

REVIEW

Blood Cytokines as Biomarkers of *In Vivo* Toxicity in Preclinical Safety Assessment: Considerations for Their Use

Jacqueline M. Tarrant¹

Safety Assessment, Genentech, Inc., South San Francisco, California 94080

¹To whom correspondence should be addressed at Safety Assessment, Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080. Fax: (650) 225-8936. E-mail: tarrant.jacqueline@gene.com.

Received February 2, 2010; accepted April 28, 2010

In the drive to develop drugs with well-characterized and clinically monitorable safety profiles, there is incentive to expand the repertoire of safety biomarkers for toxicities without routine markers or premonitory detection. Biomarkers in blood are pursued because of specimen accessibility, opportunity for serial monitoring, quantitative measurement, and the availability of assay platforms. Cytokines, chemokines, and growth factors (here referred to collectively as cytokines) show robust modulation in proximal events of inflammation, immune response, and repair. These are key general processes in many toxicities; therefore, cytokines are commonly identified during biomarker discovery studies. In addition, multiplexed cytokine immunoassays are easily applied to biomarker discovery and routine toxicity studies to measure blood cytokines. However, cytokines pose several challenges as safety biomarkers because of a short serum half-life; low to undetectable baseline levels; lack of tissue-specific or toxicity-specific expression; complexities related to cytokine expression with multiorgan involvement; and species, strain, and interindividual differences. Additional challenges to their application are caused by analytical, methodological, and study design-related variables. A final consideration is the strength of the relationship between changes in cytokine levels and the development of phenotypic or functional manifestations of toxicity. These factors should inform the integrated judgment-based qualification of novel biomarkers in preclinical, and potentially clinical, risk assessment. The dearth of robust, predictive cytokine biomarkers for specific toxicities is an indication of the significant complexity of these challenges. This review will consider the current state of the science and recommendations for appropriate application of cytokines in preclinical safety assessment.

Key Words: cytokines; toxicity; safety; risk assessment; biomarker; qualification; validation; multiplex immunoassay.

One of the most facile tests of toxicity *in vivo* are changes in blood biomarkers. The advantages of blood biomarkers include quantitative measurement, accessibility, serial monitoring, fast analytical turn-around time, and the availability of analytical

platforms. Blood-based markers have the potential for the translation of preclinical risk assessment to the human patient population as they are generally readily adapted to clinical trials. Presently, there is a shortage of adequately predictive blood biomarkers that correlate with phenotypic manifestations of several important drug toxicities. For example, there are no validated blood biomarkers of vasculitis, lung toxicity (e.g., interstitial pneumonitis), idiosyncratic liver injury, and testicular toxicity (e.g., sertoli cell toxicity and germ cell degeneration).

There is encouragement by regulatory agencies to use sensitive and predictive tests to identify early and clinically monitorable toxicity for successful development and registration of innovative and safe drugs (Woodcock and Woosley, 2008). Identification and understanding of potential toxicities during drug discovery and early development permit accelerated attrition of drug candidates and the opportunity to improve chemical selection through structure-toxicity modeling when such activities are fully supported. The Critical Path initiative by the U.S. Food and Drug Administration (FDA, 2006) has encouraged this approach in their 2006 report, "Innovation or Stagnation: Critical Path Opportunities Report and List." This report acknowledged the gaps extant in toxicity detection using the current toolbox of routine safety biomarkers (e.g., standard clinical pathology parameters) and fosters the advancement of development science in many areas, including biomarker discovery and use. A biomarker qualification process has evolved from this initiative. The process begins with a voluntary submission to regulatory authorities of a novel biomarker and the context of its application by consortia, industry, government, academic, or clinical researchers, followed by a data package (Goodsaid *et al.*, 2008). The test case for this process has been the submission of scientific data on several novel renal toxicity biomarkers by the Critical Path Institute Predictive Safety Testing Consortium to the European Medicines Agency and the FDA, which led to the acceptance of these markers for use in preclinical drug development (EMEA, 2008).

Interest in cytokines, chemokines, and growth factors (hereafter collectively referred to as cytokines) as safety biomarkers has been fueled by mechanistic and exploratory biomarker studies that insinuate inflammatory and repair processes in toxicity. Discovery techniques, such as RNA expression analysis, proteomics, and preconfigured multiplex cytokine assays, may disclose increased or decreased values for various cytokines in animals with toxicity. When there is no premonitory biomarker for a specific toxicity in routine panels, these factors are often proposed as biomarkers for the toxicity. Although there is great desire and technological capability around the potential use of cytokines as safety biomarkers, based on our current scientific knowledge and experience, what practical potential do these factors have in becoming qualified toxicity biomarkers for preclinical risk assessment and in a patient population?

This review will examine the conceptual utility of cytokines as *in vivo* biomarkers of toxicity in preclinical drug development. In particular, consideration will be given to (1) the underlying natural biology of cytokines, (2) analytical and methodological factors in biomarker qualification, (3) evaluation in the established preclinical *in vivo* study designs, and (4) relationship between changes in cytokine levels and association with morphological or functional manifestations of toxicity. These considerations will be described in the context of using systemic (blood) levels of cytokines as exploratory biomarkers for the toxicity of protein therapeutics and small molecules.

CHARACTERISTICS OF A TOXICITY BIOMARKER

A biomarker is defined as any measurable biological characteristic encompassing the detection of physiologic, pharmacologic, and pathologic processes (Biomarkers Definitions Working Group, 2001). A toxicity (or safety) biomarker will either directly reflect or predict susceptibility to a structural and/or functional consequence of exposure to a chemical or biologic therapeutic. When the toxicity biomarker is involved in the mechanism of action of the pharmacological agent, the marker is also a pharmacodynamic (PD) end point. In this instance, there will be a continuum from pharmacology to toxicology and the thresholds will be set in accordance to routine toxicity end points such as histopathology and clinical pathology (clinical chemistry, hematology, urinalysis, and coagulation) and/or demonstration of a level (i.e., a decision limit) beyond which the response is nonreversible and deleterious. Toxicity biomarkers that are not directly related to the desired pharmacologic activity of the drug constitute off-target activity or a secondary (indirect) pathologic process. In this circumstance, the biomarker is purely a signal of toxicity. When the indirect toxicity depends on complex biological interactions between organ systems *in vivo*, the marker may behave differently across species and be contingent on other study design variables.

Desirable characteristics of a toxicity biomarker include specificity and early detection of toxicity (i.e., sensitivity) with a magnitude of change sufficient to distinguish from biological

variability. The half-life of a serum biomarker should allow a practical window for detection, yet be responsive to the changing state of the injured tissue with adequate stability *in vitro*. Premonitory biomarkers, those that precede the manifestation of a histological lesion and predict severity, are clearly valuable for translation to a clinical population and could expedite risk assessment and dosing decisions, i.e., to stop or decrease the dose.

Safety biomarkers measured in blood and other body fluids provide the opportunity for serial monitoring and could potentially reduce the numbers of animals used if the detection of toxicity relied solely on microscopic examination of tissue collected at necropsy. Quantifiable biomarkers allow inference of severity and reversibility of toxicity. Markers that are continuous variables and correlate with microscopic severity and/or loss of function can be used in modeling of dose-response-toxicity relationships to better define safety margins and risk-benefit characterization.

Novel preclinical safety biomarkers are qualified by correlation to an established toxicity end point, e.g., tissue histopathology, clinical pathology parameters, and/or functional tests in animal toxicity studies. Time course and dose-response studies across species with robust positive and negative controls allow evaluation of the performance of the marker in preclinical development. In general, the applicability of a novel toxicity biomarker to clinical trials will center on a comparable biological response between nonclinical species and humans, invasiveness of the procedure, and availability of the technology and expertise.

Thus, a minimal expectation for the use of cytokines as safety biomarkers is that they are specific, sensitive, and show reliable temporal kinetics that permit detection, persistence, and resolution of toxicity.

CONSIDERATIONS RELATED TO CYTOKINE BIOLOGY

Cytokines are a diverse group of soluble peptides that signal between cells and elicit biological responses, including but not limited to cell activation, proliferation, growth, differentiation, migration, and cytotoxicity. Classically, cytokines were understood in the context of the immune response, whereby sequential cytokine secretion orchestrates inflammation and immunity. T-lymphocyte cytokines divide the Th1 (interferon gamma, IFN γ , and interleukin 12, IL12), Th2 (IL4, IL5, IL6, and IL13), Th17 (IL17, IL21, and IL22), and T-regulatory (transforming growth factor beta, TGF β , and IL10) subsets of T cells that stimulate or modulate the adaptive immune response to infectious agents and other antigens. Macrophages and injured cells secrete chemotactic (chemokines) and proinflammatory cytokines to elicit the innate immune response to sites of active inflammation. Furthermore, colony-stimulating factors and interleukins harmonize myelo- and lymphopoiesis to populate these cellular responses. Under the cytokine umbrella of 5–70 kD, soluble mediators are an extensive network of peptide and glycopeptide families comprising interleukin, interferon, chemokine, tumor

necrosis factor (TNF), and growth factors. Current understanding of cytokine biology underscores their pleiotropic and redundant functionality, widespread expression by non-hematopoietic cell types, and roles outside of immunity in development, reproduction, endocrine regulation, and metabolism (Papanicolaou *et al.*, 1998). A number of cytokines can therefore define a pathologic response but not necessarily a tissue site of toxicity.

Cytokines function as autocrine (secretion and stimulation of the same cell), paracrine and juxtacrine (stimulation of nearby cells), or endocrine signals (circulating in the peripheral blood to act on cells remote to the source of production). The majority of cytokines have induced expression and are secreted or translocated to the cell membrane upon translation (Haddad, 2002). Cytokines such as TGF β and platelet factor 4 (PF4)/CXCL4 are stored in secretory vesicles (in this case platelet granules) for immediate release upon cell activation. Generally, these factors act locally at nano- to picogram per milliliter concentrations have a short half-life and transient activity. These low concentrations and predominately local activity may produce little change in cytokine levels in the systemic circulation despite considerable local perturbation. As an example, IL13 is important in the pathogenesis of asthma and there is interest to explore its use as a systemic disease biomarker of asthma as well as a PD biomarker for therapeutics targeting IL13 (St Ledger *et al.*, 2009). An assay with improved sensitivity for detecting serum IL13 showed no difference in the systemic levels of IL13 in symptomatic asthmatics compared with asymptomatic asthmatics and healthy controls (St Ledger *et al.*, 2009). At least in the cohorts tested in this study, serum IL13 was not a biomarker of asthma. Notable exceptions to this general scheme of local activity are the hematopoietic cytokines (e.g., erythropoietin) that act as endocrine factors to maintain homeostatic set points of blood cells and have a measurable basal blood level.

The cytokine cascade mode of action is illustrated by the inflammatory response. The primary proinflammatory cytokines, comprising TNF- α , IL1, and IL6, are expressed sequentially and amplify cellular activation and recruitment to generate additional cytokines and chemokines. Anti-inflammatory cytokines, principally the IL10 family, are produced early to downregulate proinflammatory cytokines, and TGF β expression contributes to resolution and tissue repair phases. Cytokine cascades result in a staged appearance and disappearance of cytokines in the local and systemic environments, with primary cytokines that drive the early response and are more commonly detectable in peripheral blood. Dysregulation of these cascades can lead to autoimmune disease and hypersensitivity.

Apart from mediating overt immune responses, cytokines play a role in physiological processes. Cytokines participate in maintenance of organ structure and function by tissue-resident macrophages and restoration of homeostasis during “para-inflammatory” states of cell stress (Medzhitov, 2008). Prominent examples of such physiological roles include maintenance of vascular integrity, the interplay of energy metabolism with the

immune system, and neurohumoral stress (Medzhitov, 2008). These activities are theorized to occur largely on a tissue level and are mediated by resident tissue macrophages; any contribution to systemic cytokine levels is unclear.

Cytokine receptors are classified by structural similarities and are primarily cell surface receptors. Soluble receptors can have agonist and antagonist cytokine signaling activity and confer cell extrinsic signaling receptivity to cytokines, e.g., cardiomyocyte hypertrophy in response to soluble IL6 receptor ligation (Papanicolaou *et al.*, 1998). The presence of soluble receptors is also a point of regulating cytokine activity. Cytokine receptors are composed of subunits that form higher order complexes following cytokine binding. The largest receptor class are nontyrosine kinase class I and II receptors that signal through the JAK/STAT pathway. Class I receptors bind the hematopoietic cytokines and many of the interleukins. Heterodimeric class I receptors combine a ligand-specific subunit with a shared signaling receptor (e.g., common beta and gamma chains, and gp130). Other receptors include class II receptors that bind the interferon and IL10 families, the immunoglobulin superfamily receptors that bind the IL1 cytokine family, TNF family receptors, TGF β receptors, and the G-protein-coupled chemokine (or rhodopsin superfamily) receptors (Haddad, 2002). The common use of receptors by cytokine families often confers overlapping signaling outcomes.

Pleiotropy (multiple actions) and redundancy (overlapping actions) are characteristic of cytokine biology. As discussed, shared receptors are a point of integration for multiple cytokine ligands and contribute to redundancy. Pleiotropy occurs when shared receptors modulate different downstream signals determined by a cytokine’s concentration, its relative receptor subunit affinity, and the cell type acted upon. For example, crystallographic analysis of the shared use of receptors by IL4 and IL13 demonstrates that differences in the structural dynamics of receptor-cytokine engagement can specify unique signaling outcomes even when an identical receptor heterodimer in a defined cell and context is engaged (LaPorte *et al.*, 2008). Additional modifiers will include the presence and concentration of other local cytokines and the growth and activation state of the cell (Haddad, 2002). Transitory lateral communication and cross talk with noncytokine receptors also expand the signaling outcome (Bezbradica and Medzhitov, 2009). IL6 exemplifies pleiotropic action by exerting proinflammatory, anti-inflammatory, endocrine, and metabolic effects. These multiple responses include induction of the acute-phase response (e.g., C-reactive protein, fibrinogen), activation of the hypothalamic-pituitary-adrenal axis (with attendant anti-inflammatory effects), growth hormone secretion, and reduction of human serum cholesterol (Papanicolaou *et al.*, 1998). These attributes of pleiotropy and redundancy have implications for the evaluation of cytokines as biomarkers. As a number of cytokines may cause similar responses such as TNF- α , IL1, and IL6, they can be interpreted collectively as markers of inflammation. On the other hand,

each cytokine may have multiple actions, the example given is IL6, and thus lack specificity for one outcome alone.

Cytokine expression and activity are highly regulated to constrain a system that has potential for immunopathology. Only brief mention of some mechanisms is made here and the reader is referred to review articles (Bonecchi *et al.*, 2009; Haddad, 2002; Medzhitov and Horng, 2009). Cytokine networks trigger several members that downregulate the cascade to self-limit a response, i.e., IL10 and TGF β . Cytosolic regulation of receptor activity is primarily by the suppressor of cytokine signaling (SOCS) family with contribution also by protein tyrosine phosphatases and protein inhibitor of activated signal transducer and activator of transcription (PIAS) members. The onset and level of cytokine production are also influenced by posttranscriptional processing by adenine- and uridine-rich elements in the 3' untranslated region of messenger RNA (Anderson, 2008). In the extracellular environment, proteases can regulate cytokine activity and provide a checkpoint for activation of latent factors, i.e., TGF β . The short half-life (usually < 1 hr) of these peptides and the presence of binding proteins and decoy receptors also attenuate functional activity of cytokines extracellularly.

There are species homologues for the majority of cytokines; however, some surprising differences in phylogenetically highly related species are observed. Polymorphisms in the *TNF* promoter differ between nonhuman and human primates and between primate species and subspecies (Baena *et al.*, 2007). For example, divergence in the promoter can alter the binding affinity of transcription factors and transcriptional activation of the *TNF* gene in response to lipopolysaccharide (Baena *et al.*, 2007). Functional single nucleotide polymorphisms of cytokine genes in the human population can result in different levels of expression in healthy individuals and have disease associations with autoimmunity, susceptibility to infection, cardiovascular disease, and cancer (Smith and Humphries, 2009). Cytokine gene polymorphisms in preclinical toxicity species are not well described. An investigation comparing polymorphisms in the regulatory regions of the rat *TNF- α* gene and *in vitro* *TNF- α* release to splenocyte stimulation failed to show a relationship (Warle *et al.*, 2005). Cytokine genes also have different conservation across species. Rodents do not have an ortholog of the human *IL8* gene; rather, the mouse *KC* gene and the rat cytokine-induced neutrophil chemoattractants (*CINC*) family are possibly examples of convergent evolution to produce a neutrophilic chemotactant (Modi and Yoshimura, 1999). These species, subspecies, and strain differences, often not appreciated because of the emerging nature of cytokine analysis in the systemic circulation, require a circumspect approach to cytokine biomarker translatability to humans.

Biological variability of cytokines, also referred to as inter- and intraindividual or between and within subject variability, has not been extensively evaluated in healthy animals or humans. Analysis of IL13 serum levels showed a 10-fold interindividual variability in healthy subjects (< 0.07–1.02 pg/ml). The intra-

individual variability of serum IL13 in asymptomatic asthmatics (whom had no apparent association between serum IL13 and disease) was up to threefold over a 15-day period (St Ledger *et al.*, 2009). Circadian rhythm has also been exhibited by proinflammatory cytokines and linked to changes in corticosteroid and melatonin levels over the 24-h day cycle. Blood levels of IL1, IL6, IFN γ , and TNF- α are highest in the morning (de Jager and Rijkers, 2006). Fascinating is the recent work that establishes the presence of local and cell-autonomous circadian rhythms in lymphoid organs and peritoneal macrophages (Keller *et al.*, 2009). The occurrence of circadian rhythm may contribute not only to changes in measured cytokines at different times of the day but also to the effectiveness and nature of the immune response.

In the context of multiple organ toxicity, key aspects of cytokine biology (pleiotropy, redundancy, and tiered expression) may lead to similar (primary) cytokines showing measurable changes in blood yet not disclosing tissue-specific differences in toxicity. As much of the biological activity of cytokines occurs at a cellular level in local environments, elevations may not be detected in the systemic circulation.

METHODOLOGICAL CONSIDERATIONS FOR CYTOKINE MEASUREMENT

Quantification of soluble cytokines is performed largely by immuno- or bioassay. In contrast to an immunoassay that quantifies the peptide, bioassays demonstrate functional activity of the cytokine. However, as bioassays are not readily scalable or standardized and vary in specificity, the remaining discussion will focus on immunoassays as a more robust format for cytokine measurement in preclinical drug development. Technological advancements in immunoassays have led to improved detection sensitivity and attendant miniaturization, automation, and multianalyte formats. In a less concerted effort, attention has also been directed toward defining the immunoreactive component measured in the assay. The latter concept is explored in greater depth in the following section on method validation.

Platforms can be categorized into those with solid-based supports for the capture antibody, such as a plastic plates and tubes, membranes, and glass slides, or suspension systems using antibody-coupled beads. The conventional ELISA format of a two-site immunometric “sandwich” assay conducted in a microtiter plate and detecting a single analyte can optimally achieve detection limits in the low to mid picogram per milliliter over a ~2 log working range. Bead-based suspension systems address some limitations of the conventional ELISA, e.g., sample volume requirement and assay run time, by increasing the available surface area for the antigen-antibody reaction and employing faster fluid-phase reaction kinetics (Kellar and Iannone, 2002). Detection systems also differentiate platforms. Fluorescence and chemiluminescence have led to claimed sensitivity of single digit picogram per milliliter cytokine concentrations and a 4–5 log working range. A new approach uses digital counting to measure individual fluorescently tagged antigen-antibody complexes in the suspension phase to improve sensitivity (St Ledger *et al.*, 2009).

Multiplexed assays represent a major advance in the measurement of cytokines. In a multiplexed assay, many different cytokines can be measured simultaneously in one specimen aliquot. Common multiplex platforms are flow cytometric assays with fluorescent microspheres and capture antibody-spotted plate-based assays. Clear advantages are obtained through specimen conservation and labor and timesavings. These features make multiplexes a useful screening tool. In combining multiple different antigen-antibody reactions into the one assay, compromises to assay factors such as incubation time, buffers, and

specimen dilution are inevitably made to accommodate measurement of all the analytes (cytokines) together. Such compromises can result in reduced sensitivity and/or dynamic range for some markers and variation between different multiplexes. A method comparison study between a popular electrochemiluminescent plate (MSD) and fluorescent bead (Luminex) platform demonstrated superior sensitivity and accuracy with the MSD assay and better precision (reproducibility) with the Luminex assay (Chowdhury *et al.*, 2009). There have been varied conclusions regarding comparability of the findings when clinical blood specimens have been evaluated by several multiplex platforms and single analyte ELISA. Results will be influenced by the number of cytokines tested, the platforms chosen, the number of specimens analyzed, and type and severity of the disease state or response in the individual (Khan *et al.*, 2004, 2009; Toedter *et al.*, 2008). Generally, different platforms (i.e., bead, plate, single analyte, and multiplex) and assays (i.e., different manufacturers within a platform) are broadly comparable, showing similar trends and response profiles. However, absolute values will usually differ and discrepancies for individual cytokines are not unexpected. Most divergence between assays is likely to occur for cytokines with low to undetectable concentrations in the blood of healthy subjects. A robust approach to selecting the most reliable and informative assay for the toxicity under investigation is to evaluate several assays on specimens reflecting the intended study set prior to analyzing a large study.

Assay automation will influence platform and, possibly, biomarker selection. Manual assays (e.g., a plate-based sandwich ELISA) can be semiautomated with robotics for sample preparation, reagent mixing, incubation, washing, and signal-detection steps. Automation should improve precision by eliminating manual pipetting steps and reduce other sources of random error by minimizing procedural variations. An automated platform could also facilitate the transfer of the safety biomarker into clinical laboratories for clinical trial work. In a realistic scenario, discovery would take place on manual assays to decrease the time to implementation and allow comparison of various assays. The move to automation would occur at later stages of qualification when the biomarker is more promising, assay selection has been made, and the studies are larger and/or more frequent.

In summary, cytokines can be measured in standard single analyte ELISA or as multiplexed plate or bead-based assays evaluate many cytokines at once. Multiplexes provide a convenient and cost-effective approach and useful screening step in early biomarker discovery. As much of the assay performance is conditioned on the specificity and affinity of the antibody reagents, platform choice should be evaluated on the merit of the particular assay for the analyte under examination: there is no clearly superior system.

HOW ASSAY PERFORMANCE AND METHOD VALIDATION AFFECT BIOMARKER DISCOVERY

The method validation of an assay is evaluated by determining a number of parameters including accuracy and precision. This is distinguished from biomarker qualification (the commonly used terminology) that determines the association of a biomarker with a phenotypic end point. The level of assay validation, and the criteria for acceptance, will depend on the stage of drug development, so called “fit for purpose.” Less rigor is advocated in early preclinical drug development with regard to the resources and time required for advanced method validation and the criticality of the decision based on the assay (Lee *et al.*, 2005). In the context of cytokine immunoassays, several of these validation criteria should be better understood by those unfamiliar with assay development but who influence biomarker selection and interpret the data generated in preclinical studies. The following discussion highlights concepts pertinent to cytokine biomarker discovery and qualification; more detail on this topic is found in a position paper on assay validation of biomarkers in drug development (Lee and Hall, 2009; Lee *et al.*, 2005).

In comparison to routine serum biomarkers, i.e., the clinical chemistry panel, there is often a lack of assay standardization for cytokines. This is a major reason that assays differ in the measured value for a cytokine in the same specimen. Calibration of the assay is usually done with recombinant protein in buffer that does not represent the native specimen. Such assays provide relative

quantification rather than definitive or absolute values. As there is no reference standard, and calibrators will vary among assays, values obtained from different assays will not necessarily agree. Accuracy is thus a relative term in the absence of standardization. Changes in recovery and a lack of linearity may occur as a consequence of binding proteins, complex formation, and undefined interferants in the blood. It is advisable to prospectively optimize dilutions that will be applied to specimens from studies to verify manufacturer’s claims and/or establish basic assay performance criteria.

Precision, both within a run and between runs or days, should be verified prior to assay implementation. Precision is usually expressed as the coefficient of variation and can be more important than the problematic criterion of accuracy. A precise assay will provide reproducible results within and between studies and demonstrate relative response, i.e., fold change, and patterns of response for potential biomarkers. Analytical coefficient of variations ranging from 18 to 44% for different cytokines were found for one multiplex (Wong *et al.*, 2008). Yet this degree of analytical imprecision was still less than individual (biologic) variation, permitting real differences in cytokine values to be detected in the population tested (Wong *et al.*, 2008). Generally, an assay for a novel biomarker should have < 25% imprecision (Lee *et al.*, 2005), although there is no hard rule. Quality control (QC) material derived from pooled specimens aliquoted and run on each plate is very useful in both confirming precision during assay validation and bridging results between studies (Chowdhury *et al.*, 2009; Lee and Hall, 2009). These are a very useful extra layer to the QC material that, although sometimes provided by the assay manufacturer, are not entirely representative of the native matrix of study specimens and are subject to lot and source changes. Knowing or establishing the expected biologic variation for a cytokine biomarker and degree of change that corresponds to toxicity will assist in setting this assay performance goal.

The manufacturer (or developer) should provide data on the specificity of the assay for the cytokine under examination, i.e., the percentage of the principle analyte and other structurally related peptides detected by the assay. This is crucial in multiplex assays that could be measuring structurally similar analytes simultaneously. Also, knowledge of the specific form of the cytokine measured impacts evaluation of the cytokine as a biomarker in several ways: by testing the correct form of the cytokine in a hypothesis-driven experiment, accurate description of a putative cytokine biomarker discovered in a nonhypothesis-driven experiment, and comparing cytokine findings across studies when different assays are used. The measurement of TGF β is instructive in this regard. Circulating TGF β comprises several isoforms as well as latent and active peptides and heterogeneous complexes of these peptides (Grainger, 2007). Groups investigating the association of TGF β with atherosclerosis have reported increases, decreases, or no change in blood levels of “TGF β ,” likely because of their measuring different forms (Grainger, 2007). A basic approach here is to know the immunoreactive component or form of the peptide assayed. When that information is lacking, knowledge of the circulating forms and biological action of the cytokine will determine whether a greater level of detail is required to accurately define the biomarker.

Information on the species cross-reactivity of an immunoassay is sometimes lacking. Characterization of mouse cross-reactivity may be provided; however, this is no guarantee that the antibodies detect rat peptides. Similarly, reactivity with human proteins does not imply detection in nonhuman primate specimens. Species cross-reactivity should demonstrate recovery of the protein in the native matrix and the lack of detection when the analyte is absent or intentionally depleted. Experiments using recombinant proteins in buffer are an approximation and do not account for both the heterogeneity of posttranslational modifications in the protein and the interferants present in native specimens.

Reporting of the lower end of the working range of the assay is fraught with misunderstanding and vitally important for distinguishing “trends” at low concentrations. Strictly, the lowest reportable value, or lower limit of quantification (LLOQ, also referred to as functional sensitivity), should at least meet a predetermined precision and accuracy criteria, i.e., $\pm 30\%$ (Lee *et al.*, 2005). This will usually have to be established by the laboratory with a precision curve deriving the precision limits for each point of the calibration curve. Manufacturers often only provide a limit of detection (LOD, also

designated as least detectable dose, minimal detectable concentration, limit of blank, and sensitivity), which pertains to the noise of the assay and is calculated as 2 or 3 SD from the average signal of the buffer (or blank). Values for an analyte that are above but close to the LOD are imprecise and inaccurate until proven otherwise, and much care should be taken both in the reporting and in the interpreting results without knowledge of the LLOQ.

In addition to the aspects of analytical validation outlined above, controlling variability associated with specimen collection and handling should commence at early stages of safety biomarker exploration and qualification. Preanalytical aspects such as the matrix (serum or plasma) and type of anticoagulant, blood tube, collection site, and processing time should be determined and standardized as soon as feasible. Several cytokines can be degraded *in vitro* during delayed sample processing and are more stable in EDTA or citrated plasma (Niwa *et al.*, 2000). Degranulation of platelets and white blood cells (and residual contamination of cells in plasma) may also alter serum as compared with plasma concentrations of several cytokines (Boehlen and Clemetson, 2001; Grainger, 2007; Hosnijeh *et al.*, 2009; Wong *et al.*, 2008). The potential influence of anticoagulant on cytokine values measured by multiplex immunoassays is shown in Table 1. Wong *et al.* (2008) found only 3 of 10 cytokines (IL4, IL6, and IL8) had significant correlation between serum and plasma, although this comparison was made in healthy human volunteers with low levels of cytokines and thus a limited range of values. The table suggests that values in plasma often exceed serum, although there is no clearly superior matrix for cytokine measurement in published studies. Serum is often a pragmatic choice for combining with collections for clinical chemistry and applicability in the clinic.

Tube type can also affect the levels of low-abundant analytes (Ray *et al.*, 2005) and at the very least, the use of nonstandard tubes is discouraged. Specimen stability should also be considered and lengthy storage periods avoided when there is insufficient information available. Several major inflammatory cytokines show little thermal lability and resist several freeze-thaw cycles (Kenis *et al.*, 2002), although this may be epitope dependent (Ray *et al.*, 2005). It should be appreciated that the sooner a specimen matrix and tube type are specified, and stability defined, the variation within and between studies is reduced and study data are maximized. The investigator can more rigorously reanalyze previously collected specimens to bridge results for changes in assay or extend the number of markers analyzed retrospectively.

Last, the challenges of developing a multiplex assay for later regulatory work should be considered during selection of a panel of biomarkers. The most pertinent guidance the FDA (2007) has for multiplex immunoassays pertain to pharmacogenetic tests. There has been little pressure testing of immunoassay multiplexes in a regulatory environment. Ellington *et al.* (2009) performed a large-scale study with human plasma specimens run on two planar plate-based multiplexes that included several cytokines and multiple QC materials. They found potential issues with imprecision, an unidentified systematic bias between plates, and QC failures. On the other hand, Ray *et al.* (2005) validated a 5-plex bead-based assay and showed acceptable assay performance. The group drew attention to postanalytical data management requiring additional points of QC and process management for this assay format.

Method validation occurs incrementally during biomarker discovery and qualification. Understanding key features of cytokine assay performance, particularly precision, specificity, and LLOQ, will allow more informed assay selection and evaluation of data generated during biomarker discovery. Manufacturer's claims of assay performance should be verified prior to assay implementation and specimen matrix and stability defined early to increase the quality of the data generated during biomarker discovery and evaluation.

THE INFLUENCE OF PRECLINICAL TOXICITY STUDY FACTORS ON CYTOKINE INTERPRETATION

The biology of a cytokine can be altered in a toxicity study by the influences of the physicochemical properties of the biologic or pharmaceutical agent, pharmacokinetics, and/or the

TABLE 1
Matrix and Anticoagulant Affect on Cytokine Values Measured by Multiplex Immunoassay in Humans

	Higher ^a	Equivalent ^b	< 50% Reported ^c
IL1 α		C ^d , H ^d , S ^d	
IL1 β	C ^d , A ^e		H ^d , S ^e
IL2	A ^e	C ^d , H ^d , S ^d	S ^e
IL4	A ^e , (H ^e)		C ^d , H ^d , S ^d
IL5	C ^d , S ^d ; A ^e , (H ^e)	H ^d , S ^d	S ^e
IL6	C ^d ; A ^e	H ^d , S ^d , e	
IL7	A ^e , (H ^e)		
IL8	H ^e , S ^e	C ^d , H ^d , S ^d , e	
IL10	A ^e , (H ^e)	C ^d , H ^d , S ^d	
IL12	A ^e , (H ^e)	C ^d , H ^d , S ^d	
IL13	C ^d ; A ^e , (H ^e)		H ^d , S ^d
IFN γ	A ^e , (H ^e)	C ^d , H ^d , S ^d	
TNF- α	C ^d ; A ^e , (H ^e)		H ^d , S ^d
RANTES			C ^d , H ^d , S ^d
Eotaxin	H ^d		C ^d , S ^d

Note. A, acid citrate dextrose plasma; C, citrate plasma; H, lithium heparin plasma; S, serum.

^aCytokines with geometric mean or median values (depending on study) that exceed by $\geq 40\%$ the value in the next highest matrix. When the second highest value also exceeds by 40% the third matrix, this matrix is in parentheses. 40% chosen to exceed the analytical variability (except IL13 in Wong *et al.*, 2008).

^bCytokine geometric mean or median value (depending on study) is within 40% of the highest value for the cytokine.

^cValues not reported are beyond the working range of the assay, i.e., below the lower limit or above the upper limit of detection.

^dHosnijeh *et al.* (2009).

^eWong *et al.* (2008).

perturbation of the system associated with pharmacological and toxicological effects. The design of toxicity studies, in particular the primacy of collecting a standard set of end points to assess toxicity, also imposes constraints on the blood collection schedule and blood volume removed. A stepwise biomarker qualification process in the preclinical space, analogous to the fit for purpose assay validation approach (Lee *et al.*, 2005), is advocated to determine how the putative biomarker performs in different study designs, species, and in lockstep with increasing regulatory rigor during drug development (Fig. 1).

The progress of a therapeutic agent through preclinical development *in vivo* follows a path of increasing study length and more comprehensive end point analyses. Pharmacokinetic studies offer a short window into acute effects of the drug (usually over 24 h) and a view of PD and safety signals in different species. This time course study is a useful design for capturing cytokine modulations and relating changes to both PD markers and drug exposure. As the intent is not to define toxic doses, there may be little in the way of detectable cytokine changes at the exposures tested. Single dose tolerability studies for small molecules do provide an opportunity to sample the animal for toxicity biomarkers with respect to driving the system to toxic thresholds. Important considerations in the interpretation

Preclinical Toxicity Biomarker Qualification

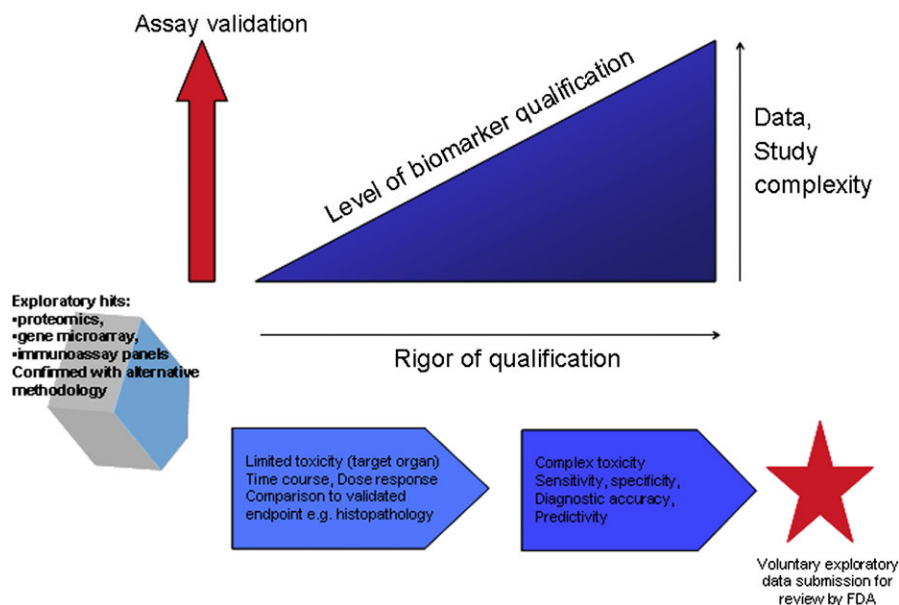


FIG. 1. Preclinical toxicity biomarker qualification.

of cytokines in these acute studies are the impact of stress, diurnal variations of the analyte, and blood volume removed. A single, large volume of blood collected has been shown to induce cytokine gene expression in the liver and lung of mice (Rajnik *et al.*, 2002). Excitement and stress triggered in toxicity studies with handling, blood collection, and tissue damage have the potential to increase expression of IL6 (Papanicolaou *et al.*, 1998). In using pharmacokinetic and tolerability studies as opportunities for biomarker discovery, it is important to control some of the study-related variables that may impact cytokine expression independent of the therapeutic agent. Baseline cytokine blood levels (predose) should be taken on all animals to control for individual differences not related to the drug. However, baselines alone may not be sufficient to control for diurnal changes, different stress levels during the study, individual biological variability unrelated to the therapeutic, and progressive blood volume reduction. Contemporaneous vehicle-treated control animals matched for age and sex, with a similar group size, are therefore a necessary additional control for these factors. Another key strategy is to reduce analytical variability by ensuring that specimens over the entire time course and between groups are randomly allocated to assay runs.

Multidose studies, usually ranging from several weeks to months, are conducted to evaluate toxicity associated with prolonged drug exposure and support human dosing of the therapeutic agent. Incorporating biomarker discovery and qualification into this study type permits longitudinal analysis and the association of the biomarker with subacute and subchronic toxicities. Biological variability in the levels of

cytokines related to aging and ovarian cycles (Brannstrom *et al.*, 1999; Cannon, 2000) may become evident in the longer multidose study format. As depicted in the schematic for preclinical toxicity biomarker qualification (Fig. 1), a study with limited, i.e., single organ, versus complex (multiorgan) toxicity necessitates a higher degree of assay validation and study rigor, i.e., controls, to support biomarker evaluation. Suggested time points for cytokine analysis are baseline (prestudy), predose, and postdose. The predose sampling is useful over longer studies to capture a shifting baseline or persistence of previous recorded changes in cytokine levels. Postdose time points are typically acute and multiple, e.g., 1–6 h and 24 h, and may be informed by the time to maximal concentration in blood of the therapeutic agent or data from previous biomarker studies of cytokine response kinetics to the toxicity under investigation. This multisampling paradigm accommodates the transience of biomarker changes and the potential for a modified cytokine response in the transition from acute to chronic (long-standing) toxicities, reflecting the altered cellular and tissue environment. Collecting sufficient blood for reanalysis of specimens by a second platform is best practice but entirely contingent on the stage of qualification of the biomarker and the size of study animals. Blood volume limitations, dependent on the animal size and species, may not accommodate the proposed cytokine biomarker collection schedule. Species selection (large vs. small animal), use of larger animals with a greater circulating blood volume, dividing time points into cohorts, and satellite groups are strategies to not exceed blood volume restrictions.

The rigor of multidose studies can also benefit biomarker discovery and qualification by decreasing variability. Standardized processes that decrease variability include specified blood collection time points in relation to both time of day, dosing, feeding, specified collection sites, anesthetic regimens (if used), specimen matrix, and blood processing and handling procedures. Use of vehicle-control animals, typically included in multidose toxicity studies, are key to account for biological variability when interpreting cytokine data. A power analysis leveraging what is known on biological variability or cytokine response can also select the appropriate number of animals to demonstrate a drug-related change in cytokine value. Similar to the recommendation for biomarker discovery in single dose studies, random assignment of specimens to assay runs is advised. In these longer studies, specimen stability should be conducted prestudy to facilitate the suggested randomization of specimens from time points and groups; in addition, QC materials in relevant species and strain matrix should be included in each run to accept runs and bridge results within a long or large study.

Misinterpretation of biomarker data can occur when pharmacological or toxicological responses *in vivo* result in factors that interfere with cytokine assays. Analytical interference attributable to the physicochemical properties of the therapeutic should be addressed prestudy to assist in selection of an assay or to develop techniques to circumvent the interference. Cytokine analogues can generate anticytokine antibodies both to the protein therapeutic itself (de Lemos Rieper *et al.*, 2009) or to the structurally similar endogenous cytokines. Antitherapeutic antibodies (ATA) to host cytokines may have physiological consequences depending on the nature of the autoantibody, e.g., neutralizing, and potentially interfere with *in vitro* cytokine measurement (de Jager and Rijkers, 2006). Therapeutic antibodies could elicit low-avidity heterophilic (nonspecific) antibodies in the animal that cross-react and bridge the capture and detection antibodies in an immunoassay and result in false-positive results. Interference by heterophilic antibodies or ATA can be verified by depletion of immunoglobulins from the specimen and rerunning the specimen, correlating ATA to cytokine levels, and showing nonlinearity upon dilution of the sample. These interferences can also be removed by similar procedures (depletion of dilution). In addition, some assay manufacturers include a diluent that blocks interference from host antibodies.

Both protein therapeutics and small molecule agents have the potential to induce autoantibodies against cytokines that undergo a chronic course of stimulation by the therapeutic. High affinity autoantibodies against cytokines are observed in both healthy and diseased human populations and have been detected in rats and mice (de Lemos Rieper *et al.*, 2009; Watanabe *et al.*, 2007). In humans, autoantibodies are found to IL1, IL2, IL6, IL8, G-CSF, TNF- α , and VEGF. These antibodies are often neutralizing and have been associated with pathology related to depletion of the cytokine (de Lemos

Rieper *et al.*, 2009). Autoantibodies most often will spuriously decrease biomarker values and can be corrected by dilution of the specimen (de Jager and Rijkers, 2006). In preclinical toxicity studies, the presence of cytokine autoantibodies that arise from chronic stimulation is not known, nor whether this could contribute to a “false-negative” test result, or modification of the toxicologic response *in vivo*.

Species and strain differences in cytokine expression can exert a considerable influence on the evaluation of cytokine biomarkers of toxicity. The differences could be as rudimentary as susceptibility to the toxicity itself. The strain resistance variation of mice to bleomycin-induced pulmonary fibrosis is attributable in part to differences in the induction of cytokine receptor expression in the lung (Cavarra *et al.*, 2004). A log-fold difference in TNF- α production in response to *in vitro* splenocyte stimulation has been demonstrated between rat strains (Warle *et al.*, 2005). The availability of reagents has allowed characterization of many of the wide ranging differences between human and mouse immune systems that impact cytokine expression. Some of the underlying differences include variations in immune cell subsets and localization, receptor expression, signal transduction, and differences in evolutionary retention and divergence of genes for cytokines and chemokines in mouse compared with human (Mestas and Hughes, 2004). Distinctions between human and rat, dog, and nonhuman primate cytokine repertoire and responses are not as well described (Piccotti *et al.*, 2009).

The TGN1412 anti-CD28 molecule exemplifies the most disastrous consequence of species differences in cytokine response. Cynomolgous macaques did not predict the exaggerated immune stimulation (cytokine storm) experienced in the phase I clinical trial (Suntharalingam *et al.*, 2006). However, studies examining species differences in the immunostimulatory response to this biologic provide important information for future guidances in preclinical safety assessment of targets of this type. In subsequent *in vitro* lymphocyte stimulation assays, macaque lymphocytes, while not showing the 2–3 log induction of TNF- α and IFN γ of human lymphocytes, exhibited a modest 18-fold increase in IL6 and IL5 (Muller and Brennan, 2009). A surrogate anti-rat-CD28 antibody (JJ316) also failed to elicit a cytokine storm in the rat (Muller *et al.*, 2008). The profound lymphopenia and lymphocyte redistribution shown by human patients was observed in the rat, in addition to the pan-T-cell activation. Furthermore, sorted naïve and adopted effector T cells from JJ316-treated rats had upregulated transcription of cytokines (IFN γ , IL17, and MCP-1) and surface expression of cell activation markers. However, cytokine levels in rat serum (IFN γ and TNF- α) were only mildly elevated. Various rat models with an immunostimulatory or autoimmune phenotype also failed to exhibit a cytokine storm in response to anti-CD28 stimulation (Muller *et al.*, 2008). Both the macaque and the rats substantially underestimated the immunostimulation of TGN1412 anti-CD28 in humans.

TABLE 2
Commonly Modulated Blood Cytokines Associated with Pathological Responses

Pathological response ^{a,b}	Cytokines ^c
Acute-phase response	IL1 β , IL6, TNF- α
Cytokine storm/release	IL2, IL6, IL8, IL10, IFN γ , TNF- α
Fibrosis	TGF β
Hemophagocytic syndrome	IFN γ , IL1 β , IL6, TNF- α
Neutrophilic inflammation	IL8, MIP-1, TNF- α
Systemic inflammatory response syndrome	IL6, MCP-1, TNF- α
Th1 immune response	IFN γ , IL2, IL12
Th2 immune response	IL4, IL5, IL6, IL10, IL13

^aPathological responses observed in toxicity studies, includes inflammation, immunity, and repair.

^bDescriptions of immunostimulatory toxic responses are in Gribble *et al.* (2007).

^cCytokines that drive the response and are most commonly at detectable systemic levels.

Cytokines can be explored as safety biomarkers in a traditional toxicity study designs and thus are an important proving ground for biomarker qualification. Toxicity study design factors can impact cytokine responses; therefore, robust study controls are necessary to accurately attribute changes in cytokine values to toxicity.

CYTOKINES AS BIOMARKERS OF TOXICITY

Cytokines have been rationally evaluated as biomarkers of intended and unintended inflammation and immunomodulation or uncovered by systems biology approaches during toxicity biomarker discovery. Biomarker exploration typically analyzes early time points to reveal the temporal response and by doing so will detect proximal pathophysiological processes. Candidate markers that are significantly changed in biomarker discovery comprise components of on-target or off-target pathways, tissue-specific response to injury, and/or represent a general process once the cell injury has occurred. For an uncomplicated toxicity limited to one organ system or one mechanism, an early marker of a general process such as inflammation may directly correlate with the toxicity and have adequate specificity for the tissue injury. Specificity of an inflammatory biomarker becomes difficult to attain when there are multiple mechanisms of toxicity and/or organ involvement that could affect cytokine values. Typical scenarios are decreased leukocyte numbers affecting cytokine production, liver injury that impairs synthesis of binding proteins and clearance, renal dysfunction affecting clearance, and compromise of the intestinal barrier leading to endotoxin translocation and stimulation of inflammatory mediators. The hallmark cytokines for several pathologic responses observed in toxicity studies (Table 2) exemplify several concepts of cytokine biology, namely that primary cytokines are key drivers of inflammation and immunity and the expected overlap reflects

cytokine pleiotropy and redundancy of action. Changes in the toxic stimulus (dose and pharmacologic variables), site of action, and coexisting toxicities may provide an additional retinue of cytokines for each pathologic response. In addition, these key cytokines are frequently but not always detected in the blood.

Routine clinical pathology analyses, comprising clinical chemistry, hematology, coagulation panels, and urinalysis, constitute a powerful set of markers that usually capture general processes of inflammation and tissue injury. On occasion, these markers may not provide adequate sensitivity for low-grade or focal inflammation or could be compromised by coexisting toxicity, i.e., myelosuppression reducing the number of neutrophils. This situation may yield a gap in markers of tissue inflammation and repair. When no such gap exists, critical evaluation of the cost benefit of a new cytokine marker in comparison with routine clinical pathology parameters is necessary. This comparison should include the additional cost of analysis, strength of association with the tissue damage, predictive value, and whether application of the marker would translate to improved patient safety.

Correlation to the severity of tissue injury is an important attribute of a toxicity biomarker. This is not always a characteristic of proximal markers of disease, such as cytokines. Proximal markers of toxicity could reflect the pharmacokinetics of the test compound (peak serum concentration and exposure) more so than sustained or distal disease processes that contribute largely to structural damage and/or organ dysfunction. Lesions with a subacute to chronic course, and a multidose regimen, may elicit cyclic fluctuations of cytokines that are not linearly related to the histological findings. Moreover, counterregulatory changes to cytokine release may result in abrogation, diminution, or an altered time course of cytokines following multiple dosing of a therapeutic. The level of IFN γ , a PD and toxicity biomarker of recombinant human IL12 administration, is maximal after the first dose and then markedly downregulated, concomitant with increased expression of IL12 receptor (Rakhit *et al.*, 1999). Furthermore, the decrease in an early mediator may not imply that the lesion is resolving but rather connote a short circulating half-life and systemic levels that do not reflect local concentrations and activity. Illustrating the latter point is the extensive variability of a cytokine panel in patients with chronic periodontal disease (Gorska *et al.*, 2003). Although tissue cytokine levels correlate to disease severity by microscopic examination of the gingiva, the overwhelming biological (interindividual) variability in the disease population for serum cytokines precluded their use as accurate diagnostic biomarkers for this disease. This inherent and often unexplained biological variability was also demonstrated in a clinical study of acute experimental endotoxemia. There was up to a 10-fold range in baseline and peak values of IL6 and TNF- α that was unassociated with *TNF* genotype (Kovar *et al.*, 2007).

The strength of association with existing toxicity end points is a key criterion in selecting single or multiple biomarkers. Cytokines in a multiplex assay can be evaluated individually

or, by using combinatorial analysis, as a group. Multivariate statistical techniques, with correction for multiple tests, can be applied to the analysis of these large data sets. In addition, bioinformatics techniques such as principal component analysis (Wong *et al.*, 2008) and hierarchical cluster analysis (Khan *et al.*, 2009; Panelli *et al.*, 2004) can be used to detect relationships between multiple analytes and between specimens. By clustering similarly responding cytokines, a biological response may become evident. Hence, these methods are a strategy to contend with the redundancy and pleiotropic action of cytokines. Quality data (controls, preanalytical and analytical variability reduced) and statistical methods to control false positives are essential prerequisites to discovering true associations.

Serum cytokines have been considered for toxicities where serum biomarkers are absent or inadequately premonitory, such as drug-induced liver and vascular injury (Lacour *et al.*, 2005) (Kerns *et al.*, 2005). There is, however, little published on the utility of cytokine biomarkers in the assessment of these toxicities to date. A study of acetaminophen overdose in a clinical population demonstrated an increase in IL8, IL6, and MCP-1 in the most severely affected patients (based on serum ALT), yet only MCP-1 had a good correlation ($R^2 = 0.607$) over the range of severity (James *et al.*, 2005). MCP-1 did not show an association with serum acetaminophen level or the Rumack-Matthew nomogram for estimating risk of hepatotoxicity after acetaminophen overdose (James *et al.*, 2005).

Several phosphodiesterase IV (PDE4) inhibitors induce inflammatory vascular injury in preclinical species. Serum levels of cytokines (IL6, CINC-1, and VEGF), acute-phase response proteins (haptoglobin and $\alpha 1$ acid glycoprotein), and neutrophils show time- and dose-related increases and a relationship to histological severity of the vasculitis caused by the two PDE4 inhibitors under examination, SCH 351591 and SCH 534385 (Weaver *et al.*, 2008). However, drug-induced vasculitis is not observed for all drugs inhibiting this target, possibly relating to differences in drug selectivity between phosphodiesterase subtypes (Dietsch *et al.*, 2006). Toxicity profiling of the PDE4 inhibitor IC542 showed inflammation in multiple tissues without prominent vasculitis, yet an overlap in the repertoire of biomarkers for IC542-induced inflammation and those previously described for SCH 351591/SCH 53438-induced vasculitis (Dietsch *et al.*, 2006). Accordingly, the utility of inflammatory markers as an indication of vasculitis cannot be generalized for all PDE4 inhibitors and is not surprising given the lack of specificity of these biomarkers.

Measurement of serum cytokines may have most utility in immunotoxicity studies that evaluate intended or unintended inflammation and immunomodulation produced by therapeutics. Toxicities associated with immunostimulatory molecules include the acute-phase response, cytokine storm (also known as cytokine release syndrome, systemic inflammatory response, and multiple organ dysfunction syndrome), vascular leak syndrome, vasculitis, antibody-mediated cytopenia, hemophagocytic syndrome, immune complex disease, local tissue injury,

e.g., liver, kidney, skin, lung, and first dose effect (Gribble *et al.*, 2007). There is a dearth of published data on the correlation and accuracy, i.e., predictive values, receiver operator characteristics, etc., of select serum cytokines and specific toxicities. Concerns in the industry have been expressed over the significance of small magnitude changes of cytokines and the lack of defined dose-response thresholds for pharmacology (if on target), reversible cell injury, and toxicity that translates to the clinic (Piccotti *et al.*, 2009). Given that adverse events with immunostimulatory agents are usually acute and predicated on changes in standard, i.e., accepted by regulatory bodies, symptomatology, physical exam, diagnostic, and laboratory parameters, the measurement of cytokines may provide *post hoc* mechanistic data more so than premonitory markers sufficient to guide intervention or dose modulation ahead of adverse events.

Successful use of a cytokine biomarker to guide a program is demonstrated by the monitoring of serum IFN γ for the immunostimulatory toxicity associated with iv recombinant human IL12 (rHuIL12) administration. In humans, cynomolgus monkeys, and mice, species-specific rIL12 resulted in large modulations of serum IFN γ corresponding to gastrointestinal toxicity, multiorgan dysfunction, and death. Cytokines were measured 24 h after each daily dose. Interestingly, mice and humans showed similar IFN γ response kinetics with a peak after the second or third dose (Leonard *et al.*, 1997). Serious adverse events in patients occurred after two daily iv doses of rHuIL12. It was not investigated whether IFN γ measurement prior to 24 h after first dose would have proven to be premonitory for these adverse events. By using IFN γ as a marker of toxicity, schedule changes, route changes, and further mechanistic investigations were made possible (Leonard *et al.*, 1997; Rakhit *et al.*, 1999). Recombinant human IL18 is an immunostimulatory therapeutic anticipated to have similar activity to IL12 and potentially similar toxicity. The toxicity profile of five daily iv infusions of recombinant IL18 in the clinic was monitored by measuring IFN γ , granulocyte-macrophage colony-stimulating factor (GM-CSF) and IL12 prior to the first dose, 24 h after each dose, and at 6 and 12 h after the first and fifth dose. Peaks of IFN γ and GM-CSF occurred 6 h after the first dose and resolved by 24 h postdose. Blood level of IFN γ was lower than that measured at a similar time point after rHuIL12 and corresponded to the milder toxicity of rHuIL18 (Robertson *et al.*, 2006).

Current guidance documents for preclinical filing of pharmaceuticals (S8: Immunotoxicity Studies for Human Pharmaceuticals) address only unintended immunomodulation and advocate additional immunotoxicity studies to characterize risk (International Conference on Harmonization, 2006). Broadly, immunotoxicity evaluations advanced by the S8 guidance include *ex vivo* immunophenotyping of blood cells, immune cell function *in vitro*, *in vivo* immune challenges (e.g., T-cell-dependent immune response), and host resistance, as well as extended histopathological examination of lymphoid tissue in standard animal toxicity studies. Serum cytokine measurement

does not have industry-wide adoption in first tier immune function evaluations (Piccotti *et al.*, 2009). Concerns over species translation (or sensitivity) and detection of immunomodulation have in part driven alternative approaches to risk assessment (Muller and Brennan, 2009). Regulatory guidance for selecting a safe starting dose in human trials now includes the minimal anticipated biological response calculated by pharmacokinetics and PD markers in addition to a no adverse effect level determined in preclinical *in vivo* toxicity studies (EMA, 2007).

CONCLUSIONS

With the encouragement of regulatory agencies to improve development of innovative and safe drugs, extant gaps in toxicity biomarkers are being addressed. Cytokines are not uncommonly identified in this push to improve biomarker repertoire, facilitated in part by the excellent technical advances in multiplexed immunoassays, and their role as integral components of inflammation, repair, and immunomodulation processes in toxicity. Ultimately, toxicity biomarkers are most useful when sensitive, specific, and predictive, and having kinetics consonant with tissue injury, tissue dysfunction, or the mechanism of toxicity. Cytokine fluctuations can be sensitive but may be too acute (proximal) to correlate with the severity of tissue injury when distal processes dominate. When cytokines are key drivers of acute toxicity, namely immunostimulation, their use as biomarkers is more successful. In spite of this, biological variability, assay sensitivity, and short half-life still remain obstacles. The early appearance of cytokines and short half-life offer definite advantages as mechanistic markers and in modeling exposure-activity-toxicity relationships.

Practically, investigators should use several platforms for cytokine biomarker discovery, only directly compare results across studies from the same assay, and validate all assays in-house to determine assay performance on the species specimen of interest. Using a multiplex assay as a screening tool is undeniably useful and a good starting point; however, an additional multiplex or single analyte assay(s) should be completed to confirm initial findings.

Multiplexes also facilitate the testing of cytokine panels as an innovative solution to toxicity biomarker discovery and mechanistic understanding of toxicity. Combinatorial analysis of cytokine biomarker panels could exploit a unique pattern of “general” markers to provide specificity. This is a departure from toxicity detection with one biomarker in isolation or subjective pattern recognition. However, there are few examples of this combinatorial approach used in preclinical risk assessment so far; therefore, analytical validation and robust qualification will require some trail blazing and commitment.

We have amassed a huge amount of knowledge that informs our interpretation of traditional serum biomarkers, including the timing of specimen collection, standardized measurement,

species differences and interpretation of changes in the context of health (biological variability), toxicity, and multiple toxicities. We need to actively seek or generate this information when assessing serum cytokines as safety biomarkers. Understanding the complex pathophysiology of multiple coexisting toxicities and taking a “whole” animal integrated perspective is vital to judge the value of a cytokine during biomarker qualification. With the expected overlap of cytokines seen during different pathologic processes and the inherent biological variability of systemic levels of cytokines, identification of a specific cytokine or panel of cytokines, together with setting a threshold (decision limit) of systemic levels for toxicity, is challenging. Few cytokine biomarkers that are predictive of specific tissue toxicities have emerged as yet.

ACKNOWLEDGMENTS

It is with the utmost gratitude that I acknowledge Wendy Halpern and Donna Dambach for their insight and suggestions during manuscript review.

REFERENCES

- Anderson, P. (2008). Post-transcriptional control of cytokine production. *Nat. Immunol.* **9**, 353–359.
- Baena, A., Mootnick, A. R., Falvo, J. V., Tsytskova, A. V., Ligeiro, F., Diop, O. M., Brieva, C., Gagneux, P., O'Brien, S. J., Ryder, O. A., *et al.* (2007). Primate TNF promoters reveal markers of phylogeny and evolution of innate immunity. *PLoS One* **2**, e621.
- Bezbradica, J. S., and Medzhitov, R. (2009). Integration of cytokine and heterologous receptor signaling pathways. *Nat. Immunol.* **10**, 333–339.
- Biomarkers Definitions Working Group. (2001). Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin. Pharmacol. Ther.* **69**, 89–95.
- Boehlen, F., and Clemetson, K. J. (2001). Platelet chemokines and their receptors: what is their relevance to platelet storage and transfusion practice? *Transfus. Med.* **11**, 403–417.
- Bonocchi, R., Galliera, E., Borroni, E. M., Corsi, M. M., Locati, M., and Mantovani, A. (2009). Chemokines and chemokine receptors: an overview. *Front. Biosci.* **14**, 540–551.
- Brannstrom, M., Friden, B. E., Jasper, M., and Norman, R. J. (1999). Variations in peripheral blood levels of immunoreactive tumor necrosis factor alpha (TNFalpha) throughout the menstrual cycle and secretion of TNFalpha from the human corpus luteum. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **83**, 213–217.
- Cannon, J. G. (2000). Inflammatory cytokines in nonpathological states. *News Physiol. Sci.* **15**, 298–303.
- Cavarra, E., Carraro, F., Fineschi, S., Naldini, A., Bartalesi, B., Pucci, A., and Lungarella, G. (2004). Early response to bleomycin is characterized by different cytokine and cytokine receptor profiles in lungs. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **287**, L1186–L1192.
- Chowdhury, F., Williams, A., and Johnson, P. (2009). Validation and comparison of two multiplex technologies, Luminex and Mesoscale Discovery, for human cytokine profiling. *J. Immunol. Methods* **340**, 55–64.
- de Jager, W., and Rijkers, G. T. (2006). Solid-phase and bead-based cytokine immunoassay: a comparison. *Methods* **38**, 294–303.

- de Lemos Rieper, C., Galle, P., and Hansen, M. B. (2009). Characterization and potential clinical applications of autoantibodies against cytokines. *Cytokine Growth Factor Rev.* **20**, 61–75.
- Dietsch, G. N., Divalpa, C. R., Eyre, R. J., Pham, T. Q., Poole, K. M., Pefaur, N. B., Welch, W. D., Trueblood, E., Kerns, W. D., and Kanaly, S. T. (2006). Characterization of the inflammatory response to a highly selective PDE4 inhibitor in the rat and the identification of biomarkers that correlate with toxicity. *Toxicol. Pathol.* **34**, 39–51.
- Ellington, A. A., Kullo, I. J., Bailey, K. R., and Klee, G. G. (2009). Measurement and quality control issues in multiplex protein assays: a case study. *Clin. Chem.* **55**, 1092–1099.
- European Medicines Agency (EMA). (2007). *Guideline on Strategies to Identify and Mitigate Risks for First-in-Human Clinical Trials with Investigational Medicinal Products*. European Medicines Agency: Committee for Medicinal Products for Human Use Available at: <http://www.ema.europa.eu/pdfs/human/swp/2836707enfin.pdf>. Accessed November, 2009.
- European Medicines Agency (EMA). (2008). *Final Report on the Pilot Joint EMEA/FDA VXDS Experience on Qualification of Nephrotoxicity Biomarkers* Available at: <http://www.ema.europa.eu/pdfs/human/biomarkers/25088508en.pdf>. Accessed November, 2009.
- Food and Drug Administration (FDA). (2006). *Innovation or Stagnation: Critical Path Opportunities Report and List* Available at: <http://www.fda.gov/ScienceResearch/SpecialTopics/CriticalPathInitiative/default.htm>. Accessed November, 2009.
- Food and Drug Administration (FDA). (2007). *Guidance on Pharmacogenetic Tests and Genetic Tests for Heritable Markers*. Available at: <http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm077862.htm>. Accessed November 2009.
- Goodsaid, F. M., Frueh, F. W., and Mattes, W. (2008). Strategic paths for biomarker qualification. *Toxicology* **245**, 219–223.
- Gorska, R., Gregorek, H., Kowalski, J., Laskus-Perendyk, A., Syczevska, M., and Madalinski, K. (2003). Relationship between clinical parameters and cytokine profiles in inflamed gingival tissue and serum samples from patients with chronic periodontitis. *J. Clin. Periodontol.* **30**, 1046–1052.
- Grainger, D. J. (2007). TGF-beta and atherosclerosis in man. *Cardiovasc. Res.* **74**, 213–222.
- Gribble, E. J., Sivakumar, P. V., Ponce, R. A., and Hughes, S. D. (2007). Toxicity as a result of immunostimulation by biologics. *Expert Opin. Drug Metab. Toxicol.* **3**, 209–234.
- Haddad, J. J. (2002). Cytokines and related receptor-mediated signaling pathways. *Biochem. Biophys. Res. Commun.* **297**, 700–713.
- Hosnijeh, F. S., Krop, E. J., Portengen, L., Rabkin, C. S., Linseisen, J., Vineis, P., and Vermeulen, R. (2009). Stability and reproducibility of simultaneously detected plasma and serum cytokine levels in asymptomatic subjects. *Biomarkers* **15**, 140–148.
- International Conference on Harmonization. (2006). *International Conference on Harmonization (ICH)—Guidance for Industry: S8 Immunotoxicity Studies for Human Pharmaceuticals* Available at: <http://www.fda.gov/RegulatoryInformation/Guidances/ucm129118.htm>. Accessed November, 2009.
- James, L. P., Simpson, P. M., Farrar, H. C., Kearns, G. L., Wasserman, G. S., Blumer, J. L., Reed, M. D., Sullivan, J. E., and Hinson, J. A. (2005). Cytokines and toxicity in acetaminophen overdose. *J. Clin. Pharmacol.* **45**, 1165–1171.
- Kellar, K. L., and Iannone, M. A. (2002). Multiplexed microsphere-based flow cytometric assays. *Exp. Hematol.* **30**, 1227–1237.
- Keller, M., Mazuch, J., Abraham, U., Eom, G. D., Herzog, E. D., Volk, H. D., Kramer, A., and Maier, B. (2009). A circadian clock in macrophages controls inflammatory immune responses. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 21407–21412.
- Kenis, G., Teunissen, C., De Jongh, R., Bosmans, E., Steinbusch, H., and Maes, M. (2002). Stability of interleukin 6, soluble interleukin 6 receptor, interleukin 10 and CC16 in human serum. *Cytokine* **19**, 228–235.
- Kerns, W., Schwartz, L., Blanchard, K., Burchiel, S., Essayan, D., Fung, E., Johnson, R., Lawton, M., Loudon, C., MacGregor, J., et al. (2005). Drug-induced vascular injury—a quest for biomarkers. *Toxicol. Appl. Pharmacol.* **203**, 62–87.
- Khan, I. H., Krishnan, V. V., Ziman, M., Janatpour, K., Wun, T., Luciw, P. A., and Tuscano, J. (2009). A comparison of multiplex suspension array large-panel kits for profiling cytokines and chemokines in rheumatoid arthritis patients. *Cytometry B Clin. Cytom.* **76**, 159–168.
- Khan, S. S., Smith, M. S., Reda, D., Suffredini, A. F., and McCoy, J. P., Jr. (2004). Multiplex bead array assays for detection of soluble cytokines: comparisons of sensitivity and quantitative values among kits from multiple manufacturers. *Cytometry B Clin. Cytom.* **61**, 35–39.
- Kovar, F. M., Marsik, C., Cvitko, T., Wagner, O. F., Jilma, B., and Endler, G. (2007). The tumor necrosis factor alpha -308 G/A polymorphism does not influence inflammation and coagulation response in human endotoxemia. *Shock* **27**, 238–241.
- Lacour, S., Gautier, J. C., Pallardy, M., and Roberts, R. (2005). Cytokines as potential biomarkers of liver toxicity. *Cancer Biomark.* **1**, 29–39.
- LaPorte, S. L., Juo, Z. S., Vaclavikova, J., Colf, L. A., Qi, X., Heller, N. M., Keegan, A. D., and Garcia, K. C. (2008). Molecular and structural basis of cytokine receptor pleiotropy in the interleukin-4/13 system. *Cell* **132**, 259–272.
- Lee, J. W., and Hall, M. (2009). Method validation of protein biomarkers in support of drug development or clinical diagnosis/prognosis. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **877**, 1259–1271.
- Lee, J. W., Weiner, R. S., Sailstad, J. M., Bowsher, R. R., Knuth, D. W., O'Brien, P. J., Fourcroy, J. L., Dixit, R., Pandite, L., Pietrusko, R. G., et al. (2005). Method validation and measurement of biomarkers in nonclinical and clinical samples in drug development: a conference report. *Pharm. Res.* **22**, 499–511.
- Leonard, J. P., Sherman, M. L., Fisher, G. L., Buchanan, L. J., Larsen, G., Atkins, M. B., Sosman, J. A., Dutcher, J. P., Vogelzang, N. J., and Ryan, J. L. (1997). Effects of single-dose interleukin-12 exposure on interleukin-12-associated toxicity and interferon-gamma production. *Blood* **90**, 2541–2548.
- Medzhitov, R. (2008). Origin and physiological roles of inflammation. *Nature* **454**, 428–435.
- Medzhitov, R., and Horng, T. (2009). Transcriptional control of the inflammatory response. *Nat. Rev. Immunol.* **9**, 692–703.
- Mestas, J., and Hughes, C. C. (2004). Of mice and not men: differences between mouse and human immunology. *J. Immunol.* **172**, 2731–2738.
- Modi, W. S., and Yoshimura, T. (1999). Isolation of novel GRO genes and a phylogenetic analysis of the CXC chemokine subfamily in mammals. *Mol. Biol. Evol.* **16**, 180–193.
- Muller, N., van den Brandt, J., Odoardi, F., Tischner, D., Herath, J., Flugel, A., and Reichardt, H. M. (2008). A CD28 superagonistic antibody elicits 2 functionally distinct waves of T cell activation in rats. *J. Clin. Invest.* **118**, 1405–1416.
- Muller, P. Y., and Brennan, F. R. (2009). Safety assessment and dose selection for first-in-human clinical trials with immunomodulatory monoclonal antibodies. *Clin. Pharmacol. Ther.* **85**, 247–258.
- Niwa, Y., Akamatsu, H., Sumi, H., Ozaki, Y., and Abe, A. (2000). Evidence for degradation of cytokines in the serum of patients with atopic dermatitis by calcium-dependent protease. *Arch. Dermatol. Res.* **292**, 391–396.
- Panelli, M. C., White, R., Foster, M., Martin, B., Wang, E., Smith, K., and Marincola, F. M. (2004). Forecasting the cytokine storm following systemic interleukin (IL)-2 administration. *J. Transl. Med.* **2**, 17.
- Papanicolaou, D. A., Wilder, R. L., Manolagas, S. C., and Chrousos, G. P. (1998). The pathophysiologic roles of interleukin-6 in human disease. *Ann. Intern. Med.* **128**, 127–137.
- Piccotti, J. R., Lebecq, H. N., Evans, E., Herzyk, D. J., Hastings, K. L., Burns-Naas, L. A., Gourley, I. S., Wierda, D., and Kawabata, T. T. (2009).

- Summary of a workshop on nonclinical and clinical immunotoxicity assessment of immunomodulatory drugs. *J. Immunotoxicol.* **6**, 1–10.
- Rajnik, M., Salkowski, C. A., Thomas, K. E., Li, Y. Y., Rollwagen, F. M., and Vogel, S. N. (2002). Induction of early inflammatory gene expression in a murine model of nonresuscitated, fixed-volume hemorrhage. *Shock* **17**, 322–328.
- Rakhit, A., Yeon, M. M., Ferrante, J., Fettner, S., Nadeau, R., Motzer, R., Bukowski, R., Carvajal, D. M., Wilkinson, V. L., Presky, D. H., *et al.* (1999). Down-regulation of the pharmacokinetic-pharmacodynamic response to interleukin-12 during long-term administration to patients with renal cell carcinoma and evaluation of the mechanism of this “adaptive response” in mice. *Clin. Pharmacol. Ther.* **65**, 615–629.
- Ray, C. A., Bowsher, R. R., Smith, W. C., Devanarayan, V., Willey, M. B., Brandt, J. T., and Dean, R. A. (2005). Development, validation, and implementation of a multiplex immunoassay for the simultaneous determination of five cytokines in human serum. *J. Pharm. Biomed. Anal.* **36**, 1037–1044.
- Robertson, M. J., Mier, J. W., Logan, T., Atkins, M., Koon, H., Koch, K. M., Kathman, S., Pandite, L. N., Oei, C., Kirby, L. C., *et al.* (2006). Clinical and biological effects of recombinant human interleukin-18 administered by intravenous infusion to patients with advanced cancer. *Clin. Cancer Res.* **12**, 4265–4273.
- Smith, A. J., and Humphries, S. E. (2009). Cytokine and cytokine receptor gene polymorphisms and their functionality. *Cytokine Growth Factor Rev.* **20**, 43–59.
- St Ledger, K., Agee, S. J., Kasaian, M. T., Forlow, S. B., Durn, B. L., Minyard, J., Lu, Q. A., Todd, J., Vesterqvist, O., and Burczynski, M. E. (2009). Analytical validation of a highly sensitive microparticle-based immunoassay for the quantitation of IL-13 in human serum using the Erenna immunoassay system. *J. Immunol. Methods* **350**, 161–170.
- Suntharalingam, G., Perry, M. R., Ward, S., Brett, S. J., Castello-Cortes, A., Brunner, M. D., and Panoskaltis, N. (2006). Cytokine storm in a phase 1 trial of the anti-CD28 monoclonal antibody TGN1412. *N. Engl. J. Med.* **355**, 1018–1028.
- Toedter, G., Hayden, K., Wagner, C., and Brodmerkel, C. (2008). Simultaneous detection of eight analytes in human serum by two commercially available platforms for multiplex cytokine analysis. *Clin. Vaccine Immunol.* **15**, 42–48.
- Warle, M. C., van der Laan, L. J., Kusters, J. G., Pot, R. G., Hop, W. C., Segeren, K. C., Ijzermans, J. N., Metselaar, H. J., and Tilanus, H. W. (2005). No association between tumor necrosis factor-alpha production and gene polymorphisms among inbred rat strains. *Transpl. Immunol.* **14**, 77–82.
- Watanabe, M., Uchida, K., Nakagaki, K., Kanazawa, H., Trapnell, B. C., Hoshino, Y., Kagamu, H., Yoshizawa, H., Keicho, N., Goto, H., *et al.* (2007). Anti-cytokine autoantibodies are ubiquitous in healthy individuals. *FEBS Lett.* **581**, 2017–2021.
- Weaver, J. L., Snyder, R., Knapton, A., Herman, E. H., Honchel, R., Miller, T., Espandiari, P., Smith, R., Gu, Y. Z., Goodsaid, F. M., *et al.* (2008). Biomarkers in peripheral blood associated with vascular injury in Sprague-Dawley rats treated with the phosphodiesterase IV inhibitors SCH 351591 or SCH 534385. *Toxicol. Pathol.* **36**, 840–849.
- Wong, H. L., Pfeiffer, R. M., Fears, T. R., Vermeulen, R., Ji, S., and Rabkin, C. S. (2008). Reproducibility and correlations of multiplex cytokine levels in asymptomatic persons. *Cancer Epidemiol. Biomarkers Prev.* **17**, 3450–3456.
- Woodcock, J., and Woosley, R. (2008). The FDA critical path initiative and its influence on new drug development. *Annu. Rev. Med.* **59**, 1–12.