Centrally Determined Standardization of Flow Cytometry Methods Reduces Interlaboratory Variation in a Prospective Multicenter Study

Liset Westera, PhD[1], Tanja van Viegen, PhD[2], Jenny Jeyarajah, PhD[3], Azar Azad, PhD[4], Janine Bilsborough, PhD[5], Gijs R. van den Brink, MD, PhD[1], Jonathan Cremer[6], Silvio Danese, MD, PhD[7], Geert D'Haens, MD, PhD[2,8], Lars Eckmann, MD[9], William Faubion, MD[10], Melissa Filice[11], Hannelie Korf, PhD[6], Dermot McGovern, MD, PhD[5], Julian Panes, MD[12], Azucena Salas, PhD[12], William J. Sandborn, MD[9,13], Mark S. Silverberg, MD, PhD[11], Michelle I. Smith, PhD[11], Severine Vermeire, MD, PhD[6], Stefania Vetrano, PhD[7], Lisa M. Shackelton, PhD[3], Larry Stitt, MSc[3], Vipul Jairath, MD, PhD[3,14], Barrett G. Levesque, MD[9,13], David M. Spencer[13], Brian G. Feagan, MD[3,14] and Niels Vande Casteele, PharmD, PhD[9,13]

OBJECTIVES: Flow cytometry (FC) aids in characterization of cellular and molecular factors involved in pathologic immune responses. Although FC has potential to facilitate early drug development in inflammatory bowel disease, interlaboratory variability limits its use in multicenter trials. Standardization of methods may address this limitation. We compared variability in FC-aided quantitation of T-cell responses across international laboratories using three analytical strategies.

METHODS: Peripheral blood mononuclear cells (PBMCs) were isolated from three healthy donors, stimulated with phorbol 12-myristate 13-acetate and ionomycin at a central laboratory, fixed, frozen, and shipped to seven international laboratories. Permeabilization and staining was performed in triplicate at each laboratory using a common protocol and centrally provided reagents. Gating was performed using local gating with a local strategy (LGLS), local gating with a central strategy (LGCS), and central gating (CG). Median cell percentages were calculated across triplicates and donors, and reported for each condition and strategy. The coefficient of variation (CV) was calculated across laboratories. Between-strategy comparisons were made using a two-way analysis of variance adjusting for donor.

RESULTS: Mean interlaboratory CV ranged from 1.8 to 102.1% depending on cell population and gating strategy (LGLS, 4.4–102.1%; LGCS, 10.9–65.6%; CG, 1.8–20.9%). Mean interlaboratory CV differed significantly across strategies and was consistently lower with CG.

CONCLUSIONS: Central gating was the only strategy with mean CVs consistently lower than 25%, which is a proposed standard for pharmacodynamic and exploratory biomarker assays.

INTRODUCTION

Development of new targeted immunologic therapies for the treatment of chronic inflammatory disease remains an important clinical goal. A necessary component of this endeavor for the inflammatory bowel diseases is a greater understanding of immune pathways associated with initiation and perpetuation of intestinal inflammation. Flow cytometry (FC) is a powerful tool for monitoring immune functions that allows for simultaneous detection of several functional characteristics in single cells, including surface and intracellular components, thus permitting detailed characterization of multiple subsets of cells in complex mixtures in blood and

tissues.[1,2] The use of FC in preclinical and early-phase clinical research has the potential for timely prediction of both the performance and risks of candidate drugs and to reduce the delays and costs associated with traditional clinical drug development programs.[3,4]

Although currently used in several stages of drug development, widespread adoption of FC in multicenter clinical trials has been historically limited by complexity, cost, and inconsistent methods for sample handling, preparation, instrument setup, and data analysis among laboratories.[5] Precise gating (the sequential identification of a cellular population of interest using a panel of fluorescent markers) is a central prerequisite

[1]Tytgat Institute for Liver and Intestinal Research, Amsterdam, Netherlands; [2]Robarts Clinical Trials, BV., Amsterdam, Netherlands; [3]Robarts Clinical Trials, London, Ontario, Canada; [4]Mount Sinai Services, Toronto, Ontario, Canada; [5]F. Widjaja Foundation Inflammatory Bowel and Immunobiology Institute, Los Angeles, CA, USA; [6]University of Leuven, Leuven, Belgium; [7]Humanitas University, Milan, Italy; [8]Academic Medical Center, Amsterdam, Netherlands; [9]Division of Gastroenterology, School of Medicine, University of California San Diego, San Diego, CA, USA; [10]Mayo Clinic, Rochester, MN, USA; [11]Mount Sinai Hospital, Toronto, Ontario, Canada; [12]Institut d'Investigacions Biomèdiques August Pi i Sunyer- Hospital Clínic Barcelona, Barcelona, Spain; [13]Robarts Clinical Trials, San Diego, CA, USA and [14]University of Western Ontario, London, Ontario, Canada
Correspondence: N. Vande Casteele, PharmD, PhD, Division of Gastroenterology, School of Medicine, University of California San Diego, IBD Center, 9500 Gilman Drive #0956, La Jolla, CA 92093, USA. E-mail: nvandecasteele@ucsd.edu
Received 12 June 2017; accepted 22 September 2017

of reliable FC, yet remains the largest single contributor to interlaboratory variability.[5] Such variability may be reduced by the use of a shared gating template among laboratories, or the use of centralized gating strategies, but the detailed procedures and degree of achievable improvement remain to be firmly established.[5,6]

Standardization of methods across laboratories will allow for the potential inclusion of FC of intestinal tissue biopsies in early-phase multicenter clinical trials. In this trial, we aimed to compare the variability in quantitation of T-cell responses from peripheral blood mononuclear cells (PBMCs) across major international laboratories using a common protocol for permeabilization and staining, and three different gating strategies.

METHODS

This study was performed at eight international laboratories associated with tertiary inflammatory bowel disease centers in North America and Europe. The Tytgat Institute (Academic Medical Center, Amsterdam, The Netherlands) acted as the central laboratory. All pre-analytical procedures described in the following section and gating for analysis of data collected by the local laboratories was performed by a single investigator (LW) at the central laboratory. The protocol for the study was developed by Robarts Clinical Trials, in collaboration with the participating expert laboratories. Figure 1 shows the overall study design.

### Central laboratory (pre-analytical) procedures

Preparation of PBMCs and master mixes. PBMCs were isolated from buffy coats of three healthy volunteers using Ficoll (Ficoll Paque PLUS, GE Healthcare/VWR, Radnor, PA, USA) gradient centrifugation and stored overnight at 4 °C in complete culture medium (RPMI 1640 supplemented with 10% fetal bovine serum, 1% penicillin/streptomycin, and L-glutamine). The following day, PBMCs were centrifuged and resuspended in either a cocktail containing phorbol 12-myristate 13-acetate and ionomycin plus protein transport inhibitors or transport inhibitors alone (Cell Stimulation Cocktail (plus protein transport inhibitors), Thermo Fisher Scientific/eBioscience, Carlsbad, CA, USA), followed by incubation in culture flasks for 4 h at 37 °C. Stimulated and unstimulated PBMCs were collected, washed, and resuspended in 1× phosphate-buffered saline containing Fixable Amine-reactive Live/Dead marker AQUA (Thermo Fisher Scientific/Life Technologies, Carlsbad, CA, USA) (except for one tube of cells to be used for instrument setup), followed by incubation in the dark for 30 min at room temperature. Cells were washed in phosphate-buffered saline and resuspended in a freshly diluted ezKine Fix/Lyse solution prepared per the manufacturer's instructions (eZKine Th1/Th17 Whole Blood Intracellular Cytokine Kit, Thermo Fisher Scientific/eBioscience), and incubated for 25 min at room temperature. Fixed PBMCs (1×10^6 cells) were then frozen and stored as 1 ml aliquots at −80 °C.

To generate antibody master mixes for staining of cell surface and cytokine targets, the eZKine kit Th1/Th17 cytokine and concentration-matched isotype cocktails (containing CD3, CD4 plus interleukin (IL)-17A, interferon (IFN)-gamma antibodies, or isotype controls) were supplemented with titrated amounts of antibodies against CD8 (Anti-human CD8a AlexaFluor700 (HIT8a, Mouse IgG1), BioLegend, San Diego, CA, USA) and γδ T-cell receptor (TCR; Anti-human γδTCR PE-Cy7 (B1, Mouse IgG1) BioLegend). Master mixes were prepared and aliquoted shortly before shipment to the different laboratories and stored at 4 °C. Frozen PBMCs, master mixes, and aliquoted materials required for setup (listed under local laboratory procedures) were shipped to the participating laboratories on dry ice (PBMCs and live-dead marker) or cold packs (antibodies and compensation beads). Experiments were conducted and reagents were used within 4 weeks of shipping to the participating laboratories.

Local laboratory procedures. Permeabilization and staining of PMBCs was performed at each laboratory in triplicate using a common protocol based on the manufacturer's instructions (eZKine Whole Blood Intracellular Cytokine Kit, Thermo Fisher Scientific/eBioscience) and centrally provided reagents. Fixed and frozen cells were thawed at 37 °C and transferred to FC tubes. Permeabilization Buffer (1×) was then added and cells were centrifuged, resuspended in 1× Permeabilization Buffer, and recentrifuged. Following decanting of supernatant and vortexing, either cytokine or isotype master mixes were added to each tube. After incubation for 30 min in the dark at room temperature, additional 1× Permeabilization Buffer was added to each tube followed by centrifugation. Following aspiration of supernatant, cells were washed, recentrifuged, resuspended in FC buffer (phosphate-buffered saline+1% bovine serum albumin), and stored at 2–8 °C until analysis, which was performed on the same day as staining.

Single-stained compensation controls were prepared using UltraComp eBeads (Thermo Fisher Scientific/eBioscience) for IL-17A-PE and TCRγδ-PE-Cy7, ArC amine-reactive/negative compensation beads (Thermo Fisher Scientific/Life Technologies) for the live/dead marker,
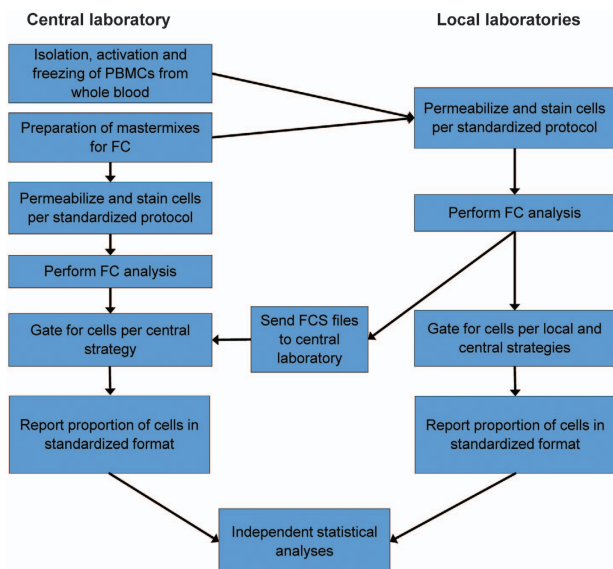
Figure 1 Overall study design.
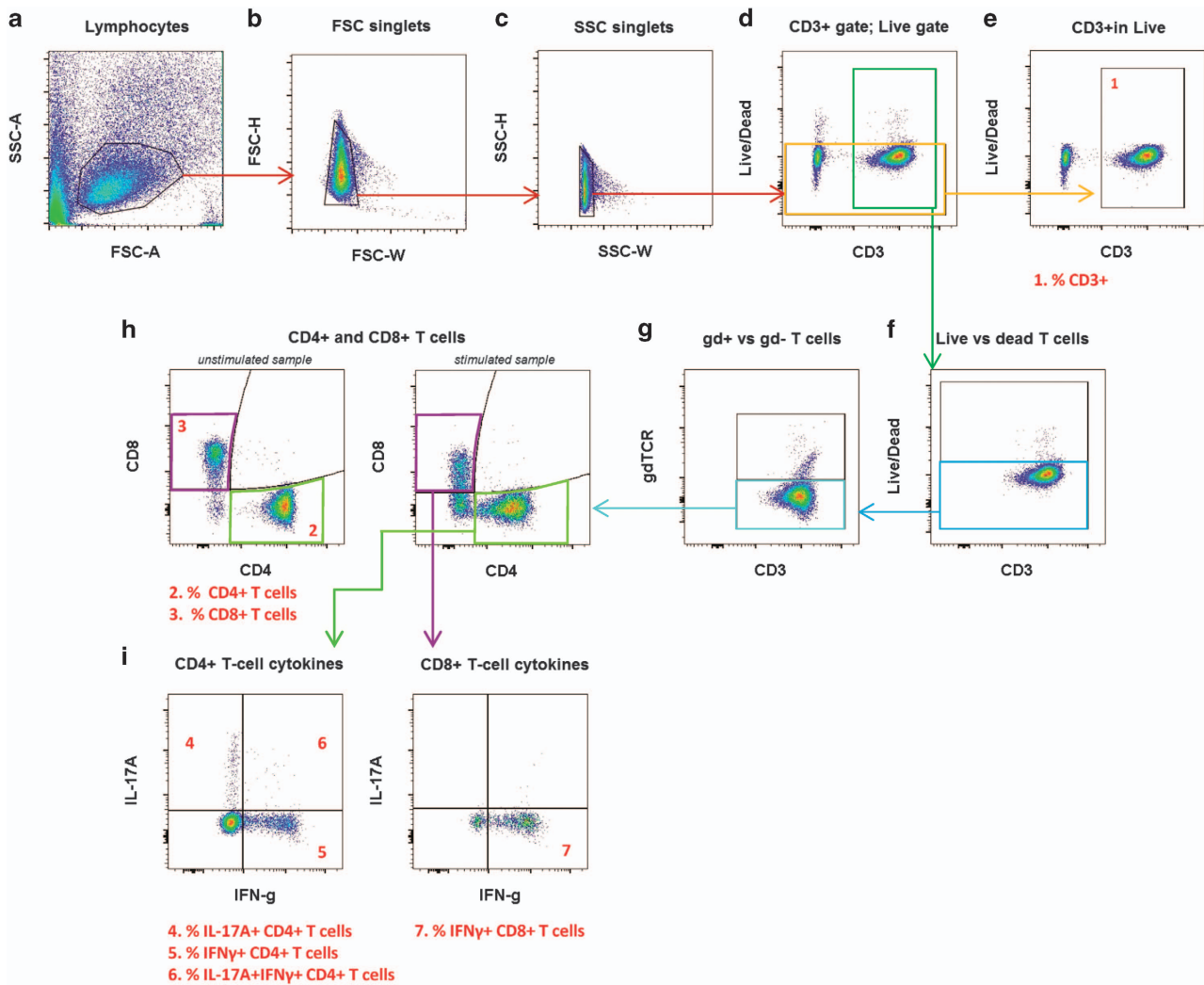
Figure 2 Central gating strategy used to distinguish (**a**) lymphocytes, singlets based on (**b**) forward scatter and (**c**) side scatter, (**d**) live lymphocytes (yellow gate) or CD3+ cells (green gate), (**e**) CD3+ cells within lymphocytes, (**f**) live CD3+ cells (blue gate), (**g**) γδTCR- (hence αβTCR+) T cells (cyan gate), (**h**) CD4+ (green gate) versus CD8+ T cells (purple gate) in unstimulated (left) and stimulated (right) samples, and (**i**) subsets of cytokine-positive cells within CD4+ (left) and CD8+ (right) T cells.

and stimulated, fixed PBMCs ("setup cells") for CD3-APC, CD4-PerCP-eFluor710, CD8α-AlexaFluor700, IFNγ-FITC (because for these markers fluorescence intensity on cells exceeded that on saturated beads), and unstained control. Antibodies (PE IL-17A or PE-Cy7 TCRγδ) or dye (AQUA Live/Dead) were added to appropriate beads and incubated in the dark for 20 min at 2–8 °C, followed by washing, centrifugation, and resuspension in FC buffer.

Although we did not require the use of specific instrumentation, parameters were generally recorded on Beckton Dickson LSRFortessa (BD Biosciences, San Jose, CA, or equivalent) cell analyzers. Unstained and single-stained cells, and single-stained beads were used for instrument setup and to calculate compensation values using the automated calculation function.

We also did not require specific software (or versions thereof) for analysis, although most laboratories used FACS-Diva (BD Biosciences) or Flowjo software (FlowJo LLC, Asland, OR). Gating was performed on singlets. Parameters were analyzed by the laboratories using local gating with a local strategy (LGLS) or local gating with a central strategy (LGCS) developed by the central laboratory in collaboration with the individual laboratories (Figure 2). In addition, original Flow Cytometry Standard files for each experiment were sent from the individual laboratories to the central laboratory for analysis (central gating (CG)) and analyzed with Flowjo software over a period of several weeks.

Statistical analyses. Median cell percentage was calculated across triplicates and donors, and reported for both stimulated and unstimulated conditions and strategies. For each FC parameter the mean across the three replicates was calculated and then the percentage coefficient of variation (CV) was calculated for each donor as $100 \times$ s.d./mean across laboratories. Among-strategy comparisons were made using a two-way analysis of variance adjusted for donor. Sample size was not determined *a priori* due to the exploratory nature and complexity of the study design. A convenience sample of three donors was used.
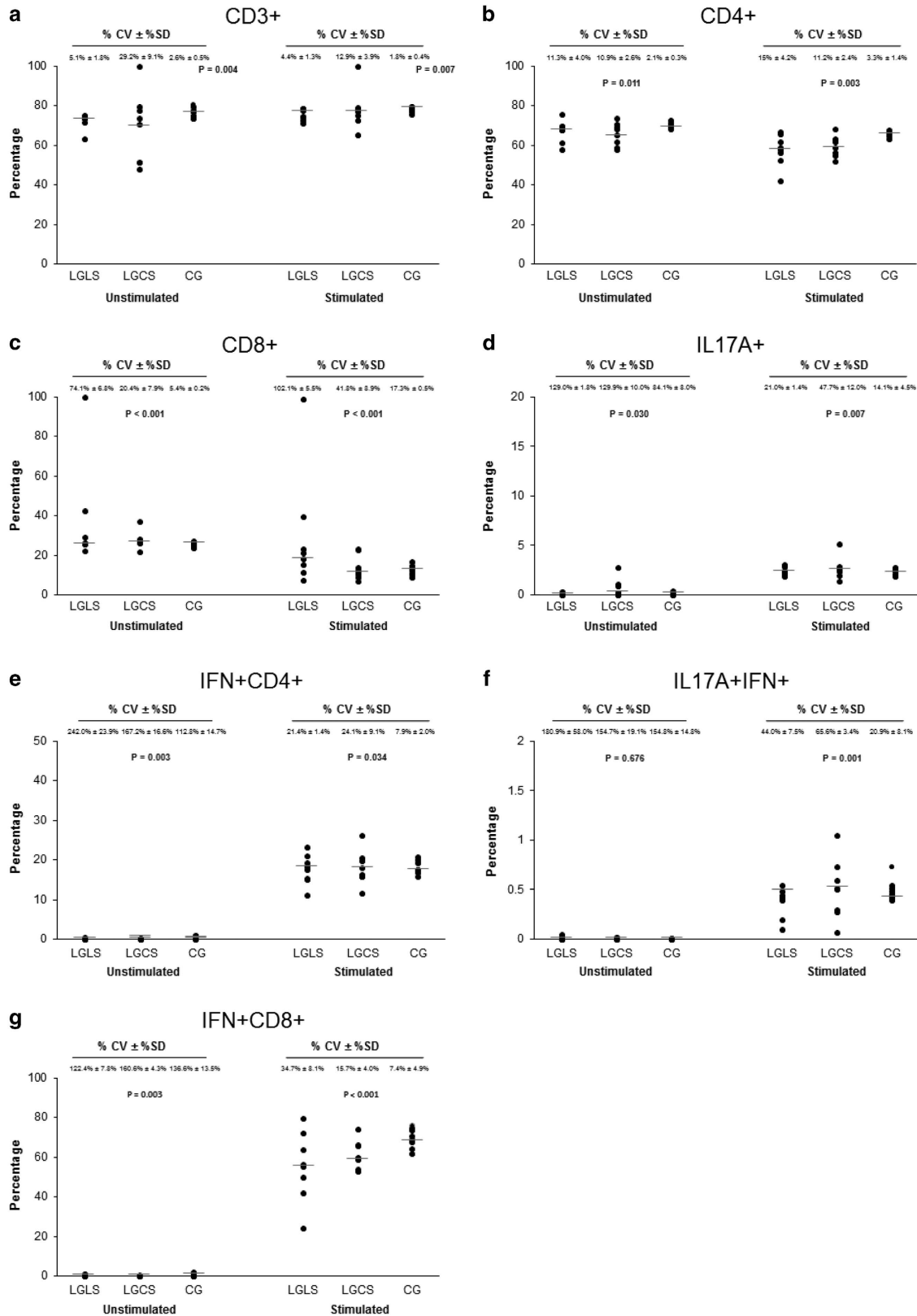
Figure 3 Median cell percentages for each parameter calculated across triplicates and donors in both unstimulated and stimulated conditions as reported by the laboratories for each gating strategy. Associated mean interlaboratory variability (% CV +/- %SD) for each gating strategy is shown at the top of the graphs.

RESULTS

Median cell percentages. The median percentage (across donors) of cells for each parameter tested (under stimulated and unstimulated conditions) reported by the individual laboratories according to the three gating strategies tested in this study is shown in Figure 3. In general, the median percentages reported for each parameter did not change substantially by gating strategy. The parameter with the highest reported median percentage (for both stimulated and unstimulated conditions) was CD3+ cells within lymphocytes and the lowest median percentage reported was for IL-17A +IFNγ+CD4+ cells within CD4+CD3+ αβ T cells. The median percentage reported for each parameter ranged from 0 to 76.9%. Several parameters were associated with very low median percentages of positive cells ($<2.5\%$) in both stimulated and unstimulated conditions (e.g., IL-17A+ and IL-17A+IFNγ+CD4+ cells within CD4+CD3+ αβ T cells). As expected, the median percentage of cytokine-expressing cells detected in unstimulated conditions was $<0.22\%$. In addition, as expected, the median percentages of CD4+ and CD8+ cells were reduced in stimulated conditions due to a decrease in CD4 and CD8 expression levels (which was not overcome by intracellular staining of surface markers) that likely led to cells falling outside the marker-positive gate.

Intralaboratory variation. To address the variation between replicate samples measured in the same laboratory, intralaboratory CVs were calculated from triplicate measurements of all parameters (Supplementary Table 1). Mean intralaboratory CV ranged from $0.95\% \pm 0.92\%$ to $37.9\% \pm 34.2\%$ depending on cell population, conditions, and gating strategy ($2.2\% \pm 1.1\%$ to $22.9\% \pm 9.9\%$ for LGLS, $1.3\% \pm 1.1\%$ to $37.9\% \pm 34.2\%$ for LGCS, and $0.95\% \pm 0.92\%$ to $19.2\% \pm 7.8\%$ for CG). The mean intralaboratory CV for cytokine-positive cells in unstimulated conditions ranged from $49.5\% \pm 23.4\%$ to $161.9\% \pm 15.1\%$ and no difference was observed based on gating strategy.

Intralaboratory CV was lowest for parameters with higher median percentages of positive cells (e.g., CD3+ cells within lymphocytes, and CD4+ or CD8+ cells within CD3+ αβ T cells). As expected, parameters with very low median percentages of positive cells were associated with the highest intralaboratory CV (e.g., IL-17A+IFNγ+CD4+ cells within CD4+CD3+ αβ T cells in unstimulated conditions).

Interlaboratory variation. To address the variation between different labs, interlaboratory CVs were calculated from triplicate measurements of all parameters (Figure 3). The mean interlaboratory CV ranged from $1.8\% \pm 0.4\%$ to $102.1\% \pm 5.5\%$ depending on cell population and gating strategy ($4.4\% \pm 1.3\%$ to $102.1\% \pm 5.5\%$ for LGLS, $10.9\% \pm 2.6\%$ to $65.6\% \pm 3.4\%$ for LGCS, and $1.8\% \pm 0.4\%$ to $20.9\% \pm 8.1\%$ for CG). The mean interlaboratory CV for cytokine-positive cells in unstimulated conditions ranged from $84.1\% \pm 8.0\%$ to $242.0\% \pm 23.9\%$.

Central analysis with a CG strategy was associated with the lowest interlaboratory CV, and with a narrower range of CV compared with either of the local strategies (using local or CG). Similar to what was observed for intralaboratory CV,

parameters with low median percentages of positive cells were associated with higher mean interlaboratory CV regardless of gating strategy (Figure 3, unstimulated conditions, d–f). However, in contrast to intralaboratory CV, the mean interlaboratory CV differed significantly across gating strategies for all of the parameters tested, and was consistently lower with CG than either local gating with local strategy or local gating with central strategy, with the exception of IL-17A+IFNγ+CD4+ within CD4+CD3+ αβ T cells and IFNγ+CD8+ within CD8 +CD3+ αβ T cells in unstimulated conditions. It should be noted, however, that the median percentages of background cytokine-expressing cells reported by the laboratories were exceedingly low ($<0.22\%$) in unstimulated conditions, which makes the estimates of inter- and intralaboratory variability for these parameters essentially meaningless. Mean interlaboratory CV for LGCS was not consistently less than for LGLS.

DISCUSSION

We have demonstrated that it is procedurally feasible for multiple international laboratories with expertise in inflammatory bowel disease research to collaborate and contribute FC data for reproducible FC analysis by a central laboratory. We also show that permeabilization and staining of previously frozen and fixed cells may be successfully performed by independent labs and results in acceptable interlaboratory variability according to industry standards ($<25\%$ interlaboratory CV),[2] when the FC analyses are performed by a central laboratory. The decrease in the interlaboratory variability observed with central FC data analysis is consistent with the observation that central assessment can reduce variability and increase precision, as seen with other measurement modalities such as endoscopic activity in inflammatory bowel disease clinical trials.[7]

The results of our study are consistent with earlier observations,[6] and confirm that gating is a major source of interlaboratory variability in the analysis of FC data, and that this variability can be significantly reduced when analyses are performed by a central laboratory using a pre-specified gating strategy. Surprisingly, despite our collaborative efforts to harmonize the gating strategy used for analysis by the laboratories, use of a CG strategy by the individual laboratories did not consistently reduce variability of the results compared to gating performed according to the various local gating strategies. Although this outcome was unexpected, application of a central laboratory's gating strategy may be less intuitive when data appear slightly different than the single gating example provided. Furthermore, the example provided to the laboratories in our pilot study was not accompanied by detailed recommendations for gating. The provision of specific instructions and/or recommendations with a gating example may aid in reducing variability when analyses are performed by individual laboratories. A marked benefit for harmonizing gating strategies to decrease variability and to increase accuracy and comparability of results across laboratories has been previously demonstrated for intracellular cytokine-staining assays.[8] Future studies may further address the subjectivity of manual analysis through the use of emerging tools such as automated analysis or machine learning.[9]

We also did not achieve a consensus among the participating laboratories before initiation of the study for methods for gating on cytokine-positive vs. -negative cells. Laboratories may have defined these cells differently based on either isotype control background, negative control background, or on the "internal negative population" of cytokine-stained cells (and where the positive population of cells in these samples is found). Although isotype controls (which address nonspecific antibody binding) have been historically used as a negative control in FC and are typically supplied in commercial cytokine kits, their added value in FC experiments has been much debated and it is now increasingly accepted that they should not be used to set gates.[10,11] This is illustrated by the finding that background staining of the isotype control was higher than the internal background (i.e., cytokine-negative cells) in cytokine-stained samples in this study (a number of cytokine-positive cells were excluded based on isotype background). Although the background staining of the unstimulated control was comparable to the internal background of cytokine-stained samples (due to potential differences in the inherent "stickiness (e.g., nonspecific binding)" of cells between stimulated and unstimulated conditions), an unstimulated control is generally also not an appropriate gating control. The advantage of an internal negative control population (provided that this can be easily identified, e.g., IFNγ – IL-17A – CD4+ T cells as negative control for cytokine-positive CD4+ T cells) is that these cells have been exposed to the same conditions as the marker-positive cells. The gating boundaries in the central laboratory were consistently based on the internal negative population.

Fluorescence-minus-one (FMO) controls (which address background by fluorescent spillover) were absent from this study, but may help in determining gating boundaries of positive and negative populations. As most background variation in multicolor experiments with pre-titrated monoclonal antibodies is spillover-induced, FMO controls are generally considered most appropriate for determining gating boundaries. In future experiments where internal negative populations (e.g., in patients with existing inflammation) cannot be identified, the inclusion of one or more FMO controls may be useful to improve consistency in the identification of positive and negative populations.

We also confirm, not unexpectedly, that intra- and interlaboratory variability in the quantitation of T-cell responses is inversely related to the percentage of positive cells detected in the parent population. In an attempt to control for factors other than gating that might influence variability and minimize inherent baseline differences in immune function/parameters between patient samples, we analyzed T-cell responses in healthy donors and examined parameters that occurred over a broad range of cell percentages. As expected, cytokine-positive cells (e.g., IL-17A or IFNγ) were detected at exceedingly low cell percentages (e.g., $< 0.22\%$) in unstimulated conditions, most likely due to nonspecific induction. Quantitation of these cells resulted in high intra- and interlaboratory variability, although variability was lowest for the majority of the parameters in either stimulated or unstimulated conditions when the analysis was performed by the central lab. Although we did not pre-specify number of events to acquire in our study, variability associated with quantitating parameters occurring at very low frequencies may be partially addressed by increasing the number of events acquired. In our study, the mean (s.d.) event number for CD3+ cells according to each laboratory ranged from 4,132 (1,393) to 16,176 (992). Future studies should critically evaluate the feasibility of reliably analyzing all populations of interest, particularly those that may occur infrequently, and ensure that instrumentation does not limit the ability of participating laboratories to sufficiently distinguish all populations of interest.

A final limitation of our study is the generalizability of the results. In addition to healthy donors, the sample size for analysis was small and homogeneous relative to what might be expected for a clinical trial, and the initial cell isolation and preparation of master mixes for staining was performed by the central laboratory. The role of a central laboratory in a large multicenter trial in inflammatory bowel disease, for example, would differ when FC analysis depends upon on-site processing of biopsy samples collected at the time of endoscopy. In this case, factors such as sample isolation, preparation (including time to processing), and storage might be expected to introduce additional variability to the results of FC analysis. Efforts on the harmonization/standardization of FC methods in a multicenter setting are ongoing in other disease areas[12] and through the International Society for Advancement of Cytometry.

In conclusion, methods to decrease interlaboratory variability associated with the outcomes of FC are essential for the implementation of this powerful translational research tool in multicenter research. Central FC data analysis overcomes the variability associated with both inherent instrument, and subjective operator differences, and is likely to be the preferred strategy in future inflammatory bowel disease clinical trials. Additional refinement and harmonization of gating strategies may facilitate reliable assessment of FC data by multiple independent laboratories and allow for accurate interpretation of integrated data from multicenter trials. Research on the best methodological approaches for standardization of FC analysis of cells isolated from biopsy tissue samples is ongoing, and should accelerate both our understanding of the cellular and molecular basis of the inflammatory bowel diseases, as well as drug discovery and personalized approaches to treatment.

CONFLICT OF INTEREST

Study Highlights

WHAT IS CURRENT KNOWLEDGE

✓ Flow cytometry (FC) is a powerful tool for monitoring immune functions.

✓ The use of FC in multicenter trials has been hampered, in part, by a lack of standardized methodology resulting in unacceptably high interlaboratory variability.

✓ Gating is the largest contributor to this variability.

WHAT IS NEW HERE

✓ Central analysis of FC data reduces interlaboratory variability to a greater extent than that achieved with the use of a shared gating template by local laboratories, or when gating is performed according to local protocols.

✓ When FC data are centrally analyzed, interlaboratory variability lies within an acceptable range (25%) for pharmacodynamic and exploratory biomarker assays.

TRANSLATIONAL IMPACT

✓ Our research demonstrates that it is feasible to conduct FC at multiple laboratories with acceptable precision using a common protocol for permeabilization and staining, and centrally performed gating and analysis.

✓ Additional research is ongoing to standardize methods for FC analysis of cells isolated from biopsy samples from patients with inflammatory bowel disease.

✓ These findings will facilitate the use of FC in global multicenter trials.

1. Maecker HT, McCoy JP, Nussenblatt R. Standardizing immunophenotyping for the Human Immunology Project. *Nat Rev Immunol* 2012; **12**: 191–200.
2. O'Hara DM, Xu Y, Liang Z *et al.* Recommendations for the validation of flow cytometric testing during drug development: II assays. *J Immunol Methods* 2011; **363**: 120–134.

3. Von Herrath MG, Nepom GT. Lost in translation: barriers to implementing clinical immunotherapeutics for autoimmunity. *J Exp Med* 2005; **202**: 1159–1162.

4. Davis MM. A prescription for human immunology. *Immunity* 2008; **29**: 835–838.

5. Maecker HT, McCoy JP Jr, Amos M *et al.* A model for harmonizing flow cytometry in clinical trials. *Nat Immunol* 2010; **11**: 975–978.

6. Maecker HT, Rinfret A, D'Souza P *et al.* Standardization of cytokine flow cytometry assays. *BMC Immunol* 2005; **6**: 13.

7. Feagan BG, Sandborn WJ, D'Haens G *et al.* The role of centralized reading of endoscopy in a randomized controlled trial of mesalamine for ulcerative colitis. *Gastroenterology* 2013; **145**: 149–157.

8. MacNeil LK, Price L, Britten CM *et al.* A harmonized approach to intracellular cytokine staining gating: Results from an international multiconsortia proficiency panel conducted by the Cancer Immunotherapy Consortium (CIC/CRI). *Cytometry A* 2013; **83**: 728–738.

9. White S, Laske K, Welters MJ *et al.* Managing multi-center flow cytometry data for immune monitoring. *Cancer Inform* 2015; **13** (Suppl 7): 111–122.

10. Maecker HT, Trotter J. Flow cytometry controls, instrument setup, and the determination of positivity. *Cytometry A* 2006; **69**: 1037–1042.

11. Hulspas R, O'Gorman MR, Wood BL *et al.* Considerations for the control of background fluorescence in clinical flow cytometry. *Cytometry B Clin Cytom* 2009; **76**: 355–364.

12. Mandruzzato S, Brandau S, Britten CM *et al.* Toward harmonized phenotyping of human myeloid-derived suppressor cells by flow cytometry: results from an interim study. *Cancer Immunol Immunother* 2016; **65**: 161–169.

Supplementary Information accompanies this paper on the Clinical and Translational Gastroenterology website (http://www.nature.com/ctg)