

# Blood Transcriptomic Stratification of Short-term Risk in Contacts of Tuberculosis

Jennifer Roe,<sup>1</sup> Cristina Venturini,<sup>1</sup> Rishi K. Gupta,<sup>2</sup> Celine Gurry,<sup>1</sup> Benjamin M. Chain,<sup>1</sup> Yuxin Sun,<sup>3</sup> Jo Southern,<sup>2</sup> Charlotte Jackson,<sup>2</sup> Marc C. Lipman,<sup>4,5</sup> Robert F. Miller,<sup>2</sup> Adrian R. Martineau,<sup>6</sup> Ibrahim Abubakar,<sup>2,a</sup> and Mahdad Noursadeghi<sup>1,7,a</sup>

<sup>1</sup>Division of Infection & Immunity, <sup>2</sup>Institute for Global Health, <sup>3</sup>Department of Computer Science, and <sup>4</sup>University College London Respiratory Medicine, University College London, <sup>5</sup>Department of Respiratory Medicine, Royal Free London National Health Service Trust, <sup>6</sup>Blizard Institute, Queen Mary University of London, and <sup>7</sup>National Institute for Health Research University College London Hospitals Biomedical Research Centre, United Kingdom

**Background.** The highest risk of tuberculosis arises in the first few months after exposure. We reasoned that this risk reflects incipient disease among tuberculosis contacts. Blood transcriptional biomarkers of tuberculosis may predate clinical diagnosis, suggesting they offer improved sensitivity to detect subclinical incipient disease. Therefore, we sought to test the hypothesis that refined blood transcriptional biomarkers of active tuberculosis will improve stratification of short-term disease risk in tuberculosis contacts.

**Methods.** We combined analysis of previously published blood transcriptomic data with new data from a prospective human immunodeficiency virus (HIV)–negative UK cohort of 333 tuberculosis contacts. We used stability selection as an alternative computational approach to identify an optimal signature for short-term risk of active tuberculosis and evaluated its predictive value in independent cohorts.

**Results.** In a previously published HIV-negative South African case-control study of patients with asymptomatic *Mycobacterium tuberculosis* infection, a novel 3-gene transcriptional signature comprising *BATF2*, *GBP5*, and *SCARF1* achieved a positive predictive value (PPV) of 23% for progression to active tuberculosis within 90 days. In a new UK cohort of 333 HIV-negative tuberculosis contacts with a median follow-up of 346 days, this signature achieved a PPV of 50% (95% confidence interval [CI], 15.7–84.3) and negative predictive value of 99.3% (95% CI, 97.5–99.9). By comparison, peripheral blood interferon gamma release assays in the same cohort achieved a PPV of 5.6% (95% CI, 2.1–11.8).

**Conclusions.** This blood transcriptional signature provides unprecedented opportunities to target therapy among tuberculosis contacts with greatest risk of incident disease.

**Keywords.** blood transcriptome; biomarker; tuberculosis; diagnosis.

The causative agent of tuberculosis, *Mycobacterium tuberculosis* (Mtb) is an obligate pathogen [1]. In the absence of an effective vaccine, the earliest possible identification of disease offers the best strategy to reduce onward transmission. Early treatment also offers the opportunity to adopt shorter and simpler treatment regimens, limit pathology, and reduce tuberculosis-associated morbidity. This rationale and the fact that risk of tuberculosis is highest within the first few months after significant exposure [2, 3] provide the basis for screening close contacts of patients with active tuberculosis in order to identify prevalent

disease and offer preventative treatment to individuals with asymptomatic infection.

Detection of immune memory for Mtb using interferon gamma release assays (IGRAs) or the tuberculin skin test (TST) is widely used to identify infected individuals. These tests have poor sensitivity, estimated at 50%–85% for identifying contacts who progress to tuberculosis. In addition, they have poor positive predictive value (PPV), estimated to be <5% for 2-year cumulative tuberculosis incidence or 1–1.5 per 100 person-years [4–6]. In the absence of clinical disease, the ability to stratify differential risk of progression to tuberculosis remains extremely limited, leading to unnecessary treatment of people at low risk. The fact that overall risk is low means that significant numbers of individuals refuse the offer of treatment, thereby undermining the overall effectiveness of contact tracing to prevent incident tuberculosis.

Blood transcriptomic biomarkers have emerged as a sensitive approach for identification of active tuberculosis [7–11]. Importantly, these may predate conventional clinical diagnosis [12, 13]. In a South African cohort of human immunodeficiency virus (HIV)–negative individuals with latent tuberculosis [12],

Received 4 January 2019; editorial decision 18 March 2019; accepted 20 March 2019; published online March 28, 2019.

<sup>a</sup>I. A. and M. N. contributed equally to this work.

Correspondence: M. Noursadeghi, Division of Infection & Immunity, Cruciform Building, University College London, London WC1E 6BT, United Kingdom (m.noursadeghi@ucl.ac.uk).

Clinical Infectious Diseases® 2019;70(5):731–7

© The Author(s) 2019. Published by Oxford University Press for the Infectious Diseases Society of America. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited. DOI: 10.1093/cid/ciz252

the sensitivity of a 16-gene blood transcriptional signature that discriminated individuals who progressed to a diagnosis of active tuberculosis from those who did not improve as the time interval between sampling and diagnosis was reduced. These data suggest that blood transcriptional signatures can identify presymptomatic incipient disease. The fact that blood transcriptional changes may not be suitable for stratification of long-term risk was highlighted in a second multicohort African study of household contacts that specifically excluded patients who progressed to active tuberculosis within 3 months of enrollment [13]. In that study, a 4-gene blood transcriptional signature discriminated between progressors and nonprogressors, with a modest receiver operating characteristic area under the curve (AUC) of 0.69 and PPV of 3%, equivalent to that of IGRAs or TST.

These studies focused on deriving blood transcriptional signatures for prospective risk of incident tuberculosis for up to 2 years. However, they may not represent the most sensitive biomarkers of preclinical incipient disease that underlie the short-term risk of tuberculosis among contacts. We recently reported that *BATF2* gene expression provided a single blood transcript that accurately discriminated active from latent tuberculosis [11]. *BATF2* was identified by comparing transcriptional profiles from patients with active and treated tuberculosis. Of note, *BATF2* is a component of the 16-gene signature reported by Zak et al [12].

In the present study, we tested the ability of *BATF2* on its own to identify incipient tuberculosis disease in the Zak cohort. In addition, we sought to improve on the predictive value of *BATF2* using stability selection as an alternative computational approach to identifying discriminating features in high-dimensional data. We then validated our findings in 2 independent datasets, first, the previously published South African cohort of HIV-negative individuals with latent tuberculosis infection [12] and then a new UK cohort of tuberculosis contacts.

## METHODS

### Analysis of Previously Published Data

Raw sequencing data from Zak et al [12] was pseudoaligned to the human transcriptome (Ensembl Human GRCh38) using Kallisto [14]. The abundance for protein-coding RNA was expressed as  $\log_2$ -transformed transcripts per million [15]. Microarray data from Roe et al [11] were used as normalized  $\log_2$ -transformed data. RNA sequencing and microarray data were standardized by subtracting the mean and dividing by the standard deviation of each dataset. Gene expression data from each study were annotated with Human Genome Organisation Nomenclature Committee gene symbols.

### Blood Transcriptional Profiling of a UK Cohort of Tuberculosis Contacts

Close contacts of patients with tuberculosis were invited to participate (Supplementary Methods). The study was approved by the UK National Research Ethics Service (reference 14/EM/1208).

All participants provided written informed consent. At enrollment, IGRAs were done using the QuantiFERON-TB Plus assay (Qiagen, Germany), and peripheral blood RNA was collected into Tempus tubes for transcriptional profiling by RNA sequencing (Supplementary Methods). These data are available in ArrayExpress [16] under accession E-MTAB-6845. At the end of the study, participants who progressed to active tuberculosis were identified by linkage with the national electronic tuberculosis register as previously described [17]. Local case notes were reviewed in order to identify individuals who had received preventative treatment.

### Stability Selection to Refine the Optimal Gene Signature for Incipient Tuberculosis

We combined stability selection with support vector machines (SVMs) [18] (Supplementary Methods) to identify a ranking of the individual transcripts independent of other genes that discriminated patients with active tuberculosis pretreatment from those who had been successfully treated [11]. After we selected a consistent subset of discriminating genes using stability selection, we used these genes to train a traditional L2 Norm learning SVM using kernlab with a linear kernel in the R statistical computing platform [19] as previously described [11]. We used this SVM to generate decision values for each dataset tested. Receiver operating characteristic (ROC) curves were plotted in GraphPad Prism, version 7 (GraphPad Software, La Jolla, CA) and used to calculate the ROC AUC. The Youden index for each ROC curve was derived from the sum of sensitivity and specificity-1 [20], and the predictive value of each transcriptional signature was estimated using Bayesian conditional probabilities [21]; 95% confidence intervals are provided for each measure of test performance.

## RESULTS

### Positive Predictive Value of *BATF2* Blood Transcript Levels for Short-term Risk of Incident Tuberculosis

We first sought to test the hypothesis that elevated blood levels of *BATF2* gene expression identified individuals with subclinical incipient disease leading to incident tuberculosis in the short term. We compared *BATF2* blood transcript levels in all samples from individuals with latent tuberculosis in the Zak cohort [12] who did not progress to active tuberculosis ( $n = 48$ ) with samples from individuals who progressed to active tuberculosis within 90 days ( $n = 12$ ), 91–360 days ( $n = 29$ ), and after an interval of greater than 360 days ( $n = 25$ ). In this analysis, we ensured that samples were not incorrectly allocated to the nonprogressor group because of inadequate follow-up by restricting the nonprogressor group to include only samples from individuals who had more than 12 months follow-up after the sample was collected. Consistent with our hypothesis, *BATF2* blood transcript levels were highest in the group of individuals who progressed to active tuberculosis within

90 days (Figure 1A). Accordingly, discrimination of individuals who progress to tuberculosis from nonprogressors using *BATF2* levels achieved the highest ROC AUC of 0.93 (0.86–1) in the group who progressed to tuberculosis within 90 days (Figure 1B). The threshold for *BATF2* levels that discriminated this group of progressors from nonprogressors with the greatest accuracy was identified using the ROC curve Youden index. This threshold achieved a sensitivity of 0.83 (0.52–0.98) and a specificity of 0.92 (0.83–0.99; Figure 1C), giving a positive likelihood ratio of 13.3 (4.3–41.1). At the same threshold, the sensitivity of elevated *BATF2* levels to identify individuals who progressed to tuberculosis after 90 days was reduced to 0.52 (0.36–0.74) in the 91–360 day interval and to 0.24 (0.12–0.49) after 360 days. The cumulative tuberculosis risk in this cohort approximated to 1.5% [12]. Using this pretest probability, we estimated the PPV of elevated *BATF2* levels for diagnosis of tuberculosis within 90 days to be 13% (Figure 1C). These were comparable to the PPV for disease within 90 days using the 16-gene signature described by Zak et al, which achieved an ROC AUC of 0.94 (0.89–1; Supplementary Figure 1).

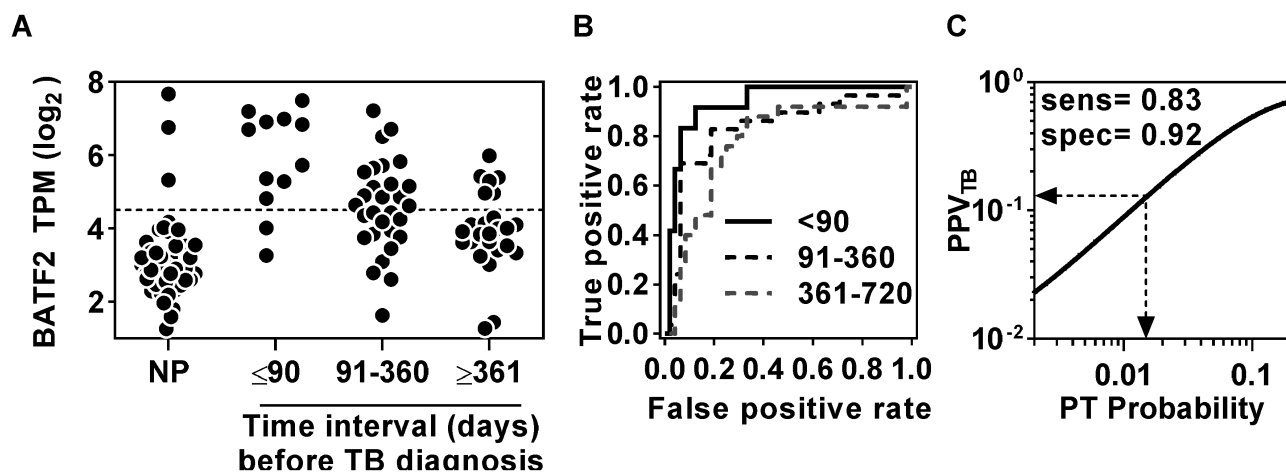
#### A Refined 3-Gene Signature to Predict Active Tuberculosis Within 90 Days

In the clinical cohort described by Zak et al, their 16-gene signature offered no advantage to measuring *BATF2* transcripts alone to discriminate progressors from nonprogressors within 90 days. Nonetheless, we hypothesized that the addition of selected genes may further improve the performance of *BATF2* alone as a biomarker of incipient tuberculosis. We reasoned that the most discriminating transcriptional biomarker may be different among subsets of cases. Therefore, a combination of the genes most frequently ranked top in

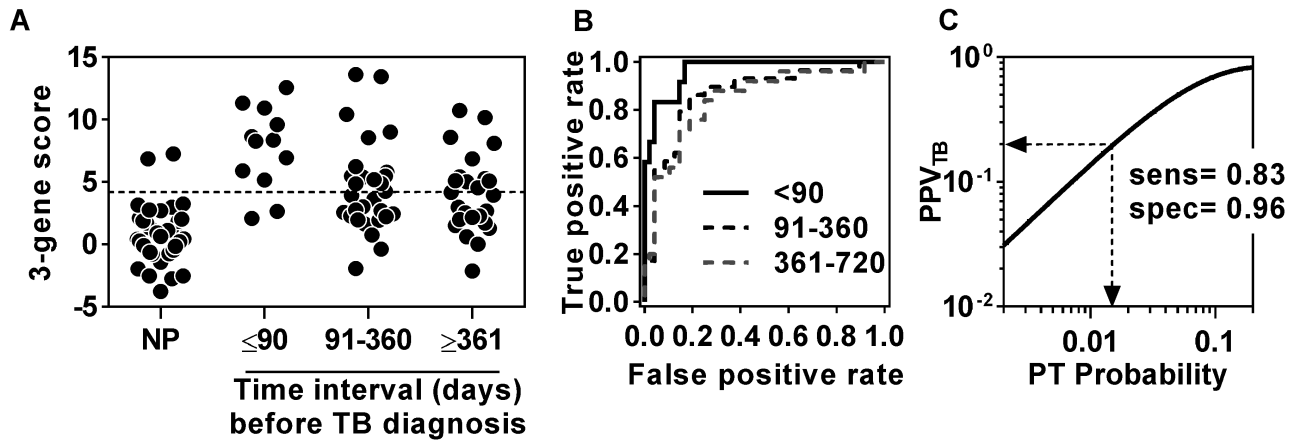
multiple subsamples of the data may give the optimal gene signature for incipient tuberculosis. This approach of featuring selection in high-dimensional data has been called stability selection [18]. We used stability selection to rank the transcripts that best discriminated subsets of our previously published active and treated tuberculosis cases [11]. Using this ranking, we trained an SVM model to discriminate active and treated tuberculosis with a cumulative number of genes and tested how accurately each SVM model correctly classified progressor and nonprogressor patients in the Zak cohort (Supplementary Figure 2B). The most accurate classification was achieved by the top 3 genes comprising *BATF2*, *GBP5*, and *SCARF1* (Supplementary Figure 2C). We represented the 3-gene score for each individual in the Zak cohort as the distance from the separating hyperplane that discriminates between 2 classes in the SVM model (Figure 2A). The 3-gene SVM model discriminated between nonprogressors and those who progressed to tuberculosis within 90 days with an ROC AUC of 0.96 (0.92–1; Figure 2B). At the Youden index, this ROC curve generated a sensitivity of 0.83 (0.52–0.98) and a specificity of 0.96 (0.83–0.99). This was equivalent to a positive likelihood ratio of 20 (5–79.5) and a PPV for active tuberculosis within 90 days of 23%, given a prior probability of 1.5% (Figure 2C).

#### Predictive Value of the 3-Gene Signature for Active Tuberculosis in a New UK Cohort of Tuberculosis Contacts

The predictive value of a tuberculosis biosignature based on *BATF2*, *GBP5*, and *SCARF1*, described above, was obtained from estimates of sensitivity, specificity derived from case-control data, and estimates of prior probability. A prospective,



**Figure 1.** Identification of incipient tuberculosis by measurement of blood *BATF2* transcript levels. *A*, *BATF2* transcript levels (TPM) are shown for blood samples from all NP patients in the Zak et al [12] cohort (with at least 12 months follow-up after the time of sampling) and for samples from patients who progressed to a diagnosis of tuberculosis within the time intervals indicated. *B*, ROC curves for discriminating between NP and progressors in each time interval shown using the *BATF2* transcript level. *C*, PPV<sub>TB</sub> for patients who progress to tuberculosis within 90 days using sensitivity and specificity values derived from the optimal Youden index (dotted line in [A]) of the ROC curve in (B) and a range of PT probabilities. Arrows highlight the PPV<sub>TB</sub> of 13% for PT probability of 1.5%. Abbreviations: NP, patients who did not progress to tuberculosis; PPV<sub>TB</sub>, positive predictive value for a diagnosis of tuberculosis; PT, pretest; ROC, receiver operating characteristic; sens, sensitivity; spec, specificity; TPM, transcripts per million.



**Figure 2.** Identification of incipient tuberculosis by a novel 3-gene model incorporating blood transcript levels of *BATF2*, *GBP5*, and *SCARF1*. **A**, Three-gene scores derived from the support vector machine model to discriminate between active and treated tuberculosis are shown for blood samples from all NP patients in the Zak et al [12] cohort (with at least 12 months follow-up) and for samples from patients who progressed to a diagnosis of tuberculosis within the time intervals indicated. **B**, ROC curves for discriminating between NP and progressors in each time interval shown using the 3-gene scores. **C**, PPV<sub>TB</sub> for patients who progress to tuberculosis within 90 days using sensitivity and specificity values derived from the optimal Youden index of the ROC curve in (B) and a range of PT probabilities. Arrows highlight the PPV<sub>TB</sub> of 23% for PT probability of 1.5%. Abbreviations: NP, patients who did not progress to tuberculosis; PPV<sub>TB</sub>, positive predictive value for a diagnosis of tuberculosis; PT, pretest; ROC, receiver operating characteristic; sens, sensitivity; spec, specificity.

independent, observational cohort was required to confirm these findings. In addition, although our 3-gene signature was discovered in data derived from a UK cohort of patients with active tuberculosis and validated in a South African cohort of patients with latent tuberculosis, additional validation in a further independent UK cohort at risk of active tuberculosis was necessary to extend the evidence for its generalizability. Therefore, we obtained blood transcriptomic data from a new observational HIV-negative cohort of 333 close contacts of cases of active tuberculosis, representing a group at highest risk of developing disease in the short term (Table 1). Median follow-up of the cohort was 346 days (interquartile range, 250–450). Six participants in the cohort progressed to a diagnosis of tuberculosis disease 3–342 days after recruitment to the study (Table 2).

We used the novel 3-gene model to calculate a decision score, as described above, for each patient in the UK tuberculosis contacts cohort. First, we sought to define the distribution of 3-gene scores in 192 IGRA-negative contacts as a control population among tuberculosis contacts with low risk of developing disease. This group was younger, included fewer non-UK born individuals, and included fewer household contacts of tuberculosis compared with the IGRA-positive individuals, reflecting known risk factors for acquisition of *Mtb* infection (Table 1). The 3-gene scores among the IGRA-negative group showed a parametric distribution. Therefore, we used a standard score of 2 ( $z_2$ ) to represent the 97.7th percentile and a standard score of 3 ( $z_3$ ) to represent the 99.9th percentile as thresholds for an elevated 3-gene score (Figure 3A). We then compared the distribution of 3-gene scores among all IGRA-positive contacts and tested the hypothesis that the 3-gene score stratifies differential risk of incident tuberculosis in this cohort. Among individuals

in the cohort who did not receive preventative therapy, we compared the incident tuberculosis rate per 100 person-years in all IGRA-positive and IGRA-negative contacts with 3-gene scores greater than the  $z_2$  and  $z_3$  thresholds (Figure 3B).

Using the  $z_2$  threshold, the tuberculosis incidence rate among individuals with a low 3-gene score was 0.76 per 100 person-years (0.19–3.05). Among 19 individuals with a high 3-gene score, the tuberculosis incidence rate was 27.7 per 100 person-years (10.4–73.8). The incidence rate ratio (incidence of tuberculosis for those with a high 3-gene score compared with those with a low 3-gene score) was 36.3 (6.6–198.1). Using the  $z_3$  threshold, the tuberculosis incidence rate among individuals with a low 3-gene score was 0.74 per 100 person-years (0.18–3.0). Among 8 individuals with a high 3-gene score, the tuberculosis incidence rate was 72.0 per 100 person-years (27.0–191.8). The incidence rate ratio (incidence of tuberculosis for those with a high 3-gene score compared with those with a low 3-gene score) was 97.4 (17.8–532.0). By comparison, incidence rates were 5.8 per 100 person-years (2.6–13.0) and 0 per 100 person-years among IGRA-positive and IGRA-negative contacts, respectively. Overall, at the  $z_2$  threshold, the 3-gene score achieved a PPV of 21.1% (6.1–46.6) and a positive likelihood ratio of 13 (6.2–27.6). At the  $z_3$  threshold, the 3-gene score achieved a PPV of 50% (15.7–84.3) and a positive likelihood ratio of 48.8 (15.8–150.5). Both of these thresholds for the 3-gene score achieved negative predictive value (NPV) of 99.3% (97.5–99.9 for the  $z_3$  threshold and 97.4–99.9 for the  $z_2$  threshold) and a negative likelihood ratio of 0.35 (0.11–1.1) for cumulative incident tuberculosis within 2 years. The IGRA result achieved a PPV of 5.6% (2.1–11.8) and an NPV of 100% (98.1–100).

**Table 1. Summary Characteristics of United Kingdom Tuberculosis Contacts Cohort**

Characteristic	All (N = 333)		IGRA positive (n = 141)		IGRA negative (n = 192)		PValue
Age, median (interquartile range), y	33 (26–47)		40.5 (29–51)		32 (25–43)		<.05
	n	%	n	%	n	%	
<b>Sex</b>							
Male	173	52.0	77	54.6	96	50.0	ns
Female	155	46.5	62	44.0	93	48.4	ns
NA	5	1.5	2	1.4	3	1.6	ns
<b>Ethnicity</b>							
White	73	21.9	24	17.0	49	25.5	ns
Indian, Pakistani, or Bangladeshi	120	36.0	51	36.2	69	35.9	ns
Black	56	16.8	32	22.7	24	12.5	<.05
Mixed	64	19.2	29	20.6	35	18.2	ns
Other	10	3.0	2	1.4	8	4.2	ns
NA	10	3.0	3	2.1	7	3.6	ns
<b>Country of birth</b>							
Non–United Kingdom	259	77.8	126	89.4	133	69.3	<.05
United Kingdom	68	20.4	12	8.5	56	29.2	<.05
NA	6	1.8	3	2.1	3	1.6	ns
<b>Tuberculosis contact type</b>							
Household	218	65.5	103	73.0	115	59.9	<.05
Nonhousehold	77	23.1	25	17.7	52	27.1	<.05
NA	38	11.4	13	9.2	25	13.0	ns
<b>Social risk factor</b>							
Yes	35	10.5	13	9.2	22	11.5	ns
No	298	89.5	128	90.8	170	88.5	ns
NA	0	0	0	0	0	0	ns
<b>Diabetes</b>							
Yes	27	8.1	16	11.3	11	5.7	ns
No	296	88.9	122	86.5	174	90.6	ns
NA	10	3	3	2.1	7	3.6	ns
<b>Smoking</b>							
Current	76	22.8	32	22.7	44	22.9	ns
Previous	43	12.9	18	12.8	25	13	ns
No	21	63.1	89	63.1	121	63	ns
NA	4	1.2	2	1.4	2	1	ns
<b>Preventative therapy</b>							
Yes	34	10.2	34	24.1	0	0	n/a
No	289	86.8	97	68.8	192	100	n/a
NA	10	3.0	10	7.1	0	0	n/a

For statistical tests, age was compared using a Mann-Whitney U test, and categorical variables were compared using  $\chi^2$  tests. Social risk factors included history of homelessness, imprisonment, or harmful drug use.

Abbreviations: IGRA, interferon gamma release assay; NA, not available; n/a, not applicable; ns, non-significant.

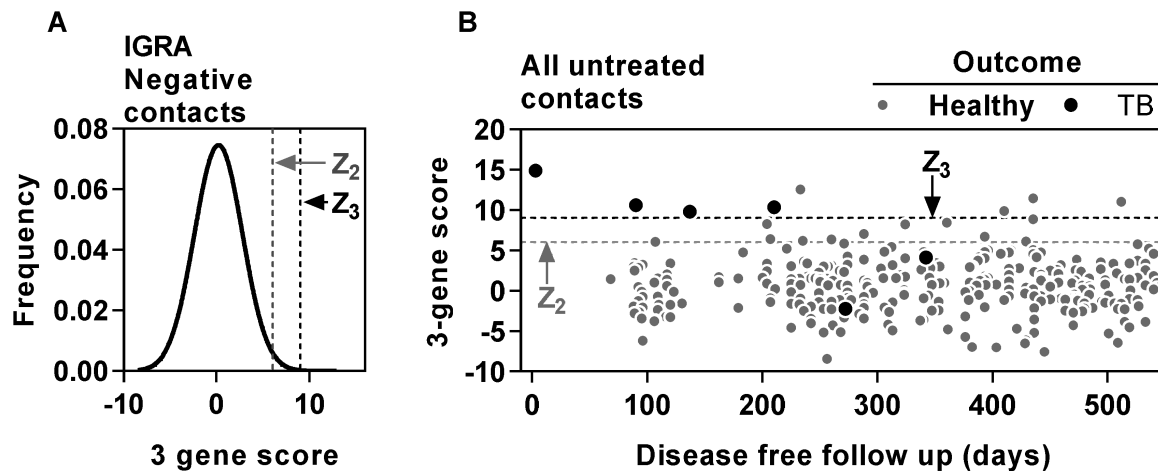
## DISCUSSION

The present study was based on the premise that blood transcriptional biomarkers of active tuberculosis predate clinical

presentation of disease and consequently serve as biomarkers of incipient tuberculosis. We confirmed this hypothesis in data from a South African case-control study using a single

**Table 2. Selected Characteristics of Patients Who Progressed to a Diagnosis of Tuberculosis in the United Kingdom Cohort Study**

Disease-free Interval, days	Age, y	Sex	Ethnicity	Interferon Gamma Release Assay	Site of Disease	Culture Confirmed	Symptoms at Screening	Screening Chest Radiograph
3	32	Female	Indian	Positive	Uterus	Yes	Yes	Normal
90	16	Male	Black	Positive	Pleural	No	No	Normal
137	21	Male	Mixed	Positive	Lymph node	No	No	Normal
210	21	Female	Mixed	Positive	Lymph node	No	No	Paratracheal adenopathy
272	52	Male	Indian	Positive	Pleural	Yes	No	Normal
342	27	Female	Black	Positive	Pleural	No	No	Normal



**Figure 3.** Blood transcriptomic 3-gene score at recruitment in contacts of active tuberculosis (TB). *A*, Frequency distribution of 3-gene scores in IGRA-negative contacts of active TB showing threshold (dashed lines) for identification of a high 3-gene score based on the mean +2 SD ( $z_2$ ) or +3 SD ( $z_3$ ) of the scores among IGRA-negative cases. *B*, Individual 3-gene scores for untreated IGRA-positive and IGRA-negative contacts who developed active TB or remained healthy on follow-up. Abbreviations: IGRA, interferon gamma release assay; SD, standard deviation.

transcriptional biomarker for active tuberculosis, *BATF2*. The addition of 2 additional genes generated a 3-gene signature derived from independent data, which further enhanced the discrimination between short-term progressors and nonprogressors in the case-control study. Finally, we validated our findings in a new UK cohort study.

We provide the first proof-of-concept data for a blood transcriptional signature that predicts the short-term risk of disease in tuberculosis contacts with substantially greater PPV than is achieved in current practice using IGRAs. The 3-gene signature in the present study was derived from comparison of patients with pulmonary tuberculosis before and after treatment, but predicted cases of extrapulmonary disease in the UK cohort. This finding is consistent with the hypothesis that blood transcriptional signatures of tuberculosis are not specific to different sites of disease, but is equally applicable to pulmonary and extrapulmonary tuberculosis. Our blood transcriptional signature predicted disease progression in 4 of 6 close contacts. The patient who progressed within 3 days was symptomatic at enrollment, but the others had no symptoms of active disease, supporting the underlying hypothesis that this signature predates clinical presentation of disease. The potential clinical impact of this approach is to enable more precise targeting of preventative antimicrobial tuberculosis treatment. On the basis of the differential PPVs among contacts of active tuberculosis, risk stratification using the blood transcriptional signature may reduce the number needed to treat to <5 compared to >20 using IGRAs. In addition, the better PPV of the blood transcriptional signature may be expected to incentivize increased treatment acceptance and completion rates. Taken together, these effects have the potential to transform the efficiency and therefore

scalability of contact tracing as part of a tuberculosis control program.

In the South African case-control study, the sensitivity of the 3-gene signature to identify individuals who progressed to tuberculosis reduced as interval time to tuberculosis increased. Consistent with this, in the UK cohort study, the 3-gene signature failed to identify 2 of 6 contacts who progressed to active tuberculosis, with the longest disease-free intervals both greater than 6 months. These data highlight the reduced sensitivity for long-term risk of incident disease. Therefore, interval follow-up measurements for IGRA-positive contacts may be needed to prevent cases beyond the first 6 months. Thereafter, the numbers needed to screen and treat with preventative therapy for the very small residual long-term risk of disease may not justify the economic cost and risk of drug toxicity. To illustrate this, in a large observational cohort of 4861 tuberculosis contacts with median follow-up of 2.9 years, 52% and 73% of incident tuberculosis cases occurred within 6 and 12 months of contact screening, respectively [6]. Therefore, a strategy of testing recent tuberculosis contacts with blood transcriptional biomarkers at baseline and after a 6-month interval may identify up to three quarters of incident tuberculosis cases among contacts.

The major limitation of our study is the low frequency of progressive disease in the UK cohort. Despite the fact that tuberculosis contacts have the highest short-term risk of disease, the absolute 2-year cumulative incidence of disease is low, necessitating very large-scale cohorts to definitively assess the correlates of progression [6]. Consequently, in the present study, the confidence intervals of the PPV for the 3-gene signature are wide, albeit significantly better than IGRA, which achieves a PPV of <5% in large-scale studies. The impact of HIV coinfection is also untested. In addition, it is clear that

a high 3-gene score is evident in individuals who do not develop active tuberculosis. This observation may reflect spontaneous resolution of subclinical tuberculosis among some individuals, inadequate follow-up of these patients, or a lack of specificity for tuberculosis. Therefore, further assessments of potential confounding of this transcriptional signature by HIV coinfection, other comorbidities, and longitudinal studies of the expression of this signature with and without tuberculosis treatment are required. Notwithstanding these limitations, we propose that the 3-gene blood transcriptional signature offers exciting new opportunities to transform risk stratification for progression to tuberculosis disease among contacts of active tuberculosis. This application of the blood transcriptional signature is consistent with proposals from the World Health Organization for a non-sputum test with sensitivity and specificity of >75% that predicts risk of disease in patients with incipient tuberculosis [22]. Our findings pave the way for extended validation of the signature “head-to-head” comparison of the performance of the different published blood transcriptional signatures of tuberculosis and development of the technology to allow scale-up of near-patient testing.

### Supplementary Data

Supplementary materials are available at *Clinical Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

### Notes

**Author contributions.** J. R., M. L., R. F. M., A. R. M., I. A., and M. N. conceived of the study. R. G., J. S., and C. J. collected samples and data. J. R. and C. G. processed samples. J. R., C. V., R. G., B. M. C., Y. S., and M. N. analyzed the data. J. R., R. G., I. A., and M. N. wrote the manuscript with input from all the authors.

**Financial support.** This work was supported by an Medical Research Council (MRC) Fellowship to J. R. (MR/L001756/1), an National Institute of Health Research (NIHR) fellowship to R. G. (DRF-2018-11-ST2-004), a Wellcome Trust Investigator Award to M. N. (207511/Z/17/Z), the British Lung Foundation (TB05/11), the University College London Technology Fund, the Rosetrees Trust, and the NIHR Biomedical Research Centre Funding to University College Hospitals NHS Foundation Trust and University College London. I. A. was funded by NIHR (SRF-2011-04-001, NF-SI-0616-10037), MRC, UK Department of Health, and the Wellcome Trust.

**Potential conflicts of interest.** J. R., A. R. M., and M. N. have a patent application pending in relation to blood transcriptomic biomarkers of tuberculosis. R. M. reports personal fees from Gilead outside the submitted work. All other authors: No reported conflicts. All authors have submitted

the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

### References

1. Gagneux S. Ecology and evolution of *Mycobacterium tuberculosis*. *Nat Rev Microbiol* **2018**; 16:202–13.
2. Sloot R, Schim van der Loeff MF, Kouw PM, Borgdorff MW. Risk of tuberculosis after recent exposure. A 10-year follow-up study of contacts in Amsterdam. *Am J Respir Crit Care Med* **2014**; 190:1044–52.
3. Behr MA, Edelstein PH, Ramakrishnan L. Revisiting the timetable of tuberculosis. *BMJ* **2018**; 362:k2738.
4. Zellweger JP, Sotgiu G, Block M, et al; TBNET. Risk assessment of tuberculosis in contacts by IFN- $\gamma$  release assays. A tuberculosis Network European Trials Group Study. *Am J Respir Crit Care Med* **2015**; 191:1176–84.
5. Haldar P, Thuraisingam H, Patel H, et al. Single-step QuantiFERON screening of adult contacts: a prospective cohort study of tuberculosis risk. *Thorax* **2013**; 68:240–6.
6. Abubakar I, Drobniewski F, Southern J, et al. Prognostic value of interferon- $\gamma$  release assays and tuberculin skin test in predicting the development of active tuberculosis (UK PREDICT TB): a prospective cohort study. *Lancet Infect Dis* **2018**.
7. 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.
8. Anderson ST, Kaforou M, Brent AJ, et al. Diagnosis of childhood tuberculosis and host RNA expression in Africa. *N Engl J Med* **2014**; 370:1712–23.
9. Maertzdorf J, McEwen G, Weiner J, et al. Concise gene signature for point-of-care classification of tuberculosis. *EMBO Mol Med* **2015**.
10. Sweeney TE, Khatri P. Blood transcriptional signatures for tuberculosis diagnosis: a glass half-empty perspective; authors' reply. *Lancet Respir Med* **2016**; 4:e29.
11. Roe JK, Thomas N, Gil E, et al. Blood transcriptomic diagnosis of pulmonary and extrapulmonary tuberculosis. *JCI Insight* **2016**; 1:e87238.
12. Zak DE, Penn-Nicholson A, Scriba TJ, et al; ACS and GC6-74 Cohort Study Groups. A blood RNA signature for tuberculosis disease risk: a prospective cohort study. *Lancet* **2016**; 387:2312–22.
13. Suliman S, Thompson E, Sutherland J, et al. Four-gene Pan-African blood signature predicts progression to tuberculosis. *Am J Respir Crit Care Med* **2018**.
14. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **2016**; 34:525–7.
15. Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci* **2012**; 131:281–5.
16. ArrayExpress < EMBL-EBI. Available at: <https://www.ebi.ac.uk/arrayexpress/>.
17. Aldridge RW, Shaji K, Hayward AC, Abubakar I. Accuracy of probabilistic linkage using the enhanced matching system for public health and epidemiological studies. *PLoS One* **2015**; 10:e0136179.
18. Meinshausen N, Bühlmann P. Stability selection. *J R Stat Soc Ser B Stat Methodol* **2010**; 72:417–473.
19. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, **2015**. Available at: <http://www.R-project.org/>. Accessed 14 March 2019.
20. Youden WJ. Index for rating diagnostic tests. *Cancer* **1950**; 3:32–5.
21. López Puga J, Krzywinski M, Altman N. Points of significance: Bayes' theorem. *Nat Methods* **2015**; 12:277–8.
22. World Health Organization. Consensus meeting report: development of a target product profile (TPP) and a framework for evaluation for a test for predicting progression from tuberculosis infection to active disease. **2017**. Available at: <https://apps.who.int/iris/handle/10665/259176>. Accessed 14 March 2019.