

**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
課題別事後評価結果について(平成 27 年度実施分)**

平成28年8月1日

バイオサイエンスデータベースセンター(NBDC)

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))において平成27年度に採択した研究開発課題について、「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、事後評価を実施した。

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

- ・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、有用な知見が見いだされたか)
- ・当初研究開発計画の進捗状況や達成度
- ・開発したツールの将来展望

2. 評価対象の研究開発課題

1) 概要

課題数: 3件
研究開発費: 100万円以内(研究チームを編成する場合、200万円以内)
実施期間: 平成27年5月15日～平成28年3月31日

2) 研究開発課題と研究代表者

研究開発課題	研究代表者	所属	役職
ChIP-seq SRA の統合的可視化とバイオデータベースとの連携	沖 真弥	九州大学 大学院医学研究院	助教
配合生薬の横断検索のためのソフトウェアツールの開発	桂樹 哲雄	豊橋技術科学大学 大学院 工学研究科	助教
高精度全電子計算に基づくレクチン-糖鎖間相互作用解析 (※研究チームを編成して実施)	中野 祥吾	静岡県立大学 食品栄養科学部	助教

*五十音順。所属・役職は平成28年3月末時点

3) 評価スケジュール

平成 28 年 3 月	研究開発成果報告会(一般公開)・事後評価会(非公開)の開催 ①被評価者からの報告、評価者・被評価者間の意見交換 ②研究アドバイザーに評価に対する意見を聴取
平成 28 年 5 月	被評価者による研究開発終了報告書の作成・提出
平成 28 年 6 月	上記報告会での意見交換と報告書を参考にした、研究総括が研究総括による評価の実施 ①事後評価結果(案)を被評価者に提示し、意見交換 ②事後評価結果の確定(全評価者の確認・合意)
平成 28 年 7 月	JST Web サイトにて公表

3. 評価者

研究総括

長洲 毅志	元エーザイ株式会社 プロダクトクリエイション本部 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長
五斗 進	京都大学 化学研究所 准教授
鹿内 俊秀	産業技術総合研究所 創薬基盤研究部門 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 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 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・役職は平成 28 年 3 月末時点

4. 事後評価結果

NBDC では、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、研究データの共有と統合化を推進するために統合化推進プログラムを実施してきた。そこで、平成 25 年度より、統合データ解析トライアルとして、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、研究開発課題の公募を開始した。平成 27 年度は 8 件の応募があり、うち 3 件を採択した。研究代表者は比較的若手の研究者であることから、研究開発の計画・実行にあたっては各課題に研究アドバイザー 1 名がメンターとなって研究指導を行った。

研究開発課題ごとの事後評価結果は以下の通り。

1) ChIP-seq SRAの統合的可視化とバイオデータベースとの連携

NGS データの公共データベース SRA に登録された膨大な ChIP-seq データについて、実験研究者の誰もが情報処理を要することなく横断的に活用できようにするため、SRA 登録データのほぼ全てである 4 万件について、メタデータの表記揺れを校正し、統合的な解析を行った上で、ChIP-Atlas (<http://chip-atlas.org>) として公開した。ChIP-Atlas の検索結果に表示される抗原名、細胞名について、PosMed、PDBj、RIKEN BRC 等の各エントリにリンクを張ることにより、関連する知見を本サービスからスムーズに得られるよう工夫されている。また、利用者からのフィードバックに基づき、利用者手持ちのデータを用いた比較解析機能が追加で実装された。本サービスを用い、非遺伝子領域から疾患に関連する転写因子の候補が得られている。

当初の研究開発計画がほぼ全て高いレベルで達成された。校正したデータを提供し、また多様な知見を利便性よく閲覧できる点で、独自性、汎用性が高い。利用者の意見や要望を得つつ進めている事も高く評価できる。今後、ChIP-Seq のデータ解釈、また、転写制御の規則性、メカニズム解明など様々な目的の研究に活用されると期待できる。

SRA 登録データの ChIP-Atlas への搭載作業は研究代表者の貢献に大きく依存しており、今後、どのように更新作業を続けていくかが重要な課題であると思われる。更新作業の簡略化、明文化を進め、協力者を得て体制化に取り組むことにより、継続的、安定的に運用されることが求められる。その上で、登録生物種の拡張や API の提供、RDF 化など、さらなる利活用拡大のための検討を期待したい。

2) 配合生薬の横断検索のためのソフトウェアツールの開発

幅広い層からの伝統的配生薬データの利用を促進するため、配合生薬の処方・効能について横断検索が可能な Android 用アプリ “Herbal Medicine Systems”を開発し、Google Play ストアで配信開始した (<https://play.google.com/store/apps/details?id=net.sonyhartono.crudedrugsystems>)。配信データに KNApSAcK KAMPO の漢方のデータ、KNApSAcK JAMU のジャムウのデータが含まれているので、一度アプリをダウンロードすればその後はインターネット通信をせずに情報検索できる。また、ユーザーが独自に収集したデータなどを追加・修正する事もできる。

当初の研究開発計画は達成され、当初計画で挙げられた機能が全て搭載された。利用者が独自のデータを登録できるという点は新しく、汎用性も確保されている。

今後、本ツールが広く利用されるためには、使い勝手の点で、さらなる開発が必要と思われる。特に、利用者などとの意見交換などによって、利用者・利用例の具体的な検討とそれを踏まえた作り込みが重要である。例えば、非研究者を対象ユーザーとする場合、アプリやマニュアルの日本語化や、他の薬の

データの配信、独自データの登録フォーマットの簡便化などが考えられる。一方、研究者や製薬会社等を対象ユーザーとする場合、独自データを登録できるという利点を活かすためには、Windows、Mac 版アプリなどへの移植等も考えられる。その他、検索結果に名称以外の情報を表示する、検索結果に他のデータベースへのリンクを張る等、利用をサポートするようなインターフェース作りが必要と思われる。

3) 高精度全電子計算に基づくレクチン-糖鎖間相互作用解析

フラグメント分子軌道法(FMO 法)を用いたタンパク質-リガンド間の相互作用の定量的な評価を容易化するため、FMO 法プログラム PAICS を用いた計算・解析をサポートするツール PyPaics を開発した。PyPaics は、タンパク質分子ビューア PyMOL のプラグインで、(1)PDBj や LfDB に登録されている構造データから PAICS の実行ファイルを作成するプログラム(PyPaics_input)、(2)PAICS の計算結果を PyMOL で視覚的に表示できるプログラム(PyPaics_output)からなる。また、PAICS の計算結果を CSV・TXT 形式及び棒グラフ・データとして出力する、Ruby スクリプトの開発も行った。上記のうち PyPaics_output と Ruby スクリプトは研究開発期間内に Web 公開され (<http://dfns.u-shizuoka-ken.ac.jp/labs/proeng/custom20.html>)、残る PyPaics_input は、執筆中の論文投稿後に公開が予定されている。本ツール群を用いてグラム陰性菌が持つレクチンの機能部位の構造データを解析したところ、機能部位の安定化に重要なアミノ酸残基の予測、レクチン-糖鎖の相互作用に寄与するアミノ酸の序列の推測に成功した。

当初の研究開発計画はほぼ達成された。独自性の高いツールであって、ターゲットの分子種やタンパク質構造を問わず解析できる点で汎用性も高い。今後、本ツールを用いて、薬剤等の合理的な設計等に有用な知見が得られることが期待される。

今後、タンパク質-リガンド間相互作用の定量化に FMO 法を用いる有用性を示し、また継続的に改良を進めて頂きたい。例えば、電子密度などの情報も PyMOL 上に表示できると良い。また、ビューア上の可視化手法は、情報可視化の専門家と連携できるとさらに改良できると考えられる。さらに、本ツール群の開発言語が複数であることがネックになる可能性があるため、改善できると良い。

以上