Special Report

Recommendations from the iSBTc-SITC/FDA/NCI Workshop on Immunotherapy Biomarkers

Lisa H. Butterfield, A. Karolina Palucka, Cedrik M. Britten, Madhav V. Dhodapkar, Leif Hakansson, Sylvia Janetzi, Yutaka Kawakami, Thomas-Oliver Kleen, Peter P. Lee, Cristina Maccalli, Giorgio Trinchieri, Sjoerd H. van Der Burg, Theresa L. Whiteside, Jon M. Wigginton, Craig L. Slingluff, Jr., Wenru Song, David F. Stroncek,Hideaki Tahara, Magdalena Thurin, Raffaele Scientific Institute, 32Unit of Immunotherapy of Human Tumors, Tumor Vaccine Group, Division of Oncology, University of Siena, Siena, Italy; 33Medical Oncology and Immunotherapy, University Hospital of Siena, Istituto di Riferimento Oncologico, Aviano, Italy; 34Department of Surgery and Bioengineering, Advanced Clinical Research Center, Institute of Medical Science, The University of Tokyo, Minato-ku, Tokyo, Japan; 35Department of Otolaryngology, Head & Neck Surgery, University of Washington, Seattle, WA 98195. Phone: 206-616-1823; Fax: 206-685-3128; E-mail: disis@uw.edu

Abstract

Purpose: To facilitate development of innovative immunotherapy approaches, especially for treatment concepts exploiting the potential benefits of personalized therapy, there is a need to develop and validate tools to identify patients who can benefit from immunotherapy. Despite substantial effort, we do not yet know which parameters of antitumor immunity to measure and which assays are optimal for those measurements.

Experimental Design: The iSBTc-SITC (International Society for Biological Therapy of Cancer-Society for Immunotherapy of Cancer), FDA (Food and Drug Administration), and NCI (National Cancer Institute) partnered to address these issues for immunotherapy of cancer. Here, we review the major challenges, give examples of approaches and solutions, and present our recommendations.

Results and Conclusions: Although specific immune parameters and assays are not yet validated, we recommend following standardized (accurate, precise, and reproducible) protocols and use of functional assays for the primary immunologic readouts of a trial; consideration of central laboratories for immune monitoring of large, multi-institutional trials; and standardized testing of several phenotypic and functional potential potency assays specific to any cellular product. When reporting results, the full QA/ QC should be conducted and selected examples of truly representative raw data and assay performance characteristics should be included. Finally, to promote broader analysis of multiple aspects of immunity, and gather data on variability, we recommend that in addition to cells and serum, RNA and DNA samples be banked (under standardized conditions) for later testing. We also recommend that sufficient blood be drawn to allow for planned testing of the primary
hypothesis being addressed in the trial, and that additional baseline and posttreatment blood is banked for testing novel hypotheses (or generating new hypotheses) that arise in the field. Clin Cancer Res; 17(10); 1–13. ©2011 AACR.

**Translational Relevance**

Progress in the field of immunotherapy has been slow, but recent clinical successes have given strong support to the potential of this approach as a treatment modality in cancer. To define biomarkers to identify patients who will have clinical benefit, and to ultimately identify appropriate patient groups to enroll, clinical trials testing immune-based interventions should conduct more thorough and standardized immunologic assays to fully study clinical responders and nonresponders, and report data and their analysis methodology in greater detail. An increased focus on immune assessments in these patients will allow us to learn more about the mechanisms of action of the tested interventions and the positive and negative immune responses in treated patients.

Immunotherapy of cancer is becoming increasingly utilized and is becoming a part of cancer management alongside more classical approaches such as chemotherapy or radiotherapy. The U.S. Food and Drug Administration (FDA) has recently granted full approval for the first therapeutic cancer vaccine "Provenge" (1) on the basis of a randomized phase III trial (2), and other recent trials have reached their primary endpoints according to the investigators (3–5). However, challenges exist that need to be resolved to facilitate development of innovative approaches, especially for treatment concepts exploiting the potential benefits of personalized therapy. There is a need for development and validation of tools to identify patients who can benefit from a particular form of immunotherapy. For example, only a fraction of patients are eligible for adoptive tumor-infiltrating lymphocyte (TIL) cell transfer (6), only a fraction of patients can achieve durable regressions in response to cell or antigen vaccination (7), or antibody therapies, and we do not know the mechanisms responsible. Despite substantial efforts from many groups, we do not know which parameters of immune responses, and which assays used to assess these parameters, are optimal for efficacy analysis. Indeed, the tumor-specific cellular immune response promoted by immunization often has not correlated with clinical cancer regression despite the induced cytotoxic T cells detected in in vitro assays (8–11). The major reason is that objective clinical response rates are usually below 10%, preventing meaningful correlations of specific T-cell response rates with clinical responses in small sized, early stage trials. Other possible reasons are:

- A lack of agreement over which immunologic parameter to measure, what time point(s) is relevant, and which assay to use to make that measurement.
- Immunologic assays used may not accurately reflect an in vitro immune response, for example cytotoxicity assays with lengthy in vitro stimulation.
- The analysis of single parameters alone may not provide sufficient insights about complex immune system–tumor interactions.
- Common immunoassays do not take into account changes: in the differentiation of immune cells, in the antigenic profile of tumors and responding T cells, in T-cell homing receptors, or the complex analysis of "private" antigen responses.

**Several Key Issues Must Be Addressed to Move the Field Forward**

- How to ensure the quality of sample processing and storage for the future studies?
- How to standardize the assays and which assays should we standardize?
- How to link immune and clinical outcomes with cellular product characteristics?
- How to utilize the existing knowledge and assays to allow for meta-analyses of laboratory and clinical results?

Following discussions between representatives of the FDA leadership, the National Cancer Institute (NCI), the NIH and Industry, as well as experts in the field of immunotherapy, the iSBTc-SITC/NCI/FDA Taskforce on immunotherapy biomarkers was created. This group led a workshop held on October 28, 2009, in conjunction with the Annual Meeting of the International Society for Biological Therapy of Cancer; iSBTc (now known as the Society for Immunotherapy of Cancer, SITC). This workshop was also a follow-up to the 2001 workshop of the iSBTc-SITC (12) and included participation from 6 partner societies and representation of 20 countries. The results of the discussions of the taskforce and recommendations from the workshop participants and from representatives of several international immunotherapy societies are presented here.

**Background**

The major goal for the field of immunotherapy is to improve the clinical efficacy of immunotherapies. To do so, we require immunologic biomarkers of efficacy. The NCI Translational Research Working Group (TRWG) has incorporated the need for these biomarkers in their developmental pathways, which are frameworks for bench to
bedside development of new therapies and which include a pathway for “Immune Response Modifiers” (13). The FDA pharmacogenomics guidance (14) defines a valid biomarker as “a biomarker that is measured in an analytic test system with well-established performance characteristics and for which there is an established scientific framework or body of evidence that elucidates the physiologic, toxicologic, pharmacologic, or clinical significance of the test results.” This clearly states the need for biomarker assay standardization and also implies the inherent complexity in accomplishing this goal in the field of immunotherapy of cancer.

The “Critical Path” is the FDA’s initiative to identify and prioritize the most pressing medical product development problems and the greatest opportunities for rapid improvement in public health benefits (15). Its primary purpose is to ensure that basic scientific discoveries translate more rapidly into new and better medical treatments by creating new tools to find answers about how the safety and effectiveness of new medical products can be shown in faster time frames with more certainty and at lower costs. The Critical Path has six areas of focus (including biomarker development and product manufacturing) and the Critical Areas for Biomarker Development are described as follows: Biospecimens, Analytical Performance, Standardization and Harmonization, Bioinformatics, Collaboration and Data Sharing, Stakeholder Education and Communication, Regulatory Issues, and Science Policy. The importance of biomarkers is clear. Good biomarkers offer the prospect for earlier diagnosis, and for which there is an established scientific framework and serum for immunologic studies.

Challenges, examples, and recommendations

Processing and storage of blood samples to bank PBMCs and serum for immunologic studies. Challenge. Multiple variables in drawing blood and obtaining tissue can impact the quality of the cells obtained and their proper testing.

Examples. This area has been addressed in depth by the Immunologic Monitoring Consortium (investigators from the University of Washington, Duke University, BD Biosciences, and Coulter), and their cryopreservation and thawing recommendations are outlined in Appendix 1A. In addition, recent studies have tested the importance of time from blood draw to PBMC processing and established that the shorter the time, the superior the viability and functionality of the PBMCs (17, 18). The use of cell preparation tubes (CPT) versus Ficoll for PBMC separation, and shipping of CPT tubes (BD Vacutainer) versus shipping whole blood, has been investigated. There are data to suggest that CPT can perform equivalently to Ficoll (19), but it remains unclear whether shipping spun CPT tubes is superior to shipping heparinized whole blood. To allow for multi-institutional (and multicontinent) trials with minimal blood sample function loss, the AIDS Network has established an Immunology Quality Assessment (IQA) Program to evaluate and enhance the comparability of immunologic laboratories handling blood samples from patients (see Appendix 1B and “Centralization of immunologic monitoring” centralizes laboratories). This approach could be of great benefit to the cancer immunotherapy community.

An additional critical issue is the volume of blood collected (20). Ideally, banking samples for future analyses using newly developed techniques would potentially allow a better understanding of the mechanisms of response or exploration of novel prognostic biomarkers. In a study of 416 blood draws each aimed at taking 250 mL of blood, a median of 200 mL was actually collected in patients with stage III or IV breast cancer. The hematocrit of these patients was not significantly decreased during the time of these blood draws, data which may facilitate Institutional Review Board (IRB) approvals of larger volume blood draws (20). PBMC samples stored for extended periods are being tested for function (ref. 21; Appendix 1C).

Recommendations. We recommend following standardized processing, cryopreservation, storage, and thawing protocols already tested by the Immunologic Monitoring Consortium, or testing the same parameters in your own laboratory and stating the extent of standardization in the associated publications. Consider drawing large (200–250 mL) volumes or carrying out pre- and posttreatment aphereses, to allow for broad assessment of multiple immune parameters including cells and serum and/or plasma.

Characterization of cellular products for therapy. Challenge. A wide variety of cellular products are being tested for therapy of cancer, from minimally manipulated autologous blood products to cultured cell lines, and antigen loaded, matured dendritic cells (DC; ref. 22). These are required to undergo FDA-mandated tests before release and
administration (21 CFR 211.65). Some are relatively straightforward (safety, identity, and purity) and others are more complex [potency (developed in phase I and II, to be utilized for phase III), stability (acceptable conditions for both short- and long-term storage), and consistency (batch to batch comparability)]. Products that do not meet the prespecified release criteria must not be administered. Autologous products can be highly variable between patients and are challenging to characterize and standardize, and such variability, often minimally characterized, can impact immune biomarkers.

Examples. We have included an example of the testing (both exploratory and for product release) conducted for a current autologous DC-based vaccine clinical trial (in Appendix 2). The methods for testing safety are well-standardized. Measures of identity and purity are necessarily specific to the product and are generally flow cytometry based. These might include lineage, activation, and differentiation markers. Potency assays remain exploratory to date and include testing cell surface and intracellular proteins, cytokines and chemokines produced, and activation of target cells [i.e., lymphocyte proliferation stimulated by DC vaccines; killing by natural killer (NK) cells or T cells].

Two examples of candidate potency assays for antigen-presenting cells (APC) are CD54 expression (23) and interleukin (IL) 12p70 production (24). CD54 upregulation on the vaccine cells seemed to correlate with overall survival in two phase 3 clinical trials (25). Spontaneous and induced IL-12p70 secretion assays have been standardized (26) and data are now being collected from multiple ongoing DC-based vaccine trials to determine whether this functional readout correlates with clinical outcome and could become validated as a potency assay. No potency assays have yet been validated and even DC, which have been tested in many different clinical trials, are sufficiently heterogeneous that each modification in antigen loading and maturation may result in a different functional profile; hence, these assays must still be considered exploratory.


Assay standardization and harmonization before testing patient samples. Challenge. Patient blood samples are hard-won resources for understanding the effects of immunotherapy, and are often extremely limited. Therefore, the primary immune response assays carried out using them must be robust and standardized before use in the context of a clinical trial.

Examples. Preclinical and clinical development phases of immunologic biomarkers have been outlined to clarify different stages involved (28). Assay standardization is described in CLIA (Clinical Laboratory Improvements Amendments) Guidelines (see Appendix 3; ref. 29) and the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH which brings together the regulatory authorities of Europe, Japan, and the United States and the pharmaceutical industry to discuss scientific and technical aspects of product registration; ref. 30). The general standardization requirements include quality sample processing and storage (see “Processing and storage of blood samples to bank PBMC and serum for immunologic studies”), a standard operating procedure (SOP) for each sample type which maintains sample integrity, and a quality control (QC) program for each sample type. SOPs need to be established for each component of the testing process, including sample processing, the immunoassay, and analysis. For several common cellular immune biomarker assays [cytokine flow cytometry, MHC tetramers, and enzyme-linked immunosorbent spot (ELISPOT)], a detailed standard procedure is available (31) and suggestions for standardization are included.

Assay standardization and validation also involve technical performance characteristics of the assay (e.g., reproducibility intraday and interday, overall precision and accuracy, specificity, sensitivity, assay range; ref. 16). Criteria for analytic performance of the assay are as follows: the accuracy, precision, reportable range, reference ranges/ intervals (normal values), turnaround time, and failure rate of the assay as it is to be conducted in the trial. Limits of acceptable performance and the QC data should be obtained. Data on positive and negative controls, calibrators, and reference standards should be available. If the assay is to be conducted at more than one site, interlaboratory variability in the measurements needs to be assessed; sources of variation should be minimized to maintain performance at all sites within acceptable limits. Prospective evaluation of assay performance is often executed on a reference population. A bank of PBMCs, serum, or leukopheresis products is helpful as a source of material useful for conducting quality assurance testing (32, 33). Commercially available ELISA or Lumexins kits are generally well standardized for use. For cellular assays, an example is characterizing cytomegalovirus (CMV)-specific T-cell responses in PBMCs, which allows the identification of low (<0.5%), intermediate (0.5%–2%), and high (>2%) level responses which could serve as a surrogate for the range of immunity that might be seen in an immunotherapy study (34).

Establishing clinical utility of the assay involves measuring the relevant response in immunotherapy trials, that is, the magnitude of the association between the assay result and the immunologic and clinical endpoint measured. Scoring procedures need to be established for quantitative, semiquantitative, or qualitative assays (see “Validation of specific assays and/or analytes as biomarkers of clinical response”).
In the last several years, harmonization efforts have been organized to understand and reduce variability in ELISPOT, ICS (intracellular cytokine staining), and MHC tetramer assays between multiple international laboratories. These efforts have involved large proficiency panel programs and include the option for each laboratory to use their own materials, reagents, and protocols, to test centrally distributed, pretested PBMCs and antigens with general logistical guidelines to allow for comparison of results obtained from participating labs. These programs allow for the identification of protocol variables which may influence assay outcome. Panel findings have been summarized in harmonization guidelines. An important aspect of these harmonization efforts is that they do not require strict protocol adherence across all laboratories. Recent findings indicate that serum is not required for \( g \) ELISPOT, according to collaborative studies of different protocols from the European CIMT Immunoguiding Program and the Cancer Immunotherapy Consortium of the Cancer Research Institute (33, 35–38), and other international collaborative studies (32, 38).

The AIDS Consortium have been leaders in interlaboratory proficiency testing and in developing programs which resolve variation between laboratories carrying out immunoassays (see Appendix 1B; ref. 39). Technician understanding of procedures and training has been hurdles to interlaboratory variation and is surmountable. Use of standard methods and reagents across laboratories also increases interlaboratory comparability (40). Indeed, standardized training procedures and reagents allowed for highly concordant IFN-\( \gamma \)-ELISPOT results to be obtained across 7 international laboratories on 3 continents (41). In another interlaboratory comparison study of the IFN-\( \gamma \) ELISPOT, 11 assay “novices” were given identical cells, reagents, and SOPs and the results showed that even without prior experience, standardized procedures and reagents can yield reproducible results across laboratories, countries, and continents (32). The tuberculosis community (including the WHO Initiative for Vaccine Research) has also been involved in ELISPOT assay and related cytokine assay standardization (42, 43).

**Recommendations.** We support the use of standardized assays (following CLIA and ICH guidelines) with full disclosure of methodology specifics for the primary immunologic readouts of a trial. Standard protocols and critical assay parameters for several most commonly used assays (particularly for the ELISPOT) have been published (32–42) and should be strongly considered for clinical trial immune response testing. Participation in external proficiency panels and use of prescreened PBMCs and other additional controls can be very important for comparing results between sites and, potentially, between trials.

**Centralization of immunologic monitoring.** Challenge.

Clinical trial blood and tissue sample processing, cGMP cellular product production, and immunologic monitoring assay standardization are extremely laborious and costly. To address these concerns, several institutions have invested in centralized laboratories. **Examples.** The benefits of these laboratories include high quality and reliability with QA/QC programs, state-of-the-art assay development, standardization, and validation, as well as decreased cost of immune monitoring (due to large discounted purchases, batched assay testing, previously developed assays, and well-trained staff). In addition, assay consultation and result interpretation in conjunction with data analysis by biostatisticians can be available; and banks of samples for normal controls exist for comparisons and normal ranges. One drawback is timing for sample handling. Shipping from a clinical site to a central laboratory necessitates a delay of up to 24 hours in sample processing, which may adversely affect PBMC functional responses as well as the expression of some labile markers on cells and in serum (also ref. 44; e.g., CD62L). Standardized shipping conditions (i.e., materials, logistics, and monitored temperature) can address some of the time-dependent alterations. A different approach is setting up a regional central laboratory network with centralized training and oversight (see AIDS network IQA in Appendix 1B).

The following are 4 examples of multiple centralized laboratories across the United States. The Laboratory of Cell Mediated Immunity (LCMI), SAIC-Frederick, Inc., is a centralized contractor laboratory performing immunologic monitoring for the NIH (A.M.). The laboratory is CLIA certified and conducts many assays including modified ELISPOT assays including peptide, whole protein, and tumor cells. Normal donors with known responses to recall antigens serve as positive controls and new assay variations undergo optimization studies. Cellular Technology Limited (CTL) operates a GLP-compliant and CLIA-certified laboratory for specimen processing, storage, and immune monitoring with an emphasis on ELISPOT. CTL was awarded multiple IDIQ contracts by the NIH for validation of T-cell–based immune monitoring approaches and has served as ELISPOT reference laboratory to the Immune Tolerance Network (ITN), and as a central laboratory to the Cancer Vaccine Consortium (CVC). The ECOG Central Immunology Laboratory which operates at the University of Pittsburgh Cancer Institute (UPCI; T.L.W. and L.H.B.) provides ECOG network investigators with processing, storage, and testing of specimens under CAP-inspected, CLIA-compliant conditions. Samples are received via protocol-specific specimen shipping kits prepared by the laboratory for clinical sites. In E1696, a phase II multiepitope peptide vaccine trial for patients with measurable metastatic melanoma, a significant difference was reported in overall survival of patients stratified by immune response status, with responders living longer than non-responders (median overall survival \( = 21.3 \) vs. 10.8 months, \( P = 0.033; \) ref. 45). However, in a Cox model AJCC (American Joint Committee on Cancer) stage at diagnosis was the most significant predictor of overall survival (\( P = 0.002 \)) and immune response status trended to significance (\( P = 0.073 \)). Examples of percent CV (coefficient of variation) and assay controls from ECOG 1696 ELISPOTs are listed in Appendix 4. Similarly, in an analysis...
of immune responses to a 12-peptide vaccine conducted in the Human Immune Therapy Center at the University of Virginia (46), there was a significant correlation between ELISPOT reactivity and disease-free survival in a univariate analysis of 48 patients with resected stages II to IV melanoma who received a multipeptide vaccine in the adjuvant setting. This analysis did not stratify for stage; so other factors could have contributed to this observed correlation. Finally, in a trial testing a polyepitope DNA prime/vaccinia boost vaccine in stage III/IV melanoma patients, in which immune monitoring was centrally conducted, a significant correlation was seen between MHC tetramer responses and median survival, and a trend toward correlation between IFN-γ ELISPOT response and median survival was seen (47).

**Recommendations.** We recommend consideration of central laboratories for immune monitoring, due to their experience in standardized assay conduct and existing infrastructure. In particular, they can be a critical part of larger scale, multicenter clinical trials for minimizing variation. It is recommended that central laboratories establish (1) the historical data on any specific standardized assay for the selected parameter as the reference and (2) provide a service to conduct a comparability test to validate the data that are generated in other study sites with the reference data as the control if applicable.

**Standardized (or standardizable) assays which should be used for clinical trial antitumor immune response determination.** Challenges. There are many assays potentially capable of measuring aspects of immune function and limited blood and tissue samples require choosing a limited number of possible assays. The field of immunologic monitoring is also constantly evolving. The small numbers of complete clinical responders, small-scale trials, and variability in assays chosen and assay conducted make identification of the crucial assay parameters to measure difficult to identify.

Examples. The choice of immunoassay will, in large part, depend on the proposed mechanisms of action of the immune intervention. A vaccine designed to generate a specific antibody response, for example, would focus on the assessment of humoral immunity rather than a cellular response. To test specific immune effectors, there are many choices for the methods such as ELISPOT, MHC-peptide multimer staining, intracellular cytokine staining, as well as soluble cytokines, NK cells, and Th (T helper) phenotype. Not only the magnitude of the response and frequency of effector cells but also the antigenic breadth and degree of multifunctionality have all been shown to be critical in model systems and specific clinical trials (48) and broad polyepitopic immune responses have been associated with complete clinical response in a small group of therapeutically vaccinated patients with vulvar intraepithelial neoplasia (49). In contrast, vaccine failure was associated with higher frequencies of disease-specific CD4+ Foxp3+–positive T cells and a low production of IFN-γ by disease-specific T cells on first vaccination (50). The need for several assays to explain the success and failure of this therapeutic vaccine illustrates the need for sufficient amounts of blood, taken at different time points.

The ELISPOT, perhaps the most thoroughly standardized assay to date (see “Standardized (or standardizable) assays which should be used for clinical trial antitumor immune response determination”), identifies the number of functional antigen-specific cells, multiple samples may be tested simultaneously, and it can be used for testing more than 1 analyte or function (51). It has been shown that 2 measures of cytotoxicity, the Granzyme B ELISPOT Assay (52) and standard 51Cr release correlated better with each other than MHC tetramer or IFN-γ ELISPOT assays, in clinical setting (53). In addition, with automated analysis methods, great reproducibility and accuracy for detecting specific T cells can be achieved. An alternative functional assay is intracellular staining for cytokines or other effector molecules using flow cytometry. Multicolor analysis thus provided can complement ELISPOT, as it can provide additional information regarding multiple cytokines that are produced by specific surface-stained cell subsets. Correlation of tetramer assay results with *in vitro* cytotoxicity in clinical trial material has been previously observed (54). However, in a multicenter cooperative group trial, MHC tetramer frequencies and differentiation stage did not correlate with clinical outcome but IFN-γ ELISPOT response did (Schaefer, submitted manuscript).

Another important issue is the target used in the assay. In two phase II trials of vaccination with a cocktail of altered HLA (human leukocyte antigen)-A2 tumor peptides in early melanoma (N = 40) and prostate carcinoma (N = 20) patients, the *ex vivo* IFN-γ ELISPOT and HLA/peptide multimer staining showed a rapid induction of peptide-specific CD8+ T cells of the majority of vaccinated patients. However, clinical efficacy only correlated with significant antitumor cell activity *in vitro*. These data clearly stress the need for including tumor cell recognition when monitoring patients treated with tumor vaccines, especially if based on modified peptides, to gain better information about tumor reactivity (55, 56).

Recent flow cytometric measures of cell-mediated immunity can evaluate both the target cell death and effector cell function simultaneously, allowing for more efficient acquisition of both tumor target cell cytolysis and CTL activation. A flow cytometry–based cytotoxicity assay has been developed to simultaneously measure NK cell cytotoxicity and NK cell phenotype (37). Another cytotoxicity assay that has been optimized to utilize low numbers of antigen-specific T cells has been described (58) in which peptide/MHC multimer-positive CD8+ T cells were purified, cloned, expanded, and tested for CD107a cell surface expression and their cytotoxicity evaluated on the basis of the frequency of dead cells in CMFMR-labeled target cell populations. This assay proved to be more sensitive than the 51Cr-release assay. Similarly, combining the measure of CD107a by CD8+ effector cells with the apoptosis marker Annexin V binding to target cells has been used (34, 59, 60).
There is increasing enthusiasm for polyfunctional flow cytometric methods, which are already standardized in the HIV community. As discussed at the 2009 workshop, CMV, HIV, and cancer can all induce endogenous T-cell responses of varying magnitudes; but only CMV responses tend to be protective. The T-cell response signatures for CMV, HIV, and cancer may be very different: CMV elicits a relatively high proportion of IFN-γ⁺ IL-2⁺ cells with heterogeneous phenotypes including many effectors. HIV elicits few CD8⁺ IL-2⁺ T cells and intermediate phenotypes. Cancer patients may show low magnitude, IL-2⁺ but not IFN-γ⁺ T cells and central memory phenotype. The mechanisms leading to these different signatures need to be further elucidated (61).

It is essential to develop well-established QC standards that can be made available among different laboratories for assay standardization to ensure assay consistency between sites. Minimally, a description of internal (e.g., positive antigen or peptide, negative controls, mitogens) and external controls need to be provided for each assay (see “Validation of specific assays and/or analytes as biomarkers of clinical response”). The taskforce members have been in discussions with the BRB [Biological Resources (Branch)] of the DTP (Developmental Therapeutics Program), NCI, about creation of a repository for assay standards. This may be feasible once sources for the agreed-upon standards are identified. Alternatively, commercial sources are currently available (SeraCare, Cellular Technology Limited, etc.).

Another important aspect of standardization is testing the extent to which an assay can be run from batched cryopreserved samples, or whether only fresh samples yield reliable results. Finally, how to balance standardized assays for immune responses (which allow the field to move forward by having some ability to compare trials) with research questions (which drive innovation and may identify novel biomarkers with greater specificity for clinical outcomes)? Larger volumes of blood drawn [without negative impact on the patient (see “Processing and storage of blood samples to bank PBMC and serum for immunologic studies”)] may allow for this balance with IRB approval. Alternatively, conducting pre- and posttreatment aphereses provides cells for monitoring as well as a resource for research questions and assay development (62, 63).

Recommendations. We support choosing a standardized, functional assay as the primary readout of the immune response, which addresses the specific hypothesis being tested and proposed mechanism of action of the intervention. There are now strong data that testing multiple functional parameters (multiple cytokines, recognition of not only peptides but also tumor cells) can yield important information.

How assay data should be analyzed for “responder” and “nonresponder” identification? Challenge. Although many assays yield data on changes in antitumor immunity, distinguishing assay variation from normal human variation and treatment-induced responses is not trivial.

Examples. All binary response definitions for individual patients are effectively “seat of pants,” statistical properties are unknown and in practice, unknowable—that is, false-positive and false-negative rates are unknown. Use of a single standardized definition could allow the results of different studies to be compared (but will not solve this problem). At the UPCI, the Biostatistics Facility (W. Gooding) developed a definition for individual IFN-γ ELISPOT response that has achieved a measure of success insofar as it has been shown to correlate with clinical response (45). The definition assumes that there are 3 wells each for identical test samples (tst) and control samples (ctrl). The variable y is set equal to mean (tst) − mean (ctrl). If the number of cells is different for tst and ctrl, the tst counts are scaled by the ratio of the number of tst cells to the number of ctrl cells, and y is set to 0 if the number of responding tst cells in any well is less than the number of responding control cells in any well. An individual’s response to treatment is then determined by the following criteria: y (posttreatment) divided by y (pretreatment) must be greater than 2; and y (posttreatment) must be greater than 10. This definition is essentially based on a factor of 2-fold increase in post- over pretreatment background-corrected ELISPOT counts, albeit with considerable protection against false-positive results due to small counts. This definition has also been employed in other multicenter trials (64), and an adjustment of it for use with stimulated ELISPOT assays has been shown to have low false-positive results, and to correlate with clinical outcome (46). It must be kept in mind that both true immune response and clinical response are continuous, not binary, variables. Different definitions of binary response for individual patients are arbitrary and will correlate differently with clinical outcome. Thus, the relationship of degree of immune response with the degree of clinical outcome might be used to refine binary response criteria. Moodie and colleagues (65) have recently compared response definitions for ELISPOT assays. Although they define response in terms of a comparison of test and control samples, the methods that they describe could generally be adapted to a comparison of pre- and posttreatment samples.

Recommendations. Obtain multiple pretherapy samples (at different times) for analysis; these can be used to assess pretherapy variability of the biomarker level. Tighten response criteria as follows: require positive responses at 2 consecutive posttherapy time points; this is useful for limiting posttherapy variability. Consider using clinical response to refine the definition of immune response. When immune response is the primary outcome of interest in a trial, use nonparametric techniques (such as the Wilcoxon signed-ranks test) to assess response of the entire sample of patients as a group (49). This avoids the problems associated with attempting to define individual responses. To date, no one knows how big an absolute or relative increase in the frequency of antigen-specific T cells between 2 time points should be considered a biologically relevant response (ref. 66 for empirical rules).

Reporting immunologic monitoring data in publications. Challenge. Evaluation of immunologic monitoring
studies is not possible without disclosure of the details of the methods that are known to influence test outcome variability, the assay conduct, and its interpretation.

Examples. Reporting of complex data sets is a challenge, as different styles exist that limit comparability. The concept of "minimal information" projects and structured reporting of data sets was pioneered by Brazma and colleagues (67). The concept has been adapted for T-cell assays (68) and multiple minimal information (MI) projects with overlap now exist. The REMARK criteria have become the standard for publication of prognostic tumor marker studies (69). The REMARK recommendations clearly outline what should be reported to interpret patient selection, sample storage, assay performance, and critical statistical analysis. These recommendations are accepted by most major journals including those published by ASCO and AACR, yet few immune-based studies adhere to them (69).

In the recent FDA draft guidance for therapeutic cancer vaccines (September 2009), recommendations include consideration of conducting at least 2 assays and that all assay parameters and controls should be clearly described (70–78). It is recommended that investigators make themselves familiar with these different projects to choose the appropriate guidelines for the assays chosen.

Recommendations. We suggest that the following study aspects should always be included when reporting results, independent of other applicable guidelines: The QA/QC conducted, reference populations included, all reagents and controls tested, at least some selected examples of truly representative raw data and the assay performance characteristics. These parameters will allow appropriate reviewer and reader evaluation of the quality and potential impact of the data.

Validation of specific assays and/or analytes as biomarkers of clinical response. Challenge. When data from basic and translational research settings (and exploratory testing in phase I trials) suggest that a specific immunoassay or other biomarker correlates with clinical outcome, standardization and validation are needed to substantiate those data and to allow possible comparisons between treatments and trials.

Examples. GLP guidelines from the FDA for general laboratory conduct for assay performance are available (79). For the ELISPOT assay, validation of the assay has been addressed (80). A recent report (81) has described many aspects of standardization of MHC tetramer, IFN-γ ELISPOT, and IFN-γ real-time PCR and the authors also attempted to validate the assays for determining the absolute frequency of antigen-specific T cells. They concluded that this could not be accomplished without a "gold standard" measure for such cells.

Quantification of polyfunctional cytokine-producing T cells (often IFN-γ, IL-2, and TNF) by multicolor flow cytometry is being used in infectious disease models and HIV patients for correlation with clinical outcome (60, 82). This is an immunologic readout which may also serve as a biomarker for clinical response and there is already standardized shared software for analysis of the flow cytometric data (SPICE, M. Roederer, VRC, NIAID, NIH). Peptide pools are commonly paired as antigen sources for this assay and they are evaluated by another shared software package (Deconvolute This; ref. 83).

Conventional response criteria may not adequately assess the activity of immunotherapeutic agents. Therefore, systemic criteria, immune-related response criteria (irRC) have been defined to capture relevant clinical response patterns observed in melanoma patients undergoing immunotherapy. Use of irRC may allow for improved comprehensive evaluation of immunotherapeutic agents in clinical trials and potentially offer guidance in clinical care, as well as being a more appropriate comparator for correlation with in vitro measures of antitumor immunity (84). Consideration of the time required for evolution of immune responses may require collection of patient blood samples over a longer time period.

Recommendations. The immunotherapy field continues to produce novel data from immune assessments in patients which correlate to clinical outcome in different diseases and treatment settings. These candidate biomarkers should first be standardized (85) and then validated by other investigators. The evolution of antitumor immunity may necessitate longer term immunologic monitoring.

Novel assays in development for immunologic testing of patients. Challenge. To move toward more sensitive, high-throughput evaluations, there must first be quality sample acquisition for analysis and hypothesis testing. Also, it is not only immunotherapy trials which must be evaluated. Most biological agents used singly or in combination with conventional drugs for cancer therapy engage the immune system. Other biological agents that specifically target growth factor receptors, blood vessels/endothelial cells, tumor cells, or tumor-associated antigens often involve immunologic mechanisms.

Examples. Directly assaysing the tumor environment, performing expression arrays (from the tumor), testing for determinant/epitope spreading, and testing genetic aspects of the host [single nucleotide polymorphism (SNP), genome-wide association studies (GWAS), HLA] are not yet commonly conducted. One example of a large-scale initiative to analyze patient tumors for individual gene expression patterns is "M2Gen," a research collaboration between H. Lee Moffitt Cancer Center and Merck & Co. Researchers are collecting tumor tissues from patients to identify the biological markers unique to each tumor (86).

Immune profiling can include the following: high-throughput molecular profiling platforms to study the human immune system, polychromatic flow cytometry, RNA profiling (mRNA, miRNA, RNAseq), SNP arrays (soon genome sequence), multiplex serum chemokines, cytokines profiles, and protein and peptide arrays (87) for serologic responses, mapping antigenic repertoire, and semiquantitative immunohistochemistry of the tumor (88). In addition to the positive effects of antitumor effector activation, critical aspects of tumor immunosuppression should be investigated as the frequency and function...
of MDSC (myeloid-derived suppressor cells) and Treg (regulatory T cells), functional defects in TIL and in circulating immune cells, cytokine imbalance (Th2 vs. Th1), failure to generate central memory T cells, persistent activation of T cells, spontaneous apoptosis of T cells, T-cell senescence, and presence of soluble factors in serum that induce death in immune cells. All of these immune deficits have been reported but are not yet part of the regular immune monitoring repertoire. This is crucial for understanding why some approaches are not successful and for personalized selection of available anticancer therapies in the future.

An important aspect of these broad assessments of immunity, particularly with newer, high-throughput approaches, is data management (89). Currently data are often stored in multiple clinical and laboratory databases requiring manual data entry and coordination. Informatics must address the following: the ability to integrate data from multiple technology platforms, the ability to integrate clinical and biomarker data from multiple projects, and include an emphasis on data dissemination/high availability (to allow for downstream analyses by biostatistics/bioinformatics teams; for access/query by investigators, and ultimately to promote insight, sharing data with study participants, collaborators, consortia members, scientific community, and streamline data export to public repositories). A goal inherent to this is defining a universal data element set to accompany all high-quality biospecimens.

**Recommendations.** We recommend that both RNA and DNA samples as well as sera and plasma be banked under standardized conditions for later testing in multiplex, molecular assays (from blood and the tumor and to study the microenvironment). Improved collection of tumor and TIL are crucial for understanding the impact of different therapeutic approaches. We also reiterate that sufficient blood be drawn to allow for the planned testing of the primary hypothesis being investigated in the trial, such that additional baseline and posttreatment blood is banked for testing novel hypotheses (or generating new hypotheses) that arise in the field during the time required for trial design, approvals, enrollment, and conclusion.

**Discussion**

Immunotherapy clinical trials can only benefit from careful study of the effects on patient immune responses and the state of immune function (and dysfunction). Because of the large variation between patients, elimination of as much variation as possible in procedures used for handling blood and tumor specimens, and in procedures for assays, is essential. Equally essential is the thorough reporting of the level of standardization and the specific methods used for specimens, assays, and analysis. We are also recommending an increased level of banking of diverse biological specimens for unspecified future research. We recognize that implementation of this goal will require discussion and cooperation between human subjects’ protection committees and patient advocate representatives along with researchers and clinicians.

As a service to the community, in addition to the present report, we propose the following steps:

a. Specific SOP recommendations (many of which have been standardized and some of which are published) with links to be posted on the iSBTc-SITC web site for easy access (and also to other society web sites).

b. Promoting greater focus on standardization and validation assay guidelines [CLSI (Clinical and Laboratory Standards Institute), CLIA, ICH].

c. Improved data reporting immune biomarker studies should adhere to published guidelines, especially those that are advocated by the journals to which the papers are submitted such as REMARK. Many efforts are ongoing to refine these recommendations MIFlowCyt, MIACA, MIATA, MIBBI, that will help to select the “MI” projects most applicable to a specific scientific question or setting.

d. Further discussion at the follow-up workshop: ‘Symposium on Immuno-Oncology Biomarkers, 2010 and Beyond: Perspectives from the iSBTc-SITC Biomarker Task Force,’ Masur Auditorium on the NIH campus, Building 10, Clinical Center, Bethesda, MD, September 30, 2010 (90).

e. Strengthening of interimmunotherapy society communication and collaborations (Disis, Fox, manuscript in preparation).

Although not directly in the hands of individual investigators and smaller teams, we also recommend greater funding levels to support the acquisition of blood and tumor samples for embedded correlative studies as well as unspecified banking for future analysis. It is only with resources of tumor, serum/plasma, PBMCs, DNA, and RNA that we will be able to learn as much as possible about the state of immunity in cancer patients, the positive effects of our interventions, and the inhibitory effects of tumor progression that we have yet to overcome. These freshly tested and banked samples, collected and assayed under standardized conditions, will also be crucial in allowing us to better understand patient-to-patient variability and take steps toward more effective and personalized approaches.

**Appendix 1A**

**Sample Collection, Cryopreservation, and Thawing (Lymphocytes/PBMCs)**

**Factors that Did Not Matter.**

- Shipping on dry ice, 24, 48, 72 hours versus liquid nitrogen ($P > 0.05$).
- Large (100 mL) versus small volume (25 mL) dilution for thaw ($P > 0.05$).
- Time (5–10 minutes) and speed of spin at wash, 1,200 to 1,500 rpm ($P > 0.05$).
- Number of cells per vial (1, 2, 3 × 10^6; $P > 0.05$).
Factors that Did Matter. The thawing method. Additives such as human serum albumin, dextran, and FBS were superior to human AB serum; washing thawed cells in medium prewarmed to 25°C to 37°C was superior to chilled (4°C) medium (91).

Open Access Protocols are available from the Immunologic Monitoring Consortium (91, 92).

Appendix 1B

Examples from the AIDS Network

IQA Program. The IQA is a resource designed to help immunologists evaluate and enhance the integrity and comparability of immunologic laboratory determinations conducted on patients enrolled in multisite HIV/AIDS investigations.

- About 83 participating laboratories.
- 6 test shipments per year.
- The viability of PBMCs and the viable yield before freezing and after thawing is tested.
- IQA reviews statistical report and identifies sites that are having difficulty performing assays.
- Poor performers are contacted to discuss specific problems.
- Laboratories fax histogram results to the IQA for central review.

Laboratories who serve both the Aids Clinical Trials Group (ACTG) and IMPAACT trials must conform to the stringent criteria of the ACTG (80% viability and 80% viable cell recovery of PBMCs; refs. 93, 94).

Appendix 1C

Large volume blood samples from cancer patients (median volume collected: 200 mL; median yield: 0.8 × 10^6 PBMCs/mL; ref. 21).

Validation in stored samples: ongoing performance
Analysis of 80 samples
Sample age range: 30–2,500 days, average = 600
Median recovery: 70%
Mean recovery: 70%
Range: 22%–130%

Appendix 2

Cellular Product Minimal Safety, Purity, and Identity Tests

The following is an example of the specific release tests which are required by the FDA for early phase trials involving autologous, in vitro manipulated cellular products (in this example, DC). This example also shows the identity/purity testing chosen for this type of product, and the candidate potency test being conducted.

Viability. The cells are counted by microscopic observation on a hemacytometer and a differential count (DCs vs. lymphocytes) is obtained using trypan blue dye. Minimum 70% viability.

Purity. The DCs must express MHC class II and CD86 by flow cytometry in a minimum of 70% of the cells. Additional phenotyping (MHC class I, CD80, CD83, CCR7, others) is conducted to fully characterize the DC and is for research proposes.

Sterility. DCs are tested by bacterial (aerobic and anaerobic) and fungal cultures at the Clinical Microbiology Laboratory. Final results of the microbial cultures are available in 14 days. Prior to release of the DC for vaccine use, a standard Gram stain is conducted and must be negative for the presence of microorganisms.

Mycoplasma testing of cell suspensions (not supernatants) is conducted using a rapid detection system, based on nucleic acid hybridization or by PCR. The cell preparation must be negative for mycoplasma.

Endotoxin testing is conducted on the cell culture at the time of harvest and prior to release of the final product. The acceptable endotoxin level is less than 5 EU/kg of body weight per dose.

Potency. To define a measure of potency for the DC, we determine their ability to produce IL-12p70 and IL-10 by Luminex assay (26). This test is conducted batched, with and without activation by CD40L and/or LPS, and is available several weeks after vaccine injection. Data will be correlated with measures of DC phenotype and clinical outcome.

In addition, a 0.5-mL sample of the final DC preparation from each vaccination time is cryopreserved for possible ancillary testing in the future. These samples are stored a minimum of 1 year after vaccine administration.

Appendix 3

Clinical Laboratory Improvements Amendments (29)

Test accuracy (close agreement to the true value).
Precision (agreement of independent results: same day, different day).
Reproducibility (intra-assay and inter assay)
Reportable range (limits of detection)
Normal ranges (pools of healthy donors, accumulated patient samples: test at least 20, include a banked healthy donor control in patient assays).
Personnel competency testing (minimally, annually)
Equipment validation, monitoring
Reagent tracking

Appendix 4

Preliminary examples from the ECOG Central Immunology Laboratory ELISPOT assays of large-scale, multi-institution clinical trials [n = 20 from each trial, from >70 (E1696) and >200 (E4697) patient ELISPOT assays]
E4697 (2008–2009)  Spontaneous  PMA/1 (±)/OKT3
Healthy control average 4.9 (54% CV) 304 (19.2% CV intra-assay) (48% CV interassay) Patient average 0.7 (35% CV) 81 (38.7% CV)

Healthy control average 5.4 (56% CV) 284 (15.5% CV intra-assay) (51% CV interassay) Patient average 19 (40% CV) 171 (18.8% CV)

Disclosure of Potential Conflicts of Interest
S. Janetzki is founder and president of ZellNet Consulting, Inc., a for-profit company offering ELISPOT plate evaluation and consulting services, and is the Coordinator of the Immune Assay Working Group of the Cancer Immunotherapy Consortium, Cancer Research Institute, a non-profit organization. The other authors disclosed no potential conflicts of interest.

Acknowledgments
The authors thank Dr. Raj Puris (Director, Division of Cellular and Gene Therapies, Office of Cellular, Tissue and Gene Therapies, FDA/Center for Biologics Evaluation and Research) and the FDA liaison to iSBTc-SITC for providing critical critique of the manuscript. They also thank the staff of iSBTc-SITC for production and technical assistance in the manuscript preparation: Tara Withington, CAE, Executive Director; Angela Kilbert, Director of Administration; Chloe Surinak, Project Manager; Roseann Marotz, Meetings Manager; and Jimmy Balwit, Scientific Communications.

Grant Support
The study was supported by NIH grants P01 CA121973 and RO1 CA14924 to L.H. Butterfield; NIH grants RO1 CA119123 and R21 CA13864 to B.A. Fox; and research grant (“Standardization of Immune Monitoring”) to C.M. Britten and S.H. van der Burg from the Wallace Coulter Foundation, Miami, FL. This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. The iSBTc-SITC/FDA/NCI Workshop on Immunotherapy Biomarkers was supported by educational grants obtained by iSBTc-SITC from Bristol-Myers Squibb Company, the Center for Cancer Research, National Cancer Institute, Merck Oncology, Biogen Idec, the Center for Human Immunology, Autoimmunity, and Inflammation, and with additional support from Pfizer and Immunex Corporation.

Received August 19, 2010; revised November 12, 2010; accepted December 10, 2010; published OnlineFirst May 10, 2011.

References

www.aacnjournals.org

Cancer Res; 17(10) May 15, 2011 OF11

Published OnlineFirst May 10, 2011; DOI: 10.1158/1078-0432.CCR-10-2234

Immunotherapy Biomarkers Recommendations

Center for Biologics Evaluation and Research) and the FDA liaison to iSBTc-SITC for providing critical critique of the manuscript. They also thank the staff of iSBTc-SITC for production and technical assistance in the manuscript preparation: Tara Withington, CAE, Executive Director; Angela Kilbert, Director of Administration; Chloe Surinak, Project Manager; Roseann Marotz, Meetings Manager; and Jimmy Balwit, Scientific Communications.

Grant Support
The study was supported by NIH grants P01 CA121973 and RO1 CA14924 to L.H. Butterfield; NIH grants RO1 CA119123 and R21 CA13864 to B.A. Fox; and research grant (“Standardization of Immune Monitoring”) to C.M. Britten and S.H. van der Burg from the Wallace Coulter Foundation, Miami, FL. This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. The iSBTc-SITC/FDA/NCI Workshop on Immunotherapy Biomarkers was supported by educational grants obtained by iSBTc-SITC from Bristol-Myers Squibb Company, the Center for Cancer Research, National Cancer Institute, Merck Oncology, Biogen Idec, the Center for Human Immunology, Autoimmunity, and Inflammation, and with additional support from Pfizer and Immunex Corporation.

Received August 19, 2010; revised November 12, 2010; accepted December 10, 2010; published OnlineFirst May 10, 2011.
design and manufacture in order to develop commercially and clinically viable products. Vaccine 2007;25 Suppl 2:847–60.


42. Kirkwood JM, Lee S, Moschos SJ, Albertini MR, Malyguine A, Strobl SL, Shafer-Weaver KA, Ulderich T, Troke A, Baseler M, Malyguine A. Application of the granzyme B ELISPOT assay for simultaneous immunophenotypes effector cells: comparisons to a cytometric panLeukogating (PLG): a NIAID-DAIDS Immunology Qu-
Clinical Cancer Research

Recommendations from the iSBTc-SITC/FDA/NCI Workshop on Immunotherapy Biomarkers


Clin Cancer Res Published OnlineFirst May 10, 2011.

Updated version Access the most recent version of this article at: doi:10.1158/1078-0432.CCR-10-2234

E-mail alerts Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.