Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: 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. DOI: 10.1056/NEJMoa1303657

Supplementary Appendix

Members of the study consortium

ILULU Consortium members: **Institute of Infectious Diseases and Molecular Medicine, University of Cape Town** Nonzwakazi Bangani, Lizl Bashe, Melina Carr, Hannah P. Gideon, Rene Goliath, Yekiwe Hlombe, Vanessa January, Bekekile Kwaza, Suzaan Marais, Marc Mendelson, Tolu Oni, Fadheela Patel, Ronnett Seldon, Relebohile Tsekela, Katalin A. Wilkinson, Robert J. Wilkinson, Kathryn Wood; **London School of Hygiene and Tropical Medicine/Karonga Prevention Study** Lyn Ambrose, Amelia C. Crampin, Hazel M. Dockrell, Neil French, Lumbani Munthali, Bagrey Ngwira, Amos Phiri, Femia Zgambo; **Red Cross War Memorial Children's Hospital, University of Cape Town** Margaret Cooper, Brian Eley, Mabel Gcuwa, Spasina King, Glynis Kossew, Karen McCabe, Wonita Petersen, Sandra Pienaar, Vashini Pillay; **Liverpool School of Tropical Medicine/Malawi-Liverpool-Wellcome Trust Clinical Research Programme**, **University of Malawi College of Medicine** Benjamin Allubha, George Chagaluka, Angeziwa Chunga, Janet Dube, Robert S. Heyderman, Annie Joabe, Martha Kalemba, Anne Kerr, Monica Matola, Rachel Mlotha, Agnes Mwale, David Mzinza; **Brighton and Sussex Medical School, University of Sussex** Suzanne T. Anderson, Gillian Baker, Claire M. Banwell, Terry Bishop, Natalie Chaplin, Julian Golland, Florian Kern, Susan Poore, Jayne Wellington; **Imperial College London** Lachlan J. Coin, Harieta Eleftherohorinou, Melissa Shea Hamilton, Myrsini Kaforou, Paul R. Langford, Michael Levin, Stephanie Menikou, Victoria J. Wright.

KIDS study group: Victor Bandika, Jay Berkley, Andrew Brent, Joyce Langat, Kath Maitland, Susan Morpeth, Daisy Mugo, Carole Mulunda, Robert Musyimi, Agnes Mutiso, Charles Newton, John Paul Odhiambo, Anthony Scott, Hemed Twahir, Joshua Wambua.

Supplementary Methods

Study sites

Children were recruited to the discovery cohort at study sites in South Africa and Malawi, and to the validation cohort at a third study site in Kenya.

Red Cross War Memorial Children's Hospital, Cape Town, South Africa

The Red Cross War Memorial Children's Hospital is a tertiary referral hospital in the Western Cape Province of South Africa (SA), serving a predominantly urban population in and around Cape Town. South Africa has one of the highest pediatric TB incidence rates worldwide (981 per 100,000)[1], as well as one of the most widespread national pediatric HIV epidemics.[2,3] Despite >98% infant BCG vaccination coverage, there is a high incidence of disseminated TB including TB meningitis.[4] Malnutrition and bacterial and helminth infections are common, but there is no malaria transmission in the Western Cape.

Queen Elizabeth Central Hospital, Blantyre, Malawi

The Queen Elizabeth Central Hospital is the tertiary public health facility for Blantyre, the major commercial centre of Malawi (MLW). The national incidence of TB in Malawi was 191 per 100,000 in 2011, with a national HIV prevalence of 11%.[5] Malaria and malnutrition are also endemic, and bacterial and helminth infections are common. Neonatal BCG vaccination coverage is estimated at 90%.[5]

Kilifi District Hospital & Coast Provincial General Hospital, Coast Province, Kenya

Coast Provincial General Hospital (CPGH) is the tertiary public health facility for Kenya's Coast Province. Situated in the port city of Mombasa, it is the second largest hospital in the country. Kilifi District Hospital (KDH) provides primary and secondary level care to Kilifi District, also in Coast Province. Both hospitals serve a mixed rural and urban population in an area where malaria[6] and malnutrition[7] are endemic, and invasive bacterial[8] and helminth infections[6] common. HIV prevalence among women attending antenatal services is 4.4%[7] and infant BCG vaccination coverage is estimated at 96%.[7]

Patient screening and recruitment

The process of patient screening and recruitment differed slightly between the study sites.

Discovery cohort

An operational approach to patient screening was instituted at the sites in South Africa and Malawi, whereby children were referred to the study team for investigation if the clinician looking after the child suspected a diagnosis of TB. All these children were eligible for inclusion in the study and further investigations and case management were as described in the main paper.

Validation cohort

Much broader screening criteria were used in the validation cohort as this formed part of a larger study of the epidemiology and diagnosis of pediatric tuberculosis. All children admitted to hospital were formally screened for a persistent cough for >2 weeks; pneumonia not responding to first line antibiotics; unexplained fever for >2 weeks; unexplained progressive weight loss or failure to thrive for >4 weeks; a history of close TB contact; and/or a doctor's clinical suspicion of TB for any other reason. Children with one or more of these features, plus those referred for outpatient investigation for TB and children <5 years old who were identified as household TB contacts of smear positive pulmonary TB, were then eligible for inclusion in the study. Further investigations and case management were as described in the main paper.

Laboratory diagnosis

Two spontaneous or induced sputum samples were examined by standard microscopy for acid-fast bacilli (AFB) and cultured for mycobacteria using the BD MGIT liquid culture technique. Isolation of MTB was confirmed by microscopic cording, MTB-64 lateral flow assays (Capilia®; TAUNS Laboratories, Inc., Numazu, Japan) and growth on p-nitrobenzoic acid (Malawi), plus specific PCR (SA and Kenya).

RNA sample extraction and processing

Whole blood (2.5ml) was collected into PAXgene™ blood RNA tubes (PreAnalytiX, Germany), incubated for 2 hours, frozen at -20^{o}C within 6 hours of collection, and stored at -80^{o}C. RNA was extracted using PAXgene™ blood RNA kits (PreAnalytiX, Germany) according to the manufacturer's instructions at one site (Cape Town for the discovery cohort or London for the Kenyan samples) to minimize sample handling bias. The integrity and yield of the total RNA was assessed using an Agilent 2100 Bioanalyser and a NanoDrop 1000 spectrophotometer. Total RNA was then shipped to the Genome Institute of Singapore. After quantification and

quality control, biotin-labeled cRNA was prepared using Illumina TotalPrep RNA Amplification kits (Applied Biosystems) from 500ng RNA. Labeled cRNA was hybridized overnight to Human HT-12 V4 Expression BeadChip arrays (Illumina). After washing, blocking and staining, the arrays were scanned using an Illumina BeadArray Reader according to the manufacturer's instructions. Using Genome Studio software the microarray images were inspected for artifacts and QC parameters were assessed. No arrays were excluded at this stage. Fourteen samples were excluded in total: 11 due to insufficient RNA after processing, 1 due to discrepant labeling, 2 removed at data QC in Principal Components Analysis (PCA).

C-reactive protein measurements

CRP was measured by ELISA (Invitrogen ELISA kit #KHA0031) on serum collected at the same time as blood for RNA expression.

Supplementary Statistical Methods

Microarray analysis

Mean raw intensity values for each probe were corrected for local background intensities and a robust spline normalisation[9] (combining quantile normalisation and spline interpolation) was applied to each array. Expression values were transformed to a logarithmic scale (base 2). PCA was used as part of the quality control process of the arrays before the split into 80%-20% for the identification of signatures. PCA is an approach that allowed us to summarize our data and reduce the dimensionality (536 arrays x 48,000 probes, down to 536 arrays x no of principal components) in order to explore variance in the expression level.[10] RNA expression profiles of most children in the discovery cohort clustered together on PCA; two outlying samples were removed from the analysis (Figure S2). At the two first principal components there was no variance introduced because of location or HIV status of the samples (Figure S2). Using the 2-dimensional equivalent of the t-statistic, the Hotelling test[11], we removed two samples before the analysis (categorized as TB/HIV+ and OD/HIV+ from Malawi). The samples were divided into a training set (n_{TB}=87 n_{OD}=134 HIV+/-; n_{TB}=56 n_{OD}= 82 HIV-, n_{TB}=31 n_{OD}=52 HIV+, n_{LTBI}=43 HIV-) and test set (n_{TB}=23 n_{OD}=34 HIV+/-; n_{TB}=14 n_{OD}= 21 HIV-, n_{TB}=9 n_{OD}=13 HIV+, n_{LTBI}=11 HIV-). Using the training set, we identified the transcripts that were differentially expressed between patient groups with $|\log_2 FC| > 0.5$, which were taken forward to variable selection with elastic net.[12] This threshold was chosen in order to ensure that differential expression for selected variables could be distinguished using the resolution of qtPCR. The **a** and **λ** parameters of elastic net, which control the

size of the selected model, were optimized via ten-fold cross-validation (CV). The weights assigned by elastic net to the trained model were used within a linear regression model to classify samples in the test set.

Disease risk score

For each individual, we calculated the disease risk score using the minimal transcript selected sets for TB vs. LTBI and TB vs. OD. The score is based on subtracting the summed intensities of the down-regulated transcripts from the summed intensities of the up-regulated transcripts. The disease risk score for individual i is:

$$Disease\ Risk\ Score^{i} = \sum_{k=0}^{n} expr.value_{k}^{i} - \sum_{l=0}^{m} expr.value_{l}^{i} \qquad (1)$$

where: **n** the number of up-regulated number of probes in the signature in disease of interest (TB) compared to comparator group(s).

m the number of down-regulated number of probes in the signature in disease of interest (TB) compared to comparator group(s).

The threshold for the classification was calculated as the weighted average of risk score within each class, with weights given as inverse of the standard deviation of the score within each class. The threshold for the classification between group *u* and *v* is shown below:

$$threshold(u,v) = \frac{\frac{\mu_u}{\sigma_u} + \frac{\mu_v}{\sigma_v}}{\frac{1}{\sigma_u} + \frac{1}{\sigma_v}} \qquad (2)$$

where: **μ** average of the disease risk score in the group.

σ standard deviation of the disease risk score in the group.

Analysis of validation dataset

The microarray analysis for the Kenyan validation cohort was done as previously described, but the raw microarray data were pre-processed (background subtracted and normalized) separately from the discovery cohort. We then calculated the disease risk scores, based on the signatures derived in the discovery cohort, for the samples of the Kenyan cohort to evaluate their performance in an independent validation cohort.

In order to evaluate the IGRA+ OD patients who may have either self resolving primary TB or latent TB infection we performed the TB vs. OD comparison both with and without inclusion of the IGRA+ patients. There were 9 IGRA+ patients in the OD group randomly selected for array. The sensitivity of DRS for TB vs. OD remained unchanged with or without the IGRA+ patients, while specificity was 1% lower for when the IGRA+ patients were included. 7 of the 9 OD patients that were IGRA positive were classified as not TB by the DRS. As sensitivity of DRS was unchanged when IGRA positive patients were included or excluded, and to exclude possibility of including self resolving primary TB in the OD group, for calculation of performance of DRS in culture negative group we used only IGRA negative OD cases.

Calculation of effective sensitivity in culture-negative TB groups

In order to obtain more realistic estimates of the test sensitivity across the culture-negative TB categories, we recognized that each category is a mixture of “actual” TB cases and OD clinically confused with TB. We therefore, modeled the observed true-positive rate (TPR) as a function of the unknown actual TPR, the false-positive rate (FPR) estimated from the OD group, and the prevalence of TB (Supplementary Methods Eq. 3), from which we calculated a corrected Receiver Operator Characteristic (ROC) curve and estimates of 'effective' sensitivity in each category. As the prevalence of TB in each category is unknown, we investigated a range of prevalence of 70%-90%, 40%-60% and 30%-50% for “highly probable”, “probable” and “possible” TB respectively and also present unadjusted results which are equivalent to assuming a TB prevalence of 100% in each category.

Application of a classifier, such as DRS, to the culture-negative TB group results in an observed estimate of the true-positive rate (TPR_{obs}), which is the proportion of all observed culture–negative TB cases(P_{obs}) scored as 'positive' by the classifier. However, these observed positives are in fact a mixture of actual true TB and false TB (i.e. OD), hence

$$TPR_{obs} = \frac{TP_{obs}}{P_{obs}} = \frac{TP_{actual} + FP_{actual}}{P_{actual} + F_{actual}} = \frac{TPR_{effective} * P_{actual} + FPR_{effective} * F_{actual}}{P_{actual} + F_{actual}}$$

$$= TPR_{effective} * \frac{P_{actual}}{P_{actual} + F_{actual}} + FPR_{effective} * \frac{F_{actual}}{P_{actual} + F_{actual}}$$

$$= TPR_{effective} * \Pr(TB) + FPR_{effective} * (1 - \Pr(TB)) \qquad (3)$$

where: F_{actual} is the number of OD and Pr(TB) is the prevalence of true TB and in the group under consideration. $FPR_{effective}$ is the false-positive rate, estimated as the proportion of OD cases that are falsely called TB by the classifier. We can re-arrange equation (3) to obtain a formula for the effective TPR in terms of the group prevalence and the FPR estimated from the OD group:

$$TPR_{effective} = \frac{TPR_{obs} - FPR_{actual} * (1 - \Pr(TB))}{\Pr(TB)} \qquad (4)$$

Positive and negative predictive value for combined culture positive and culture negative TB

We calculated the positive and negative predictive value (PPV and NPV) as a function of specificity sensitivity and prevalence according to the following formulae:

$$NPV = \frac{specificity * (1 - prevalence)}{specificity * (1 - prevalence) + (1 - sensitivity) * prevalence} \qquad (5)$$

$$PPV = \frac{sensitivity * prevalence)}{sensitivity * prevalence + (1 - sensitivity) * (1 - prevalence)} \qquad (6)$$

Given the dependency of NPV/PPV on test sensitivity, specificity and prevalence, it is important to provide estimates of these values specific to scenarios in which such a diagnostic test would be applied. We have calculated these values for a scenario in which a child presents to a clinic with symptoms consistent with TB, and thus we use the specificity as reported in Table 2 for the HIV-infected and -uninfected combined other disease group from the Kenyan validation set.

We use a test sensitivity estimate derived on the combined culture-positive and culture-negative TB groups. We estimated this as a weighted average of the 'effective' sensitivity in the culture-confirmed, “highly probable” (HP), “probable” (Pr) and “possible” (Pos) TB, with the weights given by the proportion of samples in the Kenyan prospective study which were assigned to each of these groups. The effective sensitivity in each

subgroup was calculated using equation 4, based on the same range of assumptions on the prevalence of 'actual' TB in each group used to calculate effective sensitivities in Table S7. In more detail, the scenarios considered are:

- A: TB prevalence in: culture confirmed = 100%; HP TB = 70%; Pr TB =40%; Pos TB = 30%
- B: TB prevalence in: culture confirmed = 100%; HP TB = 80%; Pr TB =50%; Pos TB = 40%
- C: TB prevalence in: culture confirmed = 100%; HP TB = 90%; Pr TB =60%; Pos TB = 50%

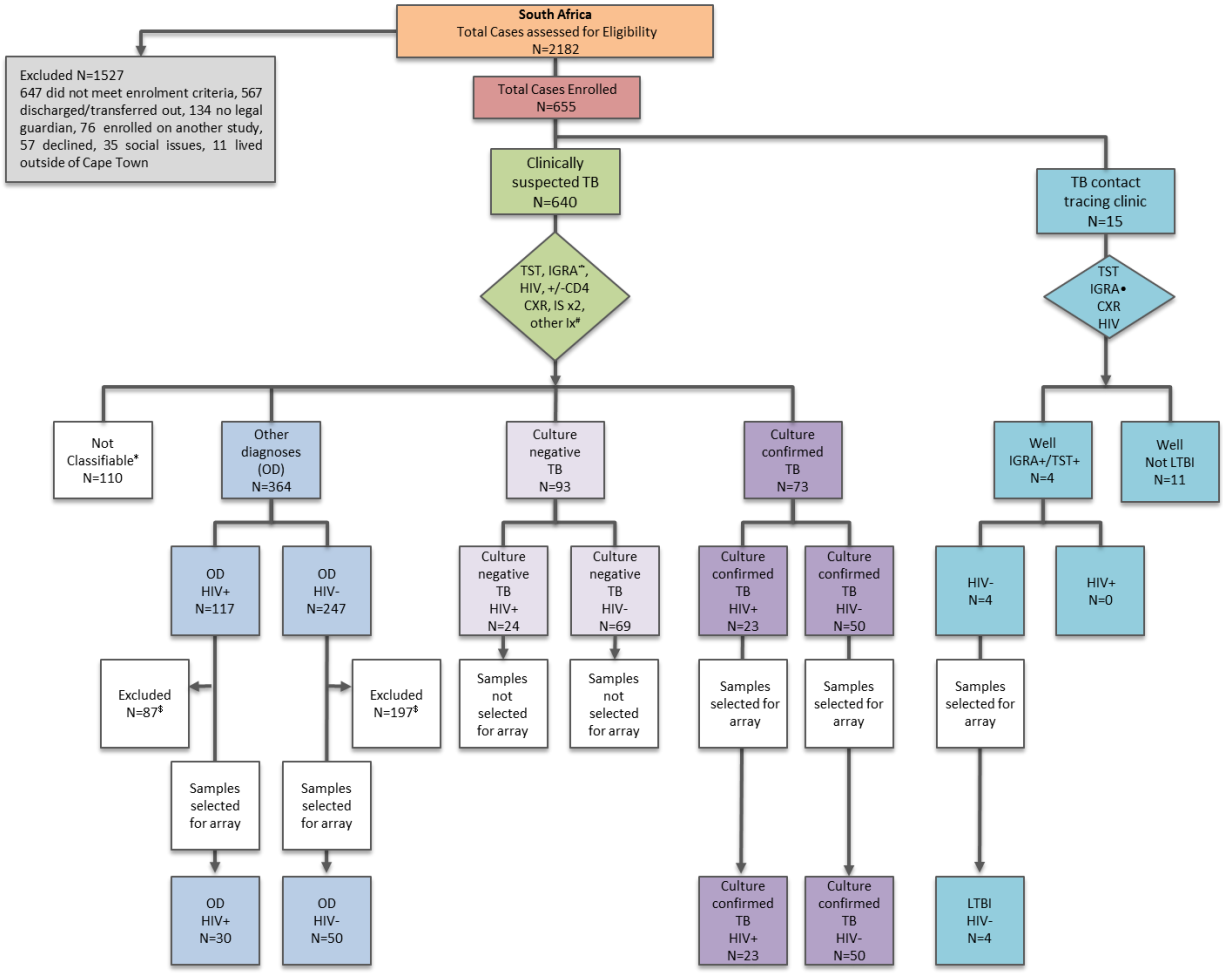
Also recognizing that the prevalence of TB among the population tested will depend on the operational clinical strategy for TB screening, we used a range of estimates for TB population prevalence in the patients screened (10%, 30% and 50%).

- 10%: reflects the prevalence of TB in the Kenyan cohort
- 30%: reflects the prevalence of TB in the South Africa and Malawi recruitment
- 50%: reflects a scenario that clinicians would do prior filtering or combining with another test

The proportion of proven TB amongst patients suspected of having TB varies depending on the strategy for TB investigation. In South Africa and Malawi, patients were investigated for TB if the clinicians responsible for the child's care considered TB to be included in the differential diagnosis. In Kenya, a systematic screening process was undertaken for all children with cough, fever or weight loss of > 2 weeks duration. The difference in approach resulted in different proportions of TB cases, with the broad criteria used in Kenya being reflected in a lower proportion of TB. As our research setting actively sought to identify TB cases, it is likely that in a non-research setting in typical African hospitals, patients selected to undergo investigation for TB may make up a higher proportion of those investigated, hence our exploration of 50% prevalence scenario.

Supplementary Figures

Figure S1a. Recruitment at Red Cross War Memorial Children's Hospital, Cape Town, SA.



¨IGRA performed at baseline and 3 months in the OD category & at baseline and where possible 3 months in the TB cases category.

#investigations done at attending clinician's discretion to diagnose OD's (urine, cerebrospinal fluid, blood cultures) as well as additional investigations performed to diagnose TB (ultrasound scans, CT-scans, histology and cytology).

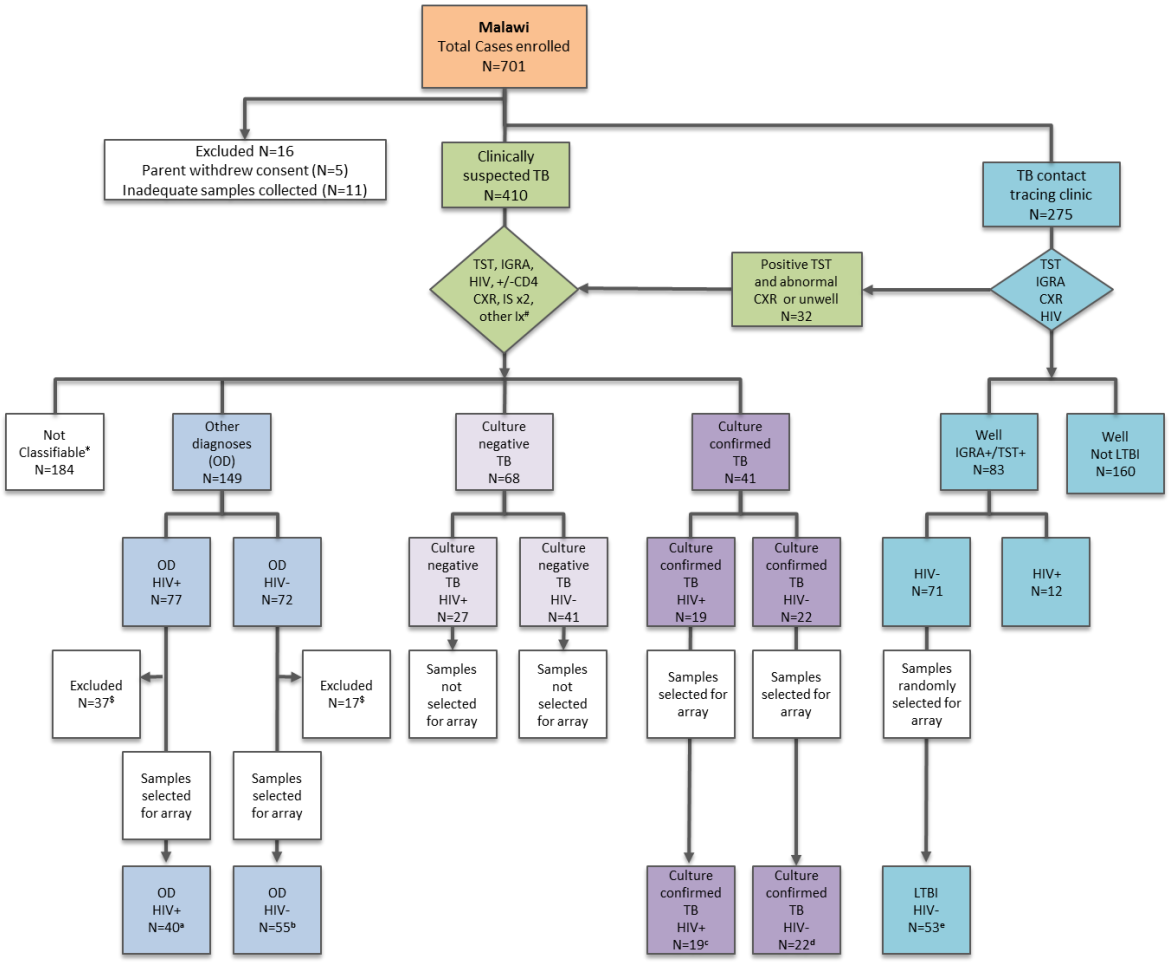
•IGRA performed at baseline and at 3 months.

*cases excluded due to inconclusive/inadequate investigations at baseline.

$cases excluded due to inconclusive diagnoses/patients lost to follow-up.

Among HIV-uninfected and -infected definite TB cases, 76.0% and 56.5% of samples respectively were smear negative on microscopy.

Figure S1b. Recruitment at Queen Elizabeth Central Hospital, Blantyre, Malawi.



#Investigations done at attending clinicians discretion to diagnose ODs (urine, CSF, blood cultures, histology, malaria thick film); additional investigations performed to diagnose TB (ultrasound scans, MRI-scans, TB blood culture, histology).

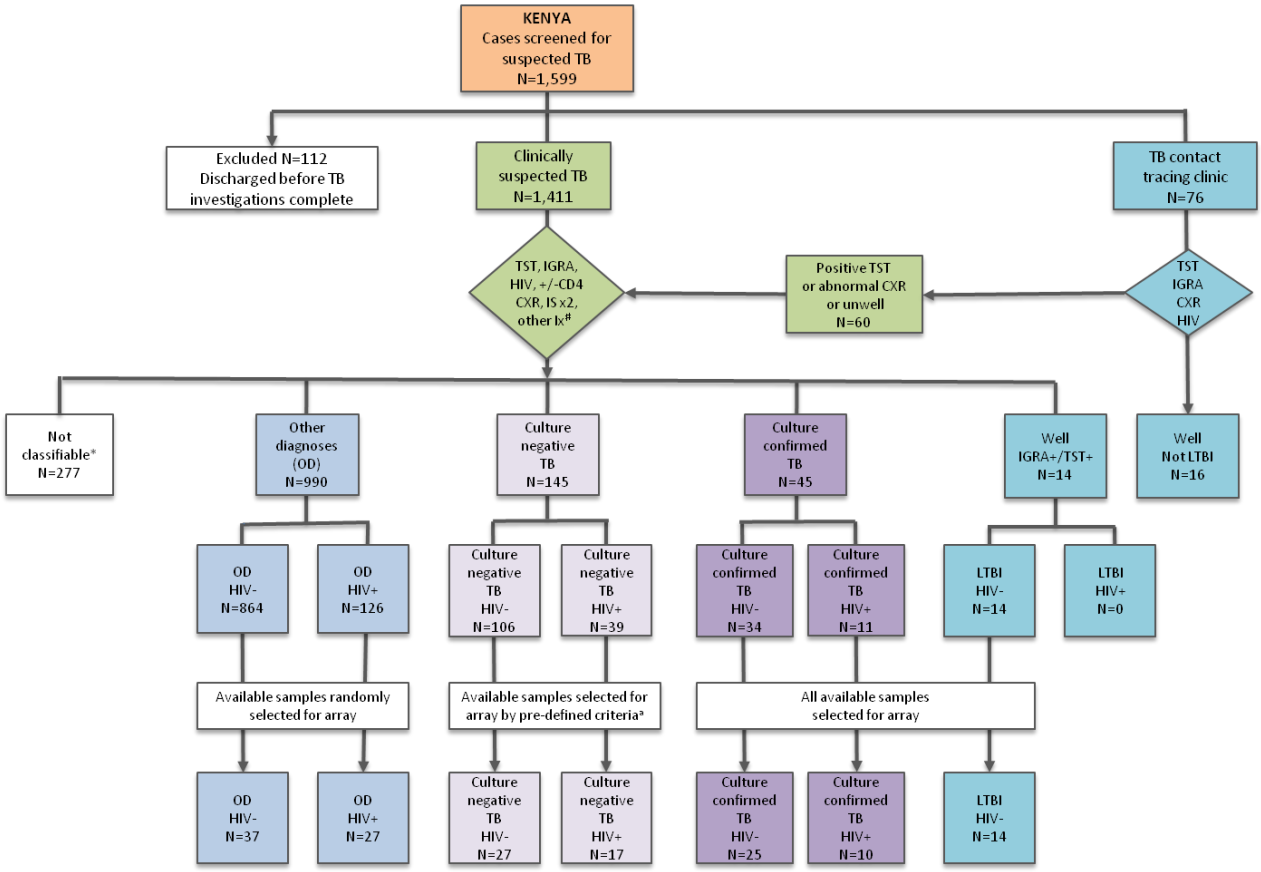
***Cases excluded due to inconclusive/inadequate investigations at baseline.

$Samples excluded because of inconclusive diagnoses.

a,b,c,d,e 4, 2, 1, 2, 3 samples respectively lost during sample processing.

Among HIV-uninfected and -infected definite TB cases, 50% and 54% of samples respectively were smear negative on microscopy.

Figure S1c. Recruitment of validation cohort at Kilifi District Hospital & Coast Provincial General Hospital, Coast Province, Kenya.



♯ Additional investigations done at the attending clinician's discretion to aid diagnosis of TB or ODs included: thick and thin films for malaria; blood cultures; urine cultures; CSF microscopy, culture, bacterial antigen tests and biochemistry; culture of pleural, peritoneal, joint and abscess fluid; bone marrow biopsy; radiological imaging including ultrasound and computed tomography scans; and tissue biopsy for histology and culture).

*Cases that were not classifiable were those not treated for TB in whom TB could be neither diagnosed nor excluded with confidence due to death or loss to follow up.

[a] see methods.

Figure S2. Principal components analysis of the microarrayed samples in the discovery cohort (South Africa and Malawi).

Principal components analysis (PCA) plot of PCA1 & PCA2 based on all genes on all of the samples after background adjustment and normalisation. Two samples indicated by the arrows (a TB/HIV+ and an OD/HIV+ case from Malawi) were removed from the analysis. Confidence ellipses calculated for the population mean are shown below (0.999 inner circle, 0.9999 outer circle).

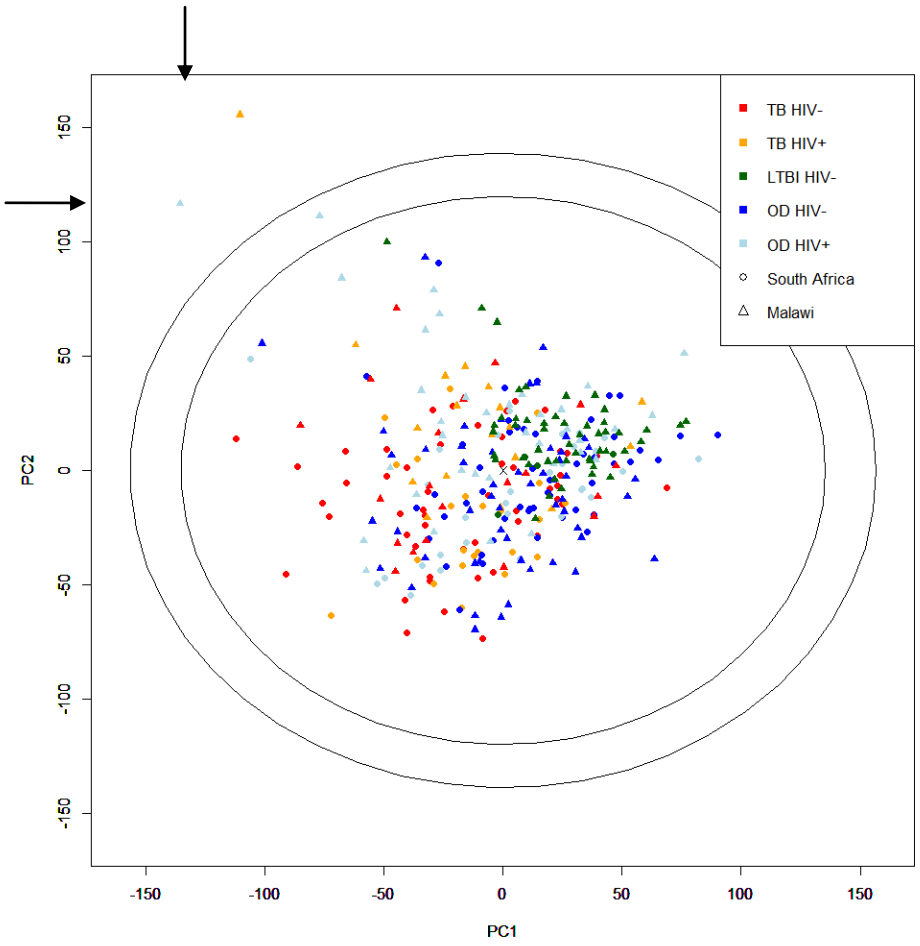


Figure S3. Disease risk scores and Receiver Operator Characteristic curves based on the TB/OD 51-transcript signature applied to the South African (SA)/Malawi HIV+/- test set and the Kenyan validation cohort.

Sensitivity, specificity are reported in Table 2. Test set: nTB=23 nOD=34; Validation cohort: nTB=35, nOD=55; Validation cohort including IGRA+ patients: nTB=35, nOD=64. 7 out of the 9 IGRA+ OD patients are classified as OD and 2 as TB (red arrows) using the DRS.

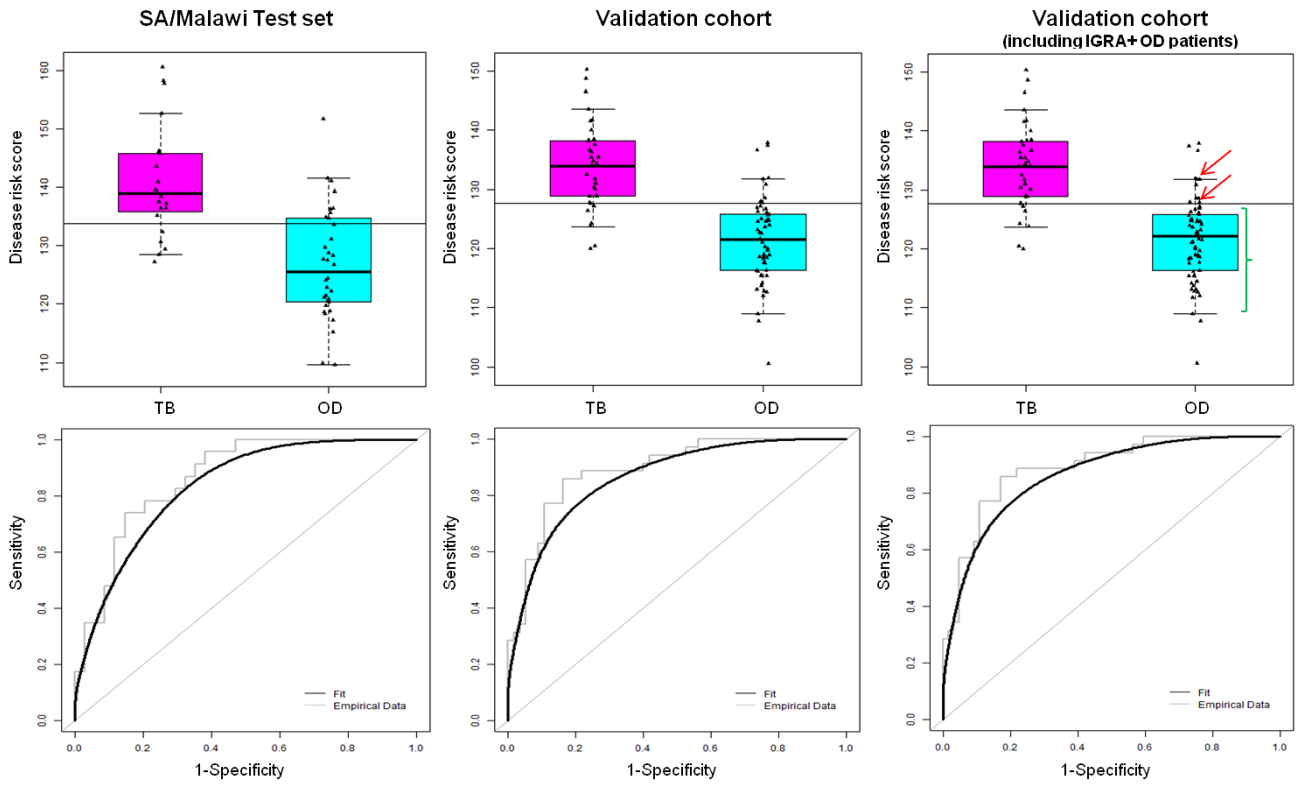


Figure S4. Disease risk scores and Receiver Operator Characteristic curves based on the TB/LTBI 42-transcript signature applied to the South African (SA)/Malawi HIV+/- test set and the Kenyan validation cohort.

Sensitivity, specificity are reported in Table S3. Test set: nTB=23 nLTBI=11; Validation cohort: nTB=35, nLTBI=14.

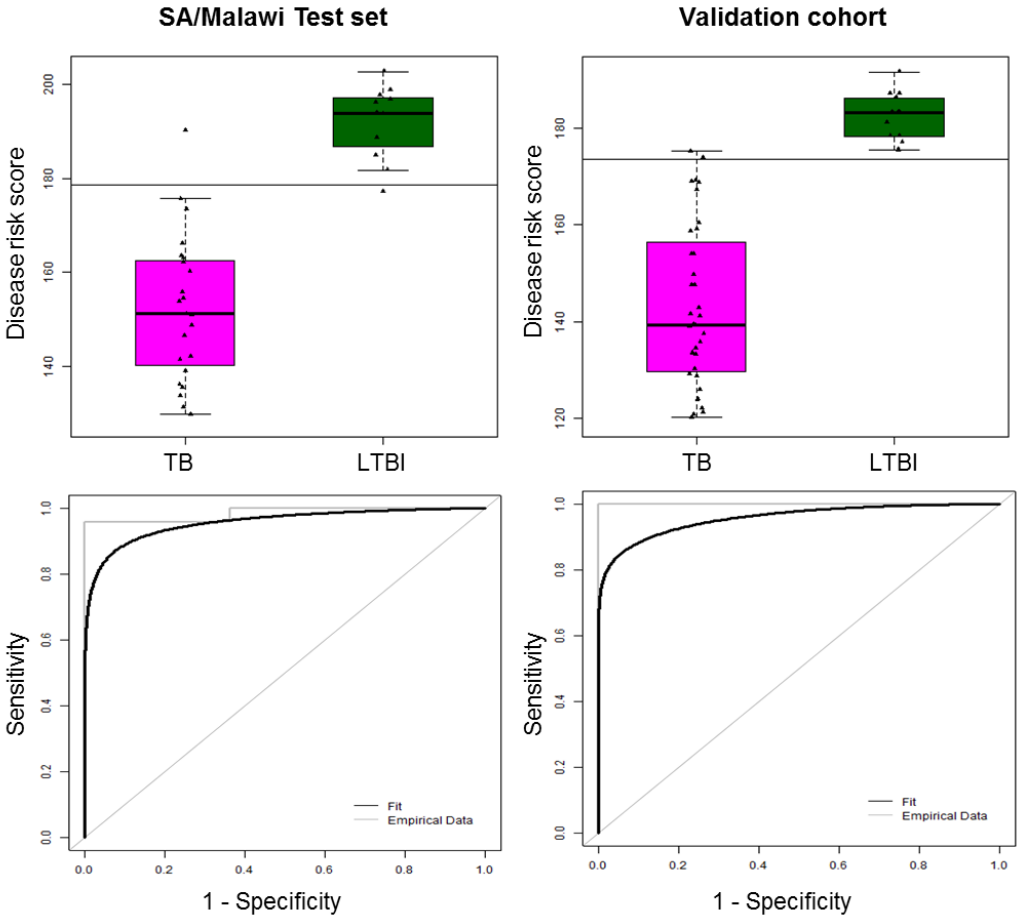


Figure S5. Disease risk scores and Receiver Operator Characteristic curves based on the TB/OD 51-transcript signature applied to the Kenyan cohort by HIV status.

Sensitivity, specificity are reported in Table S6. Validation cohort: nTB=25 nOD= 29 HIV-uninfected, nTB=10 nOD=26 HIV-infected.

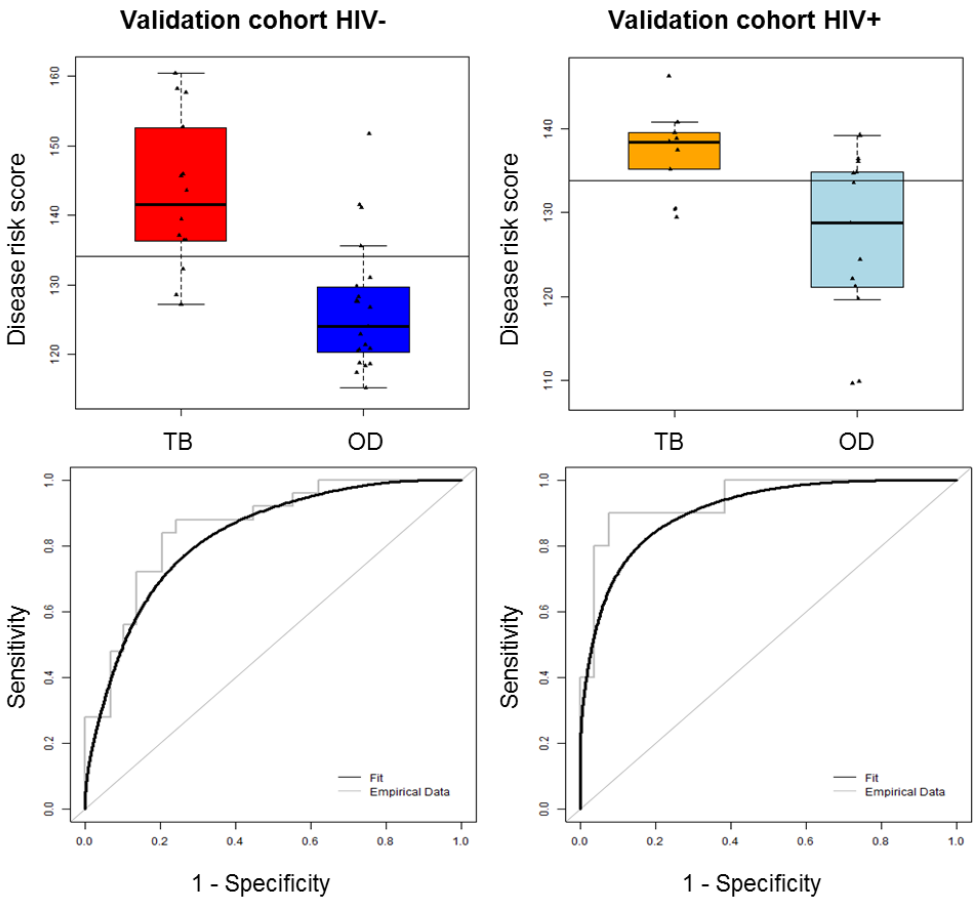
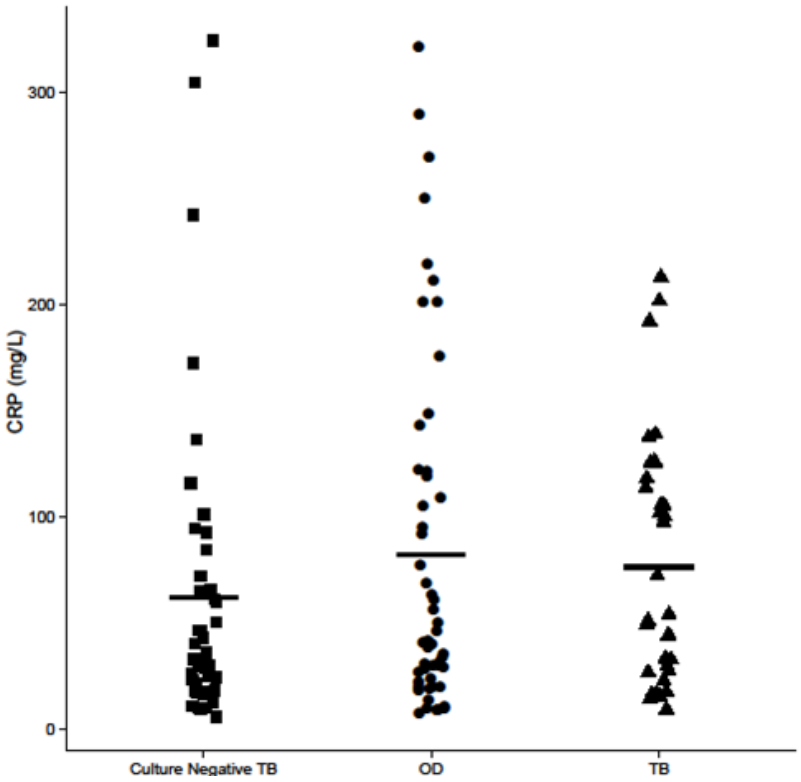


Figure S6. Serum CRP measurements from the patients which have been included in the gene expression analysis from the Kenyan validation cohort (44 culture negative TB, 52 other diseases and 34 culture positive TB).



Supplementary Tables

Table S1a. Clinical features of children in the South Africa/Malawi discovery cohort.

	TB / HIV-		TB / HIV+[†]		LTBI / HIV-		OD / HIV+[†]		OD / HIV-	
Location	SA	Malawi	SA	Malawi	SA	Malawi	SA	Malawi	SA	Malawi
No. children	50	22	23	19	4	53	30	40	50	55
Median age, months (IQR)	29 (15; 96)	99 (55; 134)	62 (43; 102)	89 (51; 130)	35 (29; 50)	48 (23; 91)	23 (13; 66)	97 (51; 156)	15 (11; 25)	47 (20; 104)
Male (%)	62	55	65	58	50	53	50	49	64	62
Median WAZ score (IQR)	-1.5 (-2.3; -0.4)	-2.7 (-3.4; -0.9)	-1.78 (-2.5; -1.1)	-2.9 (-3.2; -2.3)	-0.62 (-1.8; 0.4)	-1.1 (-2.1; 0.0)	-1.2 (-3.0; -0.6)	-2.2 (-2,8; -1.7)	-1.2 (-1.9; -0.0)	-1.1 (-2.2; -0.5)
BCG vaccinated (%)	42/45 (93%)	19/19 (100%)	17/18 (94%)	17/17 (100%)	4/4 (100%)	51/51 (100%)	25/25 (100%)	34/34 (100%)	47/48 (97.92)	52/53 (98%)
Median CD4 count/mm^3 (IQR)	NA	NA	640 (169; 812)	418 (278; 762)	NA	NA	531 (315; 805)	349 (141; 611)	NA	NA
Median % CD4 count (IQR)	NA	NA	18.7 (12.6; 26.3)	ND	NA	NA	20.9 (11.6; 22.4)	ND	NA	NA
TST positive*	38/50 (76%)	16/22 (73%)	10/20 (50%)	7/19 (37%)	4/4 (100%)	53/53 (100%)	0/30 (0%)	0/40 (0%)	0/50 (0%)	0/50 (0%)
IGRA positive	39/50 (78%)	18/22 (82%)	17/22 (77%)	13/19 (68%)	4/4 (100%)	53/53 (100%)	0/30 (0%)	0/40 (0%)	0/50 (0%)	0/50 (0%)

SA = South Africa, TB = active TB, LTBI = latent TB infection, OD = other diseases (see below), HIV- = HIV-uninfected, HIV+ = HIV-infected, IQR= interquartile range, WAZ = weight-for-age z-score, TST = tuberculin skin test, IGRA= interferon gamma release assay, ND= not done, NA= not applicable.

†12 of the HIV+ children in Malawi were on ART and 0 of the children from South Africa.

*A positive TST was defined according to WHO guidelines as an induration of ≥10mm; or ≥ 5mm in children with HIV infection or severe malnutrition with 2 TU of PPD RT23 (SSI, Denmark).

Discrepancies in total number of children in each category and number with a visible BCG scar denote cases in whom it was difficult to determine whether a scar was present.

Table S1b. Major clinical diagnoses in the 'Other Diseases' groups from each study site.

Group	HIV-infected			HIV-uninfected			
Location	**SA**	**Malawi**	**Kenya**	**SA**	**Malawi**	**Kenya**	**Total**
Pneumonia[a]	24	15	15	30	17	18	119
Bronchiectasis/chronic lung disease	2	7	-	-	1	-	10
Lymphocytic Interstitial Pneumonitis	-	2	-	-	-	-	2
Upper respiratory tract infection (URTI)	-	-	-	11	-	-	11
Inflammatory bone and joint diseases	-	1	-	-	8	2	11
Bacterial soft tissue infection	-	5	-	-	16	-	21
Gastroenteritis	2	-	-	5	-	-	7
Infection at ≥2 sites[b]	2	-	-	3	-	-	5
Sepsis without a focus[c]	-	-	4	-	-	1	5
Kaposi Sarcoma[d]	-	7	-	-	-	-	7
Other malignancy[e]	-	-	-	-	5	1	6
Malaria + severe malnutrition	-	-	-	-	1	3	4
Primary diagnosis of severe malnutrition[f]	-	1	4	-	1	3	9
Other[g]	-	2	3	1	6	1	13

[a] Includes 10 with pneumonia and bacteremia; one of whom also had empyema.

[b] Includes pneumonia and gastroenteritis (2); pneumonia and urinary tract infection and gastroenteritis (1) bacterial meningitis and URTI (1); gastroenteritis and URTI (1).

[c] Includes 3 children with bacteraemia, one child with bacteraemia and malaria, and one child with malaria and septic shock.

[d] Includes one child with Kaposi Sarcoma and septicemia.

[e] All histologically confirmed - Burkitt's lymphoma (3); rhabdomyosarcoma (1); non-Hodgkin's lymphoma (1); and metastatic carcinoma of uncertain origin (1).

[f] These are children who had a primary diagnosis of severe malnutrition; many other children in the OD group also had severe malnutrition in addition to the diagnoses listed.

[g] Includes cryptococcal meningitis (2); empyema (3); NTS septicaemia (2); congenital spinal abnormalities (2); abscess + bacteremia (1); bacterial meningitis (1); severe anemia (1); and one child with severe malnutrition and a febrile illness of uncertain etiology which resolved without TB treatment.

Table S2a. 42-transcript signature for distinguishing TB from latent TB infection.

Array ID	ILMN Gene	Transcript	Direction[$]	Description
6480059*****	ACTA2*****	ILMN_6588*****	UP	actin, alpha 2, smooth muscle, aorta (ACTA2), mRNA.
3310324	ALKBH7	ILMN_7229	DOWN	alkB, alkylation repair homolog 7 (E. coli) (ALKBH7), mRNA.
5550397*****	APOL6*****	ILMN_38312*****	UP	apolipoprotein L, 6 (APOL6), mRNA.
7400341	C11ORF2	ILMN_10940	DOWN	chromosome 11 open reading frame 2 (C11orf2), mRNA.
1500546	C20ORF201	ILMN_25727	DOWN	chromosome 20 open reading frame 201 (C20orf201), mRNA.
6380187	C21ORF57	ILMN_21121	DOWN	chromosome 21 open reading frame 57 (C21orf57), transcript variant 1, mRNA.
1470706	C8ORF55	ILMN_25304	DOWN	chromosome 8 open reading frame 55 (C8orf55), mRNA.
2030170*****	CARD16*****	ILMN_21555*****	UP	caspase recruitment domain family, member 16 (CARD16), transcript variant 2, mRNA.
6110427	CLIP1	ILMN_15054	UP	CAP-GLY domain containing linker protein 1 (CLIP1), transcript variant 1, mRNA.
5340246	CRIP2	ILMN_29728	DOWN	cysteine-rich protein 2 (CRIP2), mRNA.
4540239	**DEFA1**	**ILMN_29692**	**UP**	**defensin, alpha 1 (DEFA1), mRNA.**
4860128	DEFA1B	ILMN_176067	UP	defensin, alpha 1B (DEFA1B), mRNA.
2970747	DEFA3	ILMN_11220	UP	defensin, alpha 3, neutrophil-specific (DEFA3), mRNA.
7200274	DGCR6	ILMN_138781	DOWN	DiGeorge syndrome critical region gene 6 (DGCR6), mRNA.
3440647	DNAJC30	ILMN_30295	DOWN	DnaJ (Hsp40) homolog, subfamily C, member 30 (DNAJC30), mRNA.
3390068	E4F1	ILMN_23848	DOWN	E4F transcription factor 1 (E4F1), mRNA.
4670441	FBLN5	ILMN_29187	DOWN	fibulin 5 (FBLN5), mRNA.
1510364*	**GBP5***	**ILMN_24462***	**UP**	**guanylate binding protein 5 (GBP5), mRNA.**
3780047*	**GBP6***	**ILMN_1956***	**UP**	**guanylate binding protein family, member 6 (GBP6), mRNA.**
450632	GNG3	ILMN_7558	DOWN	guanine nucleotide binding protein (G protein), gamma 3 (GNG3), mRNA.
1500575	HS.538100	ILMN_103699	DOWN	xn24e12.x1 NCI_CGAP_Kid11 cDNA clone IMAGE:2694670 3, mRNA sequence
4590026	IMPDH2	ILMN_3439	DOWN	IMP (inosine monophosphate) dehydrogenase 2 (IMPDH2), mRNA.
7330575	KLHL28	ILMN_22112	DOWN	kelch-like 28 (Drosophila) (KLHL28), mRNA.
2810669	LCMT1	ILMN_16696	DOWN	leucine carboxyl methyltransferase 1 (LCMT1), transcript variant 1, mRNA.
5340414	LGTN	ILMN_4831	DOWN	ligatin (LGTN), mRNA.
2140541	LOC389816	ILMN_182870	DOWN	cytokeratin associated protein (LOC389816), mRNA.
620403	LOC400759	ILMN_181219	UP	similar to Interferon-induced guanylate-binding protein 1 (GTP-binding protein 1) (Guanine nucleotide-binding protein 1) (HuGBP-1) (LOC400759) on chromosome 1.
2230538	LRRN3	ILMN_306943	DOWN	leucine rich repeat neuronal 3 (LRRN3), transcript variant 1, mRNA.
5560075	MFGE8	ILMN_11368	DOWN	milk fat globule-EGF factor 8 protein (MFGE8), mRNA.
4210411*****	NDRG2*****	ILMN_19545*****	DOWN	NDRG family member 2 (NDRG2), transcript variant 6, mRNA.
6450424	NME3	ILMN_23571	DOWN	non-metastatic cells 3, protein expressed in (NME3), mRNA.
6770603	NOG	ILMN_7080	DOWN	noggin (NOG), mRNA.
4150017	PAQR7	ILMN_3765	DOWN	progestin and adipoQ receptor family member VII (PAQR7), mRNA.

2140382*****	PASK*****	ILMN_19873*****	DOWN	PAS domain containing serine/threonine kinase (PASK), mRNA.
4150100*****	PASK*****	ILMN_19873*****	DOWN	PAS domain containing serine/threonine kinase (PASK), mRNA.
7150189	PHF17	ILMN_1535	DOWN	PHD finger protein 17 (PHF17), transcript variant S, mRNA.
3400468	RAP1A	ILMN_20446	UP	RAP1A, member of RAS oncogene family (RAP1A), transcript variant 1, mRNA.
4670487	SIVA	ILMN_6846	DOWN	CD27-binding (Siva) protein (SIVA), transcript variant 2, mRNA.
6280433	SNHG7	ILMN_371358	DOWN	small nucleolar RNA host gene 7 (non-protein coding) (SNHG7), transcript variant 1, non-coding RNA.
4260189	TGIF1	ILMN_162784	DOWN	TGFB-induced factor homeobox 1 (TGIF1), transcript variant 1, mRNA.
4050059	U2AF1L4	ILMN_8757	DOWN	U2 small nuclear RNA auxiliary factor 1-like 4 (U2AF1L4), transcript variant 2, mRNA.
6550358	UBA52	ILMN_27795	DOWN	ubiquitin A-52 residue ribosomal protein fusion product 1 (UBA52), transcript variant 2, mRNA.

$ in TB patients in relation to patients with latent TB infection. 3 probes are also in the TB/OD signature (in bold), 8 probes overlap with the Berry *et al*. 393 probe adult signature (*****)

Table S2b. 51-transcript signature for distinguishing TB from other diseases.

Array ID	ILMN Gene	Transcript	Direction[$]	Description
4180768	ALAS2	ILMN_13644	UP	aminolevulinate, delta-, synthase 2 (ALAS2), nuclear gene encoding mitochondrial protein, transcript variant 3, mRNA.
1070477	ALDH1A1	ILMN_177898	UP	aldehyde dehydrogenase 1 family, member A1 (ALDH1A1), mRNA.
5910019	C1QB	ILMN_36274	UP	complement component 1, q subcomponent, B chain (C1QB), mRNA.
4290026	C20ORF103	ILMN_165304	DOWN	chromosome 20 open reading frame 103 (C20orf103), mRNA.
2600634	C3HC4	ILMN_6980	DOWN	membrane-associated ring finger (C3HC4) 8 (MARCH8), transcript variant 6, mRNA.
1580048	CAST	ILMN_163108	UP	calpastatin (CAST), transcript variant 9, mRNA.
3390564	CCDC52	ILMN_23129	UP	coiled-coil domain containing 52 (CCDC52), mRNA.
3940754	CD226	ILMN_3877	UP	CD226 molecule (CD226), mRNA.
1780440	CD79A	ILMN_37614	UP	CD79a molecule, immunoglobulin-associated alpha (CD79A), transcript variant 1, mRNA.
5890653	CDKN1C	ILMN_20689	DOWN	cyclin-dependent kinase inhibitor 1C (p57, Kip2) (CDKN1C), mRNA.
5340767	CEACAM1	ILMN_21651	DOWN	carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein) (CEACAM1), transcript variant 1, mRNA.
130086	CYB561	ILMN_8373	UP	cytochrome b-561 (CYB561), transcript variant 1, mRNA.
840446	CYB561	ILMN_20474	UP	cytochrome b-561 (CYB561), transcript variant 3, mRNA.
4540239	**DEFA1**	**ILMN_29692**	**UP**	**defensin, alpha 1 (DEFA1), mRNA.**
1050068	F2RL1	ILMN_176188	UP	coagulation factor II (thrombin) receptor-like 1 (F2RL1), mRNA.
6510707	FER1L3	ILMN_18562	UP	fer-1-like 3, myoferlin (C. elegans) (FER1L3), transcript variant 1, mRNA.
6840767	FRMD3	ILMN_11826	DOWN	FERM domain containing 3 (FRMD3), mRNA.
2350189	GBP3	ILMN_3653	UP	guanylate binding protein 3 (GBP3), mRNA.
1510364	**GBP5**	**ILMN_24462**	**UP**	**guanylate binding protein 5 (GBP5), mRNA.**
3780047	**GBP6**	**ILMN_1956**	**UP**	**guanylate binding protein family, member 6 (GBP6), mRNA.**
6220739	GRAMD1B	ILMN_308544	DOWN	GRAM domain containing 1B (GRAMD1B), mRNA.
5260484	HLA-DRB1	ILMN_20550	UP	major histocompatibility complex, class II, DR beta 1 (HLA-DRB1), mRNA.
6370315	HLA-DRB5	ILMN_3178	UP	major histocompatibility complex, class II, DR beta 5 (HLA-DRB5), mRNA.
620544	HLA-DRB6	ILMN_5312	UP	major histocompatibility complex, class II, DR beta 6 (pseudogene) (HLA-DRB6), non-coding RNA.
630619	HPSE	ILMN_165418	DOWN	heparanase (HPSE), mRNA.
5340762	HS.106234	ILMN_74965	UP	cDNA FLJ37173 fis, clone BRACE2028392
7320678	HS.171481	ILMN_80341	UP	hx21e11.y1 Human primary human ocular pericytes. Equalized (hx) Homo sapiens cDNA clone hx21e11 5, mRNA sequence
4880370	JUP	ILMN_3789	DOWN	junction plakoglobin (JUP), transcript variant 1, mRNA.
1050215	KCNJ15	ILMN_164363	DOWN	potassium inwardly-rectifying channel, subfamily J, member 15 (KCNJ15), transcript variant 1, mRNA.
2570438	KIFC3	ILMN_4695	UP	kinesin family member C3 (KIFC3), mRNA.
7560114	KLHDC8B	ILMN_6513	UP	kelch domain containing 8B (KLHDC8B), mRNA.

5310445	KREMEN1	ILMN_41914	DOWN	kringle containing transmembrane protein 1 (KREMEN1), transcript variant 4, mRNA.
4570164	LOC389386	ILMN_165610	UP	PREDICTED: misc_RNA (LOC389386), partial miscRNA.
4780044	LOC389386	ILMN_352098	UP	PREDICTED: misc_RNA (LOC389386), partial miscRNA.
2350121	LOC642678	ILMN_38908	UP	PREDICTED: similar to myeloid/lymphoid or mixed-lineage leukemia 3 isoform 2 (LOC642678), mRNA.
6480364	LOC647460	ILMN_38026	DOWN	PREDICTED: similar to Ig kappa chain V-I region HK101 precursor (LOC647460), mRNA.
6900291	LOC649210	ILMN_33006	DOWN	PREDICTED: similar to Ig lambda chain V region 4A precursor (LOC649210), mRNA.
830639	LOC653778	ILMN_32201	DOWN	PREDICTED: similar to solute carrier family 25, member 37 (LOC653778), mRNA.
2260349	MIR1974	ILMN_388657	DOWN	microRNA 1974 (MIR1974), microRNA.
830750	NCF1B	ILMN_168368	UP	neutrophil cytosolic factor 1B pseudogene (NCF1B), non-coding RNA.
6760593	OSBPL10	ILMN_11112	UP	oxysterol binding protein-like 10 (OSBPL10), mRNA.
3170246	PDCD1LG2	ILMN_3561	UP	programmed cell death 1 ligand 2 (PDCD1LG2), mRNA.
2000292	SCGB3A1	ILMN_23096	DOWN	secretoglobin, family 3A, member 1 (SCGB3A1), mRNA.
160368	SEMA6B	ILMN_21277	DOWN	sema domain, transmembrane domain (TM), and cytoplasmic domain, (semaphorin) 6B (SEMA6B), mRNA.
1400593	SIGLEC14	ILMN_309673	UP	sialic acid binding Ig-like lectin 14 (SIGLEC14), mRNA.
460463	SMARCD3	ILMN_19301	UP	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 3 (SMARCD3), transcript variant 1, mRNA.
540520	SNORD8	ILMN_366693	UP	small nucleolar RNA, C/D box 8 (SNORD8), small nucleolar RNA.
1240554	TNFRSF17	ILMN_17574	UP	tumor necrosis factor receptor superfamily, member 17 (TNFRSF17), mRNA.
4760747	TPST1	ILMN_174128	UP	tyrosylprotein sulfotransferase 1 (TPST1), mRNA.
2630195	VAMP5	ILMN_20179	DOWN	vesicle-associated membrane protein 5 (myobrevin) (VAMP5), mRNA.
3940088	ZBED2	ILMN_4927	DOWN	zinc finger, BED-type containing 2 (ZBED2), mRNA.

$ in TB patients in relation to patients with other diseases. 3 probes are also in the TB/LTBI signature (in bold), no probes overlap with the Berry *et al.* 86 probe adult signature.

Table S3. Diagnostic performance of the disease risk score based on the TB vs. LTBI signature in the SA/Malawi test set.

Test set: n_{TB}=23 n_{LTBI}=11, Validation cohort: n_{TB}=35 n_{LTBI}=14.

	Test set SA and Malawi	**Kenya Validation Cohort**
	HIV- & HIV+ combined	
TB vs. LTBI 42 transcript DRS		
Area under ROC curve (95% CI)	98.4 (94.5 - 100.0)	100.0 (100.0 - 100.0)
Sensitivity % (95% CI)	95.7 (87.0 -100.0)	94.3 (85.7 - 100.0)
Specificity % (95% CI)	90.9 (72.7 - 100.0)	100.0 (100.0 - 100.0)

Table S4. Strategy for selection of culture-negative TB.

We selected numbers of each category (highly probable, probable, and possible TB) for inclusion in the microarray study, to match the expected number of true TB contributed by each sub-group to the overall culture negative group. We assumed 80%, 50%, 40% prevalence of TB among the highly probable, probable and possible TB respectively based on the differing clinical certainty of correct diagnosis in each category. As shown in the table below, the arrayed samples in each category closely followed the predicted proportions of true TB contributed by each group.

Group	**Number of samples recruited**	**Prevalence of TB in group**	**Expected number of actual TB cases**	**Proportion of expected number of cases**	**Number of cases arrayed**	**Proportion of cases arrayed**
Highly probable TB	15	80%	12	17%	8	18%
Probable TB	64	50%	32	45%	19	43%
Possible TB	66	40%	26	38%	17	39%
Total	145		70	100%	44	100%

Table S5a. Comparison of culture-negative TB cases included in & excluded from the array analysis by diagnostic category for the validation cohort.

	Highly probable TB					Probable TB					Possible TB				
	Included (n=8)		Excluded (n=7)		p value[2]	Included (n=19)		Excluded (n=45)		p value[2]	Included (n=17)		Excluded (n=49)		p value[2]
Tuberculosis exposure[1]															
Close TB contact history	5	(63%)	2	(29%)	0.31	11	(58%)	10	(22%)	0.009	5	(29%)	8	(16%)	0.29
TST positive	8	(100%)	6	(86%)	0.47	6	(39%)	13	(29%)	1	1	(6%)	1	(2%)	0.45
Tuberculosis exposure	8	(100%)	6	(86%)	0.47	13	(68%)	20	(44%)	0.10	5	(29%)	8	(16%)	0.29
Clinical symptoms/signs of TB[1]															
Persistent cough >2 weeks	5	(63%)	5	(71%)	1	11	(58%)	30	(67%)	0.57	12	(71%)	30	(61%)	0.57
Persistent fever >2 weeks	4	(50%)	4	(57%)	1	6	(32%)	27	(60%)	0.06	8	(47%)	33	(67%)	0.16
Night sweats >2 weeks	3	(38%)	3	(43%)	1	6	(32%)	7	(16%)	0.18	3	(18%)	10	(20%)	1
Weight loss or failure to thrive	7	(88%)	6	(86%)	1	12	(63%)	28	(62%)	1	12	(71%)	34	(69%)	1
CXR features of TB[1]															
Airway compression	1	(13%)	0	(0%)	1	1	(5%)	0	(5%)	0.30	0	(0%)	0	(0%)	-
Lymphadenopathy	6	(75%)	6	(86%)	1	7	(37%)	19	(42%)	0.78	1	(6%)	4	(8%)	1
Airspace shadowing	3	(38%)	2	(29%)	1	9	(47%)	18	(40%)	0.59	3	(18%)	15	(31%)	0.36
Miliary/nodular shadowing	0	(0%)	0	(0%)	-	1	(5%)	1	(2%)	0.51	1	(6%)	1	(2%)	0.45
Pleural effusion	0	(0%)	1	(14%)	0.47	2	(11%)	2	(4%)	0.58	0	(0%)	1	(2%)	1
Cavities	0	(0%)	0	(0%)	-	2	(11%)	3	(7%)	0.63	0	(0%)	1	(2%)	1
Calcified Ghon focus	0	(0%)	0	(0%)	-	0	(0%)	0	(0%)	-	0	(0%)	0	(0%)	-
Vertebral spondylitis	0	(0%)	0	(0%)	-	0	(0%)	0	(0%)	-	0	(0%)	0	(0%)	-

[1] See reference[13]

[2] Fisher's exact 2-sided test

Table S5b. Comparison of OD cases included in & excluded from the array analysis by diagnostic category for the validation cohort.

	Included (n=64)		Excluded (n=935)		p value[2]
Tuberculosis exposure[1]					
Close TB contact history	6	(9%)	143	(15%)	0.28
TST positive	4	(6%)	118	(13%)	0.17
Tuberculosis exposure	16	(25%)	297	(32%)	0.27
Clinical symptoms/signs of TB[1]					
Persistent cough >2 weeks	23	(36%)	433	(47%)	0.12
Persistent fever >2 weeks	21	(33%)	379	(41%)	0.24
Night sweats >2 weeks	10	(16%)	143	(15%)	1
Weight loss or failure to thrive	50	(78%)	514	(56%)	<0.001
CXR features of TB[1]					
Airway compression	0	(0%)	2	(0.2%)	1
Lymphadenopathy	5	(8%)	67	(7%)	0.80
Airspace shadowing	17	(27%)	156	(17%)	0.06
Miliary/nodular shadowing	0	(0%)	0	(0%)	-
Pleural effusion	4	(6%)	17	(2%)	0.04
Cavities	0	(0%)	1	(0.1%)	1
Calcified Ghon focus	0	(0%)	0	(0%)	-
Vertebral spondylitis	0	(0%)	0	(0%)	-

[1] See reference[13]

[2] Fisher's exact 2-sided test

Table S6. Diagnostic performance of the TB/OD disease risk score in the SA/Malawi test set and the Kenyan validation cohort and comparison with Xpert MTB/RIF by HIV status.

	Test set SA and Malawi		**Kenyan Independent Validation Cohort**			
	TB/OD 51 transcript signature				**Xpert MTB/RIF**	
	HIV-	**HIV+**	**HIV-**	**HIV+**	**HIV-**	**HIV+**
Number of participants	TB=14 OD=21	TB=9 OD=13	TB=25 OD=29	TB=10 OD=26	TB=25 OD=29	TB=10 OD=26
Area under ROC curve (95% CI)	88.4 (75.9 - 97.6)	84.6 (64.0 - 96.6)	85.7 (75.0 - 94.4)	93.9 (83.9 - 100.0)	74.0 (64.0 - 84.0)	85.0 (70.0 - 95.1)
Sensitivity % (95% CI)	78.6 (57.1 - 100.0)	77.8 (55.6 - 100.0)	80.0 (64.0 - 92.0)	90.0 (70.0 - 100.0)	48.0 (28.0 - 64.1)	70.0 (40.0 - 100.0)
Specificity % (95% CI)	81.0 (61.9 - 95.2)	61.5 (30.8 - 84.6)	79.3 (65.4 - 93.1)	92.3 (80.8 - 100.0)	100.0 (100.0 - 100.0)	100.0 (100.0 - 100.0)

Table S7. Diagnostic performance of the TB/OD disease risk score and the Xpert MTB/RIF on culture-negative TB samples from the Kenyan validation cohort.

		Area under ROC curve % (95% CI)	Sensitivity % (95% CI)	Effective Sensitivity % (95% CI)		
Highly Probable TB vs. OD (n_{TB}=8 n_{OD}=55)	**Estimated "actual" TB prevalence in group**	**100%**	**100%**	**70%**	**80%**	**90%**
	DRS 51 TB vs. OD signature	77.5 (58.2 – 94.3)	62.5 (25.0 – 100.0)	82.3 (41.9 – 100.0)	74.1 (37.6 – 100.0)	67.6 (35.1 – 100.0)
	Xpert MTB/RIF®*	62.5 (50.0 – 81.3)	25.0 (0.0 – 50.0)	35.7 (1.1 – 65.7)	31.3 (1.0 – 57.6)	27.8 (1.0 – 51.3)
Probable TB vs. OD (n_{TB}=19 n_{OD}=55)	**Estimated "actual" TB prevalence in group**	**100%**	**100%**	**40%**	**50%**	**60%**
	DRS 51 TB vs. OD signature	72.3 (59.6 – 84.2)	42.1 (21.1 – 63.2)	80.8 (36.4 – 100.0)	67.9 (32.7 – 100.0)	59.3 (30.2 – 90.6)
	Xpert MTB/RIF®*	52.6 (50.0 – 57.9)	5.3 (0.0 – 17.8)	13.3 (0.0 – 36.5)	10.6 (0.0 – 29.3)	8.8 (0.0 – 24.5)
Possible TB vs. OD (n_{TB}=17 n_{OD}=55)	**Estimated "actual" TB prevalence in group**	**100%**	**100%**	**30%**	**40%**	**50%**
	DRS 51 TB vs. OD signature	64.5 (48.4 – 77.7)	35.3 (11.8 – 58.8)	79.6 (7.2 – 100.0)	63.8 (9.2 – 100.0)	54.3 (10.2 – 91.0)
	Xpert MTB/RIF®*	50.0 (50.0 – 50.0)	0.0 (0.0 – 0.0)	0.0 (0.0 – 0.0)	0.0 (0.0 – 0.0)	0.0 (0.0 – 0.0)

* The Xpert MTB/RIF test had a positive outcome for 3 out of 44 culture-negative TB cases and 0 out of 55 other diseases cases.

Specificity remains the same as in Table 2.

Table S8. Positive and Negative predictive value for the Kenyan validation cohort in different prevalence scenarios & based on the sensitivity in both culture-negative and culture-positive TB groups.

Combined sensitivity[a]	Statistic	Prevalence[b]		
		10%	30%	50%
A: 70%	PPV % (95% CI)	38.3 (23.4 – 53.3)	70.5 (57.4 – 83.7)	84.8 (76.6 – 943.0)
	NPV % (95% CI)	93.6 (94.9 – 97.7)	87.1 (82.8 – 91.5)	74.4 (67.0 – 81.8)
B: 75%	PPV % (95% CI)	41.0 (25.8 – 56.3)	72.9 (60.4 – 85.3)	86.2 (78.7 – 93.7)
	NPV % (95% CI)	96.9 (95.5 – 98.2)	89 (84.6 – 93.4)	77.6 (69.8 – 85.4)
C: 82%	PPV % (95% CI)	44.3 (28.8 – 59.8)	75.4 (63.8 – 87.1)	87.8 (81.0 – 94.5)
	NPV % (95% CI)	97.8 (96.5 – 99.0)	91.9 (87.6 – 96.1)	82.9 (74.9 – 90.9)

[a] Combined sensitivity is defined as the average sensitivity across all culture-negative and -positive TB groups. This is calculated by weighting the adjusted sensitivity calculated in each group by the relative size of each group in the Kenyan prospective cohort. This sensitivity is calculated according to three scenarios, as described in methods, and depends on an assumption as to the prevalence of 'actual' TB in each group.

A: Definite TB (100%) + HP TB (90%) +Pr TB (60%) + Pos TB (50%)

B: Definite TB (100%) + HP TB (80%) +Pr TB (50%) + Pos TB (40%)

C: Definite TB (100%) + HP TB (70%) +Pr TB (40%) + Pos TB (30%)

[b] Prevalence represents the prevalence of actual TB in the group of children to which test is given

10%: reflects the prevalence of TB in the Kenyan cohort

30%: reflects the prevalence of TB from the South Africa and Malawi recruitment

50%: reflects a scenario which includes prior filtering or a combination with another test

References

1. WHO. Global tuberculosis control. Geneva: World Health Organisation; 2011.

2. Kranzer K, van Schaik N, Karmue U, et al. High prevalence of self-reported undiagnosed HIV despite high coverage of HIV testing: a cross-sectional population based sero-survey in South Africa. PLoS One 2011;6:e25244.

3. WHO UNICEF UNAIDS. Global HIV/AIDS response, epidemic update and health sector progress towards universal access. Geneva2011.

4. van Well GT, Paes BF, Terwee CB, et al. Twenty years of pediatric tuberculous meningitis: a retrospective cohort study in the western cape of South Africa. Pediatrics 2009;123:e1-8.

5. National Statistical Office. Malawi Demographic and Health Survey 2004. Zomba, Malawi: National Statistical Office; 2005 December 2005.

6. Brooker SJ, Pullan RL, Gitonga CW, et al. Plasmodium-helminth coinfection and its sources of heterogeneity across East Africa. The Journal of infectious diseases 2012;205:841-52.

7. Kenya National Bureau of Statistics. Kenya Demographic and Health Survey 2008-09. Nairobi2010.

8. Brent AJ, Ahmed I, Ndiritu M, et al. Incidence of clinically significant bacteraemia in children who present to hospital in Kenya: community-based observational study. Lancet 2006;367:482-8.

9. Schmid R, Baum P, Ittrich C, et al. Comparison of normalization methods for Illumina BeadChip HumanHT-12 v3. BMC Genomics 2010;11:349.

10. Jolliffe IT. Principal Component Analysis: Springer-Verlag; 1986.

11. Hotelling H. The generalization of Student's ratio. Annals of Mathematical Statistics 1931;2:360-78.

12. Zou H, Hastie T. Regularization and variable selection via the elastic net. J Roy Stat Soc B 2005;67:301-20.

13. Graham SM, Ahmed T, Amanullah F, et al. Evaluation of tuberculosis diagnostics in children: 1. Proposed clinical case definitions for classification of intrathoracic tuberculosis disease. Consensus from an expert panel. The Journal of infectious diseases 2012;205 Suppl 2:S199-208.