Novel Differential Flow Cytometry Data Analyses Method Using Data Nuggets Compression and Projection Pursuit Algorithms

Author: Davit Sargsyan

Biography

Davit Sargsyan is an associate director at the Translational Medicine and Early Development Statistics (TMEDS) group, Johnson & Johnson Innovative Medicine R&D. Davit received MS in Statistics and PhD in Pharmaceutical Sciences form Rutgers University. This project is a collaboration between J&J TMEDS, and Dr. Cabrera's and Dr. Kong's laboratories at Rutgers University.

Abstract

Multicolor flow cytometry is a popular laboratory technique that measures physical and chemical properties of individual cells or particles. One of flow cytometry applications is the immune cell phenotyping. Cell size, granularity and surface proteins allow for identification and separation of specific cell types from mixed immune cell populations. Standard analysis of flow cytometry data is done by plotting it in two dimensions at the time (i.e., against two markers) in specialized software such as *FlowJo* and drawing areas of interest called gates. Gating strategies are rooted in the current understanding of biological cell differentiation mechanisms but are subjective and do not produce optimal 2D projections of the multidimensional data.

In this work we present a novel method to conduct differential analysis of flow cytometry data. First, data is compressed using Data Nuggets algorithm (Beavers 2023)[1] on data rows that contain individual cell information. This reduces each sample from a table with hundreds of thousands of rows to just a few thousand nuggets that preserve most of the information about the data structure. Next, a projection pursuit algorithm is applied to the compressed data to find projections that maximize the differences between levels of experimental conditions such as treatment. For that, we introduce a new projection pursuit index similar to the Natural Hermite index that measures differences between distributions. A new factor rotation algorithm aligns the axis with a small number of variables in each dimension.

We applied out new method to a flow cytometry dataset containing samples of HIV-exposed but uninfected (HEU) and unexposed and uninfected (UE) infants. The blood samples from each infant were either left untreated (UT) or challenged with lipopolysaccharide (LPS) to induce strong immune response. The goal of this study was to test the ability of our algorithm to detect the differences between the immune reaction in the HEU infant samples versus the UE infant samples. If HIV exposure suppressed the immune response similarly to actual HIV infection, LPS challenge would produce significantly smaller differences compared to UT samples in HEU versus the UE samples. These results were compared to the results of a conventional flow cytometry analysis on the same data.

[1] Beavers, T.; Cabrera, J.; Chen G., Duan Y; Lubomirski; Tiggler, S, M. 2023. 'Data Nuggets: A Method for Reducing Big Data While Preserving Data Structure', IN REVISION.