

**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(統合データ解析トライアル)」**  
**平成 26 年度採択課題における事後評価について**

バイオサイエンスデータベースセンター

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(統合データ解析トライアル)」における平成26年度研究開発課題の事後評価を「ライフサイエンスデータベース統合推進事業の実施に関する規則」に基づき実施した。

## 1. 事後評価の目的と基準

### (1) 評価の目的

研究開発の実施状況及び研究開発成果等を明らかにし、今後の研究開発成果の展開及び事業運営の改善に資することを目的とする。

### (2) 評価項目及び基準

- ・当初計画の達成度
- ・開発したツールの成果(開発したツールを活用したデータベースの新たな活用法や、有用な知見を発見したか)
- ・開発したツールの将来性

## 2. 評価対象となった研究開発課題

### (1) 件数と研究開発実施期間

対象課題数: 4 件

実施期間: 平成 26 年 9 月 1 日～平成 27 年 2 月 28 日まで(6ヶ月)

### (2) 研究開発課題名と研究代表者名

研究開発課題名	研究代表者名	所属	役職
PDBj タンパク質をゲノムにマップした pdbBAM の作成	城田 松之	東北大学大学院 医学系研究科	助教
RDF ストア間データ連結フレームワーク の開発およびオーソログ解析への適用	千葉 啓和	基礎生物学研究所 ゲノム情報研究室	研究員
生化学反応ネットワーク統合解析環境の 拡張	西田 孝三	理化学研究所 生命システム研究センター	テクニカル スタッフ
HLA 遺伝子完全配列決定パイプライン の構築	細道 一善	国立遺伝学研究所 人類遺伝研究部門	外来研究員

\*五十音順。所属は平成 27 年 3 月時点

### (3) 評価スケジュール

平成 27 年 3 月	研究代表者(被評価者)による研究開発終了報告書の提出
平成 27 年 3 月	研究開発成果報告会・事後評価会開催 ①被評価者からの報告、評価者・被評価者の意見交換 ②研究アドバイザーに評価に対する意見を聴取
平成 27 年 4 月～5 月	研究総括による評価 ①研究アドバイザーの意見をもとに事後評価結果(案)の作成 ②事後評価結果(案)を被評価者に提示し、意見交換 ③事後評価結果の確定(全評価者の確認・合意)
平成 27 年 6 月	業務運営会議等への報告・JST Web サイトによる公表

### 3. 評価者

#### 研究総括

長洲 毅志	エーザイ株式会社 プロダクトクリエーション本部 ポートフォリオ戦略・推進部 顧問
-------	---

#### 研究アドバイザー

伊藤 武彦	東京工業大学 大学院生命理工学研究科 教授
大浪 修一	理化学研究所 生命システム研究センター チームリーダー
小池 麻子	株式会社日立製作所 研究開発グループ ヘルスケアイノベーション統括センター センター長
五斗 進	京都大学 化学研究所バイオインフォマティクスセンター化学生命科学領域 准教授
鹿内 俊秀	産業技術総合研究所生命工学領域創薬基盤研究部門 糖鎖技術研究グループ 糖鎖科学データベース班 招聘研究員
瀬々 潤	産業技術総合研究所創薬基盤研究部門 ゲノム機能情報研究グループ 主任研究員
高橋 祥子	株式会社ジーンクエスト 代表取締役社長
高橋 眞澄	アイバイオテック株式会社 代表取締役
藤 博幸	関西学院大学理工学部生命医化学科 教授
中山 英樹	東京大学大学院情報理工学系研究科 講師

\*五十音順。所属は平成 27 年 4 月時点

#### 4. 事後評価結果

バイオサイエンスデータベースセンター (NBDC) では、データの共有と統合化を推進することにより、我が国のライフサイエンス研究から生み出されるデータの価値を最大化することを目指して統合化推進プログラムを実施している。統合化推進プログラムで統合された(または統合中の)データベースの有用性を高めるための解析ツール等を開発し、データベースのさらなる活用を促すため、統合データベース解析トライアルを実施した。

本プログラムでは、データ解析をするツール等を開発し、それをを用いて新たな知識発見を目指す研究開発提案の募集を行い、8件の応募の内、4件を採択した。6ヶ月という非常に短い研究開発期間(研究開発費:80万円程度)ではあったが、若手研究者が熱心に取り組んだ。

この度、研究開発期間を終了したことに伴い、事後評価を行った。各研究開発課題の事後評価結果については、以下の通り。

##### (1) PDBj タンパク質をゲノムにマップした pdbBAM の作成

(研究代表者:城田 松之)

PDBj に登録された全タンパク質のアミノ酸配列について、mRNA 情報を介してヒトゲノム DNA 配列にマッピングし、ゲノムブラウザで表示可能とし網羅的に俯瞰できるツールを開発した。次世代シーケンサにより大量に産出される塩基配列データから機能的に重要な部分や変異箇所を抽出・可視化することの重要性が増しており、本ツールによって、ゲノム上でタンパク質立体構造が明らかになっている遺伝子領域を容易に確認することが可能となった。

今後、例えばタンパク質の天然変性領域、アミノ酸残基のリン酸化などの修飾部位、リガンドの作用部位など、アミノ酸残基レベルの情報をゲノム情報とともに可視化することができるようになれば、さらに本ツールの活用の幅が広がると考えられる。将来的に、本ツールのアップデートや公開において、利用者のリクエストを考慮しながら開発を進めることで、さらに発展し、利用者数を伸ばしていくことが期待される。

##### (2) RDF ストア間データ連結フレームワークの開発およびオーソログ解析への適用

(研究代表者:千葉 啓和)

RDF 形式によってデータベース間の検索が可能となったが、そのためには、複雑な SPARQL 検索定義文をその都度入力する必要があり、実験研究者にとって利用しにくい環境であった。そこで研究代表者は、繰り返し入力となる部分をコマンドラインのオプションとすることによって、複雑な SPARQL コードパターンを効率よく生成することで、実験者が比較的容易に問い合わせを可能にするための SPARQL クライアントツール「SPANG」を開発した。SPARQL の活用環境を支援するようなツールがまだ少ないなかで、利用者の意見を取り入れた使いやすいツールになったことが評価できる。

「SPANG」は、SPARQL を利用したデータ解析の効率を高めることができる支援ツールとして非常に有用であり、将来的に基盤的なツールとして普及する可能性が見込まれる。今後、多くの利用者間で SPARQL ライブラリを共有できるプラットフォームとなるよう開発することで、非常に有用なツールとなることが期待できる。そのためには、普及活動にも力を入れることが重要である。

### (3) 生化学反応ネットワーク統合解析環境の拡張

(研究代表者:西田 孝三)

平成 25 年度統合データ解析トライアルにおける研究代表者の成果である KEGGscape について拡張と改善を行い、シロイヌナズナの葉と根の時系列代謝物プロファイルの組織間差異の可視化を行う解析パイプラインを構築した。本解析パイプラインを用いることにより、任意の KEGG パスウェイ上で組織間の代謝物プロファイルの差異が容易に確認できるようになった。なお、汎用性を考慮し、プロファイルのマッピングに利用する KEGG パスウェイは IPython Notebook で容易に任意のものを選択することができる。

入力となる代謝物プロファイルのアノテーションの質や、代謝物とデータベースとの対応付けのための基盤的な情報が研究開発計画時に十分に把握できていなかったため、当初計画していた RDF を利用した KEGG とシロイヌナズナのパスウェイ情報の統合については達成できず、一つの事例の解析と可視化に留まった点が残念であった。しかしながら、オミクス情報の統合には需要があり、完成できれば統合化推進プログラムの成果データベースの有用性を示すようなツールになることが期待される。代謝物プロファイルから代謝物を推測するなど、技術的に解決していかなければならない問題点も多々あるが、今後、実現可能性を見極めつつ着実に研究開発を推進していくことを期待する。

### (4) HLA 遺伝子完全配列決定パイプラインの構築

(研究代表者:細道 一善)

次世代シーケンサによる HLA (ヒト主要組織適合抗原) 遺伝子の配列決定法ならびにタイピング法の解析パイプラインを開発した。HLA 遺伝子は多型性が高く、従来のタイピング手法 (蛍光ビーズ法による PCR-SSO やダイレクトシーケンスによる PCR-SBT) では、低頻度のアレルを同定できない、タイピング結果の精度が低いという問題点があったが、本パイプラインによって、高精度な HLA 遺伝子によるアレル配列決定のためのタイピング手法を確立することができた。構築したパイプラインは、利用者の利便性を考慮し、ウェブブラウザ上で解析可能なシステムとして公開した。

HLA 領域は、リウマチや糖尿病など 100 種近くの生活習慣病、自己免疫疾患、がん、造血幹細胞移植に伴う GVH 病等の疾患や薬剤副作用に関連するため、HLA 遺伝子は予防医学上の重要性から高精度のタイピングが期待される遺伝子の一つであり、今後定量的な検証を深めることで、本ツールの臨床での診断応用が期待される。また、本ツールの手法は有用であり、ヒト以外の生物種や他領域の遺伝子多型解析等への応用が可能と考えられるため、それを考慮した今後のさらなる展開が望まれる。

以上