

# SCIENTIFIC REPORTS

OPEN

## Systems Immunology of Diabetes-Tuberculosis Comorbidity Reveals Signatures of Disease Complications

Cesar A. Prada-Medina<sup>1</sup>, Kiyoshi F. Fukutani<sup>2</sup>, Nathella Pavan Kumar<sup>3</sup>, Leonardo Gil-Santana<sup>4,5,6</sup>, Subash Babu<sup>3</sup>, Flávio Lichtenstein<sup>1</sup>, Kim West<sup>7</sup>, Shanmugam Sivakumar<sup>8</sup>, Pradeep A. Menon<sup>8</sup>, Vijay Viswanathan<sup>9</sup>, Bruno B. Andrade<sup>4,5,6,10,11</sup>, Helder I. Nakaya<sup>1</sup> & Hardy Kornfeld<sup>7</sup>

Comorbid diabetes mellitus (DM) increases tuberculosis (TB) risk and adverse outcomes but the pathological interactions between DM and TB remain incompletely understood. We performed an integrative analysis of whole blood gene expression and plasma analytes, comparing South Indian TB patients with and without DM to diabetic and non-diabetic controls without TB. Luminex assay of plasma cytokines and growth factors delineated a distinct biosignature in comorbid TBDM in this cohort. Transcriptional profiling revealed elements in common with published TB signatures from cohorts that excluded DM. Neutrophil count correlated with the molecular degree of perturbation, especially in TBDM patients. Body mass index and HDL cholesterol were negatively correlated with molecular degree of perturbation. Diabetic complication pathways including several pathways linked to epigenetic reprogramming were activated in TBDM above levels observed with DM alone. Our data provide a rationale for trials of host-directed therapies in TBDM, targeting neutrophilic inflammation and diabetic complication pathways to address the greater morbidity and mortality associated with this increasingly prevalent dual burden of communicable and non-communicable diseases.

The adverse effect of type 2 DM on TB severity at presentation, treatment response, outcomes and relapse has become a major global public health concern due to rapidly increasing DM prevalence in countries where TB prevalence was already high<sup>1</sup>. The mechanisms responsible for TB susceptibility and severity in DM are not well understood and there is a paucity of data to inform the management of TBDM patients and the design of interventional and implementation trials to address this problem. Susceptibility of diabetic mice results from a delayed innate immune response to the alveolar macrophages initially infected with inhaled *Mycobacterium tuberculosis*, resulting in several additional days of logarithmic bacterial replication<sup>2,3</sup>. Once expressed in the lung, T cell mediated immunity is qualitatively intact but with a quantitatively greater burden of inflammatory pathology due to the higher plateau bacterial burden along with possible defects in counterregulation<sup>4</sup>.

<sup>1</sup>Department of Pathophysiology and Toxicology, School of Pharmaceutical Sciences, University of São Paulo, 05508, São Paulo, Brazil. <sup>2</sup>Laboratório de Imunoparasitologia, Instituto Gonçalo Moniz, Fundação Oswaldo Cruz, Salvador, Brazil. <sup>3</sup>National Institutes of Health- NIRT - International Center for Excellence in Research, Chennai, India. <sup>4</sup>Unidade de Medicina Investigativa, Laboratório Integrado de Microbiologia e Imunoregulação, Instituto Gonçalo Moniz, Fundação Oswaldo Cruz, Salvador, Brazil. <sup>5</sup>Multinational Organization Network Sponsoring Translational and Epidemiological Research, Instituto Brasileiro para a Investigação da Tuberculose, Fundação José Silveira, Salvador, Brazil. <sup>6</sup>Curso de Medicina, Faculdade de Tecnologia e Ciências, Salvador, Brazil. <sup>7</sup>Department of Medicine, University of Massachusetts Medical School, Worcester, Massachusetts, United States of America. <sup>8</sup>National Institute for Research in Tuberculosis, Chennai, India. <sup>9</sup>Prof. M. Viswanathan Diabetes Research Center, Chennai, India. <sup>10</sup>Universidade Salvador (UNIFACS), Laureate Universities, Salvador, Brazil. <sup>11</sup>Division of Infectious Diseases, Department of Medicine, Vanderbilt University School of Medicine, Nashville, USA. Cesar A. Prada-Medina, Kiyoshi F. Fukutani and Nathella Pavan Kumar contributed equally to this work. Bruno B. Andrade, Helder I. Nakaya and Hardy Kornfeld jointly supervised this work. Correspondence and requests for materials should be addressed to H.I.N. (email: [hnakaya@usp.br](mailto:hnakaya@usp.br)) or H.K. (email: [Hardy.Kornfeld@umassmed.edu](mailto:Hardy.Kornfeld@umassmed.edu))

Investigating early events in human TB is challenging since patients typically present more than 2 months after disease onset<sup>5</sup>. Furthermore, comorbid DM specifically increases the risk for pulmonary more than extrapulmonary TB<sup>6</sup> which presents another barrier to clinical investigation since sampling the site of disease is impossible in most settings. The blood transcriptome offers a reflection of immunological events in the lung and a consensus gene expression signature of active TB is emerging from observational studies in nondiabetic TB patients in Africa, China, Europe and Indonesia<sup>7</sup>. To gain insights into mechanisms of TB susceptibility in human DM and to assess the impact of TB disease on diabetic complications, we leveraged data and blood samples from diabetic and nondiabetic TB patients enrolled in the Effects of Diabetes on Tuberculosis Severity (EDOTS) study in Chennai, India<sup>8</sup>. Integrative analysis applied to these data showed a high degree of comparability in the blood transcriptional response to TB between diabetic and nondiabetic participants but a distinct signature of plasma cytokine and growth factor levels and an association of TBDM with neutrophilic inflammation. The data further show that comorbid TB activates a range of pathways associated with diabetic complications, above the levels observed with DM alone. We conclude that neutrophilic inflammation and diabetic complication pathways may be useful targets for host-directed therapies to improve TB treatment and outcomes in this growing patient population.

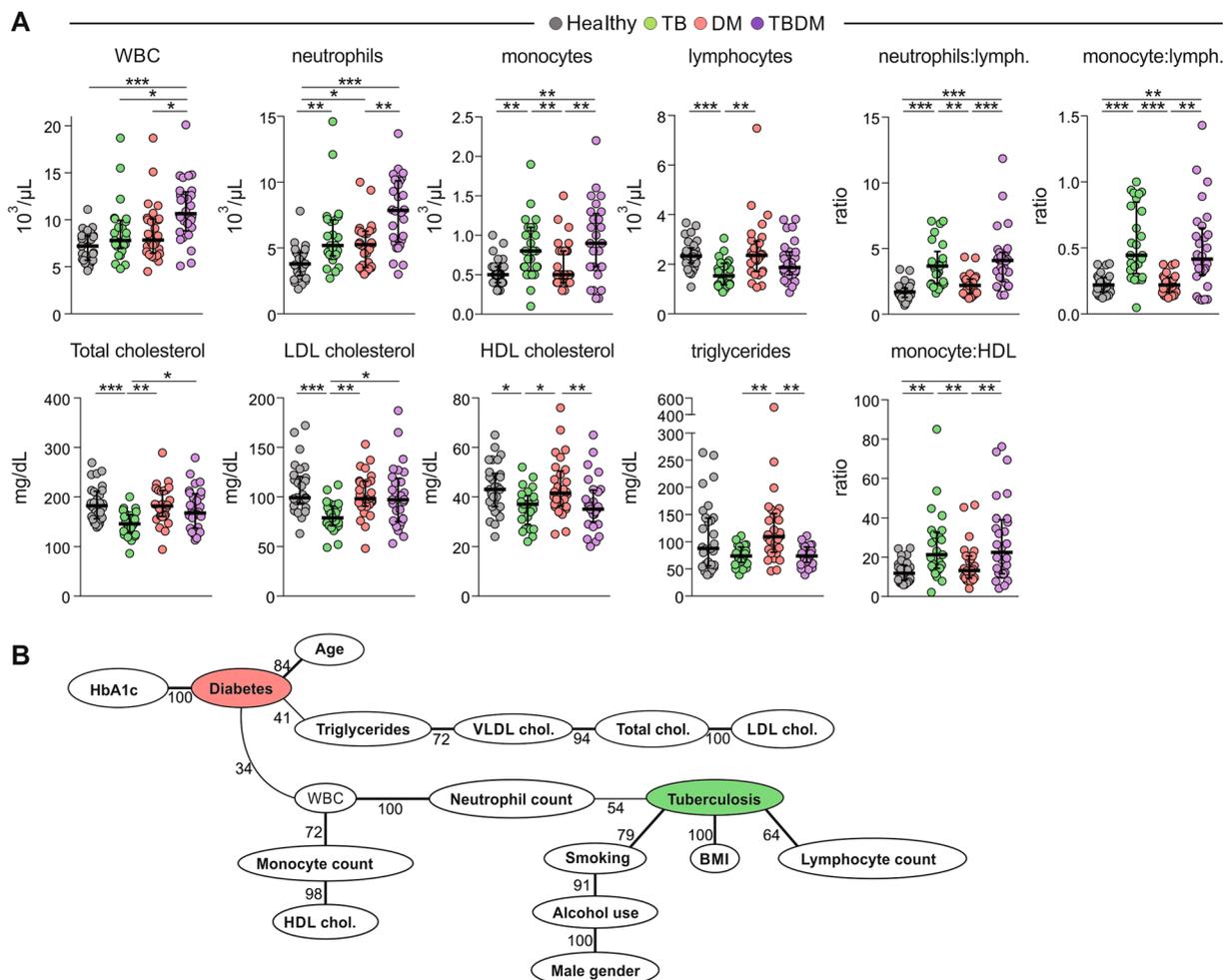
## Results

**Study population.** EDOTS is a longitudinal, observational cohort study comparing adult participants with pulmonary TB in Chennai, India, who are rigorously classified as diabetic or normoglycemic according to WHO guidelines<sup>8</sup>. The investigations reported here used demographic and clinical data as well as stored samples of plasma and whole blood RNA collected at enrollment from 60 EDOTS cohort participants (30 diabetic and 30 normoglycemic). Since the phenotype of TBDM might differ between participants who were diabetic prior to incident TB and those with an initial diagnosis of DM at the time of enrollment in EDOTS, we restricted the current investigation to participants with preexisting DM. Retreatment of TB, prior active TB diagnosis and HIV infection are exclusion criteria for EDOTS. TB diagnosis was classified by positive sputum culture for *M. tuberculosis*. Comparison was made with 60 control participants without active TB disease, of whom 30 were classified as diabetic and 30 as normoglycemic using identical case definitions. Demographic and behavioral differences were identified between participants with TBDM dual burden, TB alone, DM alone and healthy controls (Supplemental Table 1). Those with TBDM tended to be older than normoglycemic participants with TB but less likely to report tobacco or alcohol consumption. The healthy control and DM subgroups had comparable, normal median body mass index (BMI) whereas median BMI was significantly lower in the TBDM and TB subgroups.

**Demographic, behavioral and clinical laboratory parameters.** Among the 120 participants in the current investigation, median glycohemoglobin (HbA1c) was 5.5%, 7.5%, 5.7% and 10.2% in the subgroups of healthy controls, DM, TB, and TBDM, respectively (Supplemental Table 2). All subgroups were vitamin D insufficient (<20 ng/mL; Supplemental Table 2)<sup>9</sup>. Median circulating leukocytes and absolute neutrophil count were higher in TBDM patients than the other subgroups, but both DM subgroups had higher neutrophil counts than healthy controls (Fig. 1A and Supplemental Table 2). The neutrophil:lymphocyte ratio (NLR), a biomarker associated with TB severity<sup>10</sup>, insulin resistance and type 2 DM<sup>11,12</sup> and coronary artery disease risk<sup>13,14</sup> was higher in the TB and TBDM subgroups compared to DM and to healthy controls. Participants with TBDM had lower serum triglycerides and lower HDL than the DM subgroup (Fig. 1A). The monocyte:HDL ratio (MHR), a biomarker associated with coronary artery disease risk and severity<sup>15</sup> was also higher in the TBDM and TB subgroups compared to DM and to healthy controls.

We applied Bayesian Network modeling to infer causal relations between the presence of TB and/or DM and the recorded demographic, behavioral and clinical laboratory parameters (Fig. 1B). The network learned from these variables represents the minimum probabilistic dependency between the variables that explains most of the observed correlations in the data<sup>16</sup>. This confirmed the expected associations of DM with patient age, HbA1c, triglycerides and VLDL, total and LDL cholesterol. Gender (male), alcohol use and smoking formed an association chain with TB, while BMI and lymphocyte count were also strongly associated with TB. Neutrophil count was identified at the nexus between DM and TB. Altogether, these data indicate that neutrophilic inflammation is a central feature of TBDM and that incident TB disease elevates levels of biomarkers associated with diabetic vascular complications above the levels found in DM without comorbid TB.

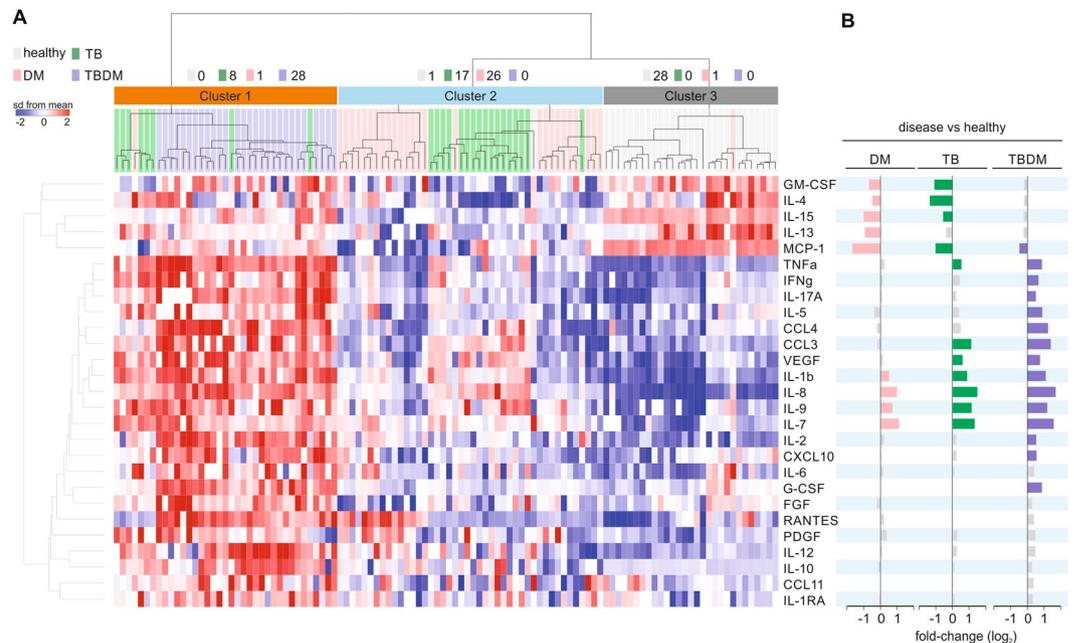
**Signature pattern of elevated plasma cytokines and growth factors in TBDM comorbidity.** Elevated plasma levels of a broad range of cytokines in TBDM patients have been reported in diverse populations<sup>17–19</sup>. To assess this relationship in our cohort of TB patients and controls, we measured plasma levels of 27 cytokines and growth factors by Luminex immunoassay (Fig. 2 and Supplemental Table 3). Primary analyses revealed that several analytes exhibited statistically different concentrations in plasma between the study groups (Supplemental File 1). Data was then z-score normalized across the entire cohort and analyzed by hierarchical clustering. Three main clusters were identified and showed 100% separation for TBDM vs. healthy controls, with more overlap between the TB and the DM subgroups. The cluster containing all TBDM cases (and 32% of TB cases) was notable for higher expression of nearly all analytes tested (Fig. 2A and Supplemental File 1). Surprisingly, another cluster contained both the majority of nondiabetic TB cases (68%) and the majority of diabetic controls without TB (93%). Growth factors elevated in TBDM included VEGF, PDGF and FGF that are all implicated in diabetic complications<sup>20–22</sup> (Supplemental File 1). Compared to healthy subjects, the levels of several analytes (IFN $\gamma$ , IL-17A, IL-5, CCL4, IL-2, CXCL10, and G-CSF) were significantly higher only in the TBDM subgroup (Fig. 2B). We conclude that a signature pattern of plasma levels of cytokines and growth factors distinguished TBDM from DM and healthy control patients, and from a majority of nondiabetic TB patients in this cohort.



**Figure 1.** Factors associated with TB and/or DM in the cohort. **(A)** Data represent median and interquartile ranges. The Kruskal Wallis test with Dunn's multiple comparisons ad hoc test were employed to compare the values detected between the study subgroups. *P*-values for the Kruskal Wallis tests are shown in Appendix File S1. *P*-values were adjusted for multiple comparisons using Holm-Bonferroni's approach as described in methods. Only comparisons with significant *P*-values after adjustments are displayed (\**P* < 0.05, \*\**P* < 0.01, \*\*\**P* < 0.001). **(B)** Bayesian network with bootstrap (100x) was used to illustrate the statistically significant associations between the parameters and the presence of TB and/or DM in the study population. Lines represent direct associations. Associations that remained statistically significant on >30 times out of 100 bootstraps are plotted. Numbers of times each association persisted during bootstrap are shown. Bold lines highlight the strongest associations, which persisted more than 60 times in the bootstrap. Vitamin D is not presented in the picture, as it would be solitary node with no arcs attached.

**Blood transcriptome profile of patients with TB and/or diabetes.** To investigate the impact of DM on human immunity to TB we performed whole blood gene expression profiling of all participants in the cohort. A total of 993 differentially expressed genes (DEGs) were identified, comparing the TBDM, TB and DM subgroups to the healthy subjects (Fig. 3A, Supplemental File 2). The expression patterns of these DEGs across all samples revealed considerable heterogeneity among individuals within each subgroup (Fig. 3B). Most of the DEGs (935) were identified comparing TB or TBDM classes vs. healthy controls. Among these, 455 genes were commonly found in TB and TBDM comparisons. These included genes found in previously reported TB signatures<sup>23,24</sup> (Fig. 3A). Only 74 DEGs were found between DM patients and healthy subjects (Fig. 3A). We conclude that the immune response to *M. tuberculosis* infection is the major driver of the blood transcriptomic changes seen in TBDM patients and this response is only mildly perturbed by DM as compared to the response in nondiabetic individuals with pulmonary TB disease.

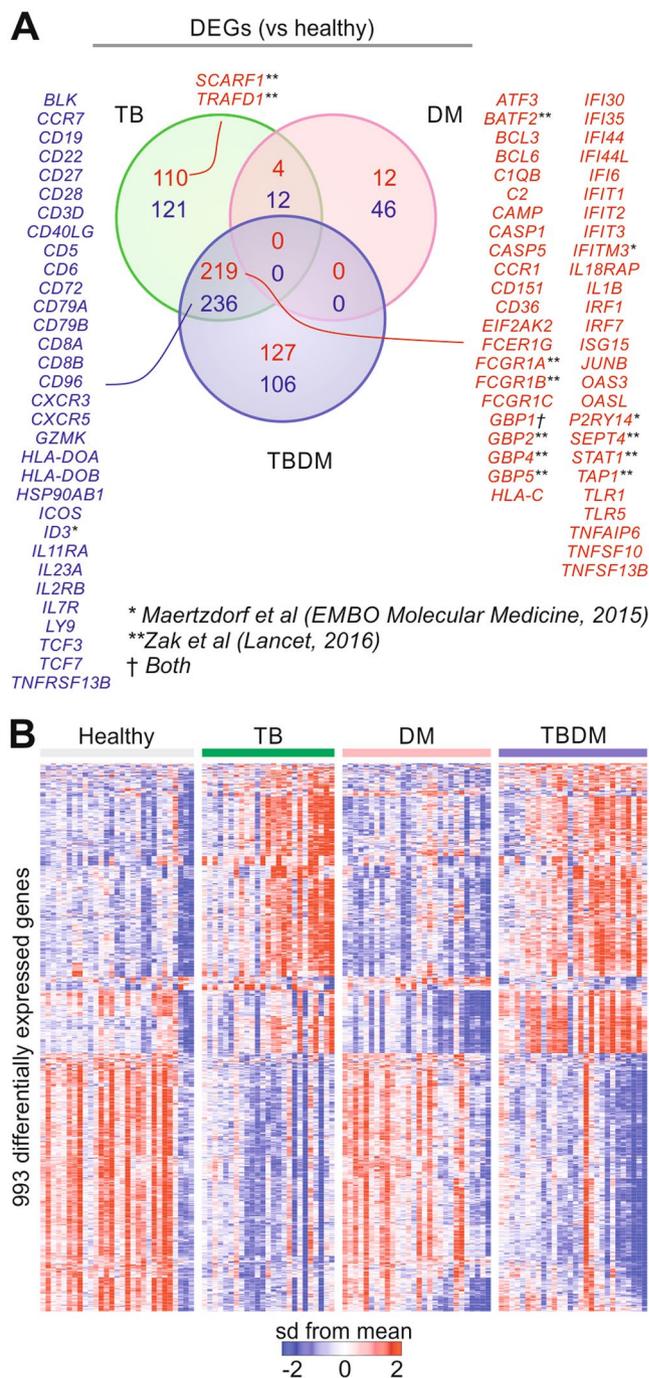
We next compared differences in the transcriptome profiles between participants with TBDM and participants with either TB or DM alone (Fig. 4A, Supplemental File 2). Whereas little change was found between the TBDM and TB subgroups (62 DEGs), a total of 1,138 genes were differentially expressed in TBDM compared to DM patients ("TBDM total"). Of those, 521 genes were also differentially expressed in the comparison between TB patients and DM patients ("TB shared"). The remaining 436 genes were uniquely differentially expressed in the TBDM vs. DM comparison, reflecting an impact of TBDM comorbidity on blood gene expression distinct from



**Figure 2.** Cytokine profiling in plasma of patients with TB and/or DM. (A) Cytokine concentrations were z-score normalized across all subjects. Each column represents one patient. The cytokine names are shown to the right of the heat map. Cytokine profiles were ordered by hierarchical clustering (Euclidean distance and clustered with ward method). The condition tree at the top shows 3 main clusters. The number of subjects from each clinical phenotype (class) on each cluster is indicated. (B) Differences in cytokine levels for each disease compared to healthy subjects. Differences which did not reach statistical significance (Adjusted  $P < 0.05$ , fold-change  $> 1.4$ ) are represented as grey bars.

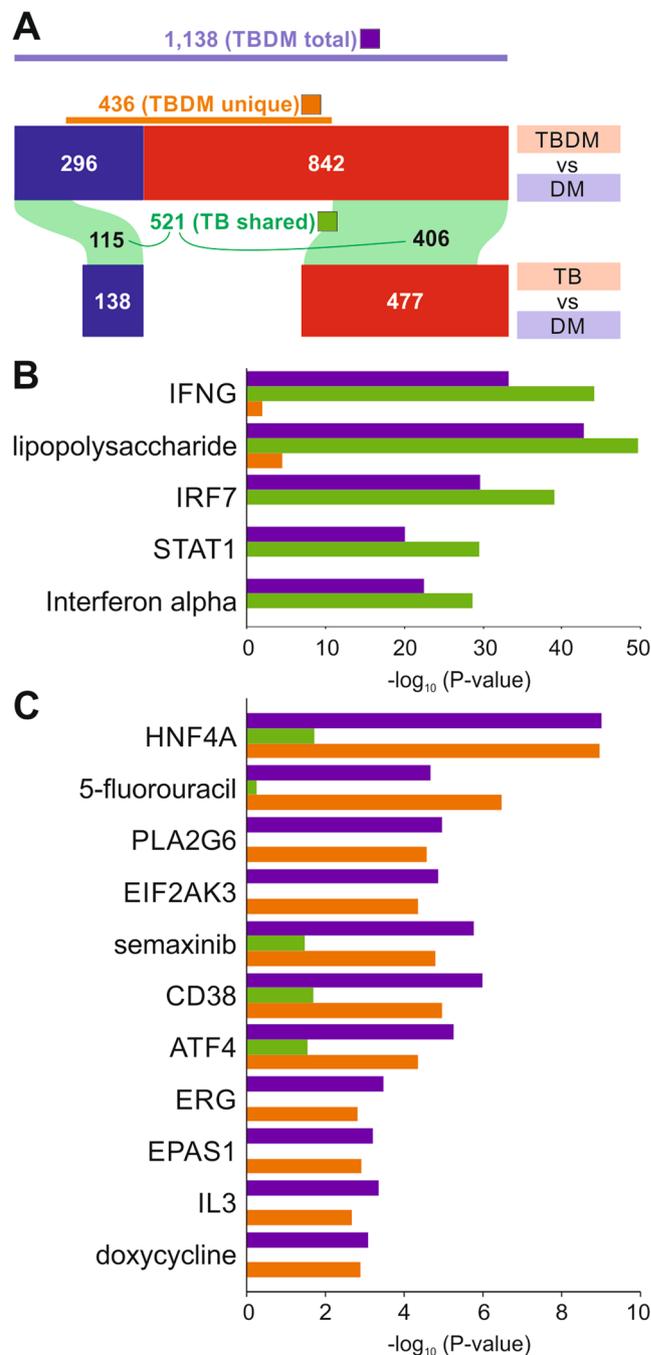
the influence of either TB or DM alone (“TBDM unique”). Ingenuity Pathway Analysis applied to these gene lists revealed the main possible transcriptional upstream regulators driving the TB (Fig. 4B) and TBDM (Fig. 4C) responses. As expected, TB disease was associated with the induction of genes related to types I and II IFN responses (Fig. 4B). Among potential regulators significantly enriched in “TBDM unique” genes, several were related to endoplasmic reticulum (ER) stress, including HNF4A, ATF4, EIF2AK3, and PLA2G6. Glucotoxicity and lipotoxicity are known to promote ER stress in type 2 DM<sup>25</sup> but our results indicate that comorbid TB further exacerbates ER stress in diabetic hosts. This may be clinically significant since ER stress is linked to loss of  $\beta$ -cell mass in type 2 DM and is implicated in diabetic complications<sup>25,26</sup>. The “TBDM unique” genes were also significantly enriched by targets of the anti-cancer drugs 5-fluorouracil and semaxanib, and the antibiotic doxycycline (Fig. 4C). These three different classes of drugs share a common link with the DNA damage response pathway<sup>27–29</sup>, which is associated with insulin resistance and microvascular complications of DM<sup>30</sup>.

**Molecular degree of perturbation is elevated in TB and diabetes alone and in combination.** The molecular distance to health (MDH) is a numerical score of transcriptional perturbation made by summing the differences in the expression levels of pre-selected genes between a given sample group and healthy control participants<sup>31</sup>. When applied to the blood transcriptome, MDH largely reflects immunological activation and it was reported to be higher in TB patients with more severe TB disease on chest x-ray<sup>32</sup>. To assess global effects of TB and DM on the blood transcriptome we quantified the molecular degree of perturbation (MDP) in the transcriptome of each patient. The MDP is adapted from the MDH and conveys similar information but has the advantages of controlling for variability in the control group and allowing for cross-dataset comparisons (see Materials and Methods). Median MDP vs. the healthy control transcriptome was elevated to a comparable degree in the TB and TBDM subgroups compared to participants with DM alone (Fig. 5A). Correlation of the expression values of each individual gene with MDP demonstrated a preponderance of downregulated transcripts in the DM subgroup vs. predominantly upregulated transcripts in TB patients with TBDM spanning these two extremes (Supplemental Figure 1). Across the total cohort, BMI, % and total lymphocytes, total cholesterol and all cholesterol subfractions were negatively correlated with MDP, while positive correlations were identified with HbA1c, total WBC, % and total neutrophils and % monocytes (Fig. 5B). Within the participant subgroups, a negative correlation of HDL cholesterol and MDP was identified only in TBDM patients, while neutrophils were positively correlated with MDP in the TB, DM and TBDM subgroups. Of note, the frequency of participants with high MDP and high absolute neutrophil count was greatest in the TBDM subgroup (Fig. 5C). These observations suggest an association of peripheral blood neutrophil numbers in with TB disease severity as reflected by MDP in patients with TB or TBDM. The data also suggest an association of low HDL cholesterol with the hyperinflammatory response of TBDM patients while high BMI was associated with a lower level of systemic inflammation in all subgroups.



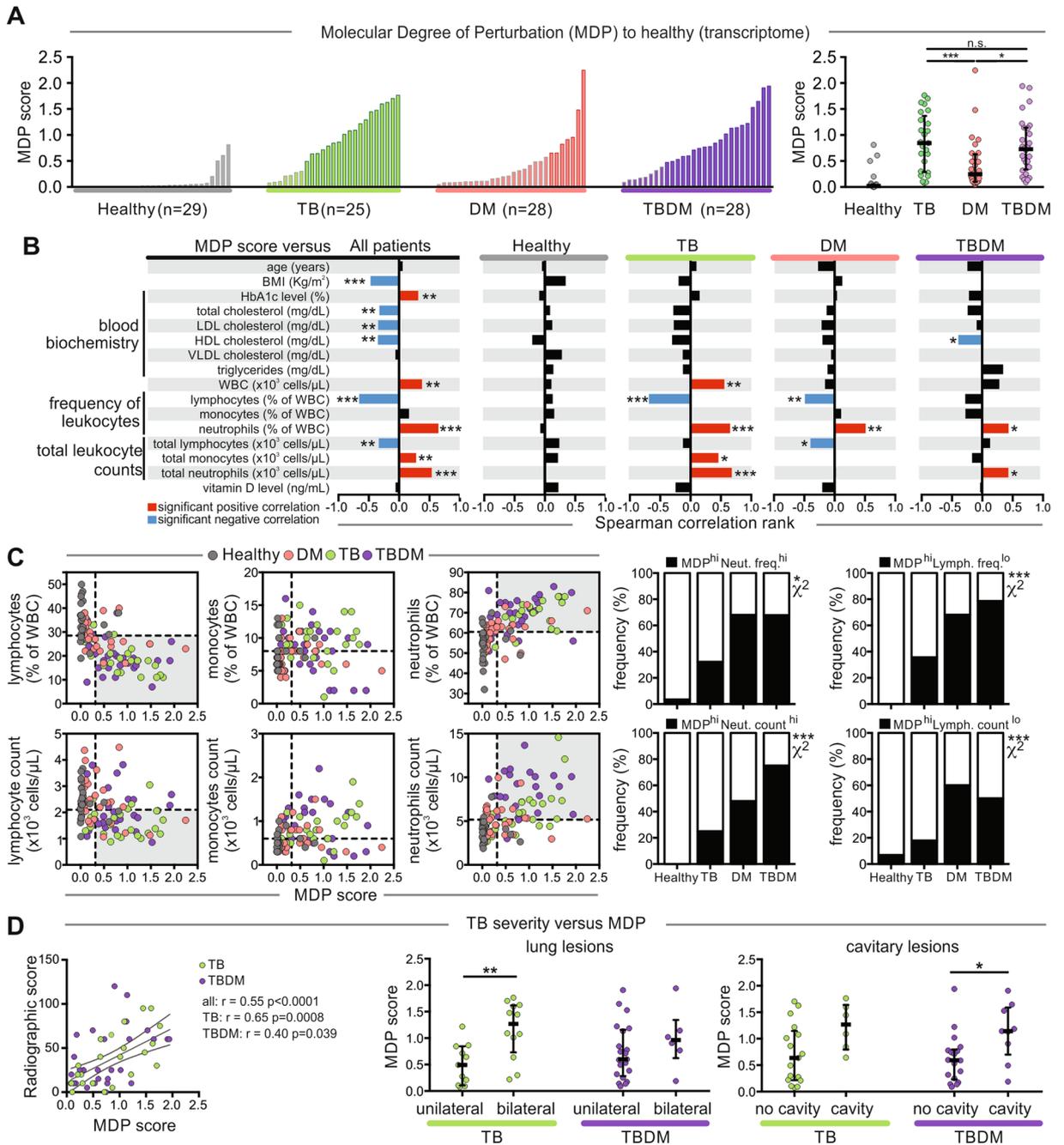
**Figure 3.** Transcriptomic changes with TB and/or diabetes compared to healthy subjects. **(A)** Differentially expressed genes (DEGs) in patients with TB and/or DM compared to healthy subjects (adjusted  $P < 0.05$ , fold-change  $> 1.4$ ). The Venn diagram shows the number of DEGs in common to two or more clinical phenotypes or unique to only one disease. **(B)** Expression patterns of the DEGs from **(A)**. Cohort subgroups are shown by the colored bars. Each column represents one subject. The genes (rows) were normalized by z-score across all samples.

The MDH was previously shown to correlate with the radiographic extent of TB disease estimated as mild, moderate or advanced using a widely adopted but non validated categorical scoring method<sup>32</sup>. In the current study, we graded chest x-ray severity using a numerical scoring method validated for smear-positive pulmonary TB<sup>33</sup>. This continuous radiographic severity variable correlated with MDP in TB and TBDM patients (Fig. 5D). Statistically significant correlation of MDP with the presence of bilateral vs. unilateral lung involvement was identified in TB patients, with a similar trend in TBDM. The presence of cavitory lesions correlated significantly with MDP in TBDM, with a similar trend in the TB subgroup. We conclude that immune activation as reflected by whole blood MDP is comparably elevated across the groups of TB and TBDM patients and correlates at the

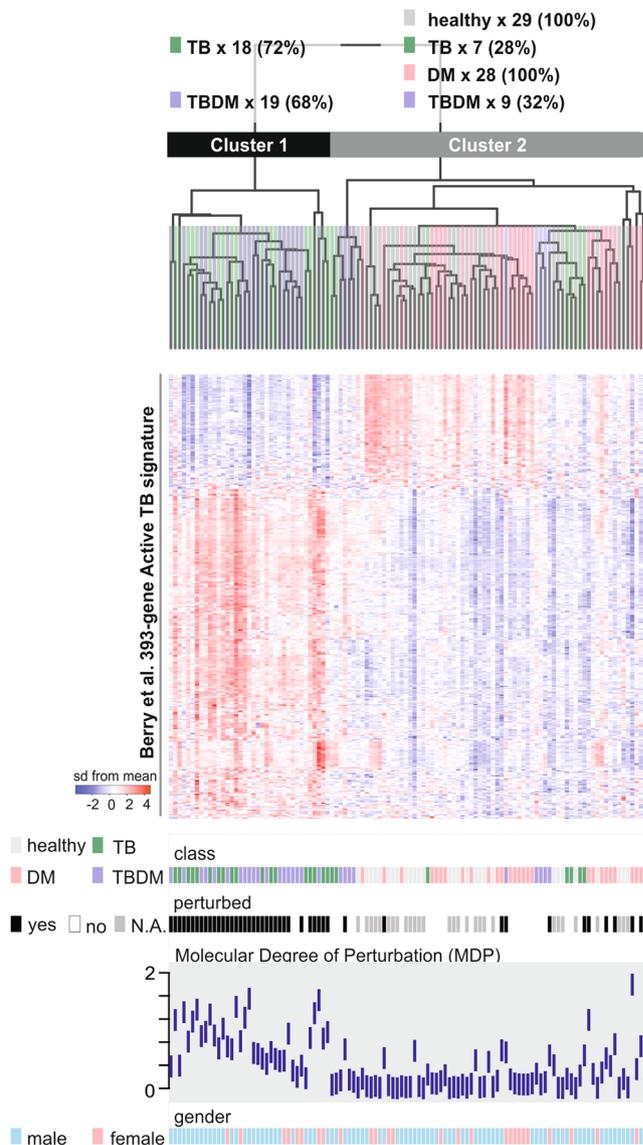


**Figure 4.** Blood transcriptome of patients with dual TB and DM burden compared to TB or DM alone. **(A)** Differentially expressed genes between patients with distinct clinical phenotypes (class). The comparison between two classes is indicated at the top of each bar. Red and blue bars represent the number of DEGs which are up- or down-regulated, respectively in the first class of the comparison. The purple bar shows the number of DEGs in the comparison TBDM vs. DM (“TBDM total”). The number of DEGs which are in common to two comparisons are shown inside green areas connecting two horizontal bars (“TB shared”). Genes which are uniquely differentially expressed in TBDM vs. DM is shown by the brown bar (“TBDM unique”). **(B)** Top transcriptional regulators in “TB shared” DEGs. Ingenuity’s Upstream Regulator Analysis was performed using “TB shared” (green bars), “TBDM unique” (brown bars), and “TBDM total” DEGs (purple bars). The x-axis indicates the significance of the enrichment for the upstream regulators on the left. **(C)** Top transcriptional regulators in “TBDM unique” DEGs identified by the same approach as in **(B)**.

individual level with immune pathology and disease severity estimated by chest x-ray. These data further support the association of neutrophilic inflammation with TB severity, a relationship that is particularly strong in the TBDM subgroup.



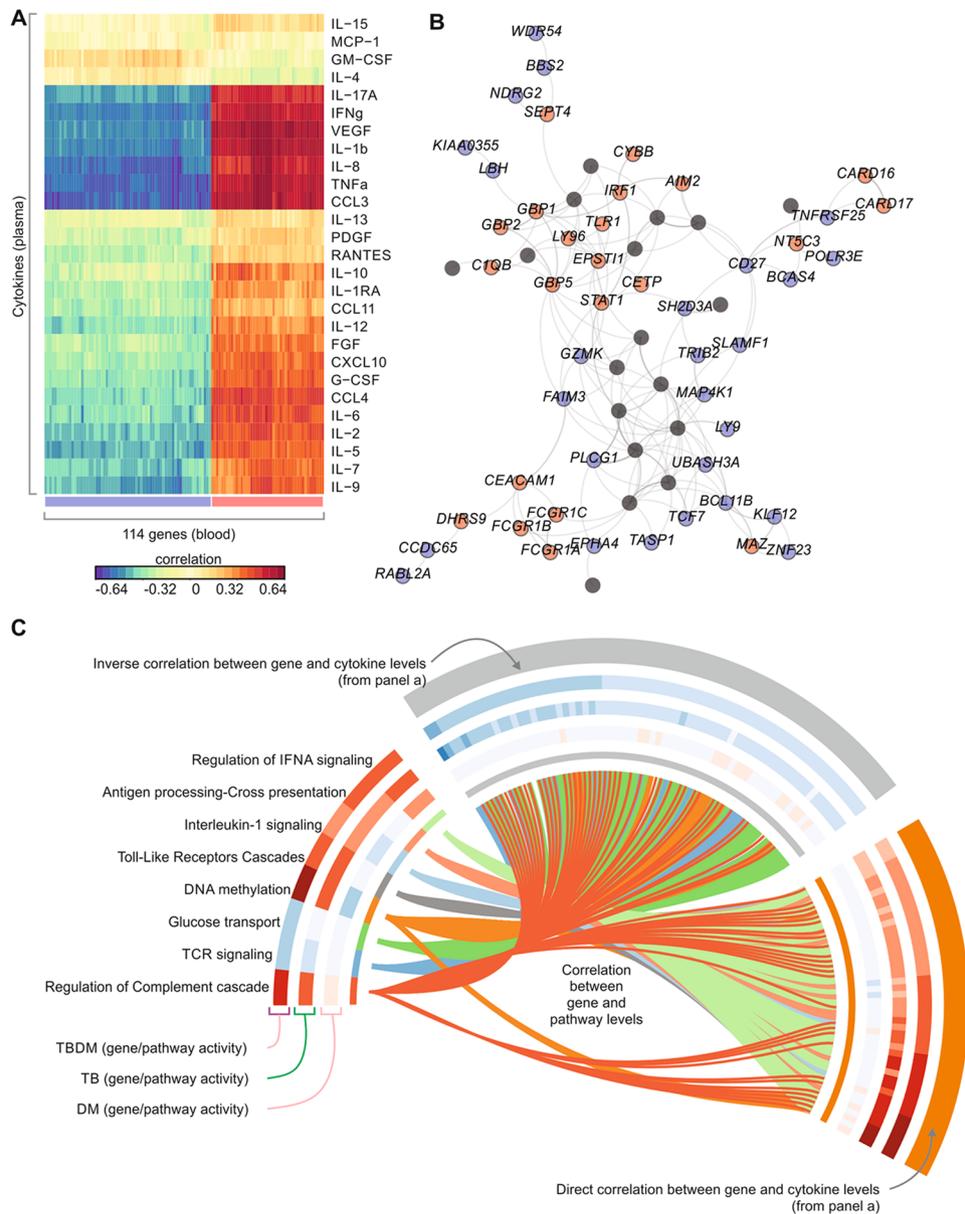
**Figure 5.** Factors associated with increased molecular degree of perturbation. **(A)** The molecular degree of perturbation (MDP) relative to healthy controls was calculated as described in Methods. Left panels show histograms of individuals in each subgroup. Right panel shows individual values with median and IQR per group. Data were compared using one-way ANOVA with Tukey’s multiple comparisons test, with single pooled variance. **(B)** MDP values were tested for correlations with the indicated parameters in each study subgroup using Spearman correlation ranks. Statistically significant correlations, after Holm-Bonferroni’s adjustment for multiple comparisons, are highlighted (red, positive correlation; blue, negative correlation; black, non-significant correlation). **(C)** Left panel shows correlation plots between differential WBC counts (neutrophils, monocytes and lymphocytes) and the MDP in all study groups. Dotted lines represent median values for MDP or leukocyte counts/frequency for the entire study population. Shaded areas indicate the individuals with the highest MDP values and the lowest lymphocyte counts/frequency or the highest neutrophil counts/frequency relative to the study population. Right panel shows comparisons of the frequencies in each study subgroup of individuals from the shaded areas on the left. Frequency comparisons were performed using the chi-squared test. **(D)** Left panel, MDP score was tested for correlation with the radiographic severity score (Spearman rank test). Right panel shows comparisons of MDP values in patients diverging in terms of TB disease distribution (unilateral vs. bilateral lung lesions) or presence of cavitory lesions. Within each study subgroup, values were compared using the Mann-Whitney U test. In all comparisons: \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .



**Figure 6.** A blood transcriptional signature of active TB classifies most patients with or without diabetes. We applied a previously defined 393-gene signature of active TB<sup>32</sup> to the 387 genes represented in the microarrays used for the present study. Expression profile of each patient and health subject (column) was ordered by hierarchical clustering (Spearman correlation with average linkage). The condition tree at the top shows 2 main clusters. The number of individuals from all each clinical subgroup (class) who segregated into each cluster is indicated at the top of the figure. Colored blocks at each profile base represent the different classes, molecular perturbation status, MDP score and gender.

**Classification of active TB by whole blood gene expression.** Blood transcriptional signatures may inform the development of rapid diagnostic tests to distinguish between TB infection and TB disease. To determine whether comorbid DM alters the expression of TB signature genes, we analyzed the expression of a previously described 393-gene TB signature<sup>32</sup> using the microarray dataset from our cohort, ordering the results by hierarchical clustering (Fig. 6). By this approach, 72% of nondiabetic participants with TB and 68% of participants with TBDM were correctly classified by the active TB disease signature. The second main cluster accurately classified 100% of healthy controls but misclassified 28% in the TB subgroup and 32% in the TBDM subgroup. The clustering results were compared at the individual level with MDP, which revealed that most patients with active TB who were misclassified in the healthy control cluster had lower MDP than the TB and TBDM subgroup members who were accurately classified as having active TB (Fig. 6A). Despite this moderate level of correct classification using the 393-gene signature, we noted that DEGs common to TB patients with or without DM in our cohort included all four genes from a 4-gene signature proposed by Maertzdorf *et al.* as the basis for developing a point-of-care diagnostic test<sup>23</sup> (Fig. 3A). Another recently described 16-gene signature was shown to predict progression from TB infection to disease in independent South African and Gambian cohorts<sup>24</sup>. In our Indian cohort, DEGs in the nondiabetic TB subgroup included 11 DEGs present in the 16-gene signature, while 9 DEGs





**Figure 8.** Integrative analysis of cytokines, gene expression and pathway activity in TBDM. **(A)** Integration of cytokine levels and microarray gene expression. Sparse group partial least square methods were used to identify the genes whose expression best correlates with most of the cytokines. Colors represent the correlation values between the expression of genes in blood (columns) and the levels of cytokines in plasma (rows). **(B)** Network containing some of the 114 genes from **(A)** GeneMania program was used to define the interactions (edges) between genes (nodes), and Gephi program to visualize the network. Colors represent is positive (red) or negative (blue) correlation of gene expression with the cytokines. **(C)** Integration of gene expression and pathway activity. Circos plot shows 8 selected pathways and 114 genes from **(A)**. Each pathway was assigned to one color (inner circle). The lines connecting pathways and genes represent a high correlation (spearman correlation above 0.6) between the activity of the pathway (defined by the single-sample enrichment score) and the gene expression. Heat maps shown in the outer circles represent the activity of pathways (mean Normalized Enrichment Score) and genes (mean log<sub>2</sub> fold-change in patients compared to healthy subjects). Colors indicate increase (red) or decrease (blue) in pathway activity or gene expression fold-change.

also enriched, albeit to a lesser extent. Regardless of magnitude, the combined effects of the complication pathways activated in TBDM and linked to TB susceptibility suggest bidirectional pathological interactions between TB and DM.

**Integrative analysis reveals the potential drivers of the plasma cytokine changes in TBDM comorbidity.** We applied the sparse partial least squares regression (sPLS) method<sup>39</sup> to identify the minimum number of genes whose expression best explains most of the observed cytokine variation in TBDM (Fig. 8A). This

approach identified 114 genes, many with key roles in innate and adaptive immunity (e.g. *BATF2*, *AIM2*, *CD27*, *FCGR1A*, *FCGR1B*, *FCGR1C*, *GZMK*) as well as histone genes (Fig. 8B, Supplemental File 4). Notably, plasma proteins associated with neutrophil recruitment and activation (IL-1b, IL-8, IL-17A, CCL3, TNFa, VEGF) were among those having strongest positive and negative correlations with gene expression values. Integrating the expression of these 114 genes with the activity of Reactome pathways, confirmed that most of the genes were associated with IFN, DNA methylation and antigen processing cross-presentation (Fig. 8C, Supplemental File 4). The patterns of gene/pathway activation were qualitatively similar for the TBDM and TB groups, and distinct from DM without TB (Fig. 8C). On a quantitative basis, the TBDM group had a greater increase in the activity of the “Regulation of complement cascade” and particularly the “DNA methylation” pathways as compared to TB alone. Taken together, these data suggest that epigenetic reprogramming and neutrophilic inflammation play a role in the signature pattern of plasma cytokines and growth factors in TBDM. While causation cannot be inferred from the available data, these may offer targets for therapeutic intervention.

## Discussion

The global population-attributable fraction of DM for adult TB cases (estimated to be 15% in 2013)<sup>40</sup> exceeds the burden attributable to HIV and will continue to rise rapidly in countries with high TB burden and increasing prevalence of DM<sup>41</sup>. The dual burden of TB and DM is particularly acute in India, which has the largest number of TB cases (23% of global burden) and the second highest estimated number of DM cases (65.1 million) in the world<sup>42</sup>. There is an urgent need to understand the mechanisms and consequences of the interactions between DM and TB in order to define research priorities and provide a rational basis for the development of interventions. To address this gap in knowledge we collected clinical data, measured plasma cytokines and growth factors, and analyzed whole blood gene expression in a prospectively recruited cohort of South Indian adult pulmonary TB patients who were rigorously classified as diabetic or normoglycemic.

The immunological basis of diabetic TB susceptibility is not completely understood but insights have been gained from animal models<sup>43</sup>. Diabetic mice exhibit a critical delay in adaptive immune priming attributable to defective sentinel function of resident alveolar macrophages<sup>3,44</sup>. Once underway, the T cell response in diabetic mice appears to be functionally intact but quantitatively excessive, possibly reflecting higher antigen load and/or defective counterregulation<sup>4</sup>. On this basis, we predicted that the immune response in TBDM patients would be qualitatively similar but quantitatively greater than that occurring in the majority of nondiabetic TB patients.

Bayesian network analysis placed neutrophils at nexus of the connection between DM and TB (Fig. 1B). An association of neutrophilic inflammation with TB severity in the general population is well recognized<sup>45</sup>. Our data indicate that comorbid DM amplifies this response, which might reflect high bacterial load and/or a specific perturbation of immune function. Elevated levels of plasma cytokines and growth factors distinguished participants with TBDM from two thirds of nondiabetic TB cases and all but one DM control whereas there was substantial overlap between the nondiabetic TB group and the diabetic controls without TB (Fig. 2). Results for the TBDM group are consistent with and extend published human and mouse data<sup>2,4,17–19</sup> while the overlap between nondiabetic participants with TB and diabetic controls without TB was unexpected. The latter result points to a need for rigorous classification of glycemic status when interpreting proteomic data, at least in South Indian populations where DM appears to be a strong confounding factor.

The first published blood transcriptomic signature of active TB in cohorts from the United Kingdom and South Africa<sup>32</sup> highlighted a role for type I IFN signaling in neutrophils. Fourteen subsequent transcriptomic studies analyzed gene expression in whole blood or isolated leukocyte fractions in TB patient cohorts from Africa, Europe, Indonesia and China<sup>46</sup>. Notably, none of these studies focused on TBDM comorbidity or included patients in India. Consistent with our prediction, the qualitative transcriptional response in TBDM closely matched that of the nondiabetic TB signature in terms of DEGs vs. healthy or diabetic participants without TB disease. An important conclusion is that comorbid DM does not appear to confound the clinical application of transcriptional profiling to TB diagnosis. Contrary to our expectation, there was no significant trend for higher MDP in TBDM vs. TB (Fig. 5A) although the TBDM subgroup had the highest proportion of individuals with combined high neutrophil count and high MDP (Fig. 5C). High MDP was positively correlated with radiographic severity, high neutrophil numbers and low BMI (Fig. 5B–D). The data are consistent with evidence that higher BMI is protective in TB<sup>47</sup> but this association is lost in diabetic obesity, presumably as a consequence of diabetic immunopathy. An association of low HDL with high MDP was unique to the TBDM subgroup. HDL cholesterol modulates innate and adaptive immunity through interactions with lipid rafts on myeloid and lymphoid cells<sup>48</sup>. The low HDL cholesterol associated with type 2 DM and pre-diabetes might be a glucose-independent factor contributing to TB susceptibility. We conclude that the transcriptional response to TB is qualitatively and quantitatively similar in the presence or absence of DM but confirmatory studies examining larger numbers of patients are warranted as the relatively small sample size and high degree of individual variation in MDP seen in all four subgroups were limiting factors.

While the blood transcriptional response did not distinguish TBDM from nondiabetic TB patients, 436 DEGs identified in the comparison of TBDM vs. DM were not shared in the comparison of nondiabetic TB to DM patients (Fig. 4). A finding with potentially high biological and clinical significance was that many of these “TBDM unique” DEGs participate in pathways that promote diabetic complications. Across a wide range of comparisons in this study, the amplification of genes, cytokines, growth factors and other biomarkers associated with diabetic complications was observed. These included but were not limited to the metrics of NLR and MHR, the plasma levels of VEGF, PDGF and FGF, and pathways associated with DNA damage and oxidative and ER stress, complement and platelet activation, RAGE signaling, and genes and pathways involved in epigenetic remodeling. We propose that a syndemic relationship exists between TB and DM that is conceptually similar to the relationship between TB and HIV<sup>49</sup>. Amplification of diabetic complication pathways during TB disease could be a factor in the increased early mortality of TBDM vs. DM<sup>50</sup> and might exacerbate cardiovascular disease, kidney

dysfunction, neuropathy and retinopathy in TB survivors. Conversely, the activation of pathways linked to beta cell loss could accelerate progression of dysglycemia. These questions are amenable to clinical investigation and TBDM patients might derive particular benefit from host-directed therapies targeting dyslipidemia, oxidative and ER stress, and other diabetic complication pathways such as RAGE signaling for which a robust drug development pipeline is already in place<sup>51</sup>.

Weaknesses of the present study include relatively small sample size, the high degree of individual variability in transcriptional response, heterogeneity within and between participant subgroups, and sample and data collection from the single time point of TB diagnosis. These deficiencies were partially compensated by the selection of a study site and patient population highly relevant to the TBDM syndemic. Results from EDOTS study screening indicate that 54% of adult pulmonary TB patients currently presenting to the government clinic system in Chennai are diabetic and another 21% are prediabetic<sup>8</sup>, making this a clinically important association regardless of causation. Additional strengths were the rigorous case definitions for glycemic status based on oral glucose tolerance test and fasting plasma glucose, the diversity of clinical parameters measured and the application of a validated radiographic scoring method.

In this systems biology investigation of TBDM comorbidity in a South Indian cohort we found no evidence for a discrete defect in the immune response to established infection. This was in keeping with studies from the mouse TBDM model that identified impaired kinetics of innate immunity following inhalation of *M. tuberculosis* and an associated delay in adaptive immune priming as key factors in diabetic TB susceptibility. These early events occur before patients with active TB seek medical attention and can only be inferred in clinical research by subsequent effects disease evolution. No unique TBDM transcriptional signature was identified in this study; future studies in multiple, larger cohorts will be required to generate training and validation datasets to definitively address the question whether a discrete transcriptomic signature of TBDM exists. Despite limitations, the data presented here confirm the association of TBDM with increased, neutrophil-rich inflammation<sup>52</sup>. An unanticipated finding was that incident TB disease upregulated a broad range of genes and pathways involved in the progression of diabetic complications, above the levels seen in diabetic participants without TB. A particularly strong signal was seen for genes and pathways associated with epigenetic regulation (Figs 6 and 7). Epigenetic reprogramming could explain the persistence of DM-related phenotypes of T cells and macrophages adoptively transferred into non-diabetic recipients<sup>4,44</sup> and epigenomic investigation in peripheral blood leukocytes is promising direction for future clinical studies. Diabetic complication mechanisms, including epigenetic reprogramming, may contribute to the adverse effects of comorbid DM on TB outcomes and offer targets for therapeutic intervention in the rapidly growing number of people living with this dual burden of communicable and non-communicable diseases.

## Materials and Methods

**Study site and population.** Participants with active pulmonary TB disease and either DM that preceded incident TB or normoglycemia were selected from those enrolled in the EDOTS cohort study in Chennai, India<sup>8</sup>. Briefly, adult smear-positive TB suspects from public clinics in Chennai were consented for screening if they conformed to the inclusion criterion of age 25–60 yr and lacked excluding criteria of treatment for prior episodes of TB disease, >7 d treatment for the current TB episode, >7 doses of a fluoroquinolone within 30 d of screening, pregnant or nursing, HIV seropositive or taking immunosuppressive drugs. Eligible candidates reporting a prior history of DM were confirmed by HbA1c  $\geq 6.5\%$ <sup>53</sup> and current use anti-diabetic drug treatment. Those without DM history were screened by fasting plasma glucose and oral glucose tolerance test (75 g glucose challenge). Glycemic status based on plasma glucose 2 h post-challenge according to World Health Organization criteria<sup>54</sup>: DM ( $\geq 200$  mg/dL) and normoglycemia ( $<140$  mg/dL). Enrolled participants classified as diabetic or normoglycemic were withdrawn if the sputum culture obtained at enrollment was final negative for *Mycobacterium tuberculosis*. Candidates with no prior DM history but meeting the screening classification for DM were enrolled in the EDOTS cohort but were not included in the analysis reported here since they may represent a discrete subpopulation and could introduce unwanted heterogeneity. Sixty TB patients (30 with DM and 30 without DM) enrolled between January, 2014, and July, 2015, were selected from the EDOTS cohort for the current investigations. Selection was based on the closest age match between the TB and TBDM subgroups, with investigators blinded to other participant characteristics. A total of 60 control participants without TB (30 diabetic and 30 normoglycemic) were recruited from persons attending a DM screening clinic operated by the M.V. Hospital for Diabetes in the same urban location as the TB clinics. The absence of TB disease in control participants was confirmed by negative chest x-ray and sputum culture. Glycemic status classification in the control participants was confirmed by identical criteria as the TB subgroups.

**Laboratory Investigations.** Complete blood counts, fasting lipid panel (total, low density lipoprotein [LDL] and high density lipoprotein [HDL] cholesterol, triglycerides) and 25-hydroxyvitamin D were measured for all participants in the laboratory of the M.V. Hospital for Diabetes. Chest x-rays were graded by two blinded readers using a validated radiographic TB severity scoring method<sup>33</sup>. All investigations reported here were conducted at the time of enrollment in the EDOTS cohort when participants had received  $<7$  d of anti-TB treatment. Characteristics of the population are shown in Supplemental Table 1. Samples of plasma were purified and stored frozen at  $-80^{\circ}\text{C}$  prior to batch-wise Luminex assays (Bio-Rad, Hercules, CA). Samples of whole blood were stored at  $-80^{\circ}\text{C}$  in Tempus tubes (Applied Biosystems, Foster City, CA) prior to batch-wise preparation of total RNA using Tempus Spin RNA Isolation Kits (Thermo Fisher Scientific). Purified total RNA samples were delivered to Bionivid Technology Private Limited (Bangalore, India) for RNA quality control, labeling and hybridization on Illumina Human HT-12 v4 Expression BeadChips (Illumina Inc., San Diego, CA).

**Clinical and Laboratory data analyses.** Median and interquartile ranges (IQR) were used as measures of central tendency. Frequencies were compared using the Pearson's chi-squared test. Continuous variables were compared using the Mann-Whitney U test (between two groups) or the Kruskal-Wallis test with Dunn's multiple comparisons (between >2 groups). Correlations were tested using the Spearman's rank correlation test. The p-values were adjusted for multiple comparisons using the Holm-Bonferroni's method<sup>55</sup>. Bayesian network learning was used to describe and visualize conditional dependencies between the multiple clinical and laboratory variables. Continuous variables were discretized by Hartemink's algorithm<sup>56</sup> accessed by "bnlearn" package in R 3.1.0<sup>57</sup>. The learning algorithm used to establish the network structure was based on the heuristic Hill climb method<sup>58</sup>. The dependencies represented qualitatively by a directed acyclic graph where each node corresponds to a variable and a direct arc between nodes represents a direct influence. Robustness of the arcs was scored by a non-parametric bootstrap test (100 × replicates)<sup>59</sup>. Arcs with more than 30% support were depicted. The strongest associations were considered those remaining statistically significant in ≥60% bootstraps. All analyses were pre-specified. Two-sided *P* value < 0.05 after adjustment for multiple comparisons were considered statistically significant. Statistical analyses were performed using SPSS 20.0 (IBM statistics), Graphpad Prism 6.0 (GraphPad Software, San Diego, CA) and JMP 12.0 (SAS, Cary, NC, USA).

**Transcriptome analyses: Data acquisition and processing.** RNA purity and integrity were evaluated by ND-1000 Spectrophotometer (NanoDrop, Wilmington, USA), Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, USA). Total RNA was amplified and purified using TargetAmp-Nano Labeling Kit for Illumina Expression BeadChip (EPICENTRE, Madison, USA) to yield biotinylated cRNA. Briefly, 300 ng of total RNA was reverse-transcribed to cDNA using a T7 oligo(dT) primer. Second-strand cDNA was synthesized, *in vitro* transcribed, and labeled with biotin-NTP. After purification, the cRNA was quantified on a ND-1000 Spectrophotometer (NanoDrop, Wilmington, USA). 750 ng of labeled cRNA samples were hybridized to each Human HT-12 v4.0 Expression Beadchip for 17 h at 58 °C, according to the manufacturer's instructions (Illumina, Inc., San Diego, USA). Array signal detection was carried out using Amersham fluorolink streptavidin-Cy3 (GE Healthcare Bio-Sciences, Little Chalfont, UK) following the bead array manual. Arrays were scanned with an Illumina BeadArray Reader confocal scanner. Hybridization quality and chip performance were monitored by visual inspection of internal quality control checks and the raw scanned data. Raw data were extracted using Illumina GenomeStudio v2011.1 software (Gene Expression Module v1.9.0). Outlier samples were identified and removed using the arrayQualityMetrics R package<sup>60</sup>. Remaining samples per groups were: 29 Healthy, 25 TB, 28 DM and 28 TBDM. Probes whose expression value was not significantly above the detection limit (*P* < 0.05) in at least 50% of all samples were removed. Data were log<sub>2</sub> transformed and corrected for batch effects using the ComBat algorithm implemented in the *sva* R package<sup>61</sup>. Probes matching the same gene symbol were collapsed by taking the one with highest expression across all samples. Mean and standard deviation for each gene using all samples were calculated and used to remove the 20% of the genes with lowest expression; and the 20% of the genes with lowest variance. The gene expression dataset used for subsequent analysis comprised 7,515 genes and 110 samples. Heat maps and "corrplots" were created using gplots, heatmap and corplot R packages.

**Transcriptome analyses: Differentially expressed gene analysis.** We defined 4 classes of samples: healthy, TB, DM and TBDM. Differential expression analyses were performed for each pair of classes using limma R package<sup>62</sup>. We considered DEGs as those genes having a limma adjusted *P*-value < 0.05 and an expression fold change > 1.4. Ingenuity Pathway Analysis (Ingenuity Systems, Qiagen, Hilden, Germany) was used to identify canonical pathways and upstream regulators which were significantly enriched in the DEG lists (*P* < 0.01).

**Transcriptome analyses: Calculating the molecular degree of perturbation.** The MDP method used for the present study is an adaptation of the MDH described by Pankla *et al.*<sup>31</sup>. Healthy controls were defined as the "reference" group, and the average expression level and standard deviation of this reference group were calculated for each gene. The MDP score of an individual gene (gMDP) in a given sample "s" was defined by taking the difference in expression level in sample "s" from the average of the gene in reference group divided by the corresponding standard deviation. Essentially, the gMDP score represents the number of standard deviations from reference. A filter was applied by replacing all absolute gMDP scores below 2 standard deviations by zero. Next, for each phenotype class (TB, DM and TBDM), we calculated the average gMDP score in the class, and used them to identify the top 25% genes with highest gMDP difference between the phenotype class and the reference group. Finally, the MDP of an individual sample (sMDP) was defined by taking the average absolute gMDP scores for all top 25% genes. This filter avoids absolute criteria while keeping the genes that best distinguish the control and test groups. To identify which sample was "perturbed", we used a threshold consisting of the average sMDP scores plus 2 standard deviations of the reference group. Any sample above this threshold was considered "perturbed" (compared to our reference).

**Transcriptome analyses: Single-sample Gene Set Enrichment Analysis.** Gene MDP scores were used to rank the genes for each participant. Gene Set Enrichment Analysis (GSEA)<sup>63</sup> was run on these pre-ranked individual gene lists using the Reactome pathways<sup>64</sup> as gene sets (GSEA, nominal *P* < 0.05; 1,000 permutations). GSEA calculated the normalized enrichment scores (NES) for the Reactome pathways based on the distribution of member genes of each module in the ranked list. We used the NES values to perform correlation analysis between two pathways or between genes and pathways. The network representing these correlations were created using Gephi.

**Transcriptome analyses: Whole blood 393-gene TB signature.** The microarray platform used for the current study included 387 genes present in a previously reported 393-gene transcriptional signature of active TB

(Berry *et al.*<sup>32</sup>). We used these we used these 387 genes as probes to cluster the samples in our cohort (Spearman correlation with average linkage).

**Integrative Analysis.** Sparse partial least squares regression (sPLS) method implemented in the sgPLS R package<sup>39</sup> was used to integrate gene expression and cytokine levels. Genes were used as ‘predictors’ and the cytokines as ‘respond variables.’ This method decomposes both predictor and response variation while identifying their correlation and the contributors to each variation component. We determined the minimum number of genes (chosen among the most informative 300 genes) of the first predictor component that best maximize the explained-cytokine variation. In this way, we identified 114 genes which were then correlated (Pearson correlation) with the cytokine profiles allowing to identify the way in which the best explanatory genes correlate with cytokines. GeneMania<sup>65</sup> was used to create an interaction network using as input the 114 genes. The information about “physical interactions”, “pathway”, “co-localization”, and “shared protein domains” was used to create the edges of the network. Gephi program was used to create the network. We then performed Spearman correlation between the expression of the 114 genes with the NES values (described in ‘Single-Sample Gene Set Enrichment Analysis’ section) of all Reactome pathways.

**Ethics Statement.** This study was conducted according to the principles expressed in the Declaration of Helsinki and approved by the Ethics Committee of the M.V. Diabetes Research Centre (ECR/51/INST/TN/2013/MVDRC/01). Informed consent was obtained in writing from all participants.

## References

1. Harries, A. D. *et al.* The looming epidemic of diabetes-associated tuberculosis: learning lessons from HIV-associated tuberculosis. *Int. J. Tuberc. Lung Dis.* **15**, 1436–44, doi:10.5588/ijtld.11.0503 (2011).
2. Martens, G. W. *et al.* Tuberculosis susceptibility of diabetic mice. *Am. J. Respir. Cell Mol. Biol.* **37**, 518–524, doi:10.1165/rcmb.2006-0478OC (2007).
3. Vallerskog, T., Martens, G. W. & Kornfeld, H. Diabetic Mice Display a Delayed Adaptive Immune Response to Mycobacterium tuberculosis. *J. Immunol.* **184**, 6275–6282, doi:10.4049/jimmunol.1000304 (2010).
4. Martinez, N. *et al.* Chromatin decondensation and T cell hyperresponsiveness in diabetes-associated hyperglycemia. *J. Immunol.* **193**, 4457–4468, doi:10.4049/jimmunol.1401125 (2014).
5. Sherman, L. F., Fujiwara, P. I., Cook, S. V., Bazerman, L. B. & Frieden, T. R. Patient and health care system delays in the diagnosis and treatment of tuberculosis. *Int. J. Tuberc. Lung Dis.* **3**, 1088–1095 (1999).
6. Viswanathan, V. *et al.* Prevalence of diabetes and pre-diabetes and associated risk factors among tuberculosis patients in India. *PLoS One* **7**, e41367, doi:10.1371/journal.pone.0041367 (2012).
7. 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**, 20130427–20130427, doi:10.1098/rstb.2013.0427 (2014).
8. 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, doi:10.1016/j.chest.2016.02.675 (2016).
9. Thacher, T. D. & Clarke, B. L. Vitamin D insufficiency. *Mayo Clin. Proc.* **86**, 50–60, doi:10.4065/mcp.2010.0567 (2011).
10. Abakay, O., Abakay, A., Sen, H. S. & Tanrikulu, A. C. The relationship between inflammatory marker levels and pulmonary tuberculosis severity. *Inflammation* **38**, 691–696, doi:10.1007/s10753-014-9978-y (2015).
11. Lou, M. *et al.* Relationship between neutrophil-lymphocyte ratio and insulin resistance in newly diagnosed type 2 diabetes mellitus patients. *BMC Endocr. Disord.* **15**, 9, doi:10.1186/s12902-015-0002-9 (2015).
12. Guo, X. *et al.* Neutrophil-lymphocyte ratio is positively related to type 2 diabetes in a large-scale adult population: a Tianjin Chronic Low-Grade Systemic Inflammation and Health cohort study. *Eur. J. Endocrinol.* **173**, 217–225, doi:10.1530/EJE-15-0176 (2015).
13. Zhan, Y., Xu, T. & Tan, X. Two parameters reflect lipid-driven inflammatory state in acute coronary syndrome: atherogenic index of plasma, neutrophil-lymphocyte ratio. *BMC Cardiovasc. Disord.* **16**, 96, doi:10.1186/s12872-016-0274-7 (2016).
14. Kim, B. J. *et al.* The combined impact of neutrophil-to-lymphocyte ratio and type 2 diabetes mellitus on significant coronary artery disease and carotid artery atherosclerosis. *J. Cardiovasc. Ultrasound* **24**, 115–122, doi:10.4250/jcu.2016.24.2.115 (2016).
15. Cetin, M. S. *et al.* Monocyte to HDL cholesterol ratio predicts coronary artery disease severity and future major cardiovascular adverse events in acute coronary syndrome. *Heart Lung Circ* **25**, 1077–1086, doi:10.1016/j.hlc.2016.02.023 (2016).
16. Deforche, K. *et al.* Analysis of HIV-1 pol sequences using Bayesian Networks: implications for drug resistance. *Bioinformatics* **22**, 2975–2979, doi:10.1093/bioinformatics/btl508 (2006).
17. Restrepo, B. I. *et al.* Tuberculosis in poorly controlled type 2 diabetes: altered cytokine expression in peripheral white blood cells. *Clin. Infect. Dis.* **47**, 634–641, doi:10.1086/592125 (2008).
18. Kumar, N. P. *et al.* Expansion of pathogen-specific T-helper 1 and T-helper 17 cells in pulmonary tuberculosis with coincident type 2 diabetes mellitus. *J. Infect. Dis.* **208**, 739–748, doi:10.1093/infdis/jit241 (2013).
19. 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, doi:10.1513/AnnalsATS.201305-112OC (2013).
20. Wong, T. Y., Cheung, C. M., Larsen, M., Sharma, S. & Simo, R. Diabetic retinopathy. *Nat. Rev. Dis. Primers* **2**, 16012, doi:10.1038/nrdp.2016.12 (2016).
21. McAuley, A. K. *et al.* Vitreous biomarkers in diabetic retinopathy: a systematic review and meta-analysis. *J. Diabetes Complications* **28**, 419–425, doi:10.1016/j.jdiacomp.2013.09.010 (2014).
22. Yilmaz, M. I. *et al.* Ramipril lowers plasma FGF-23 in patients with diabetic nephropathy. *Am. J. Nephrol.* **40**, 208–214, doi:10.1159/000366169 (2014).
23. Maertzdorf, J. *et al.* Concise gene signature for point-of-care classification of tuberculosis. *EMBO Mol. Med.* **8**, 86–95, doi:10.15252/emmm.201505790 (2016).
24. Zak, D. E. *et al.* A blood RNA signature for tuberculosis disease risk: a prospective cohort study. *Lancet* **387**, 2312–2322, doi:10.1016/S0140-6736(15)01316-1 (2016).
25. Back, S. H. & Kaufman, R. J. Endoplasmic reticulum stress and type 2 diabetes. *Annu. Rev. Biochem.* **81**, 767–793, doi:10.1146/annurev-biochem-072909-095555 (2012).
26. Cnop, M., Ladrerie, L., Igoillo-Esteve, M., Moura, R. F. & Cunha, D. A. Causes and cures for endoplasmic reticulum stress in lipotoxic beta-cell dysfunction. *Diabetes Obes. Metab.* **2**, 76–82, doi:10.1111/j.1463-1326.2010.01279.x (2010).
27. De Angelis, P. M., Svendsrud, D. H., Kravik, K. L. & Stokke, T. Cellular response to 5-fluorouracil (5-FU) in 5-FU-resistant colon cancer cell lines during treatment and recovery. *Mol. Cancer* **5**, 20, doi:10.1186/1476-4598-5-20 (2006).
28. Peiris-Pages, M., Sotgia, F. & Lisanti, M. P. Doxycycline and therapeutic targeting of the DNA damage response in cancer cells: old drug, new purpose. *Oncoscience* **2**, 696–699, doi:10.18632/oncoscience.215 (2015).
29. Mahajan, K. & Mahajan, N. P. Cross talk of tyrosine kinases with the DNA damage signaling pathways. *Nucleic Acids Res* **43**, 10588–10601, doi:10.1093/nar/gkv1166 (2015).

30. Tatsch, E. *et al.* Oxidative DNA damage is associated with inflammatory response, insulin resistance and microvascular complications in type 2 diabetes. *Mutat. Res.* **782**, 17–22, doi:10.1016/j.mrfmmm.2015.10.003 (2015).
31. Pankla, R. *et al.* Genomic transcriptional profiling identifies a candidate blood biomarker signature for the diagnosis of septicemic melioidosis. *Genome Biol.* **10**, R127, doi:10.1186/gb-2009-10-11-r127 (2009).
32. Berry, M. P. *et al.* An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature* **466**, 973–977, doi:10.1038/nature09247 (2010).
33. Ralph, A. P. *et al.* A simple, valid, numerical score for grading chest x-ray severity in adult smear-positive pulmonary tuberculosis. *Thorax* **65**, 863–869, doi:10.1136/thx.2010.136242 (2010).
34. Luna, P., Guarner, V., Farias, J. M., Hernandez-Pacheco, G. & Martinez, M. Importance of metabolic memory in the development of vascular complications in diabetic patients. *J. Cardiothorac. Vasc. Anesth.* **30**, 1369–1378, doi:10.1053/j.jvca.2016.02.008 (2016).
35. van den Berg, S. M. *et al.* Blocking CD40-TRAF6 interactions by small-molecule inhibitor 6860766 ameliorates the complications of diet-induced obesity in mice. *Int. J. Obes. (Lond)* **39**, 782–790, doi:10.1038/ijo.2014.198 (2015).
36. Zbidi, H., Jardin, I., Bartegi, A., Salido, G. M. & Rosado, J. A. Ca<sup>2+</sup> leakage rate from agonist-sensitive intracellular pools is altered in platelets from patients with type 2 diabetes. *Platelets*. **22**, 284–293, doi:10.3109/09537104.2010.528813 (2011).
37. Ramasamy, R., Yan, S. F. & Schmidt, A. M. Receptor for AGE (RAGE): signaling mechanisms in the pathogenesis of diabetes and its complications. *Ann. N. Y. Acad. Sci.* **1243**, 88–102, doi:10.1111/j.1749-6632.2011.06320.x (2011).
38. Ghosh, P., Sahoo, R., Vaidya, A., Chorev, M. & Halperin, J. A. Role of complement and complement regulatory proteins in the complications of diabetes. *Endocr. Rev.* **36**, 272–288, doi:10.1210/er.2014-1099 (2015).
39. Liquet, B., de Micheaux, P. L., Hejblum, B. P. & Thiebaut, R. Group and sparse group partial least square approaches applied in genomics context. *Bioinformatics*. **32**, 35–42, doi:10.1093/bioinformatics/btv535 (2016).
40. Lonnroth, K., Roglic, G. & Harries, A. D. Improving tuberculosis prevention and care through addressing the global diabetes epidemic: from evidence to policy and practice. *Lancet Diabetes Endocrinol* **2**, 730–39, doi:10.1016/S2213-8587(14)70109-3 (2014).
41. Chen, L., Magliano, D. J. & Zimmet, P. Z. The worldwide epidemiology of type 2 diabetes mellitus—present and future perspectives. *Nat. Rev. Endocrinol* **8**, 228–236, doi:10.1038/nrendo.2011.183 (2012).
42. World Health Organization 2015. *Global Tuberculosis Report 2015* WHO Press, Geneva (2015).
43. Martinez, N. & Kornfeld, H. Diabetes and immunity to tuberculosis. *Eur. J. Immunol.* **44**, 617–626, doi:10.1002/eji.v44.3 (2014).
44. 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, doi:10.1093/infdis/jiw436 (2016).
45. Lowe, D. M., Redford, P. S., Wilkinson, R. J., O'Garra, A. & Martineau, A. R. Neutrophils in tuberculosis: friend or foe? *Trends Immunol.* **33**, 14–25, doi:10.1016/j.it.2011.10.003 (2012).
46. Deffur, A., Wilkinson, R. J. & Coussens, A. K. Tricks to translating TB transcriptomics. *Ann. Transl. Med* **3**, S43, doi:10.3978/j.issn.2305-5839.2015.04.12 (2015).
47. Lonnroth, K., Williams, B. G., Cegielski, P. & Dye, C. A consistent log-linear relationship between tuberculosis incidence and body mass index. *Int. J. Epidemiol.* **39**, 149–155, doi:10.1093/ije/dyp308 (2010).
48. Kaji, H. High-density lipoproteins and the immune system. *J. Lipids* **2013**, 684903–8, doi:10.1155/2013/684903 (2013).
49. Kwan, C. K. & Ernst, J. D. HIV and tuberculosis: a deadly human syndemic. *Clin. Microbiol. Rev.* **24**, 351–376, doi:10.1128/CMR.00042-10 (2011).
50. Reed, G. W. *et al.* Impact of diabetes and smoking on mortality in tuberculosis. *PLoS. One* **8**, e58044, doi:10.1371/journal.pone.0058044 (2013).
51. Nickerson, H. D. & Dutta, S. Diabetic complications: current challenges and opportunities. *J. Cardiovasc. Transl. Res.* **5**, 375–379, doi:10.1007/s12265-012-9388-1 (2012).
52. Andrade, B. B. *et al.* Heightened plasma levels of heme oxygenase-1 and tissue inhibitor of metalloproteinase-4 as well as elevated peripheral neutrophil counts are associated with TB-diabetes comorbidity. *CHEST*. **145**, 1244–1254, doi:10.1378/chest.13-1799 (2014).
53. American Diabetes Association. Classification and diagnosis of diabetes. In: 2016 Standards of Medical Care in Diabetes. *Diabetes Care* **39**, S13–22, doi:10.2337/dc16-S005 (2016).
54. World Health Organization. *Definition and Diagnosis of Diabetes Mellitus and Intermediate Hyperglycaemia* WHO Press, Geneva (2006).
55. Holm, S. A simple sequentially rejective multiple test procedure. *Scand. J. Statist* **6**, 65–70 (1979).
56. Hartemink, A. J., Gifford, D. K., Jaakkola, T. S., Young, R. A. Using graphical models and genomic expression data to statistically validate models of genetic regulatory networks. *Pac. Symp. Biocomput.* 422–433 (2001).
57. Su, C., Andrew, A., Karagas, M. R. & Borsuk, M. E. Using Bayesian networks to discover relations between genes, environment, and disease. *BioData Mining* **6**, 6, doi:10.1186/1756-0381-6-6 (2013).
58. Hesar, A. S., Tabatabaee, H. & Jalali, M. *Structure learning of Bayesian networks using heuristic methods* IACSIT Press, Singapore (2012).
59. Friedman, N. G. M. & Goldszmidt, M. W. A. Data analysis with bayesian networks: a bootstrap approach. *Elsevier Science B.V. Amsterdam*. 196–205 (1999).
60. Kauffmann, A., Gentleman, R. & Huber, W. arrayQualityMetrics—a bioconductor package for quality assessment of microarray data. *Bioinformatics* **25**, 415–416, doi:10.1093/bioinformatics/btn647 (2009).
61. 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, doi:10.1093/bioinformatics/bts034 (2012).
62. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47–e47, doi:10.1093/nar/gkv007 (2015).
63. Subramanian, A., Kuehn, H., Gould, J., Tamayo, P. & Mesirov, J. P. GSEA-P: a desktop application for Gene Set Enrichment Analysis. *Bioinformatics*. **23**, 3251–3253, doi:10.1093/bioinformatics/btm369 (2007).
64. Fabregat, A. *et al.* The Reactome pathway Knowledgebase. *Nucleic Acids Res* **44**, D481–D487, doi:10.1093/nar/gkv1351 (2016).
65. Warde-Farley, D. *et al.* The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res* **38**, W214–W220, doi:10.1093/nar/gkq537 (2010).

## Acknowledgements

We thank the EDOTS study volunteers who participated in this study, the staff of the Prof. M. Viswanathan Diabetes Research Centre, the M.V. Hospital for Diabetes, the National Institute for Research in Tuberculosis and the District Tuberculosis Officer Dr. J. Lavanya and staff of the Revised National Tuberculosis Control Program in Chennai, India. This work was supported by the following funding sources: grant #USB1-31149-XX-13 (to V. Viswanathan and H. Kornfeld) jointly sponsored by the Indian Department of Biotechnology, the Indian Council of Medical Research and the National Institute for Allergy and Infectious Diseases, National Institutes of Health, and administered by CRDF Global.; FAPESP grant # 2013/08216-2 (to H. Nakaya). H. Nakaya receives a CNPq research fellowship.

### Author Contributions

Designed the study (S.B., V.V., B.B.A., H.K.); conducted experiments (N.P.K.); acquired data (N.P.K., K.W., S.S., P.A.M.); analyzed data (C.A.P.-M., K.F., L.G.-S., F.L., B.B.A., H.I.N.); wrote the manuscript (B.B.A., H.I.N., H.K.).

### Additional Information

**Supplementary information** accompanies this paper at doi:[10.1038/s41598-017-01767-4](https://doi.org/10.1038/s41598-017-01767-4)

**Competing Interests:** The authors declare that they have no competing interests.

**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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017