

（様式 1-1）

提出日：2026 年 5 月 8 日

2025 年度 大阪大学蛋白質研究所 拠点事業

研究成果報告書

（1）事業名（下記より該当事業名を選択し、ほかは削除してください。）

共同研究員

（2）研究代表者

氏名：粕川雄也

所属機関名・部局名・職名：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

（3）研究課題名（申請時に記載したものと同一課題名を記入してください。）

一細胞データによるシスエレメント活性情報同定手法の設計

（4）蛋白質研究所受入担当教員

教員名：岡田 眞里子（研究室名：細胞システム研究室）

（5）研究成果の概要

理化学研究所生命医科学研究センター生命医科学大容量データ技術研究チームでは、2022 年度より、ヒトおよびマウスを対象として、ゲノム中に存在するシスエレメント（cis-regulatory element; CRE）領域の位置情報、細胞種・細胞株・組織ごとの活性状態、ならびに近傍のゲノム変異との関連性を統合したデータベース「fanta.bio」を開発・公開している（<https://fanta.bio/>）。本データベースは、公共リポジトリ等で公開されている 5'端を標的とした RNA-seq 法（CAGE 法、NET-CAGE 法など）の配列データを用いて、転写情報に基づく活性シスエレメント領域を同定し、さらに独自に付与したアノテーション情報や細胞種等のメタデータを統合したものである。研究者がより網羅的かつ詳細なシスエレメント情報を取得し、それぞれの研究に活用できる基盤の構築を目的としている。本データベース開発では、当初、バルク 5'端 RNA-seq データを用いてシスエレメントデータの作成を行っていた。しかし近年、10X Genomics 社の Chromium 5'法や STRT 法など、5'端を標的とした一細胞 RNA-seq データの取得が可能となっている。そこで、これらの一細胞 5'端 RNA-seq データを用いたシスエレメント同定および活性量定量情報の統合を進めるため、これまで一細胞オミックス情報を用いた転写制御研究を推進してきた細胞システム研究室との共同研究を実施し、データベース機能の拡充に向けた検討を昨年度より進めている。

本年度は、昨年度の検討結果を踏まえ、公共データベースから取得した一細胞 5'端 RNA-seq データに対して実際にデータ処理を実施し、その結果の検証を行った。その結果、一部の解析手法について調整が必要となるケースが確認された。例えば、当初はシーケンスランごとにクラスタリングを実施する予定であったが、公共データには、ドナーのみが異なる同一対象サンプルの実験結果が多数含まれるケースが存在したため、データセット単位で統合的に処理する方式へ変更した。一方で、全体としては概ね当初想定どおりの結果が得られており、現在、開発中の fanta.bio データベースへのデータ取り込みを進めている。