

Transcriptional Analysis for Tuberculosis in Pregnant Women From the PRegnancy Associated Changes In Tuberculosis Immunology (PRACHITi) Study

Jyoti S. Mathad,¹ Artur T. L. Queiroz,^{2,3} Ramesh Bhosale,^{4,5} Mallika Alexander,⁵ Shilpa Naik,^{4,5} Vandana Kulkarni,⁵ Bruno B. Andrade,^{2,6} and Amita Gupta^{5,7}

¹Center for Global Health, Weill Cornell Medicine, New York, New York, USA; ²Multinational Organization Network Sponsoring Translational and Epidemiological Research Initiative, Salvador, Brazil; ³Centro de Integração de Dados e Conhecimentos para Saúde, Instituto Gonçalo Moniz, Fundação Oswaldo Cruz, Salvador, Brazil; ⁴Byramjee Jeejeebhoy Government Medical College–Sassoon Government Hospital, Pune, India; ⁵Byramjee Jeejeebhoy Government Medical College–Johns Hopkins University Clinical Trials Unit, Pune, India; ⁶Laboratório de Inflamação e Biomarcadores, Instituto Gonçalo Moniz, Fundação Oswaldo Cruz, Salvador, Brazil; and ⁷Division of Infectious Diseases, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

A new tuberculosis (TB) diagnostic cartridge assay, which detects a 3-gene TB signature in whole blood, was not diagnostic in women with maternal TB disease in India (area under the curve [AUC] = 0.72). In a cohort of pregnant women, we identified a novel gene set for TB diagnosis (AUC = 0.97) and one for TB progression (AUC = 0.96).

Keywords. tuberculosis; pregnancy; RNA signature; transcriptomics; immunology.

The highest-risk time for a woman to develop tuberculosis (TB) disease is within 90 days postpartum [1], likely related to suppression of cell-mediated immunity during pregnancy followed by relative immune reconstitution immediately postpartum [2]. These changes can mask TB symptoms, causing underdiagnosis peripartum. A test that reliably diagnoses maternal TB would decrease TB-related complications for mother and child.

Transcriptional RNA signatures are blood-based tests that diagnose TB disease or predict progression from TB infection to disease [3]. Of 47 published transcriptional TB studies, none included pregnant women. Most TB signatures, including

the 3-gene signature developed into a cartridge-based diagnostic assay [4], identify upregulated inflammatory pathways [3, 4]. Because proinflammatory pathways are suppressed during pregnancy [2], TB signatures in nonpregnant populations may not be valid during pregnancy and postpartum.

We sought to identify differentially expressed genes (DEGs) in pregnant and postpartum women, before progressing and at TB diagnosis, to determine if published signatures remain valid and to identify differences in TB pathogenesis during pregnancy.

MATERIALS AND METHODS

We conducted a case-control study nested within a prospective observational cohort of pregnant women with and without human immunodeficiency virus (HIV) (PRACHITi) at Byramjee Jeejeebhoy Government Medical College (BJGMC)–Sassoon Hospital in Pune, India.

Study Population and Procedures

We included women ≥ 18 years of age with gestational age 13–34 weeks, and TB infection detected by Quantiferon TB Gold In-Tube assay (Qiagen). We excluded women with TB disease in the last 2 years, immunosuppression, or current use of antibiotics. Women were assessed for TB disease with a symptom screen at entry, third trimester, delivery, and postpartum with chest radiograph and GeneXpert, if indicated. All women had blood collected at each visit, and if TB was suspected, in PAXgene RNA tubes, stored at -80°C .

Of 234 women in PRACHITi, 10 developed TB. TB was defined as (1) sputum GeneXpert positive or (2) TB symptoms with radiographic evidence of TB disease and response to TB treatment. Seven cases had samples from entry (pre-TB) and time of TB diagnosis. For each case, 4 controls were identified who did not develop TB disease, matched on HIV status and gestational age at entry.

RNA was extracted using commercially available PAXgene Blood RNA kits (Becton, Dickinson and Company, Franklin Lakes, New Jersey) according to the manufacturer's instructions. The extracted RNA was sequenced at MedGenome in Bengaluru, India, on Illumina HiSeq4000 to generate 100 bp paired-end reads per sample.

Data Analysis

The raw RNA-seq data were retrieved in FastQ-formatted files. For all samples, low-quality bases were removed and adapters trimmed using Trimmomatic version 0.32. After quality check, sequences were aligned to the human transcriptome (GRCh38 version), comprising messenger RNA and noncoding RNA, with

Received 11 February 2022; editorial decision 25 May 2022; published online 10 June 2022

Correspondence: J. Mathad, Center for Global Health, Weill Cornell Medicine, 402 E 67th Street, 2nd floor, New York, NY 10065 (jsm9009@med.cornell.edu).

Clinical Infectious Diseases® 2022;75(12):2239–42

© The Author(s) 2022. Published by Oxford University Press on behalf of the Infectious Diseases Society of America.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com <https://doi.org/10.1093/cid/ciac437>

Salmon version 1.2.0. All downstream analyses were performed in R version 4.0.2 (R Foundation for Statistical Computing, Vienna, Austria). After mapping, the Salmon output was converted to count tables with the tximport package. The count gene expression matrix was examined by edgeR package to identify DEGs for (1) pregnant pre-TB vs pregnant control (prediction model) and (2) postpartum TB vs postpartum control (diagnostic model). Significant changes in gene expression were defined as statistical test values (false discovery rate-adjusted P value) $< .05$ and fold change higher than ± 1.4 . Candidate DEGs were visualized in volcano plots and Venn diagrams with the VennDiagram package. The obtained DEGs were scanned by REACTOME pathway databases using the compareCluster package. The count table was variance-stabilizing transformation (VST) normalized with the DESeq and edgeR packages. The data were evaluated to identify outlier samples with clustering analysis using the Stats package. A machine-learning based random forest algorithm with leave-one-out cross-validation was applied in the VST-normalized expression data to identify the minimal variable (gene) set, which exhibited higher classification power to describe each group with the randomForest package. Variables were sorted based on their model-classified importance and accuracy. Variables greater than the third quartile were used for analysis.

Previously published gene expression signatures were obtained from the TBSignatureProfile package (<https://github.com/comptbiomed/TBSignatureProfiler>) and tested against our datasets (pre-TB and TB diagnosis). We included BATF2 [3] and applied a general linear model to gene expression values from each signature gene. Outcomes were binarized to measure the sensitivity and specificity of classification, allowing measurement of each group rate and area under the curve (AUC) to identify the best classifier. The entire gene expression of our data set is available at the GEO database (accession number GSE168519; <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE168519>).

Ethical Considerations

All women provided written informed consent. The study protocol received ethical approvals from BJGMC in India and Johns Hopkins University and Weill Cornell Medicine in the United States.

RESULTS

Clinical Characteristics

Of the 7 cases and 28 controls, the median age was 25 years (interquartile range [IQR], 22–27 years). Two cases (28%) and 2 controls (7%) had a known TB exposure ($P = .11$). Four cases and 16 controls had HIV; all were on antiretroviral therapy with $>75\%$ virally undetectable. The median CD4 count at entry in cases vs controls was 428 cells/ μL and 402 cells/ μL , respectively ($P = .82$).

All 35 women had a positive interferon-gamma (IFN- γ) release assay (IGRA) at entry. All 7 TB cases occurred postpartum; the median time from delivery to diagnosis was 60 days (IQR, 30–150 days). The majority (71%) were diagnosed by GeneXpert. There were no other significant baseline differences between cases and controls, including IFN- γ from nil, mitogen, or TB antigen IGRA tubes (Supplementary Table 1).

Transcriptional and Classification Analysis

There were 424 differentially expressed genes in cases vs controls at TB diagnosis. The log counts per million (CPM) values for *LONRF1* and *ITSN1* correctly differentiated cases from controls. These genes were used as a model (MachineLearn) and evaluated by receiver operating characteristic (ROC) with an AUC of 0.97 (Figure 1A). Leave-one-out cross-validation had an accuracy of 0.91 (95% confidence interval [CI], .71–.99), a no-information rate of 0.684, a sensitivity of 0.867, and a specificity of 1.00.

There were also 469 differentially expressed genes in cases at the visit before developing TB disease (progressors) vs controls. The log CPM values for *GGT7* and *MYOM1* composed the best predictive gene set to differentiate progressors from controls with an AUC of 0.96 (Figure 1B). Leave-one-out cross-validation had an accuracy of 0.94 (95% CI, .80–.99), a no-information rate of 0.781, a sensitivity of 0.963, and a specificity of 0.86.

Comparison With Published Signatures

We compared the performance of 39 TB predictive and diagnostic signatures to our gene sets, and to the randomForest gene models in the conditions classification, using linear models to measure the area under the ROC curve (AUC) of each gene set and its CI. The gene sets we identified showed better predictive and diagnostic performance in pregnancy and postpartum than all published signatures, whose AUCs were 0.39–0.81 (Figure 1D and 1E). Specifically, the 3-gene signature had an AUC of 0.72 for the diagnostic model and 0.73 for the predictive model.

DISCUSSION

We prospectively studied pregnant women at high risk for active TB and identified gene sets unique from published signatures in nonpregnant populations, including the signature used in the new Cepheid cartridge. Our findings suggest differences in maternal TB pathogenesis and highlight that novel diagnostics may not benefit all populations.

We found that *ITSN1* and *LONRF1* differentiated TB disease from infection peripartum. *ITSN1*, like genes identified in published signatures, is associated with adaptive inflammatory responses, including CD4⁺ T-cell activation [5]. *ITSN1* and *LONRF1*, however, are also involved in innate immune

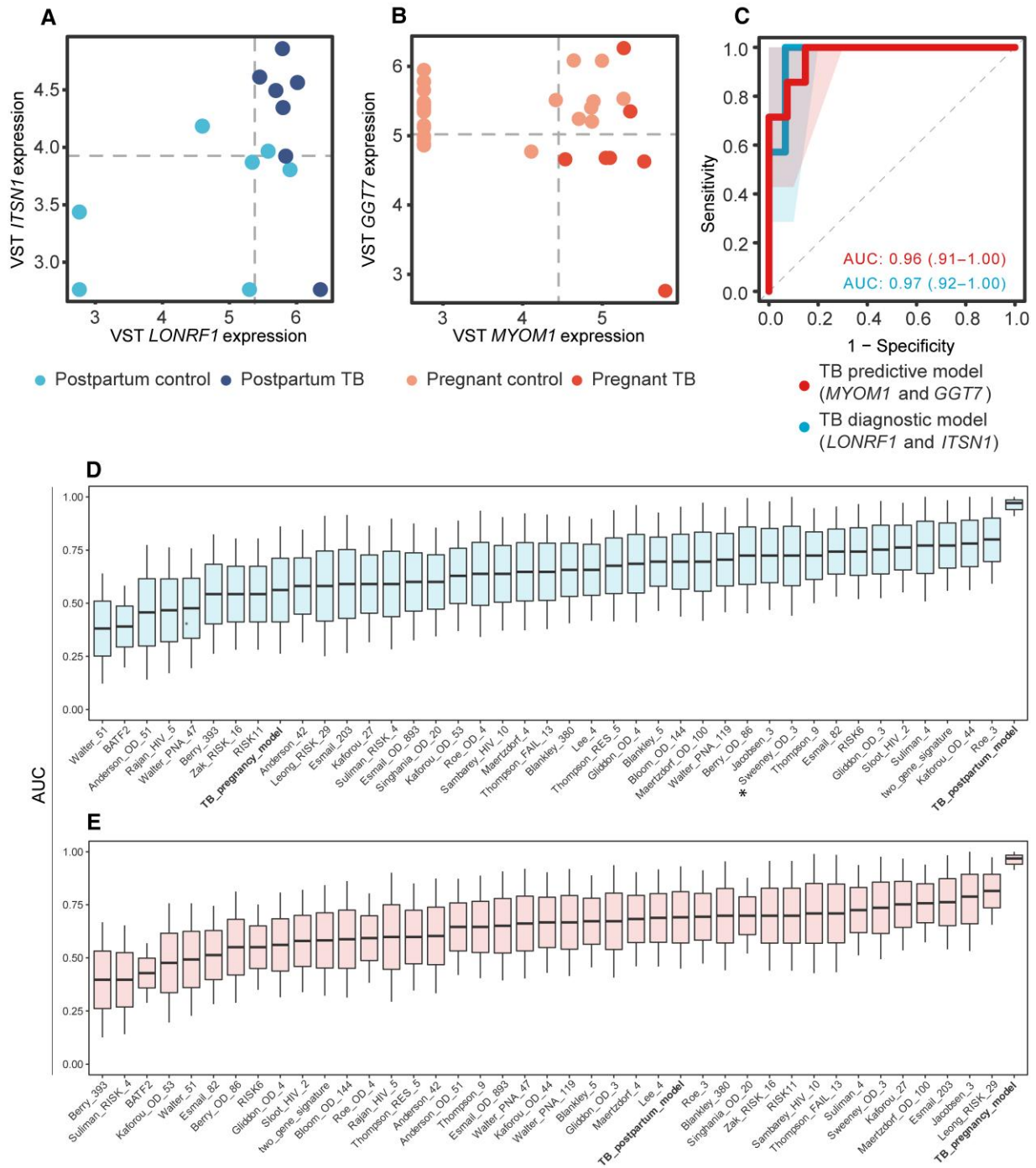


Figure 1. Biomarker identification analysis results and tuberculosis (TB) signature comparison. The dot plots from the biomarkers identified as best classifiers from cases at the time of active TB diagnosis vs controls (diagnostic model) (A) and cases *before* they developed active TB vs controls (predictive model) (B). C, Receiver operating characteristic (ROC) curve from each biomarker set. The TB predictive biomarkers are colored in red, the diagnostic model in blue. The shaded areas correspond to the standard error. The area under the curve (AUC) values for each curve are noted with the same colors. Boxplots show the AUC, measured by general linear modeling, for randomForest genes (bold), differentially expressed genes (bold), and publicly available TB gene expression signatures identifying the randomForest genes as the best TB classifier in postpartum (A) and pregnancy (B). We then compared the performance of diagnostic TB signatures (D), and predictive TB signatures (E) and the randomForest gene models in the conditions classification by using linear models to measure the area under the ROC curve (AUC) of each gene set and its confidence interval. The asterisked signature is being commercially used for rapid TB diagnosis. The randomForest gene sets we identified for pregnancy (predictive) and postpartum (diagnostic) outperforms all signatures in all comparisons. Abbreviation: VST, variance-stabilizing transformation.

pathways, including mediation of macrophage activation, proteasome degradation, and dendritic cell maturation [5, 6]. These genes, then, represent biologically plausible TB pathways slightly different from published signatures [3].

Participants with TB disease did not have significant differences in genes related to the IFN response or inflammasome pathways reported in the 3-gene signature and others [3, 7]. Known peripartum increases in IFN, documented by ourselves and others, may mask the expected increase in IFN signaling with TB disease [8, 9]. During pregnancy, inflammasome formation is also suppressed to prevent fetal rejection [10]. That neither IFN nor inflammasome pathways are upregulated in maternal TB may provide insight into maternal TB pathogenesis.

Our data suggest that the new 3-gene signature TB diagnostic assay may not improve diagnosis of maternal TB, which is especially challenging. Weight loss, a classic TB symptom, is expected postpartum, making symptom screening less specific. Consequently, postpartum women are often diagnosed with TB after their newborn. A test that accurately discriminates TB disease from other peripartum conditions would improve maternal and infant outcomes.

We found that *GGT7* levels helped predict TB progression in pregnant women before they developed symptomatic disease. *GGT7* is involved in glutathione metabolism, which is important in the TB immune response and has been associated with TB disease [11]. This gene has not been identified in several TB progression studies [3, 12], suggesting that glutathione pathways may be more important in maternal TB pathogenesis.

Our longitudinal approach allowed assessment of presymptomatic and symptomatic TB. Our study was limited by a smaller sample size and lack of a validation cohort, which is planned. The small sample size, however, is a common challenge in maternal TB studies. Including pregnant women in ongoing TB research, especially low-risk observational studies, would address this. By excluding pregnant women due to immunologic and physiologic changes, researchers have inadvertently prevented advancement of the diagnosis, prevention, and treatment of maternal TB.

Supplementary Data

Supplementary materials are available at *Clinical Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Notes

Disclaimer. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or the other listed foundations.

Financial support. This work was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (grant number R01HD08192) and the National Institute of Allergy and Infectious Diseases (NIAID) of the National Institutes of Health (NIH) (award numbers K23AI129854 to J. S. M. and UM1AI069465 to A. G.).

Potential conflicts of interest. A. G. receives support from the Gilead Foundation, the Makhija Foundation, and the Wyncote Foundation; reports grants or contracts unrelated to this work from NIH, Unitaid, and the Centers for Disease Control and Prevention; reports participation on a data and safety monitoring board or advisory board for the NIH/NIAID Advisory Council and Indo US Science Technology Governing Board; and reports a leadership or fiduciary role with the International Maternal Pediatric and Adolescent AIDS Clinical Trial Network TB Scientific Committee and World Health Organization Multidrug-Resistant Tuberculosis Guidelines Committee. J. S. M. reports grants or contracts unrelated to this work and support for attending meetings and/or travel from the NIH, paid to institution, and the Aurum Institute. All other authors report no potential conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Zenner D, Kruijshaar ME, Andrews N, et al. Risk of tuberculosis in pregnancy: a national, primary care-based cohort and self-controlled case series study. *Am J Respir Crit Care Med* **2012**; 185:779–84.
2. Kourtis AP, Read JS, Jamieson DJ. Pregnancy and infection. *N Engl J Med* **2014**; 370:2211–8.
3. 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**; 8:395–406.
4. Zimmer AJ, Schumacher SG, Södersten E, et al. A novel blood-based assay for treatment monitoring of tuberculosis. *BMC Res Notes* **2021**; 14:247.
5. Pasquinelli V, Rovetta AI, Alvarez IB, et al. Phosphorylation of mitogen-activated protein kinases contributes to interferon γ production in response to *Mycobacterium tuberculosis*. *J Infect Dis* **2013**; 207:340–50.
6. Walker LR, Jobman EE, Sutton KM, et al. Genome-wide association analysis for porcine reproductive and respiratory syndrome virus susceptibility traits in two genetic populations of pigs1. *J Anim Sci* **2019**; 97:3253–61.
7. Sweeney TE, Braviak L, Tato CM, et al. Genome-wide expression for diagnosis of pulmonary tuberculosis: a multicohort analysis. *Lancet Respir Med* **2016**; 4: 213–24.
8. Mathad JS, Bhosale R, Sangar V, et al. Pregnancy differentially impacts performance of latent tuberculosis diagnostics in a high-burden setting. *PLoS One* **2014**; 9:e92308.
9. Birku M, Desalegn G, Kassa G, et al. Effect of pregnancy and HIV infection on detection of latent TB infection by tuberculin skin test and QuantiFERON-TB Gold In-Tube assay among women living in a high TB and HIV burden setting. *Int J Infect Dis* **2020**; 101:235–42.
10. Shirasuna K, Karasawa T, Takahashi M. Role of the NLRP3 inflammasome in pre-eclampsia. *Front Endocrinol (Lausanne)* **2020**; 11:80.
11. Aiyaz M, Bipin C, Pantulwar V, et al. Whole genome response in guinea pigs infected with the high virulence strain *Mycobacterium tuberculosis* TT372. *Tuberculosis* **2014**; 94:606–15.
12. Penn-Nicholson A, Mbandi SK, Thompson E, et al. RISK6, A 6-gene transcriptomic signature of TB disease risk, diagnosis and treatment response. *Sci Rep* **2020**; 10:8629.