Differential Projection Pursuit on Flow Cytometry Data to Identify and Profile Immune Cells Compromised by Exposure to HIV

Author: Davit Sargsyan

Email: [dsargsy@its.jnj.com](mailto:dsargsy@its.jnj.com)

Biography

Davit Sargsyan is an associate director at the Immunology Discovery Statistics 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 laboratory at Rutgers University.

Abstract

Differential projection pursuit (DPP) is a machine learning technique recently developed by our team for finding informative d-dimensional projections of p-dimensional data (d<<p) with complex structure across experimental design factors. Specifically, the method was developed for visualization and analysis of cell-level flow cytometry data obtained from samples exposed to different experimental conditions such as treatments or dosage regimens. Multicolor flow cytometry (FC) is a laboratory technique that allows density estimation of specific biomarkers (proteins) on or in each cell in a sample. It is widely used in immunology to phenotype the immune cells, assess their activation status and functions, and examine the cell composition changes and differences in the samples.

In 2009, Kollmann Lab at the University of British Columbia conducted a study to evaluate computational pipelines for identification of cell populations that can discriminate between HIV-exposed (mothers diagnosed with HIV) uninfected and unexposed uninfected infants, 44 infants in total. Human immunodeficiency viruses (HIV) are species of Lentivirus that attacks human immune system and can lead to acquired immunodeficiency syndrome (AIDS) causing immune system failure. The virus can be transmitted non-sexually from infected mothers to their children during pregnancy, childbirth or breastfeeding. The samples were analyzed via flow cytometry, and data posted online on a public flow cytometry repository site (<https://flowrepository.org>, Repository ID: FR-FCM-ZZZU).

To examine and visualize the data structure, we first compressed the data (40M rows x 8 protein markers) using Data Nuggets, then classified the resulting 4,000+ nuggets based on the proportions of HEU and UE cells and clustered them. DPP was then applied to create most informative 2D projections to show the data structure. We validated the projections by mapping cells in 2D clusters back to the 8D clusters (cell-level data). Additionally, we compared DPP results with results of two commonly used techniques – tSNE and UMAP. As a result, we identified and profiled groups of immune cells that were found to be disproportionately represented in HEU vs. UE. This approach allows to uncover the effect of a specific condition using flow cytometry data without relying on gating and producing more robust results.