Journal of Immunological Methods xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Journal of Immunological Methods

journal homepage: www.elsevier.com/locate/jim

Research paper

Setting objective thresholds for rare event detection in flow cytometry

Adam J. Richards [a,b,c,*], Janet Staats [b,c,d], Jennifer Enzor [b,c,d], Katherine McKinnon [e], Jacob Frelinger [f], Thomas N. Denny [c,g], Kent J. Weinhold [b,c,d], Cliburn Chan [a,b,c]

[a] Department of Biostatistics & Bioinformatics, Duke University, Durham, NC, USA
[b] Duke Center for AIDS Research, Duke University, Durham, NC, USA
[c] Duke External Quality Assurance Program Oversight Laboratory, Duke University, Durham, NC, USA
[d] Department of Surgery, Duke University Medical Center, Durham, NC, USA
[e] Vaccine Branch, Center for Cancer Research, NCI, Bethesda, MD, USA
[f] Institute for Genome Sciences and Policy, Duke University, NC, USA
[g] Duke Human Vaccine Institute, Duke University, Durham, NC, USA

ARTICLE INFO

Article history:
Received 7 August 2013
Received in revised form 5 March 2014
Accepted 1 April 2014
Available online xxxx

Keywords:
ICS
Standardization
Reproducibility
Rare events
Positivity
Automated analysis

ABSTRACT

The accurate identification of rare antigen-specific cytokine positive cells from peripheral blood mononuclear cells (PBMC) after antigenic stimulation in an intracellular staining (ICS) flow cytometry assay is challenging, as cytokine positive events may be fairly diffusely distributed and lack an obvious separation from the negative population. Traditionally, the approach by flow operators has been to manually set a positivity threshold to partition events into cytokine-positive and cytokine-negative. This approach suffers from subjectivity and inconsistency across different flow operators. The use of statistical clustering methods does not remove the need to find an objective threshold between between positive and negative events since consistent identification of rare event subsets is highly challenging for automated algorithms, especially when there is distributional overlap between the positive and negative events ("smear"). We present a new approach, based on the F_β measure, that is similar to manual thresholding in providing a hard cutoff, but has the advantage of being determined objectively. The performance of this algorithm is compared with results obtained by expert visual gating. Several ICS data sets from the External Quality Assurance Program Oversight Laboratory (EQAPOL) proficiency program were used to make the comparisons. We first show that visually determined thresholds are difficult to reproduce and pose a problem when comparing results across operators or laboratories, as well as problems that occur with the use of commonly employed clustering algorithms. In contrast, a single parameterization for the F_β method performs consistently across different centers, samples, and instruments because it optimizes the precision/recall tradeoff by using both negative and positive controls.

1. Introduction

The classification of events as positive and negative based on the setting of a threshold has traditionally been a fundamental requirement in many flow cytometry (FCM) applications, particularly in the case when positive and negative populations overlap (Maecker and Trotter, 2006). In the context of HIV monitoring, intracellular staining (ICS) assays are often employed to track functionally active antigen-specific cells that may be exceedingly rare. The current practice in most laboratories is to set a positivity threshold for each effector function (e.g. cytokine expression) by visual comparison of negative and test/positive control data, designating events that fall above the

* Corresponding author at: Station d'Ecologie Expérimentale du CNRS à Moulis, French National Center for Scientific Research (CNRS), 09200 Moulis, France. Tel.: +33 5 61 04 05 87.
E-mail address: adam.richards@ecoex-moulis.cnrs.fr (A.J. Richards).

http://dx.doi.org/10.1016/j.jim.2014.04.002
0022-1759/

threshold as positive. However, there is no objective method for threshold determination in the FCM community and this represents a roadblock to harmonizing ICS analyses across laboratories (Maecker et al., 2005, 2010). Visual threshold determination is problematic due to its subjectivity, but also because there is poor scalability to large panels.

There is sometimes substantive overlap between the positively and negatively stimulated samples in terms of target cell subsets and it becomes necessary to either set a threshold based on expert opinion or 'tune' the algorithm or model to enable discovery of the rare events in the case of automated methods. A number of potentially viable methods to detect rare events are available through the use of clustering (Finak et al., 2009; Hahne et al., 2009; Lo et al., 2009; Pyne et al., 2009; Cron et al., 2013). An important initiative called FlowCAP (Aghaeepour et al., 2013) exists to critically evaluate the numerous methods available for automated analysis in flow cytometry. If the target population is reasonably separable from the negative events then the use of automated methods like clustering is ideal and it eliminates the need to find a threshold. The main issue with clustering methods and model-based methods in general is that of data masking, where the target population is identified as events in the tail of the negative event population rather than as a separate population. These clustering methods are also unsupervised which creates the additional challenge of labeling clusters as positive — in this case, the availability of an objective threshold can be helpful in separating positive from negative clusters.

The method we propose here provides an objective means of separating biologically meaningful categories of events that are difficult to consistently resolve with clustering. Our method essentially sets a threshold by optimizing the precision–recall trade-off through the use of both positive and negative controls (Calvelli et al., 1993; Nicholson et al., 1996) as the use of negative controls alone cannot control for false discoveries. This approach to the automatic assignment of thresholds is one-dimensional. However, the F-score threshold can serve as a generator for methods that combine univariate thresholds to identify high-dimensional cell subsets (Roederer et al., 2011; Aghaeepour et al., 2012) or as a filter for events of interest before further exploratory analysis with unsupervised algorithms (Qiu et al., 2011).

In this work, we first illustrate common scenarios where thresholding methods based on negative controls alone perform poorly. Then we compare several commonly employed clustering algorithms and discuss each method's suitability in the context of rare event detection. Finally, we compare F_β thresholds to expert visual gating, optimized using back-gating, by making use of data from the multi-center proficiency study, EQAPOL.

2. Methods

2.1. Data sets

Two 11-color data sets (11C-EQAPOL-1, 11C-EQAPOL-2) with explicit positive (SEB) stimulations were used in this study as well as a 4-color data set (4C-EQAPOL) without an explicit positively stimulated control. Negative controls for the 11-color data included co-stimulatory monoclonal antibodies (mAbs) anti-CD28 and anti-CD49d together with both Brefeldin A (BRF) and monensin, while the negative control for the 4C-EQAPOL panel used only dimethylsulfoxide (DMSO) (no Costim) and BRF. The 11C-EQAPOL-1 data were used to demonstrate the difficulties encountered with an endogenous background response, where the 11C-EQAPOL-2 data provided a data set with a more typical response. All three panels were developed as part of the External Quality Assurance Program Oversight Laboratory (EQAPOL) proficiency program. The lymphocyte subsets for these three data sets are available through http://duke.edu/~ccc14/papers/fscore.

2.2. Sample preparation and ICS assay

Normal human donors were leukapheresed in accordance with Duke University's Institutional Review Board and informed consent was obtained prior to sample collection (Jaimes et al., 2011). Sample preparation and staining were performed as previously described for the 4-color (Jaimes et al., 2011) and 11-color ICS assays (Ottinger et al., 2008; Snyder et al., 2011).

2.3. Manual gating

Gating for each data set was performed by highly trained operators in accordance with our established standard operating procedure and the process included extensive back-gating to both maximize signal and minimize noise. Uniform gates were applied within each donor. In Section 3.2 the manual gates and thresholds (see Fig. 2) from two independent experts were used to infer a range for the value of β, the principal tunable parameter in the F_β method. In Fig. 3, manual gates and thresholds from two independent labs who participated in the EQAPOL 4-color ICS EP1 Program were used.

2.4. Automated analyses

All automated data analyses were carried out using the Python programming language (http://python.org). In addition, all figures in this manuscript were produced using the Python library matplotlib (Hunter, 2007). The basic subsets for all samples were found by manual gating and exported from Flowjo as FCS files. The subsets were then imported into the Python environment using the Python package py-fcm (http://code.google.com/p/py-fcm). The axis scaling for event plots that used a biexponential transform was configured for visual clarity (Parks et al., 2006). All plotted events use a biexponential transformation unless otherwise stated with the biexponential parameters ($w = 0.5, D = 4.5, T = 262144$). The calculation of an F_β determined threshold is detailed in Section 3.1. The parameters for the positivity thresholding method were optimized in Section 3.2.

In Section 4.2, there are a number of clustering algorithms that were applied to discover cytokine subsets. These methods were realized through the use of py-fcm along with the machine learning package scikit-learn (Pedregosa et al., 2011). The parameters were tuned by hand using a basic grid search approach. We also constrained each method to a single best set of parameters that work for all three stimulations. We provide in the supplemental materials (http://duke.edu/~ccc14/papers/fscore) a description of these methods and all necessary code required to reproduce the results and accompanying figure.

Please cite this article as: Richards, A.J., et al., Setting objective thresholds for rare event detection in flow cytometry, J. Immunol. Methods (2014), http://dx.doi.org/10.1016/j.jim.2014.04.002