Title: A clustering-based machine learning approach for automated debarcoding of mass cytometry data

Authors: Vera Bockhorn[1], Axel Schulz[1], Heike Hirseland[1], Lena Teichert[2], Tobias Alexander[2], Henrik Mei[1]

[1]Deutsches Rheuma-Forschungszentrum Berlin, a Leibniz Institute, Berlin, Germany
[2]Department of Rheumatology and Clinical Immunology, Charité - Universitätsmedizin Berlin, Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

Abstract:
Mass cytometry offers the advantage to measure multiple samples from different sources at once, using sample-specific metal isotope labels, i.e. barcode tags. The debarcoding of the measured sample convolute back into separate samples is a crucial step when doing multi-sample CyTOF measurements. While traditional debarcoding relies on manual gating, several algorithmic solutions have been developed to address this challenge.

In this work, we introduce a novel computational debarcoding method based on the FlowSOM algorithm, a widely utilized clustering technique for cytometry data. Our approach distinguishes itself by offering various parametric options that enable the user to control the quality and quantity of sample assignments. Notably, our method allows for the execution of two consecutive FlowSOMs to enhance cell assignments, the flexibility to define starting points for clustering, and a parameter governing cluster-to-sample assignment based on cluster homogeneity.

To validate and optimize our FlowSOM debarcoding algorithm, we utilized a "checkerboard" mass cytometric dataset. This dataset comprised 20 Smarttube-fixed leukocyte samples barcoded with a 6-choose-3 beta-2-microglobulin based scheme, recorded both individually (ground truth) and as a pooled sample convolute. The individual samples were characterized by trackable properties, such as the presence or absence of CD8 and CD45 staining, B-cell depletion, and diverse tissue origins.

Our evaluation demonstrates that the debarcoding performed by our novel algorithm results in a higher number of correctly assigned cell events compared to existing debarcoders. Importantly, our method preserves the original compositional properties of individual samples. In summary, our FlowSOM-based debarcoding approach proves to be a versatile and scalable alternative to existing methods. Its optimization capabilities make it particularly well-suited for barcoding situations with low and heterogeneous barcode separations, as commonly encountered in surface barcoding of living cells.