Check for updates

SYSTEMATIC REVIEW

REVISED Systematic review of diagnostic and prognostic host blood transcriptomic signatures of tuberculosis disease in people living with HIV [version 2; peer review: 2 approved]

Simon C Mendelsohn, Savannah Verhage, Humphrey Mulenga, Thomas J Scriba, Mark Hatherill 📛

South African Tuberculosis Vaccine Initiative, Institute of Infectious Disease and Molecular Medicine and Division of Immunology, Department of Pathology, University of Cape Town, Cape Town, Western Cape, 7935, South Africa

Abstract

Background
HIV-associated tuberculosis (TB) has high mortality; however, current triage and prognostic tools offer poor sensitivity and specificity, respectively. We conducted a systematic review of diagnostic and prognostic host-blood transcriptomic signatures of TB in people living with HIV (PLHIV).

Methods
We systematically searched online databases for studies published in English between 1990-2020. Eligible studies included PLHIV of any age in test or validation cohorts, and used microbiological or composite reference standards for TB diagnosis. Inclusion was not restricted by setting or participant age. Study selection, quality appraisal using the QUADAS-2 tool, and data extraction were conducted independently by two reviewers. Thereafter, narrative synthesis of included studies, and comparison of signatures performance, was performed.

Results
We screened 1,580 records and included 12 studies evaluating 31 host-blood transcriptomic signatures in 10 test or validation cohorts of PLHIV that differentiated individuals with TB from those with HIV alone, latent *Mycobacterium tuberculosis* infection, or other diseases (OD). Two (2/10; 20%) cohorts were prospective (29 TB cases; 51 OD) and 8 (80%) case-control (353 TB cases; 606 controls) design. All cohorts (10/10) were recruited in Sub-Saharan Africa and 9/10 (90%) had a high risk of bias. Ten signatures (10/31; 32%) met minimum WHO Target Product Profile (TPP) criteria for TB triage tests. Only one study (1/12; 8%) evaluated prognostic performance of a transcriptomic signature for progression to TB in PLHIV, which did not meet the minimum WHO prognostic TPP.

Conclusions

Generalisability of reported findings is limited by few studies enrolling PLHIV, limited geographical diversity, and predominantly case-control design, which also introduces spectrum bias. New prospective cohort studies are needed that include PLHIV and are conducted in diverse settings. Further research exploring the effect of HIV clinical, virological, and immunological factors on diagnostic performance is necessary for development and implementation of TB transcriptomic signatures in PLHIV.

Keywords

Host, blood, diagnostic, prognostic, gene, signatures, tuberculosis, HIV

Corresponding author: Mark Hatherill (mark.hatherill@uct.ac.za)

Introduction

There were an estimated 703,000 HIV-associated incident tuberculosis (TB) cases in 2021 however only 368,600 (52%) were notified, with a resultant case fatality rate of 27%[1]. Earlier diagnosis and initiation of treatment, or disease prevention through targeted short-course TB preventive therapy (TPT), may reduce this burden. However, we lack adequate TB mass screening tools to direct confirmatory testing or prognostic tools to guide preventive therapy in the outpatient or community setting. Symptom screening, the most widely used TB triage tool, has low specificity in antiretroviral therapy (ART-) naïve and low sensitivity in ART-experienced people living with HIV (PLHIV)[2]. The addition of chest radiography improves sensitivity, at a cost of reduced specificity[2]. With almost three-quarters of the 38 million PLHIV globally now receiving ART[3], new tools should be efficacious in this group.

The WHO currently recommends that PLHIV with a positive or unknown tuberculin skin test (TST) result should receive TPT, if active TB has been excluded[4]. There is strong evidence to support such an approach[5]. However, TST and interferon-γ release assay (IGRA) reflect a memory T-cell response following *Mycobacterium tuberculosis* (Mtb) exposure (sensitisation) and not necessarily ongoing infection. In TB-endemic countries with high rates of Mtb transmission and exposure, these tests have limited utility for guiding TPT[6,7]. In addition, loss or dysfunction of Mtb-specific memory T-cells among immunocompromised PLHIV with low CD4 cell counts results in lower IGRA positivity and may reduce sensitivity for predicting progression to disease[8,9]. There is also limited evidence regarding repeat courses of TPT among PLHIV; a recent study demonstrated that universal retreatment after one year did not provide additional benefit[10].

Biomarker-guided treatment has been proposed to target therapy to those that need it most, reducing unnecessary pill burden, drug interactions, and side effects in individuals, and increasing efficacy and cost-effectiveness of mass screening[11,12]. Host-response blood transcriptomic signatures can identify those with active TB and those who are progressing to disease[13–15]. Performance of most signatures in adults without HIV meet at least one of the minimum World Health Organization (WHO) Target Product Profile (TPP) TB triage test performance criteria (sensitivity 90% and specificity 70%) for diagnosing prevalent TB[16,17]. Several signatures have been shown to meet minimum prognostic benchmarks (sensitivity 75% and specificity 75%)[18] for short-term prediction of progression to TB disease within six months of testing[19–21]. However, only a couple of signatures, Roe1 and Roe3[22,23], meet these criteria through 12 months of follow-up for progression[21]. Transcriptomic

biomarkers selected for advancement through the diagnostics pipeline for development as point-of-care assays should also perform well in PLHIV. We systematically reviewed the published literature on host-response blood transcriptomic biomarkers for diagnosing prevalent and predicting progression to incident TB disease in PLHIV, and compared performance to the WHO TPP criteria.

Methods

Protocol and registration
This review is reported in line with the Preferred Reporting Items for Systematic reviews and Meta-Analysis of Diagnostic Test Accuracy Studies (PRISMA-DTA)[24] recommendations (Table 1). The systematic review protocol was registered with the International Prospective Register of Systematic Reviews (PROSPERO) on 02 January 2021 with registration number CRD42021224155 and published in the *BMJ Open*[25].

Eligibility criteria
We considered cross-sectional and case-control studies, prospective and retrospective cohort studies, and randomised control trials evaluating diagnostic and/or prognostic performance of human host-blood transcriptomic signatures of TB (index tests). Eligible studies included PLHIV in the signature test and/or validation cohorts. Studies that only reported signature discovery cohort performance, or treatment response and failure monitoring cohorts, were not considered. PLHIV of all ages, ethnicities, and in all settings were considered. Studies which did not report any measures of signature performance (sensitivity and specificity, or reported data which enable the reconstruction of a two-by-two table for test accuracy calculation for PLHIV), did not clearly state the case definition of TB disease, did not report primary data, or did not independently report signature performance in PLHIV, were excluded.

Endpoint definitions
The primary TB disease endpoint (target condition) was defined by a positive microbiological test, such as mycobacterial culture or the Xpert MTB/RIF assay (reference standards), in sputum or other bodily fluid sample. Microbiologically-confirmed extra-pulmonary TB disease was also considered. The secondary TB disease endpoint was defined by non-microbiologically-confirmed, presumptive TB diagnosed via composite clinical features. TB disease diagnosed within one month of the index test was presumed to be prevalent disease (diagnostic studies). Prognostic studies were defined as prospective studies in which participants were followed up for progression to incident TB disease with measurement of a transcriptomic signature from blood samples collected at enrolment. Eligible studies included healthy individuals, latent Mtb-infected individuals, or individuals with other respiratory or systemic diseases as a control group. Latent Mtb infection was defined by a positive TST or IGRA.

Search strategy and information sources
We systematically searched PubMed (*MEDLINE*), *WOS Core Collection, Biological Abstracts,* and *SciELO Citation Index (via Web of Science), Africa-Wide Information* and *General Science Abstracts (via EBSCOhost), Scopus,* and *Cochrane Central Register of Controlled Trials* databases for full-text articles published in English between 1 January 1990 and 31

Table 1. PRISMA for Diagnostic Test Accuracy 2018 checklist. This checklist has been adapted from McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, The PRISMA-DTA Group (2018). Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. JAMA. 2018 Jan 23;319(4):388–396. doi: 10.1001/jama.2017.19163.

Section/topic	#	PRISMA-DTA Checklist Item	Reported on page #
TITLE / ABSTRACT			
Title	1	Identify the report as a systematic review (+/- meta-analysis) of diagnostic test accuracy (DTA) studies.	Page 1
Abstract	2	Abstract: See PRISMA-DTA for abstracts.	Page 1
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	Page 3
Clinical role of index test	D1	State the scientific and clinical background, including the intended use and clinical role of the index test, and if applicable, the rationale for minimally acceptable test accuracy (or minimum difference in accuracy for comparative design).	Page 3
Objectives	4	Provide an explicit statement of question(s) being addressed in terms of participants, index test(s), and target condition(s).	Page 3
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	Page 3
Eligibility criteria	6	Specify study characteristics (participants, setting, index test(s), reference standard(s), target condition(s), and study design) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	Page 3
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	Page 3
Search	8	Present full search strategies for all electronic databases and other sources searched, including any limits used, such that they could be repeated.	Page 3
Study selection	9	State process for selecting studies (i.e. screening, eligibility, included in systematic review, and, if applicable, included in meta-analysis).	Page 4
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	Page 4
Definitions for data extraction	11	Provide definitions used in data extraction and classifications of target condition(s), index test(s), reference standard(s) and other characteristics (e.g. study design, clinical setting).	Page 3
Risk of bias & applicability	12	Describe methods used for assessing risk of bias in individual studies and concerns regarding the applicability to the review question.	Page 6
Diagnostic accuracy measures	13	State the principal diagnostic accuracy measure(s) reported (e.g. sensitivity, specificity) and state the unit of assessment (e.g. per-patient, per-lesion).	Page 5
Synthesis of results	14	Describe methods of handling data, combining results of studies and describing variability between studies. This could include, but is not limited to: a) handling of multiple definitions of target condition. b) handling of multiple thresholds of test positivity, c) handling multiple index test readers, d) handling of indeterminate test results, e) grouping and comparing tests, f) handling of different reference standards	Page 5
Meta-analysis	D2	Report the statistical methods used for meta-analyses, if performed.	NA
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating if pre-specified.	NA
RESULTS			
Study selection	17	Provide numbers of studies screened, assessed for eligibility, included in the review (and included in meta-analysis, if applicable) with reasons for exclusions at each stage, ideally with a flow diagram.	Page 6 Figure 1

Section/topic	#	PRISMA-DTA Checklist Item	Reported on page #
Study characteristics	18	For each included study provide citations and present key characteristics including: a) participant characteristics (presentation, prior testing), b) clinical setting, c) study design, d) target condition definition, e) index test, f) reference standard, g) sample size, h) funding sources	Table 2, Table 3
Risk of bias and applicability	19	Present evaluation of risk of bias and concerns regarding applicability for each study.	Page 14, Figure 2
Results of individual studies	20	For each analysis in each study (e.g. unique combination of index test, reference standard, and positivity threshold) report 2x2 data (TP, FP, FN, TN) with estimates of diagnostic accuracy and confidence intervals, ideally with a forest or receiver operator characteristic (ROC) plot.	Figure 3–Figure 6
Synthesis of results	21	Describe test accuracy, including variability; if meta-analysis was done, include results and confidence intervals.	Pages 14–17
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression; analysis of index test: failure rates, proportion of inconclusive results, adverse events).	NA
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence.	Pages 17–19
Limitations	25	Discuss limitations from included studies (e.g. risk of bias and concerns regarding applicability) and from the review process (e.g. incomplete retrieval of identified research).	Pages 17–19
Conclusions	26	Provide a general interpretation of the results in the context of other evidence. Discuss implications for future research and clinical practice (e.g. the intended use and clinical role of the index test).	Pages 17–19
FUNDING			
Funding	27	For the systematic review, describe the sources of funding and other support and the role of the funders.	Page 2

December 2020 using Medical Subject Headings (MeSH) and keyword search terms for "Diagnosis", "Messenger RNA", "Biomarkers/blood", "Tuberculosis", and "HIV". The search strategy, including publication date range, were prespecified and published in a systematic review protocol[25]. We reviewed reference lists of eligible articles and performed forward citation tracking using *Science Citation Index* (via *Web of Science*) to identify further articles and reports missed by the electronic database search[26].

Study selection and data collection
Two reviewers (SCM and SV) independently conducted the literature search and screened the search outputs for potential inclusion using EndNote bibliographic software to manage references, as previously described[27]. After removal of duplicates, the selection process included an initial screening of titles and abstracts for relevance, followed by full text review for eligibility. The two reviewers resolved any disagreements or uncertainties by discussion. Data elements of included studies were then independently extracted by the two reviewers. Corresponding authors of potentially eligible studies were contacted to provide deidentified participant-level data to reconstruct two-by-two tables or summary performance data for the PLHIV subgroup. Studies without summary or participant-level data available for the PLHIV subgroup were excluded.

Data analysis
We performed a narrative synthesis of the eligible study cohorts and signatures, including study design, cohort and signature characteristics, and diagnostic and prognostic performance of signatures stratified by study control groups (healthy, latent-Mtb infected, or other disease), and diagnostic reference standards (microbiological or composite clinical). Studies and cohorts were designated by the first author name and year of publication (e.g. Author2019a) and signatures by first author and number of transcripts (e.g. Author11). Signature area under the curve (AUC), sensitivity, and specificity were summarised in forest plots (R *forestplot* package[28]). For studies with available participant-level data for the PLHIV subgroup, we were able to recalculate AUC (R *pROC* package[29]), and benchmark sensitivity and specificity against the WHO TPP minimum performance criteria for a triage (70% specificity and 90% sensitivity)[16] or prognostic (75% specificity and 75% sensitivity)[18] test. 95% Confidence intervals for AUCs and sensitivity and specificity, were calculated using the DeLong[30] and Wilson binomial[31] methods, respectively. For studies in which participant-level data were not available, we report summary AUC, sensitivity, and specificity estimates, and 95% confidence intervals, for the PLHIV subgroup as published in the original papers. Most of these estimates were not specifically benchmarked against the WHO TPP minimum performance criteria for a triage test.

Risk of bias, applicability, and quality of evidence

The methodological quality and applicability concerns of included studies was assessed by the two reviewers using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool[32] and graphically represented using traffic-light plots (R *robvis* package[33]). Risk of bias and applicability concerns for individual study was evaluated in four domains relating to (1) patient selection, (2) measurement of the index test, (3) measurement of the reference standard, and (4) study flow and timing of investigational and diagnostic procedures. Risks of bias are reported within each domain of each study as low risk (no risks of bias identified), some concerns (one risk or unclear risk of bias identified), or high risk (more than one risk or unclear risk of bias identified). Overall risk of bias for each study was reported as low risk (no risks of bias identified in any domains), some concerns (some concerns identified in one domain), or high risk (some concerns identified in two or more domains or high risk of bias in any domain). We assessed the cumulative quality of evidence synthesised by the systematic review using the "Grading of Recommendations Assessment, Development and Evaluation" (GRADE) approach[34] with classification based on study design and limitations, indirectness, inconsistency, imprecision, and publication bias[35,36].

Results

Search results

We performed the literature search in January 2021, identifying 1,580 unique records published between 1 January 1990 and 31 December 2020, of which 98 full-text articles were assessed for eligibility, and 12 studies[13,22,37–47] met all criteria for inclusion (Figure 1). The main reasons for study exclusion were absence of PLHIV in study cohorts (n=29), absence of an independent test or external validation cohort which included PLHIV (n=20), and inappropriate index test (n=16) or study design (n=10). Nine of 10 studies excluded for inappropriate design were commentaries or reviews. In addition, deidentified participant-level data to reconstruct two-by-two tables, or summary performance data for PLHIV, were not available and no data were received from corresponding authors for 9 records[19,48–55].
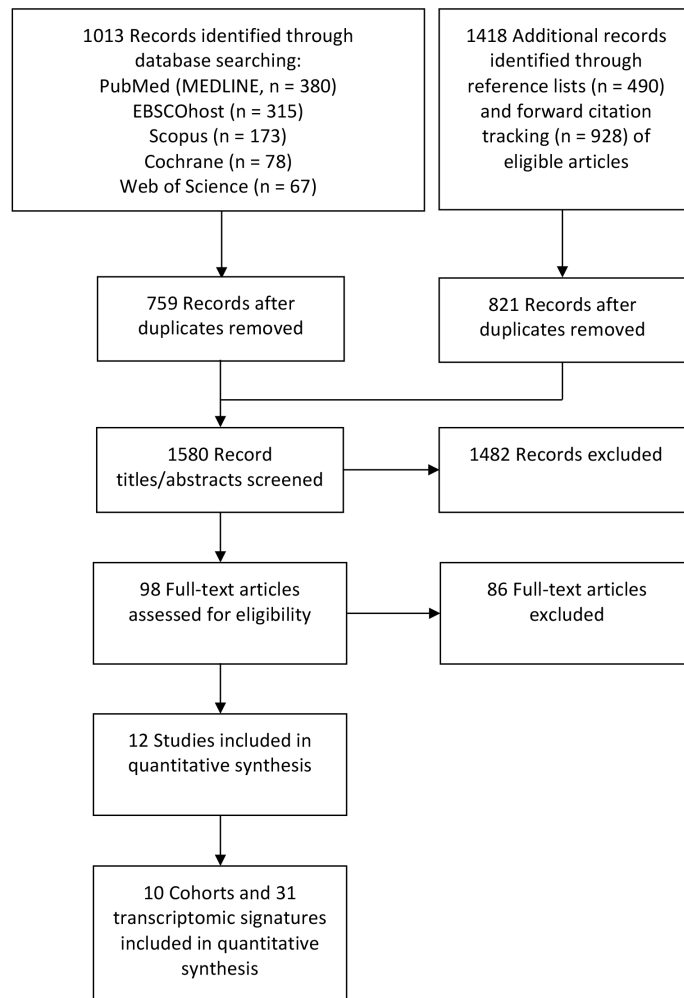

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram.

Study cohorts included in quantitative synthesis

The 12 eligible studies included 10 independent test or validation cohorts featuring PLHIV (Table 2), cumulatively evaluating diagnostic performance of 31 transcriptomic signatures (Table 3) incorporating over 700 unique transcripts. All independent test and validation cohorts enrolled PLHIV from outpatient clinics or hospital inpatients, most with suspicion (symptom-positive) or high risk (initiating antiretroviral therapy) of prevalent TB (Table 2). No studies prospectively enrolled participants without significant risk factors from community settings with a high burden of subclinical TB. Cohort sizes were generally small, with a maximum of 97 TB cases and 176 controls with HIV in the Kaforou2013[37] study, and 67 TB cases and 134 controls with HIV in the Södersten2020[46] study. All other cohorts included less than 50 TB cases with HIV. Only one of the 12 studies (Darboe2019)[42] reported transcriptomic signature prognostic performance for incident TB disease (disease recurrence after completion of TB treatment). The remainder reported diagnostic performance for differentiating patients with pulmonary TB disease from healthy controls, from controls with latent TB infection, or from other diseases. Most studies were conducted in Africa, with 8 out of 10 cohorts enrolling participants in South Africa. Only one cohort (Södersten2020)[46] included participants from another continent (Peru, South America). Many studies used the same microarray datasets, GSE37250[37] and GSE39941[38], from the Kaforou2013 (Malawi and South Africa) and Anderson2014 (Kenya, Malawi and South Africa) studies respectively, for external validation of signatures. Three cohorts measured signature scores using near-point-of-care RT-qPCR platforms (Darboe2019 and Penn-Nicholson2019a,b)[42,45], while only one study (Södersten2020)[46] used a point-of-care device, the Cepheid Xpert-MTB Host-Response (HR) Prototype, to measure the Sweeney3 signature[19,49] in biobanked whole blood RNA samples. No included studies validated signature performance for diagnosing extrapulmonary TB. However, one cohort (Anderson2014b)[38] included 17 HIV-infected children with non-microbiologically-confirmed, presumptive clinical TB with clinical and radiologic features that prompted empirical treatment. All other study cohorts, with the exception of Anderson2014, enrolled adults and used a microbiological reference standard (culture or Xpert MTB/RIF).

Transcriptomic signatures included in quantitative synthesis

All 31 signatures evaluated were discovered in case-control cohorts, eight of which were nested within prospective cohorts (Table 3). Only one of the 31 signatures (Rajan5)[44] was discovered in a cohort exclusively consisting of PLHIV. Seven signatures (Duffy10[43], Gliddon3[55], Gliddon4[55], Kaforou27[37], Kaforou44[37], Kaforou53[37], and Sweeney3[19,49]) were discovered using the Kaforou2013 cohort[37] and two signatures (Anderson42 and Anderson51)[38] using the Anderson2014 cohort[38], which included participants both with and without HIV. The 31 signatures were predominantly discovered in adult cohorts, with only 6 out 31 derived in paediatric (Gjøen7[51], Anderson42, and Anderson51) or adolescent (Darboe11[56], Penn-Nicholson6[45], and Zak16[40]) cohorts. Signature discovery cohorts were geographically diverse, with recruitment in Asia (3 signatures), Europe (10), North America (4), and South America (1);

however there was a predominance in Africa (16), particularly South Africa (15).

The 40 most frequent transcripts, all included in 3 or more signatures, are listed in Table 4. We used the INTERFEROME database[57] to classify interferon-stimulated genes (ISGs): We defined ISGs as genes significantly up or down regulated in expression (>1.5 fold change) in any human samples treated with Type-I IFN, relative to control samples. Almost all (38/39) of the most common transcripts with available gene annotations were classified as ISGs. The six transcripts most frequently included in signatures were Guanylate Binding Protein (GBP) 5 (11 signatures), GBP6 (10 signatures), Complement C1q B Chain (C1QB) and Fc fragment of IgG receptor Ia (FCGR1A) (7 signatures each), and Basic Leucine Zipper ATF-Like Transcription Factor 2 (BATF2) and GBP2 (6 signatures each).

Quality appraisal of eligible studies

In the patient selection domain, 5 of the 10 independent test and external validation cohorts consecutively or randomly enrolled participants, with the remaining cohorts not reporting sampling method (Figure 2). Eight of the 10 cohorts utilised a case-control design (Table 2), with exclusion of participants with uncertain diagnosis introducing a high risk of spectrum bias and potentially inflating diagnostic accuracy. The Penn-Nicholson2019b cohort[45] used a prospective design in recruiting symptomatic clinic attendees, but excluded probable and uncertain TB cases from analysis. Only one prospective diagnostic accuracy study, Turner2020[13], measured signature scores and tested performance in all enrolled participants with clinically suspected tuberculosis, including those with uncertain diagnosis, representative of the target population of symptomatic clinic attendees.

In the index test measurement domain, transcriptomic signature scores were interpreted without knowledge of the reference standard (i.e. blinded) in the Turner2020[13] and Penn-Nicholson2019b[45] cohorts, with unclear reporting for the other studies. Due to the early stage of biomarker development and diverse signature measurement and score calculation methodologies, no studies used pre-specified signature score thresholds. The risk of bias in the reference standard domain was deemed to be low in all cohorts with use of appropriate and standardised microbiological confirmatory TB testing (Mtb culture and Xpert MTB/RIF) likely to correctly classify the target condition. Reference standard results were interpreted without knowledge of the results of the transcriptomic signature scores (i.e. blinded) for all included studies. In the study flow and timing domain, all studies used an appropriate interval between index test and reference standard sample collection, and all participants received the same reference standard tests. However, only the Turner2020[13] study included all participants in analysis. In terms of applicability concerns, selection of participants, and measurement and interpretation of index tests and reference standard matched the review question for all included studies. Overall 9 of the 10 cohorts had a high risk of bias, and one study (Turner2020[13]) had some concerns due to lack of a pre-specified test threshold prospectively applied to each signature (Figure 2).

Table 2. Study cohorts included in quantitative synthesis.

Cohort ID*	Cohort name or accession number	Study design	Study setting	Country	Age group	Cohort type	Cases	Controls	Sample type	Signature measurement method	HIV+ TB cases	HIV+ controls
Anderson2014a[38]	GSE39941	Case-control nested within prospective cohort	Children with suspected TB (Hospital setting)	Malawi, South Africa	Children	Diagnostic	Sputum culture+	OD	Whole blood	Microarray	40	65
Anderson2014b[38]	GSE39941	Case-control nested within prospective cohort	Children with suspected TB (Hospital setting)	Kenya	Children	Diagnostic	Sputum culture+ (n=10) or composite reference standard (n=17)	OD	Whole blood	Microarray	27	27
Darboe2019[42]	TRuTH Cohort	Case-control nested within prospective cohort	Outpatients who recently completed TB treatment	South Africa	Adults	Prognostic	Sputum culture+ recurrent TB within 3 years of treatment completion	No recurrent TB	PBMC	RT-qPCR	38	84
Kaforou2013a[37]	GSE37250	Case-control study	Hospital inpatient, community, clinic	Malawi, South Africa	Adults	Diagnostic	Sputum culture+	LTBI	Whole blood	Microarray	84	84
Kaforou2013b[37]	GSE37250	Case-control study	Hospital inpatient, community, clinic	Malawi, South Africa	Adults	Diagnostic	Sputum culture+	OD	Whole blood	Microarray	97	92
Penn-Nicholson2019a[45]	CTBC Cohort	Case-control study	Clinic outpatient	South Africa	Adults	Diagnostic	Sputum culture+ or Xpert MTB/RIF+	LTBI+HC	Whole blood	RT-qPCR	44	40
Penn-Nicholson2019b[45]	ScreenTB and AE-TBC Cohorts	Prospective cohort	Symptomatic clinic attendees	South Africa	Adults	Diagnostic	Sputum culture+ or Xpert MTB/RIF+	ORD	Whole blood	RT-qPCR	12	24
Rajan2018[44,47]	Data not publicly available	Case-control nested within prospective cohort	Outpatients initiating ART	Uganda	Adults	Diagnostic	Sputum culture+	LTBI+HC	Whole blood	Microarray	40	80
Södersten2020[46]	Biobanked PAXgene samples	Nested case-control study	Symptomatic inpatients	Peru, South Africa	Adults	Diagnostic	Sputum/blood culture+ or sputum/urine Xpert MTB/RIF+	OD	Whole blood	Cepheid Xpert-MTB-HR-Prototype	67	134
Turner2020[13]	Array Express: E-MTAB-8290	Prospective cohort	Symptomatic clinic attendees	South Africa	Adults	Diagnostic	Sputum culture+ or Xpert MTB/RIF+	ORD	Whole blood	RNA sequencing	17	27

* A study may evaluate multiple signatures using several validation cohorts. Studies and cohorts are designated by the first author name and year of publication (e.g. Author2019a) and signatures by first author and number of transcripts (e.g. Author11).

GSE, Genomic Spatial Event (database). TruTH, TB Recurrence upon Treatment with HAART. CTBC, Cross-sectional TB Cohort. AE-TBC, African-European Tuberculosis Consortium. LTBI, latent tuberculosis infection. HC, healthy controls. OD, other diseases. ORD, other respiratory diseases. PBMC, peripheral blood mononuclear cell. RT-qPCR, real-time quantitative polymerase chain reaction.

Table 3. Transcriptomic signatures included in quantitative synthesis.

Signature ID*	Year published	Intended use	Signature model	Discovery cohort study design	Discovery cohort country	Discovery cohort age group	Discovery cohort HIV status	Discovery data	Discovery approach
Anderson42 (39. LTBI)[38]	2014	Diagnostic (TB vs LTBI)	Disease risk score	Case-control nested in prospective cohort	Malawi, South Africa	Children	Mixed	Genome-wide RNA Microarray	Elastic net
Anderson51 (39. OD)[38]	2014	Diagnostic (TB vs OD)	Disease risk score	Case-control nested in prospective cohort	Malawi, South Africa	Children	Mixed	Genome-wide RNA Microarray	Elastic net
Berry86[58]	2010	Diagnostic (TB vs OD)	k-nearest neighbours algorithm	Case-control	United Kingdom	Adults	Negative	Genome-wide RNA Microarray	Differential expression
Berry393[58]	2010	Diagnostic (TB vs LTBI+HC)	k-nearest neighbours algorithm	Case-control	South Africa, United Kingdom	Adults	Negative	Genome-wide RNA Microarray	Differential expression
Darboe11[56]	2018	Prognostic (Incipient TB vs HC)	Pair-wise SVM ensemble	Case-control nested in prognostic cohort	South Africa	Adolescents	Negative	RT-qPCR using selected target gene set (Zak16)	Refinement of Zak16: most reproducible and least redundant genes
de Araujo1 (NPC2)[59]	2016	Diagnostic (TB vs HC + LTBI)	Standardised expression	Case-control	Brazil	Adults	Not stated	Genome-wide RNA Sequencing	Differential expression
Duffy10[43]	2019	Diagnostic (TB vs LTBI + OD)	Six-class multinomial random forest	Case-control	Malawi, South Africa	Adults	Mixed	Genome-wide RNA Microarray	Multinomial random forest
Gjøen8 (7)[51]	2017	Diagnostic (TB vs HC)	LASSO regression	Case-control	India	Children	Negative	198 selected target genes measured by dc-RT MLPA	LASSO
Gliddon3[55]	2020	Diagnostic (TB vs LTBI)	Disease risk score	Case-control	Malawi, South Africa	Adults	Mixed	Genome-wide RNA Microarray	FS-PLS
Gliddon4[55]	2020	Diagnostic (TB vs OD)	Disease risk score	Case-control	Malawi, South Africa	Adults	Mixed	Genome-wide RNA Microarray	FS-PLS
Huang13 (11)[60]	2015	Diagnostic (TB vs HC + OD)	SVM	Case-control	United Kingdom	Adults	Negative	Genome-wide RNA Microarray	LASSO, $L_{1/2}$, and elastic net
Kaforou27 (25)[37]	2013	Diagnostic (TB vs LTBI)	Disease risk score	Case-control	Malawi, South Africa	Adults	Mixed	Genome-wide RNA Microarray	Elastic net
Kaforou44 (39)[37]	2013	Diagnostic (TB vs OD)	Disease risk score	Case-control	Malawi, South Africa	Adults	Mixed	Genome-wide RNA Microarray	Elastic net
Kaforou53 (45)[37]	2013	Diagnostic (TB vs OD+LTBI)	Disease risk score	Case-control	Malawi, South Africa	Adults	Mixed	Genome-wide RNA Microarray	Elastic net

Signature ID*	Year published	Intended use	Signature model	Discovery cohort study design	Discovery cohort country	Discovery cohort age group	Discovery cohort HIV status	Discovery data	Discovery approach
Maertzdorf4[48]	2015	Diagnostic (TB vs HC + LTBI)	Random forest	Case-control	India	Adults	Negative	360 selected target genes measured by RT-qPCR	Random forest
Penn-Nicholson6[45]	2020	Prognostic (Incipient TB vs HC)	Pair-wise ensemble structure	Case-control nested in prognostic cohort	South Africa	Adolescents	Negative	Genome-wide RNA Sequencing and targeted RT-qPCR	SVM-based gene pair ratios approach
Qian17[50]	2016	Diagnostic (TB vs HC + OD)	Sum of standardised expression	Case-control	United Kingdom	Adults	Negative	862 Nrf2-mediated genes measured by RNA Microarray	Differential expression
Rajan5[44]	2018	Diagnostic (HIV+TB vs HIV+HC)	Unsigned sums	Case-control nested in prospective cohort	Uganda	Adults	Positive	Genome-wide RNA Microarray	Differential expression
Roe1 (BATF2)[22]	2016	Diagnostic (TB vs HC and TB patients 2-4 years post recovery)	Standardised expression	Case-control	United Kingdom	Adults	Negative	Genome-wide RNA Microarray	SVM (linear kernel)
Roe3[23]	2020	Prognostic (Incipient TB vs HC)	SVM (linear kernel)	Case-control	United Kingdom	Adults	Negative	Genome-wide RNA Sequencing	Stability selection and SVM
Roe4[22]	2016	Diagnostic (TB vs OD)	SVM (linear kernel)	Case-control	United Kingdom	Adults	Negative	Genome-wide RNA Microarray	SVM (linear kernel)
Roe5[22]	2016	Diagnostic (TB vs HC + OD)	SVM (linear kernel)	Case-control	United Kingdom	Adults	Negative	Genome-wide RNA Microarray	SVM (linear kernel)
Sambarey10[41]	2017	Diagnostic (TB vs HC + LTBI)	Linear discriminant analysis	Case-control	India	Adults	Negative	Genome-wide RNA Sequencing	Human Protein–Protein Interaction Network, Condition-specific Weighted Networks, Differential expression
Singhania20[53]	2018	Diagnostic (TB vs LTBI)	Disease risk score (modified)	Case-control	United Kingdom	Adults	Negative	Genome-wide RNA Sequencing	Random forest using modular approach
Suliman2[61]	2018	Prognostic (Incipient TB vs HC)	Pair-wise ensemble structure	Case-control nested in prognostic cohort	The Gambia, South Africa	Adults	Negative	Genome-wide RNA Sequencing	SVM-based gene pair ratios approach
Suliman4[61]	2018	Prognostic (Incipient TB vs HC)	Pair-wise ensemble structure	Case-control nested in prognostic cohort	The Gambia, Ethiopia, South Africa	Adults	Negative	Genome-wide RNA Sequencing	SVM-based gene pair ratios approach

Signature ID*	Year published	Intended use	Signature model	Discovery cohort study design	Discovery cohort country	Discovery cohort age group	Discovery cohort HIV status	Discovery data	Discovery approach
Sweeney3[19,49]	2016	Diagnostic (TB vs HC + LTBI + OD)	[(GBP5 + DUSP3) / 2] – KLF2	Multicohort analysis (Case-control)	France, Malawi, South Africa, United Kingdom, USA	Adults	Mixed	Genome-wide RNA Microarray	Significance thresholding and forward search
Walter47 (32)[39]	2016	Diagnostic (TB vs OD)	SVM	Case-control	USA	Adults	Negative	Genome-wide RNA Microarray	SVM with recursive feature elimination
Walter51 (46)[39]	2016	Diagnostic (TB vs LTBI)	SVM	Case-control	USA	Adults	Negative	Genome-wide RNA Microarray	SVM with recursive feature elimination
Walter119 (101)[39]	2016	Diagnostic (TB vs LTBI + OD)	SVM	Case-control	USA	Adults	Negative	Genome-wide RNA Microarray	SVM with recursive feature elimination
Zak16[40]	2016	Prognostic (Incipient TB vs LTBI)	Pair-wise SVM ensemble	Case-control nested in prognostic cohort	South Africa	Adolescents	Negative	Genome-wide RNA Sequencing	SVM-based gene pair approach

* A study may evaluate multiple signatures using several validation cohorts. Studies and cohorts are designated by the first author name and year of publication (e.g. Author2019a) and signatures by first author and number of transcripts (e.g. Author11). Numbers in brackets indicate the subsequently reduced number of transcripts in a given signature due to duplicate transcript symbols (IDs) or where transcript sequences from an original discovery cohorts could not be mapped to a more recent reference transcriptome.

LTBI, latent tuberculosis infection. OD, other diseases. HC, healthy controls. SVM, support vector machines. LASSO, Least Absolute Shrinkage and Selection Operator. FS-PLS, Forward Selection-Partial Least Squares. RT-qPCR, real-time quantitative polymerase chain reaction. Dc-RT MLPA, dual-color-Reverse-Transcriptase-Multiplex-Ligation-dependent-Probe-Amplification.

Table 4. Forty most frequent transcripts included in 3 or more transcriptomic signatures, sorted by transcript use frequency.

Gene ID	Transcript use frequency	*Interferon-stimulated gene	Anderson42 (39)	Anderson51 (39)	Berry393	Berry86	Darboe11	de Araujo1	Duffy10	Gjøen8	Gliddon3	Gliddon4	Huang13 (11)	Kaforou27 (25)	Kaforou44 (39)	Kaforou53 (45)	Maertzdorf4
GBP5	11	Yes	1	1	1		1			1						1	
GBP6	10	Yes	1	1	1				1			1		1	1	1	
C1QB	7	Yes		1	1				1		1		1	1			
FCGR1A	7	Yes			1						1			1		1	
BATF2	6	Yes			1		1										
GBP2	6	Yes			1		1										
ANKRD22	5	Yes			1									1		1	
DUSP3	5	Yes			1									1	1	1	
FCGR1B	5	Yes			1				1					1		1	
GBP1	5	Yes			1		1										1
PRDM1	5	Yes			1							1		1			
SCARF1	5	Yes					1										
SEPT4	5	Yes			1										1	1	
SERPING1	5	Yes			1		1							1			
ALDH1A1	4	**No†**		1	1										1	1	
CYB561	4	Yes		1											1	1	
DHRS9	4	Yes			1	1											
FCGR1C	4	Yes					1							1		1	
GAS6	4	Yes			1									1		1	
IFITM3	4	Yes			1				1								1
LHFPL2	4	Yes			1									1	1	1	
SMARCD3	4	Yes		1	1									1			
STAT1	4	Yes			1		1										
VAMP5	4	Yes		1	1									1			
ACTA2	3	Yes	1		1												
BLK	3	Yes			1											1	
C20ORF103 (LAMP5)	3	Yes		1													
CD74	3	Yes				1									1	1	
CREB5	3	Yes			1										1	1	
DEFA1	3	Yes	1	1												1	
ETV7	3	Yes			1		1										
FER1L3 (MYOF)	3	Yes		1	1	1											
ID3	3	Yes			1				1								1
KREMEN1	3	Yes		1	1												
LOC389386	3	NA**		1											1	1	
OSBPL10	3	Yes		1												1	
POLB	3	Yes				1										1	
TAP1	3	Yes			1		1										
TMCC1	3	Yes										1			1	1	
TRAFD1	3	Yes			1		1										

Gene ID	Transcript use frequency	*Interferon-stimulated gene	Penn-Nicholson6	Qian17	Rajan5	Roe1	Roe3	Roe4	Roe5	Sambarey10	Singhania20	Suliman2	Suliman4	Sweeney3	Walter47 (32)	Walter51 (46)	Walter119 (101)	Zak16
GBP5	11	Yes					1							1	1		1	1
GBP6	10	Yes			1												1	
C1QB	7	Yes														1		
FCGR1A	7	Yes								1						1		1
BATF2	6	Yes				1	1		1									1
GBP2	6	Yes	1													1	1	1
ANKRD22	5	Yes										1						1
DUSP3	5	Yes												1				
FCGR1B	5	Yes	1															
GBP1	5	Yes		1														1
PRDM1	5	Yes													1		1	
SCARF1	5	Yes					1				1						1	1
SEPT4	5	Yes										1						1
SERPING1	5	Yes	1															1
ALDH1A1	4	No†																
CYB561	4	Yes												1				
DHRS9	4	Yes														1	1	
FCGR1C	4	Yes																1
GAS6	4	Yes										1						
IFITM3	4	Yes		1														
LHFPL2	4	Yes																
SMARCD3	4	Yes								1								
STAT1	4	Yes		1														1
VAMP5	4	Yes														1		
ACTA2	3	Yes			1													
BLK	3	Yes										1						
C20ORF103 (LAMP5)	3	Yes														1	1	
CD74	3	Yes																
CREB5	3	Yes																
DEFA1	3	Yes																
ETV7	3	Yes																1
FER1L3 (MYOF)	3	Yes																
ID3	3	Yes																
KREMEN1	3	Yes														1		
LOC389386	3	NA**																

Gene ID	Transcript use frequency	*Interferon-stimulated gene	Penn-Nicholson6	Qian17	Rajan5	Roe1	Roe3	Roe4	Roe5	Sambarey10	Singhania20	Suliman2	Suliman4	Sweeney3	Walter47 (32)	Walter51 (46)	Walter119 (101)	Zak16
OSBPL10	3	Yes										1						
POLB	3	Yes															1	
TAP1	3	Yes																1
TMCC1	3	Yes																
TRAFD1	3	Yes																1

* Gene significantly up or down regulated in expression (>1.5 fold change) in samples treated with Type-I interferon, relative to control samples, were defined as interferon-stimulated genes (ISGs). Data from INTERFEROME database (Samarajiwa SA, Forster S, Auchettl K, Hertzog PJ. INTERFEROME: the database of interferon regulated genes. *Nucleic Acids Res*. 2009;37:D852-7).

** No updated gene annotation available. NA, not available.

† No experimental data demonstrating >1.5 fold-change in expression in response to Type-I interferon, however, there was >1.5 fold-change in expression response to Type-II interferon.
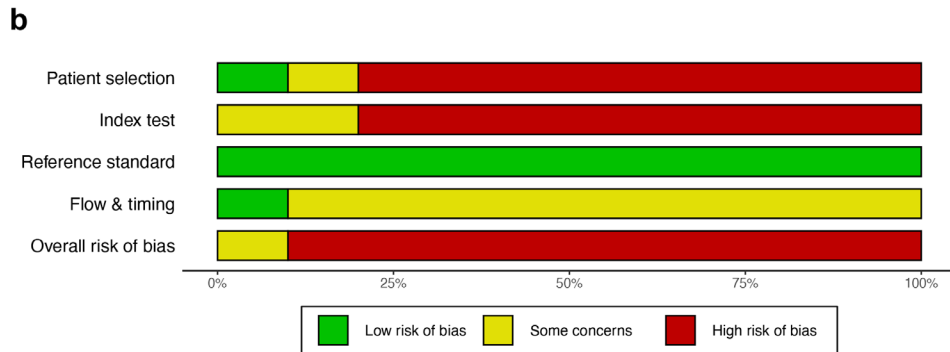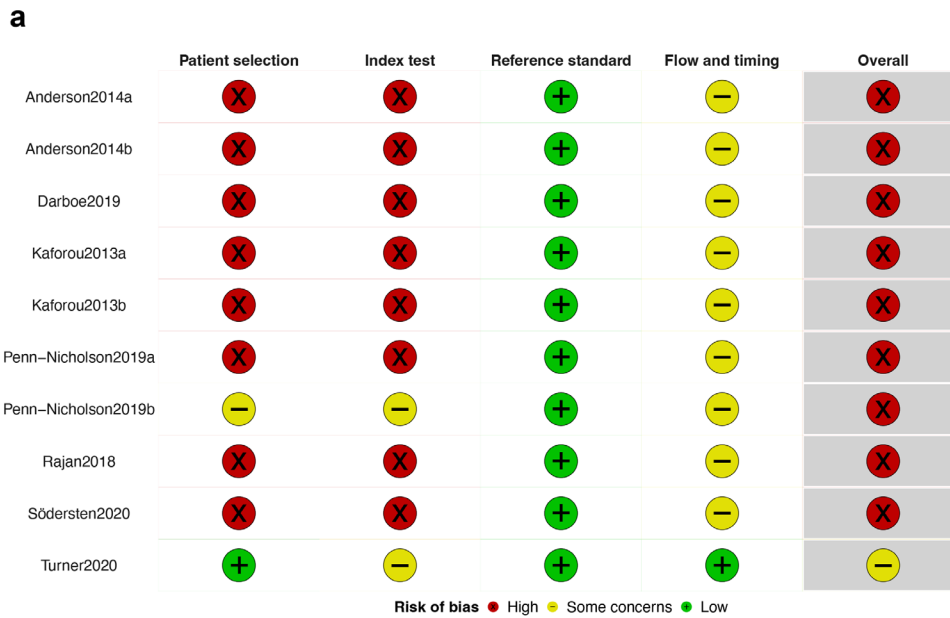

a

b

Figure 2. Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2). (**a**) Assessment of bias in individual cohorts included in the systematic review and (**b**) summary of results of the QUADAS-2 assessment in the four domains: patient selection, index test, reference standard, and flow and timing.

Transcriptomic signature diagnostic performance

Nine independent test or external validation diagnostic cohorts with PLHIV subgroups were included in the systematic review (Table 2). Three cohorts evaluated diagnostic performance of 12 signatures for discriminating HIV-infected adults with prevalent TB disease from latent-Mtb infected individuals or healthy controls with HIV (Figure 3), and 6 cohorts evaluated diagnostic performance of 29 signatures for discriminating HIV-infected adults or children with prevalent TB disease from those with HIV and other respiratory or systemic diseases (Figure 4).

Only the Kaforou27 (Kaforou2013a and Rajan2018 cohorts), Penn-Nicholson6 (Penn-Nicholson2020a cohort), and Rajan5 (Rajan2018 cohort) signatures met the WHO TPP minimum performance criteria for a triage test for differentiating adults with prevalent TB disease from latent-Mtb infected individuals or healthy controls with HIV in these case-control studies (Figure 3). However, the upper bounds of the 95% confidence intervals for 4 other signatures met the TPP benchmarks: Darboe11 (Penn-Nicholson2020a cohort), Duffy10 (Kaforou2013a cohort), Kaforou53 (Rajan2018 cohort), and Sweeney3 (Kaforou2013a and Rajan2018 cohorts).

Signatures which are able to distinguish PLHIV with prevalent TB disease from those with other respiratory or systemic diseases are more clinically useful in the inpatient or outpatient setting, as compared to the community. Seven out of 29 signatures met the WHO TPP minimum performance criteria for a triage test in this context (Figure 4): Anderson51 in the Anderson2014b paediatric cohort; Kaforou44 in the Kaforou2013b adult cohort; and Maertzdorf4, Qian17, Roe1, Roe3, and Sweeney3 in the Turner2020 adult cohort. A further 8 signatures (Anderson39.OD, de Araujo1, Kaforou25, Kaforou39, Penn-Nicholson6, Rajan5, Suliman2, and Zak16) had upper 95% confidence interval bounds for sensitivity which met the 90% TPP benchmark (with specificity set at 70%) in the Turner2020 prospective diagnostic accuracy cohort. However, the PLHIV subgroup of this cohort only included 17 TB cases and 27 controls with other respiratory diseases. In other cohorts, the Anderson51 (Anderson2014a cohort), Penn-Nicholson6 (Penn-Nicholson2020b cohort), and Sweeney3 (Kaforou2013b cohort) signatures also approached the WHO minimum TPP triage test criteria. Only one signature (Sweeney3) was tested using a point-of-care device, the Cepheid Xpert-MTB-HR Prototype in the Södersten2020 cohort on biobanked whole blood RNA samples from symptomatic

a

	TB	No TB	AUC (95%CI)	Specificity (95%CI) at 90% Sensitivity		Sensitivity (95%CI) at 70% Specificity
Kaforou2013a						
Duffy10	47	48	0.91 (0.86–0.97)		66.7 (52.5–78.3)	89.4 (77.4–95.4)
Roe1	98	84	0.82 (0.76–0.88)		52.4 (41.8–62.7)	77.6 (68.3–84.7)
Sweeney3	98	84	0.89 (0.84–0.94)		65.5 (54.8–74.8)	87.8 (79.8–92.9)
Zak16	98	84	0.86 (0.81–0.91)		48.8 (38.4–59.3)	83.7 (75.1–89.7)
Penn–Nicholson2020a						
Darboe11	44	38	0.88 (0.81–0.95)		55.3 (39.7–69.9)	81.8 (68.0–90.5)
Penn–Nicholson6	42	40	0.93 (0.87–0.98)		72.5 (57.2–83.9)	90.5 (77.9–96.2)

Specificity (%) at 90% Sensitivity   Sensitivity (%) at 70% Specificity

b

	TB	No TB	AUC (95%CI)	Specificity (95%CI)	Sensitivity (95%CI)
Kaforou2013a					
Berry393	97	84	0.88 (0.82–0.92)	80.0 (71.0–87.0)	74.0 (65.0–82.0)
Kaforou27	18	20	0.97 (0.95–1.00)	90.0 (75.0–100)	94.4 (83.0–100)
Walter51	100#	82	0.90 (0.90–0.90)	77.1 (76.4–77.8)	80.0 (79.2–80.8)
Rajan2018*					
Kaforou27*	40	80	0.93 (0.84–1.00)	88.4 (77.0–100)	92.8 (60.0–100)
Kaforou44*	40	80	0.71 (0.50–0.89)	100 (86.0–100)	42.8 (22.0–81.0)
Kaforou53*	40	80	0.94 (0.86–0.99)	96.1 (74.0–100)	71.4 (50.0–100)
Rajan5	ns	ns	0.87 (0.72–0.98)	75.0 (ns–ns)	94.0 (ns–ns)
Sweeney3*	40	80	0.91 (0.80–0.98)	92.3 (73.0–100)	71.4 (40.0–100)

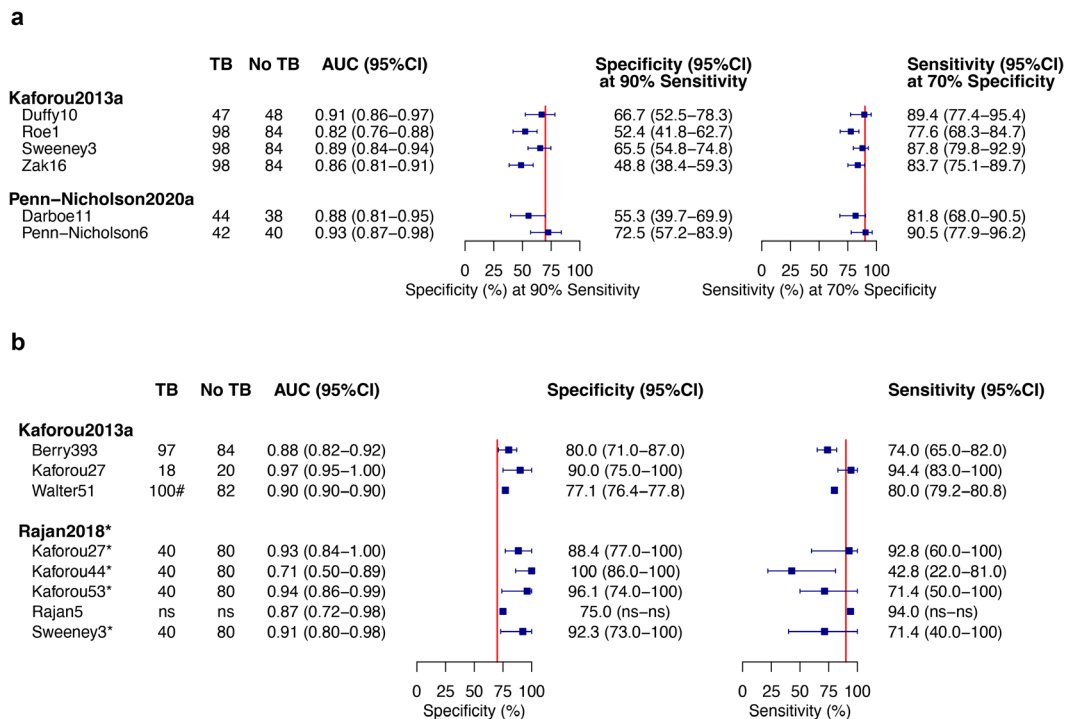Specificity (%)   Sensitivity (%)

Figure 3. Transcriptomic signature diagnostic performance for differentiation of patients with microbiologically confirmed TB disease from latent-Mtb infected individuals or healthy controls in independent test or external validation cohorts. (**a**) Signature diagnostic performance (area under the curve, AUC) recalculated using participant-level data retrieved from supplementary data or supplied by study authors, with sensitivity and specificity benchmarked against the WHO TPP minimum performance criteria for a triage test (70% specificity and 90% sensitivity, red vertical lines)[16]. 95% confidence intervals for AUC and proportions (sensitivity and specificity) were calculated using the DeLong[30] and binomial proportion (Wilson)[31] methods, respectively. (**b**) Signature diagnostic performance and 95% confidence intervals as originally reported in studies where participant-level data were not available. *Data for other signatures measured in the full Rajan2018 cohort were extracted from an ATS Conference abstract (Rajan, *AJRCCM* 2017)[47]. The number of participants included in the Rajan2018 test set was not specified (ns). #100 TB cases were reported in the paper by Walter *et al.* (*J Clin Microbiol*, 2016)[39], however only 97 TB cases were included in the original microarray analysis[37].

a

	TB	No TB	AUC (95%CI)		Specificity (95%CI) at 90% Sensitivity		Sensitivity (95%CI) at 70% Specificity
Anderson2014ab							
Duffy10	51	93	0.80 (0.73–0.88)		45.2 (35.4–55.3)		70.6 (57.0–81.3)
Sweeney3	51	93	0.80 (0.72–0.88)		34.4 (25.5–44.5)		74.5 (61.1–84.5)
Zak16	51	93	0.79 (0.71–0.87)		26.9 (18.9–36.7)		72.5 (59.1–82.9)
Kaforou2013b							
Duffy10	47	62	0.81 (0.72–0.89)		37.1 (26.2–49.5)		74.5 (60.5–84.7)
Sweeney3	98	92	0.86 (0.81–0.92)		67.4 (57.3–76.1)		87.8 (79.8–92.9)
Zak16	98	92	0.84 (0.79–0.90)		58.7 (48.5–68.2)		81.6 (72.8–88.1)
Penn–Nicholson2020b							
Penn–Nicholson6	12	24	0.80 (0.65–0.94)		62.5 (42.7–78.8)		75.0 (46.8–91.1)
Turner2020							
Anderson39.LTBI	17	27	0.55 (0.35–0.75)		0.0 (0.0–12.5)		47.1 (26.2–69.0)
Anderson39.OD	17	27	0.81 (0.68–0.94)		51.9 (34.0–69.3)		82.4 (59.0–93.8)
de Araujo1 (NPC2)	17	27	0.78 (0.65–0.92)		48.1 (30.7–66.0)		76.5 (52.7–90.4)
Duffy10	17	27	0.79 (0.66–0.92)		44.4 (27.6–62.7)		70.6 (46.9–86.7)
Gjoen8	17	27	0.52 (0.34–0.69)		11.1 (3.9–28.1)		23.5 (9.6–47.3)
Gliddon3	17	27	0.80 (0.67–0.94)		37.0 (21.5–55.8)		64.7 (41.3–82.7)
Gliddon4	17	27	0.78 (0.64–0.92)		40.7 (24.5–59.3)		70.6 (46.9–86.7)
Huang11	17	27	0.64 (0.45–0.82)		0.0 (0.0–12.5)		47.1 (26.2–69.0)
Kaforou25	17	27	0.85 (0.73–0.97)		55.6 (37.3–72.4)		76.5 (52.7–90.4)
Kaforou39	17	27	0.85 (0.71–0.98)		44.4 (27.6–62.7)		76.5 (52.7–90.4)
Kaforou45	17	27	0.66 (0.49–0.83)		29.6 (15.9–48.5)		52.9 (31.0–73.8)
Maertzdorf4	17	27	0.86 (0.74–0.97)		70.4 (51.5–84.1)		94.1 (73.0–99.0)
Penn–Nicholson6	17	27	0.83 (0.70–0.97)		44.4 (27.6–62.7)		76.5 (52.7–90.4)
Qian17	17	27	0.83 (0.70–0.96)		70.4 (51.5–84.1)		94.1 (73.0–99.0)
Rajan5	17	27	0.81 (0.67–0.96)		22.2 (10.6–40.8)		82.4 (59.0–93.8)
Roe1	17	27	0.91 (0.83–0.99)		77.8 (59.2–89.4)		94.1 (73.0–99.0)
Roe3	17	27	0.89 (0.79–1.00)		70.4 (51.5–84.1)		94.1 (73.0–99.0)
Roe4	17	27	0.60 (0.42–0.77)		14.8 (5.9–32.5)		41.2 (21.6–64.0)
Roe5	17	27	0.72 (0.56–0.88)		22.2 (10.6–40.8)		70.6 (46.9–86.7)
Singhania20	17	27	0.62 (0.45–0.79)		40.7 (24.5–59.3)		23.5 (9.6–47.3)
Suliman2	17	27	0.85 (0.74–0.97)		48.1 (30.7–66.0)		88.2 (65.7–96.7)
Suliman4	17	27	0.73 (0.57–0.88)		25.9 (13.2–44.7)		64.7 (41.3–82.7)
Sweeney3	17	27	0.95 (0.89–1.00)		85.2 (67.5–94.1)		100.0 (81.6–100.0)
Walter101	17	27	0.61 (0.44–0.79)		29.6 (15.9–48.5)		47.1 (26.2–69.0)
Walter32	17	27	0.58 (0.40–0.76)		7.4 (2.1–23.4)		41.2 (21.6–64.0)
Walter46	17	27	0.55 (0.36–0.74)		0.0 (0.0–12.5)		35.3 (17.3–58.7)
Zak16	17	27	0.80 (0.66–0.94)		55.6 (37.3–72.4)		88.2 (65.7–96.7)

Specificity (%) at 90% Sensitivity   Sensitivity (%) at 70% Specificity

b

	TB	No TB	AUC (95%CI)		Specificity (95%CI)		Sensitivity (95%CI)
Anderson2014a							
Anderson51	9	13	0.85 (0.64–0.97)		61.5 (30.8–84.6)		77.8 (55.6–100)
Anderson2014b							
Anderson51	10	26	0.94 (0.84–1.00)		92.3 (80.8–100)		90.0 (70.0–100)
Kaforou2013b							
Berry86	97	92	0.75 (0.68–0.82)		69.0 (59.0–78.0)		67.0 (58.0–75.0)
Kaforou44	20	19	0.94 (0.83–1.00)		84.2 (68.0–100)		95.0 (85.0–100)
Sambarey10	97	92	ns		85.9 (ns–ns)		80.4 (ns–ns)
Walter47	102#	88	0.89 (0.88–0.89)		76.3 (75.3–77.3)		84.8 (84.0–85.7)
Södersten2020							
Sweeney3	67	134	0.89 (0.83–0.94)		55.8 (47.2–64.1)		91.0 (81.8–95.8)

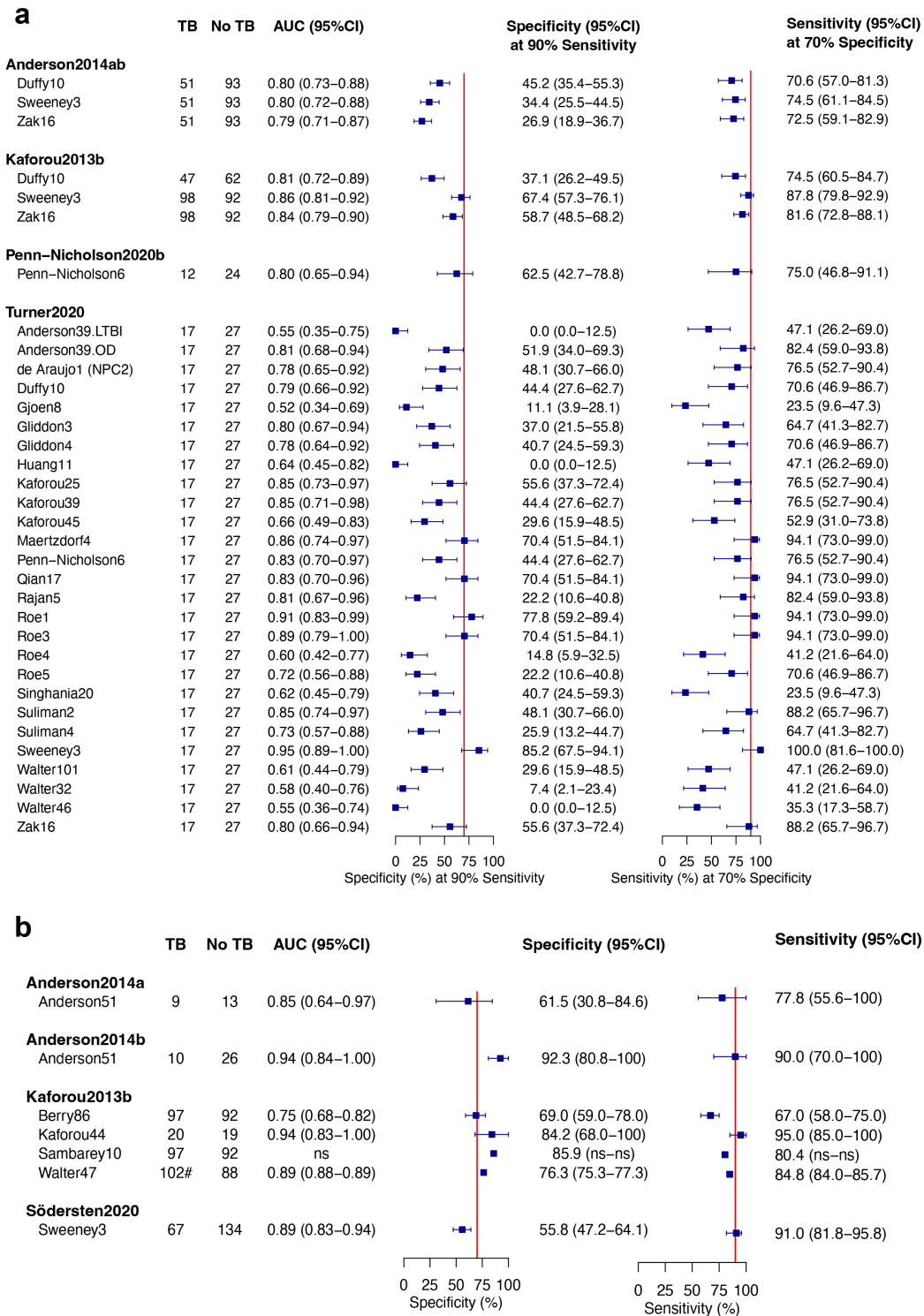Specificity (%)   Sensitivity (%)

Figure 4. Transcriptomic signature diagnostic performance for differentiation of patients with microbiologically confirmed TB disease from patients with other diseases in independent test or external validation cohorts. (**a**) Signature diagnostic performance (area under the curve, AUC) recalculated using participant-level data retrieved from supplementary data or supplied by study authors, with sensitivity and specificity benchmarked against the WHO TPP minimum performance criteria for a triage test (70% specificity and 90% sensitivity, red vertical lines)[16]. 95% confidence intervals for AUC and proportions (sensitivity and specificity) were calculated using the DeLong[30] and binomial proportion (Wilson)[31] methods, respectively. (**b**) Signature diagnostic performance and 95% confidence intervals as originally reported in studies where participant-level data were not available. #102 TB cases were reported in the paper by Walter *et al.* (*J Clin Microbiol,* 2016)[39]. however only 97 TB cases were included in the original microarray analysis[37]. ns, not specified.

inpatients with HIV. Sweeney3 achieved sensitivity of 91.0% (95%CI 81.8–95.8) and specificity of 55.8% (95%CI 47.2–64.1), falling short of the WHO TPP benchmark. No signatures met the WHO TPP benchmark criteria for a non-sputum confirmatory diagnostic test in PLHIV (98% specificity and 80% sensitivity)[16] in any cohorts.

Participant-level data were available for only 3 signatures in the Anderson2014b cohort, which also included 17 children with non-microbiologically confirmed, presumptive clinically-diagnosed TB disease (Table 2). All signatures performed poorly in differentiating TB from other diseases in this subset (Figure 5).

Only the Södersten2020 study reported signature performance stratified by CD4 cell count, with lower Sweeney3 specificity in the 56 inpatients with CD4 cell count less than 200 (66.7%, 95%CI 46.7–82.0) as compared to 129 inpatients with CD4 greater than or equal to 200 (94.7%, 95%CI 88.9–97.5)[46]. Conversely, sensitivity was higher in the inpatients with CD4 cell count less than 200 (93.8%, 95%CI 79.9–98.3) as compared to the inpatients with CD4 greater than or equal to 200 (75.0%, 95%CI 50.5–89.9). No participant-level data were available to perform further subgroup analyses by CD4 cell count, HIV

plasma viral load, TPT or ART status, or other variables in any cohorts.

There was considerable clinical and methodological heterogeneity between cohorts, with participants recruited in diverse settings, with dissimilar eligibility criteria, and distinct composition of the control groups. Studies also used different signature measurement methods (microarray, RNA sequencing, and RT-qPCR), different methods of signature score calculation, and there were no standardised signature score thresholds. Due to the significant heterogeneity in study design and index test measurement, and limited participant data, a meta-analysis was not deemed appropriate.

Transcriptomic signature prognostic performance

Only one study (Darboe2019)[42] evaluated transcriptomic signature prognostic performance for recurrent TB disease in adults with HIV who had recently completed TB therapy. There were no eligible studies evaluating performance for prediction of progression to incident TB disease in PLHIV without prior TB disease. We stratified the prognostic performance of the Darboe11 signature by time from measurement until TB disease recurrence (Figure 6). Signature prognostic performance was best within 90 days of signature measurement, and waned

	TB	No TB	AUC (95%CI)		Specificity (95%CI) at 90% Sensitivity		Sensitivity (95%CI) at 70% Specificity
Anderson2014b							
Duffy10	17	93	0.49 (0.36–0.62)		11.8 (6.7–19.9)		11.8 (3.3–34.3)
Sweeney3	17	93	0.64 (0.52–0.76)		33.3 (24.6–43.4)		41.2 (21.6–64.0)
Zak16	17	93	0.54 (0.40–0.67)		24.7 (17.1–34.4)		23.5 (9.6–47.3)
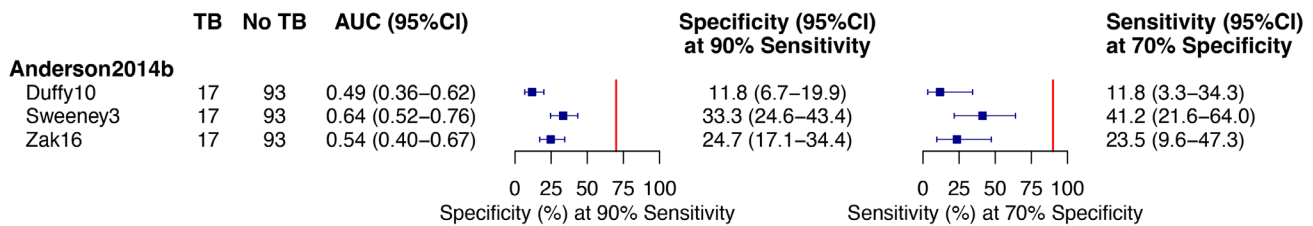
Figure 5. Transcriptomic signature diagnostic performance for non-microbiologically confirmed presumptive TB disease. Transcriptomic signature diagnostic performance for differentiation of children with non-microbiologically confirmed (i.e. presumptive clinically-diagnosed) TB disease from children with other diseases in an external validation cohort. Signature diagnostic performance (area under the curve, AUC) recalculated using participant-level data supplied by study authors, with sensitivity and specificity benchmarked against the WHO TPP minimum performance criteria for a triage test (70% specificity and 90% sensitivity, red vertical lines)[16]. 95% confidence intervals for AUC and proportions (sensitivity and specificity) were calculated using the DeLong[30] and binomial proportion (Wilson)[31] methods, respectively.

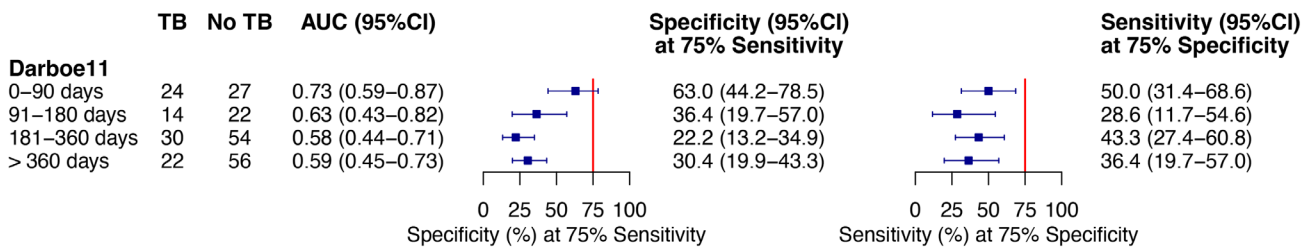|	TB	No TB	AUC (95%CI)		Specificity (95%CI) at 75% Sensitivity		Sensitivity (95%CI) at 75% Specificity
Darboe11							
0–90 days	24	27	0.73 (0.59–0.87)		63.0 (44.2–78.5)		50.0 (31.4–68.6)
91–180 days	14	22	0.63 (0.43–0.82)		36.4 (19.7–57.0)		28.6 (11.7–54.6)
181–360 days	30	54	0.58 (0.44–0.71)		22.2 (13.2–34.9)		43.3 (27.4–60.8)
> 360 days	22	56	0.59 (0.45–0.73)		30.4 (19.9–43.3)		36.4 (19.7–57.0)

Figure 6. Transcriptomic signature prognostic performance for microbiologically confirmed TB disease. Darboe11 transcriptomic signature prognostic performance for microbiologically confirmed recurrent TB disease in the TB Recurrence upon Treatment with HAART (TRuTH) cohort (Darboe2019)[42]. Signature prognostic performance (area under the curve, AUC) is stratified by time window following completion of TB treatment: 0–90, 91–180, 181–360, and >360 days. Sensitivity and specificity benchmarked against the WHO TPP minimum performance criteria for a prognostic test (75% sensitivity and 75% specificity, red vertical lines)[18].

thereafter. Sensitivity and specificity of the Darboe11 signature did not meet the minimum WHO TPP performance criteria for a prognostic test (75% sensitivity and 75% specificity)[18] in any time window. Sample size was not sufficient to perform subgroup prognostic performance analyses, however Darboe11 signature scores were higher in individuals with detectable plasma HIV viral load (>400 copies/mL) as compared to those with an undetectable plasma viral load (<400 copies/mL; p<0.0001)[42].

GRADE evidence summary

A total of 10 cohorts, 2 prospective cohorts (29 TB cases; 51 other respiratory diseases) and 8 case-control studies (353 TB cases; 606 controls), were included in this systematic review of the diagnostic and prognostic accuracy of host blood transcriptomic signatures in PLHIV. All studies used reliable reference standards for definitive TB diagnosis. However, we adjudged that there was a very serious risk of bias due to exclusion of participants with indeterminate (non-microbiologically confirmed) TB in numerous studies, removing diagnostic uncertainty, and resulting in reduced diversity of clinical TB disease. Several case-control studies also included healthy asymptomatic controls, healthy Mtb-sensitised (latent Mtb infected) individuals (IGRA or TST positive), or individuals with other uncommon diseases, not reflective of the target population or setting, further exacerbating the spectrum bias. The inclusion of severe TB cases and healthy controls (or controls with inappropriate other diseases) may have resulted in misleadingly high diagnostic accuracy in some of these studies. Other limitations included uncertainty regarding consecutive recruitment, blinding status not clearly stated, and lack of *a priori* score thresholds.

Indirectness is synonymous with applicability, generalisability, translatability, and external validity of the evidence[36]. Included studies evaluated diagnostic performance among adults or children, within clinical outpatient and hospital inpatient settings, prospectively among symptomatic clinic attendees or within matched case-control cohorts. Some of these settings and populations are not appropriate or relevant to clinical practice, and results are unlikely to be generalisable. The lack of diagnostic uncertainty, spectrum bias, and inappropriate control selection is a concern for external validity of these results. Point-of-care device translatability has only been tested for one signature (Sweeney3) among PLHIV, with unsatisfactory diagnostic accuracy[46]. Technical variability and operator reliability have not been tested on point-of-care platforms for any tests.

With regards to downstream effects, false negative signature results (patients incorrectly classified as not having TB) may have serious consequences, with delayed TB diagnosis resulting in Mtb transmission to close contacts, and increased risk of morbidity and mortality. While consequences are less serious, false positive results (individuals incorrectly classified as having TB) may result in costly further investigations, or 6 months of curative therapy with potential adverse effects and without apparent benefit. Incorrect diagnosis of TB may result in missed or delayed alternate diagnosis and treatment, with

potential downstream consequences. Misdiagnosis of TB may also result in stigmatisation from family and community, and psychological distress. There is no uncertainty regarding true positive and true negative results.

We found very serious risks of inconsistency, with significant unexplained heterogeneity of diagnostic sensitivity and specificity estimates for signatures in different validation cohorts and settings. No signatures consistently met the WHO TPP criteria in all or most cohorts, suggesting publication bias toward more optimistic signature performance, particularly in discovery cohorts. There was also a very serious risk of imprecision, with small sample sizes and wide confidence intervals for estimates of test accuracy among PLHIV, with no pooling of data. Data was not available or accessible for the PLHIV subgroup in numerous cohorts. In summary, the data included in this review provides very low quality evidence and we would not recommend any changes to clinical practice based on these results.

Discussion

TB transcriptomic biomarkers selected for advancement through the diagnostics pipeline for further development as point-of-care tests should ideally perform well in PLHIV. We systematically searched online databases for studies which evaluated the performance of host-blood transcriptomic signatures for diagnosing prevalent TB and identifying those who are progressing to incident TB in PLHIV, and compared performance to the WHO TPP criteria. We found 12 studies published prior to 2021 which included 10 independent test or validation cohorts featuring PLHIV, evaluating 31 transcriptomic signatures. Several of the signatures approached or met the WHO TPP minimum performance criteria for a triage test for differentiating people with prevalent TB disease from latent-Mtb infected individuals, healthy controls, or individuals with other respiratory or systemic diseases[16]. However, no transcriptomic signatures met the TPP benchmark criteria for a non-sputum confirmatory diagnostic test among PLHIV. The signatures also performed poorly for diagnosing non-microbiologically confirmed, presumptive TB disease.

Only one cohort evaluated a signature for predicting TB disease recurrence in individuals who recently completed TB treatment and initiated ART. Prognostic performance appeared to be superior proximally to incident TB disease, with highest AUC and specificity in the 3 months preceding TB recurrence[42]. The Darboe11 signature did not meet the WHO TPP for a prognostic test in any time window in this population[18].

Among the 31 signatures evaluated, we found that the genes most frequently incorporated in TB transcriptomic signatures were ISGs, which may also be upregulated by chronic HIV viraemia[62,63]. We hypothesised that these signatures would be less discriminatory for TB in PLHIV, particularly among viraemic ART-naïve individuals, due to an increased abundance of circulating type-I IFNs. While no studies performed subgroup analyses of diagnostic accuracy by HIV plasma viral load, Darboe and colleagues[42] reported higher signature scores in individuals with a viral load greater than 400 copies per

mL, as compared to those with an undetectable viral load (<400 copies/mL). Södersten and colleagues[46] demonstrated decreased Sweeney3 specificity in adults with a CD4 cell count less than 200, as compared to those with CD4 cell count greater than 200. Low CD4 cell count is a proxy for ART-naivety and high HIV plasma viral load. The lower specificity is possibly due to higher signature scores in the control group, either due to HIV viraemia, undiagnosed early or minimal TB, or other opportunistic infections. By specifically excluding ISGs, Esmail and colleagues[62] have demonstrated that classical complement pathway and Fc-γ receptor 1 (FCGR1) genes are also differentially expressed in individuals with subclinical HIV-associated TB. While traditional discriminant analysis yields an overabundance of ISGs and an underabundance of B- and T-cell genes in active TB patients versus latently Mtb-infected controls, Singhania and colleagues[53] have shown that a modular approach (i.e. pre-filtering genes by functional modules) results in a more diverse gene set.

The synthesised systematic review results represent an overall low quality of evidence, with lack of generalisability and external validity, inconsistency in results between studies, and imprecision in estimates. Most of the discovery and validation studies were conducted in Africa, particularly South Africa, limiting geographic diversity and generalisability of results. There were also few training and test datasets including PLHIV, with a notable overreliance on the Kaforou2013 and Anderson2014 datasets for signature discovery and validation, further limiting generalisability. Also, all eligible validation cohorts were from outpatient or inpatient settings. It is notable that signatures generally performed best in small test sets derived from the same population as the signature training cohort (e.g. Rajan5 in Rajan2018 test cohort, Kaforou27 in Kaforou2013a test cohort, Kaforou44 in Kaforou2013b, Anderson51 in Anderson2014b, and Penn-Nicholson6 in Penn-Nicholson2020a cohort), and performance waned in subsequent external validation. The differences in signature performance between cohorts, with signatures meeting WHO TPP criteria in one cohort but not others, is also likely attributable to differences in discovery and validation cohort designs, and publication bias. Multicohort gene meta-analytical methods, similar to those employed by Sweeney and colleagues[49], may help to overcome such limitations, and result in greater reproducibility of performance across cohorts.

There is insufficient evidence in the literature to suggest that signatures discovered in cohorts of PLHIV only (e.g. Rajan5) have greater diagnostic accuracy in populations with HIV than signatures discovered among HIV-uninfected (e.g. Darboe11, Penn-Nicholson6, Roe1, and Roe3) or mixed (e.g. Duffy10, Kaforou27, and Sweeney3) cohorts. However, there is a need for larger prospective cohorts, which include PLHIV, for signature discovery and validation, to confirm the utility of transcriptomic signatures among PLHIV. Such cohorts should enrol clearly defined, clinically relevant populations, such as symptomatic clinic attendees, HIV-infected outpatients and other high risk groups; or healthy, asymptomatic individuals from communities in diverse high incidence settings. To reduce spectrum bias, all participants should be enrolled and included in analysis, irrespective of diagnostic uncertainty. Prolonged follow-up over 3 to 6 months may aid diagnosis in unconfirmed cases. Most studies used microarray and RNA sequencing to measure signature scores. Only a few signatures were tested using near-point-of-care benchtop PCR instruments, and one with a point-of-care RT-qPCR device[46]. Future field evaluation studies should move towards technologies implementable at the point-of-care, such as the Cepheid Xpert-MTB-HR Prototype[46,64,65].

Strengths of this systematic review include the comprehensive search strategy, with rigorous eligibility criteria, and publication of a peer-reviewed study protocol. The review also had several limitations. The pre-specified literature search strategy only included studies published prior to 2021; there were few studies with data available for an PLHIV subgroup in test or validation sets in this period, and most included cohorts were small and underpowered. The use of published summary performance data is also problematic due to different statistical methods used in original papers. We were also unable to obtain summary diagnostic performance estimates, or participant-level data to reconstruct two-by-two tables, for PLHIV subgroups from 9 studies, despite contacting corresponding authors. Most papers did not conform to the standard reporting guidelines for diagnostic accuracy studies (STARD)[66], with missing information particularly relating to study design, participant recruitment, and blinding to reference standard result. Requisite anonymised participant data with signature scores were only available for a handful of studies. The lack of participant-level metadata precluded subgroup analyses. We were thus unable to systematically determine the effect of HIV viral load, CD4 cell count, TPT, and ART on transcriptomic signature diagnostic performance. Metanalysis was also deemed inappropriate due to the clinical and methodological heterogeneity and high risk of bias in cohort design and signature measurement method, with no predetermined score cut-offs or methods of standardising scores across platforms.

In the two years subsequent to the completion of this systematic literature review, diagnostic and prognostic performance of several transcriptomic signatures measured by qPCR were prospectively tested for mass screening in a South African community setting amongst predominantly asymptomatic PLHIV who were not seeking care[21,67]. Most transcriptomic signatures measured in this cohort met WHO triage test TPP benchmark criteria among symptomatic participants, however the signatures were upregulated by respiratory viral infection and HIV viraemia, and offered poor specificity for diagnosing sub-clinical TB. The Roe1 signature met the WHO prognostic test TPP benchmarks through 15 months of follow-up among PLHIV, and Darboe11 and Roe3 approached this threshold. These signatures could enable the screening of symptomatic adults seeking care and predict risk of progression to TB disease, thus enabling targeting of preventive therapy[68].

Sputum-free TB diagnostics remain an appealing prospect in populations who are unable to produce sputum, such as children and non-ambulant hospitalised patients, and among individuals with paucibacillary or negative sputum samples, such as disseminated TB disease cases or those with advanced

HIV[69–71]. Recent studies have evaluated transcriptomic signature diagnostic performance for extrapulmonary TB and TB immune reconstitution inflammatory syndrome (IRIS), both of which are difficult to diagnose, and have high morbidity and mortality among PLHIV. The Sweeney3 and the Penn-Nicholson6 signatures were shown to predict TB immune reconstitution inflammatory syndrome (IRIS) prior to ART initiation and at IRIS diagnosis in children, and paradoxical worsening of TB within half a week of starting ART in adults with TB disease[72]. The signatures were also able to distinguish IRIS in adults with TB meningitis from HIV-infected patients who did not develop IRIS after starting ART. In this context, transcriptomic signatures could potentially be used to delay ART initiation to prevent paradoxical worsening of TB. While performance is promising, these studies are generally underpowered with clinically inappropriate control groups, suffering from similar flaws in study design as discussed previously. Larger, prospective studies are required to validate transcriptomic signatures for diagnosis of disseminated TB and TB IRIS.

Additionally, most of the cohorts evaluated in this systematic review included adults, with only the Anderson2014 paediatric cohort eligible for inclusion. Paediatric TB is particularly difficult to diagnose due to its paucibacillary nature and difficulty in obtaining sputum samples from small children[73–76]. An accurate non-sputum diagnostic, such as host response transcriptomic signatures, would transform the diagnosis of TB in children[77]. However, young children are frequent vectors for respiratory and gastrointestinal viruses, and specificity of transcriptomic signatures may be low due to induction of ISG signalling from viral infections.

Transcriptomic signatures show promise for screening for prevalent TB to guide further investigations and predicting progression to incident TB for targeted TB preventive therapy[17]. However, evidence among PLHIV is limited and mostly from small case-control studies with high risk of spectrum bias. This review emphasises the need for larger heterogenous prospective discovery and validation cohorts exclusively consisting of PLHIV, or mixed cohorts which include PLHIV. Such cohorts should ideally comparatively test biomarker performance side-by-side to determine which signature should be advanced through the developmental pipeline. Further research exploring the effect of HIV clinical, virological, and immunological status is necessary for the design and implementation of TB transcriptomic signatures in this population who are at heightened risk of TB and its sequelae.

Data availability statement
All data underlying the results are available in the original published manuscripts or on request from corresponding authors. No original data are associated with this article.

Reporting guidelines
The PRISMA flow diagram and reporting checklist have been deposited in Zivahub (https://doi.org/10.25375/uct.21907296), an open access data repository hosted by the University of Cape Town's institutional data repository powered by Figshare for Institutions.

Authors' contributions
SCM and MH conceived the idea. SCM and SV independently performed the literature search, screened articles for eligibility, performed quality appraisal, and extracted data. SCM analysed the data, interpreted the results, and wrote the first draft under supervision from MH and TJS. All authors reviewed and approved the final manuscript.

References

1. WHO: **Global tuberculosis report 2022**. Geneva: World Health Organization, 2022; (accessed October 27, 2022).
   Reference Source

2. Hamada Y, Lujan J, Schenkel K, *et al.*: **Sensitivity and specificity of WHO's recommended four-symptom screening rule for tuberculosis in people living with HIV: a systematic review and meta-analysis.** *Lancet HIV.* 2018; **5**(9): e515–e23.
   PubMed Abstract | **Publisher Full Text**

3. UNAIDS: **Confronting inequalities: Lessons for pandemic responses from 40 years of AIDS. Global AIDS update 2021**. Geneva: Joint United Nations Programme on HIV/AIDS, 2021; (accessed November 4, 2021).
   Reference Source

4. WHO: **Latent tuberculosis infection: Updated and consolidated guidelines for programmatic management**. Geneva: World Health Organization, 2018; (accessed August 6, 2020).
   Reference Source

5. Ross JM, Badje A, Rangaka MX, *et al.*: **Isoniazid preventive therapy plus antiretroviral therapy for the prevention of tuberculosis: a systematic review and meta-analysis of individual participant data.** *Lancet HIV.* 2021; **8**(1): e8–e15.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

6. Barry CE 3rd, Boshoff HI, Dartois V, *et al.*: **The spectrum of latent tuberculosis: rethinking the biology and intervention strategies.** *Nat Rev Microbiol.* 2009; **7**(12): 845–55.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

7. Mahomed H, Hawkridge T, Verver S, *et al.*: **The tuberculin skin test versus QuantiFERON TB Gold® in predicting tuberculosis disease in an adolescent cohort study in South Africa.** *PLoS One.* 2011; **6**(3): e17984.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

8. Redelman-Sidi G, Sepkowitz KA: **IFN-γ release assays in the diagnosis of latent tuberculosis infection among immunocompromised adults.** *Am J Respir Crit Care Med.* 2013; **188**(4): 422–31.
   PubMed Abstract | **Publisher Full Text**

9. Mendelsohn SC, Fiore-Gartland A, Awany D, *et al.*: **Clinical predictors**

of pulmonary tuberculosis among South African adults with HIV. *EClinicalMedicine.* 2022; **45**: 101328.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

10. Churchyard G, Cárdenas V, Chihota V, *et al.*: **Annual Tuberculosis Preventive Therapy for Persons With HIV Infection : A Randomized Trial.** *Ann Intern Med.* 2021; **174**(10): 1367–76.
PubMed Abstract | **Publisher Full Text**

11. Fiore-Gartland A, Carpp LN, Naidoo K, *et al.*: **Considerations for biomarker-targeted intervention strategies for tuberculosis disease prevention.** *Tuberculosis (Edinb).* 2018; **109**: 61–8.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

12. Sumner T, Mendelsohn SC, Scriba TJ, *et al.*: **The impact of blood transcriptomic biomarker targeted tuberculosis preventive therapy in people living with HIV: a mathematical modelling study.** *BMC Med.* 2021; **19**(1): 252.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

13. Turner CT, Gupta RK, Tsaliki E, *et al.*: **Blood transcriptional biomarkers for active pulmonary tuberculosis in a high-burden setting: a prospective, observational, diagnostic accuracy study.** *Lancet Respir Med.* 2020; **8**(4): 407–19.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

14. Gupta RK, Turner CT, Venturini C, *et al.*: **Concise whole blood transcriptional signatures for incipient tuberculosis: a systematic review and patient-level pooled meta-analysis.** *Lancet Respir Med.* 2020; **8**(4): 395–406.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

15. Warsinske H, Vashisht R, Khatri P: **Host-response-based gene signatures for tuberculosis diagnosis: A systematic comparison of 16 signatures.** *PLoS Med.* 2019; **16**(4): e1002786.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

16. WHO: **High-priority target product profiles for new tuberculosis diagnostics: report of a consensus meeting**. Geneva: World Health Organization, 2014; (accessed May 22, 2020).
Reference Source

17. Mulenga H, Zauchenberger CZ, Bunyasi EW, *et al.*: **Performance of diagnostic and predictive host blood transcriptomic signatures for Tuberculosis disease: A systematic review and meta-analysis.** *PLoS One.* 2020; **15**(8): e0237574.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

18. WHO: **Consensus Meeting Report: Development of a Target Product Profile (TPP) and a framework for evaluation for a test for predicting progression from tuberculosis infection to active disease**. Geneva: World Health Organization, 2017; (accessed May 22, 2020).
Reference Source

19. Warsinske HC, Rao AM, Moreira FMF, *et al.*: **Assessment of Validity of a Blood-Based 3-Gene Signature Score for Progression and Diagnosis of Tuberculosis, Disease Severity, and Treatment Response.** *JAMA Netw Open.* 2018; **1**(6): e183779.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

20. Scriba TJ, Fiore-Gartland A, Penn-Nicholson A, *et al.*: **Biomarker-guided tuberculosis preventive therapy (CORTIS): a randomised controlled trial.** *Lancet Infect Dis.* 2021; **21**(3): 354–65.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

21. Mendelsohn SC, Mbandi SK, Fiore-Gartland A, *et al.*: **Prospective multicentre head-to-head validation of host blood transcriptomic biomarkers for pulmonary tuberculosis by real-time PCR.** *Commun Med (Lond).* 2022; **2**(1): 26.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

22. Roe JK, Thomas N, Gil E, *et al.*: **Blood transcriptomic diagnosis of pulmonary and extrapulmonary tuberculosis.** *JCI Insight.* 2016; **1**(16): e87238.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

23. Roe J, Venturini C, Gupta RK, *et al.*: **Blood Transcriptomic Stratification of Short-term Risk in Contacts of Tuberculosis.** *Clin Infect Dis.* 2020; **70**(5): 731–7.
PubMed Abstract | **Publisher Full Text**

24. McInnes MDF, Moher D, Thombs BD, *et al.*: **Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement.** *JAMA.* 2018; **319**(4): 388–96.
PubMed Abstract | **Publisher Full Text**

25. Mendelsohn SC, Mulenga H, Mbandi SK, *et al.*: **Host blood transcriptomic biomarkers of tuberculosis disease in people living with HIV: a systematic review protocol.** *BMJ Open.* 2021; **11**(8): e048623.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

26. Greenhalgh T, Peacock R: **Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources.** *BMJ.* 2005; **331**(7524): 1064–5.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

27. Peters MDJ: **Managing and Coding References for Systematic Reviews and Scoping Reviews in EndNote.** *Med Ref Serv Q.* 2017; **36**(1): 19–31.
PubMed Abstract | **Publisher Full Text**

28. Gordon M, Lumley T: **forestplot: Advanced Forest Plot Using 'grid' Graphics (Version 2.0.1)**. 2021; [Accessed August 22, 2022].
Reference Source

29. Robin X, Turck N, Hainard A, *et al.*: **pROC: an open-source package for R and S+ to analyze and compare ROC curves.** *BMC Bioinformatics.* 2011; **12**: 77.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

30. DeLong ER, DeLong DM, Clarke-Pearson DL: **Comparing the areas under

31. two or more correlated receiver operating characteristic curves: a nonparametric approach.** *Biometrics.* 1988; **44**(3): 837–45.
PubMed Abstract | **Publisher Full Text**

31. Wilson EB: **Probable Inference, the Law of Succession, and Statistical Inference.** *J Am Stat Assoc.* 1927; **22**(158): 209–12.
Publisher Full Text

32. Whiting PF, Rutjes AW, Westwood ME, *et al.*: **QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies.** *Ann Intern Med.* 2011; **155**(8): 529–36.
PubMed Abstract | **Publisher Full Text**

33. McGuinness L: **robvis: Visualize the Results of Risk-of-Bias (ROB) Assessments (Version 0.3.0)**. 2019; [Accessed August 22, 2022].
Reference Source

34. Guyatt GH, Oxman AD, Vist GE, *et al.*: **GRADE: an emerging consensus on rating quality of evidence and strength of recommendations.** *BMJ.* 2008; **336**(7650): 924–6.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

35. Schünemann HJ, Oxman AD, Brozek J, *et al.*: **Grading quality of evidence and strength of recommendations for diagnostic tests and strategies.** *BMJ.* 2008; **336**(7653): 1106–10.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

36. Schünemann HJ, Mustafa RA, Brozek J, *et al.*: **GRADE guidelines: 21 part 1. Study design, risk of bias, and indirectness in rating the certainty across a body of evidence for test accuracy.** *J Clin Epidemiol.* 2020; **122**: 129–41.
PubMed Abstract | **Publisher Full Text**

37. Kaforou M, Wright VJ, Oni T, *et al.*: **Detection of tuberculosis in HIV-infected and -uninfected African adults using whole blood RNA expression signatures: a case-control study.** *PLoS Med.* 2013; **10**(10): e1001538.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

38. Anderson ST, Kaforou M, Brent AJ, *et al.*: **Diagnosis of childhood tuberculosis and host RNA expression in Africa.** *N Engl J Med.* 2014; **370**(18): 1712–23.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

39. Walter ND, Miller MA, Vasquez J, *et al.*: **Blood Transcriptional Biomarkers for Active Tuberculosis among Patients in the United States: a Case-Control Study with Systematic Cross-Classifier Evaluation.** *J Clin Microbiol.* 2016; **54**(2): 274–82.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

40. Zak DE, Penn-Nicholson A, Scriba TJ, *et al.*: **A blood RNA signature for tuberculosis disease risk: a prospective cohort study.** *Lancet.* 2016; **387**(10035): 2312–22.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

41. Sambarey A, Devaprasad A, Mohan A, *et al.*: **Unbiased Identification of Blood-based Biomarkers for Pulmonary Tuberculosis by Modeling and Mining Molecular Interaction Networks.** *EBioMedicine.* 2017; **15**: 112–26.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

42. Darboe F, Mbandi SK, Naidoo K, *et al.*: **Detection of Tuberculosis Recurrence, Diagnosis and Treatment Response by a Blood Transcriptomic Risk Signature in HIV-Infected Persons on Antiretroviral Therapy.** *Front Microbiol.* 2019; **10**: 1441.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

43. Duffy FJ, Thompson EG, Scriba TJ, *et al.*: **Multinomial modelling of TB/HIV co-infection yields a robust predictive signature and generates hypotheses about the HIV+TB+ disease state.** *PLoS One.* 2019; **14**(7): e0219322.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

44. Rajan JV, Semitala FC, Mehta T, *et al.*: **A Novel, 5-Transcript, Whole-blood Gene-expression Signature for Tuberculosis Screening Among People Living With Human Immunodeficiency Virus.** *Clin Infect Dis.* 2019; **69**(1): 77–83.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

45. Penn-Nicholson A, Mbandi SK, Thompson E, *et al.*: **RISK6, a 6-gene transcriptomic signature of TB disease risk, diagnosis and treatment response.** *Sci Rep.* 2020; **10**(1): 8629.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

46. Södersten E, Ongarello S, Mantsoki A, *et al.*: **Diagnostic Accuracy Study of a Novel Blood-Based Assay for Identification of Tuberculosis in People Living with HIV.** *J Clin Microbiol.* 2021; **59**(3): e01643.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

47. Rajan J, Deng X, Seielstad M, *et al.*: **Performance Of Gene Expression Signatures In The Context Of Intensified Tuberculosis Case Finding Among People Living With HIV (PLHIV).** *Am J Respir Crit Care Med.* 2017; **195**.

48. Maertzdorf J, McEwen G, Weiner J 3rd, *et al.*: **Concise gene signature for point-of-care classification of tuberculosis.** *EMBO Mol Med.* 2016; **8**(2): 86–95.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

49. Sweeney TE, Braviak L, Tato CM, *et al.*: **Genome-wide expression for diagnosis of pulmonary tuberculosis: a multicohort analysis.** *Lancet Respir Med.* 2016; **4**(3): 213–24.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

50. Qian Z, Lv J, Kelly GT, *et al.*: **Expression of nuclear factor, erythroid 2-like 2-mediated genes differentiates tuberculosis.** *Tuberculosis (Edinb).* 2016; **99**: 56–62.
PubMed Abstract | **Publisher Full Text**

51. Gjøen JE, Jenum S, Sivakumaran D, *et al.*: **Novel transcriptional signatures for sputum-independent diagnostics of tuberculosis in children.** *Sci Rep.* 2017; **7**(1): 5839.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

52. Verma S, Du P, Nakanjako D, *et al.*: **Tuberculosis in advanced HIV infection is associated with increased expression of IFNγ and its downstream targets.** *BMC Infect Dis.* 2018; **18**(1): 220.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

53. Singhania A, Verma R, Graham CM, *et al.*: **A modular transcriptional signature identifies phenotypic heterogeneity of human tuberculosis infection.** *Nat Commun.* 2018; **9**(1): 2308.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

54. Tornheim JA, Madugundu AK, Paradkar M, *et al.*: **Transcriptomic Profiles of Confirmed Pediatric Tuberculosis Patients and Household Contacts Identifies Active Tuberculosis, Infection, and Treatment Response Among Indian Children.** *J Infect Dis.* 2020; **221**(10): 1647–58.

55. Gliddon HD, Kaforou M, Alikian M, *et al.*: **Identification of Reduced Host Transcriptomic Signatures for Tuberculosis Disease and Digital PCR-Based Validation and Quantification.** *Front Immunol.* 2021; **12**: 637164.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

56. Darboe F, Mbandi SK, Thompson EG, *et al.*: **Diagnostic performance of an optimized transcriptomic signature of risk of tuberculosis in cryopreserved peripheral blood mononuclear cells.** *Tuberculosis (Edinb).* 2018; **108**: 124–6.
PubMed Abstract | **Publisher Full Text**

57. Samarajiwa SA, Forster S, Auchettl K, *et al.*: **INTERFEROME: the database of interferon regulated genes.** *Nucleic Acids Res.* 2009; **37**(Database issue): D852–7.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

58. Berry MPR, Graham CM, McNab FW, *et al.*: **An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis.** *Nature.* 2010; **466**(7309): 973–7.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

59. de Araujo LS, Vaas LAI, Ribeiro-Alves M, *et al.*: **Transcriptomic Biomarkers for Tuberculosis: Evaluation of *DOCK9*. *EPHA4*, and *NPC2* mRNA Expression in Peripheral Blood.** *Front Microbiol.* 2016; **7**: 1586.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

60. Huang HH, Liu XY, Liang Y, *et al.*: **Identification of 13 blood-based gene expression signatures to accurately distinguish tuberculosis from other pulmonary diseases and healthy controls.** *Biomed Mater Eng.* 2015; **26 Suppl 1**: S1837–43.
PubMed Abstract | **Publisher Full Text**

61. Suliman S, Thompson EG, Sutherland J, *et al.*: **Four-Gene Pan-African Blood Signature Predicts Progression to Tuberculosis.** *Am J Respir Crit Care Med.* 2018; **197**(9): 1198–208.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

62. Esmail H, Lai RP, Lesosky M, *et al.*: **Complement pathway gene activation and rising circulating immune complexes characterize early disease in HIV-associated tuberculosis.** *Proc Natl Acad Sci U S A.* 2018; **115**(5): E964–E73.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

63. Lazear HM, Schoggins JW, Diamond MS: **Shared and Distinct Functions of Type I and Type III Interferons.** *Immunity.* 2019; **50**(4): 907–23.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

64. Moreira FMF, Verma R, Pereira Dos Santos PC, *et al.*: **Blood-based host biomarker diagnostics in active case finding for pulmonary tuberculosis: A diagnostic case-control study.** *EClinicalMedicine.* 2021; **33**: 100776.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

65. Sutherland JS, van der Spuy G, Gindeh A, *et al.*: **Diagnostic Accuracy of the Cepheid 3-gene Host Response Fingerstick Blood Test in a Prospective, Multi-site Study: Interim Results.** *Clin Infect Dis.* 2022; **74**(12): 2136–41.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

66. Bossuyt PM, Reitsma JB, Bruns DE, *et al.*: **STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies.** *BMJ.* 2015; **351**: h5527.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

67. Mendelsohn SC, Fiore-Gartland A, Penn-Nicholson A, *et al.*: **Validation of a host blood transcriptomic biomarker for pulmonary tuberculosis in people living with HIV: a prospective diagnostic and prognostic accuracy study.** *Lancet Glob Health.* 2021; **9**(6): e841–e53.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

68. Sumner T, Mendelsohn SC, Scriba TJ, *et al.*: **The impact of blood transcriptomic biomarker targeted tuberculosis preventive therapy in people living with HIV: a mathematical modelling study.** *BMC Med.* 2021; **19**(1): 252.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

69. Walzl G, McNerney R, du Plessis N, *et al.*: **Tuberculosis: advances and challenges in development of new diagnostics and biomarkers.** *Lancet Infect Dis.* 2018; **18**(7): e199–e210.
PubMed Abstract | **Publisher Full Text**

70. Tebruegge M, Ritz N, Curtis N, *et al.*: **Diagnostic Tests for Childhood Tuberculosis: Past Imperfect, Present Tense and Future Perfect?** *Pediatr Infect Dis J.* 2015; **34**(9): 1014–9.
PubMed Abstract | **Publisher Full Text**

71. Sabur NF, Esmail A, Brar MS, *et al.*: **Diagnosing tuberculosis in hospitalized HIV-infected individuals who cannot produce sputum: is urine lipoarabinomannan testing the answer?** *BMC Infect Dis.* 2017; **17**(1): 803.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

72. Mbandi SK, Painter H, Penn-Nicholson A, *et al.*: **Host transcriptomic signatures of tuberculosis can predict immune reconstitution inflammatory syndrome in HIV patients.** *Eur J Immunol.* 2022; **52**(7): 1112–9.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

73. Togun T, Kampmann B, Pai M: **Diagnosis of Childhood Tuberculosis.** *Reference Module in Biomedical Sciences.* 2017.

74. Graham SM, Cuevas LE, Jean-Philippe P, *et al.*: **Clinical Case Definitions for Classification of Intrathoracic Tuberculosis in Children: An Update.** *Clin Infect Dis.* 2015; **61Suppl 3**(Suppl 3): S179–87.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

75. Nicol MP, Zar HJ: **New specimens and laboratory diagnostics for childhood pulmonary TB: progress and prospects.** *Paediatr Respir Rev.* 2011; **12**(1): 16–21.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

76. Zar HJ, Connell TG, Nicol M: **Diagnosis of pulmonary tuberculosis in children: new advances.** *Expert Rev Anti Infect Ther.* 2010; **8**(3): 277–88.
PubMed Abstract | **Publisher Full Text**

77. Togun TO, MacLean E, Kampmann B, *et al.*: **Biomarkers for diagnosis of childhood tuberculosis: A systematic review.** *PLoS One.* 2018; **13**(9): e0204029.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 2

Reviewer Report 09 May 2023

https://doi.org/10.21956/gatesopenres.16010.r33155

✔ **Theresa Ryckman** 🆔

Bloomberg School of Public Health, Department of Epidemiology, Johns Hopkins University, Baltimore, Maryland, USA

I approve of the manuscript and have no further comments.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Infectious disease epidemiology and policy

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 09 May 2023

https://doi.org/10.21956/gatesopenres.16010.r33154

✔ **Nagasuma Chandra** 🆔

Department of Biochemistry and Center for BioSystems Science and Engineering (BSSE), Indian Institute of Science, Bengaluru, Karnataka, India

No further comments

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Systems Biology, Tuberculosis, Genomics and Bioinformatics, Diagnostic

biomarkers

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 26 April 2023

✔ **Nagasuma Chandra** [iD]

[1] Department of Biochemistry and Center for BioSystems Science and Engineering (BSSE), Indian Institute of Science, Bengaluru, Karnataka, India
[2] Department of Biochemistry and Center for BioSystems Science and Engineering (BSSE), Indian Institute of Science, Bengaluru, Karnataka, India

The authors report a review of diagnostic and prognostic host-blood transcriptomic signatures of TB in HIV background. They study 31 transcriptomic signatures collated from multiple studies reported for differentiating samples of HIV with HIV co-infections with active or latent TB as well as with other diseases.

It is indeed a very important topic as there is a great need for accurate diagnosis of TB in HIV background, so that appropriate treatment can be provided, and the spread of TB can be checked. The need for the analysis is well brought out in the review.

The authors have conducted a systematic survey for relevant transcriptomic data reported over a period of 30 years and use well-defined criteria for inclusion and exclusion of the signatures, and list clear objectives and evaluation metrics and also quality assessment of the diagnostic accuracy. The manuscript is overall well written. It would be helpful if the authors could elaborate on the following points, so as to make it more accessible to a wider audience.

1. A brief note could be added on the WHO TPP minimum performance criteria, and why some datasets did not qualify - especially to bring in the difference between 'lack of required data/comparisons' in the original study versus 'signatures not performing well enough'.

2. A brief introduction on the risk of bias and the various domains under which it is assessed and how it is addressed.

3. The statement "*Overall 9 of the 10 cohorts had a high risk of bias, and one study (Turner2020) had some concerns*" could be followed by some recommendation on how to read meaning from these studies. It would also help if the term 'concerns' were explained and placed in

the risk spectrum.

4. As some studies include more than one sub-objective of diagnosis/prognosis, it would help if a note is added in discussion on the diagnostic accuracy/potential for each comparison axis separately.

Are the rationale for, and objectives of, the Systematic Review clearly stated?
Yes

Are sufficient details of the methods and analysis provided to allow replication by others?
Yes

Is the statistical analysis and its interpretation appropriate?
Yes

Are the conclusions drawn adequately supported by the results presented in the review?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Systems Biology, Tuberculosis, Genomics and Bioinformatics, Diagnostic biomarkers

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 29 Apr 2023
Simon Mendelsohn

Thank you for your thoughtful review of our manuscript. We appreciate your positive feedback and suggestions for improvement. Please find our response to each of your points below:

1. We include the following note on the WHO TPP minimum performance criteria in the introduction: "… the minimum World Health Organization (WHO) Target Product Profile (TPP) TB triage test performance criteria (sensitivity 90% and specificity 70%) for diagnosing prevalent TB" and "…minimum prognostic benchmarks (sensitivity 75% and specificity 75%)".

Regarding why "some datasets did not qualify - especially to bring in the difference between 'lack of required data/comparisons'": per published systematic review protocol eligibility criteria (see Mendelsohn, BMJ Open, 2021), we excluded datasets which did not include PLHIV, or where it was not possible to stratify results by HIV status, where only signature discovery performance was reported (no independent validation), or where the controls or cases were not adequately defined. We also excluded studies where we were not able to retrieve participant level data, or calculated signature scores.

We did not exclude datasets on the basis of "signatures not performing well enough".

2. We appreciate your suggestion to include a brief introduction on the risk of bias and how it is assessed. We will revise the "Risk of bias, applicability, and quality of evidence" Methods section to provide clarity.

3. As per the point above, we will revise the "Risk of bias, applicability, and quality of evidence" Methods section to provide clarity. We will also add clarity as to why the Turner study had some concerns: "due to lack of a pre-specified test threshold prospectively applied to each signature".

4. We have considered diagnostic (paragraph 1) and prognostic (paragraph 2) performance of the signatures separately in the Discussion.

Thank you again for your valuable feedback. We will incorporate your suggestions into the revised manuscript to improve the clarity and accessibility of our findings.

Competing Interests: No competing interests were disclosed.

Reviewer Report 02 March 2023

✔ **Theresa Ryckman** [iD]
[1] Bloomberg School of Public Health, Department of Epidemiology, Johns Hopkins University, Baltimore, Maryland, USA
[2] Bloomberg School of Public Health, Department of Epidemiology, Johns Hopkins University, Baltimore, Maryland, USA

This study systematically reviewed publications that assessed the performance of host blood transcriptomic signatures in diagnosing active TB and/or predicting risk of progression to active TB among PLHIV. The study found that most publications were at high risk of bias, and that few signatures met the WHO criteria for a triage test for diagnosis (and none met the criteria for a prognostic test), with variability in signature performance across different cohorts of PLHIV. The study is very clearly written, and in particular the methods were well-described and easy to follow. I have only a few minor suggestions for the authors.

○ In the methods, it would be useful to clarify how the risk of bias across each of the 4 domains was translated into the "overall risk of bias" for each study as in Figure 2 (as I don't believe a method for arriving at this summary measure of bias is clearly defined in QUADAS-2?)

○ Methods: In the quality appraisal results/figure 2, it seems as though exclusion of some patients from analysis was evaluated in both the "patient selection" and "study flow and timing" domains. It would be good to clarify to what extent there was overlap across the considerations included in these 2 domains (e.g., if a patient was excluded due to diagnostic uncertainty, was this exclusion assessed only under "patient selection", or also under "study flow and timing"?), and to justify this overlap (if there was any).

○ For signatures that were evaluated across multiple cohorts of the same type (for example, Sweeney3 was evaluated in 3 cohorts in Figure 4a), given that the authors have deemed a meta-analysis not appropriate, it could be useful to expand on how the signatures may have met WHO TPP criteria under 1 cohort but not others. This is described only briefly in the last paragraph of the results. Were there any signatures tested in multiple cohorts that met the criteria in all or most cohorts? If there were any, it seems useful to call these out (despite the acknowledged risks of bias). If there weren't any, then this also seems to be an important finding (e.g., even with all the risks of bias - that mostly seem to bias studies in the direction of more optimistic performance - there weren't any tests that consistently met the criteria).

○ In the last paragraph – large heterogeneous prospective discovery and validation cohorts that include PLHIV seem like the obvious gold standard. But such studies are also expensive. Do you think there is a case to be made for discovery and initial validation using smaller cohorts/existing microarray data, and then only proceeding to larger studies if these smaller studies show promise? Given the poor performance of many of the signatures, even in the smaller studies reviewed in this paper (if again, the authors generally deem these smaller studies to result on overly optimistic results), it could probably be argued that in most cases, initial validation of a signature in a large prospective cohort will be a waste of money, because that signature is unlikely to perform well. I just wonder if the findings from this study can point the research community in a more strategic direction while still arguing that evaluation in larger prospective cohorts is probably needed before any signature should be adopted into clinical practice.

Are the rationale for, and objectives of, the Systematic Review clearly stated?
Yes

Are sufficient details of the methods and analysis provided to allow replication by others?
Yes

Is the statistical analysis and its interpretation appropriate?
Yes

Are the conclusions drawn adequately supported by the results presented in the review?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Infectious disease epidemiology and policy

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 29 Apr 2023

Simon Mendelsohn

Thank you very much for taking the time to carefully review our study and for your constructive feedback. We appreciate your positive comments regarding the clarity of our writing and the description of our methods.

1. Regarding your first suggestion, we agree that we could have been clearer about how we arrived at the overall risk of bias for each study in Figure 2. We will revise the "Risk of bias, applicability, and quality of evidence" Methods section to provide clarity.

2. Regarding your second comment, we tried to avoid overlap between the "Patient selection" and "Study flow and timing" domains in QUADAS-2. For the "Patient selection" domain, we specifically asked "Did the study avoid inappropriate exclusions?" per the eligibility criteria (i.e. were the eligibility criteria appropriate for the study design). For example, did the study inappropriately exclude individuals with comorbidities, such as diabetes, or individuals who use drugs.

For the "Study flow and timing" domain, we asked whether "...all [eligible] patients [were] included in the analysis?" (i.e. was there inappropriate exclusion of eligible/enrolled participants from analysis due to, for example, inconclusive diagnostic test results/diagnostic uncertainty or loss to follow up). However, we acknowledge that there is some subjectivity in this process.

3. Regarding your third suggestion, we appreciate your suggestion to expand on the performance of signatures across multiple cohorts. We did not identify any signatures that consistently met the WHO TPP criteria in all or most cohorts. We agree that this may suggest bias toward more optimistic performance. It is notable that signatures generally performed best in small test sets derived from the same population as the signature training cohort (e.g. Rajan5 in Rajan2018 test cohort, Kaforou27 in Kaforou2013a test cohort, Kaforou44 in Kaforou2013b, Anderson51 in Anderson2014b, and Penn-Nicholson6 in Penn-Nicholson2020a cohort), and performance waned in subsequent external validation. The differences in signature performance between cohorts, with signatures meeting WHO TPP criteria in one cohort but not others, is also likely attributable to differences in discovery and validation cohort designs. Multicohort gene meta-analytical discovery and validation methods, similar to those employed by Sweeney and colleagues, may help to overcome such limitations.

We will expand on the above in the Results and Discussion sections.

4. Regarding your final suggestion, we agree that large, heterogeneous prospective cohorts that include PLHIV are the gold standard for discovering and validating host blood transcriptomic signatures for TB diagnosis and risk prediction. We also acknowledge that such studies are expensive. This review suggests that transcriptomic/gene signatures

discovered in smaller studies are unlikely to be reproducible across multiple distinct cohorts. We agree that using multicohort gene meta-analytical methods in existing microarray or sequencing data is a more pragmatic use of resources, and may result in greater reproducibility of performance across new validation cohorts. However, we argue that there is insufficient existing microarray or sequencing data among people living with HIV, hence the need for large prospective cohorts in this population.

Thank you again for your thoughtful review and helpful suggestions. We will address these points in our revised manuscript.

Competing Interests: No competing interests were disclosed.