Mass Cytometry Marker Panel Extension

T. Abdelaal[*1,3], A. Mahfouz[1,3], T. Höllt[2,3], V.V. Unen[4], F. Koning[4], B.P.F. Lelieveldt[1,3], M.J.T. Reinders[1,3] [1]Delft Bioinformatics Lab, [2]Computer Graphics and Visualization, Delft University of Technology, The Netherlands, [3]Computational Biology Center, [4]Department of Immunohematology and Blood Transfusion, Leiden University Medical Center, The Netherlands

Introduction: High-dimensional mass cytometry (CyTOF) permits the simultaneous measurement of many cellular markers, providing a system-wide view of immune phenotypes at the single-cell level[1]. Yet, the maximum number of markers that can be measure simultaneously is limited to ~50 due to several technical challenges. We propose a new method to integrate CyTOF data from several marker panels that include an overlapping set of markers, allowing for a deeper interrogation of the cellular composition of the immune system.

Materials & Methods: Given that the maximum number of markers on a CyTOF panel is N. The goal of our study is to expand the number of markers per cell by integrating measurements from two panels which share $m<N$ markers. The remaining slots can be used to measure (N-m) markers that are unique to each panel. By combining the data, we can extend the number of markers per cell to $2N$-m. We created a simulated dataset by selecting the CD8[+] T cells lineage (~460k cells, 32 markers) from a recent study[1]. We split the dataset into two halves (A and B), with cells in A represented by m+k1 and cells in B represented by m+k2 markers. The shared markers m were identified using three methods: PCA, Auto Encoder neural network, and HSNE[2]. The remaining N-m markers are split into the non-overlapping sets k1 and k2. We used KNN (K = 20) to impute the values of the k2 markers in A (not measured) using the k2 measurements from B, and vice versa.

Results & Discussion: To evaluate our method, we calculated the Euclidian distance between the imputed and measured marker values of each cell and compared them to all the pairwise distances in the lineage (mean±std = 8.6±0.9). The obtained distances for the different values of m are: 2.3±0.8 (m=4), 2.0±0.7 (m=8), 1.6±0.7 (m=12) and 1.6±0.7 (m=15). These preliminary results illustrate the feasibility of using a smaller subset of markers to represent the CD8[+] T cells lineage, providing a basis for an approach to extend the number of markers by combining data from multiple panels.

References
1. V.V. Unen et al., *Immunity* (44): 1227-1239, 2016.
2. N. Pezzotti et al., EuroVis (35): 3, 2016.