

1 **A protein signature associated with active tuberculosis identified by plasma profiling**
2 **and network-based analysis**

3
4 Zaynab Mousavian^{1,2,3#}, Elin Folkesson^{1,4,#}, Gabrielle Fröberg^{1,5}, Fariba Foroogh^{1,2,4},
5 Margarida Correia-Neves^{1,6,7}, Judith Bruchfeld^{1,4,\$}, Gunilla Källenius^{1,2,4,\$}, Christopher
6 Sundling^{1,2,4,\$,§}

7
8 ¹ Division of Infectious Diseases, Department of Medicine Solna, Karolinska Institutet,
9 Stockholm, Sweden

10 ² Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden

11 ³ School of Mathematics, Statistics, and Computer Science, College of Science, University of
12 Tehran, Tehran, Iran

13 ⁴ Department of Infectious Diseases, Karolinska University Hospital, Stockholm, Sweden

14 ⁵ Department of Clinical Microbiology, Karolinska University Laboratory, Karolinska
15 University Hospital, Stockholm, Sweden

16 ⁶ Life and Health Sciences Research Institute, School of Medicine, University of Minho,
17 Braga, Portugal

18 ⁷ ICVS/3B's, PT Government Associate Laboratory, Braga, Portugal

19

20 # shared first author

21 \$ shared last author

22 § corresponding author: Christopher Sundling (christopher.sundling@ki.se).

23

24

25

26 **Abstract:**

27

28 **Objectives:** Tuberculosis (TB) is a bacterial infectious disease caused by *Mycobacterium*
29 *tuberculosis*. Annually, an estimated 10 million people are diagnosed with active TB, and
30 approximately 1.4 million dies of the disease. If left untreated, each person with active TB
31 will infect 10 to 15 new individuals every year. Therefore, interrupting disease transmission
32 by accurate early detection and diagnosis, paired with appropriate treatment is of major
33 importance. In this study, we aimed to identify biomarkers associated with the development
34 of active TB that can then be further developed for clinical testing.

35

36 **Methods:** We assessed the relative plasma concentration of 92 proteins associated with
37 inflammation in individuals with active TB (n=19), latent TB (n=13), or healthy controls
38 (n=10). We then constructed weighted protein co-expression networks to reveal correlations
39 between protein expression profiles in all samples. After clustering the networks into four
40 modules, we assessed their association with active TB.

41

42 **Results:** One module consisting of 16 proteins was highly associated with active TB. We
43 used multiple independent transcriptomic datasets from studies investigating respiratory
44 infections and non-TB diseases. We then identified and removed genes encoding proteins
45 within the module that were low expressed in active TB or associated with non-TB diseases,
46 resulting in a 12-protein plasma signature associated with active TB.

47

48 **Conclusion:** We identified a plasma protein signature that is highly enriched in patients with
49 active TB but not in individuals with latent TB or healthy controls and that also had minimal
50 cross-reactivity with common viral or bacterial lower respiratory tract infections.

51

52 **Keywords:** Mycobacterium Tuberculosis, Co-expression network analysis, Plasma protein
53 signature

54

55

56

57

58 **Introduction:**

59

60 Tuberculosis (TB), caused by *Mycobacterium tuberculosis* (Mtb), is an ongoing pandemic
61 responsible for approximately 10 million clinical cases and 1.4 million deaths annually,
62 making it the single deadliest infectious disease excluding the ongoing SARS-CoV2
63 pandemic. However, there are still many limitations in diagnostic methods available for
64 active TB (1, 2). Around one-fourth to one-fifth of the world's population is estimated to be
65 latently infected with Mtb (3), of which 5-10% are estimated to eventually develop active
66 disease.

67 Pulmonary TB is the most common clinical form, and it is diagnosed by detecting
68 Mtb in sputum samples by microscopy, nucleic acid amplification tests, such as PCR, and the
69 reference method mycobacterial culture. Several limitations to these methods exist; as
70 compared with culture positivity, sputum microscopy is positive only in approximately 50%
71 of cases (4) and PCR in approximately 90% of cases in respiratory samples (GeneXpert
72 RIF/TB and GeneXpert Ultra) (5, 6). Additional difficulties relate to specific patient groups,
73 such as HIV-infected patients and children, the latter commonly having low bacterial load
74 and cannot produce sputum. Mycobacterial culture can take several weeks to yield positive
75 results and requires specialized safety laboratories. In addition, it is often unavailable in
76 resource-poor settings where TB is more prevalent. In addition, the diagnosis of
77 extrapulmonary TB relies on invasive procedures to obtain samples for microbiological
78 analysis. Thus, diagnosis is frequently based on clinical and radiological findings or
79 algorithms, especially in low-resource settings. The diagnostic delays in endemic areas are
80 well-described (7) The gap between estimated and reported TB cases was more than 4 million
81 in 2020, and of those reported only 59% were microbiologically confirmed (8). No specific
82 blood test or biochemical marker has yet been introduced in the routine clinical work-up to
83 distinguish TB from other medical conditions. The need for non-sputum-based tests, both for
84 screening and diagnostic purposes is urgent and the requirements of those tests have been
85 described in detail in WHO statements for Target Product Profiles (9).

86 Over the last decades, various methods that examine the host response to Mtb
87 infection have been evaluated. Attempts at identifying TB-specific transcriptional, protein,
88 and metabolic signatures in patient blood samples were recently reviewed (10, 11). Promising
89 results have emerged for transcriptional signatures (12, 13) but no protein signature for active
90 TB has so far been validated in independent confirmatory studies. The protein signatures so
91 far identified show limited overlap and together with varying study designs and methods, this
92 makes a meta-analysis difficult (14).

93 In this study, we investigated the profile of 92 inflammatory proteins in the plasma
94 from a Swedish cohort including individuals with active TB, latent TB, and healthy controls
95 (Figure 1). Through weighted co-expression network analysis, we identified a signature that
96 was highly associated with active TB and disease severity. We refined the signature by
97 removing proteins associated with other bacterial and viral respiratory infections and
98 sarcoidosis. We then validated the signature in several independent transcriptional datasets
99 and showed it to be highly enriched in individuals with active TB.

100

101

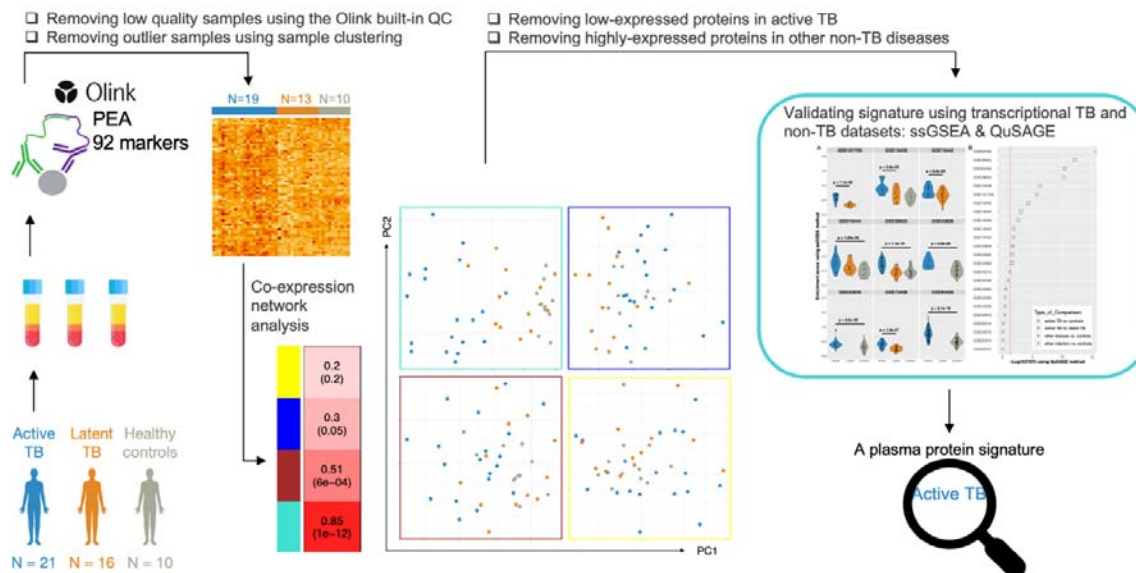


Figure 1: Flowchart of the study.

102

103

104

105 **Materials and methods:**

106

107 **Ethical considerations**

108 The study was registered and granted ethical permission from the Swedish Ethical Review
109 Board, EPN-number 2013/1347-31/2 and 2017/2262-32. All study subjects signed written
110 informed consent forms after receiving written and verbal information, in relevant cases
111 using professional interpreter services.

112

113 **Study participants**

114 Patients were recruited at the Karolinska University Hospital, Stockholm from May 2018 to
115 November 2019. Eligible for the study were: 1) Active TB patients within seven days of
116 treatment initiation. Active TB was defined through microbiological verification via Mtb
117 culture, or if culture negative, other microbiological positive result for Mtb (microscopy or
118 PCR) combined with clinical and radiological findings and response to anti-TB treatment
119 (active TB individuals are further described in Table S1); 2) Latent TB individuals with a
120 positive IGRA result (QuantiFERON-TB Gold In-Tube (QFT-GIT) or QuantiFERON-TB
121 Gold Plus (QFT-Plus) identified through contact investigation or screening of migrants, for
122 which active TB had been excluded. The cut-off between recent and remote TB was set to
123 two years after exposure; 3) Healthy controls with a negative IGRA-result and without
124 previously known TB-exposure, co-morbidities, or immunosuppression.

125

126 **Data collection**

127 Demographic, epidemiological, and clinical data for patients with active and latent TB were
128 extracted from patient charts. For all subjects this included information regarding previous
129 exposure to patients with active TB or infection, comorbidities and current medication,
130 radiological, biochemical, and microbiological test results. For the active TB cases, clinical
131 symptoms of TB disease were noted, and patients were classified according to pulmonary or
132 extrapulmonary TB. Microbiological samples for mycobacterial analysis were collected
133 independent of the study in accordance with clinical practice.

134

135 **Analysis of plasma proteins by proximity extension assay**

136 Venous blood samples were collected in heparin vacutainer tubes and transferred to the
137 laboratory for immediate preparation. The tubes were centrifuged at 670 *g* for 8 min after
138 which the plasma fraction was aliquoted and stored at -80°C .

139 Patient plasma was analysed using the Olink proximity extension assay, a qPCR
140 technology that simultaneously measures combinations of cytokines in preselected panels.
141 We used the 96-protein inflammatory panel, analysing 92 protein biomarkers and 4 controls
142 (Table S2). In this automated process, specific antibodies carrying single stranded matching
143 DNA bind in pairs to each of the target proteins allowing for DNA hybridization and
144 subsequent DNA-extension. The resulting DNA, unique for each target protein, is then
145 subject to PCR amplification and finally detection. The generated results consist of
146 normalized protein expression (NPX) values that correspond to \log_2 transformed relative
147 protein concentrations. The assay was performed by the Translational Plasma Profile Facility
148 at SciLifeLab, Stockholm, Sweden.

149

150 **Protein expression data analysis**

151 Repeated samples were used between experimental batches for running bridge normalization.
152 We used *read_NPX* and *olink_normalization* functions from the OlinkAnalyze R package
153 (<https://github.com/Olink-Proteomics/OlinkRPackage>) to read the \log_2 NPX protein
154 expression values and perform bridge normalization between batches of our data respectively.
155 Proteins with NPX values below the limit of detection (LOD) in more than 30% of samples
156 were filtered out. To remove batch effects from the final dataset, we applied the
157 *removeBatchEffect* function from the Limma R package (15) (Figure S1). We also performed
158 differential expression analysis using the Limma R package.

159

160 **Weighted co-expression network analysis**

161 We used the WGCNA R package to construct a weighted protein co-expression network
162 among the proteins of our dataset. *Pearson correlation* and the *signedhybrid* network type
163 were used in the *adjacency* function of the WGCNA package. In the *signedhybrid* network
164 type, only links associated to positive correlations were retained in the network, and negative
165 correlations were discarded. Since most of the biological networks have a scale-free
166 topology, the WGCNA package in the *pickSoftThreshold* function tries to find the best value
167 of the power parameter to make a scale-free network. After network construction, the
168 network was clustered into modules containing proteins that were highly positively co-
169 expressed using a hierarchical clustering algorithm as implemented in the *cutreeDynamic*
170 function of the WGCNA package with parameters *deepSplit* = 4 and *minClusterSize* = 5, and
171 other parameters set as default. The expression profiles of proteins in each module were
172 summarized by module eigengenes. In the *moduleEigengenes* function from the WGCNA
173 package, the first principal component of the expression data of each module is measured as a
174 module eigengene for that module. To find a module associated with TB progression, the
175 correlation between each module eigengene and the trait vector was computed to identify
176 which module that had a significant correlation with TB progression.

177

178 **Protein signature enrichment analysis**

179 To further validate the protein signature obtained from the co-expression network analysis,
180 we applied two computational methods (ssGSEA and QuSAGE) for enrichment analysis in
181 independent transcriptomic datasets. In ssGSEA, the enrichment scores of the protein
182 signature were calculated per sample based on the absolute value of proteins expression in
183 that sample to quantify how much the protein signature was overrepresented in a specific
184 sample. Moreover, to verify that the protein signature was specific to active TB and not to
185 other non-TB respiratory infections or diseases with clinical presentations similar to active

186 TB, the *quusage* function from the QuSAGE R package was applied to datasets comparing
187 active TB or non-TB disease to various control groups. The gene set differential expression
188 was calculated by combining individual probability functions obtained from individual
189 differential expression of genes in a particular comparison. A p-value, calculated by
190 QuSAGE, determined the statistical significance of a gene set for each given comparison. We
191 also used different types of single sample gene set enrichment analysis algorithms, including
192 ssGSEA, GSEA (16) and zscore (17), all implemented in the gsva R package, to compare all
193 co-expressed modules in terms of enrichment in active TB versus latent TB and healthy
194 controls.

195

196 **Results**

197

198 **Clinical characteristics of the study participants**

199 The plasma of 21 individuals with active TB, 16 individuals with latent TB, and 10 healthy
200 controls were analyzed using the Olink inflammation proximity extension assay. During data
201 processing, 5 samples (2 active TB and 3 latent TB) were excluded for either failing the
202 quality control check included in the OlinkRPackage or identified as outliers by cluster
203 analysis (Figure S2). The characteristics of the remaining study participants are further
204 described in Table 1.

205 As a reflection of the low TB incidence in Sweden, most study individuals originated
206 from other countries, mainly situated in Africa, Asia and Eastern Europe. Fifteen of 19 active
207 cases had pulmonary TB including 3 with pleuritis, 6 of which were sputum smear positive.
208 Four had extrapulmonary TB cases of which two with disseminated disease. Only 1 out of the
209 19 active TB patients was not confirmed by microbiologic culture; this patient had a positive
210 PCR for *M. tuberculosis* in a lymph node aspirate and gastric lavage as well as radiologic
211 signs of active pulmonary TB and showed a clinical response to TB-treatment. The active TB
212 cases were sampled within one week of treatment initiation except for one patient that was
213 sampled at day 9. Of the 13 latent TB patients, 10 had a known recent TB exposure 1-4
214 months prior to study inclusion. All but 2 latent TB individuals completed preventive TB
215 treatment, and none progressed to active TB during a follow up period of > 2 years. There
216 were few significant co-morbidities, with no patient in either group on immunosuppressive
217 treatment. All active TB patients were HIV negative. The latent TB and healthy controls were
218 not routinely tested for HIV infection.

219 From the 10 healthy controls, 1 had a QFT-Gold Plus result in the low range of the
220 borderline interval (0.35-0.99 IU/ml) but with no known exposure, potentially indicating a
221 false positive result (18, 19).

222

223

224 **Table 1. Clinical characteristics of the sample cohort**
225

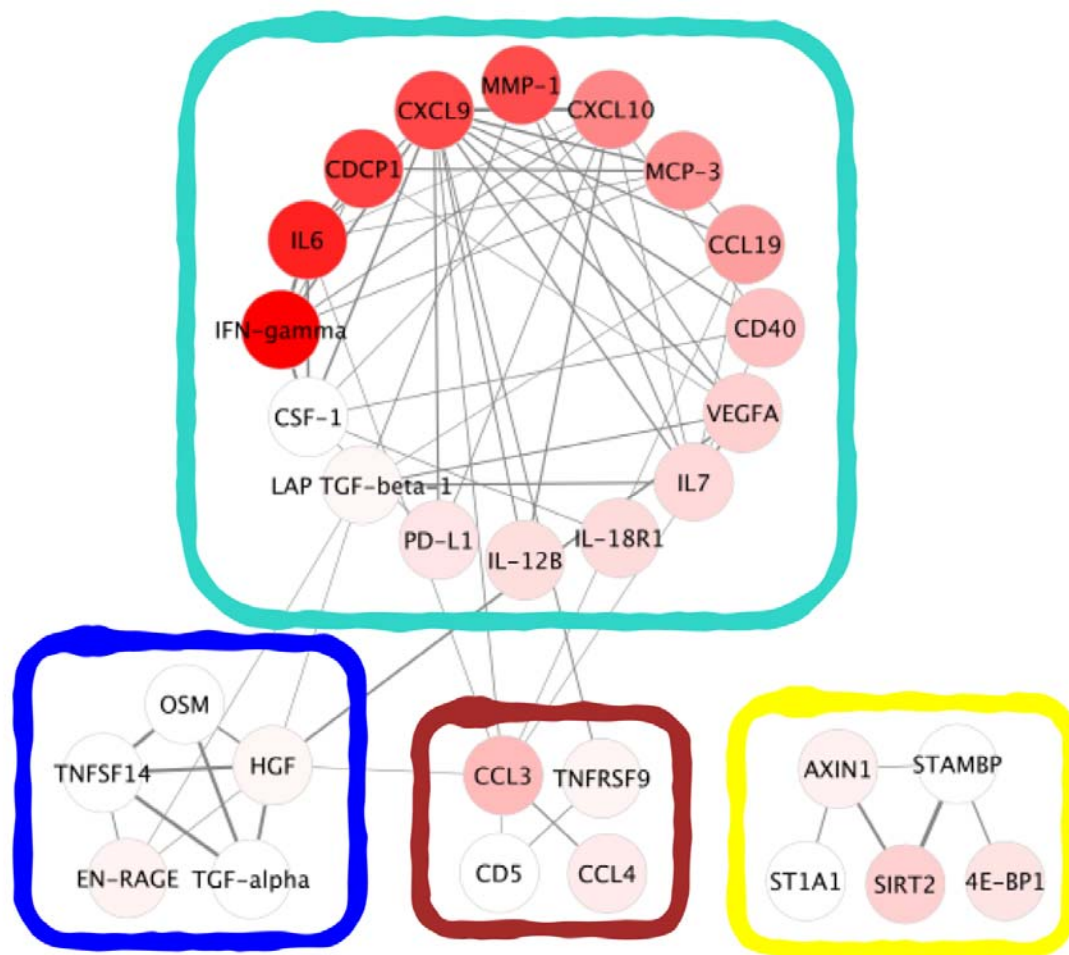
		Active TB (n=19)	Latent TB (n=13)	Healthy controls (n=10)
Men n (%)		8 (42)	8 (62)	6 (67)
Age, y (mean, range)		39 (20-71)	39 (18-72)	25 (21-37)
Origin from TB high endemic country, n		14	5	0
Time since immigration to Sweden, y		9 (0.2-45)	5 (0.5-15)	n/a
BCG-vaccination	yes/no/unknown	2/1/16	4/1/8	1/9/0
IGRA result	pos/neg/unknown	10/1/8	13/0/0	1 ⁶ /9/-
Previous Mtb infection	active ¹	3	0	
	latent	1	0	
Previous Mtb exposure ⁴	<2 years (recent)	4	10	
	>2 years (remote)	1	1	
Biochemistry ^{2,5} (mean, range)	CRP (mg/L)	25 (1-94)	1 (1-2)	
	ESR (mm)	51 (7-117)	6 (1-21)	
	WBC (10 ⁹ /L)	6.5 (3.8-11.8)	6.6 (5.6-9)	
	Hb (g/L)	126 (102-149)	145 (118-158)	
	Alb ³ (g/L)	32 (26-38)	41 (39-44)	

226 Patient origins; Active TB: Somalia (4) Eritrea (3), Sweden (3), Philippines (2), Peru, Bangladesh, Pakistan,
227 Mongolia, Poland, China, Afghanistan. Latent TB: Sweden (3), Romania (3), Moldavia (2), Mongolia, Ethiopia,
228 Eritrea, Ghana, India.
229 Comorbidities; Active TB: postpartum period, hypertension (2), intestinal schistosomiasis, chronic hepatitis B,
230 acute thyroiditis. Latent TB: chronic hepatitis B, hypertension (2)
231 C-reactive protein (CRP), Erythrocyte sedimentation rate (ESR), White blood cell count (WBC), Hemoglobin
232 (Hb), Albumin (Alb)
233 ¹Two patients previously treated for ATB, 1 and 4 years before; one patient not treated
234 ²ATB n=19 LTBI n=9
235 ³ATB n=17 LTBI n=7
236 ⁴Exposure self-reported or verified
237 ⁵Normal range; CRP<5, ESR <20, WBC 4.4-10.0, Hb >120 (F),>130 (M), Alb>38
238 ⁶Healthy controls; one QFT borderline (0.36/0.39), no exposure to TB
239

240
241 **Network construction reveals one module associated with active TB**

242 Following data pre-processing, including quality control, batch correction, and clustering the
243 relative level (NPX) of each of the 92 proteins was assessed. Proteins with more than 30% of
244 samples with NPX values below the limit of detection (n = 28) were excluded from further
245 analysis. Out of those 28 proteins, 25 had no differential expression between the groups
246 (Figure S3). A weighted protein co-expression network was constructed with the remaining
247 64 proteins to examine correlations between protein expression profiles in all included
248 samples (n = 42 individuals). We considered only the links associated with positive
249 correlations in the network reconstruction and selected power 8 to reach a scale-free topology
250 (Figure S4). After clustering the network into modules, four modules indicated by distinct
251 colours in Figure 2 were discovered (Table 2), including 34 proteins. The remaining 30
252 proteins that were not included in any of the modules, were discarded from further analysis.
253 We observed that the turquoise module out of the four modules had a significantly stronger
254 correlation with active TB (Table 2). A visual representation of the protein co-expression
255 network and the discovered protein modules were then generated using Cytoscape 3.0 (20)
256 (Figure 2).

257



258
259
260
261
262
263
264
265
266

Figure 2. A visual representation of the protein co-expression network and the protein modules: The protein modules are surrounded by the colours turquoise, blue, brown or yellow. Nodes are coloured based on protein abundance (Fold Change) in comparison of active TB vs. control, with a darker red indicating a larger fold change. Links are depicted with thickness proportional to the correlation and only links representing Pearson's $r > 0.6$ are shown.

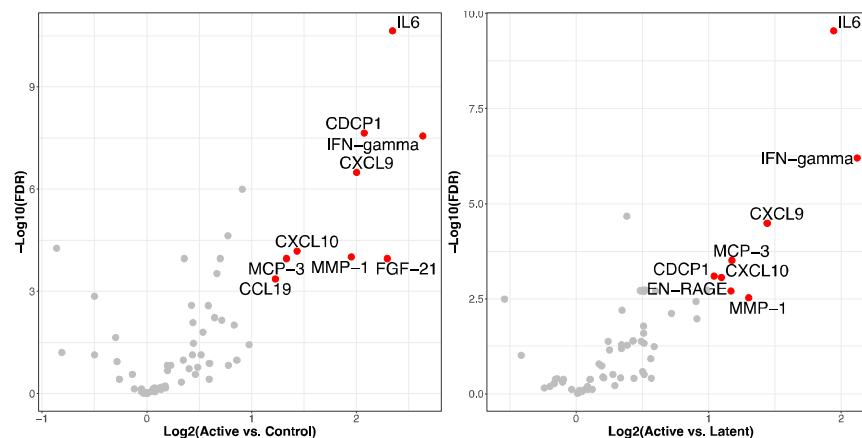
Table 2: Details about the modules of the protein co-expression network

Module	Number of proteins	Correlation	pValue
Turquoise	16	0.85	1E-12
Blue	7	0.3	5E-02
Brown	6	0.51	6E-04
Yellow	5	0.2	2E-01

267
268
269
270
271
272
273

We then assessed how the markers differed between individuals with active TB, latent TB, and healthy controls (Figure S5). All differentially expressed proteins ($\log_2(\text{FC}) \geq 1$ and FDR-corrected $p\text{-value} < 0.01$) between either active TB and latent TB or active TB and healthy controls except FGF-21 and EN-RANGE were included in the turquoise module (Figure 3). Eight of the 16 proteins in the turquoise module were highly expressed in active TB patients while the remaining 8 proteins were strongly co-expressed with those highly

274 expressed proteins (Figure 2). To illustrate how the turquoise module segregated active TB
275 individuals from latent TB and healthy controls, we performed principal component analysis
276 (PCA) on the expression profile of proteins of each module and used the first and the second
277 principal components (PC1 and PC2) to show differences in protein levels (Figure S6). All
278 active TB samples except one were clearly separated from the other samples by the proteins
279 in the turquoise module, suggesting that the included proteins can potentially serve as a
280 signature to identify individuals with active TB.
281



282
283 Figure 3. Differential expression analysis of all 64 proteins: Active TB vs. control (left) and active TB vs. latent
284 TB (right). Red points indicate the significant highly expressed proteins with $\log_2(\text{FC}) > 1$ and FDR-corrected
285 P-value < 0.01 in each comparison.
286

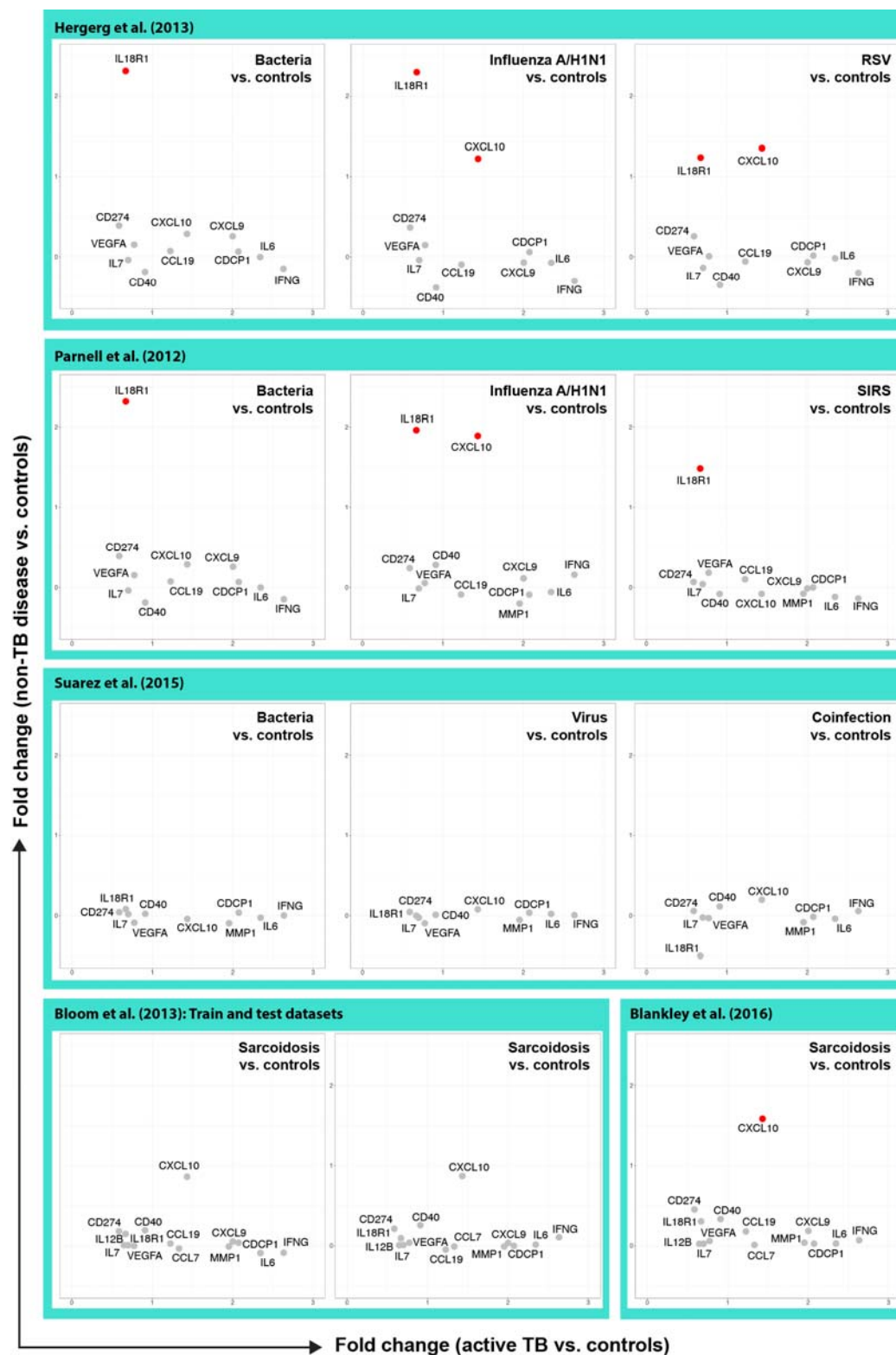
287 Validating the turquoise module in independent transcriptomic TB datasets

288 There is limited overlap in markers between studies investigating protein or gene signatures
289 aiming to discriminate active TB from latent TB, healthy controls, or other diseases (Figure
290 S7). Additionally, we did not identify any available pre-processed proteomic dataset
291 including all proteins of the turquoise module that we could use for validation. For this
292 reason, we chose to instead evaluate to what extent the proteins of the different modules
293 could distinguish active TB from latent TB and healthy controls in transcriptomic datasets.
294 We applied all four modules to several independent cohorts using multiple gene set
295 enrichment analysis methods. Nine transcriptomic TB datasets were selected based on the
296 criteria of age >15 , number of individuals per group >10 and no anti-TB treatment (Table
297 S3). The datasets were provided in the curatedTBData R package (21), which also included
298 the corresponding genes to all proteins included in the four modules. In total the datasets we
299 selected included >3000 individuals from four continents. Three enrichment methods (ssgsea,
300 gsva and zscore) demonstrated high enrichment of the turquoise module in active TB and low
301 enrichment in latent TB and the control group across most of the TB datasets (Figures S8-
302 S10). This was also the case for the blue module, although to a less extent compared with the
303 turquoise module. In contrast, neither the brown nor yellow modules were found highly
304 enriched in active TB using the same three methods on the same transcriptomic datasets
305 (Figures S8-S10).
306

307 Identification of a new 12-protein signature for active TB diagnosis

308 We identified the turquoise module as having the highest correlation with active TB.
309 However, two of the 16 proteins in the module, LAP TGF-beta-1 and CSF-1, had low fold
310 change values ($\log_2(\text{FC}) < 0.5$) when comparing active TB versus both latent TB and healthy
311 controls (Figure 3), and were therefore potentially redundant.

312 To investigate if the remaining 14 proteins were specific for active TB, we performed
313 differential expression analysis using multiple transcriptomic datasets including either viral or
314 bacterial lower respiratory tract infections (LRTI), systemic inflammatory response syndrome
315 (SIRS) or sarcoidosis (Figure 4). This allowed us to identify genes encoding the signature
316 proteins that were highly and significantly expressed ($\log_2(\text{FC}) > 1$ and $P\text{-value} < 0.05$) in
317 diseases with clinical symptoms overlapping with active TB. To this end, we used three
318 transcriptomic datasets, GSE42026 (22), GSE40012 (23) and GSE60244 (24), each
319 containing various types of lower respiratory infections. GSE42026 included (pediatric)
320 patients with severe LRTI of different etiologies; bacterial, (mostly *Streptococcus*
321 *pneumoniae*), Influenza A/H1N1/09 and Respiratory syncytial virus (RSV) infection.
322 GSE40012 included adult patients with severe community acquired pneumonia (CAP); either
323 bacterial or caused by Influenza A/H1N1, and SIRS (25), without evidence of infection.
324 Finally, GSE60244 included patients hospitalized for bacterial LRTI (*Streptococcus*
325 *pneumoniae* being most common), viral LRTI (Influenza A, B or RSV) and viral/bacterial
326 coinfection. We also conducted the same experiment using three datasets from two studies
327 containing sarcoidosis samples (26, 27). Two genes, corresponding to the proteins IL18R1
328 and CXCL10 (IP-10), stood out in these analyses (Figure 4). IL18R1 was highly expressed in
329 severe viral and bacterial LRTI and SIR, while CXCL10 was highly expressed in severe viral
330 LRTI and in one of the sarcoidosis studies. However, most of the proteins of the signature
331 were only observed to be expressed at high levels in individuals with active TB. Therefore, in
332 addition to LAP TGF-beta-1 and CSF-1 that were expressed only at very low levels, we also
333 removed IL18R1 and CXCL10 from the signature, leading to a 12-marker plasma signature
334 associated with active TB and with low cross-reactivity to other bacterial/viral lower
335 respiratory infections and sarcoidosis. The final signature consisted of the proteins IFN-
336 gamma, IL6, CDCP1, CXCL9, MMP-1, MCP-3, CCL19, CD40, VEGFA, IL7, IL-12B and
337 PDL-1.



338
339
340
341
342
343
344

Figure 4. The expression of proteins of the turquoise module in our Swedish cohort compared to other viral/bacterial infections, SIRS and sarcoidosis from different datasets. The fold-changes of proteins of the turquoise module in our cohort (active TB vs. controls; x-axis) compared to their fold-changes in other non-TB diseases (non-TB diseases vs. controls; y-axis). Red points indicate proteins of the turquoise module, which genes encoding them are significantly highly expressed in either other infections or other diseases ($\log_2(\text{FC}) > 1$ and $P\text{-value} < 0.05$). Row 1). Pediatric patients with severe respiratory tract infection: Bacterial: (n=18,

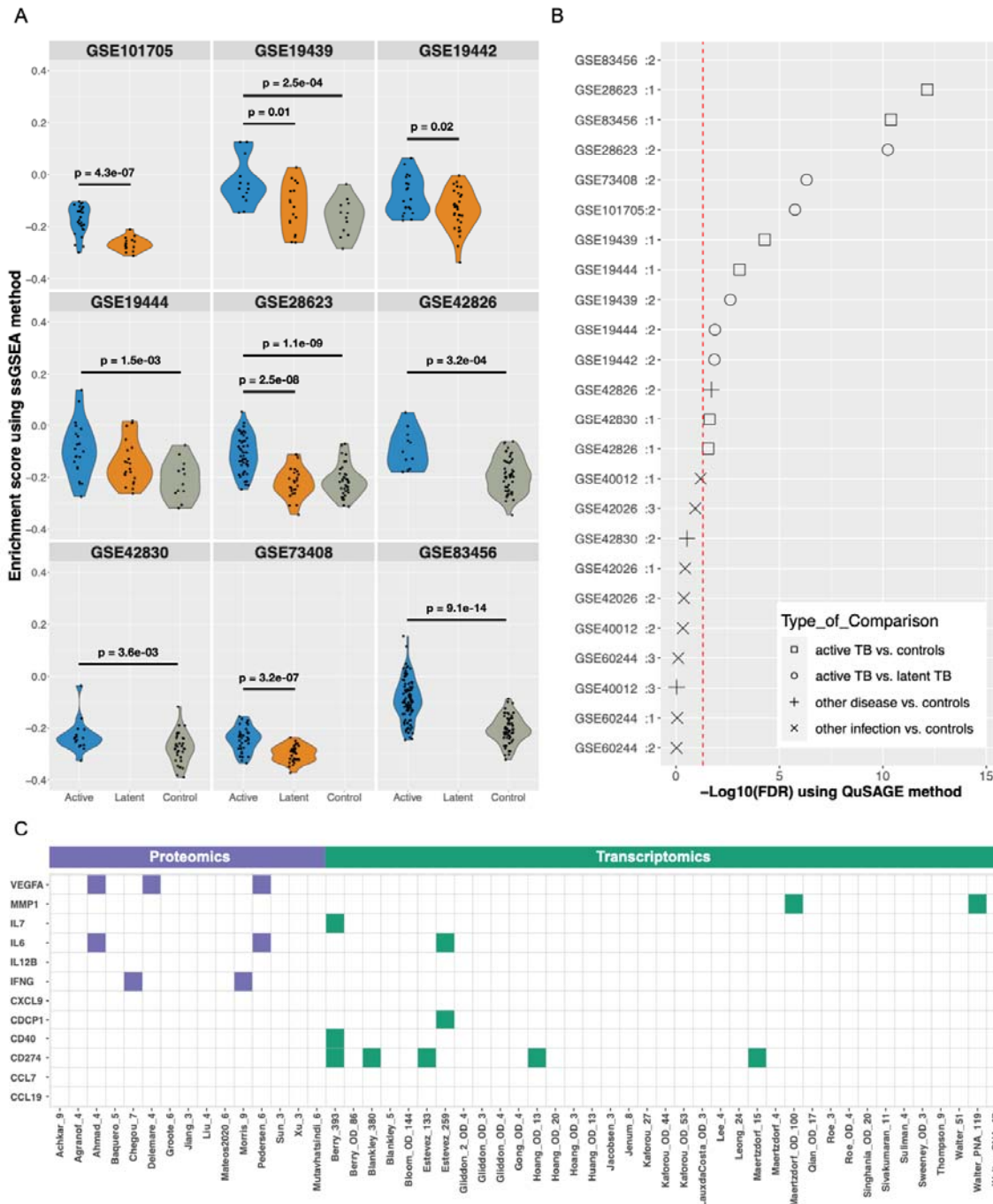
345 *Streptococcus pneumoniae* (12), *Streptococcus pyogenes* (4) *Staphylococcus aureus* (2) including 5 with viral co-
346 infection (non-Influenza A H1N1/RSV); Influenza A/H1N1 (n=19), RSV (n=22) HC (n=33). Row 2) Adult
347 patients with severe CAP or SIRS requiring ICU-care: Bacterial: (n=16, mixed etiology) Influenza
348 A/H1N1(n=8), SIRS without infection (n=12), HC (n=18). Row 3) Adult patients hospitalized for LRTI:
349 Bacterial (n=22; *S. pneumoniae* (13), *Moraxella catarrhalis* (4), *S. aureus* (2), mixed bacterial (3); viral (n=71,
350 Influenza A (32), RSV (17), Influenza B (9), HMPV (7); mixed bacterial/viral (n=25) HC (n=40). Row 4)
351 pulmonary sarcoidosis vs controls. RSV = respiratory syncytial virus; CAP = Community acquired pneumonia;
352 ICU = Intensive Care Unit; SIRS = systemic inflammatory response syndrome; LRTI = Lower respiratory tract
353 infection; HMPV = human metapneumovirus

354

355 **Validating the 12-marker signature in TB and non-TB proteomic and transcriptomic** 356 **datasets**

357 To assess the significance of the 12-marker signature in independent TB cohorts, we
358 performed ssGSEA on transcriptomic TB datasets from the curatedTBData R package (Table
359 S3). We observed significantly ($p < 0.05$) higher enrichment scores in the active TB group,
360 compared with either latent TB or controls in all the transcriptomic TB datasets except one
361 (GSE19444) (Figure 5A). We then used QuSAGE to compare the enrichment of the 12-
362 marker signature in TB infections with both other bacterial/viral infections and other
363 pulmonary diseases (Figure 5B). Three different comparisons were done using the
364 GSE42026, GSE40012 and GSE60244 datasets to assess the signature in respiratory
365 infections versus healthy controls, and in sarcoidosis disease versus healthy controls using
366 three transcriptomic datasets (GSE83456, GSE42826 and GSE42830) (Figure 5B). We
367 observed that the 12-marker signature was significantly overrepresented (FDR corrected P-
368 value < 0.05) in all comparisons between active TB and latent TB or healthy controls but was
369 not significantly enriched in those gene sets comparing other LRTIs vs. healthy controls.
370 However, in two of the three sarcoidosis datasets the signature was also enriched, indicating
371 that the signature might not, on its own, be able to distinguish active TB from sarcoidosis
372 without also weighing clinical data.

373 We then compared the 12-marker signature with the other published gene signatures
374 from the TBSignatureProfiler R package (28) and published protein signatures to investigate
375 overlap between protein and transcriptional signatures with the proteins of our signature
376 (Figure 5C). VEGFA, IL6, and IFN-gamma were identified in at least two other proteomic
377 studies (29-33), while CD274 (also called PD-L1) was observed in several published
378 transcriptional signatures. IP-10 (also called CXCL10), which was removed due to its high
379 expression in severe viral LRT and sarcoidosis appears in several proteomic studies (30-33).
380 The other markers were less common or identified in the current study. Although these
381 proteins have not been included in protein signatures before, they have been associated with
382 TB (34-37) and could potentially be generated via similar signaling pathways in response to
383 mycobacterial infection, such as has been indicated for signal transducer and activator of
384 transcription 1 (STAT1) in TB (38). To assess if this was the case, we used the StrongestPath
385 application (39) in Cytoscape to evaluate how the proteins were connected to different STAT
386 transcription factors based on data from the KEGG database (Figure S11). We observed that
387 several of the proteins were directly associated with STAT1, consistent with previous
388 literature (38) to a lesser extent with STAT3 and STAT4, and indirectly with STAT2 and
389 STAT5A.



390

391

392

393

394

395

396

397

398

399

400

401

402

403

Figure 5. Validation and overlap of the 12-marker protein signature.

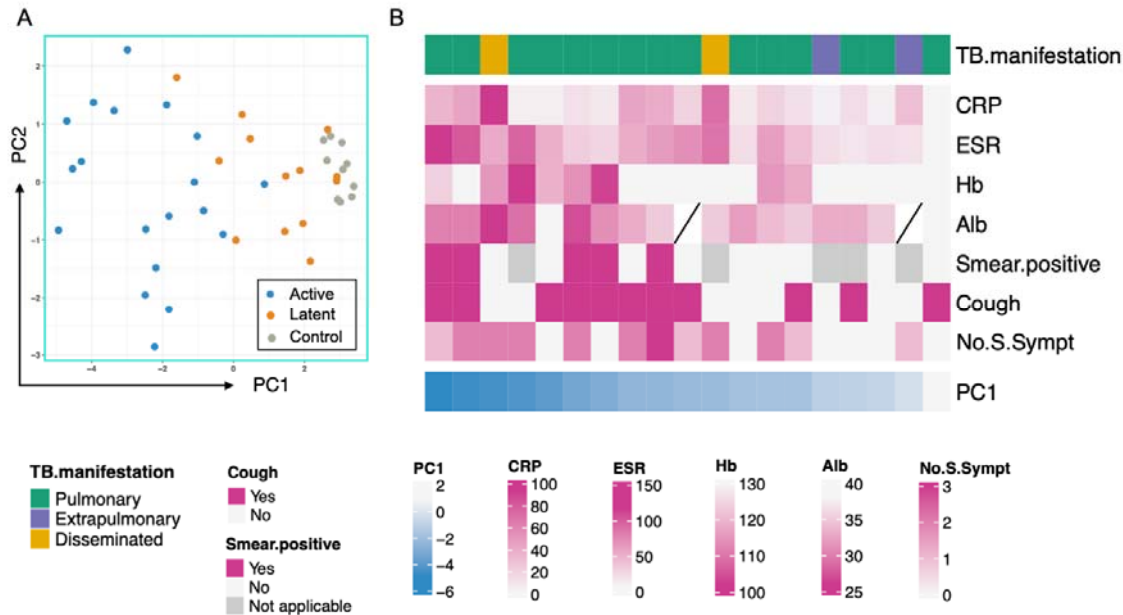
A) ssGSEA indicating enrichment score for the 12-marker signature in different transcriptomic datasets. Blue indicates individuals with active TB, orange indicates latent TB and grey indicates healthy controls. The p values show the statistical significance for the enrichment score difference between either active TB and Latent TB or Active TB and healthy controls using the wilcox.test. B) QuSAGE analysis with $-\log(\text{FDR})$ indicating capacity of the 12-marker signature in separating active TB from healthy controls (open boxes), active TB from latent TB (open circles), non-infectious inflammatory disease from healthy controls (+) and other LRTIs or SIRS from healthy controls (×). The red line indicates $\text{FDR} < 0.05$. C) Overlap between individual proteins in the 12-marker signature with other proteomic signatures (left) and transcriptomic signatures (right) from the TBSignatureProfiler R package identifying active TB.

404 **Association between the 12-marker signature and disease severity**

405 Principal component analysis of the 12-marker signature (Figure 6A) showed that, similar to
406 the turquoise module (Figure S6), active TB was primarily separated from latent TB and
407 healthy controls by principal component 1 (PC1). We therefore hypothesized that PC1 could
408 represent a proxy for the magnitude of the inflammatory response and disease severity. To
409 evaluate if this was the case, we stratified the individuals with active TB based on their PC1
410 values and a set of clinical features (Figure 6B). The heatmap shows that samples with a
411 smaller PC1 value, located towards the left side of heatmap, were associated with several
412 biochemical markers indicating more extensive inflammation, such as lower hemoglobin
413 (Hb) and albumin (Alb), and higher erythrocyte sedimentation rate (ESR) and C-reactive
414 protein (CRP). The presence of any systemic symptoms (fever, night sweats and weight loss)
415 and smear positivity were also associated with a lower PC1 value. We next performed a
416 regression analysis to see if individual or combinations of the clinical variables could explain
417 the PC1 values. A standard least squares model including ESR (estimate: -0.03 , 95% CI: $-$
418 $0.016 - -0.044$, $p = 0.0008$) and Alb (estimate: 0.21 , 95% CI: $0.08-0.34$, $p = 0.006$) were
419 highly associated with PC1 ($r^2 = 0.74$, $p < 0.0001$), clearly associating the signature with
420 clinical presentation (40).

421 The three active TB cases with the weakest PC1 signal had few symptoms and limited
422 disease activity and inflammation (Figure 6A). One patient had a culture-confirmed lymph
423 node TB. Another patient had a PCR positive but culture-negative lymph node and
424 pulmonary TB. The active TB case with the lowest inflammatory signal was found through
425 contact investigation very early in active disease progression with cough as the only symptom
426 but with no inflammation in laboratory tests. This suggests that the signature has the potential
427 to identify active TB very early on in the progression from latent TB.

428 Although the signature was not optimized for the detection of latent TB, the visually
429 apparent separation of latent TB from healthy controls (Figure 6A) could possibly indicate a
430 TB-specific immune response detectable in the plasma of these individuals. Two of the three
431 latent TB cases clustering with the healthy controls had no recent exposure to TB and were
432 clinically classified as remotely acquired infections (i.e., no known exposure to a TB case for
433 > 2 years). In summary, these data suggest that the signature could be helpful in identifying
434 TB disease progression or cure.



435

436

437 Figure 6. Association between the first principal component (PC1) and clinical characteristics of active TB. A)

438 PC1 and PC2 indicate variation in concentrations of the 12-marker signature for individuals with active TB

439 (blue dots), latent TB (orange dots) and healthy controls (grey dots). B) Clinical characteristics (rows) of all

440 active TB donors (columns) stratified by the negative value of PC1 (the more to the left indicating a stronger

441 signature). TB manifestation: Pulmonary: (including 3 with pleuritis); 2 disseminated

442 (pulmonary/abdominal/lymphnode and liver/lymphnode) and 2 extrapulmonary (lymphnode and soft

443 tissue/osteitis), CRP = C-reactive protein, ESR = erythrocyte sedimentation rate, Hb = hemoglobin, Alb =

444 albumin. Smear.positive; sputum smear microscopy positive (5 of 14 with sputum); Cough: as reported by

445 patient; Presence of systemic symptoms (fever, weight loss, night sweats); No.S.Sympt = number of systemic

446 symptoms; night sweats, fever, and weight loss.

447

448

449

450

451

452

453

454

455

456

457

458

459

460

461

462

463

464

465

466

467

468

469

470

471

472

473

474

475

476

477

478

479

480

Discussion

The purpose of this study was to identify a protein signature in plasma that was enriched in

individuals with active TB compared to those with latent infection and healthy controls. We

used the Olink PEA method to simultaneously measure 92 proteins in plasma. Using co-

expression analysis, we could identify 16 proteins that were co-expressed and highly

correlated with active TB. To test the validity of these findings, we used publicly available

transcriptomic datasets. First, we removed two proteins with very low differential expression

between active TB and controls. We then refined the signature by excluding two additional

proteins that were highly expressed in other lower respiratory tract infections. This allowed

us to generate a 12-protein signature that was highly specific when applied to independent TB

datasets. When retesting the signature on our own data it also retained a similar

discriminatory capacity as the initial 16-protein signature.

Albeit small, the study cohort is well defined, with microbiological verification of all active

TB cases, a wide range of disease severity and with large variation in patient origin,

reflecting the TB population in Sweden and thus increasing the likelihood of generalizability

to other geographic regions. Further, results were validated in nine transcriptomic data sets

including over 3000 individuals from four continents.

On an individual patient level, the signature was associated with disease activity with

a stronger protein signature significantly related to perturbations in common clinical

inflammatory markers (ESR and albumin) and reported symptoms.

469 We have shown that the identified signature is differentially expressed in independent TB
470 transcriptomic datasets and importantly, it is not expressed in patients with other lower
471 respiratory tract infections, which is essential if the signature is to be used as a diagnostic
472 screening test in a clinical setting as intended. However, when testing the signature in a
473 sarcoidosis dataset (27), we observed a significant enrichment of the proteins. Sarcoidosis
474 and TB have previously been shown to have similar gene expression patterns (27). Both
475 diseases demonstrate an interferon-driven gene up-regulation, although as shown by Blankley
476 et al, on a group level this pattern is more strongly expressed in active TB, reflecting disease
477 activity (27). We were not able to directly compare sarcoidosis versus active TB to assess if
478 the signal is higher in one group or the other. However, if the signature was to be used as a
479 screening test this potential overlap will likely not pose a significant diagnostic problem in a
480 clinical setting as symptoms and other clinical information such as radiology will help
481 separate the two conditions. Further, in TB high-endemic countries where a screening test is
482 most urgently needed, the prevalence of sarcoidosis is very low compared to pulmonary TB
483 (41, 42).

484

485 Lack of overlap between protein signatures for active TB has been previously described, and
486 the methods used to quantify proteins in different studies also vary (14). When comparing the
487 12-marker signature to recently published protein signatures (29-33, 40, 43-51) some
488 biomarkers reappear, although the overlap is limited. Additionally, in the published
489 proteomic studies there were none with processed data that included all the proteins of our
490 signature in their dataset, making us unable to use them to validate the signature proposed in
491 this study.

492 VEGF and IL-6, present in the 12-protein signature, together with IL-8 and IL-18,
493 constitute a 4-protein signature identified by Ahmad et al. (29). They analyzed 47 proteins
494 with Luminex and the 4 selected proteins were validated in serum from three different patient
495 cohorts collected through the FIND initiative. The sensitivity for active TB in TB suspects
496 was 80% (95% CI, 73 to 85%), and the specificity was 65% (95% CI, 57-71%). Interestingly,
497 there was quite a large overlap between the Olink inflammation panel we used and their 47-
498 protein Luminex panel. Although IL-8 and IL-18 were more abundant in active TB patients
499 compared with healthy controls in our dataset (Figure S5), the fold-change values were low
500 compared with the proteins included in the 12-protein signature and were not significantly
501 higher when compared with latent TB individuals.

502 VEGF was also included in a 4-protein signature (CCL1, CXCL10, ADA-2, VEGF)
503 proposed by Delemarre et al. (30), where they compared active TB to treated and untreated
504 latent TB. The signature was validated in two separate patient cohorts with a sensitivity of
505 95% and a specificity of 90%. CXCL10 was also highly expressed in our analysis but was
506 removed due to its high expression in other lower respiratory tract infections. CCL1 and
507 ADA-2 were not analyzed in the current study.

508 Although VEGF was not included in the Chegou et al. (31) 7-protein signature, it was
509 increased in active TB patients in their cohort. Of their 7 proteins, IFN-gamma overlaps with
510 our 12-protein signature. CXCL10 (IP-10) again appears in their signature while the other
511 proteins (CRP, TTR, CFH, APO-A1, SSA) were not part of the proteins evaluated in our
512 study.

513 In 2021, Morris et al. (32) and Mutavhatsindi et al. (51) investigated the same 22 proteins as
514 Chegou et al. (31) and attempted to validate the 7-protein signature in patients with suspected
515 TB. In the first study, the sensitivity was very high (98%) while the specificity was low
516 (12%). They argued that this might be due to the different patient cohorts – primary care level
517 versus hospital care level. Instead, they identified a 9-protein signature (fibrinogen, alpha-2-
518 macroglobulin, CRP, MMP-9, transthyretin, complement factor H [CFH], IFN-gamma, IP-

10, and TNF-alpha) where IFN-gamma and IP-10 reappear. In the test set, the sensitivity was 92% (95% CI, 80-98%) and specificity 71% (95% CI, 56-84%) for diagnosing culture verified TB from other diseases in TB suspects. Their second study (51) failed to validate the 7 marker signature. Instead, they proposed CRP and CCL1 as a signature that performed equally well in both HIV- and HIV+ individuals. Trying to design a protein signature based on previous findings, Pedersen et al. (33) evaluated 9 proteins in pulmonary TB patients compared to healthy controls. They found that IL-6, VEGFA, (and IP-10) are significantly increased in active TB. Their proposed signature consists of IP-10 and 4 miRNA molecules, including miR-29a, miR-146a, miR-99b and miR-221. IP-10 is also part of the 5-protein signature identified by Luo et al. (52).

Garay-Baquero et al., analyzed more than 5000 peptides using Mass spectrometry on a relatively small discovery cohort (10 individuals with active TB and 10 healthy controls) (44). They identified 46 proteins to be overexpressed in active TB and selected 9 and 7 proteins for validation in larger populations in South Africa and the UK, respectively. They used Luminex or ELISA and compared active TB to other respiratory diseases (ORD) and healthy controls. The proposed 5 protein signature; CFHR5, LRG1, CRP, LBP, and SAA1, performed well with an AUC > 0.8 in both settings for active TB vs ORD. There was no overlap with the 12-protein signature identified here, and none of the proteins were among the 136 proteins identified as associated with active TB (46 more abundant and 90 less abundant) in their discovery phase, although the use of different methods makes direct comparison difficult.

Another large proteomic study by de Groote et al., (53) used SOMAscan to measure over 4000 proteins in 1470 patient samples from pulmonary TB patients and other TB suspect cases and resulted in a 6-protein signature. Although their signature did not overlap with our 12-protein signature, IL-6 and MMP-1 were overexpressed in their active TB cases.

In summary, IFN-gamma, IL-6, and VEGF, together with IP-10 were identified as markers for active TB in several previous studies, as well as identified in our study. These are all regulated by STAT-1, which has previously been identified as an important immune response pathway in Mtb infection (38), and potentially explains why these proteins are identified in different studies. However, as found here, IP-10 is also highly expressed in other severe infectious diseases and as such is likely not specific for TB in the absence of Mtb-specific stimulation (54).

To conclude, we identified a plasma biomarker signature associated with active TB progression that was further corroborated in several independent datasets. Although the signature can likely be optimized further by testing it in independent protein datasets, the included biomarkers warrant further investigation and development for diagnostic purposes, which will be critical to stopping the TB pandemic.

557

558 **Acknowledgements**

559 We would foremost like to thank the study participants for agreeing to be included in the
560 study and donate study material. We would also like to thank the study nurses Monica
561 Modin, Jan Bellbrant, Anna Dahlberg and Lena Jansson for patient inclusion, Julius
562 Lautenbach and Victor Yman for constructive discussions about the analysis, and the
563 SciLifeLab Translational Plasma Profiling facility for their support and generating data for
564 this study. This study was supported by grants from the Swedish Research Council (2021-
565 03706), Magnus Bergvall Foundation (2019-03436), Åke Wiberg Foundation (M19-0559)
566 and the Swedish Medical Association (SLS-934363) to CS and grants from Swedish
567 Research Council (2019-04663 and 2020-03602) and the Heart- and Lung Foundation
568 (20180386 and 20200194) to GK.

569

570 **REFERENCES**

- 571 1. Pai M, Nicol MP, Boehme CC. Tuberculosis Diagnostics: State of the Art and Future
572 Directions. *Microbiol Spectr*. 2016;4(5).
- 573 2. Dorman S. Advances in the diagnosis of tuberculosis: current status and future
574 prospects. *The international journal of tuberculosis and lung disease : the official journal of*
575 *the International Union against Tuberculosis and Lung Disease*. 2015;19(5):504-16.
- 576 3. Cohen A, Mathiasen VD, Schon T, Wejse C. The global prevalence of latent
577 tuberculosis: a systematic review and meta-analysis. *Eur Respir J*. 2019;54(3).
- 578 4. Steingart KR, Henry M, Ng V, Hopewell PC, Ramsay A, Cunningham J, et al.
579 Fluorescence versus conventional sputum smear microscopy for tuberculosis: a systematic
580 review. *Lancet Infect Dis*. 2006;6(9):570-81.
- 581 5. Horne DJ, Kohli M, Zifodya JS, Schiller I, Dendukuri N, Tollefson D, et al. Xpert
582 MTB/RIF and Xpert MTB/RIF Ultra for pulmonary tuberculosis and rifampicin resistance in
583 adults. *Cochrane Database Syst Rev*. 2019;6:CD009593.
- 584 6. Zifodya JS, Kreniske JS, Schiller I, Kohli M, Dendukuri N, Schumacher SG, et al.
585 Xpert Ultra versus Xpert MTB/RIF for pulmonary tuberculosis and rifampicin resistance in
586 adults with presumptive pulmonary tuberculosis. *Cochrane Database Syst Rev*.
587 2021;2:CD009593.
- 588 7. Getnet F, Demissie M, Assefa N, Mengistie B, Worku A. Delay in diagnosis of
589 pulmonary tuberculosis in low-and middle-income settings: systematic review and meta-
590 analysis. *BMC Pulm Med*. 2017;17(1):202.
- 591 8. WHO. Global Tuberculosis
592 Report. <https://www.who.int/publications/i/item/9789240037021>. 2021.
- 593 9. WHO. Consensus meeting report: development of a target product profile (TPP) and a
594 framework for evaluation for a test for predicting progression from tuberculosis infection to
595 active disease. World Health Organization. License: CC BY-NC-SA 3.0 IGO. 2017.
- 596 10. Haas CT, Roe JK, Pollara G, Mehta M, Noursadeghi M. Diagnostic 'omics' for active
597 tuberculosis. *BMC Med*. 2016;14:37.
- 598 11. Weiner J, Kaufmann SH. High-throughput and computational approaches for
599 diagnostic and prognostic host tuberculosis biomarkers. *Int J Infect Dis*. 2017;56:258-62.
- 600 12. Sweeney TE, Braviak L, Tato CM, Khatri P. Genome-wide expression for diagnosis
601 of pulmonary tuberculosis: a multicohort analysis. *Lancet Respir Med*. 2016;4(3):213-24.
- 602 13. Blankley S, Graham CM, Levin J, Turner J, Berry MP, Bloom CI, et al. A 380-gene
603 meta-signature of active tuberculosis compared with healthy controls. *Eur Respir J*.
604 2016;47(6):1873-6.
- 605 14. MacLean E, Broger T, Yerlikaya S, Fernandez-Carballo BL, Pai M, Denkinger CM.
606 A systematic review of biomarkers to detect active tuberculosis. *Nat Microbiol*.
607 2019;4(5):748-58.
- 608 15. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers
609 differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids*
610 *Res*. 2015;43(7):e47.
- 611 16. Hanzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for
612 microarray and RNA-seq data. *BMC Bioinformatics*. 2013;14:7.
- 613 17. Lee E, Chuang HY, Kim JW, Ideker T, Lee D. Inferring pathway activity toward
614 precise disease classification. *PLoS Comput Biol*. 2008;4(11):e1000217.
- 615 18. Jonsson J, Westman A, Bruchfeld J, Sturegard E, Gaines H, Schon T. A borderline
616 range for Quantiferon Gold In-Tube results. *PLoS one*. 2017;12(11):e0187313.
- 617 19. Wikell A, Jonsson J, Dyrdak R, Henningsson AJ, Eringfalt A, Kjerstadius T, et al.
618 The Impact of Borderline Quantiferon-TB Gold Plus Results for Latent Tuberculosis

- 619 Screening under Routine Conditions in a Low-Endemicity Setting. *Journal of clinical*
620 *microbiology*. 2021;59(12):e0137021.
- 621 20. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a
622 software environment for integrated models of biomolecular interaction networks. *Genome*
623 *Res*. 2003;13(11):2498-504.
- 624 21. Wang XJ, W.E.
625 Patil, P. . curatedTBData: Curation of existing 49 tuberculosis transcriptomic studies. R
626 package version 1.0.0. <https://github.com/compbio/curatedTBData>. 2021.
- 627 22. Herberg JA, Kaforou M, Gormley S, Sumner ER, Patel S, Jones KD, et al.
628 Transcriptomic profiling in childhood H1N1/09 influenza reveals reduced expression of
629 protein synthesis genes. *The Journal of infectious diseases*. 2013;208(10):1664-8.
- 630 23. Parnell GP, McLean AS, Booth DR, Armstrong NJ, Nalos M, Huang SJ, et al. A
631 distinct influenza infection signature in the blood transcriptome of patients with severe
632 community-acquired pneumonia. *Crit Care*. 2012;16(4):R157.
- 633 24. Suarez NM, Bunsow E, Falsey AR, Walsh EE, Mejias A, Ramilo O. Superiority of
634 transcriptional profiling over procalcitonin for distinguishing bacterial from viral lower
635 respiratory tract infections in hospitalized adults. *The Journal of infectious diseases*.
636 2015;212(2):213-22.
- 637 25. Chakraborty RK, Burns B. Systemic Inflammatory Response Syndrome. *StatPearls*.
638 *Treasure Island (FL)2022*.
- 639 26. Bloom CI, Graham CM, Berry MP, Rozakeas F, Redford PS, Wang Y, et al.
640 Transcriptional blood signatures distinguish pulmonary tuberculosis, pulmonary sarcoidosis,
641 pneumonias and lung cancers. *PloS one*. 2013;8(8):e70630.
- 642 27. Blankley S, Graham CM, Turner J, Berry MP, Bloom CI, Xu Z, et al. The
643 Transcriptional Signature of Active Tuberculosis Reflects Symptom Status in Extra-
644 Pulmonary and Pulmonary Tuberculosis. *PloS one*. 2016;11(10):e0162220.
- 645 28. Johnson WE, Odom A, Cintron C, Muthaiah M, Knudsen S, Joseph N, et al.
646 Comparing tuberculosis gene signatures in malnourished individuals using the
647 TBSignatureProfiler. *BMC Infect Dis*. 2021;21(1):106.
- 648 29. Ahmad R, Xie L, Pyle M, Suarez MF, Broger T, Steinberg D, et al. A rapid triage test
649 for active pulmonary tuberculosis in adult patients with persistent cough. *Sci Transl Med*.
650 2019;11(515).
- 651 30. Delemarre EM, van Hoorn L, Bossink AWJ, Drylewicz J, Joosten SA, Ottenhoff
652 THM, et al. Serum Biomarker Profile Including CCL1, CXCL10, VEGF, and Adenosine
653 Deaminase Activity Distinguishes Active From Remotely Acquired Latent Tuberculosis.
654 *Front Immunol*. 2021;12:725447.
- 655 31. Chegou NN, Sutherland JS, Malherbe S, Crampin AC, Corstjens PL, Geluk A, et al.
656 Diagnostic performance of a seven-marker serum protein biosignature for the diagnosis of
657 active TB disease in African primary healthcare clinic attendees with signs and symptoms
658 suggestive of TB. *Thorax*. 2016;71(9):785-94.
- 659 32. Morris TC, Hoggart CJ, Chegou NN, Kidd M, Oni T, Goliath R, et al. Evaluation of
660 Host Serum Protein Biomarkers of Tuberculosis in sub-Saharan Africa. *Front Immunol*.
661 2021;12:639174.
- 662 33. Pedersen JL, Barry SE, Bokil NJ, Ellis M, Yang Y, Guan G, et al. High sensitivity
663 and specificity of a 5-analyte protein and microRNA biosignature for identification of active
664 tuberculosis. *Clin Transl Immunology*. 2021;10(6):e1298.
- 665 34. Datta M, Via LE, Kamoun WS, Liu C, Chen W, Seano G, et al. Anti-vascular
666 endothelial growth factor treatment normalizes tuberculosis granuloma vasculature and
667 improves small molecule delivery. *Proc Natl Acad Sci U S A*. 2015;112(6):1827-32.

- 668 35. Kumar NP, Moideen K, Nancy A, Viswanathan V, Shruthi BS, Sivakumar S, et al.
669 Plasma chemokines are biomarkers of disease severity, higher bacterial burden and delayed
670 sputum culture conversion in pulmonary tuberculosis. *Sci Rep*. 2019;9(1):18217.
- 671 36. Andersen P, Doherty TM, Pai M, Weldingh K. The prognosis of latent tuberculosis:
672 can disease be predicted? *Trends Mol Med*. 2007;13(5):175-82.
- 673 37. Kahnert A, Hopken UE, Stein M, Bandermann S, Lipp M, Kaufmann SH.
674 Mycobacterium tuberculosis triggers formation of lymphoid structure in murine lungs. *The*
675 *Journal of infectious diseases*. 2007;195(1):46-54.
- 676 38. Yi XH, Zhang B, Fu YR, Yi ZJ. STAT1 and its related molecules as potential
677 biomarkers in Mycobacterium tuberculosis infection. *J Cell Mol Med*. 2020;24(5):2866-78.
- 678 39. Mousavian Z, Khodabandeh M, Sharifi-Zarchi A, Nadafian A, Mahmoudi A.
679 StrongestPath: a Cytoscape application for protein-protein interaction analysis. *BMC*
680 *Bioinformatics*. 2021;22(1):352.
- 681 40. Schulman H, Niward K, Abate E, Idh J, Axenram P, Bornefall A, et al. Sedimentation
682 rate and suPAR in relation to disease activity and mortality in patients with tuberculosis. *The*
683 *international journal of tuberculosis and lung disease : the official journal of the International*
684 *Union against Tuberculosis and Lung Disease*. 2019;23(11):1155-61.
- 685 41. Collaborators GBDCRD. Prevalence and attributable health burden of chronic
686 respiratory diseases, 1990-2017: a systematic analysis for the Global Burden of Disease
687 Study 2017. *Lancet Respir Med*. 2020;8(6):585-96.
- 688 42. Labaki WW, Han MK. Chronic respiratory diseases: a global view. *Lancet Respir*
689 *Med*. 2020;8(6):531-3.
- 690 43. Agranoff D, Fernandez-Reyes D, Papadopoulos MC, Rojas SA, Herbster M,
691 Loosemore A, et al. Identification of diagnostic markers for tuberculosis by proteomic
692 fingerprinting of serum. *Lancet*. 2006;368(9540):1012-21.
- 693 44. Garay-Baquero DJ, White CH, Walker NF, Tebruegge M, Schiff HF, Ugarte-Gil C, et
694 al. Comprehensive plasma proteomic profiling reveals biomarkers for active tuberculosis. *JCI*
695 *Insight*. 2020;5(18).
- 696 45. De Groote MA, Higgins M, Hraha T, Wall K, Wilson ML, Sterling DG, et al. Highly
697 Multiplexed Proteomic Analysis of Quantiferon Supernatants To Identify Biomarkers of
698 Latent Tuberculosis Infection. *Journal of clinical microbiology*. 2017;55(2):391-402.
- 699 46. Jiang TT, Shi LY, Chen J, Wei LL, Li M, Hu YT, et al. Screening and identification
700 of potential protein biomarkers for evaluating the efficacy of intensive therapy in pulmonary
701 tuberculosis. *Biochem Biophys Res Commun*. 2018;503(4):2263-70.
- 702 47. Liu Q, Pan L, Han F, Luo B, Jia H, Xing A, et al. Proteomic profiling for plasma
703 biomarkers of tuberculosis progression. *Mol Med Rep*. 2018;18(2):1551-9.
- 704 48. Mateos J, Estevez O, Gonzalez-Fernandez A, Anibarro L, Pallares A, Reljic R, et al.
705 Serum proteomics of active tuberculosis patients and contacts reveals unique processes
706 activated during Mycobacterium tuberculosis infection. *Sci Rep*. 2020;10(1):3844.
- 707 49. Sun H, Pan L, Jia H, Zhang Z, Gao M, Huang M, et al. Label-Free Quantitative
708 Proteomics Identifies Novel Plasma Biomarkers for Distinguishing Pulmonary Tuberculosis
709 and Latent Infection. *Front Microbiol*. 2018;9:1267.
- 710 50. Xu D, Li Y, Li X, Wei LL, Pan Z, Jiang TT, et al. Serum protein S100A9, SOD3, and
711 MMP9 as new diagnostic biomarkers for pulmonary tuberculosis by iTRAQ-coupled two-
712 dimensional LC-MS/MS. *Proteomics*. 2015;15(1):58-67.
- 713 51. Mutavhatsindi H, van der Spuy GD, Malherbe ST, Sutherland JS, Geluk A, Mayanja-
714 Kizza H, et al. Validation and Optimization of Host Immunological Bio-Signatures for a
715 Point-of-Care Test for TB Disease. *Front Immunol*. 2021;12:607827.
- 716 52. Luo J, Zhang M, Yan B, Li F, Guan S, Chang K, et al. Diagnostic performance of
717 plasma cytokine biosignature combination and MCP-1 as individual biomarkers for

718 differentiating stages Mycobacterium tuberculosis infection. The Journal of infection.
719 2019;78(4):281-91.
720 53. De Groote MA, Sterling DG, Hraha T, Russell TM, Green LS, Wall K, et al.
721 Discovery and Validation of a Six-Marker Serum Protein Signature for the Diagnosis of
722 Active Pulmonary Tuberculosis. Journal of clinical microbiology. 2017;55(10):3057-71.
723 54. Ruhwald M, Dominguez J, Latorre I, Losi M, Richeldi L, Pasticci MB, et al. A
724 multicentre evaluation of the accuracy and performance of IP-10 for the diagnosis of
725 infection with M. tuberculosis. Tuberculosis (Edinb). 2011;91(3):260-7.
726
727