

Abstract

In blood alone, researchers have identified hundreds of phenotypically and functionally distinct lymphocyte subsets. Precise characterization of the function and role of these subsets plays a critical role in our understanding of the mechanisms of immunological disease and protection in diverse areas of research, from transplantations and cancers to diagnostics and stem cells.

Fluorescence Activated Cell Sorting (FACS) is one important technology that facilitates investigation of the immune response at the single-cell level. Up to 17 fluorescently labeled markers on the surface and inside a single cell can be used simultaneously to precisely differentiate populations of cells and study their role in the immune response. In a typical study, cells exhibiting similar markers are grouped into populations, and the frequency is compared between a test and control sample. Central to this analysis is the identification and comparison of corresponding populations of cells across different samples.

While the cell populations share molecular markers, there is substantial biological variability in the measurable characteristics across different samples. As a result, the distribution of corresponding cell populations are, statistically speaking, significantly different; yet researchers routinely identify cell population correspondence manually and make comparisons accordingly.

In this project we propose the development of automated techniques for comparing Hi-D FACS data from multiple samples. Specifically, we aim to: develop robust algorithms to identify corresponding cell populations across multiple samples, provide support for annotating populations to an expressive terminology, and utilize the annotations to create a knowledge-driven approach to identifying cell population correspondence.