Title	Accurate integration of single cell transcriptome replicates
Author(s)	Loza Lopez, Martin De Jesus
Citation	大阪大学, 2022, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/88181
rights	
Note	

Form 3

Abstract of Thesis

Name	(LOZA LOPEZ MARTIN DE JESUS)
Title	Accurate integration of single cell transcriptome replicates (一細胞トランスクリプトーム反復実験データの正確な統合)

Abstract of Thesis

Single-cell RNA-sequencing (scRNA-seq) technologies have revolutionized the study of biological systems by capturing gene expression profiles from thousands of cells in the same experiment. One important application of scRNA-seq data is the comparison of two or more samples to describe genetic changes between conditions, e.g. disease or stimulation conditions. In this kind of analysis, replicated samples allow investigating subtle changes in cells composition, improving the understanding and treatment of conditions. However, these analyses are hindered by *technical differences* of samples known as *batch effects*. Batch effects must be addressed in every joint analysis due to their possible correlation with the main biological components. But this task is not trivial, as non-linearities in technical differences of samples might appear in distinct ways on every experiment. In the last three years, different methods have been developed for the integration of scRNA-seq data, allowing to create cell atlases and to perform joint analyses of datasets. However, these methods could *over-correct*, merging cells from different types. This issue is particularly problematic in the analysis of replicated experiments with small batch effects, where cells with subtle changes in gene expression could be masked, affecting the conclusion of the experiment. To address this problem, I designed *Canek*, a bioinformatics tool to integrate scRNA-seq data replicates sequenced with the same technology. Assuming a *linear batch effect within a group of similar cells*, Canek uses *linear estimation* and *fuzzy logic* to obtain cell-specific correction vectors to integrate datasets. Using tests specifically designed to assess over-correction, I show that Canek integrates datasets with the smallest amount of over-correction as compared with state-of-the-art methods. To show how to implement Canek within a workflow, I performed a complete analysis using Canek for the characterization of mouse cells from the spleen. In the same analysis, I show the parameter dependence for commonly used tools in the study of scRNA-seq data, which will serve as a comprehensive guide for other researchers. Canek is computationally efficient and can integrate thousands of cells without over-correction, a special characteristic that could lead future experiments to the improvement and/or the design of *gene-specific treatments*.

様式 7

論文審査の結果の要旨及び担当者

氏名	(LOZA LOPEZ MARTIN DE JESUS)		
論文審査担当者		(職)	氏名
	主 査	教授	近藤 滋
	副 査	教授	廣瀬 哲郎
	副 査	教授	石井 優
	副 査	准教授	Diego Diez

論文審査の結果の要旨

Shingle cell transcriptomeの使用は、現在の生物学において一般的になりつつある。データが豊富に蓄積されており、各研究者が興味を持つ細胞種の研究において、過去に行われたデータを参照することが、研究の進行を進める。しかし、他の場所で行われたデータ間には、多少の誤差が含まれるため、生のデータのままでは比較することが難しい。

LOZA LOPEZ MARTIN DE JESUS氏は、この問題を解決する新しいバイオインフォ的な手法を作成した。Canekと名づけられたその手法は、異なるデータ間の平均的な違いを検知し、線形の計算処理により、統合することに成功しており、先行研究により作られた田の方法と比較しても、より良好な統合を行えることも示している。また、この成果は既に国際的な専門誌にアクセプトさせている。

以上より、**LOZA LOPEZ MARTIN DE JESUS**の生命科学における貢献は、十分に、博士号の授与に値するものと判断した。

近藤滋