# Negative Reference Extraction for Machine-Assisted Flow Cytometry Immunophenotyping

Adam Koza1, Karel Fišer1

*1 Department of Bioinformatics, Second Faculty of Medicine, Charles University, Prague, Czech Republic*

Flow cytometry is a method for quantitatively measuring multiple parameters on millions of individual cells. This capability is widely used in both research and diagnostics. For example, in hemato-oncology, identifying malignant cell populations by flow cytometry is a crucial part of diagnosis and disease monitoring. Beyond identifying cell populations, their description, known as the immunophenotype, is essential for clinical decision-making.

Currently, immunophenotypes are often reported based on individual expert judgment. We have previously developed a system for automatically deriving immunophenotypes from data with labeled cells of interest and cells serving as a negative reference. However, in most cytometric data, the appropriate negative reference population is unknown. This is either because it is unclear which of the labeled populations should serve as negative, or whether each dimension of the data requires its own reference, or the negative reference is not among the labeled populations at all.

Here, we present possible approaches to identify the negative reference in flow cytometry data. In particular: 1) selection of one of the labeled populations as a negative reference using maximal distance to other populations or by the absolute position of its mode; 2) extraction of a reference for each dimension of the data separately, either for the constitution of an artificial joined population or for subsequent per-dimension analysis; and 3) an attempt at extracting Gaussian mixtures from individual dimensions on unlabeled data under the assumption that flow cytometry data parameters are normally distributed within cell populations.

Overall, we demonstrate different approaches for the automatic detection of negative cell populations to facilitate automatic immunophenotyping of flow cytometry data.