High throughput flow cytometry analysis with Bioconductor

Nolwenn LeMeur[1,2,*], Florian Hahne[1], Deepayan Sarkar[1], Byron Ellis[3], Josef Spidlen[4], Ryan R. Brinkman[4], Robert Gentleman[1]

1. Life Sciences Department, Computational Biology Program, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. N, M2-B876, PO Box 19024, Seattle, Washington 98109-1024, USA
2. EA SeRAIC INSERM, IRISA - Symbiose, Campus Beaulieu, Université de Rennes I, 35042 Rennes Cedex, France
3. AdBrite Inc., 731 Market St., 5th Floor, San Francisco, California 94103, USA
4. Terry Fox Laboratory, British Columbia Cancer Agency Research Centre,675 West 10th Avenue, Vancouver, BC V5Z 1L3 Canada, Canada

* Contact author: nlemeur@irisa.fr

Keywords: flow cytometry, high throughput, data structure, data analysis

Recent advances in automation technologies have enabled the use of flow cytometry high content screening (FH-HCS), in both basic and clinical research, generating large complex data sets with many covariates. However, data management and data analysis methods have not yet progressed sufficiently far from the initial small-scale studies to support modeling in the presence of multiple covariates.

To those aims, we developed a set of computational tools in the R package flowCore to facilitate the analysis of these complex data. We propose R data structures to handle flow cytometry data through the main steps of importing, storing, assessing and preprocessing data from flow cytometry experiments. For example, this package provides facilities for compensation, transformation and filtering preprocessing steps. A key component of the flowCore package is to have suitable data structures that support the application of similar operations to a collection of samples or a clinical cohort. In addition, our software constitutes a shared and extensible research platform that enables collaboration between bioinformaticians, computer scientists, statisticians and biologists.

The software has been used in the analysis of various data sets and its data structures have proven to be highly efficient in capturing and organizing the analytic work flow. Finally, a number of additional Bioconductor packages successfully build on the infrastructure provided by flowCore, offers new opportunities for flow data analysis.