Cytometry
PART A
Journal of the International Society for Advancement of Cytometry

MIFlowCyt

Competitive SWIFT Cluster Templates Enhance Detection of Aging Changes

Jonathan A. Rebhahn,[1] David R. Roumanes,[1] Yilin Qi,[1] Atif Khan,[2] Juilee Thakar,[2,3] Alex Rosenberg,[4] F. Eun-Hyung Lee,[5] Sally A. Quataert,[1] Gaurav Sharma,[2,6] Tim R. Mosmann[1,3]*

[1]David H. Smith Center for Vaccine Biology and Immunology, University of Rochester Medical Center, Rochester, New York

[2]Department of Biostatistics and Computational Biology, University of Rochester, Rochester, New York

[3]Department of Microbiology and Immunology, University of Rochester

[4]Department of Medicine, University of Rochester

[5]Department of Medicine, Emory University School of Medicine, Atlanta, Georgia

[6]Department of Electrical and Computer Engineering, University of Rochester

Received 3 February 2015; Revised 21 April 2015; Accepted 5 August 2015

Grant sponsor: National Institutes of Health through the Rochester Human Immunology Center, Grant number: R24AI054953; Grant sponsor: New York Influenza Center of Excellence, National Institute for Allergy and Infectious Diseases, Grant number: HHSN266200700008C

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

*Correspondence to: Dr. Tim R. Mosmann, University of Rochester Medical Center, School of Medicine and Dentistry, 601 Elmwood Ave, Box 609 Rochester, NY 14642. E-mail: Tim_mosmann@URMC.rochester.edu

Conflict of Interest: The authors declare no financial or commercial conflict of interest.

Abstract

Clustering-based algorithms for automated analysis of flow cytometry datasets have achieved more efficient and objective analysis than manual processing. Clustering organizes flow cytometry data into subpopulations with substantially homogenous characteristics but does not directly address the important problem of identifying the salient differences in subpopulations between subjects and groups. Here, we address this problem by augmenting SWIFT—a mixture model based clustering algorithm reported previously. First, we show that SWIFT clustering using a "template" mixture model, in which all subpopulations are represented, identifies small differences in cell numbers per subpopulation between samples. Second, we demonstrate that resolution of inter-sample differences is increased by "competition" wherein a joint model is formed by combining the mixture model templates obtained from different groups. In the joint model, clusters from individual groups compete for the assignment of cells, sharpening differences between samples, particularly differences representing subpopulation shifts that are masked under clustering with a single template model. The benefit of competition was demonstrated first with a semisynthetic dataset obtained by deliberately shifting a known subpopulation within an actual flow cytometry sample. Single templates correctly identified changes in the number of cells in the subpopulation, but only the competition method detected small changes in median fluorescence. In further validation studies, competition identified a larger number of significantly altered subpopulations between young and elderly subjects. This enrichment was specific, because competition between templates from consensus male and female samples did not improve the detection of age-related differences. Several changes between the young and elderly identified by SWIFT template competition were consistent with known alterations in the elderly, and additional altered subpopulations were also identified. Alternative algorithms detected far fewer significantly altered clusters. Thus SWIFT template competition is a powerful approach to sharpen comparisons between selected groups in flow cytometry datasets. © 2015 The Authors. Published Wiley Periodicals Inc.

Key terms

flow cytometry; template; immunophenotyping; competitive clustering; EM algorithm; sample comparison; automated analysis; SWIFT

Introduction

In response to the increasing complexity and dimensionality of flow cytometry data, several algorithms have recently been developed to automate the identification and/or quantification of cell populations in flow cytometry data (1–11). Some algorithms function by automating the traditional manual gating procedure for detecting and isolating specific subpopulations of interest, others attempt to completely resolve cells within a sample into clusters that ideally correspond to biologically meaningful subpopulations. Among the latter class of techniques, algorithms based on probabilistic models can describe overlapping subpopulations commonly observed in flow cytometry. These algorithms may therefore provide a better description of the overall dataset, although individual cells are not unambiguously labeled by such methods.

ISAC
International Society for Advancement of Cytometry

Published online 6 October 2015 in Wiley Online Library (wileyonlinelibrary.com)

DOI: 10.1002/cyto.a.22740

© 2015 The Authors. Cytometry Part A Published by Wiley Periodicals, Inc. on behalf of ISAC

This is an open access article under the terms of the Creative Commons Attribution NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

We have previously described a model-based algorithm, SWIFT, that provides high-resolution analysis of high-dimensional flow cytometry samples and is capable of detecting extremely small subpopulations (8,10,11). SWIFT generates a probabilistic mixture model description for the observed data in a sample, and each mixture model component is then equated to a (soft) cluster. The model can also serve as a cluster template to which additional samples may be readily assigned by computing posterior probabilities of membership in each of the clusters for each cell in the sample. To facilitate comparisons across multiple samples when some subpopulations exist only in a subset of the samples (e.g., negative control samples may lack stimulated cells), the cluster template should be obtained from a sample that contains all the subpopulations to be analyzed. With this requirement met, the template-assign procedure ensures that all samples can be compared using identical cluster descriptions with no "missing" subpopulations (11). The cluster template approach is effective for comparing the number of cells in each cluster, across many different subjects and experimental conditions. Comparisons of cell numbers in specified subpopulations is a major outcome measured in flow cytometry analysis. Because cells are assigned to each cluster on the basis of their probability of belonging to that cluster, rather than according to a rigidly defined boundary, the cluster template approach allows meaningful comparison of cell numbers in subpopulation clusters even when uncontrolled experimental variations cause minor shifts in the fluorescence intensity of markers between different samples.

However, with clearly distinct subpopulations, changes in intensity that may also provide information about differences between sample groups can be masked by the basic cluster template assignment procedure outlined above. To address this limitation, we have developed a modification of the basic cluster template method using competition between cluster templates to sharpen distinctions between cell populations in different samples. The validity of this approach has been confirmed on semisynthetic samples and a clinical study of human T cell differences between young and elderly subjects.

Methods

SWIFT Clustering and Templates

The SWIFT algorithm (10) clusters samples via a three step process with essential details as follows: In the first step, a Gaussian mixture model with specified number of components is fit to the data using the Expectation Maximization (EM) algorithm with an iteratively weighted sampling procedure that improves scalability and resolution of smaller clusters. The second step examines each of these clusters individually, and if necessary, splits individual clusters into additional Gaussian mixtures until all clusters are unimodal along individual dimensions. The third step examines and merges pairs of clusters that appear unimodal along the axis of maximal separation identified by linear discriminant analysis. The end result after this agglomerative merging step is a hierarchical mixture model in which each mixture component is itself composed of one or more Gaussian components that were obtained from the splitting phase. The merging and splitting steps are important for three reasons: for automatically adjusting the number of clusters to mitigate the dependence on the initial operator-supplied number; for identifying very small populations; and for representing skewed non-Gaussian clusters as agglomerations of smaller Gaussian components. SWIFT generates several output files, including a cluster template that provides all relevant mixture model parameters, including means and covariance matrices for the multivariate Gaussian mixture components in the primary and split clusterings, and a list of merging indices that identifies the post-split Gaussian clusters that constitute each merged cluster. This template is critically important as multiple samples in a dataset can be assigned to the template, allowing samples and groups to be compared between matched clusters. The cluster template derived from a single SWIFT clustering analysis of a sample (or consensus sample) will be referred to as a Single Clustering Run (SCR) template.

Combining Templates

The swift_template_combine program (included in the currently-distributed version of SWIFT (10,11)) combines SCR templates from multiple SWIFT analyses into one Joint template, provided the flow cytometry input parameters are identical across the templates and the per-channel data transformations are close. Specifically, the Joint template includes all final mixture model components in each of the constituent SCR templates, in the same relative proportions, e.g., if two templates are combined, all proportions will be reduced two-fold. A CSV file is created referencing the Joint template cluster indices to prior cluster indices from the parent templates. Multiple samples can then be assigned to the Joint template. As the constituent sets of clusters will compete for cell assignment, this will be referred to as cluster template competition.

Single Cluster Isolation, Transformation, and ReInsertion

A human sample of peripheral blood mononuclear cells (PBMCs) was clustered in SWIFT, sample events were stochastically assigned to the cluster template and an individual cluster (#22) selected. All events in cluster #22 were isolated and saved in a new FCS file, with modifications that were manually

specified in a CSV file as part of the input to the cluster isolation program. After compensation (matrix left division), and transformation (inverse hyperbolic sine) using the same parameters as in the original SWIFT analysis, the cluster was modified in the transformed space by shifting the position, i.e., uniformly changing the scale of the fluorescence values by specified amounts in the CD4, CD8, CD14, and CD45RA dimensions as shown in Supporting Information Figure S2. In addition to the modified cluster shown in Supporting Information Figure S2 (CD4 median 10,000), three more modified clusters were produced with CD4 medians of 20,000, 30,000, and 40,000. Finally, the events were converted back to instrument space by untransforming (hyperbolic sine), then uncompensating (matrix multiplication), then saving as a new FCS file. The resulting modified single-cluster FCS files were then concatenated (in different proportions) with a 10% random subsample of the initial sample, so that each resulting concatenated FCS file contained all normal cell subpopulations, plus a new "target" subpopulation at a specified location and size. These semisynthetic FCS files were then clustered in SWIFT as usual.

Clinical Samples

Blood was collected from healthy human subjects, ages 19–82 years, in two studies performed in 2003 (Study 1) and 2012 (Study 2). PBMCs were isolated from sodium heparinized peripheral blood by Ficoll-Hypaque^{TM} gradient centrifugation, washed and cryopreserved in 90% FBS and 10% DMSO (Sigma-Aldrich, St. Louis, MO). Cells were frozen to −80°C using an isopropanol-filled, controlled-rate freezing device. After 24 − 48 h at −80°C, the vials were transferred into liquid nitrogen. All procedures and the consent form were approved by the Research Subjects Review Board at the University of Rochester Medical Center, Rochester, New York. Ages and genders of the subjects are shown in Supporting Information Table 1.

Flow Cytometric Analysis

PBMC were thawed in RPMI 1640 (Cellgro, Manassas, VA), supplemented with penicillin (50 IU/mL)-streptomycin (50 μg/mL) (GIBCO, Carlsbad, CA), 10 μg/mL DNase (Sigma-Aldrich, St. Louis, MO) and 8% FBS (assay medium). Cells were centrifuged and re-suspended in RPMI 1640, supplemented with 8% FBS. PBMC were labeled with a 15-color T cell phenotyping panel of antibodies (Supporting Information Table 2) using a micromethod (12,13). Cell data were acquired using an LSR II cytometer (BD Immunocytometry Systems). All samples from both studies were analyzed in a single batch, so that no normalization needed to be applied to the data.

Data Processing Workflow (Summarized in Supporting Information Fig. S1)

Compensation was examined manually, and adjusted where necessary. Consensus samples were produced by concatenating equal numbers of cells from random sub-samples from all samples in each experimental group, e.g., concatenating all samples from elderly subjects to produce a consensus Old sample. The consensus samples were clustered in SWIFT, to generate a SWIFT mixture model that clustered the events into mixture model components. The resulting templates (probability distributions for membership in each cluster) were then used to assign all events (fractionally) in all individual samples in both subject groups. In addition to assigning all samples to the individual templates, the samples were also assigned to combination templates in which the clusters derived from the two constituent templates competed for event assignment. For each cluster, the numbers of events in each sample were tested for statistical significance using the distribution-free two-tailed Wilcoxon test to compute p values for each cluster, comparing e.g., young versus elderly or male versus female subjects for each cluster. The P values were adjusted using the Benjamini-Hochberg correction (14) for multiple tests, with a false discovery rate of 5%.

The clusters were also normalized using z-scores and were used to classify young or elderly subjects using a linear Support Vector Machine (SVM), a supervised learning method (15). The SVM classifies samples by finding the hyper-plane in a high dimensional space that maximizes the margin between the two classes. The SVM is robust and effective especially when the number of features (clusters) is greater than the number of samples (16).

Availability

Raw FCS data are publically available at http://flowrepository.org (17), under repository ID FR-FCM-ZZGS.

The SWIFT suite of programs, in MATLAB (Release 2014b, The MathWorks, Inc., Natick, MA), includes the tools required for consensus sample construction and template competition, and is freely available on request from: http://www.ece.rochester.edu/projects/siplab/Software/SWIFT.html.

Results

Strategy for Detecting Size Versus Fluorescence Intensity Variations in Clusters Across Samples

The SWIFT clustering algorithm produces a cluster template that describes the probability density of the observed flow cytometry measurements as a mixture of components (or clusters), each of which is composed of one or more multivariate Gaussians. Cells in the same sample, or any other sample analyzed under the same conditions, are assigned to clusters based on their probability of belonging to each cluster (10). Therefore, the semantics of populations in similar locations across multiple assigned samples are preserved. Assignment can be fractional (partial membership in multiple clusters), or stochastic (full membership in single clusters). This method effectively quantifies the number of cells per cluster across samples, e.g., distinguishing between stimulated and unstimulated samples (11), as diagrammed in Fig. 1A, in which the clusters are represented by variable shading to emphasize the probabilistic nature of the cluster template. Provided that the template is produced from a sample (or consensus sample) containing all clusters to be analyzed, the cluster template assignment procedure is able to quantify cells in rare subpopulations, down to zero, and is robust to small differences in the position of the subpopulation in different samples.

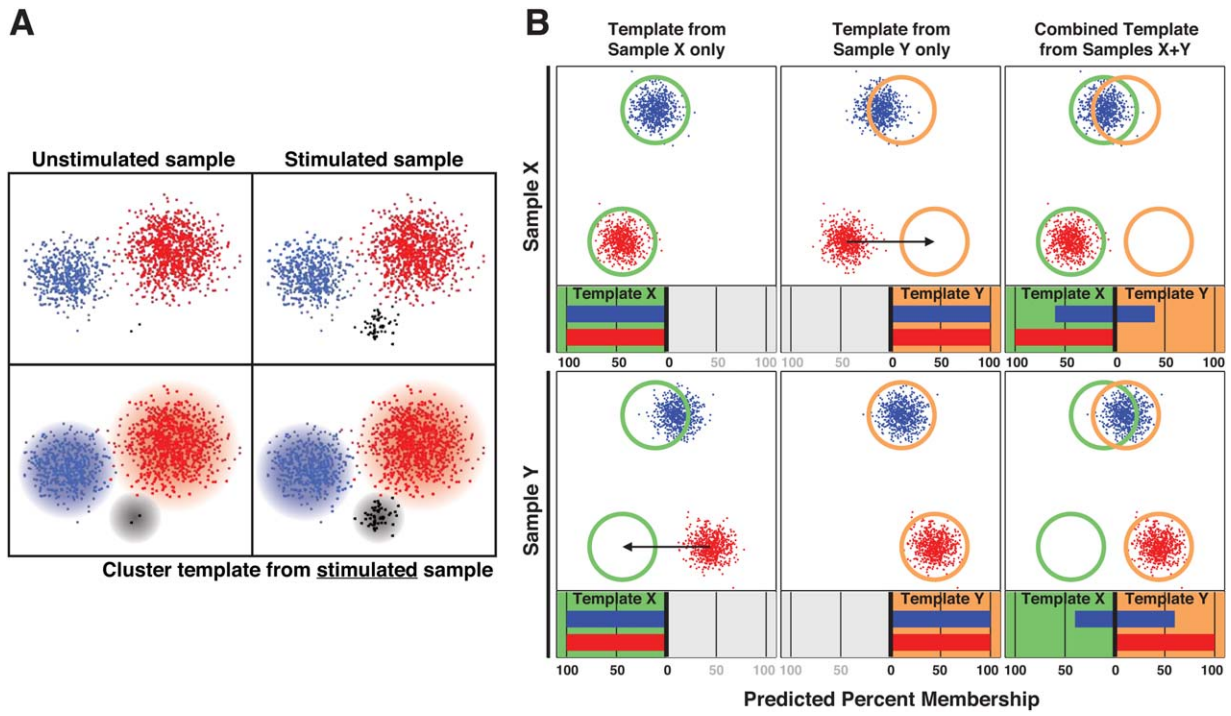

Figure 1. Strategy for competitive cluster template assignment. **A**: SWIFT identifies clusters as a mixture of unimodal probability density distributions, indicated by the shaded circles in the lower panels. Note that the small black subpopulation can only be identified in a template produced from the stimulated sample. **B**: Using a template derived from a single sample, cells are captured by the nearest cluster, even if this is not a perfect match, e.g., the red population in Samples X and Y. However, if a Joint template is produced from the X and Y Single Clustering Run (SCR) templates, each subpopulation in each sample is captured selectively by the best-fitting cluster. Bar graphs show percent allocations of cells from each subpopulation to clusters for the patterns shown in the scatterplots. SCR templates capture the two subpopulations within a single cluster without identifying the shifts in population medians, whereas the Joint template reveals the altered positions as the fraction of assignments to clusters from templates derived from X and Y.

However, if a population has a different median fluorescence intensity in one or more channels (i.e., shift in position), the population will be assigned to the nearest available cluster within the template (Fig. 1B, clusters depicted as circles for clarity). This may still be the appropriate cluster, but this process loses the information that the subpopulation has shifted. Both numbers of cells/cluster and the positions of the clusters may provide information about differences between subject groups. Therefore, we have sought to capture both types of information by combining two SCR templates into a Joint template. When a sample is assigned to the Joint template, clusters associated with each of the two templates compete for cells, so that each subpopulation will be assigned to a cluster that provides the best "fit" (Fig. 1B).

Semisynthetic Data for Evaluating Competition

Assigning multiple samples to a single SCR template was first compared to the competition method using semisynthetic datasets constructed by isolating one cluster of cells from an actual human PBMC sample, modifying the location of this cluster in four channels, and inserting the modified cluster into a normal PBMC sample from the same experiment (Supporting Information Fig. S2). This strategy was designed to keep the final sample as close as possible to real biological data, while allowing precise identification of the altered cluster. The construction of semisynthetic samples allowed clear analysis of the different requirements for resolving changes in cell numbers or positions of clusters in different samples.

Cell Number Differences Detected by SCR Templates

The semisynthetic approach was first used to confirm that SCR template assignment identified differences in the NUMBER of cells per cluster, without change in the median fluorescence values. A single SWIFT cluster (1,042 cells) was isolated from a PBMC sample, and its values in four channels (CD3, CD4, CD8, CD14) were modified (Supporting Information Fig. S2) by moving to a region (CD4+CD3− CD8+CD14+) that was relatively empty in the original dataset, allowing the behavior of SWIFT assignment to be analyzed with minimal interference from neighboring clusters. This modified target cluster was then added at different cell numbers to a normal sample, resulting in samples containing 0, 20, 40, 60, 80, and 100% of the target cluster. A SWIFT cluster template was constructed from the 100% sample, and each sample was assigned to this SCR template. Figure 2A shows that SWIFT correctly identified the number of cells in the added cluster, confirming that the SCR template/assignment strategy correctly quantifies changes in the number of cells in each population.

Moderate Shifts in Fluorescence Are Not Identified by SCR Templates

We next asked whether changes in the POSITION of the cluster could be detected by SCR templates. The same

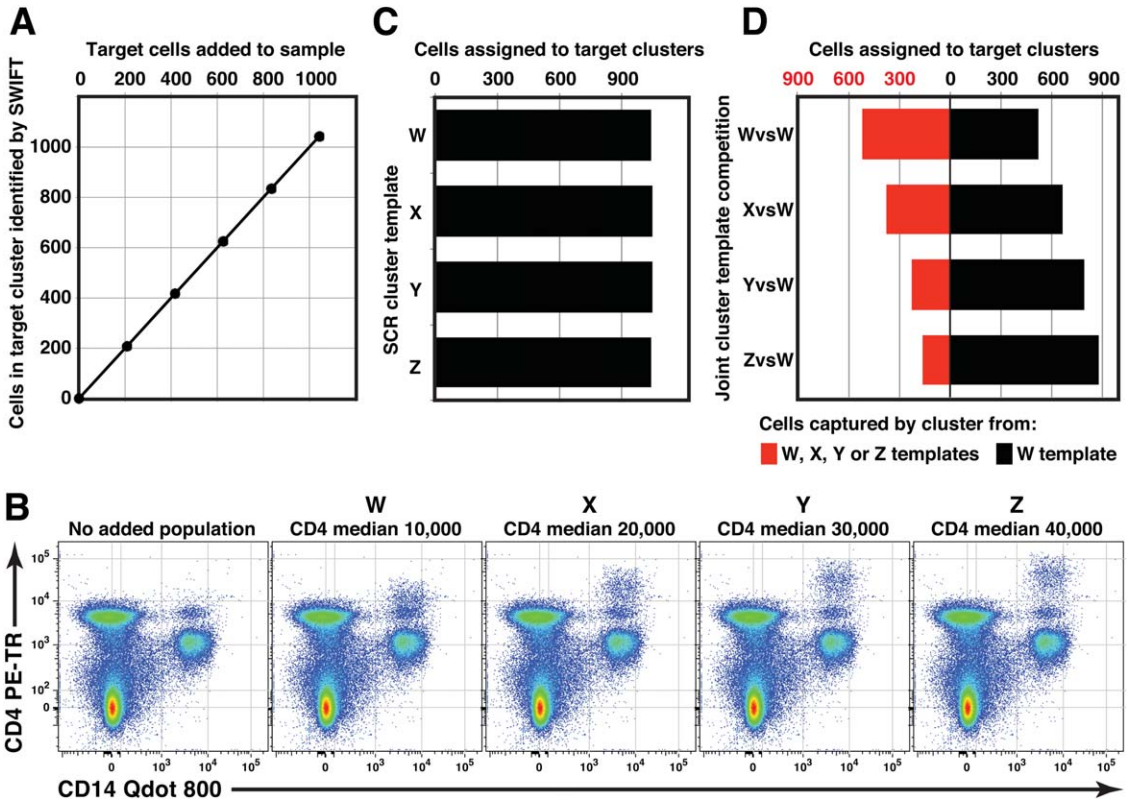

Figure 2. Performance of basic and competitive cluster assignment using semisynthetic data. A single SWIFT cluster was isolated from a human PBMC flow cytometry sample, modified by uniformly changing the scale of the CD3, CD4, CD8, and CD14 parameters, and added back to a random 10% subset of the original sample (Supporting Information Fig. S2). **A**: Graded numbers, from 0 to 1,042 cells, of the target subpopulation were added. A cluster template was produced from the sample with 1,042 target cells, and all samples were assigned to this template. The plot shows the correspondence between the number of cells assigned to the relevant cluster by swift_assign_main, versus the actual number added. **B**: The target subpopulation was modified by changing the median CD4 fluorescence intensity in four increments (W, X, Y, Z), and each was added to the random 10% subset of the original sample. **C**: SWIFT cluster templates were produced from samples W, X, Y, Z and Sample A was assigned to each template. The cluster containing the target cells in each template was identified, and the figure shows the number of cells assigned to those clusters. **D**: The templates from C were combined pairwise with the Sample W template, and sample W assigned to each Joint template. Black bars show the number of target cells captured by the relevant cluster in template W, and red bars show the target cells captured by the competing template in each Joint template pair. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

modified cluster (at 100%) was added to the full sample at different median CD4 fluorescence intensities (Fig. 2B, Samples W, X, Y, Z). SWIFT cluster templates were produced by independent clustering analyses of all four samples, and each sample was assigned to each template. During assignment to a template, events are assigned on the basis of probability of belonging to a particular cluster, therefore if no other cluster interferes, cells in a shifted population (e.g., the target cluster in sample W versus Z) will still be assigned to the same cluster unless there is a nearby cluster that could compete. Figure 2C shows that each SCR template detected the added cluster in sample A at the same numbers, i.e., the shift in cluster position was not detected.

Template Competition Detects Populations That Change Fluorescence Intensity

To detect changes in the position of equivalent clusters in different samples, we combined templates X, Y, Z pairwise with template W. When samples are assigned to a Joint template, all clusters that are described equally in the two constituent SCR templates (in this case, all clusters from the background sample) should be assigned about 50% of the cells in their relevant population. However, the added population in W should fit better (i.e., >50% capture of cells) in the relevant cluster in template W, compared to the relevant clusters in the other three templates. Figure 2D shows the expected bias in assignment, increasing progressively as the CD4 median intensity was increased, up to a 5.5-fold preference for the W template cluster to capture the target population in W, when competed with template Z. Thus the Joint template competition method identifies populations that change intensity between samples, even when the clusters overlap, and the number of cells per cluster does not change.

Aging Studies—Consensus Samples and Templates

The cluster competition method was then used to analyze the differences between PBMC samples from the young and elderly subject groups. We predicted that if the SCR templates

and the Joint templates were applied to two groups with genuine differences, then the SCR templates should reveal some group-specific alterations, and this number should be increased by the Joint template competition approach.

PBMC from 19 young (19 − 35 years old) and 20 elderly (60 − 79 years old) subjects in the Aging 1 study were analyzed with a 15-color phenotyping panel focused on T cells (Supporting Information Table 2). Consensus samples of 4 million cells were then constructed by random sampling of equal numbers of cells from the resulting data files, in three groups: Total (all subjects); Young; and Old. The construction of consensus samples was important to ensure that all subpopulations in all samples were represented in the final cluster template, and to avoid giving undue weight to any particular sample. Each consensus sample was then clustered independently in SWIFT, resulting in Total, Young and Old SCR templates with 769, 881, and 634 merged clusters, respectively. The Young and Old SCR templates were then combined using swift_template_-combine, yielding the Joint Young/Old template (1,515 clusters). All samples were then individually assigned to these four templates. A flow diagram summarizing these steps is provided in Supporting Information Figure S1.

Competition between Young and Old Templates Improves the Detection of Age-Related Differences

To evaluate whether the initial SCR templates, or the Joint template could resolve fine differences between the young and elderly subjects, each cluster was evaluated by the Wilcoxon ranksum test for significant differences between the Young and Old groups. As the SWIFT clustering algorithm has high resolution and produces large numbers of clusters, the results were analyzed by the Benjamini-Hochberg (BH) procedure (14). The Wilcoxon *P* values were adjusted by the BH procedure, with a false discovery rate of 5%.

The resulting BH-adjusted *P* values are shown in Figure 3, plotted against the ratio of the median cells/cluster in the two groups being compared. Among the 769 clusters comprising the Total SCR template (made from a consensus of all samples, Figure 3A), only 6 clusters were significantly different at $P \leq 0.05$. Substantially larger numbers of clusters were significantly different between the young and elderly subjects when samples were assigned separately to either the Young or Old SCR templates (Fig. 3B). 52 clusters were derived from the Young (27 clusters) and Old (25 clusters) templates, respectively. The competition approach with the Joint template further sharpened these differences, increasing the number of clusters showing significant young/elderly differences, to 110 clusters (71 Young and 39 Old, Fig. 3C). As expected, the majority of clusters that were more highly represented in young subjects (X-axis values >0) were derived from the Young template clusters within the Joint template, and vice versa. These results suggested (1) that competition between two cluster templates (Fig. 3C) could increase the resolution of differences between groups of samples (particularly outperforming the SCR template derived from a consensus of all samples, Fig. 3A); and (2) that substantial numbers of the clusters quantified by the SWIFT Joint template approach were different between the Young and Old groups.

Is SWIFT Cluster Competition Causing Over-Fitting?

We considered the possibility that the high resolution and large numbers of clusters identified by SWIFT might facilitate finding minor differences between the groups that were not due to the aging process. This was tested in two ways—by repeating the entire clustering and templating approach with the same subjects grouped into male and female subjects; and by analyzing subjects in a second clinical study that had not contributed to the cluster templates.

Male/female comparison. All subjects in Study 1 were regrouped into subsets of 14 male and 29 female subjects, consensus samples were prepared from the male and female subjects as described above, and the consensus samples were clustered in SWIFT. All samples were assigned to the Total SCR template, both SCR templates, and to the Joint Male/Female template. Figure 3D shows the resulting BH-corrected *P* values for Male versus Female cluster comparisons in the Total SCR template. In contrast to the young/elderly comparison, the male/female comparison showed no clusters reaching significance in the Total SCR template. When the samples were assigned to the Male and Female SCR templates, 14 clusters were significantly different between males and females (Fig. 3E), and this number was increased to 64 using the Joint Male/Female template (Fig. 3F). Importantly, the improved detection of significantly different clusters was specific for the templates used for cluster assignment and competition: The Male and Female SCR templates did not identify young/elderly differences any better than the Total SCR template (Fig. 3G), and competition using the Joint Male/Female template also did not result in any improvement in the detection of young/elderly differences (Fig. 3H). Similar results were observed in the reverse direction, e.g., competing the Joint Young/Old template did not improve the detection of male/female significant differences (data not shown).

Extension to a second study. Samples from subjects in Aging Study 2 with ages within the ranges of young and elderly in Study 1 (19 − 35 and 60 − 82 years old) were then assigned to the templates generated from Study 1, resulting in the addition of 3 young and 14 elderly subjects to the analysis. Study 2 was not evaluated separately because only three young subjects were included. If the competitive cluster template method had over-fitted the data and found mainly nonage-related differences, the *P* values of the comparisons should generally decrease in the aggregate dataset because the second dataset did not participate in the construction of the templates. However, if genuine young/elderly differences had been detected, the increase in subject numbers due to the contribution of the second study should further improve the significance and result in an increased number of clusters being judged significantly different. Figure 3I shows that the *P* values did improve substantially, and the total number of clusters

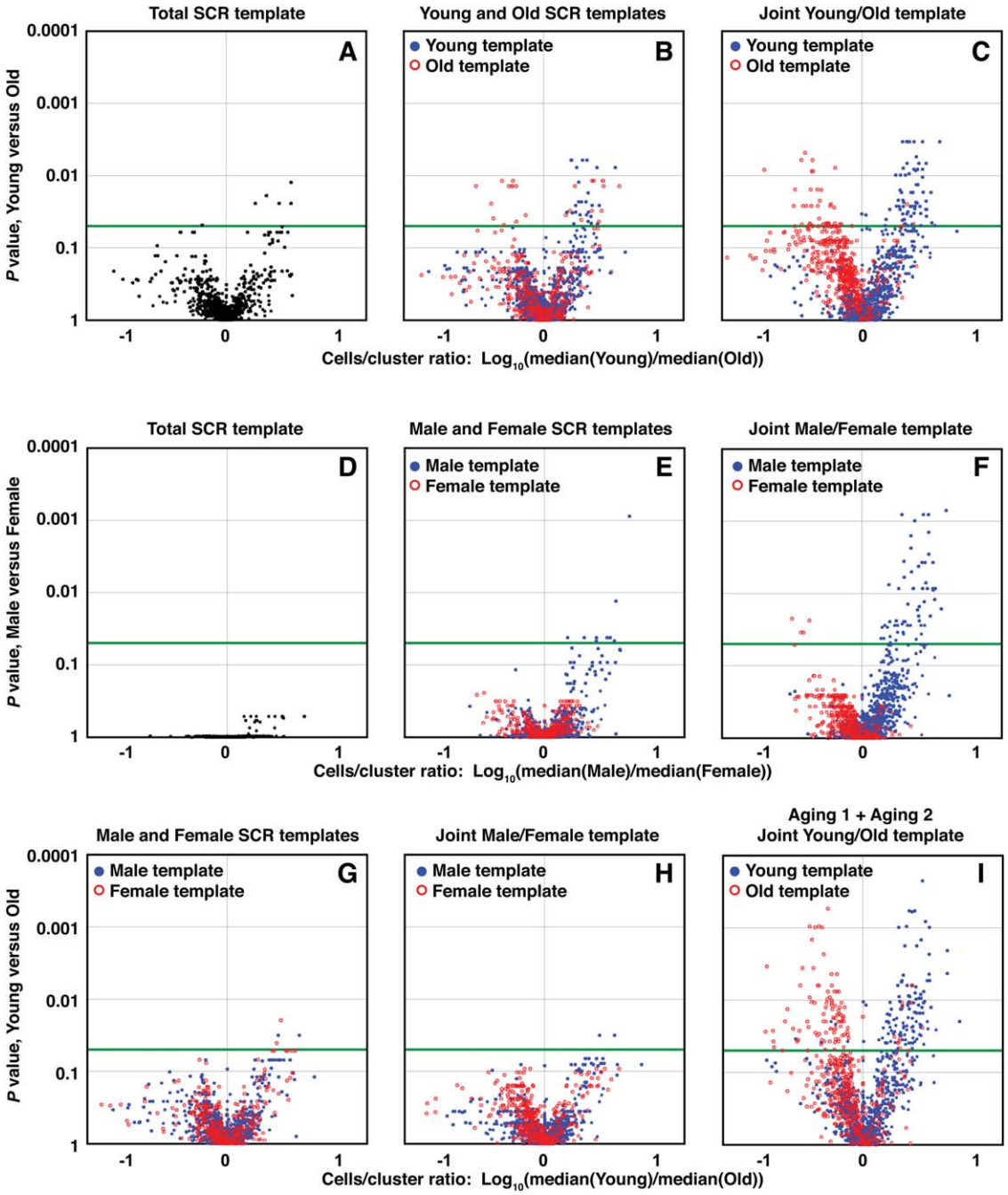

Figure 3. Competitive cluster template assignment identifies a larger number of significant clusters for discriminating young versus elderly subjects. PBMC from 19 young and 20 elderly subjects were analyzed by a 15-color phenotyping panel, and five consensus samples (Total (all subjects); Young; Old; Male; and Female) were constructed by random sampling of equal numbers of events from each subject, to obtain a total of 4 million events per consensus sample. The five consensus samples were clustered in SWIFT, and Joint templates produced by combining Young and Old SCR templates, as well as Male and Female SCR templates. All individual samples were assigned to each of the resulting templates. For each cluster, in each template, the numbers of cells assigned to the cluster were compared by a two-tailed Wilcoxon test between Young versus Old (**A, B, C, G, H, I**), or Male versus Female (**D, E, F**) subjects. The resulting *P* values were adjusted by the Benjamini-Hochberg correction, at a false discovery rate of 5%. The adjusted *P* values are plotted against the log of the ratio between the median numbers of cells assigned to that cluster in the two subject groups being compared. Significance is assumed at $P < 0.05$ (green line). Templates used for assignment were Total SCR (A, D), Young and Old SCR (B), Male and Female SCR (E, G), Joint Young/Old (C, I), Joint Male/Female (F, H). All results show the Aging 1 study only, except panel I which includes both Aging 1 and Aging 2 studies. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Table 1. Machine learning classification according to SWIFT clusters

TEMPLATE	STUDY	CLASSIFICATION ACCURACY (%)
Young and Old SCR	10-fold cross-validation on #1	89.7
	10-fold cross-validation on #1 & #2	91.1
	Trained on #1 Independently tested on #2	82.4
Joint Young/Old	10-fold cross-validation on #1	100
	10-fold cross-validation on #1 & #2	96.4
	Trained on #1 Independently tested on #2	82.4

A support vector machine was used to classify young and elderly subjects using numbers assigned to the clusters obtained by SCR or Joint templates.

deemed significant increased further, to 226. Thus the addition of the second study further suggests that genuine age-related differences were being detected by the cluster template competition approach.

Competitive SWIFT template analysis of a pregnancy dataset. The SWIFT competitive template approach was extended to an independent dataset comprising pregnant and nonpregnant women. Templates were produced from all pregnant and all nonpregnant subjects, and all samples assigned to the two SCR templates, and the joint template. Supporting Information Figure S3 shows that SWIFT identified a number of clusters that were significantly altered in pregnant subjects, and the detection of these differences was enhanced by template competition.

Machine learning classifier. Additionally, we used a SVM to determine if the template competition approach improves classification between young and elderly subjects. Three strategies were used for classification using SCR and Joint templates, after normalizing results for each cluster by *z* scores across all samples: 1) ten-fold cross-validation was performed on clustering results from Study 1; 2) ten-fold cross-validation was performed on the aggregate results from Studies 1 and 2; and 3) Study 2 results were used as an independent test set after training on the Study 1 samples. All three strategies showed a classification accuracy of >80% (Table 1), confirming that the differences detected by SWIFT represented real aging-related changes. In the first two strategies, template competition further increased the accuracy of the classification, and in the independent analysis of Study 2, the accuracy was equal with or without competition.

What Are the Clusters Identified by the High-Resolution Competitive Method?

To determine whether the clusters with significant young/elderly differences detected by the competition strategy were biologically reasonable, we examined the 24 clusters with the most significant *P* values derived from each of the Young and Old constituent templates within the Joint template. Figure 4 shows the properties of each cluster in each subject, for 24 clusters (12 each from the Young and Old templates) selected from these 48. The selected Young template clusters were all represented at higher levels in the Young samples (Fig. 4A), whereas the clusters from the Old template were more populous in the elderly subjects. Figure 4B shows that, within each cluster, the median fluorescence intensity values were normally consistent across all subjects, suggesting that the same subpopulations were being identified in each subject. There was considerable heterogeneity between clusters, particularly when all parameters were considered (Fig. 4B and Supporting Information Fig. S4). Four main patterns of marker expression were observed in these selected 24 clusters: from the Young template, eight clusters had the phenotype of naïve CD8 T cells (CD3+CD4−CD8+CD45+CCR7+) and four were CD3+CD4−CD8− (probably gamma-delta T cells). From the Old template, five clusters appeared to be memory CD4 T cells (CD3+CD4+CD8−CD45−), three expressed monocytic markers (CD3−CD14+CD11B+), and four expressed both T cell and monocyte markers and showed high FSC-W values, consistent with T cell/monocyte conjugates. The T cell results fit well with previous studies that showed a shift in frequency from naïve to memory CD8 T cells on aging (18–21), consistent with the accumulation of memory cells with continued antigen stimulation, and oligoclonal expansion of specific memory CD8 T cell populations (22,23). Similarly, a decrease in gamma-delta T cells has been reported previously (24,25). Although CD27 expression was variable in the clusters detected by SWIFT, clusters 913, 1,328, and 908 were memory CD4 T cells that expressed low levels of CD27 and were over-represented in the elderly, consistent with previous studies (26–28). Thus the competitive template assignment in SWIFT identified many potentially different clusters between young and elderly. Some of these were consistent with previous reports, and further subpopulations were identified for future analysis.

In the pregnancy dataset, several CD4 T cell clusters differed significantly between pregnant and nonpregnant groups (Figs. S3 and S5). These were mostly naive (CD45RA high), with one large cluster (#12) of CD4 memory cells. The significantly different CD8 T cell clusters were also mostly naive, except for the abundant cluster #14, and were also mostly more abundant in the samples from pregnant subjects. Finally, there were four CD14+ clusters (#407 to #413) that probably represent monocytes, and these were all decreased in the pregnant subjects. Cluster #413 is probably an aggregate of monocytes and CD8 T cells. Previous studies have shown alterations in NK cell populations during pregnancy, particularly in association with different pregnancy outcomes (29,30). The effects of pregnancy on circulating T cell populations are less known, although we and others have shown a subtle alteration of T cell cytokine patterns during pregnancy ((31–33), and our data not shown). Alterations in specific monocytes

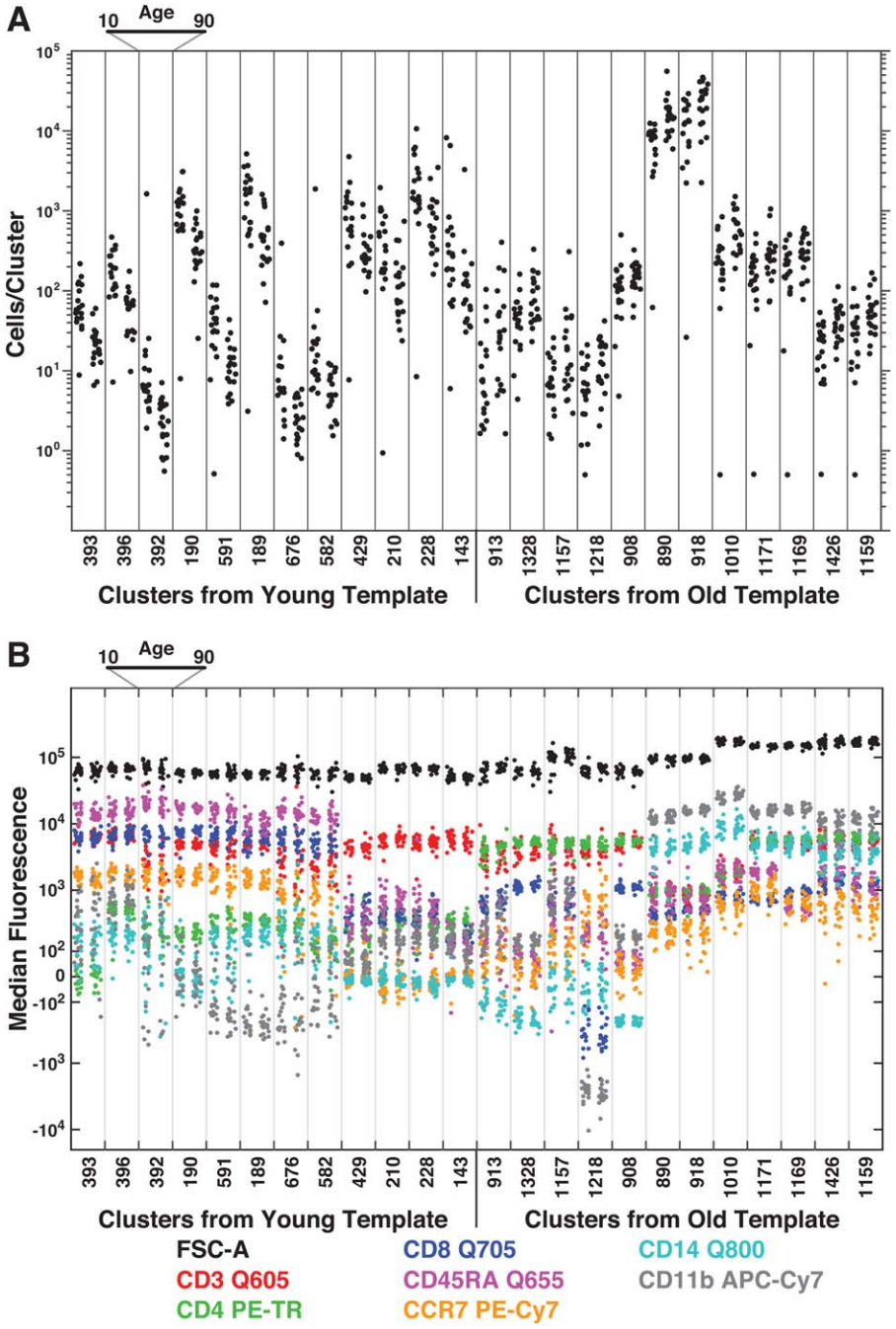

Figure 4. Properties of clusters that exhibit statistically significant differences between young and elderly subjects. Twelve clusters each were chosen from the Young and Old templates, with significant differences between young and elderly subjects in the number of cells per cluster, as evaluated by the Joint Young/Old template (**A**). For each cluster, the inverse hyperbolic sine of the median fluorescence intensity of the cluster in each subject is plotted (with labels referring to actual fluorescence values) for eight markers (**B**) (additional markers are shown in Supporting Information Fig. S4). The coarse scale on the X-axis represents different clusters, and the fine scale shows the age of each subject.

sub-populations have been described (34). Because the dataset used for the analysis presented here was from a T cell-focused phenotyping panel using resting PBMC, differences in NK cell and activated T cell patterns would not be expected to be well-resolved, whereas subtle differences in T cell populations could be identified.

Comparison of SWIFT Cluster Template Competition with Other Publicly Available Programs

We compared the SWIFT competitive template approach with alternative algorithms that could compare cluster data between groups of samples, using the available implementations on the GenePattern (35) server or in R:Bioconductor

(36). Using the Aging 1 dataset, competitive SWIFT analysis was benchmarked against seven clustering and comparison algorithms. Three of these programs [SamSpectralCluster, (37), FlowMergeCluster, and FlowClustClassify (38)], in their GenePattern versions, were not able to cluster the large samples (about 1-4 million cells, 19 parameters) used in these analyses. ImmPortFLOCK (5) (plus ImmPortCrossSample) could be used in a similar manner as SWIFT, by combining templates from Old and Young subjects. Using the same post-processing analysis as for SWIFT, FLOCK identified one cluster that was significantly different between old and young groups (after the Benjamini-Hochberg correction) (Supporting Information Fig. S6). KMeansClassifyFCS (39) + MClustClusterLabel also identified one significant cluster. Both FLOCK and KMeans clusters were naïve CD8+ T cells, consistent with the identification of naive CD8 T cells as some of the most significant clusters by SWIFT. FlowMeansCluster (40)+ MClustClusterLabel was able to cluster only 22 of the 39 Aging 1 samples, and possibly because of the reduced subject numbers, no clusters were significantly different between old and young subjects. We also tested our Aging 1 dataset on the recently published immunoClust algorithm (41) available in R:Bioconductor. Because of its relatively high computational resource requirements, immunoClust was not able to complete the analysis of all 39 samples (running on the University of Rochester BlueHive computing cluster, https://info.circ.rochester.edu/BlueHive/System_Overview.html). We selected subsets of eight Young and 8 Old subjects, and analyzed this smaller data set using both SWIFT and immunoClust. SWIFT identified 22 significant clusters ($P < 0.05$ after the Benjamini-Hochberg correction) whereas immunoClust did not identify any cluster with a BH-corrected P values <1 (Supporting Information Fig. S7). Thus alternative methods for finding differences between experimental groups gave compatible results for some methods, but the increased resolution of SWIFT and the competition approach resulted in many more significantly different clusters being identified.

Discussion

The SWIFT template competition method takes advantage of two features of the SWIFT clustering output—the high resolution of complex samples into relatively large numbers of clusters, and the cluster templates that allow assignment of many samples to the same template. Changes in the numbers of cells per cluster are detected effectively by the original SWIFT method of assigning samples to a SCR template, but this process does not readily detect small changes in the position of clusters. In fact, this property is an advantage in many applications addressing target subpopulations, as the template/assign procedure will compensate for small shifts in the positions of subpopulations, allowing the program to adapt to the small differences normally seen between subjects (11). For example, if the goal of an experiment is to enumerate the antigen-specific T cell response to antigen stimulation, the original SCR template/assign method yields excellent results. The template competition method was developed in response to a concern that in exploratory analysis, searching for subpopulations that are altered between groups defined by age, gender, disease, vaccination, etc., there might be alterations in both numbers of cells/cluster, as well as the fluorescence intensity of the cluster in one or more channels.

The competition method effectively reveals small shifts or larger shifts in the absence of competing subpopulations. Competition can reveal shifts of only twofold in median fluorescence intensity in one channel (Fig. 2D), using a relatively broadly distributed cluster as the test case. If these samples were analyzed by a single consensus template, the two populations would be merged, and therefore the difference between the samples would be obscured. As the competition method depends on the overlap between the probability distributions of the clusters, subpopulations with lower CVs would be resolved even more. However, if a subpopulation shifts so far that it overlaps with another subpopulation, cells may still be misassigned. This difficulty is shared with other analysis methods, and can only be resolved by analysis with additional markers. However, competition may improve the precision of identification in cases where a subpopulation has moved partially toward a different subpopulation—by providing a better fit, the combined template may pull cells away from the neighboring, inappropriate cluster.

The cluster template competition strategy also works well for multiple SCR templates, for example four-way competitions (data not shown). As the number of cells per cluster is diluted by multiple templates, evaluation of subpopulation memberships by fractional assignment of cells is more appropriate than stochastic assignment of whole cells—both assignment choices are available in the SWIFT config file.

Comparison of SWIFT with alternative methods showed that at least two other methods could identify a sub-population of naïve CD8 T cells that differed significantly between young and elderly subjects, but that the alternative methods did not identify several other populations identified by SWIFT, including at least two cell types that were also known from previous studies to be altered in elderly subjects. The higher resolution of SWIFT—hundreds of clusters versus e.g., 30 for alternative methods—may have helped to resolve more subpopulations, and this resolution was further increased by the competition strategy. Both in our previous studies (10,11) and in this study, at least some of the small populations identified by SWIFT were biologically significant. Thus at least SOME of the extra resolution afforded by SWIFT is valuable for identifying true biological diversity. Without exhaustive analysis of the hundreds of SWIFT sub-populations it is difficult to know whether ALL of the subpopulations are biologically meaningful, although it is increasingly clear that the full diversity of T cells, for example, is much greater than previously thought (recently reviewed in Ref. 42).

Although the extensive model-fitting, splitting and merging steps in SWIFT resolve clusters that correspond as closely as possible to biological subpopulations of cells, the subsequent template competition method introduces some redundancy into the clusters. If a cell subpopulation is identical in the two samples used to derive two SCR templates, then the subpopulation will be described by two clusters in the Joint template (one from each constituent SCR template), each capturing about half of the cells in that subpopulation. Duplicate clusters could

be removed by performing an additional SWIFT merging step after combining the cluster templates, but this might reduce resolution by merging clusters with small shifts, e.g. samples W and X in Figure 2B. As a major purpose of the cluster template competition method is to identify any populations that are altered between groups of samples, the duplication of unaltered clusters does not interfere with the interpretation of the results. The competitive SWIFT template method is ideally suited to identifying leads for further biological analysis, and providing diagnostic signatures for classification purposes. Thus the normal outcome of such projects is to identify cell subpopulations that will be targeted in follow-up experiments.

The clinical studies analyzed here were primarily targeted at identifying differences between young and elderly subjects, but as the subjects included both genders, male/female differences could be investigated in the same dataset. This provided a good reciprocal demonstration of the effectiveness of the cluster template competition procedure, as the Joint Young/Old template enhanced detection of young/elderly but not male/female differences, whereas the Joint Male/Female template did the opposite. Interestingly, more young/elderly differences were revealed by either the SCR or Joint templates, suggesting that aging has a larger effect than gender on PBMC subpopulations. Some of the altered subpopulations identified by SWIFT were consistent with known changes in the elderly, e.g. the shift from naïve to memory CD8 and T cells. However, all naïve CD8 T cell subpopulations were not affected equally, suggesting that the extra resolution of SWIFT, compared to manual analysis, may reveal more detailed and specific differences. Additional subpopulations identified by SWIFT, e.g. the monocyte clusters, may provide fruitful leads for further investigation.

The cluster competition method thus provides an important extension to the cluster template method developed in SWIFT. Competition improves the detection of subpopulations that have altered the expression of different markers, and is particularly suited to the agnostic detection of subpopulations that have changed between experimental groups. The ability to interrogate different aspects of the population, by judiciously choosing the templates to construct and compete, makes this a versatile tool for enhancing the detection of altered subpopulations.

Literature Cited

1. Bashashati A, Brinkman RR. A survey of flow cytometry data analysis methods. Adv Bioinform 2009;2009:19.
2. Lo K, Hahne F, Brinkman RR, Gottardo R. flowClust: A bioconductor package for automated gating of flow cytometry data. BMC Bioinform 2009;10:145
3. Pyne S, Hu X, Wang K, Rossin E, Lin TI, Maier LM, Baecher-Allan C, McLachlan GJ, Tamayo P, Hafler DA. and, others. Automated high-dimensional flow cytometric data analysis. Proc Natl Acad Sci USA 2009;106:8519–8524.
4. Finak G, Bashashati A, Brinkman R, Gottardo R. Merging mixture components for cell population identification in flow cytometry. Adv Bioinform 2009;247646
5. Qian Y, Wei C, Eun-Hyung Lee F, Campbell J, Halliley J, Lee JA, Cai J, Kong YM, Sadat E, Thomson E. et al. Elucidation of seventeen human peripheral blood B-cell subsets and quantification of the tetanus response using a density-based method for the automated identification of cell populations in multidimensional flow cytometry data. Cytometry B Clin Cytom 2010;78(Suppl 1):S69–S82.
6. Qiu P, Simonds EF, Bendall SC, Gibbs KD Jr, Bruggner RV, Linderman MD, Sachs K, Nolan GP, Plevritis SK. Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. Nat Biotechnol 2011;29:886–891.
7. Aghaeepour N, Jalali A, O'Neill K, Chattopadhyay PK, Roederer M, Hoos HH, Brinkman RR. RchyOptimyx: Cellular hierarchy optimization for flow cytometry. Cytometry A 2012;81A:1022–1030.
8. Aghaeepour N, Finak G, The FlowCAP Consortium, the DREAM Consortium, Hoos H, Mosmann TR, Brinkman R, Gottardo R, Scheuermann RH. Critical assessment of automated flow cytometry data analysis techniques. Nat Methods 2013;10: 228–238.
9. Cron A, Gouttefangeas C, Frelinger J, Lin L, Singh SK, Britten CM, Welters MJ, van der Burg SH, West M, Chan C. Hierarchical modeling for rare event detection and cell subset alignment across flow cytometry samples. PLoS Comput Biol 2013;9: e1003130
10. Naim I, Datta S, Rebhahn J, Cavenaugh JS, Mosmann TR, Sharma G. SWIFT-scalable clustering for automated identification of rare cell populations in large, high-dimensional flow cytometry datasets, Part 1: Algorithm design. Cytometry A 2014; 85A:408–421.
11. Mosmann TR, Naim I, Rebhahn J, Datta S, Cavenaugh JS, Weaver JM, Sharma G. SWIFT-scalable clustering for automated identification of rare cell populations in large, high-dimensional flow cytometry datasets, Part 2: Biological evaluation. Cytometry A 2014;85A:422–433.
12. Wang JC, Kobie JJ, Zhang L, Cochran M, Mosmann TR, Ritchlin CT, Quataert SA. An 11-color flow cytometric assay for identifying, phenotyping, and assessing endocytic ability of peripheral blood dendritic cell subsets in a single platform. J Immunol Methods 2009;341:106–116.
13. Weaver JM, Yang H, Roumanes D, Lee FE, Wu H, Treanor JJ, Mosmann TR. Increase in IFNgamma(-)IL-2(+) cells in recent human CD4 T cell responses to 2009 pandemic H1N1 influenza. PLoS One 2013;8:e57275
14. Benjamini Y, Hochberg Y. Controlling the false discovery rate—A practical and powerful approach to multiple testing. J Roy Stat Soc Series B-Methodological 1995; 57:289–300.
15. Wu X, Kumar V, Quinlan JR, Ghosh J, Yang Q, Motoda H, McLachlan GJ, Ng A, Liu B, Yu PS, et al. Top 10 algorithms in data mining. Knowledge Inform Syst 2008;14:1–37.
16. Meyer DJ, Leisch F, Hornik K. The support vector machine under test. Neurocomputing 2003;55:169–186.
17. Spidlen J, Breuer K, Rosenberg C, Kotecha N, Brinkman RR. FlowRepository: A resource of annotated flow cytometry datasets associated with peer-reviewed publications. Cytometry A 2012;81A:727–731.
18. Koch S, Larbi A, Derhovanessian E, Ozcelik D, Naumova E, Pawelec G. Multiparameter flow cytometric analysis of CD4 and CD8 T cell subsets in young and old people. Immun Ageing 2008;5:6
19. Rymkiewicz PD, Heng YX, Vasudev A, Larbi A. The immune system in the aging human. Immunol Res 2012;53:235–250.
20. Wertheimer AM, Bennett MS, Park B, Uhrlaub JL, Martinez C, Pulko V, Currier NL, Nikolich-Zugich D, Kaye J, Nikolich-Zugich J. Aging and cytomegalovirus infection differentially and jointly affect distinct circulating T cell subsets in humans. J Immunol 2014;192:2143–2155.
21. Nikolich-Zugich J. Aging of the T cell compartment in mice and humans: From no naive expectations to foggy memories. J Immunol 2014;193:2622–2629.
22. Schwab R, Szabo P, Manavalan JS, Weksler ME, Posnett DN, Pannetier C, Kourilsky P, Even J. Expanded CD4+ and CD8+ T cell clones in elderly humans. J Immunol 1997;158:4493–4499.
23. Posnett DN, Sinha R, Kabak S, Russo C. Clonal populations of T cells in normal elderly humans: The T cell equivalent to "benign monoclonal gammapathy". J Exp Med 1994;179:609–618.
24. Roux A, Mourin G, Larsen M, Fastenackels S, Urrutia A, Gorochov G, Autran B, Donner C, Sidi D, Sibony-Prat J, et al. Differential impact of age and cytomegalovirus infection on the gammadelta T cell compartment. J Immunol 2013;191: 1300–1306.
25. Vasudev A, Ying CT, Ayyadhury S, Puan KJ, Andiappan AK, Nyunt MS, Shadan NB, Mustafa S, Low I, Rotzschke O, et al. Gamma/delta T cell subsets in human aging using the classical alpha/beta T cell model. J Leukoc Biol 2014;96:647–655.
26. Nijhuis EW, Remarque EJ, Hinloopen B, Van der Pouw-Kraan T, Van Lier RA, Ligthart GJ, Nagelkerken L. Age-related increase in the fraction of CD27-CD4+ T cells and IL-4 production as a feature of CD4+ T cell differentiation in vivo. Clin Exp Immunol 1994;96:528–534.
27. Saurwein-Teissl M, Lung TL, Marx F, Gschösser C, Asch E, Blasko I, Parson W, Böck G, Schönitzer D, Trannoy E, et al. Lack of antibody production following immunization in old age: Association with CD8(+)CD28(-) T cell clonal expansions and an imbalance in the production of Th1 and Th2 cytokines. J Immunol 2002;168:5893–5899.
28. Kovaiou RD, Weiskirchner I, Keller M, Pfister G, Cioca DP, Grubeck-Loebenstein B. Age-related differences in phenotype and function of CD4+ T cells are due to a phenotypic shift from naive to memory effector CD4+ T cells. Int Immunol 2005;17: 1359–1366.
29. Lee SK, Na BJ, Kim JY, Hur SE, Lee M, Gilman-Sachs A, Kwak-Kim J. Determination of clinical cellular immune markers in women with recurrent pregnancy loss. Am J Reprod Immunol 2013;70:398–411.
30. Fu B, Li X, Sun R, Tong X, Ling B, Tian Z, Wei H. Natural killer cells promote immune tolerance by regulating inflammatory TH17 cells at the human maternal-fetal interface. Proc Natl Acad Sci USA 2013;110:E231–E240.
31. Wegmann TG, Lin H, Guilbert L, Mosmann TR. Bidirectional cytokine interactions in the maternal-fetal relationship: Is successful pregnancy a TH2 phenomenon? Immunol Today 1993;14:353–356.
32. Krishnan L, Guilbert LJ, Russell AS, Wegmann TG, Mosmann TR, Belosevic M. Pregnancy impairs resistance of C57BL/6 mice to Leishmania major infection and causes decreased antigen-specific IFN-gamma response and increased production of T helper 2 cytokines. J Immunol 1996;156:644–652.
33. Lissauer D, Goodyear O, Khanum R, Moss PA, Kilby MD. The profile of maternal CD4 T cell effector function during normal pregnancy and in women with a history of recurrent miscarriage. Clin Sci (Lond) 2013;126:347–354.

34. Melgert BN, Spaans F, Borghuis T, Klok PA, Groen B, Bolt A, de Vos P, van Pampus MG, Wong TY, van Goor H, et al. Pregnancy and preeclampsia affect monocyte subsets in humans and rats. PLoS One 2012;7:e45229
35. Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP. GenePattern 2.0. Nat Genet 2006;38:500–5001.
36. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo HC, Davis S, Gatto L, Girke T, et al. Orchestrating high-throughput genomic analysis with Bioconductor. Nat Methods 2015;12:115–121.
37. Zare H, Shooshtari P, Gupta A, Brinkman R. Data reduction for spectral clustering to analyze high throughput flow cytometry data. BMC Bioinform 2010;11:403
38. Lo K, Brinkman RR, Gottardo R. Automated gating of flow cytometry data via robust model-based clustering. Cytometry A 2008;73A:321–332.
39. Lloyd SP. Least-squares quantization in Pcm. IEEE Trans Inform Theory 1982;28: 129–137.
40. Aghaeepour N, Nikolic R, Hoos HH, Brinkman RR. Rapid cell population identification in flow cytometry data. Cytometry A 2011;79A:6–13.
41. Sorensen T, Baumgart S, Durek P, Grutzkau A, Haupl T. immunoClust-An automated analysis pipeline for the identification of immunophenotypic signatures in high-dimensional cytometric datasets. Cytometry A 2015;87A: 603–615.
42. Rebhahn JA, Deng N, Sharma G, Livingstone AM, Huang S, Mosmann TR. An animated landscape representation of CD4 T-cell differentiation, variability and plasticity: Insights into the behavior of populations versus cells. Eur J Immunol 2014;44: 2216–2229.