Published in final edited form as:

Int J Tuberc Lung Dis. 2018 August 01; 22(8): 950–958. doi:10.5588/ijtld.17.0855.

Blood transcriptomic markers of *Mycobacterium tuberculosis* load in sputum

Kathryn M. Dupnik¹, James M. Bean², Myung Hee Lee¹, Marc Antoine Jean Juste³, Lucy Skrabanek⁴, Vanessa Rivera¹, Charles K. Vorkas^{1,2}, Jean W. Pape^{1,3}, Daniel W. Fitzgerald¹, and Michael Glickman^{2,5}

¹Department of Medicine, Weill Cornell Medical College, New York, NY 10065, USA

²Immunology Program, Memorial Sloan Kettering Cancer Center, New York, NY 10021, USA

³GHEKIO center, Port au Prince, Haiti

⁴Applied Bioinformatics Core and Department of Physiology and Biophysics, Weill Cornell Medicine, New York, NY 10021, USA

⁵Division of Infectious Diseases, Memorial Sloan Kettering Cancer Center, New York, NY 10021, USA

SUMMARY

Background: Peripheral blood transcriptome signatures that distinguish active pulmonary tuberculosis (TB) from control groups have been reported, but correlations of these signatures with sputum mycobacterial load are incompletely defined.

Methods: We assessed the performance of published TB transcriptomic signatures in Haiti, and identified transcriptomic biomarkers of TB bacterial load in sputum as measured by GeneXpert molecular testing. People in Port au Prince, Haiti with untreated pulmonary TB (n=51) were included in this study, 19 people with low and 32 with high sputum *M. tuberculosis* load. Peripheral whole blood transcriptomes were generated using RNA-sequencing.

Results: Twenty of the differentially expressed transcripts in TB vs. no TB were differentially expressed in people with low vs high sputum mycobacterial loads. The difference between low and high bacterial load groups was independent of radiographic severity. In a published dataset of transcriptomic response to TB therapy, this 20-gene subset was more treatment-responsive at 6 months than the full active TB signature.

Conclusion: In this study, we identified genes whose transcript levels in the blood distinguish active tuberculosis with high vs. low *M. tuberculosis* loads in the sputum. These transcripts may

Corresponding author: Kathryn M. Dupnik, MD, Department of Medicine, Weill Cornell Medical College, 402 E 67th Street, 2nd floor, New York, NY 10065, kad9040@med.cornell.edu, Phone: 646-962-8140, Fax: 646-962-0285. Conceptualization, K.M.D. and M.G.; Methodology, K.M.D., J.M.B., D.W.F., and M.G.; Software, K.M.D., J.M.B., M.H.L., and L.S.; Formal analysis, K.M.D., M.Y.H., and J.M.B.; Data curation, J.M.B.; Investigation, K.M.D., M.A.J.J., V.R., and C.K.V.; Resources, J.W.P.; Data curation J.M.B.; Writing – Original Draft, K.M.D.; Writing – Review and Editing, K.M.D., J.M.B., M.Y.H., D.W.F., and M.G.; Visualization J.M.B. and M.H.L.; Supervision, J.W.P., D.W.F., and M.G.; Funding Acquisition K.M.D., J.W.P., D.W.F., and M.G.

The authors have no financial conflicts of interest to declare.

reveal mechanisms of mycobacterial control of *M. tuberculosis* during active infection, as well as identifying potential biomarkers for bacterial response to anti-TB treatment.

Keywords

pulmonary tuberculosis; RNA-Seq; GeneXpert; nucleic acid amplification test

INTRODUCTION

Peripheral blood transcriptome signatures for pulmonary tuberculosis (TB) have been reported in the literature. Several groups have reported transcriptional signatures in peripheral blood, which correlate with pulmonary TB in adults¹, children², and HIV-infected and uninfected people³. Additional studies have indicated that these signatures normalize with treatment⁴ and can predict progression from latent to active TB⁵. Maertzdorf *et al.* conducted a meta-analysis of published transcriptional studies on human TB and proposed a four-gene signature as a potential diagnostic⁶ whereas a similar analysis by Sweeney *et al.* derived a distinct three-gene signature⁷. These studies included people from Africa or Europe, so their performance across a broader geographic range is unknown. In addition, they did not include microbiologic correlates such as sputum bacterial load. To expand the utility of this approach to new geographic regions, and to search for transcriptomic markers of bacterial load, we completed transcriptome studies of people with active TB in Port au Prince, Haiti, for whom detailed microbiologic data was available.

Nucleic acid amplification tests for *M. tuberculosis* in sputum samples is a standard diagnostic tool, including in resource-limited settings. The GeneXpert system detects *M. tuberculosis* DNA and rifampicin resistance⁸. The GeneXpert system reports a binary positive or negative result for clinical use. However, the test also measures the relative abundance of *M. tuberculosis* DNA in sputum, with the magnitude of the amplification providing a semi-quantitative measure of the bacterial load in the sputum sample being tested^{9,10}. The range of bacterial concentrations that Xpert can detect is 10² to 10⁷ cfu/mL¹¹. GeneXpert is more standardizable between laboratories and correlates with smear microscopy grades¹¹.

The relation between peripheral blood transcriptome and sputum bacillary load in people with pulmonary TB is not known. This relationship could be important to develop markers of *M. tuberculosis* clearance during treatment. A “biological gradient” correlation with gene expression would validate the association of gene expression with disease state¹². Therefore, we conducted this study to identify genes whose expression varies with *M. tuberculosis* load in sputum.

STUDY POPULATION AND METHODS

Study Design.

We conducted a case-control study to determine if any of the genes in previously published TB signatures were associated with sputum mycobacterial load. This study and consent forms were reviewed and approved by the institutional review boards of the Groupe Haitien

d'Etude du Sarcome de Kaposi et des Infections Opportunistes (GHESKIO) centers and Weill Cornell Medical College. All participants provided written informed consent for participation in the study.

Study site.

Haiti has the highest rate of TB in the Americas, 188 new cases per 100,000 population in 2016¹³. GHESKIO centers in Port au Prince, Haiti is a Haitian non-profit organization that is the largest provider of HIV and TB care in the Americas. Over 3,000 people are diagnosed with TB at GHESKIO each year.

Study population and laboratory testing.

People with pulmonary TB were enrolled from TB ambulatory care centers at GHESKIO. Microbiologic diagnosis of TB was AFB smear or GeneXpert MTB/RIF (Cepheid, CA, USA) positivity with subsequent culture positivity (BACTEC MGIT™ system (Becton Dickinson, NJ, USA). All participants were HIV-negative using the Determine HIV-1/2 Antigen/Antibody test (Alere-Abbott, Illinois, USA). Chest x-rays were obtained as part of standard of care. The number of lung quadrants with abnormalities on chest x-ray and the presence or absence of a cavitary lesion was recorded. Adult community controls were HIV-uninfected and with no personal history or symptoms of TB. Controls were tested for latent TB using the QuantiFERON®-TB Gold In-Tube interferon-gamma release assay (IGRA) (Qiagen, Germantown, MD, USA). Blood was collected prior to starting anti-TB therapy.

Assignment to sputum mycobacterial load group.

The GeneXpert test uses cycle threshold (Ct), which is inversely correlated with the abundance of TB in the sample, to generate semi-quantitative results. To create a dichotomous variable for screening for gene expression differences by mycobacterial load, we binned the GeneXpert quantitation score as follows. People were classified as having “low sputum mycobacterial load” if the GeneXpert level was negative, low, or very low (Ct > 22), and “high sputum mycobacterial load” if the GeneXpert level was medium or high (Ct ≤ 22)^{11,14}. Genes that screened as differentially expressed between “low sputum mycobacterial load” and “high sputum mycobacterial load” were analyzed in the sub-categories assigned by the GeneXpert machine: “high” if Ct ≤ 16, “medium” if 16 < Ct ≤ 22, “low” if 22 < Ct ≤ 28, and “very low” if 28 < Ct ≤ 38.

RNA and Transcriptome generation.

Peripheral blood was collected into Tempus™ Blood RNA tubes (Applied Biosystems). RNA was extracted using the Tempus™ Spin RNA Isolation Kit (Ambion), with addition of on-column DNase treatment (AbsoluteRNA, Ambion). RNA was ribo-depleted by polyA selection and libraries were generated using TruSeq (Illumina, San Diego, CA). Paired-end RNA sequences (50×50PE) were generated with HiSeq 2500 (Illumina) with at least 50 million reads per sample. Sequence integrity was verified using FastQC (Babraham Bioinformatics). Sequences were aligned to the human genome (version hg38) using STAR aligner^{15,16} and transcript counts were estimated using featurecounts¹⁷. Quality of aligned and counted reads was assessed using Quality of RNA-Seq ToolSet (QoRTs)¹⁸.

Comparison of Haiti signature to published TB signatures.

Comparison signatures were the 86-gene active TB signature from Berry et al [1], the 16-gene TB progression risk signature from Zak et al [5], the three-gene TB signature from Sweeney et al [7], and the four-gene TB signature from Maertzdorf et al. Expression values from DESeq2 were normalized and variance-stabilized using pheatmap¹⁹, and the values for gene sets were subjected to unsupervised hierarchical clustering using Euclidean distance measures. Violin plots of TB scores for the three-gene and four-gene signatures were generated with the vioplot package in R²⁰. Kernel density plots were smoothed using default optimal smoothing parameters for normal curves.

Assessing gene expression changes during TB treatment.

Normalized expression values for probes from the 393-gene Berry TB signature were obtained from 29 published blood transcriptional profiles from people with TB in South Africa before, during, and after TB treatment (GSE40553; [4]). This TB signature discriminated TB from community controls⁴. We excluded those genes that were present in the signature but which are no longer mapped as genes in the National Center for Biotechnology Information (<https://ncbi.nlm.nih.gov>). Expression of genes from our 20-gene bacterial load signature were compared to expression of the remainder of genes in the 393-gene signature (n=320). For each gene, the log₂ difference in expression relative to pre-treatment was calculated for each time point. Gene expression was considered to be altered if a 2-fold or greater difference relative to pre-treatment was observed. Boxplots showing the absolute values of the post-treatment change in each sub-set at the different time points were generated, and rates were compared with the Wilcoxon rank sum test.

Statistical Analysis.

DESeq2 was used to compare transcriptome count data^{21,22} between TB vs. non-TB. Differentially expressed genes with a false discovery rate (FDR) of less than or equal to 0.05 were included in analyses²³. We identified a TB-associated gene set that “overlapped” with the published signature for people with TB in South Africa¹. To identify genes correlated with bacterial load, expression levels of genes in the “overlap” signature in high TB vs low TB were compared using DESeq2. P-values were adjusted using a Benjamini-Hochberg correction for genes in the overlap TB signature. Spearman correlation was calculated to assess the trend between gene expression and sputum mycobacterial load in the GeneXpert sub-categories (high, medium, low, very low, negative) and no TB. To assess the three-gene and four-gene TB signature scores in TB vs. non-TB, and to assess differences in response to TB treatment between the 20-gene signature found in this study and the 393-gene Berry signature, we used the Wilcoxon rank sum test.

RESULTS

Study population.

Fifty-one people with microbiologically confirmed pulmonary TB were included, of whom 19 had low *M. tuberculosis* load and 32 had high *M. tuberculosis* load. Controls were 21 people from the community with no signs, symptoms, or history of TB. Within the control

group, 15 (71%) were IGRA-positive. The difference in number of lung quadrants involved on chest x-ray did not differ significantly between low and high *M. tuberculosis* groups (1.6 vs. 1.9, respectively). There were 7 people with cavitory lesions on chest x-ray, all in the high *M. tuberculosis* group (Table 1).

Comparison of differentially expressed genes in Haitian cohort with published TB signatures.

From the 393 genes in the TB signature of Berry *et al.*, there were 320 unique genes that mapped to our transcriptome of Haitian people with TB¹. There were 246 genes differentially expressed in our population. We tested the performance of the smaller TB-specific Berry 86-gene signature in classifying active TB in our cohort, and found that active TB and controls clustered together (Figure 1A). We also tested the performance of the signature reported by Zak *et al.*⁵, and found outstanding segregation of active TB from non-TB (Figure 1B).

We next examined the predictive power of published three- and four-gene signatures associated with active pulmonary TB. The three-gene signature of Sweeney *et al.*, which includes dual specificity phosphatase 3 (DUSP3), guanylate binding protein 5 (GBP5), and kruppel-like factor 2 (KLF2) accurately classified active TB from non-TB (DUSP3 + GBP5 – KLF2; Figure 2). We calculated the TB score based on these three genes and found that the score was significantly higher in active TB compared with non-TB ($W=14$, $p=6.57 \times 10^{-16}$, Figures 2B). The four-gene signature proposed by Maertzdorf *et al.*⁶, which includes guanylate binding protein 1 (GBP1), purinergic receptor P2Y14 (P2RY14), and interferon induced transmembrane protein 3 (IFITM3), and inhibitor of DNA binding 3, HLH protein (ID3) expression (GBP1 + P2RY14 + IFITM3 – ID3), also was significantly different in TB vs non-TB groups ($W=58$, $p=5.60 \times 10^{-12}$, Figure 3). Taken together, these results indicate that previously reported active TB signatures perform well in the Americas.

Derivation of transcriptomic markers of sputum mycobacterial load.

To derive markers of sputum mycobacterial load, we identified genes in the active TB signature that were differentially expressed in blood of people with high vs low sputum *M. tuberculosis*. When comparing differentially expressed genes by sputum mycobacterial load to the signature for active TB reported by Berry *et al.* from a cohort in South Africa¹, there were 20 gene products differentially expressed between low vs. high bacterial loads (Table 2). Expression levels of three representative genes (SOCS3, CR1, OSM) are shown as violin plots in Figure 4A. The expression levels of the 20 differentially expressed genes by sputum mycobacterial load are shown in Supplemental Figure 1. To exclude the possibility that the high bacterial load was acting as a surrogate marker for pathologic severity due to the 7 cavitory cases, a sensitivity analysis looked at differential expression of genes in low vs. high bacterial loads with the 7 cavitory cases excluded. Eighteen of 20 genes that were significantly different in the primary analysis remained in the top 20 genes by adjusted p-value ranking when cavitory cases were excluded. The remaining two genes were numbers 22 and 26 by adjusted p-value.

For the 20 genes differentially expressed in the high sputum mycobacterial load group, we compared the trend in gene counts by bacterial burden using the no TB group and quantitation as generated by GeneXpert: high, medium, low, very low, and negative (Supplemental Figure 1). There were positive correlations between gene counts and bacterial load with additional subgroups. When the 7 people with cavitory disease on chest x-ray were removed in a sensitivity analysis, the correlations remained (Supplemental Figure 2).

Bacterial load transcriptomic markers resolve more quickly with treatment.

Our identification of a 20- gene signature that correlates with high bacterial load in sputum may indicate that these gene expression differences resolve more quickly with TB therapy than a broader transcriptional signature that may reflect multiple aspects of the disease process. To examine this question, we analyzed a published dataset in which the 393 gene signature resolved with therapy⁴. We asked whether the 20 gene bacterial load signature returned to baseline more quickly than the remaining 373 genes that we did not find associated with high bacterial load. We found that at six months, the 20-gene bacterial load signature had normalized more quickly than the broader signature, whereas at 12 months there was no difference (Figure 5).

DISCUSSION

Our study identifies transcripts in peripheral blood that differ between active TB cases with high vs. low *M. tuberculosis* loads in sputum, as measured by molecular testing. Some of the differentially expressed genes have been studied for their involvement in control of mycobacterial infection and replication. This finding may support further investigations into the mechanisms by which these gene products contribute to control of *M. tuberculosis* during active infection, and potential biomarkers of *M. tuberculosis* response to anti-TB treatment. Additionally, to our knowledge this is the first study of human peripheral blood transcriptome in pulmonary tuberculosis in the Western Hemisphere.

We found that differential gene expression between TB vs. non-TB in this cohort were comparable to published data^{1,5}. Both the 393-gene Berry *et al.* and 16-gene Zak *et al.* TB signatures discriminated active TB from non-TB in our population. Additionally, the 3-gene and 4-gene TB signatures were significantly different in TB vs. non-TB. This indicates that published signatures can identify active TB in a western hemisphere population. This suggests that TB diagnostics based on gene expression will be applicable to diverse geographic areas.

Genes that are differentially expressed in TB, and at different levels between low and high sputum mycobacterial loads, could be used as biomarkers of treatment response. One caveat is that sputum mycobacterial DNA load is a proxy for bacterial burden in the airways, and there may be different processes occurring in the lung parenchyma. Interestingly, the correlation between gene counts and sputum mycobacterial load persisted in a sensitivity analysis where the 7 people with cavitory disease on chest x-ray were removed from the high mycobacterial load group. This result indicates that the transcriptomic differences between high and low groups truly reflect bacterial burden rather than pathologic or radiographic

severity. In clinical care, biomarkers that correlate with bacterial burden may enable early detection of treatment failure. These biomarkers would also facilitate early bactericidal activity (EBA) studies in TB drug development²⁴. The 20 genes differentially expressed in people with high sputum *M. tuberculosis* load may be such biomarker candidates if validated in longitudinal studies of treatment response. The 20 genes differentially expressed between low and high sputum mycobacterial load were more likely to be significantly changed from baseline TB treatment initiation at 6 months, compared with other genes in the active TB gene signature¹. Future studies will examine the expression of these genes in paucibacillary and extrapulmonary *M. tuberculosis* disease. The progressively higher expression of our 20-gene signature from no TB to low-load to high-load pulmonary TB suggests that we will also find differential expression in these more difficult to diagnose forms of TB.

Several of the genes differentially expressed in the high sputum mycobacterial load group are of particular interest because of previous studies on proteins encoded by these genes in TB pathogenesis. Suppressor of cytokine signaling 3 (SOCS3), increased in the high bacterial load group, is a regulator of the Signal Transducers and Activators of Transcription 3 (STAT3) pathway (reviewed in Rottenberg *et al*²⁵). Studies have reported increased SOCS3 expression in blood T cells in TB compared with non-TB²⁶. In bronchoalveolar lavage (BAL) and pleural cells of people with pulmonary TB, moderate-severe pulmonary TB had higher SOCS3 expression. Blood-derived macrophages infected with *M. tuberculosis* H37Rv in vitro also had increased SOCS3 expression. SOCS expression was greater than 10 times higher in macrophages than in PBMC exposed to *M. tuberculosis*²⁷. This may explain the conflicting findings of other studies on PBMC in advanced pulmonary TB, which did not find increased SOCS3 expression in response to *M. tuberculosis* H37Rv^{28,29}. Our findings suggest that the increase in SOCS3 expression is also associated with increased sputum mycobacterial load. Further studies to look at SOCS3 expression in different immune cell subsets in the peripheral blood of people with high vs. low sputum TB load will be informative.

Another gene with increased expression in high sputum TB load was oncostatin M (OSM). In blood-derived monocytes and macrophages infected with *M. tuberculosis* H37Rv, OSM is released in a dose-dependent manner. This is an early response, with maximum OSM 72 hours after infection³⁰. With TNF- α , OSM stimulates matrix metalloproteinase (MMP) secretion by pulmonary fibroblasts in vitro. MMP-1 cleaves type 1 collagen in the lung³¹, likely where fibroblasts are found around tuberculous granulomas. Increase in OSM could reflect increased numbers of alveolar macrophages infected with *M. tuberculosis* in the high load sputa, or of increased OSM secretion from macrophages with a greater number of intracellular *M. tuberculosis* per macrophage. We propose future studies looking at OSM in the sputum of people with high vs. low sputum bacterial loads.

One way that *M. tuberculosis* interacts with macrophages is via human monocyte complement C3b/C4b receptor 1 (complement receptor 1 or CR1). When *M. tuberculosis* is phagocytosed in association with complement via CR1 binding, phagosome-mediated killing is diminished, possibly by modifying calcium signaling which impacts fusion with the lysosome³². The CR1 on erythrocytes bind immune complexes and facilitate their removal in liver and spleen. CR1 also includes the Knops blood group antigens³³. Although

sample sizes were relatively small, researchers have reported association between a subset of the Knops blood group genotypes and resistance to TB in Mali and in the Aché people of Brazil^{34,35}

CONCLUSIONS

We provide confirmation of active TB transcriptomic signatures in Haiti, the first TB transcriptome study of a western hemisphere population. Further, this study contributes new knowledge correlating expression of TB-associated genes in the peripheral blood with sputum mycobacterial load. Longitudinal studies of TB treatment focused on the 20 genes associated with sputum mycobacterial load will give additional insight into changes in systemic gene expression with declining bacterial load and substantiate these genes as biomarkers of early bactericidal activity.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We acknowledge the study participants, care teams, and research teams at GHESKIO Centers in Port au Prince, Haiti. We thank the WCMC Genomics Core Laboratory for RNA quality assessments, the MSKCC Genomics Core laboratory for RNA sequencing, and Mary J. Reust in the Center for Global Health Laboratory at WCMC.

This work was supported by the TRBU network of the National Institute of Allergy and Infectious Disease at the National Institutes of Health (U19-AI111143 (MG, DWF, JB)), UL1-TR000457 (KMD), T32-AI007613 (CKV), K24-AI098627 (DWF), D43-TW010062, and P30-CA008748 (MG). The funders had no role in study design or execution.

BIBLIOGRAPHY AND REFERENCES CITED

1. Berry MP, Graham CM, McNab FW, Xu Z, Bloch SA, Oni T, et al. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature* 2010;466(7309):973–7. [PubMed: 20725040]
2. Anderson ST, Kaforou M, Brent AJ, Wright VJ, Banwell CM, Chagaluka G, et al. Diagnosis of childhood tuberculosis and host RNA expression in Africa. *N Engl J Med* 2014;370(18):1712–23. [PubMed: 24785206]
3. Kaforou M, Wright VJ, Oni T, French N, Anderson ST, Bangani N, 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. [PubMed: 24167453]
4. Bloom CI, Graham CM, Berry MP, Wilkinson KA, Oni T, Rozakeas F, et al. Detectable changes in the blood transcriptome are present after two weeks of antituberculosis therapy. *PLoS One* 2012;7(10):e46191. [PubMed: 23056259]
5. Zak DE, Penn-Nicholson A, Scriba TJ, Thompson E, Suliman S, Amon LM, et al. A blood RNA signature for tuberculosis disease risk: a prospective cohort study. *Lancet* 2016;387(10035):2312–22. [PubMed: 27017310]
6. Maertzdorf J, McEwen G, Weiner J, 3rd, Tian S, Lader E, Schriek U, et al. Concise gene signature for point-of-care classification of tuberculosis. *EMBO Mol Med* 2016;8(2):86–95. [PubMed: 26682570]
7. Sweeney TE, Braviak L, Tato CM, Khatri P. Genome-wide expression for diagnosis of pulmonary tuberculosis: a multicohort analysis. *Lancet Respir Med* 2016;4(3):213–24. [PubMed: 26907218]
8. Boehme CC, Nabeta P, Hillemann D, Nicol MP, Shenai S, Krapp F, et al. Rapid molecular detection of tuberculosis and rifampin resistance. *N Engl J Med* 2010;363(11):1005–15. [PubMed: 20825313]

9. Theron G, Peter J, Calligaro G, Meldau R, Hanrahan C, Khalfey H, et al. Determinants of PCR performance (Xpert MTB/RIF), including bacterial load and inhibition, for TB diagnosis using specimens from different body compartments. *Sci Rep* 2014;4:5658. [PubMed: 25014250]
10. Shenai S, Amisano D, Ronacher K, Kriel M, Banada PP, Song T, et al. Exploring alternative biomaterials for diagnosis of pulmonary tuberculosis in HIV-negative patients by use of the GeneXpert MTB/RIF assay. *J Clin Microbiol* 2013;51(12):4161–6. [PubMed: 24108610]
11. Blakemore R, Nabeta P, Davidow AL, Vadwai V, Tahirli R, Munsamy V, et al. A multisite assessment of the quantitative capabilities of the Xpert MTB/RIF assay. *Am J Respir Crit Care Med* 2011;184(9):1076–84. [PubMed: 21836139]
12. Hill AB. The Environment and Disease: Association or Causation? *Proc R Soc Med* 1965;58:295–300. [PubMed: 14283879]
13. WHO. Tuberculosis profile: Haiti 2016 [Available from: https://extranet.who.int/sree/Reports?op=Replet&name=%2FWHO_HQ_Reports%2FG2%2FPROD%2FEXT%2FTBCountryProfile&ISO2=HT&LAN=EN&outtype=html].
14. Ocheretina O, Byrt E, Mabou MM, Royal-Mardi G, Merveille YM, Rouzier V, et al. False-positive rifampin resistant results with Xpert MTB/RIF version 4 assay in clinical samples with a low bacterial load. *Diagn Microbiol Infect Dis* 2016;85(1):53–5. [PubMed: 26915638]
15. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;29(1):15–21. [PubMed: 23104886]
16. Rosenbloom KR, Armstrong J, Barber GP, Casper J, Clawson H, Diekhans M, et al. The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res* 2015;43(Database issue):D670–81. [PubMed: 25428374]
17. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 2014;30(7):923–30. [PubMed: 24227677]
18. Hartley SW, Mullikin JC. QoRTs: a comprehensive toolset for quality control and data processing of RNA-Seq experiments. *BMC Bioinformatics* 2015;16:224. [PubMed: 26187896]
19. RCoreTeam. R: A language and environment for statistical computing R Foundation for Statistical Computing. Vienna, Austria 2016 [Available from: <https://www.R-project.org/>].
20. Adler D Package ‘vioplot’ 2005 [Available from: <https://cran.r-project.org/web/packages/vioplot/vioplot.pdf>].
21. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol* 2010;11(10):R106. [PubMed: 20979621]
22. 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. [PubMed: 25516281]
23. Benjamini YHY. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society* 1995;57(1).
24. Bonnett LJ, Ken-Dror G, Koh G, Davies GR. Comparing the efficacy of drug regimens for pulmonary tuberculosis: meta-analysis of endpoints in early phase clinical trials. *Clin Infect Dis* 2017.
25. Rottenberg ME, Carow B. SOCS3 and STAT3, major controllers of the outcome of infection with *Mycobacterium tuberculosis*. *Semin Immunol* 2014;26(6):518–32. [PubMed: 25458989]
26. Jacobsen M, Repsilber D, Kleinstaub K, Gutschmidt A, Schommer-Leitner S, Black G, et al. Suppressor of cytokine signaling-3 is affected in T-cells from tuberculosis TB patients. *Clin Microbiol Infect* 2011;17(9):1323–31. [PubMed: 20673263]
27. Ashenafi S, Aderaye G, Bekele A, Zewdie M, Aseffa G, Hoang AT, et al. Progression of clinical tuberculosis is associated with a Th2 immune response signature in combination with elevated levels of SOCS3. *Clin Immunol* 2014;151(2):84–99. [PubMed: 24584041]
28. Masood KI, Rottenberg ME, Carow B, Rao N, Ashraf M, Hussain R, et al. SOCS1 gene expression is increased in severe pulmonary tuberculosis. *Scand J Immunol* 2012;76(4):398–404. [PubMed: 22670716]
29. Masood KI, Rottenberg ME, Salahuddin N, Irfan M, Rao N, Carow B, et al. Expression of M. tuberculosis-induced suppressor of cytokine signaling (SOCS) 1, SOCS3, FoxP3 and secretion of IL-6 associates with differing clinical severity of tuberculosis. *BMC Infect Dis* 2013;13:13. [PubMed: 23320781]

30. O’Kane CM, Elkington PT, Jones MD, Caviedes L, Tovar M, Gilman RH, et al. STAT3, p38 MAPK, and NF-kappaB drive unopposed monocyte-dependent fibroblast MMP-1 secretion in tuberculosis. *Am J Respir Cell Mol Biol* 2010;43(4):465–74. [PubMed: 19915152]
31. O’Kane CM, Elkington PT, Friedland JS. Monocyte-dependent oncostatin M and TNF-alpha synergize to stimulate unopposed matrix metalloproteinase-1/3 secretion from human lung fibroblasts in tuberculosis. *Eur J Immunol* 2008;38(5):1321–30. [PubMed: 18398932]
32. Malik ZA, Denning GM, Kusner DJ. Inhibition of Ca(2+) signaling by Mycobacterium tuberculosis is associated with reduced phagosome-lysosome fusion and increased survival within human macrophages. *J Exp Med* 2000;191(2):287–302. [PubMed: 10637273]
33. Storry JR, Castilho L, Daniels G, Flegel WA, Garratty G, de Haas M, et al. International Society of Blood Transfusion Working Party on red cell immunogenetics and blood group terminology: Cancun report (2012). *Vox Sang* 2014;107(1):90–6. [PubMed: 24372289]
34. Noumsi GT, Tounkara A, Diallo H, Billingsley K, Moulds JJ, Moulds JM. Knops blood group polymorphism and susceptibility to Mycobacterium tuberculosis infection. *Transfusion* 2011;51(11):2462–9. [PubMed: 21569042]
35. Lindenau JD, Salzano FM, Hurtado AM, Hill KR, Hutz MH The role of variants from the innate immune system genes in tuberculosis and skin test response in a Native American population. *Hum Immunol* 2016;77(10):981–4. [PubMed: 27346696]

Figure 1.
 Performance of two published active TB transcriptomic signatures in a Haitian population.
 A) The 86-gene signature reported in Berry et al.¹ was applied to the cohort of active TB cases (grey) or controls (blue) by unsupervised hierarchical clustering. B) The active TB signature reported by Zak et al.⁵, was applied to the active TB cohort and controls.

Figure 2. Discriminatory power of the TB three gene score reported by Sweeney *et al*⁷ in people with TB in Haiti. A) Unsupervised hierarchical clustering of expression values for each of the three genes in the score for TB cases (grey) and controls (blue). B) Violin plots of the calculated three gene score in control and active TB cases. The internal box is the interquartile range, with the white dot being the median. Whiskers extend to 1.5 times the interquartile range or the maximum/minimum, whichever is closer to the median. The p-value for the difference in scores is 6.6×10^{-16} . KLF2 = kruppel like factor 2, DUSP3 = dual specificity phosphatase 3, GBP5= guanylate binding protein 5

Figure 3.

Discriminatory power of the TB four gene score reported by Maertzdorf *et al*⁶ in people with TB in Haiti. A) Unsupervised hierarchical clustering of expression values for each of the four genes in the score for TB cases (grey) and controls (blue). B) Violin plots of the calculated four gene score in control and active TB cases. The internal box is the interquartile range, with the white dot being the median. Whiskers extend to 1.5 times the interquartile range or the maximum/minimum, whichever is closer to the median. The p-value for the difference in scores is 5.6×10^{-12} . P2RY14=purinergic receptor P2Y14, GBP1 = guanylate binding protein 1, IFITM3 = interferon-induced transmembrane protein 3, ID3 = inhibitor of DNA binding 3.

Figure 4.

Differences in transcript counts in peripheral blood for three representative genes found to be related to sputum *M. tuberculosis* load as determined by GeneXpert testing. No TB, low *M. tb*, and high *M. tb* groups were compared using ANOVA with adjustment for multiple comparisons (Figure 4A). P-values for the differences in groups for Suppressor of cytokine signaling 3 (SOCS3) ($p=2.5 \times 10^{-11}$), complement C3b/C4b receptor 1 (CR1) ($p=1.99 \times 10^{-11}$), and oncostatin M (OSM) ($p=2.41 \times 10^{-11}$) were statistically significant. There were no significant differences in low sputum mycobacterial load vs. high sputum mycobacterial load for the Maertzdorf *et al*⁶ 4-gene signature score (Figure 4B, $p=0.1087$) or the Sweeney *et al*⁷ 3-gene signature score (Figure 4C, $p=0.1662$).

Figure 5.

Boxplots showing the post-treatment change in gene expression for the Berry (light grey) and the bacterial load 20-gene signature (dark grey) at each time point after starting TB treatment. There was a significant difference in the expression change from baseline at the 2 month, 6 month, and 12 month time points. The box is the interquartile range, with the horizontal line being the median. Whiskers extend to 1.5 times the interquartile range or the maximum/minimum, whichever is closer to the median.

Table 1.

Demographics of study participants.

	Pulmonary TB Cases (n=51)		Community Control s (n=21)
	Low <i>M. tubercu- losis</i>[*] (n=19)	High <i>M. tubercu- losis</i>⁺ (n=32)	
Mean age (years)	36.3	32.2	34.4
Male sex	14/19 (73%)	22/32 (69%)	10/21 (47.6%)
Quadrants of chest x-ray involved (mean)	1.6	1.9	—
Cavitary lesion on chest x-ray	0/19 (0%)	7/32 (22%)	—
IGRA[‡] Positive	—	—	15/21 (71.4%)

^{*} Low *M. tuberculosis* is negative, very low, or low on sputum GeneXpert testing.

⁺ High *M. tuberculosis* is medium or high on sputum GeneXpert testing.

[‡] IGRA is Interferon Gamma Release Assay using Quantiferon (Qiagen) test for *M. tuberculosis*.

Table 2.Differentially expressed genes in pulmonary TB comparing low vs. high sputum *M. tuberculosis* loads.

Gene Name	Gene ID	Fold-change* (log ₂)	P _{adj} ⁺
Annexin A3	ANXA3	0.508	0.022
Calcium channel, voltage-dependent, R type, alpha 1E subunit	CACNA1E	0.517	0.022
Complement C3b/C4b receptor 1	CR1	0.443	0.022
cAMP responsive element binding protein 5	CREB5	0.488	0.022
Dysferlin	DYSF	0.394	0.043
Ectonucleoside triphosphate diphosphohydrolase 1	ENTPD1	0.266	0.035
Glycerol kinase	GK	0.444	0.022
Glycogenin 1	GYG1	0.394	0.036
LIM domain kinase 2	LIMK2	0.413	0.022
Mitogen-activated protein kinase 14	MAPK14	0.404	0.035
NLR family apoptosis inhibitory protein	NAIP	0.476	0.035
Oncostatin M	OSM	0.539	0.017
6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase 3	PFKFB3	0.479	0.022
Peptidoglycan recognition protein 1	PGLYRP1	0.512	0.022
Putative homeodomain transcription factor 1	PHTF1	0.401	0.022
RAB20, member RAS oncogene family	RAB20	0.444	0.030
Signal induced proliferation associated 1 like 2	SIPA1L2	0.597	0.009
Suppressor of cytokine signaling 3	SOCS3	0.576	0.017
Serine/arginine-rich domain protein kinase 1	SRPK1	0.370	0.045
WD repeat and FYVE domain containing 3	WDFY3	0.431	0.022

* The fold-change is for high sputum *M. tuberculosis* load compared to low load as baseline.

⁺ The p_{adj} is the p-value adjusted for multiple comparisons using a Benjamini-Hochberg correction.