

研究開発課題別中間評価結果

➤ 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

➤ 中間評価結果

公開されている ChIP-seq データを網羅的に収集し、個別の ChIP-seq データのアライメントやピークコールデータを閲覧できるだけでなく、共局在解析やエンリッチメント解析などのデータマイニングが出来る統合データベースの構築を目指す研究開発課題である。

ChIP-Atlas は、2019 年 10 月時点で、ほぼ全ての公開 ChIP-seq データ、DNase-seq データを格納している。収録生物種としては、既存のヒト、マウス、ハエ、線虫、酵母に加え、新たにラットを追加した。これらの解析データは、DBKERO のほか、外部機関の DeepBlue、Regulator Trail などにも連係されている。このほか、UCSC Genome Browser に対応し、データの可用性向上に努めた。また、サンプルメタデータのキュレーション作業を補助するための自動化プログラムを開発したことで、ChIP-Atlas の更新体制が大幅に改善された。また、学会・シンポジウムでの展示・発表、大学研究機関などでのハンズオン講習会を多数開催し、ユーザー拡大と利活用推進を図った。さらに、ユーザーからの要望を踏まえ、リファレンスゲノム更新、ATAC-seq データ収録を追加的に実施することとした一方、Bisulfite-seq データについては、パイプラインを構築したもののデータ収録を後ろ倒した。以上のように、ユーザーからの要望や外部状況の変化に柔軟に対応し、適切に計画変更を行った。

ChIP-Atlas は、公開されている研究データを再利用するための技術的障壁を排除し、可用性を高めることに大きく貢献した。2017 年以降、被引用数が毎年倍増し、2019 年 10 月時点で 109 件が確認されている。海外からのアクセスが半分以上を占め、幅広い研究分野で活用されている。

上述の通り、進捗状況と今後の成果見込みは大変優れている。しかし、今後も ChIP-Atlas を安定的に維持・更新し、かつ有用性をさらに高めていくためには、研究計画を見直し人員体制を拡張すべきである。特に、シロイヌナズナのデータ追加については、植物学の専門家と密接な協力体制を構築することが望ましい。また、RDF 化については、他の研究開発課題とも連携しつつ、今後も着実に進めていって欲しい。

以上