# Formalized machine-assisted flow cytometry immunophenotyping

Sieger Tomáš1, Koblížek Miroslav2, Fišer Karel3, 4

1 Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University in Prague, Prague, Czech Republic

2 Department of Pathology and Molecular Medicine, Second Faculty of Medicine, Charles University and University Hospital Motol, Prague, Czech Republic

3 CLIP-Childhood Leukaemia Investigation Prague, Department of Pediatric Hematology and Oncology, Second Faculty of Medicine, Charles University and University Hospital Motol, Prague, Czech Republic

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

Flow cytometry is a method for quantitative measuring of multiple parameters on millions of individual cells. Such ability is widely used both in research and diagnostics. For example, in hemato-oncology, the identification of malignant cell population by flow cytometry is a part of diagnostics and disease monitoring. Apart from idenitification of cell population its description, so call immunophenotype, is important for clinical decision making.

However the guidelines for description of the immunophenotype vary between laboratories despite multiple efforts for standardization. It is mainly due to the fact that the descriptions are formally imprecise and so leave space for subjective interpretations.

Here we present a system of deriving immunophenotypes automatically from data with labeled cells of interest and cells serving as negative reference. First we formalized several guidelines described in the literature. However, more importantly we also formalized immunophenotyping further by providing continuous measures of cell population positivities for individual analysed parameters in terms of distances between population of interest and reference negative cell population. We first defined a set of formal criteria (n = 9) for such measures. Then we considered known distance measures (n = 7) including Separation Index, Kullback-Leibler divergence and Marker Enrichment Modeling score, and added our own measures (n = 13) based mostly on ROC-AUC and Earth movers distances, and tested all measures for fulfilling the previously set criteria. To test the performance of selected measures we used synthetic data sampled from known distributions and selected best performing measures. To be able to compare the best performing measure to expert immunophenotype, the output of the measure was discretized into three categories (negative, medium and positive) . When used on two real world data sets, such categories were in agreement with expert-generated categories in 27 out of 28 parameters used to define 7 cell populations. As a last step we created a framework to automatically generate pdf or html reports from provided labeled data.

Overall we provide: 1) formalized versions of immunophenotype reporting rules, 2) a set of measures of population positivity based on distribution distances and evaluation of their peformance, and 3) R-based tools to generate immunophenotype report automatically from labeled data.

This work was supported by UNCE/MED/015.