Cytometry
PART A

High-Throughput Flow Cytometry Data Normalization for Clinical Trials

Greg Finak,[1*] Wenxin Jiang,[1] Kevin Krouse,[2] Chungwen Wei,[3] Ignacio Sanz,[3] Deborah Phippard,[4] Adam Asare,[4] Stephen C. De Rosa,[1,5] Steve Self,[1,6] Raphael Gottardo[1,7]

[1]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109

[2]LabKey Software, Seattle, Washington 98105

[3]Division of Rheumatology, Lowance Center for Human Immunology, Emory University School of Medicine, Atlanta, Georgia 30322

[4]Immune Tolerance Network, Bethesda, Maryland 20814

[5]Department of Laboratory Medicine, University of Washington, Seattle, Washington 98195

[6]Department of Biostatistics, University of Washington, Seattle, Washington 98195

[7]Department of Statistics, University of Washington, Seattle, Washington 98195

Conflicts of interest: Kevin Krouse is an employee of LabKey Software.

Received 11 September 2013; Revised 18 November 2013; Accepted 13 December 2013

Grant sponsor: NIH, Grant numbers: R01 EB008400, U01 AI068635, UM1 AI068618, N01 AI15416, P01 AI078907, P30 AI027757

Grant sponsor: Bill and Melinda Gates Foundation [through the Collaboration for Aids Vaccine Discovery (CAVD)]; Grant number: OPP1032317; Grant sponsors: NIAID, Public Health Service, University of Washington Center for AIDS Research.

Additional Supporting Information may be found in the online version of this article.

Correspondence to: Greg Finak, Vaccine and Infectious Disease Division, Fred

ISAC
International Society for Advancement of Cytometry

• **Abstract**

Flow cytometry datasets from clinical trials generate very large datasets and are usually highly standardized, focusing on endpoints that are well defined apriori. Staining variability of individual makers is not uncommon and complicates manual gating, requiring the analyst to adapt gates for each sample, which is unwieldy for large datasets. It can lead to unreliable measurements, especially if a template-gating approach is used without further correction to the gates. In this article, a computational framework is presented for normalizing the fluorescence intensity of multiple markers in specific cell populations across samples that is suitable for high-throughput processing of large clinical trial datasets. Previous approaches to normalization have been global and applied to all cells or data with debris removed. They provided no mechanism to handle specific cell subsets. This approach integrates tightly with the gating process so that normalization is performed during gating and is local to the specific cell subsets exhibiting variability. This improves peak alignment and the performance of the algorithm. The performance of this algorithm is demonstrated on two clinical trial datasets from the HIV Vaccine Trials Network (HVTN) and the Immune Tolerance Network (ITN). In the ITN data set we show that local normalization combined with template gating can account for sample-to-sample variability as effectively as manual gating. In the HVTN dataset, it is shown that local normalization mitigates false-positive vaccine response calls in an intracellular cytokine staining assay. In both datasets, local normalization performs better than global normalization. The normalization framework allows the use of template gates even in the presence of sample-to-sample staining variability, mitigates the subjectivity and bias of manual gating, and decreases the time necessary to analyze large datasets.

• **Key terms**

bias; BioConductor; immunology; template gating; staining variability

Introduction

Flow cytometry studies in clinical trials often generate large quantities of data. However, these studies are usually focused on specific, predefined cell subpopulations and endpoints. Consequently, a manual gating approach to data analysis is often tedious, repetitive, and time consuming. However, these same qualities make the data well suited to some form of automated analysis (1–3). The maturity of automated gating approaches for flow cytometry data has recently been demonstrated (3). Unfortunately, these fully automated algorithms are not entirely suitable for application in a clinical setting, where standardization and interpretability of cell subsets are important. Flow assays (which includes the data analysis process) in a clinical setting often need to be qualified to provide some guarantees of accuracy (4). Unless an assay is well standardized and well controlled, the gating will be difficult to fully automate due to technical and biological variation that manifests as changes in

Hutchinson Cancer Research Center, 1100 Fairview Ave N, Seattle, WA 98109, USA. E-mail: gfinak@fhcrc.org

Published online 27 December 2013 in Wiley Online Library (wileyonlinelibrary.com)

DOI: 10.1002/cyto.a.22433

the position, separation, and density of cell populations from sample to sample. This can negatively impact the robustness of fully automated (i.e., high-dimensional) gating methods that need to match cell populations across samples after gating.

This variability can be caused by suboptimal antibodies and fluorescence marker combinations, instrument variability over time, changes in reagent lots, poor sample handling, or any other number of variables. In general, it is an effect that is difficult to perfectly control and is often an unavoidable nuisance in flow cytometry studies of significant size. Here, we refer to these effects in general as sample-to-sample variation in observed staining intensity. These effects preclude not only fully automated analysis but also simpler, less computationally intensive approaches to speedup data analysis such as template gating, where a single set of gates can be copied and reused across samples (5).

Consequently, variability across samples complicates the data analysis process. Regardless of the approach, the analyst must manually verify and adjust the gates for each sample to ensure that subpopulations of interest are correctly gated. This consequence is an increased workload for the data analyst and an increased chance to introduce errors into the downstream analysis through misplaced gates.

There are few tools to mitigate these problems. An algorithm for flow cytometry data normalization based on the ideas behind image warping was recently introduced (6,7). The algorithm works with the one-dimensional marginal densities of the channels, identifies regions of significant curvature (i.e., peaks), and applies techniques from functional data analysis to match and align the peaks via a nonlinear transformation. A shortcoming of this approach is that it operates on the marginal (i.e., independent of the gating scheme) distributions of the fluorescence intensities of each channel. Normalization is global, applied to all cells in a channel, and ignores the gating structure used to define distinct cell subsets of interest. Consequently, the algorithm can fail to correctly normalize cell populations appearing lower in the gating hierarchy as, marginally, their distributions are masked by other cells. This effect can be significant even if the data are pregated to remove debris. Other technical issues are related to the implementation, which is limited to working with data in memory and restricts the maximum size of datasets that can be effectively processed. Although it is natural to combine data normalization with template gating and to normalize a dataset to a specific, accurately gated target sample, the current implementation aligns features to their average position across samples. This implies that data need to be regated after normalization.

Here, we describe extensions to the *fdaNorm* normalization algorithm that address these limitations and make the application flow data normalization more practical, robust, and applicable to large, real-world datasets (7). Most importantly, we address the issue of robustness by making normalization local rather than global (i.e., we apply normalization to specific gates) by integrating it into the gating procedure. Second, we extended the implementation to allow normalization to a template sample thus facilitating the use of template gates, and finally, we enabled the algorithm to work with large, real-world datasets by implementing support for disk-based storage of data via Network Common Data Form (NetCDF) files (8). These improvements enable the use of template gates to rapidly analyze large studies where many samples need to be gated in an identical manner, even in the presence of substantial sample-to-sample variability. We demonstrate the effectiveness of this improved algorithm on two real-world clinical trial datasets and compare it with both manual gating and global normalization.

Materials and Methods

Datasets

We analyzed two datasets for this study. The first is a B-cell phenotyping dataset from the Immune Tolerance Network examining the abundances of different B-cell phenotypes in healthy subjects. The dataset consisted of 33 FCS files with a total size of 11 Gb. The data were stained with a panel of 12 antibody-fluorochrome conjugates and acquired on a LSRII flow cytometer (BD Bioscience). The data were exported in a FCS 3.0 format and had a bin resolution of 262,144 (2^{18}). Our normalization algorithm was applied to three channels, CD3, CD19, and CD27, and Mito Tracker Green FM (MTG; defining six gates, B-cells, unstimulated memory, stimulated memory, naive and transitional, double-negative, and MTG+), which exhibited variability across subjects. We compared the variability and bias of cell subpopulations defined using template gates, template gates with local normalization and template gates with global normalization, against manual gates that were carefully adjusted by hand. For this dataset, global normalization was applied to the same channels as local normalization; however, the data were pregated data to the level of the lymphocyte cell subset as recommended in the original publication (7). The dataset is available on flowrepository.org under ID FR-FCM-ZZ7P.

The second dataset is a flow cytometry dataset from a clinical trial conducted within the HIV Vaccine Trials Network (HVTN). The dataset is a subset of a 10-color intracellular cytokine staining (ICS) dataset from a Phase I HIV vaccine trial (HVTN080) (9). We refer readers to that publication for further details on the study. Technical details on instrument settings for data collection and the ICS assay used by the HVTN are relevant for both the 10-color and 12-color panels used by the HVTN (10). We examined data from 48 healthy individuals assayed at Day 0 (Visit Code 2) and at 40 days

postvaccination (Visit Code 5). We focused on three antigen challenge pools (one GAG and two POL peptide pools) for a total of 470 FCS files and 13.2 GB in size. These data were gated using manually curated template gates for all but one channel, <APC-A> perforin, which exhibited substantial sample-to-sample variability in staining intensity and was thus ignored. We examined the effect of normalization on the variability of cell population statistics based on the <APC-A> perforin gate, as well as the effect on subsequent positivity calls using those cell subsets. We also compared our approach with global normalization, which was applied to pregated data on the lymphocyte subset. The dataset is available on flowrepository.org under ID FR-FCM-ZZ7U. Both studies were approved by institutional ethical review boards (IRB).

Evaluation of the B-Cell Phenotyping Data

We evaluated the B-cell phenotyping dataset by looking at the bias in the extracted population statistics relative to manual gating. For each FCS file, we computed the difference between the extracted population proportion (relative to the parent population) from manual gating and from template gating without normalization, with local normalization, and with global normalization. A desirable outcome is to have low bias and low variability relative to the manual gates.

Evaluation and Positivity Calls in the ICS Data

In the ICS data, we could not look at the bias for the perforin gate because this gate was ignored in the original dataset and was never manually adjusted to account for the variability in the data. However, we chose instead to look at the effect of normalization on positivity calls using the normalized and unnormalized cell subset. Positivity calls were made for each subject within each combination of cytokine $\times$ stimulation using a one-sided Fisher's exact test on the positive cell counts in negative controls and paired stimulated samples (11). Positive responders were identified as observations with q-values below the 1% false discovery rate threshold. Multiplicity was adjusted globally (over all subjects, cytokines, and stimulations).

Extensions to the Normalization Algorithm

We have extended the normalization algorithm of Hahne et al. as described below to make normalization more robust when applied to small cell subsets (6,7).

Integration with manual gating. We have extended the normalization routine to integrate with manual gates that can be imported from external tools (i.e., FlowJo > 9.6.3, TreeStar, Ashland, OR) using our `flowWorkspace` (version 3.8.0) BioConductor package. The routine uses the structure of the gating hierarchy to alternately perform normalization and gating on each subpopulation in the data. This refinement moves away from global normalization and allows local normalization (conditional on cell subset and channel of interest) of specific channels for specific cell subpopulations in gates that are adversely affected by sample-to-sample variability.

Normalization against a reference. A further extension allows the normalization routine to use a reference or template sample to define a target distribution to which other samples should be normalized. This can be an FCS file that has been carefully and accurately gated before normalization is applied to the data. This sample provides both a set of template gates and a set of reference distributions for cell subpopulations at each gate that are used as a target distribution for peak matching and alignment. After normalization, the template gates are copied over to the normalized data.

Handling large datasets. We have integrated our normalization algorithm into the `flowStats` BioConductor package, which makes use of the `flowWorkspace` and `ncdfFlow` infrastructure to handle large flow datasets with minimal memory overhead by leveraging the disk-based NetCDF scientific data format and allowing for high-throughput data processing. This infrastructure has been previously described (12).

Normalization Algorithm

Our normalization algorithm builds on the density-based, per-channel normalization approach (*fdaNorm* from the *flowStats* BioConductor Package) (7). Briefly, that approach estimates the density for each channel and each sample to be normalized, locates major peaks for each sample, then matches the peaks across samples, and aligns them using techniques from functional data analysis. This peak identification and registration step is global, in the sense that it is performed once per FCS file per channel on the ungated data, but does not take into account the structure of the gating hierarchy. Consequently, overlapping subpopulations in the data have the potential to confound the peak matching step in some instances. Indeed, the authors mention this problem in their publication and suggest preliminary gating of debris as a potential solution; however, they do not provide a mechanism to do so in practice (7). Nonetheless, we have undertaken this latter approach in our comparison.

Local normalization. Here, we make use of an available gating hierarchy for the data to target normalization to local regions of the gating hierarchy (specific cell subsets and channels) as needed. We use the gating hierarchy of manual gates, together with a predefined reference sample as a template. The user specifies which cell subpopulations and which channels within the gating hierarchy should be normalized during the gating process. The algorithm then performs the stepwise, hierarchical gating of each cell subpopulation in the gating scheme on each sample. When the algorithm reaches a cell subpopulation in the gating hierarchy that requires normalization, it applies the normalization routine to the parent population and uses the distribution of the reference sample as a target to identify, match, and align peaks in the other samples as originally described (7). By performing this landmark identification and registration locally, the distributions of cells within each channel are more comparable across samples, as cells upstream in the gating hierarchy are not masking the location of the cell subsets of interest. This local approach makes the normalization algorithm more robust. After normalization, the algorithm applies the template gates from the reference sample to the

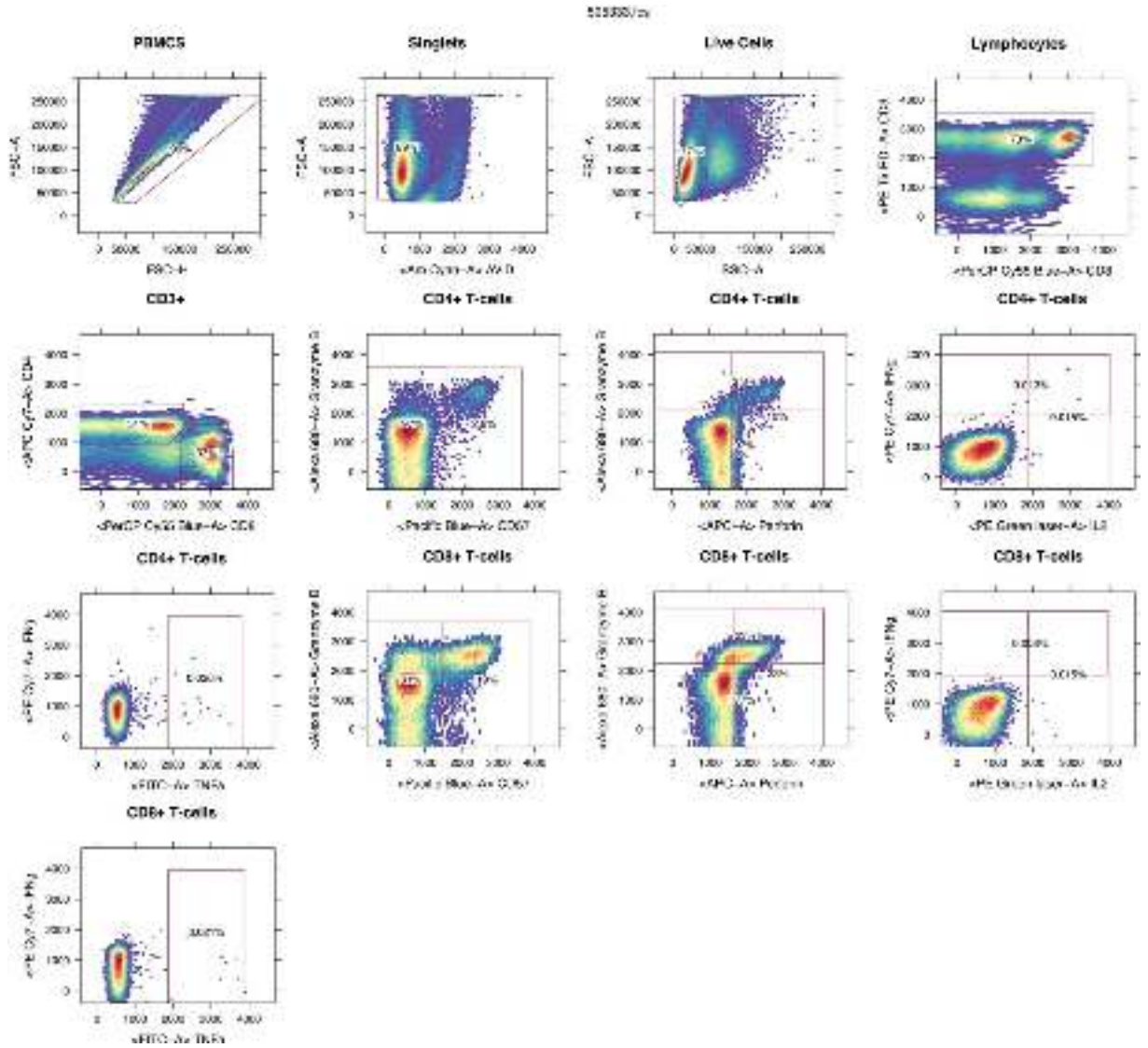

Figure 1. Gating scheme for an FCS file from the ICS dataset. In this example, cells are stimulated with POL-1-PTEG peptide pool. Problematic gating of the perforin channel for both CD4 and CD8 subsets is clearly visible. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

normalized cell subpopulation, recomputes the cell subset membership, and continues the gating process.

Assumptions and caveats. The normalization approach described here is based on assumptions and has limitations that must be clearly understood. The basic assumption is that the data to be normalized were properly acquired and compensated. The normalization algorithm cannot correct for poor quality data. Rather it can correct for day-to-day variation in MFI due to less-than-optimal assay standardization or other sources of technical variation that manifest as MFI variability, where the cell populations are nonetheless properly stained and resolved. These are the types of issues that arise in the datasets analyzed here. Normalization is meant to standardize the data, making it amenable to high-throughput analysis via template gating or automated gating approaches.

Results

We applied our local normalization algorithm to the ICS and B-cell phenotyping datasets described in the "Materials and Methods" section and compared it with global normalization and template gating without normalization.

Template and Manual Gating of Datasets

The ICS data were gated using template gates with minor manual adjustments for all but the <APC-A> perforin channel, which was deemed too variable to be usable in the original analysis of the data. In contrast, the B-cell phenotyping data were gated using manual gates with careful adjustments to account for sample-to-sample variability. A representative sample from the ICS dataset showing the complete gating scheme is shown in Figure 1, and the gating scheme for the B-cell is shown in Supporting Information Figure S1.

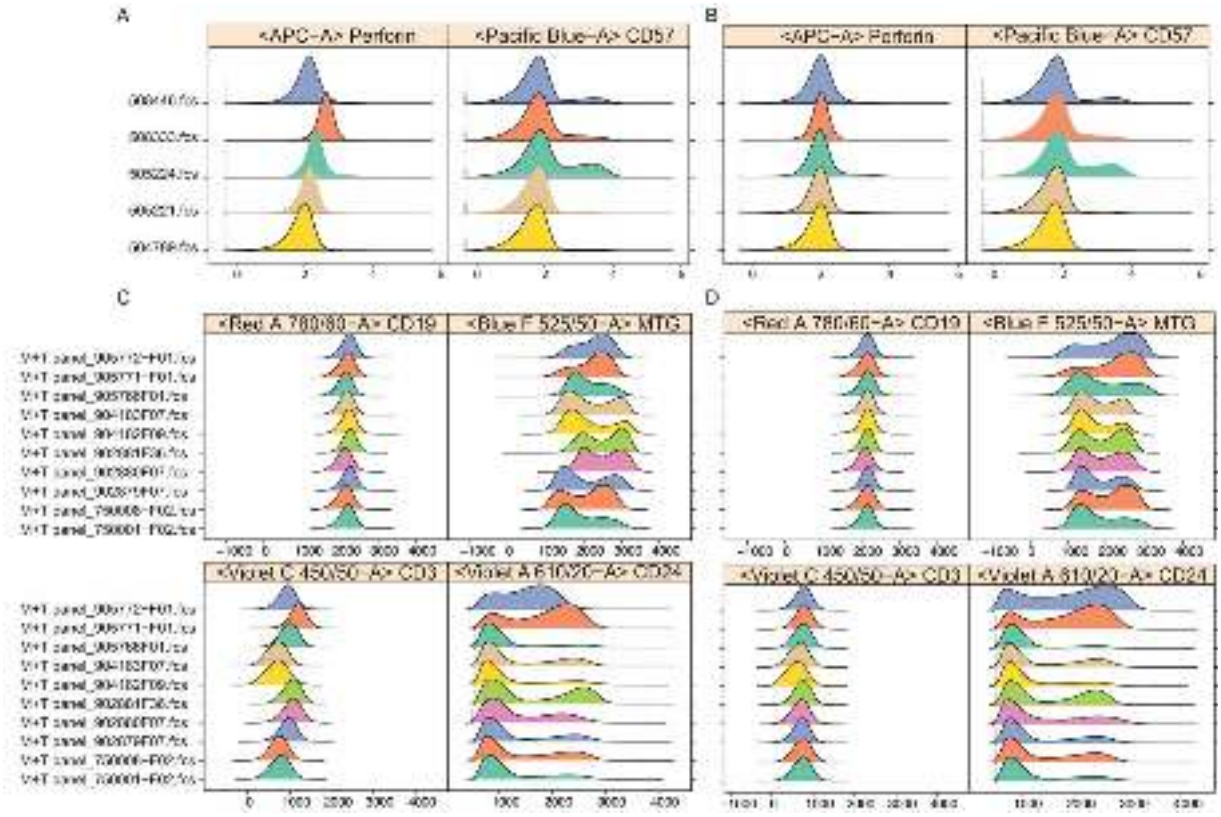

Figure 2. Variability in the staining intensity of channels in the ICS and B-cell phenotyping datasets. In the ICS data, the <APC-A> perforin channel is compared with <Pacific Blue-A> CD57, before and after normalization. A random sampling of 12 FCS files showing the variability in staining on the <APC-A> channel (**A**) before and (**B**) after normalization, contrasted against the staining of <Pacific Blue-A>, which does not exhibit staining variability. The density is shown for the CD4+ cell subpopulation, right before the perforin+ gate is applied. In the B-cell phenotyping data, the density of different channels in the (**C**) locally normalized and the (**D**) raw data is shown. Variability across subjects is evident for the Red-A, Violet-C, Violet-A, and Blue-F channels, and their variability is reduced across subjects after normalization. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Examination of the ICS data revealed that the <APC-A> perforin channel exhibited substantial variability in the MFI of perforin-negative cells across samples, whereas other channels did not, as shown in Figure 2A. The variability in MFI was decreased after local normalization of the CD4 and CD8 T-cell parent populations (Fig. 2B). Similarly, for the B-cell phenotyping data, the Red-A, Violet-C, Violet-A, and Blue-F channels exhibited substantial variability across subjects (Fig. 2D) and were normalized to a reference sample (M+T panel_750008-F02.fcs) using our local normalization algorithm applied at the CD19 parent population level resulting in decreased variation across subjects postnormalization for those markers (Fig. 2C). The decrease in variability of MFI makes the data more amenable to template gating.

Local Normalization Improves Accuracy of Template Gating

The effect of local normalization on the proportion of perforin-positive T cells can be observed by examining the pairwise dotplots of perforin gates. Representative samples are shown in Figures 3A–3D. Before normalization, the template perforin+ gate captures perforin-negative T cells in both the CD4+ and CD8+ subsets (Figs. 3A and 3B) for some samples, but not for the reference sample. After normalization to the reference sample, the perforin+ template gates are correctly placed for both subsets of T cells (Figs. 3C and 3D).

Normalization Reduces Variability and Bias of Template-Gated Cell Population Statistics

Perforin is a constitutively expressed, CD8-T-cell-specific protein that is not significantly expressed on CD4 T cells (13). Local normalization reduced the variability of the measured proportion of CD4+ and CD8+ perforin+ T cells, as shown in Figure 4A. This is particularly evident for the CD4+ T cells, where perforin is not expected to show significant expression. In contrast to local normalization, global normalization (Fig. 4B) inflates the observed number of CD4-positive, perforin-positive T cells. The poor performance of global normalization is caused by differences in the placement of CD4 and CD8 perforin template gates, which requires conditioning on the cell population for accurate peak alignment.

In the B-cell phenotyping dataset, we compared cell population statistics of template-gated, locally normalized, and globally normalized data against manually gated data (with

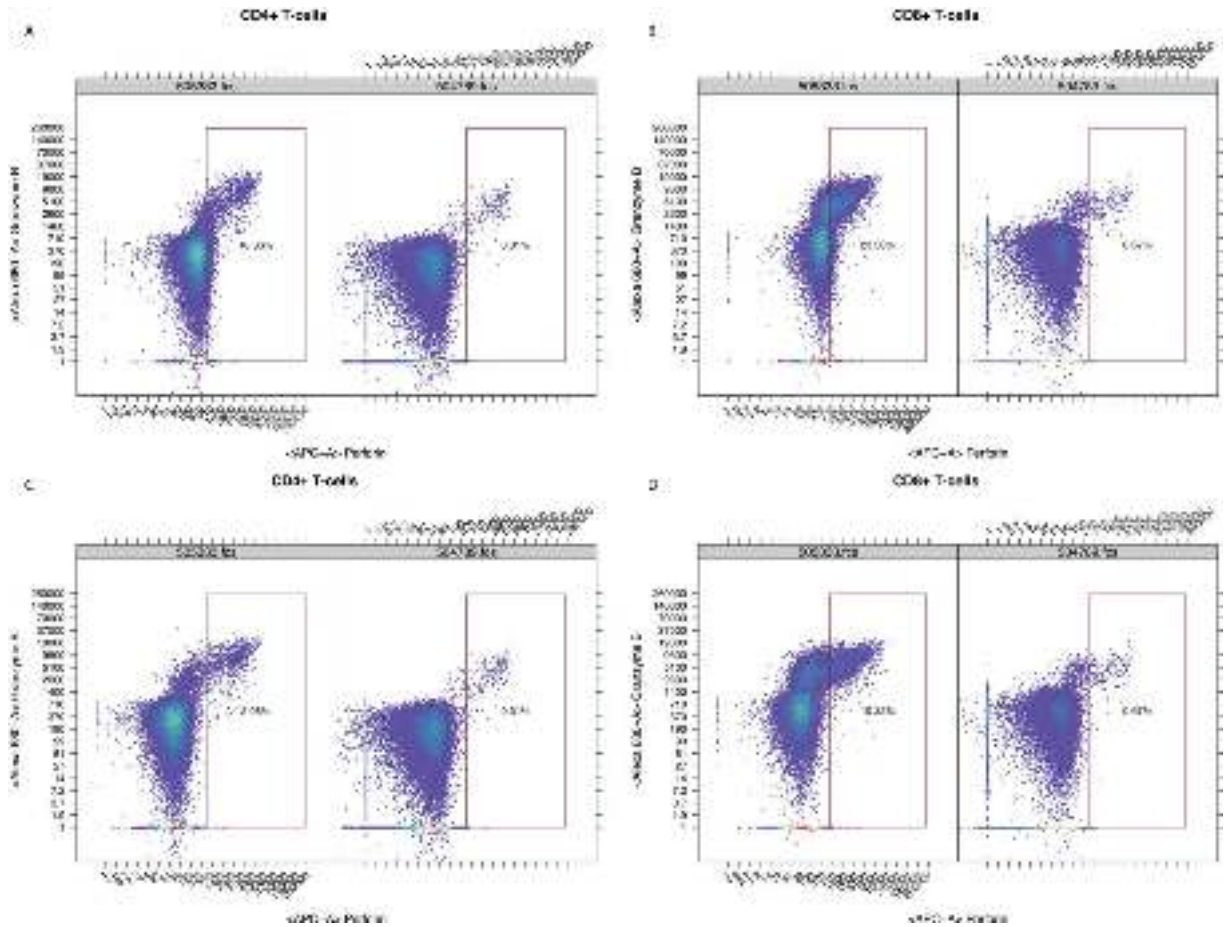

Figure 3. Percent of perforin-positive CD4+ and CD8+ T cells before (**A** and **B**) and after (**C** and **D**) normalization in a sample FCS file (505333.fcs) and the target reference file (504789.fcs). The reference is unchanged after normalization. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

local adjustments), evaluating the bias and variability. Boxplots of these statistics are shown in Supporting Information Figure S2A. Local normalization with template gating shows a clear improvement over both global normalization and template gating alone, particularly for the MTG gate. This effect is even more evident when examining paired differences between manual gating and the competing methods (Fig. 4C), where local normalization with template gating is the only approach that exhibits both low variability and low bias relative to manual gating. These results demonstrate that local, gate-specific normalization combined with a template gating approach can be used effectively to account for sample-to-sample variability in a manner analogous to careful manual gating.

Normalizaton Reduces False Positives in ICS Assay Positivity Calls

Although low variability is a desirable quality, we can also evaluate more objective criteria for the ICS dataset. Here, we compared positivity calls (see "Materials and Methods" section) for CD4+ and CD8+ T-cell subsets coexpressing perforin with other cytokines (IL2, IFN-γ, or TNF-α), for each antigen stimulation at two time points (Visit 2, Day 0, prevaccination, and Visit 5, Day 40, 2 weeks postvaccination). Positive responses were called at a q-value of 1% (14) from Fisher's exact test, adjusted across all subjects. Because global normalization exhibited poor template gating, we did not evaluate it further for positivity calls.

The difference in the proportion of positive cells between stimulated and unstimulated samples, conditional on normalization, stimulation, cytokine, T-cell subset, and time point is shown in Table 1 and Supporting Information Figure S3. We do not expect to observe substantial positivity in the CD4+ T-cell subset or positivity on Day 0 (before vaccination). For the unnormalized data, there are a total of 20 false-positive response calls, 18 in the CD4+ subset at Day 40, and 2 in the CD8+ subset at Day 0. After normalization, these positive calls are eliminated. The number of positive response calls observed before normalization on Day 40 in the CD8+ subset is reduced (from 14 prenormalization to 9 postnormalization; Supporting Information Fig. S3). However, the lower overall false-positive rate gives us higher confidence in these remaining calls.

To verify that the samples called positive in the CD4+ T-cell subset prior to normalization are indeed false positives,

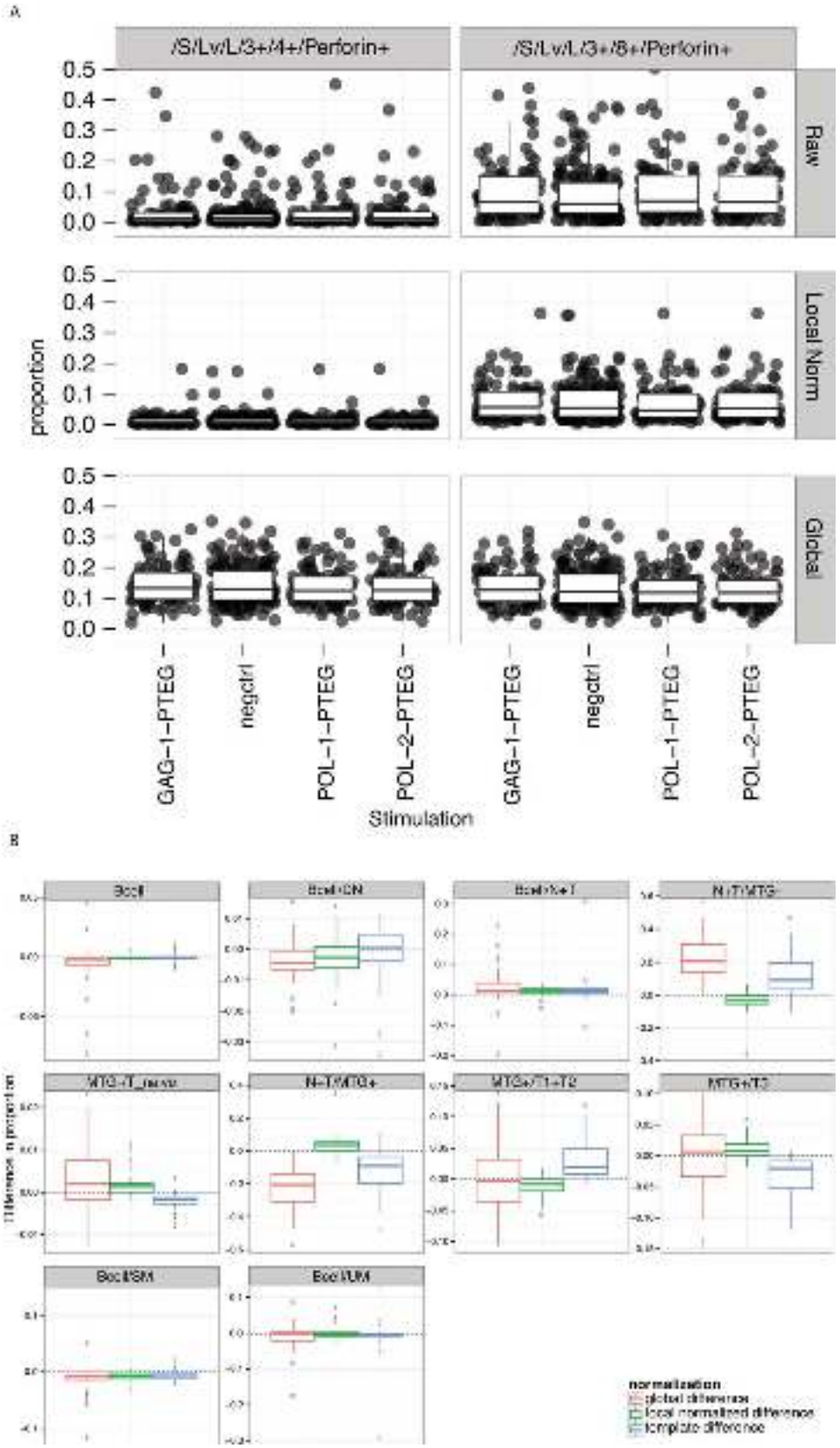

Figure 4. Comparison of global and local normalization approaches for the ICS data and B-cell phenotyping data. (**A**) The proportion of perforin-positive CD4- and CD8-positive T cells before and after local and global normalization. Local normalization leads to accurate gating of the CD4 and CD8 perforin-positive T cells, whereas global normalization inflates the number of CD4 and CD8 perforin-positive T cells. Variability is expected to decrease within stimulation groups and T-cell subsets for this dataset on accurate gating. (**B**) Paired differences between manually gated and normalized, globally normalized, or template-gated cell populations in the B-cell phenotyping data. Dotted gray line at zero is a guide to the eye. Differences closer to zero are better because they imply less bias. Narrower boxes imply greater precision. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Table 1. Response rates for different perforin+ cytokine+ (double positive) T-cell subset combinations in the HVTN ICS dataset before and after normalization

NORMALIZATION	ANTIGEN	CYTOKINE (PERFORIN DOUBLE POSITIVE)	CD4 DAY 0	CD4 DAY 40	CD8 DAY 0	CD8 DAY 40
Normalized	GAG-1-PTEG	IFN-γ	0.00	0.00	0.00	0.00
Normalized	GAG-1-PTEG	IFN-γ or IL2	0.00	0.00	0.00	0.00
Normalized	GAG-1-PTEG	IL2	0.00	0.00	0.00	0.00
Normalized	GAG-1-PTEG	TNF-α	0.00	0.00	0.00	0.00
Normalized	POL-1-PTEG	IFN-γ	0.00	0.00	0.00	0.00
Normalized	POL-1-PTEG	IFN-γ or IL2	0.00	0.00	0.00	0.00
Normalized	POL-1-PTEG	IL2	0.00	0.00	0.00	0.00
Normalized	POL-1-PTEG	TNF-α	0.00	0.00	0.00	7.69
Normalized	POL-2-PTEG	IFN-γ	0.00	0.00	0.00	6.67
Normalized	POL-2-PTEG	IFN-γ or IL2	0.00	0.00	0.00	5.88
Normalized	POL-2-PTEG	IL2	0.00	0.00	0.00	0.00
Normalized	POL-2-PTEG	TNF-α	6.25	0.00	11.11	9.09
Raw	GAG-1-PTEG	IFN-γ	9.09	0.00	10.00	0.00
Raw	GAG-1-PTEG	IFN-γ or IL2	5.56	4.76	7.69	0.00
Raw	GAG-1-PTEG	IL2	0.00	7.14	0.00	0.00
Raw	GAG-1-PTEG	TNF-α	5.00	5.56	0.00	0.00
Raw	POL-1-PTEG	IFN-γ	7.14	15.00	0.00	25.00
Raw	POL-1-PTEG	IFN-γ or IL2	11.11	12.50	0.00	23.08
Raw	POL-1-PTEG	IL2	0.00	15.38	0.00	0.00
Raw	POL-1-PTEG	TNF-α	18.75	14.29	9.09	17.65
Raw	POL-2-PTEG	IFN-γ	6.67	0.00	12.50	13.33
Raw	POL-2-PTEG	IFN-γ or IL2	8.70	5.00	8.33	11.11
Raw	POL-2-PTEG	IL2	0.00	9.09	0.00	0.00
Raw	POL-2-PTEG	TNF-α	12.00	5.26	16.67	15.38

No response is observed for Gag at the Day 0 or Day 40 time points after normalization.

we examined the gating of these samples before and after normalization. The comparative gates for two potential false positives in the POL-1-PTEG stimulation are shown in Figure 5. The gates of the raw data clearly include cells that are negative for perforin, whereas the gates of the normalized data include only perforin-positive cells, indicating that the responder calls before normalization are false positive.

Integration into LabKey

We have integrated the normalization pipeline into the LabKey tool (LabKey Software, Seattle, WA), thus making it accessible to a wider array of potential users who may lack the expertise to implement a normalization pipeline in R and BioConductor (15,16). Users can select gates and channels for normalization from within the LabKey flow cytometry module import pipeline.

Discussion

We have developed an improved flow cytometry data normalization algorithm and pipeline capable of handling the large datasets commonly encountered in vaccine clinical trials. The algorithm and pipeline are implemented in R and make use of the BioConductor infrastructure packages available for processing large volumes of flow cytometry data (*ncdfFlow*) as well as the `flowWorkspace` package for importing manual gates into the R framework. By making manual gates and large datasets available to the normalization algorithm, we have been able to improve its performance and practical applicability as demonstrated on two flow datasets from clinical trials. Neither the global nor the local normalization approaches could have been applied to the datasets in this study without the NetCDF support built into the infrastructure (8).

Our normalization algorithm makes use of the manual gating hierarchy to perform local normalization of cell subsets within specific gates, rather than globally as previously described. Performing normalization locally makes the peak identification and alignment steps more robust by eliminating multiple peaks from overlapping cell subsets higher in the hierarchy.

A further benefit of our approach is an improved workflow which allows an analyst to carefully gate only a single sample. This sample can then be used as a reference for normalization algorithm and acts as a source of template gates for the remaining samples in the dataset. We demonstrated this approach on the B-cell phenotyping dataset and showed that local normalization is unbiased and has low variability when compared against manual gating. In contrast, global normalization and template gating without normalization exhibited biased population statistics with larger variability.

Although our approach decreases staining variability across samples for specific channels and cell subsets, the cell

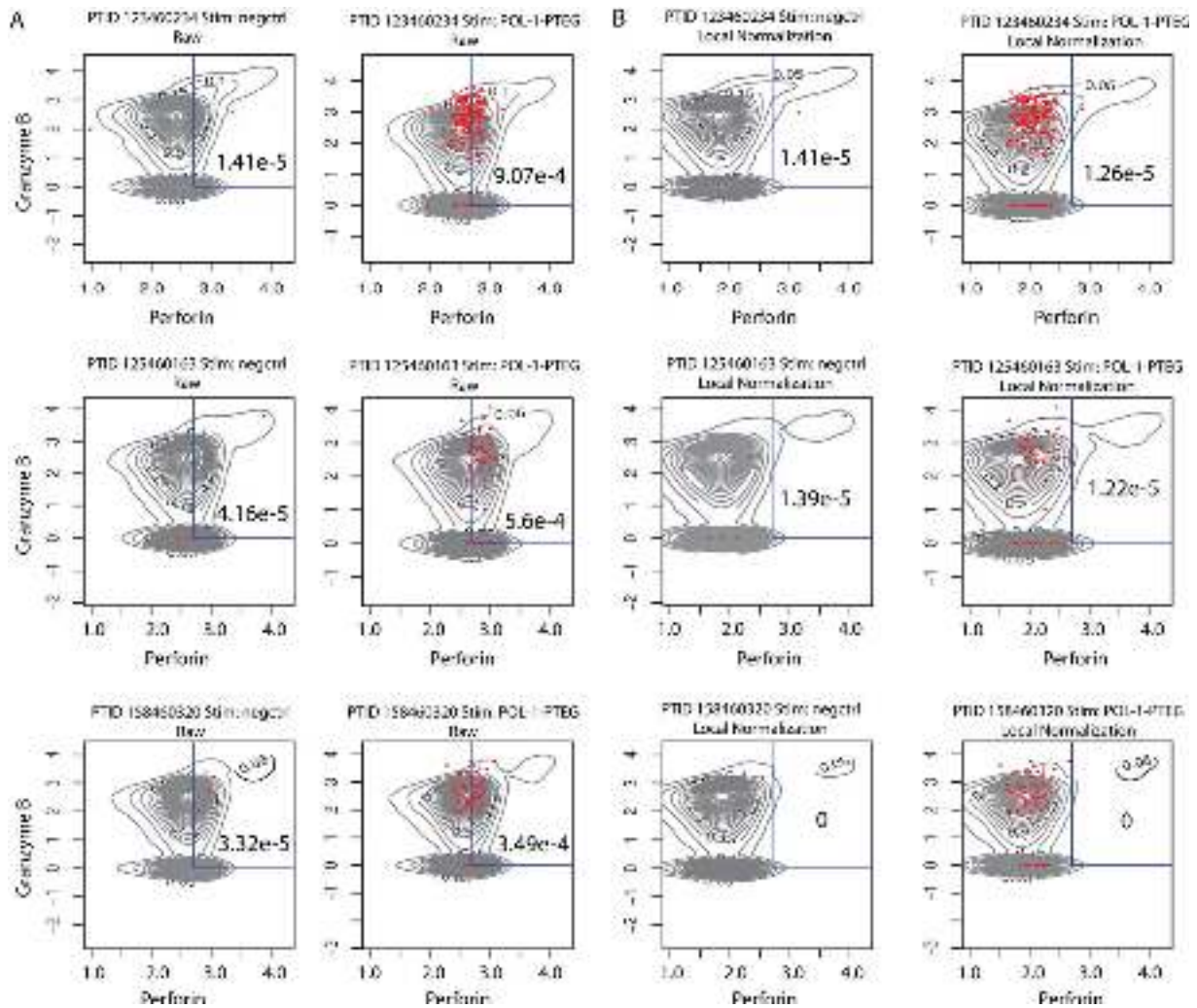

Figure 5. Comparative gating of three potential false-positive responders in the IFN-γ-producing CD4+ T-cell subset stimulated with POL-1-PTEG. Gating of the (**A**) normalized data and (**B**) raw data with associated proportions of perforin-positive cells. Red dots indicate IFN-γ-positive T cells. [Color figure can be viewed in the online issue which is available at wileyonlinelibrary.com.]

subpopulation statistics extracted from template gating have also decreased variability indicating that the algorithm is performing as expected. This is critically important for downstream analysis where increased variation can decrease statistical power. However, care is warranted when normalizing data to ensure basic assumptions are met, namely that the MFI distribution of the cell subset to normalize is relatively consistent across samples (i.e., apart from shifts in the MFI, a bimodal distribution does not become trimodal in some samples), otherwise peak matching may still be error prone. To account for such (potentially biological) effects, the algorithm can be applied separately to subsets of the data (e.g., by treatment condition). In ICS assays, where the increased production of antigen-specific T-cell subsets is measured in response to antigen stimulation (17), careful gating is critical as cell subsets can be rare, and small changes in cell counts can lead to significant differences cell proportions. We have shown that our local, gate-specific normalization approach is able to decrease the bias in template gating of perforin-positive T-cell populations, producing meaningful and measurable reductions in the number of false-positive assays across multiple T-cell subsets and time points. Specifically, the absence of detectable response in the CD4 T-cell subset after normalization is consistent with the CD8 specificity of perforin expression, and the loss of responses at the prevaccination time points after normalization is consistent with the expected lack of antigen specificity prior to vaccination. Finally, although the number of postvaccination CD8 responses decreased after normalization, the increased specificity of the assay overall gave higher confidence to those responses. The lost responses were confirmed as false positives by examining dotplots of the raw data, which demonstrated the presence of perforin-negative cells within the perforin+ gates. Importantly, we note that if the MFI of a cell population (after gating) is of primary interest in an assay, then this normalization approach is likely not suitable to correct for technical artifacts.

Finally, the framework presented here could be used to support an automated, hierarchical gating pipeline. Such an approach could be extremely useful to support efforts such as the Human Immune Phenotyping Consortium Lyplate flow data standardization effort currently underway (18). We are exploring such extensions as ongoing work.

Our algorithm has been implemented in the updated `flowStats` package (version 1.19.9) available at https://github.com/RGLab/flowStats and is integrated with the LabKey tool.

Acknowledgments

The authors thank the James B. Pendleton Charitable Trust for a generous equipment donation. This research was performed with the support of the Immune Tolerance Network, an international clinical research consortium headquartered at the University of California San Francisco and supported by the National Institute of Allergy and Infectious Diseases and the Juvenile Diabetes Research Foundation.

Literature Cited

1. Lo K, Hahne F, Brinkman R, Gottardo R. flowClust: A bioconductor package for automated gating of flow cytometry data. BMC Bioinformatics 2009;10:145.
2. Finak G, Bashashati A, Brinkman R, Gottardo R. Merging mixture components for cell population identification in flow cytometry. Adv Bioinformatics 2009;2009:247646.
3. Aghaeepour N, Finak G, Hoos H, Mosmann TR, Brinkman R, Gottardo R, Scheuermann RH. Critical assessment of automated flow cytometry data analysis techniques. Nat Methods 2013;10:445.
4. Kaltsidis H, Cheeseman H, Kopycinski J, Ashraf A, Cox MC, Clark L, Anjarwalla I, Dally L, Bergin P, Spentzou A, Higgs C, Gotch F, Gazzard B, Gomez R, Hayes P, Kelleher P, Gill DK, Gilmour J. Measuring human T cell responses in blood and gut samples using qualified methods suitable for evaluation of HIV vaccine candidates in clinical trials. J Immunol Methods 2011;370:43–54.
5. Shulman N, Bellew M, Snelling G, Carter D, Huang Y, Li H, Self SG, McElrath MJ, De Rosa SC. Development of an automated analysis system for data from flow cytometric intracellular cytokine staining assays from clinical vaccine trials. Cytometry Part A 2008;73A:847–856.
6. Glasbey CA, Mardia KV. A review of image-warping methods. J Appl Stat 1998;25: 155–171.
7. Hahne F, Khodabakhshi AH, Bashashati A, Wong CJ, Gascoyne RD, Weng AP, Seyfert-Margolis V, Bourcier K, Asare A, Lumley T, Gentleman R, Brinkman RR. Per-channel basis normalization methods for flow cytometry data. Cytometry Part A 2010;77A:121–131.
8. Rew R, Davis G. NetCDF: An interface for scientific data access. IEEE Comput Graphics Appl 1990;10:76–82.
9. Kalams SA, Parker SD, Elizaga M, Metch B, Edupuganti S, Hural J, De Rosa S, Carter DK, Rybczyk K, Frank I, Fuchs J, Koblin B, Kim DH, Joseph P, Keefer MC, Baden LR, Eldridge J, Boyer J, Sherwat A, Cardinali M, Allen M, Pensiero M, Butler C, Khan AS, Yan J, Sardesai NY, Kublin JG, Weiner DB. Safety and comparative immunogenicity of an HIV-1 DNA vaccine in combination with plasmid interleukin 12 and impact of intramuscular electroporation for delivery. J Infect Dis 2013;208:818–829.
10. De Rosa SC, Carter DK, McElrath MJ. OMIP-014: Validated multifunctional characterization of antigen-specific human T cells by intracellular cytokine staining. Cytometry Part A 2012;81:1019–1021.
11. Horton H, Thomas EPE, Stucky JA, Frank I, Moodie Z, Huang Y, Chiu YLL, McElrath MJ, De Rosa SC. Optimization and validation of an 8-color intracellular cytokine staining (ICS) assay to quantify antigen-specific T cells induced by vaccination. J Immunol Methods 2007;323:39–54.
12. Finak G, Jiang W, Pardo J, Asare A, Gottardo R. QUAliFiER: An automated pipeline for quality assessment of gated flow cytometry data. BMC Bioinformatics 2012;13: 252.
13. Pipkin ME, Rao A, Lichtenheld MG. The transcriptional control of the perforin locus. Immunol Rev 2010;235:55–72.
14. Storey J. A direct approach to false discovery rates. J R Stat Soc Ser B Stat Methodol 2002;64:479–498.
15. Gentleman R, Carey V, Bates D, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J. Bioconductor: Open software development for computational biology and bioinformatics. Genome Biol 2004;5:R80.
16. Nelson EK, Piehler B, Eckels J, Rauch A, Bellew M, Hussey P, Ramsay S, Nathe C, Lum K, Krouse K, Stearns D, Connolly B, Skillman T, Igra M. LabKey server: An open source platform for scientific data integration, analysis and collaboration. BMC Bioinformatics 2011;12:71.
17. Trigona WL, Clair JH, Persaud N, Punt K, Bachinsky M, Sadasivan-Nair U, Dubey S, Tussey L, Fu TM, Shiver J. Intracellular staining for HIV-specific IFN-γ production: Statistical analyses establish reproducibility and criteria for distinguishing positive responses. J Interferon Cytokine Res 2003;23:369–377.
18. Maecker HT, McCoy JP, Nussenblatt R. Standardizing immunophenotyping for the Human Immunology Project. Nat Rev Immunol 2012;12:191–200.