



Plasma host protein biomarkers correlating with increasing *Mycobacterium tuberculosis* infection activity prior to tuberculosis diagnosis in people living with HIV

Sarah N. Singer,^{a,1} Okechukwu C. Ndumego,^{b,1,2} Ryung S. Kim,^c Thumbi Ndung'u,^{b,d,e,f,g} Kathryn Anastos,^{a,c} Audrey French,^h Gavin Churchyard,^{i,j,k} Eustache Paramithiothis,^l Victoria O. Kasprowicz,^{b,d,e} and Jacqueline M. Achkar^{a,m,*}

^aDepartments of Medicine, Albert Einstein College of Medicine, Bronx, NY 10461, USA

^bAfrica Health Research Institute, Durban 4013, South Africa

^cEpidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY 10461, USA

^dHIV Pathogenesis Programme, Nelson R. Mandela School of Medicine, University of KwaZulu-Natal, Durban, South Africa

^eRagon Institute of MGH, MIT and Harvard University, Cambridge, MA, USA

^fMax Planck Institute of Infection Biology, Berlin, Germany

^gDivision of Infection and Immunity, University College London, London, UK

^hDepartment of Medicine, Stroger Hospital of Cook County, Chicago, IL, USA

ⁱAurum Institute, Johannesburg, South Africa

^jSchool of Public Health, University of Witwatersrand, Johannesburg, South Africa

^kDepartment of Medicine, Vanderbilt University, Nashville, TN, USA

^lCellCarta Biosciences Inc, 201 President-Kennedy Ave., Suite 3900 Montreal, H2×3Y7, Quebec, Canada

^mMicrobiology and Immunology, Albert Einstein College of Medicine, Bronx, NY 10461, USA

Summary

Background Biomarkers correlating with *Mycobacterium tuberculosis* infection activity/burden in asymptomatic individuals are urgently needed to identify and treat those at highest risk for developing active tuberculosis (TB). Our main objective was to identify plasma host protein biomarkers that change over time prior to developing TB in people living with HIV (PLHIV).

Methods Using multiplex MRM-MS, we investigated host protein expressions from 2 years before until time of TB diagnosis in longitudinally collected (every 3-6 months) and stored plasma from PLHIV with incident TB, identified within a South African (SA) and US cohort. We performed temporal trend and discriminant analyses for proteins, and, to assure clinical relevance, we further compared protein levels at TB diagnosis to interferon-gamma release assay (IGRA; SA) or tuberculin-skin test (TST; US) positive and negative cohort subjects without TB. SA and US exploratory data were analyzed separately.

Findings We identified 15 proteins in the SA (n=30) and 10 in the US (n=24) incident TB subjects which both changed from 2 years prior until time of TB diagnosis after controlling for 10% false discovery rate, and were significantly different at time of TB diagnosis compared to non-TB subjects (p<0.01). Five proteins, CD14, A2GL, NID1, SCTM1, and A1AG1, overlapped between both cohorts. Furthermore, after cross-validation, panels of 5 – 12 proteins were able to predict TB up to two years before diagnosis.

Interpretation Host proteins can be biomarkers for increasing *Mycobacterium tuberculosis* infection activity/burden, incident TB, and predict TB development in PLHIV.

Funding NIH/NIAID AI17927, AI146329, and AI127173 to JMA.

*Corresponding author at: Departments of Medicine (Div. of Infectious Diseases), and Microbiology and Immunology, Albert Einstein College of Medicine, 1300 Morris Park Ave, Block Bld., Rm. 115, Bronx, New York 10461, USA.

E-mail address: jacqueline.achkar@einsteinmed.edu (J.M. Achkar).

¹ Shared first authorship.

² Current address: Moredun Scientific, Pentlands Science Park, Edinburgh EH26 0PZ, UK.

Copyright © 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Keywords: Tuberculosis; Incipient TB; Subclinical TB; Biomarker; Diagnosis; HIV; Proteomics; Prediction

Research in context

Evidence before this study

Prior to the detection of *M. tuberculosis* (*Mtb*) in body fluids (e.g. sputum) or tissue at the time of diagnosing active tuberculosis (TB), *Mtb* infection activity or burden cannot be determined *in vivo* without biopsy and histology or complex radiographic imaging tools such as Positron Emission Tomography – Computed Tomography (PET-CT). Therefore, it is not possible to determine who among asymptomatic individuals presumed to be infected with *Mtb* will remain asymptomatic/latently infected and who will, or is, progressing to TB. A few well-designed longitudinal studies have recently identified transcriptomic, metabolic, and proteomic signatures predicting progression to TB in HIV uninfected individuals. Albeit with cohort-specific variability in performance, TB predictive transcriptomic signatures have been validated by different groups up to 3 - 24 months before diagnosis, with best performance in predicting short-term risk. However, transcriptomics require expertise and tools for RNA-sequencing. Furthermore, studies have not investigated host protein markers for increasing *Mtb* infection activity/burden in people living with HIV (PLHIV) who are at particularly high risk for developing TB and would benefit from regular screening of easily detectable markers integrated into routine HIV care.

Added value of this study

We identified human host proteins correlating with increasing *Mtb* burden from up to two years prior until time of TB diagnosis in PLHIV. A set of 5 candidate proteins overlapped between South African and US HIV-infected cohort subjects with incident TB, indicating that these markers are upregulated irrespective of TB endemic region and level of CD4 counts. Moreover, for the South African cohort, a panel of 5 - 6 protein markers met the WHO target for a predictive TB test up to one year prior to TB diagnosis. These plasma host proteins can, in contrast to gene signatures, be detected with simple and rapid tests, and can be markers for increasing *Mtb* burden, detection of incipient TB, and prediction of TB development prior to clinical signs and symptoms in PLHIV.

Implications of all available evidence

Complementary to transcriptomic, metabolic, and proteomic signatures predicting progression to TB in HIV uninfected individuals, our data provide important insights into the host responses associated with increased *Mtb* infection activity/burden in PLHIV. We show that proteins can serve as easily detectable plasma biomarkers for increasing *Mtb* burden prior to TB diagnosis in PLHIV. If further validated, our results

can inform the development of simple blood tests that could be integrated into routine outpatient follow-up visits of PLHIV and trigger screening and monitoring for TB when proteins are elevated. Used as a single platform, or, when combined with other biomarkers, as a multi-platform tool, such tests can aid in both optimizing the timing and increasing the effectiveness of TB preventive or therapeutic therapy in PLHIV.

Introduction

Active tuberculosis (TB) is a major global public health problem and the leading cause of death among people living with HIV (PLHIV).¹ To fill gaps in the current TB diagnostic armamentarium, accurate and easily detectable non-sputum-based biomarkers correlating with increasing *Mycobacterium tuberculosis* (*Mtb*) burden in asymptomatic individuals with remote or recent *Mtb* infection are urgently needed.^{2,3} Such biomarkers are critical for: i) predicting the risk of TB development; ii) detecting the early onset of disease; and iii) monitoring antituberculous treatment response.³ Furthermore, the need for such biomarkers is particularly high in the setting of HIV co-infection, a major risk factor for TB development.³⁻⁶ Although asymptomatic latent *Mtb* infection (LTBI) and TB are often seen as binary states, disease due to reactivation of endogenous remote infection is preceded by a continuum of increasing *Mtb* replication progressing to incipient, subclinical and clinical TB disease (reviewed in⁷⁻¹¹). Depending on the level of immunosuppression, this process can last months to years before developing symptomatic and clinically diagnosable TB.⁷⁻¹⁰ Similarly, TB due to exogenous new or reinfection is preceded by increasing infection activity, albeit at an often faster rate than that of more remote infection.²

Prior to the detection of *Mtb* in body fluids (e.g. sputum) at the time of diagnosing active tuberculosis (TB), *Mtb* infection activity cannot be determined *in vivo*, to date, without complex radiographic imaging tools such as Positron Emission Tomography – Computed Tomography (PET-CT).¹¹⁻¹³ Because HIV infection increases the risk for TB development 30 - 60 fold, the World Health Organization (WHO) recommends TB preventative drug therapy for all PLHIV living in TB endemic regions.^{6,14} However, the effect of preventative therapy is often not durable, and there are associated side-effects.¹⁵⁻¹⁸ More so, TB preventative therapy may be inadequate in treating PLHIV with undiagnosed

incipient or subclinical TB.^{12,19} Because only a portion of those infected with *Mtb* develop TB, easily detectable non-sputum-based biomarkers correlating with *Mtb* burden in asymptomatic individuals could serve as screening or triage tools to identify both PLHIV at risk for progressing to TB and those with early disease.

The only currently available simple (for use by first contact providers) point-of-care (POC) test for the rapid diagnosis of TB is based on the detection of the *Mtb* cell wall glycolipid lipoarabinomannan (LAM) in urine. Because this lateral flow test (AlereLAM) has an overall test sensitivity of below 50% for TB, which changes to around 56% in patients with advanced HIV,²⁰ the World Health Organization (WHO) has endorsed its use only for HIV-infected patients with CD4 counts < 100 cells/ μ l.²¹ A recently developed new generation urine lateral flow POC test (FujiLAM) has shown enhanced sensitivity in PLHIV.²² Nevertheless, the detection of LAM in urine is dependent on a sufficiently high *Mtb* burden which typically coincides with disease manifestations. A handful of well-designed longitudinal studies have recently investigated human host biomarker signatures predicting TB but have focused predominantly on HIV-uninfected adolescents and household contacts. These show mostly transcriptomic, but also metabolic and proteomic signatures predicting TB between 3 - 24 months before diagnosis, albeit with cohort-specific performance variability and higher accuracy in the months closer to TB diagnosis.²³⁻²⁷ One study in PLHIV reported a transcriptomic signature that was able to differentiate between recurrent TB progressors and non-progressors 3 months prior to TB diagnosis,²⁸ and another found transcripts of complement pathways and Fc γ receptors upregulated in PLHIV with subclinical TB.²⁹ However, studies have not investigated host protein markers of increasing *Mtb* burden years prior to TB diagnosis in PLHIV, who i) can differ considerably from HIV uninfected individuals in their TB presentation and host responses (reviewed in^{7,30}); ii) are at particularly high risk for developing TB; and iii) would benefit from easily detectable markers that could be integrated into routine HIV follow-up visits.

In our prior cross-sectional studies, using liquid chromatography and mass spectrometry (MS), followed by multiplex multiple reaction monitoring MS (MRM-MS), we identified host proteins that were significantly differentially expressed in the sera of US HIV uninfected and co-infected patients with TB compared to those who were *Mtb* uninfected, had LTBI, or other respiratory diseases.³¹ In a subsequent study with Ugandan TB household contacts, we further found several overlapping host proteins to be associated with tuberculin skin-test (TST) conversion, indicative of initial *Mtb* infection.³² In the current study, using a panel of 163 TB host protein biomarkers identified in our prior studies,^{31,32} we investigated longitudinally collected

(every 3-6 months) and stored plasma samples from prospectively followed HIV-infected subjects from two distinct cohorts followed in South Africa (SA) and the US. Our main objectives were to i) identify plasma host protein biomarkers that increase (or decrease) over time in PLHIV prior to TB diagnosis and, to assure clinical relevance, are also significantly different at time of TB diagnosis compared to non-TB subjects; and ii) enhance our understanding of host proteins involved in pathogenesis during the loss of *Mtb* control. Our secondary objective was to identify panels of biomarkers that can predict TB development up to two years prior to diagnosis. Our secondary differs from our primary objective in not requiring a significant change of biomarkers over time. We hypothesized that plasma host proteins can serve as biomarkers for increasing *Mtb* burden, incident TB, and predict TB development in PLHIV.

Materials and methods

Subjects and study design

To address our specific objectives in the absence of existing reference standards for *Mtb* infection activity/burden prior to TB diagnosis, we tested longitudinally collected and stored plasma samples from adult SA and US HIV-infected cohort subjects up to two and a half years (30 months) pre and one and a half years (18 months) post TB diagnosis. Inclusion criterion was a diagnosis of TB during follow-up time (as defined below) and exclusion criterion was a lack of a sample within 6 months prior to TB diagnosis. Sample size was based on all subjects identified with incident TB in each cohort. Our longitudinal analysis focused on host protein expression changes from two years before until time of TB diagnosis, a time-frame for which we had samples from all incident TB subjects available. When available, time-points up to 30 months pre and 18 months post diagnosis were included in the testing for better visualization of protein expression over time and to be able to examine whether significant protein expression changes reversed in the year post TB diagnosis (tertiary objective). To assure clinical relevance, we tested and compared expressions of host proteins between TB subjects at time of diagnosis and interferon-gamma release assay (IGRA; SA) or TST (US) positive and negative cohort subjects who had not developed TB. Inclusion criterion for non-TB subjects was an available IGRA (SA) or TST (US) result, and exclusion criterion was less than 3 year follow-up duration. For each subject with incident TB in each cohort, we sampled two non-TB subjects from the same cohort with equal numbers of IGRA (SA) or TST (US) positive and negative subjects. This was done because TB, especially in the setting of HIV co-infection, can be the result of reactivation of a remote or new *Mtb* infection. Non-TB subjects were not matched, but to avoid

overrepresentation of subjects with high CD4 counts among those without TB, we stratified TB subjects in each cohort by CD4 count quartiles at the time of TB diagnosis (< 100, 101–200, 201–300, and > 300 cells/mm³), and randomly sampled, depending on availability, equal or similar numbers of IGRA positive and negative (SA Sinikithemba) and TST positive and negative (US WHIS) non-TB subjects from each CD4 quartile in each cohort. Each cohort was investigated and exploratory data analysed separately.

Cohort descriptions. Subjects from SA were enrolled within the Sinikithemba cohort which was based in Durban and followed over 500 HIV-infected men and women from 2003–2011 to study HIV clade C virus infection.³³ Subjects from the US were enrolled within the Women's Interagency HIV Study (WIHS) cohort, one of the largest and longest US-wide cohorts. This cohort was established in 1993 to investigate the impact of HIV infection on women in the US, is ongoing, and follows several thousand HIV-infected women.^{34–36} All subjects in both cohorts had documented HIV infection and were seen either around every 3 months (Sinikithemba) or every 6 months (WIHS). None of the SA cohort subjects included in this study were on antiretroviral therapy (ART) against HIV, whereas, depending on year of enrollment and follow-up period, US subjects were on ART with at least two drugs against HIV. During follow-up visits, routine laboratory (e.g. CD4 counts and HIV viral load (VL)), physical exams and interviews were performed, and blood drawn, aliquoted, and frozen at -80°C until needed.

Incident TB and Non-TB definitions. TB and non-TB subjects were selected at the end of follow-up. Data sources were medical records (SA) and the WHIS cohort database. Cohort subjects with incident TB were defined as those who, as per medical records (SA) or WHIS database (US) were diagnosed with TB and treated with antituberculous drugs during the follow-up duration of the original cohort study. For the SA TB subjects, TB diagnosis was established through a combination of TB-associated symptoms and physical findings with supporting radiographic imaging, detection of acid fast bacilli in sputum, mycobacterial cultures, and/or body fluid analysis or histology consistent with TB. For the US TB subjects, a cohort subject was entered into the database as having TB when reported by their health care provider or a microbiology lab to the respective US state department of health as being diagnosed with TB. Because subjects in both cohorts were followed for HIV progression and not specifically for TB, the time of TB diagnosis among SA Sinikithemba cohort subjects was considered equivalent to the date a physician recorded the diagnosis of TB in the medical records. Medical records were reviewed by study personnel on a quarterly basis. For the US WIHS subjects, for whom medical records were not available, we used the date a subject was reported to a US state department

of health as having TB. Non-TB subjects were defined as cohort subjects who never developed TB during their follow-up in the original study. Although it cannot be guaranteed that non-TB subjects did not become a case, they had at least one-year follow-up time beyond the plasma time-point analyzed to assure TB did not occur. Non-TB subjects from the SA Sinikithemba cohort were tested by an interferon-gamma release assay (IGRA). These were performed at several follow-up visits in the form of enzyme-linked immunosorbent spot (ELISpot) assays, measuring interferon-gamma secretion from peripheral blood mononuclear cells in response to stimulation with peptide pools from the RD1 proteins ESAT6 and CFP10, and were performed and interpreted as described.³⁷ Only subjects with concordant positive or negative results were included. Non-TB subjects from the US WHIS cohort had Tuberculin skin test (TST) results available. TSTs were only performed once at baseline and, as per guidelines for HIV-infected individuals, were considered positive at a cut-off of 5 mm.³⁸ As per guidelines, IGRA and TST positive subjects were considered *Mtb* infected³⁸; none of the subjects received TB preventative therapy.

Ethics statement. Human blood sample collections for the performed studies were approved by the Biomedical Research Ethics Committee (BREC) of the University of KwaZulu-Natal (E028/99) and the Institutional Review Boards (IRBs) of each WHIS site.^{34–36} All patients studied gave written informed consent. The studies were further approved by the IRB of the Albert Einstein College of Medicine (2006-428).

Sample processing and data acquisition

All plasma samples were aliquoted on the day of blood draw and stored at -80°C until testing. For the incident TB subjects, consecutive longitudinal samples were processed. Due to restricted funds, testing of non-TB subjects was limited to a single time-point around a recent IGRA or TST. The plasma samples from each cohort were processed and analyzed independently.

Blinding. All plasma samples were randomly coded by a computer program and all personnel involved in testing were blinded as to the clinical information associated with each sample. Samples were grouped into blocks containing each of the clinical groups and the order of the samples within each block was randomized. Each processing batch contained a balanced number of disease groups comprising around 21 test plasma and 3 standard human plasma quality control samples inserted at the beginning, middle, and end of each batch.

Sample processing. Samples were processed for LC-MS/MS as described.^{31,32} Briefly, samples were

depleted of abundant proteins using an HSA/IgG column (Agilent Technologies, Mississauga, ON) in tandem with an IgY14 and Supermix (Sigma, Oakville, Ontario) column. The flow through was digested with trypsin, freeze-dried, desalted, distributed into triplicate 96-well plates, and vacuum evaporated. Sample plates were stored at -20°C until injection by LC-MS/MS.

Data acquisition. Acquisition of MRM-MS data was done as described.^{31,32} Briefly, peptide separation was achieved using a BioBasic C18 column (Thermo; 320 µm × 150 mm, 5 µm particle size). The optimal two transitions per peptide were determined using selected reaction monitoring (SRM)-triggered MS/MS on a QTRAP 5500 instrument (AB Sciex). The two most intense fragment ions (b or y fragment ions only) in the MS/MS spectrum and its elution time were determined for each acquired peptide and the collision energy (CE) was optimized for all of the chosen transitions. Sample processing variability was measured using aliquots of pooled plasma that were inserted at regular intervals among the study samples and taken through the entire analysis. For the SA cohort the process quality control sample CV was 16.8% and for the US cohort the process quality control sample CV was 16.1%.

Protein biomarker data acquisition was done using a previously defined MRM-MS assay containing 163 proteins. These proteins have been previously shown to be differentially expressed in either US immigrants with TB compared to LTBI and other respiratory diseases stratified by HIV-infection status,³¹ or in newly *Mtb*-infected HIV negative Ugandan TB household contacts³². Expression analysis of MRM-MS data was performed using R version 3.3.1. An intensity threshold (IT) below which the measure is deemed less reliable was determined empirically and set to 10,000 pre-normalization. A detection rate (DR) was defined as the proportion of samples within a group with a raw intensity value greater or equal to the IT. Transitions with DR below 50% in both incident TB and asymptomatic groups were excluded from expression analysis. Differential intensity (DI) ratios were calculated in pair wise comparisons for each transition as the ratio of the average normalized intensities of each group.

Statistical methods

Analysis of temporal trend for each protein.

Data from incident TB subjects ranging from 24.5 months before to 12.5 months after TB diagnosis and, due to costs, one instead of several time-points for all non-TB subjects were included separately for SA and US cohorts for trend analyses. Because we expected that proteins correlating with *Mtb* infection activity/burden would show changes in temporal patterns of the

intensities, we fit a linear spline model with the change-points at TB diagnosis. To account for within-person correlation for repeated measures, random intercept of subject-specific effect was included.³⁹ We fit the following model for each protein:

$$y_{ij} = \mu_k + z_i\beta_o + z_i\beta_1t_j + z_i\beta_2 t_jI(t_j > 0) + \theta_i + \epsilon_{ij}$$

where y_{ij} was the normalized plasma protein level of i^{th} subject at j^{th} visit which occurred t_j months before or after TB diagnosis, z_i was the binary indicator of TB subjects (1 for TB, 0 for non-TB), μ_k ($k=1,2,3,4$) was the mean intensities among the non-TB subjects with the same CD4 level at baseline as i^{th} subject: the lowest (<100 cells/mm³), second (101-200), third (201-300), and the highest level (> 300) of CD4. θ_i was the subject-specific Gaussian random intercept, and ϵ_{ij} was the Gaussian mean-zero residual with a common variance.

For each protein, we tested two main hypotheses: $H_o: \beta_o=0$ and $H_o: \beta_1=0$, and one secondary hypothesis $H_o: \beta_2 = 0$. The test for the first main hypothesis was the comparison of protein intensities between the non-TB subjects and the TB subjects at TB diagnosis. Because both IGRA/TST positive and negative PLHIV are clinically relevant non-TB subjects, they were combined for comparison of protein levels to TB subjects at time of TB diagnosis. The second main hypothesis examined whether the slope during the 24 months period prior to TB diagnosis among the TB subjects was flat. And, as an additional (tertiary) objective, the third test examined whether the slope after TB diagnosis among the TB subjects was different from the pre-diagnosis slope. We controlled for type 1 error at 0.01 for the between-group (TB vs. non-TB) comparison, and for false discovery rate (FDR) at 10% for the pre-diagnosis trend test using Benjamini & Hochberg's method.⁴⁰ The FDR is the rate that significant features are truly null. For example, an FDR of 10% means that among all features called significant, 10% of these are truly null on average. Because the post-diagnosis test was a tertiary objective, the results were not adjusted for multiple testing.

Prediction of TB. We used linear support vector machines (SVMs)^{41,42} recursively to identify the proteins contributing the most to distinguish TB subjects from non-TB subjects at 0 – 6 months, 6 – 12 months, 12 – 18 months, and 18 – 24 months prior to TB diagnosis. Recursive use of SVM, a machine-learning method for binary outcomes to select features and build a diagnostic or prognostic model, is well established in the field of genetic data, MRI images, and patient-level clinical data.⁴³⁻⁴⁶ It can be used to select features and build the models when there are large number of potential predictors with collinearity. Rather than analyzing individual proteins, this method finds a set of proteins that work together in distinguishing two sample groups.⁴⁷

To avoid overfitting and information leakage, both the variable selection and the model fitting steps were included in leave-one-out cross-validation (supplemental figure 1). First, we randomly selected one test sample and fit SVM with remaining training samples with all host proteins as predictors and calculated the probability that the test sample would be TB. Then, using backward selection approach, we removed the protein with the least absolute linear coefficient at each iteration until no proteins remained in the model. As we removed each protein, we recorded the TB probability of the test sample and which proteins remained in the model. We repeated this until each sample was once selected as a test sample. Thus, for each number of proteins (between 1 and 143/163 for SA and 138/163 for US, the targeted proteins detected in greater than 50% of samples), we calculated the TB probability. Then, using true TB and non-TB status, the area under the curve (AUC) of ROC was calculated. We then chose the number of proteins (i.e. model complexity, say k) that corresponded to the highest AUC and selected the k proteins that were most frequently selected in the models with size k . These analyses were performed separately for the SA and US cohorts. Each protein intensity was scaled by the standard deviation of the non-TB subjects and the incident TB subjects between 30.5 months before and 12.5 months after TB diagnosis. All analyses were performed using statistical software R.⁴⁸

Role of funding source

The funders had no role in study design, data collection, data analyses, interpretation, or writing of the paper.

Results

Protein biomarkers predictive of onset of HIV-associated TB were identified using plasma samples from the two distinct cohorts - the Sinikithemba cohort, representing SA as a region with high TB incidence and the WHIS cohort, representing the US as a region with low

TB incidence. Within the SA cohort, 31 subjects with incident TB were identified with one subject excluded due to lack of a sample within 6 months prior to TB diagnosis (Table 1). Of the 30 SA TB patients, 15 were diagnosed with pulmonary TB, and 15 with extrapulmonary TB or both pulmonary and extrapulmonary TB. Twelve of the 30 TB patients had a history of prior TB. Diagnosis was established through a combination of TB-associated symptoms with supporting radiographic imaging, detection of acid fast bacilli in sputum, mycobacterial cultures, and/or body fluid analysis or histology consistent with TB. Within the US cohort, 24 subjects with incident TB were identified (Table 2). Information on organ involvement was not available but none of the US TB subjects was reported to have had a history of prior TB.

For the SA cohort, 382 samples and, for the US cohort, 147 samples were analyzed. For the SA samples, 143/163 (90%) and, for US samples, 138/163 (85%) of the targeted proteins were detected in greater than 50% of samples, the threshold for their inclusion in statistical analyses. Both cohorts displayed significant differential expression changes in a subset of proteins leading up to TB diagnosis with many of these reversing significantly in the months post TB diagnosis, presumably due to treatment response. We considered proteins to be markers of increasing *Mtb* infection activity/burden if they met significance at 10% FDR in the pre-diagnosis trend test among the incident TB subjects during the 24 months period prior to TB diagnosis. To assure this was clinically meaningful, the protein level at time of TB diagnosis needed to be also significantly different from the non-TB subjects at a more stringent $p < 0.01$. Because both IGRA/TST positive and negative PLHIV are clinically relevant non-TB subjects, they were combined for comparison of protein levels to TB subjects at time of TB diagnosis. The magnitude of temporal changes in plasma protein levels (slope) for both SA and US cohort over the 24 months prior to TB diagnosis is depicted in Figure 1. Detailed values for all proteins, including those for changes up to 18 months post TB

Characteristics	SA TB+ (n = 30)	SA Non-TB (n = 62)	P value
Female (%)	22 (73)	47 (76)	0.80 ¹
Age, mean years (±SD)	37 (12)	35 (9)	0.30 ²
IGRA+ (%)	NA	31 (50)	NA
ART (%)	0	0	NA
CD4 cells/mm ³ , median (IQR)	306 (239-375)	293 (231-439)	0.98 ³
VL log median, copies/ml (IQR)	5.03 (4.32-5.35)	4.29 (3.42-4.79)	0.002 ³

Table 1: South African incident TB cases and controls.

Subjects were all HIV+ and enrolled within the Sinikithemba cohort. CD4 cell counts listed for TB patients are those closest to time of TB diagnosis.

IGRA: Interferon-gamma release assay; ART: antiretroviral therapy (with at least two drugs against HIV); IQR: Interquartile range; VL: Viral load.

¹ Chi-square test

² t test

³ Mann-Whitney U test.

Characteristics	US TB+ (n = 24)	US Non-TB (n = 48)	P value
Age, mean years (\pm SD)	39 (7)	38 (6)	0.71 ¹
TST+ (%)	NA	24 (50)	NA
ART (%)	10 (42)	25 (52)	0.40 ²
CD4 cells/mm ³ , median (IQR)	196 (95-338)	213 (123-433)	0.21 ³
VL log median, copies/ml (IQR)	4.49 (3.74-5.51)	3.79 (2.77-4.93)	0.004 ³

Table 2: US incident TB cases and controls.

Subjects were all HIV+ and enrolled within the Women's Interagency HIV Study (WIHS) cohort. CD4 cell counts listed for TB patients are those closest to time of TB diagnosis.

TST: Tuberculin skin test; ART: antiretroviral therapy (with a minimum of two drugs against HIV at least 6 months prior to TB diagnosis or time of control subjects' visit); IQR: Interquartile range; VL: Viral load.

¹ *t* test

² Chi-square test

³ Mann-Whitney *U* test.

diagnosis, are provided in **supplemental Table 1**. The SA cohort displayed a larger number of proteins meeting our significance criteria than the US cohort ($n = 15$ versus 10, respectively), and the set of candidate biomarkers differed in part between SA and US incident TB subjects. Five markers overlapped between both cohorts. Specifically, these were soluble CD14, Alpha-2-glycoprotein (A2GL), Nidogen-1 (NID1), secreted and transmembrane protein 1 (SCTM1), and Alpha-1-acid glycoprotein 1 (A1AG1). Applying a local polynomial regression (LOESS) fit,⁴⁹ the temporal expression patterns for proteins meeting significance in both cohorts are shown in **Figure 2**. Temporal expression patterns of

representative examples of host proteins meeting criteria for increasing *Mtb* infection activity/burden in only one of the cohorts are shown in **Figure 3**. We note that the expression of only two proteins in the SA (BGH3 and TENX) and one in the US cohort (PLSL) were significantly different at a $p < 0.01$ [Mann-Whitney *U* test] between IGRA or TST positive and negative non-TB subjects, respectively, and that none of the protein expressions were significantly different between these non-TB subjects at 10% FDR in either cohort. Detailed information for all proteins for pre and post diagnosis slopes, as well as comparison of TB subjects (at time of diagnosis) to non-TB subjects, and between IGRA or

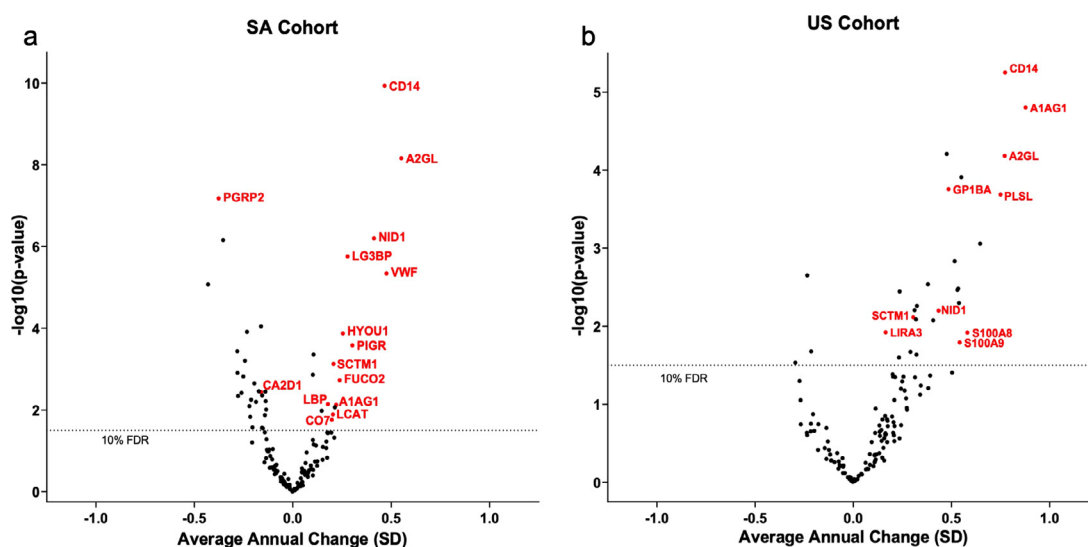


Figure 1. Volcano plots of trend analysis showing the significance and annual change for the differential expression of host proteins over time identified within the SA (A) and US (B) incident TB subjects. The negative \log_{10} -transformed *p* value corresponds with the significance of the pre-diagnosis slope during the 24 months prior to TB diagnosis. The horizontal axis shows the slope per year per standard deviation of intensities. Proteins highlighted in red are those meeting criteria of both significant pre-diagnosis slope (during the 24 months prior to TB diagnosis; $FDR < 10\%$) and significantly different expression ($p < 0.01$) in the incident TB subjects at time of TB diagnosis compared to the asymptomatic non-TB subjects (IGRA positive and negative combined for SA, and TST positive and negative combined for US cohort).

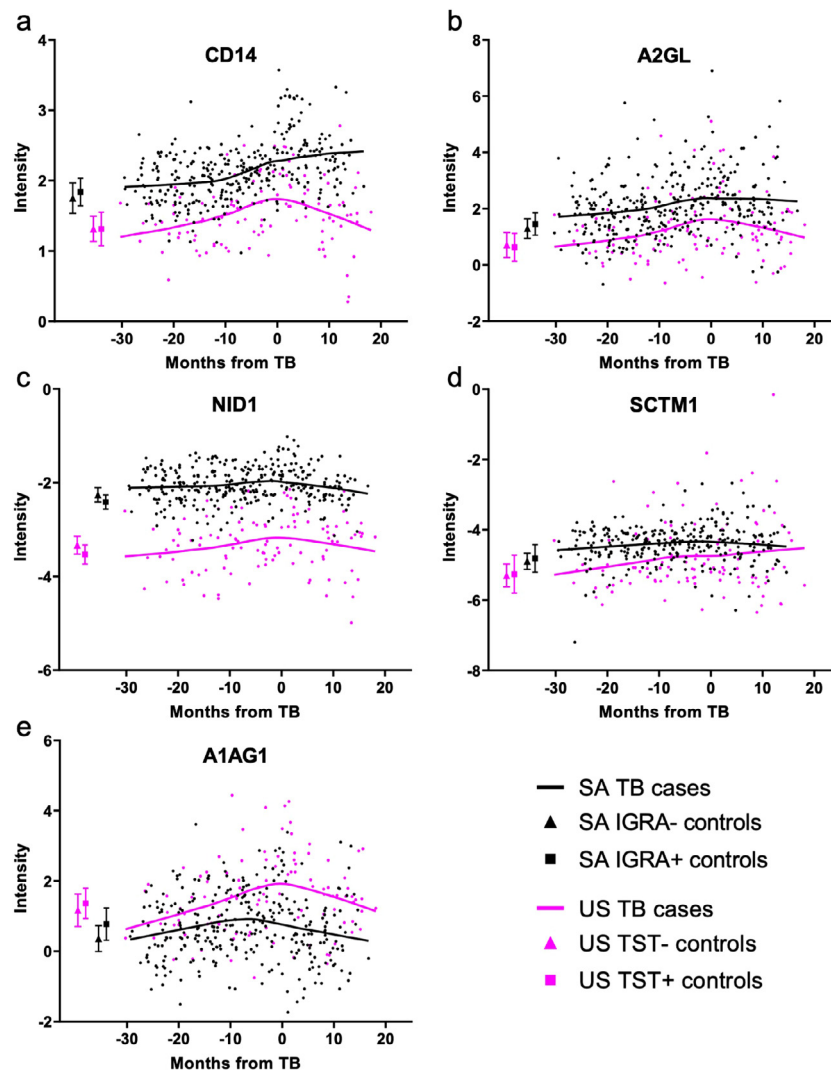


Figure 2. Temporal expression patterns for individual host proteins meeting criteria for markers of increasing *Mtb* infection activity/burden in both SA and US cohorts. Graphs show local polynomial regression (LOESS) fit of the incident TB subjects (n=30 for SA, and n=24 for US) and the mean and 95% confidence intervals for the non-TB subjects (n=62 for SA, and n=48 for US). Criteria consisted of i) a statistically significant slope (from 24 months prior to time of TB diagnosis until time of TB diagnosis) after adjusting for a false discovery rate (FDR) of 10%, and ii) being significantly differentially expressed ($p < 0.01$) in the incident TB subjects at time of TB diagnosis compared to a single baseline time-point for the asymptomatic non-TB cohort subjects (IGRA positive and negative combined for SA, and TST positive and negative combined for US cohort). A2GL: leucine-rich alpha-2-glycoprotein, NID1: nidogen-1, SCTM1: secreted and transmembrane protein 1, A1AG1: alpha-1-acid glycoprotein 1.

TST positive and negative non-TB subjects for both cohorts are listed in [supplemental Table 1](#).

Pathway analysis showed that among the molecular and cellular functions of the proteins correlating with increasing *Mtb* infection activity, cellular compromise was one of the most prominent in both cohorts ([Figure 4](#)). However, other top functional categories differed between the two cohorts, indicating potentially different interactions between *Mtb* and the host in PLHIV from SA versus those from the US.

To assess the prognostic performance of host proteins, we used SVMs^{41,42} with cross-validation to avoid overestimation of predictive accuracy. We identified panels of 5 - 8 proteins predicting TB development with excellent accuracy 0 - 6 (AUC 0.93) and 6 - 12 months (AUC 0.86) prior to TB diagnosis for the SA cohort, and good accuracy 0 - 6 (AUC 0.74) and 6 - 12 months (AUC 0.76) prior to TB diagnosis for the US cohort ([Figure 5](#); [Table 3](#)). Based on the SA data, 5 - 6 candidate proteins (CD248, HYOU1, NID1, PON1, VCAM1

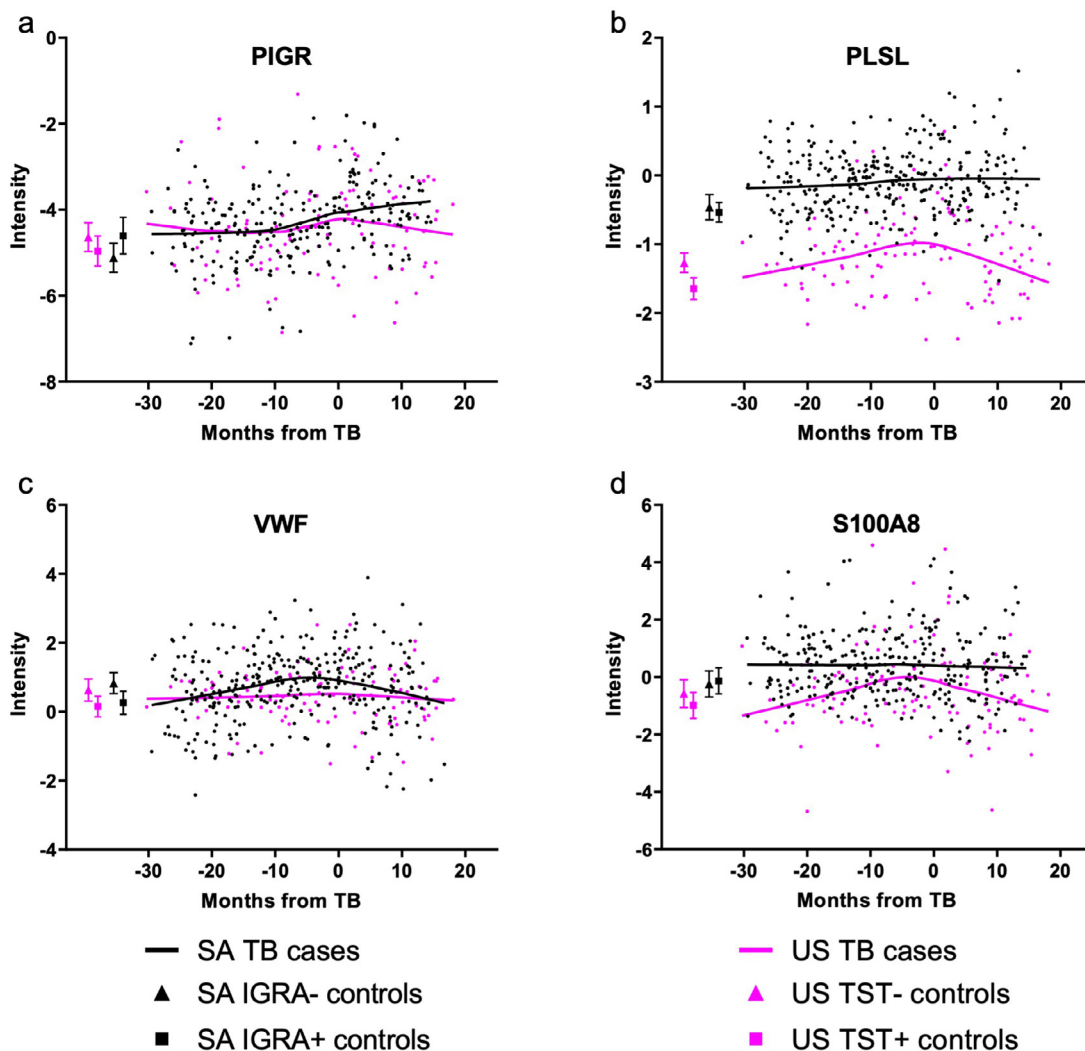


Figure 3. Temporal expression patterns for examples of individual host proteins meeting criteria for markers of increasing *Mtb* infection activity/burden in either only SA (A & C) or only US cohort (B & D). Graphs show local polynomial regression (LOESS) fit of the incident TB subjects (n=30 for SA, and n=24 for US) and the mean and 95% confidence intervals for the non-TB subjects (n=62 for SA, and n=48 for US). Criteria consisted of i) a statistically significant slope (from 24 months prior to time of TB diagnosis until time of TB diagnosis) after adjusting for a false discovery rate (FDR) of 10%, and ii) being significantly differentially expressed ($p < 0.01$) in the incident TB subjects at time of TB diagnosis to a single baseline time-point for the asymptomatic non-TB subjects (IGRA positive and negative combined for SA, and TST positive and negative combined for US cohort). PIGR: polymeric immunoglobulin receptor, PLSL: plastin-2, VWF: von Willebrand factor

and APOA1) can predict/detect TB with high accuracy up to 12 months prior to diagnosis - at a specificity of 80% with a sensitivity of 94% - 6 months prior to TB, and 86% 6 - 12 months prior to TB. Predictive accuracy 12 - 18 and 18 - 24 months prior to TB diagnosis remained mostly high (AUC 0.75 and 0.87 for SA, respectively; and 0.84 and 0.80 for US, respectively) albeit with different sets of up to 12 markers.

Discussion

Using MRM-MS and plasma samples from prospectively followed HIV-infected subjects, we identified host proteins that increase significantly from two years prior to until time of TB diagnosis, indicating that these proteins correlate with increasing *Mtb* infection activity/burden in PLHIV. Subjects from both cohorts, representing SA as a region with high TB incidence and the

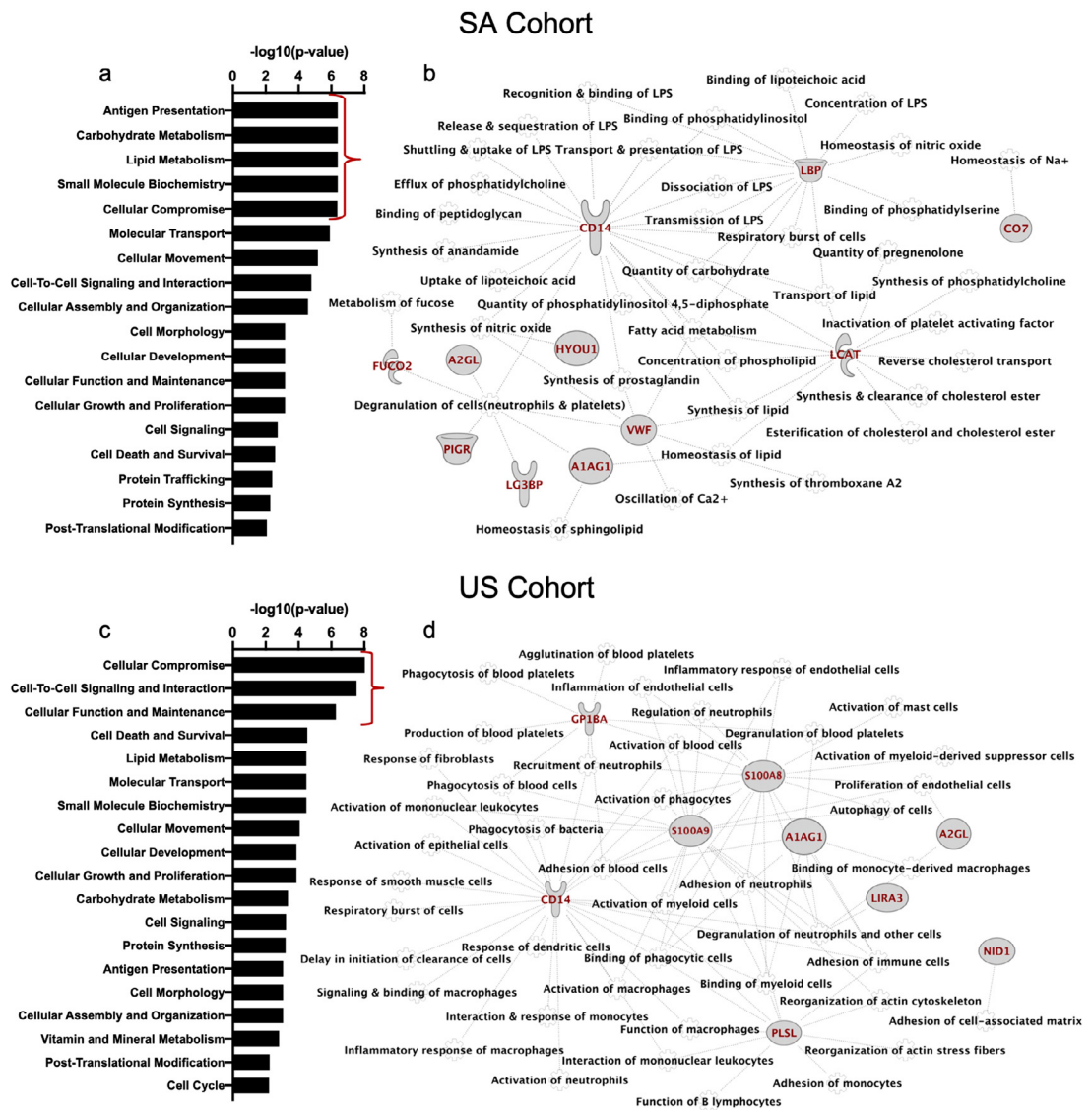


Figure 4. Ingenuity Pathway Analysis (IPA) of most significant molecular and cellular functions for host proteins associated with *Mtb* infection activity/burden in SA (A & B) or US cohort (C & D). A & C. Top cellular and molecular functions ($p < 0.01$) associated with host proteins generated by IPA's Diseases and Functions analysis. P -values refer to the association between a given protein and function and were calculated using a Right Tailed Fisher's Exact test. B & D. Detailed networks are shown for proteins within most significantly associated functions ($-\log(p\text{-value}) > 6$).

US as a region with low TB incidence, displayed significant differential expression changes in partially overlapping subsets of proteins leading up to TB diagnosis. We identified 15 candidate proteins in the SA and 10 in the US cohort that met our selection criteria for increasing *Mtb* burden - significance for both change over time after adjusting for 10% FDR and highly significant difference ($p < 0.01$) at time of TB diagnosis compared to asymptomatic IGRA or TST positive and negative cohort subjects who did not develop TB. Many of the upregulated candidate proteins reversed significantly in

the months following TB diagnosis, presumably due to treatment response. Five proteins, CD14, A2GL, NID1, SCTM1, and A1AG1, comprising 30 % of the SA and 50 % of the US markers for increasing *Mtb* infection activity/burden, overlapped between both cohorts.

Although not in this combination, 4 of the 5 proteins overlapping in both cohorts have been identified in other, predominantly cross-sectional TB biomarker studies, corroborating their relevance as candidate markers for TB. Our most significant marker, CD14, is a well-characterized pattern recognition receptor for

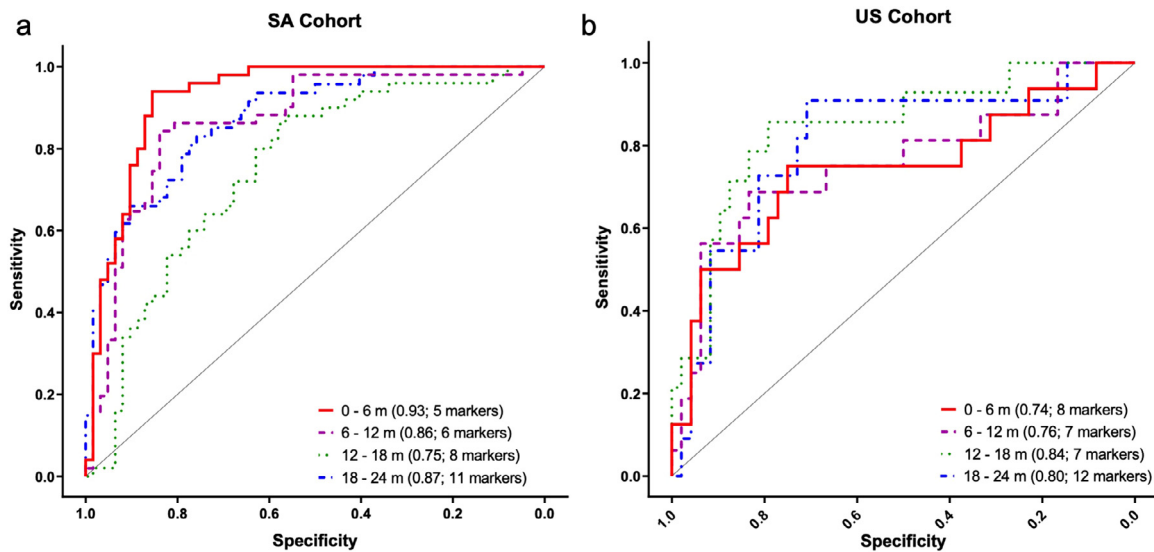


Figure 5. Discrimination of patients with incident TB from asymptomatic non-TB subjects at various time intervals prior to TB diagnosis. Area under the receiver operating curve (AUC) is shown for the (A) SA cohort and (B) US cohort. Non-TB subjects consisted of IGRA positive and negative combined for the SA, and TST positive and negative combined for the US cohort. Performance of protein panels was assessed using SVM with cross-validation to avoid overfitting. AUC values and number of markers are listed in parentheses; m refers to months prior to TB diagnosis).

microbial ligands, expressed on the surface of monocytes and polymorphonuclear cells.^{50,51} Its binding to bacterial and mycobacterial lipopolysaccharide and other cell wall components induces cellular activation and secretion of inflammatory cytokines.^{51–53} Upon activation, CD14 is shed in soluble form into serum. Elevated serum concentrations of soluble CD14 have been found in HIV-uninfected pulmonary TB patients compared to those with LTBI or no *Mtb* infection.^{54–55} Importantly, in our prior unbiased MS-based screening with biologically unrelated sera from HIV uninfected as well as co-infected patients from the US, soluble CD14 was one of the key markers to

distinguish TB from both LTBI and other respiratory diseases.³¹ Through use of isobaric tags for relative and absolute quantitation (iTRAQ) combined with MALDI-TOF-MS, CD14 was further identified among a set of 26 host proteins significantly increased in Chinese patients with TB versus those with other pneumonias.⁵⁶ Using sandwich ELISA, we recently further showed significantly elevated soluble CD14 levels in biologically independent sera from SA HIV-infected patients with smear-negative TB compared to asymptomatic HIV-infected subjects with and without signs of LTBI.⁵⁷ Collectively, these data suggest that CD14 is a highly promising marker for

Time pre-TB	SA cohort	US cohort
0 – 6 mo	CD248, HYOU1, NID1, PON1, VCAM1	CD14, CD166, DSG2, LAMP1, LRP1, NCAM2, R4RL2, VASN
6 – 12 mo	CD248, HYOU1, NID1, PON1, VCAM1, APOA1	APOA1, APOA4, GP1BA, LUM, NID1, PLSL, SCTM1
12 – 18 mo	APOA4, CBPN, CD14, CPN2, LCAT, LUM, LYAM1, PNPH	APOA4, CATA, COL11, LUM, MEGF8, PLSL, SCTM1
18 – 24 mo	CD248, CSTN1, FUCO2, HYOU1, MINP1, MMP2, MYOC, NID1, PON1, VASN, VCAM1	AMPN, CA2D1, CBPQ, CNTN1, COL11, DPP4, HYOU1, ICAM1, LAMP1, LUM, PCOC1, PIGR

Table 3: Composition of host protein biomarker panels distinguishing incident TB cases prior to TB diagnosis from controls.

HYOU1: hypoxia up-regulated protein 1; NID1: nidogen-1; PON1: serum paraoxonase/arylesterase 1; VCAM1: vascular cell adhesion protein 1; DSG2: desmoglein-2; LAMP1: lysosome-associated membrane glycoprotein 1; LRP1: prolow-density lipoprotein receptor-related protein 1; NCAM2: neural cell adhesion molecule 2; R4RL2: reticulon-4 receptor-like 2; VASN: vasorin; APOA1: apolipoprotein A-I; APOA4: apolipoprotein A-IV; GP1BA: platelet glycoprotein Ib alpha chain; LUM: lumican; PLSL: plastin-2; SCTM1: secreted and transmembrane protein 1; CBPN: carboxypeptidase N catalytic chain; CPN2: carboxypeptidase N subunit 2; LCAT: lecithin-cholesterol acyltransferase; LUM: lumican; LYAM1: L-selectin; PNPH: purine nucleoside phosphorylase; CATA: catalase; COL11: collectin-11; MEGF8: multiple epidermal growth factor-like domains protein 8; CSTN1: calsyntenin-1; FUCO2: plasma alpha-L-fucosidase; MINP1: multiple inositol polyphosphate phosphatase 1; MMP2: matrix metalloproteinase-2; MYOC: myocilin; AMPN: aminopeptidase N; CA2D1: voltage-dependent calcium channel subunit alpha-2/delta-1; CBPQ: carboxypeptidase Q; CNTN1: contactin-1; DPP4: dipeptidyl peptidase 4; ICAM1: intercellular adhesion molecule 1; LAMP1: lysosome-associated membrane glycoprotein 1; PCOC1: procollagen C-endopeptidase enhancer 1; PIGR: polymeric immunoglobulin receptor.

increasing *Mtb* infection activity/burden and the early detection of HIV-associated TB.

Less is known about the TB biomarker performance of NID1 and SCTM1. Also referred to as entactin, NID1 is a multifunctional glycoprotein present in basement membranes, the sheet of extracellular matrix underlying epithelial and endothelial cells.⁵⁸ NID1 is involved in epithelial cell apoptosis,⁵⁹ stimulates neutrophil adhesion and chemotaxis,⁶⁰ and was further recently shown to regulate natural killer (NK) cell function by modulating cytotoxicity or interferon-gamma (IFN- γ) production.⁶¹ In our prior unbiased MS-based screening with biologically unrelated sera, NID1 distinguished TB from both LTBI and other respiratory diseases in HIV co-infected US patients.³¹ Recently, it was also identified by iTRAQTM-coupled LC-MS/MS among a subset of proteins elevated in tuberculous compared to malignant pleural effusions.⁶² Even less is known about the TB biomarker performance of SCTM1, a glycoprotein expressed by neutrophils, monocytes, and epithelial cells.^{63,64} Its expression is inducible by IFN- γ , leads to further IFN- γ production, and it costimulates CD4 and CD8 proliferation.^{63,64} It has been shown to modulate neutrophil and lymphocyte activation in airway epithelial cell infection with *Streptococcus pneumoniae* and human respiratory syncytial virus.^{65,66}

Of the remaining two candidates overlapping between SA and US cohorts, we identified both A2GL (alpha-2-glycoprotein) and A1AG1 (alpha-1-acid glycoprotein 1) in our prior unbiased MS-based screening with biologically unrelated sera as proteins able to distinguish TB from other respiratory diseases in HIV-uninfected subjects.³¹ A2GL is involved in signal transduction and cell adhesion.⁶⁷ Also referred to as LRG1 (leucine-rich glycoprotein 1), it has been shown to be increased in different inflammatory conditions, such as rheumatoid arthritis and inflammatory bowel diseases.^{68–70} Serum LRG1 levels were found to be significantly higher in TB patients than healthy non-TB subjects and declined one month post-treatment.⁷¹ A1AG1, also referred to as orosomucoid1 (ORM1), is an acute-phase protein which has immune-modulating effects including inhibition of leukocyte adhesion/migration, decreased neutrophil chemotaxis and superoxide production.^{72–74} A1AG1 can be a potent inducer of the largely suppressive M2b macrophage subset leading to increased susceptibility to opportunistic infections.^{75–77} Production of the protein by alveolar macrophages and type 2 pneumocytes in experimental pulmonary TB in Balb/c mice suppressed cell mediated immunity and facilitated bacillary growth.⁷⁸ A1AG1 has been shown to distinguish TB from LTBI, and is significantly more expressed in severe compared to mild TB or uninfected states in HIV-uninfected subjects.^{79–81} Despite their promise as biomarker for TB and *Mtb* burden, the increased expression of both A2GL and A1AG1 in a range of inflammatory conditions warrants further

assessment of their capability to distinguish TB from other respiratory diseases in PLHIV.

Several of our identified host proteins met our criteria for increasing *Mtb* infection activity/burden in one but not the other cohort, indicating differences in TB pathogenesis and immune responses between the incident TB subjects from SA and the US. For example, the polymeric immunoglobulin receptor (PIGR), which binds and transports locally produced IgA and IgM from the basolateral to the apical surface of mucosal epithelial cells and enhances the immune functions of secretory IgA and IgM,^{82,83} was significantly associated with *Mtb* infection activity/burden in the SA but not US cohort subjects. Little is known of PIGR as a biomarker for TB. PIGR knockout compared to wildtype mice are i) more susceptible to airway mycobacterial infection,⁸⁴ and ii) lose the protective efficacy of respiratory mucosal immunization with the *Mtb* whole cell inactivated vaccine MTBVAC.⁸⁵ Collectively, the data suggest that the upregulation of PIGR might be an important component of the humoral immune response in the airways of TB patients to combat the disease. The less pronounced increase of PIGR expression in the US versus SA incident TB subjects could be due to their lower ongoing *Mtb* exposure and/or lower median CD4 count of 196 versus 306, respectively, and their thus higher likelihood of disseminated TB subjects within this cohort, both hypotheses that warrant further investigation.

By contrast, expression of S100A8 and S100A9, two highly related proteins, also referred to as Calgranulin A and B, or migration inhibitory factor-related proteins 8 and 14 (MRP-8 and MRP-14), respectively, was significantly associated with *Mtb* infection activity/burden in the US but not the SA incident TB subjects. Released by primed phagocytes, these proteins promote further recruitment of neutrophils and monocytes by inducing production of proinflammatory cytokines and chemokines (reviewed in^{86,87}). Serum levels have been associated with extent of inflammatory lung damage and disease severity in TB.^{81,88,89} Consistent with our data in HIV co-infected incident TB subjects, expression of S100A8/9 mRNA was increased in HIV uninfected people who progressed to TB compared to those who did not, and returned to baseline levels upon TB treatment.⁹⁰ Interestingly, we observed no increase in S100A8/9 expression in the SA incident TB subjects, likely due to the already higher baseline levels of these proteins in the SA compared to the US incident TB subjects up to two years prior to TB diagnosis. Whether a history of prior TB, which half of the SA TB subjects had, could have contributed to the higher S100A8/9 baseline expression, warrants further investigation.

Our data provide important insights into the pathogenesis associated with increasing *Mtb* infection activity/burden in PLHIV. Among the molecular and cellular functions of candidate proteins, cellular compromise was one of the most prominent in both cohorts.

Other top functional categories differed between the two cohorts, indicating different host-pathogen interactions in SA versus US PLHIV who develop TB. For example, antigen presentation and lipid and carbohydrate metabolic functions were highly prominent among the SA, whereas cell-to-cell signaling and interactions and cellular maintenance were more prominent among the biomarker proteins identified in US incident TB subjects. Among the contributing factors for these differences could be the lack of ART in the SA versus treatment in 40-50% of the US subjects, the higher exposure to *Mtb* in SA than the US subjects, and the higher median CD4 counts in the SA compared to the US incident TB subjects (~300 vs 200 cells/mm³), suggesting that ART, remote versus recent *Mtb* exposure/infection, and the level of CD4 count could drive the difference in host-pathogen interactions in PLHIV who develop TB.

While our primary objective was the identification of biomarkers for increasing *Mtb* infection activity/burden, we also identified panels of 5 - 12 host proteins predictive of TB development. The key difference between these two approaches was the focus on significantly increasing expression of proteins from two years prior until time of TB diagnosis in the former, whereas predictive markers would not need to fulfill this criterion in the latter. Anticipated and similar to the markers of *Mtb* infection activity/burden, these proteins also differed in part between SA and US cohorts and predictive panels changed over time. Panels of 5 - 8 proteins predicted TB development with excellent accuracy for the SA cohort 0 - 6 and 6 - 12 months prior to TB diagnosis (AUC 0.93 and 0.86, respectively), and good accuracy for the US cohort (AUC 0.74 and 0.76, respectively). For the SA cohort, at a specificity of 80% with a sensitivity of 94% 0 - 6 months prior to TB, and 86% 6 - 12 months prior to TB, the combination of 5 - 6 markers met the WHO target for a predictive TB test.⁹¹ The predictive accuracy of panels of 7 - 12 proteins for 12 - 18 and 18 - 24 months prior to TB diagnosis was lower for some time-frames (AUC 0.75 and 0.87 for SA, and 0.84 and 0.80 for US, respectively), but with overall higher predictive values than those reported for 12 - 24 months prior to TB from other TB predictive biomarker studies in HIV uninfected cohorts.²³⁻²⁷ To account for the lack of validation in biologically independent samples, we performed cross-validation for AUC assessments. However, these data should be considered hypothesis generating, with the need to assess and validate candidate markers in other prospective cohorts.

Collectively, we show that plasma host proteins can be biomarkers for increasing *Mtb* burden in PLHIV. Measurement of such markers in the form of a simple blood test could be integrated into routine outpatient follow-up visits of PLHIV and trigger screening and monitoring for TB when proteins are elevated. Given the likely regional influence with differences in *Mtb* infection pressure and host factors combined with the

broad spectrum of clinical manifestations, such a blood test would probably require more than 1-2 proteins but ideally should consist of no more than a handful. Although we have here identified a range of promising biomarkers, validation studies are needed to determine which specific host proteins will ultimately show the best performance. Of particular interest would be further assessments of biomarker panels including proteins such as soluble CD14 and NID1, which have shown ability to distinguish TB from other respiratory diseases in prior studies,³¹ and in the current study are both part of predictive panels and markers of increasing *Mtb* infection activity/burden prior to TB diagnosis in both SA and US cohorts.

A few other recent large-scale longitudinal studies have identified blood transcriptomic, metabolic, and proteomic signatures predicting TB prior to diagnosis but have focused mostly on HIV-uninfected adolescents and adults.²³⁻²⁷ Related to our proteomics focus, a panel of 5 host proteins predicted TB in a cohort of South African HIV uninfected adolescents with an AUC of 0.96 0 - 6 months and 0.76 6 - 12 months prior to TB diagnosis.²⁶ However, when validated in HIV uninfected Gambian cohort subjects, AUC decreased to 0.66 within one year prior to TB diagnosis, showing cohort-specific performance variability even within TB endemic African countries. Predictive RNA signatures have been the most studied and validated, albeit with cohort-specific variability in performance.^{23-25,28} These gene signatures are attractive options for predicting who among household contacts will develop TB within a few years post contact. However, due to their need for RNA sequencing and thus laboratory infrastructure, as well as their limit to short-term risk prediction when analysed across cohorts,⁹² gene signatures are impractical for regular screening of PLHIV who are at high risk for TB development but in whom neither the time of initial infection nor the time to TB development would be known. By contrast, plasma host proteins can be detected by simple methods and their monitoring can be easily integrated into routine clinical follow-up visits of PLHIV.

With 30 incident TB subjects in the SA and 24 in the US cohort, our study was limited by a relatively small sample size. Furthermore, because subjects in both cohorts were followed for HIV progression, our study was limited by lack of clinical details for TB subjects. While medical information was available for SA TB subjects, we had limited clinical information for the US TB subjects and used the date a subject was reported as having TB to a US state department of health as the time of TB diagnosis. Since we investigated changes of TB markers over time and identified significant and in part overlapping markers in both cohorts, we do not believe that this limitation had a major impact on our results. Due to restricted funds, testing of non-TB subjects was limited to a single time-point around a recent IGRA or TST. We do however have unpublished data on

individual markers (e.g. CD14 and NID-1) measured via targeted immunoassays in sera over two years of consecutive visits from the non-TB subjects who did not develop TB and did not see significant changes over time (unpublished) Although we performed cross-validation for AUC assessments, another limitation was the lack of validation in cohorts from various TB endemic regions. Larger studies investigating our identified markers for their association with *Mtb* infection activity/burden and their ability to distinguish TB from other respiratory diseases in PLHIV in various TB endemic regions are now warranted.

In conclusion, a set of human host proteins correlate with increasing *Mtb* infection activity/burden in PLHIV in high and low TB endemic regions. Complementary to transcriptomic, metabolic, and proteomic signatures predicting progression to TB in HIV uninfected individuals, our data provide important insights into the host responses associated with increasing *Mtb* burden in PLHIV. If further validated, our results can inform the development of simple blood tests that could be integrated into existing algorithms for monitoring PLHIV, and trigger screening and monitoring for TB when proteins are elevated. Used as a single platform, or when combined with other biomarkers as a multi-platform tool, such tests can aid in optimizing the timing of the initiation of TB preventive or therapeutic therapy in PLHIV and in increasing its effectiveness.

Contributors

Singer SN: data collection, data verification, and data curation, literature search, figures, data analysis, data interpretation, writing – original draft, review & editing; Ndumnego OC: data collection, literature search, data interpretation, writing – original draft, review & editing; Kim RS: data curation, formal analysis, figures, data interpretation, writing – original draft, review & editing; Ndung'u T: resources, writing – review & editing; Anastos K: resources, data collection, writing – review & editing; French A: resources, data collection, writing – review & editing; Churchyard G: data interpretation, writing – review & editing; Paramithiothis E: data collection, data verification, resources, project administration, supervision, writing – review & editing; Kasproicz VO: resources, data collection, writing – review & editing; Achkar JM: conceptualisation, funding acquisition, resources, project administration, supervision, methodology, investigation, data collection, data verification, formal analysis, data interpretation, visualisation, literature search, writing – original draft, and writing – review & editing. All authors read and approved the final version of this manuscript.

Declaration of interests

EP is the Vice President, Research and Development at CellCarta Biosciences Inc, formerly Caprion Biosciences

Inc. The host protein biomarkers described herein are the subject of an international patent application entitled “Methods for assessing the risk of developing active tuberculosis” (PCT/US20/54083) with EP and JMA as co-inventors. Both EP and JMA report no other conflict of interests related to this report, and all other authors declared no conflict of interest.

Data sharing statement

SNS, RK, EP, and JMA had full access to the study data. MRM-MS data generated in this study are deposited in the PeptideAtlas database (Institute for Systems Biology), a publically accessible repository for MS proteomics data (reference# PASS01693).

Acknowledgements

This work was supported by funds from the National Institutes of Health (NIH)/National Institute of Allergy and Infectious Diseases (NIAID) to JMA (A117927, A1146329, and A1127173). Data in this manuscript were collected by WIHS, now the MACS/WHIS Combined Cohort Study (MWCCS). The MWCCS is funded by the National Institute of Health, primarily by the National Heart, Lung, and Blood Institute (NHLBI), with additional co-funding from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), National Institute on Aging (NIA), National Institute of Dental and Craniofacial Research (NIDCR), NIAID, National Institute of Neurological Disorders and Stroke (NINDS), National Institute of Mental Health (NIMH), National Institute on Drug Abuse (NIDA), National Institute of Nursing Research (NINR), National Cancer Institute (NCI), National Institute on Alcohol Abuse and Alcoholism (NIAAA), National Institute on Deafness and Other Communication Disorders (NIDCD), National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute on Minority Health and Health Disparities (NIMHD), and in coordination and alignment with the research priorities of the National Institutes of Health, Office of AIDS Research (OAR). TN was supported in part by the Strategic Health Innovation Partnerships (SHIP) Unit of the South African Medical Research Council with funds received from the South African Department of Science and Innovation. Other support came from the South African Research Chairs Initiative and the Victor Daitz Foundation. The work was further supported through a post-doctoral grant to OCN by the Sub-Saharan African Network for TB/HIV Research Excellence (SANTHE), a DELTAS Africa Initiative (DEL-15-006). The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and

Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust (107752/Z/15/Z) and the UK government. The views expressed in this publication are those of the author(s) and not necessarily those of AAS, NEPAD Agency, Wellcome Trust or the UK government.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ebiom.2021.103787.

References

- Corbett EL, Watt CJ, Walker N, et al. The growing burden of tuberculosis: global trends and interactions with the HIV epidemic. *Arch Intern Med* 2003;163(9):1009–21.
- Salgame P, Geadas C, Collins L, Jones-Lopez E, Ellner JJ. Latent tuberculosis infection - revisiting and revising concepts. *Tuberculosis (Edinb)* 2015;95(4):373–84.
- Denkinger CM, Kik SV, Cirillo DM, et al. Defining the needs for next generation assays for tuberculosis. *J Infect Dis* 2015;211 Suppl 2:S29–38.
- Nahid P, Kim PS, Evans CA, et al. Clinical research and development of tuberculosis diagnostics: moving from silos to synergy. *J Infect Dis* 2012;205 Suppl 2:S159–68.
- Sterling TR, Pham PA, Chaisson RE. HIV infection-related tuberculosis: clinical manifestations and treatment. *Clin Infect Dis* 2010;50: S223–S30.
- Gonzalez Fernandez L, Casas EC, Singh S, et al. New opportunities in tuberculosis prevention: implications for people living with HIV. *J Int AIDS Soc* 2020;23(1):e25438.
- Achkar JM, Jenny-Avital ER. Incipient and subclinical tuberculosis: defining early disease states in the context of host immune response. *J Infect Dis* 2011;204 Suppl 4:S179–86.
- Barry 3rd CE, Boshoff HI, Dartois V, et al. The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. *Nat Rev Microbiol* 2009;7(12):845–55.
- Lawn SD, Wood R, Wilkinson RJ. Changing concepts of "latent tuberculosis infection" in patients living with HIV infection. *Clin Dev Immunol* 2011;2011.
- Drain PK, Bajema KL, Dowdy D, et al. Incipient and subclinical tuberculosis: a clinical review of early stages and progression of infection. *Clin Microbiol Rev* 2018;31(4).
- Lin PL, Flynn JL. The end of the binary era: revisiting the spectrum of tuberculosis. *J Immunol* 2018;201(9):2541–8.
- Esmail H, Lai RP, Lesosky M, et al. Characterization of progressive HIV-associated tuberculosis using 2-deoxy-2-(18F)fluoro-D-glucose positron emission and computed tomography. *Nat Med* 2016;22(10):1090–3.
- Boom WH, Schaible UE, Achkar JM. The knowns and unknowns of latent Mycobacterium tuberculosis infection. *J Clin Invest* 2021;131(3).
- World Health Organization. Guidelines for intensified tuberculosis case-finding and isoniazid preventive therapy for people living with HIV in resource-constrained settings. Geneva, Switzerland http://www.who.int/tb/challenges/hiv/ICF_IPTguidelines/en/.
- Churchyard GJ, Fielding KL, Lewis JJ, et al. A trial of mass isoniazid preventive therapy for tuberculosis control. *N Engl J Med* 2014;370(4):301–10.
- Grant AD, Charalambous S, Fielding KL, et al. Effect of routine isoniazid preventive therapy on tuberculosis incidence among HIV-infected men in South Africa: a novel randomized incremental recruitment study. *JAMA* 2005;293(22):2719–25.
- Johnson JL, Okwera A, Hom DL, et al. Duration of efficacy of treatment of latent tuberculosis infection in HIV-infected adults. *AIDS* 2001;15(16):2137–47.
- Hermans SM, Grant AD, Chihota V, et al. The timing of tuberculosis after isoniazid preventive therapy among gold miners in South Africa: a prospective cohort study. *BMC Med*. 2016;14:45.
- Samandari T, Agizew TB, Nyirenda S, et al. 6-month versus 36-month isoniazid preventive treatment for tuberculosis in adults with HIV infection in Botswana: a randomised, double-blind, placebo-controlled trial. *Lancet* 2011;377(9777):1588–98.
- Shah M, Hanrahan C, Wang ZY, et al. Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults. *Cochrane Database Syst Rev* 2016(5):CD011420.
- World Health Organization. The use of lateral flow urine lipoarabinomannan assay (LF-LAM) for the diagnosis and screening of active tuberculosis in people living with HIV. Geneva, Switzerland: World Health Organization; 2015.
- Broger T, Sossen B, du Toit E, et al. Novel lipoarabinomannan point-of-care tuberculosis test for people with HIV: a diagnostic accuracy study. *Lancet Infect Dis* 2019;19(8):852–61.
- Zak DE, Penn-Nicholson A, Scriba TJ, et al. A blood RNA signature for tuberculosis disease risk: a prospective cohort study. *Lancet* 2016.
- Warsinske HC, Rao AM, Moreira FMF, et al. Assessment of validity of a blood-based 3-gene signature score for progression and diagnosis of tuberculosis, disease severity, and treatment response. *JAMA Netw Open* 2018;1(6):e183779.
- Suliman S, Thompson EG, Sutherland J, et al. Four-Gene Pan-African blood signature predicts progression to tuberculosis. *Am J Respir Crit Care Med* 2018;197(9):1198–208.
- Penn-Nicholson A, Hraha T, Thompson EG, et al. Discovery and validation of a prognostic proteomic signature for tuberculosis progression: a prospective cohort study. *PLoS Med* 2019;16(4): e1002781.
- Weiner 3rd J, Maertzdorf J, Sutherland JS, et al. Metabolite changes in blood predict the onset of tuberculosis. *Nat Commun* 2018;9(1):5208.
- Darboe F, Mbandi SK, Naidoo K, et al. Detection of tuberculosis recurrence, diagnosis and treatment response by a blood transcriptomic risk signature in HIV-infected persons on antiretroviral therapy. *Front Microbiol* 2019;10:1441.
- Esmail H, Lai RP, Lesosky M, et al. Complement pathway gene activation and rising circulating immune complexes characterize early disease in HIV-associated tuberculosis. *Proc Natl Acad Sci U S A* 2018;115(5):E964–E73.
- Tornheim JA, Dooley KE. Tuberculosis associated with HIV infection. *Microbiol Spectr* 2017;5(1).
- Achkar JM, Cortes L, Croteau P, et al. Host protein biomarkers identify active tuberculosis in HIV uninfected and co-infected individuals. *EBioMedicine* 2015;2(9):1160–8.
- Bark CM, Manceur AM, Malone LL, et al. Identification of host proteins predictive of early stage mycobacterium tuberculosis infection. *EBioMedicine* 2017;21:150–7.
- Brumme Z, Wang B, Nair K, et al. Impact of select immunologic and virologic biomarkers on CD4 cell count decrease in patients with chronic HIV-1 subtype C infection: results from Sinikithemba Cohort, Durban, South Africa. *Clin Infect Dis* 2009;49(6):956–64.
- Bacon MC, von Wyl V, Alden C, et al. The Women's Interagency HIV Study: an observational cohort brings clinical sciences to the bench. *Clin Diagn Lab Immunol* 2005;12(9):1013–9.
- Barkan SE, Melnick SL, Preston-Martin S, et al. The women's interagency HIV study. WIHS Collaborative Study Group. *Epidemiology* 1998;9(2):117–25.
- Adimora AA, Ramirez C, Benning L, et al. Cohort profile: the Women's interagency HIV study (WIHS). *Int J Epidemiol* 2018;47(2):393–41.
- Kasprowicz VO, Mitchell JE, Chetty S, et al. A molecular assay for sensitive detection of pathogen-specific T-cells. *PLoS One* 2011;6(8): e20606.
- Lewinsohn DM, Leonard MK, LoBue PA, et al. Official American Thoracic Society/Infectious Diseases Society of America/Centers for Disease Control and Prevention clinical practice guidelines: diagnosis of tuberculosis in adults and children. *Clin Infect Dis* 2017;64(2):111–5.
- Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Softw* 2014;67(1):1–48.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B (Methodol)* 1995;57(1):289–300.
- Cortes C, Vapnik V. Support-vector networks. *Mach Learn* 1995;20(3):273–97.
- Chang C-C, Lin C-J. LIBSVM. A library for support vector machines. *ACM Trans Intell Syst Technol (TIST)* 2011;2(3):1–27.
- Kia A, Timsina P, Joshi HN, et al. MEWS++: enhancing the prediction of clinical deterioration in admitted patients through a machine learning model. *J Clin Med* 2020;9(2).

- 44 Merali ZG, Witiw CD, Badhiwala JH, Wilson JR, Fehlings MG. Using a machine learning approach to predict outcome after surgery for degenerative cervical myelopathy. *PLoS One* 2019;14(4):e0215133.
- 45 Wottschel V, Chard DT, Enzinger C, et al. SVM recursive feature elimination analyses of structural brain MRI predicts near-term relapses in patients with clinically isolated syndromes suggestive of multiple sclerosis. *Neuroimage Clin* 2019;24:102011.
- 46 Zhang X, Lu X, Shi Q, et al. Recursive SVM feature selection and sample classification for mass-spectrometry and microarray data. *BMC Bioinform* 2006;7:197.
- 47 Bachoo RM, Kim RS, Ligon KL, et al. Molecular diversity of astrocytes with implications for neurological disorders. *Proc Natl Acad Sci U S A* 2004;101(22):8384–9.
- 48 Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing; 2018.
- 49 Cleveland W, Grosse E, Shyu W. Local regression models. Chapter 8 in *Statistical models in S* (JM Chambers and TJ Hastie eds.), 608 p. Wadsworth & Brooks/Cole, Pacific Grove, CA1992.
- 50 Pugin J, Heumann ID, Tomasz A, et al. CD14 is a pattern recognition receptor. *Immunity* 1994;1(6):509–16.
- 51 Frey EA, Miller DS, Jahr TG, et al. Soluble CD14 participates in the response of cells to lipopolysaccharide. *J Exp Med* 1992;176(6):1665–71.
- 52 Goyert SM, Ferrero EM, Seremetis SV, Winchester RJ, Silver J, Mattison AC. Biochemistry and expression of myelomonocytic antigens. *J Immunol* 1986;137(12):3909–14.
- 53 Goyert SM, Ferrero E, Rettig WJ, Yenamandra AK, Obata F, Le Beau MM. The CD14 monocyte differentiation antigen maps to a region encoding growth factors and receptors. *Science* 1988;239(4839):497–500.
- 54 Feruglio SL, Troseid M, Damas JK, Kvale D, Dyrhol-Riise AM. Soluble markers of the Toll-like receptor 4 pathway differentiate between active and latent tuberculosis and are associated with treatment responses. *PLoS One* 2013;8(7):e69896.
- 55 Zambuzi FA, Cardoso-Silva PM, Espindola MS, et al. Identification of promising plasma immune biomarkers to differentiate active pulmonary tuberculosis. *Cytokine* 2016;88:99–107.
- 56 Li C, He X, Li H, et al. Discovery and verification of serum differential expression proteins for pulmonary tuberculosis. *Tuberculosis (Edinb)* 2015;95(5):547–54.
- 57 Liu Y, Nduemego OC, Chen T, et al. Soluble CD14 as a diagnostic biomarker for smear-negative HIV-associated tuberculosis. *Pathogens* 2018;7(1).
- 58 Dedhar S, Jewell K, Rojiani M, Gray V. The receptor for the basement membrane glycoprotein entactin is the integrin alpha 3/beta 1. *J Biol Chem* 1992;267(26):18908–14.
- 59 Alexander CM, Howard EW, Bissell MJ, Werb Z. Rescue of mammary epithelial cell apoptosis and entactin degradation by a tissue inhibitor of metalloproteinases-1 transgene. *J Cell Biol* 1996;135(6 Pt 1):1669–77.
- 60 Senior RM, Gresham HD, Griffin GL, Brown EJ, Chung AE. Entactin stimulates neutrophil adhesion and chemotaxis through interactions between its Arg-Gly-Asp (RGD) domain and the leukocyte response integrin. *J Clin Invest* 1992;90(6):2251–7.
- 61 Gaggero S, Bruschi M, Petretto A, et al. Nidogen-1 is a novel extracellular ligand for the NKp44 activating receptor. *Oncoimmunology* 2018;7(9):e1470730.
- 62 Shi J, Li P, Zhou L, et al. Potential biomarkers for antidiastole of tuberculous and malignant pleural effusion by proteome analysis. *Biomark Med* 2019;13(2):123–33.
- 63 Huyton T, Gottmann W, Bade-Doding C, Paine A, Blasczyk R. The T/NK cell co-stimulatory molecule SECTM1 is an IFN "early response gene" that is negatively regulated by LPS in human monocyte cells. *Biochimica et biophysica acta* 2011;1810(12):1294–301.
- 64 Wang T, Huang C, Lopez-Coral A, et al. K12/SECTM1, an interferon-gamma regulated molecule, synergizes with CD28 to costimulate human T cell proliferation. *J Leukoc Biol* 2012;91(3):449–59.
- 65 Touzelet O, Broadbent L, Armstrong SD, et al. The secretome profiling of a pediatric airway epithelium infected with hRSV identified aberrant apical/basolateral trafficking and novel immune modulating (CXCL6, CXCL16, CSF3) and antiviral (CEACAM1) proteins. *Mol Cell Proteomics* 2020;19(5):793–807.
- 66 Kamata H, Yamamoto K, Wasserman GA, et al. Epithelial cell-derived secreted and transmembrane 1a signals to activated neutrophils during pneumococcal pneumonia. *Am J Respir Cell Mol Biol* 2016;55(3):407–18.
- 67 Buchanan SG, Gay NJ. Structural and functional diversity in the leucine-rich repeat family of proteins. *Prog Biophys Mol Biol* 1996;65(1-2):1–44.
- 68 Rainer TH, Leung LY, Chan C, et al. Circulating human leucine-rich alpha-2-glycoprotein 1 mRNA and protein levels to detect acute appendicitis in patients with acute abdominal pain. *Clin Biochem* 2017;50(9):485–90.
- 69 Serada S, Fujimoto M, Ogata A, et al. iTRAQ-based proteomic identification of leucine-rich alpha-2 glycoprotein as a novel inflammatory biomarker in autoimmune diseases. *Annals of the rheumatic diseases* 2010;69(4):770–4.
- 70 Serada S, Fujimoto M, Terabe F, et al. Serum leucine-rich alpha-2 glycoprotein is a disease activity biomarker in ulcerative colitis. *Inflammatory Bowel Dis* 2012;18(11):2169–79.
- 71 Fujimoto M, Matsumoto T, Serada S, et al. Leucine-rich alpha 2 glycoprotein is a new marker for active disease of tuberculosis. *Sci Rep* 2020;10(1):3384.
- 72 Laine E, Couderc R, Roch-Arveiller M, Vasson M, Giroud J, Raichvarg D. Modulation of human polymorphonuclear neutrophil functions by alpha 1 acid glycoprotein. *Inflammation* 1990;14(1):1–9.
- 73 Mestriner F, Spiller F, Laure H, et al. Acute-phase protein alpha-1-acid glycoprotein mediates neutrophil migration failure in sepsis by a nitric oxide-dependent mechanism. *Proc Natl Acad Sci* 2007;104(49):19595–600.
- 74 Spiller F, Carlos D, Souto FO, et al. alpha-Acid glycoprotein decreases neutrophil migration and increases susceptibility to sepsis in diabetic mice. *Diabetes* 2012;61(6):1584–91.
- 75 Bories P, Guenounou M, Feger J, Kodari E, Agneray J, Durand G. Human alpha 1-acid glycoprotein-exposed macrophages release interleukin 1 inhibitory activity. *Biochem Biophys Res Commun* 1987;147(2):710–5.
- 76 Kobayashi M, Jeschke MG, Shigematsu K, et al. M2b monocytes predominated in peripheral blood of severely burned patients. *J Immunol* 2010;185(12):7174–9.
- 77 Nakamura K, Ito I, Kobayashi M, Herndon DN, Suzuki F. Orosomucoid 1 drives opportunistic infections through the polarization of monocytes to the M2b phenotype. *Cytokine* 2015;73(1):8–15.
- 78 Cordero EM, Gonzalez M, Aguilar L, Orozco E, Pando RH. Alpha-1-acid glycoprotein, its local production and immunopathological participation in experimental pulmonary tuberculosis. *Tuberculosis* 2008;88(3):203–11.
- 79 Zhang J, Wu X, Shi L, et al. Diagnostic serum proteomic analysis in patients with active tuberculosis. *Clin Chim Acta* 2012;413(9-10):883–7.
- 80 Sun H, Pan L, Jia H, et al. Label-free quantitative proteomics identifies novel plasma biomarkers for distinguishing pulmonary tuberculosis and latent infection. *Front Microbiol* 2018;9:1267.
- 81 Liu Q, Pan L, Han F, et al. Proteomic profiling for plasma biomarkers of tuberculosis progression. *Mol Med Rep* 2018;18(2):1551–9.
- 82 Johansen FE, Kaetzel CS. Regulation of the polymeric immunoglobulin receptor and IgA transport: new advances in environmental factors that stimulate pIgR expression and its role in mucosal immunity. *Mucosal Immunol* 2011;4(6):598–602.
- 83 Turula H, Wobus CE. The role of the polymeric immunoglobulin receptor and secretory immunoglobulins during mucosal infection and immunity. *Viruses* 2018;10(5).
- 84 Tjarnlund A, Rodriguez A, Cardona PJ, et al. Polymeric IgR knockout mice are more susceptible to mycobacterial infections in the respiratory tract than wild-type mice. *Int Immunol* 2006;18(5):807–16.
- 85 Aguilo N, Uranga S, Mata E, et al. Respiratory immunization with a whole cell inactivated vaccine induces functional mucosal immunoglobulins against tuberculosis in mice and non-human primates. *Front Microbiol* 2020;11:1339.
- 86 Foell D, Wittkowski H, Vogl T, Roth J. S100 proteins expressed in phagocytes: a novel group of damage-associated molecular pattern molecules. *J Leukoc Biol* 2007;81(1):28–37.
- 87 Sunahori K, Yamamura M, Yamana J, et al. The S100A8/A9 heterodimer amplifies proinflammatory cytokine production by macrophages via activation of nuclear factor kappa B and p38 mitogen-activated protein kinase in rheumatoid arthritis. *Arthritis Res Ther* 2006;8(3):R69.
- 88 Gopal R, Monin L, Torres D, et al. S100A8/A9 proteins mediate neutrophilic inflammation and lung pathology during tuberculosis. *Am J Respir Crit Care Med* 2013;188(9):1137–46.
- 89 Pechkovsky DV, Zalutskaya OM, Ivanov GI, Misuno NI. Calprotectin (MRP8/14 protein complex) release during mycobacterial

- infection in vitro and in vivo. *FEMS Immunol Med Microbiol* 2000;29(1):27–33.
- 90 Scott NR, Swanson RV, Al-Hammadi N, et al. S100A8/A9 regulates CD11b expression and neutrophil recruitment during chronic tuberculosis. *J Clin Invest* 2020;130(6):3098–112.
- 91 Consensus meeting report: development of a Target Product Profile (TPP) and a framework for evaluation for a test for predicting progression from tuberculosis infection to active disease. Geneva: World Health Organization; 2017 (WHO/HTM/TB/2017.18).
- 92 Gupta RK, Turner CT, Venturini C, et al. Concise whole blood transcriptional signatures for incipient tuberculosis: a systematic review and patient-level pooled meta-analysis. *Lancet Respir Med* 2020;8(4):395–406.