

Supplementary file 3

PERFORMANCE OF HOST BLOOD TRANSCRIPTOMIC SIGNATURES FOR DIAGNOSING AND PREDICTING PROGRESSION TO TUBERCULOSIS DISEASE IN HIV-NEGATIVE ADULTS AND ADOLESCENTS: A SYSTEMATIC REVIEW PROTOCOL.

Data Extraction Form

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: _____ Threshold value _____

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 | Signature Name | Number of Genes | 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

Supplementary file 3

gave rise to the cases (d) at the end of follow-up period.

Description of negative population

^aHealthy Controls: Healthy endemics Healthy non-endemics TB Contacts Other,
describe: _____

^bOther diseases: Non-respiratory diseases Respiratory diseases

^cLTBI : TST>5mm TST>10mm IGRA N/A

If IGRA, IGRA type : QFT Elispot Other _____

^dTB 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 | Signature Name | # of Genes | 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