

RESEARCH ARTICLE

The Transcriptional Signature of Active Tuberculosis Reflects Symptom Status in Extra-Pulmonary and Pulmonary Tuberculosis

Simon Blankley¹, Christine M. Graham¹, Jacob Turner², Matthew P. R. Berry^{1,3}, Chloe I. Bloom¹, Zhaohui Xu⁴, Virginia Pascual⁴, Jacques Banchereau⁵, Damien Chaussabel⁶, Ronan Breen⁷, George Santis^{7,8}, Derek M. Blankenship², Marc Lipman^{9,10}, Anne O'Garra^{1,11*}



OPEN ACCESS

Citation: Blankley S, Graham CM, Turner J, Berry MPR, Bloom CI, Xu Z, et al. (2016) The Transcriptional Signature of Active Tuberculosis Reflects Symptom Status in Extra-Pulmonary and Pulmonary Tuberculosis. PLoS ONE 11(10): e0162220. doi:10.1371/journal.pone.0162220

Editor: Olivier Neyrolles, Institut de Pharmacologie et de Biologie Structurale, FRANCE

Received: June 15, 2016

Accepted: August 18, 2016

Published: October 5, 2016

Copyright: © 2016 Blankley et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data are available from GEO GSE83456.

Funding: AOG and this work is supported by the Francis Crick Institute, which receives its core funding from Cancer Research UK (FC001126), the U.K. Medical Research Council (FC001126) and the Wellcome Trust (FC001126). SB was jointly funded by the UK Medical Research Council (MRC) and the UK Department for International Development (DFID) under the MRC/DFID Concordat agreement (MR/J010723/1). GS was supported in part by the Department of Health via the National Institutes of

1 Laboratory of Immunoregulation and Infection, The Francis Crick Institute, Mill Hill Laboratory, London, United Kingdom, **2** Baylor Research Institute, Baylor Scott & White, Dallas, Texas, United States of America, **3** Department of Respiratory Medicine, Imperial College Healthcare NHS Trust, St Mary's Hospital, London, United Kingdom, **4** Baylor Institute for Immunology Research/ANRS Center for Human Vaccines, INSERM, Dallas, Texas, United States of America, **5** The Jackson Laboratory for Genomic Medicine, 10 Discovery Drive, Farmington, 06032, Connecticut, United States of America, **6** Sidra Medical and Research Center, Doha, Qatar, **7** Department of Respiratory Medicine, King's College London, London, United Kingdom, **8** Division of Asthma, Allergy and Lung Biology, King's College London, London, United Kingdom, **9** Department of Respiratory Medicine, Royal Free London NHS Foundation Trust, London, United Kingdom, **10** Division of Medicine, University College London, London, United Kingdom, **11** Department of Medicine, NHLI, Imperial College, London, United Kingdom

* anne.ogarra@crick.ac.uk

Abstract

Background

Mycobacterium tuberculosis infection is a leading cause of infectious death worldwide. Gene-expression microarray studies profiling the blood transcriptional response of tuberculosis (TB) patients have been undertaken in order to better understand the host immune response as well as to identify potential biomarkers of disease. To date most of these studies have focused on pulmonary TB patients with gene-expression profiles of extra-pulmonary TB patients yet to be compared to those of patients with pulmonary TB or sarcoidosis.

Methods

A novel cohort of patients with extra-pulmonary TB and sarcoidosis was recruited and the transcriptional response of these patients compared to those with pulmonary TB using a variety of transcriptomic approaches including testing a previously defined 380 gene meta-signature of active TB.

Results

The 380 meta-signature broadly differentiated active TB from healthy controls in this new dataset consisting of pulmonary and extra-pulmonary TB. The top 15 genes from this meta-signature had a lower sensitivity for differentiating extra-pulmonary TB from healthy controls as compared to pulmonary TB. We found the blood transcriptional responses in

Health Research comprehensive Biomedical Research Centre award to Guy's and St Thomas' National Health Service Foundation Trust in partnership with King's College London. This study received financial support from SILA. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: Patents have been filed against distinguishing active TB from latent TB and healthy controls, WO2009/158521, WO2011/066008; and active TB from sarcoidosis and other lung diseases (61/736,908t) (inventors, AOG, MPRB, CIB, JB, DC, VP). We confirm that this does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

pulmonary and extra-pulmonary TB to be heterogeneous and to reflect the extent of symptoms of disease.

Conclusions

The transcriptional signature in extra-pulmonary TB demonstrated heterogeneity of gene expression reflective of symptom status, while the signature of pulmonary TB was distinct, based on a higher proportion of symptomatic individuals. These findings are of importance for the rational design and implementation of mRNA based TB diagnostics.

Introduction

Mycobacterium tuberculosis (*Mtb*) continues to be a significant cause of mortality and morbidity worldwide [1]. In 2013 there were 9 million cases of active tuberculosis (TB) resulting in 1.5 million deaths worldwide, of these active TB cases approximately 15% were notified as extra-pulmonary TB [1]. Common sites of extra-pulmonary TB infection include the pleura, lymph nodes, meninges, bone and genito-urinary tract [2]. Extra-pulmonary TB can be challenging to diagnose compared to pulmonary TB as it is often not considered during clinical evaluation and diagnostic samples can be more difficult to obtain [3, 4].

Whole genome expression analysis of human whole blood as well as peripheral blood mononuclear cells (PBMC) has been widely used to profile the host transcriptional response in active TB and identify potential biomarkers for use in diagnostics (reviewed in [5]). An interferon-dominated 393 transcript signature was identified from human whole blood which was present in active TB and absent in the majority of healthy and latently infected individuals and was shown to correlate with the extent of radiographic lung disease [6]. This finding of enrichment for interferon signalling has since been reported in multiple subsequent studies [7–11]. Other studies have, in addition to interferon signalling, identified immunological pathways that may be relevant to the pathogenesis of active TB such as TLR signalling, complement as well as enrichment for T- and B-cell function gene expression [6–9, 11–15]. Two large studies have included extra-pulmonary TB patients within their active TB cohort, but no sub-group analysis was undertaken to determine if there was a transcriptional signature unique to extra-pulmonary TB [16, 17]. Therefore the question of whether there is a difference in the blood transcriptional signature between pulmonary and extra-pulmonary TB has not been answered.

We recently published a 380 gene meta-signature (S1 Table) of active TB compared to healthy controls [18], which represents an attempt to identify the most consistently differentially expressed genes across the published publicly available datasets [18]. Similar to the individual published studies the 380 gene meta-signature was found to be enriched for immune response pathways including multiple pattern recognition receptors, cytokines, the inflammatory, complement and immunoglobulin pathways [18].

We herein used this published derived gene set as well as other transcriptional analysis tools to test similarities and differences between pulmonary TB and extra-pulmonary TB in a novel dataset and now report heterogeneity driven mainly by the extent of symptoms.

Materials and Methods

Ethics

This study was approved by the Central London 3 Research Ethics Committee (09/H0716/41). All participants gave written informed consent.

Clinical cohort

Recruitment took place between 2011 and 2014. Sites for recruitment were the Royal Free and Guy's & St Thomas' Hospitals, London, United Kingdom. Healthy controls and the majority of pulmonary TB patients were chosen from an existing bank of globin reduced RNA samples. These subjects had consented to their samples being used in further approved studies. Patients identified from this bank were selected on cohort group, age, sex and ethnicity. Diagnosis of active TB was based on the following: positive mycobacterial culture result from the site of disease; or caseating granuloma on biopsy and/or clinical/radiological features consistent with active TB and a good response to therapy. Sarcoidosis patients all had mediastinal disease with compatible histology and clinical/radiological features to support a diagnosis of sarcoidosis and negative mycobacterial cultures. Patients were excluded if they were immunosuppressed either through disease (such as HIV, diabetes or autoimmune diseases), medication or had significant co-morbidities affecting the pulmonary system.

Microarray

Human whole blood RNA was isolated, globin reduced and amplified as described previously [6, 19]. 750 ng of cRNA was hybridized to Illumina Human HT-12 V4 BeadChip arrays and scanned on Illumina iScan. GenomeStudio was used to perform quality control and generate signal intensity. Microarray analysis was undertaken using GeneSpring GX 13.0. Per sample normalisation (75th centile) and per transcript normalisation (median of all samples) was performed. Microarray data were deposited in the NCBI Gene Expression Omnibus (GEO) with series accession number (GSE83456). All data collected and analysed in these experiments adhere to the Minimal Information About a Microarray Experiment (MIAME) guidelines.

Expression analysis

Molecular Distance to Health (MDTH), was calculated using methodology previously described [20], for defined groups of genes (group of genes defined in legend of Figures) relative to a control group (detailed in legend of Figure). Additionally, the MDTH was calculated using the same methodology but now using 3409 transcripts, which represent the transcripts from 38 annotated modules [21]. Z-scores were derived from the respective MDTH's and calculated relative to the control group.

Modular analysis was undertaken using log₂ transformed normalised data [21].

Differentially expressed genes identified from new dataset for TB and sarcoidosis involved groups being compared to the healthy control cohort (transcripts filtered which were not significantly detected from background in at least 10% of samples, low expressed transcripts filtered (less than 2 fold change from median in 10% of samples), followed by statistical testing (independent t-test with Benjamini Hochberg multiple testing correction (*q-value* <0.05) between groups of interest). Transcripts were then matched to Entrez Gene identifiers; duplicates (retaining those with the largest fold change difference) and non-matched transcripts were filtered.

Statistical analyses

GraphPad Prism 6 or Microsoft Excel (2010) were used for statistical analysis, details of statistical testing given in figure legends.

Results

Testing the meta-signature in a new dataset reveals differences in the transcriptional response of pulmonary and extra-pulmonary patients

A new cohort of extra-pulmonary and sarcoidosis patients was recruited together with an existing bank of pulmonary TB and healthy controls formed the dataset for analysis (Cohort details; S2 and S3 Tables). There was no difference in gender frequency (Fig 1A) between groups however there were differences in the group composition with regard to ethnicity and age, with the sarcoidosis group being significantly older than the other groups and tending to have less patients of Indian subcontinent background (Fig 1B and 1C). Total white cell count was

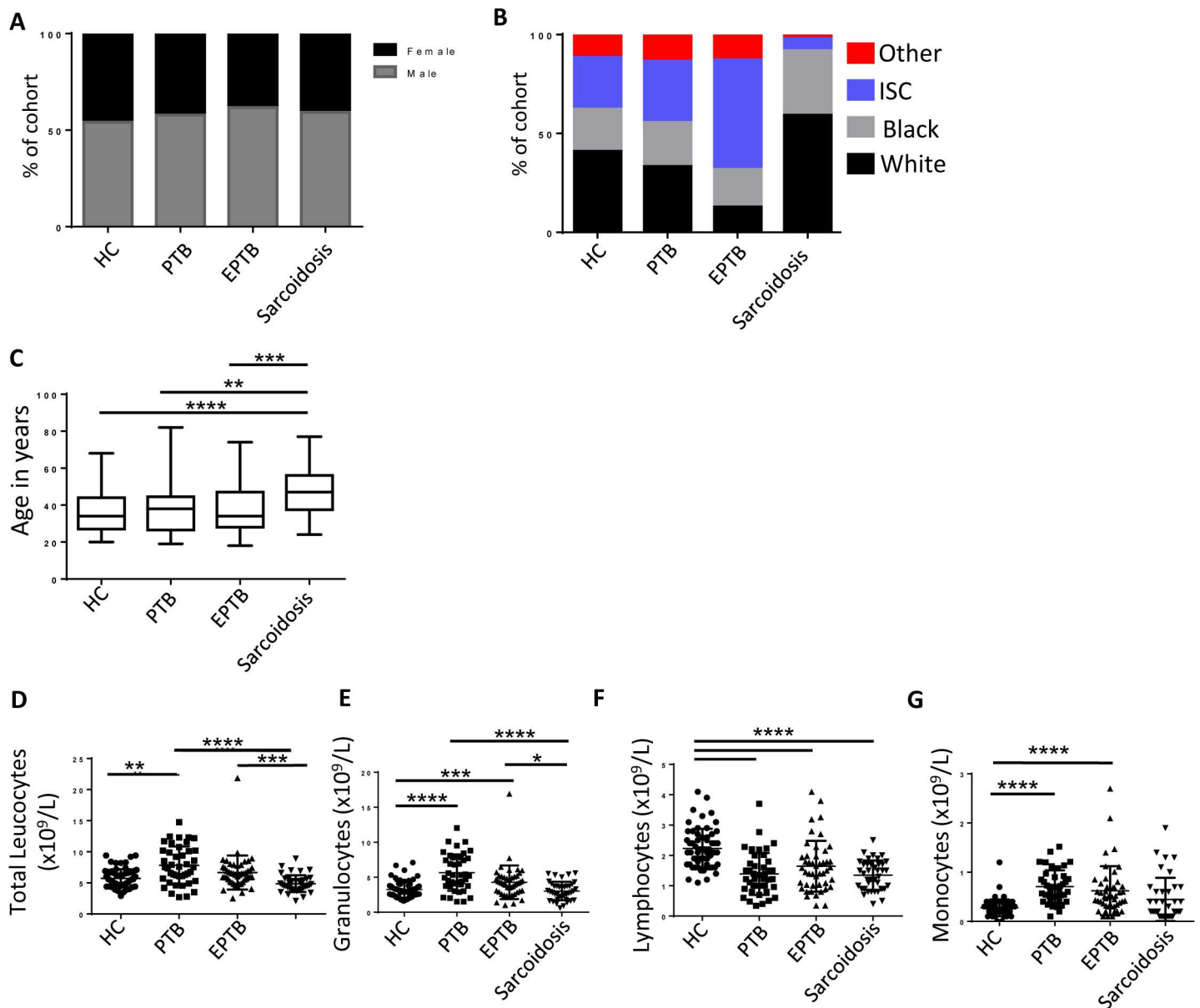


Fig 1. Clinical parameters of patients included in microarray dataset. (A) percentage of cohort by sex (B) percentage of cohort by ethnicity (ISC; Indian sub-continent) and (C) age (mean; min-max bars). Whole blood composition: (D) total leukocytes (E) Granulocytes (F) lymphocytes and (G) monocytes. Statistical tests: Kruskal Wallis with Dunn's multiple testing correction.

doi:10.1371/journal.pone.0162220.g001

significantly elevated in Pulmonary TB patients compared to the other groups (Fig 1D), this was mainly due to increased numbers of granulocytes compared to the other groups (Fig 1E). Total lymphocyte count was significantly higher in healthy controls compared to all the groups (Fig 1F) and monocytes were significantly elevated in Pulmonary and Extra-pulmonary TB patients compared to Healthy controls (Fig 1G).

The previously defined 380 meta-signature genes [18] (S1 Table, mapped to 687 Illumina transcripts, of which 113 transcripts were excluded as they were not significantly detected from background in 10% of samples) were used for analysis in this new dataset. Hierarchical clustering of the dataset (healthy controls, pulmonary and extra-pulmonary TB patients only) revealed that the 380 meta-signature genes (S1 Table) were able to separate healthy controls from the majority of both pulmonary and extra-pulmonary TB patients, although a small number of TB patients clustered together with the healthy controls (Fig 2A). Using the most consistently identified genes from the meta-analysis as potential biomarkers (15 genes which were identified in 15 or more meta-analysis data-sets, S1 Table) and calculating a molecular score for these genes, there was a greater sensitivity in identifying pulmonary TB patients than extra-pulmonary TB patients from healthy controls (Fig 2B).

Transcriptional signatures in TB reflect symptom status of individuals

There was a significant difference in the molecular distance to health (MDTH) of both pulmonary TB and extra-pulmonary TB patients as compared to healthy controls (Fig 3A).

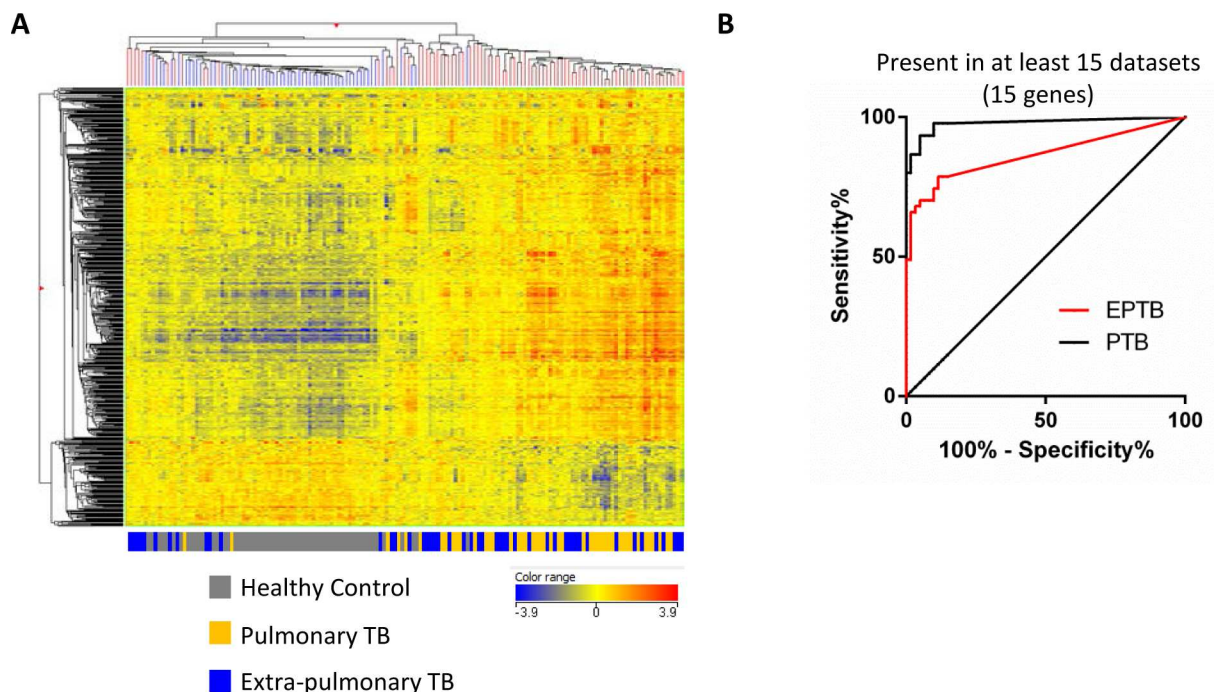


Fig 2. Testing the meta-signature and the most consistently identified genes in a new dataset. (A) The 380 gene meta-signature were mapped to 687 Illumina transcripts, of which 113 transcripts were excluded as they were not significantly detected from background in 10% of samples. Hierarchical clustering (Pearson uncentred (cosine) with averaged linkage) on individuals and transcripts broadly cluster healthy controls from pulmonary and extra-pulmonary TB patients. (B) Receiver operator curves for extra-pulmonary and pulmonary cohorts against healthy controls using MDTH derived from the most consistently identified genes (15 genes identified in at least 15 of the meta-analysis datasets [18])—representing the most robustly identified genes in that analysis, healthy controls used as control group for MDTH) as potential diagnostic biomarker. Both prediction results were validated using k-fold cross validation with k equal to 10 with 1,000 iterations. The mean AUC for the EPTB and PTB validation results are 0.865 (95% confidence interval: 0.857–0.872) and 0.977 (95% confidence interval: 0.974–0.981) respectively.

doi:10.1371/journal.pone.0162220.g002

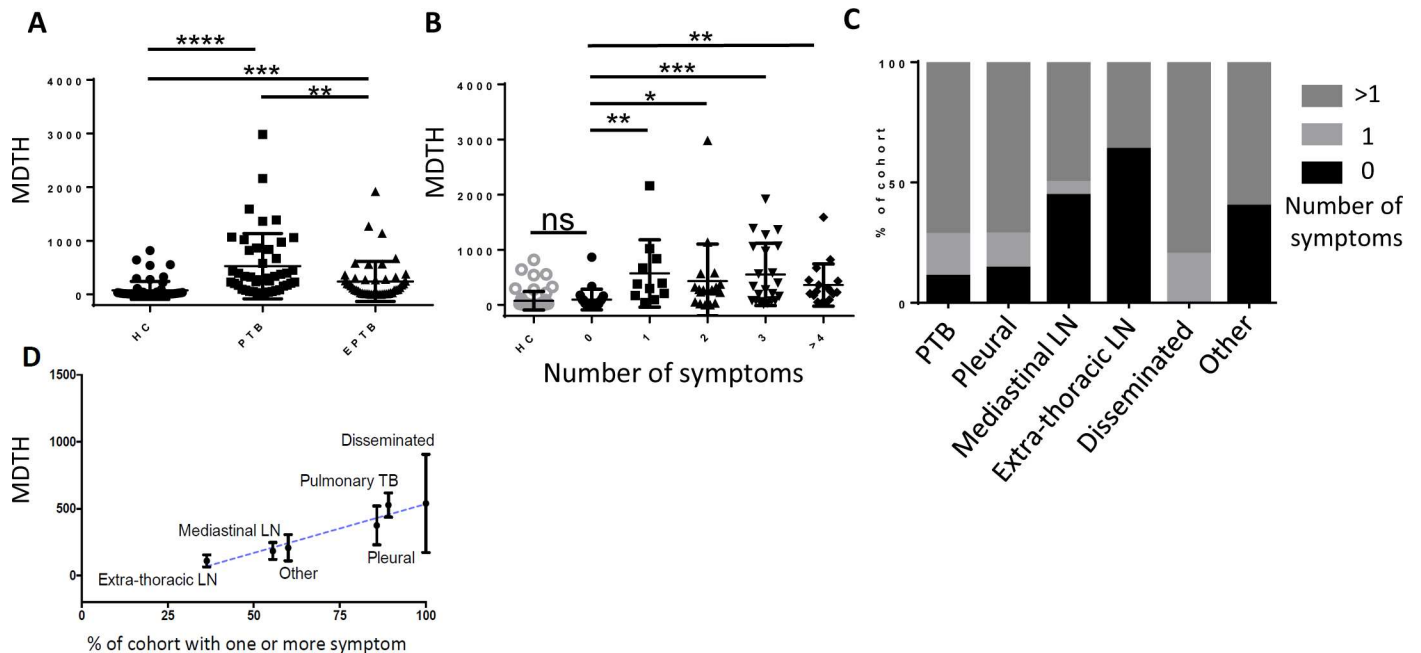


Fig 3. Molecular distance to health is linked to symptoms status and site of disease. (A) Molecular distance to health (MDTH) calculated for each individual (from 3409 transcripts which represent the transcripts of the 38 annotated modules shown in Fig 2A, healthy controls used as the control group). Individuals grouped by disease status, (median value, error bars SD) statistical testing Kruskal Wallis with Dunn's multiple testing correction. (B) Patients grouped by number of reported symptoms; night sweats, fever, weight loss, chest pain or cough (median value, error bars SD) statistical testing Kruskal Wallis with Dunn's multiple testing correction. (C) Patients grouped by site of disease and number of reported symptoms (as previous). (D) Mean MDTH (error bars, Standard error of the mean) plotted for each site of disease against % of the cohort suffering from one or more symptom (from list of symptoms described previously). Blue line represents Pearson's correlation ($R^2 = 0.95, p = 0.0090$).

doi:10.1371/journal.pone.0162220.g003

Pulmonary TB had a significantly higher MDTH compared to extra-pulmonary TB (Fig 3A). The MTDH was not influenced by the ethnicity of the TB patients (S1A Fig) or by *M. tuberculosis* culture status in extra-pulmonary TB patients (S1B Fig).

Presence of one of the following symptoms: night sweats, fever, weight loss, chest pain or cough resulted in a significantly higher MDTH compared to healthy controls; there was no additive effect of increasing number of symptoms in terms of magnitude of MDTH (Fig 3B). Absence of any of the above symptoms resulted in no significant difference in MDTH score from healthy controls. There were differences in symptom prevalence depending on the site of disease (Fig 3C). The mean MDTH of the patients grouped by site of disease significantly correlated with the percentage of patients within the group having one or more symptom (Fig 3D). There was no difference in MDTH score of extra-pulmonary patients dependent on culture status (S1C Fig).

Upon individual modular analysis, there was no apparent site-specific modular pattern. Instead, presence of any of the five symptoms seemed to influence the modular pattern observed (Fig 4A). Molecular scoring of the genes within modules annotated as "inflammation" revealed no significant difference between healthy controls and the asymptomatic group of TB patients, in contrast to the significant difference between asymptomatic and symptomatic populations (Fig 4A and 4B). For the interferon modules there was a significantly higher score for the symptomatic group compared to the asymptomatic group (Fig 4A and 4B). While there were significantly higher lymphocyte and significantly lower granulocyte and monocyte counts in the asymptomatic group compared to the symptomatic TB patients, there was considerable overlap between the two populations (Fig 4C).

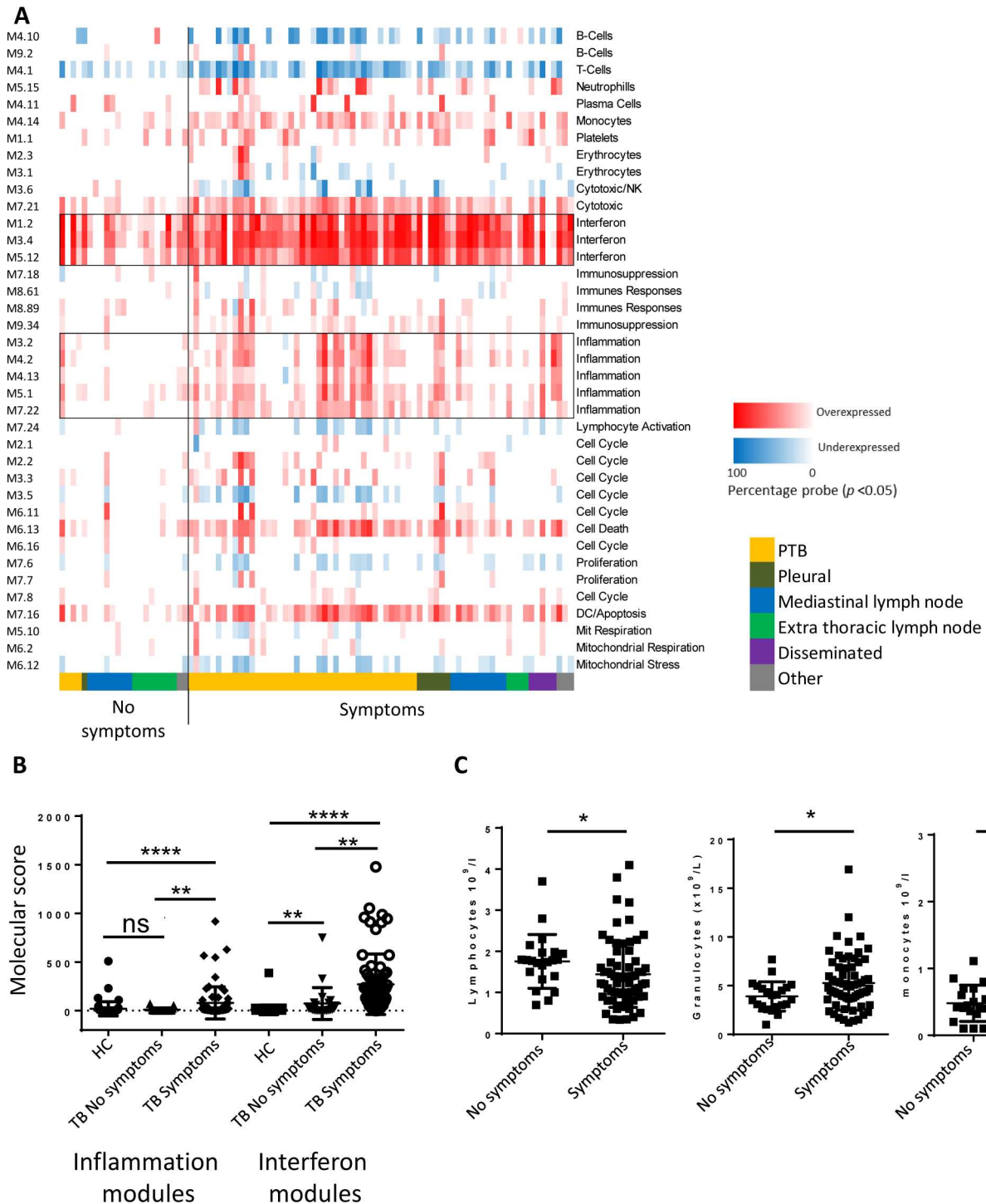


Fig 4. Symptom status linked to alteration in modular profile and blood counts. (A) Individual modular analysis for all TB patients grouped by symptom status and site of disease, boxed are modules annotated as interferon or inflammation. 38 annotated modules are displayed as a heatmap with red indicates significant over-abundance of transcripts and a blue indicating significant under-abundance ($p < 0.05$). The colour intensity represents the percentage of genes in that module that are significantly differentially expressed. (B) MDTH calculated for all interferon module transcripts and all inflammatory module transcripts (healthy controls used as control group). TB patients grouped according to symptom status (median value, error bars SD) statistical testing Kruskal Wallis with Dunn's multiple testing correction. (C) TB patients grouped according to symptom status with lymphocyte, granulocyte and monocytes counts shown (median value, error bars SD) statistical testing Mann-Whitney U.

doi:10.1371/journal.pone.0162220.g004

Overlap in expression profiles between sarcoidosis and extra-pulmonary TB patients

Analysis of all the TB patients from this new dataset identified 927 genes (S4 Table) as being differentially regulated compared to healthy controls and 883 genes (S4 Table) differentially regulated between sarcoidosis patients and healthy controls (Fig 5A). Overlapping these two gene lists revealed that 709 genes were commonly expressed between the two conditions compared to healthy controls, of which 15 had been previously identified from the meta-analysis as those most consistently expressed in TB (Fig 5A, S4 Table). Nine of the top 10 upregulated genes by fold change of TB compared with healthy controls were also found within the top 10 upregulated genes by fold change of sarcoidosis compared with healthy controls (Table 1).

Of those genes which were identified in only TB or sarcoidosis compared with healthy controls, only one (*RNF182*) was two-fold different between TB and sarcoidosis (S4 Table). Using the Bloom *et al* 144 transcript list [19] by hierarchical clustering it was possible again to separate the majority of sarcoidosis patients from those with pulmonary TB (Fig 5B), but not from mediastinal TB (Fig 5C).

Calculating the molecular scores for the 380 meta-signature for all patients and ranking them in order of z score (compared to healthy controls) revealed that there was an overlap between the three groups, with pulmonary TB patients tending towards the higher end and extra-pulmonary and sarcoidosis patients the lower end of the spectrum (Fig 5D). Indeed, for all three outcomes the mean of pulmonary TB patients was significantly higher than both the extra-pulmonary TB and sarcoidosis patients. However, any individual regardless of disease type could lie anywhere within this range. This observation was consistent for MDTH and molecular scores for the interferon modules. Patients with none of the 5 listed symptoms, regardless of their disease, tended to have the lowest z scores (Fig 5D).

Discussion

This study is the first to our knowledge to attempt to identify blood transcriptional signatures associated with the site of *Mtb* infection. We show that the transcriptional response is similar across sites of disease as measured from the blood although the magnitude of response varies and this is mainly associated with the presence or absence of symptoms and probably the site of the infection.

There were differences in the demographic composition of the cohorts in this study. Patients diagnosed with sarcoidosis were older and more likely to be of white or black ethnic background. This skewing in terms of age and ethnicity is well described in sarcoidosis [22, 23]. TB in the UK tends to be diagnosed from the immigrant population, in particular those from the Indian subcontinent [24], with extra-pulmonary TB being more common in those of Asian and African origin [25–27]. Hence the differences in ethnic composition between our clinical groups are likely to be due to a combination of these factors. These differences in group composition did not affect the transcriptional responses observed in our study. Similarly, whole blood cell composition in this study at the group level was found to be altered in pulmonary and extra-pulmonary TB compared to healthy controls with increased monocytes and granulocytes and decreased lymphocytes as has been previously described [28].

Extra-pulmonary TB represents *Mtb* infection which has spread haematogenously or via the lymphatic system from the lung. It is hypothesised that this spread happens during initial infection and that extra-pulmonary disease may represent reactivation rather than primary disease [4]. It has long been recognised that there is variation in symptoms and bacterial load dependent on site of disease, hence differences in the transcriptional signature may reflect variation in bacterial burden at the site of disease [4, 29–31], or in the site specific host immune

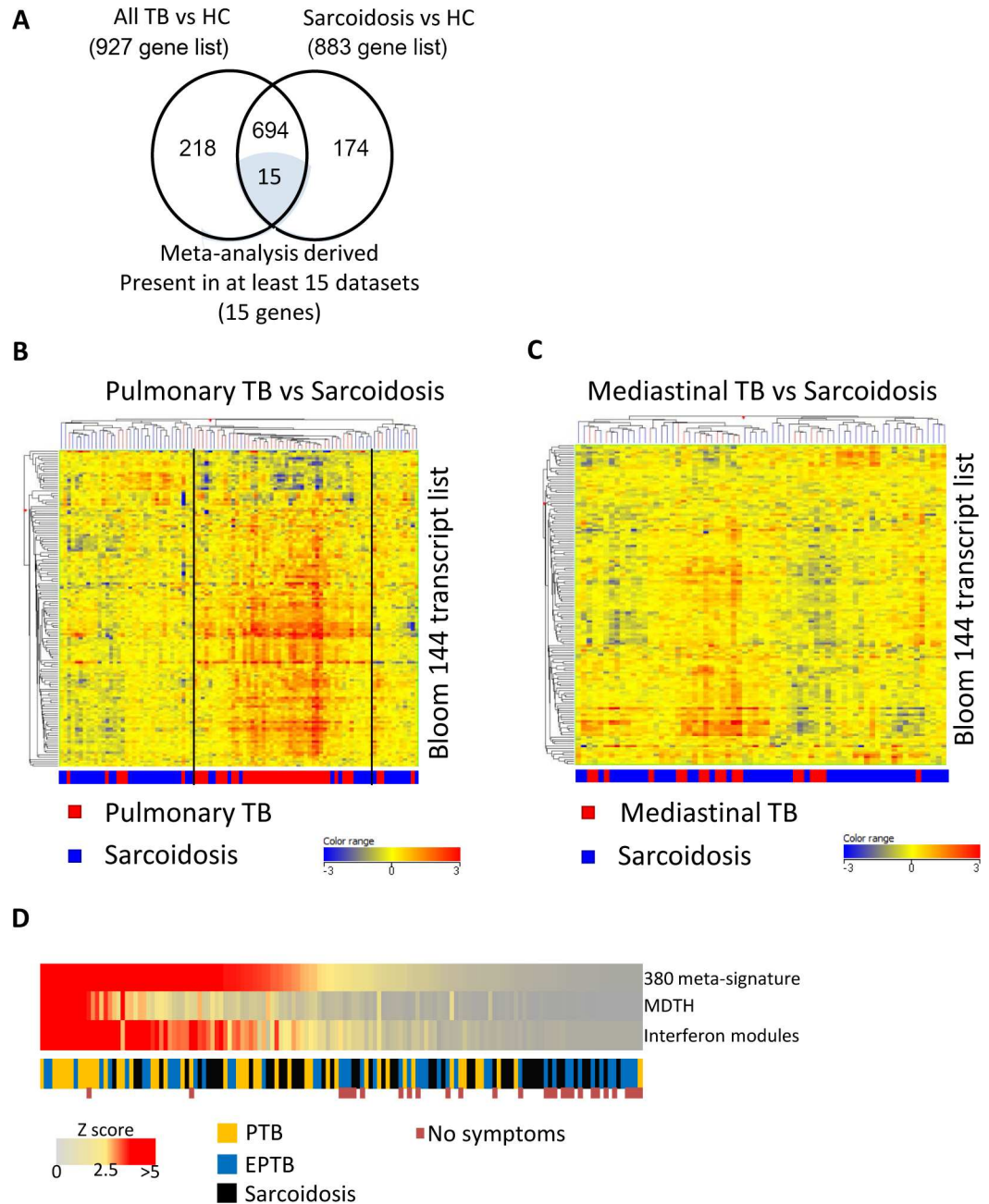


Fig 5. Tuberculosis and sarcoidosis have similar differentially regulated genes. (A) Differentially expressed genes identified from new dataset for TB and sarcoidosis groups compared to healthy controls (transcripts filtered which were not significantly detected from background in at least 10% of samples, low expressed transcripts filtered (less than 2 fold change from median in 10% of samples), followed by statistical testing (independent t-test with Benjamini Hochberg multiple testing correction (q -value < 0.05) between groups of interest). Transcripts were matched to Entrez gene identifiers and duplicates and non-matched transcripts filtered. Venn diagram showing the overlap of DEGs between these two diseases and the most significant 15 gene list from the meta-analysis. (B) Heatmap of normalised expression using the Bloom et al 144 transcript list are able to broadly differentiate pulmonary TB patients from sarcoidosis patients. Clustering (Pearson's uncentred (Cosine) with averaged linkage) on transcripts (rows) and individual patient blood samples (columns). (C) The same 144 transcript list (16) is unable to differentiate between mediastinal TB patients and sarcoidosis patients by clustering (as before) (D) Molecular scores calculated for transcripts from 380 gene meta-signature, MDTH and transcripts representing the interferon modules (healthy controls acting as control group). Z scores calculated (with healthy controls used as reference group for calculating mean and SD) and then TB patients and sarcoidosis patients ranked according to Z score of the 380 gene meta-signature. For all three outcomes the

mean of pulmonary TB patients was significantly higher than both the extra-pulmonary TB and sarcoidosis patients ($p < 0.05$). Differences among means were tested with a generalized linear model assuming a normal distribution and a Bonferroni multiple testing correction. No symptoms indicates absence of any of the five symptoms listed previously.

doi:10.1371/journal.pone.0162220.g005

response which may be reflected in the blood. Gene expression signatures of TB patients have previously been linked with bacterial burden, with a prior study showing reduced sensitivity of a whole genome expression derived disease risk score with decreased culture positivity in children [17]. Equally, with respect to complement which has been identified as one of the key findings in several microarray studies, *C1q* expression levels correlated with sputum smear positivity and diminished with treatment [12, 32]. Our findings of differences in MDTH (and symptoms) based on the site of disease may therefore be linked to differences in bacterial burden or the host site specific immune response. Integration of both host transcriptional response and more accurate quantitation of total mycobacterial burden may result in a better understanding of what drives the blood host transcriptional response.

Whilst it was possible to again here to distinguish pulmonary TB from sarcoidosis using our previously described gene set [19], it was not possible to distinguish sarcoidosis and mediastinal TB by hierarchical clustering. The inability to distinguish sarcoidosis and mediastinal TB, is possibly due to the lower symptom status of mediastinal extra-pulmonary TB cohorts versus pulmonary TB cohorts. This would fit with our findings that the magnitude of the transcriptional response as measured from the blood is mainly associated with the presence or absence of symptoms. Previous studies which had identified differentially expressed genes between sarcoidosis and TB had used only pulmonary TB patients as their comparator group [9, 19, 33]. We show that compared to healthy controls the gene expression signature of TB (including pulmonary and extra-pulmonary patients) is very similar to sarcoidosis. We show that individuals from each disease state (Pulmonary TB, Extra-pulmonary TB or sarcoidosis) can lie anywhere along this spectrum of gene expression including some clustering together with healthy controls. However when taken at the group level differences in magnitude of expression can be observed between pulmonary TB and the other disease states.

Both TB and sarcoidosis are granulomatous diseases, and together with another granulomatous disease melioidosis they have been shown to have similar gene expression signatures to TB [9, 19, 33, 34]. Despite these similarities, it may still be possible using clinical parameters

Table 1. Top 10 differentially upregulated genes in TB and sarcoidosis^a.

Tuberculosis		Sarcoidosis	
Gene	Log2 FC	Gene	Log2 FC
<i>SERPING1</i>	3.83	<i>ANKRD22</i>	3.35
<i>ANKRD22</i>	3.82	<i>SERPING1</i>	3.20
<i>BATF2</i>	3.61	<i>BATF2</i>	3.02
<i>FCGR1C</i>	3.04	<i>IFIT3</i>	2.67
<i>LOC728744</i>	3.04	<i>FCGR1A</i>	2.54
<i>GBP6</i>	3.03	<i>FCGR1C</i>	2.52
<i>C1QB</i>	3.02	<i>GBP6</i>	2.52
<i>FCGR1A</i>	3.02	<i>FCGR1B</i>	2.50
<i>FCGR1B</i>	2.90	<i>RAP1GAP</i>	2.44
<i>IFIT3</i>	2.84	<i>LOC728744</i>	2.42

^a Significantly upregulated genes as measured by fold change compared to healthy controls.

doi:10.1371/journal.pone.0162220.t001

and risk stratification to utilise mRNA expression based diagnostics to differentiate between diseases as has been done for active and latent TB [6–9, 15–18, 35].

We show here for the first time that blood based transcriptional signatures in pulmonary and extra-pulmonary TB differ as a result of the symptom status and site of the disease in each individual. These findings have implications for design and implementation of mRNA expression tools to support diagnostics and treatment monitoring of TB.

Supporting Information

S1 Fig. Molecular Distance to health. Molecular distance to health (MDTH) calculated for each individual (from 3409 transcripts which represent the transcripts of the 38 annotated modules shown in Fig 2A (Control group for molecular scores–healthy controls). (A) MDTH for TB patients by ethnicity (no significant difference between groups) (ISC; Indian sub-continent). (B) MDTH for extra-pulmonary patients—there was no significant difference in MDTH status (Mann-whitney U) dependent on culture status.

(TIFF)

S1 Table. Identity of 380 genes expressed in blood of active TB patients from metaanalysis distinguish active TB patients from healthy Controls. List of genes identified in earlier study [18] now by hierarchical clustering of the dataset (healthy controls, pulmonary and extra-pulmonary TB patients only) revealed that this 380 meta-signature genes separate healthy controls from the majority of both pulmonary and extra-pulmonary TB patients.

(XLSX)

S2 Table. Numbers of patient within the cohorts of pulmonary TB, extra-pulmonary TB and sarcoidosis patients and healthy controls.

(DOCX)

S3 Table. Numbers of patients with different forms of TB in different body sites, their *Mtb* culture status, histology and clinical diagnosis status.

(DOCX)

S4 Table. Lists of 694 commonly expressed genes in the blood of TB and sarcoid patients as compared to healthy controls; unique TB genes and unique sarcoid genes expressed in blood as compared to healthy controls.

(XLS)

Acknowledgments

With thanks to all of the patients who volunteered to participate in this study. We thank Abdul Sesay and Deborah Jackson, Advanced Sequencing Facility at The Francis Crick Institute, for help with microarray samples. We thank Fotini Rozakeas (MRC-NIMR) for help with recruitment of volunteers. We thank Lúcia Moreira-Teixeira for critical review and Krzysztof Potempa for analysis of data and critical review of the manuscript and Leona Gabrýsova for graphics help.

Author Contributions

Conceptualization: AOG SB MPRB CIB VP JB DC ML.

Data curation: SB.

Formal analysis: SB JT ZX DMB.

Funding acquisition: AOG SB.

Investigation: SB.

Methodology: SB CMG MPRB CIB.

Project administration: SB AOG.

Resources: SB RB GS DMB ML.

Supervision: AOG ML.

Validation: SB JT DMB.

Visualization: SB AOG.

Writing – original draft: SB AOG.

Writing – review & editing: SB AOG CMG MPRB CIB JB.

References

1. World Health Organization. World Health Organisation: Global tuberculosis report 2014. 2014.
2. Peto HM, Pratt RH, Harrington TA, LoBue PA, Armstrong LR. Epidemiology of extrapulmonary tuberculosis in the United States, 1993–2006. *Clin Infect Dis*. 2009; 49(9):1350–7. Epub 2009/10/02. doi: [10.1086/605559](https://doi.org/10.1086/605559) PMID: [19793000](https://pubmed.ncbi.nlm.nih.gov/19793000/).
3. Solovic I, Jonsson J, Korzeniewska-Kosela M, Chiotan DI, Pace-Asciak A, Slump E, et al. Challenges in diagnosing extrapulmonary tuberculosis in the European Union, 2011. *Euro surveillance: bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin*. 2013; 18(12). PMID: [23557946](https://pubmed.ncbi.nlm.nih.gov/23557946/).
4. American Thoracic Society. Diagnostic Standards and Classification of Tuberculosis in Adults and Children. This official statement of the American Thoracic Society and the Centers for Disease Control and Prevention was adopted by the ATS Board of Directors, July 1999. This statement was endorsed by the Council of the Infectious Disease Society of America, September 1999. *American journal of respiratory and critical care medicine*. 2000; 161(4 Pt 1):1376–95. doi: [10.1164/ajrccm.161.4.16141](https://doi.org/10.1164/ajrccm.161.4.16141) PMID: [10764337](https://pubmed.ncbi.nlm.nih.gov/10764337/).
5. Blankley S, Berry MP, Graham CM, Bloom CI, Lipman M, O'Garra A. The application of transcriptional blood signatures to enhance our understanding of the host response to infection: the example of tuberculosis. *Philosophical transactions of the Royal Society of London Series B, Biological sciences*. 2014; 369(1645):20130427. doi: [10.1098/rstb.2013.0427](https://doi.org/10.1098/rstb.2013.0427) PMID: [24821914](https://pubmed.ncbi.nlm.nih.gov/24821914/).
6. 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. Epub 2010/08/21. doi: [10.1038/nature09247](https://doi.org/10.1038/nature09247) PMID: [20725040](https://pubmed.ncbi.nlm.nih.gov/20725040/); PubMed Central PMCID: [PMC3492754](https://pubmed.ncbi.nlm.nih.gov/PMC3492754/).
7. Ottenhoff TH, Dass RH, Yang N, Zhang MM, Wong HE, Sahiratmadja E, et al. Genome-wide expression profiling identifies type 1 interferon response pathways in active tuberculosis. *PloS one*. 2012; 7(9):e45839. doi: [10.1371/journal.pone.0045839](https://doi.org/10.1371/journal.pone.0045839) PMID: [23029268](https://pubmed.ncbi.nlm.nih.gov/23029268/); PubMed Central PMCID: [PMC3448682](https://pubmed.ncbi.nlm.nih.gov/PMC3448682/).
8. Maertzdorf J, Ota M, Reipsilber D, Mollenkopf HJ, Weiner J, Hill PC, et al. Functional correlations of pathogenesis-driven gene expression signatures in tuberculosis. *PloS one*. 2011; 6(10):e26938. Epub 2011/11/03. doi: [10.1371/journal.pone.0026938](https://doi.org/10.1371/journal.pone.0026938) PMID: [22046420](https://pubmed.ncbi.nlm.nih.gov/22046420/); PubMed Central PMCID: [PMC3203931](https://pubmed.ncbi.nlm.nih.gov/PMC3203931/).
9. Maertzdorf J, Weiner J 3rd, Mollenkopf HJ, Network TB, Bauer T, Prasse A, et al. Common patterns and disease-related signatures in tuberculosis and sarcoidosis. *Proceedings of the National Academy of Sciences of the United States of America*. 2012; 109(20):7853–8. Epub 2012/05/02. doi: [10.1073/pnas.1121072109](https://doi.org/10.1073/pnas.1121072109) PMID: [22547807](https://pubmed.ncbi.nlm.nih.gov/22547807/); PubMed Central PMCID: [PMC3356621](https://pubmed.ncbi.nlm.nih.gov/PMC3356621/).
10. Lu C, Wu J, Wang H, Wang S, Diao N, Wang F, et al. Novel biomarkers distinguishing active tuberculosis from latent infection identified by gene expression profile of peripheral blood mononuclear cells. *PloS one*. 2011; 6(8):e24290. Epub 2011/09/10. doi: [10.1371/journal.pone.0024290](https://doi.org/10.1371/journal.pone.0024290) PMID: [21904626](https://pubmed.ncbi.nlm.nih.gov/21904626/); PubMed Central PMCID: [PMC3164189](https://pubmed.ncbi.nlm.nih.gov/PMC3164189/).
11. Lesho E, Forestiero FJ, Hirata MH, Hirata RD, Cecon L, Melo FF, et al. Transcriptional responses of host peripheral blood cells to tuberculosis infection. *Tuberculosis*. 2011; 91(5):390–9. Epub 2011/08/13. doi: [10.1016/j.tube.2011.07.002](https://doi.org/10.1016/j.tube.2011.07.002) PMID: [21835698](https://pubmed.ncbi.nlm.nih.gov/21835698/).

12. Cliff JM, Lee JS, Constantinou N, Cho JE, Clark TG, Ronacher K, et al. Distinct phases of blood gene expression pattern through tuberculosis treatment reflect modulation of the humoral immune response. *The Journal of infectious diseases*. 2013; 207(1):18–29. Epub 2012/08/09. doi: [10.1093/infdis/jis499](https://doi.org/10.1093/infdis/jis499) PMID: [22872737](https://pubmed.ncbi.nlm.nih.gov/22872737/).
13. 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. Epub 2012/10/12. doi: [10.1371/journal.pone.0046191](https://doi.org/10.1371/journal.pone.0046191) PMID: [23056259](https://pubmed.ncbi.nlm.nih.gov/23056259/); PubMed Central PMCID: [PMC3462772](https://pubmed.ncbi.nlm.nih.gov/PMC3462772/).
14. Mistry R, Cliff JM, Clayton CL, Beyers N, Mohamed YS, Wilson Pa, et al. Gene-expression patterns in whole blood identify subjects at risk for recurrent tuberculosis. *The Journal of infectious diseases*. 2007; 195(3):357–65. doi: [10.1086/510397](https://doi.org/10.1086/510397) PMID: [17205474](https://pubmed.ncbi.nlm.nih.gov/17205474/)
15. Jacobsen M, Repsilber D, Gutschmidt A, Neher A, Feldmann K, Mollenkopf HJ, et al. Candidate biomarkers for discrimination between infection and disease caused by *Mycobacterium tuberculosis*. *Journal of molecular medicine*. 2007; 85(6):613–21. Epub 2007/02/24. doi: [10.1007/s00109-007-0157-6](https://doi.org/10.1007/s00109-007-0157-6) PMID: [17318616](https://pubmed.ncbi.nlm.nih.gov/17318616/).
16. 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 medicine*. 2013; 10(10):e1001538. Epub 2013/10/30. doi: [10.1371/journal.pmed.1001538](https://doi.org/10.1371/journal.pmed.1001538) PMID: [24167453](https://pubmed.ncbi.nlm.nih.gov/24167453/); PubMed Central PMCID: [PMC3805485](https://pubmed.ncbi.nlm.nih.gov/PMC3805485/).
17. Anderson ST, Kaforou M, Brent AJ, Wright VJ, Banwell CM, Chagaluka G, et al. Diagnosis of childhood tuberculosis and host RNA expression in Africa. *The New England journal of medicine*. 2014; 370(18):1712–23. doi: [10.1056/NEJMoa1303657](https://doi.org/10.1056/NEJMoa1303657) PMID: [24785206](https://pubmed.ncbi.nlm.nih.gov/24785206/).
18. Blankley S, Graham CM, Levin J, Turner J, Berry MP, Bloom CI, et al. A 380-gene meta-signature of active tuberculosis compared with healthy controls. *The European respiratory journal: official journal of the European Society for Clinical Respiratory Physiology*. 2016; 47(6):1873–6. doi: [10.1183/13993003.02121-2015](https://doi.org/10.1183/13993003.02121-2015) PMID: [27076596](https://pubmed.ncbi.nlm.nih.gov/27076596/); PubMed Central PMCID: [PMC4892351](https://pubmed.ncbi.nlm.nih.gov/PMC4892351/).
19. Bloom CI, Graham CM, Berry MP, Rozakeas F, Redford PS, Wang Y, et al. Transcriptional blood signatures distinguish pulmonary tuberculosis, pulmonary sarcoidosis, pneumonias and lung cancers. *PloS one*. 2013; 8(8):e70630. Epub 2013/08/14. doi: [10.1371/journal.pone.0070630](https://doi.org/10.1371/journal.pone.0070630) PMID: [23940611](https://pubmed.ncbi.nlm.nih.gov/23940611/); PubMed Central PMCID: [PMC3734176](https://pubmed.ncbi.nlm.nih.gov/PMC3734176/).
20. Pankla R, Buddhisa S, Berry M, Blankenship DM, Bancroft GJ, Banchereau J, et al. Genomic transcriptional profiling identifies a candidate blood biomarker signature for the diagnosis of septicemic melioidosis. *Genome Biol*. 2009; 10(11):R127. doi: [10.1186/gb-2009-10-11-r127](https://doi.org/10.1186/gb-2009-10-11-r127) PMID: [19903332](https://pubmed.ncbi.nlm.nih.gov/19903332/); PubMed Central PMCID: [PMC3091321](https://pubmed.ncbi.nlm.nih.gov/PMC3091321/).
21. Chaussabel D, Quinn C, Shen J, Patel P, Glaser C, Baldwin N, et al. A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. *Immunity*. 2008; 29(1):150–64. doi: [10.1016/j.immuni.2008.05.012](https://doi.org/10.1016/j.immuni.2008.05.012) PMID: [18631455](https://pubmed.ncbi.nlm.nih.gov/18631455/); PubMed Central PMCID: [PMC2727981](https://pubmed.ncbi.nlm.nih.gov/PMC2727981/).
22. Milman N, Selroos O. Pulmonary sarcoidosis in the Nordic countries 1950–1982. II. Course and prognosis. *Sarcoidosis*. 1990; 7(2):113–8. PMID: [2255786](https://pubmed.ncbi.nlm.nih.gov/2255786/).
23. Rybicki BA, Major M, Popovich J Jr., Maliarik MJ, Iannuzzi MC. Racial differences in sarcoidosis incidence: a 5-year study in a health maintenance organization. *American journal of epidemiology*. 1997; 145(3):234–41. doi: [10.1093/oxfordjournals.aje.a009096](https://doi.org/10.1093/oxfordjournals.aje.a009096) PMID: [9012596](https://pubmed.ncbi.nlm.nih.gov/9012596/).
24. Public Health England. Tuberculosis in the UK 2014 report. 2014.
25. Kruijshaar ME, Abubakar I. Increase in extrapulmonary tuberculosis in England and Wales 1999–2006. *Thorax*. 2009; 64(12):1090–5. doi: [10.1136/thx.2009.118133](https://doi.org/10.1136/thx.2009.118133) PMID: [19850965](https://pubmed.ncbi.nlm.nih.gov/19850965/).
26. Yang Z, Kong Y, Wilson F, Foxman B, Fowler AH, Marrs CF, et al. Identification of risk factors for extrapulmonary tuberculosis. *Clin Infect Dis*. 2004; 38(2):199–205. doi: [10.1086/380644](https://doi.org/10.1086/380644) PMID: [14699451](https://pubmed.ncbi.nlm.nih.gov/14699451/).
27. Forssbohm M, Zwahlen M, Loddenkemper R, Rieder HL. Demographic characteristics of patients with extrapulmonary tuberculosis in Germany. *The European respiratory journal: official journal of the European Society for Clinical Respiratory Physiology*. 2008; 31(1):99–105. doi: [10.1183/09031936.00020607](https://doi.org/10.1183/09031936.00020607) PMID: [17804450](https://pubmed.ncbi.nlm.nih.gov/17804450/).
28. Veenstra H, Baumann R, Carroll NM, Lukey PT, Kidd M, Beyers N, et al. Changes in leucocyte and lymphocyte subsets during tuberculosis treatment; prominence of CD3dimCD56+ natural killer T cells in fast treatment responders. *Clinical and experimental immunology*. 2006; 145(2):252–60. doi: [10.1111/j.1365-2249.2006.03144.x](https://doi.org/10.1111/j.1365-2249.2006.03144.x) PMID: [16879244](https://pubmed.ncbi.nlm.nih.gov/16879244/); PubMed Central PMCID: [PMC1809688](https://pubmed.ncbi.nlm.nih.gov/PMC1809688/).
29. Gonzalez OY, Adams G, Teeter LD, Bui TT, Musser JM, Graviss EA. Extra-pulmonary manifestations in a large metropolitan area with a low incidence of tuberculosis. *The international journal of tuberculosis and lung disease: the official journal of the International Union against Tuberculosis and Lung Disease*. 2003; 7(12):1178–85. PMID: [14677893](https://pubmed.ncbi.nlm.nih.gov/14677893/).

30. 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. *Scientific reports*. 2014; 4:5658. doi: [10.1038/srep05658](https://doi.org/10.1038/srep05658) PMID: [25014250](https://pubmed.ncbi.nlm.nih.gov/25014250/).
31. Breen RA, Leonard O, Perrin FM, Smith CJ, Bhagani S, Cropley I, et al. How good are systemic symptoms and blood inflammatory markers at detecting individuals with tuberculosis? *The international journal of tuberculosis and lung disease: the official journal of the International Union against Tuberculosis and Lung Disease*. 2008; 12(1):44–9. PMID: [18173876](https://pubmed.ncbi.nlm.nih.gov/18173876/).
32. Cai Y, Yang Q, Tang Y, Zhang M, Liu H, Zhang G, et al. Increased complement c1q level marks active disease in human tuberculosis. *PloS one*. 2014; 9(3):e92340. doi: [10.1371/journal.pone.0092340](https://doi.org/10.1371/journal.pone.0092340) PMID: [24647646](https://pubmed.ncbi.nlm.nih.gov/24647646/); PubMed Central PMCID: PMC3960215.
33. Koth LL, Solberg OD, Peng JC, Bhakta NR, Nguyen CP, Woodruff PG. Sarcoidosis blood transcriptome reflects lung inflammation and overlaps with tuberculosis. *American journal of respiratory and critical care medicine*. 2011; 184(10):1153–63. Epub 2011/08/20. doi: [10.1164/rccm.201106-1143OC](https://doi.org/10.1164/rccm.201106-1143OC) PMID: [21852540](https://pubmed.ncbi.nlm.nih.gov/21852540/); PubMed Central PMCID: PMC3262024.
34. Koh GC, Schreiber MF, Bautista R, Maude RR, Dunachie S, Limmathurotsakul D, et al. Host responses to melioidosis and tuberculosis are both dominated by interferon-mediated signaling. *PloS one*. 2013; 8(1):e54961. doi: [10.1371/journal.pone.0054961](https://doi.org/10.1371/journal.pone.0054961) PMID: [23383015](https://pubmed.ncbi.nlm.nih.gov/23383015/); PubMed Central PMCID: PMC3558488.
35. Sweeney TE, Braviak L, Tato CM, Khatri P. Genome-wide expression for diagnosis of pulmonary tuberculosis: a multicohort analysis. *The Lancet Respiratory medicine*. 2016; 4(3):213–24. doi: [10.1016/S2213-2600\(16\)00048-5](https://doi.org/10.1016/S2213-2600(16)00048-5) PMID: [26907218](https://pubmed.ncbi.nlm.nih.gov/26907218/); PubMed Central PMCID: PMC4838193.