

2025 年度年次報告書

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

【研究課題名】「統合的な転写制御データ基盤の構築」

研究代表者：

粕川 雄也（理化学研究所 生命医科学研究センター チームディレクター）

研究分担者：

川路 英哉（東京都医学総合研究所 ゲノム医学研究センター 副センター長）

榎屋 啓志（理化学研究所 バイオリソース研究センター 室長）

沖 真弥（熊本大学 生命資源研究・支援センター 教授）



©2025 粕川 雄也(理化学研究所) licensed under CC表示4.0国際

1. 研究進捗状況と成果

【シスエレメントデータの構築と維持】一細胞 5' RNA-seq データからシスエレメント領域を同定するため、ヒト 4 データセット (50 クラスター) およびマウス 3 データセット (45 クラスター) のデータ構築と品質管理 (QC) を進行している。さらに、新規の 5' 端 RNA-seq 公共データセット (ヒト 12、マウス 2) を追加して CRE (シスエレメント) 領域を再同定した結果、ヒト 513,895 個、マウス 307,827 個へとデータが大幅に拡充された。対象サンプルもヒト 2,644 サンプル、マウス 1,269 サンプルへと拡大し、外部 DB (SilencerDB、VISTA) との連携強化を図った上で、2026 年 2 月 24 日に「fanta.bio ver1.2.0」としてウェブサイトおよび独自の検索インタフェース、UCSC TrackHub にて公開した。また、マウス変異情報について、公共リポジトリの探索や長鎖解読由来の構造変異分布の注釈追加・可視化を推進し、論文発表および外部公開を達成した (Takada et al. 2025 Mammalian Genome)。

【ChIP-Atlas 高度化によるエピゲノミクス・トランス因子データの構築と維持】NCBI SRA の月次更新に合わせ、遺伝研スパコンでのデータ解析パイプラインの構築を進めた。2025 年 3 月の新スパコンへのリプレイスおよびジョブスケジューラ変更 (Grid Engine から Slurm へ) に伴い、自動ジョブ投入システム等の大幅な修正を実施した。CUT&RUN や CUT&TAG 等の新規手法データ (21,891 個の SRX) の抽出クライテリア策定や、シロイヌナズナのエピゲノミクスデータ (計 25,294 件) の抽出を行い、順次データ更新プログラムへの組み込みを進めている。また、データクオリティの可視化セクションの新設や、生データのまま解析可能な新たなエンリッチメント解析ツールの開発など大幅な機能改善を行い、ChIP-Atlas のアップデート論文を報告予定である。

【データベース連携】fanta.bio と ChIP-Atlas のサンプル間の対応づけを精緻化し、人手によるキュレーションや LLM を用いた手法を用いて細胞株、初代培養細胞、組織への分類と ID (Cellosaurus、Cell Ontology、Uberon) の付与を行った。これにより、ヒト・マウスの各組織・細胞種にわたる多数のレコードの対応づけを完了した。また、シスエレメントが制御する遺伝子・転写産物の予測法として ABC モデルの妥当性を確認したほか、一細胞データの根拠情報として「scPortalen2」との連携を拡張した。さらに、ChIP-Atlas の Peak Browser への fanta.bio データの閲覧機能 (Annotation track) の追加や、各種細胞タイプのエンハンサーに結合がエンリッチする転写因子の網羅的探索など、実用的なアプリケーション・機能開発を進めた。

2. 代表的な成果事例

1. Zou Z, Ohta T, Kasukawa T, Oki S. ChIP-Atlas 2025 Update: 10-year anniversary of a data-mining platform for exploring epigenomic landscape. *Nucleic Acids Research*, **54**, gkag378 (2026).

概要：ChIP-Atlas の公開 10 周年とメジャーアップデートに関して報告した。収録データの拡充の他に、個々の実験データの信頼性を可視化する新機能に加え、遺伝子発現を網羅的に測る RNAseq データの上流解析をより簡便に行える新しいオンラインツールを実装した。これにより、利用者はデータの確からしさを直感的に確認しながら、真核生物における遺伝子制御の仕組みをより柔軟に解析できる。このような改良を通じて、ChIP-Atlas は利用者の多様な需要により即したエピゲノム解析インフラへと進化した。

以上