Immune Profiling Mass Cytometry Assay Harmonization: Multicenter Experience from CIMAC-CIDC

Bitá Sahaf¹, Mina Pichavant², Brian H. Lee³, Caroline Duault², Emily M. Thrash⁴, Melanie Davila⁵, Nicolas Fernandez⁵, Karen Millerchip⁶, Salah-Eddine Bentebibel⁶, Cara Haymaker⁷, Natalia Sigal², Diane M. Del Valle⁸, Srinika Ranasinghe⁴, Sarah Fayle⁷, Beatriz Sanchez-Espiridion⁷, Jiexin Zhang⁹, Chantale Bernatchez⁶, Catherine J. Wu^{4,10,11}, Ignacio I. Wistuba⁷, Seunghee Kim-Schulze⁸, Sacha Gnjjatic³, Sean C. Bendall¹², Minkyung Song¹³, Magdalena Thurin¹³, J. Jack Lee¹⁴, Holden T. Maecker², and Adeeb Rahman^{3,5}

ABSTRACT

Purpose: The Cancer Immune Monitoring and Analysis Centers – Cancer Immunologic Data Commons (CIMAC-CIDC) Network is supported by the NCI to identify biomarkers of response to cancer immunotherapies across clinical trials using state-of-the-art assays. A primary platform for CIMAC-CIDC studies is cytometry by time of flight (CyTOF), performed at all CIMAC laboratories. To ensure the ability to generate comparable CyTOF data across labs, a multistep cross-site harmonization effort was undertaken.

Experimental Design: We first harmonized standard operating procedures (SOPs) across the CIMAC sites. Because of a new acquisition protocol comparing original narrow- or new wide-bore injector introduced by the vendor (Fluidigm), we also tested this protocol across sites before finalizing the harmonized SOP. We then performed cross-site assay harmonization experiments using

five shared cryopreserved and one lyophilized internal control peripheral blood mononuclear cell (PBMC) with a shared lyophilized antibody cocktail consisting of 14 isotype-tagged antibodies previously validated, plus additional liquid antibodies. These reagents and samples were distributed to the CIMAC sites and the data were centrally analyzed by manual gating and automated methods (Astrolabe).

Results: Average coefficients of variation (CV) across sites for each cell population were reported and compared with a previous multisite CyTOF study. We reached an intersite CV of under 20% for most cell subsets, very similar to a previously published study.

Conclusions: These results establish the ability to reproduce CyTOF data across sites in multicenter clinical trials, and also highlight the importance of quality control procedures, such as the use of spike-in control samples, for tracking variability in this assay.

Introduction

The Cancer Immune Monitoring and Analysis Centers - Cancer Immunologic Data Commons (CIMAC-CIDC) Network was launched in September 2017 through the NCI Cancer Moonshot Initiative. Its goal is to identify biomarkers for optimizing immunotherapeutic strategies for patients with cancer. The four CIMACs are composed of the Icahn School of Medicine at Mount Sinai (MSSM), the Dana-Farber Cancer Institute (DFCI), MD Anderson Cancer Center (MDACC), and Stanford University (<https://cimac-network.org/>).

To accomplish its goal, the CIMACs offer a variety of harmonized immune-monitoring and genomic assays, which are applied to study

responses to immunotherapies across multiple clinical trials. The sites have adopted cytometry by time of flight (CyTOF) as the assay of choice for high-dimensional single-cell immune monitoring and biomarker discovery of cells circulating in the blood. Mass cytometry largely avoids the issues of overlap between detection channels, as seen in high-dimensional flow cytometry, and thus allows the routine use of over 40 simultaneous markers to perform high-throughput single-cell analyses (1–3).

Since the first publication of comprehensive immunological data generated by CyTOF (4), most translational studies adopting mass cytometry have been performed in a single lab at a single site (5). Few studies have tested machine-to-machine variation, although in recent

¹Stanford Cancer Institute, Stanford Medicine, Stanford University, California.

²Stanford Institute for Immunity, Transplantation and Infection, Stanford Medicine, Stanford University, California. ³Human Immune Monitoring Center, Icahn School of Medicine at Mount Sinai, New York, New York. ⁴Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts.

⁵Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York. ⁶Department of Melanoma Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas. ⁷Department of Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center, Houston, Texas. ⁸Human Immune Monitoring Center, Tisch Cancer Institute and the Precision Immunology Institute, Icahn School of Medicine at Mount Sinai, New York, New York.

⁹Department of Bioinformatics & Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, Texas. ¹⁰Harvard Medical School, Harvard University, Boston, Massachusetts. ¹¹Broad Institute of MIT and Harvard, Cambridge, Massachusetts. ¹²Department of Pathology, Stanford Medicine, Stanford University, Stanford, California. ¹³Division of Cancer Treatment and Diagnosis, National Cancer Institute, Bethesda, Maryland.

¹⁴Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, Texas.

Note: Supplementary data for this article are available at Clinical Cancer Research Online (<http://clincancerres.aacrjournals.org/>).

B. Sahaf and M. Pichavant contributed equally to this article.

Corresponding Author: Bitá Sahaf, Stanford Cancer Institute, Stanford Medicine, Stanford University, Lorry Lokey Building/SIM 1265 Campus Drive, Ste G2103, Stanford, CA 94305-5456. Phone: 650-721-0780; Fax: 650-387-7921; E-mail: bsahaf@stanford.edu

Clin Cancer Res 2021;27:5062–71

doi: 10.1158/1078-0432.CCR-21-2052

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 International (CC BY-NC-ND).

©2021 The Authors; Published by the American Association for Cancer Research

Translational Relevance

Deep phenotyping and functional analysis of blood cell subsets can identify biomarkers that correlate with clinical outcome in response to immunotherapy. CyTOF is one of the most comprehensive tools currently available for cell subset phenotyping at the single-cell level. Comparison of CyTOF data across trials may be necessary to reveal the consistency or differences of biomarkers, across tumor types and immunotherapy modalities. Harmonization is an essential process that enables comparison of data across different sites and trials. Reproducibility of CyTOF performance across multiple centers was addressed through integrating laboratory-specific protocols and tracking performance across sites.

years the use of normalization beads to account for intrainstrument day-to-day variation has become common practice (6, 7). At least two multicenter studies have been reported (8), and one multicenter CyTOF assessment compared staining and instrument performance consistency (9).

Although challenging to coordinate, multicenter immune-monitoring studies have been successfully performed using fluorescent flow cytometry (10–12). To ensure consistency in sample collection and processing and reproducibility of the data across the Network, the CIMACs have validated and harmonized sample collection protocols, immune-phenotyping panels, and staining and acquisition protocols across sites. A reference panel for all major immune-cell populations and best practices for tracking batch variability have been previously reported by our group (13).

The data presented in this publication represent the completion of iterative efforts to integrate laboratory-specific mass cytometry protocols across CIMAC sites, followed by a multisite experiment designed to evaluate reproducibility using the harmonized protocol. Our aim here was to at least match the level of concordance reached using immune monitoring by flow cytometry (12). The experiment utilized a combination of healthy-donor peripheral blood mononuclear cells (PBMCs) and lyophilized reference standards and was designed to evaluate both intrasite and intersite reproducibility. The harmonization process also focused on identification of variables that affect assay performance, including those attributable to sample preparation, staining step, or those associated with mass cytometry data acquisition.

We demonstrated that cross-site harmonization is feasible by means of integration of laboratory-specific protocols, shown in **Figure 1**, allowing identification and improvement of variables relevant to assay performance. We observed that, in CyTOF analysis, experimental rigor, gating strategies, and data reporting play an important role in data comparability. In addition, we present the metrics [coefficient of variation (CV) of cell frequencies, variance component analysis] that proved to be essential in validating the performance of the assay across different sites. Finally, we further discussed the need for internal and cross-site controls in each experiment for continued monitoring of concordance between sites. The harmonized CyTOF protocol is provided at <https://cimac-network.org/> for access and potential application by the larger scientific community.

Materials and Methods

Reagents

The thawing media consisted of RPMI 1640 (Hyclone) supplemented with L-Glutamine–Penicillin (100 U/mL)–Streptomycin

(100 µg/mL; Hyclone) and 10% inactivated FBS (Hyclone). The Rh103 Intercalator, 500 µmol/L (No. 201103A), Cell-ID Intercalator-Ir, 500 µmol/L (No. 201192B, aliquots kept at -20°C), EQ four element calibration beads (No. 201078), MaxPar Cell Acquisition Solution (CAS, No. 201237) for Wide Bore (WB) or MaxPar Water for Narrow Bore (NB) users, and Tuning solution (No. 201072) were purchased from Fluidigm Corporation. Fc receptor blocking solution Human TruStainFcX and lyophilized Tantalum 181 (181Ta)-labeled PBMC Veri-cells were purchased from Biolegend. Benzonase (Pierce Universal Nuclease for Cell Lysis, No. 88701) and 16% formaldehyde (w/v, methanol-free, Pierce No. 28906) were acquired from Thermo Fisher. 10X CyPBS (No. MB-008) and 10X PBS (No. 70011-044) were purchased from Rockland and Gibco (Thermo Fisher) respectively. BSA (No. A1933) was obtained from Sigma-Aldrich. Five-milliliter tubes with cell strainer caps (No. 352235) were from Corning.

Generation of cell-surface staining lyophilized panel

The harmonization experiments were performed using a previously described 14-marker antibody panel composed of antibodies to allow identification of all major immune cell types in peripheral blood (ref. 14; Supplementary Table I), along with additional liquid “drop-in” antibodies (CD45, CD45RA, PD-1, PD-L1, and CD69). We chose this 14-marker panel to cover the major immune subsets, while leaving the rest of the panel available for trial specific antibodies. The antibodies in this panel were first individually conjugated to their corresponding isotope using commercial MaxPar X8 conjugation kits (Fluidigm) and titrated on PBMCs to establish an optimal usage concentration. The antibodies were then combined as a cocktail and revalidated, after which the cocktail was lyophilized and packaged as single-test “lyospheres” (Biolyph). Staining performance of the lyophilized antibody cocktail was validated by direct comparison on parallel aliquots of PBMCs stained with reserved aliquots of the original liquid antibodies. The ultimate performance of the lyophilized cocktail was found to be highly concordant with that of the liquid antibodies, while offering advantages in long-term stability, thermal resistance, and reduced potential experiment variability due to multiple pipetting steps, making it well-suited for shipment to multiple sites for our harmonization experiments.

PBMCs

The PBMCs from four healthy donors' blood were isolated by density gradient using Ficoll-Paque (GE-HealthCare, LifeSciences), then divided into equal aliquots. Part of the samples were directly cryopreserved and stored in liquid nitrogen. Other aliquots were stained with the lyophilized antibody cocktail consisting of 14 isotype-tagged antibodies at site at Mount Sinai (as described below) prior to freezing. A single vial of each unstained and prestained PBMC sample was distributed to each of the four CIMACs.

In parallel, another batch of healthy-donor PBMCs was prepared at DFCI and split into two conditions, unstimulated (US) and stimulated (Stim) cells. To stimulate cells, CD3/CD28 Dynabeads (Thermo Fisher) were incubated with the cells in RPMI1640 media supplemented with 10% FBS and 1X Antibiotic-Antimycotic (Thermo Fisher) for 3 days. On the fourth day, CD3/CD28-stimulated cells were treated with phorbol 12-myristate 13-acetate (PMA; 50 ng/mL) and ionomycin (1 µg/mL) for 2 hours. After 2 hours of incubation with PMA + ionomycin, cells were frozen at 5 million viable cells in 0.5 mL of freezing media and stored in liquid nitrogen. US and Stim cells were again divided into an equal number of samples and distributed to each CIMAC site.

To guide investigators through sample distribution options among a variety of assays and provide standardized methods for specimen

Figure 1.

Material used for the second round of the cross-site harmonization study. All sites stained cells with lyophilized 14-marker cocktail provided with addition of liquid CD45, CD45RA, PD1, PD-L1, and CD69. In addition, Ta-tagged Veri-cells were added to each sample. Prestained and unstained cells from each healthy donor were sent to each site. A mixed Unstim+PMA/ionomycin stimulated sample from an external batch was added as a longitudinal control.

collection and handling, including for CyTOF, the Network and NCI developed the CIMAC Specimen Collection “Umbrella” protocol (found at <https://cimac-network.org/documents/>). The Umbrella protocol addresses various steps in the “sample flow” from tissue or blood-sample acquisition at trial sites, to immediate processing and storage at biorepositories, to subsequent processing and downstream distribution to the CIMAC laboratories.

Shipment to sites

All CIMAC sites were provided with cryopreserved PBMCs from the same four donor lots; antibody cocktail 2X lysopherses; additional liquid antibody conjugates for CD45, CD45RA, PD-1, PD-L1, and CD69 markers; and cells centrally prestained with the lyophilized 14-marker cocktail (all from the MSSM CIMAC). Each site was also provided with aliquots of US- and Stim- control PBMCs from the DFCI CIMAC. Veri-cells (Ta-labelled cells, Biolegend) were provided to all sites from the same batch (Fig. 1).

Thawing healthy-donor and US/Stim PBMCs

Each site stained PBMCs from the four donor lots, along with the stimulation control (50:50 mixture of US- and Stim- control PBMCs). The frozen PBMC vials were kept on dry ice until time to thaw. Each cryopreserved PBMC vial was thawed and added to 10-mL thawing media containing 1:10,000 benzonase warmed in a 37°C water bath. After centrifugation at 300 g for 5 minutes, the supernatant was removed, and the cell pellet was resuspended in 5 mL of RPMI with 10% FBS. The cells were then counted and recorded for post-thaw count and viability and centrifuged again at 300 g for 5 minutes. Aliquots of 2 million cells from each sample were prepared in RPMI containing 10% FBS for staining purposes and kept on ice. The US and Stim samples were combined 1:1 for a total of 2 million cells.

Reconstitution of lyophilized Veri-cells

As an internal reference control, lyophilized ¹⁸¹Ta-tagged Veri-cells (Biolegend) were used. To reconstitute the Veri-cells, both the reconstitution buffer and Veri-cells were brought to room temperature for 5 minutes. In a glass vial, 1.3 mL of reconstitution buffer was added to the cells and incubated at room temperature for 10 minutes. The Veri-cells were then centrifuged for 1 minute at 1,500 g and the supernatant removed. The cell pellet was resuspended in 1 mL of Cell Staining Media (CSM) made of PBS + 0.2% BSA + 0.02% NaN₃ and centrifuged at 1,500 g. Once the supernatant was aspirated, the cell pellet was then resuspended in 1 mL of CSM. The cell count was taken to ensure an amount of approximately 6 million cells. Into each 2 million PBMC sample, 200,000 reconstituted Veri-cells were spiked. Veri-cells will also serve as a standard reference material for quality-control assessment of assay performance within and across different CIMACs over time. Even though we are describing the use of Vericells as internal controls, a similar workflow could be implemented using cryopreserved aliquots of a PBMC reference sample labeled with a distinct metal barcode to serve as internal reference controls during analysis.

Reconstitution of the antibody cocktail lysosphere

To reconstitute the 2X antibody cocktail from the lyophilized panel, each lysosphere was reconstituted using 50 µL of CSM per lysosphere while avoiding touching the lysosphere with the pipette tip. After a 5-minute incubation at room temperature, the reconstituted antibody cocktail was transferred to 0.1 µmol/L polyvinylidene difluoride (PVDF) filter tubes (Millipore UFC30VV00), and 1 µL of each additional liquid antibody (per sample) and 2 µL of PD1-153Eu were added to the antibody cocktail (CD45-89Y, CD45RA-143Nd, PDL1-156Gd, CD69-164Dy). The filter tube containing the antibody cocktail was centrifuged at 12,000 g for 1 minute.

Cell-surface staining

1X Rh103 viability staining media was obtained by diluting the stock to 1 $\mu\text{mol/L}$ staining solution in prewarmed cell culture medium (RPMI+10% FBS). Then, 1 mL of 1X Rh103 staining medium was added and mixed with the unstained PBMCs. Cells were incubated at 37°C for 20 minutes and spun down at $300 \times g$ for 5 minutes. Once the pellet was formed, the supernatant was aspirated, and the cells were washed by adding 1 mL of CSM and centrifuging at $300 \times g$ for 5 minutes.

For the cell-surface staining, 2X TruStain Fc block (FcX) was prepared by adding 1 μL of FcX to 50 μL of CSM for each sample. The mixture of aliquoted cells with the spiked-in reconstituted Veri-cells were then spun down at $300 \times g$ for 5 minutes to aspirate supernatant. The cells were resuspended in 50 μL of CSM + FcX and mixed with 50 μL of filtered antibody cocktail. The cells were stained for 30 minutes on ice and washed with 1 mL of CSM. A last spin was performed at $300 \times g$ for 5 minutes. Prestained cells from each healthy donor were sent with the unstained cells to each site for CyTOF analysis.

To fix and stain cell nuclei with iridium (Ir), the cells were resuspended in 200 μL of PBS. A series of reagents dilutions were prepared as depicted in Supplementary Protocol and described below. The 14.4% isotonic formaldehyde was prepared by mixing 1800 μL of 16% formaldehyde with 200 μL of 10X PBS. The fixation buffer was prepared by mixing 2,428 μL of PBS, 320 μL of 2% saponin in PBS solution, and 1,332 μL of 14.4% isotonic formaldehyde. The 2X fixative-Ir solution was made by mixing 2 μL of stock 500 $\mu\text{mol/L}$ Intercalator-Ir solution diluted in 1 mL Fix Buffer (final concentration of 1 $\mu\text{mol/L}$). We used 200 μL per sample. This 1- $\mu\text{mol/L}$ solution was then diluted into 1:16 and 200 μL was added to each sample, for a final concentration of Ir of 0.03 $\mu\text{mol/L}$. The samples were incubated for 30 minutes at room temperature. The cells were then washed once in 1 mL CSM, spun down at 1,500 g for 1 minute, and stored in 200 μL of

CSM containing 0.03 $\mu\text{mol/L}$ Intercalator-Ir. The stained cells were stored for up to 1 week at 4°C. To perform the wash, the cells were centrifuged in a microcentrifuge at 1,500 $\times g$ for 1 minute, and the tube then rotated 180° and centrifuged at 1,500 $\times g$ for an additional 15 seconds so the pellet moved to the bottom of the tube. The supernatant was carefully aspirated and discarded, and the pellet gently vortexed and resuspended in 1 mL of MilliQ Water (for NB) or MaxPar CAS (WB). The cells were then washed a second time in MaxPar Water (for NB) or MaxPar CAS (for WB). The steps from the centrifugation at 1,500 $\times g$ for 1 minute were repeated to resuspend the cells in 300 μL of MaxPar Water (NB) or MaxPar CAS (WB). The sample was filtered through a cell strainer cap into a FACS tube, and the cells counted. A 1:10 dilution of EQ beads were added and acquired using the NB or WB injector, aiming for a target acquisition of a minimum of 300,000 events.

Data analysis

FCS files were centrally processed using an analysis workflow utilizing a combination of the Cytobank (www.cytobank.org) and Astrolabe Diagnostics (<https://astrolabediagnostics.com>) analysis platforms. Both liquid and lyophilized antibody-defined populations were analyzed. FCS files first underwent bead-based normalization, followed by exclusion of Ce140⁺ beads and bead-cell doublets, Gaussian ion cloud multiplet fusion events, and Rh103⁺ dead cells (15). Spiked-in Veri-cells were identified based on Ta181⁺ signal, and this population, together with the Ta181-viable CD45⁺ events, were then selected as the primary populations for downstream analyses and represent the parent-cell populations for all downstream subset frequencies (Fig. 2). These pregated populations were then clustered and annotated using a hierarchical flowSOM approach as implemented in the Astrolabe Diagnostics platform. The population annotations in the Astrolabe-labeled FCS were validated and visualized using tSNE visualization in Cytobank,

Figure 2.

Data processing workflow. **A**, Data processing pipeline for preanalytical processing included panel naming, bead normalization, cell identification, doublet exclusion, and live CD45⁺ cell gating. **B**, Manual gating strategy.

and the frequencies of major immune-cell subsets were then exported for statistical analyses and graphing using Microsoft Excel and Prism 8 (GraphPad).

The CIDC provides bioinformatics methods and the computational expertise and resources to facilitate the analysis of immunology trial data for the Network. Using a unified workflow management system, the CIDC has established several uniform bioinformatics processing pipelines including licensing the Astrolabe platform, which uses an automated gating strategy to determine cell populations. CIDC will facilitate future cross-trial analyses for CyTOF data.

Statistical considerations

Descriptive statistics and summary statistics such as mean, SD, and CV (population SD divided by mean) were calculated to report the results of the harmonization. Intersite CVs were calculated as unit-free measures to quantify the variability across the four CIMAC sites. Intrastate CVs were calculated for the controls (Veri-cells). Variance component analysis was used to partition the variability into three sources: due to site, subject, or random variation. A good harmonization would show that most of the variability comes from sample variation while the site variability and the residual error constitute small portions of the total variability compared with the sample variation.

Results

Consensus SOP generation

Our harmonization efforts began with a detailed evaluation of the experimental PBMC thawing and staining protocols currently in use at each CIMAC site. After identifying differences in independent protocols that could potentially contribute to experimental variability at each step of the experimental workflow, we developed a single consensus SOP (Supplementary Protocol I). Some of the key factors that were harmonized included (i) the centrifugation speed during the wash steps; (ii) the use of fresh formaldehyde to ensure optimal fixation of the samples prior to data acquisition; and (iii) the inclusion of saponin during intercalator staining to ensure effective entry of the Ir intercalator into cell nuclei.

Acquisition protocol testing

After establishing a consensus SOP for sample staining, we next evaluated data acquisition protocols in use at each site. A major consideration in this regard was the recent introduction by Fluidigm of an alternate WB sample injector and sample acquisition buffer (CAS) offered for use on the Helios mass cytometer. We performed extensive testing of sample acquisition using the conventional NB injector and MilliQ water protocol in comparison with the alternate WB and CAS protocol, the results of which have been published in detail elsewhere (16). While the WB and CAS protocol was shown to result in an overall reduction in interinstrument variability in that study, the relative impact of the two protocols was found to be highly instrument-specific. Our data pertaining to head-to-head comparison of the two methods on the same instrument clearly showed that there were no significant differences between the data acquired using either protocol. Therefore, it was ultimately decided that each CIMAC site should utilize the data acquisition protocol that resulted in optimal data quality on their specific instruments.

We next designed an experiment to evaluate cross-site experimental reproducibility using our consensus SOP shown in Fig. 1. Replicate

aliquots of cryopreserved PBMCs from four donors were shipped to each site together with a lyophilized metal-labeled reference sample (Biolegend Ta-labeled Veri-cells), a lyophilized 14-marker antibody cocktail, and supplemental liquid antibodies. The experimental protocol specified that each vial of the cryopreserved PBMCs would be spiked with an aliquot of the Ta-tagged lyophilized Veri-cells to serve as an internal reference control, and the samples would then be processed for staining with the supplied lyophilized antibody cocktail in accordance with the consensus SOP.

To compare intrasite and cross-site concordance, CyTOF data acquisition by each CIMAC site was measured on both site-stained and centrally stained aliquots of the same PBMC samples. The staining patterns of samples were acquired using a supplied acquisition template, and a minimum 300,000 cells were acquired for each sample. The resulting FCS files were centrally processed using an analysis pipeline that included bead-based normalization, bead exclusion, cell doublet exclusion, live/dead exclusion, and gating all CD45⁺ cells as parent population for all subsets (Fig. 2). Major immune-cell subsets were identified using hierarchical clustering approaches (Astrolabe) and further confirmed using manual gating strategies.

Analysis of spiked Veri-cells

The internal Veri-cell reference samples were first distinguished from biological PBMC samples based on Ta181 labeling. Cell distributions were visualized in the prestained (Fig. 3A) and the site-stained samples (Fig. 3B) using tSNE, and the frequencies of major immune-cell types were calculated for both prestained (Fig. 3C) and site-stained samples (Fig. 3D). While we did note some differences in absolute signal intensity for some mass channels between sites as shown by the changes in relative distribution of populations on the tSNE plots (Fig. 3A), we observed largely uniform percentages for B cells, CD4⁺ T cells, CD8⁺ T cells, CD14⁺ monocytes, and natural killer (NK) cells in prestained and site-stained PBMC samples (Fig. 3C and 3D). Because these Veri-cell population frequencies are expected to be uniform, we evaluated these frequencies across samples to determine the intrasite reproducibility for each population (Fig. 3E and 3F). Intrastate CVs were slightly lower for prestained samples but were generally more than 5% for all cell populations in both prestained and site-stained samples. We next determined the intersite concordance (Fig. 3G and 3H). While higher than the intrasite CVs, intersite CVs were still more than 10% for most populations, with the exception of NK cells in the site-stained samples, which was partly attributable to poorly resolved CD56 staining on the Veri-cells.

Analysis of donor PBMCs

We next performed an analogous analysis for each of the donor's PBMC samples. Major immune-cell populations were visualized by tSNE (Fig. 4A and 4B) and their frequencies determined for both prestained (Fig. 4C) and site-stained (Fig. 4D) PBMC samples. Despite some differences in relative mass channel signal intensity between sites, overall relative population frequencies were highly concordant for both the prestained and the site-stained samples across all sites. Population-specific intersite CVs showed a generally expected trend towards higher CVs for rarer populations. Intersite CVs were more than 10% for all populations in the prestained samples (Fig. 4E). While intersite CVs were somewhat higher in site-stained populations (Fig. 4F), suggesting a contribution of sample processing and thawing procedures towards experimental variability, they were still more than 15% for all populations, with the exception of basophils, a relatively rare cell type.

Figure 3.

Final harmonized protocol and pipeline evaluated across sites based on Veri-cells. Major cell populations were gated, and overall phenotypic distributions were visualized using tSNE (**A** and **B**). Population frequencies are shown for each sample across each site for prestained samples (**C**) and site-stained samples (**D**); given the low relative cell numbers of the internal Veri-cells, analysis was restricted to more abundant major immune cell types). Linked points represent internal reference cells from the same donor sample analyzed at each site. The corresponding average intrasite CV is shown in **E** and **F**, calculated by determining the intrasite CV for each site and then averaging across sites. The corresponding intersite CV is shown in **G** and **H**, calculated by first determining the intersite CV for each sample and then taking the average across all samples.

Variance component analysis

In addition to evaluating overall performance and data concordance of CyTOF analysis between CIMAC sites, a key consideration was the ability to consistently resolve biological variation between samples. To

assess this, variance component analysis was used to evaluate the relative contribution of site and samples to the overall observed variance. The stacked bar represents the percentage of total variance that can be attributed to the site, the sample, or random variation. In

Figure 4.

Final harmonized protocol and pipeline evaluated across sites based on PBMCs. Major cell populations were gated, and overall phenotypic distributions were visualized using tSNE (**A-B**). Population frequencies are shown for each sample across each site for prestained samples (**C**) and site-stained samples (**D**). Linked lines represent the same donor sample analyzed across sites. The corresponding intersite CV is shown in **E** and **F**, calculated by first determining the intersite CV for each sample and then taking the average across all samples.

the case of both the prestained (**Fig. 5A**) and site-stained (**Fig. 5B**) samples, the biological characteristics of the sample are the primary source of variance for all cell populations evaluated. For the prestained samples, biological variation associated with the different donor samples accounted for more than 90% of the variance for all cell types except for plasmacytoid dendritic cells. For the site-stained samples, the biological variability accounted for more than 80% of the variance for all cell types except for B cells, basophils, and pDCs.

Discussion

Mass cytometry offers a powerful approach to conducting highly-multiplexed single-cell immune-profiling to identify immune changes induced by cancer immunotherapy. Here, we report a harmonization effort to establish consistent CyTOF phenotyping of clinical samples

across sites. We report concordance that is very similar to previous data for multisite flow and mass cytometry studies (1, 6, 9). In addition, we describe the methodologic harmonization of this multi-omic assay. We divided our approach into (i) technical harmonization, (ii) harmonizing standards, (iii) testing internal controls, and (iv) testing intersite controls.

As part of this harmonization effort across the four CIMAC sites, we evaluated the immunostaining standard operating protocols from each of the sites and developed a consensus protocols covering PBMC thawing, immunostaining, and CyTOF data acquisition. Pilot studies were then designed to identify key factors that may potentially contribute to experimental variability, including the need for fresh formaldehyde for effective fixation, consistent centrifugation velocity to prevent cell loss, and optimal intercalator staining procedures to ensure effective cell identification. Ultimately, the consensus protocol

Figure 5.

Variance component analysis based on PBMCs. For both the prestained samples (A) and the site-stained samples (B), variance component analysis was used to evaluate the relative contribution of site and samples to the overall observed variance. The stacked bar represents the percentage of total variance that can be attributed to the site, the sample, or random variation.

was used for a final experiment to map the cross- and intrasite variability as performed by other multisite CyTOF harmonization efforts described earlier (9). The study findings include the following: First, acquisition of insufficient cell numbers was found to contribute to higher CVs, particularly for rarer cell populations. Second, inclusion of a viability marker and CD45 as a pan-immune cell marker was determined to be important to ensure a consistent denominator for calculating cell frequencies. Third, uniform data preprocessing, including bead-based normalization, stringent DNA gates, and time-based gating to remove instrument-dependent time-based artifacts were also found to be important factors upstream of cell-subset definition. Developing the consensus protocol and leveling the proficiency of all sites to eliminate such discrepancies, ultimately allowed to obtain optimally standardized data from all four sites.

Several key concepts informed the design of our final experiment:

- (i) Centrally stained samples are ideally suited to evaluating experimental variability specifically associated with data acquisition and instrument performance.
- (ii) Stable, biologically homogenous samples (e.g., lyophilized reference standards, Veri-cells, or any viably frozen vials of metal-tagged PBMCs) are well suited for evaluating variability associated with staining and data acquisition protocols.
- (iii) In addition to homogenous reference samples, biologically heterogeneous samples (e.g., PBMCs from multiple individuals with known differences in relative cell composition and/or phenotype) should also be evaluated for reproducibility to allow an evaluation of both absolute data concordance and the ability to accurately resolve relevant biological heterogeneity between samples.
- (iv) The inclusion of a viability marker (Rh103) and ubiquitous cell surface marker (e.g., CD45) are important to define a clear-base parent population that can be used as a consistent denominator for expressing more granular cell-subset frequencies.
- (v) Low cell numbers can introduce considerable variability when evaluating the frequency of rare cell types. Hence, a minimum cell acquisition threshold for all samples is necessary to ensure sufficient sampling of the rarest cell populations intended for comparison.

- (vi) Use of a standard data acquisition template, including uniform and panel/channel names and a consistent file-naming structure, can greatly facilitate downstream data analyses.
- (vii) All experimental goals and protocols should be clearly defined, and consensus protocols should be developed following an iterative process of review and discussions among all study participants.

The conducted experiments utilized a combination of PBMCs from distinct donors and a commercial metal-labeled lyophilized Veri-cells spike-in reference control to evaluate intra and intersite data reproducibility using both prestained and site-stained samples (Figs. 3 and 4). Evaluation of overall staining consistency using Veri-cells in prestained samples showed very high intrasite reproducibility (<5% CV), and slightly lower but acceptable levels of intersite reproducibility (mostly <10% CV). The intersite reproducibility in prestained donor PBMCs was very similar to that of the Veri-cells, whereas site-stained PBMCs had slightly higher variability (<15% CV for most cell populations). This level of precision is consistent with a previous reference study (9).

Our study also included a control sample that was a mixture of Stim and US PBMCs. Although we do not present data on that sample here, it served as a useful staining control, particularly for activation markers whose performance might otherwise be difficult to judge on resting PBMCs. This was also a drawback of the Veri-cells, which, due to being fixed and lyophilized, do not stain well for all markers of potential immunophenotyping interest. We therefore anticipate continued use of both the US + Stim control, as well as the Veri-cells, in our ongoing CIMAC studies. As a large-scale source of cells that was distributed to all centers, and since Veri-cells express the major lineage markers, the addition of these cells as controls by each site and in each experiment will serve to continuously monitor intrasite variability. This protocol has already been implemented in two studies completed by the DFCI CIMAC postharmonization (17, 18). These data will be tracked by the CIDC to ensure continued harmonization.

While a lyophilized antibody cocktail formed the basis of this study, we added liquid antibodies as well, mimicking the expected customization to be used in the CIMAC network going forward. These liquid antibodies performed well overall, but the need for careful control of reagent performance is increased with the introduction of liquid antibodies.

Each new custom antibody is validated and titrated with relevant positive and negative control-cell populations for the marker being tested. To facilitate this, replicate vials of healthy-donor PBMC or PMA/Ionomycin-stimulated PBMC samples depending on the marker are used. To validate the coexpression patterns of the new marker, we stain in the context of the full antibody panel at the identified titer. Finally, the reproducibility of the antibody expression profile is validated using a second screen performed using an additional donor.

We propose that CIMAC and other studies use frozen aliquots of antibody cocktails for longitudinal experiments that need to be compared, as recently validated by Schulz and colleagues (19). In fact, other than temperature stability, lyophilized cocktails offer no great advantage to frozen aliquots, even for a “core” cocktail.

This study used commercial solutions for lyophilized antibodies (BioLyph Lyospheres), spike-in control cells (VeriCells), and analysis software (Astrolabe). However, noncommercial options for these can also be employed. As mentioned previously, frozen antibody cocktails have been shown to be stable and are a good alternative to lyophilized cocktails (19). Any mass cytometry lab could produce their own metal-tagged control cells and freeze aliquots. Some groups have produced their own analytical packages for standardized cell-subset identification that are freely available (20, 21). Although commercial solutions offer convenience and generally good reproducibility, they can have cost barriers and other limitations, and should not be considered the only route to standardization.

Moving forward, the CIMAC/CIDC consortium will use the consensus SOP developed here, but with a larger lyophilized “core” cocktail commercially available (Maxpar Direct Immune Profiling Assay, Fluidigm). This 30-antibody panel can still be augmented with multiple additional antibodies of interest to particular studies. The network will collect data from both Veri-Cells and US + Stim PBMC controls to determine the level of ongoing harmonization between centers and studies.

Conclusions

- We determined the variations in frequencies of gated populations (CVs) to be the most appropriate read-out for performance analyses.
- Our final CIMAC-wide analysis showed a high level of concordance in major cell populations (under 20% CV) as presented in Fig. 4.
- Regarding the operator-driven differences in specimen preparation, it is recommended that the same, trained operators are involved to secure data consistency; new operators be thoroughly trained, and their work checked before running clinical samples; and any deviations to the protocol be documented.
- In addition, analytical harmonization in either through adopting the same manual gating strategy or unsupervised gating will ensure consistency and reduced variability.
- As a conclusion of our harmonization efforts, this multisite comparison demonstrates that all sites are able to generate highly concordant data that should allow for effective comparative analyses of CyTOF results generated by each site across clinical studies comparing cancer immunotherapeutics and derive meaningful biomarker datasets, the ultimate goal of the CIMAC-CIDC network.
- Evaluation of assay concordance is ultimately anticipated to be a consistent ongoing effort across all CIMAC-CIDC studies, which will be facilitated by inclusion of shared spike-in reference controls across all study samples.

- The harmonized CIMAC CyTOF protocol can be found at <https://cimac-network.org/>.

Recommendations

- Future studies involving longitudinal or cross-site CyTOF collection should use a well developed consensus protocol shared across sites (Supplementary protocol I), and test that protocol for reproducibility prior to study initiation.
- Optimal controls are stained for all relevant markers that are spiked into samples under study. These could be metal-labeled Stim + US PBMC, which can be frozen in aliquots and thus take the place of lyophilized Veri-Cells.
- Although lyophilized cocktails are not required, at a minimum, frozen aliquots of antibody cocktails should be used to avoid the variabilities of liquid cocktails that need to be created for each experiment.

Authors' Disclosures

E.M. Thrash reports personal fees from Fluidigm Corporation during the conduct of the study and personal fees from Fluidigm Corporation outside the submitted work; in addition, E.M. Thrash is an employee at Current Fluidigm. N. Fernandez reports personal fees from Vizgen outside the submitted work. C. Haymaker reports personal fees from Briacell Therapeutics outside the submitted work. C. Bernatchez reports grants from NCI during the conduct of the study and other support from Myst Therapeutics outside the submitted work. C.J. Wu reports other support from Pharmacocyclics outside the submitted work. I.I. Wistuba reports grants and personal fees from Genentech/Roche, Bayer, AstraZeneca, Pfizer, HTG Molecular, Merck, Guardant Health, and Novartis; personal fees from Bristol Myers Squibb, GlaxoSmithKline, Oncocyte, and MSD; grants from Adaptive, Adaptimmune, EMD Serono, Takeda, Amgen, Karus, Johnson & Johnson, Iovance, 4D, and Lilly; and grants from Akoya outside the submitted work. S. Gnjatic reports grants from Regeneron, BMS, Takeda, Genentech, and Janssen R&D; personal fees from OncoMed; and personal fees from Merck outside the submitted work. S.C. Bendall reports personal fees from Ionpath, and Purigen and grants from Biogen outside the submitted work. H.T. Maecker reports grants from NIH during the conduct of the study. A. Rahman reports grants from NIH/NCI during the conduct of the study; in addition, A. Rahman has a patent for metal-labeled reference standards for mass cytometry licensed to Biolegend. No disclosures were reported by the other authors.

Authors' Contributions

B. Sahaf: Data curation, formal analysis, supervision, validation, methodology, writing—original draft, project administration, writing—review and editing. **M. Pichavant:** Supervision, writing—original draft, project administration, writing—review and editing. **B.H. Lee:** Data curation, software, formal analysis, visualization. **C. Duault:** Data curation, formal analysis, methodology. **E.M. Thrash:** Data curation, methodology. **M. Davila:** Validation. **N. Fernandez:** Methodology. **K. Millerchip:** Methodology. **S.E. Bentebibel:** Methodology. **C. Haymaker:** Supervision, methodology, writing—review and editing. **N. Sigal:** Methodology. **D.M. Del Valle:** Project administration. **S. Ranasinghe:** Project administration. **S. Fayle:** Methodology, project administration. **B. Sanchez-Espiridon:** Project administration. **J. Zhang:** Data curation, software, formal analysis. **C. Bernatchez:** Project administration. **C.J. Wu:** Conceptualization, supervision, funding acquisition, validation, methodology, writing—original draft, writing—review and editing. **I.I. Wistuba:** Conceptualization, resources, supervision, methodology, project administration, writing—review and editing. **S. Kim-Schulze:** Data curation, methodology. **S. Gnjatic:** Conceptualization, resources, data curation, formal analysis, supervision, methodology, writing—original draft, writing—review and editing. **S.C. Bendall:** Conceptualization, resources, data curation, formal analysis, supervision, visualization, methodology, writing—review and editing. **M. Song:** Project administration. **M. Thurin:** Conceptualization, resources, supervision, investigation. **J.J. Lee:** Data curation, software, formal analysis, supervision. **H.T. Maecker:** Conceptualization, resources, formal analysis, supervision, methodology, writing—original draft, writing—review and editing. **A. Rahman:** Conceptualization, data curation, formal analysis, supervision, validation, visualization, methodology, writing—original draft, writing—review and editing.

Acknowledgments

Scientific and financial support for the CIMAC-CIDC Network are provided through the NCI Cooperative Agreements U24CA224319 (to the Icahn School of Medicine at Mount Sinai CIMAC), U24CA224331 (to the Dana-Farber Cancer Institute CIMAC), U24CA224285 (to the University of Texas MD Anderson Cancer Center CIMAC), U24CA224309 (to the Stanford University CIMAC), and U24CA224316 (to the CIDC at Dana-Farber Cancer Institute). Additional support is made possible through the NCI CTIMS Contract HHSN261201600002C (to the Emmes Company, LLC). MDACC received support from NIH Cancer Center Support Grant P30CA016672 and the University of Texas SPORE NCI P50CA70907. Scientific and financial support for the PACT project are made possible through funding support provided to the FNIH by AbbVie Inc., Amgen Inc., Boehringer-

Ingelheim Pharma GmbH & Co. KG., Bristol Myers Squibb, Celgene Corporation, Genentech Inc., Gilead, GlaxoSmithKline plc, Janssen Pharmaceutical Companies of Johnson & Johnson, Novartis Institutes for Biomedical Research, Pfizer Inc., and Sanofi.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received June 4, 2021; revised July 10, 2021; accepted July 12, 2021; published first July 15, 2021.

References

- Bendall SC, Nolan GP, Roederer M, Chattopadhyay PK. A deep profiler's guide to cytometry. *Trends Immunol* 2012;33:323–32.
- Ornaty O, Baranov VI, Bandura DR, Tanner SD, Dick J. Multiple cellular antigen detection by ICP-MS. *J Immunol Methods* 2006;308:68–76.
- Gadalla R, Noamani B, MacLeod BL, Dickson RJ, Guo M, Xu W, et al. Validation of cytoF against flow cytometry for immunological studies and monitoring of human cancer clinical trials. *Front Oncol* 2019;9:415.
- Bendall SC, Simonds EF, Qiu P, Amir ED, Krutzik PO, Finck R, et al. Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science* 2011;332:687–96.
- Brodin P, Jovic V, Gao T, Bhattacharya S, Angel CJL, Furman D, et al. Variation in the human immune system is largely driven by non-heritable influences. *Cell* 2015;160:37–47.
- Tricot S, Meyrand M, Sammiceli C, Elhmouzi-Younes J, Corneau A, Bertholet S, et al. Evaluating the efficiency of isotope transmission for improved panel design and a comparison of the detection sensitivities of mass cytometer instruments. *Cytometry A* 2015;87:357–68.
- Finck R, Simonds EF, Jager A, Krishnaswamy S, Sachs K, Fantl W, et al. Normalization of mass cytometry data with bead standards. *Cytometry A* 2013;83:483–94.
- Nasaar A, Carter B, Lannigan J, Montgomery R, Paul N, Poulin M, et al. The first multi-center comparative study using a novel technology mass cytometry time-of-flight mass spectrometer (CyTOF2) for high-speed acquisition of highly multi-parametric single cell data: a status report. *Talk at CYTO; 2005;Glasgow, Scotland*. p. 1100–20.
- Leipold MD, Obermoser G, Fenwick C, Kleinstuber K, Rashidi N, McNeven JP, et al. Comparison of CyTOF assays across sites: results of a six-center pilot study. *J Immunol Methods* 2018;453:37–43.
- Finak G, Langweiler M, Jaimes M, Malek M, Taghiyar J, Korin Y, et al. Standardizing flow cytometry immunophenotyping analysis from the human immunophenotyping consortium. *Sci Rep* 2016;6:20686.
- Kalina T, Flores-Montero J, Lecomte Q, Pedreira CE, van der Velden VHJ, Novakova M, et al. Quality assessment program for EuroFlow protocols: summary results of four-year (2010–2013) quality assurance rounds. *Cytometry A* 2015;87:145–56.
- Maecker HT, Rinfret A, D'Souza P, Darden J, Roig E, Landry C, et al. Standardization of cytokine flow cytometry assays. *BMC Immunol* 2005;6:13.
- Sahaf B, Rahman A, Maecker HT, Bendall SC. High-parameter immune profiling with CyTOF. *Methods Mol Biol* 2020;2055:351–68.
- Amir E-AD, Lee B, Badoual P, Gordon M, Guo XV, Merad M, et al. Development of a comprehensive antibody staining database using a standardized analytics pipeline. *Front Immunol* 2019;10:1315.
- Lee BH, Rahman AH. Acquisition, processing, and quality control of mass cytometry data. *Methods Mol Biol* 1989, 2019;13–31.
- Lee BH, Kelly G, Bradford S, Davila M, Guo XV, Amir E-AD, et al. A modified injector and sample acquisition protocol can improve data quality and reduce inter-instrument variability of the helios mass cytometer. *Cytometry A* 2019;95:1019–30.
- Penter L, Zhang Y, Savell A, Huang T, Cieri N, Thrash EM, et al. Molecular and cellular features of CTLA-4 blockade for relapsed myeloid malignancies after transplantation. *Blood* 2021;137:3212–17</bib>
- Monjazebe AM, Giobbie-Hurder A, Lako A, Thrash EM, Brennick RC, Kao KZ, et al. A randomized trial of combined PD-L1 and CTLA-4 inhibition with targeted low-dose or hypofractionated radiation for patients with metastatic colorectal cancer. *Clin Cancer Res* 2021;27:2470–80.
- Schulz AR, Baumgart S, Schulze J, Urbicht M, Grützkau A, Mei HE. Stabilizing antibody cocktails for mass cytometry. *Cytometry A* 2019;95:910–6.
- Nowicka M, Krieg C, Crowell HL, Weber LM, Hartmann FJ, Guglietta S, et al. CyTOF workflow: differential discovery in high-throughput high-dimensional cytometry datasets. [version 3; peer review: 2 approved]. *F1000Res* 2017;6:748.
- Chen H, Lau MC, Wong MT, Newell EW, Poidinger M, Chen J. Cytokit: A bioconductor package for an integrated mass cytometry data analysis pipeline. *PLoS Comput Biol* 2016;12:e1005112.