FACTORS AFFECTING DIAGNOSTIC AND PROGNOSTIC PERFORMANCE OF A TRANSCRIPTOMIC SIGNATURE OF RISK OF TUBERCULOSIS IN HIV-UNINFECTED SOUTH AFRICAN ADULTS

HUMPHREY MULENGA

MLNHUM001

BSc, MPH (EPIDEMIOLOGY)

Thesis submitted for the degree of

Doctor of Philosophy

in the Department of Pathology

Faculty of Health Sciences

University of Cape Town

Supervisors:

Professor Mark Hatherill

Professor Thomas J. Scriba

17-OCTOBER-2022

Declaration

I Humphrey Mulenga, (Student #: MLNHUM001) declare that the work presented in this thesis is my own and where the work of others has been used, (whether quoted verbatim, paraphrased or referred to) it has been attributed and acknowledged and has not, in whole or in part, been submitted towards another degree, at this University or elsewhere.

The University is empowered to reproduce either the whole or any portion of the contents for the purposes of research.

Permission to include published manuscripts

"I confirm that I have been granted permission by the University of Cape Town's Doctoral Degrees Board to include the following publications in my thesis, and where co-authorships are involved, my co-authors have agreed that I may include the publications:

- **Mulenga H**, Bunyasi EW, Mbandi SK, Mendelsohn SC, Kagina B, Penn-Nicholson A, Scriba T, and Hatherill M. *Performance of host blood transcriptomic signatures for diagnosing and predicting progression to tuberculosis disease in HIV-negative adults and adolescents: a systematic review protocol*. BMJ Open, 2019. 9(5): p. e026612.
- **Mulenga H**, Zauchenberger CZ, Bunyasi EW, Mbandi SK, Mendelsohn SC, Kagina B, Penn-Nicholson A, Scriba T, and Hatherill M. *Performance of diagnostic and predictive host blood transcriptomic signatures for tuberculosis disease: a systematic review and meta-analysis*. PLoS One, 2020. **15**(8): p. e0237574.
- **Mulenga H**, Fiore-Gartland A, Mendelsohn SC, Penn-Nicholson A, Mbandi SK, Borate B, Musvosvi M, Tameris M, Walzl G, Naidoo K, Churchyard G, Scriba TJ, and Hatherill M. *The effect of host factors on discriminatory performance of a transcriptomic signature of tuberculosis risk*. EBioMedicine, 2022. 77:103886.
- **Mulenga H**, Fiore-Gartland A, Mendelsohn SC, Penn-Nicholson A, Mbandi SK, Nemes M, Borate B, Musvosvi M, Tameris M, Walzl G, Naidoo K, Churchyard G, Scriba TJ, and Hatherill M. *Evaluation of a transcriptomic signature of tuberculosis risk in combination with an interferon gamma release assay: a diagnostic test accuracy study*. EclinicalMedicine, 2022. 47:101396

Note that permission was not sought to include the publication below, and hence it is not presented verbatim because the manuscript to which these data contributed was jointly first co-authored with another PhD student S.C Mendelsohn. Therefore chapter 5 contains only my contribution to that manuscript. **Mulenga H**, Musvosvi M, Mendelsohn SC, Penn-Nicholson A, Mbandi SK, Fiore-Gartland A, Tameris M, Mabwe S, Africa H, Bilek N, Kafaar F, Khader SA, Carstens B, Hadley K, Hikuam C, Erasmus M, Jaxa L, Raphela R, Nombida O, Kaskar M, Nicol MP, Mbhele S, Van Heerden J, Innes C, Brumskine W, Hiemstra A, Malherbe ST, Hassan-Moosa R, Walzl G, Naidoo K, Churchyard G, Hatherill M, Scriba TJ. *Longitudinal Dynamics of a Blood Transcriptomic Signature of Tuberculosis.* Am J Respir Crit Care Med, 2021. **204**(12): p. 1463-1472..

Thesis abstract

Background

Host blood transcriptomic signatures, such as RISK11, have potential as tests for diagnosing and predicting tuberculosis. This thesis aimed to review the literature, evaluate host and non-host factors associated with variability of the RISK11 signature and impact on discriminatory performance and evaluate RISK11 performance in combination with tests of *Mycobacterium tuberculosis* sensitization.

Methods

A systematic review of discriminatory performance of transcriptomic signatures for tuberculosis was conducted. RISK11, QuantiFERON-TB Gold-Plus and host factors were analysed in a prospective cohort, in which a cross-sectional study of upper respiratory organisms was nested. Effects on RISK11 were quantified using multivariable generalised regression. Discriminatory performance of RISK11, and RISK11/QuantiFERON combinations, were quantified by area under the curve and/or sensitivity and specificity.

Results

In the literature, one signature (90% sensitivity; 74% specificity) met the minimal criteria for a triage test; one signature (86% sensitivity; 84% specificity) met the minimal criteria for a predictive test.

In the prospective cohort, RISK11 scores were higher among individuals with prevalent tuberculosis (+18.90%), night sweats (+14.65%) and incident tuberculosis (+7.29%). Cough was associated with 72.55% higher RISK11 score in prevalent tuberculosis cases. Stratification by cough improved diagnostic performance from area under curve of 0.74 overall, to 0.97 in cough-positive participants. Adjustment for host factors affecting controls did not change RISK11 discriminatory performance.

In the cross-sectional study, RISK11 scores were higher by +16.7%, +67.8% and +13.5% in participants with coronavirus, influenza and rhinovirus, respectively, such that RISK11 could not differentiate prevalent tuberculosis from upper respiratory viruses.

Compared to RISK11, the Either-Positive test combination decreased diagnostic negative likelihood ratio from 0.7 to 0.3, and prognostic negative likelihood ratio from 0.9 to 0.3, but did not improve upon QuantiFERON alone. Compared to QuantiFERON, the Both-Positive test combination increased diagnostic positive likelihood ratio from 1.3 to 4.7, and prognostic positive likelihood ratio from 1.4 to 2.8, but did not improve upon RISK11 alone.

Conclusion

RISK11 holds promise as a triage test for tuberculosis. Further optimisation, or development of new signatures is needed to improve discrimination of subclinical tuberculosis, without cough, and to mitigate the impact of viral co-infection. RISK11/QuantiFERON combination testing is not recommended.

Abstract word count = 350

Table of contents

Declaration ii
Permission to include published manuscripts iii
Thesis abstract iv
List of tables xii
List of figures xiii
List of abbreviations and terms xiv
Acknowledgements xv
Dedication xvi
Expanded summary xvii
Chapter 1 22
1 General introduction 22
1.1 Background 22
1.2 Study design and setting 24
1.3 Thesis structure and aims 25
1.4 Ethics 26
1.5 References 26
Chapter 2 30
2 Performance of host blood transcriptomic signatures for diagnosing and predicting progression to tuberculosis disease in HIV-negative adults and adolescents: a systematic review protocol. 30
2.1 Abstract 32
2.2 Introduction 33
2.3 Research question and aims 35
2.3.1 Research question 35
2.3.2 Objectives 35
2.4 Methods and analysis 36
2.4.1 Definitions and study inclusion criteria 36
2.4.2 Literature search 37
2.4.3 Data management 37
2.4.4 Study selection 38
2.4.5 Data extraction 38
2.4.6 Quality appraisal 39
2.4.7 Data analysis and synthesis 39
2.4.8 Ethics and dissemination 40

2.4.9 Assessing cumulative evidence 41
2.4.10 Patient and public involvement 41
2.5 Discussion 41
2.6 References 42
Chapter 3 46
3 Performance of diagnostic and predictive host blood transcriptomic signatures for tuberculosis disease: a systematic review and meta-analysis 46
3.1 Abstract 48
3.2 Introduction 49
3.3 Materials and methods 50
3.3.1 Study inclusion/exclusion criteria 50
3.3.2 Literature search 51
3.3.3 Study selection 51
3.3.4 Data extraction and management 51
3.3.5 Appraisal of methodological quality of studies 52
3.3.6 GRADE quality of evidence 52
3.3.7 Data synthesis and analysis 52
3.4 Results 53
3.4.1 Search results 53
3.4.2 Quality of diagnostic studies 54
3.4.3 Quality of studies for prediction of progression to TB disease 55
3.4.4 Performance of mRNA signatures for TB Diagnosis 55
3.4.5 Meta-analysis of performance of mRNA signatures for diagnosis of TB disease 57
3.4.6 Performance of mRNA signatures for prediction of progression to TB disease 58
3.4.7 GRADE evidence profile 59
3.5 Discussion 60
3.5.1 Strengths and limitations of the study 63
3.6 Conclusion 63
3.7 References 64
Chapter 4 70
4 The effect of host factors on discriminatory performance of a transcriptomic signature of tuberculosis risk 70
4.1 Abstract 73
4.2 Research in context 74

4.3 Introduction 75
4.4 Methods 77
4.4.1 Ethics approval 77
4.4.2 Study design and participants 77
4.4.3 Statistical analysis 78
4.4.4 Role of the funding source 80
4.5 Results 81
4.5.1 Baseline characteristics 81
4.5.2 Factors associated with RISK11 in all participants 82
4.5.3 Factors associated with RISK11 in prevalent tuberculosis cases 83
4.5.4 Factors associated with RISK11 in incident tuberculosis cases 84
4.5.5 Tuberculosis-independent factors affecting RISK11 in controls 85
4.5.6 Effect of host factors on RISK11 signature performance 86
4.5.7 Combination of RISK11 with host factors 88
4.6 Discussion 90
4.7 References 94
Chapter 5 97
5 The effect of upper respiratory tract organisms on RISK11 score. 97
5.1 Abstract 99
5.2 Introduction 100
5.3 Methods 101
5.3.1 Study design 101
5.3.2 Study population 101
5.3.3 Study procedures 102
5.3.4 Statistical analysis 102
5.4 Results 104
5.4.1 Baseline characteristics 104
5.4.2 Upper respiratory tract organisms 106
5.4.3 Symptoms 107
5.4.4 TB disease 108
5.4.5 Effect of upper respiratory organisms on RISK11 scores 109
5.4.6 RISK11 performance for differentiating between individuals with and without prevalent TB and upper respiratory tract organisms. 110
5.5 Discussion 111

5.6 Conclusion 113
5.7 References 114
6 Evaluation of a transcriptomic signature of tuberculosis risk in combination with an interferon gamma release assay: a diagnostic test accuracy study 116
6.1 Summary 119
6.2 Research in context 120
6.3 Introduction 121
6.4 Methods 122
6.4.1 Study design and participants 122
6.4.2 Statistical analysis 123
6.4.3 Ethics approval 125
6.4.4 Role of the funding source 125
6.5 Results 125
6.5.1 Associations between RISK11 and QFTPlus 126
6.5.2 Prevalent TB 127
6.5.3 Incident TB 130
6.6 Discussion 132
6.7 References 137
Chapter 7 140
7 Discussion and conclusion 140
7.1 Summary of the primary findings 140
7.2 The spectrum of TB disease 140
7.3 Diagnostic and prognostic performance of published mRNA signatures of TB 141
7.4 Effect of host factors on discriminatory performance of the RISK11 mRNA signature 142
7.5 Combining host factors with RISK11 to improve discriminatory performance 143
7.6 Effect of upper respiratory viruses on discriminatory performance of RISK11 and risk of incident TB 143
7.7 Probability and risk of TB for individuals who test RISK11 and IGRA positive 144
7.8 Diagnostic and prognostic performance of a RISK11/IGRA combination testing 145
7.9 Limitations and strengths 146
7.10 Potential future research areas 147
7.11 Conclusion 148
7.12 References 149
8 Appendices 153

8.1 Appendix 2.1: PRISMA-P 2015 checklist: Items to be reported in a systematic review 153

8.2 Appendix 2.2: Initial search strategy..... 154

8.3 Appendix 2.3: Data extraction form 155

8.4 Appendix 2.4: Individual Study Quality Assessment Tool (QUADAS-2) 157

8.5 Appendix 3.1: PRISMA checklist: Items reported in the systematic review 158

8.6 Appendix 3.2: A matrix of all identified signatures with their corresponding gene composition. 160

8.7 Appendix 3.3: Column chart of forty-two most frequently utilised genes in transcriptomic signatures of either TB diagnosis or prediction to TB disease. 161

8.8 Appendix 3.4: Characteristics of studies for diagnosis of TB included in the systematic review. 162

8.9 Appendix 3.5: Methodological quality summary of all diagnostic studies with independent validation cohort. 170

8.10 Appendix 3.6: Methodological quality summary of all diagnostic studies with independent validation cohort 171

8.11 Appendix 3.7: Characteristics of studies for predicting progression to TB disease included in the systematic review. 172

8.12 Appendix 3.8: Methodological quality summary of studies of prediction to TB disease in independent validation cohorts. 173

8.13 Appendix 3.9: Forest plots of sensitivity and specificity of mRNA transcriptomic signatures for diagnosis of TB disease in independent validation cohorts (all studies 174

8.14 Appendix 3.10: Forest plots of sensitivity and specificity of mRNA transcriptomic signatures for diagnosis of TB disease for studies without independent validation cohorts. 175

8.15 Appendix 3.11: Summary receiver operating characteristic (SROC) curves for the Berry86_2010, Sweeney3_2016 and Zak16_2016 signatures. 176

8.16 Appendix 3.12: Performance of Signatures for predicting progression to TB disease. 177

8.17 Appendix 3.13: GRADE evidence profile: mRNA signatures for the diagnosis of TB (all studies). 178

8.18 Appendix 4.1: The 11 transcripts comprising the RISK11 signature of TB risk and their putative functions. 179

8.19 Appendix 4.2: Study design..... 180

8.20 Appendix 4.3: Baseline characteristics of enrolled participants by TB status adjusted to reflect screening population. 181

8.21 Appendix 4.4: Comparison of distribution of RISK11 scores by baseline characteristics in (a) Prevalent TB cases, (b) Incident TB cases, and (c) Controls. 182

8.22 Appendix 4.5: Distribution of baseline RISK11 scores in (a) prevalent and (b) incident TB cases stratified by chest radiograph positivity at diagnosis. 183

8.23 Appendix 4.6: Distribution of baseline RISK11 scores among controls, one-sample sputum positive cases, and two-sample sputum positive cases unadjusted for sampling weights. 184

8.24 Appendix 4.7: Univariable and multivariable regression analyses of baseline predictors of RISK11 score in participants with prevalent TB at enrolment. 185

8.25 Appendix 4.8: Univariable and multivariable regression analyses of baseline predictors of RISK11 score in participants who progressed to incident TB. 186

8.26 Appendix 4.9: Univariable and multivariable regression analyses of baseline predictors of RISK11 score in participants without prevalent TB and did not progress to incident TB (controls). . 187

8.27 Appendix 4.10a: Initial exploratory ROC regression analysis for the effect of covariates on discriminatory accuracy of RISK11. 188

8.28 Appendix 4.10b: Final ROC regression analysis for the effect of covariates on discriminatory accuracy of RISK11 derived from appendix table 10a 189

8.29 Appendix 5.1: Exploratory univariable and multivariable generalised linear models for upper respiratory organisms and socio-demographic predictors of RISK11 score in participants who were not investigated for TB..... 190

8.30 Appendix 6.1: Parent study (CORTIS) trial profile..... 191

8.31 Appendix 6.2: Current study flow of participants included in the analysis. 192

8.32 Appendix 6.3: Scatter plots for the relationship between RISK11 and QFTPlus scores. 193

8.33 Appendix 6.4: Test agreement between QFTPlus and RISK11..... 194

8.34 Appendix 6.5: RISK11-positivity rates stratified by either TB status or QFTPlus results. 195

8.35 Appendix 6.6: Prevalent and incident TB disease by risk category using alternative RISK11 and QFTPlus cut-offs. 196

8.36 Appendix 6.7: Prevalent and incident TB disease by risk category at optimal test thresholds 197

8.37 Appendix 6.8: Performance of RISK11 and QFTPlus alone and in combination for diagnosis of prevalent and prognosis of incident TB in the enrolled population. 198

8.38 Appendix 6.9: Simulated performance of RISK11 and QFTPlus alone and in combination for diagnosis of prevalent and prognosis of incident TB for 10000 tests conducted..... 199

8.39 Appendix 6.10: Performance of RISK11 and QFTPlus alone and in combination for diagnosis of prevalent and prognosis of incident TB at optimal thresholds..... 200

8.40 Appendix 6.11: Performance of RISK11 and QFTPlus as continuous biomarkers. 201

8.41 Appendix 12: RISK of Incident TB disease by risk category and excluding participants that received preventive therapy for TB. 202

8.42 Appendix 6.13: Additional methods and results..... 203

8.43 References 205

List of tables

Table 3.1: Meta-analysis of performance of transcriptomic signatures for diagnosis of TB disease in independent validation cohorts that included LTBI or other disease populations 58

Table 4.1: Baseline characteristics of enrolled participants by TB status. 81

Table 4.2: Univariable and multivariable generalised linear models of the predictors of RISK11 score in all participants. 83

Table 4.3: Multivariable generalised linear models of the predictors of RISK11 score in prevalent TB cases and controls. 84

Table 4.4: Crude and cough-specific performance estimates for RISK11 at various thresholds. 88

Table 4.5: Performance of RISK11, base models, and combination models for TB disease. 89

Table 5.1: Demographic characteristics of participants. 105

Table 5.2: Frequency of upper respiratory tract organisms detected by PCR and distribution of RISK11 scores in all participants. 106

Table 5.3: Association between upper respiratory viral organisms and risk of TB 108

Table 5.4: Univariable and multivariable generalised linear models of the effect of socio-demographic and upper respiratory organisms on RISK11 score in participants investigated for TB. 109

Table 6.1: Comparison of baseline characteristics by RISK11/QFTPlus risk categories 126

Table 6.2: Prevalent and incident tuberculosis disease by risk category 128

Table 6.3: Performance of RISK11 and QFTPlus alone and in combination for diagnosis of prevalent and prognosis of incident tuberculosis. 129

List of figures

Figure 3.1: Flow of studies in the review of transcriptomic signatures for diagnosing and predicting progression to TB disease. 54
Figure 3.2: Summary of results of QUADAS-2 assessment of diagnostic studies in clinically relevant independent validation cohorts with other diseases. 55
Figure 3.3: Summary of results of QUADAS-2 assessment in studies of prediction to TB disease in independent validation cohorts. 55
Figure 3.4: Forest plots of sensitivity and specificity of transcriptomic signatures for diagnosis of TB disease that met at least one minimum TPP performance criterion in independent validation cohorts. 56
Figure 3.5: Forest plots of sensitivity and specificity of transcriptomic signatures for diagnosis of TB disease in independent validation cohorts of clinically relevant populations with other diseases. 57
Figure 3.6: Forest plots of sensitivity and specificity of transcriptomic signatures for prediction of progression to TB disease in independent validation cohorts. 59
Figure 3.7: GRADE evidence profile: transcriptomic signatures for the diagnosis of TB in clinically relevant populations with other diseases only. 60
Figure 4.1: Predicted marginal effects on RISK11 by different host factors in (a) all participants, (b) participants with prevalent tuberculosis, and (c) participants without tuberculosis. 82
Figure 4.2: Crude and covariate-adjusted ROC curves for the discrimination of (a) prevalent TB from controls and (b) incident TB from controls. 86
Figure 4.3: Performance of RISK11 when stratified by cough status and when combined with host factors. 87
Figure 5.1: Study design. Of the 286 participants co-enrolled in the CORTIS parent study and investigated for TB, 11 had prevalent TB and 9 progressed to incident TB. 105
Figure 5.2: Viral or bacterial upper respiratory tract organisms, and uninfected individuals. 107
Figure 5.3: Predicted marginal effects on RISK11 of the different upper respiratory organisms and socio-demographic factors in (a) participants investigated for TB and (b) participants not investigated for TB. 110
Figure 5.4: Diagnostic performance of RISK11 for differentiating between participants with TB, viral or bacterial upper respiratory tract organisms, and uninfected individuals. 111

List of abbreviations and terms

CAPRISA	Centre for the AIDS Programme of Research in South Africa
CORTIS	Correlates of Risk Targeted Intervention Study
DTA	Diagnostic Test Accuracy
EPTB	Extra Pulmonary Tuberculosis
FIND	Foundation for Innovative New Diagnostics
HIV	Human Immunodeficiency Virus
IGRA	Interferon Gama Release Assay
Incident TB	New cases of active TB occurring in a specified time interval
ISG	Interferon Stimulated Gene
KZN	KwaZulu-Natal
LTBI	Latent TB Infection
mRNA	Messenger Ribonucleic Acid
MTB	*Mycobacterium Tuberculosis*
NDWG	New Diagnostics Working Group
Prevalent TB	Active TB cases occurring at a specific time point, as a proportion of total number at risk
PCR	Polymerase-Chain Reaction
PTB	Pulmonary Tuberculosis
QFT	QuantiFERON
QFTPlus	QuantiFERON-TB Gold-Plus
RISK11	The 11-gene transcriptomic signature test for diagnosis of and prediction of progression to TB disease used in this study
SATVI	South African Tuberculosis Vaccine Initiative
STARD	Standards for the Reporting of Diagnostic Accuracy Studies
TPP	Target Product Profile
UCT	University of Cape Town
UN	United Nations
TB	Tuberculosis
TST	Tuberculin Skin Test
WHO	World Health Organisation

Acknowledgements

The success of a thesis involves support from several individuals, whom I would like to thank as follows;

- The Director of the South African Tuberculosis Vaccine Initiative (SATVI) for according me time to do my PhD studies as well as letting me use data from one of the trials, protocol CORTIS, which was funded by GATES.
- My supervisors, Professors Mark Hatherill and Thomas Scriba for their guidance, mentorship, unwavering support including funding for testing of respiratory organisms, and leadership in both conducting my work, studies, as well as the write up of this thesis.
- Doctors Michele Tameris, Simon Mendelson, Gerhard Walzl, Kogieleum Naidoo and Gavin Churchyard, who were investigators on the CORTIS trial.
- The various study teams at all sites: clinical, field, laboratory, administrative and logistics, who were responsible for the successful conduct of the CORTIS trial and those at SATVI for the conduct of the Respiratory Organisms sub-study.
- The study participants themselves who gave up their time and willingly gave samples for both studies.
- Mrs. Mary Shelton, the Health Sciences' reference librarian, at the University of Cape Town, for her assistance in developing the search strategies for the systematic review and meta-analysis

Dedication

To Janet, Richard, Rosina and Sophie who were robbed of quality time during this period.

To the South African TB Vaccine Initiative staff who conducted both studies and those who may not have been involved directly, all of whom I will forever be indebted to, and forever be entrenched in my memory.

Expanded summary

Background

Tuberculosis (TB) is a major global health problem. Development of new rapid non-sputum biomarker-based tests based on target product profiles (TPP) published by the World Health Organisation (WHO) for the diagnosis of TB, and prediction of progression from latent infection to TB disease, is part of the strategy to address this problem. Multiple host blood transcriptomic messenger ribonucleic acid (mRNA) signatures offer promise as diagnostic and prognostic TB tests. An 11-gene mRNA signature (RISK11) for diagnosis and prediction of TB up to one year before onset of disease is one such biomarker.

This thesis aimed to review the literature on mNRA signatures of TB, to evaluate host and non-host factors associated with variability of the RISK11 signature, to evaluate impact of these factors on discriminatory performance, and to evaluate whether discriminatory performance can be improved by combination with other tests of *Mycobacterium tuberculosis* (MTB) sensitization.

Chapter 1 provides a general introduction, **Chapters 2-6** address the specific aims, and **Chapter 7** discusses the importance of the findings.

The specific aims were as follows:

1. Synthesise the diagnostic and prognostic performance of published mRNA transcriptomic signatures for TB disease; addressed in **Chapters 2 and 3.**
2. Evaluate the effect of host factors on discriminatory performance of the RISK11 transcriptomic signature for TB disease; addressed in **Chapter 4**.
3. Evaluate the effect of upper respiratory tract organisms on RISK11 score and discriminatory performance; addressed in **Chapter 5**.
4. Evaluate the diagnostic and prognostic performance of RISK11 in combination with a test of MTB sensitisation (interferon-gamma release assay, IGRA); addressed in **Chapter 6**.

Methods

A systematic review and meta-analysis of the diagnostic and prognostic performance of host blood mRNA signatures was conducted to accomplish **Aim 1 (Chapters 2 and 3**). Medline, Scopus, Web of Science, and EBSCO libraries were searched for articles published between January 2005 and May 2019. Data were extracted to construct 2 x 2 contingency tables of reference test against index test

results. Using RevMan 5.3 (Cochrane Collaboration), forest plots of sensitivity and specificity with 95% confidence intervals for each signature were plotted.

Effect of host factors on discriminatory performance of RISK11 (**Aim 2, Chapter 4**) and diagnostic and prognostic performance of RISK11 in combination with QuantiFERON-TB Gold Plus (QFTPlus) (**Aim 4, Chapters 6**) were evaluated using data from a prospective study of RISK11 (The "Correlates of Risk Targeted Intervention Study" or CORTIS). HIV-negative adult volunteers aged 18–60 years were enrolled and surveilled for TB through 15 months. Participants were tested for both QFTPlus and RISK11 positivity. To accomplish **Aim 2**, Generalised linear models and receiver-operating characteristic (ROC) regression were used to estimate effect of host factors on RISK11 score (%marginal effect) and on discriminatory performance for tuberculosis disease (area under the curve, AUC), respectively.

To evaluate the effect of upper respiratory tract organisms on RISK11 score and on discriminatory performance of RISK11 (**Aim 3, Chapter 5**), a cross-sectional sub-study of the upper respiratory microbiome was nested in CORTIS, in which adult volunteers aged 18–60 years were consecutively enrolled. Participants provided one nasopharyngeal and one oropharyngeal swab for detection of upper respiratory organisms using multiplex real-time polymerase-chain reaction (RT-PCR) and; a PAXgene blood sample for measurement of RISK11. A multivariable generalised linear model was used to estimate the effect of upper respiratory organisms on RISK11 score (% marginal effect). The AUC was used to differentiate participants with and without TB or other upper respiratory organisms; in ROC analysis.

To accomplish **Aim 4** (**Chapter 6**), qualitative RISK11 and QFTPlus results were combined using the two rules: 'Either-Positive' (positive combination test = +/- or -/+ or +/+; negative combination test = -/-) and 'Both-Positive' (positive combination test = +/+; negative combination test = -/- or +/- or -/+). Prevalence and incidence-rate ratios were used to evaluate probability of prevalent and risk of incident TB respectively; and positive (LR+) and negative (LR-) likelihood ratios were used to compare individual tests versus Both-Positive (RISK11+/QFTPlus+) and Either-Positive (RISK11+ or QFTPlus+) combinations.

Results

Aim 1, Chapters 2 and 3: Twenty studies evaluating 25 diagnostic or prognostic signatures for TB disease in 68 cohorts were included in the systematic review. Pooled sensitivity was 84% (95%CI 68–

93), 74% (95%CI 57–86) and 90% (95%CI 85–93), and pooled specificity 79% (95%CI 73–84), 71% (95%CI 49–86) and 74% (95%CI 56–86), respectively, for three diagnostic signatures validated in clinically relevant cohorts to differentiate TB from other diseases. Thus, only one of the three signatures met the WHO minimal TTP for a triage test. One prognostic signature met the minimal TPP for a test to predict progression to TB disease with sensitivity of 86% (95%CI 72–95) and specificity of 84% (95%CI 76–91).

Aim 2, Chapter 4: Among 2,923 participants evaluated in the parent prospective study, including 74 prevalent and 56 incident TB cases, percentage marginal effects on RISK11 score were predicted to be higher among those with prevalent TB (+18.90%, 95%CI 12.66–25.13), night sweats (+14.65%, 95%CI 5.39–23.91), incident TB (+7.29%, 95%CI 1.46–13.11), flu-like symptoms (+5.13%, 95%CI 1.58–8.68), and smoking history (+2.41%, 95%CI 0.89–3.93) than those without; and reduced in males (-6.68%, 95%CI -8.31–-5.04) and with every unit increase in BMI (-0.13%, 95%CI -0.25–-0.01). Adjustment for host factors only affecting controls did not change RISK11 discriminatory performance for prevalent or incident TB. However, in prevalent cases, presence of cough was associated with 72.55% higher RISK11 score. Stratification by cough improved diagnostic performance from AUC=0.74 (95%CI 0.67–0.82) overall, to 0.97 (95%CI 0.90–1.00, p<0.001) in cough-positive participants. Combining host factors with RISK11 improved prognostic performance, compared to RISK11 alone, (AUC=0.76, 95%CI 0.69–0.83 versus 0.56, 95%CI 0.46–0.68, p<0.001) over a 15-month predictive horizon.

.

Aim 3, Chapter 5: Among the 1,000 participants enrolled in the upper respiratory organisms sub-study, non-HIV viral and non-tuberculous bacterial organisms were detected in 7.2% and 38.9%, respectively. In the 286 participants co-enrolled in the prospective study and investigated for TB, 3.8% (11/286) and 3.2% (9/286) were diagnosed with prevalent and incident TB, respectively. The proportion of individuals with upper respiratory viral organisms was significantly higher (p=0.02) in those who progressed to incident TB (44.4%, 4/9) than in those who remained healthy (12.4%, 33/266). Participants with a virus detected were five times more likely to progress to TB than those without a virus detected (Incident Rate Ratio; IRR 5.0, 95% CI 1.0–23.2). In multivariable generalised linear regression, percent marginal effects on RISK11 score were predicted to be higher by +16.7% (95%CI 4.1%–29.4%), +67.8% (95%CI 52%–83.5%) and +13.5% (95%CI 3.5%–23.5%) in participants with positive PCR for coronavirus, influenza and rhinovirus, respectively, compared to those without these viruses. Presence of upper respiratory tract viruses negatively impacted signature performance such that RISK11 could not discriminate prevalent TB from viruses (AUC=48.4%; 95%CI 27.5%–69.5%).

Aim 4, Chapter 6: Among 2912 participants with both RISK11 and QFTPlus results, risk of prevalent TB in RISK11+/QFTPlus+ participants was 13.3-fold (95%CI 4.2–42.7) higher than RISK11-/QFTPlus-; 2.4–fold (95%CI 1.2-4.8) higher than RISK11+/QFTPlus-; and 4.5-fold (95%CI 2.5–8.0) higher than RISK11-/QFTPlus+ participants, respectively. Risk of incident TB in RISK11+/QFTPlus+ participants was 8.3-fold (95%CI 2.5–27.0) higher than RISK11-/QFTPlus-; 2.5-fold (95%CI 1.0–6.6) higher than RISK11+/QFTPlus-; and 2.1-fold (95%CI 1.2–3.4) higher than RISK11-/QFTPlus+ participants, respectively. The Both-Positive RISK11/QFTPlus test combination increased diagnostic LR+ from 1.3 (95%CI 1.2–1.5) to 4.7 (95%CI 3.2–7.0), and prognostic LR+ from 1.4 (95%CI 1.2–1.5) to 2.8 (95%CI 1.5–5.1), compared to QFTPlus, but did not improve upon RISK11 alone. The Either-Positive test combination decreased diagnostic LR- from 0.7 (95%CI 0.6–0.9) to 0.3 (95%CI 0.2–0.6), and prognostic LR- from 0.9 (95%CI 0.8–1.0) to 0.3 (0.1–0.7), compared to RISK11, but did not improve upon QFTPlus alone

Conclusion

Host blood mRNA signatures including RISK11 hold promise as triage tests for TB, but further optimisation is needed if mRNA signatures are to be used as standalone diagnostic, triage, or predictive tests for therapeutic decision-making. Host factors affecting controls do not affect RISK11 performance, but cough status does affect performance in prevalent TB cases. RISK11 could not discriminate between TB and upper respiratory tract viruses, likely due to induction of interferon-signalling genes by upper respiratory tract viruses, which appear to be important confounding factors that negatively affect performance of transcriptomic signatures of TB and pose a major challenge for implementation of these biomarkers as new tests for TB. New signatures are needed to improve discrimination of subclinical TB, particularly without cough, and to mitigate the impact of viral co-infection. Further, to improve prognostic performance, combining host factors with RISK11 might be considered. Although clinical utility of all possible RISK11/QFTPlus test combinations would depend on whether the goal of testing is to rule-in, or rule-out, TB risk; RISK11/QFTPlus combination testing is not recommended without reservation, because it does not improve overall discriminatory performance, relative to the individual tests.

Important notes.

Note on the thesis author

The author of this thesis was responsible for the design, conduct and analysis of the systematic review. He was primarily responsible for all data management activities for the CORTIS study, the parent study on which this thesis is based, including oversight of data collection, query resolution and data quality management, and specifically responsible for the design, conduct and interpretation of all analyses based on the CORTIS dataset that are presented in this thesis. He was the investigator primarily responsible for the design, protocol development, operational conduct, data collection, analysis and reporting of the respiratory organisms sub-study, under supervision of his supervisors. He was the first author and analysed the data and wrote the first drafts of all published manuscripts included in this thesis under supervision of his supervisors. Specific contributions by the author and all co-authors are given under each chapter.

Note on included publications:

The final accepted versions of the manuscripts are included in this thesis verbatim, reformatted to match the format, themes and styles used in the thesis as a whole and with tables and figures renumbered, as prescribed by the doctoral degrees board. All journals for the included manuscripts used the “numbered referencing style”, therefore the entire thesis has adopted the same style for consistency. The tables and figures for the manuscripts are shown in the text, rather than at the end of each chapter, for readability. This layout may vary from the applicable journal rules. Where such variation exists, it will be stated as a “variation note” in the applicable chapter.

Note on references

Each chapter has its own reference list.

Chapter 1

1 General introduction

1.1 Background

Tuberculosis (TB) is a major public health problem. An estimated 10 million new cases of TB and 1.5 million deaths attributable to TB were reported by the World Health Organisation (WHO) in 2020.[1] Effective TB control requires that individuals who develop TB disease are quickly identified and treated before transmitting to others. However, current TB diagnostic tools perform suboptimally.[2-5] *Mycobacterium tuberculosis* (MTB) culture is the gold standard but takes too long (14 to 42 days) to get a confirmatory result, which is detrimental to rapid patient care.[6, 7] Sputum smear microscopy, commonly used in many TB endemic countries, misses a substantial amount of pulmonary TB (PTB) cases because of low sensitivity.[4, 5] Sputum Xpert MTB/RIF and Xpert Ultra are relatively better diagnostic tests[8, 9], but are also reliant on a sufficient sputum sample, and require specialised equipment and consistent power supply, which hampers routine screening in TB-endemic resource-limited settings.[10] A chest radiograph (CXR) has low specificity and may give false-negative results in early stages of TB disease.[11]

People with latent TB infection (LTBI) identified by showing immunological sensitisation to MTB have an increased risk of progression to TB disease compared to MTB-unsensitised people. The risk of progression is highest in individuals within two years of MTB-sensitisation.[12-14] Prevention of active TB disease is critical for TB elimination and requires that MTB-sensitised individuals are promptly identified and treated to interrupt progression.[15] The tuberculin skin test (TST) and interferon gamma release assay (IGRA) form the mainstay for diagnosis of MTB-sensitisation.[3, 16] However, both IGRA and TST have poor positive predictive value[17] and low specificity[16] for incident TB disease; of people that test IGRA or TST positive, only around 10-15% will progress to TB disease.[15] This means that relying on IGRA or TST for mass preventive therapy, would require treating majority of the 1.7 billion MTB-sensitised people, most of them needlessly, because they will remain healthy. This approach would be unaffordable and potentially ineffective, because re-infection would likely occur before programmatic coverage was complete.[18] Thus, the most effective way is to identify and target preventive therapy to the 10-15% of the people that will progress to TB; and thus, benefit from treatment.

Management of TB is currently undertaken on binary approach, where individuals with ideally microbiologically confirmed active TB disease are treated with a standard multi-drug 6-month regimen and those with latent TB infection without evidence of disease, treated with a one or two-drug prophylactic regimen.[19] However, this dual system is a simplification of the TB continuum for clinical and public health management purposes.[20] Recent evidence indicates that TB exists as a spectrum of disease states ranging from TB infection, incipient TB, subclinical TB, and active TB disease. [19-22] LTBI refers to a state in which the individual shows immune sensitisation to MTB, is asymptomatic, has a normal CXR, is bacteriologically negative, and progression to TB disease. Incipient TB is infection with viable MTB and a positive biomarker of progression to TB disease in the absence of intervention, but without clinical symptoms, CXR abnormalities, or microbiologic evidence of TB, at the time of biomarker sampling. Note that the state of incipient TB is only confirmed retrospectively by observation of progression to TB disease and that many individuals identified by a positive biomarker do not ultimately progress to TB disease. It follows that incipient TB can only be diagnosed by prospective follow-up and cannot be diagnosed in a cross-sectional study. Subclinical TB refers to a disease state in which the individual is asymptomatic, but has microbiologic evidence of TB, with or without CXR abnormalities. Active TB disease refers to a disease in which the individual is symptomatic, with either CXR abnormalities, or bacteriological confirmation or both.[19]

WHO has set out a strategy to combat TB globally, aiming to reduce new cases of TB by 90% and deaths due to TB by 95%, by the year 2035 compared to 2015 levels.[23] Included in the third and final pillar of this strategy is the need for intensified research and innovation; to support the discovery, development and rapid implementation of new tools, interventions and approaches to fight TB. Discovery and development of an accurate, biomarker-based, and rapid point-of-care (POC) test that can be used under field settings for diagnosis, triaging or predicting progression to TB disease, is one such strategy highlighted under the third pillar. A test that is quicker, easily implemented, more affordable and which can accurately discriminate between TB disease and other illnesses, in addition to predicting who will develop TB disease after MTB-infection would be beneficial in the fight for TB. Consequently, WHO has identified optimal target product profiles (TPPs) for rapid non-sputum-based biomarker-based diagnostic, triage, and predictive tests of TB disease. The TPPs stipulate the following performance metrics: for a diagnostic test (designed for making a therapeutic decision), minimum 65% sensitivity and 98% specificity, optimal >80% sensitivity in all case types and ≥98% specificity; for a triage test (designed to identify individuals that should undergo further confirmatory testing), minimum 90% sensitivity and 70% specificity, optimal >95% sensitivity overall and >80% specificity; and for a test to predict progression to TB disease within two years of MTB-sensitisation (designed to

identify individuals that will develop TB in future), minimum 75% sensitivity and 75% specificity, optimal ≥90% sensitivity and ≥90% specificity.[24, 25] Several host blood transcriptional signatures offer promise as diagnostic, triage and prognostic tests of TB.[26-28] These transcriptional signatures will especially be important for the diagnosis of incipient TB as it is asymptomatic and bacteriologically negative.[21]

Scientists at the South African Tuberculosis Vaccine Initiative (SATVI), University of Cape Town (UCT), have previously discovered and validated a transcriptomic signature of risk (RISK11), based on messenger Ribonucleic Acid (mRNA) expression of 11 interferon-stimulated gene (ISG) signatures.[29] This smaller 11-gene signature has equivalent performance to a previously validated 16-gene signature, from which it was derived.[28]

The RISK11 signature is a model of multiple transcript pairs, each functioning as a "vote" for or against TB risk. RISK11 scores are computed from cycle threshold (Cq) values for each mRNA transcript representing the 11 genes, measured by microfluidic qRT-PCR. The RISK11 score is the proportion of votes for risk of TB; and a score threshold can be set for the RISK11 assay to function as a qualitative (positive/negative) test for TB risk. In prior studies, RISK11 prospectively differentiated between incident TB cases and healthy controls. In receiver operating characteristic (ROC) curve analysis, RISK11 classified progressors and control samples with an area under the curve (AUC) of 0.84. RISK11 also has diagnostic utility and could be used as a triage test to identify undiagnosed TB; it showed a high diagnostic performance, distinguishing TB disease cases from healthy QuantiFERON-positive (QFT+) controls with AUCs of 0.97 (95%CI 91–100) and 0.98 (95%CI 95–100) in whole blood and PBMC samples respectively[29] In another study, the RISK11 signature had 100% and 65% sensitivity in HIV-negative and HIV-positive individuals respectively, with a specificity of 80% for both groups.[30] Thus, RISK11 has potential both as a test to identify progressors to TB disease, and as a triage test to identify undiagnosed TB.

1.2 Study design and setting

Three studies contributed data to this thesis: a systematic review and meta-analysis of the mRNA transcriptomic signature literature was used to address Aim 1 in Chapters 2 and 3; a randomised partially blinded clinical trial (parent study) was used to address Aims 2 and 4 in Chapters 4 and 6, respectively; and a cross-sectional sub-study of upper respiratory organisms nested in the parent study was used to address Aim 3 in Chapter 5. The parent longitudinal study, "The Correlates of Risk Targeted Intervention Study" (CORTIS), was conducted at five sites across South Africa. The five areas

from which participants were recruited for CORTIS were Worcester for SATVI, Khayelitsha for Stellenbosch University (SUN), Rustenburg and Klerksdorp for Aurum, and Durban for CAPRISA. All these five sites are in TB endemic areas. The respiratory organisms sub-study only recruited participants from the SATVI site at Worcester. All studies including the systematic review and meta-analysis were conducted in HIV-uninfected people.

1.3 Thesis structure and aims

The systematic review (Chapters 2 & 3) provides a background perspective for subsequent chapters that describe effect of host factors on discriminatory performance of RISK11 for TB disease (Chapter 4) and effect of upper respiratory organisms on RISK11 score and discriminatory performance (Chapter 5). Chapter 6 focuses on evaluating the diagnostic and prognostic performance of RISK11 in combination with QFTPlus. The rationale for the individual chapters is described below:

Prior to the conduct of this study, several host blood transcriptomic signatures for diagnosing and predicting progression to TB disease had been developed, including RISK11. Therefore, it was considered important to perform a synthesis of the published literature on the performance of all known mRNA signatures for TB disease published at the time through a systematic review and meta-analysis to accumulate strong evidence (Aim 1, Chapters 2 and 3).

Despite the potential of RISK11 as a screening test to rule out TB disease, to direct further investigations, and as a test for predicting incident TB disease and to inform therapy, it was not known whether or which epidemiological and other host factors might affect RISK11; and the direction of possible associations were not known. It was also not known if RISK11 score increased or decreased as a result of variation in one or more covariates. Therefore, identifying epidemiological covariates that might be associated with variation in RISK11 score was key to understanding the factors that may affect discriminatory performance of RISK11. Thus, Aim 2 (Chapter 4), was designed to address this gap in the knowledge.

Detectable HIV viral load has been associated with raised transcriptomic signature scores, compared to undetectable viral load, possibly because of induction of type I interferon (IFN) and raised expression of IFN-stimulated genes (ISG), which are preferentially included in RISK11 and similar TB signatures.[26, 30] Other viruses such as influenza are also known to induce ISGs and have been shown to affect transcriptomic signatures scores.[31] However, the effect of other common viral and bacterial

upper respiratory organisms on TB signatures, and specifically on RISK11, remains largely unexplored. Aim 3 (Chapter 5), was therefore designed to answer this question.

Studies have shown that improvements in test performance may be attained by using tests in combination. It was considered important to test whether combining the RISK11 signature with IGRA would increase diagnostic and/or prognostic performance for TB and improve the utility of these tests for rule-in or rule-out clinical scenarios in which risk of TB is suspected. Thus, the objective of Aim 4 (Chapter 6), was to answer this question.

The aims of this thesis were to:

1. Synthesise the diagnostic and prognostic performance of published mRNA transcriptomic signatures for TB disease; addressed in Chapters 2 and 3.
2. Evaluate the effect of host factors on discriminatory performance of the RISK11 transcriptomic signature for TB disease; addressed in Chapter 4.
3. Evaluate the effect of upper respiratory tract organisms on RISK11 score and discriminatory performance; addressed in Chapter 5.
4. Evaluate the diagnostic and prognostic performance of RISK11 in a combination with a test of MTB sensitisation (IGRA); addressed in Chapter 6.

1.4 Ethics

The parent study, CORTIS, was approved by Institutional Human Research Ethics Committees of each of the five participating sites and registered with ClinicalTrials.gov (NCT02735590). The PhD study protocol was approved by the UCT-Human Research Ethics Committee (HREC REF# 327/2017). The systematic review protocol was designed, registered on PROSPERO, and published prior to conduct of the review (PROSPERO registration number CRD42017073817).

1.5 References

1 WHO. Global Tuberculosis Report 2019. Geneva, Switzerland: World Health Organization; 2019.

2 Andrews JR, Noubary F, Walensky RP, Cerda R, Losina E, Horsburgh CR. Risk of progression to active tuberculosis following reinfection with Mycobacterium tuberculosis. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America.* 2012;54(6):784-91.

3 Andrews JR, Hatherill M, Mohamed H, et al. The Dynamics of QuantiFERON-TB Gold In-Tube Conversion and Reversion in a Cohort of South African Adolescents. *American journal of respiratory and critical care medicine.* 2015;191(5):584-91.

4 Harries AD, Maher D, Nunn P. An approach to the problems of diagnosing and treating adult smear-negative pulmonary tuberculosis in high-HIV-prevalence settings in sub-Saharan Africa. *Bulletin of the World Health Organization.* 1998;76(6):651-62.

5 Davies PD, Pai M. The diagnosis and misdiagnosis of tuberculosis. *Int J Tuberc Lung Dis*. 2008;12(11):1226-34.

6 Ghodbane R, Raoult D, Drancourt M. Dramatic reduction of culture time of Mycobacterium tuberculosis. *Sci Rep.* 2014;4:4236.

7 Norbis L, Miotto P, Alagna R, Cirillo DM. Tuberculosis: lights and shadows in the current diagnostic landscape. *New Microbiol.* 2013;36(2):111-20.

8 Li S, Liu B, Peng M, et al. Diagnostic accuracy of Xpert MTB/RIF for tuberculosis detection in different regions with different endemic burden: A systematic review and meta-analysis. *PloS one.* 2017;12(7):e0180725.

9 Dorman SE, Schumacher SG, Alland D, et al. Xpert MTB/RIF Ultra for detection of Mycobacterium tuberculosis and rifampicin resistance: a prospective multicentre diagnostic accuracy study. *The Lancet Infectious diseases.* 2018;18(1):76-84.

10 Maertzdorf J, Kaufmann SH, Weiner J, 3rd. Toward a unified biosignature for tuberculosis. *Cold Spring Harbor perspectives in medicine.* 2014;5(1):a018531.

11 Harris M, Qi A, Jeagal L, et al. A systematic review of the diagnostic accuracy of artificial intelligence-based computer programs to analyze chest x-rays for pulmonary tuberculosis. *PloS one.* 2019;14(9):e0221339.

12 Getahun H, Matteelli A, Chaisson RE, Raviglione M. Latent Mycobacterium tuberculosis infection. *The New England journal of medicine.* 2015;372(22):2127-35.

13 Trauer JM, Moyo N, Tay EL, et al. Risk of Active Tuberculosis in the Five Years Following Infection . . . 15%? Chest. 2016;149(2):516-25.

14 Behr MA, Edelstein PH, Ramakrishnan L. Revisiting the timetable of tuberculosis. *BMJ.* 2018;362:k2738.

15 Dye C, Glaziou P, Floyd K, Raviglione M. Prospects for tuberculosis elimination. *Annu Rev Public Health.* 2013;34:271-86.

16 Mahomed H, Hawkridge T, Verver S, et al. The tuberculin skin test versus QuantiFERON TB Gold(R) in predicting tuberculosis disease in an adolescent cohort study in South Africa. *PLoS One.* 2011;6(3):e17984.

17 Abubakar I, Drobniewski F, Southern J, et al. Prognostic value of interferon-gamma release assays and tuberculin skin test in predicting the development of active tuberculosis (UK PREDICT TB): a prospective cohort study. *The Lancet Infectious diseases.* 2018;18(10):1077-87.

18 Sumner T, Bozzani F, Mudzengi D, et al. Estimating the Impact of Tuberculosis Case Detection in Constrained Health Systems: An Example of Case-Finding in South Africa. *Am J Epidemiol.* 2019;188(6):1155-64.

19 Esmail H, Macpherson L, Coussens AK, Houben R. Mind the gap - Managing tuberculosis across the disease spectrum. *EBioMedicine.* 2022;78:103928.

20 Pai M, Behr MA, Dowdy D, et al. Tuberculosis. Nat Rev Dis Primers. 2016;2:16076.

21 Migliori GB, Ong CWM, Petrone L, D'Ambrosio L, Centis R, Goletti D. The definition of tuberculosis infection based on the spectrum of tuberculosis disease. *Breathe (Sheff).* 2021;17(3):210079.

22 Drain PK, Bajema KL, Dowdy D, et al. Incipient and Subclinical Tuberculosis: a Clinical Review of Early Stages and Progression of Infection. Clinical microbiology reviews. 2018;31(4).

23 WHO. The End TB Strategy: Global strategy and targets for tuberculosis prevention, care and control after 2015. Geneva, Switzerland: *World Health Organization*; 2015.

24 WHO. Meeting Report : High-priority target product profiles for new tuberculosis diagnostics: report of a consensus meeting. Geneva, Switzerland: *World Health Organization;* 2014. Contract No.: WHO/HTM/TB/2014.18.

25 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. Report. Geneva: *World Health Organization*; 2017. Contract No.: WHO/HTM/TB/2017.18.

26 Sweeney TE, Braviak L, Tato CM, Khatri P. Genome-wide expression for diagnosis of pulmonary tuberculosis: a multicohort analysis. *The Lancet Respiratory medicine.* 2016;4(3):213-24.

27 Warsinske H, Vashisht R, Khatri P. Host-response-based gene signatures for tuberculosis diagnosis: A systematic comparison of 16 signatures. *PLoS medicine.* 2019;16(4):e1002786.

28 Zak DE, Penn-Nicholson A, Scriba TJ, et al. A blood RNA signature for tuberculosis disease risk: a prospective cohort study. *Lancet (London, England).* 2016;387(10035):2312-22.

29 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.* 2018;108:124-6.

30 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. *Frontiers in microbiology.* 2019;10:1441.

31 Singhania A, Verma R, Graham CM, et al. A modular transcriptional signature identifies phenotypic heterogeneity of human tuberculosis infection. *Nature communications.* 2018;9(1):2308.

Chapter 2

2 Performance of host blood transcriptomic signatures for diagnosing and predicting progression to tuberculosis disease in HIV-negative adults and adolescents: a systematic review protocol.

Chapter overview

This chapter is a protocol that provides the methods for the systematic review and meta-analysis. It was written, registered with PROSPERO, and published prior to conducting the systematic review. The systematic review protocol is presented as published.

Mulenga H, Bunyasi EW, Mbandi SK, Mendelsohn SC, Kagina B, Penn-Nicholson A, Scriba T, and Hatherill M. *Performance of host blood transcriptomic signatures for diagnosing and predicting progression to tuberculosis disease in HIV-negative adults and adolescents: a systematic review protocol*. BMJ Open, 2019. 9(5): p. e026612. (http://dx.doi.org/10.1136/bmjopen-2018-026612).

Chapter contribution to the thesis

The systematic review protocol provides the methods for the literature review and addresses Aim 1.

Contributions of the candidate

The candidate designed the search strategy, data extraction forms and study quality assessment tool that was to be used in the systematic review. Additionally, the candidate was the first author and planned, wrote, reviewed, revised, and approved the final version of the manuscript for publication. Furthermore, the candidate registered the protocol on PROSPERO, the online register for systematic review protocols. MH and TS conceived the idea and provided supervision to the candidate. All co-authors, EWB, SKM, SCM, BK, AP-N, TS and MH were involved in the review, revision, and approval of the final version of the protocol.

Publication

Performance of host blood transcriptomic signatures for diagnosing and predicting progression to tuberculosis disease in HIV-negative adults and adolescents: a systematic review protocol.

Humphrey Mulenga[1], Erick W. Bunyasi[1], Stanley Kimbung Mbandi[1], Simon C. Mendelsohn[1], Benjamin Kagina[2], Adam Penn-Nicholson[1], Thomas J. Scriba[1*], Mark Hatherill[1*]

Author affiliations:

[1]South African Tuberculosis Vaccine Initiative (SATVI), Institute of Infectious Disease & Molecular Medicine and Division of Immunology, Department of Pathology, University of Cape Town, Cape Town, South Africa

[2]Vaccines for Africa Initiative (VACFA), Institute of Infectious Disease & Molecular Medicine and Division of Medical Microbiology, University of Cape Town, Cape Town, South Africa.

Correspondence: Professor Mark Hatherill, South African Tuberculosis Vaccine Initiative (SATVI), Institute of Infectious Disease & Molecular Medicine and Division of Immunology, Department of Pathology, S2.11 Werner Beit South Building, Faculty of Health Sciences, University of Cape Town, Anzio Road, Observatory, 7925, South Africa.

Tel. +27 21 406 6791

Email. Mark.Hatherill@uct.ac.za

2.1 Abstract

Introduction

One quarter of the global population, including the majority of adults in tuberculosis (TB) endemic countries, are estimated to be *Mycobacterium tuberculosis* (MTB) infected. An estimated 10 million new TB cases occurred in 2017. One of the biggest challenges confronting TB control is the lack of accurate diagnosis and prediction of prevalent and incident TB disease respectively. Several host blood transcriptomic (mRNA) signatures that reflect the host immune response following infection with MTB and progression to TB disease in different study populations have recently been published, but these TB biomarkers have not been systematically described. We will conduct a systematic review of the performance of host blood transcriptional signatures for TB diagnosis and prediction of progression to TB disease.

Methods and analysis

This systematic review will involve conducting a comprehensive literature search of cohort, case-control, cross-sectional, and randomised-controlled studies of the performance of host blood transcriptomic signatures for TB diagnosis and prediction of progression to TB disease. We will search Medline via PubMed, Scopus, Web of Science, and EBSCO libraries, complemented by a search of bibliographies of selected articles for other relevant articles. The literature search will be restricted to studies published in English from 2005 to 2018 and conducted in HIV-uninfected adults and adolescents (≥ 12 years old). Forest plots and a narrative synthesis of the findings will be provided. The primary outcomes will be sensitivity, specificity, as well as true/false positives and true/false negatives. Heterogeneity resulting from differences in the design, composition and structure of individual signatures will preclude meta-analysis and pooling of results.

Ethics and dissemination

Ethics approval is not required for this systematic review protocol. The results of this review will be disseminated through a peer-reviewed journal as well as conference presentations.

PROSPERO registration number CRD42017073817

Strengths and limitations of this study

This will be the first systematic review of the performance of host blood transcriptomic signatures for the diagnosis of prevalent TB and prediction of incident TB disease in adults and adolescents.

- Data reporting will adhere to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines for reviews and protocols.
- Included studies will be restricted to those published in English, which may introduce publication and language bias.
- The design/composition/structure of individual signatures are expected to be significantly heterogeneous, precluding meta-analysis and pooling of results.

2.2 Introduction

Tuberculosis (TB) is the most common cause of infectious disease mortality worldwide, yet TB control remains a major public health challenge, because it is difficult to predict and prevent, diagnose, and treat. TB disease is caused by the bacillus *Mycobacterium tuberculosis* (MTB) and is transmitted by inhaled droplet spread from individuals with active disease. Healthy individuals who are exposed to aerosolised MTB bacilli may develop infection, which may be cleared, contained as latent MTB infection, or, if containment is unsuccessful, progress to active TB disease known as primary TB. Latent MTB infection may also progress to active TB disease at a later stage known as post primary TB.[1] The ultimate result of exposure to MTB bacilli is determined by a range of environmental, sociological, mycobacterial, and host immune factors.[2]

An estimated 1.7 billion individuals or 23% of the world population, including the majority of adults in TB endemic countries, are MTB infected.[3] There were 10 million new cases of TB disease in 2017, of which 90% occurred in adults.[4] One of the 2030 Sustainable Development Goals adopted by the United Nations in 2015 is to end the global TB epidemic. The End TB Strategy demands that new cases of TB should be reduced by 80% from 2015 levels by the year 2030, and deaths occurring due to TB should be reduced by 90% for the same period.[5] In order to reduce new TB cases and deaths to meet the set targets, major advances in TB drugs, vaccines and diagnostics are critical.

Currently available TB diagnostic tests have important drawbacks especially if applied as a screening test, thus making TB diagnosis difficult.[6-9] Sputum smear microscopy, still used in many high burden TB countries, has low sensitivity[9] ranging from 32% to 89%[10] resulting in a considerable number of active pulmonary TB (PTB) patients being missed.[8] Xpert MTB/RIF has considerably better diagnostic performance with sensitivity of 77% and specificity of 99%.[11] However, Xpert MTB/RIF is relatively unaffordable in resource-limited settings and has technical limitations such as the need for special equipment as well as a reliable power supply, thereby impeding routine screening in TB-endemic

resource-limited settings.[12] MTB culture, the gold standard, delays TB diagnosis as it usually takes more than 2 weeks (up to 42 days) to get a confirmatory result, and this is not ideal for rapid patient management.[1, 13] Furthermore, MTB culture requires a reference laboratory and is relatively costly. A chest radiograph (CXR) is inconclusive for PTB diagnosis as it may yield false-negative results particularly when the disease is in its initial phases.[14] It may also yield false-positives in individuals with lung damage from prior TB disease or other lung diseases. The inability of CXR to accurately differentiate between the many abnormalities consistent with TB from those of other lung pathologies restricts its specificity, which ranges between 46% and 89%.[14-16] Furthermore, readout of a CXR is highly dependent on a skilled interpretation and a level of subjectivity, which is problematic for low resource settings. Symptom screening alone has a low specificity in diagnosis of PTB, especially in HIV-infected individuals. In HIV-uninfected individuals and individuals of unknown HIV-status, symptom screening has a sensitivity and specificity of about 77% and 68% respectively.[16]

Latently MTB infected individuals, identified by a positive tuberculin skin test (TST) or interferon-gamma release assay (IGRA), have a higher risk of developing TB disease than uninfected people.[7, 17] However, IGRA and TST have poor specificity for incident TB disease (49.3% and 45% respectively) and hence predicting incident TB disease remains difficult.[17] This problem is further compounded by the fact that, in TB endemic populations, up to 90% of people who test IGRA or TST positive will not go on to develop active TB disease.[6, 18] A systematic review showed that the positive predictive values (PPVs) for these current predictive tools are too low to have clinical utility in directing use of preventive therapy[19] for high TB burden settings. The PPVs for progression from latent MTB infection to TB disease in all settings were 2.7% and 1.5% for IGRA and TST respectively. In high-risk groups, the PPVs increased marginally to 6.8% for IGRA, and 2.4% for TST.[19] Although prevention of TB disease arising from latent MTB infection is key to achieving WHO elimination targets[20], mass preventive therapy based on IGRA or TST screening in TB endemic countries would need to treat 50% to 80% of the population, most of them unnecessarily. Many incident TB cases would also be missed due to poor sensitivity (IGRA=75% and TST=77%).[17] Mass preventive therapy for all MTB infected people using current tools would not be feasible, affordable, or effective, because reinfection would likely occur before programmatic coverage was complete. More specific predictive tools are needed to identify those individuals who would most benefit from preventive therapy. Given the inadequacies of current diagnostic tools, more sensitive, highly specific, quicker, and much more affordable tests that differentiate active TB from healthy individuals, latent MTB infection, and other diseases, as well as predict progression from latent MTB infection to active disease, are needed. Advances in TB

prevention, prediction, diagnosis and treatment are impeded by the fact that the immunological basis for progression from MTB infection to disease is poorly understood.[2, 21]

In recent years, host blood transcriptomic (mRNA) signatures have provided a promising alternative for both TB disease diagnosis and prediction of progression to TB disease. Transcriptional signatures of TB have also provided better understanding of the TB-specific immune mechanisms[22] in individuals with MTB infection[23] and those with active TB disease.[24] Several studies of host blood transcriptomic signatures have shown that individuals with prevalent TB disease can be discriminated from those who are uninfected, latently MTB infected, or suffering from another disease.[25] Diagnostic sensitivity has ranged between 61% and 100% while specificity has ranged between 75% and 97% for active TB versus latent MTB infection, or active TB versus other diseases.[12, 26] Recent work has also shown that transcriptomic signatures can predict the development of TB disease in individuals with MTB infection. A 16 gene signature of risk predicted progression from latent MTB infection to TB disease with a sensitivity of 66.1% and a specificity of 80.6% in the 12 months prior to TB diagnosis.[27] Validation of this signature in an independent cohort of household contacts of active TB patients, predicted progression to TB disease with a sensitivity of 53.7% and a specificity of 82.8% in the 12 months preceding TB diagnosis.[27] Recently, Suliman *et al* reported that a four-gene signature of risk predicted progression to TB disease in household contacts of active TB disease with an area under the curve (AUC) of 0.66 in the 12 months prior to TB diagnosis.[28] However, the performance of host blood transcriptomic signatures for diagnosis of prevalent and prediction of incident TB disease has not been synthesised and examined systematically. Consequently, we will conduct a systematic review aiming to describe and summarise the performance of the currently available host blood transcriptomic signatures for diagnosing and predicting TB disease.

2.3 Research question and aims

2.3.1 Research question

What are the performance characteristics of host blood transcriptomic signatures for diagnosing prevalent TB disease and predicting incident TB disease in HIV-negative adolescents and adults?

2.3.2 Objectives

1. To describe the performance of host blood transcriptomic signatures for diagnosis of TB disease in HIV-negative adolescents and adults
2. To describe the performance of host blood transcriptomic signatures for predicting progression to TB disease in HIV-negative adolescents and adults

2.4 Methods and analysis

This protocol conforms to the Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P)[29] (Appendix 2.1) and the Cochrane Collaboration's diagnostic test accuracy methods for evidence searching and synthesis.[30, 31] Our review methodology will include a thorough literature search, examination of studies identified, and selection of studies using predefined criteria. We will then extract data from the included studies, evaluate methodological quality, summarise it, and rate the quality of evidence from our systematic review. The statistical analysis, evidence synthesis, and reporting of findings will be performed according to the guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA).[32] The study will be conducted and completed between February 2019 and June 2019. This review is registered with the International Prospective Register of Systematic Reviews (PROSPERO), registration number CRD42017073817.

2.4.1 Definitions and study inclusion criteria

(i) Definitions

- Predictive studies: prospective studies of progression to TB disease in which biomarker status is assigned at enrolment and TB case/control status is assigned at a later time point, after at least 6 months of follow-up.
- Diagnostic studies: studies in which both biomarker status and TB case/control status are assigned at the same time point.
- Latent MTB infection is defined as a positive TST ≥ 5mm; or positive IGRA conforming to the manufacturer's instructions and cut-off point.
- TB disease is defined as either or both pulmonary TB (PTB) and extrapulmonary TB (EPTB) diagnosed with microbiological confirmation by MTB culture or Xpert MTB/RIF or smear microscopy.
- A TB contact is defined as a person in close contact with, or living in the same household as, an individual diagnosed with active TB disease within the past 6 months.

We will select studies meeting all the following criteria:

(ii) General inclusion criteria

- Cohort, case-control, cross-sectional and randomised control studies conducted in HIV-uninfected humans.
- Studies using host blood transcriptomic (mRNA) signatures for diagnosis or prediction of TB disease with a microbiological reference standard of either MTB culture or Xpert MTB/RIF or smear microscopy.
- Studies using either TST or IGRA for the diagnosis of latent MTB infection.

- Studies comparing TB disease cases versus controls with or without other diseases, and with or without latent MTB infection. Both PTB and EPTB cases will be included.
- Studies reporting either discovery or validation of a host blood transcriptomic (mRNA) signature.
- Studies conducted in adults or adolescents (≥ 12 years old).
- Studies published both as abstracts and full articles after 2005–2018.
- Studies published in English regardless of location or country of origin.
- Studies reporting sensitivity and specificity; or reporting results enabling the recreation of a 2x2 table for test performance calculation, or studies where we receive a response on test performance data within 4 weeks of inquiry.

(iii) Additional inclusion criteria for TB predictive studies

- Studies with a follow-up period of at least 6 months from enrolment.
- Studies enrolling TB contacts, latently MTB infected individuals or healthy individuals.
- Randomised controlled trials or prospective cohort studies

2.4.2 Literature search

The primary electronic searches will be conducted in Medline via PubMed, Scopus, Web of Science, and EBSCO databases. The search strategy will employ a combination of database specific Medical Subject Heading terms and other key words that include but not limited to TB, Tuberculosis, *Mycobacterium tuberculosis*, *M. tuberculosis*, MTB, diagnosis, diagnostic, detect, prognosis, prognostic, predict, blood, host, human, biomarker, signature, bio-signature, transcriptome, transcriptomic, RNA, sensitivity and specificity, accuracy, diagnostic accuracy, performance, area under the curve, AUC, receiver operating characteristic, and ROC. The initial PubMed search strategy is availed as online S2 (Appendix 2.2). The finalised PubMed search strategy will be adapted to other databases and will be published in the systematic review. Furthermore, bibliographies of included papers will be scrutinised for potential papers to include in the review that would otherwise have been missed by the search term. Unpublished reports and conference proceedings/papers will not be included due to absence of peer review and difficulties in obtaining data. We recognise that this shortcoming may result in publication bias.

2.4.3 Data management

The first author (HM) will conduct the data management activities. A google drive account will be created and maintained for the systematic review. All documents relating to the conduct of this review, such as a record of the search strategy and identified articles, protocol, individual study quality assessment records, and other supplementary material will be uploaded to this google drive folder.

Additionally, a database will be developed using Microsoft SqlServer 2012 as the back-end and forms in Microsoft Access 2010 as the user interface, to manage individual data metrics extracted from the articles. This will enable electronic and quick comparison of the extracted data as well as inclusion/exclusion decisions between HM and EWB. EndNote referencing software will be used to manage the titles of identified articles and references during study selection and write up. A backup of all the records will also be kept on SATVI's server as well as the laptop from which this work will be carried out.

2.4.4 Study selection

Two reviewers, HM and EWB will independently screen the search outputs for potentially qualifying studies. The selection process will initially involve importing all articles returned by the search strategy into EndNote software using distinct groups (folders) for each literature source. Once all articles have been imported into their respective groups, another group will be created which will contain all articles from these subgroups, including duplicates. Duplicates will then be removed by creating the final group into which distinct titles will be stored. HM will then import all distinct studies into the Microsoft SqlServer database and assign a unique study identification number. Only the article title, first-author name and publication year will be imported into the database, while the rest of the information will be captured as the studies are screened. HM and EWB will separately screen titles and abstracts first, and thereafter, read the full text of all potentially qualifying studies to assess eligibility. Only studies meeting all the inclusion criteria will be included in the systematic review. HM and EWB will independently categorise articles into one of the three groups; (1) selected, (2) not selected, and (3) pending. Thereafter, the two reviewers, HM and EWB, will compare their results and resolve any disagreements by discussion. Articles categorised as 'selected' and 'not selected' by both reviewers will be included and excluded in the review respectively, while articles categorised as pending will be discussed by both reviewers in order to reach consensus. If consensus cannot be reached, discrepancies will be discussed with a third reviewer (BK). The search process and selection of studies will be summarised and presented as a flow chart in conformance with PRISMA guidelines for reviews.

2.4.5 Data extraction

Data from selected studies will be recorded into an electronic data extraction form (Appendix 2.3) developed using Microsoft Access-2010 forms, in order to enable assessment of study quality and evidence synthesis. Because the reviewers will independently extract the data, this form will be piloted on a sample of at least five randomly selected studies to assess the concordance level between the two data extractors. HM and EWB will then compare the results of the extracted data and resolve

any differences by discussion, with arbitration from (BK) for any unresolved differences. We will request missing data from study authors through email and exclude studies where the author does not respond to two email requests over a period of 4 weeks. Data elements to be extracted will include, but are not limited to the following;

- *Study characteristics*: first author, title, publication year, sample type, country, design, type (diagnostic vs predictive) case definition, and specimen used for reference standard tests.
- *Population characteristics*: age category, number of study participants, cohort type (test vs validation) proportion of adolescents, gender composition, and number by disease status (TB disease, latent MTB infection, healthy control, or other disease).
- *Transcriptomic signature characteristics:* signature name and number of genes, sample type, signature discovery method (microarray, RNA Sequencing and or PCR), model (random forest, pairwise, SVM ad so on) and, threshold score.
- Gold Standard: The reference standard tests used to diagnose TB disease or latent MTB infection will include MTB culture, smear microscopy, or Xpert MTB/RIF for the diagnosis of TB; and TST or IGRA for the diagnosis of latent MTB infection.
- Outcomes: Primary outcome measures will include sensitivity and specificity, true/false positives and true/false negatives. Secondary outcome measures will include positive likelihood ratio (LR+) and negative likelihood ratio (LR−), and area under the receiver operating characteristic curve (AUC).

2.4.6 Quality appraisal

The quality of the studies included in the systematic review will be evaluated using a customised form (Appendix 2.4) based on the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2)[33] assessment tool as well as the 'Standards for the Reporting of Diagnostic Accuracy Studies'[34] guidelines. HM and EWB will independently assess the risk of bias and compare their evaluations. If the two reviewers cannot reach consensus, a third reviewer (BK) will adjudicate. The reviewers will not be blinded to the journal titles, study authors, or institutions.

2.4.7 Data analysis and synthesis

We will provide a narrative synthesis of the findings from the included studies, focused on the performance of the transcriptomic signatures for diagnosis of prevalent TB disease (diagnostic performance) or prediction of incident TB disease (predictive performance), and the target population characteristics. Data on diagnostic performance will be analysed and presented separately from that of predictive performance. The individual index tests (transcriptomic signatures) will be compared against a reference test (MTB culture or Xpert MTB/RIF or Smear Microscopy). For each test, the

reported true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) will be retrieved. If these values are not reported, but data are available within the paper or can be obtained from the authors, we will calculate these values based on the reported sensitivity and specificity. Similarly, sensitivity and specificity will be computed where it is not reported but data allowing the calculation of such performance values is available. Separate evidence summary tables and forest plots for diagnostic and predictive studies will be reported for each signature. Signatures used for both diagnosis and prediction of TB disease will be presented in both sets of analysis. The forest plots will be produced using Review Manager (RevMan)[35], the Cochrane Collaboration's software for preparation and maintenance of reviews. The forest plot will show the data for the sensitivity and specificity of the index test and the corresponding 95% CI. Data from the test and validation sets, including for previously published signatures being evaluated in a different cohort, will be analysed and compared separately for each signature and cohort.

Each study will likely have used a different transcriptomic signature incorporating a number of different genes for the diagnosis or prediction of TB disease and consequently, we anticipate considerable clinical and methodological heterogeneity. We therefore anticipate that meta-analysis of synthesised data would not usually be appropriate.[36] However, if the same signature is evaluated in more than one study, then meta-analysis might be possible. For such analysis, a bivariate random effects model will be used to calculate pooled sensitivity and specificity with the corresponding 95% CI.[37] The Higgins (I^2) test, which quantifies the degree of inconsistency in the results of studies, will be used to assess statistical heterogeneity[38] This test explains the percentage of total variation across studies that is attributable to heterogeneity rather than chance. I^2 values less than 25% and those between 26% and 50% will be considered as low and moderate, respectively, while those between 51% and 75% and above 75% will be considered high and very high respectively. Very high inconsistency will preclude meta-analysis. For prospective studies of predictive performance of mRNA transcriptomic signatures, we will categorise performance over prospective time points reported in the studies. Rate ratios with corresponding 95% CI will also be shown for the predictive studies. If there is sufficient data available, we will conduct subgroup analyses by age category, signature discovery method, signature model, infection/exposure category (household contacts or latent MTB infection), and for predictive studies only, time to TB disease diagnosis.

2.4.8 Ethics and dissemination

Given that this is a systematic review that will use peer-reviewed, published and publicly available anonymised data, ethical approval of this protocol is not required. This review will be reported as much as possible in conformance with the PRISMA statement. The findings from this study will be

published in a peer-reviewed journal and presented at conferences. This review will also form part of a doctoral thesis at the University of Cape Town.

2.4.9 Assessing cumulative evidence

Assessing the quality of the body of evidence is recommended by PRISMA-P. We will attempt to rate the quality of our review evidence as either high, moderate, low or very low, by applying 'The Grading of Recommendations Assessment, Development and Evaluation' (GRADE) methodology.[39] Using the grade development software, GRADEpro, we will create a summary table of the evidence. HM and EWB will independently evaluate the body of evidence for each gradable outcome with regard to study design, risk of bias, directness and precision,[40, 41] and then compare the results afterwards so as to arrive at a grading decision. The quality of the evidence will be applied to test performance estimates of true positives, true negatives, false positives and false negatives using a previously published GRADE guideline.[41]

2.4.10 Patient and public involvement

Patients and or public were not involved in this systematic review protocol of published peer-reviewed articles.

2.5 Discussion

Considering the massive global burden of TB disease, low-cost, rapid, and easy-to-use tests that will accurately diagnose TB disease are urgently needed to ensure early diagnosis and treatment of patients, improve treatment outcomes, and interrupt transmission. Similarly, tests that will accurately predict which individuals with latent MTB infection will develop active TB disease are also urgently needed to ensure that treatment is targeted to those people at increased risk of incident TB disease, thereby avoiding the huge expense and unfeasibility of treating a quarter of the world's population as well as the unnecessary treatment and adverse events in people that will otherwise remain healthy. Host blood transcriptomic signatures are a candidate for such rapid biomarker-based non-sputum-based tests.

This systematic review will generate up-to-date information on the performance levels of present host blood transcriptomic signatures for diagnosis of prevalent TB and prediction of incident TB and help us compare the performance characteristics of the individual signatures to the target product profiles (TPPs) for new rapid non-sputum TB diagnostic and predictive tests proposed by WHO, Foundation for Innovative New Diagnostics and the New Diagnostics Work Group.[42, 43] Comparing the performance levels of these host blood transcriptomic signatures for diagnosis of prevalent TB and prediction of

incident TB to the TPP is important because it will allow scientists to focus their research and development efforts on the signatures that are closest to meeting the TPP and may translate into practice and have impact on the epidemic. Based on the evidence from this systematic review, we will discuss the differences and similarities of the signatures, as well as knowledge gaps identified.

2.6 References

1 Norbis L, Miotto P, Alagna R, Cirillo DM. Tuberculosis: lights and shadows in the current diagnostic landscape. *New Microbiol*. 2013;36(2):111-20.

2 O'Garra A, Redford PS, McNab FW, Bloom CI, Wilkinson RJ, Berry MP. The immune response in tuberculosis. *Annu Rev Immunol*. 2013;31:475-527.

3 Houben RM, Dodd PJ. The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. *PLoS medicine*. 2016;13(10):e1002152.

4 WHO. Global Tuberculosis Report 2018. Geneva, Switzerland: World Health Organization; 2018.

5 WHO. The End TB Strategy. Geneva, Switzerland: World Health Organization; 2015.

6 Andrews JR, Noubary F, Walensky RP, Cerda R, Losina E, Horsburgh CR. Risk of progression to active tuberculosis following reinfection with Mycobacterium tuberculosis. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2012;54(6):784-91.

7 Andrews JR, Hatherill M, Mohamed H*, et al.* The Dynamics of QuantiFERON-TB Gold In-Tube Conversion and Reversion in a Cohort of South African Adolescents. *American journal of respiratory and critical care medicine*. 2015;191(5):584-91.

8 Harries AD, Maher D, Nunn P. An approach to the problems of diagnosing and treating adult smear-negative pulmonary tuberculosis in high-HIV-prevalence settings in sub-Saharan Africa. *Bulletin of the World Health Organization*. 1998;76(6):651-62.

9 Davies PD, Pai M. The diagnosis and misdiagnosis of tuberculosis. *Int J Tuberc Lung Dis*. 2008;12(11):1226-34.

10 Steingart KR, Henry M, Ng V*, et al.* Fluorescence versus conventional sputum smear microscopy for tuberculosis: a systematic review. *The Lancet infectious diseases*. 2006;6(9):570-81.

11 Li S, Liu B, Peng M*, et al.* Diagnostic accuracy of Xpert MTB/RIF for tuberculosis detection in different regions with different endemic burden: A systematic review and meta-analysis. *PLoS One*. 2017;12(7):e0180725.

12 Maertzdorf J, Kaufmann SH, Weiner J, 3rd. Toward a unified biosignature for tuberculosis. *Cold Spring Harb Perspect Med*. 2014;5(1):a018531.

13 Ghodbane R, Raoult D, Drancourt M. Dramatic reduction of culture time of Mycobacterium tuberculosis. *Sci Rep*. 2014;4:4236.

14 Al-Zamel FA. Detection and diagnosis of Mycobacterium tuberculosis. *Expert Rev Anti Infect Ther*. 2009;7(9):1099-108.

15 WHO. Chest radiography in tuberculosis detection : Summary of current WHO recommendations and guidance on programmatic approaches Geneva, Switzerland: World Health Organization; 2016.

16 van't Hoog AH, Langendam MW, Mitchell E*, et al.* A systematic review of the sensitivity and specificity of symptom and chest-radiography screening for active pulmonary tuberculosis in HIV-negative persons and persons with unknown HIV status. Geneva, Switzerland: World Health Organization; 2013.

17 Mahomed H, Hawkridge T, Verver S*, et al.* The tuberculin skin test versus QuantiFERON TB Gold(R) in predicting tuberculosis disease in an adolescent cohort study in South Africa. *PLoS One*. 2011;6(3):e17984.

18 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.

19 Diel R, Loddenkemper R, Nienhaus A. Predictive value of interferon-gamma release assays and tuberculin skin testing for progression from latent TB infection to disease state: a meta-analysis. *Chest*. 2012;142(1):63-75.

20 Dye C, Glaziou P, Floyd K, Raviglione M. Prospects for tuberculosis elimination. *Annu Rev Public Health*. 2013;34:271-86.

21 Cliff JM, Kaufmann SH, McShane H, van Helden P, O'Garra A. The human immune response to tuberculosis and its treatment: a view from the blood. *Immunol Rev*. 2015;264(1):88-102.

22 Burel JG, Peters B. Discovering transcriptional signatures of disease for diagnosis versus mechanism. *Nat Rev Immunol*. 2018;18(5):289-90.

23 Blankley S, Berry MPR, Graham CM, Bloom CI, Lipman M, O'Garra A. The application of transcriptional blood signatures to enhance our understanding of the host response to infection: the example of tuberculosis. *Philos T R Soc B*. 2014;369(1645).

24 Berry MPR, Graham CM, McNab FW*, et al.* An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature*. 2010;466(7309):973-U98.

25 Sweeney TE, Braviak L, Tato CM, Khatri P. Genome-wide expression for diagnosis of pulmonary tuberculosis: a multicohort analysis. *Lancet Respir Med*. 2016;4(3):213-24.

26 Walzl G, McNerney R, du Plessis N*, et al.* Tuberculosis: advances and challenges in development of new diagnostics and biomarkers. *The Lancet infectious diseases*. 2018.

27 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.

28 Suliman S, Thompson E, Sutherland J*, et al.* Four-gene Pan-African Blood Signature Predicts Progression to Tuberculosis. *American journal of respiratory and critical care medicine*. 2018.

29 Moher D, Shamseer L, Clarke M*, et al.* Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev*. 2015;4:1.

30 Virgili G, Conti AA, Murro V, Gensini GF, Gusinu R. Systematic reviews of diagnostic test accuracy and the Cochrane collaboration. *Intern Emerg Med*. 2009;4(3):255-8.

31 Cochrane_Collaboration. Handbook for DTA Reviews: Cochrane; 2017 [Cochrane DTA Working Group]. Available from: http://methods.cochrane.org/sdt/handbook-dta-reviews.

32 Moher D, Altman DG, Liberati A, Tetzlaff J. PRISMA Statement. *Epidemiology*. 2011;22(1):128-31.

33 Whiting PF, Rutjes AWS, Westwood ME*, et al.* QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. *Ann Intern Med*. 2011;155(8):529-U104.

34 Bossuyt PM, Cohen JF, Gatsonis CA, Korevaar DA, Grp S. STARD 2015: updated reporting guidelines for all diagnostic accuracy studies. *Ann Transl Med*. 2016;4(4).

35 Review Manager (RevMan) [Computer program]. Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014.

36 Ioannidis JP, Patsopoulos NA, Rothstein HR. Reasons or excuses for avoiding meta-analysis in forest plots. *BMJ*. 2008;336(7658):1413-5.

37 Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol*. 2005;58(10):982-90.

38 Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327(7414):557-60.

39 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.

40 Singh S, Chang SM, Matchar DB, Bass EB. Grading a Body of Evidence on Diagnostic Tests. In: Chang SM, Matchar DB, Smetana GW, Umscheid CA, editors. Methods Guide for Medical Test Reviews. AHRQ Methods for Effective Health Care. Rockville (MD)2012.

41 Schunemann 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.

42 Seshadri P, Denkinger C. Target Product Profile: Test for Incipient Tuberculosis. Geneva, Switzerland: Foundation for Innovative New Diagnostics; 2016. Contract No.: Vs8.2; 20 October 2016.

43 WHO. High-priority target product profiles for new tuberculosis diagnostics: report of a consensus meeting. Geneva, Switzerland: World Health Organization; 2014. Contract No.: 28–29 April 2014.

Contributors MH and TS conceived the study. HM wrote the protocol under supervision from MH and TS. HM, EWB, SKM, SCM, BK, AP-N, TS and MH, reviewed, revised and approved the final version of protocol and will be involved in analysis and interpretation of the results. *MH and TS contributed equally.

Funding This research received no specific grant from any funding agency in the public, commercial or not for profit sectors.

Competing interests TS and AP-N are inventors of blood transcriptomic signatures of risk of TB.

Data sharing statement The authors declare that this research protocol is original work and findings from completed study, using this protocol will be published in a peer-reviewed journal.

Acknowledgements The authors acknowledge the contribution of Mrs. Mary Shelton, the Health Sciences' reference librarian, at the University of Cape Town, for her assistance in developing the search strategies.

Chapter 3

3 Performance of diagnostic and predictive host blood transcriptomic signatures for tuberculosis disease: a systematic review and meta-analysis

Chapter overview

This chapter provides a summary of the available evidence at the time, on the performance of mRNA signatures for the diagnosis of and prediction of progression to TB disease through a systematic review and meta-analysis. The protocol in the previous chapter acts as the methods section for this systematic review. The systematic review and meta-analysis are presented as published, verbatim.

Mulenga H, Zauchenberger CZ, Bunyasi EW Mendelsohn SC, Kagina B, Penn-Nicholson A, Scriba T, and Hatherill M. *Performance of diagnostic and predictive host blood transcriptomic signatures for tuberculosis disease: a systematic review and meta-analysis*. PLoS One, 2020. **15**(8): p. e0237574. (https://doi.org/10.1371/journal.pone.0237574)

Chapter contribution to the thesis

The systematic review and meta-analysis provide the literature review component of the thesis and addresses Aim 1. This was the first systematic review of host blood transcriptomic signatures to include a meta-analysis of signature performance as reported by the original studies, in addition to reporting on biomarkers for progression to TB disease.

Contributions of the candidate

The candidate planned the study, performed the literature search, study selection and assessment, data extraction, data management, data analysis, and wrote the manuscript with editorial input and guidance from his supervisors (TS and MH). CZZ contributed to study selection, data extraction and analysis, and all contributed to research group discussions and reviewed the final draft manuscript (HM, CZZ, EWB, SKM, SCM, BK, APN, TS, and MH)

Publication

Performance of diagnostic and predictive host blood transcriptomic signatures for tuberculosis disease: a systematic review and meta-analysis.

Humphrey Mulenga[1], Chambrez-Zita Zauchenberger[1], Erick W. Bunyasi[1], Stanley Kimbung Mbandi[1], Simon C. Mendelsohn[1], Benjamin Kagina[2], Adam Penn-Nicholson[1], Thomas J. Scriba[1¶], Mark Hatherill[1¶*]

Author affiliations:

[1]South African Tuberculosis Vaccine Initiative (SATVI), Institute of Infectious Disease & Molecular Medicine and Division of Immunology, Department of Pathology, University of Cape Town, Cape Town, South Africa

[2]Vaccines for Africa Initiative (VACFA), School of Public Health & Family Medicine, University of Cape Town, Cape Town, South Africa.

***Correspondence:** Professor Mark Hatherill, South African Tuberculosis Vaccine Initiative (SATVI), Institute of Infectious Disease & Molecular Medicine and Division of Immunology, Department of Pathology, S2.11 Werner Beit South Building, Faculty of Health Sciences, University of Cape Town, Anzio Road, Observatory, 7925, South Africa.

Tel. +27 21 406 6791

Email. Mark.Hatherill@uct.ac.za

[¶]These authors contributed equally

3.1 Abstract

Introduction

Host blood transcriptomic biomarkers have potential as rapid point-of-care triage, diagnostic, and predictive tests for Tuberculosis disease. We aimed to summarise the performance of host blood transcriptomic signatures for diagnosis of and prediction of progression to Tuberculosis disease; and compare their performance to the recommended World Health Organisation target product profile.

Methods

A systematic review and meta-analysis of the performance of host blood mRNA signatures for diagnosing and predicting progression to Tuberculosis disease in HIV-negative adults and adolescents, in studies with an independent validation cohort. Medline, Scopus, Web of Science, and EBSCO libraries were searched for articles published between January 2005 and May 2019, complemented by a search of bibliographies. Study selection, data extraction and quality assessment were done independently by two reviewers. Meta-analysis was performed for signatures that were validated in ≥3 comparable cohorts, using a bivariate random effects model.

Results

Twenty studies evaluating 25 signatures for diagnosis of or prediction of progression to TB disease in a total of 68 cohorts were included. Eighteen studies evaluated 24 signatures for TB diagnosis and 17 signatures met at least one TPP minimum performance criterion. Three diagnostic signatures were validated in clinically relevant cohorts to differentiate TB from other diseases, with pooled sensitivity 84%, 87% and 90% and pooled specificity 79%, 88% and 74%, respectively. Four studies evaluated signatures for progression to TB disease and performance of one signature, assessed within six months of TB diagnosis, met the minimal TPP for a predictive test for progression to TB disease.

Conclusion

Host blood mRNA signatures hold promise as triage tests for TB. Further optimisation is needed if mRNA signatures are to be used as standalone diagnostic or predictive tests for therapeutic decision-making.

3.2 Introduction

The World Health Organisation (WHO) has targeted 2035 to end tuberculosis (TB) and aims for 90% reduction in new TB cases and 95% reduction in TB deaths compared to 2015 levels.[1] Non-sputum triage, diagnostic, and predictive tests for TB may play a role in advancing TB control efforts. The WHO, in conjunction with the Foundation for Innovative New Diagnostics and the New Diagnostics Working Group of the Stop TB Partnership, has published Target Product Profiles (TPPs) for non-sputum biomarker triage, diagnostic, and predictive tests for progression from latent TB infection (LTBI) to TB disease.[2-4] The TPPs require minimum 90% sensitivity and 70% specificity for a triage test; 65% sensitivity and 98% specificity for a diagnostic test [2], and 75% sensitivity and 75% specificity for a test to predict progression from LTBI to active TB disease within two years.[3, 4] A new predictive test should also achieve a positive predictive value (PPV) of 5.8% given a 2% pre-test probability.[3]

Current commercially available TB diagnostic tests are not optimal. *Mycobacterium tuberculosis* (MTB) culture, considered the gold standard, requires days to weeks to obtain a result from a reference laboratory, and is thus not ideal for rapid patient management.[5, 6] Sputum smear microscopy has low sensitivity[7] ranging from 32% to 89% [8], resulting in a considerable proportion of active pulmonary TB patients being missed.[9] Sputum Xpert MTB/RIF and Xpert MTB/RIF Ultra have considerably better diagnostic performance[10, 11] than smear microscopy, but are similarly dependant on obtaining an adequate sputum sample; and need specialised laboratory equipment and a reliable power supply, which impedes routine screening in TB-endemic resource-limited settings.[12]

Individuals with LTBI, defined by a positive tuberculin skin test (TST) or interferon-gamma release assay (IGRA), have a higher risk of progression to TB disease than MTB-uninfected people.[13, 14] However, only about 10-15% of people who test IGRA or TST positive will go on to develop TB disease.[15, 16] Predictive specificity of IGRA and TST for incident TB disease is poor (49.3% and 45% respectively).[14, 17] While prevention of TB disease arising from LTBI is key to achieving WHO elimination targets[18], mass preventive therapy based on IGRA or TST screening in TB-endemic countries would need to treat a significant proportion of the population[19], most of them unnecessarily, which would be unaffordable and potentially ineffective, because re-infection would likely occur before programmatic coverage was complete.

In recent years, host blood transcriptomic signatures have offered a promising alternative as tests for both diagnosis of and prediction of progression to TB disease. These signatures have also improved our understanding of inflammatory processes associated with progression[20, 21] in individuals with MTB

infection[22] and those with TB disease.[23] Host blood transcriptomic signatures have been shown to discriminate prevalent TB disease cases from MTB-uninfected and latently MTB-infected individuals, and individuals with other respiratory ailments [24]; and predict progression to TB disease in individuals with LTBI.[25, 26]

Two systematic reviews have evaluated biomarkers for diagnosis of TB disease in children[27] and all age groups.[28] These systematic reviews did not include biomarkers for progression to TB disease; and did not include a meta-analysis, owing to heterogenous study designs, patient selection, and biomarker composition. Two additional studies performed re-analysis of patient-level data but did not perform a systematic review of signature performance as reported by the original studies. Warsinske *et al*[29] compared 16 signatures for TB diagnosis by recreating the original model of each signature and evaluated each signature across the datasets they had identified, and Gupta *et al*[30] conducted a meta-analysis of patient-level pooled data for signatures of incipient TB. We present a systematic review and meta-analysis of transcriptomic signatures which have been evaluated in independent validation cohorts. We aimed to summarise the performance of host blood transcriptomic signatures for diagnosis of and prediction of progression to TB disease in HIV-negative adults and adolescents; and to compare individual signature performance to the WHO TPP. This review is registered with the International Prospective Register of Systematic Reviews (PROSPERO), registration number CRD42017073817.

3.3 Materials and methods

We conducted a systematic review according to standard guidelines[31, 32] (Appendix 3.1). and designed the protocol prior to conducting the review (http://dx.doi.org/10.1136/bmjopen-2018-026612).[33]

3.3.1 Study inclusion/exclusion criteria

We included studies evaluating host blood mRNA signatures for diagnosis of or prediction of progression to TB disease in HIV-negative adults and/or adolescents (≥ 12 years old) and published in English. Studies were restricted to those published between January 2005 and May 2019 to concentrate on recent evidence because of the fast pace at which the field of transcriptomics is changing and advancing. Only studies comparing TB disease cases versus MTB-uninfected controls, individuals with other diseases (ODs), or with LTBI; and using a microbiological reference standard of either sputum MTB culture, Xpert MTB/RIF, or smear microscopy for TB disease diagnosis, were eligible for inclusion. Studies of prediction of incident TB disease were required to be prospective, with a follow-up period of at least six months; and enrolling either TB contacts, latently MTB-infected, or uninfected individuals. Studies were excluded if: conducted in animals; were in children younger than

12 years, did not report sensitivity and specificity, did not allow recreation of a 2 x 2 contingency table for calculation of test performance, and where authors did not respond to enquiries for data within four weeks of inquiry. Unpublished reports and conference proceedings were excluded due to absence of peer review and difficulty in obtaining data.

3.3.2 Literature search

Medline via PubMed, Scopus, Web of Science, and EBSCO databases were searched for relevant studies. The search strategy developed on PubMed was adapted to other databases and was as follows:

((((((Tuberculosis [MeSH] OR Mycobacterium tuberculosis [MeSH] OR (Tuberculosis OR TB OR Mycobacterium tuberculosis OR MTB))) AND ((Diagnosis [MeSH] OR Diagnosis [subheading] OR Prognosis [MeSH] OR (Diagnosis OR diagnostic OR detect* OR predict* OR prognosis OR prognostic OR screen*)))) AND ((Biomarkers/Blood [MeSH] OR RNA/Blood [MeSH] OR Transcription, Genetic [MeSH] /etiology/genetics/immunology OR (Blood Biomarker OR blood biomarkers OR bio-signature OR gene expression OR genetic transcription OR host blood OR immune marker OR immunologic marker OR Ribonucleic Acid OR RNA OR signature OR surrogate endpoint OR surrogate marker OR transcriptome OR transcriptomic)))) AND ((Area under Curve [MeSH] OR Sensitivity and Specificity [MeSH] OR (Area under curve OR area under curves OR AUC OR receiver operating characteristic OR ROC OR Accuracy OR Performance OR sensitivity OR specificity)))) AND (Humans[Mesh])) AND (("2005/01/01"[Date–Publication] : "2019/05/31"[Date–Publication])). Additionally, bibliographies of included papers were scrutinised for potential papers that were missed by the search terms.

3.3.3 Study selection

Two reviewers (HM and CZZ) independently screened search outputs for eligible studies. Publications were first screened by title and abstract, and thereafter by full text. Articles were independently categorised as either (i) selected, (ii) not selected, or (iii) pending. The two reviewers conferred to resolve any disagreements about pending publications, and if a consensus could not be reached, discrepancies were adjudicated by a third reviewer (BK). Figure 3.1 depicts the study selection process.

3.3.4 Data extraction and management

Data metrics were extracted separately from the relevant articles and double-entered into a Microsoft SQL Server 2012 database to facilitate electronic comparison of data between reviewers. For each biomarker test, we extracted the reported performance data; sensitivity, specificity, true positives (TP) false positives (FP), true negatives (TN), and false negatives (FN); as well as details of the study design,

population, and index and reference test characteristics using a customised data extraction form (Appendix 2.3). We requested missing data from authors by email.

3.3.5 Appraisal of methodological quality of studies

HM and CZZ independently assessed the quality of included studies using a customised form based on the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2)[34] assessment tool (Appendix 2.4). BMK provided adjudication where a disagreement occurred. For each of the four domains in QUADAS-2, namely "patient selection", "index test", "reference standard" and "flow and timing"; risk of bias was scored as "low" if all responses in that domain were answered as "yes", "high" if any of the responses were answered as "no" or "unclear" and "unclear" if it was unclear for all the responses. We judged "low applicability concerns" for the patient selection domain if a clinically relevant control population was used; OD for diagnostic studies and LTBI or household TB contacts for predictive studies and "high applicability concerns" if other populations were used. For other domains, we judged "low applicability concerns" if all signalling questions in that domain were answered as "yes", "high applicability concerns" if any question was answered as "no" and "unclear applicability concerns" if answered as such.

3.3.6 GRADE quality of evidence

We used the "Grading of Recommendations Assessment, Development and Evaluation" (GRADE) approach to judge the quality of evidence. Classification of the quality of evidence was based on study design in conjunction with the five factors that affect study quality; study limitations, indirectness, inconsistency, imprecision, and publication bias.[35]

3.3.7 Data synthesis and analysis

We extracted data to construct 2 x 2 contingency tables of reference test versus index test results. TB positive and TB negative were defined as participants with and without TB disease respectively, based on the reference standard. Forest plots of sensitivity and specificity with 95% confidence intervals for each signature were created using RevMan 5.3.[36] Each entry in the forest plot represents a signature that was evaluated in a distinct cohort. Several signatures were tested in multiple cohorts and reported in multiple studies. In naming individual studies, we used the first author name, year of manuscript publication, and a sequential letter representing a specific cohort. Similarly, the signature naming convention was first author name, number of genes, and year of publication.

We reported the index test results as TP, FP, TN and FN. If not explicitly reported, TP, FP, TN, and FN were estimated from the reported sensitivity and specificity and total number of TB positive individuals and controls. Similarly, sensitivity and specificity were reported, or calculated if the data were available or obtained from the authors. We also calculated the PPV and negative predictive values (NPV) at 2% pre-test probability for predictive signatures.

Signatures used for both diagnosis and prediction of TB disease are presented in separate forest plots. Each evaluation of a signature in a different population, or of different signatures in the same population, is shown as a separate entry (or entries) in the forest plot. If the same signature was reported using different models, the best performing model was included in the analysis. Similarly, if a study reported several signatures with the same number of genes in the same population, only the best performing signature was chosen.

Considerable clinical and methodological heterogeneity was anticipated due to reporting of multiple signatures in different populations. Therefore, we did not perform meta-analysis on all signatures in all studies. In order to address this heterogeneity and compare relative performance of signatures, we performed meta-analysis only for signatures that were evaluated in at least three comparable cohorts with the same control population. Meta-analysis was conducted in STATA 11, using hierarchical logistic regression. We used the bivariate random effects model to calculate summary sensitivity and specificity with the corresponding 95% CI[37], and to create summary receiver operating characteristic curves for each signature. Heterogeneity in the diagnostic or predictive performance of the signatures was assessed by visual inspection of the forest plots and the I^2 statistic.

3.4 Results

3.4.1 Search results

Our search term returned 2,313 reports of which 27[23-26, 38-62] satisfied all the pre-specified inclusion criteria (Fig 3.1). Twenty-three of the 27 studies reported exclusively on diagnostic performance of mRNA signatures for discriminating TB disease cases from controls with or without ODs, and controls with or without LTBI. Two studies reported exclusively on predictive performance for progression to TB disease;[26, 55] and two studies reported both diagnostic and predictive performance.[25, 60] A total of 35 transcriptomic signatures incorporating 1,027 genes were identified (Appendix 3.2). Forty-two of the 1,027 genes were employed in at least three or more transcriptomic signatures and Fc gamma receptor 1A (FCGR1A) was the most frequently utilised gene (Appendix 3.3)

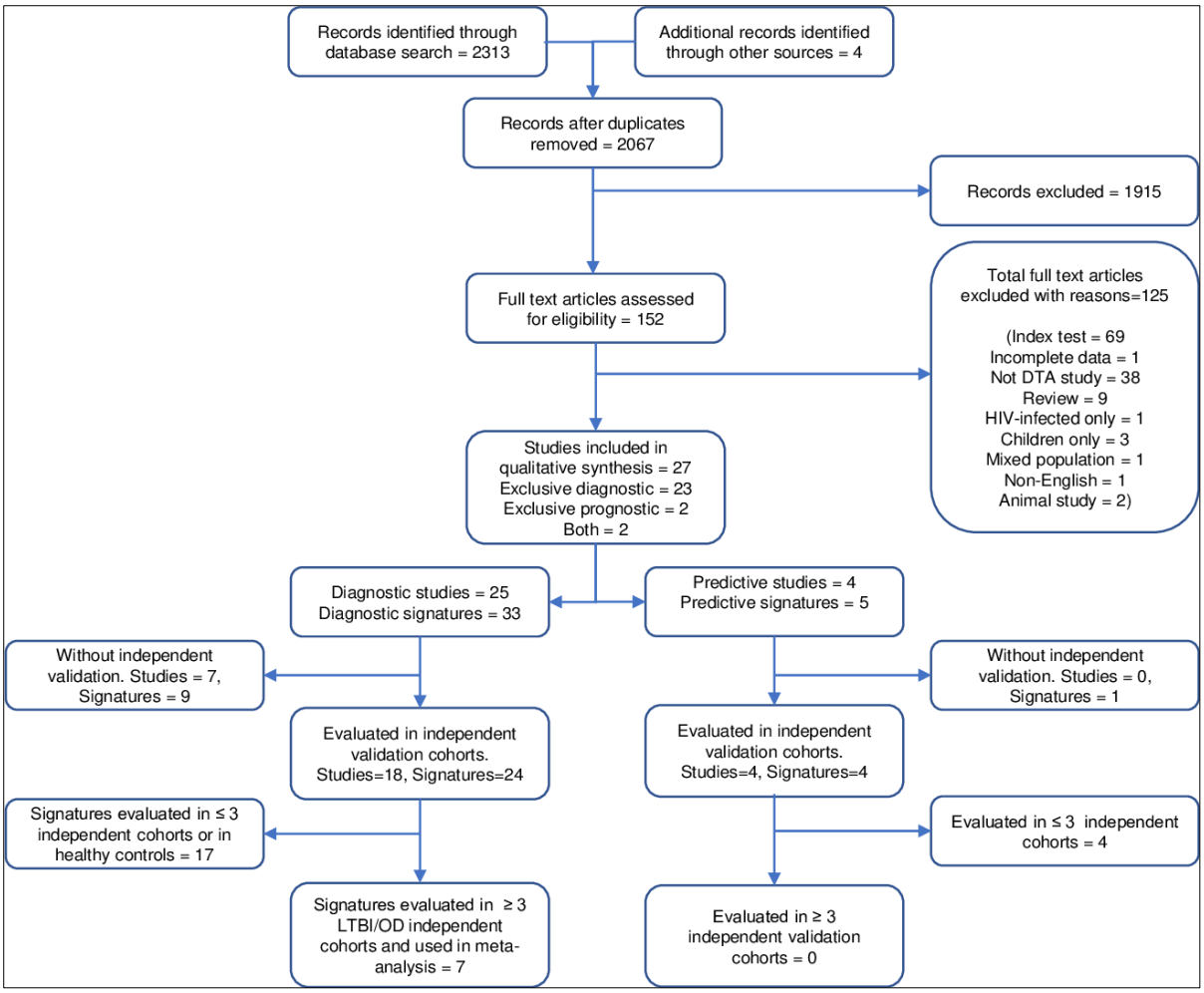


Figure 3.1: Flow of studies in the review of transcriptomic signatures for diagnosing and predicting progression to TB disease.

DTA; diagnostic test accuracy, LTBI; Latent TB infection, OD; Other diseases

3.4.2 Quality of diagnostic studies

Four studies[42, 44, 58, 60] employed a cohort design, four[23, 38, 51, 52] a cross sectional design, and the remainder employed a case-control design. All studies had a reference standard of either smear microscopy [50], Xpert/MTB RIF[41] or MTB culture; (Appendix 3.4). One study[46] did not specify the type of reference test used. Eight [23-25, 38, 44, 45, 49, 56, 59, 60] of the 18 studies evaluated signatures in populations with ODs. Only one study[43] made reference to blinding of the index test readers. Bias in patient selection arose in most studies due to case-control design and non-reporting of sampling method, resulting in a non-representative spectrum of patients. Bias in the index test resulted from the lack of reported blinding in the index test interpretation. Bias in the reference standard was minimal since 97% (68/70) of the entries used MTB culture as a reference standard which can be considered objective (Fig 3.2; Appendices 3.5 and 3.6).

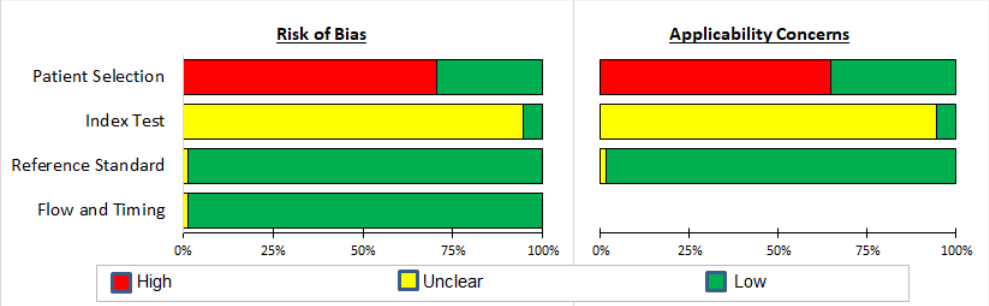


Figure 3.2: Summary of results of QUADAS-2 assessment of diagnostic studies in clinically relevant independent validation cohorts with other diseases.

3.4.3 Quality of studies for prediction of progression to TB disease

All studies[25, 26, 55, 60] validated signatures in participants from prospective independent cohorts, although two studies[25, 26] were case-control studies nested in prospective cohorts (Appendix 3.7). Participants from three studies did not all receive the same reference standard[26, 55, 60] and blinding was only stated in two studies[25, 26] (Fig 3.3; Appendix 3.8).

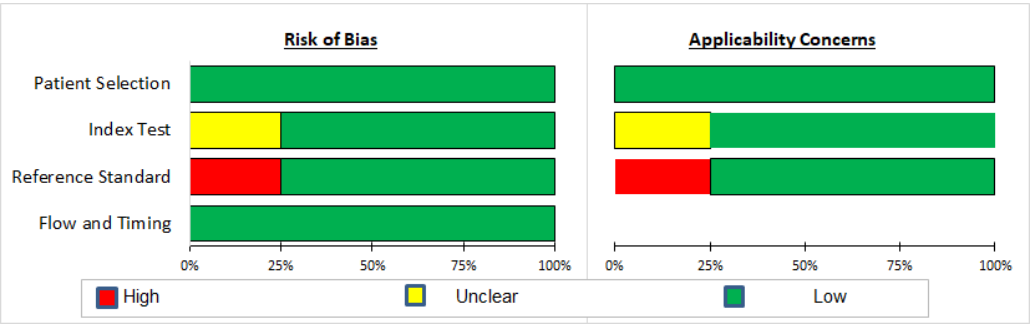


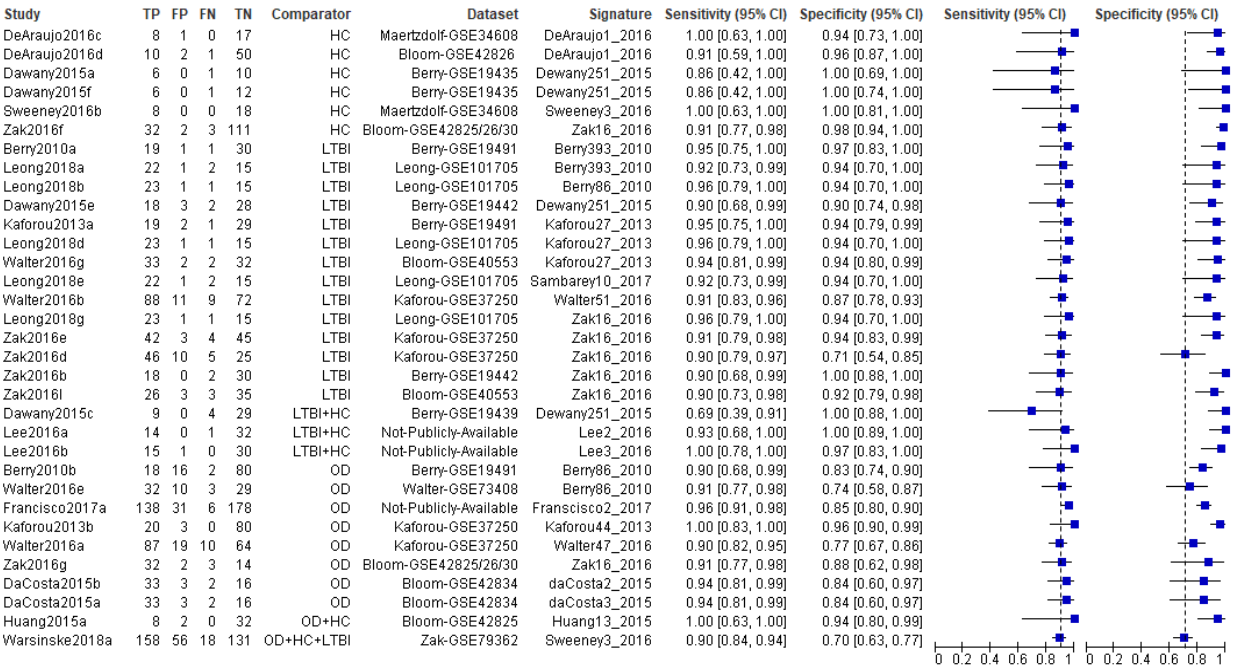
Figure 3.3: Summary of results of QUADAS-2 assessment in studies of prediction to TB disease in independent validation cohorts.

3.4.4 Performance of mRNA signatures for TB Diagnosis

Eighteen studies evaluated 24 transcriptomic signatures for diagnosis of TB disease in 70 different independent validation cohorts (Appendix 3.9). Individual signatures displayed substantial variation in diagnostic performance with sensitivity ranging from 50%-100% and specificity ranging from 32%-100%. The observed heterogeneity (I^2 = 0.99) in study design and signature performance precluded pooling of diagnostic accuracy estimates for all signatures. Nine signatures were not evaluated in independent validation cohorts and hence excluded (Appendix 3.10).

Thirty-three entries (46.5%) representing 17 different signatures in 12 studies met at least one TPP minimum performance criterion in independent validation sets containing uninfected controls, LTBI,

ODs, or a combination of these populations (Fig 3.4). Signature performance ranged between 69%-100% for sensitivity, and between 70%-100% for specificity.



Study	TP	FP	FN	TN	Comparator	Dataset	Signature	Sensitivity (95% CI)	Specificity (95% CI)
DeAraujo2016c	8	1	0	17	HC	Maertzdolf-GSE34608	DeAraujo1_2016	1.00 [0.63, 1.00]	0.94 [0.73, 1.00]
DeAraujo2016d	10	2	1	50	HC	Bloom-GSE42826	DeAraujo1_2016	0.91 [0.59, 1.00]	0.96 [0.87, 1.00]
Dawany2015a	6	0	1	10	HC	Berry-GSE19435	Dewany251_2015	0.86 [0.42, 1.00]	1.00 [0.69, 1.00]
Dawany2015f	6	0	1	12	HC	Berry-GSE19435	Dewany251_2015	0.86 [0.42, 1.00]	1.00 [0.74, 1.00]
Sweeney2016b	8	0	0	18	HC	Maertzdolf-GSE34608	Sweeney3_2016	1.00 [0.63, 1.00]	1.00 [0.81, 1.00]
Zak2016f	32	2	3	111	HC	Bloom-GSE42825/26/30	Zak16_2016	0.91 [0.77, 0.98]	0.98 [0.94, 1.00]
Berry2010a	19	1	1	30	LTBI	Berry-GSE19491	Berry393_2010	0.95 [0.75, 1.00]	0.97 [0.83, 1.00]
Leong2018a	22	1	2	15	LTBI	Leong-GSE101705	Berry393_2010	0.92 [0.73, 0.99]	0.94 [0.70, 1.00]
Leong2018b	23	1	1	15	LTBI	Leong-GSE101705	Berry86_2010	0.96 [0.79, 1.00]	0.94 [0.70, 1.00]
Dawany2015e	18	3	2	28	LTBI	Berry-GSE19442	Dewany251_2015	0.90 [0.68, 0.99]	0.90 [0.74, 0.98]
Kaforou2013a	19	2	1	29	LTBI	Berry-GSE19491	Kaforou27_2013	0.95 [0.75, 1.00]	0.94 [0.79, 0.99]
Leong2018d	23	1	1	15	LTBI	Leong-GSE101705	Kaforou27_2013	0.96 [0.79, 1.00]	0.94 [0.70, 1.00]
Walter2016g	33	2	2	32	LTBI	Bloom-GSE40553	Kaforou27_2013	0.94 [0.81, 0.99]	0.94 [0.80, 0.99]
Leong2018e	22	1	2	15	LTBI	Leong-GSE101705	Sambarey10_2017	0.92 [0.73, 0.99]	0.94 [0.70, 1.00]
Walter2016b	88	11	9	72	LTBI	Kaforou-GSE37250	Walter51_2016	0.91 [0.83, 0.96]	0.87 [0.78, 0.93]
Leong2018g	23	1	1	15	LTBI	Leong-GSE101705	Zak16_2016	0.96 [0.79, 1.00]	0.94 [0.70, 1.00]
Zak2016e	42	3	4	45	LTBI	Kaforou-GSE37250	Zak16_2016	0.91 [0.79, 0.98]	0.94 [0.83, 0.99]
Zak2016d	46	10	5	25	LTBI	Kaforou-GSE37250	Zak16_2016	0.90 [0.79, 0.97]	0.71 [0.54, 0.85]
Zak2016b	18	0	2	30	LTBI	Berry-GSE19442	Zak16_2016	0.90 [0.68, 0.99]	1.00 [0.88, 1.00]
Zak2016l	26	3	3	35	LTBI	Bloom-GSE40553	Zak16_2016	0.90 [0.73, 0.98]	0.92 [0.79, 0.98]
Dawany2015c	9	0	4	29	LTBI+HC	Berry-GSE19439	Dewany251_2015	0.69 [0.39, 0.91]	1.00 [0.88, 1.00]
Lee2016a	14	0	1	32	LTBI+HC	Not-Publicly-Available	Lee2_2016	0.93 [0.68, 1.00]	1.00 [0.89, 1.00]
Lee2016b	15	1	0	30	LTBI+HC	Not-Publicly-Available	Lee3_2016	1.00 [0.78, 1.00]	0.97 [0.83, 1.00]
Berry2010b	18	16	2	80	OD	Berry-GSE19491	Berry86_2010	0.90 [0.68, 0.99]	0.83 [0.74, 0.90]
Walter2016e	32	10	3	29	OD	Walter-GSE73408	Berry86_2010	0.91 [0.77, 0.98]	0.74 [0.58, 0.87]
Francisco2017a	138	31	6	178	OD	Not-Publicly-Available	Franscisco2_2017	0.96 [0.91, 0.98]	0.85 [0.80, 0.90]
Kaforou2013b	20	3	0	80	OD	Kaforou-GSE37250	Kaforou44_2013	1.00 [0.83, 1.00]	0.96 [0.90, 0.99]
Walter2016a	87	19	10	64	OD	Kaforou-GSE37250	Walter47_2016	0.90 [0.82, 0.95]	0.77 [0.67, 0.86]
Zak2016g	32	2	3	14	OD	Bloom-GSE42825/26/30	Zak16_2016	0.91 [0.77, 0.98]	0.88 [0.62, 0.98]
DaCosta2015b	33	3	2	16	OD	Bloom-GSE42834	daCosta2_2015	0.94 [0.81, 0.99]	0.84 [0.60, 0.97]
DaCosta2015a	33	3	2	16	OD	Bloom-GSE42834	daCosta3_2015	0.94 [0.81, 0.99]	0.84 [0.60, 0.97]
Huang2015a	8	2	0	32	OD+HC	Bloom-GSE42825	Huang13_2015	1.00 [0.63, 1.00]	0.94 [0.80, 0.99]
Warsinske2018a	158	56	18	131	OD+HC+LTBI	Zak-GSE79362	Sweeney3_2016	0.90 [0.84, 0.94]	0.70 [0.63, 0.77]

Figure 3.4: Forest plots of sensitivity and specificity of transcriptomic signatures for diagnosis of TB disease that met at least one minimum TPP performance criterion in independent validation cohorts.

HC; Healthy controls, LTBI; Latent TB infection, OD; Other diseases. Vertical dashed lines correspond to 90% sensitivity and 70% specificity.

Ten studies evaluated 12 signatures for diagnosis of TB disease in clinically relevant populations with ODs (Fig 3.5). Signature performance ranged between 50%–100% for sensitivity, and between 47%–96% for specificity. Seven of these signatures; Berry86_2010, daCosta2_2015, daCosta3_2015, Francisco2_2017, Kaforou44_2015, Walter47_2016, and Zak16_2016 met the WHO-recommended minimal TPP for a triage test. None of these signatures met the WHO minimal TPP for a diagnostic test in the OD population.

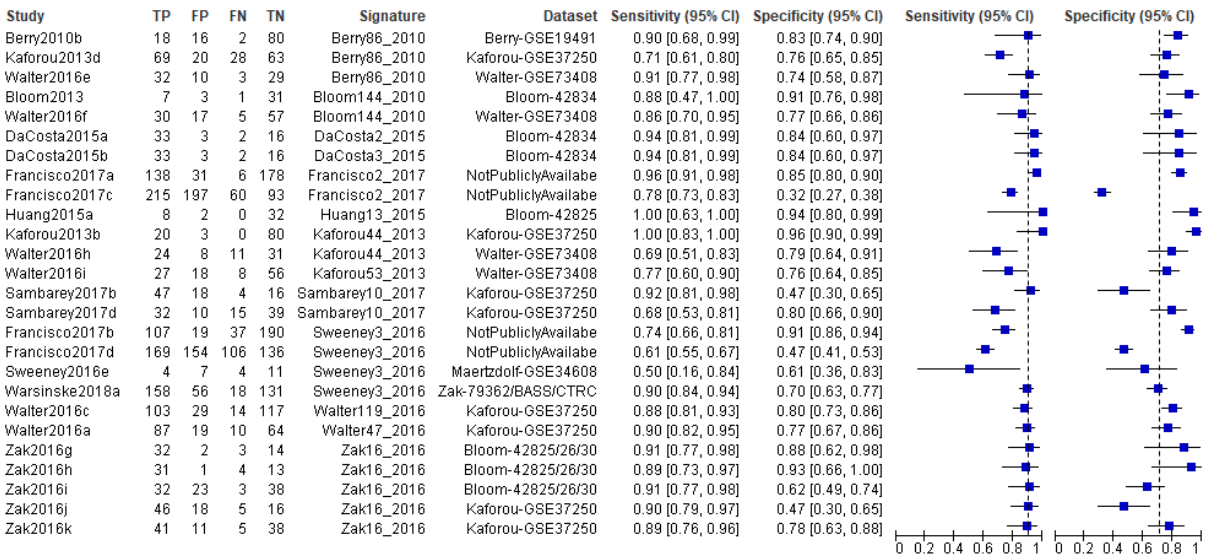


Figure 3.5: Forest plots of sensitivity and specificity of transcriptomic signatures for diagnosis of TB disease in independent validation cohorts of clinically relevant populations with other diseases.

Vertical dashed lines correspond to 90% sensitivity and 70% specificity.

3.4.5 Meta-analysis of performance of mRNA signatures for diagnosis of TB disease

Ten studies included seven signatures that were validated for diagnosis of TB in at least three comparable cohorts including either LTBI or OD populations. Signature diagnostic performance results are shown in Table 3.1. We excluded signatures validated exclusively in populations with uninfected controls, but included signatures evaluated in LTBI or OD populations that also included uninfected controls. Four signatures were validated in exclusively LTBI populations and six signatures were validated exclusively in OD populations; and two signatures in both LTBI and OD populations. Among the six signatures validated in LTBI cohorts, three signatures with similar diagnostic accuracy (Berry393_2010, Kaforou27_2013, Zak16_2016) showed pooled sensitivity and specificity that met the minimal WHO TPP for a non-sputum biomarker triage test in this population. Summary receiver operating characteristic curves for the signatures in OD populations are shown as Figures a, b and c in Appendix 3.11. Heterogeneity in the signature performance was not explained by comparison group (I^2=0.98 for all OD, I^2=0.95 for all LTBI, I^2=0.90 for all healthy controls,). Zak16_2016 which used the TPP bench-mark of 90% sensitivity had a lower I^2 of 0.56 compared to Sweeny3_2016's I^2 of 0.93. Some of the heterogeneity was explained by whether the evaluation used the TPP benchmarks and by composition of the control population.

Table 3.1: Meta-analysis of performance of transcriptomic signatures for diagnosis of TB disease in independent validation cohorts that included LTBI or other disease populations

Signature	Study Entry ID	TB/ Control	Control Group	Sensitivity % (95% CI)	Specificity % (95% CI)
Berry393_2010	Berry2010a	20/31	LTBI	90 (83, 94)	92 (82,96)
	Kaforou2013c	97/83	LTBI		
	Leong2018a	24/16	LTBI		
	Walter2016d	35/35	LTBI		
Dawany251_2015	Dawany2015b	29/38	LTBI	82 (70, 90)	95 (88, 98)
	Dawany2015c	13/29	LTBI*		
	Dawany2015d	21/33	LTBI*		
	Dawany2015e	20/31	LTBI		
Kaforou27_2013	Kaforou2013a	20/31	LTBI	95 (87, 98)	93 (85, 97)
	Leong2018d	24/16	LTBI		
	Walter2016g	35/35	LTBI		
Samabarey10_2017	Leong2018e	24/16	LTBI	83 (75, 88)	92 (85, 96)
	Sambarey2017a	51/36	LTBI		
	Sambarey2017c	47/47	LTBI		
Sweeney3_2016	Leong2018f	24/16	LTBI	89 (84, 92)	72 (66, 78)
	Warsinske2018a	176/187	LTBI+		
	Sweeney2016d	46/25	LTBI		
Zak16_2016	Leong2018g	24/16	LTBI	91 (86, 94)	90 (72, 97)
	Zak2016a	21/21	LTBI		
	Zak2016b	20/30	LTBI		
	Zak2016d	51/35	LTBI		
	Zak2016e	46/48	LTBI		
	Zak2016l	29/38	LTBI		
Berry86_2010	Berry2010b	20/96	OD	84 (68, 93)	79 (73, 84)
	Kaforou2013d	97/83	OD		
	Walter2016e	35/39	OD		
Sweeney3_2016	Francisco2017b	275/290	OD	74 (57, 86)	71 (49, 86)
	Francisco2017d	144/209	OD		
	Sweeney2016e	8/18	OD		
	Warsinske2018a	176/187	OD+		
Zak16_2016	Zak2016g	35/16	OD	90 (85, 93)	74 (56, 86)
	Zak2016h	35/14	OD		
	Zak2016i	35/61	OD		
	Zak2016j	51/34	OD		
	Zak2016k	46/49	OD		

LTBI; Latent TB infection, OD; Other diseases, HC; Health control.

*Includes some HCs, +Consists of HCs, ODs and LTBIs.

3.4.6 Performance of mRNA signatures for prediction of progression to TB disease

Four studies evaluated five signatures for prediction of progression to TB disease in independent validation cohorts with LTBI and uninfected controls. Two of these studies evaluated LTBI populations only; one study evaluated TB contacts only; and one study evaluated both TB contacts and LTBI populations (Fig 3.6). The time window between signature measurement and TB diagnosis reported

in these studies was not consistent and ranged from 6 months to 24 months before TB diagnosis. Since differential gene expression becomes more pronounced as individuals approach TB diagnosis[21, 25, 26], this variable precluded direct comparison of signature performance for prediction of progression to TB disease. In LTBI populations, signature performance ranged between 76%–86% for sensitivity; and between 55%–84% for specificity. Signatures evaluated in TB contacts showed between 53%–67% sensitivity and 83%–99% specificity. The Sweeney3_2016 signature met the TPP performance criterion (PPV ≥ 5.8% and 75% sensitivity and 75% specificity) for a test to predict progression when measured within 6 months of TB diagnosis. Performance of the other signatures was not reported for this 6-month predictive interval (time to TB disease) and none of the other signatures met the TPP performance criterion over longer predictive intervals. Regardless, two other signatures, assessed within 12 months of TB diagnosis, also achieved a PPV ≥ 5.8% (Appendix 3.12).

Figure 3.6: Forest plots of sensitivity and specificity of transcriptomic signatures for prediction of progression to TB disease in independent validation cohorts.

Vertical dashed lines correspond to 75% sensitivity and 75% specificity. Prediction time is the time to TB disease used in each study

3.4.7 GRADE evidence profile

We followed a previously published GRADE guideline[35] to assess the quality of the body of evidence and produced a GRADE evidence profile (Fig 3.7, Appendix 3.13). The overall quality of evidence supporting the estimates of sensitivity and specificity of mRNA signatures for the diagnosis of TB disease was rated as "very low". Consequently, very low confidence is placed in the estimates obtained from pooling studies in meta-analysis. Similarly, the quality of evidence for studies of progression to TB disease was also very low.

Outcome	Study design	# Studies (Sample size)	Risk of bias (Study limitations)	Inconsistence	Indirectness	Imprecision	Publication bias	Final quality	Effect per 100,000	Importance[1]
True Positives	Cohort, Case-Control, Cross-Section	70 (7,765)	Very serious[2] (-2)	Very serious[3] (-2)	No serious indirectness[4]	Serious[5] (-1)	Likely[6]	Very low ⨁◯◯◯	Prevalence[7] 2% : 1,780	Critical
True Negative	Cohort, Case-Control, Cross-Section	70 (7,765)	Very serious[2] (-2)	Very serious[3] (-2)	No serious indirectness[4]	Serious[5] (-1)	Likely[6]	Very low ⨁◯◯◯	Prevalence[7] 2% : 88,200	Critical
False Positives	Cohort, Case-Control, Cross-Section	70 (7,765)	Very serious[2] (-2)	Very serious[3] (-2)	No serious indirectness[4]	Serious[5] (-1)	Likely[6]	Very low ⨁◯◯◯	Prevalence[7] 2% : 9,800	Critical
False Negatives	Cohort, Case-Control, Cross-Section	70 (7,765)	Very serious[2] (-2)	Very serious[3] (-2)	No serious indirectness[4]	Serious[5] (-1)	Likely[6]	Very low ⨁◯◯◯	Prevalence[7] 2% : 220	Critical

1. The Importance of outcomes was classified as either "Critial", "Important" or "Not important" according to their relative importance

2. Majority of studies were case-controls and lacked both a representative patient spectrum and blinding

3. Substantial heterogeneity was observed in the study results

4. No downgrade was applied for indirectness though diagnostic accuracy is considered a surrogate for paitent-important outcomes.

5. Diagnostic accuracy estimates were not pooled. There was a considerable number of studies with wide 95% CI. We down-graded by one point only, as there were a large number of studies and we had already down-graded for inconsistency

6. We did not down-grade for publication bias. The data in this systematic review did not allow for formal assessment of publication bias with methods such as funnel plots or regression analysis. Publication bias can not be ruled out as studies may not have been published in which mRNA signatures showed poor diagnostic accuracy for TB

7. What is the meaning of these results among people being screened for TB disease, given a 2% prevalance of disease

Explanation: Based on a sample size of 7,765, median sensitity=89.5% and median specificity=90%, we rated the quality of the evidence as high when no points were subtracted, moderate when only one point was subtracted, low when two points were subtracted and very low when more than two points were subtracted. Deduction of points was based on the five factors that decrease study quality; study limitations, incosistence in results accross studies, indirectness in evidence, imprecision in summary estimates, and possibility of reporting bias. For each outcome, evidence from cohort and cross-sectional studies started as high quality while that from case-control studies started as moderate quality. We deducted two, one and zero points for "very serious", "serious", and "no serious" issues identified respectively, and the deducted points are shown in brackets. Reporting bias was classfied as either "very likely", "likely" and "not likely".

Figure 3.7: GRADE evidence profile: transcriptomic signatures for the diagnosis of TB in clinically relevant populations with other diseases only.

3.5 Discussion

The New Diagnostics Working Group Strategic Framework 2016–2020 aims to "achieve early and universal diagnosis of all patients with all forms of TB to foster progress towards TB elimination, by making appropriate and affordable diagnostic solutions available at the right setting".[63] Comparison of signature performance in similar populations and under similar conditions is critical to down-select candidate biomarkers for development as rapid point-of-care (POC) tests. Future implementation decisions hinge on test performance in clinically relevant populations under field conditions, in addition to feasibility and cost considerations that contribute to the assessment of impact and public health value.

We show that of the 17 mRNA signatures that met at least one minimal TPP performance criterion, 11 were validated in populations including either uninfected controls, LTBI, or both, and four were validated exclusively in populations with ODs. Four signatures validated in the uninfected control category (Dewany251_2015, Sweeney3_2016, Zak16_2016, and Lee2_2016) have potential as diagnostic tests while the remainder show potential as triage tests. The signatures in the LTBI/uninfected control category that are yet to be validated in populations with ODs such as DeAraujo1_2016, Lee2_2016, Lee3_2016, and Sambarey10_2017 should also be validated in such populations to confirm robustness of diagnostic accuracy and allow comparison with the signatures above.

Although TB triage tests might be used in community mass screening campaigns or contact investigations that include uninfected individuals or asymptomatic individuals with LTBI; the greatest need is for TB diagnostic tests that discriminate TB from ODs among symptomatic individuals seeking health care. We also recognise that evaluation of novel biomarkers in cohorts of carefully selected TB cases and uninfected controls tends to over-estimate performance. In this regard, diagnostic evaluation of signatures in populations that exclusively included LTBI or uninfected controls would be considered inferior to validation of performance in clinically relevant cohorts that included individuals with other respiratory diseases. The OD category in several of these studies did not include symptomatic individuals presenting to healthcare facilities with suspected TB, but rather individuals with systemic lupus erythematosus, sarcoidosis, or other infrequently encountered conditions. None of the signatures validated in cohorts with ODs met the minimum WHO TPP for sensitivity and specificity of a diagnostic test. However, seven signatures; Berry86_2010, daCosta2_2015, daCosta3_2015, Francisco2_2017, Kaforou44_2015, Walter47_2016, and Zak16_2016 met the minimum WHO TPP for a triage test. The findings suggest that these seven signatures could be further evaluated as TB triage tests under field conditions; and that feasibility and unit cost-effectiveness as potential rapid POC tests would be a consideration for further clinical development. Other signatures that approached but did not meet the TPP may also warrant further validation in side-by-side comparison studies. Signatures with smaller number of genes which may be more adaptable to a POC device could be given preferential consideration for clinical validation and development. Our findings also suggest that further improvement in performance of existing signatures, or even further discovery of new signatures with improved performance, would be necessary for clinical development of a non-sputum TB diagnostic test. For instance, incorporating covariates such as age and sex in the models.

Validation of diagnostic tests in several geographically different populations is important to confirm robustness. Relatively few signatures were validated in at least three comparable cohorts of a similar population and were eligible for meta-analysis. Only one of the signatures (Zak16_2016), validated in multiple cohorts that included ODs such as pneumonia, lung cancer, sarcoidosis, or systemic lupus erythematosus, met the minimum WHO TPP performance criteria for a triage test in meta-analysis and none met the minimal TPP for a diagnostic test in the meta-analyses. The finding suggests that this signature has potential as a rapid POC test and should be considered for clinical development upon validation under field conditions. Similarly, other signatures approaching the minimum TPP target in meta-analysis should be considered for field validation.

Tests that will accurately predict which individuals with LTBI will develop TB disease are needed to ensure that preventive treatment can be targeted for those individuals at increased risk of incident TB disease, while saving those individuals at lowest risk from the cost, burden, and side-effects of unnecessary intervention. Current tests for MTB infection, including IGRA and TSTs, are poor predictors because of their low specificity for incident TB disease. It is not cost-effective to treat the estimated two billion individuals latently infected with MTB worldwide and therefore preventive therapy targeted with a more specific biomarker may be a more feasible alternative.[64] Predictive signatures might be used in community-level TB mass-screening campaigns or contact investigations, but might also be useful in symptomatic individuals who have been investigated and found not to have active TB disease at the time of testing. Only one of the four signatures (Sweeney3_2016) met the minimum TPP criteria for both PPV ≥ 5.8% and 75% sensitivity and 75% specificity for a 6-month period prior to TB diagnosis, although this was the only signature for which performance within six-months of TB diagnosis was reported. The other studies reported signature performance one or two years prior to TB diagnosis. It is thus not clear how these signatures would perform during the six-month period before TB diagnosis. These results highlight that more studies of predictive performance are necessary. It is also evident that a two-year predictive horizon may be overly optimistic for prediction of progression to TB disease as progression is very variable in occurrence.

One of the challenges of this systematic review and meta-analysis was that most studies did not specify whether the signature being tested was intended for triage, diagnostic, or predictive use, neither did they benchmark performance of signatures against the WHO TPPs. For example, if the goal is to discover and validate a TB triage test, a study should report specificity at 90% sensitivity or higher, to allow comparison with other novel biomarkers against this standard. Similarly, if the goal is to discover and validate a TB diagnostic test, the study should ideally report sensitivity at 98% specificity or higher; and studies aiming to discover and validate a test to predict progression to TB disease should report sensitivity at 75% specificity or higher, or PPV and NPV along with the prediction time horizon. However, it must be noted that the differences in the benchmarks are partially because some of the studies were completed before publication of the TPPs. Secondly, we found that many studies were sub-optimally designed and used control populations without clinical relevance. As observed in a previous systematic review[28], we found that several of transcriptomic signatures for TB diagnosis were discovered but not validated in independent representative cohorts. We also found that a number of diagnostic accuracy studies did not conform to the reporting guidelines for diagnostic test accuracy (DTA) studies stipulated in the "Standards for the Reporting of Diagnostic Accuracy Studies" (STARD).[65] In several studies, cardinal data on study design, patient selection, numbers of participants in each

group, and diagnostic performance data such as sensitivity and specificity with their corresponding confidence intervals (CIs) that would enable reproduction of the study were not reported. This is a major drawback in synthesising the body of evidence on DTA studies and thus compliance to STARD in designing DTA studies and reporting their findings cannot be over-emphasised.

3.5.1 Strengths and limitations of the study

We used an inclusive time frame of January 2005 to May 2019 to include the period in which we believe all transcriptomic TB biomarker studies were published. We also developed a protocol prior to performing the systematic review that explicitly stated a rigorous search strategy and clear inclusion/exclusion criteria. Unlike previous systematic reviews, our review includes evaluation of signatures for predicting progression to TB disease and a meta-analysis.

Some signatures were designed to optimise sensitivity while others were designed to optimise specificity. This may have introduced bias in the pooled estimates of sensitivity and specificity in the meta-analysis, and difficult to compare signature performance. Restricting included studies to those conducted in HIV-negative adults and adolescents may have excluded signatures with superior diagnostic performance in studies conducted in children or in HIV-positive individuals. Additionally, language selection bias cannot be ruled out since we only included studies reported in English. We did not formally assess publication bias as current methods are not suitable for DTA studies.[66]

Heterogeneity of study design and reliance on reported data makes it impossible to fairly compare signature performance. A major finding of this study and limitation is the very low quality of evidence: preponderance of case control studies, spectrum bias and narrow geography. This highlights the need for high quality, prospective studies, with relevant populations of symptomatic clinic attendees, mass screening endemic community population or high-risk populations such as household TB contacts which minimise spectrum bias, and from multiple geographies.

3.6 Conclusion

Host blood mRNA signatures show considerable promise as triage tests for TB. Signatures designed for TB diagnosis meeting at least one TPP minimum performance criterion in independent validation sets containing healthy controls or LTBI populations should be further optimised in populations with ODs. Similarly signatures for TB diagnosis validated in populations with ODs and signatures for prediction of progression from LTBI to TB disease meeting the minimum TPP should be further optimised and validated under field conditions to confirm their accuracy for use as standalone diagnostic or predictive tests for therapeutic decision-making. There is also need for signature discovery in large “real-world” clinically appropriate populations, without spectrum bias, need for head-to-head comparison of signatures and adaptation and implementation towards a POC test.

Acknowledgements

The authors acknowledge the contribution of Mrs. Mary Shelton, the Health Sciences' reference librarian, at the University of Cape Town, for her assistance in developing the search strategies.

3.7 References

1 WHO. The End TB Strategy. Geneva, Switzerland: World Health Organization; 2015.

2 WHO. Meeting Report : High-priority target product profiles for new tuberculosis diagnostics: report of a consensus meeting. Geneva, Switzerland: World Health Organization; 2014. Contract No.: WHO/HTM/TB/2014.18.

3 Seshadri P, Denkinger C. Target Product Profile: Test for Incipient Tuberculosis. Geneva, Switzerland: Foundation for Innovative New Diagnostics; 2016. Contract No.: Vs8.2; 20 October 2016.

4 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. Report. Geneva: World Health Organization; 2017. Contract No.: WHO/HTM/TB/2017.18.

5 Ghodbane R, Raoult D, Drancourt M. Dramatic reduction of culture time of Mycobacterium tuberculosis. *Scientific reports*. 2014;4:4236.

6 Norbis L, Miotto P, Alagna R, Cirillo DM. Tuberculosis: lights and shadows in the current diagnostic landscape. *New Microbiol*. 2013;36(2):111-20.

7 Davies PD, Pai M. The diagnosis and misdiagnosis of tuberculosis. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease*. 2008;12(11):1226-34.

8 Steingart KR, Henry M, Ng V*, et al.* Fluorescence versus conventional sputum smear microscopy for tuberculosis: a systematic review. *The Lancet Infectious diseases*. 2006;6(9):570-81.

9 Harries AD, Maher D, Nunn P. An approach to the problems of diagnosing and treating adult smear-negative pulmonary tuberculosis in high-HIV-prevalence settings in sub-Saharan Africa. *Bull World Health Organ*. 1998;76(6):651-62.

10 Li S, Liu B, Peng M*, et al.* Diagnostic accuracy of Xpert MTB/RIF for tuberculosis detection in different regions with different endemic burden: A systematic review and meta-analysis. *PloS one*. 2017;12(7):e0180725.

11 Dorman SE, Schumacher SG, Alland D*, et al.* Xpert MTB/RIF Ultra for detection of Mycobacterium tuberculosis and rifampicin resistance: a prospective multicentre diagnostic accuracy study. *The Lancet Infectious diseases*. 2018;18(1):76-84.

12 Maertzdorf J, Kaufmann SH, Weiner J, 3rd. Toward a unified biosignature for tuberculosis. *Cold Spring Harbor perspectives in medicine*. 2014;5(1):a018531.

13 Andrews JR, Hatherill M, Mohamed H*, et al.* The Dynamics of QuantiFERON-TB Gold In-Tube Conversion and Reversion in a Cohort of South African Adolescents. *American journal of respiratory and critical care medicine*. 2015;191(5):584-91.

14 Mahomed H, Hawkridge T, Verver S*, et al.* The tuberculin skin test versus QuantiFERON TB Gold(R) in predicting tuberculosis disease in an adolescent cohort study in South Africa. *PloS one*. 2011;6(3):e17984.

15 Andrews JR, Noubary F, Walensky RP, Cerda R, Losina E, Horsburgh CR. Risk of progression to active tuberculosis following reinfection with Mycobacterium tuberculosis. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2012;54(6):784-91.

16 Barry CE, 3rd, Boshoff HI, Dartois V*, et al.* The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. *Nature reviews Microbiology*. 2009;7(12):845-55.

17 Diel R, Loddenkemper R, Nienhaus A. Predictive value of interferon-gamma release assays and tuberculin skin testing for progression from latent TB infection to disease state: a meta-analysis. *Chest*. 2012;142(1):63-75.

18 Dye C, Glaziou P, Floyd K, Raviglione M. Prospects for tuberculosis elimination. *Annu Rev Public Health*. 2013;34:271-86.

19 Sumner T, Scriba TJ, Penn-Nicholson A, Hatherill M, White RG. Potential population level impact on tuberculosis incidence of using an mRNA expression signature correlate-of-risk test to target tuberculosis preventive therapy. *Scientific reports*. 2019;9.

20 Burel JG, Peters B. Discovering transcriptional signatures of disease for diagnosis versus mechanism. *Nature reviews Immunology*. 2018;18(5):289-90.

21 Scriba TJ, Penn-Nicholson A, Shankar S*, et al.* Sequential inflammatory processes define human progression from M. tuberculosis infection to tuberculosis disease. *PLoS pathogens*. 2017;13(11):e1006687.

22 Blankley S, Berry MPR, Graham CM, Bloom CI, Lipman M, O'Garra A. The application of transcriptional blood signatures to enhance our understanding of the host response to infection: the example of tuberculosis. *Philos T R Soc B*. 2014;369(1645).

23 Berry MPR, Graham CM, McNab FW*, et al.* An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature*. 2010;466(7309):973-U98.

24 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(10):e1001538.

25 Zak DE, Penn-Nicholson A, Scriba TJ*, et al.* A blood RNA signature for tuberculosis disease risk: a prospective cohort study. *Lancet (London, England)*. 2016;387(10035):2312-22.

26 Suliman S, Thompson EG, Sutherland J*, et al.* Four-Gene Pan-African Blood Signature Predicts Progression to Tuberculosis. *American journal of respiratory and critical care medicine*. 2018;197(9):1198-208.

27 Togun TO, MacLean E, Kampmann B, Pai M. Biomarkers for diagnosis of childhood tuberculosis: A systematic review. *PloS one*. 2018;13(9):1-19.

28 MacLean E, Broger T, Yerlikaya S, Fernandez-Carballo BL, Pai M, Denkinger CM. A systematic review of biomarkers to detect active tuberculosis. *Nat Microbiol*. 2019;4(5):748-58.

29 Warsinske H, Vashisht R, Khatri P. Host-response-based gene signatures for tuberculosis diagnosis: A systematic comparison of 16 signatures. *PLoS medicine*. 2019;16(4):e1002786.

30 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(4):395-406.

31 Cochrane_Collaboration. Handbook for DTA Reviews: Cochrane; 2017 [Cochrane DTA Working Group]. Available from: http://methods.cochrane.org/sdt/handbook-dta-reviews.

32 Moher D, Shamseer L, Clarke M*, et al.* Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev*. 2015;4:1.

33 Mulenga H, Bunyasi EW, Mbandi SK*, et al.* Performance of host blood transcriptomic signatures for diagnosing and predicting progression to tuberculosis disease in HIV-negative adults and adolescents: a systematic review protocol. *BMJ Open*. 2019;9(5):e026612.

34 Whiting PF, Rutjes AWS, Westwood ME*, et al.* QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. *Annals of internal medicine*. 2011;155(8):529-U104.

35 Horvath AR. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *Clinical chemistry*. 2009;55(5):853-5.

36 Review Manager (RevMan) [Computer program]. Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014.

37 Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol*. 2005;58(10):982-90.

38 Bloom CI, Graham CM, Berry MP*, et al.* Transcriptional blood signatures distinguish pulmonary tuberculosis, pulmonary sarcoidosis, pneumonias and lung cancers. *PloS one*. 2013;8(8):e70630.

39 Satproedprai N, Wichukchinda N, Suphankong S*, et al.* Diagnostic value of blood gene expression signatures in active tuberculosis in Thais: a pilot study. *Genes and immunity*. 2015;16(4):253-60.

40 Cai Y, Yang Q, Tang Y*, et al.* Increased complement C1q level marks active disease in human tuberculosis. *PloS one*. 2014;9(3):e92340.

41 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*. 2018;108:124-6.

42 Dawany N, Showe LC, Kossenkov AV*, et al.* Identification of a 251 gene expression signature that can accurately detect M. tuberculosis in patients with and without HIV co-infection. *PloS one*. 2014;9(2):e89925.

43 de Araujo LS, Vaas LA, Ribeiro-Alves M*, et al.* Transcriptomic Biomarkers for Tuberculosis: Evaluation of DOCK9. EPHA4, and NPC2 mRNA Expression in Peripheral Blood. *Frontiers in microbiology*. 2016;7:1586.

44 Francisco NM, Fang YM, Ding L*, et al.* Diagnostic accuracy of a selected signature gene set that discriminates active pulmonary tuberculosis and other pulmonary diseases. *The Journal of infection*. 2017;75(6):499-510.

45 Hai-Hui H, Xiao-Ying L, Yong L, Hua C, Liang-Yong X. Identification of 13 blood-based gene expression signatures to accurately distinguish tuberculosis from other pulmonary diseases and healthy controls. *Bio-Medical Materials & Engineering*. 2015;26:S1837-S43.

46 Huang ZK, Yao FY, Xu JQ*, et al.* Microarray Expression Profile of Circular RNAs in Peripheral Blood Mononuclear Cells from Active Tuberculosis Patients. *Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology*. 2018;45(3):1230-40.

47 Jacobsen M, Repsilber D, Gutschmidt A*, et al.* Candidate biomarkers for discrimination between infection and disease caused by Mycobacterium tuberculosis. *Journal of Molecular Medicine-Jmm*. 2007;85(6):613-21.

48 Joosten SA, Goeman JJ, Sutherland JS*, et al.* Identification of biomarkers for tuberculosis disease using a novel dual-color RT-MLPA assay. *Genes and immunity*. 2012;13(1):71-82.

49 Laux da Costa L, Delcroix M, Dalla Costa ER*, et al.* A real-time PCR signature to discriminate between tuberculosis and other pulmonary diseases. *Tuberculosis (Edinburgh, Scotland)*. 2015;95(4):421-5.

50 Lee SW, Wu LS, Huang GM, Huang KY, Lee TY, Weng JT. Gene expression profiling identifies candidate biomarkers for active and latent tuberculosis. *BMC bioinformatics*. 2016;17 Suppl 1:3.

51 Leong S, Zhao Y, Joseph NM*, et al.* Existing blood transcriptional classifiers accurately discriminate active tuberculosis from latent infection in individuals from south India. *Tuberculosis*. 2018;109:41-51.

52 Lu C, Wu J, Wang H*, et al.* Novel biomarkers distinguishing active tuberculosis from latent infection identified by gene expression profile of peripheral blood mononuclear cells. *PloS one*. 2011;6(8):e24290.

53 Maertzdorf J, Repsilber D, Parida SK*, et al.* Human gene expression profiles of susceptibility and resistance in tuberculosis. *Genes and immunity*. 2011;12(1):15-22.

54 Pan L, Wei N, Jia H*, et al.* Genome-wide transcriptional profiling identifies potential signatures in discriminating active tuberculosis from latent infection. *Oncotarget*. 2017;8(68):112907-16.

55 Roe J, Venturini C, Gupta RK*, et al.* Blood transcriptomic stratification of short-term risk in contacts of tuberculosis. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2019.

56 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.

57 Serrano CJ, Cuevas-Cordoba B, Macias-Segura N*, et al.* Transcriptional profiles discriminate patients with pulmonary tuberculosis from non-tuberculous individuals depending on the presence of non-insulin diabetes mellitus. *Clinical immunology (Orlando, Fla)*. 2016;162:107-17.

58 Sweeney TE, Braviak L, Tato CM, Khatri P. Genome-wide expression for diagnosis of pulmonary tuberculosis: a multicohort analysis. *The Lancet Respiratory medicine*. 2016;4(3):213-24.

59 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. *Journal of clinical microbiology*. 2016;54(2):274-82.

60 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.

61 Wu B, Huang C, Kato-Maeda M*, et al.* Messenger RNA expression of IL-8, FOXP3, and IL-12beta differentiates latent tuberculosis infection from disease. *Journal of immunology (Baltimore, Md : 1950)*. 2007;178(6):3688-94.

62 Lesho E, Forestiero FJ, Hirata MH*, et al.* Transcriptional responses of host peripheral blood cells to tuberculosis infection. *Tuberculosis (Edinburgh, Scotland)*. 2011;91(5):390-9.

63 New Diagnostics Working Group (NDWG). The Roadmap to new TB diagnostics: Stop TB Partnership; [Available from: http://www.stoptb.org/wg/new_diagnostics/objectives.asp.

64 Sumner T, Bozzani F, Mudzengi D*, et al.* Estimating the Impact of Tuberculosis Case Detection in Constrained Health Systems: An Example of Case-Finding in South Africa. *Am J Epidemiol*. 2019;188(6):1155-64.

65 Bossuyt PM, Cohen JF, Gatsonis CA, Korevaar DA, Grp S. STARD 2015: updated reporting guidelines for all diagnostic accuracy studies. *Ann Transl Med*. 2016;4(4).

66 Macaskill P, Gatsonis C, Deeks J, Harbord R, Takwoingi Y. Chapter 10 Analysing and Presenting Results-Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.: The Cochrane Collaboration; 2010 [Available from: http://srdta.cochrane.org/.

Chapter 4

4 The effect of host factors on discriminatory performance of a transcriptomic signature of tuberculosis risk

Chapter overview

Non-TB pathogens, such as HIV and respiratory viruses, are thought to affect the readout of signatures like RISK11 that include interferon-stimulated genes. Only one previous study compared differential diagnostic accuracy among predefined population subgroups but did not evaluate the effect of several factors on the diagnostic performance of the signatures. Thus, a critical deficiency in transcriptomic biomarker development is the lack of understanding as to whether, and to what extent, host factors negatively affect discriminatory accuracy and contribute to a transcriptomic biomarker 'performance ceiling.

This chapter analyses the effect of host factors on diagnostic and prognostic performance of RISK11 for TB disease. Specifically, the chapter identifies and quantifies the effect of tuberculosis-independent and tuberculosis-dependent host factors on RISK11 score in tuberculosis cases and healthy controls; and quantifies the effect of adjustment for host factors on diagnostic performance of RISK11 for both prevalent and incident tuberculosis. Further, the chapter evaluates the value of combining RISK11 with baseline risk factors for the diagnosis or prognosis of tuberculosis disease.

Mulenga H, Fiore-Gartland A, Mendelsohn SC, Penn-Nicholson A, Mbandi SK, Borate B, Musvosvi M, Tameris M, Walzl G, Naidoo K, Churchyard G, Scriba TJ, and Hatherill M. *The effect of host factors on discriminatory performance of a transcriptomic signature of tuberculosis risk*. EBioMedicine. 2022;77:103886. (https://doi.org/10.1016/j.ebiom.2022.103886)

Chapter contribution to the thesis

This chapter addresses Aim 2 of the thesis. Prior to this analysis, few data were available on the effect of host factors on diagnostic and prognostic performance of RISK11; and similarly other signatures that contain interferon signaling genes.

Contributions of the candidate

The candidate provided data management and operational support for the study, analysed, and interpreted the data, and wrote the manuscript with editorial input and guidance from his supervisors (TJS and MH). Co-authors were responsible for recruitment of participants, clinical management, and clinical data collection at study sites (MT, GW, KN, and GC); and operational or laboratory support and project management (HM, SKM, AP-N, MM). AF-G provided statistical support. All co-authors reviewed the final draft of the manuscript (HM, AF-G, SCM, AP-N, SKM, BB, MM, MT, GW, KN, GC, TJS, and MH).

Publication

The effect of host factors on discriminatory performance of a transcriptomic signature of tuberculosis risk

Humphrey Mulenga,[a] Andrew Fiore-Gartland,[b] Simon C. Mendelsohn,[a] Adam Penn-Nicholson,[a] Stanley Kimbung Mbandi,[a] Bhavesh Borate,[b] Munyaradzi Musvosvi,[a] Michèle Tameris,[a] Gerhard Walzl,[c] Kogieleum Naidoo,[d,e] Gavin Churchyard,[f,g,h] Thomas J. Scriba,[a] Mark Hatherill,[a*] and the CORTIS Study Team

[a] South African Tuberculosis Vaccine Initiative, Institute of Infectious Disease and Molecular Medicine, Division of Immunology, Department of Pathology, University of Cape Town, Anzio Road, Observatory, 7925, South Africa.

[b] Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Fairview Ave. N., Seattle, WA 98109-1024, USA.

[c] DST/NRF Centre of Excellence for Biomedical TB Research and SAMRC Centre for TB Research, Division of Molecular Biology and Human Genetics, Department of Biomedical Sciences, Faculty of Medicine and Health Sciences, Stellenbosch University, Francie Van Zijl Dr, Parow, 7505, South Africa.

[d] Centre for the AIDS Programme of Research in South Africa (CAPRISA), Durban, Doris Duke Medical Research Institute, University of KwaZulu-Natal, 719 Umbilo Road, Durban 4001, South Africa.

[e] MRC-CAPRISA HIV-TB Pathogenesis and Treatment Research Unit, Doris Duke Medical Research Institute, University of KwaZulu-Natal, 719 Umbilo Road, Durban 4001, South Africa.

[f] The Aurum Institute, 29 Queens Rd, Parktown, Johannesburg, 2194, Gauteng, South Africa.

[g] School of Public Health, University of Witwatersrand, 27 St Andrews Road, Parktown 2193, Johannesburg, South Africa.

[h] Department of Medicine, Vanderbilt University, Nashville, TN, USA

***Corresponding Author:** Professor Mark Hatherill, South African Tuberculosis Vaccine Initiative, Institute of Infectious Disease and Molecular Medicine, Division of Immunology, Department of Pathology, University of Cape Town, Anzio Road, Observatory, 7925, South Africa.
Email mark.hatherill@uct.ac.za; Telephone +27214066791.

4.1 Abstract

Background

We aimed to understand host factors that affect discriminatory performance of a transcriptomic signature of tuberculosis risk (RISK11).

Methods

HIV-negative adults aged 18–60 years were evaluated in a prospective study of RISK11 and surveilled for tuberculosis through 15 months. Generalised linear models and receiver-operating characteristic (ROC) regression were used to estimate effect of host factors on RISK11 score (%marginal effect) and on discriminatory performance for tuberculosis disease (area under the curve, AUC), respectively.

Findings

Among 2923 participants including 74 prevalent and 56 incident tuberculosis cases, percentage marginal effects on RISK11 score were increased among those with prevalent tuberculosis (+18.90%, 95%CI 12.66–25.13), night sweats (+14.65%, 95%CI 5.39–23.91), incident tuberculosis (+7.29%, 95%CI 1.46–13.11), flu-like symptoms (+5.13%, 95%CI 1.58–8.68), and smoking history (+2.41%, 95%CI 0.89–3.93) than those without; and reduced in males (-6.68%, 95%CI -8.31–-5.04) and with every unit increase in BMI (-0.13%, 95%CI -0.25–-0.01). Adjustment for host factors affecting controls did not change RISK11 discriminatory performance. Cough was associated with 72.55% higher RISK11 score in prevalent tuberculosis cases. Stratification by cough improved diagnostic performance from AUC=0.74 (95%CI 0.67–0.82) overall, to 0.97 (95%CI 0.90–1.00, p<0.001) in cough-positive participants. Combining host factors with RISK11 improved prognostic performance, compared to RISK11 alone, (AUC=0.76, 95%CI 0.69–0.83 versus 0.56, 95%CI 0.46–0.68, p<0.001) over a 15-month predictive horizon.

Interpretation

Several host factors affected RISK11 score, but only adjustment for cough affected diagnostic performance. Combining host factors with RISK11 should be considered to improve prognostic performance.

Funding

Bill and Melinda Gates Foundation, South African Medical Research Council.

Keywords

Mycobacterium tuberculosis, transcriptomic, signature, RNA, host factors, performance.

4.2 Research in context

Evidence before this study

We conducted a systematic review, by searching Medline, Scopus, Web of Science, and EBSCO with comprehensive search terms for “tuberculosis”, “diagnosis”, “prognosis”, “transcriptional”, “blood”, “signatures” and “performance”, for studies conducted in English and published between January 2005 and May 2019. Several host blood transcriptomic signatures with promising diagnostic and prognostic performance for tuberculosis disease have been developed. Non-TB pathogens, such as HIV and respiratory viruses, are thought to affect the readout of signatures like RISK11 that include interferon-stimulated genes. Only one previous study compared differential diagnostic accuracy among predefined population subgroups but did not evaluate the effect of several factors on the diagnostic performance of the signatures. Thus, a critical deficiency in transcriptomic biomarker development is the lack of understanding as to whether, and to what extent, host factors negatively affect discriminatory accuracy and contribute to a transcriptomic biomarker ‘performance ceiling’. It is not known whether tuberculosis signatures such as RISK11 should be adjusted for host factors that limit performance, or whether host factors that improve performance should be included in a combination clinical-transcriptomic signature to improve accuracy.

Added value of this study

To our knowledge, this is the first report that comprehensively evaluates the effect of host factors on discriminatory performance of transcriptomic signatures in HIV-negative individuals. We present evidence from a large longitudinal study in a tuberculosis-endemic setting that several tuberculosis-independent host factors affected the RISK11 signature readout in people without tuberculosis, but adjustment for these host factors did not alter discriminatory performance. By contrast, a tuberculosis-dependent host factor, cough, was associated with a 72.55% marginal increase in RISK11 score in prevalent tuberculosis cases and stratification by cough status showed significantly improved signature performance in people with cough. A combination signature including both host factors and RISK11, compared to RISK11 alone, did improve prognostic accuracy for incident tuberculosis that occurred at least 12 months after testing, a predictive horizon beyond which transcriptomic signature performance is known to deteriorate. However, the combination signature did not improve diagnostic

or short-term prognostic accuracy of RISK11 within six months of tuberculosis disease, a time-frame within which transcriptomic signature performance is thought to be optimal.

Implications of all the available evidence

Although the host blood transcriptomic signature RISK11 is affected by tuberculosis-independent host factors, implementation as a triage or prognostic test for tuberculosis would not require adjustment for these host characteristics. By contrast, the tuberculosis-dependent host factor cough has a major impact on signature performance. Discriminatory performance in people with cough is excellent (AUC=0.97); poor performance (AUC=0.72) in people without cough suggests that RISK11 and similar transcriptomic signatures that detect interferon signalling genes may not be useful as triage tests for active case-finding of subclinical tuberculosis disease. This finding illustrates that cough status is a major contributor to the performance ceiling of host blood transcriptomic biomarkers of tuberculosis. Inclusion of host factors with the transcriptomic biomarker in a combination signature may mitigate the deterioration in signature performance over distant prognostic horizons, but does little to improve short-term performance. Future discovery and validation studies should examine the impact of host factors on performance of host blood transcriptomic signatures of tuberculosis across different populations and geographical settings.

4.3 Introduction

Tuberculosis is a major public health problem, killing approximately 1.4 million people in 2019.[1] Effective tuberculosis prevention and control requires rapid diagnosis and treatment of tuberculosis cases and early identification of individuals likely to develop tuberculosis so that they can be treated timeously to interrupt disease progression. However, quick and accurate tuberculosis diagnosis and prediction of progression to tuberculosis disease is hindered by inadequate available tests.[2-4]

Several studies have shown that host blood transcriptomic signatures can be used for both tuberculosis disease diagnosis and to predict progression to tuberculosis disease.[5-7] However, tuberculosis-independent host factors may affect signature readout in individuals without tuberculosis (controls), whereas tuberculosis-dependent host factors would only affect signature readout in individuals with tuberculosis (cases), and potentially affect discriminatory performance. For example, factors independent of tuberculosis risk that increase biomarker scores among controls, such as viral infections[8], might increase the false-positive test rate, whilst host factors that decrease signature score in tuberculosis cases, such as increasing age and male sex[9, 10], might reduce sensitivity.

Tuberculosis-specific characteristics such as disease severity, might also influence the classification performance of the test.[11] Therefore, host factors that shift biomarker distribution among cases and controls should be accounted for when evaluating discriminatory performance of diagnostic tests[12], to allow calculation of covariate-specific performance metrics (outcomes based on stratification of a specific covariate), or covariate-adjusted performance metrics (averaged outcomes which account for each covariate so that a better assessment of classification is obtained than the crude result).[13] Host factors that contribute to accurate classification might be combined with the biomarker to create a combination risk score with improved diagnostic or predictive performance.[14, 15]

We previously discovered and validated a 16-gene RNA host blood transcriptomic signature of risk that identified individuals with prevalent tuberculosis disease and predicted progression from *Mycobacterium tuberculosis* infection to tuberculosis disease.[16] This 16-gene signature was adapted to real-time (RT) quantitative polymerase chain reaction (qPCR) platform and refined to an 11-gene (RISK11) signature with equivalent performance.[17] RISK11 is comprised of interferon signalling pathway genes BATF2, ETV7, FCGR1C, GBP1, GBP2, GBP5, SCARF1, SERPING1, STAT1, TAP1, and TRAFD1 (Appendix 4.1). RISK11 was recently validated in a large multi-centre longitudinal study.[18] We have previously shown that people living with HIV (PLHIV) have significantly higher RISK11 scores, compared to HIV-uninfected individuals[19]; and male sex, older age, and lower HIV viral load are associated with reduced RISK11 score in PLHIV.[20] RISK11 score may also be increased in the presence of upper respiratory viral pathogens[8] However, it is not known whether or which host factors affect RISK11 score; and the direction of associations is not known in HIV-negative individuals. Furthermore, the extent to which host factors affect discriminatory performance has not been quantified in HIV-uninfected populations. Specifically, it is not known whether RISK11 should be adjusted for host characteristics, or whether host characteristics should be included in a combination signature to improve performance.

If transcriptomic biomarkers are to be implemented as tuberculosis triage tests it would be important to account for confounding factors that affect signature performance. This study aimed to (i) identify and quantify the effect of tuberculosis-independent and tuberculosis-dependent host factors on RISK11 score in tuberculosis cases and healthy controls, (ii) quantify the effect of adjustment for host factors on diagnostic performance of RISK11 for prevalent and incident tuberculosis; and (iii) evaluate the effect on discriminatory performance of combining RISK11 with baseline risk factors for tuberculosis.

4.4 Methods

4.4.1 Ethics approval

This analysis is based on the dataset from a randomised clinical trial (CORTIS)[18], conducted between September 2016 and December 2019 in South Africa, which evaluated the performance of RISK11 for diagnosis of prevalent tuberculosis and prediction of incident tuberculosis. The study was approved by Institutional Human Research Ethics Committees of the five participating sites and was also registered on ClinicalTrials.gov (NCT02735590). Written informed consent was sought and obtained from all participants.

4.4.2 Study design and participants

The methodology and main results have been reported previously.[18] In brief, HIV-uninfected adults between the ages of 18 and 60 years, with no history of tuberculosis disease within the last three years or other co-morbidities, were tested for RISK11 at baseline. RISK11 scores were measured by microfluidic RT-qPCR in whole blood RNA as previously described.[17, 18] Briefly, RISK11 is a model of multiple transcript pairs, each functioning as a "vote" for tuberculosis risk. The RISK11 score is the continuous proportion of positive transcript pair votes for tuberculosis risk, ranging from 0–100%. A score threshold can be set for the RISK11 assay to function as a qualitative (positive/negative) test for tuberculosis risk. All participants were screened for prevalent tuberculosis at baseline; those without prevalent tuberculosis were followed for up to 15 months for incident tuberculosis disease. Prevalent tuberculosis was defined as tuberculosis disease diagnosed within 30 days of enrolment (baseline); thereafter, any tuberculosis disease diagnosed was classified as incident disease. Controls were defined as participants without prevalent or incident tuberculosis including those with an unknown outcome at the end of study. Participants provided two expectorated sputum samples for tuberculosis investigation at baseline and end of study (Xpert MTB/RIF or Xpert MTB/RIF Ultra; Cepheid, Franklin Lakes, NJ); interim sputum investigation was symptom-triggered (liquid mycobacterial culture (MGIT, Becton-Dickinson, USA) and Xpert MTB/RIF or Xpert MTB/RIF Ultra). Participants presenting with any one or more symptoms of persistent unexplained cough, weight loss, chest pains, night sweats, fever, for two weeks or more; or any haemoptysis within the last two weeks, were defined as symptomatic. Flu-like symptoms other than those compatible with tuberculosis were also recorded. For this analysis, the microbiologically-confirmed tuberculosis disease endpoint was defined as one or more positive sputum samples by Xpert MTB/RIF, Xpert MTB/RIF Ultra, or MGIT culture. One-sample positive cases were tuberculosis cases in which collection of confirmatory sputum samples did not yield a confirmatory positive result within 30 days. Two-sample positive cases were tuberculosis cases in which collection of confirmatory sputum samples yielded a confirmatory positive result within 30

days. Note that the primary endpoint in the parent study (CORTIS) was two or more positive sputum samples and for this reason a sensitivity analysis was done for prediction of tuberculosis risk using baseline characteristics. A chest radiograph was performed in a sub-set of participants with microbiologically-confirmed tuberculosis and was interpreted by a clinical trial investigator. A positive chest radiograph was defined by any of the following features: hilar or paratracheal lymphadenopathy, miliary pattern, alveolar consolidation, cavitation, pleural effusion, apical shadows, Ghon focus, or calcified nodules.

4.4.3 Statistical analysis

Statistical analyses were performed using STATA/IC version 16.1 (StataCorp., College Station, TX, USA) and MedCalc 20.023 (MedCalc Software Ltd, Ostend, Belgium). Descriptive statistics were computed as either mean and standard deviation (SD), or median and interquartile range (IQR), depending on the distribution, or as frequencies and percentages for categorical variables. The Wilcoxon Rank Sum and Kruskal Wallis tests were used to compare the distribution of RISK11 scores and other numerical variables between two and more than two groups, respectively. Categorical variables were compared using the Chi-squared test, or Fisher's exact test when the expected frequencies were <5. Correlation between RISK11 score and continuous variables was measured using Spearman's rank correlation coefficient.

To quantify the associations between the dependent variable, RISK11 score and each of the predictor variables (host factors) in all participants and specific subgroups of interest (controls and prevalent and incident tuberculosis), univariable generalised linear models (GLMs) were employed (STATA *glm* command). A multivariable GLM was used to estimate the effect of baseline covariates on RISK11 score. Since RISK11 score is a continuous percentage ranging from 0-100%, which was scaled down to a proportion ranging between 0 and 1 for modelling purposes, the logit link function, binomial distribution family and a robust error term (vce-robust) were used in the models.[21, 22] The outcome measure was the percent marginal effect (increase/decrease) on RISK11 score associated with each predictor variable in the model (*margins* command). The model was built using the likelihood ratio test method. First, an initial model with just RISK11 was fitted. Next, nested models were fitted and compared to the initial model with likelihood ratios. The variable with the smallest additional Akaike Information Criterion (AIC) and biggest likelihood ratio, thus making the most significant contribution, was then added to the initial model. The process was repeated until no variable made a significant ($p>0.05$) contribution to the previous model.

To adjust for covariates and evaluate the effect of covariates on the discriminatory performance of RISK11 for either prevalent or incident tuberculosis, ROC regression, using the *rocreg* command in STATA, was performed. The parametric option of the *rocreg* command was employed to allow adjustment for covariates and incorporation of sampling weights in the analysis. First, all variables significantly associated with RISK11 score among tuberculosis-negative controls and variables significantly associated with RISK11 among prevalent and incident tuberculosis cases in multivariable generalised linear regression were included in the ROC regression analyses. All these variables were included in both the "control population" (adjustment for control distribution) and "roc model" (covariates affecting ROC curve) parameters of the ROC regression analysis (Supplementary table S7a). Next, all non-significant variables were removed from each respective section, one at a time until only significant variables remained. Significant variables in the control population as confirmed in ROC regression were included in covariate-adjusted ROC analyses. Variables significant ($p<0.05$) in the roc model section of ROC regression analysis and therefore affecting the ROC curve and by extension discriminatory accuracy, were included in covariate-specific subgroup ROC analyses.[11, 12, 23] Covariate specific subgroup analyses were first performed at a 60% threshold level, which was the RISK11-positivity cut-off point, and thereafter performed at the optimal threshold levels for each subgroup to evaluate whether diagnostic performance improved with optimal covariate-specific thresholds compared to the original 60% thresholds. The Youden Index was used to compute RISK11's optimal covariate-specific thresholds. Outcome measures for the ROC regression were the adjusted AUC and the host factors' effect magnitude on the ROC curve.

To assess the use of baseline characteristics for the diagnosis and prediction of prevalent and incident tuberculosis, ROC analysis was performed. First, logistic and Cox proportional hazards regression models were constructed from covariates that were significant predictors of prevalent and incident tuberculosis (base models) respectively, using the likelihood ratio test method described above. A base model of risk for prevalent tuberculosis was constructed from age, BMI, and cough. A base model of risk for incident tuberculosis was constructed from BMI, smoking history, and previous tuberculosis history using the 15-month follow-up period; and the same base model was applied to the 6- and 12-month follow-up periods. Binary predictor variables included in the models had at least 10 tuberculosis events. The base models were then combined with RISK11 using logistic regression (incremental value method)[11] and Cox regression for prevalent and incident tuberculosis respectively; in order to evaluate the improvement in classification performance. Thus, the outcome measure of interest was the AUC resulting from the combination risk score of RISK11 and the base model. A risk score (combination risk score or predicted risk of disease) was computed for each of the models that

was used to construct the ROC curves. Optimal risk score threshold was based on the maximal Youden index. The area under the curve (AUC) for RISK11 alone and the base model alone were compared to that of the combined RISK11-base model AUC to assess the improvement in AUC. The AUCs were compared using the Delong *et al* method within MedCalc.[24] The univariable and multivariable models as well as ROC analysis were adjusted with probability weights to reflect the CORTIS screening population.

Clarification is made here that covariate-adjustment in ROC analysis is different from using covariates in a predictive model or in incremental value analysis that also evaluate classification performance.[23] When covariates are added in a model evaluating incremental value or prediction, such covariates contribute to the predicted probability of the outcome (combination risk score); usually computed with logistic regression. Thus, the resultant ROC curve for this combination score differs from a covariate-adjusted ROC curve of the test. Because covariates in a combination score contribute to classification, the combination score, may perform well even when the test is a poor classifier, provided that the covariate is a good classifier. In contrast, in covariate adjustment, the classification accuracy of the test is characterised conditional on the covariate.

Enrolment into CORTIS was dependant on RISK11 status and for purposes of conducting the study efficiently, roughly 79% of all eligible RISK11+ and only 13% of all eligible RISK11- participants were enrolled. Thus, the enrolled population was enriched with RISK11+ participants which required assignment of probability weights of 1.263 to RISK11+ and 7.920 to RISK11- individuals to obtain estimates of the screened population. A 0.05 significance level was used for statistical significance in all analyses. Unless otherwise stated, all analyses in this manuscript are based on the microbiologically-confirmed tuberculosis disease endpoint definition (≥1 positive sputum sample) to leverage the increased number of tuberculosis cases relative to the double-positive endpoint (≥2 positive sputum sample) used in the parent study (CORTIS). A sensitivity analysis was performed for baseline predictors of tuberculosis risk. Sample size calculation was performed to ensure that the Primary Objectives of the study could be addressed and did not consider the secondary analyses described here. For this reason, the third aim focused on testing whether baseline host characteristics could be used to improve performance and not necessarily to develop a validated model for prediction.

4.4.4 Role of the funding source

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of this manuscript.

4.5 Results

4.5.1 Baseline characteristics

20,207 volunteers were screened and 2,923 enrolled as previously described (Appendix 4.2). Prevalence of tuberculosis at baseline was 1.4% (74/2923, adjusted to reflect the screened population, see Methods) and cumulative tuberculosis incidence was 1.6% (56/2,849, adjusted) over 15 months. Participant baseline characteristics by tuberculosis status are shown in Table 4.1 (Appendix 4.3).

Table 4.1: Baseline characteristics of enrolled participants by TB status.

Variable	Total	A) Prevalent TB	B) Incident TB	C) Control	A vs C	B vs C	A vs B vs C
	n=2,923	n=74	n=56	n=2,793	P-value	P-value	P-value
Age (median, IQR)	26 (22–33)	29 (24–36)	28 (22–37)	26 (22–33)	0.01	0.17	0.01
BMI (median, IQR)	23 (20–28)	21 (18–24)	20 (19–23)	23 (20–28)	<0.001	<0.001	<0.001
RISK11 Score (median, IQR)	26 (8–77)	87 (61–96)	67 (15–81)	24 (8–75)	<0.001	0.01	<0.001
Male sex (n, %)	1,338 (45.8)	47 (63.5)	33 (58.9)	1258 (45.1)	0.01	0.04	0.01
Ethnicity (n, %)							
Caucasian	4 (0.1)	0 (0)	0 (0)	4 (0.1)			
Mixed	968 (33.1)	34 (45.9)	26 (46.4)	908 (32.5)	0.08	0.18	0.06
Black	1,947 (66.6)	40 (54.1)	30 (53.6)	1,877 (67.2)			
Asian	4 (0.1)	0 (0)	0 (0)	4 (0.1)			
Smoking history (n, %)	1,478 (50.6)	45 (60.8)	41 (73.2)	1,392 (49.8)	0.08	0.01	<0.001
Prior TB (n, %)	230 (7.9)	19 (25.7)	8 (14.3)	203 (7.3)	<0.001	0.06	<0.001
TB contact history (n, %)	462 (15.8)	15 (20.3)	9 (16.1)	438 (15.7)	0.33	0.85	0.53
Flu-like symptoms (n, %)	134 (4.6)	4 (5.4)	1 (1.8)	129 (4.6)	0.78	0.52	0.66
TB Symptoms							
Chest pains (n, %)	30 (1.0)	4 (5.4)	0 (0)	26 (0.9)	0.01	1.00	0.01
Cough (n, %)	58 (2.0)	12 (16.2)	0 (0)	46 (1.6)	<0.001	1.00	<0.001
Fever (n, %)	3 (0.1)	1 (1.4)	0 (0)	2 (0.1)	0.08	1.00	0.13
Haemoptysis (n, %)	2 (0.1)	1 (0)	0 (0)	2 (0.1)	1.00	1.00	1.00
Loss of weight (n, %)	41 (1.4)	5 (6.8)	0 (0)	36 (1.3)	0.01	0.48	0.01
Night sweats (n, %)	32 (1.1)	7 (9.5)	1 (1.8)	24 (0.9)	<0.001	0.39	<0.001
Any symptom (n, %)	123 (4.2)	13 (17.6)	1 (1.8)	109 (3.9)	<0.001	0.72	<0.001

For continuous data, p values were computed using Wilcoxon Rank Sum test between two groups and Kruskal Wallis test for more than two groups. For categorical data, p values were computed using Fischer's exact test. P-values are not corrected for multiple comparisons. Participants that were not diagnosed with tuberculosis and did not complete follow-up for any reason were included in controls. Point estimates are computed using the enrolled population. See Appendix 4.3 for adjusted point estimates to reflect screening population. IQR, inter-quartile range. BMI, body-mass index.

There were no significant baseline differences in ethnicity, tuberculosis contact history, and presence of flu-like symptoms among prevalent tuberculosis cases, those who progressed to incident tuberculosis, and controls without tuberculosis. Compared to controls, participants with prevalent tuberculosis and those who progressed to incident tuberculosis had lower BMI, higher RISK11 scores and majority were males. Additionally, compared to controls, participants with prevalent tuberculosis were older and had higher proportions of prior tuberculosis and symptoms; while those who progressed to incident tuberculosis had a higher proportion of smoking history.

4.5.2 Factors associated with RISK11 in all participants

First, the factors affecting RISK11 score in all participants, including tuberculosis cases and controls were assessed. In a multivariable generalised linear model, the percent marginal changes in RISK11 score were higher among participants with prevalent tuberculosis, those who progressed to incident tuberculosis, those with a smoking history, flu-like symptoms, or night sweats; and lower in males, and with every unit increase in BMI (Figure 4.1a; Multivariable GLM, $p<0.05$). Age, ethnicity, prior tuberculosis disease, tuberculosis contact history, chest pains, cough, fever, haemoptysis, and subjective loss of weight were not independently associated with RISK11 score (Table 4.2).

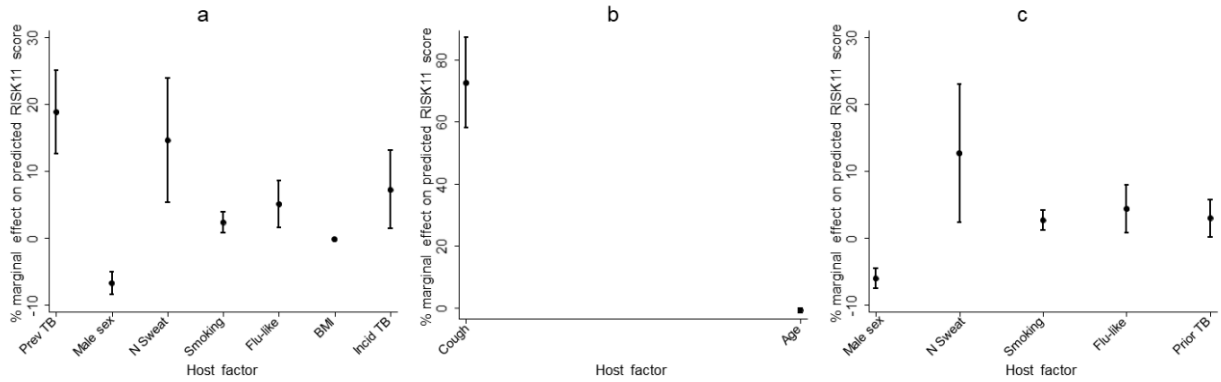


Figure 4.1: Predicted marginal effects on RISK11 by different host factors in (a) all participants, (b) participants with prevalent tuberculosis, and (c) participants without tuberculosis.

Prev TB, Prevalent tuberculosis. N Sweat, Night sweats. Smoking, Smoking history. Flu-like, Flu-like symptoms. BMI, Body-mass index. Incid TB, Incident tuberculosis. Prior TB, Prior tuberculosis. The midline indicates the percentage marginal effect and the error bars indicate the 95% CIs.

Table 4.2: Univariable and multivariable generalised linear models of the predictors of RISK11 score in all participants.

Variable	n =2,923	Univariable Analysis		Multivariable Analysis		
		β. Coef. (95%CI)	P-value	β. Coef. (95%CI)	% Marginal effect (95%CI)	P-value
Age (median, IQR)	26 (22-33)	0.01 (0.001–0.01)	0.02	-		-
BMI (median, IQR)	23 (20-28)	-0.01 (-0.01–0.01)	0.56	-0.01 (-0.01–0.01)	-0.13 (-0.25–-0.001)	0.04
Male sex	1,338 (45.8)	-0.27 (-0.36–-0.19)	<0.001	-0.40 (-0.49–-0.30)	-06.68 (-8.31–-5.04)	<0.001
Race: Black (Reference)	1,947 (66.6)	-	-	-		-
Asian	4 (0.1)	-0.16 (-0.96–0.64)	0.69	-		-
Caucasian (n, %)	4 (0.1)	-0.49 (-1.34–0.35)	0.25	-		-
Mixed (n, %)	968 (33.1)	0.37 (0.29–0.46)	<0.001	-		-
Smoking history (n, %)	1,478 (50.6)	0.06 (-0.02–0.14)	0.16	0.14 (0.05–0.23)	2.41 (0.89–3.93)	<0.001
Prior TB (n, %)	230 (7.9)	0.24 (0.08–0.40)	0.01	-		-
TB contact history (n, %)	462 (15.8)	-0.06 (-0.16–0.05)	0.30	-		-
Flu-like symptoms (n, %)	134 (4.6)	0.32 (0.11–0.54)	0.01	0.30 (0.09–0.52)	5.13 (1.58–8.68)	0.01
Chest pains (n, %)	30 (1.0)	0.17 (-0.28–0.61)	0.46	-		-
Cough (n, %)	58 (2.0)	0.45 (0.13–0.78)	0.01	-		-
Fever (n, %)	3 (0.3)	0.89 (-0.29–2.07)	0.14	-		-
Haemoptysis (n, %)	2 (0.1)	-0.26 (-1.35–-1.16)	<0.001	-		-
Loss of weight (n, %)	41 (1.4)	0.1 (-0.25–0.45)	0.59	-		-
Night sweats (n, %)	32 (1.1)	0.92 (0.36–1.48)	0.01	0.87 (0.32–1.42)	14.65 (5.39–23.91)	0.01
Prevalent TB (n, %)	74 (2.5)	1.12 (0.75;1.49)	<0.001	1.12 (0.75;1.49)	18.90 (12.66–25.13)	<0.001
Incident TB (n, %)	56 (1.9)	0.41 (0.06–0.77)	0.02	0.43 (0.09–0.78)	7.29 (1.46–13.11)	0.01

IQR, inter-quartile range. BMI, body-mass index. β. Coef., Beta coefficient.

% Marginal effect. Percentage marginal change in RISK11 score associated with each respective predictor variable.

Next, the factors affecting RISK11 score in specific groups of interest were assessed, i.e., tuberculosis-dependant factors in prevalent and incident tuberculosis cases; and tuberculosis-independent factors in controls without tuberculosis.

4.5.3 Factors associated with RISK11 in prevalent tuberculosis cases

Among participants with prevalent tuberculosis, RISK11 scores were significantly higher in symptomatic patients (13/74; median=97.0%, IQR=93.1%–98.3%) compared to asymptomatic patients (median=81.8%, IQR=41.6%–93.5%; Wilcoxon rank sum test, p<0.001). No differences were observed in RISK11 scores in participants with or without a history of smoking, prior tuberculosis disease, or household tuberculosis contact history (Appendix 4.4a). Of the 13 symptomatic patients, 10 were QuantiFERON-positive (QFT+). Median RISK11 scores were not different (Wilcoxon rank sum

test, p=0.80) between the 10 QFT+ (96.54%, IQR=82.68%–100.00%) and the three QFT- (96.97%, IQR=94 .37%–99.57%) patients.

Among the 28 prevalent tuberculosis cases in whom a chest radiograph was done, RISK11 scores were significantly higher in the 19 patients with a chest radiograph suggestive of tuberculosis (median=97.8%, IQR=68.4–99.6 vs median=65.8, IQR=15.6%–85.3%; Wilcoxon rank sum test, p=0.03, Appendix 4.5a). One-sample sputum positive cases had lower RISK11 scores (Appendix 4.6a), and a significantly lower proportion of chest radiographs suggestive of tuberculosis, compared to two-sample positive cases (33% vs 86%; Fisher's exact, p=0.02).

In the analysis of factors affecting RISK11 score in the 74 prevalent tuberculosis cases, multivariable regression identified cough as the only factor affecting RISK11 score. Participants with a baseline cough were predicted to have a RISK11 score that was higher by 72.55% (95%CI 58.06–87.03) compared to those without cough (Figure 4.1b, Table 4.3a, Appendix 4.7).

Table 4.3: Multivariable generalised linear models of the predictors of RISK11 score in prevalent TB cases and controls.

	Variable		Multivariable Analysis		
			β. Coef. (95%CI)	% Marginal effect (95%CI)	P-value
A: Prevalent TB; N=74	Age (median, IQR)	29 (24–36)	-0.03 (-0.06–0.01)	-0.65 (-1.34–0.04)	0.07
	Cough (n, %)	12 (16.2)	3.23 (2.45–3.94)	72.55 (56.08–87.03)	0.01
B: Controls; n=2793	Male sex	1,258 (45.0)	-0.36 (-0.45– -0.27)	-5.99 (-0.749– -4.50)	<0.001
	Smoking history (n, %)	1,392 (49.8)	0.16 (0.07–0.25)	2.74 (1.24–4.24)	<0.001
	Prior TB (n, %)	203 (7.3)	0.18 (0.01–0.35)	3.03 (0.25–5.82	0.03
	Flu-like symptoms (n, %)	129 (4.6)	0.26 (0.05–0.48)	4.39 (0.80–7.98)	0.02
	Night sweats (n, %)	24 (0.9)	0.76 (0.14–1.38)	12.69 (2.34–23.04)	0.02

Multivariable models for the factors associated with RISK11 score in (A) prevalent TB cases and (B) controls not diagnosed with either prevalent or incident TB. Complete univariable and multivariable models of this table are shown in Appendices 4.7 and 4.9.

IQR, inter-quartile range. BMI, body-mass index. β. Coef., beta coefficient.

% Marginal effect. Percentage marginal change in RISK11 score associated with each respective predictor variable.

4.5.4 Factors associated with RISK11 in incident tuberculosis cases

Participants with incident tuberculosis were predominantly asymptomatic at baseline (Table 8) and showed no significant differences in baseline RISK11 score among those with or without a smoking history (median scores 66% vs 69%; Wilcoxon rank sum test, p=0.39), prior tuberculosis disease (42%

vs 70%; Wilcoxon rank sum test, p=0.26), or a household tuberculosis contact history (66% vs 67%; Wilcoxon rank-sum test, p=0.93, Appendix 4.4b). A chest radiograph was performed at diagnosis in 36 participants and baseline RISK11 scores were not significantly different between the 25 patients with a positive chest radiograph and the 11 with a negative chest radiograph (median=67.1%, IQR=9.1%–90.0% vs 48.9%, IQR=15.2%–72.7%; Wilcoxon rank sum test, p=0.58; Appendix 4.5b). However, overall, one-sample sputum positive cases had lower RISK11 scores (Appendix 4.6e and 4.6f), and a significantly lower proportion of chest radiographs suggestive of tuberculosis, compared to two-sample positive cases (43% vs 86%; Wilcoxon rank sum test, p=0.01). Stratification by diagnostic window showed that one-sample sputum positive incident cases diagnosed between months 2–6 and 7–12 (Appendix 4.6b and 4.6c, respectively) had lower RISK11 scores compared to two-sample positive cases but there was no difference in RISK11 score distribution between the one-sample and two-sample positive cases diagnosed between months 13–15 (Wilcoxon rank sum test, p=0.89; Appendix 4.6d).

In the analysis of factors affecting RISK11 score in the 56 incident tuberculosis cases, flu-like symptoms and night sweats showed an association with RISK11 in univariable analysis (Appendix 4.8). However, a multivariable model could not be fitted due to insufficient positive observations of participants with flu-like symptoms or night sweats (n=1).

4.5.5 Tuberculosis-independent factors affecting RISK11 in controls

Among controls who remained tuberculosis-free through 15 months, those with any baseline symptom (109/2,793; median RISK11 scores of 48% vs 23%; Wilcoxon rank sum test, p<0.001), baseline night sweats (71% vs 24%; Wilcoxon rank sum test, p=0.01), flu-like symptoms (61% vs 23%; Wilcoxon rank sum test, p<0.001), prior tuberculosis disease (43% vs 23%; Wilcoxon rank sum test, p=0.02), and females (34% vs 16%; Wilcoxon rank sum test, p<0.001) had significantly higher baseline RISK11 scores (Appendix 4.3c). 2,603 of the 2,793 tuberculosis-free controls were asymptomatic and also free from flu-like symptoms compatible with tuberculosis, of which 960 (960/2,603) had elevated (≥60%) RISK11 scores (median=84.0%, IQR=70.1%–94.4%).

Analysis of factors affecting RISK11 score in the 2,793 controls without tuberculosis using multivariable regression identified smoking history, prior tuberculosis, flu-like symptoms, night sweats, and sex as significant factors affecting RISK11 in controls. The percentage marginal effect on RISK11 score were higher in participants with a smoking history, prior tuberculosis, flu-like symptoms, or night sweats than in those without these characteristics; and lower in males than females (Multivariable GLM, p<0.05; Figure 4.1c, Table 4.3b, Appendix 4.9).

4.5.6 Effect of host factors on RISK11 signature performance

Next, the effect of these baseline covariates on diagnostic and prognostic performance of RISK11 for tuberculosis was assessed using ROC regression analysis. RISK11 diagnostic performance (AUC 0.74, 95%CI 0.67–0.82) and prognostic performance through 15-months follow-up (AUC 0.56, 95%CI 0.46–0.68) for the one sample-positive tuberculosis cases was previously reported using nonparametric methods.[18] In the current analysis, a parametric method was used to allow adjustment for covariates and incorporation of sampling weights; the unadjusted parametric ROC analysis yielded results that were similar to the published non-parametric analysis with diagnostic AUC of 0.72 (95%CI 0.65–0.80) and prognostic AUC of 0.59 (95%CI 0.51–0.66).

Adjustment for tuberculosis-independent host factors that significantly altered RISK11 distribution in controls (i.e. BMI, sex, night sweats, haemoptysis, flu-like symptoms, and smoking history) did not significantly alter the AUC for diagnostic performance for prevalent tuberculosis (AUC 0.72 vs 0.72; Delong method, p=0.98). Similarly, the covariate-adjusted AUC for prognostic performance through 15-months follow-up (0.60, 95%CI: 0.52–0.67; Delong method, p=0.94) did not significantly differ from the crude AUC (Figure 4.2).

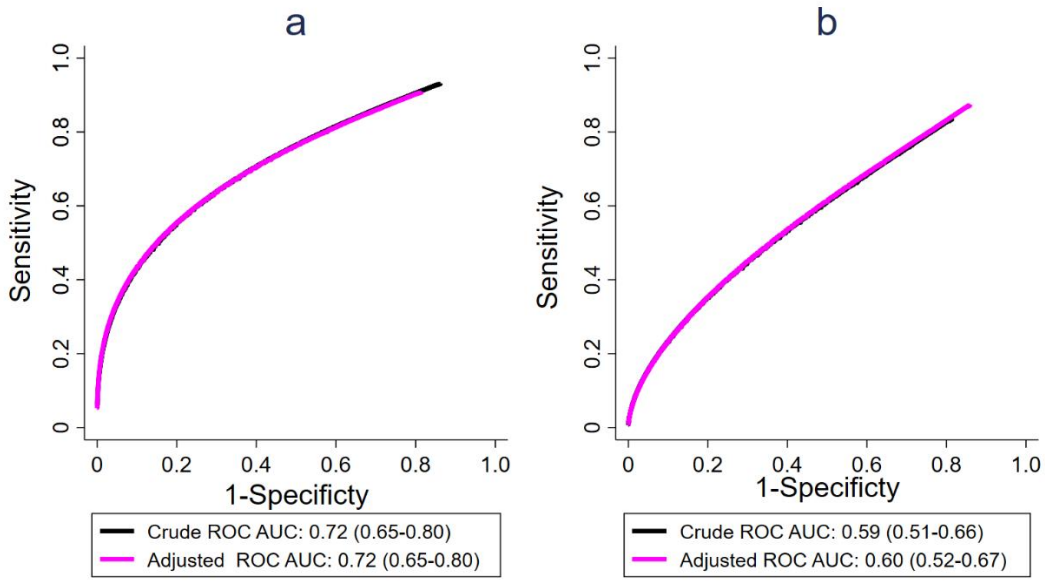


Figure 4.2: Crude and covariate-adjusted ROC curves for the discrimination of (a) prevalent TB from controls and (b) incident TB from controls.

The crude and covariate-adjusted ROC curves are superimposed in both figures (a) and (b). The ROC curves are adjusted for BMI, sex, night sweats, haemoptysis, flu-like symptoms, and smoking history in both instances.

Baseline cough and flu-like symptoms were the two factors that affected (Multivariable ROC regression, p<0.001) discriminatory performance of RISK11 in prevalent and incident tuberculosis cases, respectively, in ROC regression (Appendix 4.10a and 4.10b). Covariate-specific ROC curves computed for cough showed a high discrimination between cough-positive prevalent tuberculosis cases and cough-positive controls (AUC 0.97, 95%CI 0.90–1.00). In contrast, the AUC for discriminating cough-negative prevalent tuberculosis cases from cough-negative controls was 0.72 (95%CI 0.65–0.79; Delong method, p<0.001, Figure 4.3a, Table 4.4c). Diagnostic accuracy improved from 62.2% in all to 94.6% in cough-positive individuals, at the 60% RISK11-positivity threshold (Table 4.4a vs 4.4c). The optimal cough-specific RISK11-positivity thresholds were 76% and 26% for cough-positive and cough-negative individuals respectively; and using these thresholds marginally improved covariate-specific performance of RISK11 (Table 4.4c vs 4.4d). Covariate-specific ROC curves were not computed for discriminating incident tuberculosis cases from controls in participants with and without flu-like symptoms, because only one individual with incident tuberculosis had flu-like symptoms.

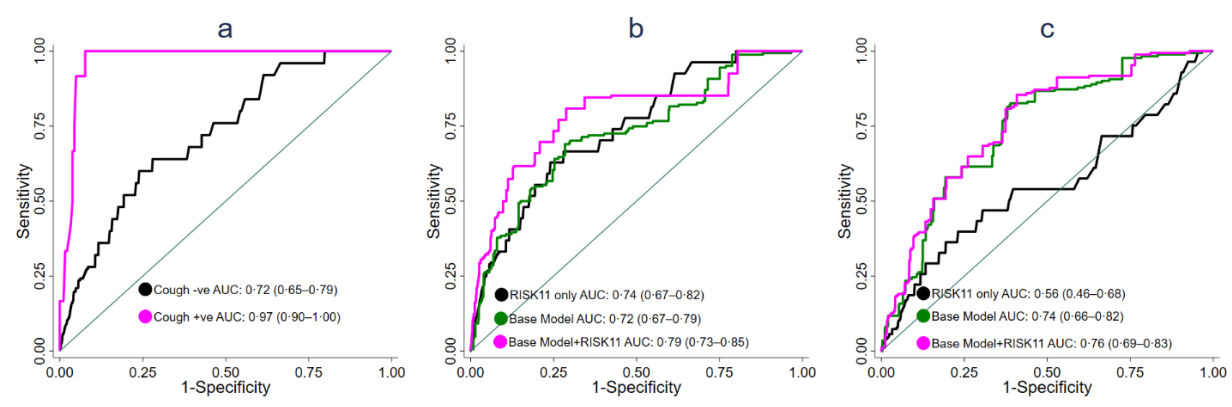


Figure 4.3: Performance of RISK11 when stratified by cough status and when combined with host factors.

(a) Covariate-specific ROC curves for the diagnostic performance of RISK11 in cough-positive (n=58) and cough-negative (n=2,865) individuals. Crude (RISK11 only), and combination ROC curves for discriminating (b) prevalent TB versus controls, and (c) incident TB versus controls through 15 months follow-up. Baseline model AUCs were derived from predictive models containing age, BMI, and cough for prevalent TB; and BMI smoking history, and previous TB history for incident TB.

Table 4.4: Crude and cough-specific performance estimates for RISK11 at various thresholds.

Statistic	a) Crude	b) Crude	c) Covariate-Specific		d) Covariate-Specific	
	RISK11(60)† n=2,923	RISK11(26)† n=2,923	Cough+ RISK11(60)† n=58	Cough- RISK11(60)† n=2,865	Cough+ RISK11(76)† n=58	Cough- RISK11(26)† n=2,865
PR, (95%CI)	4.8 (2.9–8.2)	4.86 (2.57–11.70)	10.7 (6.3–18.1)	3.9 (2.23–6.77)	13.4 (8.2–22.0)	4.4 (2.0–9.4)
AUC, (95%CI)	0.74 (0.67–0.82)		0.97 (0.90–100.0)	0.72 (0.65–0.79)	0.97 (0.90–100.0)	0.72 (0.65–0.79)
Sensitivity, % (95%CI)	33.2 (23.4–47.6)	62.9 (47.0–80.5)	100.0 (73.5–100.0)	28.1 (16.9–40.2)	100.0 (73.5–100.0)	60.0 (46.4–71.9)
Specificity, % (95%CI)	91.1 (91.0–91.2)	74.7 (73.1–76.4)	89.2 (76.4–96.4)	91.1 (90.0–92.1)	92.4 (79.2–97.6)	74.9 (73.3–76.5)
PPV % (95%CI)	4.9 (3.7–6.2)	3.3 (2.3–4.6)	37.5 (21.1–56.3)	4.0 (2.9–5.3)	46.2 (26.6–66.6)	3.0 (2.2–4.1)
NPV % (95%CI)	99.0 (98.5–99.4)	99.3 (98.9–99.7)	100 (86.8–100.0)	99.0 (98.4–99.4)	100 (89.1–100.0)	99.3 (98.7–99.7)
Accuracy, % (95%CI)	62.2 (60.4–64.0)	68.8 (67.1–70.5)	94.6 (85.6–98.9)	59.6 (57.8–61.4)	96.2 (88.1–99.6)	67.5 (65.8–69.2)

Covariate-specific subgroup analyses were first performed at 60% threshold level, which was the RISK11-positivity cut-off point, and thereafter performed at the optimal threshold levels for each subgroup, to evaluate whether diagnostic performance improved with optimal covariate-specific thresholds compared to the original 60% thresholds.

†Numbers in brackets following RISK11 denote the RISK11-positivity threshold. Cough-specific thresholds in (d) were the optimal thresholds at the maximal Youden index.

PR, prevalence ratio. AUC, area under the curve. PPV, positive predictive value. NPV, negative predictive value.

4.5.7 Combination of RISK11 with host factors

An assessment of whether combining RISK11 score with other baseline variables would improve discriminatory performance was made. Univariable and multivariable baseline predictors of either prevalent or incident tuberculosis and performance of these models is shown in Table 6. In multivariable analyses, the significant host predictors of prevalent tuberculosis were age, BMI, and cough, which formed the base model for prevalent tuberculosis (Table 4.5a). Combining the base model with RISK11 discriminated prevalent tuberculosis from controls with an AUC of 0.79 (95%CI 0.77–0.87), a non-significant (p=0.06) increase of 5% compared to the 0.74 (95%CI 0.67–0.82) for RISK11 alone (Figure 4.3b).

Similarly, host predictors included in the incident tuberculosis base model were BMI, smoking history, and previous tuberculosis history (Table 4.5b-d). Combining the incident tuberculosis base model with RISK11 significantly improved discrimination between incident tuberculosis and controls, from AUCs of 0.62 (95%CI 0.51–0.73) and 0.56 (95%CI 0.46–0.68) for RISK11 alone, to AUCs of 0.80 (95%CI 0.70–0.90; Delong method, p=0.02) and 0.76 (95%CI 0.69–0.83; Delong method, p<0.001) for the

combination model, over 12- and 15-month prognostic horizons respectively (Table 4.5c & 4.5d, Figure 4.3c). However, combination of the incident tuberculosis base model with RISK11 did not significantly improve prediction compared to RISK11 alone (Delong method, p=0.11) through a 6-month follow-up period (Table 4.5b).

Table 4.5: Performance of RISK11, base models, and combination models for TB disease.

Group	Model	Variable	Univariable Analysis		Multivariable Analysis		≥1 positive sputum sample) AUC (95%CI)	≥2 positive sputum sample) AUC (95%CI) †	P
			OR (95%CI)	P-value	aOR (95%CI)	P-value			
(a) Prevalent TB	RISK11 only	RISK11	1.03 (1.02–1.04)	<0.001	–		0.74 (0.67–0.82)	0.77 (0.68–0.86)	0.54
	Base Model	Age	1.06 (1.03–1.08)	<0.001	1.06 (1.04–1.10)	<0.001	0.72 (0.65–0.79)	0.76 (0.69–0.84)	0.42
		BMI	0.91 (0.85–0.98)	0.01	0.90 (0.83–0.97)	0.01			
		Cough	5.01 (2.37–10.57)	<0.001	2.63 (1.10–6.27)	0.03			
	Base Model + RISK11	Age	1.06 (1.03–1.08)	<0.001	1.06 (1.03–1.09)	<0.001	0.79 (0.73–0.85)	0.83 (0.77–0.89)	0.38
		BMI	0.91 (0.85–0.98)	0.01	0.90 (0.84–0.97)	0.01			
		Cough	5.01 (2.37–10.57)	<0.001	2.23 (1.05–4.73)	0.04			
		RISK11	1.03 (1.02–1.04)	<0.001	1.03 (1.02–1.03)	<0.001			
Group	Model	Variable	Univariable Analysis		Multivariable Analysis		≥1 positive sputum sample) AUC (95%CI)	≥2 positive sputum sample) AUC (95%CI) †	P
			HR (95%CI)	P-value	aHR/ (95%CI)	P-value			
(b) Incident TB through 6 months	RISK11 only	RISK11	1.02 (1.00–1.04)	0.13	–		0.62 (0.45–0.79)	0.95 (0.92–1.00)	0.01
	Base Model	BMI	0.90 (0.77–1.05)	0.17	0.90 (0.80–1.02)	0.09	0.66 (0.49–0.83)	0.58 (0.34–0.82)	0.64
		Smoking	1.73 (0.34–8.73)	0.51	1.11 (0.29–4.18)	0.88			
		Prior TB	4.85 (0.71–33.12)	0.11	4.50 (0.71–28.47)	0.11			
	Base Model + RISK11	BMI	0.90 (0.77–1.05)	0.17	0.90 (0.80–1.02)	0.10	0.73 (0.57–0.89)	0.96 (0.85–1.00)	0.02
		Smoking	1.74 (0.34–8.80)	0.5	1.10 (0.29–4.19)	0.89			
		Prior TB	4.89 (0.7–33.87)	0.11	4.16 (0.64–26.91)	0.13			
		RISK11	1.02 (1.00–1.04)	0.13	1.02 (0.99–1.04)	0.17			
(c) Incident TB through 12 months	RISK11 only	RISK11	1.02 (1.00–1.03)	0.01	–		0.62 (0.51–0.73)	0.80 (0.65–0.94)	0.04
	Base Model	BMI	0.84 (0.76–0.93)	0.01	0.87 (0.80–0.94)	0.01	0.77 (0.67–0.87)	0.76 (0.63–0.89)	0.92
		Smoking	3.69 (1.21–11.26)	0.02	2.22 (0.76–6.49)	0.15			
		Prior TB	4.02 (1.26–12.84)	0.02	3.38 (1.01–11.31)	0.05			
	Base Model + RISK11	BMI	0.84 (0.76–0.93)	0.01	0.87 (0.80–0.94)	0.01	0.80 (0.70–0.90)	0.87 (0.76–0.98)	0.33
		Smoking	3.69 (1.21–11.26)	0.02	2.21 (0.76–6.44)	0.15			
		Prior TB	4.02 (1.26–12.84)	0.02	3.15(0.94–10.54)	0.06			
		RISK11	1.02 (1.00–1.03)	0.01	1.01 (1.00–1.03)	0.02			
(d) Incident TB through 15 months	RISK11 only	RISK11	1.01 (1.00–1.02)	0.01	–		0.56 (0.46–0.68)	0.63 (0.47–0.80)	0.30
	Base Model	BMI	0.87 (0.80–0.94)	0.01	0.90 (0.83–0.97)	0.01	0.74(0.66–0.82)	0.80 (0.70–0.90)	0.34
		Smoking	3.92 (1.76–8.74)	0.01	2.62 (1.13–6.08)	0.03			
		Prior TB	3.09 (1.27–7.51)	0.01	2.61 (1.03–6.57)	0.04			
	Base Model + RISK11	BMI	0.87 (0.80–0.94)	0.01	0.90 (0.82–0.97)	0.01	0.76 (0.69–0.83)	0.82 (0.73–0.92)	0.32
		Smoking	3.92 (1.76–9.08)	0.01	2.61 (1.14–6.24)	0.03			
		Prior TB	3.09 (1.27–7.51)	0.01	2.46 (1.00–6.20)	0.05			
		RISK11	1.01 (1.00–1.02)	0.01	1.01 (1.00–1.02)	0.02			

All modelling data shown in this table is based on the one-sample positive endpoint definition (≥1 positive sputum sample) which was the primary endpoint for this analysis. Point estimates for prevalent tuberculosis are adjusted odds ratios (aOR) and those for incident (Tables b, c and d) tuberculosis are adjusted hazard ratios (aHR). AUCs shown with '†' are based on a sensitivity analysis computed using the two-sample positive endpoint definition (≥2 positive sputum samples) which was the primary endpoint in CORTIS. The corresponding model output data for the two-sample positive endpoint are not shown. BMI, body-mass index. AUC, area under the curve. P values comparing the AUCs for the one-sample versus two-sample positive endpoint definition were computed using the Delong method in MedCalc.

4.6 Discussion

We have shown that although several host factors affected RISK11 readout, adjustment for tuberculosis-independent host factors affecting controls did not change diagnostic or prognostic performance of the RISK11 transcriptomic signature. However, stratification for cough status, a tuberculosis-dependent factor that was associated with a 72.55% marginal increase in RISK11 score in those with prevalent tuberculosis, significantly improved discriminatory accuracy in individuals with cough. However, diagnostic performance in individuals without cough was poor.

We also showed that although certain host factors affecting RISK11 score are also associated with tuberculosis risk, incorporation of these host factors into a combination signature did not significantly improve diagnostic performance for prevalent tuberculosis. By contrast, combining baseline host factors with RISK11 significantly improved discrimination of incident tuberculosis from controls compared to RISK11 alone over the longer 12- and 15-month predictive horizons. These findings may also be generalisable to other transcriptomic signatures that, like RISK11, include interferon-stimulated genes.

These findings build upon our previous work that showed the effect of HIV infection[19] and upper respiratory viral pathogens on RISK11 score[8]; and on the work of others who have evaluated the effect of host factors on performance of transcriptomic signatures[25] and combined biomarkers and clinical variables to improve prediction of tuberculosis risk and treatment outcomes.[15, 26, 27] In addition to identifying host characteristics associated with changes in RISK11 score and tuberculosis risk, we have quantified the effect of these host factors on the ability of RISK11 to discriminate between participants with prevalent tuberculosis or incident tuberculosis from controls without tuberculosis.

Viral or other infections in participants without tuberculosis cannot be excluded as the cause of the raised RISK11 scores since we did not test for respiratory or other pathogens. In a sub-study that co-enrolled 286 participants and tested for upper respiratory tract pathobionts in nasopharyngeal and

oropharyngeal swabs, RISK11 was able to differentiate between participants with prevalent tuberculosis and those with no pathobionts detected or only bacterial pathobionts. However, RISK11 could not differentiate between participants with prevalent tuberculosis and those with upper respiratory tract viruses.[8] Similarly, HIV infection has been associated with raised RISK11 scores, especially in those with uncontrolled viral load, and was associated with diminished performance in one study where most participants were not on antiretroviral therapy,[19] but associated with good performance in another study where most participants were on stable antiretroviral therapy.[20]

In a recent study, it was shown that transcriptomic signatures have good discriminatory capacity for tuberculosis disease in participants presenting with symptoms compatible with tuberculosis.[28] We also previously showed that diagnostic performance of RISK11 for prevalent tuberculosis was superior in symptomatic tuberculosis cases, compared to cases of subclinical tuberculosis[18], which form a large proportion of tuberculosis cases in community prevalence surveys.[29] Here, we evaluated which symptom component underlies this difference in performance. Cough was the only host factor that was significantly associated with a raised RISK11 score in participants with prevalent tuberculosis, with a 72.55% marginal increase in RISK11 score compared to cough-negative cases, and cough affected discriminatory performance in ROC regression. Cough is the most common manifesting symptom of symptomatic tuberculosis disease and thus typically distinguishes symptomatic from subclinical presentation, which may be associated with less severe disease, as suggested by higher Xpert/MTB RIF Ct values in subclinical disease.[30] The superior performance of RISK11 in participants with symptomatic tuberculosis disease which may be associated with severe inflammation, suggests induction of interferon signalling resulting in elevated signature scores that drive the superior discriminatory performance. We found that RISK11 had excellent diagnostic performance at a 76% RISK11-positivity threshold in a small number of cough-positive individuals and similar performance at 26% RISK11-positivity threshold in cough-negative individuals to that of the crude estimates at a 26% RISK11-positivity threshold (Table 4.4). Although discriminatory performance for screening of asymptomatic individuals might be improved by using a different threshold than for symptomatic patients, the use of multiple thresholds for different populations would complicate interpretation and likely hinder implementation in the field.

Several host characteristics, including some factors associated with tuberculosis risk, were associated with a significantly increased or decreased RISK11 score in controls. Although we surmised it might be important to incorporate covariate information in assessing discriminatory performance, we showed

that RISK11 performance was not different between covariate-adjusted and crude ROC curves (Figure 4.2).

Several studies have shown that combining biomarkers with host risk factors may significantly improve classification capacity.[14, 15, 26, 27, 31] Sivakumaran *et al* found that signature performance for predicting tuberculosis treatment outcomes was improved when they combined host-derived biomarkers with patient characteristics.[27] We demonstrated that combining the significant clinical predictors of incident tuberculosis through 12- and 15-months with RISK11 significantly improved discriminatory capacity compared to RISK11 alone, but not for incident tuberculosis through 6-months, or for prevalent tuberculosis. This important finding demonstrates that a classification model consisting of RISK11 plus baseline characteristics may improve discriminatory capacity for predicting incident tuberculosis through longer predictive horizons at which transcriptomic signature performance deteriorates. Alternatively, a signature discovered in asymptomatic participants might be required in a classification model to improve short term classification of incident tuberculosis.[32]

Weaknesses of our study may include the fact we used a tuberculosis disease endpoint definition based on one positive sputum sample, as used in the public health system. However, one-sample sputum positive cases were predominantly subclinical, with fewer chest radiographs suggestive of tuberculosis, and lower RISK11 scores compared to the two-sample sputum positive cases (Appendix 4.6). It is therefore not surprising that the one-sample-positive endpoint showed poorer RISK11 performance compared to a two-sample-positive endpoint used in the CORTIS trial, which increases the potential for host factors to improve performance in a combination model. Furthermore, although we found that RISK11 performance was better in participants with a cough, this was based on a relatively small sample size. Baseline predictors performed well relative to RISK11 for prediction of tuberculosis risk over longer time-frames. It should be noted that this study reports the training cohort for these host factors and validated performance would require testing in an independent cohort. Strengths of this study include the large study sample, large number of tuberculosis cases, and the fact that the study recruited from five geographically distinct areas throughout South Africa with unique population demographics. These findings from five geographically distinct sites are broadly representative of community settings with high prevalence of undiagnosed subclinical tuberculosis in South Africa. They may not be applicable to other countries with low rates of prevalent and incident tuberculosis.

This study highlights that the discriminatory performance of RISK11 and potentially other transcriptomic signatures may be affected by host factors and the tuberculosis endpoint definition. Although this study showed that only cough influenced discriminatory performance of RISK11, further work may be warranted in high-risk populations, for example PLHIV or other co-morbidities such as diabetes mellitus. Future transcriptomic signature discovery studies should not ignore host characteristics in their design. Evidence from this study suggests that presence or absence of cough has a major impact on diagnostic performance for tuberculosis disease, which might severely limit the utility of transcriptomic biomarkers for triage and active case-finding approaches for subclinical tuberculosis, which forms a large proportion of prevalent tuberculosis in endemic communities.[29]

Contributors

MH and TJS conceived and directed the study. MT, GW, KN, and GC were responsible for site-level activities, including recruitment, clinical management, and data collection. HM, SCM, SKM, and MM provided operational or laboratory support and project management. AFG provided statistical support. HM analysed the data and wrote the first draft of the manuscript. HM, AFG, SCM, APN, SKM, BB, MM, MT, GW, KN, GC, TJS and MH had full access to the data, and reviewed, revised, and approved the manuscript before submission.

Declaration of interests

AP-N, GW, GC, TJS, and MH report grants from the Bill & Melinda Gates Foundation, during the conduct of the study; AP-N and GW report grants from the South African Medical Research Council, during the conduct of the study; GW and TJS report grants from the South African National Research Foundation, during the conduct of the study. In addition, AP-N and TJS have patents of the RISK11 and RISK6 signatures pending; GW has a patent “TB diagnostic markers” (PCT/IB2013/054377) issued and a patent “Method for diagnosing TB” (PCT/IB2017/052142) pending. All other authors had nothing to disclose.

Acknowledgements

The authors gratefully acknowledge the study participants and funders. The CORTIS-01 study was funded by the Bill and Melinda Gates Foundation (BMGF; OPP1116632 and OPP1137034) and the Strategic Health Innovation Partnerships (SHIP) Unit of the South African Medical Research Council with funds received from the South African Department of Science and Technology.

SCM is a recipient of PhD funding from the Fogarty International Center of the National Institutes of Health (NIH) under Award Number D43 TW010559, the Harry Crossley Clinical Research Fellowship, and the South African Medical Research Council (SAMRC) through its Division of Research Capacity Development under the SAMRC Clinician Researcher Programme.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of this manuscript.

Data sharing statement

The study protocol, de-identified RISK11 signature scores, TB endpoint data, and clinical metadata for all participants is available on Zivahub (https://doi.org/10.25375/uct.13573337.v1), an open access data repository hosted by the University of Cape Town's data repository powered by Figshare for Institutions.

4.7 References

1 STATA. Fractional response regression Geneva, Switzerland: StataCorp., College Station, TX, USA; 2015 [Available from: https://www.stata.com/manuals14/rfracreg.pdf.

2 Hamada Y, Cirillo DM, Matteelli A, Penn-Nicholson A, Rangaka MX, Ruhwald M. Tests for tuberculosis infection: landscape analysis. *The European respiratory journal*. 2021.

3 Treatment-Action-Group. Pipeline Report 2019-Tuberculosis Diagnostics. New York, USA: Treatment Action Group; 2019.

4 Nema V. Tuberculosis diagnostics: Challenges and opportunities. *Lung India*. 2012;29(3):259-66.

5 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;6(3):e17984.

6 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(4):395-406.

7 Warsinske H, Vashisht R, Khatri P. Host-response-based gene signatures for tuberculosis diagnosis: A systematic comparison of 16 signatures. *PLoS medicine*. 2019;16(4):e1002786.

8 Mulenga H, Musvosvi M, Mendelsohn SC*, et al.* Longitudinal Dynamics of a Blood Transcriptomic Signature of Tuberculosis. *American journal of respiratory and critical care medicine*. 2021.

9 Molony RD, Nguyen JT, Kong Y, Montgomery RR, Shaw AC, Iwasaki A. Aging impairs both primary and secondary RIG-I signaling for interferon induction in human monocytes. *Sci Signal*. 2017;10(509).

10 Klein SL, Flanagan KL. Sex differences in immune responses. *Nature reviews Immunology*. 2016;16(10):626-38.

11 Janes H, Longton G, Pepe M. Accommodating Covariates in Receiver Operating Characteristic Analysis. *Stata Journal*. 2009;9:17-39.

12 Janes H, Pepe MS. Adjusting for covariate effects on classification accuracy using the covariate-adjusted receiver operating characteristic curve. *Biometrika*. 2009;96(2):371-82.

13 Pardo-Fernandez JC, Rodríguez-Álvarez M, Keilegom I. A review on ROC curves in the presence of covariates. *Revstat - Statistical Journal*. 2014;12:21-41.

14 Patel VB, Singh R, Connolly C*, et al.* Comparison of a clinical prediction rule and a LAM antigen-detection assay for the rapid diagnosis of TBM in a high HIV prevalence setting. *PloS one*. 2010;5(12):e15664.

15 Tenforde MW, Gupte N, Dowdy DW*, et al.* C-reactive protein (CRP), interferon gamma-inducible protein 10 (IP-10), and lipopolysaccharide (LPS) are associated with risk of tuberculosis after initiation of antiretroviral therapy in resource-limited settings. *PloS one*. 2015;10(2):e0117424.

16 Zak DE, Penn-Nicholson A, Scriba TJ*, et al.* A blood RNA signature for tuberculosis disease risk: a prospective cohort study. *Lancet (London, England)*. 2016;387(10035):2312-22.

17 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*. 2018;108:124-6.

18 Scriba TJ, Fiore-Gartland A, Penn-Nicholson A*, et al.* Biomarker-guided tuberculosis preventive therapy (CORTIS): a randomised controlled trial. *The Lancet Infectious diseases*. 2021.

19 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. *Frontiers in microbiology*. 2019;10:1441.

20 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.

21 Baum CF. Stata tip 63: Modeling proportions. *Stata Journal*. 2008;8(2):299-303.

22 Chen K, Cheng Y, Berkout O, Lindhiem O. Analyzing Proportion Scores as Outcomes for Prevention Trials: a Statistical Primer. *Prev Sci*. 2017;18(3):312-21.

23 Janes H, Pepe MS. Adjusting for covariates in studies of diagnostic, screening, or prognostic markers: an old concept in a new setting. *Am J Epidemiol*. 2008;168(1):89-97.

24 DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44(3):837-45.

25 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. *The Lancet Respiratory medicine*. 2020;8(4):407-19.

26 Gupta RK, Calderwood CJ, Yavlinsky A*, et al.* Discovery and validation of a personalized risk predictor for incident tuberculosis in low transmission settings. *Nature medicine*. 2020;26(12):1941-9.

27 Sivakumaran D, Jenum S, Vaz M*, et al.* Combining host-derived biomarkers with patient characteristics improves signature performance in predicting tuberculosis treatment outcomes. *Commun Biol*. 2020;3(1):359.

28 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. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2021.

29 Van der Walt M, Moyo S. The First National TB Prevalence Survey - South Africa 2018. South Africa: South African National Department of Health; 2020.

30 Du Bruyn E, Ruzive S, Lindestam Arlehamn CS*, et al.* Mycobacterium tuberculosis-specific CD4 T cells expressing CD153 inversely associate with bacterial load and disease severity in human tuberculosis. *Mucosal Immunol*. 2020.

31 Bedell RA, van Lettow M, Meaney C*, et al.* Predictive value of C-reactive protein for tuberculosis, bloodstream infection or death among HIV-infected individuals with chronic, non-specific symptoms and negative sputum smear microscopy. *Tropical medicine & international health : TM & IH*. 2018;23(3):254-62.

32 Kwan PKW, Periaswamy B, De Sessions PF*, et al.* A blood RNA transcript signature for TB exposure in household contacts. *BMC infectious diseases*. 2020;20(1):403.

Chapter 5

5 The effect of upper respiratory tract organisms on RISK11 score.

Chapter overview

Viral infection is one factor that may affect host blood transcriptomic signature scores. Amongst individuals without TB disease, detectable HIV viral load has been associated with raised signature scores compared to undetectable viral load, likely because of induction of type I interferon (IFN) and raised expression of IFN-stimulated genes (ISG), which are included in RISK11 and other similar TB signatures. Influenza is another viral infection that induces ISGs and has been shown to affect transcriptomic signatures scores. However, the effect of other common viral and bacterial upper respiratory organisms on TB signatures remains largely unexplored. This chapter evaluated the effect of upper respiratory tract organisms on RISK11 score and discriminatory performance. The socio-demographic, viral and microbiologic factors affecting RISK11 score in the respiratory organisms cohort are presented as evaluated through a multivariable analysis.

The data presented in this chapter has been published in a peer-reviewed journal. It is not presented verbatim because the manuscript to which these data contributed was jointly first co-authored with another PhD student S.C Mendelsohn. Thus, only my work and contribution to that manuscript is included in this chapter (**Mulenga H**, Musvosvi M, Mendelsohn SC, *et al*. *Longitudinal Dynamics of a Blood Transcriptomic Signature of Tuberculosis.* Am J Respir Crit Care Med, 2021. **204**(12): p. 1463-1472. https://doi.org/10.1164/rccm.202103-0548oc).

Chapter contribution to the thesis

This chapter addresses Aim 3 of the thesis. Prior to this analysis, few data were available on the effect of upper respiratory tract viruses or bacteria on RISK11 score, or other signatures that contain interferon signaling genes.

Contributions of the candidate

The candidate designed and planned the study and wrote the protocol under supervision from the PhD supervisors. Furthermore, the candidate was the responsible investigator for implementation and training of study personnel and provided overall study oversight. The candidate was also responsible

for coordinating the transportation of samples between the satellite laboratory at SATVI and testing laboratory at UCT and coordinated the testing of samples and collection of results. Additionally, the candidate created the database for the study and performed the data management activities such as query resolution and data quality management and validation, giving rise to quality data for this analysis. The candidate also prepared the data for analysis, analysed and interpreted the data, and wrote the chapter. The PhD supervisors, MH and TJS conceived the study.

5.1 Abstract

Background

HIV and influenza have been shown to affect transcriptomic signatures. However, effects of common upper respiratory tract organisms on transcriptomic signatures of tuberculosis (TB) have not been systematically studied. This study aimed to evaluate the effect of upper respiratory tract organisms on an 11-gene blood transcriptomic TB signature, RISK11, and to test whether RISK11 differentiates between individuals with and without TB or other respiratory organisms.

Methods

A cross-sectional sub-study of upper respiratory tract bacterial, viral, and fungal organisms was nested in a larger prospective TB biomarker cohort (CORTIS) at screening. A convenience sample of consecutive participants provided one nasopharyngeal, one oropharyngeal swab and; a PAXgene blood sample for measurement of RISK11. Multiplex real-time PCR was used to detect a panel of 33 upper respiratory organisms. A subset of participants was co-enrolled into CORTIS and underwent evaluation for microbiologically-confirmed TB at baseline and through 15 months of follow-up. A multivariable generalised linear model was used to estimate the effect of upper respiratory organisms on RISK11 score (% marginal effect). The area under the receiver operating characteristic curve (AROC) was used to differentiate participants with and without TB or other upper respiratory organisms.

Findings

1,000 HIV-negative volunteers with median age of 27 years (interquartile range, IQR, 22–34) were enrolled. Prevalence of all respiratory organisms was 42.9%: 4% were viruses only, 35.7% bacteria only, and 3.2% were a combination of viruses and bacteria. Overall, RISK11 scores were higher in participants with viral organisms only (46.7%) or viral and bacterial organisms (42.8%) than participants with bacterial organisms other than TB (13.4%), or no organisms (14.2%). Among the 286 participants investigated for TB, in whom 3.8% (11/286) and 3.2% (9/286) were diagnosed with prevalent and incident TB, respectively, RISK11 scores were significantly higher in participants with prevalent TB (85.7%), or incident TB and viruses (82.7%), or viruses only (77.5%) or both viruses and bacteria (89.2%) compared to participants without these organisms (24.6%) or participants with bacteria only (20.4%) or participants with incident TB only (28.6%). In multivariable linear regression and controlling for prevalent TB, percent marginal effects on RISK11 score were predicted to be higher by 16.7% (95%CI 4.1%–29.4%), 67.8% (95%CI 52%–83.5%) and 13.5% (95%CI 3.5%–23.5%) in participants with coronaviruses, influenza and rhinoviruses, respectively. RISK11 scores differentiated

participants as follows: prevalent TB versus bacteria (AUC=0.70; 95%CI 0.51–0.86), prevalent TB versus no-organism (AUC=0.70; 95%CI 0.51–0.88), virus versus bacteria only (AUC=0.71; 95%CI 0.64–0.78) virus versus no-organism (AUC=071; 95%CI 0.64–0.78) and bacteria versus no-organism (AUC=0.51; 95%CI 0.47–0.54). RISK11 could not discriminate prevalent TB from viruses (AUC=0.48; 95%CI 0.28–0.70).

Conclusion

RISK11 could not discriminate between TB and viral upper respiratory tract organisms; emphasising the problem that viral upper respiratory tract organisms are important confounding factors of transcriptomic signatures of TB and highlights a challenge for implementation of these biomarkers as new tests for TB.

5.2 Introduction

An estimated 1.7 billion people worldwide show immunological sensitisation to *Mycobacterium tuberculosis* (MTB) but do not have disease.[1] Although most individuals who are sensitised to MTB will remain healthy for life, about 10% may progress to tuberculosis (TB) disease during their lifetime.[2, 3] Screening for symptoms compatible with TB, and microbiological testing of sputum samples, form the mainstay for diagnosis and treatment of TB.

There is an urgent need for biomarkers that can identify individuals with both subclinical and clinical disease. Clinical TB is disease that presents with symptoms compatible with TB, with either CXR abnormalities, or bacteriological confirmation or both. In contrast subclinical TB refers to disease that is asymptomatic, bacteriologically positive and with or CXR abnormalities compatible with TB.[4] As such biomarkers may be beneficial to guide confirmatory testing, or for screen-and-treat strategies to direct short-course TB preventive therapy (TPT) to persons at highest risk of progression to TB. There are several reported[5, 6] host blood transcriptomic TB signatures which include interferon (IFN) signalling genes that show promise as TB triage tests[7], and as tests for predicting progression to TB disease.[8]

SATVI developed a 16-gene transcriptomic signature of TB risk, from which an abbreviated 11-gene version (RISK11) has been transferred to a PCR platform. In case control studies, this signature predicted progression to TB disease in individuals showing sensitisation up to 12 months prior to TB diagnosis; and showed promising diagnostic performance for TB.[9-11] RISK11 showed excellent diagnostic performance for symptomatic TB and good prognostic performance for short-term prediction of incident TB among HIV-uninfected adults in a TB endemic setting, but RISK11-guided TPT

using Rifapentine and Isoniazid (3HP) given weekly for 3 months did not reduce progression to TB disease over 15 months.[12]

Viral infection is one factor that may affect host blood transcriptomic signature scores. Amongst individuals without TB disease, detectable HIV viral load has been associated with raised signature scores compared to undetectable viral load, possibly because of induction of type I IFN and raised expression of IFN-stimulated genes (ISG), which are included in RISK11 and other signatures.[11] Influenza is another viral infection that induces ISGs[13] and has been shown to affect transcriptomic signatures scores. However, the effect of other common viral and bacterial upper respiratory organisms on TB signatures remains largely unexplored. We evaluated the effect of upper respiratory tract organisms on RISK11 score and discriminatory performance of RISK11 for TB.

5.3 Methods

5.3.1 Study design

This nested cross-sectional sub-study was conducted between 09-Feb-2018 and 10-Sep-2018 in South Africa. The sub-study was nested in screening procedures for a clinical trial (CORTIS, ClinicalTrials.gov: NCT02735590).[12] The study was approved by the University of Cape Town Faculty of Health Sciences Human Research Ethics Committee (HREC REF. 327/2017). The study aimed to (i) identify respiratory organisms other than MTB associated with, and estimate their effect on, the RISK11 score and (ii) evaluate whether the RISK11 score differentiates participants with and without respiratory organisms or MTB.

5.3.2 Study population

Participants were adult healthy community volunteers aged between 18–60 years, living in and around the rural town of Worcester. Worcester is a TB endemic area, approximately 120 km from Cape Town in the Cape Winelands district. The district has a high TB incidence rate of approximately 880 per 100 000.[14] Participants were recruited during screening for the parent CORTIS study. Following written informed consent, participants eligible for the CORTIS main study, and for whom a PAXgene blood sample was obtained for the measurement of RISK11, were contemporaneously and consecutively enrolled, regardless of symptoms or signs of upper respiratory tract infection or TB.

5.3.3 Study procedures

Symptoms compatible with TB were recorded, including persistent unexplained cough for two weeks or more, persistent unexplained weight loss for two weeks or more, persistent unexplained fever for two weeks or more, persistent unexplained chest pains for two weeks or more, persistent unexplained night sweats for two weeks or more and haemoptysis in the last two weeks. Flu-like symptoms in the last two weeks, other than those compatible with TB, were also recorded.

RISK11 scores were computed from Ct values for each of the 11 genes, measured by microfluidic qRT-PCR as previously reported.[12] Briefly, RISK11 is a model of multiple transcript pairs, each functioning as a "vote" for or against TB risk; and the RISK11 score is the proportion of votes for TB risk. A score threshold can be set for the RISK11 assay to be utilised as a qualitative (positive/negative) test for TB risk. For analyses in this chapter, a score threshold of 60% was used to categorise a result as RISK11 positive (RISK11+). For detection of upper respiratory organisms, one nasopharyngeal and one oropharyngeal swab was collected using flocked Copan swabs (FLOQSwabs; COPAN Diagnostics Inc.) and kept in 1.5ml of Primestore buffer (Longhorn Vaccines and Diagnostics), a DNA preservation media for nucleic acid extraction, at -80 °C. The Qiasymphony Virus/Bacteria Mini Kit (Qiagen) was used to extract the nucleic acid from the swab samples. A multiplex real-time PCR qualitative assay for in vitro diagnostics, "FTD respiratory pathogens 33", (Fast Track Diagnostics, Luxembourg S.à.r.l.) was used to detect a broad panel of 33 respiratory pathogens, in accordance with the manufacturer's instructions, on the CFX96 Touch System light cycler platform (Bio-Rad) by the UCT-based Nicol Laboratory, on a fee-for-service basis.

A subset of participants was co-enrolled in the CORTIS parent study and were additionally investigated for prevalent TB disease at baseline and incident TB disease through 15 months. For this analysis, a TB endpoint was defined as one or more positive sputum samples microbiologically-confirmed by Xpert MTB/RIF, Xpert MTB/RIF Ultra, or MGIT culture. Prevalent TB disease was defined as disease diagnosed within 30 days of enrolment (baseline) and any TB disease diagnosed afterwards was classified as incident disease. Controls were defined as participants without prevalent or incident TB including those with an unknown outcome.

5.3.4 Statistical analysis

Statistical analyses were performed using STATA/IC version 16 (StataCorp. College Station, Texas). Histograms and box and whisker plots were used to explore numerical variables while frequency distributions were used to explore categorical variables. Descriptive statistics were computed as either

mean and standard deviation (SD), or median, interquartile range (IQR), minimum and maximum values for continuous variables, depending on the distribution. To test whether there is a significant difference in median RISK11 scores between participants with and without MTB, viruses, or bacteria, p-values were computed using the Wilcoxon rank sum test.

To quantify the relationship between RISK11 score and each organism, univariable generalised linear models were fitted using the STATA *glm* command. To estimate the adjusted effect of organisms on RISK11 score, a multivariable generalised linear model was employed. Because RISK11 score is a continuous percentage that ranges between 0–100%, which was scaled down to a continuous proportion ranging from 0–1 for modelling purposes, the logit link function, binomial distribution family and a robust error term were used in the models to ensure that predictions from the model also fall between the values 0 and 1 (0–100%).[15, 16] The outcome measure was the percent marginal effect (increase/decrease) on RISK11 score associated with each organism (STATA's *margins* command). The likelihood ratio test method was used to construct the multivariable model. Initially a model with only RISK11, without any explanatory variable (empty model), was fitted. Thereafter, nested models of RISK11 plus individual organisms were fitted and compared with the initial model to assess whether the variable being added made a significant contribution to the initial model. The Akaike's information criterion (AIC) value was used to evaluate the contribution, and the variable with the lowest AIC value was chosen. This process was repeated until no variable made a significant contribution to the model.

To test whether RISK11 score can differentiate between participants with and without upper respiratory organisms, ROC curves were used to compute the area under the curve (AUC). AUC discriminatory capacity was categorised as follows; of 0.5 = no discrimination, 0.5–0.7 = poor discrimination, 0.7–0.8 = acceptable discrimination, 0.8–0.9= excellent discrimination, and >0.9 = outstanding discrimination. In conducting the ROC curve analysis, participants with viruses only and those with both viruses and bacteria were combined to form the viral group, because they showed equivalent distribution of RISK11 scores. ROC curve analyses for evaluating RISK11 performance to discriminate between TB and the various organism categories were performed on the subset of participants co-enrolled in CORTIS who were also investigated for TB; while analyses evaluating RISK11 performance to discriminate between viruses and bacteria, viruses and no organisms, or bacteria and no organisms included all participants except those diagnosed with TB.

Enrollment into the parent CORTIS study was based on RISK11 status using a RISK11-positivity threshold of 60%. For purposes of conducting the study efficiently, about 79% of all eligible RISK11+ and only 13% of all eligible RISK11- participants were enrolled. Because of this enrichment of RISK11+ participants in the CORTIS co-enrolled population, probability weights of 1.263 and 7.920 were assigned to RISK11+ and RISK11- co-enrolled participants, respectively. Thus, analyses in the subset of participants co-enrolled into CORTIS were adjusted with probability weights to obtain estimates of the screened population. A 0.05 significance level was used for statistical significance in all analyses.

5.4 Results

5.4.1 Baseline characteristics

1000 participants with median age of 27 years (IQR 22–36) were enrolled in this sub-study. One participant was excluded from the analysis because of an indeterminate RISK11 result. 286 of the 999 remaining participants were co-enrolled in CORTIS and also underwent investigation for prevalent and incident TB at baseline and during follow-up, respectively (Figure 5.1).

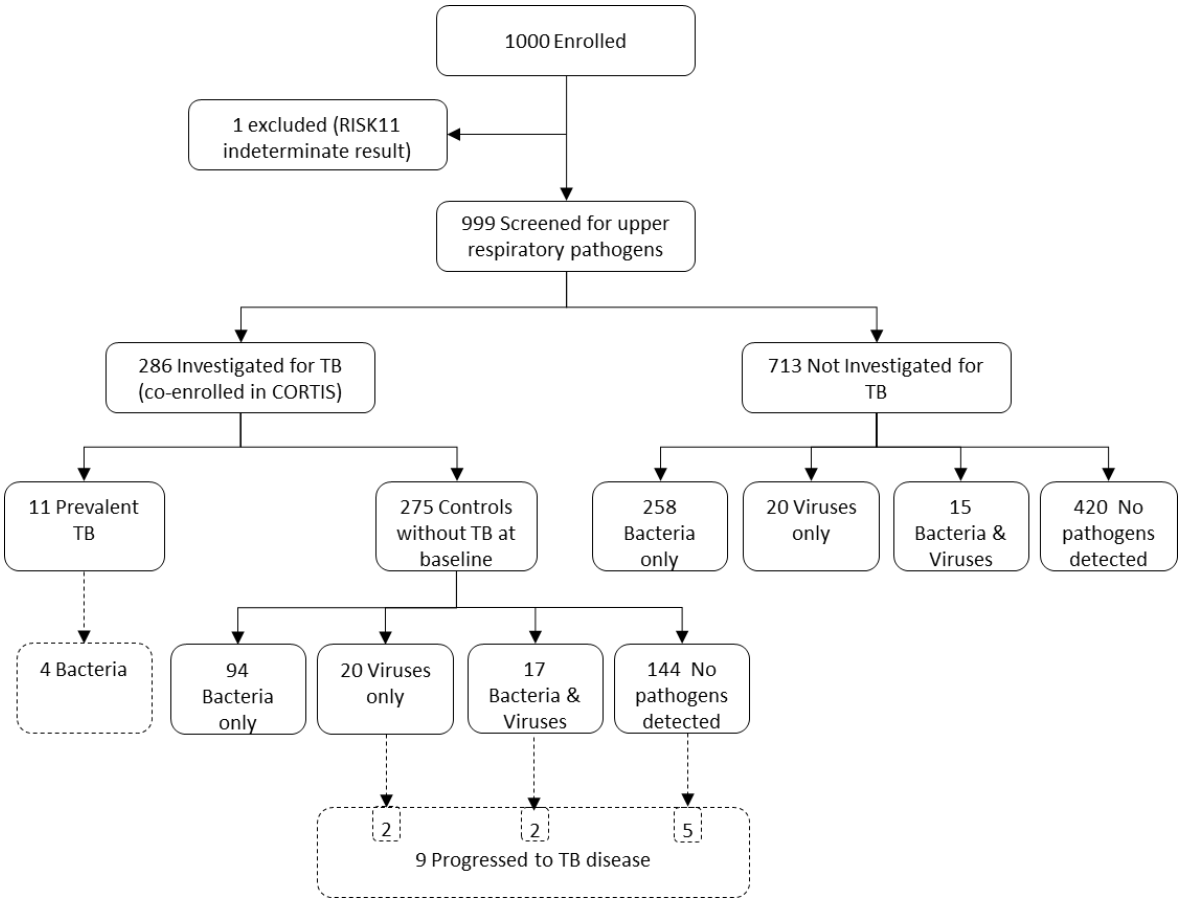


Figure 5.1: Study design. Of the 286 participants co-enrolled in the CORTIS parent study and investigated for TB, 11 had prevalent TB and 9 progressed to incident TB.

Participants investigated for TB had significantly lower body-mass index (BMI), but higher RISK11 scores and RISK11 positivity rate, compared to those not investigated for TB. No differences were observed in the distribution of sex, age, ethnicity, education level, employment status, smoking history, prior TB history and household TB contact history between those investigated and not investigated for TB (Table 5.1).

Table 5.1: Demographic characteristics of participants.

Variable name	All participants (N=999)	Participants investigated for TB (N=286)	Participants not investigated for TB (N=713)	P value[1]
Sex, n (%)				
Male	437 (43.7)	116 (40.6)	321 (45.0)	
Female	562 (56.3)	170 (59.4)	392 (55.0)	0.21
Median age (IQR)	27 (22–36)	27 (22–34)	28 (22–37)	0.28
Ethnicity, n (%)				
Caucasian	3 (0.3)	1 (0.4)	2 (0.3)	
Mixed ancestry	621 (62.2)	177 (61.9)	444 (62.3)	0.98
Black African	375 (37.5)	108 (37.8)	267 (37.5)	
Education, n (%)				
Primary or no school	74 (7.4)	18 (6.3)	56 (7.9)	
Secondary	921 (92.2)	267 (93.4)	654 (91.7)	0.76
Tertiary	4 (0.4)	1 (0.4)	3 (0.4)	
Employment status, n (%)				
Unemployed	786 (78.7)	234 (81.8)	552 (77.4)	
Casual employment	98 (9.8)	23 (8)	75 (10.5)	0.31
Formal employment	115 (11.5)	29 (10.1)	86 (12.1)	
Smoking history, n (%)	634 (63.5)	192 (67.1)	442 (62.0)	0.15
Prior TB, n (%)	114 (11.4)	39 (13.6)	75 (10.5)	0.17
TB household contacts, n (%)	106 (10.6)	37 (12.9)	69 (9.7)	0.14
IGRA result, n (%)				
Negative	58 (20.3)	58 (20.3)	NA[2]	NA
Positive	228 (79.7)	228 (79.7)	NA[2]	
Median BMI (IQR)	23.5 (20.1–29.6)	22.5 (19.9–28)	23.9 (20.2–30)	0.02
TB symptoms positive, n (%)	3 (0.3)	3 (1.1)	0	NA
Flu-like symptoms, n (%)	50 (5.0)	20 (7.0)	30 (4.2)	0.08
Median RISK11 score at enrolment (IQR)	14.7 (7.9–32.9)	29.4 (12–86.2)	12.6 (7.6–23.4)	<0.001
RISK11 positive, n (%)	148 (14.8)	123 (43.0)	25 (3.5)	<0.001

IQR, inter-quartile range. IGRA, interferon-gamma release assay. BMI, body-mass index.

[1]P-values were calculated using Wilcoxon Rank Sum test (continuous data) or Fisher's Exact Test (categorical data), comparing group investigated for TB versus group not investigated for TB.

[2]IGRA was not measured in participants who were not enrolled in the CORTIS parent study or not investigated for TB.

[3]All participants who were co-enrolled into the CORTIS trial were investigated for TB.

IQR, inter-quartile range. IGRA, interferon-gamma release assay. BMI, body-mass index.

5.4.2 Upper respiratory tract organisms

All 999 participants with a valid RISK11 result were screened for upper respiratory tract organisms and 42.9% (429/999) tested positive for at least one organism (Table 5.2). Bacterial and viral organisms were detected in 38.9% (389/999) and 7.2% (72/999) of the participants respectively, with bacterial-viral co-detection in 3.2% (32/999).

Table 5.2: Frequency of upper respiratory tract organisms detected by PCR and distribution of RISK11 scores in all participants.

Organism	Category	Frequency n (%)	Median RISK11 score (IQR)	P value[1]
Bacteria				
Any bacteria	No	610 (61.1)	14.7 (7.8-33.2)	
	Yes	389 (38.9)	14.3 (8.2-33.5)	0.74
Chlamydia pneumonia	No	998 (99.9)	14.7 (7.9-33.0)	
	Yes	1 (0.1)	3.9	0.1
Haemophilus influenza	No	786 (78.7)	14.5 (7.9-32.5)	
	Yes	213 (21.3)	15.2 (7.8-33.5)	0.86
Haemophilus influenza B	No	998 (99.9)	14.7 (7.9-33.0)	
	Yes	1 (0.1)	58.0	0.23
Klebsiella pneumonia	No	981 (98.2)	14.7 (7.8-33)	
	Yes	18 (1.8)	13.1 (8.1-33.4)	0.8
Legionella spp.	No	997 (99.8)	14.7 (7.9-33.0)	
	Yes	2 (0.2)	14.7 (14.7-14.7)	0.98
Moraxella catarrhalis	No	941 (94.2)	14.3 (7.8-32.5)	
	Yes	58 (5.8)	20.0 (10.0-42.9)	0.03
Mycoplasma pneumonia	No	998 (99.9)	14.7 (7.9-32.7)	
	Yes	1 (0.1)	98.3	0.09
Staphylococcus aureus	No	897 (89.8)	14.7 (7.9-33.0)	
	Yes	102 (10.2)	13.1 (7.8-32.5)	0.57
Streptococcus pneumonia	No	907 (90.9)	14.7 (7.9-32.6)	
	Yes	92 (9.1)	14.9 (8.2-37.2)	0.77
Viruses				
Any virus	No	927 (92.8)	14 (7.8-30.3)	
	Yes	72 (7.2)	45.7 (16.7-91.3)	<0.001
Influenza (types A, B and C)	No	995 (99.6)	14.7 (7.9-32.6)	
	Yes	4 (0.4)	89.1 (47.5-95.3)	0.07
Cytomegalovirus	No	998 (99.9)	14.7 (7.9-33.0)	
	Yes	1 (0.1)	4.8	0.13
Rhinovirus	No	955 (95.6)	14.3 (7.8-31.4)	
	Yes	44 (4.4)	37.2 (16.9-89.6)	0.001
Parainfluenza (types 1,2,3 and 4)	No	996 (99.7)	14.7 (7.9-32.5)	
	Yes	3 (0.3)	51.5 (7.8-97.0)	0.29
Coronavirus (NL63, 229E, OC43, and HKU1)	No	981 (98.2)	14.3 (7.9-32.5)	
	Yes	18 (1.8)	56.4 (16.9-93.9)	0.003
Respiratory syncytial virus (A or B)	No	995 (99.8)	14.7 (7.9-32.9)	
	Yes	2 (0.2)	50.2 (22.9-77.5)	0.17
Adenovirus	No	998 (99.9)	14.7 (7.9-32.7)	
	Yes	1 (0.1)	95.2	0.11
Mycobacteria				
Mycobacterium tuberculosis[2,3]	No	266 (93.0)	26.6 (11.7-84.8)	
	Prevalent	11 (3.8)	85.7 (37.2-97.4)	0.1
	Incident[4]	9 (3.2)	67.5 (28.6-79.6)	

[1]P values for comparing the distributions of RISK11 scores between those with and without specific organisms using the Wilcoxon rank sum test.

[2]Presence of *Mycobacterium tuberculosis* was ascertained through sputum Xpert MTB/RIF, Xpert Ultra, or liquid culture (mycobacterial growth inhibition assay).

[3]*Mycobacterium tuberculosis* was ascertained in 286 individuals co-enrolled in CORTIS.

[4]Follow up for incident tuberculosis was over a 15-month period. IQR, inter-quartile range.

Overall, RISK11 scores were significantly higher (p<0.05) in participants who tested positive for any virus (median=46.7%, IQR: 16.7%–93.9%) or a combination of viruses and bacteria (median=42.8%, IQR: 13.0%–89.2%) than participants with only bacteria (median=13.4%, IQR: 8.1%–9.0%), or who were negative for any organism. When stratified by TB investigation status, RISK11 scores were significantly higher (p<0.001) in participants investigated for TB compared to participants not investigated for TB, for all organism categories (Figures 5.2a & 5.2b). The median scores in investigated compared to uninvestigated participants, respectively, were as follows: viruses only, 77.5% (IQR: 23.7%–94.8%) versus 18.2% (IQR: 8.9%–51.3%; p=0.01); viruses and bacteria, 89.2% (IQR: 26.6%–94.8%) versus 19.1% (IQR: 7.8%–47.6%; p=0.01); Bacteria only, 20.7% (IQR: 9.1%–78.0%) versus 12.3% (IQR: 7.8%–22.1%; p<0.001); and negative for any organism, 24.6% (IQR: 11.3%–76.2%) versus 12.6% (IQR: 7.4%–22.5%; p<0.001).

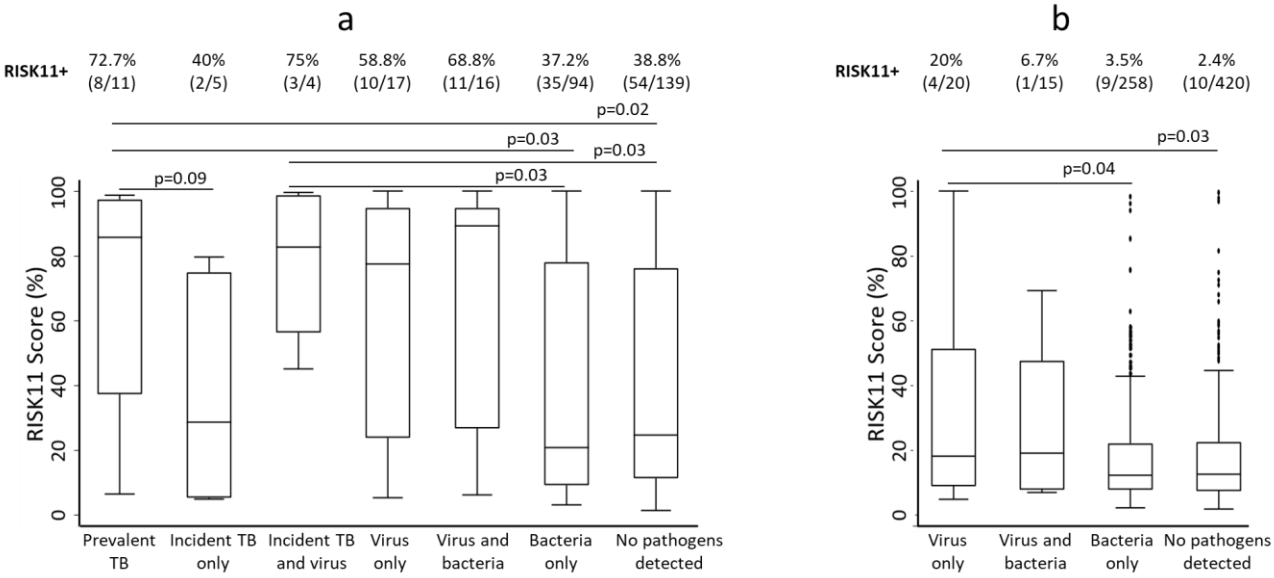


Figure 5.2: Viral or bacterial upper respiratory tract organisms, and uninfected individuals.

(a) Distributions of RISK11 scores in participants investigated for TB (a, n=286) and (b) participants not investigated for TB (n=713) Only P values below 0.1 are shown. Boxes represent the interquartile range, and horizontal lines in the boxes represent medians. The whiskers represent the lowest and highest RISK11 score within 1.5 times the interquartile range from the lower quantile and upper quantile, respectively. The proportions of RISK11+ participants using a threshold of 60% at each time point are indicated above each plot.

5.4.3 Symptoms

Only three of the 999 (0.3%) participants presented with at least one symptom compatible with TB, including night sweats (0.3%, 3/999), cough (0.2%, 2/999), and loss of weight (0.2%, 2/999). Median RISK11 scores were significantly higher (p=0.02) in symptomatic participants (65.4%, IQR: 45%–98.3%) compared to asymptomatic participants (14.6%, IQR: 7.9%–32.6%). Among the three symptomatic

participants, all of whom were investigated for TB, one participant with a chest radiograph (CXR) suggestive of TB was diagnosed with prevalent TB disease (microbiologically confirmed) and had the highest RISK11 score (98.7%).

Fifty of the 999 (5.0%) participants reported a flu-like illness at screening, out of which three (6.0%), five (10.0%) and 19 (38.0%) participants tested positive for viruses only, viruses and bacteria, and bacteria only, respectively. Median RISK11 scores were not significantly different (p=0.14) between participants with flu-like symptoms (20.5%, IQR: 8.7%–47.5%) and participants without flu-like symptoms (14.5%, IQR: 7.9%–32.5%).

5.4.4 TB disease

Among the 286 (28.6%) participants co-enrolled into the parent CORTIS study and investigated for TB at baseline, 11 (3.8%) had prevalent TB; and 9 of the remaining 275 subsequently developed incident TB.

Participants with prevalent TB had significantly higher (p=0.04) RISK11 scores (median=85.7%, IQR: 37.2%–97.4%) compared to controls (median=26.6%, IQR: 11.7%–84.9%). In contrast, RISK11 scores were not significantly different (p=0.46) between those who progressed to incident TB (median=67.5%, IQR: 28.6%–79.6%) and controls. Eight (72.7%) of the 11 prevalent TB cases and five (66.7%) of nine incident TB cases had a RISK11+ result. Furthermore, four (36.4%) of the prevalent and two (22.2%) of the incident TB cases had co-detection of upper respiratory bacterial organisms.

Viruses were not detected in participants with prevalent TB but were detected in four participants that progressed to incident TB. The proportion of individuals with upper respiratory viral organisms was significantly higher (p=0.02) in those who progressed to incident TB (44.4%, 4/9) than in those who remained healthy (12.4%, 33/266). Participants with a viral organism were five times more likely to progress to pulmonary TB than those without a viral organism (Incident Rate Ratio; IRR 5.0, 95% CI 1.0–23.2, Table 5.3).

Table 5.3: Association between upper respiratory viral organisms and risk of TB

Viral upper respiratory tract organism	At-Risk of Incident TB n=275	Incident TB Cases n=9	Observation Time (person-years)	Incidence Rate per 100 person-years (95% CI)	Cumulative Incidence (95% CI)	Incidence Rate Ratio (95% CI)	P-Value
Absent	238	5	2493.3	0.2 (0.08–0.48)	0.02 (0.01–0.05)	5.0 (1.0–23.2)	0.03
Present	37	4	399.9	1 (0.38–2.67)	0.11 (0.03–0.25)		

5.4.5 Effect of upper respiratory organisms on RISK11 scores

The role and effect magnitude of individual organisms on RISK11 score was investigated. A multivariable generalised linear model was fitted in the 286 participants investigated for both TB and upper respiratory tract organisms to assess their effect on RISK11 score, while controlling for prevalent TB (Table 5.4).

Three organisms, influenza, rhinoviruses, and coronaviruses, independently predicted RISK11 score after accounting for interaction (p<0.05). Independent of TB risk, RISK11 scores were predicted to be higher by 16.7% (95%CI 4.1%–29.4%), 67.8% (95%CI 52%–83.5%) and 13.5% (95%CI 3.5%–23.5%) in participants with coronaviruses, influenza, and rhinoviruses, respectively, compared to those in whom such viruses were not detected (Table 5.4 and Figure 5.3). An exploratory multiple generalised linear model in participants not investigated for TB showed similarly that the presence of viral rhinoviruses was significantly associated with higher marginal effects on RISK11 score (Figure 5.3b and Appendix 5.1).

Table 5.4: Univariable and multivariable generalised linear models of the effect of socio-demographic and upper respiratory organisms on RISK11 score in participants investigated for TB.

Variable	N=286	Univariable Analysis			Multivariable Analysis		
		β. Coefficient (95% CI)	% Marginal Effect (95% CI)	P	β. Coefficient (95% CI)	% Marginal Effect (95% CI)	P
Age, (median, IQR)	27 (22–34)	0 (-0.01–0.02)	0.1 (-0.1–0.3)	0.54	-	-	-
BMI (median, IQR)	22.5 (19.1–28)	-0.01 (-0.02–0.01)	-0.1 (-0.4–0.2)	0.51	-	-	-
Sex (male) (n, %)	116 (40.6)	-0.19 (-0.44–0.07)	-3.3 (-7.9–1.2)	0.15	-	-	-
Adenovirus (n, %)	1 (0.4)	4.17 (4.05–4.29)	75.2 (72.5–77.8)	<0.001	-	-	-
Coronavirus (NL63, 229E, OC43 & HKU1) (n, %)	10 (3.5)	0.86 (0.14–1.58)	15.5 (2.6–28.4)	0.02	0.95 (0.23–1.67)	16.7 (4.1–29.4)	0.01
Influenza (A, B C & H1N1) (n, %)	3 (1.1)	3.73 (2.84–4.63)	67.0 (51.0–83.0)	<0.001	3.84 (2.94–4.73)	67.8 (52–83.5)	<0.001
Rhinoviruses (n, %)	21 (7.3)	0.69 (0.13–1.26)	12.5 (2.3–22.6)	0.02	0.76 (0.2–1.33)	13.5 (3.5–23.5)	0.01
Prevalent TB (n, %)	11 (3.9)	0.72 (-0.23–1.66)	13.0 (-4.1–30.0)	0.14	0.81 (-0.13–1.76)	14.4 (-2.4–31.1)	0.09
Incident TB (n, %)	9 (3.1)	0.4 (-0.45–1.25)	7.2 (-8.2–22.6)	0.36	-	-	-
Haemophilus Influenzae (n, %)	65 (22.7)	-0.05 (-0.35–0.24)	-1.0 (-6.2–4.3)	0.72	-	-	-
Klebsiella Pneumoniae (n, %)	8 (2.8)	-0.54 (-1.24–0.17)	-9.7 (-22.4–3.0)	0.13	-	-	-
Moraxella Catarrhalis (n, %)	18 (6.3)	0.35 (-0.08–0.79)	6.4 (-1.4–14.1)	0.11	-	-	-
Parainfluenza (1–4) (n, %)	1 (0.4)	4.64 (4.52–4.76)	83.6 (80.4–86.8)	<0.001	-	-	-
Respiratory Syncytial Virus (A or B) (n, %)	2 (0.7)	0.35 (-0.51–1.2)	6.3 (-9.2–21.7)	0.43	-	-	-
Staphylococcus Aureus (n, %)	28 (9.8)	0.01 (-0.45–0.46)	0.5 (-45.3–46.3)	0.98	-	-	-
Streptococcus Pneumoniae (n, %)	32 (11.2)	-0.19 (-0.57–0.19)	-3.4 (-10.3–3.4)	0.33	-	-	-

IQR, inter-quartile range. BMI, body-mass index. The β coefficients can be exponentiated to obtain Odds Ratios of the association between each predictor variable and RISK11 score.

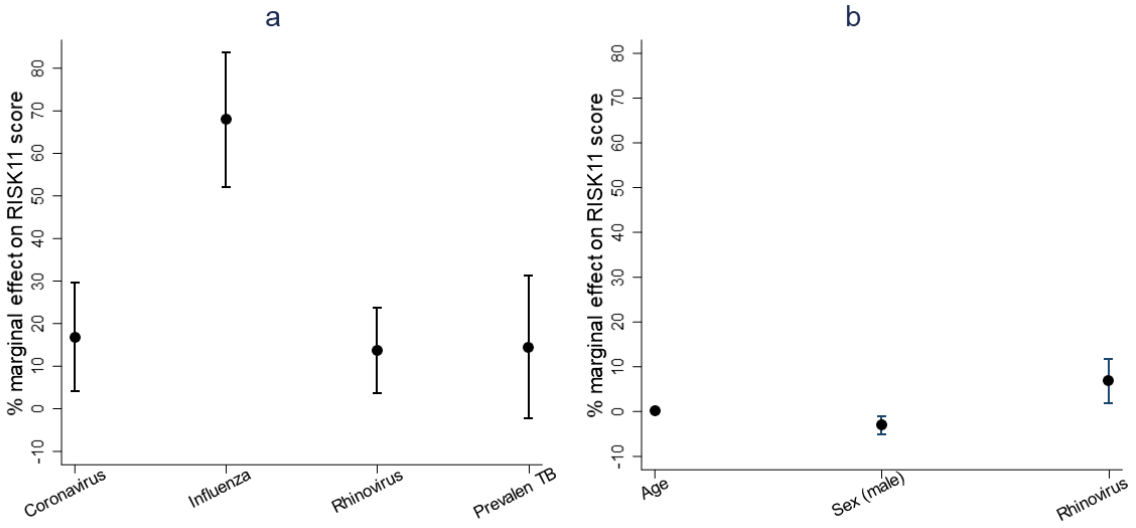


Figure 5.3: Predicted marginal effects on RISK11 of the different upper respiratory organisms and socio-demographic factors in (a) participants investigated for TB and (b) participants not investigated for TB.

Dots represent the point estimate of percent increase/decrease in RISK11 score attributable to each organism or host factor and whiskers represent the 95% confidence interval for the point estimate.

5.4.6 RISK11 performance for differentiating between individuals with and without prevalent TB and upper respiratory tract organisms.

Amongst the 286 participants investigated for TB and upper respiratory tract organisms, the discriminatory capacity of RISK11 to differentiate individuals with and without various organisms was evaluated. RISK11 moderately differentiated prevalent TB (n=11) from bacterial organisms (n=94) or no organisms (n=139) with AUCs of 0.70 (95%CI 0.51–0.86) and 0.70 (95%CI 0.51–0.88), respectively (Figures 5.4a and 5.4b). However, RISK11 did not discriminate prevalent TB from viral organisms (n=37, AUC 0.48; 95%CI 0.28–0.70; Figure 5.4c).

Amongst the 979 participants without prevalent or incident TB, RISK11 discriminated between 68 participants with viral organisms and 352 participants with only bacterial organisms with an AUC of 0.71 (95%CI 0.64–0.78; Figure 5.4d), and the 559 participants without organisms with an AUC of 0.71 (95%CI 0.64–0.78; Figure 5.4e). RISK11 did not distinguish between individuals with bacteria only and individuals with no organisms (AUC 0.51, 95 CI 0.47–0.54; Figure 5.4f).

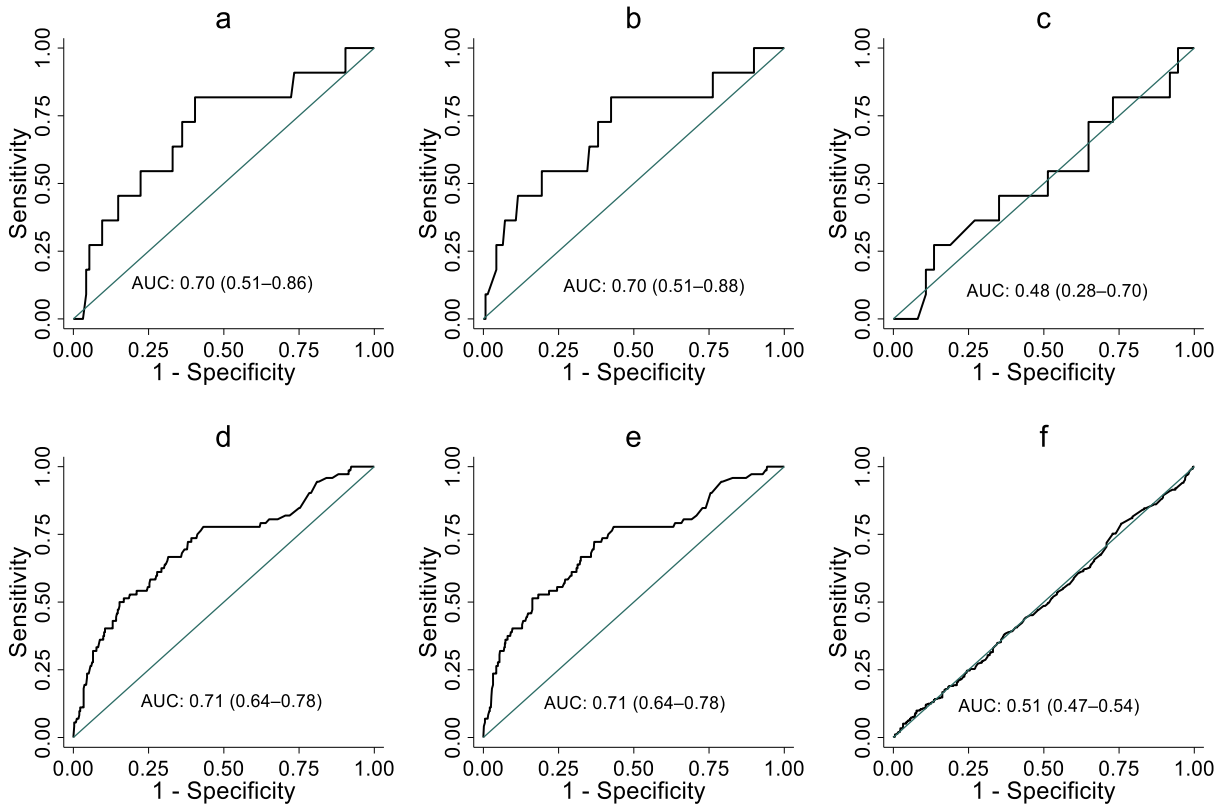


Figure 5.4: Diagnostic performance of RISK11 for differentiating between participants with TB, viral or bacterial upper respiratory tract organisms, and uninfected individuals.

Performance of RISK11 in differentiating between participants with TB and participants with a bacterial upper respiratory organism (a), between participants with TB and uninfected participants (b), between participants with TB and participants with a viral upper respiratory organism (c), between participants with a viral upper respiratory organism and participants with a bacterial upper respiratory organism (d), between participants with a viral upper respiratory organism and uninfected participants (e), and between participants with a bacterial organism and uninfected participants (f). AUC=area under the curve. Numbers in brackets represent the 95% confidence interval (CI) of the AUC.

5.5 Discussion

This cross-sectional study aimed to identify upper respiratory organisms other than MTB that might also associate with RISK11 score, and to estimate their effect on RISK11 score readout. Further, the study aimed to evaluate whether the RISK11 score differentiates between individuals with and without respiratory organisms and MTB.

The results indicate that several upper respiratory viruses, namely, coronaviruses, influenza and rhinoviruses affected RISK11 score. The presence of upper respiratory tract viruses resulted in elevated RISK11 scores and diminished RISK11 discriminatory performance, such that RISK11 could

not differentiate TB from viruses. Controlling for MTB, participants with these respiratory viral organisms had elevated RISK11 scores, predicted to be higher by 12.5%–67.0%, compared to participants without the corresponding respiratory viral organisms. These results suggest that intercurrent respiratory viral infections, such as coronavirus, and influenza and rhinovirus, drive signature conversion and transient false-positive TB risk status.

Elevated RISK11 scores in individuals with viral upper respiratory organisms are likely due to an IFN-inducible gene profile, that may consist of both IFN-y and type I IFN-αβ signalling.[17] Both TB and respiratory viral infection can lead to overlapping symptoms, including cough, and increasing severity of both conditions is associated with an increase in cough – so it is not surprising that higher IFN-y responses are associated with more symptoms. RISK11 and other transcriptomic signatures of TB that include IFN signalling genes may give false positive results in participants with infection or colonization with upper respiratory tract viruses, which will have a significant negative impact on the specificity of such diagnostics, as shown by the inability of RISK11 to differentiate TB from viruses. This finding shows the importance of confirmatory testing and/or combined TB and viral screening, such as that implemented in some TB endemic settings for COVID-19. Alternatively, there is need to develop and validate signatures that are less affected by ISG modulation, such as those proposed by Singhania *et al.*, and Esmail *et al*.[13, 18].

Although upper respiratory tract viral organisms were shown to affect transcriptomic signatures in controls without TB, ie. independent of TB risk, a vital question is whether upper respiratory tract viral organisms might also induce changes in immune control of MTB infection and trigger progression to TB disease.[19] This putative effect would be analogous to that of HIV infection, which has both TB-independent effects on signature score in controls without TB, and TB-dependent effects on signature score as a consequence of increased risk of progression to TB disease. Preliminary evidence of a possible association between upper respiratory viral organisms and risk of incident TB is presented here, supporting the hypothesis that respiratory viral infections may trigger progression to TB disease, or that immune dysfunction increases susceptibility to both viral and MTB infection. Participants with respiratory viral organisms were five times more likely to develop TB compared to participants without viral organisms. Although this result anchors on small numbers, thus limiting the significance and generalisability, the finding that viral co-infection may be associated with an elevated risk of progression to TB requires rigorous testing in future studies. Prior studies have previously reported that viral respiratory co-infections are associated with quicker progression or more severe TB disease.[20, 21] TB patients with nasopharyngeal viral-bacterial co-infection were also likely to have more

severe TB disease[22], while murine influenza infection led to impaired control of MTB infection via a Type I IFN-dependent mechanism.[23] However, a study in a Cape Town paediatric cohort did not find an association between upper respiratory organisms and pulmonary TB.[24]

This study had several limitations. First, the study included too few participants that progressed to incident TB disease to allow a robust evaluation of the association between upper respiratory viral infection and progression to TB disease. Second, only 28.6% of participants, who were co-enrolled in the parent CORTIS study, underwent TB investigation. Thus, TB disease could not be excluded in the remaining 71.4%. Further, alternative diagnoses were not sought in symptomatic participants without detected respiratory tract organisms or TB disease; nor were lower respiratory or gastrointestinal viruses sought, which might also modulate RISK11 scores. Therefore, it is hypothesised that the variation in RISK11 score which is not explained by the model might be explained by other viruses and host factors that were not measured. Finally, it is acknowledged that several of the upper respiratory organisms observed are not typically pathogenic and occur as commensals.

The study had several design characteristics that strengthen these results. First, participants were enrolled consecutively as they presented for screening and enrolment into the parent study, thereby minimising selection bias. Second, this study used both oropharyngeal and nasopharyngeal samples, thus boosting the detection rate of organisms. Third, this may be the first study to investigate perturbation of an ISG-based signature by common upper respiratory viral organisms in individuals at risk for incident TB disease.

5.6 Conclusion

This study provides insight into the association between common upper respiratory viruses and one of several published TB signatures[7, 8, 25]; highlighting the challenge that coronaviruses, influenza and rhinoviruses significantly increase signature scores, such that RISK11 cannot differentiate between viruses and TB, which may hinder implementation of such biomarkers as new tools for TB control. Therefore, accounting for confounding factors associated with raised host blood transcriptomic signature scores, including viral infection, or parallel investigation for TB and viral infection, may be important if these biomarkers are to be implemented as TB tests. It is not yet known to what degree these results are generalisable to other host blood TB transcriptomic signatures, a question that needs to be addressed.

5.7 References

1 Houben RM, Dodd PJ. The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. *PLoS Med.* 2016;13(10):e1002152.

2 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.

3 Andrews JR, Noubary F, Walensky RP, Cerda R, Losina E, Horsburgh CR. Risk of progression to active tuberculosis following reinfection with Mycobacterium tuberculosis. *Clin Infect Dis*. 2012;54(6):784-91.

4 Esmail H, Macpherson L, Coussens AK, Houben R. Mind the gap - Managing tuberculosis across the disease spectrum. *EBioMedicine*. 2022;78:103928.

5 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.

6 MacLean E, Broger T, Yerlikaya S, Fernandez-Carballo BL, Pai M, Denkinger CM. A systematic review of biomarkers to detect active tuberculosis. *Nat Microbiol*. 2019;4(5):748-58.

7 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. *The Lancet Respiratory medicine*. 2020;8(4):407-19.

8 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.

9 Zak DE, Penn-Nicholson A, Scriba TJ*, et al.* A blood RNA signature for tuberculosis disease risk: a prospective cohort study. *Lancet (London, England)*. 2016;387(10035):2312-22.

10 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*. 2018;108:124-6.

11 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. *Frontiers in microbiology*. 2019;10:1441.

12 Scriba TJ, Fiore-Gartland A, Penn-Nicholson A*, et al.* Biomarker-guided tuberculosis preventive therapy (CORTIS): a randomised controlled trial. *The Lancet Infectious diseases*. 2021.

13 Singhania A, Verma R, Graham CM*, et al.* A modular transcriptional signature identifies phenotypic heterogeneity of human tuberculosis infection. *Nature communications*. 2018;9(1):2308.

14 TBFacts.Org. TB Statistics South Africa 2021 [updated Sep-2021. Available from: https://tbfacts.org/tb-statistics-south-africa/.

15 Baum CF. Stata tip 63: Modeling proportions. *Stata Journal*. 2008;8(2):299-303.

16 STATA. Fractional response regression Geneva, Switzerland: StataCorp., College Station, TX, USA; 2015 [Available from: https://www.stata.com/manuals14/rfracreg.pdf.

17 Berry MPR, Graham CM, McNab FW*, et al.* An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature*. 2010;466(7309):973-U98.

18 Esmail H, Lai RP, Lesosky M*, et al.* Complement pathway gene activation and rising circulating immune complexes characterize early disease in HIV-associated tuberculosis. *Proceedings of the National Academy of Sciences of the United States of America*. 2018;115(5):E964-E73.

19 Esmail H, Barry CE, 3rd, Young DB, Wilkinson RJ. The ongoing challenge of latent tuberculosis. *Philos Trans R Soc Lond B Biol Sci*. 2014;369(1645):20130437.

20 Dangor Z, Izu A, Moore DP*, et al.* Temporal association in hospitalizations for tuberculosis, invasive pneumococcal disease and influenza virus illness in South African children. *PloS one*. 2014;9(3):e91464.

21 Walaza S, Cohen C, Tempia S*, et al.* Influenza and tuberculosis co-infection: A systematic review. *Influenza Other Respir Viruses*. 2020;14(1):77-91.

22 Mhimbira F, Hiza H, Mbuba E*, et al.* Prevalence and clinical significance of respiratory viruses and bacteria detected in tuberculosis patients compared to household contact controls in Tanzania: a cohort study. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*. 2019;25(1):107 e1- e7.

23 Redford PS, Mayer-Barber KD, McNab FW*, et al.* Influenza A virus impairs control of Mycobacterium tuberculosis coinfection through a type I interferon receptor-dependent pathway. *The Journal of infectious diseases*. 2014;209(2):270-4.

24 Dube FS, Kaba M, Robberts FJ*, et al.* Respiratory microbes present in the nasopharynx of children hospitalised with suspected pulmonary tuberculosis in Cape Town, South Africa. *BMC infectious diseases*. 2016;16(1):597.

25 Warsinske H, Vashisht R, Khatri P. Host-response-based gene signatures for tuberculosis diagnosis: A systematic comparison of 16 signatures. *PLoS medicine*. 2019;16(4):e1002786.

Chapter 6

6 Evaluation of a transcriptomic signature of tuberculosis risk in combination with an interferon gamma release assay: a diagnostic test accuracy study

Chapter overview

Current evidence suggests that host blood transcriptomic signatures may be valuable for identifying prevalent symptomatic TB and individuals at highest risk of progressing to disease. Approaches to further improve diagnostic and prognostic performance, especially for subclinical disease, are needed. Studies have shown that improvements in performance may be attained by using tests in combination. It is not yet known whether using a host blood transcriptomic signature in combination with an interferon gamma release assay (IGRA), such as QuantiFERON Gold-plus (QFTPlus), might improve discriminatory performance for both subclinical and clinical TB disease. Therefore, this chapter analyses the use of RISK11 and QFTPlus in combination as diagnostic and prognostic tests of TB risk.

Mulenga H, Fiore-Gartland A, Mendelsohn SC, Penn-Nicholson A, Mbandi SK, Nemes M, Borate B, Musvosvi M, Tameris M, Walzl G, Naidoo K, Churchyard G, Scriba TJ, and Hatherill M. *Evaluation of a transcriptomic signature of tuberculosis risk in combination with an interferon gamma release assay: a diagnostic test accuracy study*. EclinicalMedicine. (Accepted)

Chapter contribution to the thesis

This chapter addresses Aim 4 of the thesis. Prior to this analysis, no data were available on evaluation of the effect of using IGRA in combination with RISK11 on diagnostic and prognostic test utility. However, few data were available on using a combination of other biomarkers to improve diagnostic and/or prognostic test utility.

Contributions of the candidate

The candidate provided data management and operational support for the study, analysed, and interpreted the data, and wrote the manuscript with editorial input and guidance from his supervisors (TJS and MH). Co-authors were responsible for recruitment of participants, clinical management, and clinical data collection at study sites (MT, GW, KN, and GC); and operational or laboratory support and project management (HM, SKM, AP-N, MM). AF-G provided statistical support. All co-authors

reviewed the final draft of the manuscript (HM, AF-G, SCM, AP-N, SKM, BB, MM, MT, GW, KN, GC, TJS, and MH).

Publication

Evaluation of a transcriptomic signature of tuberculosis risk in combination with an interferon gamma release assay: a diagnostic test accuracy study

Humphrey Mulenga,[a] Andrew Fiore-Gartland,[b] Simon C. Mendelsohn,[a] Adam Penn-Nicholson,[a] Stanley Kimbung Mbandi,[a] Elisa Nemes, [a] Bhavesh Borate,[b] Munyaradzi Musvosvi,[a] Michèle Tameris,[a] Gerhard Walzl,[c] Kogieleum Naidoo,[d,e] Gavin Churchyard,[f,g,h] Thomas J. Scriba,[a] Mark Hatherill,[a*] and the CORTIS Study Team.

[a] South African Tuberculosis Vaccine Initiative, Institute of Infectious Disease and Molecular Medicine, Division of Immunology, Department of Pathology, University of Cape Town, Anzio Road, Observatory, 7925, South Africa.

[b] Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Fairview Ave. N., Seattle, WA 98109-1024, USA.

[c] DST/NRF Centre of Excellence for Biomedical TB Research and SAMRC Centre for TB Research, Division of Molecular Biology and Human Genetics, Department of Biomedical Sciences, Faculty of Medicine and Health Sciences, Stellenbosch University, Francie Van Zijl Dr, Parow, 7505, South Africa.

[d] Centre for the AIDS Programme of Research in South Africa (CAPRISA), Durban, Doris Duke Medical Research Institute, University of KwaZulu-Natal, 719 Umbilo Road, Durban 4001, South Africa.

[e] MRC-CAPRISA HIV-TB Pathogenesis and Treatment Research Unit, Doris Duke Medical Research Institute, University of KwaZulu-Natal, 719 Umbilo Road, Durban 4001, South Africa.

[f] The Aurum Institute, 29 Queens Rd, Parktown, Johannesburg, 2194, Gauteng, South Africa.

[g] School of Public Health, University of Witwatersrand, 27 St Andrews Road, Parktown 2193, Johannesburg, South Africa.

[h] Department of Medicine, Vanderbilt University, Nashville, TN, USA

*Study team members are listed in the supplementary appendix

Corresponding Author: Professor 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, Anzio Road, Observatory, 7925, South Africa.
Email mark.hatherill@uct.ac.za; Telephone +27214066791.

6.1 Summary

Background

We evaluated the diagnostic and prognostic performance of a transcriptomic signature of tuberculosis (TB) risk (RISK11) and QuantiFERON-TB Gold-plus (QFTPlus) as combination biomarkers of TB risk.

Methods

Healthy South Africans who were HIV-negative aged 18–60 years with baseline RISK11 and QFTPlus results were evaluated in a prospective cohort. Prevalence and incidence-rate ratios were used to evaluate risk of TB. Positive (LR+) and negative (LR-) likelihood ratios were used to compare individual tests versus Both-Positive (RISK11+/QFTPlus+) and Either-Positive (RISK11+ or QFTPlus+) combinations.

Findings

Among 2912 participants, prevalent TB in RISK11+/QFTPlus+ participants was 13.3-fold (95% CI 4.2–42.7) higher than RISK11-/QFTPlus-; 2.4–fold (95% CI 1.2–4.8) higher than RISK11+/QFTPlus-; and 4.5-fold (95% CI 2.5–8.0) higher than RISK11-/QFTPlus+ participants. Risk of incident TB in RISK11+/QFTPlus+ participants was 8.3-fold (95% CI 2.5–27.0) higher than RISK11-/QFTPlus-; 2.5-fold (95% CI 1.0–6.6) higher than RISK11+/QFTPlus-; and 2.1-fold (95% CI 1.2–3.4) higher than RISK11-/QFTPlus+ participants, respectively. Compared to QFTPlus, the Both-Positive test combination increased diagnostic LR+ from 1.3 (95% CI 1.2–1.5) to 4.7 (95% CI 3.2–7.0), and prognostic LR+ from 1.4 (95% CI 1.2–1.5) to 2.8 (95% CI 1.5–5.1), but did not improve upon RISK11 alone. Compared to RISK11, the Either-Positive test combination decreased diagnostic LR- from 0.7 (95% CI 0.6–0.9) to 0.3 (95% CI 0.2–0.6), and prognostic LR- from 0.9 (95% CI 0.8–1.0) to 0.3 (0.1–0.7), but did not improve upon QFTPlus alone.

Interpretation

Combining two tests such as RISK11 and QFTPlus, with discordant individual performance characteristics does not improve overall discriminatory performance, relative to the individual tests.

Funding

Bill and Melinda Gates Foundation, South African Medical Research Council.

Keywords

Mycobacterium tuberculosis, transcriptomic, signature, QuantiFERON, combination, performance.

Abstract word count = 275; **Manuscript body word count** = 5128

6.2 Research in context

Evidence before this study

It is not known whether using a host blood transcriptomic signature in combination with an interferon gamma release assay, such as QuantiFERON-TB Gold-plus (QFTPlus), might improve discriminatory performance for both subclinical and clinical TB disease.

Medline through PUBMED was searched for studies published between Jan 1, 2005, and Dec 1, 2021, without language restrictions, that evaluated tests in combination to improve diagnostic and/or prognostic performance using the following search term: ((((((Combining) AND (tests)) AND (improve)) AND (diagnostic OR prognostic)) AND (performance OR accuracy)) AND (tuberculosis)) AND (("2005/01/01"[Date - Publication] : "2021/12/01"[Date - Publication])) . The search returned 265 articles, out of which 16 were identified to have combined tests in the diagnosis of all forms of TB disease and showed improved sensitivity or specificity or both in two studies. No study combined a transcriptomic signature with another test for either TB diagnosis or prognosis.

Added value of this study

In a large prospective study of HIV-negative individuals in a TB-endemic setting, individuals with a double-positive RISK11+/QFTPlus+ result are at 13- and 8-times higher risk of prevalent and incident TB disease, respectively, compared to RISK11-/QFTPlus- individuals. However, no simultaneous improvement in the positive and negative likelihood ratios (LR+ and LR-) of the combined test was observed relative to individual tests. The Both-Positive test combination increased diagnostic LR+ and prognostic LR+, compared to QFTPlus, but did not improve upon RISK11 alone. Conversely, the Either-Positive test combination decreased diagnostic and prognostic LR-, compared to RISK11, but did not improve upon QFTPlus alone and the expected increase in False Positive results outweighed the benefit of identifying few additional True Positives.

Implications of all the available evidence

The findings suggest that combining two tests such as RISK11 and QFTPlus, with discordant individual performance characteristics (high sensitivity/low specificity and low sensitivity/high specificity), does not improve overall discriminatory performance, since there is no simultaneous improvement in the LR+ and LR- relative to the individual tests. The inadequate sensitivity of the Both-Positive and

inadequate specificity of the Either-Positive approaches would preclude RISK11/QFTPlus combination tests from use in a generic screening strategy. However, the expected increase in True Negative results with few additional False Negatives suggests the Both-Positive approach might have benefit as a rule-in RISK11/QFTPlus combination test.

6.3 Introduction

Approximately one quarter of the global population shows immunological sensitisation to *Mycobacterium tuberculosis (Mtb)* and may be at risk of TB disease, highlighting the importance of diagnosis and treatment of those who are more likely to progress to TB disease for TB elimination.[1] Historically, tuberculin skin tests (TSTs) have been used to identify those with immunological sensitisation to *Mtb*. Interferon gamma (IFN-γ) release assays (IGRA) were developed to improve specificity in BCG-vaccinated populations and are thought to have better predictive ability for incident TB disease than TST.[2, 3] The QuantiFERON-TB Gold-Plus (QFTPlus, Qiagen, Hilden, Germany) is one such commercially available IGRA.

The World Health Organization (WHO) has called for development of rapid non-sputum biomarker-based diagnostic and triage tests; and prognostic (incipient TB) tests that can predict progression from *Mtb* infection to incident TB. Performance specifications for these tests are stipulated in the target product profile (TPP) document and specifies minimum 90% sensitivity and 70% specificity for a triage test; 65% sensitivity and 98% specificity for a diagnostic test; and 75% sensitivity and 75% specificity for a prognostic test that can predict progression to TB disease within 2 years.[4, 5] In response, many such tests have been developed.[6] Some of the promising tests include host blood transcriptomic signatures, which may have multiple uses as diagnostic, triage, and prognostic tests of TB disease.[7] We previously developed and validated RISK11, a transcriptomic signature of TB risk based on mRNA expression of 11 interferon-stimulated genes with diagnostic sensitivity of 34.9% and specificity of 91.0% for prevalent TB disease; and prognostic sensitivity of 25.0% and specificity of 91.1% for progression to incident disease 15 months before onset, in individuals who were HIV-uninfected.[8-10]

Improvements in diagnostic and prognostic performance may be achieved by using tests in combination.[11-13] Common rules for combining diagnostic tests are 'Either-Positive' (combined test is positive if either individual test is positive) and 'Both-Positive' (combined test is positive if both individual tests are positive). Performance of individual diagnostic tests is usually compared with sensitivity and specificity. However, using sensitivity and specificity to compare performance of a

combined test to that of individual tests is problematic, because of the unavoidable trade-off between these measures.[14, 15] For example, when an Either-Positive combination is used, sensitivity of the combined test will be greater, and specificity will be less, than that of the individual tests. When the Both-Positive combination is used, the opposite applies.[16] Positive (LR+) and Negative (LR-) likelihood ratios provide a way to account for the trade-off in sensitivity and specificity, which is helpful for comparing individual versus combination tests.[17, 18] In this instance, the LR+ is the probability of an individual with TB testing positive divided by the probability of an individual without TB testing positive, given by the formula LR+ = sensitivity/(1–specificity); the LR- is similar, but in reference to a negative test and is given by LR- = (1–sensitivity)/specificity. If likelihood ratios fail to provide a clear choice between the individual and combined tests, then the trade-off in the expected number of extra true- and false-positives (TP or FP) or true- and false-negatives (TN or FN) may be used to decide whether tests should be used in combination.

It is not known whether using a transcriptomic signature such as RISK11 and IGRA in combination would increase diagnostic or prognostic performance and improve the utility of these tests for rule-in or rule-out clinical scenarios in which risk of TB is suspected. This analysis aimed to estimate the probability of prevalent and risk of incident TB, and diagnostic and prognostic performance, using a combination of RISK11 and QFTPlus results.

6.4 Methods

6.4.1 Study design and participants

We analysed data from a multi-center, randomised, partially blinded, clinical trial (CORTIS) conducted between Sept 20, 2016 and Dec 20, 2019 in South Africa. Study methods and main results were reported previously.[10] Briefly, the study assessed the diagnostic and prognostic discriminatory performance of RISK11 for TB disease, and treatment efficacy of high dose isoniazid and rifapentine (3HP). It was designed to have 90% power to reject the null hypothesis of a RISK11+ and RISK11- cumulative risk ratio less than 2 with one-sided alpha of 0.025. For treatment efficacy, there was 80% power to reject the null hypothesis of efficacy less than 20%, with one-sided alpha of 0.05. HIV-uninfected adults from five TB endemic sites aged between 18 and 60 years, without prior TB disease in the preceding 3 years or other co-morbidities (known diabetes mellitus, liver disease, porphyria, peripheral neuropathy, epilepsy, psychosis, or alcoholism), underwent simultaneous RISK11 and QFTPlus testing at baseline. Participants were screened for HIV using the Determine HIV-1/2 (Abbot Laboratories, Germany) and Uni-Gold Recombigen HIV-1/2 (Trinity Biotech PLC, Ireland) tests. The

QFTPlus was interpreted according to manufacturer's instructions; with a positive QFTPlus (QFTPlus+) defined as either a TB1 minus Nil or TB2 minus Nil IFN-γ result of ≥0.35 IU/mL and ≥25% of Nil. RISK11 scores were computed from the quantification cycle (Cq) values for each of the 11 genes measured by microfluidic qRT-PCR as reported previously.[9, 10] In brief, RISK11 is a model of multiple transcript pairs, each functioning as a vote for TB risk. The RISK11 score ranges from 0 to 100% and is the continuous proportion of positive transcript pair votes for TB risk. A positivity threshold for the score can be set for the RISK11 assay to be used as a qualitative (positive/negative) test for TB risk. A positive RISK11 (RISK11+) result was predefined as a RISK11 score of 60% for the main analysis; and as RISK11 score of 26%, which was deemed the optimal cut-off, for the sensitivity analyses.

All participants were screened for prevalent TB at baseline; those without prevalent TB were followed for a median of 15 months for incident TB disease. Participant evaluation for incident TB was symptom-triggered at each of six scheduled visits (months 1, 2, 3, 6, 9, and 12) and the symptoms were actively asked for using a symptom questionnaire. All participants were evaluated at the final visit, month 15, regardless of symptom status. Standardized evaluation of suspected TB disease included symptom history, TB contact history, and sputum collection for Xpert MTB/RIF (Cepheid, Franklin Lakes, NJ) for prevalent TB; and symptom history, TB contact history, and sputum collection for liquid mycobacterial culture (Mycobacteria Growth Indicator Tube, MGIT, Becton-Dickinson, USA), and Xpert MTB/RIF or Xpert MTB/RIF Ultra (Cepheid, Franklin Lakes, NJ) for incident TB. TB cases diagnosed within 30 days of enrolment (baseline) were classified as prevalent and those diagnosed after 30 days were classified as incident. Participants without a prevalent or incident TB diagnosis, including those with an unknown outcome at the end of study due to withdrawal or lost to follow-up, were classified as controls because excluding them did not change the prognostic performance measures. Participants presenting with any one or more symptoms of persistent unexplained cough, weight loss, chest pains, night sweats, or fever for two weeks or more, or any haemoptysis were defined as symptomatic. Sputum samples were all spontaneously expectorated. In this analysis, the microbiologically-confirmed TB disease endpoint was defined as a positive Xpert MTB/RIF, Xpert MTB/RIF Ultra, or MGIT culture on one or more sputum samples.

6.4.2 Statistical analysis

Statistical analyses were performed using STATA/IC version 16.1 (StataCorp. College Station, TX, USA) and R version 3.6.3 (R Foundation, Vienna, Austria). A significance level (α) of 0.05 was used for all analyses. Only participants with valid QFTPlus and RISK11 results were included in the analyses. The median and interquartile range (IQR); and proportions were used as descriptive statistics for continuous and categorical variables, respectively.

The Fisher's exact test was used to compare categorical variables. The Wilcoxon rank-sum and Kruskal-Wallis tests were used to compare continuous variables among two, or more than two groups, respectively. Test agreement for qualitative RISK11 and QFTPlus results was evaluated by Cohen's kappa (κ) coefficient and proportion of concordant results. Spearman's correlation coefficient (ρ) was used to assess the relationship between continuous RISK11 and QFTPlus readouts.

To evaluate the probability of prevalent or risk of incident TB for a double-positive RISK11+ and QFTPlus+ result compared to other risk categories, each participant was grouped in one of the four risk categories as follows: RISK11+/QFTPlus+, RISK11+/QFTPlus-, RISK11-/QFTPlus+, and RISK11-/QFTPlus-. The prevalence ratio (PR) was used to evaluate probability of prevalent TB and included all participants. The incidence-rate ratio (IRR) was used to evaluate the risk of incident TB; and excluded participants with prevalent TB and those who did not attend any visit after enrolment. RISK11+ participants that received the CORTIS trial intervention (3HP) were not excluded from this analysis, since provision of 3HP to RISK11+ participants did not affect rate of progression to TB disease over 15 months.[10]

To evaluate the diagnostic and prognostic performance of a combination RISK11/QFTPlus test compared to individual tests, likelihood ratios were computed. Confidence intervals for likelihood ratios were computed following the method of Koopman.[19] To combine the tests, two rules were used: 'Either-Positive' (positive combination test = +/- or -/+ or +/+; negative combination test = -/-) and 'Both-Positive' (positive combination test = +/+; negative combination test = -/- or +/- or -/+). For the main analysis, tests were combined at the prespecified RISK11 positivity threshold of 60% and the manufacturer positivity threshold of 0.35 IU/mL for QFTPlus. Thereafter, in sensitivity analysis, tests were combined at optimal thresholds of each test, computed with the Youden Index method.

Enrolment into the parent study (CORTIS) was based on RISK11 status. The enrolled population was enriched with RISK11+ participants compared to the screened population by design. Approximately 79% of all eligible RISK11+ participants and 13% of all eligible RISK11- participants were enrolled (Appendix 6.1). Therefore, to obtain estimates that reflect the screened population, enrolled participants were assigned sampling weights of 1.263 and 7.920 for RISK11+ and RISK11- participants, respectively. Sample size calculation was performed for the parent CORTIS study for which this study utilised the entire dataset.

6.4.3 Ethics approval

The CORTIS study received approval from all the five institutional human research ethics committees of the sites that participated and was also registered with ClinicalTrials.gov (NCT02735590). The protocol for the current analysis was approved by the University of Cape Town Human Research Ethics Committee. All study participants provided written informed consent.

6.4.4 Role of the funding source

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of this manuscript. All authors had full access to all the data reported here and approved to submit the manuscript for publication.

6.5 Results

Of the 2923 participants enrolled, 11 were excluded for a missing or indeterminate RISK11 or QFTPlus result. Of the 2912 participants evaluated, the proportions in each RISK11/QFTPlus category, adjusted to the screening population were 6.3% (778) for RISK11+/QFTPlus+, 2.9% (356) for RISK11+/QFTPlus-, 57.1% (1117) for RISK11-/QFTPlus+, and 33.8% (661) for RISK11-/QFTPlus- (Table 6.1, Appendix 6.2). Median BMI was similar among RISK11+/QFTPlus+, RISK11+/QFTPlus-, RISK11-/QFTPlus+, and RISK11-/QFTPlus- individuals (p=0.50). Median age in RISK11+/QFTPlus+ was older than RISK11+/QFTPlus-, RISK11-/QFTPlus+, and RISK11-/QFTPlus- individuals (p<0.001). The proportion of individuals with a TB contact history, fever and loss of weight was similar among all the RISK11/QFTPlus groups (p>0.05). Significant differences (p<0.05) among the four RISK11/QFTPlus groups were observed in the proportion of males, race groups, individuals with a smoking history, prior TB, flu-like symptoms, chest pains, haemoptysis, night sweats, prevalent and incident TB.

Table 6.1: Comparison of baseline characteristics by RISK11/QFTPlus risk categories

	Total (n=2912)	Total Adjusted	RISK11+/ QFTPlus+ (n=778)	RISK11+/ QFTPlus– (n=356)	RISK11– /QFTPlus+ (n=1117)	RISK11– /QFTPlus– (n=661)	P-Value
Age (median, IQR)	26.0 (22.0–33.0)	26.0	26.0 (22.0–34.0)	24.0 (21.0–30.0)	24.0 (21.0–31.0)	24.0 (21.0–31.0)	<0.001
BMI (median, IQR)	22.6 (20.0–27.7)	22.7	22.5 (20.1–27.1)	22.2 (19.9–27.7)	22.5 (19.9–27.9)	23.1 (20.0–28.1)	0.50
Sex (n, %)							
Male	1331 (45.7)	48.4	309 (39.72)	146 (41.01)	576 (51.57)	300 (45.39)	<0.001
Female	1581 (54.3)	51.6	469 (60.28)	210 (58.99)	541 (48.43)	361 (54.61)	
Race (n, %)							
Mixed	965 (33.1)	30.7	365 (46.92)	67 (18.82)	432 (38.68)	101 (15.28)	
Black African	1939 (66.6)	69.0	412 (52.96)	288 (80.90)	681 (60.97)	558 (84.42)	<0.001
Caucasian	4 (0.1)	0.2	0 (0.00)	1 (0.28)	2 (0.18)	1 (0.15)	
Asian	4 (0.1)	0.2	1 (0.13)	0 (0.00)	2 (0.18)	1 (0.15)	
History of smoking (n, %)	1476 (50.7)	49.9	447 (57.46)	146 (41.01)	628 (56.22)	255 (38.58)	<0.001
Prior tuberculosis (n, %)	230 (7.9)	7.0	86 (11.05)	25 (7.02)	102 (9.13)	17 (2.57)	<0.001
Tuberculosis contact history (n, %)	461 (15.8)	16.0	126 (16.20)	49 (13.76)	190 (17.01)	96 (14.56)	0.37
Flu-like symptoms (n, %)	133 (4.6)	3.7	46 (5.91)	25 (7.02)	47 (4.21)	15 (2.27)	0.01
Tuberculosis Symptoms							
Any Symptom	121 (4.2)	3.3	43 (5.53)	23 (6.46)	35 (3.13)	20 (3.03)	0.01
Chest pains (n, %)	30 (1)	0.8	13 (1.67)	3 (0.84)	5 (0.45)	9 (1.36)	0.04
Cough (n, %)	57 (2)	1.5	24 (3.08)	8 (2.25)	15 (1.34)	10 (1.51)	0.05
Fever (n, %)	3 (0.1)	0.1	2 (0.01)	0 (0.00)	1 (0.09)	0 (0.00)	0.64
Haemoptysis (n, %)	1 (0.1)	0.1	1 (0.00)	0 (0.00)	1 (0.09)	0 (0.00)	1.00
Loss of weight (n, %)	40 (1.4)	1.3	12 (1.54)	6 (1.69)	14 (1.25)	8 (1.21)	0.85
Night sweats (n, %)	32 (1.1)	0.7	17 (2.19)	6 (1.69)	4 (0.36)	5 (0.76)	<0.001
Tuberculosis Disease							
Prevalent (n, %)	74 (2.5)	1.4	47 (6.04)	9 (2.53)	15 (1.34)	3 (0.45)	<0.001
Incident (n, %) [#]	56 (2.0)	1.5	28 (3.83)	5 (1.44)	20 (1.81)	3 (0.46)	<0.001

RISK11 and QFTPlus thresholds used in this table were 60% and 0.35 IU/mL, respectively.

#Percentages for incident tuberculosis were computed using the 'at risk' population, i.e., excluding prevalent tuberculosis cases.

For continuous data, p values were computed using the Kruskal-Wallis test. For categorical data, p values were computed using the Fisher's exact test. Point estimates (proportions and medians) were computed using the enrolled population. Because the point estimates are within the same RISK11 groupings, the adjusted and unadjusted medians and proportions will be the same except for the 'Totals" which cross RISK11 groupings, hence the "adjusted total" column.

BMI, Body Mass Index. IQR, Interquartile Range. QFTPlus, QuantiFERON-TB Gold-Plus.

6.5.1 Associations between RISK11 and QFTPlus

Quantitative and qualitative associations between RISK11 and QFTPlus were evaluated. A very weak correlation was observed between RISK11 and QFTPlus scores among prevalent TB cases (ρ=–0.23; p=0.05) but not among controls (ρ=0.05; p=0.004 and incident TB cases (ρ=–0.03; p=0.80; Appendix 6.3). Qualitative test result agreement between RISK11 and QFTPlus was poor (40.1%; κ=0.02) using pre-specified positivity thresholds of 60% and 0.35 IU/mL, respectively.

Test result agreement improved from 40.1% to 67.3% when using a QFTPlus threshold of > 4 IU/mL and RISK11 threshold of 60%. Similar estimates in qualitative test result agreement between RISK11 and QFTPlus were observed when the RISK11 threshold was lowered to 26%, for both the 0.35 and 4 IU/mL QFTPlus thresholds (45.6% and 61.2% respectively, Appendix 6.4). Cohen's kappa statistics (κ) were below 0.2 for all RISK11/QFTPlus groupings (Appendix 6.4).

RISK11+ rates were higher in those with prevalent and incident TB compared to controls without disease (Appendix 6.5a). When stratified into QFTPlus ≤4 and >4 IU/mL groups, RISK11+ rates were similar (p>0.05) between the two QFTPlus groups in all subgroups of those with prevalent or incident TB and controls without TB Appendix 6.5b, 6.5d and 6.5f). Upon QFTPlus stratification into <0.35, 0.35–4, and >4 IU/mL groups, RISK11+ rates among controls were higher in individuals with QFTPlus values of 0.35–4 or >4 IU/mL compared to those with QFTPlus values of <0.35 IU/mL; no differences were observed in RISK11+ rates among the prevalent and incident TB cases (Appendix 6.5c, 6.5e and 6.5g).

6.5.2 Prevalent TB

Probability of prevalent TB for double-positive RISK11+/QFTPlus+ individuals

Probability of prevalent TB was assessed in participants with a double-positive RISK11+/QFTPlus+ result compared to other risk categories. Prevalence of TB disease was 0.45% (3/661) in RISK11-/QFTPlus-; 2.53% (9/356) in RISK11+/QFTPlus-; 1.34% (15/1117) in RISK11-/QFTPlus+; 6.04% (47/778) in RISK11+/QFTPlus+ individuals and 1.37% overall (74/2912; adjusted to reflect screening population, see Methods). Participants with a double-positive RISK11+/QFTPlus+ result were 13.31 (95% CI 4.16–42.7; p<0.001) times more likely to have TB disease at baseline compared to participants with a RISK11-/QFTPlus- result (Table 6.2a); and 2.39 (95% CI 1.18–4.82, p=0.02) and 4.50 (95% CI 2.53–7.99, p<0.001) times more likely to have TB disease at baseline than RISK11+/QFTPlus- and RISK11-/QFTPlus+ participants, respectively.

Table 6.2: Prevalent and incident tuberculosis disease by risk category

Risk Group	a) Prevalent Tuberculosis			b) Incident Tuberculosis		
	Prevalence, % (95% CI)	Prevalence Ratio (95% CI)	P-Value	Incidence rate per 100 person-years (95% CI)	Incidence Rate Ratio (95% CI)	P-Value
RISK11-/QFTPlus-	0.45 (0.09–1.32)	Reference		0.46 (0.14–2.23)	Reference	
RISK11+/QFTPlus-	2.53 (1.16–4.74)	5.57 (1.52–20.45)	0.01	1.47 (0.63–4.36)	3.24 (0.78–13.48)	0.11
RISK11-/QFTPlus+	1.34 (0.75–2.21)	2.96 (0.86–10.18)	0.09	1.78 (1.17–2.85)	3.90 (1.16–13.09)	0.03
RISK11+/ QFTPlus+	6.04 (4.47–7.95)	13.31 (4.16–42.58)	<0.001	3.75 (2.63–5.55)	8.23 (2.51–27.01)	0.001

RISK11 and QFTPlus thresholds used in this table were 60% and 0.35 IU/mL, respectively. QFTPlus, QuantiFERON-TB Gold-Plus.

Probability of prevalent TB at alternative RISK11 and QFTPlus test thresholds

Using a 60% threshold for RISK11-positivity, and stratifying QFTPlus into <0.35, 0.35–4, and >4 IU/mL risk groups, probability of prevalent TB was highest in participants with a RISK11+/QFTPlus >4 IU/mL result, with 14.77-fold (95% CI 4.47–48.87) higher probability of TB at baseline compared to RISK11-/QFTPlus- (<0.35 IU/mL) participants (Table S2a), but this risk was not significantly higher than participants with RISK11+/QFTPlus 0.35–4 IU/mL (PR=1.22, 95% CI 0.70–2.12, p=0.49). When the RISK11 threshold was lowered to 26% for the same QFTPlus thresholds, lower probability of prevalent TB was observed (Appendix 6.6). Optimal diagnostic thresholds for RISK11 and QFTPlus were computed using the Youden Index and found to be 26% and 0.92 IU/mL for RISK11 and QFTPlus, respectively. Using these optimal thresholds to categorise participants improved the probability of prevalent TB from 13% to 17% (Table 6.2a versus Appendix 6.7).

Diagnostic performance of RISK11/QFTPlus test combinations

Diagnostic and prognostic performance of a combined RISK11/QFTPlus test for TB disease was compared to individual test performance. Individual diagnostic and prognostic performance of RISK11 and QFTPlus was previously reported for all participants in the CORTIS trial.[10] This analysis includes only participants with a valid result for both tests; and the endpoint definition requires microbiological confirmation of TB disease on one or more, rather than two, sputum samples. Performance of individual and combined tests for prevalent and incident TB are shown in Tables 6.3 and Appendix 6.8 for adjusted and unadjusted performance metrics, respectively.

Table 6.3: Performance of RISK11 and QFTPlus alone and in combination for diagnosis of prevalent and prognosis of incident tuberculosis.

Statistic	(a) Prevalent Tuberculosis			
	RISK11 (60)	QFTPlus (0.35)	RISK11/QFTPlus (Both-Positive)	RISK11/QFTPlus (Either-Positive)
PR (95% CI)	4.88 (2.88–8.25)	2.93 (1.23–6.97)	5.70 (3.42–9.50)	4.06 (1.25–13.18)
Sensitivity (95% CI)	33.1 (23.2–45.7)	83.5 (73.4–91.3)	27.7 (18.5–40.0)	88.8 (79.8–95.2)
Specificity (95% CI)	91.1 (90.1–92.2)	36.9 (35.1–38.7)	94.0 (93.0–94.8)	34.1 (32.3–35.9)
PPV (95% CI)	4.9 (3.8–6.4)	1.8 (1.6–2.0)	6.0 (4.5–8.0)	1.8 (1.7–2.0)
NPV (95% CI)	99.0 (98.4–99.4)	99.4 (98.7–99.8)	98.9 (98.4–99.3)	99.5 (98.7–99.9)
LR+ (95% CI)	3.7 (2.6–5.2)	1.3 (1.2–1.5)	4.7 (3.2–7.0)	1.3 (1.2–1.5)
LR– (95% CI)	0.7 (0.6–0.9)	0.4 (0.3–0.7)	0.8 (0.6–0.9)	0.3 (0.2–0.6)
Statistic	(b) Incident Tuberculosis			
	RISK11 (60)	QFTPlus (0.35)	RISK11/QFTPlus (Both-Positive)	RISK11/QFTPlus (Either-Positive)
IRR (95% CI)	2.36 (1.39–4.00)	3.69 (1.38–9.84)	2.88 (1.69–4.96)	4.27 (1.31–13.95)
Sensitivity (95% CI)	19.0 (8.9–30.4)	86.7 (73.8–93.6)	15.8 (7.6–28.3)	89.4 (78.1–96.0)
Specificity (95% CI)	91.4 (90.3–92.4)	37.2 (35.5–39.1)	94.2 (93.3–95.0)	34.5 (32.7–36.3)
PPV (95% CI)	3.1 (2.1–4.3)	2.0 (1.4–2.8)	3.9 (2.6–5.6)	2.0 (1.2–2.3)
NPV (95% CI)	98.7 (98–99.2)	99.5 (98.8–99.8)	98.7 (98.1–99.1)	99.5 (98.7–99.9)
LR+ (95% CI)	2.3 (1.3–3.9)	1.4 (1.2–1.5)	2.8 (1.5–5.1)	1.4 (1.2–1.5)
LR– (95% CI)	0.9 (0.8–1.0)	0.4 (0.2–0.7)	0.9 (0.8–1.0)	0.3 (0.1–0.7)

PR, Prevalence ratio. IRR, Incidence-rate ratio. TP, True positive. TN, True negative. FP, False positive. FN, False negative. LR+, Positive likelihood ratio. LR–. Negative likelihood ratio. QFTPlus, QuantiFERON-TB Gold-Plus.

'Either-Positive' combination outcomes defined as: positive test = +/– or –/+ or +/+; negative test = –/–

'Both-Positive' combination outcomes defined as: positive test= +/+; negative test = –/– or +/– or –/+.

RISK11 and QFTPlus thresholds used in this table were 60% and 0.35 IU/mL, respectively. Performance measures are adjusted to the screening population.

Compared to QFTPlus, the Both-Positive test increased the diagnostic LR+ from 1.3 (95% CI 1.2–1.5) to 4.7 (95% CI 3.2–7.0). However, this improvement in LR+ were associated with deterioration in diagnostic LR- from 0.3 (95% CI 0.2–0.6) to 0.8 (95% CI 0.6–0.9). By using the Both-Positive test there would be approximately 5727 additional TN versus 76 FN (Missed cases) results compared to QFTPlus alone; and 281 additional TN versus 7 FN results compared to RISK11 for every 10,000 tests done (Appendix 6.9a). The LR+ and LR- for the Both-Positive were not significantly changed compared to RISK11 alone.

The Either-Positive test improved upon RISK11 alone, with decreased diagnostic LR- from 0.7 (95% CI 0.6–0.9) to 0.3 (95% CI 0.2–0.6). The improvement in LR- was accompanied by deterioration in diagnostic LR+ from 3.7 (95% CI 2.6–5.2) to 1.3 (95% CI 1.2–1.5). This change in performance would yield approximately 77 additional TP versus 5629 FP results for the Either-Positive test compared to RISK11 alone; and 8 additional TP versus 282 FP results compared to QFTPlus, for every 10,000 tests conducted (Appendix 6.9a). The LR+ and LR- for the Either-Positive test was not significantly changed compared to QFTPlus alone.

Although using optimal thresholds showed an improvement in diagnostic sensitivity for the Both-Positive test, there was a reduction in the LR+. Similar diagnostic performance estimates to those observed using the 60% and 0.35 IU/mL thresholds for RISK11 and QFTPlus were observed for the Either-Positive test (Table 6.3a versus Appendix 6.10a).

When RISK11 and QFTPlus were treated as continuous variables, there was no improvement in performance upon that of RISK11 alone (AUCs= 0.74 vs 0.75), but improvement was observed upon that of QFTPlus alone (AUCs=0.63 vs 0.75; Appendix 6.11a). The optimal combination risk score was a predicted probability for prevalent TB of 1.17%; and this risk score achieved diagnostic sensitivity of 70.25% (95% CI 58.52–80.34), specificity of 71.88% (95% CI 70.19–73.53), positive predictive value (PPV) of 3.36% (95% CI 2.30–4.86) and negative predictive value (NPV) of 99.43% (95% CI 98.98–99.70).

6.5.3 Incident TB

Risk of incident TB for double-positive RISK11+/QFTPlus+ individuals

Risk of incident TB was assessed in the 2838 participants at risk of progressing to incident TB. 20 of the 2838 participants were excluded from the analysis because they did not attend any follow-up visit. Incident TB was diagnosed in 56 of the remaining 2818 participants (adjusted incidence rate 1.47 per 100 person-years, 95% CI 1.04–2.07). TB incidence per 100 person-years was 0.46 in RISK11-/QFTPlus-; 1.47 in RISK11+/QFTPlus-; 1.78 in RISK11-/QFTPlus+ and 3.75 in RISK11+/QFTPlus+ individuals. Double-positive RISK11+/QFTPlus+ participants were 8.23 times (95% CI 2.51–27.01; p=0.001) more likely to develop TB disease compared to RISK11-/QFTPlus- participants (Table 2b); and at 2.53 times (95% CI 1.00–6.55, p=0.05) and 2.11 times (95% CI 1.19–3.72, p=0.01) higher risk of progressing to TB disease than RISK11+/QFTPlus- and RISK11-/QFTPlus+ participants, respectively (Table 6.2b).

Upon excluding 372 participants that received the intervention drug 3HP, incident TB was found in 49 of the remaining 2446 participants. (Incidence rate, 1.43 per 100 person-years, 95% CI 1.02–2.07). TB

incidence per 100 person-years was 0.46 in RISK11-/QFTPlus-; 1.83 (95% CI in RISK11+/QFTPlus-; 1.78 in RISK11-/QFTPlus+ and 4.24 in RISK11+/QFTPlus+ individuals. Double-positive RISK11+/QFTPlus+ participants were 9.31 times (95% CI 2.80–31.03; p<0.001) more likely to progress to incident disease compared to RISK11-/QFTPlus- participants (Appendix 6.12a); and at 2.39 times (95% CI 1.16–13.10, p=0.01) higher risk of progressing to incident disease than RISK11-/QFTPlus+ participants, but this risk was not significantly higher than RISK11+/QFTPlus- participants (IRR 2.32, (95% CI 0.81–6.67; p=0.12).

Risk of incident TB at alternative RISK11 and QFTPlus test thresholds

Similar to prevalent TB, risk of progressing to incident TB was highest in participants with a RISK11+/QFTPlus >4 IU/mL result, who had a 10.09-fold (95% CI 2.93–34.76, p<0.001) higher risk of progression compared to QFTPlus-/RISK11- participants (Appendix 6.6b), but which was not significantly higher than participants with RISK11+/QFTPlus 0.35–4 IU/mL (IRR =1.49, 95% CI 0.71–3.10, p=0.29). When the RISK11 threshold was lowered to 26% for the same QFTPlus thresholds, lower risk of incident TB was observed (Appendix 6.6b). Disregarding RISK11 results but using the same QFTPlus stratifications (<0.35, 0.35–4, and >4 IU/mL), risk of progression to incident TB was 4.38-fold (95% CI 1.56–12.32) higher in participants with QFTPlus >4 IU/mL than QFTPlus- (<0.35 IU/mL) participants (Table S8), but not significantly higher than QFTPlus+ participants with values between 0.35–4 IU/mL (IRR=1.41, 95% CI 0.68–2.92, p=0.35).

Using the optimal thresholds of 26% and 0.92 IU/mL for RISK11 and QFTPlus, respectively, risk of incident TB halved from 8% to 4% per 100 person-years for RISK11+/QFTPlus+ individuals, compared to RISK11-/QFTPlus- individuals (Table 6.2b versus Appendix 6.9b). Similar estimates were observed when participants in the intervention group were excluded. The risk of developing TB reduced from 9% to 4% per 100 person-years for RISK+/QFTPlus+ individuals compared to RISK11-/QFTPlus- individuals (Appendix 6.12a vs 12b).

Prognostic performance of RISK11/QFTPlus test combinations

Compared to QFTPlus, the Both-Positive test increased prognostic LR+ from 1.4 (95% CI 1.2–1.5) to 2.8 (1.5–5.1). The improvement in LR+ was associated with deterioration in prognostic LR- from 0.4 (95% CI 0.3–0.7) to 0.9 (95% CI 0.8–1.0). By using the Both-Positive test, an additional 5766 TN versus 100 FN results compared to QFTPlus alone; and an additional 4 TN versus 276 FN compared to RISK11 alone, would be expected, for every 10,000 tests conducted. The LR+ and LR- for the Both-Positive test were not significantly changed compared to RISK11 alone (Appendix 6.9b).

The Either-Positive test improved upon RISK11 alone, with decreased prognostic LR- from 0.9 (95% CI 0.8–1.0) to 0.3 (95% CI 0.1–0.7). The improvement in LR- was accompanied by deterioration in prognostic LR+ from 2.3 (95% CI 1.3–3.9) to 1.4 (95% CI 1.2–1.5). This change in performance would yield approximately 104 TP versus 5607 FP, compared to RISK11 alone; and approximately 8 TP versus 241 FP compared to QFTPlus alone. The LR+ and LR- for the Either-Positive test were not significantly changed compared to QFTPlus alone.

Using optimal thresholds showed a slight improvement in prognostic sensitivity for the Both-Positive test, which was accompanied by a reduction in the LR+. Similar prognostic performance estimates to those observed using the 60% and 0.35 IU/mL thresholds for RISK11 and QFTPlus were observed for the Either-Positive test (Table 6.3b versus Appendix 6.10b).

Combining RISK11 and QFTPlus as continuous variables did not improve upon the performance of QFTPlus alone (AUCs= 0.65 vs 0.67) but improved upon that of RISK11 alone (AUCs=0.55 vs 0.67; Figure S5b). The optimal combination risk score was a predicted hazard for incident TB of 1.88; and this risk score achieved prognostic sensitivity of 68.13% (95% CI 54.04–79.71), specificity of 63.86% (95% CI 62.52–66.13), PPV of 2.73% (95% CI 1.79–3.86) and NPV of 99.26% of (95% CI 87.88–90.64).

6.6 Discussion

We have shown in a large prospective cohort of adults who are HIV-uninfected in a TB endemic country that correlation and agreement between RISK11 and QFTPlus was poor, suggesting that a combination test might improve discriminatory accuracy in clinical scenarios in which the benefits of one test may mitigate the deficiencies of the other. Indeed, the combination of a positive RISK11 test with a positive QFTPlus test was associated with significantly increased risk for both prevalent and incident TB. However, the use of RISK11 with QFTPlus in a combination test did not add to the overall performance for both prevalent and incident TB, since no simultaneous improvement in the LR+ and LR- was observed, relative to one of the individual tests. This was also confirmed when tests were treated as continuous variables; no overall improvement in performance of the combination test compared to the individual tests was observed. Our analysis also confirmed that risk of progression to incident TB was highest in those QFTPlus+ individuals with IFN-γ values >4 IU/mL, as has been shown previously in both children and adults.[20-22]

Probability of prevalent and risk incident TB has previously been shown to be higher in participants with RISK11+ results compared to those with RISK11- results[10, 23] Probability of prevalent and risk of incident TB has also been shown to be higher in QFTPlus+ individuals compared to QFTPlus- individuals.[10, 23-25] We demonstrated that testing positive for both RISK11 and QFTPlus poses an even higher probability of prevalent and greater risk of incident TB, compared to testing positive for one test alone. Individuals who tested RISK11+/QFTPlus+ double-positive had the highest probability of prevalent TB and highest risk of progressing to incident TB, compared to those who tested RISK11+/QFTPlus-, RISK11-/QFTPlus+, or RISK11-/QFTPlus-. Although double-positive RISK11+/QFTPlus+ individuals showed significantly higher probability for prevalent and higher risk for incident TB disease compared to other risk groups, the highest risk was observed in individuals who tested RISK11+ and QFTPlus+ at IFN-γ values >4.00 IU/mL. It follows that this category of individuals should be the highest priority for investigation for TB; and those without prevalent TB should be offered preventive therapy to interrupt progression to TB disease.

These findings build upon our previous work on the association between positive RISK11 or QFTPlus tests and probability of prevalent and risk of incident TB disease[10, 23]; and on the work of others who have combined diagnostic tests for TB to improve accuracy.[26, 27] Fan *et al* found that using a single test of either culture, Xpert MTB/RIF, or simultaneous amplification testing method for TB (SAT-TB) resulted in lower sensitivity compared to a combined parallel testing method.[27] Similarly, Theron *et al* evaluated the diagnostic accuracy of adjunct tests, individually and in combination with Xpert MTB/RIF, and found that the combined tests improved diagnostic accuracy.[26] This study also found that either sensitivity or specificity may be improved upon compared to the individual tests, depending on the test combination chosen and whether the main aim is to improve sensitivity or specificity.

Risk of progression to TB disease has been shown to be significantly higher in individuals with IFN-γ values ≥ 4 IU/mL. Andrews *et al* reported that IFN-γ values >4 IU/mL in infants with recent infection were associated with increased risk of progression to TB disease compared to IFN-γ values of between 0.35–4 or <0.35 IU/mL.[20] Similarly, this study also found that individuals with IFN-γ values >4 IU/mL had 4.4-fold higher risk of incident TB disease compared to those with IFN-γ value < 0.35 IU/mL, although this risk was not significantly higher than those with IFN-γ values between 0.35–4 IU/mL.

Previous validation studies have shown that RISK11 has high diagnostic sensitivity and specificity for symptomatic TB; and high prognostic sensitivity and specificity for TB disease diagnosed within six

months of testing.[8, 10] Although QFTPlus is not routinely used as a diagnostic test, it has been shown to have reasonable diagnostic sensitivity and specificity.[24, 28] However, QFTPlus has poor prognostic specificity for incident TB.[29] We showed that comparison of RISK11 or QFTPlus alone to the combined tests, using likelihood ratios, did not show clear superiority of the combination compared to the individual tests, as there was no simultaneous improvement in both the LR+ and LR-. Similarly, treating the tests as continuous variables showed no overall improvement in the AUCs of the combined test compared to RISK11 and QFTPlus for prevalent and incident TB, respectively.

The use of the Both-Positive test improved the diagnostic and prognostic likelihood of TB, compared to QFTPlus alone, while not improving upon RISK11 alone. Similarly, the Either-Positive combination test improved the diagnostic and prognostic likelihood of absence of TB, compared to RISK11, but did not improve upon QFTPlus alone. For both test combinations, improvement in likelihood of prevalent and incident TB was associated with deterioration in the likelihood of absence of TB, and vice versa. This finding suggests that the trade-off between the expected increase in TP and FP, or TN and FN, may guide a decision on whether to use these tests in isolation or in combination.

For prevalent TB the Both-Positive test would result in approximately 40 more TN results for every FN result identified, compared to RISK11 alone; and 75 more TN results for every FN result identified, compared to QFTPlus alone. Similarly, the Either-Positive test would result in approximately 73 more FP results for every TP result identified, compared to RISK11 alone, and 35 more FP results for every TP result identified, compared to QFTPlus. Similar results for the expected number of additional TP/FP and TN/FN were obtained for the combination tests for discriminating incident TB from controls.

For the Either-Positive test, the expected increased number of FP results clearly outweighs the benefits of identifying few TP results. However, for the Both-Positive combination, the expected increase in TN results compared to the few FN results might suggest this combination as beneficial for a rule-in combination test, though the increase in LR+ for the combined test over RISK11 was only modest. Since the sensitivity and specificity of the individual tests are discordant rather than complementary, the use of combination testing could be tailored to the clinical setting, and specifically, to whether the primary goal of testing is to rule-in or rule-out risk of TB. The Both-Positive approach would be best suited as a TB rule-in test, because of the high specificity of this combination. Conversely, the Either-Positive approach would be best suited as a TB rule-out test, because of the moderately high sensitivity of this combination. However, because the improvements from combining the tests are not substantial, a rule-out test with these performance characteristics might not be

applied routinely in a high TB setting. There is need for a feasibility and cost-effectiveness evaluation to determine the most suitable approach to testing.

Weaknesses of this study include the fact that all participants were recruited in South Africa and hence the results may not be generalisable to other countries with different TB transmission dynamics and disease prevalence. Particularly, these results are unlikely to be representative of low TB incidence settings where greater emphasis is placed on latent TB infection screening programmes; and 3HP is likely to be effective for prevention of TB in such settings. Although we replicated the public health approach of using Xpert MTB/RIF for the diagnosis of prevalent TB, the absence of culture may have resulted in missing some Xpert MTB/RIF negative prevalent cases. Further work is warranted in high-risk populations, for example, people living with HIV or other co-morbidities, such as diabetes mellitus. Strengths of the study include the large study sample and large number of microbiologically-confirmed prevalent and incident TB cases. Furthermore, recruitment from five geographically distinct areas within South Africa with unique population demographics may aid generalisability of the study findings. These strengths, combined with the fact that all participants were tested for both RISK11 and QFTPlus, allow rigorous evaluation of the potential for combination testing to add diagnostic or prognostic value.

The findings highlight that individuals with elevated RISK11 and QFTPlus results have a higher probability of prevalent TB and are at greatly increased risk of incident TB disease. Evidence from this study suggests that either sensitivity or specificity may be improved upon further, relative to the individual tests, depending on the selected test combination and whether the primary goal is to maximise sensitivity or specificity. Given the inadequate sensitivity of the Both-Positive, and inadequate specificity of the Either-Positive approaches, RISK11/QFTPlus combination testing would not be suitable as a generic screening strategy. However, the expected increase in TN results compared to few additional FN results using the Both-Positive approach suggests that this RISK11/QFTPlus combination might have application as a rule-in test for TB.

Contributors

MH and TJS conceived and directed the study. MT, GW, KN, and GC were responsible for site-level activities, including recruitment, clinical management, and data collection. HM, SCM, AP-N, SKM, and MM provided operational or laboratory support and project management. AF-G provided statistical support. HM analysed the data and wrote the first draft of the manuscript. HM, AF-G, SCM, AP-N, SKM, EN, BB, MM, MT, GW, KN, GC, TJS, and MH had full access to the data, and reviewed, revised,

and approved the manuscript before submission. HM, AF-G, and BB accessed and verified the underlying data.

Declaration of interests

AP-N, GW, GC, TJS, and MH report grants from the Bill & Melinda Gates Foundation, during the conduct of the study; AP-N and GW report grants from the South African Medical Research Council, during the conduct of the study; GW and TJS report grants from the South African National Research Foundation, during the conduct of the study. In addition, AP-N and TJS have a patent of the RISK11 and RISK6 signatures issued; GW has a patent "TB diagnostic markers" (PCT/IB2013/054377) issued and a patent "Method for diagnosing TB" (PCT/IB2017/052142) pending. All other authors had nothing to disclose.

Acknowledgments

The authors gratefully acknowledge the study participants and funders. The CORTIS study was funded by the Bill and Melinda Gates Foundation (BMGF; OPP1116632 and OPP1137034) and the Strategic Health Innovation Partnerships (SHIP) Unit of the South African Medical Research Council with funds received from the South African Department of Science and Technology. SCM is a recipient of PhD funding from the Fogarty International Center of the National Institutes of Health (NIH) under Award Number D43 TW010559, the Harry Crossley Clinical Research Fellowship, and the South African Medical Research Council (SAMRC) through its Division of Research Capacity Development under the SAMRC Clinician Researcher Programme.

Funding

Bill and Melinda Gates Foundation, South African Medical Research Council.

Data sharing

The study protocol, de-identified RISK11 signature scores, TB endpoint data, and clinical metadata for all participants is available on Zivahub (https://doi.org/10.25375/uct.13573337.v1), an open access data repository hosted by the University of Cape Town's data repository powered by Figshare for Institutions.

6.7 References

1 Houben RM, Dodd PJ. The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. *PLoS medicine*. 2016;13(10):e1002152.

2 Zhou G, Luo Q, Luo S*, et al.* Interferon-gamma 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(12):1457-69.

3 Pai M, Denkinger CM, Kik SV*, et al.* Gamma interferon release assays for detection of Mycobacterium tuberculosis infection. *Clinical microbiology reviews*. 2014;27(1):3-20.

4 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. Report. Geneva: World Health Organization; 2017. Contract No.: WHO/HTM/TB/2017.18.

5 WHO. Meeting Report : High-priority target product profiles for new tuberculosis diagnostics: report of a consensus meeting. Geneva, Switzerland: World Health Organization; 2014. Contract No.: WHO/HTM/TB/2014.18.

6 MacLean E, Broger T, Yerlikaya S, Fernandez-Carballo BL, Pai M, Denkinger CM. A systematic review of biomarkers to detect active tuberculosis. *Nat Microbiol*. 2019;4(5):748-58.

7 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.

8 Zak DE, Penn-Nicholson A, Scriba TJ*, et al.* A blood RNA signature for tuberculosis disease risk: a prospective cohort study. *Lancet (London, England)*. 2016;387(10035):2312-22.

9 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*. 2018;108:124-6.

10 Scriba TJ, Fiore-Gartland A, Penn-Nicholson A*, et al.* Biomarker-guided tuberculosis preventive therapy (CORTIS): a randomised controlled trial. *The Lancet Infectious diseases*. 2021;21(3):354-65.

11 Camp BW, Bonnell LN. Combining Two Developmental Screening Tests to Improve Predictive Accuracy. *Acad Pediatr*. 2020;20(3):413-20.

12 Pepe MS, Thompson ML. Combining diagnostic test results to increase accuracy. *Biostatistics*. 2000;1(2):123-40.

13 Xu T, Fang Y, Rong A, Wang J. Flexible combination of multiple diagnostic biomarkers to improve diagnostic accuracy. *BMC Med Res Methodol*. 2015;15:94.

14 Franco EL, Ferenczy A. Assessing gains in diagnostic utility when human papillomavirus testing is used as an adjunct to Papanicolaou smear in the triage of women with cervical cytologic abnormalities. *Am J Obstet Gynecol*. 1999;181(2):382-6.

15 Macaskill P, Walter SD, Irwig L, Franco EL. Assessing the gain in diagnostic performance when combining two diagnostic tests. *Stat Med*. 2002;21(17):2527-46.

16 Thompson ML. Assessing the diagnostic accuracy of a sequence of tests. *Biostatistics*. 2003;4(3):341-51.

17 Bandos AI, Rockette HE, Gur D. Use of likelihood ratios for comparisons of binary diagnostic tests: Underlying ROC curves. *Medical physics*. 2010;37(11):5821-30.

18 Biggerstaff BJ. Comparing diagnostic tests: a simple graphic using likelihood ratios. *Statistics in Medicine*. 2000;19(5):649-63.

19 Koopman PAR. Confidence Intervals for the Ratio of Two Binomial Proportions. *Biometrics*. 1984;40(2):513-7.

20 Andrews JR, Nemes E, Tameris M*, et al.* Serial QuantiFERON testing and tuberculosis disease risk among young children: an observational cohort study. *The Lancet Respiratory medicine*. 2017;5(4):282-90.

21 Gupta RK, Lipman M, Jackson C*, et al.* Quantitative IFN-gamma Release Assay and Tuberculin Skin Test Results to Predict Incident Tuberculosis. A Prospective Cohort Study. *American journal of respiratory and critical care medicine*. 2020;201(8):984-91.

22 Winje BA, White R, Syre H*, et al.* Stratification by interferon-gamma release assay level predicts risk of incident TB. *Thorax*. 2018(73):10.

23 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.

24 Sotgiu G, Saderi L, Petruccioli E*, et al.* QuantiFERON TB Gold Plus for the diagnosis of tuberculosis: a systematic review and meta-analysis. *The Journal of infection*. 2019;79(5):444-53.

25 Altet N, Dominguez J, Souza-Galvao ML*, et al.* Predicting the Development of Tuberculosis with the Tuberculin Skin Test and QuantiFERON Testing. *Annals of the American Thoracic Society*. 2015;12(5):680-8.

26 Theron G, Pooran A, Peter J*, et al.* Do adjunct tuberculosis tests, when combined with Xpert MTB/RIF, improve accuracy and the cost of diagnosis in a resource-poor setting? *European Respiratory Journal*. 2012;40(1):161-8.

27 Fan L, Li DF, Zhang SJ*, et al.* Parallel Tests Using Culture, Xpert MTB/RIF, and SAT-TB in Sputum Plus Bronchial Alveolar Lavage Fluid Significantly Increase Diagnostic Performance of Smear-Negative Pulmonary Tuberculosis. *Frontiers in microbiology*. 2018;9.

28 Takasaki J, Manabe T, Morino E*, et al.* Sensitivity and specificity of QuantiFERON-TB Gold Plus compared with QuantiFERON-TB Gold In-Tube and T-SPOT.TB on active tuberculosis in Japan. *Journal of infection and chemotherapy : official journal of the Japan Society of Chemotherapy*. 2018;24(3):188-92.

29 Mahomed H, Hawkridge T, Verver S*, et al.* The tuberculin skin test versus QuantiFERON TB Gold(R) in predicting tuberculosis disease in an adolescent cohort study in South Africa. *PloS one*. 2011;6(3):e17984.

Chapter 7

7 Discussion and conclusion

Chapter overview

This chapter describes the importance and implications of the primary findings.

7.1 Summary of the primary findings

This thesis set out to synthesise the literature on diagnostic and prognostic performance of published mRNA transcriptomic signatures for TB disease; to evaluate the effect of host factors and external factors, such as co-infection or colonization by upper respiratory tract organisms, on RISK11 score and discriminatory performance for TB; and to evaluate the effect of using QFTPlus in combination with RISK11 on diagnostic and prognostic test utility. Host blood mRNA signatures have considerable potential as screening tests for TB, but further optimisation is needed if these signatures are to be used as standalone diagnostic, triage, or predictive tests for therapeutic decision-making.[1] Several host factors affected RISK11 score, but only adjustment for cough status affected diagnostic performance.[2] To improve prognostic performance, combining host factors with RISK11 should be considered.[2] Additionally, the presence of upper respiratory tract viruses negatively affected diagnostic performance of RISK11 for TB; and accordingly, RISK11 could not discriminate between TB and upper respiratory tract viruses.[3] Finally, testing double-positive for RISK11 and QFTPlus was associated with a very high probability of prevalent TB and greater risk for incident TB. However, using a combination test of RISK11 and QFTPlus did not improve overall discriminatory performance, relative to the individual tests.[4]

The subsequent sections provide a summary of the spectrum of TB disease and importance and implications of the primary findings by chapter.

7.2 The spectrum of TB disease

To aid progress towards tuberculosis (TB) elimination, early diagnosis of all patients with all forms of TB through provision of suitable and inexpensive diagnostic solutions as envisioned by the New Diagnostics Working Group Strategic Framework (NDWG) 2018–2022 is critical.[5] However, development and provision of suitable TB diagnostic solutions is hampered partly by the limited

knowledge of the 'clinical pathogenic spectrum of infection and disease'. Rather than a disease with two states, infection and active disease, recent research has shown that TB exists as a continuum spectrum of disease in which latent TB infection (LTBI) and active disease can be divided along that clinical spectrum of the disease.[6] Clinicians often encounter individuals that do not fit into the binary LTBI and active TB groups and fall in between. Thus, patients may present with or without symptoms compatible with TB, a chest radiograph (CXR) compatible with TB, but with a bacteriologically negative result. The management of such individuals falling in between LTBI and active disease may lead to over treatment, under treatment, invasive investigation or loss to follow-up as these individual may not see the need for prolonged treatment when they are basically feeling well. The clinical spectrum of TB disease as currently understood comprises LTBI, incipient TB, subclinical TB and active TB disease.[6, 7] Because LTBI and incipient TB both present without symptoms coupled with negative bacteriological results; identification of individuals that are most likely to progress to subclinical and clinical active TB disease from this group is critical, if interruption of progression and transmission is to be achieved. The tuberculin skin test (TST) and interferon gamma release assay (IGRA) have poor positive predictive value[8] and specificity[9] for incident TB, and a CXR has poor specificity for TB disease.[10] These performance deficiencies of IGRA, TST and a CXR present a great opportunity for development of non-sputum biomarker-based tests that can identify individuals with LTBI and incipient TB who are most likely to progress to subclinical or active TB disease so that therapeutic interventions can be targeted to this group in order to interrupt disease progression and transmission. Host blood mRNA signatures of TB such as RISK11 are one such type of non-sputum biomarker-based tests that have been developed to aid TB elimination.[1, 11]

7.3 Diagnostic and prognostic performance of published mRNA signatures of TB

Comparison of mRNA signature performance in similar populations and under similar conditions is important to down-select candidate biomarkers for point of care (POC) tests. Evaluating biomarkers in carefully selected cohorts of TB cases and uninfected controls tends to over-estimate test performance due to spectrum bias and therefore future implementation decisions depend on additional testing of biomarker performance in clinically relevant populations under field conditions, in addition to feasibility and cost considerations. The primary need is for TB diagnostic tests that differentiate active TB from other diseases among symptomatic individuals seeking health care. However, none of the signatures validated in cohorts with other diseases in the systematic review met the minimum World Health Organisation (WHO) target product profile (TPP) for sensitivity and

specificity of a diagnostic test; although seven did meet the minimum WHO TPP for a triage test. Similar results were also observed in a study by Mendelsohn and colleagues, in which they performed head-to-head validation of nine host blood transcriptomic signatures for pulmonary TB (PTB) by PCR.[12] Several signatures in that study met the minimal TPP performance targets for a triage test. While only one signature met the minimum WHO TPP performance targets for a prognostic test to predict progression from LTBI to TB disease in the systematic review, several signatures met the TPP performance targets for a prognostic test in the Mendelson *et al* study.[12] These findings suggest that the greatest potential of host blood transcriptomic signatures is for use as triage and prognostic tests for TB rather than as diagnostic tests. This further implies that for mRNA signatures to meet the TPP performance targets of a diagnostic test, further optimisation and validation under field conditions is necessary. Furthermore, as has been shown here, discovery, validation and optimisation of these tests should account for host and external factors, including viral and other co-infections that may affect the discriminatory capacity of these tests. For prognostic signatures, a rethink of the TPP criteria may be required. Previous studies including the CORTIS study have shown that transcriptomic signatures perform markedly better and may meet the TPP criteria over shorter time-periods of between 3 to 12 months prior to TB diagnosis rather than the longer 2-year period stipulated in the TPP.[11, 13, 14]

7.4 Effect of host factors on discriminatory performance of the RISK11 mRNA signature

Although several host factors affected RISK11 readout, adjustment for TB-independent host factors affecting controls did not change diagnostic or prognostic performance of the RISK11 transcriptomic signature. However, stratification by cough status, a TB-dependent factor that was associated with a 72.55% marginal increase in RISK11 score in those with prevalent TB, significantly improved discriminatory accuracy in individuals with cough, whereas diagnostic performance in individuals without cough was poor. The superior performance of RISK11 in individuals with symptomatic TB disease, which is likely associated with more severe inflammation, suggests induction of interferon signalling pathways that results in elevated signature scores in TB cases, which in turn drives superior discriminatory performance. This finding implies that implementation of RISK11 as a triage or prognostic test for TB would not need adjustment for the measured TB-independent host factors, which did not alter diagnostic or prognostic performance, despite affecting the RISK11 read out. Conversely, the TB-dependent host factor cough, has a major impact on signature performance and should be accounted for when interpreting the test result. Since the presence or absence of a cough symptom has a major impact on diagnostic performance for TB disease, this finding might severely limit the utility of transcriptomic biomarkers for triage and active case-finding approaches for

subclinical TB, which forms a large proportion of prevalent TB in South Africa and endemic communities elsewhere.[15] One approach to overcome some of the performance limitations that may be attributed to host factors might be inclusion of those host factors that contribute to discrimination into a classification model.[16]

7.5 Combining host factors with RISK11 to improve discriminatory performance

Several studies have shown that combining biomarkers with host risk factors may significantly improve classification capacity.[17-21] Patel *et al* developed a clinical prediction rule from risk factors associated with TB meningitis (TBM) and combined it with the lipoarabinomannan (LAM) antigen detection test; and found that the combination of risk factors for TBM in a clinical prediction rule and LAM markedly improved the sensitivity from 31% to 63% while still maintaining a high specificity of 93% (specificity for LAM alone was 94%) compared to LAM alone.[17] Another study by Tenforde *et al*, found that combining the biomarkers C-reactive protein (CRP), lipopolysaccharide (LPS) and albumin with the baseline patient characteristics CD4 count, body mass index, and prior TB, modestly improved discrimination for TB risk.[18] In this study, it has been shown that although certain host factors that affect RISK11 score are also associated with TB risk, incorporation of these host factors into a combination signature did not significantly improve diagnostic performance for prevalent TB, or for incident TB over the short 6-month predictive horizon, compared to RISK11 alone. This is perhaps indicative of the biomarker and host factor identifying the same pathological phenomenon, over the short time-periods. By contrast, combining baseline host factors with RISK11 significantly improved discrimination of incident TB from controls over the longer 12- and 15-month predictive horizons, compared to RISK11 alone. This finding illustrates that inclusion of host factors with the transcriptomic biomarker RISK11 in a combination signature may mitigate the time-dependent deterioration in signature performance over distant prognostic horizons, but does little to improve short-term performance. These findings may also be generalisable to other transcriptomic signatures that, like RISK11, include interferon stimulated genes (ISGs) that may be induced by coinfection with respiratory and other viruses.

7.6 Effect of upper respiratory viruses on discriminatory performance of RISK11 and risk of incident TB

Like HIV infection[22], common upper respiratory viral infections may affect host blood transcriptomic signature scores, likely due to induction of type 1 interferon (IFN) and raised expression of ISGs, which are included in RISK11 and many other mRNA signatures.[23] Presence of coronaviruses, influenza, and

rhinoviruses in the upper respiratory tract resulted in elevated RISK11 scores and diminished RISK11 discriminatory performance, such that RISK11 could not differentiate TB from viruses. Controlling for presence or absence of MTB, participants with upper respiratory viruses had elevated RISK11 scores, which were predicted to be 12.5%–67.0% higher compared to participants without such viruses. These results suggest that intercurrent respiratory viral infections, such as coronavirus, influenza, and rhinovirus, drive mRNA signature conversion and transient false-positive TB risk status, which impacts the specificity of such diagnostics during and immediately following such intercurrent respiratory viral illnesses. This finding demonstrates the importance of confirmatory testing and/or combined TB and viral screening, such as that implemented in some TB endemic settings for COVID-19. Alternatively, there is need to develop and validate mRNA signatures that are less affected by ISG modulation, such as the Thompson5 signature which showed that it is not affected by viral infection in a head-to-head comparison of signatures[12] and those proposed by Singhania *et al*., and Esmail *et al.*[24, 25]

Previous studies have reported that viral respiratory co-infections are associated with more rapid progression or more severe TB disease.[26, 27] A vital question is whether upper respiratory tract viral organisms might also induce changes in immune control of MTB infection and trigger progression to TB disease.[28] Although based on small numbers, participants with upper respiratory viral organisms in this study were five times more likely to develop TB compared to participants without such organisms. This preliminary evidence suggests that, in addition to the TB-independent effect of viruses on mRNA signature score discussed above, there may also be an association between upper respiratory viral organisms and risk of incident TB, supporting the hypothesis that respiratory viral infections may trigger progression to TB disease,[29, 30] or that immune dysfunction increases susceptibility to both viral and MTB infection. Regardless of any possible association between respiratory viral infection and progression to TB, it will be necessary to mitigate the TB-independent effect of viral co-infection on signature performance, for example by seeking to develop new mRNA signatures of TB that do not induce ISGs and which might not be affected by intercurrent viral infections[12] or by using other biomarkers such as IGRA in combination with mRNA signatures to improve performance.[17, 18]

7.7 Probability and risk of TB for individuals who test RISK11 and IGRA positive

Probability of prevalent TB and risk of incident TB has previously been shown to be higher in participants with RISK11+ results compared to those with RISK11- results.[11, 31] Similar findings have been shown for IGRA.[11, 31, 32] This study has demonstrated that testing positive for both RISK11 and

QFTPlus presents an even higher probability of prevalent TB, and greater risk of incident TB, compared to testing positive for one test alone. Although RISK11+/QFTPlus+ individuals showed significantly higher probability for prevalent TB and higher risk for incident TB disease, compared to other risk groups, the highest risk was observed in individuals who tested RISK11+ and QFTPlus+ at IFN-γ values >4.00 IU/mL. This finding suggests that highest priority for TB investigation should be given to this category of individuals; and those in this high-risk group who are investigated and not found to have prevalent TB should be offered preventive therapy (PT) to interrupt progression to TB disease.

7.8 Diagnostic and prognostic performance of a RISK11/IGRA combination testing

Previous studies have shown that RISK11 has high diagnostic sensitivity and specificity for symptomatic TB and high prognostic sensitivity and specificity for incident TB diagnosed within six months of testing[11, 13], whereas IGRA has been shown to have reasonable diagnostic sensitivity and specificity[33, 34]. However, in this study comparison of test combinations to RISK11 or QFTPlus alone, using likelihood ratios, did not show clear superiority of the test combinations compared to the individual tests, as there was no simultaneous improvement in both the LR+ and LR-. Similarly, no overall improvement in the AUCs of the combined test compared to RISK11 and QFTPlus alone for prevalent and incident TB, respectively, was observed when the tests were treated as continuous variables.

While the use of the Both-Positive test combination did indeed improve the diagnostic and prognostic likelihood of presence of TB, compared to QFTPlus alone, this test combination did not improve upon RISK11 alone. Similarly, the use of the Either-Positive test combination improved the diagnostic and prognostic likelihood of absence of TB, compared to RISK11, but did not improve upon QFTPlus alone. Thus, lack of overall superiority in performance of the combination tests compared to individual tests suggests that the trade-off between the expected increase in TP and FP, or TN and FN, should guide the decision on whether to use these tests in isolation or in combination.[35] The Either-Positive test combination did not appear beneficial on this basis, identifying few additional TP at the cost of a significant increase in number of FP results. However, benefit might be derived from using the Both-Positive combination as a rule-in test, based on the expected increase in TN results at the cost of few additional FN results. Thus, owing to the discordant rather than complementary sensitivity and specificity of the individual RISK11 and IGRA tests, the use of combination testing could be tailored to the clinical setting, and specifically, to whether the primary goal of testing was to rule in or rule out risk of TB. While the Both-Positive approach would be best suited as a TB rule-in test, on the basis of

specificity, the Either-Positive approach would be best suited as a TB rule-out test, because of its moderately high sensitivity. However, it must be acknowledged that the improvement in performance is not substantial and might preclude routine application as a rule-out test in a TB endemic setting, on the basis of higher costs and the need for more complex laboratory resources.

7.9 Limitations and strengths

This thesis has a number of limitations and strengths that affect interpretation and generalisability of the findings. Weaknesses of the systematic review include the fact that heterogeneity of study design and reliance on reported data made it difficult to compare mRNA signature performance directly and precluded some signatures from inclusion in the meta-analysis. Further, there was a lack of high quality, prospective studies in clinically relevant populations, which led to low quality of evidence. A further limitation is that new signatures published after May 2019 are not included in the thesis as they were not part of the published systematic review which is included verbatim. For both the longitudinal and cross-sectional respiratory organisms sub-study, a moderate weakness was that all participants were recruited in South Africa and hence the results may not be generalisable to other countries with different TB transmission dynamics and TB disease prevalence. Second, both these sub-studies used a TB disease endpoint definition based on one positive sputum sample, as used in the public health system, yet one-sample sputum positive cases were predominantly subclinical. This preponderance of subclinical TB cases contributed to poorer RISK11 performance compared to a two-sample-positive endpoint used in the CORTIS trial. Furthermore, although we replicated the public health approach of using Xpert MTB/RIF for the diagnosis of prevalent TB, the absence of culture in most participants may have resulted in missing some Xpert MTB/RIF-negative prevalent TB cases. Further work is warranted in high-risk populations, for example, people living with HIV or other co-morbidities, such as diabetes mellitus to evaluate whether effects on signature scores would be similar or greater in such populations. In the longitudinal study, we found that RISK11 performed significantly better in participants with the symptom of cough, but this finding was based on a relatively small sample size. Furthermore, although baseline predictors performed well relative to RISK11 for prediction of TB risk over long time-frames, it should be noted that this study reports the training cohort for these host factors and validated performance of host factors for TB prediction would require additional testing in an independent cohort. In the cross-sectional respiratory organisms sub-study, weaknesses included the fact that too few participants progressed to incident TB disease to allow a robust evaluation of the intriguing association between upper respiratory viral infection and progression to TB disease. Furthermore, because only the subset of participants co-enrolled in the parent CORTIS study underwent TB investigation, TB disease could not be excluded in the remainder.

Alternative diagnoses were not sought in symptomatic participants without detected respiratory tract organisms or TB disease; nor were participants investigated for lower respiratory or gastrointestinal viruses, which might also modulate RISK11 scores. Therefore, it is hypothesised that the variation in RISK11 score which is not explained by the model might be explained by other viruses and host factors that were not measured. Finally, it is acknowledged that several of the upper respiratory organisms observed are not typically pathogenic and occur as commensals.

Strengths of the study include the use of an inclusive time frame for the systematic review of January 2005 to May 2019, the period in which most transcriptomic TB biomarker studies were published. The systematic review protocol was developed and registered on PROSPERO prior to performing the systematic review and the protocol explicitly stated the rigorous search strategy and clear inclusion/exclusion criteria. Furthermore, in contrast to previous systematic reviews, this review included evaluation of signatures for predicting progression to TB disease and a meta-analysis. Strengths of the longitudinal study and the parent CORTIS study included the large sample size, relatively large number of TB cases, and the fact that the study recruited from five geographically distinct areas throughout South Africa with unique population demographics, which should allow some generalisability to other settings. Furthermore, all participants were tested for both RISK11 and QFTPlus, which allows for rigorous evaluation of the potential for combination testing to add diagnostic or prognostic value. In the respiratory organisms sub-study, participants were enrolled consecutively, thereby minimising selection bias; and both oropharyngeal and nasopharyngeal sampling was performed to optimize the detection rate of organisms in the upper respiratory tract. In terms of novelty, to our knowledge this is the first study to investigate perturbation of an ISG-based mRNA signature by common upper respiratory viral organisms in individuals at risk for incident TB disease.

Based on the primary findings of this thesis, some potential research areas have been identified.

7.10 Potential future research areas

The fact that several mRNA signatures for TB diagnosis were validated in clinically relevant cohorts and met the minimum TPP for a triage test, warrants further optimisation of such signatures towards meeting the optimal TPP; or if this is not attained, then the use case for such signatures may be better defined such as using them as rule-in/rule-out tests in those with Xpert-Ultra trace+ results, or for extra PTB disease or paediatric TB. Further, mRNA signatures should be validated under field

conditions, accounting for both host factors and extrinsic factors such as common upper respiratory organisms, to confirm their accuracy for use as standalone diagnostic or prognostic tests for therapeutic decision-making in settings where TB and multiple other respiratory pathogens are endemic.

The finding that the symptom of cough influenced discriminatory performance of RISK11, may warrant further work in high-risk populations, such as PLWH or people with other co-morbidities like diabetes mellitus, to confirm whether the presence or absence of such symptoms also affects these populations. Very large studies, or studies involving multi-centre collaborations that include cohorts from different studies and countries, would be ideal for such evaluations in high-risk sub-populations. Further, the observation that host factors performed well relative to RISK11 for prediction of TB risk over longer time-frames requires validation in an independent cohort, given that this observation was made in a training cohort. This finding is potentially important, since host factors are often cheap and easy to document, and we recommend that these data should always be studied together with relatively complex laboratory assays such as mRNA signatures. Further, the finding that viral co-infection may be associated with an elevated risk of progression to TB would require rigorous testing in large longitudinal studies that sample and investigate for both TB and multiple viral pathogens prospectively and at serial time-points. Additionally, urgent discovery and validation of more mRNA signatures similar to the Thompson5[12, 36] and those suggested by Singhania *et al*. that do not induce interferon pathway to avoid interference by viruses and improve specificity are needed.[12, 25, 36] Finally, 66% (37/56) of the incident TB cases in the parent CORTIS study were asymptomatic representing subclinical disease. There is need to discover or optimise existing signatures for identification of incipient and subclinical TB. It may be assumed that participants who progressed to incident TB may have had incipient TB at baseline. Therefore, baseline samples from the progressors may be used for discovery or validation of signatures for incipient TB while samples at the time of diagnosis may be used for the discovery or validation of signature for subclinical TB. Samples for symptomatic TB cases may then be used for discovery and validation of signatures for active TB. Because TB is a spectrum, it must be borne in mind that there is uncertainty about whether prevalent cases and early incident cases are indeed different in terms of their underlying TB pathogenesis.

7.11 Conclusion

Results from this thesis project indicate that host blood mRNA signatures do hold promise as triage tests for TB, but will require further optimisation if they are to be used as standalone diagnostic or

prognostic tests for therapeutic and preventive decision-making. Future transcriptomic signature discovery studies should not ignore host characteristics in their design. Evidence from this study suggests that presence or absence of cough has a major impact on diagnostic performance for TB disease, which might severely limit the utility of transcriptomic biomarkers for triage and active case-finding approaches for subclinical TB. Since subclinical TB forms a large proportion of prevalent TB in endemic communities, this is likely to be a major challenge for commercial development of current mRNA signature-based triage tests. Similarly, common upper respiratory viruses significantly increase signature scores and lead to considerable false-positive results, such that RISK11 could not differentiate between viruses and TB, which may also hinder programmatic implementation of such biomarkers as new tools for TB control, unless new signatures are developed that are less affected by intercurrent viral infection. It appears that combining two tests such as RISK11 and IGRA, which have discordant individual performance characteristics, does not further improve overall discriminatory performance, relative to the individual tests. However, either sensitivity or specificity may be improved upon further, relative to the individual tests, depending on the selected test combination and whether the primary goal is to maximise sensitivity or specificity. In summary, although current mRNA signatures show the potential of these biomarkers as commercial TB triage tests, this potential may only be realized if new signatures that include both subclinical and active TB, and which are not affected by intercurrent viral infection, are discovered and validated.

7.12 References

1 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.

2 Mulenga H, Fiore-Gartland A, Mendelsohn SC, et al. The effect of host factors on discriminatory performance of a transcriptomic signature of tuberculosis risk. *EBioMedicine.* 2022;77:103886.

3 Mulenga H, Musvosvi M, Mendelsohn SC, et al. Longitudinal Dynamics of a Blood Transcriptomic Signature of Tuberculosis. *American journal of respiratory and critical care medicine.* 2021;204(12):1463-72.

4 Mulenga H, Fiore-Gartland A, Mendelsohn SC, et al. Evaluation of a transcriptomic signature of tuberculosis risk in combination with an interferon gamma release assay: A diagnostic test accuracy study. *EClinicalMedicine.* 2022;47:101396.

5 New_Diagnostics_Working_Group. New Diagnostics Strategic Framework 2018–2022. Geneva: *Stop TB Partnership: FIND;* 2018.

6 Drain PK, Bajema KL, Dowdy D, et al. Incipient and Subclinical Tuberculosis: a Clinical Review of Early Stages and Progression of Infection. *Clinical microbiology reviews.* 2018;31(4).

7 Migliori GB, Ong CWM, Petrone L, D'Ambrosio L, Centis R, Goletti D. The definition of tuberculosis infection based on the spectrum of tuberculosis disease. *Breathe (Sheff)*. 2021;17(3):210079.

8 Abubakar I, Drobniewski F, Southern J, et al. Prognostic value of interferon-gamma release assays and tuberculin skin test in predicting the development of active tuberculosis (UK PREDICT TB): a prospective cohort study. *The Lancet Infectious diseases.* 2018;18(10):1077-87.

9 Mahomed H, Hawkridge T, Verver S, et al. The tuberculin skin test versus QuantiFERON TB Gold(R) in predicting tuberculosis disease in an adolescent cohort study in South Africa. *PloS one*. 2011;6(3):e17984.

10 Harris M, Qi A, Jeagal L, et al. A systematic review of the diagnostic accuracy of artificial intelligence-based computer programs to analyze chest x-rays for pulmonary tuberculosis. *PloS one.* 2019;14(9):e0221339.

11 Scriba TJ, Fiore-Gartland A, Penn-Nicholson A, et al. Biomarker-guided tuberculosis preventive therapy (CORTIS): a randomised controlled trial. The Lancet Infectious diseases. 2021;21(3):354-65.

12 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).

13 Zak DE, Penn-Nicholson A, Scriba TJ, et al. A blood RNA signature for tuberculosis disease risk: a prospective cohort study. *Lancet (London, England).* 2016;387(10035):2312-22.

14 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.

15 Van der Walt M, Moyo S. The First National TB Prevalence Survey - South Africa 2018. South Africa: *South African National Department of Health*; 2020.

16 Janes H, Longton G, Pepe M. Accommodating Covariates in Receiver Operating Characteristic Analysis. *Stata Journal.* 2009;9:17-39.

17 Patel VB, Singh R, Connolly C, et al. Comparison of a clinical prediction rule and a LAM antigen-detection assay for the rapid diagnosis of TBM in a high HIV prevalence setting. *PloS one*. 2010;5(12):e15664.

18 Tenforde MW, Gupte N, Dowdy DW, et al. C-reactive protein (CRP), interferon gamma-inducible protein 10 (IP-10), and lipopolysaccharide (LPS) are associated with risk of tuberculosis after initiation of antiretroviral therapy in resource-limited settings. *PloS one.* 2015;10(2):e0117424.

19 Gupta RK, Calderwood CJ, Yavlinsky A, et al. Discovery and validation of a personalized risk predictor for incident tuberculosis in low transmission settings. *Nature medicine.* 2020;26(12):1941-9.

20 Bedell RA, van Lettow M, Meaney C, et al. Predictive value of C-reactive protein for tuberculosis, bloodstream infection or death among HIV-infected individuals with chronic, non-specific symptoms and negative sputum smear microscopy. *Tropical medicine & international health : TM & IH.* 2018;23(3):254-62.

21 Sivakumaran D, Jenum S, Vaz M, et al. Combining host-derived biomarkers with patient characteristics improves signature performance in predicting tuberculosis treatment outcomes. *Commun Biol.* 2020;3(1):359.

22 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. *Frontiers in microbiology.* 2019;10:1441.

23 Berry MP, Graham CM, McNab FW, et al. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature.* 2010;466(7309):973-7.

24 Esmail H, Lai RP, Lesosky M, et al. Complement pathway gene activation and rising circulating immune complexes characterize early disease in HIV-associated tuberculosis. *Proceedings of the National Academy of Sciences of the United States of America.* 2018;115(5):E964-E73.

25 Singhania A, Verma R, Graham CM, et al. A modular transcriptional signature identifies phenotypic heterogeneity of human tuberculosis infection. *Nature communications*. 2018;9(1):2308.

26 Walaza S, Cohen C, Tempia S, et al. Influenza and tuberculosis co-infection: A systematic review. *Influenza Other Respir Viruses.* 2020;14(1):77-91.

27 Dangor Z, Izu A, Moore DP, et al. Temporal association in hospitalizations for tuberculosis, invasive pneumococcal disease and influenza virus illness in South African children. *PloS one.* 2014;9(3):e91464.

28 Esmail H, Barry CE, 3rd, Young DB, Wilkinson RJ. The ongoing challenge of latent tuberculosis. *Philos Trans R Soc Lond B Biol Sci.* 2014;369(1645):20130437.

29 Cobelens F, Nagelkerke N, Fletcher H. The convergent epidemiology of tuberculosis and human cytomegalovirus infection. *F1000Res*. 2018;7:280.

30 Martinez L, Nicol MP, Wedderburn CJ, et al. Cytomegalovirus acquisition in infancy and the risk of tuberculosis disease in childhood: a longitudinal birth cohort study in Cape Town, South Africa. *Lancet Glob Health.* 2021;9(12):e1740-e9.

31 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.

32 Altet N, Dominguez J, Souza-Galvao ML, et al. Predicting the Development of Tuberculosis with the Tuberculin Skin Test and QuantiFERON Testing. *Annals of the American Thoracic Society.* 2015;12(5):680-8.

33 Takasaki J, Manabe T, Morino E, et al. Sensitivity and specificity of QuantiFERON-TB Gold Plus compared with QuantiFERON-TB Gold In-Tube and T-SPOT.TB on active tuberculosis in Japan. *Journal of infection and chemotherapy : official journal of the Japan Society of Chemotherapy*. 2018;24(3):188-92.

34 Sotgiu G, Saderi L, Petruccioli E, et al. QuantiFERON TB Gold Plus for the diagnosis of tuberculosis: a systematic review and meta-analysis. *The Journal of infection.* 2019;79(5):444-53.

35 Macaskill P, Walter SD, Irwig L, Franco EL. Assessing the gain in diagnostic performance when combining two diagnostic tests. *Stat Med*. 2002;21(17):2527-46.

36 Thompson EG, Du Y, Malherbe ST, et al. Host blood RNA signatures predict the outcome of tuberculosis treatment. *Tuberculosis (Edinburgh, Scotland)*. 2017;107:48-58.

8 Appendices

8.1 Appendix 2.1: PRISMA-P 2015 checklist: Items to be reported in a systematic review

Section and Topic	Check list item & PRISMA-P code	Page
Section 1: Administrative information		
1. Title:	Identification [1a]	1
	Update [1b]	NA
2. Registration	Registration [2]	2
3. Authors:		
	Contact [3a]	1
	Contributions [3b]	14
4. Amendments	Amendments [4]	NA
5. Support:		
	Sources [5a]	NA
	Sponsor [5b]	NA
	Role of sponsor or funder [5c]	NA
Section 2: Introduction		
6. Rationale	Rationale [6]	4-6
7. Objectives	Objectives [7]	6
Section 3: Methods		
8. Eligibility	Eligibility criteria [8]	7
9. Information	Information sources [9]	7
10. Search	Search strategy [10]	7
11. Study records:		
	Data management [11a]	8
	Selection process [11b]	8
	Data collection process [11c]	9
12. Data	Data items [12]	9
13. Outcomes	Outcomes and prioritization [13]	9
14. Bias	Risk of bias in individual studies [14]	10
15. Data synthesis		
	Quantitative synthesis criteria [15a]	10
	Appropriateness of data for synthesis [15b]	10
	Sensitivity and subgroup analyses [15c]	10
	Qualitative synthesis? [15d]	11
16. Meta-bias(es)	Meta-bias(es) [16]	11
17. Confidence in cumulative evidence	Assessment of strength of cumulative evidence [17]	11

(PRISM-P=Preferred Reporting Items for Systematic Review-Protocol)

8.2 Appendix 2.2: Initial search strategy.

Performance of host blood transcriptomic signatures for diagnosis and prediction of progression to tuberculosis disease in HIV-negative adults and adolescents.

Search Item	Search Term	Connecting Operator
TB DISEASE	(Tuberculosis [MeSH] OR Mycobacterium tuberculosis [MeSH] OR (Tuberculosis OR TB OR Mycobacterium tuberculosis OR MTB))	AND
DIAGNOSTIC AND PROGNOSTIC	(Diagnosis [MeSH] OR Diagnosis [subheading] OR Prognosis [MeSH] OR (Diagnosis OR diagnostic OR detect* OR predict* OR prognosis OR prognostic OR screen*))	AND
HOST BLOOD TRANSCRIPTOMIC SIGNATURES	(Biomarkers/Blood [MeSH] OR RNA/Blood [MeSH] OR Transcription, Genetic [MeSH] /etiology/genetics/immunology OR (Blood Biomarker OR blood biomarkers OR bio-signature OR gene expression OR genetic transcription OR host blood OR immune marker OR immunologic marker OR Ribonucleic Acid OR RNA OR signature OR surrogate endpoint OR surrogate marker OR transcriptome OR transcriptomic))	AND
PERFORMANCE	(Area under Curve [MeSH] OR Sensitivity and Specificity [MeSH] OR (Area under curve OR area under curves OR AUC OR receiver operating characteristic OR ROC OR Accuracy OR Performance OR sensitivity OR specificity))	AND
*ADULTS AND ADOLESCENTS	Humans[Mesh]	AND
TIME PERIOD	Between 01-Jan-2005 and 31-May-2019	

* The search strategy did not filter by adults and adolescents as this may not be clearly indexed and may result in some articles being missed. Also, we did not filter by HIV status as some articles may contain both HIV negative and positive cohorts.

8.3 Appendix 2.3: Data extraction form

Performance of host blood transcriptomic signatures for diagnosis and prediction of progression to tuberculosis disease in HIV-negative adults and adolescents: a systematic review and meta-analysis:

Reviewer initials □ HM □ EWB □ Other_____

Part A: Study characteristics

Author's last name: _____ Publication year: _____

Study title: _____

Country of study population: _____ Study #_____

TB Burden in study population: □Low □Intermediate □ High

Study design: □Cross-sectional □Cohort □ Case control □ RCT □ Other _____

Study type: □ Diagnostic □ Predictive *if predictive; follow-up time_____months* **Study purpose:** □ Discovery □ Validation

Sampling: □ Consecutive □Convenient □Random □Other_____

Part C: Characteristics of tests

Index sample type : □ Whole blood □ PBMC

Index test (signature) type : □mRNA □Other_____

Signature name : _____ **# of genes:** _____

Signature discovery method: □ RNA Seq □Micro array □PCR □Other_____

Signature model : □ Pairwise □Random forest □SVM □Other_____

TB disease gold standard : □Culture □Xpert MTB/RIF □ Smear □ Other_____

Part B: Population characteristics

Population: □Adults □Adolescents □ Children □Mixed □ Undefined

Age range: _____

Participant's cohort (test/validation) and disease status**

Cohort Type	Healthy Controls[a]	LTBI[b]	Other Diseases[c]	TB Disease[d]	Total Enrolled

*** For diagnostic studies, the number of participants in each category represents the numbers at enrolment while for predictive studies; a, b and c represent the number of participants in each category at enrolment that gave rise to the cases (d) at the end of follow-up period.*

Description of negative population

[a]Healthy Controls: □ Healthy endemics □ Healthy non-endemics □ TB Contacts □ Other, describe:_____

[b]Other diseases: □ Non-respiratory diseases □ Respiratory diseases

[c]LTBI : □TST>5mm □TST>10mm □IGRA □ N/A

*If IGRA, IGRA type : □QFT □Elispot □Other*_____

[d]TB disease: For diagnostic studies, this number represents the number of participants included in the study at enrolment while for predictive studies; this number represents the number of participants diagnosed with TB at the end of follow-up period.

Part D: Outcomes

Cohort Type	Sens	Spec	AUC	TP	TN	FP	FN	LR+	LR-	RR

Key: *Sens= Sensitivity; Spec= Specificity; AUC= Area under the curve; TP = True positives; FP = False positives; TN = True negatives; FN = False negatives; LR+ = Positive likelihood ratio; LR- = Negative likelihood ratio; RR = Rate Ratio for predictive studies only.*

Inclusion/Exclusion Decision

□Included □Excluded □ Pending

If excluded, reason:

□ Excluded by title

□ Children only

□ Different index test

□ Reference test neither Culture, Xpert/MTB RIF nor Smear Microscopy

□ HIV infected only

□ Mixed HIV infected and uninfected and could not desegregate the participants

□ Mixed children and adults and could not desegregate the participants

□ Other study design

□ Diagnostic performance data not reported and unable to get from authors

8.4 Appendix 2.4: Individual Study Quality Assessment Tool (QUADAS-2)

Performance of host blood transcriptomic signatures for diagnosing and predicting progression to tuberculosis disease in HIV-negative adults and adolescents: a systematic review protocol:

Study Title: _____

Author: _____ Publication year: _____ Study #_____

Study type: □ Diagnostic □ Predictive

Domain 1. Patient selection

(a) Risk of bias: L/H/U

- Was a consecutive or random sample of patients enrolled? Y/N/U
- Was a case-control design avoided? Y/N/U

(b) Applicability Concerns: L/H/U

- Is there concern that the included patients do not match the review question? Y/N/U

Domain 2. Index Test (Transcriptomic Signature)

(a) Risk of bias: L/H/U

- Were the Transcriptomic signature test results interpreted without knowledgeof the results of the MTB culture or XPert/MTB RIF or Smear Microscopy? Y/N/U

(b) Applicability Concerns: L/H/U

- Is there concern that the Transcriptomic signature test, its conduct, or interpretation differ from the review question? Y/N/U

Domain 3. Reference Standard (MTB Culture or Xpert/MTB RIF or Smear Microscopy)

(a) Risk of bias: L/H/U

- Is the reference standard likely to correctly classify the target condition? Y/N/U
- Was MTB Culture or Xpert/MTB RIF or Smear Microscopy test results interpreted without knowledge of the results of Transcriptomic signature test? Y/N/U

(b) Applicability Concerns: L/H/U

- Is there concern that the MTB Culture or Xpert/MTB RIF or Smear Microscopy, its conduct, or interpretation differ from the review question? Y/N/U

Domain 4. Flow and timing

(a) Risk of bias: L/H/U

- Was there an appropriate interval between the index test and reference standard tests? Y/N/U
- Did patients in the study receive the same reference standard? Y/N/U
- Were all patients included in the analysis? Y/N/U

Overall Study Rating L/H/U

For each domain, 'risk of bias' or 'concerns regarding applicability' will be scored as 'L' if all responses in that domain are scored as 'Y' and 'H' if any of the responses is 'N' and 'U' if we are unclear for all the responses.

Legend: Y=Yes; N=No; U=Unclear ***H=high; L=Low; U=Unclear***

8.5 Appendix 3.1: PRISMA checklist: Items reported in the systematic review

	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	3-4
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	4
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.	4
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	4
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.	5
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	5
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	6
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.	6
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	6; Appendix 2.3
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	6; Appendix 2.4
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	7
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	7-8
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	7
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	NA
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	8
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	8; Appendices 3.4 and 3.7
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	7-8
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	7; Appendices 3.9 and 3.10

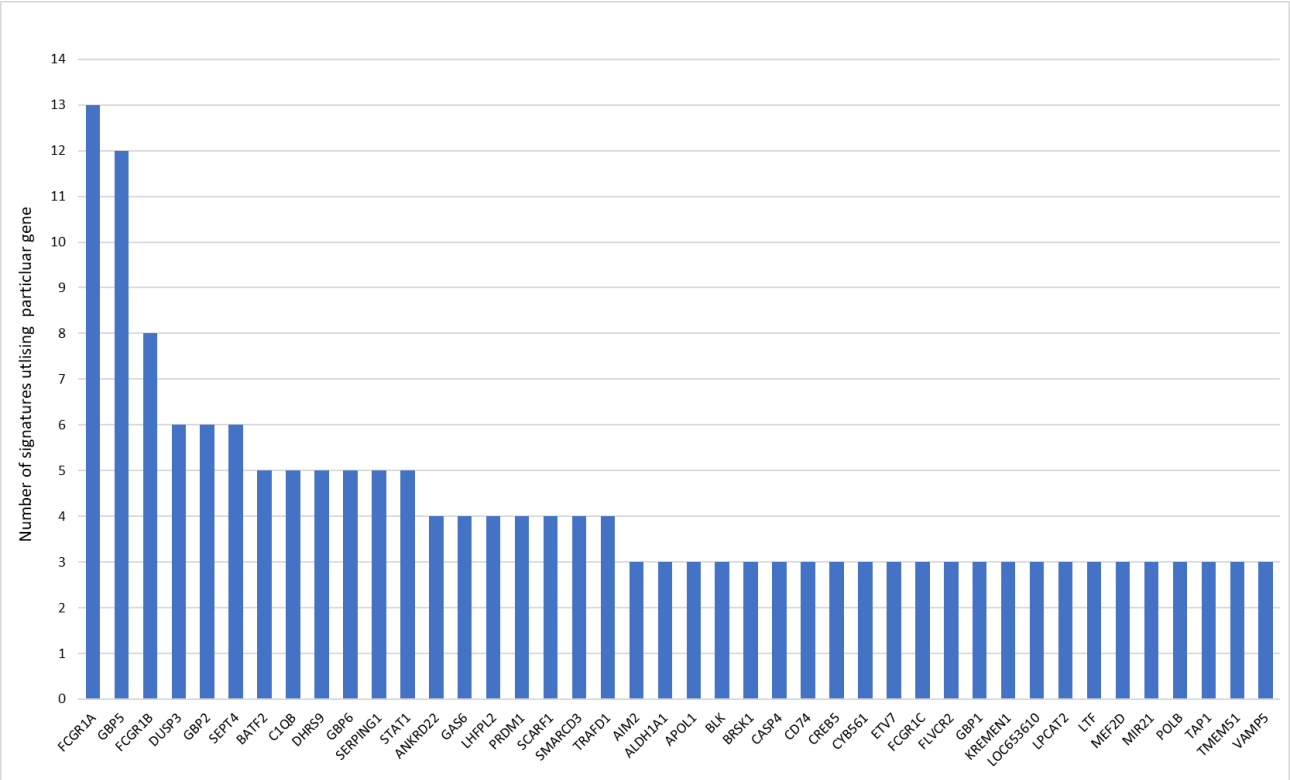
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	10
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	12
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	NA
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers)	12-16
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	16
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	17
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	44

8.6 Appendix 3.2: A matrix of all identified signatures with their corresponding gene composition.

This is a large excel spreadsheet file that cannot fit within the thesis and hence the DOI to the file is provided. It is a matrix of gene signatures and their frequency of use within the manuscripts included in the systematic review. The file contains 36 columns and 1027 rows. The file can be found on the DOI address given below.

https://doi.org/10.1371/journal.pone.0237574.s002

8.7 Appendix 3.3: Column chart of forty-two most frequently utilised genes in transcriptomic signatures of either TB diagnosis or prediction to TB disease.



8.8 Appendix 3.4: Characteristics of studies for diagnosis of TB included in the systematic review.

Ref	Study Name	Study Entry ID	Signature Name	Biomarker Description	Dataset	Signature discovery model *	Control Type	TB	Controls	Sens [95% CI]	Spec [95% CI]	AUC	Validation Set	Index Sample	TB reference standard	Population	Sample Country
23	Berry 2010	Berry2010a	Berry393_2010	393 transcript signature	Berry-GSE19491	K nearest neighbour	LTBI	20	31	0.95 [0.75, 1.00]	0.97 [0.83, 1.00]	NR	Y	Whole Blood	Culture	Adults	South Africa, Malawi
		Berry2010b	Berry86_2010	86 transcript signature	Berry-GSE19491		OD	20	96	0.90 [0.68, 0.99]	0.83 [0.74, 0.90]	NR	Y				
38	Bloom 2013	Bloom2013	Bloom144_2010	144 transcript signature	Bloom-GSE42834	Support vector machine	OD+HC	8	34	0.88 [0.47, 1.00]	0.91 [0.76, 0.98]	NR	Y	Whole Blood	Culture	Adults	South Africa
49	DaCosta 2015	DaCosta2015a	daCosta2_2015	GBP5; FCGR1A	Bloom-GSE42834	Random forest	OD	35	19	0.94 [0.81, 0.99]	0.84 [0.60, 0.97]	0.96	Y	Whole Blood	Culture	Adults	South Africa
		DaCosta2015b	daCosta3_2015	GBP5; FCGR1A; GZMA	Bloom-GSE42834		OD	35	19	0.94 [0.81, 0.99]	0.84 [0.60, 0.97]	0.96	Y				
42	Dawany 2015	Dawany2015a	Dewany251_2015	251 transcript signature	Berry-GSE19435	Support vector machine	HC	7	10	0.86 [0.42, 1.00]	1.00 [0.69, 1.00]	1.00	Y	Whole Blood	Culture	Adults	UK
		Dawany2015b	Dewany251_2015		Bloom-GSE40553		LTBI	29	38	0.86 [0.68, 0.96]	0.92 [0.79, 0.98]	0.97	Y				South Africa
		Dawany2015c	Dewany251_2015		Berry-GSE19439		LTBI+HC	13	29	0.69 [0.39, 0.91]	1.00 [0.88, 1.00]	0.94	Y				UK
		Dawany2015d	Dewany251_2015		Berry-GSE19444		LTBI+HC	21	33	0.76 [0.53, 0.92]	0.97 [0.84, 1.00]	0.89	Y				UK

		Dawany2015e	Dewany251_2015		Berry-GSE19442		LTBI	20	31	0.90 [0.68, 0.99]	0.90 [0.74, 0.98]	0.92	Y				South Africa
		Dawany2015f	Dewany251_2015		Berry-GSE19435		HC	7	12	0.86 [0.42, 1.00]	1.00 [0.74, 1.00]	0.98	Y				UK
43	DeAraujo 2016	DeAraujo2016a	DeAraujo1_2016	NPC2	Berry-GSE19491	Classification tree	HC	34	24	0.82 [0.65, 0.93]	0.96 [0.79, 1.00]	0.95	Y	Whole Blood	Culture	Adults	UK
		DeAraujo2016b	DeAraujo1_2016	NPC2	Berry-GSE19491		LTBI	20	31	0.85 [0.62, 0.97]	0.97 [0.83, 1.00]	0.97	Y				South Africa
		DeAraujo2016c	DeAraujo1_2016	NPC2	Maertzdolf-GSE34608		HC	8	18	1.00 [0.63, 1.00]	0.94 [0.73, 1.00]	0.99	Y				Germany
		DeAraujo2016d	DeAraujo1_2016	NPC2	Bloom-GSE42826		HC	11	52	0.91 [0.59, 1.00]	0.96 [0.87, 1.00]	0.99	Y				UK, France
44	Francisco 2017	Francisco2017a	Franscisco2_2017	GBP5; KLF2	NR	Random forest	OD	144	209	0.96 [0.91, 0.98]	0.85 [0.80, 0.90]	0.89	Y	Whole Blood	Culture & Smear	Adults	China
		Francisco2017b	Sweeney3_2016	GBP5; DUSP3; KLF2	NR		OD	144	209	0.74 [0.66, 0.81]	0.91 [0.86, 0.94]	0.71	Y	Whole Blood			
		Francisco2017c	Franscisco2_2017	GBP5; KLF2	NR		OD	275	290	0.78 [0.73, 0.83]	0.32 [0.27, 0.38]	0.54	Y	PBMC			
		Francisco2017d	Sweeney3_2016	GBP5; DUSP3; KLF2	NR		OD	275	290	0.61 [0.55, 0.67]	0.47 [0.41, 0.53]	0.53	Y	PBMC			
		Francisco2017e	Franscisco2_2017	GBP5; KLF2	NR		HC	144	209	0.78 [0.70, 0.84]	0.87 [0.82, 0.91]	0.86	Y	Whole Blood			
		Francisco2017f	Sweeney3_2016	GBP5; DUSP3; KLF2	NR		HC	144	209	0.85 [0.79, 0.91]	0.71 [0.64, 0.77]	0.85	Y	Whole Blood			
45	Haung 2015	Huang2015a	Huang13_2015	13 transcript signature	Bloom-GSE42825	Support vector machine	OD+HC	8	34	1.00 [0.63, 1.00]	0.94 [0.80, 0.99]	NR	Y	PBMC	Culture	Adults	UK, France
46	Huang 2018	Huang2018	Huang1_2018	hsa_circRNA_001937	NR	Hierarchical clustering	HC	115	90	0.72 [0.63, 0.80]	0.90 [0.82, 0.95]	0.85	Y	PBMC	Microbiological	Adults	China

24	Kaforou 2013	Kaforou2013a	Kaforou27_2013	27 transcript signature	Berry-GSE19491	Difference of means	LTBI	20	31	0.95 [0.75, 1.00]	0.94 [0.79, 0.99]	0.99	Y	Whole Blood	Culture	Adults	South Africa, Malawi
		Kaforou2013b	Kaforou44_2013	44 transcript signature	Kaforou-GSE37250		OD	20	83	1.00 [0.83, 1.00]	0.96 [0.90, 0.99]	1.00	Y				
		Kaforou2013c	Berry393_2010	393 transcript signature	Kaforou-GSE37250		LTBI	97	83	0.88 [0.79, 0.93]	0.84 [0.75, 0.91]	0.94	Y				
		Kaforou2013d	Berry86_2010	86 transcript signature	Kaforou-GSE37250		OD	97	83	0.71 [0.61, 0.80]	0.76 [0.65, 0.85]	0.78	Y				
50	Lee 2016	Lee2016a	Lee2_2016	PTPRC; ASUN	NR	Naïve Bayes	LTBI+HC	15	32	0.93 [0.68, 1.00]	1.00 [0.89, 1.00]	0.94	Y	PBMC	Smear	Adults	Taiwan
		Lee2016b	Lee3_2016	PTPRC; ASUN; DHX29	NR		LTBI+HC	15	31	1.00 [0.78, 1.00]	0.97 [0.83, 1.00]	0.98	Y	PBMC			
51	Leong 2018	Leong2018a	Berry393_2010	393 transcript signature	Leong-GSE101705	Ridge logistic regression	LTBI	28	16	0.93 [0.76, 0.99]	0.94 [0.70, 1.00]	0.99	Y	Whole Blood	Culture	Adults	India
		Leong2018b	Berry86_2010	86 transcript signature	Leong-GSE101705		LTBI	28	16	0.93 [0.76, 0.99]	0.94 [0.70, 1.00]	0.97	Y				
		Leong2018c	Jacobsen3_2007	RAB33A; FCGR1A; LTF	Leong-GSE101705		LTBI	28	16	0.86 [0.67, 0.96]	0.88 [0.62, 0.98]	0.90	Y				
		Leong2018d	Kaforou27_2013	27 transcript signature	Leong-GSE101705		LTBI	28	16	0.93 [0.76, 0.99]	0.94 [0.70, 1.00]	0.96	Y				
		Leong2018e	Sambarey10_2017	IFI44L; CYP4F3;	Leong-GSE101705		LTBI	28	16	0.89 [0.72, 0.98]	0.94 [0.70, 1.00]	0.96	Y				

				FCGR1A; TIMM10; BCL6; HK3; SMARCD3; RBBP8; RAB13; SLP1													
		Leong2018f	Sweeney3_2016	GBP5; DUSP3; KLF2	Leong-GSE101705		LTBI	28	16	0.86 [0.67, 0.96]	0.81 [0.54, 0.96]	0.90	Y				
		Leong2018g	Zak16_2016	16 transcript signature	Leong-GSE101705		LTBI	28	16	0.93 [0.76, 0.99]	0.94 [0.70, 1.00]	0.96	Y				
		Leong2018h	Leong24_2018	24 transcript signature	Leong-GSE101705		LTBI	28	16	0.93 [0.76, 0.99]	0.94 [0.70, 1.00]	0.98	N				
52	Lu 2011	Lu2011	Lu3_2011	CXCL10; ATP10A; TLR6	NR	Decision tree	LTBI	17	19	0.71 [0.44, 0.90]	0.89 [0.67, 0.99]	NR	Y	PBMC	Culture	Adolescents and Adults	China
54	Pan 2017	Pan2017a	Pan2_2017	RETN; KLK1	NR	Pairwise comparison	HC	66	86	0.85 [0.74, 0.92]	0.90 [0.81, 0.95]	0.94	Y	PBMC	Culture or 2 Smears	Adults	China
		Pan2017b	Pan2_2017	RETN; KLK1	NR		LTBI	66	78	0.71 [0.59, 0.82]	0.94 [0.86, 0.98]	0.92	Y				
56	Sambarey 2017	Sambarey2017a	Sambarey10_2017	I0 transcript signature (IFI44L; CYP4F3; FCGR1A; TIMM10; BCL6; HK3; SMARCD3;	Kaforou-GSE37250	Linear-discriminant analysis	LTBI	51	36	0.82 [0.69, 0.92]	0.92 [0.78, 0.98]	NR	Y	Whole Blood	Culture	Adults	Malawi
		Sambarey2017b	Sambarey10_2017		Kaforou-GSE37250		OD	51	34	0.92 [0.81, 0.98]	0.47 [0.30, 0.65]	NR	Y				Malawi
		Sambarey2017c	Sambarey10_2017		Kaforou-GSE37250		LTBI	47	47	0.79 [0.64, 0.89]	0.91 [0.80, 0.98]	NR	Y				South Africa
		Sambarey2017d	Sambarey10_2017		Kaforou-GSE37250		OD	47	49	0.68 [0.53, 0.81]	0.80 [0.66, 0.90]	NR	Y				South Africa

				RBBP8; RAB13; SLPI)													
58	Sweeney 2016	Sweeney2016a	Sweeney3_2016	GBP5; DUSP3; KLF2	Maertzdolf-GSE28623	Forward search	HC	46	37	0.85 [0.71, 0.94]	0.81 [0.65, 0.92]	NR	Y	Whole Blood	Smear + CXR	Adults	Gambia
		Sweeney2016b	Sweeney3_2016		Maertzdolf-GSE34608		HC	8	18	1.00 [0.63, 1.00]	1.00 [0.81, 1.00]	NR	Y		Culture	Adults	Germany
		Sweeney2016c	Sweeney3_2016		Ottenhoff-GSE56153		HC	18	18	0.61 [0.36, 0.83]	0.67 [0.41, 0.87]	NR	Y		Smear	Adolescents and adults	Indonesia
		Sweeney2016d	Sweeney3_2016		Maertzdolf-GSE28623		LTBI	46	25	0.87 [0.74, 0.95]	0.84 [0.64, 0.95]	NR	Y		Smear + CXR	Adults	Gambia
		Sweeney2016e	Sweeney3_2016		Maertzdolf-GSE34608		OD	8	18	0.50 [0.16, 0.84]	0.61 [0.36, 0.83]	NR	Y		Culture	Adults	Germany
59	Walter 2016	Walter2016a	Walter47_2016	47 transcript signature	Kaforou-GSE37250	Support vector machine	OD	97	83	0.90 [0.82, 0.95]	0.77 [0.67, 0.86]	0.91	Y	Whole Blood	Culture	Adults	South Africa, Malawi
		Walter2016b	Walter51_2016	51 transcript signature	Kaforou-GSE37250		LTBI	97	83	0.91 [0.83, 0.96]	0.87 [0.78, 0.93]	0.95	Y				
		Walter2016c	Walter119_2016	119 transcript signature	Kaforou-GSE37250		OD+LTBI	117	146	0.88 [0.81, 0.93]	0.80 [0.73, 0.86]	0.91	Y				
		Walter2016d	Berry393_2010	393 transcript signature	Walter-GSE73408		LTBI	35	35	0.89 [0.73, 0.97]	0.94 [0.81, 0.99]	0.94	Y	Whole Blood	Culture + Smear	Adults	USA
		Walter2016e	Berry86_2010	86 transcript signature	Walter-GSE73408		OD	35	39	0.91 [0.77, 0.98]	0.74 [0.58, 0.87]	0.90	Y				

		Walter2016f	Bloom144_2010	144 transcript signature	Walter-GSE73408		OD+LTBI	35	74	0.86 [0.70, 0.95]	0.77 [0.66, 0.86]	0.91	Y				
		Walter2016g	Kaforou27_2013	27 transcript signature	Walter-GSE73408		LTBI	35	35	0.94 [0.81, 0.99]	0.91 [0.77, 0.98]	0.98	Y				
		Walter2016h	Kaforou44_2013	44 transcript signature	Walter-GSE73408		OD	35	39	0.69 [0.51, 0.83]	0.79 [0.64, 0.91]	0.83	Y				
		Walter2016i	Kaforou53_2013	53 transcript signature	Walter-GSE73408		OD+LTBI	35	74	0.77 [0.60, 0.90]	0.76 [0.64, 0.85]	0.83	Y				
60	Warsinske 2018	Warsinske2018a	Sweeney3_2016	GBP5; DUSP3; KLF2	Zak-GSE79362	Forward search	OD+HC+LTBI	176	187	0.90 [0.84, 0.94]	0.70 [0.63, 0.77]	NR	Y	Whole Blood	Culture	Adolescents and adults	China
25	Zak 2016	Zak2016a	Zak16_2016	16 transcript signature (GBP2; FCGR1A; FCGR1B; GBP5; STAT1; SERPING1; ANKRD22; BATF2; GBP1; APOL1; TRAFD1; SCARF1; SEPT4;	Berry-GSE19444	Support vector machine	LTBI	21	21	0.90 [0.70, 0.99]	0.52 [0.30, 0.74]	0.86	Y	Whole Blood	Culture	Adults	UK
		Zak2016b	Zak16_2016		Berry-GSE19442		LTBI	20	30	0.90 [0.68, 0.99]	1.00 [0.88, 1.00]	0.99	Y				South Africa
		Zak2016c	Zak16_2016		Berry-GSE19444		HC	21	12	0.90 [0.70, 0.99]	0.67 [0.35, 0.90]	0.91	Y				UK
		Zak2016d	Zak16_2016		Kaforou-GSE37250		LTBI	51	35	0.90 [0.79, 0.97]	0.71 [0.54, 0.85]	0.91	Y				Malawi
		Zak2016e	Zak16_2016		Kaforou-GSE37250		LTBI	46	48	0.91 [0.79, 0.98]	0.94 [0.83, 0.99]	0.94	Y				SA
		Zak2016f	Zak16_2016		Bloom-GSE42825/26/30		HC	35	113	0.91 [0.77, 0.98]	0.98 [0.94, 1.00]	0.99	Y				UK, France

		Zak2016g	Zak16_2016	TAP1; ETV7; GPB4)	Bloom-GSE42825/26/30		OD	35	16	0.91 [0.77, 0.98]	0.88 [0.62, 0.98]	0.95	Y				UK, France
		Zak2016h	Zak16_2016		Bloom-GSE42825/26/30		OD	35	14	0.89 [0.73, 0.97]	0.93 [0.66, 1.00]	0.91	Y				UK, France
		Zak2016i	Zak16_2016		Bloom-GSE42825/26/30		OD	35	61	0.91 [0.77, 0.98]	0.62 [0.49, 0.74]	0.83	Y				UK, France
		Zak2016j	Zak16_2016		Kaforou-GSE37250		OD	51	34	0.90 [0.79, 0.97]	0.47 [0.30, 0.65]	0.74	Y				Malawi
		Zak2016k	Zak16_2016		Kaforou-GSE37250		OD	46	49	0.89 [0.76, 0.96]	0.78 [0.63, 0.88]	0.83	Y				South Africa
		Zak2016l	Zak16_2016		Bloom-GSE40553		LTBI	29	38	0.90 [0.73, 0.98]	0.92 [0.79, 0.98]	0.98	Y				UK, France
40	Cai 2014	Cai2014a	Cai1_2014	C1QC	Cai-GSE54992	Support vector machine	HC	162	45	0.83 [0.76, 0.88]	0.89 [0.76, 0.96]	0.93	N	PBMC	Culture	Adults	China
		Cai2014b	Cai1_2014		Cai-GSE54992		LTBI	165	162	0.74 [0.67, 0.80]	0.82 [0.75, 0.88]	0.84	N				
41	Darboe 2018	Darboe2018	Darboe11_2018	(GBP2; FCGR1B; GBP5; STAT1; SERPING1; BATF2; GBP1; TRAFD1; SCARF1; TAP1;ETV7	NR	Support vector machine	LTBI	30	30	1.00 [0.88, 1.00]	0.80 [0.61, 0.92]	0.97	N	PBMC	Xpert MTB/RIF	Adults	Indonesia
62	Lesho 2011	Lesho2011	Lesho127_2011	127 transcript signature	NR	Supervised learning algorithms	LTBI+HC	5	18	1.00 [0.48, 1.00]	1.00 [0.81, 1.00]	NR	N	PBMC	Culture	Adults	USA

53	Maertzdorf 2011	Maertzdorf3_2011	Maertzdorf3_2011	CD64; LTF; RAB33A	Maertzdolf-GSE25534	Random Forest	LTBI	32	34	0.88 [0.71, 0.96]	0.91 [0.76, 0.98]	NR	N	Whole Blood	Culture	Adolescents and adults	SA
		Maertzdorf5_2011	Maertzdorf5_2011	FCGR1B; CD64;GBP5 LTF; GZMA	Maertzdolf-GSE25534		LTBI	32	34	0.94 [0.79, 0.99]	0.97 [0.85, 1.00]	NR	N				
39	Satproedprai 2015	Satproedprai2015	Satproedprai7_2015	FCGR1A; FCGR1B variant 1; FCGR1B variant 2; MAFB; APOL1; STAT1; KAZN	NR	Logistic regression	HC	40	56	0.82 [0.67, 0.93]	1.00 [0.94, 1.00]	0.97	N	Whole Blood	Culture + Smear	Adults	Thailand
57	Serrano 2016	Serrano2016	Serrano2_2016	NCF1; ORM	NR	Analysis of variance	LTBI+HC	10	20	0.90 [0.55, 1.00]	0.80 [0.56, 0.94]	NR	N	PBMC	Culture + Smear	Adults	Mexico
61	Wu 2007	Wu2007	Wu3_2007	IL-8; FOXP3; IL-12β,	NR	Logistic regression	LTBI	30	24	0.97 [0.83, 1.00]	0.88 [0.68, 0.97]	0.97	N	PBMC	Culture or CXR	Adults	USA

AUC; Area under the curve, UK; United Kingdom, USA; United States of America, Sens; Sensitivity, Spec; Specificity, LTBI; Latent TB infection, HHC; Household TB Contact. * Signature discovery method applies to the signature that was discovered in that article.

8.9 Appendix 3.5: Methodological quality summary of all diagnostic studies with independent validation cohort.

Study Name	Signature	Patient Selection		Index Test		Reference Standard		Flow & Timing
		Risk of Bias	Applicability Concerns	Risk of Bias	Applicability Concerns	Risk of Bias	Applicability Concerns	Risk of Bias
Berry2010a	Berry393_2010	LR	HR	UC	UC	LR	LR	LR
Berry2010b	Berry86_2010	LR	LR	UC	UC	LR	LR	LR
Bloom2013	Bloom144_2010	HR	LR	UC	UC	LR	LR	LR
daCosta2015a	daCosta2_2015	HR	LR	UC	UC	LR	LR	LR
daCosta2015b	daCosta3_2015	HR	LR	UC	UC	LR	LR	LR
Dawany2015a	Dewany251_2015	HR	HR	UC	UC	LR	LR	LR
Dawany2015b	Dewany251_2015	HR	HR	UC	UC	LR	LR	LR
Dawany2015c	Dewany251_2015	HR	HR	UC	UC	LR	LR	LR
Dawany2015d	Dewany251_2015	HR	HR	UC	UC	LR	LR	LR
Dawany2015e	Dewany251_2015	HR	HR	UC	UC	LR	LR	LR
Dawany2015f	Dewany251_2015	HR	HR	UC	UC	LR	LR	LR
DeAraujo2016a	DeAraujo1_2016	HR	HR	LR	LR	LR	LR	LR
DeAraujo2016b	DeAraujo1_2016	HR	HR	LR	LR	LR	LR	LR
DeAraujo2016c	DeAraujo1_2016	HR	HR	LR	LR	LR	LR	LR
DeAraujo2016d	DeAraujo1_2016	HR	HR	LR	LR	LR	LR	LR
Francisco2017a	Franscisco2_2017	HR	HR	UC	UC	LR	LR	LR
Francisco2017b	Sweeney3_2016	HR	LR	UC	UC	LR	LR	LR
Francisco2017c	Franscisco2_2017	HR	HR	UC	UC	LR	LR	LR
Francisco2017d	Sweeney3_2016	HR	LR	UC	UC	LR	LR	LR
Francisco2017e	Franscisco2_2017	HR	LR	UC	UC	LR	LR	LR
Francisco2017f	Sweeney3_2016	HR	LR	UC	UC	LR	LR	LR
Haung2015a	Huang13_2015	HR	LR	UC	UC	LR	LR	LR
Huang2018	Huang1_2018	HR	HR	UC	UC	UC	UC	UC
Kaforou2013a	Kaforou27_2013	HR	HR	UC	UC	LR	LR	LR
Kaforou2013b	Kaforou44_2013	HR	LR	UC	UC	LR	LR	LR
Kaforou2013c	Berry393_2010	HR	HR	UC	UC	LR	LR	LR
Kaforou2013d	Berry86_2010	HR	LR	UC	UC	LR	LR	LR
Lee2016a	Lee2_2016	HR	HR	UC	UC	LR	LR	LR
Lee2016b	Lee3_2016	HR	HR	UC	UC	LR	LR	LR
Leong2018a	Berry393_2010	LR	HR	UC	UC	LR	LR	LR
Leong2018b	Berry86_2010	LR	HR	UC	UC	LR	LR	LR
Leong2018c	Jacobsen3_2007	LR	HR	UC	UC	LR	LR	LR
Leong2018d	Kaforou27_2013	LR	HR	UC	UC	LR	LR	LR
Leong2018e	Sambarey10_2017	LR	HR	UC	UC	LR	LR	LR
Leong2018f	Sweeney3_2016	LR	HR	UC	UC	LR	LR	LR
Leong2018g	Zak16_2016	LR	HR	UC	UC	LR	LR	LR
Lu2011	Lu3_2011	HR	HR	UC	UC	LR	LR	LR
Pan2017a	Pan2_2017	HR	HR	UC	UC	LR	LR	LR
Pan2017b	Pan2_2017	HR	HR	UC	UC	LR	LR	LR
Sambarey2017a	Sambarey10_2017	HR	HR	UC	UC	LR	LR	LR
Sambarey2017b	Sambarey10_2017	HR	LR	UC	UC	LR	LR	LR
Sambarey2017c	Sambarey10_2017	HR	HR	UC	UC	LR	LR	LR
Sambarey2017d	Sambarey10_2017	HR	LR	UC	UC	LR	LR	LR
Sweeney2016a	Sweeney3_2016	HR	HR	UC	UC	LR	LR	LR
Sweeney2016b	Sweeney3_2016	HR	HR	UC	UC	LR	LR	LR
Sweeney2016c	Sweeney3_2016	HR	HR	UC	UC	LR	LR	LR
Sweeney2016d	Sweeney3_2016	HR	HR	UC	UC	LR	LR	LR
Sweeney2016e	Sweeney3_2016	HR	LR	UC	UC	LR	LR	LR
Walter2016a	Walter47_2016	HR	LR	UC	UC	LR	LR	LR
Walter2016b	Walter51_2016	HR	HR	UC	UC	LR	LR	LR
Walter2016c	Walter119_2016	HR	LR	UC	UC	LR	LR	LR
Walter2016d	Berry393_2010	HR	HR	UC	UC	LR	LR	LR
Walter2016e	Berry86_2010	HR	LR	UC	UC	LR	LR	LR
Walter2016f	Bloom144_2010	HR	LR	UC	UC	LR	LR	LR
Walter2016g	Kaforou27_2013	HR	HR	UC	UC	LR	LR	LR
Walter2016h	Kaforou44_2013	HR	LR	UC	UC	LR	LR	LR
Walter2016i	Kaforou53_2013	HR	LR	UC	UC	LR	LR	LR
Warsinske2018a	Sweeney3_2016	HR	LR	UC	UC	LR	LR	HR
Zak2016a	Zak16_2016	LR	HR	UC	UC	LR	LR	LR
Zak2016b	Zak16_2016	LR	HR	UC	UC	LR	LR	LR
Zak2016c	Zak16_2016	LR	HR	UC	UC	LR	LR	LR
Zak2016d	Zak16_2016	LR	HR	UC	UC	LR	LR	LR
Zak2016e	Zak16_2016	LR	HR	UC	UC	LR	LR	LR
Zak2016f	Zak16_2016	LR	HR	UC	UC	LR	LR	LR
Zak2016g	Zak16_2016	LR	LR	UC	UC	LR	LR	LR
Zak2016h	Zak16_2016	LR	LR	UC	UC	LR	LR	LR
Zak2016i	Zak16_2016	LR	LR	UC	UC	LR	LR	LR
Zak2016j	Zak16_2016	LR	LR	UC	UC	LR	LR	LR
Zak2016k	Zak16_2016	LR	LR	UC	UC	LR	LR	LR
Zak2016l	Zak16_2016	LR	HR	UC	UC	LR	LR	LR

LR; Low risk, HR; High risk, UC; Unclear

8.10 Appendix 3.6: Methodological quality summary of all diagnostic studies with independent validation cohort

Study Name	Signature	Patient Selection		Index Test		Reference Standard		Flow & Timing
		Risk of Bias	Applicability Concerns	Risk of Bias	Applicability Concerns	Risk of Bias	Applicability Concerns	Risk of Bias
Cai2014a	Cai1_2014	HR	HR	UC	UC	LR	LR	LR
Cai2014b	Cai1_2014	HR	HR	UC	UC	LR	LR	LR
Darboe2018	Darboe11_2018	HR	HR	UC	UC	LR	LR	LR
Leong2018h	Leong24_2018	HR	HR	UC	UC	LR	LR	LR
Lesho2011	Lesho127_2011	HR	HR	UC	UC	LR	LR	LR
Maertzdolf2011a	Maertzdolf3_2011	HR	HR	UC	UC	LR	LR	LR
Maertzdolf2011b	Maertzdolf5_2011	HR	HR	UC	UC	LR	LR	LR
Satproedprai2015	Satproedprai7_2015	HR	HR	UC	UC	LR	LR	LR
Serrano2016	Serrano2_2016	HR	HR	UC	UC	LR	LR	LR
Wu2007	Wu10_2007	HR	HR	UC	UC	LR	LR	LR

LR; Low risk, HR; High risk, UC; Unclear

8.11 Appendix 3.7: Characteristics of studies for predicting progression to TB disease included in the systematic review.

Ref.	Study Name	Study Entry ID	Signature Name	Biomarker Description	Dataset	Signature discovery model	Population type	Total enrolled	Progressed to TB	Did not Progress to TB	AUC	Index Sample	TB reference standard	Population	Sample Country
55	Roe 2019	Roe2019b	Roe3_2019	BATF2; GBP5; SCARF1;	Roe-E-MTAB6385	Stability selection	HHC	333	6	327	NR	Whole blood	Culture	Adults	UK
26	Suliman 2018	Suliman2018b	Suliman4_2018	GAS6; SEPT4; CD1C; BLK	Zak-GSE79362	Pair ratio	LTBI	145	41	104	0.69	Whole blood	Culture or 2 Smears	Adults	South Africa
60	Warsinske 2018	Warsinske2018b	Sweeney3_2016	GBP5; DUSP3; KLF2	Zak-GSE79362	Forward search	LTBI	144	43	101	0.86	Whole blood	Culture or 2 Smears	Adolescents	South Africa
25	Zak 2016	Zak2016p	Zak16_2016	Refer to 25 in Appendix 3.1	Suliman-GSE94438	Support vector machine	HHC	374	73	301	0.72	Whole blood	Culture & Smear	Adults	USA

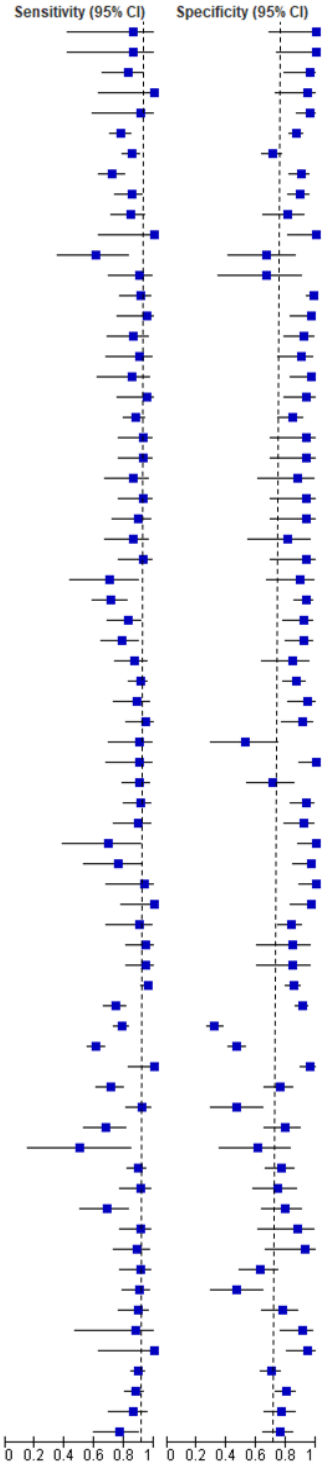
AUC; Area under the curve, UK; United Kingdom, USA; United States of America, LTBI; Latent TB infection, HC; Healthy control, OD; Other diseases

8.12 Appendix 3.8: Methodological quality summary of studies of prediction to TB disease in independent validation cohorts.

Study Name	Signature	Patient Selection		Index Test		Reference Standard		Flow & Timing
		Risk of Bias	Applicability Concerns	Risk of Bias	Applicability Concerns	Risk of Bias	Applicability Concerns	Risk of Bias
Roe2019b	Roe3_2019	LR	LR	UC	UC	HR	HR	LR
Suliman2018b	Suliman4_2018	LR	LR	LR	LR	LR	LR	LR
Warsinske2018b	Sweeney3_2016	LR	LR	LR	LR	LR	LR	LR
Zak2016p	Zak16_2016	LR	LR	LR	LR	LR	LR	LR

8.13 Appendix 3.9: Forest plots of sensitivity and specificity of mRNA transcriptomic signatures for diagnosis of TB disease in independent validation cohorts (all studies

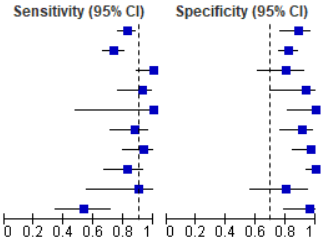
Study	TP	FP	FN	TN	Comparator	Dataset	Signature	Sensitivity (95% CI)	Specificity (95% CI)
Dawany2015a	6	0	1	10	HC	Berry-GSE19435	Dewany251_2015	0.86 [0.42, 1.00]	1.00 [0.69, 1.00]
Dawany2015f	6	0	1	12	HC	Berry-GSE19435	Dewany251_2015	0.86 [0.42, 1.00]	1.00 [0.74, 1.00]
DeAraujo2016a	28	1	6	23	HC	Berry-GSE19491	DeAraujo1_2016	0.82 [0.65, 0.93]	0.96 [0.79, 1.00]
DeAraujo2016c	8	1	0	17	HC	Maertzdolf-GSE34608	DeAraujo1_2016	1.00 [0.63, 1.00]	0.94 [0.73, 1.00]
DeAraujo2016d	10	2	1	50	HC	Bloom-GSE42826	DeAraujo1_2016	0.91 [0.59, 1.00]	0.96 [0.87, 1.00]
Francisco2017e	112	27	32	182	HC	Not-Publicly-Available	Franscisco2_2017	0.78 [0.70, 0.84]	0.87 [0.82, 0.91]
Francisco2017f	123	61	21	148	HC	Not-Publicly-Available	Sweeney3_2016	0.85 [0.79, 0.91]	0.71 [0.64, 0.77]
Huang2018	83	9	32	81	HC	Not-Publicly-Available	Huang1_2018	0.72 [0.63, 0.80]	0.90 [0.82, 0.95]
Pan2017a	56	9	10	77	HC	Not-Publicly-Available	Pan2_2017	0.85 [0.74, 0.92]	0.90 [0.81, 0.95]
Sweeney2016a	39	7	7	30	HC	Maertzdolf-GSE28623	Sweeney3_2016	0.85 [0.71, 0.94]	0.81 [0.65, 0.92]
Sweeney2016b	8	0	0	18	HC	Maertzdolf-GSE34608	Sweeney3_2016	1.00 [0.63, 1.00]	1.00 [0.81, 1.00]
Sweeney2016c	11	6	7	12	HC	Ottenhoff-GSE56153	Sweeney3_2016	0.61 [0.36, 0.83]	0.67 [0.41, 0.87]
Zak2016c	19	4	2	8	HC	Berry-GSE19444	Zak16_2016	0.90 [0.70, 0.99]	0.67 [0.35, 0.90]
Zak2016f	32	2	3	111	HC	Bloom-GSE42825/26/30	Zak16_2016	0.91 [0.77, 0.98]	0.98 [0.94, 1.00]
Berry2010a	19	1	1	30	LTBI	Berry-GSE19491	Berry393_2010	0.95 [0.75, 1.00]	0.97 [0.83, 1.00]
Dawany2015b	25	3	4	35	LTBI	Bloom-GSE40553	Dewany251_2015	0.86 [0.68, 0.96]	0.92 [0.79, 0.98]
Dawany2015e	18	3	2	28	LTBI	Berry-GSE19442	Dewany251_2015	0.90 [0.68, 0.99]	0.90 [0.74, 0.98]
DeAraujo2016b	17	1	3	30	LTBI	Berry-GSE19491	DeAraujo1_2016	0.85 [0.62, 0.97]	0.97 [0.83, 1.00]
Kaforou2013a	19	2	1	29	LTBI	Berry-GSE19491	Kaforou27_2013	0.95 [0.75, 1.00]	0.94 [0.79, 0.99]
Kaforou2013c	85	13	12	70	LTBI	Kaforou-GSE37250	Berry393_2010	0.88 [0.79, 0.93]	0.84 [0.75, 0.91]
Leong2018a	26	1	2	15	LTBI	Leong-GSE101705	Berry393_2010	0.93 [0.76, 0.99]	0.94 [0.70, 1.00]
Leong2018b	26	1	2	15	LTBI	Leong-GSE101705	Berry86_2010	0.93 [0.76, 0.99]	0.94 [0.70, 1.00]
Leong2018c	24	2	4	14	LTBI	Leong-GSE101705	Jacobsen3_2007	0.86 [0.67, 0.96]	0.88 [0.62, 0.98]
Leong2018d	26	1	2	15	LTBI	Leong-GSE101705	Kaforou27_2013	0.93 [0.76, 0.99]	0.94 [0.70, 1.00]
Leong2018e	25	1	3	15	LTBI	Leong-GSE101705	Sambarey10_2017	0.89 [0.72, 0.98]	0.94 [0.70, 1.00]
Leong2018f	24	3	4	13	LTBI	Leong-GSE101705	Sweeney3_2016	0.86 [0.67, 0.96]	0.81 [0.54, 0.96]
Leong2018g	26	1	2	15	LTBI	Leong-GSE101705	Zak16_2016	0.93 [0.76, 0.99]	0.94 [0.70, 1.00]
Lu2011	12	2	5	17	LTBI	Not-Publicly-Available	Lu3_2011	0.71 [0.44, 0.90]	0.89 [0.67, 0.99]
Pan2017b	47	5	19	73	LTBI	Not-Publicly-Available	Pan2_2017	0.71 [0.59, 0.82]	0.94 [0.86, 0.98]
Sambarey2017a	42	3	9	33	LTBI	Kaforou-GSE37250	Sambarey10_2017	0.82 [0.69, 0.92]	0.92 [0.78, 0.98]
Sambarey2017c	37	4	10	43	LTBI	Kaforou-GSE37250	Sambarey10_2017	0.79 [0.64, 0.89]	0.91 [0.80, 0.98]
Sweeney2016d	40	4	6	21	LTBI	Maertzdolf-GSE28623	Sweeney3_2016	0.87 [0.74, 0.95]	0.84 [0.64, 0.95]
Walter2016b	88	11	9	72	LTBI	Kaforou-GSE37250	Walter51_2016	0.91 [0.83, 0.96]	0.87 [0.78, 0.93]
Walter2016d	31	2	4	33	LTBI	Walter-GSE73408	Berry393_2010	0.89 [0.73, 0.97]	0.94 [0.81, 0.99]
Walter2016g	33	3	2	32	LTBI	Walter-GSE73408	Kaforou27_2013	0.94 [0.81, 0.99]	0.91 [0.77, 0.98]
Zak2016a	19	10	2	11	LTBI	Berry-GSE19444	Zak16_2016	0.90 [0.70, 0.99]	0.52 [0.30, 0.74]
Zak2016b	18	0	2	30	LTBI	Berry-GSE19442	Zak16_2016	0.90 [0.68, 0.99]	1.00 [0.88, 1.00]
Zak2016d	46	10	5	25	LTBI	Kaforou-GSE37250	Zak16_2016	0.90 [0.79, 0.97]	0.71 [0.54, 0.85]
Zak2016e	42	3	4	45	LTBI	Kaforou-GSE37250	Zak16_2016	0.91 [0.79, 0.98]	0.94 [0.83, 0.99]
Zak2016l	26	3	3	35	LTBI	Bloom-GSE40553	Zak16_2016	0.90 [0.73, 0.98]	0.92 [0.79, 0.98]
Dawany2015c	9	0	4	29	LTBI+HC	Berry-GSE19439	Dewany251_2015	0.69 [0.39, 0.91]	1.00 [0.88, 1.00]
Dawany2015d	16	1	5	32	LTBI+HC	Berry-GSE19444	Dewany251_2015	0.76 [0.53, 0.92]	0.97 [0.84, 1.00]
Lee2016a	14	0	1	32	LTBI+HC	Not-Publicly-Available	Lee2_2016	0.93 [0.68, 1.00]	1.00 [0.89, 1.00]
Lee2016b	15	1	0	30	LTBI+HC	Not-Publicly-Available	Lee3_2016	1.00 [0.78, 1.00]	0.97 [0.83, 1.00]
Berry2010b	18	16	2	80	OD	Berry-GSE19491	Berry86_2010	0.90 [0.68, 0.99]	0.83 [0.74, 0.90]
daCosta2015a	33	3	2	16	OD	Bloom-GSE42834	daCosta2_2015	0.94 [0.81, 0.99]	0.84 [0.60, 0.97]
daCosta2015b	33	3	2	16	OD	Bloom-GSE42834	daCosta3_2015	0.94 [0.81, 0.99]	0.84 [0.60, 0.97]
Francisco2017a	138	31	6	178	OD	Not-Publicly-Available	Franscisco2_2017	0.96 [0.91, 0.98]	0.85 [0.80, 0.90]
Francisco2017b	107	19	37	190	OD	Not-Publicly-Available	Sweeney3_2016	0.74 [0.66, 0.81]	0.91 [0.86, 0.94]
Francisco2017c	215	197	60	93	OD	Not-Publicly-Available	Franscisco2_2017	0.78 [0.73, 0.83]	0.32 [0.27, 0.38]
Francisco2017d	169	154	106	136	OD	Not-Publicly-Available	Sweeney3_2016	0.61 [0.55, 0.67]	0.47 [0.41, 0.53]
Kaforou2013b	20	3	0	80	OD	Kaforou-GSE37250	Kaforou44_2013	1.00 [0.83, 1.00]	0.96 [0.90, 0.99]
Kaforou2013d	69	20	28	63	OD	Kaforou-GSE37250	Berry86_2010	0.71 [0.61, 0.80]	0.76 [0.65, 0.85]
Sambarey2017b	47	18	4	16	OD	Kaforou-GSE37250	Sambarey10_2017	0.92 [0.81, 0.98]	0.47 [0.30, 0.65]
Sambarey2017d	32	10	15	39	OD	Kaforou-GSE37250	Sambarey10_2017	0.68 [0.53, 0.81]	0.80 [0.66, 0.90]
Sweeney2016e	4	7	4	11	OD	Maertzdolf-GSE34608	Sweeney3_2016	0.50 [0.16, 0.84]	0.61 [0.36, 0.83]
Walter2016a	87	19	10	64	OD	Kaforou-GSE37250	Walter47_2016	0.90 [0.82, 0.95]	0.77 [0.67, 0.86]
Walter2016e	32	10	3	29	OD	Walter-GSE73408	Berry86_2010	0.91 [0.77, 0.98]	0.74 [0.58, 0.87]
Walter2016h	24	8	11	31	OD	Walter-GSE73408	Kaforou44_2013	0.69 [0.51, 0.83]	0.79 [0.64, 0.91]
Zak2016g	32	2	3	14	OD	Bloom-GSE42825/26/30	Zak16_2016	0.91 [0.77, 0.98]	0.88 [0.62, 0.98]
Zak2016h	31	1	4	13	OD	Bloom-GSE42825/26/30	Zak16_2016	0.89 [0.73, 0.97]	0.93 [0.66, 1.00]
Zak2016i	32	23	3	38	OD	Bloom-GSE42825/26/30	Zak16_2016	0.91 [0.77, 0.98]	0.62 [0.49, 0.74]
Zak2016j	46	18	5	16	OD	Kaforou-GSE37250	Zak16_2016	0.90 [0.79, 0.97]	0.47 [0.30, 0.65]
Zak2016k	41	11	5	38	OD	Kaforou-GSE37250	Zak16_2016	0.89 [0.76, 0.96]	0.78 [0.63, 0.88]
Bloom2013	7	3	1	31	OD+HC	Bloom-GSE42834	Bloom144_2010	0.88 [0.47, 1.00]	0.91 [0.76, 0.98]
Huang2015a	8	2	0	32	OD+HC	Bloom-GSE42825	Huang13_2015	1.00 [0.63, 1.00]	0.94 [0.80, 0.99]
Warsinske2018a	158	56	18	131	OD+HC+LTBI	Zak-GSE79362	Sweeney3_2016	0.90 [0.84, 0.94]	0.70 [0.63, 0.77]
Walter2016c	103	29	14	117	OD+LTBI	Kaforou-GSE37250	Walter119_2016	0.88 [0.81, 0.93]	0.80 [0.73, 0.86]
Walter2016f	30	17	5	57	OD+LTBI	Walter-GSE73408	Bloom144_2010	0.86 [0.70, 0.95]	0.77 [0.66, 0.86]
Walter2016i	27	18	8	56	OD+LTBI	Walter-GSE73408	Kaforou53_2013	0.77 [0.60, 0.90]	0.76 [0.64, 0.85]



HC; Healthy Controls, LTBI; Latent TB infection, OD; Other diseases. Each signature validated in a different population is represented as a separate entry. Vertical dotted lines correspond to 90% sensitivity and 70% specificity

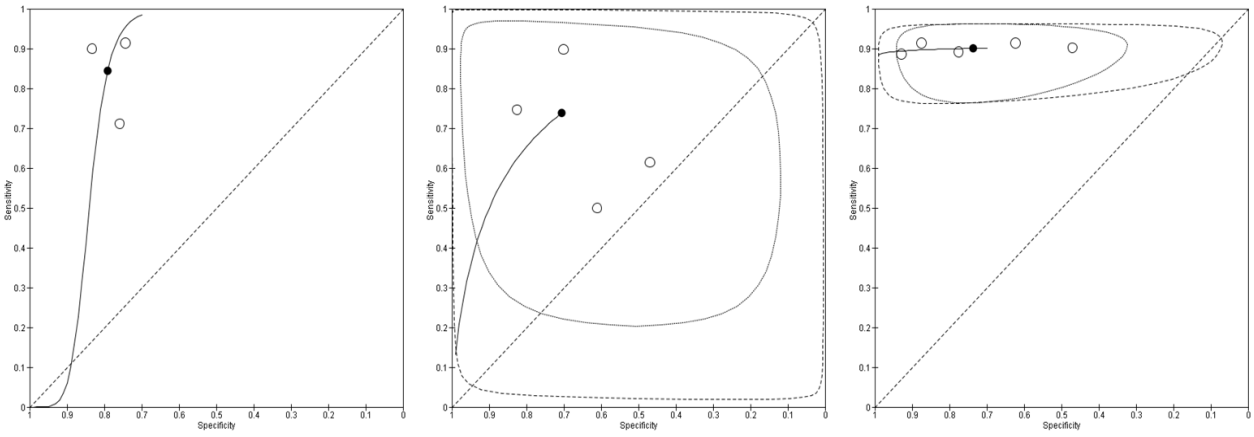
8.14 Appendix 3.10: Forest plots of sensitivity and specificity of mRNA transcriptomic signatures for diagnosis of TB disease for studies without independent validation cohorts.

Study	TP	FP	FN	TN	Comparator	Dataset	Signature	Sensitivity (95% CI)	Specificity (95% CI)
Cai2014a	134	5	28	40	HC	Cai-GSE54992	Cai1_2014	0.83 [0.76, 0.88]	0.89 [0.76, 0.96]
Cai2014b	122	29	43	133	LTBI	Cai-GSE54992	Cai1_2014	0.74 [0.67, 0.80]	0.82 [0.75, 0.88]
Darboe2018	30	6	0	24	LTBI	Not-Publicly-Available	Darboe11_2018	1.00 [0.88, 1.00]	0.80 [0.61, 0.92]
Leong2018h	26	1	2	15	LTBI	Leong-GSE101705	Leong24_2018	0.93 [0.76, 0.99]	0.94 [0.70, 1.00]
Lesho2011	5	0	0	18	LTBI+HC	Not-Publicly-Available	Lesho127_2011	1.00 [0.48, 1.00]	1.00 [0.81, 1.00]
Maertzdorf2011a	28	3	4	31	LTBI	Maertzdolf-GSE25534	Maertzdorf3_2011	0.88 [0.71, 0.96]	0.91 [0.76, 0.98]
Maertzdorf2011b	30	1	2	33	LTBI	Maertzdolf-GSE25534	Maertzdorf5_2011	0.94 [0.79, 0.99]	0.97 [0.85, 1.00]
Satproedprai2015	33	0	7	56	OD+HC+LTBI	Berry-GSE19491	Satproedprai7_2015	0.82 [0.67, 0.93]	1.00 [0.94, 1.00]
Serrano2016	9	4	1	16	LTBI+HC	Not-Publicly-Available	Serano2_2016	0.90 [0.55, 1.00]	0.80 [0.56, 0.94]
Wu2007	16	1	14	23	LTBI	Not-Publicly-Available	Wu10_2007	0.53 [0.34, 0.72]	0.96 [0.79, 1.00]



HC; Healthy Controls, LTBI; Latent TB infection, OD; Other diseases. Each signature validated in a different population is represented as a separate entry. Vertical dotted lines correspond to 90% sensitivity and 70% specificity

8.15 Appendix 3.11: Summary receiver operating characteristic (SROC) curves for the Berry86_2010, Sweeney3_2016 and Zak16_2016 signatures.



(a) Berry86_2010 (b) Sweeney3_2016 (c) Zak16_2016

Black dot is the summary estimate, white circles are individual study estimates, solid line passing through summary estimate is SROC curve. Solid line around estimates is 95% confidence region, dashed line around estimates is 95% prediction region

8.16 Appendix 3.12: Performance of Signatures for predicting progression to TB disease.

Study Name	Signature	Time window before TB (months)	Sensitivity	Specificity	PPV	NPV
Roe2019b	Roe3_2019	12	66.7%	98.8%	52.7%	99.3%
Suliman2018b	Suliman4_2018	24	75.6%	54.8%	3.3%	99.1%
Warsinske2018b	Sweeney3_2016	6	86.0%	84.2%	10.0%	99.7%
Zak2016p	Zak16_2016	12	53.4%	82.7%	5.9%	98.9%

PPV and NPV where calculated at 2% pre-test probability. PPV; Positive predictive value, NPV; Negative predictive value

8.17 Appendix 3.13: GRADE evidence profile: mRNA signatures for the diagnosis of TB (all studies).

Outcome	Study design	# Studies (Sample size)	Risk of bias (Study limitations)	Inconsistence	Indirectness	Imprecision	Publication bias	Final quality	Effect per 100,000	Importance[1]
True Positives	Cohort, Case-Control, Cross-Section	26 (4,182)	Very serious[2] (-2)	Very serious[3] (-2)	No serious indirectness[4]	Serious[5] (-1)	Likely[6]	Very low □OOO	Prevalence[7] 2% : 1.790	Critical
True Negative	Cohort, Case-Control, Cross-Section	26 (4,182)	Very serious[2] (-2)	Very serious[3] (-2)	No serious indirectness[4]	Serious[5] (-1)	Likely[6]	Very low □OOO	Prevalence[7] 2% : 76,930	Critical
False Positives	Cohort, Case-Control, Cross-Section	26 (4,182)	Very serious[2] (-2)	Very serious[3] (-2)	No serious indirectness[4]	Serious[5] (-1)	Likely[6]	Very low □OOO	Prevalence[7] 2% : 21,070	Critical
False Negatives	Cohort, Case-Control, Cross-Section	26 (4,182)	Very serious[2] (-2)	Very serious[3] (-2)	No serious indirectness[4]	Serious[5] (-1)	Likely[6]	Very low □OOO	Prevalence[7] 2% : 210	Critical

1. The Importance of outcomes was classified as either "Critial", "Important" or "Not important" according to their relative importance

2. Majority of studies were case-controls and lacked both a representative patient spectrum and blinding

3. Substantial heterogeneity was observed in the study results

4. No downgrade was applied for indirectness though diagnostic accuracy is considered a surrogate for paitent-important outcomes.

5. Diagnostic accuracy estimates were not pooled. There were a considerable number of studies with wide 95% CI. We down-graded by one point only, as there were a large number of studies and we had already down-graded for inconsistency

6. We did not down-grade for publication bias. The data in this systematic review did not allow for formal assessment of publication bias with methods such as funnel plots or regression analysis. Publication bias can not be ruled out as studies may not have been published in which mRNA signatures showed poor diagnostic accuracy for TB

7. What is the meaning of these results among people being screened for TB disease, given a 2% prevalance of disease

Explanation: Based on a sample size of 4,170, median sensitity=89.5% and median specificity=78.5%, we rated the quality of the evidence as high when no points were subtracted, moderate when only one point was subtracted, low when two points were subtracted and very low when more than two points were subtracted. Deduction of points was based on the five factors that decrease study quality; study limitations, incosistence in results accross studies, indirectness in evidence, imprecision in summary estimates, and possibility of reporting bias. For each outcome, evidence from cohort and cross-sectional studies started as high quality while that from case-control studies started as moderate quality. We deducted two, one and zero points for "very serious", "serious", and "no serious" issues identified respectively, and the deducted points are shown in brackets. Reporting bias was classfied as either "very likely", "likely" and "not likely".

8.18 Appendix 4.1: The 11 transcripts comprising the RISK11 signature of TB risk and their putative functions.

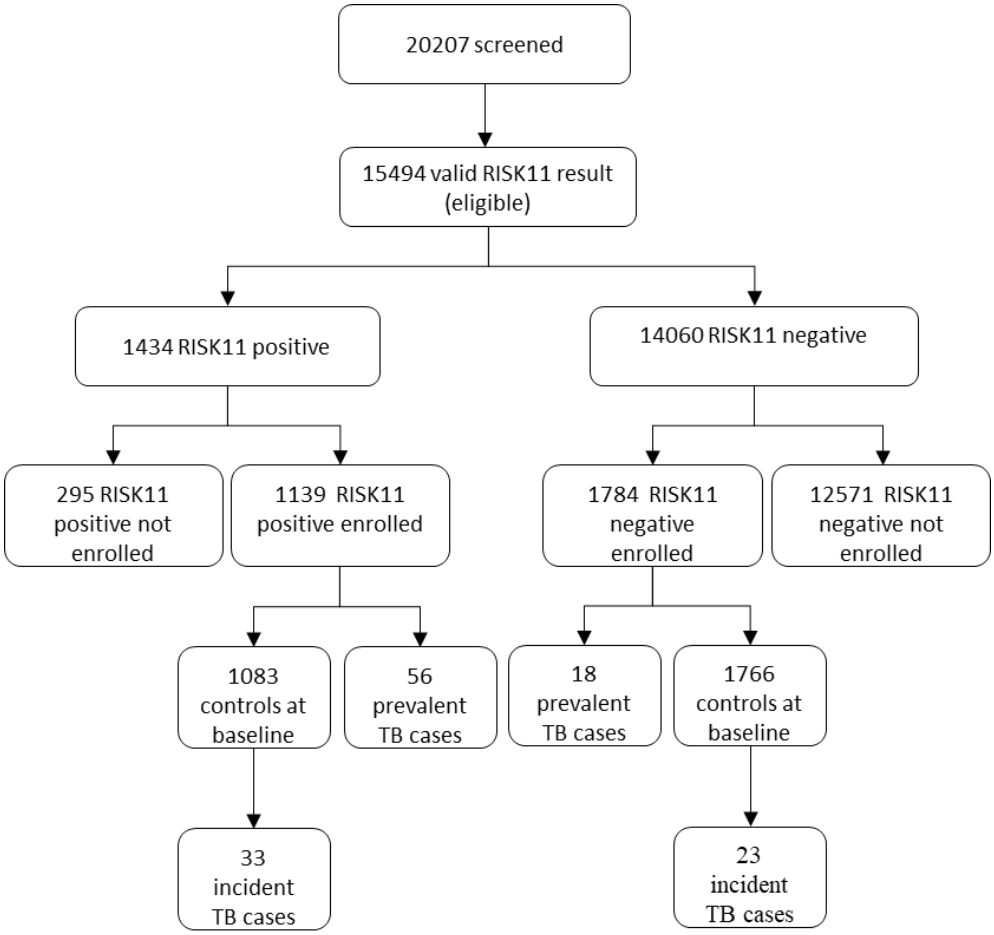
Symbol	Description/ putative function(s)
BATF2	Homo sapiens basic leucine zipper transcription factor, ATF-like 2 (BATF2), mRNA; protein dimerization activity [goid 46983]v [evidence IEA]; sequence-specific DNA binding [goid 43565] [evidence IEA]; transcription factor activity [goid 3700] [evidence IEA]; regulation of transcription, DNA-dependent [goid 6355] [evidence IEA]; nucleus [goid 5634] [evidence IEA]
ETV7	Homo sapiens ets variant gene 7 (TEL2 oncogene) (ETV7), mRNA; sequence-specific DNA binding [goid 43565] [evidence IEA]; transcription factor activity [goid 3700] [evidence IEA]; specific RNA polymerase II transcription factor activity [goid 3704] [pmid 10828014] [evidence TAS]; transcription [goid 6350] [evidence IEA]; transcription from RNA polymerase II promoter [goid 6366] [pmid 10828014] [evidence TAS]; regulation of transcription, DNA-dependent [goid 6355] [evidence IEA]; nucleus [goid 5634] [pmid 10828014] [evidence TAS]
FCGR1C‡	Homo sapiens Fc Fragment Of IgG Receptor Ic, Pseudogene (FCGR1CP), mRNA; The gene represents one of three related immunoglobulin gamma Fc receptor genes located on chromosome 1. This family member lacks the transmembrane and coiled-coiled domains found in other family members and is thought to be a pseudogene of Fc-gamma-receptor 1A.
GBP1	Homo sapiens guanylate binding protein 1, interferon-inducible, 67kDa (GBP1), mRNA; GTP binding [goid 5525] [pmid 1715024] [evidence TAS]; GTPase activity [goid 3924] [evidence IEA]; nucleotide binding [goid 166] [evidence IEA]; immune response [goid 6955] [evidence IEA]; membrane [goid 16020] [evidence IEA]
GBP2	Homo sapiens guanylate binding protein 2, interferon-inducible (GBP2), mRNA; GTP binding [goid 5525] [pmid 1715024] [evidence TAS]; GTPase activity [goid 3924] [evidence IEA]; nucleotide binding [goid 166] [evidence IEA];immune response [goid 6955] [pmid 1715024] [evidence TAS]; membrane [goid 16020] [evidence IEA]
GBP5	Homo sapiens guanylate binding protein 5 (GBP5), mRNA; GTP binding [goid 5525] [evidence IEA]; GTPase activity [goid 3924] [evidence IEA]; nucleotide binding [goid 166] [evidence IEA]; immune response [goid 6955] [evidence IEA]; membrane [goid 16020] [evidence IEA]
SCARF1	Homo sapiens scavenger receptor class F, member 1 (SCARF1), transcript variant 3, mRNA; low-density lipoprotein binding [goid 30169] [pmid 9395444] [evidence IDA]; transmembrane receptor activity [goid 4888] [pmid 9395444] [evidence TAS]; protein binding [goid 5515] [evidence IEA]; low-density lipoprotein catabolism [goid 45192] [pmid 9395444] [evidence TAS]; cell adhesion [goid 7155] [evidence IEA]; receptor mediated endocytosis [goid 6898] [pmid 9395444] [evidence TAS]; membrane [goid 16020] [evidence IEA]; integral to membrane [goid 16021] [pmid 9395444] [evidenc
SERPING1	Homo sapiens serpin peptidase inhibitor, clade G (C1 inhibitor), member 1, (angioedema, hereditary) (SERPING1), transcript variant 2, mRNA; serine-type endopeptidase inhibitor activity [goid 4867] [pmid 1363816] [evidence TAS]; innate immune response [goid 45087] [evidence IEA]; blood coagulation [goid 7596] [evidence IEA]; circulation [goid 8015] [pmid 2563376] [evidence TAS]; complement activation, classical pathway [goid 6958] [evidence IEA]; extracellular region [goid 5576] [pmid 14718574] [evidence NAS]
STAT1	Homo sapiens signal transducer and activator of transcription 1, 91kDa (STAT1), transcript variant alpha, mRNA; transcription factor activity [goid 3700] [pmid 10848577] [evidence TAS]; hematopoietin/interferon-class (D200-domain) cytokine receptor signal transducer activity [goid 5062] [pmid 8608597] [evidence TAS]; signal transducer activity [goid 4871] [evidence IEA]; protein binding [goid 5515] [pmid 12867595] [evidence IPI]; calcium ion binding [goid 5509] [evidence IEA]; transcription from RNA polymerase II promoter [goid 6366] [pmid 9630226] [evidence TA
TAP1	Homo sapiens transporter 1, ATP-binding cassette, sub-family B (MDR/TAP) (TAP1), mRNA; ATPase activity, coupled to transmembrane movement of substances [goid 42626] [evidence IEA]; ATPase activity [goid 16887] [evidence IEA]; transporter activity [goid 5215] [evidence IEA]; nucleotide binding [goid 166] [evidence IEA]; protein heterodimerization activity [goid 46982] [pmid 11133832] [evidence IPI]; ATP binding [goid 5524] [evidence IEA]; oligopeptide transporter activity [goid 15198] [evidence IEA]; protein bin
TRAFD1	Homo sapiens TRAF-type zinc finger domain containing 1 (TRAFD1), mRNA; zinc ion binding [goid 8270] [evidence IEA]; nucleic acid binding [goid 3676] [evidence IEA]; intracellular [goid 5622] [evidence IEA]

Sources:

(a) Gene symbols and descriptions/function(s) are based on the Chaussabel et al gene set modules except for FCGR1C.[1]

(b) FCGR1C‡ pseudogene: Gene symbol and description/function(s) are based on the GeneCards database (https://www.genecards.org/cgi-bin/carddisp.pl?gene=FCGR1CP).[2] The most recent (third) iteration for the Chaussabel gene set modules (CM; 'BloodGen3') does not include FCGR1C.[3, 4]

8.19 Appendix 4.2: Study design



8.20 Appendix 4.3: Baseline characteristics of enrolled participants by TB status adjusted to reflect screening population.

Variable	Total	a) Prevalent TB	b) Incipient TB	c) Control	a vs b vs c
	n=2923	n=74	n=56	n=2793	P-value
Age (median, IQR)	26 (22–33)	29 (24–36)	28 (22–37)	26 (22–33)	0.01
BMI (median, IQR)	23 (20–28)	21 (18–24)	20 (19–23)	23 (20–28)	0.01
RISK11 Score (median, IQR)	26 (8–77)	87 (61–96)	67 (15–81)	24 (8–75)	0.01
Sex (males) (n, %)	1338 (48.5)	47 (62.2)	33 (69.2)	1258 (48)	<0.001
Race (n, %)					
Caucasian	4 (0.2)	0 (0)	0 (0)	4 (0.2)	0.02
Mixed	968 (30.7)	34 (45.1)	26 (47.4)	908 (30.3)	
Black	1947 (69.0)	40 (54.9)	30 (52.6)	1877 (69.4)	
Asian	4 (0.2)	0 (0)	0 (0)	4 (0.2)	
Smoking history (n, %)	1478 (49.8)	45 (64.1)	41 (79.6)	1392 (49.2)	<0.001
Prior TB (n, %)	230 (7.0)	19 (20.6)	8 (19.4)	203 (6.6)	<0.001
TB contact history (n, %)	462 (16.0)	15 (21.4)	9 (11)	438 (16)	0.23
Flu-like symptoms (n, %)	134 (3.8)	4 (2.4)	1 (0.6)	129 (3.9)	0.87
TB Symptoms					
Chest pains (n, %)	30 (0.8)	4 (2.4)	0 (0)	26 (0.8)	0.20
Cough (n, %)	58 (1.6)	12 (7.1)	0 (0)	46 (1.5)	0.01
Fever (n, %)	3 (0.1)	1 (0.6)	0 (0)	2 (0.1)	1.00
Haemoptysis (n, %)	2 (0.1)	0 (0)	0 (0)	2 (0.1)	1.00
Loss of weight (n, %)	41 (1.3)	5 (3)	0 (0)	36 (1.3)	0.39
Night sweats (n, %)	32 (0.6)	7 (4.2)	1 (3.5)	24 (0.6)	0.01
Any symptom (n, %)	123 (3.4)	13 (7.7)	1 (3.5)	109 (3.4)	0.08

For continuous data, p values were computed using Wilcoxon Rank Sum test between two groups and Kruskal Wallis test for more than two groups. For categorical data, p values were computed using Fischer's exact test.

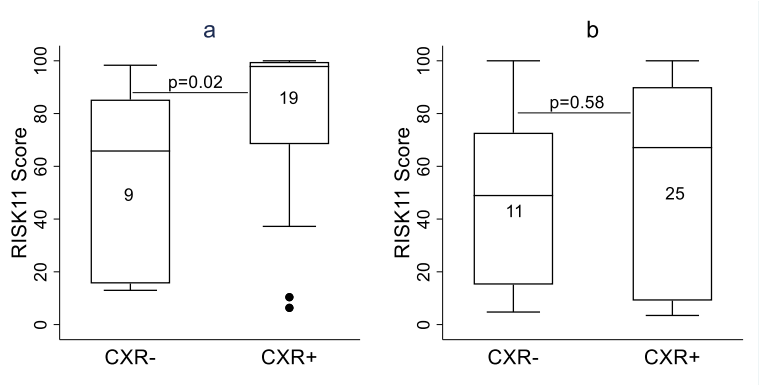
IQR, inter-quartile range. BMI, body mass index.

8.21 Appendix 4.4: Comparison of distribution of RISK11 scores by baseline characteristics in (a) Prevalent TB cases, (b) Incident TB cases, and (c) Controls.

Variable	Category	(a) Prevalent TB, N = 74			(b) Incident TB, N = 56			(c) Controls, N = 2,793		
		n (%)	RISK11 score median (IQR)	P-value	n (%)	RISK11 score median (IQR)	P-value	n (%)	RISK11 score median (IQR)	P-value
Race	Caucasian	0	-	-	0	-	-	4 (0.1)	12.5 (5.6; 47.0)	ref
	Mixed	34 (46.0)	84.2 (61.4–94.8)		26 (46.6)	67.3 (33.3–80.1)		908 (32.5)	39.7 (12.1–76.2)	0.18
	Black	40 (54.0)	91.1 (58.9–97.0)	0.37	30 (53.4)	67.4 (8.2–90.0)	0.28	1,877(67.2)	18.2 (7.4–74.5)	0.42
	Asian	0	-	-	0	-	-	4 (0.1)	15.8 (8.8–60.6)	0.39
Sex	Female	27 (36.5)	90.9 (40.3–95.2)		23 (41.1)	69.7 (55.8–82.3)		1,535 (55.0)	34.1 (10.0–77.5)	
	Male	47 (63.5)	86.6 (61.4–96.9)	0.78	33 (58.9)	48.9 (7.8–80.1)	0.85	1,258 (45.0)	16.9 (6.9–71.4)	<0.001
Chest pains	No	70 (94.6)	85.7 (54.5–95.2)		56 (100)	67.3 (15.4–81.2)		2,767 (99.1)	24.2 (8.2–74.9)	
	Yes	4 (5.4)	98.3 (95.5–99.1)	0.02	0	-	-	26 (0.9)	28.4 (9.5–93.5)	0.22
Cough	No	62 (83.8)	83.5 (41.6–94.8)		56 (100)	67.3 (15.4–81.2)		2,747 (98.3)	23.8 (8.2–74.9)	
	Yes	12 (16.2)	96.1 (92.9–98.3)	0.001	0	-	-	46 (1.7)	41.3 (12.3–89.6)	0.08
Fever	No	73 (98.6)	86.6 (61.4–96.1)		56 (100)	67.3 (15.4–81.2)		2,791 (99.9)	24.2 (8.2–75.3)	
	Yes	1 (1.4)	92.6 (NA)	0.64	0	-	-	2 (0.1)	55.2 (24.7–85.7)	0.42
Flu-like symptoms	No	70 (94.6)	85.7 (54.5–95.7)		55 (98.2)	67.1 (15.2–80.1)		2,664 (95.4)	22.9 (8.2–74.2)	
	Yes	4 (5.4)	96.3 (93.5–98.9)	0.05	1 (1.8)	99.6 (NA)	0.13	129 (4.6)	61.0 (12.1–88.3)	<0.001
Haemoptysis	No	74 (100)	87.0 (61.4–96.1)		56 (100)	67.3 (15.4–81.2)		2,791 (99.9)	24.2 (8.2–75.3)	
	Yes	0	-	-	0	-	-	2 (0.1)	7.4 (6.9–7.8)	0.15
Loss of weight	No	69 (93.2)	85.7 (54.5–95.2)		56 (100)	67.3 (15.4–81.2)		2,757 (98.7)	24.2 (8.2–74.9)	
	Yes	5 (6.8)	97.0 (94.4–98.3)	0.02	0	-	-	36 (1.3)	19.3 (10.0–82.0)	0.66
Night sweats	No	67 (90.5)	85.7 (51.1–94.8)		55 (98.2)	67.5 (15.2–82.3)		2,769 (99.1)	23.7 (8.2–74.9)	
	Yes	7 (9.5)	98.3 (92.6–100)	0.01	1 (1.8)	45 (NA)	0.64	24 (0.9)	71.4 (28.6–94.6)	0.003
Smoking history	No	29 (39.2)	92.6 (68.4–97.8)		15 (26.8)	68.8 (55.8–90.0)		1,401 (50.2)	22.0 (8.2–74.5)	
	Yes	45 (60.8)	82.7 (51.1–93.5)	0.13	41 (73.2)	66.1 (8.2–80.1)	0.37	1,392 (49.8)	26.4 (8.2–75.8)	0.31
Prior TB	No	55 (74.3)	89.5 (51.1–97.4)		48 (85.7)	70.1 (14.9–86.1)		2,590 (92.7)	22.7 (8.2–74.9)	
	Yes	19 (25.7)	85.7 (65.8–90.9)	0.29	8 (14.3)	42.2 (15.4–65.2)	0.27	203 (7.3)	42.9 (11.7–77.7)	0.02
TB contact history	No	59 (79.7)	85.7 (61.4–95.7)		47 (83.9)	67.5 (15.2–90.0)		2,355 (84.3)	25.1 (8.2–75.3)	
	Yes	15 (20.3)	91.3 (40.3–100.0)	0.28	9 (16.1)	66.1 (61.9–79.6)	0.92	438 (15.7)	21.2 (8.2–73.6)	0.72
Any symptom	No	61 (82.4)	81.8 (41.6–93.5)		55 (98.2)	67.5 (15.2–82.3)		2,684 (96.1)	23.4 (8.2–74.5)	
	Yes	13 (17.6)	97.0 (93.1–98.3)	<0.001	1 (1.8)	45 (NA)	0.64	109 (3.9)	48.5 (10.8–89.2)	0.002

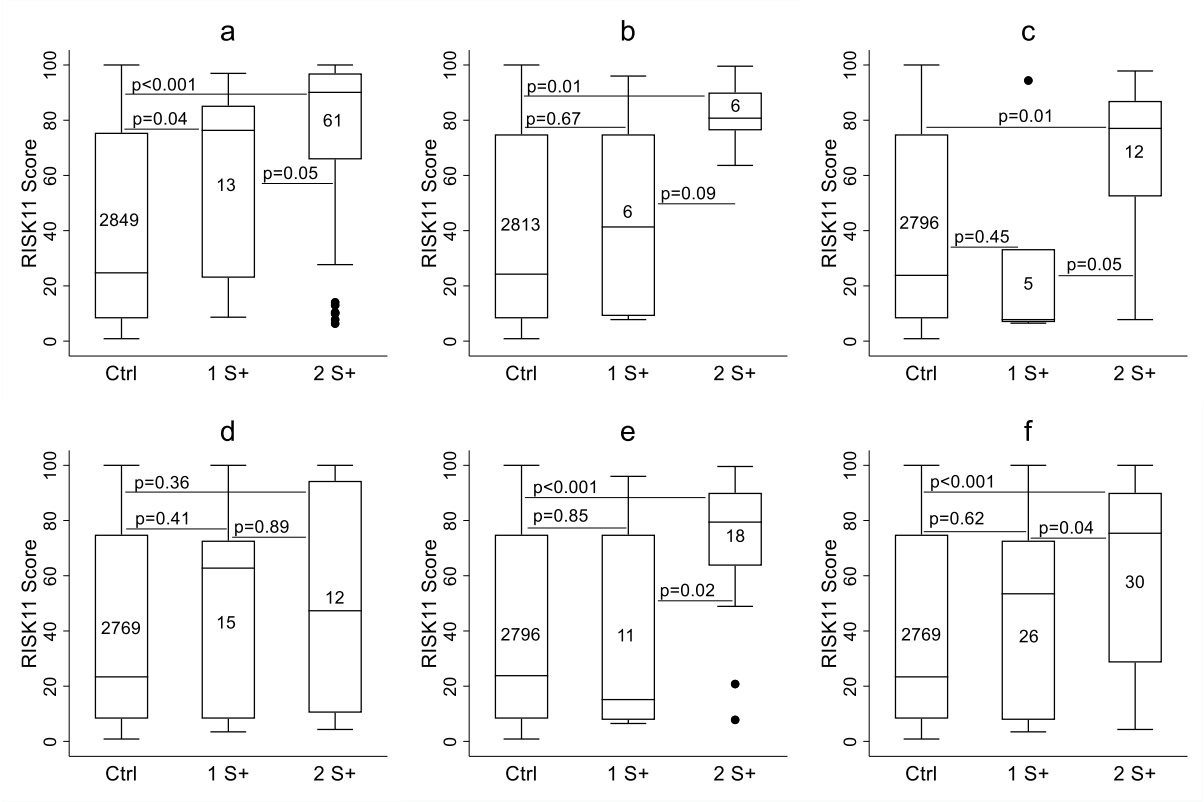
P values were computed using Wilcoxon Rank Sum test for continuous data and Fischer's exact for categorical data. Data are not adjusted to reflect screening population.

8.22 Appendix 4.5: Distribution of baseline RISK11 scores in (a) prevalent and (b) incident TB cases stratified by chest radiograph positivity at diagnosis.



CXR+, chest radiograph suggestive of TB disease. CXR-, chest radiograph not suggestive of TB disease. P values were computed using Wilcoxon Rank Sum.

8.23 Appendix 4.6: Distribution of baseline RISK11 scores among controls, one-sample sputum positive cases, and two-sample sputum positive cases unadjusted for sampling weights.



a) RISK11 score distribution in controls and prevalent TB cases (TB cases at baseline). b) RISK11 score distribution in controls and incident TB cases diagnosed between months two and six. c) RISK11 score distribution in controls and incident TB cases diagnosed between months seven and twelve. d) RISK11 score distribution in controls and incident TB cases diagnosed between months thirteen and fifteen. e) RISK11 score distribution in controls and cumulative incident TB cases diagnosed between months one and twelve. f) RISK11 score distribution in controls and cumulative incident TB cases diagnosed between months one and fifteen. Numbers in b, c, d, e and f excludes 24 participants who did not attend a subsequent visit and hence have unknown TB outcome. Numbers in the boxplots indicate the number of participants in that stratum. P values were computed using Wilcoxon Rank Sum.

Ctrl, controls without TB. 1 S+, one-sample sputum positive TB cases. 2 S+, two-sample sputum positive TB cases.

8.24 Appendix 4.7: Univariable and multivariable regression analyses of baseline predictors of RISK11 score in participants with prevalent TB at enrolment.

Variable	n=74	Univariable Analysis		Multivariable Analysis		
		β. Coef. (95% CI)	P-value	β. Coef. (95% CI)	% Marginal effect (95% CI)	P-value
Age (median, IQR)	29 (24–36)	-0.03 (-0.07–0.01)	0.04	-0.03 (-0.06–0.01)	-0.65 (-1.34–0.04)	0.07
BMI (median, IQR)	20.7 (18.5–23.8)	-0.07 (-0.16–0.01)	0.08	-	-	-
Male sex (n, %)	47 (63.5)	0.24 (-0.52–1)	0.53	-	-	-
Race:						
Black (n, %)	40 (54)	Ref	-	-	-	-
Mixed (n, %)	34 (46)	0.1 (-0.62–0.83)	0.78	-	-	-
Smoking history (n, %)	45 (60.8)	-0.24 (-1.05–0.56)	0.56	-	-	-
Prior TB (n, %)	19 (25.7)	-0.12 (-1.15–0.91)	0.82	-	-	-
TB contact history (n, %)	15 (20.7)	0.08 (-0.77–0.93)	0.85	-	-	-
Flu-like symptoms (n, %)	4 (5.4)	3.46 (2.62–4.31)	<0.001	-	-	-
Chest pains (n, %)	4 (5.4)	3.81 (2.71–4.91)	<0.001	-	-	-
Cough (n, %)	12 (16.2)	3.31 (2.64–3.98)	<0.001	3.23 (2.51–3.94)	72.55 (58.06–87.03)	<0.001
Fever (n, %)	1 (1.4)	2.72 (2.36–3.09)	<0.001	-	-	-
Loss of weight (n, %)	5 (6.8)	3.54 (2.77–4.32)	<0.001	-	-	-
Night sweats (n, %)	7 (9.5)	3.33 (2.26–4.4)	<0.001	-	-	-

IQR, inter-quartile range. BMI, body-mass index. β. Coef., Beta coefficient. % Marginal effect, predicted marginal change in RISK11 score associated with each predictor variable in the model. P values are reported from the model output.

8.25 Appendix 4.8: Univariable and multivariable regression analyses of baseline predictors of RISK11 score in participants who progressed to incident TB.

Variable	n=56	Univariable Analysis		Multivariable Analysis		
		β. Coef. (95% CI)	P-value	β. Coef. (95% CI)	% Marginal effect (95% CI)	P-value
Age (median, IQR)	28 (22-37)	0.02 (0–0.05)	0.10	-	-	-
BMI (median, IQR)	19.8 (18.7-23.4)	0.02 (-0.07–0.12)	0.61	-	-	-
Male sex (n, %)	33 (58.9)	-0.75 (-1.51–0.01)	0.05	-	-	-
Race:						
Black (n, %)	30 (53.4)	Ref	-	-	-	-
Mixed (n, %)	26 (46.6)	0.5 (-0.2–1.19)	0.16	-	-	-
Smoking history (n, %)	41 (73.2)	-0.67 (-1.56–0.23)	0.14	-	-	-
Prior TB (n, %)	8 (14.3)	-0.03 (-0.87–0.8)	0.94	-	-	-
TB contact history (n, %)	9 (16.1)	0.03 (-1.25–1.31)	0.97	-	-	-
Flu-like symptoms (n, %)	1 (1.8)	6.33 (5.97–6.69)	<0.001	-	-	-
Night sweats (n, %)	1 (1.8)	0.7 (0.33–1.06)	<0.001	-	-	-

IQR, inter-quartile range. BMI, body-mass index. β. Coef., Beta coefficient. % Marginal effect, predicted marginal change in RISK11 score associated with each predictor variable in the model. P values are reported from the model output.

8.26 Appendix 4.9: Univariable and multivariable regression analyses of baseline predictors of RISK11 score in participants without prevalent TB and did not progress to incident TB (controls).

Variable	n=2,793	Univariable Analysis		Multivariable Analysis		
		β. Coef. (95% CI)	P-value	β. Coef. (95% CI)	% Marginal effect (95% CI)	P-value
Age (median, IQR)	26 (22-33)	0.01 (-0.01–0.01)	0.07	-		-
BMI (median, IQR)	22.7 (20-27.9)	0.01 (-0.01–0.01)	0.88	-		-
Male sex (n, %)	1258 (45)	-0.3 (-0.38–-0.21)	<0.001	-0.36 (-0.45—0.27)	-5.99 (-7.49—4.5)	<0.001
Race: Black (n, %;)	1877 (67.2)	Ref	-	-		-
Asian (n, %)	4 (0.1)	-0.14 (-0.94–0.66)	0.73	-		-
Caucasian (n, %)	4 (0.1)	-0.47 (-1.32–0.37)	0.27	-		-
Mixed (n, %)	908 (32.5)	0.36 (0.27–0.45)	<0.001	-		-
Smoking history (n, %)	1392 (49.8)	0.05 (-0.03–0.14)	0.20	0.16 (0.07–0.25)	2.74 (1.24–4.24)	<0.001
Prior TB (n, %)	20 (7.3)	0.22 (0.05–0.38)	0.01	0.18 (0.01–0.35)	3.03 (0.25–5.82)	0.03
TB contact history (n, %)	438 (15.7)	-0.07 (-0.18–0.04)	0.23	-		-
Flu-like symptoms (n, %)	129 (4.2)	0.31 (0.1–0.52)	<0.001	0.26 (0.05–0.48)	4.39 (0.08–7.98)	0.02
Chest pains (n, %)	26 (0.9)	0.03 (-0.4–0.46)	0.89	-		-
Cough (n, %)	46 (1.7)	0.26 (-0.06–0.59)	0.11	-		-
Fever (n, %)	2 (0.1)	0.60 (-0.31–1.51)	0.19	-		-
Haemoptysis (n, %)	2 (0.1)	-1.23 (-1.33—1.13)	<0.001	-		-
Loss of weight (n, %)	36 (1.3)	-0.01 (-0.36–0.34)	0.95	-		-
Night sweats (n, %)	24 (0.9)	0.68 (0.08–1.29)	0.03	0.76 (0.14–1.38)	12.69 (2.34–23.04)	0.02

IQR, inter-quartile range. BMI, body-mass index. β. Coef., Beta coefficient. % Marginal effect predicted marginal change in RISK11 score associated with each predictor variable in the model. P values are reported from the model output.

8.27 Appendix 4.10a: Initial exploratory ROC regression analysis for the effect of covariates on discriminatory accuracy of RISK11.

Group/Variable	Prevalent TB		Incident TB	
	β. Coef. (95% CI)	P-value	β. Coef. (95% CI)	P-value
Case population				
BMI	-1.69 (-3.60–0.22)	0.08	-0.65 (-2.78–1.48)	0.55
Male sex	9.32 (-7.31–25.95)	0.27	-8.57 (-29.85–12.7)	0.43
Night sweats	-6.00 (-27.70–15.71)	0.59	10.6 (-6.61–27.81)	0.23
Flu-like symptoms	19.19 (-16.9–55.28)	0.30	60.14 (41.22–79.06)	<0.001
Haemoptysis	Omitted	-	Omitted	-
Smoking history	-9.39 (-24.01–5.23)	0.21	-13.24 (-37.84–11.36)	0.29
Cough	35.31 (14.78–55.83)	0.001	Omitted	-
Control population				
BMI	-0.11 (-0.21—0.01)	0.04	-0.09 (-0.19–0.01)	0.06
Male sex	-6.70 (-8.30—5.09)	<0.001	-6.54 (-8.13—-4.94)	<0.001
Night sweats	12.92 (-1.38–27.22)	0.08	11.51 (-3.79–26.8)	0.14
Flu-like symptoms	5.11 (0.98–9.23)	0.02	5.26 (1.09–9.43)	0.01
Haemoptysis	-12.82 (-20.07—-5.57)	0.001	-13.24 (-20.62—-5.86)	<0.001
Smoking history	2.67 (1.16–4.18)	0.001	2.71 (1.2–4.22)	<0.001
Cough	3.63 (-3.6–10.86)	0.33	4.23 (-3.14–11.6)	0.26
ROC model				
BMI	-0.06 (-0.12–0.01)	0.09	-0.02 (-0.1–0.05)	0.55
Male sex	0.32 (-0.26–0.89)	0.28	-0.31 (-1.11–0.48)	0.44
Night sweats	-0.2 (-0.94–0.53)	0.59	0.39 (-0.26–1.04)	0.24
Flu-like symptoms	0.65 (-0.57–1.87)	0.28	2.21 (1.4–3.01)	<0.001
Haemoptysis	Omitted	-	Omitted	-
Smoking history	-0.32 (-0.82–0.18)	0.21	-0.49 (-1.4–0.43)	0.30
Cough	1.19 (0.76–1.93)	0.001	Omitted	-

Covariates under 'case population' are those that affect TB cases and variables under 'control population' are those that affect controls without TB. The covariates under the 'ROC model' are those that affect discrimination between cases and controls. See appendix 4.10b for the final ROC regression model derived from this analysis by removing the non-significant variables in each category. P values are reported from the model output.

BMI, body mass index.

8.28 Appendix 4.10b: Final ROC regression analysis for the effect of covariates on discriminatory accuracy of RISK11 derived from appendix table 10a

Group/Variable	Prevalent TB		Incident TB	
	β. Coef. (95% CI)	P-value	β. Coef. (95% CI)	P-value
Case population				
Cough	43.5 (30.64–56.37)	<0.001	-	-
Flu-like symptoms	-	-	50.38 (40.26–60.49)	<0.001
Control population				
BMI	-0.11 (-0.21–0)	0.04	-0.15 (-0.27–-0.03)	0.02
Male sex	-6.57 (-8.17–-4.96)	<0.001	-7.72 (-9.76–-5.68)	<0.001
Night Sweats	15.02 (2.12–27.91)	0.02	17.71 (2.71–32.7)	0.02
Flu-like symptoms	5.17 (1.05–9.29)	0.014	8.4 (2.89–13.91)	0.003
Haemoptysis	-9.31 (-10.67–-7.96)	<0.001	-12.88 (-14.67–-11.1)	<0.001
Smoking history	2.64 (1.14–4.14)	0.001	3.15 (1.21–5.08)	0.001
ROC model				
Cough	1.40 (1.00–1.87)	<0.001	-	-
Flu-like symptoms	-	-	1.58 (1.14–2.03)	<0.001

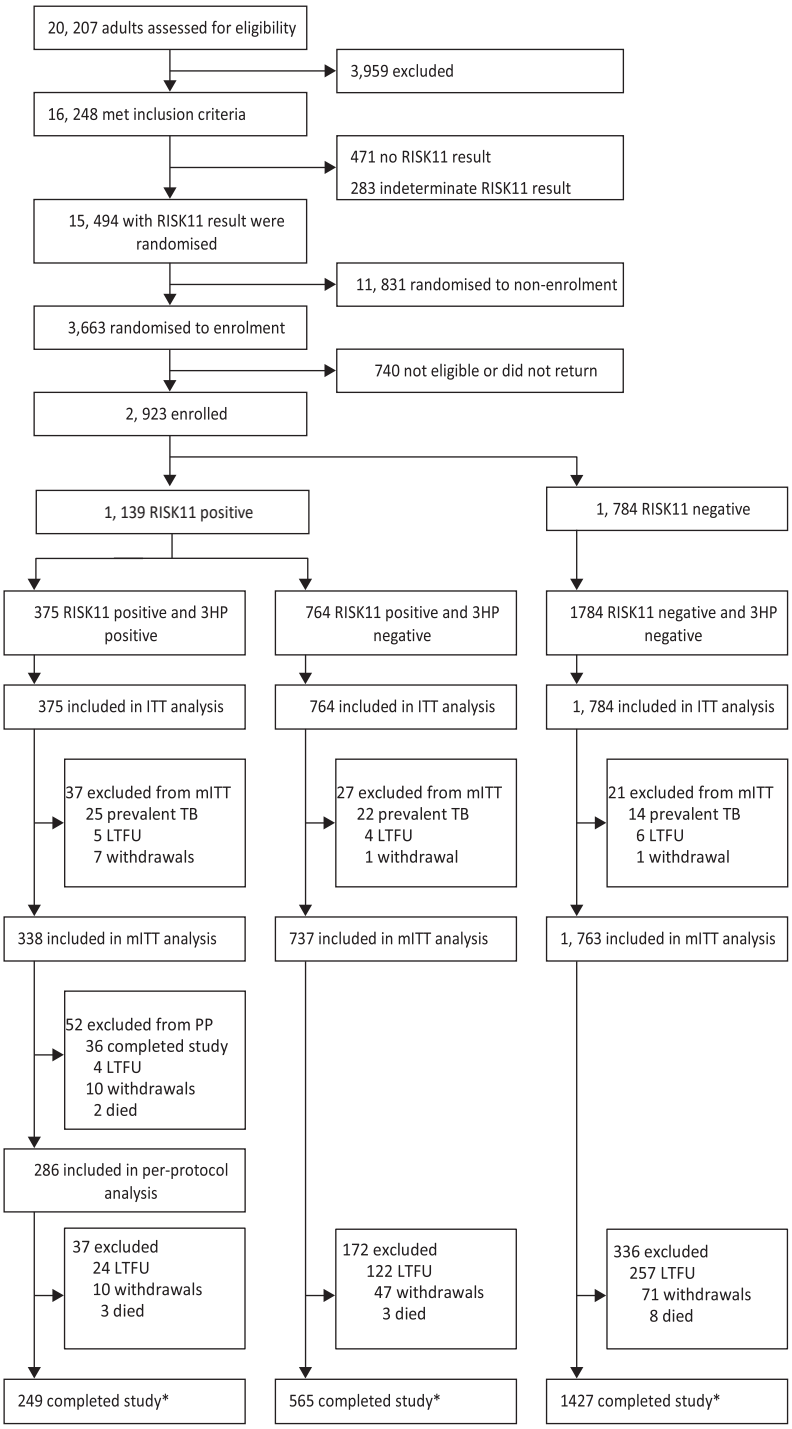
Covariates under 'case population' are those that affect TB cases and variables under 'control population' affect controls without TB. The covariates under the 'ROC model' are those that affect discrimination between cases and controls. See online supplementary table 6a for detailed model on which this model is based. P values are reported from the model output.
BMI, body-mass Index.

8.29 Appendix 5.1: Exploratory univariable and multivariable generalised linear models for upper respiratory organisms and socio-demographic predictors of RISK11 score in participants who were not investigated for TB.

Variable	N=713	Univariable Analysis			Multivariable Analysis		
		β. Coefficient (95% CI)	% Marginal Effect (95% CI)	P	β. Coefficient (95% CI)	%Marginal Effect (95% CI)	P
Age (median, IQR)	29 (22–37)	0.01 (0–0.02)	0.1 (0–0.2)	0.01	0.01 (0–0.02)	0.1 (0–0.2)	0.02
BMI (median, IQR)	23.9 (20.2–30)	0 (0–0.01)	0.1 (-0.1–0.2)	0.29	-		-
Sex (Male) (n, %)	321 (45)	-0.24 (-0.38–-0.1)	-3.4 (-5.3–-1.4)	0.01	0.01 (-0.01–0)	-3.1 (-5.1–-1.1)	0.01
Coronavirus (NL63, 229E, OC43 & HKU1) (n, %)	8 (1.1)	0.27 (-0.61–1.15)	3.8 (-8.7–16.3)	0.55	-		-
Influenza (A, B C & H1N1) (n, %)	1 (0.1)	-0.96 (-1.03–-0.89)	-13.6 (-15.2–-12)	<0.001	-		-
Rhinoviruses (n, %)	23 (3.3)	0.48 (0.13–0.82)	6.7 (1.9–11.6)	0.01	0.02 (0.12–0)	6.8 (1.9–11.7)	0.01
Mycoplasma pneumonia (n, %)	1 (0.1)	5.61 (5.55–5.68)	79.7 (77.1–82.3)	<0.001	-		-
Cytomegalovirus (n, %)	1 (0.1)	-1.41 (-1.48–-1.34)	-20 (-21.9–-18.1)	<0.001	-		-
Haemophilus Influenzae (n, %)	148 (20.8)	-0.01 (-0.17–0.16)	-0.1 (-2.5–2.3)	0.95	-		-
Klebsiella Pneumoniae (n, %)	10 (1.4)	-0.24 (-0.57–0.09)	-3.4 (-8.1–1.3)	0.15	-		-
Moraxella Catarrhalis (n, %)	40 (5.6)	0.21 (-0.07–0.5)	3 (-1.0–7.0)	0.14	-		-
Parainfluenza (types 1,2,3 and 4) (n, %)	2 (0.3)	0.71 (-0.74–2.17)	10.1 (-10.5–30.8)	0.34	-		-
Chlamydia Pneumoniae (n, %)	1 (0.1)	-1.63 (-1.7–-1.56)	-23.2 (-25.2–-21.1)	<0.001	-		-
Staphylococcus Aureus (n, %)	74 (10.4)	-0.04 (-0.28–0.19)	-0.6 (-3.9–2.8)	0.73	-		-
Streptococcus Pneumoniae (n, %)	60 (8.4)	0.16 (-0.09–0.41)	2.3 (-1.3–5.8)	0.21	-		-
Legionella (n, %)	2 (0.3)	-0.18 (-0.25–-0.11)	-2.6 (-3.7–-1.5)	<0.001	-		-

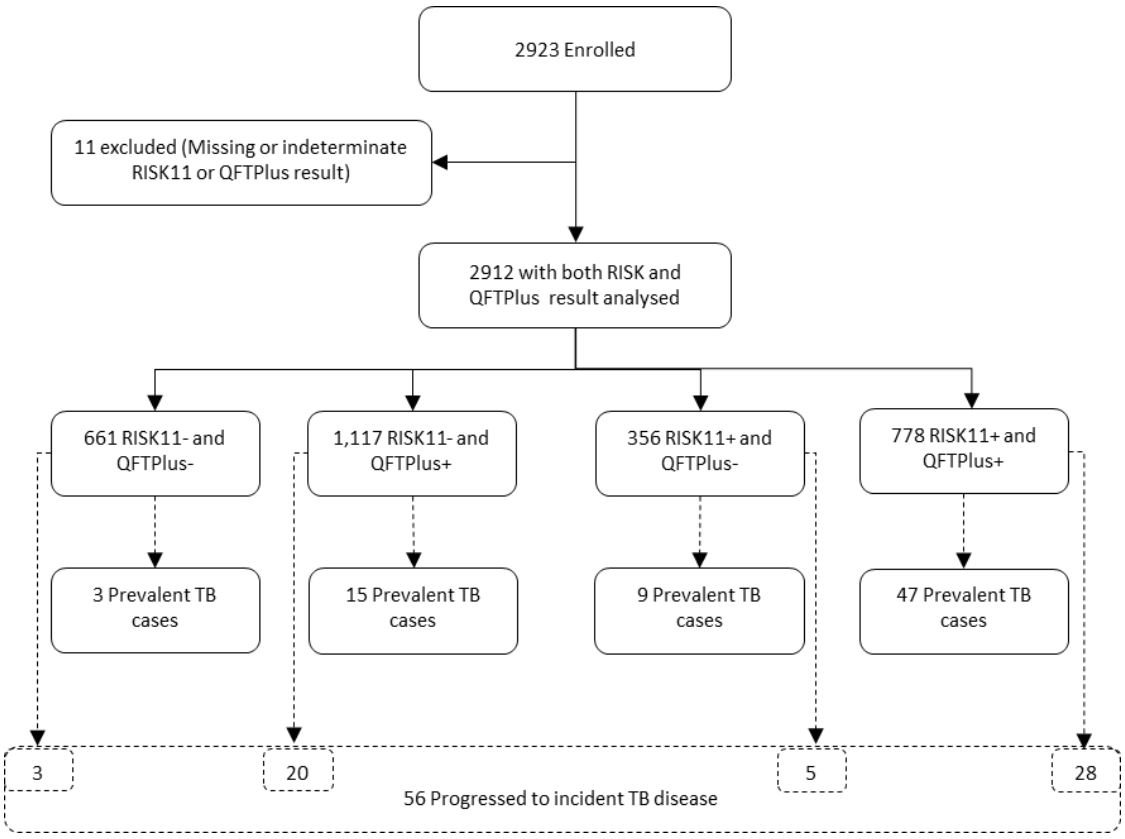
IQR, inter-quartile range. BMI, body-mass index. P, P-value. Percent (%) marginal effect is the predicted increase/decrease in RISK11 score associated with the presence of each socio-demographic factor or upper respiratory organism. The β coefficients can be exponentiated to obtain Odds Ratios of the association between each predictor variable and RISK11 score.

8.30 Appendix 6.1: Parent study (CORTIS) trial profile



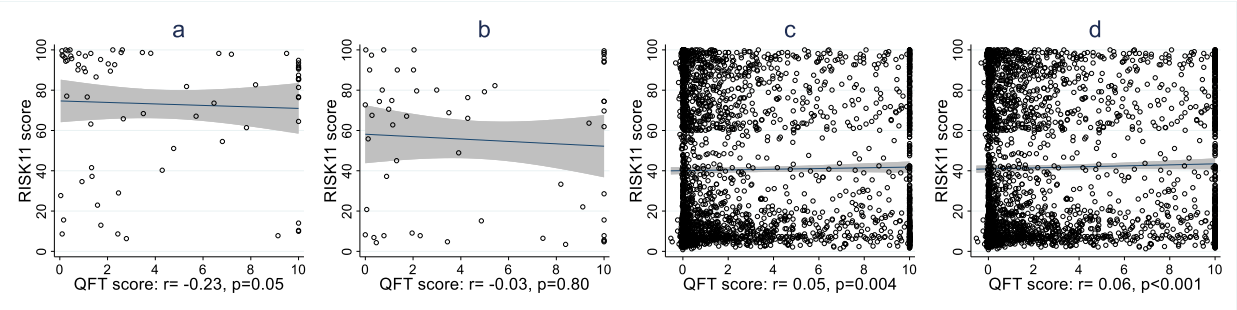
The figure was adapted from the parent study.[5] ITT, Intention to treat. mITT, Modified intention to treat. LTFU, Lost to follow-up. PP, Per protocol analysis. *585 participants did not complete the trial for reasons including: 53 (9%) pregnancies, 22 (4%) investigator withdrawals, 46 (8%) consent withdrawals, 26 (4%) HIV infections, 422 (72%) LTFU, and 16 (3%) deaths.

8.31 Appendix 6.2: Current study flow of participants included in the analysis.



QFTPlus, QuantiFERON-TB Gold-Plus. TB, Tuberculosis.

8.32 Appendix 6.3: Scatter plots for the relationship between RISK11 and QFTPlus scores.



Figures a, b, c and d depicts relationships in participants with prevalent TB, incident TB, controls without TB and all participants, respectively. r, spearman's rho coefficient. p, p-value.

QFTPlus, QuantiFERON-TB Gold-Plus. TB, Tuberculosis.

8.33 Appendix 6.4: Test agreement between QFTPlus and RISK11

Test	Group	Category	N	+/+ (n, %)	-/- (n, %)	+/- (n, %)	-/+ (n, %)	% Agreement	% Agreement#	κ
RISK11 (60)†	QFTPlus cut-off (0.35)	All	2912	778 (26.7)	661 (22.7)	356 (12.2)	1117 (38.4)	49.2	40.1	0.05
		Controls	2782	703 (25.3)	655 (23.5)	342 (12.3)	1082 (38.9)	48.9	40.3	0.04
		Prevalent TB	74	47 (63.5)	3 (4.1)	9 (12.2)	15 (20.3)	67.6	38.9	0.01
		Incident TB	56	28 (50)	3 (5.4)	5 (8.9)	20 (35.7)	55.4	26.4	-0.02
	QFTPlus cut-off (4)	All	2912	343 (11.8)	1263 (43.4)	791 (27.2)	515 (17.7)	55.2	67.3	0.02
		Controls	2782	305 (11)	1240 (44.6)	740 (26.6)	497 (17.9)	55.5	67.7	0.01
		Prevalent TB	74	23 (31.1)	11 (14.9)	33 (44.6)	7 (9.5)	46	54.5	0.01
		Incident TB	56	15 (26.8)	12 (21.4)	18 (32.1)	11 (19.6)	48.2	50.9	-0.02
RISK11 (26)†	QFTPlus cut-off (0.35)	All	2912	994 (34.1)	553 (19)	464 (15.9)	901 (30.9)	53.3	45.6	0.06
		Controls	2782	907 (32.6)	549 (19.7)	448 (16.1)	878 (31.6)	52.3	45.4	0.05
		Prevalent TB	74	54 (73)	2 (2.7)	10 (13.5)	8 (10.8)	75.7	61.2	0.04
		Incident TB	56	33 (58.9)	2 (3.6)	6 (10.7)	15 (26.8)	62.5	40.5	-0.04
	QFTPlus cut-off (4)	All	2912	445 (15.3)	1041 (35.7)	1013 (34.8)	413 (14.2)	51	61.2	0.02
		Controls	2782	402 (14.5)	1027 (36.9)	953 (34.3)	400 (14.4)	51.4	61.6	0.02
		Prevalent TB	74	26 (35.1)	6 (8.1)	38 (51.4)	4 (5.4)	43.2	47.1	0.003
		Incident TB	56	17 (30.4)	8 (14.3)	22 (39.3)	9 (16.1)	44.6	43.8	-0.08

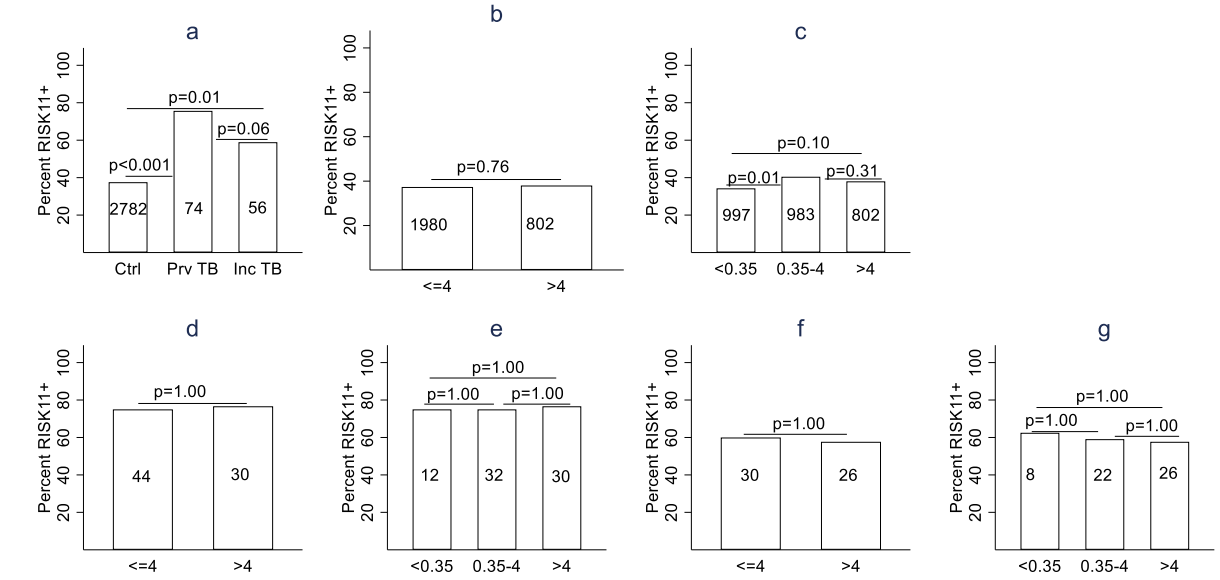
#Adjusted to reflect screening population.

†Numbers in brackets denote the RISK11 positivity threshold.

The numbers for controls include participants with unknown TB outcome.

QFTPlus, QuantiFERON-TB Gold-Plus. TB, Tuberculosis.

8.34 Appendix 6.5: RISK11-positivity rates stratified by either TB status or QFTPlus results.



Figures 'a' are RISK11 positivity rates in controls, prevalent and incident TB cases, 'b' and 'c' are positivity rates in controls, 'd' and 'e' are positivity rates in prevalent TB cases and 'f' and 'g' are positivity rates in incident TB cases. Controls in figures 'c' and 'd' include participants that who were controls at baseline but with undetermined outcome status at end of study because they never attended any subsequent visit or were lost to follow-up. Ctrl, Control. Prv TB, Prevalent tuberculosis. Inc TB, Incident tuberculosis. Numbers in the bars represent the number of participants in that QFTPlus group. P-values were computed using Fisher's exact.

QFTPlus, QuantiFERON-TB Gold-Plus. TB, Tuberculosis.

8.35 Appendix 6.6: Prevalent and incident TB disease by risk category using alternative RISK11 and QFTPlus cut-offs.

Category	Risk Group	a) Prevalent T			b) Incident TB		
		Prevalence (n, %)	Prevalence Ratio (95% CI)	P	Incident Rate (n , 95% CI)	Incident Rate Ratio (95% CI)	P
RISK11 (60)†	QFTPlus<0.35/RISK11-	3 (0.45)	Reference	-	0.46 (0.14–2.23)	Reference	
	QFTPlus<0.35/RISK11+	9 (2.53)	5.57 (1.52–20.45)	0.01	1.3 (0.51–4.33)	3.24 (0.78–13.48)	0.11
	QFTPlus0.35-4/RISK11-	8 (1.33)	2.93 (0.78–10.99)	0.11	1.49 (0.8–3.15)	3.27 (0.89–12.05)	0.08
	QFTPlus0.35-4/RISK11+	24 (5.52)	12.16 (3.68–40.13)	<0.001	2.94 (1.68–5.60)	6.79 (1.94–23.74)	0.01
	QFTPlus>4/RISK11-	7 (1.36)	2.99 (0.78–11.53)	0.11	2.11 (1.2–4.08)	4.63 (1.3–16.53)	0.02
	QFTPlus>4/RISK11+	23 (6.71)	14.77 (4.47–48.87)	<0.001	4.17 (2.48–7.55)	10.09 (2.93–34.76)	<0.001
RISK11 (26)†	QFTPlus<0.35/RISK11-	2 (0.36)	Reference	-	0.36 (0.08–3.62)	Reference	
	QFTPlus<0.35/RISK11+	10 (1.48)	4.09 (0.79–21.16)	0.09	1.12 (0.40–4.31)	3.08 (0.51–18.65)	0.22
	QFTPlus0.35-4/RISK11-	4 (0.82)	2.27 (0.42–12.32)	0.34	1.23 (0.56–3.22)	3.38 (0.68–16.71)	0.14
	QFTPlus0.35-4/RISK11+	28 (4.27)	11.8 (2.69–51.9)	0.01	2.79 (1.57–5.46)	7.71 (1.64–36.28)	0.01
	QFTPlus>4/RISK11-	4 (0.97)	2.68 (0.49–14.55)	0.25	2.14 (1.15–4.50)	5.9 (1.28–27.19)	0.02
	QFTPlus>4/RISK11+	26 (4.25)	11.8 (2.66–52.13)	0.01	3.09 (1.77–5.90)	7.92 (1.69–37.05)	0.01

†Numbers in brackets denote the RISK11 positivity threshold. QFTPlus, QuantiFERON-TB Gold-Plus. TB, Tuberculosis.

8.36 Appendix 6.7: Prevalent and incident TB disease by risk category at optimal test thresholds

Risk Group	a) Prevalent TB			b) Incident TB		
	Prevalence, % (95% CI)	Prevalence Ratio (95% CI)	P	Incidence Rate per 100 person-years (95% CI)	Incidence Rate Ratio (95% CI)	P
RISK11-/QFTPlus-	0.28 (0.03–1.02)	Reference	-	0.71 (0.30–2.10)	Reference	-
RISK11+/QFTPlus-	1.52 (0.76–2.88)	5.41 (1.16–25.27)	0.03	1.50 (0.60–4.81)	2.13 (0.61–7.36)	0.23
RISK11-/QFTPlus+	1.07 (0.46–2.10)	3.81 (0.81–17.87)	0.09	1.59 (0.93–2.99)	2.26 (0.8–6.39)	0.13
RISK11+/ QFTPlus+	4.79 (3.39–6.42)	16.99 (4.02–71.81)	<0.001	2.88 (1.72–5.20)	4.08 (1.47–11.30)	0.01

RISK11 and QFTPlus positivity thresholds used in this table were 26% for RISK11 and 0.92 UI/mL for QFTPlus, respectively. QFTPlus, QuantiFERON-TB Gold-Plus. TB, Tuberculosis.

8.37 Appendix 6.8: Performance of RISK11 and QFTPlus alone and in combination for diagnosis of prevalent and prognosis of incident TB in the enrolled population.

Statistic	(a) Prevalent TB				(b) Incident TB			
	RISK11	QFTPlus	Both-Positive	Either-Positive	RISK11	QFTPlus	Both-Positive	Either-Positive
True Positives	56	62	47	71	33	48	28	53
False Positives	1078	1833	731	2180	1029	1773	2068	2108
True Negatives	1760	1005	2107	658	1733	989	694	654
False Negatives	18	12	27	3	23	8	28	3
Total	**2912**	**2912**	**2912**	**2912**	**2818**	**2818**	**2818**	**2818**

RISK11 and QFTPlus positivity thresholds used in this table were 60% for RISK11 and 0.35 UI/mL for QFTPlus, respectively. The performance metrics in this table are not adjusted to the screening population. See table 3 in the manuscript for adjusted performance metrics.

QFTPlus, QuantiFERON-TB Gold-Plus. TB, Tuberculosis.

8.38 Appendix 6.9: Simulated performance of RISK11 and QFTPlus alone and in combination for diagnosis of prevalent and prognosis of incident TB for 10000 tests conducted.

Statistic	(a) Prevalent TB				(b) Incident TB			
	RISK11	QFTPlus	Both-Positive	Either-Positive	RISK11	QFTPlus	Both-Positive	Either-Positive
True Positives	45	114	38	122	27	123	23	131
False Positives	874	6221	593	6503	848	6214	572	6455
True Negatives	8989	3642	9270	3360	9005	3639	9281	3398
False Negatives	92	23	99	15	120	24	124	16
Total	**10000**	**10000**	**10000**	**10000**	**10000**	**10000**	**10000**	**10000**

RISK11 and QFTPlus positivity thresholds used in this table were 60% for RISK11 and 0.35 UI/mL for QFTPlus, respectively. QFTPlus, QuantiFERON-TB Gold-Plus. TB, Tuberculosis.

8.39 Appendix 6.10: Performance of RISK11 and QFTPlus alone and in combination for diagnosis of prevalent and prognosis of incident TB at optimal thresholds.

	(a) Prevalent TB			
Statistic	**RISK11 (26)**	**QFTPlus (0.92)**	**RISK11/QFTPlus (Both-Positive)**	**RISK11/QFTPlus (Either-Positive)**
PR (95% CI)	4.88 (2.40–9.92)	3.62 (1.69–7.73)	6.00 (3.18–11.32)	7.07 (1.71–29.32)
Sensitivity (95% CI)	62.81 (50.13–73.19)	79.99 (68.78–88.19)	50.24 (38.14–61.86)	92.56 (83.18–96.97)
Specificity (95% CI)	74.88 (73.24–76.46)	47.93 (46.07–49.78)	86.09 (84.75–87.34)	36.61 (34.83–38.41)
PPV (95% CI)	3.36 (2.25–4.98)	2.09 (1.4–3.02)	4.79 (3.39–6.42)	1.99 (1.45–2.67)
NPV (95% CI)	99.31 (98.85–99.61)	99.42 (98.91–99.74)	99.2 (98.7–99.53)	99.72 (98.98–99.97)
LR+ (95% CI)	2.50 (2.05–2.99)	1.54 (1.36–1.73)	3.61 (2.81–4.59)	1.46 (1.35–1.56)
LR- (95% CI)	0.50 (0.38–0.68)	0.42 (0.27–0.67)	0.58 (0.46–0.73)	0.2 (0.1–0.48)

	(b) Incident TB			
Statistic	**RISK11 (26)**	**QFTPlus (0.92)**	**RISK11/QFTPlus (Both-Positive)**	**RISK11/QFTPlus (Either-Positive)**
IRR 100 person-yrs (95% CI)	2.01 (1.1–3.68)	2.17 (1.03–4.59)	2.44 (1.25–4.77)	2.69 (1.05–6.94)
Sensitivity (95% CI)	39.8 (38.14–61.86)	70.71 (57.79–82.7)	28.22 (17.3–42.21)	82.29 (69.6–91.09)
Specificity (95% CI)	75.09 (73.43–76.69)	48.19 (46.31–50.07)	86.32 (84.98–87.58)	36.97 (35.16–38.8)
PPV (95% CI)	2.32 (1.59–3.26)	1.99 (1.35–2.84)	2.98 (1.91–4.46)	1.91 (1.36–2.57)
NPV (95% CI)	98.82 (98.12–99.31)	99.1 (98.98–99.97)	98.78 (98.2–99.21)	99.29 (98.36–99.77)
LR+ (95% CI)	1.6 (1.13–2.2)	1.36 (1.16–1.63)	2.06 (1.37–3.19)	1.31 (1.15–1.48)
LR- (95% CI)	0.8 (0.65–1)	0.61 (0.39–0.9)	0.83 (0.7–0.98)	0.48 (0.27–0.85)

RISK11 and QFTPlus positivity thresholds used in this table were 26% for RISK11 and 0.92UI/mL for QFTPlus, respectively. PR, Prevalence ratio. IRR, Incidence-rate ratio. PPV, Positive predictive value. NPV, Negative predictive value. LR+, Positive likelihood ratio. LR–. Negative likelihood ratio. QFTPlus, QuantiFERON-TB Gold-Plus. TB, Tuberculosis.

8.40 Appendix 6.11: Performance of RISK11 and QFTPlus as continuous biomarkers.

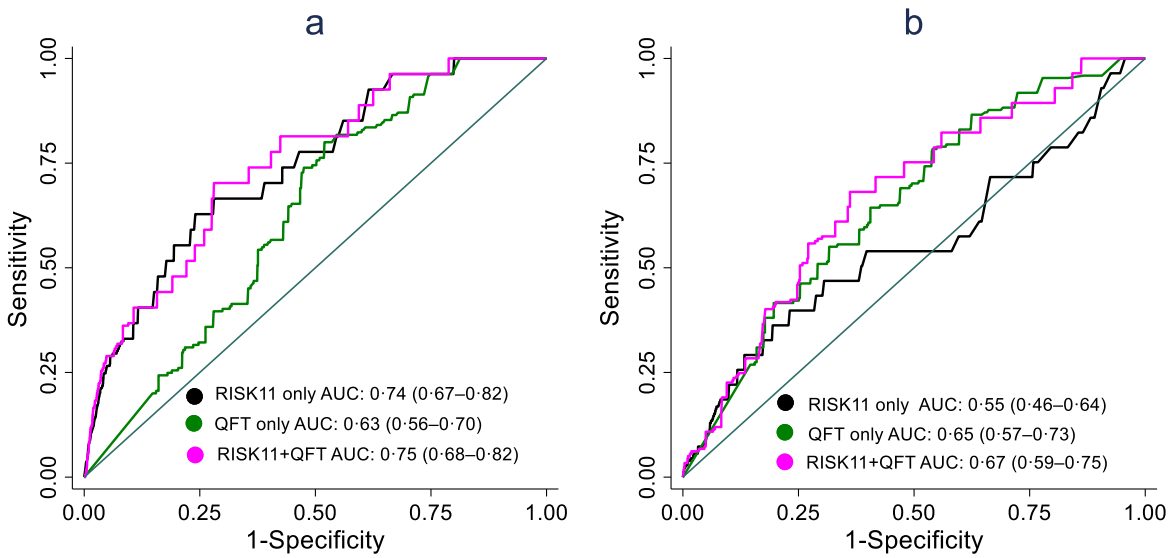


Figure 'a' is performance of RISK11 only, QFTPlus only and RISK11 plus QFTPlus for discriminating prevalent TB cases from controls and figure 'b' is performance of RISK11 only, QFTPlus only and RISK11 plus QFTPlus for discriminating incident TB cases from controls.

QFTPlus, QuantiFERON-TB Gold-Plus. TB, Tuberculosis.

8.41 Appendix 12: RISK of Incident TB disease by risk category and excluding participants that received preventive therapy for TB.

Risk Group	a) Incident TB†			b) Incident TB‡		
	Incidence Rate per 100 person-years (95% CI)	Incidence Rate Ratio (95% CI)	P	Incidence Rate per 100 person-years (95% CI)	Incidence Rate Ratio (95% CI)	P
RISK11-/QFTPlus-	0.46 (0.14–2.23)	Reference	-	0.71 (0.3–2.1)	Reference	-
RISK11+/QFTPlus-	1.83 (0.69–6.47)	4.02 (0.9–17.83)	0.07	1.53 (0.56–5.74)	2.17 (0.59–8)	0.24
RISK11-/QFTPlus+	1.78 (1.17–2.85)	3.9 (1.16–13.1)	0.03	1.59 (0.93–2.99)	2.26 (0.8–6.39)	0.13
RISK11+/ QFTPlus+	4.24 (2.84–6.63)	9.31 (2.8–31.03)	<0.001	2.91 (1.65–5.61)	4.12 (1.45–11.72)	0.01

†RISK11 and QFTPlus positivity thresholds used in (a) were 60% for RISK11 and 0.35 UI/mL for QFTPlus, respectively.

‡RISK11 and QFTPlus positivity thresholds used in (b) were optimal thresholds of 26% for RISK11 and 0.92 UI/mL for QFTPlus, respectively.

QFTPlus, QuantiFERON-TB Gold-Plus. TB, Tuberculosis.

8.42 Appendix 6.13: Additional methods and results.

Additional methods

To assess the diagnostic and prognostic performance of a RISK11/QFTPlus combination as continuous variables, a multivariable logistic and Cox proportional hazards regression model was constructed for prevalent and incident TB, respectively. Thereafter, ROC analysis was performed on the resultant continuous risk scores computed from the models.

Results of additional analysis

Distribution of severity of disease in TB cases

Of the 74 prevalent TB cases, 13 were symptomatic. In the 13 symptomatic cases, RISK11 and the Either-Positive test were positive in all, while QFTPlus and the Both-Positive test were positive in 10, respectively.

There was only one symptomatic incident case. RISK11 and the Both-Positive test were negative in the symptomatic case while QFTPlus and the Either-Positive test were positive in this one symptomatic case.

Impact of serial testing approach on number of tests conducted

An analysis was performed using the serial testing approach to assess the impact on the number of tests conducted if 10000 participants were tested.

Prevalent TB

Performing the RISK11 test first, followed by a QFTPlus only if the RISK11 result was negative, RISK11 would be expected to be negative in 9081 individuals; thus 9801 QFTPlus tests would be done on the participants with negative RISK11 results. This approach would result in a total of 19081 tests conducted. On the other hand, performing a QFTPlus first and RISK11 only in QFTPlus negative individuals would require 3665 RISK11 tests, because QFTPlus would be negative in 3665 individuals. The total number of tests to be conducted would be 13665. It must be noted that the results from this testing approach will be the same as for the Either-Positive test combination, except for the number of tests conducted.

Conversely, performing the RISK11 test first, followed by a QFTPlus only if the RISK11 result was positive, would require 981 QFTPlus tests done, as this will be the number of expected positive RISK11 results. A total of 10981 tests would need to be conducted using this approach. On the other hand, performing a QFTPlus first and RISK11 only in QFTPlus positive individuals would need 6335 RISK11 tests to be performed, because QFTPlus would be positive in 6335 individuals. The total number of

tests to be conducted would be 16335. It must be noted that the results from this testing approach will be the same as for the Both-Positive test combination, except for the number of tests conducted.

Incident TB

When RISK11 is performed first, followed by QFTPlus only if the RISK11 result is negative; 9125 QFTPlus would be expected to be conducted, because RISK11 would be negative in 9125 individuals. This would result in a total of 19125 tests done. Conversely, performing a QFTPlus first and RISK11 only in QFTPlus negative individuals would require 3663 RISK11 tests, because QFTPlus would be negative in 3663 individuals. A total of 13663 tests would need to be conducted.

If RISK11 is performed first, followed by QFTPlus only if the RISK11 result is positive, 875 QFTPlus tests would need to be done, since RISK11 would be positive in 875 individuals. Thus, a total of 10875 tests would need to be conducted. Contrastingly, performing a QFTPlus first and RISK11 only in QFTPlus positive individuals would need 6337 RISK11 tests to be performed, since QFTPlus would be positive in 6337 individuals. The resultant number of tests to be conducted would be 16337.

8.43 References

1 Chaussabel D, Quinn C, Shen J*, et al.* A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. *Immunity*. 2008;29(1):150-64.

2 Stelzer G, Rosen N, Plaschkes I*, et al.* The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Curr Protoc Bioinformatics*. 2016;54:1 30 1-1 3.

3 Rinchai D, Roelands J, Toufiq M*, et al.* BloodGen3Module: blood transcriptional module repertoire analysis and visualization using R. *Bioinformatics (Oxford, England)*. 2021;37(16):2382-9.

4 Altman MC, Rinchai D, Baldwin N*, et al.* Development of a fixed module repertoire for the analysis and interpretation of blood transcriptome data. *Nature communications*. 2021;12(1):4385.

5 Scriba TJ, Fiore-Gartland A, Penn-Nicholson A*, et al.* Biomarker-guided tuberculosis preventive therapy (CORTIS): a randomised controlled trial. *The Lancet Infectious diseases*. 2021;21(3):354-65.