



Universiteit  
Leiden  
The Netherlands

## **Tuberculosis and diabetes: biomarkers and drug candidates from a host perspective**

Doorn, C.L.R. van

### **Citation**

Doorn, C. L. R. van. (2022, November 16). *Tuberculosis and diabetes: biomarkers and drug candidates from a host perspective*. Retrieved from <https://hdl.handle.net/1887/3486320>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3486320>

**Note:** To cite this publication please use the final published version (if applicable).



## Chapter 2

# Impaired resolution of blood transcriptomes through tuberculosis treatment with diabetes comorbidity

Clare Eckold\*, Cassandra L.R. van Doorn\*, Rovina Ruslami, Katharina Ronacher, Anca-Lelia Costache, Suzanne van Veen, Ji-Sook Lee, Vinod Kumar, Sarah Kerry-Barnard, Stephanus T. Malherbe, Léanie Kleynhans, Kim Stanley, Simone A. Joosten, Julia A. Critchley, Philip C. Hill, Reinout van Crevel, Cisca Wijmenga, Mariëlle C. Haks, Mihai Loana, Bachtı Alisjahbana, Gerhard Walzl, Tom H.M. Ottenhoff, Hazel M. Dockrell, Eleonora Vianello<sup>#</sup> and Jacqueline M. Cliff<sup>#</sup>, on behalf of the TANDEM Consortium

*\*These authors contributed equally*

*<sup>#</sup>These authors contributed equally*

MedRxiv. 2022.02.07.22269422

## **Abstract**

### **Background**

People with diabetes are more likely to develop tuberculosis (TB) and to have poor TB treatment outcomes than those without. We previously showed that blood transcriptomes in people with TB-diabetes (TB-DM) co-morbidity have excessive inflammatory and reduced interferon responses at TB diagnosis. It is unknown whether this persists through treatment, potentially underlying adverse outcomes.

### **Methods**

Pulmonary TB patients were recruited in South Africa, Indonesia and Romania, and classified as having TB-DM, TB with prediabetes, TB-related hyperglycaemia or uncomplicated TB, based on glycated haemoglobin (HbA1c) concentration at TB diagnosis and after 6 months of TB treatment. Gene expression in blood samples collected at diagnosis and at regular intervals throughout treatment was measured by unbiased RNASeq and targeted Multiplex Ligation-dependent Probe Amplification.

### **Results**

Gene expression was modulated by TB treatment in all groups but to different extents, such that differences remained in people with TB-DM relative to TB-only throughout, including genes involved in innate responses, anti-microbial immunity and the inflammasome. People with prediabetes or with TB-related hyperglycaemia had gene expression more similar to people with TB-DM than TB-only throughout treatment. The overall pattern of change was similar across clinical groups irrespective of glycaemic index, permitting models predictive of TB treatment to be developed.

### **Conclusions**

The exacerbated transcriptome changes seen in TB-DM take longer to resolve during TB treatment, indicating that prolonged treatment or host-directed therapy may be needed to improve TB treatment outcomes. Development of transcriptome-based biomarker signatures of TB-treatment response should include people with diabetes to be useful across populations.

## Introduction

Diabetes mellitus (DM) negatively impacts TB control by increasing the risk of *Mycobacterium tuberculosis* infection [1] and of progression to active TB disease three-fold [2, 3]. The growing prevalence of DM, particularly in countries with high TB burden, means DM now underlies around 15% of TB cases globally [4], accounting for 10% of TB deaths in HIV-negative people. Concomitant DM negatively impacts TB-treatment outcomes, with increased risk of delayed sputum conversion, relapse, treatment failure and death: relative risk for each poor outcome is ~2 to ~5 in meta-analyses [5, 6]. It is unknown whether extending standard TB treatment would improve outcome for TB-DM comorbid patients, or whether alternative treatment is required, such as host-directed therapy.

Worldwide DM prevalence is ~463 million people, estimated to rise to 700 million by 2045 [7]. The majority of people have type-2 DM, caused by a reduction in insulin's ability to control target cell metabolism triggering an increase in insulin production, pancreatic damage through exhaustion, and impaired glucose tolerance. There is a spectrum from normal through to full DM via intermediate hyperglycaemia (IH). People with IH are more likely to develop DM in the future [8]. HbA1c concentration can indicate an individual's position on this spectrum [8]. Infectious diseases, including TB, can cause temporary stress hyperglycaemia, which carries a higher risk of adverse events than longer-term pre-diabetes [9]. TB-induced stress hyperglycaemia also makes DM diagnosis difficult: some people with apparent newly-diagnosed DM at TB diagnosis no longer reach DM diagnostic criteria after TB treatment [10]. TB incidence and TB-DM treatment outcomes are worse in people with poorly controlled DM with higher HbA1c concentrations [11].

People with TB-DM comorbidity have altered immunity compared to uncomplicated TB, with both innate and adaptive immune responses affected [12]. In plasma, various inflammatory cytokines such as IL-1 $\beta$ , IL-17A, IFN $\gamma$  and TNF $\alpha$  are more elevated in TB-DM [13, 14] and TB-pre-diabetes [15] than in uncomplicated TB. People with TB-DM have more circulating Th1 and Th17 cells and fewer Tregs. In uncomplicated TB, peripheral immune responses typically resolve to normal levels during successful TB treatment [14]. In contrast, the excessive inflammatory plasma cytokine responses in TB-DM are still evident after treatment completion [16], and dendritic cell, monocyte [17] and T cell differentiation [18] aberrations are still present at 2 months, although resolved by 6 months, indicating a delayed response to TB treatment in TB-DM patients.

Transcriptomic technologies have delineated altered peripheral immunity in TB in multiple studies, revealing an enhanced circulating inflammatory and type 1 interferon response [19-22]. With successful TB treatment, the transcriptomic signature is rapidly downregulated, has largely diminished after two months of treatment and mostly

disappears by 12 months [20, 22, 23], mirroring clinical resolution and chest X-ray improvement; however transcriptomes do not fully resolve with poor TB treatment outcome [24]. We recently showed [25] that DM comorbidity, as well as IH, significantly affects the TB diagnosis biosignature, causing an enhanced inflammatory but reduced interferon response. The impact of DM on transcriptomes through TB treatment has not been described. The aim of this study was to determine whether transcriptomic biosignatures resolve normally in TBDM, or whether changes during TB treatment are kinetically or qualitatively different to those in uncomplicated TB.

## Methods

### Patient recruitment and classification.

Newly diagnosed patients with bacteriologically confirmed pulmonary TB, with or without concomitant DM, were recruited in three locations: Bandung, Indonesia (UNPAD), Cape Town, South Africa (SUN) and Craiova, Romania (UMFCV), as part of the TANDEM project [26]. Exclusion criteria were multi-drug-resistant TB, HIV positivity, pregnancy, other serious co-morbidity or corticosteroid use. All participants gave written informed consent. The study was approved by LSHTM Observational Research Ethics Committee (6449/July2013), SUN Health Research Ethics Committee (N13/05/064/July2013), UNPAD Health Research Ethics Committee (377/UN6.C2.1.2/ KEPK/PN/2012), and UMFCV Committee of Ethics and Academic and Scientific Deontology (94/September2013).

All TB patients underwent first line TB treatment according to WHO guidelines. Most patients diagnosed with DM received local standard of care treatment, and medication taken was noted. A TB-DM subgroup within the Indonesian cohort had intensive HbA1c monitoring as part of a pragmatic randomised control trial, with DM medication changed accordingly [27]. Participants were classified by DM/glycaemia status at TB diagnosis and after 6 months TB treatment (Supplementary Table S1). The “TB-DM” group included patients with both pre-existing and newly diagnosed DM (Supplementary Table S2) Newly diagnosed TB-DM had laboratory HbA1c test  $\geq 6.5\%$  with confirmatory HbA1c test  $\geq 6.5\%$  or fasting blood glucose  $\geq 7$  mmol/L at TB diagnosis [26, 28], followed by a further HbA1c test  $\geq 6.5\%$  after 6 months of TB treatment. Patients whose HbA1c test results were  $\geq 5.7\%$  and  $< 6.5\%$  at both TB diagnosis and at 6 months were deemed to have pre-diabetes (“TB-preDM”). Patients whose HbA1c result was  $\geq 5.7\%$  at TB diagnosis but below  $< 5.7\%$  at 6 months were deemed to have TB-related intermediate hyperglycaemia at TB diagnosis (“TBrel-IH”). All groups were evenly sex balanced, except for male predominance in TB-PreDM. Age ranges were similar across groups. In the Indonesian TB-DM group, there was a highly significant decrease in HbA1c through TB treatment, likely due to intensive DM follow-up (Supplementary Figure S1 & Supplementary Table S1); this was not evident in

South Africa or Romania.

### **Sample collection and RNA extraction**

Venous blood samples (2.5ml) were collected into PAXgene Blood RNA Tubes (PreAnalytiX) from TB patients prior to TB treatment initiation (W0) and at intervals through treatment (W2,4,8,16,26) up to 12 months post diagnosis (W52), and stored at -80oC prior to analysis. Total RNA was extracted using RNeasy spin columns (Qiagen) and quantified by Nanodrop (Agilent).

### **Unbiased whole genome RNA-Seq**

A detailed description of RNA-Seq analysis is given in Supplementary Methods. Samples were processed using the poly-A tail Bioscientific NEXTflex-Rapid-Directional mRNA-seq method and single-end sequenced. Longitudinal differential gene expression analyses were performed on normalised data using the MaSigPro (v1.62.0) [29] package in R. MaSigPro follows a two-step regression method to find genes with significant temporal expression changes and significant differences between groups. Modular analysis was performed on genes that were differentially expressed between clinical groups.

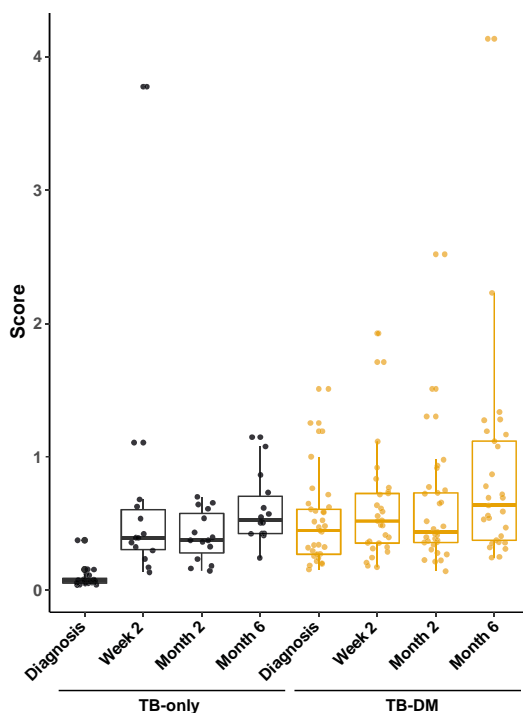
### **Targeted gene expression profiling**

Targeted gene expression profiling was performed using dual-color Reverse-Transcriptase Multiplex Ligation-dependent Probe Amplification (dcRT-MLPA) [30]: a detailed description is given in Supplementary Methods. Primers and half-probes were designed for 4 housekeeping genes and 144 selected key immune TB-related genes (Supplementary Table S3). Longitudinal differentially expressed genes (DEGs) within groups were identified using linear mixed models. Signatures with the best discriminatory capability were identified using logistic regression with lasso regularization (glmnet R package).

## **Results**

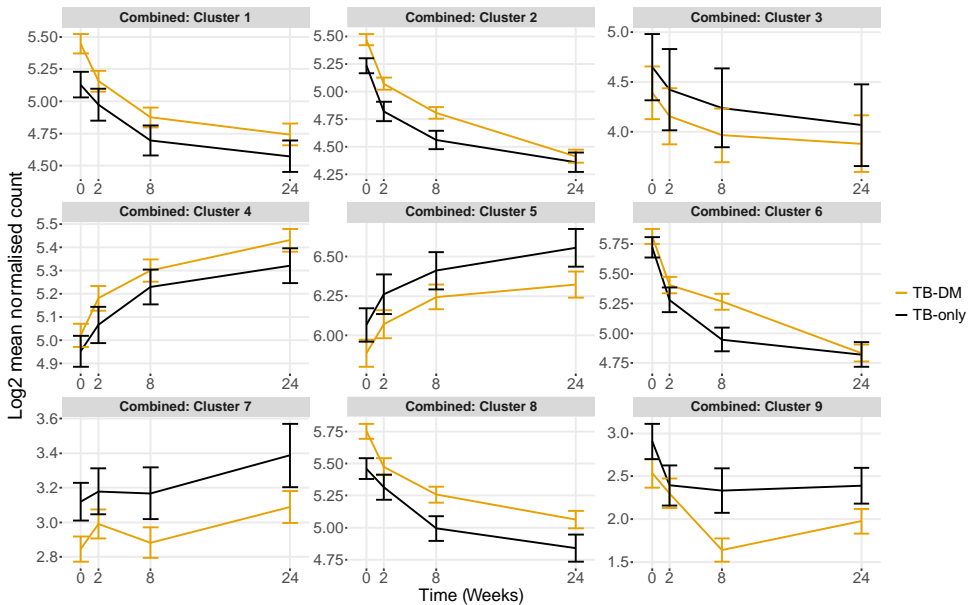
### **Global longitudinal transcriptomes in TB-DM**

Gene expression was determined in venous blood by RNA-Seq in a subgroup of study participants from the four TB patient clinical groups (Supplementary Table S1; Supplementary Figure S1). The molecular degree of perturbation (MDP) of gene expression in individual samples from patients with TB-only or TB-DM over time was calculated relative to mean gene expression at diagnosis in TB-only (Figure 1). As expected, there were gene expression changes during TB treatment in the TB-only group, evident by week 2 and continuing throughout treatment. Global gene expression was perturbed in patients with TB-DM relative to TB-only at TB diagnosis, and while there was some resolution through time, the changes were reduced such that global gene expression in the TB-DM group remained different to the TB-only group at all time points (Figure 1).



**FIGURE 1 | Molecular Degree of Perturbation Plots representing change in global gene expression in blood relative to patients with TB-only at TB diagnosis.** Gene expression was determined by RNA-Seq of whole venous blood from pulmonary TB patients from all three clinical locations with (TB-DM: n=34) or without (TB-only: n= 18) concomitant diabetes, at TB diagnosis and during TB treatment. The bars show the median and 1.5\*inter-quartile range.

The MaSigPro analysis package identified 167 genes with significantly different changed expression between TB-DM and TB-only groups through TB treatment, in the combined dataset from South Africa, Indonesia and Romania. Hierarchical clustering of these genes based on similar expression patterns yielded 9 clusters (Figure 2; Supplementary Table S4). Clusters which were more highly expressed in TB-DM patients throughout treatment (clusters 1,2,4&8) were enriched for genes involved in the innate immune response, IL-4 signalling, protein dimerisation and neutrophil chemotaxis, determined using the DAVID Functional Annotation Tool [31] (Table 1). Cluster 6 exhibited divergence between TB and TB-DM patients only at week 8 of treatment: this cluster was enriched for genes involved in anti-viral and IFN signalling responses. Clusters more highly expressed in TB-only patients (clusters 5,7,9) were smaller and enriched for alternative splice variants and immunoglobulins.



**FIGURE 2 |** MaSigPro analysis of change in gene expression through TB treatment in blood samples from patients in all 3 populations combined (South Africa, Indonesia and Romania). MaSigPro identified genes that behave similarly between patient groups using hierarchical clustering. Results are shown for log-transformed normalised count for the TB-only group or TB-DM. Bars show mean  $\pm$  1 SEM. Data were filtered to remove lowly abundant transcripts prior to analysis.

### Aberrant longitudinal transcriptomes in TB patients with intermediate hyperglycaemia

Previously [25] we showed gene expression in TBrel-IH is more similar to people with diagnosed DM than with TB-only at TB diagnosis. We repeated the MaSigPro analyses separately for South Africa and Indonesia, combining all those patients with raised glycaemic indices but not DM at TB diagnosis, to determine how transcriptomes changed through TB treatment in intermediate groups (Supplementary Figure S2). In South Africa, the analysis resulted in 1,179 transcripts separated into three hierarchical clusters, which changed through treatment differently across clinical groups (Supplementary Figure S2A; Supplementary Table S5), with the combined intermediate group behaving more similarly to TB-DM. Similar results were obtained with the Indonesian cohort, with 2,354 transcripts across 4 hierarchical clusters behaving differently between clinical groups (Supplementary Figure S2B; Supplementary Table S6).

A core list of 102 genes overlapped between MaSigPro analyses for the combined cohort from Romania, South Africa and Indonesia, and from the latter two populations separately (Supplementary Figure S3; Supplementary Table S7). Gene ontology and pathway analyses of this core list using the g:profiler webtool revealed functional enrichment of genes involved in the immune response, response to biotic

**TABLE 1 | Clusters of genes differentially expressed between TB-DM and TB-only patients in MaSigPro analysis of the combined RNA-Seq dataset from South Africa, Indonesia and Romania.**

| Cluster | Overall pattern                                    | Number of transcripts | Gene Function  |                      |             |                              | Top Non-redundant Functions from DAVID <sup>#</sup>              |
|---------|--|-----------------------|----------------|----------------------|-------------|------------------------------|--|
|         |  |                       | Protein Coding | Processed Transcript | Pseudo-gene | Regulatory RNAs <sup>*</sup> |  |
| 1       | Higher in TB-DM throughout; decreasing with time   | 27                    | 19             | 0                    | 2           | 6                            | Innate Immunity<br>Antimicrobial<br>RAGE receptor binding        |
| 2       | Higher in TB-DM throughout; decreasing with time   | 50                    | 28             | 5                    | 4           | 13                           | IPAF inflammasome<br>IL-4 signalling<br>Transmembrane helices    |
| 3       | Lower in TB-DM throughout; decreasing with time    | 6                     | 6              | 0                    | 0           | 0                            | Disulphide bond<br>Inflammation/fibrosis                         |
| 4       | Higher in TB-DM throughout; increasing with time   | 17                    | 11             | 3                    | 0           | 3                            | Coiled coil<br>Protein homo-dimerisation                         |
| 5       | Lower in TB-DM throughout; increasing with time    | 9                     | 8              | 1                    | 0           | 0                            | Collagen-binding<br>Alternative splicing<br>phosphoprotein       |
| 6       | Higher in TB-DM at week 8, otherwise similar       | 28                    | 22             | 1                    | 0           | 5                            | GTPase activity<br>Anti-viral defence<br>IFN $\gamma$ signalling |
| 7       | Lower in TB-DM throughout; rising end treatment    | 10                    | 4              | 1                    | 0           | 5                            | Splice variant   |
| 8       | Higher in TB-DM throughout; decreasing with time   | 16                    | 13             | 2                    | 0           | 1                            | Secreted<br>Neutrophil chemotaxis<br>Transmembrane helix         |
| 9       | Lower in TB-DM at week 8 and 24, otherwise similar | 4                     | 3              | 0                    | 0           | 1                            | Immunoglobulins  |

<sup>#</sup>DAVID [31] analysis of GO terms BP, MF, CC direct UP-Keywords

<sup>\*</sup>Retained introns, Antisense, LncRNA, miRNA, nonsense-mediated decay, sense overlapping, sense intronic, snoRNA,

stimuli, and gene products localising to intracellular vesicles (Supplementary Figure S4). We hypothesised there would be differences between the TB-preDM and TBrel-IH groups, due to persistence or resolution of hyperglycaemia through TB treatment; however, both TB-preDM and TBrel-IH patients responded more similarly to TB-DM than to TB-only patients (Figure 3). Longitudinal mixed effects model analysis of mean expression within core gene list clusters showed highly significant changes across all four clinical groups throughout treatment, with differences between the groups in larger gene clusters (Supplementary Table S8). Importantly, there was no interaction between clinical group and time, showing there was resolution of expression in all groups through treatment, albeit from different starting points and at different rates.

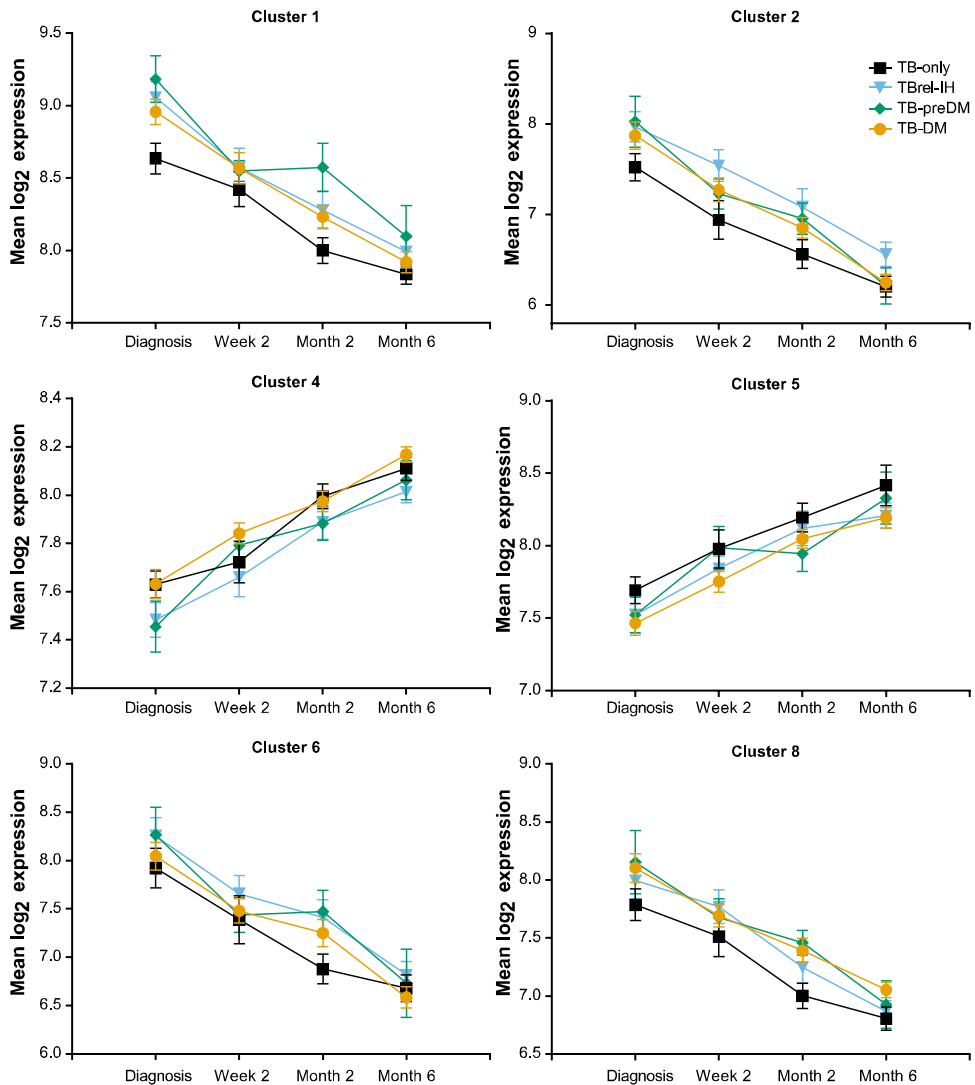
### **Modular analysis of differentially expressed genes (DEGs)**

DEGs identified in MaSigPro analyses were used in modular analyses to understand biological differences between clinical groups in South Africa and Indonesia (Supplementary Table 9 and 10 respectively). The most statistically significant modules were investigated further by calculating their modular activity in TB-DM relative to TB-only through time. The top module in both populations was immune activation, which was upregulated in TB-DM compared to TB-only throughout treatment. In both populations, different modules fluctuated between TB-DM and TB, and behaved inversely to one another through treatment (Figure 4).

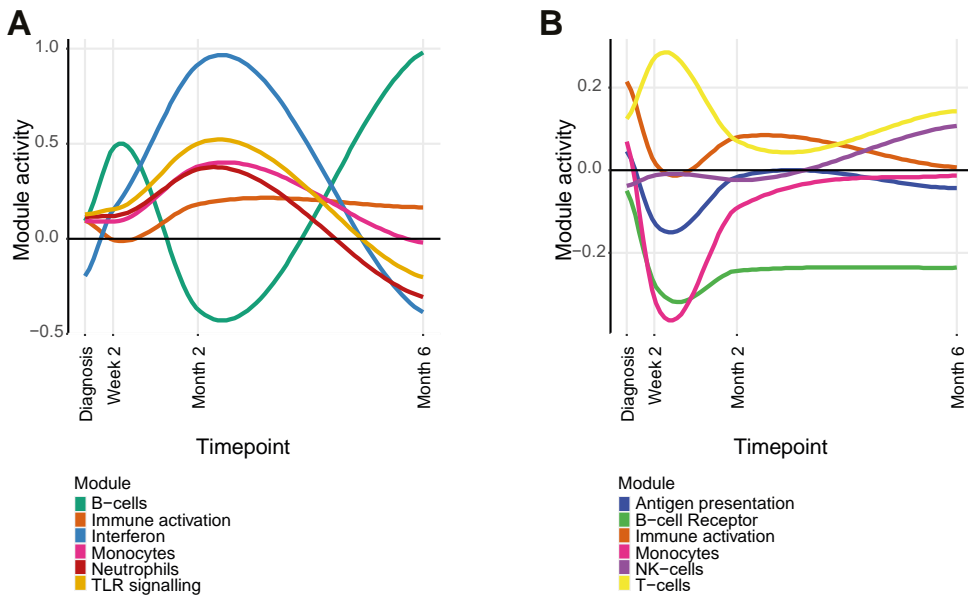
### **Impact of DM on TB treatment response using targeted gene expression profiling**

We performed targeted profiling of TB-relevant immune gene expression in an expanded cohort from South Africa with more intensive sampling, using dcRT-MLPA (Supplementary Table S11A). At baseline, the overall MDP was similar in all study groups compared to healthy controls (Supplementary Figure S5A); Partial Least Squares – Discriminant Analysis separated DM-only patients from other groups including healthy controls, suggesting distinct genes are perturbed in DM-only (Supplementary Figure S5B; [25]). Gene expression was strongly correlated between TB-only and TB-DM, TB-preDM or TBrel-IH, but with some outlier genes which were affected by glycaemic status (Figure 5A). There were more DEGs relative to healthy controls in the TB-DM, TB-preDM and TBrel-IH groups than TB-only, at TB diagnosis and throughout treatment (Figure 5B). Normalisation of expression of genes such as *GNLY* and *GBP1* occurred by 2 weeks in the TB-only group but was delayed in TBDM, TB-preDM and TBrel-IH. Results from targeted dcRT-MLPA analysis were thus in accordance with the global RNA-Seq analysis.

Longitudinal MDP analysis in the South African and Indonesian cohorts indicated the magnitude of transcriptomic response to TB treatment was dependent on diabetes/glycaemia with TB-DM patients displaying the largest gene expression perturbation over time (Figure 6A, Supplementary Figure S6A). Gene expression changes



**FIGURE 3 | Gene expression through treatment in TB patients with Pre-diabetes or TB-related intermediate hyperglycaemia, relative to TB-DM and TB-only patients.** The expression of genes in the Core 102 genelist (Supplementary Table S7) was summed for those genes within each MaSigPro gene cluster (Figure 2) for individual patients (log<sub>2</sub> scale). Only MaSigPro clusters with > 3 genes in the core gene list are shown. Points show the mean  $\pm$  SEM for each of the four clinical groups at each timepoint.

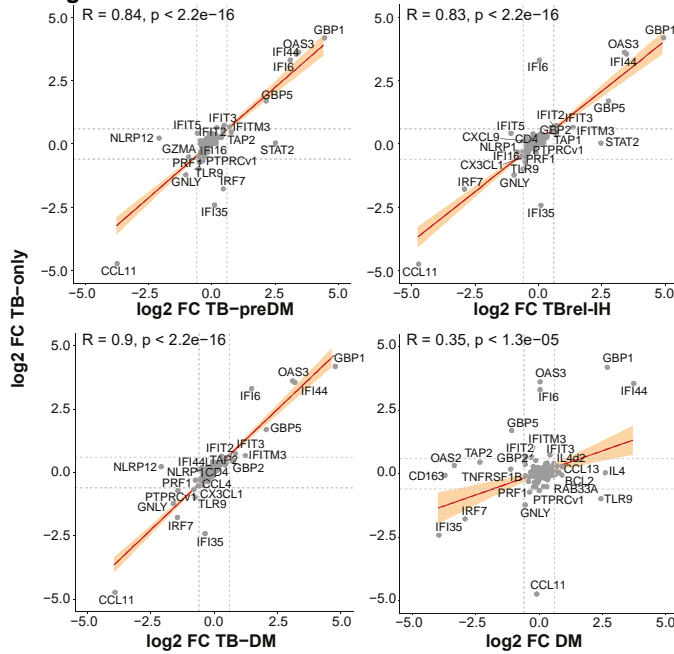


**FIGURE 4 | Modular activity of the most significant modules in TB-DM relative to TB-only in A) South Africa and B) Indonesia.** Modular analysis was performed between TB-DM and TB-only patients and the most statistically significant were chosen ( $p$ -value  $< 0.05$ ). Modular activity calculated by summing the expression of genes within a module and dividing by the number of genes within that module.

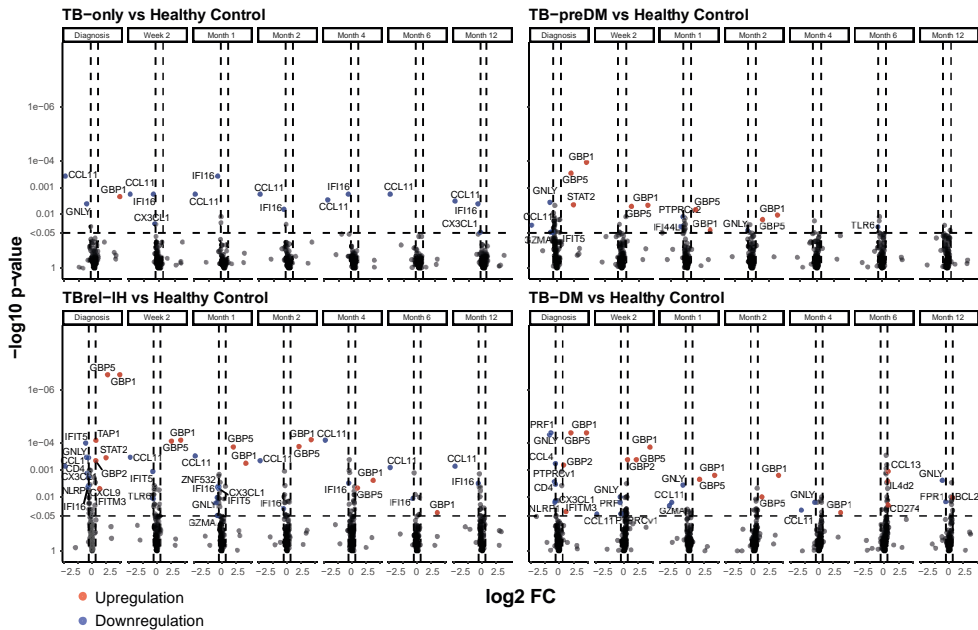
through treatment, identified by Linear Mixed Models, showed some consistency across TB groups, with the South African cohort exhibiting downregulation of *GBP5*, *GBP1*, and *IFITM3* (Figure 6B) and the Indonesian cohort showing downregulation of *GBP5* and *IFITM3* and upregulation of *GNLY* (Supplementary Figure 6B) from diagnosis to 6 months. Importantly, the number of upregulated DEGs in response to TB treatment increased with glycaemic index in both cohorts (South Africa: TB-only:6 DEGs, TB-preDM:10 DEGs, TBrel-IH:12 DEGs, TB-DM:14 DEGs; Indonesia: TB-only: 9 DEGs, TBrel-IH:13 DEGs, TB-DM:22 DEGs). Notably, no DEGs were detected between 6 and 12 months in the South African cohort, except for *GBP5* ( $p < 1e-10$ ) in patients with TBrel-IH (Supplementary Figure S7).

Ingenuity Pathway Analysis showed the majority of treatment-response DEGs in TB-only and TB-preDM were Interferon-signaling genes (ISGs) (Figure 6C, Supplementary Figure S6C). In contrast, in TBrel-IH and TB-DM patients, although downregulation of ISGs through treatment was observed, the major change was upregulation of genes associated with adaptive immunity (T-cell subset markers, Th1-associated genes, Treg-associated genes, cytotoxicity markers). Overall, the dcRT-MLPA confirmed that although TB-associated gene profiles show similar pattern and rate of change in TB patients and TB-DM, the magnitude is different.

**A Pearson at diagnosis**



**B Volcano plots: Study groups versus healthy controls - South Africa**



**FIGURE 5 | Gene expression profiles in TBrel-IH and TB-DM are not completely normalised to healthy control profiles at the end of TB treatment. A) Scatter plots representing Pearson correlations between expression of all genes in targeted dcRT-MLPA panel in TB patients relative to healthy controls (y-axes) versus the other study groups relative to healthy controls (x-axes), plotted**

as log<sub>2</sub> FC. Red line corresponds to line of best fit and shaded bands indicate confidence intervals. Genes regulated log<sub>2</sub> FC <-0.6 or > 0.6 are annotated. B) Differential Expression Analysis was performed on GAPDH-normalized log<sub>2</sub>-transformed targeted gene expression data of the South African cohort. Volcano plots representing DEGs at diagnosis and at different timepoints post TB treatment initiation of TB patients categorized based on their diabetes/glycaemia status compared to the healthy controls. The y-axis scales of all plots are harmonized per study group. P-values, -log<sub>10</sub>-transformed for better visualization, are plotted against log<sub>2</sub> FC. Genes with p <0.05 and log<sub>2</sub> FC <-0.6 or >0.6 were labelled as DEGs.

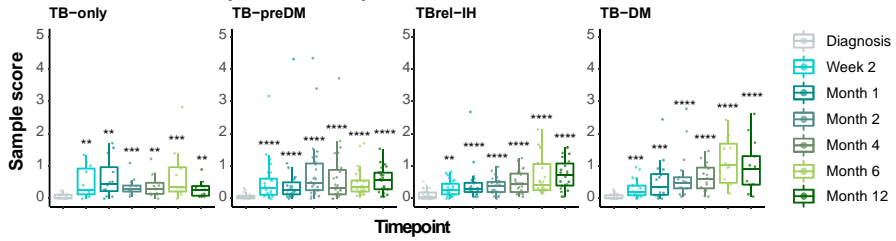
### Identification of a signature for TB treatment-response

As TB transcriptomic signatures were altered in people with DM or IH, we identified signatures with the highest classifying power to discriminate between patients at diagnosis and end of TB treatment irrespective of diabetes/glycaemia by pooling all TB patients, using logistic regression with lasso regularisation. Initially, signatures were developed in the South African and Indonesian cohorts separately (Table 2; Supplementary Table S12). The classifying capability of each signature against the training (AUC range: 0.73 – 1.0) and validation (AUC range: 0.69 – 0.92) cohorts for each clinical group was reasonably good (Figure 7A&B, Supplementary Figure 8A&B). To improve the classification performance and reduce cohort dependency, the datasets of both cohorts were pooled, and a combined two cohort 15-gene signature developed. This showed enhanced classification performance across the cohorts, with ROC analysis showing AUCs of 0.88 for TB-only, 0.96 for TBrel-IH and 0.85 for TB-DM, with excellent classification retained in individual cohorts (Figure 7C; Supplementary Figure 8C). The kinetic profiles of 6 representative genes are shown in Supplementary Figure S9.

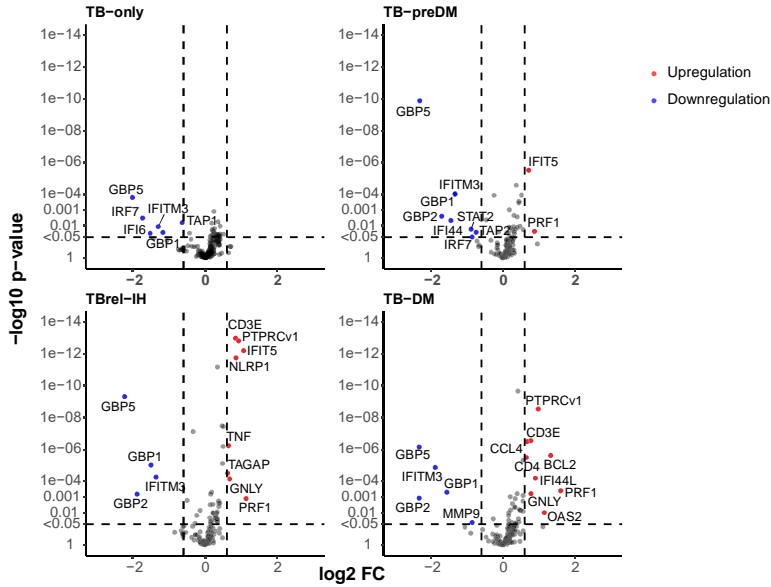
## Discussion

In this longitudinal analysis of blood transcriptomes, excessive gene expression perturbation previously described at TB diagnosis [13, 25] continued throughout six months of TB treatment in pulmonary TB patients with diabetes co-morbidity. However, qualitatively and kinetically similar changes occurred in patients with or without diabetes, suggesting prolonged TB treatment might be sufficient to restore normal transcriptomes. TB patients with either pre-diabetes or TB-related IH also exhibited greater magnitudes of gene expression perturbation throughout treatment, similar to patients with diagnosed diabetes. The overall consistency in change of gene expression through treatment, irrespective of glycaemic index, enabled derivation of accurate predictive models of TB treatment response, which could be used effectively in populations with or without diabetes.

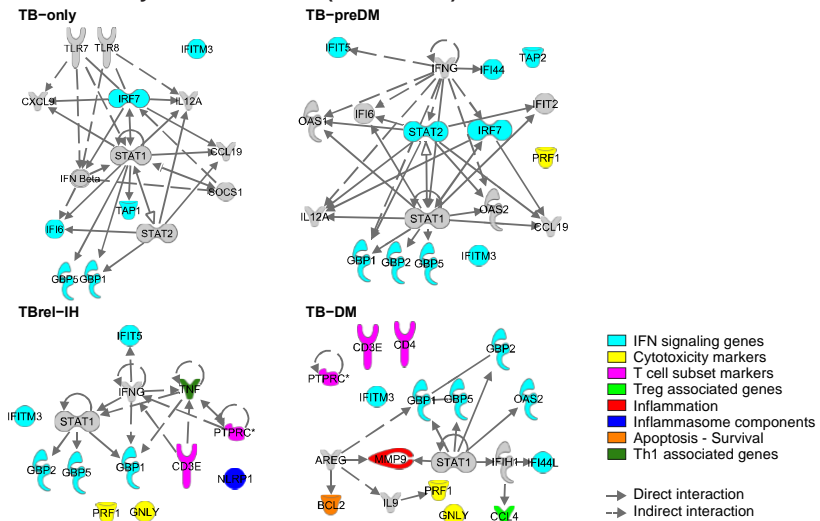
**A MDP South Africa (0-12 months)**



**B Differentially expressed genes South Africa (0-6 months)**



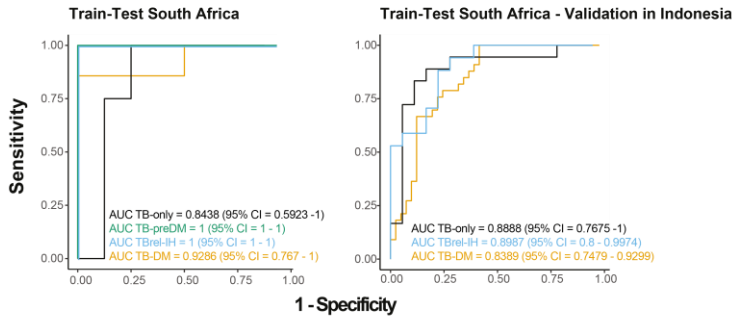
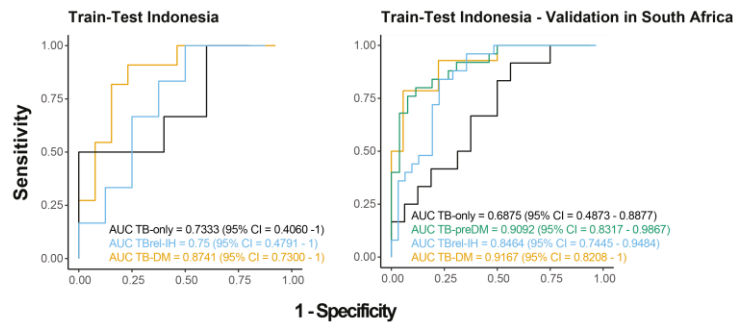
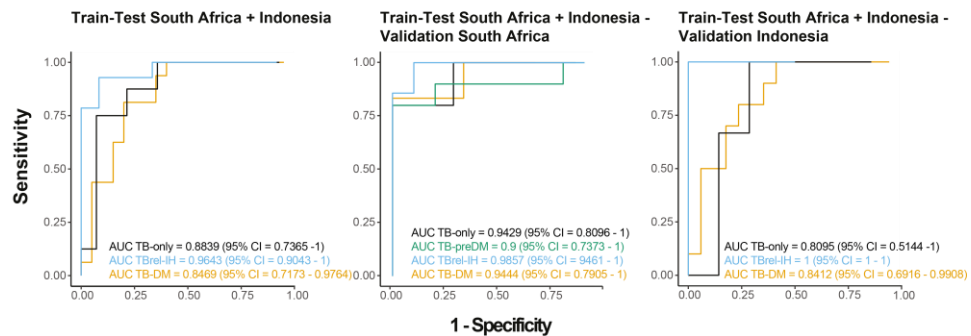
**C IPA network analysis South Africa (0-6 months)**



**FIGURE 6 | TB treatment response in TB patients is dependent on diabetes/glycaemia status.** MDP and Differential Expression Analyses were performed on GAPDH-normalized log<sub>2</sub>-transformed targeted gene expression data of the South African cohort. (A) MDP analyses of the different study groups showing the impact of TB treatment on the overall gene perturbation over time. Samples of patients at diagnosis were used as baseline controls. (B) Volcano plots representing DEGs regulated during TB treatment of TB patients categorized based on their diabetes/glycaemia status. The y-axis scales of all plots are harmonized per study group. P-values, -log<sub>10</sub>-transformed for better visualization, are plotted against log<sub>2</sub> FC. Genes with  $p < 0.05$  and log<sub>2</sub> FC  $< -0.6$  or  $> 0.6$  were labelled as DEGs. (C) IPA interactive network analyses of DEGs regulated during TB treatment. The various shapes of the nodes represent the functional classes of the gene products. Gene modules are indicated by distinctive colours.

**TABLE 2 | Gene expression signature predicting month 6 versus diagnosis, obtained by pooling the study groups and cohorts (South Africa + Indonesia).**

| Gene      | Module                                | Coefficient |
|-----------|---------------------------------------|-------------|
| Intercept |                                       | -14.99967   |
| BLR1      | G protein-coupled receptors           | 0.07841     |
| CCL13     | Chemokines                            | 0.40415     |
| CCL4      | Treg associated genes                 | 0.66015     |
| CD19      | Immune cell subset markers - B cells  | 0.2371      |
| CD3E      | T cell subset markers                 | 0.35071     |
| CD4       | T cell subset markers                 | 0.00263     |
| FCGR1A    | IFN signaling genes                   | -0.48224    |
| FPR1      | Myeloid-associated genes              | -0.1725     |
| GBP5      | IFN signaling genes                   | -0.16957    |
| IFIT5     | IFN signaling genes                   | 0.0172      |
| NLRP1     | Inflammasome components               | 0.3498      |
| PTPRCv1   | T cell subset markers                 | 0.6625      |
| TAP1      | IFN signaling genes                   | -0.51107    |
| TNF       | Th1 associated genes                  | 0.06539     |
| ZNF532    | Transcriptional regulators/activators | 0.06389     |

**A South Africa****B Indonesia****C South Africa and Indonesia**

**FIGURE 7 | Identification of common host biomarker signatures associated with TB treatment response irrespective of population heterogeneity and diabetes/glycaemia severity.** South African, Indonesian, or pooled cohort transcriptomic datasets of TB patients independent of their diabetes/glycaemia status were used to train the models. Receiver Operating Characteristic (ROC) curves (Sensitivity plotted against 1-Specificity) and Area Under the Curve (AUC) with 95% Confidence Intervals (CI) show the classifying performance of the trained models. (A) The model trained on 70% of the South African dataset was tested in the remaining 30% of the South African dataset split into the different TB study groups (left panel) and validated using the complete dataset of the Indonesian cohort split into the different TB study groups (right panel). (B) The model trained on 70% of the Indonesian dataset was tested in the remaining 30% of the Indonesian dataset split into the different TB study groups (left panel) and validated using the complete dataset of the South

African cohort split into the different TB study groups (right panel). (C) The model trained on 70% of the pooled (South African and Indonesian) dataset was tested in the remaining 30% of the pooled dataset split into the different TB study groups that both cohorts have in common (left panel) and validated using the complete dataset of the South African cohort split into the different TB study groups (middle panel) or the complete dataset of the Indonesian cohort split into the different TB study groups (right panel).

Diabetes has a negative effect on TB treatment outcomes [5, 6], for unclear reasons. One explanation could be a qualitatively different immune response in diabetes, leaving people persistently susceptible to bacterial replication and disease reactivation. An alternative explanation is that excessive inflammation and immune activation at diagnosis in TB-DM means patients require longer or, more likely, different treatment to reach the same endpoint as people with uncomplicated TB, so that they are not left susceptible to TB recurrence. Our data support the latter model, as all gene clusters differentially expressed between clinical groups exhibited similar changes, but of different magnitude. Bronchial spread often persists beyond treatment initiation, with new or expanding cavities appearing on PET-CT scans 4 weeks into treatment in one-fifth of pulmonary TB patients [32]. Plausibly, increased ongoing bacterial spread in patients with diabetes comorbidity causes persistent pro-inflammatory responses: the peripheral transcriptome correlates with lung inflammatory activity in TB patients [33]. Restoration of normal transcriptomes, and presumably improved lung resolution, could potentially also be achieved by coadministration of host-directed therapy alongside standard treatment. Therapy which dampens pro-inflammatory responses, such as corticosteroids or matrix metalloproteinase inhibitors [34], would have added benefit by reducing lung damage, which often persists after microbiological cure [35]. Anti-hyperglycaemic therapy, such as metformin, leads to more balanced, less inflammatory responses to *M. tuberculosis* [36], and has been suggested as adjunctive therapy for TB, particularly in patients with diabetes [37]. Our transcriptomic data suggest that patients with either pre-diabetes or TB-related IH would also benefit from prolonged or adjunctive host-directed therapy, in alignment with observed worse TB treatment outcomes in people with transient hyperglycaemia [38].

The ability to monitor TB treatment and predict outcome would be beneficial for clinical management. We show that transcriptomic models can be derived from host blood which reflect TB treatment-response irrespective of glycaemia. The best models include genes involved in interferon signalling, known to be suppressed at TB diagnosis in TB-DM patients [25], which we found were enhanced mid-way through treatment but did eventually resolve by 6 months. People protected against TB development display balanced prostaglandin 2 and lipoxin expression in lungs, preventing TB disease

progression following infection [39]. Drugs which target 5-lipoxygenase restrict lung pathology and reduce bacterial replication in murine models, by lowering the type 1 interferon response; the increases through treatment in the TB-DM cohort may relate to sustained infection and accompanying inflammation. In TB-DM patients the inflammation-related genes resolved more linearly through TB treatment, but remained elevated to the end of TB treatment, persisting until 12 months post-diagnosis in the South African cohort. In future studies, it would be important to test whether prolonged treatment with standard therapy impacts blood transcriptomes beyond the 6 month time point. Increased doses of anti-TB drugs might also lead to better treatment outcomes in TB-DM. In a companion paper (van Doorn *et al*, [40] unpublished), transcriptomic signatures indicative of treatment outcome have been derived that can be used in patients with either DM or IH. Together, these papers show that signatures related to poor TB outcome are distinct from the excessive and prolonged inflammation observed in TB-DM.

These findings further illustrate how comorbidity with diabetes affects the host response to *M.tuberculosis* infection, and how a better understanding of these interactions could be exploited to reduce poor TB treatment outcomes associated with TB and diabetes comorbidity.

## Funding

The research leading to these results, as part of the TANDEM Consortium, has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no. 305279.

## Conflicts of Interest

GW has patents about methods of tuberculosis diagnosis and tuberculosis biomarkers which are unrelated to the current study. No other authors have any declared conflicts of interest.

## Acknowledgments

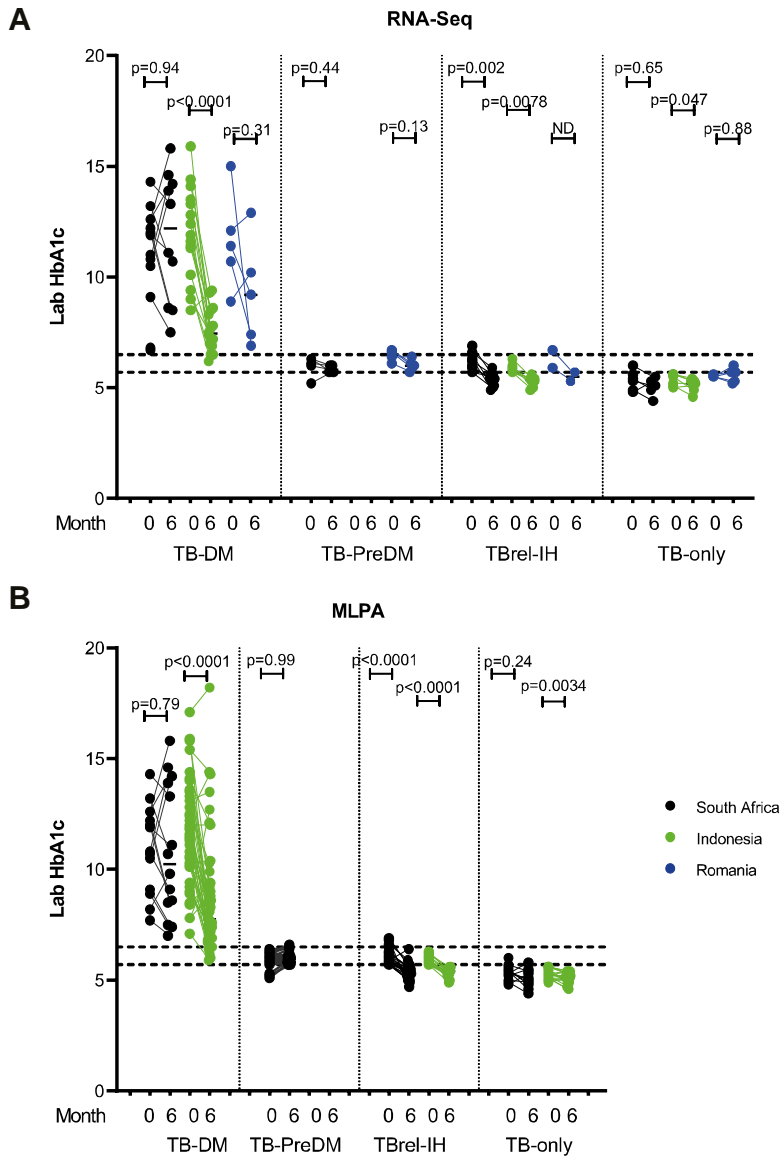
We are grateful to all the study participants for blood and data donations. We acknowledge Bahram Sanjabi, Desiree Brandenburg-Weening, and Pieter van der Vlies for assistance with the RNA-Seq, Evelien Temminck for providing technical assistance with dcRT-MLPA experiments, and Erni Durdevic for providing statistical and machine learning advice.

## References

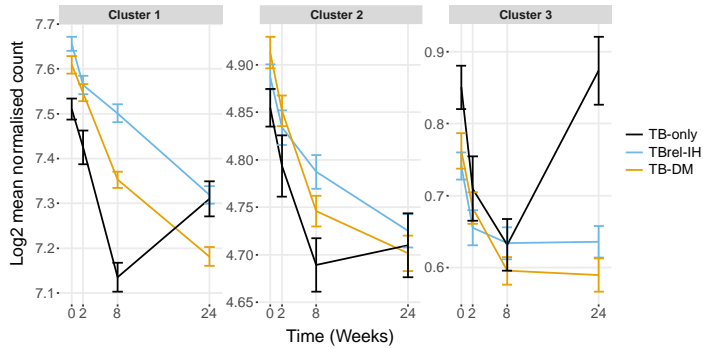
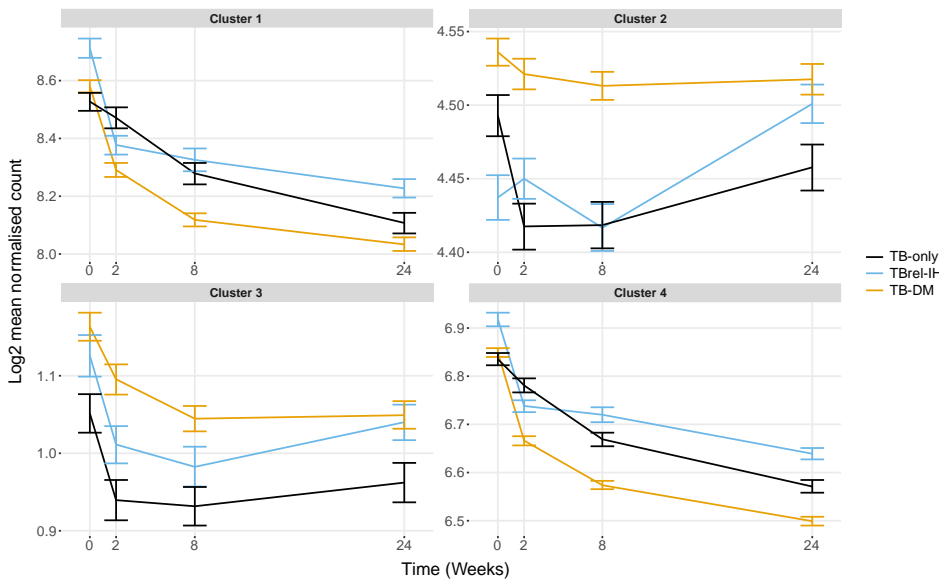
1. Koesoemadinata RC, McAllister SM, Soetedjo NNM, et al. Latent TB infection and pulmonary TB disease among patients with diabetes mellitus in Bandung, Indonesia. *Trans R Soc Trop Med Hyg* 2017; 111(2): 81-9.
2. Al-Rifai RH, Pearson F, Critchley JA, Abu-Raddad LJ. Association between diabetes mellitus and active tuberculosis: A systematic review and meta-analysis. *PLoS One* 2017; 12(11):e0187967.
3. Jeon CY, Murray MB. Diabetes mellitus increases the risk of active tuberculosis: a systematic review of 13 observational studies. *PLoS Med* 2008; 5(7): e152.
4. Noubiap JJ, Nansseu JR, Nyaga UF, et al. Global prevalence of diabetes in active tuberculosis: a systematic review and meta-analysis of data from 2.3 million patients with tuberculosis. *Lancet Glob Health* 2019; 7(4): e448-e60.
5. Baker MA, Harries AD, Jeon CY, et al. The impact of diabetes on tuberculosis treatment outcomes: a systematic review. *BMC Med* 2011; 9: 81.
6. Huangfu P, Ugarte-Gil C, Golub J, Pearson F, Critchley J. The effects of diabetes on tuberculosis treatment outcomes: an updated systematic review and meta-analysis. *Int J Tuberc Lung Dis* 2019; 23(7): 783-96.
7. International Diabetes Federation. *IDF Diabetes Atlas Ninth Edition*, 2019.
8. American Diabetes Association. Diagnosis and classification of diabetes mellitus. *Diabetes Care* 2010; 33 Suppl 1: S62-9.
9. Dungan KM, Braithwaite SS, Preiser JC. Stress hyperglycaemia. *Lancet* 2009; 373(9677):1798-807.
10. Boillat-Blanco N, Ramaiya KL, Mganga M, et al. Transient Hyperglycemia in Patients With Tuberculosis in Tanzania: Implications for Diabetes Screening Algorithms. *J Infect Dis* 2016;213(7): 1163-72.
11. Yoon YS, Jung JW, Jeon EJ, et al. The effect of diabetes control status on treatment response in pulmonary tuberculosis: a prospective study. *Thorax* 2017; 72(3): 263-70.
12. Ronacher K, Joosten SA, van Crevel R, Dockrell HM, Walzl G, Ottenhoff TH. Acquired immunodeficiencies and tuberculosis: focus on HIV/AIDS and diabetes mellitus. *Immunol Rev* 2015; 264(1): 121-37.
13. Prada-Medina CA, Fukutani KF, Pavan Kumar N, et al. Systems Immunology of Diabetes-Tuberculosis Comorbidity Reveals Signatures of Disease Complications. *Sci Rep* 2017; 7(1):1999.
14. Ronacher K, Chegou NN, Kleynhans L, et al. Distinct serum biosignatures are associated with different tuberculosis treatment outcomes. *Tuberculosis (Edinb)* 2019; 118: 101859.
15. Kumar NP, Banurekha VV, Nair D, et al. Coincident pre-diabetes is associated with dysregulated cytokine responses in pulmonary tuberculosis. *PLoS One* 2014; 9(11): e112108.
16. Kumar NP, Fukutani KF, Shruthi BS, et al. Persistent inflammation during anti-tuberculosis treatment with diabetes comorbidity. *Elife* 2019; 8.
17. Kumar NP, Moideen K, Sivakumar S, et al. Modulation of dendritic cell and monocyte subsets in tuberculosis-diabetes co-morbidity upon standard tuberculosis treatment. *Tuberculosis (Edinb)* 2016; 101: 191-200.
18. Kumar NP, Moideen K, Viswanathan V, Kornfeld H, Babu S. Effect of standard tuberculosis treatment on naive, memory and regulatory T-cell homeostasis in tuberculosis-diabetes comorbidity. *Immunology* 2016; 149(1): 87-97.
19. Berry MP, Graham CM, McNab FW, et al. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature* 2010; 466(7309): 973-7.
20. Cliff JM, Lee JS, Constantinou N, et al. Distinct phases of blood gene expression pattern through tuberculosis treatment reflect modulation of the humoral immune response. *J Infect Dis* 2013; 207(1): 18-29.
21. Kaforou M, Wright VJ, Oni T, et al. Detection of tuberculosis in HIV-infected and -uninfected

- African adults using whole blood RNA expression signatures: a case-control study. *PLoS Med* 2013; 10(10): e1001538.
22. Ottenhoff TH, Dass RH, Yang N, et al. Genome-wide expression profiling identifies type 1 interferon response pathways in active tuberculosis. *PLoS One* 2012; 7(9): e45839.
  23. Bloom CI, Graham CM, Berry MP, et al. Detectable changes in the blood transcriptome are present after two weeks of antituberculosis therapy. *PLoS One* 2012; 7(10): e46191.
  24. Thompson EG, Du Y, Malherbe ST, et al. Host blood RNA signatures predict the outcome of tuberculosis treatment. *Tuberculosis* 2017; 107: 48-58.
  25. Eckold C, Kumar V, Weiner J, et al. Impact of Intermediate Hyperglycemia and Diabetes on Immune Dysfunction in Tuberculosis. *Clin Infect Dis* 2021; 72(1): 69-78.
  26. Ugarte-Gil C, Alisjahbana B, Ronacher K, et al. Diabetes Mellitus Among Pulmonary Tuberculosis Patients From 4 Tuberculosis-endemic Countries: The TANDEM Study. *Clin Infect Dis* 2020; 70(5): 780-8.
  27. Ruslami R, Koesoemadinata RC, Soetedjo NNM, et al. The effect of a structured clinical algorithm on glycemic control in patients with combined tuberculosis and diabetes in Indonesia: A randomized trial. *Diabetes research and clinical practice* 2021; 173: 108701.
  28. Grint D, Alisjahbana B, Ugarte-Gil C, et al. Accuracy of diabetes screening methods used for people with tuberculosis, Indonesia, Peru, Romania, South Africa. *Bull World Health Organ* 2018; 96(11): 738-49.
  29. Nueda MJ, Tarazona S, Conesa A. Next maSigPro: updating maSigPro Bioconductor package for RNA-seq time series. *Bioinformatics* 2014; 30(18): 2598-602.
  30. Joosten SA, Goeman JJ, Sutherland JS, et al. Identification of biomarkers for tuberculosis disease using a novel dual-color RT-MLPA assay. *Genes Immun* 2012; 13(1): 71-82.
  31. Dennis G, Jr., Sherman BT, Hosack DA, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* 2003; 4(5): P3.
  32. Chen RY, Yu X, Smith B, et al. Radiological and functional evidence of the bronchial spread of tuberculosis: an observational analysis. *The Lancet Microbe* 2021; 2(10): e518-e26.
  33. Odia T, Malherbe ST, Meier S, et al. The Peripheral Blood Transcriptome Is Correlated With PET Measures of Lung Inflammation During Successful Tuberculosis Treatment. *Frontiers in immunology* 2020; 11: 596173.
  34. Young C, Walzl G, Du Plessis N. Therapeutic host-directed strategies to improve outcome in tuberculosis. *Mucosal Immunology* 2020; 13(2): 190-204.
  35. Krug S, Parveen S, Bishai WR. Host-Directed Therapies: Modulating Inflammation to Treat Tuberculosis. *Frontiers in immunology* 2021; 12: 660916.
  36. Lachmandas E, Eckold C, Bohme J, et al. Metformin Alters Human Host Responses to *Mycobacterium tuberculosis* in Healthy Subjects. *J Infect Dis* 2019; 220(1): 139-50.
  37. Singhal A, Jie L, Kumar P, et al. Metformin as adjunct antituberculosis therapy. *Science translational medicine* 2014; 6(263): 263ra159.
  38. Liu Q, You N, Pan H, et al. Glycemic Trajectories and Treatment Outcomes of Patients with Newly Diagnosed Tuberculosis: A Prospective Study in Eastern China. *Am J Respir Crit Care Med* 2021; 204(3): 347-56.
  39. Mayer-Barber KD, Andrade BB, Oland SD, et al. Host-directed therapy of tuberculosis based on interleukin-1 and type I interferon crosstalk. *Nature* 2014; 511(7507): 99-103.
  40. van Doorn CLR, Eckold C, Ronacher K, et al. Transcriptional profiles predict treatment outcome in patients with tuberculosis and diabetes at diagnosis and at two weeks after initiation of anti-tuberculosis treatment. *MedRxiv* 2022; (MEDRXIV/2022/269796).

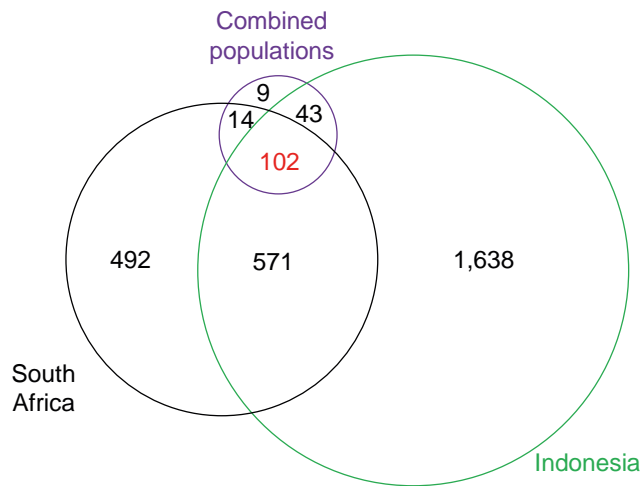
## Supplementary Figures



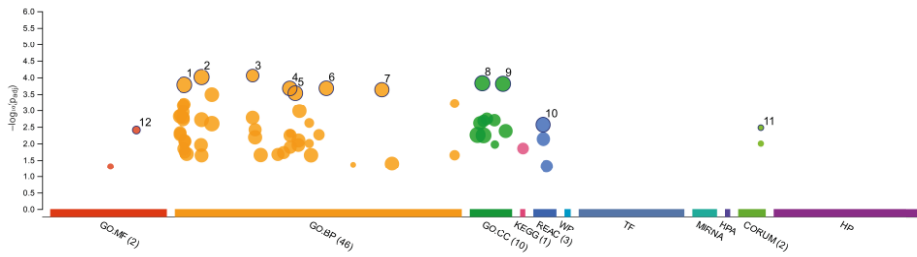
**SUPPLEMENTARY FIGURE S1 | Changes in HbA1c in pulmonary TB patients through TB drug treatment in people with concomitant diabetes, pre-diabetes or TB-related intermediate hyperglycaemia.** Laboratory HbA1c recorded at month 0, prior to initiation of treatment, and at the end of treatment at month 6 in patients included in A) RNA-Seq analysis and B) IMLPA analysis, for cohorts from South Africa (Black), Indonesia (Green) and Romania (Blue). The dotted lines show the threshold for diagnosis of diabetes (6.5%) and pre-diabetes (5.7%) [8]. Timepoints in each cohort were compared by Wilcoxon signed rank test.

**A South Africa****B Indonesia**

**SUPPLEMENTARY FIGURE S2 | MaSigPro analysis of change in gene expression through TB treatment in blood samples from patients in A) South Africa and B) Indonesia.** MaSigPro identified genes that behaved similarly through time and between patient groups using hierarchical clustering. Results are shown for log-transformed normalised transcript counts for the TB-only, TB-DM and the combined TB-PreDM and TBrel-IH groups. Bars show mean  $\pm$  1 SEM. Data were filtered to remove lowly abundant transcripts prior to analysis. NB: There were insufficient study participants to conduct an analysis separately for the Romanian population.



**SUPPLEMENTARY FIGURE S3 | Comparison of results from MaSigPro analyses for individual and combined populations.** Genes which were identified as behaving differently between TB-DM, TBrel-IH/TB-PreDM or TB-only in South Africa (Black), Indonesia (Green) or combined populations from Romania, South Africa and Indonesia (Purple) in MaSigPro analyses (Figures 2, S2; Supplementary Tables S4, S5, S6) were compared. The Venn Diagram shows the number of identified genes which overlap between analyses

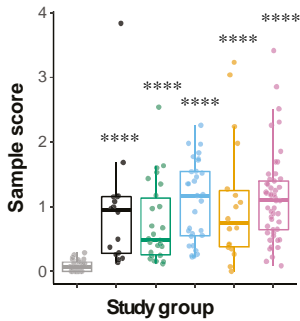
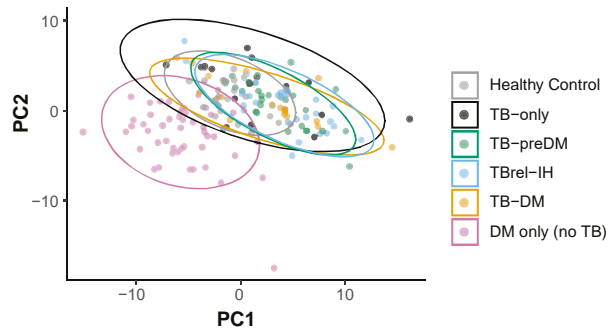


| ID | Source | Term ...         | Term Name   | Padj (query_1)           |
|----|--------|------------------|---|--------------------------|
| 1  | GO:BP  | GO:0002376       | immune system process                                   | 1.664 × 10 <sup>-4</sup> |
| 2  | GO:BP  | GO:0006955       | immune response   | 9.729 × 10 <sup>-5</sup> |
| 3  | GO:BP  | GO:0031349       | positive regulation of defense response                 | 8.788 × 10 <sup>-5</sup> |
| 4  | GO:BP  | GO:0043207       | response to external biotic stimulus                    | 2.126 × 10 <sup>-4</sup> |
| 5  | GO:BP  | GO:0044419       | biological process involved in interspecies interact... | 2.954 × 10 <sup>-4</sup> |
| 6  | GO:BP  | GO:0051707       | response to other organism                              | 2.101 × 10 <sup>-4</sup> |
| 7  | GO:BP  | GO:0098542       | defense response to other organism                      | 2.320 × 10 <sup>-4</sup> |
| 8  | GO:CC  | GO:0031410       | cytoplasmic vesicle                                     | 1.489 × 10 <sup>-4</sup> |
| 9  | GO:CC  | GO:0097708       | intracellular vesicle                                   | 1.548 × 10 <sup>-4</sup> |
| 10 | REAC   | REAC:R-HSA-16... | Immune System   | 2.721 × 10 <sup>-3</sup> |
| 11 | CORUM  | CORUM:6826       | Calprotectin heterotetramer                             | 3.346 × 10 <sup>-3</sup> |
| 12 | GO:MF  | GO:0050786       | RAGE receptor binding                                   | 3.919 × 10 <sup>-3</sup> |

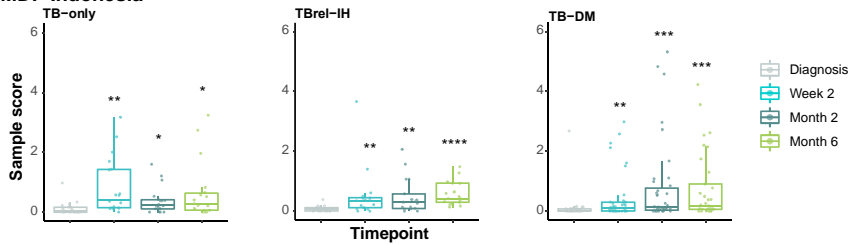
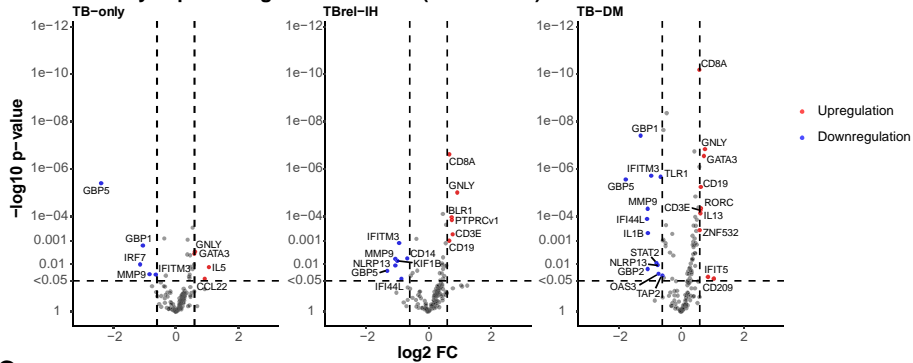
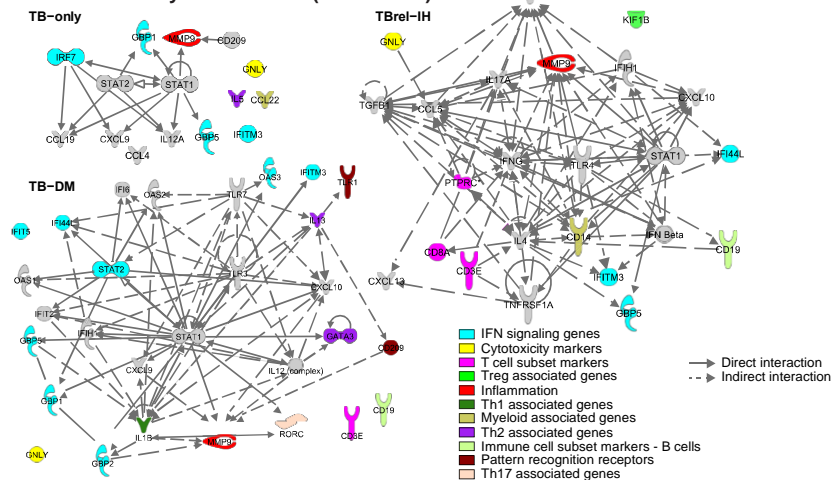
version e104\_eg01\_p10\_3922dba  
 date 13/12/2021, 13:09:13  
 organism hsapiens

g:Profiler

**SUPPLEMENTARY FIGURE S4 | G:profiler analysis of 102 gene MSP core gene list, identifying top Gene Ontology (GO) molecular function (MF), biological process (BP) and cellular component (CC) as well as biological pathways in KEGG and Reactome databases.**

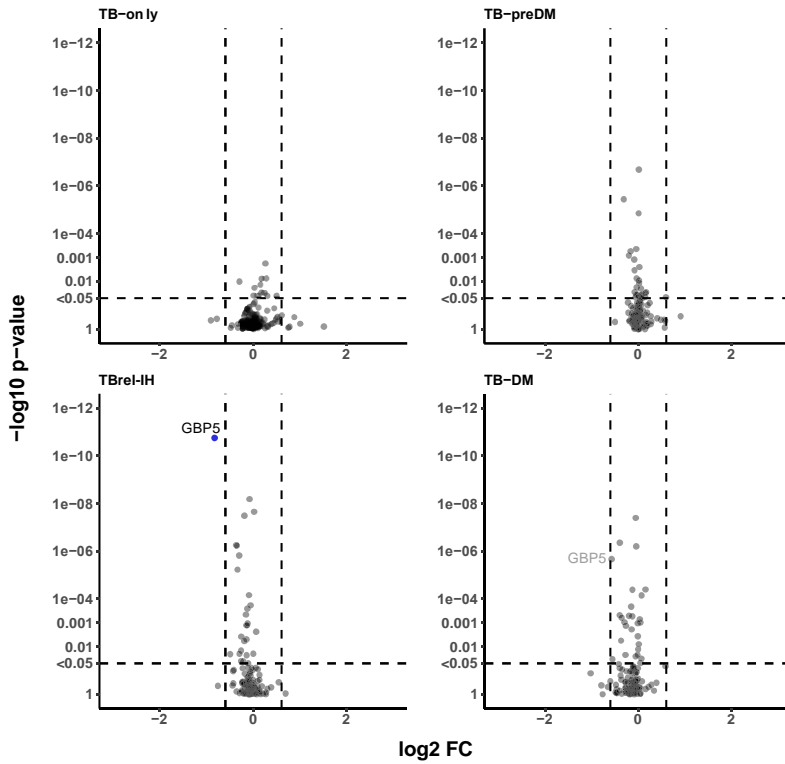
**A MDP at diagnosis****B PLS-DA at diagnosis**

**SUPPLEMENTARY FIGURE S5 | Gene expression profiles in TB patients is dominant over the gene perturbation caused by DM.** Molecular Degree of Perturbation (MDP) and Partial Least Squares – Discriminant Analysis (PLS-DA) were performed on GAPDH-normalized log<sub>2</sub>-transformed targeted gene expression data of the South African cohort. (A) MDP analysis of the different study groups showing the impact of TB and DM on the overall gene perturbation at diagnosis. Samples of healthy individuals were used as controls. (B) PLS-DA analysis of the different study groups at diagnosis. Samples of healthy individuals were used as controls.

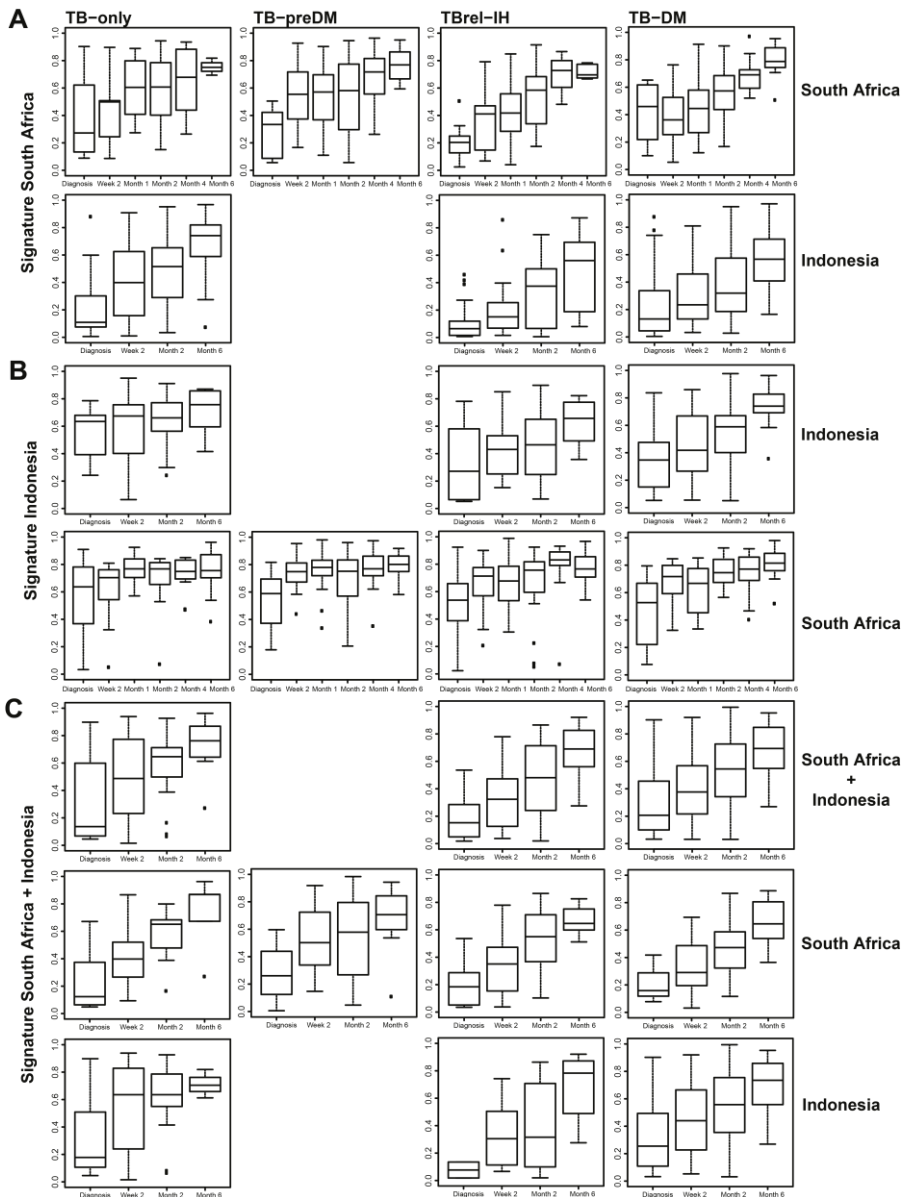
**A MDP Indonesia****B Differentially expressed genes Indonesia (0-6 months)****C IPA network analysis Indonesia (0-6 months)**

**SUPPLEMENTARY FIGURE S6 | TB treatment response in TB patients is dependent on diabetes/glycaemia status in the Indonesian cohort.** MDP and Differential Expression Analyses were performed on GAPDH-normalized  $\log_2$ -transformed targeted gene expression data of the Indonesian cohort. (A) MDP analyses of the different study groups showing the impact of TB treatment on the overall gene perturbation over time. Samples of patients at diagnosis were used as baseline controls. (B) Volcano plots representing DEGs regulated during treatment of TB patients categorized based on their diabetes/glycaemia status. The y-axis scales of all plots are harmonized per study group.  $-\log_{10}$ -transformed p-values are plotted against  $\log_2 FC$ . Genes with  $p < 0.05$  and  $\log_2 FC < -0.6$  or  $> 0.6$  were labelled as DEGs. (C) IPA interactive network analyses of DEGs regulated during treatment. The various shapes of the nodes represent the functional classes of the gene products. Gene modules are indicated by distinctive colors.

## Differentially expressed genes South Africa after anti-TB treatment (6-12 months)

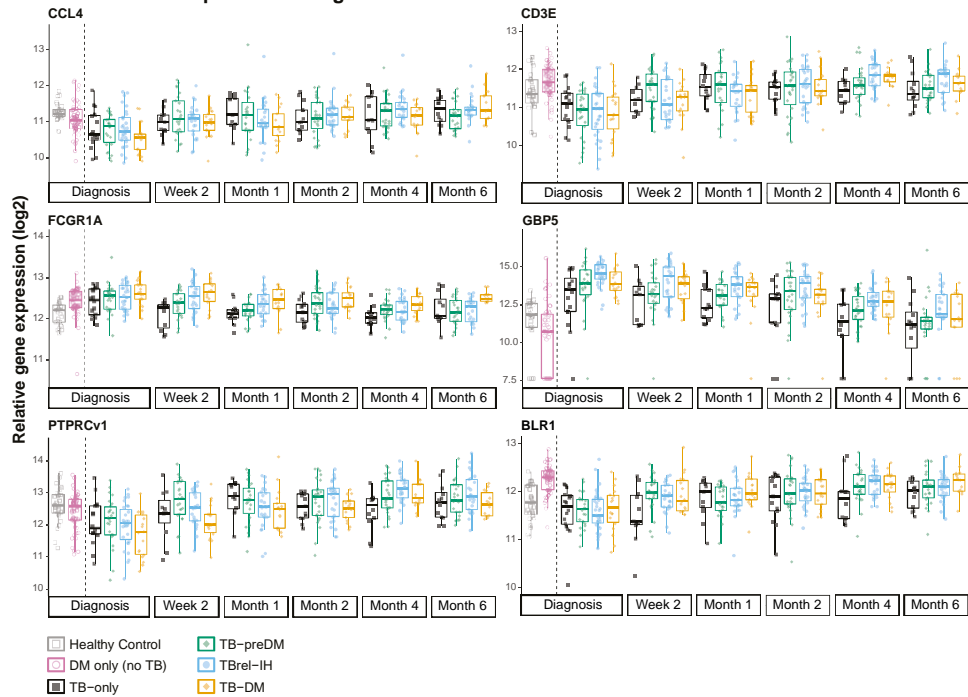


**SUPPLEMENTARY FIGURE S7 | Transcriptional response after anti-TB treatment in the South African cohort.** Volcano plots representing DEGs regulated after TB treatment (6-12 months) of TB patients categorized based on their diabetes/glycaemia status. The y-axis scales of all plots are harmonized per study group. P-values,  $-\log_{10}$ -transformed for better visualization, are plotted against  $\log_2 \text{FC}$ . Genes with  $p < 0.05$  and  $\log_2 \text{FC} < -0.6$  or  $> 0.6$  were labelled as DEGs.

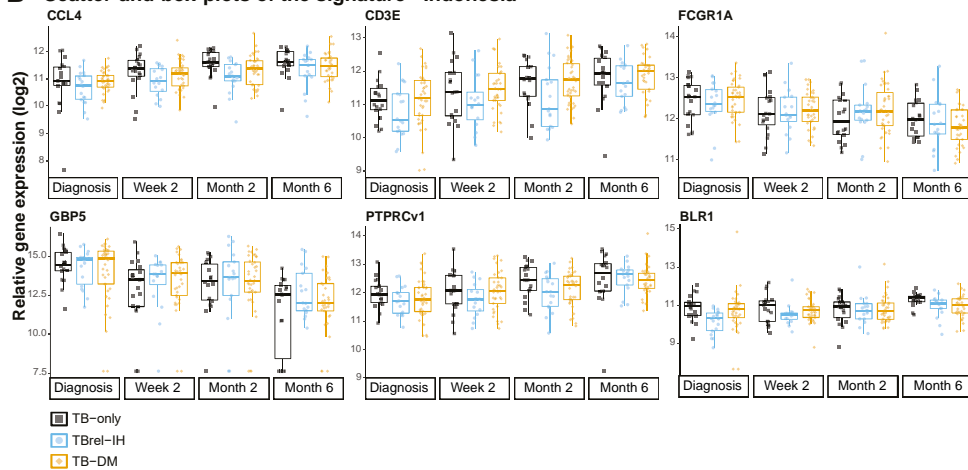


**SUPPLEMENTARY FIGURE S8 | Classifying capability of identified TB treatment response biomarker signatures.** Predicted probability plots showing the accuracy of TB treatment response biomarker signatures identified in (A) South Africa or (B) Indonesia across timepoints in box-and-whiskers plots (5-95 percentiles) either in the cohort in which the Train Test Split (TTS) was performed or in the validation cohort, split into TB study groups based on their diabetes/glycaemia status. (C) Predicted probability plots showing the accuracy of the identified pooled (South Africa and Indonesia) biomarker signature across timepoints in box-and-whiskers plots (5-95 percentiles) either in the pooled cohort in which Train-Test Split (TTS) was performed (left panel) or the single validation cohorts (right panels), split into TB study groups based on their diabetes/glycaemia status.

**A Scatter-and-box-plots of the signature - South Africa**



**B Scatter-and-box-plots of the signature - Indonesia**



**SUPPLEMENTARY FIGURE S9 | The kinetic profiles of 6 representative genes of the 15-gene TB treatment response signature of the pooled cohort dataset.** Expression kinetics of single genes in the (A) South African cohort or (B) Indonesian cohort at diagnosis and during TB treatment. Box plots depict log<sub>2</sub>-transformed median gene expression values and the inter quartile range (IQR), while the whiskers represent the data within the Q<sub>1</sub>-1.5xIQR and Q<sub>3</sub>+1.5xIQR interval. Outliers are reported as symbols.

## Supplementary Tables

**SUPPLEMENTARY TABLE S1 | List of target genes for dcfRT-MLPA.**

| Module                               | Gene     |
|--------------------------------------|----------|
| Immune cell subset markers - B cells | CD19     |
|                                      | NCAM1    |
| NK cells                             |          |
|                                      |          |
| T cell subset markers                | CD3E     |
|                                      | CD4      |
|                                      | CD8A     |
|                                      | CCR7     |
|                                      | PTPRCv1  |
|                                      | PTPRCv2  |
|                                      | AIRE     |
|                                      | IL7R     |
| Th1 associated / IFN signaling genes | CXCL10   |
| Th1 associated genes                 | IFNG     |
|                                      | IL1B     |
|                                      | IL2      |
|                                      | IL15     |
|                                      | TBX21    |
|                                      | TNF      |
| Th2 associated genes                 | GATA3    |
|                                      | IL4      |
|                                      | IL452    |
|                                      | IL5      |
|                                      | IL6      |
|                                      | IL10     |
|                                      | IL13     |
| Th9 associated genes                 | IL9      |
| Th17 associated genes                | IL17A    |
|                                      | RORC     |
|                                      | IL22RA1  |
| Treg associated genes                | CCL4     |
|                                      | CTLA4    |
|                                      | FOXP3    |
|                                      | IL2RA    |
|                                      | LAG3     |
|                                      | TGFB1    |
|                                      | TNFRSF18 |
| Cytotoxicity markers                 | GNLY     |
|                                      | GZMA     |
|                                      | GZMB     |
|                                      | PRF1     |
| Apoptosis / Survival                 | CASP8    |
|                                      | BCL2     |
|                                      | FASLG    |
|                                      | FLCN1    |
|                                      | TNFRSF1A |
|                                      | TNFRSF1B |
| Myeloid-associated genes             | CD14     |
|                                      | CD163    |

|                               |        |
|-------------------------------|--------|
|                               | CCL2   |
|                               | CCL3   |
|                               | CCL5   |
|                               | CCL22  |
|                               | CXCL13 |
|                               | IL12A  |
|                               | IL12B  |
|                               | IL23A  |
|                               | FPR1   |
| Chemokines                    | CCL11  |
|                               | CCL13  |
|                               | CCL19  |
|                               | CXCL9  |
|                               | CX3CL1 |
| Pattern recognition receptors | CD209  |
|                               | CLEC7A |
|                               | MRC1   |
|                               | MRC2   |
|                               | NOD1   |
|                               | NOD2   |
|                               | TLR1   |
|                               | TLR2   |
|                               | TLR3   |
|                               | TLR4   |
|                               | TLR5   |
|                               | TLR6   |
|                               | TLR7   |
|                               | TLR8   |
|                               | TLR9   |
|                               | TLR10  |
| Inflammasome components       | NLR4   |
|                               | NLRP1  |
|                               | NLRP2  |
|                               | NLRP3  |
|                               | NLRP4  |
|                               | NLRP6  |
|                               | NLRP7  |
|                               | NLRP10 |
|                               | NLRP11 |
|                               | NLRP12 |
|                               | NLRP13 |
| IFN signaling genes           | CD274  |
|                               | FCGR1A |
|                               | GBP1   |
|                               | GBP2   |
|                               | GBP5   |
|                               | IFI6   |
|                               | IFI16  |
|                               | IFI35  |
|                               | IFI44  |
|                               | IFI44L |

|  |   |
|--|---|
|  | IFIH1<br>IFIT2<br>IFIT3<br>IFIT5<br>IFITM1/3<br>INDO<br>IRF7<br>OAS1<br>OAS2<br>OAS3<br>SOCS1<br>STAT1<br>STAT2<br>TAP1<br>TAP2 |
| Inflammation                                     | DSE<br>MMP9<br>SPP1<br>TIMP2<br>TNIP1   |
| Cell growth / Proliferation                      | BMP6<br>TGFB2<br>AREG<br>EGF<br>VEGF  |
| Cell activation                                  | HCK<br>LYN<br>SLAMF7  |
| Small GTPases / (Rho) GTPase activating proteins | ASAP1<br>RAB13<br>RAB24<br>RAB33A<br>TAGAP<br>TBC1D7  |
| Anti-microbial activity                          | BPI<br>LTF  |
| E3 ubiquitin protein ligases                     | NEDD4L  |
| Scavenger receptors                              | MARCO   |
| G protein-coupled receptors                      | BLR1  |
| Transcriptional regulators/activators            | CAMTA1<br>TWIST1<br>ZNF331<br>ZNF532  |
| Intracellular transport                          | SEC14L1<br>KIF1B  |
| Mitochondrial Stress / Proteasome                | HPRT  |
| Housekeeping                                     | ABR<br>B2M<br>GAPDH<br>GUSB   |

**SUPPLEMENTARY TABLE S2 | Demographics of study participants in the three locations in RNA-Seq and MLPA study arms.**

| Characteristic                       | Country    | TB-DM                  |                        | TB-PreDM             |                      | TBrel-IH             |                      | TB-only              |                      | DM-only                | HC                   |
|--------------------------------------|------------|------------------------|------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|------------------------|----------------------|
|                                      |            | RNA-Seq                | MLPA                   | RNA-Seq              | MLPA                 | RNA-Seq              | MLPA                 | RNA-Seq              | MLPA                 | MLPA                   | MLPA                 |
| Number of participants               | S Africa   | 13                     | 19                     | 5                    | 28                   | 12                   | 33                   | 6                    | 18                   | 48                     | 28                   |
|                                      | Indonesia  | 16                     | 48                     | 0                    | 0                    | 8                    | 19                   | 7                    | 22                   | -                      | -                    |
|                                      | Romania    | 5                      | 0                      | 5                    | 0                    | 2                    | 0                    | 5                    | 0                    | -                      | -                    |
|                                      | <b>All</b> | <b>34</b>              | <b>67</b>              | <b>10</b>            | <b>28</b>            | <b>22</b>            | <b>52</b>            | <b>18</b>            | <b>40</b>            | <b>48</b>              | <b>28</b>            |
| Age in years, median (range)         | S Africa   | 46 (27-57)             | 46 (27-57)             | 46 (43-56)           | 47 (24-68)           | 40 (25-55)           | 39 (22-62)           | 42 (32-56)           | 39 (19-56)           | 54 (19-69)             | 42 (30-70)           |
|                                      | Indonesia  | 50 (33-62)             | 50 (25-73)             | -                    | -                    | 44 (37-62)           | 43 (37-64)           | 48 (28-61)           | 45 (28-64)           | -                      | -                    |
|                                      | Romania    | 46 (22-60)             | -                      | 54 (44-75)           | -                    | 29 (22-36)           | -                    | 43 (30-64)           | -                    | -                      | -                    |
|                                      | <b>All</b> | <b>48 (22-62)</b>      | <b>50 (25-73)</b>      | <b>54 (43-75)</b>    | <b>47 (24-68)</b>    | <b>40 (22-62)</b>    | <b>40 (22-64)</b>    | <b>45 (28-64)</b>    | <b>43 (19-64)</b>    | <b>54 (19-69)</b>      | <b>42 (30-70)</b>    |
| Sex, % male (No. male/female)        | S Africa   | 38 (5/8)               | 47 (9/10)              | 80 (4/1)             | 79 (22/6)            | 33 (4/8)             | 58 (19/14)           | 17 (1/5)             | 50 (9/9)             | 48 (23/25)             | 50 (14/14)           |
|                                      | Indonesia  | 50 (8/8)               | 54 (26/22)             | -                    | -                    | 50 (4/4)             | 58 (11/8)            | 71 (5/2)             | 59 (13/9)            | -                      | -                    |
|                                      | Romania    | 80 (4/1)               | -                      | 80 (4/1)             | -                    | 100 (2/0)            | -                    | 60 (3/2)             | -                    | -                      | -                    |
|                                      | <b>All</b> | <b>50 (17/17)</b>      | <b>52 (35/32)</b>      | <b>80 (8/2)</b>      | <b>79 (22/6)</b>     | <b>45 (10/12)</b>    | <b>58 (30/22)</b>    | <b>50 (9/9)</b>      | <b>55 (22/18)</b>    | <b>48 (23/25)</b>      | <b>50 (14/14)</b>    |
| HbA1c TB diagnosis, % median (range) | S Africa   | 11.9 (6.7-14.3)        | 10.7 (6.7-14.3)        | 6.1 (5.2-6.3)        | 6.0 (5.1-6.4)        | 6.0 (5.7-6.9)        | 5.9 (5.7-6.9)        | 5.4 (4.8-6.0)        | 5.4 (4.8-6.0)        | 10.5 (4.7-14.9)        | 5.3 (4.8-6.5)        |
|                                      | Indonesia  | 11.8 (8.5-15.9)        | 11.5 (7.1-17.1)        | -                    | -                    | 5.8 (5.7-6.3)        | 5.9 (5.7-6.3)        | 5.5 (5.0-5.6)        | 5.3 (4.9-5.6)        | -                      | -                    |
|                                      | Romania    | 11.4 (8.9-15.0)        | -                      | 6.6 (6.1-6.7)        | -                    | 6.3 (5.9-6.7)        | -                    | 5.5 (5.5-5.6)        | -                    | -                      | -                    |
|                                      | <b>All</b> | <b>11.8 (6.7-15.9)</b> | <b>11.1 (6.7-17.1)</b> | <b>6.3 (5.2-6.7)</b> | <b>6.0 (5.1-6.4)</b> | <b>5.9 (5.7-6.9)</b> | <b>5.9 (5.7-6.9)</b> | <b>5.5 (4.8-6.0)</b> | <b>5.3 (4.8-6.0)</b> | <b>10.5 (4.7-14.9)</b> | <b>5.3 (4.8-6.5)</b> |
| HbA1c at 6 months, % median (range)  | S Africa   | 12.2 (7.5-15.8)        | 10.25 (7.0-15.8)       | 5.8 (5.7-6.0)        | 6.0 (5.7-6.6)        | 5.4 (4.9-5.9)        | 5.4 (4.7-6.4)        | 5.0 (4.4-5.3)        | 5.0 (4.4-5.8)        | -                      | -                    |
|                                      | Indonesia  | 7.5 (6.2-9.4)          | 7.7 (5.9-12.3)         | -                    | -                    | 5.4 (4.9-5.6)        | 5.4 (4.9-5.6)        | 5.2 (4.6-5.4)        | 5.1 (4.6-5.4)        | -                      | -                    |
|                                      | Romania    | 9.2 (6.9-12.9)         | -                      | 6.0 (5.7-6.4)        | -                    | 5.5 (5.3-5.7)        | -                    | 5.7 (5.2-6.0)        | -                    | -                      | -                    |
|                                      | <b>All</b> | <b>8.5 (6.2-15.8)</b>  | <b>8.3 (5.9-18.2)</b>  | <b>6.0 (5.7-6.4)</b> | <b>6.0 (5.7-6.6)</b> | <b>5.4 (4.9-5.9)</b> | <b>5.4 (4.7-6.4)</b> | <b>5.3 (4.4-6.0)</b> | <b>5.1 (4.4-5.8)</b> | <b>-</b>               | <b>-</b>             |

**SUPPLEMENTARY TABLE S3 | Diabetes diagnoses and treatment in TB-DM patients.**

|                           |                      | S Africa        |            | Indonesia       |             | Romania          |            | All             |             |
|---------------------------|----------------------|-----------------|------------|-----------------|-------------|------------------|------------|-----------------|-------------|
|                           |                      | Pre-existing DM | New DM     | Pre-existing DM | New DM      | Pre-existing DM  | New DM     | Pre-existing DM | New DM      |
| <b>Number (%)</b>         |                      | 14<br>(74%)     | 5<br>(26%) | 35<br>(73%)     | 13<br>(27%) | 3 (60%)<br>(40%) | 2<br>(40%) | 52<br>(72%)     | 20<br>(28%) |
| <b>Medication, number</b> | Insulin              | 2               | 0          | 9               | 3           | 3                | 0          | 14              | 3           |
|                           | Metformin            | 6               | 0          | 19              | 8           | 0                | 0          | 25              | 8           |
|                           | Other or combination | 6               | 0          | 7               | 2           | 0                | 0          | 13              | 2           |

**SUPPLEMENTARY TABLE S8 | Fixed effects (type III) longitudinal modelling of changes in gene expression in individual patients through time across 4 clinical groups (TB-DM, TB-preDM, TBrel-IH, TB-only).**

| Combined MaSigPro Cluster                       | Cluster 1 | Cluster 2 | Cluster 4 | Cluster 5 | Cluster 6 | Cluster 8 | Cluster 3 | Cluster 7 | Cluster 9 |
|---|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| <b>Number of genes in original Combined MSP</b> | 27        | 50        | 17        | 9         | 28        | 16        | 6         | 10        | 4         |
| <b>Number of genes in core list in combined</b> | 18        | 36        | 9         | 4         | 19        | 11        | 3         | 1         | 1         |
| <b>Clinical Group P-value</b>                   | 0.0217    | 0.0322    | 0.0439    | 0.244     | 0.2466    | 0.2283    |           |           |           |
| <b>Time P-value</b>                             | <0.0001   | <0.0001   | <0.0001   | <0.0001   | <0.0001   | <0.0001   |           |           |           |
| <b>Clinical Group x Time P-value</b>            | 0.4996    | 0.8123    | 0.6796    | 0.8598    | 0.6779    | 0.8211    |           |           |           |

Supplementary Table S4-S7 and S9-S11 can be downloaded at:

<https://www.medrxiv.org/content/10.1101/2022.02.07.22269422v1.supplementary-material>

**SUPPLEMENTARY TABLE S12 | Signature month 6 versus diagnosis obtained by pooling the study groups in the South African cohort or the Indonesian cohort.****Signature South Africa**

| Gene             | Module                      | Coefficient |
|------------------|-----------------------------|-------------|
| <b>Intercept</b> |                             | -31.46231   |
| BLR1             | G protein-coupled receptors | 0.52423     |
| CCL4             | Treg associated genes       | 0.74179     |
| CD3E             | T cell subset markers       | 0.64186     |
| GBP5             | IFN signaling genes         | -0.39161    |
| IFIT5            | IFN signaling genes         | 0.07225     |
| IFITM1/3         | IFN signaling genes         | -0.00135    |
| NLRP1            | Inflammasome components     | 0.033       |
| PRF1             | Cytotoxicity markers        | 0.0287      |
| PTPRCv1          | T cell subset markers       | 0.39243     |
| SEC14L2          | Intracellular transport     | 0.6171      |
| TNF              | Th1 associated genes        | 0.05957     |

**Signature Indonesia**

| Gene             | Module                   | Coefficient |
|------------------|--------------------------|-------------|
| <b>Intercept</b> |                          | -1.15273    |
| CCL13            | Chemokines               | 0.34316     |
| CCL4             | Treg associated genes    | 0.10669     |
| CD3E             | T cell subset markers    | 0.1203      |
| CD4              | T cell subset markers    | 0.06428     |
| FCGR1A           | IFN signaling genes      | -0.77562    |
| FPR1             | Myeloid-associated genes | -0.42241    |
| GBP2             | IFN signaling genes      | -0.09952    |
| GNLY             | Cytotoxicity markers     | 0.34373     |
| KIF1B            | Intracellular transport  | -0.11987    |
| MMP9             | Inflammation             | -0.1168     |
| RORC             | Th17 associated genes    | 0.05677     |
| TBX21            | Th1 associated genes     | 0.75817     |

## Supplementary Methods

### RNASeq data analysis

RNA Samples processed for RNA-Seq analysis were quality-assessed by LabChip GX HiSens RNA system (PerkinElmer). Total RNA samples were processed using the poly-A tail Bioscientific NEXTflex-Rapid-Directional mRNA-seq method with the Caliper SciClone to generate libraries, which were single-end sequenced using the NextSeq500 High Output kit V2 (Illumina) for 75 cycles. Data are deposited in the NCBI-GEO database, accession number GSE193978. STAR (v2.5.1b) [1] was used to align the sequence data from FASTQ files, to the Human g1kv37 reference genome, and quality control performed with FastQC [2]. Downstream analysis was performed in R [3]. HTseq-count (v0.61) was used for transcript quantification [4]. Lowly expressed transcripts for which counts did not exceed 50 across all samples were removed. Data were normalised using R package DESeq2 (1.30.0) [5]. For the MaSigPro [6] analysis, due to the number of timepoints, a quadratic regression model (degrees of freedom = 2) was executed. False discovery correction was done using Benjamini-Hochberg method with an adjusted p-value of < 0.05 deemed significant. Molecular Degree of Perturbation (MDP) analysis was performed using the R package mdp [7]. The R package tmod [8] was used to perform modular analysis on the genes found to be differentially expressed between clinical groups using its HGtest function against all genes as a background. Modules with an adjusted p-value <0.05 were deemed significant. Modular activity was calculated by summing the differential expression of genes in TB/DM relative to TB-only within a module and then dividing by the number of genes within that module. The g:profiler webtool [9] was used for gene ontology and pathway analyses gene lists.

### Reverse-Transcriptase Multiplex Ligation-dependent Probe Amplification (dcRT-MLPA)

Reverse-Transcriptase Multiplex Ligation-dependent Probe Amplification (dcRT-MLPA) was performed using the SALSA MLPA kit (MRC-Holland) as described elsewhere [10]. RT primers and half-probes were designed by Leiden University Medical Centre (LUMC, Leiden, the Netherlands) and comprised sequences for 4 housekeeping genes and 144 selected key immune-related genes to profile specific compartments of the human immune response (Supplementary Table S1). Briefly, 125 ng RNA was reverse transcribed to cDNA by incubation at 37°C for 15 min using Moloney Murine Leukemia Virus (M-MLV) reverse transcriptase (Promega, Leiden, The Netherlands), gene-specific RT-primers (Sigma-Aldrich, Zwijndrecht, The Netherlands), followed by inactivation of the enzyme by heating at 98°C for 2 min. The left- and right-hand half probes were hybridized to the cDNA at 60°C overnight, followed by ligation at 54°C for 15 minutes using ligase-65 (MRC-Holland), and inactivation by heating at 98°C for 5 min. Ligated probes were amplified by

PCR (33 cycles at 95°C for 30 seconds, 58°C for 30 seconds and 72°C for 60 seconds, followed by one cycle at 72°C for 20 minutes). PCR products were 1:10 diluted in Highly deionized (Hi-Di) formamide (ThermoFisher) containing 400HD Rhodamine X (ROX) fluorophore size standard (ThermoFisher). PCR products were denatured at 95°C for 5 min, stored immediately at 4°C and analyzed on an Applied Biosystems 3730 capillary sequencer in GeneScan mode (BaseClear, Leiden, The Netherlands). Trace data were analyzed using GeneMapper software 5 (Applied Biosystems, Warrington, UK). The areas of each assigned peak (arbitrary units) were exported for analysis in R (version 3.6.3). Data were normalized to housekeeping gene glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and signals below the threshold value for noise cutoff in GeneMapper ( $\log_2$  transformed peak area 7.64) were assigned the threshold value for noise cutoff.

dcRT-MLPA data were analysed to identify differentially expressed genes (DEGs) between groups at diagnosis by the non-parametric Mann-Whitney U-test with Benjamini-Hochberg correction for multiple testing. Ingenuity Pathway Analysis (IPA-60467501) (QIAGEN, Hilden, Germany) was used to explore interactive networks between the DEGs. Molecular Degree of Perturbation (MDP) analysis (mdp R package) [7], Partial Least Squares – Discriminant Analysis (PLS-DA) (mixOmics R package) [11] and Pearson correlations of gene expression ( $\log_2$  FC) versus healthy control were performed in R version 4.0.2. Longitudinal changes in gene expression levels from diagnosis (baseline) to 6 months (Indonesian cohort) or 12 months (South African cohort) were assessed by means of linear mixed models for repeated measurements over time. Models were fitted to base 2 logarithm transformed measurements using lme4 package in R using the lmer function [12]. Group-time interactions were included as fixed effects and the patients ID-time interactions were included as random effects. For the South African cohort, we forced a b-spline at 6 months, which enabled us to identify altered gene expressions during treatment (0-6 months) as well as altered gene expression after treatment (6-12 months). Time was coded as 0 for the first timepoint (diagnosis) and as a continuous variable for the time difference between the two time points. P-values were adjusted for multiple testing using the False Discovery Rate (FDR) method of Benjamini-Hochberg [13]. An adjusted p-value < 0.05 and a  $\log_2$ -fold change (FC) < -0.6 and > 0.6 were set as threshold for the identification of differentially expressed genes (DEGs). Genes that were below the detection limit in >90% of the samples per cohort were excluded from analysis. Signatures with the best discriminatory capability were identified using logistic regression with lasso regularization (glmnet R package) [14]. Leave-One-Out Cross Validation (LOOCV) and Train-Test Split (TTS) were used to assess the performance of the trained regression models. The classifying performance of the models were assessed by evaluating sensitivity, specificity, Receiver Operating Characteristic (ROC) curve, and Area Under the

ROC Curve (AUC) with 95% Confidence Interval (CI), and box-and-whiskers-plots representing the predicted probability for each class were used to evaluate the classifying performance of the models.

## Supplementary References

1. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013; 29(1): 15-21.
2. Andrews S. FastQC: a quality control tool for high throughput sequence data. Available at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
3. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2013.
4. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* 2015; 31(2): 166-9.
5. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014; 15(12): 550.
6. Nueda MJ, Tarazona S, Conesa A. Next maSigPro: updating maSigPro bioconductor package for RNA-seq time series. *Bioinformatics* 2014; 30(18): 2598-602.
7. Lever M RP, Nakaya H. mdp: Molecular Degree of Perturbation calculates scores for transcriptome data samples based on their perturbation from controls. R package version 1.12.0. Available at: <https://mdp.sysbio.tools/>.
8. Weiner 3rd J, Domaszewska T. tmod: an R package for general and multivariate enrichment analysis. *PeerJ Preprints* 2016; 4: e2420v1
9. Raudvere U, Kolberg L, Kuzmin I, et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res* 2019; 47(W1): W191-W8.
10. Joosten SA, Goeman JJ, Sutherland JS, et al. Identification of biomarkers for tuberculosis disease using a novel dual-color RT-MLPA assay. *Genes Immun* 2012; 13(1): 71-82.
11. Kim-Anh Le Cao FR, Ignacio Gonzalez, Sebastien Dejean. mixOmics: Omics Data Integration Project. R package version 6.1.1. Available at: <https://CRAN.R-project.org/package=mixOmics>.
12. Bates D, Mächler M, Bolker BM, Walker SC. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 2015; 67.
13. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* 1995; 57(1): 289-300.
14. Friedman J HT, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software* 2010; 33(1), 1–22.