UC Irvine
UC Irvine Previously Published Works

Title

Blood Transcriptional Biomarkers for Active Tuberculosis among Patients in the United States: a Case-Control Study with Systematic Cross-Classifier Evaluation

Permalink

Journal

ISSN

Authors

Walter, Nicholas D
Miller, Mikaela A
Vasquez, Joshua
et al.

Publication Date

DOI

Peer reviewed

Blood Transcriptional Biomarkers for Active Tuberculosis among Patients in the United States: a Case-Control Study with Systematic Cross-Classifier Evaluation

Nicholas D. Walter,[a,b,c] Mikaela A. Miller,[c] Joshua Vasquez,[d] Marc Weiner,[e] Adam Chapman,[b] Melissa Engle,[e] Michael Higgins,[f] Amy M. Quinones,[g] Vanessa Rosselli,[c] Elizabeth Canono,[h] Christina Yoon,[d] Adithya Cattamanchi,[d] J. Lucian Davis,[d*] Tzu Phang,[b] Robert S. Stearman,[b*] Gargi Datta,[i] Benjamin J. Garcia,[i] Charles L. Daley,[j] Michael Strong,[i] Katerina Kechris,[c] Tasha E. Fingerlin,[c*] Randall Reves,[c,f] Mark W. Geraci[b*]

Pulmonary Section, Veterans Administration Medical Center, Denver, Colorado, USA[a]; Division of Pulmonary Sciences and Critical Care Medicine, University of Colorado Denver, Aurora, Colorado, USA[b]; Colorado School of Public Health, University of Colorado Denver, Aurora, Colorado, USA[c]; Division of Pulmonary and Critical Care Medicine, University of California, San Francisco, California, USA[d]; Department of Medicine, Veterans Administration Medical Center, San Antonio, Texas, USA[e]; Denver Metro Tuberculosis Control Program, Denver, Colorado, USA[f]; Denver Health Medical Center, Denver, Colorado, USA[g]; Clinical Translational Research Center, National Jewish Health, Colorado, USA[h]; Center for Genes, Environment, and Health, National Jewish Health, Denver, Colorado, USA[i]; Division of Mycobacterial and Respiratory Infections, National Jewish Health, Denver, Colorado, USA[j]

Blood transcriptional signatures are promising for tuberculosis (TB) diagnosis but have not been evaluated among U.S. patients. To be used clinically, transcriptional classifiers need reproducible accuracy in diverse populations that vary in genetic composition, disease spectrum and severity, and comorbidities. In a prospective case-control study, we identified novel transcriptional classifiers for active TB among U.S. patients and systematically compared their accuracy to classifiers from published studies. Blood samples from HIV-uninfected U.S. adults with active TB, pneumonia, or latent TB infection underwent whole-transcriptome microarray. We used support vector machines to classify disease state based on transcriptional patterns. We externally validated our classifiers using data from sub-Saharan African cohorts and evaluated previously published transcriptional classifiers in our population. Our classifier distinguishing active TB from pneumonia had an area under the concentration-time curve (AUC) of 96.5% (95.4% to 97.6%) among U.S. patients, but the AUC was lower (90.6% [89.6% to 91.7%]) in HIV-uninfected Sub-Saharan Africans. Previously published comparable classifiers had AUC values of 90.0% (87.7% to 92.3%) and 82.9% (80.8% to 85.1%) when tested in U.S. patients. Our classifier distinguishing active TB from latent TB had AUC values of 95.9% (95.2% to 96.6%) among U.S. patients and 95.3% (94.7% to 96.0%) among Sub-Saharan Africans. Previously published comparable classifiers had AUC values of 98.0% (97.4% to 98.7%) and 94.8% (92.9% to 96.8%) when tested in U.S. patients. Blood transcriptional classifiers accurately detected active TB among U.S. adults. The accuracy of classifiers for active TB versus that of other diseases decreased when tested in new populations with different disease controls, suggesting additional studies are required to enhance generalizability. Classifiers that distinguish active TB from latent TB are accurate and generalizable across populations and can be explored as screening assays.

Early and accurate diagnosis of tuberculosis (TB) is critical for control of the global TB epidemic. However, the sensitivity of existing tests is inadequate (1). Most existing tests for active pulmonary TB are based on the detection of *Mycobacterium tuberculosis* in sputum. However, the bacterium is not always detectable in sputum, and sputum is not always obtainable. Since blood can be obtained from nearly all patients undergoing evaluation for pulmonary and/or extrapulmonary TB, a diagnostic test based on human markers in blood would be optimal (1).

Human immune responses to *M. tuberculosis* may lead to transcriptional patterns in blood that are not present in other conditions (2–4). Selected sets of mRNA transcripts have been used in prediction models to classify patient samples as having active TB or not (5–11). In early phase studies (12), blood transcriptional classifiers have shown promise for active TB diagnosis (3, 10, 11, 13, 14) and for monitoring treatment response (15).

However, to move from proof of concept to an assay that is useful for TB control, important practical questions must be confronted. First, heterogeneity in diagnostic performance between populations is a well-known barrier to the development of diagnostic tests (16). Many genetic, environmental, and technical fac-

tors affect transcriptional biomarkers (3, 17). Would a transcriptional classifier developed among sub-Saharan Africans (11) be similarly accurate among ethnically diverse U.S. patients? Existing blood transcriptional classifiers for active TB vary in the transcripts included, the populations from which they were derived, and the prediction models used to derive them (5–11). To develop assays that can be implemented widely to improve TB control, it is critical to ascertain whether there is a universal, generalizable transcriptional pattern characteristic of active TB (3).

A second important practical question concerns the clinical usefulness of different types of blood transcriptional classifiers for active TB. Three distinct classifier types have been based on different reference groups. The first type compared active TB patients with systemically ill patients with other diseases (OD) that may mimic active TB (7, 11). The second type compared symptomatic TB patients with healthy subjects with latent TB infection (LTBI) (7, 11). The third type compared active TB patients with combined groups of systemically ill patients and healthy persons (including individuals with LTBI) (5, 8, 11). To develop transcriptional classifiers as a patient care and public health tool, it is critical to evaluate the limitations and potential clinical applications of each of these three classifier types.

We sought to develop a minimally invasive blood test to accurately diagnose active TB in racially and ethnically diverse populations. Therefore, we conducted a case-control study of blood microarray data among U.S. adults with and without active TB and compared the transcriptional classifiers we identified with previously published classifiers. We identified and tested three novel transcriptional classifiers that distinguish active TB from pneumonia (type 1), LTBI (type 2), and a combined group of LTBI or pneumonia (type 3). Additionally, to assess generalizability, we validated our transcriptional classifiers in an externally derived cohort of sub-Saharan Africans and systematically evaluated the accuracy of previously published transcriptional classifiers when tested in our U.S. patients. Finally, we assessed the limitations and potential clinical applications of the three classifier types.

MATERIALS AND METHODS

Study design. The expression in pneumonia and tuberculosis (ePAT) study was a prospective case-control study of HIV-uninfected adults with and without pulmonary TB in Colorado (Denver Health Medical Center) and in TB control programs in Texas (see the supplemental material). Three groups were enrolled. Adults with active TB were positive by sputum acid-fast bacillus (AFB) smear and *M. tuberculosis* culture. Adults with community-acquired pneumonia had cough or dyspnea, fever or leukocytosis, an infiltrate on chest radiograph, and a negative QuantiFERON-TB Gold In-Tube (Cellestis) (QFT) result. Adults with LTBI had a positive QFT result and no cough, fever, weight loss, or radiographic evidence of active TB. All patients were enrolled as outpatients; some were subsequently hospitalized. The Colorado Multiple Institutions Review Board and the University of Texas Health Sciences Center, San Antonio Institutional Review Board approved this project. All subjects provided written informed consent.

Laboratory analysis and data preprocessing. Blood was drawn in PAXgene Blood RNA tubes (Qiagen). RNA was extracted using the PAXgene Blood RNA kit (Qiagen). Specimens with an RNA integrity number of >7 via Bioanalyzer (Agilent) were considered acceptable. We selected 109 of 136 acceptable samples for microarray at random with stratification to achieve an approximate 1:1:1 ratio between groups. RNA was hybridized to an Affymetrix GeneChip Human Gene 1.1 ST array. After

TABLE 1 Clinical and demographic characteristics of U.S. patients in the ePAT study

Characteristic	No. (%) active TB (n = 35)	No. (%) LTBI (n = 35)	P value[a]	No. (%) pneumonia (n = 39)	P value[a]
Male	25 (71)	18 (51)	0.08	23 (59)	0.3
Foreign-born	22 (63)	30 (86)	0.06	13 (33)	0.01
Race/ethnicity					
Asian	6 (17)	14 (40)	0.06	2 (5)	0.1
Black	2 (6)	6 (17)	0.3	5 (13)	0.3
Latino	23 (66)	9 (26)	<0.001	16 (41)	0.03
White	3 (9)	1 (3)	0.6	11 (28)	0.04
Other	1 (3)	5 (14)	0.2	5 (13)	0.02
Age					
18–34	10 (29)	15 (43)		3 (8)	
35–49	7 (20)	13 (37)		16 (41)	
50–64	10 (29)	5 (14)		10 (26)	
>65	8 (23)	2 (6)		10 (26)	
Diabetes	12 (34)	3 (9)	0.02	9 (23)	0.3
Asthma or COPD	6 (17)	4 (11)	0.04	11 (28)	0.2
Current smoker	10 (29)	8 (23)	0.6	16 (41)	0.3

[a] χ^2 or Fisher's exact P value relative to TB group.

quality inspection, expression data were normalized via Robust Multichip Average (10) and were log-transformed and adjusted for batch effect (11). Differential expression between groups was estimated using the R limma package (version 3.20.1) with a Benjamini-Hochberg correction for multiple comparisons. Transcripts with an adjusted P value of <0.01 and an absolute fold change of >1.2 were considered significant.

Identification and evaluation of novel expression classifiers. We identified three different classifiers that distinguish active TB from pneumonia (type 1), LTBI (type 2), and a combined group of pneumonia or LTBI (type 3). Samples were randomly partitioned into training (2/3) and test (1/3) sets. Transcript (feature) selection methods are detailed in the supplemental material. Briefly, support vector machines with recursive feature elimination (SVM-RFE) (18) were applied to training set samples to identify the transcripts most predictive of active TB (e1071 package, version 1.6.3). To determine the number of transcripts in each classifier, we identified the point at which adding additional transcripts did not substantially improve classification (see the supplemental material) (19, 20). Accuracy in the training set was estimated across 20 iterations of 6-fold cross-validation. After transcript selection, a new SVM was fit to the training samples and was used to predict the class of test set samples (see the supplemental material).

For external validation, we used publically available data from the largest published whole-blood transcriptional classifier study by Kaforou et al. (11). Briefly, Affymetrix probe set identifiers were converted to Illumina transcript identifiers. Classification accuracy for an SVM using converted transcripts was estimated in 20 iterations of 6-fold cross-validation.

Evaluation of previously published expression classifiers. We systematically reviewed the published literature for transcriptional classifiers in adult human whole blood based on machine-learning classification of whole-genome microarray data (see the supplemental material). To determine if transcriptional patterns observed in previously published classifiers were also evident in our data, we compared the median expression of transcripts from each published classifier with the median expression of the same transcripts in our ePAT patients. To test previously published classifiers in our cohort, we converted transcript identifiers to Affymetrix probe set identifiers and estimated classification accuracy in cross-validation (6-fold, repeated 20 times) using the same classification method used in respective original manuscripts.

Microarray data accession number. Data are available in NCBI's Gene Expression Omnibus (accession number GSE73408).

TABLE 2 Accuracy of three types of whole-blood transcriptional classifiers for active TB[a]

Classifier/Set	No. of samples		Sensitivity, % (95% CI)	Specificity, % (95% CI)	AUC, % (95% CI)
	TB	Other diseases			
Type 1: Active TB versus pneumonia classifier					
ePAT type 1 classifier in ePAT patients					
Training set	24	24 (pneumonia)	89.8 (87.5–92.1)	87.9 (85.7–90.1)	96.5 (95.3–97.7)
Test set	11	15 (pneumonia)	100 (61.5–100)	80.0 (51.9–95.7)	90.1 (78.8–100)
External validation in African patients					
HIV negative	97	83 (OD)	90.2 (89.5–90.9)	77.7 (76.9–78.4)	90.6 (90.4–90.9)
HIV positive	102	88 (OD)	84.8 (84.0–85.7)	76.3 (75.3–77.3)	88.6 (88.1–89.1)
Previously published type 1 classifiers in ePAT patients					
Kaforou	35	39 (pneumonia)	69.7 (69.0–7–0.4)	79.1 (78.7–79.5)	82.9 (82.3–83.6)
Berry	35	39 (pneumonia)	91.3 (90.4–92.2)	73.7 (72.1–75.3)	90.0 (89.2–90.8)
Type 2: Active TB versus LTBI classifier					
ePAT type 2 classifier in ePAT patients					
Training set	24	24 (LTBI)	90.8 (90.0–91.6)	90.4 (88,4–92.4)	95.9 (95.2–96.6)
Test set	11	11 (LTBI)	100 (61.5–100)	81.8 (48.2–97.7)	98.4 (94.5–100)
External validation in African patients					
HIV negative	97	83 (LTBI)	90.4 (90.0–90.9)	86.4 (85.9–86.8)	95.3 (95.1–95.6)
HIV positive	100	82 (LTBI)	80.0 (79.2–80.8)	77.1 (76.4–77.8)	89.9 (89.6–90.2)
Previously published type 2 classifiers in ePAT patients					
Kafrorou	35	35 (LTBI)	93.9 (93.4–94.3)	92.4 (91.6–93.2)	98.0 (97.8–98.3)
Berry	35	35 (LTBI)	89.7 (88.9–90.5)	94.3 (94.3–94.3)	94.8 (94.3–95.4)
Type 3: Active TB versus LTBI or pneumonia classifier					
ePAT type 3 classifier in ePAT patients					
Training set	24	48 (LTBI or pneumonia)	81.9 (80.3–83.5)	79.0 (77.2–80.7)	85.9 (84.7–87.0)
Test set	11	26 (LTBI or pneumonia)	90.9 (58.7–99.8)	76.9 (56.4–91.0)	94.1 (86.5–100)
External validation in African patients					
HIV negative	117	146 (LTBI or OD)	87.7 (86.9–88.6)	80.3 (79.3–81.4)	91.1 (90.7–91.5)
HIV positive	121	153 (LTBI or OD)	79.6 (78.3–81.0)	74.0 (73.1–74.9)	85.4 (84.6–86.1)
Previously published type 3 classifiers in ePAT patients					
Kaforou	35	74 (LTBI or pneumonia)	77.4 (76.6–78.3)	75.7 (75.2–76.3)	82.8 (82.4–83.2)
Bloom	35	74 (LTBI or pneumonia)	85.7 (84.7–86.8)	76.8 (75.5–78.0)	90.5 (89.9–91.1)
Maertzdorf	35	74 (LTBI or pneumonia)	84.9 (83.6–86.1)	86.8 (85.9–87.6)	90.4 (89.8–90.9)

[a] For each classifier type, results are shown for ePAT training and test sets and external validation in Sub-Saharan African patients with and without HIV infection. Additionally, results are shown for previously published classifiers when tested in ePAT patients.

RESULTS

Enrollment. We screened 184 adults and enrolled 136, and we conducted microarray for 109 samples identified via stratified random selection (see Fig. S1 in the supplemental material). Similar to the demographics of TB reported in the United States (21), 63% of TB patients were foreign-born, from Southeast Asia, Africa, and Latin America (Table 1). Active TB patients more commonly reported diabetes (34%) than did patients with LTBI (9%; $P = 0.02$). Active TB patients were more commonly foreign-born (63%) than pneumonia patients (33%; $P = 0.01$).

Identification of novel ePAT transcriptional classifiers. (i) ePAT TB versus pneumonia classifier. Using pneumonia as a disease reference group, our novel ePAT type 1 classifier included 47 transcripts (see data set S1 in the supplemental material). Sensitivity was 89.8% in the ePAT training set and 100% in the test set (95% confidence intervals (CI) are presented in Table 2). Specificity was 87.9% in the training set and 80% in the test set. Figures 1A through D show receiver operating characteristic (ROC) curves.

For external validation, we applied our type 1 ePAT classifier to Sub-Saharan African subjects with and without HIV infection (11). Among HIV-uninfected Africans, sensitivity of the ePAT type 1 classifier was 90.2% and specificity was 77.7% (Table 2). The area under the concentration-time curve (AUC) was significantly lower among HIV-uninfected Sub-Saharan Africans (90.6%) than that among subjects in the ePAT training set (95.9%) ($P < 0.001$). Among Sub-Saharan Africans with HIV infection, sensitivity was 84.8% and specificity was 76.3%. AUC was 88.6%, which is significantly lower than that in the ePAT training set ($P < 0.001$).

Although the type 1 classifier was designed to distinguish active TB from pneumonia, we additionally tested how it would classify LTBI samples. It distinguished active TB from LTBI with 100% (95% CI, 85.5 to 100%) sensitivity but with only 40% (95% CI, 23.9 to 57.9%) specificity. AUC was 86.0% (95% CI, 76.5 to 95.5%).

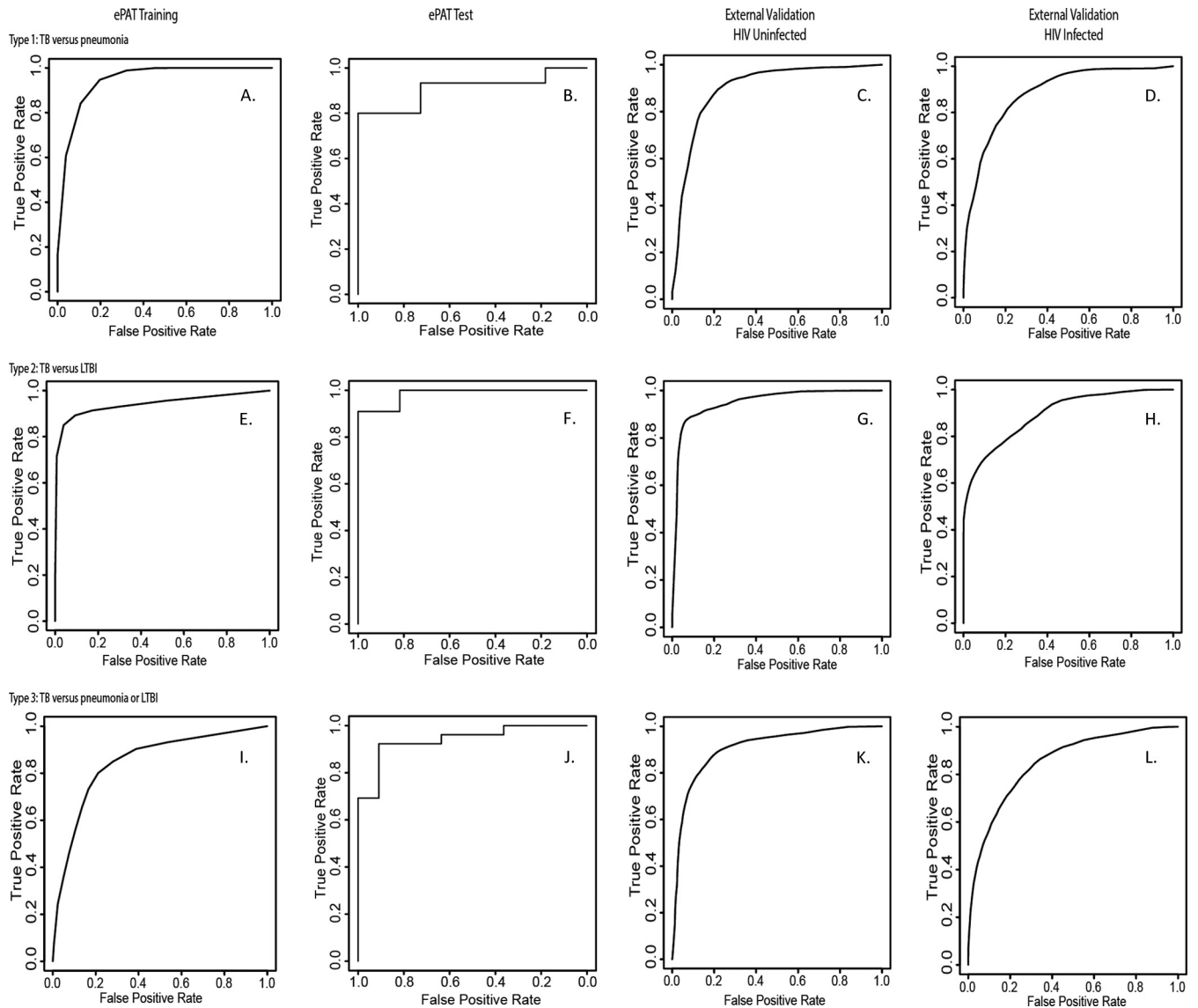**(ii) ePAT active TB versus LTBI classifier.** Our novel ePAT

FIG 1 ROC curves for three different transcriptional classifier types. Active TB versus pneumonia classifier in (A) ePAT training set, (B) ePAT test set, (C) external validation in HIV-uninfected Sub-Saharan Africans, and (D) HIV-infected Sub-Saharan Africans. Active TB versus LTBI classifier in (E) ePAT training set, (F) ePAT test set, (G) external validation in HIV-uninfected Sub-Saharan Africans, and (H) HIV-infected Sub-Saharan Africans. Active TB versus pneumonia or LTBI in (I) ePAT training set, (J) ePAT test set, (K) external validation in HIV-uninfected Sub-Saharan Africans, and (L) HIV-infected Sub-Saharan Africans.

active TB versus LTBI classifier included 51 transcripts (see data set S1 in the supplemental material). Sensitivity was 90.8% in the ePAT training set and 100% in the test set. Specificity was 90.4% in the training set and 81.8% in the test set. AUC was 95.9% in the training set and 98.4% in the test set. Figures 1E through H show ROC curves.

In external validation among Sub-Saharan African subjects without HIV infection (11), sensitivity of the ePAT type 2 classifier was 90.4% and specificity was 86.4% (Table 2). AUC was 95.3% and was not significantly different from the accuracy observed in U.S. ePAT subjects ($P = 0.2$) (Table 2). Among African subjects with HIV infection, AUC was 89.9%, which is significantly lower than that among U.S. ePAT subjects ($P < 0.001$).

To test the specificity of the type 2 classifier (designed to dis-

tinguish active TB from LTBI), we applied it to pneumonia and LTBI patients. The ePAT type 2 classifier categorized 35 (90%) of 39 pneumonia patients as active TB, suggesting that it detects the presence of systemic illness rather than a signature unique to active TB. When the ePAT type 2 classifier was applied to active TB and pneumonia patients, 34 (87%) of 39 pneumonia patients were classified as active TB. The transcripts differentially expressed in active TB relative to LTBI are highly overlapping with the transcripts altered in pneumonia relative to LTBI (Fig. 2). Of the 1,611 transcripts with higher expression in active TB than that in LTBI, 1,279 (79%) were also upregulated in pneumonia relative to LTBI.

(iii) ePAT active TB versus LTBI or pneumonia classifier. Our final classifier included 119 transcripts (see data set S1 in the supplemental material) that distinguished active TB from a com-
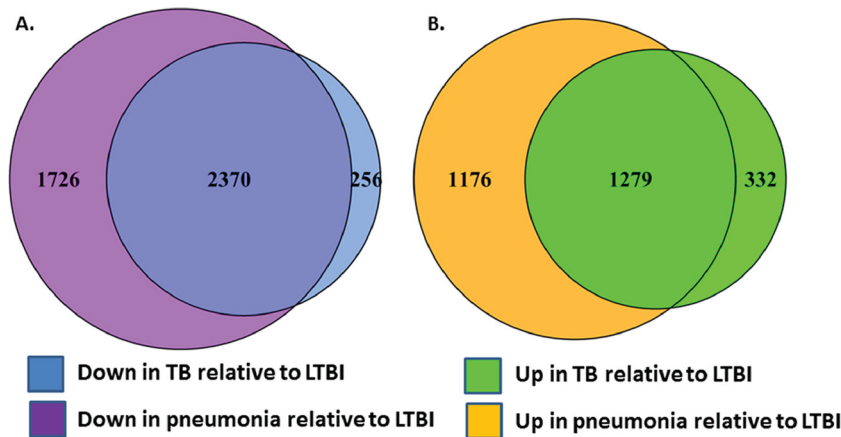
FIG 2 Venn diagram of transcripts differentially expressed in active TB relative to LTBI and pneumonia relative to LTBI. (A) Transcripts with significantly lower expression in active TB and pneumonia than that in LTBI (Benjamini-Hochberg adjusted P value of <0.01 and fold change of >1.2). (B) Transcripts with significantly higher expression in active TB and pneumonia than that in LTBI.

bined group of pneumonia or LTBI in ePAT samples. Sensitivity was 81.9% in the ePAT training set and 90.9% in the test set. Specificity was 79.0% in the training set and 76.9% in the test set (Table 2). The AUC was 85.6%, which is significantly lower than the AUC for classifier type 1 ($P < 0.001$) or type 2 ($P < 0.0001$).

In external validation, the type 3 ePAT classifier had higher accuracy among HIV-uninfected Sub-Saharan Africans than that among the U.S. ePAT subjects used for classifier development (AUC, 91.1% and 85.6%, respectively; $P < 0.001$) Among Sub-Saharan Africans with HIV infection, the AUC was 85.4%, which is not significantly different than that of the ePAT type 3 classifier in ePAT patients.

Evaluation of previously published transcriptional classifiers. A systematic literature review identified four published arti-

cles that developed classifiers for active TB based on adult human whole-blood microarray data (5, 7–9, 11) (see Table S1 and data set S2 in the supplemental material).

These articles included two type 1 classifiers. Kaforou et al. (11) compared active TB with a heterogeneous group of infections and neoplasms of the respiratory, genitourinary, and gastrointestinal tracts. When tested in U.S. ePAT patients, the Kaforou classifier had an AUC of 82.9%, which is significantly lower than that of the ePAT TB versus pneumonia classifier ($P < 0.001$) (Table 2). Berry et al. (7) compared active TB with various infectious and rheumatologic diseases. The Berry classifier had an AUC of 90.0%, which is also significantly lower than that of the ePAT active TB versus pneumonia classifier ($P < 0.001$).

There was little overlap (4.5%) in the transcripts included in

	Berry LTBI	Kaforou LTBI	ePAT LTBI	Berry OD	Kaforou OD	ePAT PNA	Kaforou LTBI+OD	Bloom LTBI+OD	Maertzdorf LTBI+OD	ePAT LTBI+OD
Berry LTBI	317	17	9	11	12	2	15	41	0	5
Kaforou LTBI		26	4	0	3	0	10	6	1	1
ePAT LTBI			51	3	0	0	3	5	0	13
Berry OD				80	1	3	2	0	0	3
Kaforou OD					42	2	19	2	0	2
ePAT PNA						47	2	0	0	14
Kaforou LTBI+OD							44	6	1	3
Bloom LTBI+OD								117	0	0
Maertzdorf LTBI+OD									69	2
ePAT LTBI+OD										119

FIG 3 Overlap in the transcripts included in seven previously published classifiers and the three novel transcriptional classifiers developed in this study. Gray cells represent the number of unique transcripts in each classifier after conversion to Affymetrix probe identifiers. Yellow shading indicates between-study overlap in classifiers for active TB versus those for LTBI. Pink shading indicates between-study overlap in classifiers for active TB versus those for other diseases (OD) or pneumonia (PNA). Blue shading indicates overlap in classifiers for active TB versus those for the combination of LTBI and other diseases.

different type 1 classifiers (Fig. 3). To determine whether these different transcripts might nonetheless reflect the same biologic processes, we used DAVID Bioinformatics Resources (22) to quantify the enrichment of gene ontology (GO) terms BP_Fat, CC_Fat, and MF_Fat in the classifier transcript lists. This did not reveal shared GO terms (see data set S3 in the supplemental material). Finally, side-by-side tornado plots provide visualization of global differences between studies in the expression of transcripts included in classifiers. Figure 4A displays the mean fold change for transcripts in the Kaforou type 1 classifier in Kaforou data. Figure 4B displays change in the same transcripts in ePAT data. The transcriptional change that distinguishes active TB from OD in Africans is not clearly discernible among ePAT patients. Side-by-side tornado plots for the Berry classifier are included in Fig. S2A and B in the supplemental material.

Our systematic review identified two type 2 classifiers (7, 11) that compared active TB to healthy controls with or without LTBI. When applied to ePAT data, the sensitivity and specificity of Berry (7) and Kaforou (11) type 2 classifiers were similar to that of the ePAT type 2 classifier (Table 2). The AUC for the Kaforou type 1 classifier in our ePAT data was 98.0%, which is significantly higher than that of the ePAT classifier in ePAT data ($P < 0.001$). The AUC for the Berry type 1 classifier was 94.8%, which is significantly lower than that of the ePAT classifier ($P < 0.001$).

The average overlap in the list of transcripts included in type 2 classifiers was 33%, which is higher than that for type 1 classifiers. Data set S3 in the supplemental material shows that type 2 transcript lists are enriched for similar GO terms, including "innate immune response," "response to wounding," and "defense response." In contrast to the type 1 classifiers, side-by-side tornado plots showed that the transcriptional changes that distinguished active TB from LTBI among Sub-Saharan African subjects in the Kaforou study were clearly observable among ePAT patients (Fig. 4C and D). Similarly, the transcriptional changes observed in the Berry data set were also evident in ePAT subjects (see Fig. S2C and D in the supplemental material).

Finally, to determine whether misclassification of LTBI patients as active TB is random or whether the same individual subjects are consistently misclassified by different classifiers, we evaluated the frequency with which each subject was misclassified in repeated cross-validation. In Fig. 5, the colored vertical bars indicate that 2 of 35 LTBI patients were consistently misclassified by all classifiers, suggesting that misclassification may result from an intrinsic difference in the patient sample rather than from random error.

Systematic review identified three previous type 3 classifiers (5, 8, 11) that compared active TB to a combined group with healthy/LTBI and other diseases. When tested in U.S. ePAT patients, the Kaforou type 3 classifier had an AUC of 82.8%, which is significantly lower than that of the ePAT type 3 classifier ($P < 0.001$). Bloom et al. (5) compared a group of patients with active TB with a combined group of healthy persons and pneumonia, cancer, and sarcoidosis patients. When tested in U.S. ePAT patients, the Bloom classifier had significantly higher AUC than that of the ePAT type 3 classifier (90.5% versus 85.6%; $P < 0.001$). Maertzdorf et al. (8) compared patients with active TB and sarcoidosis. When tested in ePAT patients, the AUC was 90.4%, which is also significantly higher than that of the ePAT type 3 classifier ($P < 0.001$).

The average overlap in the list of transcripts included in type 3
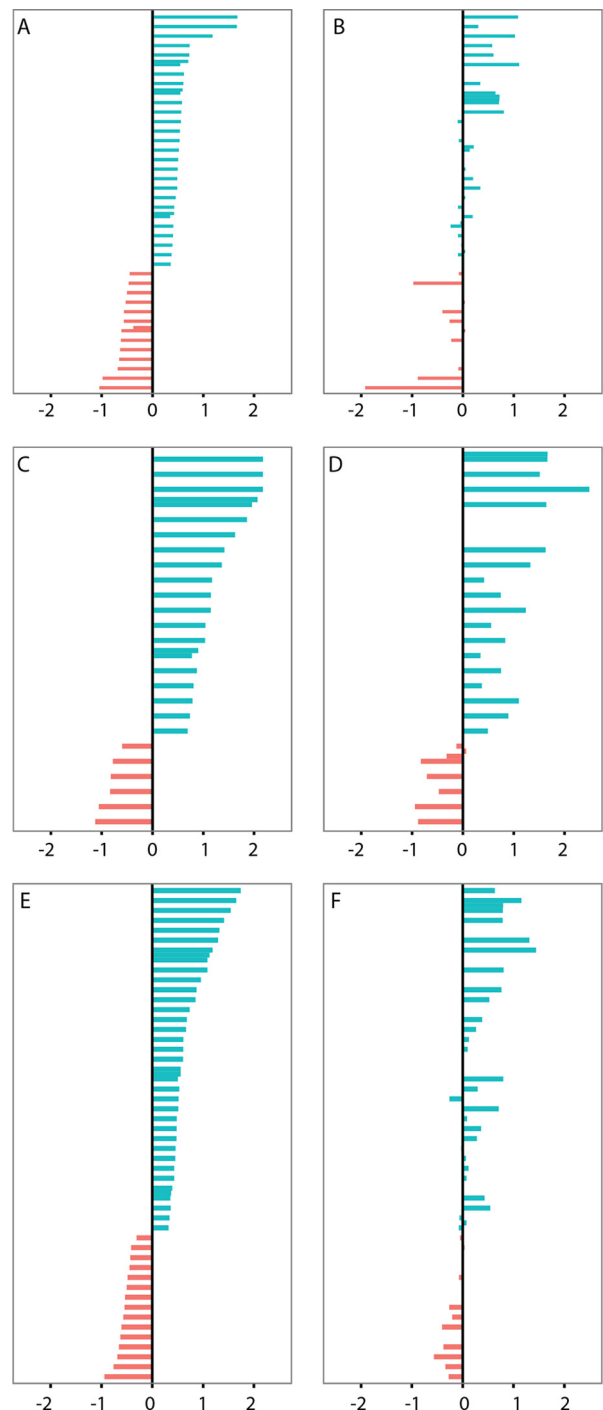

FIG 4 Comparison of transcriptional changes in Kaforou classifiers in the Kaforou study and ePAT. (A) Mean fold change (log₂ scale) of 44 transcripts in the Kaforou active TB versus other diseases classifier in the original Kaforou study. Rows represent individual transcript identifiers. Blue bars indicate transcripts with increased expression in active TB in the original manuscript and red represents decreased expression in active TB in the original manuscript. (B) Mean fold change for the same 44 transcripts among ePAT patients with active TB and pneumonia. (C) Mean fold change for 27 transcripts in the Kaforou active TB versus LTBI classifier in the original Kaforou study. (D) Mean fold change for the same 27 transcripts among ePAT patients. (E) Mean fold change for 53 transcripts in the Kaforou active TB versus LTBI or other diseases classifier in the original Kaforou study. (F) Mean fold change for the same 53 transcripts among ePAT patients.
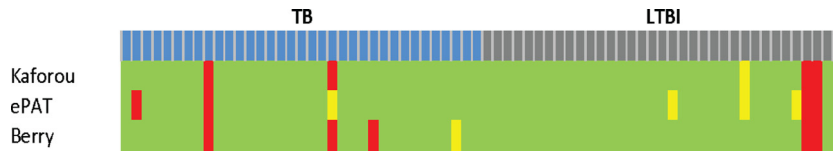
FIG 5 Consistency of misclassification of active TB and LTBI patients by previously published TB/LTBI classifiers and our ePAT classifiers. Columns represent individual subjects. Colored headers indicate the subjects' true class: blue, active TB; gray, LTBI. Green cells represent instances in which the classifier assigned the correct class with a high degree of certainty (probability of the correct class estimated as >0.66). Yellow cells represent instances in which the classifier did not have high certainty (probability of correct class estimated as between 0.33 and 0.66). Red cells represent instances in which the classifier assigned the incorrect class (misclassification) with a high degree of certainty (probability of correct class estimated as <0.33). The presence of columns of red and yellow suggest that different classifiers consistently misclassify certain individuals.

classifiers was 4.1%. As was the case for type 1 classifiers, side-by-side tornado plots indicated that the transcriptional change observed in previous studies was not clearly discernible among ePAT patients (Fig. 4A and B; see also Fig. S4A and B in the supplemental material).

DISCUSSION

We found that blood transcriptional classifiers accurately detected active TB among ethnically/racially diverse U.S. adults. Our type 1 classifier was highly accurate in a common clinical challenge of distinguishing active TB from pneumonia. However, the accuracy of classifiers for active TB versus that for other diseases decreased when applied to new populations with different combinations of other diseases. Type 2 classifiers that distinguish active TB from LTBI were highly accurate and generalizable across diverse populations. However, type 2 classifiers do not distinguish between active TB and pneumonia. They are nonspecific markers for systemic illness rather than signatures that are unique to active TB. Transcriptional classifiers have the potential to improve TB care and control, but additional discovery and validation studies in diverse populations with different disease references are needed to identify the most predictive transcripts and generalizable signatures.

Biomarker development is a multiphase process, starting with preclinical exploratory studies, expanding to case-control and cohort studies, and optimally culminating in randomized trials that demonstrate public health impact (12). It is essential to determine early in this process whether transcriptional classifiers are generalizable beyond the studies in which they were developed (12). In this study, we not only independently developed novel transcriptional classifiers in a previously unstudied population, but we also systematically evaluated the generalizability of our classifiers and previously published classifiers.

Type 1 classifiers address the high-priority clinical need of distinguishing active TB from other diseases that may mimic TB. We addressed the common clinical challenge of determining whether a patient with lower respiratory tract infection has active TB. Our classifier was highly accurate among U.S. patients. Accuracy diminished modestly (from an AUC of 96.5% to an AUC of 90.6%) when applied to Sub-Saharan African patients with different disease reference groups. Previously published classifiers were also less than optimally generalizable; accuracy was lower in ePAT patients than that reported in original manuscripts. Type 1 classifiers developed in different studies (with different "other disease" reference groups) had very little overlap in transcript sets. Analysis of GO terms did not identify shared biological processes in different classifier transcript sets.

The suboptimal generalizability represents a key challenge for identifying a robust and universal transcriptional signature for active TB versus other diseases (3). Unlike type 2 classifiers that compare active TB with a single reference condition (LTBI), a useful type 1 classifier would need to distinguish active TB from the protean range of infectious, neoplastic, and rheumatologic conditions that can mimic active TB. A classifier optimized to distinguish active TB from a single disease reference (such as lung cancer) may be less accurate in a different comparison (such as pneumonia). An essential next step will be identifying specific transcripts that are consistently predictive of active TB versus the range of clinical mimics across different populations. This will require additional independent whole-transcriptome classifier development in additional settings with additional disease comparison groups.

Type 2 classifiers distinguish active TB from LTBI. Symptomatic active TB is associated with a massive change in the blood transcriptome relative to healthy persons with LTBI, enabling accurate classification despite racial/ethnic diversity and medical comorbidities. Type 2 classifiers are remarkably accurate and generalizable across populations. When tested across different African, European, and U.S. settings, the AUC was consistently ~95% or higher. Studies conducted independently in different populations identified similar transcripts, representing similar immune and inflammatory processes. Tornado plots showed that the transcriptional patterns distinguishing active TB from LTBI were remarkably consistent across studies.

Unfortunately, type 2 classifiers do not represent a signature that is unique to active TB. The transcriptional changes observed in active TB relative to LTBI overlap nearly entirely with those observed in pneumonia relative to LTBI. It is therefore unsurprising that our type 2 classifier categorized nearly all pneumonia patients as having active TB. Type 2 classifiers essentially distinguish sick from healthy. Therefore, type 2 classifiers are not useful in evaluation of systemically ill patients; for sick patients, the relevant question is not whether the patient is sick or healthy but whether the patient has active TB or an alternative disease process (i.e., the task of type 1 classifiers).

Can the type 2 classifier be used to identify individuals with incipient, subclinical active TB? We found that several LTBI patients were consistently misclassified as TB in all type 2 models. It has been proposed that LTBI patients who are misclassified as having active TB may be at a transitional point between LTBI and active TB (7, 23). If a molecular signature of illness precedes the development of TB symptoms (24), type 2 classifiers might have value as a screening test for incipient active TB among persons such as household contacts of TB patients, health care workers in

high TB incidence settings, or HIV-infected persons. Since the type 2 classifier appears to be a nonspecific marker of illness, positive screening results would lead to more intensive TB evaluation. This important potential application has yet to be tested.

The third classifier type is an "all purpose" classifier designed to identify active TB among patients that are either healthy or systemically ill. As a composite of the previous two classifier types, type 3 classifiers combine but do not resolve the challenges outlined above. As was observed in the Kaforou study (11), our type 3 classifier was less accurate than our type 1 or type 2 classifier.

This study has several limitations. First, the limited sample size led to wide uncertainty intervals in the test set. Second, comparison between studies required conversion between microarray platforms, a methodological difference that tends to reduce generalizability (17). However, the use of different platforms makes the between-study consistency we observed in TB-LTBI classifiers an even stronger finding. Some transcript identifiers may not be converted because they have been retired from current annotation. Our conversion process therefore likely eliminated noise. Third, we did not enroll active TB patients that were sputum AFB smear-negative. Future studies should evaluate the accuracy of classifiers in this important population. Fourth, diabetes was more common among patients with active TB than it was among those with LTBI, potentially confounding the type 2 TB versus LTBI classifier. However, for 37 (73%) of the 51 transcripts in the type 2 classifier, adding diabetes as a covariate in limma models resulted in a <10% change in the TB/LTBI parameter, indicating no discernible confounding. Adding diabetes changed the TB/LTBI parameter by >25% for only one transcript, indicating that confounding was modest and present for only a subset of transcripts in the classifier. Finally, an obvious practical limitation is that there is currently no diagnostic platform that would make assays of blood transcriptional patterns feasible in settings with high TB incidence. Identification of an accurate, generalizable transcriptional classifier with a clear clinical application can motivate novel platform development.

In conclusion, blood transcriptional classifiers are capable of accurately identifying active TB. A remaining challenge for classifiers designed to detect TB among systematically ill patients is the breadth and heterogeneity of diseases that may mimic TB. Additional studies of systemically ill patients in diverse settings with systematic assessment of generalizability are needed. Classifiers comparing active TB and LTBI do not identify a signature that is unique to active TB but nonetheless should be explored as screening tools in high-risk asymptomatic or minimally symptomatic persons. Blood transcriptional assays that enable early accurate TB diagnosis may have an important impact on control of the global TB epidemic.

ACKNOWLEDGMENTS

REFERENCES

1. **Wallis RS, Kim P, Cole S, Hanna D, Andrade BB, Maeurer M, Schito M, Zumla A.** 2013. Tuberculosis biomarkers discovery: developments, needs, and challenges. Lancet Infect Dis **13:**362–372. http://dx.doi.org/10.1016/S1473-3099(13)70034-3.
2. **Joosten SA, Fletcher HA, Ottenhoff THM.** 2013. 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. http://dx.doi.org/10.1371/journal.pone.0073230.
3. **Maertzdorf J, Kaufmann SH, Weiner J, III.** 2014. Toward a unified biosignature for tuberculosis. Cold Spring Harb Perspect Med **5:**a018531.
4. **Kaforou M, Wright VJ, Levin M.** 2014. Host RNA signatures for diagnostics: an example from paediatric tuberculosis in Africa. J Infect **69**(Suppl):S28–S31. http://dx.doi.org/10.1016/j.jinf.2014.08.006.
5. **Bloom CI, Graham CM, Berry MPR, Rozakeas F, Redford PS, Wang Y, Xu Z, Wilkinson KA, Wilkinson RJ, Kendrick Y, Devouassoux G, Ferry T, Miyara M, Bouvry D, Valeyre D, Gorochov G, Blankenship D, Saadatian M, Vanhems P, Beynon H, Vancheeswaran R, Wickremasinghe M, Chaussabel D, Banchereau J, Pascual V, Ho LP, Lipman M, O'Garra A.** 2013. Transcriptional blood signatures distinguish pulmonary tuberculosis, pulmonary sarcoidosis, pneumonias and lung cancers. PLoS One **8:**e70630. http://dx.doi.org/10.1371/journal.pone.0070630.
6. **Bloom CI, Graham CM, Berry MPR, Wilkinson KA, Oni T, Rozakeas F, Xu Z, Rossello-Urgell J, Chaussabel D, Banchereau J, Pascual V, Lipman M, Wilkinson RJ, O'Garra A.** 2012. Detectable changes in the blood transcriptome are present after two weeks of antituberculosis therapy. PLoS One **7:**e46191. http://dx.doi.org/10.1371/journal.pone.0046191.
7. **Berry MPR, Graham CM, McNab FW, Xu Z, Bloch SA, Oni T, Wilkinson KA, Banchereau R, Skinner J, Wilkinson RJ, Quinn C, Blankenship D, Dhawan R, Cush JJ, Mejias A, Ramilo O, Kon OM, Pascual V, Banchereau J, Chaussabel D, O'Garra A.** 2010. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. Nature **466:**973–979. http://dx.doi.org/10.1038/nature09247.
8. **Maertzdorf J, Weiner J, III, Mollenkopf H-J, Network T, Bauer T, Prasse A, Müller-Quernheim J, Kaufmann SHE.** 2012. Common patterns and disease-related signatures in tuberculosis and sarcoidosis. Proc Nat Acad Sci U S A **109:**7853–7858. http://dx.doi.org/10.1073/pnas.1121072109.
9. **Lesho E, Forestiero FJ, Hirata MH, Hirata RD, Cecon L, Melo FF, Paik SH, Murata Y, Ferguson EW, Wang Z, Ooi GT.** 2011. Transcriptional responses of host peripheral blood cells to tuberculosis infection. Tuberculosis **91:**390–399. http://dx.doi.org/10.1016/j.tube.2011.07.002.
10. **Anderson ST, Kaforou M, Brent AJ, Wright VJ, Banwell CM, Chagaluka G, Crampin AC, Dockrell HM, French N, Hamilton MS, Hibberd ML, Kern F, Langford PR, Ling L, Mlotha R, Ottenhoff THM, Pienaar S, Pillay V, Scott JAG, Twahir H, Wilkinson RJ, Coin LJ, Heyderman RS, Levin M, Eley B.** 2014. Diagnosis of childhood tuberculosis and host RNA expression in Africa. N Engl J Med **370:**1712–1723. http://dx.doi.org/10.1056/NEJMoa1303657.
11. **Kaforou M, Wright VJ, Oni T, French N, Anderson ST, Bangani N, Banwell CM, Brent AJ, Crampin AC, Dockrell HM, Eley B, Heyderman RS, Hibberd ML, Kern F, Langford PR, Ling L, Mendelson M, Ottenhoff TH, Zgambo F, Wilkinson RJ, Coin LJ, Levin M.** 2013. Detection of tuberculosis in HIV-infected and -uninfected African adults using whole blood RNA expression signatures: a case-control study. PLoS Med **10:**e1001538. http://dx.doi.org/10.1371/journal.pmed.1001538.
12. **Pepe MS.** 2005. Evaluating technologies for classification and prediction in medicine. Stat Med **24:**3687–3696. http://dx.doi.org/10.1002/sim.2431.
13. **Satproedprai N, Wichukchinda N, Suphankong S, Inunchot W, Kuntima T, Kumpeerasart S, Wattanapokayakit S, Nedsuwan S, Yanai H, Higuchi K, Harada N, Mahasirimongkol S.** 2015. Diagnostic value of blood gene expression signatures in active tuberculosis in Thais: a pilot study. Genes Immun **16:**253–260. http://dx.doi.org/10.1038/gene.2015.4.
14. **Laux da Costa L, Delcroix M, Dalla Costa ER, Prestes IV, Milano M, Francis SS, Unis G, Silva DR, Riley LW, Rossetti MLR.** 2015. A real-time PCR signature to discriminate between tuberculosis and other pulmonary diseases. Tuberculosis **95:**421–425. http://dx.doi.org/10.1016/j.tube.2015.04.008.
15. **Cliff JM, Lee J-S, Constantinou N, Cho J-E, Clark TG, Ronacher K, King EC, Lukey PT, Duncan K, Van Helden PD, Walzl G, Dockrell HM.**

2013. Distinct phases of blood gene expression pattern through tuberculosis treatment reflect modulation of the humoral immune response. J Infect Dis **207:**18–29. http://dx.doi.org/10.1093/infdis/jis499.

16. **Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y.** 2012. Analyzing and presenting results. Cochrane handbook for systematic reviews of diagnostic test accuracy. The Cochrane Collaboration, Oxford, UK.

17. **Justice AC, Covinsky KE, Berlin JA.** 1999. Assessing the generalizability of prognostic information. Ann Intern Med **130:**515–524. http://dx.doi.org/10.7326/0003-4819-130-6-199903160-00016.

18. **Guyon I, Weston J, Barnhill S, Vapnik V.** 2002. Gene selection for cancer classification using support vector machines. Machine Learning **46:**389–422. http://dx.doi.org/10.1023/A:1012487302797.

19. **Hastie T, Tibshirani R, Friedman J.** 2009. Elements of statistical learning: data mining, inference and prediction, 2nd ed. Springer-Verlag, New York, NY.

20. **Krstajic D, Buturovic L, Leahy D, Thomas S.** 2014. Cross-validation pitfalls when selecting and assessing regression and classification models. J Cheminform **6:**10. http://dx.doi.org/10.1186/1758-2946-6-10.

21. **Centers for Disease Control.** 2014. Reported tuberculosis in the United States, 2013. U.S. Department of Health and Human Services. Centers for Disease Control, Atlanta, GA.

22. **Huang DW, Sherman BT, Lempicki RA.** 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc **4:**44–57. http://dx.doi.org/10.1038/nprot.2008.211.

23. **Barry CE, III, Boshoff HI, Dartois V, Dick T, Ehrt S, Flynn J, Schnappinger D, Wilkinson RJ, Young D.** 2009. The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. Nat Rev Microbiol **7:**845–855.

24. **Chaussabel D, Pulendran B.** 2015. A vision and a prescription for big data-enabled medicine. Nat Immunol **16:**435–439. http://dx.doi.org/10.1038/ni.3151.