



Published in final edited form as:

Cytometry A. 2008 May ; 73(5): 411–420. doi:10.1002/cyto.a.20556.

Nine-Color Flow Cytometry for Accurate Measurement of T Cell Subsets and Cytokine Responses. Part II: Panel Performance Across Different Instrument Platforms

Bridget E. McLaughlin¹, Nicole Baumgarth², Martin Bigos³, Mario Roederer⁴, Stephen C. De Rosa⁵, John D. Altman⁶, Douglas F. Nixon⁷, Janet Ottinger⁸, Judy Li⁹, Laurel Beckett⁹, Barbara L. Shacklett¹⁰, Thomas G. Evans¹¹, David M. Asmuth^{1,*}

¹University of California Davis, Internal Medicine, Infectious Diseases, Davis, California

²University of California Davis, Center for Comparative Medicine, Davis, California

³J. David Gladstone Institute of Virology and Immunology, San Francisco, California

⁴National Institutes of Health (NIH), NIAID, Bethesda, Maryland

⁵University of Washington, Fred Hutchinson Cancer Research Center, Seattle, Washington

⁶Emory University, Emory Vaccine Center at Yerkes, Atlanta, Georgia

⁷University of California San Francisco, Division of Experimental Medicine, San Francisco, California

⁸Duke University Medical Center, Duke Center for AIDS Research, Durham, North Carolina

⁹University of California Davis, Public Health Sciences, Division of Biostatistics, Davis, California

¹⁰University of California Davis, Medical Microbiology and Immunology, Davis, California

¹¹Novartis Institute of Biological Research, Cambridge, Massachusetts

Abstract

Cellular immune responses elicited by vaccination are complex and require polychromatic analysis to accurately characterize the phenotype and function of rare, responding cells. Technical challenges and a lack of instrument standardization between research sites have limited the application of polychromatic cytometry in multicenter clinical trials. Two previously developed six-color T cell subset immunophenotyping reagent panels deliberately designed to accommodate three additional low frequency functional measurements were compared for their reproducibility of staining across three different flow cytometers. We repeatedly measured similar T cell subset frequencies between the two reagent panels and across the three different cytometers. Spectral overlap reduced sensitivity in two of the three open measurement channels (PE [IL-2] and APC [IFN γ]) for one reagent combination, particularly in subsets with low cytokine expression. There was no significant interassay variation for measurements across instrument platforms. Careful panel design will identify reagent combinations that minimize spectral spillover into channels

*Correspondence to: David M. Asmuth, University of California-Davis Medical Center, Division of Infectious Diseases, 4150 V Street, PSSB G500, Sacramento, CA 95817-1460, USA.

reserved for cytokine measurement and comparable results can be achieved using different cytometers, however, it is important to establish standardized quality control procedures for each instrument to minimize variation between cytometers.

Keywords

polychromatic flow cytometry; multicolor flow cytometry; high-resolution immunophenotyping; fluorochrome conjugated antibody; compensation

POLYCHROMATIC flow cytometry (PFC) allows multiple, coordinately expressed antigens to be quantitatively measured on individual lymphocytes. It is a powerful tool for accurately detecting and characterizing antigen specific lymphocytes. Recent studies using PFC have shown that antigen-specific T cell responses to vaccination are complex (1) and that rare, polyfunctional T cell subsets can be identified that potentially confer enhanced control and protection from chronic viral infections such as SIV and human immunodeficiency virus infection (2–4). Clinical trials for vaccines designed to elicit cellular immunity would benefit greatly from broader measures of antigen-specific responses. As such, there is need to maximize the number of measurements for each sample to better characterize the phenotype and function of individual cellular responses. Although advances in commercial flow cytometer design now allow for the simultaneous measurement of nine or more fluorochromes, several challenges have delayed the widespread use of PFC in multicenter trials. These challenges result chiefly from an increased number of spectral overlaps and measurement errors generated by the presence of multiple fluorochromes on each cell (5–11). Today, there exists no standard platform of lasers, filters, and optical design upon which to correlate the results of polychromatic data acquired at multiple research centers using different cytometers.

We have described the development of two 6-color reagent panels that afford robust T-cell subset determination, while minimizing background expansion into three channels reserved for low intensity and low frequency functional measurements (12). We now expand these two empirically derived reagent panels to include three cytokine stains (TNF α : FITC, IL-2: PE, and IFN γ : APC), and compare their performance using three different cytometers to determine if similar and reproducible results can be achieved with either panel across multiple instruments. Although the cytometers were similarly equipped to measure nine or more separate fluorochromes, each instrument differed slightly in its optical and flow geometry, laser power, and signal processing electronics. Our results suggest that similar and reliable results can be expected across instrument platforms, although careful attention must be given to reagent panel design and instrument calibration to achieve comparable results.

MATERIALS AND METHODS

Cell Isolation and Stimulation

Heparinized whole blood was collected from three healthy volunteers at a single time point to reduce intradonor variability in replicate comparison experiments carried out across instrument platforms. Peripheral blood mononuclear cells (PBMCs) from each donor were

separated by Ficoll centrifugation (Histopaque-1077, Sigma Aldrich, St. Louis, MO) and immediately cryopreserved in 90% FBS (Omega Scientific, Tarzana, CA) with 10% DMSO (Sigma Aldrich) and stored in liquid nitrogen (-196°C) until use within 2-months time. For each experiment, thawed PBMCs from each donor were treated with 50 units ($5\ \mu\text{l}$ of $10\ \text{U}/\mu\text{l}$ of DNase 1 (RNase-free, Roche Applied Science, Indianapolis, IN) for 5 min in a 37°C water bath, washed and resuspended in 20% RPMI + supplements (Sodium pyruvate [1 mM], HEPES [10 mM], NEAA [1X], L-Glutamine [2 mM], Pen-Strep [100 U/mL–100 $\mu\text{g}/\text{mL}$], all Gibco), and allowed to rest in culture overnight at 37°C . Approximately 2×10^6 rested cells were stimulated at 37°C for 6 h in the presence of brefeldin A ($10\ \mu\text{g}/\text{mL}$, Sigma) with or without Staphylococcus Enterotoxin-B (SEB, $5\ \mu\text{g}/\text{mL}$, Sigma-Aldrich) (13). Stimulated cells were stored at 4°C for convenience and stained the following day.

Reagent Panels

In a series of tests, 30 permutations of six-color T-lymphocyte surface stains were compared to identify reagent combinations that afforded robust determination of memory and effector T-cell subsets (12). These “anchor panels” were comprised of dimmer dyes so that brighter dyes such as FITC, PE, and APC were reserved for the measurement of cytokines or other low intensity markers. These initial comparisons were carried out using PBMCs stimulated in the presence of brefeldin A, but did not include FITC, PE, or APC cytokine stains. We identified two anchor panels, listed in Table 1, that maximized T-cell subset measurements while minimizing background spread of the compensated data into the three cytokine channels. Three cytokine stains (TNF α : FITC, IL-2: PE, and IFN γ : APC) were added to these panels and compared across cytometer platforms.

Cell Staining

Separate aliquots of stimulated and unstimulated cells from each donor were washed in cold buffer ($1\times$ PBS with 1% bovine serum albumin [BSA]), pelleted, and stained with reagent Panels I and II. Surface antigens, including CD4, CD8, CD14, CD19, CD45RA, and CCR7, were stained in the presence of $0.5\ \mu\text{g}/\text{mL}$ ethidium monoazide to identify dead cells (EMA, Molecular Probes, Eugene, OR) for 15 min at room temperature in the dark (14). Immediately thereafter, all samples were exposed (15 min) to a bright white light source to photochemically crosslink EMA in nonviable cells. Unincorporated EMA and surface antibodies were washed away prior to fixation (BD Cytotfix/Cytoperm kit, BD Biosciences, San Jose, CA) and subsequent intracytoplasmic staining for CD3, IL-2 PE, IFN γ APC, and TNF α FITC. After two washes in BD Perm/Wash and a final PBS/BSA wash, the cells were resuspended in 0.5% paraformaldehyde solution (PBS/BSA buffer). Stained cells from each donor were then divided into three separate aliquots prior to analysis on each cytometer.

Flow Cytometers and Data Acquisition

The optical components of the three instruments used in this comparison are summarized in Table 2. Each cytometer was configured with similar excitation laser lines and bandpass filter sets, but the optical paths to respective PMTs, laser-cell interrogation points (cuvette vs. stream-in-air), and laser output power varied between instruments. Uncompensated data were recorded for all experiments on all instruments. Since the design and goal of these experiments are not meant to find which platform is “better”, but rather to describe the

variability and bias when using different platforms while examining similar samples, we feel that it is a diversion to name the manufacturers of the specific instruments.

Separate single fluorochrome stained cells or antibody capture beads (BD Compbeads, BD, San Jose, CA) for each measured fluorochrome were acquired for each experiment on all instruments as controls for post-hoc software compensation. In some experiments, separate “fluorescence minus one” (FMO) controls were prepared for each donor and reagent panel using stimulated cells. FMO controls assist in the accurate definition of negative populations (5,9,12,15). A separate FMO control was prepared for each cytokine and consisted of all stains in Panel I or II, except for one of the cytokines under study.

Individual aliquots of stimulated and unstimulated cells were prepared and stained with two nine-color PFC reagent panels, and then refrigerated and protected from light prior to acquisition on each of three cytometers within a 24-h period after staining. At least 1×10^5 events were collected per condition on each instrument. Data were acquired using three different cytometers on three separate occasions. Samples were transported to each cytometer on ice, protected from light. The cytometers were located on the UC Davis campus and sample transport time to each cytometer was less than 5 min.

Flow Cytometric Data Analysis

All data were analyzed using FlowJo analysis software (Tree Star, Ashland, OR). Samples were analyzed with a single gating strategy (Fig. 1) in which naïve (N) CCR7⁺, CD45RA⁺; central memory (CM) CCR7⁺, CD45RA⁻; effector memory (EM) CCR7⁻, CD45RA⁻; and RA⁺ memory (RAM) CCR7⁻, CD45RA⁺ populations of CD4⁺ and CD8⁺ T cells were derived from scatter-gated, dump negative, CD3⁺ events (16,17). The total amount of IL-2:PE, IFN γ :APC, and TNF α :FITC cytokines produced by each of the CD4⁺ and CD8⁺ maturational subsets was calculated for stimulated and unstimulated cells. Spontaneous or background cytokine production in the unstimulated cultures was subtracted from the cytokine values reported for the stimulated samples. FMO controls were used to identify gate boundaries. Bivariate histograms are displayed using compensated data with 0 and negative log scaling (18).

We compared the mean percentages of CD4 and CD8 maturational subtypes measured by a 6-color anchor panel of T-cell surface antigens separately from the total amount of cytokine (IL-2, IFN γ , or TNF α) expressed by each maturational subtype. Background cytokine production in unstimulated cultures was subtracted from the mean percentages of total cytokine reported in each cell subtype. Interlaboratory variation was minimized by preparing the samples and acquiring the data at one location. Intradonor variation was controlled by using blood drawn at a single time point. Samples were prepared by different operators. Analysis was performed by a single operator using a manual gating process.

Statistical Analysis

Results were expected to vary among individuals, and since the number of donors in this study was small, person was treated as a fixed effect in all statistical analyses. To assess systematic differences in mean lymphocyte percentages, we carried out analysis of variance (ANOVA) of the effects of person, instrument and reagent, separately for data of

stimulated and unstimulated cells. We began with a full model with three-way interaction, all two-way interactions, and main effects for each of the three factors. A significant three-way interaction would suggest that some instrument–reagent combinations gave different estimates of the difference between people in average lymphocyte percentage. A significant two-way interaction between person and instrument, for example, would suggest that the estimates of between-person difference were not identical across instruments. We would expect significant between-person differences, but between-instrument or between-reagent differences might suggest some systematic bias. The results of ANOVA analyses did not identify any significant three-way or two-way interactions between lymphocyte subset and cytokine percentages within any person, reagent, and instrument combination in unstimulated or stimulated samples. Thus, in subsequent analyses, we considered only the main effects of reagent combination or instrument design on the mean percentages and cytokine expression of T cell subsets.

Differences in mean subset and cytokine percentages were assessed between instruments, between reagent combinations (Panel I vs. Panel II) and between donors. ANOVA tests were applied to data of stimulated and unstimulated cells to assess potential differences between donors, reagent combinations, and across instrument platforms in the mean percentages of T cell subsets, and to stimulated samples to compare mean cytokine percentages within T cell subsets. To assess variability of results within instruments or reagent panels, the means for each donor for the replicates measured by the same instrument and same reagent panel were computed. Then, the absolute deviations from that mean were calculated and compared between reagent panels using Wilcoxon signed rank tests and instruments using Kruskal–Wallis tests.

To provide experiment-wise protection against falsely reporting significant effects of the experimental factors, we tested the overall ANOVA at level 0.05 and tested the individual factors varied in the experiment, using prespecified orthogonal contrasts, only when the F test was significant (19). To preserve sensitivity to possible measurement problems and avoid missing important design aspects where care should be taken in measurement, we did not adjust for multiple comparisons, beyond the overall experiment-wise F test.

All statistical analyses were carried out using SAS/STAT software, version 9.0 (SAS Institute, 2004) and R statistical computing program (R Foundation for Statistical Computing, 2006).

RESULTS

Similar T cell Subset Frequencies Measured Between Reagent Panels

Reagent Panels I and II identified similar percentages of T cell surface phenotypes within each donor, across all instruments and replicates, in both stimulated (data not shown) and unstimulated samples (Fig. 2, upper panel). As expected, total percentages of CD4⁺ and CD8⁺ major and maturational subsets were different among the three donors. After accounting for interdonor variation, the frequency of total CD4 and CD8 cells and maturational subsets were virtually identical in unstimulated and stimulated cultures. Only minimal interpanel differences were observed in the mean percentages of isolated

maturational subsets. Specifically, Panel II measured slightly higher percentages of CD4 CM T cells than Panel I in unstimulated (Fig. 2, upper CD4 CM, Panel I vs. Panel II, respectively: donor $a = 27\%$ vs. 29% , donor $b = 26\%$ vs. 29% , donor $c = 29\%$ vs. 35% , $P = 0.001$) and SEB stimulated (data not shown, $P = 0.001$) cultures. No other differences were observed between the two reagent panels using unstimulated samples. In stimulated cultures, Panel I identified slightly higher percentages of CD4 N T cells than Panel II (Panel I vs. Panel II, respectively: donor $a = 53\%$ vs. 50% , donor $b = 53\%$ vs. 48% , donor $c = 49\%$ vs. 46% , $P = 0.039$).

Similar T Cell Subset Frequencies Measured Between Instruments

Only minor differences in total lymphocyte percentages were measured between the three different cytometers. After taking into account interdonor and between panel variation, the mean percentage of total CD4⁺ and CD8⁺ T cells were identical across all three instruments in unstimulated (Fig. 2, lower, shown for donor b only) and stimulated samples (data not shown). Small differences in the mean percentages of all CD4⁺ and two CD8⁺ maturational subtypes were measured in unstimulated samples, but were not consistently associated with a particular instrument. For example, using cells from any of the three donors stained with either panel, instrument C measured more CD4 N cells (cytometer A, B and C, respectively: 47, 51, 55% $P = 0.001$) and CD8 N cells (cytometer A, B, and C, respectively: 40, 42, 45%, $P = 0.001$), whereas instrument A measured slightly more CD4 RAM cells (cytometer A, B, and C, respectively: 5, 2, 2%, $P = 0.001$) and CD8 CM cells (cytometer A, B, and C, respectively: 12, 8, 7%, $P = 0.001$), but instrument B measured more CD4 CM cells (cytometer A, B, and C, respectively: 29, 31, 27%, $P = 0.028$). Fewer differences were noted in stimulated samples (data not shown) and were restricted to CD4 N cells (cytometer A, B, and C, respectively: 46, 51, 54%, $P = 0.003$) and CD4 RAM cells (cytometer A, B, and C, respectively: 6, 2.5, 2.5%, $P = 0.001$). Percentages represent summed data from both panels, all three donors, and three replicates per donor. Taken as a whole, both reagent panels gave similar results across donors on each instrument platform with only small differences observed in the actual mean percentages of unstimulated T cell subsets.

Reproducibility of T cell Subset Measurements Between Panels and Instruments

To assess reproducibility between reagents and across instruments, we calculated the mean percentages of T cell subsets for each donor for the three replicates measured by the same instrument and same reagent (triplet mean), and then computed the absolute deviations of each individual triplet measurement from the global triplet mean.

There was little evidence of interpanel variability in replicate experiments. CD4 and CD8 lineage subset frequencies were measured with comparable reproducibility with either reagent panel (unstimulated samples, data not shown, interquartile ranges: CD4: Panel I = 2%, Panel II = 2% and CD8 Panel I = 1.5%, Panel II = 1.5%) and in the maturational subsets of each T cell lineage (interquartile range Panel I vs. Panel II: CD4 CM (5 vs. 4.5%), N (4 vs. 6%) RAM (2 vs. 1.5%) EM (4.5 vs. 4.5%), and CD8 CM (2 vs. 3%), N (3 vs. 3.8%), RAM (1.5 vs. 2.5%), EM (2.5 vs. 4%) However, results for all CD4 maturational subsets were somewhat less consistent than CD8 memory subtypes with either staining panel

(CM, N, RAM, and EM: average CD4 interquartile range = 3.7, vs. average CD8 interquartile range = 2), particularly in CCR7+ CD4 memory subtypes (CM, N).

Variability within a particular instrument and between instruments was low in repeated measurements using unstimulated samples (Fig. 3). After allowing for interdonor and interpanel variation, mean percentages of CD4⁺ and CD8⁺ major and virtually all maturational subtypes varied by less than 4% on all cytometers. Interinstrument measurements were stable in CD8⁺ bulk and maturational subsets and no significant differences in CD8 subset frequencies were measured between instruments. However, CD4 memory subset measurements varied somewhat between cytometers, and in our system instrument C gave more consistent results than instrument A or B in repeated measurements of CD4 EM (cytometer A, B, and C: 4.4, 4.5, 2.5%, $P = 0.02$), RAM (cytometer A, B, and C, respectively: 3, 1, 1%, $P = 0.01$) and N (cytometer A, B, and C, respectively: 7, 5.6, 3.8%, $P = 0.01$) cell types.

Cytokine Detection Between Reagent Panels

We compared the total frequencies of IL-2, IFN γ , and TNF α producing cells within each of the four CD4⁺ and CD8⁺ maturational subtypes to determine if cytokine measurement varied significantly between reagent panels. The total mean percentage of IL-2, IFN γ , and TNF α producing CD4 and CD8 maturational subsets was similar between reagent panels in subsets in which cytokine expression levels exceeded 5% (Fig. 4). However, in subsets where cytokine expression intensity was low (< 5%, 11 total subsets), Panel I identified more cytokine positive cells than Panel II in seven subsets, in particular many IL-2 (6 of the 7) subsets and one IFN γ subset (Panel I vs. Panel II: CD4 N IL-2, 3.5 vs. 1.9%, $P = 0.02$, RAM IL-2, 4.2 vs. 2.7%, $P = 0.04$; CD8 CM IL-2, 4.2 vs. 1.5% $P = 0.001$, N IL-2, 1.5 vs. 0.1%, $P = 0.003$, RAM IL-2, 2.2 vs. 0.3%, $P = 0.003$, EM IL-2, 4.5 vs. 3.1%, $P = 0.05$, N IFN γ , 1 vs. 0%, $P < 0.0001$). As previously reported, Panel I generated lower levels of background expansion in the PE and APC channels than Panel II (12). The current findings show that the measurement of low intensity staining for IL-2 and IFN γ is enhanced when background expansion is minimized.

Cytokine Measurements Between Instruments

Mean frequencies of cytokine-producing cells were similar across all instruments and there were no significant differences noted in most maturational subsets (14 of total 24). However, in the remaining 10 subsets, instruments B and C consistently detected higher percentages of cytokine-positive cells than instrument A (Fig. 5). Interestingly, these differences tended to cluster within IL-2 and IFN γ positive subsets, although in contrast to interpanel comparisons, instrument C or B often demonstrated greater sensitivity regardless of cytokine expression intensity, or stated differently, the apparent enhanced sensitivity associated with instruments B and C occurred not only in subsets where cytokine expression was of low intensity, but also in subsets in which cytokine expression was robust. Interinstrument differences were more apparent in CD4 maturational subsets (7 subsets) than in CD8 (3 subsets). Although significant interinstrument differences in TNF α (FITC channel) production were restricted to a single CD4 subtype (CD4 CM, cytometer A, B, and C, respectively: 11.0, 16.3, 17.2%, $P = 0.002$), trends in the mean percentages of

TNF α (FITC channel) expressed by many other CD4 and CD8 subtypes also suggest that instruments B and C detect greater numbers of FITC+ events.

Reproducibility of Cytokine Measurements Between Panels and Across Instruments

Interpanel cytokine measurements generally were stable in repeated trials across all donors and instruments (data not shown). Variability between panels was observed in selected subsets (CD4, 3 subsets, CD8, 4 subsets), and was typically measured in low-frequency IL-2 and IFN γ positive subtypes stained with Panel I. In all other 17 subsets, a range of cytokine values were often measured per cytokine in each subset, but there was little evidence of instability associated with a particular reagent panel.

Cytokines were measured reliably in most CD4⁺ and CD8⁺ maturational cell types (17 total subtypes) regardless of instrument design. Some variation was evident, after allowing for interpanel and donor variability, in seven subsets measured by instrument C (data not shown). Specifically, instrument C measured a broader range of cytokine responses in these few subtypes, including CD4 N TNF α (75th percentile, cytometer A, B, and C, respectively: 4.6, 6.0, 8.2, $P=0.05$), RAM TNF α : 5.7, 3.0, 8, $P=0.001$) and often in the same subtypes (5 of 7) in which maximum responses were also recorded by this cytometer, including CD4 CM IFN γ (75th percentile, cytometer A, B, and C, respectively: 0.6, 1.4, 1.6, $P=0.01$, EM IFN γ : 2.2, 4.3, 5.2, $P=0.01$, EM IL-2: 3.5, 6.7, 9.7, $P=0.01$; CD8 RAM IL-2: 1.3, 0.6, 3, $P=0.001$), and EM IL-2: 2.2, 2.5, 6.5 $P=0.01$). Variability was not restricted to subsets where total cytokine production was low (both CD8 subsets, 2–3% responses overall), but often occurred in subsets in which cytokine responses were abundant (>5%). Newly described instrument characterization procedures would help to standardize PMT voltage ranges and further reduce variability in replicate experiments (15).

DISCUSSION

This study was conducted to investigate polychromatic reagent panel performance across various instrument platforms using samples prepared at a single research site. As such, it is the first report of bias and variability of optimized nine-color reagent panels using differently configured cytometers and provides information for the design of larger, interlaboratory comparisons of polychromatic reagent panels aimed at determining the usefulness of PFC in clinical trials. In a prior, comprehensive multicenter study of antigen-specific T cell cytokine responses conducted by Maecker et al., 4-color reagent panels (3 phenotypic markers plus IFN γ) were used to assess intersite variability and concordance in sample preparation types (20). In their study, the authors compared sample preparation and data analysis methods to refine the results obtained at various research sites. Although they used at least two different 4-color reagent panels and acquired data across multiple cytometer platforms, they did not discuss potential interpanel variation or interinstrument differences between research sites. In the current study, our analytic goals were to investigate the sensitivity and reproducibility of nine individual T-cell subset and cytokine measurements between reagent panels and across instruments to determine if results were influenced by (1) the unique set of spectral overlaps and measurement errors

associated with each reagent panel and (2) differences in the optical and electronic design of the cytometers.

Our results demonstrate that similar and reliable T-cell subset measurements can be made using appropriately designed and optimized reagent panels. Because the two panels tested in this study had been selected based on superior performance in a prior comparison of thirty 6-color T-cell surface antigen anchor panels, little difference in T-cell subset frequencies between panels were anticipated (12). Our current results further confirm and extend our previous findings to show that carefully optimized reagent panels will measure major and maturational T-cell subsets with similar precision across different instrument platforms, though possible insights into subtle differences in results between the two reagent panels will be discussed below to illustrate relevant points.

Interinstrument comparisons detected small differences in the mean frequency of CD4 and CD8 maturational subtypes (but not in total CD4 and CD8 percentages) in unstimulated samples. The cytometers used in this study were configured with similar bandpass filter sets to detect the same fluorochromes, but differed in their optical and flow geometry. Specifically, instrument (C) was a bench-top analyzer whereas the other two cytometers were cell sorters [jet-in-air (A), cuvette (B)]. Despite these differences in instrument design, the magnitude of the difference in T cell subtype percentages measured between cytometers was small and not confined to a particular instrument. Thus in the context of our system, there is little evidence upon which to conclude that one instrument measured T cell subset frequencies with more sensitivity than another, and overall results were comparable between cytometers. On the other hand, one explanation for the slight variability we detected between cytometers in selected CD4+ maturational subsets, (N, EM, RAM) could be due to differences in instrument design, since the cell sorters, especially the jet-in-air cytometer, often required more effort to align and optimize before each acquisition session. Even greater efforts to achieve pre-acquisition instrument standardization could further optimize the results (21).

The total mean percentages of cytokines expressed per T cell maturational subset were similar between the two reagent panels, especially in subsets in which cytokine responses to SEB stimulation were abundant (>5%), as reported previously (20). However, in subsets in which total cytokine expression was low (<5%), higher frequencies of IL-2 PE and IFN γ APC+ cells were measured in samples that were stained with reagent Panel I. To determine if Panel I also measured more nonspecific background cytokine expression in these subsets, we investigated cytokine expression between panels using data for unstimulated cells, and found only four differences (CD4 N IFN γ , $P=0.03$; CD4 RAM IFN γ $P=0.05$, CD8 CM TNF α , $P=0.04$, CD8 N TNF α , $P=0.01$) that did not correspond with the subsets in which Panel I detected more cytokine than Panel II. As demonstrated in a companion manuscript comparing multiple 6-color reagent panels, we measured less background expansion in the PE and APC channels in samples stained with reagent Panel I (or Panel I minimized background expansion into the PE and APC channels more than other panels). It is likely that because Panel I minimized background expansion into the PE (IL-2) and APC (IFN γ) channels, a greater number of low-intensity, cytokine positive cells were identified in stimulated samples stained with this reagent combination.

Interpanel cytokine measurements were stable in the majority of maturational cell types. The results obtained with Panel I varied somewhat more than Panel II in most (6 out of 7) subsets in which significant variation was computed. In four of the six subsets in which significant interpanel variation was calculated, Panel I also detected a higher percentage of cytokine-positive cells. It is possible that a greater number of replicate experiments on a larger sample of individuals could have improved precision.

Interinstrument cytokine measurements were generally similar across cytometer platforms, although in selected maturational subtypes (10 of 24 total subtypes), instruments B or C measured a greater total mean percentage of cytokine positive cells, particularly within CD4+ IL-2 (PE) and IFN γ (APC) positive subsets, regardless of whether cytokine expression intensity was low or high. Significant background cytokine production was only observed in one subset (CD4 CM IFN γ , $P=0.01$), instruments B and C>A) in which significant interinstrument differences were also measured in stimulated samples, excluding the possibility that these results reflect the contribution of excessive measurement artifacts in these subsets. One solution to normalize these differences between instruments further would include a more rigorous preacquisition quality control procedure to minimize interinstrument bias in replicate experiments beyond the standard procedures used in core facility maintenance procedures (15,22,23). For example, median fluorescent intensity targets could have been developed for each fluorochrome using bright singly stained cells to set regions of optimal PMT voltage for all stains in each panel. Before each acquisition, uniformly fluorescent particles (e.g., single-peak Rainbow Beads) could then be used to adjust PMT voltages such that peak bead fluorescence falls within the target regions established for each detector/fluorochrome with the singly stained cells. In this way, it would be possible to determine if a particular cytometer needed additional adjustment prior to data collection. The use of such a “global” quality control procedure would likely assure a more equal level of sensitivity between cytometers, and also may have helped to reduce the variability in cytokine measurements recorded on instrument C.

In summary, our results highlight the need to carefully titrate and test candidate polychromatic panel reagents to achieve reliable identification of cell subsets, while maximizing measurement sensitivities. Once an optimal reagent combination is identified, it is possible to achieve good agreement of results obtained with different cytometers. Careful quality control procedures, however, must be in place for all instruments in multiplatform studies. These results support the use of different instruments in multicenter clinical trials when measurement of fresh samples or a distribution of labor is sought. A greater number of donors and the application of global instrument quality control procedures should help to minimize differences between instruments further and hasten the application of polychromatic approaches in clinical trials.

Acknowledgments

The authors thank Carol Oxford, Abigail Spinner, Gil Reinin (Becton Dickinson), and Dr. Taiwo Akande for technical advice and assistance.

Grant sponsor: NIH; Grant number: AI051999; Grant sponsors: Research Facilities Improvement Program, National Center for Research Resources, National Institutes of Health; Grant number: C06RR-12088-01.

LITERATURE CITED

1. De Rosa SC, Lu FX, Yu J, Perfetto SP, Falloon J, Moser S, Evans TG, Koup R, Miller CJ, Roederer M. Vaccination in humans generates broad T cell cytokine responses. *J Immunol* 2004;173:5372–5380. [PubMed: 15494483]
2. Betts MR, Nason MC, West SM, De Rosa SC, Migueles SA, Abraham J, Lederman MM, Benito JM, Goepfert PA, Connors M, Roederer M, Koup RA. HIV nonprogressors preferentially maintain highly functional HIV-specific CD8+ T cells. *Blood* 2006;107:4781–4789. [PubMed: 16467198]
3. Precopio ML, Betts MR, Parrino J, Price DA, Gostick E, Ambrozak DR, Asher TE, Douek DC, Harari A, Pantaleo G, Bailer R, Graham BS, Roederer M, Koup RA. Immunization with vaccinia virus induces polyfunctional and phenotypically distinctive CD8(+) T cell responses. *J Exp Med* 2007;204:1405–1416. [PubMed: 17535971]
4. Mattapallil JJ, Douek DC, Buckler-White A, Montefiori D, Letvin NL, Nabel GJ, Roederer M. Vaccination preserves CD4 memory T cells during acute simian immunodeficiency virus challenge. *J Exp Med* 2006;203:1533–1541. [PubMed: 16735692]
5. Baumgarth N, Roederer M. A practical approach to multicolor flow cytometry for immunophenotyping. *J Immunol Methods* 2000;243:77–97. [PubMed: 10986408]
6. Perfetto SP, Chattopadhyay PK, Roederer M. Seventeen-colour flow cytometry: Unravelling the immune system. *Nat Rev Immunol* 2004;4:648–655. [PubMed: 15286731]
7. Roederer M, Brenchley JM, Betts MR, De Rosa SC. Flow cytometric analysis of vaccine responses: How many colors are enough? *Clin Immunol* 2004;110:199–205. [PubMed: 15047198]
8. Roederer M Compensation is not dependent on signal intensity or on number of parameters. *Cytometry* 2001;46:357–359. [PubMed: 11754206]
9. Roederer M Spectral compensation for flow cytometry: Visualization artifacts, limitations, and caveats. *Cytometry* 2001;45:194–205. [PubMed: 11746088]
10. Roederer M, De RS, Gerstein R, Anderson M, Bigos M, Stovel R, Nozaki T, Parks D, Herzenberg L, Herzenberg L. 8 Color, 10-parameter flow cytometry to elucidate complex leukocyte heterogeneity. *Cytometry* 1997;29:328–339. [PubMed: 9415416]
11. Bigos M, Baumgarth N, Jager GC, Herman OC, Nozaki T, Stovel RT, Parks DR, Herzenberg LA. Nine color eleven parameter immunophenotyping using three laser flow cytometry. *Cytometry* 1999;36:36–45. [PubMed: 10331625]
12. McLaughlin BE, Baumgarth N, Bigos M, Roederer M, De Rosa SC, Altman JD, Nixon DF, Ottinger J, Oxford C, Evans TG, Asmuth DM. Nine color flow cytometry for accurate measurement of T cell subsets and cytokine responses. Part I: Panel design by an empiric approach. *Cytometry A* 2008; 73A: in press. doi: 10.1002/cyto.a.20555 (this issue).
13. Horton H, Thomas EP, Stucky JA, Frank I, Moodie Z, Huang Y, Chiu YL, 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. [PubMed: 17451739]
14. Riedy MC, Muirhead KA, Jensen CP, Stewart CC. Use of a photolabeling technique to identify nonviable cells in fixed homologous or heterologous cell populations. *Cytometry* 1991;12:133–139. [PubMed: 2049970]
15. Maecker HT, Trotter J. Flow cytometry controls, instrument setup, and the determination of positivity. *Cytometry Part A* 2006;69A:1037–1042.
16. Sallusto F, Lenig D, Forster R, Lipp M, Lanzavecchia A. Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. *Nature* 1999;401:708–712. [PubMed: 10537110]
17. Geginat J, Lanzavecchia A, Sallusto F. Proliferation and differentiation potential of human CD8+ memory T-cell subsets in response to antigen or homeostatic cytokines. *Blood* 2003;101:4260–4266. [PubMed: 12576317]
18. Parks DR, Roederer M, Moore WA. A new “Logicle” display method avoids deceptive effects of logarithmic scaling for low signals and compensated data. *Cytometry Part A* 2006;69A:541–551.
19. Lehmann E Testing Statistical Hypotheses. New York: Springer; 1986.

20. Maecker HT, Rinfret A, D'souza P, Darden J, Roig E, Landry C, Hayes P, Birungi J, Anzala O, Garcia M, Harari A, Frank I, Baydo R, Baker M, Holbrook J, Ottinger J, Lamoreaux L, Epling CL, Sinclair E, Suni MA, Punt K, Calarota S, El-Bahi S, Alter G, Maila H, Kuta E, Cox J, Gray C, Altfeld M, Nougarede N, Boyer J, Tussey L, Tobery T, Brecht B, Roederer M, Koup R, Maino VC, Weinhold K, Pantaleo G, Gilmour J, Horton H, Sekaly RP. Standardization of cytokine flow cytometry assays. *BMC Immunol* 2005;6:13. [PubMed: 15978127]
21. Wood JC, Hoffman RA. Evaluating fluorescence sensitivity on flow cytometers: An overview. *Cytometry* 1998;33:256–259. [PubMed: 9773888]
22. Perfetto SP, Ambrozak D, Nguyen R, Chattopadhyay P, Roederer M. Quality assurance for polychromatic flow cytometry. *Nat Protoc* 2006;1:1522–1530. [PubMed: 17406444]
23. Establishing Optimum Baseline PMT Gains to Maximize Resolution on BD Biosciences Digital Flow Cytometers. Becton Dickinson Hotlines 2006; 10:winter 2006. Available at www.bdbiosciences.com/pdfs/whitePapers/23-8389-00.pdf.

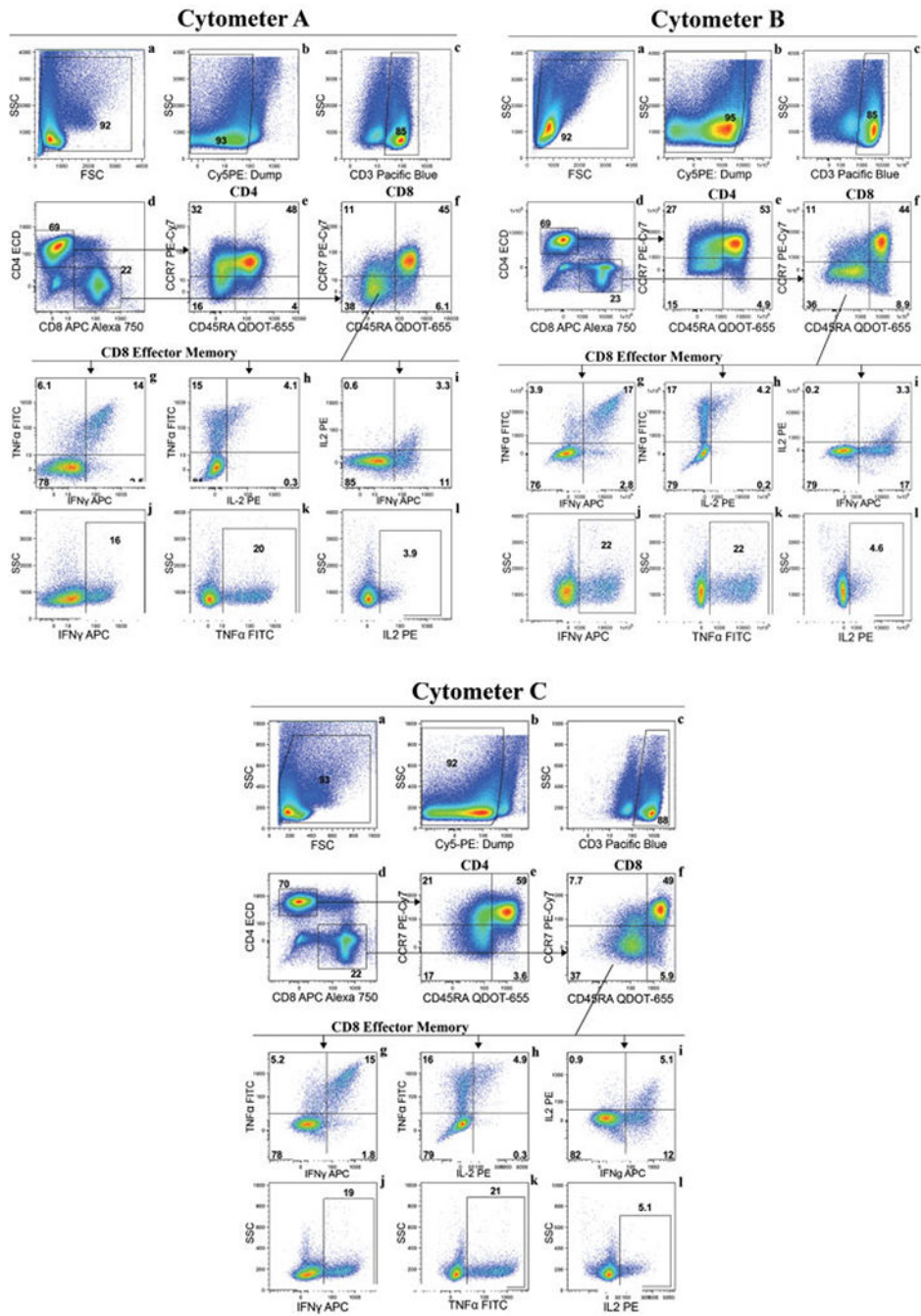


Figure 1. Gating strategy used to identify cytokine producing T cells between panels and across instrument platforms. Shown are pseudocolor plots for cytokine-producing CD8+ EM T cells from donor b stained with Panel II and recorded on three different cytometers (cytometer A, cytometer B, and cytometer C). A standardized gating strategy was used to analyze samples acquired on each cytometer with either Panel I or II. Unstimulated samples and FMO controls for each cytokine and for CD45RA and CCR7 maturational markers were used to determine optimal gate placement for each experiment. Numbers

within gates or quadrant regions indicate the percentage of cells. Gates were drawn to identify (a) lymphocytes by side and forward scatter, (b) exclude “dump” positive cells, (c) select only CD3⁺ T cells, (d) divide CD3⁺ T cells into CD4⁺ and CD8⁺ populations, and further subdivide CD4⁺ cells (e), and CD8⁺ cells (f) into naïve (N) CCR7⁺, CD45RA⁺; CM CCR7⁺, CD45RA⁻; effector memory (EM) CCR7⁻, CD45RA⁻; and RA⁺ memory (RAM) CCR7⁻, CD45RA⁺ subpopulations. Cytokine-producing CD4⁺ and CD8⁺ cells within each memory subpopulation were compared for selected pairs of cytokines as shown for CD8⁺ EM T cells in panels (g) TNF α FITC vs. IFN γ APC, (h) TNF α FITC vs. IL-2 PE, and (i) IL-2 PE vs. IFN γ APC. The total amount of each cytokine produced by CD8⁺ EM T cells as measured by each cytometer is shown in (j) IFN γ , (k) TNF α , and (l) IL-2.

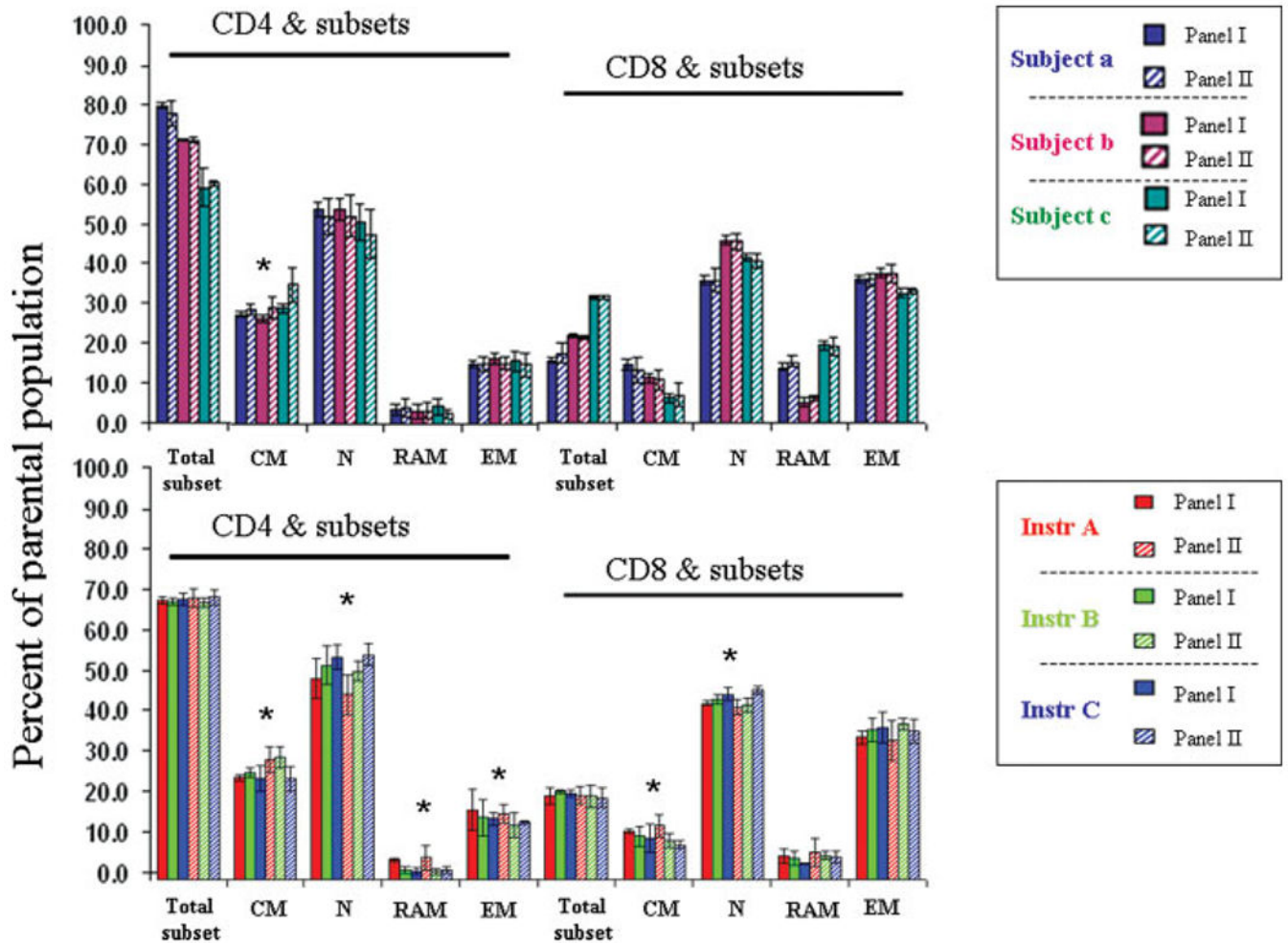


Figure 2.

Bias and variability of T cell subset measurements between panels and across instruments. Shown are the means (\pm SE) of T cell subset frequencies measured by reagent Panels I and II using unstimulated samples. Similar subset frequencies were recorded within individual donors using either panel (a). Estimates of subset frequency varied between cytometers in most maturational subsets, but this variability was low and not consistently associated with a particular instrument (shown in (b), donor b only).

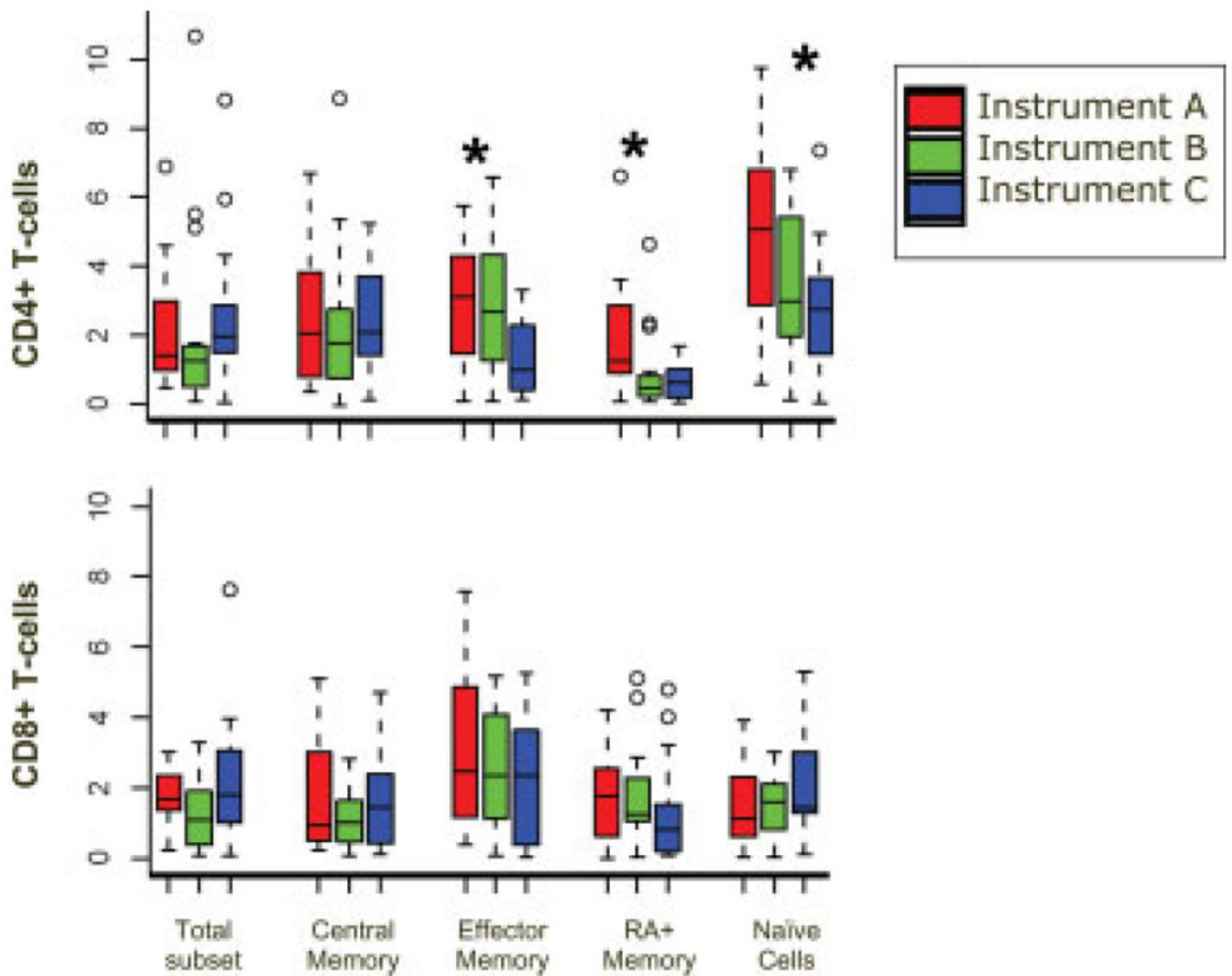


Figure 3.

Interinstrument variation is confined to three CD4 maturational subsets. Boxplot diagrams illustrate the absolute variance from the mean for CD4 (a) and CD8 (b) subset percentages measured by each instrument (cytometer A, red bars; cytometer B, green bars; cytometer C, blue bars) using unstimulated samples. Deviations across instruments were compared using Kruskal–Wallis tests. The boxplot diagrams illustrate the magnitude of the variance from the mean for each instrument: the colored bar represents the middle 50% of the variance, ranging from the 25th to 75th percentile, the median variance is shown as a horizontal black bar within the colored bars, dashed vertical lines capped with a horizontal bar indicate the extreme minimum and maximum data values, with open circles used to show outliers. The results indicate that the variance in T cell subset measurements is low across all instruments, with median values ~2% above or below the mean in most subsets. Significant differences (indicated by *) in the variance from the mean between instruments were confined to only three CD4 maturational subsets, CD4 EM ($P = 0.02$), CD4 RAM ($P = 0.01$), and CD4 N ($P = 0.01$).

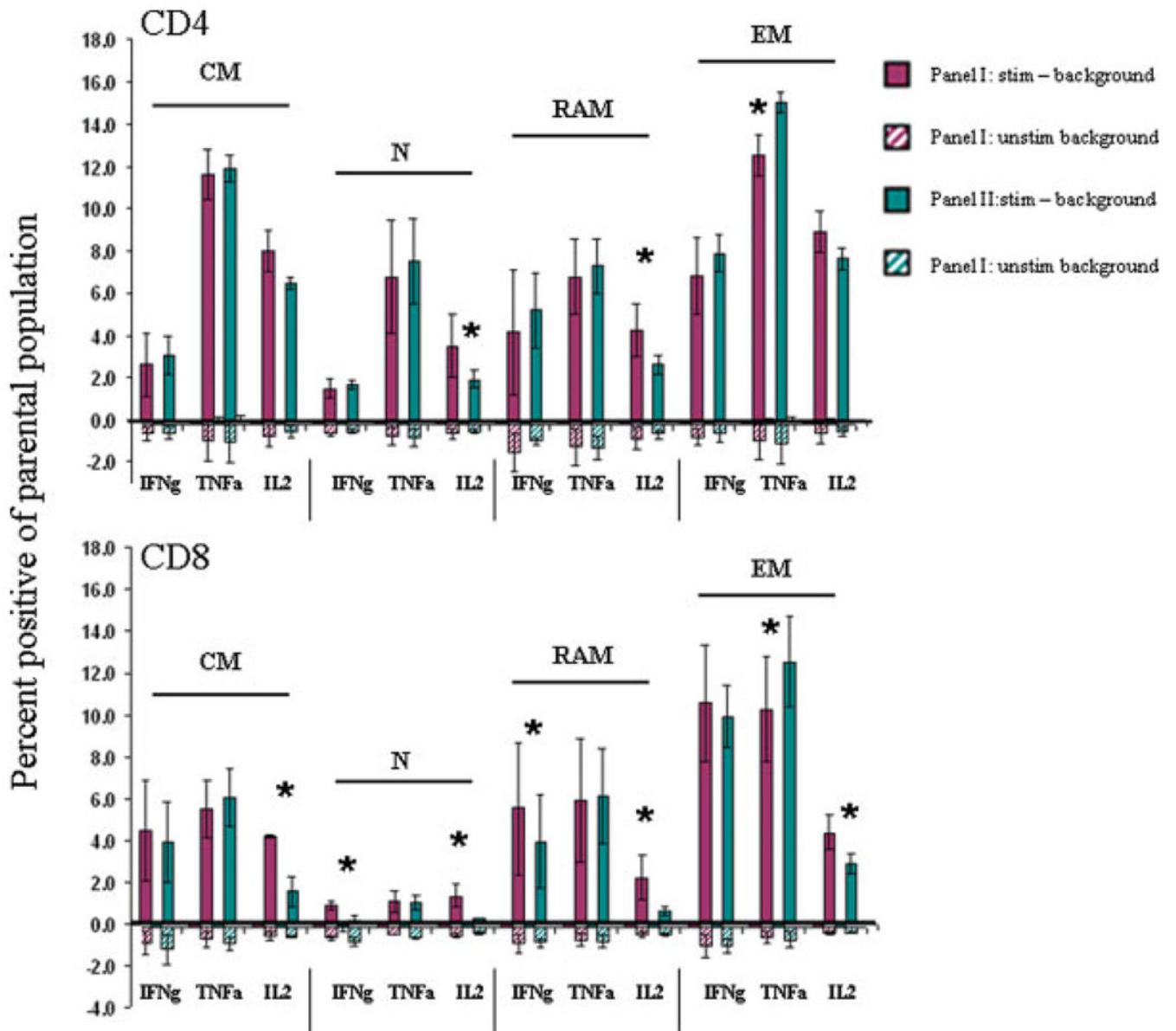


Figure 4.

Cytokine measurements are similar between reagent Panels I and II. Total mean (\pm SE) cytokine expression of each of the maturational subsets (CD4 upper panel, CD8 lower panel) following SEB stimulation. Since it was expected that cytokine expression between individuals would vary, the analysis of variance between reagent panels was calculated using results of the three replicates of each donor across all three instruments, treating each donor as a fixed effect. Significant differences were observed in selected subsets, as indicated (*). With the exception of the CD8 N IFN γ ($P=0.001$) subset, these differences tended to cluster in IL-2+ subsets (CD8: CM ($P=0.004$), N ($P=0.01$), RAM [$P=0.04$], and CD4 N ($P=0.02$)). In each of these subsets, higher cytokine production was measured in samples stained with Panel I.

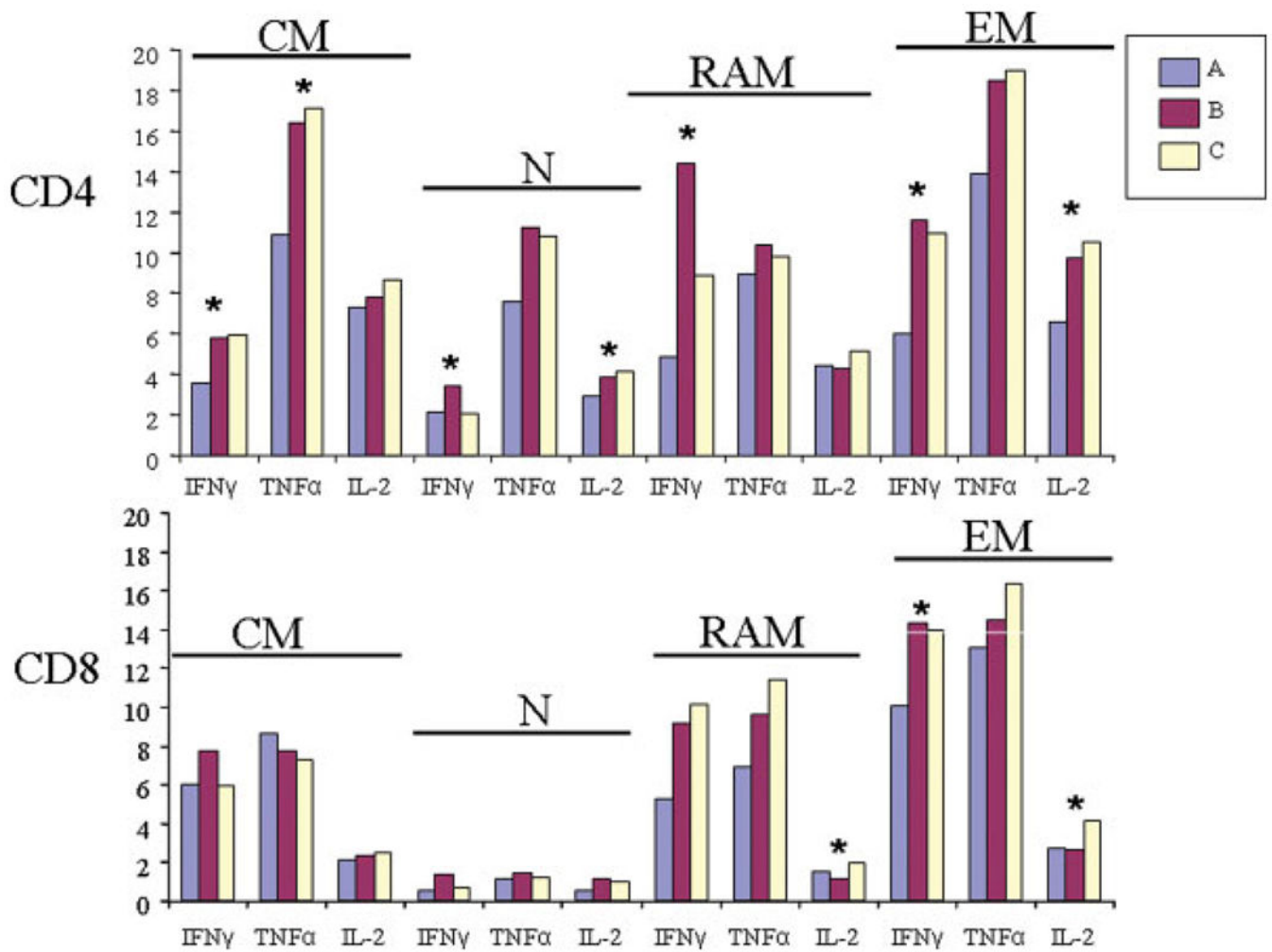


Figure 5.

Interinstrument cytokine measurements. The total mean percentage of cytokine-positive CD4⁺ (top) and CD8⁺ (bottom) T cells from donor b are summarized for each cytometer (instrument A, blue bars; instrument B, red bars; and instrument C, yellow bars). Cytokine measurements were similar between instruments for most CD8⁺ maturational subsets (CM, central memory; N, naïve; RAM, CD45RA⁺ effector memory; EM, effector memory), but instrument B and/or C often recorded higher percentages of cytokine+ cells in CD4⁺ maturational subsets. Significant differences were observed in selected subsets, as indicated (*).

Table 1.

Six-color T Cell immunophenotyping “anchor” reagent panels

	CD14 ^a	CD19 ^b	EMA	CD3	CD4	CD8	CD45RA	CCR7
Panel 1	PE-Cy5	Alexa 700 ^c	Pacific Blue ^d	APC-Alexa 750 ^f	Qdot 655 ^g	PE-Cy7 ^h		
Panel 2	PE-Cy5	Pacific Blue ^c	ECD (PE-TR) ^e	APC-Alexa 750	Qdot 655	PE-Cy7		

Monoclonal antibody clones:

^aTüK4.^bH1B19.^cUCHT-1.^dRPA-T4.^eSFC112T4D11 (T4).^f3B.5.^gMEM-56.^h3D12.

Table 2.

Cytometer laser lines, output power, and bandpass filter characteristics (FWHM)*

	LASER EMISSION AND OUTPUT POWER								
	INSTRUMENT A			INSTRUMENT B			INSTRUMENT C		
	407 nm	488 nm	647 nm	405 nm	488 nm	633 nm	403 nm	488 nm	637 nm
	150 mW	200 mW	200 mW	15 mW	100 mW	20 mW	50 mW	100 mW	25 mW
FLUOROCROME	BANDWIDTH TRANSMITTED TO PMT								
Pacific Blue	417.5–482.5			425–475			425–475		
QD 655	655–685			645–675			645–665		
FITC		505–555			505–555			500–550	
PE		565–595			563–589			564–606	
PE-Texas Red		599.5–624.5			600–620			605–622	
PE-Cy5		655–685			645–685			655–685	
PE-Cy5.5		697.5–742.5			697.5–742.5			685–705	
PE-Cy7		750 LP			750 LP			750–810	
APC			655–685			640–680			663–677
Alexa 700			685–735			685–735			685–740
APC Alexa 750			750 LP			740 LP			750–810

* Full width at half maximum (FWHM).