We are very appreciative of all the reviewer’s suggestion and criticisms.

As for reviewer 1’s

|  |
| --- |
| Reviewer 1: Regarding: Single Cell Explorer, collaboration-driven tools to leverage large-scale single cell RNAseq Data. The authors report the release of a web based tool for the rapid and simple analysis of single cell data from sources such as the sequencing of X10 Chromium generated libraries. While the authors make ambitious claims for the web application, it is not possible to determine if it is really as good as they say. Consequently, before any meaningful review of the manuscript can be made the authors should: 1: Produce a manual explaining how to use the website as it is not immediately obvious how to use it and what the results it produce's mean. 2: Describe in both the paper and especially in a manual how to use it, what the application does, how it does it and why you'd want it done. 3: Clean up the Github site as numerous links to files that don't seem to provide any information. For instance it contains files like nohup.out and .RData which in my experience contain information produced by an application when its output is disconnected from the terminal (nohup.out) or gives the state of an R environment when it closes (.RData) the value of either is not apparent to me. 4: Rewrite the paper as it doesn't really describe what the application does etc. it just uses keeps saying its written in python, uses a MongoDB database and CCS and HTML5 etc. etc. which is interesting only in passing to most bioinformatics users, but doesn't give any insight in to what a user really wants to know like what it does, how it does it and how to use it. |

1: For the reviewer’s concern, we produced instruction manual webpages including a video tutorial of using the web app, which includes including gene search, manual cluster selection, differentially gene expression, cell annotation.

2. We are appreciative of the suggestion. We added a section of using software in the revised manuscript.

3: We are very sorry that non-essential information in Github. We removed .Rdata and nohup.out. To encourage R users to use the system, we added example R codes to enable users to utilize R Seurat object and Loom files. We moved Rdata, loom files, Seurat to a data folder for users to download when necessary.

4. We added contents about insight about using the tool for collaborative research.

|  |
| --- |
| Reviewer 2: As single-cell -omics datasets become more common, reliable annotation of cell subpopulations is a vital task. Packages such as Seurat provide sophisticated methods for clustering and visualization, but refining these automated analyses often requires manual intervention. The authors describe Single Cell Explorer, an application that facilitates this process and allows for some collaboration between computational and experimental biologists. The scope of this tool is limited, but it serves to fill an important role. Their manuscript is worth publishing in BMC Genomics.  A major strength of this application lies in allowing users to designate clusters based on freehand selection, and sharing these assignments. I also appreciated that a number of known cell markers were readily available. The features for gene expression visualization were useful as well, allowing users to make plots without needing familiarity with R or Python. However, it took some experimentation to understand how to use these features, and I feel that a guided visual tutorial would be helpful.   From a technical perspective, I have some apprehension about installing the dependencies and setting up the mongodb database. This area could be improved if the authors provided a conda environment and could confirm its successful use on some other popular Linux distributions and perhaps MacOS.   Minor corrections: -Would benefit from careful reading by a native English speaker. -A lot of inconsistent capitalization, e.g. UMAP vs. uMap vs. umap. -Figures contain wavy underlines from spell checking. -Might be useful for one of the figures to reflect an actual screenshot of the tool. |

1. To address the review’s concern about installation and mongoDB setup, we created an installation manual that lists all detailed steps to install the web application and MongoDB set up.

2. The manuscript has been thoroughly revised and edited by native speakers to address the spelling and capitalization.

3. A screenshot of the software was added in revised manuscript.

|  |
| --- |
| Reviewer 3: The manuscript presented by Feng, Whithurst, Shan, and Yue describes a new tool they have built to facilitate analysis of single cell RNAseq datasets. The authors develop Single Cell Explorer, a web server application, with the intention of enabling collaborative exploration and annotation of scRNA-seq data. This work does not establish new analytical tools, but rather provides a platform and a user interface through which researchers can work together to explore data and analyses generated offline.  In general the authors need to provide more details throughout the paper in order to more accurately capture the potential and any possible limitations of their approach. Perhaps a vignette or two to describe how the authors intend the application to be used (e.g. by genomics centers, single-PI labs, multi-lab collaborations, model organism databases, etc). It would also help to more directly specify the manner in which Single Cell Explorer can be integrated into existing pipelines (e.g. once researchers have data in some specified format, the data can be loaded into a python notebook, etc). This would also help to delineate what is actually being provided in this git repository vs what analyses need to be performed offline before loading into the application. |

We agreed. A vignette or two to describe how the authors intend the application to Genomics centers, single-PI labs, multi-lab collaborations, model organism databases,

The web application was designed for integrating with other analytic systems. In the instruction manual on our website, in addition to demonstrating data analysis using python notebook, we also published example code for python notebook. The code includes examples for single cell RNAseq data analysis, FASTQ processing by 10X Genomic Cell Ranger, data integration of Seurat object, and adding Loom files to the system.

|  |
| --- |
| The authors should more explicitly delineate what they see as the possibilities as well as the limitations of this new software. For example, will it handle data from all genomes equally well, or are some data sources more streamlined (e.g. ease of implementation for a non-standard model organism vs, say, drosophila). Are there certain platforms or data analytical pipelines for which the data have been tested? The authors mention Chromium input files. Are there other pipelines that have not been tested and may or may not work? Would they all work and integrate eventually through modification of which standard procedures described? |

The current application is able to host data from different species with different genomes. In addition to human blood and tissue samples, we added the drosophila mid brain sample in the demo website to demonstrate this. We added non-standard model organisms as features in discussion section since BMC Genomic readers worked on a variety of organisms.

|  |
| --- |
| Furthermore, there are unsubstantiated statements that can and should be expanded upon. For example in the background section (Pg4 Ln41) the sentence that begins "In contrast to the existing R-based frameworks..." demands more context. How does the current tool being python-based explicitly contrast with R-based frameworks? If this contrast is important it needs to be more clearly justified (i.e. what are the limitations of an R application that are solved by using the python framework), otherwise such rhetorical statements should be omitted. |

We are very sorry for not elaborating more about the choice. We added the rational of using Django as our web frame work versus R Shiny in the revised manual.

|  |
| --- |
| The authors should also address the many small omissions, typos and other errors that dramatically hinder the interpretability of their work. For example there is an unlabeled table (Pg 10 Ln 40) that I believe to be the python API functions designed for data retrieval. If so, the table needs a proper legend to communicate what is being presented and how it integrates into the work. Furthermore there are very many typos throughout the manuscript (verb tense issues, missing articles, subject-verb agreements, etc) that need to be addressed.  Overall, once their work is described more thoroughly I think this software will be a useful tool for the growing number of scientists pursuing single cell RNAseq approaches and warrants publication in BMC Genomics |

We are so sorry about the omissions and typos. The revised manuscript has been thoroughly revised and edited by native speakers. We added API table titles in the revised manuscript. We also published example code to explain how to use these API methods, which can be found in the manual on our website.

|  |
| --- |
| Reviewer 4: The authors present a report of a python-based web server application developed for the researchers to annotate cell expression phenotypes and share findings, named "Single Cell Explorer". Looking at the website, GitHub website, the app is well documented. The application seems useful for users to make their results available. The authors presented a superficial contrast their website to the website presented by the Cellxgene(Initiative). |

As for this concern, we added a comparison with Cellxgene’s features in the revised manuscript and on the website. Currently, Cellxgene needs to load h5ad files for a particular dataset or map each time, which hinders its ability to serve as a data portal for multiple datasets with concurrent users. We tested Cellxgene before and watched the development. It used flask web framework, but little progress was made in the backend, and the previously mentioned problem has not been solved. Current development efforts are all around front end UI with the addition of testing a few JavaScript packages. No database support or other database management has been planned.

|  |
| --- |
| On its current state the website is functional, but I was not able to access a local installation and the performance of their software on an independent database. |

We published codes and installation methods in READMe.md in Single Cell Explorer Github. Now we created a website as a digital manual. We also provided detailed step by step tutorial for local installation (Ubuntu).

|  |
| --- |
| While there is increasing generation of transcriptome data from thousands of single cells it is unclear that this software is of "broad utility that represents a significant advance over previously published software", as stated on the journal's website. There might be other venues that are better suitable for this type of paper, such as BioData Mining. |

We are appreciative of the comments. Our system is built to meet the needs of hosting an increased number of single cell datasets generated by the research community. Our system is suitable for use in biotech, pharmaceuticals, and university core facilities or individual labs have bioinformatics support. We believe that it could be broadly used due to the increased interest from both academic and industry laboratory scientists to explorer public and own single cell datasets. When bioinformatical support is not available, high quality figures can directly downloaded for publication purposes. As far as we know, the only platform that supports hosting multiple datasets is single cell portal (Ruby on Rails) from Broad Institute. However, their tool required users to use Google Cloud (Google Compute Engine virtual machine) and Broad Institute's FireCloud platform. That means users need to adopt the specific ecosystem from Broad Institute and Google Cloud. In contrast, our solution does not rely on any particular cloud ecosystem. You can install it in your Linux server, AWS and other cloud platforms. We demonstrated the significant advance over previously published software such as Cellxgenes. We also compared our software to Scope, which is Node.js based and only supports data loading, not databases. We believe the BMC Genomic audience will be interested in a general purpose web application system to study a variety of organisms. The articles from Biodata Mining are very specific to address specific problems, but our system is indeed built for general purpose in single cell genomic field.

|  |
| --- |
| Reviewer 5: In this short paper, Feng et al describe Single Cell Explorer, a python-based we server application for the annotation and exploration of single cell RNAseq data. The application focuses on data sharing , visualization and exploration using a visually intuitive and easy to use understand interface . The link to github was not working for me, so I had to rely on the figures the authors provided for the review and the demo environment. Overall the screenshots appear appealing but due to my inability to access the server could not verify the user responsiveness. Implementation appears to rely on state-of-the-art approaches, yet the large number of dependencies suggest that installation and deployment may not straightforward for the novice user. In that regards, I hope the authors could provide detailed installation instructions (or even an installation script) for users who want to go beyond the demo environment. |

We are sorry that you have trouble checking the github. You can also use a domain name [www.singlecellexplorer.org](http://www.singlecellexplorer.org) to access our website. We added the installation manual in website to list all detailed steps for installing and launching the web application and for MongoDB set up. A video tutorial of using the web app was produced. Additional, we also published notebook example code for single cell RNAseq data analysis, FASTQ processing by 10X Genomic Cell Ranger, data integration of Seurat object, and adding Loom files to the system.

|  |
| --- |
| I also feel that there should be a reasonable disclaimer about the size of the dataset that can be analyzed in this environment without running into memory issues. The analytic plots selected (t-SNE or u-Map) appear reasonable, but is there the option of adding more conventional cluster analyses plots (e.g. based on Kmeans?) |

We agree. The ratio in MongoDB of working set to available memory in the system has a major impact on system performance. We added the minimal requirement in FAQ on our website and in the revised manuscript as well. The system is able to plot heatmap for multiple genes. The conventional cluster analyses can be done in Python notebook.

|  |
| --- |
| I personally disagree mildly with the authors that statistical exploration based on the Wilcoxon test is adequate, yet I feel that going above this test may compromise the exploratory nature of the application. Perhaps the authors should comment about functionality (if it exists) to add statistical analyses above and beyond the Wilcoxon? |

We are very appreciative of the suggestion. According to Soneson et al (Nature Methods volume 15, pages 255–261, 2018), for non-complex study design, the Wilcoxon test had a comparable (good) performance in comparison to many other methods. Currently, Seurat, a popular R package, also uses the Wilcoxon test to identify cell markers. For the exploratory nature of the application, we allow users to verify DE results by manually checking the percentage of the expression in each cell cluster as well as the average expression levels. If specificity is the priority for cell markers, users probably will choose genes with higher percentages in the clusters of interest. The purpose is to identify cell types and save marker genes or genes of interest in the database. The information can be called by the API. This is not the end of the analysis. Since our app is designed for collaborative efforts, we tend to leave other statistical tests to computational biologists or users who are willing to run scripts on Python notebook or R studio. We published Python code for users in this revised manuscript. The data can be retrieved using our published API. We will continue to implement the methods widely adopted by research community.