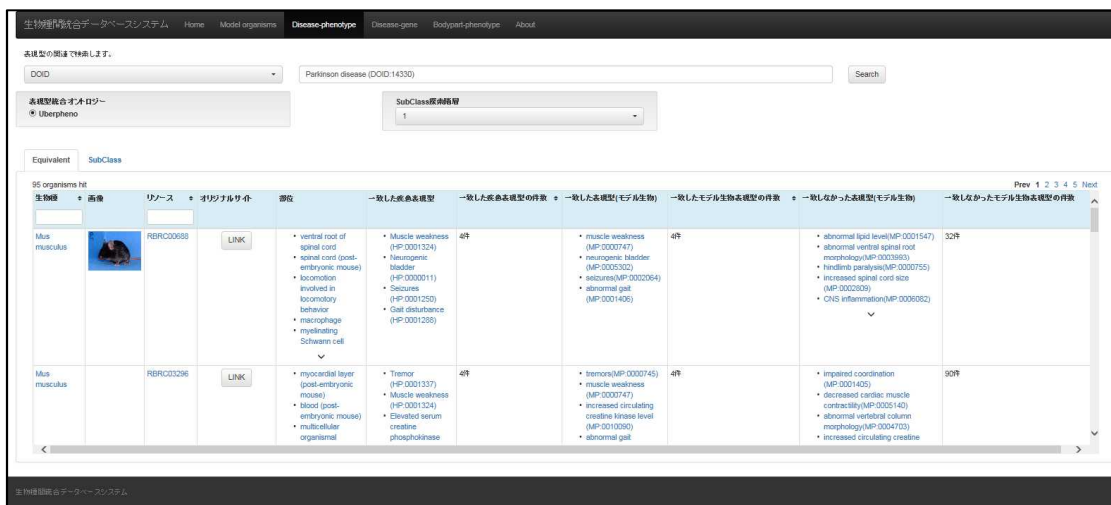


図 1 『パーキンソン病』の疾患モデル動物候補となる系統一ノックアウトマウスを検索した例

2. 遺伝子オルソログデータを元にヒト疾患研究のモデル動物の候補を検索する機能
3. 部位名から、関係するヒト疾患研究のモデル動物の候補を検索する機能

### §3. 成果発表等

#### (3-1) 原著論文発表

- ① 発行済論文数 (国内 (和文) 0 件、国際 (欧文) 3 件)
- ② 未発行論文数 (国内 (和文) 0 件、国際 (欧文) 0 件)
- ③ 論文詳細情報

1. Karp N A, Meehan T F, Morgan H, Mason J C, Blake A, Kurbatova N, Smedley D, Jacobsen J, Mott R F, Iyer V, Matthews P, Melvin D G, Wells S, Flenniken A M, Masuya H, Wakana S, White J K, Lloyd K C K, Reynolds C L, Paylor R, West D B, Svenson K L, Chesler E J, de Angelis M H, Tocchini-Valentini G P, Sorg T, Herault Y, Parkinson H, Mallon A M, Brown S D M, “Applying the ARRIVE Guidelines to an In Vivo Database” PLOS Biology, 13, e1002151, 2015 (DOI:10.1371/journal.pbio.1002151)
2. Maier H, Schütt C, Steinkamp R, Hurt A, Schneltzer E, Gormanns P, Lengger C, Griffiths M, Melvin D, Agrawal N, Alcantara R, Evans A, Gannon D, Holroyd S, Kipp C, Raj NP, Richardson D, LeBlanc S, Vasseur L, Masuya H, Kobayashi K, Suzuki T, Tanaka N, Wakana S, Walling A, Clary D, Gallegos J, Fuchs H, de Angelis MH, Gailus-Durner V, Principles and application of LIMS in mouse clinics, Mammalian Genome, 26, 467-481, 2015 (DOI 10.1007/s00335-015-9586-7)
3. Kimura M, Ichimura S, Sasaki K, Masuya H, Suzuki T, Wakana S, Ikegawa S, Furuichi T, Endoplasmic reticulum stress-mediated apoptosis contributes to a skeletal dysplasia resembling platyspondylic lethal skeletal dysplasia, Torrance type, in a novel Col2a1 mutant mouse line, Biochemical and Biophysical Research Communications, 468, 86-91, 2015 (doi:10.1016/j.bbrc.2015.10.160)

#### (3-2) データベースおよびウェブツール等の構築と公開

別紙1参照。

#### (3-3) その他の著作物(総説、書籍など)

該当なし

#### (3-4) 国際学会および国内学会発表

- ① 招待講演 (国内 0 件、国際 2 件)

〈国内〉

該当なし

〈国際〉

1. Masuya H, Informational technologies for data sharing in biology, The Fourth Sino-Japan Short Summer Course, Nanjing, China, July 28
2. Masuya H, Development of an integrated phenotype database of experimental animals, Incheon, Korea, August 12

② 口頭講演 (国内 1 件、国際 1 件)

〈国内〉

1. 榎屋啓志、高山英紀、古崎晃司、今井健、大江和彦、溝口理一郎、生物表現型情報と、疾患情報をつなげるデータベース、2015 年度 人工知能学会全国大会、函館、5 月 30 日

〈国際〉

1. Masuya H, Takatsuki T, Saito M, Mori H, Kawashima S, Data coordination of bio-resources related information using resource description framework technology, The 7th ANRRC International Meeting, Incheon, Korea, September 18

③ ポスター発表 (国内 9 件、国際 4 件)

〈国内〉

1. 高月照江、矢田有加里、齋藤実香子、高田豊行、真下知士、成瀬清、若菜茂晴、田中信彦、榎屋啓志、生物種を超えた表現型情報統合のためのデータ作成、第 62 回日本実験動物学会総会、京都、5 月 28 日
2. 廣田和之、小澤恵代、榎屋啓志、田村勝、若菜茂晴、日本マウスクリニックにおける骨形態評価の国際標準化への取り組み、第 62 回日本実験動物学会総会、京都テルサ、5 月 28 日
3. 三浦郁生、篠木晶子、臼田大輝、鈴木智広、金田秀貴、古瀬民生、小林喜美男、山田郁子、田村勝、榎屋啓志、若菜茂晴、C57BL/6 亜系統間を検出可能な SNPs マーカーセットの構築、第 62 回日本実験動物学会総会、京都テルサ、5 月 29 日
4. 榎屋啓志、セマンティック Web 技術を用いたバイオリソースデータの標準化と整備、第 1 回理化学研究所・産業技術総合研究所 共同シンポジウム、産業技術総合研究所 臨海副都心センター、6 月 29 日
5. 榎屋啓志、高月照江、齋藤実香子、高山英紀、大島和也、大城望、田中信彦、モデル生物の表現型統合データベースの作成に向けて、第 29 回モロシヌス研究会、かんぼの宿有馬、7 月 3 日
6. 高月照江、齋藤実香子、大島和也、高山英紀、金子裕代、成瀬清、若菜茂晴、田中信彦、榎屋啓志、表現型情報統合のためのデータ作成、トーゴーの日シンポジウム 2015、東京大学弥生講堂、10 月 6 日
7. 高月照江、齋藤実香子、大島和也、高山英紀、金子裕代、成瀬清、若菜茂晴、田中信彦、榎屋啓志、表現型情報統合データベースのためのデータ作成、BMB2015、神戸ポートアイランド、12 月 3 日
8. 田中信彦、土岐秀明、鈴木智広、金田秀貴、三浦郁生、山田郁子、古瀬民生、小林喜美男、井上麻紀、美野輪治、若菜茂晴、榎屋啓志、表現型の異常を高感度で検出する統計解析ワークフローの開発、BMB2015、神戸ポートアイランド、12 月 3 日
9. 榎屋啓志、小林紀郎、生物表現型ビッグデータの形成に向けて:セマンティック Web 技術を用いた表現型データの統合、第 2 回理化学研究所・産業技術総合研究所 共同シ

ンポジウム, 産業技術総合研究所つくばセンター, 2月2日

〈国際〉

1. Masuya H, Takatsuki T, Saito M, Lents K, Takayama E, Ohshima K, Tanaka N, Naruse K, Takada T, Kuramoto T, Wakana S, Kobayashi N, Development of Semantic Web/RDF based integrated database of experimental animals, 29th International Mammalian Genome Conference, Yokohama, Japan, November 9
2. Tanaka N, Masuya H, Exploring of novel mouse models for human disease with comprehensive mouse phenotyping data, 29th International Mammalian Genome Conference, Yokohama, Japan, November 9
3. Masuya H, Takatsuki T, Saito M, Takayama E, Ohshima K, Ohshiro N, Lenz K, Tanaka N, Mori H, Kawashima S, Kobayashi N, RDF-based data sharing of bio-resource related information, SWAT4LS International Conference, Cambridge, England, December 8
4. Kobayashi N, Lenz K, Masuya H, RIKEN Meta Database: a life-science metadata database based on the Semantic Web, SWAT4LS International Conference, Cambridge, England, December 8

### (3-5) 知的財産権の出願

該当なし

### (3-6) 受賞・報道等

該当なし

## §4. 研究開発期間中に主催した活動(主催したワークショップ等)

年月日	名称	場所	参加人数	概要
2015年4月17日	チーム内ミーティング(非公開)	理研筑波研究所 研究棟 121 室	6人	研究進捗報告のためのミーティング
2015年4月8日	研究打ち合わせ(非公開)	理研情報基盤センター	3人	理研内 RDF データ基盤との擦り合わせ
2015年4月20日	表現型データベース勉強会(非公開)	理研筑波研究所 研究棟 121 室	5人	Rat Genome Database (RGD)
2015年4月21日	研究打ち合わせ(非公開)	理研情報基盤センター	3人	理研内 RDF データ基盤との擦り合わせ
2015年4月27日	チーム内ミーティング(非公開)	理研筑波研究所 研究棟 121 室	6人	研究進捗報告のためのミーティング
2015年5月11日	研究打ち合わせ(非公開)	Jackson 研究所 (米国)	5人	MGI データの RDF 化に関する打ち合わせ
2015年5月18日	表現型データベース勉強会(非公開)	理研筑波研究所 研究棟 121 室	5人	IMPC データの統計解析方法について
2015年5月19日	研究打ち合わせ(非公開)	広島大学	2人	ゲノム編集データの RDF 化に関する打ち合わせ



2015年5月20日	チーム内ミーティング(非公開)	理研筑波研究所 研究棟 121 室	5 人	研究進捗報告のための ミーティング
2015年6月3日	チーム内ミーティング(非公開)	理研筑波研究所 研究棟 121 室	7 人	研究進捗報告のための ミーティング
2015年6月8日	表現型データベース勉強会(非公開)	理研筑波研究所 研究棟 121 室	5 人	Web Ontology Language (OWL)
2015年6月30日	表現型データベース勉強会(非公開)	理研筑波研究所 研究棟 121 室	5 人	RDF 作成ツールの利用
2015年7月1日	チーム内ミーティング(非公開)	理研筑波研究所 研究棟 121 室	6 人	研究進捗報告のための ミーティング
2015年7月13日	表現型データベース勉強会(非公開)	理研筑波研究所 研究棟 121 室	5 人	統計解析環境 R
2015年7月17日	研究打ち合わせ(非公開)	理研バイオリソースセンター	4 人	中国科学院との微生物 RDF に関する打ち合わせ
2015年7月21日	チーム内ミーティング(非公開)	理研筑波研究所 研究棟 121 室	6 人	研究進捗報告のための ミーティング
2015年7月22日	研究打ち合わせ(非公開)	東京工業大学	4 人	中国科学院、東工大、 DBCLS との微生物 RDF に関する打ち合わせ
2015年8月5日	チーム内ミーティング(非公開)	理研筑波研究所 研究棟 121 室	6 人	研究進捗報告のための ミーティング
2015年8月11日	研究打ち合わせ(非公開)	Seoul National University (韓国)	4 人	Korean Mouse Phenotyping Consortium との連携 に関する打ち合わせ
2015年9月7日	表現型データベース勉強会(非公開)	理研筑波研究所 研究棟 121 室	5 人	SQL の初歩
2015年9月9日	チーム内ミーティング(非公開)	理研筑波研究所 研究棟 121 室	7 人	研究進捗報告のための ミーティング
2015年9月11日	ANRRC の IT に関する打ち合わせ	Incheon (韓国)	6 人	アジアにおけるバイオリ ソース情報の標準化に 関する意見交換
2015年9月24日	研究打ち合わせ(非公開)	東北大学医学部	5 人	マウス表現型及び、疾 患語彙の共通利用に 関する意見交換
2015年9月28日	表現型データベース勉強会(非公開)	理研筑波研究所 研究棟 121 室	5 人	JCM のデータの詳細 解説
2015年9月1日	研究打ち合わせ(非公開)	理研情報基盤センター	3 人	理研内 RDF データ基 盤との打ち合わせ
2015年10月7日	チーム内ミーティング(非公開)	理研筑波研究所 研究棟 121 室	6 人	研究進捗報告のための ミーティング
2015年10月16日	研究打ち合わせ(非公開)	理研横浜研究所	3 人	理研内 RDF データ基 盤を用いた情報統合に 関する打ち合わせ
2015年10月19日	表現型データベース勉強会(非公開)	理研筑波研究所 研究棟 121 室	5 人	IMPC ウェブサイトのデ ータの詳細解説
2015年10月22日	理研データベース基盤に関する意見	理研情報基盤センター	6 人	理研内データベース基 盤における RDF の利

	交換			点に関する意見交換
2015年11月4日	チーム内ミーティング(非公開)	理研筑波研究所 研究棟 121 室	6 人	研究進捗報告のためのミーティング
2015年11月6日	研究打ち合わせ(非公開)	慶応大学	3 人	AMED-IRUD との連携に関する意見交換
2015年12月11日	IMPC データ連携 打ち合わせ	European Bioinformatics Institute (EBI)	5 人	IMPC との RDF 連携に関する打ち合わせ
2016年1月8日	SBI とのデータ連携 打ち合わせ	システム・バイオ ロジー研究機構 (東京)	4 人	RDF を用いた、データ連携及び利用に関する意見交換、打ち合わせ
2016年1月13日	チーム内ミーティング(非公開)	理研筑波研究所 研究棟 121 室	6 人	研究進捗報告のためのミーティング
2015年1月25日	表現型データベース勉強会(非公開)	理研筑波研究所 研究棟 121 室	4 人	非生物学者のための遺伝学解説
2016年2月9日	チーム内ミーティング(非公開)	理研筑波研究所 研究棟 121 室	6 人	研究進捗報告のためのミーティング
2015年2月22日	表現型データベース勉強会(非公開)	理研筑波研究所 研究棟 121 室	4 人	MySQL の初歩
2015年3月1日	細胞データベース 打ち合わせ	慶応大学	8 人	培養細胞のデータ共有に関する意見交換
2015年3月7日	表現型データベース勉強会(非公開)	理研筑波研究所 研究棟 121 室	5 人	統計解析の初歩
2016年3月22日	チーム内ミーティング(非公開)	理研筑波研究所 研究棟 121 室	5 人	研究進捗報告のためのミーティング

以上

別紙1 既公開のデータベース・ウェブツール等

No.	状態	研究開発課題名	データベース、ウェブツール等の名称	概要 (150字程度)	URL	公開開始日	関連論文 (論文リストに記載があれば、その番号でも可)
1	新規	生命と環境のフェノーム統合データベース	BRC マウスリソース表現型メタデータ(ベータ版)	生物遺伝材料としてのマウス系統を収録したデータベースです。系統の持つ遺伝子の変異や表現型が公共データやオントロジーにリンクされています。ここに登録されているマウス系統は、理研バイオリソースセンターより提供されています。	<a href="http://metadb.riken.jp/db/rikenbrc_mouse">http://metadb.riken.jp/db/rikenbrc_mouse</a>	2015年6月1日	
2	新規	生命と環境のフェノーム統合データベース	BRC細胞リソース特性メタデータ(ベータ版)	本DBは、生物遺伝材料としての培養細胞株を収録しています。ヒトやマウス等の哺乳類をはじめ様々な生物種にわたる、多種の細胞株を公開しています。ここに登録されている細胞株は、理研バイオリソースセンターより提供されています。	<a href="http://metadb.riken.jp/db/rikenbrc_cell">http://metadb.riken.jp/db/rikenbrc_cell</a>	2015年6月1日	
3	新規	生命と環境のフェノーム統合データベース	JCM微生物リソース特性メタデータ(ベータ版)	本DBは、生物遺伝材料としての微生物株を収録しています。様々な種の微生物について、株としての情報を特性や表現型情報を含めて提供しています。ここに登録されている微生物株は、理研バイオリソースセンターより提供されています。	<a href="http://metadb.riken.jp/db/rikenbrc_jcm_microbe">http://metadb.riken.jp/db/rikenbrc_jcm_microbe</a>	2015年6月1日	
4	新規	生命と環境のフェノーム統合データベース	シロイヌナズナフェノームデータベース	文献キュレーションによって収集したシロイヌナズナフェノーム情報および理研内で開発されたシロイヌナズナのバイリソース(変異体)のフェノームを統合しました。	<a href="http://jphenome.info/?page_id=95">http://jphenome.info/?page_id=95</a>	2015年6月1日	
5	新規	生命と環境のフェノーム統合データベース	J-phenome	本サイトは、遺伝子の多様性の結果として現れる生命の表現型情報を、モデル動物(マウス、ラット、ゼブラフィッシュ、メダカ)、ゲノム編集研究など、幅広い研究コミュニティから収集し、研究分野の垣根を超えて標準化・統合化・体系化して公開するデータベースへのリンクするポータルサイトです。	<a href="http://jphenome.info">http://jphenome.info</a>	2015年5月15日	
6	新規	生命と環境のフェノーム統合データベース	NBRPメダカ表現型メタデータ	生物学、医学、食糧、環境に関する研究および技術開発に用いられるメダカ系統の持つ遺伝子の変異や表現型が公共データやオントロジーにリンクされて整理されています。ここに登録されているマウス系統は、NBRPプロジェクト(基礎生物学研究所)より提供されています。	<a href="http://metadb.riken.jp/metadb/db/NBRP_medaka">http://metadb.riken.jp/metadb/db/NBRP_medaka</a>	2015年7月16日	
7	新規	生命と環境のフェノーム統合データベース	JCGGDB: ノックアウトマウスを用いた機能糖鎖科学データベースのメタデータ	本データは、日本糖鎖科学統合データベース(JCGGDB)の一部である「ノックアウトマウスを用いた機能糖鎖科学データベース」のRDFメタデータ(オントロジーや公共データでアノテーションした情報共有のためのデータ)です。	<a href="http://metadb.riken.jp/metadb/db/Glycomics_mouse">http://metadb.riken.jp/metadb/db/Glycomics_mouse</a>	2015年10月1日	
8	新規	生命と環境のフェノーム統合データベース	IMPC RDFデータ	国際マウス表現型解析コンソーシアム(International Mouse Phenotyping Consortium: IMPC)の網羅的表現型データ	<a href="http://metadb.riken.jp/metadb/db/IMPC_RDF">http://metadb.riken.jp/metadb/db/IMPC_RDF</a>	2015年10月14日	