

Reviewer's report

Title: Flow: Statistics, visualization and informatics for flow cytometry

Version: 1 **Date:** 10 May 2008

Reviewer number: 1

Reviewer's report:

Major Compulsory Revisions:

- 1) The authors frequently use the terms "clustering" and "classification" in different contexts. For the mixture modeling and KDE approaches the term clustering is appropriate. As far as I can tell, classification in the machine learning sense is not performed at all (this would involve some form of supervised learning). A term better suited for assignment of OBO terms to subsets of the data might be "annotation".
- 2) In flow cytometry data analysis, the term "filtering" is used for the process of applying gates on the data. In this manuscript, filtering describes a potentially random subsetting operations. The authors should try to disambiguate the term.
- 3) The manuscript does not address how to treat groups of flow measurements in a complex experiment. There doesn't seem to be a concept of sample annotation or coordinated analysis of common samples within a sample group, or specific data structures do deal with that. Does the hierarchical tree structure allow for sample groups? And can meta data be edited?. Since annotation and keeping track of experimental meta data is one crucial point in high throughput experiments, the authors should point out how this can be handled within their software.
- 4) The tree doesn't make use of inheritance, that is, features are only available at a certain level in the tree and not for all downstream leaves. Is this a design decision or a shortcoming of the underlying data structure.
- 5) I was not able to make the projection feature work on my

installation. Neither could I install the software from the source code provided in the supplementary materials. When creating 3D plots for transformed data, the software doesn't seem to pick up the scale changes as all transformed channels are stuck on the axes.

6) The separation between basic features and plug in functionality is not very clear. E.g., adding ellipse outlines to a plot only makes sense for normal or at least symmetric mixture components. While the focus is obviously on automated gating, there still seems to be a need for the classical flow data analysis tools. The very rudimentary tools provided with the base annotation might not be sufficient.

Minor Essential Revisions:

1) Page 2, first paragraph: "The basic ideas is that if a subset..." This might be true in theory, however in practise many gate selections rely either on expert knowledge by the investigator or information that is borrowed from other samples like negative or positive controls. Either way, the selection is not exclusively based on the features of the underlying data.

2) Page 6, second paragraph: "All the clustering routines..." delete one of the two "alls"

3)Page 9, fourth paragraph: "One limitation of flow is..." The authors should mention the available markup languages here, which are the recommended way to interface between programs on a file level. The FCS format is strictly limited to the raw measurement data and mostly instrument specific meta data.

Discretionary Revisions:

1) The installation process is very tedious since there are many dependencies on other software. The documentation on the project web page is helpful, but providing binaries would be crucial for the intended target audience to make use of the software. A debian meta package containing all dependencies would speed up the installation process for Linux users.

2) QA/QC is not mentioned at all in the manuscript, but seems to be a

major deal for high throughput data analysis. The authors' views on how to implement that using their software would be helpful.

3) There seems to be an issue with the security certificate of the project web page. The latest Firefox 3.0 did not display the page before setting a security check exception.

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.