FlowCAP - Flow Cytometry: Critical Assessment of Population Identification Methods

Flow cytometry has been widely used by immunologists and cancer biologists for more than 30 years as a biomedical research tool to distinguish different cell types in mixed populations based on the expression of cellular markers. It has also become a widely used diagnostic tool for clinicians to identify abnormal cell populations associated with disease. In the last decade, advances in instrumentation and reagent technologies have enabled simultaneous single-cell measurement of tens of surface and intracellular markers, as well as tens of signaling molecules, positioning flow cytometry to play an even bigger role in medicine and systems biology. However, the rapid expansion of flow cytometry applications has outpaced the functionality of traditional analysis tools used to interpret flow cytometry data such that scientists are faced with the daunting prospect of manually identifying interesting cell populations in 20 dimensional data from a collection of millions of cells. For these reasons a reliable automated approach to flow cytometric analysis is desirable. While there has been a growing interest among the scientific community in developing these methods, guidance for end users about appropriate use and application of these methods is difficult to come by.

In response to this need, we are pleased to announce the **Flow Cytometry: Critical Assessment of Population Identification Methods (FlowCAP)** project. The goal of FlowCAP is to advance the development of computational methods for the identification of cell populations of interest in flow cytometry data. FlowCAP will provide the means to objectively test these methods, primarily by comparison to manual analysis by experts using common shared datasets, and secondly by comparison to synthetic data sets with known properties.

FlowCAP will consist of three parts:

1) The collection of de-identified data sets for prediction from the experimental community that will be shared among the algorithm development community as a common reference for analysis;
2) The collection of population subset predictions (gates) from the computational biology community derived from these common reference data sets using existing and novel algorithmic approaches; and
3) The assessment and discussion of the results in comparison with the manual gating gold standard.

We plan to hold a summit of participants within one year to present and discuss the results and future directions in this area.

The flow cytometry community is hereby invited to provide de-identified data representing typical use cases in flow cytometry for both diagnosis and discovery types of applications. Data sets that represent both important scientific use cases and technical challenges are sought.

Datasets requested include, but are not limited to:

Scientific use cases

- Detection of rare cell populations (*e.g.,* minimal residual disease in cancer);
- Enumeration of large numbers of distinct cell populations in high dimensional flow cytometry data (*e.g.,* >10 colors);
- Discovery of clinically relevant cell populations in patient cohorts associated with disease states (e.g. markers of autoimmune disease, survival indicators in lymphoma);
- Measurement of DNA quantities (*e.g.,* Flow-FISH);
- Diagnostic panel assessments (*e.g.,* CD4/CD8 counts, chronic lymphoma/leukemia).
- Kinetic analysis of changes in cell population proportions (*e.g.,* immune response to infection or vaccination)
- Kinetic analysis of marker expression levels (*e.g.,* up-regulation of activation markers, changes in the phosphorylation of signaling proteins

Technical challenges

- Specimens with excessive cell debris
- Samples with slight differences in marker staining intensities due to sample processing differences
- Segregation of overlapping cell populations

Data sets should also be accompanied by metadata descriptions about the specimens and staining procedures used compliant with the MIFlowCyt data standard. Be aware that by providing these data sets you are agreeing that these data will be freely distributed to any interested party and so should not contain any proprietary or sensitive information.

Those able to provide data are asked to contact the organizing committee at flowcap@flowsite.org. We also encourage developers of computational techniques to contact the organizing committee to register their interest in participation.

We believe that this effort will be of benefit to the whole flow community, and thank you for your interest and participation.

FlowCAP Organizing Committee

Ryan Brinkman, British Columbia Cancer Agency
Robert Gentleman, Fred Hutchison Cancer Research Center
Raphael Gottardo, Clinical Research Institute of Montreal;
Richard H. Scheuermann, University of Texas Southwestern Medical Center
Jill Schoenfeld, TreeStar Inc.