

OPEN

Serum proteomics of active tuberculosis patients and contacts reveals unique processes activated during *Mycobacterium tuberculosis* infection

Jesús Mateos^{1*}, Olivia Estévez², África González-Fernández², Luis Anibarro^{2,3,4}, Ángeles Pallarés³, Rajko Reljic⁵, Tufária Mussá^{6,7}, Cremildo Gomes-Maueia⁶, Artur Ngulichane⁸, José M. Gallardo¹, Isabel Medina¹ & Mónica Carrera^{1*}

Tuberculosis (TB) is the most lethal infection among infectious diseases. The specific aim of this study was to establish panels of serum protein biomarkers representative of active TB patients and their household contacts who were either infected (LTBI) or uninfected (EMI-TB Discovery Cohort, Pontevedra Region, Spain). A TMT (Tandem mass tags) 10plex-based quantitative proteomics study was performed in quintuplicate containing a total of 15 individual serum samples per group. Peptides were analyzed in an LC-Orbitrap Elite platform, and raw data were processed using Proteome Discoverer 2.1. A total of 418 proteins were quantified. The specific protein signature of active TB patients was characterized by an accumulation of proteins related to complement activation, inflammation and modulation of immune response and also by a decrease of a small subset of proteins, including apolipoprotein A and serotransferrin, indicating the importance of lipid transport and iron assimilation in the progression of the disease. This signature was verified by the targeted measurement of selected candidates in a second cohort (EMI-TB Verification Cohort, Maputo Region, Mozambique) by ELISA and nephelometry techniques. These findings will aid our understanding of the complex metabolic processes associated with TB progression from LTBI to active disease.

Tuberculosis (TB) is currently a global pandemic that kills approximately 1.3 million people worldwide each year¹. Although the incidence of TB is decreasing each year, the rate of that decline is not high enough to accomplish the TB Global Strategy for the 2015–2035 period^{2,3}. *Mycobacterium tuberculosis* (MTB) is the causal agent of the pulmonary TB⁴, which is air-transmitted by nasal/oral inhalation of aerosol droplets carrying the pathogen from an active TB patient to a healthy individual usually through coughing or sneezing⁵. Small droplets that are able to reach the lower lung induce formation of a granuloma⁶, which is a host-defensive structure providing, a fibrotic physical barrier between the infected and the healthy neighboring tissues⁷. Latently infected individuals (LTBI) represent roughly a quarter of the global population⁸. Tuberculin Skin Test (TST) and Interferon- γ Release Assay (IGRA) are commonly used for diagnosis^{9,10} of LTBI individuals. Although they cannot infect a healthy person, MTB activation occurs in roughly 10% of cases due to recurrent infections or a compromised immune system in the host¹¹.

¹Spanish National Research Council (CSIC), Vigo, Pontevedra, Spain. ²Biomedical Research Centre (CINBIO), Galician Singular Center of Research, Galicia Sur Health Research Institute (IIS-GS), Campus Universitario de Vigo S/N, University of Vigo, 36310, Vigo, Pontevedra, Spain. ³Tuberculosis Unit, Infectious Diseases, Internal Medicine Service, Complejo Hospitalario Universitario de Pontevedra, Galicia Sur Health Research Institute (IIS-GS, Pontevedra, Spain. ⁴Mycobacterial Infections Study Group (GEIM) of the Spanish Society of Infectious Diseases and Clinical Microbiology (SEIMC), Madrid, Spain. ⁵St. George's, University of London, London, UK. ⁶Department of Technologic Platforms, Instituto Nacional de Saúde, Maputo Province, Maputo, Mozambique. ⁷Department of Microbiology, Faculty of Medicine, Eduardo Mondlane University, Maputo, Mozambique. ⁸Centro de Saúde da Machava II, Maputo Province, Maputo, Mozambique. *email: jmateos@iim.csic.es; mcarrera@iim.csic.es

Patient Type	Sex	Age (years; mean \pm SD)	Contact score (mean \pm SD)*	BCG vaccination	MTB culture
Active TB patient (n = 26)	Female (15.4%), male (84.6%)	41.3 \pm 13.9	n.a.	Yes: 15.4% (n = 4) No: 84.6% (n = 22)	Positive: 100% (n = 26) Negative: 0% (n = 0)
LTBI contact (n = 29)	Female (41.3%), male (58.7%)	47.4 \pm 14.7	10.9 \pm 2.6	Yes: 27.6% (n = 8) No: 72.4% (n = 21)	n.a.
Uninfected contact (n = 44)	Female (52.3%), male (47.7%)	40.0 \pm 15.2	9.6 \pm 2.3	Yes: 32.8% (n = 14) No: 68.2% (n = 30)	n.a.

Table 1. Demographic summary of the patients included in the EMI-TB shotgun study cohort (Pontevedra, Spain). *“Contact Score” was assigned to the contacts following this criteria: acid-fast bacilli (AFB) microscopy index: from 0 to 4; cavity X-rays: No = 0, Yes = 1; contact: No = 0, Yes = 1; hours of exposure/day: 0–3 hours: 1, 4–7 hours: 2, 8–11 hours: 3, \geq 12 hours: 4; type of exposure: Outdoors = 0.25, Different room: 1, As a bar: 2, As a class: 2.5, As an office: 3, As a room or car: 4; weeks in contact with Index Case: $<$ 12 weeks: 0, \geq 12 weeks: 1; sleeps in the same room: No = 0, Yes = 1; first-degree family relationship: No = 0, Yes = 1.

The safest and most cost-effective way to fight against infectious diseases is vaccination¹². TB and other poverty-related and neglected infectious diseases still lack effective vaccines to protect the entire population¹³. BCG (Bacillus Calmette–Guerin), the only licensed vaccine currently in use against TB, was developed eighty years ago and displays an efficacy of 80% or more in children under 4 years old¹⁴. However, the efficacy of BCG in adolescents and adults is variable, 0 to 80% depending on the geographical region¹⁵. The TB Global Strategy for the period of 2015–2035 draws special attention to the urgent need to develop a new, more efficient vaccine against TB^{16,17}. Efforts are being made in different directions. While some research strategies have focused on boosting the existing BCG vaccine¹⁸ or exploring the use of new BCG strains¹⁹, others are focused on developing a mucosal vaccine, as is the case with the Horizon2020 EMI-TB consortium^{20,21}.

Proteomics has been used in the last decade as a tool to globally analyze, at the protein level, cellular and whole-organism processes related to disease and its progression. Proteomic profiling allows the elucidation of connections between various cellular pathways and thus complements both the genomics and traditional biochemical approaches. Several studies have focused on the MTB proteome, all of which were recently reviewed²². Alternatively, for the study of the molecular basis of health and disease, biological fluids from patients and controls are a reliable source for the identification of protein markers²³. Thanks to the extensive and continuous progress of analytical proteomics and because of the emergence of Mass Spectrometry-based techniques that facilitate the identification, quantification and characterization of proteins, proteomics is expected to be, in the near future, the tool of choice for diagnosing patients and searching for therapeutic biomarkers^{24,25}.

In the present study our focus was on identifying the specific protein signatures in sera of active TB patients and their contacts, including latently infected (LTBI) and uninfected contacts.

Experimental Procedures

Patient selection and database management. Collection (discovery cohort) of samples from TB patients and their contacts (LTBI and uninfected) started in Pontevedra (Spain) on September 2015, within the framework of the EMI-TB project (Eliciting Mucosal Immunity to Tuberculosis) following previously established clinical criteria²⁶. Briefly, patients were diagnosed using the highest standards for clinical TB research, including tuberculin skin test (TST) and/or the Quantiferon-TB-Gold test (QFT) and again after 8–12 weeks after the last possible exposure to the index case if the first test was negative. Additionally, chest radiographies were done to exclude active TB patients from the LTBI group. Lowenstein-Jensen and Colletos culture were done for all TB-diagnosed patients. The study was approved by the Spanish National Bioethics committee (Project ID: 643558). Exclusion criteria were age $<$ 18 y, co-infection with the human immunodeficiency virus (HIV), any other immunosuppressive medical condition or concomitant use of immunosuppressive drugs. Patients with previous TB or LTBI were also excluded for the study. Serum samples were collected right after final diagnosis and before commencement of the anti-TB drug treatment. The datasets were analysed in blind fashion using codes and keeping the anonymity of the volunteers. Summary of the demographic data considered relevant to the study is shown in Table 1. Informed consent was obtained, and all patients and contacts received a detailed explanation of the project and its objectives, as well as the assurance of the confidentiality.

In parallel, and also under the framework of the EMI-TB project, samples from TB patients (verification cohort) and their contacts (LTBI and uninfected) were also collected in the Maputo Region (Mozambique) using the same standard operating procedures (SOPs). The study was approved by the Mozambican National Bioethics committee (IRB:00002657; ID: 298/CNBS/15).

Additionally, all the experimental procedures described in this section were performed in accordance with relevant guidelines and regulations (both national and the EU).

Sample collection and storage of quantitative shotgun proteomics samples. EDTA (ethylene diamine tetra acetic acid) blood samples (40 mL) from each participant were obtained by the TB unit and immediately transferred to the University of Vigo for processing. During the whole sample collection period the same SOP (Standard Operational Procedure), based on the recommendations of the Human Proteome Organization (HUPO), was applied. Once obtained, the serum samples were distributed into 1000 μ L aliquots (Protein LoBind Tube 1.5 mL, Eppendorf, Germany) and stored at -80 °C for later analysis.

TMT 1 Serum	SAMPLE	PO-20	PO-28	PO-38	PO-24	PO-27	PO-36	PO-21	PO-26	PO-44	Standard
	Group	TB	TB	TB	LTBI	LTBI	LTBI	Uninfected	Uninfected	Uninfected	Mix
	TMT-Label	126	127N	127C	128N	128C	129N	129C	130N	130C	131
	BCG vaccination	No	No	Yes	No	Yes	Yes	No	No	No	n.a.
TMT 2 Serum	SAMPLE	PO-16	PO-40	PO-51	PO-30	PO-41	PO-54	PO-34	PO-57	PO-65	Standard
	Group	TB	TB	TB	LTBI	LTBI	LTBI	Uninfected	Uninfected	Uninfected	Mix
	TMT-Label	126	127N	127C	128N	128C	129N	129C	130N	130C	131
	BCG vaccination	No	No	No	No	No	No	No	No	No	n.a.
TMT 3 Serum	SAMPLE	PO-32	PO-63	PO-88	PO-31	PO-42	PO-55	PO-68	PO-73	PO-72	Standard
	Group	TB	TB	TB	LTBI	LTBI	LTBI	Uninfected	Uninfected	Uninfected	Mix
	TMT-Label	126	127N	127C	128N	128C	129N	129C	130N	130C	131
	BCG vaccination	Yes	No	No	No	Yes	No	No	No	No	n.a.
TMT 4 Serum	SAMPLE	PO-76	PO-90	PO-53	PO-66	PO-101	PO-87	PO-75	PO-77	PO-79	Standard
	Group	TB	TB	TB	LTBI	LTBI	LTBI	Uninfected	Uninfected	Uninfected	Mix
	TMT-Label	126	127N	127C	128N	128C	129N	129C	130N	130C	131
	BCG vaccination	No	No	No	No	No	Yes	No	Yes	No	n.a.
TMT 5 Serum	SAMPLE	PO-102	PO-67	PO-29	PO-80	PO-92	PO-86	PO-82	PO-72	PO-78	Standard
	Group	TB	TB	TB	LTBI	LTBI	LTBI	Uninfected	Uninfected	Uninfected	Mix
	TMT-Label	126	127N	127C	128N	128C	129N	129C	130N	130C	131
	BCG vaccination	No	Yes	No	Yes	Yes	No	Yes	No	No	n.a.

Table 2. Individual serum samples used in the EMI-TB shotgun study.

Protein preparation for TMT labeling. Total protein in the serum samples was precipitated by the addition of six volumes of cold acetone and incubated overnight at -20°C . After centrifugation, the dried protein pellet was resuspended in 0.1 M TEAB buffer solution (Triethyl ammonium bicarbonate: Stock solution: 1 M). The protein concentration was measured by the bicinchoninic acid (BCA) assay. The protein integrity was confirmed in 1 μg aliquots by SDS-PAGE and silver staining. Samples with protein degradation were detected and discarded from the proteomic analysis.

Experimental design and statistical rationale. For the shotgun study, we analyzed individual serum samples rather than pools and later applied a robust statistical analysis of quantification ratios (Kruskal-Wallis). For this reason, we increased the number of TMT 10plex labels to five (Table 2). Three active TB patients (channels 126, 127N and 127C), three infected LTBI contacts (channels 128N, 128C and 129N), three uninfected contacts (channels 129C, 130N and 130C), making a total of 45 individual samples. One standard mix sample were included within each TMT experiment, prepared by mixing equal amounts of proteins for the nine samples included (channel 131).

TMT 10plex labeling. Samples were selected for the quantitative proteomics study according to their protein concentration and integrity. Labelling and fractionation of the protein samples were done as previously described²⁷. Briefly, 100 μg of each individual sample was resuspended in a final volume of 100 μL of 0.1 M TEAB buffer solution, reduced/alkylated and digested with trypsin for 16 h at 37°C . Labeling with TMT 10plex reagents (Thermo Fisher Scientific, San Jose, CA, US) was performed following the manufacturer's indications.

Peptide fractionation by high-pH reversed phase. Dried aliquots of 100 μg were reconstituted in 300 μL of trifluoroacetic acid (TFA) and the peptides were fractionated using the High-pH Reversed Phase Fractionation Kit (Thermo Fisher Scientific) following the manufacturer's instructions. The peptide concentration in the resulting eight fractions was determined using the Quantitative Colorimetric Peptide Assay (Thermo Fisher Scientific).

LC analysis and orbitrap-elite settings. One microgram of each fraction was injected and analyzed by LC-MS/MS using a Proxeon EASY-nLC II liquid chromatography system (Thermo Fisher Scientific) coupled to an LTQ-Orbitrap Elite (Thermo Fisher Scientific). Separation of the peptides was performed on an RP column (EASY-Spray column, 50 cm \times 75 μm ID, PepMap C18, 2 μm particles, 100 \AA pore size, Thermo Fisher Scientific) with a 10 mm precolumn (Accucore XL C18, Thermo Fisher Scientific) using 0.1% formic acid (mobile phase A) and 98% ACN with 0.1% formic acid (mobile phase B). A 240 min linear gradient from 5 to 35% B was applied at a flow rate of 300 nL per min. Ionization was performed in an NSI source using a spray voltage of 1.95 kV and a capillary temperature of 275°C . The peptides were analyzed in positive mode (1 μs scan; 400–1600 amu), followed by 10 data-dependent HCD (high-energy collision dissociation) MS/MS scans (1 μs scans), using a normalized collision energy of 38% at a resolution of 30,000 and an isolation width of 1.5 amu. Dynamic exclusion was enabled with a repeat count of 1, a repeat duration of 30 s, a duration of the exclusion of 80 s, and a relative exclusion width of 10 ppm. Unassigned charged ions were excluded from the analysis.

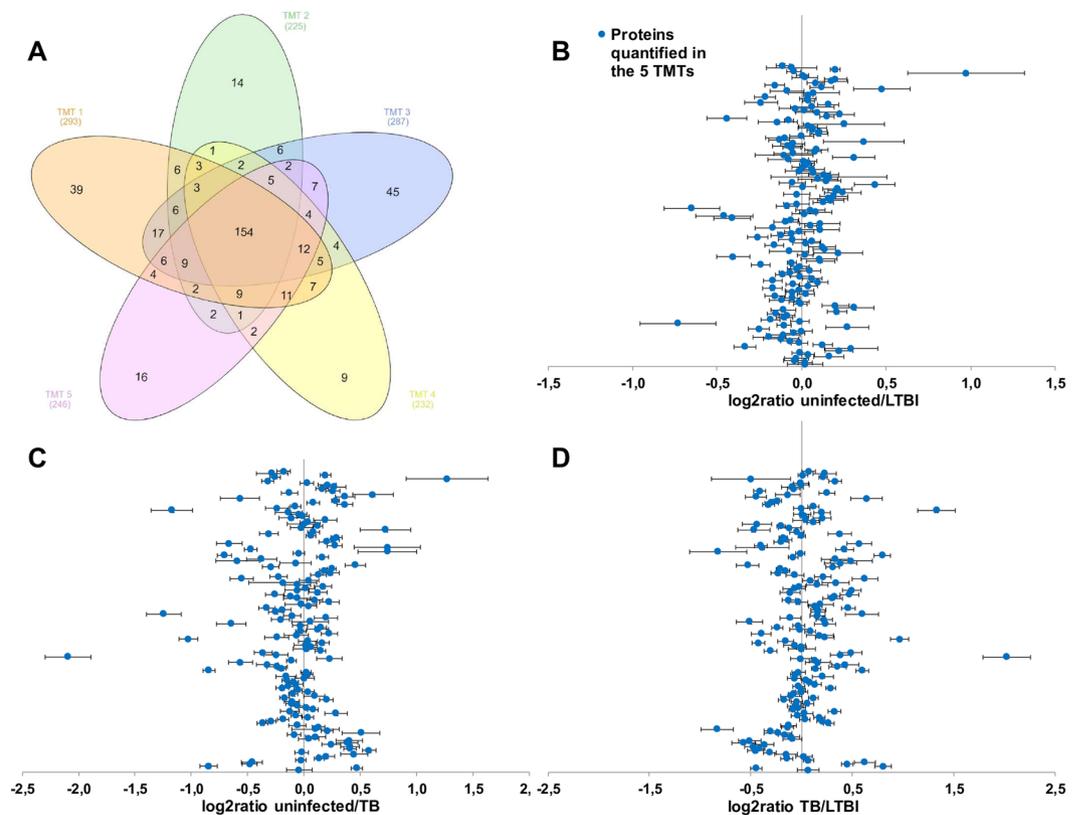


Figure 1. Summary of the quantitative shotgun proteomic study. Venn diagram representation of the five TMT experiments (A). A total of 418 proteins were identified and quantified with at least one unique peptide in the whole study. 154 of them were common for the five TMT experiments. Dispersion diagrams for the 154 common proteins (B–D) representing the mean of the log₂ ratio and the standard error of the mean (SEM).

Mass spectrometry data processing. Xcalibur 3.1 software (Thermo Fisher Scientific) was used for data acquisition, raw file generation, inspection of the chromatography profile and confirmation of the labeling of the peptides. After that, protein identification and quantification were performed using Proteome Discoverer 2.1 software (Thermo Fisher Scientific). Peak lists were generated with a precursor signal-to-noise ratio of 1.5, and default settings were used to search the latest Human UniProtKB/SwissProt Release (2018_10 with 551,681 entries) with the Sequest algorithm. The enzyme specificity was set to trypsin, and one missed cleavage was tolerated. TMT-labeling and carbamidomethylation of cysteines were set as fixed modifications, whereas oxidation of methionines and N-terminal acetylation were set as variable modifications. The precursor ion mass tolerance was set to 7 ppm, and the product ion mass tolerance was set to 0.06 Da. A decoy database search was performed to determine the peptide false discovery rate (FDR) with the Target Decoy PSM (peptide-spectrum matches) Validator Module. Quantification was performed using a Quantification Module, and normalization was performed against the total peptide amount. A 1% peptide FDR threshold was applied.

Samples were categorized by the patient type (TB, LTBI and uninfected contacts and internal standard). Quantification jobs were alternatively launched using a) the patient type option for the global analysis and b) the individual ratios option for the nonparametric statistical analysis.

Protein identification and quantification. For the Venn diagram representation, a list of proteins from each TMT experiment was quantified and at least one unique peptide was loaded in the InteractiVenn tool²⁸. For ratio dispersion representation, all the ratios for all the proteins quantified in the five TMT experiments were extracted from the Proteome Discoverer 2.1 software using the “export to Excel” option, and dispersion diagrams were performed using XLSTAT software.

Statistical analysis of the TMT quantification ratios. Only proteins identified and quantified in the five TMT experiments were considered for statistical analysis. Briefly, the protein ratios (45 for each comparison; 9 ratios per TMT experiment, 5 TMT experiments) were imported into R commander console (statistical package included in R 3.5.1 software) and represented in box diagrams. For each protein, data were used to analyze the differences between the three studied ratios (45 ratios uninfected/LTBI, 45 ratios uninfected/TB and 45 ratios TB/LTBI) by applying a Kruskal-Wallis test, since we observed that the quantification ratios did not follow a normal distribution. Differences in the modulation were considered significant when $p\text{-value} \leq 0.001$.

The final list (UniProt accession number) of significantly modulated proteins was analyzed using String 10.1 software for the determination of pathways and biological processes modulated in each group of participants.

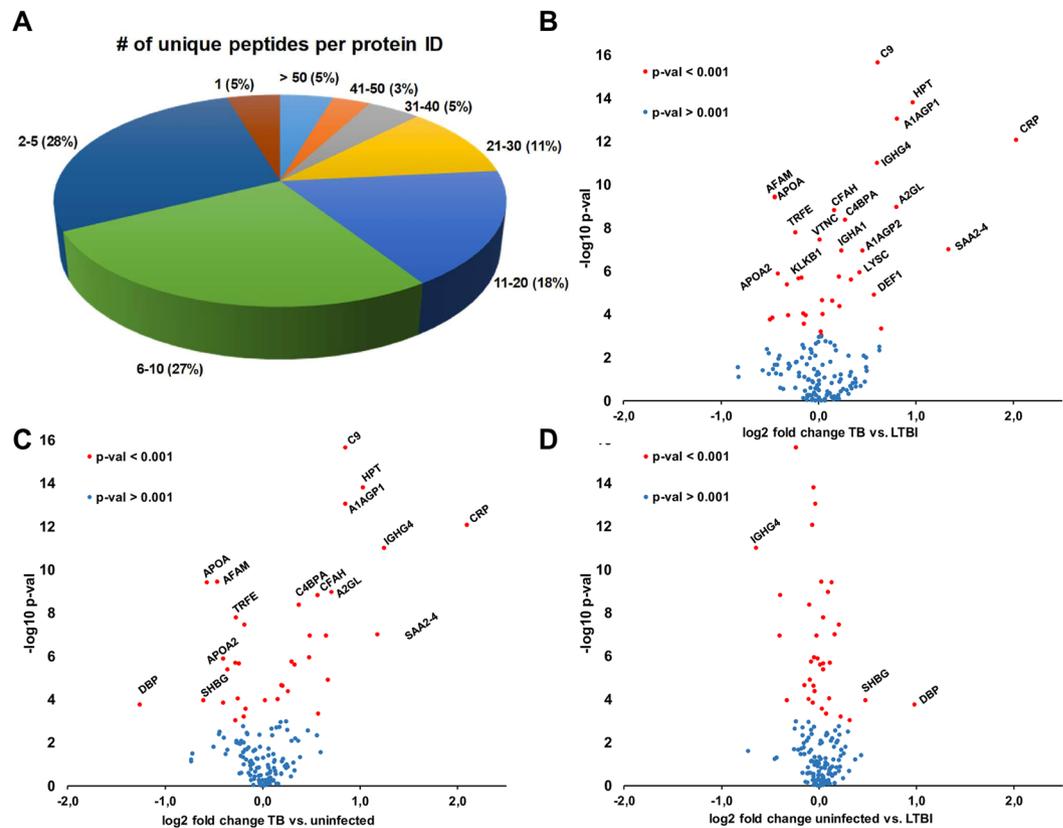


Figure 2. Summary of the number of unique peptides quantified per protein (A). Ninety-five percent of the proteins were quantified with at least two unique peptides. Volcano-plot representations of the statistical analysis of the quantification ratios. A specific proteomic signature is detected when comparing active TB patients versus both LTBI (B) and uninfected (C) contacts. No apparent specific signature is detected when comparing uninfected versus LTBI contacts (D). Significance is considered when $p\text{-value} \leq 0.001$.

ELISA of selected proteins. Serum levels of selected proteins were measured in the verification cohort using the following ELISA kits (Abcam): ab99995 (CRP), ab108852 (A1AGP1), ab219048 (HPT), ab108911 (TRFE) and ab202405 (KLKB1). Briefly, serum samples (twenty-one active TB patients and fifteen LTBI contacts) were diluted accordingly in the supplied sample diluent, and immunodetection was performed following the manufacturer's instructions. Absorbance was recorded in triplicate at 450 nm in a Multiskan GO Plate Reader (Thermo Fisher Scientific). Statistical analysis (Kruskal-Wallis) and diagrams were performed using R commander console and GraphPad software.

Nephelometry analysis of APOA1 in serum samples. APOA1 serum levels were measured in the verification cohort using an Immage™ nephelometer (Beckman-Coulter) and the corresponding APOA1 reagent kit (#4464110, Beckman Instruments, Inc.) following the manufacturer's instructions. Statistical analysis and diagrams were performed using R commander console and GraphPad software.

Receiver operating characteristic (ROC) analysis of the targeted analysis of selected proteins. The diagnostic accuracy of the six selected proteins were assessed by the receiver operator characteristics (ROC) curve analysis. The cut-off values for each parameter were determined by the highest Youden Index in order to maximize the sensitivity and specificity of the test. Statistical analysis was performed using GraphPad Prism version 6.0 for Windows (GraphPad Software; CA, USA) and IBM SPSS version 23 for windows (SPSS Inc., Chicago, Ill., USA).

Results

Protein identification and quantification. Proteomic datasets were deposited in the MassIVE repository (www.massive.ucsd.edu). Raw and processed files (SERUM_DATA_FILES_EMI_TB_PROTEOMICS; #MSV000083645) are public and freely accessible. Figure 1 represents the global identification results. The number of proteins identified in each TMT with at least one unique peptide range from 293 in TMT1 to 225 in TMT2 (Fig. 1A), whereas a total of 154 proteins were identified and quantified in the five TMT experiments with at least one unique peptide, 147 of them with at least two unique peptides (Fig. 2A and Supplementary Information). Dispersion diagrams of the quantification ratios show slight differences between uninfected versus LTBI participants (Fig. 1B) and more evident differences between active TB patients versus both uninfected and LTBI contacts (Fig. 1C,D).

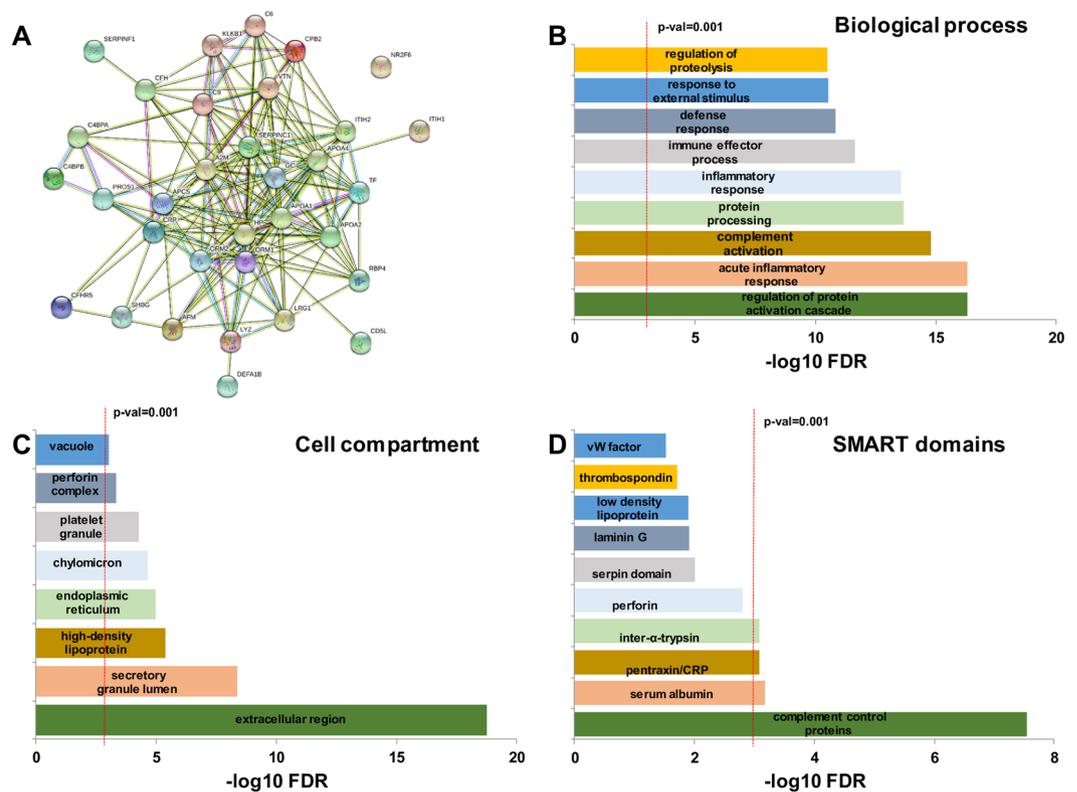


Figure 3. Interaction and pathway analysis of the proteins detected as modulated in active TB patients. String 10.1 analysis show a strong interaction network (A) between those proteins. Statistical pathway analysis (B–D) show that most of the proteins play roles in defense against pathogens, complement activation and inflammation.

Specific serum proteomic signature of active TB patients, LTBI contacts and uninfected contacts. A nonparametric Kruskal-Wallis statistical analysis was applied to the quantification ratios for all the proteins. Volcano-plot representations are shown in Fig. 2. Differences were detected between active TB patients and both uninfected and LTBI contacts (Fig. 2A,B), whereas no detectable differences were detected when comparing LTBI versus uninfected contacts.

Examples of proteins that accumulated in the serum of active TB patients include C-reactive protein (CRP), haptoglobin (HPT) alpha-1-acid glycoprotein 1 (A1AGP1), complement component C9 (C9), neutrophil defensin 1 (DEF1) and serum amyloid P component (SAA2–4). Volcano-plot representation of the quantification ratios showed no apparent differences in the serum protein content when comparing LTBI versus uninfected contacts.

In contrast, proteins such as apolipoprotein A (APOA1 and 2), serotransferrin (TRFE) and plasma kallikrein (KLKB1) were decreased in the serum of active TB patients (Fig. 2A,B). Similarly, the proteins that were diminished in the serum of active TB patients presented no differences when comparing LTBI versus uninfected contacts (Fig. 2).

Pathway and interaction analysis of modulated proteins. String 10.1 analysis showed a strong interaction network between the set of modulated proteins (Fig. 3A). Enrichment analyses by biological processes, cell compartment and SMART domains were performed (Fig. 3B–D), showing that most of the proteins modulated in active TB patients are secreted and play roles in complement activation, acute-phase response and modulation of inflammation. Box plot diagrams of the quantification ratios show that this set of proteins is specifically modulated in the serum of active TB patients, while no differences are detected when comparing LTBI and uninfected contacts (Figs. 4 and 5).

Targeted measurement of selected candidates in serum samples. ELISA and nephelometry techniques were used for verification in a second cohort due to the clinical relevance of identified proteins. A1AGP1, CRP and HPT were selected as proteins overrepresented in active TB patients versus both LTBI and uninfected contacts, whereas KLKB1, TRFE and APOA were selected as proteins diminished in active TB patients versus both LTBI and uninfected contacts. Statistical analysis showed similar results to those obtained for the six proteins in the shotgun proteomics phase (Fig. 6A). Significant differences ($p\text{-value} \leq 0.01$) were observed for A1AGP1, CRP, TRFE and APOA. Less significant differences ($p\text{-value} \leq 0.05$) were obtained for KLKB1, and no significant differences ($p\text{-value} = 0.08$) were observed for HPT.

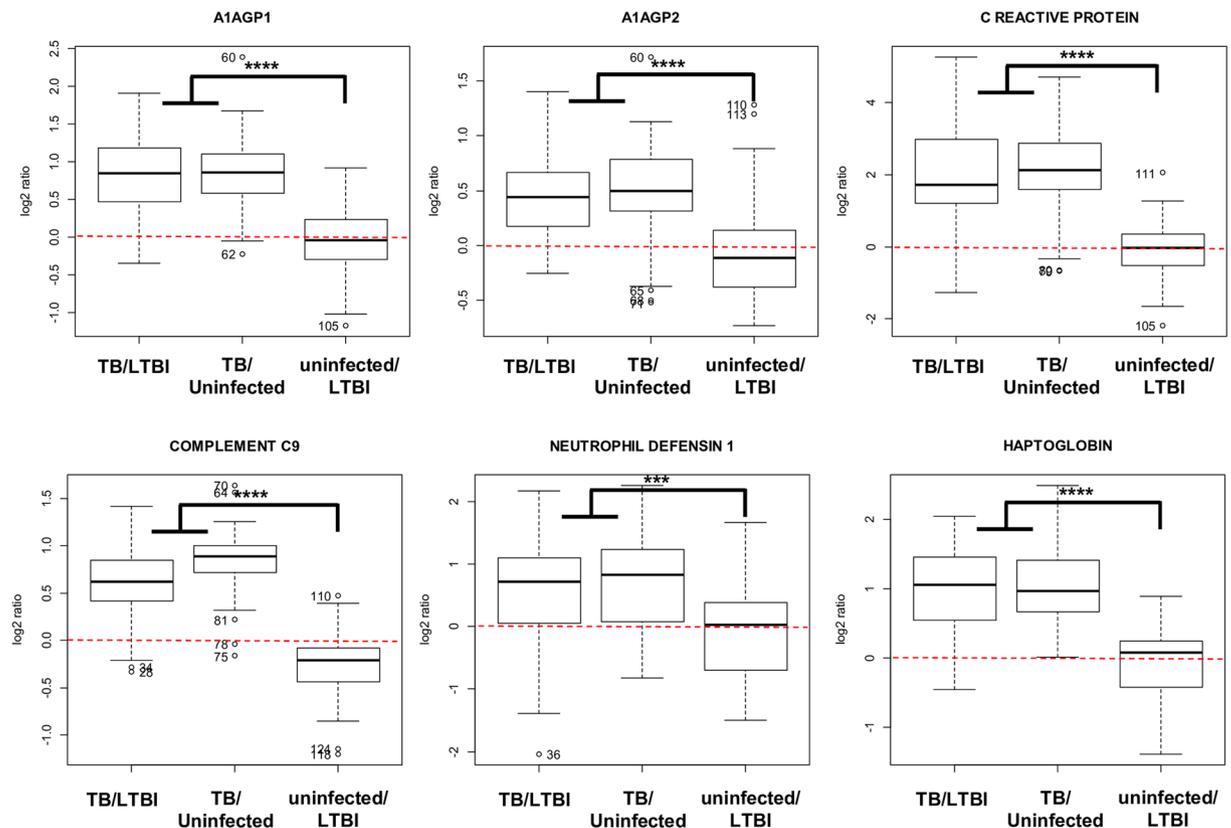


Figure 4. Box-plot representation of selected proteins that are detected as over-represented in the serum of active TB patients versus both LTBI and uninfected contacts but are not modulated when comparing uninfected versus LTBI contacts. ****p-value ≤ 0.00001 ; ***p-value ≤ 0.001 .

Receiver Operating Characteristic (ROC) analysis showed that five out of the six proteins present AUC (area under the curve) values higher than 0.7 (Fig. 6B). Two of them, CRP (AUC= 0.92) and TRFE (AUC= 0.83), presented the best balance between sensitivity and specificity, as revealed by Youden index analysis (Fig. 6C).

Discussion

The host innate and acquired immune processes activated during MTB infection and disease progression are still not fully understood. Biomarker discovery in biological fluids using high-resolution proteomics allows researchers to identify the modulation of cellular pathways when comparing different disease progression states. In the previous work, we focused on the analysis of nasal/oral fluid samples to obtain a deeper understanding of the innate immune processes implicated in resistance to the initial MTB infection²⁷. In the present study, by analyzing serum samples of active TB patients and their contacts, our aim was to focus on the mechanisms activated systematically in the host during the progression of the disease.

A total of 418 serum proteins were quantified in our study. At first, it might seem that the number of total quantified proteins using our approach was not high, but this result is typical of non-depleted serum samples. Depletion of abundant proteins using commercially available antibody-coated LC columns or cartridges²⁹ as well as other approaches, such as those based on the use of combinatorial hexapeptide ligand libraries³⁰, enables a higher coverage of the serum and plasma proteome in mass spectrometry-based studies but at the cost of losing information from highly abundant proteins and protein complexes. In a previous multilaboratory experiment (Proteored Multicenter Experiment 6), several groups were able to identify more than one hundred different proteins in the “bound” fraction of an immune-depleted serum sample with a commercial column supposed to bind only the 20 most abundant serum proteins³¹. In the present study, we took the view that information coming from those abundant proteins is important and should be also analysed. In future studies, we also think that a larger coverage of the serum proteome could be obtained by increasing the level of peptide fractionation, using off-gel-based techniques³². However, the number of proteins quantified in the present study are in concordance with very recent serum-based studies. Using MALDI-TOF mass spectrometry (no tryptic digestion and no fragmentation) Zhang *et al.*³³ found 236 serum “protein peaks” differentially expressed between patients with Intestinal Tuberculosis (ITB) and Crohn’s Disease (CD). In another study by Arya *et al.*³⁴, 132 and 68 serum extracellular vesicle (EVs) proteins from active TB patients were identified when compare to non-tuberculosis and healthy patients. The number of identified proteins were higher (more than seven hundred with at least one unique peptide) in the study by Chen *et al.*³⁵ by analyzing immune-depleted serum samples. Finally, Liu *et al.*³⁶ were able to identify one thousand proteins in plasma samples using label-free quantification of severe and mild TB patients versus healthy individuals with no exposure to MTB.

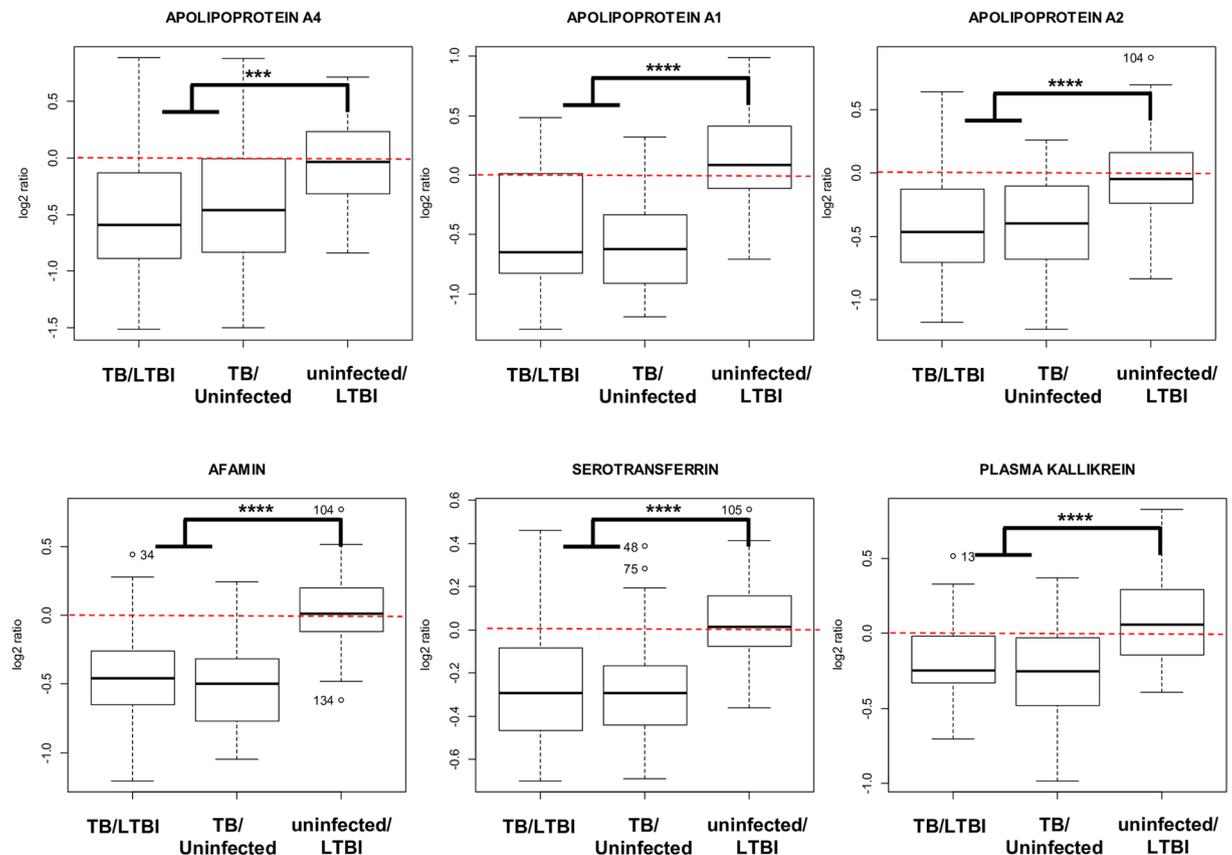


Figure 5. Box-plot representation of selected proteins that are detected as decreased in the serum of active TB patients versus both LTBI and uninfected contacts but are not modulated when comparing uninfected versus LTBI contacts. ****p-value ≤ 0.00001 ; ***p-value ≤ 0.001 .

Regarding protein identification and quantification, we obtained good reproducibility when comparing the data from five different TMT experiments. A total of 154 proteins are common to the five biological replicates and 147 of them were quantified with at least two unique peptides (Fig. 2A and supplementary info). The standard error of the mean (SEM) of the ratios is low for most of them (Fig. 1B–D).

Kruskal–Wallis test was applied to detect significantly modulated proteins when comparing different progression states of the disease (Fig. 2B–D). Uninfected contacts remained free of MTB infection despite high degree exposure to an active TB patient (Table 1). We have recently identified some components of the innate immune response that specifically accumulate in the sputum of these uninfected individuals when compared with LTBI contacts²⁷, indicating the importance of nasal/oral secretions as a first barrier to fighting the initial entry of MTB. In the present study, using the same approach, we were unable to detect any serum protein significantly accumulated in uninfected contacts versus LTBI contacts (Fig. 2D). In contrast, a set of proteins were significantly accumulated in active TB patients versus both LTBI and uninfected contacts (Fig. 2B,C).

Interaction network and pathway analysis showed that the specific protein signature of active TB patients was characterized by an accumulation of proteins related to complement activation, inflammation and modulation of the immune response (Fig. 3). Individual box plot diagrams representing the three kind of ratios show the specificity for modulated proteins (Figs. 4 and 5). For instance, proteins, such as C-reactive protein, haptoglobin and alpha-1 acid glycoprotein-1 and -2 (ORM-1 and -2), were specifically accumulated in active TB patients, whereas no differences were detected when comparing LTBI and uninfected contacts (Fig. 4). All these proteins were previously identified as accumulated in serum samples of TB patients with antibody-based techniques^{37,38}, by label-free mass spectrometry³⁶, and more recently in saliva and sputum samples²⁷. To the best of our knowledge, other proteins, such as neutrophil defensin 1 and lysozyme C, have not been previously described as being accumulated in the serum of active TB patients.

Furthermore, a decrease in a small subset of proteins, including apolipoproteins A1, A2 and A4, serotransferrin and plasma kallikrein, was revealed as characteristic of active TB patients (Fig. 5). The ApoA complex is a central component of high density lipoproteins (HDL) and is one of the main components responsible for lipid transport from tissues to the liver³⁹. A decrease in ApoA1 was associated with HDL deficiency type 2 or hypercholesterolemia more than two decades ago⁴⁰; interestingly, a decrease in ApoA1 has been more recently reported to cause an impairment of the immune response against TB⁴¹. In relation to this finding, ApoA was detected to be increased during rifampicin treatment of active TB in humans⁴² and formed a complex with isoniazid during the treatment in a TB mouse model⁴³. At this point, it is important to remark that the serum samples used in the present study were collected before patients underwent anti-TB treatment, so the serum levels of ApoA1 are not

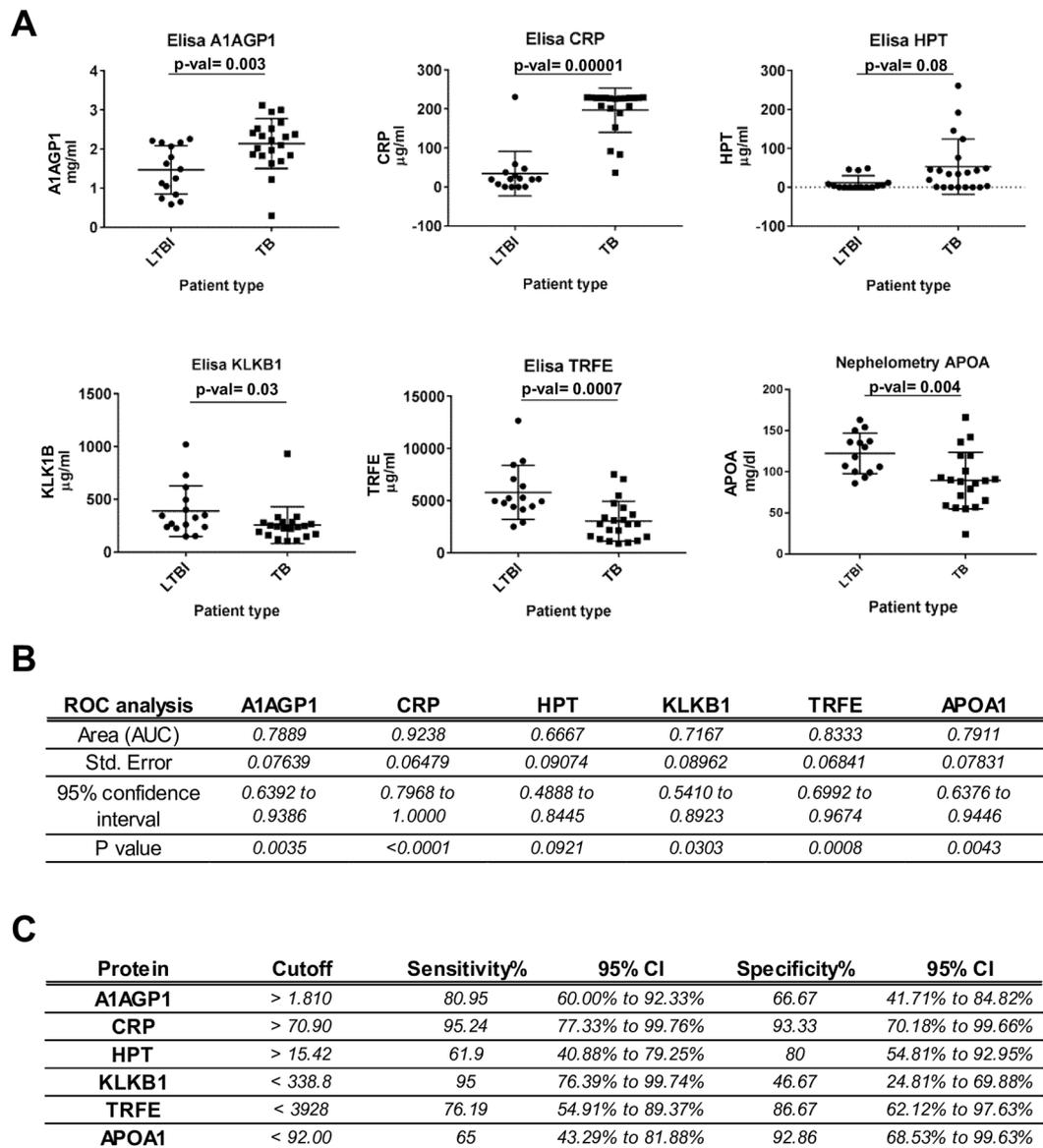


Figure 6. Targeted measurement of the serum levels of selected candidates (three over-represented and three decreased in active TB patients) in an independent cohort (verification cohort, Maputo region, Mozambique). Significant ($p\text{-val} \leq 0.05$) modulation is detected for five out of the six proteins (A). ROC (B) and Youden index (C) analysis was done for the six targets to investigate their discriminatory potential between active TB patients and latently infected contacts.

influenced by the uptake of anti-TB or any other drugs. This finding is in concordance with a previous quantitative proteomics study in which ApoA was found to be increased after anti-TB drug treatment in cured *versus* untreated tuberculosis patients⁴⁴, thus strongly indicating that rifampicin/isoniazid treatment increases the serum levels of this protein. A possibility is that LTBI contacts with decreased levels of ApoA1 could be more susceptible to developing active TB than those presenting normal levels of this protein (Fig. 7). Further investigations are needed to explore the putative implications of this result in the treatment of the disease.

In addition, iron assimilation of MTB from the host has been revealed as a key pathway for the progression of the disease⁴⁵. Specifically, MTB acquisition of iron from serotransferrin has been reported to be decreased by IFN-gamma treatment⁴⁶. Patients with hemochromatosis, also known as iron overload, have been previously associated with more severe TB disease⁴⁷. The decrease in serotransferrin levels as a host mechanism of defense against TB progression *via* iron deprivation represents an attractive possibility to be investigated in the future.

Plasma kallikreins are a group of serine proteases involved in inflammation and autoimmune response processes^{48,49}. KLK1 was previously described by genome-wide transcriptome profiling as downregulated in MTB-stimulated peripheral blood mononuclear cells (PBMCs) from patients with TB; thus, KLK1 was previously used in combination with other markers to discriminate active TB patients from LTBI individuals and healthy controls⁵⁰.

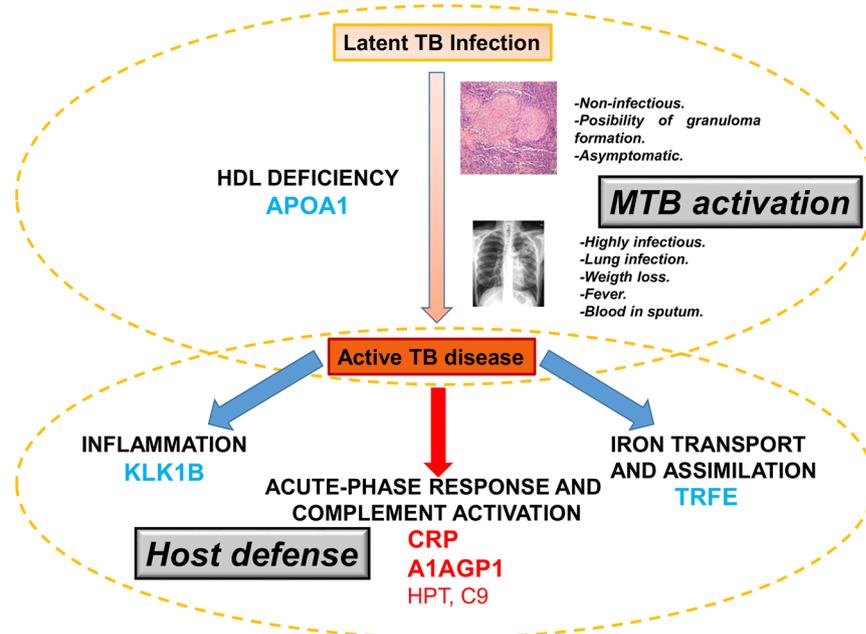


Figure 7. Schematic representation of the processes and proteins detected as modulated during the progression of the TB from a latent asymptomatic infection to an active disease. Proteins in blue are decreased in active TB patients whereas proteins in red are increased in this group versus both LTBI and uninfected contacts. Proteins in bold were also detected as significantly modulated in an independent cohort by targeted antibody-based techniques.

Verification of selected proteins in a second cohort was performed using antibody-based targeted techniques since these approaches are still considered the gold standard for clinical validation despite the promising development of multiplexing immunoassays⁵¹. The targeted measurement of the selected proteins was in agreement with the shotgun proteomic data, and five out of the six proteins were detected as significantly modulated in active TB patients versus LTBI contacts (Fig. 6A). ROC analysis showed that the serum levels of those five proteins could have a potential for discriminating between active TB and latent infected contacts. In particular, CRP and TRFE display a good balance between sensitivity and specificity (Fig. 6B,C and Supplementary Data 2). However, the putative discriminatory potential of those targets is still based on preliminary data and further investigation using bigger cohorts including other non-TB respiratory diseases are needed.

TB burden is especially dramatic in sub-Saharan countries⁵²; Mozambique, as the country of origin in the second cohort is from this region and, therefore, verification of the initial quantitative proteomics study in this second group of participants from a distant geographical region increases the strength of our findings and suggests a common host-acquired immune response to TB. Recently, mass spectrometry has been used to study the effect of phosphorylation in mycobacterial proteins known to be involved in MTB virulence⁵³. Although, MTB strain identification was not performed during patient diagnosis, it would be of interest to investigate whether or not different MTB strains possessing distinct pathogenicity traits promote similar defense mechanisms in the host.

We think that the findings presented in our study could have important implications for treating MTB and contribute to our understanding of the host defense mechanisms against TB. To summarize, lipid transport, iron assimilation acute-phase response and inflammation have all been identified as modulated pathways when comparing latent TB infection versus active TB disease (Fig. 7). We believe that these findings carry significant biological relevance and will help to better understand MTB pathogenesis and the host innate response, potentially aiding the rational vaccine design and testing for future studies.

Data availability

Raw and processed files (SERUM_DATA_FILES_EMI_TB_PROTEOMICS; #MSV000083645) and statistical analysis are publicly and freely accessible from the MassIVE Repository (www.massive.ucsd.edu).

Received: 23 July 2019; Accepted: 17 February 2020;

Published online: 02 March 2020

References

- Floyd, K., Glaziou, P., Zumla, A. & Raviglione, M. The global tuberculosis epidemic and progress in care, prevention, and research: an overview in year 3 of the End TB era. *The Lancet Respiratory Medicine* **6**, 299–314, [https://doi.org/10.1016/S2213-2600\(18\)30057-2](https://doi.org/10.1016/S2213-2600(18)30057-2) (2018).
- Uplekar, M. & Raviglione, M. WHO's End TB Strategy: From stopping to ending the global TB epidemic. *Indian J Tuberc* **62**, 196–199, <https://doi.org/10.1016/j.ijtb.2015.11.001> (2015).

3. Lytras, T. & Kalkouni, O. The global tuberculosis epidemic: turning political will into concrete action. *Journal of thoracic disease* **10**, S3149–S3152, <https://doi.org/10.21037/jtd.2018.08.73> (2018).
4. Sandgren, A. *et al.* Initiation and completion rates for latent tuberculosis infection treatment: a systematic review. *BMC Infect Dis* **16**, 204, <https://doi.org/10.1186/s12879-016-1550-y> (2016).
5. Fennelly, K. P. *et al.* Cough-generated aerosols of *Mycobacterium tuberculosis*: a new method to study infectiousness. *Am J Respir Crit Care Med* **169**, 604–609, <https://doi.org/10.1164/rccm.200308-1101OC> (2004).
6. Via, L. E. *et al.* Tuberculous granulomas are hypoxic in guinea pigs, rabbits, and nonhuman primates. *Infect Immun* **76**, 2333–2340, <https://doi.org/10.1128/IAI.01515-07> (2008).
7. Adams, D. O. The granulomatous inflammatory response. A review. *Am J Pathol* **84**, 164–192 (1976).
8. Houben, R. M. & Dodd, P. J. The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. *PLoS Med* **13**, e1002152, <https://doi.org/10.1371/journal.pmed.1002152> (2016).
9. Rao, M. *et al.* Latent TB Infection (LTBI) - *Mycobacterium tuberculosis* pathogenesis and the dynamics of the granuloma battleground. *Int J Infect Dis* **80S**, S58–S61, <https://doi.org/10.1016/j.ijid.2019.02.035> (2019).
10. Rana, A., Thakur, S., Kumar, G. & Akhter, Y. Recent Trends in System-Scale Integrative Approaches for Discovering Protective Antigens Against *Mycobacterial* Pathogens. *Front Genet* **9**, 572, <https://doi.org/10.3389/fgene.2018.00572> (2018).
11. Ahmad, S. Pathogenesis, immunology, and diagnosis of latent *Mycobacterium tuberculosis* infection. *Clin Dev Immunol* **2011**, 814943, <https://doi.org/10.1155/2011/814943> (2011).
12. Delany, I., Rappuoli, R. & De Gregorio, E. Vaccines for the 21st century. *EMBO Mol Med* **6**, 708–720, <https://doi.org/10.1002/emmm.201403876> (2014).
13. Bethony, J. M. *et al.* Vaccines to combat the neglected tropical diseases. *Immunol Rev* **239**, 237–270, <https://doi.org/10.1111/j.1600-065X.2010.00976.x> (2011).
14. Hawn, T. R. *et al.* Tuberculosis vaccines and prevention of infection. *Microbiol Mol Biol Rev* **78**, 650–671, <https://doi.org/10.1128/MMBR.00021-14> (2014).
15. Roy, A. *et al.* Effect of BCG vaccination against *Mycobacterium tuberculosis* infection in children: systematic review and meta-analysis. *BMJ* **349**, g4643, <https://doi.org/10.1136/bmj.g4643> (2014).
16. Zumla, A. *et al.* The WHO 2014 global tuberculosis report—further to go. *Lancet Glob Health* **3**, e10–12, [https://doi.org/10.1016/S2214-109X\(14\)70361-4](https://doi.org/10.1016/S2214-109X(14)70361-4) (2015).
17. Weber, W. P. *et al.* The association of preoperative anemia and perioperative allogeneic blood transfusion with the risk of surgical site infection. *Transfusion* **49**, 1964–1970, <https://doi.org/10.1111/j.1537-2995.2009.02204.x> (2009).
18. Penn-Nicholson, A. *et al.* Safety and immunogenicity of the novel tuberculosis vaccine ID93 + GLA-SE in BCG-vaccinated healthy adults in South Africa: a randomised, double-blind, placebo-controlled phase 1 trial. *Lancet Respir Med* **6**, 287–298, [https://doi.org/10.1016/S2213-2600\(18\)30077-8](https://doi.org/10.1016/S2213-2600(18)30077-8) (2018).
19. Nieuwenhuizen, N. E. & Kaufmann, S. H. E. Next-Generation Vaccines Based on Bacille Calmette-Guerin. *Front Immunol* **9**, 121, <https://doi.org/10.3389/fimmu.2018.00121> (2018).
20. Stylianou, E. *et al.* Mucosal delivery of antigen-coated nanoparticles to lungs confers protective immunity against tuberculosis infection in mice. *Eur J Immunol* **44**, 440–449, <https://doi.org/10.1002/eji.201343887> (2014).
21. Diogo, G. R. & Reljic, R. Development of a new tuberculosis vaccine: is there value in the mucosal approach? *Immunotherapy* **6**, 1001–1013, <https://doi.org/10.2217/imt.14.62> (2014).
22. Banaei-Esfahani, A., Nicod, C., Aebersold, R. & Collins, B. C. Systems proteomics approaches to study bacterial pathogens: application to *Mycobacterium tuberculosis*. *Curr Opin Microbiol* **39**, 64–72, <https://doi.org/10.1016/j.mib.2017.09.013> (2017).
23. Csoz, E. *et al.* Quantitative body fluid proteomics in medicine - A focus on minimal invasiveness. *J Proteomics* **153**, 30–43, <https://doi.org/10.1016/j.jprot.2016.08.009> (2017).
24. Legrain, P. *et al.* The human proteome project: current state and future direction. *Mol Cell Proteomics* **10**, M111 009993, <https://doi.org/10.1074/mcp.M111.009993> (2011).
25. Rai, A. J. *et al.* HUPO Plasma Proteome Project specimen collection and handling: towards the standardization of parameters for plasma proteome samples. *Proteomics* **5**, 3262–3277, <https://doi.org/10.1002/pmic.200401245> (2005).
26. González-Martín, J. *et al.* [Consensus document on the diagnosis, treatment and prevention of tuberculosis]. *Enferm Infecc Microbiol Clin* **28**, 297.e291–220, <https://doi.org/10.1016/j.eimc.2010.02.006> (2010).
27. Mateos, J. *et al.* High-resolution quantitative proteomics applied to the study of the specific protein signature in the sputum and saliva of active tuberculosis patients and their infected and uninfected contacts. *J Proteomics* **195**, 41–52, <https://doi.org/10.1016/j.jprot.2019.01.010> (2019).
28. Heberle, H., Meirelles, G. V., da Silva, F. R., Telles, G. P. & Minghim, R. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* **16**, 169, <https://doi.org/10.1186/s12859-015-0611-3> (2015).
29. Yadav, A. K. *et al.* A systematic analysis of eluted fraction of plasma post immunoaffinity depletion: implications in biomarker discovery. *PLoS One* **6**, e24442, <https://doi.org/10.1371/journal.pone.0024442> (2011).
30. Li, S. *et al.* Digging More Missing Proteins Using an Enrichment Approach with ProteoMiner. *J Proteome Res* **16**, 4330–4339, <https://doi.org/10.1021/acs.jproteome.7b00353> (2017).
31. Asanov, K. *et al.* Multi-Site Assessment of ProteoRed Plasma Reference Sample for Benchmarking LC-MS Platform Performance. *Journal of Biomolecular Techniques: JBT* **22**, S66–S67 (2011).
32. Franchin, C., Pivato, M., Rattazzi, M., Arrighini, G. & Millioni, R. OFFGEL fractionation of peptides: where really is your sample? *J Chromatogr A* **1355**, 278–283, <https://doi.org/10.1016/j.chroma.2014.05.077> (2014).
33. Zhang, F. *et al.* Exploration of Serum Proteomic Profiling and Diagnostic Model That Differentiate Crohn's Disease and Intestinal Tuberculosis. *PLoS One* **11**, e0167109, <https://doi.org/10.1371/journal.pone.0167109> (2016).
34. Arya, R. *et al.* Serum Small Extracellular Vesicles Proteome of Tuberculosis Patients Demonstrated Deregulated Immune Response. *Proteomics Clin Appl*, e1900062, <https://doi.org/10.1002/prca.201900062> (2019).
35. Chen, C., Yan, T., Liu, L., Wang, J. & Jin, Q. Identification of a Novel Serum Biomarker for Tuberculosis Infection in Chinese HIV Patients by iTRAQ-Based Quantitative Proteomics. *Front Microbiol* **9**, 330, <https://doi.org/10.3389/fmicb.2018.00330> (2018).
36. Liu, Q. *et al.* Proteomic profiling for plasma biomarkers of tuberculosis progression. *Mol Med Rep* **18**, 1551–1559, <https://doi.org/10.3892/mmr.2018.9134> (2018).
37. De Groote, M. A. *et al.* Discovery and Validation of a Six-Marker Serum Protein Signature for the Diagnosis of Active Pulmonary Tuberculosis. *J Clin Microbiol*. <https://doi.org/10.1128/JCM.00467-17> (2017).
38. Phalane, K. G. *et al.* Differential expression of host biomarkers in saliva and serum samples from individuals with suspected pulmonary tuberculosis. *Mediators Inflamm* **2013**, 981984, <https://doi.org/10.1155/2013/981984> (2013).
39. Pankhurst, G. *et al.* Characterization of specifically oxidized apolipoproteins in mildly oxidized high density lipoprotein. *J Lipid Res* **44**, 349–355, <https://doi.org/10.1194/jlr.M200256-JLR200> (2003).
40. Brooks-Wilson, A. *et al.* Mutations in ABC1 in Tangier disease and familial high-density lipoprotein deficiency. *Nat Genet* **22**, 336–345, <https://doi.org/10.1038/11905> (1999).
41. Martens, G. W. *et al.* Hypercholesterolemia impairs immunity to tuberculosis. *Infect Immun* **76**, 3464–3472, <https://doi.org/10.1128/IAI.00037-08> (2008).
42. Albanna, A. S., Bachmann, K., White, D., Valiquette, C. & Menzies, D. Serum lipids as biomarkers for therapeutic monitoring of latent tuberculosis infection. *Eur Respir J* **42**, 547–550, <https://doi.org/10.1183/09031936.00064713> (2013).

43. Sumenkova, D. V., Poliakov, L. M. & Panin, L. E. [Influence of isoniazid complex with A-I apolipoprotein on activity of lysosomal enzymes in mice with tuberculous inflammation model]. *Eksp Klin Farmakol* **75**, 28–30 (2012).
44. Wang, C. *et al.* Screening and identification of five serum proteins as novel potential biomarkers for cured pulmonary tuberculosis. *Sci Rep* **5**, 15615, <https://doi.org/10.1038/srep15615> (2015).
45. Banerjee, S., Farhana, A., Ehtesham, N. Z. & Hasnain, S. E. Iron acquisition, assimilation and regulation in mycobacteria. *Infect Genet Evol* **11**, 825–838, <https://doi.org/10.1016/j.meegid.2011.02.016> (2011).
46. Olakanmi, O., Schlesinger, L. S., Ahmed, A. & Britigan, B. E. Intraphagosomal Mycobacterium tuberculosis acquires iron from both extracellular transferrin and intracellular iron pools. Impact of interferon-gamma and hemochromatosis. *J Biol Chem* **277**, 49727–49734, <https://doi.org/10.1074/jbc.M209768200> (2002).
47. Abreu, R., Essler, L., Loy, A., Quinn, F. & Giri, P. Heparin inhibits intracellular Mycobacterium tuberculosis bacterial replication by reducing iron levels in human macrophages. *Sci Rep* **8**, 7296, <https://doi.org/10.1038/s41598-018-25480-y> (2018).
48. Sotiropoulou, G., Pampalakis, G. & Diamandis, E. P. Functional roles of human kallikrein-related peptidases. *J Biol Chem* **284**, 32989–32994, <https://doi.org/10.1074/jbc.R109.027946> (2009).
49. Liu, K. *et al.* Kallikrein genes are associated with lupus and glomerular basement membrane-specific antibody-induced nephritis in mice and humans. *J Clin Invest* **119**, 911–923, <https://doi.org/10.1172/JCI36728> (2009).
50. Pan, L. *et al.* Genome-wide transcriptional profiling identifies potential signatures in discriminating active tuberculosis from latent infection. *Oncotarget* **8**, 112907–112916, <https://doi.org/10.18632/oncotarget.22889> (2017).
51. Tighe, P. J., Ryder, R. R., Todd, I. & Fairclough, L. C. ELISA in the multiplex era: potentials and pitfalls. *Proteomics Clin Appl* **9**, 406–422, <https://doi.org/10.1002/prca.201400130> (2015).
52. Soares, N. C., Bou, G. & Blackburn, J. M. Editorial: Proteomics of Microbial Human Pathogens. *Front Microbiol* **7**, 1742, <https://doi.org/10.3389/fmicb.2016.01742> (2016).
53. Calder, B., Albeldas, C., Blackburn, J. M. & Soares, N. C. Mass Spectrometry Offers Insight into the Role of Ser/Thr/Tyr Phosphorylation in the Mycobacteria. *Front Microbiol* **7**, 141, <https://doi.org/10.3389/fmicb.2016.00141> (2016).

Acknowledgements

We are very grateful for the specialized personnel of “Complexo Hospitalario Universitario de Pontevedra” (SERGAS), specially Alberto Peña and Laura Calviño, for collecting and processing samples and for the patients and their contacts for participating in the present study and their altruistic donation of serum samples. The picture of the granuloma was kindly provided by Carlos Álvarez-Álvarez (SERGAS). We also thank Lelia Pedrós and Josefa San José-Vega (Hospital Meixoeiro, Vigo, Pontevedra, Spain) for nephelometry measurements of ApoA. Shotgun data acquisition was performed by Manuel Marcos (Proteomics Facility, University of Vigo, CACTI, Vigo, Spain). CSIC (M.C.), CINBIO (A.G.F.), EMU (T. M.) and SGUL (R.R.) groups are part of the Eliciting Mucosal Immunity in Tuberculosis (EMI-TB) consortium. EMI-TB is a European Union Horizon2020-funded action (Grant No. 643558) focused on selecting and developing a novel vaccine candidate for TB. We thank all the members of the EMI-TB, especially Elina Garet and Silvia Lorenzo-Abalde, for their collaboration and helpful discussions and suggestions.

Author contributions

J.M. and O.E. performed the experiments and analyzed data. L.A. and A.P., T.M., A.N. and C.G.M. managed the patient cohorts and collected/distributed serum samples. J.M. and M.C. designed the study, analyzed data, processed data, prepared figures and wrote the manuscript. M.C., R.R., L.A., A.G.F., I.M. and J.M.G. got funding and critically reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-60753-5>.

Correspondence and requests for materials should be addressed to J.M. or M.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020