**Transcriptomic signatures of progression to TB disease among close contacts in Brazil: Public dataset**

Simon C Mendelsohn,‡#1 Bruno B Andrade,#2 Stanley Kimbung Mbandi,1 Alice MS Andrade,2 Vanessa M Muwanga,1 Marina C Figueiredo,3 Mzwandile Erasmus,1 Valeria C Rolla,4 Prisca K Thami,1 Marcelo Cordeiro-Santos,5 Adam Penn-Nicholson,1,6 Afranio L Kritski,7 Mark Hatherill,1 Timothy R Sterling,#3 Thomas J Scriba,#1 and the RePORT-South Africa and RePORT-Brazil Consortia

**Author affiliations:**

1South 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, 7935, Cape Town, South Africa.

2Laboratório de Pesquisa Clínica e Translacional, Instituto Gonçalo Moniz, Fundação Oswaldo Cruz, Salvador, Bahia 40296-710, Brazil.

3Vanderbilt University Medical Center, Nashville, TN 37232, USA.

4Laboratorio de Pesquisa Clinica em Micobacterioses, Instituto Nacional de Infectologia Evandro Chagas, Fiocruz, Rio de Janeiro 21040-360, Brazil.

5Fundação de Medicina Tropical Doutor Heitor Vieira Dourado

6FIND, 9 Chemin des Mines, 1202, Geneva, Switzerland.

7Universidade Federal do Rio de Janeiro, Rua Prof Rodolpho Paulo Rocco, 255, Centro de Pesquisa em Tuberculose, Cidade Universitária, 21941-913, Rio de Janeiro, RJ, Brazil.

**‡Corresponding author:** Simon C Mendelsohn

**Email:** Simon.Mendelsohn@uct.ac.za **Telephone:** +27 (0)21 406 6791

**Postal address:** 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, 7935, Cape Town, South Africa.

**#Contributed equally**

**Contents**

[Abstract 2](#_Toc155273926)

[Overview 3](#_Toc155273927)

[Table 1: TaqMan PCR primer-probe panel for Fluidigm 96.96 gene expression integrated fluidic circuit. 4](#_Toc155273928)

[Table 2: Characteristics of host blood transcriptomic signatures included in the analysis. 6](#_Toc155273929)

[Delta Ct File Data Dictionary 7](#_Toc155273930)

[Signature Scores File Data Dictionary 8](#_Toc155273931)

[References 10](#_Toc155273932)

# 

# Abstract

**Background**

Approximately 5% of people infected with *Mycobacterium tuberculosis* progress to tuberculosis (TB) disease without preventive therapy. There is a need for a prognostic test to identify those at highest risk of incident TB, so that therapy can be targeted. We evaluated host blood transcriptomic signatures for progression to TB disease.

**Methods**

Close contacts (≥4 hours exposure per week) of adult patients with culture-confirmed pulmonary TB were enrolled in Brazil. Investigation for incident, microbiologically-confirmed or clinically-diagnosed pulmonary or extra-pulmonary TB disease through 24 months of follow-up was symptom-triggered. Twenty previously validated blood TB transcriptomic signatures were measured at baseline by real-time quantitative PCR. Prognostic performance for incident TB was tested using receiver operating characteristic curve (ROC) analysis at 6, 9, 12, and 24 months of follow-up.

**Results**

Between June 2015 and June 2019, 1,854 close contacts were enrolled; Twenty-five progressed to incident TB, of whom 13 had microbiologically-confirmed disease. Baseline transcriptomic signature scores were measured in 1,789 close contacts. Prognostic performance for all signatures was best within 6 months of diagnosis. Seven signatures (Gliddon4, Suliman4, Roe3, Roe1, Penn-Nicholson6, Francisco2, and Rajan5) met the minimum World Health Organization target product profile (TPP) for a prognostic test through 6 months; three (Gliddon4, Rajan5, and Duffy9) through 9 months. None met the TPP threshold through 12 or more months of follow-up.

**Conclusions**

Blood transcriptomic signatures may be useful for predicting TB risk within 9 months of measurement among TB-exposed contacts, to target preventive therapy administration.

**Funding**

This work was supported by the U.S. National Institutes of Health, CRDF Global, and Departamento de Ciência e Tecnologia (DECIT) - Secretaria de Ciência e Tecnologia (SCTIE) – Ministério da Saúde (MS), Brazil.

# Overview

This is a public, subject-level dataset containing key variables necessary to reconstruct the study findings. A data dictionary is provided below. This dataset includes all available Fluidigm RT-qPCR data (**Table 1**) and transcriptomic signature scores (**Table 2**) from close contacts enrolled in the study with data at the baseline and/or month 6 timepoints.

**Files included**

1. TaqMan gene expression assay delta cycle threshold (Ct) data: normalized Ct data (not batch corrected) [RePORT-Brazil\_CohortB\_deltaCts.csv]
2. Signature score data [RePORT-Brazil\_CohortB\_signatureScores.csv]:
   1. Original signature scores (non-batch corrected, non-normalized)
   2. Batch-corrected signature scores (non-normalized)
   3. Batch corrected, normalized signature scores

# Table 1: TaqMan PCR primer-probe panel for Fluidigm 96.96 gene expression integrated fluidic circuit.

| **Gene Symbol** | **TaqMan Assay ID** | **Assay excluded\*\*** | **Reference probe** | **# of signatures** | **da Costa3** | **de Araujo1** | **Duffy9 (10)†** | **Francisco2** | **Gjøen7** | **Gliddon3** | **Gliddon4** | **Jacobsen3** | **Kaforou22 (27)‡** | **Maertzdorf4** | **Penn-Nicholson6** | **Rajan5** | **Roe1** | **Roe3** | **Sambarey10** | **Satproedprai7** | **Suliman2** | **Suliman4** | **Sweeney3** | **Thompson5** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ACTR3 | Hs01029159\_g1 |  | X | 14 | X | X | X | X | X | X | X | X | X |  |  | X | X | X | X | X |  |  |  |  |
| TMBIM6 | Hs00162661\_m1 |  | X | 14 | X | X | X | X | X | X | X | X | X |  |  | X | X | X | X | X |  |  |  |  |
| USF2 | Hs01100994\_g1 |  | X | 14 | X | X | X | X | X | X | X | X | X |  |  | X | X | X | X | X |  |  |  |  |
| ACTA2 | Hs00426835\_g1 |  |  | 1 |  |  |  |  |  |  |  |  |  |  |  | X |  |  |  |  |  |  |  |  |
| ANKRD22 | Hs00944015\_m1 |  |  | 2 |  |  |  |  |  |  |  |  | X |  |  |  |  |  |  |  | X |  |  |  |
| APOL1 | Hs01066280\_m1 |  |  | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | X |  |  |  |  |
| ARG1 | Hs00968979\_m1 |  |  | 1 |  |  |  |  |  |  | X |  |  |  |  |  |  |  |  |  |  |  |  |  |
| BATF2 | Hs00912737\_m1 |  |  | 2 |  |  |  |  |  |  |  |  |  |  |  |  | X | X |  |  |  |  |  |  |
| BCL6 | AREPURD\* |  |  | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  | X |  |  |  |  |  |
| BLK | Hs01017452\_m1 |  |  | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | X |  |  |
| C1QA | Hs00706358\_s1 | X |  | 0 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| C1QB | Hs00608019\_m1 |  |  | 3 |  |  | X |  |  | X |  |  | ‡ |  |  |  |  |  |  |  |  |  |  |  |
| C1QC | Hs00757779\_m1 | X |  | 1 |  |  |  |  |  |  |  |  | ‡ |  |  |  |  |  |  |  |  |  |  |  |
| C4ORF18 / FAM198B | Hs00259260\_s1 |  |  | 1 |  |  |  |  |  |  |  |  | X |  |  |  |  |  |  |  |  |  |  |  |
| C5 | Hs00156197\_m1 |  |  | 1 |  |  |  |  |  |  |  |  | X |  |  |  |  |  |  |  |  |  |  |  |
| CCR6 | Hs01890706\_s1 |  |  | 1 |  |  |  |  |  |  |  |  | X |  |  |  |  |  |  |  |  |  |  |  |
| CD160 | Hs00199894\_m1 |  |  | 1 |  |  | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CD1C | Hs00957534\_g1 |  |  | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | X |  |  |
| CD36.2 | Hs01567186\_m1 |  |  | 1 |  |  | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CD3E | APDJ3NT\* |  |  | 1 |  |  |  |  | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CD40L | Hs00163934\_m1 |  |  | 1 |  |  | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CD74-j1 | Hs04983808\_s1 | X |  | 0 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CD74-j2 | AP9HN47\* | X |  | 0 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CD79A | Hs00998120\_g1 |  |  | 1 |  |  |  |  |  |  |  |  | X |  |  |  |  |  |  |  |  |  |  |  |
| CD79B | Hs01058826\_g1 |  |  | 1 |  |  |  |  |  |  |  |  | X |  |  |  |  |  |  |  |  |  |  |  |
| CDCA7 | Hs00230589\_m1 | X |  | 0 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CDKN1C | Hs00175938\_m1 | X |  | 0 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| CXCR5 | Hs00540548\_s1 |  |  | 1 |  |  |  |  |  |  |  |  | X |  |  |  |  |  |  |  |  |  |  |  |
| CYP4F3 | Hs01587860\_mH |  |  | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  | X |  |  |  |  |  |
| DUSP3 | Hs01115776\_m1 |  |  | 2 |  |  |  |  |  |  |  |  | X |  |  |  |  |  |  |  |  |  | X |  |
| F2RL1 | Hs00608346\_m1 | X |  | 0 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| FAM20A | Hs01034066\_m1 |  |  | 1 |  |  |  |  |  |  |  |  | X |  |  |  |  |  |  |  |  |  |  |  |
| FCGR1A | Hs00174081\_m1 |  |  | 5 | X |  |  |  |  | X |  | X | X |  |  |  |  |  | X |  |  |  |  |  |
| FCGR1A | AP2XDMK\* |  |  | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | X |  |  |  |  |
| FCGR1B | Hs02341825\_m1 |  |  | 3 |  |  | X |  |  |  |  |  | X |  | X |  |  |  |  |  |  |  |  |  |
| FCGR1B\_ VARIANT1 | AP3267H\* |  |  | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | X |  |  |  |  |
| FCGR1B\_ VARIANT2 | AP47ZTF\* |  |  | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | X |  |  |  |  |
| FCGR1C | Hs00417598\_m1 |  |  | 1 |  |  |  |  |  |  |  |  | X |  |  |  |  |  |  |  |  |  |  |  |
| FLVCR2 | Hs00900390\_m1 |  |  | 1 |  |  |  |  |  |  |  |  | X |  |  |  |  |  |  |  |  |  |  |  |
| GAS6 | AR47XGZ\* |  |  | 2 |  |  |  |  |  |  |  |  | X |  |  |  |  |  |  |  |  | X |  |  |
| GAS6 | AR323W3\* |  |  | 1 |  |  |  |  |  |  |  |  | X |  |  |  |  |  |  |  |  |  |  |  |
| GBP1 | Hs00977005\_m1 |  |  | 1 |  |  |  |  |  |  |  |  |  | X |  |  |  |  |  |  |  |  |  |  |
| GBP2 | Hs00894846\_g1 |  |  | 1 |  |  |  |  |  |  |  |  |  |  | X |  |  |  |  |  |  |  |  |  |
| GBP5 | Hs00369472\_m1 |  |  | 5 | X |  |  | X | X |  |  |  |  |  |  |  |  | X |  |  |  |  | X |  |
| GBP6 | Hs01584201\_m1 |  |  | 3 |  |  | X |  |  |  | X |  | ‡ |  |  | X |  |  |  |  |  |  |  |  |
| GNG7 | Hs00192999\_m1 |  |  | 1 |  |  |  |  |  |  |  |  | X |  |  |  |  |  |  |  |  |  |  |  |
| GYG1 | Hs00907542\_g1 |  |  | 1 |  |  |  |  |  |  |  |  |  |  |  | X |  |  |  |  |  |  |  |  |
| GZMA | Hs00989184\_m1 |  |  | 1 | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| HK3 | Hs01092850\_m1 |  |  | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  | X |  |  |  |  |  |
| ID3 | Hs00954037\_g1 |  |  | 2 |  |  | X |  |  |  |  |  |  | X |  |  |  |  |  |  |  |  |  |  |
| IFI44L | Hs00915292\_m1 |  |  | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  | X |  |  |  |  |  |
| IFITM3 | Hs03057129\_s1 |  |  | 2 |  |  |  |  | X |  |  |  |  | X |  |  |  |  |  |  |  |  |  |  |
| IFITM3\_G | APMF24C\* | X |  | 0 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| KAZN | AP7DVDD\* |  |  | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | X |  |  |  |  |
| KIF1B | Hs01114512\_g1 |  |  | 1 |  |  |  |  | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| KLF2 | Hs00360439\_g1 |  |  | 2 |  |  |  | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  | X |  |
| KLF2 | ARH6CF4\* | X |  | 0 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| KLHDC8B | Hs00293902\_m1 | X |  | 0 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| LAG3 | Hs00958444\_g1 |  |  | 1 |  |  | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| LHFPL2 | Hs00299613\_m1 |  |  | 1 |  |  |  |  |  |  |  |  | X |  |  |  |  |  |  |  |  |  |  |  |
| LTF | Hs00158924\_m1 |  |  | 1 |  |  |  |  |  |  |  | X |  |  |  |  |  |  |  |  |  |  |  |  |
| MAFB | APZTH2N\* |  |  | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | X |  |  |  |  |
| MAP7D3 | Hs00226257\_m1 |  |  | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | X |
| MMP9 | APEPW9P\* |  |  | 1 |  |  |  |  | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| MPO | Hs00165162\_m1 |  |  | 1 |  |  |  |  |  |  |  |  | X |  |  |  |  |  |  |  |  |  |  |  |
| MTRF1L | Hs01097882\_g1 |  |  | 1 |  |  |  |  |  |  |  |  |  |  |  | X |  |  |  |  |  |  |  |  |
| NOD2 | Hs01550759\_g1 |  |  | 1 |  |  |  |  | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| NPC2 | Hs01119244\_m1 |  |  | 1 |  | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| OSBPL10 | Hs00215016\_m1 |  |  | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | X |  |  |  |
| P2RY14 | Hs01848195\_s1 |  |  | 1 |  |  |  |  |  |  |  |  |  | X |  |  |  |  |  |  |  |  |  |  |
| PRDM1 | Hs00153357\_m1 |  |  | 1 |  |  |  |  |  |  | X |  |  |  |  |  |  |  |  |  |  |  |  |  |
| RAB13 | Hs04400188\_g1 |  |  | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  | X |  |  |  |  |  |
| Rab33A | Hs00191243\_m1 |  |  | 1 |  |  |  |  |  |  |  | X |  |  |  |  |  |  |  |  |  |  |  |  |
| RABL2A | Hs00255244\_m1 |  |  | 1 |  |  |  |  |  |  |  |  |  |  |  | X |  |  |  |  |  |  |  |  |
| RBBP8 | Hs01090329\_m1 |  |  | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  | X |  |  |  |  |  |
| RP11-295G20.2 | Hs01373568\_m1 |  |  | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | X |
| S100A8 | Hs00374264\_g1 |  |  | 1 |  |  |  |  |  |  |  |  | X |  |  |  |  |  |  |  |  |  |  |  |
| SCARF1 | Hs01092483\_m1 |  |  | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  | X |  |  |  |  |  |  |
| SDR39U1 | Hs01016970\_g1 |  |  | 1 |  |  |  |  |  |  |  |  |  |  | X |  |  |  |  |  |  |  |  |  |
| SEPT4 | Hs00910208\_g1 |  |  | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | X |  |  |
| SERPING1 | Hs00934329\_m1 |  |  | 1 |  |  |  |  |  |  |  |  |  |  | X |  |  |  |  |  |  |  |  |  |
| SLPI | Hs00268206\_m1 |  |  | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  | X |  |  |  |  |  |
| SMARCD3 | Hs01088251\_g1 |  |  | 3 |  |  |  |  |  |  |  |  | X |  |  |  |  |  | X |  |  |  |  | X |
| STAT1 | Hs01013996\_m1 |  |  | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | X |  |  |  |  |
| STT3A | Hs00967491\_m1 |  |  | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | X |
| TIMM10 | APT2DEC\* |  |  | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  | X |  |  |  |  |  |
| TMCC1 | Hs01037666\_s1 |  |  | 1 |  |  |  |  |  |  | X |  |  |  |  |  |  |  |  |  |  |  |  |  |
| TNIP1 | AP47ZYZ\* |  |  | 1 |  |  |  |  | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| TRMT2A | Hs01000041\_g1 |  |  | 1 |  |  |  |  |  |  |  |  |  |  | X |  |  |  |  |  |  |  |  |  |
| TUBGCP6 | Hs00363509\_g1 |  |  | 1 |  |  |  |  |  |  |  |  |  |  | X |  |  |  |  |  |  |  |  |  |
| UCP2 | Hs01075224\_g1 |  |  | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | X |
| VAMP5 | Hs01105383\_g1 |  |  | 1 |  |  |  |  |  |  |  |  | X |  |  |  |  |  |  |  |  |  |  |  |
| WARS-j1 | Hs00998737\_m1 | X |  | 0 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| WARS-j2 | APAAFHX\* | X |  | 0 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| ZDHHC19 | Hs00376118\_m1 |  |  | 1 |  |  | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| ZNF296 | Hs00377132\_m1 |  |  | 2 |  |  |  |  |  | X |  |  | X |  |  |  |  |  |  |  |  |  |  |  |

Signatures are named by first author and number of transcripts included in the model (e.g. Author11). Numbers in brackets indicate the original number of transcripts in the published model. Some signatures have a reduced number of transcripts when translated to RT-qPCR due to duplicate transcript symbols (IDs), high-primer probe failure rate (\*\*poor amplification efficiency), or where transcript sequences from an original discovery cohorts could not be mapped to a more recent reference transcriptome.

\* Custom designed primer-probe assays; available on request to corresponding author.

† The CERKL1 primer-probe assay failed during panel optimisation and was removed from the Duffy10 signature.

‡ Three transcripts (C1QB, C1QC, and GBP6) were excluded from the Kaforou27 signature due to a high primer-probe assay failure rate (\*\*). Unique primer-probe assays could not be designed for both of the FCGR1B transcript variants; only one FCGR1B primer-probe was included in the Kaforou22 model. We were unable to design a primer-probe for LOC728744; all information had been withdrawn from both the NCBI and RefSeq databases.

# Table 2: Characteristics of host blood transcriptomic signatures included in the analysis.

| **Signature \*** | **Reference** | **Signature model** | **Discovery cohort country** | **Discovery cohort age group** | **Discovery cohort application** |
| --- | --- | --- | --- | --- | --- |
| da Costa3 | *Tuberculosis* (2015)1 | Random forest | Brazil | Adults | Diagnostic; TB vs LTBI, HC, and OD |
| de Araujo1 (NPC2) | *Front Microbiol* (2016)2 | Standardised expression | Brazil | Adults | Diagnostic; TB vs LTBI and HC |
| Duffy9 (10)† | *PLoS One* (2019)3 | Six-class multinomial random forest | South Africa,  Malawi | Adults | Diagnostic, TB vs LTBI and OD |
| Francisco2 | *J Infect* (2017)4 | Random forest | China | Adults | Diagnostic; TB vs OD and HC |
| Gjøen7 | *Sci Rep* (2017)5 | LASSO regression | India | Children | Diagnostic; TB vs HC |
| Gliddon3 | *Front Immunol* (2021)6 | Disease risk score | South Africa,  Malawi | Adults | Diagnostic; TB vs LTBI |
| Gliddon4 | *Front Immunol* (2021)6 | Disease risk score | South Africa,  Malawi | Adults | Diagnostic; TB vs OD |
| Jacobsen3 | *J Mol Med* (2007)7 | Linear discriminant analysis | Germany | Adults | Diagnostic; TB vs LTBI and HC |
| Kaforou22 (27)‡ | *PLoS Med* (2013)8 | Disease risk score | South Africa,  Malawi | Adults | Diagnostic; TB vs LTBI |
| Maertzdorf4 | *EMBO Mol Med* (2016)9 | RawCTID3 – [ (RawCTGBP1 + RawCTIFITM3 + RawCTP2RY14) / 3]  (adapted from Suliman et al. 10) | India | Adults | Diagnostic; TB vs HC and LTBI |
| Penn-Nicholson6 | *Sci Rep* (2020)11 | Pair-wise ensemble structure | South Africa | Adolescent | Prognostic; TB progressors vs non-progressors |
| Rajan5 | *Clin Infect Dis* (2019)12 | Unsigned sums | Uganda | Adults | Diagnostic; HIV+ TB vs HIV+ HC |
| Roe1 (BATF2) | *JCI Insight* (2016)13 | Standardised expression | United Kingdom | Adults | Diagnostic; TB vs HC and TB patients 2-4 years post recovery |
| Roe3 | *Clin Infect Dis* (2020)14 | SVM (linear kernel) | United Kingdom | Adults | Prognostic; TB progressors vs non-progressors |
| Sambarey10 | *EBioMedicine* (2017)15 | Linear discriminant analysis | India | Adults | Diagnostic; TB vs HC and LTBI |
| Satproedprai7 | *Genes Immun* (2015)16 | LASSO regression | Thailand | Adults | Diagnostic; TB vs HC and previous TB patients |
| Suliman2 | *Am J Respir Crit Care Med* (2018)10 | Pair-wise ensemble structure | The Gambia,  South Africa | Adults | Prognostic; TB progressors vs non-progressors |
| Suliman4 | *Am J Respir Crit Care Med* (2018)10 | Pair-wise ensemble structure | The Gambia,  South Africa | Adults | Prognostic; TB progressors vs non-progressors |
| Sweeney3 | *Lancet Respir Med* (2016)17; *JAMA Netw Open* (2018)18 | (RawCTGBP5 + RawCTDUSP3) / 2 – RawCTKLF2 | France, Malawi, South Africa, United Kingdom, USA | Adults | Diagnostic; TB vs HC + LTBI and OD |
| Thompson5 | *Tuberculosis* (2017)19 | Pair-wise ensemble structure | South Africa | Adults | Monitoring TB treatment response |

\* Signatures are named by first author and number of transcripts included in the model (e.g. Author11). Numbers in brackets indicate the original number of transcripts in the published model. Some signatures have a reduced number of transcripts when translated to RT-qPCR due to duplicate transcript symbols (IDs), high-primer probe failure rate (poor amplification efficiency), or where transcript sequences from an original discovery cohorts could not be mapped to a more recent reference transcriptome.

† The CERKL1 primer-probe assay failed during panel optimisation and was removed from the Duffy10 signature.

‡ Three transcripts (C1QB, C1QC, and GBP6) were excluded from the Kaforou27 signature due to a high primer-probe assay failure rate. Unique primer-probe assays could not be designed for both of the FCGR1B transcript variants; only one FCGR1B primer-probe was included in the Kaforou22 model. We were unable to design a primer-probe for LOC728744; all information had been withdrawn from both the NCBI and RefSeq databases.

LTBI, latent tuberculosis infection. OD, other diseases. HC, healthy controls. SVM, support vector machines. LASSO, Least Absolute Shrinkage and Selection Operator.

# Delta Ct File Data Dictionary

The csv file, “RePORT-Brazil\_CohortB\_deltaCts.csv”, contains delta Ct results: Ct values for each study participant, that have been normalized to the geometric mean of three housekeeping genes on the panel (TMBIM6.Hs00162661\_m1, USF2.Hs01100994\_g1, and ACTR3.Hs01029159\_g1).

| **Variable Name** | **Variable Description** | **Format** | **Example Values** |
| --- | --- | --- | --- |
| 1. pubid | Unique anonymized public participant ID number | char | 1. 101312,  2. 972478…. |
| 1. TBstatus | TB endpoint status | char | 1. Non-progressor,  2. Progressor |
| 1. microConfirmation | Microbiological confirmation of TB diagnosis (1), or clinical diagnosis of TB (0) | binary | 1. 0,  2. 1 |
| 1. diagnosis | Pulmonary TB (PTB) or extrapulmonary TB (EPTB) | char | 1. PTB, 2. EPTB |
| 1. TimeToTB | Time in months from enrolment and sample collection until progression to incident TB disease (Progressors only) | num | 1. NA (non-progressor), 2. 4, 3. 15… |
| 1. Visit | Study visit (enrolment or month 6) | char | 1. M0, 2. M6 |
| 1. QFT\_result | Qualitative QuantiFERON TB-Gold In-Tube assay (QFT) result. QFT result is positive if the max of QFT score for tb1-nil and tb2-nil antigens is ≥ 0.35. | char | 1. Negative, 2. Positive, 3. Indeterminate |
| 1. QFT\_numeric | Quantitative QuantiFERON TB-Gold In-Tube assay (QFT) result. Max of QFT score for tb1-nil and tb2-nil antigens | num | 1. 0 2. 3.4 3. 1.3… |
| 1. chipBarcode | Unique identification number for each Fluidigm microfluidic Gene Expression (GE) Integrated Fluidic Circuit (IFC) run | num | 1. 1362633279,  2. 1362633136… |
| 1. ThermoFisher TaqMan Gene Expression Assays   e.g. ACTA2.Hs00426835\_g1  ACTA2= Gene symbol  Hs00426835\_g1 = Assay ID | Delta Ct value for each TaqMan assay  “NA” implies a failed reaction or reaction which did not pass assay or sample quality control checks. | num | 1. 7.112558,  2. 6.652142,  3. -1.972579… |

# Signature Scores File Data Dictionary

The csv file, “RePORT-Brazil\_CohortB\_signatureScores.csv”, contains signature scores for each study participant, at each time-point/visit:

1. Original signature scores (non-batch corrected, non-normalized)
2. Batch-corrected signature scores (non-normalized)
3. Batch corrected, normalized signature scores

Original signature scores were batch corrected to account for differences in signature score distribution between Fluidigm chip runs (chipBarcode) using a quantile regression method with the adjust\_batch function in the batchtma R package.

Normalized scores were derived from batch-corrected z-scores using the cumulative distribution function (pnorm function, stats R package).

| **Variable Name** | **Variable Description** | **Format** | **Example Values** |
| --- | --- | --- | --- |
| 1. pubid | Unique anonymized public participant ID number | char | 1. 101312,  2. 972478…. |
| 1. TBstatus | TB endpoint status | char | 1. Non-progressor,  2. Progressor |
| 1. microConfirmation | Microbiological confirmation of TB diagnosis (1), or clinical diagnosis of TB (0) | binary | 1. 0,  2. 1 |
| 1. diagnosis | Pulmonary TB (PTB) or extrapulmonary TB (EPTB) | char | 1. PTB, 2. EPTB |
| 1. TimeToTB | Time in months from enrolment and sample collection until progression to incident TB disease (Progressors only) | num | 1. NA (non-progressor), 2. 4, 3. 15… |
| 1. Visit | Study visit (enrolment or month 6) | char | 1. M0, 2. M6 |
| 1. QFT\_result | Qualitative QuantiFERON TB-Gold In-Tube assay (QFT) result. QFT result is positive if the max of QFT score for tb1-nil and tb2-nil antigens is ≥ 0.35. | char | 1. Negative, 2. Positive, 3. Indeterminate |
| 1. QFT\_numeric | Quantitative QuantiFERON TB-Gold In-Tube assay (QFT) result. Max of QFT score for tb1-nil and tb2-nil antigens | num | 1. 0 2. 3.4 3. 1.3… |
| 1. chipBarcode | Unique identification number for each Fluidigm microfluidic Gene Expression (GE) Integrated Fluidic Circuit (IFC) run | num | 1. 1362633279,  2. 1362633136… |
| 1. da Costa3, de Araujo1 Duffy9, Francisco2, Gjøen7, Gliddon3, Gliddon4, Jacobsen3, Kaforou22, Maertzdorf4, Penn-Nicholson6, Rajan5, Roe1, Roe3, Sambarey10, Satproedprai7, Suliman2, Suliman4, Sweeney3, and Thompson5   Suffix:  No suffix = Original signature scores (non-batch corrected, non-normalized)  “\_batchCorrected” = Batch-corrected, non-normalized signature scores for each sample  “\_normalised” = Batch-corrected, normalized signature scores for each sample (range 0-1) | Signature scores for each sample  “NA” implies that the signature score could not be calculated due to a missing transcript Ct value. | num | 1. 0.13203337, 2. 1.7556480, … |

# References

1. da Costa LL, Delcroix M, Costa ERD, Prestes IV, Milano M, Francis SS, et al. A real-time PCR signature to discriminate between tuberculosis and other pulmonary diseases. Tuberculosis. 2015;95(4):421-5. doi: 10.1016/j.tube.2015.04.008.

2. de Araujo LS, Vaas LA, Ribeiro-Alves M, Geffers R, Mello FC, de Almeida AS, et al. Transcriptomic Biomarkers for Tuberculosis: Evaluation of DOCK9. EPHA4, and NPC2 mRNA Expression in Peripheral Blood. Front Microbiol. 2016;7:1586. doi: 10.3389/fmicb.2016.01586.

3. Duffy FJ, Thompson EG, Scriba TJ, Zak DE. Multinomial modelling of TB/HIV co-infection yields a robust predictive signature and generates hypotheses about the HIV+TB+ disease state. PLoS ONE. 2019;14(7):e0219322. doi: 10.1371/journal.pone.0219322.

4. Francisco NM, Fang YM, Ding L, Feng S, Yang Y, Wu M, et al. Diagnostic accuracy of a selected signature gene set that discriminates active pulmonary tuberculosis and other pulmonary diseases. Journal of Infection. 2017;75(6):499-510. doi: 10.1016/j.jinf.2017.09.012.

5. Gjoen JE, Jenum S, Sivakumaran D, Mukherjee A, Macaden R, Kabra SK, et al. Novel transcriptional signatures for sputum-independent diagnostics of tuberculosis in children. Sci Rep. 2017;7(1):5839. doi: 10.1038/s41598-017-05057-x.

6. Gliddon HD, Kaforou M, Alikian M, Habgood-Coote D, Zhou C, Oni T, et al. Identification of Reduced Host Transcriptomic Signatures for Tuberculosis Disease and Digital PCR-Based Validation and Quantification. Front Immunol. 2021;12:637164. doi: 10.3389/fimmu.2021.637164.

7. Jacobsen M, Repsilber D, Gutschmidt A, Neher A, Feldmann K, Mollenkopf HJ, et al. Candidate biomarkers for discrimination between infection and disease caused by Mycobacterium tuberculosis. J Mol Med (Berl). 2007;85(6):613-21. doi: 10.1007/s00109-007-0157-6.

8. Kaforou M, Wright VJ, Oni T, French N, Anderson ST, Bangani N, 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. doi: 10.1371/journal.pmed.1001538.

9. Maertzdorf J, McEwen G, Weiner J, 3rd, Tian S, Lader E, Schriek U, et al. Concise gene signature for point-of-care classification of tuberculosis. EMBO Molecular Medicine. 2016;8(2):86-95. doi: 10.15252/emmm.201505790.

10. Suliman S, Thompson EG, Sutherland J, Weiner J, 3rd, Ota MOC, Shankar S, 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. doi: 10.1164/rccm.201711-2340OC.

11. Penn-Nicholson A, Mbandi SK, Thompson E, Mendelsohn SC, Suliman S, Chegou NN, et al. RISK6, a 6-gene transcriptomic signature of TB disease risk, diagnosis and treatment response. Sci Rep. 2020;10(1):8629. doi: 10.1038/s41598-020-65043-8.

12. Rajan JV, Semitala FC, Mehta T, Seielstad M, Montalvo L, Andama A, et al. A Novel, 5-Transcript, Whole-blood Gene-expression Signature for Tuberculosis Screening Among People Living With Human Immunodeficiency Virus. Clinical Infectious Diseases. 2019;69(1):77-83. doi: 10.1093/cid/ciy835.

13. Roe JK, Thomas N, Gil E, Best K, Tsaliki E, Morris-Jones S, et al. Blood transcriptomic diagnosis of pulmonary and extrapulmonary tuberculosis. JCI Insight. 2016;1(16):e87238. doi: 10.1172/jci.insight.87238.

14. Roe J, Venturini C, Gupta RK, Gurry C, Chain BM, Sun Y, et al. Blood Transcriptomic Stratification of Short-term Risk in Contacts of Tuberculosis. Clinical Infectious Diseases. 2020;70(5):731-7. doi: 10.1093/cid/ciz252.

15. Sambarey A, Devaprasad A, Mohan A, Ahmed A, Nayak S, Swaminathan S, et al. Unbiased Identification of Blood-based Biomarkers for Pulmonary Tuberculosis by Modeling and Mining Molecular Interaction Networks. EBioMedicine. 2017;15:112-26. doi: 10.1016/j.ebiom.2016.12.009.

16. Satproedprai N, Wichukchinda N, Suphankong S, Inunchot W, Kuntima T, Kumpeerasart 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. doi: 10.1038/gene.2015.4.

17. 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. doi: 10.1016/S2213-2600(16)00048-5.

18. Warsinske HC, Rao AM, Moreira FMF, Santos PCP, Liu AB, Scott M, 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. doi: 10.1001/jamanetworkopen.2018.3779.

19. Thompson EG, Du Y, Malherbe ST, Shankar S, Braun J, Valvo J, et al. Host blood RNA signatures predict the outcome of tuberculosis treatment. Tuberculosis (Edinb). 2017;107:48-58. doi: 10.1016/j.tube.2017.08.004.