The value of transcriptomics to advance knowledge of the immune response and diagnosis in TB

Akul Singhania[1], Robert J. Wilkinson[2,3,4], Marc Rodrigue[5], Pranabashis Haldar[6], Anne O'Garra[1,7]

[1]Laboratory of Immunoregulation and Infection, The Francis Crick Institute, London NW1 1AT, United Kingdom.
[2]Laboratory of Tuberculosis, The Francis Crick Institute, London NW1 1AT, United Kingdom.
[3]Department of Medicine, Imperial College London, London W2 1PG, United Kingdom.
[4]Wellcome Centre for Infectious Diseases Research in Africa, University of Cape Town, Observatory 7925, Republic of South Africa.
[5]Medical Diagnostic Discovery Department, bioMérieux SA, 69280 Marcy l'Etoile, France.
[6]Respiratory Biomedical Research Centre, Institute for Lung Health, Department of Infection Immunity and Inflammation, University of Leicester, Leicester LE3 9QP, United Kingdom.
[7]National Heart and Lung Institute, Imperial College London, London W2 1PG, United Kingdom.

Correspondence should be addressed to AOG (email: Anne.OGarra@crick.ac.uk)

ORCID IDs: 0000-0001-9845-6134 (AOG); 0000-0002-6941-3618 (AS)

Abstract

Blood transcriptomics in tuberculosis have revealed an interferon-inducible signature that diminishes upon successful treatment, promising improved diagnostics and treatment monitoring, which are essential to eradicate tuberculosis. Sensitive radiography revealing lung abnormalities and blood transcriptomics have demonstrated heterogeneity in active tuberculosis patients and exposed asymptomatic latent individuals, suggesting a continuum of infection and immune states. Here, we describe the immune response to *M. tuberculosis* infection revealed using transcriptomics, and differences between clinical phenotypes of infection that may inform temporal changes in host immunity associated with evolving infection. We also review the diverse reduced blood transcriptional gene signatures that have been proposed for tuberculosis diagnosis and identification of at-risk asymptomatic individuals, and suggest novel approaches for developing such biomarkers for clinical use.

Tuberculosis (TB) remains a major health problem worldwide and is the leading cause of mortality from a single infectious agent, with 1.67 million reported deaths in 2016[1]. The complexity of the immune response upon airborne transmission of the causative agent *Mycobacterium tuberculosis* and during progressive disease remains poorly characterised and understood[2,3]. Confirmation of active TB is based on the combination of symptoms and pathology (radiographically or histologically identified), as well as microbiological evidence of infection in sputum, typically by culture, which can take up to 6 weeks, and/or a nucleic acid amplification test [1,4] (Table 1). However, a sputum sample from patients can be hard to obtain, and although bronchoalveolar lavage can be used as a substitute, this is prohibitive in countries with limited resources and difficult in children[4,5]. Furthermore, *M. tuberculosis* can disseminate from the lung and cause disease throughout the body. Thus, alternative tests are required to improve and support the diagnosis of TB.

It is estimated that one fourth of all individuals worldwide have been infected by *M. tuberculosis*. The majority of infected individuals generate an effective immune response to possibly eliminate or control the infection and remain clinically asymptomatic, termed latent TB infection (LTBI), which is not transmissible. A small proportion of about 5–15% of latent individuals, however, go on to develop active TB disease at some stage during their lifetime[1]. Current diagnosis for LTBI involves testing reactivity to mycobacterial antigens, determined by a tuberculin skin test (TST), or an *M. tuberculosis*-specific interferon-γ (IFN-γ) release assay (IGRA), which can demonstrate whether a T cell-mediated immune response has been elicited in response to the infection[6], but both tests have poor prognostic value. These tests cannot determine whether the infection has been cleared, whether the individual is controlling the infection or may have subclinical disease, or whether the individual will go on to develop active TB (Fig. 1). Thus, these methods incompletely capture the spectrum of infectious states observed after exposure to *M. tuberculosis* infection.

Heterogeneity of LTBI was recognised by epidemiological differences in the risk of TB between recent and remote infection[7]. In a cohort of 35 asymptomatic LTBI individuals with HIV-1 co-infection, combined positron emission and computed tomography (PET-CT) identified ten individuals with pulmonary abnormalities suggestive of subclinical active disease who were substantially more likely to progress to clinical disease[8]. These findings challenge the classical view that divides TB into two states -latent infection or active disease- and urge the identification of biomarkers predictive of progression[8]. Progression from LTBI to active TB disease can be clinically subtle and individuals with subclinical TB have been reported to transmit the organism to others[9]. Earlier identification of active TB in individuals with undiagnosed disease is needed to initiate early treatment essential to limit onward transmission. A means of screening high-risk populations to identify people with early disease or those with latent infection at high-risk of developing TB is essential for early application of prophylactic therapy to prevent TB.

The dynamic relationships that exist between the proposed states of latent TB or subclinical TB and immune factors that influence possible transition between states are not known. Although protective factors have been described, including

IL-12, IFN-γ and TNF, our understanding of the early phase of *M. tuberculosis* infection or progression to disease in humans is very limited [2,10-13,14,15]. Risk factors responsible for a large proportion of TB cases in the general population[3] include HIV-coinfection[16], anti-TNF therapy[15,17], vitamin D deficiency[18], protein energy malnutrition[19], pregnancy[20] and intercurrent viral infections[21,22]. A better understanding of the early immune response to *M. tuberculosis* infection in individuals who control the infection after recent contact, remain sublclinical or go on to develop disease would greatly advance the development of improved diagnostics.

Blood transcriptomics elucidate the host response in TB

Blood transcriptomic profiling has provided an unbiased analysis and comprehensive overview of host factors perturbed upon infection and in active TB. A transcriptional signature dominated by IFN-inducible genes was identified in the whole blood of active TB patients, but not in healthy controls and the majority of individuals with LTBI[23]. This IFN-inducible gene signature included genes downstream of both IFN-γ and type I IFN, and was diminished upon successful treatment[23,24]. This transcriptomic signature has been confirmed in several studies worldwide with independent clinical cohorts[24-34], and in meta-analyses combining several of these cohorts[35-38]. An under-abundance of a type II IFN response in the transcriptional blood signature in TB patients, with downregulation of *IFNG* and *TBX21* has also been found[38].

Type I IFN has a deleterious effect in the control of TB in mouse models[2,21,39-45], consistent with a correlation between an IFN-inducible transcriptional signature in the blood and radiographic lung disease in human TB[23] and in non-human primate models[46]. Production of type I IFN by macrophages infected with different strains of *M. tuberculosis* can result from differential activation of the pattern recognition receptors TLR2 or TLR4 and its downstream adaptor protein TRIF[47]. The cytosolic DNA sensor cGAS has a central role in the detection of mycobacterial DNA[48-50] or mitochondrial DNA[51] released in the host cytosol, and the induction of type I IFN transcription in macrophages. Although *M. tuberculosis* can induce type I IFN in macrophages by these diverse pathways, various studies[22,43,52,39-42,45,53] have shown that elevated type I IFN, resulting from either virulent strains of *M. tuberculosis*[40,41,47], genetic deletion of type I IFN-regulatory genes such as *tpl-2*[43] or adjuvant[42,53] or viral coinfection[22], are required to induce detrimental effects to the host upon *M. tuberculosis* infection (Fig. 2). High amounts of type I IFN could result from differences in the genetic background[39-41,54], the mycobacterial challenge dose or strain[40,41,47] or microbiome composition[55,56,2]. An association between impaired type I IFN signaling and increased resistance to TB has been reported[57]. Patients with an inherited deficiency in the gene encoding ISG15[58] are more susceptible to mycobacterial infections[59], although there is some debate as to whether it is the increase in type I IFN that is responsible for the susceptibility to TB.

Various mechanisms account for the adverse effects of type I IFN in TB (reviewed in[44]), including inhibition of IL-1 and prostaglandin E2 (PGE2), which are critical for the defence against *M. tuberculosis* infection[42,52,60,61]. The adverse effects of type I

IFN on TB may be explained by the induction of IL-10, which suppresses the production of proinflammatory cytokines required for TB control[52,62]. Elevated IL-10 observed in mouse models of TB and human disease[2] contribute to increased bacterial loads[2,12,63,64,65]. It is tempting to speculate that blockade of IFNαβR signaling, which is currently in clinical trials for autoimmunity[66] could be applied in conjunction with anti-mycobacterial drugs to reduce high expression of type I IFN in the treatment of TB, especially in individuals with very severe disease and/or multi-drug resistance TB. The use of biologics as immunemodulators is supported by findings that individuals with mutations in *IL12RB or IFNG*[14] have been successfully treated with a combination of an anti-mycobacterial drugs and/or IFN-γ or IL-12 respectively.

In some cases, type I IFN may have a protective role against mycobacterial diseases[67,68], indicating context-specificity in the pathogenesis of TB[44]. Low amounts of type I IFN are required for the production of IL-12 and TNF[52], suggesting that low amounts of type I IFN may be protective against TB in the context of low *M. tuberculosis* burden. Conversely, high and sustained type I IFN signaling, potentially resulting from different genetic or context-specific effects, including coinfection, may contribute to TB pathogenesis, in part by induction of IL-10 and blockade of the protective factors required to control the mycobacterial infection (Fig. 2).

Blood transcriptomics reveal heterogeneity in LTBI and TB progression

A longitudinal transcriptomic analysis in cynomolgus macaques[46], which recapitulate the spectrum of clinical outcomes observed in human TB[69,70], reported increased transcriptional activity in innate and adaptive pathways early during infection, including an IFN signature. The blood transcriptome correlated with lung inflammation, as measured by PET-CT at early time points post-infection, and with the extent of disease[46,71].

Blood transcriptomics of latent individuals who have pulmonary abnormalities suggestive of subclinical active disease and are co-infected with HIV, identified an over-abundance of the classical complement pathway and Fcγ receptor 1, and increased amounts of circulating immune complexes in individuals with evidence of subclinical disease[8,72]. The increased expression of classical complement components in TB may be a response to increased production of immune complexes at the site of disease and may allow the inhibition of immune complex precipitation by C1q to minimize lung damage [8,72]. The increase in complement components was also observed in a cohort of 6,363 healthy adolescents that were followed for 24 months or more[73]. Individuals who ultimately developed microbiologically-confirmed TB disease more than 6 months after enrolment (n=44)[34] were compared to 106 matched controls who remained healthy during two years of follow up. Transcriptomic analysis of blood collected every 6 months until diagnosis showed a sequential modulation of immunological processes that preceded the manifestation of TB and subsequent clinical diagnosis[34]. Type I and II IFN signaling, and genes involved in the complement cascade, were observed up to 18 months before diagnosis, while changes in other inflammatory genes were observed closer to disease manifestation[34]. However, reinfection is prevalent in

high TB incidence countries[74-81] making it challenging to separate processes arising as a result of reactivation of infection from those caused by reinfection. Independent reanalysis of the same dataset suggested heterogeneity of the complement and Fcγ-receptor genes at an individual level[72]. Collectively, these studies[72,34] suggested that there may be a state consistent with subclinical TB, consisting of a specific increase in IFN response genes and activation of the complement cascade, which can be revealed in blood in individuals with no other signs of disease. Both studies restricted their analysis to $IGRA^+$ LTBI, assuming $IGRA^-$ individuals do not have latent infection. IGRAs have an overall sensitivity of approximately 85% in microbiologically-confirmed active TB, indicating that a proportion of latent infections will be missed using this test alone[82].

Although high TB incidence settings have often been referred to as "real world" TB, TB in low incidence settings remains a burden on public health, and both settings need to be addressed in order to eradicate TB. There are clear differences in the priorities, needs and goals for TB control between high and low TB incidence settings (Supplementary Table 1). High-burden, low-income settings have fragile health service frameworks with scarce resources and limited availability of either standard or advanced diagnostics, which allows onward transmission of infection that perpetuates poor TB control. Consideration of TB prevention strategies will be complicated in very high-incidence settings by the high risk of re-infection. Low-burden, high-income settings have well-resourced health service frameworks and extensive access to diagnostic tools. A biomarker sampled from an easily accessible part of the body, that identifies latent infection at high risk of TB progression with sensitivity and specificity greater than IGRAs and TST would greatly advance earlier TB diagnosis. Biomarkers of TB risk may best be validated reliably in low incidence TB settings where the risk of re-infection is low, unless study design in high-burden countries verifies that disease did not arise from reinfection by comparing the *M. tuberculosis* sequence from the index TB case with that of the LTBI contact who seemingly reactivates TB.

A proportion of LTBI individuals across each of the cohorts from London, South Africa and Leicester cluster with active TB patients on the basis of their transcriptional blood signature, similar to that observed in active TB; such individuals were termed LTBI outliers[23,38]. Modular analysis indicated that genes co-expressed in LTBI outliers and active TB patients represented biological processes linked to the IFN response, complement system, myeloid and pattern recognition receptors genes[38]. In addition, a reduced abundance of *IFNG* and *TBX21* in these LTBI outliers suggested a host response evolving towards that of active TB[38]. Because these LTBI outliers represented static instances of latent infection, transcriptional profiles of individuals recently exposed to *M. tuberculosis* by contact with active TB patients who either remained healthy (n=31) or developed active TB disease (n=9) were evaluated over time in Leicester, a setting with minimal risk of reinfection. Most $IGRA^-$ individuals who remained healthy showed few perturbations in their modular transcriptional signature over time. A proportion of the $IGRA^+$ individuals had profiles similar to that observed in TB, although in most cases this was transient[38]. In contrast, a modular signature comparable to that of active TB was observed in the majority (67%) of those who progressed to TB before diagnosis[38]. The blood transcriptome thus provides a sensitive approach to characterise between-subject heterogeneity and within-

subject variability following TB exposure and provide the hypothesis of indicators of transition in the host immune response that signal progression of *M. tuberculosis* infection[38]. It also appears that early events after exposure, measured as patterns of dynamic change in the transcriptional immune response, may influence the fate of infection[38,46].

Host transcriptional gene signatures in the diagnosis of TB

A key advantage of developing the blood transcriptome as a biomarker for TB progression is the ease of blood testing. This is relevant for groups in which microbiological diagnosis is constrained by poor capability for sample acquisition including pulmonary TB associated with little or no sputum production, typically seen in early disease, prior to cavitation; extra-pulmonary TB, where microbiological diagnosis requires examination of samples from the infected tissue site using invasive procedures; paediatric TB, which is paucibacillary and minimally productive of sputum; and HIV associated TB, where pathology leading to sputum production is diminished.

Use of transcriptomics as diagnostics specific for TB relies on the ability to identify commonalities and differences in the host response between TB and other infections and diseases[17,29,38,83-85]. Although TB and sarcoidosis patients show a big overlap in the blood transcriptome, by sharing IFN signalling and proinflammatory pathways[27,29,86], a subset of differentially regulated genes discriminated between the two pathologies, as well as between TB and lung cancer and pneumonia[23]. Two sets of IFN-inducible genes were also shared between TB and viral infections, albeit at different enrichment levels[38]. While enrichment of complement system and myeloid genes was greater in TB, the IFN-inducible gene set containing pattern recognition receptors and virally-induced genes was higher in viral infections[38]. Conversely, perturbations in cell proliferation, metabolism and haematopoiesis were observed in viral infections, but not in TB[38].

The development of gene signatures that could be used as diagnostic biomarkers for TB requires the definition of a small gene set with high diagnostic accuracy in multiplex testing. Currently there is no consensus on these diagnostic signatures. Studies have reported distinct sets of genes, each developed using standard machine learning algorithms, most with similar performance (Table 2 and Supplementary Table 2)[33,87-92]. These signatures cannot discriminate between TB and diseases such as pneumonia[93,94] and also identify acute viral infections[38]. This is a potential problem in children and some adults, where primary TB can present with clinical and radiological features often indistinguishable from respiratory viral illness[95,96]. In HIV-coinfected persons, TB frequently presents as a rapid onset of non-specific respiratory and systemic illness. Tuberculous meningitis, where the outcome critically depends on early intervention, requires an average of 3 health care practitioner visits before it is even suspected[97]. In the context of an LTBI screening programme, the prevalence of intercurrent viral illness at the time of testing may be significant, and will present a confounder, lowering the specificity of existing gene signatures for this purpose.

A 20-gene signature composed of genes perturbed in TB, but not in influenza was developed to circumvent these problems based on a modular approach (Fig. 3),

followed by machine-learning algorithms[38]. This 20-gene signature captures multiple biological pathways and is able to discriminate, albeit with lower sensitivity than less discriminant signatures, between TB and LTBI (Supplementary Table 2), and importantly, does not detect influenza, as an example of viral infections[38], and provides a proof of principle for new approaches to develop reduced signatures. The 20-gene signature was detected in the majority of individuals who progressed to TB in the Leicester cohort weeks or even months before clinical diagnosis[38], but was only minimally enriched in most $IGRA^-$ contacts and only transiently enriched in the $IGRA^+$ group who did not progress to disease. Other low number-gene signatures have been identified in asymptomatic LTBI individuals and patients with subclinical TB who progressed to active TB[75,100] (Table 2 and Supplementary Table 2). A 16-gene risk signature was evident up to six months before clinical presentation with disease in a South African adolescent cohort [73]. This 16-gene signature inadvertently detected influenza against healthy controls with high specificity and sensitivity[38]. In multiple sub-Saharan African cohorts of exposed, HIV-negative contacts, a 4-gene-transcript signature identified individuals at high-risk of developing TB up to two years before the onset of disease[98].

These findings suggest that there might be a trade-off between achieving a diagnostic TB signature with high sensitivity against LTBI, as well as high specificity against other diseases, and that alternative and complementary approaches, beyond machine algorithms should be considered for signature development. For example, applying a modular approach to inform gene expression changes across the global immune response, observed in TB, but not in LTBI or other potentially confounding diseases (Fig. 3), followed by machine-learning algorithms, to select the most discriminant genes across multiple differentially expressed modules, may allow identification of a more specific low-number gene signature. Pooling such a signature with a second signature, characterised by high sensitivity for TB detection, and applying combined yet discriminatory algorithms, could allow the development of a test to diagnose TB with greater confidence. Additional use of gene sets that detect and rule out confounding diseases, such as intercurrent viral infections, could be used to supplement these gene sets. The inclusion of the IFN-inducible genes that diminish upon successful treatment, as early as 2 weeks[23,24,30,99], may provide added clinical utility for determining optimal treatment duration. The diminished blood transcriptomic signature observed during successful TB treatment could also help in monitoring the response to treatment and in the development of new drugs, considering that current tests for monitoring drug efficacy, such as the early bactericidal assays and 2-month sputum conversion are both time-consuming and lack specificity, even when sputum can be obtained[99]. Such diagnostic biomarkers will need to be carefully tested in a multitude of TB cohorts from distinct geographical locations and optimised for specificity using cohorts of other infections. New molecular platforms with increasing capacity of multiplexing could be of help in facilitating the use of such tests in the clinic. Furthermore, it is anticipated that different contexts, goals and clinical applications in high-incidence, low-income countries, or low-incidence, high-income countries (Supplementary Table 1), will dictate the use of a transcriptomic based diagnostic or prognostic, in addition to a tool for monitoring drug treatment.

Conclusions and future perspectives

There is still limited understanding of the complete spectrum of infectious states evident in latently infected individuals. High sensitivity radiographic imaging together with blood transcriptomic signatures have revealed the heterogeneity of latent TB in both humans and non-human primate models. However the events that determine whether an exposed individual will control the infection or go on to develop TB are unknown. It is critical to understand the host response in the lung directly following exposure to *M. tuberculosis* infection to determine how this may influence the outcome of infection. This could be achieved using transcriptomic and complementary immunological approaches in well-defined and carefully curated clinical cohorts, longitudinally profiling blood as well as lung samples (e.g. bronchoalveolar lavage) from individuals exposed to TB. This will advance our knowledge of the local host immune response involved in the control of infection or progression to disease.

Transcriptomic approaches also show promise with respect to the development of biomarkers for diagnosis and prognosis of TB, and for drug treatment monitoring. Biomarker signatures for clinical use would need to be downsized to facilitate a multiplex type test, be rapid and automated, with a turnaround time of 2-3 hours, and inexpensive, to be feasible for implementation testing in a field or bedside setting. This would facilitate effective and early treatment which is essential for the eradication of TB.

References

1 World Health Organisation. Global Tuberculosis Report. (2017).
2 O'Garra, A. *et al.* The immune response in tuberculosis. Annu Rev Immunol 31, 475-527 (2013).
3 Pai, M. *et al.* Tuberculosis. Nat Rev Dis Primers 2, 16076 (2016).
4 Davies, P. D. & Pai, M. The diagnosis and misdiagnosis of tuberculosis. Int J Tuberc Lung Dis 12, 1226-1234 (2008).
5 Pfyffer, G. E., Cieslak, C., Welscher, H. M., Kissling, P. & Rusch-Gerdes, S. Rapid detection of mycobacteria in clinical specimens by using the automated BACTEC 9000 MB system and comparison with radiometric and solid-culture systems. J Clin Microbiol 35, 2229-2234 (1997).
6 Trajman, A., Steffen, R. E. & Menzies, D. Interferon-Gamma Release Assays versus Tuberculin Skin Testing for the Diagnosis of Latent Tuberculosis Infection: An Overview of the Evidence. Pulm Med 2013, 601737 (2013).
7 Barry, C. E., 3rd *et al.* The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. Nat Rev Microbiol 7, 845-855 (2009).
8 Esmail, H. *et al.* Characterization of progressive HIV-associated tuberculosis using 2-deoxy-2-[18F]fluoro-D-glucose positron emission and computed tomography. Nat Med 22, 1090-1093 (2016).
9 Dowdy, D. W., Basu, S. & Andrews, J. R. Is passive diagnosis enough? The impact of subclinical disease on diagnostic strategies for tuberculosis. Am J Respir Crit Care Med 187, 543-551 (2013).
10 Orme, I. M., Robinson, R. T. & Cooper, A. M. The balance between protective and pathogenic immune responses in the TB-infected lung. Nat Immunol 16, 57-63 (2015).
11 Cooper, A. M. Cell-mediated immune responses in tuberculosis. Annu Rev Immunol 27, 393-422 (2009).
12 Cooper, A. M., Mayer-Barber, K. D. & Sher, A. Role of innate cytokines in mycobacterial infection. Mucosal Immunol 4, 252-260 (2011).
13 Flynn, J. L. & Chan, J. Immunology of tuberculosis. Annu Rev Immunol 19, 93-129 (2001).
14 Casanova, J. L. & Abel, L. Genetic dissection of immunity to mycobacteria: the human model. Annu Rev Immunol 20, 581-620 (2002).
15 Keane, J. *et al.* Tuberculosis associated with infliximab, a tumor necrosis factor alpha-neutralizing agent. N Engl J Med 345, 1098-1104 (2001).
16 Esmail, H. *et al.* The Immune Response to Mycobacterium tuberculosis in HIV-1-Coinfected Persons. Annu Rev Immunol 36, 603-638 (2018).
17 Lin, P. L. *et al.* PET CT Identifies Reactivation Risk in Cynomolgus Macaques with Latent M. tuberculosis. PLoS Pathog 12, e1005739 (2016).
18 Martineau, A. R. Old wine in new bottles: vitamin D in the treatment and prevention of tuberculosis. Proc Nutr Soc 71, 84-89 (2012).
19 Dye, C. After 2015: infectious diseases in a new era of health and development. Philos Trans R Soc Lond B Biol Sci 369, 20130426 (2014).
20 Llewelyn, M., Cropley, I., Wilkinson, R. J. & Davidson, R. N. Tuberculosis diagnosed during pregnancy: a prospective study from London. Thorax 55, 129-132 (2000).
21 McNab, F., Mayer-Barber, K., Sher, A., Wack, A. & O'Garra, A. Type I interferons in infectious disease. Nat Rev Immunol 15, 87-103 (2015).

22 Redford, P. S. *et al.* Influenza A virus impairs control of Mycobacterium tuberculosis coinfection through a type I interferon receptor-dependent pathway. J Infect Dis 209, 270-274 (2014).
23 Berry, M. P. *et al.* An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. Nature 466, 973-977 (2010).
24 Bloom, C. I. *et al.* Detectable changes in the blood transcriptome are present after two weeks of antituberculosis therapy. PLoS One 7, e46191 (2012).
25 Maertzdorf, J. *et al.* Human gene expression profiles of susceptibility and resistance in tuberculosis. Genes Immun 12, 15-22 (2011).
26 Maertzdorf, J. *et al.* Functional correlations of pathogenesis-driven gene expression signatures in tuberculosis. PLoS One 6, e26938 (2011).
27 Maertzdorf, J. *et al.* Common patterns and disease-related signatures in tuberculosis and sarcoidosis. Proc Natl Acad Sci U S A 109, 7853-7858 (2012).
28 Ottenhoff, T. H. *et al.* Genome-wide expression profiling identifies type 1 interferon response pathways in active tuberculosis. PLOS One 7, e45839 (2012).
29 Bloom, C. I. *et al.* Transcriptional blood signatures distinguish pulmonary tuberculosis, pulmonary sarcoidosis, pneumonias and lung cancers. PLoS One 8, e70630 (2013).
30 Cliff, J. M. *et al.* Distinct phases of blood gene expression pattern through tuberculosis treatment reflect modulation of the humoral immune response. J Infect Dis 207, 18-29 (2013).
31 Berry, M. P., Blankley, S., Graham, C. M., Bloom, C. I. & O'Garra, A. Systems approaches to studying the immune response in tuberculosis. Curr Opin Immunol 25, 579-587 (2013).
32 Blankley, S. *et al.* The Transcriptional Signature of Active Tuberculosis Reflects Symptom Status in Extra-Pulmonary and Pulmonary Tuberculosis. PLoS One 11, e0162220 (2016).
33 Roe, J. K. *et al.* Blood transcriptomic diagnosis of pulmonary and extrapulmonary tuberculosis. JCI insight 1, e87238 (2016).
34 Scriba, T. J. *et al.* Sequential inflammatory processes define human progression from M. tuberculosis infection to tuberculosis disease. PLoS Pathog 13, e1006687 (2017).
35 Joosten, S. A., Fletcher, H. A. & Ottenhoff, T. H. A helicopter perspective on TB biomarkers: pathway and process based analysis of gene expression data provides new insight into TB pathogenesis. PLoS One 8, e73230 (2013).
36 Blankley, S. *et al.* A 380-gene meta-signature of active tuberculosis compared with healthy controls. Eur Respir J 47, 1873-1876 (2016).
37 Sambarey, A. *et al.* Meta-analysis of host response networks identifies a common core in tuberculosis. NPJ Syst Biol Appl 3, 4 (2017).
38 Singhania, A. *et al.* A modular transcriptional signature identifies phenotypic heterogeneity of human tuberculosis infection. Nature Communications 9 (2018).
39 Dorhoi, A. *et al.* Type I IFN signaling triggers immunopathology in tuberculosis-susceptible mice by modulating lung phagocyte dynamics. Eur J Immunol 44, 2380-2393 (2014).
40 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-a/b. Proc Natl Acad Sci U S A 98, 5752-5757 (2001).

41 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 25, 694-701 (2005).
42 Mayer-Barber, K. D. *et al.* Host-directed therapy of tuberculosis based on interleukin-1 and type I interferon crosstalk. Nature 511, 99-103 (2014).
43 McNab, F. W. *et al.* TPL-2-ERK1/2 signaling promotes host resistance against intracellular bacterial infection by negative regulation of type I IFN production. J Immunol 191, 1732-1743 (2013).
44 Moreira-Teixeira, L., Mayer-Barber, K., Sher, A. & O'Garra, A. Type I interferons in tuberculosis: Foe and occasionally friend. J Exp Med (2018).
45 Ordway, D. *et al.* The hypervirulent Mycobacterium tuberculosis strain HN878 induces a potent TH1 response followed by rapid down-regulation. J Immunol 179, 522-531 (2007).
46 Gideon, H. P., Skinner, J. A., Baldwin, N., Flynn, J. L. & Lin, P. L. Early Whole Blood Transcriptional Signatures Are Associated with Severity of Lung Inflammation in Cynomolgus Macaques with Mycobacterium tuberculosis Infection. J Immunol 197, 4817-4828 (2016).
47 Carmona, J. *et al.* Mycobacterium tuberculosis Strains Are Differentially Recognized by TLRs with an Impact on the Immune Response. PLoS One 8, e67277 (2013).
48 Collins, A. C. *et al.* Cyclic GMP-AMP Synthase Is an Innate Immune DNA Sensor for Mycobacterium tuberculosis. Cell Host Microbe 17, 820-828 (2015).
49 Wassermann, R. *et al.* Mycobacterium tuberculosis Differentially Activates cGAS- and Inflammasome-Dependent Intracellular Immune Responses through ESX-1. Cell Host Microbe 17, 799-810 (2015).
50 Watson, R. O. *et al.* The Cytosolic Sensor cGAS Detects Mycobacterium tuberculosis DNA to Induce Type I Interferons and Activate Autophagy. Cell Host Microbe 17, 811-819 (2015).
51 Wiens, K. E. & Ernst, J. D. The Mechanism for Type I Interferon Induction by Mycobacterium tuberculosis is Bacterial Strain-Dependent. PLoS Pathog 12, e1005809 (2016).
52 McNab, F. *et al.* Type I IFN induces IL-10 production in an IL-27-independent manner and blocks responsiveness to IFN-γ for production of IL-12 and bacterial killing in Mycobacterium tuberculosis-infected macrophages. J Immunol (2014).
53 Antonelli, L. R. *et al.* Intranasal Poly-IC treatment exacerbates tuberculosis in mice through the pulmonary recruitment of a pathogen-permissive monocyte/macrophage population. J Clin Invest 120, 1674-1682 (2010).
54 Domaszewska, T. *et al.* Concordant and discordant gene expression patterns in mouse strains identify best-fit animal model for human tuberculosis. Sci Rep 7, 12094 (2017).
55 Namasivayam, S. *et al.* Longitudinal profiling reveals a persistent intestinal dysbiosis triggered by conventional anti-tuberculosis therapy. Microbiome 5, 71 (2017).
56 Wipperman, M. F. *et al.* Antibiotic treatment for Tuberculosis induces a profound dysbiosis of the microbiome that persists long after therapy is completed. Sci Rep 7, 10767 (2017).
57 Zhang, G. *et al.* A proline deletion in IFNAR1 impairs IFN-signaling and underlies increased resistance to tuberculosis in humans. Nat Commun 9, 85 (2018).
58 Zhang, X. *et al.* Human intracellular ISG15 prevents interferon-alpha/beta over-amplification and auto-inflammation. Nature 517, 89-93 (2015).

59 Bogunovic, D. *et al.* Mycobacterial disease and impaired IFN-gamma immunity in humans with inherited ISG15 deficiency. Science 337, 1684-1688 (2012).
60 Chen, D. Y. *et al.* Single-chain antibody against human lipocalin-type prostaglandin D synthase: construction, expression, purification, and activity assay. Biochemistry (Mosc) 73, 702-710 (2008).
61 Divangahi, M., King, I. L. & Pernet, E. Alveolar macrophages and type I IFN in airway homeostasis and immunity. Trends Immunol 36, 307-314 (2015).
62 Moreira-Teixeira, L. *et al.* T Cell-Derived IL-10 Impairs Host Resistance to Mycobacterium tuberculosis Infection. J Immunol 199, 613-623 (2017).
63 Beamer, G. L. *et al.* Interleukin-10 promotes Mycobacterium tuberculosis disease progression in CBA/J mice. J Immunol 181, 5545-5550 (2008).
64 Redford, P. S., Murray, P. J. & O'Garra, A. The role of IL-10 in immune regulation during M. tuberculosis infection. Mucosal Immunol 4, 261-270 (2011).
65 Huynh, J. P. *et al.* Bhlhe40 is an essential repressor of IL-10 during Mycobacterium tuberculosis infection. J Exp Med (2018).
66 Furie, R. *et al.* Anifrolumab, an Anti-Interferon-alpha Receptor Monoclonal Antibody, in Moderate-to-Severe Systemic Lupus Erythematosus. Arthritis Rheumatol 69, 376-386 (2017).
67 Moreira-Teixeira, L. *et al.* Type I IFN Inhibits Alternative Macrophage Activation during Mycobacterium tuberculosis Infection and Leads to Enhanced Protection in the Absence of IFN-gamma Signaling. J Immunol 197, 4714-4726 (2016).
68 Ward, C. M. *et al.* Adjunctive treatment of disseminated Mycobacterium avium complex infection with interferon alpha-2b in a patient with complete interferon-gamma receptor R1 deficiency. Eur J Pediatr 166, 981-985 (2007).
69 Lin, P. L. *et al.* Quantitative comparison of active and latent tuberculosis in the cynomolgus macaque model. Infect Immun 77, 4631-4642 (2009).
70 Cadena, A. M., Fortune, S. M. & Flynn, J. L. Heterogeneity in tuberculosis. Nat Rev Immunol 17, 691-702 (2017).
71 Capuano, S. V., 3rd *et al.* Experimental Mycobacterium tuberculosis infection of cynomolgus macaques closely resembles the various manifestations of human M. tuberculosis infection. Infect Immun 71, 5831-5844 (2003).
72 Esmail, H. *et al.* Complement pathway gene activation and rising circulating immune complexes characterize early disease in HIV-associated tuberculosis. Proc Natl Acad Sci U S A 115, E964-E973 (2018).
73 Zak, D. E. *et al.* A blood RNA signature for tuberculosis disease risk: a prospective cohort study. The Lancet 387, 2312-2322 (2016).
74 Charalambous, S. *et al.* Contribution of reinfection to recurrent tuberculosis in South African gold miners. Int J Tuberc Lung Dis 12, 942-948 (2008).
75 Uys, P. *et al.* The Risk of Tuberculosis Reinfection Soon after Cure of a First Disease Episode Is Extremely High in a Hyperendemic Community. PLoS One 10, e0144487 (2015).
76 van Helden, P. D., Warren, R. M. & Uys, P. Predicting reinfection in tuberculosis. J Infect Dis 197, 172-173; author reply 173-174 (2008).
77 van Rie, A. *et al.* Reinfection and mixed infection cause changing Mycobacterium tuberculosis drug-resistance patterns. Am J Respir Crit Care Med 172, 636-642 (2005).
78 van Rie, A. *et al.* Exogenous reinfection as a cause of recurrent tuberculosis after curative treatment. N Engl J Med 341, 1174-1179 (1999).
79 van Rie, A. *et al.* Transmission of a multidrug-resistant Mycobacterium tuberculosis strain resembling "strain W" among noninstitutionalized, human

immunodeficiency virus-seronegative patients. J Infect Dis 180, 1608-1615 (1999).
80 Verver, S. *et al.* Rate of reinfection tuberculosis after successful treatment is higher than rate of new tuberculosis. Am J Respir Crit Care Med 171, 1430-1435 (2005).
81 Warren, R. M. *et al.* Patients with active tuberculosis often have different strains in the same sputum specimen. Am J Respir Crit Care Med 169, 610-614 (2004).
82 Diel, R., Loddenkemper, R. & Nienhaus, A. Evidence-based comparison of commercial interferon-gamma release assays for detecting active TB: a metaanalysis. Chest 137, 952-968 (2010).
83 Elkington, P., Tebruegge, M. & Mansour, S. Tuberculosis: An Infection-Initiated Autoimmune Disease? Trends Immunol 37, 815-818 (2016).
84 Clayton, K., Polak, M. E., Woelk, C. H. & Elkington, P. Gene Expression Signatures in Tuberculosis Have Greater Overlap with Autoimmune Diseases Than with Infectious Diseases. Am J Respir Crit Care Med 196, 655-656 (2017).
85 Mourik, B. C., Lubberts, E., de Steenwinkel, J. E. M., Ottenhoff, T. H. M. & Leenen, P. J. M. Interactions between Type 1 Interferons and the Th17 Response in Tuberculosis: Lessons Learned from Autoimmune Diseases. Front Immunol 8, 294 (2017).
86 Koth, L. L. *et al.* Sarcoidosis blood transcriptome reflects lung inflammation and overlaps with tuberculosis. Am J Respir Crit Care Med 184, 1153-1163 (2011).
87 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 medicine 10, e1001538 (2013).
88 Anderson, S. T. *et al.* Diagnosis of childhood tuberculosis and host RNA expression in Africa. N Engl J Med 370, 1712-1723 (2014).
89 Maertzdorf, J. *et al.* Concise gene signature for point-of-care classification of tuberculosis. EMBO Mol Med 8, 86-95 (2016).
90 Sweeney, T. E., Braviak, L., Tato, C. M. & Khatri, P. Genome-wide expression for diagnosis of pulmonary tuberculosis: a multicohort analysis. Lancet Respir Med 4, 213-224 (2016).
91 Leong, S. *et al.* Existing blood transcriptional classifiers accurately discriminate active tuberculosis from latent infection in individuals from south India. Tuberculosis (Edinb) 109, 41-51 (2018).
92 Pan, L. *et al.* Genome-wide transcriptional profiling identifies potential signatures in discriminating active tuberculosis from latent infection. Oncotarget 8, 112907-112916 (2017).
93 Walter, N. D. *et al.* Blood Transcriptional Biomarkers for Active Tuberculosis among Patients in the United States: a Case-Control Study with Systematic Cross-Classifier Evaluation. J Clin Microbiol 54, 274-282 (2016).
94 Walter, N. D., Reves, R. & Davis, J. L. Blood transcriptional signatures for tuberculosis diagnosis: a glass half-empty perspective. Lancet Respir Med 4, e28 (2016).
95 Cox, H. *et al.* Delays and loss to follow-up before treatment of drug-resistant tuberculosis following implementation of Xpert MTB/RIF in South Africa: A retrospective cohort study. PLoS Med 14, e1002238 (2017).
96 Cox, H. S. *et al.* The need to accelerate access to new drugs for multidrug-resistant tuberculosis. Bull World Health Organ 93, 491-497 (2015).

608 97 Bang, N. D. *et al.* Clinical presentations, diagnosis, mortality and prognostic
609 markers of tuberculous meningitis in Vietnamese children: a prospective
610 descriptive study. BMC Infect Dis 16, 573 (2016).
611 98 Suliman, S. *et al.* Four-gene Pan-African Blood Signature Predicts Progression to
612 Tuberculosis. Am J Respir Crit Care Med (2018).
613 99 Cliff, J. M., Kaufmann, S. H., McShane, H., van Helden, P. & O'Garra, A. The
614 human immune response to tuberculosis and its treatment: a view from the
615 blood. Immunol Rev 264, 88-102 (2015).

616

Acknowledgements

AOG and AS were funded by The Francis Crick Institute, (Crick 10126; Crick 10468), which receives its core funding from Cancer Research UK, the U.K. Medical Research Council, and the Wellcome Trust. RJW was supported by The Francis Crick Institute, (Crick 10128), which receives its core funding from Cancer Research UK, the U.K. Medical Research Council, and Wellcome; by Wellcome (104803; 203135); MRC South Africa under strategic health innovation partnerships EDCTP SR1A 2015-1065; and NIH 019 AI 111276 and UO1AI115940. Recent work contributing to this review was from the NIHR Leicester Biomedical Research Centre for their support of the study at Leicester; BIOASTER Microbiology Technology Institute, Lyon, France; Medical Diagnostic Discovery Department, bioMérieux SA, Marcy l'Etoile, France; with BIOASTER investment, receiving funding from the French Government through the Investissement d'Avenir program (Grant NO. ANR-10-AIRT-03).. The views expressed are those of the author(s) and not necessarily those of the NHS the NIHR or the Department of Health.

Competing interests

The authors declare no competing interests and note that previous patents held by Anne O'Garra on the use of the blood transcriptomic for diagnosis of tuberculosis have lapsed and discontinued. Marc Rodrigue is an employee of BioMérieux. BioMérieux has not filed patents related to this study. Furthermore, the authors also confirm that this does not alter their adherence to all the Nature Immunology's policies.

Figure legends

Figure 1. Heterogeneity in outcomes upon exposure to *M. tuberculosis.* Upon contact with an active TB patient (red), an individual with recent exposure to *M. tuberculosis* (grey) can manifest a range of infectious states. The majority of the exposed individuals will remain asymptomatic with the possible scenarios: remain uninfected or eliminate the bacteria (purple); become infected but control the bacteria either by innate immune responses (purple) or by *M. tuberculosis* antigen-specific T cell response as detected by the IGRA test (gradation from purple to black); develop subclinical TB and show pulmonary abnormalities by advanced radiographic approaches and a transient blood signature (black). A small proportion of exposed individuals will progress to active TB (red) and further represent a spectrum of infection states based on the *M. tuberculosis* load as measured in sputum by a smear test (indicative of high bacterial load); *M. tuberculosis* culture or nucleic acid amplification test; or if negative in sputum, measured in BAL, when possible (indicative of lower bacterial load) and may manifest different degrees of symptoms (different degrees of red). Adapted from Pai et al., 2016 (Ref. 3)

Figure 2. The immune response to *M. tuberculosis* **infection.** The immune response generated in the host upon exposure to *M. tuberculosis* is complex and remains incompletely understood, with limited information about host factors that determine control versus progression. The cytokines IL-12, IL-1 and TNF, produced by innate immune cells, as well as IFN-γ produced by T cells, are protective against TB. Upon infection with *M. tuberculosis*, resident lung alveolar macrophages can become infected. **(a)** Early and low levels of type I IFN from macrophages, inflammatory monocytes and myeloid dendritic cells (DCs) and other innate immune cells at low mycobacterial loads can induce IL-1, IL-12 and TNF. **(b)** High and sustained levels of type I IFN from the macrophage and other sources (e.g. paracrine type I IFN produced by DCs upon infection with virus), can be harmful and lead to the production of the suppressive cytokine IL-10 leading to the inhibition of the production of IL-1, IL-12 and TNF by macrophages and DC, and inhibition of their activation by IFN-γ. Thus in the context of low mycobacterial loads type I IFN may be protective, whereas high mycobacterial loads and increased and sustained levels of type I IFN may result in disease progression.

Figure 3. Modular host gene signatures in tuberculosis and in other infections and diseases. Modular approaches can be utilized to tease out subtle differences between TB and other diseases and infections, by profiling blood from patients using transcriptomics approaches, such as RNA-sequencing, to capture the entire transcriptome. Each gene within the transcriptome is expressed at a particular level across each individual sample, and genes involved in similar biological pathways are co-ordinately expressed. These groups of co-ordinately expressed genes constitute individual modules that represent discrete biological pathways and can be identified using unbiased approaches such as weighted gene co-expression network analysis (WGCNA). Perturbation as a response to infection with *M. tuberculosis* or other pathogens, can be measured within each module of co-expressed genes, compared to healthy controls. Using such an approach, modular signatures can be identified for TB and other infections and diseases, to inform on

690 the immune response, and this information can also be utilized to develop reduced
691 gene signatures that are more specific to TB to develop biomarkers for diagnosis.

692

693 **Table 1.** Diagnostics for TB currently in clinical use
694

Type of measurement	Objective	Tests available	Sample type	Measure	Advantages	Disadvantages
Detect presence of bacteria	To confirm active tuberculosis	Smear microscopy	Expectorated sputum Bronchoalveolar lavage (in developed countries)	Presence of mycobacteria	Simple, rapid and inexpensive Highly indicative in high tuberculosis incidence areas Allows identification of highly infectious patients	Operator dependent and labour intensive Poor sensitivity Difficult in extra-pulmonary, pediatric, and HIV co-infected tuberculosis Cannot distinguish viable from non-viable organisms
		Bacterial culture	Expectorated sputum Bronchoalveolar lavage (in developed countries)	Confirmation of *M. tuberculosis* Evaluation of drug sensitivity	High sensitivity and specificity Enables determination of phenotypic and genotypic drug sensitivity	Culture not successful in all cases (70% in pulmonary TB and <50% in all forms of extra-pulmonary TB) Results can take up to 6 weeks or more
		Nucleic acid amplification tests (eg. GeneXpert® MTB/RIF assay)	Expectorated sputum Bronchoalveolar lavage (in developed countries)	Direct detection of *M. tuberculosis* Evaluation of certain drug sensitivities	High sensitivity and specificity Rapid turnaround time (~2 hours)	Requires sputum that can be hard to obtain from 30% of adults and most children Expensive for resource-poor settings Cannot distinguish viable from non-viable organisms
Detect host response to infection	To confirm history of *M. tuberculosis* infection	Tuberculin skin test (TST)	Skin sensitization	Memory response to mycobacterial antigens	Relatively simple test Cheap	Cannot distinguish active from latent disease Cannot distinguish remote from recent infection Cannot distinguish from other mycobacteria or BCG Operator dependent and subjective assessment of induration size
		Interferon gamma release assay (IGRA)	Blood	Memory response to *M. tuberculosis* antigen	Specific for *M. tuberculosis*	Cannot distinguish active from latent disease Cannot distinguish remote from recent infection Expensive Can be practically challenging

695
696

697 **Table 2.** Blood transcriptional reduced gene signatures proposed for TB diagnosis
698

Gene List from 8 published studies**	**Frequency of gene in each proposed published** signature**	**Singhania et al. 2018****	**Suliman et al. 2018****	**Zak et al. 2016****	**Maertzdorf et al. 2016****	**Roe et al. 2016****	**Sweeney et al. 2016****	**Kaforou et al. 2013**** (TB vs. LTBI)	**Kaforou et al. 2013**** (TB vs. Other Diseases)
DUSP3	3	-	-	-	-	-	DUSP3	DUSP3	DUSP3
FCGR1A	3	-	-	FCGR1A	FCGR1A	-	-	FCGR1A	-
GBP5	3	-	-	GBP5	GBP5	-	GBP5	-	-
SEPT4	3	-	SEPT4	SEPT4	-	-	-	-	SEPT4
ANKRD22	2	-	-	ANKRD22	-	-	-	ANKRD22	-
BATF2	2	-	-	BATF2	-	BATF2	-	-	-
FCGR1B	2	-	-	FCGR1B	-	-	-	FCGR1B[#]	-
FCGR1C	2	-	-	-	FCGR1C	-	-	FCGR1C	-
GAS6	2	-	GAS6	-	-	-	-	GAS6[#]	-
GBP1	2	-	-	GBP1	GBP1	-	-	-	-
GBP6	2	-	-	-	-	-	-	GBP6	GBP6
LHFPL2	2	-	-	-	-	-	-	LHFPL2	LHFPL2
S100A8	2	-	-	-	S100A8	-	-	S100A8	-
SCARF1	2	SCARF1	-	SCARF1	-	-	-	-	-
SERPING1	2	-	-	SERPING1	-	-	-	-	SERPING1
AAK1	1	-	-	-	-	-	-	-	AAK1
ALDH1A1	1	-	-	-	-	-	-	-	ALDH1A1[#]
APOL1	1	-	-	APOL1	-	-	-	-	-
APOL4	1	APOL4	-	-	-	-	-	-	-
ARG1	1	-	-	-	-	-	-	-	ARG1
ARHGEF9	1	ARHGEF9	-	-	-	-	-	-	-
ARNTL2	1	ARNTL2	-	-	-	-	-	-	-
BACH2	1	BACH2	-	-	-	-	-	-	-
BDH1	1	BDH1	-	-	-	-	-	-	-
BLK	1	-	BLK	-	-	-	-	-	-
BTN3A1	1	-	-	-	-	-	-	-	BTN3A1
C19ORF12	1	-	-	-	-	-	-	-	C19ORF12
C1QB	1	-	-	-	-	-	-	C1QB	-
C1QC	1	-	-	-	-	-	-	C1QC	-
C4ORF18	1	-	-	-	-	-	-	C4ORF18	-
C5	1	-	-	-	-	-	-	C5	-
CALML4	1	-	-	-	-	-	-	-	CALML4
CASC1	1	-	-	-	-	-	-	-	CASC1
CCDC120	1	CCDC120	-	-	-	-	-	-	-
CCR6	1	-	-	-	-	-	-	CCR6	-
CD177	1	-	-	-	-	CD177	-	-	-
CD1C	1	-	CD1C	-	-	-	-	-	-
CD274	1	-	-	-	CD274	-	-	-	-
CD74	1	-	-	-	-	-	-	-	CD74
CD79A	1	-	-	-	-	-	-	CD79A	-
CD79B	1	-	-	-	-	-	-	CD79B	-
CD96	1	-	-	-	CD96	-	-	-	-
CERKL	1	-	-	-	-	-	-	-	CERKL
CLC	1	-	-	-	-	CLC	-	-	-
CNIH4	1	-	-	-	CNIH4	-	-	-	-
COL4A4	1	COL4A4	-	-	-	-	-	-	-
CREB5	1	-	-	-	-	-	-	-	CREB5
CTSB	1	CTSB	-	-	-	-	-	-	-
CXCR5	1	-	-	-	-	-	-	CXCR5	-
CYB561	1	-	-	-	-	-	-	-	CYB561[#]
DHRS9	1	-	-	-	DHRS9	-	-	-	-
EBF1	1	-	-	-	-	-	-	-	EBF1
ETV7	1	-	-	ETV7	-	-	-	-	-
FAM20A	1	-	-	-	-	-	-	FAM20A	-
FAM26F	1	-	-	-	FAM26F	-	-	-	-
FBXL5	1	-	-	-	FBXL5	-	-	-	-
FLVCR2	1	-	-	-	-	-	-	FLVCR2	-
GBP2	1	-	-	GBP2	-	-	-	-	-
GBP4	1	-	-	GBP4	-	-	-	-	-
GJA9	1	-	-	-	-	-	-	-	GJA9
GNG7	1	-	-	-	-	-	-	GNG7	-
HLA-DPB1	1	-	-	-	-	-	-	-	HLA-DPB1
HM13	1	-	-	-	-	-	-	-	HM13[#]
HP	1	-	-	-	-	HP	-	-	-
HS.131087	1	-	-	-	-	-	-	-	HS.131087
HS.162734	1	-	-	-	-	-	-	-	HS.162734
ICAM1	1	ICAM1	-	-	-	-	-	-	-
ID3	1	-	-	-	ID3	-	-	-	-
IFITM3	1	-	-	-	IFITM3	-	-	-	-
IGJ	1	-	-	-	-	IGJ	-	-	-
IMPA2	1	-	-	-	-	-	-	-	IMPA2
KCNC4	1	KCNC4	-	-	-	-	-	-	-
KLF2	1	-	-	-	-	-	KLF2	-	-
LIMK1	1	LIMK1	-	-	-	-	-	-	-
LOC100133800	1	-	-	-	-	-	-	-	LOC100133800
LOC196752	1	-	-	-	-	-	-	-	LOC196752

LOC389386	1	-	-	-	-	-	-	-	LOC389386
LOC728744	1	-	-	-	-	-	-	LOC728744	-
MAK	1	-	-	-	-	-	-	-	MAK
MAP7	1	-	-	-	-	-	-	-	MAP7[#]
MIR1974	1	-	-	-	-	-	-	-	MIR1974
MPO	1	-	-	-	-	-	-	MPO	-
ORM1	1	-	-	-	-	-	-	-	ORM1
P2RY14	1	-	-	-	P2RY14	-	-	-	-
PAIP2B	1	PAIP2B	-	-	-	-	-	-	-
PCNXL2	1	-	-	-	PCNXL2	-	-	-	-
PDK4	1	-	-	-	-	-	-	-	PDK4
PGA5	1	-	-	-	-	-	-	-	PGA5
PPPDE2	1	-	-	-	-	-	-	-	PPPDE2
PRDM1	1	-	-	-	-	-	-	-	PRDM1
RBM12B	1	-	-	-	-	-	-	-	RBM12B
RNF19A	1	-	-	-	-	-	-	-	RNF19A
RP5-1022P6.2	1	-	-	-	-	-	-	-	RP5-1022P6.2
SMARCD3	1	-	-	-	-	-	-	SMARCD3	-
SMYD5	1	SMYD5	-	-	-	-	-	-	-
SPHK1	1	SPHK1	-	-	-	-	-	-	-
STAT1	1	-	-	STAT1	-	-	-	-	-
TAP1	1	-	-	TAP1	-	-	-	-	-
TMCC1	1	-	-	-	-	-	-	-	TMCC1
TMEM25	1	TMEM25	-	-	-	-	-	-	-
TRAF4	1	TRAF4	-	-	-	-	-	-	-
TRAFD1	1	-	-	TRAFD1	-	-	-	-	-
TRIM47	1	TRIM47	-	-	-	-	-	-	-
UGP2	1	-	-	-	-	-	-	-	UGP2
USP54	1	USP54	-	-	-	-	-	-	-
VAMP5	1	-	-	-	-	-	-	VAMP5	-
VEGFB	1	VEGFB	-	-	-	-	-	-	-
VPREB3	1	-	-	-	-	-	-	-	VPREB3
ZNF296	1	-	-	-	-	-	-	ZNF296	-

Abbreviations: TB, tuberculosis; LTBI, latent TB infection

#, the gene appears twice in the signature

Supplementary Table 1. TB in high and low incidence settings

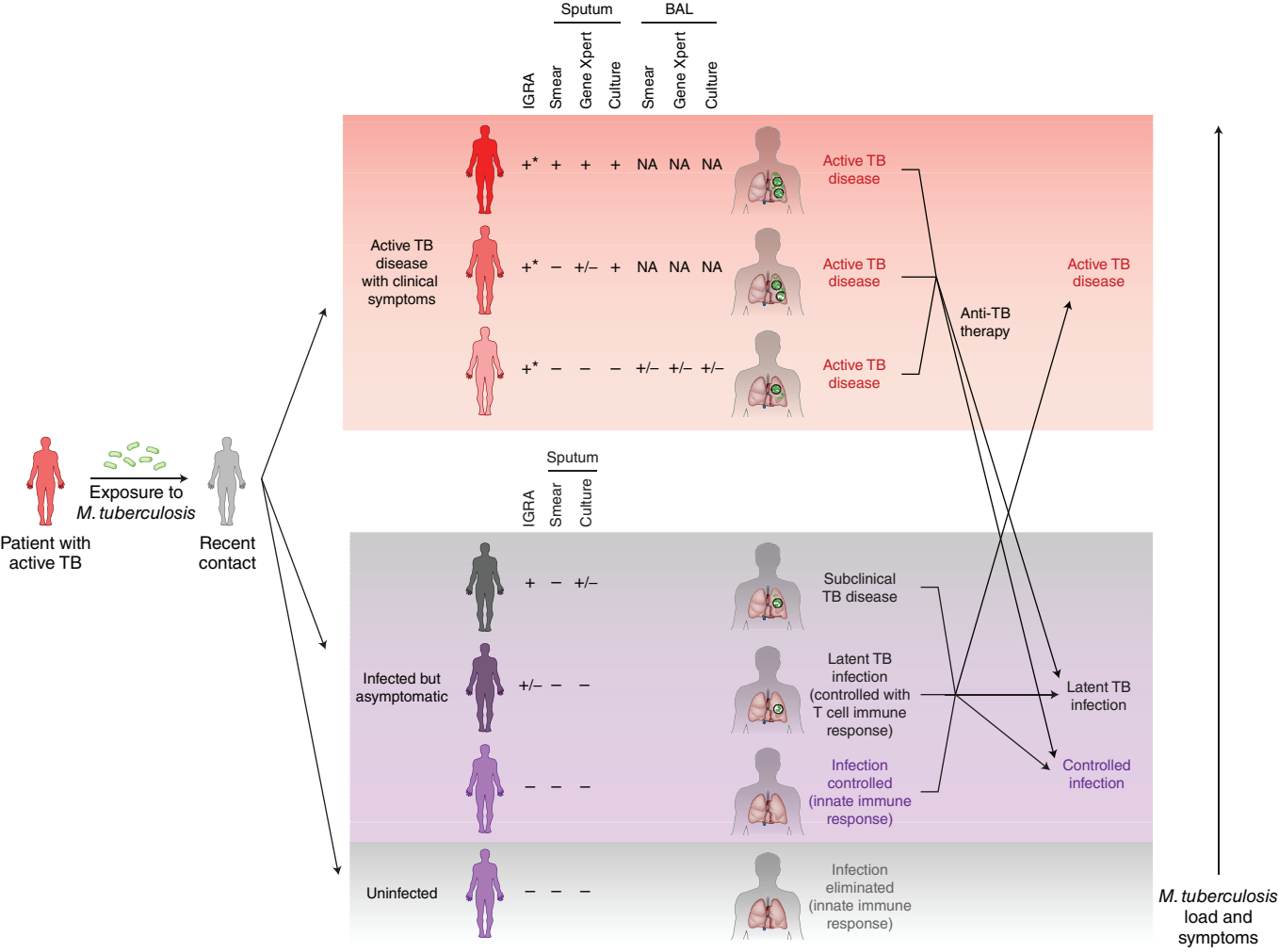
	Setting	
	High incidence, low income country	**Low incidence, high income country**
Context	Paucity of healthcare resources and infrastructure. Requirement for automated, point of care tests to support investigation and TB management	Extensive access to diagnostic tools within a well organised healthcare framework
Goals	To reduce onward transmission of infection by early identification of active TB	Progress toward TB elimination through TB prevention programmes and early identification of active TB
Clinical applications	TB diagnostic used alone or in conjunction with sputum microbiology for pulmonary TB (samples and resource permitting) to inform early initiation of TB treatment Screening tool in active case finding programmes to identify individuals with possible active TB for treatment or further investigation	TB diagnostic for supporting diagnosis of difficult cases As a screening tool to identify individuals with latent TB infection at significant risk of developing TB Screening tool for active case finding programmes in underserved populations
	Test requirements	
	Key features	**Comments**
All	Sampling from easily accessible site	Blood offers a readily accessible, minimally invasive tissue compartment for universal sampling
	Point of care or rapid inexpensive laboratory-based hardware, with automation	Automated platforms supporting rapid detection of specified reduced gene signatures are in development
TB diagnostic	High specificity to avoid inappropriate TB diagnosis	A highly specific transcriptional signature that effectively discriminates from confounding illnesses may have lower sensitivity that risks missing TB. This can be overcome by use as a follow-on test after ruling in the possibility of TB with a highly sensitive transcriptional signature developed for active case finding A biomarker that comprises a combination of gene sets and algorithms in a multiplex assay to achieve high sensitivity and high specificity in one test
Screening in active case finding	High sensitivity to avoid missing early cases of active TB	A highly sensitive test may not be sufficiently specific to discriminate from confounding illness but can effectively rule out TB in screening programmes A biomarker that comprises a combination of gene sets and algorithms in a multiplex assay to achieve high sensitivity and high specificity in one test
Screening in latently infected populations	High specificity to improve cost-effectiveness of targeted chemopreventative therapy	Transcriptional signatures with a higher specificity than TST or IGRAs for identifying individuals at risk of TB progression may be insufficiently sensitive to identify latent infection. In this context, they may be developed for use in two-step screening programmes after TST or IGRA

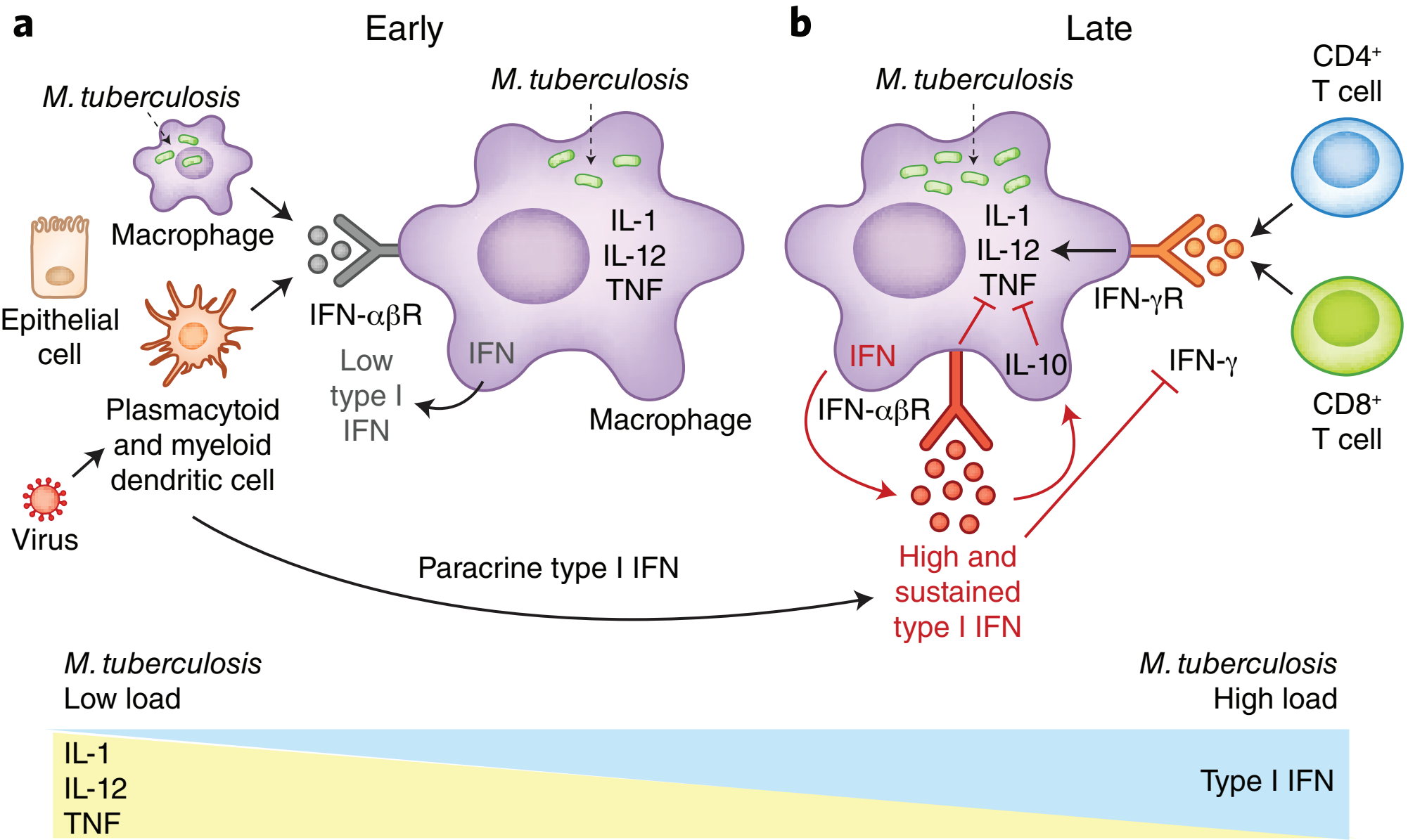
Supplementary Table 2. Accuracy of proposed blood transcriptional reduced gene signatures in diagnosing adult TB

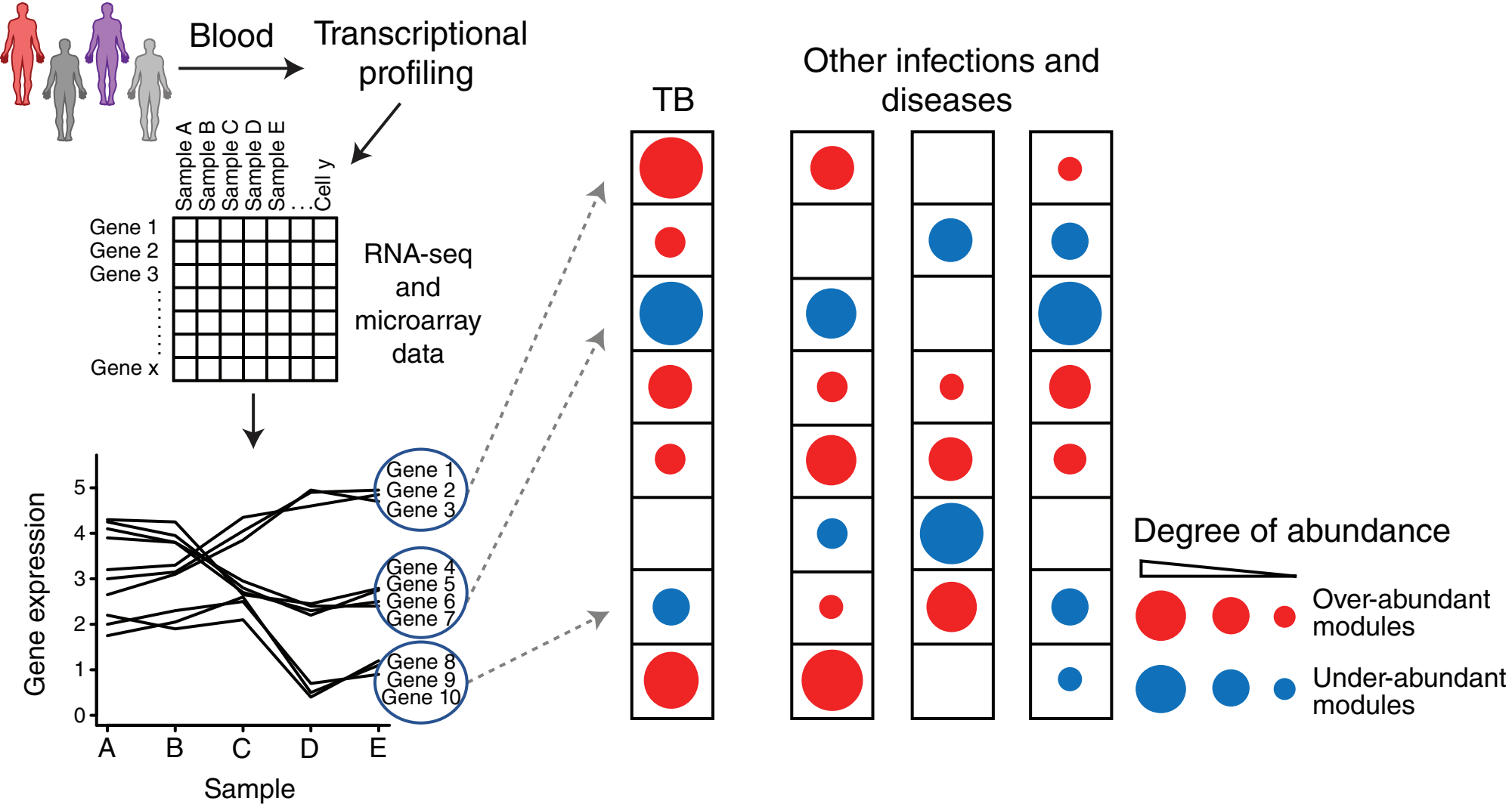
Study	Type of signature	Number of genes	Classification	Accuracy
Singhania et al. 2018	TB vs. LTBI/Other diseases	20	TB vs. LTBI	AUC 0.92-1
			TB vs. Other diseases	AUC 0.74-0.79
Suliman et al. 2018	Risk of TB progression	4	Risk of TB progression within a year of TB diagnosis	AUC 0.66
Zak et al. 2016	Risk of TB progression	16	Risk of TB progression in the 12 months preceding TB diagnosis	AUC 0.779; Sensitivity 66.1%, Specificity 80.6%
Maertzdorf et al. 2016	TB vs. Healthy individuals	4, 15	TB vs. Healthy individuals	AUC 0.98
Roe et al. 2016	TB vs. Healthy individuals/Other febrile infections	5	TB vs. Healthy individuals and other febrile infections	AUC 0.951
Sweeney et al. 2016	TB vs. LTBI/Healthy individuals/Other diseases	3	TB vs. LTBI	AUC 0.88
			TB vs. Healthy individuals	AUC 0.9
			TB vs. Other diseases	AUC 0.84
Kaforou et al. 2013	TB vs. LTBI	27	TB vs. LTBI	Sensitivity 95%, Specificity 90%
	TB vs. Other diseases	44	TB vs. Other diseases	Sensitivity 93%, Specificity 88%

Abbreviations: TB, tuberculosis; LTBI, latent TB infection; AUC, area under the curve

715


Sputum
BAL
IGRA
Smear
Gene Xpert
Culture
Smear
Gene Xpert
Culture
Active TB disease with clinical symptoms
+*
+
+
+
NA
NA
NA
Active TB disease
+*
–
+/–
+
NA
NA
NA
Active TB disease
+*
–
–
–
+/–
+/–
+/–
Active TB disease
Anti-TB therapy
Active TB disease
Patient with active TB
Exposure to M. tuberculosis
Recent contact
Sputum
IGRA
Smear
Culture
Infected but asymptomatic
+
–
+/–
Subclinical TB disease
+/–
–
–
Latent TB infection (controlled with T cell immune response)
Latent TB infection
–
–
–
Infection controlled (innate immune response)
Controlled infection
Uninfected
–
–
–
Infection eliminated (innate immune response)
M. tuberculosis load and symptoms


a
Early
M. tuberculosis
Macrophage
Epithelial cell
Plasmacytoid and myeloid dendritic cell
Virus
IFN-αβR
Low type I IFN
M. tuberculosis
IL-1
IL-12
TNF
IFN
Macrophage
Paracrine type I IFN
b
Late
M. tuberculosis
IL-1
IL-12
TNF
IFN-γR
IFN
IL-10
IFN-γ
IFN-αβR
CD4+ T cell
CD8+ T cell
High and sustained type I IFN
M. tuberculosis
Low load
M. tuberculosis
High load
IL-1
IL-12
TNF
Type I IFN


Blood
Transcriptional profiling
Sample A
Sample B
Sample C
Sample D
Sample E
…
Cell y
Gene 1
Gene 2
Gene 3
Gene x
RNA-seq and microarray data
Gene expression
5
4
3
2
1
0
A
B
C
D
E
Sample
Gene 1
Gene 2
Gene 3
Gene 4
Gene 5
Gene 6
Gene 7
Gene 8
Gene 9
Gene 10
TB
Other infections and diseases
Degree of abundance
Over-abundant modules
Under-abundant modules