Published in final edited form as:

Cancer Immunol Immunother. 2015 May ; 64(5): 585–598. doi:10.1007/s00262-014-1649-1.

Data analysis as a source of variability of the HLA-peptide multimer assay: from manual gating to automated recognition of cell clusters

Cécile Gouttefangeas^{1,*,\$}, Cliburn Chan^{2,\$}, Sebastian Attig^{1,#}, Tania T. Køllgaard^{3,†}, Hans-Georg Rammensee¹, Stefan Stevanovi¹, Dorothee Wernet⁴, Per thor Straten³, Marij J.P. Welters⁵, Christian Ottensmeier⁶, Sjoerd H. van der Burg^{5,+}, and Cedrik M. Britten^{7,+} for the CIP group

¹Institute for Cell Biology, Department of Immunology, Eberhard Karls University, Tübingen, Germany ²Department of Biostatistics and Bioinformatics, Duke University Medical Center, Durham, NC, USA ³Center for Cancer Immune Therapy, Copenhagen University Hospital, Herlev, Denmark ⁴Center for Clinical Transfusion Medicine, University Hospital, Tübingen, Germany ⁵Department of Clinical Oncology, Leiden University Medical Center, Leiden, the Netherlands ⁶Cancer Sciences Division, Southampton University Hospitals, Southampton, UK ⁷Translational Oncology, University Medical Center, Johannes Gutenberg University (TRON gGmbH), Mainz, Germany

Abstract

Multiparameter flow cytometry is an indispensable method for assessing antigen-specific T cells in basic research and cancer immunotherapy. Proficiency panels have shown that cell sample processing, test protocols and data analysis may all contribute to the variability of the results obtained by laboratories performing *ex vivo* T-cell immune monitoring. In particular, analysis currently relies on a manual, step-by-step strategy employing serial gating decisions based on visual inspection of one- or two-dimensional plots. It is therefore operator dependent and subjective. In the context of continuing efforts to support inter-laboratory T-cell assay harmonization, the Immunoguiding Program organized its third proficiency panel dedicated to the detection of antigen-specific CD8⁺ T cells by HLA-peptide multimer staining. We first assessed the contribution of manual data analysis to the variability of reported T-cell frequencies within a group of laboratories staining and analyzing the same cell samples with their own reagents and protocols. The results show that data analysis is a source of variation in the multimer assay outcome. To evaluate if an automated analysis approach can reduce variability of proficiency panel data, we used a hierarchical statistical mixture model to identify cell clusters. Challenges for automated analysis were the need to process non-standardized data sets from multiple centers, and

*Corresponding author: Cécile Gouttefangeas, Institute for Cell Biology, Department of Immunology, Eberhard Karls University, Auf der Morgenstelle 15, 72076 Tübingen, Germany. Tel +49 7071 2980994; Fax +49 7071 295653; cecile.gouttefangeas@uni-tuebingen.de.

#Current address: TRON gGmbH, University Medical Center, Johannes Gutenberg University Mainz, Mainz, Germany

†Current address: Institute for Inflammation Research (IIR), Copenhagen University Hospital, Denmark

§,+Both authors contributed equally to this work

Conflict of interest: the authors declare that they have no conflict of interest

the fact that the antigen-specific cell frequencies were very low in most samples. We show that this automated method can circumvent difficulties inherent to manual gating strategies and is broadly applicable for experiments performed with heterogeneous protocols and reagents.

Keywords

HLA-peptide multimer; proficiency panel; data analysis; flow cytometry gating; automated analysis

Introduction

HLA-class I peptide multimers (HLA-multimers) have evolved as essential tools for T-cell research [1, 2]. The relatively easy access to such reagents, together with recent technological developments in multiparameter flow cytometry (FCM), allow the enumeration and characterization of pathogen- or tumor-specific CD8⁺ cytotoxic T lymphocytes with improved precision and sensitivity [3–6]. HLA-multimers are also widely used to monitor T-cell responses induced by experimental vaccination against cancer [7–9]. So far however, only a few studies have demonstrated a positive correlation between the induction of anti-vaccine T cells as measured with HLA-multimers and clinical benefit [10–12].

A pre-requisite for evaluating the usefulness of HLA-multimer staining as a biomarker of clinical efficacy is that the assay is sufficiently sensitive to detect antigen-specific cells, and to do so robustly when performed in different laboratories. Several technical publications have provided optimized protocols and guidelines for cell staining, instrument set-up, test analysis and data reporting for FCM that can improve sensitivity and robustness [13–18]. Validation and standardization issues are also being addressed [19, 20].

Despite these efforts, methods still vary considerably between institutions and sometimes even between different investigators from the same institution. The results of immune monitoring studies performed at different laboratories are consequently difficult to compare and this delays progress in the field. One approach to improve the reproducibility of FCM results is to establish core facilities which can perform centralized immune monitoring for single- and multi-center studies [21]. Alternatively, one can coordinate inter-laboratory projects that aim to optimize and harmonize T-cell assays between institutions, so that similar results are achieved using laboratory-specific protocols and reagents [22].

The Immunoguiding Program of the Association for Cancer Immunotherapy (CIP, www.CIMT/workgroups/CIP) is an international working group which is actively developing the concept of assay harmonization. An important instrument of harmonization is the coordination of iterative proficiency panels for assays commonly used in immune monitoring; CIP proficiency panels have so far focused on the monitoring of antigen-specific CD8⁺ T cells by HLA-multimers, IFN γ ELISPOT and intracellular cytokine staining (ICS). Such panels organized by CIP and others [23–29] have shown that there is considerable heterogeneity of assay protocols and performance among laboratories (labs) with respect to many steps of the assay, including cell handling (freezing, thawing, culture),

test conditions, and data analysis. CIP and the Cancer Immunotherapy Consortium (CIC) demonstrated in two independent studies that inter-center variability in ICS decreased significantly when common gating instructions were provided to operators analyzing the same data sets [26, 30]. Centralized analyses also reduced heterogeneity [29, 31]. In light of these results, algorithms for automated recognition of cell clusters constitute attractive tools that are increasingly accessible to flow users [32–34].

Impact of the gating step on the outcome of the multimer assay has never been systematically addressed, and it is not known if different gating strategies would affect variability in the same way as demonstrated for the ICS. Hence, in continuation of our efforts to harmonize the HLA-multimer assay, we report here on the comparison of different analysis methods applied to data generated from the same cell samples by a heterogeneous group of seventeen laboratories. Both the detection rate and the frequencies of CD8⁺HLA-multimer⁺ T cells were assessed. The analysis methods evaluated consisted of two pre-selected gating procedures applied by the participants, a manual centralized analysis by a single investigator, and the utilization of a statistical mixture model for identification of rare cell clusters [35].

Materials and methods

The study was divided into three successive steps 1) inter-laboratory testing i.e. proficiency panel, 2) analysis of selected data sets by a single operator i.e. central analysis, and 3) computer-guided analysis (Figure 1). The methods section is structured according to the guidelines of the Minimal Information About T-cell Assays framework [36].

Proficiency panel

Preparation of reagents at the central laboratory—Samples: leukapheresis samples were obtained from healthy volunteers at the Department of Transfusion Medicine of the University Hospital of Tübingen after informed consent. Low resolution DNA HLA-class I typing and human cytomegalovirus (HCMV) serological status were known. The products were transported to the laboratory at room temperature (RT) and processed within 8 hrs. After dilution ¼ with sterile PBS, peripheral mononuclear cells (PBMC) were isolated by standard density gradient centrifugation (PAA, Pasching, Austria). PBMC were washed twice in PBS and counted using Trypan blue. For freezing, cells were resuspended gently in cold 90% heat-inactivated bovine serum (Hyclone, Bonn, Germany; serum was pre-tested for *in vitro* cell proliferation) plus 10% DMSO, and distributed in cryovials at $15\text{--}20 \times 10^6$ cells/1 ml on ice. Samples were transferred in freezing containers at -80°C then to a liquid nitrogen tank.

HLA-peptide multimer production: synthetic peptides representing two immunodominant, HLA-A*0201 restricted, virus-derived epitopes were used for HLA-monomer refolding, i.e. HCMV (pp65 495–503 NLVPMVATV) and Influenza A (Flu Matrix 58–66 GILGFVFTL) [23]. Fluorescent HLA-multimers were generated by co-incubating monomers with streptavidin-PE or -APC (Invitrogen, Darmstadt, Germany) at a 4:1 molar ratio. They were used for screening experiments either directly or after a freezing step at -80°C (in Tris 20

mM, 16% glycerol, 0.5% human serum albumin and 1X Complete Protease Inhibitor, Roche Diagnostics, Mannheim, Germany).

Pretesting and selection of the PBMC samples: PBMC were screened *ex vivo* with CMV or Flu HLA-multimers at the central lab, with an additional test being performed at the co-organizing lab. Stainings at the central lab were done in two steps following the CIP guidelines (www.cimt.eu/workgroups/CIP), with CD3-FITC or CD4-FITC (OKT3- or HP2/6-FITC, in-house labelling) and CD8-PE-Cy7 (clone SFCI21Thy2D3, Beckman Coulter, Krefeld, Germany) at pretested concentrations. Acquisition was performed on a FACS Canto II (BD Biosciences, Heidelberg, Germany) using Diva software. PMT channels and compensations were adjusted using unstained PBMC and fluorescent beads (BD Biosciences). Analysis was done with FlowJo version 7.2. PBMC from 5 donors (D1 to D5) with a total of 7 CMV- and Flu-specific T cell responses showing different levels of reactivity (n= 4 low i.e. < 0.1%, n=1 intermediate, and n= 2 high i.e. > 1% multimer⁺ in the CD8⁺ subset) were selected. One donor was HLA-A*02 negative and HCMV seropositive (D5), one was HLA-A*02 positive and HCMV seronegative (D1) and the remaining three were HLA-A*02 positive and HCMV seropositive (D2, D3, D4).

Inter-laboratory testing—Seventeen European labs (ID01 to ID17) participated. All labs self-reported as being experienced with at least 3 (up to 10)-color staining and the majority (n=16) had performed HLA-multimer staining assays before. Every laboratory received 2 PBMC vials from each donor (D1 to D5) as well as two HLA-A*0201-multimers aliquots (Flu- and CMV-multimers labeled with PE) shipped on dry ice and delivered via courier (n=14 labs) or collected at the central lab (n=3). Upon receipt, reagents had to be stored in liquid nitrogen (cells) or at -80°C (HLA-multimers) until use, and experiments had to be performed within 6 weeks. A panel guideline and a report form were distributed.

The main objective of the proficiency panel was to address the influence of data analysis on the variability of HLA-multimer frequency estimates. Two gating strategies were compared; gating strategy 1 was: lymphocyte gate (FSC/SSC), CD3⁺ gate in lymphocytes (CD3/CD8), CD8/multi mer in CD3⁺; gating strategy 2 was: low granularity CD3⁺ gate (CD3/SSC), lymphocyte gate (FSC/SSC) in CD3⁺, CD8/multimer in lymphocytes (Suppl. Figure 1). PBMC had to be tested twice, either immediately after thawing (15/17 laboratories counted living cells by Trypan blue exclusion, and two used a Guava counter), or after a cell-resting phase; results of immediate testing are shown in the manuscript. In addition, the utility of including a Fluorescence Minus One (FMO) staining (all mAb, but no multimer) for setting quadrants or gates was assessed. In total, each laboratory was expected to perform 15 *ex vivo* stainings (FMO, Flu-multimer and CMV-multimer i.e. 3 tests × 5 donors), each analyzed with the two predefined gating strategies.

Test conditions: reagents (except HLA-multimers), staining protocols and flow cytometer setup were not standardized but some procedures were mandatory following the recommendations of previous CIP proficiency panels [23]. Participants had to 1) use at least 1×10^6 (up to 2×10^6) PBMC per stain and acquire all cells contained in the sampling tubes, 2) include CD3 and CD8 mAb, 3) include a FMO control sample and 4) stain cells with the multimers for 30 min at RT before adding mAb (recommended concentration of

multimer was 5 µg/ml). Participants were free to 5) include or exclude a dump channel and/or a dead cell dye, 6) choose the mAb clones and fluorescent dyes and 7) choose the staining protocol and buffers used (a suggested protocol was provided).

Data reporting: parameters collected for central analysis were the thawing conditions, cell recovery and viability, the number of PBMC distributed per test, and the number of CD3⁺, CD8⁺ and multimer⁺ cells counted. The mAb (specificity, clone, company, fluorochrome), the use of a dump channel or of a dye for excluding dead cells, and the cytometer type were also recorded (Suppl. Table 1). Frequencies of antigen-specific cells were expressed as the number and % of multimer⁺ among CD3⁺CD8⁺ cells using analysis strategies 1 or 2.

Central data analysis

Analysis of data generated by the individual participants: all dot- or pseudocolor-plots of HLA-multimer stains were analyzed by a group of 5 experts, who scored the tests with “0” (no CD8⁺multimer⁺ cell population), “1” (plausible CD8⁺multimer⁺ cell population) or “2” (obvious CD8⁺multimer⁺ cell population) for a maximum aggregate score of 10. Stains with an aggregate score of 6 or more were considered positive (i.e. detection of CD8⁺multimer⁺ cells). In addition, the central analysis also evaluated if the gating strategies used conformed to the provided guideline and if appropriate intermediate steps for the dump or dead cell exclusion were added when applicable. Data sets were excluded from the final analysis if the test results were not accurately reported by the lab, or if the frequencies reported were obvious outliers, suggesting technical problems with performance of the assay (Suppl. Table 2). One lab recorded “false positive” stains (CMV- and Flu-multimer binding cells in the HLA-A*02⁻ donor D5), presumably arising from artifacts due to technical problems during cell staining or sample acquisition (not shown).

Final results were generated based on the information provided by the participants. Detection rates were defined as: [number of detected responses as per central review/number of expected responses = 7] * 100; antigen-specific T-cell frequencies are given in 1 CD8⁺multimer⁺ T cell in × CD3⁺CD8⁺ lymphocytes. For statistical analysis, mean, median, standard deviations and % coefficient of variation (% CV = SD/mean * 100) were calculated. For group comparisons, a two-tailed paired Student t-test was performed using GraphPad Prism 4.02. Relative differences are expressed as $\frac{(x-y)}{y} * 100$.

Central analysis of FCS files: re-analysis of selected data sets was performed at the central lab by one investigator on the same day, using FCS files provided by 14 of the participants (ID04, ID13 and ID24 were excluded) and Flow Jo v7.6.3. Gating strategy 1 was applied as per panel guideline. The number of parameters included was identical to that used by the individual laboratories, i.e. gating steps for dump channel and exclusion of dead cells were included when applicable. Five representative specificities, i.e. D2 Flu and CMV, D3 Flu and CMV, and D4 Flu were examined. The Fluorescence compensation was slightly modified for two labs. FSC/SSC lymphocyte gates were kept constant within each donor and CD8⁺multimer⁺ lymphocytes were identified using quadrants or gates set on the CD8 vs. multimer dot-plots.

Automated analysis of FCS files

We used a hierarchical statistical Gaussian mixture model that we recently described [35] to analyse FCS files (FMO and multimer tests) obtained from 14/17 participants. File pre-processing included compensation, the application of the same biexponential transformation to all data sets, and extraction of data of only the subset of channels common to mAb panels from all laboratories for the statistical model fitting. The hierarchical mixture model was fitted to all data samples from a single laboratory simultaneously, and the same procedure was then repeated for each separate laboratory. A complete description of the pre-processing and processing pipeline is provided in Supplementary Materials (see download link below). More than 98% of the FCS files (207/210) could be analyzed, demonstrating the feasibility of applying such program for non-standardized data. Three files from one laboratory were excluded because they were recorded on a different scale (max range = 1024) than the other twelve files (max range = 262144), violating the assumptions of the hierarchical modeling software.

The automated analysis differed from manual/central analysis in the following ways: 1) data annotation needed to be preprocessed, 2) the original fluorescence compensation was used for all laboratories, 3) only information from markers common to all laboratories was used for the model fitting, although viability information was used to screen out dead cell events in preprocessing for laboratories that had included a marker for viability, 4) positivity was determined using an arbitrary event count rather than by expert consensus. The positivity threshold was set at 10 events detected, so that the rate of false positives was very low and close to that obtained by manual analysis (for this data set, 2 and 3 false positives were recorded for individual and automated approaches, respectively), possibly at the risk of missing very rare events.

Identification of antigen-specific cell clusters was performed using a small set of conservative heuristic rules: 1) the probability density evaluated on the fitted mixture model at cluster center is at least 5-fold higher for CMV- or Flu-multimer than for FMO for the *same* donor, 2) the FSC intensity at cluster center is less than the mean across all samples, 3) the SSC intensity at cluster center is less than the mean across all samples, 4) the CD3 fluorescence intensity at cluster center is greater than the mean across all samples, 5) the CD8 fluorescence intensity at cluster center is greater than the mean across all samples, 6) the multimer fluorescence intensity at cluster center is greater than the mean + 2 SD across all samples. Rule 1) filters for events that are likely to be antigen-specific, rules 2) and 3) exclude clusters that are more likely to be monocytes or macrophages, and rules 4), 5) and 6) filter for clusters that are likely to be comprised of CD3⁺CD8⁺multimer⁺ cells. However, because of a large degree of variability in the separation of multimer⁺ from multimer^{neg} events across samples from different laboratories, visual inspection was used to select only multimer positive clusters that were consistently present in multiple samples. In particular, a CD8⁺multimer^{low} cluster was sometimes observed in CMV and Flu-multimer stained samples but not in the FMO samples; these were considered by the experts to be non-antigen specific even though they met the criteria listed above. These non-specific background multimer events were typically assigned into different clusters from truly multimer positive events, allowing simple exclusion after visual inspection.

An executable IPython Notebook (CIP_Analysis.ipynb) and a read-only version (CIP_Analysis.html) showing the full analysis pipeline from the FCS files to the frequency estimates are provided as Suppl. Materials (<https://duke.box.com/s/9kphs2300kumjyowjq06>), along with instructions for installing the pre-requisite software (README.md file). The FCS files from one participating lab are uploaded at the FlowRepository website (www.flowrepository.org, experiment: CIP ID03 panel, ID: FR-FCM-ZZER).

Laboratory environment

The central lab operated under exploratory research conditions using established protocols and trained personal. Participants followed their own procedures and most of them (n= 14) had taken part in previous HLA-multimer proficiency panels organized by CIP or CIC. Three other labs had never been engaged in such activities.

Results

General panel characteristics

Centers used local protocols for thawing, counting, resting and staining the cells. The median number of living cells after thawing was 12.5×10^6 per vial with 85% cell viability, and most laboratories used at least 1×10^6 cells per stain as per panel requirement (mean: 1.6×10^6 cells/test, range $0.9 - 2.7 \times 10^6$). In addition to the mandatory markers (CD3, CD8, multimers), some included a dump channel mAb and/or a dye for excluding dead cells (Suppl. Tables 1a–c).

In previous panels, we had noticed that inaccurate placement of quadrants or gates on the CD8/multimer dot-plots generally results in erroneous overestimation of the percentage of multimer⁺ cells among the CD8⁺ cell subset [23, 37]. A fluorescence Minus One staining (FMO for the PE i.e. multimer channel) was therefore introduced to guide participants in placing markers or gates. Central review of CD8/FMO-PE dot-plots showed that all but two labs analyzed the same number of cells in the FMO and in the multimer-containing tests; quadrants/gates were adjusted either “low” (close to the CD8⁺multimer^{neg} population) or “high” (close to the CD8⁺multimer⁺ population) on the PE channel (see examples in Suppl. Figure 3); no serious mistakes were observed, except for one lab. Of note, a CD8⁺multimer^{low} population was often seen in the multimer-containing tests that was absent from the FMO stains, probably due to weak binding of the multimers independently of the antigen-specificity. Thus, a FMO test may not constitute an ideal negative control for the multimer assay.

HLA-multimer assay outcome may be influenced by the gating strategy

From the screening experiments, it was expected that a maximum of 7 antigen-specificities ($4 \times$ Flu and $3 \times$ CMV) could be detected in all five PBMC samples. The majority of the labs were able to detect multimer⁺CD8⁺ T cells for the 5 T-cell populations present at the highest frequencies (D3 CMV, D4 CMV, D2 Flu, D1 Flu and D4 Flu) using either of the two gating strategies (Figure 2a and Suppl. Table 2c). The group performance was, however, drastically decreased for the two lowest frequencies, i.e. D3 Flu and D2 CMV (both <

0.04%, Figure 2a). For both specificities, the best detection rates were obtained when applying gating strategy 1. A detailed analysis (Figure 2b) revealed that many labs performed equally well with the two analysis conditions; however, 5 participants missed between 1 and 2 T-cell responses when applying strategy 2 as compared to strategy 1 (e.g. ID08, ID14). Of note, gating strategy 1 led to significantly higher CD8⁺ T cell numbers than gating strategy 2 (mean and median for the group = 106,500 vs. 102,000 and 99,000 vs 90,500 respectively, Suppl. Figure 2). However, we found no correlation between the mean numbers of CD8⁺ T cells and the individual lab performances, suggesting that sufficient numbers of cells were included in most tests to support detection of even low-frequency T cells. These observations suggest that operator analysis decisions may impact on the test outcome and sensitivity of the multimer assay.

Next, we assessed the frequency of antigen-specific cells (for absolute numbers, see Suppl. Table 2a–c). As shown in Figure 3, the inter-laboratory variation (% CV) was limited for antigen-specific CD8⁺ T cells present at high (>1%, D3 CMV and D4 CMV) or intermediate (0.1% < D2 Flu < 1%,) frequencies among PBMC, and calculated mean frequencies appear to be similar with the two gating strategies ($p > 0.05$). Variability was increased for most low-frequency T cells (< 0.1%), especially in one case (D3 Flu: % CV = 104 with gating strategy 1), confirming our previous observations [23]. Interestingly, for the lowest T-cell frequency that could be compared, i.e. D2 CMV, the variability within the group was smaller after analysis with gating strategy 1 as compared to gating strategy 2 (Figure 3: % CV 53 to 87, = 64%).

Overall, superior results were obtained when applying a classical 3-step analysis strategy including a standard FSC/SSC scatter plot for identifying the lymphocyte subset. More importantly, we concluded that the gating procedure used for data analysis is a source of inter-center variability in detecting and enumerating HLA-multimer binding cells. This is especially apparent for rare antigen-specific cells (less than 0.04% in this panel), which is of high relevance for tumor immunologists since the frequencies of tumor-directed T cells are expected to be low in many settings.

Central gating analysis decreases inter-laboratory variability

To further evaluate the contribution of gating decisions on variability, we performed a central re-analysis of selected FCS files. A single user used gating strategy 1 to analyze 5 representative tests with low (D2 CMV, D3 Flu, D4 Flu), intermediate (D2 Flu) and high (D3 CMV) frequencies of antigen-specific T cells provided by 14 panel participants. Frequencies of CD8⁺multimer⁺ T cells were determined only for stains defined as positive during the central review process. Figure 4 displays the results of this central analysis (C) as compared to those compiled from the individual participants (I) for the same tests. Strikingly, 3 out of the 5 calculated % CV were substantially decreased (= 72–114 %) after central analysis, and all 3 were related to detection of low frequency T cells (D2 CMV, D3 Flu and D4 Flu). For the two other antigen/donor combinations, the CV remained essentially unchanged. One obvious reason for discrepancy between the deduced T-cell frequencies was the setting of the lower bound of the multimer-channel gate to be distant from the CD8⁺multimer^{neg} subset in the central analysis; some participants chose a position

closer to this multimer^{neg} cell population, thereby including diffuse CD8⁺PE^{low} events in the multimer⁺ population that most likely reflect non-specific binding. Hence, the mean frequencies were generally slightly lower after central analysis.

These analyses demonstrate that not only protocol parameters, but also individual gating choices contribute to the inter-laboratory variability of the HLA-multimer assay and corroborate previous findings for the ICS assay [26, 30].

Automated analysis of heterogeneous data sets is feasible, allows detection of even rare antigen-specific T cells, and leads to results comparable to that of manual gating

We next explored the possibility of identifying multimer-binding cells using a hierarchical statistical mixture model for automated, multidimensional recognition of cell clusters that allows robust, linear and sensitive (< 0.02% of all cells) detection of antigen-specific CD8⁺ T cells spiked at defined numbers in autologous PBMC [35]. Examples of complete automated analyses and deduced CD8⁺multimer⁺ T-cell frequencies are given in Suppl. Figure 4a and Suppl. Table 3.

The results obtained after automated processing of the FCS files were consistent with those generated by individual or central manual analyses applying gating strategy 1 (Figure 5a and Table 1). The detection rates for CD8⁺multimer⁺ T cells were equivalent for 4/7 low to high frequency specificities; for the two rare specificities, it was increased in one case and decreased in the other (D2 CMV: 38% vs. 54% and D3 Flu: 38 vs. 23% for the central vs. automated assessments, respectively). The mean and median frequencies assessed by the automated analysis were in essence lower than for the manually generated data, suggesting a more stringent selection of multimer⁺ T cells by the statistical mixture model. As compared to the lab-individual results, the % CV was notably decreased for 1 stain (D2 CMV, = 58%), essentially unchanged for 4 tests (D2 Flu, D3 Flu, D4 CMV and D4 Flu, < 40%), and increased for 2 stains (D1 Flu and D3 CMV, > 50%). Notably, for D3 CMV, test of one lab (i.e. ID10) contributed greatly to this effect, with HLA-multimer⁺ cells of low intensity which were included in the individual (and central) analysis but ignored from the software-guided identification (% CV 29 vs. 158 and 29 vs. 60 with or without including ID10 staining, respectively (Table 1 and Suppl. Figure 4b). Further dot-plot inspection revealed that what can be interpreted as low intensity, non-specific background is more easily “ignored” by the automated analysis than by visual inspection. Also, within “diffuse” cell populations disregarded by visual analysis, distinct cell clusters may be captured by automated recognition, as illustrated in Figure 5b and Suppl. Figure 4b–c. Importantly, this suggests that the automated analysis may be not only more objective (non-biased by operator), but also more selective than the manual analysis.

In summary, these results highlight the considerable potential of computer-based approaches in general and of the hierarchical modeling software in particular for objective, centralized analysis of rare cell events in FCM data generated by single laboratories over time or by multicenter studies.

Discussion

Fluorescent HLA-multimers are excellent tools to detect antigen-specific CD8⁺ T cells independently of their function. The method is considered to be robust and has been shown to be very sensitive in the hands of experts (described limit of detection is below 0.02% = 1 multimer⁺ among 5000 CD8⁺ cells) [6, 38]. Considering these favorable characteristics, it was rather unexpected that earlier proficiency panels organized by CIP, CIC and others, reported considerable inter-laboratory variability in detecting rare antigen-specific T cells. Some parameters of cell handling that influence assay performance were identified, including the number of cells per test and whether a dump channel was used [23, 28, 37, 39]. Furthermore, gating strategies and marker settings varied, including cases of incorrect cell-subset gating that may contribute to the variability of the results [23]. Based on these findings, our study focused on the impact of individual gating decisions on assay outcome. Furthermore, we evaluated the feasibility of an alternative, automated method for analyzing data generated by a group of non-standardized laboratories.

The inter-laboratory variability (% CV) in this panel was decreased as compared to our previous experience; we believe that this is because the majority of the participants had taken part in proficiency panels before, underlining the value of inter-laboratory harmonization. Despite an excellent overall performance, differences (detection rates and frequencies of CD8⁺ HLA-multimer⁺ T cells) were clearly discernible between participants, especially for very low (< 0.04%) frequencies of antigen-specific T cells.

Manual gating relies on the skill and experience of individual users, with few basic rules (such as gating sequences) that may vary between labs and often between operators. Analysis of FCS files by a single operator (central analysis) reduced the inter-center variability, indicating that gating indeed plays an essential role in the HLA-multimer staining outcome. These results reinforce previous observations made for ICS, where it was shown that single-operator analysis or harmonization of the gating strategy among labs analyzing the same set of FCS files reduced variability considerably [26, 29–31].

However, central analysis remains highly dependent on operator judgment, is time-consuming especially for multiparametric data sets, and may simply not be feasible. Automated methods in which cell subsets are directly quantified by machine algorithms are therefore attractive tools to overcome the problem of operator subjectivity. A number of algorithms for cell-subset identification have been recently proposed [32, 34]. With such approaches, analysis of cell populations is in principle independent of any sequential gating strategy, since clusters are identified using the full set of data dimensions simultaneously.

We have developed a hierarchical statistical mixture that allows detection of low frequency antigen-specific T cells in spiked samples with a sensitivity and reproducibility comparable with that of cytometry experts [35]. Here we show that this program is also suited for analyzing FCS files generated at different centers using unmanipulated PBMC samples and HLA-peptide multimers, but different protocols, staining reagents, and cytometer settings. Statistical mixture modeling may therefore be used not only for individual purposes, but also for analyzing data generated by proficiency panels or multi-centric studies. A similar, non-

hierarchical Gaussian mixture model was also recently shown to identify rare antigen-specific cell subsets with high sensitivity [40].

The detection rates obtained with the modeling were overall in the range of that of the manual analysis; however, the automated analysis can occasionally result in higher variability as inclusion/exclusion applies to entire clusters rather than individual events potentially amplifying differences between samples. Also, mean frequencies were mainly decreased after automated analysis. Unlike manual gating where events must be discriminated using only two markers at a time, the multivariate clustering method simultaneously uses information from multiple markers in order to determine cluster membership: this facilitates the separation of cell subsets that can only be distinguished on the basis of multiple markers simultaneously. At the same time, events with similar characteristics in all dimensions will be clustered together, and not excluded by “hard” gating boundaries because they are relative outliers in a 2- dimensional dot-plot. In line with this, we observed that low-intensity multimer⁺ events that are often interpreted as background by immunologists are classified to different clusters than the distinct high intensity events, allowing a clean separation either using a rule-based method or by visual inspection (Suppl. Figure 4c). Conversely, in some cases (Figure 5b, Suppl. Figure 4c), the automated analysis revealed CD3⁺CD8⁺multimer⁺ events that would be excluded by a manual scatter gating (FSC/SSC) as non-specific binding of the reagents to debris or dying/dead cells. Notably, these events only appeared for samples that also contain multimer-binding CD8⁺ cells within the standard FSC/SSC lymphocyte gate. It is known that binding of HLA-multimers to cognate TCR may induce activation-driven cell death, a process which is temperature and likely protocol-dependent [41, 42]. If so, these events correspond to “true positive” antigen-specific cells and should not be excluded from the analysis.

Our data support current efforts to improve the availability of robust and sensitive automated tools for flow cytometry analysis and to make these accessible to the broader community. The potential payoff is clear: currently, data analysis takes up a large fraction of time for staff members in immune monitoring laboratories; automated tools have the potential to dramatically lower the need for manual analysis of standard Ab panels, freeing up time for other activities. However, automated analysis still faces significant challenges before it can be widely adopted for routine analysis of multicenter data.

First, in the absence of consistent and sufficiently complete annotation, there is a large preprocessing burden before the data can be fed to automated analysis pipelines. In this panel, laboratories were allowed to use their own annotation, leading to an extensive variation in the naming of FCS files and channels. In some cases, the FCS metadata did not provide information on the channel marker, and cross-referencing with data provided by the laboratories was necessary. Such problems can be overcome if laboratories develop consistent data annotation protocols that minimize human error (e.g. use of barcoding), and bioinformaticians develop tools to harmonize lab annotations, for example via the intermediary of appropriate ontologies.

Second, it is challenging to map event clusters to their biologically meaningful nomenclature for non-standardized data from multiple laboratories. We described a heuristic set of filter

criteria based on statistical comparison across data sets and controls to identify candidate antigen-specific clusters, but final visual inspection was still necessary to disambiguate clusters that were CD8⁺ and multimer (low) into non-specific or antigen-positive subsets. We suggest that a pragmatic solution is to move away from a fully automated solution, and instead use the automated analysis pipeline to detect candidate clusters for targeted cell subsets using heuristic rules, and then use visual inspection to accept or reject the candidates based on subtle properties that are currently not captured by our algorithms (e.g. shape, compactness, consistency across different data sets, known biological context of the sample). Third, the statistical mixture model software must be operated by a computer expert both for initial parameter settings and for the following analysis process; a browser-based version that is more accessible to average flow analysts, both in the context of multicenter studies or for center-specific projects, is currently being evaluated by CIP core laboratories [44].

Finally, some recommendations for data analysis can be added to our HLA-multimer guideline (www.CIMT/workgroup/CIP): 1) run systematic comparisons of gating strategies when establishing or optimizing new mAb panels [39, 43], 2) carefully set and audit quadrant/gate placement for selecting CD8⁺multimer⁺ cells. A FMO staining -performed on the same number of cells as the test stains- may help inexperienced users but cannot be utilized for setting thresholds accurately because it provides no information on non-specific multimer binding. As an alternative, irrelevant HLA-multimers have been introduced as controls in subsequent CIP proficiency panels, in line with recently published recommendations [1, 28].

In summary, the results of this proficiency panel extend our knowledge on parameters which impact performance of the HLA-multimer assay, and underline the importance of data analysis strategies. Automated methods are attractive tools that bear the promise to reduce inter-laboratory variability of flow cytometry and increase robustness, accuracy and comparability of results, but much needs to be done before fully automated methods can be used for routine analysis of non-standardized multi-center data.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

CIP and C. Chan are supported by a grant of the Wallace Coulter Foundation (Miami, Florida, USA). C. Gouttefangeas receives a grant from the Deutsche Forschungsgemeinschaft SFB 685/Z5. M. J. P. Welters is supported by a grant from the Dutch Cancer Society (UL 2009-4400). We thank the Association for Cancer Immunotherapy (CIMT) for its patronage and financial support, all panel participants listed below, and lab ID01 for providing FCS files for the FlowRepository database.

List of abbreviations

CIP	CIMT Immunoguiding Program
CIMT	Association for Cancer Immunotherapy

CV	coefficient of variation
FCM	flow cytometry
FCS file	Flow Cytometry Standard format for data files
mAb	monoclonal antibody
PBMC	peripheral blood mononuclear cells

Reference List

1. Davis MM, Altman JD, Newell EW. Interrogating the repertoire: broadening the scope of peptide-MHC multimer analysis. *Nat Rev Immunol.* 2011; 11:551–558. [PubMed: 21760610]
2. Altman JD, Moss PA, Goulder PJ, Barouch DH, McHeyzer-Williams MG, Bell JI, McMichael AJ, Davis MM. Phenotypic analysis of antigen-specific T lymphocytes. *Science.* 1996; 274:94–96. [PubMed: 8810254]
3. Appay V, Dunbar PR, Callan M, Klenerman P, Gillespie GM, Papagno L, Ogg GS, King A, Lechner F, Spina CA, Little S, Havlir DV, Richman DD, Gruener N, Pape G, Waters A, Easterbrook P, Salio M, Cerundolo V, McMichael AJ, Rowland-Jones SL. Memory CD8+ T cells vary in differentiation phenotype in different persistent virus infections. *Nat Med.* 2002; 8:379–385. [PubMed: 11927944]
4. Almeida JR, Price DA, Papagno L, Arkoub ZA, Sauce D, Bornstein E, Asher TE, Samri A, Schnuriger A, Theodorou I, Costagliola D, Rouzioux C, Agut H, Marcelin AG, Douek D, Autran B, Appay V. Superior control of HIV-1 replication by CD8+ T cells is reflected by their avidity, polyfunctionality, and clonal turnover. *J Exp Med.* 2007; 204:2473–2485. [PubMed: 17893201]
5. Baitsch L, Baumgaertner P, Devedre E, Raghav SK, Legat A, Barba L, Wieckowski S, Bouzourene H, Deplancke B, Romero P, Rufer N, Speiser DE. Exhaustion of tumor-specific CD8(+) T cells in metastases from melanoma patients. *J Clin Invest.* 2011; 121:2350–2360. [PubMed: 21555851]
6. Andersen RS, Kvistborg P, Frosig TM, Pedersen NW, Lyngaa R, Bakker AH, Shu CJ, Straten P, Schumacher TN, Hadrup SR. Parallel detection of antigen-specific T cell responses by combinatorial encoding of MHC multimers. *Nat Protoc.* 2012; 7:891–902. [PubMed: 22498709]
7. Slingluff CL Jr, Petroni GR, Olson WC, Smolkin ME, Ross MI, Haas NB, Grosh WW, Boisvert ME, Kirkwood JM, Chianese-Bullock KA. Effect of granulocyte/macrophage colony-stimulating factor on circulating CD8+ and CD4+ T-cell responses to a multi-peptide melanoma vaccine: outcome of a multicenter randomized trial. *Clin Cancer Res.* 2009; 15:7036–7044. [PubMed: 19903780]
8. Coulie PG, Karanikas V, Colau D, Lurquin C, Landry C, Marchand M, Dorval T, Brichard V, Boon T. A monoclonal cytolytic T-lymphocyte response observed in a melanoma patient vaccinated with a tumor-specific antigenic peptide encoded by gene MAGE-3. *Proc Natl Acad Sci U S A.* 2001; 98:10290–10295. [PubMed: 11517302]
9. Filipazzi P, Pilla L, Mariani L, Patuzzo R, Castelli C, Camisaschi C, Maurichi A, Cova A, Rigamonti G, Giardino F, Di Florio A, Ascoli M, Frati P, Sovena G, Squarcina P, Maio M, Danielli R, Chiarion-Sileni V, Villa A, Lombardo C, Tragni G, Santinami M, Parmiani G, Rivoltini L. Limited induction of tumor cross-reactive T cells without a measurable clinical benefit in early melanoma patients vaccinated with human leukocyte antigen class I-modified peptides. *Clin Cancer Res.* 2012; 18:6485–6496. [PubMed: 23032742]
10. Walter S, Weinschenk T, Stenzl A, Zdrojowy R, Pluzanska A, Szczylik C, Staehler M, Brugger W, Dietrich PY, Mendrzyk R, Hilf N, Schoor O, Fritsche J, Mahr A, Maurer D, Vass V, Trautwein C, Lewandrowski P, Flohr C, Pohla H, Stanczak JJ, Bronte V, Mandruzzato S, Biedermann T, Pawelec G, Derhovanessian E, Yamagishi H, Miki T, Hongo F, Takaha N, Hirakawa K, Tanaka H, Stevanovic S, Frisch J, Mayer-Mokler A, Kirner A, Rammensee HG, Reinhardt C, Singh-Jasuja H. Multi-peptide immune response to cancer vaccine IMA901 after single-dose cyclophosphamide associates with longer patient survival. *Nat Med.* 2012; 18:1254–1261. [PubMed: 22842478]

11. De Vries IJ, Bernsen MR, Lesterhuis WJ, Scharenborg NM, Strijk SP, Gerritsen MJ, Ruiter DJ, Figdor CG, Punt CJ, Adema GJ. Immunomonitoring tumor-specific T cells in delayed-type hypersensitivity skin biopsies after dendritic cell vaccination correlates with clinical outcome. *J Clin Oncol*. 2005; 23:5779–5787. [PubMed: 16110035]
12. Van Tendeloo VF, Van DV, Van Driessche A, Cools N, Anguille S, Ladell K, Gostick E, Vermeulen K, Pieters K, Nijs G, Stein B, Smits EL, Schroyens WA, Gadisseur AP, Vrelust I, Jorens PG, Goossens H, De Vries IJ, Price DA, Oji Y, Oka Y, Sugiyama H, Berneman ZN. Induction of complete and molecular remissions in acute myeloid leukemia by Wilms' tumor 1 antigen-targeted dendritic cell vaccination. *Proc Natl Acad Sci U S A*. 2010; 107:13824–13829. [PubMed: 20631300]
13. Wooldridge L, Lissina A, Cole DK, van den Berg HA, Price DA, Sewell AK. Tricks with tetramers: how to get the most from multimeric peptide-MHC. *Immunology*. 2009; 126:147–164. [PubMed: 19125886]
14. Herzenberg LA, Tung J, Moore WA, Herzenberg LA, Parks DR. Interpreting flow cytometry data: a guide for the perplexed. *Nat Immunol*. 2006; 7:681–685. [PubMed: 16785881]
15. Mahnke Y, Chattopadhyay P, Roederer M. Publication of optimized multicolor immunofluorescence panels. *Cytometry A*. 2010; 77:814–818. [PubMed: 20722004]
16. Lee JA, Spidlen J, Boyce K, Cai J, Crosbie N, Dalphin M, Furlong J, Gasparetto M, Goldberg M, Goralczyk EM, Hyun B, Jansen K, Kollmann T, Kong M, Leif R, McWeeney S, Moloshok TD, Moore W, Nolan G, Nolan J, Nikolich-Zugich J, Parrish D, Purcell B, Qian Y, Selvaraj B, Smith C, Tchuvatkina O, Wertheimer A, Wilkinson P, Wilson C, Wood J, Zigon R, Scheuermann RH, Brinkman RR. MIFlowCyt: the minimum information about a Flow Cytometry Experiment. *Cytometry A*. 2008; 73:926–930. [PubMed: 18752282]
17. Chattopadhyay PK, Melenhorst JJ, Ladell K, Gostick E, Scheinberg P, Barrett AJ, Wooldridge L, Roederer M, Sewell AK, Price DA. Techniques to improve the direct ex vivo detection of low frequency antigen-specific CD8+ T cells with peptide-major histocompatibility complex class I tetramers. *Cytometry A*. 2008; 73:1001–1009. [PubMed: 18836993]
18. Perfetto SP, Ambrozak D, Nguyen R, Chattopadhyay PK, Roederer M. Quality assurance for polychromatic flow cytometry using a suite of calibration beads. *Nat Protoc*. 2012; 7:2067–2079. [PubMed: 23138348]
19. Horton H, Thomas EP, Stucky JA, Frank I, Moodie Z, Huang Y, Chiu YL, McElrath MJ, De Rosa SC. Optimization and validation of an 8-color intracellular cytokine staining (ICS) assay to quantify antigen-specific T cells induced by vaccination. *J Immunol Methods*. 2007; 323:39–54. [PubMed: 17451739]
20. Maecker HT, McCoy JP, Nussenblatt R. Standardizing immunophenotyping for the Human Immunology Project. *Nat Rev Immunol*. 2012; 12:191–200. [PubMed: 22343568]
21. Maecker HT, McCoy JP Jr, Amos M, Elliott J, Gaigalas A, Wang L, Aranda R, Banchereau J, Boshoff C, Braun J, Korin Y, Reed E, Cho J, Hafler D, Davis M, Fathman CG, Robinson W, Denny T, Weinhold K, Desai B, Diamond B, Gregersen P, Di Meglio P, Nestle FO, Peakman M, Villanova F, Ferbas J, Field E, Kantor A, Kawabata T, Komocsar W, Lotze M, Nepom J, Ochs H, O'Lone R, Phippard D, Plevy S, Rich S, Roederer M, Rotrosen D, Yeh JH. A model for harmonizing flow cytometry in clinical trials. *Nat Immunol*. 2010; 11:975–978. [PubMed: 20959798]
22. van der Burg SH, Kalos M, Gouttefangeas C, Janetzki S, Ottensmeier C, Welters MJ, Romero P, Britten CM, Hoos A. Harmonization of immune biomarker assays for clinical studies. *Sci Transl Med*. 2011; 3:108ps44.
23. Britten CM, Gouttefangeas C, Welters MJ, Pawelec G, Koch S, Ottensmeier C, Mander A, Walter S, Paschen A, Muller-Berghaus J, Haas I, Mackensen A, Kollgaard T, thor SP, Schmitt M, Giannopoulos K, Maier R, Veelken H, Bertinetti C, Konur A, Huber C, Stevanovic S, Wolfel T, van der Burg SH. The CIMT-monitoring panel: a two-step approach to harmonize the enumeration of antigen-specific CD8+ T lymphocytes by structural and functional assays. *Cancer Immunol Immunother*. 2008; 57:289–302. [PubMed: 17721783]
24. Mander A, Gouttefangeas C, Ottensmeier C, Welters MJ, Low L, van der Burg SH, Britten CM. Serum is not required for ex vivo IFN-gamma ELISPOT: a collaborative study of different

- protocols from the European CIMT Immunoguiding Program. *Cancer Immunol Immunother.* 2010; 59:619–627. [PubMed: 20052465]
25. Filbert H, Attig S, Bidmon N, Renard BY, Janetzki S, Sahin U, Welters MJ, Ottensmeier C, van der Burg SH, Gouttefangeas C, Britten CM. Serum-free freezing media support high cell quality and excellent ELISPOT assay performance across a wide variety of different assay protocols. *Cancer Immunol Immunother.* 2013; 62:615–627. [PubMed: 23138872]
 26. Welters MJ, Gouttefangeas C, Ramwadhoebe TH, Letsch A, Ottensmeier CH, Britten CM, van der Burg SH. Harmonization of the intracellular cytokine staining assay. *Cancer Immunol Immunother.* 2012; 61:967–978. [PubMed: 22714399]
 27. Janetzki S, Panageas KS, Ben Porat L, Boyer J, Britten CM, Clay TM, Kalos M, Maecker HT, Romero P, Yuan J, Kast WM, Hoos A. Results and harmonization guidelines from two large-scale international Elispot proficiency panels conducted by the Cancer Vaccine Consortium (CVC/SVI). *Cancer Immunol Immunother.* 2008; 57:303–315. [PubMed: 17721781]
 28. Attig S, Price L, Janetzki S, Kalos M, Pride M, McNeil L, Clay T, Yuan J, Odunsi K, Hoos A, Romero P, Britten CM. A critical assessment for the value of markers to gate-out undesired events in HLA-peptide multimer staining protocols. *J Transl Med.* 2011; 9:108. [PubMed: 21745365]
 29. Jaimes MC, Maecker HT, Yan M, Maino VC, Hanley MB, Greer A, Darden JM, D'Souza MP. Quality assurance of intracellular cytokine staining assays: analysis of multiple rounds of proficiency testing. *J Immunol Methods.* 2011; 363:143–157. [PubMed: 20727897]
 30. McNeil LK, Price L, Britten CM, Jaimes M, Maecker H, Odunsi K, Matsuzaki J, Staats JS, Thorpe J, Yuan J, Janetzki S. A harmonized approach to intracellular cytokine staining gating: Results from an international multicenter proficiency panel conducted by the Cancer Immunotherapy Consortium (CIC/CRI). *Cytometry A.* 2013; 83:728–738. [PubMed: 23788464]
 31. Maecker HT, Rinfret A, D'Souza P, Darden J, Roig E, Landry C, Hayes P, Birungi J, Anzala O, Garcia M, Harari A, Frank I, Baydo R, Baker M, Holbrook J, Ottinger J, Lamoreaux L, Epling CL, Sinclair E, Suni MA, Punt K, Calarota S, El-Bahi S, Alter G, Maila H, Kuta E, Cox J, Gray C, Altfeld M, Nougarede N, Boyer J, Tussey L, Tobery T, Bredt B, Roederer M, Koup R, Maino VC, Weinhold K, Pantaleo G, Gilmour J, Horton H, Sekaly RP. Standardization of cytokine flow cytometry assays. *BMC Immunol.* 2005; 6:13. [PubMed: 15978127]
 32. Chan C, Feng F, Ottinger J, Foster D, West M, Kepler TB. Statistical mixture modeling for cell subtype identification in flow cytometry. *Cytometry A.* 2008; 73:693–701. [PubMed: 18496851]
 33. Frelinger J, Ottinger J, Gouttefangeas C, Chan C. Modeling flow cytometry data for cancer vaccine immune monitoring. *Cancer Immunol Immunother.* 2010; 59:1435–1441. [PubMed: 20563720]
 34. Aghaepour N, Finak G, Dougall D, Khodabakhshi AH, Mah P, Obermoser G, Spidlen J, Taylor I, Wuensch SA, Bramson J, Eaves C, Weng AP, Iii ES, Ho K, Kollmann T, Rogers W, De Rosa S, Dalal B, Azad A, Pothan A, Brandes A, Bretschneider H, Bruggner R, Finck R, Jia R, Zimmerman N, Linderman M, Dill D, Nolan G, Chan C, Khettabi FE, O'Neill K, Chikina M, Ge Y, Sealfon S, Sugar I, Gupta A, Shooshtari P, Zare H, De Jager PL, Jiang M, Keilwagen J, Maisog JM, Luta G, Barbo AA, Majek P, Vilcek J, Manninen T, Huttunen H, Ruusuvoori P, Nykter M, McLachlan GJ, Wang K, Naim I, Sharma G, Nikolic R, Pyne S, Qian Y, Qiu P, Quinn J, Roth A, Meyer P, Stolovitzky G, Saez-Rodriguez J, Norel R, Bhattacharjee M, Biehl M, Bucher P, Bunte K, Di Camillo B, Sambo F, Sanavia T, Trifoglio E, Toffolo G, Dimitrieva S, Dreos R, Ambrosini G, Grau J, Grosse I, Posch S, Guex N, Keilwagen J, Kursu M, Rudnicki W, Liu B, Maienschein-Cline M, Manninen T, Huttunen H, Ruusuvoori P, Nykter M, Schneider P, Seifert M, Strickert M, Vilar JM, Hoos H, Mosmann TR, Brinkman R, Gottardo R, Scheuermann RH. Critical assessment of automated flow cytometry data analysis techniques. *Nat Methods.* 2013; 10:228–238. [PubMed: 23396282]
 35. Cron A, Gouttefangeas C, Frelinger J, Lin L, Singh SK, Britten CM, Welters MJ, van der Burg SH, West M, Chan C. Hierarchical modeling for rare event detection and cell subset alignment across flow cytometry samples. *PLoS Comput Biol.* 2013; 9:e1003130. [PubMed: 23874174]
 36. Britten CM, Janetzki S, Butterfield LH, Ferrari G, Gouttefangeas C, Huber C, Kalos M, Levitsky HI, Maecker HT, Melief CJ, O'Donnell-Tormey J, Odunsi K, Old LJ, Ottenhoff TH, Ottensmeier C, Pawelec G, Roederer M, Roep BO, Romero P, van der Burg SH, Walter S, Hoos A, Davis MM. T cell assays and MIATA: the essential minimum for maximum impact. *Immunity.* 2012; 37:1–2. [PubMed: 22840835]

37. Britten CM, Janetzki S, Ben Porat L, Clay TM, Kalos M, Maecker H, Odunsi K, Pride M, Old L, Hoos A, Romero P. Harmonization guidelines for HLA-peptide multimer assays derived from results of a large scale international proficiency panel of the Cancer Vaccine Consortium. *Cancer Immunol Immunother.* 2009; 58:1701–1713. [PubMed: 19259668]
38. Speiser DE, Pittet MJ, Guillaume P, Lubenow N, Hoffman E, Cerottini JC, Romero P. Ex vivo analysis of human antigen-specific CD8+ T-cell responses: quality assessment of fluorescent HLA-A2 multimer and interferon-gamma ELISPOT assays for patient immune monitoring. *J Immunother.* 2004; 27:298–308. [PubMed: 15235391]
39. Heijnen IA, Barnett D, Arroiz MJ, Barry SM, Bonneville M, Brando B, D’hautcourt JL, Kern F, Totterman TH, Marijt EW, Bossy D, Preijers FW, Rothe G, Gratama JW. Enumeration of antigen-specific CD8+ T lymphocytes by single-platform, HLA tetramer-based flow cytometry: a European multicenter evaluation. *Cytometry B Clin Cytom.* 2004; 62:1–13. [PubMed: 15468327]
40. Mosmann TR, Naim I, Rebhahn J, Datta S, Cavanaugh JS, Weaver JM, Sharma G. SWIFT-scalable clustering for automated identification of rare cell populations in large, high-dimensional flow cytometry datasets, part 2: biological evaluation. *Cytometry A.* 2014; 85:422–433. [PubMed: 24532172]
41. Knabel M, Franz TJ, Schiemann M, Wulf A, Villmow B, Schmidt B, Bernhard H, Wagner H, Busch DH. Reversible MHC multimer staining for functional isolation of T-cell populations and effective adoptive transfer. *Nat Med.* 2002; 8:631–637. [PubMed: 12042816]
42. Whelan JA, Dunbar PR, Price DA, Purbhoo MA, Lechner F, Ogg GS, Griffiths G, Phillips RE, Cerundolo V, Sewell AK. Specificity of CTL interactions with peptide-MHC class I tetrameric complexes is temperature dependent. *J Immunol.* 1999; 163:4342–4348. [PubMed: 10510374]
43. Bergeron M, Nicholson JK, Phaneuf S, Ding T, Soucy N, Badley AD, Hawley Foss NC, Mandy F. Selection of lymphocyte gating protocol has an impact on the level of reliability of T-cell subsets in aging specimens. *Cytometry.* 2002; 50:53–61. [PubMed: 12116346]
44. White S, Laske K, Welters MJP, Bidmon N, van der Burg SJ, Britten CM, Enzor J, Staats J, Weinhold KJ, Gouttefangeas C, Chan C. Managing multi-center flow cytometry data for immune monitoring. *Cancer Informatics.* 2014 accepted for publication.

Panel participants

E. Inderberg, K. Lislerud, AM. Rasmussen, G. Gaudernack, Institute of Cancer Research, Radium Hospital, Oslo, Norway

K. Giannopoulos, Clinical Immunology Department, Medical University of Lublin, Lublin, Poland

S. Attig*, C. Gouttefangeas, Department of Immunology, University of Tübingen, Tübingen, Germany.

M. Schmitt-Händle, E. Kämpgen, Department of Dermatology, University Hospital Erlangen, Erlangen, Germany

A. Konur, Third Medical Department, Johannes-Gutenberg University, Mainz, Germany

A. Letsch, Department for Haematology and Oncology, Charite Hospital, Berlin, Germany

A. Mackensen, M. Aigner, Department for Haematology and Oncology, Erlangen, Germany

R. Maier, Institute of Immunobiology, Cantonal Hospital St. Gallen, Switzerland

*Now TRON gGmbH, University Medical Center, Johannes Gutenberg University Mainz, Mainz, Germany.

L. Low, C. Ottensmeier, Cancer Science Division, Southampton University Hospitals, Southampton, UK

E. Derhovanessian, G. Pawelec, Center for Medical Research, University of Tübingen, Tübingen, Germany

H. Pohla, Laboratory for Tumor Immunology, Ludwig-Maximilians University, Munich, Germany

D. Riemann, B. Seliger, Institute for Medical Immunology, Martin Luther University, Halle, Germany

T. Køllgaard*, P. Thor Straten, Center for Cancer Immune Therapy, Herlev, Denmark.

M. Welters, SH. van der Burg, Department of Clinical Oncology, Leiden University Medical Center, Leiden, The Netherlands.

CM. Britten*, Department of Immunohematology and Blood Transfusion, Leiden University Medical Center, Leiden, The Netherlands.

S. Koch*, R. van Lier, Department for Experimental Immunology, University of Amsterdam, Amsterdam, The Netherlands.

M. Navarette*, AK. Kaskel, H. Veelken⁺, Department of Haematology and Oncology, Freiburg University Medical Center, Freiburg, Germany.

R. Mendrzyk, S. Walter, Immatics biotechnologies GmbH, Tübingen, Germany

*Now Institute for Inflammation Research (IIR), Copenhagen University Hospital, Denmark.

*Now TRON, Johannes-Gutenberg University, Mainz, Germany.

*Now Curevac GmbH, Tübingen, Germany.

*Now Magallanes University Medical School, Punta Arenas, Chile.

⁺Now Department of Haematology, Leiden University Medical Center, Leiden, The Netherlands; MN

Figure 1. Flow chart of the study

Experimental settings and conclusions of the three successive steps are indicated.

Figure 2a.

Figure 2b.

Figure 2. Detection rates for FLU and CMV-specific CD8⁺ T cells following the two gating strategies

- Group analysis for each of the 7 antigen specificities to be detected separately.
- Single-center and group analysis for all 7 antigen specificities together. (for details, see Suppl. Table 2).

Figure 3. Frequencies of Flu- and CMV-specific CD8⁺ T cells - group analysis

Scatter plots display lab-individual frequencies of CD8⁺multimer⁺ lymphocytes obtained for each of the 7 expected antigen/donor combinations with the two gating strategies (positive stainings only). Means (black lines) and % CV are indicated.

Figure 4. Frequencies of antigen-specific cells determined by individual and central analyses
 Scatter plots display the frequencies of CD8⁺multimer⁺ lymphocytes obtained by the individual labs (black circles, I) or by the central analysis (gray triangles, C) of FCS files. Means (black lines) and % CV are indicated.

Five antigen specificities are included in the analysis (numbers of positive stainings: 5, 13, 14, 5 and 11 for D2 CMV, D2 Flu, D3 CMV, D3 Flu and D4 Flu respectively)

Figure 5a.

Figure 5b.

Figure 5. Automated analysis of heterogeneous FCS files

a) Comparison of the frequencies of antigen-specific cells determined by individual and automated analyses.

Scatter plots display the frequencies of CD8⁺multimer⁺ lymphocytes for all antigen specificities obtained by the n=14 individual labs (black circles, I) or by the automated analysis of the same FCS files (gray diamonds, A). Means (black lines) and % CV are indicated.

b) Examples of individual and automated analyses.

(A) automated analysis with back-gated multimer⁺ events shown in black. (I) CD8/HLA-multimer dot-plots provided by the participants. The upper staining was scored as being positive whereas the middle and lower ones were considered to be negative after central inspection of proficiency panel data.

Table 1

Comparison of the three analysis methods

	D1 Flu	D2 CMV	D2 Flu	D3 CMV	D3 Flu	D4 CMV	D4 Flu
Detection rate							
I/C	71.4 (10/14)	38.5 (5/13)	100 (13/13)	100 (14/14)	38.5 (5/13)	100 (14/14)	84.6 (11/13)
A	57.1 (8/14)	53.8 (7/13)	92.3 (12/13)	100 (14/14)	23.1 (3/13)	100 (14/14)	92.3 (12/13)
Mean							
I	1627	3299	676	34	3208	47	1699
C	n.d.	3562	843	36	4951	n.d.	2071
A	3795	7064	1201	73	4170	47	2852
Median							
I	1616	3148	610	32	1286	44	1586
C	n.d.	3490	754	34	6463	n.d.	2120
A	2743	7311	968	40	4801	43	2082
% CV							
I	36	60	39	29	105	30	60
C	n.d.	28	39	36	61	n.d.	34
A	70	38	54	158 (60)*	77	41	58

Results are given for the seven specificities analyzed for n=14 data sets (gating strategy 1), for individual (I) or central (C) manual analyses, or for the automated approach (A).

Mean and median frequencies of antigen-specific cells are given as 1 multimer⁺ cell/ × CD8⁺ lymphocytes.

n.d. = not determined (D1 Flu and D4 Flu were not assessed by central analysis).

* CV after excluding test from lab ID 10