CORTIS-HR Public Dataset

1. Overview and Abstract

The **Validation of Correlates of Risk of TB Disease in High Risk Populations (CORTIS-HR) Study**, a companion study of the CORTIS-01 Trial (ClinicalTrials.gov: NCT02735590), was conducted to test the diagnostic and prognostic performance of the RISK11 biomarker for tuberculosis (TB) disease in people living with HIV (PLHIV) in an ambulant community setting.

The "CORTIS-HR pubdata.csv" is a public, subject-level dataset for the CORTIS-HR study containing key variables necessary to reconstruct the study findings. A data dictionary is provided below. The "CORTIS-HR PCR data.csv" provides subject-level TaqMan qPCR probe raw CT (cycle threshold) gene expression data from the Fluidigm microfluidic 96.96 Gene Expression Integrated Fluidic Circuits (chips) with sample quality control ("SAMPLE_QC") results. Analyses of the qPCR probe data are ongoing; the embargo on this data ends 1 July 2021 when the data will be published on ZivaHub. "CORTIS-HR Protocol Version 1.0.pdf" and "CORTIS-HR SAP Version 1.0.pdf" are the protocol and the statistical analysis plan for the study respectively and have been included for reference. The following is quoted from the abstract of the accepted manuscript:

Prospective validation of a host blood transcriptomic biomarker for pulmonary tuberculosis in people living with HIV: a diagnostic and prognostic accuracy study

Simon C. Mendelsohn, MBChB[1], Andrew Fiore-Gartland, PhD[2], Adam Penn-Nicholson, PhD[1], Humphrey Mulenga, MPH[1], Stanley Kimbung Mbandi, PhD[1], Bhavesh Borate, MS[2], Katie Hadley, PhD[1], Chris Hikuam, PhD[1], Munyaradzi Musvosvi, PhD[1], Nicole Bilek, PhD[1], Mzwandile Erasmus, BSc[1], Lungisa Jaxa, BTech[1], Rodney Raphela, BSc[1], Onke Nombida, GT[1], Masooda Kaskar, MBChB[1], Tom Sumner, PhD[3], Richard G. White, PhD[3], Craig Innes, MBChB[4], William Brumskine, MBChB[4], Andriëtte Hiemstra, MBChB[5], Stephanus T. Malherbe, MBChB[5], Razia Hassan-Moosa, MPH[6,7], Michèle Tameris, MBChB[1], Gerhard Walzl, MBChB[5], Kogieleum Naidoo, PhD[6,7], Gavin Churchyard, PhD[4,8], Thomas J. Scriba, DPhil[1*], Mark Hatherill, MD[1*] and the CORTIS-HR Study Team.

*TJS and MH contributed equally

[1]South African Tuberculosis Vaccine Initiative, Institute of Infectious Disease and Molecular Medicine, Division of Immunology, Department of Pathology, University of Cape Town, South Africa.
[2]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA.
[3]TB Modelling Group, TB Centre, Centre for Mathematical Modelling of Infectious Diseases, Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, United Kingdom.
[4]The Aurum Institute, Johannesburg, Gauteng, South Africa.
[5]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, Cape Town, South Africa.
[6]Centre for the AIDS Programme of Research in South Africa (CAPRISA), Durban, South Africa.
[7]MRC-CAPRISA HIV-TB Pathogenesis and Treatment Research Unit, Doris Duke Medical Research Institute, University of KwaZulu-Natal, Durban, South Africa.
[8]School of Public Health, University of Witwatersrand, Johannesburg, South Africa.

Corresponding author: Mark Hatherill **Email:** Mark.Hatherill@uct.ac.za **Telephone:** +27 (0) 21 406 6080

Postal address: 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, 7935, Cape Town, South Africa

Abstract

Background

We tested diagnostic and prognostic performance of a host blood transcriptomic signature of tuberculosis (RISK11) for screening of people living with HIV (PLHIV) in a prospective, community-based cohort.

Methods

Ambulant adult volunteers living with HIV were enrolled at five South African sites. RISK11 status was assessed at baseline by real-time PCR; Participants and study staff were blinded to the result. Participants underwent active surveillance for microbiologically-confirmed tuberculosis from enrolment through 15 months. The primary outcomes were prevalence and cumulative incidence of two-sputum-sample positive tuberculosis in RISK11+ versus RISK11- participants.

Findings

Among 820 participants with valid RISK11 results, eight (1%) tuberculosis cases were identified at baseline. Risk of prevalent tuberculosis was 13·1-fold (95%CI 2·1-81·6) greater in RISK11+ than RISK11- participants, with tuberculosis prevalence of 2·5% and 0·2%, respectively. RISK11 had diagnostic area under the receiver-operating-characteristic curve (AUC) 88·2%; sensitivity 87·5% and specificity 65·8% at the pre-defined threshold (60%).

Thereafter, eight tuberculosis cases were identified through median 15 months follow-up. Tuberculosis incidence was 2·5 vs 0·2 per 100 person-years in RISK11+ compared to RISK11- participants with cumulative incidence ratio 16·0 (95%CI 2·0-129·5); AUC 80·0%; sensitivity 88·6% and specificity 68·9%. By comparison, QuantiFERON TB Gold-Plus (QFT) had a cumulative incidence ratio of 2·0 (95%CI 0·5-8·4); AUC 70·8%; sensitivity 62·1% and specificity 56·2%.

Interpretation

RISK11 identified prevalent tuberculosis and predicted risk of progression to incident tuberculosis within 15 months in ambulant PLHIV. Performance approached the World Health Organization Target Product Profile benchmarks for screening and prognostic tests for tuberculosis. QFT performance fell short of the prognostic benchmark.

Funding

Bill & Melinda Gates Foundation, South African Medical Research Council

Sponsor

University of Cape Town

2. Privacy of Information Disclosed

While the Fred Hutchinson Cancer Research Center and by extension the Statistical Center for HIV/AIDS Research and Prevention (SCHARP) is not a covered entity regarding HIPAA requirements, SCHARP uses the 'limited dataset' model as described in HIPAA when publishing public use datasets. This limited Dataset is used for public health and research purposes only. No personal identifying information is used in the Datasets, such as names or social security numbers. The Dataset uses its own unique identifiers that cannot identify the research participants (i.e., the Dataset identifiers are not participant IDs used in the research). SCHARP and South African Tuberculosis Vaccine Initiative (SATVI) at the University of Cape Town (UCT) staff are trained on and sign an agreement to follow SCHARP's and the UCT's privacy and confidentiality policies.

These procedures comply with relevant models required in FDA regulations such as Title 21 CFR Parts 20, 21, 803, Title 45 CFR Part 164.514(e)(3)(i), and as recommended in Good Clinical Data Management Practices (GCDMP) defined by the Society of Data Quality Management.

3. Subject-Level Dataset Data Dictionary with Example Values

Variable Name	Variable Description	Format	Example Values
1) pubid	Unique identification number for each subject in trial	char	1. 100001, 2. 356282 …
2) RISK11_D0_score	RISK11 score at enrolment	num	1. 19.895, 2. 60.234, …
3) RISK11_D0_status	RISK11_D0_status is positive if RISK11 score is ≥60	char	1. “RISK11+”, 2. “RISK11-”, 3. “INDETERMINATE”
4) IGRA_D0_score	IGRA (QuantiFERON TB Gold-Plus) score at enrolment	num	1. 0.02, 2. 4.62, 3. 9.41, …
5) IGRA_D0_status	IGRA status is positive if the max of IGRA score for tb1 and tb2 antigens is ≥ 0.35.	char	1. “IGRA+”, 2. “IGRA-“, 3. “INDETERMINATE”
6) LAM_D0	Urine Alere lipoarabinomannan (LAM) assay result at enrolment	char	1. “Positive”, 2. “Negative”
7) tbsymptoms_at_baseline	Any TB symptoms at baseline (at enrollment visit)	char	1. “Positive”, 2. “Negative”
8) endpoint	Indicator for sample being positive via the sputum Xpert MTB/RIF assay and/or MGIT assay	char	1. “MGIT+”, 2. “MGIT+/MGIT+”, 3. “MGIT+/Xpert+”, 4. “Xpert+”, 5. “Xpert+/Xpert+”
9) tbsymptoms_at_diagnosis	Any TB symptoms at diagnosis of prevalent or incident tuberculosis (at time of positive sputum sample collection)	char	1. “Positive”, 2. “Negative”
10) endpoint_osts	Binary indicator for secondary endpoint (≥ 1 sample positive)	int	1. 0, 2. 1
11) tevent_osts	Follow-up time for secondary endpoint analysis (days)	num	1. 0, 2. 84, 3. 449, …
12) endpoint_ts	Binary indicator for primary endpoint (≥ 2 sample positive)	int	3. 0, 4. 1
13) tevent_ts	Follow-up time for primary endpoint analysis (days)	num	4. 0, 5. 84, 6. 449, …
14) sex	Participant sex	char	1. “MALE”, 2. “FEMALE”
15) age	Participant age	num	1. 18, 2. 30.26, 3. 55
16) ethnicity	Participant ethnicity	char	1. “ASIAN”, 2. “BLACK”, 3. “CAUCASIAN”, 4. “MIXED RACE”,

Variable Name	Variable Description	Format	Example Values
			5. “OTHER”
17) bmi	Body-mass index at enrolment	num	1. 14.2, 2. 21.01, 3. 24.76
18) weight	Weight at enrolment	num	1. 92.3, 2. 68.1, 3. 44.8
19) smoking_history	History of smoking by participant	char	1. “YES”, 2. “NO”
20) tb_history	History of prior TB disease in participant	char	1. “YES”, 2. “NO”
21) tb_hhc	Tuberculosis household contact	char	1. “YES”, 2. “NO”
22) IPT	Isoniazid preventive therapy (IPT) duration at enrolment. “No IPT recorded” implies that the participant did not take IPT on study. "Started after enrolment" implies participant started IPT during the conduct of the study.	char	1. "<6 months", 2. ">12 months", 3. "6-12 months", 4. "No IPT recorded", 5. "Started after enrollment"
23) ART	Antiretroviral therapy (ART) duration at enrolment. “No ART recorded” implies that the participant did not take ART on study. "Started after enrolment" implies participant started ART during the conduct of the study.	char	1. "<6 months", 2. ">12 months", 3. "6-12 months", 4. "No ART recorded", 5. "Started after enrollment"
24) CD4	CD4-positive cell count at enrolment	num	1. 45, 2. 834, 3. 345, …
25) viral_load	HIV plasma viral load at enrolment	char	1. “<100”, 2. “100-999”, 3. “>=1000”