Tuberculosis Biomarker and Surrogate Endpoint Research Roadmap

Payam Nahid¹, Jussi Saukkonen², William R. Mac Kenzie³, John L. Johnson⁴, Patrick P. J. Phillips⁵, Janet Andersen⁶, Erin Bliven-Sizemore³, John T. Belisle⁷, W. Henry Boom⁴, Annie Luetkemeyer⁸, Thomas B. Campbell⁹, Kathleen D. Eisenach¹⁰, Richard Hafner¹¹, Jeffrey L. Lennox¹², Mamodikoe Makhene¹³, Susan Swindells¹⁴, M. Elsa Villarino³, Marc Weiner¹⁵, Constance Benson¹⁶, and William Burman¹⁷

¹Division of Pulmonary and Critical Care, Curry International Tuberculosis Center, San Francisco General Hospital, University of California, San Francisco, San Francisco, California; ²Pulmonary Center, Boston University School of Medicine; Boston, Massachusetts; ³Centers for Disease Control and Prevention, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, Division of Tuberculosis Elimination, Clinical Research Branch, Atlanta, Georgia; ⁴Tuberculosis Research Unit, Department of Medicine, Division of Infectious Diseases, Case Western Reserve University, Cleveland, Ohio; ⁵British Medical Research Council Clinical Trials Unit, London, United Kingdom; ⁶Center for Biostatistics in AIDS Research, Harvard School of Public Health, Boston, Massachusetts; ⁷Mycobacteria Research Laboratories, Department of Microbiology, Immunology, and Pathology, Colorado State University, Fort Collins, Colorado; ⁸HIV/AIDS Division, San Francisco General Hospital, University of California, San Francisco, San Francisco, California; ⁹Division of Infectious Diseases, Department of Medicine, University of Colorado Denver, Denver, Colorado; ¹⁰Departments of Microbiology and Pathology, University of Arkansas for Medical Sciences, Little Rock, Arkansas; ¹¹Division of AIDS, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland; ¹²Department of Internal Medicine, Division of Infectious Disease, Emory University School of Medicine, Atlanta, Georgia; ¹³Division of Microbiology and Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland; ¹⁴Department of Internal Medicine, Division of Infectious Disease, University of Nebraska Medical Center, Omaha, Nebraska; ¹⁵Department of Medicine, Infectious Diseases, University of Texas Health Science Center at San Antonio, San Antonio, Texas; ¹⁶Division of Infectious Diseases, University of California, San Diego School of Medicine, San Diego, California; ¹⁷Denver Public Health, Denver, Colorado

The Centers for Disease Control and Prevention and National Institutes of Health convened a multidisciplinary meeting to discuss surrogate markers of treatment response in tuberculosis. The goals were to assess recent surrogate marker research and to provide specific recommendations for (1) the qualification and validation of biomarkers of treatment outcome; (2) the standardization of specimen and data collection for future clinical trials, including a minimum set of samples and collection time points; and (3) the creation of a specimen repository to support biomarker testing. This article summarizes these recommendations and provides a roadmap for their implementation.

In April 2010, the Centers for Disease Control and Prevention (CDC) and National Institutes of Health (NIH) brought together tuberculosis experts, physicians, bench and clinical research scientists, biostatisticians, and microbiologists to discuss the identification and evaluation of surrogate markers of tuberculosis treatment response that could accelerate the clinical testing of new tuberculosis drugs and regimens. The current recommended combination regimen for the treatment of active tuberculosis is more than 40 years old, requires a minimum of 6 months to complete, and is

often hampered by nonadherence and drug-related toxicity. There is renewed interest in shortening treatment duration and a number of new agents for tuberculosis treatment are under investigation in clinical trials (1). Combining these new agents with existing antituberculosis drugs offers the hope for regimens that may be better tolerated, shorter in duration, and with fewer drug-drug interactions as compared with existing regimens. Some of these agents are in novel classes (e.g., bedaquiline/TMC207, delamanid/OPC67683, PA824), making them useful for both rifampin-susceptible and -resistant tuberculosis. Others are in established classes (e.g., moxifloxacin, rifapentine), and their use may improve and shorten the treatment of drug-susceptible latent tuberculosis infection and active disease (2).

Establishing efficacy of a candidate drug within a regimen that includes other existing or new agents represents a significant challenge. Whereas bactericidal activity, safety, and tolerability of single agents can be assessed with relative simplicity, such determinations for multidrug regimens are difficult. Moreover, early bactericidal activity does not measure sterilizing activity of a candidate drug against bacilli that persist despite effective treatment (3). Nonetheless, preclinical studies provide an indication of improved efficacy and suggest which combination of drugs to pursue further in clinical trials necessary to establish sterilizing capability (4). One measure commonly used to assess efficacy in phase 2 clinical trials, the sputum culture status on solid media at 2 months after therapy initiation (i.e., 2-mo sputum culture conversion), however, has come under significant scrutiny (5–12). Despite its presumed potential in tuberculosis drug development, 2-month culture status is a problematic endpoint. There is mounting uncertainty about its ability, as a putative intermediate endpoint, to adequately assess, through phase 2 trials, appropriate drug combinations, doses, and dosing frequencies to be selected for evaluation in phase 3 trials. In the large series of studies by the British Medical Research Council, differences in 2-month culture status on solid media (primarily Lowenstein-Jensen media) correlated

(Received in original form May 10, 2011; accepted in final form June 20, 2011)

Supported partially by the National Institutes of Health through National Heart, Lung, and Blood Institute funding K23HL092629 (P.N.), National Institute of Allergy and Infectious Diseases funding AI068636 (A.L., T.B.C., J.L.L., S.S. and C.B.) and AI068634 (J.A.), and also through the Centers for Disease Control and Prevention, Division of Tuberculosis Elimination, Tuberculosis Trials Consortium.

Summary of Joint CDC/NIH Tuberculosis Biomarkers Workshop, April, 2010, Denver, Colorado

Correspondence and requests for reprints should be addressed to Payam Nahid, M.D., M.P.H., University of California, San Francisco, Pulmonary and Critical Care Medicine, San Francisco General Hospital, 1001 Potrero Avenue, 5K1, San Francisco, CA 94110. E-mail: pnahid@ucsf.edu

Am J Respir Crit Care Med Vol 184, pp 972–979, 2011

Originally Published in Press as DOI: 10.1164/rccm.201105-0827WS on July 7, 2011

Internet address: www.atsjournals.org

with the sterilizing activity of regimens (13). However, existing regimens achieve high culture conversion rates after 2 and 3 months of therapy, and as a dichotomous endpoint at a single time point, the use of 2-month culture status as the efficacy endpoint results in the need for relatively large sample sizes (75 to 250 patients per arm, depending on assumptions) for the evaluation of a new regimen. By comparison, in the field of HIV therapeutics the availability of a quantitative biomarker of treatment effect with a wide dynamic range (i.e., change in HIV RNA level) and detailed knowledge of the pharmacodynamics of antiretroviral drugs allow phase 2 studies of new agents to be completed with approximately 30 to 40 patients per arm (14). For the development of new antituberculosis drugs, and with the multiple permutations of drug combinations, doses, and dosing intervals needing evaluation in phase 2 trials, the use of 2-month culture conversion as the principle endpoint will significantly prolong the time to move these new drugs into clinical practice. Therefore, there is a pressing need to identify tuberculosis biomarkers that, if qualified, will improve efficiency of phase 2 tuberculosis drug testing.

In recognition of this need, numerous potential alternative biomarkers of treatment response have been proposed in the literature. However, none have been qualified or validated as surrogate endpoints of failure and recurrence. A major roadblock toward developing biomarkers into validated surrogate endpoints has been the lack of well-characterized repositories with biospecimens from patients who have had adequate follow-up to quantify recurrent disease. An optimal setting in which putative surrogates of treatment response could be effectively evaluated is within a clinical drug trial in which well-characterized patients are monitored in a rigorous and standardized fashion and receive verified dosing of drugs at defined time points at which biological specimens can be collected for banking. Phase 2 and 3 clinical trials inherently provide these critical components. If banking of biological specimens were integrated into such trials, a comprehensive and integrated biomarker discovery and validation program would be feasible. A recently established Consortium for TB Biomarkers (CTB2) among the Food and Drug Administration, the Global Alliance for TB Drug Development, National Institute of Allergy and Infectious Diseases/AIDS Clinical Trials Group, and CDC/Tuberculosis Trials Consortium represents a significant step forward in supporting such a program.

DEFINITIONS, QUALIFICATION, AND VALIDATION OF BIOMARKERS AND SURROGATES

The terms “biomarker” and “surrogate endpoint” are not synonymous. The U.S. National Institutes of Health Biomarkers Definitions Working Group define a biomarker as “a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention” (15). Biomarkers reflect the activity of a disease process and can be used to aid the development of new drugs, contributing information on safety, efficacy, drug interactions, or appropriate dosing in preclinical and early phases of drug development. A biomarker can also be used as a substitute for the clinical endpoint in a phase 2 or phase 3 trial, wherein it can be described as a surrogate endpoint. Surrogate endpoint and surrogate marker are used interchangeably, but the former is preferred as it emphasizes that this surrogate is a substitute endpoint rather than just a marker giving an indication of the treatment effect (16). Other terms, such as intermediate endpoint (17) or auxiliary endpoint (18), refer to endpoints that might augment, but not replace, the clinical endpoint information.

Temple defined a surrogate endpoint as “a laboratory measurement or a physical sign [a biomarker] used as a substitute

for a clinically meaningful endpoint that measures directly how a patient feels, functions or survives. Changes induced by a therapy on a surrogate endpoint are expected to reflect changes in a clinically meaningful endpoint” (19). This definition emphasizes the need to have a clearly defined final endpoint for which the surrogate will substitute; the surrogate is expected to predict the effect of the therapy and thus the final endpoint (19). It is only in this context that a biomarker can rightly be described as a surrogate endpoint. Fleming and De Mets defined a surrogate as needing to be “in the only causal pathway of the disease process, and the intervention’s entire effect on the true clinical outcome is mediated through its effect on the surrogate” (20). Prentice, in the first attempt to provide a clear statistical definition of a surrogate endpoint, argues that a surrogate endpoint should be considered in the context of the treatment comparison, as “a response variable for which a test of the null hypothesis of no relationship to the treatment groups under comparison is also a valid test of the corresponding null hypothesis based on the true endpoint” (21). This widely accepted definition is the most common starting point for a discussion on the evaluation of surrogate endpoints, but its stringency may limit its use as a benchmark. As noted by Prentice, it places “some strong restrictions on the relationship of the treatment to the surrogate and true out-

U.S. Food and Drug Administration definitions of validated versus qualified biomarkers (32, 33)

Valid Biomarker: A “test result may be considered a valid biomarker if (1) it is measured in an analytical test system with well-established performance characteristics, and (2) there is an established scientific framework or body of evidence that elucidates the physiologic, pharmacologic, toxicologic, or clinical significance of the test results.” A fully validated (vs. probable valid) “biomarker would meet criteria (1) and (2) above, and its association with a meaningful outcome would have been demonstrated in more than one experiment.”

Qualified Biomarker: “Qualification is a conclusion that within the stated context of use, the results of assessment with a (biomarker) can be relied upon to have a specific interpretation and application in drug development and regulatory decision-making.” That is, “analytically valid measurements of it can be relied upon to have a specific use and interpretable meaning in drug development...industry can use the (biomarker) for the qualified purpose during drug development, and Center for Drug Evaluation and Research reviewers can be confident in applying the (biomarker) for the qualified use”.

comes” and “results in limited potential to convincingly validate a surrogate outcome, especially for use in the evaluation of a range of future treatments” (22).

At a minimum, a biomarker of treatment response most useful for drug development would need to (1) correspond closely with treatment outcomes, (2) have a wide dynamic range that would allow analysis as a continuous variable, and (3) provide this information from a limited number of early time points. Ideally, it should be measurable in biological specimens that are simple and inexpensive to collect, process, store, and ship. Finally, the analytic technique used to measure the biomarkers should be adaptable for routine clinical use, as the biomarker may prove to be useful for diagnosis and clinical monitoring as

well as in clinical trials. Whereas Temple, Fleming, Prentice, and others have provided the field with operational criteria for validating a putative surrogate endpoint, we believe there is also significant value in identifying a qualified, but not validated, biomarker of treatment response, particularly given the significant challenges to convincingly validate a surrogate endpoint. A qualified biomarker could also be very useful in providing the basis for “go” and “no go” decisions early in the development of a candidate drug or regimen as part of phase 2 trials (*see* Text Box).

RECOMMENDATIONS FOR THE DEVELOPMENT OF A TUBERCULOSIS SPECIMEN AND DATA REPOSITORY

A primary objective of the workshop was to reach consensus on an approach to developing a tuberculosis specimen and data repository. The Working Group sought to establish a minimum set of specimens and collection time points that could be integrated into future clinical trials. Due to the low frequency of treatment failure and relapse with modern tuberculosis treatment regimens (23), multiple trials will be needed to capture enough failure and relapse events to permit qualification and subsequent validation of promising biomarkers. Harmonization and standardization of procedures for data and specimen collection would permit pooling of trial data and samples for future analyses, providing not only significantly increased power in evaluating surrogacy but also permitting inclusion of patients from diverse geographies and ethnicities in the qualification and validation steps.

Requirements for Clinical Trial Inclusion

The types and designs of trials for inclusion is an important component of designing a quality repository that will aid in the development of biomarkers of treatment response. First, the minimum requirement from a participating clinical trial is that subjects must have culture-confirmed *Mycobacterium tuberculosis* at enrollment. Patients with positive sputum smears are more likely to be culture positive and, by having higher numbers of tubercle bacilli in their sputum, increase opportunities to detect and correlate changes in bacillary load and candidate biomarkers. Although molecular techniques have been used to provide rapid detection of pathogen and presence of drug resistance, these techniques do not provide isolates for storage and future testing.

Pretreatment *M. tuberculosis* isolates should be stored for all subjects for later genotyping if needed to help distinguish between relapse from the patient's initial strain and exogenous reinfection (24). Treatment trials of extrapulmonary tuberculosis may be acceptable as long as there are rigorous and standardized approaches to obtaining isolates for culture, to documenting the extent of clinical disease at baseline and the clinical response to therapy, and to defining the outcomes. This principal requirement for culture-proven disease and well-documented treatment response should also apply to pediatric trials, wherein microbiologic endpoints can be difficult to obtain. Second, the repository should be built with samples collected from persons enrolled in randomized treatment trials. Randomization provides the optimal design for analysis of surrogacy and the evaluation of a putative marker's ability to distinguish treatment effect between regimens. Third, the use of directly observed therapy throughout treatment is preferred over self-administered therapy, as it provides verified dosing at the well-defined time points when specimens are collected. Fourth, the duration of follow-up after treatment completion is a critical component of any banking and surrogate marker development effort. Historically, the duration of follow-up after treatment completion was 24 months; recently it has been noted that 70% of recurrence occurs within the first 6 months of treatment completion and 90% within the first 12 months (25). For the purposes of a repository, the Working Group recommended a minimum of 12 months of follow-up after treatment completion with active evaluation for recurrence.

Given that biomarkers will be assessed against microbiologically confirmed treatment failure and/or relapse, large trials (>500 participants) or a combination of trials with harmonized data and specimen collection standards will be needed to detect adequate numbers of poor outcomes to permit nested case-control designs within which putative surrogates can be tested. This is because the existing short-course regimen for active, drug-susceptible tuberculosis achieves cure in more than 95% of cases (26). In this regard, trials in drug-resistant tuberculosis have been proposed as an attractive alternative from which to develop a bank, as participants from such trials have a higher likelihood of failure or relapse. However, there are several key drawbacks to using trials in multidrug-resistant (MDR) tuberculosis, including (1) the duration of treatment is long (up to 24 months or more); (2) likelihood of drug toxicity is higher with second-line

TABLE 1. SAMPLE SET OF INDIVIDUAL DEMOGRAPHIC, CLINICAL, RADIOGRAPHIC, AND LABORATORY CLINICAL TRIAL DATA REQUIRED FOR BIOMARKER DISCOVERY AND EVALUATION

Individual-Level Data	Time Point for Collection
Demographic data	Baseline
Age at time of enrollment	
Sex	
Place of birth	
Ethnicity	
Enrollment site (city, country)	
Enrollment date	
Clinical data	
HIV status	Baseline in all individuals
Antiretroviral use, CD4 count, viral load	Baseline, Week 8, end of treatment, and at failure and relapse in HIV-infected individuals
Microbiologic data	
Smear status with WHO-based grading	Baseline, Weeks 1 and 2, and every 2 weeks subsequently
Culture status on solid and liquid media	
Drug susceptibility tests	Baseline, failure and relapse
Radiographic data	
Cavitation, yes/no	Baseline, Week 8, end of treatment, and at failure and relapse
Bilateral disease, yes/no	
Number of lobes involved	

Definition of abbreviation: WHO = World Health Organization.

Figure 1. Sources for potential host and pathogen biomarkers of treatment response.

drugs, resulting in increased likelihood of drop-outs; (3) patients with MDR tuberculosis may have a history of noncompliance and may be difficult to recruit, enroll, and follow; and (4) patients with MDR tuberculosis often represent a very select group of survivors who have lived without adequate treatment for long periods of time. An alternate to using MDR tuberculosis trials for the development of a biospecimen collection would be to select study subpopulations with drug-susceptible disease but at higher risk of treatment failure and relapse. There is substantial information regarding which patients are at greatest risk of poor outcomes at baseline ($>3+$ smear, cavitory disease, etc.). In contrast, persons with low-grade acid-fast bacillus smears and limited infiltrates (i.e., those with lower disease burden) are less likely to have poor outcomes. Overall, the principle of enhancing the study population for potential poor outcomes is a viable option, although the generalizability of results obtained must also be considered (see Text Box).

Data Standards

Integral parts of a biological specimen bank designed for the evaluation of putative surrogate markers are the clinical, radiographic,

and laboratory data accompanying the samples (Table 1). It is strongly recommended that data standards be adopted by tuberculosis clinical trial groups to allow pooling of clinical data and specimens across clinical trial evaluations of putative surrogate markers. The international Clinical Data Interchange Standards Consortium or CDISC (<http://www.cdisc.org/>) has established international standards to support the acquisition, exchange, submission, and archiving of clinical research data and metadata; these standards should be adopted whenever possible.

Types of Specimens and Time Points of Collection

The sample types and the time points at which specimens are to be collected during a treatment trial are a source of debate. Figure 1 shows a schematic of possible sources for host and pathogen biomarkers of treatment response for tuberculosis. Table 2 highlights the advantages and challenges of working with these sources. Table 3 provides a list of potential biomarkers of tuberculosis treatment effect stratified into three categories: markers that quantitate *M. tuberculosis*, markers of inflammation, and markers of specific immune reactivity to *M. tuberculosis*. Depending on the biomarker of interest, there may be substantial variation in the dynamic range and kinetic change in response to treatment, and consequently optimal timing of collection may be unknown. For example, without knowing the assay, the platform for the assay and the biospecimen kinetics of response to treatment for any given biomarker *a priori*, it is difficult to know whether sampling a specimen at baseline and Week 4 are the optimal time points for assessing a given biomarker's surrogacy. Consequently, sampling is generally integrated into time points at which there are already routine clinical evaluations as part of the parent trial. *M. tuberculosis* isolates, sputa, sera, plasma, urine, peripheral blood mononuclear cells (PBMC), and host DNA are the most commonly referenced specimen types. Each has a set of advantages and challenges. The Working Group reviewed these issues and recommended a set of biological samples divided into two tiers, a "minimum" set and an "optional" set (Tables 4 and 5). The recommended specimen types and associated time points for collection were suggested as a minimum and could be expanded as needed by clinical trial networks for individual trials. The proposed tiered approach took into account four key

TABLE 2. ADVANTAGES AND CHALLENGES BY SOURCE OF BIOMARKER

Biomarker Source	Advantages	Challenges
Gene expression	Large data sets Targeting of selected genes	Limited for pathogen* Technical expertise Resource restricted†
Macromolecules	Large data sets Targeting of selected products Measure by "simple" technologies	Limited for pathogen* Host variability Rate of change
Metabolites	Large data sets Targeting of selected products Rapid change Clinical sample	Methods of detection Host variability
Cell function/viability	Culture is gold standard for pathogen Provides information on host biology	Clinical sample Resource restricted for host† Storage requirements Assuring viability
Cell/tissue architecture	Strong measure of host	Feasibility for pathogen May require invasive procedure Resource restricted†

The advantages and challenges are for the application of specific types of biomarker. It is recognized that the discovery of biomarkers of each type has separate advantages and challenges.

*This limitation relates to the pathogen load in a clinical sample and the ability to detect small numbers of pathogen products among a large number of host products.

†Resource limitations include the need for specific technical expertise and/or specialized equipment or computer software.

TABLE 3. EXAMPLES OF POTENTIAL TUBERCULOSIS BIOMARKERS OF TREATMENT EFFECT

Type of Biomarker of Treatment Effect	Suggested Biomarkers
Quantitation of <i>Mycobacterium tuberculosis</i>	Sputum smear Quantitative sputum culture on solid media (single specimen vs. serial sputum colony counting) Time to detection of growth in an automated broth culture system from sputum specimens Quantitative nucleic acid amplification of <i>M. tuberculosis</i> in sputum or blood Concentration of a tuberculosis antigen in blood, sputum, or urine
Markers of inflammation/immune activation	Erythrocyte sedimentation rate (ESR) Soluble IL-2 receptor- α (sIL-2R α) Soluble tumor necrosis factor- α receptors 1 and 2 (sTNF-R1 and sTNF-R2) Soluble intercellular adhesion molecule type 1 (sICAM-1) C-reactive protein (CRP) Soluble urokinase receptor (suPAR) Neopterin Granzyme B IFN- γ -inducible protein 10 (IP-10/CXCL10) Pentraxin-3, 25(OH)D LL-37 (cathelicidin) Osteopontin
Markers of specific immune memory and activity against <i>M. tuberculosis</i>	IFN- γ release assays Quantity and quality of activated T cell types (CD4 or CD8) specific for <i>M. tuberculosis</i> antigens B cells and antibodies
Metabolomic/proteomic approaches to biomarker identification	Cast a broad net to evaluate the association between changes in concentrations of small molecules in blood or urine and tuberculosis treatment effect

characteristics in selecting which biospecimens to store, including (1) plausibility of identifying pathogen-related markers within a specimen type; (2) degree of complexity involved in handling the specimen type, including site-level expertise required to collect, process, and store requisite biological specimens; (3) logistical aspects of collecting a given sample type and building a cross-trial specimen repository from multiple international clinical trial sites; and (4) published evidence in support of candidate

biomarkers and their associated requisite biospecimen type. For tier 1, the group recommended sputum, serum or plasma, and urine samples, in addition to the collection and storage of *M. tuberculosis* isolates. For tier 2, the group recommended collection of PBMC, host peripheral mRNA, stimulated plasma, and host DNA, to be collected at select sites that have the capacity and expertise in collecting, processing, and storing such biospecimens. Despite nominally increased complexity in handling some of these

TABLE 4. PROPOSED TIER 1 OR MINIMUM SET OF HOST BIOLOGICAL SPECIMEN TYPES, COLLECTION TIME POINTS IN CLINICAL TRIALS OF ACTIVE TUBERCULOSIS

Tier 1 Minimum Set Specimen Types Collected at All Enrollment Sites				
Host Specimen Type	Time Points Relative to Initiation of Drug Treatment*	Potential Commercial Collection Systems	Suggested Volume	Examples of Common Uses
Sputum for banking	Baseline Week 2 Week 4 Week 8 Plus at failure and recurrence	BD Falcon Sputum Collection System (BD Diagnostics)	Total volume available	mRNA, microbiologic measures, host immune markers
Serum or plasma [†]	Baseline Week 4 Week 8 End of treatment Plus at failure and recurrence with time-matched nonrelapsed control subjects	BD Vacutainer SST Tube for serum (BD Diagnostics) BD EDTA Vacutainer Tube or BD Vacutainer PST for plasma (BD Diagnostics)	6–10 ml (two 8-ml SSTs) per time point	Proteomics, metabolomics, lipidomics, noncellular measures of immune response (i.e. cytokines, chemokines, and related reactants)
Urine	Baseline Week 4 Week 8 End of treatment Plus at failure and recurrence with time-matched nonrelapsed control subjects	BD Vacutainer Urine Collection Kits (BD Diagnostics)	10 ml or more	Metabolomics, measures of microbial markers

*Baseline or time zero samples are defined as samples collected preferably before any administration of tuberculosis medications. If medications are administered before blood draw, this information should be collected and reported. Time points are provided as weeks, but should also be linked with number of treatment doses administered.

[†] The type should be noted.

TABLE 5. PROPOSED TIER 2 OR OPTIONAL SET OF HOST BIOLOGICAL SPECIMEN TYPES, COLLECTION TIME POINTS IN CLINICAL TRIALS OF ACTIVE TUBERCULOSIS

Tier 2 Optional Set Specimen Types Collected at Sites with Requisite Expertise and Facilities				
Host Specimen Type	Time Points Relative to Initiation of Drug Treatment*	Potential Commercial Collection Systems	Suggested Volume	Examples of Common Uses
Peripheral blood mononuclear cells	Baseline Week 8 End of treatment Plus at failure and recurrence with time-matched nonrelapsed control subjects	Cell Preparation Tubes (CPT, BD Diagnostics)	10 ml (2 CPTs) per time point	Immunologic measures of treatment response
Supernatant from whole blood stimulated with mycobacterial antigens	Baseline Week 4 Week 8 End of treatment Plus at failure and recurrence with time-matched nonrelapsed control subjects	Quantiferon-Gold InTube (Cellestis, Inc).	3 ml (per manufacturer guidelines) per time point	Noncellular measures of immune response (i.e. cytokines, chemokines and related reactants)
Whole blood transcriptome	Baseline Week 4 Week 8 End of treatment Plus at failure and recurrence with time-matched nonrelapsed control subjects	PAXgene Blood RNA System (Qiagen Inc); Tempus Blood RNA Tubes, (Applied Biosystems)	5 ml (two 2.5-ml PAXgene tubes) per time point	Transcriptomics
Host DNA	Baseline	Oragene DNA kits (DNA Genotek)	2 ml of expectorated saliva	Pharmacogenomics, genotype–phenotype relationships

* Baseline or “time zero” samples are defined as samples collected preferably before any administration of tuberculosis medications. If medications are administered before blood draw, this information should be collected and reported. Time points are provided as weeks, but should also be linked with number of treatment doses administered.

biospecimens, we believe they warrant collection at select sites as part of an early look into the genomics, transcriptomics, proteomics, and host–pathogen immunology of treatment response (12). Advanced imaging techniques using simultaneous positron emission and computerized axial tomography scanning with *M. tuberculosis*–specific positron emission tomography probes are also being investigated as possible tools to monitor response to treatment. These sophisticated techniques, along with other specialized specimen types, including gastric lavage and dried blood spots, were believed not to be optimized adequately to be included at this time in a recommended set of sample types and data sets, but could be integrated in the future.

Last, use of harmonized and standardized procedures in the collection, processing, shipping, and storing of all biological samples was strongly recommended. Seemingly minor differences in processing or handling of a specimen can have significant effects in analytical reliability and reproducibility of biomarker measures (27). In addition, quality assurance methods that include testing the viability of immune cells and the presence of certain biomarkers during the development of the repository is also critical, so as to assure samples are being collected and stored appropriately.

Tier 1, or Minimum Set Sample Types

Sputum for culture and banking. Sputum microscopy and culture are key components of all tuberculosis treatment trials. Both enrollment eligibility and trial endpoints are determined by these microbiologic measures. Single site studies or consortia that opt to use a central laboratory for processing of sputum for smear and culture may not have as many challenges in harmonizing procedures to minimize variability that might alter microbiologic

measures (28). Using similar techniques in processing sputum for biobanking is critical, particularly when one attempts to link multiple consortia or consortia with geographically dispersed enrollment sites. Despite the numerous complexities involved in collecting, processing, and storing sputum, including issues related to preparing the patient, collecting the sample (timing of collection, expectorated vs. induced, single sample vs. pooled, etc.), and processing the sample (splitting for culture and banking vs. obtaining separate sputa for banking, effect of additives and decontamination procedures, etc.), sputum is considered a part of the minimum set because it is the sample with the highest likelihood of providing a pathogen-based biomarker. Given the importance of sputum for the clinical trial microbiologic endpoints, it is recommended that a separate sample be collected for banking, in place of splitting samples.

Serum and plasma. Compared with sputum, serum and plasma have lower complexity in terms of collection, processing, and storage. Proteomic, lipidomic, metabolomic, and other unbiased and targeted multiplexed array approaches are well suited to these specimen types (29). Serum requires a low to moderate degree of on-site processing when commercially available serum separator tubes are used. Serum separating tubes are preferred over *in vitro* blood clotting methods as the latter have a higher likelihood of introducing *ex vivo* changes. As an alternative to serum, plasma is considered more stable over time. Most “-omic” approaches are applicable to both specimen types, although there is greater familiarity with using plasma. Whether the use of anti-coagulants in obtaining plasma and other differences between serum and plasma are important in terms of biomarker research in tuberculosis is unknown and requires further research. For most assays, plasma and serum can be used interchangeably.

Urine. The relative ease of urine collection makes it a convenient specimen type from which biomarkers of treatment response could readily be measured. Mycobacterial lipoarabinomannan and urine mycobacterial DNA have been studied extensively as diagnostic tests for tuberculosis (30, 31). Proteomic, lipidomic, metabolomic and other multiplexed array approaches have also been evaluated using urine. The optimal timing of collection during the day, the volume collected, and the optimal time points for collection remain unknown.

Tier 2, or Optional Set Sample Types

Host DNA. Common sources of DNA include peripheral blood and oral specimens. Although DNA yields from oral specimens are generally less than from whole blood, oral specimens may be preferred over venipuncture because oral specimens can be obtained away from the clinic. Host DNA provides researchers the opportunity to study polymorphisms, such as in the cytochrome P isoenzyme system, which could influence pharmacokinetics of drugs and consequently treatment response. In addition, polymorphisms that might identify populations at risk for hepatotoxicity in whom frequent monitoring during treatment may be indicated to prevent interruption in therapy would also be valuable, although not critical to the study of biomarkers of treatment response.

Host RNA and microRNA. There exist commercial systems for the collection and stabilization of RNA that are easy to use. These collection systems provide a standardized method for the stabilization and isolation of RNA from whole blood for later transcriptomic studies. Changes in the whole blood transcriptome might be a tool for monitoring response to antituberculosis therapy. The transcriptome along with host DNA can provide critical information on pharmacogenomics and pharmacokinetics as well as risk for hepatotoxicity during treatment. The relative simplicity and high-value yield of transcriptomic signals makes this platform attractive for biospecimen banking in tuberculosis.

PBMCs. PBMCs are a challenging specimen type in terms of collection, processing, and storage. Nonetheless, PBMCs provide a rich resource for biomarker development and should be included as an optional set of samples, wherever feasible. Study of measures of immunity through PBMC samples may afford advantages over study of expectorated specimens, as tuberculosis treatment can reduce a patient's cough significantly by the end of the intensive phase. In addition, by the end of treatment, it is unlikely that pathogen-related factors will be measurable in any of the other specimen types proposed, but immunologic measures may provide a measure of treatment effect by regimen, particularly given that immune clearance plays a critical role in achieving nonrelapsing cure. As with other specimen types, the optimal time points for collection of PBMCs are unknown, but collection at baseline (to ascertain immunologic status in relation to severity of disease presentation), at a relatively early time point (e.g., at 2 months, or the end of the intensive phase), and at the end of treatment (when there may be subtle immune indicators of persisting infection) seems reasonable. Three key issues, however, make the inclusion of PBMCs in a repository program challenging: (1) processing requirements to isolate PBMCs, (2) the volume of blood generally required for immunologic studies, and (3) the requirement for liquid nitrogen for long-term storage of cells. Consequently, not all clinical trial sites will be able to participate in this activity.

Stimulated plasma. Food and Drug Administration–approved IFN- γ release assays provide a commercially available and relatively standardized approach to collecting and storing plasma for biomarker discovery using multiplexed assays. Such assays provide a medium in which a variety of released cytokines and chemokines (in addition to IFN- γ) can be measured after whole

blood stimulation with antigenic peptides (early secreted antigen target-6, culture filtrate protein-10, and TB7.7) that are substantially more specific for *M. tuberculosis* than PPD and encoded by genes located within the region of difference 1 (RD1) segment of the *M. tuberculosis* genome.

SUMMARY

Establishing efficacy of a new drug candidate agent for treatment of tuberculosis when used in combination with other novel compounds or with existing drugs is a significant challenge. In this workshop we describe a roadmap toward discovery, qualification, and validation of biomarkers of treatment response that would improve the efficiency of phase 2 and 3 clinical trials of new regimens for active tuberculosis. A major roadblock in developing biomarkers into validated surrogate endpoints of treatment response has been the lack of well-characterized repositories with biospecimens obtained from patients who have had adequate follow-up for failure and relapse. We provide recommendations on a minimum set of data and biological specimens to be collected in phase 2 and 3 clinical trials to permit a comprehensive and integrated biomarker discovery and validation program in the future.

Author Disclosure: P.N. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript. J.S. has received travel support from Denver Health. W.R.M. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript. J.L.J. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript. P.P.J.P. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript. J.A. has received consultancy fees from PATH and Tibotec. E.B.S. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript. J.T.B. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript. W.H.B. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript. A.L.'s institution has received grants from Gilead Sciences, Pfizer, Merck, and Tobira. T.C.'s institution has received grants from Pfizer, Tibotec, Merck, Gilead, Bristol Myers Squibb, GlaxoSmithKline (GSK), Boehringer Ingelheim, and Wyeth. K.D.E. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript. R.H. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript. J.L.L. has received consultancy fees from Merck; his institution has received grants from Merck, Viiv Pharmaceuticals, Gilead, Pfizer, and Tibotec. M.M. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript. S.S. has received consultancy fees from Gilead and Abbott Diagnostics; her institution has received grants from Pfizer, GSK, and Bristol Myers Squibb. M.E.V. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript. M.W. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript. C.B. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript. W.B.'s institution has received fees for participating in review activities from Tibotec.

References

1. Ma Z, Lienhardt C, McIlleron H, Nunn AJ, Wang X. Global tuberculosis drug development pipeline: the need and the reality. *Lancet* 2010;375:2100–2109.
2. Spigelman M, Gillespie S. Tuberculosis drug development pipeline: progress and hope. *Lancet* 2006;367:945–947.
3. Jindani A, Dore CJ, Mitchison DA. Bactericidal and sterilizing activities of antituberculosis drugs during the first 14 days. *Am J Respir Crit Care Med* 2003;167:1348–1354.
4. Nuermberger E. Using animal models to develop new treatments for tuberculosis. *Semin Respir Crit Care Med* 2008;29:542–551.
5. Davies GR, Phillips PP, Nunn AJ. Biomarkers and surrogate end points in clinical trials of tuberculosis treatment. *J Infect Dis* 2007;196:648–649, author reply 649–650.
6. Perrin FM, Lipman MC, McHugh TD, Gillespie SH. Biomarkers of treatment response in clinical trials of novel antituberculosis agents. *Lancet Infect Dis* 2007;7:481–490.
7. Wallis RS, Pai M, Menzies D, Doherty TM, Walzl G, Perkins MD, Zumla A. Biomarkers and diagnostics for tuberculosis: progress, needs, and translation into practice. *Lancet* 2010;375:1920–1937.

8. Walzl G, Ronacher K, Djoba Siawaya JF, Dockrell HM. Biomarkers for TB treatment response: challenges and future strategies. *J Infect* 2008; 57:103–109.
9. Davies GR. Early clinical development of anti-tuberculosis drugs: science, statistics and sterilizing activity. *Tuberculosis (Edinb)* 2010;90: 171–176.
10. Walzl G, Ronacher K, Hanekom W, Scriba TJ, Zumla A. Immunological biomarkers of tuberculosis. *Nat Rev Immunol* 2011;11:343–354.
11. Doherty TM, Wallis RS, Zumla A. Biomarkers of disease activity, cure, and relapse in tuberculosis. *Clin Chest Med* 2009;30:783–796. (x.).
12. Wallis RS, Doherty TM, Onyebujoh P, Vahedi M, Laang H, Olesen O, Parida S, Zumla A. Biomarkers for tuberculosis disease activity, cure, and relapse. *Lancet Infect Dis* 2009;9:162–172.
13. Mitchison DA. Assessment of new sterilizing drugs for treating pulmonary tuberculosis by culture at 2 months. *Am Rev Respir Dis* 1993;147:1062–1063.
14. Markowitz M, Nguyen BY, Gotuzzo E, Mendo F, Ratanasuwan W, Kovacs C, Prada G, Morales-Ramirez JO, Crumpacker CS, Isaacs RD, et al. Rapid and durable antiretroviral effect of the HIV-1 integrase inhibitor raltegravir as part of combination therapy in treatment-naïve patients with HIV-1 infection: results of a 48-week controlled study. *J Acquir Immune Defic Syndr* 2007;46:125–133.
15. De Gruttola VG, Clax P, DeMets DL, Downing GJ, Ellenberg SS, Friedland L, Gail MH, Prentice R, Wittes J, Zeger SL. Considerations in the evaluation of surrogate endpoints in clinical trials. Summary of a National Institutes of Health workshop. *Control Clin Trials* 2001;22: 485–502.
16. Molenberghs G, Buyse ME, Burzykowski T. The evaluation of surrogate endpoints. New York: Springer; 2005.
17. Freedman LS, Graubard BI, Schatzkin A. Statistical validation of intermediate endpoints for chronic diseases. *Stat Med* 1992;11:167–178.
18. Fleming TR, Prentice RL, Pepe MS, Glidden D. Surrogate and auxiliary endpoints in clinical trials, with potential applications in cancer and aids research. *Stat Med* 1994;13:955–968.
19. Temple R. Are surrogate markers adequate to assess cardiovascular disease drugs? *JAMA* 1999;282:790–795.
20. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Ann Intern Med* 1996;125:605–613.
21. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med* 1989;8:431–440.
22. Prentice RL. Surrogate and mediating endpoints: current status and future directions. *J Natl Cancer Inst* 2009;101:216–217.
23. Blumberg HM, Burman WJ, Chaisson RE, Daley CL, Etkind SC, Friedman LN, Fujiwara P, Grzemska M, Hopewell PC, Iseman MD, et al. American Thoracic Society/Centers for Disease Control and Prevention/Infectious Diseases Society of America: treatment of tuberculosis. *Am J Respir Crit Care Med* 2003;167:603–662.
24. Jasmer RM, Bozeman L, Schwartzman K, Cave MD, Saukkonen JJ, Metchock B, Khan A, Burman WJ. Recurrent tuberculosis in the United States and Canada: relapse or reinfection? *Am J Respir Crit Care Med* 2004;170:1360–1366.
25. Nunn AJ, Phillips PP, Mitchison DA. Timing of relapse in short-course chemotherapy trials for tuberculosis. *Int J Tuberc Lung Dis* 2010;14: 241–242.
26. Fox W, Ellard GA, Mitchison DA. Studies on the treatment of tuberculosis undertaken by the British Medical Research Council tuberculosis units, 1946–1986, with relevant subsequent publications. *Int J Tuberc Lung Dis* 1999;3:S231–279.
27. Brunner N, Schrohl AS, Wurtz S, Kohn E, Banks RE, Nielsen HJ, Sweep FCGJ. Banking of biological fluids for studies of disease-associated protein biomarkers. *Mol Cell Proteomics* 2008;7:2061–2066.
28. Mac Kenzie WR, Heilig CM, Bozeman L, Johnson JL, Muzanye G, Dunbar D, Jost KC Jr, Diem L, Metchock B, Eisenach K, et al. Geographic differences in time to culture conversion in liquid media: tuberculosis trials consortium study 28. Culture conversion is delayed in Africa. *PLoS ONE* 2011;6:e18358.
29. Walzl G, Ronacher K, Hanekom W, Scriba TJ, Zumla A. Immunological biomarkers of tuberculosis. *Nat Rev Immunol* 2011;11:343–354.
30. Lawn SD, Edwards DJ, Kranzer K, Vogt M, Bekker LG, Wood R. Urine lipoarabinomannan assay for tuberculosis screening before anti-retroviral therapy diagnostic yield and association with immune reconstitution disease. *AIDS* 2009;23:1875–1880.
31. Green C, Huggett JF, Talbot E, Mwaba P, Reither K, Zumla AI. Rapid diagnosis of tuberculosis through the detection of mycobacterial DNA in urine by nucleic acid amplification methods. *Lancet Infect Dis* 2009;9:505–511.
32. The Food and Drug Administration. Guidance for industry: pharmacogenomic data submissions. Rockville, Maryland: Food and Drug Administration; 2005.
33. The Food and Drug Administration. Draft guidance for industry, qualification process for drug development tools. Rockville, Maryland: Food and Drug Administration; 2010.