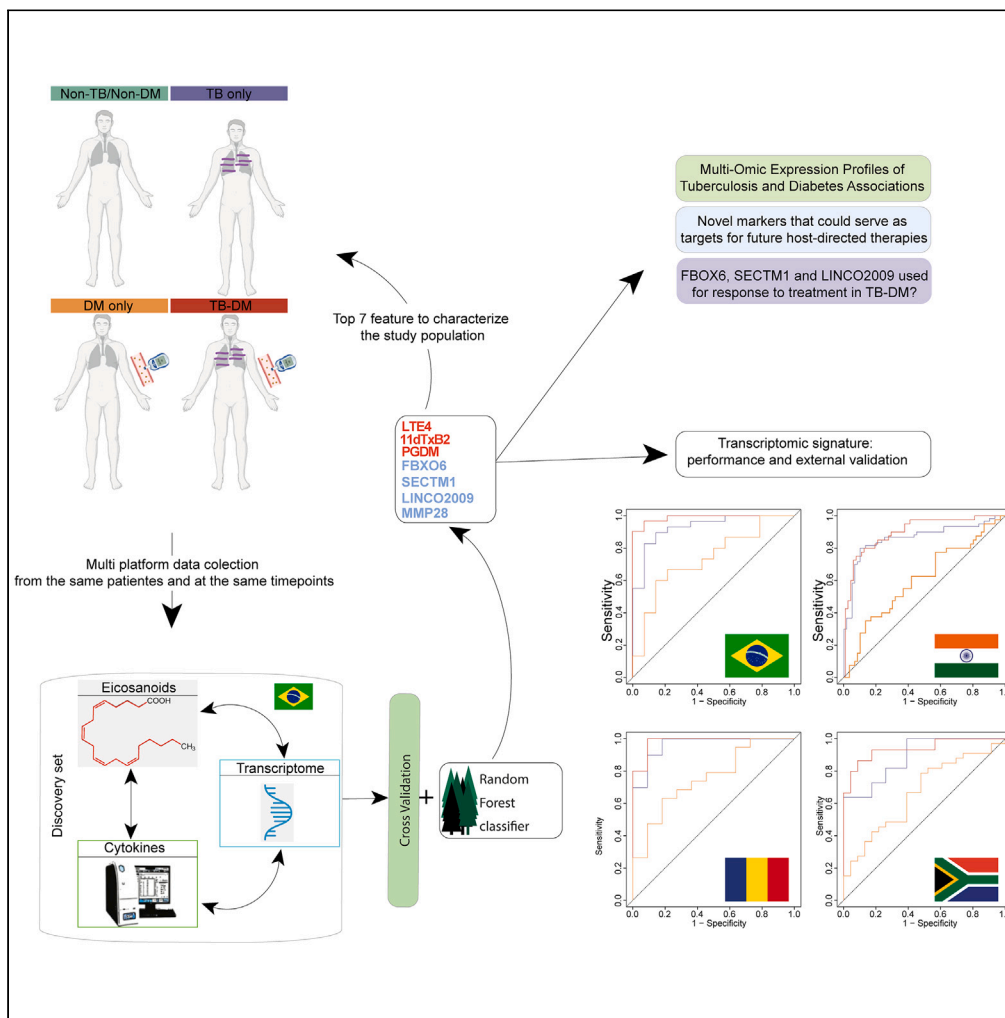


Article

# An integrative multi-omics approach to characterize interactions between tuberculosis and diabetes mellitus



Caian L. Vinhaes, Eduardo R. Fukutani, Gabriel C. Santana, ..., Bruno B. Andrade, Artur T.L. Queiroz, for the RePORT Brazil Consortium

bruno.andrade@fiocruz.br

Highlights

A distinct multi-omic signature characterizes TB regardless of DM status

Anti-tubercular therapy leads to decreases in expression of the multi-omic signature

The biomarkers identified may serve as candidates for host-directed therapies

Vinhaes et al., iScience 27, 109135  
March 15, 2024 © 2024 The Author(s).  
<https://doi.org/10.1016/j.isci.2024.109135>



## Article

## An integrative multi-omics approach to characterize interactions between tuberculosis and diabetes mellitus

Caian L. Vinhaes,<sup>1,2,3,4,5,16</sup> Eduardo R. Fukutani,<sup>2,5,6,16</sup> Gabriel C. Santana,<sup>1,2,7</sup> María B. Arriaga,<sup>8</sup> Beatriz Barreto-Duarte,<sup>1,2,5,7,9</sup> Mariana Araújo-Pereira,<sup>1,2,5,10</sup> Mateus Maggitti-Bezerril,<sup>2,5</sup> Alice M.S. Andrade,<sup>1,2,5</sup> Marina C. Figueiredo,<sup>8</sup> Ginger L. Milne,<sup>11</sup> Valeria C. Rolla,<sup>12</sup> Afrânio L. Kristki,<sup>9</sup> Marcelo Cordeiro-Santos,<sup>13,14,15</sup> Timothy R. Sterling,<sup>8</sup> Bruno B. Andrade,<sup>1,2,3,5,7,8,10,17,18,\*</sup> Artur T.L. Queiroz,<sup>2,5,6,17</sup> and for the RePORT Brazil Consortium

## SUMMARY

**Tuberculosis-diabetes mellitus (TB-DM) is linked to a distinct inflammatory profile, which can be assessed using multi-omics analyses. Here, a machine learning algorithm was applied to multi-platform data, including cytokines and gene expression in peripheral blood and eicosanoids in urine, in a Brazilian multi-center TB cohort. There were four clinical groups: TB-DM (n = 24), TB only (n = 28), DM (HbA1c  $\geq$  6.5%) only (n = 11), and a control group of close TB contacts who did not have TB or DM (n = 13). After cross-validation, baseline expression or abundance of MMP-28, LTE-4, 11-dTxB2, PGDM, FBXO6, SECTM1, and LINCO2009 differentiated the four patient groups. A distinct multi-omic-derived, dimensionally reduced, signature was associated with TB, regardless of glycemic status. SECTM1 and FBXO6 mRNA levels were positively correlated with sputum acid-fast bacilli grade in TB-DM. Values of the biomarkers decreased during the course of anti-TB therapy. Our study identified several markers associated with the pathophysiology of TB-DM that could be evaluated in future mechanistic investigations.**

## INTRODUCTION

Tuberculosis (TB) affects approximately 10 million people annually, and despite being a preventable and curable disease, 1.6 million people died from TB in 2021.<sup>1</sup> About one-quarter of the world's population is estimated to be infected with *Mycobacterium tuberculosis* (Mtb).<sup>2</sup>

After infecting persons via the lungs, Mtb can disseminate and affect several other organ systems, resulting in a large potential range of clinical manifestations and varying degrees of severity depending on the interaction between Mtb and host immune responses. Factors that affect immune responses can influence the inflammatory reactions against Mtb; these include infection with HIV,<sup>3</sup> and genetic,<sup>4</sup> environmental,<sup>5</sup> nutritional,<sup>6</sup> and metabolic features such as dysglycemia and diabetes mellitus (DM).

DM is one of the most common metabolic diseases worldwide. Currently, 422 million people are living with DM,<sup>7</sup> most of whom reside in low and middle-income countries. Approximately 1.5 million deaths have been attributed to DM annually.<sup>7</sup> Sustained and uncontrolled hyperglycemia leads to changes in immune regulation<sup>8</sup>; therefore, people living with DM are considered immunosuppressed hosts.<sup>9</sup> The

<sup>1</sup>Laboratório de Pesquisa Clínica e Translacional, Instituto Gonçalo Moniz, Fundação Oswaldo Cruz, Salvador 40296-710, Brazil

<sup>2</sup>Multinational Organization Network Sponsoring Translational and Epidemiological Research (MONSTER) Initiative, Salvador 41810-710, Brazil

<sup>3</sup>Programa de Pós-Graduação em Medicina e Saúde Humana, Escola Bahiana de Medicina e Saúde Pública (EBMSP), Salvador 40290-150, Brazil

<sup>4</sup>Departamento de Infectologia, Hospital Português da Bahia, Salvador 40140-901, Brazil

<sup>5</sup>Instituto de Pesquisa Clínica e Translacional, Faculdade de Tecnologia e Ciências, Salvador 41741-590, Brazil

<sup>6</sup>Centro de Integração de Dados e Conhecimentos para Saúde, Instituto Gonçalo Moniz, Fundação Oswaldo Cruz, Salvador, Brazil

<sup>7</sup>Curso de Medicina, Universidade Salvador, Salvador, Brazil

<sup>8</sup>Division of Infectious Diseases, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA

<sup>9</sup>Programa Acadêmico de Tuberculose, Faculdade de Medicina, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

<sup>10</sup>Faculdade de Medicina, Universidade Federal da Bahia, Salvador, Brazil

<sup>11</sup>Division of Clinical Pharmacology, Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN, USA

<sup>12</sup>Instituto Nacional de Infectologia Evandro Chagas, Fiocruz, Rio de Janeiro, Brazil

<sup>13</sup>Fundação Medicina Tropical Doutor Heitor Vieira Dourado, Manaus, Brazil

<sup>14</sup>Programa de Pós-Graduação em Medicina Tropical, Universidade do Estado do Amazonas, Manaus, Brazil

<sup>15</sup>Universidade Nilton Lins, Manaus, Brazil

<sup>16</sup>These authors contributed equally

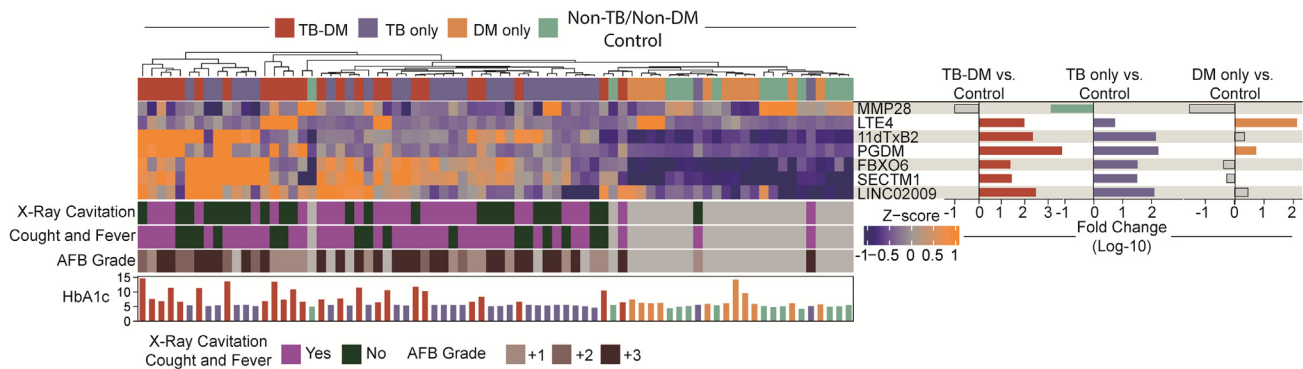
<sup>17</sup>These authors contributed equally

<sup>18</sup>Lead contact

\*Correspondence: [bruno.andrade@fiocruz.br](mailto:bruno.andrade@fiocruz.br)

<https://doi.org/10.1016/j.isci.2024.109135>





**Figure 1. Distinct multi-omic expression profiles identified tuberculosis regardless the glycemic status**

*Right panel.* A hierarchical cluster analysis (Ward method with 100 × bootstrap) was employed to test the overall expression of plasma cytokines, gene expression and urinary eicosanoids in the study population. Dendrograms represent Canberra distance. *Left panel.* Differential expression analysis was used to calculate the fold-changes and show differences in biomarkers levels for each clinical group (TB, TB-DM, and DM) versus control. Differences that reached statistical significance after adjustment for multiple comparisons (adjusted  $p < 0.05$ ) are represented in colored bars.

association between TB and DM has important health consequences. Previous studies have revealed that DM increases the risk of incident TB 3-fold and increases the risk of TB unfavorable outcomes 2-fold.<sup>10,11</sup>

Our group has studied TB-DM interactions, including activation of transcriptional pathways in disease complications,<sup>12</sup> persistent inflammation during anti-TB treatment (ATT),<sup>13</sup> increased Mtb transmission to close contacts,<sup>14</sup> more severe TB clinical presentation,<sup>15</sup> and increased risk of unfavorable ATT outcomes.<sup>16–18</sup> However, while most of the studies used different assays, which were analyzed separately to characterize TB-DM, the use of multiple omics platforms remains poorly explored in the context of TB-DM interactions. Multi-omics provides the opportunity to gain insights into disease pathogenesis, given that it simultaneously explores several components of immune responses through multiple assay platforms. This is a powerful tool to generate hypotheses pertaining to the intricate molecular relationships that may be driving disease. To our knowledge, no previous investigation has examined whether multi-omics analyses can help characterize TB-DM pathogenesis. To fill this knowledge gap, we leveraged samples and data collected in the multi-center prospective cohort of the Regional Prospective Observational Research in Tuberculosis (RePORT)-Brazil ([www.reportbrazil.org](http://www.reportbrazil.org)). We used transcriptomic and cytokine data from peripheral blood and eicosanoids measured in urine, collected at the same time points from persons with TB-DM, TB only, DM only, and non-TB/non-DM close contacts (control group). We applied a random forest model to investigate a possible multi-omic signature that could characterize TB-DM. Furthermore, we tracked the expression of the components of the multi-omic signature in the study groups during the course of ATT. Our results identified novel pathways that may drive disease pathogenesis and serve as targets for future host-directed therapies for TB-DM and as markers of progression to cure after antitubercular therapy initiation. Additionally, the transcriptomic markers identified here have shown a robust accuracy to detect TB-DM in an external validation study using cohorts from India, South Africa, and Romania.

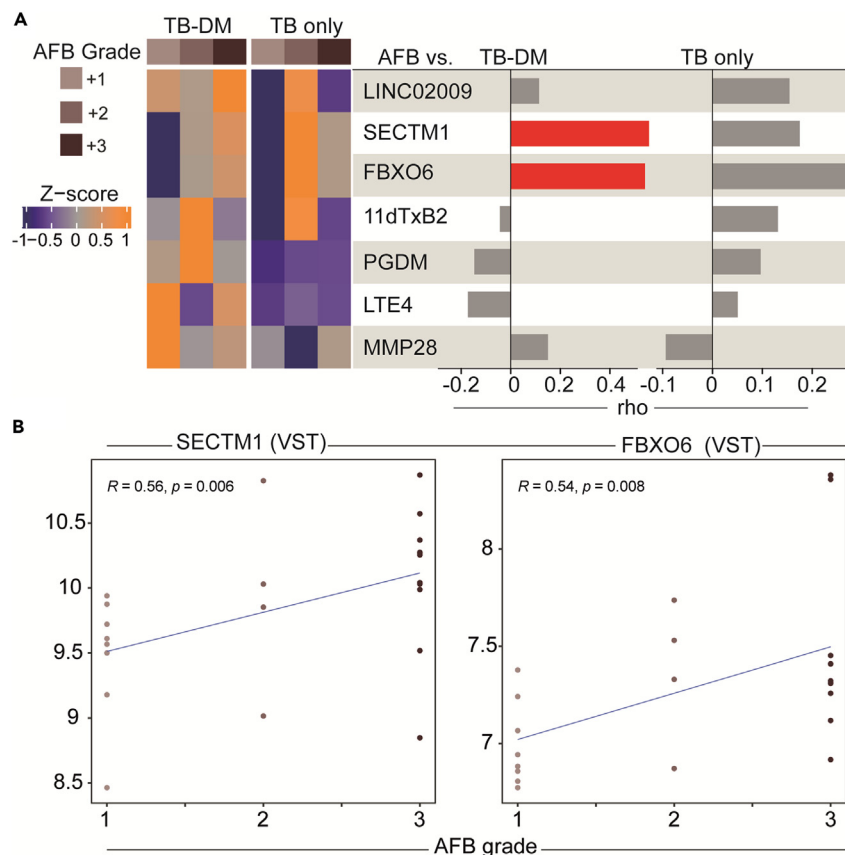
## RESULTS

### Characteristics of study population

All patients included in our analysis were classified according to their status regarding TB disease and DM. The groups were composed of 24 with TB-DM, 28 TB only, 11 DM only, and 13 non-TB/non-DM (controls). TB-DM and DM only participants were older, with a median age of 46.5 (37–55.7) and 56 (IQR:51–59) years, respectively ( $p < 0.001$ ; [Table S1](#)). In addition, controls and participants with DM were more frequently female than those with TB-DM and TB only. Higher BMI was found in the controls and in DM only groups, with a median (IQR) of 29 (24–33) and 29 (27–32), respectively, compared to 20 (18–22) and 22 (18–26) in TB only and TB-DM patients ( $p < 0.001$ ) ([Table S1](#)). The levels of HbA1c were higher in those with TB-DM (median [IQR] 8.1 [6.6–11.3]), followed by DM only 7.1 [6.9–8.4] ( $p < 0.001$ ). Of note, TB only patients were not significantly different regarding AFB (acid-fast bacilli) smear grade and the proportion with cavities on chest X-ray ([Table S1](#)).

### Multi-omic expression profiles of tuberculosis and diabetes associations

To address the aim of finding a distinct profile that distinguishes TB regardless of DM, our first step was to study the multi-omic alterations in TB and DM using a well-established dimensionality reduction approach ([Figure S1](#)). This approach selected 7 parameters that could most contribute to discrimination between all the study groups ([Figure S1](#)). The Gini score and mean accuracy for each marker used in our model to discriminate between the clinical groups were provided in [Table S2](#) and [Figure S2](#). To further evaluate the baseline multi-omic profiles of the clinical groups, we first examined the expression of the seven selected features ([Figure 1](#)), as described in the [STAR methods](#) section. Using a heatmap with log-10 transformed values to assess the overall sample expression, we discovered that TB patients (TB-DM and TB only) had a distinct expression profile compared to non-TB groups, including controls and DM only, with a tendency toward higher expression



**Figure 2. A distinct multi-omic expression and correlation profile between AFB grade and markers among TB groups**

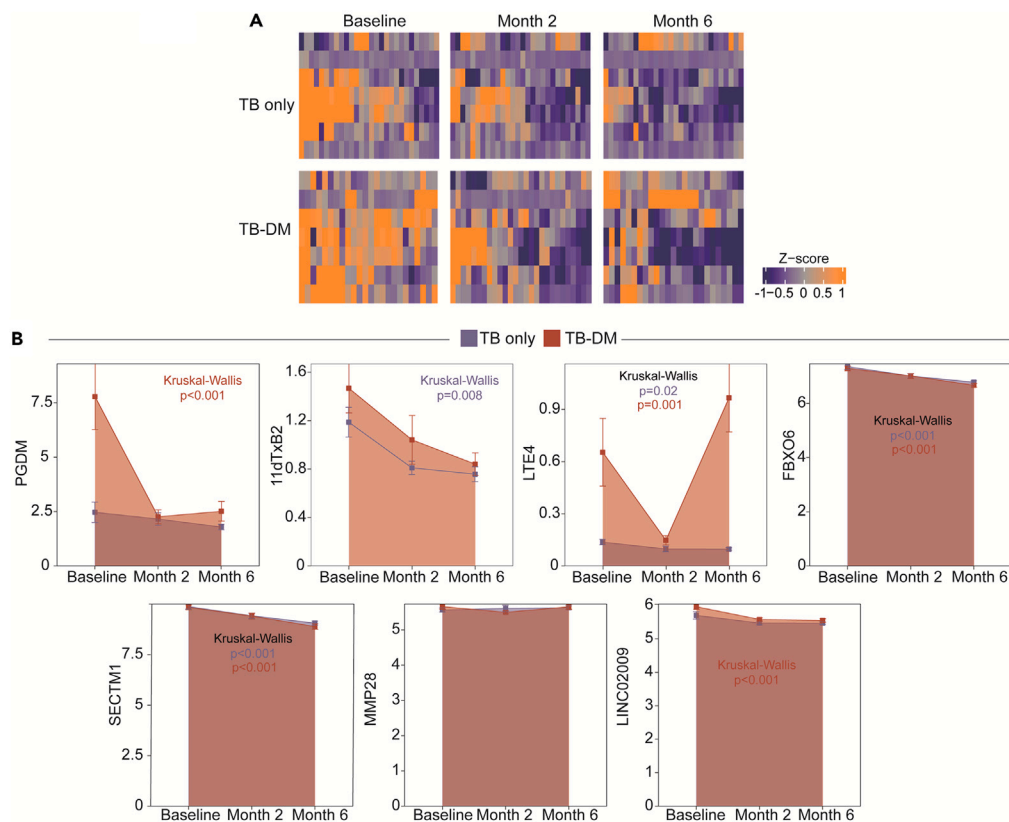
(A) *Left panel.* A hierarchical cluster analysis (Ward method with 100 × bootstrap) was employed to evaluate multi-omic marker expression according to the AFB smear grade in TB and TB-DM, as indicated. *Right panel.* A Spearman correlation analysis was used to study the influence of the AFB smear grade on multi-omic marker expression. The rho values are shown. Red bars indicate correlation with p value < 0.05.

(B) Spearman correlation plots demonstrating the associations between the expression levels of the indicated genes and the AFB smear grade. Line represent linear regression. R: Spearman rho value.

of 11dTxB2, PGDM, FBXO6, SECTM1, and LINC02009 in TB patients (Figure 1, left panel; Table S3). Of note, the heatmap demonstrated that the impact of TB disease on the biosignature expression values was able to cluster TB-disease participants independent of DM status (Figure 1, left panel; Table S2). The fold-difference of the multi-omic markers between the clinical groups versus control group is summarized in Figure 1, right panel. When the TB only group was compared with controls, the profile was similar to that observed between TB-DM vs. control group, with higher expressions observed in the majority of the markers in both TB groups, except for MMP28, which was higher in the control group when compared with the TB only group. (Figure 1, right panel). Extending the approach to DM only and control differences, we found higher levels of LTE4 (leukotriene E4) and PGDM (prostaglandin D metabolite) in DM only participants (Figure 1, right panel). These findings reveal a higher multi-omic expression of biomarkers capable of clustering TB patients, regardless of DM status, which suggests differential expression triggered by TB disease.

### Variation in the multi-omic activation according to AFB grade and clinical group

After identifying a multi-omic profile that characterized the TB only and TB-DM groups, our next step was to evaluate whether glycemic status affected the expression of markers according to AFB smear grade (Figure 2). Using hierarchical clustering on the Z score normalized data at baseline, we observed that patients with TB (with and without DM) and AFB smear grade equal to 2+ tended to express higher levels of SECTM1, FBXO6, and 11dTxB2 compared with other AFB grades (Figure 2A). Those marker levels were lower among those with TB with AFB grade 1+ (Figure 2A). On the other hand, the TB-DM group showed higher expression of LTE4 and MMP28 in those with AFB grade 1+, PGDM and 11dTxB2 in AFB grade 2+, and finally FBXO6, SECTM1, and LINC02009 in AFB grade 3+ (Figure 2A). Next, using Spearman correlation analysis we found that the AFB smear grade was correlated with SECTM1 and FBXO6 expression in TB-DM participants. No significant correlation was observed in the TB only group (Figures 2A and 2B). No significant correlation was observed in the TB only group (Figure 2A). This finding suggests that the presence of TB was associated with changes in multi-omic expression according to AFB smear grade.



**Figure 3. Changes in multi-omic expression after anti-tuberculosis therapy initiation**

(A) A hierarchical cluster analysis (Ward method with 100 × bootstrap) was employed to evaluate multi-omic marker expression in TB and TB-DM after anti-tuberculosis therapy initiation.

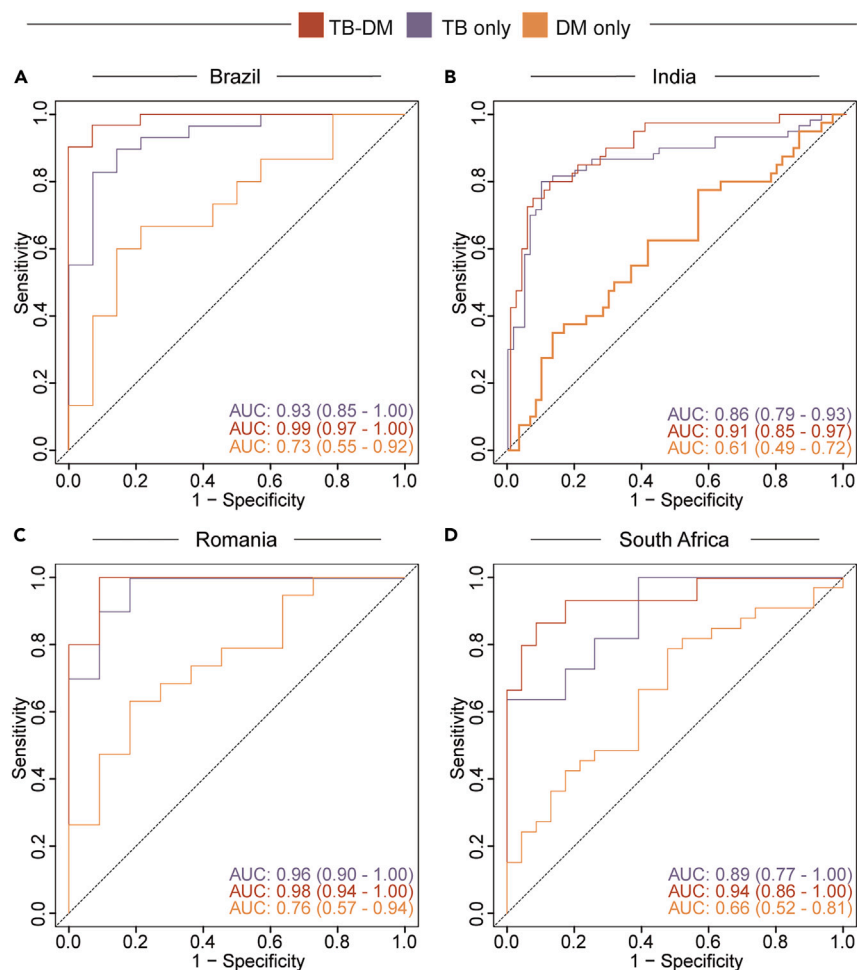
(B) A boxplot was used to test the changes in multi-omic levels in months 2 and 6 after ATT. In the graphs, dots represent median and whiskers represent interquartile range values.

### Changes in multi-omic expression after anti-tuberculosis therapy initiation

We next evaluated changes in the biomarker expressions or concentrations after ATT initiation (Figure 3). Using a heatmap approach, we were able to show an overall decreasing tendency in marker values, in both the TB and TB-DM groups (Figure 3A). Next, we prospectively assessed marker expressions to evaluate the dynamics of multi-omic expression during ATT (Figure 3B). We discerned a reduced level of 11dTxB2 in the TB group, as well as a reduced level of FBXO6 and SECTM1 expression in both the TB and TB-DM groups, and reduced LINC02009 expression in the TB-DM group.

### The multi-omic approach resulted in a transcriptomic signature to detect TB-DM

After evaluating the multi-omic profile of TB and TB-DM interaction, we extended our analysis to verify the expression of the transcriptomic markers identified here. Using data from a multi-center, prospective cohort study of our group,<sup>19</sup> which assessed transcriptomic data from 290 participants, recruited from two sites in India, one in Brazil and publicly available RNA-seq data from two sites of a TANDEM consortium (Romania and South Africa),<sup>20</sup> we test the accuracy of the transcriptomic markers identified here by our machine learning analysis (Figures 4 and S3). First, we used a heatmap approach applied to the four countries, verifying the expression of MMP28, SECTM1, FBXO6, and LINC02009 (Figure S3). Our results identified that these genes were capable of detecting TB regardless of the glycemic status in Brazil, India, and Romania (Figure S3), marked by higher expression of the markers in TB-DM and TB only (Figure S3). Additionally, we used a fold-change Log2 to study the expression of the markers in each clinical group and in each country (Figure S3). Of interest, our results revealed a similar expression profile in TB-DM and TB only in Brazil and India, with higher expression of SECTM1, FBXO6, and LINC02009 when compared with non-TB/non-DM controls, whereas MMP28 was higher in the controls (Figure S3). Next, we used the receiver operating characteristic (ROC) curve, employing the genes MMP28, SECTM1, FBXO6, and LINC02009 to verify the accuracy of these markers, together, to identify TB only, TB-DM, and DM only (Figure 4), shown a robust accuracy in detecting TB-DM, area under curve (AUC) 0.99, 0.91, 0.98, and 0.94 in Brazil, India, Romania, and South Africa, respectively (Figure 4). Our result identifies a transcriptomic signature, composed by expression levels of MMP28, SECTM1, FBXO6, and LINC02009, and identified using machine learning analysis applied to multi-platform data, with a great accuracy to detect TB-DM.



**Figure 4. Accuracy of a new transcriptomic signature to detect TB-DM**

Accuracy of the gene signature detected in the random forest analysis to classify TB, with or without DM in Brazil (A), India (B), Romania (C) and South Africa (D). Receiver operator characteristics (ROC) curve analysis was used to check the accuracy of the signature genes identified by the random forest model to classify the TB, TB-DM, and DM groups in each clinical site as indicated with respect to TB disease.

## DISCUSSION

To achieve the World Health Organization targets for TB elimination, new strategies in the TB field are required. The study of omics has been extensively applied to provide insights into the pathophysiological changes induced by *Mtb* infection. Several mRNA transcriptomic signatures have been proposed in recent years as new tools for the diagnosis, prognosis, and treatment monitoring of *Mtb* infection and TB disease.<sup>21–24</sup> Nevertheless, advances have been made using isolated omics, and the employment of an integrative analysis could better dissect the changes triggered by *Mtb* infection. Here, we performed an integrated omics analysis, using transcriptomic data; eicosanoid measurements in urine; and soluble inflammatory biomarkers in blood, to identify a multi-omic profile that characterized TB and DM.

Our first step was to perform dimensionality data reduction. A random forest model was applied to the integrated multi-omic data (Luminex and RNAseq from peripheral blood and eicosanoids from urine) to select the most informative features. The random forest algorithm works based on summarizing multiple decision trees, evaluating the importance of markers in classifying the samples within the data. The variables were evaluated by using the mean decrease of accuracy and Gini indexes. This algorithm has been used by our group in TB and HIV coinfection studies.<sup>25,26</sup> This approach was also applied to the conceptualization of predictive models in cardiovascular diseases<sup>27</sup> and to predict outcomes in neurosurgery.<sup>28</sup> After cross-validation, MMP-28, LTE-4, 11- dTxB2, PGDM, FBXO6, SECTM1, and LINCO2009 were selected. Importantly, despite the crucial role of cytokines in TB inflammatory activation, none of these markers were selected by this approach.

MMP-28 is a protein coding gene that encodes a secreted enzyme associated with casein degradation and is involved in the breakdown of extracellular matrix.<sup>29</sup> The role of MMPs in TB pathogenesis has been largely studied and linked with tissue remodeling.<sup>30</sup> After *Mtb* infection, inflammatory activation leads to a host response that increases tissue destruction involving MMPs, and disrupts homeostasis, which is required for *Mtb* dissemination. Some evidence has shown the expression of MMP28 in normal tissues, indicating its role in the preservation



or maintenance of homeostatic conditions.<sup>29,31</sup> In accordance, our results identified higher expression of MMP-28 in the control group when compared with TB only patients, which reinforces the marker activity in homeostasis. If confirmed in other cohorts and future studies, the MMP-28 pathway could be used in the future as a target for host-directed therapies, to improve tissue healing or recovery or limiting the damage caused by the responses against the Mtb.

The role of eicosanoids in TB pathogenesis has been largely explored,<sup>32,33</sup> and the changes in the arachidonic acid pathways persist after ATT.<sup>34</sup> Here, we identified three eicosanoids which, in TANDEM, could help future studies that develop a host-based therapy. LTE4 is a lipid inflammatory mediator and the unique cysteinyl leukotriene stable and abundant *in vivo*.<sup>35</sup> LTE4 has been described as having a critical role in pulmonary inflammation,<sup>35</sup> through leukocyte activation and triggering cytokine production and macrophage necrosis.<sup>36</sup> Their action in TB pathogenesis remains unclear, but LTE-4 was associated with mucosal eosinophilia and airway hyperresponsiveness in asthma diseases<sup>35</sup> and seems increased in persons with DM.<sup>37</sup> In accordance, our results demonstrated higher levels of LTE4 in TB only, TB-DM and DM only when compared with non-TB/non-DM control group. Of note, we showed a decrease in the levels of LTE4 after ATT initiation, mainly in those with TB only, suggesting an indirect attenuation of lung damage after the induction phase of ATT, the first two months. However, in TB-DM participants, the levels of LTE4 returned to higher expression in month 6, corroborating a hypothesis from one of our previous publications that showed persistent inflammation during ATT in TB-DM participants<sup>13</sup> (11dTxB2 is a urinary metabolite produced from the breakdown of the thromboxane A2). Their levels, measured in urine, can be used to evaluate the response to aspirin therapy in heart disease,<sup>38</sup> asthma,<sup>39</sup> and in diseases associated with higher platelet activation<sup>40</sup> as an indirect measure of platelet activity. The role of 11dTxB2 in TB pathogenesis remains unclear, but TB patients usually present with higher platelet counts that correlate with disease severity and hypercoagulability.<sup>41</sup> Platelets contribute to MMP-mediated tissue damage through their effects on monocytes, leading to upregulation of activation markers, increases in MMP secretion and enhanced phagocytosis.<sup>41</sup> Our results reveal higher levels of 11dTxB2 associated with TB, independent of glycemic status, with decreases in the expression after ATT that could be associated with platelet activation and tissue injury consequences. Of note, considering the effects of platelets on TB pathogenesis, some anti-platelet drugs such as aspirin could be a potential target for host-directed therapy in pulmonary TB.<sup>42</sup> Finally, PGDM is a urinary metabolite from PGD2, and little is known about its function in TB and in infected alveolar macrophages. A recent study in *Histoplasma capsulatum* infection demonstrated an immunostimulatory effect of PGD2, contributing to fungicidal mechanisms and controlling the inflammatory damage.<sup>43</sup> Our results showed higher levels of PGD2 in all three groups when compared with controls, suggesting lung damage triggered by TB disease, but the real role of the molecule in the pathophysiology of the disease needs to be better evaluated. Our finding highlights the potential role of eicosanoids as candidates for host-directed therapy in TB and TB-DM. The blockage of LTE4 pathways could attenuate the cytokine storm, inflammatory activation,<sup>44</sup> and tissue damage. Leukotriene blockers are being used in asthma treatment, and 11dTxB2 has the potential to be used in the future as a marker of platelet activation in TB, as in disease progression, as well as to measure response to the host-directed therapies using anti-platelet agents.

FBXO6 is a protein code gene associated with phosphorylation-dependent ubiquitination, largely studied in cancer,<sup>45</sup> without robust evidence in infectious diseases. In a murine study of Influenza A infection, deficiencies in FBXO6 expression were associated with decreased pulmonary viral replication, inflammatory responses, and mortality.<sup>46</sup> Despite the absence of robust publications involving the gene in the TB field, we found an interesting correlation between the gene expression levels of FBXO6 and the AFB smear grade in TB-DM patients. Additionally, after ATT initiation, we identified a decrease in the expression in both TB and TB-DM, highlighting the possibilities of a new molecule associated with bacilli proliferation and target for host directed therapies. To test this hypothesis, future studies are required to delineate the mechanisms of a potential role for FBXO6 in TB.

SECTM1 is a gene that encodes a transmembrane and secreted protein. In humans, SECTM1 works as a T/NK cell co-stimulatory molecule that has been associated with CD4 and CD8 T cell proliferation and interferon-gamma (IFN- $\gamma$ ) production<sup>47</sup> and reported as an IFN early response gene.<sup>48</sup> In TB, IFN- $\gamma$  displays a pivotal role in macrophage activation and immune protection against the bacilli.<sup>49,50</sup> Importantly, SECTM1 was identified in a recent study that evaluated host protein features in persons with HIV, as a marker to evaluate Mtb infection prior to TB diagnosis.<sup>51</sup> In accordance with the literature, our findings revealed increased expression of SECTM1 in TB participants, independent of the glycemic status. Additionally, we showed a direct correlation between SECTM1 expression level and AFB smear grade in the TB-DM group, highlighting it as a possible marker to evaluate disease severity in TB-DM. Finally, during ATT the levels of SECTM1 decreased, in both the TB and TB-DM groups. Our findings emphasize the role of SECTM1 as a marker of TB, disease progression in TB-DM, and prognosis, as recently proposed.<sup>52</sup> The last selected marker by our random forest model was the LINCO2009, a non-coding RNA with unknown function in TB. The non-coding RNA displays a high potential to modulate biochemical pathways, and our results could represent a potential topic of interest for further investigations in TB-DM pathology. Future studies are required to better evaluate the role and potential of LINCO2009 in TB and TB-DM.

The final step of our study was to analyze the accuracy of the transcriptomic markers to detect TB-DM. Several studies around the world have been conducted to identify a transcriptomic profile that characterizes TB-DM interaction, but none of these obtained robust results.<sup>12,20</sup> We recently published a multi-center, prospective cohort study of whole blood gene expression in TB-DM (the MSTDI study).<sup>19</sup> We applied a random forest model to a whole RNA-seq from two sites in India<sup>53,54</sup> and one site in Brazil, as a discovery set, and used public RNA-seq data from two sites of the TANDEM consortium<sup>20</sup> as a validation set (Romania and South Africa), to verify the dynamicity across population. We identified four genes as more informative features to detect TB-DM, SMARCD3, BATF2, VAMP5, and ANKRD22, with an accuracy of 0.97 in Brazil and South Africa, 0.92 in India and 0.89 in Romania.<sup>19</sup> Here, using a similar machine learning analysis applied to multi-platform data from Brazil, we identified another four genes as more informative features to detect TB-DM, LINCO2009, SECTM1, FBXO6, and MMP28, with a concise and robust accuracy in discrimination TB-DM. We extended our approach and tested this new transcriptomic signature in the sites

previously studied in MSTDI and found a better performance in all four sites using the transcriptomic markers identified here, with an accuracy of 0.99, 0.91, 0.98, and 0.94 in Brazil, India, Romania, and South Africa, respectively. This finding highlights the potential applicability of multi-omic platforms and machine learning approaches in the study of inflammatory and/or infectious diseases since the findings obtained through the application of the decision tree on the three platforms were better and more concise than what was found when we applied the model only to transcriptomic data.

Despite some limitations, to the best of our knowledge, this is the first study to evaluate the multi-omic analysis of TB and TB-DM, with data retrieved from samples collected from the same patients and time points. By identifying novel markers that could serve as targets for future host-directed therapies, our results expand the current knowledge regarding TB-DM pathogenesis and that can be useful to improve anti-TB treatment outcomes in this population.

### Limitations of the study

This study had some limitations. First, the populations were different regarding age and BMI, factors that could influence inflammatory activation. A validation in an independent cohort is necessary to ensure the robustness of our findings. In the pre-analytical phase of the current study, we defined that all RePORT Brazil participants who had data on all the omics platforms investigated here would be included, and no pre-analytical matching of the study sample was performed. This approach may have introduced bias, which will require further investigation in additional cohorts to clarify. Second, the measurement of HbA1c levels was performed only at baseline, and changes in glycemia could also affect the immune responses after therapy initiation. We are currently developing a prospective investigation to assess HbA1c levels in a novel cohort from the RePORT Brazil attended population.

### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Ethics statement
  - Study design and population
- METHOD DETAILS
  - RNA sequencing
  - Collection, processing, and analysis of eicosanoid metabolites
  - Immunoassays
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Gene expression analysis
  - Feature selection analysis using machine learning
  - Statistical analysis

### SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2024.109135>.

### ACKNOWLEDGMENTS

We thank the study participants. We recognize the important contributions of the RePORT-Brazil Consortium, who participated at different stages of the consortium and helped to build the databases explored in the present study: Vanessa Nascimento and Michael S. Rocha (Instituto Brasileiro para Investigação da Tuberculose, Fundação José Silveira, Salvador, Brazil, Multinational Organization Network Sponsoring Translational and Epidemiological Research Initiative, Salvador, Brazil); Saulo R. N. Santos and André Ramos (Instituto Brasileiro para Investigação da Tuberculose, Fundação José Silveira, Salvador, Brazil); Juan Manuel Cubillos-Ângulo (Multinational Organization Network Sponsoring Translational and Epidemiological Research Initiative, Salvador, Brazil); Jéssica Reboças-Silva (Faculdade de Medicina, Universidade Federal da Bahia, Salvador, Brazil, Laboratório de Inflamação e Biomarcadores, Instituto Gonçalo Moniz, Fundação Oswaldo Cruz, Salvador, Brazil); Pedro Brito (Multinational Organization Network Sponsoring Translational and Epidemiological Research Initiative, Salvador, Brazil, Laboratório de Inflamação e Biomarcadores, Instituto Gonçalo Moniz, Fundação Oswaldo Cruz, Salvador, Brazil); Carolina A. S. Schmaltz, Adriano Gomes-Silva, Francine P. Ignácio, Maria C. Lourenço, Aline Benjamin, Jamile G. de Oliveira, Solange Cavalcante, and Betina Durovni (Laboratório de Pesquisa Clínica em Micobacteriose, Instituto Nacional de Infectologia Evandro Chagas, Fiocruz, Rio de Janeiro, Brazil); Alysson G. Costa (Fundação Medicina Tropical Dr Heitor Vieira Dourado, Manaus, Brazil, Programa de Pós-Graduação em Medicina Tropical, Universidade do Estado do Amazonas, Manaus, Brazil, Universidade Federal do Amazonas, Manaus, Brazil); Leandro Sousa Garcia, Brenda



K. de Sousa Carvalho, Bruna P. de Loiola, and Alexandra B. Souza (Fundação Medicina Tropical Dr Heitor Vieira Dourado, Manaus, Brazil, Programa de Pós-Graduação em Medicina Tropical, Universidade do Estado do Amazonas, Manaus, Brazil); Elisangela C. Silva, Mayla Mello, and Adriana S. R. Moreira (Programa de Pós-Graduação em Medicina Tropical, Universidade do Estado do Amazonas, Manaus, Brazil); José R. Lapa-e-Silva (Programa Acadêmico de Tuberculose da Faculdade de Medicina, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil); Megan Turner, John R Koethe, and Carlos Henrique Serezani (Division of Infectious Diseases, Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN, United States). A special thanks to Elze Leite (FIOCRUZ, Salvador, Brazil), Eduardo Gama (FIOCRUZ, Rio de Janeiro, Brazil), Elcimar Junior (FMT-HVD, Manaus, Brazil), and Hilary Vansell (VUMC, Nashville, USA) for administrative and logistical support. Some data used here were originated from publicly available databases previously reported by the RePORT India and the TANDEM consortium, and thus we thank the research teams that generated the primary data. This work was supported by the Departamento de Ciência e Tecnologia (DECIT)—Secretaria de Ciência e Tecnologia (SCTIE)—Ministério da Saúde, Brazil [25029.000507/2013-07 to V.R.], the National Institute of Allergy and Infectious Diseases [U01-AI069923] at NIH, and by The Civilian Research and Development Foundation (CRDF) Global #DAA3-17-63144. The study was partially supported by the Intramural Research Program of the Fundação Oswaldo Cruz and the Intramural Research Program of the Fundação José Silveira. B.B.A., V.C.R., A.K., and M.C.-S. are senior scientists from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). M.B.A. is Researcher Level I from The Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica (CONCYTEC) and is supported by Project Number 5 R01-AI147765-04 from NIH. G.C.S. is a scientific initiation fellow from CNPq. M.A.-P. and B.B.-D. received a research fellowship from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, finance code: 001). The funders had no role in study design, data collection, and analysis, the decision to publish, or the preparation of the manuscript.

## AUTHOR CONTRIBUTIONS

A.T.L.Q., B.B.A., T.S., M.C.-S., A.K., V.R., and M.F. designed the study and mentored the work; M.B.A., B.B.-D., M.A.-P., A.A., and B.B.A. performed the experiments and data collections; C.L.V., E.R.F., G.C.S., B.B.A., and A.T.L.Q. analyzed the data; M.B.A., M.M.B., B.B.-D., M.A.-P., V.C.R., A.L.K., M.C.-S., and T.R.S. helped with data interpretation; C.L.V., G.C.S., B.B.A., and A.T.L.Q. wrote the manuscript. All authors contributed to the article and approved the submitted version.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 29, 2023

Revised: January 2, 2024

Accepted: February 1, 2024

Published: February 5, 2024

## REFERENCES

- World Health Organization (2022). *Global Tuberculosis Report*.
- Houben, R.M.G.J., and Dodd, P.J. (2016). The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. *PLoS Med.* 13, e1002152. <https://doi.org/10.1371/journal.pmed.1002152>.
- Bell, L.C.K., and Noursadeghi, M. (2018). Pathogenesis of HIV-1 and Mycobacterium tuberculosis co-infection. *Nat. Rev. Microbiol.* 16, 80–90. <https://doi.org/10.1038/nrmicro.2017.128>.
- Kathirvel, M., and Mahadevan, S. (2016). The role of epigenetics in tuberculosis infection. *Epigenomics* 8, 537–549. <https://doi.org/10.2217/epi.16.1>.
- Murray, M., Oxlade, O., and Lin, H.H. (2011). Modeling social, environmental and biological determinants of tuberculosis. *Int. J. Tuberc. Lung Dis.* 15, 64–70. <https://doi.org/10.5588/ijtld.10.0535>.
- Gupta, K.B., Gupta, R., Atreja, A., Verma, M., and Vishvkarma, S. (2009). Tuberculosis and nutrition. *Lung India.* 26, 9–16. <https://doi.org/10.4103/0970-2113.45198>.
- World Health Organization (2016). *Global Report on Diabetes*.
- Mantovani, A., and Garlanda, C. (2023). Humoral Innate Immunity and Acute-Phase Proteins. *N. Engl. J. Med.* 388, 439–452. <https://doi.org/10.1056/NEJMra2206346>.
- Berbudi, A., Rahmadika, N., Tjahjadi, A.I., and Ruslami, R. (2020). Type 2 Diabetes and its Impact on the Immune System. *Curr. Diabetes Rev.* 16, 442–449. <https://doi.org/10.2174/1573399815666191024085838>.
- Restrepo, B.I. (2016). Diabetes and Tuberculosis. *Microbiol. Spectr.* 4, 1–11. <https://doi.org/10.1128/microbiolspec.TNMI7-0023-2016>.
- Jeon, C.Y., and Murray, M.B. (2008). Diabetes mellitus increases the risk of active tuberculosis: a systematic review of 13 observational studies. *PLoS Med.* 5, e152. <https://doi.org/10.1371/journal.pmed.0050152>.
- Prada-Medina, C.A., Fukutani, K.F., Pavan Kumar, N., Gil-Santana, L., Babu, S., Lichtenstein, F., West, K., Sivakumar, S., Menon, P.A., Viswanathan, V., et al. (2017). Systems Immunology of Diabetes-Tuberculosis Comorbidity Reveals Signatures of Disease Complications. *Sci. Rep.* 7, 1999. <https://doi.org/10.1038/s41598-017-01767-4>.
- Kumar, N.P., Fukutani, K.F., Shrutthi, B.S., Alves, T., Silveira-Mattos, P.S., Rocha, M.S., West, K., Natarajan, M., Viswanathan, V., Babu, S., et al. (2019). Persistent inflammation during anti-tuberculosis treatment with diabetes comorbidity. *Elife* 8, e46477. <https://doi.org/10.7554/eLife.46477>.
- Arriaga, M.B., Rocha, M.S., Nogueira, B.M.F., Nascimento, V., Araújo-Pereira, M., Souza, A.B., Andrade, A.M.S., Costa, A.G., Gomes-Silva, A., Silva, E.C., et al. (2021). The Effect of Diabetes and Prediabetes on Mycobacterium tuberculosis Transmission to Close Contacts. *J. Infect. Dis.* 224, 2064–2072. <https://doi.org/10.1093/infdis/jiab264>.
- Gil-Santana, L., Almeida-Junior, J.L., Oliveira, C.A.M., Hickson, L.S., Daltro, C., Castro, S., Kornfeld, H., Netto, E.M., and Andrade, B.B. (2016). Diabetes Is Associated with Worse Clinical Presentation in Tuberculosis Patients from Brazil: A Retrospective Cohort Study. *PLoS One* 11, e0146876. <https://doi.org/10.1371/journal.pone.0146876>.
- Arriaga, M.B., Araújo-Pereira, M., Barreto-Duarte, B., Nogueira, B., Freire, M.V.C.N.S., Queiroz, A.T.L., Rodrigues, M.M.S., Rocha, M.S., Souza, A.B., Spener-Gomes, R., et al. (2022). The Effect of Diabetes and Prediabetes on Antituberculosis Treatment Outcomes: A Multicenter Prospective Cohort Study. *J. Infect. Dis.* 225, 617–626. <https://doi.org/10.1093/infdis/jiab427>.
- Barreda, N.N., Arriaga, M.B., Aliaga, J.G., Lopez, K., Sanabria, O.M., Carmo, T.A., Frões Neto, J.F., Lecca, L., Andrade, B.B., and

- Calderon, R.I. (2020). Severe pulmonary radiological manifestations are associated with a distinct biochemical profile in blood of tuberculosis patients with dysglycemia. *BMC Infect. Dis.* 20, 139. <https://doi.org/10.1186/s12879-020-4843-0>.
18. Calderon, R.I., Arriaga, M.B., Aliaga, J.G., Barreda, N.N., Sanabria, O.M., Barreto-Duarte, B., Franco, J.P.D., Lecca, L., Andrade, B.B., Carvalho, A.C.C., and Kritski, A.L. (2022). Persistent dysglycemia is associated with unfavorable treatment outcomes in patients with pulmonary tuberculosis from Peru. *Int. J. Infect. Dis.* 116, 293–301. <https://doi.org/10.1016/j.ijid.2022.01.012>.
19. Queiroz, A.T.L., Vinhaes, C.L., Fukutani, E.R., Gupte, A.N., Kumar, N.P., Fukutani, K.F., Arriaga, M.B., Sterling, T.R., Babu, S., Gaikwad, S., et al. (2023). A multi-center, prospective cohort study of whole blood gene expression in the tuberculosis-diabetes interaction. *Sci. Rep.* 13, 7769. <https://doi.org/10.1038/s41598-023-34847-9>.
20. Eckold, C., Kumar, V., Weiner, J., Alisjahbana, B., Riza, A.L., Ronacher, K., Coronel, J., Kerry-Barnard, S., Malherbe, S.T., Kleynhans, L., et al. (2021). Impact of Intermediate Hyperglycemia and Diabetes on Immune Dysfunction in Tuberculosis. *Clin. Infect. Dis.* 72, 69–78. <https://doi.org/10.1093/cid/ciaa751>.
21. Hamada, Y., Penn-Nicholson, A., Krishnan, S., Cirillo, D.M., Matteelli, A., Wyss, R., Denkinger, C.M., Rangaka, M.X., Ruhwald, M., and Schumacher, S.G. (2022). Are mRNA based transcriptomic signatures ready for diagnosing tuberculosis in the clinic? - A review of evidence and the technological landscape. *EBioMedicine* 82, 104174. <https://doi.org/10.1016/j.ebiom.2022.104174>.
22. Singhania, A., Wilkinson, R.J., Rodrigue, M., Haldar, P., and O'Garra, A. (2018). The value of transcriptomics in advancing knowledge of the immune response and diagnosis in tuberculosis. *Nat. Immunol.* 19, 1159–1168. <https://doi.org/10.1038/s41590-018-0225-9>.
23. Burel, J.G., Babor, M., Pomaznoy, M., Lindestam Arlehamn, C.S., Khan, N., Sette, A., and Peters, B. (2019). Host Transcriptomics as a Tool to Identify Diagnostic and Mechanistic Immune Signatures of Tuberculosis. *Front. Immunol.* 10, 221. <https://doi.org/10.3389/fimmu.2019.00221>.
24. Warsinske, H., Vashisht, R., and Khatri, P. (2019). Host-response-based gene signatures for tuberculosis diagnosis: A systematic comparison of 16 signatures. *PLoS Med.* 16, e1002786. <https://doi.org/10.1371/journal.pmed.1002786>.
25. Krishnan, S., Queiroz, A.T.L., Gupta, A., Gupte, N., Bisson, G.P., Kumwenda, J., Naidoo, K., Mohapi, L., Mave, V., Mngqibisa, R., et al. (2021). Integrative Multi-Omics Reveals Serum Markers of Tuberculosis in Advanced HIV. *Front. Immunol.* 12, 676980. <https://doi.org/10.3389/fimmu.2021.676980>.
26. Kulkarni, V., Queiroz, A.T.L., Sangle, S., Kagal, A., Salvi, S., Gupta, A., Ellner, J., Kadam, D., Rolla, V.C., Andrade, B.B., et al. (2021). A Two-Genes Signature for Tuberculosis Diagnosis in Persons With Advanced HIV. *Front. Immunol.* 12, 631165. <https://doi.org/10.3389/fimmu.2021.631165>.
27. Yang, L., Wu, H., Jin, X., Zheng, P., Hu, S., Xu, X., Yu, W., and Yan, J. (2020). Study of cardiovascular disease prediction model based on random forest in eastern China. *Sci. Rep.* 10, 5245. <https://doi.org/10.1038/s41598-020-62133-5>.
28. Hanko, M., Grendár, M., Snopko, P., Opšėnák, R., Šutovský, J., Benčo, M., Soršák, J., Zelenák, K., and Kolarovszki, B. (2021). Random Forest-Based Prediction of Outcome and Mortality in Patients with Traumatic Brain Injury Undergoing Primary Decompressive Craniectomy. *World Neurosurg.* 148, e450–e458. <https://doi.org/10.1016/j.wneu.2021.01.002>.
29. Ra, H.J., and Parks, W.C. (2007). Control of matrix metalloproteinase catalytic activity. *Matrix Biol.* 26, 587–596. <https://doi.org/10.1016/j.matbio.2007.07.001>.
30. Amaral, E.P., Vinhaes, C.L., Oliveira-de-Souza, D., Nogueira, B., Akrami, K.M., and Andrade, B.B. (2021). The Interplay Between Systemic Inflammation, Oxidative Stress, and Tissue Remodeling in Tuberculosis. *Antioxid. Redox Signal.* 34, 471–485. <https://doi.org/10.1089/ars.2020.8124>.
31. Sabir, N., Hussain, T., Mangi, M.H., Zhao, D., and Zhou, X. (2019). Matrix metalloproteinases: Expression, regulation and role in the immunopathology of tuberculosis. *Cell Prolif.* 52, e12649. <https://doi.org/10.1111/cpr.12649>.
32. Arriaga, M.B., Karim, F., Queiroz, A.T.L., Araújo-Pereira, M., Barreto-Duarte, B., Sales, C., Moosa, M.Y.S., Mazibuko, M., Milne, G.L., Maruri, F., et al. (2022). Effect of Dysglycemia on Urinary Lipid Mediator Profiles in Persons With Pulmonary Tuberculosis. *Front. Immunol.* 13, 919802. <https://doi.org/10.3389/fimmu.2022.919802>.
33. Chen, M., Divangahi, M., Gan, H., Shin, D.S.J., Hong, S., Lee, D.M., Serhan, C.N., Behar, S.M., and Remold, H.G. (2008). Lipid mediators in innate immunity against tuberculosis: opposing roles of PGE2 and LXA4 in the induction of macrophage death. *J. Exp. Med.* 205, 2791–2801. <https://doi.org/10.1084/jem.20080767>.
34. Vinhaes, C.L., Oliveira-de-Souza, D., Silveira-Mattos, P.S., Nogueira, B., Shi, R., Wei, W., Yuan, X., Zhang, G., Cai, Y., Barry, C.E., 3rd, et al. (2019). Changes in inflammatory protein and lipid mediator profiles persist after antitubercular treatment of pulmonary and extrapulmonary tuberculosis: A prospective cohort study. *Cytokine* 123, 154759. <https://doi.org/10.1016/j.cyto.2019.154759>.
35. Paruchuri, S., Tashimo, H., Feng, C., Maekawa, A., Xing, W., Jiang, Y., Kanaoka, Y., Conley, P., and Boyce, J.A. (2009). Leukotriene E4-induced pulmonary inflammation is mediated by the P2Y12 receptor. *J. Exp. Med.* 206, 2543–2555. <https://doi.org/10.1084/jem.20091240>.
36. Salimi, M., Stöger, L., Liu, W., Go, S., Pavord, I., Klenerman, P., Ogg, G., and Xue, L. (2017). Cysteinyl leukotriene E(4) activates human group 2 innate lymphoid cells and enhances the effect of prostaglandin D(2) and epithelial cytokines. *J. Allergy Clin. Immunol.* 140, 1090–1100.e11. <https://doi.org/10.1016/j.jaci.2016.12.958>.
37. Rafnsson, A., and Bäck, M. (2013). Urinary leukotriene E4 is associated with renal function but not with endothelial function in type 2 diabetes. *Dis. Markers* 35, 475–480. <https://doi.org/10.1155/2013/370461>.
38. Lordkipanidzé, M., Pharand, C., Schampaert, E., Turgeon, J., Palisaitis, D.A., and Diodati, J.G. (2007). A comparison of six major platelet function tests to determine the prevalence of aspirin resistance in patients with stable coronary artery disease. *Eur. Heart J.* 28, 1702–1708. <https://doi.org/10.1093/eurheartj/ehm226>.
39. Oosaki, R., Mizushima, Y., Mita, H., Shida, T., Akiyama, K., and Kobayashi, M. (1997). Urinary leukotriene E4 and 11-dehydrothromboxane B2 in patients with aspirin-sensitive asthma. *Allergy* 52, 470–473. <https://doi.org/10.1111/j.1398-9995.1997.tb01032.x>.
40. Catella, F., Healy, D., Lawson, J.A., and FitzGerald, G.A. (1986). 11-Dehydrothromboxane B2: a quantitative index of thromboxane A2 formation in the human circulation. *Proc. Natl. Acad. Sci. USA* 83, 5861–5865. <https://doi.org/10.1073/pnas.83.16.5861>.
41. Kirwan, D.E., Chong, D.L.W., and Friedland, J.S. (2021). Platelet Activation and the Immune Response to Tuberculosis. *Front. Immunol.* 12, 631696. <https://doi.org/10.3389/fimmu.2021.631696>.
42. Cubillos-Angulo, J.M., Nogueira, B.M.F., Arriaga, M.B., Barreto-Duarte, B., Araújo-Pereira, M., Fernandes, C.D., Vinhaes, C.L., Villalva-Serra, K., Nunes, V.M., Miguez-Pinto, J.P., et al. (2022). Host-directed therapies in pulmonary tuberculosis: Updates on anti-inflammatory drugs. *Front. Med.* 9, 970408. <https://doi.org/10.3389/fmed.2022.970408>.
43. Pereira, P.A.T., Assis, P.A., Prado, M.K.B., Ramos, S.G., Aronoff, D.M., de Paula-Silva, F.W.G., Sorgi, C.A., and Faccioli, L.H. (2018). Prostaglandins D(2) and E(2) have opposite effects on alveolar macrophages infected with *Histoplasma capsulatum*. *J. Lipid Res.* 59, 195–206. <https://doi.org/10.1194/jlr.M078162>.
44. Tobin, D.M., Roca, F.J., Oh, S.F., McFarland, R., Vickery, T.W., Ray, J.P., Ko, D.C., Zou, Y., Bang, N.D., Chau, T.T.H., et al. (2012). Host genotype-specific therapies can optimize the inflammatory response to mycobacterial infections. *Cell* 148, 434–446. <https://doi.org/10.1016/j.cell.2011.12.023>.
45. Merry, C., Fu, K., Wang, J., Yeh, I.J., and Zhang, Y. (2010). Targeting the checkpoint kinase Chk1 in cancer therapy. *Cell Cycle* 9, 279–283. <https://doi.org/10.4161/cc.9.2.10445>.
46. Cen, M., Ouyang, W., Lin, X., Du, X., Hu, H., Lu, H., Zhang, W., Xia, J., Qin, X., and Xu, F. (2023). FBXO6 regulates the antiviral immune responses via mediating alveolar macrophages survival. *J. Med. Virol.* 95, e28203. <https://doi.org/10.1002/jmv.28203>.
47. Wang, T., Huang, C., Lopez-Coral, A., Slentz-Kesler, K.A., Xiao, M., Wherry, E.J., and Kaufman, R.E. (2012). K12/SECTM1, an interferon-gamma regulated molecule, synergizes with CD28 to costimulate human T cell proliferation. *J. Leukoc. Biol.* 91, 449–459. <https://doi.org/10.1189/jlb.1011498>.
48. Huyton, T., Göttmann, W., Bade-Döding, C., Paine, A., and Blasczyk, R. (2011). The T/NK cell co-stimulatory molecule SECTM1 is an IFN "early response gene" that is negatively regulated by LPS in human monocyte cells. *Biochim. Biophys. Acta* 1810, 1294–1301. <https://doi.org/10.1016/j.bbagen.2011.06.020>.
49. Khan, T.A., Mazhar, H., Saleha, S., Tipu, H.N., Muhammad, N., and Abbas, M.N. (2016). Interferon-Gamma Improves Macrophages Function against *M. tuberculosis* in Multidrug-Resistant Tuberculosis Patients. *Chemother. Res. Pract.* 2016, 7295390. <https://doi.org/10.1155/2016/7295390>.

50. Flynn, J.L., Chan, J., Triebold, K.J., Dalton, D.K., Stewart, T.A., and Bloom, B.R. (1993). An essential role for interferon gamma in resistance to Mycobacterium tuberculosis infection. *J. Exp. Med.* 178, 2249–2254. <https://doi.org/10.1084/jem.178.6.2249>.
51. Singer, S.N., Ndumego, O.C., Kim, R.S., Ndung'u, T., Mlamla, Z., French, A., Churchyard, G., Paramithiothis, E., Kasprovicz, V.O., and Achkar, J.M. (2022). Plasma host protein biomarkers correlating with increasing Mycobacterium tuberculosis infection activity prior to tuberculosis diagnosis in people living with HIV. *EBioMedicine* 75, 103787. <https://doi.org/10.1016/j.ebiom.2021.103787>.
52. Young, B.L., Mlamla, Z., Gqamana, P.P., Smit, S., Roberts, T., Peter, J., Theron, G., Govender, U., Dheda, K., and Blackburn, J. (2014). The identification of tuberculosis biomarkers in human urine samples. *Eur. Respir. J.* 43, 1719–1729. <https://doi.org/10.1183/09031936.00175113>.
53. Kornfeld, H., West, K., Kane, K., Kumpatla, S., Zacharias, R.R., Martinez-Balzano, C., Li, W., and Viswanathan, V. (2016). High Prevalence and Heterogeneity of Diabetes in Patients With TB in South India: A Report from the Effects of Diabetes on Tuberculosis Severity (EDOTS) Study. *Chest* 149, 1501–1508. <https://doi.org/10.1016/j.chest.2016.02.675>.
54. Gupte, A., Padmapriyadarsini, C., Mave, V., Kadam, D., Suryavanshi, N., Shivakumar, S.V.B.Y., Kohli, R., Gupte, N., Thiruvengadam, K., Kagal, A., et al. (2016). Cohort for Tuberculosis Research by the Indo-US Medical Partnership (CTRIUMPH): protocol for a multicentric prospective observational study. *BMJ Open* 6, e010542. <https://doi.org/10.1136/bmjopen-2015-010542>.
55. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. <https://doi.org/10.1093/bioinformatics/bts635>.
56. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. <https://doi.org/10.1186/s13059-014-0550-8>.
57. Breiman, L. (2001). Random Forest. *Mach. Learn.* 45, 5–32. <https://doi.org/10.1023/A:1010933404324>.
58. Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *J. Stat. Softw.* 28, 26. <https://doi.org/10.18637/jss.v028.i05>.
59. Arriaga, M.B., Amorim, G., Queiroz, A.T.L., Rodrigues, M.M.S., Araújo-Pereira, M., Nogueira, B.M.F., Souza, A.B., Rocha, M.S., Benjamin, A., Moreira, A.S.R., et al. (2021). Novel stepwise approach to assess representativeness of a large multicenter observational cohort of tuberculosis patients: The example of RePORT Brazil. *Int. J. Infect. Dis.* 103, 110–118. <https://doi.org/10.1016/j.ijid.2020.11.140>.
60. Geadas, C., Stoszek, S.K., Sherman, D., Andrade, B.B., Srinivasan, S., Hamilton, C.D., and Ellner, J. (2017). Advances in basic and translational tuberculosis research: Proceedings of the first meeting of RePORT international. *Tuberculosis (Edinb)* 102, 55–67. <https://doi.org/10.1016/j.tube.2016.11.006>.
61. van der Heijden, Y.F., Abdullah, F., Andrade, B.B., Andrews, J.R., Christopher, D.J., Croda, J., Ewing, H., Haas, D.W., Hatherill, M., Horsburgh, C.R., Jr., et al. (2018). Building capacity for advances in tuberculosis research; proceedings of the third RePORT international meeting. *Tuberculosis (Edinb)* 113, 153–162. <https://doi.org/10.1016/j.tube.2018.09.009>.
62. Arriaga, M.B., Araújo-Pereira, M., Barreto-Duarte, B., Sales, C., Miguez-Pinto, J.P., Nogueira, E.B., Nogueira, B.M.F., Rocha, M.S., Souza, A.B., Benjamin, A., et al. (2021). Prevalence and Clinical Profiling of Dysglycemia and HIV Infection in Persons With Pulmonary Tuberculosis in Brazil. *Front. Med.* 8, 804173. <https://doi.org/10.3389/fmed.2021.804173>.
63. Broekmans, J.F., Migliori, G.B., Rieder, H.L., Lees, J., Ruutu, P., Lodenkemper, R., Raviglione, M.C.; World Health Organization International Union Against Tuberculosis and Lung Disease and Royal Netherlands Tuberculosis Association Working Group, and Lung, D.; Royal Netherlands Tuberculosis Association Working Group (2002). European framework for tuberculosis control and elimination in countries with a low incidence. Recommendations of the World Health Organization (WHO), International Union Against Tuberculosis and Lung Disease (IUATLD) and Royal Netherlands Tuberculosis Association (KNCV) Working Group. *Eur. Respir. J.* 19, 765–775. <https://doi.org/10.1183/09031936.02.00261402>.
64. American Diabetes Association Professional Practice Committee, Draznin, B., Arora, V.R., Bakris, G., Benson, G., Brown, F.M., Freeman, R., Green, J., Huang, E., Isaacs, D., et al. (2022). 6. Glycemic Targets: Standards of Medical Care in Diabetes-2022. *Diabetes Care* 45, S83–S96. <https://doi.org/10.2337/dc22-S006>.
65. Song, W.L., Wang, M., Ricciotti, E., Fries, S., Yu, Y., Grosser, T., Reilly, M., Lawson, J.A., and FitzGerald, G.A. (2008). Tetranor PGDM, an abundant urinary metabolite reflects biosynthesis of prostaglandin D2 in mice and humans. *J. Biol. Chem.* 283, 1179–1188. <https://doi.org/10.1074/jbc.M706839200>.
66. Murphey, L.J., Williams, M.K., Sanchez, S.C., Byrne, L.M., Csiki, I., Oates, J.A., Johnson, D.H., and Morrow, J.D. (2004). Quantification of the major urinary metabolite of PGE2 by a liquid chromatographic/mass spectrometric assay: determination of cyclooxygenase-specific PGE2 synthesis in healthy humans and those with lung cancer. *Anal. Biochem.* 334, 266–275. <https://doi.org/10.1016/j.ab.2004.08.019>.
67. Oliveira-de-Souza, D., Vinhaes, C.L., Arriaga, M.B., Kumar, N.P., Cubillos-Angulo, J.M., Shi, R., Wei, W., Yuan, X., Zhang, G., Cai, Y., et al. (2019). Molecular degree of perturbation of plasma inflammatory markers associated with tuberculosis reveals distinct disease profiles between Indian and Chinese populations. *Sci. Rep.* 9, 8002. <https://doi.org/10.1038/s41598-019-44513-8>.

## STAR★METHODS

## KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Raw RNAseq data from TANDEM	Eckold et al. <sup>20</sup>	Bioproject ID: PRJNA470512
Raw RNAseq from RePORT	Kornfeld et al. <sup>53</sup> ; Gupte et al. <sup>54</sup> ; Queiroz et al. <sup>19</sup>	GEOncbi ID: GSE181143
Software and algorithms		
R version 4.2.2.	R Core Team	<a href="https://cran.r-project.org/">https://cran.r-project.org/</a>
STAR version 2.7.10	Dobin et al. <sup>55</sup>	<a href="https://code.google.com/archive/p/rna-star/">https://code.google.com/archive/p/rna-star/</a>
DESeq2 version 1.40.2	Love et al. <sup>56</sup>	<a href="https://www.bioconductor.org/packages/release/bioc/html/DESeq2.html">https://www.bioconductor.org/packages/release/bioc/html/DESeq2.html</a>
randomForest version 4.7–1.1	Breiman et al. <sup>57</sup>	<a href="https://cran.r-project.org/web/packages/randomForest/index.html">https://cran.r-project.org/web/packages/randomForest/index.html</a>
caret version 6.0–94	Kuhn et al. <sup>58</sup>	<a href="https://cran.r-project.org/web/packages/caret/index.html">https://cran.r-project.org/web/packages/caret/index.html</a>
complexheatmap	<a href="https://doi.org/10.1093/bioinformatics/btw313">https://doi.org/10.1093/bioinformatics/btw313</a>	<a href="https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html">https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html</a>
Hmisc version 5.1–1	<a href="https://cran.rproject.org/web/packages/Hmisc/Hmisc.pdf">https://cran.rproject.org/web/packages/Hmisc/Hmisc.pdf</a>	<a href="https://hbiostat.org/r/hmisc/">https://hbiostat.org/r/hmisc/</a>

## RESOURCE AVAILABILITY

## Lead contact

Further information and requests regarding the packages employed for the analysis performed in this study should be directed to and will be fulfilled, by the lead contact, Bruno B. Andrade ([bruno.andrade@fiocruz.br](mailto:bruno.andrade@fiocruz.br)).

## Materials availability

- This study did not generate new unique reagents.

## Data and code availability

- Data used here were from two previous works. The gene expression from the main population, Salvador site from RePORT Brazil, and Indian population (used only in the validation of the transcriptomic signature) were from the MSTDI study and have been deposited at the GEO ncbi database and is publicly available as of the date of publication (Accession number GSE181143, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE181143>).

The validation of the transcriptomic signature identified here was performed using an external cohort, with population from South Africa and Romania, previously published by the TANDEM consortium and have been previously deposited at the SRA database and are publicly available as of the date of publication (BioProject ID PRJNA470512).

- All employed packages' references are available at the [key resources tables](#) and methodology.
- This paper does not report original code. Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

## EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

## Ethics statement

The study was conducted according to the Declaration of Helsinki. Data and specimens were obtained from the RePORT-Brazil observational cohort study. The study was approved by the Institutional Review Boards of the Instituto Nacional de Infectologia Evandro Chagas, Fundação Oswaldo Cruz, Brazil (CAAE: 25102412.3.1001.5262). Written informed consents were obtained from all voluntary participants.

## Study design and population

This was a retrospective analysis of a prospective observational study with data and specimens collected within the scope of RePORT-Brazil,<sup>59–61</sup> at the Salvador site. All participants were enrolled between June 2015 and June 2019 and followed for up to 24 months. Several studies from RePORT-Brazil have explored different aspects of TB-DM.<sup>14,16,32,62</sup> However, the evaluation of multi-platform data has not been previously explored. Additionally, to evaluate the accuracy of the transcriptomic signature identified in this study we used public data from a previously published work from the RePORT-India<sup>53,54</sup> and TANDEM consortium.<sup>20</sup>

Participants from the RePORT-Brazil were at least 18 years old, had an Mtb culture-positive sputum at enrollment (for those in the TB groups) and were evaluated at three in-person visits: (i) ATT initiation (baseline), (ii) two months after initiating ATT, and (iii) after completing treatment (approximately month 6). The sputum smears were prepared by the Ziel–Nielsen method using 1% carbol-fuchsin and scored using the International Union Against Tuberculosis and Lung Disease (IUATLD) scale.<sup>63</sup> Non-TB/Non-DM controls and DM patients (without TB) were selected from close TB contacts who agreed to participate in the study. These patients were evaluated at two visits: baseline and 6 months after enrollment. We selected those participants who tested negative, at both visits, for Mtb infection by QuantiFERON-TB Gold. The selection of the participants included in the sub-groups was based on availability of databases on each assay platform assessed in the present study. As a result, only participants who had data on all the platforms evaluated were included. This was pre-specified in the analysis plan to allow data integration and multi-omic analyses. Sociodemographic, clinical, and epidemiologic data such as age, sex, race/ethnicity (self-reported), and body mass index (BMI) were collected at baseline and shown in Table S1. Of note, gender variable was not collected. Regarding the biological sex and race/ethnicity (self-reported) our analysis found no differences between the groups. Glycated hemoglobin (HbA1c) was measured in all participants at enrollment.<sup>32</sup>

Diabetes was defined according to baseline HbA1c, following American Diabetes Association (ADA) guidelines.<sup>64</sup> Individuals were classified as having DM (HbA1c  $\geq$  6.5%) or normoglycemia (HbA1c < 5.7%).

## METHOD DETAILS

### RNA sequencing

Samples from TB patients (with and without DM) were collected at baseline, 2, and 6 months of treatment. Samples for the non-TB/non-DM control and DM only groups were collected at baseline. Total RNA from venous blood (2.5 mL) was collected in PAXgene Blood RNA Tubes (PreAnalytiX), extracted using a PAXgene blood miRNA kit (Qiagen) with the semi-automated QIAcube (Qiagen) and quantified using the LabChip GX HiSens RNA system (PerkinElmer). RNA-seq libraries were prepared using the Bioscientific NEXTflex-Rapid-Directional mRNA-seq sample preparation with the Caliper SciClone. Samples were sequenced using the NextSeq500 High Output kit V2 (Illumina) for 75 cycles. For all samples, RNA was sequenced by Illumina HiSeq 2500 at MedGenome in Bangalore, India.

### Collection, processing, and analysis of eicosanoid metabolites

Urine samples from patients were collected at baseline, 2 months, and 6 months of treatment. Urine was stored at  $-80^{\circ}\text{C}$  within 1 h of sample collection.<sup>65,66</sup> PGE-M and PGD-M are unstable if urine is not stored at  $-80^{\circ}\text{C}$  within 90 min of collection. Concentrations of the major urinary prostaglandin (PG)E<sub>2</sub> metabolite, PGE-M; the tetranor-PGE<sub>1</sub>, TN-E, the major urinary metabolite of PGD<sub>2</sub>, PGD-M; 11-dehydro-thromboxane-B<sub>2</sub>, dTxB<sub>2</sub>; the metabolite of PGI<sub>2</sub>, PGI-M; and leukotriene (LT)E<sub>4</sub> were measured in urine at each time point in all study participant. The assays were performed at the Eicosanoid Core Laboratory at Vanderbilt University Medical Center, in the USA. Samples were shipped on dry ice and stored at  $-80^{\circ}\text{C}$  until analysis.

[<sup>2</sup>H<sub>6</sub>]-PGE-M and [<sup>2</sup>H<sub>11</sub>]-TN-E were synthesized as previously described.<sup>32</sup> [<sup>2</sup>H<sub>4</sub>]-PGI-M and [<sup>2</sup>H<sub>4</sub>]-11dTxB<sub>2</sub> were purchased from Cayman Chemicals (Ann Arbor, MI USA), and [20,20,20-<sup>2</sup>H<sub>3</sub>]-LTE<sub>4</sub> from Enzo Life Sciences (Farmingdale, NY USA). Sep-Pak C18 and Oasis HLB (3cc/60mg) extraction cartridges were obtained from Waters Corporation (Milford, MA USA). The organic reagents were of high-performance Liquid Chromatograph (LC) quality and purchased from Sigma Aldrich (St. Louis, MO USA).

### Immunoassays

Plasma was stored at  $-80^{\circ}\text{C}$  for immunology assays. Using the Bio-Plex Pro Human Cytokine Standard 27-Plex kit (Group I) on the Biorad Bioplex 200 Luminex platform the following analytes were measured: interleukin (IL)-1 $\beta$ , IL-1 receptor antagonist (IL-1Ra), IL-2, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-12p70, IL-13, IL-15 IL-17A, eotaxin, granulocyte colony stimulating factor (G-CSF), granulocyte-macrophage colony stimulating factor (GM-CSF/CSF2), interferon gamma (IFN- $\gamma$ ), monocyte chemoattractant protein-1 (MCP-1)/C-C motif chemokine ligand 2 (CCL2), macrophage inflammatory protein-1 alpha (MIP-1 $\alpha$ /CCL3), MIP-1 beta (MIP-1 $\beta$ /CCL4), platelet-derived growth factor-BB (PDGF), regulated on activation, normal T cell expressed and secreted (RANTES/CCL5), tumor necrosis factor-alpha (TNF- $\alpha$ ), and vascular endothelial growth factor (VEGF). Biomarkers were chosen based on commercially available pre-mixed kits and included parameters previously described to be involved in TB pathogenesis.<sup>34,67</sup>

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Gene expression analysis

After the quality check, sequences were aligned to the human transcriptome (GRCh38 version 100), comprising mRNA and ncRNA, using STAR.<sup>55</sup> Count gene expression matrix was examined using the DESeq2 R package,<sup>56</sup> version 4.2.2, to identify differentially expressed genes

(DEG) across the four patient groups. Changes in gene expression with false discovery rate (FDR)-adjusted p value <0.05 and log<sub>2</sub> fold-difference +1.4 were considered significant.

### Feature selection analysis using machine learning

We employed a feature selection approach to identify the most relevant features from the multi-omic factors used here associated with TB-DM. The integrated multi-omic data, composed by all three platforms measured in the study, transcriptomic data, cytokines, and eicosanoids, were used to train the model and cross validation, applying the random forest algorithm, using the randomForest package.<sup>57</sup> 1000 trees were used in the model and at each split the number of features was 41.88, the square root of the total of variables in the dataset plus one. Furthermore, a leave-one-out cross-validation with 50-folds, and 5 repetitions was performed to assess the markers' accuracy, using the caret package.<sup>58</sup> The cross-validation allowed the estimation of the model's accuracy and the No information rate, which was 1(1–0.95) and 0.36 respectively. The p value for accuracy > No information rate was 0.0000000000000022. With this method, we aim to reduce the dimension of data and obtain the most accurate model. Seven features were selected after applied the random forest model and included in the analysis presented here. [Figure S1](#) illustrates the model design. [Table S2](#) and [Figure S2](#) show the variable importance of the top 20 features.

### Statistical analysis

The median values with interquartile ranges (IQR) were used as measures of central tendency and dispersion. Features were compared between the study groups using the Kruskal-Wallis test with Dunn's multiple comparisons. Categorical variables were compared using Chi-square test. Hierarchical cluster analyses (Ward's method), with 100X bootstrap of Z score normalized data were employed to depict the overall multi-omic expression profile in the study groups. Dendrograms represent Manhattan distance.

Spearman correlation analysis was used to evaluate the interaction between sputum smear AFB grade and multi-omic expression data, and between HbA1c levels and the multi-omic expression data. All univariate comparisons with p values <0.05 after adjustments were considered statistically significant. The statistical analyses were performed using R 4.2.2. The R package used to perform the analyses here is described in [Table S4](#).