What is the key question?

Are there serum host marker signatures, which are suitable for point-of-care tests that differentiate between active pulmonary TB and other conditions in individuals presenting with signs and symptoms suggestive of TB in primary health care settings in Africa?

What is the bottom line?

A seven-marker host serum protein biosignature consisting of CRP, transthyretin, IFN-γ, complement factor H, apolipoprotein-A1, IP-10 and serum amyloid A, is promising as a diagnostic biosignature for TB disease, regardless of HIV infection status or African country of sample origin.

Why read on?

The 7 serum marker biosignature identified in this large multi-centered study on 716 individuals with signs and symptoms suggestive of TB could form the basis of a rapid, point-of-care screening test, and with a sensitivity of 94% and negative predictive value of 96%, such a test would render about 75% of the currently performed GeneXpert or TB cultures unnecessary.

Diagnostic Performance of a Seven-marker Serum Protein Biosignature for the Diagnosis of Active TB Disease in African Primary Health Care Clinic Attendees with Signs and Symptoms Suggestive of TB

Novel N. Chegou[1], Jayne S. Sutherland[2], Stephanus Malherbe[1], Amelia C. Crampin[3], Paul L.A.M. Corstjens[4], Annemieke Geluk[5], Harriet Mayanja-Kizza[6], Andre G. Loxton[1], Gian van der Spuy[1], Kim Stanley[1], Leigh A. Kotzé[1], Marieta van der Vyver[7], Ida Rosenkrands[8], Martin Kidd[9], Paul D. van Helden[1], Hazel M. Dockrell[10], Tom H.M. Ottenhoff[5], Stefan H.E. Kaufmann[11], and Gerhard Walzl[1#] on behalf of the AE-TBC consortium

[1]DST/NRF Centre of Excellence for Biomedical Tuberculosis Research and SAMRC Centre for Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, PO Box 241, Cape Town, 8000, South Africa

[2]Vaccines and Immunity, Medical Research Council Unit, Fajara, The Gambia

[3]Karonga Prevention Study, Chilumba, Malawi

[4]Department of Molecular Cell Biology, Leiden University Medical Centre, PO Box 9600, 2300 RC Leiden, The Netherlands

[5]Department of Infectious Diseases, Leiden University Medical Centre, PO Box 9600, 2300 RC Leiden, The Netherlands

[6]Department of Medicine, Makerere University, Kampala, Uganda

[7]School of Medicine, Faculty of Health Sciences, University of Namibia, Namibia

[8]Department of Infectious Disease Immunology, Statens Serum Institut, Copenhagen 2300s, Denmark

[9]Centre for Statistical Consultation, Department of Statistics and Actuarial Sciences, Stellenbosch University, Cape Town, South Africa

[10]Department of Immunology and Infection, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK

[11]Department of Immunology, Max Planck Institute for Infection Biology, Charitéplatz 1, 10117 Berlin, Germany

#Corresponding Author: Gerhard Walzl, DST/NRF Centre of Excellence for Biomedical Tuberculosis Research and SAMRC Centre for Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch

University, PO Box 241, Cape Town, 8000, South Africa, Telephone: +27219389158, Fax: +27219389863, E-mail: gwalzl@sun.ac.za

Alternate (Pre-publication) corresponding Author: Novel N. Chegou, DST/NRF Centre of Excellence for Biomedical Tuberculosis Research and SAMRC Centre for Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, PO Box 241, Cape Town, 8000, South Africa, Telephone: +27219389069, Fax: +27219389863, E-mail: novel@sun.ac.za/novel01@gmail.com

Keywords:

Sensitivity, specificity, tuberculosis, biomarker, diagnosis

Word Count: 3648

ABSTRACT

Background

User-friendly, rapid, inexpensive yet accurate TB diagnostic tools are urgently needed at points-of-care in resource-limited settings. We investigated host biomarkers detected in serum samples obtained from adults with signs and symptoms suggestive of TB at primary health care clinics in five African countries (Malawi, Namibia, South Africa, The Gambia, and Uganda), for the diagnosis of TB disease.

Methods

We prospectively enrolled individuals presenting with symptoms warranting investigation for pulmonary TB, prior to assessment for TB disease. We evaluated 22 host protein biomarkers in stored serum samples using a multiplex cytokine platform. Using a pre-established diagnostic algorithm comprising of laboratory, clinical and radiological findings, participants were classified as either definite TB, probable TB, questionable TB status or non-pulmonary TB.

Results

Of the 716 participants enrolled, 185 were definite and 29 were probable TB cases, six had questionable TB disease status, whereas 487 had no evidence of TB. A seven-marker biosignature of CRP, transthyretin, IFN-γ, CFH, apolipoprotein-A1, IP-10 and SAA identified on a training sample set (n=491), diagnosed TB disease in the test set (n=210) with sensitivity of 93.8% (95% CI, 84.0-98.0%), specificity of 73.3% (95% CI, 65.2-80.1%), and positive and negative predictive values of 60.6% (95% CI, 50.3-70.1) and 96.4% (95% CI, 90.5-98.8%) respectively, regardless of HIV infection status or study site.

Conclusion:

We have identified a seven-marker host serum protein biosignature for the diagnosis of TB disease irrespective of HIV infection status or ethnicity in Africa. These results hold promise for the development of a field-friendly point-of-care screening test for pulmonary TB.

INTRODUCTION:

Tuberculosis (TB) remains a global health problem with an estimated 9.6 million people reported to have fallen ill with the disease and 1.5 million deaths in 2014[1]. Sputum smear microscopy, which has well described limitations, particularly sensitivity[2], remains the most commonly used diagnostic test for TB in resource-constrained settings. *Mycobacterium tuberculosis* (*M.tb*) culture, the reference standard test, has a long turn-around time[2], is expensive, prone to contamination and is not widely available in resource-limited settings. The GeneXpert MTB/RIF sputum test (Cepheid Inc, Sunnyvale, CA), arguably the most important commercial recent advance in the TB diagnostic field yields results within 2hours, coupled with the detection of rifampicin resistance. The Xpert test has been massively rolled out in developed countries but limitations, including relatively high operating costs and infrastructural requirements[3], hamper its use in resource-constrained settings. An important limitation of diagnostic tests based on sputum, is that they are unsuitable in individuals, particularly children, who have difficulty in providing good quality sputum[4], and also in individuals with extrapulmonary TB. There is an urgent need for alternative diagnostic tests that are suitable for use in all patient types, especially in resource-poor settings. Tests based on the detection of host inflammatory molecules[5;6] may be beneficial, especially when applied to easily available samples such as finger-prick blood or serum.

In search of immunodiagnostic tools that could be useful for the diagnosis of active TB, attempts are being made to identify novel antigens[7-9]. Those currently used in the Interferon-gamma (IFN-γ) release assays (ESAT-6/CFP-10/TB7.7) cannot differentiate between latent and active TB. There is also a search for host markers other than IFN-γ, that are produced after overnight stimulation of blood cells with ESAT-6/CFP-10/TB7.7[10-14], and antibodies against novel *M.tb* antigens[15;16].

Although some T-cell-based approaches[17] are promising for the diagnosis of active TB, overnight culture-based assays are not optimal as point-of-care tests. The importance of diagnosis of individuals with TB disease at the first patient contact and real-time notification to TB programs cannot be overemphasized, as delays in these steps lead to delays in the initiation of treatment and substantial loss to follow-up[18]. Therefore, diagnostic tests that can be easily performed at points-of-care by healthcare providers, without the need for

sophisticated laboratory equipment will contribute significantly to the management of TB disease.

We conducted a study investigating the potential of protein serum host markers to identify pulmonary TB in primary health care clinic attendees from five African countries. Our aim was to further investigate the diagnostic potential of biosignatures identified in our own unpublished pilot studies in a relatively large cohort of study participants, from different regions of the African continent, as such biosignatures might be useful as point-of-care tests for TB disease.

METHODS

Study participants

We prospectively recruited adults who presented with symptoms requiring investigation for pulmonary TB disease at primary health care clinics at five field sites in five African countries. The clinics served as field study sites for researchers at Stellenbosch University (SUN), South Africa; Makerere University (UCRC), Uganda; Medical Research Council Unit (MRC), The Gambia; Karonga Prevention Study (KPS), Malawi; and the University of Namibia (UNAM), Namibia, as part of the African European Tuberculosis Consortium (AE-TBC) for TB Diagnostic Biomarkers (www.ae-tbc.eu). Study participants were recruited between November 2010 and November 2012. All study participants presented with persistent cough lasting ≥2 weeks and at least one of either fever, malaise, recent weight loss, night sweats, knowledge of close contact with a TB patient, haemoptysis, chest pain or loss of appetite. Participants were eligible for the study if they were 18 years or older and willing to give written informed consent for participation in the study, including consent for HIV testing. Patients were excluded if they were pregnant, had not been residing in the study community for more than 3 months, were severely anaemic (haemoglobin <10g/l), were on anti-TB treatment, had received anti-TB treatment in the previous 90 days or if they were on quinolone or aminoglycoside antibiotics during the past 60 days. The study protocol was approved by the Health Research Ethics Committees of the participating institutions.

Sample collection and microbiological diagnostic tests

Harmonized protocols were used for collection and processing of samples across all study sites. Briefly, blood samples were collected at first contact with the patient, in 4-ml plain BD vacutainer serum tubes (BD Biosciences) and transported within 3 hours at ambient

temperature to the laboratory, where tubes were centrifuged at 2500 rpm for 10 minutes, after which serum was harvested, aliquoted and frozen (–80˚C) until use. Sputum samples were collected from all participants and cultured using either the MGIT method (BD Biosciences) or Lowenstein–Jensen media, depending on facilities available at the study site. Specimens demonstrating growth of microorganisms were examined for acid-fast bacilli using the Ziehl-Neelsen method followed by either Capilia TB testing (TAUNS, Numazu, Japan) or standard molecular methods, to confirm the isolation of organisms of the *M.tb* complex, before being designated as positive cultures.

Classification of study participants and reference standard

Using a combination of clinical, radiological, and laboratory findings, participants were classified as definite TB cases, probable TB cases, participants without pulmonary TB (no-PTB) or questionable disease status as described in table 1. Briefly, No-PTB cases had a range of other diagnoses, including upper and lower respiratory tract infections (viral and bacterial infections, although attempts to identify organisms by bacterial or viral cultures were not made), and acute exacerbations of chronic obstructive pulmonary disease or asthma. In assessing the accuracy of host biosignatures in the diagnosis of TB disease, all the definite and probable TB cases were classified as "TB", and then compared to the no-PTB cases, whereas questionables were excluded from the main analysis (Figure 1).

Table 1: **Harmonized definitions used in classifying study participants**

Classification	Definition
Definite TB	Sputum culture positive for MTB OR 2 positive smears and symptoms responding to TB treatment OR 1 Positive smear plus CXR suggestive of PTB
Probable TB	1 positive smear and symptoms responding to TB treatment OR CXR evidence and symptoms responding to TB treatment
Questionable	Positive smear(s), but no other supporting evidence OR CXR suggestive of PTB, but no other supporting evidence. OR Treatment initiated by healthcare providers on clinical suspicion only. No other supporting evidence
No-PTB	Negative cultures, negative smears, negative CXR and treatment never initiated by healthcare providers

Abbreviations: CXR, chest X ray; MTB, *Mycobacterium tuberculosis;* TB, pulmonary tuberculosis, No-PTB, non-"pulmonary tuberculosis".

Multiplex immunoassays

Using the Luminex technology, we evaluated the levels of 22 host biomarkers including interleukin-1 receptor antagonist (IL-1ra), transforming growth factor (TGF)-α, IFN-γ, IFN-γ-inducible protein (IP)-10, tumour necrosis factor (TNF)-α, IFN-α2, vascular endothelial growth factor (VEGF), matrix metallo-proteinase (MMP)-2, MMP-9, apolipoprotein A-1 (ApoA-1), Apo-CIII, transthyretin, complement factor H (CFH) (Merck Millipore, Billerica, MA, USA), and C-reactive protein (CRP), serum amyloid A (SAA), serum amyloid P (SAP), fibrinogen, ferritin, tissue plasminogen activator (TPA), procalcitonin (PCT), haptoglobulin and alpha-2-macroglobulin (A2M) (Bio-Rad Laboratories, Hercules, CA, USA). Prior to testing, samples for MMP-2 and MMP-9 were pre-diluted 1:100 following optimization experiments. Samples for all other analytes were evaluated undiluted, or diluted as recommended by the different manufacturers in the package inserts. The laboratory staff performing the experiments were blinded to the clinical groups of study participants. All assays were performed and read in a central laboratory (SUN) on the Bio-Plex platform (Bio-Rad), with the Bio-Plex Manager™ Software version 6.1 used for bead acquisition and analysis.

Statistical analysis

Differences in analyte concentrations between participants with TB disease and those without TB were evaluated by the Mann–Whitney U-test for non-parametric data analysis. The diagnostic accuracy of individual analytes was investigated by receiver operator characteristics (ROC) curve analysis. Optimal cut-off values and associated sensitivity and specificity were selected based on the Youden's index[19]. The predictive abilities of combinations of analytes were investigated by General discriminant analysis (GDA)[20] and random forests[21], following the training/test set approach. Briefly, patients were randomly assigned into the training set (70% of study participants, n=491) or test set (30%, n=210), regardless of HIV infection status or study site by the software used in data analysis (Statistica, Statsoft, Ohio, USA). These training and test sets were selected using random sampling, stratified on the dependent (TB) variable. The most accurate of the top 20 marker combinations identified in the training set were then evaluated on the test sample set.

RESULTS

A total of 716 individuals were prospectively evaluated in the current study. One study participant was found to be pregnant at the time of recruitment, and data for 8 other participants were not appropriately captured. These 9 individuals were excluded from further analysis (Figure 1). Table 2 shows participant characteristics.

Table 2: Clinical and demographic characteristics of study participants. The number and characteristics of participants enrolled from the different study sites are shown

Study site	SUN	MRC	UCRC	KPS	UNAM	Total
Participants (n)	161	209	171	117	49	707
Age, mean±SD, yr	37.4±11.3	34.9±12.1	33.1±10	39.9±13.6	36.5±9.6	36.0±11.8
Males, n(%)	68(42)	123(59)	87(51)	59(50)	28(57)	365(52)
HIV pos, n (%)	28(17)	20(10)	28(16)	67(57)	27(55)	170(24)
QFT pos, n(%)	105(69)	83(41)	119(70)	44(38)	35(71)	386(56)
Definite TB, n(%)	22(14)	53(25)	59(35)	18(15)	33(67)	185(26)
Probable TB, n(%)	4(2)	13(6)	4(2)	3(3)	5(10)	29(4)
Total TB[#], (n)	26	66	63	21	38	214
No-PTB, n (%)	133(83)	140(67)	108(63)	96(82)	10(20)	487(69)
Questionable, n(%)	2(1)	3(1)	0(0)	0(0)	1(2)	6 (1)

Table notes: SUN, Stellenbosch University, South Africa; KPS, Karonga Prevention Study, Malawi; MRC, Medical Research Council Unit, The Gambia; UCRC, Makerere University, Uganda; UNAM, University of Namibia, Namibia; SD, standard deviation; QFT, Quantiferon TB Gold In Tube; pos, positive; neg, negative; indet, indeterminate. ♯Total TB cases = all the Definite TB + Probable TB cases; TB, Pulmonary TB; No-PTB, non-"pulmonary tuberculosis".

Using pre-established and harmonized case definitions (Table 1), 185 (26.2%) of the study participants were classified as definite pulmonary TB cases, 29 (4.1%) were probable TB cases, representing the active TB group (214 participants; 30.3%), whereas 487 (68.9%) were No-PTB cases and 6 (0.8%) had an uncertain diagnosis (Table 2). The characteristics of the different patient subgroups are shown in Table 3.

Table 3: Characteristics of TB and no-PTB cases and individuals with "Questionable TB" disease status.

	Definite TB (n=185)	Probable TB (n=29)	ALL TB (n=214)	No-PTB (n=487)	Questionable TB (n=6)
Age, mean±SD, yr	33.8±9.6	36.3±9.6	34.1±9.6	36.8±12.6	36.5±12.0
Males, n(%)	118(64)	14(48)	132(62)	229(47)	4(67)
HIV pos, n(%)	47 (25)	8(28)	55(26)	114(23)	1(17)
QFT pos, n(%)	144 (78)	19(66)	164(78)	221(47)	2(33)
QFT neg, n(%)	28 (15)	10(34)	38(18)	235(49)	3(50)
QFT Indet, n(%)	8 (4)	0(0)	8(4)	19(4)	1(17)

SD, standard deviation; QFT= Quantiferon TB Gold In Tube; pos, positive; neg, negative; indet, indeterminate.

Utility of individual serum biomarkers in the diagnosis of TB disease

All serum markers investigated showed significant differences ($p<0.05$) between the TB cases and No-PTB cases except A2M and MMP-2 (Supplementary Table 1), irrespective of HIV infection status. Concentrations of CFH, CRP, ferritin, fibrinogen, haptoglobulin, IFN-α2, IFN-γ, IL-1ra, IP-10, MMP-9, PCT, SAA, SAP, TGF-α, TNF-α, TPA, and VEGF were significantly higher in the TB cases while those of ApoA-1, Apo-CIII, and transthyretin were higher in the no-PTB cases (Supplementary Table 1). When the accuracy for the diagnosis of TB disease was investigated by ROC curve analysis, the areas under the ROC curve (AUC) were between 0.70 and 0.84 for 10 analytes: CRP, ferritin, fibrinogen, IFN-γ, IP-10, TGF-α, TPA, transthyretin, SAA and VEGF (Figure 2). Sensitivity and specificity were both >70% for six of these analytes, namely; CRP, ferritin, IFN-γ, IP-10, transthyretin and SAA (Supplementary Table 1).

Supplementary Table 1: Median levels of analytes detected in serum samples from individuals with pulmonary TB disease (n=214) or no-PTB disease (n=487), and accuracies in the diagnosis of TB disease

Host marker	No-PTB (IQR)	TB (IQR)	P-value	AUC	Cut-off value	Sensitivity (%)	Specificity (%)
IL-1ra	8 (0-40)	35 (0-77)	**<0.0001**	0.63 [0.58-0.68]	>33.9	52.2 [45.1-59.2]	71.9 [67.6-75.9]
TGF-α	3 (1-6)	7 (3-13)	**<0.0001**	0.73 [69.1-77.4]	>5.6	62.8 [55.8-69.4]	76.0 [72.0-79.8]

IP-10	368 (209-652)	1712 (808-3558)	**<0.00001**	0.82 [0.79-0.86]	>651.7	81.2 [75.2-86.3]	75.0 [71.0-78.8]
TNF-α	7 (3-12)	14 (8-27)	**<0.0001**	0.69 [0.65-0.74]	>9.5	67.2 [60.3-73.5]	65.0 [60.6-69.3]
IFN-α2	0 (0-6)	7 (0-19)	**<0.0001**	0.67 [0.62-0.71]	>2.9	59.4 [52.4-66.2]	71.3 [67.0-75.3]
IFN-γ	1 (0-3)	9 (3-21)	**<0.0001**	0.80 [0.76-0.84]	>2.8	78.3 [72.0-83.7]	74.2 [70.0-78.0]
VEGF	158 (19-286)	341 (144-624)	**<0.0001**	0.70 [0.65-74]	>269.8	60.4 [53.4-67.1]	72.5 [68.3-76.5]
MMP-2	175792 (28693-474927)	92881 (22348-312697)	0.091	0.54 [0.49-0.59]	<254965	64.3 [57.3-70.8]	44.6 [40.1-49.2]
MMP-9	401540 (155072-756297)	651549 (43831-1299700)	**0.0004**	0.59 [0.53-0.64]	> 525174	0.56 [0.49-0.63]	0.62 [0.58-0.66]
ApoA-1	2593900 (2101500-3847700)	1999800 (1493900-2604300)	**<0.0001**	0.69 [0.65-0.73]	< 2.17e+006	0.57 [0.50-0.64]	0.72 [0.68-0.76]
Apo C-III	261321 (178708-418395)	180967 (115790-297779)	**<0.0001**	0.65 [0.61-0.70]	< 265480	0.70 [0.63 - 0.76]	0.50 [0.45-0.55]
Transthyretin	411528 (261059-591773)	184107 (107526-291488)	**<0.0001**	0.78 [0.74-0.82]	< 280585	0.73 [0.66-0.79]	0.73 [0.68-0.76]
CFH	663345 (515681-929872)	760622 (599474-1008200)	**0.0013**	0.58 [0.53-0.62]	> 683022	0.61 [0.54 - 0.68]	0.53 [0.49 -0.58]
A2M	1770000 (712956-3273400)	1380700 (501530-3284700)	0.141	0.54 [0.49-0.58]	< 1.26e+006	0.48 [0.41-0.55]	0.61 [0.57-0.66]
Haptoglobulin	955718 (287186-26796000)	2774400 (443581-60000000)	**0.0001**	0.62 [0.57-0.66]	> 6.17e+006	0.47 [0.40-0.54]	0.71 [0.67 -0.75]
CRP	1731 (321-9686)	59195 (14047-136520)	**<0.0001**	0.84 [0.81-0.87]	> 7251	0.82 [0.76-0.87]	0.73 [0.68-0.76]
SAP	46609 (23028-81115)	63664 (20776-129181)	**0.0011**	0.58 [0.53-0.63]	> 63321	0.50 [0.43-0.57]	0.67 [0.63-0.71]
PCT	4259 (2474-6776)	6807 (4399-10000)	**<0.0001**	0.68 [0.63-0.72]	> 5245	0.70 [0.63-0.76]	0.61 [0.56-0.65]
Ferritin	33894 (13921-83571)	158610 (61712-365165)	**<0.0001**	0.78 [0.75-0.82]	> 69684	0.71 [0.64-0.77]	0.70 [0.66-0.74]
TPA	1638 (931-2604)	2977 (1949-4317)	**<0.0001**	0.72 [0.68-0.76]	> 2163	0.70 [0.63-0.76]	0.66 [0.61-0.70]

Fibrinogen	2466 (1804-4182)	3987 (2991-6555)	**0<0.0001**	0.73 [0.69-0.77]	> 2854	0.80 [0.74-0.85]	0.57 [0.53-0.62]
SAA	771 (279-3985)	6778 (4265-9689)	**<0.0001**	0.83 [0.80-0.86]	> 3113	0.86 [0.80-0.90]	0.71 [0.67-0.75]

Abbreviations: CFH, complement factor H; A2M, alpha-2-macroglobulin; CRP, C-reactive protein; SAP, serum amyloid P; SAA, serum amyloid A; PCT, procalcitonin; TPA, tissue plasminogen activator; AUC, area under the ROC curve; ROC, receiver operator characteristics. Both HIV-infected and -uninfected individuals were included in the analysis. The values shown for IFN-α2, IFN-γ, IL-1ra, IP-10, TGF-α, TNF-α, VEGF, ferritin, PCT and TPA are in pg/ml. All other analyte concentrations are in ng/ml. The values in brackets under AUC, sensitivity and specificity are the 95% Confidence Intervals.

Accuracy of individual host markers in HIV-uninfected study participants

We stratified the study participants according to HIV infection status and repeated the ROC curve analysis. No differences were observed in the AUCs for ApoA-1, PCT and MMP-9 in HIV-positive versus HIV-negative participants. However, the AUCs for some of the acute-phase proteins including A2M, CRP, ferritin, haptoglobulin, SAP and TPA, were higher in HIV-positive individuals. This was in contrast to the observations for the classical pro-inflammatory host markers (IFN-γ, IP-10, TNF-α); the growth factors (TGF-α and VEGF); the blood clotting protein fibrinogen, the thyroxin and retinol transporting protein; transthyretin and CFH, which performed best in HIV-uninfected individuals (Figure 3).

Utility of serum multi-analyte models in the diagnosis of TB disease

General discriminant analysis (GDA) models showed optimal prediction of pulmonary TB disease with seven-marker combinations. The most accurate seven-marker biosignature for the diagnosis of TB disease, regardless of HIV infection status, was a combination of ApoA-1, CFH, CRP, IFN-γ, IP-10, SAA and transthyretin. Without any model "supervision", this biosignature ascertained TB disease with a sensitivity of 86.7% (95% CI, 79.9-91.5%) and specificity of 85.3% (95% CI, 81.0-88.8%) in the training dataset (n=491; 168 TB and 323 no-PTB), and a sensitivity of 81.3% (95% CI, 69.2-89.5%) and specificity of 79.5% (95% CI, 71.8-85.5%) in the test dataset (n=210, 77 TB and 133 No-PTB). To improve test performance, we optimised the model for higher sensitivity at the expense of lower specificity, which would allow the test to be used as a screening tool. The amended cut-off values ascertained TB disease with a sensitivity of 90.7% (95% CI, 84.5-94.6%) and specificity of 74.8% (95% CI, 69.8-

79.2%) in the training dataset (n=491), and sensitivity of 93.8% (95% CI: 84.0-98.0) and specificity of 73.3% (95% CI, 65.2-80.1%) in the test dataset (n=210). The positive and negative predictive values (NPV) of the biosignature were 60.6% (95% CI, 50.3-70.1 %) and 96.4% (95% CI, 90.5-98.8%), respectively (Table 4). The AUC for the seven-marker biosignature (determined on the training sample set) was 0.91 (95% CI, 0.88-0.94) (Figure 4).

The random forest modelling approach gave similar prediction accuracies for TB and no-PTB as GDA (87% sensitivity and 83% specificity in the training sample set, and 83% sensitivity and 89% specificity in the test sample set), without selection of any preferred cut-off values. In addition to the seven analytes included in the optimal GDA biosignature, Apo-CIII, ferritin, fibrinogen, MMP-9 and TNF-α were also identified as important contributors to top models by the random forest analysis (Figure 4).

Table 4: Accuracy of the seven-marker serum protein biosignature (ApoA-1, CFH, CRP, IFN-γ, IP-10, SAA, transthyretin) in the diagnosis of TB disease regardless of HIV infection status.

	Sensitivity	**Specificity**	**PPV**	**NPV**
Training set (n=491)				
%, (n/N)	86.7 (130/150)	85.3 (291/341)	72.2	93.6
95% CI	(79.9-91.5)	(81.0-88.8)	(65.0-78.5)	(90.1-95.9)
Test set (n=210)				
%, (n/N)	81.3(52/64)	79.5(116/146)	63.4	90.6
95% CI	(69.2-89.5)	(71.8-85.5)	(52.0-73.6)	(83.9-94.8)
Accuracy of biosignature after selection of cut-off values optimized for sensitivity				
Training set (n=491)				
	Sensitivity	**Specificity**	**PPV**	**NPV**

%, (n/N)	90.7 (136/150)	74.8 (255/341)	61.3	94.8
95% CI	(84.5-94.6)	(69.8-79.2)	(54.5-67.6)	(91.2-97.0)
Test set (n=210)				
%, (n/N)	93.8 (60/64)	73.3 (107/146)	60.6	96.4
95% CI	(84.0-98.0)	(65.2-80.1)	(50.3-70.1)	(90.5-98.8)

Accuracy of the seven-marker biosignature in smear and culture negative patients

We evaluated the accuracy of the biosignature in classifying all study participants as TB disease or "no-TB" regardless of the results of the reference standard, and particularly focused on patients who were missed by the microbiological tests (smear and culture) but diagnosed with TB disease based on clinical features including chest X-rays and response to TB treatment (Table 1). The biosignature correctly classified 74% (17/23) of patients who were smear negative but culture positive, and 67% (6/9) of patients who were both smear and culture negative. However, the biosignature only correctly classified 88% (86/98) of all the smear positive TB patients, but correctly diagnosed 91% (80/88) of these patients if the smear results were culture confirmed.

Accuracy of serum biosignatures in individuals without HIV infection

In the absence of HIV infection the GDA procedure indicated optimal diagnosis of TB disease when markers were used in combinations of four with ApoA-1, IFN-γ, IP-10 and SAA constituting the top model with sensitivity of 76.5% (95% CI, 67.5–83.7%) and specificity of 91.1% (95% CI 86.7–94.1) in the training sample set (n=372, 115 TB and 257 no-PTB), and a sensitivity of 77.3% (95% CI, 61.8–88.0) and specificity of 87.1% (95% CI, 79.3–92.3%) in the test dataset (n=160, 44 TB and 116 no-PTB). The positive and NPV of the four-marker model in the test set were 69.4% (95% CI, 54.4–81.3%) and 91.0% (95% CI, 83.7–95.4%), respectively.

DISCUSSION

We investigated the potential value of 22 host serum protein biomarkers in the diagnosis of TB disease in individuals presenting with symptoms suggestive of pulmonary TB disease at peripheral-level healthcare clinics in five different African countries. Although most of the analytes showed promise individually, the most optimal discriminatory profile was a seven-marker biosignature comprised of ApoA-1, CFH, CRP, IFN-γ, IP-10, SAA and transthyretin,

which might be useful in the rapid diagnosis of TB disease regardless of HIV infection status or ethnicity in Africa.

Diagnostic tests based on the detection of host protein biomarkers in *ex vivo* samples might be more beneficial than antigen stimulation assays as results can potentially be obtained rapidly if lateral flow technologies are employed. Besides the markers that were included in our final seven-marker biosignature (ApoA-1, CFH, CRP, IFN-γ, IP-10, SAA and transthyretin), other analytes including ferritin, fibrinogen, PCT, TGF-α, TNF-α, TPA, and VEGF showed diagnostic potential for TB disease and could have equally been included in the final model in place of any of the seven selected markers. Most of these markers are well known, disease non-specific markers of inflammation and have been extensively investigated in diverse disease conditions.

IFN-γ, IP-10 and TNF-α, together with other markers including IL-2 (reviewed in[17]), are amongst the most investigated host immunological biomarkers for the diagnosis of *M.tb* infection and disease. Both IFN-γ and IP-10 showed potential in this study. The inclusion of these markers in the seven-marker model is not surprising, given their widely accepted roles in the pathogenesis of *M.tb* infection.

CRP, ferritin, fibrinogen, SAA, and TPA are acute-phase proteins. The circulating levels of these proteins, as well as those of complement and clotting factors, are known to change by at least 25% in response to inflammatory stimuli, in keeping with their roles in host defense[22]. CRP (reviewed in[22]) is predominantly produced by hepatocytes. The association between serum levels of CRP, SAA and TB has long been established, including for treatment response[23]. Ferritin is widely known as a biomarker for iron deficiency [24], and is essential in iron homeostasis in *M.tb*[25]. Although high levels of ferritin have been observed in many non-communicable diseases including cancers, disseminated *M.tb* disease is a common cause of hyperferritinemia[26;27].

PCT, the precursor molecule of calcitonin is a general inflammatory response marker that is secreted in healthy individuals by the C cells of the thyroid and by leukocytes via alternate pathways, including induction by cytokines and bacterial products after microbial infection[28]. Although mainly known as a diagnostic marker for bacteremia[29], PCT levels have been shown

to be potentially useful in discriminating between pulmonary TB and community-acquired pneumonia in HIV-positive individuals[30].

ApoA-1, the major protein component of high-density lipoproteins, and CFH, a crucial regulator of the alternative complement pathway, were also amongst the markers included in our final seven-marker biosignature. ApoA-1 is one of the most important biomarkers for cardiovascular disease[31], but is rarely investigated as a biomarker in TB. Like the other markers investigated in this study, ApoA-1 may not play any specific role in the pathogenesis of TB. The low levels obtained in TB patients may be a result of the many changes in lipid metabolism, which are believed to occur after the generation of the acute-phase response following an inflammatory condition[31]. One of the ways that CFH recognizes host cells is by binding to host markers expressed on the surfaces of cells undergoing apoptosis[32]. With the help of these markers, including CRP and pentraxin 3, CFH ensures proper opsonization of these cells for efficient removal without excessive complement activation during the process, thus limiting immunopathology[32]. This process is however believed to be exploited by *M.tb*, to limit opsonization and therefore avoid killing[33]. Like ApoA-1, lower levels of CFH were observed in the TB cases in this study.

Transthyretin (reviewed in[34]) is a protein that is secreted by the liver into the blood and by the choroid plexus into the cerebrospinal fluid and has been widely investigated as a biomarker for nutritional status[34]. In previous TB studies, higher levels of transthyretin were observed in TB patients in comparison to uninfected controls[35], whereas lower levels were obtained in TB patients as compared to patients with lung cancer, with serum concentrations in TB patients increasing over the course of treatment[36]. Our observation is in agreement with these reports.

Combinations between transthyretin, CRP, SAA and neopterin ascertained TB disease with 78% accuracy in a previous proteomic finger-printing study[37]. In our study, a biosignature containing transthyretin, CRP, SAA and markers involved in Th1-related immunity to TB (IFN-γ, IP-10), an apolipoprotein and CFH showed excellent promise as a diagnostic tool for TB. Although most of these markers are promising individually[17;23;26;27;30;35-37], single host markers have many shortcomings in predicting TB disease due to poor specificity. As observed in this large multi-centered pan-African study, the accuracy of different host markers is affected differentially by HIV infection. A biosignature containing different classes of biomarkers, produced by different cell types such as the classical Th1 immune-related markers plus acute-

phase proteins, complement and apolipoproteins appears to offset the non-specific response patterns of individual or smaller groups of analytes. As a result, markers that perform relatively well in HIV-infected individuals such as the acute phase proteins, help in identifying patients who are missed by markers that may be more often affected by HIV infection such as IFN-γ and IP-10. The resultant test performance with relatively high sensitivity (93.8 %) and high NPV (96.4 %) appears promising as a screening test for active TB disease. Our data indicates that a test based on this biosignature will be superior to smear microscopy and may identify some patients who might be missed by culture.

The current study stands out in that the investigations were performed in a large number of individuals recruited from peripheral level health care clinics in high-burden settings in multiple countries from different ethnic regions of the African continent. Although there is a need to evaluate the performance of the biosignature in other high TB burdened regions, the inclusion of study participants from these different ethnic regions of the African continent implies that the signature identified in this study may be highly relevant across Africa and perhaps even globally. A limitation of this study was the lack of firmly established alternate diagnoses in the no-PTB group, which is difficult in primary health care settings. This however has no bearing on the importance of our findings as the goal of any TB diagnostic test is to distinguish individuals with TB disease from those presenting with similar symptoms due to conditions other than TB. The utility of this approach in difficult-to-diagnose TB groups such as paediatric and extra-pulmonary TB has to be investigated in future studies. As the HIV infected individuals in this study were not extensively staged with CD4 counts and viral loads, it is not certain whether severe HIV infection might have any influence on the performance of the biosignature. Therefore the influence of severe HIV infection on test performance as well as the effect of anti-retroviral therapy should be investigated in future studies. Future studies should also include samples from confirmed non-TB infectious or inflammatory diseases such as non-TB pneumonia and patients with sarcoidosis and other systemic inflammatory disorders, as such patient groups will be important in ascertaining the specificity of the biosignature for TB.

The biosignature identified in the current study warrants further development into a field-friendly point-of-care screening test for active TB, potentially based on lateral flow technology[38;39] and adapted for finger-prick blood. To allow appropriate point-of-care testing in remote settings, the final prototype would include a lightweight portable strip reader with

built-in software including an algorithm to interpret results obtained with LF strips comprising multiple cytokine test lines. Such a device is an improvement of the recently investigated UCP-LF platform in a multi-site evaluation study in Africa[40]. A cheap point-of-care test, with a high NPV of 96.4%, would identify patients who require confirmatory testing with gold standard tests such as culture and the GeneXpert, which are technically more demanding and have to be conducted in a centralized manner. A test with performance characteristics as demonstrated here would render about 75% of the GeneXpert tests currently performed in presumed TB cases for example in South Africa unnecessary, as most of the 70 to 75% of individuals that present with symptoms, are tested, and in whom TB disease is ruled out, would be identified by the point-of-care test, thereby leading to cost savings. The GeneXpert and culture tests could then be used as confirmatory tests in individuals with positive point-of-care test results and for drug susceptibility testing.

Conclusion

We have identified a promising seven-marker serum host protein biosignature for the diagnosis of active pulmonary TB disease in adults regardless of HIV infection status or ethnicity. These results hold promise for further development into a field-friendly point-of-care test for TB.

ACKNOWLEDGEMENTS

We are grateful to all our study participants, and support staff at the different laboratories that participated in the project. The following present or past members of the AE-TBC Consortium contributed to this work:

Stellenbosch University, **South Africa**: Gerhard Walzl, Novel N. Chegou, Magdalena Kriel, Gian van der Spuy, Andre G. Loxton, Kim Stanley, Stephanus Malherbe, Belinda Kriel, Leigh A Kotzé, Dolapo O. Awoniyi, Elizna Maasdorp

MRC Unit, The Gambia: Jayne S Sutherland, Olumuyiwa Owolabi, Abdou Sillah, Joseph Mendy, Awa Gindeh, Simon Donkor, Toyin Togun, Martin Ota

Karonga Prevention Study, Malawi: Amelia C Crampin, Felanji Simukonda, Alemayehu Amberbir, Femia Chilongo, Rein Houben

Ethiopian Health and Nutrition Research Institute, Ethiopia: Desta Kassa, Atsbeha Gebrezgeabher, Getnet Mesfin, Yohannes Belay, Gebremedhin Gebremichael, Yodit Alemayehu

University of Namibia, Namibia: Marieta van der Vyver, Faustina N Amutenya, Josefina N Nelongo, Lidia Monye, Jacob A Sheehama, Scholastica Iipinge,

Makerere University, Uganda: Harriet Mayanja-Kizza, Ann Ritah Namuganga, Grace Muzanye, Mary Nsereko, Pierre Peters

Armauer Hansen Research Institute, Ethiopia: Rawleigh Howe, Adane Mihret, Yonas Bekele, Bamlak Tessema, Lawrence Yamuah

Leiden University Medical Centre, The Netherlands: Tom H.M. Ottenhoff, Annemieke Geluk, Kees L.M.C. Franken, Paul L.A.M. Corstjens, Elisa M. Tjon Kon Fat, Claudia J. de Dood, Jolien J. van der Ploeg-van Schip

Statens Serum Institut, Copenhagen, Denmark: Ida Rosenkrands, Claus Aagaard

Max Planck Institute for Infection Biology, Berlin, Germany: Stefan H.E. Kaufmann, Maria M. Esterhuyse

London School of Hygiene and Tropical Medicine, London, UK: Jacqueline M. Cliff, Hazel M. Dockrell

COMPETING INTERESTS

Chegou NN, Walzl G and Mihret A are listed as inventors on an international patent application on the work reported in this manuscript, application no: PCT/IB2015/051435, Filing date: 2015/02/26; Chegou NN and Walzl G are listed as co-inventors on other patents related to diagnostic biosignatures for TB disease including: PCT/IB2015/052751, Filing date: 15/04/2015; PCT/IB2013/054377/US14/403,659, Filing date: 2014/11/25.

FUNDING

This work was supported by the European and Developing Countries Clinical Trials Partnership (EDCTP), grant number IP_2009_32040), through the African European Tuberculosis Consortium (AE-TBC, www.ae-tbc.euwww.ae-tbc.eu), with Prof. Gerhard Walzl as Principal Investigator.

Reference List

(1) World Health Organisation. Global Tuberculosis Report 2015.

(2) Chegou NN, Hoek KG, Kriel M, Warren RM, Victor TC, Walzl G. Tuberculosis assays: past, present and future. Expert Rev Anti Infect Ther 2011; 9(4):457-469.

(3) Trebucq A, Enarson DA, Chiang CY, Van DA, Harries AD, Boillot F et al. Xpert((R)) MTB/RIF for national tuberculosis programmes in low-income countries: when, where and how? Int J Tuberc Lung Dis 2011; 15(12):1567-1572.

(4) Marais BJ, Pai M. New approaches and emerging technologies in the diagnosis of childhood tuberculosis. Paediatr Respir Rev 2007; 8(2):124-133.

(5) Chegou NN, Walzl G, Bolliger CT, Diacon AH, van den Heuvel MM. Evaluation of adapted whole-blood interferon-gamma release assays for the diagnosis of pleural tuberculosis. Respiration 2008; 76(2):131-138.

(6) Phalane KG, Kriel M, Loxton AG, Menezes A, Stanley K, van der Spuy GD et al. Differential expression of host biomarkers in saliva and serum samples from individuals with suspected pulmonary tuberculosis. Mediators Inflamm 2013; 2013:981984.

(7) Chegou NN, Black GF, Loxton AG, Stanley K, Essone PN, Klein MR et al. Potential of novel Mycobacterium tuberculosis infection phase-dependent antigens in the diagnosis of TB disease in a high burden setting. BMC Infect Dis 2012; 12(1):10.

(8) Essone PN, Chegou NN, Loxton AG, Stanley K, Kriel M, van der SG et al. Host cytokine responses induced after overnight stimulation with novel M. tuberculosis infection phase-dependent antigens show promise as diagnostic candidates for TB disease. Plos One 2014; 9(7):e102584.

(9) Goletti D, Vincenti D, Carrara S, Butera O, Bizzoni F, Bernardini G et al. Selected RD1 peptides for active tuberculosis diagnosis: comparison of a gamma interferon whole-blood enzyme-linked immunosorbent assay and an enzyme-linked immunospot assay. Clin Diagn Lab Immunol 2005; 12(11):1311-1316.

(10) Chegou NN, Black GF, Kidd M, van Helden PD, Walzl G. Host markers in Quantiferon supernatants differentiate active TB from latent TB infection : preliminary report. BMC Pulm Med 2009; 9(1):21.

(11) Chegou NN, Essone PN, Loxton AG, Stanley K, Black GF, van der Spuy GD et al. Potential of host markers produced by infection phase-dependent antigen-stimulated cells for the diagnosis of tuberculosis in a highly endemic area. PLoS ONE 2012; 7(6):e38501.

(12) Chegou NN, Detjen AK, Thiart L, Walters E, Mandalakas AM, Hesseling AC et al. Utility of host markers detected in quantiferon supernatants for the diagnosis of tuberculosis in children in a high-burden setting. PLoS ONE 2013; 8(5):e64226.

(13) Frahm M, Goswami ND, Owzar K, Hecker E, Mosher A, Cadogan E et al. Discriminating between latent and active tuberculosis with multiple biomarker responses. Tuberculosis (Edinb) 2011; 91(3):250-256.

(14) Kellar KL, Gehrke J, Weis SE, Mahmutovic-Mayhew A, Davila B, Zajdowicz MJ et al. Multiple cytokines are released when blood from patients with tuberculosis is

stimulated with Mycobacterium tuberculosis antigens. PLoS ONE 2011; 6(11):e26545.

(15) Baumann R, Kaempfer S, Chegou NN, Oehlmann W, Loxton AG, Kaufmann SH et al. Serologic diagnosis of tuberculosis by combining Ig classes against selected mycobacterial targets. J Infect 2014; 69(6):581-589.

(16) Baumann R, Kaempfer S, Chegou NN, Oehlmann W, Spallek R, Loxton AG et al. A Subgroup of Latently Mycobacterium tuberculosis Infected Individuals Is Characterized by Consistently Elevated IgA Responses to Several Mycobacterial Antigens. Mediators of Inflammation 2015. 2015.

(17) Chegou NN, Heyckendorf J, Walzl G, Lange C, Ruhwald M. Beyond the IFN-gamma horizon: biomarkers for immunodiagnosis of infection with Mycobacterium tuberculosis. Eur Respir J 2014; 43(5):1472-1486.

(18) Claassens MM, du TE, Dunbar R, Lombard C, Enarson DA, Beyers N et al. Tuberculosis patients in primary care do not start treatment. What role do health system delays play? Int J Tuberc Lung Dis 2013; 17(5):603-607.

(19) Fluss R, Faraggi D, Reiser B. Estimation of the Youden Index and its Associated Cutoff Point. Biometrical Journal 2005; 47(4):458-472.

(20) Statsoft. General Discriminant Analysis (GDA). Statsoft Electronic Statistics Textbook. http://www.statsoft.com/textbook/general-discriminant-analysis/button/1 Date Accessed: 07 January 2016

(21) Breiman L. Random Forests. Machine Learning 2001; 45(1):5-32.

(22) Black S, Kushner I, Samols D. C-reactive Protein. J Biol Chem 2004; 279(47):48487-48490.

(23) de Beer FC, Nel AE, Gie RP, Donald PR, Strachan AF. Serum amyloid A protein and C-reactive protein levels in pulmonary tuberculosis: relationship to amyloidosis. Thorax 1984; 39(3):196-200.

(24) Kotru M, Rusia U, Sikka M, Chaturvedi S, Jain AK. Evaluation of serum ferritin in screening for iron deficiency in tuberculosis. Ann Hematol 2004; 83(2):95-100.

(25) Pandey R, Rodriguez GM. A ferritin mutant of Mycobacterium tuberculosis is highly susceptible to killing by antibiotics and is unable to establish a chronic infection in mice. Infect Immun 2012; 80(10):3650-3659.

(26) Opolot JO, Theron AJ, Anderson R, Feldman C. Acute phase proteins and stress hormone responses in patients with newly diagnosed active pulmonary tuberculosis. Lung 2015; 193(1):13-18.

(27) Visser A, van d, V. Severe hyperferritinemia in Mycobacteria tuberculosis infection. Clin Infect Dis 2011; 52(2):273-274.

(28) Singh M, Anand L. Bedside procalcitonin and acute care. Int J Crit Illn Inj Sci 2014; 4(3):233-237.

(29) Hoeboer SH, van der Geest PJ, Nieboer D, Groeneveld AB. The diagnostic accuracy of procalcitonin for bacteraemia: a systematic review and meta-analysis. Clin Microbiol Infect 2015; 21(5):474-481.

(30) Schleicher GK, Herbert V, Brink A, Martin S, Maraj R, Galpin JS et al. Procalcitonin and C-reactive protein levels in HIV-positive subjects with tuberculosis and pneumonia. Eur Respir J 2005; 25(4):688-692.

(31) Montecucco F, Favari E, Norata GD, Ronda N, Nofer JR, Vuilleumier N. Impact of systemic inflammation and autoimmune diseases on apoA-I and HDL plasma levels and functions. Handb Exp Pharmacol 2015; 224:455-482.

(32) Ferreira VP, Pangburn MK, Cortes C. Complement control protein factor H: the good, the bad, and the inadequate. Mol Immunol 2010; 47(13):2187-2197.

(33) Carroll MV, Lack N, Sim E, Krarup A, Sim RB. Multiple routes of complement activation by Mycobacterium bovis BCG. Mol Immunol 2009; 46(16):3367-3378.

(34) Buxbaum JN, Reixach N. Transthyretin: the servant of many masters. Cell Mol Life Sci 2009; 66(19):3095-3101.

(35) Wang C, Li YY, Li X, Wei LL, Yang XY, Xu DD et al. Serum complement C4b, fibronectin, and prolidase are associated with the pathological changes of pulmonary tuberculosis. BMC Infect Dis 2014; 14:52.

(36) Luo H, Zhu B, Gong L, Yang J, Jiang Y, Zhou X. The value of serum prealbumin in the diagnosis and therapeutic response of tuberculosis: a retrospective study. Plos One 2013; 8(11):e79940.

(37) Agranoff D, Fernandez-Reyes D, Papadopoulos MC, Rojas SA, Herbster M, Loosemore A et al. Identification of diagnostic markers for tuberculosis by proteomic fingerprinting of serum. Lancet 2006; 368(9540):1012-1021.

(38) Corstjens PL, Chen Z, Zuiderwijk M, Bau HH, Abrams WR, Malamud D et al. Rapid assay format for multiplex detection of humoral immune responses to infectious disease pathogens (HIV, HCV, and TB). Ann N Y Acad Sci 2007; 1098:437-445.

(39) Corstjens PL, Zuiderwijk M, Tanke HJ, van der Ploeg-van Schip JJ, Ottenhoff TH, Geluk A. A user-friendly, highly sensitive assay to detect the IFN-gamma secretion by T cells. Clin Biochem 2008; 41(6):440-444.

(40) Corstjens PL, Tjon Kon Fat EM, de Dood CJ, van der Ploeg-van Schip JJ, Franken KL, Chegou NN et al. Multi-center evaluation of a user-friendly lateral flow assay to determine IP-10 and CCL4 levels in blood of TB and non-TB cases in Africa. Clin Biochem. Published Online First: 15 August 2015. doi: 10.1016/j.clinbiochem.2015.08.013

FIGURE LEGENDS

Figure 1: STARD diagram showing the study design and classification of study participants. CRF, case report form; TB, Pulmonary tuberculosis; No-PTB, Individuals presenting with symptoms and investigated for pulmonary TB but in whom TB disease was ruled out; ROC, Receiver operator characteristics.

Figure 2: Levels of host markers detected in serum samples from pulmonary TB cases (n=214) and individuals without TB disease (n=487) and receiver operator characteristics (ROC) plots showing the accuracies of these markers in the diagnosis of pulmonary TB disease, regardless of HIV infection status. Representative plots for CRP, SAA, IP-10, ferritin, IFN-γ and transthyretin are shown. Error bars in the scatter-dot plots indicate the median and Inter-quartile ranges.

Figure 3: Areas under the ROC curve for individual analytes. AUCs obtained after data from pulmonary TB and no-PTB patients were analysed after stratification according to HIV infection status is shown as histograms (A) or 'Before and after' graphs (B). Host markers that performed better in HIV infected individuals are indicated by an asterix.

Figure 4: **Inclusion of different analytes into host biosignatures for the diagnosis of TB disease.** (A) Frequency of analytes in the top 20 most accurate GDA seven-marker biosignatures for diagnosis of TB disease regardless of HIV infection status. (B) Importance of analytes in diagnostic biosignatures for pulmonary TB disease, irrespective of HIV infection as revealed by random forests analysis. (C) ROC curve showing the accuracy of the finally selected seven-marker GDA biosignature in the diagnosis of pulmonary TB disease irrespective of HIV status. (D) Frequency of analytes in the top 20 GDA biosignatures for diagnosis of TB disease in HIV-uninfected individuals. The ROC curve for TB Vs. No-PTB, regardless of HIV (C) was generated from the training dataset.