

1 **Title page**

2 **Title**

3 Single-gene transcripts for subclinical TB: an individual participant data meta-analysis

4 **Authors**

5 James Greenan-Barrett¹, Simon C. Mendelsohn², Thomas J. Scriba², Mahdad Noursadeghi^{*3}, Rishi K.
6 Gupta^{*1}

7 *Co-senior authors

8 **Affiliations**

9 1. UCL Respiratory, Division of Medicine, University College London, London, United Kingdom

10 2. South African Tuberculosis Vaccine Initiative, Division of Immunology, Department of Pathology
11 and Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town,
12 South Africa

13 3. UCL Division of Infection and Immunity, University College London, London, United Kingdom

14 **Correspondence**

15 Dr Rishi K Gupta, UCL Respiratory, 1st Floor Rayne Building, 5 University Street, London, WC1E 6JF,
16 United Kingdom. r.gupta@ucl.ac.uk

17 **Abstract**

18 **Background:**

19 Translation of blood RNA signatures may be accelerated by identifying more parsimonious biomarkers.
20 We tested the hypothesis that single-gene transcripts provide comparable accuracy for detection of
21 subclinical TB to multi-gene signatures and benchmarked their clinical utility to interferon- γ release
22 assays (IGRAs).

23 **Methods:**

24 We identified datasets where participants underwent RNA sampling and at least 12 months of follow-
25 up for progression to TB. We performed a one-stage individual participant data meta-analysis to
26 compare multi-gene signatures against single-gene transcripts to detect subclinical TB, defined as
27 asymptomatic prevalent or incident TB (diagnosed ≥ 21 days from enrolment, irrespective of symptoms)
28 over a 12-month interval. We performed decision curve analysis to evaluate the net benefit of using
29 RNA biomarkers and IGRA, alone or in combination, compared to treating all or no individuals with
30 preventative treatment.

31 **Results:**

32 We evaluated 80 single-genes and eight multi-gene signatures in a pooled analysis of four RNAseq and
33 three qPCR datasets, comprising 6544 total samples and including 283 samples from 214 individuals
34 with subclinical TB. Five single-gene transcripts were equivalent to the best-performing multi-gene
35 signature over 12 months, with areas under the receiver operating characteristic curves ranging from
36 0.75-0.77, but none met the WHO minimum target product profile (TPP) for a TB progression test. IGRA
37 demonstrated much lower specificity in higher burden settings, while sensitivity and specificity of RNA
38 biomarkers were consistent across settings. In higher burden settings, RNA biomarkers had greater net
39 benefit than IGRA, which offered little clinical utility over treating all with preventative therapy. In low
40 burden settings, IGRA approximated the TPP and offered greater clinical utility than RNA biomarkers,
41 but combining both tests provided the highest net benefit for services aiming to treat <50 people to
42 prevent a single case.

43 **Interpretation:**

44 Single-gene transcripts are equivalent to multi-gene signatures for detection of subclinical TB, with
45 consistent performance across settings. Single transcripts demonstrate potential clinical utility to stratify
46 treatment, particularly when used in combination with IGRA in low burden settings.

47 **Funding:**

48 National Institute for Health Research, Wellcome Trust.

49

50 **Research in Context**

51 ***Evidence before this study***

52 With increasing recognition of the spectrum of tuberculosis and the importance of subclinical TB, the
53 International Consensus for Early TB (ICE-TB) group recently developed a framework to classify
54 disease states and called for diagnostic development to detect the subclinical states. We performed a
55 systematic search of PubMed from inception to June 10, 2024 using terms for ‘tuberculosis’, ‘subclinical’
56 and ‘RNA’, without language restrictions. Multiple studies have discovered or validated multi-gene blood
57 RNA signatures for TB, including to discriminate individuals who progress to clinical TB from non-
58 progressors. In our previous head-to-head evaluation of these signatures, we demonstrated that eight
59 signatures had equivalent performance to predict progression to TB. More parsimonious biomarkers,
60 such as single-gene transcripts, may facilitate clinical translation. However, no previous studies have
61 systematically compared the performance of single-gene transcripts to multi-gene signatures.
62 Moreover, the clinical utility of RNA biomarkers to detect subclinical TB and guide treatment decisions,
63 compared to alternative strategies including using interferon-gamma release assays (IGRAs), remains
64 untested.

65 ***Added value of this study***

66 To our knowledge, this is the largest pooled analysis of RNA biomarkers to predict progression to clinical
67 TB, the first head-to-head comparison of single-gene transcripts to multi-gene signatures and the first
68 RNA analysis to align with the ICE-TB definitions. We tested 80 single-genes and eight multi-gene
69 signatures to detect subclinical TB in a pooled dataset from four RNAseq and three qPCR datasets,
70 comprising over 6500 RNA samples. We show that five co-correlated single-gene transcripts were
71 equivalent to the best-performing multi-gene signature to detect subclinical TB over a 12-month interval,
72 but none met the WHO minimum target product profile (TPP) for a TB progression test. Discriminative
73 performance of single-gene transcripts waned over increasing time intervals from sampling. BATF2, the
74 transcript with the highest discrimination point estimate, showed consistent performance across
75 settings. In contrast, IGRA performance was heterogenous, approximating the WHO minimum TPP in
76 low TB burden settings, but demonstrating poor specificity in higher burden settings. Decision curve
77 analysis comparing the clinical utility of BATF2 and IGRA showed that BATF2 offered highest net benefit
78 in high burden settings. In low burden settings, IGRA offered greater net benefit, while a two-step
79 approach combining both tests achieved the greatest net benefit for services aiming to treat fewer than
80 50 people to prevent a TB case.

81 ***Implications of all the available evidence***

82 Single-gene transcripts perform as well as multi-gene signatures for subclinical TB. These findings may
83 simplify RNA biomarker testing, encourage commercial competition and facilitate translation of this
84 technology to clinical practice. RNA biomarkers demonstrate clinical utility to direct treatment decisions,
85 as a stand-alone test in high burden settings and in combination with IGRA in low burden settings.
86 Further interventional studies are required to evaluate the clinical and cost-effectiveness of serial RNA
87 biomarker testing to stratify delivery of preventative therapy, for example among high-risk contacts in
88 high burden countries, or using a two-step testing approach in combination with IGRA in low burden
89 settings.

90

91 Introduction

92 Despite global efforts, tuberculosis (TB) remains a leading cause of morbidity and mortality worldwide,
93 causing 10.6 million cases and 1.3 million deaths in 2022, with a disproportionate burden on
94 disadvantaged communities^{1,2}. In response to increased recognition of the spectrum of TB, the
95 International Consensus for Early TB (ICE-TB) group recently developed a framework to classify
96 disease states. The ICE-TB framework divides TB disease into subclinical or clinical based on signs
97 and symptoms, with further subdivisions into infectious and non-infectious based on detection of
98 aerosolised or expectorated *M tuberculosis*³. Targeting the subclinical disease state, potentially with
99 truncated treatment regimens⁴, may prevent progression to clinical disease and reduce risk of onward
100 transmission.

101 Current prognostic tests for TB, such as the tuberculin skin test (TST) and interferon- γ release assay
102 (IGRA), have low positive predictive values (PPVs) for progression to clinical disease⁵⁻⁷, resulting in
103 unnecessary TB preventative therapy (TPT) for most individuals. This is financially burdensome in
104 healthcare systems and adverse effects are estimated to affect 3.7% of TPT-treated individuals⁸.
105 Measurement of blood RNA levels can be used to detect changes in host gene expression in response
106 to TB disease. Multiple RNA signatures have been discovered that have promising diagnostic accuracy
107 for clinical TB⁹⁻²⁰ or to predict progression to clinical disease²¹⁻²⁴. In a previous analysis comparing the
108 performance of 17 RNA signatures to predict progression to clinical disease, we demonstrated that
109 eight signatures performed equivalently, were co-correlated and shared common upstream pathways²⁵.

110 Development of near-patient, cartridge-based prototype platforms, such as the Cepheid MTB-Host
111 Response (MTB-HR) assay have paved the way for clinical translation of blood RNA signatures.
112 However, this progress may be further accelerated by further simplification of multi-gene signatures to
113 single-gene biomarkers. In view of the co-correlation and common regulators of the genes that comprise
114 the most accurate signatures to date, multiple genes may not offer orthogonal value. In this study, we
115 used a pooled dataset of studies with blood RNA sampling and longitudinal follow-up for TB to
116 hypothesise that single-gene transcripts will perform as well as multi-gene signatures for detection of
117 subclinical TB. We also sought to benchmark the diagnostic performance and clinical utility of RNA
118 biomarkers to IGRA, stratified by TB burden.

119 Methods

120 *Data sources and preparation*

121 We performed a systematic search to identify datasets where participants underwent whole blood RNA
122 sampling with at least 12 months of follow-up for development of clinical TB (supplementary methods).
123 Studies using genome-wide (RNAseq or microarray) or targeted transcriptional profiling (qPCR or
124 NanoString quantification) were included. We included four RNAseq datasets from our previous
125 individual participant data meta-analysis (IPD-MA)²⁵ and three subsequent studies using qPCR²⁶⁻²⁸. All
126 included RNA datasets were publicly available.

127 The four previously included RNAseq datasets were mapped, batch corrected, and integrated into a
128 single pooled dataset using transcripts per million measurements, as previously described²⁵. Our data
129 preparation pipeline for qPCR studies is described in the supplementary methods. We included 80
130 single-genes that were present in the RNAseq dataset and at least one qPCR study (CORTIS-
131 01/HR^{26,27} or REPORT-Brazil²⁸). We calculated scores for eight existing RNA signatures that were
132 included in our previous analysis (supplementary methods)²⁹. We standardised signatures and single-
133 gene transcripts within each RNAseq and qPCR dataset by converting to z-scores (supplementary
134 methods), before combining the z-score transformed datasets into a pooled dataset.

135 *Outcome definitions*

136 We used the original study definitions as a reference standard for TB in each cohort. We sought to align
137 our terminology with the ICE-TB consensus classification³. Clinical TB was defined as symptomatic
138 prevalent TB cases. Subclinical TB was defined as asymptomatic prevalent TB cases (approximates to
139 'subclinical TB, infectious') or incident TB cases (approximates to 'subclinical TB, non-infectious'). The
140 'subclinical, non-infectious' state is based on the presence of "macroscopic pathology", for which there

141 is no gold standard. We therefore used progression to incident TB as a reference standard to
142 approximate this, based on the assumption that macroscopic pathology would have been detectable at
143 the time of blood RNA sampling if investigated with sufficient resolution. Prevalent TB was defined as
144 a TB diagnosis made less than 21 days after RNA sampling whereas incident TB was defined as a TB
145 diagnosis ≥ 21 days. Non-progressors were defined as individuals who remained TB-free during follow-
146 up.

147 **Analysis**

148 All analyses were done using R (version 4.4.0). We performed a one-stage IPD-MA to calculate the
149 accuracy of candidate signatures and transcripts to discriminate subclinical TB from non-progressors,
150 stratified by interval from sampling to disease. Since our initial analyses demonstrated similar accuracy
151 of RNA biomarkers for each study, our primary analysis assumed common accuracy across studies, as
152 previously²⁵. The primary analysis was over a 12-month interval from sampling, with further stratified
153 analyses over 0-3, 0-6, 0-15, 6-12, and 12-15 months. For the primary analysis, non-progressor
154 samples with less than 12 months of follow-up from sampling were excluded. We excluded individuals
155 who received TPT from the primary analysis since this affects progression risk and may be differential
156 among those with higher, compared to lower, RNA biomarker scores. Where datasets included serial
157 samples from the same individuals, we considered serial samples as independent since intra-individual
158 variance was similar to inter-individual variance (supplementary figure 7). Where candidate signatures
159 were originally derived from included datasets, we excluded these datasets when evaluating the
160 accuracy of that signature.

161 Accuracy of candidate signatures and transcripts was quantified by the area under the receiver
162 operating curve (AUROC) with 95% confidence intervals (CIs). Sensitivity and specificity were
163 calculated at the maximum Youden index, giving equal weighting to sensitivity and specificity, and
164 benchmarked against the WHO minimum Target Product Profile (TPP) parameters for predicting
165 progression to TB over 2 years ($\geq 75\%$ sensitivity and $\geq 75\%$ specificity)^{04/07/2024 14:57:00}. AUROCs
166 of single-gene transcripts were compared to the best performing multi-gene signature using the pairwise
167 Delong test, with multiple testing correction using the Benjamini-Hochberg approach. Signatures and
168 transcripts with adjusted p values > 0.05 were considered equivalent. Correlation of equivalent
169 transcripts was assessed using Spearman rank correlation. We also evaluated expression and
170 AUROCs of the best-performing single gene transcript across different disease states; clinical TB,
171 'subclinical TB, infectious', and 'subclinical TB, non-infectious'.

172 We then compared the diagnostic performance and clinical utility of the single-gene transcript with the
173 highest AUROC point estimate to IGRA, in a head-to-head analysis among participants for whom results
174 of both tests were available, over a 12-month interval from sampling. These analyses were stratified by
175 setting, defined as low TB burden where incidence was < 50 per 100,000, and high TB burden above
176 this. Among individuals with serial IGRA samples, low intra-individual variance indicated high correlation
177 between serial samples (supplementary figure 7); we therefore included only one sample per individual,
178 sampled at random. We used thresholds of the maximum Youden index for the transcript and the
179 standard cut-off of 0.35 IU/ml for the QuantiFERON-TB assay. We also explored an approach using a
180 combination of the transcript and IGRA, where only those positive for both tests are offered treatment.
181 We compared the sensitivity, specificity and PPV of these approaches based on a 1% and 2% prior
182 probability, with the latter based on 1-year incidence rates in high-risk close contacts^{30,31}.

183 To evaluate the clinical utility of the RNA biomarker, IGRA and combined testing approaches, we
184 performed decision curve analysis. Decision curve analysis quantifies the trade-off between correctly
185 identifying true TB progressors and incorrectly identifying false-positives, as 'net benefit'³². Net benefit
186 is calculated across a range of weightings for the false-positives, defined as the threshold probability.
187 Threshold probability is the risk of disease at which a clinician or patient would opt for an intervention
188 such as treatment and relates to the number-willing-to-treat to prevent a single case of disease. We
189 calculated net benefit using the best performing transcript or IGRA to guide preventative treatment
190 compared to the default strategies of treating all individuals or no individuals, across a range of
191 threshold probabilities. Since the contributing datasets included case-control analyses, we fixed the
192 cumulative TB risk as 1% and 2% in our decision curve analyses.

193 Finally, we estimated the number-needed-to-treat with preventative therapy to prevent a single TB case
194 for these testing approaches compared to a default strategy of treating all individuals. We assumed a
195 constant treatment effect of 80% and fixed the cumulative TB risk as 1% and 2%.

196 **Sensitivity analyses**

197 We performed four sensitivity analyses. Firstly, we performed a two-stage IPD-MA where we calculated
198 accuracy of signatures and transcripts for each contributing cohort, explored between study
199 heterogeneity, and performed a random-effects meta-analysis to calculate pooled AUROCs (using the
200 *metafor* package in R³³). Secondly, we included participants commencing TPT in the analysis. Thirdly,
201 we included only one sample per individual, sampled at random. Finally, we included datasets from
202 which signatures were originally derived in the accuracy calculation for that signature.

203 **Role of the funding source**

204 The funder had no role in study design, data collection, data analysis, data interpretation, writing of the
205 report, or decision to submit for publication. The corresponding authors had full access to all the data
206 in the study and had final responsibility for the decision to submit for publication.

207 **Results**

208 **Overview of contributing studies**

209 Four RNAseq datasets and three qPCR datasets were included (table 1). The RNAseq datasets
210 included the Adolescent Cohort Study (ACS) of IGRA/TST positive individuals from South Africa²¹, the
211 Grand Challenges 6-74 (GC6-74) study of household contacts of TB patients from South Africa, The
212 Gambia, and Ethiopia²², and two UK close contact studies from London²⁴ and Leicester³⁴. The qPCR
213 datasets included the Correlate of Risk Targeted Intervention Study (CORTIS-01) study of healthy
214 volunteers from TB endemic communities in South Africa²⁶, the CORTIS-HR study consisting of people
215 living with HIV (PLHIV) from TB endemic communities in South Africa²⁷, and the Regional Prospective
216 Observational Research for Tuberculosis Brazil (REPORT-Brazil) study of close contacts of TB patients
217 from Brazil²⁸.

218 ACS and GC6-74 studies were nested case-control studies within larger prospective cohort studies.
219 CORTIS-01 was a randomised control trial where a cohort of healthy participants first underwent
220 measurement of the RISK11 signature (Darboe11). RISK11-positive individuals were randomised to
221 either receive TPT or not, whilst the RISK11-negative individuals did not receive TPT. Only a subset of
222 randomly sampled RISK11-negative participants were included to enrich the study cohort for RISK11-
223 positive participants. All other studies were observational cohort studies. All studies performed RNA
224 sampling at baseline; ACS, GC6-74, and REPORT-Brazil also performed serial, 6-monthly sampling.
225 REPORT-Brazil, Leicester contacts, and London contacts were from countries with a low TB burden,
226 whereas the remainder were from countries with a high TB burden. CORTIS-01 and CORTIS-HR
227 performed 15 months of follow-up whereas the remaining studies performed 22.8-24 months of follow-
228 up. TB case definitions varied by study; all included microbiologically confirmed cases (culture +/-
229 smear/PCR) while GC6-74, London contacts, and REPORT-Brazil also included clinically diagnosed
230 cases.

231 **Overview of included samples**

232 In total, 6530 samples from 5185 individuals were included in the primary analysis, with a total of 283
233 subclinical TB samples – 39 asymptomatic prevalent samples and 244 incident samples
234 (supplementary figure 8). Baseline characteristics and risk of TB disease for each study are shown in
235 supplementary table 1.

236 Three of the included eight signatures were derived from included datasets – Darboe11 and Penn-
237 Nicholson6 from ACS, and Suliman4 from GC6-74 – so the original datasets were excluded from the
238 evaluation of signature performance in the primary analysis. Distributions of gene z-scores after
239 standardization were similar between datasets and there was little heterogeneity in AUROCs of

240 transcripts between datasets (supplementary figures 4B & 10). We therefore proceeded to a one-stage
241 IPD-MA as the primary analysis.

242 ***Accuracy of signatures and single transcripts for subclinical TB***

243 The multi-gene signature with the highest AUROC for discrimination of subclinical TB from non-
244 progressors over 12 months from sampling was Roe3 (AUC 0.77 [95% CI 0.73-0.81]), which was then
245 used for pairwise comparison with the single-gene transcripts. Five single-gene transcripts had
246 equivalent AUROCs to Roe3; BATF2 (0.77 [0.73-0.81]), FCGR1A/B (0.77 [0.73-0.81]), ANKRD22 (0.77
247 [0.72-0.81]), GBP2 (0.75 [0.71-0.79]), and SERPING1 (0.75 [0.71-0.79]) (table 2). These transcripts
248 demonstrated moderate-strong correlation (0.54-0.81) using Spearman rank (supplementary figure 9).
249 Using maximum Youden index thresholds, none of the transcripts met the WHO minimum TPP
250 parameters for progression tests over 12 months. Discriminative performance of single-gene transcripts
251 diminished gradually over increasing time intervals from sampling to disease, with poor discrimination
252 over 12-15 months (supplementary table 2). BATF2 expression and discriminative performance
253 compared to non-progressors across different disease states is shown in figure 1. Notably,
254 discriminative performance of BAFT2 was highest for clinical TB (AUROC 0.93 [0.87-0.99]).

255 ***Comparison of accuracy to IGRA***

256 Comparison of the diagnostic performance of BATF2, IGRA, and a combined approach of BATF2 and
257 IGRA, stratified by setting and benchmarked against the WHO minimum TPP performance criteria
258 (figure 2A) showed that whilst the sensitivity of IGRA was similar across settings (87-88%), specificity
259 was markedly lower in high burden settings (32% [95% CI 30-35%]) compared to low burden settings
260 (74% [72-76%]) resulting in PPVs of 1.3% (1.1-1.4%) and 3.3% (2.4-3.9%) respectively (supplementary
261 table 4). In contrast, BATF2 performance was consistent across settings, with PPVs of 2.3% (1.8-2.7%)
262 and 2.4% (1.6-3.1%) in high and low burden settings respectively. In high burden settings, the combined
263 approach resulted in a slight increase in specificity compared to BATF2 alone (77% [54-76%] versus
264 67% [64-69%]) but similar PPVs (2.8% [2.1-3.5%] versus 2.3% [1.8-2.7%]). In contrast, a large increase
265 in specificity (92% [CI 90-93%] versus 72% [70-74%]) and PPV (6.8% [3.8-9.9%] versus 2.4% [1.6-
266 3.1%]) was achieved in the low burden settings when using the combined approach compared to BATF2
267 alone, albeit with some loss of sensitivity (58% [39-76%] versus 67% [47-82%]). None of the testing
268 approaches met WHO minimum TPP parameters over 12 months, although IGRA accuracy approached
269 the benchmarks in the low burden settings. Using a 2% prior probability, PPVs approximately doubled
270 without changing the overall pattern of results (supplementary figure 12).

271 ***Clinical utility analysis***

272 Decision curve analysis of these testing approaches varied by setting (figure 2B). In high burden
273 settings, IGRA offered minimal additional net benefit over a treat all strategy. BATF2 offered greater net
274 benefit than IGRA and, using prior probability of 1%, had the highest net benefit across threshold
275 probabilities between 0.4% and 2.2%, equating to a number-willing-to-treat to prevent a single case
276 (NWT) of 45-250. Above this threshold, where NWT was less than 45, treating none was best. The
277 combined approach (where only those positive for both tests are offered treatment) achieved only
278 slightly greater net benefit than BATF2 alone at threshold probabilities of 1-3%, equating to a NWT of
279 33-100. In contrast, in low burden settings, IGRA outperformed BATF2 across all threshold probabilities.
280 IGRA offered the highest net benefit at threshold probabilities under 2%, equating to a NWT of over 50,
281 whereas the combined approach offered the highest net benefit at threshold probabilities of 2-7%,
282 equating to a NWT of 14-50. Using a higher prior probability of 2%, findings were similar, but net benefit
283 for all strategies was shifted to the right on the threshold probability scale. For example, in higher burden
284 settings, BATF2 offered greater net benefit over IGRA and a treat all strategy at threshold probabilities
285 of 0.7-4.5%, equating to a NWT 22-143 (supplementary figure 12).

286 Number-needed-to-treat (NNT) estimates to prevent a single TB case for the testing approaches,
287 compared to a treat all strategy, are shown in figure 3A. In high burden settings, performing IGRA testing
288 resulted in slightly lower NNT estimates than treating all (98 [88-114] versus 125). Compared to IGRA,
289 NNTs were significantly lower using BATF2 (54 [46-68]) or a combined approach (44 [34-59]). In low
290 burden settings, NNTs of IGRA and BATF2 were similar (38 [32-51] and 53 [40-80] respectively). Using

291 the combined approach resulted in a lower NNT of 18 (13-33). Due to superior specificity, lower NNT
292 estimates were achieved by IGRA (either alone or in combination with BATF2) in low burden versus
293 high burden settings. Using a 2% prior probability roughly halved the NNT with a similar pattern (figure
294 3B).

295 **Sensitivity analyses**

296 A two-stage meta-analysis to calculate pooled AUROCs resulted in similar findings to the primary
297 analysis (supplementary table 5). Similarly, we observed similar AUROCs when including recipients of
298 TPT (excluding the Darboe-11 positive participants randomised to systematically receive TPT in
299 CORTIS-01), including only one RNA sample per individual, or when including datasets from which
300 signatures were originally derived in the accuracy calculation for that signature (supplementary table
301 6).

302 **Discussion**

303 To the best of our knowledge, we report the largest pooled RNA biomarker analysis in subclinical TB to
304 date, including over 6500 RNA samples, and the first comprehensive head-to-head analysis comparing
305 single-gene transcripts to multi-gene signatures. This is also the first evaluation of RNA biomarkers to
306 align with the ICE-TB classification. We showed that five single-gene transcripts perform equivalently
307 to the best multi-gene signature for discrimination of subclinical TB from non-progressors over 12
308 months. RNA biomarker performance was consistent across settings. In contrast, IGRA performance
309 varied markedly, with poor performance in higher burden settings. In decision curve analysis we
310 demonstrate that, in high burden settings, RNA biomarker testing offers clinical utility over IGRA, which
311 has minimal benefit over treating all with preventative therapy. In low burden settings, IGRA was the
312 best single test and approximated the WHO TPP, with greater net benefit than RNA biomarker testing.
313 However, for services aiming to treat fewer than 50 people to prevent a TB case, a two-step combined
314 testing approach improves specificity and is superior.

315 Development of the Cepheid MTB-HR prototype has demonstrated that translation to a near-patient
316 platform is feasible. Whilst the cost of such platforms is currently unknown, it is unlikely to exceed the
317 WHO TPP maximum of \$100 per test, based on the cost of an IGRA³⁵. Our findings may facilitate
318 translation of RNA biomarker technology to clinical practice by encouraging commercial competition
319 using measurement of any one of best performing single genes at lower cost. Nonetheless, none of the
320 transcripts met the WHO TPP minimum sensitivity and specificity, even over a 12-month interval,
321 although this was almost achieved by IGRA in the low burden setting (74% specificity and 88%
322 sensitivity). In the high burden setting, the WHO TPP seem an unrealistic aim that is unlikely to be
323 achieved over two years with a biomarker targeting early disease, although serial testing may improve
324 overall performance. Combining different modalities of tests may improve specificity, albeit at greater
325 cost. A universal testing strategy may be challenging to achieve if test performance is heterogenous
326 across settings; rather tailored strategies may be required based on TB burden. We demonstrate this
327 by showing that in low burden settings IGRA remains a useful test; a combined approach using IGRA
328 and blood RNA biomarkers shows additional promise and warrants further evaluation. However, in high
329 burden settings, the high prevalence of *Mtb* sensitisation means that IGRA has poor specificity and thus
330 minimal utility. Greater specificity may be achieved with better measures of recent *Mtb* infection, for
331 example *Mtb*-specific T cell activation³⁶. Alternatively, combining molecular approaches with
332 radiological testing, such as digital chest radiographs³⁷, as a method to detect macroscopic pathology,
333 may also improve performance.

334 Our findings may provide some insights into host immune responses in early TB. The co-correlation of
335 best performing single-gene transcripts is consistent with previous findings of shared upstream
336 interferon and tumour necrosis factor signalling pathways²⁵. This explains why a single transcript is a
337 sufficient measure of this immune response and combining these transcripts into multigene signatures
338 does not offer orthogonal value. The consistent performance of RNA biomarkers across settings
339 suggests that this a common host response across populations. Likewise, the similar performance in
340 CORTIS-HR suggests that this pathway is preserved in PLHIV, though previous data have
341 demonstrated likely upregulation of common type-I interferon responses in untreated HIV³⁸. The fall in
342 discriminative performance after 12 months may be reflective of de novo infection following signature

343 measurement in high burden settings, or that there is a minimal host response in earlier subclinical
344 disease. The equivalent performance of multiple co-correlated RNA biomarkers, which has also been
345 reported previously^{25,27,38}, suggests that future discovery and validation of signatures using similar
346 approaches is unlikely to yield a test with better performance. To date, discovery approaches have
347 largely focused on identifying differentially expressed transcripts in TB, before combining these to form
348 a discriminating multi-gene signature. Simplifying these to single gene biomarkers has the added benefit
349 of facilitating their integration into panels of blood RNA biomarkers for multinomial classification³⁹ that
350 may overcome the specificity limitations of the current binomial approach, for example by combining
351 with the best performing single gene biomarkers of viral infections⁴⁰.

352 An important strength of our analysis is that we have adopted ICE-TB terminology for subclinical TB, to
353 facilitate comparisons across studies and biomarker domains. We included 6530 samples including 283
354 samples from subclinical TB cases, making it the largest analysis of RNA biomarkers for subclinical TB
355 to date. Our data processing pipeline ensured batch correction within studies and integrated RNAseq
356 and qPCR data into a pooled dataset. We also performed the first decision curve analysis, to our
357 knowledge, to quantify the clinical utility of RNA biomarkers for subclinical TB and compare to existing
358 tests. We additionally stratified our decision curve analysis by TB burden, which allowed more granular
359 assessment of clinical utility by setting. We also performed a number of sensitivity analyses, including
360 a two-stage IPD-MA, to ensure our primary findings were robust.

361 A limitation of our study is that, whilst we included cohorts from high and low burden settings, there
362 were few contributing countries (South Africa, UK, Brazil, The Gambia, and Ethiopia) with no
363 representation from Asia, although large proportions of the UK studies were individuals of South Asian
364 ethnicity. There were variations in case definitions and baseline screening between studies which may
365 have resulted in misclassification, however this is reflective of real-world variations in clinical practice
366 according to resource availability. Furthermore, with the exception of CORTIS-01/HR, evaluation of TB
367 disease during follow-up was symptom-triggered so additional cases of subclinical TB may have been
368 missed. There were low numbers of subclinical cases over a 12-month interval in some studies, with
369 both Leicester Contacts and CORTIS-HR reporting 5 cases each, however this reflects the reality that
370 TB is a relatively rare outcome in longitudinal cohort studies. We also acknowledge that there is no gold
371 standard for the 'subclinical, non-infectious' state and high-resolution investigations for macroscopic
372 pathology, such as PET-CT⁴¹, were not performed. We therefore assumed that participants who
373 developed TB within 12 months would have had macroscopic pathology at baseline, had high-resolution
374 investigation been performed. However ongoing *Mtb* exposure during follow-up, particularly in high
375 transmission settings, may mean that disease cases within the primary 12-month interval may be
376 attributable to new infection and may have led to underestimation of sensitivity for subclinical TB. Since
377 subclinical TB may regress or undulate without treatment⁴², it is also possible that we underestimated
378 specificity for 'subclinical, non-infectious' TB that did not progress to clinical disease within 12 months.
379 Future studies will be required to further evaluate the accuracy of candidate biomarkers for the
380 'subclinical, non-infectious' state, once a scalable and widely accepted reference standard is
381 established.

382 In summary, we have demonstrated that several single-gene transcripts perform equivalently to multi-
383 gene signatures to detect subclinical TB, which may simplify assays and encourage commercial
384 competition. RNA biomarker performance is consistent across settings and exceeds performance of
385 IGRA in high burden settings, but falls short of WHO benchmarks. A combination strategy with IGRA
386 shows promise to enable more targeted preventative treatment in low incidence settings.

387

388 **Footnotes**

389 ***Acknowledgements***

390 We thank the study teams and participants from all of the primary studies included in this analysis (ACS,
391 CORTIS-01, CORTIS-HR, GC6-74, Leicester Contacts, London Contacts and REPORT-Brazil).

392 ***Author contributions***

393 JGB, MN, and RKG conceived the study. JGB and RKG wrote the analysis plan with input from all
394 authors. JGB performed the analyses and wrote the first draft of the manuscript, supported by RKG and
395 MN. All authors contributed to the methods and interpretation and approved the final submitted
396 manuscript.

397 ***Funding***

398 RKG is funded by National Institute for Health Research (NIHR303184), the BMA Foundation for
399 Medical Research and by NIHR Biomedical Research Funding to UCL and UCLH. MN was supported
400 by the Wellcome Trust (207511/Z/17/Z) and by NIHR Biomedical Research Funding to UCL and UCLH.

401 ***Declaration of interests***

402 MN holds a patent in relation to blood transcriptomic biomarkers of tuberculosis. TJS is co-inventor of
403 patents of the Penn-Nicholson⁶ and Suliman⁴ signatures. All other authors declare no competing
404 interests.

405 ***Data availability***

406 All data analysed in this study are publicly available through the original contributing studies.

407

References

- 1 World Health Organization. Global Tuberculosis Report 2023. <https://www.who.int/publications-detail-redirect/9789240083851> (accessed May 28, 2024).
- 2 Litvinjenko S, Magwood O, Wu S, Wei X. Burden of tuberculosis among vulnerable populations worldwide: an overview of systematic reviews. *The Lancet Infectious Diseases* 2023; **23**: 1395–407.
- 3 Coussens AK, Zaidi SMA, Allwood BW, *et al.* Classification of early tuberculosis states to guide research for improved care and prevention: an international Delphi consensus exercise. *The Lancet Respiratory Medicine* 2024; **0**. DOI:10.1016/S2213-2600(24)00028-6.
- 4 Paton Nicholas I., Cousins Christopher, Suresh Celina, *et al.* Treatment Strategy for Rifampin-Susceptible Tuberculosis. *New England Journal of Medicine* 2023; **388**: 873–87.
- 5 Mahomed H, Hawkrige 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**: e17984.
- 6 Kik SV, Franken WPJ, Mensen M, *et al.* Predictive value for progression to tuberculosis by IGRA and TST in immigrant contacts. *Eur Respir J* 2010; **35**: 1346–53.
- 7 Zhou G, Luo Q, Luo S, *et al.* Interferon- γ release assays or tuberculin skin test for detection and management of latent tuberculosis infection: a systematic review and meta-analysis. *The Lancet Infectious Diseases* 2020; **20**: 1457–69.
- 8 Melnychuk L, Perlman-Arrow S, Lisboa Bastos M, Menzies D. A Systematic Review and Meta-Analysis of Tuberculous Preventative Therapy Adverse Events. *Clin Infect Dis* 2023; **77**: 287–94.
- 9 Berry MPR, Graham CM, McNab FW, *et al.* An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature* 2010; **466**: 973–7.
- 10 Bloom CI, Graham CM, Berry MPR, *et al.* Transcriptional blood signatures distinguish pulmonary tuberculosis, pulmonary sarcoidosis, pneumonias and lung cancers. *PLoS One* 2013; **8**: e70630.
- 11 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 Medicine* 2013; **10**: e1001538.
- 12 Anderson ST, Kaforou M, Brent AJ, *et al.* Diagnosis of childhood tuberculosis and host RNA expression in Africa. *N Engl J Med* 2014; **370**: 1712–23.
- 13 Sweeney TE, Braviak L, Tato CM, Khatri P. Genome-wide expression for diagnosis of pulmonary tuberculosis: a multicohort analysis. *Lancet Respir Med* 2016; **4**: 213–24.
- 14 Roe JK, Thomas N, Gil E, *et al.* Blood transcriptomic diagnosis of pulmonary and extrapulmonary tuberculosis. *JCI Insight* 2016; **1**. DOI:10.1172/jci.insight.87238.
- 15 Maertzdorf J, McEwen G, Weiner J, *et al.* Concise gene signature for point-of-care classification of tuberculosis. *EMBO Mol Med* 2016; **8**: 86–95.
- 16 Gjøen JE, Jennum S, Sivakumaran D, *et al.* Novel transcriptional signatures for sputum-independent diagnostics of tuberculosis in children. *Sci Rep* 2017; **7**: 5839.
- 17 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**: 77–83.

- 18 Kulkarni V, Queiroz ATL, Sangle S, *et al.* A Two-Gene Signature for Tuberculosis Diagnosis in Persons With Advanced HIV. *Front Immunol* 2021; **12**. DOI:10.3389/fimmu.2021.631165.
- 19 Natarajan S, Ranganathan M, Hanna LE, Tripathy S. Transcriptional Profiling and Deriving a Seven-Gene Signature That Discriminates Active and Latent Tuberculosis: An Integrative Bioinformatics Approach. *Genes* 2022; **13**: 616.
- 20 Chendi BH, Jooste T, Scriba TJ, *et al.* Utility of a three-gene transcriptomic signature in the diagnosis of tuberculosis in a low-endemic hospital setting. *Infectious Diseases* 2023; **55**: 44–54.
- 21 Zak DE, Penn-Nicholson A, Scriba TJ, *et al.* A blood RNA signature for tuberculosis disease risk: a prospective cohort study. *Lancet* 2016; **387**: 2312–22.
- 22 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**: 1198–208.
- 23 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**: 8629.
- 24 Roe J, Venturini C, Gupta RK, *et al.* Blood Transcriptomic Stratification of Short-term Risk in Contacts of Tuberculosis. *Clin Infect Dis* 2020; **70**: 731–7.
- 25 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. *The Lancet Respiratory Medicine* 2020; **8**: 395–406.
- 26 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**: 354–65.
- 27 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. *The Lancet Global Health* 2021; **9**: e841–53.
- 28 Mendelsohn SC, Andrade BB, Mbandi SK, *et al.* Transcriptomic signatures of progression to TB disease among close contacts in Brazil. *The Journal of Infectious Diseases* 2024; : jiae237.
- 29 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* 2022; **2**: 1–13.
- 30 Martinez L, Cords O, Horsburgh CR, Andrews JR, Pediatric TB Contact Studies Consortium. The risk of tuberculosis in children after close exposure: a systematic review and individual-participant meta-analysis. *Lancet* 2020; **395**: 973–84.
- 31 Du Y, He Y, Zhang H, *et al.* Declining incidence rate of tuberculosis among close contacts in five years post-exposure: a systematic review and meta-analysis. *BMC Infectious Diseases* 2023; **23**: 373.
- 32 Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006; **26**: 565–74.
- 33 Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software* 2010; **36**: 1–48.
- 34 Singhania A, Verma R, Graham CM, *et al.* A modular transcriptional signature identifies phenotypic heterogeneity of human tuberculosis infection. *Nat Commun* 2018; **9**: 2308.

- 35 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 (WHO/HTM/TB/2017.18). Licence: CC BY-NC-SA 3.0 IGO. .
- 36 Mpande CAM, Musvosvi M, Rozot V, *et al.* Antigen-Specific T-Cell Activation Distinguishes between Recent and Remote Tuberculosis Infection. *Am J Respir Crit Care Med* 2021; **203**: 1556–65.
- 37 Tan Q, Huang C-C, Becerra MC, *et al.* Chest Radiograph Screening for Detecting Subclinical Tuberculosis in Asymptomatic Household Contacts, Peru. *Emerg Infect Dis* 2024; **30**: 1115–24.
- 38 Mann T, Gupta RK, Reeve BWP, *et al.* Blood RNA biomarkers for tuberculosis screening in people living with HIV before antiretroviral therapy initiation: a diagnostic accuracy study. *The Lancet Global Health* 2024; **12**: e783–92.
- 39 Habgood-Coote D, Wilson C, Shimizu C, *et al.* Diagnosis of childhood febrile illness using a multi-class blood RNA molecular signature. *Med* 2023; **4**: 635-654.e5.
- 40 Gupta RK, Rosenheim J, Bell LC, *et al.* Blood transcriptional biomarkers of acute viral infection for detection of pre-symptomatic SARS-CoV-2 infection: a nested, case-control diagnostic accuracy study. *Lancet Microbe* 2021; **2**: e508–17.
- 41 Esmail H, Lai RP, Lesosky M, *et al.* Characterization of progressive HIV-associated tuberculosis using 2-deoxy-2-[18F]fluoro-D-glucose positron emission and computed tomography. *Nat Med* 2016; **22**: 1090–3.
- 42 Richards AS, Sossen B, Emery JC, *et al.* Quantifying progression and regression across the spectrum of pulmonary tuberculosis: a data synthesis study. *Lancet Glob Health* 2023; **11**: e684–92.

Inserts

Table 1: Characteristics of studies included in meta-analysis

	ACS	CORTIS-01	CORTIS-HR	GC6-74	Leicester Contacts	London Contacts	REPORT-Brazil
Method of RNA profiling	RNAseq	RT-qPCR	RT-qPCR	RNAseq	RNAseq	RNAseq	RT-qPCR
Study design	Nested case-control	Randomised controlled trial*	Cohort	Nested case-control	Cohort	Cohort	Cohort
Population	Positive IGRA/TST	Endemic community	PLHIV in endemic community	Household contacts	Close contacts	Close contacts	Close contacts
Age	12-18	≥18	≥18	10-60	≥16	≥18	≥18
TB burden of setting	High burden	High burden	High burden	High burden	Low burden	Low burden	Low burden
Country	South Africa	South Africa	South Africa	South Africa, The Gambia, Ethiopia	UK	UK	Brazil
Baseline screening	Clinical evaluation	2x sputum	2x sputum	Clinical evaluation	Clinical evaluation + CXR	Clinical evaluation + CXR	Clinical evaluation
Study case definition	2x positive smear or 1x positive culture	2x positive Xpert MTB/RIF or Xpert Ultra or culture	2x positive Xpert MTB/RIF or Xpert Ultra or culture	Positive culture or clinically diagnosed	Positive culture or Xpert MTB/RIF	Positive culture or clinically diagnosed	Positive culture or clinically diagnosed
Duration of follow-up (months)	24	15	15	24	24	22.8	24
RNA sample timing	Baseline & 6/12/18/24 months	Baseline	Baseline	Baseline & 6/18 months	Baseline	Baseline	Baseline & 6 months
Total participants	144	2496	404	334	104	324	1379
Total RNA samples	318	2496	404	412	104	324	2472

*The cohort was enriched for RISK11-positive individuals. The RISK11-positive individuals were randomised to TPT or no TPT. The RISK11-negative individuals did not receive TPT.

Legend: CXR, chest x-ray; PLHIV, people living with HIV

Table 2: Performance metrics of equivalent single-gene transcripts for subclinical TB

Performance metrics of single-gene transcripts with equivalent performance to the best performing multi-gene signature (Roe3) to discriminate subclinical TB cases from non-progressors at an interval of 12 months from sampling to disease are shown. Equivalence to Roe3 was defined as an adjusted p value >0.05 in the pairwise Delong test. Performance metrics include estimates and 95% confidence intervals for the Area Under the Receiver Operating Curve (AUROC) as well as sensitivity and specificity at the maximum Youden index calculated from the one-stage meta-analysis.

RNA biomarker	AUROC	Sensitivity	Specificity	N	Cases	Controls	p
Roe3	0.77 (0.73 - 0.81)	0.74 (0.67 - 0.79)	0.7 (0.69 - 0.71)	5,171	189	4,982	Reference
BATF2	0.77 (0.73 - 0.81)	0.74 (0.67 - 0.8)	0.69 (0.67 - 0.7)	5,171	189	4,982	0.715
FCGR1A/B	0.77 (0.73 - 0.81)	0.65 (0.58 - 0.72)	0.79 (0.78 - 0.8)	5,170	189	4,981	0.867
ANKRD22	0.77 (0.72 - 0.81)	0.57 (0.48 - 0.65)	0.86 (0.85 - 0.87)	3,381	138	3,243	0.881
GBP2	0.75 (0.71 - 0.79)	0.74 (0.67 - 0.79)	0.65 (0.64 - 0.66)	5,171	189	4,982	0.064
SERPING1	0.75 (0.71 - 0.79)	0.66 (0.59 - 0.73)	0.75 (0.73 - 0.76)	5,168	189	4,979	0.069

Figure 1: BATF2 expression and diagnostic accuracy for different disease states

Expression (Z-score transformed) of the best performing single-gene transcript, BATF2, is shown as a combined boxplot and scatterplot, stratified by disease state. Disease state is classified according to ICE-TB consensus definitions and descriptive terms are also shown. Boxes represent interquartile range and median values. Grey dots represent individual samples and black dots represent outliers. Area Under the Curve (AUC) with 95% confidence intervals of BATF2 to discriminate the disease state from non-progressors is also shown.

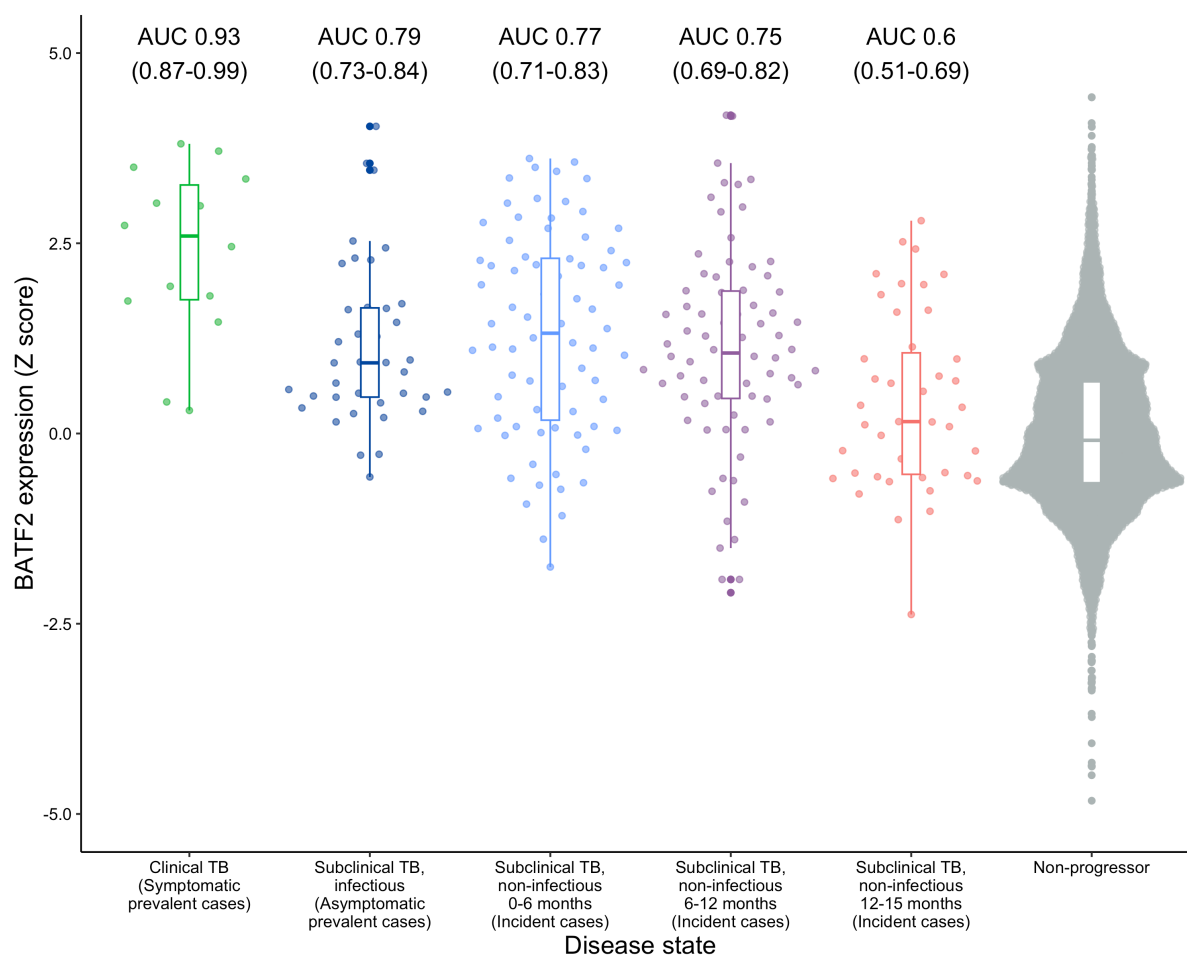


Figure 2: Diagnostic performance for subclinical TB of BATF2, IGRA and a combined approach shown in receiver operating space and in a decision curve analysis, by setting

Figure 2A compares the diagnostic performance for subclinical TB over a 0-12 month interval of BATF2 (threshold set at maximum Youden Index), IGRA (threshold set at standard cutoff of 0.35 IU/ml) and a combined approach of BATF2 and IGRA in the receiver operating space with sensitivity on the y axis and 1-specificity on the x axis, stratified by setting. Only participants with results for both tests were included in this analysis. Shown as point estimates with 95% confidence intervals (boxes). Dotted lines represent the 75% WHO minimum Target Product Profile sensitivity and specificity for a TB progression test. Dashed lines represent positive predictive values of 1%, 3% and 5%, based on a 1% prior probability. Underlying data are shown in supplementary table 4.

Figure 2B is a decision curve analysis where each test is compared to default strategies of treating all or treating no persons, stratified by setting. Threshold probability is the risk of TB disease at which a clinician or patient would opt for preventative therapy and is the reciprocal of the number-willing-to-treat to prevent a single case. Net benefit is calculated at a range of threshold probabilities as the true positive rate minus a weighted false positive rate, where the weighting is the threshold probability. Since the contributing datasets included case-control analyses, the cumulative TB risk was fixed at 1%.

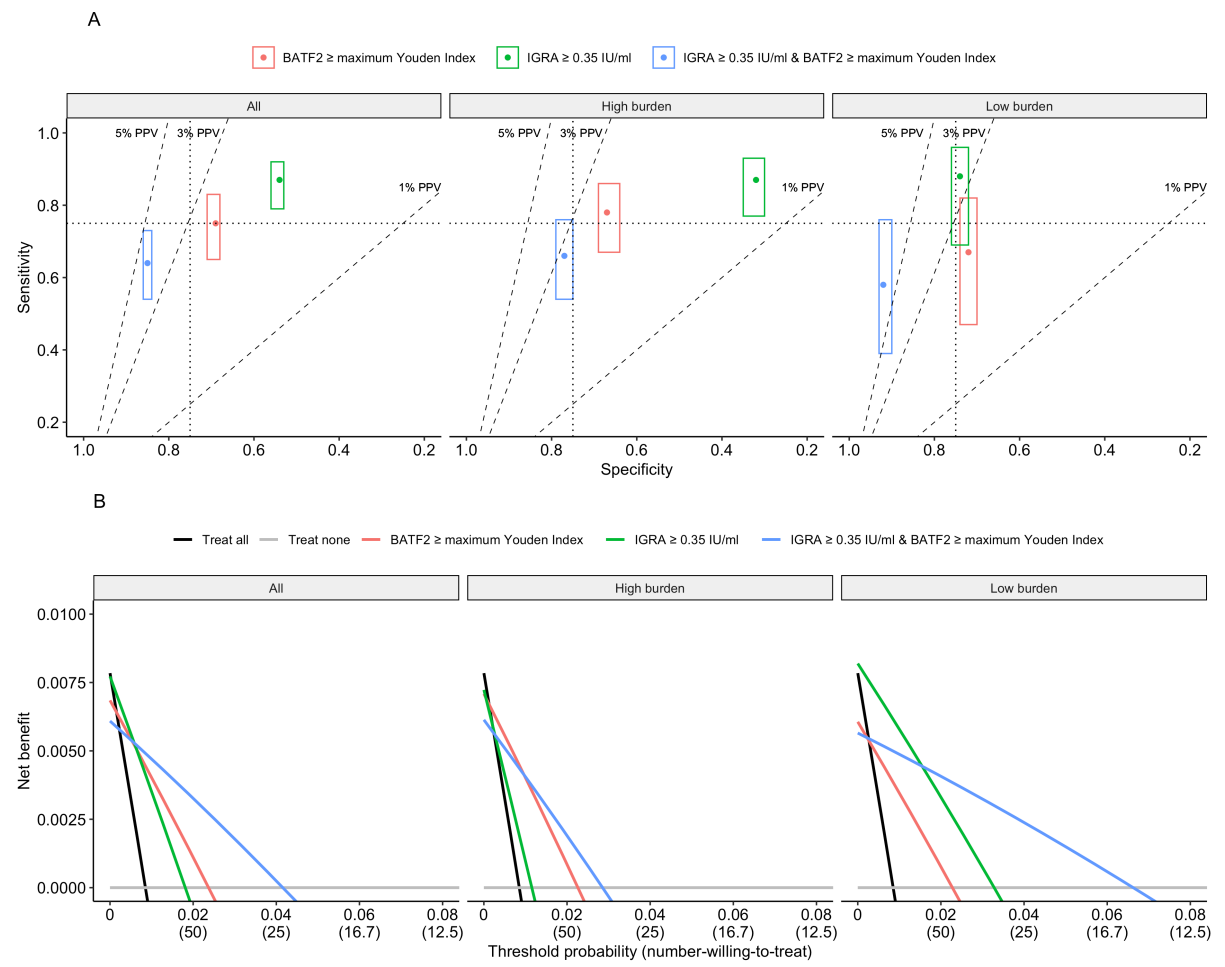


Figure 3: Number-needed-to-treat to prevent a single TB case using different testing strategies, by setting, using 1% and 2% prior probabilities

Estimated number-needed-to-treat (NNT) with preventative therapy to prevent a single TB case is shown for the default strategy of treating all compared to test stratified treatment using BATF2 (threshold set at maximum Youden Index), IGRA (threshold set at standard cutoff of 0.35 IU/ml) and a combined approach of BATF2 and IGRA. Shown as point estimates and 95% confidence intervals (bars). An estimated treatment effect of 80% was used. Since the contributing datasets included case-control analyses, the cumulative TB risk was fixed at 1% (A) and 2% (B).

