September 24, 2019

Dear Editor,

I am grateful for the opportunity to revise my manuscript. I thank all reviewers for their comments, suggestions and critiques of my manuscript and the software tool. Below I provide the point-to-point responses

**Associate Editor**

*Comments:*

*1. While there is some value in providing integrated visual environments for scRNA-seq analysis, and MatLab reimplementations can be potentially useful for certain users, reviewers considered the features of the toolbox to be currently too limited to warrant publication. In particular, I concur with Reviewer 3 that the functionalities accessible through the GUI should be extended, to offer a interactive graphical environment for a wider range of use-cases. Moreover, the reimplementation of critical functions should be validated and shown to perform comparably (exactly or similarly in the case of stochastic methods) as previous implementation (within a supp mat). Also, Reviewer 2 mentioned comparable Matlab software, which should be mentioned and positioned against.*

I have expanded the functions accessible from the GUI scGEAToolbox. I added three more tab panels: Cluster, Trajectory and Network, to the GUI. Under each of the panels, 2 or 3 methods are added. These include functions implementing commonly used algorithms (e.g., sc\_tscan and sc\_sc3) as well as new algorithms (e.g., sc\_trajectory and sc\_pcnetpar). I have carefully checked all implementations and compared the numerical outputs of my implementation with those produced by original software tools. I provide a summary report as supplementary data to show the results of the comparison.

**Reviewer #1.**

*Comments:*

*1. Since this is a single-author paper, it should be "I" rather than "We" throughout the text.*

I have made the change.

*2. Please add a little more information to explain Figure 1.*

I made a new Figure 1 and more descriptive legend text, which are more informative.

**Reviewer #2.**

*Major comments:*

*1. This note briefly describes a toolbox for analysis of scRNAseq data in MATLAB. A key features of this toolbox is its GUI, however the algorithms do not seem to be original, primarily wrappers for existing code in R. Even without original algorithms, the toolbox may be useful for some readers. However if it contains original algorithms, that should be mentioned as it would increase the paper's impact.*

I appreciate the positive comment from the reviewer. I modified the main text to suggest that the biggest value of scGEAToolbox is to provide a comprehensive set of existing functions. In the revised manuscript, I do mention two new functions I created that have not been seen anywhere else.

*Also note that there is other MATLAB code for transcriptomic analysis already in existence, forexample at https://urldefense.proofpoint.com/v2/url?u=https-3A\_\_github.com\_cortex-2Dlab\_Transcriptomics&d=DwIFaQ&c=u6LDEWzohnDQ01ySGnxMzg&r=dzRP0h5ZWyh3FOHMTgCOAg&m=vJlip4uybg8wkXmMgaF6rMAGsLpcsgPpTr0MR5nI4ck&s=q7Q\_W8hcjxk1NN9aSPDhPIDGIjjKxcT55N3uaw9WPp8&e= ,*

I added two additional citations for Matlab-based tools (one for SIMLR https://doi.org/10.1002/pmic.201700232 and another for SoptSC https://doi.org/10.1093/nar/gkz204) in the revised manuscript. I wanted to cite the code the reviewer mentioned; however, I found the paper is not a methodology paper. Given the limit of the space, I decided not to include it. Hope the reviewer understands my choice.

**Reviewer #3.**

*General comments:*

1. *Single-cell RNA sequencing (scRNA-seq) offers gene expression measurements at single-cell resolution and makes it possible to study molecular mechanisms at the single-cell level. A large number of methods for the analysis of scRNA-seq data have been developed (mainly in R or Python). Here, the authors developed a Matlab toolbox for scRNA-seq data analysis. This toolbox includes many categories of functions that can be used for scRNA-seq data analysis, such as gene filtering, data normalization and visualization. The authors implemented a GUI application for facilitating users to perform scRNA-seq data analysis. However, only a small subset of functions included in the tool box can be accessed through GUI. Therefore, many of the functions (such as SC3 for clustering) still need to be run from command line, thus not providing any additional convenience compared to directly using the native tool (say the SC3 R package). In summary, the contribution of this toolbox to the field is very limited. My major concerns are detailed in the following.*

I thank the reviewer for point out the issue in the original design of scGEAToolbox. Now I have expanded the set of functions that can be accessed through the GUI. Briefly, I added three more tab panels to the GUI. These include *Cluster*, *Trajectory* and *Network*. By including these function groups, the scGEAToolbox has become more easy-to-use than before.

*Major comments:*

*1. As described in Fig. 1, only six types of functions (filter, normalization, batch correction, imputation, feature selection and visualization) can be accessed through GUI, with many of the functions included in the toolbox such as cell clustering, trajectory analysis and network construction are not implemented in the GUI. This makes the value of this work very limited. Taking SC3 as an example, I believe that it would be difficult to persuade users to use the re-implemented Matlab version of SC3 instead of the native SC3 R package.*

As said, I have expanded the functions accessible from the GUI scGEAToolbox. I added three more tab panels: Cluster, Trajectory and Network, to the GUI. Under each of the panels, 2 or 3 methods are added. Matlab implementation of SC3 in the function of sc\_sc3 is much faster than its original implementation in R. The results generated are comparable. For example, using an example data, I found that the normalized mutual information (NMI) value between the cluster outputs produced by Matlab and R implementations is greater than 0.92, indicating a strong consistency. This comparison has been included in a summary report in the supplementary data of this manuscript.

*2. I would suggest to add more functions to aid data analysis. For example, the authors may consider adding functions to compare the data before and after the removal of batch effect, which may help users to investigate whether batch effects are removed or not.*

I included a new Figure 1 to show that the combination of scGEAToolbox functions can produce new workflows, which is main value of the toolbox. I updated the demo script file #2 by adding the comparison of data sets before and after batch effects. The visualization of comparisons is to the file.

*3. The authors are suggested to move the description of the algorithm of each functional category In the Results section to the Method section.*

This has been done. I revised the whole results of the manuscript.

*4. I also suggest to add more demo examples/scripts to illustrate how to use the toolbox.*

I added one additional demo code file to showcase the usage of trajectory/pseudotime analysis and single-cell gene regulatory network construction functions.

*5. As some algorithms such as the “sc\_sc3.m” is re-implemented in Matlab. The authors need to prove that the re-implemented version can reproduce the results obtained by the native one.*

I have carefully checked all implementations and compared the numerical outputs of my implementation with those produced by original software tools. I provide a summary report as supplementary data to show the results of the comparison. These include the comparison between sc\_sc3 (my reimplementation of SC3) and R/SC3 (the original package).

*6. In addition to the commonly used analysis tools, the most recent algorithms for each category of functions may be added. For example, the authors re-implemented two feature selection methods but both of them were proposed before 2016.*

I am not aware of the new development of feature selection algorithms. From real-data analyses, I found that the model-based algorithm although published before 2016 still produce reasonably good results. scGEAToolbox will be further developed and upgraded. If the reviewer has strong opinion on what exact algorithm(s) I should include in the current version, I would like to know and implement it.

*7. Figure 1 is of poor quality. The screen shots of the GUI are obscure. In addition, it is necessary to add figure captions to describe each sub-figure..*

I provided a new Figure 1. A PDF version of the figure is provided now.

Again, I appreciate all these valuable and constructive comments. Thank you for taking the time and energy to help me improve the paper.

Sincerely,



James Cai, Ph.D.

Associate Professor

Department of Veterinary Integrative Biosciences

Department of Electrical & Computer Engineering (affiliated)

Texas A&M University

4458 TAMU

College Station, TX 77843-4458

Tel (979) 458-5482 | Fax (979) 847-8981

jcai@tamu.edu