

Needle in a haystack: rare cell subtypes in flow cytometry

Ioanna Manolopoulou, Cliburn Chan and Mike West

Flow cytometry is a powerful technology for identifying and characterizing microscopic particles flowing within a stream of fluid. It uses the principles of light scattering and emission of fluorochrome molecules to provide measurements from particle (often cells) characteristics (see Figure 1).

Cytometry allows various types of information about the physical and chemical structure of the cells to be extracted, using beams of several different wavelengths. Until recently, cells were classified according to a few ‘gating’ markers that are associated with specific cell subtypes, leading to a high degree of subjectivity. In a recent paper, Chan et. al.¹ employed flexible Gaussian mixture models in order to classify cells based on all markers simultaneously.

In flow cytometry, a vast number of observations can be generated in each experiment, and unfortunately, the computational cost of Gaussian mixture models becomes prohibitively high for very large data sets. In our particular application, which arises in immunology, we are specifically interested in identifying and characterizing certain very rare cell subtypes pre- and post-vaccination within a vast number of cells.

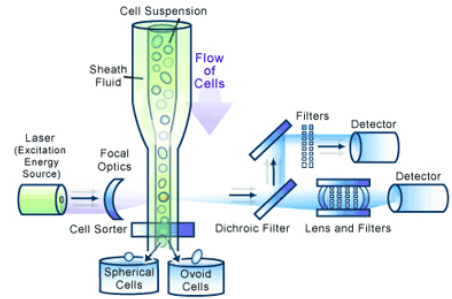


Figure 1: *Generic setup of cytometry experiment (figure courtesy of Science Creative Quarterly, www.scq.ubc.ca, by Jane Wang).*

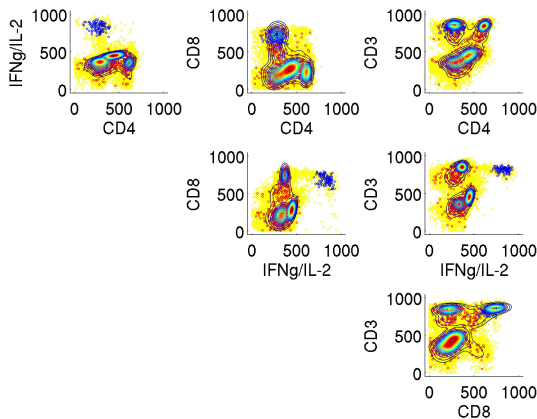


Figure 2: *Pairwise plot of four cell markers: CD3, CD4, CD8 and IFNg/IL-2. Yellow datapoints represent the whole data, red a random sample and blue the ‘gated’ subset.*

The ‘big data’ problem is usually addressed by using computationally manageable subsets of the data. Here, in fact the cell subtypes of interest are very rare, implying that a random subset of the data may not contain enough of those rare cells in order to draw reliable inferences. We develop a targeted re-sampling approach using Sequential Monte Carlo, so that the subset of the data used contains information specifically relevant to the rare cell subtypes. More specifically, using an initial random subset of the data we obtain an estimate about ‘gating’ the remaining observations into the various cell subtypes, and augment our subset with observations fitting the rare cell subtypes well. Our method provides fast and reliable inferences without any additional assumptions to the model.

A variety of similar datasets of clinical interest would greatly benefit from our method, and this is increasingly apparent as experimental technologies advance. We implement our procedure on an assessment of the immune response to cytomegalovirus (CMV)-peptide stimulation using 4-color flow cytometry, which is part of the Intracellular Cytokine Staining Quality Assurance Program (ICS QAP) proficiency panel for laboratories who routinely measure vaccine-induced antigen-specific responses in support of HIV/AIDS clinical trials.

¹C. Chan, F. Feng, J. Ottinger, D. Foster, M. West, T. Kepler Statistical mixture modeling for cell subtype identification in flow cytometry. *Cytometry Part A*, 73:693–701, 2008.