A novel unbiased method links variability of co-expression between multiple proteins on single cells to a clinical phenotype

Gerjen H. Tinnevelt[1,2], Selma van Staveren[2,3], Kristiaan Wouters[4], Bart Hilvering[3], Rita Folcarelli[1], Leo Koenderman[3], Lutgarde M.C. Buydens[1], Jeroen J. Jansen[1]

[1]Radboud University, Institute for Molecules and Materials, (Analytical Chemistry/Chemometrics), P.O. Box 9010, 6500 GL Nijmegen, The Netherlands
[2]TI-COAST, Science Park 904, 1098 XH Amsterdam, The Netherlands
[3]Department of Respiratory Medicine and laboratory of translational immunology (LTI), University Medical Center Utrecht, Heidelberglaan 100, 3584CX, Utrecht, The Netherlands
[4]Dept. of Internal Medicine, Laboratory of Metabolism and Vascular Medicine, P.O. Box 616 (UNS50/14), 6200 MD Maastricht, The Netherlands
Email: gtinnevelt@science.ru.nl

Abstract: The immune system has evolved to combat a wide variety of infections, regulate healing processes, and displays many functions in homeostasis. This leads to an enormous complex cell mixture with a high cellular and functional diversity, which can be studied using flow cytometry. A typical multicolour flow cytometry sample may contain a large number of cells (>10,000), of which specific protein expressions are measured at the single-cell level. Individuals that exhibit an immune response, may have both changes in protein expressions on individual cells and changes in ratios of similar cells. Our new data analysis method Discriminant Analysis of Multi-Aspect flow Cytometry (DAMACY) uses Principal Component Analysis to describe the cellular distribution and subsequently uses Orthogonal Partial Least Squares – Discriminant Analysis to show systematic changes. Thus, DAMACY merges and quantitatively integrates all the relevant characteristics on protein co-expression, the specific cells on which these are expressed, the distribution of these cells within all samples and the systematic change in this distribution upon changes in homeostasis of the host such as immune responses. The resulting model is comprehensively statistically validated to optimize the model information content and robustness.

We also show how DAMACY may be used to quantitatively integrate different multicolour flow cytometry tubes. The multiple tubes are needed because the number of proteins per measurement is technologically limited. This is unfortunate, as many high-impact studies reveal that interrogation of more proteins simultaneously leads to a more comprehensive view on the immune system. We show how data fusion of all tubes may find all the relevant cells in an activated immune states in obese versus lean people. The resulting model may not find the best biomarker, but will show how very different cells may function together in the individual development of type 2 diabetes.

Acknowledgements: This research received funding from the Netherlands Organization for Scientific Research (NWO) in the framework of the Technology Area COAST of the Fund New Chemical Innovations.