HiPPO and PANDA: two bioinformatics tools to support analysis of high-dimensional mass cytometry data

Stefano Pirrò¹, Andrea Cerquone Perpetuini¹, Milica Marinkovic¹, Livia Perfetto¹, Lucia Lisa Petrilli¹, Alessio Reggio¹, Marco Rosina¹ Filomena Spada¹, Luisa Castagnoli¹, Gianni Cesareni¹

¹ Department of Biology, University of Rome Tor Vergata, Rome, Italy

Corresponding Authors:

Stefano Pirrò¹
Via della Ricerca Scientifica, Rome, 00133, Italy
Email address: stefano.pirro@uniroma2.it

Gianni Cesareni¹
Via della Ricerca Scientifica, Rome, 00133, Italy
Email address: cesareni@uniroma2.it
Biological processes are modulated by the interaction of different cell types and their study requires technologies able to analyse single cells in heterogeneous populations.

High-dimensional mass cytometry enables the single cell analysis of more than 40 parameters. Several computational approaches have been proposed to reduce the multidimensionality of the datasets produced by this technology (i.e. SPADE and viSNE for clustering and visualization).

We have developed two new bioinformatics tools that help overcoming some of the limitations of the available toolboxes by defining quasi homogeneous cell populations (HiPPO) and by matching their expression profiles with that of cell populations described in the literature.

HiPPO uses a supervised quantitation approach to discretize the expression distribution curves generated for each marker monitored in the experiments. Cells in the continuous, multidimensional dataset are converted into a bi-dimensional matrix where row and columns are events (cells) and markers, respectively.

In order to characterize cell populations, HiPPO queries PANDA, a manually-curated database which stores expression profiles for selected markers of primary cells. The comparison and PANDA discrete expression profiles with those identified by PANDA in the populations under study, allow to monitor cell type abundance.

Moreover, given a set of experiments in different conditions, HiPPO can evaluate the variation of protein expression levels, for any identified population. This is performed by the use of Kolmogorov-Smirnov non-parametric test, evaluating empirical distribution differences of two samples. The analysis is conducted interactively, through a user-friendly web application.

HiPPO ability to identify populations and rapidity of execution have been evaluated and compared with the popular SPADE and viSNE tools. For benchmarking we have used a published dataset of healthy human bone marrow biopsies analysed with the CyTOF platform by simultaneously measuring 34 parameters in single cells (Qiu et al., 2011). As use case, we focused on 5 cell populations (Mature/Naïve T-helper cells, Mature/Naïve T-cytotoxic cells, B cells). The abundance of cell populations identified by HiPPO is highly-comparable to SPADE and viSNE analysis conducted by the authors (Figure 1). On the other hand HIPPO is 2,15X faster in completing the analysis.

HiPPO is a resource that can be easily used to analyze high-throughput multi-dimensional data. The synergy with PANDA represent a substantial improvement in the analysis pipeline and helps overcoming some of the shortcomings of other tools.

HiPPO is freely available at http://160.80.35.248/hippo

PANDA at http://160.80.35.248/panda
Figure 1: HiPPO has been used on mass cytometry data of human hematopoietic cells (Qiu et al., 2011). Accuracy in the identification of T and B cells is more comparable to manual gating than SPADE and viSNE.