Graduate School of
Information Science
and Technology
IST

大阪大学
OSAKA UNIVERSITY

August 3rd, 2021

Dear Editors

We thank the editor and all the reviewers very much again, for their constructive comments on our revised manuscript, “SC-JNMF: Single-cell clustering integrating multiple quantification methods based on joint non-negative matrix factorization” by Shiga et al. In response to the comments, we performed additional experiments and revised the manuscript. Please check the point-by-point response to the reviewers’ comments in the following pages.

We deeply appreciate for the opportunity to improve our manuscript according to the reviewers’ valuable comments and suggestions. We hope that the revised version is suitable for publication.

Shigeto Seno, Ph.D.
Associate Professor of Graduate School of Information Science and Technology, Osaka University

On behalf of all authors.

Reviewer 1

Basic reporting

no comment

Experimental design

no comment

Validity of the findings

no comment

Additional Comments

Authors have solved all my comments, I have no further comments.

Reviewer 2

Basic reporting

see comments

Experimental design

see comments

Validity of the findings

see comments

Additional Comments

The authors tried to resolve all of my concerns, but I still have a few concerns.

1. I cannot follow why the authors used Monaco and CellBench dataset only in Fig.2. In my opinion, these datasets will be useful for other analyses (such as Fig.3).

[Our response:]:

We thank the reviewer for the suggestion. We added the results of Monaco and CellBench datasets in Fig.3, Fig.4 and Fig.5, and revised the manuscript.

There were some difficulties for us to execute the quantification pipelines including

“demultiplexing barcodes”. In CellBench dataset, RNA-seq reads of all cells are stored in a single file using barcoding technology. In the dataset we used at the first submission, RNA-seq reads were stored in separate fastq file for each cell.

Now we can handle the quantification pipelines including demultiplexing.

2. The ARI values of “salmon and kallisto” are lower than those of “kallisto” or “salmon” for the Pollen dataset. In my opinion, it is important to analyze and discuss this case for clarifying the usefulness of “joint” NMF.

[Our response:]:

We thank the reviewer for this very important comment. For the past month, we have been examining various situations to answer this comment. However, we could not come up with a clear answer.

The reason for this may be the effect of the loss function of NMF, the update algorithm, and the initial values, in addition to the dataset dependence. For "salmon & kallisto" in Pollen's dataset, our Joint-NMF can give a high ARI, however, the variance of the ARI is large. In this case, it may be difficult to obtain stable solutions by joint-NMF. This can be improved by adjusting some parameters, but originally, we do not assume to input two similar expression profiles. In an extreme case, there would be no joint effect if the same expression profiles are used as inputs.

Based on the reviewers' comments and additional experiments, the cases in which our method is useful can be summarized as follows:

1. The number of cells is sufficient large (the accuracy decreases when the subsample rate decreases).
2. Different quantification methods yield expression profiles with different characteristics.

These discussions had been added to “Discussion” as limitations.

NMF is a widely used method, but the regularization techniques and update algorithms are still being studied. This comment is a very essential and difficult issue, so we would like to work for it as a future work.

3. It is difficult to see and compare the values of different methods in Fig.5. I want the authors to revise Fig.5 so that the boxplot are arranged for the same dataset and same subsample rate, likewise Fig.3.

[Our response:]:

The reason for the arrangement of the previous Fig.5 was that we wanted to show that the higher the subsample rate, the better the accuracy for both LSNMF and our Joint-NMF method.

Now we followed the comments and revised Fig.5.