



Published in final edited form as:

Cytometry A. 2012 May ; 81(5): 353–356. doi:10.1002/cyto.a.22037.

FCSTrans: an open source software system for FCS file conversion and data transformation*

Yu Qian^{#1,2}, Yue Liu^{#3}, John Campbell³, Elizabeth Thomson³, Y. Megan Kong², and Richard H. Scheuermann^{1,2,*}

¹Division of Biomedical Informatics, Department of Clinical Sciences, University of Texas Southwestern Medical Center at Dallas, Dallas, TX 75390

²Department of Pathology, University of Texas Southwestern Medical Center at Dallas, Dallas, TX 75390

³Health Information Systems, Northrop Grumman, Inc., Rockville, MD 20850

These authors contributed equally to this work.

In flow cytometry (FCM) experiments, investigators usually rely on instrument manufacturers and “black box” commercial software to transform cellular marker expressions into cell populations on 2D dot plots. Techniques behind these systems and their limitations have not been sufficiently addressed or disclosed. Currently, an FCS file can be FCS2.0 or FCS3.0 (N.B. Newer standards like FCS3.1 and ACS1.0 have been proposed and might be generated by some manufacturers in the future.). According to the Becton Dickinson (BD, <http://www.bd.com>) acquisition software manual [1], by default FCS2.0 fluorescence data are log-transformed while FCS3.0 files keep the raw outputs from the instrument in a linear mode. Therefore FCS3.0 format provides more control to bioinformaticians and FCM software developers in data processing (e.g., changing data compensation after acquisition). However, it remains unclear how the linear-mode fluorescence data in FCS3.0 files should be transformed before events can be plotted for population identification.

Figure 1A–C shows a comparison on FCS3.0 data conversion and transformation we conducted between FlowJo (Tree Star Inc., <http://www.flowjo.com>), flowTrans [2] of Bioconductor, and FCS2CSV [3]¹ on three FCS3.0 files collected from different labs (details of FCS files and software systems can be found in Supplementary File 1; headers of the three BD FCS file can be found in Supplementary File 2). It seems that under the default setting FCS2CSV and flowTrans² seemed to generate insufficient data modality for population segregation in these FCS files. Also the conversion techniques (and transformation parameters) used in the three systems seem to be inconsistent, reflected by the output data range and the data distribution characteristics. These issues have brought in uncertainty not only in FCM data analysis but also in data sharing and interpretation, and

*This work is supported by NIH N01AI40076.

*correspondence should be sent to: richard.scheuermann@utsouthwestern.edu Address: 5323 Harry Hines Blvd., Dallas, Texas, USA 75390-9072 Telephone: +1 214 648 4115 Fax: +1 214 648 4070.

The authors do not have a conflict of interest to declare.

¹We have also tested flowCore [4], FCSEExtract [5], and LLData [6]: results of flowCore and FCSEExtract are similar to those of FCS2CSV, while LLData could not read the FCS3.0 files we have.

²There are four options in flowTrans: arcsinh, biexponential, linlog, and Box-Cox. We chose the results of arcsinh because it generated relatively better segregation of populations than the other three options, whose results can be found in Supplementary Figure 1.

therefore made us believe it was necessary to develop a single open source software system that would transform different types of FCS files appropriately, robustly, and consistently.

Here we report the development of *FCSTrans*, an open source FCS file converter and transformation system that generates numeric data matrix into .txt files from binary FCS files. *FCSTrans* is written in R. Its source code and technical report can be found at <http://importflock.sourceforge.net>. Results of *FCSTrans* on the three BD FCS files, as shown in Figure 1D, are highly consistent with those of FlowJo in Figure 1A. It supports both BD and Accuri (<http://www accuricytometers.com>) FCS files. Important technical details of *FCSTrans* including method description, transformation equations, and identification of transformation parameters can be found in Supplementary File 1. It has also been implemented in the FCM data analysis pipeline of the Immunology Database and Analysis Portal (ImmPort, <http://import.niaid.nih.gov>).

We have compared the results of *FCSTrans* with those of FlowJo and flowCore [4] of Bioconductor on the three most commonly used transformation methods including linear, logarithmic, and logicle transformation. The advantages of *FCSTrans*, when compared with FlowJo, are focused on processing negative inputs, supporting 24-bit data, and being open source. In order to study the behavior of different transformation methods, we have developed an FCS data simulator to write values uniformly selected from 2^{-32} to 2^{32} into binary FCS format. Our full range input simulation has identified that FlowJo linear transformation converted negative inputs into 4095 and reset values larger than $2^{18} = 262144$ to start from zero (data plot can be found in Supplementary Figure 2A). This can lead to problematic transformation for 24-bit data used by Accuri cytometers or when scatter parameters have negative values. In contrast, *FCSTrans* linear transformation converts negative inputs to 0 and values larger than 262144 to 4095 (data plot can be found in Supplementary Figure 2B). Both logarithmic and logicle transformation of *FCSTrans* generate essentially the same output as FlowJo (method details can be found in Supplementary File 1), as Figure 1 has shown. Logicle transformation in *FCSTrans* segregates populations better than the default logicle setting in flowCore in our experiments with real FCS data, as Figure 1 has shown. The full range input data simulation also supports our conclusion when comparing results from default parameters used in the logicle transformation, as in flowCore¹ and in *FCSTrans*. While the two sets of parameters perform similarly for large input values (Supplementary Figure 2C), the default parameters used in flowCore seem to be problematic, which does not preserve a linear-like relationship for small values (Supplementary Figure 2D). The effect of using different w , a critical parameter setting in the logicle transformation can be found in Supplementary Figure 3.

Data compensation needs to be done before transformation can be performed. *FCSTrans* is able to look for the keyword \$COMP (FCS3.0 standard), SPILL (BD instruments), or SPILLOVER (Accuri instruments) to retrieve the compensation matrix from the FCS3.0 file header, and automatically applies it to compensate the FCS3.0 data before further transformation. However, there is no standard indicator on whether an FCS file has been compensated or not. Therefore *FCSTrans* provides the automated compensation function as an option. The version we have deployed at ImmPort assumes the submitted FCS files have been compensated and does not compensate them automatically, while the version used in this paper (also released at <http://importflock.sourceforge.net>) automatically applies the spillover matrix, if found in BD and Accuri FCS files, to compensate the data before applying the transformation, which seems convenient for most BD FCS3.0 files we have tested.

¹using R command line and the default parameters specified in [7]: logicleTransform (transformationId = "default-LogicleTransform", w = 0, t = 262144, m = 4.5, a=0)

Data compensation may generate negative values. *FCSTrans* follows the traditional -111 cutoff used in FlowJo and flowCore, so that dot plots from different samples and from different software platforms can be directly compared. Future work is to allow users to change the cut-off when necessary. For example, when the cut-off is too high, a large number of events will be truncated to zero and pile up on the axis. Decreasing the cut-off is necessary for disclosing the expression patterns in the negative area.

While the transformation methods in *FCSTrans* are general, we have identified different transformation parameters for different instrument manufacturers based on their data characteristics (e.g., number of bits in data representation). The current implementation of *FCSTrans* automatically supports FCS files from BD and Accuri Cytometers. Details on supporting different manufacturer files can be found in Supplementary File 1 and our technical report online (<http://importflock.sourceforge.net>). Our experiments (dot plots of one example Accuri FCS file can be found in Supplementary Figure 4) have shown that results generated by *FCSTrans* are highly consistent with those from Accuri CFlow software [7].

In summary, based on the study on the behavior of different transformation methods with both simulated and real FCS files, we have developed an open source software system *FCSTrans* that can convert and transform FCS files from BD and Accuri cytometers. Compared with existing systems, *FCSTrans* has: a) avoided the linear transformation limitation on negative and 18-bit data in FlowJo; b) identified a set of logicle transformation parameters for effective population segregation consistent with FlowJo; 3) been open source and free; 4) supported both BD and Accuri FCS files within one single system. Results of *FCSTrans* can be used in better segregating cell populations and consistent cross-sample comparison and data sharing with commercial software platforms and different parties. We hope that *FCSTrans* can help remove the preprocessing obstacle of FCS file conversion and data transformation, and provide a starting point for independent data analysts, statisticians, and software developers to develop advanced and customized FCM data analysis and visualization software.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We sincerely appreciate Josef Spidlen and Ryan Brinkman for providing FCS2CSV, and the useful discussion with Florian Hahne on flowCore. We also sincerely thank Chungwen Wei and Iñaki Sanz Lab (University of Rochester), Adam Seegmiller and Nitin Karandikar Lab (University of Texas Southwestern Medical Center), Lisa Beck Lab (University of Rochester), and Doris Wiener (University of South Florida) for providing different FCS files for us to study data transformation and test software systems.

LITERATURE CITED

- [1]. BD FACS Diva Software 6.0 Reference Manual. Beckton and Dickinson:
- [2]. Greg, Finak; Juan-Manuel, Perez; Andrew, Weng; Raphael, Gottardo. Optimizing transformations for automated, high throughput analysis of flow cytometry data. *BMC Bioinformatics*. 2010; 11:546. [PubMed: 21050468]
- [3]. <http://sourceforge.net/projects/flowcyt/files/GenePatternFlowCytometrySuite/FCS2CSV/>, last visited on February 02, 2012
- [4]. <http://bioconductor.org/packages/2.6/bioc/manuals/flowCore/man/flowCore.pdf>, last visited on February 02, 2012

- [5]. <http://research.stowers-institute.org/efg/ScientificSoftware/Utility/FCSExtract/>, last visited on February 02, 2012
- [6]. <http://www.cyto.purdue.edu/archive/flowcyt/software/DATA/PURDUE/LLDATA.DOC>, last visited on February 02, 2012
- [7]. [http://accuricytometers.com/files/Accuri Revolutionizes Flow Cytometry.pdf](http://accuricytometers.com/files/Accuri%20Revolutionizes%20Flow%20Cytometry.pdf), last visited on February 02, 2012

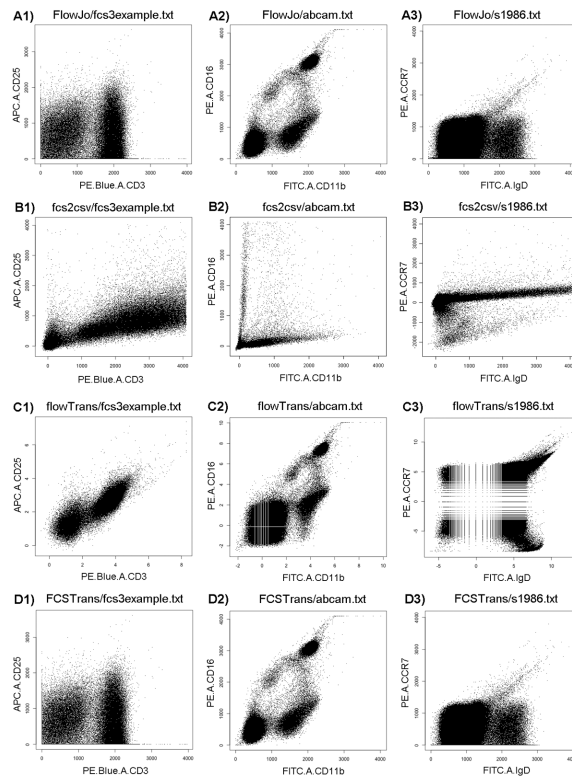


Figure 1.

2D plots of three BD FCS3.0 files after being transformed by four software systems. Each row (Rows A, B, C, and D, from top to bottom) corresponds to one method: A) FlowJo (version 8.8.6, MacOSX); B) FCS2CSV (for comparison with the other two methods, values larger than 4096 not plotted); C) FlowTrans (ArcSinh transformation option used); D) FCSTrans. Each column (Columns 1, 2, and 3, from left to right) uses one FCS3.0 file: 1) CD3 vs CD25 of FCS3example.fcs; 2) CD11b vs CD16 of abcam.fcs; 3) IgD vs CCR7 of s1986.fcs. Detailed information about the software and the files used can be found in Supplementary File 1.