

# A 380-gene meta-signature of active tuberculosis compared with healthy controls

*To the Editor:*

*Mycobacterium tuberculosis* is estimated to have infected one third of the world's population and continues to be a significant cause of mortality and morbidity [1]. There is a need for new and improved diagnostics or treatment-monitoring tools and blood-based mRNA diagnostics are a potential solution [2]. Gene expression microarray analysis of human blood has been widely used to profile the host transcriptional response in active tuberculosis (TB) to identify potential biomarkers and better understand the host immune response [2]. So far, there has been a relative lack of concordance in the actual genes being identified from the published studies [2, 3], although there has been agreement in some of the pathways identified. Interferon (IFN) signalling has been identified as a dominant signature in many of the individual studies [2, 4]; however, when significant gene lists were combined from eight publicly available TB datasets, TREM1 (triggering receptor expressed on myeloid cells 1) signalling became the most significant pathway [5].

We collectively reanalysed the publicly available datasets using differing methodologies to identify robustly differentially expressed genes that could distinguish active TB from controls. These genes are potential candidates for blood-based mRNA biomarkers of active disease and could provide valuable information regarding the immune and inflammatory response underlying TB pathogenesis.

We undertook a comprehensive search of PubMed and microarray depositories. Publicly available datasets that had active TB patients *versus* healthy controls, latently infected or patients post-treatment were identified and retained. The latter three cohorts are synonymous transcriptionally at the group level [4, 6, 7]. HIV-infected individuals were excluded from this analysis.

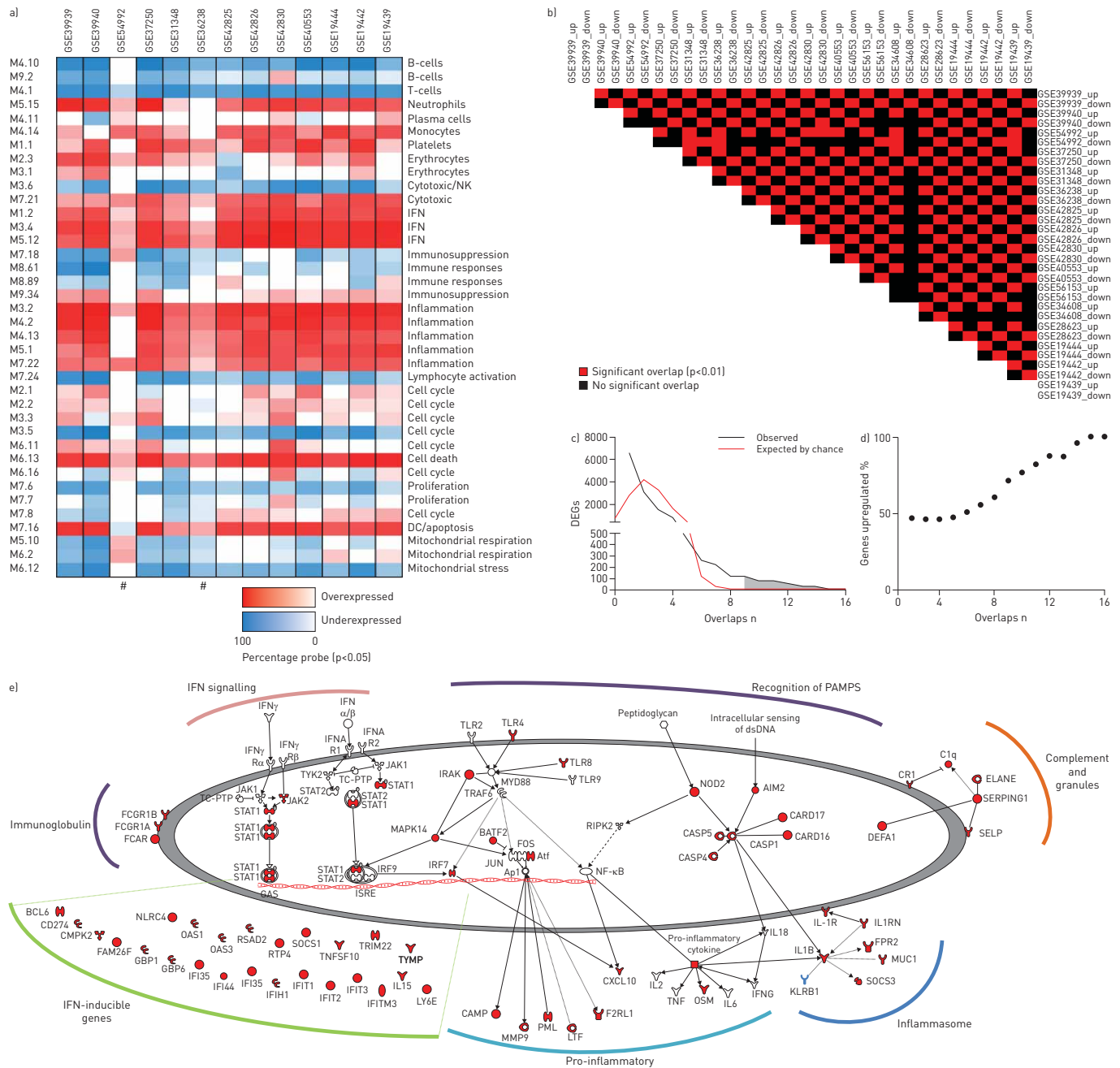
Where possible, data were imported in their raw format, Illumina and Agilent data were 75th centile normalised, and Affymetrix data were RMA (robust multichip average) quantile normalised. If raw data were not available then the authors' normalisation was used. All datasets were then filtered for low-expression transcripts (transcripts with expression 2 fold-change from the median in  $\geq 10\%$  of all samples retained) followed by statistical filtering (independent t-test with Benjamini-Hochberg multiple testing correction q-value  $< 0.05$ ). Probe/transcript IDs were matched to Entrez gene identifiers for each dataset. Multiple represented genes were filtered and the most significant (by q-value) retained. Venn Mapping [8] was used to check significance of the overlaps between any two datasets. Meta-profiling [9] of the significant gene lists was undertaken to identify the number of overlaps required for inclusion in the meta-signature. Only those genes that were expressed in a consistent direction of regulation for at least the number of determined overlaps were retained as the meta-signature.

Modular analysis [10] was undertaken of compatible datasets; it was not possible to analyse GSE56153, GSE34608 and GSE28623 with this method as their technology platforms were not supported by the tool. Canonical pathway, gene network analysis, gene function annotation and upstream analyses were generated through the use of IPA (Ingenuity Pathway Analysis; Ingenuity Systems, Qiagen, Redwood City, CA, USA).

16 datasets were included in the meta-analysis (GSE39939, GSE39940, GSE54992, GSE37250, GSE31348, GSE36238, GSE42825, GSE42826, GSE42830, GSE40553, GSE56153, GSE34608, GSE28623, GSE19444, GSE19442 and GSE19439). Modular analysis of these datasets revealed similarities, with overexpression of modules annotated as cytotoxic, IFN, inflammation and dendritic cells/apoptosis, and underexpression of modules annotated as B-cells, T-cells, lymphocyte activation and mitochondrial stress (figure 1a). Two datasets with the smallest cohort sizes (GSE54992 and GSE36238) had different weaker modular patterns with fewer modules identified as significantly different from the control group.

Independently for each dataset, differentially expressed genes between controls and TB groups were identified. There was significant overlap between the differentially expressed gene lists across the datasets, with no difference in degree of overlap dependent on the choice of control group (figure 1b). Meta-profiling identified 380 genes that were identified in nine or more datasets in a consistent direction of regulation (figure 1c). Upregulated genes were more consistently identified across datasets than downregulated genes (figure 1d). Five genes were identified in all 16 datasets: *AIM2*, *BATF2*, *FCGR1B*, *HP* and *TLR5*.

IFN- $\gamma$  was the top predicted upstream regulator of these 380 meta-signature genes, with 54 genes directly or indirectly linked to IFN- $\gamma$  within the IPA database (data not shown). The 380-gene meta-signature had



**FIGURE 1** Modular and meta-profiling identifying common transcriptional response. **a)** Modular analysis of 13 publicly available and compatible datasets of tuberculosis (TB) compared with control groups. 38 annotated modules are displayed as a heatmap, with red indicating significant over-abundance of transcripts and blue indicating significant under-abundance ( $p < 0.05$ ). The colour intensity represents the percentage of genes in that module that are significantly differentially expressed. NK: natural killer; IFN: interferon; DC: dendritic cell. #: two datasets with different modular profiles. **b)** Venn Mapper was used to identify significance of overlap between any two differentially expressed gene lists. Significance of overlap between significant gene lists was calculated by the Venn Mapper programme, with methodology adapted from SMID *et al.* [8]. Significant genes were identified following filtering of probes for low expression (probes were eliminated if expression had  $< 2$  fold-change from median normalised value in  $\geq 10\%$  of samples), followed by statistical filtering (independent t-test with Benjamini-Hochberg multiple testing correction between TB and control group). Probes were then matched to Entrez gene IDs, which was used as the reference "array" for analysis. The fold-change representation for each gene was the mean fold-change of TB group compared with control group; where genes were multiply represented on an array the fold-change associated with the most significant q-value was chosen. **c)** Differentially expressed genes (DEGs) were identified for each of the 16 datasets (probes filtered for low expression, followed by statistical filtering (independent t-test with Benjamini-Hochberg multiple testing correction between TB and control group)) and then meta-profiling was used to simulate the data and identify the number of overlaps, to define the meta-signature (shaded grey). **d)** Genes were grouped by the number of datasets in which they were significantly identified, and the percentage that were upregulated was calculated [direction of regulation relative to control group; where there were inconsistencies in direction of regulation across datasets the direction most often observed was used to determine direction]. **e)** Curated cartoon of the 380-gene meta-signature generated using IPA (Ingenuity Pathway Analysis; Ingenuity Systems, Qiagen, Redwood City, CA, USA) and IPA knowledge base. Red: significantly upregulated in the 380-gene meta-signature; blue: significantly downregulated; white: not significantly identified, included to provide context. PAMPs: pathogen-associated molecular patterns.

enrichment for IPA canonical pathways involved in pattern recognition, IFN signalling, interleukin-6 signalling, TREM1 signalling and complement (data not shown). Based on these findings, a curated cartoon summarising the major functional groups of genes and their relationships was created (figure 1e).

In this study, we have identified a 380-gene meta-signature of active TB compared with healthy controls, patients post-treatment and asymptomatic latently infected individuals, which showed enrichment for both innate and adaptive immune functions.

Two main methodologies were used to analyse the publicly available data: modular analysis and meta-profiling. Modular analysis depends on identifying differences in coordinately expressed groups of genes (modules) rather than individual genes [10]. We identified remarkable similarity between the datasets, with overexpression of modules annotated with IFN, inflammation functions, monocyte and neutrophil functions, and underexpression of B- and T-cell modules. These findings are in keeping with the individual studies that have included grouped modular analysis [4, 7]. Where grouped modular profiles were less consistent, this may have resulted from the small cohort sizes used and emphasises the need for individual studies to be appropriately powered to detect all differentially expressed genes.

Using meta-profiling we found that 380 genes were consistently differentially expressed in nine or more datasets, with five genes identified as differentially expressed in all 16 datasets. These five genes were *AIM2*, *BATF2*, *FCGR1B*, *TLR5* and *HP*, which have been shown to potentially play a role following *M. tuberculosis* infection [11–14]. A role for *TLR5* has yet to be described and identification of such differentially regulated genes may therefore be part of a programmatic response rather than specifically targeted for a tailored response to the pathogen.

Analysis of the 380 genes comprising the meta-signature identified IFN- $\gamma$  as the most significant potential upstream regulated molecule, with a large network of IFN- $\gamma$ -regulated genes present within the 380 genes. IFN- $\gamma$  is critical for control of mycobacterial disease in humans, with mutations either in the IFN- $\gamma$  receptor or STAT1 (signal transducer and activator of transcription 1) resulting in increased susceptibility [15]. However, upregulation of gene expression molecules downstream of type I IFN signalling was also observed, which is of relevance to TB exacerbation since type I IFN has been shown to antagonise signalling downstream of IFN- $\gamma$  [15]. Thus, capturing the overall picture of significant enrichment may be more informative than identification of one individual pathway, as shown in the summary cartoon (figure 1e). A number of immune pathways/functions are enriched for within the meta-signature, including multiple pattern recognition receptors, cytokines, the inflammasome, complement and immunoglobulin. This supports diverse findings obtained from both mouse and human studies that the immune response following *M. tuberculosis* infection is complex and can be cross-regulatory [15].

This study confirms the reproducibility of blood-based transcriptional analysis to identify the innate and adaptive host response in TB. It also identifies that upregulated mRNA transcripts are more reliably identified and highlights several mRNA candidates that could collectively be used as potential biomarkers of active disease. These findings have implications for the design and implementation of mRNA expression tools to support diagnostics and treatment monitoring of TB.



@ERSpublications

**Modular and meta-profiling identify a common transcriptional response of patients with TB versus healthy controls** <http://ow.ly/YvEP7>

**Simon Blankley<sup>1</sup>, Christine M. Graham<sup>1</sup>, Joe Levin<sup>1</sup>, Jacob Turner<sup>2</sup>, Matthew P.R. Berry<sup>1,3</sup>, Chloe I. Bloom<sup>1</sup>, Zhaohui Xu<sup>4</sup>, Virginia Pascual<sup>4</sup>, Jacques Banchereau<sup>5</sup>, Damien Chaussabel<sup>6</sup>, Ronan Breen<sup>7</sup>, George Santis<sup>7,8</sup>, Derek M. Blankenship<sup>2</sup>, Marc Lipman<sup>9,10</sup> and Anne O'Garra<sup>1,11</sup>**

<sup>1</sup>Laboratory of Immunoregulation and Infection, The Francis Crick Institute, Mill Hill Laboratory (formerly The MRC National Institute for Medical Research), London, UK. <sup>2</sup>Baylor Research Institute, Baylor Scott & White, Dallas, TX, USA.

<sup>3</sup>Dept of Respiratory Medicine, Imperial College Healthcare NHS Trust, St Mary's Hospital, London, UK. <sup>4</sup>Baylor Institute for Immunology Research/ANRS Center for Human Vaccines, INSERM, Dallas, TX, USA. <sup>5</sup>The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA. <sup>6</sup>Sidra Medical and Research Center, Doha, Qatar. <sup>7</sup>Dept of Respiratory Medicine, King's College London, London, UK. <sup>8</sup>Division of Asthma, Allergy and Lung Biology, King's College London, London, UK. <sup>9</sup>Dept of Respiratory Medicine, Royal Free London NHS Foundation Trust, London, UK. <sup>10</sup>Division of Medicine, University College London, London, UK. <sup>11</sup>Dept of Medicine, NHLI, Imperial College, London, UK.

Correspondence: Anne O'Garra, Laboratory of Immunoregulation and Infection, The Francis Crick Institute, Mill Hill Laboratory, The Ridgeway, Mill Hill, London, NW7 1AA, UK. E-mail: [anne.ogarra@crick.ac.uk](mailto:anne.ogarra@crick.ac.uk)

Received: Sept 20 2015 | Accepted after revision: Feb 10 2016 | First published online: April 13 2016

Support statement: S. Blankley and A. O'Garra were supported by the Medical Research Council (MRC), UK (grant U117565642), now The Francis Crick (A. O'Garra: Crick Budget 10126). S. Blankley was jointly funded by the UK

MRC and the UK Dept for International Development (DFID) under the MRC/DFID Concordat agreement (grant MR/J010723/1). G. Santis was supported in part by the Dept of Health via the National Institutes of 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. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Funding information for this article has been deposited with FundRef.

Conflict of interest: Disclosures can be found alongside the online version of this article at [erj.ersjournals.com](http://erj.ersjournals.com)

## References

- 1 World Health Organization. Global tuberculosis report 2014. Geneva, World Health Organization, 2014. Available from: [http://apps.who.int/iris/bitstream/10665/137094/1/9789241564809\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/137094/1/9789241564809_eng.pdf)
- 2 Blankley S, Berry MP, Graham CM, *et al*. The application of transcriptional blood signatures to enhance our understanding of the host response to infection: the example of tuberculosis. *Philos Trans R Soc Lond B Biol Sci* 2014; 369: 20130427.
- 3 Kaforou M, Wright VJ, Oni T, *et al*. Detection of tuberculosis in HIV-infected and -uninfected African adults using whole blood RNA expression signatures: a case-control study. *PLoS Med* 2013; 10: e1001538.
- 4 Berry MP, Graham CM, McNab FW, *et al*. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature* 2010; 466: 973-977.
- 5 Joosten SA, Fletcher HA, Ottenhoff TH. A helicopter perspective on TB biomarkers: pathway and process based analysis of gene expression data provides new insight into TB pathogenesis. *PLoS One* 2013; 8: e73230.
- 6 Bloom CI, Graham CM, Berry MP, *et al*. Detectable changes in the blood transcriptome are present after two weeks of antituberculosis therapy. *PLoS One* 2012; 7: e46191.
- 7 Cliff JM, Lee JS, Constantinou N, *et al*. Distinct phases of blood gene expression pattern through tuberculosis treatment reflect modulation of the humoral immune response. *J Infect Dis* 2013; 207: 18-29.
- 8 Smid M, Dorsers LC, Jenster G. Venn Mapping: clustering of heterologous microarray data based on the number of co-occurring differentially expressed genes. *Bioinformatics* 2003; 19: 2065-2071.
- 9 Rhodes DR, Yu J, Shanker K, *et al*. Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. *Proc Natl Acad Sci USA* 2004; 101: 9309-9314.
- 10 Chaussabel D, Quinn C, Shen J, *et al*. A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. *Immunity* 2008; 29: 150-164.
- 11 Saiga H, Kitada S, Shimada Y, *et al*. Critical role of AIM2 in *Mycobacterium tuberculosis* infection. *Int Immunol* 2012; 24: 637-644.
- 12 Tussiwand R, Lee WL, Murphy TL, *et al*. Compensatory dendritic cell development mediated by BATF-IRF interactions. *Nature* 2012; 490: 502-507.
- 13 Jacobsen M, Repsilber D, Gutschmidt A, *et al*. Candidate biomarkers for discrimination between infection and disease caused by *Mycobacterium tuberculosis*. *J Mol Med* 2007; 85: 613-621.
- 14 Kasvosve I, Gomo ZA, Mvundura E, *et al*. Haptoglobin polymorphism and mortality in patients with tuberculosis. *Int J Tuberc Lung Dis* 2000; 4: 771-775.
- 15 O'Garra A, Redford PS, McNab FW, *et al*. The immune response in tuberculosis. *Annu Rev Immunol* 2013; 31: 475-527.

Eur Respir J 2016; 47: 1873-1876 | DOI: 10.1183/13993003.02121-2015 | Copyright ©ERS 2016

# Implementation of digital technology solutions for a lung health trial in rural Malawi

To the Editor:

There is a global epidemic of non-communicable disease including chronic lung disease [1-3]. In the case of chronic lung disease, this challenge plays out in the context of a high burden of communicable diseases and pulmonary tuberculosis in particular.

The World Health Organization Practical Approach to Lung Health (PAL) provides an approach to address the dual challenges of chronic lung disease and tuberculosis [4]. We designed the Triage Plus cluster randomised controlled trial to determine the effect of using an adaptation of the PAL strategy in combination with engagement of informal health providers on the case detection and treatment of chronic lung disease and tuberculosis in rural Malawi (trial registration number: PACTR201411000910192).

In this research letter, we describe the ways in which we used digital technology solutions to prepare for the Triage Plus trial including mapping the study area, obtaining a sample of participants and collecting and managing baseline data from study participants.



CrossMark