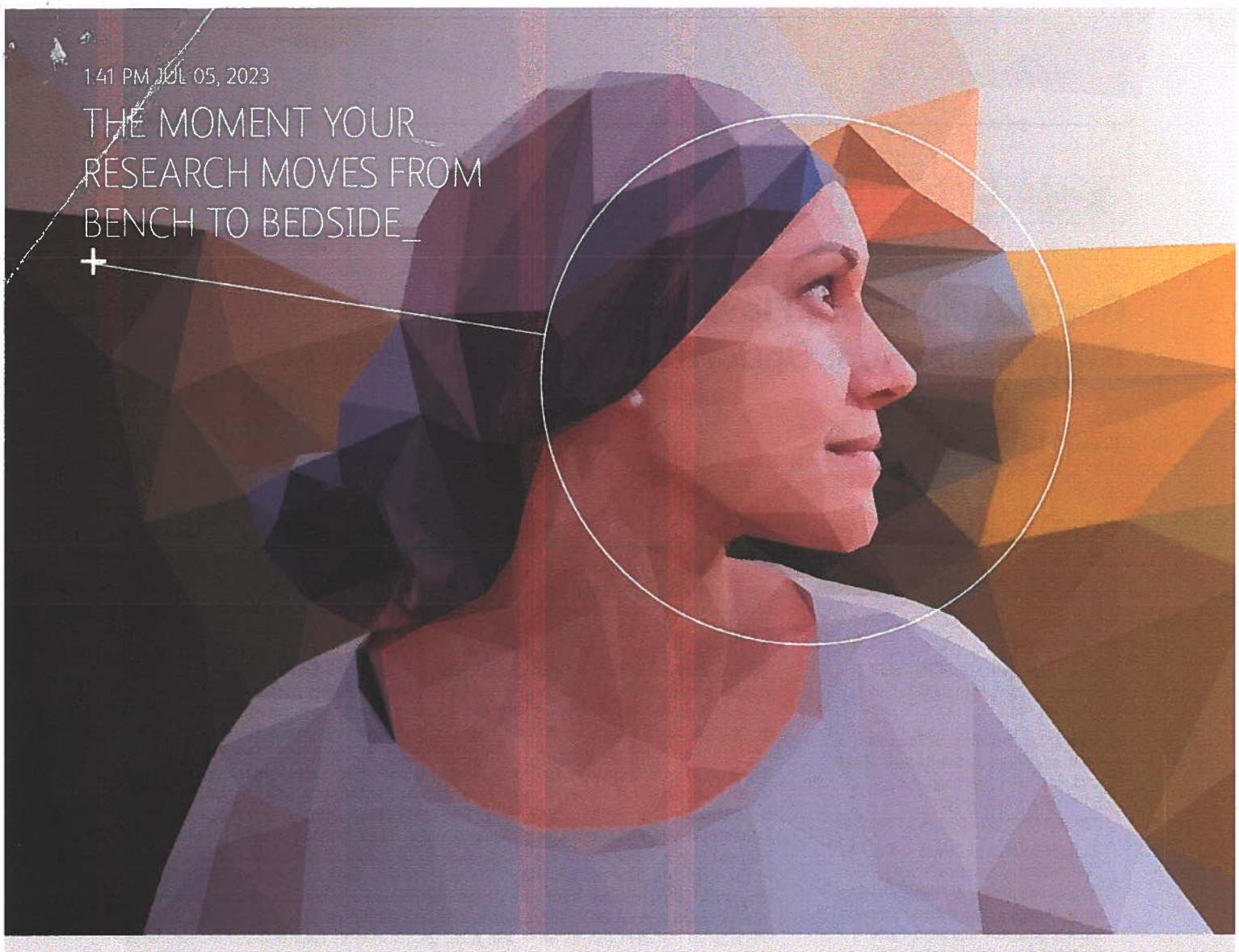

THE DIFFERENCE OF BREAKTHROUGH RESEARCH TO ENABLE MORE CURES

WITH BD AS YOUR PARTNER IN IMMUNO-ONCOLOGY FROM BENCH TO BEDSIDE. BD is your partner in immuno-oncology from discovery research to clinical applications, with a range of high-performance solutions designed to give you high-quality, reliable research data for even the most complex experiments. More solutions. More answers. More data you can trust. Discover more with BD at your side. **Discover the new BD.**

Learn how you can advance your research >

BD
Advancing the world of health

ISAC CYTOMETRY
Journal of Quantitative Cell Science PART A

Reproducibility of Flow Cytometry Through Standardization: Opportunities and Challenges

Tomas Kalina*

CLIP-Childhood Leukemia Investigation Prague, Department of Pediatric Hematology and Oncology, 2nd Medical School, Charles University and University Hospital Motol, Prague, Czech Republic

Received 21 March 2019; Revised 4 September 2019; Accepted 11 September 2019

Grant sponsor: Ministry of Education, Youth and Sports, Czech Republic, Grant number LO1604; Grant sponsor: EU-Prague, Grant numbers CZ.2.16/3.1.00/24505, CZ.2.16/3.1.00/21540

*Correspondence to: Tomas Kalina, CLIP-Childhood Leukemia Investigation Prague, Department of Paediatric Hematology/Oncology, 2nd Medical School, Charles University Prague, V Uvalu 84, 150 06 Praha 5, Czech Republic. Email: tomas.kalina@lfmotol.cuni.cz

Published online 8 October 2019 in Wiley Online Library (wileyonlinelibrary.com)

DOI: 10.1002/cyto.a.23901

Abstract

There is an agreement in the field that interlaboratory reproducibility of flow cytometry measurements as well as the whole studies might be improved by a consensual use of methodological approach. Typically, a consensus is made on a crucial markers needed in the immunostaining panel, sometimes on the particular fluorochrome conjugates and rarely on a complete set of methods for sample preparation. The term "standardization" is used to describe the complete set of methodical steps, while "harmonization" is used for partial agreement on the method. Standardization can provide a platform for improved reproducibility of cytometry results over prolonged periods of time, across different sites and across different instruments. For the purpose of structured discussion, several desired aims are described: common interpretation of the immunophenotype definition of a target subset, accurate quantification, reproducible pattern of a multicolor immunophenotype, and reproducible intensity of all measured parameters. An overview of how standardization was approached by several large consortia is provided: EuroFlow, The ONE Study, Human Immunology Project Consortium (HIPC), and several other groups. Their particular aims and the tools adopted to reach those aims are noted. How those standardization efforts were adopted in the field and how the resulting outcome was evaluated is reviewed. Multiple challenges in the instrument hardware design, instrument setup tools, reagent design, and quality features need to be addressed to achieve optimal standardization. Furthermore, the aims of different studies vary, and thus, the reasonable requirements for standardization differ. A framework of reference for the reasonable outcomes of different approaches is offered. Finally, it is argued that complete standardization is important not only for the reproducibility of measurements but also for education, for quality assessment and for algorithmic data analysis. The different standardized approaches can and in fact should serve as benchmarking reference tools for the development of future flow cytometry studies. © 2019 International Society for Advancement of Cytometry

Key terms

standardization; flow cytometry; EuroFlow; data analysis

Why, at What Level and to What Extent We Need Reproducibility

A central question to any laboratory method is its reproducibility. Flow cytometry maintains a high overall reproducibility potential within a particular laboratory whenever an appropriate sample preparation protocol with a quality antibody reagent, a stable instrument and proper analysis is deployed. This has led to the widespread use of flow cytometry in immunology, hematology, oncology, cell therapy and other disciplines, ranging from research applications to translational research involving human samples to diagnostic applications and cell therapy products. However, the reproducibility of findings is not as straightforward when a particular assay is adopted in a different laboratory, which is mostly performed using a locally preferred staining protocol, an experimental design relying on locally chosen reagents, a local instrument setup with locally chosen settings and local gating strategies. Diverse applications pose different stringency requirements on reproducibility

parameters. For the purpose of a structured discussion on reproducibility, one should acknowledge that it is a rather broad concept that may have a range of different meanings. In general, we can ask for reproducibility at different levels: (a) interpretation, (b) enumeration (Fig. 1), (c) staining pattern, and (d) staining intensity (1). The focus of this review will be aimed at interlaboratory reproducibility.

Most applications of clinical cytometry aim for interpretation (e.g., "leukemia of B-cell precursor origin is present in the sample") and aim for precise enumeration (CD4 counts, CD34 counts). Translational and research applications typically aim for the enumeration of multiple subsets at once from a given specimen to provide a broad description of cellular composition (e.g., percentage of various immune system cells in blood) (2). Staining pattern is mostly categorized by the positivity or negativity of a given marker on a cell population of interest, although it can also be approached bioinformatically across multiple parameters at once. Often, intensity of expression is annotated as overexpression (strong expression), normal expression or dim (weak) expression as exemplified by AIEOP-BFM Consensus Guidelines 2016 (3). However, this is where reproducibility issues arise. Experimental design plays a major role in the outcome of staining pattern (e.g., choice of markers, choice of fluorochromes, choice of monoclonal antibody clones, choice of instrument and instrument settings, and sample preparation protocol) as does the experimental execution (e.g., adherence to the protocols and experimental robustness). Not surprisingly, manual analysis influences the end result to a great extent. When interpretation and enumeration is the intended outcome, multiple variants of staining protocols and instrument setups may yield comparable results in CD4+ T-cells (Fig. 1A), but would yield larger discrepancy in CD8+ T-cells (Fig. 1B), stressing the need for subset specific benchmarking. To complicate the matter, particular diseases present with high proportions of subsets that are very rare in healthy (4,5) necessitating disease specific benchmarking. Definitions that influence staining patterns may impact the ability to clearly define and/or resolve particular subsets, which in turn will impact enumeration. Both, practical and theoretical considerations of validation of cell-based assays are summarized in a series of articles by International Council for Standardization of Haematology (ICSH) and International Clinical Cytometry Society (ICCS) (6–10). Excellent overview of cytometry

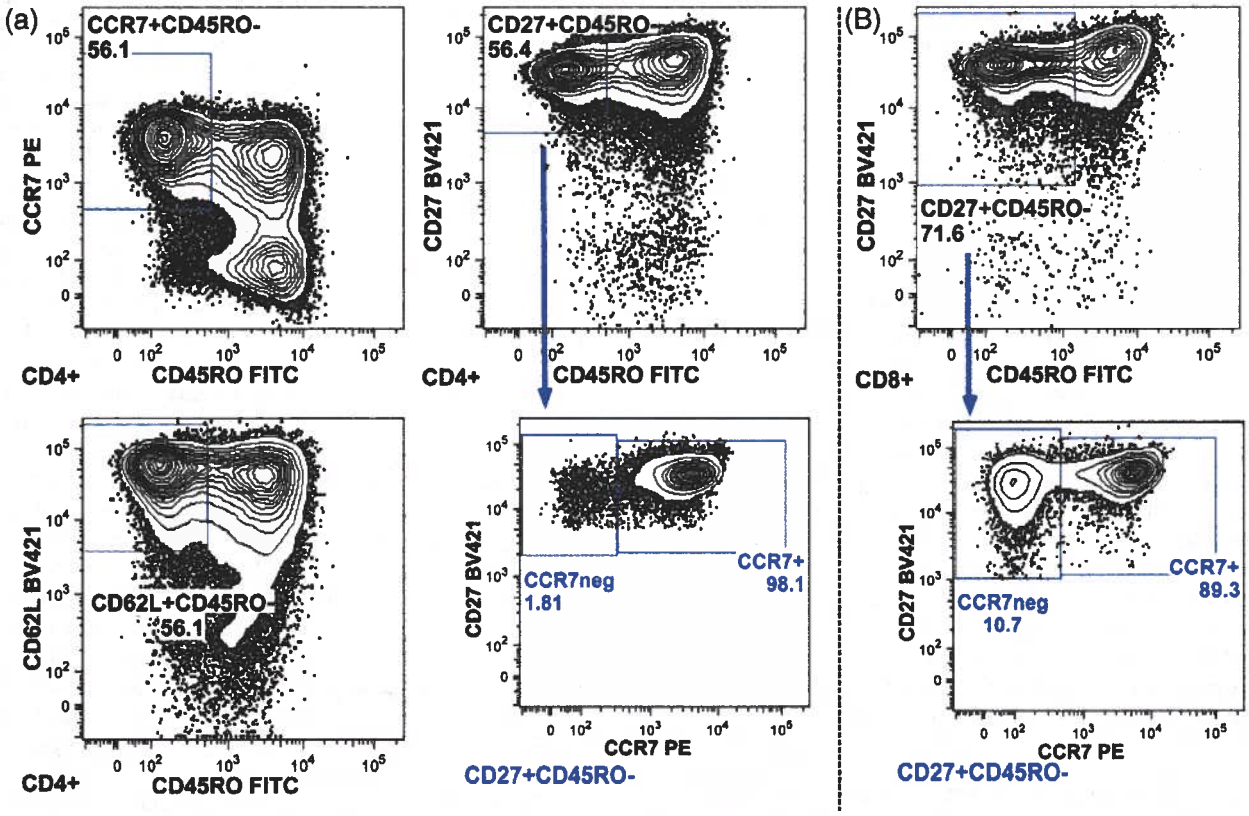

Figure 1. Immunophenotype interpretation and enumeration reproducibility benchmarking. (A) Three different combinations of surface markers (CD27 + CD45RO−, CD62L + CD45RO−, CD197(CCR7) + CD45RO−) allow for Naïve T-cell definition in CD4+ T cell subset and consequently for comparable enumeration (56%) of Naïve in CD4+ T cells. Note that small proportion (<2%) of CD27 + CD45RO− forms CCR7negative subset (B) Naïve T-cell definition using CD27 + CD45RO− in CD8+ T-cell subset would however capture 10.7% CCR7negative subset making the enumeration discrepancy of CD8 Naïve T-cell by CD27 + CD45RO− versus CD197(CCR7) + CD45RO− notable. Different disease states might pronounce this discrepancy, necessitating disease specific benchmarking. [Color figure can be viewed at wileyonlinelibrary.com]

methods, protocols, and discussion of technical consideration is presented in a recent guidelines for the use of flow cytometry and cell sorting in immunological studies (11).

Thoughtful definition of reproducibility aim calls for particular methodological approach. The term standardization is used for a complete and detailed set of linked methodical steps, supported by standard operating protocols (SOP), where any recognized sources of variation are removed. This includes pre-analytical (sample preparation and immunostaining), analytical (instrument setup), and post-analytical (analysis and interpretation) steps. In contrast, the term harmonization is used to describe an approach where only key steps of the whole approach are performed in a similar manner while other parts are performed by locally preferred method. In general, while standardization can achieve all above mentioned reproducibility aims, harmonization can yield reproducible interpretation, enumeration and in some instances (e.g. harmonized immunostaining panel with locally optimized sample preparation and instrument setup) even a reproducible pattern. These terms are not strictly exclusive, they in fact present a spectrum that can develop in the course of the interlaboratory studies. Harmonization can be a sensible and achievable first step and may gradually develop into more complex set of methods that can be considered standardization. Either approach should be transparent about the desired outcome, should provide measurable outcome parameters and thus enable for a benchmarking. Only then it can actually contribute to the overall increased reproducibility.

Flow cytometry presents a paradox as all generated data are in a digital format from their inception; however, the data interpretation is mostly visual by an experienced analyst (12). The processing of large datasets of multiple measurements by computational tools is thus relatively accessible (the data are available) but is not employed at large by the cytometry community. Several studies have indicated that manual analysis is a major source of variation (13,14), and the best documented example of improving reproducibility by gating strategy unification was the ISHAGE guidelines on CD34+ progenitor enumeration (15,16). Computational tools are clearly of major importance to extend our understanding of the complex multicolor data we acquire (12), but such tools also present the possibility to analyze large datasets acquired in multiple laboratories. However, this requires that staining patterns are not compromised by technical variability. This is relatively easy to adopt in a single laboratory with single instrument settings and data collection over prolonged time periods (month or years) but poses challenges in coordinated studies involving multiple institutions and multiple (even diverse) instruments.

Existing Efforts to Achieve Reproducibility

How Large-Scale Interlaboratory Studies Approach Reproducibility and Standardization?

Different consortia listed below aimed to address diverse levels of reproducibility: (a) the reproducibility of interpretation (all consortia; some argue that more is not necessary for their purpose such as Harmonemia (17), AIEOP-BFM group (3,18–23), and ERIC (21–23)), (b) the reproducibility of enumeration (prominently in immunology: The ONE Study (24,25), HIPC (14,26,27), but also in the leukemia residual disease detection: EuroFlow (28–30), COG group (31), and ERIC (21,22)), (c) the reproducibility of staining pattern is typically achieved when the same panel of reagents is used (The ONE Study (25), HIPC (26), EuroFlow (32,33), and COG (31)), which allows definition of the manual analysis strategy and also automated analysis tools, and (d) the reproducibility of patterns including the staining intensity that allows analysis standardization and database-assisted comparison to previous cases in interlaboratory settings: EuroFlow (34) or The Canadian National Transplant Research Program that analyzed multiple immune cells in multiple centers by automated analysis (35).

In the hematology field, the EuroFlow consortium of eight European hematology laboratories was funded by the European Commission Program in 2006 and decided to develop an approach for standardized hematological diagnostics that would describe the complete preanalytical procedures and design analytical software that would integrate all data measured by the consortium and use this dataset to analyze and interpret any new cases (36). Thus, a standard operating protocol (SOP) was developed that would set up all cytometers in the consortium to the same intensity readout (using hard dyed beads as a standard) (37). Uniform sets of reagents (antibody panels) were assembled and tested together with the staining SOP (33), and software tools were built that enabled the analysis of the standardized data. Alternative reagents providing the same staining patterns were tested side-by-side with the original reagents, and the updated reagent list is publicly shared on the euroflow.org website. Staining intensity-based evaluation of the quality of the locally acquired data file in comparison with the expected variation has been built into a quality assurance scheme (1). An analysis of a dataset of 1,438 cases of acute leukemia was published by Lhermitte et al. (34) that utilized automated analysis tools applied to standardized data collected in 12 laboratories. Furthermore, the consortium developed a minimal residual disease detection panel for multiple myeloma (MM) (28) and B-cell precursor leukemia (BCP-ALL) (38) with corresponding data analysis tools that are based on multiparameter evaluations (39). The Swiss Cytometry Society (SCS) published a feasibility study in 10 clinical laboratories that documented the improvement in EuroFlow LST tube performance when proper feedback was given based on a local quality assurance program (40). A guide on setting up the hard dyed-bead target values on different cytometry platforms made by different manufacturers has been published and allows full intensity standardization (41).

The European Research Initiative on chronic lymphocytic leukemia (CLL) (ERIC, www.ericll.org) has been focused on improving the outcome of patients with CLL. As part of these efforts, ERIC developed and tested the performance of a consensus 4-color flow cytometry antibody panel for the detection of CLL minimal residual disease in 2007 (23). This approach was later extended to six colors with enhanced

sensitivity (21), and it was recently determined to be complementary to next-generation sequencing-based methods (22).

The Harmonemia project (supported by Beckman Coulter) focused on confirming that the interpretation, enumeration and staining patterns of subsets analyzed by EuroFlow lymphocytosis screening tube can be achieved using alternative reagents with the same outcome quality (42) and that different cytometers can yield data with similar patterns (17,43).

The Children's Oncology Group (COG) study developed a unified minimal residual disease detection approach (44) deployed in two centers for the COG study AALL0232 that used the same antibody panel, although on different cytometers, and demonstrated excellent interlaboratory enumeration correlation (31). Recently, COG developed a QA scheme and educational program for BCP-ALL MRD (45).

The Associazone Italiana Ematologia Oncologia Pediatrica (AIEOP) and the Berlin-Frankfurt-Munster (BFM) group have developed a common, standardized protocol for BCP-ALL MRD evaluation by four-color flow cytometry (19). This protocol was followed by an 8-color approach tested for stratification concordance with the PCR-based assessment (20). This group uses a range of diverse instruments, so the proposed panels offer some flexibility in antibody reagent choice. The 2016 consensus guidelines on immunophenotyping focused primarily on the delivery of a reproducible interpretation without the requirement to use particular antibody panels or instruments (although minimum requirements of markers were stated and general requirements of gating strategy were defined) (3). A proposal for a short and machine readable format of diagnostic summary reporting (FDE code) was presented (18).

In the immunology field, the expansion of the number of functional subsets revealed by expanding the number of available parameters led to a need for consensus on which set of markers could be used to define and thus enumerate different cell subsets (46,47). Multiple partially overlapping but unidentical definitions are used to define naïve T-cells (Fig. 1), transitional B cells, and dendritic cells. In some cases, the reasoning was based on confusion caused by poor performance of particular clones (as in the case of CCR7 detection (26)).

To resolve some of these conflicting issues, several collaborative studies were initiated. The Human Immunology Project Consortium (HIPC) (26) has comprehensively summarized the variables and challenges in the assessment of the cellular composition of immune cells. In 2012, they proposed five 8-color panels that defined T-cells, B-cells and innate cells with surface markers only and prepared commercially available lyophilized reagent cocktails. A further focus was aimed at data analysis (48), where automated population identification and naming was achieved with flowCL (27), and superiority of automated analysis over individual analyst gating was shown (14).

The ONE Study consortium aimed to describe the immune system after solid organ transplantation (25). They designed a panel of six tubes (seven to nine colors) and had those prepared as dried reagents formulated per single test. A detailed guide to their approach was published in 2016 (24). Later, they performed a single laboratory study reporting age and gender differences on a cohort of 98 healthy adults (49). The Canadian National Transplant Research Program (CNTRP) adopted The ONE study panels and performed a follow-up study using the dried reagents in which the CNTRP compared sample preparation variables (PBMC versus immediate vs. 24-h delayed staining) and developed and benchmarked an automated analysis approach (35).

The PRECISAIDS project published an initial study on standardization quality for a planned interlaboratory study using nine 8-color panels that reached very good MFI and cell subset enumeration variability (50).

For application in the immunology field, the EuroFlow consortium built six panels describing lymphocytes in 2012 with the aim of describing abnormalities in the immune system in primary immunodeficiency (PID). A fully standardized system including instrument setup, sample preparation and data analysis was built in a stepwise manner to aid in diagnostics and translational research in PID. This standardized system provides a first tier test of lymphocyte subsets (PID orientation tube; including naïve vs memory compartments) (51,52) and is a powerful approach that was confirmed to be applicable to PID with a large international cohort of 99 PID patients and 250 healthy controls (52). A further detailed panel that aims to dissect T-cell and B-cell subsets was assembled and used by the consortium (52). T-cell tubes, including naïve forms, and activation and definition of recent thymic emigrants provide a very sensitive and specific instrument to diagnose severe immunodeficiency (SCID, CID, including activated T-cells in Omenn syndrome cases and cases with maternal engraftment). B cells were further phenotyped by 8-color panels to describe PreGC and PostGC development and by 12-color panel revealing the IgH-isotype expressing cells (53,54). The same panel was used by the EuroFlow PID consortium to investigate primary antibody deficiency (55).

Summary of the studies conducted by above mentioned consortia is provided in Table 1.

What are the commonalities and disparities between clinical diagnostic and clinical research or monitoring studies? Apart from the different regulatory status, they all need to define the cells of interest, interpret (simplify) their identity from the immunophenotype and enumerate their proportion in the sample. Then, data analysis should be performed without any subjective and sample specific bias. While the diagnostic tests should use mature and validated methods, the clinical research studies may have multiple objectives, where particular cell subsets need to be enumerated reproducibly with a high priority, but other subsets may have rather limited priority and their evaluation is rather exploratory. Thus, building more extensive standardization is more appropriate for diagnostic use, while harmonization might offer more flexibility for clinical research studies.

When approached from the data analysis perspective, the diagnostic use requires reproducible analysis of a singular sample (with the frame of reference of large cohort of control samples), while clinical research studies need to be analyzed as a whole cohort without any bias, but with a possibility to react to unexpected new findings, to accommodate for added

Table 1. Overview of interlaboratory studies aiming at standardization

CONSORTIUM	STANDARDIZATION ASPECT	VALIDATION OR BENCHMARKING	REFERENCE
Hematology			
EuroFlow, www.euroflow.org	EuroFlow concept		(36)
	Standardization procedures		(37)
	Panels of reagents for leukemia diagnostics		(33)
	QA scheme	QA scheme	(1)
	Automated leukemia classification	652 reference cases	(34)
	Minimal residual disease for multiple myeloma	385 samples vs conventional flow and vs outcome	(28)
	Minimal residual disease for BCP-ALL	Enumeration performance benchmarked by NGS	(38)
	Adoption by the field	QA scheme	(40)
	Cross platform standardization	Canto, Navios, MACS Quant; QA scheme	(41)
ERIC, www.ericll.org	4-color panel for detection of CLL MRD		(23)
	6-color panel for detection of CLL MRD		(21)
	6-color panel for detection of CLL MRD	Enumeration performance benchmarked by NGS	(22)
Harmonemia	Panel for Lymphocytosis	Benchmarked by EuroFlow panel	(42)
	Cross platform standardization	FACS Canto vs Navios	(17,43)
Children's Oncology Group (COG)	MRD for BCP-ALL	inter-laboratory enumeration	(31)
	QA scheme	QA scheme	(45)
The AIEOP-BFM group	4-color panel MRD for BCP-ALL		(19)
	8-color panel MRD for BCP-ALL	Stratification performance benchmarked by PCR	(20)
	Immunophenotyping panels and diagnostic interpretations for pediatric ALL		(3)
Immunology			
Human Immunology Project, www.immuneprofiling.org	five 8-color panels		(26)
	Automated subset naming		(27)
	Automated gating	Benchmarked by manual analysis	(14)
The ONE Study	Panel of six tubes (7 to 9-color)	Inter-lab, intra-assay, inter-operator	(25) (24)
CNTRP (One Study panels)	Multicenter study of inter-center variability, subset enumeration variability, automated data analysis	Automated vs manual, fresh vs stored viable samples	(35)
PRECISAIDS	nine 8-color panels, 11 sites, 7 instrument types	Inter-laboratory test, enumeration and MFI	(50)
EuroFlow	8-color panel for lymphocytes	Performance in 99 PID patients (250 controls)	(52)
	12-color B cell phenotype	234 healthy controls	(53)
	12-color B cell phenotype	139 primary antibody deficiency	(55)

markers in further research stages to improve cell definitions. Example of the former approach is the EuroFlow database-guided diagnosis and classification of acute leukemia (34), while the latter is documented by HIPC (14) and CNTRP (35).

How Good Is Good Enough?

A key initial parameter of any reproducibility effort is the question: How good is good enough? Validation and performance of laboratory tests uses terms accuracy, trueness, and precision (Box 1) (56). Definition of the lower border of good

reproducibility should take into account the results of quality assessment (QA) studies. There is a summary of experiences with UKNEQAS (57,58) and Dutch SIHON (59). However, both groups measured consensus on the interpretation (also in a per marker fashion) and enumeration. The general message of CD34 ISHAGE guidelines and their deployment is that regardless of instrument and reagent manufacturer, the enumeration reproducibility can be as high as CV <10,8%, when an appropriate choice of bright reagents is made for CD34 and a proper gating scheme is followed (60). This is similar to CV < 13.7% for CD4 T cell counting (61). At the same time, it also summarized evidence indicating that adherence to good practices and guidelines is frequently poor in the large field, so education is a crucial part of improvement for the future (16,62). A recent French study evaluated the real-life intra-assay repeatability of enumeration of lymphocyte subsets and found that frequent subsets (T-lymphocytes) can be quantitated with a CV of 1% (T-lymphocytes), while other subsets were less precise (NK cells with a CV of 4.78%) (63). Note that, precise enumeration not only requires that sufficient numbers of events are collected, but it also requires assay validation with a set of controls (64). This gap between interlaboratory reproducibility and intra-assay repeatability might be closed by standardization. The consensus recommendations by ICSH and ICCS for intra-assay repeatability is that desirable target is a CV of less than 10%, but for less abundant populations 20% can be acceptable (10).

Reproducibility of patterns was not evaluated in the literature systematically, however studies building automated analysis tools quote technical variability as likely reason for failure to recognize biological differences (14,35).

When evaluating intensity, only EuroFlow QA data are available for a single panel (1,65). The Lymphocytosis Screening Tube was chosen as a QA model since it uses conservatively expressed molecules in most channels (CD4, CD3, CD8, CD19, CD20, CD45, CD5, and CD81) and allows the gating of a subset that is clearly positive for each of them, and its MFI could be compared across the whole dataset of 123 samples from 11 laboratories. A principal component analysis projection resolves all subsets well in a distinct space of that projection. The coefficient of variation of MFI was as low as 10.9% and as high as 52.3% for the above-named molecules in individual years (27 to 33 samples per year) in the standardized yet interlaboratory settings. These MFI CVs are well comparable to a PRECISESADS study, that used single sample measured in 11 different laboratories with standardized approach (50). For less consistent molecules (kappa and lambda light chains), the CV was in the range of 43–189%. Translated into a visual difference, this means one can expect the median fluorescence intensity of the molecules in the group to be within half of a decade on a dotplot in any laboratory, during any year, for any instrument and even with similarly performing alternative reagents.

Importantly, the most appropriate measurable indicator of reproducibility outcome should be identified for each type of study (i.e., interpretation, enumeration, and staining pattern), and any approach should be evaluated against those criteria in a postimplementation study or in a quality assurance scheme. Any level of precision achieved should be interpreted in relation to the purpose of each flow cytometry assay. In addition, individual measurements (their normal or abnormal values) should be interpreted in relation to the general performance of the assay in question.

Box 1

Accuracy is closeness of agreement between a quantity value obtained by measurement and the true value of the measurand.

Trueness is defined as the closeness of agreement between the average value obtained from a large series of test results and an accepted reference value.

Precision is the dispersion of replicate measurements, assessed intra-assay (“***repeatability***”) or inter-assay (interlaboratory: “***reproducibility***”). Precision is quantified by standard deviation and coefficient of variation (CV).

How Reproducibility Can Be Achieved Between Projects, Between Instruments and Between Reagents?

The collaborative studies experienced that multiple hurdles remain in the way of standardization efforts. Those can be summarized as the interplay between these factors: purpose, resources, technology and training.

The purpose of each collaborative study is different in its ambition, intended complexity, concerns on applicability in a given field and time, demand for fast or long term solutions, size of the group and its heterogeneity, and intended regulatory status.

Resources limit each of those studies financially (e.g., how large the benchmarking or comparison studies can be and whether there is an infrastructure for data sharing and computational analysis) and physically (e.g., how much manpower is available, what instrumentation is available, and how many samples can be processed).

Technology limits studies in the context of diversity of instruments and their ability to measure particular fluorochrome combinations with intended quality.

For the purpose of this review, only the technological limitations will be discussed, as those can be overcome by the efforts of the current and future investigators in collaboration with the industry, and consequently, the resource and purpose issues can be redefined when needed.

Hardware

One frequently voiced objection is that flow cytometry instrumentation differences pose a major obstacle to full standardization. However, when an 8-color panel is designed for a three laser instrument, it can be measured on any of the 13 instruments from nine manufacturers (41). A similar conclusion was reached by the Harmonemia group (43) and the

CNTRP group (35). It is thus feasible to reach reproducibility of interpretation and enumeration of different instrument types with the same panel, with a caveat that some differences in sensitivity to weak expression can be introduced by small differences in excitation wavelength and emission filter specifications or other hardware differences (detector type). The feasibility of a multi-instrument interlaboratory setup was documented by EuroFlow (Canto II, LSR II, and Cyan ADP) (37), COG (Canto and LSR II instrument) (31), and Harmonemia (Canto II and Navios instrument) (42). It was argued by Solly et al. (43), that a setup of Canto II and Navios instruments with Rainbow beads produces comparable data for cells stained with CD16 Pacific Blue after a simple rescaling to accommodate the differences in the parameter scaling. However, the differences can be more prominent in channels where a larger difference in filter wavelength specification exists. In those cases, adjustment of the Rainbow target values must be performed for each fluorochrome and filter pair, as recently demonstrated by Novakova for Canto II and Navios instruments (41). This results in a full standardization down to the same pattern and intensity. However, all of those attempts rely on two crucial assumptions: there is a linearity of measurement, and there is no difference in sensitivity to a low signal. Also, unequal measurement noise of different instrument types will impact on the low end of the scale. To reach similar resolution, CNTRP has adopted an interesting approach to reach the same Stain index settings for two differently designed instruments (35). It remains to be seen how similar the patterns can be between instruments working on a different photon detection principle (PMT detector, semiconductor detector arrays "avalanche diode"). Methods for the objective comparison of the capability of instruments to perform fluorescence measurements are being developed and will be necessary for making assumptions about the "similar" performance of diverse cytometers (66).

Sample preparation method

Another factor that can influence the standardization of results is a sample preparation method. This is critical primarily for the standardization of enumeration because cell loss is an inevitable consequence of suspension processing that involves centrifugation. The total number of cells lost during sample prep differed by a factor of two in the test performed by EuroFlow, while the intensity of individual fluorochromes decreased only by up to 25% (37). Another challenging variable is cell type-specific ex vivo longevity. Plasmablasts, for instance, have significantly reduced longevity when sample processing is delayed by 24 h (35). Fluorophores are sensitive to light exposure (67), fixatives and even elements such as copper (in the case of quantum dots) (68). Thus, sample preparation methods must be evaluated for either a loss of target cells or a loss of fluorescence signal due to either epitope loss, damage, masking or fluorophore destruction. An overview of the variables that require optimization particularly in intracellular cytokine staining was presented by Nomura et al (69). Careful assay validation (6–10), appropriate staff training and external quality assessment are the tools that need to be employed to reduce sample preparation validation. Unfortunately, a protocol nonadherence is a frequent cause for failure in the QA (16,65).

Analytical specificity

The first keystone of flow cytometry reproducibility is the choice of proper definition markers for a given subset of interests ("Analytical specificity" (10)). To this end, international consensus guidelines were issued for leukemia and lymphoma immunophenotyping (70), and proposal seeking unification of immune cell definition for flow cytometry monitoring was made by HIPC (26). In 2010, the new publication format, Optimized Multicolor Immunofluorescence Panel (OMIP), was introduced in Cytometry Part A (71). The OMIP format facilitates the exchange of knowledge on the optimal detection of particular cell subsets and their surface or intracellular proteins and importantly presents the optimization experiments to the field. Up to July 2019, 59 OMIPs were published, describing a focused cell subset (e.g., dendritic cells (72)) or broad leukocyte profiling with 21-parameter flow cytometry (73) or 26-parameter mass cytometry (74). All consortia mentioned in the section on reproducibility in the large-scale studies invested significant effort into the selection and performance testing of the optimal antibody panels.

Reagents

An additional critical component of the standardization procedure is the antibody conjugate. Misleading results can be obtained for several reasons: Antibody clones do not recognize their intended target or bind to other targets at the same time. Cluster of differentiation (CD nomenclature) was therefore developed based on independent testing by HLDA workshops to provide clear consensual data on antibody clones' reactivities (75). CD number designation is used for both, the target protein and antibody clones recognizing that target. Nevertheless, controversies may result from nonuniform data (e.g., shedding/recirculation of surface proteins under some conditions or epitope masking) (26,76) or can be caused by different antibodies belonging to the same CD but recognizing different epitopes. Notably, some epitopes are decreased or lost upon cryopreservation (e.g., in CD62L) (35); in other cases, cell fixation can decrease epitope availability for particular antibody clones (77,78), resulting in alteration in the measured signal intensity. Accompanying article in this Special issue is dedicated to antibody validation (79).

Differences in the resolution of dimly stained cells between different fluorochromes are another possible source of disparities. Thus, interpretation of dimly positive or negative staining can be achieved when a bright or dim fluorochrome is used on the same molecule and the same cell, respectively (11,80). This could also hamper the enumeration of cells with low expression of the marker, and it will certainly affect the overall staining pattern.

Prerequisite for reproducible intensity of staining is the use of antibody conjugates that are produced with no substantial lot-to-lot differences. Unfortunately, there is no data

on the stringency of manufacturers for this parameter thus far. Recently, Böttcher et al. performed a pilot study on 1,323 consecutive antibody lots of 157 different monoclonal antibodies (81), and he found that while in general the variability is very low (median CV of 3.8%), 8,8% of reagents exceeded 20% (including 3.6% of reagents that exceeded 30% lot-to-lot intensity difference). When stringency criteria of 20% were applied, different fluorochromes performed with different failure rates (0% to 37.5%), and he did observe diverse failure rates in different manufacturers. The method used (analysis of old vs new lot in parallel on the antibody capture bead) cannot resolve a production problem from a stability problem and does not test for binding to the epitope. This study shows that antibody reagents can be made very reproducibly (91.2% reagents less than 20% different) and thus should not be a major source of variability. Declaration of particular staining intensity performance criteria by the manufacturer would guide the choice of reagents for standardized assays.

EuroFlow has also tested all reagents accepted later as alternative reagents for use with the standardized approach for both staining pattern and intensity of fluorescence on normal as well as abnormal (leukemic) cells, and the results indicated that most reagents yielded the same pattern and intensity even when different clones were used and different manufacturers produced them. However, nothing can replace a side-by-side test on the intended sample type with a standardized protocol in the assessment whether two reagents provide the same staining pattern and intensity. This is also evidenced in the EuroFlow QA assessment, where one of the reasons for QA failure is the use of locally chosen reagents that were never compared to the original reagents and thus were never listed as a EuroFlow-tested alternative reagent (65).

Of course, pipetting accuracy and reagent shelf stability influence the staining pattern and intensity. One possible solution is the use of stabilized reagent mixtures that not only provide ease of use but also limit room for pipetting errors (14,51,82). These are available either as commercially supplied ready to use mixtures or dried reagents or can be custom made by several manufacturers.

Data analysis

Data analysis plays a major role in the reproducibility of any flow cytometry study. Any cytometry study should be published with a full explanation of the gating strategy as described in the "The minimum information about a flow cytometry experiment" (MIFlowCyt) (83), which summarizes that a complete set of information about a published cytometry experiment is mandatory for Cytometry Part A journal submissions.

Automated analysis with software tools was shown to bring more objectivity into the gating procedure and thus increase the reproducibility of gating and enumeration (12,14,35). Indeed, one of the prerequisites for automated analyses is high-quality data that is fully and correctly annotated without ad hoc changes and without any accidentally missing reagents.

Standardized instrument setup has facilitated the interlaboratory comparisons. EuroFlow has thus far collected over 6,000 anonymized FCS files acquired in the standardized fashion on its internal database server. This enables automated, database driven comparisons of phenotypes and grouping of samples with similar features (34) as well as using the stored data to aid in the interpretation of a newly acquired case.

HIPC has used a dataset acquired in standardized interlaboratory settings, using lyophilized reagents on stored (fixed or cryopreserved) PBMCs for an automated data analysis (14). They evaluated two series of replicates and statistically evaluated performance of manual, central manual and automated analysis. They found a lower coefficient of variation in 20 of 21 subsets in central analysis. Power analysis allowed them to identify subsets where automated centralized analysis performed better then local gating. Not surprisingly, poorly resolved populations and protocol non-adherence was found to be a major contributor to variable results.

CNTRP has designed a successful automated analysis pipeline for The ONE study panels (35). Again, subsets that were poorly resolved by gating markers yielded higher variability in automated detection compared to clearly separated clusters of cells.

How Would Standardization Develop in the Near Future?

This review documents that the room for consensus is much less constricted by the instrumentation and reagents than previously claimed. Furthermore, to close the gaps between different consensus studies, benchmarking one approach by using alternative approaches is necessary for future progress, the reproducibility of interpretation and enumeration (42) and perhaps even for evaluating staining patterns.

All of the above-named large studies serve as benchmarks of their kind. It would be extremely valuable to the field if each new panel proposal that aspires to serve as a benchmark and potentially a new consensus would perform a side-by-side benchmarking comparison to existing large studies and document the differences obtained. Additionally, a deposition of at least a small cohort of well-chosen example files (anonymized but annotated as controls and disease) to a public domain (FlowRepository or similar) would allow further virtual comparisons by data analysis tools.

Standardization will be facilitated with several types of pressure, where data analysis pipeline developers will push for consistent (standardized) data for successful implementation and a regulatory pressure on IVD diagnostics (EU Medical Device 2017/746 effective from 2022) will ask higher level of stringency even for Laboratory Devised Tests in clinical diagnostics.

The technical advancements (automated solutions for portability of assay setup from instrument to instrument) will enable deployment of the standardization. Since most currently manufactured clinical and mid-range cytometers can detect at least 8 common fluorochromes, standardization across different instruments is also feasible (41). Further

development of multicolor cytometry studies poses also logistical and sample preparation challenges. For large studies spanning long time periods, it is essential that all reagents needed for a panel are available in the laboratory stock, are not compromised by shelf life and are pipetted to each sample exactly as required by the SOP. This is not a simple logistical task, and it is error prone. Thus, the availability of stable reagent mixes on the market that contain an entire panel or allow at least a building block approach to a panel design (backbone reagents in the mix and few reagents added individually) would help to reduce the design and performance variability. This benefit is documented by some of the newer technologies, for example, mass cytometry (84). Reagent sets can be supplied in blocs, thus allowing analysis tools to be tailored to a given panel (85) and thus allow for direct comparability of the resulting immunophenotype patterns between projects and between laboratories. Paradoxically, progress in this direction is slowed by the requirements of national regulatory bodies that require formal clinical studies for antibody reagent mixtures to declare them fit for diagnostic purpose, resulting in the virtual absence of commercially available reagent mixtures. However, where legally allowed, the on-demand antibody reagent mixtures or their dried alternatives might contribute reduction of the variation of the signal intensity (as exemplified by PIDOT and LST tubes (51)). Fortunately for the field, EuroFlow, HIPC, and The ONE study panels are developing into complete and standardized solutions (from instrument setup to reagent panels to data analysis) manufactured by established vendors.

Thus, in the future, standardized and benchmarked immunostaining panels might be more often supplied as ready-made reagent mixtures (either as a catalog items or as an on demand made products). This might simplify sample preparation and reduce errors. Ideally, the demand for reproducibility would also motivate reagent manufacturers to develop and declare intensity based quality controls of the antibody conjugates (79). Reduction of technical errors together with the broader use of a particular immunostaining panel might motivate the development of data analysis tools, that will be more accessible also to non-programmers.

High parameter cytometry (>20) allows for benchmarking alternative definitions of cell subsets together with cell classification and phenotype reporting tools (27,86,87). These approaches have the potential to reduce the current cell subset nomenclature issues. An interesting approach to learn cell identity from flow cytometry datasets was developed by the Irish group, called "Marker Enrichment Modeling (MEM) (87,88). This kind of aggregated immunophenotype analysis approach could potentially bring more benchmarking options into comparisons of different immunophenotyping approaches for the same cell subsets in a high-content manner by reanalyses of various existing datasets.

Conclusions

Standardized approaches are already technically feasible as documented by the studies listed above. Within one interlaboratory study, achieving reproducible results is possible. Is there any chance that different groups and studies would build on each other's advancements and eventually converge? Thus far, the greatest challenge to achieving reproducible results seems to be our ability to reach a consensus and eventually adhere to that consensus. This can be understood to some extent since any consensus is by definition not the most up-to-date solution because further development is achieved in any field before a consensus is reached and tested. A concern has been raised that adherence to a standardized assay practically blocks any further innovation. Additionally, a divergence in the purpose and available resources limit the consensual use of standardized procedures. It is therefore likely that early phases of the interlaboratory studies will be less standardized (or rather harmonized), while later, more mature stages will be standardized.

In conclusion, standardized approaches offer not only the benefit of consensual solutions that can be benchmarked and evaluated in diverse applications but also the structured education of users and quality assessment tools. Furthermore, a fully standardized solution to the level of patterns and intensity lends itself to automated analyses by academically designed algorithms or by commercial software solutions. All of these factors combined will contribute to the enhanced interlaboratory reproducibility of the cytometry studies.

Acknowledgments

TK is a founding member of EuroFlow consortium. Author wishes to thank to Ondrej Hrusak, Ester Mejstrikova and Ruud Hulspas for fruitful discussions. TK has been supported by project of Ministry of Education, Youth and Sports, Czech Republic nr.LO1604 and by EU-Prague project CZ.2.16/3.1.00/24505 and CZ.2.16/3.1.00/21540.

Literature Cited

1. Kalina T, Flores-Montero J, Lecrevisse Q, Pedreira CE, van der Velden VHJ, Novakova M, Mejstrikova E, Hrusak O, Böttcher S, Karsch D, et al. Quality assessment program for EuroFlow protocols: Summary results of four-year (2010–2013) quality assurance rounds. Cytometry A 2015;87:145–156.
2. Maecker HT, McCoy JP, Amos M, Elliott J, Gaigalas A, Wang L, Aranda R, Banchereau J, Boshoff C, Braun J, et al. A model for harmonizing flow cytometry in clinical trials. Nat Immunol 2010;11:975–978.
3. Dworzak MN, Buldini B, Gaipa G, Ratei R, Hrusak O, Luria D, Rosenthal E, Bourquin J, Sartor M, Schumich A, et al. AIEOP-BFM consensus guidelines 2016 for flow cytometric immunophenotyping of pediatric acute lymphoblastic leukemia. Cytometry B Clin Cytom 2018;93:82–93.
4. Holzelova E, Vonarbourg C, Stolzenberg M-C, Arkwright PD, Selz F, Prieur A-M, Blanche S, Bartunkova J, Vilmer E, Fischer A, et al. Autoimmune lymphoproliferative syndrome with somatic *Fas* mutations. N Engl J Med 2004;351: 1409–1418.
5. Stuchlý J, Kanderová V, Vlková M, Heřmanová I, Slámová L, Pelák O, Taraldsrud E, Jílek D, Králíčková P, Fevang B, et al. Common variable immunodeficiency patients with a phenotypic profile of immunosenescence present with thrombocytopenia. Sci Rep 2017;7:39710.
6. Davis BH, Wood B, Oldaker T, Barnett D. Validation of cell-based fluorescence assays: Practice guidelines from the ICSH and ICCS—part i—rationale and aims. Cytometry B Clin Cytom 2013;84:282–285.
7. Davis BH, Dasgupta A, Kussick S, Han JY, Estrellado A. Validation of cell-based fluorescence assays: Practice guidelines from the ICSH and ICCS—part II—Preanalytical issues. Cytometry B Clin Cytom 2013;84:286–290.
8. Tanqri S, Vall H, Kaplan D, Hoffman B, Purvis N, Porwit A, Hunsberger B, Shankey TV. Validation of cell-based fluorescence assays: Practice guidelines from the ICSH and ICCS—part III—analytical issues. Cytometry B Clin Cytom 2013;84: 291–308.
9. Barnett D, Louzao R, Gambell P, De J, Oldaker T, Hanson CA. Validation of cell-based fluorescence assays: Practice guidelines from the ICSH and ICCS—part IV—postanalytic considerations. Cytometry B Clin Cytom 2013;84:309–314.

10. Wood B, Jevremovic D, Béné MC, Yan M, Jacobs P, Litwin V. Validation of cell-based fluorescence assays: Practice guidelines from the ICSH and ICCS—part V—assay performance criteria. Cytometry B Clin Cytom 2013;84:315–323.
11. Cossarizza A, Chang H-D, Radbruch A, Akdis M, Andrä I, Annunziato F, Bacher P, Barnaba V, Battistini L, Bauer WM, et al. Guidelines for the use of flow cytometry and cell sorting in immunological studies. Eur J Immunol 2017;47:1584–1797.
12. Kvistborg P, Gouttefangeas C, Aghaeepour N, Cazaly A, Chattopadhyay PK, Chan C, Eckl J, Finak G, Hadrup SR, Maecker HT, et al. Thinking outside the gate: Single-cell assessments in multiple dimensions. Immunity 2015;42:591–592.
13. Mcneil LK, Price L, Britten CM, Jaimes M, Maecker H, Odunsi K, Matsuzaki J, Staats JS, Thorpe J, Yuan J, 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.
14. Finak G, Langweiler M, Jaimes M, Malek M, Taghiyar J, Korin Y, Raddassi K, Devine L, Obermoser G, Pekalski ML, et al. Standardizing flow cytometry immunophenotyping analysis from the Human ImmunoPhenotyping Consortium. Sci Rep 2016;6:20686.
15. Levering WHBM, Preijers FWMB, van Wieringen WN, Kraan J, van Beers WAM, Sintnicolaas K, van Rhenen DJ, Gratama JW. Flow cytometric CD34+ stem cell enumeration: Lessons from nine years' external quality assessment within the Benelux countries. Cytometry B Clin Cytom 2007;72:178–188.
16. Whitby A, Whitby L, Fletcher M, Reilly JT, Sutherland DR, Keeney M, Barnett D. ISHAGE protocol: Are we doing it correctly? Cytometry B Clin Cytom 2012; 82B:9–17.
17. Lacombe F, Bernal E, Bloxham D, Couzens S, Porta MGD, Johansson U, Kern W, Macey M, Matthes T, Morilla R, et al. Harmonemia: a universal strategy for flow cytometry immunophenotyping-A European LeukemiaNet WP10 study. Leukemia 2016;30:1762–1772.
18. Hrušák O, Basso G, Ratei R, Gaipa G, Luria D, Mejstříková E, Karawajew L, Buldini B, Rozenthal E, Bourquin JP, et al. Flow diagnostics essential code: a simple and brief format for the summary of leukemia phenotyping. Cytometry B Clin Cytom 2014;86:288–291.
19. Dworzak MN, Gaipa G, Ratei R, Veltroni M, Schumich A, Maglia O, Karawajew L, Benetello A, Pötschger U, Husak Z, et al. Standardization of flow cytometric minimal residual disease evaluation in acute lymphoblastic leukemia: Multicentric assessment is feasible. Cytometry B Clin Cytom 2008;74:331–340.
20. Karawajew L, Dworzak M, Ratei R, Rhein P, Gaipa G, Buldini B, Basso G, Hrusak O, Ludwig W-D, Henze G, et al. Minimal residual disease analysis by eight-color flow cytometry in relapsed childhood acute lymphoblastic leukemia. Haematologica 2015;100:935–944.
21. Rawstron A, Böttcher S, Letestu R, Villamor N, Fazi C, Kartsios H, de Tute RM, Shingles J, Ritgen M, Moreno C, et al. Improving efficiency and sensitivity: European Research Initiative in CLL (ERIC) update on the international harmonised approach for flow cytometric residual disease monitoring in CLL. Leukemia 2013; 27:142–149.
22. Rawstron AC, Fazi C, Agathangelidis A, Villamor N, Letestu R, Nomdedeu J, Palacio C, Stehlikova O, Kreuzer K-A, Liptrot S, 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;30:929–936.
23. Rawstron AC, Villamor N, Ritgen M, Böttcher S, Ghia P, Zehnder JL, Lozanski G, Colomer D, Moreno C, Geuna M, et al. International standardized approach for flow cytometric residual disease monitoring in chronic lymphocytic leukaemia. Leukemia 2007;21:956–964.
24. Schlickeiser S, Streitz M, Sawitzki B. Standardized multi-color flow cytometry and computational biomarker discovery. Methods Mol Biol 2016;1371:225–238.
25. Streitz M, Miloud T, Kapinsky M, Reed MR, Magari R, Geissler EK, Hutchinson JA, Vogt K, Schlickeiser S, Kverneland A, et al. Standardization of whole blood immune phenotype monitoring for clinical trials: Panels and methods from the ONE study. Transplant Res 2013;2:17.
26. Maecker HT, McCoy JP, Nussenblatt R. Standardizing immunophenotyping for the human immunology project. Nat Rev Immunol 2012;12:191–200.
27. Courtot M, Meskas J, Diehl AD, Droumeva R, Gottardo R, Jalali A, Taghiyar MJ, Maecker HT, McCoy JP, Ruttenberg A, et al. flowCL: Ontology-based cell population labelling in flow cytometry. Bioinformatics 2015;31:1337–1339.
28. Flores-Montero J, Sanoja-Flores L, Paiva B, Puig N, García-Sánchez O, Böttcher S, van der Velden VHJ, Pérez-Morán J-J, Vidriales M-B, García-Sanz R, et al. Next generation flow for highly sensitive and standardized detection of minimal residual disease in multiple myeloma. Leukemia 2017;31:2094–2103.
29. van Dongen JJM, van der Velden VHJ, Brüggemann M, Orfao A. Minimal residual disease diagnostics in acute lymphoblastic leukemia: Need for sensitive, fast, and standardized technologies. Blood 2015;125:3996–4009.
30. Theunissen P, Mejstrikova E, Sedek L, van der Sluijs-Gelling AJ, Gaipa G, Bartels M, Sobral da Costa E, Kotrova M, Novakova M, Sonneveld E, et al. Standardized flow cytometry for highly sensitive MRD measurements in B-cell acute lymphoblastic leukemia. Blood 2017;129:347–357.
31. Borowitz MJ, Wood BL, Devidas M, Loh ML, Raetz EA, Salzer WL, Nachman JB, Carroll AJ, Heerema NA, Gastier-Foster JM, et al. Prognostic significance of minimal residual disease in high risk B-ALL: a report from Children's Oncology Group study AALL0232. Blood 2015;126:964–971.
32. van Dongen JJM, van der Burg M, Kalina T, Perez-Andres M, Mejstrikova E, Vlkova M, Lopez-Granados E, Wentink M, Kienzler A-K, Philippé J, et al. EuroFlow-based flowcytometric diagnostic screening and classification of primary immunodeficiencies of the lymphoid system. Front Immunol 2019; 10:1–21.
33. van Dongen JJM, Lhermitte L, Böttcher S, Almeida J, van der Velden VHJ, Flores-Montero J, Rawstron A, Asnafi V, Lécrevisse Q, Lucio P, et al. EuroFlow antibody panels for standardized n-dimensional flow cytometric immunophenotyping of normal, reactive and malignant leukocytes. Leukemia 2012;26:1908–1975.
34. Lhermitte L, Mejstrikova E, Van Der Sluijs-Gelling AJ, Grigore GE, Sedek L, Bras AE, Gaipa G, Sobral Da Costa E, Novakova M, Sonneveld E, et al. Automated database-guided expert-supervised orientation for immunophenotypic diagnosis and classification of acute leukemia. Leukemia 2018;32:874–881.
35. Ivison S, Malek M, Garcia RV, Broady R, Halpin A, Richaud M, Brant RF, Wang S-I, Goupil M, Guan Q, et al. A standardized immune phenotyping and automated data analysis platform for multicenter biomarker studies. JCI Insight 2018;3:e121867.
36. van Dongen JJM, Orfao A. EuroFlow: Resetting leukemia and lymphoma immunophenotyping. Basis for companion diagnostics and personalized medicine. Leukemia 2012;26:1899–1907.
37. Kalina T, Flores-Montero J, van der Velden VHJ, Martin-Ayuso M, Böttcher S, Ritgen M, Almeida J, Lhermitte L, Asnafi V, Mendonça A, et al. EuroFlow standardization of flow cytometer instrument settings and immunophenotyping protocols. Leukemia 2012;26:1986–2010.
38. Theunissen P, Mejstrikova E, Sedek L, Van Der Sluijs-Gelling AJ, Gaipa G, Bartels M, Sobral da Costa E, Kotrová M, Novakova M, Sonneveld E, et al. Standardized flow cytometry for highly sensitive MRD measurements in B-cell acute lymphoblastic leukemia. Blood 2017;129:347–357.
39. Pedreira CE, Costa ES, Arroyo ME, Almeida J, Orfao A. A multidimensional classification approach for the automated analysis of flow cytometry data. IEEE Trans Biomed Eng 2008;55:1155–1162.
40. Glier H, Heijnen I, Hauwel M, Dirks J, Quarroz S, Lehmann T, Rovo A, Arn K, Matthes T, Hogan C, et al. Standardization of 8-color flow cytometry across different flow cytometer instruments: A feasibility study in clinical laboratories in Switzerland. J Immunol Methods 2017. https://doi.org/10.1016/j.jim.2017.07.013.
41. Nováková M, Glier H, Brdičková N, Vlková M, Santos AH, Lima M, Roussel M, Flores-Montero J, Szczepanski T, Böttcher S, et al. How to make usage of the standardized EuroFlow 8-color protocols possible for instruments of different manufacturers. J Immunol Methods 2017. https://doi.org/10.1016/j.jim.2017.11.007.
42. Preijers FWMB, Huys E, Favre C, Moshaver B. Establishment of harmonization in immunophenotyping: A comparative study of a standardized one-tube lymphocyte-screening panel. Cytometry B Clin Cytom 2014;86:418–425.
43. Solly F, Rigollet L, Baseggio L, Guy J, Borgeot J, Guérin E, Debliquis A, Drenou B, Campos L, Lacombe F, et al. Comparable flow cytometry data can be obtained with two types of instruments, canto II, and navios. A GEIL study. Cytometry A 2013;83: 1066–1072.
44. Borowitz MJ, Devidas M, Hunger SP, Bowman WP, Carroll AJ, Carroll WL, Linda S, Martin PL, Pullen DJ, Viswanatha D, et al. Clinical significance of minimal residual disease in childhood acute lymphoblastic leukemia and its relationship to other prognostic factors : A Children's Oncology Group study. Blood 2008;111: 5477–5485.
45. Keeney M, Wood BL, Hedley BD, DiGiuseppe JA, Stetler-Stevenson M, Paietta E, Lozanski G, Seegmiller AC, Greig BW, Shaver AC, et al. A QA program for MRD testing demonstrates that systematic education can reduce discordance among experienced interpreters. Cytometry B Clin Cytom 2018;94:239–249.
46. Appay V, van Lier RAW, Sallusto F, Roederer M. Phenotype and function of human T lymphocyte subsets: Consensus and issues. Cytometry A 2008;73:975–983.
47. 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;43:2797–2809.
48. Brusic V, Gottardo R, Kleinstein SH, Davis MM, Committee H steering, Davis MM, Hafler DA, Quill H, Palucka AK, Poland GA, et al. Computational resources for high-dimensional immune analysis from the human immunology project consortium. Nat Biotechnol 2014;32:146–148.
49. Kverneland AH, Streitz M, Geissler E, Hutchinson J, Vogt K, Bo??s D, Niemann N, Pedersen AE, Schlickeiser S, Sawitzki B. Age and gender leucocytes variances and references values generated using the standardized ONE-study protocol. Cytometry A 2016;89:543–564.
50. Jamin C, Le Lann L, Alvarez-Errico D, Barbarroja N, Cantaert T, Ducreux J, Dufour AM, Gerl V, Kniesch K, Neves E, et al. Multi-center harmonization of flow cytometers in the context of the European "PRECISESADS" project. Autoimmun Rev 2016;15:1038–1045.
51. van der Velden VHJ, Flores-Montero J, Perez-Andres M, Martin-Ayuso M, Crespo O, Blanco E, Kalina T, Philippé J, Bonroy C, de Bie M, et al. Optimization and testing of dried antibody tube: The EuroFlow LST and PIDOT tubes as examples. J Immunol Methods 2017. https://doi.org/10.1016/j.jim.2017.03.011.
52. van der Burg M, Kalina T, Perez-Andres M, Vlkova M, Lopez-Granados E, Blanco E, Bonroy C, Sousa AE, Kienzler A-K, Wentink M, et al. The EuroFlow PID orientation tube for flow cytometric diagnostic screening of primary immunodeficiencies of the lymphoid system. Front Immunol 2019;10:246.
53. Blanco E, Pérez-Andrés M, Arriba-Méndez S, Contreras-Sanfeliciano T, Criado I, Pelak O, Serra-Caetano A, Romero A, Puig N, Remesal A, et al. Age-associated distribution of normal B-cell and plasma cell subsets in peripheral blood. J Allergy Clin Immunol 2018;141:2208–2219.e16.
54. Blanco E, Perez-Andres M, Sanoja-Flores L, Wentink M, Pelak O, Martín-Ayuso M, Grigore G, Torres-Canizales J, López-Granados E, Kalina T, et al. Selection and validation of antibody clones against IgG and IgA subclasses in switched memory B-cells and plasma cells. J Immunol Methods 2017. https://doi.org/10.1016/j.jim.2017.09.008.
55. Blanco E, Pérez-Andrés M, Arriba-Méndez S, Serrano C, Criado I, Del Pino-Molina L, Silva S, Madruga I, Bakardjieva M, Martins C, et al. EuroFlow PID group. Defects in memory B-cell and plasma cell subsets expressing different immunoglobulin-subclasses in CVID and Ig-subclass deficiencies. J Allergy Clin Immunol 2019;144:809–824.

56. Menditto A, Patriarca M, Magnusson B. Understanding the meaning of accuracy, trueness and precision. Accredit Qual Assur 2007;12:45–47.
57. Bainbridge J, Rountree W, Louzao R, Wong J, Whitby L, Denny TN, Barnett D. Laboratory accuracy improvement in the UKNEQAS leucocyte Immunophenotyping immune monitoring program: An eleven-year review via longitudinal mixed effects modeling. Cytometry B Clin Cytom 2018;94:250–256.
58. Reilly JT, Barnett D. UKNEQAS for leucocyte immunophenotyping: The first 10 years. J Clin Pathol 2001;54:508–511.
59. Preijers FWMB, Van Der Velden VHJ, Preijers T, Brooimans RA, Marijt E, Homburg C, Van Montfort K, Gratama JW. Fifteen years of external quality assessment in leukemia/lymphoma immunophenotyping in The Netherlands and Belgium: A way forward. Cytometry B Clin Cytom 2016;90:267–278.
60. Barnett D, Granger V, Kraan J, Whitby L, Reilly JT, Papa S, Gratama JW. Reduction of intra- and interlaboratory variation in CD34+ stem cell enumeration using stable test material, standard protocols and targeted training. DK34 Task Force of the European Working Group of Clinical Cell Analysis (EWGCCA). Br J Haematol 2000;108:784–792.
61. Barnett D, Granger V, Whitby L, Storie I, Reilly JT. Absolute CD4+ T-lymphocyte and CD34+ stem cell counts by single-platform flow cytometry: The way forward. Br J Haematol 1999;106:1059–1062.
62. Whitby L, Granger V, Storie I, Goodfellow K, Sawle A, Reilly JT, Barnett D. Quality control of CD4+ T-lymphocyte enumeration: Results from the last 9 years of the United Kingdom National External Quality Assessment Scheme for immune monitoring (1993–2001). Cytometry 2002;50:102–110.
63. Ticchioni M, Brouzes C, Durrieu F, Lambert C. Acceptable "real-life" variability for lymphocyte counts by flow cytometry. Cytometry B Clin Cytom 2019;96B:379–388.
64. Roederer M. How many events is enough? Are you positive? Cytometry A 2008; 73A:384–385.
65. Kalina T, Brdickova N, Glier H, Fernandez P, Bitter M, Flores-Montero J, van Dongen JJM, Orfao A. Frequent issues and lessons learned from EuroFlow QA. J Immunol Methods 2018. https://doi.org/10.1016/j.jim.2018.09.008.
66. Parks DR, Moore WA, Brinkman RR, Chen Y, Condello D, El Khettabi F, Nolan JP, Perfetto SP, Redelman D, Spidlen J, et al. Methodology for evaluating and comparing flow cytometers: A multisite study of 23 instruments. Cytometry A 2018;93: 1087–1091.
67. Hulspas R, Dombkowski D, Preffer F, Douglas D, Kildew-Shah B, Gilbert J. Flow cytometry and the stability of phycoerythrin-tandem dye conjugates. Cytometry A 2009;75:966–972.
68. Zarkowsky D, Lamoreaux L, Chattopadhyay P, Koup RA, Perfetto SP, Roederer M. Heavy metal contaminants can eliminate quantum dot fluorescence. Cytometry A 2011;79:84–89.
69. Nomura L, Maino VC, Maecker HT. Standardization and optimization of multiparameter intracellular cytokine staining. Cytometry A 2008;73A:984–991.
70. Wood BL, Arroz M, Barnett D, DiGiuseppe J, Greig B, Kussick SJ, Oldaker T, Shenkin M, Stone E, Wallace P. 2006 Bethesda International Consensus recommendations on the immunophenotypic analysis of hematolymphoid neoplasia by flow cytometry: Optimal reagents and reporting for the flow cytometric diagnosis of hematopoietic neoplasia. Cytometry B Clin Cytom 2007;72(Suppl 1):S14–S22.
71. Roederer M, Tárnok A. OMIPs—Orchestrating multiplexity in polychromatic science. Cytometry A 2010;77:811–812.
72. Mair F, Prlic M. OMIP-044: 28-color immunophenotyping of the human dendritic cell compartment. Cytometry A 2018;93:402–405.
73. Staser KW, Eades W, Choi J, Karpova D, DiPersio JF. OMIP-042: 21-color flow cytometry to comprehensively immunophenotype major lymphocyte and myeloid subsets in human peripheral blood. Cytometry A 2018;93:186–189.
74. Baumgart S, Peddinghaus A, Schulte-Wrede U, Mei HE, Grützkau A. OMIP-034: Comprehensive immune phenotyping of human peripheral leukocytes by mass cytometry for monitoring immunomodulatory therapies. Cytometry A 2016;91: 34–38.
75. Engel P, Boumsell L, Balderas R, Bensussan A, Gattei V, Horejsi V, Jin B-Q, Malavasi F, Mortari F, Schwartz-Albiez R, et al. CD nomenclature 2015: Human leukocyte differentiation antigen workshops as a driving force in immunology. J Immunol 2015;195:4555–4563.
76. Mahnke YD, Roederer M. Optimizing a multicolor immunophenotyping assay. Clin Lab Med 2007;27:469–485.
77. Davis KA, Abrams B, Iyer SB, Hoffman RA, Bishop JE. Determination of CD4 antigen density on cells: Role of antibody valency, avidity, clones, and conjugation. Cytometry 1998;33:197–205.
78. Wang L, Abbasi F, Ornatsky O, Cole KD, Misakian M, Gaigalas AK, He HJ, Marti GE, Tanner S, Stebbings R. Human CD4+ lymphocytes for antigen quantification: Characterization using conventional flow cytometry and mass cytometry. Cytometry A 2012;81 A:567–575.
79. Kalina T, Lundsten K, Engel P. Relevance of antibody validation for flow cytometry. Cytometry A 2019; in press.
80. Maecker HT, Frey T, Nomura LE, Trotter J. Selecting fluorochrome conjugates for maximum sensitivity. Cytometry A 2004;62:169–173.
81. Böttcher S, van der Velden V, Villamor N, Ritgen M, Flores-Montero J, Escobar HM, Kalina T, Brüggemann M, Grigore G, Martin-Ayuso M, et al. Lot-to-lot stability of antibody reagents for flow cytometry. J Immunol Methods 2017. https://doi.org/10.1016/j.jim.2017.03.018.
82. Hedley BD, Keeney M, Popma J, Chin-Yee I. Novel lymphocyte screening tube using dried monoclonal antibody reagents. Cytometry B Clin Cytom 2015;88: 361–370.
83. Lee JA, Spidlen J, Boyce K, Cai J, Crosbie N, Dalphin M, Furlong J, Gasparetto M, Goldberg M, Goralczyk EM, et al. International Society for Advancement of Cytometry Data Standards Task Force, Scheuermann RH, Brinkman RR. MIFlowCyt: The minimum information about a flow cytometry experiment. Cytometry A 2008; 73:926–930.
84. Schulz AR, Baumgart S, Schulze J, Urbicht M, Grützkau A, Mei HE. Stabilizing antibody cocktails for mass cytometry. Cytometry A 2019;95:910–916.
85. Li SKH, Majonis D, Bagwell CB, Hunsberger BC, Baranov V, Ornatsky O. An efficient human whole blood workflow using CyTOF technology: a lyophilized 30-plex antibody panel coupled with automated data analysis. J Immunol 2018;200:120.2 LP-120.2.
86. Cosma A. Universal cell type identifier based on number theory. Cytometry A 2018; 93:649–652.
87. Diggins KE, Greenplate AR, Leelatian N, Wogsland CE, Irish JM. Characterizing cell subsets using marker enrichment modeling. Nat Methods 2017;14:275–278.
88. Diggins KE, Gandelman JS, Roe CE, Irish JM. Generating quantitative cell identity labels with marker enrichment modeling (MEM). Curr Protoc Cytom 2018;83: 10.21.1–10.21.28.