

研究開発課題別事後評価結果

➤ 課題情報

研究開発課題名	エピゲノミクス統合データベースの開発と機能拡充
研究代表名	沖 真弥（京都大学 大学院医学研究科 特定准教授）

➤ 事後評価結果

1. 総合評価	<p>当初の実実施計画が全て実施され、期待を大きく上回る成果が得られた。</p> <p>利用者の意見を反映した機能拡張やさまざまな改良が実施され、世界でも有数の高度なエピゲノミクス統合データベースが確立された。活用状況も大変優れており、「使えるデータベース」として国内外から広く認知され、研究インフラとしての地位を固めた。生命科学研究やイノベーションに大きな波及効果をもたらし、数多くの学術論文で ChIP-Atlas が引用された。以上、短期間で本プログラムの看板となるような期待以上の大きな成果が得られたことを高く評価する。</p>
2. 研究コミュニティを含むデータ提供者やDB利用者との連携・協業	<p>ChIP-Atlas について EMBO Reports 誌に論文報告し、海外も含め広く周知がされた。また、国内外の学会や研究集会、大学や研究機関での講演やハンズオンセミナーの開催等により積極的に広報活動が展開し、利用者の拡大につながった点を評価する。利用者からの要請に応え、当初計画になかった ATAC-seq データを追加収載するなど、研究コミュニティと十分な連携がされた点も評価する。</p>
3. 利用者にとって有用なデータ基盤、持続的な DB 運用体制構築に向けた取り組み	<p>公開されているほぼ全てのエピゲノミクス測定データを統一プロトコールで再計算し、サンプル間の比較解析ができる再解析データが収載された。これにより、これらのエピゲノミクス情報を統合的に俯瞰できるようになっただけでなく、これらのデータを使って、転写因子の共局在解析や、ユーザーのクエリに対するエンリッチメント解析等、高度な利活用が可能となった。また最新のゲノムアセンブリに対応し、UCSC ゲノムブラウザとも連携できるようにされ利便性が向上した。さらに jPOST と連携してプロテオームデータから直接データを取り込んでエピゲノム解析できるようにしたり、DBKERO で他のオーミクスデータと並列して閲覧するためにピークデータを提供したりする等、他のデータベースとの連携も強化された。また ChIP-Atlas の全ピークデータが RDF 化され、エピゲノミクスデータの他のデータとの統合利用の利便性が高まった。上記のように、新たな知識発見につながる利用者にとって有用なエピゲノミクス情報基盤が構築された点を高く評価する。</p>
4. 生命科学研究やイノベ	<p>ChIP-Atlas の利用者数・論文引用数ともに増加し、十分に利活用された。またアカデミアや企業との共同研究も多く実施され、生命科学研究やイノベーションの創</p>

ーションへの波及効果（DBを利用して得られた研究成果）	出にも大きく貢献した点を評価する。
5. その他特記すべき事項	研究代表者がPIとして独立ポジションを得るとともに、複数の研究員を採用するなど、人材の育成にも貢献した点を評価する。
6. 実施計画の達成状況	<p>ChIP-Atlasを研究開発対象とし、1) Bisulfite-seqデータの追加、2) エピゲノミクス統合データベースとしての確立、3) 持続可能な運用体制の構築、4) 利活用の拡大を目的として、研究開発が実施された。</p> <p>1)については、Bisulfite-seqデータの解析パイプラインが構築され、所与のゲノム領域のメチル化率と高/低メチル化領域が可視化された。</p> <p>2)については、当初計画にはなかったATAC-seqデータが利用者の声を受けて新たに追加された。2022年1月25日時点で、ChIP-seq 140,959件、ATAC-seq 52,135件、DNase-seq 2,869件、Bisulfite-seq 28,696件、合計224,659件のデータが収載された。これにより、ゲノム-タンパク質相互作用(ChIP-seq)、オープンクロマチン情報(DNase-seq, ATAC-seq)、メチローム情報(Bisulfite-seq)を統合的に俯瞰し利活用できる、世界有数のエピゲノミクス統合データベースが確立された。</p> <p>3)については、最も手間を要していたサンプルメタデータのキュレーション作業に機械学習を導入して半自動化されたことで、ChIP-Atlasの持続可能性が大幅に向上された。</p> <p>4)については、毎月4~5千人のユニークユーザの訪問、20万ページ閲覧のサイトへと成長し、学術論文でも2018年12月の論文発表以降、300報以上に引用された。</p> <p>上記の通り、当初の実施計画はすべて達成されただけでなく、当初計画にはなかったATAC-seqデータの収載が実施された。</p>