Exploring HIV Exposure Impact on Immune Cell Composition and Activation Using Differential Projection Pursuit

Mahan Dastgiri1, Javier Cabrera1\*, Davit Sargsyan2,3,4, Ge Cheng1, Yajie Duan1, Jocelyn Sendecki4, Xuechen Wang4, Nicholas Hein4, Kanaka Tatikola4, Surya Mohanty4

1Department of Statistics, School of Arts and Sciences, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

2Department of Pharmaceutics, Ernest Mario School of Pharmacy, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

3Graduate Program in Pharmaceutical Science, Ernest Mario School of Pharmacy, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

4Janssen Pharmaceuticals, Johnson and Johnson, Spring House, PA, USA, and Beerse, BE

**\*Correspondence**

Professor Javier Cabrera

Rutgers, the State University of New Jersey

Dept of Statistics, Hill Center Room 471

110 Frelinghuysen Rd, Piscataway, NJ 08854

Phone: +1-848-445-7665

Email: cabrera@stat.rutgers.edu

# **Abstract**

XXX

**Keywords:** machine learning; differential projection pursuit; flow cytometry; HIV, immunology.

1 Introduction

Differential projection pursuit (DPP) is a statistical method 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. A flow cytometer contains an array of tubes through which cells travel one at the time. The cells are stained with fluorochrome-conjugated antibodies in advance. As the stained cells make their way through the tubes, they are hit with lasers which excite the fluorochromes. In response, fluorochromes begin to emit a narrow spectrum of visible light (e.g., green, red, and so on). The emitted light is then picked up by detectors, amplified and converted to from analog to digital before being stored as data files. The intensity of light detected from each cell in each spectrum is interpreted as the relative abundance of specific proteins on the surface of or inside that cell. In addition to proteins, flow cytometer also provides information about cell’s physical properties such as size and granularity via measurements like the forward and side light scattering (FSC and SSC, respectively) (Figure 1). The data is then stored in a special Flow Cytometry Standard (FCS) format that is maintained by the International Society for Advancement of Cytometry (add REF using ISAC website: <https://isac-net.org/page/Publications>, <https://currentprotocols.onlinelibrary.wiley.com/journal/19349300>, ) . Each FSC file contains meta data about the sample and the instrument settings as well as a 2-dimensional array, with rows representing individual cells, columns - the markers being measured including FSC, SSC and the fluorochromes, and the values corresponding to light intensity measured by the detectors. Since the number of cells in each sample and the total number of samples can be large, the combined data can contain measurements of millions to billions of cells. Additionally modern flow cytometers are capable of detecting up to 48 markers simultaneously. Because of its size of, it becomes crucial to compress the data before any kind of analysis. Our workflow includes a compression algorithm called Data Nuggets (REF) that reduces cell-level data to a few thousand representative data nuggets while preserving data structure as much as possible.

Human immunodeficiency viruses (HIV) are species of Lentivirus that attacks human immune system. If left unchecked, the infection can lead to acquired immunodeficiency syndrome (AIDS) causing immune system failure. The virus can be transmitted sexually or non-sexually. Perinatal transmission is a non-sexual route allowing the virus particles or infected immune cells to pass from infected mothers to their children during pregnancy, childbirth or breastfeeding. Even though the mother-to-child (MTC) transmission rate reduced dramatically over the past couple of decades from 25%-42% to 1% or less [1], exposure to the virus can still affect the infants by weakening their immune system during the first few months of their life <is this true? Find a reference>. At that stage, infants’ adaptive system is still poorly developed, and they rely heavily on breast milk for immune components including B and T cells and antibodies [2]. Therefore, weakened immune system of a mother will directly affect her infant’s ability to mount an adequate immune response against pathogens.

In this work, we applied the DPP algorithm to a set of flow cytometry data that was generated by the …(**@Mahan**: follow-up on the UBC lab JCVI Terry Fox lab, as well as Nima’s paper for details. Ask for protocol IRB approval, IACUC protocol, SOP. Statistical Analysis Plan (SAP), any other supporting document and experimental details that they can share with us. Also, how do we reference their work – main paper, group name, etc.). Discuss the experimental design here: samples from HIV exposed (HEU)and unexposed (UE) infants were treated with compounds that cause strong immune response in healthy individuals. Out of 6 treatments used by the experimenting team, we only chose to examine lipopolysaccharide (LPS) challenge is this should produce a very strong response. LPS is a bacterial coating that activates human immune cells on contact. This is a non-specific response by the acquired and innate immune system cells (**@Davit:** READ MORE IN THE GREEN BOOK AND WRITE DETAILS HERE).

First, we test samples within Unstimulated HEU and within UE for homogeneity (Hotelling test based on Hotelling T-square statistic). Same for LPS-stimulated samples. If the difference the distances withing groups are large compared to between groups, then we need to normalize te data in some way. To compare all 4 groups at once, we can use 2-way MANOVA.

Second, we examined compositional changes in HEU and UE unstimulated samples. Hypothesis #1: certain types of immune cells will be in lower abundance in the HEU samples because a) the mothers did not have enough of them to pass on; b) the exposure to HIV impacted infants’ cells.

Hypothesis #2: in healthy subjects LPS should activate several types of immune cells as well as differentiate naïve immune cells compared to unstimulated samples from the same individual. We calculated the differences of transformed light intensity measurements in each LPS-exposed sample cell with the median of the corresponding unexposed sample, separately for each marker. We hypothesized that the response magnitude would be lower in the HIV-exposed infant samples. Therefore, we should observe difference in HIV-Exposed LPS-Unstim and HIV-unexposed LPS - Unstim measurements. First, we look for clusters in *p*-dimensional space, then run DPP and try to match the clusters in the projections to the *p*-dimensional clusters. This is done to decide how many 2D projections are needed to observe all clusters. Once the projections are selected, we profile each cluster in each projection in terms of the protein markers – was a particular marker up, down or no-change in HIV-exposed vs UE group? What is the clinical meaning of the findings? We need an immunologist and a virologist to explain the findings.

# 2 Materials and Methods

## 2.1 Data Source and Experimental Design

Describe the experimental design (same as above, maybe move it here completely)

Describe the workflow:

1. FSC vs SSC: Separate lymphocytes using landmarks and convex hull (“lymphocyte finder”, option to find lymphocytes)
2. FSC vs SSC: Separate monocytes (same function, different option)
3. Save the lymphocytes and monocytes and delete all other rows
4. Transformation of all markers except FSC and SSC using automated optimal normalization search.
5. Hotelling test of transformed data to see if we need to further normalize the data. If the groups are not homogeneous (Hotelling statistics is very different between the subjects in the same group compared to between-groups. Also/Or withing subject LPS vs unstim.), and think about the results.
6. Taking the differences between unstim and LPS-stim samples (per subject). Right now, we do LPS –mean(unstim) on transformed data. Javier proposed to take quantile differences. Explore this.
7. Data nuggets to compress the data
8. Run DPP

# **3 Results**

1. Profile the cells in the regions that are differentially populated
2. Explain the results. E.g., macrophages got hyperactivated in the UE (high amounts of TNF produced) compared to HEU; number of specific type of immune cells is different; activation marker concentrations are different, … Make sure to map the results back to log-fold changes for easy interpretation of the results.

# 4 Discussion

Discuss the Stone Garden of Ryoanji**.**

# Figures

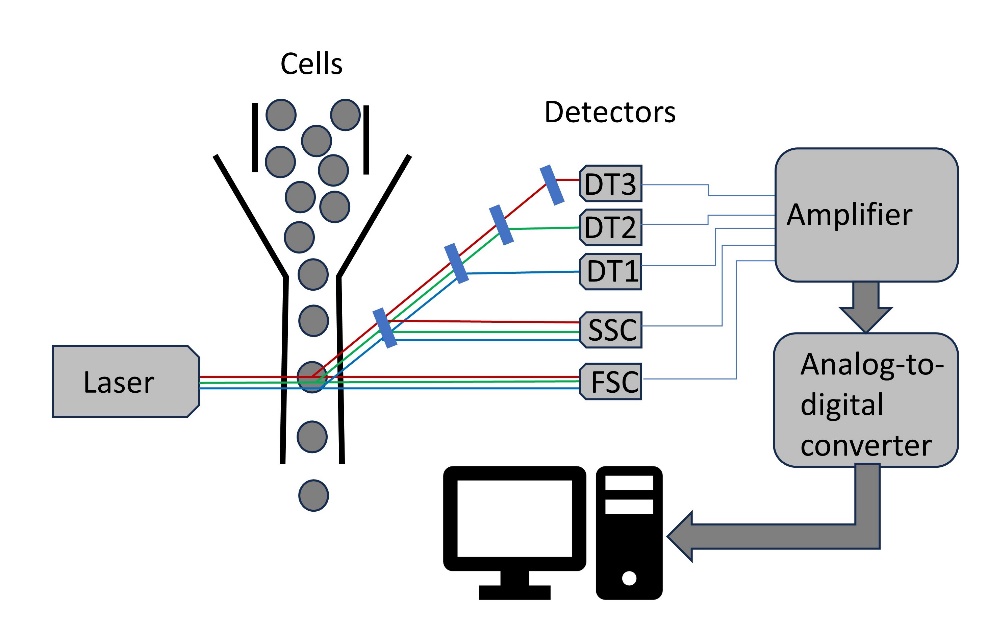


Figure 1: Schematics of a flow cytometer. As the cells pass through a narrow tube, the fluorochromes attached to the cells get excited by the laser. The light is deflected by a series of optical filters into the detectors. The signal from the detectors is amplified, converted from analog to digital, and stored on the computer.

# References

1 Hurst, S.A., Appelgren, K.E., and Kourtis, A.P.: ‘Prevention of mother-to-child transmission of HIV type 1: the role of neonatal and infant prophylaxis’, Expert Rev Anti Infect Ther, 2015, 13, (2), pp. 169-181

2 Lokossou, G.A.G., Kouakanou, L., Schumacher, A., and Zenclussen, A.C.: ‘Human Breast Milk: From Food to Active Immune Response With Disease Protection in Infants and Mothers’, Front Immunol, 2022, 13, pp. 849012