



OPEN

## A multi-center, prospective cohort study of whole blood gene expression in the tuberculosis-diabetes interaction

Artur T. L. Queiroz<sup>1,2,3,25</sup>, Caian L. Vinhaes<sup>2,3,4,25</sup>, Eduardo R. Fukutani<sup>1</sup>, Akshay N. Gupte<sup>5</sup>, Nathella Pavan Kumar<sup>6</sup>, Kiyoshi F. Fukutani<sup>2</sup>, María B. Arriaga<sup>2,3</sup>, Timothy R. Sterling<sup>7</sup>, Subash Babu<sup>6</sup>, Sanjay Gaikwad<sup>8</sup>, Rajesh Karyakarte<sup>9</sup>, Vidya Mave<sup>10,11</sup>, Mandar Paradhkar<sup>10,11</sup>, Vijay Viswanathan<sup>12</sup>, Amita Gupta<sup>5</sup>, Bruno B. Andrade<sup>2,3,4,13,26</sup>, Hardy Kornfeld<sup>14,15,26</sup>, the RePORT Brazil\* & RePORT India Consortia\*

Diabetes mellitus (DM) increases tuberculosis (TB) severity. We compared blood gene expression in adults with pulmonary TB, with or without diabetes mellitus (DM) from sites in Brazil and India. RNA sequencing (RNAseq) performed at baseline and during TB treatment. Publicly available baseline RNAseq data from South Africa and Romania reported by the TANDEM Consortium were also analyzed. Across the sites, differentially expressed genes varied for each condition (DM, TB, and TBDM) and no pattern classified any one group across all sites. A concise signature of TB disease was identified but this was expressed equally in TB and TBDM. Pathway enrichment analysis failed to distinguish TB from TBDM, although there was a trend for greater neutrophil and innate immune pathway activation in TBDM participants. Pathways associated with insulin resistance, metabolic dysfunction, diabetic complications, and chromosomal instability were positively correlated with glycohemoglobin. The immune response to pulmonary TB as reflected by whole blood gene expression is substantially similar with or without comorbid DM. Gene expression pathways associated with the microvascular and macrovascular complications of DM are upregulated during TB, supporting a syndemic interaction between these coprevalent diseases.

Diabetes mellitus (DM) has been associated with increased risk for tuberculosis (TB) progression and adverse TB treatment outcomes in most clinical studies<sup>1</sup>. Mirroring the human data, animal models combining chronic

<sup>1</sup>Centro de Integração de Dados e Conhecimentos para Saúde, Instituto Gonçalo Moniz, Fundação Oswaldo Cruz, Salvador, Brazil. <sup>2</sup>Laboratório de Inflamação e Biomarcadores, Instituto Gonçalo Moniz, Fundação Oswaldo Cruz, Salvador, Brazil. <sup>3</sup>Multinational Organization Network Sponsoring Translational and Epidemiological Research (MONSTER) Initiative, Salvador 41810-710, Brazil. <sup>4</sup>Escola Bahiana de Medicina e Saúde Pública (EBMSP), Salvador 40290-150, Brazil. <sup>5</sup>Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. <sup>6</sup>National Institutes of Health- NIRT - International Center for Excellence in Research, Chennai, India. <sup>7</sup>Division of Infectious Diseases, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. <sup>8</sup>Department of Pulmonary Medicine, Byramjee-Jeejeebhoy Government Medical College and Sassoon General Hospitals, Pune, India. <sup>9</sup>Department of Microbiology, Byramjee-Jeejeebhoy Government Medical College and Sassoon General Hospitals, Pune, India. <sup>10</sup>Byramjee-Jeejeebhoy Government Medical College-Johns Hopkins University Clinical Research Site, Pune, India. <sup>11</sup>Johns Hopkins Center for Infectious Diseases in India, Pune, India. <sup>12</sup>Prof. M. Viswanathan Diabetes Research Centre, Chennai, India. <sup>13</sup>Faculdade de Tecnologia e Ciências, Instituto de Pesquisa Clínica e Translacional, Salvador 41741-590, Brazil. <sup>14</sup>Department of Medicine, University of Massachusetts Medical School, Worcester, MA, USA. <sup>15</sup>UMass Chan Medical School, Worcester, MA, USA. <sup>25</sup>These authors contributed equally: Artur T. L. Queiroz and Caian L. Vinhaes. <sup>26</sup>These authors jointly supervised this work: Bruno B. Andrade and Hardy Kornfeld. \*List of authors and their affiliations appears at the end of the paper. ✉email: bruno.andrade@fiocruz.br; hardy.kornfeld@umassmed.edu

hyperglycemia with *Mycobacterium tuberculosis* challenge showed higher lung bacterial burden and more TB immune pathology<sup>2</sup>. The global population-attributable fraction of TB associated with DM is comparable to that of HIV/AIDS<sup>3</sup>. Despite its significance as a barrier to TB elimination<sup>4</sup>, the mechanisms whereby DM impairs host defense against *M. tuberculosis* are not well understood<sup>5</sup>.

Mechanistic studies of human immunity to TB are limited to accessible tissue samples. Blood transcriptomic studies have described a consensus TB signature of increased type I interferon (IFN) signaling<sup>6</sup>. In a prior blood transcriptome study using microarrays to evaluate a South Indian cohort, we found no pattern of immune gene expression that distinguished TB-DM comorbidity from TB in participants without DM<sup>7</sup>. The TANDEM Consortium later published a whole blood RNA sequencing (RNAseq) study from four national sites (South Africa, Peru, Indonesia, Romania)<sup>8</sup>. Their analysis revealed increased inflammatory pathway gene expression but reduced type I interferon signaling in TB combined with dysglycemia compared to TB in people with normoglycemia. To further explore the impact of DM on the host response to TB in diverse populations, we performed RNAseq on whole blood RNA sampled at TB diagnosis (baseline) and treatment months 2 and 6 from adults with pulmonary TB, with or without DM, at sites in India and Brazil.

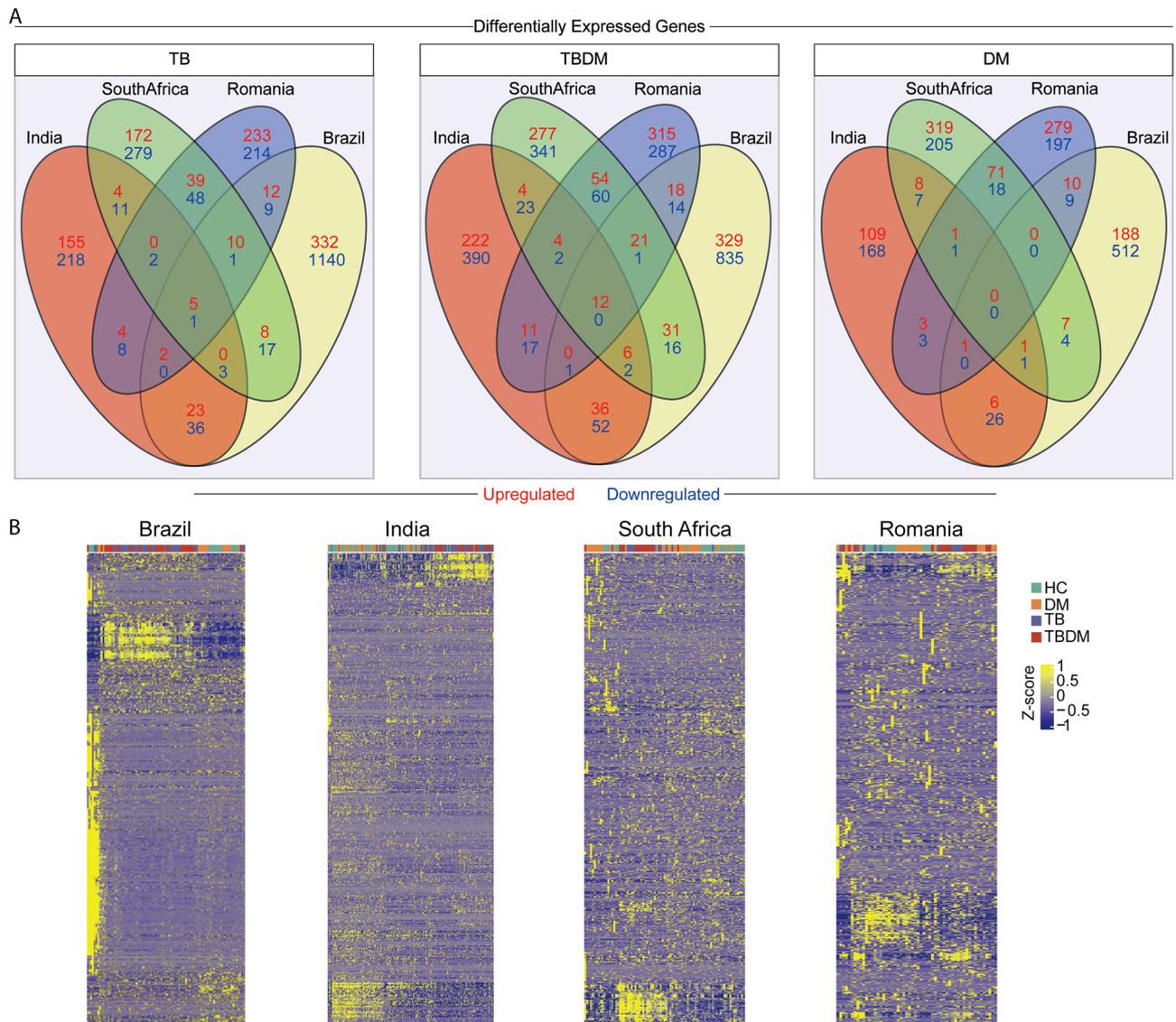
The Molecular Signatures of Tuberculosis-Diabetes Interaction (MSTDI) study reported here leveraged unreported participant data and whole blood RNA samples from prospective observational pulmonary TB cohorts from two sites of the RePORT India and one from RePORT Brazil consortia. A single vendor performed RNAseq on baseline samples from all three sites and longitudinal samples during TB treatment from one Indian site and Brazil. To check the variability of the blood transcriptome in TB-DM interaction, comparison was made with published blood RNAseq data from two TANDEM consortium sites (Romania and South Africa). Our analysis was aimed to compare the intensity and quality of inflammatory activation between the clinical conditions and sites and evaluate the impact of HbA1c levels in the biological pathways. This investigation did not reveal insights to the mechanisms of TB susceptibility in DM but the data support the existence of a syndemic interaction between TB and DM<sup>9</sup>.

## Results

**Differential gene expression between clinical conditions and sites.** The MSTDI cohort, comprising 290 participants, was recruited from two sites in India and one site in Brazil. Adults newly diagnosed with drug-sensitive pulmonary TB, with or without DM (TB and TBDM, respectively), and control group participants without TB, with or without DM (DM and HC, respectively) were enrolled. Characteristics of the population are shown in Supplementary Fig. S1 and Supplementary Tables S1 and S2. Additional comparison was made with publicly available RNAseq data from two sites (Romania and South Africa) of a TANDEM consortium gene expression study with similar group structure and where data were available for site-specific healthy control (HC) participants.

Gene expression within the disease condition groups (DM, TB, TBDM) was compared against site-specific HC participants. The raw data on DEGs are shown in Supplementary File S1. Across all four sites, the DM groups exhibited no DEGs in common, while the TB and TBDM groups shared six and twelve DEGs, respectively (Fig. 1A). A z-score normalized heatmap using the combined DEG gene expression values further demonstrated the variability within and between groups and sites, finding no pattern that classified any clinical group across all sites (Fig. 1B). A principal component analysis model (PCA) applied to DEGs differentiated between groups by the presence or absence of TB disease but did not discriminate between TB and TBDM (Supplementary Fig. S2A,B). Furthermore, we calculated the molecular degree of perturbation (MDP) score<sup>10</sup> to estimate the overall level of inflammation within disease groups at each site. The highest individual and median MDP values were present in the TB and TBDM groups and there was a non-significant trend for higher MDP in the TBDM than in the TB groups at all sites (Fig. 2A–D). Statistically significant differences in MDP scores across the four study sites were identified in the HC, DM, TB, and TBDM groups (Fig. 2E). To evaluate possible effects of clinical and epidemiological features in the differential gene expression, we performed a PCA, labelling the participants according to presence or absence of cavitation (Supplementary Fig. S3A,B). Furthermore, a Spearman correlation between BMI and the MDP values was performed (Supplementary Fig. S3C,D). No association between BMI values and gene expression variation was observed in the TB and TBDM groups in both sites. The results showed that we could not segregate the participants according to the presence or absence of cavitation, and that the BMI values were not associated with the degree of inflammatory activation, highlighting that the population-specific differences for all these clinical conditions were likely not associated with clinical and epidemiological features.

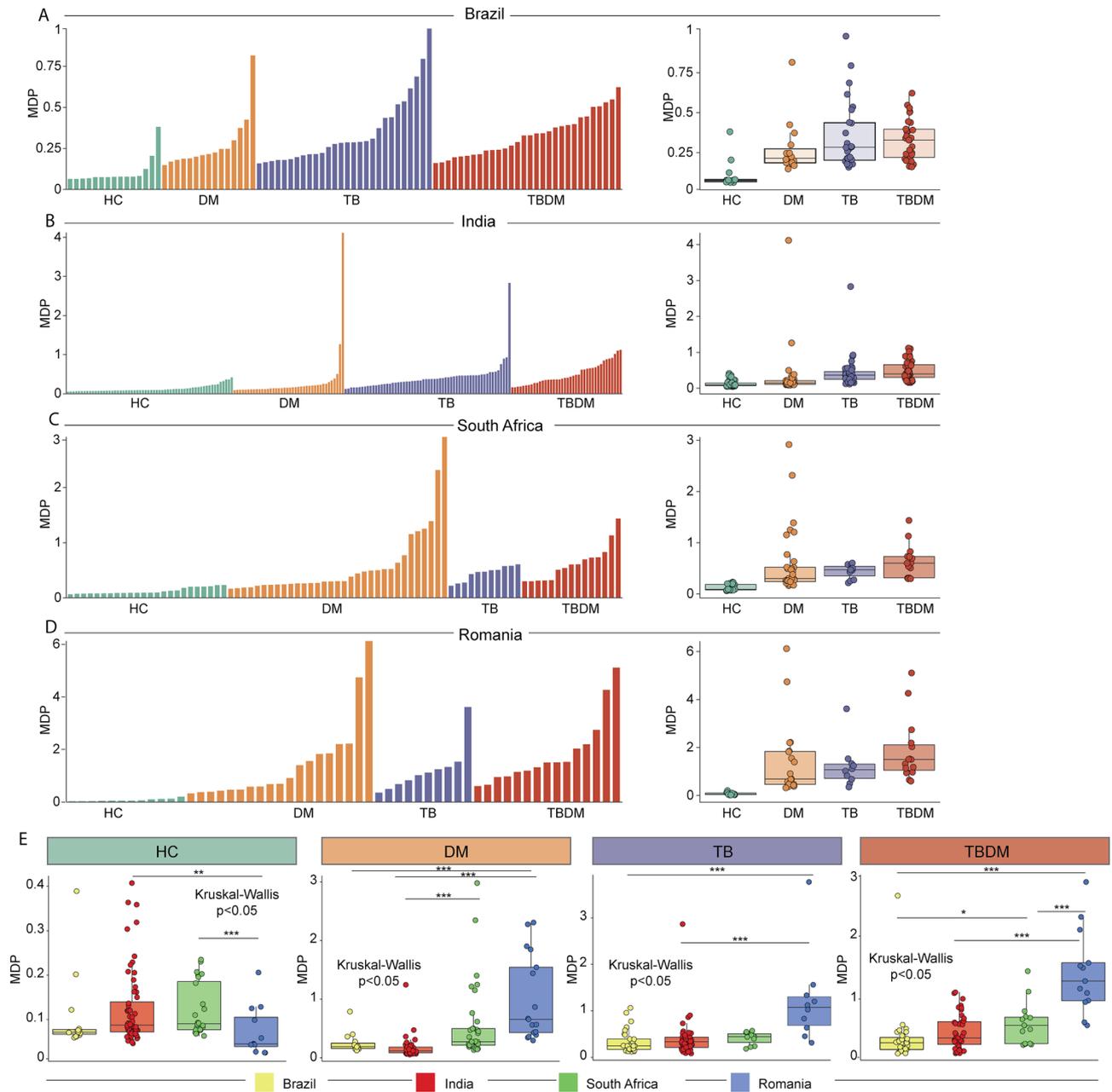
**Discovery and validation of a TB-associated gene expression signature.** As an alternative approach to identify gene expression patterns associated with specific groups across all sites, we inputted expression values of the combined total of 3427 DEGs for a random forest model (Supplementary Fig. S4). The South Africa and Romania sites were used as a discovery set for each of the four conditions (HC, DM, TB, TBDM) and the model was tested in a validation set composed of samples from India and Brazil (Supplementary Fig. S4). Model validation resulted in the accuracy of 58.97 (63.32–74.84);  $p = 0.001658$ . Gene expression data were used in a feature selection analysis with a random forest algorithm to rank the variables according to their model importance and those above the third quartile were selected, identifying SMARCD3, VAMP5, ANKRD22, and BATF2 as most informative genes to distinguish the clinical conditions (Supplementary Fig. S4A). An unsupervised cluster analysis of z-score normalized expression data identified higher baseline expression of these four genes in participants with TB at all four sites, irrespective of DM status (Fig. 3). Consistent with that observation, Spearman correlation analysis performed between HbA1c levels and expression of the most informative genes did not reveal a consistent association between the expression of these genes and HbA1c levels in participants from Brazil or India (Supplementary Fig. S4B,C). While the 4-gene signature did not discriminate between TB



**Figure 1.** Distribution of differentially expressed genes (DEGs) between clinical groups in the sites from MSTDI and TANDEM cohorts. **(A)** Venn Diagrams show the DEGs defined by the threshold corrected  $p$ -value  $< 0.05$  and log twofold change  $> \pm 1.4$  from each comparison of TB vs, TBDM vs, and DM vs HC from the subjects within the cohort sites. **(B)** Heat maps of z-score normalized data of all DEGs from each cohort site. Each heat map depicts all DEGs from TB vs, TBDM vs, and DM vs HC comparison from subjects within each cohort site.

and TBDM, receiver operator characteristic (ROC) curve analysis demonstrated that the signature presented good accuracy to classify TB and TBDM from DM and HC at all four sites with  $\text{ROC} \geq 0.85$  (Fig. 4). Additionally, we tested the accuracy of previously published gene biosignatures in our study sites, with relatively high accuracy of most of signatures to identify either TB or TBDM participants (Supplementary Figs. S5, S6, respectively). The temporal expression of these signature genes over the course of TB treatment differed among the individual genes, the condition (TB or TBDM), and between the Indian and Brazilian cohorts (Fig. 5). Expression levels of all four signature genes tended to be higher in TBDM than TB in the Brazil cohort at month-6, rising from a nadir at month-3. In the India cohort, expression of *BATF2*, *VAMP5*, and *ANKRD22* tended to be higher in TBDM than TB at baseline and month-2, with *VAMP5* and *ANKRD22* continuing that trend to month-6. These patterns might reflect persistent inflammation in TBDM, which was identified in prior study measuring plasma cytokines<sup>11</sup>.

**Pathway enrichment and interaction analysis across conditions and populations.** While comparison of DEGs failed to discriminate between TB and TBDM, we questioned whether pathway enrichment had the potential to reveal condition-specific differences. Reactome pathways of interest were identified within significant ( $p < 0.05$ ) and false discovery rate (FDR)-corrected DEGs from all conditions and sites. No pathway was uniquely enriched within the TB or TBDM groups at all sites, but interferon signaling that has been identi-

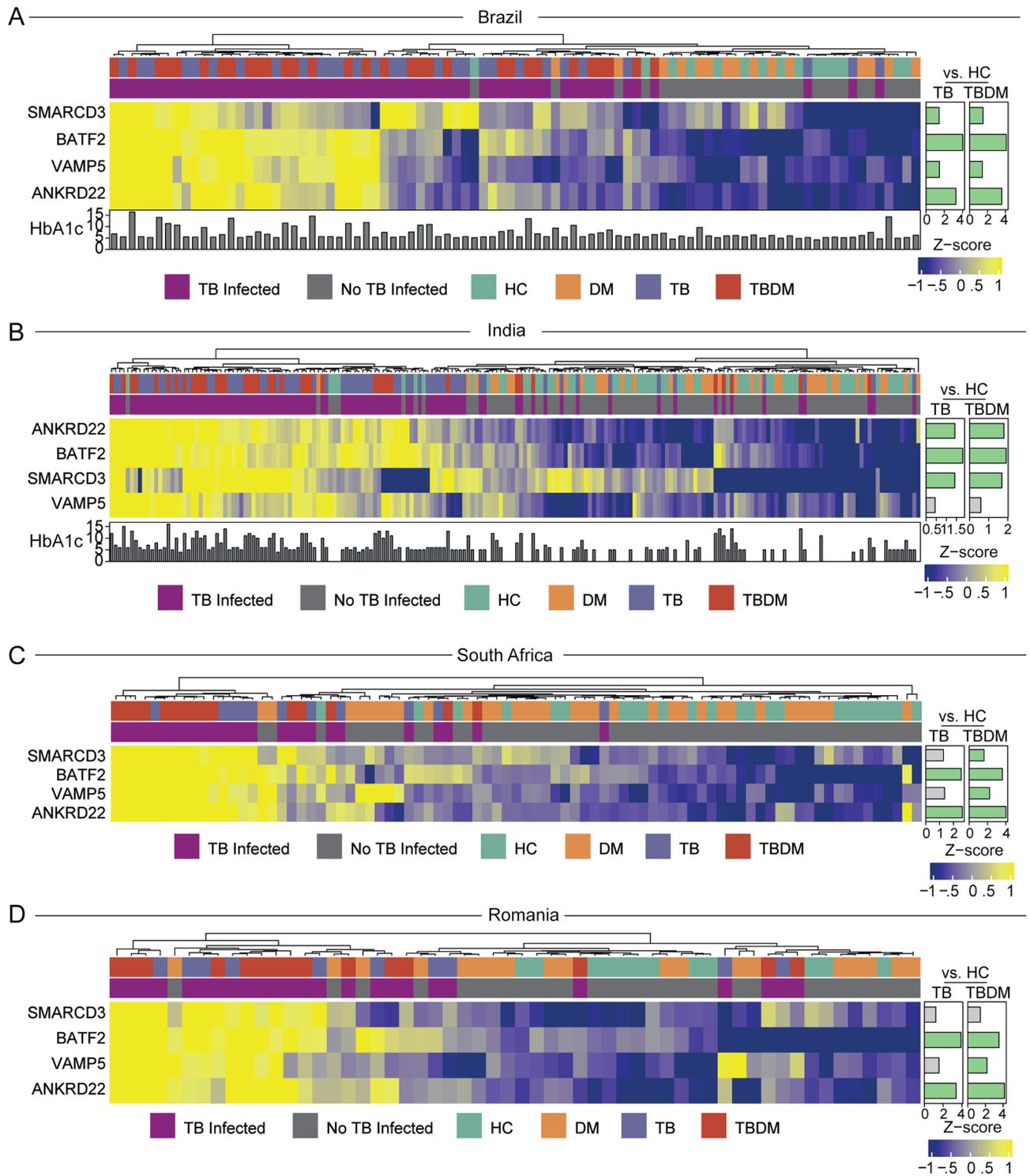


**Figure 2.** Molecular degree of perturbation stratified by clinical group and site. Histograms show molecular degree of perturbation (MDP) scores of each sample (left) and boxplots of each group (right) are shown for Brazil (A), India (B), South Africa (C) and Romania (D), using Kruskal–Wallis test. (E) Boxplots of MDP by of each clinical group (DM, TB, TBDM and HC) compared between sites. To estimate which groups are different from others, Tukey’s post test was employed. Thus, \*refers to  $p \leq 0.05$ , \*\*refers to  $p \leq 0.01$ , \*\*\*refers to  $p \leq 0.001$  and \*\*\*\*refers to  $p \leq 0.0001$ .

fied in many TB gene expression studies<sup>12</sup> was enriched among TB participants from Brazil, South Africa, and Romania (Fig. 6A). The India and Brazil TBDM groups shared enrichment of the Neutrophil.

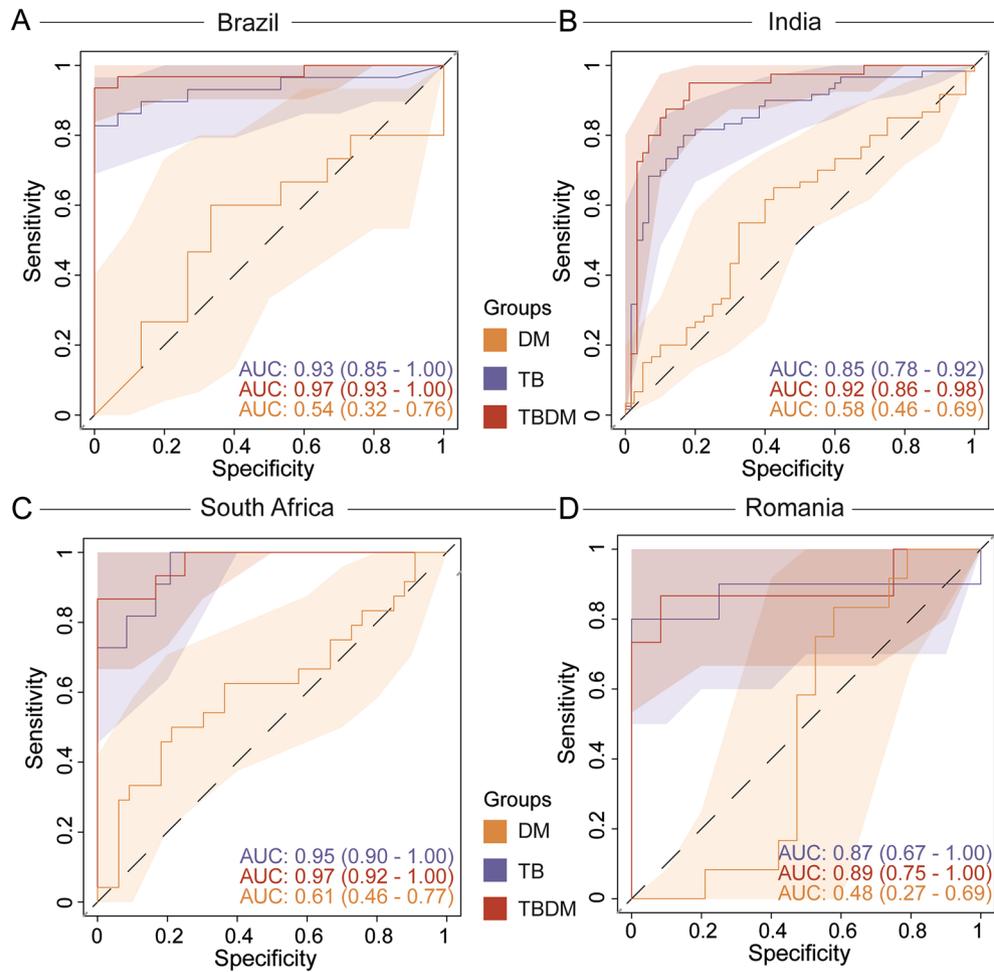
Degranulation, Antimicrobial Peptides, and Extracellular Matrix Organization pathways. Interferon signaling was enriched only in the Romanian TBDM group but not in TBDM participants from the other sites (Fig. 6B). Mirroring the DEG analysis, the changes in pathway predominance during TB treatment were markedly different between sites (Supplementary Fig. S7). Although pathway enrichment did not consistently distinguish TBDM from TB across all sites, the trend for increased Neutrophil Degranulation, Antimicrobial Peptides, and Matrix Organization fits with prior reports of neutrophilic inflammation and elevated circulating levels of cathelicidin, human beta defensin-2, human neutrophil peptides 1–3, and matrix metalloproteinases in TB-DM comorbidity<sup>7,13,14</sup>.

The Neutrophil Degranulation, Antimicrobial Peptides, Interferon Signaling, Regulation of Complement Cascade, Complement Cascade, and Interferon alpha/beta Signaling pathways were evaluated using a hierarchical

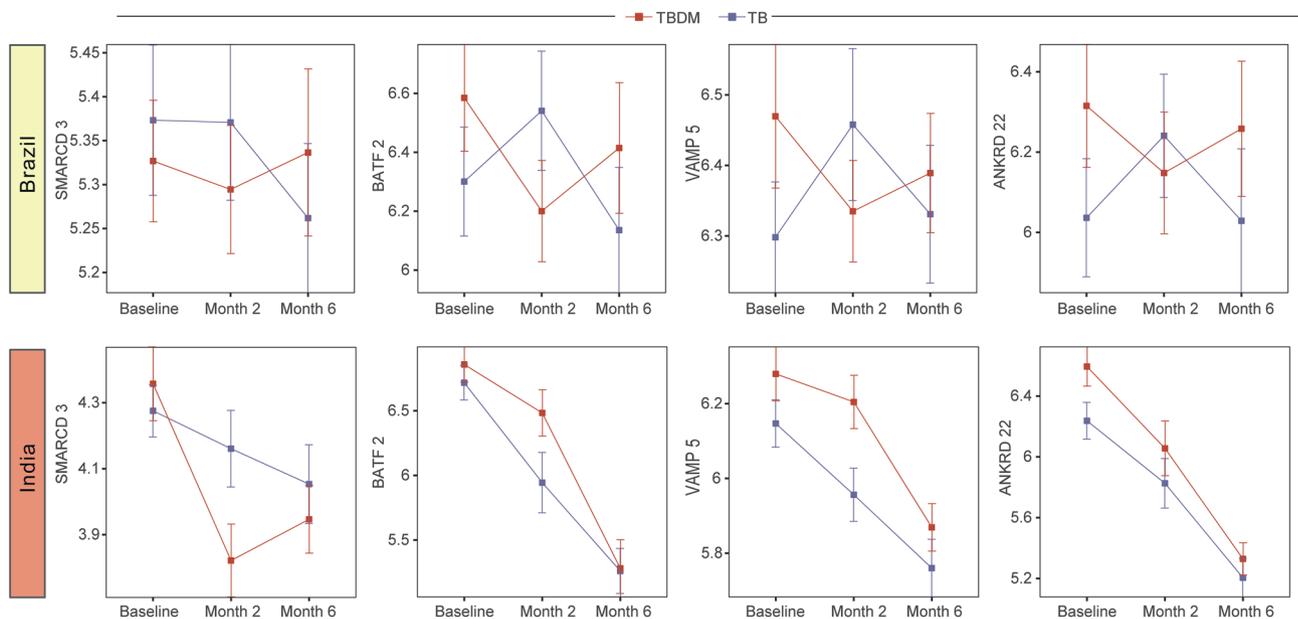


**Figure 3.** Relative expression of four TB signature genes at each site. Using a machine learning analysis, we defined four top genes among the clinical sites (as showed in Supplementary Material). A z-score normalized heatmap was employed to depict overall trends in gene expression among the clinical groups each study site, as indicated. Panels to the right of heatmaps show the average fold-difference between the signature gene expression in the HC group versus TB and TBDM (log-transformed values).

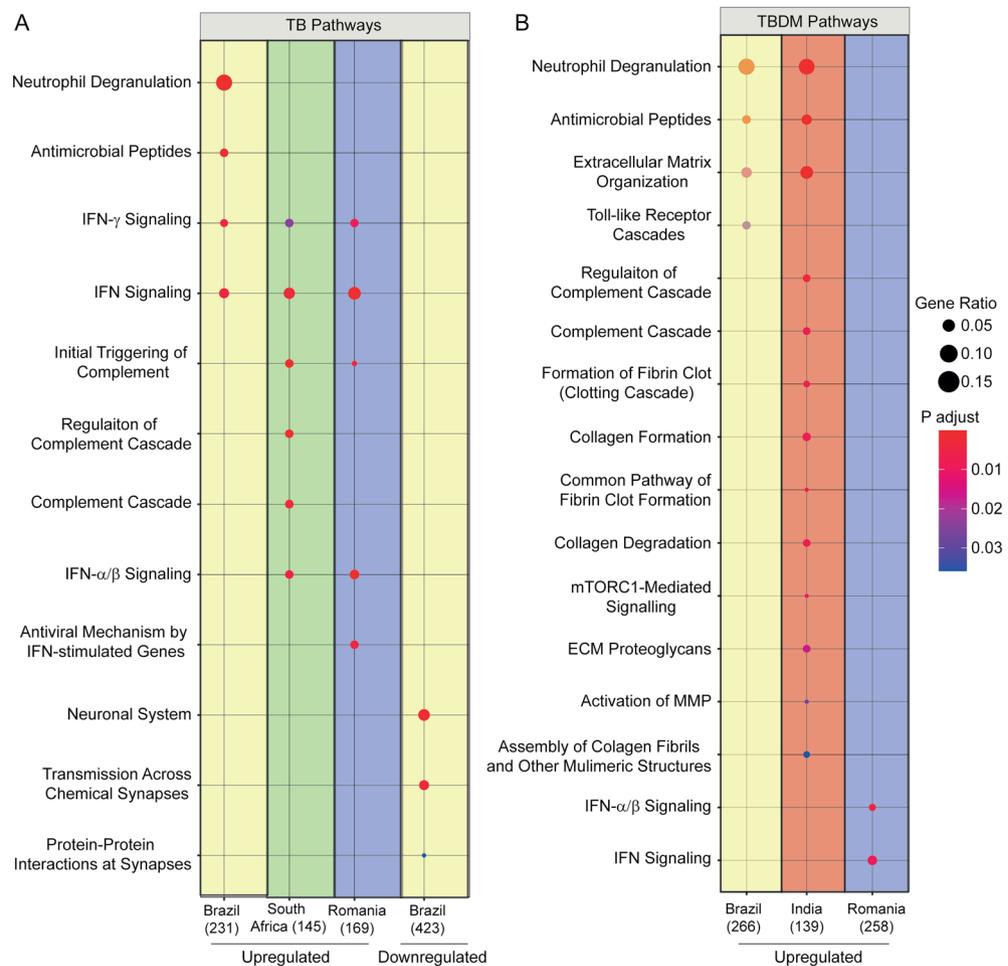
cluster analysis ordinating participants by MDP levels in TB and TBDM groups (Fig. 7A,B). This analysis was limited by the absence of some pathway enrichment data from South Africa and Romania, thus box plots from only Brazil and India are displayed. We observed a higher expression of the Interferon Signaling and Neutrophil Degranulation pathways in Brazilian TB participants when compared with the TB group from India (Fig. 7A).



**Figure 4.** Accuracy of top the 4-gene signature to classify TB, with or without DM. Receiver operator curve (ROC) analysis was used to check the accuracy of the signature genes identified by the random forest model to classify the TB, TBDM, and DM groups in each clinical site as indicated with respect to TB disease.



**Figure 5.** A distinctive pattern of gene expression during TB treatment. Expression of the four TB signature genes was analyzed at baseline, month-2 and month-6 in MSTDI participants in India and Brazil. Nemenyi's non-parametric all-pairs comparison was used.

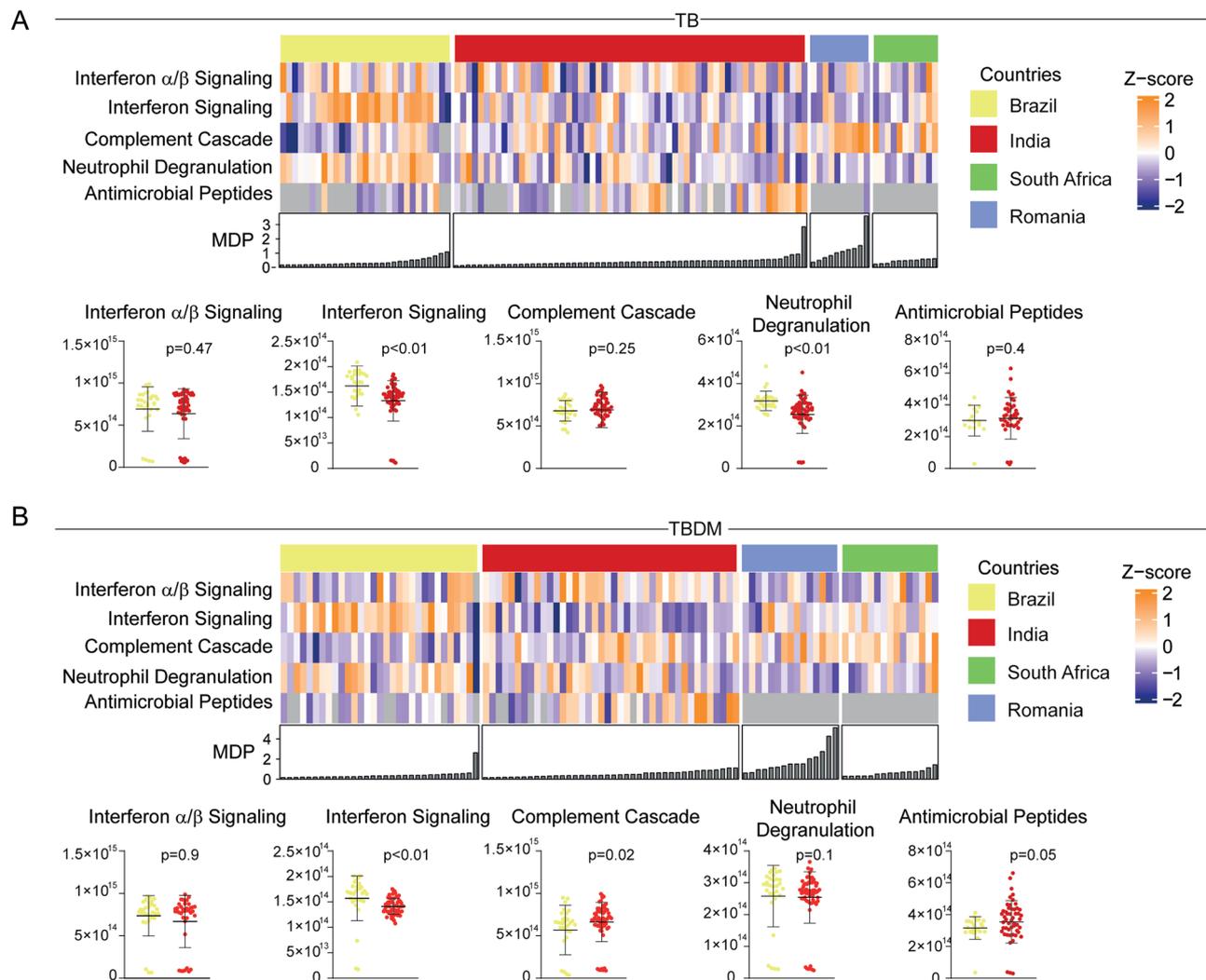


**Figure 6.** Pathway enrichment by condition and site. The colored spots indicate selected upregulated and downregulated Reactome pathways of interest identified by the combined significant ( $p < 0.05$ ) and FDR-corrected DEGs from the TB groups (A) and TBDM groups (B) in Brazil (yellow), India (red), South Africa (green), and Romania (blue). The hue of spots corresponds to p-value and the area corresponds to gene ratio. The method did not detect any pathways in TB participants from India.

Among TBDM groups, the Interferon Signaling pathway was more highly expressed in Brazil while the Complement Cascade pathway was more highly expressed in India (Fig. 7B).

To further evaluate the interaction between DM and TB, we examined the interactions between pathways identified from DEGs compared to Reactome in the TBDM groups (Fig. 8) and TB groups (Supplementary Fig. S8) which were observed at least two sites. This approach differed from our initial pathway analysis by estimating enrichment scores computed in each sample instead of a set of DEGs, enabling further analyses such as correlation or fold change score. Correlation network analysis and fold-change analysis using single sample gene set enrichment analysis (ssGSEA) to generate normalized enrichment scores (NES) identified differences across sites and clinical groups, along with highly correlated pathways at some sites that were not identified by the prior enrichment analysis (Fig. 6), which was not suitable to investigate correlation. As shown in Fig. 8, higher density of network pathways in TBDM was found in South Africa (0.108), India (0.091), and Brazil (0.069). Network pathway density was markedly lower in the Romanian TBDM group (0.013) despite a higher number of vertices found in these correlations (Fig. 8). Excluding the Romanian cohort, the top correlated pathway in TBDM was Complement Cascade, with 22 connections in Brazil, 30 correlations in India and 25 in South Africa, however, the pathway was not uniformly up or downregulated in TBDM at all the three sites compared with the site-specific healthy control group. Correlation analysis for the normoglycemic TB groups using ssGSEA-NES values showed comparable diversity between sites (Supplementary Fig. S8).

**Correlation of ssGSEA pathways with HbA1c.** Finally, to assess the influence of average blood glucose levels with pathway engagement, we employed a correlation model using HbA1c levels and ssGSEA NES values with available data from Brazil and India (Fig. 9). Among TB group participants, the India cohort data showed a substantially higher number correlations (68 positive and 6 negative) than Brazil (9 positive and only one negative) (Fig. 9). The TBDM comparison showed a lower total number of correlations than was seen for TB, again

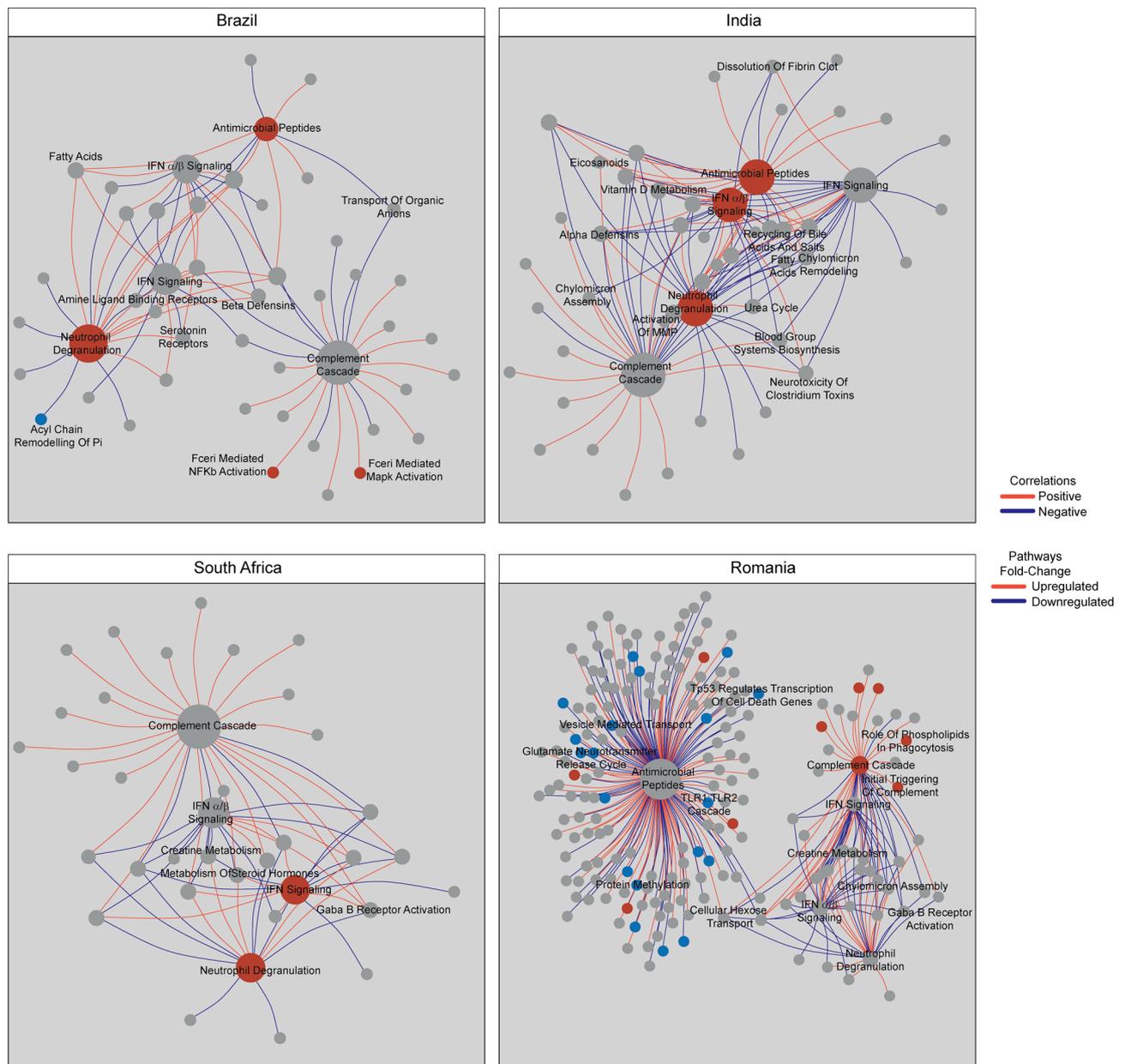


**Figure 7.** Changes in the pathways expression across clinical sites and diseases. A heatmap Z-score normalized was employed using the overlapped pathways identified in the enrichment analysis from DEGs (as described in Fig. 3). Box plots represents median and interquartile ranges, Mann Whitney test was used.

with more in the Indian cohort (7 positive and 7 negative correlations) compared to Brazil (6 positive and 2 negative correlations) (Fig. 9). There was no overlap in these correlated pathways between the sites or conditions. Notably, many of the positive correlations from the India TB group were for pathways associated with insulin resistance and metabolic syndrome (e.g. Passive Transport by Aquaporins, Yap1 and WWTR1 Taz Stimulated Gene Expression, Complex I Biogenesis, Mitochondrial Translation), diabetic complications (Axon Guidance, Regulation of Kit Signaling, Kinesins, MAPK Family Signaling Cascades, Mitochondrial Translation, VEGF Signaling, Factors Involved in Megakaryocyte Development and Platelet Production, Unblocking of NMDA Receptors Glutamate Binding and Activation), and pathways associated with chromosomal instability (Mitotic Prometaphase, Centrosome Maturation, G2 M DNA Damage, M Phase, Chromatin Organization, Mitotic Metaphase and Anaphase). Correlation with immune and inflammatory pathways were less prominent but included positive correlations with the RORA Activates Gene Expression and the Innate Immune System pathways in the India TB group, and negative correlations with B Cells Memory and Potassium Channels Pathways. The universally high level of HbA1c prevented correlation of these pathways in TBDM individuals.

## Discussion

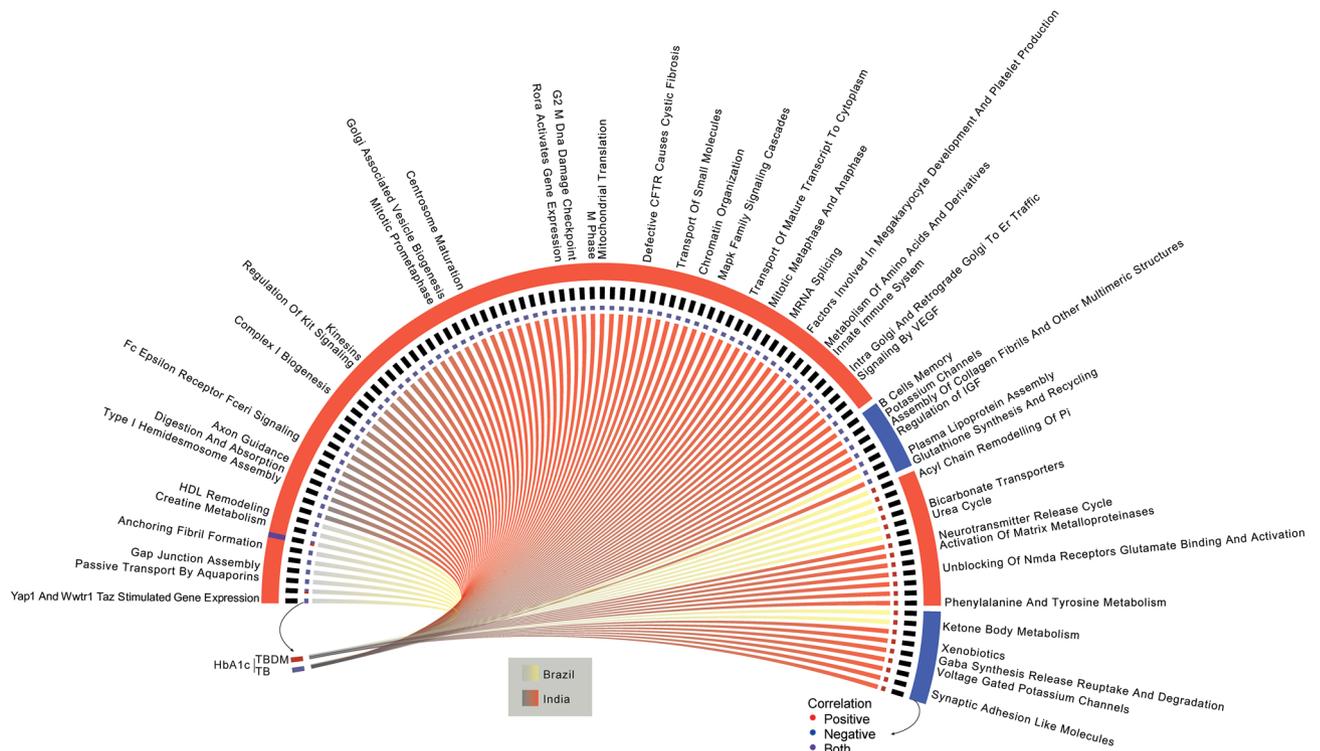
Whole blood RNAseq potentially offers a window into biological processes at the primary site of TB disease in the lung. In an earlier blood gene expression study at Chennai with a different group of participants, we found no differences in immune response pathway activation between TB and TBDM<sup>7</sup>. To address shortcomings of that study done at one site and using microarrays, the MSTDI study included participants from Chennai and Pune in India and Salvador in Brazil and used RNAseq performed on the same platform. Results of the MSTDI study supported our prior findings. There was no evident signature pattern of immune response gene expression or pathway activation that offered mechanistic insight to the basis of TB susceptibility in people living with DM. However, trends for increased neutrophil and innate immune pathway activation in TBDM were noted. Studies



**Figure 8.** Pathway dynamicity across the sites and disease groups. A Spearman correlation analysis was performed using the pathways from TBDM participants in each clinical site, as indicated. Each node indicates a pathway, blue nodes indicate downregulation compared to the site-specific HC group, while red nodes indicate upregulation. Grey nodes represent pathways that were not statistically significant compared to the site-specific healthy control group. Red lines infer positive correlation and blue lines negative interaction.

in hyperglycemic mice identified a defect in alveolar macrophage sentinel function as the key immune mechanism of TB susceptibility in DM<sup>15–17</sup>. This impacts the initial encounter of alveolar macrophages with inhaled *M. tuberculosis*, delaying the transition from innate to adaptive immunity during the period of logarithmic increase in lung bacterial load. Once the adaptive immune response is expressed in diabetic mice, it effectively limits *M. tuberculosis* replication and is qualitatively indistinguishable from the response of normoglycemic control mice with TB. If a similar mechanism operates in human DM, then the key immunological events contributing to susceptibility will have occurred months before TB diagnosis. Gene expression at baseline in human studies corresponds to the later time points of mouse experiments when the cell-mediated immune response is fully activated.

Based on animal model data, we anticipated that immune pathology in human TBDM would be qualitatively similar to euglycemic TB but quantitatively more severe<sup>18</sup>. That prediction was supported by plasma cytokine and radiographic studies<sup>7,19</sup>, but not so clearly by differences in whole blood gene expression<sup>7</sup>. The TANDEM study included four geographic sites and reported an overall increase of innate immune gene expression with decreased expression of genes linked to adaptive immunity in participants with TB combined with DM or pre-DM<sup>20</sup>. That result supports the general notion of increased immune pathology in TBDM but does not explain the broadly



**Figure 9.** Integrative analysis of HbA1c and Pathways expression in Brazil and India. Levels of HbA1c were correlated with the pathways according to disease group as indicated in the left box (red refers to TBDM whereas purple to TB). The external-colored bars indicated the direction of correlation, as indicated in right circles (red for positive correlation, blue for negative correlation, and purple for positive and negative correlations). Lines in center of the figure refers to the country (red for India and yellow for Brazil).

elevated plasma levels of Th1, Th2, and Th17 cytokines reported in this condition<sup>7,21</sup>. The MSTDI data presented here showed a non-significant trend for higher MDP values in TBDM than TB participants, whereas the differences in MDP between sites were statistically significant. Pathway enrichment, interaction, and correlation analyses likewise failed to identify a universally consistent TBDM-specific pattern of engagement. Overall, the data suggest greater severity of immune pathology in TBDM, which more likely reflects a consequence rather than cause of TB susceptibility in people living with DM.

Results from the MSTDI study, notably the correlation of pathway engagement with HbA1c levels, further support the potential for TB to exacerbate non-communicable disease processes<sup>7,9</sup>. Bacterial pneumonia and COVID-19 are both associated with increased risk of cardiovascular disease events<sup>22,23</sup>. A similar association has been made with TB<sup>24</sup> where the longer duration of inflammation, particularly with comorbid DM<sup>11</sup>, could be particularly damaging. Similar reasoning would apply to the microvascular complications of DM, but this has not been studied.

An unexpected finding in our analyses was the high degree of variability in baseline gene expression between all four populations evaluated. Of note, some differences in the clinical presentation and disease severity were found in our study, as in TB, with a higher percentage of cavitations in India when compared with Brazilian TBDM participants, as well in DM, with higher HbA1c levels detected in India. The heterogeneity in this proportion could affect the gene expression and to verify its influence we performed an additional analysis that demonstrated an apparent lack of association between such parameters and differences in gene expression profiles among the countries. Equally remarkable were the different temporal patterns of gene expression in the India and Brazil cohorts from baseline through TB treatment completion at month 6. This variability presumably reflects differences in host and microbe genetics, behaviors, and environmental exposures between sites. Similar issues have hampered the application of concise gene expression signatures for TB diagnosis and treatment response at different sites<sup>25</sup>. Further research will be required to identify the fundamental processes influencing the TB-DM interaction and their expression in diverse populations.

This work has some limitations. First, not all clinical and epidemiological data are available from all datasets used. The data from South Africa and Romania did not present other information such as HbA1c levels, Cavitation or alcohol, and smoking usage. For this reason, only age and sex were used to adjust the models in differential expression analysis. The second limitation was the differences observed in the BMI values, age, sex, smoking, and alcohol use between the population of Brazil and India (Supplementary Tables S1 and S2). Additionally, some patients were under treatment with metformin and statin, that could affect the inflammatory responses. We have applied a negative binomial model for the differential gene expression analysis. This modeling has limitations regarding multiple variable adjustments. Despite we have evaluated the Cavitory TB and BMI association with

gene expression, their systemic influence could not be fully corrected in the model. Moreover, use of metformin and/or statins, and frequency of cavitary TB were significantly different in TB-infected individuals between those countries. Cavitary TB and BMI values are known to affect immune activation and thus could influence the gene expression profile of those subjects and may explain the absence of overlapping pathways in Brazil and India sites. Another limitation was associated with the different sample sizes observed in the cohorts (Supplementary Fig. S1). South Africa data presented 11 TB and 15 TBDM samples, Romania 10 TB and 15 TBDM samples, Brazil 29 TB and 31 TBDM samples, and India 60 TB and 40 TBDM samples. This unbalanced sample size could have influenced both gene expression analysis and signature performance. DM is a metabolic disease, and it is reasonable to speculate that using other platforms, such as metabolomics, could increase the odds of detecting discrepancies between TB and TBDM. We are currently performing new investigations using multi-omics to portray the TB-DM interaction.

This work has some differences compared with the TANDEM work which makes the straightforward comparison of the results difficult. First, the TANDEM has used data from 4 different sites (South Africa, Romania, Peru, and Indonesia). However, all the analysis was performed with whole data, without performing the comparisons within the sites, as was performed in the MSTDI. Also, two TANDEM sites do not have healthy controls, thus all comparisons from the case subjects from Peru and Indonesia were performed with foreign control, which may have inserted variation. Second, the enrichment analysis was performed using different approaches. In TANDEM, the transcriptional module enrichment analysis was performed with tmod package while two different approaches were performed in the MSTDI, using clusterProfiles and Reactome database, and the single sample gene set enrichment analysis with ssGSEA package. This shows that the cohorts explored the disease dynamics in a different optic, with the first exploring the overall impact and the second one investigating the populational contribution on this impact.

In summary, we found substantial variability in whole blood gene expression between HC, DM, TB and TBDM participants across the two MSTDI study sites in Brazil and India and the TANDEM study South African and Romanian cohorts. No mechanistically informative signature of immune pathway gene expression distinguished TBDM vs TB in all populations, although increased innate immune and vascular complication pathway activation was common across some sites. Our findings lend evidence in support of adjunctive anti-inflammatory and antioxidant therapies during TB treatment. In that regard, retrospective evidence demonstrated that metformin added to antibiotic treatment reduces the mortality risk in TBDM independent of glycemic control<sup>26</sup> and a recently completed randomized controlled trial of TBDM patients in India showed that metformin reduced inflammation and radiographic severity in disease<sup>27</sup>.

## Materials and methods

**Ethics statement and study population.** The MSTDI study used unpublished data and samples from the RePORT India and the RePORT Brazil consortia. Samples from the RePORT India were enrolled under protocols approved by the Ethics Committee of the Prof. M. Viswanathan Diabetes Research Center and the Institutional Review Boards of Byramjee Jeejeebhoy Government Medical College, Pune and National Institute for Research in Tuberculosis and Johns Hopkins University. The samples enrolled from the RePORT Brazil had their protocols approved by the institutional review boards of the Instituto Gonçalo Moniz, Fundação Oswaldo Cruz and Vanderbilt University Medical Center. Written informed consent was obtained from all participants prospectively enrolled at two sites of the RePORT India and one site of the RePORT Brazil consortia, with organizational support from RePORT International<sup>28</sup>. The study was conducted according to the principles of the Declaration of Helsinki. The Indian sites were in Chennai (EDOTS study)<sup>29</sup> and Pune (CTRIUMPH study<sup>30</sup>), while the Brazilian site was in Salvador (RePORT International Common Protocol)<sup>28</sup>. The combined MSTDI cohort of 290 individuals comprised 120 participants from Chennai, 80 from Pune and 90 from Salvador. These participants were selected from a larger cohort of the RePORT protocols based of sample availability. Participant groups included pulmonary TB disease with or without DM (TBDM and TB groups, respectively) and two control groups without TB, with or without DM (DM and HC groups, respectively). Inclusion criteria were age 18–65 and new diagnosis of pulmonary TB (or absence of pulmonary TB for the control group participants). Drug-resistant TB, retreatment TB, treatment of incident TB for > 7 days prior to enrollment, pregnancy, immunosuppressive medications, and HIV infection were exclusions. All TB data used here were obtained from participants accompanied during the 6 months of anti-TB therapy according to the RePORT common protocol and who were successfully treated according to WHO criteria<sup>31</sup>. Participant characteristics are presented in Supplementary Tables S1 and S2 of the Supplementary Material. Secondary data were used from the TANDEM study<sup>32</sup> that explored the TB-DM interactions at clinical sites in Indonesia, Peru, Romania, and South Africa. Patients from TANDEM were also on TB treatment according to the respective local TB program. We used TANDEM data from South Africa and Romania where site-specific healthy control participant data were available. This data set comprised 83 participants from South Africa (24 HC, 33 DM, 11 TB, 15 TBDM) 56 from Romania (12 HC, 19 DM, 10 TB, 15 TBDM). The group sizes for all sites used in our analysis are summarized in Supplementary Fig. S1.

**Classification and blood sampling.** TB diagnosis was based on positive sputum culture for *M. tuberculosis* with a compatible chest x-ray at enrollment, while negative culture and x-ray defined the control groups. Classification with DM was based on self-reported medical and anti-diabetic medication history or glycohemoglobin (HbA1c)  $\geq 6.5\%$ . Classification as euglycemic was based on self-reported medical history and HbA1c < 5.7% or 75-g oral glucose tolerance test 2-h blood glucose < 140 mg/dL. Baseline blood samples were collected in RNA storage tubes from all participants at the time of enrollment and no later than 7 days after the initiation of anti-tubercular treatment.

**Library preparation and RNA sequencing.** The samples from TB-infected individuals from both India and Brazil sites were collected at baseline, 2, and 6 months of treatment. Samples for the HC and DM groups were collected at baseline. Whole blood (5 mL) was collected PAXgene Blood RNA tubes (Qiagen, catalog #762165) and frozen at  $-80^{\circ}\text{C}$ . RNA was extracted using the PAXgene Blood RNA kit (Qiagen, catalog #762174) and quantified using Qubit RNA assay HS (Invitrogen, Cat #Q32852). RNA purity was checked using QIAxpress, and RNA integrity was assessed on TapeStation using RNA HS ScreenTapes (Agilent, Cat #5067-5579). NEB Ultra II Directional RNA-Seq Library Prep kit protocol was used to prepare libraries for total RNA sequencing. Prepared libraries were quantified using Qubit High Sensitivity Assay (Invitrogen, Cat #Q32852), pooled and diluted to final optimal loading concentration before cluster amplification on Illumina flow cell. Once the cluster generation was completed, the cluster flow cell was loaded on Illumina HiSeq 2500 instrument to generate paired end reads at MedGenome in Bangalore, India.

**RNA-seq data analysis.** Raw RNA-seq data from the MSTDI cohort were retrieved from Illumina HiSeq 2500 platform. Sequence data from the TANDEM cohort<sup>8</sup> was retrieved from the SRA database using BioProject PRJNA470512 using the SRA tools. Sequence data from MSTDI and TANDEM were quality control processed by removing low-quality bases and adapters using *Trimmomatic V0.32*. After the quality check, sequences were pseudo-aligned against the human transcriptome (GRCh38 version) comprising both mRNA and miRNA with *salmon v0.8.2*<sup>33</sup> and presented a mean mapping rate of  $68.03 \pm 2.28$ . After mapping, the output was converted to a count table using *tximport* package<sup>34</sup> from *R 4.1.3*. Count gene expression matrix was examined using the *edgeR* package<sup>35</sup> from *R 4.1.3* to identify differentially expressed genes (DEGs). MSTDI gene expression data are available at the GEO database (Accession number GSE181143, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE181143>).

Differential expression analysis in the MSTDI cohort compared three conditions (TB, TBDM, DM) at two sites (Brazil and India) to the healthy control (HC) group at the respective site to determine the fold change and *p*-value of each gene. For instance, the TB, TBDM, and DM subjects from one country are compared with the HC subjects from the same country, and not from another country, to avoid inserting operational variation in the analysis. From the TANDEM cohort, only two sites (South Africa and Romania) were used in the differential expression analysis. The other sites (Peru and Indonesia) lack the HC group at the respective sites. Multiple testing correction was performed using the Benjamini & Hochberg false discovery rate (FDR) method<sup>36</sup>. Changes in gene expression were considered significant when corrected *p*-values remained  $< 0.05$  after FDR adjustment and if the fold change differences were higher than  $\pm 1.4$ . The complete list of genes is available in Supplementary File 1. The DEGs were visualized using the *VennDiagram* package<sup>37</sup> from *R 4.1.3*. The *compareCluster* package<sup>38</sup> from *R 4.1.3* was used with the obtained DEGs to scan the *REACTOME* pathway database<sup>39</sup> to perform the pathway enrichment analyses and the Benjamini & Hochberg false discovery rate (FDR) method<sup>36</sup> was used to correct the *p*-values for multiple testing.

Participant characteristics tables present the median and interquartile range or percentage for the nominal variables. All comparisons were performed with Kruskal–Wallis with Tukey's post test and Chi-squares tests using *R 4.1.3*.

**Population heterogeneity evaluation and feature selection analysis.** Sample variation within and between sites was evaluated using the molecular degree of perturbation (MDP) package<sup>10</sup> applied in the gene expression values after variance-stabilizing transformation (VST), for Brazil, South Africa and Romania data, and batch effect correction with *sva* package<sup>40</sup> in the case of India site data. A gene was classified as perturbed when its variation compared to HC was  $> 2$  standard deviations.

To evaluate the sample clustering and classification across the sites we performed one-sided unsupervised hierarchical clustering Ward's method<sup>41</sup>, Heatmaps<sup>42</sup> and the Principal component analysis (PCA) plot in the VST gene expression values from each cohort (Supplementary Figs. S2, S3A,B). This approach allows the visualization of sample dispersion across the groups in the sites. To maximize the performance in the TB and TBDM classification, we employed a dimensionality reduction approach to reduce the number of genes associated with TB. Thus, we retrieved the VST gene expression of all 3427 DEGs from the comparisons. The data from South Africa and Romania were used as a discovery set in a random forest algorithm with leave-one-out cross-validation with the *caret*<sup>43</sup> and *randomForest*<sup>44</sup> packages. The data from Brazil and India was used as a validation set (Supplementary Fig. S4). The minimal gene set exhibiting higher classification power to describe the groups was defined by the variable importance in the random forest model. Gene expression values from each were retrieved from each site and the Receiver Operator Characteristics (ROC)<sup>45</sup> were used to assess the accuracy of the gene set to distinguish between comparison groups specified in the TANDEM and MSTDI datasets.

**Performance analysis with previously identified signatures.** We conducted a performance comparison using 69 previously published gene expression signatures for TB diagnosis, progression, and treatment provided by the *TBSignatureProfiler* package (<https://github.com/compbioMed/TBSignatureProfiler>). In addition, we have included RISK6, RISK11 and BATF2 signatures for comparison (Supplementary Figs. S5 and S6). The signature classification performance was performed using data from each signature data from TB and HC (Supplementary Fig. S5), and TBDM and HC (Supplementary Fig. S5) within each country, similar to the analysis performed to identify the DEGs. We estimated the area under the curve (AUC) values of each signature with its confidence interval (CI), by applying a general linear model to gene expression values from each comparison (TB and HC or TBDM vs HC, within the countries). The detailed performance of each signature in each clinical conditions is shown in the Supplementary File S2. Moreover, linear modeling allows the comparison of performance from signatures composed by DEGs and composed by scores, as well as a fair comparison between

them<sup>46,47</sup>. The outcomes were binarized to measure the sensitivity and specificity of classification, allowing us to measure each group rate and plot each signature AUC and CI value, from each country. This allows the direct performance comparison of each signature in data from different countries and different conditions (TB or TBDM).

**Single sample gene set enrichment analysis (ssGSEA).** The normalized enrichment scores (NES) from each sample were calculated with ssGSEA<sup>48</sup> using the Reactome database<sup>39</sup>. Only significant NES values were used (FDR < 0.05 and 100 permutations) to perform correlation analysis with the enriched pathways, HbA1c values and gene expression levels. Correlation relationships were depicted as chord diagrams and networks, performed by *circlize*<sup>49</sup> and *igraph*<sup>50</sup> package.

### Data availability

The dataset from the TANDEM cohort analyzed during the current study is available at the BioProject data repository, identified by the accession code PRJNA470512 (<https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA470512>). The MSTDI gene expression data is available at the geoNCBI data repository, identified by the accession number GSE181143 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE181143>).

Received: 27 February 2023; Accepted: 9 May 2023

Published online: 12 May 2023

### References

- Critchley, J. A. *et al.* Defining a research agenda to address the converging epidemics of tuberculosis and diabetes: Part 1: Epidemiology and clinical management. *Chest* **152**, 165–173 (2017).
- Ronacher, K. *et al.* Defining a research agenda to address the converging epidemics of tuberculosis and diabetes: Part 2: Underlying biologic mechanisms. *Chest* **152**, 174–180 (2017).
- Jeon, C. Y. & Murray, M. B. Diabetes mellitus increases the risk of active tuberculosis: A systematic review of 13 observational studies. *PLoS Med.* **5**, e152 (2008).
- Odone, A., Houben, R. M., White, R. G. & Lonnroth, K. The effect of diabetes and undernutrition trends on reaching 2035 global tuberculosis targets. *Lancet Diab. Endocrinol.* **2**, 754–764 (2014).
- Martinez, N. & Kornfeld, H. Diabetes and immunity to tuberculosis. *Eur. J. Immunol.* **44**, 617–626 (2014).
- Blankley, S. *et al.* The application of transcriptional blood signatures to enhance our understanding of the host response to infection: the example of tuberculosis. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **369**, 20130427 (2014).
- Prada-Medina, C. A. *et al.* Systems immunology of diabetes-tuberculosis comorbidity reveals signatures of disease complications. *Sci. Rep.* **7**, 1999 (2017).
- Eckold, C. *et al.* Impact of intermediate hyperglycemia and diabetes on immune dysfunction in tuberculosis. *Clin. Infect. Dis.* **72**, 69–78 (2021).
- Magee, M. J. *et al.* Convergence of non-communicable diseases and tuberculosis: A two-way street?. *Int. J. Tuberc. Lung Dis.* **22**, 1258–1268 (2018).
- Oliveira-de-Souza, D. *et al.* Molecular degree of perturbation of plasma inflammatory markers associated with tuberculosis reveals distinct disease profiles between Indian and Chinese populations. *Sci. Rep.* **9**, 8002 (2019).
- Kumar, N. P. *et al.* Persistent inflammation during anti-tuberculosis treatment with diabetes comorbidity. *eLife* <https://doi.org/10.7554/eLife.46477> (2019).
- Singhania, A., Wilkinson, R. J., Rodrigue, M., Haldar, P. & O'Garra, A. The value of transcriptomics in advancing knowledge of the immune response and diagnosis in tuberculosis. *Nat. Immunol.* **19**, 1159–1168 (2018).
- Kumar, N. P. *et al.* Heightened circulating levels of antimicrobial peptides in tuberculosis-diabetes co-morbidity and reversal upon treatment. *PLoS One* **12**, e0184753 (2017).
- Kumar, N. P. *et al.* Association of plasma matrix metalloproteinase and tissue inhibitors of matrix metalloproteinase levels with adverse treatment outcomes among patients with pulmonary tuberculosis. *JAMA Netw. Open* **3**, e2027754 (2020).
- Martens, G. W. *et al.* Tuberculosis susceptibility of diabetic mice. *Am. J. Respir. Cell Mol. Biol.* **37**, 518–524 (2007).
- Vallerskog, T., Martens, G. W. & Kornfeld, H. Diabetic mice display a delayed adaptive immune response to *Mycobacterium tuberculosis*. *J. Immunol.* **184**, 6275–6282 (2010).
- Martinez, N., Ketheesan, N., West, K., Vallerskog, T. & Kornfeld, H. Impaired recognition of *Mycobacterium tuberculosis* by alveolar macrophages from diabetic mice. *J. Infect. Dis.* **214**, 1629–1637 (2016).
- Martinez, N. & Kornfeld, H. Diabetes and immunity to tuberculosis. *Eur. J. Immunol.* **44**, 617–626 (2014).
- Barreda, N. N. *et al.* Severe pulmonary radiological manifestations are associated with a distinct biochemical profile in blood of tuberculosis patients with dysglycemia. *BMC Infect. Dis.* **20**, 139 (2020).
- Eckold, C. *et al.* Impact of intermediate hyperglycaemia as well as diabetes on immune dysfunction in tuberculosis. *Clin. Infect. Dis.* **72**, 69–78 (2020).
- Kumar, N. P. *et al.* Type 2 diabetes mellitus coincident with pulmonary tuberculosis is associated with heightened systemic type 1, type 17 and other pro-inflammatory cytokines. *Ann. Am. Thorac. Soc.* **10**, 441–449 (2013).
- Corrales-Medina, V. F. *et al.* Association between hospitalization for pneumonia and subsequent risk of cardiovascular disease. *JAMA* **313**, 264–274 (2015).
- Xie, Y., Xu, E., Bowe, B. & Al-Aly, Z. Long-term cardiovascular outcomes of COVID-19. *Nat. Med.* **28**, 583–590 (2022).
- Basham, C. A., Smith, S. J., Romanowski, K. & Johnston, J. C. Cardiovascular morbidity and mortality among persons diagnosed with tuberculosis: A systematic review and meta-analysis. *PLoS One* **15**, e0235821 (2020).
- Penn-Nicholson, A. *et al.* RISK6, a 6-gene transcriptomic signature of TB disease risk, diagnosis and treatment response. *Sci. Rep.* **10**, 8629 (2020).
- Degner, N. R., Wang, J. Y., Golub, J. E. & Karakousis, P. C. Metformin use reverses the increased mortality associated with diabetes mellitus during tuberculosis treatment. *Clin. Infect. Dis.* **66**, 198–205 (2018).
- Padmapriyadarsini, C. *et al.* Randomized trial of metformin with anti-tuberculosis drugs for early sputum conversion in adults with pulmonary tuberculosis. *Clin. Infect. Dis.* <https://doi.org/10.1093/cid/ciab964> (2021).
- Hamilton, C. D. *et al.* RePORT international: Advancing tuberculosis biomarker research through global collaboration. *Clin. Infect. Dis.* **61**(Suppl 3), S155–159 (2015).
- Kornfeld, H. *et al.* 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 (2016).

30. Gupte, A. *et al.* Cohort for tuberculosis research by the Indo-US medical partnership (CTRIUMPH): Protocol for a multicentric prospective observational study. *BMJ Open* **6**, e010542 (2016).
31. Linh, N. N. *et al.* World health organization treatment outcome definitions for tuberculosis: 2021 update. *Eur. Respir. J.* **58**, 2100804 (2021).
32. van Crevel, R. & Dockrell, H. M. TANDEM: Understanding diabetes and tuberculosis. *Lancet Diab. Endocrinol.* **2**, 270–272 (2014).
33. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
34. Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: Transcript-level estimates improve gene-level inferences. *F1000Research* **4**, 1521 (2015).
35. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
36. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc.* **57**, 289–300 (1995).
37. Chen, H. & Boutros, P. C. VennDiagram: A package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinform.* **12**, 35 (2011).
38. Guangchuang, Y., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
39. Yu, G. & He, Q. Y. ReactomePA: An R/Bioconductor package for reactome pathway analysis and visualization. *Mol. Biosyst.* **12**, 477–479 (2016).
40. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).
41. Ward, J. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* **58**, 236–244 (1963).
42. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
43. Kuhn, M. Building predictive models in R using the caret package. *J. Stat. Softw.* **28**, 1–26 (2008).
44. Breiman, L. Random forests. *Mach. Learn.* **45**, 2–32 (2001).
45. Zou, K. H., O'Malley, A. J. & Mauri, L. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation* **115**, 654–657 (2007).
46. Kulkarni, V. *et al.* A two-gene signature for tuberculosis diagnosis in persons with advanced HIV. *Front. Immunol.* **12**, 631165 (2021).
47. Mathad, J. S. *et al.* Transcriptional analysis for tuberculosis in pregnant women from the PRACHITi study. *Clin. Infect. Dis.* **75**, 2239–2242 (2022).
48. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).
49. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. Circlize implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811–2812 (2014).
50. Csardi, G. & Nepusz, T. The Igraph software package for complex network research. *Interf. Complex Syst.* **1695**, 1–9 (2006).

## Acknowledgements

We thank Ms. Daphne Martin and Ms. Samyra Cox for outstanding administrative support.

## Author contributions

K.F.F., A.T.Q.L., C.L.V., E.R.F. and M.B.A. performed the data curation, analysis, interpretation, and draft of the first version of the manuscript. A.N.G., S.G., R.K., V.M., M.P. and A.G. performed the data interpretation, revising manuscript critically for important intellectual content, final approval of the manuscript. N.P.K. performed sample preparation and curation for the Chennai cohort. T.R.S. supervised the Brazilian study and helped with data interpretation. S.B. and V.V. coordinated the clinical study in Chennai. B.B.A. performed the study conceptualization, data analysis and interpretation, and draft of the manuscript. H.K. performed the study conceptualization, data curation, analysis and interpretation, revising manuscript critically for important intellectual content and coordinated all sites studies.

## Funding

This work was funded by: OISE-17-63459-1 from the National Institutes of Health, administered by CRDF Global; DAA3-18-64718-1, formerly USB1-31149-XX-13 from the Indo-US Vaccine Action Initiative on TB Research, administered by CRDF Global. The Brazilian site was supported by the National Institutes of Health (NIH U01AI069923 and R01AI120790), CCASAnet, RePORT-Brazil Tennessee Center for AIDS Research (TNC-FAR). The study was also supported by the Intramural Research Program of the Fundação José Silveira, and the Intramural Research Program of the Oswaldo Cruz Foundation, Brazil. BBA is a senior investigator from the Brazilian Council for Science and Technology (CNPq) and MBA received research fellowship from the Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB).

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-34847-9>.

**Correspondence** and requests for materials should be addressed to B.B.A. or H.K.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

## the RePORT Brazil

Alice M. S. Andrade<sup>2,3,13</sup>, Marina C. Figueiredo<sup>7</sup>, Vanessa Nascimento<sup>16</sup>,  
 Juan Manuel Cubillos-Angulo<sup>3</sup>, Hayna Malta-Santos<sup>3</sup>, Jéssica Rebouças-Silva<sup>2,17</sup>,  
 Adriano Gomes-Silva<sup>18</sup>, Saulo R. N. Santos<sup>16</sup>, André Ramos<sup>16</sup>, Pedro Brito<sup>2,3</sup>,  
 Carolina A. S. Schmaltz<sup>5</sup>, Alysson G. Costa<sup>19,20,21</sup>, Leandro Sousa Garcia<sup>19,20</sup>,  
 Brenda K. de Sousa Carvalho<sup>19,20</sup>, Bruna P. de Loiola<sup>19,20</sup>, Francine P. Ignácio<sup>18</sup>,  
 Maria C. Lourenço<sup>18</sup>, Elisangela C. Silva<sup>19</sup>, Mayla Mello<sup>19</sup>, Alexandra B. Souza<sup>19,20</sup>,  
 Michael S. Rocha<sup>3,16</sup>, Aline Benjamin<sup>18</sup>, Adriana S. R. Moreira<sup>20</sup>, Jamile G. de Oliveira<sup>18</sup>,  
 Solange Cavalcante<sup>18</sup>, Betina Durovni<sup>18</sup>, Marcelo Cordeiro-Santos<sup>19,20</sup>, Afrânio L. Kristki<sup>22</sup>,  
 Valeria C. Rolla<sup>23</sup> & José R. Lapa-e-Silva<sup>24</sup>

<sup>16</sup>Instituto Brasileiro para Investigação da Tuberculose, Fundação José Silveira, Salvador, Brazil. <sup>17</sup>Faculdade de Medicina, Universidade Federal da Bahia, Salvador, Brazil. <sup>18</sup>Laboratório de Pesquisa Clínica em Micobacteriose, Instituto Nacional de Infectologia Evandro Chagas, Fiocruz, Rio de Janeiro, Brazil. <sup>19</sup>Fundação Medicina Tropical Dr Heitor Vieira Dourado, Manaus, Brazil. <sup>20</sup>Programa de Pós-Graduação em Medicina Tropical, Universidade do Estado do Amazonas, Manaus, Brazil. <sup>21</sup>Universidade Federal do Amazonas, Manaus, Brazil. <sup>22</sup>Programa Acadêmico de Tuberculose da Faculdade de Medicina, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil. <sup>23</sup>Instituto Nacional de Infectologia Evandro Chagas, Fiocruz, Rio de Janeiro, Brazil. <sup>24</sup>Faculdade de Medicina, Programa Acadêmico de Tuberculose, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil.

## RePORT India Consortia

Kim West<sup>15</sup>, Vandana Kulkarni<sup>9</sup> & Nikhil Gupte<sup>5</sup>