Europe PMC Funders Group
Author Manuscript
Nature. Author manuscript; available in PMC 2012 November 08.

Published in final edited form as:
Nature. 2010 August 19; 466(7309): 973–977. doi:10.1038/nature09247.

An Interferon-Inducible Neutrophil-Driven Blood Transcriptional Signature in Human Tuberculosis

Matthew P. R. Berry[1], **Christine M. Graham**[1,*], **Finlay W. McNab**[1,*], **Zhaohui Xu**[6], **Susannah A.A. Bloch**[3], **Tolu Oni**[4,5], **Katalin A. Wilkinson**[2,4], **Romain Banchereau**[9], **Jason Skinner**[6], **Robert J. Wilkinson**[2,4,5], **Charles Quinn**[6], **Derek Blankenship**[7], **Ranju Dhawan**[8], **John J. Cush**[6], **Asuncion Mejias**[10], **Octavio Ramilo**[10], **Onn M. Kon**[3], **Virginia Pascual**[6], **Jacques Banchereau**[6], **Damien Chaussabel**[6], and **Anne O'Garra**[1,#]

[1]Division of Immunoregulation, MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London, NW7 1AA, UK.

[2]Division of Mycobacterial Research, MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London, NW7 1AA, UK.

[3]Department of Respiratory Medicine, Imperial College Healthcare NHS Trust.

[4]Institute of Infectious Diseases and Molecular Medicine Institute of Infectious Diseases and Molecular Medicine, University of Cape Town

[5]Division of Medicine, Imperial College London.

[6]Baylor Institute for Immunology Research-ANRS Center for Human Vaccines, INSERM U899, 3434 Live Oak St., Dallas, Texas 75204

[7]Institute for Health Care Research and Improvement, Baylor Health Care System, Dallas, Texas 75204.

[8]Department of Radiology, Imperial College Healthcare NHS Trust.

[9]UT Southwestern Medical Center, Dallas, Texas.

[10]Center for Vaccines and Immunity, The Research Institute at Nationwide Children's Hospital, 700 Children's Drive, Columbus, OH 43205, USA.

Abstract

Tuberculosis (TB), caused by infection with *Mycobacterium tuberculosis* (*M. tuberculosis*), is a major cause of morbidity and mortality worldwide and efforts to control TB are hampered by difficulties with diagnosis, prevention and treatment 1,2. Most people infected with *M. tuberculosis* remain asymptomatic, termed latent TB, with a 10% lifetime risk of developing active

[#]Please address correspondence to AOG..
[*]CG & FMcN contributed equally to this study.

Author contributions

M.B, D.C, O.M.K, and A.OG. designed the study on TB with input from JB and RW and for other diseases with input from V.P and O.R; M.B, S.B, T.O, K.W, J.C, A.M, R.B. and O.M.K recruited, sampled and collected patient data; M.B, R.B, A.M. and C.G processed whole blood for microarray experiments with help from J.S; C.G performed blood cell subset separations and processing for microarray experiments with help from J.S; M.B, C.G and Z.X performed microarray data analysis, with advice and input from JS, DC and VP; M.B. and Z.X. performed Ingenuity, Modular and Molecular Distance to Health Analyses; M.B. performed multiplex serum analyses; F.McN performed flow cytometry analysis; D.C., V.P and A.OG supervised data analysis; M.B and D.B performed statistical analysis; M.B, S.B, R.D and O.M.K performed analyses of radiology; A.OG and M.B wrote the manuscript, with early input from C.G, F.McN, J.B, D.C. and J.S, and subsequently all authors provided advice and approved the final manuscript.

All microarray data are deposited in GEO under Accession Numbers GSE19491, GSE19444, GSE19443, GSE19442, GSE19439, GSE19435. Some of the work has been submitted as US Patent Application PCT 371: Blood Transcriptional Signature of Mycobacterium Tuberculosis Infection: Serial No: 12/602,488.

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

TB disease, but current tests cannot identify which individuals will develop disease 3. The immune response to *M. tuberculosis* is complex and incompletely characterized, hindering development of new diagnostics, therapies and vaccines 4,5. We identified a whole blood 393 transcript signature for active TB in intermediate and high burden settings, correlating with radiological extent of disease and reverting to that of healthy controls following treatment. A subset of latent TB patients had signatures similar to those in active TB patients. We also identified a specific 86-transcript signature that discriminated active TB from other inflammatory and infectious diseases. Modular and pathway analysis revealed that the TB signature was dominated by a neutrophil-driven interferon (IFN)-inducible gene profile, consisting of both IFN-γ and Type I IFNαβ signalling. Comparison with transcriptional signatures in purified cells and flow cytometric analysis, suggest that this TB signature reflects both changes in cellular composition and altered gene expression. Although an IFN signature was also observed in whole blood of patients with Systemic Lupus Erythematosus (SLE), their complete modular signature differed from TB with increased abundance of plasma cell transcripts. Our studies demonstrate a hitherto under-appreciated role of Type I IFNαβ signalling in TB pathogenesis, which has implications for vaccine and therapeutic development. Our study also provides a broad range of transcriptional biomarkers with potential as diagnostic and prognostic tools to combat the TB epidemic.

Blood transcriptional profiling has improved diagnosis and understanding of disease pathogenesis 6-9. Such a comprehensive unbiased survey will provide insights into the immunopathogenesis of TB, leading to advances in control of this complex disease. Genome-wide transcriptional profiles were generated from blood from active TB patients (before treatment), latent TB patients and healthy controls (Supplementary Fig. 1, Tables 1, 2). A distinct 393-transcript signature was defined in active TB patients (Training Set, London), using a combination of expression level and statistical filters and hierarchical clustering (Supplementary Fig. 2b(i), Table 3, Methods). We then applied the 393-transcript list to two independent cohorts (UK Test Set; South African Validation Set). Hierarchical clustering of transcriptional profiles showed active TB patients cluster independently of latent TB and healthy controls, in both intermediate (London) and high burden (South Africa) regions with a significant association between cluster and study group (Fisher's exact test, $p = 0.00001365$, UK, Fig. 1a; and $p = 5.79 \times 10^{-10}$, South Africa, Fig. 1b). This was independent of ethnicity, age or gender (Supplementary Figs. 2b(ii), c and d). The transcriptional profiles of 10 – 25% of latent TB patients (5/21 Test Set, 3/31 Validation Set) clustered with active TB patients (Fig. 1a, 1b). K-nearest neighbours class prediction using the 393-transcript list, gave a sensitivity of 61.67%, specificity 93.75%, and an indeterminate rate of 1.9% for the Test Set, (Supplementary Table 4), with 5 latent TB patients classified as active TB and 4 active TB patients misclassified. In the Validation Set the sensitivity was 94.12%, specificity 96.67%, and indeterminate rate 7.8%. The UK patients were of diverse ethnicity, potentially infected with different *M. tuberculosis* lineages, suggesting the signature may be independent of bacterial clade, although molecular typing was not available. The proportion of latent patients having a transcriptional signature similar to that of active TB was equal to the expected frequency of patients at risk of progression to active disease 3, potentially identifying latent TB patients with sub-clinical active disease or higher burden latent infection.

Four out of 21 active TB patients in the Test Set, also misclassified by class prediction, clustered with healthy controls and latent TB patients (●, #, ■, ◆ Fig. 1a), demonstrating molecular heterogeneity that could reflect clinical variance. To address this, radiographic extent of disease was assessed by 3 physicians, blinded to clinical diagnosis and transcriptional profile (Supplementary Fig. 3)10. The median “Molecular Distance to Health”11, a composite of the number of transcripts in a profile that significantly differ from

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

the healthy control baseline, and the degree of that difference for advanced disease, was significantly higher than for those with minimal or no disease (Fig. 1c). We show for the first time that the transcriptional signature in blood correlates with extent of disease in active TB patients, and reflects changes at the site of disease. The transcriptional signature was diminished in active TB patients after 2 months, and completely extinguished by 12 months after treatment, with “Molecular Distance to Health” at 12 months significantly lower than at baseline pretreatment (Fig. 1d and Supplementary Fig. 4), reflecting radiographic improvement. Thus the blood transcriptional signature of active TB patients could be used to monitor efficacy of treatment, and is reflective of the host response to *M. tuberculosis* infection.

The 393-gene active TB signature may reflect common inflammatory responses evoked during many diseases. We therefore identified a TB-specific 86-gene whole blood signature through analysis of significance 12, compared to patients with other bacterial and inflammatory diseases (Supplementary Fig. 5, Tables 5 and 6). This 86-gene signature was then tested against patients normalized to their own controls from 7 independent datasets by class prediction (K-nearest neighbours) (Fig. 2a). Sensitivities in the TB Training and Validation Sets were 92% and 90% respectively, distinguishing active TB from other diseases with a pooled specificity of 83% (Supplementary Table 7). As with the 393-gene signature this 86-gene signature was diminished in response to treatment (Fig. 2b), and reflected the same heterogeneity in identical patient samples (Supplementary Fig. 6).

To identify functional components of the transcriptional host response during active TB, we employed a modular data mining strategy, using sets of genes that are coordinately expressed in different diseases and defined as specific modules, often demonstrating coherent functional relationships through unbiased literature profiling 7. The blood modular signature of active TB patients as compared to healthy controls (filtering out only undetected transcripts, α=0.01, in at least 2 individuals) was similar in all 3 TB datasets (Fig. 3a, Supplementary Fig. 7), confirming the reproducibility of the transcriptional signature. The modular TB signature revealed decreased abundance of B cell (Module, M1.3) and T cell (M2.8) transcripts and increased abundance of myeloid related transcripts (M1.5 and M2.6). The largest proportion of transcripts changing in a given module in TB, were within the IFN-inducible module (M3.1; 75 - 82% of IFN-module transcripts; Fig. 3a). Since a Type I IFN-inducible signature, linked with disease pathogenesis, has been demonstrated in PBMC from patients with SLE 13,14, we compared whole blood modular signatures from patients with other diseases. SLE patients demonstrated over-representation of the IFN-inducible module (M3.1, Fig. 3a, quantitated in Supplementary Fig. 8), but displayed a plasma-cell related module absent in TB (M1.1, Fig. 3a; Supplementary Fig. 8). The blood modular signature from patients with Group A Streptococcus or Staphylococcus infection, or Still's disease showed minimal to no change in the IFN-inducible module (M3.1) but marked over-representation of the neutrophil-related module (M2.2), distinguishing these diseases from TB (Fig. 3a, Supplementary Fig. 8). Thus the IFN-inducible signature is not common to all inflammatory responses, but is preferentially induced during some diseases, potentially reflecting protection or pathogenesis. Although SLE and TB share common inflammatory components such as an IFN-inducible response, the overall pattern of transcriptional changes (Fig. 3a) and their amplitude (Supplementary Fig. 8) distinguishes one disease from another.

The TB blood transcriptional signature could represent altered cell composition or changes in gene expression in discrete cellular populations. Percentages of $CD4^+$ and $CD8^+$ T cells and B cells, assessed by flow cytometry were significantly diminished in active TB patients, with reduced numbers of total and central memory $CD4^+$ T cells (Fig. 3b; Supplementary Figs. 9a and b), in keeping with previous studies 15. That the reduction in T cell transcripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

revealed by the modular analysis (Fig. 3a) resulted from changes in cell numbers in the blood, was further confirmed since expression of these transcripts in purified T cells from the same individuals did not differ between TB patients and healthy controls (Supplementary Fig. 9c). In contrast, the increase in myeloid transcripts (M1.5, M2.6, Fig. 3a) in the blood of active TB patients was not accounted for by changes in monocytes ($CD14^+$, $CD16^-$) or neutrophils ($CD16^+$, $CD14^-$) although inflammatory monocytes ($CD14^+$, $CD16^+$) were increased (Fig. 3c; Supplementary Fig. 10a), as in other diseases 16. Increased abundance of myeloid transcripts was less pronounced in purified monocytes ($CD14^+$) (Supplementary Figure 10b), suggesting involvement of other cells.

Pathway analysis confirmed IFN signalling as the most significantly over-represented pathway in the 393-gene signature (Fischer's Exact test, Benjamini-Hochberg multiple test correction, $p<0.0000001$, Supplementary Fig. 11). Genes downstream of both IFN-γ and Type I IFN α/β receptor signalling were significantly over-represented in blood from active TB patients (Fig. 4a, b, c). IFN-α2a and IFN-γ proteins were not elevated in serum from active TB patients, although the IFN-inducible chemokine CXCL10 (IP10) was (Supplementary Fig. 11c, d, e).

Although IFN-γ is protective during immune responses to intracellular pathogens, including mycobacteria 4,17,18, the role of Type I IFNαβ is less clear. Type I IFN signalling is crucial for defense against viral infections but may be detrimental during bacterial 19, including mycobacterial infections20,21. Absence of IFNαβ signalling in mice improved outcome after infection with highly virulent 20-22 but not less virulent strains of *M. tuberculosis* 23. Highly virulent strains of *M. tuberculosis* induce higher levels of Type I IFNs 20. There are reports of TB reactivation during IFN-α treatment for hepatitis C viral infection 24. The increase in Type I IFNαβ-inducible transcripts in the blood of active TB patients (Fig. 4c), correlating with disease severity, provides the first data in human disease to support a role for Type I IFNs in TB pathogenesis. These IFN-inducible transcripts were over-expressed in purified blood neutrophils and to a lesser extent monocytes, but not $CD4^+$ and $CD8^+$ T cells, from active TB patients, compared to healthy controls (Fig. 4d; OAS1, IFI6, IFI44, IFI44L, OAS3, IRF7, IFIH1, IFI16, IFIT3, IFIT2, OAS2, IFITM3, IFITM1, GBP1, GBP5, STAT1, GBP2, TAP1, STAT1, STAT2, IFI35, TAP2, CD274, SOCS1, CXCL10, IFIT5). Neutrophils are the predominant cell type infected with rapidly replicating *M. tuberculosis* in TB patients 25. Evidence from genetically susceptible mice suggests that neutrophils contribute to pathology during *M. tuberculosis* infection 26. Our studies support a role for neutrophils in TB pathogenesis, which may result from over-activation by IFN-γ and Type I IFNs.

Earlier microarray studies, limited by small numbers of patients and custom microarrays, reported a small number of genes in blood associated with TB 27,28. Here we provide the first complete description of the human blood transcriptional signature of TB. The signature of active TB, observed in 10 – 20% of latent TB patients, may identify those individuals who will develop active disease, facilitating targeted preventative therapy, but longitudinal studies are needed to assess this. That the TB signature is dominated by Type I IFN-signalling and reflects extent of lung disease, may indicate the process leading to disease susceptibility. These data improve our understanding of the fundamental biology of TB and may offer future leads for diagnosis and treatment.

METHODS SUMMARY

Whole blood was collected into Tempus tubes (Applied Biosystems, CA, USA) from active TB patients (confirmed by culture for *M. tuberculosis)*; Latent TB patients (defined by a positive tuberculin-skin test (TST) (London) and/or a positive *M. tuberculosis* antigen-

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

specific IFN-γ release assay, IGRA; healthy controls (recruited in London; TST/IGRA negative). RNA was extracted from whole blood and purified (by Dynabeads, Invitrogen) neutrophils, monocytes, CD4+ and CD8+ T cells and genome-wide transcriptional profiles were generated using Illumina HT12 beadarrays, and analysed using Genespring GX, (see main Methods section). Calculation of "Molecular Distance to Health" 11, transcriptional modular analysis 7, and analysis of significance 12 were performed as previously described. Pathway analysis was performed using Ingenuity (Ingenuity Systems, Inc., CA, USA). Multiplex Serum Protein Measurement was performed using Milliplex Multi-Analyte Profiling System by Millipore UK, Ltd, Dundee. Flow Cytometry was performed on a Beckman Coulter Cyan using Summit Software Version 3.02, followed by FlowJo analysis.

METHODS (for online publication)

Participant Recruitment and Patient Characterization

The local Research Ethics Committees at St. Mary's Hospital London, UK (REC 06/Q0403/128) and University of Cape Town, Cape Town, Republic of South Africa (REC 012/2007) approved the study. All participants were aged over 18 years old and gave written informed consent. Participants were recruited from St. Mary's Hospital and Hammersmith Hospital, Imperial College Healthcare NHS Trust, London, UK, Hillingdon Hospital, The Hillingdon Hospitals NHS Trust, Uxbridge, UK and the Ubuntu TB/HIV clinic, Khayelitsha, Cape Town, South Africa. Patients were prospectively recruited and sampled, before any anti-mycobacterial treatment was initiated, but only included in the final analysis if they met the full clinical criteria for their relevant study group. A subset of active TB patients recruited into the first cohort recruited in London was also sampled at 2 and 12 months after the initiation of therapy. Patients who were pregnant, immunosuppressed, or who had diabetes, or autoimmune disease were ineligible and excluded from this study. In South Africa, all participants had routine HIV testing using the Abbott Determine® HIV1/2 rapid antibody assay test kit (Abbott Laboratories, Abbott Park, Illinois, USA). Active TB patients were confirmed by laboratory isolation of *M. tuberculosis* on mycobacterial culture of a respiratory specimen (either sputum or bronchoalvelolar lavage fluid) with sensitivity testing performed by The Royal Brompton Hospital Mycobacterial Reference Laboratory, London, UK or The Reference Lab of the National Health Laboratory Service, Groote Schuur Hospital, Cape Town. In the UK, latent TB patients were recruited from those referred to the TB clinic with a positive TST, together with a positive result using an IGRA. Latent TB participants in South Africa were recruited from individuals self-referring to the voluntary testing clinic at the Ubuntu TB/HIV clinic, and IGRA positivity alone was used to confirm the diagnosis, irrespective of TST result (although this was still performed). Healthy control participants were recruited from volunteers at the National Institute for Medical Research (NIMR), Mill Hill, London, UK. To meet the final criteria for study inclusion healthy volunteers had to be negative by both TST and IGRA.

Tuberculin Skin Testing

This was performed according to the UK guidelines 29 using 0.1ml (2TU) tuberculin PPD (RT23, Serum Statens Institute, Copenhagen, Denmark). A positive TST was termed ≥6mm if BCG unvaccinated, ≥15mm if BCG vaccinated, as per the UK national guidelines 30.

Interferon Gamma Release Assay Testing

The QuantiFERON® Gold In-Tube assay (Cellestis, Carnegie, Australia) was performed according to the manufacturers instructions.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Total and Differential Leucocyte Counts

2mls of whole blood was collected into Terumo Venosafe 5ml K2-EDTA tubes (Terumo Europe, Leuven, Belgium). Samples were then analysed within 4 hours using the Nihon Kohden MEK-6400 Automated Hematology Analyzer (Nihon Kohden Corporation, Tokyo, Japan).

Assessment of Radiographic Extent of Disease

Plain chest radiographs were obtained for all patients recruited in London as digital images and graded by three independent clinicians, blinded to the transcriptional profiles and the clinical data, using a modified version of the classification system of the U.S. National Tuberculosis and Respiratory Disease Association 10. This system characterises the radiographic extent of disease into “Minimal”, “Moderately advanced” or “Far advanced” stages, according to criteria based upon the density and extent of lesions and presence of absence of cavitation. We modified the system for use in our study so that it also included a classification of “No disease, and accounted for the presence of pleural disease or lymphadenopathy. The system was then converted into a decision tree to aid classification (Supplementary Figure 3a).

RNA Sampling, Extraction and Processing for Microarray Analysis

3mls of whole blood was collected into Tempus tubes (Applied Biosystems, Foster City, CA, USA), vigorously mixed immediately after collection, and stored between −20°C and −80°C before RNA extraction. RNA was isolated from Training Set samples using 1.5mls whole blood and the PerfectPure RNA Blood kit (5 PRIME Inc, Gaithersburg, MD, USA). Test and Validation (SA) Set samples were extracted from 1ml of whole blood using the MagMAX™-96 Blood RNA Isolation Kit (Applied Biosystems/Ambion, Austin, TX, USA) according to the manufacturer's instructions. 2.5mg of isolated total RNA was then globin reduced using the GLOBINclear™ 96-well format kit (Applied Biosystems/Ambion, Austin, TX, USA) according to the manufacturer's instructions. Total and globin-reduced RNA integrity was assessed using an Agilent 2100 Bioanalyzer showing a quality of RIN of 7 – 9.5 (Agilent Technologies, Santa Clara, CA, USA). RNA yield was assessed using a Nanodrop 1000 spectrophotometer (NanoDrop Products, The rmo Fisher Scientific Inc, Wilmington, DE, USA). Biotinylated, amplified antisense complementary RNA targets (cRNA) were then prepared from 200 - 250ng of the globin-reduced RNA using the Illumina CustomPrep RNA amplification kit (Applied Biosystems/Ambion, Austin, TX, USA). 750ng of labelled cRNA was hybridized overnight to Illumina Human HT-12 BeadChip arrays (Illumina Inc, San Diego, CA, USA), which contain more than 48,000 probes. The arrays were then washed, blocked, stained and scanned on an Illumina BeadStation 500 following the manufacturer's protocols. Illumina BeadStudio v2 software (Illumina Inc, San Diego, CA, USA) was used to generate signal intensity values from the scans.

Separated cells isolation and RNA extraction

Whole blood was collected in EDTA. Neutrophils ($CD15^+$), monocytes ($CD14^+$), $CD4^+$ T cells and $CD8^+$ T cells were isolated sequentially using Dynabeads according to manufacturers instructions. RNA was extracted from whole blood (5′ Prime Perfect Pure kit) or separated cell populations (Qiagen RNEasy Mini Kit) and stored at −80°C until use.

Microarray Data Analysis

Normalisation—Illumina BeadStudio v2 software was used to subtract background, and scale average signal intensity for each sample to the global average signal intensity for all samples, except for Figure 5 and Supplementary Figures 5c and 6b, where signals are normalised to the median of the control. A gene expression analysis software program,

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

GeneSpring GX, version 7.1.3 (Agilent Technologies, Santa Clara, CA, USA, hereafter referred to as GeneSpring), was used to perform further normalisation. All signal intensity values less than 10 were set to equal 10. Next, per-gene normalisation was applied, by dividing the signal intensity of each probe in each sample by the median intensity for that probe across all samples. These normalised data were used for all downstream analyses except the assessment of molecular distance to health detailed below.

Filtering—Using GeneSpring, all transcripts were filtered first to select detected transcripts; those called present in greater than 10% of all samples. Present calls are selected if the signal precision is <0.01. The remaining transcripts are filtered to select the most variable probes - those that have a minimum of 2-fold expression change compared to the median intensity across all samples, in greater than 10% of all samples.

Unsupervised analysis: hierarchical clustering and class discovery—This approach aims to create an unbiased grouping of samples on the basis of their molecular profiles, independently of any other phenotypic or clinical classification. Transcripts meeting the filtering criteria are then subjected to hierarchical clustering using GeneSpring. For hierarchical clustering of genes, we use a clustering algorithm based upon Pearson correlation, creating a vertical dendrogram of genes, where transcripts with a similar expression pattern across all samples are grouped together. The distances between branches of the tree relate to the similarity of the expression patterns, and the distance between clusters is determined by the average of the distance between all points in each cluster, known as average linkage. The vertical expression profiles so generated can then be subjected to the same hierarchical clustering algorithm, now grouping individual participants into horizontally presented clusters on the basis of the similarity of their expression profiles. For this stage, we base the clustering algorithm on Spearman Rank correlation. By examining the cluster membership we can assess both whether the samples are grouping according to known factors (clinical diagnosis, demographic features) and also discover if there are unknown subclasses within the dataset.

Supervised analysis: statistical filtering and class comparison—The aim of the supervised analysis is to identify transcripts which are differentially expressed between study groups and that might serve as classifiers or yield insight into immunopathogenesis, i.e. class comparison. The filtered list of transcripts generated for unsupervised analysis was used as the starting point for the supervised analysis, i.e. those transcripts that were both detected, and had at least 2 fold change in expression compared to the median, in greater than 10% of all samples. Using GeneSpring, these transcripts were then tested using the Kruskal-Wallis test for comparisons across all study groups, with a=0.01. Adjustment for multiple testing was applied using the Benjamini-Hochberg False Discovery Rate (FDR) set at 1%. Lists of transcripts generated in this way were then used for hierarchical clustering as described above. Interpretation of functional roles of individual transcripts was established by searching the database at the National Center for Biotechnology Information Gene database at http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene

Class Prediction—We utilised one of the class prediction tools available within GeneSpring. The prediction model employed the K-nearest neighbours algorithm, with 10 neighbours and a p value ratio cut off of 0.5. All genes from the 393 transcript list were used for the prediction. The prediction model was refined by cross-validation on the training set, with the one Active outlier excluded. This model was then used to predict the classification of the samples in the independent Test and Validation Sets. Where no prediction was made, this was recorded as an indeterminate result. Sensitivity, specificity and 95% confidence

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

intervals (95% CI) were determined using GraphPad Prism version 5.02 for Windows. P-values were determined using two-sided Fisher's Exact test

Supervised analysis: (i) "Molecular Distance to Health"—This technique was performed as previously described11. It aims to convert transcript abundance values into a representative score indicating the degree of transcriptional perturbation of a given sample compared to a healthy baseline. This is performed by determining whether the expression values of a given sample lie inside or outside two standard deviations from the mean of the healthy controls.

Supervised analysis: (ii) Pathway analysis—Additional functional analysis of differentially expressed genes was performed using Ingenuity Pathways Analysis (Ingenuity® Systems, Inc., Redwood, CA, USA, www.ingenuity.com). Canonical pathways analysis identified the pathways from the Ingenuity Pathways Analysis that were most significantly represented in the dataset. The significance of the association between the dataset and the canonical pathway was measured using Fisher's Exact test to calculate a p-value representing the probability that the association between the transcripts in the dataset and the canonical pathway is explained by chance alone, with a Benjamini-Hochberg correction for multiple testing applied. The program can also be used to map the canonical network and overlay it with expression data from the dataset.

Supervised analysis: (iii) Transcriptional modular analysis—This analysis was performed as described previously7,11. In the context of the present study, since the modular framework was derived using Affymetrix HG U133A&B GeneChips, it was necessary to translate the probes comprising the modules into their equivalents on the Illumina platform. RefSeq IDs were used to match probes between the Affymetrix HG U133 and Illumina HT-12 V3 platforms. Unambiguous matches were found for 2,109 out of the 5,348 Affymetrix probe sets, and these were used in the present modular analysis. The matching probes were preserved in their original modules. To graphically present the global transcriptional changes, for the disease group as a whole versus the healthy control group as a whole, spots are aligned on a grid, with each position corresponding to a different module based on their original definition. Spot intensity indicates the percentage of differentially expressed transcripts changing in the direction shown, from the total number of transcripts detected for that module, while spot colour indicates the polarity of the change (red = over-represented, blue = under-represented).

Significance Analysis12—Transcriptional changes in whole blood were evaluated through statistical group comparison performed systematically for active TB (Test Set), Staphylococcus infection, Still's syndrome, and adult and pediatric SLE versus their respective healthy controls, which allows the normalization of each disease group to its own matched healthy control group, thus avoiding biological or technical confounding factors. A TB-specific whole blood signature composed of 86 genes was identified (Supplementary Fig. 5 and Table 5 and 6, $p<0.01$) not in the four other datasets ($p>0.05$) using Mann-Whitney T-Test with Benjamini and Hochberg False Discovery Rate multiple testing correction. Class prediction was performed using K-Nearest Neighbors algorithm, as before.

Multiplex Serum Protein Measurement

1 – 4ml blood was collected into serum clot activator tubes (either Greiner BioOne 1ml vacuette tubes, ref 454098, Greiner BioOne, Kremsmünst, Austria; or BD 4ml vacutainer tubes, ref 368975; Becton Dickinson). Tubes were centrifuged at 2000*g* for 5 minutes at room temperature and the serum portion extracted and frozen at −80°C pending analysis. Analysis was performed by multiplexed cytokine bead-based immunoassay by Millipore UK

Europe PMC Funders Author Manuscripts

(Millipore UK Ltd, Dundee, UK) using the Milliplex® Multi-Analyte Profiling system (Millipore, Billerica, MA, USA). The serum levels of 63 cytokines, chemokines, soluble receptors, growth factors, adhesion molecules and acute phase proteins were measured in this way in each sample. Samples were assayed for levels of MMP-9, C-reactive protein, serum amyloid A, EGF, Eotaxin, FGF-2, Flt-3 Ligand, Fractalkine, G-CSF, GM-CSF, GRO, IFN-a2, IFN-g, IL-10, IL-12p40, IL-12p70, IL-13, IL-15, IL-17, IL-1ą, IL-1b, IL-1Ra, IL-2, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, CXCL10 (IP10), MCP-1, MCP-3, MIP-1 a, MIP-1b, PDGF-AA, PDGF-AB/BB, RANTES, soluble CD40 ligand, soluble IL-2RA, TGF-a, TNF-a, VEGF, MIF, soluble Fas, soluble Fas Ligand, tPAI-1, soluble ICAM-1, soluble VCAM-1, soluble CD30, soluble gp130, soluble IL-1RII, soluble IL-6R, soluble RAGE, soluble TNF-RI, soluble TNF-RII, IL-16, TGF-b1, TGF-b2 and TGFb-3.

Flow Cytometry

200μl of whole blood (collected in Sodium-Heparin tubes) per staining panel was incubated with the appropriate antibodies for 20 minutes at room temperature in the dark. Red blood cells were then lysed using BD FACS lysing solution (BD Biosciences), incubating for 10 minutes at room temperature in the dark. Cells were spun down and washed in 2ml FACS buffer (PBS/ BSA/ Azide) before being fixed in 1% paraformaldehyde. Samples were then run on a Beckman Coulter Cyan using Summit Software Version 3.02. Analysis was carried out using FlowJo Version 8.7.3 for Macintosh (Tree Star, Inc.). Gating strategies used are set out in Supplementary Figures 9 and 10. Flow cytometric data is presented as dot plots (Figures 3, and Supplementary 9 and 10). Where appropriate pooled flow cytometry data was tested for significance using the Mann-Whitney Rank Sum U-test. All antibodies were purchased from BD Pharmingen or Caltag Laboratories (Invitrogen) except for CD45RA, which was purchased from Beckman Coulter.

Statistical Analysis

Molecular distance to health and Modular Framework analysis calculations were performed using Microsoft Excel 2003 (Microsoft Corporation, Redmond, WA, USA). Statistical analysis of continuous variables and correlation analysis was performed using GraphPad Prism version 5.02 for Windows (GraphPad Software, San Diego California USA, www.graphpad.com). Analysis of categorical variables was performed using SPSS version 14 for Windows (Chicago, Illinois, USA).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the patients and volunteer participants. We thank D. Kioussis (NIMR) and D. Young (NIMR) for discussion and input. We thank N. Baldwin (BIIR) for advice and support on bioinformatics analysis, Q-A. Nguyen (BIIR) and colleagues provided technical assistance with microarray processing, S. Caidan (NIMR), J. Wills (NIMR) and S. Phillips (BIIR) for help and advice with sample storage and transport. We thank the TB service at Imperial College Healthcare NHS Trust, Dr. B.M. Haselden and the TB service at Hillingdon Hospital, Uxbridge UK. We thank H. Giedon and R. Seldon for help in laboratory analyses, and Y. Hlombe for patient recruitment and follow up in South Africa. A. Rae (NIMR), T. Dipucchio (BIIR) and K. Palucka (BIIR) provided advice on flow cytometry. We thank J. Brock (NIMR) for help with graphics. MB was supported by an MRC career development fellowship and a grant from the Dana Foundation Program in Human Immunology. The research was funded by the Medical Research Council, UK and The Dana Foundation Program in Human Immunology. AOG, C.G, F.McN are funded by the MRC, UK. VP is supported by NIH R01 AR050770-01, NIH P50 ARO54083, NIH 1 U19 AI082715-01. The work of JB, DC, VP is supported by the Baylor Health Care System Foundation and the National Institutes of Health (U19 AIO57234-02, U01 AI082110, P01 CA084512).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

References

1. Dye, C.; Floyd, K.; Uplekar, M. WHO report. World Health Organization; Geneva: 2008.
2. Kaufmann SH, McMichael AJ. Annulling a dangerous liaison: vaccination strategies against AIDS and tuberculosis. Nat Med. 2005; 11:S33–44. [PubMed: 15812488]
3. Barry CE 3rd, et al. The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. Nat Rev Microbiol. 2009; 7:845–55. [PubMed: 19855401]
4. Cooper AM. Cell-mediated immune responses in tuberculosis. Annu Rev Immunol. 2009; 27:393–422. [PubMed: 19302046]
5. Young DB, Perkins MD, Duncan K, Barry CE 3rd. Confronting the scientific obstacles to global control of tuberculosis. J Clin Invest. 2008; 118:1255–65. [PubMed: 18382738]
6. Ardura MI, et al. Enhanced monocyte response and decreased central memory T cells in children with invasive Staphylococcus aureus infections. PLoS One. 2009; 4:e5446. [PubMed: 19424507]
7. Chaussabel D, et al. A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. Immunity. 2008; 29:150–64. [PubMed: 18631455]
8. Pascual V, Chaussabel D, Banchereau J. A genomic approach to human autoimmune diseases. Annu Rev Immunol. 28:535–71. [PubMed: 20192809]
9. Ramilo O, et al. Gene expression patterns in blood leukocytes discriminate patients with acute infections. Blood. 2007; 109:2066–77. [PubMed: 17105821]
10. Falk A, O'Connor JB. Classification of pulmonary tuberculosis: Diagnosis standards and classification of tuberculosis. National tuberculosis and respiratory disease association. 1969; 12:68–76.
11. Pankla R, et al. Genomic Transcriptional Profiling Identifies a Candidate Blood Biomarker Signature for the Diagnosis of Septicemic Melioidosis. Genome Biol. 2009; 10:R127. [PubMed: 19903332]
12. Allantaz F, et al. Blood leukocyte microarrays to diagnose systemic onset juvenile idiopathic arthritis and follow the response to IL-1 blockade. J Exp Med. 2007; 204:2131–44. [PubMed: 17724127]
13. Baechler EC, et al. Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. Proc Natl Acad Sci U S A. 2003; 100:2610–5. [PubMed: 12604793]
14. Bennett L, et al. Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. J Exp Med. 2003; 197:711–23. [PubMed: 12642603]
15. Beck JS, Potts RC, Kardjito T, Grange JM. T4 lymphopenia in patients with active pulmonary tuberculosis. Clin Exp Immunol. 1985; 60:49–54. [PubMed: 3874015]
16. Auffray C, Sieweke MH, Geissmann F. Blood monocytes: development, heterogeneity, and relationship with dendritic cells. Annu Rev Immunol. 2009; 27:669–92. [PubMed: 19132917]
17. Casanova JL, Abel L. Genetic dissection of immunity to mycobacteria: the human model. Annu Rev Immunol. 2002; 20:581–620. [PubMed: 11861613]
18. Flynn JL, Chan J. Immunology of tuberculosis. Annu Rev Immunol. 2001; 19:93–129. [PubMed: 11244032]
19. Decker T, Muller M, Stockinger S. The yin and yang of type I interferon activity in bacterial infection. Nat Rev Immunol. 2005; 5:675–87. [PubMed: 16110316]
20. Manca C, et al. Hypervirulent M. tuberculosis W/Beijing strains upregulate type I IFNs and increase expression of negative regulators of the Jak-Stat pathway. J Interferon Cytokine Res. 2005; 25:694–701. [PubMed: 16318583]
21. Ordway D, et al. The hypervirulent Mycobacterium tuberculosis strain HN878 induces a potent TH1 response followed by rapid down-regulation. J Immunol. 2007; 179:522–31. [PubMed: 17579073]
22. Manca C, et al. Virulence of a Mycobacterium tuberculosis clinical isolate in mice is determined by failure to induce Th1 type immunity and is associated with induction of IFN-alpha /beta. Proc Natl Acad Sci U S A. 2001; 98:5752–7. [PubMed: 11320211]

Europe PMC Funders Author Manuscripts

23. Cooper AM, Pearl JE, Brooks JV, Ehlers S, Orme IM. Expression of the nitric oxide synthase 2 gene is not essential for early control of Mycobacterium tuberculosis in the murine lung. Infect Immun. 2000; 68:6879–82. [PubMed: 11083808]

24. Telesca C, et al. Interferon-alpha treatment of hepatitis D induces tuberculosis exacerbation in an immigrant. J Infect. 2007; 54:e223–6. [PubMed: 17307255]

25. Eum SY, et al. Neutrophils are the predominant infected phagocytic cells in the airways of patients with active pulmonary tuberculosis. Chest. 2009

26. Eruslanov EB, et al. Neutrophil responses to Mycobacterium tuberculosis infection in genetically susceptible and resistant mice. Infect Immun. 2005; 73:1744–53. [PubMed: 15731075]

27. Jacobsen M, et al. Candidate biomarkers for discrimination between infection and disease caused by Mycobacterium tuberculosis. J Mol Med. 2007; 85:613–21. [PubMed: 17318616]

28. Mistry R, et al. Gene-expression patterns in whole blood identify subjects at risk for recurrent tuberculosis. J Infect Dis. 2007; 195:357–65. [PubMed: 17205474]

29. Salisbury, D.; Ramsay, M. Immunization against infectious diseases - the Green Book. *D.O.Health*. The Stationery Office; London: 2006. p. 391-408.

30. National Institute for Health and Clinical Excellence. Royal College of Physicians; UK: 2006.

Europe PMC Funders Author Manuscripts

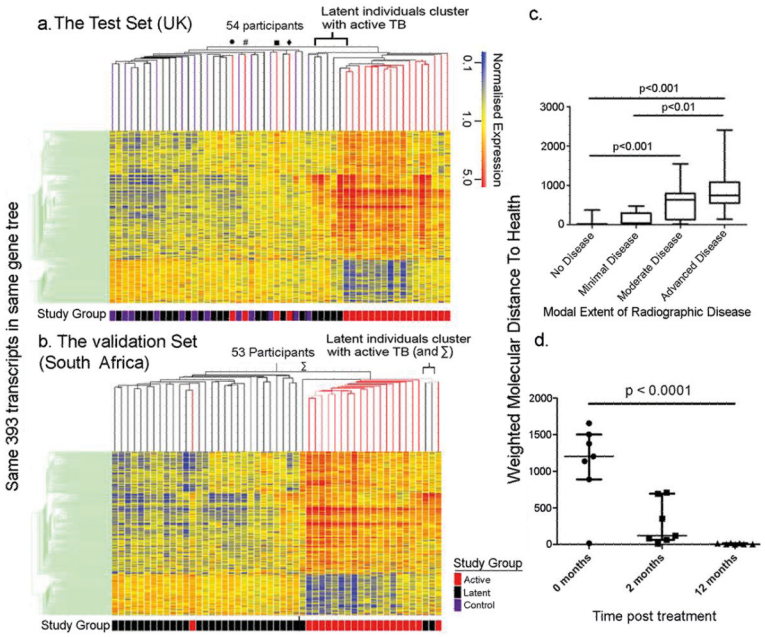

Figure 1. A distinct whole blood 393-gene transcriptional signature of active TB

393-transcripts differentially expressed in whole blood of active and latent TB patients and healthy controls **a,** Test Set. **b,** Validation Set (SA) profiles, ordered by hierarchical clustering (Spearman correlation with average linkage) creating a condition tree, upper horizontal edge of heatmap; study grouping (clinical phenotype) coloured blocks at each profile base. Heatmap rows = genes, columns = participants. **c,** Profiles were grouped according to radiographic extent of disease and the mean "Molecular Distance to Health" compared between groups (Methods) (Kruskal-Wallis ANOVA, Dunn's multiple comparison, *** = p <0.0001). **d,** Active TB patients at 0, 2 and 12 months after initiation of anti-mycobacterial treatment. The mean "Molecular Distance to Health" for each timepoint was compared (Friedman's repeated measures test, Dunn's multiple comparison). Horizontal bars indicate the median, 5th and 95th percentiles.

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

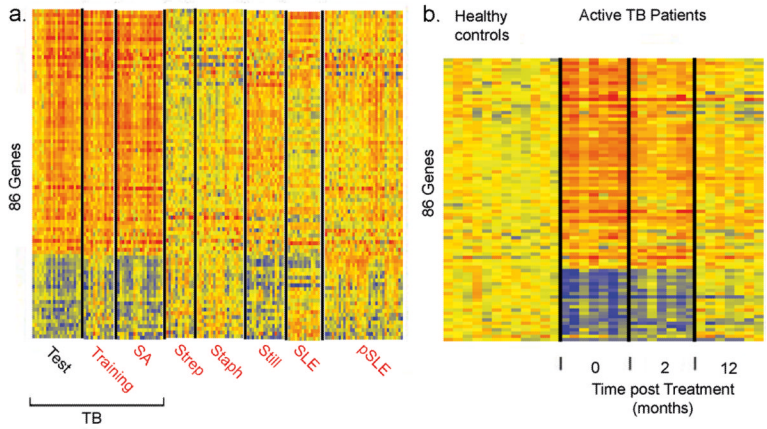

Figure 2. A distinct whole blood 86-gene transcriptional signature of active TB is distinct from other diseases

a, Comparison of 86-gene signature in patients with TB and other diseases normalized to their own controls; TB (Training, n=13; Control, n=12), TB (SA, n=20; Control = 12), Group A Streptococcus (Strep; n=23; Control=12), Staphylococcus (Staph; n=40; Control= 12), Still's disease (Still; n=31; Control=22), Adult (SLE; n=29; Control= 16) and paediatric SLE (pSLE; n=49; Control=11) patients. **b,** Expression levels of 86 gene signature after 2 and 12 months of treatment in TB patients.

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

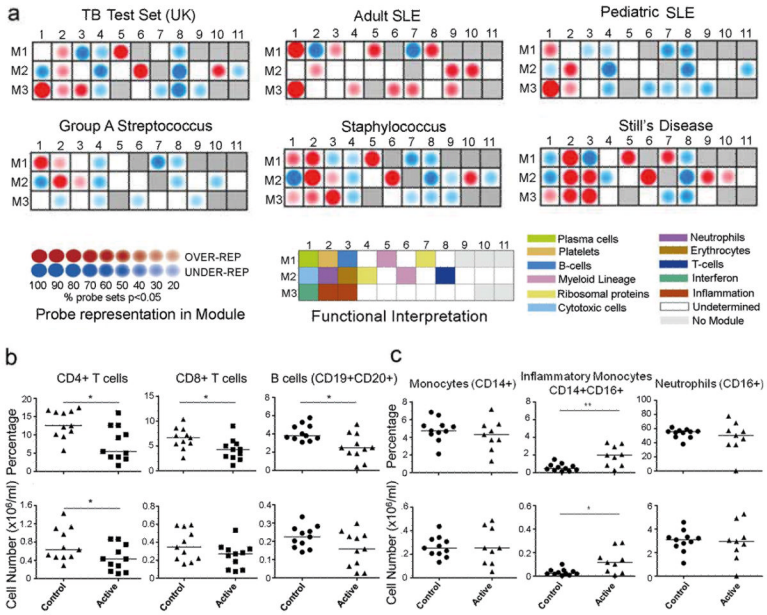

Figure 3. Whole blood transcriptional signature of active TB reflects distinct changes in cellular composition and gene expression

a, Gene expression (disease versus healthy controls) of TB (Test Set) and different diseases mapped within a pre-defined modular framework. Spot intensity (red = increased, blue = decreased) indicates transcript abundance. Functional interpretations previously determined by unbiased literature profiling shown by colour-coded grid. Whole blood (Test Set active TB patients and controls) analysed by flow cytometry for (**b**) $CD3^+CD4^+$ and $CD3^+CD8^+$ T cells and $CD19^+CD20^+$ B cells (**c**) $CD14^+$ monocytes, $CD14^+CD16^+$ inflammatory monocytes and $CD16^+$ neutrophils. Error bars = median, ** = p<0.01, * = p<0.05, Mann-Whitney test.

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

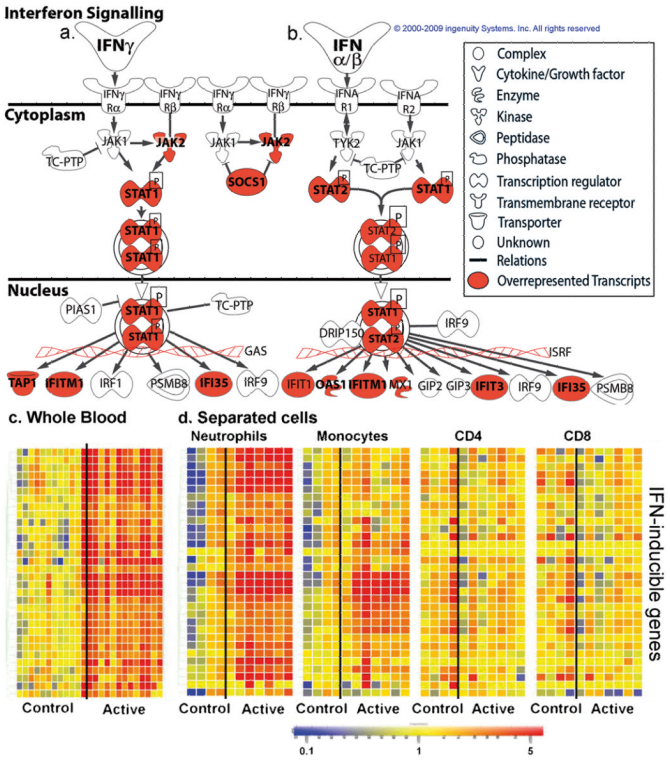

Figure 4. Interferon-inducible gene expression in active TB

Ingenuity Pathways analysis canonical pathway for interferon signalling symbol indicates gene function (legend on right). Transcripts over-represented in Test Set active TB patients shaded red **a,** Type II IFN-γ. **b,** Type I IFNαβ signalling. Transcript abundance of representative IFN-inducible genes in active TB (Test Set) from **c,** whole blood, **d,** separated blood leucocyte population. Transcript abundance/expression is normalised to the median of the healthy controls.

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts