

平成 26 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

[榊屋 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]

§1. 研究実施体制

平成 26 年度の本研究課題は、下記の代表者グループのみで遂行した。

(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. 【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

1.1. 新たな表現型データの収集: 第1期(H23~25年度)における、マウス系統、細胞株、微生物株の表現型情報に加え、新たに各コミュニティの表現型データベースから「分子表現型」、「単純表現型」、「複合表現型」の階層に分類して収集し、標準化・統合化・体系化して公開する事を目的として、平成26年度は、各データベースとの連携の大枠の確立、データインポートのワークフローの確立と、RDF化に必要な各オントロジーの選定や、基盤として必要なマッピングデータの作成を下記のように行った。

■ 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:(目標:約100系統)
約132系統、1716表現型を熊本大学より収集した。
- ゼブラフィッシュ遺伝子トラップ系統マーカー遺伝子の発現情報:(目標:約500系統)
約330系統、1400表現型のデータを国立遺伝学研究所より収集した。
研究コミュニティに資するRDFスキーマについて、データ提供元と協議中。

■ 単純表現型:

- マウス網羅的表現型解析データ:(目標:約100系統)
約100系統、5000表現型を、理研バイオリソースセンター・マウスクリニックおよび、国立遺伝学研究所・コンソミック系統表現型データより収集した。
- ラット網羅的表現型解析データ:(目標:約1000系統)
175系統の網羅的表現型データを、京都大学・NBRPラットデータベースより収集した。
- メダカ表現型:(目標:約約500系統)
約300系統、430表現型を、基礎生物学研究所・NBRPメダカ表現型データベースおよび、基礎生物学研究所において論文より新たにキュレーションした表現型データとして収集した。

■ 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:約400疾患。厚生労働省の指定する130難病に含まれる400疾患について、国際的に用いられている難病オントロジーORDOとの関連性を整理した。

1.2. 既存の表現型データの改良および公開基盤の変更: 第1期に収集したデータに関して、より世界での流通、統合化推進プロジェクト間での連携に適したRDFへの変更を、下記のように行った。

- 第1期の微生物株データ約15000株について、微生物系統のRDFスキーマであるMicrobial Culture Collection Vocabularyに準拠したRDFデータとして修正した。
- 第1期の細胞株表現型データ3800株について、海外の提供機関とともに国際国際コンソーシアムとして、世界標準の候補となるRDFスキーマ:Cell Line Ontology

(CLO)を提案し、論文発表を行った(論文 1)。

- 生命データの RDF 共有において、遺伝子の知識モデルが確立されていない点が大きな問題である。この問題を将来的に解決するための先進的なオントロジー「遺伝学オントロジー」を提案すべく、論文発表を行った(論文 2)。また、表現型データを公開する際のメタデータ記載に関する国際的な調整に参加した(論文 3)。

2. 【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

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を、主に「ゲノム編集技術」による多様な生物の変異体の表現型を対象として開発することを目的として、平成 26 年度は仕様の策定を行った。このシステムでは、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、1)で公開されるフェノーム情報と統合されるよう設計されている。

3. 【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

データ利用に直結するような表現型データの可視化の開発として、異常の起こる器官組織や、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)の生物学的対応を、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方法を開発、およびこの推論アルゴリズムを用いて、疾患研究者が、大量のデータに基づいて、実験解析に最適なモデル系を探ることができるようなアプリケーション開発を目的として、平成 26 年度は、表現型センターやゲノム編集コンソーシアムからも利用できる Web アプリケーションの仕様の策定を行った。このアプリケーションは、RDF レポジトリをバックエンドとして、表現型の現れた部位や、表現形質による検索が行なえる他、近年公開された、UberPheno (<http://f1000research.com/articles/2-30/v1>), PhenoDigm (<http://www.sanger.ac.uk/resources/databases/phenodigm/>)等の生物種横断的な表現型同士の関連性データを活用し、生物種を超えて、表現型同士の相同性も推論、また疾患との関連性等も示すことができるように設計を行った。

§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) 1件、国際(欧文) 1件):
- ② 未発行論文数(“accepted”、“in press”等)(国内(和文) 0件、国際(欧文)1件)
- ③ 論文詳細情報
 1. Sarntivijai S, Lin Y, Xiang Z, Meehan TF, Diehl AD, Vempati UD, Schürer SC, Pang C, Malone J, Parkinson H, Liu Y, Takatsuki T, Saijo K, **Masuya H**, Nakamura Y, Brush MH, Haendel MA, Zheng J, Stoeckert CJ, Peters B, Mungall CJ, Carey TE, States DJ, Athey BD, He Y. “CLO: The cell line ontology”, J Biomed Semantics. vol. 5 No. 37. 2014 (DOI: 10.1186/2041-1480-5-37)
 2. 榎屋啓志, 溝口理一郎 “遺伝学オントロジー”, 人工知能学会論文誌 vol. 29, No. 3, pp. 311-327, 2014 (DOI: <http://doi.org/10.1527/tjsai.29.311>)
 3. Karp NA, Meehan TF, Morgan H, Mason JC, Blake A, Kurbatova N, Smedley D, Jacobsen J, Mott RF, Iyer V, Matthews P, Melvin DG, Wells S, Flenniken AM, **Masuya H**, Wakana S, White JK, Lloyd KCK., Reynolds CL, Paylor R, West DB, Svenson KL, Chesler EJ, Hrabe de Angelis MH, Tocchini-Valentini GP, Sorg-Guss T, Herault Y, Parkinson H, Mallon AM, and Brown SD, Applying the ARRIVE guidelines to an in vivo database. PLOS Biology (in press).

(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)

なし。

(3-4) 国際学会および国内学会発表

なし。

(3-5) 国際学会および国内学会発表

- ① 招待講演 (国内 1件、国際 2件)

〈国内〉

1. 榎屋啓志. “モデル生物をより使いやすくするための表現型情報整備”、第11回生命資源研究・支援センターシンポジウム、熊本、2015年3月13日

〈国際〉

1. Masuya H. “Bioinformatics -Mouse Phenotyping Databases”, AMMRA-AMPC Student Workshop, Canberra Australia, October 23, 2014
2. Masuya H. “Development of the mouse phenotype database”, The 11th KOGO Winter Symposium 2015, Hongcheon Korea, February 5, 2015

② 口頭講演 (国内 4件、国際 0件)

〈国内〉

1. 榎屋啓志、高月照江、齋藤実香子、高山英紀、吉田有子、蒔田由布子、望月芳樹、土井考爾、小林紀郎、豊田哲郎. “セマンティック Web 技術を用いた、生物表現型統合データベース”、JSAI2014 人口知能学会全国大会(第 28 回)、松山、2014 年 5 月 12 日
2. 榎屋啓志、古崎晃司、高山英紀、高田豊行、真下知士、若菜茂晴、大江和彦、溝口理一郎. “マウス・ラットの表現型と関連するヒト疾患を調べられるデータベース”、日本実験動物科学技術 さっぽろ 2014、札幌、2014 年 5 月 17 日
3. 榎屋啓志、池内達郎、布山喜章. “遺伝学用語集の改訂における課題と問題”、日本遺伝学会第 86 回大会、長浜、2014 年 9 月 20 日
4. 榎屋啓志. “表現型情報で生物をつなぐフェノーム統合データベース”、トーゴーの日シンポジウム 2014、東京、2014 年 10 月 5 日

③ ポスター発表 (国内 4件、国際 1件)

〈国内〉

1. 田中信彦、土岐秀明、茂木浩未、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、野田哲生、若菜茂晴、榎屋啓志. “新規の疾患モデルマウスや生命現象の発見を目的としたデータマイニング:ワークフローの検討”、日本実験動物科学技術 さっぽろ 2014、札幌、2014 年 5 月 15 日
2. 田中信彦. “網羅的マウス表現型データの解析ワークフローの検討”、第 28 回モロシヌス研究会、三島、2014 年 6 月 27 日
3. 高月照江、矢田有加里、齋藤実香子、榎屋啓志. “表現型情報でつなぐフェノーム統合データベース”、トーゴーの日シンポジウム 2014、東京、2014 年 10 月 5 日
4. 田中信彦、土岐秀明、茂木浩未、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. “網羅的マウス表現型データの解析ワークフローの検討”、第 37 回日本分子生物学会年会、横浜、2014 年 11 月 26 日

〈国際〉

1. Takatsuki,T. Yada,Y. Saito,M. Masuya,H. 、 **Toward development of integrated phenotype database across biological species**、Towards precision medicine, Phenotyping human diseases in mice. The 16th Frank & Bobbie Fenner Conference、Canberra、20-21 October 2014

(3-6) 知的財産権の出願

- ① 当該年度の特許出願件数(国内 0件、海外 0件)
- ② 研究開始から当該年度末までの累積特許出願件数(国内 0件、海外 0件)
- ④ その他の知的財産権

なし

(3-7) 受賞・報道等

特になし

§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
2014年4月18, 19日	理化学研究所 バイオリソースセンター 一般公開 ミニ講座「生き物の研究でかつやくするコンピューター」	理研筑波研究所 バイオリソース棟小会議室	30人	一般の方向けに、生物学研究におけるデータベースの重要性を講義。
2014年5月9日	田畑グループ合同ミーティング(非公開)	かずさDNA研究所	7人	動物、植物での RDF 構成の調整のためのミーティング。田畑グループと共催
2014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	研究室内の進捗ミーティング:疾患データ ICD10 の状況確認
2014年6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	研究室内の進捗ミーティング:RDF 概要説明。
2014年7月8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	研究室内の進捗ミーティング:RDF データ不具合の検証
2014年7月11日	EBI マウスデータ統合グループとの打ち合わせ(非公開)	European Bioinformatics Institute (EBI)	10人	EBI グループとの共催。マウス表現型データの RDF 国際共有のための打ち合わせ
2014年7月24日	チーム内ミーティング(非公開)	理研情報基盤センター	3人	理研内 RDF データ基盤を利用した、表現型データベース開発について打ち合わせ
2014年8月19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	研究室内の進捗ミーティング:RDF 変換ツールの動作確認
2014年9月12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
2014年9月30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	研究室内の進捗ミーティング:ラットデータの確認
2014年11月20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
2014年12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	研究室内の進捗ミーティング:微生物 RDF データスキーマ確認
2014年12月18日	ゼブラフィッシュゲ	国立遺伝学研	3人	ゼブラフィッシュゲ

日	ータ研究打ち合わせ(非公開)	究所		タのコミュニティへの貢献について議論
2014年12月19日	マウス毛色データ収集に関するミーティング(非公開)	国立遺伝学研究所	10人	遺伝研に保存されている各国のマウス毛色表現型のデータ化に関する議論
2015年1月13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1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	ゼブラフィッシュデータベース ZFIN の全体像を確認
2015年2月4日	韓国マウス表現型解析グループとの打ち合わせ	Seoul National University	5人	ソウル大学グループと共催:韓国におけるマウス表現型解析解析の状況確認と、データ化についての議論
2015年3月2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	国際マウス表現型解析コンソーシアムデータベースの全体像を確認
2015年3月16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スキーマの詳細について確認
2015年3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	7人	研究室内の進捗ミーティング:メダカデータの表現型アノテーション状況確認

以上

別紙1 既公開のデータベース・ウェブツール等

No.	研究開発課題	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)	備考
01	生命と環境のフェノーム統合データベース	BRG 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。	http://metadb.riken.jp/db/rikenbr c.mouse (URL変更)	2011.5月 2015.6.1(新たなURLによる更新版)	Yoshiki A, Ike F, Mekada K, Kitaura Y, Nakata H, Hiraiwa N, Mochida K, Ijuin M, Kadota M, Murakami A, Ogura A, Abe K, Moriwaki K, Obata Y. "The mouse resources at the RIKEN BioResource center." Exp Anim. 2009 Apr;58(2):85-96. (PMID: 19448331)	
02	生命と環境のフェノーム統合データベース	BRG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。	http://metadb.riken.jp/db/rikenbr c.cell (URL変更)	2011.9月 2015.6.1(新たなURLによる更新版)	(論文リスト番号 1)	
03	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。	http://metadb.riken.jp/db/rikenbr c.jcm_microbe (URL変更)	2013.8月 2015.6.1(新たなURLによる更新版)		
04	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。	http://jphenome.info/arabidopsis (URL変更)	2013.11月 2015.6.1(新たなURLによる更新版)	Nakazawa M, Ichikawa T, Ishikawa A, Kobayashi H, Tsubura Y, Kawashima M, Suzuki K, Muto S, Matsui M "Activation tagging, a novel tool to dissect the functions of a gene family". Plant J 2003; 741-750 (PMID: 12787254) Sakurai T, Satou M, Akiyama K, Iida K, Seki M, Kuromori T, Ito T, Konagaya A, Toyoda T, Shinozaki K. "RARGE: a large-scale database of RIKEN Arabidopsis resources ranging from transcriptome to phenome." Nucleic Acids Research 2005; 33: D647-50 (PMID: 15608280) Kondou Y, Higuchi M, Takahashi S, Sakurai T, Ichikawa T, Kuroda H, Yoshizumi T, Tsumoto Y, Horii Y, Kawashima M et al "Systematic approaches to using the FOX hunting system to identify useful rice genes." Plant J 2009; 883-894 (doi: 10.1111/j.1365-	
05	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。	http://jphenome.info	2015.05.15		