Supplementary data

A modular transcriptional signature identifies phenotypic heterogeneity of human tuberculosis infection

Akul Singhania1,#, Raman Verma2,#, Christine M. Graham1, Jo Lee2, Tran Trang3, Matthew Richardson2, Patrick Lecine3, Philippe Leissner3, Matthew P.R. Berry4, Robert J. Wilkinson5,6,7, Karine Kaiser8, Marc Rodrigue8, Gerrit Woltmann2, Pranabashis Haldar2,¶, Anne O’Garra1,9,¶

1 Laboratory of Immunoregulation and Infection, The Francis Crick Institute, London, United Kingdom.
2 Respiratory Biomedical Research Centre, Institute for Lung Health, Department of Infection Immunity and Inflammation, University of Leicester, Leicester, United Kingdom.
3 BIOASTER Microbiology Technology Institute, Lyon, France.
4 Department of Respiratory Medicine, Imperial College Healthcare NHS Trust, St Mary’s Hospital, London, United Kingdom.
5 Wellcome Centre for Infectious Diseases Research, Africa; Institute for Infectious Diseases and Molecular Medicine, University of Cape Town, Observatory 7925, South Africa.
6 Department of Medicine, Imperial College London, London, United Kingdom.
7 Tuberculosis Laboratory, The Francis Crick Institute, London, United Kingdom.
8 Medical Diagnostic Discovery Department, bioMérieux SA, Marcy l’Etoile, France.
9 National Heart and Lung Institute, Imperial College London, London, United Kingdom.

# contributed equally to this work.
¶ co-supervised the work equally.

Correspondence and requests for materials should be addressed to AOG (email: Anne.OGarra@crick.ac.uk)
Supplementary Figures

Supplementary Figure 1. Schematic diagram documenting the objectives undertaken in the study. The cohorts analysed as part of each objective are shown, with microarray and RNA-seq cohorts represented in different colours. Table numbers (for the patient numbers/clinical characteristics for each dataset), and figure numbers associated with each objective are shown.
Supplementary Figure 2. Whole-blood transcriptional gene signatures in TB using RNA-seq and validation of microarray findings. a) Samples from the test (Berry London) and validation (Berry South Africa) sets from Berry et al. 2010, originally profiled using microarray in the publication, were profiled in this present study using RNA-sequencing. b) Heatmaps depicting unsupervised hierarchical clustering of active TB (red), LTBI (black) and control samples (purple) using the microarray 393-transcript signature of TB from Berry et al. 2010 in RNA-seq samples from Berry London and Berry South Africa cohorts. c) Volcano plots depicting the 373-gene signature of TB derived independently using RNA-seq Berry London cohort. d) Outliers identified by unsupervised hierarchical clustering in the Berry London and Berry South Africa cohorts, and (e) Leicester cohort.
Supplementary Figure 3. Cellular deconvolution analysis in whole-blood RNA-seq TB samples. 

**a** Stacked bar plots depicting *in silico* cellular composition of whole-blood RNA-seq samples from Berry London, 

**b** Berry South Africa and 

**c** Leicester cohorts derived using CIBERSORT. Percent fractions for 11 representative cell types for each sample are shown, with colours representing the different cell types. 

**d** CIBERSORT fractions for CD4, CD8, B cells, Monocytes/Macrophages and Neutrophils compared between active TB, LTBI and controls in Berry London cohort. Mean values are shown as a line in each group. p-values were calculated using a one-way ANOVA, with Tukey multiple test correction. *, p≤0.05; **, p≤0.01; ***, p≤0.001.
Supplementary Figure 4. Differential expression of published gene signatures in TB compared to severe influenza infection over time. a Log2 fold changes in Berry London cohort (derived from active TB vs. controls; y-axis) compared to log2 fold changes derived from severe influenza vs. controls from Parnell et al.20112 (x-axis) for genes from the 16-gene signature from Zak et al. 20163, and (b) the 27-gene signature (TB vs. LTBI) and (c) 44-gene signature (TB vs. other diseases (OD)) from Kaforou et al. 20134. Colours represent the WGCNA module membership of the gene. Shapes represent significance associated with the fold changes (FDR p-value < 0.05) in either Berry London only (squares), respective dataset only (diamonds), both (circles) or neither (triangles).
Supplementary Figure 5. Differential expression of published gene signatures in TB compared to influenza A infection over time. a Log\(_2\) fold changes in Berry London cohort (derived from active TB vs. controls; y-axis) compared to log\(_2\) fold changes derived from influenza A vs. controls from Zhai et al. 2015\(^3\) (x-axis) for genes from the 16-gene signature from Zak et al.2016\(^3\), and (b) the 27-gene signature (TB vs. LTBI) and (c) 44-gene signature (TB vs. other diseases (OD)) from Kaforou et al.2013\(^2\). Colours represent the WGCNA module membership of the gene. Shapes represent significance associated with the fold changes (FDR p-value < 0.05) in either Berry London only (squares), respective dataset only (diamonds), both (circles) or neither (triangles).
Supplementary Figure 6. Development of a reduced TB-specific gene signature. a Representative TB modular profile depicting modules that were perturbed in TB. b Venn diagram depicting genes significantly differentially expressed in TB only (active TB vs. controls in Berry London cohort) and not in influenza (influenza vs. controls in Parnell et al. 2011, Zhai et al. 2015 and Herberg et al. 2013). c The 61-gene signature identified as predictive of active TB compared to LTBI using the Boruta algorithm. d The set of 20 genes with the highest GINI score (using random forest) in the Boruta 61-gene signature. Module membership of the 20-gene signature is also shown.
Supplementary Figure 7. Modular transcriptional profiles of TB contacts followed over time. Longitudinal modular profiles of IGRA-ve TB contacts who remained healthy (n=15), IGRA+ve TB contacts who remained healthy (n=16), and TB contacts who developed TB during the study (n=9). Enrichment scores derived using ssGSEA compared to the average enrichment scores of IGRA-ve controls are depicted, with red and blue indicating modules over- or under-expressed compared to controls. Colour intensity and size represent the degree of enrichment. The x-axis depicts the time in months of recruitment for the study from Baseline for each patient. For the contacts who developed TB during the study, the time point when the contact was diagnosed with active TB in the clinic is represented in red letters. Representative modular TB profiles depicting modules that were perturbed in TB are shown for visual comparison on either side of each modular figure.
Supplementary Figure 7.

IGRA\textsuperscript{ve} contacts

Months from baseline

Contact

12 24 31 33 39 47 70 81 134 135 207 103 *209* 041

IGRA\textsuperscript{ve} contacts

Months from baseline

Contact

189 239 166 178 *196* 238 237 *45* 96 199 *49* 132 170 35 185 40

TB contacts developing TB

Annotation

DNA binding
T cells
Inflammasome
T cells
Innate/Hemopoietic mediators
Ubiquilatation
Cell proliferation/Metabolism
Innate immunity/PRR/C
Coagulation/Granulocytes
Th2/Eosinophils/Mast cells
Immunoglobulins
Interferon/PRR\textsuperscript{*}
Cell cycle
Myeloid/C/Adhesion
B cells
Hematopoiesis
Extracellular matrix
PRR\textsuperscript{*}/Enzymes
Cell structures
Microtubule/Enzymes
NK & T cells
Ion channels/Cell cycle
Interferon/C/Myeloid

Module

Black
Blue
Brown
Cyan
Darkgreen
Darkred
Darkturquoise
Green
Lightcyan
Lightgreen
Lightyellow
Magenta
Midnightblue
Pink
Purple
Royalblue
Salmon
Tan
Turquoise

Enrichment score

TB profile

Months from baseline

Contact

087 002 227 245 348 278 258 266 294

*Pattern recognition receptors
\textsuperscript{*}Complement system
Supplementary Figure 8. Baseline visit scores in TB contacts followed over time. Scatter plots depicting modified Disease Risk scores at Baseline visit in IGRA\textsuperscript{ve} and IGRA\textsuperscript{+ve} TB contacts who remained healthy, and in TB contacts who developed TB during the study \((n=9)\) using the 20-gene TB-specific signature. The dispersion is expressed as median with 95\% CI. Outliers are indicated as empty circles (IGRA\textsuperscript{-ve}) or squares (IGRA\textsuperscript{+ve}) with a dot inside.
Supplementary Figure 9. Blood transcriptional profile of TB contacts followed over time using the Zak et al. 2016 signature. a Bar plots depicting the modified Disease Risk Scores using the Zak et al. 20161 16-gene signature, in TB contacts who remained IGRA\textsuperscript{ve} and did not develop TB (n=15), TB contacts who remained IGRA\textsuperscript{ve} and did not develop TB (n=16), and TB contacts who developed TB during the study (n=9). For TB contacts who developed TB during the study, the time point when the contact was diagnosed with active TB in the clinic is represented by a red bar. Baseline in the barplot is set at 1824.35, average of all Baseline time-point modified Disease Risk Scores from all IGRA\textsuperscript{ve} contacts (n=15).

Zak et al. 2016 16 gene signature in TB contacts followed over time
### Supplementary Table 1. Clinical characteristics of IGRA\textsuperscript{+ve} TB contacts, IGRA\textsuperscript{-ve} TB contacts and Active TB patients recruited in Leicester.

| DEMOGRAPHICS                  | ACTIVE TB  
| (n = 54\textsuperscript{*}) | TB CONTACTS |                  |
|------------------------------|-------------|-----------------|
|                              |             | IGRA Positive  |
|                              |             | IGRA Negative  |
| GENDER                       |             | (n = 50\textsuperscript{*}) | (n = 50) |
| Male (%)                     | 35 (65)     | 29 (58)        | 30 (60) |
| Female (%)                   | 19 (35)     | 21 (42)        | 20 (40) |
| AGE                          | 40 (± 15)   | 39 (± 15)      | 35 (± 16) |
| ETHNICITY                    |             |                 |
| Indian Subcontinent (%)      | 34 (63)     | 39 (78)        | 45 (90) |
| African (%)                  | 11 (20)     | 7 (14)         | 4 (8)   |
| White Caucasian (%)          | 9 (17)      | 4 (8)          | 1 (2)   |
| ORIGIN                       |             |                 |
| UK born (%)                  | 8 (15)      | 4 (8)          | 10 (20) |
| Foreign born (%)             | 46 (75)     | 46 (92)        | 40 (80) |
| DISEASE TYPE\textsuperscript{\$} |             |                 |
| Smear positive pulmonary (%) | 20 (37)     | 16 (32)        | 10 (20) |
| Smear negative pulmonary (%) | 24 (44)     | 21 (42)        | 0       |
| Non-pulmonary (%)            | 10 (19)     | 13 (26)        | 40 (80) |
| DURATION OF SYMPTOMS         |             | 2.13 (± 1.62)  |         |
| CONTACT TYPE                 |             |                 |
| Household, partner (%)       | 8 (16)      | 8 (16)         |         |
| Household, other (%)         | 36 (72)     | 39 (78)        |         |
| Non-household (%)            | 6 (12)      | 3 (6)          |         |

\*One patient removed as part of quality control
\#One patient not sequenced
\$For TB contacts disease type refers to the phenotype of TB in the index case to which the contact was exposed

SD – Standard deviation
Supplementary Table 2. TB gene signatures from Zak et al. 2016 and Kaforou et al. 2013.

| Genes     | IFN inducible (n=16) | Genes     | IFN inducible (n=18) | Genes     | IFN inducible (n=22) |
|-----------|----------------------|-----------|----------------------|-----------|----------------------|
| Zak et al. 2016 (n=16) | | Kaforou et al. 2013 TB vs. LTBI (n=27) | | Kaforou et al. 2013 TB vs. OD (n=44) | |
| ANKRD22 * | ANKRD22 * | AAK1 * | |
| APOL1 * | C1QB * | ALDH1A1 * | |
| BATF2 * | C1QC * | ALDH1A1 * | |
| ETV7 * | C4orf18 * | ARG1 * | |
| FCGR1A * | C5 * | BTN3A1 * | |
| FCGR1B * | CCR6 * | C19orf12 * | |
| GBP1 * | CD79A * | CALML4 * | |
| GBP2 * | CD79B * | CASC1 | |
| GBP4 * | CXC5 * | CD74 * | |
| GBP5 * | DUSP3 * | CEPLK | |
| SCARF1 * | FAM20A * | CREB5 * | |
| SEPT4 * | FCGR1A * | CYB561 | |
| SERPINC1 * | FCGR1B * | CYB561 | |
| STAT1 * | FCGR1B * | DUSP3 * | |
| TAP1 * | FCGR1C * | EB1 | |
| TRAFD1 * | FLVCR2 * | GBP6 * | |
| GAS6 * | GJA9 | | |
| GAS6 * | HLA-DPB1 * | | |
| GBP6 * | HM13 | | |
| GN57 | HM13 | | |
| LHFPL2 | HS.131077 | | |
| LOC728744 | HS.162734 | | |
| MPO | IMPA2 | | |
| S100A8 | LHFPL2 | | |
| SMARCD3 | LOC100133800 | | |
| VAMP5 | LOC196752 | | |
| ZNF296 * | LOC389386 | | |

*IFN inducible genes, as identified by the Interferome database v2.01
Supplementary Table 3. Numbers of subjects in the external datasets downloaded from GEO.

| Group                        | Days before diagnosis | Kaforou et al. 2013 (GSE37250) | Zak et al. 2016 (GSE79362) | Parnell et al. 2011 (GSE20346) | Zhai et al. 2015 (GSE68310) | Herberg et al. 2013 (GSE42026) | Suarez et al. 2015 (GSE60244) | Bloom et al. 2013 (GSE42834) |
|------------------------------|-----------------------|--------------------------------|-----------------------------|-----------------------------|-------------------------------|--------------------------------|--------------------------------|-------------------------------|
| Active TB (HIV−/+)           | 1 to 180              | 195                            | 19                          | 19                          | 19                            | 19                             | 71                             | 35                            |
| Latent TB (HIV+/+)           | 181 to 360            | 195                            | 29                          | 29                          | 29                            | 29                             | 19                             | 61                            |
| Severe Influenza             | 361 to 720            | 11                             | 11                          | 11                          | 11                            | 11                             | 11                             | 11                            |
| Non-progressors              | >720                  | 4                              | 4                           | 4                           | 4                             | 4                              | 4                              | 4                             |
| Bacterial Pneumonia          | Day 1                 | 6                              | 6                           | 6                           | 6                             | 6                              | 6                              | 6                             |
|                             | Day 2                 | Controls                       | 6                           | Controls                    | 6                             | Controls                      | Controls                      | Controls                      |
|                             | Day 3                 | 6                              | 6                           | 6                           | 6                             | 6                              | 6                              | 6                             |
|                             | Day 4                 | 6                              | 6                           | 6                           | 6                             | 6                              | 6                              | 6                             |
|                             | Day 5                 | 6                              | 6                           | 6                           | 6                             | 6                              | 6                              | 6                             |

Day 0: 4 days before diagnosis; Day 1-4: 4 days after diagnosis; H1N109: H1N1 virus; Influenza A: Influenza A virus; Controls: Healthy controls; Lung cancer: Lung cancer patients; Bacterial Pneumonia: Bacterial pneumonia; Sarcoidosis: Sarcoidosis patients; Coinfection: Coinfection patients; Malaria: Malaria patients.
Supplementary Table 4. Annotation of 23 modules of co-expressed genes derived using WGCNA from combined Berry London & South Africa cohorts.

| Module       | Number of genes | Annotation                                                                 |
|--------------|-----------------|-----------------------------------------------------------------------------|
| Black        | 203             | DNA-binding proteins, Ubiquitination                                         |
| Blue         | 902             | T-lymphocytes, Antigens                                                     |
| Brown        | 428             | Inflammasome pathway, IL-10 signaling, TLR signaling, Monocytes              |
| Cyan         | 121             | T-lymphocytes, Molecular conformation                                        |
| Darkgreen    | 27              | Lymphotoxin β Receptor Signalling, Endosomes                                 |
| Darkred      | 38              | Ubiquitination, Mutagenesis                                                  |
| Darkturquoise| 27              | Cell proliferation & Metabolism                                              |
| Green        | 267             | Chemokine signaling, Complement system, Neutrophils, Monocytes, Interleukins, Pattern recognition receptors |
| Greencyellow | 159             | Coagulation system, Granulocyte Adhesion and Diapedesis                      |
| Grey60       | 94              | Th2 pathway, Eosinophils, Mast cells                                         |
| Lightcyan    | 86              | Immunoglobulins                                                             |
| Lightgreen   | 86              | Interferon signaling, Interferon Type-I, Pattern recognition receptors       |
| Lightyellow  | 59              | Cell cycle, cytoskeleton proteins                                           |
| Magenta      | 165             | Glycoproteins/Antigens, Macrophages/Monocytes, Complement, Cell adhesion    |
| Midnightblue | 108             | B cell receptor signaling                                                   |
| Pink         | 167             | Hematopoiesis, B cell development, Ubiquitination, Hemoglobins              |
| Purple       | 165             | Extracellular matrix proteins, cell differentiation                         |
| Red          | 236             | Pattern recognition receptors, Enzymes and Coenzymes                        |
| Royalblue    | 38              | Cellular structures, Cell surface receptors, Membrane proteins              |
| Salmon       | 136             | Microtubule proteins, Regulatory elements, Enzymes and Coenzymes           |
| Tan          | 157             | Antigens and Receptors, Lymphocytes, Th1 and Th2 activation pathway, NK cell signaling |
| Turquoise    | 959             | Ion channels, cell cycle proteins                                           |
| Yellow       | 360             | Interferon signaling, Complement system, MHC & Antigen presentation, Dendritic cells/Macrophages |
Supplementary Table 5. Clinical characteristics of the TB contacts identified with TB during prospective observation in the Leicester Cohort. The table summarises longitudinal clinical observations recorded for each subject until TB diagnosis (marked in red on longitudinal time course).

| Study Number | 002 | 227 | 087 | 245 | 348 | 278 | 258 | 266 | 294 |
|--------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| **Clinical and Demographic Characteristics** |     |     |     |     |     |     |     |     |     |
| Gender       | Female | Female | Male | Male | Female | Male | Male | Male | Male |
| Ethnicity    | Indian Subcontinent | African | Indian Subcontinent | European | White Caucasian | White Caucasian | White Caucasian | White Caucasian | White Caucasian |
| Birth Origin | Foreign Born | Foreign Born | Foreign Born | Foreign Born | UK Born | UK Born | UK Born | UK Born | UK Born |
| **Index Disease Type** | Smear positive Pulmonary TB | Smear positive Pulmonary TB | Smear negative Pulmonary TB | Smear negative Pulmonary TB | Smear negative Pulmonary TB | Smear negative Pulmonary TB | Smear negative Pulmonary TB | Smear negative Pulmonary TB | Smear negative Pulmonary TB |
| **Type of Contact** | Partner | Household | Household | Household | Household | Partner | Partner | Household | Household |
| **Contact Disease Type** | Non-Pulmonary TB | Pulmonary TB | Pulmonary TB | Pulmonary TB | Pulmonary TB | Pulmonary TB | Pulmonary TB | Pulmonary TB | Pulmonary TB |
| **Category of TB Infection** | True progressor | True progressor | Subclinical active TB | Subclinical active TB | Active TB | Active TB | Active TB | Active TB | Active TB |
| **IGRA Result** | Positive | Positive | Positive | Positive | Positive | Positive | Positive | Positive | Negative |
| **Culture status** | Positive | Positive | Positive | Positive | Positive | Positive | Positive | Positive | Negative |
| **Culture positive sample TTPC (days)** | 21 | 14 | 6 | 21 | 20 | 20 | 8 | 12 | 21 |
| **Xpert MTB/RIF assay** | Negative | Positive | Positive | Positive | Positive | Negative | Positive | Positive | Positive |

| Longitudinal time course | B | 1 | 3 | 4 | B | 2.5 | 2.75 | B | 2.5 | 3.5 | 3.75 | B | 0.5 | B | 0.1 | B | 0.1 |
|---------------------------|---|---|---|---|---|-----|-----|---|-----|-----|-----|---|-----|---|-----|---|-----|
| Reported Symptoms         | X | ☐ | ☐ | X | ☐ | X | ☐ | ☐ | ☐ | ☐ | ☐ | X | ☐ | ☐ | ☐ | ☐ | ☐ |
| CXR Appearance            | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N |
| Other Clinical Information| Nil | Nil | Also sputum culture positive for S.Pneumoniae | Hatched antibiotics for suspected bacterial pneumonia | Nil | Nil | Nil | Nil | Nil | Nil | Nil | Nil | Nil | Nil | Nil | Nil | Nil |

* TTPC - Time to positive culture
^ BAL - Bronchoalveolar lavage
X - No reported symptoms at clinical assessment
☒ - Reported symptom compatible with active TB (including cough, fever and weight loss).
Each symbol refers to a single symptom type.
N - Normal chest X-ray (reported by Consultant thoracic radiologist)
‡ - Abnormal chest X-ray with TB associated features
References

1 Berry, M. P. et al. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature* **466**, 973-977, doi:10.1038/nature09247 (2010).

2 Parnell, G. et al. Aberrant cell cycle and apoptotic changes characterise severe influenza A infection—a meta-analysis of genomic signatures in circulating leukocytes. *PLOS One* **6**, e17186 (2011).

3 Zak, D. E. et al. A blood RNA signature for tuberculosis disease risk: a prospective cohort study. *The Lancet* **387**, 2312-2322 (2016).

4 Kaforou, M. et al. Detection of tuberculosis in HIV-infected and-uninfected African adults using whole blood RNA expression signatures: a case-control study. *PLOS Med* **10**, e1001538 (2013).

5 Zhai, Y. et al. Host transcriptional response to influenza and other acute respiratory viral infections—a prospective cohort study. *PLOS Pathog* **11**, e1004869 (2015).

6 Herberg, J. A. et al. Transcriptomic profiling in childhood H1N1/09 influenza reveals reduced expression of protein synthesis genes. *J Infect Dis* **208**, 1664-1668 (2013).