Editorial comments #3:

1. Please provide the experimental values in the data processing. I have commented on what values are necessary in the attached manuscript. Please include these values as we need them in order to script and film. We cannot script and film a generalized protocol.

We have re-done the protocol section based on the above instruction and believe that they now better conform to what is described above with respect to the exact values/parameters used in the example.

2. Please use the attached link to upload the ExCYT files and the example data. This will ensure that users will be able to download and access both the software presented in the article as well the example data to fully replicate the protocol:

https://www.jove.com/account/file-uploader?src=17529313

We do not believe we can re-post the data from the site we got the data from directly in JOVE as this would likely constitute and breach of the data use agreement. We have provided the link to were the data for the example can be freely downloaded on line 111.

We generally prefer to use version control capable approaches to release software, such as is freely available in github. We have made ExCyt software freely available at the github link shown on line 99 of the manuscript.

Editorial comments #2:

1. As currently written, it is not clear how to replicate the protocol. Many essential step-by-step details are not included. We cannot film a generalized protocol and we need a specific example in order to film. Furthermore, we need explicit user input commands throughout in order to film, especially for computational protocols in a GUI. It is best to include all user input commands: File | Save | etc. There should be greater emphasis on Figure 2 and Table 1 in how the protocol is performed so the experiment is replicated.

3. The Protocol should contain only action items that direct the reader to do something. Please move the discussion about the protocol to the Discussion. Step 4.2 is only a discussion of the possible clustering methods. We cannot film this.

4. Please ensure that the entire protocol is written in the imperative tense as if showing someone how to perform the protocol. The Protocol should be made up almost entirely of discrete steps without large paragraphs of text between sections. Please simplify the Protocol so that individual steps contain only 2-3 actions per step and a maximum of 4 sentences per step.

The inclusion of more explicit user commands would also help allay reviewer concerns regarding the replication of the protocol.

We have re-done the protocol section based on the above instruction and believe that they now better conform to what is described above

Can the Chevrier data be provided as an example and used in the protocol as the example? This would greatly enhance the reproducibility of the protocol.

Yes, we have added a link to the data in the manuscript (lines 109-111)

2. Additional comments are in the manuscript.

We have addressed these in the manuscript.

5. Please place the superscripted numbered references before the punctuation.

We believe we have corrected these errors now.

6. Please ensure that all graphs have units and labels on the x and y axes: t-SNE heatmaps, Figure 3B, etc.

The axes of t-SNE scatter plots are arbitrary and in arbitrary units, which in many cases not noted in the figure. We noted this in the manuscript and have added more description for the coloring of the data points (lines 102-106) for clarity. We believe this is in line with how these type of data are shown in the literature.

7. Please use the following format for references. Please do not abbreviate journal titles.

[Lastname, F.I., LastName, F.I., LastName, F.I. Article Title. Source. Volume (Issue), FirstPage – LastPage (YEAR).]

We believe we have corrected this now.

**Editorial comments #1:**

*We have attempted to address all editorial comments. We have added clarification to the editorial comments and the changes we made were needed below.*

1. Please take this opportunity to thoroughly proofread the manuscript to ensure that there are no spelling or grammar issues. The JoVE editor will not copy-edit your manuscript and any errors in the submitted revision may be present in the published version.  
  
2. Please provide an email address for each author on the first page.  
  
3. Please define all abbreviations before use.  
  
4. Please remove all commercial language from your manuscript and use generic terms instead. All commercial products should be sufficiently referenced in the Table of Materials and Reagents.  
  
5. Please ensure that all text in the protocol section is written in the imperative tense as if telling someone how to do the technique (e.g., “Do this,” “Ensure that,” etc.). The actions should be described in the imperative tense in complete sentences wherever possible. Any text that cannot be written in the imperative tense may be added as a “Note.” However, notes should be concise and used sparingly. Please include all safety procedures and use of hoods, etc.  
  
6. The Protocol should be made up almost entirely of discrete steps without large paragraphs of text between sections. The Protocol steps should contain only 2-3 actions per step and a maximum of 4 sentences per step.  
  
7. Please ensure you answer the “how” question, i.e., how is the step performed? Alternatively, add references to published material specifying how to perform the protocol action.  
  
8. For steps that involve software or analyzing tools, please make sure to provide all the details such as “click this”, “select that”, “observe this”, etc. Please mention all the steps that are necessary to execute the action item. Please provide details so a reader may replicate your analysis including buttons clicked, inputs, screenshots, etc. This is the level of detail we’re looking for. Please keep in mind that software steps without a graphical user interface cannot be filmed.  
  
9. We cannot film the steps that are not fully described. Please do not highlight those steps which actions are only explained in the external references. Please keep those steps in the protocol but not highlighted. Please attention that your manuscript after acceptance will be the source for our script writers for the video production.  
  
10. Please avoid usage of phrases such as “could be”, “should be”, and “would be” throughout the Protocol.  
  
11. Please leave a blank line between all protocol steps as well as Notes.  
  
12. Protocol: 1-6: Please use the imperative tense for all sentences in the Protocol steps. Please avoid usage of phrases such as “could be”, “should be”, and “would be”, “may be”, etc. throughout the Protocol. Please revise the protocol text to avoid the use of any personal pronouns (e.g., "we", "you", "our" etc.).  
  
13. Protocol 1.2: Please avoid using personal pronouns in the protocol steps.  
  
14. Protocol: 2.3: Please highlight complete sentences (not parts of sentences). Please ensure that the highlighted part of the step includes at least one action that is written in imperative tense.  
  
15. Protocol: 2.4, 2.5: Please avoid the use of the personal pronouns in protocol steps.  
  
16. Protocol: 3.1: Please specify the computer characteristics.  
  
17. Protocol: 3.2: Please avoid the use of the personal pronouns in protocol steps.  
  
18. Protocol: 4.2.1, 4.2.8: Please move the discussions to either the introduction or the Discussion section. Please ensure that all the protocol steps are written in the imperative tense. Please avoid using any personal pronouns in the protocol steps.

* In regards to details provided with clustering algorithms, this is not really a discussion. We believe that as a user is following the protocol, they should be aware of important details in the choice of the algorithm and should not have to go to the discussion of the manuscript to decide which algorithm to use.

19. Protocol: 4.3: Please move this step to the Discussion section.  
  
20. Protocol: 5.3: Please highlight complete sentences (not parts of sentences).  
  
21. Protocol: 6.1: please avoid using any personal pronouns in the protocol steps.  
  
22. Protocol: 6.3: Please move the discussions to the Discussion section.  
  
23. Protocol: 6.4: Please highlight the complete sentences.  
  
24. After revising the protocol, please highlight 2.75 pages or less of the Protocol (including headings and spacing) that identifies the essential steps of the protocol for the video, i.e., the steps that should be visualized to tell the most cohesive story of the Protocol. Remember that non-highlighted Protocol steps will remain in the manuscript, and therefore will still be available to the reader.  
  
25. Please include all relevant details that are required to perform the step in the highlighting. For example: If step 2.5 is highlighted for filming and the details of how to perform the step are given in steps 2.5.1 and 2.5.2, then the sub-steps where the details are provided must be highlighted.  
  
Figure and Table legends:  
  
26. Figures 1-4: Please add the corresponding colorbars to the graphs. What is each cluster?

* We tried multiple ways to denote cluster identification. Ultimately, since the number of clusters can get very high, and as we have seen in literature, color bars tend to be not very helpful in tracking clusters to t-SNE plots, we decided instead within the program for a user to identify a cluster by scrolling the mouse over the actual cluster. When one does this, the cluster number comes up. Also, when the user creates high-dimensional flow-plots, the color used in the plot is the same as the one used in the t-SNE plot. This has been how we have chosen to ‘color-code’ the outputs of the program. In a most recent submitted publication, our colleagues found the combination of t-SNE plots and high dimensional flow plots, they were able to accurately convey cluster identity.

27. Figures 1-4: Please add a label and the corresponding unit to each axes.

* The units of t-SNE analyses have no physical meaning. As t-SNE is dimensionality reduction technique, compressing 48 dimensions with units to 2 without dimensions means we cannot include units to axes.

28. If you are reusing figures from a previous publication, you must obtain explicit permission to re-use the figure from the previous publisher (this can be in the form of a letter from an editor or a link to the editorial policies that allows you to re-publish the figure). Please upload the text of the re-print permission (may be copied and pasted from an email/website) as a Word document to the Editorial Manager site in the "Supplemental files (as requested by JoVE)" section. Please also cite the figure appropriately in the figure legend, i.e. "This figure has been modified from [AUTHOR] et al.[REFERENCE]”.  
  
29. Please ensure that the references appear as the following: [Lastname, F.I., LastName, F.I., LastName, F.I. Article Title. Source. Volume (Issue), FirstPage – LastPage, doi: DOI (YEAR).] For more than 6 authors, list only the first author then et al.  
  
30. Please abbreviate all journal titles.  
  
31. Please revise the table of the essential supplies, reagents, and equipment. The table should include the name, company, and catalog number of all relevant materials in separate columns in an xls/xlsx file. Please list all the materials, equipment, instrument, and software used in your work.  
  
  
**Reviewers' comments:**

*We have address each reviewers comment below.*  
**Reviewer #1:**  
Major Concerns:  
The manuscript describes a pipeline for analysing multidimensional flow cytometry data, implemented as a "Software package" named Excyt.  
While the manuscript seems to give a good description of how to use the pipeline, it is impossible to verify and use the software.  
I see no link where it can be downloaded. It has not been published previously, as far as I can tell.  
If this manuscript is the publication for this software, some details about implementation (in Materials and Methods) should be provided, as well as information  
how to obtain it, e.g. a web page. At present, it is impossible to judge the software. At present only a Table 3 in Excel format exists, which can be dropped  
anyway and incorporated into the main text, given how little info it has.  
Also, I presume a video will be made how to use the software. It would be  
nice if that were available so as to judge how well the pipeline is described (in words and video).  
Overall, the paper looks good, but in the absence of the software, it is impossible to really review it.  
  
Minor Concerns:  
Name of the Package "Excyt". When searching for this, a lot of hits come up with "Excyte", a flow cytometry dedicated web site.  
Something more unique might be better.  
Table 1 and 2, especially table 1, should and can also be provided as PDF. I.e. what are the explanations of the Software User interface  
made as Excel file? A 1-2 page word/pdf document would be better, text within cells can also be wrapped, doesn't need to be all single rows.

* While we understand that the name may be confused with other types of cytometry services, there is no software package that is termed ExCyt or a variation of it and since we have had a few people already use the software and know it by ‘ExCyt,’ we would prefer to keep it as this. In regards to how we provided information on how to use the software, we used a template from a previously accepted article in Jove that was describing a MATLAB GUI. (<https://www.ncbi.nlm.nih.gov/pubmed/28745622>)

COMMENTS AFTER REVIEWING SOFTWARE:  
First about the “exe” file.  
In the manuscript the authors state:  
"80 we have    
81 developed ExCYT, a graphical-user interface (GUI) that can be easily installed on a PC/Mac that    
82 pulls many of the latest techniques including dimensionality reduction for intuitive visualization,    
ExCYT is a graphical user interface built in MATLAB and therefore can either be run within    
86 MATLAB directly or an installer is provided that can be used to install the software on any    
87 PC/Mac.  "  
  
If the authors can only provide an .exe file for PC, this sentence is misleading (as it requires at least MatLab), and it cannot be installed without MATLAB on "any PC/Mac”.

* We have modified the text to reflect that ExCYT is only available right now as a MATLAB script and PC executable. Many well known scientific packages are written to be only supported by one operating system in binary format (with source code provided) and we think this is acceptable, given we are providing this as a free tool, largely to academics who will either have a PC computer or a MATLAB license through their university.

Regarding the instructions in the manuscript.  
Alas, with the information given there, I was not able to generate any plots.  
Part of the problem is that text is overwritten in the interface, here is a screenshot.  
Changing font preferences in MATLAB does not help.  
  
Following the instructions (after importing facs files) I come to  
"Finally, the user will be asked to select the channels they want to use for the analysis.  
Generally, we recommend using stained channels as the ones to perform the downstream analyses on. However, if the user believes the physical parameters of the cells are important to them, they can also select the forward and side-scatter channels.”  
But when clicking on the “t-SNE” button, nothing happens.  
I cannot find any place where one is ask to select the channels one want to use.  
  
At this point, the description is not adequate.  
The authors should go through the whole manuscript with a user that has never used the software, and  
provide all steps necessary to really get this to work.

* Without screenshots and command line output from MATLAB, it is really difficult for us to address this bug as we cannot replicate it ourselves. We have given the program to several of our colleagues within the lab to use the software and they have not had these issues. They have been able to run through whole analyses easily without any major problems.

**Reviewer #2:**  
Manuscript Summary:  
Sidhom et al present a new cytometry data analysis software, ExCYT, that is a graphical user interface for MATLAB. They claim to offer a uniquely accessible tool for high dimensional flow cytometry data analysis.  
  
Major Concerns:  
The issues with the manuscript and the software fall broadly under three categories: 1) the rationale presented for development and use of ExCYT is severely dated or simply inaccurate; it likely reflects the authors limited experience with cytometry data analysis tools and their history. 2) the software is not really that unique. 3) its accessibility and usefulness will be compromised by how slow it is, how buggy it is, and the fact that it doesn't account for real world differences in how flow data will be presented to the software (it is not flexible enough to be used in different places, and doesn't have good error handling/user interaction).

* While our experience maybe different than the reviewer’s, we do believe the area of cytometry is changing dramatically in the utilization of high-dimensional assays such as CyTOF and more capable flow cytometers, able to stain up to 48 colors simultaneously. Therefore, we believe the current state of the art in terms of analysis packages has much room for improvement and innovation in terms of how to analyze the larger dimensionality of these newly generated datasets. While there are implementation of t-SNE and clustering methods within other software packages, we believe our novelty does not necessarily lie within these capabilities but within how to explore the clustering solutions through ExCYT’s ability to sort/filter clusters by user preferences and visualize their characteristics through heatmaps and novel high-dimensional flow plots. These features have already been appealing to colleagues who have already used ExCYT to create visualizations for presentations and have submitted manuscripts with ExCYT visualizations.

Specific issues:  
1). Long abstract: Authors assert that it is difficult to analyze data objectively with traditional manual gating. This is incorrect. Manual gating is actually the gold-standard against which automated approaches are compared, manual gating steps are key parts of high dimensional cytometry workflows that ensure compensation is correct, that antibodies are working, etc. So, the problem is not that manual gating is inaccurate, it is that it is inconvenient, and it is challenging to visualize cell populations in multidimensional space using 2D plots and manual gates.

* We appreciate the reviewers comment and their insights. However, we feel manual gating is by the nature of human annotation not an objective data driven process; however, we accept that despite this difficulty experienced flow cytometry analysts can produce reasonable results and interpretations from high dimension data given adequate training. Additionally, it would important to note that we have geared this tool to analyze data from technologies like CyTOF with 40+ channels (as demonstrated in the manuscript) for which manual gating is likely to be problematic.

2). Long abstract: Authors state that methods for high dimensional data analysis are not available in easy to use packages. This is fundamentally wrong. tSNE is implemented as an extremely to use platform in FlowJo v9. SPICE is a tool, available for many years, that also performs statistics for high dimensional data analysis. FlowJo v9 also has multiple easy to use clustering and binning tools. Gemstone and FCS express offer unique and easy implementations of other data visualizations. Further, it is trivial to export flow data to a text file and analyze it with GUI for graphing and statistical software (like JMP, and other MATLAB plugins). Cyotbank and Cytoflex also offer easy visualization. So the authors fundamental rationale for this tool - that it is an unmet need - ignores a myriad (really a myriad!) of tools that address high dimensional cytometry.

* We have included all these in the manuscript for completeness as recommended by the reviewer. However, we feel that in terms of opensource / freely available software we have yet to see a tool that allows the user to not only implement these routine analytics such as t-SNE and clustering algorithms but also provides a novel interface to sort/filter these clustering solutions while visualizing their complete characterization via heatmaps and novel high-dimensional flow cytometry plots. While nuanced as well (and reflected in Table 2), many of these packages that offer the ability to cluster usually just provide a simple K-means clustering algorithm which assumes ‘spheroidal’ clusters in high-dimensional space. This often is a simplistic clustering algorithm which fails to capture the non-spheroidal nature of the data in high-dimensional space. ExCYT provides alternatives to K-Means such as graph-based and gaussian mixed-models methods to better deal with the nature of high-dimensional single cell data. However, despite what we believe are improvements to the clustering options to the end user, we do believe our true novelty lies in ExCYT’s ability to explore the nature of these clusters in ways not provided by other software packages.

3). Long abstract: The authors claim that their compensation calculations will be more accurate than the instrument-calculated ones. What is the basis for this assertion?

* We have softened this comment in the manuscript by stating that we apply a method for more objective calculation of the compensation matrix as we determine the positive population in an algorithmic manner as oppose to asking the user to define the positive population manually.

4). Introduction: "manual gating (has) sufficed for experiments (with) a few markers ... that give visually discernible populations." This is incorrect. Manual gating works also for markers that are not clearly separated from negative, provided that proper controls are run. Manual gating is the basis of many data analyses from experiments using even 8-30 different markers (not just a few markers). Again, the problem is not necessarily gating, it is having a means to interpret and analyze the data.

* We have modified this section of the manuscript to clarify that manual gating is more straightforward when there are visually discernable populations as these populations do not need isotype controls to determine what is positive/negative. However, we do maintain that in higher dimensional space, manual gating is problematic especially since many of these markers in large panels are often not canonical lineage markers but rather functional markers that will often have unimodal, gaussian distributions where isotype controls are needed to place appropriate gates in order to detect biologically meaningful shifts. As panels get larger and larger, as large as 48 colors, isotype controls for the majority of these channels would quickly become overwhelming and infeasible. As can be noted by many recent publications where high-dimensional single cell data is analyzed, such as the one used in this manuscript, manual gating is only used to identify live cells while the rest of the downstream analysis is done via automated clustering algorithms.

5). Introduction: "manual gating has failed to generated reproducible results..." This is an ABSURD statement. Manual gating has been used for years, on multiparametric datasets in fact. In fact, one might argue that newer high dimensional clustering algorithms are less reliable (or reliability has not been demonstrated). The GOLD STANDARD for gating is manual gating. Are authors aware of flowCAP effort? It compares automated algorithms to the consensus manual gating of humans, which has always been deemed the gold standard. How could automated or clustering-based methods be reliable any way if they are not comparable to each other, and often don't separate biologically distinct populations? The authors own data and the design of their software (to include many different clustering approaches) actually argue against their own point.

* In order to address the reviewer’s concerns with the phrasing of the sentence, we have rephrased it to say, ‘manual gating **can** fail to generate reproducible results.’ This being said, the publication that was referenced that demonstrated this problem was on a very low dimensional flow cytometry measurement where the result should not be variable (measurement of quantity of antigen-specific T-cells based on low number of markers). We used this to highlight the problems with manual gating on a simple problem. The author is correct, we do cite that the clustering process in itself can yield different results when using various algorithms. However, this is stated in context of the type of data ExCYT is designed to analyze (high-dimensional flow cytometry geared at 20+ channels like the BD FACSymphony and CyTOF) where many functional markers can allow different clustering solutions depending on the algorithm employed. The intention of this statement in the manuscript is to highlight the problems with manual gating even in a simple setting and the need for more objective methods of segmenting flow cytometry data.

6). Introduction: the article referenced to support the claim that gating introduced variation is old. Closer analysis of this data, using a more modern interpretation, is that the instruments could not be sufficiently standardized to allow consistent manual gating. Modern automated gating or clustering approaches - which were not tested in that article - would actually fare worse.

* As previously mentioned in the prior question, we use this article to highlight the problem with manual gating approaches and the need to move towards more automated workflows (as other manuscripts on this topic have).

7). Introduction: The authors mention gene-expression studies as a model for a new generation of cytometry data analysis tools. They fail to recognize that most of the experience with those tools has been on bulk approaches rather than single cell technology. This is a fundamental difference... tools generated for bulk analysis ignore the added power and complications of single cell analysis. In fact, single cell gene expression studies rely on approaches developed for cytometry data originally. There is a fundamental lack of understanding of the landscape and history in the rationale they present for their tool.

* Respectfully, we are acutely aware of this distinction between single cell data and data derived from homogenates of cells. I think it is important to understand that these algorithms, while initially applied to bulk gene expression data, are equally applicable in the single-cell scenario with a proper understanding of the algorithms. In the case of the manuscript, t-SNE and clustering methods are initially used at the single-cell level to abstract the data into higher-concept clusters, on which traditional heatmaps are then applied to understand the cluster to cluster relationships. This is in fact a workflow adopted by many manuscripts, including the one used as the demonstration for ExCYT.

8). Introduction: Authors claim that high dimensional bioinformatic approaches for flow cytometry data remain largely unused. This is an overstatement... I could probably come up with a list of 20 papers doing this.

* While there are many high-impact manuscripts that utilize these advanced techniques, there are thousands of more papers that use simple gating strategies and delineate these in order to quantify and track populations of cells between samples. Furthermore, day to day analysis within the lab environment typically does not consist of using these more advanced algorithms to analyze data. In fact, these types of analyses are usually conducted by a bioinformatics team when a publication is being assembled and not part of a routine data analysis pipeline done on a daily basis. Our hope is that ExCYT is easily enough used that it could become a tool for daily and easy to use to explore data right off the cytometer.

9). Introduction: The authors reference a paper from 18 years ago (!) to explain the reasons why high dimensional approaches are supposedly under-used in the literature. 18-years ago was ages ago... the most advanced cytometry labs were still working in no more than 10-12 color space, not 30-color space as they are now. Flow cytometry was the only single cell technology back then, not as now where it is one of dozens.

* This particular reference, while indeed is an old one is not just commenting on the use of flow cytometry but rather attempting to understand the acceptance of new technologies. The concepts this particular manuscript is describing are generalizable to understanding the hurdles in disseminating new technologies.

10). Introduction: Interestingly, the authors cite flowjo but do not include in the table comparing exCYT to other tools. I suspect this is because many of the features of exCYT are available in FlowJo and better implemented.

* In fact, we did include Flowmeans (a flowjo add-on) which is a limited cluster package for FlowJo relative to what we present in the manuscript. We would interpret the above comments as referring to this particular FlowJo add-on as the base FlowJo is unlikely to be adequately geared to handle very high dimensional flow data. Furthermore, our tool was developed because while we have licenses for FlowJo in our lab, it proved inadequate to do the depth of the analysis we needed, and therefore, developed ExCYT first for ourselves and then considered that it might be of interest to others. We noted and compared to FCSexpress that had a similar profile of add-ons with respect to SPADE, t-SNE, PCA, etc. Again, the salient distinction in our tool is the ability to explore, characterize, and visualize the clusters, not just perform the clustering step.

11). Protocol: How does the software handle errors generally, what feedback is provided to users? I found this lacking in my use of the software.

* We accept this was the first version of the software and have since implemented and will continue to make more robust implementations while maintaining a current Github repository (as can be seen in the latest version of the software).

12). Protocol: The authors do not use the standardized method for naming detectors on the cytometer, so they cannot account for when people use different synonymous dyes across different experiments. This will complicate how compensation tubes are imported into the tool. I expect the software will have trouble across different types of cytometers or different experimental designs.

* We don’t explicitly name the detectors within the ExCYT package. They are imported from the name metadata within the FCS files.

13). Protocol: What is the method for down-sampling? If it is random, then wouldn't evaluation of low frequency populations by exCYT be hampered?

* While true, this is a universal issue with physical and computational down-sampling. We will consider more robust technique to re-introduce rare population in later versions of the software; however, we focused on clustering and cluster exploration, characterization, and visualization for this manuscript and the version of the tool associated with it.

14). Protocol: Only a single gate can be set for compensation tubes... if the population shifts or if different beads or cells are needed for compensation, compensation will be wrong.

* The single gate is set on the fsc/ssc gate, where we would not expect shifts if the samples were run at the same time on the cytometer.

15). Protocol: The authors assert that a threshold at 99th percentile is objective. How is this objective? It is actually consistent more than objective, and it is not really necessary because of the mathematics of compensation.

* We changed the wording to reflect that this threshold is consistent, rather than objective across samples. We do disagree however that it is necessary to set a consistent gate when selecting the positive population for compensation. It directly affects the computation of the compensation matrix as it can change the average value for the population which is used to measure the ‘bleed’ across channels.

16). Protocol: The times claimed for tsne analysis are really slow. FlowJo is much faster (on the same computer). I tested tsne time on exCYT and found it to be very slow, with poor feedback about progress.

* It is unclear what comparison in compute time is being made because of difference in gating and down-sampling using in FlowJo vs what ExCYT did for downstream t-SNE. We accept that the algorithm speed could be improved; we will consider for later version of the program; however, this is the standard t-SNE implementation within MATLAB which has not been reported as being “slow” by others colleagues that have been using it fairly routinely.

17). Protocol: There are a lot of parameters that customize tSNE that are not available in this implementation.

* We have added distance metric and perplexity to the user interface, with suggested default values.

18). Protocol: I have never seen anyone cluster tsne data, and I have reviewed many applications of tsne and algorithm development that includes tsne. Essentially the tsne process is already clustering the data. Populations should be easy to identify at this point - if tsne worked well and was the best tool for the dataset - simply by gating or overlaying markers. There is no need for additional clustering.

* Respectfully, t-SNE is a dimensionality reduction technique and not a clustering solution, typically used for visualization purposes. The notion above regarding "no need for additional clustering" is problematic. Again, we focus this tool on the exploration, characterization, and visualization of clustering solutions as derived from high dimensional representation of the data and lower dimension representations such as t-SNE.

19). Protocol: The number of choices for clustering is good, but they don't provide data showing performance of each on a multi parameter dataset. How do I know other clustering methods work and what their output looks like? Also, when I used software tsne graphic never showed up, nor did the other clustering tools give any output.

* Given what is known about clustering algorithms, especially as applied to single-cell high-dimensional data, there is little yet known about general optimal parameters that can be applied across all experiments. If so, we would not give the user options to use various clustering algorithms and thresholds for those algorithms. We believe the evaluation of a clustering solution needs to be informed by an understanding of the biology, and this is why the exploration of these clusters is vitally important to implement a clustering solution. This is why we have focused the attention of the algorithm to exploration, characterization, and visualization of the clustering solutions because we believe ultimately, the scientist doing the experiments needs to guide the analysis with their understanding of the experiment and the biology. As for the bug the reviewer is citing, it is unclear to us where this is happening. We have not experienced this issue. If the user is running ExCYT in MATLAB, we would like to see the command line output for when they attempt to cluster the data.

20). Figure with ExCYT pipeline: The initial pre-processing gate (A) makes no sense at all. It appears to be "selecting" almost all of the events, without regard to any underlying distribution, which can't even be seen in the histogram.

* This figure is largely just meant to be a schematic; interestingly it does highlight our perceived difficulties with gating in high dimensional data. In our hands, we use a liberal gating strategy at first in the FSC/SSC domain, and let the dimensionality reduction and clustering algorithms remove noise, if it is present.

21). Figure with ExCYT pipeline: I don't see good separation in Tsne between CD4 and CD8 cells which all appear on the same "island" in the visualization. Without information on expression of each cluster (which is not visible at this resolution) I can't judge accuracy of clusters. Notably though, some clusters are dispersed widely in different parts of Tsne plot (like the purple one). So, they are clusters by clustering algorithms but not by tsne. which is right? I think you need a dataset with more clearly defined cell types to first validate this software, then test it on more complex datasets. For informatics experts like the authors, I am very disappointed to see a lack of attention to validation.

* The example the reviewer is citing is in regards to a CD3+ population of T-cells. Given the fact that this is whole tumor data, in regards to the heterogenous nature of the tumor (myeloid, tumor, stroma, and lymphocyte cells), the algorithm will naturally create these large islands first by major cell type. This is seen commonly in many publications, including the one we pulled this data from. At the risk of detracting from the novelty in this manuscript, we are not really developing a new algorithm, and therefore do not see the need to approach the manuscript with that rigor. Instead we are developing a tool to aid in the exploration, characterization, and visualization of high-dimensional cytometry with currently established techniques in a reasonable UI. As for the validation data set, we felt this was one of the larger datasets we could find that had a high-impact manuscript to complement it with vetted conclusions. As shown in the manuscript, we validated the use of ExCYT in coming to these same conclusions.

22). High dimensional flow/box plots provide little information, nor do the authors provide any guidance on why they are valuable or how they should be interpreted.

* Respectfully, various non-informatics users of the same tools noted above by the reviewer at our institution have them informative and have used them to create figures for presentations and manuscripts currently in review.

**Reviewer #3:**  
Manuscript Summary:  
This articles describes ExCYT, a toolbox written in Matlab with a GUI to analyze flow cytometry data. Many cluster algorithms are provided.  
The manuscript is well written and will likely provide some useful tools for biologists.  
  
Major Concerns:  
None  
  
Minor Concerns:  
- Not all discussed cluster methods have references. Perhaps provide a reference to the methods and an application in FC for each?  
- Only a limited number of data transformation methods are implemented (linear, log10, arcsinh). Some other commonly-used transformations are discussed in "Optimizing transformations for automated, high throughput analysis of flow cytometry data", i.e. linlog, gen arcsinh, biexponential, gen Box-Cox. I would be useful to add these as well, with hyperparameters to be selected by the user.  
- Likewise, only the Euclidean distance is implemented for clustering. Some others might work better, such as the squared Euclidean, Mahalanobis, Manhattan etc. These should ideally also be provided.

We appreciate the suggestions from this reviewer. While our goal is to maintain simplicity of the program for non-technical users, we have added some more functionality by allowing the user to specify the distance metric and perplexity of the t-SNE analysis. In terms of the provided transformations, we have selected those that are commonly found and used by biologists in programs such as FlowJo. Our goal was to create an environment with some of the familiarities from FlowJo with the capabilities of more sophisticated tools.   
  
  
**Reviewer #4:**  
The authors provide a graphic user interface for exploring multi-parameter cytometry datasets and apply a large variety of clustering algorithms. Using the interface does not require any knowledge in programming, making the powerful dimensionality reduction and clustering techniques available to a broad spectrum of researchers. With the continuing increase of high-dimensional cytometry the open access "ExCyt"-Software will prove useful in the visualization and discovery of underlying cell population structures.  
Overall the manuscript is well-written, clearly detailing the use of the program in a step-wise protocol, thereby, enabling new users to quickly understand the data analysis process. A special focus is put on introducing short introduction of the implemented clustering algorithms that are likely unfamiliar to the intended users of "ExCyt".  
In the second part of the manuscript the authors employ their software to re-analyze a CyTOF dataset and convincingly recreate the published results, highlighting the usability of "ExCyt" in the analysis of real-world high-dimensional cytometry data.  
The ExCyt GUI presents itself clearly organized and user-friendly. It allows the straightforward creation of complex figures, especially the novel high-dimensional flow and box plots provide a powerful visualization of differences between clusters. The manual refinement of automatically generated clusters by manual gating is a valuable feature for users accustomed to the traditional analysis of cytometric data.  
  
Major:  
Bugs:  
-Canceling manual tSNE-gating erases graph

* Unfortunately, we cannot replicate this bug.

-Inappropriately named single stain samples can be selected without warning

* We included further error-checking functionality in the importation of single-stain files. The program now checks that the user selected fcs files, then checks for an ‘Unstained.fcs’ file, and then finally checks to make sure the single-stain file names are in concordance with the panel in the main file for analysis.

-Non-integer/non-numeric arguments are allowed as input for clustering arguments (e.g. # of clusters = 1.2)

* We have fixed this and now the program throws back an error to the user if a non-integer value is given for a clustering algorithm that requires an integer input.

Suggestions:  
-Add color legend for clusters  
-Improve documentation of auto-compensation  
oNaming Single Stain samples by Fluorophore does not work  
oName has to match $PnN Keyword in .FCS header  
-Editing of parameter names after data import  
-Allow users to set tSNE parameters (Perplexity, Iterations, …)  
-Display automatically calculated compensation matrix for user evalution  
-Indicator for status of auto-compensation  
-Add warning for unsaved workspace when exiting  
-Add Import of .lmd files  
Minor:  
-The sentence in line 261-263 needs clarification. Furthermore in line 151 "speculations" should be changed to "specifications".

* We thank the reviewer for suggestions. We have given the user the ability to set tSNE parameters as well corrected the typo found by the reviewer. We will work on including more of these features in future releases.