nature microbiology

ANALYSIS

https://doi.org/10.1038/s41564-019-0380-2

A systematic review of biomarkers to detect active tuberculosis

Emily MacLean[1,4], Tobias Broger[2,4], Seda Yerliyaka[2], B. Leticia Fernandez-Carballo[2], Madhukar Pai[3] and Claudia M. Denkinger[2*]

Millions of cases of tuberculosis (TB) go undiagnosed each year. Better diagnostic tools are urgently needed. Biomarker-based or multiple marker biosignature-based tests, ideally performed on blood or urine, for the detection of active TB might help to meet target product profiles proposed by the World Health Organization for point-of-care testing. We conducted a systematic review to summarize evidence on proposed biomarkers and biosignatures and evaluate their quality and level of evidence. We screened the titles and abstracts of 7,631 citations and included 443 publications that fulfilled the inclusion criteria and were published in 2010–2017. The types of biomarkers identified included antibodies, cytokines, metabolic activity markers, mycobacterial antigens and volatile organic compounds. Only 47% of studies reported a culture-based reference standard and diagnostic sensitivity and specificity. Forty-four biomarkers (4%) were identified in high-quality studies and met the target product profile minimum criteria, of which two have been incorporated into commercial assays. Of the 44 highest-quality biomarkers, 24 (55%) were multiple marker biosignatures. No meta-analyses were performed owing to between-study heterogeneity. In conclusion, TB biomarker discovery studies are often poorly designed and findings are rarely confirmed in independent studies. Few markers progress to a further developmental stage. More validation studies that consider the intended diagnostic use cases and apply rigorous design are needed. The extracted data from this review are currently being used by FIND as the foundation of a dynamic database in which biomarker data and developmental status will be presented.

Tuberculosis (TB) is the leading infectious disease cause of death globally. Traditional TB diagnostic methods, such as culture or smear microscopy, are slow or low in sensitivity. Molecular techniques, such as GeneXpert MTB/RIF, are costly and often unavailable in primary-care settings because of their infrastructure needs. Providing high-quality patient-centred diagnostics and care when patients first present is critical to reaching the End-TB targets[1]. This should increase the number of TB cases detected and reduce patient losses at the initial stage of the care cascade[2]. New solutions are needed. The World Health Organization (WHO) defined the performance and operational characteristics of a test suitable for primary care or at the point of care (POC) in its high-priority target product profiles (TPPs)[3] (Supplementary Methods).

Biomarkers are defined as 'objective characteristics that indicate a normal or pathogenic biological process'[4]. An acceptable diagnostic biomarker or multiple marker biosignature for TB would be a pathogen or host marker that is necessary and specific to the disease's underlying process[5]. With TB, a necessary but not perfectly specific marker to support a rule-out or triage test is also a diagnostic need[3].

To meet the TPPs, a biomarker test would ideally be instrument free or feasible with limited instrumentation and would utilize easily accessible samples, such as blood, urine or breath. Non-DNA-based biomarker tests are more likely to meet the operational and cost targets of TPPs than are DNA-based tests. Although non-DNA biomarkers have the disadvantage of not providing information about drug susceptibility profiles, an easy-to-perform first test without drug susceptibility testing capabilities is in line with the TPPs.

TB biomarkers research is an area of high activity, but its impact thus far has been limited. Except culture, the only WHO-endorsed tests for active TB detection are based on DNA detection in sputum (Cepheid's GeneXpert MTB/RIF, Eiken's LAMP, and Hain and Nipro's line probe assays) or are approved only for a limited use case (Abbott's Determine TB-LAM lateral flow assay, which detects lipoarabinomannan (LAM) antigen in the urine for TB diagnosis in people living with human immunodeficiency virus (PLHIV) with CD4 counts < 100 cells per μl). Even for these approved tests, uptake and implementation have been slow[6], which can partially be attributed to the tests not perfectly fitting the needed TPPs.

Systematic reviews of biomarkers, such as interferon-γ (IFN-γ)[7–9], LAM[10,11], antibody-based assays[12] and incipient TB markers[13] have already been published. Narrative reviews regarding the state of TB biomarkers research are available and informative, but do not present the full body of research[5,14,15].

In this systematic review, we synthesize the published work on biomarkers and multiple biomarker biosignatures for the detection of active TB. Our primary objective is to capture the existing biomarkers and evaluate the quality and level of evidence around them. Our secondary objective is to identify the most promising biomarkers and biomarker categories for the development of a POC test for standalone detection or triage for active TB.

Results

Results of the search. After deduplication, 4,470 publications were identified. Figure 1 illustrates the selection process. A total of 443 publications were included in this review.

Most publications were excluded for not addressing the target condition of active TB detection. For example, many publications reported biomarkers for latent TB infection detection or were basic science studies. Furthermore, 308 studies that utilized IFN-γ release

[1]Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, Québec, Canada. [2]FIND, Geneva, Switzerland. [3]McGill International TB Centre, Research Institute of the McGill University Health Centre, Montreal, Québec, Canada. [4]These authors contributed equally: Emily MacLean, Tobias Broger. *e-mail: cdenki@gmail.com

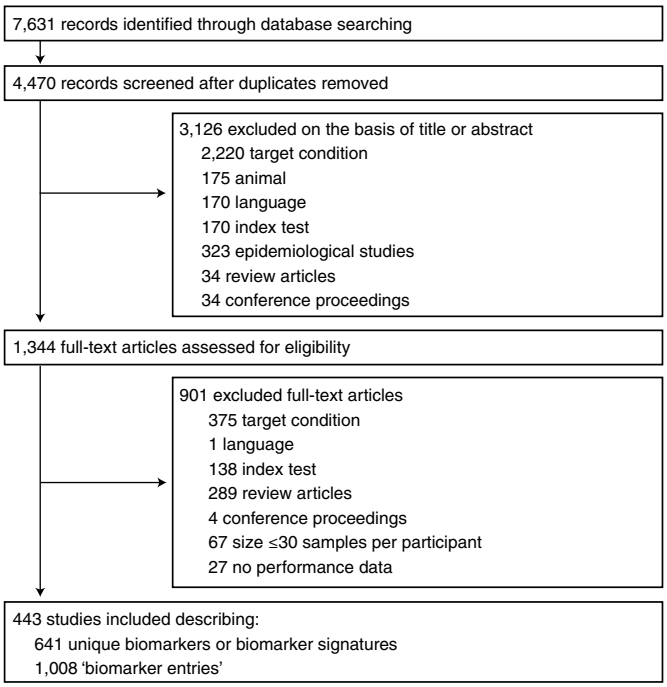

Fig. 1 | PRISMA flow chart of publication selection.

assays for active TB or growth-based or imaging-based index tests were also excluded.

Characteristics and quality of included studies. Summarized results for study quality and risk of bias based on Quality Assessment of Diagnostic Accuracy Score 2 (QUADAS-2) questions are shown in Fig. 2, with full results for each study detailed in Supplementary Table 1. The risk of bias was high in most publications, primarily because of retrospective and case–control designs, lack of consecutive sampling and use of controls that can inflate accuracy estimates.

Most studies (264 out of 443, 60%) sampled individuals in high TB burden countries as defined by the WHO[16]. The majority (70%) of publications reported results in adults, whereas demographic information was not given in 16% of publications, 8% used mixed adult and paediatric patients, and 6% reported results solely in children.

Entries per biomarker category and analysis group. The 1,008 biomarker entries from the 443 studies were classified into 12 predefined biomarker categories of either host or mycobacterial origin, and then into 2 analysis groups (Fig. 3). Fewer than half (472 out of 1,008) of the entries were classified into the α-group, which presumes a culture-based reference standard and diagnostic performance data. Although most entries in the β-group utilized a culture-based reference standard, many (448 out of 1,008) only reported *P* values for biomarker performance.

The majority (909 out of 1,008, 90%) of entries were host biomarkers, with antibody and cytokine categories being dominant (Fig. 3). Only 10% (99 out of 1,008) of all entries were pathogen biomarkers; of those, 63% (62 out of 99) provided data suitable for analysis. The most frequently reported pathogen marker was LAM (35 entries, 34 in the α-group) (Fig. 3).

Study sizes varied greatly. Few biomarkers have been assayed using large sample sizes (28 entries ≥ 500 samples) and most were evaluated in relatively small samples (614 entries ≤ 100 samples).

Biomarkers with reported diagnostic performance. All α-group biomarker entries (472 out of 1,008) were plotted on a receiver operating characteristic (ROC) plot by category (Fig. 4a). Overall, the reported diagnostic performances generally show higher specificity, whereas sensitivity varies across a wider range. The selected negative control group was considered one of the most important sources of potential bias in diagnostic performance. Thus, Fig. 4b displays biomarkers that were assayed against a clinically relevant negative control population (other respiratory disease (ORD) for pulmonary TB, other diseases for extrapulmonary TB or either group for paediatric TB), leaving only 151 out of 472 (32%) biomarkers and biosignatures. The antibody (54 out of 151, 36%) and cytokine and chemokine (24 out of 151, 16%) groups remain the largest categories, with LAM (18 out of 151, 12%) remaining well represented. Host RNA (8 out of 151, 5%), mycobacterial other markers (5 out of 151, 3%) and volatile organic compounds (5 out of 151, 3%) also remain but in much smaller numbers.

Biomarkers that meet TPP diagnostic performance criteria. Forty-four entries from Fig. 4b met at least one TPP minimum performance criteria (purple-shaded area, enlarged in Supplementary Fig. 2), the details of which are presented in Fig. 5. Thus, biomarkers included in Fig. 5 have been assayed against a clinically relevant negative control population and meet the minimum diagnostic performance targets of the TPPs.

Overall, 24 of 44 (55%) entries in Fig. 5 were multiple marker biosignatures. Contrastingly, biosignatures comprised only 19% (187 out of 1,008) of total entries, suggesting that biosignatures have better potential to reach TPP performance targets. Ten (23%) markers were assessed for extrapulmonary TB detection. The majority (73%, 32 out of 44) of entries shown in Fig. 5 were isolated from blood samples. The chosen study design of approximately half (52%, 23 out of 44) of entries in Fig. 5 was at low risk of bias, that is, cohort or cross-sectional. The remaining 48% of studies are still at relatively high risk of bias, although clinically relevant populations were used.

Host: antibody detection. Figure 5 includes 16 entries from the antibody class. Performance was notably high in three cohort studies: a study describing antibodies against the *Mycobacterium tuberculosis* protein A60 (ref. [17]); an eight-antibody signature (anti-Ag85B, anti-Ag85A, anti-Ag85C, anti-Rv0934-P38, anti-Rv3881, anti-BfrB, anti-Rv3873 and anti-Rv2878c antibodies)[18]; and a seven-antibody signature (anti-BCG (Bacillus Calmette–Guérin) IgG, anti-LAM IgG, anti-TB15.3 IgG, anti-TB51a IgG, anti-10 kDa culture filtrate protein (CFP-10)/ESAT-6 IgG, anti-CFP IgG and anti-CW IgG)[19]. The remaining antibody detection studies[20–26] did not use a cohort study design.

Host: cytokines and chemokines, proteins and metabolic activity markers. All but 3 of the 16 entries were multiple marker biosignatures. One small, early-phase study showed promise for an eight-marker biosignature (granzyme A, growth/differentiation factor 15 (GDF15), serum amyloid A (SAA), interleukin-21 (IL-21), C-X-C motif chemokine 5 (CXCL5), IL-12(p40), IL-13 and plasminogen activator inhibitor-1) measured in the saliva[27]. In the blood, a seven-marker signature was described in a study at low risk of bias (C-reactive protein (CRP), transthyretin, IFN-γ, complement factor H, apolipoprotein-A1, IFN-γ-induced protein 10 (IP-10; also known as CXCL10) and SAA) with promising performance for a triage test[28]. One study reported the biosignature of neural cell adhesion molecule, serum amyloid P, ferritin, complement factor H and extracellular matrix protein 1 (ECM-1) in the blood, as well as three promising six-marker biosignatures[29]. Plasma proteomic fingerprinting showed promise in an early-phase study to detect a combination of protein peaks between *m*/*z* 6,000 and *m*/*z* 12,000 (*m*/*z* represents the mass divided by the charge number in mass spectrometry)[30]. Monocyte chemotactic protein 1 (MCP-1; also known as CCL2)[31] and several biosignatures[32–35] detected in the blood showed promising performance, but must be further

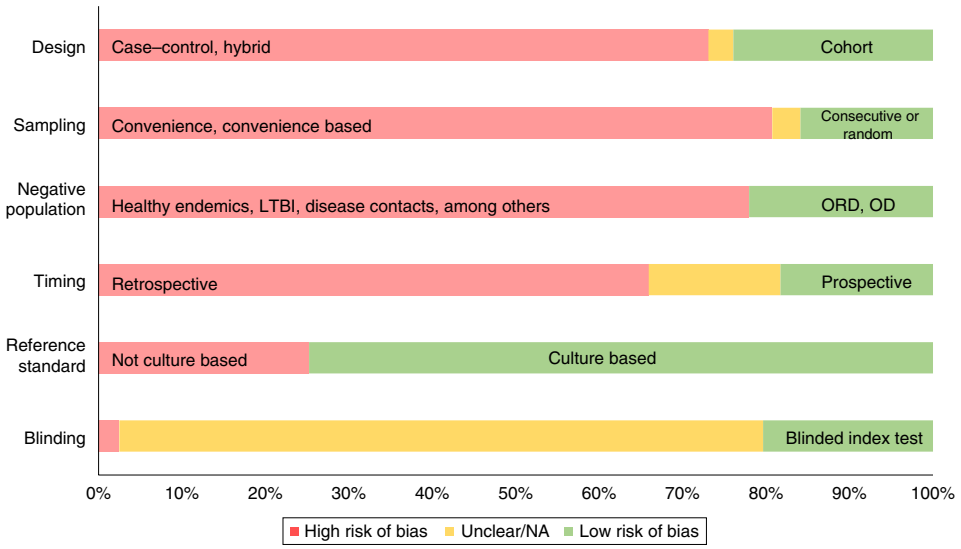

Fig. 2 | Summary of the results of QUADAS-2 assessment for the risk of bias to assess study quality. Answers to QUADAS-2 questions are given. The bar length is representative of the proportion of answers to each question. LTBI, latent TB infection; NA, not applicable; OD, other disease.

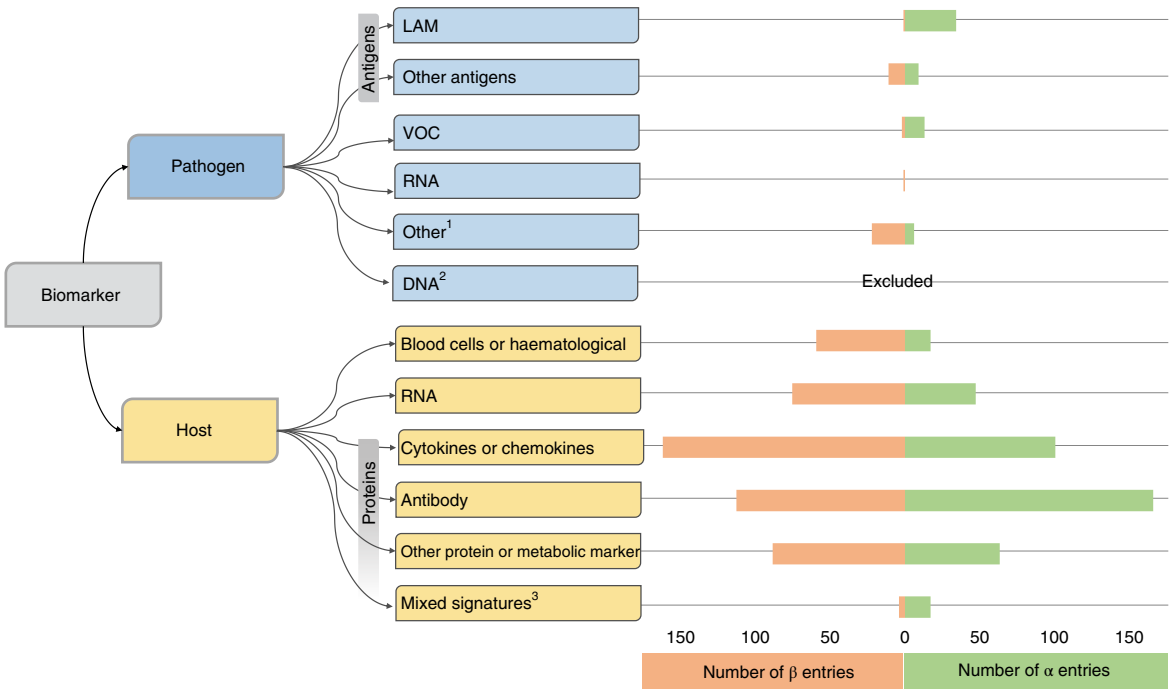

Fig. 3 | Classification and number of entries per biomarker category and analysis group. In total, 1,008 biomarker entries from the set of 443 studies were included. α-Group entries were reported with sensitivity and specificity values and were evaluated against a culture-based reference standard, whereas β-group entries lacked either or both of these characteristics. [1]Other pathogen markers include whole bacilli, *M. tuberculosis* metabolites and mycolic acids. [2]Pathogen DNA biomarkers were excluded a priori (see the Methods section). [3]Mixed signatures consist of biomarkers from multiple categories. VOC, volatile organic compound.

validated as the studies were small or at overall high risks of bias. With regards to more invasive samples, two studies reported promising performance for adenosine deaminase (ADA) for extrapulmonary TB detection in pleural fluid and cerebrospinal fluid[36,37], and two studies detected host marker signatures in pleural fluid[38,39].

Host: RNA. All RNA signatures were observed in the blood, including a 44-transcript signature in a case–control study of 102 subjects in a validation cohort[40]; more compact transcriptional signatures of guanylate-binding protein 5 (GBP5) and CD64, as well as GBP5, high-affinity IgG Fc receptor I (FCGR1A) and granzyme A were observed in one study of 49 subjects[41]; and a signature of GBP5 and Krueppel-like factor 2 (KLF2) was observed in a larger study of 353 participants[42].

Host: haematological. IFN-γ^+, tumour necrosis factor (TNF)$^+$, CD4$^+$ T cells in pleural fluid had promising performance in a small ($n=41$) study of patients with pleural TB[39].

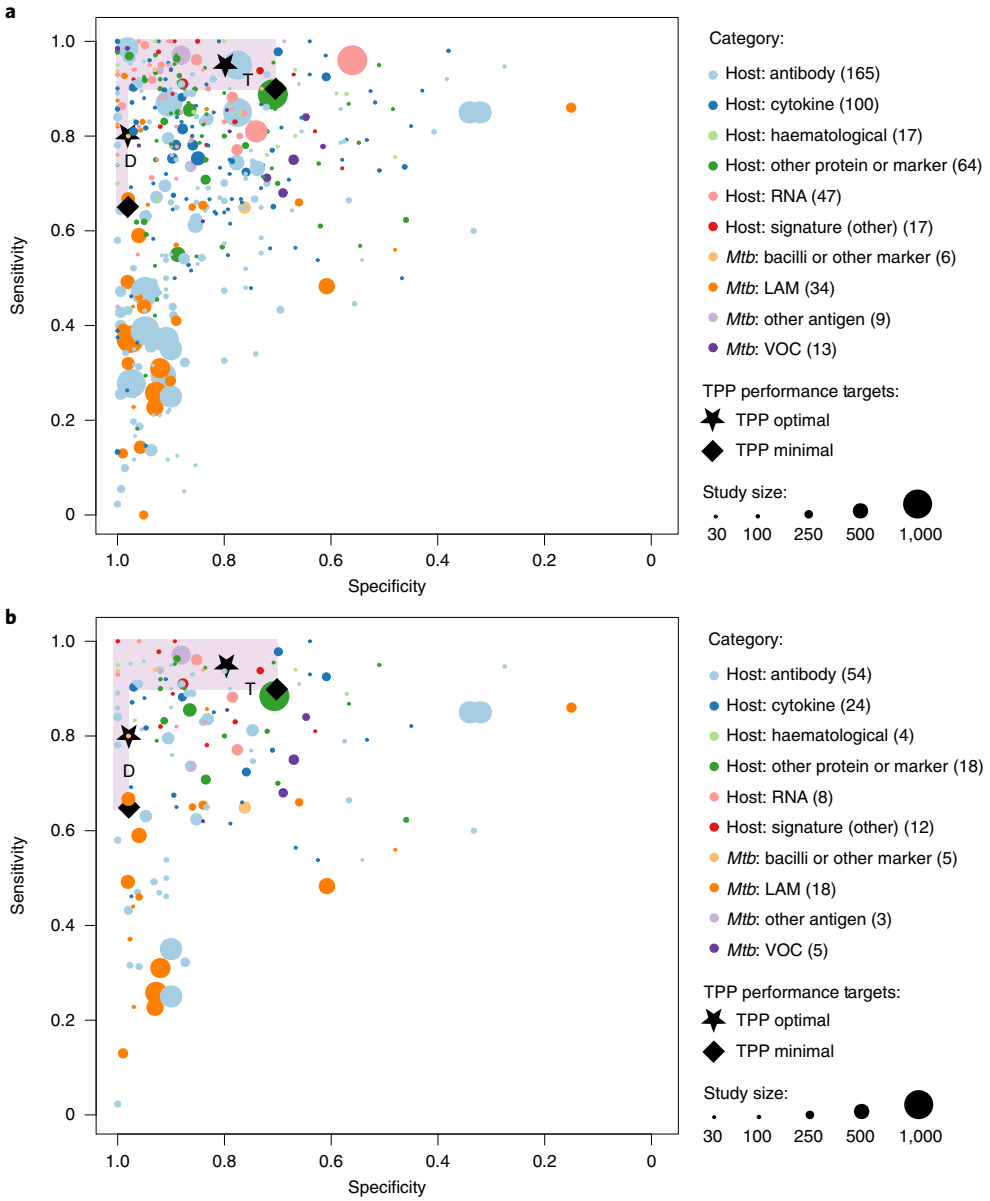

Fig. 4 | Biomarker entries per category in ROC plots. Symbol colours represent different biomarker categories, and the size of the markers represents the study sample size. The purple-shaded area represents areas with sensitivity and specificity combinations that meet at least the minimal target of one of the TPPs. 'D' refers to the TPP criteria for a detection test and 'T' refers to the TPP criteria for a triage test. **a**, All α-group entries ($n = 472$). **b**, α-Group entries assayed against a clinically relevant negative control group ($n = 151$). See Fig. 5 and Supplementary Fig. 2 for details on studies in the purple-shaded area.

Pathogen: LAM. One large, high-quality study[43] reported LAM with high performance in immunocompromised HIV-positive patients with CD4 counts of <50 cells per µl, but performance in patients with higher CD4 counts did not reach the TPP criteria.

Pathogen: other antigen. An immunoassay detecting a non-characterized *M. tuberculosis* antigen (smTB-Ag) with an unknown molecular weight was measured in a case–control study of 154 patients[22]. CFP was identified in the blood using an agglutination test in a study with 638 participants[44].

Pathogen: other markers. Four studies reported high accuracies for the detection of other pathogen biomarkers in sputum: two described whole *M. tuberculosis* bacilli detection[45,46], one reported an unspecified mycolic acid using mass spectrometry but only included 44 subjects[47], and one concerned a signature of mycolic acids, mycobacterial fatty acids and mycolic acids[48].

Markers with limited performance. Entries that were assayed against clinically relevant control populations (Fig. 4b) but did not meet either of the TPP performance targets (unshaded area of Fig. 4b) are presented in Supplementary Table 2.

Briefly, 107 biomarker and biosignature entries from Fig. 4b did not reach TPP targets. Antibody markers represent 36% (38 out of 107) of this group, including those generated against well-characterized *M. tuberculosis* antigens, such as antigen A60, LAM, PstS1 and 38 kDa protein[22,49–52]. Cytokines constituted 21% of this group (22 out of 107), with the majority being individual cytokines or chemokines (17 out of 22, 77%). Examples include CXCL11, CXCL9, IL-10, IP-10 and IFN-γ-containing signatures[31,39,53–57]. With respect

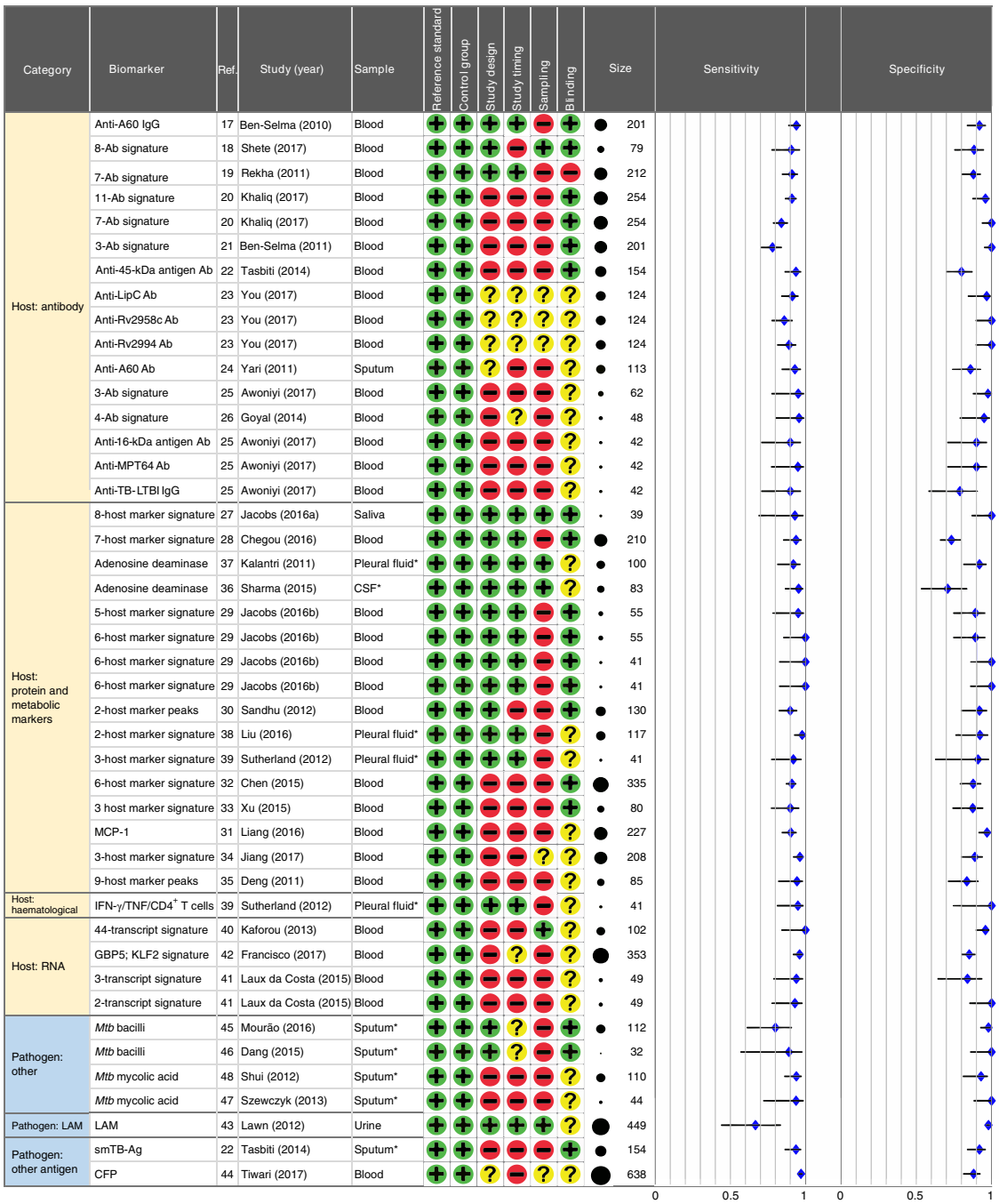

Fig. 5 | Details of the 44 entries that meet the TPP criteria in clinically relevant populations. Sample types that are not considered by the TPPs are marked with an asterisk (*). For simplicity, the host categories cytokines, other metabolic markers and mixed signatures have been condensed into one group. Ab, antibody; CSF, cerebrospinal fluid; LipC, lipase C; *Mtb*, *M. tuberculosis*; LTBI, latent TB infection.

to host metabolic markers and other proteins, general inflammation markers, such as CRP, ADA, SAA and presepsin[36,58–62], accounted for 9% (11 out of 107). Most (8 out of 11, 80%) were assessed in cohort studies with a lower risk of selection bias. In the RNA category, a 51-host transcript signature, complement component 1, C chain (C1qc) transcript, GBP5 and dual-specificity protein phosphatase 3 (DUSP3) signature, and GBP5 did not reach TPP targets[42,63,64].

The five multiple category signatures with sub-TPP target performance were CRP, ferritin, serum amyloid P, MCP-1, α-2-macroglobulin, fibrinogen and tissue plasminogen activator[65]; IL-1β, IL-23, ECM-1, hemofiltrate CC chemokine 1 (HCC1) and fibrinogen[27]; IFN-γ and ADA[36]; IP-10, IFN-γ, ferritin, SAA, CRP and antigen-stimulated IP-10 (ref. [58]); and a CRP and anti-A60 antibody signature[66]. We identified a large study ($n = 1{,}117$) with low risk of bias that evaluated the use of CRP for active TB screening in consecutive PLHIV with CD4 counts of <350 cells per µl (Fig. 4b and Supplementary Table 2). Compared to culture-confirmed TB, CRP displayed a sensitivity of 0.89 and a specificity of 0.94, approaching the triage TPP minimum target[67].

Mycobacterial antigen 85A had a low diagnostic performance[68]. Regarding LAM, all 17 high-quality studies showed its (already well-documented) low sensitivity. Five volatile organic compound entries had diagnostic performance below the TPP criteria[69–72] and were taken from breath samples, except one measured in urine[72].

Discussion

Existing biomarkers research for TB thus far has yielded limited success. The present systematic review assesses the performance of biomarkers and biosignatures for active TB detection and appraises the quality of the studies.

Few 'front-runners' have been identified that could serve as the basis of a diagnostic test. The 44 biomarker and biosignature entries judged to be of higher quality with promising diagnostic performance were mainly of host origin (37 out of 44, 84%) and included antibodies, cytokines and chemokines, and RNA signatures. Only 7 of the 44 entrees were of mycobacterial origin, and only 2 (LAM and CFP) were from easily accessible urine or blood samples (Fig. 5).

Many studies, but limited or biased evidence. This review shows the considerable extent of activity in TB biomarker discovery, while highlighting some limitations of the field that at least partially explain the limited success; relatively few well-designed studies exist and even fewer studies are performed to validate previously discovered biomarkers. These limitations are present in biomarker research at large[73]: discovery studies are performed on insufficiently representative samples and similar efforts are repeated unnecessarily, without 'nominating' promising biomarkers to a further phase of validation[74].

As the immunological processes of TB are complex, it is unsurprising that only a small fraction (44 out of 1,008) of identified biomarkers may be useful for disease detection; more troubling is that we were able to consider only a minority of entries (472 out of 1,008, 47%) for analysis. For example, owing to the absence of sensitivity and specificity estimates, 448 entries that solely reported *P* values were not analysed. This represents 35,897 patient samples and substantial research costs. Certainly, some entries were from exploratory studies whose objective was to identify new candidate biomarkers, but *P* values should not be considered adequate for decision-making regarding biomarker performance. The American Statistical Association warns that *P* values are often misused and misinterpreted, and often lead to irreproducible research findings. Thus, we strongly recommended compliance with STARD reporting guidelines and reporting of 2 × 2 tables, sensitivity and specificity, and related confidence intervals[75] (Table 1).

We found that most included studies are at high risk of bias and thus the reported biomarkers' diagnostic performances may be inaccurate. Specifically, a large proportion of studies did not utilize a clinically relevant negative control population, employed a case–control study design and sampled participants by convenience. Such design choices often lead to over-inflated diagnostic performances[76,77]. This is consistent with conclusions from a previous systematic review of serological TB tests[12]. Again, we recommend consulting the widely established STARD tool when designing, planning and reporting studies to minimize study bias and clearly inform readers of limitations. Although restricted access to samples in the early discovery period might explain some of these findings, biorepositories that support such research do exist[78,79].

Host biomarkers. Of the most promising host entries (Fig. 5), 24 out of 37 were biosignatures. It seems that multiple markers are typically required to reach TPP-level specificity, as diseases other than TB can influence circulating biomarker concentrations. Nevertheless, these host biosignatures primarily have potential for a triage test because specificity would not be high enough for a detection test.

Antibodies remain the most researched biomarker category, with 277 out of 1,008 entries (27%), probably because of their ease of translation into simple-to-use assays. However, only 16 (Fig. 5) of all antibody biomarkers and biosignatures were of relatively high quality and fulfilled TPP criteria, indicating their current inaptness for TB diagnostic tests. Of these 16 entries, only three had a study design with a low risk of selection bias (Fig. 5). Anti-A60 IgG was measured in a cohort study using the Anda-TB assay[17], but a previous meta-analysis clearly showed its insufficient accuracy[12]. A seven-antibody signature was reported in a validation study using the analytically complex antibodies in lymphocyte supernatant assay[19]. One recently recorded eight-marker signature was evaluated in a random subset of samples from a cohort study with promising results and further confirmation should be pursued[18]. Overall, our findings mainly agree with previous work that antibody detection will probably not meet TPP criteria for active TB detection[12,80] and there are several well-designed studies that confirm this (Supplementary Table 2). However, given the rapidly evolving understanding of the TB disease continuum, there may be a role for antibody-based markers of disease progression or reactivation[81].

The host protein and metabolic markers are also a frequently observed category, probably because these proteins are easily measured by readily available commercial multiplex panels. For many studies, researchers used similar, commercially available multiplex inflammation marker panels, which results in the repeated probing of a limited range of markers. Unbiased approaches to biomarker identification, such as in proteomics, are very rarely taken.

We identified one noteworthy larger study with a low risk of bias reporting on a promising seven-host marker signature[28]. Interestingly, the best-performing marker in the seven-host marker signature is CRP[28]. However, the specificity of general inflammation markers, including CRP, is often low[82]. This is not surprising as they are upregulated during pro-inflammatory states[83,84]. In addition, baseline concentrations are higher in low socioeconomic populations[85], and often people with TB are from such groups. Interestingly, a large study by Yoon and colleagues that evaluated CRP for active TB screening in PLHIV showed accuracy approaching the TPP minimal target (Supplementary Table 2). Performance was better in PLHIV with CD4 counts of <200 cells per μl, suggesting that, similar to urine LAM, it may be most useful as a test in severely ill PLHIV[67]. However, it remains unclear whether general inflammation markers will reach sufficient specificity to rule out disease in patients who present with TB-like symptoms[86].

Most host RNA studies were published after 2014, indicating that this category is a growing area of discovery-phase research. It is driven by commercially available microarrays (for example, Illumina and Affymetrix) that facilitate unbiased, comprehensive gene expression analysis. There were four blood biosignatures (Fig. 5) that met the triage test TPP performance criteria in case–control studies. Importantly, although measurement of transcript signatures on a portable platform (for example, Cepheid Omni) seems feasible, cost targets and operational characteristics of a triage test are unlikely to be met[87]. However, these assays could be considered for incipient and subclinical TB, for which there is a large market in high-resource settings that is less price sensitive[81,88–91]. This use case was beyond the scope of this study and has been discussed elsewhere[92].

Going forwards, it seems clear that a multiplex assay will almost certainly be required to reach sufficient detection and triage test performance with host markers. However, the quantitative measurement of several markers at clinically relevant concentrations on a simple, low-cost, multiplex POC platform will be a development hurdle to overcome (Table 1).

Pathogen biomarkers. In contrast to host markers, *M. tuberculosis* 'pathogen' marker detection has the potential to be highly specific and useful for the development of detection tests for the purpose of initiating treatment (Supplementary Methods). Detection of pathogen DNA (excluded from this review) is the basis of the WHO-endorsed PCR-based diagnostic tests, for example, Gene Xpert MTB/RIF[93], but requires complex and expensive technology. From a POC test perspective, the detection of non-DNA *M. tuberculosis*

Table 1 | Key findings and suggestions for further work

Topic area	Key findings	Suggestions for future research
Antibody category	- Largest number of biomarker entries - Current evidence suggests that it is unlikely that a test on the basis of antibody detection that meets one of the TPPs can be developed - Few prospective validation studies but many studies with case–control design with non-relevant negative control groups	Assessment of antibodies as markers of TB disease progression
Cytokine or other proteins and metabolic marker categories	- Well-populated biomarker category - Many single cytokines assayed - Studies often use existing pre-established multiplex panels with known inflammation markers and there are few unbiased discovery efforts (such as proteomics) that aim to find new host markers - Signatures are more promising than single markers, and signatures that meet the TPP triage performance criteria have been identified - Specificity of general inflammation markers is expected to be lower in patients with other diseases - The challenge is to transfer multiplex assays to simple, affordable POC platforms	- Study host marker signatures in different and relevant patient populations (that is, high-risk groups, such as PLHIV and patients with respiratory symptoms) - Use a comprehensive reference standard that also detects extrapulmonary TB to avoid underdiagnosis - Stop biased biomarker discovery using off-the-shelf cytokine panels that have previously been studied
Host RNA category	- Large proportion of this category is biosignatures - Increasingly popular owing to the availability and decreasing cost of commercial transcriptomics platforms - Probably more relevant for the triage use case owing to the expected moderate specificity - Measuring these signatures at the target triage TPP cost (≤US$2 per test) on a simple platform with currently available technology is not possible	- Validation and confirmation of identified signatures in clinically relevant populations (that is, by enrolling patients with presumptive TB) - Development of targeted PCR assays and related platforms to measure the identified signatures at lower cost and closer to the patient
M. tuberculosis antigen or LAM categories	- Detection of *M. tuberculosis* pathogen markers will probably lead to tests with higher specificity than host markers - *M. tuberculosis* DNA-based diagnosis is a success story (not part of this review) but is too costly and complex, therefore simple antigen detection has potential - Very few 'non-LAM' studies, which suggests that the detection of antigens is difficult - Few promising early-stage LAM studies suggest that LAM is present in general TB populations, including HIV-negative patients - Existing assays detect antigen concentrations in the nanomolar range and might miss patients with lower analyte concentrations	- Development of high-affinity binding reagents to detect *M. tuberculosis* antigens - Establish concentration ranges in easily accessible samples by using highly sensitive detection platforms - Invest in sample preparation approaches and in the development of simple POC platforms that can detect antigens at concentrations of ≤10 pM
Biosignatures	- Out of all the entries in this review, 19% were biosignatures - Of the best-performing entries, 55% were multiple marker biosignatures - Biosignatures seem to have a greater chance to meet the accuracy targets but will be challenging to measure on simple POC platforms	- Fewer 'one-marker' studies and more collaborative efforts, such as the formation of study consortia to assess biomarkers more efficiently - Use biobanks and distribute the same aliquots to several research groups working on different biomarkers so that biomarker results can be analysed in combination
Reporting	- In many publications, it is difficult to find important information regarding study design and timing, patient recruitment and diagnostic performance - Blinding is rarely specified	- Improved reporting is necessary - Following STROBE, STARD or CONSORT checklists, as appropriate, will ensure critical information is included in publications - A patient flow diagram, a table summarizing the characteristics of included patients, a 2×2 table showing index test performance versus the reference standard, ROC curves and a detailed description of included patients, including the reference standard used, are the absolute minimum
'Discovery' studies versus 'validation and confirmation' studies	- Many biomarkers have been reported numerous times in case–control studies using selected patients with TB and healthy controls - Researchers often rely on poorly characterized specimen from non-relevant patient populations - The performance of many biomarkers is measured only with a *P* value and studies are often underpowered -Discovery studies often overemphasize the significance of novel biomarker candidates, and efforts to confirm or validate markers in continuation with previous research are rare	- Consult the literature to avoid repeatedly 'discovering' biomarkers that have already been documented - Consult http://www.bm2dx.org - Use well-characterized samples and reference materials from well-established biobanks (for example, https://www.finddx.org/specimen-banks/ or https://www.tbbiorepository.org/) - Increase the number of follow-up studies of biomarkers that demonstrated promising performance in discovery studies - Utilize prospective study designs and blinding to lower the risk of bias, and enrol diverse and clinically relevant patient populations - Use a comprehensive, culture-based reference standard
TPPs	- Few studies use TPP targets as guiding principles	- Consult the WHO-endorsed TPPs (see Supplementary Methods and https://www.who.int/tb/publications/tpp_report/en/) during the design and analysis phases of studies to focus research and move towards the development of clinically useful tests - Benchmark results against TPP targets

biomarkers is probably simpler. Despite the potential of this category, we found very few studies of *M. tuberculosis* antigen markers. LAM is the *M. tuberculosis* antigen with the strongest evidence in our review (Fig. 3). However, our results confirm the already well-documented[94] low sensitivity of urinary LAM detection methods in patients other than severely immunocompromised PLHIV (Fig. 4b). Recently, three studies[95–97] showed improved clinical sensitivity when using innovative assays. These results suggest that improved assay methods and reagents for LAM detection lead to increased clinical accuracy. However, all three studies need to be further confirmed, and sophisticated assay protocols need to be translated into a POC format to reach TPPs. Currently available POC platforms for antigen detection cannot detect antigen concentrations in the low picomolar range. For future diagnostic development using antigen detection, researchers should consider (1) investing in the generation of high-affinity binding reagents with well-understood epitope specificity to *M. tuberculosis* antigens; (2) using sensitive detection platforms to establish antigen concentration ranges in urine and blood specimens; and (3) supporting the development of simple POC platforms with a low limit of detection.

Strengths of the study. This systematic review, with over 400 included studies, has several strengths. The search strategy was rigorously validated. Clear inclusion and exclusion criteria were decided a priori by all authors and were simple to apply during screening. QUADAS-2, a recognized and validated tool, was applied for quality assessment of all included publications. There were no restrictions on patient populations that were eligible for inclusion in this review; consequently, the list of biomarkers and biosignatures generated is comprehensive with respect to pulmonary and extrapulmonary TB, in adults and children, regardless of HIV or comorbidity status.

Limitations of the study. Owing to the high number of publications, we focused this review on biomarkers for the detection of active TB. However, the distinction between latent infection and active disease is not dichotomous as previously postulated[88], and future reviews could be extended to include markers of disease progression. No formal assessment of publication bias was performed, as existing methods are not helpful for diagnostic accuracy studies[98]. The time frame was limited to 2010–2017; although this may have led to excluding promising biomarkers only reported before this window, we are not aware of any critical misses and confirmed this using selected previously published biomarker reviews[5,14]. As we only used six relevant QUADAS-2 questions to evaluate each study, we could have missed other biases that were not covered by our questions.

Suggestions and future directions. Based on our observations throughout the screening, data extraction and data analysis processes, we have made various key suggestions for future research in Table 1.

The findings from this systematic review are being used as the foundation for a dynamic TB biomarkers database, Bm2Dx.org, developed by FIND[74]. Recognizing that much research in the field is repetitive and therefore somewhat wasteful, the database aims to amplify research impact by facilitating follow-up and validation of identified biomarkers, encouraging research with good reporting and a low potential for bias, and catalysing the formation of research consortia, such as the AE-TBC or the RePORT International Coordinating Center[99]. The FIND sample bank and clinical trials platform will be linked to Bm2Dx, enabling good discovery-level research and biomarker confirmation and validation.

Conclusion

The amount of interest in biomarkers and biosignatures for TB detection is encouraging. However, as only a small proportion of markers is assayed in a clinically relevant manner, progress towards a clinically useful POC test has been slow. We expect this situation to change over the next few years as researchers start to investigate biomarkers in relation to the WHO TPPs and design studies for specific use cases, with more targeted efforts into high-quality follow-up studies. A comprehensive and systematically updated biomarker pipeline, such as that available at Bm2Dx.org, will accelerate the process and help to focus limited resources.

Methods

We conducted a systematic review of non-DNA biomarkers for the diagnosis of active TB in accordance with PRISMA standards (see the PRISMA checklist in Supplementary Methods).

Data sources and search strategy. We searched the academic databases PubMed, EMBASE and Web of Science. Publications were limited to those published between 2010 and 2017 to focus on recent evidence in a rapidly evolving field, written in English and concerning human participants or banked human samples. In the case of PubMed, searches including both medical subject headings (MeSH) and 'text words' were used. In PubMed, the 'English' filter was not used. For EMBASE and Web of Science, 'English' and 'Human' filters were used. The search term was transposed to fit the parameters of each database (Supplementary Methods). The Cochrane Library of Systematic Reviews was searched using the term 'tuberculosis AND biomarker' as title, abstract or keyword entries, without other filters. An 'up-and-down' approach was used to search reviews. Citations identified from reviews were included for further screening. Authors were contacted when relevant for missing diagnostic performance data. Neither patent databases nor the grey literature were searched.

Eligibility criteria. *Time period.* Studies published from 1 January 2010 to 31 December 2017 in English were included to focus on current research in the field, recent methods and to contain the number of studies.

Biomarker types, index tests and sample type. Studies focusing on individual biomarkers or multiple marker biosignatures of host or Mycobacterial origin were included. Studies assessing the performance of a culture-based index test or using imaging-based detection methods (for example, chest X-ray, microscopy or PET-CT, among others) were excluded, as these tests were not considered suitable for translation into a test to fit the WHO TPPs or have been reviewed elsewhere[100]. Studies reporting on DNA biomarkers, including cell-free DNA, were also excluded, as comprehensive reviews of Xpert MTB/RIF[101], TB-LAMP[102] and line probe assays[103] have already been published. Furthermore, studies investigating IFN-γ release assays for active TB detection were excluded owing to existing systematic reviews[104]. Studies that did not report statistical significance or quantitative diagnostic measures (that is, sensitivity and specificity, or area under the curve) of the biomarker's diagnostic performance were excluded. Although the focus of the TPPs is easily accessible samples (such as urine, blood, saliva or exhaled air)[3], the sample type was not used as an eligibility criteria, and studies measuring biomarkers in other, less accessible sample types were included as the detection of a biomarker in these sample types can indicate its presence in more accessible sample types.

Study population. There were no exclusion criteria regarding patient characteristics. Sample size calculations suggest that a solid assessment of biomarker performance should include at least 100 TB-positive and 100 non-TB participants. However, to minimize the risk of excluding promising early-stage research studies, we decided to use a less-stringent inclusion cut-off and included studies with >30 human participants or samples from >30 individuals in total. There were no criteria for the number of participants per group, but studies that included small sample sizes may not be adequately representative of the diversity of patient populations.

Study types. Studies of any design (that is, cohort, case–control and cross-sectional studies or trials) were considered, independent of whether testing was performed on fresh samples in prospective studies or retrospectively on frozen or biobanked samples. Conference proceedings and abstracts were excluded.

Screening and data extraction. *Study screening and selection.* All publications captured by the search were collected in Endnote. After removing duplicates, publications were screened by title and abstract by one reviewer (E.M.) before full-text screening. A second reviewer (T.B.) screened any studies whose appropriateness for inclusion was not immediately obvious.

Data extraction. A Google form was piloted for data extraction and tested with a subgroup of eligible publications before being finalized for use. From each selected publication, information was collected in a standardized manner by one author (either E.M. or B.L.F.-C.) using the form. A list of the fields for data extraction is included in Supplementary Methods. Independent double data entry was

completed for 5% of publications (S.Y., T.B. and E.M.). In instances when a field was unclear (for example, study timing, study design and/or control group), the reviewers conferred to reach consensus. Agreement on extraction results between authors was excellent.

Data quality control and validation. Data quality control was performed by two independent reviewers (T.B. and S.Y.). Details of the quality control protocol are described in Supplementary Methods. For all entries, study design, study timing and biomarker category were independently verified by a second reviewer (T.B.). Data permitting, sensitivity and specificity were recalculated based on 2×2 tables and compared to reported values. We used Wilson's method for estimating the credible 95% confidence interval as this method performs well even when the probability or the sample size is small[105]. In the case of papers that only reported area under the curves or ROC curves without sensitivity, specificity or 2×2 table values, authors were contacted for relevant data.

Assessment of quality and risk of bias of individual studies. All included studies were subject to a quality and bias assessment consisting of six selected questions from the QUADAS-2 tool[106] (Supplementary Methods). The selected questions represent the four domains of QUADAS-2 to assess the risk of bias of studies. As per QUADAS-2 guideline, the selected questions were those deemed most relevant for identifying biases for those studies included in the review. Items were scored as 'yes', 'no' or 'unclear'.

Data synthesis and analysis approach. Data were synthesized and analysed using Excel (Microsoft) and MATLAB (MathWorks) to evaluate the level and quality of evidence for biomarker categories and individual biomarkers. TB diagnostic experts were consulted to develop an analysis approach. Biomarkers that were assayed using a culture-based reference standard and included diagnostic sensitivity and specificity or epidemiological 2×2 tables were included in analysis group α. Those biomarkers that did not fulfil these two criteria were classified into analysis group β and were not considered for any further analysis. Because we expected substantial heterogeneity in markers, design, quality and results, no meta-analysis was planned.

Biomarker entry. A biomarker entry was defined as a unique biomarker or a biosignature (composed of several biomarkers) with performance data. Many eligible publications reported diagnostic performance for more than one biomarker, leading to several biomarker entries per publication. All biomarker entries were categorized into one of two analysis groups (Supplementary Fig. 1).

Negative control groups. In some publications, biomarker performance was reported separately for more than one non-TB control group. According to the TPPs, biomarker-based tests must be assayed in clinically relevant populations, that is, in populations representative of those who would present to a routine health-care setting with a clinical presentation similar to that of TB (for example, due to ORD or other systemic diseases, as in the case of extrapulmonary TB). We followed the example of Steingart et al.[107] and preferentially extracted data for the negative control group with the highest clinical relevance in the following order (most to least relevant): (1) patients who were initially suspected of having TB, but who ultimately were diagnosed with ORD; (2) patients enrolled with known ORD; (3) patients enrolled with other diseases; (4) patients with latent TB infection; (5) individuals without clinical symptoms who had contact with a patient with TB; (6) healthy individuals from endemic countries; and (7) healthy individuals from non-endemic countries. For paediatric or extrapulmonary patients, the preferential control group was either other diseases or ORD. In summary, when studies presented a biomarker's performance using multiple negative control groups, data for the most relevant group were used.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data that support the findings of this study are available on www.Bm2Dx.org and are available from the corresponding author on request. A complete list of the included studies is provided in Supplementary Table 1.

Received: 21 February 2018; Accepted: 18 January 2019;
Published online: 25 February 2019

References

1. *The WHO End TB Strategy* (World Health Organization, 2015).
2. Cazabon, D. et al. Quality of tuberculosis care in high burden countries: the urgent need to address gaps in the care cascade. *Int. J. Infect. Dis.* **56**, 111–116 (2017).
3. Kik, S. V., Denkinger, C. M., Casenghi, M., Vadnais, C. & Pai, M. Tuberculosis diagnostics: which target product profiles should be prioritised? *Eur. Respir. J.* **44**, 537–540 (2014).
4. Biomarkers Definitions Working Group Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin. Pharmacol. Ther.* **69**, 89–95 (2001).
5. Wallis, R. S. et al. Tuberculosis biomarkers discovery: developments, needs, and challenges. *Lancet Infect. Dis.* **13**, 362–372 (2013).
6. *Out of Step 2017: TB Policies in 29 Countries* (MSF, 2017).
7. Cattamanchi, A. et al. Interferon-γ release assays for the diagnosis of latent tuberculosis infection in HIV-infected individuals: a systematic review and meta-analysis. *J. Acquir. Immune Defic. Syndr.* **56**, 230–238 (2011).
8. Diel, R. et al. Interferon-γ release assays for the diagnosis of latent *Mycobacterium tuberculosis* infection: a systematic review and meta-analysis. *Eur. Respir. J.* **37**, 88–99 (2011).
9. Metcalfe, J. Z. et al. Interferon-γ release assays for active pulmonary tuberculosis diagnosis in adults in low- and middle-income countries: systematic review and meta-analysis. *J. Infect. Dis.* **204**, S1120–S1129 (2011).
10. Minion, J. et al. Diagnosing tuberculosis with urine lipoarabinomannan: systematic review and meta-analysis. *Eur. Respir. J.* **38**, 1398–1405 (2011).
11. Shah, M. et al. Urine lateral flow lipoarabinomannan assay for diagnosing active tuberculosis in adults living with HIV. *Cochrane Database Syst. Rev.* **12**, CD011420 (2014).
12. Steingart, K. R. et al. Commercial serological tests for the diagnosis of active pulmonary and extrapulmonary tuberculosis: an updated systematic review and meta-analysis. *PLoS Med.* **8**, e1001062 (2011).
13. Petruccioli, E. et al. Correlates of tuberculosis risk: predictive biomarkers for progression to active tuberculosis. *Eur. Respir. J.* **48**, 1751–1763 (2016).
14. Wallis, R. S. et al. Biomarkers and diagnostics for tuberculosis: progress, needs, and translation into practice. *Lancet* **375**, 1920–1937 (2010).
15. Wallis, R. S. et al. Tuberculosis—advances in development of new drugs, treatment regimens, host-directed therapies, and biomarkers. *Lancet Infect. Dis.* **16**, e34–e46 (2016).
16. *Use of High Burden Country Lists for TB by WHO in the Post-2015 Era* (World Health Organization, 2015).
17. Ben-Selma, W. et al. Evaluation of the diagnostic value of measuring IgG, IgM, and IgA antibodies to mycobacterial A60 antigen in active tuberculosis. *Diagn. Microbiol. Infect. Dis.* **68**, 55–59 (2010).
18. Shete, P. B. et al. Evaluation of antibody responses to panels of *M. tuberculosis* antigens as a screening tool for active tuberculosis in Uganda. *PLoS ONE* **12**, e0180122 (2017).
19. Rekha, R. S. et al. Validation of the ALS assay in adult patients with culture confirmed pulmonary tuberculosis. *PLoS ONE* **6**, e16425 (2011).
20. Khaliq, A. et al. Field evaluation of a blood based test for active tuberculosis in endemic settings. *PLoS ONE* **12**, e0173359 (2017).
21. Ben-Selma, W., Harizi, H. & Boukadida, J. Immunochromatographic IgG/IgM test for rapid diagnosis of active tuberculosis. *Clin. Vaccine Immunol.* **18**, 2090–2094 (2011).
22. Tasbiti, A. H. et al. Evaluation of antigen detection test (chromatographic immunoassay): potential to replace the antibody assay using purified 45-kDa protein for rapid diagnosis of tuberculosis. *J. Clin. Lab. Anal.* **28**, 70–76 (2014).
23. You, X. et al. Evaluation of Rv0220, Rv2958c, Rv2994 and Rv3347c of *Mycobacterium tuberculosis* for serodiagnosis of tuberculosis. *Microb. Biotechnol.* **10**, 604–611 (2017).
24. Yari, S. et al. Purification of modified mycobacterial A60 antigen by affinity chromatography and its use for rapid diagnostic tuberculosis infection. *J. Microbiol. Methods* **87**, 184–188 (2011).
25. Awoniyi, D. O. et al. Detection of a combination of serum IgG and IgA antibodies against selected mycobacterial targets provides promising diagnostic signatures for active TB. *Oncotarget* **8**, 37525–37537 (2017).
26. Goyal, B. et al. Utility of B-cell epitopes based peptides of RD1 and RD2 antigens for immunodiagnosis of pulmonary tuberculosis. *Diagn. Microbiol. Infect. Dis.* **78**, 391–397 (2014).
27. Jacobs, R. et al. Diagnostic potential of novel salivary host biomarkers as candidates for the immunological diagnosis of tuberculosis disease and monitoring of tuberculosis treatment response. *PLoS ONE* **11**, e0160546 (2016).
28. Chegou, N. N. et al. Diagnostic performance of a seven-marker serum protein biosignature for the diagnosis of active TB disease in African primary healthcare clinic attendees with signs and symptoms suggestive of TB. *Thorax* **71**, 785–794 (2016).
29. Jacobs, R. et al. Identification of novel host biomarkers in plasma as candidates for the immunodiagnosis of tuberculosis disease and monitoring of tuberculosis treatment response. *Oncotarget* **7**, 57581–57592 (2016).
30. Sandhu, G. et al. Discriminating active from latent tuberculosis in patients presenting to community clinics. *PLoS ONE* **7**, e38080 (2012).
31. Liang, Y. et al. Evaluation of a whole-blood chemiluminescent immunoassay of IFN-γ, IP-10, and MCP-1 for diagnosis of active pulmonary tuberculosis and tuberculous pleurisy patients. *APMIS* **124**, 856–864 (2016).

32. Chen, T. et al. Cytokine and antibody based diagnostic algorithms for sputum culture-positive pulmonary tuberculosis. *PLoS ONE* **10**, e0144705 (2015).
33. Xu, D. et al. Serum protein S100A9, SOD3, and MMP9 as new diagnostic biomarkers for pulmonary tuberculosis by iTRAQ-coupled two-dimensional LC–MS/MS. *Proteomics* **15**, 58–67 (2015).
34. Jiang, T. T. et al. Serum amyloid A, protein Z, and C4b-binding protein β chain as new potential biomarkers for pulmonary tuberculosis. *PLoS ONE* **12**, e0173304 (2017).
35. Deng, C. W. et al. Establishing a serologic decision tree model of extrapulmonary tuberculosis by MALDI–TOF MS analysis. *Diagn. Microbiol. Infect. Dis.* **71**, 144–150 (2011).
36. Sharma, D., Dhiman, P., Rajendiran, S., Ravikumar, N. & Krishna, M. H. Osteoarticular tuberculosis: in search of new biomarkers. *Eur. J. Orthop. Surg. Traumatol.* **6**, 195–200 (2015).
37. Kalantri, Y., Hemvani, N. & Chitnis, D. S. Evaluation of real-time polymerase chain reaction, interferon-γ, adenosine deaminase, and immunoglobulin A for the efficient diagnosis of pleural tuberculosis. *Int. J. Infect. Dis.* **15**, e226–e231 (2011).
38. Liu, Y. Y. et al. A combination of the QuantiFERON-TB Gold In-Tube assay and the detection of adenosine deaminase improves the diagnosis of tuberculous pleural effusion. *Emerg. Microbes Infect.* **5**, e83 (2016).
39. Sutherland, J. S. et al. Highly accurate diagnosis of pleural tuberculosis by immunological analysis of the pleural effusion. *PLoS ONE* **7**, e30324 (2012).
40. Kaforou, M. et al. Detection of tuberculosis in HIV-infected and -uninfected African adults using whole blood RNA expression signatures: a case–control study. *PLoS Med.* **10**, e1001538 (2013).
41. Laux da Costa, L. et al. A real-time PCR signature to discriminate between tuberculosis and other pulmonary diseases. *Tuberculosis* **95**, 421–425 (2015).
42. Francisco, N. M. et al. Diagnostic accuracy of a selected signature gene set that discriminates active pulmonary tuberculosis and other pulmonary diseases. *J. Infect.* **75**, 499–510 (2017).
43. Lawn, S. D., Kerkhoff, A. D., Vogt, M. & Wood, R. Diagnostic accuracy of a low-cost, urine antigen, point-of-care screening assay for HIV-associated pulmonary tuberculosis before antiretroviral therapy: a descriptive study. *Lancet Infect. Dis.* **12**, 201–209 (2012).
44. Tiwari, D. et al. Fast and efficient detection of tuberculosis antigens using liposome encapsulated secretory proteins of *Mycobacterium tuberculosis*. *J. Microbiol. Immunol. Infect.* **50**, 189–198 (2017).
45. Mourão, M. P. B. et al. Direct detection of *Mycobacterium tuberculosis* in sputum: a validation study using solid phase extraction-gas chromatography–mass spectrometry. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **1012–1013**, 50–54 (2016).
46. Dang, N. A. et al. Direct detection of *Mycobacterium tuberculosis* in sputum using combined solid phase extraction-gas chromatography–mass spectrometry. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **986–987**, 115–122 (2015).
47. Szewczyk, R., Kowalski, K., Janiszewska-Drobinska, B. & Druszczynska, M. Rapid method for *Mycobacterium tuberculosis* identification using electrospray ionization tandem mass spectrometry analysis of mycolic acids. *Diagn. Microbiol. Infect. Dis.* **76**, 298–305 (2013).
48. Shui, G. et al. Mycolic acids as diagnostic markers for tuberculosis case detection in humans and drug efficacy in mice. *EMBO Mol. Med.* **4**, 27–37 (2012).
49. Reither, K. et al. Evaluation of Diagnos TB AG, a flow-through immunoassay for rapid detection of pulmonary tuberculosis. *Int. J. Tuberc. Lung Dis.* **14**, 238–240 (2010).
50. Khurshid, S., Khalid, R., Afzal, M. & Waheed Akhtar, M. Truncation of PstS1 antigen of *Mycobacterium tuberculosis* improves diagnostic efficiency. *Tuberculosis* **93**, 654–659 (2013).
51. Ben-Selma, W. et al. Rapid detection of immunoglobulin G against *Mycobacterium tuberculosis* antigens by two commercial ELISA kits. *Int. J Tuberc. Lung Dis.* **14**, 841–846 (2010).
52. He, X. Y. et al. Assessment of five antigens from *Mycobacterium tuberculosis* for serodiagnosis of tuberculosis. *Clin. Vaccine Immunol.* **18**, 565–570 (2011).
53. Chung, W. et al. Serum CXCR3 ligands as biomarkers for the diagnosis and treatment monitoring of tuberculosis. *Int. J. Tuberc. Lung Dis.* **19**, 1476–1484 (2015).
54. Chung, W. Y. et al. A TB antigen-stimulated CXCR3 ligand assay for the diagnosis of active pulmonary TB. *Chest* **146**, 283–291 (2014).
55. Yang, Q. T. et al. IP-10 and MIG are compartmentalized at the site of disease during pleural and meningeal tuberculosis and are decreased after antituberculosis treatment. *Clin. Vaccine Immunol.* **21**, 1635–1644 (2014).
56. You, E., Kim, M. H., Lee, W. I. & Kang, S. Y. Evaluation of IL-2, IL-10, IL-17 and IP-10 as potent discriminative markers for active tuberculosis among pulmonary tuberculosis suspects. *Tuberculosis* **99**, 100–108 (2016).
57. Sun, Q., Wei, W. & Sha, W. Potential role for *Mycobacterium tuberculosis* specific IL-2 and IFN-γ responses in discriminating between latent infection and active disease after long-term stimulation. *PLoS ONE* **11**, e0166501 (2016).
58. Awoniyi, D. O. et al. Evaluation of cytokine responses against novel Mtb antigens as diagnostic markers for TB disease. *J. Infect.* **73**, 219–230 (2016).
59. Drain, P. K. et al. Diagnostic accuracy and clinical role of rapid C-reactive protein testing in HIV-infected individuals with presumed tuberculosis in South Africa. *Int. J. Tuberc. Lung Dis.* **18**, 20–26 (2014).
60. Meldau, R. et al. Comparison of same day diagnostic tools including Gene Xpert and unstimulated IFN-γ for the evaluation of pleural tuberculosis: a prospective cohort study. *BMC Pulm. Med.* **14**, 58 (2014).
61. Ogata, Y. et al. Is adenosine deaminase in pleural fluid a useful marker for differentiating tuberculosis from lung cancer or mesothelioma in Japan, a country with intermediate incidence of tuberculosis? *Acta Med. Okayama* **65**, 259–263 (2011).
62. Qi, Z. J., Yu, H., Zhang, J. & Li, C. S. Presepsin as a novel diagnostic biomarker for differentiating active pulmonary tuberculosis from bacterial community acquired pneumonia. *Clin. Chim. Acta* **478**, 152–156 (2018).
63. Anderson, S. T. et al. Diagnosis of childhood tuberculosis and host RNA expression in Africa. *N. Engl. J. Med.* **370**, 1712–1723 (2014).
64. Cai, Y. et al. Increased complement C1q level marks active disease in human tuberculosis. *PLoS ONE* **9**, e92340 (2014).
65. Jacobs, R. et al. Host biomarkers detected in saliva show promise as markers for the diagnosis of pulmonary tuberculosis disease and monitoring of the response to tuberculosis treatment. *Cytokine* **81**, 50–56 (2016).
66. Hanafiah, K. H. et al. Serological biomarker screening and host factor analysis elucidating immune response heterogeneity in active pulmonary tuberculosis. *Trop. Biomed.* **34**, 556–569 (2017).
67. Yoon, C. et al. Point-of-care C-reactive protein-based tuberculosis screening for people living with HIV: a diagnostic accuracy study. *Lancet Infect. Dis.* **17**, 1285–1292 (2017).
68. Singh, N. et al. Diagnosis of pulmonary and extrapulmonary tuberculosis based on detection of mycobacterial antigen 85B by immuno-PCR. *Diagn. Microbiol. Infect. Dis.* **83**, 359–364 (2015).
69. Kolk, A. H. J. et al. Breath analysis as a potential diagnostic tool for tuberculosis. *Int. J. Tuberc. Lung Dis.* **16**, 777–782 (2012).
70. Phillips, M. et al. Breath biomarkers of active pulmonary tuberculosis. *Tuberculosis* **90**, 145–151 (2010).
71. Kolk, A. et al. Electronic-nose technology using sputum samples in diagnosis of patients with tuberculosis. *J. Clin. Microbiol.* **48**, 4235–4238 (2010).
72. Lim, S. H. et al. Rapid diagnosis of tuberculosis from analysis of urine volatile organic compounds. *ACS Sensors* **1**, 852–856 (2016).
73. Poste, G. Bring on the biomarkers. *Nature* **469**, 156–157 (2011).
74. Yerlikaya, S., Broger, T., MacLean, E., Pai, M. & Denkinger, C. M. A tuberculosis biomarker database: the key to novel TB diagnostics. *Int. J. Infect. Dis.* **56**, 253–257 (2017).
75. Goodman, S. N. Toward evidence-based medical statistics. 1: The *P* value fallacy. *Ann. Intern. Med.* **130**, 995–1004 (1999).
76. Lijmer, J. G. et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* **282**, 1061–1066 (1999).
77. Rutjes, A. W. et al. Evidence of bias and variation in diagnostic accuracy studies. *CMAJ* **174**, 469–476 (2006).
78. Tessema, B. et al. FIND Tuberculosis Strain Bank: a resource for researchers and developers working on tests to detect *Mycobacterium tuberculosis* and related drug resistance. *J. Clin. Microbiol.* **55**, 1066–1073 (2017).
79. Vincent, V. et al. The TDR Tuberculosis Strain Bank: a resource for basic science, tool development and diagnostic services. *Int. J. Tuberc. Lung Dis.* **16**, 24–31 (2012).
80. *Commercial Serodiagnostic Tests for Diagnosis of Tuberculosis — Policy Statement* (World Health Organization, 2011).
81. *Target Product Profile: Test for Incipient Tuberculosis* (FIND, 2017).
82. Yoon, C. et al. Diagnostic accuracy of C-reactive protein for active pulmonary tuberculosis: a meta-analysis. *Int. J. Tuberc. Lung Dis.* **21**, 1013–1019 (2017).
83. Kapasi, A. J., Dittrich, S., Gonzalez, I. J. & Rodwell, T. C. Host biomarkers for distinguishing bacterial from non-bacterial causes of acute febrile illness: a comprehensive review. *PLoS ONE* **11**, e0160278 (2016).
84. Black, S., Kushner, I. & Samols, D. C-reactive protein. *J. Biol. Chem.* **279**, 48487–48490 (2004).
85. Camelo, L. V. et al. Life course socioeconomic position and C-reactive protein: mediating role of health-risk behaviors and metabolic alterations. The Brazilian Longitudinal Study of Adult Health (ELSA-Brasil). *PLoS ONE* **9**, e108426 (2014).
86. Geluk, A. & Corstjens, P. CRP: tell-tale biomarker or common denominator? *Lancet Infect. Dis.* **17**, 1225–1227 (2017).

87. MacLean, E. & Broger, T. A 10-gene signature for the diagnosis and treatment monitoring of active tuberculosis using a molecular interaction network approach. *EBioMedicine* **16**, 22–23 (2017).
88. Drain, P. K. et al. Incipient and subclinical tuberculosis: a clinical review of early stages and progression of infection. *Clin. Microbiol. Rev.* **31**, e00021-18 (2018).
89. Sweeney, T. E., Braviak, L., Tato, C. M. & Khatri, P. Genome-wide expression for diagnosis of pulmonary tuberculosis: a multicohort analysis. *Lancet Respir. Med.* **4**, 213–224 (2016).
90. Zak, D. E. et al. A blood RNA signature for tuberculosis disease risk: a prospective cohort study. *Lancet* **387**, 2312–2322 (2016).
91. *Diagnostics Pipeline Tracker: Tuberculosis* (FIND, 2017).
92. Schumacher, S. G. & Denkinger, C. M. Diagnostic test for incipient tuberculosis: a step forward, many more to go. *Am. J. Respir. Crit. Care Med.* **197**, 1106–1107 (2018).
93. Boehme, C. C. et al. Feasibility, diagnostic accuracy, and effectiveness of decentralised use of the Xpert MTB/RIF test for diagnosis of tuberculosis and multidrug resistance: a multicentre implementation study. *Lancet* **377**, 1495–1505 (2011).
94. Shah, M. et al. Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults. *Cochrane Database Syst. Rev.* **5**, CD011420 (2016).
95. Paris, L. et al. Urine lipoarabinomannan glycan in HIV-negative patients with pulmonary tuberculosis correlates with disease severity. *Sci. Transl Med.* **9**, eaal2807 (2017).
96. Amin, A. G. et al. Detection of lipoarabinomannan in urine and serum of HIV-positive and HIV-negative TB suspects using an improved capture-enzyme linked immuno absorbent assay and gas chromatography/mass spectrometry. *Tuberculosis (Edinb.)* **111**, 178–187 (2018).
97. Sigal, G. B. et al. A novel sensitive immunoassay targeting the MTX-lipoarabinomannan epitope meets the WHO's performance target for tuberculosis diagnosis. *J. Clin. Microbiol.* **56**, e01338-18 (2018).
98. Macaskill, P., Gatsonis, C., Deeks, J., Harbord, R. & Takwoingi, Y. in *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* (The Cochrane Collaboration, 2010); https://methods.cochrane.org/sites/methods.cochrane.org.sdt/files/public/uploads/Chapter%2010%20-%20Version%201.0.pdf
99. Hamilton, C. D. et al. RePORT International: advancing tuberculosis biomarker research through global collaboration. *Clin. Infect. Dis.* **61**, S155–S159 (2015).
100. Piccazzo, R., Paparo, F. & Garlaschi, G. Diagnostic accuracy of chest radiography for the diagnosis of tuberculosis (TB) and its role in the detection of latent TB infection: a systematic review. *J. Rheumatol. Suppl.* **91**, 32–40 (2014).
101. Steingart, K. R. et al. Xpert® MTB/RIF assay for pulmonary tuberculosis and rifampicin resistance in adults. *Cochrane Database Syst. Rev.* **1**, CD009593 (2014).
102. Nagai, K. et al. Diagnostic test accuracy of loop-mediated isothermal amplification assay for *Mycobacterium tuberculosis*: systematic review and meta-analysis. *Sci. Rep.* **6**, 39090 (2016).
103. Nathavitharana, R. R. et al. Accuracy of line probe assays for the diagnosis of pulmonary and multidrug-resistant tuberculosis: a systematic review and meta-analysis. *Eur. Respir. J.* **49**, 1601075 (2017).
104. Rangaka, M. X. et al. Predictive value of interferon-γ release assays for incident active tuberculosis: a systematic review and meta-analysis. *Lancet Infect. Dis.* **12**, 45–55 (2012).
105. Wilson, E. B. Probable inference, the law of succession, and statistical inference. *J. Am. Stat. Assoc.* **22**, 209–212 (1927).
106. Whiting, P. F. et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann. Intern. Med.* **155**, 529–536 (2011).
107. Steingart, K. R. et al. Commercial serological antibody detection tests for the diagnosis of pulmonary tuberculosis: a systematic review. *PLoS Med.* **4**, e254 (2007).

Acknowledgements

We thank T. Togun (data quality checking), S. Huddart (graphics work) and R. Wyss (data curation). We also thank the New Diagnostic Working Group Biomarkers Task Force (for feedback on the design of this systematic review and on the manuscript). The New Diagnostics Working Group: D. M. Cirillo (San Raffaele Scientific Institute, Italy), C.M.D. (FIND, Switzerland), M. Doherty (SSI, Denmark), J. L. Gardiner (BMGF, USA), M. L. Gennaro (Rutgers, USA), S. A. Joosten (Leiden University, The Netherlands), M. Kaforou (Imperial College London, UK), E.M. (McGill University, Canada), P. Nahid (UCSF, USA), M.P. (McGill University, Canada), M. Schito (C-PATH, USA), T. J. Scriba (University of Cape Town, South Africa), R. S. Wallis (Aurum Institute, South Africa), G. Walzl (Stellenbosch University, South Africa), S.Y. (FIND, Switzerland), A. Zumla (UCL, UK) and T.B. (FIND, Switzerland). The work was funded by the Dutch Ministry of Foreign Affairs and the New Diagnostics Working Group of the Stop TB Partnership.

Author contributions

E.M., T.B., M.P. and C.M.D. designed and conceptualized the study. E.M. screened all of the studies. E.M. and B.L.F.-C. performed the primary data extraction. T.B., B.L.F.-C. and S.Y. validated the data. E.M. and T.B. created the analysis plan. T.B. performed the formal data analysis. E.M. and T.B. wrote the original draft of the manuscript. B.L.F.-C., S.Y., M.P. and C.M.D. provided critical editing and review.

Competing interests

T.B., S.Y., B.L.F.-C. and C.M.D. are employed by FIND. FIND is a not-for-profit foundation that supports the evaluation of publicly prioritized TB assays and the implementation of WHO-approved (guidance and prequalification) assays using donor grants. FIND has product evaluation agreements with several private sector companies that design diagnostics for TB and other diseases. These agreements strictly define FIND's independence and neutrality vis-a-vis the companies whose products get evaluated and describe roles and responsibilities. M.P. serves on the WHO SAGE IVD Group and is a member of the Scientific Advisory Committee of FIND. M.P. and E.M. have no industry or financial conflicts.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41564-019-0380-2.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to C.M.D.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature research

Corresponding author(s): Claudia M. Denkinger

Last updated by author(s): Jan 8, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	This is a systematic review. We searched the academic databases PubMed, EMBASE, and Web of Science as described.
Data analysis	Data were synthesized and analysed using Excel (Microsoft, USA) and MATLAB (Mathworks, USA)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data that support the findings of this study are available on www.Bm2Dx.org

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	This is a systematic review and sample size of included studies are reported in detail.
Data exclusions	Eligibility and exclusion criteria for this systematic review are described in detail in the methods section: Time Period Studies published from 1 January 2010 to 31 December 2017 in English were included to focus on current research in the field, recent methods, and to contain the number of studies. Biomarker types, index tests, and sample type Studies focusing on individual biomarkers or multi-marker biosignatures of host or Mycobacterial origin were included. Studies assessing the performance of a culture-based index test, or using imaging-based detection methods (e.g. chest X-ray, microscopy, PET-CT, etc.) were excluded as these tests were not considered suitable for translation into a test to fit the WHO TPPs or have been reviewed elsewhere. Studies reporting on DNA biomarkers, including cell-free DNA, were also excluded, as comprehensive reviews of Xpert MTB/RIF, TB-LAMP, and line probe assays have already been published. Further, studies investigating IGRAs for active TB detection were excluded due to existing systematic reviews. Studies that did not report statistical significance or quantitative diagnostic measures (i.e. sensitivity and specificity, or AUC) of the biomarker's diagnostic performance were excluded. Although the focus of the TPPs are easily accessible samples (such as urine, blood, saliva, or exhaled air), the sample type was not used as an eligibility criteria and studies measuring biomarkers in other, less accessible sample types were included since the detection of a biomarker in these sample types can indicate its presence in more accessible sample types. Study population There were no exclusion criteria regarding patient characteristics. Sample size calculations suggest that a solid assessment of biomarker performance should include at least 100 TB positive and 100 Non-TB participants. However, to minimize the risk of excluding promising early-stage research studies we decided to use a less-stringent inclusion cut-off and included studies with more than 30 human participants or samples from more than 30 individuals in total. There were no criteria for number of participants per group, but studies that included small sample sizes may not be adequately representative of the diversity of patient populations. Study types Studies of any design (i.e. cohort, case-control, cross-sectional studies, or trials) were considered, independent of whether testing was performed on fresh samples in prospective studies or retrospectively on frozen or biobanked samples. Conference proceedings and abstracts were excluded.
Replication	We followed "Preferred Reporting Items for Systematic Reviews and Meta-Analyses" (PRISMA) guidelines and attached the PRISMA-Checklist in the Supplementary Methods (including search term). This ensures that findings can be reproduced. Independent double data entry was completed for 5% of publications to ensure consistency and that findings can be replicated.
Randomization	Randomization is not relevant for this systematic review.
Blinding	Blinding is not relevant for this systematic review.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	In this systematic review we reviewed the population characteristics for all included studies and report on it.
Recruitment	In this systematic review we reviewed the recruitment timing for all included studies and report on it.

Ethics oversight	Not relevant for systematic reviews.

Note that full information on the approval of the study protocol must also be provided in the manuscript.