Published in final edited form as:

Cytometry A. 2010 December ; 77(12): 1126–1136. doi:10.1002/cyto.a.20987.

Optimization of a highly standardized carboxyfluorescein succinimidyl ester flow cytometry panel and gating strategy design using Discriminative Information Measure Evaluation (DIME)

Cliburn Chan^{1,*}, Lin Lin², Jacob Frelinger³, Valerie Hebert⁵, Dominic Gagnon⁵, Claire Landry⁵, Rafick-Pierre Sékaly⁵, Jennifer Enzor⁴, Janet Ottinger⁴, Kent J Weinhold⁴, Maria Jaimes⁶, and Mike West²

¹Department of Biostatistics and Bioinformatics, Duke University Medical Center, Durham, NC 27710, USA

²Department of Statistical Science, Duke University, Durham NC 27708, USA

³Program in Computational Biology and Bioinformatics, Institute for Genome Sciences and Policy, Duke University, Durham, NC 27708, USA

⁴Center for AIDS Research, Duke University Medical Center, Durham, NC 27708, USA

⁵National Immune Monitoring Laboratory (NIML), Université de Montreal, Génome Québec, Montreal, Canada

⁶BD Biosciences, San Jose, CA, 95131 USA

Abstract

The design of a panel to identify target cell subsets in flow cytometry can be difficult when specific markers unique to each cell subset do not exist, and a combination of parameters must be used to identify target cells of interest and exclude irrelevant events. Thus, the ability to objectively measure the contribution of a parameter or group of parameters towards target cell identification independent of any gating strategy could be very helpful for both panel design and gating strategy design. In this paper, we propose a *Discriminative Information Measure Evaluation* (DIME) based on statistical mixture modeling; DIME is a numerical measure of the contribution of different parameters towards discriminating a target cell subset from all the others derived from the fitted posterior distribution of a Gaussian mixture model. Informally, DIME measures the “usefulness” of each parameter for identifying a target cell subset. We show how DIME provides an objective basis for inclusion or exclusion of specific parameters in a panel, and how ranked sets of such parameters can be used to optimize gating strategies. An illustrative example of the application of DIME to streamline the gating strategy for a highly standardized carboxyfluorescein succinimidyl ester (CFSE) assay is described.

Keywords

CFSE standardization; Discriminative Information Measure Evaluation (DIME); Gating strategy optimization; Mixture model; Panel design

*Corresponding author: Cliburn Chan, Department of Biostatistics and Bioinformatics, Duke University Medical Center, Durham, NC 27710, USA. cliburn.chan@duke.edu.

Introduction

Multi-parameter flow cytometry (FCM) technology has seen dramatic advances in recent years, with 5 or more color assays now performed routinely in many basic and translational research laboratories (1). Standardization of all aspects of FCM, from instrument setup to data analysis, is an ongoing effort by multiple organizations, since standardization is necessary for consistent data comparison across sites (2). Multiple investigators have pioneered the use of advanced techniques and technologies to accurately evaluate immune cell subsets in multi-center programs for well over two decades, including the use of *backgating* for measuring *purity* and *recovery* combined with *checksums* (3), use of CD45 in three-color (4) and four-color assays (5), use of a single platform technology for absolute counts (6,7), panleukogating (8) and the use of pre-aliquoted lyophilized reagents (2).

In the context of FCM assays performed in Good Clinical Laboratory Practice (GCLP)-compliant laboratories, demonstration of reproducibility is critical for clinical acceptance (9-11). The reproducibility of flow cytometry (FCM) assays relies on key elements of the assay being standardized and well-characterized, including instrument and reagent qualification, sample preparation processes and analysis protocols (12-15). Over the past few years, multi-center standardization studies for many types of flow assays have consistently shown that sub-optimal data analysis methods are one of the most significant source of variability (2,16,17). Variability can be reduced by collection of sufficient events, use of appropriate controls, careful parameter selection and optimized gating strategies (18-21). In the context of analysis, the use of highest purity and lowest contamination measures as well as backgating can aid in the design of appropriate gates.

In the drive to maximize *recovery* and *purity* (22), gating strategies can sometimes become increasingly complex even when there are relatively few parameters being measured. While this is effective for a single laboratory, it is difficult to apply complex gating strategies consistently across different instruments and operators across multiple laboratories. Thus, the ability to objectively measure the contribution of a specific parameter or combination of parameters towards target cell identification independent of any gating strategy could be very helpful for both panel and gating strategy design.

Recent developments in computational statistics allow us to discover and monitor target cell subsets directly in multiple dimensions without use of a sequence of gates. Several groups, including ours, have recently published gating-free model-based approaches to cell subset identification using statistical mixtures of Gaussian, T or skewed distributions (23-26). Here, we show that the predictive density resulting from such model-based approaches can be exploited to perform a Discriminative Information Measure Evaluation (DIME) for FCM parameters. DIME analysis allows us evaluate parameter usefulness for identifying a target cell subset that can be specified as some collection of mixture components. From a biological perspective, DIME provides insight into optimal parameter combinations that characterize a cell subset in a way that is independent of any particular gating strategy. Practically, DIME provides an objective basis for standardizing the analysis of flow cytometry panels in multi-center clinical trials, and can contribute to improved assay reproducibility.

We show the application of DIME to the design of a simplified gating strategy for a CFSE-based assay designed to measure CD4 and CD8 T lymphocyte proliferation following antigen challenge. The context for this proof-of-concept analysis was a three center pilot study (BD Biosciences, Université de Montreal/NIML and Duke University) sponsored by DAIDS to standardize the assessment of T lymphocyte proliferation using a panel for CD3, CD4, CD8, CFSE and an amine viability stain. Experts at the three centers had, through

careful evaluation of their collective data, developed a standard consensus gating strategy that was designed to reduce background and enhance detection of specific proliferation.

Materials and Methods

Sample preparation

PBMCs used in this study were provided by SeraCare Bioservices. Briefly, concentrated leukocytes were prepared by machine leukopheresis with anticoagulant ACD-A by BRT Laboratory (Baltimore, MD). PBMC were isolated within 8 hr post collection using Ficoll procedure. Cell concentration was determined using Guava ViaCount assay (Guava Technologies Inc, Hayward, CA) and PBMCs were frozen at 15×10^6 cells/mL in freezing media (22% FCS, 7.5% DMSO and 70.5% RPMI). Vials of the cryopreserved PBMC for each donor were shipped to the three different institutions (NIML, BDB, Duke) using a liquid nitrogen dry shipper.

CFSE assay

Two vials of frozen PBMCs were quickly thawed at a time at 37°C in a water bath, and washed with RPMI containing 20% FBS and 50U/ml of benzonase nuclease (Novagen). Cells were then washed twice and resuspended in PBS at a concentration of 20×10^6 cells/mL. Just before use, a 2x CFSE working solution in PBS was prepared from a 5 mM CFSE (Invitrogen) stock solution. An equal volume of the CFSE 2x solution was then added to the cell suspension. Cells were incubated in the dark for 8 min at room temperature. At the end of the incubation, CFSE labeling was stopped by adding the same volume of 100% human AB serum (Gemini Bio-products, West Sacramento, CA) and incubating the cells for an additional 2 minutes. The CFSE labeled cells were then washed twice with PBS and resuspended in RPMI containing 10% human AB serum, 1% HEPES, 1% L-Glutamine and 1% penicillin-streptomycin at a concentration of 1.5 million cells/800 µl. Cells were then transferred to 5 ml sterile polypropylene tubes and 200 µl of different antigen preparations, resuspended from their lyophilized form, were added to the corresponding tubes. Three to five replicates of each condition were included in each experiment. Cells were incubated in a 5% CO₂ incubator at 37°C for 144 hours. Cultured cells were then washed with PBS and transfer to wells in a 96 well V bottom plate (BD Falcon), for antibody staining.

Lyophilized stimulation and staining reagents

Preconfigured lyophilized stimulation and staining plates that have been previously validated (27) were used to provide simplified assay setup. Stimuli were provided in lyophilized form in appropriate wells of a polypropylene V-bottom 96-well plate. CMV pp65 peptide pool (138 peptides consisting of 15 amino acid residues, overlapping by 11 aa residues each) was used at 1.7 µg/ml. As a positive control, a mixture of anti-CD3 (10 ng/ml) (OKT3) and anti-CD28 (L293, BD Biosciences, 1 µg/ml) was used. Dimethyl sulfoxide (DMSO) matching the concentration present in the CMVpp65 peptide stimulus was used as a negative control. Lyophilized staining antibody cocktail was provided in the corresponding wells of a second plate. All antibodies and CMV peptides were obtained from BD Biosciences (San Jose, CA). The following antibody mixture was used: anti-CD8 (SK1) PE/anti-CD4 (SK3) PerCP-Cy5.5/anti-CD3 (SK7) APC.

Cells were initially kept in PBS for staining with optimal titer of the viability dye Live/Dead fixable Aqua Dead Cell stain (Invitrogen, Carlsbad, CA). Cells were incubated at room temperature in the dark for 20 minutes, washed with PBS 2% FBS and stained using the lyophilized monoclonal antibodies. Cells were then fixed using 2% paraformaldehyde and acquired on LSR II (BD Biosciences, San Jose, CA) flow cytometers with slightly different configuration across participating institutions. Laboratories acquired all events in each

sample. Data was analyzed using either FACSDiva (BD Biosciences, San Jose, CA) or FlowJo (Tree Star Inc, Ashland, OR).

Optimization of conventional gating strategy

A gating strategy was designed taking into account the experience of the participating laboratories related to CFSE data analysis. Of note, the goal of the analysis in this study was to assess the percentage of proliferative (CFSElow) CD4 and CD8+ T-cell subsets, hence a semi quantitative method. Calculation of various proliferation indexes was beyond the scope of this work. The gates that were considered to be useful in order to reduce background, increase signal and allow for detection of aberrant samples were: doublet discrimination gate (FSC-H vs FSC_A), CFSE vs SSC-A (also known as Mississippi gate), lymphocyte gate FSC_A vs SSC_A, CD3 vs SSC-A and CD4 vs CD8 (in order to discriminate the CD4+ and CD8+ T cell subsets) The stained but unstimulated (DMSO) control was used to set the boundary for enumeration of proliferating cells. See Figure 1.

Use of DIME to simplify gating strategy

The original expert-determined standardized gating strategy made use of all 8 parameters in at least one dot plot over a total of 9 gating generations. To explore alternative, unconventional gating sequences, we looked for projections of the most informative subset of parameters as ranked by DIME in which the target clusters were well separated in the remaining parameters. Using this approach, a revised and considerably simplified manual gating scheme was found and comparative evaluations of the original and revised strategies performed.

Statistical models

Mixture models with non-Gaussian cell subtypes

To flexibly represent the data density of parameters in M dimensions we use an encompassing Gaussian mixture model with a relatively large number J of components, namely

$$g(x) = \sum_{j=1:J} \pi_j N(x | \mu_j, \Sigma_j)$$

where x is an M dimensional vector of parameters, the π_j are probabilities summing to 1 and $N(x | \mu_j, \Sigma_j)$ denotes the density of a Gaussian distribution for x with mean vector μ_j and $M \times M$ covariance matrix Σ_j . The analysis estimates the parameters $\Theta \{J; \pi_{1:J}, \mu_{1:J}, \Sigma_{1:J}\}$ using Bayesian computation methods as previously described (23). The current analyses are based on standard Bayesian Dirichlet mixture modeling methods (28,29) under which some of the mixture probabilities may be zero, so allowing the model to cut back to fewer components than the upper bound J as relevant for the data set at hand. We do not interpret the fitted Gaussian mixture components, but regard the model as a global, flexible representation of what are often quite heterogeneous data configurations. Given a fitted model, in terms of estimates of all parameters, we then identify subtypes by clustering the Gaussian mixture components into groups; this assigns each of the J components of the mixture to one of C groups, or clusters, with cluster index sets I_c containing component indices j for each cluster $c = 1: C$. These clusters are then identified with biological cell subtypes, reflected in the re-expression of the overall data density as

$$g(x) = \sum_{c=1:C} \gamma_c f_c(x)$$

where the mixing probabilities $\gamma_c = \sum_{j \in I_c} \pi_j$ give relative proportions of subtypes $c=1:C$, and the parameter distributions have subtype-specific parameter densities

$$f_c(x) = \sum_{j \in I_c} (\pi_j / \gamma_c) N(x | \mu_j, \Sigma_j).$$

The $f_c(x)$ may be markedly non-Gaussian, each flexibly fitted via the overall mixture modeling strategy. Fitted parameter values Θ and the cluster index sets $I_{1:C}$ defined in the applied data analysis study of this paper are available at <http://ftp.stat.duke.edu/WorkingPapers/10-10.html>.

Mode search for clustering components into cell subtype clusters

Modes (and antimodes) of $g(x)$ are found by an efficient numerical optimization using the mode trace function for Gaussian mixtures. To simplify notation, define the precision

matrices $\Omega_j = \Sigma_j^{-1}$. Modal search start with iteration index $i = 0$ and a point x_0 in parameter space and then, for $i = 1, 2, \dots$, iteratively computes

$$x_{i+1} = A(x_i)^{-1} \sum_{j=1:J} \alpha_j(x_i) \Omega_j \mu_j$$

where $A_j(x) = \sum_{j=1:J} \alpha_j(x) \Omega_j$ and $\alpha_j(x) = \pi_j N(x | \mu_j, \Sigma_j)$ at any point x . This defines a rapidly convergent local mode search that quickly identify modes, antimodes and ridge lines between them in the contours of Gaussian mixtures, and typically takes just a few iterates. A second derivative of $g(x)$ evaluated at any identified stationary point then identifies it as a mode or antimode. Rather than being interested in all modes of $g(x)$, we are here only interested in those that define basins of attraction for the mixture components in order to find the sets I_c of component indicators related to different modes a_c . We run this numerical search J times, initializing at $x_0 = \mu_j$, $j=1:J$ in turn, and record the unique modes a_c so identified as well as the sets I_c of Gaussian components attracted to each in this search. Matlab code for these computations is available <http://ftp.stat.duke.edu/WorkingPapers/10-10.html>.

Discriminative information measure

To assess discrimination of cell subtype c from the rest, we compare the expected value of $f_c(x)$ for cells that are *not* of subtype c relative to that for cells of all subtypes. These two expected values are, respectively,

$$\delta_c = (1 - \gamma_c)^{-1} \sum_{e=1:C, e \neq c} \gamma_e f_{c,e} \quad \text{and} \quad \Delta_c = \sum_{e=1:C} \gamma_e f_{c,e}$$

where $f_{c,e} = \int f_c(x) f_e(x) dx$ is easily computed based on the encompassing mixture model parameters. The discriminative information measure for subtype c is then simply

$$d_c = \delta_c / \Delta_c.$$

This measures information discriminating cell subtype c from the rest on the (0-1) scale. A *low* value of d_c represents a *high* level of discrimination of cell subtype c from the rest. In comparing discrimination of cell subtypes based on different subsets of parameters, we need to make explicit in the notation which parameters are used. For any subset of parameters $h \subseteq \{1 \dots M\}$ we therefore use the notation $d_c(h)$ for the discriminative information measure when restricting to only those parameters. Computationally, this simply uses the marginal mixture distribution for the parameters in index set h . Further, we use the notation $d_c(-h)$ for the measure evaluated at all parameters *excluding* those in the subset h . Finally, comparing any subset of parameters in their discriminative ability relative to the full set of M parameters can be done in terms of the difference in values $d_c(h) - d_c(1 \dots M)$.

If one parameter is independent of the rest and has the same distribution over all cell subtypes, then d_c will take the same value when computed in the mixture model analysis with or without that redundant parameter, showing the irrelevance of that parameter. We can automatically generate ranked sets of parameters by clusters, to assess how substantial changes in discriminative information are in practice, and to identify small groups of subtype-characterizing parameters as well as parameters that contribute negligibly to discrimination of some or all cell subtypes.

Parameter discrimination on simulated data

A simple *simulated* data set illustrates the principles of DIME using synthetic data drawn from a specific mixture of 3 normal distributions for $M=3$ parameters. As shown in Figure 2, DIME provides clear and unambiguous measures of the ability of a given parameter to discriminate a specific mixture component, and is able to identify non-informative parameters.

Evaluation and “gold standards”

Our analysis also evaluates parameters based on sensitivity and specificity calculations relative to a chosen “gold standard” for cell subtype classification of all the data points. Using the full set of M parameters, a cell with parameter values x is classified as of subtype c if $\gamma_c f_c(x)$ gives the maximum value over subtypes. When using all parameters, we define this as the “gold standard” classification, i.e., a hypothetical “true” classification for the purposes of comparing classification based on fewer. Recomputing the classifications based on the marginal mixture on any reduced set of h parameters, we can then evaluate which cells are misclassified with respect to each cell subtype relative to this gold standard. The resulting empirical sensitivity and specificity for any subset of parameters provides a useful practical guide to the impact of reducing to parameter subsets defined as optimal with respect to the discriminative information measure.

Results

Parameter panel and target data sets

All 3 centers used the following five color panel to assess T lymphocyte proliferation (CFSE, CD8 PE, CD4 PerCP-Cy5.5, CD3 APC and viability Aqua Amine dye). Together with FSC-A and FSC-H (to distinguish singlets from cell aggregates) and SSC-A, the typical data set comprised 1 million events in 8 dimensions. In each case, the objective was to evaluate the relative frequency of proliferating (CFSE low) and non-proliferating (CFSE

high) CD4+ and CD8+ lymphocytes. Our four target data sets are therefore CFSE low CD4+, CFSE high CD4+, CFSE low CD8+ and CFSE high CD8+ lymphocytes.

Finding proliferating lymphocytes with Bayesian mixture modeling

The data are modeled using an 8-dimensional mixture of Gaussians with Bayesian Markov chain Monte Carlo (MCMC) for model fitting, followed by modal clustering to identify cell subsets (23). This identified 4 modal clusters corresponding to CD4 non-dividing, CD4 dividing, CD8 non-dividing and CD8 dividing T lymphocytes; see Figure 3 for the CD4/CFSE and FSC-A/SSC-A projections. While the non-proliferating lymphocytes had classical scatter characteristics, the proliferating blasts were not in the standard “lymphocyte gate” location on FSC-A/SSC-A, and would be impossible to distinguish on the basis of scatter alone. Instead, they were identified on the basis of their CD3, CD4 and CD8 properties.

Finding the optimal combination of k parameters to identify target cell subsets

To identify parameter combinations that are best able to discriminate the target cell subsets (dividing and non-dividing CD4 and CD8 lymphocytes), we calculated DIME for all possible subsets of size k . Since there are 8 parameters and each can be absent or present, there are 255 ($2^8 - 1$) subsets to evaluate. The *maximally informative* subsets of each size k , where k ranged from 8 to 1, is shown in Figure 4. Surprisingly, the exclusion of the T cell receptor complex component CD3, which is a classical marker for T lymphocytes, resulted in minimal loss of discriminative information for all 4 target subsets, indicating that the CD3 marker is statistically redundant in the presence of CD4 and CD8. FSC-H and the amine viability parameter also contribute little unique discriminative information for these subsets.

A complementary use of DIME is to identify the critical discriminatory parameters by examining the sequence of parameters dropped for the *minimally informative* subsets of size k , as shown in Figure 5. The results show that CD4, CD8 are important for all subsets, and CFSE is critical for the dividing cell subsets, while the scatters are more important for non-dividing lymphocyte subsets

Evaluating the drop in sensitivity and specificity when using fewer parameters

To validate the DIME results, we evaluated the sensitivity, specificity and accuracy of each target cell subset by calculating the classification of every event in all 255 parameter combinations. The classification of events with all parameters present was used as the reference standard. In other words, we first assume that the classification of events into dividing CD4, non-dividing CD4, dividing CD8, non-dividing CD8 or “none of the above” subsets with all 8 parameters present is correct and hence serve as our reference cell subsets. We then calculate the number of true positives, false positives, true negatives and false negatives for each subset to derive the sensitivity, specificity and accuracy shown in Figure 6 for the most informative subsets identified by the DIME analysis. Figure 6 shows that this brute-force evaluation supports the usefulness of DIME since the optimal parameter subsets retain high values for these parameters down to subsets of size 4. An obvious advantage of using DIME is that it is computationally trivial compared with exhaustive evaluation by classification of events, and hence much more scalable to larger parameter subsets, as well as automatic.

Simplification of gating strategy

The information provided by DIME as to the relative usefulness of each parameter, and in particular, the revelation that CD3 did not contribute materially to the discrimination of proliferating and non-proliferating CD4+ and CD8+ lymphocytes, motivated us to explore alternative, unconventional gating strategies that excluded CD3, by simply looking for

projections in which the target clusters were well separated in the remaining parameters. We found an extremely simple gating strategy that had only 3 gating generations using 5 parameters that was effective in separating our target cell clusters, and this was mapped to a manual gating strategy (Figure 7).

Evaluation of simplified gating strategy

The reduced manual gating strategy was subsequently evaluated on all the data samples from a single CFSE standardization test panel performed by all 3 laboratories. The data samples were drawn from multiple patients treated under 3 different stimulation conditions (unstimulated, CMV pp65 and anti-CD3/CD28 mAbs). For all the quantities of interest (counts of CD4 and CD8 proliferating and non-proliferating populations), the evaluated cell subset frequencies agreed with the original expert manual analysis performed at the NIML (Figure 8). While the results suggest that consistent analysis can be achieved using different gating strategies, the gating strategy suggested by DIME analysis is the simplest and hence a good candidate for multi-institution standardization studies when using this specific panel.

Discussion

We have described a novel and effective statistical method for evaluation of the usefulness of any given parameter or group of parameters for discriminating a target cell subset; the method uses a discriminative measure calculated from the density of a fitted mixture model. To the best of our knowledge, this is the first description of an automated, quantitative approach to evaluating how useful any given parameter in a flow cytometric panel is for identifying cell subsets of interest. We also demonstrated using a CFSE proliferation assay as an example that such a measure is potentially useful for both panel optimization and gating strategy, especially in the context of assay standardization.

In the chosen example, as expected, CFSE, CD4, and CD8 were the most informative parameters. That scatter and aAmine parameters were among the least informative is also reasonable since these parameters are used primarily to reduce interference and background by negatively selecting dead/dying cells and cellular debris and are not part of positive selection for the target CD4 or CD8 proliferating cell populations. However, a surprising and very interesting result of our analysis was that in the presence of CD4 and CD8, CD3 did not contribute materially to the discrimination of proliferating and non-proliferating lymphocytes in a CFSE assay. From a technical perspective, the samples used in the CFSE assays were peripheral blood mononuclear cells (PBMC) that would have included general populations of lymphocytes and monocytes. In this experiment, PBMC samples were analyzed after a 6 day stimulation with antigen-specific peptide or anti-CD3/anti-CD28 mAbs. Potentially contaminating monocytes, NK, and B cells would be either strongly adherent or dead by the time of sample staining and acquisition. In hindsight, therefore, it is biologically reasonable that CD3 was not necessary to discriminate CD4+ and CD8+ lymphocytes in our CFSE experiments since it is likely that only CD4+ and CD8+ T cells remained after the 6 day stimulation. However we strongly caution that the lack of information provided by CD3 in this experimental example should not be generalized outside of this highly standardized CFSE assay. A great amount of effort to optimize and standardize each aspect of the assay used in our analysis yields highly reproducible results that were not previously possible (27)(Landry et. al.) and applying the same parameter usefulness analysis to an identical data set that has been acquired in a less standardized setting may not yield similar results. Rather, the point of this example is that commonly used parameters may not add value to the panel and the potential usefulness of each parameter used in a given panel is best determined with an objective measure like DIME.

The principle underlying DIME is conceptually simple. In contrast to the sequential 2D approach of conventional gating, multi-dimensional statistical analysis identifies cell populations *using all parameters simultaneously*. In a fitted model, DIME basically formalizes the idea that if we can “drop” one or more parameters at a time, and re-evaluate the likelihood that any given event in the target cell population still belongs to the original population it was in when all parameters were present, we will know how much the “dropped” parameter contributes to target cell subset identification. DIME thus further extends the multi-dimensional analysis to provide a quantitative measure of the contribution each parameter makes towards each identified cell subset of interest, giving an objective basis for panel and gating strategy design.

An objective, automatically computable measure of parameter usefulness has many benefits. In the context of clinical FCM assays in GCLP-compliant laboratories, DIME is useful for reducing the amount of trial-and-error in optimal gating strategy design by identifying the most informative parameters. As the number of parameters increases, the availability of DIME can significantly reduce the effort to find a target cell subset by identifying the subgroups of maximally informative parameters. The use of an objective basis for rationally designing a gating strategy is also likely to increase acceptance of that strategy by flow experts at different institutions. In turn, acceptance and use of a common gating strategy contributes to reducing the variability in multi-center studies. In some cases, DIME may reveal that certain parameters provide no additional information and can be dropped from the panel when cost is an issue. Panel reduction may also result in increasing the applicability of an assay by making the assay feasible on less sophisticated cytometers.

DIME can also be used to evaluate if different markers perform equivalently, and the potential loss or gain of sensitivity and specificity afforded with swapping markers. This can then be used to inform decisions on panel construction – for example, in cell sorting applications, it is necessary to use live cells, and hence a useful question that can be answered using discriminative measures could be “What is the impact of swapping an intracellular marker, that requires permeabilization and fixation to identify, with one or more cell surface markers, that may be used to identify viable cells?” Computationally, the discriminative measure provides a natural mechanism for *feature selection*, and we are currently investigating the extent to which this can be used for adaptive dimension reduction in clustering applications.

An obvious caveat is that DIME can only tell us about the usefulness of parameters for the particular data sets actually analyzed. If a parameter makes no contribution in the majority of data samples, but is critical for occasional anomalous samples, this will not be reflected by DIME unless the anomalous data samples are also analyzed. Clearly, DIME should not be used to exclude such parameters that are known to be of high informational value in anomalous data samples, but are otherwise redundant in typical data sets. However, if the test data set on which DIME is assessed is representative of the universe of data samples, then DIME provides robust information to guide decisions on marker inclusion, exclusion or exchange. Similarly, the measure of parameter usefulness is always with respect to a particular target cell subset. For different target cells, different parameters may be relevant, and this will be reflected in the DIME analysis of that target cell subset. In particular, this paper has evaluated the usefulness of DIME for a highly standardized data set, and the practicality of this statistical tool for less standardized “real world” data needs further confirmation.

It is also clear that DIME is only as good as the clustering algorithm, and will only provide a reliable guide if the cluster of interest has high sensitivity and specificity with respect to the “true” target cells. Finally, we admit that the use of DIME is highly restricted at present,

since it relies on an existing fitted statistical model of the data, and such analysis is at present only carried out in specialized research settings. However, research in statistical mixture modeling of flow data is rapidly progressing, and will become increasingly important and relevant as data dimensionality and throughput increase.

Fully automated gating techniques are currently being intensively researched at several institutions including ours, but manual gating is likely to remain the standard practice for some time. We show here that model-based approaches to automated cell subset identification can be complementary to manual gating, and provide useful information to guide both parameter selection and gating strategy design.

Acknowledgments

Research Support

Research partially supported by the National Science Foundation (DMS-0342172, Mike West) and the National Institutes of Health (RC1AI086032-01, UL1RR024128, NIH P50-GM081883, Lin Lin, Mike West, Cliburn Chan). Any opinions, findings and conclusions or recommendations expressed in this work are those of the authors and do not necessarily reflect the views of the NSF or NIH. The Center for Aids Research (CFAR) Flow Cytometry Core is supported by NIH grant 1P30 AI 64518.

References

1. Perfetto SP, Chattopadhyay PK, Roederer M. Seventeen-colour flow cytometry: unravelling the immune system. *Nat Rev Immunol.* 2004; 4(8):648–55. [PubMed: 15286731]
2. Maecker HT, Rinfret A, D'Souza P, Darden J, Roig E, Landry C, Hayes P, Birungi J, Anzala O, Garcia M, Harari A, Frank I, Baydo R, Baker M, Holbrook J, Ottinger J, Lamoreaux L, Epling CL, Sinclair E, Suni MA, Punt K, Calarota S, El-Bahi S, Alter G, Maila H, Kuta E, Cox J, Gray C, Altfeld M, Nougarede N, Boyer J, Tussey L, Tobery T, Brecht B, Roederer M, Koup R, Maino VC, Weinhold K, Pantaleo G, Gilmour J, Horton H, Sekaly RP. Standardization of cytokine flow cytometry assays. *BMC Immunol.* 2005; 6:13. [PubMed: 15978127]
3. Schenker EL, Hultin LE, Bauer KD, Ferbas J, Margolick JB, Giorgi JV. Evaluation of a dual-color flow cytometry immunophenotyping panel in a multicenter quality assurance program. *Cytometry.* 1993; 14(3):307–17. [PubMed: 8472607]
4. Nicholson J, Kidd P, Mandy F, Livnat D, Kagan J. Three-color supplement to the NIAID DAIDS guideline for flow cytometric immunophenotyping. *Cytometry-Communications in Clinical Cytometry.* 1996; 26(3):227–30. [PubMed: 8889396]
5. Schnitzlein-Bick CT, Mandy FF, O'Gorman MR, Paxton H, Nicholson JK, Hultin LE, Gelman RS, Wilkening CL, Livnat D. Use of CD45 gating in three and four-color flow cytometric immunophenotyping: guideline from the National Institute of Allergy and Infectious Diseases, Division of AIDS. *Cytometry.* 2002; 50(2):46–52. [PubMed: 12116345]
6. Schnitzlein-Bick CT, Spritzler J, Wilkening CL, Nicholson JK, O'Gorman MR. Evaluation of TruCount absolute-count tubes for determining CD4 and CD8 cell numbers in human immunodeficiency virus-positive adults. Site Investigators and The NIAID DAIDS New Technologies Evaluation Group. *Clin Diagn Lab Immunol.* 2000; 7(3):336–43. PMID: 95876. [PubMed: 10799443]
7. Reimann KA, O'Gorman MR, Spritzler J, Wilkening CL, Sabath DE, Helm K, Campbell DE. Multisite comparison of CD4 and CD8 T-lymphocyte counting by single- versus multiple-platform methodologies: evaluation of Beckman Coulter flow-count fluorospheres and the tetraONE system. The NIAID DAIDS New Technologies Evaluation Group. *Clin Diagn Lab Immunol.* 2000; 7(3): 344–51. PMID: 95877. [PubMed: 10799444]
8. Denny TN, Gelman R, Bergeron M, Landay A, Lam L, Louzao R, Mandy FF, Schmitz J, Spira T, Wilkening C, Glencross DK. A North American multilaboratory study of CD4 counts using flow cytometric panLeukogating (PLG): a NIAID-DAIDS Immunology Quality Assessment Program Study. *Cytometry B Clin Cytom.* 2008; 74B(Suppl 1):S52–64. [PubMed: 18351622]

9. Sarzotti-Kelsoe M, Cox J, Cleland N, Denny T, Hural J, Needham L, Ozaki D, Rodriguez-Chavez IR, Stevens G, Stiles T, Tarragona-Fiol T, Simkins A. Evaluation and recommendations on good clinical laboratory practice guidelines for phase I-III clinical trials. *PLoS Med.* 2009; 6(5):e1000067. [PubMed: 19536325]
10. Ezzelle J, Rodriguez-Chavez IR, Darden JM, Stirewalt M, Kunwar N, Hitchcock R, Walter T, D'Souza MP. Guidelines on good clinical laboratory practice: Bridging operations between research and clinical research laboratories. *J Pharmaceut Biomed.* 2008; 46(1):18–29.
11. Stiles T, Grant V, Mawbey T. Good clinical laboratory practice (GCLP). A quality system for laboratories that undertake the analyses of samples from clinical trials. 2003:1–18.
12. Mahnke YD, Roederer M. Optimizing a multicolor immunophenotyping assay. *Clin Lab Med.* 2007; 27(3):469–85. v. [PubMed: 17658403]
13. Owens M, Vall H, Hurley A, Wormsley S. Validation and quality control of immunophenotyping in clinical flow cytometry. *Journal of immunological methods.* 2000; 243(1-2):33–50. [PubMed: 10986405]
14. McLaughlin BE, Baumgarth N, Bigos M, Roederer M, De Rosa SC, Altman JD, Nixon DF, Ottinger J, Li J, Beckett L, Shacklett BL, Evans TG, Asmuth DM. Nine-color flow cytometry for accurate measurement of T cell subsets and cytokine responses. Part II: Panel performance across different instrument platforms. *Cytometry A.* 2008; 73A(5):411–20. [PubMed: 18383309]
15. McLaughlin BE, Baumgarth N, Bigos M, Roederer M, De Rosa SC, Altman JD, Nixon DF, Ottinger J, Oxford C, Evans TG, Asmuth DM. Nine-color flow cytometry for accurate measurement of T cell subsets and cytokine responses. Part I: Panel design by an empiric approach. *Cytometry A.* 2008; 73A(5):400–10. [PubMed: 18383316]
16. Maecker HT. Multiparameter flow cytometry monitoring of T cell responses. *Methods Mol Biol.* 2009; 485:375–91. [PubMed: 19020838]
17. Britten CM, Janetzki S, Ben-Porat L, Clay TM, Kalos M, Maecker H, Odunsi K, Pride M, Old L, Hoos A, Romero P. Harmonization guidelines for HLA-peptide multimer assays derived from results of a large scale international proficiency panel of the Cancer Vaccine Consortium. *Cancer Immunol Immunother.* 2009; 58(10):1701–13. PMID: 2714899. [PubMed: 19259668]
18. Roederer M. Spectral compensation for flow cytometry: visualization artifacts, limitations, and caveats. *Cytometry.* 2001; 45(3):194–205. [PubMed: 11746088]
19. Roederer M. How many events is enough? Are you positive? *Cytometry A.* 2008; 73A(5):384–5. [PubMed: 18307257]
20. Perfetto SP, Ambrozak D, Nguyen R, Chattopadhyay P, Roederer M. Quality assurance for polychromatic flow cytometry. *Nat Protoc.* 2006; 1(3):1522–30. [PubMed: 17406444]
21. Van Craenenbroeck EMF, Conraads VMA, Van Bockstaele DR, Haine SE, Vermeulen K, Van Tendeloo VF, Vrints CJ, Hoymans VY. Quantification of circulating endothelial progenitor cells: a methodological comparison of six flow cytometric approaches. *J Immunol Methods.* 2008; 332(1-2):31–40. [PubMed: 18255093]
22. Loken MR, Brosnan JM, Bach BA, Ault KA. Establishing optimal lymphocyte gates for immunophenotyping by flow cytometry. *Cytometry.* 1990; 11(4):453–9. [PubMed: 1693112]
23. Chan C, Feng F, Ottinger J, Foster D, West M, Kepler TB. Statistical mixture modeling for cell subtype identification in flow cytometry. *Cytometry A.* 2008; 73A(8):693–701. PMID: 2840702. [PubMed: 18496851]
24. Boedigheimer MJ, Ferbas J. Mixture modeling approach to flow cytometry data. *Cytometry A.* 2008; 73A(5):421–9. [PubMed: 18383311]
25. Pyne S, Hu X, Wang K, Rossin E, Lin TI, Maier LM, Baecher-Allan C, McLachlan GJ, Tamayo P, Hafler DA, De Jager PL, Mesirov JP. Automated high-dimensional flow cytometric data analysis. *Proc Natl Acad Sci U S A.* 2009; 106(21):8519–24. PMID: 2682540. [PubMed: 19443687]
26. Lo K, Brinkman RR, Gottardo R. Automated gating of flow cytometry data via robust model-based clustering. *Cytometry A.* 2008; 73A(4):321–32. [PubMed: 18307272]
27. Landry, CJM.; Ottinger, J.; Gagnon, D.; Enzor, J.; Hebert, V.; Sparks, S.; Yassine-Diab, B.; Sekaly, RP. Efforts toward validation of a standardized CFSE proliferation assay through inter-laboratory precision and robustness evaluation. . *CMI Standardization Techniques in Evaluation of Vaccine Response.* D'Souza, P., editor. Annecy; France: 2008.

28. Escobar MD, West M. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*. 1995; 90:577–88.
29. Ishwaran H, James LF. Approximate Dirichlet Process computing in finite normal mixtures. *Journal of Computational and Graphical Statistics*. 2002; 11(3):508–32.

Figure 1. (Top) Original group gating strategy. (Bottom) Location of non-proliferating CD4 (cyan), proliferating CD4 (blue), non-proliferating CD8 (violet) and proliferating CD8 (purple) cell subset clusters found with an MCMC algorithm are plotted on dot plot sequence of original gating strategy. Background PBMC events are shown in grey.

Figure 2.

Discriminative measure on simulated 3D data. The first 3 panels show red, blue and green clusters with 2 informative markers (markers 1 & 2) and 1 non-informative marker (marker 3). Marker 1 allows separation of the blue cluster from the others; marker 2 allows separation of the green cluster from the others; marker 3 is non-informative. Panel 1 shows the projection onto the informative marker 1 and 2 axes, and all 3 clusters are separable. Panel 2 shows the projection onto informative marker 1 and non-informative marker 3 – only the blue cluster can be uniquely identified. Panel 3 shows the projection onto informative marker 2 and non-informative marker 3 – only the green cluster can be uniquely identified. Panel 4 shows the value of the discriminative measure d_c with all markers, and after exclusion of markers 1, 2 and 3 respectively. Note that exclusion of marker 3 gives the same value as when all markers are present, confirming that marker 3 is non-informative.

Figure 3.

Modal clusters identified as CD4 non-proliferating (26), CD4 proliferating (29), CD8 non-proliferating (25) and CD8 proliferating (30) T lymphocytes obtained by clustering on all 8 dimensions. Proliferating (CFSE low) cells are in orange, while non-proliferating (CFSE high) cells are in reds. While the non-proliferating lymphocytes have classical small scatter characteristics, proliferating cells fall far outside the traditional “lymphocyte gate” on the FSC/SSC plot, and this adds significant complexity to the design of a gating strategy. The vertical histogram on the left panel illustrates how the ability to separate the CD4 from CD8 lymphocytes is lost when the CD4 marker is excluded, but dividing and non-dividing cells can still be separated on CFSE alone. DIME captures quantitatively the loss of ability to separate any given cell subset from the others when a parameter or set of parameters is excluded.

Figure 4.

A schematic showing the parameter subsets with the best discriminative ability for each target cell subset. DIME values were calculated for every combination of k parameters, and the parameter subset of size k with the *best* discriminative ability for each of the target cell subsets is shown. The top panel shows the color and positional encoding for the glyph used to represent parameters in the analysis. The next 4 panels show the change in discrimination for each target cell subset when the best combination of k markers is used (with k going from 8 to 1 horizontally). For each value of k , there are k parameters used to calculate DIME, chosen so that this particular combination gives the best discrimination out of all possible combinations of k markers. The particular k markers that give the best discrimination is indicated by the colored glyph, where *included* parameters are above the horizontal line, and *excluded* parameters below. For example, looking at the glyph in the Dividing CD4 panel (second from top), for $k=7$, the parameter excluded is FSC-H. In other words, to identify dividing CD4 events with 7 parameters, our analysis suggests that the optimal parameters are (FSC-A, SSC-A, CFSE, CD4, Amine, CD3, CD8). Note that surprisingly good discrimination is possible with only 4 or 5 parameters out of the original 8. Note that the level of discrimination is based on the target cell subset against *all* other subsets and not only among the CD4+ and CD8+ T cell subsets. For example, if CD8+ alone is sufficient to identify the majority of CD8+ T cells, and CD8+ proliferating cells constitute the majority of CD8+ cells, the discriminatory value can still be significant. In addition, it is even possible to discriminate to some extent between CFSE-high and CFSE-low CD8+ since the CFSE-low (proliferating) CD8+ T cells have on average higher expression levels of CD8 as compared with the non-proliferating population (same pattern for CD4+ T cells), as can be seen in the middle panel of Figure 7.

Figure 5.

DIME values were calculated for every combination of k parameters as in Figure 4, but now the parameter subset with the *least* discriminative information is shown. This shows how rapidly the ability to discriminate events from these target cell subsets is degraded when critical parameters are dropped.

Figure 6.

The effectiveness of DIME was checked by exhaustively calculating the sensitivity, specificity and accuracy of every parameter combination, using the event classification when all parameters are present as the reference set (“gold standard”). The figure shows the values for $k=8$ to $k=1$ optimal parameters using the same parameter sets shown in Figure 4, and confirms that sensitivity, specificity and accuracy are maintained at high levels and only begin to degrade rapidly below 4 parameters when the optimal parameter subsets given by DIME are used. Sensitivity is calculated as $\text{True Positive}/(\text{True Positive} + \text{False Negative})$, specificity as $\text{True Negative}/(\text{True Negative} + \text{False Positive})$ and accuracy as $\text{Subset Fraction} \times \text{Sensitivity} + \text{Specificity} \times (1 - \text{Subset Fraction})$. Note that the scale on the y-axis is not the same for the three plots.

Figure 7. Design of a simplified 3-generation gating strategy using 5 parameters suggested by use of DIME and cell subsets identified with mixture model analysis. In each step, the red labeled clusters were discarded and yellow labeled clusters retained.

Figure 8.

Comparison of simplified gating strategy (red symbols) with the original 9-generation gating strategy (blue symbols) on all the CFSE data sets used in a single experiment across 3 different labs (BD=square, Duke=triangle, NIML=circle) under 3 different treatment conditions (unstimulated, CMVpp65 peptide mix (15mers overlapping by 11aa) and anti-CD3/CD28 mAbs). The percentages of dividing and non-dividing CD4+ and CD8+ T lymphocyte subsets found were similar even though the gating strategies were markedly different. Samples labeled “Unstim” are prepared in the same way as the other samples but without CMVpp65 or anti-CD3/CD28 beads.