

研究開発課題別中間評価結果

➤ 課題情報

研究開発課題名 「統合的な転写制御データ基盤の構築」

研究代表名 粕川 雄也

➤ 中間評価結果：

遺伝子発現調節に関わるシスエレメントについて、その領域や細胞の種類・状態毎のシスエレメント領域の活性状態の情報を提供するデータベース fanta.bio を構築するとともに、シスエレメントに結合するトランス因子や関連するエピゲノム情報を提供するデータベース ChIP-Atlas をさらに充実させ、これらを相互に連携させることで、転写制御の総合的理解に資するデータ基盤 INTRARED を構築する研究開発課題である。

fanta.bio については、公共の 5'端 RNA-seq データを収集・処理して作成した、シスエレメント領域やそのアノテーション情報、ならびにシスエレメントのサンプルごとの活性量（転写量）などの情報を収載し、それらを探索するためのインターフェースを実装した ver.1.0.0 を構築し、当初計画を前倒して 2024 年 2 月 26 日に公開した。

ChIP-Atlas については、増大するデータ量に対応するため解析パイプラインを見直し、定期的に最新データへの更新を実施した。また 2023 年 10 月に、転写制御エレメント、Hi-C、疾患関連 SNP、eQTL など、1,000 個以上の機能注釈情報を含むアノテーショントラックを追加するとともに、当初計画にはなかった、ChIP-Atlas データセット間で差のあるゲノム領域を抽出する Diff Analysis ツールを実装し、ユーザーインターフェースを刷新した v3.0 を公開した。

また、fanta.bio と ChIP-Atlas との連携については、両データベースが提供するデータを統合的に探索できるようにするため、両データベースのサンプル情報を対応づけ、指定した細胞種・細胞株・組織ごとにシスエレメント領域とその領域に結合する転写因子の比較や、ChIP-Atlas の様々な実験によるトランス因子・エピゲノムに関するゲノム領域情報とシスエレメント領域を、ゲノム上で並べて比較できるようにした。さらに、ゲノム変異のデータ統合を進めるため、fanta.bio とヒトのゲノム変異に関する TogoVar データベースやマウスの系統間のゲノム変異に関する MoG+データベースの連携・整備を行った。ChIP-Atlas v3.0 については 2024 年 6 月の NAR Web Server Issue に論文掲載し、fanta.bio については NAR Database Issue に論文投稿し受理された。

上述の通り、進捗状況と今後の成果見込みは優れている。新規に開発していた fanta.bio の正式版を、当初の予定を前倒して公開したことを高く評価する。fanta.bio と ChIP-Atlas の間でオントロジーも含めたサンプル情報の表記を早い段階で統一した点は、今後、両データベースのコンテンツの統合的な利活用を進めていく上で非常に重要な点であり、ここから新たな研究開発が生まれてくることを期待する。fanta.bio と ChIP-Atlas のさらなる連携を図りながら、今後も概ね現計画通りに推進すべきである。

以上