Transcriptional analysis for tuberculosis in pregnant women from the 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], Amita Gupta[5,7]

[1]Weill Cornell Medicine, Center for Global Health, New York, United States of America
[2]Multinational Organization Network Sponsoring Translational and Epidemiological Research (MONSTER) Initiative, Salvador 41810-710, Brazil
[3]Centro de Integração de Dados e Conhecimentos para Saúde (CIDACS), Instituto Gonçalo Moniz, Fundação Oswaldo Cruz, Salvador 40296-710, Brazil
[4]Byramjee Jeejeebhoy Government Medical College/Sassoon Government Hospital, Pune, India
[5] Byramjee Jeejeebhoy Government Medical College-Johns Hopkins University Clinical Trials Unit, Pune, India
[6]Laboratório de Inflamação e Biomarcadores, Instituto Gonçalo Moniz, Fundação Oswaldo Cruz, Salvador 40296-710, Brazil
[7] Johns Hopkins University School of Medicine, Baltimore, United States of America

***Corresponding Author:**
Jyoti Mathad jsm9009@med.cornell.edu

Abstract

A new tuberculosis diagnostic cartridge assay, which detects a 3-gene tuberculosis signature in whole blood, was not diagnostic in women with maternal tuberculosis disease in India (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).

Key words: **tuberculosis**, **pregnancy**, **RNA signature**, **transcriptomics, immunology**

The highest risk time for a woman to develop tuberculosis disease (TB) 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 (TBI) 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 pro-inflammatory pathways are suppressed during pregnancy [2] TB signatures in non-pregnant 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.

Methods

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

Study population and procedures. We included women ≥18 years with gestational age 13-34 weeks, and TBI detected by QuantiFERON TB Gold In-tube assay (QGIT, Qiagen). We excluded women with TB disease in the last two years, immunosuppression, or current use of antibiotics. Women were assessed for TB disease with a symptom screen at entry, 3rd 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 (postpartum). For each case, four 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, NJ, USA) according to manufacturer instructions. The extracted RNA was sequenced at MedGenome in Bengaluru, India on Illumina HiSeq4000 to generate 100bp 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 V0.32. After quality check, sequences were aligned to the human transcriptome (GRCh38 version), comprising mRNA and ncRNA, with Salmon v1.2.0. All downstream analyses were performed in R v4.0.2 (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 versus pregnant control (TB prediction) and (2) postpartum TB versus postpartum control (TB diagnosis). Significant changes in gene expression were defined as statistical test values (FDR adjusted p-value) lower than 0.05 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 package. 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 > 3rd quartile were used for analysis.

Previously published gene expression signatures were obtained from the TBSignatureProfile package (https://github.com/compbiomed/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).

Oversight/Safety. All women provided written informed consent. The study protocol received ethical approvals from BJGMC in India, 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 (IQR 22-27). Two cases (28%) and two controls (7%) had a known TB exposure (p=0.11). Four cases and 16 controls had HIV; all were on antiretroviral therapy with >75% virally undetectable. The median CD4 at entry in cases versus controls was 428 and 402 cells/mm^3, respectively (p=0.82).

All 35 women had a positive IGRA at entry. Among the 7 cases, the median time of TB diagnosis was 60 days postpartum (IQR 30-150). 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 (**Supplemental Table 1).**

Transcriptional and classification analysis. There were 424 differentially expressed genes in cases versus controls at TB diagnosis. The log 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% CI 0.71-0.99), 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) versus 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 0.80-0.99), 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 Random Forest gene models in the conditions classification, using linear models to measure the area under the ROC curve (AUC) of each gene set and its confidence interval. 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&E)**. Specifically, the 3-gene signature had an AUC of 0.72 for the diagnostic 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 non-pregnant 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 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 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 interferon response or inflammasome pathways reported in the 3-gene signature and others [3, 7]. Known peripartum increases in interferon, documented by ourselves and others, may mask the expected increase in interferon 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 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.

Funding

This work was supported by the Eunice Shriver National Institutes of Child Health and Development [R01HD08192] and the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award number K23AI129854(JSM) and UM1AI069465 (AG). AG receives support from the Gilead Foundation, Makhija Foundation, and the Wyncote Foundation and reports grants or contracts unrelated to this work from NIH, UNITAID, and CDC; participation on a Data Safety Monitoring Board or Advisory Board for NIH/NIAID Advisory Council and Indo US Science Technology Governing Board; and a leadership or fiduciary role with IMPAACT Network TB Scientific Committee and WHO MDR TB Guidelines Committee. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the other listed foundations. JSM reports grants or contracts unrelated to this work and support for attending meetings and/or travel from National Institutes of Health and paid to institution.

The authors have no conflicts of interest.

References

1. Zenner D, Kruijshaar ME, Andrews N, Abubakar I. 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(7): 779-84.
2. Kourtis AP, Read JS, Jamieson DJ. Pregnancy and infection. N Engl J Med **2014**; 370(23): 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(4): 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(1): 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(2): 340-50.
6. Walker LR, Jobman EE, Sutton KM, Wittler J, Johnson RK, Ciobanu DC. Genome-wide association analysis for porcine reproductive and respiratory syndrome virus susceptibility traits in two genetic populations of pigs1. J Anim Sci **2019**; 97(8): 3253-61.
7. 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.
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(3): e92308.
9. Birku M, Desalegn G, Kassa G, Tsegaye A, Abebe M. 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 Preeclampsia. 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 (Edinb) **2014**; 94(6): 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(1): 8629.

Figure Legend:

Figure: Biomarkers identification analysis results and TB signature comparison. The dot plots from the biomarkers identified as best classifiers from (B) cases versus controls (diagnostic model) (A) cases before they developed active TB (progressors) versus controls (predictive model). The (C) Receiver Operating Characteristic (ROC) curve from each biomarker set. The TB predictive biomarkers are colored as red, the diagnostic model in blue. The shaded areas correspond to standard error. The Area Under the Curve (AUC) values for each curve are colored 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 (blue, panel D), and predictive TB signatures (pink, panel E) and the Random Forest 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.

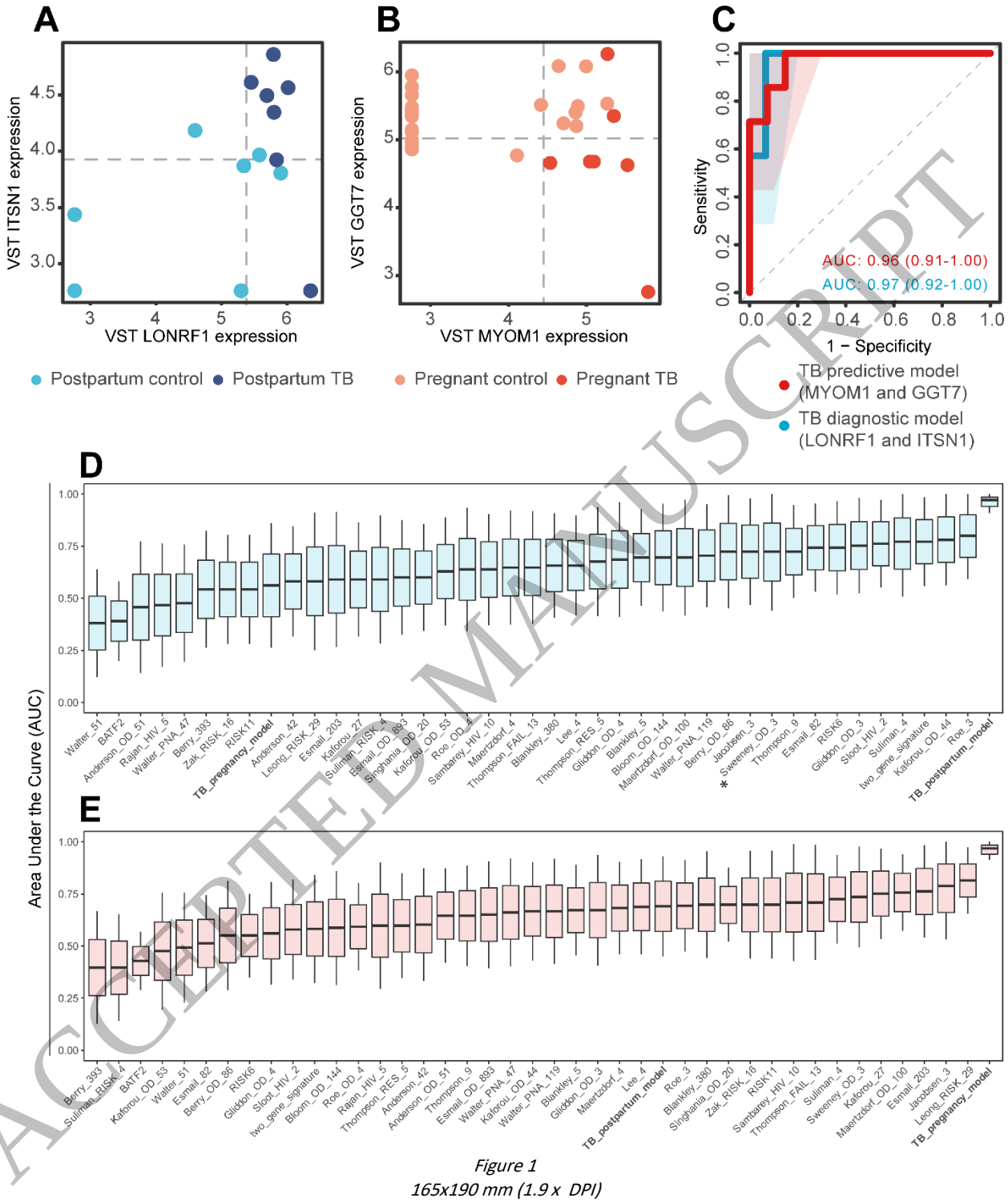

Figure 1
165x190 mm (1.9 x DPI)