Discerning divergent tuberculosis endotypes: A meta-analysis and systematic review of individual patient data

Andrew R. DiNardo1, Kimal Rajapakse2,3, Tanmay Gandhi2, Sandra L. Grimm2, Tomoki Nishiguchi1, Jan Heyckendorf4, Jaqueline Kahari, Qiniso Dlamini, Christoph Langes5, Jeffrey D. Cirillo7, Stefan HE Kaufmann8,9, Reinout van Crevel10, Mihai G. Netea10,11, Anna M. Mandalakas1, Cristian Coarfa12

*Co-first authors contributing equally. $Co-senior authors contributing equally.

1 The Global Tuberculosis Program, Texas Children’s Hospital, Immigrant and Global Health, Department of Pediatrics, Baylor College of Medicine, Houston, USA
2 Dan L Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, USA
3 Molecular and Cellular Biology Department, Baylor College of Medicine, Houston, USA
4 Division of Clinical Infectious Diseases, Research Center Borstel; German Center for Infection Research (DZIF) Clinical Tuberculosis Unit, Borstel, Germany; Respiratory Medicine &International Health, University of Lübeck, Lübeck, Germany
5 Baylor-Swaziland Children’s Foundation, Mbabane, Swaziland
6 Department of Medicine, Karolinska Institute, Stockholm, Sweden
7 Department of Microbial Pathogenesis and Immunology, Texas A&M Health Science Center, Bryan, TX, USA
8 Max Planck Institute for Infection Biology, Berlin, Germany
9 Hagler Institute for Advanced Study at Texas A&M University, College Station, TX, USA
10 Department of Internal Medicine and Radboud Center for Infectious Diseases, Radboud University Medical Center, Nijmegen, Netherlands
11 Genomics and Immunoregulation, Life & Medical Sciences Institute, University of Bonn, 53115 Bonn, Germany
12 Center for Precision Environmental Health, Baylor College of Medicine, Houston, USA

ABSTRACT

**Rationale:** Host response is a critical factor determining susceptibility to tuberculosis (TB). A delicate balance should be maintained between intracellular immunity against *Mycobacterium tuberculosis* (*Mtb*) and minimizing detrimental immunopathology. Studies have identified incongruous immune responses that can lead to a similar TB disease phenotype. Instead of envisioning that susceptibility to TB follows a singular path, we propose the hypothesis that varied host endotypes exist within the TB clinical phenotype.

**Methods and Results:** Unbiased clustering analysis from 12 publicly available gene expression datasets consisting of data from 717 TB patients and 527 controls, identified 4 TB patient endotypes with distinct immune responses. The two largest endotypes exhibit divergent metabolic, epigenetic and immune pathways. TB patient endotype A, comprising 333 TB patients (46.4%), is characterized by increased expression of genes important for i) glycolysis, ii) IL-2-STAT5, IL-6-STAT3, Type I and II Interferon IFN-γ and TNF signaling and iii) epigenetic-modifying genes. In contrast, TB patient endotype B, comprising 313 TB patients (43.6%), is characterized by i) upregulated NFAT and hormone metabolism, and ii) decreased glycolysis, IFN-γ and TNF signaling. *In silico* evaluation suggests therapies beneficial for endotype A could be detrimental to endotype B, and vice versa. Multiplex ELISA completed from an external validation cohort confirmed a TB patient sub-group with decreased immune upregulation.

**Conclusions:** Host immunity to TB is heterogenous. Unbiased clustering analysis identified distinct TB endotypes with divergent metabolic, epigenetic and immune gene expression profiles that may enable stratified or personalized treatment management in the future.
INTRODUCTION:
For almost half a century, treatment recommendations for patients with drug-susceptible tuberculosis (TB) have not changed. With existing drug combination therapy lasting at least six months, there has been interest in identifying host-directed therapies (HDT) that could both improve the efficacy of and shorten existing treatment regimens. To date, a singular immune correlate of protection against TB has not been identified. In contrast, asthma, COPD and most cancers are managed based on their endotype or molecular sub-classification \(^1\)-\(^3\). For example, individuals with lung cancer are sub-categorized by histopathology (small-cell vs. non-small cell), by genotype (EGFR Exon 18-21 mutations, ALK translocations, ROS1 translocations, BRAF V600 mutations, NTRK fusion products), by PDL-1 expression and stage (localized vs. metastatic) \(^3\). Treatment depends on these sub-categorizations. A classification system for guiding TB therapy has not yet been devised; however new data suggests that simple clinical information may help stratify the duration of TB therapy \(^4\). Identification of host immune endotypes in TB would further guide a stratified based approach for host-directed therapeutics.

Immune studies of TB have identified incongruous immune responses that can lead to a similar TB disease phenotype \(^5\)-\(^9\). Therefore, instead of envisioning TB as an infectious disease with a stereotypical immune response, we sought to assess whether diverse host immune responses could result in a similar clinical phenotype. Previous studies have identified a narrow host therapeutic window with both suppressed immunity or exuberant immunity resulting in decreased \(Mtb\) killing capacity \(^8\),\(^10\),\(^11\). Anti-mycobacterial immunity is complex, requiring a multicomponent approach including multiple cell types. For example, upstream defects in the IL-12-IFN-\(\gamma\) signaling pathway result in decreased IFN-\(\gamma\) production and susceptibility to TB. In contrast, downstream defects in the IL-12-IFN-\(\gamma\) pathway result in excess IFN-\(\gamma\) but decreased IFN-\(\gamma\)-induced gene expression and increased susceptibility to TB \(^7\),\(^12\). Similarly complex is the immune response to TNF: decreased TNF results in intracellular mycobacterial survival, while excess TNF induces immune cell death via necroptosis and extracellular mycobacterial survival \(^5\),\(^11\),\(^13\). A narrow therapeutic window of host mycobacterial immunity has also been described for the MyD88-adaptor-like protein (MAL) and immune checkpoint inhibitors \(^6\),\(^8\),\(^10\). We also reasoned that some individuals would progress to TB despite an apparently well-functioning host immune response if they had been infected with a high bacillary burden, experienced prolonged exposure, or were infected with a hyper-virulent strain of \(Mtb\). Therefore, to identify potential endotypes of host immune response during TB, we implemented an unbiased clustering of publicly available microarray gene expression datasets.

METHODS
Study inclusion
We implemented a systematic review and individual participant meta-analysis according to the PRISMA-IPD guidelines. An a priori hypothesis was established and shared with co-authors prior to implementation of bioinformatics. Publicly available data was identified using PubMed, and the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) repository. Participants without microbiologic confirmation were not included. Studies that did not describe the methods of microbiologic confirmation, or evaluated cells other than whole blood, were excluded. Studies using microarray were used for a discovery cohort and genome-wide (RNA-seq) data sets were used for a validation cohort (Table 1). The Remove Unwanted Variation (RUV) method was used to normalize the data.

### Table 1: Summary of the gene expression studies included.

<table>
<thead>
<tr>
<th>Reference</th>
<th>GEO Identifier</th>
<th>Platform</th>
<th>Healthy Controls used</th>
<th>TB Patients used</th>
<th>Countries Included</th>
<th>Age Range (years)</th>
<th>Method of Microbiologic Confirmation</th>
<th>Inclusion of HIV co-infected individuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Berry 2010</td>
<td>GSE19491</td>
<td>GPL6947</td>
<td>45</td>
<td>103</td>
<td>UK and South Africa</td>
<td>&gt;18</td>
<td>Culture</td>
<td>No</td>
</tr>
<tr>
<td>Bloom 2012</td>
<td>GSE40553</td>
<td>GPL10558</td>
<td>38</td>
<td>166</td>
<td>UK and South Africa</td>
<td>&gt;17</td>
<td>Culture</td>
<td>No</td>
</tr>
<tr>
<td>Cif 2012</td>
<td>GSE31348, GSE36238</td>
<td>GPL573</td>
<td>0</td>
<td>36</td>
<td>South Africa</td>
<td>18 – 65</td>
<td>Smear and culture</td>
<td>No</td>
</tr>
<tr>
<td>Maertzdorf 2012</td>
<td>GSE34608</td>
<td>GPL6480</td>
<td>0</td>
<td>8</td>
<td>Germany</td>
<td>21 – 65</td>
<td>??</td>
<td>No</td>
</tr>
<tr>
<td>Bloom 2013</td>
<td>GSE42834</td>
<td>GPL10558</td>
<td>113</td>
<td>35</td>
<td>UK</td>
<td>&gt;18</td>
<td>Culture</td>
<td>No</td>
</tr>
<tr>
<td>Verhagen 2013</td>
<td>GSE41055</td>
<td>GPL5175</td>
<td>0</td>
<td>2</td>
<td>Venezuela</td>
<td>1.1 – 14.5</td>
<td>Culture</td>
<td>No</td>
</tr>
<tr>
<td>Andersson 2014</td>
<td>GSE39393, GSE39440</td>
<td>GPL10558</td>
<td>68</td>
<td>35</td>
<td>South Africa, Malawi, Kenya</td>
<td>&lt; 15</td>
<td>Culture</td>
<td>Yes</td>
</tr>
<tr>
<td>Kafrou 2014</td>
<td>GSE37250</td>
<td>GPL10558</td>
<td>167</td>
<td>195</td>
<td>South Africa, Malawi, Kenya</td>
<td>19 – 53</td>
<td>Culture</td>
<td>Yes</td>
</tr>
<tr>
<td>Lai 2015</td>
<td>GSE8411</td>
<td>GPL10558</td>
<td>0</td>
<td>31</td>
<td>South Africa</td>
<td>&gt;18</td>
<td>Smear or culture</td>
<td>Yes</td>
</tr>
<tr>
<td>Tiant che 2015</td>
<td>GSE62147</td>
<td>GPL6480</td>
<td>0</td>
<td>26</td>
<td>The Gambia and UK</td>
<td>15 – 79</td>
<td>Culture</td>
<td>No</td>
</tr>
<tr>
<td>Blankeley 2016</td>
<td>GSE33456</td>
<td>GPL10558</td>
<td>61</td>
<td>45</td>
<td>UK</td>
<td>20 – 80</td>
<td>Culture</td>
<td>No</td>
</tr>
<tr>
<td>Walier 2016</td>
<td>GSE73408</td>
<td>GPL11532</td>
<td>35</td>
<td>35</td>
<td>United States</td>
<td>18 – 65</td>
<td>Smear and culture</td>
<td>No</td>
</tr>
</tbody>
</table>

**RNA-Seq Validation Cohort**

<table>
<thead>
<tr>
<th>Reference</th>
<th>GEO Identifier</th>
<th>Platform</th>
<th>Healthy Controls used</th>
<th>TB Patients used</th>
<th>Countries Included</th>
<th>Age Range (years)</th>
<th>Method of Microbiologic Confirmation</th>
<th>Inclusion of HIV co-infected individuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sultan 2018</td>
<td>GSE94438</td>
<td>GPL11154</td>
<td>314</td>
<td>0</td>
<td>The Gambia, South Africa, Ethiopia, Uganda</td>
<td>10 – 18</td>
<td>Symptoms and culture</td>
<td>No</td>
</tr>
<tr>
<td>Leong 2018</td>
<td>GSE101705</td>
<td>GPL18737</td>
<td>16</td>
<td>28</td>
<td>India</td>
<td>16 – 65</td>
<td>Culture</td>
<td>No</td>
</tr>
<tr>
<td>Singha 2018</td>
<td>GSE107995</td>
<td>GPL20301</td>
<td>192</td>
<td>90</td>
<td>UK</td>
<td>17 – 82</td>
<td>Culture</td>
<td>No</td>
</tr>
<tr>
<td>Turner 2020</td>
<td>E-MTAB-8299; (Array Express)</td>
<td>0</td>
<td>54</td>
<td>South Africa</td>
<td>27 – 48</td>
<td>Culture and Xpert</td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>

**Clustering**

To identify distinct TB endotypes, we used the Louvain network-based clustering algorithm implemented in the single-cell RNA-seq analysis package Seurat. Clusters were detected at a wide range of resolutions, from R = 0 to R = 1.2. Transcriptomes of identified clusters were compared to healthy controls; genes with an FDR-adjusted p-value<0.05 and fold change exceeding 2x were considered significant. Pathway enrichment analysis was implemented using Gene Set Enrichment Analysis (GSEA). Chemical compounds associated with the gene signatures were determined using the Library of Integrated Network-based Cellular Signatures (LINCS).

**Gene classifier**
Given the fundamental difference between the microarray and RNA-seq profiling platforms, we normalized independently the microarray data and the RNA-seq data and used the former as the discovery cohort and the latter as the validation cohort. Binary classifier for endotype A and B, which consisted of 89% of the discovery cohort, was built using Logistic Regression.

**Pseudotime**

Transcriptome profile trajectory analysis (pseudotime) was implemented on the 1244 TB patients and healthy controls using Monocle with data for selected gene groups plotted in the R statistical package.

**External Multiplex ELISA Validation Cohort**

In order to validate the gene expression clustering findings, a cohort of TB patients and their asymptomatic household contacts from Eswatini were evaluated (n = 79) \(^{14}\). The study protocol was reviewed by appropriate ethical and institutional review boards and all participation was voluntary and in concordance with the Declaration of Helsinki. Pulmonary TB patients were defined by both the presence of symptoms and all included TB cases had microbiologic confirmation. Twenty-two (55%) and nine (23%) of TB cases and household contacts were HIV co-infected. Included household contacts remained asymptomatic for 12-months after enrollment. Forty-eight percent of controls had a positive Quantiferon test (Qiagen, Hilden, Germany).

**Statistics**

Fisher’s exact test assessed incidence of clinical variables between endotypes. Rank-sum of the cytokines/chemokines from the ELISA validation was used to stratify the top and bottom 50% of TB patients. Differences between ELISA sub-groups were analyzed using Mann-Whitney rank sum test.

**RESULTS**

**Systematic selection of TB patient cohorts with whole blood transcriptomic profiles**

Twenty-two gene expression studies were identified that included microbiologically confirmed pulmonary TB from whole blood. Studies evaluating less than 10,000 genes were excluded. Twelve studies applied microarrays that included 12,468 commonly evaluated genes, including 717 individuals with microbiologically confirmed TB and 527 asymptomatic healthy controls \(^{15-25}\). These 12 studies were used to complete clustering identification of TB-endotypes. Four additional studies used RNA-seq transcriptome profiling and included 172 TB cases and 522 controls. These four studies were reserved for creation of a validation cohort \(^{26-29}\).

**Identification and clinical data characterization of TB endotypes**

In order to identify potential TB endotypes, unbiased clustering of the 717 patients with microarray gene expression data was implemented (Figure 1). To better evaluate the structure of the TB transcriptomes, Louvain network-based clustering was run using a range of
resolutions spanning from $R=0$ to $R=1.2$, and a clustering tree was generated to identify the number of potential sub-clusters. Using resolution $R=0$ four clusters were identified, at $R=0.1-0.6$ five clusters were identified, whereas at $R=1.2$ seven clusters were identified (Figure 2A). To maximize the detection and characterisation power, the resolution $R=0$ was selected for further analysis. At resolution 0, the largest endotypes included 333 (46.4% of all TB patients) and 313 individuals (43.6%), respectively (Figure 2B). The identified endotypes were evaluated by available epidemiologic data such as sex, age, HIV status and country of origin (Supplemental Figure 1). There were no statistical differences in the sex of the different endotypes. In the discovery cohort, only three studies included individuals co-infected with HIV (GSE37250, GSE39939 and GSE58411; Supplemental Figure 1). All individuals co-infected with HIV clustered in endotype A (32.4% of endotype A). At resolution $R=0.6$, endotype A splits into endotype A1 and endotype A2 and, interestingly, all HIV patients clustered into endotype A1 (49% of endotype A2). Deconvolution of whole-blood cell types was performed using CIBERSORT on the TB transcriptomes; endotype A had a relative increase in neutrophils compared to endotype B (Supplemental Figure 3). Only one study included children under 15 years of age (GSE39939), with all children clustering into endotype A (Supplemental Figure 1).

**Characterizing TB endotypes**

We posited that while patients with pulmonary TB have a relatively similar clinical presentation, the host immune response would consist of potentially distinct and divergent endotypes. To test this hypothesis, differentially expressed genes from each cluster were used to determine pathway enrichment, assessed by Gene Set Enrichment Analysis (GSEA) using the Hallmark, REACTOME, GO Biological Process, and KEGG pathway compendia. Hierarchical clustering of GSEA Normalized Enrichment Scores (NES) results showed that endotypes A, C and D had
similar gene expression responses defined by upregulation of type I and type II interferon, TNF, IL-2-STAT5 signaling, glycolysis, oxidative phosphorylation, and Myc targets (Figure 2C). In contrast, these biologic pathways were downregulated in endotype B. The TNF and IFN-γ signaling pathways are necessary, but not sufficient, components of the antimycobacterial immune response. However both pathways have been shown to have narrow therapeutic windows, with either too little or too exuberant TNF and IFN-γ resulting in detrimental immunopathology.

Individuals in TB endotypes A, C and D had upregulation of the IFN-γ signal pathway, while this was downregulated in TB endotype B (Figure 3A). IFN-γ itself was not differentially expressed between the clusters, however
endotype B had decreased expression of the IFN-γ signaling pathway, akin to a downstream Mendelian susceptibility to mycobacterial diseases (MSMD) genetic defect (Figure 3B). Similar to IFN-γ, TNF is necessary, but not sufficient, and TB in patients taking TNF antagonists are at increased risk of disease progression 30. In contrast to TB patient endotype A, endotype B had both decreased expression of TNF as well as decreased TNF via NFκB signaling (Figure 4A,B).

After activation, to increase metabolic demands for proliferation and effector functions, immune cells undergo metabolic shifts including increases in their metabolic activity via transitions towards glycolysis, termed the “Warburg effect” 31,32. Using pathway activity scores, TB patient endotype A had increased oxidative phosphorylation, glycolysis, TCA cycle activation and glucose metabolism, while in contrast these pathways were significantly downregulated in endotype B (Figure 5, Supplemental Figure 2). After brief or chronic antigenic stimulation, this shift in glycolysis induces activation of epigenetic enzymes, such as DNMT and EZH2, and long-lasting chromatin conformation changes 33-38. Therefore, to simultaneously evaluate metabolic, epigenetic and immune gene expression, we implemented trajectory analysis (pseudotime) using Monocle. Pseudotime demonstrated that endotype A expressed simultaneous upregulation of genes related to immunity (IL1B, IL12RB, GZMA, IRF1, STAT1), metabolism and epigenetics, while in contrast endotype B downregulated these same genes (Figure 6).

Multiplex ELISA external cohort validation
To further evaluate the presence of different endotypes within TB patients, multiplex protein data was re-analyzed from an independent cohort of 40 TB patients and 39 healthy controls from Eswatini. Rank-sum analysis was implemented to divide the patients into immune hypo-responsive versus responsive groups based on their production of cytokines and chemokines (IFN-γ, TNF, IL-1β, IL-6, CXCL9, and CXCL10) after whole blood stimulation to mitogen (phytohemagglutinin) (Figure 7A-B). The TB patient sub-groups (top 50% versus bottom 50%) were compared to healthy controls. The hypo-responsive group demonstrated statistically lower levels of IFN-γ, TNF, IL-1β, IL-6, CXCL9, and CXCL10 induction, while the immune-responsive group was statistically similar to the healthy controls (Figure 7C; Mann-Whitney < 0.007).

Analysis of chemical compounds signatures for TB endotypes

A strategy to determine potentially effective drugs against a transcriptome signature is the use of the Library of Integrated Network-based Cellular Signatures (LINCS). Using the gene signatures against healthy controls for endotypes A and B, we searched LINCS ranked lists of
over 5,000 chemical compounds to see which were similar or dissimilar. Similar to the contradictory metabolic, epigenetic and immune signatures, exogenous cytokines IL-2 and IFN-γ demonstrated contradictory connectivity scores between endotypes A and B. Similarly, other previously identified candidates for host-directed therapy (HDT), Vitamin D, glucocorticoids, non-steroidal anti-inflammatory drugs (NSAIDS), cyclooxygenase inhibitors, retinoids and metformin, demonstrated contradictory connectivity scores between TB patient endotypes A and B (Figure 8).

**Validation of endotypes A and B in TB blood RNA-seq cohorts**

Given the importance of the IFN-γ pathway distinguishing endotype A from endotype B, we attempted to infer a classifier using the 200 Hallmark IFN-γ pathway genes. We applied the endotype A vs B IFN-γ gene classifier on an RNA-Seq validation cohort comprised of 172 TB patients and 522 healthy controls (Table 1). We employed the logistic regression method, and applying a 10-fold cross validation in the discovery cohort led to >99%
accuracy. Given the discovery process of endotypes A and B, i.e. using the Louvain network community detection methods, such effective separations of the endotypes via machine learning are expected. Logistic regression classified 152 TB samples as endotype A and 20 as endotype B. We computed pathway enrichment using GSEA between each of the predicted validation subgroups and control samples. Logistic Regression captured the significant divergent gene expression enrichments for immune related pathways (Supplemental Figure 4). IL2/STAT5, IL6/JAK/STAT3, inflammatory response, IFN-α signaling, IFN-γ signaling, and allograft rejection were positively enriched for endotype A, and they were negatively enriched in endotype B in the discovery cohort and in the validation cohort for both classifier methods.

**Figure 8.** Heatmap of connectivity score for select chemical compounds within the TB endotypes A and B based on the Library of Integrated Network-based Cellular Signatures (LINCS). Positive connectivity scores represent chemical inducing gene expression profiles similar to the endotype, while negative connectivity scores represent chemical compounds inducing gene expression profiles antithetical to the endotype.
DISCUSSION

In order to identify different distinct molecular sub-categories (endotypes) of host immune response against *Mtb*, we explored publicly available gene expression datasets to identify at least two distinct endotypes of host gene expression in TB with opposing immune, metabolic, and epigenetic characteristics. The two most predominant TB endotypes display contradictory responses to pathways previously thought to be required for mycobacterial immunity. Endotype A has upregulation of the IFN-γ and TNF, while both of these cytokine-signaling pathways are downregulated in endotype B.

The most commonly evaluated candidates for correlates of protection against *Mtb* are IFN-γ and TNF and their corresponding signaling pathways. Therefore, it is significant that they are contradictory between the two most commonly observed endotypes. Deficiencies in IFN-γ and TNF increase the risk for TB progression. However, animal models have demonstrated that both IFN-γ and TNF have narrow therapeutic windows with exuberant responses resulting in immune-mediated pathology..

Determining appropriateness of the host immune response identified in the two largest divergent endotypes is unclear with the limited available clinical information and lack of quantified bacillary burden. Theoretically, some individuals with endotype A TB have appropriate upregulation of IFN-γ and TNF targeting effector mechanisms against *Mtb*. In contrast, theoretically, some individuals with endotype A have detrimental, excessive IFN-γ and TNF inducing pathology. Combining gene expression analysis with functional immunology and quantitative measures of *Mtb* bacillary burden will clarify the appropriateness of the immune response between endotypes.

The link between metabolism, particularly glycolysis, and immune function has been appreciated for at least 93 years. More recently, metabolism has been demonstrated as a mediator of the epigenetic mechanisms driving immune function. Therefore, it is of particular interest that endotype A displayed upregulation of genes related to metabolism, epigenetics and the IFN-γ and TNF signaling, while in contrast endotype B displayed downregulation of these pathways. Many candidates for host-directed therapies target these pathways. For example, metformin mediates the AKT-mTOR pathway, blunting cellular glycolysis leading to inhibition of chromatin conformational changes that ultimately drive antigen-induced immune function. These data suggest that drugs modulating metabolic and epigenetic mechanisms may need to be evaluated in an endotype-specific manner.

Previously identified candidates for HDTs include exogenous IFN-γ, exogenous GM-CSF, exogenous TNF, TNF inhibitors, NSAIDS, Vitamin D, glucocorticoids, mTOR modulators (rapamycin, metformin), retinoids, and statins. The divergent gene expression profiles identified between the two largest TB endotypes suggest a “one-size fits all” HDT approach may not be feasible. The in silico analysis demonstrated that previously identified HDTs will have contradictory predicted responses in endotype A versus endotype B. If functional studies validate one endotype to have decreased immune responsiveness, then vitamin D or exogenous recombinant IFN-γ would be appropriate HDT. In contrast, if future validation
studies demonstrate one endotype to have pathologic, exuberant immunity, then an NSAID, TNF inhibitor or glucocorticoid would be appropriate. Evaluation of candidate HDTs will likely need to be endotype-specific and animal and in vitro models are needed to recapitulate the clinically relevant endotypes to better evaluate the appropriateness of candidate HDTs. All included studies evaluated host gene expression at baseline, without antigenic or other stimulation. TB is a chronic infection and induces host immune suppression. While many genes down-stream of IFN-γ are elevated in TB patients at baseline, the TB-induced immune exhaustion decreases antigen-induced immune upregulation. HIV suppresses host immunity. Therefore, the fact that all HIV participants clustered into endotype A (resolution 0.2 and below; Supplemental Figure 1) highlights the trouble with over-inferring immune function based on baseline, non-stimulated gene expression measurements. Therefore, future studies will need to evaluate the functional capacity of each endotype upon antigenic stimulation.

This study was predominantly derived from publicly deposited data and therefore has multiple inherent limitations. Progression to TB is related to the interaction between host, pathogen and environmental factors. The progression to a specific endotype of TB is also likely related to similar host, pathogen and environment interactions. However, few pathogen characteristics and very limited epidemiology are available in existing public data repositories. Epidemiologic predispositions likely to drive the divergent host gene expression endotypes include malnutrition, HIV, helminth, tobacco use and or indoor biomass fuel exposure. For example, despite successful deworming, previous schistosomiasis infection ablates mycobacterial immunity and leaves a long-lasting detrimental epigenetic-mediated immune exhaustion. Another limitation is the lack of cell-specific data. Endotype A had increased neutrophils compared to endotype B, one potential explanation why all HIV cases clustered into Endotype A1. The defining feature of HIV infection is decreased cell-mediated immunity, particularly decreased IFN-γ production from lymphocytes. Considering previous evidence that neutrophils drive the gene expression signatures of TB patients, this suggests that cell-specific gene expression evaluations of endotype A1 would identify decreased lymphocyte IFN-γ signaling, but increased non-lymphocyte IFN-γ signaling. Based on the current data, it is unclear if the identified endotypes represent dichotomous groups or a continuous spectrum of host immunity. Similarly, the current data in unable to address if individuals are able to transition from one endotype to another during the disease process.

In summary, unbiased in silico analysis of publicly available data on markers of host immunity in patients with TB provides detailed evidence about the heterogeneity of immune responses in TB patients. Specifically, host gene expression in TB consists of at least two major, distinct endotypes with opposing immune, metabolic and epigenetic transcriptomic signatures. These observations suggest that despite similar clinical phenotypes, different TB patient endotypes display contradictory responses that likely have clinical and pathologic relevance. These unique endotypes are likely to differentially benefit from stratified host-directed therapies.
Supplemental Figure 1. Evaluation of HIV status (A), sex (B), age (C), geographic location (D) by endotype status at resolution 0. (E) HIV status is also shown at Resolution 0.6.

A. HIV status

B. Sex

C. Age (years)

D. Geographic Location

E. HIV status: Res 0.6
Supplemental Figure 2. Metabolic pathways show robust and distinct gene expression profiles in TB patient endotypes. Pathway enrichment scores for KEGG oxidative phosphorylation (A), Reactome glycolysis (B), Reactome glucose metabolism (C), and KEGG TCA cycle (D) in endotypes A and B compared to healthy controls. *** denotes p<0.001 in Mann-Whitney unpaired non-parametric two-tailed test. Whiskers denote minimum-maximum and dotted line represents the healthy control median.

Supplemental Figure 3. Deconvolution of whole-blood cell types using CIBERSORT.
**Supplemental Figure 4.** Enrichment of Hallmark Immune Pathways in the validation RNA-Seq cohort as compared to the discovery microarray cohort.

<table>
<thead>
<tr>
<th>Endotype:</th>
<th>Microarray Discovery</th>
<th>RNA-Seq Validation</th>
<th>Microarray Discovery</th>
<th>RNA-Seq Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>TB Samples:</td>
<td>333 313</td>
<td>152 20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interferon-γ Response</td>
<td>7.32</td>
<td>-4.76</td>
<td>8.83</td>
<td>-2.22</td>
</tr>
<tr>
<td>Interferon-α Response</td>
<td>6.65</td>
<td>-4.84</td>
<td>6.92</td>
<td>-1.52</td>
</tr>
<tr>
<td>Allograft Reaction</td>
<td>3.90</td>
<td>-4.50</td>
<td>3.32</td>
<td>-2.97</td>
</tr>
<tr>
<td>Inflammatory Response</td>
<td>3.38</td>
<td>-2.26</td>
<td>5.29</td>
<td>-1.64</td>
</tr>
<tr>
<td>IL2-STAT5 Signaling</td>
<td>2.62</td>
<td>-2.79</td>
<td>2.30</td>
<td>-1.56</td>
</tr>
<tr>
<td>IL6-JAK-STAT3 Signaling</td>
<td>2.48</td>
<td>-1.66</td>
<td>5.04</td>
<td>-1.82</td>
</tr>
<tr>
<td>Apical Junction</td>
<td>-1.59</td>
<td>1.78</td>
<td>-1.57</td>
<td>3.29</td>
</tr>
<tr>
<td>Spermatogenesis</td>
<td>-2.56</td>
<td>2.50</td>
<td>-1.62</td>
<td>2.21</td>
</tr>
<tr>
<td>Myogenesis</td>
<td>-2.67</td>
<td>3.42</td>
<td>-2.07</td>
<td>4.38</td>
</tr>
<tr>
<td>KRAS Signaling down</td>
<td>-4.22</td>
<td>4.53</td>
<td>-1.81</td>
<td>2.22</td>
</tr>
</tbody>
</table>

The NES (Normalized Enrichment Score) values are depicted in the bar chart, with colors indicating the comparison between A and B for each pathway in the discovery and validation cohorts.
REFERENCES


accurately discriminate active tuberculosis from latent infection in individuals from south India. Tuberculosis (Edinb) 2018;109:41-51.


