Average Overlap Frequency: A simple metric to evaluate staining quality and community identification in high dimensional mass cytometry experiments

El-ad David Amir, Xinzheng V. Guo, Oksana Mayovska, Adeeb Rahman

PII: S0022-1759(17)30383-6
DOI: doi: 10.1016/j.jim.2017.08.011
Reference: JIM 12362

To appear in: *Journal of Immunological Methods*

Received date: 4 April 2017
Revised date: 23 August 2017
Accepted date: 30 August 2017

Please cite this article as: El-ad David Amir, Xinzheng V. Guo, Oksana Mayovska, Adeeb Rahman , Average Overlap Frequency: A simple metric to evaluate staining quality and community identification in high dimensional mass cytometry experiments, *Journal of Immunological Methods* (2017), doi: 10.1016/j.jim.2017.08.011

Average Overlap Frequency: A Simple Metric to Evaluate Staining Quality and Community Identification in High Dimensional Mass Cytometry Experiments

El-ad David Amir, Xinzheng V. Guo, Oksana Mayovska, Adeeb Rahman

Human Immune Monitoring Center, Icahn School of Medicine at Mt. Sinai, New York, NY 10029

Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mt. Sinai, New York, NY 10029

Abstract

High dimensional cytometry now allows measurement of over 50 parameters in a single sample, and is typically visualized using sophisticated dimensionality-reducing methods and analyzed with automated clustering algorithms. While these tools facilitate the identification and presentation of key findings, it remains challenging to effectively monitor and report the staining quality of individual markers. We present the Average Overlap Frequency (AOF), a simple and efficient metric to evaluate and quantify the robustness of staining and clustering quality in high-dimensional data. We leverage the AOF to compare and determine the optimal storage conditions for stained whole blood samples prior to mass cytometry analysis. We also show that the AOF can be easily incorporated as part of automated analysis pipelines in large scale immune monitoring studies and used to flag and exclude samples with poor staining quality. We propose that the AOF may be incorporated as an essential quality control metric to better identify and report the underlying sample quality in all CyTOF and other high-dimensional cytometry experiments.

Keywords: mass cytometry, cytof, flow cytometry, computational biology, bioinformatics, single cell, high dimensionality.

1. Introduction

The growing fields of mass cytometry and highly-multiparametric flow cytometry now allow the measurement of over 40 parameters on millions of cells in a single sample, offering new opportunities to unravel the cellular heterogeneity of complex biological samples [Bendall et al., 2011, Bendall et al., 2012]. Given their high dimensionality, analyzing these datasets using traditional biaxial gating approaches is inefficient and often ineffective. Instead, investigators typically employ a growing range of sophisticated visualization methods [Amir et al., 2013, Tang et al., 2016], automated clustering algorithms [Qian et al., 2010, Qiu et al., 2011, Levine et al., 2015, Van Gassen et al., 2015], and other tools [Bruggner et al., 2014, Spitzer et al., 2015]. These can greatly facilitate the analysis of high-dimensional cytometric datasets and provide lower-dimensional representations of the data that are needed for publication and presentation. However, they can sometimes mask underlying issues with the