

（様式 1-1）

提出日：2025 年 4 月 25 日

2024 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・チームリーダー

（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開発

（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

*背景および目的、方法と結果について、公開して差し支えない範囲で記載。

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・マウス等の生物種を対象にゲノム中に存在するシスエレメント（CRE）領域の場所、細胞種や cell line などの CRE 活性状態、CRE とゲノム変異やフェノタイプとの関連性をまとめたデータベース fanta.bio を開発している (<https://fanta.bio/>)。このデータベースは公共リポジトリ等にある 5' 端をターゲットとした RNA-seq 法（CAGE 法や NET-CAGE 法など）の配列データを再処理して転写情報から活性のあるシスエレメント領域を同定し、さらに同定されたシスエレメントへ独自に付与したアノテーション情報、使用した 5' 端 RNA-seq データの元となった細胞種等のメタデータをまとめたもので、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各自の研究に活用できるようにすることを目指したものである。本データベース開発は、当初バルクの 5' 端 RNA-seq を使用してシスエレメントデータの作成を行っているが、現在 10X genomics 社の Chromium 5' 法や STRT 法など、5' 端をターゲットとした一細胞 RNA-seq のデータが取得可能となっており、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や活性量の定量化が可能であると考えられる。そこで、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行ってきた細胞システム研究室との共同研究を実施し、データベースの充実化につなげるための検討を行った。

本年度は、特に一細胞 RNA-seq データを細胞種に対応するクラスタに分類し、それぞれの細胞種を推定する手法の検討を行うべく検討を行った。その結果、クラスタリング手法としては scVI+kNN+Leiden algorithm を組み合わせた方法を、細胞種の推定は scANVI による方法が望ましいとの結論を得た。今後はこれらの手法を実際の大規模データに適用しさらなる評価を行うとともに、現在開発中の fanta.bio データベースへの取り込みを行う予定である。