

1 **What is the key question?**

2 Are there serum host marker signatures, which are suitable for point-of-care tests that  
3 differentiate between active pulmonary TB and other conditions in individuals presenting with  
4 signs and symptoms suggestive of TB in primary health care settings in Africa?  
5

6 **What is the bottom line?**

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

11 **Why read on?**

12 The 7 serum marker biosignature identified in this large multi-centered study on 716 individuals with  
13 signs and symptoms suggestive of TB could form the basis of a rapid, point-of-care screening test, and  
14 with a sensitivity of 94% and negative predictive value of 96%, such a test would render about 75% of  
15 the currently performed GeneXpert or TB cultures unnecessary.  
16

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

20

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

26

27 <sup>1</sup>DST/NRF Centre of Excellence for Biomedical Tuberculosis Research and SAMRC Centre  
28 for Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of  
29 Medicine and Health Sciences, Stellenbosch University, PO Box 241, Cape Town, 8000, South  
30 Africa

31 <sup>2</sup>Vaccines and Immunity, Medical Research Council Unit, Fajara, The Gambia

32 <sup>3</sup>Karonga Prevention Study, Chilumba, Malawi

33 <sup>4</sup>Department of Molecular Cell Biology, Leiden University Medical Centre, PO Box 9600,  
34 2300 RC Leiden, The Netherlands

35 <sup>5</sup>Department of Infectious Diseases, Leiden University Medical Centre, PO Box 9600, 2300  
36 RC Leiden, The Netherlands

37 <sup>6</sup>Department of Medicine, Makerere University, Kampala, Uganda

38 <sup>7</sup>School of Medicine, Faculty of Health Sciences, University of Namibia, Namibia

39 <sup>8</sup>Department of Infectious Disease Immunology, Statens Serum Institut, Copenhagen 2300s,  
40 Denmark

41 <sup>9</sup>Centre for Statistical Consultation, Department of Statistics and Actuarial Sciences,  
42 Stellenbosch University, Cape Town, South Africa

43 <sup>10</sup>Department of Immunology and Infection, London School of Hygiene and Tropical  
44 Medicine, Keppel Street, London WC1E 7HT, UK

45 <sup>11</sup>Department of Immunology, Max Planck Institute for Infection Biology, Charitéplatz 1,  
46 10117 Berlin, Germany

47

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

51 University, PO Box 241, Cape Town, 8000, South Africa, Telephone: +27219389158, Fax:  
52 +27219389863, E-mail: [gwalzl@sun.ac.za](mailto:gwalzl@sun.ac.za)

53

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

60

61

62 **Keywords:**

63 Sensitivity, specificity, tuberculosis, biomarker, diagnosis

64

65 **Word Count:** 3648

66

67

68

69 **ABSTRACT**

70 **Background**

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

76

77 **Methods**

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

84

85 **Results**

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

93

94 **Conclusion:**

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

98

99

100

101 **INTRODUCTION:**

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

119

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

126

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

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

135

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

142

## 143 **METHODS**

### 144 **Study participants**

145 We prospectively recruited adults who presented with symptoms requiring investigation for  
146 pulmonary TB disease at primary health care clinics at five field sites in five African countries.  
147 The clinics served as field study sites for researchers at Stellenbosch University (SUN), South  
148 Africa; Makerere University (UCRC), Uganda; Medical Research Council Unit (MRC), The  
149 Gambia; Karonga Prevention Study (KPS), Malawi; and the University of Namibia (UNAM),  
150 Namibia, as part of the African European Tuberculosis Consortium (AE-TBC) for TB  
151 Diagnostic Biomarkers ([www.ae-tbc.eu](http://www.ae-tbc.eu)). Study participants were recruited between November  
152 2010 and November 2012. All study participants presented with persistent cough lasting  $\geq 2$   
153 weeks and at least one of either fever, malaise, recent weight loss, night sweats, knowledge of  
154 close contact with a TB patient, haemoptysis, chest pain or loss of appetite. Participants were  
155 eligible for the study if they were 18 years or older and willing to give written informed consent  
156 for participation in the study, including consent for HIV testing. Patients were excluded if they  
157 were pregnant, had not been residing in the study community for more than 3 months, were  
158 severely anaemic (haemoglobin  $< 10\text{g/l}$ ), were on anti-TB treatment, had received anti-TB  
159 treatment in the previous 90 days or if they were on quinolone or aminoglycoside antibiotics  
160 during the past 60 days. The study protocol was approved by the Health Research Ethics  
161 Committees of the participating institutions.

162

### 163 **Sample collection and microbiological diagnostic tests**

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

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

175

### 176 **Classification of study participants and reference standard**

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

186

187 **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

188 Abbreviations: CXR, chest X ray; MTB, *Mycobacterium tuberculosis*; TB, pulmonary  
189 tuberculosis, No-PTB, non-“pulmonary tuberculosis”.

190

### 191 **Multiplex immunoassays**

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

206

### 207 **Statistical analysis**

208 Differences in analyte concentrations between participants with TB disease and those without  
209 TB were evaluated by the Mann–Whitney U-test for non-parametric data analysis. The  
210 diagnostic accuracy of individual analytes was investigated by receiver operator characteristics  
211 (ROC) curve analysis. Optimal cut-off values and associated sensitivity and specificity were  
212 selected based on the Youden’s index<sup>19</sup>. The predictive abilities of combinations of analytes  
213 were investigated by General discriminant analysis (GDA)<sup>20</sup> and random forests<sup>21</sup>, following  
214 the training/test set approach. Briefly, patients were randomly assigned into the training set  
215 (70% of study participants, n=491) or test set (30%, n=210), regardless of HIV infection status  
216 or study site by the software used in data analysis (Statistica, Statsoft, Ohio, USA). These  
217 training and test sets were selected using random sampling, stratified on the dependent (TB)  
218 variable. The most accurate of the top 20 marker combinations identified in the training set  
219 were then evaluated on the test sample set.

220

## 221 **RESULTS**



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

226

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

<b>Study site</b>	<b>SUN</b>	<b>MRC</b>	<b>UCRC</b>	<b>KPS</b>	<b>UNAM</b>	<b>Total</b>
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 <sup>#</sup> , (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)

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

235

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

241

242

243

244

245

246

247

248

249 **Table 3: Characteristics of TB and no-PTB cases and individuals with “Questionable**  
 250 **TB” disease status.**

	<b>Definite TB (n=185)</b>	<b>Probable TB (n=29)</b>	<b>ALL TB (n=214)</b>	<b>No-PTB (n=487)</b>	<b>Questionable TB (n=6)</b>
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)

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

253

#### 254 **Utility of individual serum biomarkers in the diagnosis of TB disease**

255 All serum markers investigated showed significant differences ( $p < 0.05$ ) between the TB cases  
 256 and No-PTB cases except A2M and MMP-2 (Supplementary Table 1), irrespective of HIV  
 257 infection status. Concentrations of CFH, CRP, ferritin, fibrinogen, haptoglobin, IFN- $\alpha$ 2, IFN-  
 258  $\gamma$ , IL-1ra, IP-10, MMP-9, PCT, SAA, SAP, TGF- $\alpha$ , TNF- $\alpha$ , TPA, and VEGF were significantly  
 259 higher in the TB cases while those of ApoA-1, Apo-CIII, and transthyretin were higher in the  
 260 no-PTB cases (Supplementary Table 1). When the accuracy for the diagnosis of TB disease  
 261 was investigated by ROC curve analysis, the areas under the ROC curve (AUC) were between  
 262 0.70 and 0.84 for 10 analytes: CRP, ferritin, fibrinogen, IFN- $\gamma$ , IP-10, TGF- $\alpha$ , TPA,  
 263 transthyretin, SAA and VEGF (Figure 2). Sensitivity and specificity were both  $>70\%$  for six  
 264 of these analytes, namely; CRP, ferritin, IFN- $\gamma$ , IP-10, transthyretin and SAA (Supplementary  
 265 Table 1).

266

#### 267 **Supplementary Table 1: Median levels of analytes detected in serum samples from** 268 **individuals with pulmonary TB disease (n=214) or no-PTB disease (n=487), and** 269 **accuracies in the diagnosis of TB disease**

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

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

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

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

277  
278

279 Accuracy of individual host markers in HIV-uninfected study participants

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

288

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

290 General discriminant analysis (GDA) models showed optimal prediction of pulmonary TB  
291 disease with seven-marker combinations. The most accurate seven-marker biosignature for the  
292 diagnosis of TB disease, regardless of HIV infection status, was a combination of ApoA-1,  
293 CFH, CRP, IFN- $\gamma$ , IP-10, SAA and transthyretin. Without any model “supervision”, this  
294 biosignature ascertained TB disease with a sensitivity of 86.7% (95% CI, 79.9-91.5%) and  
295 specificity of 85.3% (95% CI, 81.0-88.8%) in the training dataset (n=491; 168 TB and 323 no-  
296 PTB), and a sensitivity of 81.3% (95% CI, 69.2-89.5%) and specificity of 79.5% (95% CI,  
297 71.8-85.5%) in the test dataset (n=210, 77 TB and 133 No-PTB). To improve test performance,  
298 we optimised the model for higher sensitivity at the expense of lower specificity, which would  
299 allow the test to be used as a screening tool. The amended cut-off values ascertained TB disease  
300 with a sensitivity of 90.7% (95% CI, 84.5-94.6%) and specificity of 74.8% (95% CI, 69.8-

301 79.2%) in the training dataset (n=491), and sensitivity of 93.8% (95% CI: 84.0-98.0) and  
 302 specificity of 73.3% (95% CI, 65.2-80.1%) in the test dataset (n=210). The positive and  
 303 negative predictive values (NPV) of the biosignature were 60.6% (95% CI, 50.3-70.1 %) and  
 304 96.4% (95% CI, 90.5-98.8%), respectively (Table 4). The AUC for the seven-marker  
 305 biosignature (determined on the training sample set) was 0.91 (95% CI, 0.88-0.94) (Figure 4).  
 306

307 The random forest modelling approach gave similar prediction accuracies for TB and no-PTB  
 308 as GDA (87% sensitivity and 83% specificity in the training sample set, and 83% sensitivity  
 309 and 89% specificity in the test sample set), without selection of any preferred cut-off values.  
 310 In addition to the seven analytes included in the optimal GDA biosignature, Apo-CIII, ferritin,  
 311 fibrinogen, MMP-9 and TNF- $\alpha$  were also identified as important contributors to top models by  
 312 the random forest analysis (Figure 4).  
 313

314 **Table 4: Accuracy of the seven-marker serum protein biosignature (ApoA-1, CFH,**  
 315 **CRP, IFN- $\gamma$ , IP-10, SAA, transthyretin) in the diagnosis of TB disease regardless of HIV**  
 316 **infection status.**  
 317

**Training set  
(n=491)**

	<b>Sensitivity</b>	<b>Specificity</b>	<b>PPV</b>	<b>NPV</b>
%, (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)**

	<b>Sensitivity</b>	<b>Specificity</b>	<b>PPV</b>	<b>NPV</b>
--	--------------------	--------------------	------------	------------

%, (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)

318

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

320

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

330

331 Accuracy of serum biosignatures in individuals without HIV infection

332

333 In the absence of HIV infection the GDA procedure indicated optimal diagnosis of TB disease  
 334 when markers were used in combinations of four with ApoA-1, IFN- $\gamma$ , IP-10 and SAA  
 335 constituting the top model with sensitivity of 76.5% (95% CI, 67.5–83.7%) and specificity of  
 336 91.1% (95% CI 86.7–94.1) in the training sample set (n=372, 115 TB and 257 no-PTB), and a  
 337 sensitivity of 77.3% (95% CI, 61.8–88.0) and specificity of 87.1% (95% CI, 79.3–92.3%) in  
 338 the test dataset (n=160, 44 TB and 116 no-PTB). The positive and NPV of the four-marker  
 339 model in the test set were 69.4% (95% CI, 54.4–81.3%) and 91.0% (95% CI, 83.7–95.4%),  
 340 respectively.

341

342

343 **DISCUSSION**

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

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

351

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

361

362 IFN- $\gamma$ , IP-10 and TNF- $\alpha$ , together with other markers including IL-2 (reviewed in<sup>17</sup>), are  
363 amongst the most investigated host immunological biomarkers for the diagnosis of *M.tb*  
364 infection and disease. Both IFN- $\gamma$  and IP-10 showed potential in this study. The inclusion of  
365 these markers in the seven-marker model is not surprising, given their widely accepted roles in  
366 the pathogenesis of *M.tb* infection.

367

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

377

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

382 to be potentially useful in discriminating between pulmonary TB and community-acquired  
383 pneumonia in HIV-positive individuals<sup>30</sup>.

384

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

399

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

406

407 Combinations between transthyretin, CRP, SAA and neopterin ascertained TB disease with  
408 78% accuracy in a previous proteomic finger-printing study<sup>37</sup>. In our study, a biosignature  
409 containing transthyretin, CRP, SAA and markers involved in Th1-related immunity to TB  
410 (IFN- $\gamma$ , IP-10), an apolipoprotein and CFH showed excellent promise as a diagnostic tool for  
411 TB. Although most of these markers are promising individually<sup>17;23;26;27;30;35-37</sup>, single host  
412 markers have many shortcomings in predicting TB disease due to poor specificity. As observed  
413 in this large multi-centered pan-African study, the accuracy of different host markers is affected  
414 differentially by HIV infection. A biosignature containing different classes of biomarkers,  
415 produced by different cell types such as the classical Th1 immune-related markers plus acute-



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

424

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

445

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

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

462

### 463 **Conclusion**

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

467

468

469

470

### 471 **ACKNOWLEDGEMENTS**

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

475

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

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

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

483 **Ethiopian Health and Nutrition Research Institute, Ethiopia:** Desta Kassa, Atsbeha  
484 Gebrezgabher, Getnet Mesfin, Yohannes Belay, Gebremedhin Gebremichael, Yodit  
485 Alemayehu

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

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

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

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

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

496 **Max Planck Institute for Infection Biology, Berlin, Germany:** Stefan H.E. Kaufmann,  
497 Maria M. Esterhuysen

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

500

#### 501 **COMPETING INTERESTS**

502 Chegou NN, Walzl G and Mihret A are listed as inventors on an international patent application  
503 on the work reported in this manuscript, application no: PCT/IB2015/051435, Filing date:  
504 2015/02/26; Chegou NN and Walzl G are listed as co-inventors on other patents related to  
505 diagnostic biosignatures for TB disease including: PCT/IB2015/052751, Filing date:  
506 15/04/2015; PCT/IB2013/054377/US14/403,659, Filing date: 2014/11/25.

507

#### 508 **FUNDING**

509 This work was supported by the European and Developing Countries Clinical Trials  
510 Partnership (EDCTP), grant number IP\_2009\_32040), through the African European  
511 Tuberculosis Consortium (AE-TBC, [www.ae-tbc.eu](http://www.ae-tbc.eu)), with Prof. Gerhard Walzl  
512 as Principal Investigator.

513

514

515 Reference List

516

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

- 518 (2) Chegou NN, Hoek KG, Kriel M, Warren RM, Victor TC, Walzl G. Tuberculosis  
519 assays: past, present and future. *Expert Rev Anti Infect Ther* 2011; 9(4):457-469.
- 520 (3) Trebucq A, Enarson DA, Chiang CY, Van DA, Harries AD, Boillot F et al. Xpert((R))  
521 MTB/RIF for national tuberculosis programmes in low-income countries: when,  
522 where and how? *Int J Tuberc Lung Dis* 2011; 15(12):1567-1572.
- 523 (4) Marais BJ, Pai M. New approaches and emerging technologies in the diagnosis of  
524 childhood tuberculosis. *Paediatr Respir Rev* 2007; 8(2):124-133.
- 525 (5) Chegou NN, Walzl G, Bolliger CT, Diacon AH, van den Heuvel MM. Evaluation of  
526 adapted whole-blood interferon-gamma release assays for the diagnosis of pleural  
527 tuberculosis. *Respiration* 2008; 76(2):131-138.
- 528 (6) Phalane KG, Kriel M, Loxton AG, Menezes A, Stanley K, van der Spuy GD et al.  
529 Differential expression of host biomarkers in saliva and serum samples from  
530 individuals with suspected pulmonary tuberculosis. *Mediators Inflamm* 2013;  
531 2013:981984.
- 532 (7) Chegou NN, Black GF, Loxton AG, Stanley K, Essone PN, Klein MR et al. Potential  
533 of novel *Mycobacterium tuberculosis* infection phase-dependent antigens in the  
534 diagnosis of TB disease in a high burden setting. *BMC Infect Dis* 2012; 12(1):10.
- 535 (8) Essone PN, Chegou NN, Loxton AG, Stanley K, Kriel M, van der SG et al. Host  
536 cytokine responses induced after overnight stimulation with novel *M. tuberculosis*  
537 infection phase-dependent antigens show promise as diagnostic candidates for TB  
538 disease. *Plos One* 2014; 9(7):e102584.
- 539 (9) Goletti D, Vincenti D, Carrara S, Butera O, Bizzoni F, Bernardini G et al. Selected  
540 RD1 peptides for active tuberculosis diagnosis: comparison of a gamma interferon  
541 whole-blood enzyme-linked immunosorbent assay and an enzyme-linked immunospot  
542 assay. *Clin Diagn Lab Immunol* 2005; 12(11):1311-1316.
- 543 (10) Chegou NN, Black GF, Kidd M, van Helden PD, Walzl G. Host markers in  
544 Quantiferon supernatants differentiate active TB from latent TB infection :  
545 preliminary report. *BMC Pulm Med* 2009; 9(1):21.
- 546 (11) Chegou NN, Essone PN, Loxton AG, Stanley K, Black GF, van der Spuy GD et al.  
547 Potential of host markers produced by infection phase-dependent antigen-stimulated  
548 cells for the diagnosis of tuberculosis in a highly endemic area. *PLoS ONE* 2012;  
549 7(6):e38501.
- 550 (12) Chegou NN, Detjen AK, Thiart L, Walters E, Mandalakas AM, Hesselning AC et al.  
551 Utility of host markers detected in quantiferon supernatants for the diagnosis of  
552 tuberculosis in children in a high-burden setting. *PLoS ONE* 2013; 8(5):e64226.
- 553 (13) Frahm M, Goswami ND, Owzar K, Hecker E, Mosher A, Cadogan E et al.  
554 Discriminating between latent and active tuberculosis with multiple biomarker  
555 responses. *Tuberculosis (Edinb )* 2011; 91(3):250-256.
- 556 (14) Kellar KL, Gehrke J, Weis SE, Mahmutovic-Mayhew A, Davila B, Zajdowicz MJ et  
557 al. Multiple cytokines are released when blood from patients with tuberculosis is

- 558 stimulated with Mycobacterium tuberculosis antigens. PLoS ONE 2011;  
559 6(11):e26545.
- 560 (15) Baumann R, Kaempfer S, Chegou NN, Oehlmann W, Loxton AG, Kaufmann SH et  
561 al. Serologic diagnosis of tuberculosis by combining Ig classes against selected  
562 mycobacterial targets. J Infect 2014; 69(6):581-589.
- 563 (16) Baumann R, Kaempfer S, Chegou NN, Oehlmann W, Spallek R, Loxton AG et al. A  
564 Subgroup of Latently Mycobacterium tuberculosis Infected Individuals Is  
565 Characterized by Consistently Elevated IgA Responses to Several Mycobacterial  
566 Antigens. Mediators of Inflammation 2015. 2015.  
567
- 568 (17) Chegou NN, Heyckendorf J, Walzl G, Lange C, Ruhwald M. Beyond the IFN-gamma  
569 horizon: biomarkers for immunodiagnosis of infection with Mycobacterium  
570 tuberculosis. Eur Respir J 2014; 43(5):1472-1486.
- 571 (18) Claassens MM, du TE, Dunbar R, Lombard C, Enarson DA, Beyers N et al.  
572 Tuberculosis patients in primary care do not start treatment. What role do health  
573 system delays play? Int J Tuberc Lung Dis 2013; 17(5):603-607.
- 574 (19) Fluss R, Faraggi D, Reiser B. Estimation of the Youden Index and its Associated  
575 Cutoff Point. Biometrical Journal 2005; 47(4):458-472.
- 576 (20) Statsoft. General Discriminant Analysis (GDA). Statsoft Electronic Statistics  
577 Textbook. <http://www.statsoft.com/textbook/general-discriminant-analysis/button/1>  
578 Date Accessed: 07 January 2016
- 579 (21) Breiman L. Random Forests. Machine Learning 2001; 45(1):5-32.
- 580 (22) Black S, Kushner I, Samols D. C-reactive Protein. J Biol Chem 2004; 279(47):48487-  
581 48490.
- 582 (23) de Beer FC, Nel AE, Gie RP, Donald PR, Strachan AF. Serum amyloid A protein and  
583 C-reactive protein levels in pulmonary tuberculosis: relationship to amyloidosis.  
584 Thorax 1984; 39(3):196-200.
- 585 (24) Kotru M, Rusia U, Sikka M, Chaturvedi S, Jain AK. Evaluation of serum ferritin in  
586 screening for iron deficiency in tuberculosis. Ann Hematol 2004; 83(2):95-100.
- 587 (25) Pandey R, Rodriguez GM. A ferritin mutant of Mycobacterium tuberculosis is highly  
588 susceptible to killing by antibiotics and is unable to establish a chronic infection in  
589 mice. Infect Immun 2012; 80(10):3650-3659.
- 590 (26) Opolot JO, Theron AJ, Anderson R, Feldman C. Acute phase proteins and stress  
591 hormone responses in patients with newly diagnosed active pulmonary tuberculosis.  
592 Lung 2015; 193(1):13-18.
- 593 (27) Visser A, van d, V. Severe hyperferritinemia in Mycobacteria tuberculosis infection.  
594 Clin Infect Dis 2011; 52(2):273-274.

- 595 (28) Singh M, Anand L. Bedside procalcitonin and acute care. *Int J Crit Illn Inj Sci* 2014;  
596 4(3):233-237.
- 597 (29) Hoeboer SH, van der Geest PJ, Nieboer D, Groeneveld AB. The diagnostic accuracy  
598 of procalcitonin for bacteraemia: a systematic review and meta-analysis. *Clin*  
599 *Microbiol Infect* 2015; 21(5):474-481.
- 600 (30) Schleicher GK, Herbert V, Brink A, Martin S, Maraj R, Galpin JS et al. Procalcitonin  
601 and C-reactive protein levels in HIV-positive subjects with tuberculosis and  
602 pneumonia. *Eur Respir J* 2005; 25(4):688-692.
- 603 (31) Montecucco F, Favari E, Norata GD, Ronda N, Nofer JR, Vuilleumier N. Impact of  
604 systemic inflammation and autoimmune diseases on apoA-I and HDL plasma levels  
605 and functions. *Handb Exp Pharmacol* 2015; 224:455-482.
- 606 (32) Ferreira VP, Pangburn MK, Cortes C. Complement control protein factor H: the good,  
607 the bad, and the inadequate. *Mol Immunol* 2010; 47(13):2187-2197.
- 608 (33) Carroll MV, Lack N, Sim E, Krarup A, Sim RB. Multiple routes of complement  
609 activation by *Mycobacterium bovis* BCG. *Mol Immunol* 2009; 46(16):3367-3378.
- 610 (34) Buxbaum JN, Reixach N. Transthyretin: the servant of many masters. *Cell Mol Life*  
611 *Sci* 2009; 66(19):3095-3101.
- 612 (35) Wang C, Li YY, Li X, Wei LL, Yang XY, Xu DD et al. Serum complement C4b,  
613 fibronectin, and prolidase are associated with the pathological changes of pulmonary  
614 tuberculosis. *BMC Infect Dis* 2014; 14:52.
- 615 (36) Luo H, Zhu B, Gong L, Yang J, Jiang Y, Zhou X. The value of serum prealbumin in  
616 the diagnosis and therapeutic response of tuberculosis: a retrospective study. *Plos One*  
617 2013; 8(11):e79940.
- 618 (37) Agranoff D, Fernandez-Reyes D, Papadopoulos MC, Rojas SA, Herbster M,  
619 Loosemore A et al. Identification of diagnostic markers for tuberculosis by proteomic  
620 fingerprinting of serum. *Lancet* 2006; 368(9540):1012-1021.
- 621 (38) Corstjens PL, Chen Z, Zuiderwijk M, Bau HH, Abrams WR, Malamud D et al. Rapid  
622 assay format for multiplex detection of humoral immune responses to infectious  
623 disease pathogens (HIV, HCV, and TB). *Ann N Y Acad Sci* 2007; 1098:437-445.
- 624 (39) Corstjens PL, Zuiderwijk M, Tanke HJ, van der Ploeg-van Schip JJ, Ottenhoff TH,  
625 Geluk A. A user-friendly, highly sensitive assay to detect the IFN-gamma secretion  
626 by T cells. *Clin Biochem* 2008; 41(6):440-444.
- 627 (40) Corstjens PL, Tjon Kon Fat EM, de Dood CJ, van der Ploeg-van Schip JJ, Franken  
628 KL, Chegou NN et al. Multi-center evaluation of a user-friendly lateral flow assay to  
629 determine IP-10 and CCL4 levels in blood of TB and non-TB cases in Africa. *Clin*  
630 *Biochem*. Published Online First: 15 August 2015. doi: 10.1016/j.clinbiochem.2015.08.013  
631  
632  
633

634 **FIGURE LEGENDS**

635

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

640

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

647

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

652

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

662