Identification of novel salivary candidate protein biomarkers for tuberculosis diagnosis

Hygon Mutavhatsindi
Stellenbosch University
Bridget Calder
University of Cape Town
Shirley McAnda
University of Stellenbosch
Stephanus T. Malherbe
Stellenbosch University
Kim Stanley
Stellenbosch University
Martin Kidd
Stellenbosch University
Gerhard Walzl
Stellenbosch University
Novel N Chegou (novel@sun.ac.za)
Stellenbosch University https://orcid.org/0000-0002-8690-1699

Research article

Keywords: saliva, proteomics, tuberculosis, biomarkers, diagnosis

Posted Date: October 28th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-96812/v1

Abstract

Background: The diagnosis of TB remains one of the major challenges in the control of the disease, due to limitations in the currently available diagnostic tests. There is an urgent need for new, accurate, rapid, and affordable diagnostic tests. The aim of the present study was to use mass spectrometry to identify new candidate TB diagnostic protein biomarkers in saliva obtained from individuals with TB, and patients with other respiratory diseases (ORD).

Methods: Saliva samples were collected from 22 individuals who self-presented with symptoms requiring investigation for TB as part of a larger TB biomarker project. Eleven of the participants were finally diagnosed with TB using routine clinical, radiological and laboratory tests and 11 with ORD. Salivary proteins were concentrated and purified, followed by tryptic digestion. Peptides were analyzed using a QExactive Orbitrap MS coupled to a Dionex liquid chromatography system. Raw files were processed using MaxQuant software against the database of human proteins. Identified proteins were subjected to gene ontology and ingenuity pathway analysis for functional enrichment analysis.

Results: We identified 1176 protein groups across all samples, of which 46 (3.91%) were contaminants, 12 (1.02%) were reverse hits and 170 (14.46%) were single-peptide protein groups. After removal of the contaminants, reverse hits and protein groups represented by single peptides, 26 of the remaining 948 proteins significantly discriminated individuals with TB from those with ORD after Benjamini Hochberg correction, with five of these proteins diagnosing TB with AUC ³ 0.80. A 5-protein biosignature comprising of P01011, Q8NCW5, P28072, A0A2Q2TTZ9 and Q99574 diagnosed TB with an AUC of 1.00 (95% CI, 1.00-1.00), sensitivity of 100% (95% CI, 76.2-100%) and specificity of 90.9% (95% CI, 58.7-99.8%) after leave-one-out cross validation.

Conclusions: We identified novel salivary protein biomarkers and biosignatures with strong potential as TB diagnostic candidates. Our results are preliminary and require validation in larger studies.

Background

The global tuberculosis (TB) epidemic continues to be driven, in part, by undiagnosed TB cases or delays in disease diagnosis, leading to delays in initiation of treatment and increased transmission. The need for rapid and accurate tools to diagnose TB remains a priority for global disease control (1). Current diagnostic tools have well-documented limitations. The widely used Ziehl Neelsen smear microscopy test has low sensitivity, and the gold standard (sputum culture) is often unavailable in resource-limited settings, and has a long turnaround time (2, 3). With the roll-out of the automated gene amplification test XpertMtB/RIF (Cepheid Inc., Sunnyvale, USA), the diagnosis of TB has improved considerably, as the test significantly reduces detection time and is coupled with the identification of rifampicin resistance. A new version of the test (Xpert Ultra) has recently been introduced and is reported to have higher sensitivity than the first generation of the XpertMTB/RIF test. However, this test is both expensive and requires infrastructure that is not readily available in resource-limited settings, and is therefore not suitable for many of the areas that have the highest burden of TB disease (4). Currently, all available routine diagnostic tests that are suitable for use in all patient groups are sputum-based and require good quality sputum. These tests are not very useful in individuals who have difficulty in producing good quality sputum, including children who typically present with paucibacillary disease (5), and also individuals with extra-pulmonary TB. Immunodiagnostic techniques utilizing host biomarkers from easily obtainable biological fluids such as saliva could be valuable in such cases, especially if developed into rapid point-of-care tests (6, 7). An ideal diagnostic test would be one that is suitable for use in all patient groups (adults and children, regardless of HIV infection status, and people with extrapulmonary TB) and should be easily implementable in resource-poor settings.

Over the past decade, technological advances have enabled expansion of the use of whole-saliva in disease diagnosis (8). Saliva has a high protein concentration, typically 1–2 mg/mL, and consists mainly of glycoproteins, enzymes, immunoglobulins, and a wide range of antimicrobial activity-rich peptides (9–11). Saliva collection is simple, non-invasive, and does not present the same inconveniences or risks that apply to the collection of other biofluids, such as blood (12).

Proteomics technologies analyze the protein-level global activity of the cell and organism. Proteomic profiling enables the elucidation of the links between wide cellular pathways and individual molecules using bioinformatics analysis. Previous studies carried out in our research group (13–17) identified putative protein biomarkers for the diagnosis of TB, or as tools for monitoring the response to treatment. Furthermore, a recent study by Mateos and colleagues (18) identified protein signatures in saliva and sputum that distinguished active TB patients from their infected and uninfected contacts using labelled quantification. However, most of these studies were Luminex technology-based, and depended largely on targeted approaches relying on the presence of capture and detection antibodies or specific protein tags to measure each protein. Given the potential shown by previous studies, an unbiased approach has the potential to identify novel biomarkers which may be validated in future studies as tools for the diagnosis of TB. The purpose of the

present study was therefore to use an unbiased, label free proteomics approach to identify potential biomarker signatures in human saliva for the diagnosis of TB disease.

Methods

Study participants

Participants included in the current study were recruited from the Fisantekraal Community Clinic, in the outskirts of Cape Town, South Africa as part of the African European Tuberculosis Consortium (AE-TBC) study (15). The criteria for inclusion of participants in the study were cough persisting for more than 2 weeks and any one of the following: fever, recent weight-loss, night-sweats, hemoptysis, chest-pain or anorexia. Participants were eligible for the study if they were 18 years or older, willing to give written informed consent, including for HIV testing using a rapid test (Abbott, Germany), and sample storage. Exclusion criteria for the study included severe anemia (hemoglobin <10 g/L), pregnancy, other known chronic diseases such as diabetes mellitus, current anti-TB treatment, anti-TB treatment in the last 3 months, use of quinolone or aminoglycoside antibiotics in the past 2 months, or residency in the study area for less than 3 months. All study participants provided written informed consent, and were recruited between November 2010 and November 2012. Ethical approval for the study was granted by the Health Research Ethics Committee of Stellenbosch University (Reference: N10/08/274).

Sample collection and processing

Study participants fasted for at least one-hour prior to saliva collection. Briefly, participants were asked to chew a sterile cotton swab (Salivette saliva collection kit, Sarstedt, Numbrecht, Germany), for about 45 seconds. The cotton swab was then removed from the participant's mouth with sterile forceps, inserted into a sterile tube provided by the manufacturer, and then transported to the laboratory at 4–8°C. Upon arrival in the laboratory, the saliva samples were centrifuged at 1000 x *g* for 2 minutes and the supernatant was harvested and stored at -80°C until used. Specimen processing typically occurred within 2 hours of collection.

Diagnostic tests performed on study participants

Routine diagnostic tests, including mycobacterial culture, sputum smears and chest radiography were performed on all study participants as reported in previous studies (16,19). Sputum samples collected from all study participants were cultured using the Mycobacteria Growth Indicator Tube (MGIT) method (BD Biosciences, Franklin Lakes, NJ, USA), after which positive MGIT cultures were examined for acid fast bacilli using the Ziehl-Neelsen technique, followed by Capilia TB testing (TAUNS, Numazu, Japan), to confirm the isolation of organisms of the *Mycobacterium tuberculosis* (*M.tb*) complex, before being designated as positive cultures. Participants were classified as definite TB patients, probable TB patients, "questionable" or individuals with other respiratory diseases (ORD) using a combination of clinical, radiological, and laboratory findings, as previously described(16,19). Individuals with ORD had a range of other diagnoses including upper and lower respiratory tract bacterial or viral infections and acute exacerbations of chronic obstructive pulmonary disease or asthma. As reported in previous studies (15,16), the study did not attempt to identify the causative agents of diseases in the patients classified as ORD. Such activities would have necessitated various bacterial and viral cultures and different biochemical tests to be done to identify these organisms, which was beyond the scope of the main study.

Sample preparation for LC-MS/MS analysis

For this biomarker discovery study, 11 participants with confirmed TB were randomly selected and gender and age-matched with 11 participants with ORD that were recruited as part of the main prospective study. A summary of the protocol on sample preparation is shown in Figure 1. Briefly, proteins from saliva samples were precipitated using the methanol chloroform precipitation method as previously described (20) and quantified using a modified Bradford assay as previously described (21). Quantified samples were digested with trypsin and desalted using C18 spin columns (Thermo Fischer, Massachusetts, USA) prior to Liquid chromatography–mass spectrometry/mass spectrometry (LC MS/MS) analysis in order to remove any potential contaminants and to reconstitute the peptides in an MS-appropriate buffer. Samples were analyzed individually.

Identification of potential candidate protein biomarkers using LC–MS/MS

The peptides were separated using a Dionex 3000 Ultimate Ultra High Performance Liquid Chromatography (UHPLC) system running a gradient of solvent A [Water, 0.1% Formic Acid] and solvent B [Acetonitrile, 0.1% Formic Acid] from 6 to 40% solution B for 60 minutes at a flow rate of 0.4 µL/minute with a column oven temperature of 40°C. The peptides were first bound to a trap column (packed in-house, Luna C18; Phenomenex, California, USA, 2 cm) and washed in 2% solution B for 10 minutes at a flow rate of 5 µL/min. The trap column was then switched in line with the analytical column (packed in house, Aeris peptide 3.6 µm beads; Phenomenex, California, USA, 30cm),

for the gradient. Spectra were acquired in top 10 mode using a QExactive hybrid Orbitrap instrument (Thermo Fischer, Massachusetts, USA) with MS1 scans acquired at a resolution of 70,000 and an AGC target of 3 x 10^6 or an injection time of 250 milliseconds. MS/MS scans were acquired at a resolution of 17,500 and an AGC target of 1x 10^5. A dynamic exclusion window of 30 seconds was applied. Resulting raw files were analyzed using MaxQuant version 1.3.5.12. Default settings for label free quantitative (LFQ) intensity were applied and the database used was acquired from Uniprot in 2019 for *Homo sapiens*.

Pathway analysis

To better understand the pathways in which the identified proteins are involved, a list of differentially expressed proteins (P-value < 0.05) was compiled. Information from the Gene ontology (GO) database and PANTHER (Protein Analysis THrough Evolutionary Relationships; http://www.pantherdb.org/) classification systems was used to categorize these proteins according to their (GO) descriptions. The proteins were categorized according to their molecular functions, biological processes and protein classes. The canonical pathways and protein-protein interactions were analyzed using Ingenuity Pathway Analysis (IPA) Software (IPA trial version, Ingenuity Systems, http://www.ingenuity.com) (Qiagen). The predicted protein-protein interaction networks and canonical pathways were generated using inputs of gene identifiers, log10 ratio and p-values between TB and ORD group comparisons.

Statistical analysis

Differences in the label free quantitative (LFQ) intensities of the identified protein groups between TB patients and individuals with ORD were analysed using the t-test. The diagnostic abilities of the significantly different majority protein groups were assessed by receiver operator characteristics (ROC) curve analysis. Optimal cut off values and associated sensitivity and specificity were determined based on the Youden's Index (22). The predictive abilities of combinations of protein groups were investigated by general discriminant analysis (GDA), with leave-one-out cross validation. A P-value of < 0.05 was set as the significance threshold following Benjamini Hochberg multiple testing correction. The data was analysed using Statistica (TIBCO Software Inc., CA, USA) and GraphPad Prism version 8 (GraphPad Software Inc., CA, USA).

Results

Of the 22 study participants randomly selected from the main study's database, 11 were culture positive TB patients. All study participants (including those with ORD) were QuantiFERON TB Gold In Tube positive and HIV negative. The mean age of all study participants was 40.3 ± 8.9. The clinical and demographic characteristics of study participants are shown in Table 1.

Protein identification by shotgun proteomics

Representative total ion chromatograms (TICs) were obtained by pooling samples from the TB and ORD groups (Figure 2). These pooled samples were used as quality control (QC) references for the remainder of the individual clinical runs. The raw data for these references were processed according to Tyanova and colleagues (23). We identified 1176 protein groups across all samples, of these, 46 (3.91%) were contaminants, 12 (1.02%) were reverse hits and 170 (14.46%) were single peptide protein groups. The contaminants, reverse hits and protein groups represented by single peptides were cleaved and not included in further analysis. Amongst the remaining 948 proteins, 26 of them had intensities that were significantly different between the TB and ORD groups (Table 2). These differentially expressed proteins were subjected for GO and IPA.

Gene Ontology of the differentially expressed proteins

To get an overall picture of the differentially expressed salivary proteome in the two groups, the differentially expressed proteins were subjected to PANTHER classification and GO database to categorize them according to their molecular functions, biological processes and protein classes (Figure 3). The largest fraction of the biological process ontology of the identified proteins was the cellular process, followed by biological regulation and metabolic process (Figure 3A). Binding, catalytic activity and molecular function regulator were identified as the main function of the proteins followed by structural molecule activity, molecular transducer activity and transcription regulator activity by molecular function ontology (Figure 3B). Additionally, protein class ontology showed that the majority of these differentially expressed proteins belonged to the enzyme modulator protein class followed by oxidoreductase, cytoskeletal proteins and hydrolase classes (Figure 3C). The remaining smaller percentages of the differentially expressed proteins fall in the calcium-binding protein, cell adhesion molecule, defense/immunity protein, lyase, receptor, signaling molecule, surfactant and transferase protein classes.

Pathway analysis by IPA software

The protein-protein relationships and putative networks pathways analysis of differentially expressed proteins were performed by IPA (Ingenuity Inc.). The selected proteins (Table 2) were uploaded to the IPA software with their corresponding UniProt IDs and respective log10 ratios to map proteins into key pathways and to retrieve protein-protein interactions. We identified association of 26 canonical pathways with the differentially expressed proteins. Amongst these, IL-8 signaling pathway, which involves IL-8, a chemokine associated with inflammation which plays an essential role in neutrophil recruitment and neutrophil degranulation, appears as the top hit (Figure 4A). IL-8 induces chemotaxis mainly in neutrophils as well as other granulocytes, resulting in the migration of these target cells to the site of inflammation (24). As shown in Figure 4A, we further identified several important pathways, mainly signaling pathways in which these differentially expressed proteins were involved in.

Using the IPA software, we also explored protein-protein interaction networks. The protein-protein interaction analysis showed significant interactions among the differentially expressed proteins. We identified two interaction networks, with the major network including 18 of the 26 differentially expressed proteins. Of the 18 proteins in the protein-protein interaction network, only two were upregulated in individuals TB in comparison to those with ORD (Figure 4B). Within this interaction network, we identified multiple central nodes, namely PLG, IL1B, TNF, P38 MAPK, ERK 1/2, HGF, ITGAM and ITGB2. These nodes were located in different cell compartments, with the majority of them localized in the cytoplasm and extracellular space (Figure 4B). Of these multiple central nodes identified, only PLG, ITGAM and ITGB2 were proteins identified by the proteomics analysis, and the rest appeared as additional proteins of this network.

Utility of individual proteins in the diagnosis of TB

After removal of contaminants, reverse hits, and proteins represented by single peptides, data were log10 transformed and the LFQ intensities of the remaining proteins between individuals with TB and those with ORD using unpaired t-test. Intensities of 26 proteins were significantly different between the two groups (Table 2).

When data for individual proteins were assessed by ROC curve analysis, the area under the ROC curve (AUC) was above ≥ 0.75 for 14 of the 26 proteins. Notably, the AUCs for 5 of these proteins, including macrophage-capping protein (P40121), plasminogen (P00747), profilin-1 (P07737), f-actin-capping protein subunit beta (P47756) and alpha-1-antichymotrypsin (P01011) were ≥ 0.80 (Figure 5, Table 2).

Performance of combinations of proteins in the diagnosis of TB

To investigate whether the ability of proteins to discriminate between TB and ORD groups would be enhanced when used in combinations, intensity data from the 26 proteins that were found to be significantly different between the TB and ORD groups were fitted into GDA models. With the number of combined variables restricted to a maximum five, optimal prediction of TB was found to be achieved with a combination of either four or five proteins (Figure 6A).

The most accurate five-protein biosignature comprised of alpha-1-antichymotrypsin (P01011), NAD(P)H-hydrate epimerase (Q8NCW5), proteasome subunit beta type-6 (P28072), immunoglobulin kappa variable 1-33 (A0A2Q2TTZ9) and neuroserpin (Q99574). This 5-marker biosignature diagnosed TB disease with an AUC of 1.00 (95% CI, 1.00-1.00) (Figure 6C&D), corresponding to a sensitivity of 100% (95% CI, 76.2-100%) and a specificity of 100% (95% CI, 76.2-100%). After leave-one-out cross-validation, there was no change in the sensitivity (100%), however, the specificity dropped to 90.9% (95% CI, 58.7-99.8%). The negative predictive value (NPV) obtained for the 5-marker model after leave-one-out cross-validation was 100% whereas the positive predictive value (PPV) was 90.91% (95% CI, 60.49-98.49%) (Table 3).

The most accurate four-protein combination identified by best subsets GDA comprised of flavin reductase (NADPH) (P30043), myosin-9 (P35579), neuroserpin (Q99574) and protein S100-A11 (P31994) which accurately classified 90.9% of both the TB and ORD participants after leave-one-out cross-validation (Table 3).

The most frequently occurring proteins in the best 20 protein combinations that most accurately predicted TB disease or ORD were P30043, appearing in all the 20 protein combinations, Q99574 and P31949 appearing in 9 of the 20 protein combinations (Figure 6B).

Discussion

In the present study, we analysed the protein expression profiles in saliva from patients with TB, compared to what was obtained in patients suspected of having TB, but who were finally diagnosed with ORD using an unbiased protein biomarker discovery approach (QExactive mass spectrometry). We identified 26 proteins with TB diagnostic potential individually, but optimal diagnosis of TB was achieved when different proteins were combined. A 5-protein biosignature diagnosed TB with a sensitivity of 100% and a specificity of

90.9% after leave-one-out cross-validation, while a 4-protein biosignature ascertained TB disease with both sensitivity and specificity of 90.9%. Both biosignatures had positive and negative predictive values above 90%.

The analysis of the differentially expressed proteins by GO and PANTHER classification showed that the majority of these proteins are involved in the regulation of cellular processes and molecular functions, protein binding, catalytic activity, and an enzyme modulator protein class. Cellular process can be any process that is carried out at the cellular level. This includes cell communication that occurs within one cell or among more than one cell but occurs at the cellular level. This is interesting as the response to inflammation involves the combined functioning of many of the immune cells, as well as processes that occur within individual cells.

In order to get further insights of the 26 differentially expressed proteins, protein-protein interaction networks and predominant canonical pathways were generated by IPA software, which is based on a database consisting interactions between proteins selected from scientific literature that is regularly updated. The most significant canonical pathways based on p-value were IL-8 signaling and Actin cytoskeleton signaling pathways. Furthermore, the protein-protein interaction network predicted interaction between the differentially expressed proteins with pro-inflammatory cytokines, including IL-1b, TNF and IL-15. Pro-inflammatory cytokines play a significant role in inflammatory diseases of infectious origin (25). TNF has also been reported to trigger a cytokine cascade of the anti-inflammatory cytokines that block pro-inflammatory cytokine production, probably as regulatory mechanism to limit immunopathology, as well as cytokine inhibitors that block pro-inflammatory cytokine actions (25). This suggest that some of the differentially expressed proteins may play a role in the regulation of these inflammatory cytokines.

Despite TB being a treatable disease, the development of tools that are optimal in the diagnosis of the disease in all patient groups remains elusive. We identified five proteins with strong potential for use TB single biomarkers for the diagnosis of TB disease, including Macrophage-capping protein, Plasminogen, Profilin-1, F-actin-capping protein subunit beta and Alpha-1-antichymotrysin, which predicted TB disease individually with AUC > 0.80. Macrophage-capping protein, also known as actin regulatory protein CAP-G (CapG) in humans is encoded by the CAPG gene and is localized in the nucleus and cytoplasm (26). CapG was first identified as a member of the gelsolin superfamily of actin regulatory proteins, which modulates actin length by capping the barbed ends in a Ca^{2+}-dependent manner and generates propulsive force (27, 28). Increased expression of CapG plays a vital role in the regulation of cell motility in endothelial and fibroblastic cells, receptor-mediated membrane ruffling and phagocytosis, and increases metastasis in breast cancer and ovarian carcinoma (26, 29, 30). Its overexpression has not been investigated in individuals with TB. In the present study we found that salivary CapG levels were lower in individuals with TB compared to in those with ORD, a novel finding in the TB diagnostic field.

Plasminogen is a proenzyme of plasmin that is released from the liver into the systemic circulation (31). The conversion process of plasminogen to plasmin is essential to blood clot removal after healing of an injury and tissue remodeling (32, 33). Consequently, a wide range of pathogenic bacteria express virulence factors that seize the plasminogen/plasmin system in order to enhance tissue damage and to promote pathogen invasion (34, 35). Plasminogen plays an important role in tumor metastasis through degradation of the cell membrane and the extracellular matrix (36, 37). Patients with severe plasminogen deficiency suffer from a variety of difficult-to-treat inflammatory conditions, including gingivitis and ligneous conjunctivitis (38, 39). There is limited research on plasminogen and TB, however, *M.tb* possesses plasminogen binding and activating molecules (40). In the present study, the expression of salivary plasminogen was increased in individuals with ORD compared to those with TB. Interestingly, in our TB group, some individuals had a complete lack of plasminogen.

Profilin-1 is an actin-binding protein and was first identified more than 40 years ago in the calf thymus (41). Intermediate and low levels of Profilin-1 have been reported in some human solid cancers, including breast, pancreas, and liver carcinomas, while the overexpression of this protein can inhibit growth and multiplication as well as migration of cancerous cells, suggesting that it might be a tumor suppressor protein (42–44). Profilin-1 has been reported to interact with the 6 kDa *M.tb* produced early-secreted antigenic target (ESAT-6) and has been suggested that it may contribute to the mycobacterial virulence (45). Our findings however show that Profilin-1 is more expressed in the individuals suffering from ORD rather than those with TB.

F-actin-capping protein subunit beta (CapZβ) is a protein that in humans is encoded by the CAPZB gene, which is expressed in normal tissue cells such as lymphoid cells and seminiferous ducts. It is expressed in certain types of tumors including lymphoma and testicular cancer (46, 47). Some proteomics studies have identified differential expression of CapZβ in different forms of cancer (48–50). The functional roles and clinical impacts of CapZβ remains unknown in these tumors. There are no studies implicating CapZβ with *M.tb* and our data show that this protein is overexpressed in individuals with TB compared to those with ORD.

Alpha-1-antichymotrypsin, encoded by the gene SERPINA3(51), is a member of the serine protease inhibitor (serpin) family of acute phase proteins. Although antichemotrypsin is predominantly synthesized in the liver, it is also produced in the brain by astrocytes (52).

Circulating levels of acute phase proteins, transport, and clotting proteins are known to change in response to inflammatory stimuli, and these changes are essential for proteins to maintain host defenses (53). Our findings support this statement, as we show that the expression of Alpha-1-antichymotrypsin is heightened in individuals with TB.

It is interesting that the majority of the proteins identified in the present study have been extensively investigated in a range of cancers, so it is perhaps not surprising that they were found to be differentially expressed between individuals with TB and those with ORD as they may be responding to general inflammation caused by the nature of TB disease. However, in relation to the TB field, the findings about these proteins, particularly in saliva are novel and our data indicate that changes in the concentrations of these proteins in saliva may represent a novel approach for the diagnosis of active TB.

In general, the use of individual proteins as diagnostic tools for infectious diseases such as TB has not been seen favorably, because of the often poor specificity of single inflammatory biomarkers (54–56). Our findings are in agreement with this statement, as the use of individual proteins appears to either have a higher sensitivity and lower specificity or vice versa. As expected, combinations between different proteins performed much better than individual markers in the diagnosis of TB disease, with the optimal salivary 5-protein signature diagnosing TB with 100% sensitivity after leave-one-out cross validation. This is in agreement with previous TB diagnostic studies that have shown the superior accuracy of combinations of biomarkers compared to individual markers in the diagnosis of TB disease (16, 57). Agranoff et al poised that markers with low individual diagnostic specificities are capable of boosting diagnostic accuracies when used in combination (57). Biomarker combinations identified in our study are made up of combinations of both top-performing individual markers in combination with some markers with weak diagnostic potential individually, thus in line with this statement. This phenomenon was also observed in other diagnostic studies done on serum and plasma samples in our research group (15, 16).

The main limitation of the current study was the small sample size. The main reason for using such a relatively small sample size was because of the costs of experiments. Such low numbers are consistent with the numbers used in previous “omics”-based discovery studies. Despite this limitation, we identified protein-based biosignatures with potential as TB diagnostic candidates. Of note, this study was performed in individuals that presented with signs and symptoms requiring investigation for TB disease, prior to the establishment of a clinical diagnosis, and at a community level health care clinic in a high-burden setting, and not TB cases versus healthy controls, and all the individuals with ORD had latent TB infection, according to the QuantiFERON test. This implies that candidate biosignatures identified from this study have the potential to be replicated in further phase III diagnostic studies. Further studies are required to validate these new findings with larger numbers of participants presenting with signs and symptoms requiring investigation for TB disease, preferably from different geographical regions and using targeted proteomics where markers identified in the current studies can be robustly validated. As the present study was carried out in HIV uninfected individuals, the biosignatures identified in this study require further investigation also in HIV-infected individuals. The present proof-of-concept study only investigated adult pulmonary TB patients. The performance of the biosignatures also needs to be assessed in difficult to diagnose TB cases such as pediatric and extrapulmonary TB, smear and culture negative TB, and also in patients presenting with confirmed diseases that are similar to TB, including bronchiectasis and non-TB pneumonia, among others. Validated biosignatures could then be incorporated into a point-of-care triage test for TB, which may preferably be based on the lateral flow technology as recently demonstrated in a multi-center African study (58).

Conclusions

In conclusion, we have identified candidate host protein biomarkers in saliva as diagnostic candidates for TB disease. To the best of our knowledge, our study’s findings are a first in the TB field and require further validation in larger multi-site prospective studies. Due to the ease of collection of saliva and the fact that collection of such a specimen will be possible and easy in all patient groups including children, our findings require further validation, with the aim of eventually developing field-friend or point-of-care tests for TB diagnosis.

Abbreviations

TB : Tuberculosis

AETBC : African European Tuberculosis Consortium

MGIT : Mycobacteria Growth Indicator Tube

ORD : Other respiratory diseases

LC MS/MS : Liquid chromatography–mass spectrometry/mass spectrometry

GO : Gene ontology

UHPLC : Ultra High-Performance Liquid Chromatography

PANTHER : Protein Analysis THrough Evolutionary Relationships

IPA : Ingenuity Pathway Analysis

LFQ : Label free quantitative

ROC : Receiver operator characteristics

GDA : General discriminant analysis

TIC : Total ion chromatograms

QC : Quality control

AUC : Area under the ROC curve

NPV : Negative predictive value

PPV : Positive predictive value

ESAT-6 : Early-secreted antigenic target

SD : Standard deviation

CI : Confidence interval

BH : Benjamini Hochberg

FASP : Filter aided sample preparation

Declarations

Ethics approval and consent to participate

All study participants provided written informed consent prior to being recruited for the study, in their home language. Ethical approval for the study was granted by the Health Research Ethics Committee of Stellenbosch University (Reference: N10/08/274).

Approval for the study was granted by the Health Research Ethics Committee of Stellenbosch University (Reference: N10/08/274).

Consent for publication

Not applicable

Funding

This project was part of the EDCTP1 programme supported by the European Union (grant number IP_2009_32040-AE-TBC). HM was funded by the South African Medical Research council through its Division of Research Capacity Development under the Internship Scholarship Program. The content hereof is the sole responsibility of the authors and does not necessarily represent the official views of the SAMRC and EDCTP. Furthermore, the funders played no role in the study design, in the collection, analysis and interpretation of data; in the writing of the manuscript; and in the decision to submit the manuscript for publication.

Acknowledgments

We are grateful to all our study participants, and staff of the Stellenbosch University Immunology Research Group, who contributed to this study.

Competing Interests:

HM, GW and NC are listed as inventors on a provisional patent application based on the data presented in this study (South Africa Provisional Patent Application No: 2020/04135, Filing date: 2020/07/07).

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Author's contributions

HM, BC, GW and NC designing the study. STM, SM and GW recruited study participants, collected samples and/or were responsible for participant clinical classification. KS managed the database. HM and BC performed the experiments. HM, MK and NC analysed and interpreted the results. HM wrote the first draft of the manuscript. HM, BC, STM, GW and NC critically revised the manuscript. All authors read and approved the final manuscript.

References

1. WHO | The End TB Strategy [Internet]. WHO. World Health Organization; [cited 2020 Sep 15]. Available from: http://www.who.int/tb/strategy/end-tb/en/
2. Chegou NN, Hoek KG, Kriel M, Warren RM, Victor TC, Walzl G. Tuberculosis assays: past, present and future. Expert Rev Anti Infect Ther. 2011 Apr 1;9(4):457–69.
3. Desikan P. Sputum smear microscopy in tuberculosis: Is it still relevant? Indian J Med Res. 2013 Mar;137(3):442–4.
4. Piatek AS, Van Cleeff M, Alexander H, Coggin WL, Rehr M, Van Kampen S, et al. GeneXpert for TB diagnosis: planned and purposeful implementation. Glob Health Sci Pract. 2013 Mar 21;1(1):18–23.
5. Marais BJ, Pai M. New approaches and emerging technologies in the diagnosis of childhood tuberculosis. Paediatr Respir Rev. 2007 Jun;8(2):124–33.
6. Chegou NN, Walzl G, Bolliger CT, Diacon AH, van den Heuvel MM. Evaluation of adapted whole-blood interferon-gamma release assays for the diagnosis of pleural tuberculosis. Respir Int Rev Thorac Dis. 2008;76(2):131–8.
7. Munk ME, Arend SM, Brock I, Ottenhoff TH, Andersen P. Use of ESAT-6 and CFP-10 antigens for diagnosis of extrapulmonary tuberculosis. J Infect Dis. 2001 Jan 1;183(1):175–6.
8. Choo RE, Huestis MA. Oral fluid as a diagnostic tool. Clin Chem Lab Med CCLM. 2011;42(11):1273–1287.
9. Schipper RG, Silletti E, Vingerhoeds MH. Saliva as research material: biochemical, physicochemical and practical aspects. Arch Oral Biol. 2007 Dec;52(12):1114–35.
10. Zalewska A, Zwierz K, Zółkowski K, Gindzieński A. Structure and biosynthesis of human salivary mucins. Acta Biochim Pol. 2000;47(4):1067–79.
11. Humphrey SP, Williamson RT. A review of saliva: normal composition, flow, and function. J Prosthet Dent. 2001 Feb;85(2):162–9.
12. Gustafsson A, Ajeti V, Ljunggren L. Detection of suPAR in the Saliva of Healthy Young Adults: Comparison with Plasma Levels. Biomark Insights. 2011 Oct 25;6:119–25.
13. Jacobs R, Maasdorp E, Malherbe S, Loxton AG, Stanley K, van der Spuy G, et al. Diagnostic Potential of Novel Salivary Host Biomarkers as Candidates for the Immunological Diagnosis of Tuberculosis Disease and Monitoring of Tuberculosis Treatment Response. Pai M, editor. PLOS ONE. 2016 Aug 3;11(8):e0160546.
14. Jacobs R, Tshehla E, Malherbe S, Kriel M, Loxton AG, Stanley K, et al. Host biomarkers detected in saliva show promise as markers for the diagnosis of pulmonary tuberculosis disease and monitoring of the response to tuberculosis treatment. Cytokine. 2016 May;81:50–6.
15. Jacobs R, Malherbe S, Loxton AG, Stanley K, van der Spuy G, Walzl G, et al. Identification of novel host biomarkers in plasma as candidates for the immunodiagnosis of tuberculosis disease and monitoring of tuberculosis treatment response. Oncotarget [Internet]. 2016 Sep 6 [cited 2019 Feb 15];7(36). Available from: http://www.oncotarget.com/fulltext/11420
16. Chegou NN, Sutherland JS, Malherbe S, Crampin AC, Corstjens PLAM, Geluk A, et al. Diagnostic performance of a seven-marker serum protein biosignature for the diagnosis of active TB disease in African primary healthcare clinic attendees with signs and symptoms suggestive of TB. Thorax. 2016 Sep;71(9):785–94.

17. Phalane KG, Kriel M, Loxton AG, Menezes A, Stanley K, van der Spuy GD, et al. Differential Expression of Host Biomarkers in Saliva and Serum Samples from Individuals with Suspected Pulmonary Tuberculosis. Mediators Inflamm. 2013;2013:1–10.
18. Mateos J, Estévez O, González-Fernández Á, Anibarro L, Pallarés Á, Reljic R, 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. 2019 Mar 20;195:41–52.
19. Chegou NN, Sutherland JS, Namuganga A-R, Corstjens PL, Geluk A, Gebremichael G, et al. Africa-wide evaluation of host biomarkers in QuantiFERON supernatants for the diagnosis of pulmonary tuberculosis. Sci Rep. 2018 Feb 8;8(1):2675.
20. Fic E, Kedracka-Krok S, Jankowska U, Pirog A, Dziedzicka-Wasylewska M. Comparison of protein precipitation methods for various rat brain structures prior to proteomic analysis. ELECTROPHORESIS. 2010;31(21):3573–9.
21. Ramagli LS, Rodriguez LV. Quantitation of microgram amounts of protein in two-dimensional polyacrylamide gel electrophoresis sample buffer. ELECTROPHORESIS. 1985 Jan 1;6(11):559–63.
22. Fluss R, Faraggi D, Reiser B. Estimation of the Youden Index and its Associated Cutoff Point. Biom J. 2005;47(4):458–72.
23. Tyanova S, Temu T, Cox J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nat Protoc. 2016 Dec;11(12):2301–19.
24. Hedges JC, Singer CA, Gerthoffer WT. Mitogen-Activated Protein Kinases Regulate Cytokine Gene Expression in Human Airway Myocytes. Am J Respir Cell Mol Biol. 2000 Jul;23(1):86–94.
25. Srinivasan L, Harris MC, Kilpatrick LE. 128 - Cytokines and Inflammatory Response in the Fetus and Neonate. In: Polin RA, Abman SH, Rowitch DH, Benitz WE, Fox WW, editors. Fetal and Neonatal Physiology (Fifth Edition) [Internet]. Elsevier; 2017 [cited 2020 Jul 18]. p. 1241-1254.e4. Available from: http://www.sciencedirect.com/science/article/pii/B9780323352147001281
26. Pellieux C, Desgeorges A, Pigeon CH, Chambaz C, Yin H, Hayoz D, et al. Cap G, a Gelsolin Family Protein Modulating Protective Effects of Unidirectional Shear Stress. J Biol Chem. 2003 Aug 1;278(31):29136–44.
27. Silacci P, Mazzolai L, Gauci C, Stergiopulos N, Yin HL, Hayoz D. Gelsolin superfamily proteins: key regulators of cellular functions. Cell Mol Life Sci CMLS. 2004 Oct 1;61(19):2614–23.
28. Impe KV, Corte VD, Eichinger L, Bruyneel E, Mareel M, Vandekerckhove J, et al. The Nucleo-cytoplasmic Actin-binding Protein CapG Lacks a Nuclear Export Sequence Present in Structurally Related Proteins. J Biol Chem. 2003 May 16;278(20):17945–52.
29. Dahl E, Sadr-Nabavi A, Klopocki E, Betz B, Grube S, Kreutzfeld R, et al. Systematic identification and molecular characterization of genes differentially expressed in breast and ovarian cancer. J Pathol. 2005 Jan;205(1):21–8.
30. Yun D-P, Wang Y-Q, Meng D-L, Ji Y-Y, Chen J-X, Chen H-Y, et al. Actin-capping protein CapG is associated with prognosis, proliferation and metastasis in human glioma. Oncol Rep. 2018 Mar;39(3):1011–22.
31. Forsgren M, Råden B, Israelsson M, Larsson K, Hedén L-O. Molecular cloning and characterization of a full-length cDNA clone for human plasminogen. FEBS Lett. 1987;213(2):254–60.
32. Law RHP, Caradoc-Davies T, Cowieson N, Horvath AJ, Quek AJ, Encarnacao JA, et al. The X-ray Crystal Structure of Full-Length Human Plasminogen. Cell Rep. 2012 Mar 29;1(3):185–90.
33. Lijnen HR. Elements of the Fibrinolytic System. Ann N Y Acad Sci. 2001;936(1):226–36.
34. Sun H. The Interaction Between Pathogens and the Host Coagulation System. Physiology. 2006 Aug 1;21(4):281–8.
35. Waisman DM. Plasminogen: Structure, Activation, and Regulation. Springer Science & Business Media; 2012. 306 p.
36. Castellino FJ, Ploplis VA. Structure and function of the plasminogen/plasmin system. Thromb Haemost. 2005;93(4):647–54.
37. Rijken DC, Lijnen HR. New insights into the molecular mechanisms of the fibrinolytic system. J Thromb Haemost. 2009;7(1):4–13.
38. Mehta R, Shapiro AD. Plasminogen deficiency. Haemophilia. 2008;14(6):1261–8.
39. Schuster V, Hügle B, Tefs K. Plasminogen deficiency. J Thromb Haemost. 2007;5(12):2315–22.
40. Monroy V, Amador A, Ruiz B, Espinoza-Cueto P, Xolalpa W, Mancilla R, et al. Binding and Activation of Human Plasminogen by Mycobacterium tuberculosis. Infect Immun. 2000 Jul;68(7):4327–30.
41. Carlsson L, Nyström LE, Sundkvist I, Markey F, Lindberg U. Actin polymerizability is influenced by profilin, a low molecular weight protein in non-muscle cells. J Mol Biol. 1977 Sep 25;115(3):465–83.
42. Wu N, Zhang W, Yang Y, Liang Y-L, Wang L-Y, Jin J-W, et al. Profilin 1 obtained by proteomic analysis in all-trans retinoic acid-treated hepatocarcinoma cell lines is involved in inhibition of cell proliferation and migration. PROTEOMICS. 2006;6(22):6095–106.
43. Janke J, Schlüter K, Jandrig B, Theile M, Kölble K, Arnold W, et al. Suppression of Tumorigenicity in Breast Cancer Cells by the Microfilament Protein Profilin 1. J Exp Med. 2000 May 15;191(10):1675–86.

44. Wittenmayer N, Jandrig B, Rothkegel M, Schlüter K, Arnold W, Haensch W, et al. Tumor Suppressor Activity of Profilin Requires a Functional Actin Binding Site. Mol Biol Cell. 2004 Feb 6;15(4):1600–8.
45. Singh V, Kaur C, Chaudhary VK, Rao KVS, Chatterjee S. M. tuberculosis Secretory Protein ESAT-6 Induces Metabolic Flux Perturbations to Drive Foamy Macrophage Differentiation. Sci Rep [Internet]. 2015 Aug 7 [cited 2019 Jun 26];5. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5388048/
46. Park HW, Shin J-S, Kim C-W. Proteome of mesenchymal stem cells. Proteomics. 2007 Aug;7(16):2881–94.
47. Zhou Q, Chaerkady R, Shaw PG, Kensler TW, Pandey A, Davidson NE. Screening for therapeutic targets of vorinostat by SILAC-based proteomic analysis in human breast cancer cells. Proteomics. 2010 Mar;10(5):1029–39.
48. Bai SW, Herrera-Abreu MT, Rohn JL, Racine V, Tajadura V, Suryavanshi N, et al. Identification and characterization of a set of conserved and new regulators of cytoskeletal organization, cell morphology and migration. BMC Biol. 2011 Aug 11;9:54.
49. Shimada K, Uzawa K, Kato M, Endo Y, Shiiba M, Bukawa H, et al. Aberrant expression of RAB1A in human tongue cancer. Br J Cancer. 2005 May 23;92(10):1915–21.
50. Hong WX, Ye JB, Chen MT, Yan Y, Zhou GF, Yang XF, et al. Trichloroethylene induces biphasic concentration-dependent changes in cell proliferation and the expression of SET-associated proteins in human hepatic L-02 cells. Biomed Environ Sci BES. 2013 Jul;26(7):618–21.
51. Silverman GA, Bird PI, Carrell RW, Church FC, Coughlin PB, Gettins PGW, et al. The Serpins Are an Expanding Superfamily of Structurally Similar but Functionally Diverse Proteins EVOLUTION, MECHANISM OF INHIBITION, NOVEL FUNCTIONS, AND A REVISED NOMENCLATURE. J Biol Chem. 2001 Sep 7;276(36):33293–6.
52. Kalsheker NA. α1-antichymotrypsin. Int J Biochem Cell Biol. 1996 Sep 1;28(9):961–4.
53. Black S, Kushner I, Samols D. C-reactive Protein. J Biol Chem. 2004 Nov 19;279(47):48487–90.
54. Wang C, Li Y-Y, Li X, Wei L-L, Yang X-Y, Xu D-D, et al. Serum complement C4b, fibronectin, and prolidase are associated with the pathological changes of pulmonary tuberculosis. BMC Infect Dis. 2014 Jan 31;14(1):52.
55. Luo H, Zhu B, Gong L, Yang J, Jiang Y, Zhou X. The Value of Serum Prealbumin in the Diagnosis and Therapeutic Response of Tuberculosis: A Retrospective Study. PLOS ONE. 2013 Nov 19;8(11):e79940.
56. Chegou NN, Heyckendorf J, Walzl G, Lange C, Ruhwald M. Beyond the IFN-γ horizon: biomarkers for immunodiagnosis of infection with Mycobacterium tuberculosis. Eur Respir J. 2014 May 1;43(5):1472–86.
57. Agranoff D, Fernandez-Reyes D, Papadopoulos MC, Rojas SA, Herbster M, Loosemore A, et al. Identification of diagnostic markers for tuberculosis by proteomic fingerprinting of serum. The Lancet. 2006 Sep 16;368(9540):1012–21.
58. Corstjens PLAM, Tjon Kon Fat EM, de Dood CJ, van der Ploeg-van Schip JJ, Franken KLMC, Chegou NN, et al. Multi-center evaluation of a user-friendly lateral flow assay to determine IP-10 and CCL4 levels in blood of TB and non-TB cases in Africa. Clin Biochem. 2016 Jan;49(1–2):22–31.

Tables

Table 1: Clinical and demographic characteristics of study participants.

Number of participants	All (n=22)	TB (n=11)	ORD (n=11)
Males, n (%)	10 (45.5)	3 (27.3)	7 (63.6)
QuantiFERON Positive, n (%)	22 (100)	11 (100)	11 (100)
HIV negative, n (%)	22 (100)	11 (100)	11 (100)
Mean age, years ± SD	40.3 ± 8.9	40.3 ± 8.8	40.3 ± 9.3

Abbreviations: TB= tuberculosis, SD=standard deviation, ORD=other respiratory diseases

Table 2: Individual proteins that significantly differentiated participants with TB from those with ORD.

Uniprot Accession number	Protein Name	Gene name	AUC (95% CI)	Sensitivity	Specificity	t-test p-value	S. according to BH
P40121	Macrophage-capping protein	CAPG	0.83 (0.65-1.00)	90.9%	63.6%	0.02	yes
P00747	Plasminogen	PAG	0.82 (0.61-1.00)	81.8%	81.8%	<0.01	yes
P07737	Profilin-1	PFN1	0.82 (0.64-1.00)	63.6%	90.9%	0.01	yes
P47756	F-actin-capping protein subunit beta	CAPZB	0.81 (0.63-0.99)	81.8%	63.6%	0.01	yes
P01011	Alpha-1-antichymotrypsin	SERPINA3	0.81 (0.63-0.99)	72.7%	72.7%	0.01	yes
Q8NCW5	NAD(P)H-hydrate epimerase	NAXE	0.79 (0.59-0.99)	72.7%	72.7%	<0.01	yes
Q01518	Adenylyl cyclase-associated protein 1	CAP1	0.79 (0.58-0.99)	81.8%	72.7%	0.02	yes
P61158	Actin-related protein 3	ACTR3	0.79 (0.57-1.00)	81.8%	72.7%	0.04	yes
Q9Y6R7	IgGFc-binding protein	FCGBP	0.76 (0.54-0.98)	81.8%	72.7%	0.03	yes
P31949	Protein S100-A11	S100A11	0.76 (0.55-0.97)	81.8%	72.7%	0.03	yes
Q9HC84	Mucin-5B	MUC5B	0.75 (0.52-0.98)	90.9%	72.7%	0.03	yes
P29401	Transketolase	TKT	0.75 (0.53-0.98)	81.8%	72.7%	0.02	yes
P28072	Proteasome subunit beta type-6	PSMB6	0.75 (0.58-0.93)	54.5%	100.0%	0.02	yes
Q9ULZ3	Apoptosis-associated speck-like protein containing a CARD	PYCARD	0.75 (0.56-0.94)	63.6%	90.9%	0.03	yes
P62140	Serine/threonine-protein phosphatase PP1-beta catalytic subunit	PPP1CB	0.74 (0.53-0.95)	81.8%	63.6%	0.03	yes
P15153	Ras-related C3 botulinum toxin substrate 2	RAC2	0.74 (0.52-0.95)	81.8%	72.7%	0.01	yes
P20160	Azurocidin	AZU1	0.74 (0.49-0.98)	100.0%	63.6%	0.01	yes
A0A2Q2TTZ9	Immunoglobulin kappa variable 1-33	IGKV1D-33	0.74 (0.52-0.95)	63.6%	72.7%	0.03	yes

P35579	Myosin-9	MYH9	0.74 (0.52-0.95)	81.8%	63.6%	0.03	yes
O95881	Thioredoxin domain-containing protein 12	TXNDC12	0.73 (0.57-0.88)	45.5%	100.0%	0.01	yes
Q99574	Neuroserpin	SERPINI1	0.71 (0.51-0.91)	63.6%	81.8%	0.03	yes
P04899	Guanine nucleotide-binding protein G(i) subunit alpha-2	GNAI2	0.69 (0.44-0.94)	90.9%	63.6%	0.01	yes
B0YIW2	Apolipoprotein C-III	APOC3	0.68 (0.53-0.83)	36.4%	100.0%	0.03	yes
P05107	Integrin beta-2	ITGB2	0.66 (0.41-0.91)	63.6%	72.7%	0.03	yes
P30043	Flavin reductase (NADPH)	BLVRB	0.66 (0.40-0.92)	100.0%	54.5%	0.03	yes
P11215	Integrin alpha-M	ITGAM	0.62 (0.36-0.88)	100.0%	45.5%	0.01	yes

Abbreviations: AUC= area under the ROC curve, CI= confidence interval, S= significance BH= Benjamini Hochberg

Table 3: Utility of combinations of proteins in general discriminant analysis (GDA) models in discriminating between individuals with TB and ORD.

Protein Model	Resubsitution classification matrix		Leave-one-out cross validation		Positive predictive value	Negative predictive value
	TB %	ORD %	TB%	ORD%		
5-protein model P01011, Q8NCW5, P28072, Q99574, A0A2Q2TTZ9,	100 (11/11)	100 (11/11)	100 (11/11)	90.9(10/11)	100%	90.9% (95% CI, 60.5-98.5%)
4-protein model P30043, P35579, Q99574, P31949	100 (11/11)	100 (11/11)	90.9 (10/11)	90.9(10/11)	90.9% (95% CI, 60.5-98.5%)	90.9% (95% CI, 60.5-98.5%)

Figures

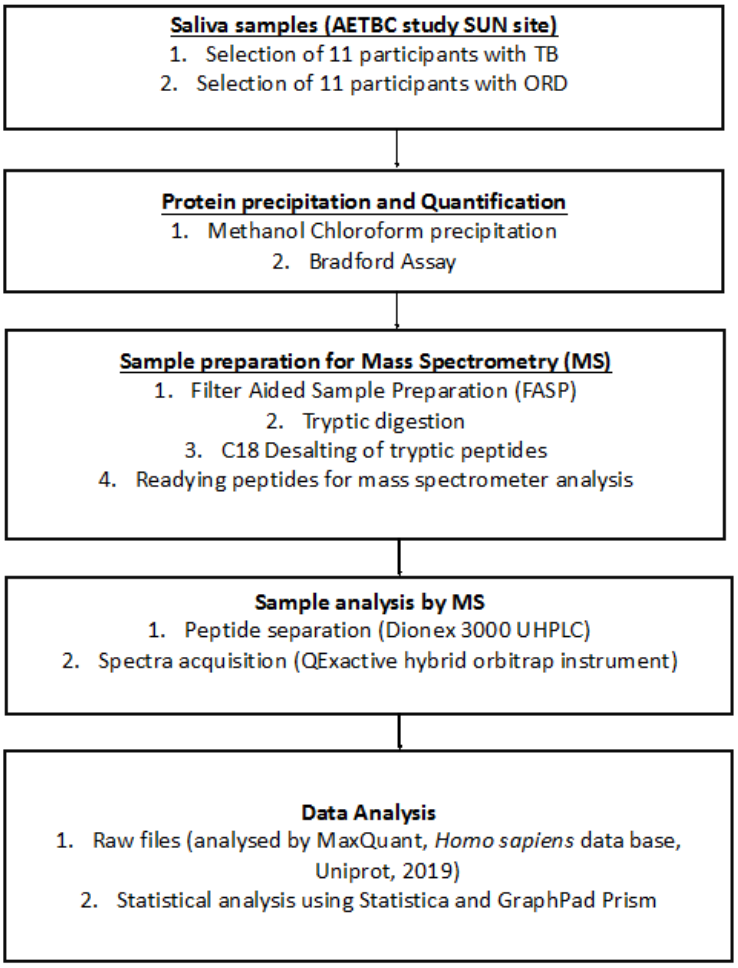

Figure 1

Schematic representation of the study design and workflow.

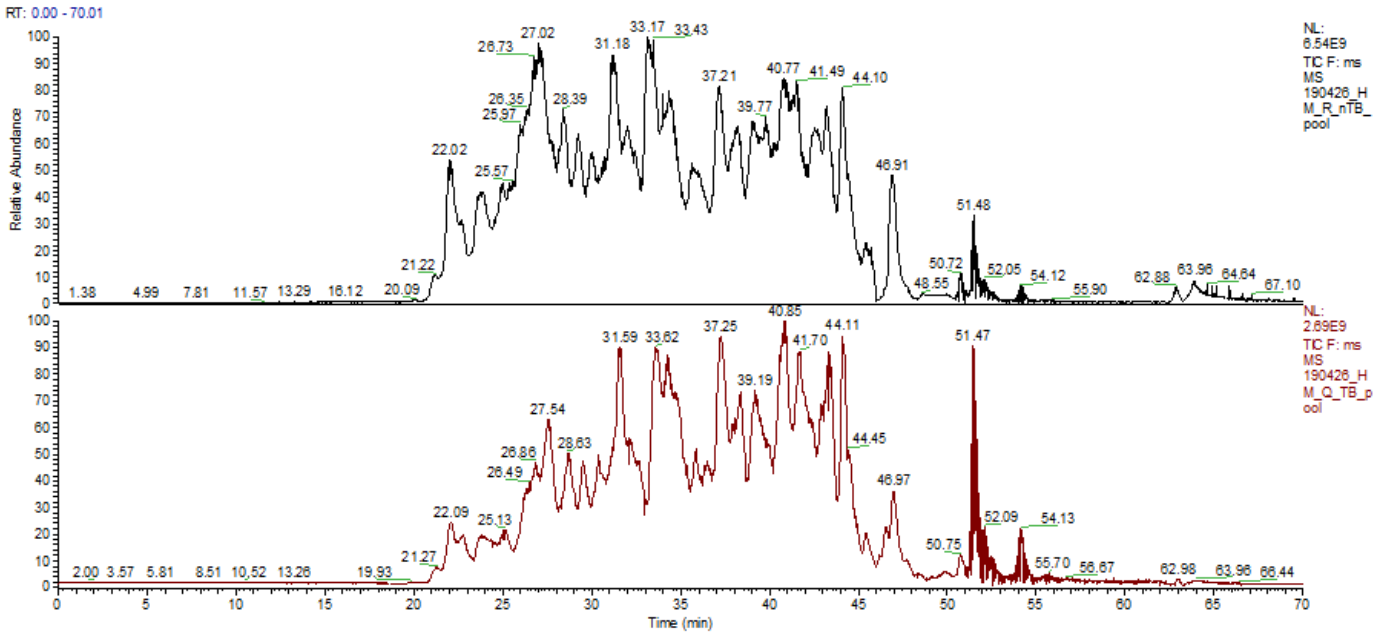

Figure 2

Representative total ion chromatograms of the TB and ORD groups. The ORD group is shown on the top panel chromatogram and the TB is on the bottom panel.

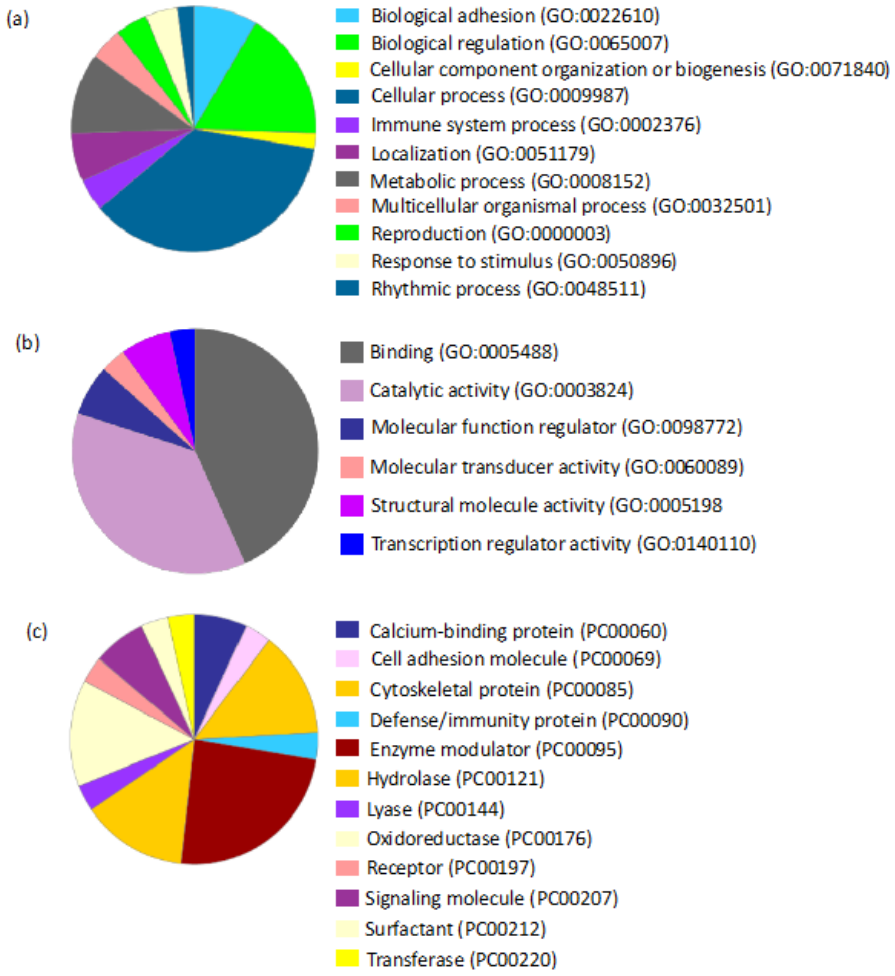

Figure 3

Gene Ontology analysis of differentially expressed proteins in the TB and ORD groups. Proteins were classified according to their PANTHER classification and GO description systems and they were analysed on the basis of biological process (a), molecular functions (b) and protein classes (c).

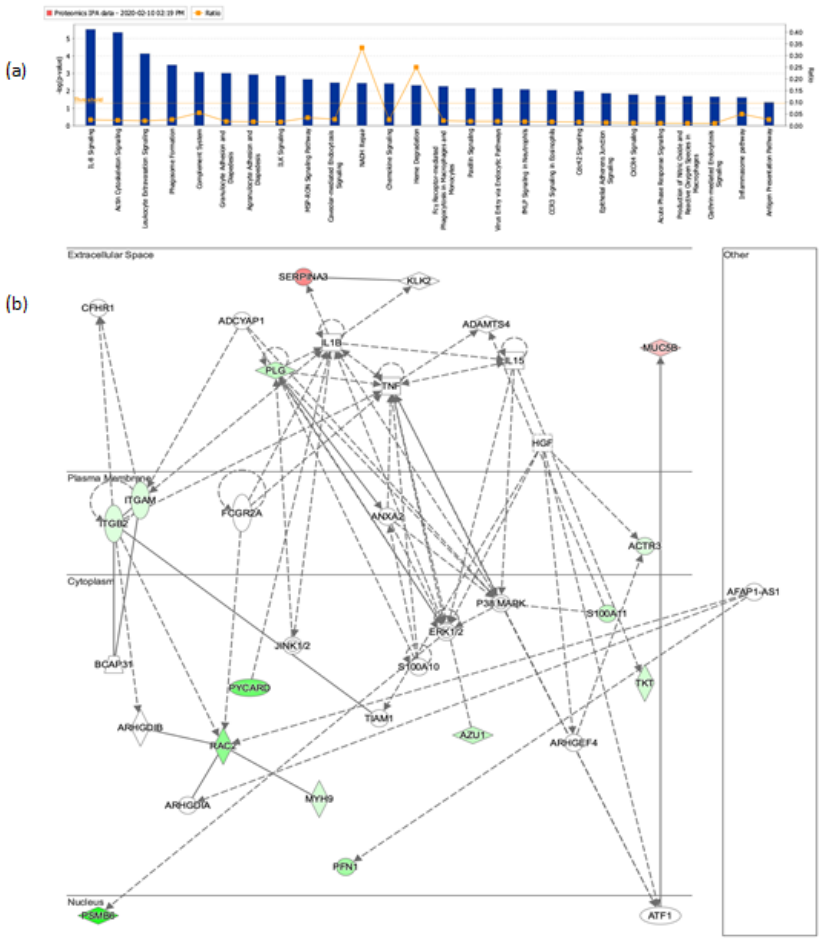

Figure 4

IPA analysis of proteins that were significantly different between the TB and ORD groups. Proteins that significantly differentiated the TB group from ORD group were subjected to IPA analysis. (a) The top 26 canonical pathways were identified as significantly by these selected proteins. (b) Protein-protein interaction by IPA analysis. IPA was also used to determine the protein-protein interactions among the differentially expressed proteins. Robust correlation with partner proteins are represented by solid lines, and less frequent but statistically significant correlations are indicated by dashed lines. Red color represents upregulated proteins in individuals with TB in comparison with individuals with ORD whereas the green color indicates downregulated proteins. Un-colored nodes indicate additional proteins of involved in the network that were not identified during the proteomics analysis. Arrows indicate the protein-protein. The shapes of the nodes indicate protein function: oval (transcription regulators); diamond (enzymes); rectangle (nuclear receptors); square (cytokines); and circle (others).

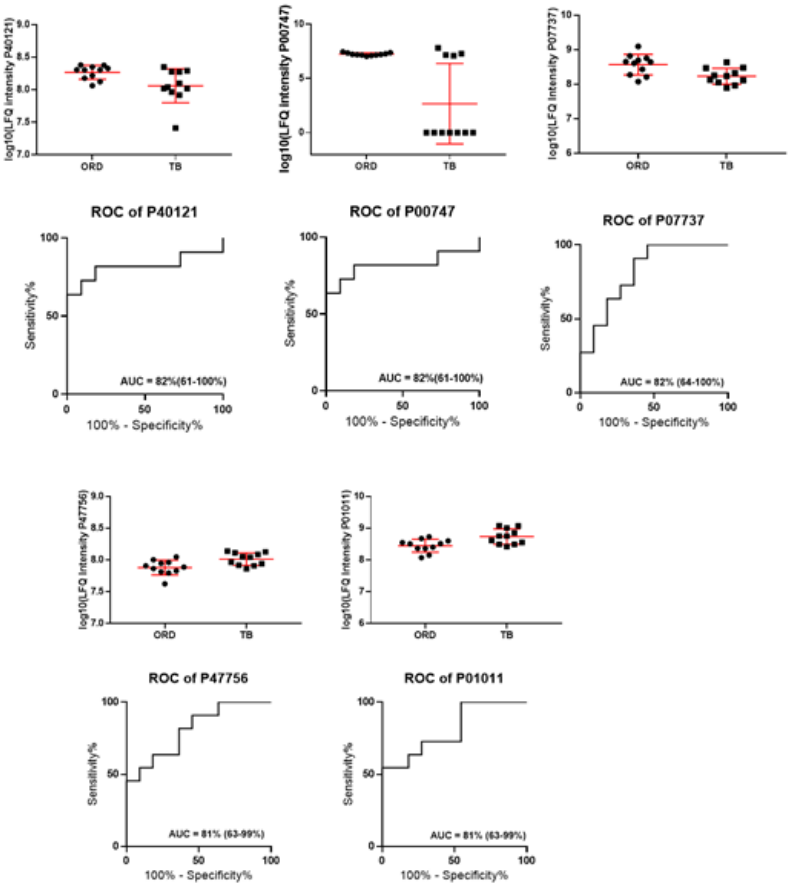

Figure 5

Representative plots showing the log10 transformed LFQ intensities of proteins identified from saliva samples of individuals with TB and ORD, and ROC curves showing the accuracies of these proteins in the diagnosis of TB. Error bars in the scatter-dot plots indicate the mean with standard deviation. Representative plots for five proteins with AUC ≥ 0.80 are shown.

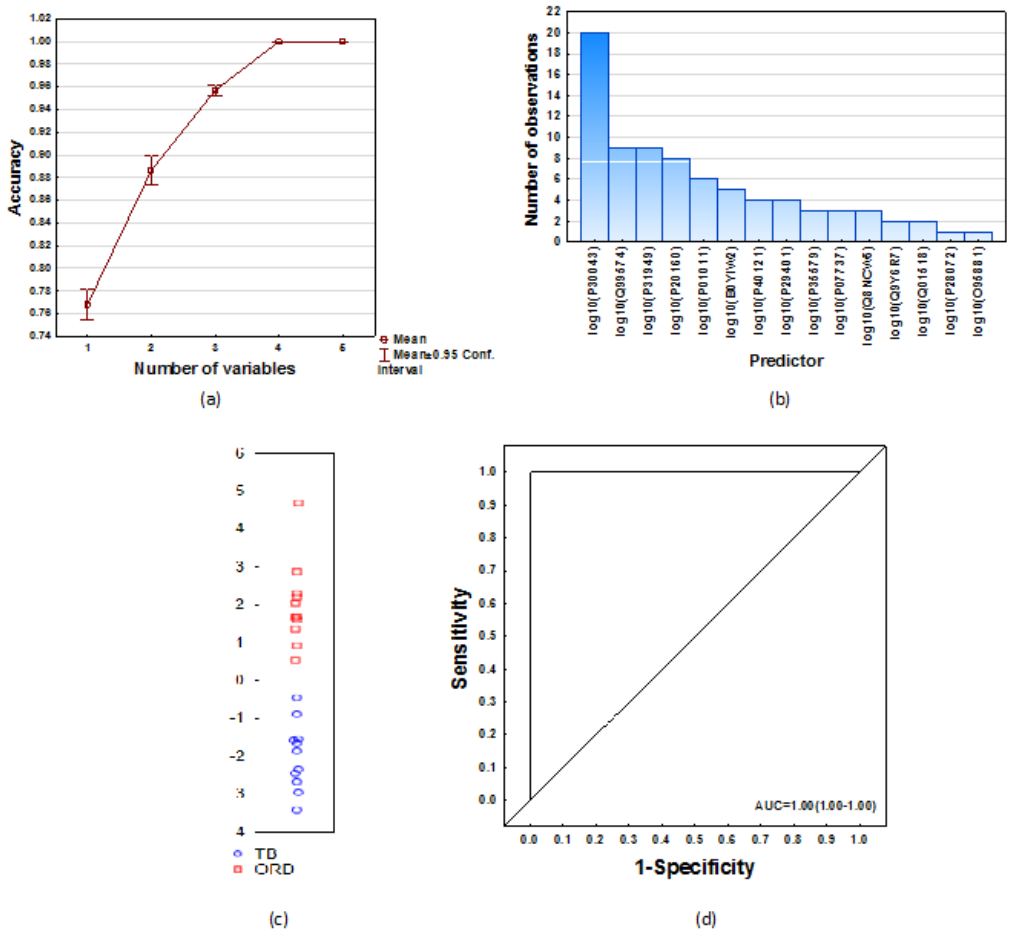

Figure 6

Inclusion of different proteins into host biosignatures for the diagnosis of TB disease. (a) Number of variables required for optimal accuracy of discriminating TB from ORD. Black squares indicate the mean accuracy and error bars indicate the 95% confidence interval. (b) Frequency of proteins in the top 20 most accurate general discriminant analysis models in the diagnosis of TB disease. (c) Scatter plot showing the ability of the 5-protein signature to classify individuals as TB or ORD. (d) ROC curve showing the accuracy of the 5-protein biosignature. Red squares: Participants with ORD. Blue circles: Participants with TB.