Opencyto: An Innovative Framework for Flow-Cytometry Data Analysis

October 20, 2014

L Pattacini

Flow cytometry is a tool increasingly used in large clinical research studies. While the technological advancement in the instrumentation has improved rapidly, the analytical tools are still time-consuming and require subjective decisions. Although different platforms have been developed, they have failed to attract widespread use, due to the lack of confidence in automatically generated results and a perceived difficulty in learning the use of such tools. A recent paper published in *PLOS Computational Biology by* Drs. Greg Finak and Raphael Gottardo in the Vaccine and Infectious Disease Division and their collaborators describes OpenCyto, a new infrastructure for flow cytometry data analyses.

The framework of OpenCyto is a collection of integrated open-source R/BioConductor packages that allows processing of large datasets, importing gates that identify specific cell populations from other programs, and easily sharing data between computational and non-computational researchers.

In addition to describing the software, the authors also show the overlap between manual and automated analyses. Dr Finak and coworkers utilized the data from the HIV vaccine trial HVTN080 to compare the results obtained by the automated OpenCyto framework and by manual gating using FlowJo, a software widely used for flow cytometry analyses. They showed that the results obtained by manual and computationally-generated gates were very similar and shared a common hierarchical structure. Furthermore, they looked at the HIV-specific responses generated by the vaccine, and observed that the results obtained by automatic and manual analyses strongly correlated. Very importantly, OpenCyto results were useful in gating markers with high variability that could not be manually analyzed.

"The tool could be a 'game-changer' for large trials" said Dr. Finak. "When assays are highly standardized, the tool allows rapid, robust, and reproducible analysis of flow data, thereby making data more comparable over time and between different data generating centers. The time-savings allow experimentalists to focus their attention on more important matters than drawing boxes around cell subsets. As cytometry data becomes increasingly high-dimensional through the inclusion of more markers and new technologies like CyTOF, automated tools will become indispensable since manual analysis of large trials will be an untenable analysis paradigm."

"We are gaining users in centers across the US," adds Dr. Finak. "Common use will come as early adopters demonstrate its utility through publications." Those that use flow cytometry enough to know how the analytical part is time-consuming and subjective, will likely wish that this useful tool will soon become widely utilized.

Finak G, Frelinger J, Jiang W, Newell EW, Ramey J, Davis MM, Kalams SA, De Rosa SC, Gottardo R. 2014. OpenCyto: An open source infrastructure for scalable, robust, reproducible, and automated, end-to-end flow cytometry data analysis. *PLoS Comput Biol*. 10(8):e1003806.
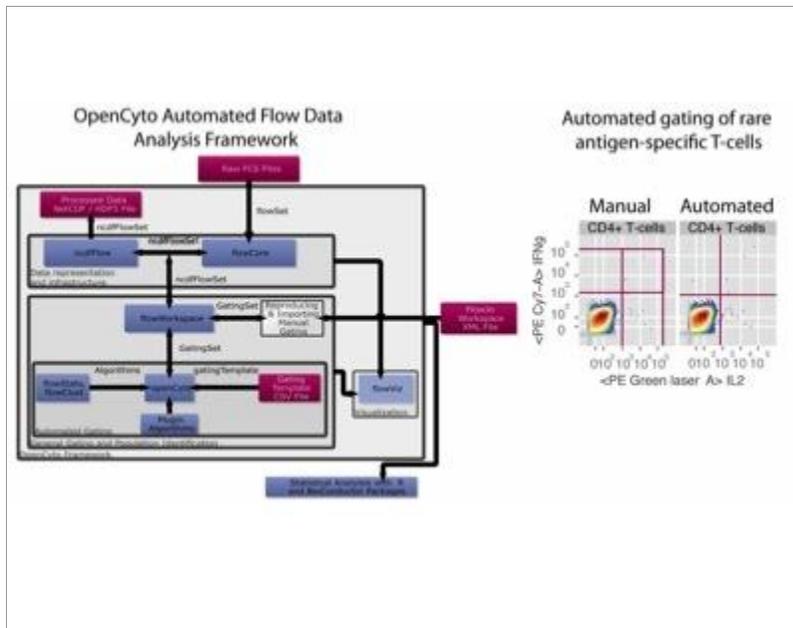
See also: www.opencyto.org

Image provided by Dr. Greg Finak

On the left: Overview of the OpenCyto infrastructure. Data (red boxes), software packages (blue boxes), framework functionality (gray boxes) and data flow/data structures (arrows) are represented. On the right: comparison between a manual and automated flow plot of a rare population.