Transformative tools for tackling tuberculosis

Jennifer L. Gardiner and Christopher L. Karp

Discovery and Translational Sciences, Global Health, Bill & Melinda Gates Foundation, Seattle, WA 98102

The world is in need of more effective approaches to controlling tuberculosis. The development of improved control strategies has been hampered by deficiencies in the tools available for detecting Mycobacterium tuberculosis and defining the dynamic consequences of the interaction of M. tuberculosis with its human host. Key needs include a highly sensitive, specific nonspu- tum diagnostic; biomarkers predictive of responses to therapy; correlates of risk for disease development; and host response–independent markers of M. tuberculosis infection. Tools able to sensitively detect and quantify total body M. tuberculosis burden might well be transformative across many needed use cases. Here, we review the current state of the field, paying particular attention to needed changes in experimental paradigms that would facilitate the discovery, validation, and development of such tools.

Mycobacterium tuberculosis has been performing its deadly dance with Homo sapiens since before the duo emerged together from Africa (Comas et al., 2013). Tuberculosis (TB) continues to cause an outsized burden of morbidity and mortality. Currently, one third of the global population is thought to harbor latent infection, on the basis of sustained lymphocyte reactivity with M. tuberculosis proteins in the absence of symptoms. Those with latent infection have an ~10% lifetime risk of developing active disease. TB claims ~1.4 million lives annually, and often leaves disability behind in survivors (Murray et al., 2014; WHO, 2014a). These global numbers mask the profoundly inequitable distribution of this scourge of poverty, the burden of which falls heaviest on low- and middle-income countries (WHO, 2014a).

Approaches to decreasing TB morbidity and mortality, along with M. tuberculosis transmission, can be conceived as falling into three general categories, the first of which is treatment of disease. Effective treatment is central to any strategy for control. In addition to preventing morbidity and mortality, timely effective diagnosis and treatment of TB are critically important for decreasing transmission because pulmonary disease and transmission are tightly linked (Russell, 2011). Important barriers to effective treatment include the lack of availability of diagnostics that afford timely, effective ascertainment of TB, accurate tools to monitor responses to treatment, and the need for dramatically shorter courses of chemotherapy, which could potentially increase patient compliance and decrease the emergence of drug resistance.

The second approach is the prevention of disease. As preventing disease prevents morbidity, mortality, and transmission, this approach has obvious pride of place in terms of potential. It provides the underlying rationale for vaccination, as well as for secondary prophylaxis strategies. Unfortunately, the current barriers here are just as obvious. BCG, brought into the world some 90 yr ago, remains the only licensed vaccine for TB. Although efficacy is high against disseminated disease in infants (Trunz et al., 2006), a benefit that may well be provided by M. tuberculosis infection itself in older individuals (Karp et al., 2015), BCG’s efficacy in preventing pulmonary TB is highly variable, ranging from negative numbers to 80%, with the lowest efficacy seen in areas of high transmission and endemicity (Colditz et al., 1994; Dye, 2013; Mangtani et al., 2014). Further, whereas the efficacy of secondary prophylaxis with the antimycobacterial drug isoniazid (focused on those at highest clinical risk, including those in the first year after infection, and those at life stages and/or with intermittent conditions that predispose to disease progression) has been demonstrable in low transmission settings such as the global North, it has not been demonstrated in high transmission settings such as South Africa (Churchyard et al., 2014; Wood and Bekker, 2014).
A final approach to limiting TB morbidity and mortality is the prevention of infection. Raising the bar to the establishment of sustained infection—a term that will be used throughout this Perspective to refer to the sustained presence of viable \textit{M. tuberculosis} bacilli, with or without ongoing replication and subclinical or clinical disease—has promise for preventing both disease and transmission. Successful infection likely involves a single airborne droplet nucleus capable of reaching the alveolus, containing a single—or at most a few—organisms (Mills et al., 1960; Riley et al., 1995). Abortive infection, defined operationally by evidence of robust T cell reactivity, followed by the reversion of such, is seen in longitudinal studies of humans (Shah et al., 2011) and guinea pigs infected by exposure to air vented from the rooms of humans with active TB disease (Dharmadhikari et al., 2011). Prevention of infection through better ventilation, and prevention of sustained infection through better nutrition, may well have played a role in decreasing \textit{M. tuberculosis} transmission in urban North America in the early 20th century. BCG vaccination also appears to have some efficacy in preventing sustained infection in both adults and children (Soyyal et al., 2005; Eisenhut et al., 2009; Barreto et al., 2011; Basu Roy et al., 2012; Chan et al., 2013; Roy et al., 2014). For this and other reasons (Karp et al., 2015), developing vaccines aimed at preventing sustained infection represents a promising approach. That said, such vaccines are not yet known, and practical ways of testing more general methods of raising the barrier to infection (e.g., improved ventilation and improved nutrition), and of translating such interventions to practice, remain to be devised.

In sum, current approaches for TB control remain insufficient. Further, as with the burden of TB itself, existing approaches also fail any reasonable equity agenda—whereas considerable success has been attained in much of the industrial North, these approaches have not led to control in endemic areas with high transmission pressures.

The development of novel approaches to TB control has been hampered by deficiencies in the available tools for detecting and quantifying the presence of \textit{M. tuberculosis} in, and the dynamic consequences of the interactions of \textit{M. tuberculosis} with, individual human hosts. Diagnosis of disease largely relies on direct detection of \textit{M. tuberculosis} (via microscopy and/or culture) or \textit{M. tuberculosis} DNA (via amplification techniques) in sputum. Detection of infection relies on evidence of sustained T cell reactivity with \textit{M. tuberculosis} antigens (via tuberculin skin tests [TSTs] or IFN-\gamma release assays [IGRAs]) in peripheral blood. These tools provide neither an optimal foundation for TB control nor a clear path to the development of improved strategies. Fundamentally, the available tools fall well short of what is needed to reflect the underlying pathogenic complexities. These complexities deserve underscoring. Clinically silent, sustained infection actually represents a broad spectrum of responses, ranging from something close to sterilizing immunity to subclinical active disease. In turn, latency is itself part of a broader spectrum that encompasses active and fulminating disease (Barry et al., 2009). The overall spectrum is highly dynamic. Functional imaging and histological analysis of infected humans and nonhuman primates has shown that individual foci of infection can pursue diametrically opposite trajectories at the same time in the same lung (Russell et al., 2010; Lin et al., 2013, 2014). The response to therapeutic intervention appears to be just as complicated and dynamic (Russell et al., 2010; Lin et al., 2013). Improved tools that could provide actionable insight into these complexities have transformative potential.

**TB diagnosis**

Current diagnostics for pulmonary TB rely on sputum samples and suffer several shortcomings. To begin, detection of \textit{M. tuberculosis} in airway secretions depends on the presence of necrotic foci of infection in contiguity with the airway. As a result, by the time diagnosis in sputum becomes possible, most patients will have had active disease for quite some time, often with considerable lung damage. The practical necessity for \textit{M. tuberculosis} to be present in airway secretions underscores that diagnosis through sputum, perforce, only occurs after individuals are likely capable of transmitting \textit{M. tuberculosis} to others.

Several alternative methods for diagnostic testing of sputum exist, but they all have shortcomings. A molecular test, the Xpert MTB/RIF test (Cepheid), is sensitive (98% in microscopy positive subjects; 72% in microscopy negative subjects) and specific (99%; Boehme et al., 2010), but requires instrumentation not broadly available in endemic settings. Microscopy is broadly available and highly specific, but lacks sensitivity, missing the diagnosis in over one third of patients seeking care (Davis et al., 2013). Mycobacterial culture is the gold standard for TB diagnosis, but it provides results after considerable delay. In some endemic settings, clinicians are disinclined to obtain sputum in the first place (Satyanarayana et al., 2015), and in these settings diagnosis and misdiagnosis proceeds from history, physical exam, and chest x-ray alone.

Other challenges for TB diagnostics include HIV infection, which markedly lowers the sensitivity of sputum diagnostic tests (Lawn and Wood, 2011), and age. In infants, sputum is sparse and/or difficult to obtain, and analysis of repetitive early morning gastric aspirates or induced sputa has sensitivities that remain problematic (Starke, 2003; Zar et al., 2005). Obviously, sputum diagnostics have no utility in extrapulmonary disease, the diagnosis of which relies on samples (tissue or cerebrospinal fluid) collected invasively.

For all of these reasons, there is need for the development of a highly sensitive and specific diagnostic test for TB that is able to rapidly identify the presence of active disease in endemic settings and is performable at reasonable cost on an easily obtainable nonsputum sample such as blood, urine, or breath. Such a diagnostic might be based on the quantitative detection of \textit{M. tuberculosis} or its products. In this vein, it remains an open question whether mycobacterial burden has a direct correlation with an individual's location along the spectrum from latent infection to active disease to effectively treated—a question that will remain open until accurate methods for quantifying mycobacterial burden are developed (vide infra).
for whom confirmatory testing should follow (WHO, 2014b). The nonsputum diagnostic TPP is presented (with permission) in Table 1, with the purpose of providing a useful exemplar for those currently involved in biomarker research. The published TPPs outline the criteria a biomarker would likely need to meet, along with those of associated diagnostic testing platforms, to result in substantial impact. Progress toward meeting these TPPs is in brief summarized below. The TPP criterion for fast turnaround (<1 h) necessitates a focus on biomarkers that are present and can be directly measured in patient specimens; as such, cell stimulation and other culture-based methods are not discussed here.

In terms of the direct detection of \textit{M. tuberculosis} or its products, detection of \textit{M. tuberculosis} DNA in blood and urine from patients with pulmonary TB has better sensitivity than culture of these specimens, but remains well below required thresholds (Cannas et al., 2008; da Cruz et al., 2011; Theron et al., 2014). Among proteins and cell wall components, lipoarabinomannan (LAM) has received the most experimental attention. A commercially available urine LAM diagnostic test is available; however, poor sensitivity has led to limited use (Minion et al., 2011). There may be potential for improved sensitivity with other LAM assays (Mukundan et al., 2012; Chan et al., 2015; Hamasur et al., 2015). In HIV-positive patients, both \textit{M. tuberculosis} DNA and LAM detection in urine hold promise due to the relatively high yield of identifying TB patients missed by sputum diagnostics, ease of collection, and nonspecific symptoms resulting from immune suppression (Aceti et al., 1999; Nakiyingi et al., 2014; Lawn et al., 2013).
with specificity may of course be caused by the presence of these same products in individuals with latent infection with *M. tuberculosis* and/or nontuberculous mycobacteria. However, these candidate biomarkers represent hypotheses that require evaluation in appropriately designed and powered studies.

On the host side, analysis of antibodies has not proved promising due to heterogeneity of the antibody response to *M. tuberculosis* (Kunnath-Velayudhan et al., 2010). Some traction has been achieved with transcriptomic profiling. Unbiased systems biology techniques revealed a type I IFN/inflammatory response (Napolitano et al., 2008; Young et al., 2014). Overall, the ability to directly detect diverse *M. tuberculosis* products in blood or urine is promising. The generally observed low detection sensitivity may be improvable through advances in sample processing (Kashyap et al., 2007), reagent development, and/or the development of novel analytical platforms. It may also be possible to increase specificity, although problems

### Table 1. TPP: A rapid biomarker-based non-sputum-based test for detecting TB

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Optimal requirements</th>
<th>Minimal requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Scope</strong></td>
<td>To develop a rapid biomarker-based test that can diagnose pulmonary TB and optimally also extrapulmonary TB using non-sputum samples (for example, urine, blood, oral mucosal transudates, saliva, exhaled air) for the purpose of initiating TB treatment during the same clinical encounter or on the same day</td>
<td></td>
</tr>
<tr>
<td><strong>Goal</strong></td>
<td>Target groups are adults and children including those who are HIV-positive and suspected of having active pulmonary TB or extrapulmonary TB in countries with a medium to high prevalence of TB as defined by WHO</td>
<td></td>
</tr>
<tr>
<td><strong>Target user of test</strong></td>
<td>Health-care workers with a minimum of training</td>
<td>Trained microscopy technicians</td>
</tr>
<tr>
<td><strong>Target population</strong></td>
<td>Primary health-care clinics with attached laboratories; peripheral microscopy centers or higher levels of the health-care system</td>
<td></td>
</tr>
<tr>
<td><strong>Setting (level of the healthcare system)</strong></td>
<td>Health posts without attached laboratories (that is, levels below microscopy centers) or higher levels of the health-care system</td>
<td></td>
</tr>
<tr>
<td><strong>Performance characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Diagnostic sensitivity for pulmonary TB in adults</strong></td>
<td>≥98% for smear positive culture positive TB, ≥68% for smear negative culture positive TB (that is, similar to Xpert MTB/RIF assay)</td>
<td>Overall ≥65% but should be ≥98% among patients with smear positive culture positive TB (that is, similar to smear microscopy), Overall pooled sensitivity should be ≥80% in adults with HIV infection</td>
</tr>
<tr>
<td><strong>Diagnostic sensitivity for extrapulmonary TB in adults</strong></td>
<td>Ideally, should be ≥80% for all forms of microbiologically confirmed TB</td>
<td>Overall pooled sensitivity should be greater than smear microscopy in adults with HIV infection</td>
</tr>
<tr>
<td><strong>Diagnostic sensitivity in children</strong></td>
<td>Sensitivity for intrathoracic TB ≥66% for microbiologically confirmed TB (that is, similar to Xpert MTB/RIF)</td>
<td>No lower range of sensitivity was defined</td>
</tr>
<tr>
<td><strong>Diagnostic specificity</strong></td>
<td>At least as specific as Xpert MTB/RIF for pulmonary, extrapulmonary, and childhood TB (that is, 98% specificity compared against microbiological reference standard); test should distinguish between active TB and latent or past infection</td>
<td></td>
</tr>
<tr>
<td><strong>Operational characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sample type</strong></td>
<td>Not invasive or minimally invasive, non-sputum samples</td>
<td></td>
</tr>
<tr>
<td><strong>Manual preparation of samples</strong></td>
<td>Sample prep should be integrated or manual prep should not be required</td>
<td>Limited number of steps only; precise measuring should not be required</td>
</tr>
<tr>
<td><strong>Time to result</strong></td>
<td>&lt;20 min including time spent preparing sample</td>
<td>&lt;1 h including time spent preparing sample</td>
</tr>
<tr>
<td><strong>Instrument and power requirement</strong></td>
<td>No instrument needed</td>
<td>Small, portable or hand-held instrument that can operate on battery or solar power</td>
</tr>
<tr>
<td><strong>Maintenance and calibration</strong></td>
<td>Disposable, no maintenance</td>
<td>Minimal maintenance required with automatic alert and remote calibration</td>
</tr>
<tr>
<td><strong>Operating temperature and humidity level</strong></td>
<td>+5°C to +50°C with 90% humidity</td>
<td>+5°C to +40°C with 70% humidity</td>
</tr>
<tr>
<td><strong>Results capturing, documentation, data display</strong></td>
<td>Instrument free test with ability to save test results using separate, attachable reader</td>
<td>Test menu must be simple to navigate; integrated screen, simple keypad or touch screen, ability to save results using instrument or separate reader</td>
</tr>
<tr>
<td><strong>Internal quality control</strong></td>
<td>Internal controls included for processing sample and detecting TB</td>
<td>Internal control only for processing sample</td>
</tr>
<tr>
<td><strong>Price</strong></td>
<td>Price of individual test &lt;US $4.00</td>
<td>&lt;US $6.00</td>
</tr>
</tbody>
</table>
transcriptional signature in peripheral blood during pulmonary TB (Berry et al., 2010), a signature that was subsequently validated in a variety of different populations (Lesho et al., 2011; Lu et al., 2011; Maertzdorf et al., 2011, 2012; Bloom et al., 2013). Importantly, the signature has reasonable ability to differentiate TB from other respiratory and inflammatory diseases (Berry et al., 2010; Maertzdorf et al., 2012; Bloom et al., 2013; Kafourou et al., 2013). The signature also increases with disease activity and diminishes with treatment (Bloom et al., 2012). To date, however, sensitivity and specificity have not met minimum TPP requirements. Further, the complexities inherent in quantitative measurement of the expression levels of a large number of genes make this approach challenging as well as expensive.

Unbiased transcriptional profiling has also been applied to the development of a diagnostic test for childhood TB (Anderson et al., 2014). In children, diagnostic options are currently limited by the difficulty in obtaining samples for microbiological confirmation. As such, the validation of a new test is more difficult in the absence of a gold standard. In the aforementioned study, a transcriptional signature with sensitivity of 83% and specificity of 83% in culture-confirmed TB was reported, with a sensitivity estimated at 35–80% in children with possible or probable TB (i.e., without microbiological confirmation).

Beyond transcriptional signatures, host markers have generally shown significant overlap between symptomatic patients confirmed positive or negative for TB, leading to poor specificity for the markers tested to date. This has been true for serum micro-RNAs (Miotto et al., 2013; Zhang et al., 2013) and proteins (Phalane et al., 2013; Huang et al., 2014; Hur et al., 2015). Volatile organic compounds, likely arising from the host response, have been explored in breath. Using gas chromatography and mass spectroscopy for detection and quantification, a sensitivity of 84% and specificity of 65% has been reported, but without an independent test of performance (Phillips et al., 2010). In sum, even highly complex gene signatures have, to date, failed to meet the high bar required for a diagnostic test. The consistent overlap between TB and non-TB patients across the majority of the host markers suggests the host response may indeed encompass a spectrum of activity that is variable within active TB and overlaps with latent TB infection as well as other illnesses. The exception may be pediatric patients, for whom receiver operator curves for host markers may actually improve on the accuracy of current diagnostic tests and algorithms. The complexity of reducing a high dimensional host signature into a practical test remains a key barrier to the application of this work in TB clinics in endemic settings.

In 2011, the Bill & Melinda Gates Foundation announced a Grand Challenge request for proposals to investigate biomarkers that could enable the development of a nonsputum-based diagnostic for TB. This resulted in investment in a diverse portfolio of projects investigating both pathogen and host biomarkers, including direct measurement of M. tuberculosis products in serum, plasma, and urine analysis of exosomes in blood as a potential enriched source of M. tuberculosis products, a targeted effort to characterize and develop new detection reagents for LAM, and broad efforts to profile host proteins, lipids, and metabolites for diagnostic potential. Data generated to date has underscored the aforementioned challenges, including the conclusion that antibody responses are unlikely to provide useful diagnostics for TB.

**Monitoring therapeutic responses and assessing bacterial burden**

Adequate tools for monitoring responses to existing TB therapeutics are also lacking. This poses a challenge for clinical management and imposes a barrier to the development of novel therapeutics. Similar to diagnostics, tools to monitor treatment must undergo prospective validation and translation to an appropriate testing platform, as depicted in Fig. 1. Standard therapy for treating active TB is for 6 mo, although data from clinical trials of shorter regimens suggest that 80–85% of patients can be successfully treated in 4 mo (Gillespie et al., 2014). However, whereas sputum culture conversion at 2 mo correlates with cure at the end of 6 mo of treatment and lack of relapse, the positive predictive value is low (Wallis, 2015). Moreover, there is no available biomarker that allows prediction of which individuals could be treated for 4 mo with a subsequent relapse rate equivalent to that of 6 mo therapy in nonstratified subjects. This has motivated a search for better biomarkers to identify individual patients who could be effectively treated with a shortened course.

There is a similar need for better biomarkers as surrogate endpoints in testing novel drugs and regimens, as the poor predictive power of sputum culture conversion to determine needed treatment duration, or to rank the relative potency of new regimens, necessitates follow up for at least 18 mo after randomization to measure the impact on relapse in the first year after finishing treatment. In addition to expense and risk, the reliance on sputum culture for drug development will potentially result in the discarding of novel agents with activity against the specific mycobacterial subpopulations most refractory to current drugs. The predictive link from sputum culture to patient outcome is likely weakened by the obvious bias toward identification of bacteria that have access to the airway. In addition to foci of infection in the lung that are occult with respect to sampling, M. tuberculosis can also be found outside the respiratory tract (Neyrolles et al., 2006; Das et al., 2013). There is also a clear bias toward the ascertainment of M. tuberculosis subpopulations that can be grown with standard culture methods, missing other viable populations (Garton et al., 2008; Mukamolova et al., 2010). Tools to quantify the total body burden of M. tuberculosis, both culturable and not, as well as the host response to any remaining mycobacteria, could enable more tailored individual treatment decisions as well as facilitate drug development. There is thus a clear need for the discovery and validation of biomarkers that can provide timely insights into therapeutic responses and cure, including methods for quantifying mycobacterial burden and biomarkers of the host response.
The development of more useful tools is made more challenging by the diverse use cases involved—including both clinical trials and routine clinical use and predictive biomarkers that allow stratification of patients and biomarkers that track response to treatment. For biomarkers to be used in clinical trials, a higher level of cost and complexity could be tolerated, but routine clinical use requires a field-deployable, low-cost test. These different use cases must be clearly defined during the research and development stages, and an appropriate test platform is required before moving to final validation and regulatory approval (Fig. 1). The vast majority of publications investigating treatment response biomarkers fail to articulate the intended use case and underlying TPP. Furthermore, most studies compare changes in proposed biomarkers over time during treatment without testing, or being powered to test, the correlation with patient outcome, i.e., relapse-free cure.

Opportunities to improve sputum analysis include the practical (reducing time and/or broadening application in low resource settings) and the more exploratory (providing information on bacterial populations that are currently unculturable). In the former vein, Friedrich et al. (2013) compared Xpert MTB/RIF results to smear microscopy and culture on solid and liquid media and found the Xpert MTB/RIF assay to have high sensitivity (97%) but poor specificity (49%) to identify culture positive specimens when Xpert is used as a binary readout. When using the quantitative measurement from the Xpert MTB/RIF assay, the change in quantitative sputum bacterial load correlated with smear grades, solid culture grades, and time to liquid culture positivity (all \( \rho = 0.73–0.74; P < 0.001 \)), leaving the possibility that this quantitative data, if used in the analysis, may have additional utility in identifying patients who are cured (Friedrich et al., 2013). The Xpert MTB/RIF test detects DNA from dead as well as live mycobacteria. A recent study tested the value of adding propidium monoazide (which quenches PCR-mediated detection of DNA from dead mycobacteria) to allow for the specific detection of viable bacilli in sputum using Xpert MTB/RIF (Nikolayevskyy et al., 2015). Whereas a positive correlation (\( r = 0.61 \)) with time to positivity of TB cultures on liquid media was reported, performance would need to be improved to allow practical use. Currently, results comparing DNA measurements with culture show relatively poor correlations. Further investigation will be needed to understand whether these correlations can be improved by accounting for live but unculturable bacteria in the sample. The contribution of such bacteria to patient outcomes, the need for continued treatment, and risk of relapse, must also be defined to understand the potential clinical utility of such analyses.

Beyond sputum, our fundamental lack of understanding of the total body burden and distribution of \( M. \) \( \text{tuberculosis} \) during latent infection, in active disease, and through the course of treatment should motivate studies to characterize total mycobacterial load. Such studies will need to start in fit-for-purpose animal models that display a broad spectrum of infection and disease states, such as cynomolgus macaques infected with low-dose \( Mtb \) (Capuano et al., 2003; Lin et al., 2012) and guinea pigs infected in natural transmission models (Dharmadhikari et al., 2011). Linked analysis of clinical specimens from well-characterized, longitudinal cohorts would be required to test hypotheses arising from such studies. Though challenging, progress in characterizing the total body \( M. \) \( \text{tuberculosis} \) burden will likely be critical for making progress in transforming the diagnosis, treatment, and prevention of TB.

On the host side, investigations into biomarkers that reflect TB burden remain at the hypothesis-generating stage. Higher baseline levels of antibodies to specific \( M. \) \( \text{tuberculosis} \) proteins were reported to correlate with slower sputum conversion, but antibody levels overlap with those seen in fast responders, hampering practical utility (Baumann et al., 2013). Plasma vascular endothelial growth factor concentrations at 2 wk have been positively correlated with time to sputum conversion (Riou et al., 2012). Decreases in frequencies of \( M. \) \( \text{tuberculosis} \) antigen-stimulated IFN-\( \gamma \)-secreting CD4+ T cells expressing CD38 and HLA-DR, and Ki-67 were associated with resolution of sputum smear and culture positivity in a preliminary study (Adekambi et al., 2015). Detection of systemic heme oxygenase-1 (HO-1) levels showed some promise, with a larger number of patients who were successfully treated showing decreased HO-1 levels, and levels remaining unchanged in five patients who were culture positive at the end of 6 mo of treatment (Andrade et al., 2013). Nahid et al. (2014) measured hundreds of proteins in serum samples from 39 patients and identified a five-marker signature (IL-11, R\( \alpha \), \( \alpha_2 \)-Antiplasmin, PSME1, and SAA) that predicted conversion with an estimated 80 ± 11% sensitivity and 80 ± 7% specificity upon internal cross-validation. All of these studies require confirmation in independent cohorts and—most importantly—correlation with relapse-free cure.

High-resolution 3-dimensional imaging provides unique opportunities here (Chen et al., 2014). In a population of patients with multidrug-resistant TB undergoing 2 yr of therapy, quantitative, volumetric changes in the uptake of 2-deoxy-2-[\( ^{18} \)F]-fluoro-D-glucose (FDG) at 2 and 6 mo, as quantified by positron emission tomography/computed tomography (PET/CT), was correlated with treatment outcome. Additional studies in drug-sensitive cohorts should provide additional data on the potential utility of such modalities in monitoring TB patients. \( M. \) \( \text{tuberculosis} \)-specific reporter molecules (Backus et al., 2011) compatible with PET/CT imaging hold potential for providing complementary data. The potential ability of the latter techniques to enable quantification of mycobacterial load in vivo is particularly exciting. Because of the requirement for specialized equipment, complexity, and cost, however, the use of such methods will likely be restricted to clinical trials.

Thus, although there are promising developments, significant work remains to be done to identify and validate therapeutic response biomarkers that can provide timely, practical, and actionable information on patient outcomes. To achieve this, longitudinal studies to characterize pathogen and host biomarkers, alongside imaging-based methods, will likely be required.
Prevention of disease and infection

TSTs and IGRAAs provide presumptive evidence of sustained infection but are unable to differentiate between latent infection and active disease. More importantly, these assays provide no predictive power for the risk of progression to active disease. The identification of biomarkers that facilitate practical stratification of latently infected individuals for risk of progression to active disease could allow for rational targeting of secondary prophylactic interventions, both existing and novel. Furthermore, mechanistic insight into the biology of progression provided by such biomarkers holds promise for the development of novel host-directed therapeutic and/or preventive approaches. Finally, such stratification has promise for aiding TB vaccine development. Any such biomarker would need to undergo confirmation and prospective validation and a move to an appropriate detection platform for use in clinical trials or, more broadly, in endemic settings (Fig. 1). As recently reviewed (Karp et al., 2015), the TB vaccine enterprise is at something of a crossroads. Lacking animal models known to be predictive, or any validated correlate of protection, the field lacks a clearly rational way to up- or down-select vaccine candidates for clinical development. Moreover, as prevention of disease has been the standard test of efficacy, and only 10% of (immunocompetent) hosts with sustained infection develop TB over a lifetime, efficacy trial sizes are large and prohibitively costly. Dramatic decreases in trial size and length might be secured by targeting such trials to those at highest risk for progression to active disease. Progress made in identifying and validating such biomarkers, on the host side, is noted below.

Studies to identify biomarkers that provide a quantitative assessment of the risk of progression to active disease are difficult to perform due to the low rate of progression, leading to the need to prospectively collect samples from large populations over time. Validation similarly requires a large prospectively followed cohort. A recent study identified a gene expression signature comprising 16–21 signal genes and 10 reference genes—mirroring the type I IFN/inflammatory signature reported in cross-sectional studies of active TB noted above—that predictively (months prior) discriminates progressors from controls with a sensitivity and specificity of 67 and 81%, respectively (unpublished data; W. Hanekom, personal communication). Such performance holds out the possibility of identifying individuals at risk for TB well before they become symptomatic, although specificity may need to be improved and gene number decreased to make practical use of such a test.

Although it remains to be determined whether type I IFN activity is causally linked to progression and disease, biological plausibility is high (Dorhoi et al., 2014; Mayer-Barber et al., 2014; McNab et al., 2014; Redford et al., 2014). Indeed, this link may well underlie the early susceptibility to TB seen after HIV infection—itself associated with sustained increases in type I IFN production (Hycrzka et al., 2007; Byrnes et al., 2008). It might also contribute to the spring peak in TB diagnosis observed across latitudes in temperate climes (Willis et al., 2012), as a result of winter respiratory virus-induced type I IFNs amplifying the dynamic flux of granuloma activity in latent infection. Defining causality could be achieved either through testing whether pulmonary delivery of exogenous type I IFNs can induce reactivation in a fit-for-purpose animal model of latency (Lin et al., 2014) or, more directly, through testing whether monoclonal antibody–mediated blockade of type I IFN activity prevents progression to disease in individuals at high risk. If causal, how might this be exploited? For one, it suggests targets for host-directed therapy. It also suggests the potential for immunizing against other common pathogens that drive robust type I IFN activity as a strategy for preventing TB.

A further impediment to the effective prevention of TB is the lack of a validated correlate of vaccine-induced protection to aim for, given the lack of a vaccine with demonstrated efficacy. Moreover, there is growing realization in the field that the presumptive mechanistic correlate underlying TB vaccine development for the past decades is likely to be inadequate or incorrect (Karp et al., 2015). The nomination of novel mechanistic correlates of protection for vaccine candidates to aim at, deriving from a deeper understanding of TB immunopathogenesis, from the study of outlier populations (e.g., those with persistent exposure in the absence of apparent sustained infection), and/or from insights derived from the definition of correlates of risk has real promise (Karp et al., 2015). However, these biomarker issues are beyond the scope of the current Perspective.

Prevention of sustained infection appears to be a rational goal for novel vaccine strategies (Karp et al., 2015). Testing such vaccines with current tools is somewhat problematic. Sustained TST and/or IGRA positivity does not provide endpoints as firm as disease endpoints. The development of sensitive ways for detecting and quantifying low M. tuberculosis burdens, and/or the presence, sustained or not, of low levels M. tuberculosis infection, discussed above, would be a major advance for the development of vaccines aimed at preventing sustained infection.

The definition of correlates of protection is beyond the purview of this Perspective. It bears noting, however, that our available tools for identifying outlier individuals whose immunological interrogation might nominate a novel vaccine concept based on eliciting a type of uncommon immunity—highly exposed individuals without apparent sustained infection—have a major, illustrative limitation. The lack of demonstrable TST and/or IGRA reactivity may reflect a lack of infection, the development of a conventional T cell response with test reversion after the achievement of sterile protection, or the development of an uncommon, protective immune response that is not measurable by assays that depend on T cell reactivity with proteins.

Moving the field forward

There is a pressing need for new tools. That said, these needs are often not well served by the current study paradigms, which are generally not guided by the clarity provided by
careful consideration of intended use case and the TPPs that derive from such. Under each use case, a detailed TPP provides the guidance needed to evaluate biomarker performance and utility against intended use. Targets within the TPP should, when possible, be informed by modeling the potential impact of a specific set of product attributes. This approach provides clear targets for researchers and product developers, and actionable results when targets are met. The articulation of clear use cases allows studies and clinical cohorts to be designed against the intended use, with the relevant control groups and key follow-up clinical data collected. Study size is often limited for early, exploratory biomarker projects. When moving to validation, appropriate power calculations are needed. It is noted that transformative progress may require funders to align behind a definitive approach to these challenges rather than enabling the proliferation of multiple, redundant, suboptimal studies.

Though publication of new biomarkers is common, refinement, validation and independent confirmation of such is not. Further, once discovered, the performance of proposed biomarkers is likely to benefit from focused efforts to improve the sensitivity and specificity and/or reduce the assay to manufacturability and practicality—bridging the gap between exploratory identification and reduction to practice by developers. A framework for these efforts is outlined in Fig. 1.

The authors thank C. Wilson, L. Stuart, K. Duncan, G. Kaplan, A. Kasmarr, and W. Hannekom for helpful discussions.

The authors declare no competing financial interests.

REFERENCES


Eskola, J., M. Tynkkynen, O. Wilkins, M. Ostrowski, and S.D. Der. 2007. Distinct transcriptional profiles in ex vivo CD4+ and CD8+ T cells are established early in human immunodeficiency virus type 1 infection and are characterized by a chronic interferon response as well as extensive transcriptional changes in CD8+ T cells. *J. Virol.* 81:3477–3486. http://dx.doi.org/10.1128/JVI.01552-06


