

A standardized immune phenotyping and automated data analysis platform for multicenter biomarker studies

Sabine Ivison,^{1,2} Mehrnosh Malek,³ Rosa V. Garcia,^{1,2} Raewyn Broady,⁴ Anne Halpin,⁵ Manon Richaud,⁶ Rollin F. Brant,² Szu-I Wang,⁵ Mathieu Goupil,⁶ Qingdong Guan,⁷ Peter Ashton,⁸ Jason Warren,⁹ Amr Rajab,¹⁰ Simon Urschel,⁵ Deepali Kumar,⁸ Mathias Streitz,¹¹ Birgit Sawitzki,¹¹ Stephan Schlickeiser,¹¹ Janetta J. Bijl,⁶ Donna A. Wall,⁷ Jean-Sebastien Delisle,⁶ Lori J. West,⁵ Ryan R. Brinkman,^{3,12} and Megan K. Levings,^{1,2}

¹Department of Surgery, University of British Columbia, Vancouver, British Columbia, Canada. ²BC Children's Hospital Research Institute, Vancouver, British Columbia, Canada. ³Terry Fox Laboratory, BC Cancer, Vancouver, British Columbia, Canada. ⁴Department of Medicine, University of British Columbia, Vancouver, British Columbia, Canada. ⁵Alberta Transplant Institute, University of Alberta, Edmonton, Alberta, Canada. ⁶Hôpital Maisonneuve-Rosemont, University of Montreal, Montreal, Quebec, Canada. ⁷Department of Pediatrics and Child Health/Internal Medicine, University of Manitoba/Cancer Care Manitoba, Winnipeg, Manitoba, Canada. ⁸Toronto General Research Institute, University of Toronto, Toronto, Ontario, Canada. ⁹Health Sciences Centre, Diagnostic Services Manitoba, Winnipeg, Manitoba, Canada. ¹⁰Department of Laboratory Medicine, Toronto General Hospital, Toronto, Ontario, Canada. ¹¹Institute of Medical Immunology, Charité, Universitätsmedizin Berlin, Berlin, Germany. ¹²Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada.

The analysis and validation of flow cytometry-based biomarkers in clinical studies are limited by the lack of standardized protocols that are reproducible across multiple centers and suitable for use with either unfractionated blood or cryopreserved PBMCs. Here we report the development of a platform that standardizes a set of flow cytometry panels across multiple centers, with high reproducibility in blood or PBMCs from either healthy subjects or patients 100 days after hematopoietic stem cell transplantation. Inter-center comparisons of replicate samples showed low variation, with interindividual variation exceeding inter-center variation for most populations (coefficients of variability <20% and interclass correlation coefficients >0.75). Exceptions included low-abundance populations defined by markers with indistinct expression boundaries (e.g., plasmablasts, monocyte subsets) or populations defined by markers sensitive to cryopreservation, such as CD62L and CD45RA. Automated gating pipelines were developed and validated on an independent data set, revealing high Spearman's correlations ($r_s > 0.9$) with manual analyses. This workflow, which includes pre-formatted antibody cocktails, standardized protocols for acquisition, and validated automated analysis pipelines, can be readily implemented in multicenter clinical trials. This approach facilitates the collection of robust immune phenotyping data and comparison of data from independent studies.

Conflict of interest: RRB has an ownership interest in Cytapex Bioinformatics Inc.

License: Copyright 2018, American Society for Clinical Investigation.

Submitted: May 1, 2018

Accepted: October 29, 2018

Published: December 6, 2018

Reference information:

JCI Insight. 2018;3(23): e121867.

<https://doi.org/10.1172/jci.insight.121867>.

<https://doi.org/10.1172/jci.insight.121867>.

Introduction

Flow cytometric analysis of peripheral blood is often used to track complex changes in leukocyte phenotype, number, and proportion during the course of disease or in response to therapy. In the era of precision medicine, there is an increasing need to rapidly assess complex leukocyte phenotypes in order to stratify patients, assign individualized treatments, and monitor response to therapy. Many studies have used flow cytometry to discover leukocyte-based biomarkers (1–3) in different disease contexts, but validation of these discoveries and their ultimate transition into routine clinical practice are limited by poor standardization of research-based flow cytometric methods. Consequently, biomarker and immune monitoring studies often rely on centralized and batched analyses of cryopreserved PBMCs to overcome site- and time-dependent variation in reagents, sample processing, and instrument performance. Moreover, the highly subjective and

Table 1. Summary of DuraClone panels used in this study: panel design with markers and fluorochromes used

Laser	Fluorochrome	Panel					
		Basic	TCR	T-ACT	T-MEM-REG	B cell	DC
Blue 488 nm	FITC	CD16	TCR $\gamma\delta$	CD57	CD127	IgD	BDCA3
	PE	CD56	TCR $\alpha\beta$	CD28	CCR7	CD21	LIN
	ECD	CD19	CD45RO	HLA-DR	CD62L	CD19	CD123
	PC7	CD14		CD27	CD25	CD27	CD11c
Red 633 nm	APC	CD4	CD4	CD4	CD4	CD24	BDCA2
	A700	CD8	CD8	CD8	CD8		
	APC-A750	CD3	CD3	CD3	CD3	CD38	CD16
Violet 405 nm	PacBlue	CD64		CD45RA	CD45RA	IgM	HLA-DR
	KrOrange	CD45	CD45	CD45	CD45	CD45	CD45

labor-intensive nature of manual data analysis introduces a further source of variation (4, 5). Therefore, methodology to standardize the collection and analysis of flow cytometry data in multicenter validation studies is needed to facilitate the clinical use of flow cytometry–based precision medicine.

Previous work to standardize complex flow cytometry was driven by the need to detect and classify hematological malignancies (6–9). Accordingly, several consensus-driven flow cytometry panels (with 8 or more colors) were developed to classify blood cancers, some of which are now certified for in vitro diagnostic use (10). Several research groups have also developed standardized flow cytometry panels for comprehensive leukocyte phenotyping, including the Human Immunology Project Consortium (HIPC) (4, 11) and the ONE Study (12, 13). Most recently, HIPC used lyophilized antibody mixes to analyze lyophilized control cells or cryopreserved PBMCs from 3 healthy subjects at 9 sites (14). Standardization using PBMCs was found to be feasible, but only for abundant and well-defined subsets. Populations in the T helper cell panel, as well as poorly resolved, low-abundance populations, could not be reliably measured in replicate aliquots (14).

Taking an alternate approach, the ONE Study used liquid antibody cocktails to phenotype unfractionated blood within 4 hours of collection (12, 13, 15). This method resulted in good standardization, but in the context of multicenter studies, it may not always be financially or logistically feasible for each center to analyze blood within such a short time frame. Thus, there is the need for a standardized platform that can be used to analyze either cryopreserved PBMCs or unfractionated blood, resulting in data that are sufficiently robust to enable inter-center and inter-study comparisons.

The Canadian National Transplant Research Program (CNTRP) is a national network of clinicians and researchers who aim to improve the access to, and success of, transplantation (16). As part of the CNTRP clinical trials initiatives, we aimed to implement a flexible, yet robust, platform for standardized immune monitoring that could readily be implemented across multiple centers. Here we describe an approach that uses commercially available, pre-formulated reagents and enables standardized collection of flow cytometry data from unfractionated blood or cryopreserved PBMCs from healthy subjects or lymphopenic hematopoietic stem cell transplant (HSCT) recipients. Automated gating pipelines were also developed, eliminating the need for labor-intensive and subjective manual data analysis. This methodology can be readily implemented in single or multi-center studies and sets the stage for inter-study comparisons and the translation of research findings into clinical practice.

Results

Standardized multi-center immune phenotyping of cryopreserved PBMCs from healthy adults. Immediate analysis of unfractionated blood is typical in clinical immunology labs, but is costly and, in the context of multi-center clinical trials, logistically complex. The ONE Study previously developed a set of leukocyte phenotyping panels and protocols for standardized immune monitoring of unfractionated blood (12). Here, we first investigated the ability to implement similar methods for analysis of cryopreserved PBMCs. Whereas the ONE Study used liquid antibody cocktails, we took advantage of DuraClone dry reagent technology (Beckman Coulter) to manufacture pre-formatted antibody cocktails. The panels, shown in Table 1, were constructed using consensus definitions for the most common and well-defined subsets of mononuclear leukocytes (11, 12).

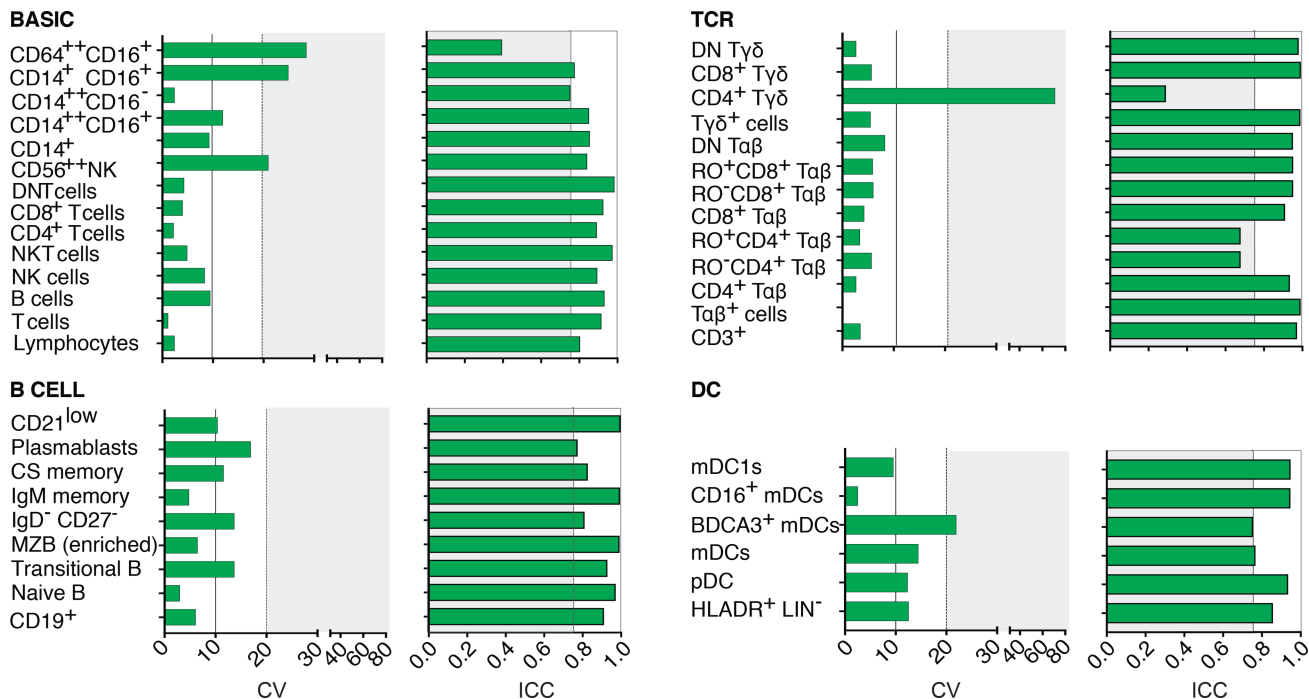


Figure 1. Inter-center comparison of populations quantified in cryopreserved PBMCs from healthy subjects with the Basic, TCR, B cell, or DC panels. Replicate aliquots of cryopreserved PBMCs from 5 subjects were analyzed at 5 different sites using the indicated DuraClone panels. Raw data (LMD files) were analyzed centrally, and the reproducibility of population proportions was determined by statistical analysis. All calculations were based on the cell proportion relative to the parent gate; the identity of parent gates for all variables is listed in Supplemental Table 1. Shown are CV and ICC values for each population. CV and ICC values indicative of poor reproducibility (i.e., CV >20% and/or ICC <0.75) are shaded in gray. CS memory, class-switched memory; DNT cells, double-negative T cells; mDCs, myeloid DCs; MZB, marginal zone B cells; pDCs, plasmacytoid DCs; RO, CD45RO.

We first measured the degree of center-to-center variability using blood collected at 3 different centers (centers 1–3), which was processed and cryopreserved in replicate aliquots using a consensus standard operating procedure (SOP) derived from previous publications (17, 18). Aliquots from 5 subjects were distributed to 5 centers for staining and acquisition of a total of 25 samples for each panel on identically calibrated flow cytometers (see Methods for details). As intra-center variation is lower than inter-center variation (14), replicates within a center were not included. To minimize variability introduced by manual analysis (4, 19), the data were initially analyzed by a single analyst following the gating strategy shown in Supplemental Figure 1; supplemental material available online with this article; <https://doi.org/10.1172/jci.insight.121867DS1> (12).

Center-to-center variability in the proportions of each cell population in the Basic, TCR (T cell receptor-expressing cells), B cell, and DC panels for each individual was quantified as coefficients of variability (CVs) or intra-class correlation coefficients (ICCs) (Figure 1). For the majority of populations, the data obtained at each center were similar, with CVs <10% in the Basic and TCR panels and <15% for the B cell and DC panels. As CVs are calculated by dividing SD by population size, rare populations (e.g., BDCA3+ DCs, CD4+ $\gamma\delta$ T cells, both <0.05% of PBMCs) tended to have higher CVs (22% and 64%, respectively). This effect was exacerbated in populations with low event counts or poorly defined boundaries, such as CD14++ or CD56++ cells, leading to high CVs (between 20% and 30%) for some low-abundant subtypes of monocytes and NK cells.

We next quantified center-to-center variability using the ICC, a statistical method that is independent of population size (20). The ICC estimates the ratio of biological to total variability, with an ICC of 0.5 being the threshold above which biological variation exceeds technical variability. As shown in Figure 1, most populations in the Basic (12 of 13), TCR (10 of 13), B cell (9 of 9), and DC (5 of 6) panels had ICC values >0.75, demonstrating that this method of standardized analysis of cryopreserved PBMCs results in low relative variation even for rare cell populations.

Cryopreservation and cell handling affect reproducible phenotyping of selected cell populations. In contrast to the excellent center-to-center reproducibility for the 4 panels described above, there was considerable variability

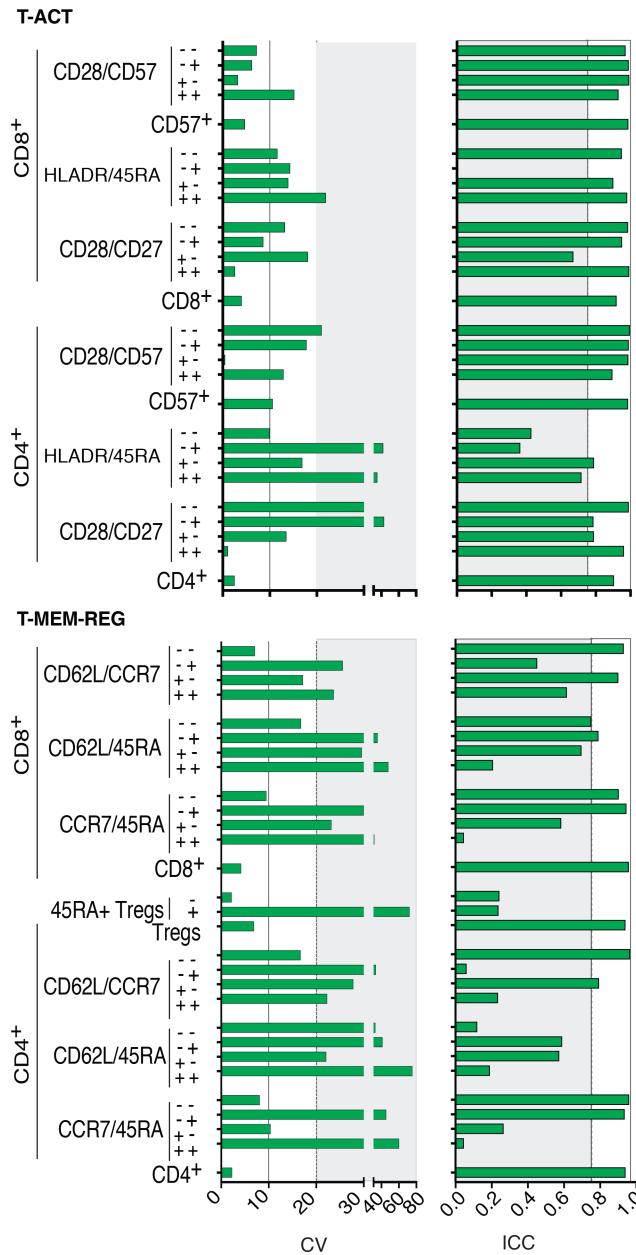


Figure 2. Inter-center comparison of populations quantified in cryopreserved PBMCs from healthy subjects with the T-ACT and T-MEM-REG panels. Replicate aliquots of cryopreserved PBMCs from 5 subjects were analyzed at 5 different sites using the indicated DuraClone panels. Raw data were analyzed centrally, and the reproducibility of population proportions was determined by statistical analysis. All calculations are based on the cell proportion relative to the parent gate; the identity of parent gates for all variables is listed in Supplemental Table 1. Shown are CV and ICC values for each population. CV and ICC values indicative of poor reproducibility (i.e., CV > 20% and/or ICC < 0.75) are shaded in gray. 45RA, CD45RA.

(CVs > 20% and ICCs < 0.75) in several T cell subsets quantified in the T cell activation (T-ACT) and memory/regulatory T cell (T-MEM-REG) panels (Figure 2). Closer examination revealed that populations with high CVs and low ICCs were primarily limited to those defined by CD45RA, with particularly high CVs and low ICCs for subsets defined by both CD45RA and CD62L.

To further explore the source of variation in these T cell populations, we compared data obtained from cryopreserved PBMCs with those obtained from analysis of unfractionated blood obtained from the same blood draw. We first compared the results between blood and PBMCs for populations with good reproducibility within replicate PBMC samples (CV < 20%) and found that they were highly comparable, with no significant differences, except in the case of CD4⁺CD25⁺CD127^{lo} Tregs (Figure 3A), which were significantly reduced in all PBMC samples (*P* = 0.003). However, as this reduction was highly consistent, it did not lead to decreased reproducibility in the inter-center comparison.

We then compared populations with poor reproducibility (i.e., CV > 20%), focusing on those that were scarce (CD4⁺γδ T cells, BDCA3⁺ mDCs, CD4⁺CD28⁻ T cells), defined by poorly resolved gates (monocyte and NK cell subtypes) or by cryopreservation-sensitive markers (CD45RA, CD62L) (14, 21–23) (Figure 3B). Of all of the populations with CVs > 20%, only those defined by the cryopreservation-sensitive

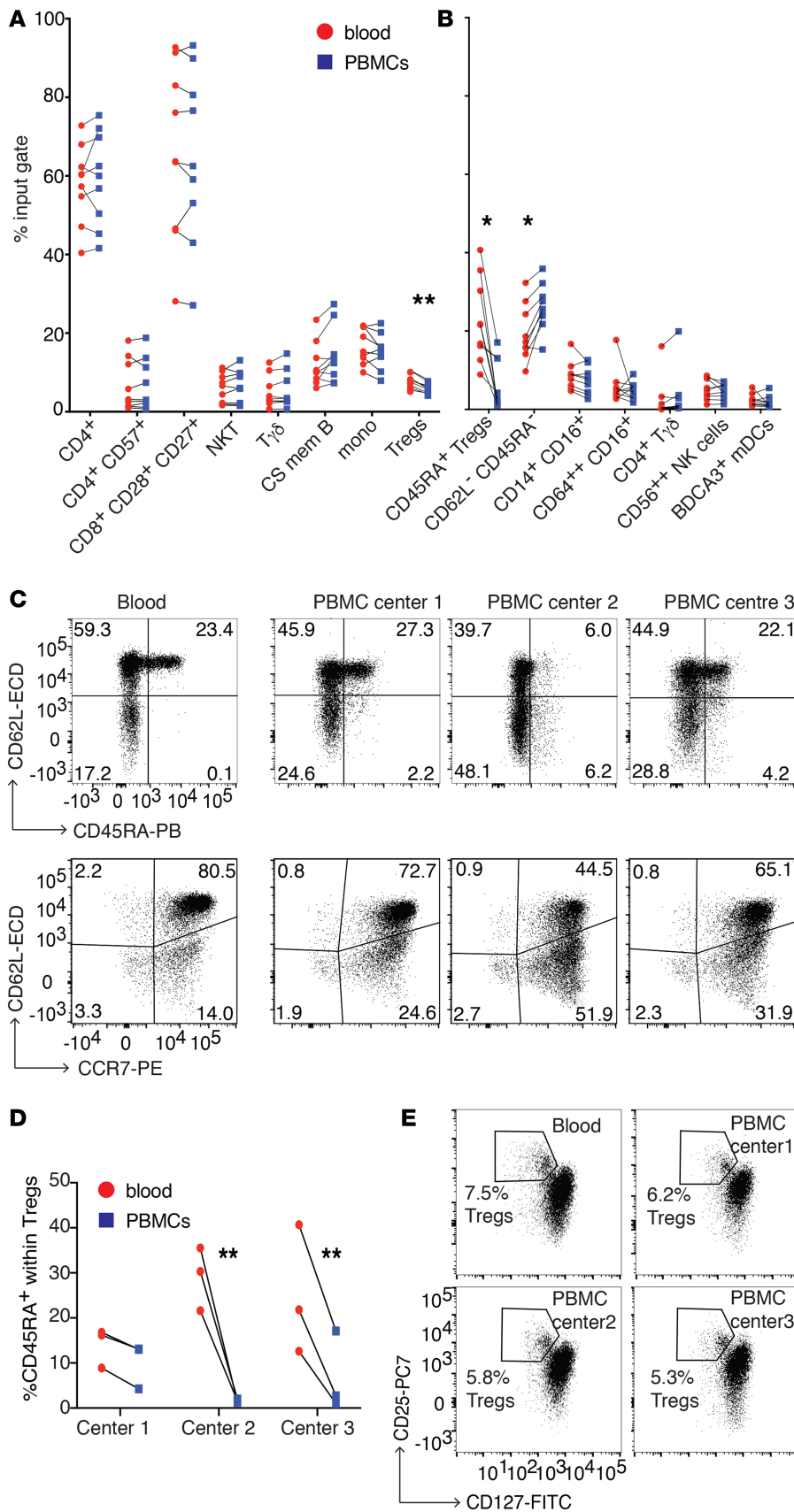


Figure 3. Effects of cryopreservation on standardization. Peripheral blood from healthy subjects was either analyzed immediately or processed into PBMCs and cryopreserved for later analysis with the same DuraClone panels. Raw data were analyzed manually at one center. Data are from 3 subjects analyzed at 3 different sites ($n = 9$); unfractionated blood was only analyzed at the collection site. The identity of the parent gates is shown in Supplemental Table 1. Representative populations as measured in blood versus PBMCs with (A) CV <20% or (B) CV >20% as shown in Figures 1 and 2. * $P < 0.05$, ** $P < 0.01$, multiple t test with FDR adjustment according to Benjamini, Hochberg, and Yekutieli. (C) Representative data from 1 individual for naive/memory CD4⁺ T cell proportions detected in blood or in replicate samples of cryopreserved PBMCs analyzed at 3 different centers (sites 1-3). (D) The proportion of CD45RA⁺ Tregs (of total Tregs) in 3 different individuals was measured in unfractionated blood or PBMCs. ** $P < 0.01$, 2-way repeated-measures ANOVA with Šidák's multiple comparison test. (E) Representative data from Tregs quantified in blood or replicate samples of cryopreserved PBMCs analyzed at 3 different centers. CS mem B, class-switched memory B cells; mono, monocytes.

markers CD45RA and/or CD62L were significantly different when data were obtained from blood versus PBMCs. Cryopreservation did not significantly impact the levels of monocytes, NK cells, CD4⁺ γδ T cells, or BDCA3⁺ mDCs.

To further explore the poor reproducibility of populations expressing CD45RA and/or CD62L, we carried out a more detailed examination of cryopreserved PBMC data from the T-MEM-REG panel. Figure 3C shows a comparison of data obtained from blood versus PBMCs for a single individual's CD4⁺ T cells defined by expression of CD45RA and CD62L, or CCR7 and CD62L. While PBMCs analyzed at centers 1 and 3 retained the expected level of CD45RA expression, this signal was completely lost at center 2. A similar result was found for CD62L. The effect is quantified in Figure 3D using the example of CD45RA⁺ Tregs.

The poor reproducibility in detection of CD45RA- and/or CD62L-expressing cells could be due to variable loss of protein expression and/or loss of the cell type after cryopreservation. As the loss of the total Treg proportions (defined as CD25^{hi}CD127^{lo} cells per CD4⁺ T cells) was not center-dependent (Figure 3E), variation in detection of cryopreservation-sensitive markers may be driven by center-specific cell handling that results in variable loss of these markers.

Optimized timing for analysis of unfractionated blood. With the observation that certain populations (e.g., Tregs) and markers (e.g., CD45RA and CD62L) are strongly affected by cryopreservation, we next carried out experiments to test the range of time over which blood can be reproducibly analyzed. Previous studies determined that analysis within 4–6 hours of blood collection was optimal (12, 24), but this may not always be feasible for multi-center studies electing to ship samples to a central site. We compared data obtained from blood that was stained and acquired within 4 hours with data obtained from blood that was immediately stained but then acquired after 24 hours, or blood that was left unmanipulated for 24 hours and then stained and acquired. The results showed that, for most panels (Basic, TCR, T-ACT, T-MEM-REG), leaving the blood unmanipulated for 24 hours and then staining and acquiring was preferable to immediate staining followed by acquisition 24 hours later, as this protocol resulted in data that were most similar to those from samples stained and acquired within 4 hours. (Supplemental Figure 2A). For the B cell and DC panels, there was no advantage to holding blood for 24 hours before staining versus immediate staining and later analysis. In fact, plasmablasts were undetectable in samples incubated 24 hours before staining, whereas samples that were stained prior to 24-hour incubation and acquisition retained levels similar to those found in their freshly analyzed counterparts (Supplemental Figure 2, B and C).

Application of standardized flow cytometry to post-HSCT patient samples. Studies of standardized flow cytometry often use samples from healthy individuals (12, 14, 19, 24), yet the utility of this method is with clinical samples, often coming from patients with abnormal leukocyte counts and/or proportions. To test the applicability of our standardized protocols on clinical samples, we carried out inter-center comparisons using cryopreserved PBMCs from patients 100 days after HSCT. Figure 4 shows CV and ICC values derived from replicate analysis of PBMCs from 5 patients at 3 different sites. While the CVs tended to be higher (average increase, 4.2%), the overall data were largely comparable to those from healthy control samples (Figures 1 and 2), with the same problematic populations (e.g., monocyte subtypes, rare cells, and/or those defined by cryosensitive markers) having CVs >20%. A notable exception was the B cell panel, which showed poor reproducibility for all subpopulations due to the expected low B cell reconstitution at this time after transplant (25) and thus the low absolute numbers of B cells. Specifically, populations with CVs >20% were those with less than 0.4% of total PBMCs for the healthy subjects and less than 2.5% of total PBMCs for the post-HSCT subjects (Figure 5). A notable exception to this rule was observed with some monocyte subsets in healthy samples: although they were ~1% of PBMCs, their CVs were >20%. The same monocyte subsets had CVs >50% in post-HSCT samples.

We next compared cell population frequencies in healthy versus post-HSCT subjects. Focusing on blood (Figure 6A), significant differences in population proportions were found in the Basic, B cell, and DC panels. Specifically, in comparison to healthy controls, HSCT subjects had lower proportions of lymphocytes, B cells, and BDCA3⁺ mDCs, but higher proportions of monocytes and CD56⁺⁺ NK cells. B cell numbers were often too low to accurately define B cell subpopulations; only 5 of 11 samples met the criterion of >1,000 events in the CD19⁺ gate. Despite the small sample size, we found that HSCT subjects had significantly increased proportions of transitional B cells and plasmablasts, but decreased IgM memory and marginal zone B cells, consistent with the well-documented immature B cell phenotype in this patient population (26). Although T cell populations in blood from healthy controls and 3-months-post-HSCT patients are known to differ in terms of

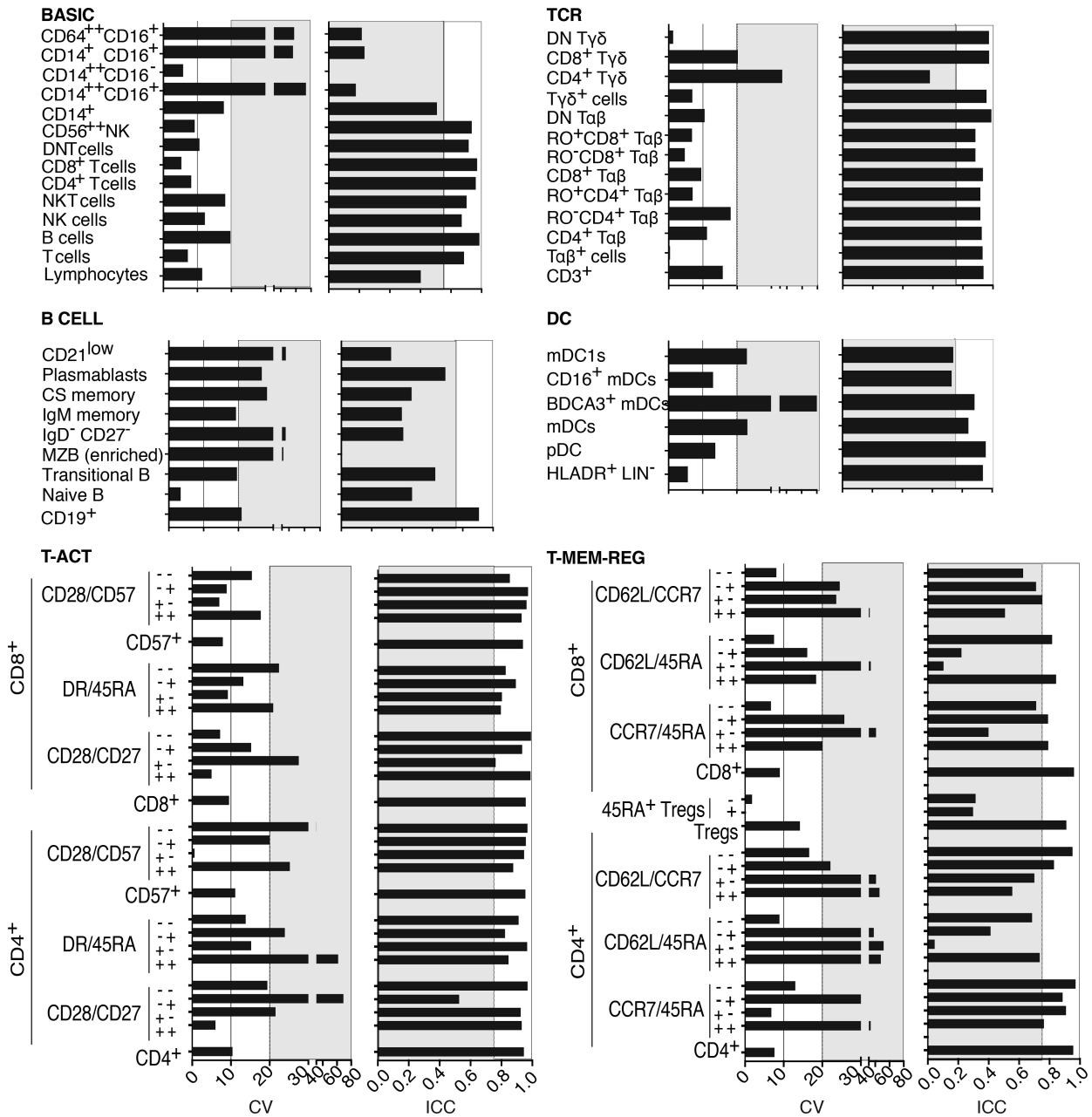


Figure 4. Inter-center comparison of cryopreserved PBMC data from post-HSCT subjects. Replicate aliquots of cryopreserved PBMCs from 5 subjects were analyzed at 3 different sites. Raw data were analyzed centrally, and the reproducibility of population proportions was determined by statistical analysis. All calculations were based on the cell proportion relative to the input gate; the identity of parent gates for all variables is listed in Supplemental Table 1. Shown are CV and ICC values for each population. CV and ICC values indicative of poor reproducibility (i.e., CV >20% and/or ICC <0.75) are shaded in gray. 45RA, CD45RA.

absolute cell counts (25), none of the T cell population proportions examined showed significant differences.

When findings in blood were compared with data obtained from cryopreserved PBMCs, we found that the changes in lymphocytes, CD56⁺ NK cells, and most B cell populations remained consistent (Figure 6B). Exceptions included plasmablasts, which, as described above, did not survive a 24-hour delay in isolation of PBMCs, and a low-abundance BDCA3⁺ DC population. On the other hand, monocyte subsets such as CD14⁺CD16⁺ cells showed significant differences in the PBMCs that were not present in whole blood.

Development of automated pipelines to analyze standardized data. Previous attempts to develop standardized flow cytometry methods found that the most significant source of variation is introduced during manual analysis (14, 19). To minimize this effect, data are often analyzed by a single operator, but this is time-consuming and difficult to implement across multiple independent studies, and results remain subjective.

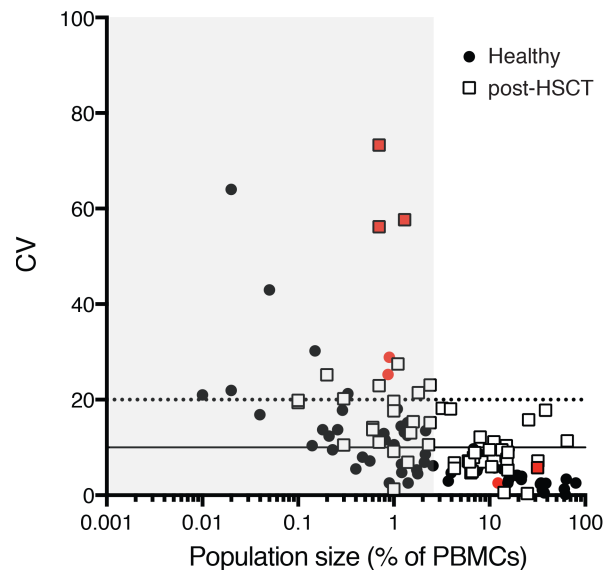


Figure 5. The effect of population size on CVs. CVs were plotted against the log of median population sizes from all inter-site comparisons. All population sizes are expressed as percent of CD45⁺ PBMCs. Solid and dotted lines indicate CVs of 10% and 20%, respectively. Gray box indicates populations that are less than 2.5% of PBMCs. Red symbols indicate monocyte subtypes. Populations shown to be affected by cryopreservation (defined by CD45RA or CD62L) were excluded.

To streamline the workflow, we developed an automated analysis workflow based on flowCore (27) and flowDensity (5). flowDensity was previously shown to outperform unsupervised algorithms in a FlowCAP (Flow Cytometry: Critical Assessment of Population Identification Methods) consortium study, with its performance matched by only one other supervised approach (28). We developed flowDensity automated pipelines customized for the 2 panels currently available as off-the-shelf products (Basic and B cell). Automated analysis aimed to replicate manual gate placement, with pipelines developed using data from the unfractionated blood samples of healthy adults and then further refined on unfractionated blood from the post-HSCT subjects. The proportions of key cell populations were determined as before (Supplemental Figure 3 and Supplemental Table 1).

The performance of automated pipelines was assessed by their ability to match on a per-event basis values obtained by an expert manual analyzer (i.e., the reference manual), currently considered the “gold standard” approach. In addition, we compared the reference manual values with those obtained by 2 additional manual analyzers who followed an identical gating strategy. Data were compared using Spearman’s correlation as well as the F_1 score to measure the performance of automated analysis. F_1 is the harmonic mean of precision and recall, with a score of 1 indicating that all individual events were placed in the same series of gates by both analysis methods. The Basic panel was highly amenable to automated gating, as most cell populations were defined by clear boundaries. Accordingly, manual and automated analyses produced results that were highly correlated with those obtained in the reference manual analysis (Spearman’s rank correlation coefficient [ρ , r_s] >0.8) (Figure 7A). Most median F_1 scores were >0.9 (Figure 7B), with an overall F_1 average of 0.93.

As an example, comparison of automated with corresponding manual reference values for total monocytes revealed a high correlation ($r_s >0.99$) (Figure 7C). For a manual versus manual comparison, an alternative manual (Manual 1) plotted against the reference manual ($r_s >0.99$) is shown in the same graph. The precision and recall of automated versus manual placement of events within the same gates was also excellent for this population, as shown by the F_1 score of 0.99.

We next examined correlations between automated and manual gating for populations with lower F_1 scores. Compared with the other population, nonclassical (CD14⁺CD16⁺) monocytes (29, 30) had particularly low r_s values and F_1 scores: both 0.83, driven by indistinct boundaries between CD14⁺ and CD14⁺⁺ (particularly in the CD16⁺ population), as well as between CD16⁻ and CD16⁺ (Figure 7D). These unclear boundaries led to variability in manual gating (Manual 1 and Manual 2 vs. reference manual, both $r_s = 0.83$), as well in automated versus reference manual ($r_s = 0.83$). For simplicity, only Manual 1 versus reference manual is shown. The effect of unclear boundaries was most apparent for low-abundance monocyte subtypes; the larger CD14⁺⁺CD16⁻ population had higher correlation values and F_1 scores, of 0.90 and 0.97, respectively.

In contrast to monocytes, CD56⁺⁺ NK cells, which are also rare and defined by a poorly resolved marker (CD56⁺⁺), were reproducibly detected by both automated and manual gating ($r_s = 0.98$, $F_1 = 0.88$)

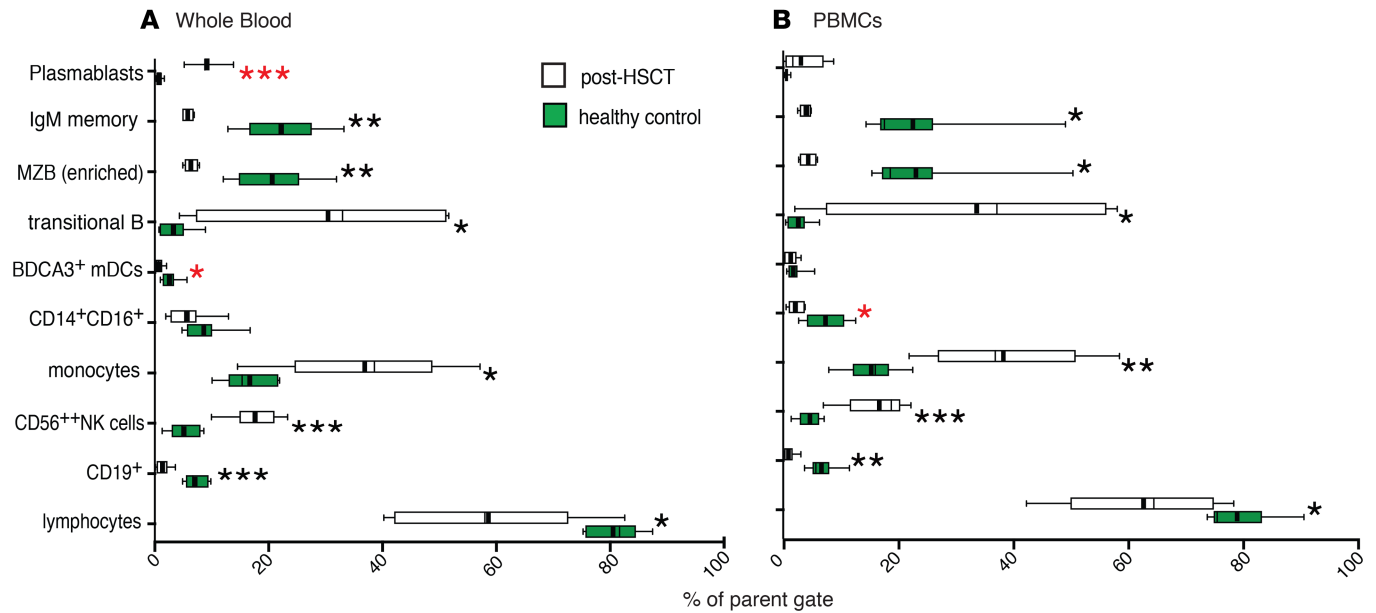


Figure 6. Detecting differences between 2 cohorts using unmanipulated blood versus cryopreserved PBMCs. Populations from all panels that significantly differed between healthy controls and patients 100 ± 20 days after HSCT in either (A) whole blood or (B) cryopreserved PBMCs. Note: Post-HSCT, but not healthy control, PBMCs were incubated 24 hours before cryopreservation. Proportions of parent gates for each population are shown as box-and-whisker plots; midline is the median, box is the interquartile range, and whiskers show minimum and maximum values. Means are indicated by a thick black band. See Supplemental Table 1 for a list of parent gates. Only samples that were evaluable in both unmanipulated blood and PBMCs were compared; $n = 9$ healthy and $n = 8$ post-HSCT subjects (except $n = 4$ for B cell-derived populations). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$; unpaired t test with Holm-Sidak corrections.

(Figure 7E). In this case a second marker (coincidentally also CD16) helped to clearly define the (largely CD16⁻) CD56⁺⁺ population.

We also developed an automated pipeline for the B cell panel, which quantifies several low-abundance populations defined by markers with indistinct boundaries such as CD24, CD27, and CD38 (see the gating strategy in Supplemental Figure 1E). Although the correlation values were typically lower than for the Basic panel, all but one (IgD⁻CD27⁻) had an r_s value >0.8 (Figure 8A), and most mean F_1 scores were >0.8 (Figure 8B), with an overall average of 0.79. Plasmablasts, which are typically $<1\%$ of total B cells and are defined by high expression of multiple markers (CD27, CD24, and CD38), showed one of the lowest correlation values ($r_s = 0.81$) and an exceedingly low F_1 score (0.37). Although automated analysis accurately identified the single sample with abnormally high plasmablast proportions (Figure 8C, red box), subtler differences may be lost. Moreover, there was good correlation between automated and manual gating for class-switched memory cells ($r_s >0.85$), which are defined by the same markers and gates as are the plasmablasts but represent a greater proportion of the population. Similarly, despite the low r_s value for IgD⁻CD27⁻ cells ($r_s = 0.65$), the one sample with high proportions was identified by both manual and automated analysis (Figure 8D, red boxes). The typically more abundant naive B cells (IgD⁺CD27⁻), defined by the same gates, had $r_s >0.9$. Interestingly, the F_1 score for the IgD⁻CD27⁻ population was fairly high (0.81), despite the low correlation score. As Spearman's correlation assesses similarity by rank rather than absolute value, samples that have many similar values (such as the IgD⁻CD27⁻ proportions) may have lower r_s . Nevertheless, the high F_1 score shows that the majority of the cells were being ordered into the same gates in manual and automated analyses.

Evaluation of data from healthy versus HSCT subjects using automated pipelines. Flow cytometry is often used to find biomarkers that discriminate between disease and control groups. We thus next compared the ability of automated versus manual analyses to identify significant differences between healthy and HSCT subjects (Table 2). Automated and manual analyses identified the same populations as being significantly different between the 2 groups with one exception: the marginal zone-enriched B cell population ($P = 0.06$). Half of the P values were less significant after automated than after manual gating — especially in the case of B cells, for which the sample numbers were very low ($n = 5$). Conversely, automated but not manual analysis found that CD14⁺⁺CD16⁺ and CD64⁺⁺CD16⁺ monocytes were significantly increased in HSCT patients. However, for CD14⁺⁺CD16⁺ monocytes, both data sets showed significant differences when tested alone (Supplemental Figure 4; manual $P = 0.0042$, automated $P = 0.0012$).

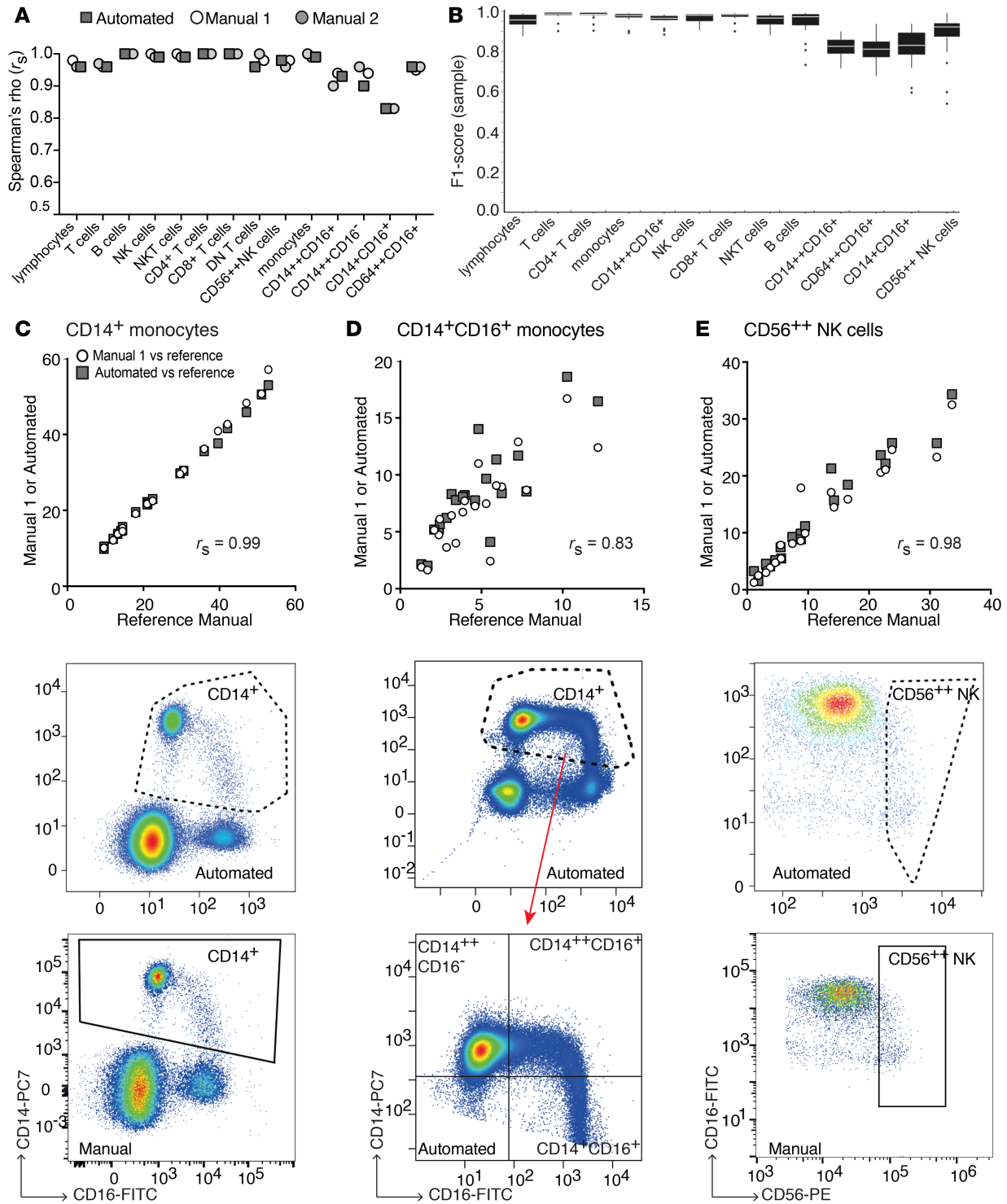


Figure 7. Generation of automated data analysis pipelines for the DuraClone IM Basic panel. The proportions of cell populations indicated in Supplemental Table 1 for the Basic panel were determined in unfractionated blood from healthy ($n = 9$) and HSCT ($n = 11$) subjects using automated gating, 2 different manual analyzers (Manual 1, Manual 2), or a reference manual analyzer. Data from Manual 1 or 2, or automated gating were compared with those from the reference manual to determine (A) Spearman's correlation coefficients (r_s) or (B) F₁ scores (harmonization of precision and recall, maximum value 1.0) ordered by population size (highest on the left). Box-and-whisker plot: midline is the median, box is the IQR, and the whiskers extend to 1.5 times the IQR. (C) Representative automated or manually gated data from 1 healthy subject and correlation graphs for (C) CD14+ total monocytes, (D) CD14+CD16+ monocytes, and (E) CD56++ NK cells. White circles indicate Manual 1, and gray squares automated, both plotted against the reference manual data.

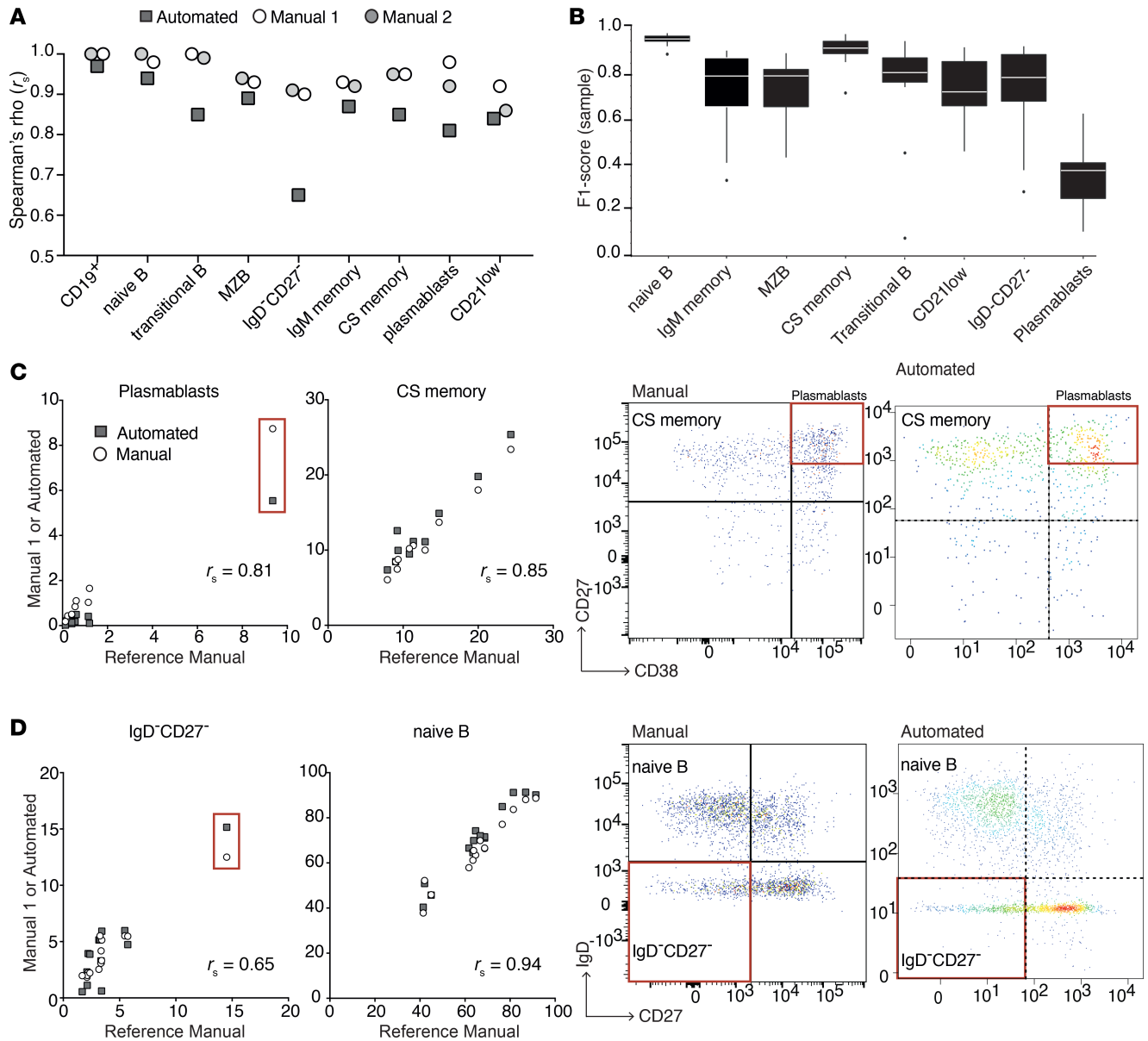


Figure 8. Generation of automated data analysis pipelines for the DuraClone B cell panel. The proportions of cell populations indicated in Supplemental Table 1 for the B cell panel were determined in unfractionated blood from healthy ($n = 9$) and HSCT ($n = 5$) subjects using automated gating, 2 different manual analyzers (Manual 1, Manual 2), or a reference manual analyzer. Data with event counts under threshold values for key gates were not included (see Methods, *Flow cytometry data analyses*). Data from Manual 1 or 2, or automated gating were compared with those from the reference manual to determine (A) Spearman's correlation coefficients (r_s) or (B) F_1 scores. Box-and-whisker plot: midline is the median, box is the IQR, and the whiskers extend to 1.5 times the IQR. (C and D) Correlation graphs and representative manual and automated plots and gates for (C) plasmablasts and class-switched (CS) memory B cells, both derived from the CD27 versus CD38 plot, pre-gated on IgM⁺IgD⁻ B cells; or (D) naive and IgD⁺CD27⁻ B cells. Red boxes indicate data points and manual or automated gates from which they were derived for the samples with outlying proportions of plasmablasts or IgD⁺CD27⁻ cells.

Validation of automated pipelines using an independent data set. Standardized flow cytometry should facilitate direct comparisons of data collected at different centers and/or from different cohorts. To assess whether this was the case for our approach, we obtained an independent data set from the ONE Study, which used the same antibody panel and fluorescence intensity settings (12), and reanalyzed these data manually and with our automated pipelines. Remarkably, we found that the correlations between manual and automated results were similar to those from the data sets used for development of the automated pipelines. When all populations analyzed were combined for the ONE Study or CNTRP data, correlation values between automated and manual gating were all >0.9 (Figure 9; Supplemental Table 2 shows the

Table 2. Cell population differences in healthy versus HSCT subjects as determined by manual or automated gating

	Manual			Automated		
	Healthy	HSCT	Adj. <i>P</i> value	Healthy	HSCT	Adj. <i>P</i> value
Basic panel						
Lymphocytes	80.6	60.6	0.040	82.4	64.0	0.028
B cells	10.2	2.0	0.000	10.0	1.8	0.000
CD56 ⁺ NK cells	5.1	18.2	0.000	5.6	19.3	0.003
CD14 ⁺ monocytes	16.6	34.8	0.030	16.9	33.8	0.026
CD64 ⁺ CD16 ⁺	6.6	13.9	<i>0.174</i>	6.6	13.1	0.018
CD14 ⁺ CD16 ⁺	6.0	10.9	<i>0.663</i>	6.6	13.1	0.018
B cell panel						
Naive B cells	57.9	80.6	0.010	61.9	85.9	0.016
Transitional B cells	3.3	28.7	0.010	3.2	29.3	0.029
MZB (enriched)	20.6	5.7	0.002	16.3	6.3	<i>0.064</i>
IgM memory	22.2	4.9	0.001	20.3	4.5	0.013

The mean proportions of the indicated cell populations were determined using automated or manual gating. Statistically significant differences between healthy ($n = 9$) and HSCT subjects ($n = 11$ for Basic, $n = 5$ for B cells) were determined by multiple t tests and adjusted for multiplicity by the Holm-Sidak method. Nonsignificant differences are in italics. Class-switched memory and plasmablast results are not shown due to low numbers in the post-HSCT samples; see Methods, *Flow cytometry data analyses*, for details.

individual r_s values). Thus, automated gating pipelines developed with one set of data can readily be used to accurately analyze independent data if they are collected using the same standardized methodology.

Analysis of data acquired on different flow cytometers. Ideally, standardized pipelines for data acquisition and analyses should be broadly useful and not restricted to a specific flow cytometry platform. To test the applicability of our SOPs and analysis pipelines to alternate instrument platforms, we analyzed parallel samples acquired on a Navios (Beckman Coulter, 3 lasers) or a Fortessa X20 (BD Biosciences, 4 lasers). Due to the differences in laser configuration and output fluorescence intensity scales, the 2 cytometers were calibrated on the basis of stain indices rather than by standardized fluorescent beads (see Methods for details).

Blood from 5 healthy volunteers was stained using off-the-shelf DuraClone Basic and B cell panels, run on both cytometers and analyzed manually and by automated pipelines. As shown in Figure 10A, manual analysis revealed that the two cytometers detected similar population proportions. An exception was lymphocytes, which tended to be a slightly smaller percentage of mononuclear cells on the Navios cytometer due to a difference in how the 2 platforms acquire forward scatter–low events. After an additional correction for platform-specific singlet gating, the automated pipelines were equally effective at quantifying population proportions acquired on the Fortessa or the Navios (Figure 10, B and C).

Discussion

Flow cytometric analysis of peripheral blood has the potential to diagnose, stratify, and monitor patients. Very few flow cytometric biomarkers (with notable exceptions related to HIV and hematological malignancies) have been validated for clinical implementation due to the logistical complexity of flow cytometry and the variation inherent in cryopreserved PBMCs. Here we report a unified workflow for standardized flow cytometry for analysis of cryopreserved PBMCs or whole blood coupled to automated analysis pipelines. Testing of the automated pipelines on an independent data set revealed the power of standardization and enabled direct comparison of data from different studies and/or centers, collected over different time intervals.

Our approach builds upon several previous efforts to develop standardized flow cytometry methods for blood leukocyte characterization. In an approach similar to that used by the HIPC and others (14, 31, 32), we utilized pre-formatted, dried-down antibody cocktails to eliminate variation due to pipetting of antibody cocktails and batch-to-batch reagent variation. Similar to Streitz et al. (12), we used a clinical instrument platform to minimize variation inherent in research-grade instruments. We extended both of these approaches by including a direct comparison of data generated from cryopreserved PBMCs versus whole blood, and showed the feasibility of using the standardized platform to analyze blood from a patient population with abnormal proportions of several leukocyte populations.

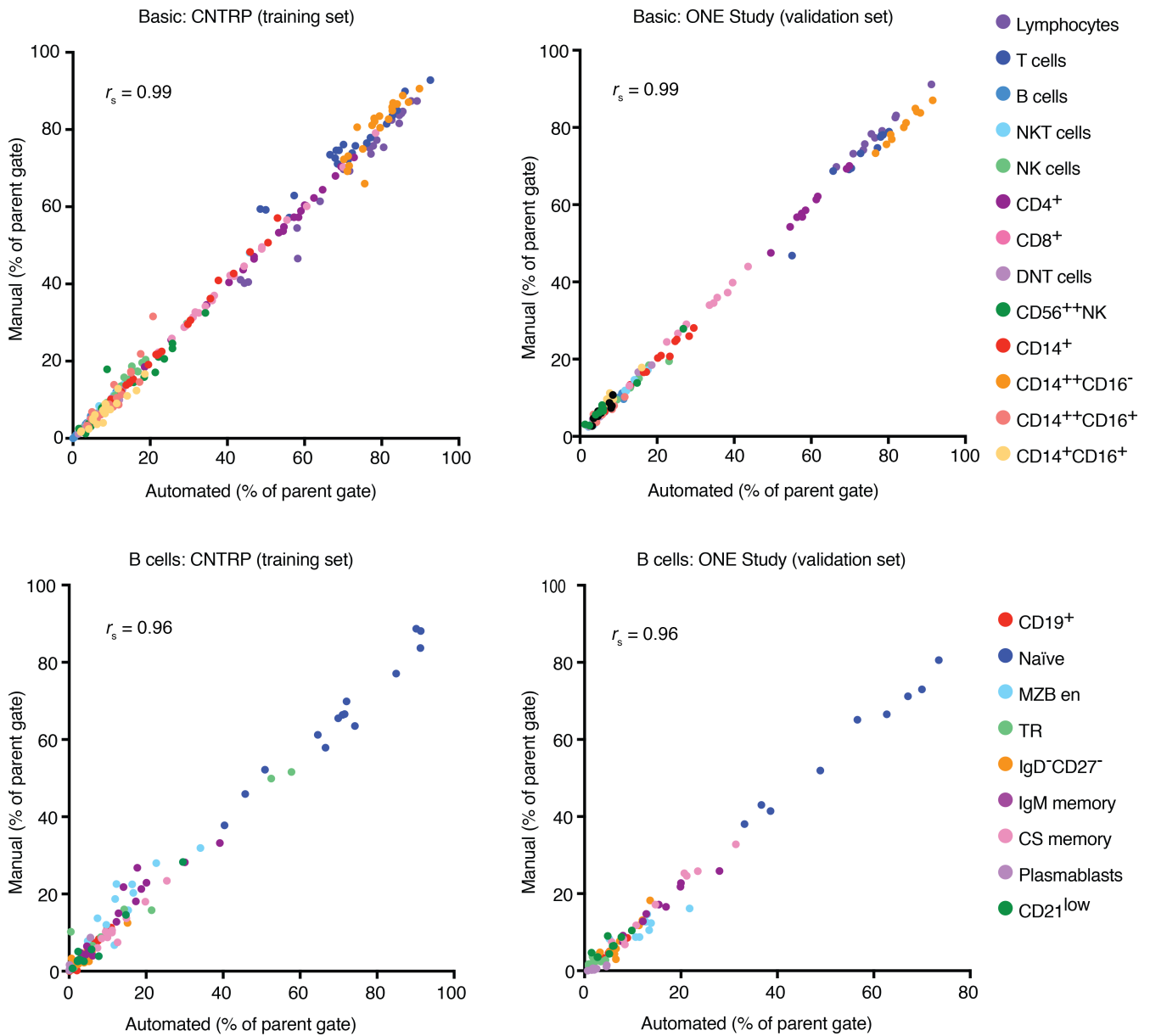


Figure 9. Validation of automated gating pipelines in an independent dataset. Blood from healthy subjects enrolled in the ONE ($n = 9$) or CNTRP ($n = 20$ for Basic, $n = 10$ – 14 for B cells) study was analyzed with the Basic or B cell panels, and data were analyzed manually or using automated pipelines developed using the CNTRP data. All population proportions listed in Supplemental Table 1 for the Basic and B cell panels were measured, and results from automated versus manual analyses were compared. Shown are correlation graphs merging all data from either the CNTRP (training set) or ONE (validation set) study quantified by automated versus manual gating in relation to input gate. Each population proportion is a different color, and each data point represents an individual. Spearman's correlation coefficient is shown at the upper left. See Supplemental Table 2 for individual r_s values.

Clinical utilization of flow cytometry–based biomarkers is likely to necessitate their rapid quantification in whole blood. Although previous studies have shown that use of whole blood facilitates flow cytometry standardization (12, 33), the logistical complexity and expense of real-time analysis is prohibitive for the large sample sizes required for biomarker discovery and validation. Moreover, use of cryopreserved PBMCs enables batching and retrospective analyses of samples after clinical endpoints have been identified and as new markers of interest emerge. For the majority of common leukocyte populations, analyses in whole blood or cryopreserved PBMCs gave equivalent results. Examples of populations that were less reliably measured in PBMCs include T cell populations defined by CD45RA and/or CD62L. Direct comparison of data obtained from matched samples of blood versus cryopreserved PBMCs showed loss of these markers in a center-specific manner. This finding is in accordance with previous reports showing that CD62L

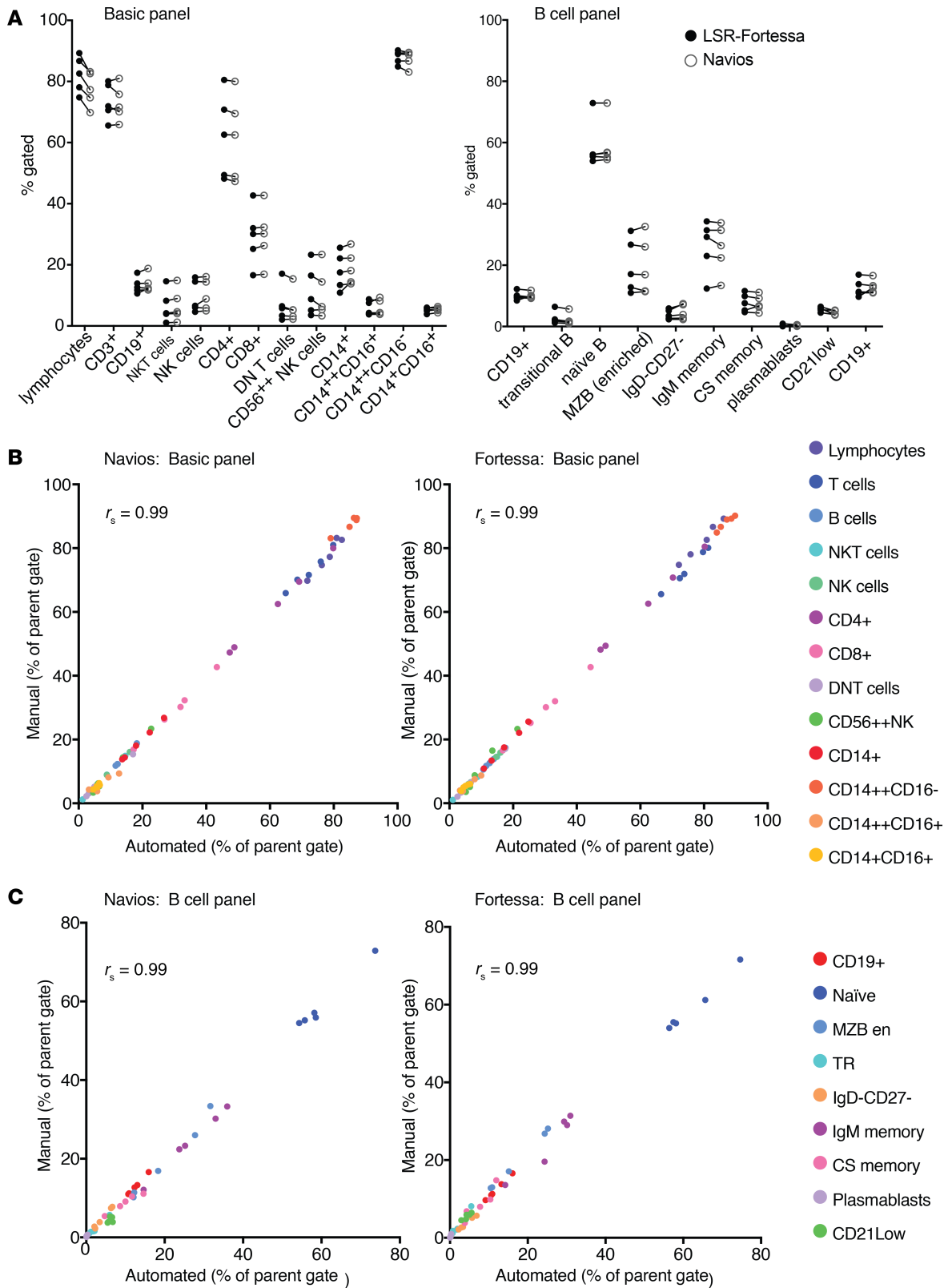


Figure 10. Application of approach to alternate flow cytometer platform. Blood from 5 healthy subjects was stained using off-the-shelf Basic and B cell DuraClone tubes and acquired on a Navios (3 lasers) or an LSR-Fortessa (4 lasers) cytometer. LMD or FSC3.0 files were analyzed manually or with automated pipelines. (A) Comparison of data from cytometers. Each variable from 5 donors is shown as an individual data point, with the LSR-Fortessa data linked to the same data point obtained on the Navios. (B and C) Comparison of results from manual or automated analysis on the indicated cytometer. Each population proportion is a different color, and each data point is an individual. Spearman's correlation coefficient is shown at the upper left.

is unstable in cryopreserved cells (22, 23), and that T cell populations defined by CCR7⁺ and CD45RA⁺ are subject to center-specific variability in analyses of cryopreserved PBMCs (14). Some variation may be mitigated by intense on-site training (4, 34) or by selection of alternative antibody clones (e.g., use of an alternate anti-CD45RA clone [L48] to define naive T cells resulted in CVs <10%) (14). Technical factors related to cell processing and cryopreservation might not be the only factors leading to poor reproducibility of these populations, as some CD45RA-expressing populations in unfractionated blood also showed high CVs (>30%). These results are in line with previous studies, particularly in combination with CCR7 (11, 12, 14). Overall, as part of the development of PBMC-based studies, it is important to compare results with blood to avoid studying populations with poor reproducibility in cryopreserved samples.

Other populations that had high CVs (between 20% and 30%) were either low in abundance (<0.5% of freshly frozen PBMCs) and/or defined by heterogeneously expressed markers such as CD14, CD16, and CD56. Low-abundance populations were more subject to variability, as small shifts in gating boundaries led to large differences in gated proportions and higher operator-dependent variability during analysis. This limitation was most evident in monocyte subpopulations, which have previously been shown to be subject to high technical variability in both whole blood (12) and cryopreserved PBMCs (14). These populations exhibited higher CVs than other populations of similarly low abundance, especially in the post-HSCT samples. Solutions to these issues should be sought, since monocyte subpopulations may, for example, predict response to cancer immunotherapy (35). In some cases, standardization may be improved by including additional markers or by refining gating strategies; for instance, the exclusion of HLA-DR^{neg} cells allows improved identification of monocyte subtypes based on CD14 and CD16 expression (36). In addition, the selection of optimal clones and fluorochromes for each marker can also improve population resolution (4, 24). Human error can be further reduced by the use of robotics for cell handling (24) and automated gating (14).

An important question is how good is “good enough” for standardized assessment of a given flow cytometric parameter? In clinical cytometry labs, CV targets tend to be <10% (37), but since these are skewed by the population proportion, CVs of up to 25% may be acceptable for flow cytometric determination of low-abundance populations (37–39). Clinically meaningful CVs will depend on the underlying biological variability (40), and the ICC may be a better indicator of the degree to which biological variation is masked by technical variation. It is also important to note that lower reproducibility can be tolerated in biomarker development and validation phases than in the final clinical assay. For comparisons between groups, ICCs of 0.6–0.8 are adequate (41), but when making patient-specific decisions, ICCs should be >0.9 (42). Overall, for clinical application, some flow cytometry-based biomarkers may need to undergo further protocol optimization to limit technical variation.

Well-defined populations were effectively gated by automated pipelines, with correlation values >0.9, even when the data were collected in an independent study or on a different flow cytometer. Moreover, the average of F_1 scores for the Basic panel (in which the majority of populations were defined by non-diffuse gates) was 0.93, substantially higher than those obtained by a recent review of state-of-the-art automated unsupervised gating (43). Importantly, automated pipelines could be applied to samples acquired with a different set of fluorochrome target intensities, and across different patient populations and centers. They could also be applied to data acquired on a cytometer with a different laser configuration without adjusting the thresholds that define population proportions. The use of beads spectrally matched to the panel fluorophores will help to further harmonize platforms in the future (44). It should be noted that pipeline performance may vary on some datasets, especially if they are derived from patient samples whose leukocyte populations are different from normal. Furthermore, our pipelines were developed using unfractionated blood and would need to be adapted for use on PBMC-based datasets.

In addition to equal or higher reproducibility, automated gating enables automation of the entire flow cytometry data analysis process, from analysis to quality control and statistical analysis to report generation (45). When comparing manual and automated gating, the current paradigm is to use an expert manual analyzer as a “gold standard” reference; however, it should be noted that there is no evidence that manual gating is actually more accurate than automated. In general, we found that automated gating agreed slightly less with the reference manual than did other manual gating, especially for low-abundance, poorly defined populations (e.g., plasmablasts), as defined in our study. Balanced against this are the advantages of rapid analyses (~1 minute of computer time per set of 25 samples from raw data to final spreadsheet versus 10–20 expert hours for the equivalent manual analyses) and the minimization of errors that can be introduced during manual manipulation of large data quantities.

Data obtained by manual and automated analyses identified almost identical populations as significantly different between healthy controls and post-transplant patients. In the case of monocyte subtypes, automated but not manual analyses showed a significant increase in some CD16⁺ monocyte subtypes (CD14⁺⁺ and CD64⁺⁺), a finding supported by at least one recent publication (46). However, results obtained for these difficult-to-gate populations should be interpreted with caution and better identification strategies sought for future studies. These populations also show increased variability in the manual inter-site comparisons, illustrating that effective standardization depends on non-ambiguous gating strategies regardless of whether automated or manual analyses are used.

Effective standardization enables data from different sites and studies to be merged, facilitating the study of heterogeneous patient populations. The 3 key aspects of our standardized platform are the use of commercially available, pre-formatted antibody cocktails; identically calibrated and clinically certified flow cytometers; and automated analysis pipelines customized for commercially available reagents. All 3 of these features are easy and rapid to implement, and, for the majority of common immune cell populations, they eliminate the need for extensive SOP training and proficiency testing. Moreover, our approach allowed for direct comparison of data from different sites, regardless of whether it was generated from analysis of blood or PBMCs. Widespread implementation of these or similarly standardized acquisition and analysis protocols is required for the discovery and validation of flow cytometry-based biomarkers and to further the establishment of precision medicine for immune-related diseases.

Methods

Blood samples. Blood from 3 healthy adults was collected at each of 3 centers (BC Children's Hospital Research Institute, Alberta Transplant Institute, and Hôpital Maisonneuve-Rosemont) ($n = 9$). Blood from 11 HSCT patients approximately 100 days (± 20 days) after transplant was collected at Vancouver General Hospital ($n = 11$). Samples were collected with the approval of the research ethics board at each site, and written informed consent was received from each donor (see *Study approval* below for details). Blood was either collected in sodium-heparin Vacutainer tubes (BD) for PBMC isolation or sodium-EDTA Vacutainer tubes for immediate staining of unfractionated blood.

Design of inter-center experiments. For cross-center comparisons using blood from healthy subjects, blood was collected, processed into PBMCs, and cryopreserved in aliquots of 10^6 cells at the 3 research centers listed above (3 blood draws per site, 9 samples in total). Unfractionated blood from the same collection was stained and acquired within 4 hours to enable direct blood versus cryopreserved PBMC comparisons. Two additional centers stained PBMCs only (Toronto General Research Institute and CancerCare Manitoba). Frozen aliquots of PBMCs were shipped on dry ice between all participating centers, so that each had at least 1 aliquot from the same 5 blood draws.

For the cross-center comparisons using blood from post-HSCT patients, 1 ml unfractionated blood was stained at one site (Vancouver) within 4 hours of collection. The remainder (~ 45 ml) was incubated for 24 hours at room temperature (RT) to emulate real-world conditions in which samples are shipped to a central site for processing. PBMCs were isolated 1 day after collection, and aliquots of 10^7 per vial were frozen as described above. Replicate aliquots from 5 different patients were shipped on dry ice from Vancouver to Edmonton and Montreal to enable a 3-center comparison of post-HSCT PBMCs.

Staining of unfractionated blood using DuraClone dry antibody cocktail tubes. All flow cytometry reagents were obtained from Beckman Coulter unless otherwise stated. Six ONE Study DuraClone panels were used; see Table 1 for a list of markers in each panel and Supplemental Table 3 for a list of clones. 100 μ l anticoagulated blood was added into each of the following tubes: Basic, TCR, T-ACT, T-MEM-REG, and DC tubes. For the B cell tube, 300 μ l blood was first washed twice with PBS to remove plasma (and the soluble IgM therein); after resuspending the washed cells in a total volume of 300 μ l PBS, 100 μ l was added to the B cell DuraClone tube. Liquid antibodies were added to the DC tubes as drop-ins: BDCA3-FITC and BDCA2-APC (Miltenyi; catalog 130-090-513 and 130-090-905, respectively). Tubes were vortexed for 10 seconds to resuspend dried antibodies with the cell sample and incubated for 15 minutes at RT in the dark. Two milliliters VersaLyse with 2.5% IOTest 3 Fixative was added to each tube, immediately vortexed for 10 seconds, and incubated for a further 15 minutes in the dark at RT. Cells were washed in cold IFN buffer (IsoFlow + 2% heat-inactivated FBS [NorthBio Inc.] + 0.1% Na₂S₂O₃) and centrifuged for 5 minutes at 300 g and 4°C. This wash step was repeated, and cells were resuspended in 300 μ l cold IFN buffer, transferred to 1.2 ml FACs tube inserts (VWR International), and stored at 4°C until acquired (within 12 hours).

Isolation, cryopreservation, and thawing of PBMCs. See the supplemental material for detailed SOPs. Briefly, all reagents and centrifuges were used at RT until the final cryopreservation step. Blood was diluted with 1 volume of PBS and mixed by inversion, and 25 ml was layered over 15 ml Lymphoprep in a SepMate tube (STEMCELL Technologies) following the manufacturer's instructions. PBMCs were counted by trypan blue staining or with an automated counter, Cellometer Auto 2000 (Nexcelom Bioscience), after staining with acridine orange/propidium iodide (AO/PI) (catalog CS2-0106) according to the manufacturer's instructions using the setting "Immune Cells, low RBCs."

After counting, cell pellets were resuspended in freezing medium (10% DMSO in FBS) at final concentration of 10×10^6 cells/ml at RT. Frozen aliquots of PBMCs were shipped to other centers on dry ice. Cryopreserved PBMCs were thawed at 37°C for 1 minute and transferred to 10 ml thawing solution (1.5 ml FBS, 250 μ l 1 M HEPES, 50 μ l 7.5 % sodium bicarbonate in 10 ml RPMI; all reagents were from Life Technologies) and centrifuged for 10 minutes at RT at 453 g. After resuspension in thawing solution, cells were counted and resuspended in PBS + 2% FBS to a final concentration of 10^7 cells/ml. The average viability of thawed PBMCs from healthy and post-HSCT subjects was 94% (range 85%–100%) and 80% (range 52%–94%), respectively.

Staining of PBMCs using DuraClone dry antibody cocktail tubes. See supplemental material for detailed SOPs. Briefly, 10^6 PBMCs in 100 μ l of PBS + 2%FBS were added to each DuraClone tube; the drop-in antibodies for BDCA-2 and -3 were added as described above for unfractionated blood staining. Tubes were vortexed for 10 seconds and incubated for 15 minutes in the dark. To remove unbound antibodies prior to fixing, 3 ml FP (2% FBS in PBS) buffer was added to each tube, cells were spun down at 453 g for 5 minutes, and the supernatant was discarded. Pellets were resuspended in 1 ml of 1 \times IOTest 3 Fixative solution for 15 minutes in the dark. Cells were then washed with 3 ml cold IFN buffer and centrifuged, and pellets were resuspended in 300 μ l cold IFN buffer until acquisition.

Instrument standardization (Navios). All flow cytometry data for the multi-site studies were acquired on 10-color/3-laser Navios flow cytometers (Beckman Coulter). Flow-Check Pro beads were run for daily quality control, and the manufacturer's criteria were followed to assess whether the instrument was in good working order. To produce consistent fluorescence outputs between the CNTRP Navios instruments and those used in the ONE Study, Flow-Set Pro beads were used to establish the voltages required for each detector on each Navios. For daily quality control, Flow-Set Pro beads were run daily at these voltages, and the instrument was considered to yield stable fluorescence outputs if the MFIs of the beads were $\pm 10\%$ of the initially established MFIs. Voltage recalibration to lot-specific FSP standard intensities was either done using the Navios-intrinsic Autoseup algorithm or manually (for phase I, healthy control samples, Autoseup was used; for phase II, post-HSCT patient samples, the manual method was used). Both procedures are described in detail in the supplemental material (SOP sections 9 and 11). Autoseup was also used to calculate a 9-color compensation matrix using VersaComp antibody capture beads stained with single CD4 antibodies conjugated to each of the 9 fluorochromes. The same initial compensation matrix was used for each of the 6 panels due to the standardized nature of the fluorochromes and instrument settings. Technical staff at each center were trained to perform daily hardware quality control, sample preparation of cells with DuraClone tubes, and data acquisition, based on a written SOP and onsite training (see supplemental material). Effectiveness of training was evaluated by acquisition of a normal blood sample and comparison of data with the central site.

Comparative analyses on the Navios and the LSR-Fortessa. Some second-generation DuraClone panels incorporated a 10th parameter (FL4/PECy5.5), so this study optimized fluorescence intensities for 10 colors. Due to differences in dynamic range and emission filters, the 2 cytometers could not be calibrated by absolute fluorescence intensities, so stain index-based calibration was used. Briefly, PBMCs stained with single anti-CD4 mAbs conjugated to one of 10 fluorochromes were acquired on the Navios, and stain indices were determined (47, 48). The same samples were then run at 30-V increments on the Fortessa from 230 to 800 V. Stain indices for each fluorochrome at each voltage were determined, and the voltage resulting in the stain index most similar to that on Navios was selected. Blood from 5 healthy volunteers was stained using the DuraClone Basic and B cell tubes (Beckman Coulter, catalog B53309 and B53318, respectively), divided into 2 and run on the same day on the Fortessa and Navios.

Flow cytometry data analyses. For inter-center comparisons, raw data (LMD files) were sent to a central site for analysis. Compensation for each panel was created from the center-specific acquisition matrix and adjusted for individual panels at the time of central analysis. Samples were gated as previously

described (12); an example of gating for each panel is shown in Supplemental Figure 1. All values are provided as a proportion of parent gate, and a list of parent gates for each variable is shown in Supplemental Table 1. Flow cytometry data were manually analyzed using with FlowJo software v. 10.2. or Kaluza Analysis Software v.1.5a (Beckman Coulter).

Threshold event numbers for flow cytometry analysis: For B cells, samples with >1000 events in the CD19⁺ gate were evaluated for subpopulation proportions; samples with >100 events in the IgD⁻IgM⁻ gate were further subdivided into class-switched memory B cells and plasmablasts. Only one sample, just over the cutoff for total B cells by manual gating but under the cutoff for automated gating, was removed to have an equal number of samples for comparison. In accordance with MIFlowCyt (49), all LMD and FCS3.0 files were uploaded to FlowRepository, experiment ID FR-FCM-ZYQT.

Development and evaluation of automated gating pipelines. The flowDensity algorithm uses a supervised, sequential bivariate clustering approach to generate a set of predefined cell populations using customized cut-offs defined by density distributions for each marker. FlowDensity is freely available from Bioconductor (<https://www.bioconductor.org>). The customized code used in this study can be downloaded from GitHub (<https://github.com/mehrnoushmalek/DuraClone-gating>; commit a8f440) for academic use; see license file for full details). Parameters for flowDensity were set globally, customized at every step of the gating hierarchy to replicate the manual approach. Three manual operators independently analyzed all data using the gating strategies outlined in Supplemental Figure 1; one operator was designated as the “gold standard” reference manual and used for the comparison to automated gating. Data similarity was assessed by correlation or F_1 scores (see *Statistics*).

For the analysis of ONE Study data with the automated pipelines, raw data from 10 healthy subjects and manual analyses thereof were provided by the ONE Study group. Two CNTRP operators reanalyzed the raw data. Automated versus manual as well as manual versus manual correlations were done as described above. One sample (HC43) was excluded due to a high level of unexplained background staining.

For analyses of data acquired on the LSR-Fortessa, the pre-processing component of the pipeline was adjusted due to differences in the amount of forward scatter–low event collection, and the singlet gating was changed from FS-A versus FS-W (Navios method) to FS-A versus FS-H (Fortessa method). For the Basic panel, which expresses some populations as a percentage of total PBMCs, exported FCS files containing only the CD45⁺ subpopulations were used for the automated analysis.

Statistics. For inter-center variation, CV estimates were obtained by taking the median of the subject-specific CVs. ICCs were obtained by first fitting a variance components model with random effects for site, donor, and residual error. The ICC was calculated by the ratio of biological variability (individual effect) to the sum of the 2 technical variability components (center, residual error). Paired blood and PBMC data were compared using a multiple t test with FDR adjustment by the method of Benjamini, Hochberg, and Yekutieli. Comparisons of staining methods were based on the absolute magnitude of differences (log scale) between the 24-hour delay methods and the optimal method (immediate staining and acquisition) by applying the Wilcoxon’s rank-sum test to cell population values for each panel as the unit of analysis.

Data from healthy or 100-days-post-HSCT subjects were used to compare the results obtained with unfractionated blood versus cryopreserved PBMCs using multiple t tests with Holm-Šidák correction for multiplicity. A P value less than 0.05 was considered significant. To compare data generated by manual or automated analyses, Spearman’s ρ (degree of linear correlation with the reference manual) and F_1 scores were calculated. The F_1 score (harmonic mean of precision and recall or sensitivity) provides a value in the range of 0 to 1 for each population, with 1 indicating a perfect reproduction of a manually gated population by automated gating. Analyses were performed using GraphPad Prism for MacOSX version 7.0 b and R version 3.3.3 (50). For all tests, $*P < 0.05$, $**P < 0.01$, and $***P < 0.001$.

Study approval. The studies presented in this article involved the collection of human tissue and were reviewed and approved by the appropriate institutional review boards: Montreal, Hôpital Maisonneuve-Rosemont — Centre intégré universitaire de santé et de services sociaux de l’Est-de-l’Île-de-Montréal (CIUSSS-EMTL), Comité d’éthique de la recherche. Vancouver: University of British Columbia/Children’s and Women’s Health Centre of British Columbia (UBC/C&W) Research Ethics Board. Edmonton: Research Ethics Office. All subjects provided written informed consent prior to their participation in the study.

Author contributions

SI, MKL, RRB, and JSD designed the research study. SI, RVG, SW, AH, MR, MG, QG, PA, JW, AR, and SU conducted experiments and acquired data. SI, RRB, MM, and RVG analyzed data. RFB provided statistical support. SI and MKL wrote the manuscript, with critical review by MS, BS, SS, and JJB. JSD and RB recruited HSCT patients. LJW, MKL, JSD, RRB, DK, and DAW secured funding.

Acknowledgments

We gratefully acknowledge the valuable contribution of Salima Janmohamed-Anastasakis (employee of Beckman Coulter), whose technical expertise and help with study design, site training, and manual data analyses were integral to the success of this project. We thank Kaija Strautins for support in figure formatting. This work was supported by the Canadian Institutes of Health Research (CIHR) through the Canadian National Transplant Research Program (TFU 127880 to MKL, DAW, JSD, and LJW), the National Institute of Allergy and Infectious Diseases of the NIH (NIH 1R01GM118417-01A1), and the Natural Sciences and Engineering Research Council (NSERC) (MM, RRB). MKL receives a Scientist Salary Award from the BC Children's Hospital Research Institute. LJW holds a Tier 1 Canadian Research Chair and is supported by the Gordon English family through the Stollery Children's Hospital Foundation and Women and Children's Health Research. JSD is supported by a career award from the Fonds de recherche du Québec-Santé (FRQS).

Address correspondence to: Megan K. Levings, A4-186, 950 West 28th Avenue, Vancouver, British Columbia, Canada, V5Z 4H4. Phone: 604.875.2000 ext 4686; Email: mlevings@bccchr.ca.

- Schreiber K, Nocturne G, Cornec D, Daïen CI. Lymphocytes as biomarkers of therapeutic response in rheumatic autoimmune diseases, is it a realistic goal? *Clin Rev Allergy Immunol*. 2017;53(2):277–290.
- Béné MC, Le Bris Y, Robillard N, Wuillème S, Fouassier M, Eveillard M. Flow cytometry in hematological nonmalignant disorders. *Int J Lab Hematol*. 2016;38(1):5–16.
- Wieland E, Shipkova M. Lymphocyte surface molecules as immune activation biomarkers. *Clin Biochem*. 2016;49(4-5):347–354.
- Maecker HT, et al. A model for harmonizing flow cytometry in clinical trials. *Nat Immunol*. 2010;11(11):975–978.
- Malek M, Taghiyar MJ, Chong L, Finak G, Gottardo R, Brinkman RR. flowDensity: reproducing manual gating of flow cytometry data by automated density-based cell population identification. *Bioinformatics*. 2015;31(4):606–607.
- Kalina T, et al. EuroFlow standardization of flow cytometer instrument settings and immunophenotyping protocols. *Leukemia*. 2012;26(9):1986–2010.
- van Dongen JJ, et al. EuroFlow antibody panels for standardized n-dimensional flow cytometric immunophenotyping of normal, reactive and malignant leukocytes. *Leukemia*. 2012;26(9):1908–1975.
- Theunissen P, et al. Standardized flow cytometry for highly sensitive MRD measurements in B-cell acute lymphoblastic leukemia. *Blood*. 2017;129(3):347–357.
- Rawstron AC, et al. A complementary role of multiparameter flow cytometry and high-throughput sequencing for minimal residual disease detection in chronic lymphocytic leukemia: an European Research Initiative on CLL study. *Leukemia*. 2016;30(4):929–936.
- Hedley BD, Keeney M, Popma J, Chin-Yee I. Novel lymphocyte screening tube using dried monoclonal antibody reagents. *Cytometry B Clin Cytom*. 2015;88(6):361–370.
- Maecker HT, McCoy JP, Nussenblatt R. Standardizing immunophenotyping for the Human Immunology Project. *Nat Rev Immunol*. 2012;12(3):191–200.
- Streitz M, et al. Standardization of whole blood immune phenotype monitoring for clinical trials: panels and methods from the ONE study. *Transplant Res*. 2013;2(1):17.
- Kverneland AH, et al. Age and gender leucocytes variances and references values generated using the standardized ONE-Study protocol. *Cytometry A*. 2016;89(6):543–564.
- Finak G, et al. Standardizing flow cytometry immunophenotyping analysis from the Human ImmunoPhenotyping Consortium. *Sci Rep*. 2016;6:20686.
- Schlickeiser S, Streitz M, Sawitzki B. Standardized multi-color flow cytometry and computational biomarker discovery. *Methods Mol Biol*. 2016;1371:225–238.
- Hébert MJ, Hartell D, West L. Transdisciplinary tour-de-force: The Canadian National Transplant Research Program. *Transplantation*. 2016;100(3):466–470.
- Disis ML, et al. Maximizing the retention of antigen specific lymphocyte function after cryopreservation. *J Immunol Methods*. 2006;308(1-2):13–18.
- Ramos TV, Mathew AJ, Thompson ML, Ehrhardt RO. Standardized cryopreservation of human primary cells. *Curr Protoc Cell Biol*. 2014;64:A.3I.1–A.3I.8.
- Maecker HT, et al. Standardization of cytokine flow cytometry assays. *BMC Immunol*. 2005;6:13.
- Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420–428.
- Mahnke YD, Brodie TM, Sallusto F, Roederer M, Lugli E. The who's who of T-cell differentiation: human memory T-cell subsets. *Eur J Immunol*. 2013;43(11):2797–2809.

22. Weinberg A, et al. Optimization and limitations of use of cryopreserved peripheral blood mononuclear cells for functional and phenotypic T-cell characterization. *Clin Vaccine Immunol.* 2009;16(8):1176–1186.
23. Costantini A, et al. Effects of cryopreservation on lymphocyte immunophenotype and function. *J Immunol Methods.* 2003;278(1-2):145–155.
24. Hasan M, et al. Semi-automated and standardized cytometric procedures for multi-panel and multi-parametric whole blood immunophenotyping. *Clin Immunol.* 2015;157(2):261–276.
25. Bosch M, Khan FM, Storek J. Immune reconstitution after hematopoietic cell transplantation. *Curr Opin Hematol.* 2012;19(4):324–335.
26. Bemark M, Holmqvist J, Abrahamsson J, Mellgren K. Translational Mini-Review Series on B cell subsets in disease. Reconstitution after haematopoietic stem cell transplantation — revelation of B cell developmental pathways and lineage phenotypes. *Clin Exp Immunol.* 2012;167(1):15–25.
27. Hahne F, et al. flowCore: a Bioconductor package for high throughput flow cytometry. *BMC Bioinformatics.* 2009;10:106.
28. Aghaepour N, et al. A benchmark for evaluation of algorithms for identification of cellular correlates of clinical outcomes. *Cytometry A.* 2016;89(1):16–21.
29. Wong KL, Yeap WH, Tai JJ, Ong SM, Dang TM, Wong SC. The three human monocyte subsets: implications for health and disease. *Immunol Res.* 2012;53(1-3):41–57.
30. Ziegler-Heitbrock L. Blood monocytes and their subsets: established features and open questions. *Front Immunol.* 2015;6:423.
31. Pitoiset F, et al. A standardized flow cytometry procedure for the monitoring of regulatory T cells in clinical trials. *Cytometry B Clin Cytom.* 2018;94(5):621–626.
32. Correia RP, et al. A ten-color tube with dried antibody reagents for the screening of hematological malignancies. *Int J Lab Hematol.* 2018;40(2):136–143.
33. Chen X, et al. Automated flow cytometric analysis across large numbers of samples and cell types. *Clin Immunol.* 2015;157(2):249–260.
34. Dyer WB, et al. Substantial improvements in performance indicators achieved in a peripheral blood mononuclear cell cryopreservation quality assurance program using single donor samples. *Clin Vaccine Immunol.* 2007;14(1):52–59.
35. Krieg C, et al. High-dimensional single-cell analysis predicts response to anti-PD-1 immunotherapy. *Nat Med.* 2018;24(2):144–153.
36. Abeles RD, et al. CD14, CD16 and HLA-DR reliably identifies human monocytes and their subsets in the context of pathologically reduced HLA-DR expression by CD14(hi) /CD16(neg) monocytes: expansion of CD14(hi)/CD16(pos) and contraction of CD14(lo) /CD16(pos) monocytes in acute liver failure. *Cytometry A.* 2012;81(10):823–834.
37. Green CL, et al. Recommendations for the development and validation of flow cytometry-based receptor occupancy assays. *Cytometry B Clin Cytom.* 2016;90(2):141–149.
38. O'Hara DM, Xu Y, Liang Z, Reddy MP, Wu DY, Litwin V. Recommendations for the validation of flow cytometric testing during drug development: II assays. *J Immunol Methods.* 2011;363(2):120–134.
39. Beck SC, Lock RJ. Uncertainty of measurement: an immunology laboratory perspective. *Ann Clin Biochem.* 2015;52(Pt 1):7–17.
40. Davis BH, et al. Determination of optimal replicate number for validation of imprecision using fluorescence cell-based assays: proposed practical method. *Cytometry B Clin Cytom.* 2013;84(5):329–337.
41. Kottner J, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *Int J Nurs Stud.* 2011;48(6):661–671.
42. Owens MA, Vall HG, Hurley AA, Wormsley SB. Validation and quality control of immunophenotyping in clinical flow cytometry. *J Immunol Methods.* 2000;243(1-2):33–50.
43. Weber LM, Robinson MD. Comparison of clustering methods for high-dimensional single-cell flow and mass cytometry data. *Cytometry A.* 2016;89(12):1084–1096.
44. Novakova M, et al. How to make usage of the standardized EuroFlow 8-color protocols possible for instruments of different manufacturers [published online ahead of print November 20, 2018]. *J Immunol Methods.* <https://doi.org/10.1016/j.jim.2017.11.007>.
45. Kvistborg P, et al. Thinking outside the gate: single-cell assessments in multiple dimensions. *Immunity.* 2015;42(4):591–592.
46. Döring M, et al. Human leukocyte antigen DR surface expression on CD14+ monocytes during adverse events after hematopoietic stem cell transplantation. *Ann Hematol.* 2015;94(2):265–273.
47. Maecker HT, Frey T, Nomura LE, Trotter J. Selecting fluorochrome conjugates for maximum sensitivity. *Cytometry A.* 2004;62(2):169–173.
48. Maecker HT, Trotter J. Flow cytometry controls, instrument setup, and the determination of positivity. *Cytometry A.* 2006;69(9):1037–1042.
49. Lee JA, et al. MIFlowCyt: the minimum information about a flow cytometry experiment. *Cytometry A.* 2008;73(10):926–930.
50. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>. Accessed November 8, 2018.