DISSERTATION

EXOSOMES: A POTENTIAL NOVEL SOURCE OF BIOMARKERS FOR TUBERCULOSIS

Submitted by

Gustavo Diaz

Department of Microbiology, Immunology and Pathology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2017

Doctoral Committee:

Advisor: Karen M. Dobos
Co-Advisor: Nicole Kruh-Garcia

John Belisle
Mercedes Gonzalez-Juarrero
Santiago Di Pietro

ABSTRACT

EXOSOMES: A POTENTIAL NOVEL SOURCE OF BIOMARKERS FOR TUBERCULOSIS

One of the major problems in the control of Tuberculosis (TB) is the lack of an effective diagnostic tool. The most commonly used test for TB diagnosis is the examination of sputum samples using light microscopy. This test has shown sensitivity rates as low as 40%. The current gold standard for TB diagnosis is the bacterial culture from sputum samples, which usually takes between 3 to 12 weeks on solid media. Inaccurate and delayed diagnosis increases the likelihood of transmission of *Mycobacterium tuberculosis (Mtb)* and impairs proper treatment. Consequently, 10.4 million TB cases and 1.4 million deaths were reported in 2015. The need for new TB biomarkers is urgent. Strategies avoiding the use of sputum will improve the current capacity to diagnose TB, specifically in children and HIV co-infected patients. The analysis of serum-derived TB biomarkers represents a promising alternative, however, the highly abundant proteins found in human serum (albumin, immunoglobulins, and transferrin among others) hinder the identification of the TB biomarkers, drastically affecting the development of potential routine tests.

Exosomes (a type of extracellular vesicles) have emerged as a great alternative source of biomarkers for several diseases since they can be obtained from almost all biological fluids. Virtually all nucleated cells in the human body produce and release exosomes to the extracellular space in a constitutive manner. Interestingly, the exosome composition depends on the cells of origin as well as the physiological status of the host. Several studies have demonstrated that exosomes released from *Mtb*-infected cells contained mycobacterial proteins.

More importantly, previous studies in our laboratory, suggested that exosomes from TB patients could carry peptides derived from *M. tuberculosis*. Considering the biology of the exosomes, it is possible that the host exosomal proteome can change because of infectious processes.

The host proteome has not been explored in the context of *Mtb* infection. Our first set of experiments demonstrate that the infection with *Mtb* influences change in the protein composition of exosomes released from infected cells. Comparative proteomic analysis revealed significant differences between exosomes from infected and control cells. Forty-one proteins were significantly more abundant in exosomes from infected cells. We sought to explore the differential abundance of membrane associated proteins since they represent a more accessible set of targets for the downstream development of a biomarker assay; 63% (26/41) of the proteins that were significantly more abundant in exosomes from infected cells were also membrane associated. These results were obtained from an *in vitro* model of infection.

In the next step, we aimed to discover proteins showing significantly different abundances amongst different TB disease states in serum-derived exosomes. We analyzed three groups of samples from TB endemic regions: smear and culture positive, smear negative and culture positive, and TB suspects without microbial evidence of disease, as well as a healthy group from a non-TB endemic area (as a control). For these experiments, we used a novel proteomic approach known as Hyper Reaction Monitoring-Sequential Windowed Acquisition of All Theoretical Fragment Ions Mass Spectrometry (HRM-SWATH-MS). In HRM, spectral libraries containing information of the peptides present in the samples were generated. Each sample was processed by SWATH-MS. The spectral libraries, as well as the fragment ion maps obtained from SWATH-MS, contained a set of the standard synthetic peptides which elute across the

whole range of retention times (RT) of the chromatographic column. The information from the standard peptides (iRT) was used to generate normalized RT values for each peptide identified. Nine proteins showed a very distinct dynamic profile across the study groups, FCGR3A, lysozyme C and allograft inflammatory factor 1, showed a step-wise increase, while fetuin-A, angiotensinogen, coagulation factor XII, PGRP-L, CBG, and CPN2, showed a step-wise decrease. This study demonstrated that a set of human proteins concentrated in serum exosomes follow a dynamic pattern that is linked with TB disease. Our findings suggest that exosomes from human serum are a source for TB biomarkers.

Considering the large number of potential TB biomarker candidates obtained from proteomic studies and the challenges of selecting some of them for validation studies, in chapter #4, we aimed to develop a predictive model to discriminate TB positive samples from TB negative samples using proteomic analysis from serum-derived exosomes. We tested the regression model developed by Tibshirani in 2006, least absolute shrinkage and selection operator (Lasso). Additionally, we tested a modified version known as adaptive Lasso which improves the stability in the selection of predictors. Lasso generates a subset selection producing simpler models that are relatively easy to interpret. We obtained a regression model with nine predictors that allowed us to segregate TB positive from TB negative samples. To further evaluate the discriminatory capacity of this protein signature, we calculated the area under the curve (AUC) of the receiver operating characteristic curves (ROC). The AUC-ROC was 0.75. The predictors selected using adaptive Lasso included three proteins of the complement system, two immunoglobulin chains, the acute-phase plasma protein A1AG1, the anti-inflammatory metalloproteinase CBPN, the glucose binding protein glucokinase GCK, and the protein Sex hormone-binding globulin SHBG. Additionally, we tested a group of the suspect samples which had previously been

classified negative for TB based on microscopic examination of sputum and culture. However, these samples demonstrated one or more *Mtb* peptides by MRM/MS assay in parallel studies conducted by our laboratory. Nine samples were categorized as TB-negative by the adaptive Lasso regression model.

Two of the proteins identified in this study (Fetuin A and SHGB) have shown promising results as TB biomarkers in previous studies. Fetuin A (chapter #3) and SHGB (chapter#4) have shown differential expression in different discovery studies (2D-electrophoresis and LC-MS/MS, respectively) further validated by ELISA. Based on our findings we hypothesize that fetuin A and SHGB could be concentrated in the exosome-rich fraction of human serum. If this hypothesis is true, the evaluation of exosomes instead of whole serum will increase the predictive power of these proteins.

Finally, our results suggest that exosomes derived from serum samples carry information that could improve the identification of TB patients. However, the current evidence suggests that there is not a single approach to find the "perfect biosignature" for TB diagnosis. Perhaps the design of algorithms combining bacterial and host derived markers from serum-derived exosomes can result in a stronger tool that definitively helps to improve the current situation of TB worldwide.

ACKNOWLEDGMENTS

My doctoral experience has allowed me the interaction with great people who have contributed during my education, professional and personal life. First, I want to thank my advisor Dr. Karen Dobos that gave me the opportunity to be part of her team and provided me excellent training and tools to successfully complete my doctoral degree in a very welcoming environment. I also want to thank Dr. Nicole Kruh-Garcia, my coadvisor, who was fundamental in the development of my research project and gave me the best advices during these years. Both always supported me in every way they could and made me feel part of the Dobos lab family. I also appreciate the time that both Dr. Karen and Nicole spent with me, that really refined my educational process. Additionally, I want to acknowledge Dr. Carolina Mehaffy who opened the door of her home and gave my family and me the opportunity to know this beautiful town and workplace that I can call my second home now, Fort Collins.

The completion of my doctoral degree was possible in great part because of the contribution of my committee members Dr. John Belisle, Dr. Mercedes Gonzalez-Juarrero and Dr. Santiago Di Pietro, who provided me constructive comments during my program. They facilitated the development of my research project providing me their specific comments and suggestions, reflected in this work.

The daily work was also possible thanks to the interaction with very kind and admirable people in the Dobos laboratory. Former and current members of this family, gave me the opportunity to work in an always friendly, helpful and respectful environment. Special thanks to Megan Lucas, Kala Early, Danny Hesser, Charlie Hoxmeier, Anne Simpson, Philip Knabenbauer, and all the students who I had the privilege to mentor and from I learned a lot during this experience.

Thanks to another “dobocyte” member, my wife Dr. Luisa Nieto who was my 24 hours work and life partner during this experience and who made my life happier during these years. Special thanks also to my family and friends in Cali, who have contributed in any stage of my life and motivated me to pursue my career and dreams.

DEDICATION

To my wife Luisa Maria, my daughter Violeta my motivation to keep going dreaming, to my mother Maria who imprinted in my soul the love for studying, and to my sisters Daniela y Lorena who have supported me in my shadiest nights.

TABLE OF CONTENTS

Chapter 1: Literature review and overview of the dissertation

1.1 Diagnosis of Tuberculosis (TB): from history to current strategies

Diagnosis of Tuberculosis (TB) was for many centuries based on clinical observations. One of the first reports of a clinical diagnosis of TB are described in the Hippocratic aphorisms (~ 400 BC) in which a set of pulmonary symptoms—pleurisy with suppuration, expectoration of blood and pus—affecting individuals between 18 to 35 years of age, were collectively referred as phthisis or consumption (1). In the middle ages, though there were no significant advances in the clinical diagnosis of TB, there were a pertinent number of reports about scrofula, a highly prevalent manifestation of extrapulmonary TB that was described by many medieval physicians as inflammatory tumors principally affecting the neck (2, 3). Significant advances in the clinical diagnosis of TB did not occur until 1816 when Rene Laennec discovered the stethoscope. This improved the accuracy of clinical diagnosis since Laennec could associate the chest sounds of TB patients with subsequent post-mortem pathological findings (4). In this way, Laennec enabled the discrimination between pulmonary and extra pulmonary TB.

The next advance in the diagnosis of TB was made possible by two amazing achievements in the history of human science: the development and improvement of the microscope and the acceptance of the germ theory of disease. Based on the previous works of Galileo Galilei, two different types of microscope were invented during the last decades of the 17th century. Their inventors Robert Hook and Antonj van Leeuwenhoek, working independently, could describe and publish the observations of minute structures from corks, seeds and insects. More importantly, van Leeuwenhoek's studies included the report of animalcule, confirming for the first time the existence of microorganism in nature (5). Soon after the Leeuwenhoek's

observations, Benjamin Marten (a medical doctor often neglected by historians) stated, perhaps for the first time, that consumption or phthisis was caused by an animalcule infecting the lungs (6). A fact confirmed only 163 years later by Robert Koch. On April 10th of 1882 the *Die Ätiologie der Tuberkulose* was published, where Dr. Koch demonstrated in the most elegant manner that the "tubercle bacillus" was the causal agent of TB (7). More importantly for the purpose of this dissertation, Dr. Koch described the first technique to stain and microscopically visualize *Mycobacterium tuberculosis (M. tuberculosis)* (7). The demonstration that *M. tuberculosis* was visible by microscopy opened the door for brilliant scientists to improve this visualization. Franz Ziehl and Friedrich Neelsen followed the work of Dr. Koch, using a mordant, heat and a decolorizing reagent to reproducibly stain the bacillus. Even though this staining technique is generally known as Ziehl-Neelsen, it is important to mention that, this technique was the result of independent efforts initiated by Paul Ehrlich, then, Ziehl, and finally, the combination of these previous works by Neelsen (8). Microscopic detection of *M. tuberculosis* using this technique, developed from 1882 to 1887, is still in use in many countries for TB diagnosis. Though the basis of this staining technique has remained the same, several modifications have been made. One such modification made by Joseph J. Kinyoun around 1896 (formally published in 1914) (9) is currently the most utilized. The Kinyoun method differs from the Ziehl-Neelsen method in that it is a cold staining procedure for acid-fast bacilli. Later in 1938, Hagemann developed a fluorescent method for detecting *M. tuberculosis* staining with auramine (10). After several improvements of the original Hagemann's method, Truant developed one the most important alternative methods for TB diagnosis: the auramine-rhodamine staining in 1962. This fluorescent staining of *M. tuberculosis,* follows the same principles of acid-fast staining but replaces fuchsine dye with the fluorochrome dyes auramine

O and rhodamine B (11). Microscopic examination of sputum provides notable advantages for diagnosis of TB: low cost, modest infrastructure requirements and relatively rapid results. However, some of its disadvantages are the low sensitivity (ranging from 20% to 60%) and the expertise needed to interpret results (12).

1.1.1 Culture-based methods for TB diagnosis

The findings of Dr. Koch also established the basis for two additional tools essential for the control of TB: the tuberculin skin test and bacterial culture which allow confirmation of exposure to the microorganism (7). The tuberculin skin test will be discussed in greater detail later in this chapter. The culture of *M. tuberculosis* remains the gold standard for the diagnosis of TB and the evaluation of treatment response (13). Dr. Koch in his initial experiments used a very simple solid medium made of cow or sheep serum to grow *M. tuberculosis.* The only special component of Dr. Koch's protocol was the patience to wait at least 10 days for the mycobacteria to form macroscopically visible colonies (7). Several modifications to the original medium used by Dr. Koch have since been developed. Initially, *M. tuberculosis* was grown in multiple formulations of agar-based and egg-based solid media (14). The egg-based medium, originally proposed by Wessely and Lowenstein in 1931, and modified for Jensen in 1932; Lowenstein- Jensen (LJ media) is currently one of the most widely used culture media worldwide (15). LJ medium contains malachite green which inhibits the growth of microorganisms other than mycobacteria making this medium very useful for specific evaluation of sputum samples in clinical settings. Two modifications of LJ media are commonly used in developing countries. The first one, developed by Ogawa in 1950 (Ogawa media) and the second one, a modified version of Ogawa (Ogawa-Kudoh) developed by Kudoh and Kudoh in 1974. In both of these modified versions, the addition of asparagine is omitted

making the preparation less expensive (16). Proskauer and Beck set the basis for the development of a liquid synthetic medium (PB media) that contained asparagine as the sole nitrogen source and glycerol as the sole carbon source in 1894. PB media is still used, mostly for experiments where albumin and/or Tween 80 need to be avoided (17). Finally, one of the most significant contributions for the improvement of the culture of *M. tuberculosis,* is the work done by Rene Jules Dubos and Gardner Middlebrook during the 1940's. They developed the formulation for the agar-based medium Middlebrook 7H11, and the liquid medium Middlebrook 7H9 (18, 19). Both media are widely used in research and clinical settings today. The development of a liquid medium for growth of *M. tuberculosis* allowed the design of several semi and fully-automated methodologies used in TB diagnosis. The SEPTI-CHEK™ AFB Mycobacteria Culture System was developed in late 1990's. This system combined 7H9 and three solid media: 7H11, egg-based medium and chocolate-agar in a closed system (20). Later, a semi-automated, radiometric system, BACTEC 460, was launched by Becton-Dickinson. In this system, the growth of *M. tuberculosis* is determined by quantifying $^{14}CO_2$ released after the metabolism of a ^{14}C-labeled substrate in the medium (21). Similar to the BACTEC 460, three additional culture-based systems using different sensors to detect mycobacterial growth have been developed. First, the Versa TREK (Trek Diagnostic Systems) is a semi-automated system in which *M. tuberculosis* growth is evaluated by detecting changes in the pressure inside the culture tube (22). Second, the BacT/ALERT MB (bioMérieux) which contains a gas permeable sensor that changes color as CO_2 is produced by growing microorganisms (23). Finally, the BACTEC Mycobacteria Growth Indicator Tube-MGIT-960 is a fully-automated system that uses a ruthenium pentahydrate oxygen sensor that fluoresces in the presence of aerobic metabolizing bacteria (24). Automated systems are extensively used in

hospital settings and have been adapted to perform drug susceptibility tests. Unfortunately, due to the slow-growth of *M. tuberculosis*, culture-based TB diagnosis is still a lengthy process even when using the MGIT method (~ 4 days).

1.1.2 The use of molecular biology in TB diagnosis

Molecular biology-based methods represent a promising alternative for rapid and specific detection of *M. tuberculosis* infection. Several different types of Nucleic Acid Amplification Tests (NAAT) have been developed. First, a transcription mediated amplification assay, known as Amplified *M. tuberculosis* Direct Test (MTD; Hologic Gen-Probe), that targets bacterial rRNA and produces results in 2.5 to 3.5 hours on direct sputum samples (25). The second NAAT is a loop mediated isothermal amplification test (TB-LAMP) which targets a species-specific DNA sequence (*gyrB*) and the universal 16S ribosomal DNA for *Mycobacterium* genus (rDNA); the test generates results as rapidly as 1 hour (26). In 2016, TB- LAMP was recommended by the World Health Organization (WHO) as a replacement for microscopy (TB-LAMP performance was 40% better than microscopy). Additionally, this test can be used as a follow up test for sputum-smear negative patients. Finally, one of the most advanced and widely used NAAT for TB diagnosis is the Xpert MTB/RIF. This test utilizes a hemi-nested polymerase chain reaction targeting the *M. tuberculosis* specific-sequence of the *rpoB* gen. A great advantage of the Xpert MTB/RIF is its capacity for detecting rifampicin resistance in the same assay (12). A completely different molecular-based approach is the implementation of matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) to detect a spectrometric pattern specific for *M. tuberculosis.* In the MALDI- TOF MS assay, pure strains of *M. tuberculosis* are used to generate spectral libraries. Then, the unknown sample is processed, and the spectra is compared to the library for identification (27). The

problem with this approach is the necessity for culturing the specimen before analysis. Overall, the high cost and need for complex infrastructure are the biggest obstacles in implementation of these molecular-based approaches to diagnosis TB in a general setting.

1.1.3 Immunology-based methods for TB diagnosis

Most of the approaches for TB diagnosis mentioned above depend on the microscopic visualization and culturing of *M. tuberculosis*. Additionally, all these tests require sputum samples, where the living bacteria are concentrated. Unfortunately, obtaining a sputum sample imposes the risk of spreading the bacteria to health care personnel and is practically impossible to obtain from children younger than 5 years old (28). Additionally, *Mtb*-HIV-coinfected patients normally have very low bacterial load (paucibacillary) in their sputum (29). There are alternative approaches where, instead of looking directly for the microorganism, infection with or exposure to *M. tuberculosis* is determined by studying the host immune response. The bases for immunology-based tests were stablished by Dr. Koch in 1890 (30). Initially, Dr. Koch considered that the injection of a preparation from *M. tuberculosis* culture (tuberculin) could be a potential treatment for TB. Soon after, it was demonstrated to be inefficacious as a treatment, however, tuberculin was considered a potential diagnostic strategy due to the strong reaction produced in some individuals after its injection (30). The findings from tuberculin injection were further studied by Von Pirquet and resulted in the development of a skin test (tuberculin was scratched in the skin) in 1907 (31). Thereafter, Charles Mantoux proposed the intradermal injection of tuberculin (30). Several studies to standardized tuberculin composition were done to obtain the final composition, currently known as Purified Protein Derivative (PPD), by Florence Seibert in 1934 (30, 32). Tuberculin Skin Test (TST) is based on a type-IV hypersensitivity reaction where CD4+ T lymphocytes, previously challenged with

mycobacterial antigens, migrate to the site of PPD injection (33). TST is used to determine *M. tuberculosis* infection, however, TST could yield a positive reaction in non-infected individuals vaccinated with Bacillus Calmette-Guérin (BCG) and/or exposed to non-tuberculous mycobacteria (NTM). Another limitation of TST is the inability to discriminate active from latent *M. tuberculosis* infection (32, 34). Interferon gamma released assays (IGRAs), which provide more specificity, can be used as alternatives for TST. Two commercially available IGRAs are widely used in clinical settings: QuantiFERON-TB Gold In-Tube test (QFT-GIT) and T-SPOT TB test (T-Spot). QTF-GIT uses a combination of antigens of the mycobacterial proteins ESAT-6, CFP10 and TB7.7. These antigens are not present in BCG and most of the NTM. For the test, the blood of an individual is mixed with the antigens and the exposure to *M. tuberculosis* is determined based on interferon gamma concentration (35, 36). The T-Spot test determines *M. tuberculosis* infection based on the number of cells producing interferon gamma using an enzyme-linked immuno-spot assay. This test uses the antigens ESAT-6 and CFP10, separately (37). In addition to IGRAs, several serum-based tests have been developed looking for antibodies specific for a variety of *Mtb* proteins: 38KDa, HspX, ESAT-6, Ag85 among others (38, 39). Most of these serological tests have shown poor performance for TB diagnosis with sensitivities ranging from 0.97% to 59%. Even worse, in 2011, WHO declared that the data associated with the evaluation of serological tests have low quality and the amount of false-positive and false-negative results had an adverse impact on patient outcome. Eventually, WHO recommended that clinicians not use serological tests for the evaluation of TB suspects (40).

A different approach for diagnosis of TB is the direct evaluation of bacterial antigens in patient samples. Lipoarabinomannan (LAM), a major constituent of the cell envelope of *M.*

tuberculosis, has been found in urine samples of TB patients showing different patterns of sensitivity depending on the HIV status of the patients (41). Clearview TB is an enzyme-linked immunosorbent assay (ELISA) that allows for a quantitative detection of LAM in urine with higher sensitivity when the urine is 100-fold concentrated (42). An alternative version of a LAM- based detection assay is the Alere Determine™ TB-LAM-Ag. This is a lateral flow assay that can be used as point-of-care test due to its simplicity: removing the need for specific infrastructure, while providing rapid results with a facile protocol (41, 43). The sensitivity of LAM tests is low in HIV-negative patients (10% to 20%) but is increased in HIV-positive patients with an average of 56% (43). The sensitivity of LAM test is even higher (66.7%) in HIV-positive patients with very low T-cell CD4+ counts (50 cells/µl) (44, 45). Overall, other than IGRA, most diagnostic tests have focused on *M. tuberculosis* detection or mycobacterial derivatives such as proteins, lipoglycans and/or nucleic acids. The evaluation of host-derived markers could increase the capacity for TB diagnosis and the identification of other stages of infection.

1.2 TB burden and current diagnostic challenges

According to WHO, about 5.2 million pulmonary TB cases were reported in 2015 by the national TB programs worldwide. A little over half of these cases were bacteriologically confirmed while the rest were clinically diagnosed without bacterial confirmation (WHO 2016 Global report). Since the number of estimated TB cases for 2015 was 10.4 million, the significant gap between reported cases and the actual number of cases is quite evident. Two main reasons account for this gap: lack of efficient reporting systems and more accessible, sensitive and accurate diagnostic tools. As of 2015, the two main tools for the bacteriological confirmation of TB cases continued to be sputum microscopy and culture with a majority of the

3 million TB cases being confirmed using microscopy only. Another important component contributing to the TB epidemic is the large proportion of latent TB infection (LTBI) cases. In fact, the data obtained from the consensus study done by Christopher Dye *et al.,* back in 1999, remains valid: "it is estimated that 32% of the global population is infected with *M. tuberculosis"* (46). Currently, there is no test for the diagnosis of asymptomatic people latently infected with the bacterium. LTBI detection is particularly important in HIV positive patients since they are at higher risk of TB reactivation (47). Unfortunately, the two available identification strategies for LTBI, TST and IGRA, are T- cell dependent immunological assays and immunosuppression severely affects their accuracy making them insufficient for detecting LTBI. Indeed, as the CD4+ cell count decreases, the effectiveness of IGRA declines (48). WHO has recommended the use of the lateral flow test for LAM detection in urine to screen for LTBI in HIV infected individuals (13). Another focal group for LTBI testing is children younger than 5 years old living in proximity to active TB patients (49). For children living in endemic areas, LTBI diagnosis is more complicated since TST often generates false positive results from BCG-vaccinated individuals. This is important, considering that about 90% of countries have a policy for universal BCG vaccination (50). Neither TST nor IGRA are useful to discriminate active TB from LTBI. It is evident that diagnosis of active disease mainly depends on *M. tuberculosis* while the identification of LTBI cases is based on host response. The combination of host and bacterial markers could improve both active TB diagnosis as well as diagnosis of LTBI individuals.

The control, reduction, and elimination of TB is a multidimensional task, requiring the development of new efficacious treatments and protective vaccines in addition to overcoming the described challenges related to diagnosis. The necessity for new drugs to treat TB is urgent,

mainly, due to the spread of drug resistant mycobacteria and the lack of treatment adherence due in part to the length of the current regimens (51). Accordingly, clinical trials to test efficacy of new drug-candidates demand the development of tests to accurately assess treatment response. An important contribution to the number of TB cases each year is derived from people with LTBI who convert to active disease. It is expected that 5-10% of individuals with LTBI will progress to active TB during their lifetime (52). A test to identify these LTBI cases with a higher risk of progression to active disease is a limiting factor in accurately providing preventive treatment.

Table 1.1 Current assays in development for the diagnosis of TB.

Test Description	**Status**
Quantitative PCR for *M. tuberculosis*. Automated system (BD MAX)	Initial validation: 100% sensitivity, 97.1% specificity (53)
Portable real time PCR (Genedrive MTB/RIF) from Epistem	Showed 45.4% sensitivity in 2016 (54)
Line probe assay for isoniazid and rifampicin resistance (Genotype MTBDR*plus*) from Hain Lifescience	Recommended by WHO after FIND evaluation (55)
Line probe assay for resistance to fluoroquinolones and second-line injectable drugs (Genotype MTBDR*sl*) from Hain Lifescience	Recommended by WHO (available at: http://www.who.int/tb/ WHOPolicyStatementSLLPA.pdf?ua=1)
Closed-tube real time PCR (MeltPro) from Zeesan Biotech	Sensitivity to detect resistance to rifampicin (94.2%), isoniazid (84.9%), ofloxacin (83.3%), amikacin (75%), kanamycin (63.5%). (56)
Line probe assay for isoniazid and rifampicin resistance (NTM+MDRTB) from Nipro	Recommended by WHO after FIND evaluation (55)
Automated real time PCR for Mtb (RealTime MTB/TB MDx m2000) from Abbott	Sensitivity 100% in smear-positive samples (57)
Chip-based NAAT with real time PCR on handheld device for *Mtb* (Truenat MTB)	In current evaluation by FIND and ICMR
Next-generation cartridge-based detection Mtb + rifampicin resistance (Xpert MTB /RIF Ultra)	Recommended by WHO (March 2017) (available at: http://who.int/tb/features_archive/Xpert-Ultra/en/)
Single –cartridge mobile platform (Xpert Omni) Cepheid	FIND study pending

NAAT (Xpert XDR)-Cepheid	FIND study pending (anticipated end: 2018)
Urine dipstick for TB Lipoarabinomannan-Alere	Recommended by WHO in people with HIV with CD4 count ≤100 cells/µl (Available at: http://www.who.int/tb/areas-of-work/laboratory/policy_statement_lam_web.pdf)

1.3 TB biomarkers: discovery, applications and current state

The detection of a *M. tuberculosis* molecules (protein, lipid, nucleic acid) in a biological sample derived from a clinical suspect can be used to diagnose TB. In the same way, immunological markers unequivocally linked to *M. tuberculosis* infection can also be used for diagnostic purposes. Collectively, biological markers that allow for the identification of patients undergoing a pathological process are denoted diagnostic *biomarkers* (58). A comprehensive definition of biomarkers according to the Office of Science Policy-National Institutes of Health (NIH) is: "Biological marker (biomarker): a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention" (58). The principal features of an ideal biomarker are: objectivity, accuracy and reproducibility. Biomarkers have been widely studied as surrogate endpoints in clinical trials (59). However, they can also play important roles in different scenarios of biomedical research including the evaluation of disease progression (prognosis) or disease state. In a simplistic explanation of the "life cycle" of *M. tuberculosis*, the initial infection is established by the inhalation of aerosolized sputum droplets loaded with bacteria from an active TB patient (60). It is hypothesized that a very few individuals can clear the infection and continue, as if they have never encountered the bacterium (61). Usually, *M. tuberculosis* invades alveolar macrophages, in the lower lung (60), establishes an intracellular infection and is fully contained by the immune response leading to LTBI. The individual can maintain the latent state of infection for decades. However, a person with LTBI can become an

active TB patient with the potential of spreading the mycobacteria, again in sputum droplets (60, 61). A third possibility happens when after the establishment of the intracellular infection, the immune system of the patient is not able to control the bacterium, and the patient develops active TB disease. Active TB patients after receiving treatment can become LTBI cases or get a complete sterilizing cure (62). None of the states during *M. tuberculosis* infection fit a perfect dichotomous classification. Instead, the transition from LTBI to active disease and from active TB to definitive cure or LTBI again (after treatment), is a continuum of stages (62). To improve the control of TB, it is important to discover diagnostic biomarkers: to identify active TB patients, early prediction of successful treatment, LTBI cases at a higher risk of reactivation, and LTBI cases in general (Figure 1.1).

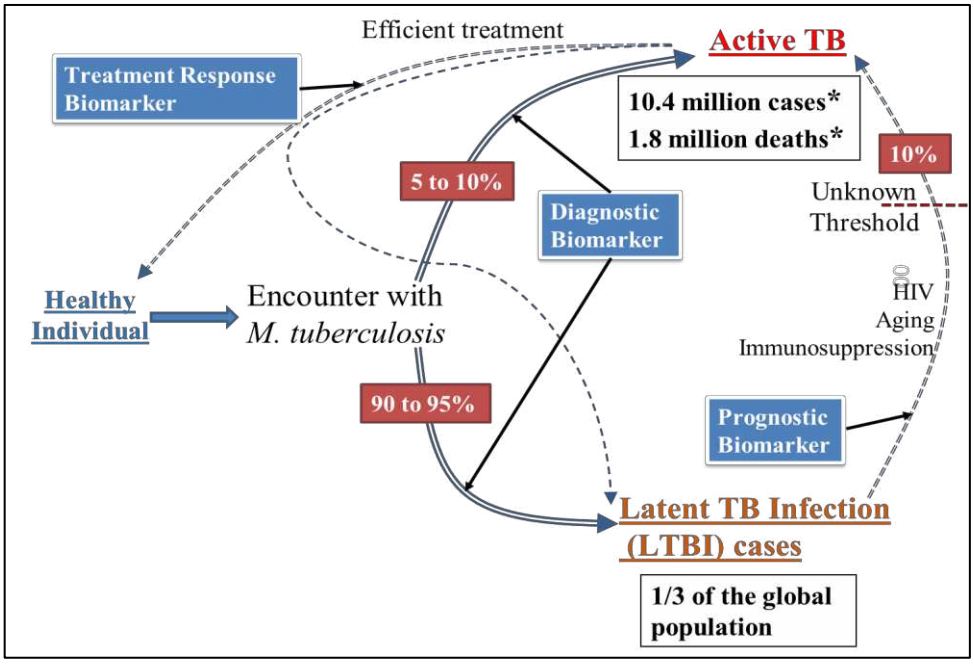

Figure 1.1 Biomarkers for Tuberculosis**.** To effectively control TB, there is a need for the discovery of new diagnostic biomarkers not only for active disease but for the identification of Latent TB infection (LTBI). Additionally, new prognostic biomarkers to identify LTBI cases at higher risk to become active TB are imperative. Finally, with the ascending number of drug resistant TB cases, the discovery of surrogate endpoint biomarkers to evaluate treatment response, is a worldwide priority. *Estimated numbers in 2016 according to WHO.Biomarkers for active disease and latent infection.

In the context of infectious diseases biomarkers can be pathogen-derived, as well as, host-derived. In addition to the already mentioned pathogen-derived diagnostic biomarker: *M. tuberculosis* and mycobacterial DNA from sputum and LAM from urine, several studies have tested mycobacterial biomolecules as potential markers of active disease or latent infection. The fibronectin-binding mycolyltransferases (Ag85) were found in plasma of TB patients complexed to immunoglobulins and fibronectin (63). Surprisingly, there are no recent studies testing for presence of Ag85 in blood from active TB patients, except from a recent study conducted by Kruh-Garcia *et al* (64). From the host side, Sartain *et al.,* proposed an improved version of a serological assay to discriminate different stages of TB disease. Sartain and colleagues, developed a protein microarray assay for the simultaneous detection of serological reactivity to several proteins from *M. tuberculosis*. In this strategy, they generated a library of 960 simple fractions from cytosol and supernatant of *M. tuberculosis*, that were tested against sera samples from a heterogeneous group: PPD+ patients (as LTBI cases), cavitary TB (an advanced form of active TB characterized by an open granuloma from which the necrotic center was ejected via the bronchial tree (65)), non-cavitary TB, HIV/TB coinfection and HIV+ TB negative patients. They found four antigens exclusively associated with cavitary TB (Psts1, HspX, Mpt64 and TrxC) and 11 antigens (including SodC and BfrB) that generated the strongest response in cavitary and non-cavitary TB (66). A different serum-based approach to identify active TB or LTBI is based on the characterization of circulating microRNAs (67, 68). One study, showed that 59 microRNAs (including miR93* and miR29a) were upregulated in active TB patients. Further evaluation demonstrated that miR29a had an area under the receiving operator characteristic curve (ROC) of 0.83, suggesting it as a great potential diagnostic biomarker (69). In the context of host-derived diagnostic biomarkers, it was found

that serum amyloid A and transthyretin showed diagnostic accuracy ranging from 78% to 90%, using surface-enhanced laser desorption ionization time of flight-MS (SELDI-MS) and immunoassays (70). Another study compared TB patients (HIV+ and HIV-) with LTBI, patients with other respiratory diseases (ORD) and negative for TB. They initially found 165 proteins differentially expressed in TB patients (HIV-) compared to LTBI/uninfected by shotgun proteomics. Interestingly, they removed from the analysis proteins known to be part of the acute phase response in various infectious diseases. Finally, they performed a multiplexed SRM study, searching for 87 proteins from the discovery phase and from other studies with similar designs. The final analysis showed that 10 proteins could discriminate TB/HIV- and from ORD (area under ROC 0.96), while 8 proteins discriminated TB/HIV+ from ORD (area under ROC 0.96). CD14 and the extracellular glycoprotein SEPP1 were common to both groups (71). Finally, one of the most recent fields for discovery of new biomarkers is the identification and quantification of small molecules (<1.5KDa) involved in all stages of cellular function, in a process known as metabolomics (72). Metabolites from the pathogen and from the host, have been studied as potential biomarkers of TB. Regarding metabolites from *M. tuberculosis*, several authors have used *in vitro* culture for initial identification to avoid the difficulties associated with host matrix interference. The compound tuberculostearic acid (TBSA) was identified in different samples of active TB patients (73). Unfortunately, TBSA showed low sensitivity and specificity (54% and 80%, respectively, in the Sezguin's study) and high cost for its detection (73, 74). Overall, cell wall lipids (specifically mycolic acids) are under intense research in this field. Presence/absence of different classes of mycolic acids has showed potential to discriminate between *M. tuberculosis* and other important group of clinically relevant NTM (Mycobacterial ID System (MYCO-LCS) (72, 75). Regarding host-derived

metabolites, plasma metabolites showed promising capacity to discriminate active TB patients from uninfected controls and patients with pneumonia. Specifically, the presence of ceramide showed sensitivity and specificity higher than 0.85 for TB diagnosis (76). One study compared serum metabolites from TB patient and healthy controls, using nuclear magnetic resonance spectroscopy (NMR). Thirteen metabolites were significantly upregulated and four down regulated in the patient's group. According to this metabolic- signature the most affected process during infection was protein biosynthesis (77). One well characterized method to detect metabolites as diagnostic biomarker of TB involved the development of an "electronic nose" able to distinguish mycobacteria species in laboratory settings, however, the test showed low sensitivity (75%) and specificity (67%) in clinical settings (72). One of the major challenges of metabolomics studies in the field of TB diagnosis is their translational application due to the use of complex and expensive technologies.

Several researchers have tried to apply a diverse type of proteomic approaches to find TB biomarkers in serum/plasma and other alternative samples to sputum (78). The concentration of proteins in different body fluids can reflect pathological processes. Several proteins are currently used as markers of normal/disease conditions in routine clinical settings: albumin, hemoglobin, and liver transaminases, among others. These proteins are normally present in high abundance in human plasma, which has allowed the development of simple, cheap, and high-throughput detection assays (ELISA, colorimetric reactions, etc.) (79). Many proteins involved in cellular processes can circulate in the blood stream, consequently, the study of the plasma (or serum) proteome, can reflect clinically relevant disease conditions (80). Mass spectrometry represent a robust tool to complete a comprehensive analysis of proteins in plasma, generating biosignatures (arrays of protein/peptides) associated to disease status. Unfortunately, due to the

extremely wide dynamic range of protein concentration in the human plasma (or serum) (>10 orders of magnitude) the identification of low abundant proteins is a significant challenge (81). A general preclinical workflow biomarker discovery includes biomarker identification, often in a cell culture or animal system, qualification-verification and validation phases, before the clinical evaluation (81). The discovery phase is frequently developed using "shotgun proteomics". This phase normally includes the unbiased analysis of a low number of samples, and the identification of hundreds or thousands of protein candidates (81). A shotgun approach for discovery of biomarkers will include, the digestion of proteins into peptides (commonly using trypsin), that will be resolved by reverse phase liquid chromatography. The resulting peptides are injected in the mass spectrometer into gas phase, via electrospray ionization. Then, the mass/charge ratio of the peptides is determined (full MS scan of precursor ions) followed by the selection of a set of precursor ions, based on its abundance, to undergo fragmentation (82). The resulting fragments provide information of amino acid sequence, this information, and the one from the precursor ions are used for protein identification and relative quantification, using specialized software that compares the in-silico processing of a date base with the experimental data. The new generation of instruments can survey several thousand of individual peptides in a fraction of second. However, several limitations make shotgun proteomics not to be suitable for accurate quantification of specific targets with clinical relevance. First, there is a bias in shotgun proteomics towards the more abundant peptides in a sample, hampering the identification the very low abundant components in the sample. Second, only a fraction of the excessive number of peptides in a complex sample is analyzed. This, in addition to the stochastic nature of the selection of peptides for analysis decreases reproducibility of the results (82). Overall, the discovery phase allows the identification of a set of candidate proteins significantly more or less

abundant between two or more conditions. In 2014, Kruh-Garcia *et al.,* reported several mycobacterial proteins in serum with the potential of identifying TB positive patients: Ag85b, Ag85c, Mpt32 (Apa), BfrB, GlcB, HspX, KatG and Mpt64. The identification of *M. tuberculosis* proteins in serum was not significantly affected due to HIV status of the patient. In this study, shotgun proteomics was used to select 76 peptides corresponding to 33 candidate proteins. From these, 29 peptides corresponding to 17 proteins were identified by Multiple Reaction Monitoring (MRM) assay in TB positive patients (64). In a recent study, the MRM methods utilized by Kruh-Garcia were optimized into two multiplexed MRM assays. Based on previous discoveries the MRM assays were developed to include isotopic-labeled peptides. They identified 35/40 patients with active TB, based on the presence of at least one peptide out of 18. When comparing TB positive with healthy individuals, four peptides from the proteins, Cfp2, Mpt32, Mpt64 and BfrB were significantly associated with TB patients (83). The last two studies used a unique fraction from sera samples enriched with nanovesicles secreted by the cells known as exosomes. The use of exosomes not only aided with the elimination of most of the heavily abundant human serum proteins but concentrated the bacterial proteins (64, 83).

1.4 Exosomes as potential source of TB biomarkers

The relatively easy sampling of exosomes from several biofluids such as serum, urine, bronchoalveolar lavage and milk among others (84-86), and the modification of their composition based on the cell of origin, make exosomes an attractive biomarker of disease. Even though several proteins are found in most of the exosomes: tetraspanins, actins, and annexins, the total composition of the vesicles is dependent on the cells of origin (85). In 2007, Bhatnagar *et al.,* showed that cells infected with different intracellular pathogens: *M. tuberculosis, M. bovis, Toxoplasma gondii* and *Salmonella typhimurium*, released exosomes

loaded with pathogen associated molecular patterns (87). The proteomic analysis of exosomes released from the mouse-macrophages J774 infected with *M. tuberculosis,* showed the presence of 41 bacterial proteins, including, KatG, Ag85, Cfp10, HspX and others (88). These studies, set the basis for the hypothesis that exosomes from TB patients could be potential source of biomarkers. Initial studies of Kruh-Garcia evaluated the presence of mycobacterial proteins in bronchoalveolar lavage of mice infected with *M. tuberculosis,* at several time points during infection. Some proteins like Ag85A, Ag85B, HspX and Mpt64 were present at all time points (89). From that, the two studies described above at the end of session 1.3.1 have stablished the utility of exosome as a source of bacterial molecules with potential biomarker application. The host proteome has not been explored in the context of *M. tuberculosis* infection, it is possible that a combination of pathogen-derived and host-derived proteins loaded in exosomes from TB patients could represent a stronger predictive power for diagnostic purposes.

1.5 Hypothesis and specific aims

The need for new TB biomarkers is urgent. Strategies avoiding the use of sputum will improve the current capacity to diagnose TB, specifically in children and HIV co-infected patients. Exosomes have two unique features, their composition changes depending of the status of the patient and they are retrievable from blood and urine, which make them promising candidates to fill the gap in knowledge described above. The study of the proteome of exosomes from serum of TB patients could reveal a novel set of host-derived diagnostic biomarkers. Additionally, the use of exosomes will aid to overcome the problem associated with the large dynamic range of proteins present in serum. For those reasons, we stated the following hypothesis for this doctoral project: *Exosomes released from Mtb-infected patients have a distinct set of proteins that can be exploited as biomarkers for TB. The analysis of the*

significantly abundant host proteins in exosome from infected patients will reveal new aspects of the host pathogen interaction during TB infection. To test the hypothesis, we defined the following three aims:

Aim 1: to demonstrate that exosomes released from *Mtb*-infected cells exhibit a characteristic proteome compared to uninfected cells.

Aim 2: to evaluate the host protein variation in serum-derived exosomes from TB positive patients, TB suspects, TB negative patients, and healthy controls.

Aim 3: to develop a predictive model to discriminate TB positive human samples from TB suspects, TB negative, and healthy controls.

References

1. Hipocrates the Aphorisms of Hippocrates: With a Translation into Latin and English: A. J. Valpy; 1822.
2. Duarte G I, Chuaqui F C. [History of scrofula: from humoral dyscrasia to consumption]. Rev Med Chil. 2016;144(4):503-7.
3. Grzybowski S, Allen EA. History and importance of scrofula. Lancet. 1995;346(8988):1472-4.
4. Roguin A. Rene Theophile Hyacinthe Laënnec (1781-1826): the man behind the stethoscope. Clin Med Res. 2006;4(3):230-5.
5. Wollman AJ, Nudd R, Hedlund EG, Leake MC. From Animaculum to single molecules: 300 years of the light microscope. Open Biol. 2015;5(4):150019.
6. Doetsch RN. Benjamin Marten and his "New Theory of Consumptions". Microbiol Rev. 1978;42(3):521-8.
7. Koch R. Classics in infectious diseases. The etiology of Tuberculosis: Robert Koch. Berlin, Germany 1882. Rev Infect Dis. 1982;4(6):1270-4.
8. Bishop PJ, Neumann G. The history of the Ziehl-Neelsen stain. Tubercle. 1970;51(2):196-206.
9. Kinyoun JJ. A note on uhlenhuths method for sputum examination, for tubercle bacilli. Am J Public Health (N Y). 1915;5(9):867-70.
10. Darzins E. The bacteriology of Tuberculosis. Minneapolis,: University of Minnesota Press; 1958. 488 p. p.
11. Truant Jp, Brett Wa, Thomas W. Fluorescence microscopy of tubercle bacilli stained with auramine and rhodamine. Henry Ford Hosp Med Bull. 1962;10:287-96.
12. Steingart KR, Schiller I, Horne DJ, Pai M, Boehme CC, Dendukuri N. Xpert® MTB/RIF assay for pulmonary Tuberculosis and rifampicin resistance in adults. Cochrane Database Syst Rev. 2014;1:CD009593.
13. World Health Organization. Global Tuberculosis Report. 2016.
14. Schwabacher H. - The pitfalls in the laboratory diagnosis of urinary Tuberculosis: A Report of the Address given at the Urological Section of the Royal Society of Medicine. 1937;- 9(- 3):- 275.
15. Essawy TS, Saeed AM, Fouad NA. Comparative study between using Lowenstein Jensen, Bio-FM media and mycobacteria growth indicator tube (MGIT) system in identification of *Mycobacterium tuberculosis*. Egyptian Journal of Chest Diseases and Tuberculosis. 2014;63(2):8.
16. Kudoh S, Kudoh T. A simple technique for culturing tubercle bacilli. Bull World Health Organ. 1974;51(1):71-82.
17. Youmans AS, Youmans GP. Studies on the metabolism of *Mycobacterium tuberculosis*. III. The growth of *Mycobacterium tuberculosis* var. hominis in the presence of various intermediates of the dissimilation of glucose to pyruvic acid. J Bacteriol. 1953;65(1):100-2.
18. Dubos RJ, Middlebrook G. Media for tubercle bacilli. Am Rev Tuberc. 1947;56(4):334- 45.
19. Middlebrook G. Dubos RJ. The effect of tubercle bacilli on the antigenicity of a

synthetic ester of oleic acid. J Immunol. 1947;56(4):301-6.
20. Sewell DL, Rashad AL, Rourke WJ, Poor SL, McCarthy JA, Pfaller MA. Comparison of the Septi-Chek AFB and BACTEC systems and conventional culture for recovery of mycobacteria. J Clin Microbiol. 1993;31(10):2689-91.
21. Pfyffer GE, Cieslak C, Welscher HM, Kissling P, Rüsch-Gerdes S. Rapid detection of mycobacteria in clinical specimens by using the automated BACTEC 9000 MB system and comparison with radiometric and solid-culture systems. J Clin Microbiol. 1997;35(9):2229-34.
22. Caulfield AJ, Wengenack NL. Diagnosis of active Tuberculosis disease: From microscopy to molecular techniques. Journal of Clinical Tuberculosis and Other Mycobacterial Diseases. 2016;4:33-43.
23. Mattei R, Savarino A, Fabbri M, Moneta S, Tortoli E. Use of the BacT/Alert MB mycobacterial blood culture system for detection of mycobacteria in sterile body fluids other than blood. J Clin Microbiol. 2009;47(3):711-4.
24. Tortoli E, Cichero P, Piersimoni C, Simonetti MT, Gesu G, Nista D. Use of BACTEC MGIT 960 for recovery of mycobacteria from clinical specimens: multicenter study. J Clin Microbiol. 1999;37(11):3578-82.
25. Lin CM, Lin SM, Chung FT, Lin HC, Lee KY, Huang CD, et al. Amplified *Mycobacterium tuberculosis* direct test for diagnosing tuberculous pleurisy--a diagnostic accuracy study. PLoS One. 2012;7(9):e44842.
26. Iwamoto T, Sonobe T, Hayashi K. Loop-mediated isothermal amplification for direct detection of *Mycobacterium tuberculosis* complex, M. avium, and M. intracellulare in sputum samples. J Clin Microbiol. 2003;41(6):2616-22.
27. El Khéchine A, Couderc C, Flaudrops C, Raoult D, Drancourt M. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry identification of mycobacteria in routine clinical practice. PLoS One. 2011;6(9):e24720.
28. Cuevas LE. The urgent need for new diagnostics for symptomatic Tuberculosis in children. Indian J Pediatr. 2011;78(4):449-55.
29. Steingart KR, Ng V, Henry M, Hopewell PC, Ramsay A, Cunningham J, et al. Sputum processing methods to improve the sensitivity of smear microscopy for Tuberculosis: a systematic review. Lancet Infect Dis. 2006;6(10):664-74.
30. Daniel TM. The history of Tuberculosis. Respir Med. 2006;100(11):1862-70.
31. Menzies D. Interpretation of repeated tuberculin tests. Boosting, conversion, and reversion. Am J Respir Crit Care Med. 1999;159(1):15-21.
32. Nayak S, Acharjya B. Mantoux test and its interpretation. Indian Dermatol Online J. 2012;3(1):2-6.
33. Vukmanovic-Stejic M, Reed JR, Lacy KE, Rustin MH, Akbar AN. Mantoux Test as a model for a secondary immune response in humans. Immunol Lett. 2006;107(2):93-101.
34. Burl S, Adetifa UJ, Cox M, Touray E, Whittle H, McShane H, et al. The tuberculin skin test (TST) is affected by recent BCG vaccination but not by exposure to non-Tuberculosis mycobacteria (NTM) during early life. PLoS One. 2010;5(8):e12287.
35. Lempp JM, Zajdowicz MJ, Hankinson AL, Toney SR, Keep LW, Mancuso JD, et al. Assessment of the QuantiFERON-TB Gold In-Tube test for the detection of *Mycobacterium tuberculosis* infection in United States Navy recruits. PLoS One. 2017;12(5):e0177752.
36. Arlehamn CS, Sidney J, Henderson R, Greenbaum JA, James EA, Moutaftsi M, et al. Dissecting mechanisms of immunodominance to the common Tuberculosis antigens ESAT-6, CFP10, Rv2031c (hspX), Rv2654c (TB7.7), and Rv1038c (EsxJ). J Immunol.

2012;188(10):5020-31.
37. Nicol MP, Davies MA, Wood K, Hatherill M, Workman L, Hawkridge A, et al. Comparison of T-SPOT.TB assay and tuberculin skin test for the evaluation of young children at high risk for Tuberculosis in a community setting. Pediatrics. 2009;123(1):38-43.
38. Baghaei P, Tabarsi P, Sabour H, Dehghani S, Marjani M, Shamaei M, et al. Detection of Antibodies Against 6, 16 and 38 kDa Antigens of *Mycobacterium tuberculosis* as a Rapid Test for Diagnosis of Tuberculosis. Tanaffos. 2011;10(4):17-22.
39. Imaz MS, Schmelling MF, Kaempfer S, Spallek R, Singh M. Serodiagnosis of Tuberculosis: specific detection of free and complex-dissociated antibodies anti-*Mycobacterium tuberculosis* recombinant antigens. Braz J Infect Dis. 2008;12(3):234-44.
40. Steingart KR, Ramsay A, Dowdy DW, Pai M. Serological tests for the diagnosis of active Tuberculosis: relevance for India. Indian J Med Res. 2012;135(5):695-702.
41. Shah M, Hanrahan C, Wang ZY, Dendukuri N, Lawn SD, Denkinger CM, et al. Lateral flow urine lipoarabinomannan assay for detecting active Tuberculosis in HIV-positive adults. Cochrane Database Syst Rev. 2016(5):CD011420.
42. Savolainen L, Kantele A, Sandboge B, Sirén M, Valleala H, Tuompo R, et al. Modification of clearview Tuberculosis (TB) enzyme-linked immunosorbent assay for TB patients not infected with HIV. Clin Vaccine Immunol. 2013;20(9):1479-82.
43. Minion J, Leung E, Talbot E, Dheda K, Pai M, Menzies D. Diagnosing Tuberculosis with urine lipoarabinomannan: systematic review and meta-analysis. Eur Respir J. 2011;38(6):1398- 405.
44. Lawn SD, Wood R. Point-of-care urine antigen screening tests for Tuberculosis and cryptococcosis: potential for mortality reduction in antiretroviral treatment programs in Africa. Clin Infect Dis. 2012;54(5):739-40.
45. Lawn SD, Kerkhoff AD, Vogt M, Wood R. Diagnostic accuracy of a low-cost, urine antigen, point-of-care screening assay for HIV-associated pulmonary Tuberculosis before antiretroviral therapy: a descriptive study. Lancet Infect Dis. 2012;12(3):201-9.
46. Dye C, Scheele S, Dolin P, Pathania V, Raviglione MC. Consensus statement. Global burden of Tuberculosis: estimated incidence, prevalence, and mortality by country. WHO Global Surveillance and Monitoring Project. JAMA. 1999;282(7):677-86.
47. Kwan CK, Ernst JD. HIV and Tuberculosis: a deadly human syndemic. Clin Microbiol Rev. 2011;24(2):351-76.
48. Aabye MG, Ravn P, PrayGod G, Jeremiah K, Mugomela A, Jepsen M, et al. The impact of HIV infection and CD4 cell count on the performance of an interferon gamma release assay in patients with pulmonary Tuberculosis. PLoS One. 2009;4(1):e4220.
49. Starke JR, Diseases COI. Interferon-γ release assays for diagnosis of Tuberculosis infection and disease in children. Pediatrics. 2014;134(6):e1763-73.
50. Zwerling A, Behr MA, Verma A, Brewer TF, Menzies D, Pai M. The BCG World Atlas: a database of global BCG vaccination policies and practices. PLoS Med. 2011;8(3):e1001012.
51. van Cutsem G, Isaakidis P, Farley J, Nardell E, Volchenkov G, Cox H. Infection Control for Drug-Resistant Tuberculosis: Early Diagnosis and Treatment Is the Key. Clin Infect Dis. 2016;62 Suppl 3:S238-43.
52. Hauck FR, Neese BH, Panchal AS, El-Amin W. Identification and management of latent Tuberculosis infection. Am Fam Physician. 2009;79(10):879-86.
53. Rocchetti TT, Silbert S, Gostnell A, Kubasek C, Widen R. Validation of a

Multiplex Real-Time PCR Assay for Detection of Mycobacterium spp., *Mycobacterium tuberculosis* Complex, and Mycobacterium avium Complex Directly from Clinical Samples by Use of the BD Max Open System. J Clin Microbiol. 2016;54(6):1644-7.
54. Shenai S, Armstrong DT, Valli E, Dolinger DL, Nakiyingi L, Dietze R, et al. Analytical and Clinical Evaluation of the Epistem Genedrive Assay for Detection of *Mycobacterium tuberculosis*. J Clin Microbiol. 2016;54(4):1051-7.
55. Nathavitharana RR, Hillemann D, Schumacher SG, Schlueter B, Ismail N, Omar SV, et al. Multicenter Noninferiority Evaluation of Hain GenoType MTBDRplus Version 2 and Nipro NTM+MDRTB Line Probe Assays for Detection of Rifampin and Isoniazid Resistance. J Clin Microbiol. 2016;54(6):1624-30.
56. Pang Y, Dong H, Tan Y, Deng Y, Cai X, Jing H, et al. Rapid diagnosis of MDR and XDR Tuberculosis with the MeltPro TB assay in China. Sci Rep. 2016;6:25330.
57. Wang SF, Ou XC, Li Q, Zheng HW, Wang YF, Zhao YL. The Abbott RealTime MTB assay and the Cepheid GeneXpert assay show comparable performance for the detection of *Mycobacterium tuberculosis* in sputum specimens. Int J Infect Dis. 2016;45:78-80.
58. Group. BDW. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther. 2001;69(3):89-95.
59. Strimbu K, Tavel JA. What are biomarkers? Curr Opin HIV AIDS. 2010;5(6):463-6.
60. Cambier CJ, Falkow S, Ramakrishnan L. Host evasion and exploitation schemes of *Mycobacterium tuberculosis*. Cell. 2014;159(7):1497-509.
61. Pai M, Behr MA, Dowdy D, Dheda K, Divangahi M, Boehme CC, et al. Tuberculosis. Nat Rev Dis Primers. 2016;2:16076.
62. Wallis RS, Pai M, Menzies D, Doherty TM, Walzl G, Perkins MD, et al. Biomarkers and diagnostics for Tuberculosis: progress, needs, and translation into practice. Lancet. 2010;375(9729):1920-37.
63. Bentley-Hibbert SI, Quan X, Newman T, Huygen K, Godfrey HP. Pathophysiology of antigen 85 in patients with active Tuberculosis: antigen 85 circulates as complexes with fibronectin and immunoglobulin G. Infect Immun. 1999;67(2):581-8.
64. Kruh-Garcia NA, Wolfe LM, Chaisson LH, Worodria WO, Nahid P, Schorey JS, et al. Detection of *Mycobacterium tuberculosis* peptides in the exosomes of patients with active and latent *M. tuberculosis* infection using MRM-MS. PLoS One. 2014;9(7):e103811.
65. Gadkowski LB, Stout JE. Cavitary pulmonary disease. Clin Microbiol Rev. 2008;21(2):305-33, table of contents.
66. Sartain MJ, Slayden RA, Singh KK, Laal S, Belisle JT. Disease state differentiation and identification of Tuberculosis biomarkers via native antigen array profiling. Mol Cell Proteomics. 2006;5(11):2102-13.
67. Fu Y, Yi Z, Wu X, Li J, Xu F. Circulating microRNAs in patients with active pulmonary Tuberculosis. J Clin Microbiol. 2011;49(12):4246-51.
68. Miotto P, Mwangoka G, Valente IC, Norbis L, Sotgiu G, Bosu R, et al. miRNA signatures in sera of patients with active pulmonary Tuberculosis. PloS one. 2013;8(11):e80149.
69. Harapan H, Fitra F, Ichsan I, Mulyadi M, Miotto P, Hasan NA, et al. The roles of microRNAs on Tuberculosis infection: meaning or myth? Tuberculosis (Edinb). 2013;93(6):596- 605.
70. 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. Lancet. 2006;368(9540):1012-21.
71. Achkar JM, Cortes L, Croteau P, Yanofsky C, Mentinova M, Rajotte I, et al. Host Protein Biomarkers Identify Active Tuberculosis in HIV Uninfected and Co-infected Individuals. EBioMedicine. 2015;2(9):1160-8.
72. Preez ID, Luies L, Loots DT. Metabolomics biomarkers for Tuberculosis diagnostics: current status and future objectives. Biomark Med. 2017;11(2):179-94.
73. Yorgancioğlu A, Akin M, Dereli S, Aktoğu S, Ilis Z, Sezgin A. The diagnostic value of tuberculostearic acid in tuberculous pleural effusions. Monaldi Arch Chest Dis. 1996;51(2):108- 11.
74. Traunmüller F, Zeitlinger MA, Stoiser B, Lagler H, Abdel Salam HA, Presterl E, et al. Circulating tuberculostearic acid in Tuberculosis patients. Scand J Infect Dis. 2003;35(11- 12):790-3.
75. Olivier I, Loots dT. A metabolomics approach to characterise and identify various Mycobacterium species. J Microbiol Methods. 2012;88(3):419-26.
76. Lau SK, Lee KC, Curreem SO, Chow WN, To KK, Hung IF, et al. Metabolomic Profiling of Plasma from Patients with Tuberculosis by Use of Untargeted Mass Spectrometry Reveals Novel Biomarkers for Diagnosis. J Clin Microbiol. 2015;53(12):3750-9.
77. Zhou A, Ni J, Xu Z, Wang Y, Lu S, Sha W, et al. Application of (1)h NMR spectroscopy-based metabolomics to sera of Tuberculosis patients. J Proteome Res. 2013;12(10):4642-9.
78. Haas CT, Roe JK, Pollara G, Mehta M, Noursadeghi M. Diagnostic 'omics' for active Tuberculosis. BMC Med. 2016;14:37.
79. Crutchfield CA, Thomas SN, Sokoll LJ, Chan DW. Advances in mass spectrometry- based clinical biomarker discovery. Clin Proteomics. 2016;13:1.
80. Geyer PE, Kulak NA, Pichler G, Holdt LM, Teupser D, Mann M. Plasma Proteome Profiling to Assess Human Health and Disease. Cell Syst. 2016;2(3):185-95.
81. Parker CE, Borchers CH. Mass spectrometry based biomarker discovery, verification, and validation--quality assurance and control of protein biomarker assays. Mol Oncol. 2014;8(4):840-58.
82. Domon B, Aebersold R. Options and considerations when selecting a quantitative proteomics strategy. Nat Biotechnol. 2010;28(7):710-21.
83. Mehaffy C, Dobos KM, Nahid P, Kruh-Garcia NA. Second generation multiple reaction monitoring assays for enhanced detection of ultra-low abundance *Mycobacterium tuberculosis* peptides in human serum. Clin Proteomics. 2017;14:21.
84. Mathivanan S, Ji H, Simpson RJ. Exosomes: extracellular organelles important in intercellular communication. J Proteomics. 2010;73(10):1907-20.
85. Simpson RJ, Jensen SS, Lim JW. Proteomic profiling of exosomes: current perspectives. Proteomics. 2008;8(19):4083-99.
86. Hoorn EJ, Pisitkun T, Zietse R, Gross P, Frokiaer J, Wang NS, et al. Prospects for urinary proteomics: exosomes as a source of urinary biomarkers. Nephrology (Carlton). 2005;10(3):283-90.
87. Bhatnagar S, Shinagawa K, Castellino FJ, Schorey JS. Exosomes released from macrophages infected with intracellular pathogens stimulate a proinflammatory response in vitro and in vivo. Blood. 2007;110(9):3234-44.
88. Giri PK, Kruh NA, Dobos KM, Schorey JS. Proteomic analysis identifies highly

antigenic proteins in exosomes from *M. tuberculosis*-infected and culture filtrate protein-treated macrophages. Proteomics. 2010;10(17):3190-202.

89. Nicole A. Kruh-Garcia, Jeff S. Schorey and Karen M. Dobos. Exosomes: New Tuberculosis Biomarkers – Prospects from the Bench to the Clinic, Understanding Tuberculosis- Global Experiences and Innovative Approaches to the Diagnosis, Dr. Pere-Joan Cardona (Ed.), InTech, DOI: 10.5772/30720. (2012). Available from: https://www.intechopen.com/books/understanding-tuberculosis-global-experiences-and-innovative-approaches-to-the-diagnosis/tuberculosis-biomarkers-prospects-from-the-bench-to-the-clinic

Chapter 2: Changes in the host proteome of exosomes released from human macrophages after *Mycobacterium tuberculosis* infection[1]

2.1 Introduction

Since the original description of exosomes, it was noticeable that the protein composition of these extracellular vesicles was related to the cell of origin. Initial studies demonstrated that reticulocyte-derived exosomes were enriched with the transferrin receptor (1, 2). Several studies regarding exosome biogenesis have established that these vesicles are formed from several cycles of inward budding of the limiting membrane of late endosomes to form multi-vesicular bodies (MVB), in a process assisted by the endosomal sorting complex for transport (ESCRT) (3). Additionally, exosomes can be also originated from an ESCRT-independent mechanism, in which sphingolipids–concentrated in membrane microdomains—are converted to ceramide inducing membrane budding (4). Subsequently, the membrane of MVB fuses with the plasma membrane releasing the exosomes to the extracellular milieu (3). The transport of the MVB to the plasma membrane is mainly mediated by Rab GTPases: Rab11, Rab27a, Rab27b and rab35 (3, 5). While the fusion of MVB membrane and plasma membrane is assisted by proteins of the soluble-NSF-attachment-protein-receptor (SNARE) complex, localized in the plasma membrane (pSNAREs) and in the vesicle membrane (VAMP7) (6). The mechanisms mediating exosome cargo remain unknown. The presence of cellular chaperons such as Hsc-70 and Hsp-90 in exosomes, suggests that protein loading could be mediated by the interaction of chaperons with the exosomal membrane (3). Additionally, several members of ESCRT-complex and other

[1] This chapter is partially presented in Diaz G, Wolfe LM, Kruh-Garcia NA, Dobos KM. Changes in the Membrane-Associated Proteins of Exosomes Released from Human Macrophages after *Mycobacterium tuberculosis* Infection. Sci Rep. 2016; 6:37975.

proteins involved in exosome biogenesis, such as TSG101, Vsp4, and Allix, are normally found in exosomes (7). Further studies demonstrated that exosomes from different cell types carry cell/tissue-specific proteomic signature (7, 8). A comparative study analyzed the proteomic composition of exosomes from urine, mast cells and different colorectal cancer cell types, identifying that, 20 proteins are exclusively present in exosomes from colorectal cancer cells and seven proteins are exclusively present in urine derived exosomes (9). Other studies showed that exosomes derived from professional antigen presenting cells carry MHC II molecules (10) and CD86 (B7.2) (11). In a similar way, the composition of exosomes can be affected by pathological conditions.

Exosomes isolated from urine of prostate cancer patients were loaded with transcripts encoding the prostate biomarkers PCA-3 and TMPRSS2:ERG (12). As stated in Chapter 1, one scenario in which exosome composition changes is during intracellular infections; exosomes can be modified by carrying pathogen-derived molecules (13-15) and second, the host exosomal proteome could change as a result of the infectious process (16).

Regarding changes in the host proteome of cells infected with *M. tuberculosis,* it is important to consider that during the intracellular infection, several organelles such as: mitochondria, endoplasmic reticulum, and the phagosome, have shown proteomic changes (17-19). Particularly, *M. tuberculosis* alters phagosomal maturation by several ways including: 1. inhibition of the "Rab conversion" a process characterized by the shifting from a phagosome enriched with Rab5 to Rab7 (20), 2. stimulation the fusion of early endosome with the phagosome, in a process mediated by the bacterial phosphatidylinositol mannoside (PIM) (21), and 3. inhibition of the phagosome-lysosome fusion affecting the regulatory trafficking molecule, phosphatidylinositol-3 phosphate (PIP-3) by either: the mycobacterial acid

phosphatase SapM (19) and/or lipoarabinomannan (LAM) (22). Considering the exosome plasticity regarding protein content under pathological conditions and the evident effect of *M. tuberculosis* on the endocytic pathway of infected macrophages (MΦ), we hypothesized that intracellular *M. tuberculosis* infection will impact the protein composition of exosomes. Additionally, since protein localization within the exosome can play an important role in function of these vesicles, we implemented a novel biotinylation scheme to differentiate the surface-exposed proteins within the exosome membrane from those protected within the vesicle. This work is presented as a proof-of-concept that intracellular *M. tuberculosis* affects the protein composition of exosomes released from host cells. In view of that, the host protein composition of exosomes from individuals infected with *M. tuberculosis* could be evaluated as a source of biomarkers for the different stages of the *M. tuberculosis* infection.

2.2 Materials and methods

2.2.1 Human monocytes growth and activation

THP-1 human monocytes (American Type Culture Collection, ATCC/TIB-202) were cultured at 37 °C at 5% CO2 in complete RPMI (cRPMI) media. This media contained RPMI 1640 base medium (ATCC) supplemented with 2-mercaptoethanol (Gibco) at a final concentration of 0.05 mM and 10% exosome-depleted fetal bovine serum (EXO-FBS, SBI). Monocytes (2.5 × 105 cells/ml) were activated to MΦ using 200 nM of Phorbol 12-myristate 13-acetate (PMA) (Sigma-Aldrich) for 72 h. After the activation, unbound cells and remaining media were removed, and the adherent cells were washed three times using phosphate buffer saline (PBS).

The cells were incubated in cRPMI for 12 to 14 hours before infection.

2.2.2 *M. tuberculosis* strain and macrophage infection

All procedures involving live *M. tuberculosis* were completed in a biosafety level 3 laboratory at Colorado State University. Infectivity stocks of *M. tuberculosis* strain H37Rv were thawed and spun down at 1,200 × *g* for 10 minutes. The pellet was suspended in cRPMI and bath-sonicated for 1 minute to disrupt bacteria clumps. The THP-1 MΦ previously prepared were infected with *M. tuberculosis* at a 1:5 ratio (MΦ: Bacteria) for 4 h at 37 °C/5% CO_2. After infection, the THP-1 MΦ were washed three times with PBS and then 25 ml of fresh cRPMI was added. The cells were returned to culture conditions for 24 h for the production of exosomes. Identical flasks of THP-1 MΦ were treated in the same way but without *M. tuberculosis* to produce control exosomes. Every experiment with *M. tuberculosis* infected and uninfected THP-1 MΦ was done in triplicate and three independent experiments (at different days) were done as biological replicates. In order to obtain the best number of viable cells after exosome harvesting, the infection model of THP-1 MΦ was standardized by testing: 1) different concentrations of monocyte for activation, and 2) different times of incubation for exosome production.

2.2.3 Exosome purification

Approximately, 25 ml of supernatant from infected and control THP-1 MΦ were collected after 24 h of infection and filtered through a 0.2 μm membrane to remove cellular debris, potential membrane fragments from lysed cells, large vesicles (>500 nm), and whole bacteria. The collected material was filtered and concentrated using an Amicon centrifugal filter unit with a molecular weight cut-off (MWCO) of 100 KDa (EMD Millipore) to 2 ml to remove most of the soluble proteins. The exosome-rich retentate was diluted to 15 ml with PBS and filtered again through the 100 KDa to elute remaining soluble proteins. After this, the retained sample was

centrifuged at 18,000 × *g* for 30 min to pellet larger vesicles. The resultant supernatant was mixed with 400 µl of ExoQuick-TC (SBI) and incubated at 4 °C overnight to precipitate exosomes. After this, the mix was centrifuged at 2,600 × *g* for 30 min. The exosome rich pellet was suspended in 1 ml of PBS and the total protein concentration was measured using the microbicinchoninic acid assay (microBCA, Thermo Scientific). The exosomes were aliquoted (50 µg/vial) and stored at −20 °C until further analysis.

2.2.4 THP-1 MΦ viability test after infection with *M. tuberculosis*

To guarantee the accurate comparison between exosomes from *M. tuberculosis*-infected and control cells, the viability of cells in both conditions, was evaluated after exosome collection. After each experiment, the cells were washed with PBS and detached from the flask using 5 ml of 0.25% trypsin-EDTA (Gibco) at 25 °C for 10 min. Five ml of cRPMI were added, to neutralize the trypsin reaction, then, the cells were centrifuged at 600 × *g* per 10 min and washed with cRPMI once. Afterward, cells were resuspended in cRPMI to perform the viability assay Alamar Blue (Invitrogen). For this, 100 µl of control and infected cell suspension were added to a 96 wells plate in triplicate. Subsequently, 10 µl of Alamar Blue reagent was added to each well and the plate was incubated at 37 °C/5%CO_2 for 4 h in the dark. After incubation, the plate was read in spectrophotometer at 570 nm and 600 nm. The viability was calculated as a function of the amount of resazurin that was reduced to resorufin for the metabolically active cells. To do this the percentage of reduction of Alamar blue was calculated using the following equation:

$$\%\text{Reduced} = \frac{(\varepsilon_{OX})\lambda_2 A\lambda_1 \; - \; (\varepsilon_{OX})\lambda_1 A\lambda_2}{(\varepsilon_{RED})\lambda_1 A'\lambda_2 \; - \; (\varepsilon_{RED})\lambda_2 A'\lambda_1}$$

λ_1: 570; λ_2: 600; $A\lambda_1$: Absorbance of test well at 570 nm; $A\lambda_2$: Absorbance of test well at 600 nm; $A'\lambda_1$: Absorbance of negative control well at 570 nm; $A'\lambda_2$: Absorbance of negative control well at 600 nm; $(\varepsilon_{OX})\lambda_1$: Molar extinction coefficient for the oxidized form at 570 nm = 80,586; $(\varepsilon_{OX})\lambda_2$: Molar extinction coefficient for the oxidized form at 600 nm = 117,216; $(\varepsilon_{RED})\lambda_1$: Molar extinction coefficient for the reduced form at 570 nm = 155,677; $(\varepsilon_{RED})\lambda_2$: Molar extinction coefficient for the reduced form at 600 nm = 14,652. The viability of the infected cells was reported relative to the control cells. Negative control only contains media and Alamar blue reagent.

2.2.5 Confirmation of *M. tuberculosis* infection after exosome purification

Ten µl of infected cells suspension were pipetted onto a microscope slide, fixed with paraformaldehyde 4% for 24 h, stained with the Kinyoun acid-fast method and microscopically evaluated. Then, the remaining infected THP-1 MΦ were pelleted and lysed using 0.05% SDS for 3 min. The SDS treatment was enough to lyse the THP-1 MΦ but not the intracellular bacteria. The resulting suspension was centrifuged at 1,200 × *g* for 10 min and the pellet was diluted in Middlebrook 7H9 medium. Seven 10-fold serial dilutions were plated on 7H11 quad-plates and incubated at 37 °C for 3 weeks for colony forming units (CFU) enumeration to verify the number of infecting bacteria after exosome production.

2.2.6 Characterization of exosomes

2.2.6.1 Light scattering analysis

The concentration and size distribution of the exosomes were evaluated by nanoparticle tracking analysis (NTA) using the NanoSight NS300 (Malvern Instruments). The instrument utilizes a laser that passes through the sample and the particles in suspension scatter light. The Brownian motion of the particles is used for the software to determine the size of each particle. The analysis was standardized at 5 µg/ml of total protein for each exosome-sample. Each sample was analyzed in triplicate.

2.2.6.2 Western blot (WB) analysis

Exosome samples (50 μg) were resuspended in Laemmli buffer (which contains: 2-Mercaptoethanol 5%, Bromophenol blue 0.01%, Glycerol 10%, SDS 2% and Tris-HCl, 63 mM) and heated at 100 °C for 5 minutes. Then, the samples were resolved in a polyacrylamide gel electrophoresis using NuPAGE Novex 4–12% Bis-Tris Gel (Life Technologies) and transferred to a nitrocellulose membrane, 0.2 μm (Bio-Rad). Afterward, the membrane was blocked with bovine serum albumin 2% for 1 h, washed, and incubated with the primary antibody for 1 h. Subsequently, the conjugated secondary antibody was added and finally the membrane was exposed to the developer reagent. For exosome characterization the primary antibodies: anti-CD63 (SBI), anti-CD81 (SBI) and anti-Rab5B (A-20) (Santa Cruz Biotechnology) were used. For validation of the proteomics results, the primary antibodies: anti-Coronin 1 C (G-R2) (Santa Cruz Biotechnology), anti-Moesin (Life Technologies), anti-Vimentin Antibody (9E7E7) (Santa Cruz Biotechnology) and anti-HSP 90 (F8) (Santa Cruz Biotechnology) were used. Two different HRP-conjugated antibodies were used: Goat Anti-Rabbit F(ab)2 fragment (Thermo Scientific) and Goat anti-Mouse IgG (H + L) (Thermo Scientific) depending on the source of the primary antibody. For the WB detecting moesin and HSP 90, the colorimetric substrate 4-Chloro-1-Naphtol (Sigma) was used. For the analysis of the other proteins the chemiluminescent substrate Super Signal West Pico (Thermo Scientific) was used. WB images and band intensities were determined using the Chemi-Doc XRS+ with Image Lab software version 3.0 (Bio-Rad). For some of the assays, after transferring the protein to the nitrocellulose membrane, a Ponceau S stain was done to verify the efficiency of the transfer. For this, the membrane was soaked in Ponceau S solution (Ponceau S 0.2%, Acetic acid 3% in distilled water) for 5 minutes, washed with distilled water for approximately 5 minutes to remove unspecific staining and the scan of

the membrane was recorded. After this, the membrane was washed with PBS 1X for 5 minutes, 3 times. Then, the WB procedure was followed as described above.

2.2.6.3 Transmission electron microscopy

Exosome samples from *M. tuberculosis*-infected and control cells were prepared following the protocol described by Théry *et al.*, (23) with some modifications. Briefly, exosome samples were fixed using 4% paraformaldehyde (PFA) (1:1 ratio PFA:sample) overnight at 4 °C. Then, a formvar/carbon coated grid (Electron Microscopy Science) was placed on top of a drop of 10 µl of fixed-exosomes for 20 min. The grid was then washed by transferring it to 100 µl of PBS, and then, a second fixation step was done by transferring the grid to a 50 µl drop of glutaraldehyde 1% for 5 min. After this, the grid was washed 7 times with 100 µl drop of distilled water. Next, the grid was negatively stained on top of a 50 µl drop of uranyl-oxalate for 5 min. Finally, the grid was transferred to a 50 µl drop of methylcellulose/uranyl acetate (9:1 ratio) for 10 min on ice. After this, excess of methylcellulose/uranyl acetate was removed by blotting on Whatman #1. The grids were air dried and observed in the transmission electron microscope JEOL 2100 F at 200 kV.

2.2.7 Biotinylation of exosome proteins

In order to determine the localization of the proteins either exposed to the external leaflet or in the lumen of the exosome, a double biotinylation labelling protocol was developed. Two different types of biotin reagents were used in this experiment. Sulfo-NHS-Biotin (Thermo Scientific), containing a shorter, 13.5 Å spacer arm biotin was used, to label the proteins in the lumen or in the internal leaflet of the exosomal membrane and Sulfo-NHS-LC-LC-Biotin (Thermo Scientific) was used, to label the proteins exposed to the external leaflet of intact exosomes and contains a larger, 30.5 Å spacer arm between the biotin and amine reactive linker.

The size of this linker helps to overcome steric hindrance and increases labeling efficiency at the crowded exosome surface. Two hundred µg of intact exosomes were mixed with 10 mM Sulfo-NHS-LC-LC-Biotin at room temperature for 30 min. Four conditions were taken into account during this experiment: (a) an excess of Sulfo-NHS-LC-LC-Biotin was used to favor a complete saturation of exposed lysine residues and potential N-terminus, (b) the presence of the sulfonate group in Sulfo-NHS-LC-LC-Biotin blocks the reagent from penetrating the exosomal membrane, (c) Sulfo-NHS-LC-LC-Biotin has an spacer arm of 30.5 Å which improves the biotinylation of proteins in their natural conformation, and (d) amino acids labeled with Sulfo-NHS-LC-LC-Biotin will have an increase in mass of 452 Da. After incubation, the excess of Sulfo-NHS-LC-LC-Biotin was removed using a 10 KDa MWCO filtration device. Biotinylated exosomes were washed with 10X volumes of 1X PBS and concentrated to a final volume of 30 µl. Biotinylated exosomes were lysed with 300 µl of RIPA buffer (Thermo Scientific) for 1 h, followed by three freeze and thaw cycles. After lysis, buffer exchange was done to replace RIPA buffer with 1X PBS; RIPA buffer contains primary amines that interfere with the next biotinylation step. Lysed-biotinylated exosomes were exposed to 10 mM Sulfo-NHS-Biotin (Thermo Scientific), at room temperature for 30 min, for labeling of remaining free amines. Excess of biotin was removed as mentioned above. Peptides modified with the second biotin will have an increase in mass of 226.3 Da.

2.2.8 In gel digestion of exosomal proteins

2.2.8.1 Unlabeled exosomes

Exosomes (50 µg) were mixed with Laemmli SDS-PAGE buffer and heated at 100 °C for 5 minutes. After this, the samples were resolved in a NuPAGE Novex 4–12% Bis-Tris Gel (Life Technologies) for 5 min. Then, the gels were stained with Coomassie Blue (Invitrogen) for 1 h to

visualize the localization of the proteins and destained briefly in water to clarify protein bands. Afterwards, the entire lane of gel containing the proteins was excised and cut into 1 mm^3 pieces that were transferred to an Eppendorf tube and mixed with distaining solution (60% acetonitrile (ACN) in ammonium bicarbonate 0.2 M) for 1 hour at 37 °C, twice. After distaining, the gel pieces were vacuum dried. Dried gel pieces were mixed with trypsin (Roche) in 0.2 M ammonium bicarbonate at a ratio of 50:1 (sample: trypsin) at 37 °C, overnight. The next day, the tryptic peptides were extracted by adding 100 µl of 60% ACN, 0.1% trifluoroacetic acid in HPLC-grade water and incubated at 37 °C for 1 hour, twice. The extracted peptides were vacuum dried, suspended in Solvent A (3% ACN and 0.1% formic acid in HPLC-grade water), centrifuged at 13,000 × g for 5 min to pellet larger debris, the supernatant was carefully transferred into auto-sampler vials (Agilent technologies), and stored at −20 °C until analysis by LC-MS/MS.

2.2.8.2 Biotinylated exosomes

Labeled exosome lysates (50 µg) were subject to gel electrophoresis and staining as described above. The biotin label binds to the free amine group of lysine residues, thus, interfering with one of the cleavage sites for trypsin; as an alternative, the endoproteinase AspN (Roche) was used for protein digestion. After obtaining the dried and destained gel pieces as described above, protein samples were reduced with 5 mM dithiothreitol (Sigma) for 20 min at 50 °C and alkylated with 15 mM iodoacetamide (Sigma) at 25 °C for 15 min in the dark, following the AspN manufacturer's recommendations. Then, proteins were digested with AspN in 0.2 M ammonium bicarbonate using a 50:1 ratio (w/w-sample: enzyme). Subsequent steps for peptide extraction after enzymatic digestion were similarly performed as described before.

2.2.9 Liquid chromatography-tandem mass spectrometry (LC-MS/MS)

Approximately 1 μg of digested peptides for each sample was injected using an EASY nanoLC-II system (Thermo Scientific, San Jose, CA). Peptides were purified and concentrated using an on-line enrichment column (EASY-Column, 100 μm ID × 2 cm ReproSil-Pur C18). Subsequent chromatographic separation was performed on a reverse phase nanospray column (EASY-Column, 3 μm, 75 μm ID × 100 mm ReproSil-Pur C18) using a 90 min linear gradient from 5–45% solvent B (100% ACN, 0.1% formic acid) at a flow rate of 400 nanoliters/min. Peptides were eluted directly into the mass spectrometer (Thermo Scientific Orbitrap Velos). The instrument was operated in Orbitrap-LTQ mode where precursor measurements were acquired in the Orbitrap (60,000 resolution) and MS/MS spectra (top 20) were acquired in the LTQ ion trap with a normalized collision energy of 35%. Mass spectra were collected over a m/z range of 400–2000 Da using a dynamic exclusion limit of 2 MS/MS spectra of a given peptide mass for 30 s (exclusion duration of 90 s). Compound lists of the resulting spectra were generated using Xcalibur 2.2 software (Thermo Scientific) with an S/N threshold of 1.5 and 1 scan/group.

2.2.10 Data analysis

Tandem mass spectra were extracted, charge state deconvoluted and deisotoped by ProteoWizard (MSConvert version 3.0). Raw data files were converted to mzXML format and submitted to the Sorcerer2 integrated data analysis platform (Sage-N Research, version 5.0.1); subsequent MS/MS analysis was performed using SEQUEST (Sage-N Research, Milpitas, CA, USA; version v. 3.5). SEQUEST was set up to search the Uni-Prot Homo sapiens Reference Proteome (ID: UP000005640, 70076 entries) assuming the enzymatic digestion with trypsin (after Arg or Lys) or AspN (before Asp or Glu) depending on which enzyme was used. SEQUEST was searched with a fragment ion mass tolerance of 1.00 Da and a parent ion tolerance of 50 PPM.

Oxidation of methionine (+15.99) and carbamidomethyl of cysteine (+57.02) were specified in SEQUEST as variable modifications for the unlabeled experiment. Biotin (+226) and LC-LC Biotin (+452) in lysine and N-termini were also included as variable modification for the labeling experiment. Scaffold (version Scaffold_4.5.1, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identification thresholds were set such that a peptide FDR of 1% and a peptide confidence threshold of 95% was achieved based on hits to the reverse database (24). Protein identifications were accepted if they could be established at greater than 99.0% probability to achieve an FDR less than 1.0% and contained at least 2 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm (25). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principle of parsimony. Proteins were annotated with GO terms from NCBI (downloaded Dec 31, 2015) (26). The mass spectrometry proteomics data were deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository (27), with the dataset identifier PXD004062 and DOI 10.6019/PXD004062. The presence of membrane associated proteins were also explored using the free online software TMHMM (Version 2.0) (28, 29). Differences in protein abundances between exosomes from *M. tuberculosis*-infected MΦ versus control cells were evaluated by t-test, using the normalized spectral abundance factor (NSAF) (30). P values < 0.05 were accepted as statistically significant. For validation of the proteomics results a subset of proteins that were significantly higher in exosomes from infected cells were evaluated by western blot as described in section 2.2.6.2.

2.3 Results and discussion

2.3.1 Standardization of culture conditions to obtain exosomes from *M. tuberculosis* infected cells

Decreasing the initial number of seeded cells from $1x10^6$ to $5x10^5$ monocytes/ml improved the number of viable cells obtained for exosome production (Figure 2.1). We hypothesized that a lower number of cells in the flask allowed a better attachment of the growing MΦ.

2.3.2 General characterization of exosomes

The Nanoparticle Tracking Analysis (NTA) allowed the quantification and size measurement of vesicles in suspension in five micrograms of isolated exosomes. The nanovesicle concentration when normalized by total protein was not statistically different between samples, ranging from 2×10^8 to 6×10^8 particles per ml (t-test, $p = 0.299$) (Figure 2.2a). Exosomes from infected MΦ appeared to be slightly larger; however, the size difference was not statistically significant (t-test, $p = 0.236$) (Figure 2.2b).

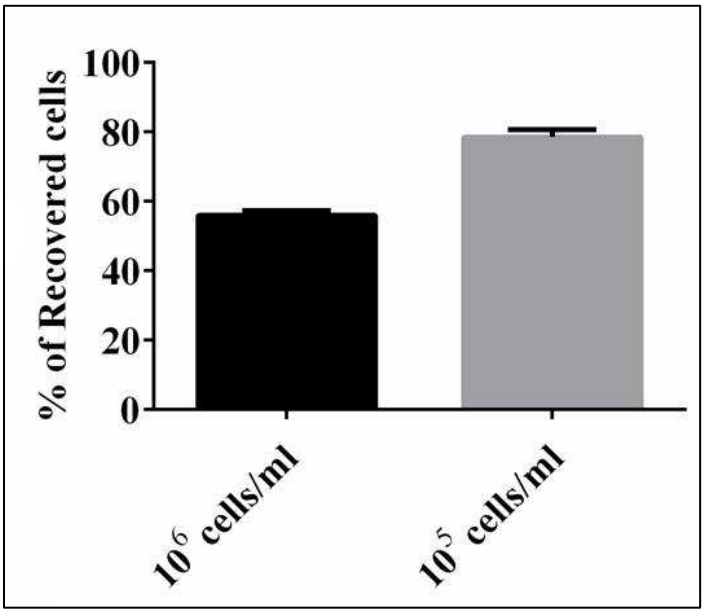

Figure 2.1 Percentage of recovered THP-1 cells after 72 hours of activation. The proportion of recovered cells is significantly higher (t-test, p=0.006) when the initial number of monocytes is 50% lower. Results from two independent experiments.

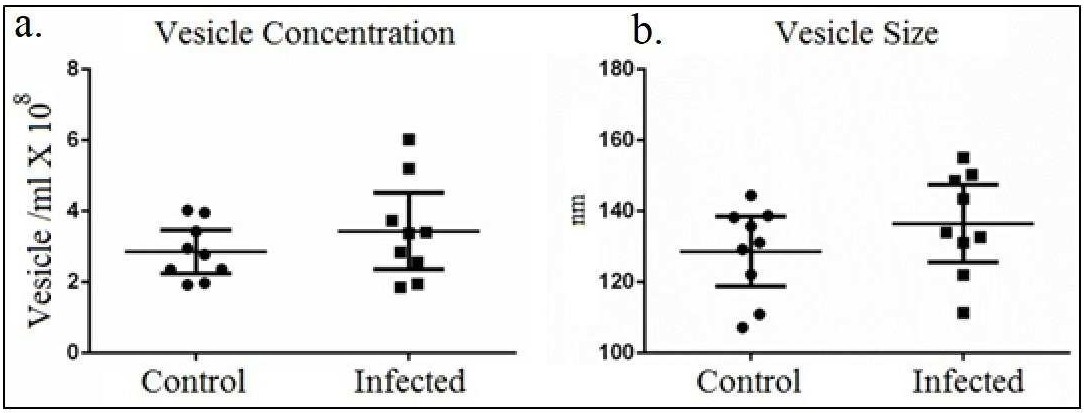

Figure 2.2 Concentration (a) and size (b) distribution of the vesicles obtained from *M. tuberculosis*-infected cells and uninfected controls. Results from 3 technical replicates from three independent experiments. Five micrograms of protein per sample were evaluated.

In addition to size and concentration, the presence of the exosome hallmark proteins: CD63, CD81, Hsp70 and Rab-5B was confirmed by western blot in all biological replicates (Fig. 2.3 a/b). Abundance of these proteins was invariant amongst all samples, except for Hsp70 which appeared more abundant in exosomes from infected cells (Figure 2.3a). A major limitation with most of the protocols used for exosome isolation is the loss of integrity of the vesicles during the process. Here, the integrity of the vesicles was confirmed by electron microscopy (Figure 2.3c). Overall, these findings confirmed that the nanovesicles in this study were consistent with exosomes.

To determine the viability of *M. tuberculosis*-infected cells and controls cells after exosome collection, we did an assay based on the reducing capacity of viable cells. The proportion of viable infected cells relative to their corresponding control ranged from 92% to 114% confirming that the viability of infected cells was comparable to that from controls.

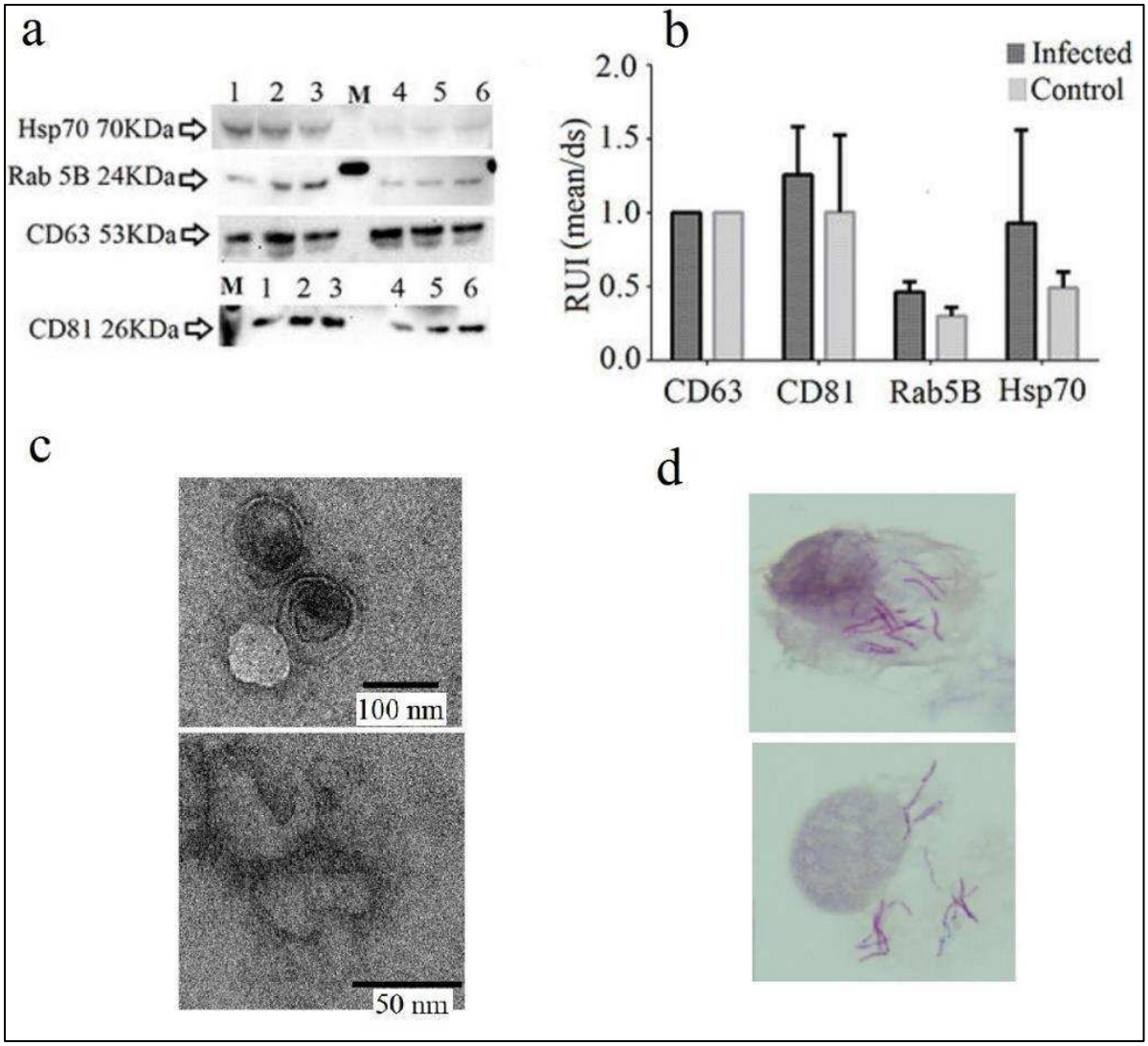

Figure 2.3 Characterization of exosomes. a. Western blot showing the presence of CD63, Hsp70, CD81 and Rab 5B in exosomes. b. Densitometry analysis of the relative abundance of each protein. c. Representative TEM images of exosomes from infected cells (upper picture) and control cells (lower picture). d. THP-1 MΦ infected with *M. tuberculosis* stained after exosome collection 100X magnification (Modified Kinyoun and hematoxylin staining). Exosomes from infected cells, lanes: 1, 2 and 3. M: Molecular weight marker. Exosomes from control cells lanes: 4, 5 and 6. RUI: relative units of intensity.

Finally, we confirmed the extent of bacterial infection after exosome production microscopically (Figure 2.3d) and by bacterial enumeration. The average number of colony forming units (CFUs) was $1.05 \times 10^7 \pm 5 \times 10^6$ CFU/ml. This average represents the 33.8% resident bacilli from the initial bacterial inoculum. Both results allowed us to confirm that the isolated exosomes were produced from *M. tuberculosis*-infected cells.

2.3.3 Proteome of exosomes released from MΦ infected with *M. tuberculosis*

A total of 355 proteins were identified in the exosomes from *M. tuberculosis* infected and control cells (a comprehensive list of all the proteins identified can be found in the publication of Diaz *et al.* (16)). More than 60% of the identified proteins are predicted to be membrane-associated, and the majority of them, are involved in binding, immunological and metabolic processes according to the Gen Ontology-GO annotations from the NCBI database (Figure 2.4a/b). These processes have been previously associated with potential roles for exosomes (31-33).

2.3.4 Comparative proteomic analysis reveals significant differences between exosomes from infected and control cells

As we predicted, the infection of MΦ with *M. tuberculosis* impacted the protein composition of exosomes. The relative abundance of each protein was determined using the spectral counts divided by the length of each individual protein (Spectral Abundance Factor-SAF), then, each SAF is divided by the sum of the SAF for all proteins in the experiment to obtain the Normalized SAF (NSAF) as previously described (30). This normalization process allowed the accurate comparison of individual protein abundances among the two sample categories: infected versus control. Forty-one proteins were significantly more abundant in exosomes from infected cells (Table 2.1), a subset of these were confirmed by western blot (Figure 2.5), including: HSP90, vimentin, Coronin 1 C and moesin. Previous studies have shown that some of these proteins play important roles during *M. tuberculosis* infection. Shekhawat *et al.*, showed that human HSP proteins (including Hsp90) were increased in sera from individuals with a high risk of being latently infected with *M. tuberculosis*, suggesting HSP proteins as potential biomarkers for LTBI (34). We observed that HSP90 was significantly higher in exosomes from infected cells. In this study, we also found that infected MΦ produced exosomes with a higher concentration of

vimentin compared with control cells. Vimentin is a ligand for NKp46, a receptor in Natural Killer (NK) cells.

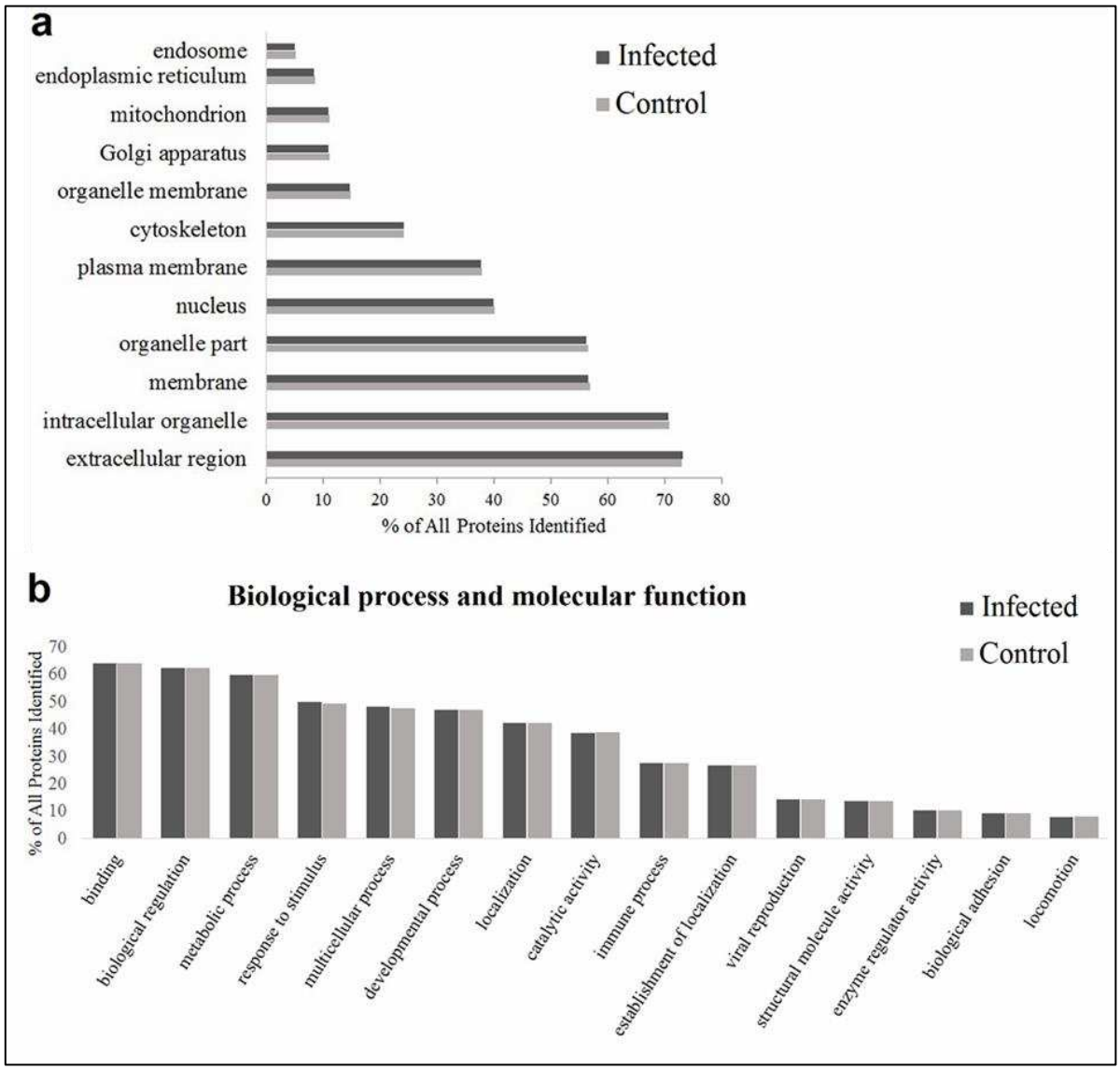

Figure 2.4 Function and localization of the proteins found in exosomes. a. Subcellular localization of most of the proteins found in the MΦ-derived exosomes. b. The top 15 identified molecular functions and biological process associated with the proteins found in exosomal samples. Information obtained directly from the NCBI database.

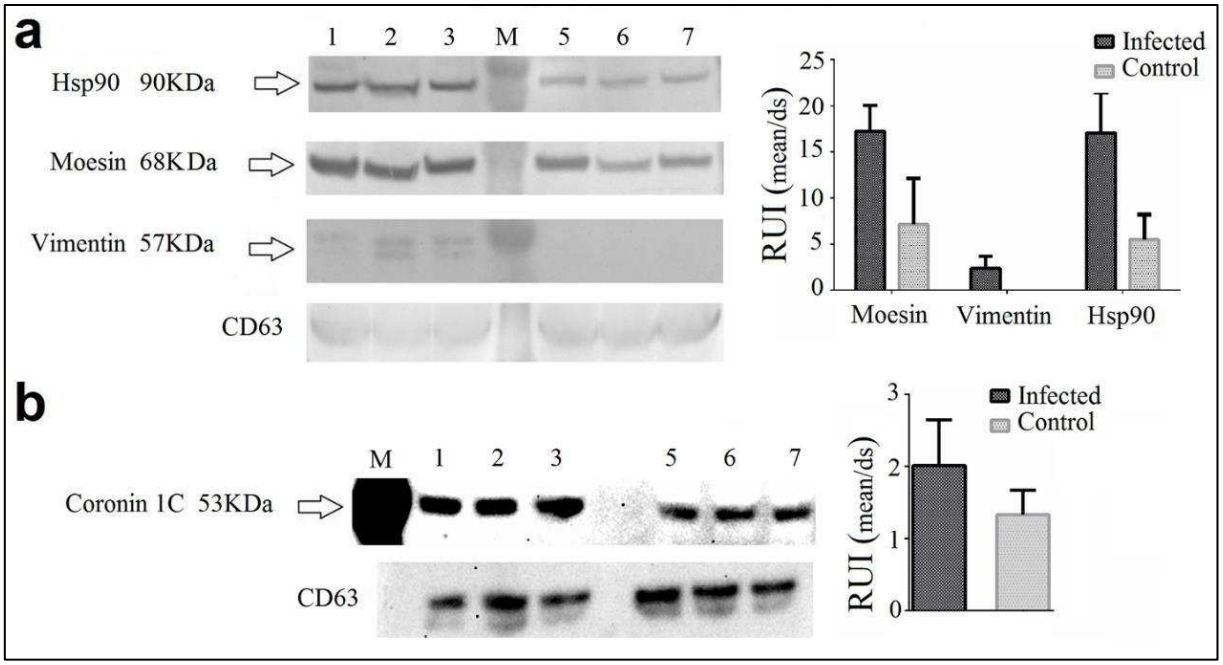

Figure 2.5 Western blot validation for proteins significantly more abundant in exosomes from *M. tuberculosis* infected cells by LC-MS/MS. a. Hsp90, Moesin and Vimentin were detected using a chromogenic substrate. The intensity of the bands was evaluated relative to the intensity of CD63. b. Coronin 1C was detected using a chemiluminescent substrate. The intensity of the bands was evaluated relative to the intensity of CD63. Results from three independent experiments. Exosomes from infected cells, lanes: 1, 2 and 3. M: Molecular weight marker. Exosomes from control cells lanes: 5, 6 and 7. RUI: relative units of intensity.

The interaction between NKp46 and vimentin mediates the lysis of *M. tuberculosis*-infected NK cells (35). Exosomes enriched with vimentin could interfere in the interaction of NKp46 with the cell membrane associated vimentin (from MΦ), thus delaying the killing of *M. tuberculosis* infected cells. In this way, intracellular mycobacteria could be modulating the loading of vimentin in exosomes as a defense mechanism. Further investigations into this phenomenon could give us new insights about the complex host-pathogen interaction in TB. Lastly, L-amino acid oxidase (LAAO) was more abundant in exosomes from *M. tuberculosis*-infected MΦ in our study. This protein plays important roles in the innate immune response acting as an antibacterial enzyme that catabolizes the deamination of L-amino acids producing H_2O_2 and ammonia (36). Our results suggest that *M. tuberculosis*-infected cells could be using exosomes to export endogenous antimicrobial molecules as a defense mechanism. This phenomenon has been

previously observed with IFNα that could be transported between cells via exosomes (31). Collectively, these changes in exosomes because of *M. tuberculosis* infection reveals important information regarding the host-pathogen interaction.

Table 2.1 Proteins significantly more abundant between exosomes from infected and control MΦ.

Identified Proteins	**Infected NSAF**	**Control NSAF**	**p value* NSAF-inf versus NSAF-**
60S acidic ribosomal protein P0	0.064	0.000	0.00012
Coronin-1C	0.023	0.000	0.00017
Lupus La protein	0.023	0.000	0.00019
Heterogeneous nuclear ribonucleoprotein K	0.075	0.003	0.00029
Heat shock 70 kDa protein 4	0.013	0.000	0.00031
Alanine--tRNA ligase, cytoplasmic	0.006	0.000	0.00035
Calreticulin	0.017	0.000	0.001
Protein disulfide-isomerase A3	0.040	0.000	0.002
L-amino-acid oxidase	0.018	0.000	0.003
Moesin	0.151	0.062	0.0032
Nucleolin	0.063	0.007	0.0032
Vimentin	0.251	0.072	0.0034
Protein disulfide-isomerase A6	0.046	0.003	0.0035
Spliceosome RNA helicase DDX39B	0.027	0.000	0.0039
Fermitin family homolog 3	0.046	0.002	0.0047
Programmed cell death 6-interacting protein	0.005	0.000	0.0047
S-adenosylmethionine synthase isoform type-	0.029	0.000	0.0048
Glyceraldehyde-3-phosphate dehydrogenase	0.293	0.201	0.0059
ATP-dependent RNA helicase A	0.005	0.000	0.0068
60 kDa heat shock protein, mitochondrial	0.013	0.000	0.0082
Cytosol aminopeptidase	0.041	0.000	0.0084
Ubiquitin-like modifier-activating enzyme 1	0.056	0.007	0.0089
ITIH4 protein	0.011	0.000	0.01
Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A alpha	0.011	0.002	0.011
Tryptophan--tRNA ligase, cytoplasmic	0.031	0.000	0.011
Transketolase	0.082	0.015	0.012
Zyxin (Fragment)	0.007	0.000	0.012
Heat shock protein HSP 90-beta	0.361	0.221	0.014
Tyrosine--tRNA ligase, cytoplasmic	0.014	0.000	0.017

6-phosphogluconate dehydrogenase, decarboxylating	0.075	0.021	0.024
X-ray repair cross-complementing protein 6	0.061	0.004	0.026
78 kDa glucose-regulated protein	0.109	0.047	0.028
Eukaryotic initiation factor 4A-I	0.115	0.038	0.028
Thrombospondin-4	0.011	0.002	0.028
Bifunctional purine biosynthesis protein	0.028	0.000	0.028
Staphylococcal nuclease domain-containing protein 1	0.012	0.001	0.031
Heat shock cognate 71 kDa protein	0.443	0.273	0.036
Integrin beta-1	0.006	0.000	0.046
UDP-glucose 6-dehydrogenase	0.013	0.000	0.046
Purine nucleoside phosphorylase	0.039	0.000	0.048
Lamin-B1	0.021	0.003	0.049

**p value* of the *t*-test comparing the averages of the normalized spectral abundance factor-NSAF (30) of exosomal proteins from infected and control cells. Results from three independent experiments.

2.3.5 Significant changes of the exosome membrane proteome after infection with *M. tuberculosis*

Exosomes are an important source of biomarkers for TB by virtue of the discovery of mycobacterial molecules packaged within these vesicles (14, 15, 37, 38). Here, the effect of the infection on host exosomal protein composition illustrates another potential opportunity to exploit exosomes as biomarkers of TB. In this regard, we sought to explore the differential abundance of membrane associated proteins since they represent a more accessible set of targets for downstream development of a biomarker assay. Using GO-term annotation analysis, we found that 63% (26/41) of the proteins that were significantly more abundant in exosomes from infected cells were also membrane associated. To better characterize this subset of proteins we used transmembrane helix prediction software and found that 31% of the 26 proteins contained at least one probable transmembrane domain or residues that were potentially membrane associated (Table 2.2). To further ascertain the localization of proteins significantly more abundant in exosomes from infected cells, we conducted an experiment to differentially label proteins based

on their localization within the exosome. First, intact exosomes were exposed to NHS-Sulfo-LC-LC-Biotin, a negatively charged molecule unable to permeate biological membranes, to label proteins that were exclusively facing-out of the exosomal membrane. NHS-Sulfo-LC-LC-Biotin binds amino-terminus and lysine residues, increasing the molecular mass of the labeled peptide by 452 Da. The fact that this molecule has a spacer arm of 30.5 Å improves its accessibility to folded proteins in the membrane of intact exosomes. After cleaning the excess of NHS-Sulfo-LC-LC-Biotin, the biotinylated exosomes were lysed followed by secondary labeling of proteins on the internal-face of the exosomal membrane or within the lumen of the exosome, using a different biotin reagent (Sulfo-NHS-Biotin) that increases the molecular weight of the labelled residues by 226 Da. This labeling strategy allowed for the identification and differentiation of protein populations by mass spectrometry analysis, including determination of protein domains exposed to the external side of the exosome. Our labeling studies demonstrated that 6 of 26 differentially abundant membrane proteins identified by GO-term analysis were differentially labeled with LC-LC biotin, and thus are surface exposed (Table 2.2). In addition, this technique afforded the identification of one additional surface exposed protein, a nucleoside diphosphate kinase (Ndk) (Table 2.2). Interestingly, Mycobacterium-derived Ndk has been associated with a greater survival of infected MΦ (39). Ndk metabolizes extracellular ATP which plays important roles during the inflammatory response and MΦ activation via P2X7 receptor (40). Although human derived-Ndk has been mostly related with intracellular vesicle trafficking (40), the human-derived Ndk secreted in exosomes could be acting as an ectoenzyme that modulates MΦ activation and survival in an ATP-dependent manner; favoring mycobacterial persistence. Three of the seven LC-LC biotin labeled proteins (Table 2.2) were also labeled with the shorter biotin (226 Da label) (Table 2.2); these are indicative of transmembrane proteins with both external and

internal segments. In the same way, the presence of proteins exhibiting only the small biotin implies that the protein was confined to the exosome lumen. Our biotinylation strategy represents an original approach that allows for a more detailed characterization of exosomes, specifically the identification of the peptide present in the outer leaflet of the exosomal membrane could drive targeted assays to produce epitope-specific antibodies to capture a selected population of exosomes. However, it is important to underscore the low number of peptides that were identified with our strategy, the extra weight of peptides with multiple biotin molecules could be negatively impacting the selection of those peptides for identification, additional standardization of the methods is needed.

Table 2.2 Membrane associated proteins significantly more abundant in exosomes from *M. tuberculosis*-infected cells and their biotinylation pattern showing the specific peptide labelled with LC-LC biotin.

Membrane associated protein‡	**Proteins with AA residues in transmembrane domains****	**LC-LC Biotinylated peptides**	**Biotin**
60S acidic ribosomal protein P0	X		
Coronin-1C	X		
Heterogeneous nuclear ribonucleoprotein K			
Alanine-tRNA ligase, cytoplasmic			
Calreticulin	X		
Protein disulfide-isomerase A3	X		
Moesin		EEAkEALLQASR	
Nucleolin			
Vimentin		DVRQQYESVAAkNLQEA	X
Protein disulfide-isomerase A6	X		
Fermitin family homolog 3			
Programmed cell death 6-interacting protein			
Glyceraldehyde-3-P-dehydrogenase		DNFGIVEGLMTTVHAITA	X
ATP-dependent RNA helicase A	X		
60 kDa heat shock protein, mitochondrial			
Cytosol aminopeptidase			
Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A alpha			
Transketolase	X		

Heat shock protein HSP 90-beta		ERIMkAQALR	
78 kDa glucose-regulated protein			
Eukaryotic initiation factor 4A-I		EVQkLQMEAPHIIVGTPGR	X
Bifunctional purine biosynthesis protein PURH			
Staphylococcal nuclease domain-containing protein 1			
Heat shock cognate 71 kDa protein		DPVEkALR	
Integrin beta-1	X		
Lamin-B1			
Nucleoside diphosphate kinase		ERTFIAIkP	

‡Classification based on the Go-term annotations from the NCBI database. **Based on TMHMM Server V. 2.0 prediction of transmembrane helices in proteins (AA: Aminoacids). k: designates a lysine residue with the LC-LC biotin modification.

2.4 Conclusions

Our study demonstrates that the infection with *M. tuberculosis* influences changes in the protein composition of exosomes released from infected cells. Even though our findings do not have an immediate translational application they represent the proof-of-concept that *M. tuberculosis*-infected cells will produce exosomes with a characteristic proteome. The confirmation of these phenomena in a clinically relevant sample set (i.e. TB-patient sera derived exosomes) will advance our knowledge about *M. tuberculosis*-host interactions and will offer a potential source for new TB biomarkers.

References

1. Harding C, Heuser J, Stahl P. Endocytosis and intracellular processing of transferrin and colloidal gold-transferrin in rat reticulocytes: demonstration of a pathway for receptor shedding. Eur J Cell Biol. 1984;35(2):256-63.
2. Pan BT, Teng K, Wu C, Adam M, Johnstone RM. Electron microscopic evidence for externalization of the transferrin receptor in vesicular form in sheep reticulocytes. J Cell Biol. 1985;101(3):942-8.
3. Raposo G, Stoorvogel W. Extracellular vesicles: exosomes, microvesicles, and friends. J Cell Biol. 2013;200(4):373-83.
4. Trajkovic K, Hsu C, Chiantia S, Rajendran L, Wenzel D, Wieland F, et al. Ceramide triggers budding of exosome vesicles into multivesicular endosomes. Science. 2008;319(5867):1244-7.
5. Ostrowski M, Carmo NB, Krumeich S, Fanget I, Raposo G, Savina A, et al. Rab27a and Rab27b control different steps of the exosome secretion pathway. Nat Cell Biol. 2010;12(1):19- 30; sup pp 1-13.
6. Cocucci E, Meldolesi J. Ectosomes and exosomes: shedding the confusion between extracellular vesicles. Trends Cell Biol. 2015;25(6):364-72.
7. Stoorvogel W, Kleijmeer MJ, Geuze HJ, Raposo G. The biogenesis and functions of exosomes. Traffic. 2002;3(5):321-30.
8. Simpson RJ, Jensen SS, Lim JW. Proteomic profiling of exosomes: current perspectives. Proteomics. 2008;8(19):4083-99.
9. Mathivanan S, Lim JW, Tauro BJ, Ji H, Moritz RL, Simpson RJ. Proteomics analysis of A33 immunoaffinity-purified exosomes released from the human colon tumor cell line LIM1215 reveals a tissue-specific protein signature. Mol Cell Proteomics. 2010;9(2):197-208.
10. Denzer K, van Eijk M, Kleijmeer MJ, Jakobson E, de Groot C, Geuze HJ. Follicular dendritic cells carry MHC class II-expressing microvesicles at their surface. J Immunol. 2000;165(3):1259-65.
11. Segura E, Nicco C, Lombard B, Véron P, Raposo G, Batteux F, et al. ICAM-1 on exosomes from mature dendritic cells is critical for efficient naive T-cell priming. Blood. 2005;106(1):216-23.
12. Nilsson J, Skog J, Nordstrand A, Baranov V, Mincheva-Nilsson L, Breakefield XO, et al. Prostate cancer-derived urine exosomes: a novel approach to biomarkers for prostate cancer. Br J Cancer. 2009;100(10):1603-7.
13. Bhatnagar S, Shinagawa K, Castellino FJ, Schorey JS. Exosomes released from macrophages infected with intracellular pathogens stimulate a proinflammatory response in vitro and in vivo. Blood. 2007;110(9):3234-44.
14. Giri PK, Kruh NA, Dobos KM, Schorey JS. Proteomic analysis identifies highly antigenic proteins in exosomes from *M. tuberculosis*-infected and culture filtrate protein-treated macrophages. Proteomics. 2010;10(17):3190-202.
15. Xu S, Cooper A, Sturgill-Koszycki S, van Heyningen T, Chatterjee D, Orme I, et al. Intracellular trafficking in *Mycobacterium tuberculosis* and Mycobacterium avium-infected macrophages. J Immunol. 1994;153(6):2568-78.
16. Diaz G, Wolfe LM, Kruh-Garcia NA, Dobos KM. Changes in the membrane-associated

proteins of exosomes released from human macrophages after *Mycobacterium tuberculosis* infection. Sci Rep. 2016;6:37975.
17. Saquib NM, Jamwal S, Midha MK, Verma HN, Manivel V. Quantitative Proteomics and Lipidomics Analysis of Endoplasmic Reticulum of Macrophage Infected with *Mycobacterium tuberculosis*. Int J Proteomics. 2015;2015:270438.
18. Jamwal S, Midha MK, Verma HN, Basu A, Rao KV, Manivel V. Characterizing virulence-specific perturbations in the mitochondrial function of macrophages infected with *Mycobacterium tuberculosis*. Sci Rep. 2013;3:1328.
19. Vergne I, Chua J, Lee HH, Lucas M, Belisle J, Deretic V. Mechanism of phagolysosome biogenesis block by viable *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A. 2005;102(11):4033-8.
20. Via LE, Deretic D, Ulmer RJ, Hibler NS, Huber LA, Deretic V. Arrest of mycobacterial phagosome maturation is caused by a block in vesicle fusion between stages controlled by rab5 and rab7. J Biol Chem. 1997;272(20):13326-31.
21. Vergne I, Fratti RA, Hill PJ, Chua J, Belisle J, Deretic V. *Mycobacterium tuberculosis* phagosome maturation arrest: mycobacterial phosphatidylinositol analog phosphatidylinositol mannoside stimulates early endosomal fusion. Mol Biol Cell. 2004;15(2):751-60.
22. Deretic V, Singh S, Master S, Harris J, Roberts E, Kyei G, et al. *Mycobacterium tuberculosis* inhibition of phagolysosome biogenesis and autophagy as a host defence mechanism. Cell Microbiol. 2006;8(5):719-27.
23. Théry C, Amigorena S, Raposo G, Clayton A. Isolation and characterization of exosomes from cell culture supernatants and biological fluids. Curr Protoc Cell Biol. 2006;Chapter 3:Unit 3.22.
24. Käll L, Storey JD, MacCoss MJ, Noble WS. Assigning significance to peptides identified by tandem mass spectrometry using decoy databases. J Proteome Res. 2008;7(1):29-34.
25. Nesvizhskii AI, Keller A, Kolker E, Aebersold R. A statistical model for identifying proteins by tandem mass spectrometry. Anal Chem. 2003;75(17):4646-58.
26. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25(1):25-9.
27. Vizcaíno JA, Côté RG, Csordas A, Dianes JA, Fabregat A, Foster JM, et al. The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. Nucleic Acids Res. 2013;41(Database issue):D1063-9.
28. Sonnhammer EL, von Heijne G, Krogh A. A hidden Markov model for predicting transmembrane helices in protein sequences. Proc Int Conf Intell Syst Mol Biol. 1998;6:175-82.
29. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001;305(3):567-80.
30. Zhang Y, Wen Z, Washburn MP, Florens L. Refinements to label free proteome quantitation: how to deal with peptides shared by multiple proteins. Anal Chem. 2010;82(6):2272-81.
31. Li J, Liu K, Liu Y, Xu Y, Zhang F, Yang H, et al. Exosomes mediate the cell-to-cell transmission of IFN-α-induced antiviral activity. Nat Immunol. 2013;14(8):793-803.
32. Théry C, Zitvogel L, Amigorena S. Exosomes: composition, biogenesis and

function. Nat Rev Immunol. 2002;2(8):569-79.
33. Anand PK. Exosomal membrane molecules are potent immune response modulators. Commun Integr Biol. 2010;3(5):405-8.
34. Shekhawat SD, Purohit HJ, Taori GM, Daginawala HF, Kashyap RS. Evaluation of heat shock proteins for discriminating between latent Tuberculosis infection and active Tuberculosis: A preliminary report. J Infect Public Health. 2015.
35. Garg A, Barnes PF, Porgador A, Roy S, Wu S, Nanda JS, et al. Vimentin expressed on *Mycobacterium tuberculosis*-infected human monocytes is involved in binding to the NKp46 receptor. J Immunol. 2006;177(9):6192-8.
36. Puiffe ML, Lachaise I, Molinier-Frenkel V, Castellano F. Antibacterial properties of the mammalian L-amino acid oxidase IL4I1. PLoS One. 2013;8(1):e54589.
37. Cheng Y, Schorey JS. Exosomes carrying mycobacterial antigens can protect mice against *Mycobacterium tuberculosis* infection. Eur J Immunol. 2013;43(12):3279-90.
38. Kruh-Garcia NA, Wolfe LM, Chaisson LH, Worodria WO, Nahid P, Schorey JS, et al. Detection of *Mycobacterium tuberculosis* peptides in the exosomes of patients with active and latent *M. tuberculosis* infection using MRM-MS. PLoS One. 2014;9(7):e103811.
39. Zaborina O, Li X, Cheng G, Kapatral V, Chakrabarty AM. Secretion of ATP-utilizing enzymes, nucleoside diphosphate kinase and ATPase, by *Mycobacterium bovis* BCG: sequestration of ATP from macrophage P2Z receptors? Mol Microbiol. 1999;31(5):1333-43.
40. Spooner R, Yilmaz Ö. Nucleoside-diphosphate-kinase: a pleiotropic effector in microbial colonization under interdisciplinary characterization. Microbes Infect. 2012;14(3):228-37.

Chapter 3: Walking through the spectrum of tuberculosis, what can the proteome of host exosomes tell us?

3.1 Introduction

Historically, Tuberculosis (TB) was classified as active disease or latent TB infection (LTBI), and TB patients after receiving complete treatment were declared "cured". However, the fact that some of the LTBI and "cured" individuals develop active TB (reactivation and recurrence, respectively), suggests that LTBI and "cured" stages represent a range of disease states (1). Recent studies based on positron emission tomography (PET) combined with computed tomography (CT), demonstrated that individuals with LTBI exhibited a wide range of pathological features that correlate with TB reactivation (2). Similar results were observed in patients who completed treatment, were declared "cured", and developed recurrent disease (3). In most imaging-based studies, the differentiation of each state of disease was determined by the activity of inflammatory cells, which was evaluated by the uptake of ^{18}F-fluorodeoxyglucose (^{18}F-FDG). It is possible that pathological manifestations occurring at the cellular level are reflected in serum markers with a potential to predict reactivation or treatment response. In this regard, Jacobs *et al.,* recently published that a set of plasma markers including C-reactive protein, serum amyloid protein, IP10 and the vascular endothelial growth factor among others, were significantly higher in active TB patients compared with patients with other respiratory diseases. The same proteins were significantly reduced after TB treatment completion (4). Kruh-Garcia and colleagues used a unique approach to fractionate serum by polyethylene glycol (PEG) precipitation allowing the detection of *M. tuberculosis* peptides indicative of LTBI (5, 6). PEG precipitates an exosome-rich pellet partially purified from heavily abundant serum proteins

(7, 8). A similar approach was used in Chapter 2 to demonstrate that the proteome of exosomes derived from *M. tuberculosis*-infected macrophages was significantly different compared to uninfected cells (9). Together these studies support the notion that serum exosomes from TB patients show a set of proteins reflective of the stage of infection. Proteomic mass spectrometry is a powerful tool for the discovery and subsequent evaluation of serum-derived biomarkers. Discovery frequently uses data dependent acquisition (DDA) methods commonly called shotgun proteomics. However, in DDA methods, the selection of peptides for fragmentation is based on a preselected set of conditions (normally choosing the more abundant precursors). The selected precursors are a small fraction of the total amount of peptides generated from a sample; this makes the analysis inherently stochastic, impairing the reproducibility and sensitivity of the method (10). An alternative to overcome the limitations of shotgun proteomics is data independent acquisition (DIA). In DIA, the fragmentation is independent of the precursor, which means that all peptides in a determined range of m/z (isolation window) will be fragmented in a cyclic manner across the whole LC retention time range (Figure 3.1) (11).

In 2012, Gillet *et al.,* developed a novel DIA-based proteomic analysis called Sequential Windowed Acquisition of All Theoretical Fragment Ions Mass Spectrometry (SWATH-MS). In SWATH-MS a predefined spectral library is used to get information about a targeted peptide; the DIA- derived fragmentation map is queried looking for matching information (11). The spectral library is obtained from a shotgun proteomic experiment. Ideally, the same samples that will be analyzed by SWATH-MS should be processed by LC-MS-MS in DDA mode to generate information for the libraries. To partially overcome the already known limitations of shotgun proteomics, each sample is highly fractionated and multiple injections (up to ten injections) to get a plateau in the peptide protein identification (12)

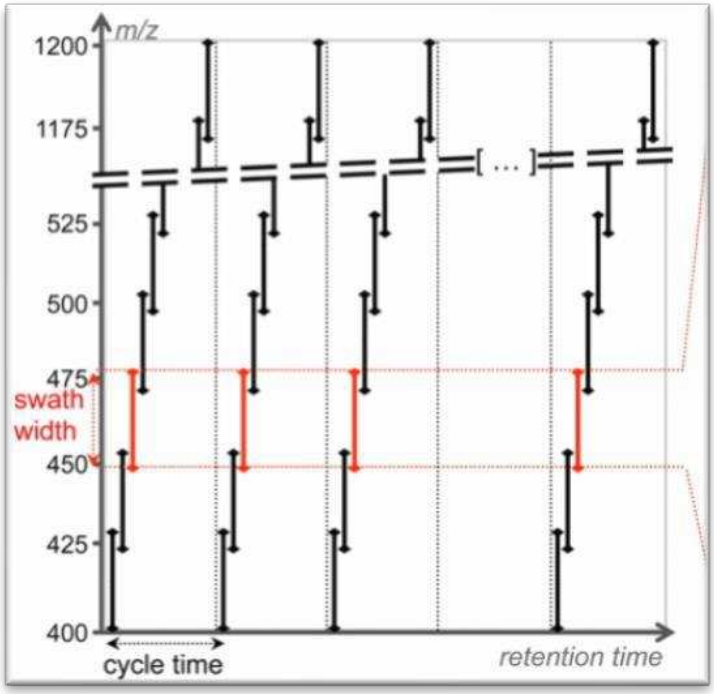

Figure 3.1 Data-independent acquisition (DIA). All precursors in the predefined SWATH (precursor isolation window) are fragmented. In the figure, the SWATH is 26 Da width and increases consecutively covering a precursor m/z range from 400 to 1,200 Da, to complete a cycle which is repeated across the LC retention time. Adapted from (11), the final version of the figure is free according to Creative Commons CC-BY license.

It is noteworthy to understand that the data that can be targeted for a DIA assay is dependent on the data present in the spectral library. A newer version of SWATH-MS known as Hyper Reaction Monitoring (HRM) showed a higher capacity for peptides/proteins identification, higher reproducibility, and higher precision to quantify and detect differentially abundant proteins when compared to shotgun proteomics (12). In HRM, the DIA fragmentation map is interrogated against the whole spectral library. The completeness of the data set generated for HRM was higher than 98% while with shotgun proteomics it was about 49%. This is particularly important because missing values challenge the statistical analysis of data sets (Figure 3.2) (12). In this chapter, we aimed to discover serum exosome proteins using HRM SWATH-MS to show significantly different abundances amongst different TB disease states.

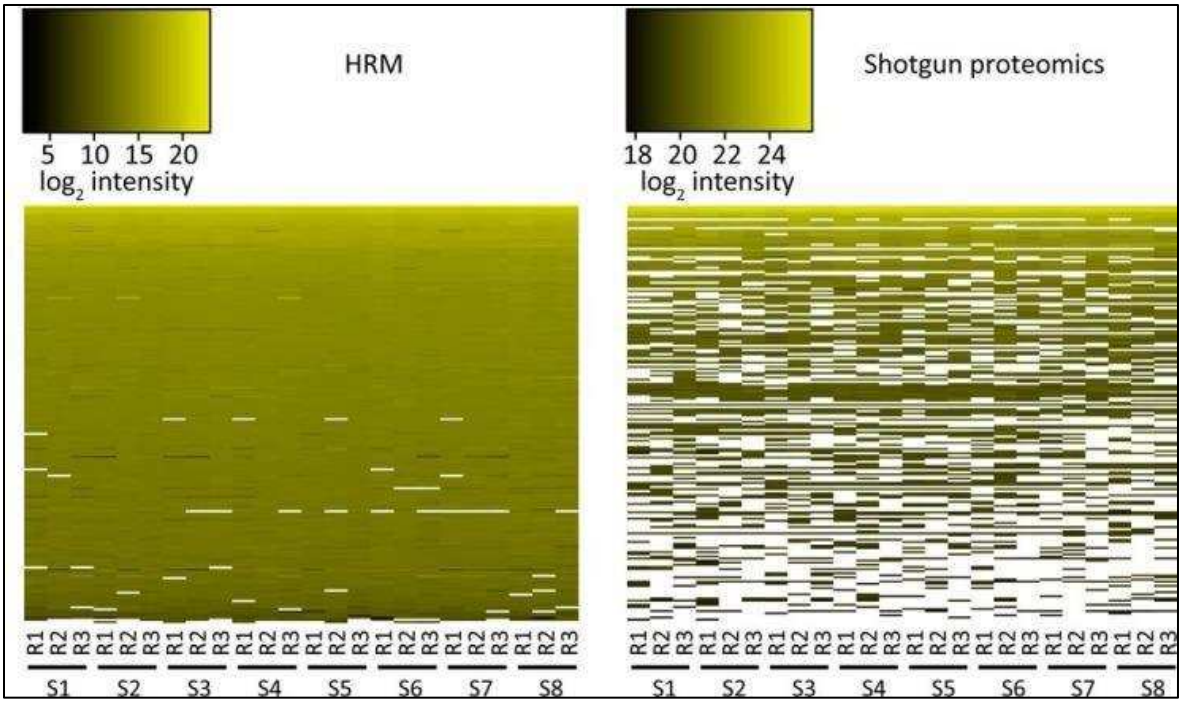

Figure 3.2 Comparison of data completion with hyper reaction monitoring (HRM) and shotgun proteomics. The intensity of the peptides is presented vertically across the 8 samples (each with 3 replicates). Adapted from: Bruderer R, *et al.* (12), the final version of the figure is free according to Creative Commons CC-BY license.

For this purpose, we analyzed three groups of samples from TB endemic regions: smear and culture positive, smear negative and culture positive, and TB suspects without microbial evidence of disease, as well as a healthy group from a non-TB endemic area (as a control).

3.2 Materials and methods

3.2.1 Sample classification

Serum samples from TB endemic regions were obtained from the Foundation for Innovative Novel Diagnostics (FIND) specimen repository (Geneva, Switzerland) and included three categories: TB patients-1 (TB-1) smear negative and culture positive, TB patient-2 (TB-2) whose sputum contained detectable *M. tuberculosis* by microscopy and culture, and TB suspects. TB suspects are defined as people who visited healthcare facilities and were found to have pulmonary symptoms indicative of TB; consequently, their sputum was collected, screened and

found to be negative for *M. tuberculosis* by either microscopy or culture evaluation of sputum samples. An additional set of samples were donated by healthy individuals from non-endemic regions (Table 3.1); these samples were collected and provided by the laboratory of Dr. David Lewinsohn at Oregon Health & Science University.

Table 3.1 Clinical classification of the samples in the study.

Sample Group (N=10 per group)	**Smear**	**Culture**	**HIV+**	**HIV-**
TB-1	positive	negative	3	7
TB-2	positive	positive	3	7
TB suspects	negative	negative	4	6
Healthy	N/A	N/A	0	10

3.2.2 Exosome enrichment from human sera samples

Sera samples were defrosted at 4 °C and spun at 18,000 X g per 15 min. twice, to pellet largerparticles. Then, the concentration of protein in the sample was quantified by Bicinchoninic acid assay (BCA, Thermo). 20 mg of sample were used for exosome isolation. Exosomes were purified using CaptoCore 700 (GE Healthcare Life Sciences). Briefly, 1.3 ml of CaptoCore 700 slurry was added to an empty column and washed with 10 ml of PBS. Samples were diluted in 300 µl of PBS and run through CaptoCore 700 column by gravity. The collected material was reapplied to the column which was washed with 500 µl of PBS. The eluate was quantified using MicroBCA (Thermo). The exosome enriched samples were dried by centrifugal evaporation under vacuum, resuspended at 2 mg/ml in a lysis salt buffer provided by Biognosys and shipped frozen to Biognosys laboratories for HRM-MS analysis.

3.2.3 Spectral library generation

Samples were reduced using Biognosys' reduction solution for 1 h at 37 °C and alkylated using Biognosys' alkylation solution for 30 min at room temperature in the dark. Subsequently,

digestion of proteins was carried out using trypsin (Promega) overnight at 37 °C at a ratio of 50:1 (protein: protease). Then, the peptides were desalted using a C18 Micro Spin-plate (The Nest Group) according to the manufacturer's instructions and dried down using a SpeedVac system. Peptides were resuspended in 15 µl solvent A (1 % acetonitrile, 0.1% formic acid (FA)) and spiked with Biognosys' HRM kit calibration peptides. Then, the samples were pooled by TB status (Healthy and TB suspects vs. TB patients group 1 and 2) by combining equal volumes of each sample, and the resulting pools were fractionated by high pH reversed-phase chromatography (HPRP). For HPRP fractionation, equal sample volumes were pooled according to sample group (Healthy/TB suspects and TB patient groups 1 and 2). The two pools were each diluted 4x in 0.2 M ammonium formate (pH 10) and applied to a C18 Micro spin column (The Nest Group). The peptides were then eluted with buffers containing 0.05 M ammonium formate and increasing acetonitrile concentrations (5, 10, 12, 14, 16, 18, 20, 22, 24, 26, 30 and 70%). The eluates were dried down, resolved in 17 µl solvent A and spiked with Biognosys' HRM kit calibration peptides prior to mass spectrometry analyses. The final peptide concentrations in all samples and fractions were determined using a UV/VIS Spectrometer (Spectro STARnano, BMG Labtech).

For the LC-MS/MS shotgun measurements, 1 µg of peptides (3 µg for fractions 5%) were injected to an in-house packed C18 column (Dr. Maisch ReproSilPur, 1.9 µm particle size, 120 Å pore size; 75 µm inner diameter, 50 cm length, New Objective) on a Thermo Scientific Easy nLC 1200 nano-liquid chromatography system connected to a Thermo Scientific Q ExactiveHF mass spectrometer equipped with a standard nano-electrospray source. Twelve fractions of each pool were measured once in shotgun MS mode (24 measurements in total). A modified TOP15 method was used as described elsewhere (13).

3.2.4 Hyper reaction monitoring/SWATH-MS

One µg of peptides per sample was injected into an identical in-house packed C18 column as the one used for shotgun analysis. Similarly, the same LC and MS systems were used. A DIA method with one full range survey scan and 14 DIA isolation windows were used. HRM mass spectrometric data were analyzed using Spectronaut 11 software (Biognosys). The false discovery rate on peptide level was set to 1%, data was filtered using row based extraction. The total peptide inventory obtained from the spectral library was searched in the DIA fragmentation maps from the entire sample set. An additional high-quality SWATH spectral library for *M. tuberculosis* developed by Schubert *et. al* 2015 (14) was used to increase the detection of *M. tuberculosis* proteins, the spectral library is available at: http://www.swathatlas.org/ LC solvents were A: 1% acetonitrile in water with 0.1% FA; B:15% water in acetonitrile with 0.1% FA. The nonlinear LC gradient was 1–52% solvent B in 60 minutes followed by 52-90% B in 10 seconds and 90% B for 10 minutes. All solvents were HPLC-grade from Sigma-Aldrich and all chemicals where not stated otherwise were obtained from Sigma-Aldrich.

3.2.5 Data analysis

The mass spectrometric data from the shotgun proteomic strategy were analyzed using MaxQuant 1.5.6.5 software (maxquant.org), the false discovery rate on peptide and protein level was set to 1%. A human UniProt.fasta database (Homo sapiens, 2015-08-28) was used together with a *M. tuberculosis* database (MtbRv_Rev_R20.fasta), allowing for 2 missed cleavages and variable modifications (N-term acetylation, methionine oxidation, lysine/arginine carbamylation, asparagine/glutamine deamidation). The HRM measurements were analyzed with the software Spectronaut 11 (Biognosys) peptide intensities were normalized using local regression normalization (15). Data post processing was carried out in the statistical package R. The heat

map was generated using Spectronaut post analysis perspective and distance was calculated using the "Manhattan" method with clustering using "ward.D". Differences in protein abundances between groups were evaluated by t-test using GraphPad Prism 7.03.

3.3 Results and discussion

3.3.1 Spectral library generation

The samples were pooled by TB status (Healthy and TB suspects vs. TB patient 1 and 2) by combining equal volumes of each sample, and the resulting pools were fractionated by high pH reversed-phase chromatography (HPRP). Twelve fractions of each pool were each measured once in shotgun MS mode (24 measurements in total). Manual inspection of the total ion current chromatograms shows stable acquisition duration over the whole gradient and strongest peaks were above expected intensity (Figure 3.3). The number of identified proteins and peptides per fraction corresponded to expected sample complexity (Table 3.2). In total 9,360 precursors and 994 proteins were included in the library. An additional spectral library specific for *M. tuberculosis* previously generated by Schubert *et al.* (14), was used simultaneously for the analysis of the DIA-fragments maps of the samples in this study. From this comprehensive spectral library, Spectronaut 11 software identified: 48,036 precursors and 3,826 proteins.

3.3.2 Proteomic profiling of exosome-enriched samples obtained from serum of patients with different TB status and healthy individuals

We generated a sequential window acquisition of all theoretical mass spectra-SWATH-MS fragment map, from each sample described in table 3.2. A total of 40 samples were randomly processed in DIA mode. Then, using Spectronaut 11 (Biognosys AG, Switzerland), the SWATH-MS maps were mined with all the peptides present in the spectral libraries. The information contained in the spectral libraries for each peptide, such as the relative intensity

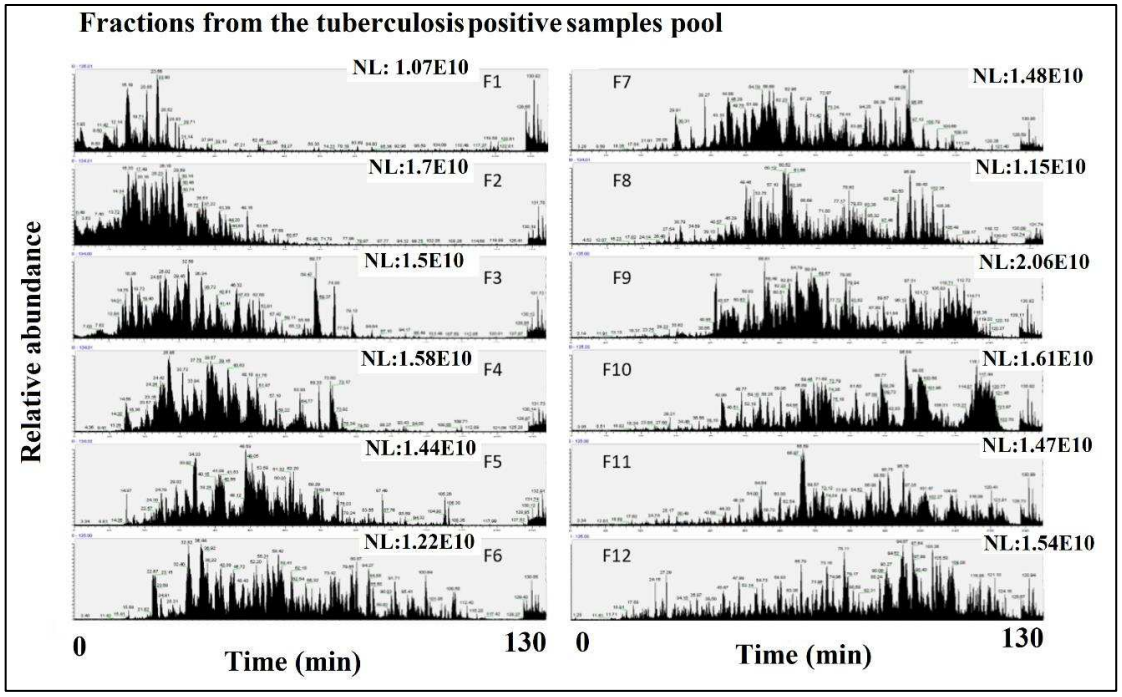

Figure 3.3 Total ion current chromatograms of the 12 fractions obtained from TB-1 and TB-2 pooled samples. The X axis shows retention time and Y axis relative abundance. Strongest peaks were above expected intensity. The relative abundance in each chromatogram depicts a scale from 0 to 100. NL: Normalized Level.

Table 3.2 Number of proteins and peptides identified in each fraction.

Sample pool	ID	F01	F02	F03	F04	F05	F06	F07	F08	F09	F10	F11	F12
TB Positive	Proteins	66	113	144	171	194	208	183	197	217	203	208	223
	Peptides	142	277	377	477	578	648	668	697	820	816	916	1231
TB Negative	Proteins	ND	122	151	217	232	290	241	239	241	177	222	216
	Peptides	ND	242	393	724	868	1109	1036	1062	1118	553	1074	1274

ND: No data

and fragment ion signal as well as the retention time (RT) were used to identify and relatively quantify, in a very specific manner, the targeted peptides in the SWATH-MS fragment map (11). Considering that RT is an important variable for peptide identification, the samples used to generate spectral libraries, or the samples processed by SWATH-MS, were spiked with a set of standard non-naturally occurring synthetic peptides (iRT standards-Biognosys AG, Switzerland). The standard peptides have a stable RT and elute across the whole range of RT of a chromatographic column. The information from the iRT standards was used to generate normalized RT for each

peptide identified (16). The combination of the SWATH-MS and the use of iRT for data extraction was defined by Bruderer *et al.,* as Hyper Reaction Monitoring (HRM) (12). A false discovery rate was set to 1% at the peptide level, only proteins present in all the samples (Q-sparse filter in Spectronaut software) and proteins identified with at least two peptides were considered for further analysis. Normalization of the peptide intensity data was performed to compensate for loading and instrument performance fluctuations. Acquisition stability was high and only minor normalization adjustments were carried out (Figure 3.4). The normalized response was used for analysis. Each sample group (Healthy, TB suspects, TB-1, and TB-2) had ten biological replicates; the healthy group showed the lowest variability in protein identification across biological replicates, however, the four groups showed a highly variable response (Figure 3.5). Overall, 369 Q-sparse filtered proteins were identified (present in all 40 samples), 19% (69/361) of these were identified by a single peptide, and consequently, were not considered for further analysis. With the resulting 278 proteins, a heat map was generated, and unsupervised clustering was used for two dimensions (protein level data set, sparse filtering). Clusters obtained by unsupervised clustering partially reconstruct the sample groups (Figure 3.6)

3.3.3 Proteins significantly different between sample categories

We pursued the identification of proteins showing significant differential abundances between sample groups by pair wise t-test. The number of significantly different proteins between healthy and the other groups increased with the detectability of *M. tuberculosis*. When comparing healthy versus TB suspects, (TB suspects were patients from TB endemic regions showing undetectable *M. tuberculosis* by either smear or culture) 118 proteins were significantly different. In the case of healthy versus TB-1 patients, which had a bacterial infection undetectable by microscopy (paucibacillary TB) but culture positive, 154 were significantly different.

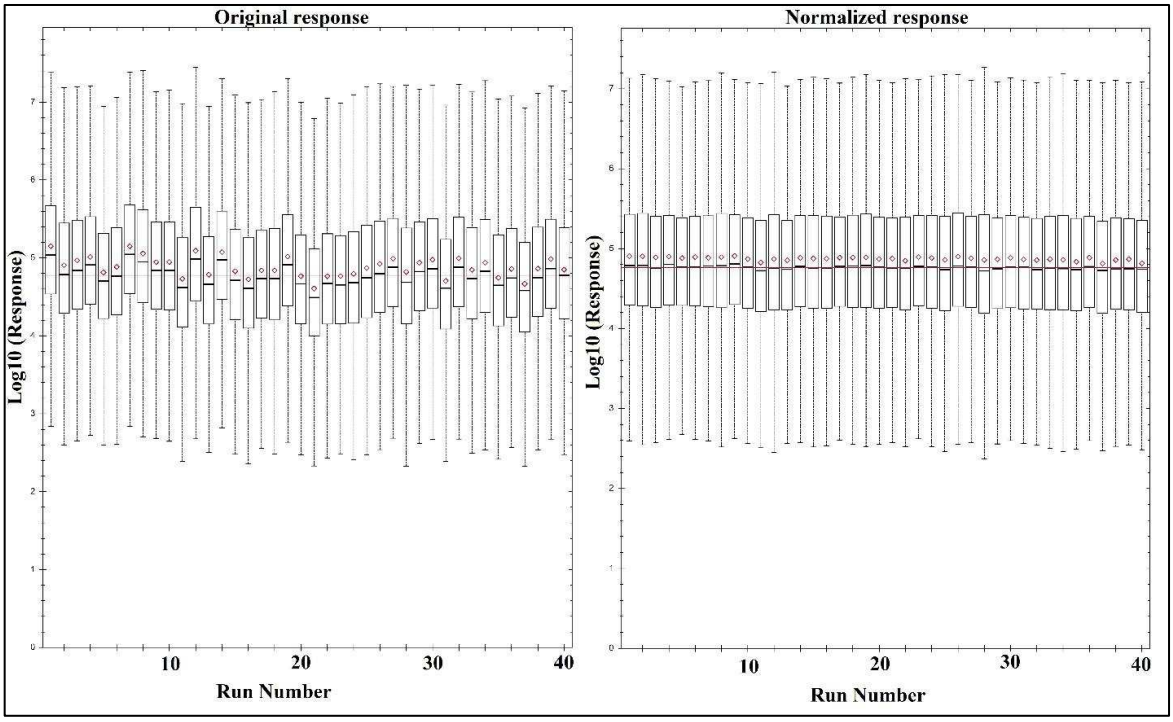

Figure 3.4 Normalization of intensities for the peptides identified. The left panel shows normal fluctuations in the intensities of identified peptides due to variability in the sample loading and/or machine performance. Right panel shows that normalization corrected such variations.

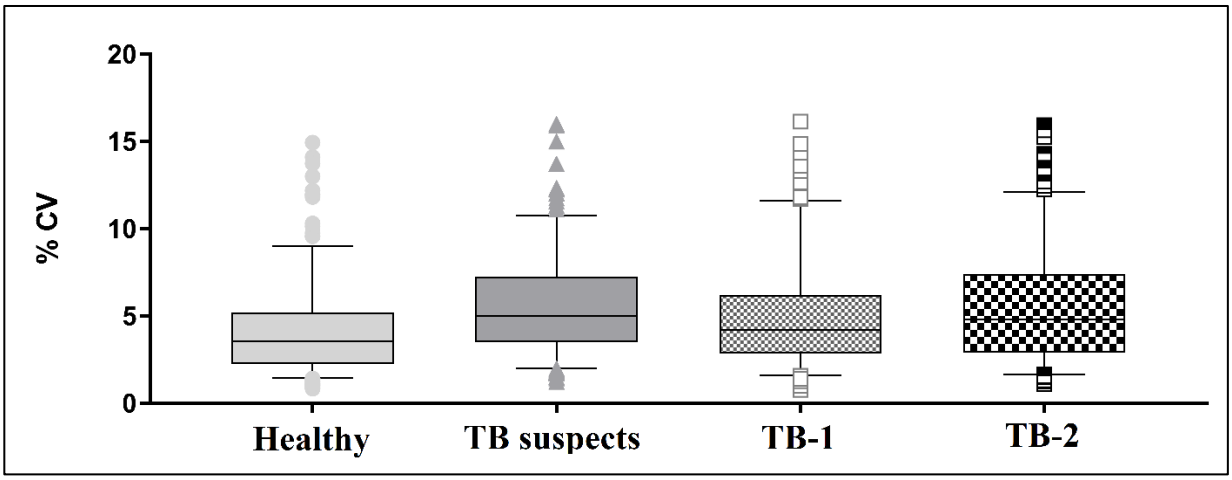

Figure 3.5 Protein coefficient of variation (CV) distribution per condition. There was no statistically significant difference between the distributions of CV for protein identification across conditions. The protein identification was variable among the ten biological replicates of each condition. Box plots show the 5-95 percentile.

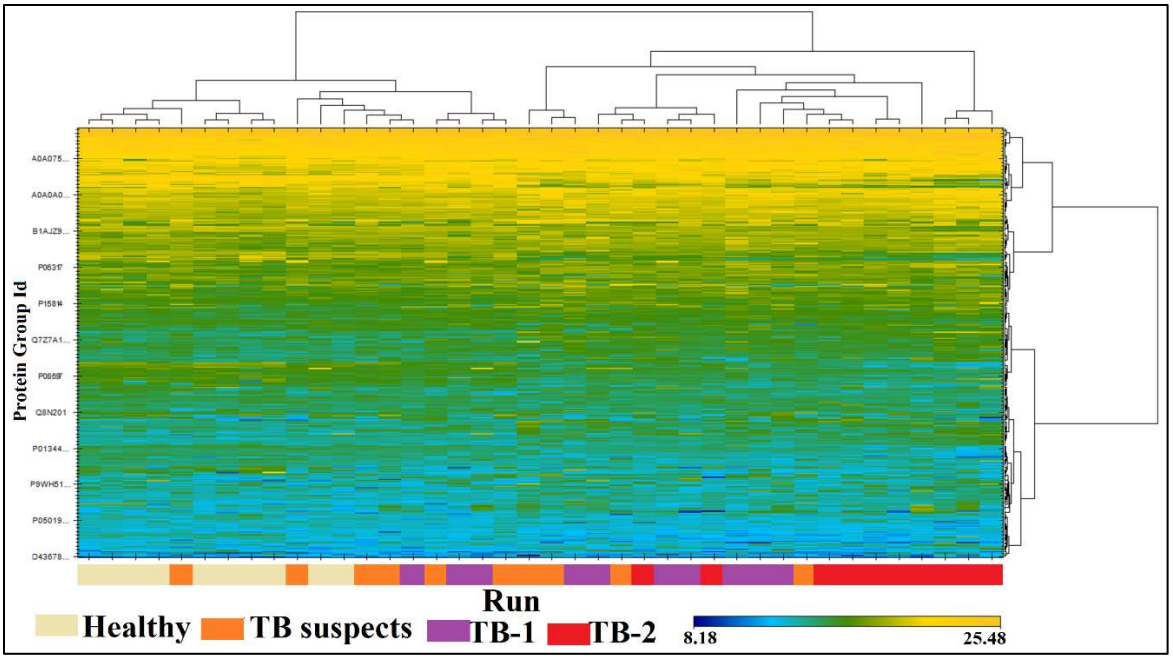

Figure 3.6 Heat map of protein intensity. The $\log_2$ of sparse-filtered protein intensity was plot for every sample run, the unsupervised clustering, partially reconstructed the sample group as depicted in the color bar.

Finally, from the comparison between healthy versus the group of patients with smear and culture positive (TB-2), there were 167 differential abundant proteins was. It is important to emphasize, that these differences are only considering human proteins, this could be inferred by the Q-sparse filter that only included proteins present in all sample groups. The comparison between TB suspects versus TB-1 and TB-2 showed a similar trend with 15 and 26 different proteins respectively. As expected, the two TB patient groups were the least variable (Table 3.3).

Table 3.3 Pair wise comparison showing the number of proteins statistically different. The higher and lower columns refer to the first group of each pair. Difference was considered statistically significant by t-test with a $p<0.05$.

Groups pairs	**Higher abundance**	**Lower abundance**
Healthy - TB suspects	45	73
Healthy - TB1	59	95
Healthy - TB2	66	101
TB suspects - TB1	6	9
TB suspects - TB2	18	12
TB1 - TB2	9	4

From all the proteins (more than 160) that showed a significantly different abundance between healthy individuals versus TB suspects, TB1 and TB2 (t-test, $p<0.05$), the five showing the greatest distance (up and down, determined by fold change) between each pair were individually analyzed. Clusterin (CLU) and Galectin-3 binding protein, were constantly lower in healthy individuals compared with TB suspects and TB patients (Figure 3.7a). Previous studies identified CLU in exosomes derived from human cells (17-20). CLU is a secreted chaperone glycoprotein activated under cellular stress, recognized for its cyto-protective function, it prevents apoptosis and is involved in inflammatory processes (18). One study showed that the levels of secreted CLU increased when blood cells of TB patients were exposed to *M. tuberculosis* antigens (21). Zhang *et al.,* demonstrated that the plasma levels of CLU increased in patients with pulmonary TB complicated by diabetes mellitus (22). These findings suggest a potential association between increased CLU plasma levels and TB. Regarding Galectin 3 binding protein, several studies showed that this is a protein commonly found in exosomes, in fact, it is heavily concentrated in ovarian carcinoma cell-derived exosomes and prostasomes (a type of exosome) (19, 20, 23). Galectin 3bp is a sialoglycoprotein with immunomodulatory functions, such as inhibition of neutrophil activation (24) and it has been suggested as an inflammatory mediator during acute dengue viral infection (25). There are no reports about a potential role of Galectin 3bp during *M. tuberculosis* infection. On the other hand, the levels of transthyretin (TTR) were constantly higher in healthy individuals compared with TB suspects and TB patients (Figure 3.7b). TTR has been identified in bovine serum exosomes (19). TTR also known as pre-albumin is normally measured to evaluate nutritional status, since its plasma levels are associated with malnutrition (26). TTR among other functions is a plasma transporter for retinol binding protein, interestingly, in our analysis, the levels of retinol binding protein correlate with TTR levels (figure 3.7).

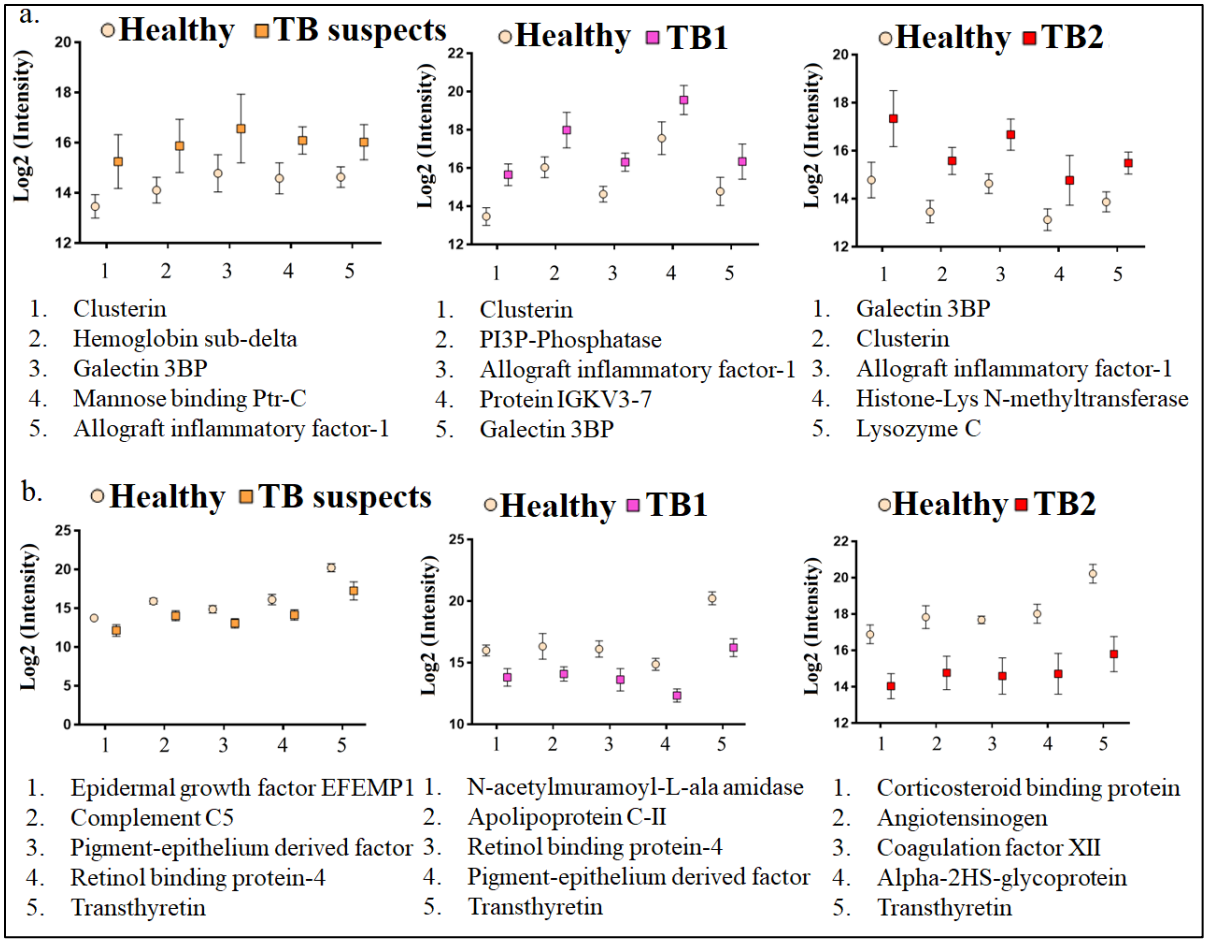

Figure 3.7 Top five proteins showing higher and lower abundances between healthy individuals and TB suspects, TB-1 and TB-2. The $\log_2$ of sparse-filtered protein intensity were used for analysis. a) Proteins significantly lower in healthy individuals. b) Proteins significantly higher in healthy individuals. Differences were considered significant for $p< 0.05$, the error bars represent mean with IC 95%. N=10, per category.

A similar pair wise analysis was developed between TB suspects and TB-1 and TB-2 patients (Figure 3.8). Surprisingly, the differences among TB suspects and TB-1 patients were considerably low, none of the proteins increased in TB-1 compared to TB suspects have been previously related to TB infection. However, kallistatin a protein involved in vascular remodeling that has been found in urine-derived exosomes (19), was significantly higher in TB suspects versus TB-1 (Figure 3.8b), additionally, this protein was significantly higher in healthy individuals compared to TB-1 ($p=0.0007$) and TB-2 patients ($p<0.0001$). De Groote *et al.,* recently published a set of serum biomarkers to detect active pulmonary TB; kallistatin was

among the top markers, compared to TB suspects patients (27). In another study from the same groups, kallistatin showed a significant increase after 8 weeks of TB treatment (28). Our results are in line with these previous reports. The evaluation between TB patient groups (TB-1 versus TB-2) showed that the adhesion molecule CD44 was statistically elevated in TB-2 patients ($p=0.03$). This molecule also was significantly higher in TB-2 compared to TB suspects (Figure 3.8a). It is possible that plasma levels of CD44 are related with the severity of mycobacterial infection. Several studies have showed that CD44 is involved in leukocytes migration to the site of infection and mediates the phagocytosis of *M. tuberculosis* (29, 30). In fact, this molecule was highly expressed in lymphocytes T accumulated in granulomas of the lungs of *M. tuberculosis* infected mice (29). Additionally, is important to note that CD44 has been identified in exosomes released from several cell types (Lymphocytes B, T and dendritic cells among others) (19, 20).

3.3.4 Proteins showing a persistent decrease (or increase) profile, from healthy individuals to TB-2 patients

To find a set of proteins that could be indicative of the dynamics of TB progression we set an analysis to evaluate proteins increasing or decreasing in a step-wise manner along each group of study. We assumed that the four groups in this study partially resemble the spectrum of TB disease. The healthy group represent the base line of protein intensity in the total absence of *M. tuberculosis* exposure. The TB suspects group were samples from patients living in TB endemic areas with high likelihood of exposure to *M. tuberculosis*. TB-1 were samples from patients with confirmed infection (culture positive) but negative smear evaluation suggesting that these patients had a very low bacterial count. Finally, TB-2 samples were from patients with smear and culture positive, which is indicative of a high bacterial burden. With this consideration, we found that three proteins (FCGR3A, lysozyme C and allograft inflammatory factor 1) showed

increasing levels in a step wise manner from the base line (healthy group) up to TB-2 (Figure 3.9a). FCGR3A or low affinity immunoglobulin gamma Fc region receptor III-A (also known as CD16), showed a significant increase from healthy to TB suspects ($p=0.02$), TB-1 ($p=0.01$) and TB-2 ($p<0.0001$). One study showed that TB patients had high levels of circulating CD16+ monocytes (31). These cells seemed to be important for the control of *M. tuberculosis* infection since they produced high levels of interferon-gamma and were highly susceptible to apoptosis (31). If TB patients have high levels of circulating CD16+ cells, this could be associated with a higher number of exosomes containing CD16+. Lysozyme C, not only was significantly different between healthy versus TB-1 and TB-2 ($p=0.0016$ and $p<0.0001$, respectively) but also showed a significant difference among TB suspects and TB-2 ($p=0.0019$). Lysozyme C is a bacteriolytic enzyme, indicator of macrophage activation, and it has been found highly expressed in granulomas of mice infected with the Bacillus Calmette-Guerin (32). Finally, allograft inflammatory protein factor 1, was significantly higher in TB suspects ($p=0.0022$), TB-1 ($p=0.0002$) and TB-2 ($p<0.0001$) compared to the healthy group. This protein is involved in macrophage activation and inflammatory responses (33).

The opposite tendency (step wise decreasing values from healthy to TB-2) was observed in six proteins: Alpha2-HS-glycoprotein (AHSG, or fetuin-A), angiotensinogen, coagulation factor XII, N-acetylmuramoyl-L-alanine amidase (PGRP-L), corticosteroid-binding globulin (CBG), and Carboxypeptidase N subunit 2 (CPN2) (Figure 3.9b). Fetuin-A is highly abundant plasma glycoprotein, which mainly functions as scavenger/carrier protein. Fetuin-A has anti-inflammatory properties, inhibits the tumor necrosis factor alpha, and is considered a negative acute phase protein showing an inverse relationship with C-reactive protein (34).

Our results showed that, fetuin-A was significantly lower in TB suspects ($p=0.012$),

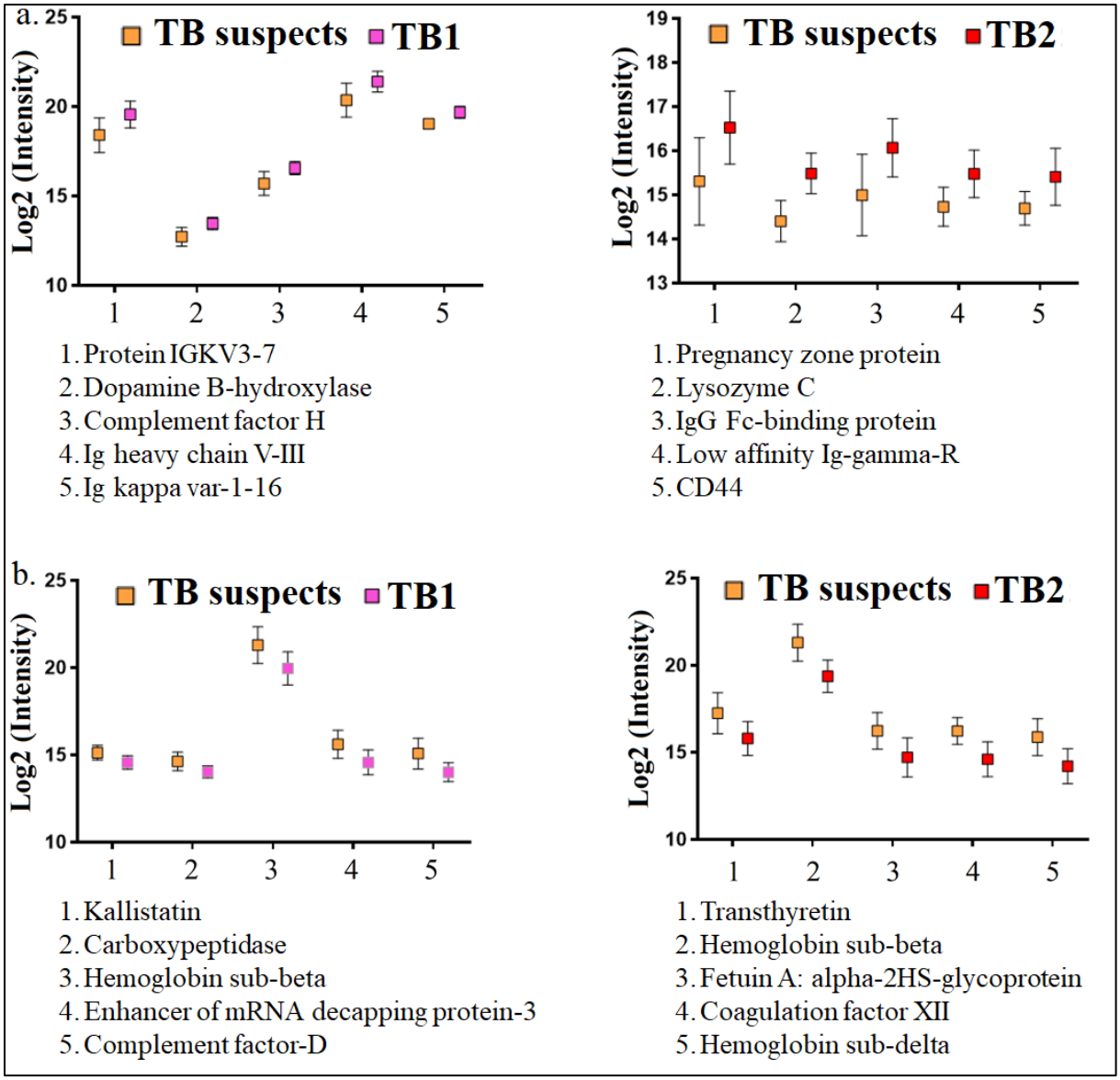

Figure 3.8 Top five proteins showing higher and lower abundances between TB suspects versus TB-1 and TB-2 patients. The $\log_2$ of sparse-filtered protein intensity were used for analysis. a) Proteins significantly lower and b) Proteins significantly higher in TB suspects patients. Differences were considered significant for $p < 0.05$, the error bars represent mean with IC 95%. N=10, per category.

TB-1 (0.002), and TB-2 ($p<0.0001$) compared to the healthy group. The same trend was observed between TB suspects and TB-2 patients ($p=0.04$) (Figure 3.9b). Our findings agree with the results obtained by Tanaka *et al.,* in 2011, they found that blood levels of fetuin-A were significantly lower in TB patients versus uninfected controls by ELISA (21). PGRP-L is a

pattern recognition molecule, which acts as peptidoglycan hydrolase with no bacteriolytic function. There are no reports about the plasma concentration of this protein but it is known that is highly expressed in liver cells. PGRP-L is mainly localized in the cell membrane or in intracellular vesicles (35). It is possible that PGRP-L is secreted in exosomes or another type of intracellular vesicles reaching systemic circulation. PGRP-L was significantly lower in TB suspects ($p=0.012$), TB-1 ($p<0.0001$) and TB-2 patients ($p<0.0001$), compared to healthy controls. The protein was also significantly lower in TB-2 compared to TB suspects patients ($p=0.0418$). Finally, coagulation factor XII was significantly lower in TB suspects ($p=0.0097$), TB-1 ($p=0.0002$) and TB-2 ($p<0.0001$) to healthy group. It was suggested by Levi *et al.,* that potential low levels of coagulation factor XII could be resembling a "negative acute phase" effect, during meningococcal septicemia (36). There is no information regarding TB disease and plasma levels of coagulation factor XII.

3.4 Conclusions

To date, this study is the first evaluation of exosome-enriched samples from the serum of TB patients searching for proteins potentially associated with the state of TB disease. We did a comprehensive proteomic profile of the samples using a state of the art methodology known as Hyper Reaction Monitoring (HRM), that combines the DIA based SWATH-MS method with the use of synthetic peptides to generate retention time normalized spectral libraries. Our sample set included three different groups across the TB disease spectrum. First, samples from a group of TB suspects patients from TB endemic areas, which suggests a high probability that these patients have been in contact with TB patients. Second (TB-1), samples from a group of patients with active TB, undetectable by microscopy which implies a very low bacterial concentration in the sputum (less than 10,000 organisms/ml sputum are practically undetectable by microscopic

examination (37). Finally, (TB-2) samples from patients with higher bacterial load: smear and culture positive for TB. Additionally, we included a control group of healthy individuals from non TB endemic regions. Nine proteins showed a very distinct dynamic profile along the study groups, FCGR3A, lysozyme C and allograft inflammatory factor 1, showed a step wise increase, while fetuin-A, angiotensinogen, coagulation factor XII, PGRP-L, CBG, and CPN2, showed a step wise decrease. From the healthy group to the TB-2 group. Several proteins that showed a profile associated with TB disease or health status (CLU, TTR, kallistatin, CD16, and fetuin-A) in the present study, have been found in previous studies, exhibiting similar profiles among TB and control patients.

Although proteins showing statistically different mean values among disease and control group could be considered potential biomarker candidates, it is important to recognize that many of these significantly different proteins could fail to discriminate the two conditions evaluated due to overlapping distributions. The pathway to discover a true biomarker requires the evaluation of its sensitivity and specificity based on statistically derived cut-off values (38).

This study demonstrated that a human set of proteins potentially concentrated in exosomes derived from sera samples, followed a dynamic tendency that is linked with TB disease. Our findings, suggests that exosome from human serum could be a source for TB biomarkers; further validation of these results is forthcoming.

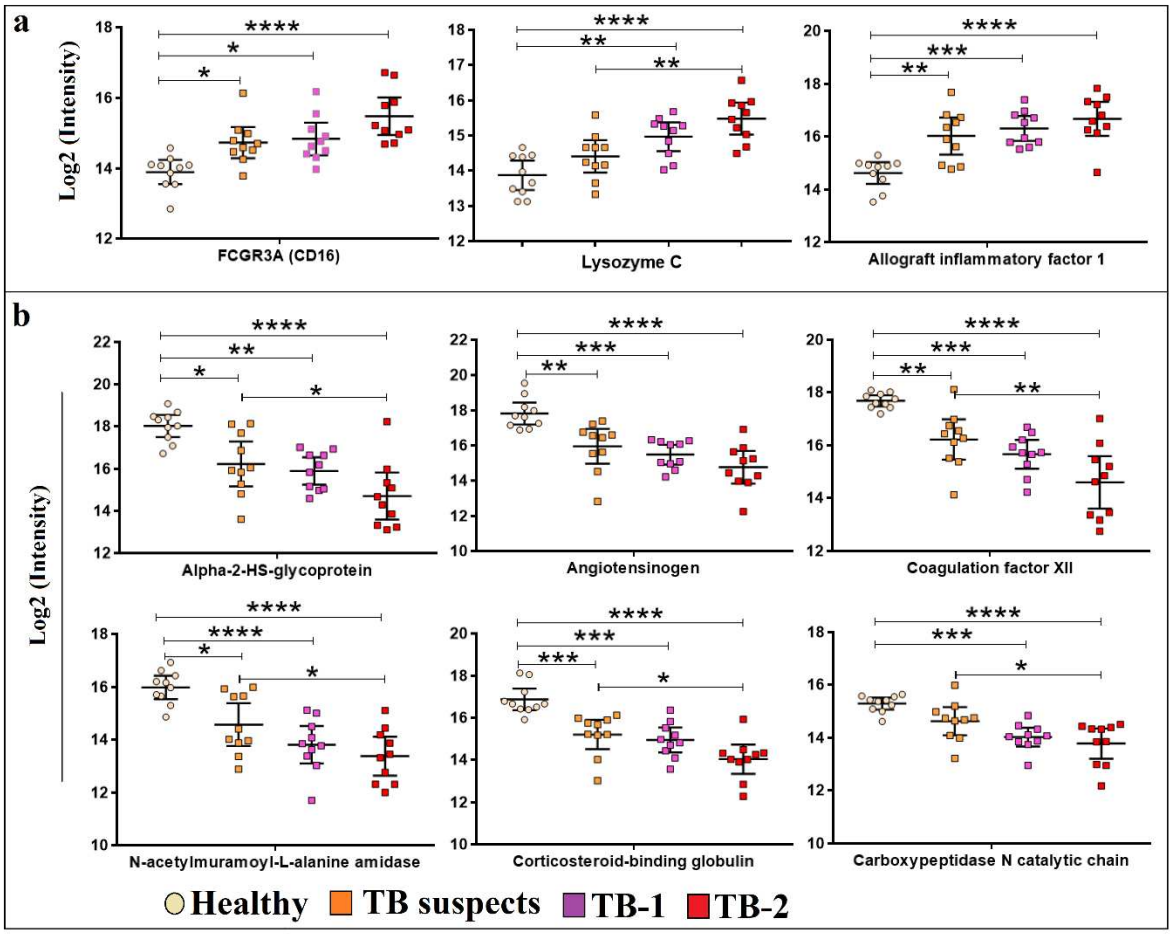

Figure 3.9 Proteins showing a step wise decrease or increase along the study groups**.** The log_2 of sparse-filtered protein intensity were used for analysis. a) Proteins showing increased b) decreased, levels in step wise manner. Differences were calculated by one-way ANOVA and the Tukey's multiple comparisons test. Significance was accepted for $p< 0.05$, the error bars represent mean with IC 95%. N=10, per category. * $0.01<p<0.05$; ** $0.001<p<0.01$; ***$0.0001<p<0.001$ and **** $p<0.0001$.

References

1. Schnappinger D, Ehrt S. A broader spectrum of tuberculosis. Nat Med. 2016;22(10):1076-7.
2. Esmail H, Lai RP, Lesosky M, Wilkinson KA, Graham CM, Coussens AK, et al. Characterization of progressive HIV-associated tuberculosis using 2-deoxy-2-[(18)F]fluoro-D-glucose positron emission and computed tomography. Nat Med. 2016;22(10):1090-3.
3. Malherbe ST, Shenai S, Ronacher K, Loxton AG, Dolganov G, Kriel M, et al. Persisting positron emission tomography lesion activity and *Mycobacterium tuberculosis* mRNA after tuberculosis cure. Nat Med. 2016;22(10):1094-100.
4. 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. 2016;7(36):57581-92.
5. Mehaffy C, Dobos KM, Nahid P, Kruh-Garcia NA. Second generation multiple reaction monitoring assays for enhanced detection of ultra-low abundance *Mycobacterium tuberculosis* peptides in human serum. Clin Proteomics. 2017;14:21.
6. Kruh-Garcia NA, Wolfe LM, Chaisson LH, Worodria WO, Nahid P, Schorey JS, et al. Detection of *Mycobacterium tuberculosis* peptides in the exosomes of patients with active and latent *M. tuberculosis* infection using MRM-MS. PLoS One. 2014;9(7):e103811.
7. Sohel MM, Hoelker M, Noferesti SS, Salilew-Wondim D, Tholen E, Looft C, et al. Exosomal and Non-Exosomal Transport of Extra-Cellular microRNAs in Follicular Fluid: Implications for Bovine Oocyte Developmental Competence. PLoS One. 2013;8(11):e78505.
8. Helwa I, Cai J, Drewry MD, Zimmerman A, Dinkins MB, Khaled ML, et al. A Comparative Study of Serum Exosome Isolation Using Differential Ultracentrifugation and Three Commercial Reagents. PLoS One. 2017;12(1):e0170628.
9. Diaz G, Wolfe LM, Kruh-Garcia NA, Dobos KM. Changes in the Membrane-Associated Proteins of Exosomes Released from Human Macrophages after *Mycobacterium tuberculosis* Infection. Sci Rep. 2016;6:37975.
10. Domon B, Aebersold R. Options and considerations when selecting a quantitative proteomics strategy. Nat Biotechnol. 2010;28(7):710-21.
11. Gillet LC, Navarro P, Tate S, Röst H, Selevsek N, Reiter L, et al. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Mol Cell Proteomics. 2012;11(6):O111.016717.
12. Bruderer R, Bernhardt OM, Gandhi T, Miladinović SM, Cheng LY, Messner S, et al. Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. Mol Cell Proteomics. 2015;14(5):1400-10.
13. Scheltema RA, Hauschild JP, Lange O, Hornburg D, Denisov E, Damoc E, et al. The Q Exactive HF, a Benchtop mass spectrometer with a pre-filter, high-performance quadrupole and an ultra-high-field Orbitrap analyzer. Mol Cell Proteomics. 2014;13(12):3698-708.
14. Schubert OT, Ludwig C, Kogadeeva M, Zimmermann M, Rosenberger G, Gengenbacher M, et al. Absolute Proteome Composition and Dynamics during Dormancy and Resuscitation of *Mycobacterium tuberculosis*. Cell Host Microbe. 2015;18(1):96-108.

15. Callister SJ, Barry RC, Adkins JN, Johnson ET, Qian WJ, Webb-Robertson BJ, et al. Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics. J Proteome Res. 2006;5(2):277-86.
16. Escher C, Reiter L, MacLean B, Ossola R, Herzog F, Chilton J, et al. Using iRT, a normalized retention time for more targeted measurement of peptides. Proteomics. 2012;12(8):1111-21.
17. Hosseini-Beheshti E, Pham S, Adomat H, Li N, Tomlinson Guns ES. Exosomes as biomarker enriched microvesicles: characterization of exosomal proteins derived from a panel of prostate cell lines with distinct AR phenotypes. Mol Cell Proteomics. 2012;11(10):863-85.
18. Foglio E, Puddighinu G, Fasanaro P, D'Arcangelo D, Perrone GA, Mocini D, et al. Exosomal clusterin, identified in the pericardial fluid, improves myocardial performance following MI through epicardial activation, enhanced arteriogenesis and reduced apoptosis. Int J Cardiol. 2015;197:333-47.
19. Mathivanan S, Fahner CJ, Reid GE, Simpson RJ. ExoCarta 2012: database of exosomal proteins, RNA and lipids. Nucleic Acids Res. 2012;40(Database issue):D1241-4.
20. Simpson RJ, Kalra H, Mathivanan S. ExoCarta as a resource for exosomal research. J Extracell Vesicles. 2012;1.
21. Tanaka T, Sakurada S, Kano K, Takahashi E, Yasuda K, Hirano H, et al. Identification of tuberculosis-associated proteins in whole blood supernatant. BMC Infect Dis. 2011;11:71.
22. Zhang X, Ma A, Sun S, Sun Y. Proteomic Analysis of Plasma in Adult Active Pulmonary Tuberculosis Patients with Diabetes Mellitus. Clin Lab. 2015;61(10):1481-90.
23. Block AS, Saraswati S, Lichti CF, Mahadevan M, Diekman AB. Co-purification of Mac- 2 binding protein with galectin-3 and association with prostasomes in human semen. Prostate. 2011;71(7):711-21.
24. Läubli H, Alisson-Silva F, Stanczak MA, Siddiqui SS, Deng L, Verhagen A, et al. Lectin galactoside-binding soluble 3 binding protein (LGALS3BP) is a tumor-associated immunomodulatory ligand for CD33-related Siglecs. J Biol Chem. 2014;289(48):33481-91.
25. Liu KT, Liu YH, Chen YH, Lin CY, Huang CH, Yen MC, et al. Serum Galectin-9 and Galectin-3-Binding Protein in Acute Dengue Virus Infection. Int J Mol Sci. 2016;17(6).
26. Dellière S, Cynober L. Is transthyretin a good marker of nutritional status? Clin Nutr. 2017;36(2):364-70.
27. De Groote MA, Sterling DG, Hraha T, Russell T, Green LS, Wall K, et al. Discovery and Validation of a Six-Marker Serum Protein Signature for the Diagnosis of Active Pulmonary Tuberculosis. J Clin Microbiol. 2017.
28. De Groote MA, Nahid P, Jarlsberg L, Johnson JL, Weiner M, Muzanyi G, et al. Elucidating novel serum biomarkers associated with pulmonary tuberculosis treatment. PLoS One. 2013;8(4):e61002.
29. Leemans JC, Florquin S, Heikens M, Pals ST, van der Neut R, Van Der Poll T. CD44 is a macrophage binding site for *Mycobacterium tuberculosis* that mediates macrophage recruitment and protective immunity against tuberculosis. J Clin Invest. 2003;111(5):681-9.
30. DeGrendele HC, Estess P, Siegelman MH. Requirement for CD44 in activated T cell extravasation into an inflammatory site. Science. 1997;278(5338):672-5.
31. Castaño D, García LF, Rojas M. Increased frequency and cell death of CD16+ monocytes with *Mycobacterium tuberculosis* infection. Tuberculosis (Edinb). 2011;91(5):348-60.
32. Keshav S, Chung P, Milon G, Gordon S. Lysozyme is an inducible marker of

macrophage activation in murine tissues as demonstrated by in situ hybridization. J Exp Med. 1991;174(5):1049-58.
33. Deininger MH, Meyermann R, Schluesener HJ. The allograft inflammatory factor-1 family of proteins. FEBS Lett. 2002;514(2-3):115-21.
34. Jahnen-Dechent W, Heiss A, Schäfer C, Ketteler M. Fetuin-A regulation of calcified matrix metabolism. Circ Res. 2011;108(12):1494-509.
35. Liu C, Xu Z, Gupta D, Dziarski R. Peptidoglycan recognition proteins: a novel family of four human innate immunity pattern recognition molecules. J Biol Chem. 2001;276(37):34686-94.
36. Levi M, Keller TT, van Gorp E, ten Cate H. Infection and inflammation and the coagulation system. Cardiovasc Res. 2003;60(1):26-39.
37. Desikan P. Sputum smear microscopy in tuberculosis: is it still relevant? Indian J Med Res. 2013;137(3):442-4.
38. LaBaer J. So, you want to look for biomarkers (introduction to the special biomarkers issue). J Proteome Res. 2005;4(4):1053-9.

Chapter 4: Identification and evaluation of an exosome-derived proteomic biosignature associated with active tuberculosis disease

4.1 Introduction

The diagnosis of Tuberculosis (TB) remains a great obstacle to control this disease worldwide. The need for a novel, accurate, easy-to-do, and inexpensive test is imperative especially in low and middle-income countries where TB burden accounts for more than 95% of the total number of cases and deaths globally. At the same time, the main tool for routine TB diagnosis in these countries is the microscopic examination of sputum (1). The microscopic detection of *Mycobacterium tuberculosis* in sputum has several limitations including its low and variable sensitivity (from 20% to 60%) (2, 3). Recently, a comprehensive review from the World Health Organization (WHO) demonstrated that an alternative sputum-based test for TB diagnosis, TB-LAMP, performs better than microscopy (sensitivity 78% and 63%, respectively) at identifying pulmonary TB suspects. TB-LAMP detects bacterial DNA from sputum samples using a loop mediated isothermal amplification reaction. WHO has recommended the use of TB-LAMP as a replacement of microscopy for the diagnosis of pulmonary TB (4). While an improvement in detection of TB cases is expected, the inherent limitations of using sputum will remain. Consequently, the diagnosis of children who do not produce sputum, and patients with miliary TB or HIV-coinfection, which normally have low/no bacterial load in their sputum, is going to continue to be a great challenge to overcome (5) without a TB diagnostic test independent of a sputum sample.

The study of systemic changes in the human proteome associated to TB status represents a promising opportunity for the discovery of new biomarkers and the design of alternative

diagnostic tools that can be based on serum or plasma samples. In the last decade, several studies have reported proteomic changes in human serum proteins associated with TB (6-11). Unfortunately, none have been translated to a diagnostic test (12, 13). Thus, the bridge that allows these biomarkers to go from "bench to bedside" has not been completed. Several limitations, such as variability among study designs, differences among technologies, geographical localization of patients and controls, and improper statistical analyses have affected the results, leading to a lack of consensus of TB biomarkers among different studies (7, 13, 14). The evolution of MS technologies allowed the identification of thousands of proteins from a single analysis, however, most studies—specifically during discovery phase—are designed with too few samples to accurately capture the number of potential predictors (15). Usually, the selection of candidate proteins is achieved by analyzing differences among group means and fold change (12). This strategy for candidate selection is well accepted, although all potential candidates must be validated (Figure 4.1).

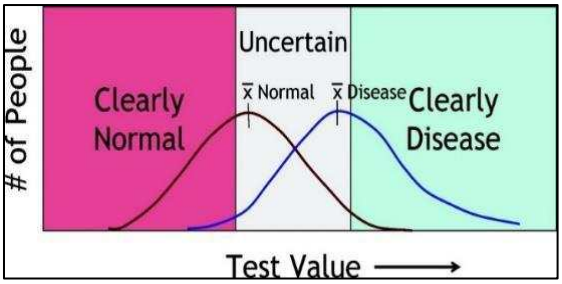

Figure 4.1 Differences of means and biomarkers**.** In the discovery phase of one study, even though the mean of a protein is significantly different among disease and control groups, the distribution of the marker among the population can be very wide, then the range of uncertainty will make very difficult to use that marker to distinguish a test patient. "Reprinted with permission from Joshua LaBaer; J. Proteome Res. 2005, 4, 1053-1059. Copyright 2005 American Chemical Society."

The number of biomarker proteins to predict a specific condition (i.e. TB case) must be trimmed from hundreds of candidates to a finite number prior to translation to a relevant clinical assay. Mathematical models are crucial for the selection of a subset of predictors with higher

discriminatory power. One strategy is the application of subset selection methods, but it has been proven that these methods perform poorly in large data sets when significant multicollinearity is present (16, 17). In 1996, Dr. Tibshirani developed a regression model that shrinks some of the variables while others are set to zero, this method is known as the least absolute shrinkage and selection operator (Lasso). Lasso generates a subset selection producing simpler models that are relatively easy to interpret (17).

Recent studies showed that Lasso can generate an inconsistent selection of predictors from the same data sets, thus, generating inconsistent predictive models. In 2006, Zou developed a variation of Lasso, the adaptive Lasso, which generates more stable models (18). We hypothesize that Lasso and adaptive Lasso can aid in the selection of potential TB biomarker candidates.

In Chapter 3 of this dissertation, we found a set of proteins with significantly different expression between four groups: healthy, TB-suspects, TB-1, and TB-2. These results strengthened the idea that exosome-enriched fractions from human serum are a source of biomarkers to discriminate TB patients from healthy individuals and/or from TB-suspects living in TB endemic regions. In this chapter, we identified two predictive models, lasso and adaptive lasso which were able to discriminate TB patients from healthy individuals and TB-suspects. To select the proteins. The regression models were then applied to an unique set of TB suspect samples in which *Mtb* proteins had been previously identified.

4.2 Materials and methods

4.2.1 Development of predictive regression models using lasso and adaptive lasso, to discriminate TB from TB suspects, TB from Healthy, TB-1 versus TB-2

The 276 proteins identified from the HRM-MS analysis of the 40 samples of chapter 3 were utilized to develop the Lasso regression models. The strategy is summarized in figure 4.2.

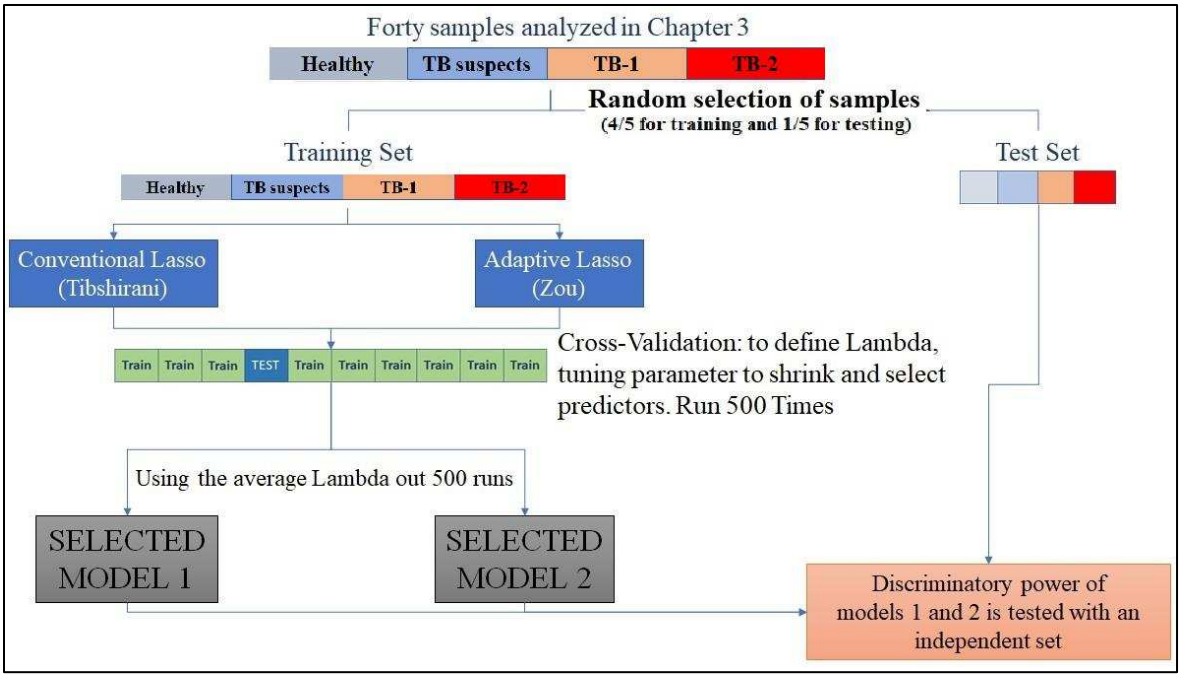

Figure 4.2 General description of model regression development. The sample groups were split into training (80% of each group N=8 per group) and test (20% of each group N=2 per group) using arbitrary randomization.

Initially, the forty samples were randomly assigned to a training (80%) and test (20%) sets. The training set was used to develop both conventional and adaptive Lasso. All procedures were done following the algorithms developed by Tibshirani in 1996 (17). In both cases, the selection of predictors depends on a tunable factor (Lambda). Briefly, if all the available predictors were used to solve the model, the sum of the absolute value of the coefficients would be $\mathbf{t_0}$, if that sum were restricted to a **t** value lower than $\mathbf{t_0}$, some of the predictors would be shrunk towards zero and some of them will be zero. For larger values of **t**, many predictors will be included in the model (overfit model), and the predictive error will be high. Conversely, with very small values of **t**, few predictors will be included in the model (under fit model) and again the predictive error will be high. There is a value for **t** that minimizes the predictive error. That value can be calculated by a cross-validation analysis. For the cross-validation process, the training set was divided into 10 parts; 9 parts were used to set a model determined by a lambda value that allowed a maximum

number of predictors (overfit model) and the remaining 1 part ($1/10^{th}$ of the complete set) was used to test the prediction error of such model (Figure 4.2). The process was repeated changing the 1 "test" part of the complete sample set every time, until all iterations of sample sets for predicting and testing had been applied. The result is a distribution of prediction errors for the specific value of lambda. Then, the value of Lambda was consecutively changed, to produce each time models with fewer predictors, and the whole validation process was repeated. Finally, the lambda that minimizes the prediction error (best lambda) was selected and the corresponding coefficients determined. Since we anticipated that different randomization of the training data will produce a variation of the selected predictors, the whole process was run 500 times, selecting each time the best lambda; the average of "best lambda" was used to define the final predictive model. The main difference in the adaptive Lasso is that, before the cross-validation process, different weights are assigned to each coefficient, following the algorithm proposed by Zou (18), then, the resulting weighted model is processed as the conventional Lasso. Finally, both models 1 and 2 were evaluated with the test set to determine their discriminatory power (figure 4.2).

4.2.2 Description of TB suspects and serum processing to obtain an exosome-enriched fraction

Ten serum samples from TB-suspects, as defined in chapter 3, were obtained from the FIND specimen repository and exosomes isolated as described in chapter 3. Mycobacterial peptides were identified in all 10 of these samples by Multiple Reaction Monitoring Mass Spectrometry (MRM-MS) as described previously, with modifications (manuscript in preparation). Exosomes were also resuspended in a lysis buffer provided by Biognosys and shipped frozen to Biognosys AG (Schlieren, Switzerland) for HRM-MS analysis.

4.2.3 Hyper reaction monitoring/SWATH-MS of TB suspect samples

The TB-suspect sample set was processed following the same procedure described in chapter 3. One μg of peptides per sample was injected into a C18 column. Similarly, the same LC and MS systems were used. A DIA method with one full range survey scan and 14 DIA isolation windows were used. The false discovery rate on peptide level was set to 1%, data was filtered using row based extraction. The total peptide inventory obtained from the spectral library generated in chapter 3, was searched in the DIA fragmentation maps of this sample set.

4.2.4 Data analysis

The HRM measurements were analyzed with the software Spectronaut 11 (Biognosys) peptide intensities were normalized using local regression normalization (21). The regression models Lasso and adaptive Lasso and the Receiving Operating Characteristic (ROC) curves to test the predictive models, were developed using R version 3.2.4 (2016, The R foundation for Statistical Computing), with the package "glmnet" ver. 2.0-5 developed by Friedman J. *et al* (22).

4.3 Results and discussion

4.3.1 Regression model development

The development of a regression model that accurately predicts the condition of an unknown sample requires a minimum of three important groups of data: training data, test data, and a confirmation/validation data set. In this study, a pre-classification of the samples was achieved by gold standard methodologies (smear microscopy and culture); this information was used to define which set of proteins (predictors) would be able to discriminate one condition from the other. Previously (in chapter 3) we identified several proteins that were significantly different by pair wise analysis of variance, among the four evaluated groups. Using the 276 proteins identified by HRM-MS with FDR 1% and minimum 2 peptides per protein, we calculated

regression models to identify protein signatures with the strongest discriminatory power to classify TB cases.

4.3.2 Regression model to individually differentiate the groups: healthy, TB suspects, TB-1 and TB-2

We initially attempted to find a model able to discriminate each group of study (healthy, TB suspects, TB-1 and TB-2). Unfortunately, due to the small number of replicates per group (n=10), the cross-validation process was unable to find a fitting model (Figure 4.3). To run this analysis, the data was split in 80% for training and 20% for testing. Consequently, the number of replicates for the training set in each category was eight. For the cross-validation process, each group of eight was randomly reassigned for model development or testing while error model was calculated. There was not any value of lambda that produced a model with a minimized error. Similarly, it was not possible to obtain a model using pair-wise data, for instance, healthy versus TB suspects or TB suspects versus (TB-1+TB-2 as a single group). Overall, every time we used one of the groups separately, the low number of replicates impaired the capacity for cross-validation analysis to produce an ideal lambda (Figure 4.3).

4.3.3 Model regression to discriminate TB from not TB

Due to the limited number of samples per condition, we decided to regroup them into two major categories: TB-neg (healthy and TB suspects) and TB-pos (TB-1 and TB-2) to increase the strength of statistical analysis. We randomly split the samples into training and test sets. First, we ran a cross-validation process to find the lambda that produced the model with minimal predictive error (best lambda). Since it has been proven that Lasso can generate inconsistent selection of predictors we decided to run the cross-validation process 500 times to estimate the range of variability of "best lambda" values (Figure 4.4a). It is important to emphasize that for

each lambda a slightly different model is expected. The figure 4.4a shows that the variation of model selection based on the number of predictors ranges from models with 1 predictor (under fitted) to models with up to 13 predictors. In the case of adaptive Lasso (Figure 4.4b), most selected models are in the range of 8 to 15 predictors. A higher number of selected overlapping predictors is observed in this adaptive model, suggesting a higher stability in model selection by using the adaptive Lasso, which is in accordance with previously published work (18).

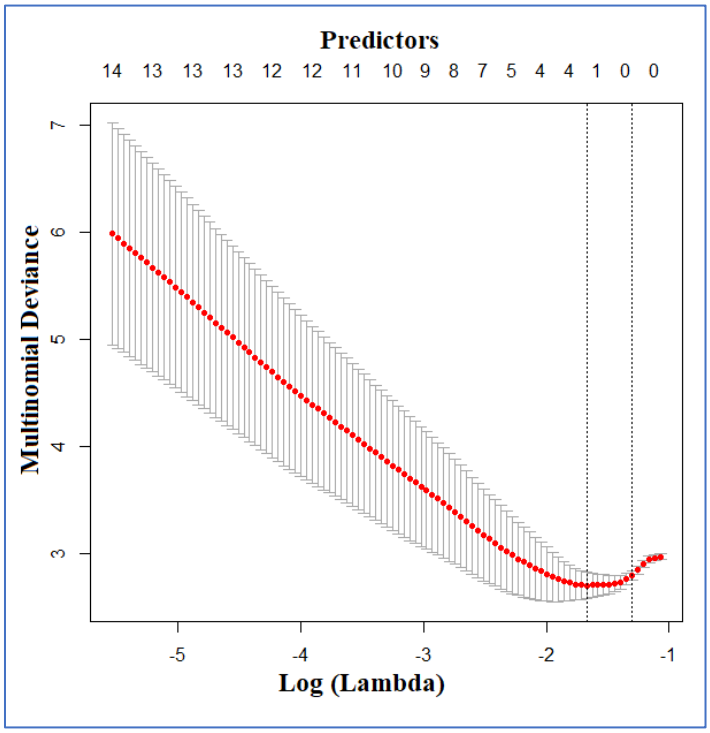

Figure 4.3 Cross-validation process to find the best lambda to generate a model to discriminate among healthy, TB suspects, TB-1 and TB-2 samples. The y axis shows the multinomial deviance as a measure of the cross-validation error. The upper side of the figure shows the polynomial grade of each model tested. The x axis shows the value for lambda (Lasso shrinkage parameter) lower values of lambda generate a model with higher polynomial grade.

From each analysis, the mean of "best lambda" value was picked to generate the corresponding predictive model. The coefficients of each model are listed in table 4.1. Conventional Lasso generated a model with 11 predictors while the Adaptive Lasso generated a model with nine predictors.

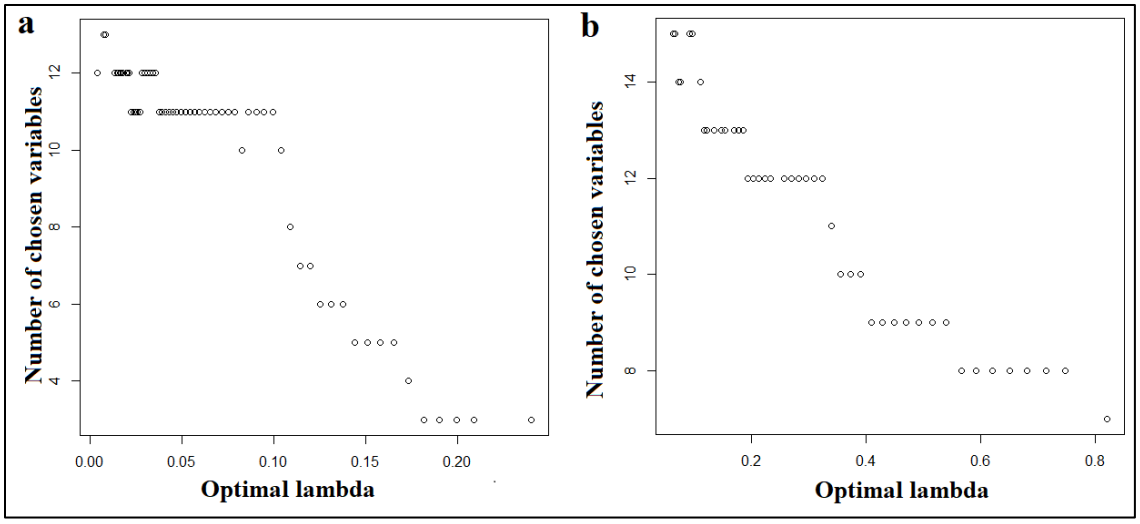

Figure 4.4 Cross-validation process distribution of "best lambda" values. The y axis shows the number of predictors for each value of lambda. **a.** Conventional Lasso and **b.** Adaptive Lasso.

Five proteins were common to both (Table 4.1). Eleven out 15 proteins included in either model have been found in exosomes in previous studies (Table 4.1). Both models selected a number of proteins associated with the humoral immune response. Interestingly, recent evidence has pointed out that individuals with active TB can be differentiated from LTBI cases by the type of circulating antibodies (23). Differences in the glycosylation pattern and the binding selectivity of immunoglobulins in LTBI individuals account for the more relevant differences compared to the antibodies from active TB individuals (23). Our results suggest a potential role of immunoglobulin fractions to identify active TB. A deeper analysis of the protein KV116 (P04430, the variable domain of immunoglobulins that participates in antigen recognition), may reveal information associated with TB infection. The adaptive Lasso model included the alpha1-acid-glycoprotein1 (A1AG1), which is a strong acute phase protein associated with anti-TB treatment response (24). In experimental TB models, it was shown that A1AG1 may be involved in disease progression by having a negative modulatory effect on cellular immunity (25). Additionally, adaptive Lasso selected the sex hormone-binding globulin (SHBG). This protein was found to be differentially expressed in TB patients in a previous serum-based biomarker

study using MALDI-TOF-MS. An independent evaluation of SHBG in serum by ELISA showed promising results classifying positive TB patients (26).

Table 4.1 Coefficients of the two predictive models generated by Lasso and adaptive Lasso methodologies.

Predictors (protein name)	**Conventional Lasso**	**Adaptive Lasso**	**Previously found in exosomes**
Complement factor H (CFAH)		-2.7474625	Yes (27)
Alpha-1-acid glycoprotein 1 (A1AG1)		0.7089016	Yes (27, 28)
Phosphotransferase (C9JQD1) Glucokinase GCK		0.4066641	No
Sex hormone-binding globulin (I3L145)		-0.7959483	Yes (29)
Complement component C8 gamma chain (CO8G)	-0.042171855	-0.8973755	Yes (27, 28)
Immunoglobulin kappa variable 3-20 (P04206)	0.015242731	0.1546919	Yes (27, 28, 30)
Carboxypeptidase N catalytic chain (CBPN)	-0.507612713	-0.3402092	No
Immunoglobulin kappa variable 1-16 (KV116)	1.295639168	3.9621907	Yes (28)
Complement factor H-related protein 1 (FHR1)	0.052614187	0.1113752	Yes (28)
Coagulation factor XII (FA12)	-0.491144475		Yes (27, 28)
Immunoglobulin lambda variable 3-25 (IGLV3-25)	0.026861277		Yes (28)
Complement factor D (K7ERG9)	-0.175230434		No
Plasma protease C1 inhibitor (SERPING1, IC1)	-0.058153259		Yes (27, 28)
Immunoglobulin kappa variable 3-7 (IGKV3-7)	0.11675631		No
Immunoglobulin lambda variable 2-11 (LV211)	0.002149055		Yes (28)
Intercept	-9.98	-16.53	

The predictive power of each model was evaluated using the test group of samples. The test set was defined randomly and included samples from each category: healthy, TB suspects, TB-1 and 2. The output of each prediction represents the probability of a sample to be TB-pos. Both models partially discriminated the samples according to their actual classification (Table 4.2 and Figure 4.5). The sample TB suspects_1 was misclassified by both models generating a wide distribution of response in the TB-neg group (Figure 4.5). The sample TB suspects_1 may be an undiagnosed patient. Due to the low number of samples evaluated it is not possible to derive further conclusions. Additionally, the conventional Lasso model classified the sample TB-1 in the borderline probability between TB-pos and TB-neg. Overall, adaptive Lasso showed stronger power to discriminate the test-set into the two categories: TB-neg and TB-pos (Figure 4.5).

Table 4.2 Probability of being TB-pos of each sample in the testing group.

		PROBABILITY OF BEING TB-POS	
	Sample group	Conventional Lasso	Adaptive Lasso
TB-NEG	Healthy_3	0.1351899	0.04651185
	Healthy_7	0.06130411	0.04693872
	TB suspects_1	0.91642564	0.99597535
	TB suspects_6	0.17668854	0.03327371
TB-POS	TB-1_2	0.87502483	0.93547903
	TB-1_7	0.4941328	0.73388035
	TB-2_9	0.59158918	0.98346325
	TB-2_10	0.95269967	0.99139639

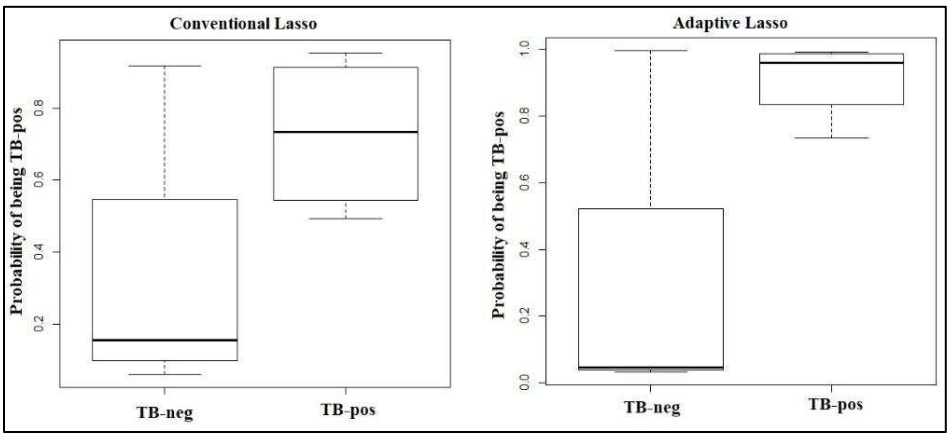

Figure 4.5 Discrimination of test-set samples by TB-status using two different predictive models. Box-plots show the mean and max and min values. TB-pos: samples from TB-1 and TB-2 Groups. TB-neg: samples from Healthy and TB suspects groups. N=4 per groups.

Even though both models were designed to perform a binomial distribution of the samples as either TB-neg or TB-pos, we explored the use of these models to calculate the probability of each test sample to be TB-pos and initially plotted the result according to the four groups of this study (Figure 4.6). Despite the very low number of samples per category, surprisingly, the adaptive Lasso model showed a very strong power to separate the healthy group from the TB-2 group (Figure 4.6). The proteins selected by each model could segregate the samples according to their clinical classification.

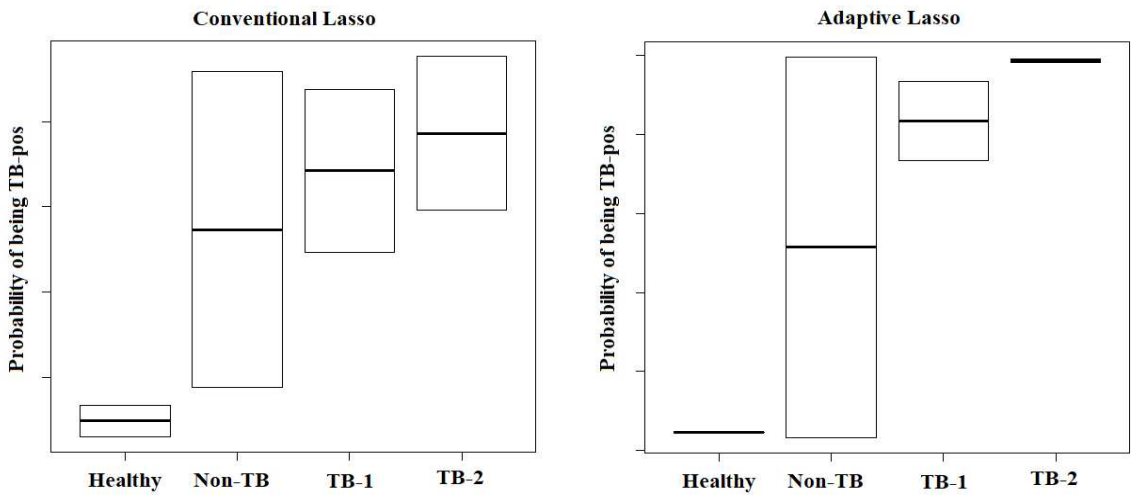

Figure 4.6 Classification of the test-samples into the four groups of study by the two predictive models. The boxplots show the average and max and min values. N=2 samples per group.

Specifically, the adaptive Lasso model could correctly categorize all TB positive samples. To further evaluate the discriminatory capacity of each protein signature, we calculated the AUC-ROC. In ROC curves, true positive results are plotted against the false positive rate (1-specificity). The AUC-ROC for the conventional Lasso was 0.81 and 0.75 for the adaptive Lasso. The lower AUC for the adaptive Lasso model is explained by the sample TB suspects_1 which had a very high probability of being TB-pos. We evaluated the training set with the adaptive Lasso model to verify how samples were classified using only nine proteins. This procedure is not ideal, since the model will have a bias towards a correct discrimination of the samples that were used to develop the model. As expected the discrimination of the samples was 100% accurate (Figure 4.7)

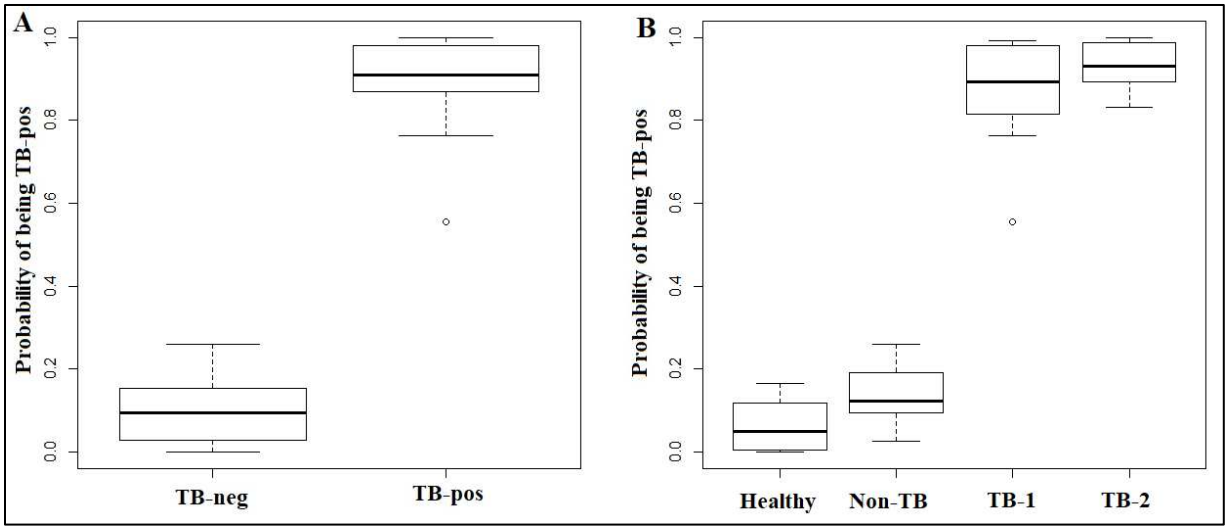

Figure 4.7 Classification of the training-set into different categories using the adaptive Lasso model. The boxplots show the average and max and min values. A. N=16 samples per group. B. N=8 samples per group.

4.3.4 Proteomic characterization of MRM+-TB-suspect group by HRM-SWATH-MS

HRM-SWATH-MS of the 10 MRM+-TB-suspect samples identified 223 proteins. Acquisition of spectral maps, identification, and quantification of proteins were generated as previously described in chapter 3 of this dissertation. Data were normalized to correct for variations among biological replicates (Figure 4.8). The normalized response was used for analysis.

4.3.5 Classification of MRM+-TB suspects according to the linear regression model

The adaptive Lasso model was used to evaluate the probability of the 10 MRM+-TB suspect samples being classified as TB-pos. The suspect samples were obtained from patients with clinical manifestations of TB who visited health care facilities and were remitted for healthcare personnel to routine TB diagnosis. They were classified as negative for TB based on microscopic examination of sputum and culture. However, these samples demonstrated one or more *M. tuberculosis* peptides by MRM/MS assay (data not shown). Nine samples were categorized as TB-neg by the adaptive Lasso regression model (Figure 4.9), five samples showed less than 1% probability of being TB-pos and only one sample showed 56% probability of being TB-pos.

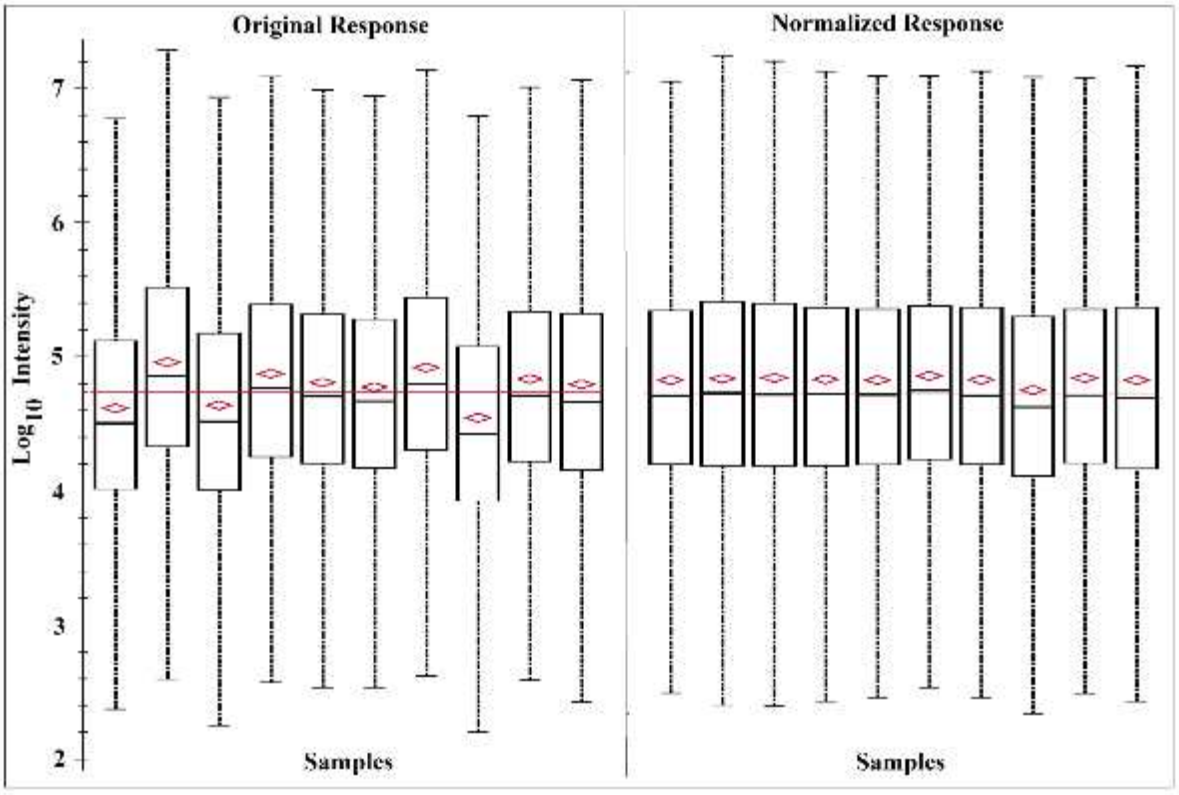

Figure 4.8 Normalization of intensities of identified peptides. The left panel shows the fluctuations in the intensities of identified peptides due to variability in the sample loading and/or machine performance. Right panel shows the normalized values.

4.4 Conclusions

Here we proposed a complementary modeling approach that can be applied very early in research and development studies, to select candidate TB biomarkers from proteomic experiments. Theoretically, the Lasso regression analysis allows the selection of the best predictors from large data sets. However, when the data contain collinear variables Lasso will select one of them. We tested here that changes in the randomization of the data led us to select different models, which can introduce selection bias. This variation in model selection was decreased by using adaptive Lasso (18).

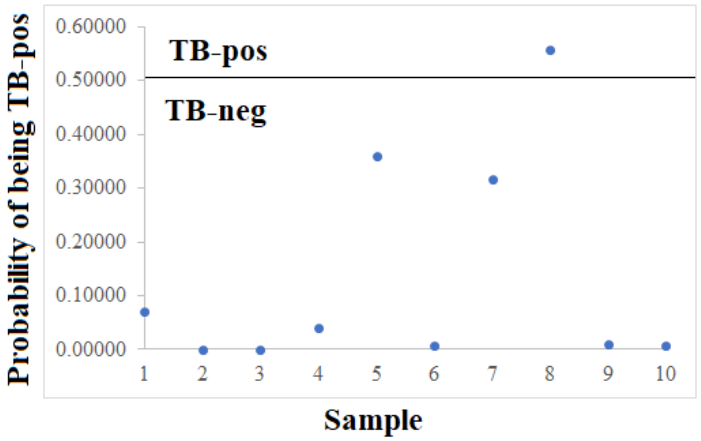

Figure 4.9 Prediction of MRM+-TB suspects by a model regression generated with adaptive Lasso method. The y axis showed the probability of a sample to be TB positive. The line at 0.5 is an arbitrary threshold based on the dichotomic classification of the model. TB-pos defined by TB-1 and TB-2 samples and TB-neg defined by healthy and TB suspects samples.

A major limitation of our study was the low number of samples, especially for the training set. Still, the adaptive Lasso model correctly classified 87.5% of the test samples (7/8) and within this, 100% (4/4) of the TB positive samples.

The selection of potential candidates is a complicated task in proteomic studies, because of the large number of explanatory variables. In the case of serum-derived biomarkers, the physiological interaction among the proteins could lead to high collinearity. Some proteins are in higher abundance because they are transported by (or interacting with) other candidates. For example, the retinol binding protein (RBP) is transported in plasma by transthyretin (TTR); thus, high levels of RBP could be in direct proportionality to TTR levels (31). Consequently, an informative look for collinear predictors and evaluation of their relevant biological meaning should be included prior to final selection of the Lasso predicted candidates to be tested in validation studies. Validation of the candidates found in this study is warranted. MRM-MS will allow the testing of all candidate markers found here in an independent cohort of patients. To this end, it is important to consider the evaluation of protein variants (defined by proteomics studies) rather than the canonical protein, as advised by Agranoff in 2006 (7).

Next steps should involve the combination of two or more sets of models predicting active TB patients or LTBI individuals. For instance, the combination of the human protein signature (as the one found in the present study) with *Mycobacterium* proteins identified by MRM-MS assays could improve the overall predictive capacity of such tests. In a preliminary attempt of this strategy, we decided to evaluate a sample set of TB suspects that were negative by sputum and culture and positive by an MRM-MS assay detecting mycobacterial peptides. Our results showed that the predictive model confirmed the diagnosis obtained by routine diagnostic tests in 9 of 10 samples.

References

1. World Health Organization. Global Tuberculosis Report. 2016.
2. Steingart KR, Schiller I, Horne DJ, Pai M, Boehme CC, Dendukuri N. Xpert® MTB/RIF assay for pulmonary tuberculosis and rifampicin resistance in adults. Cochrane Database Syst Rev. 2014;1:CD009593.
3. Steingart KR, Ng V, Henry M, Hopewell PC, Ramsay A, Cunningham J, et al. Sputum processing methods to improve the sensitivity of smear microscopy for tuberculosis: a systematic review. Lancet Infect Dis. 2006;6(10):664-74.
4. Organization WH. The use of loop-mediated isothermal amplification (TB-LAMP) for the diagnosis of pulmonary tuberculosis: policy guidance. Geneva; 2016.
5. Sharma SK, Mohan A, Sharma A. Challenges in the diagnosis & treatment of miliary tuberculosis. The Indian Journal of Medical Research. 2012;135(5):703-30.
6. Achkar JM, Cortes L, Croteau P, Yanofsky C, Mentinova M, Rajotte I, et al. Host Protein Biomarkers Identify Active Tuberculosis in HIV Uninfected and Co-infected Individuals. EBioMedicine. 2015;2(9):1160-8.
7. 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. Lancet. 2006;368(9540):1012-21.
8. De Groote MA, Nahid P, Jarlsberg L, Johnson JL, Weiner M, Muzanyi G, et al. Elucidating novel serum biomarkers associated with pulmonary tuberculosis treatment. PLoS One. 2013;8(4):e61002.
9. Nahid P, Bliven-Sizemore E, Jarlsberg LG, De Groote MA, Johnson JL, Muzanyi G, et al. Aptamer-based proteomic signature of intensive phase treatment response in pulmonary tuberculosis. Tuberculosis (Edinb). 2014;94(3):187-96.
10. De Groote MA, Sterling DG, Hraha T, Russell T, Green LS, Wall K, et al. Discovery and Validation of a Six-Marker Serum Protein Signature for the Diagnosis of Active Pulmonary Tuberculosis. J Clin Microbiol. 2017.
11. Thompson AM, Blaser A, Palmer BD, Anderson RF, Shinde SS, Launay D, et al. 6-Nitro-2,3- dihydroimidazo[2,1-b][1,3]thiazoles: Facile synthesis and comparative appraisal against tuberculosis and neglected tropical diseases. Bioorg Med Chem Lett. 2017;27(11):2583-9.
12. LaBaer J. So, you want to look for biomarkers (introduction to the special biomarkers issue). J Proteome Res. 2005;4(4):1053-9.
13. Haas CT, Roe JK, Pollara G, Mehta M, Noursadeghi M. Diagnostic 'omics' for active tuberculosis. BMC Med. 2016;14:37.
14. Ratzinger F, Bruckschwaiger H, Wischenbart M, Parschalk B, Fernandez-Reyes D, Lagler H, et al. Rapid diagnostic algorithms as a screening tool for tuberculosis: an assessor blinded cross-sectional study. PLoS One. 2012;7(11):e49658.
15. Su Y, Shi Q, Wei W. Single cell proteomics in biomedicine: High-dimensional data acquisition, visualization, and analysis. Proteomics. 2017;17(3-4).
16. Morozova O, Levina O, Uusküla A, Heimer R. Comparison of subset selection methods in linear regression in the context of health-related quality of life and substance abuse in Russia. BMC Med Res Methodol. 2015;15:71.
17. Tibshirani R. Regression Shrinkage and Selection via the Lasso. Journal of the Royal

Statistical Society Series B. 1996
58(1):267-88.
18. Zou H. The Adaptive Lasso and Its Oracle Properties. Journal of the American Statistical Association. 2006;101(476):1418-29.
19. Mehaffy C, Dobos KM, Nahid P, Kruh-Garcia NA. Second generation multiple reaction monitoring assays for enhanced detection of ultra-low abundance *Mycobacterium tuberculosis* peptides in human serum. Clin Proteomics. 2017;14:21.
20. Kruh-Garcia NA, Wolfe LM, Chaisson LH, Worodria WO, Nahid P, Schorey JS, et al. Detection of *Mycobacterium tuberculosis* peptides in the exosomes of patients with active and latent *M. tuberculosis* infection using MRM-MS. PLoS One. 2014;9(7):e103811.
21. Callister SJ, Barry RC, Adkins JN, Johnson ET, Qian WJ, Webb-Robertson BJ, et al. Normalization approaches for removing systematic biases associated with mass spectrometry and label- free proteomics. J Proteome Res. 2006;5(2):277-86.
22. Friedman JH, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. 2010. 2010;33(1):22.
23. Lu LL, Chung AW, Rosebrock TR, Ghebremichael M, Yu WH, Grace PS, et al. A Functional Role for Antibodies in Tuberculosis. Cell. 2016;167(2):433-43.e14.
24. Almeida ML, Barbieri MA, Gurgel RQ, Abdurrahman ST, Baba UA, Hart CA, et al. alpha1-acid glycoprotein and alpha1-antitrypsin as early markers of treatment response in patients receiving the intensive phase of tuberculosis therapy. Trans R Soc Trop Med Hyg. 2009;103(6):575-80.
25. Martìnez Cordero E, Gonzàlez MM, Aguilar LD, Orozco EH, Hernàndez Pando R. Alpha-1-acid glycoprotein, its local production and immunopathological participation in experimental pulmonary tuberculosis. Tuberculosis (Edinb). 2008;88(3):203-11.
26. Li C, He X, Li H, Zhou Y, Zang N, Hu S, et al. Discovery and verification of serum differential expression proteins for pulmonary tuberculosis. Tuberculosis (Edinb). 2015;95(5):547-54.
27. Principe S, Jones EE, Kim Y, Sinha A, Nyalwidhe JO, Brooks J, et al. In-depth proteomic analyses of exosomes isolated from expressed prostatic secretions in urine. Proteomics. 2013;13(10- 11):1667-71.
28. Prunotto M, Farina A, Lane L, Pernin A, Schifferli J, Hochstrasser DF, et al. Proteomic analysis of podocyte exosome-enriched fraction from normal human urine. J Proteomics. 2013; 82:193-229.
29. Meckes DG, Gunawardena HP, Dekroon RM, Heaton PR, Edwards RH, Ozgur S, et al. Modulation of B-cell exosome proteins by gamma herpesvirus infection. Proc Natl Acad Sci U S A. 2013;110(31):E2925-33.
30. Gonzalez-Begne M, Lu B, Han X, Hagen FK, Hand AR, Melvin JE, et al. Proteomic analysis of human parotid gland exosomes by multidimensional protein identification technology (MudPIT). J Proteome Res. 2009;8(3):1304-14.
31. Naylor HM, Newcomer ME. The structure of human retinol-binding protein (RBP) with its carrier protein transthyretin reveals an interaction with the carboxy terminus of RBP. Biochemistry. 1999;38(9):2647-53.

Chapter 5: Concluding remarks and future direction

The findings of Dr. Robert Koch set the basis for the current gold standard diagnostic tool (culture of sputum) and for the most widely used routine diagnostic test for TB in low-and-middle-income countries (microscopic examination of sputum) (1, 2). These tests fulfill two important considerations for the diagnosis of TB, they are relatively easy-to-do and inexpensive. Unfortunately, several studies have demonstrated that sputum-based diagnostic tests for TB have significant disadvantages including low sensitivity (2, 3). Novel sputum independent TB diagnostic tools are urgently needed. Several reasons support the study of circulating exosomes as a promising alternative to overcome the need for sputum in TB diagnosis. First, the composition of exosomes changes depending on the health status of the patient and depending on the cell of origin (4-6). Second, the separation of exosomes from serum aids in removing most of the highly abundant proteins in this biofluid, and improves biomarker discovery. Finally, recent studies demonstrated that exosomes carry mycobacterial proteins (7, 8). In this dissertation, I decided to focus on the changes of the host proteome of exosomes upon *Mycobacterium tuberculosis* infection, a gap of knowledge. We found that *M. tuberculosis* altered the protein composition of exosomes released from *in vitro* infected human macrophages. These findings suggest that a similar phenomenon occurs during *in vivo* infection. Therefore, we decided to explore the changes in the exosomal proteome as source of biomarkers to identify active TB patients. An important limitation of the work proposed in this dissertation was the lack of a robust method to isolate a pure exosome sample from serum. In our initial in vitro experiments, we attempted to concentrate sub-populations of exosomes using immunoprecipitation assays targeting abundant proteins including Cathepsin D and Coronin 1A and 1C. Unfortunately, none of our results from immunoprecipitation assays were successful. Figure 5.1 and table 5.1,

summarize an example of one of the immunocapture assays. Proteomic analysis revealed that most of the captured material were contaminants such as keratin. Alternatively, to purify exosome rich fractions from sera samples, we decided to use size exclusion chromatography, to avoid most of the common limitations observed with ultracentrifugation and precipitation methods.

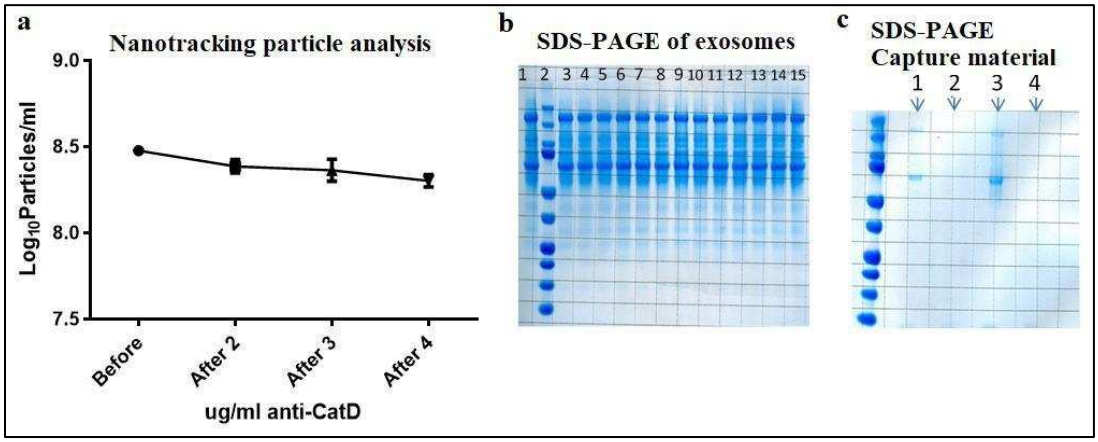

Figure 5.1 Immunocapture of exosomes targeting cathepsin D. Exosomes from *M. tuberculosis*-infected THP1 cells (100µg/well) were seeded in a 96 wells plate pre-coated with three different concentrations of CatD (2µg/ml, 3µg/ml and 4µg/ml). **a.** After 24 hours of incubation at 4 °C the supernatant was retrieved for analysis, the number of vesicles in the sample before and after the incubation was not significantly different**. b.** Approximately 25 µg of exosomes were run in SDS-PAGE and stained with simply-blue Coomassie stain. There was not difference in the protein load before (Lane 1) and after binding (Lanes 3, 4, 10 and 11: 2µg/ml, lane 5, 6, 12 and 13: 3µg/ml and lane 7, 8, 14 and 15: 4µg/ml). lane 2: protein molecular weight ladder. **c.** After retrieving the supernatant from the plates, the potential bound exosomes were detached from the plate using a solution of acetic acid 1M (pH=2.49) the material was further lysed using RIPA buffer. An aliquot was run in SDS-PAGE and stained with simply-blue.

Table 5.1 Proteins found by LC-MS/MS in the captured material using a plate coated with anti-cathepsin D.

Identified Proteins	Accession #	C4-TSC	TB3-TSC	TB4-TSC
Keratin, type II cytoskeletal 1	P04264	13	33	35
Keratin, type I cytoskeletal 9	P35527	1	24	37
Keratin, type I cytoskeletal 10	P13645	25	14	13
Serum albumin	P02768	35	20	25
Hemoglobin subunit alpha	P69905	14	14	6
Fibronectin	P02751-17	6	0	14
Keratin, type II cytoskeletal 2 epidermal	35908	6	2	7
Alpha-2-macroglobulin *	P01023	0	0	15

Actin alpha skeletal muscle*	A6NL76	0	0	10
Trypsin inhibitor heavy chain H2	P19823	0	0	7
Matrix metalloproteinase-9 *	P14780	0	0	5
Adenylyl cyclase-associated protein 1 *	Q01518	0	0	2

*proteins previously found in exosomes. C4: Control (no capture antibody). TB3 and TB4: Samples seeded in antibody-coated wells. TSC: total spectral counts.

In the future, the immunocapture assays may be improved by adapting the labeling strategy that we developed in chapter 2 to identify protein epitopes readily available on the surface of intact exosomes. This information could lead the design of antibodies to develop a capture platform of specific subpopulation of exosomes of interest. The fact that the label molecule binds to the naturally folded protein in the membrane of intact exosomes, suggests that labeled peptides are accessible for antibody targeting.

An alternative scenario in the study of exosomes in the context of TB infection could be to evaluate the impact of *M. tuberculosis* on exosomes biogenesis during macrophage infection. It was demonstrated that *M. tuberculosis* interrupts the fusion of phagosomes with lysosomes in a process mediated by lipoarabinomannan (LAM). On the other hand, the bacterium stimulates the fusion of early endosomes with the phagosome, using phosphatidylinositol mannoside (PIM) (14, 15). It was also demonstrated that PIM stimulates the accumulation of membrane proteins in the phagosome (14). Since LAM and PIM evidently affect the vesicle trafficking of infected cells they may affect exosome biogenesis. The study of these effects could reveal novel mechanisms of exosomes regulation during intracellular infection.

One limitation while studying exosomes during mycobacterial infection is the fact that *M. tuberculosis* actively releases membrane vesicles (16) that could co-isolate with exosomes. We considered identifying proteins exclusively loaded in *M. tuberculosis* membrane vesicles (MTB-EVs) to either rule out co-purification with host exosomes or allow the separation of MTB-EVs via affinity purification. We ran a set of preliminary experiments to identify proteins exclusively

loaded in MTB-EVs and other soluble proteins from supernatant of *M. tuberculosis* cultures. These results are presented at the end of this dissertation (Appendix I). Further confirmation of the presented findings and future work to evaluate the role of MTB-EVs during TB infection is currently ongoing.

In chapter 3, we reported for the first time, changes in the host proteome of serum-derived exosomes from active TB patients compared to TB-negative patients and healthy individuals. We evaluated three different groups, TB1: paucibacillary TB (sputum-, culture+), TB2: active TB (sputum+, culture+), TB suspects (sputum-, culture-), and a group of healthy individuals from non-endemic regions for TB. Three proteins, FCGR3A, lysozyme C and allograft inflammatory factor 1, showed a step wise increase and six proteins: fetuin-A, angiotensinogen, coagulation factor XII, PGRP-L, CBG, and CPN2, showed a step-wise decrease, from the healthy group to the TB-2 group. An important limitation of our study was the low number of samples per category, due to this, we could not develop separated analysis based on HIV status. As a partial solution to this limitation, we included HIV positive and negative samples in each group.For the selection of potential candidate TB biomarkers, we must consider their associated clinical applications. For instance, here we found that the protein transthyretin (TTR) was always significantly higher in healthy individuals versus the other three groups. TTR has previously been reported as potential TB-biomarker (17). However, considering the clinical application of the TTR measurement (marker to evaluate nutritional status (18)), this protein must be taken with care. Since most people suffering TB will have a weakened nutritional status, TTR will be decreased, however extreme poverty and many other conditions common in TB endemic areas will also affect the nutritional status of the evaluated population, and consequently TTR would be also decreased.

An influential factor to improve the finding of accurate TB biosignatures will be the characterization of a "universal proteomic profile of TB" as was stated by Haas in 2016 (19). However, I would hypothesize that there is not a unique universal proteomic profile for TB patients, instead, several profiles could be defined based on different geographical localizations, HIV-status, and even age groups. The use of HRM-SWATH-MS, proposed by Bruderer (20) and used in chapter 3, could represent an interesting alternative to solve the problem. The HRM approach involves the generation of retention time-normalized spectral libraries using a set of synthetic peptides (iRT peptides) (21). Since these libraries are obtained by shotgun proteomics (Data Dependent Acquisition-DDA analysis) the proteome- coverage of such selected TB sample groups will depend on fractionation and multiple-run processing (20). A suitable strategy could be achieved by pooling samples from epidemiologically compatible patients (for instance, Spectral library X: active TB/HIV+/Peru- South America/5 to 10 years old). Then, each pool could be fractionated (10 to 20 fractions) and run 6 times maximizing the peptide-protein coverage. This sounds expensive, technically challenging, and time consuming. However, the information generated could be stored in internet-based repositories (virtually forever) such as, http://www.peptideatlas.org/speclib/ and could be used in multiple studies, to extract information from SWATH-MS ion-maps obtained using standardized protocols and iRT normalization. Overall, the development of iRT-spectral libraries will improve comparability across laboratories and identification of potential TB-biomarkers. This strategy could decrease the high variability of TB biomarkers among studies. An important problem for the discovery of disease predictors using proteomics studies is the large amount of proteins that are found per sample. Due to the complex interaction among several groups of proteins, the selection of the "best predictors" for further evaluation is challenging. In this dissertation, we used two approaches to identify

potential biomarkers for active TB. In chapter 3 we conducted ANOVA analysis to identify proteins showing statistically different intensities among the groups of study. This approach is widely used, however the number of potential candidates selected is very large. For example, in our data set the proteins significantly different between healthy and TB2 were 167. Out of this number, we normally select the proteins with higher distance among means by fold change analysis. Alternatively, in chapter 4, we proposed the use of a regression model initially developed by Tibshirani and later improved by Zou (Lasso and Adaptive respectively) to select a subset of predictor (22, 23). Lasso regression allowed–in a continuous manner—the elimination of collinear predictors with smaller predictive power. We finally obtained a regression model with nine predictors that allowed to segregate TB positive from TB negative samples. A limitation with this approach is the variation in predictor selection at different randomization patterns. We confirmed that adaptive Lasso decreased the variation among models. We hypothesize that including a larger sample set will stabilize even further the model selection. After this, we concluded that the combination of the two approaches will generate the best set of potential candidates for independent testing experiments. The final model included three proteins of the complement system, two immunoglobulin chains, the acute-phase plasma protein A1AG1, the anti- inflammatory metalloproteinase CBPN, the glucose binding protein glucokinase GCK, and the protein Sex hormone-binding globulin SHBG. In addition to the previous proteins, the nine proteins found in chapter #3 could be further tested as possible TB biomarkers, using a multiplex approach such as SRM-MS, in an independent sample group. Two of the proteins identified in this study have shown promising results as TB biomarkers in previous studies. Fetuin A and SHGB showed differential expression in different discovery studies (2D-electrophoresis and LC- MS/MS, respectively) and both proteins were validated in each study by

ELISA (24, 25). Our findings suggest that Fetuin A and SHGB could be concentrated in the exosome-rich fraction of human serum, if this hypothesis is true, the evaluation of exosomes instead of whole serum, will increase the predictive power of these proteins. A next step could be the evaluation of these proteins (found in three independent studies) in exosome samples from TB patients and controls by ELISA.

There are two relevant scenarios where exosomes from serum could play a significant role, identification of individuals with latent TB infection-LTBI, (especially those at higher risk of conversion to active TB), and surrogate end points to evaluate treatment response. Reactivation of LTBI is one the major sources of active TB cases every year. A major cause of LTBI came from active TB patients who received efficacious treatment but did not achieve complete sterilization which means that some bacilli remained in the lungs in a quiescent state contained in granulomas. Another common cause for LTBI is when healthy individuals get in contact with the bacillus and without the development of symptoms, but the individual remains infected for decades. It is also known that people with a weakened immunological system (HIV-AIDS, pregnancy, aging among others) are at higher risk of TB reactivation. However, the definition of a biomarker to identify individuals at higher risk of reactivation is still a worldwide priority.

Regarding treatment response, the most widely used surrogate endpoint biomarker for treatment effect in TB is the 2-month sputum conversion (26). It could be a single measurement in solid media (LJ) at 8 weeks, or several time points during treatment. Recent data demonstrated the limited capacity of this biomarker to predict relapse-free cure (27). Alternatively, liquid media-based systems such as MGIT have been used, showing similar results to LJ (28). Another alternative is the detection of mycobacterial DNA in sputum samples by Xpert MTB/RIF as a replacement of culture. However, Friedrich *et al.,* evaluated Xpert MTB/RIF assay in the clinical

trial REMoxTB, showing that Xpert MTB/RIF had very low specificity (47%) compared to culture (29). Alternatively, the evaluation of host-derived biomarkers to predict treatment response is under intense research. One study showed that, Apolipoprotein B could be a potential biomarker of treatment adherence and had potential to predict clinical outcome (30). A transcriptomic study revealed an increased expression of genes involved in cell-mediated cytotoxicity: perforin, granulysin, rab27A, and fas-ligand, in patients who initially had TB treatment success but experienced relapse within the two following years (31). The evaluation of immune markers had shown that the concentration of interferon-inducible protein 10 (IP-10), and vascular endothelial growth factor (VEGF), significantly decreased after treatment completion (32, 33). An important number of additional host-derived biomarkers has been tested: levels of heme oxygenase-1 (lower following anti-TB therapy) (34), CD27 and serum amyloid A (35). Even urine metabolites have shown capacity to predict early response to TB treatment (36). Overall, there are not published studies evaluating serum-derived exosomes as surrogate endpoints of treatment response. Our findings from chapter 3 showed that several proteins seem to be associated with different stages of TB disease as well as health status. It could be possible that proteomic changes in exosomes of TB patient receiving treatment, have the potential to predict treatment outcome.

Finally, our results suggest that exosomes derived from serum samples carry information that could improve the identification of TB patients. However, the current evidence suggests that there is not a single approach to find the "perfect biosignature" for TB diagnosis. The design of algorithms including two or more approaches, when is technically feasible, for instance, the combination of bacterial and host-derived markers from serum-derived exosomes, could result in a stronger tool that definitively help to improve the current situation of TB worldwide.

References

1. World Health Organization. Global Tuberculosis Report. 2016.
2. Koch R. Classics in infectious diseases. The etiology of tuberculosis: Robert Koch. Berlin, Germany 1882. Rev Infect Dis. 1982;4(6):1270-4.
3. Flores LL, Steingart KR, Dendukuri N, Schiller I, Minion J, Pai M, et al. Systematic review and meta-analysis of antigen detection tests for the diagnosis of tuberculosis. Clin Vaccine Immunol. 2011;18(10):1616-27.
4. Stoorvogel W, Kleijmeer MJ, Geuze HJ, Raposo G. The biogenesis and functions of exosomes. Traffic. 2002;3(5):321-30.
5. Rabinowits G, Gerçel-Taylor C, Day JM, Taylor DD, Kloecker GH. Exosomal microRNA: a diagnostic marker for lung cancer. Clin Lung Cancer. 2009;10(1):42-6.
6. Simpson RJ, Jensen SS, Lim JW. Proteomic profiling of exosomes: current perspectives. Proteomics. 2008;8(19):4083-99.
7. Cheng Y, Schorey JS. Exosomes carrying mycobacterial antigens can protect mice against *Mycobacterium tuberculosis* infection. Eur J Immunol. 2013;43(12):3279-90.
8. Kruh-Garcia NA, Wolfe LM, Chaisson LH, Worodria WO, Nahid P, Schorey JS, et al. Detection of *Mycobacterium tuberculosis* peptides in the exosomes of patients with active and latent *M. tuberculosis* infection using MRM-MS. PLoS One. 2014;9(7):e103811.
9. de Menezes-Neto A, Sáez MJ, Lozano-Ramos I, Segui-Barber J, Martin-Jaular L, Ullate JM, et al. Size-exclusion chromatography as a stand-alone methodology identifies novel markers in mass spectrometry analyses of plasma-derived vesicles from healthy individuals. J Extracell Vesicles. 2015;4:27378.
10. Lobb RJ, Becker M, Wen SW, Wong CS, Wiegmans AP, Leimgruber A, et al. Optimized exosome isolation protocol for cell culture supernatant and human plasma. J Extracell Vesicles. 2015;4:27031.
11. Welton JL, Webber JP, Botos LA, Jones M, Clayton A. Ready-made chromatography columns for extracellular vesicle isolation from plasma. J Extracell Vesicles. 2015;4:27269.
12. Taylor DD, Shah S. Methods of isolating extracellular vesicles impact down-stream analyses of their cargoes. Methods. 2015.
13. Kowal J, Arras G, Colombo M, Jouve M, Morath JP, Primdal-Bengtson B, et al. Proteomic comparison defines novel markers to characterize heterogeneous populations of extracellular vesicle subtypes. Proc Natl Acad Sci U S A. 2016;113(8):E968-77.
14. Vergne I, Chua J, Lee HH, Lucas M, Belisle J, Deretic V. Mechanism of phagolysosome biogenesis block by viable *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A. 2005;102(11):4033-8.
15. Vergne I, Fratti RA, Hill PJ, Chua J, Belisle J, Deretic V. *Mycobacterium tuberculosis* phagosome maturation arrest: mycobacterial phosphatidylinositol analog phosphatidylinositol mannoside stimulates early endosomal fusion. Mol Biol Cell. 2004;15(2):751-60.
16. Prados-Rosales R, Baena A, Martinez LR, Luque-Garcia J, Kalscheuer R, Veeraraghavan U, et al. Mycobacteria release active membrane vesicles that modulate immune responses in a TLR2-dependent manner in mice. J Clin Invest. 2011;121(4):1471-83.
17. 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. Lancet. 2006;368(9540):1012-21.
18. Dellière S, Cynober L. Is transthyretin a good marker of nutritional status? Clin Nutr. 2017;36(2):364-70.
19. Haas CT, Roe JK, Pollara G, Mehta M, Noursadeghi M. Diagnostic 'omics' for active tuberculosis. BMC Med. 2016;14:37.
20. Bruderer R, Bernhardt OM, Gandhi T, Miladinović SM, Cheng LY, Messner S, et al. Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. Mol Cell Proteomics. 2015;14(5):1400-10.
21. Escher C, Reiter L, MacLean B, Ossola R, Herzog F, Chilton J, et al. Using iRT, a normalized retention time for more targeted measurement of peptides. Proteomics. 2012;12(8):1111-21.
22. Zou H. The Adaptive Lasso and Its Oracle Properties. Journal of the American Statistical Association. 2006;101(476):1418-29.
23. Tibshirani R. Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society Series B. 1996

58(1):267-88.
24. Tanaka T, Sakurada S, Kano K, Takahashi E, Yasuda K, Hirano H, et al. Identification of tuberculosis-associated proteins in whole blood supernatant. BMC Infect Dis. 2011;11:71.
25. Li C, He X, Li H, Zhou Y, Zang N, Hu S, et al. Discovery and verification of serum differential expression proteins for pulmonary tuberculosis. Tuberculosis (Edinb). 2015;95(5):547-54.
26. Nahid P, Saukkonen J, Mac Kenzie WR, Johnson JL, Phillips PP, Andersen J, et al. CDC/NIH Workshop. Tuberculosis biomarker and surrogate endpoint research roadmap. Am J Respir Crit Care Med. 2011;184(8):972-9.
27. Phillips PP, Mendel CM, Burger DA, Crook AM, Crook A, Nunn AJ, et al. Limited role of culture conversion for decision-making in individual patient care and for advancing novel regimens to confirmatory clinical trials. BMC Med. 2016;14:19.
28. Gillespie SH, Crook AM, McHugh TD, Mendel CM, Meredith SK, Murray SR, et al. Four-month moxifloxacin-based regimens for drug-sensitive tuberculosis. N Engl J Med. 2014;371(17):1577-87.
29. Friedrich SO, Rachow A, Saathoff E, Singh K, Mangu CD, Dawson R, et al. Assessment of the sensitivity and specificity of Xpert MTB/RIF assay as an early sputum biomarker of response to tuberculosis treatment. Lancet Respir Med. 2013;1(6):462-70.
30. Albanna AS, Bachmann K, White D, Valiquette C, Menzies D. Serum lipids as biomarkers for therapeutic monitoring of latent tuberculosis infection. Eur Respir J. 2013;42(2):547-50.
31. Cliff JM, Cho JE, Lee JS, Ronacher K, King EC, van Helden P, et al. Excessive Cytolytic Responses Predict Tuberculosis Relapse After Apparently Successful Treatment. J Infect Dis. 2016;213(3):485-95.
32. Riou C, Perez Peixoto B, Roberts L, Ronacher K, Walzl G, Manca C, et al. Effect of standard tuberculosis treatment on plasma cytokine levels in patients with active pulmonary tuberculosis. PLoS One. 2012;7(5):e36886.
33. Tonby K, Ruhwald M, Kvale D, Dyrhol-Riise AM. IP-10 measured by Dry Plasma Spots as biomarker for therapy responses in *Mycobacterium Tuberculosis* infection. Sci Rep.

2015;5:9223.
34. Andrade BB, Pavan Kumar N, Mayer-Barber KD, Barber DL, Sridhar R, Rekha VV, et al. Plasma heme oxygenase-1 levels distinguish latent or successfully treated human tuberculosis from active disease. PLoS One. 2013;8(5):e62618.
35. Goletti D, Petruccioli E, Joosten SA, Ottenhoff TH. Tuberculosis Biomarkers: From Diagnosis to Protection. Infect Dis Rep. 2016;8(2):6568.

Appendix I: Alternative application of a method to isolate exosomes to purify extracellular vesicles released from *Mycobacterium tuberculosis*.

Introduction

The study of membrane vesicles (MV) derived from the genus *Mycobacterium* is relatively new. MVs released from *M. ulcerans* were first described in 2007 (1). Later Prados-Rosales *et al.,* developed a comprehensive characterization of MVs from a variety of mycobacteria species, including *M. bovis* and *M. tuberculosis* (2). The function of MVs is unknown but it has been demonstrated that MVs or extracellular vesicles (EV) from *M. tuberculosis* (MTB-EVs) have agonist effect on TLR2-dependent signaling which is related to their high content of lipoproteins (2). The function of MTB-EVs has been linked with iron acquisition (3). One study in 2013, demonstrated that the biogenesis of MTB-EVs could be genetically regulated (4). The isolation of MTB-EVs has been done following consecutive centrifugation steps (5). Considering our experience with isolating exosomes, we found that ultracentrifugation–a very common method used to isolate exosomes—generates exosome samples highly contaminated with soluble proteins what also are pelleted at high *g* force. Here, we tested the isolation of MTB-EVs using a size exclusion chromatographic resin which has two features: size exclusion and irreversible binding. We have previously tested this resin (CaptoCore 700) to isolate exosomes with promising results (unpublished data). Additionally, to evaluate which proteins were exclusively present in the soluble part of the secreted proteome of *M. tuberculosis* a MTB-EV-depleted fraction was obtained.

Materials and methods

Production of culture filtrate from *M. tuberculosis* H37Rv

Four different replicates of stocks of *M. tuberculosis* strain H37Rv were grown in solid media 7H11 for 21 days. Later, bacteria from each replicate were inoculated in one litter of Glycerol alanine salts (Gas) media and incubated at 37 °C at constant agitation. The culture filtrate (CF) was collected after 10 days of incubation to minimize the amount of lysis products in the collected material. The CF was filtered through 0.2 µm membrane for sterilization. The total protein concentration of CF samples was measured using the bicinchoninic acid assay (BCA, Thermo Scientific).

Isolation of membrane vesicle (MTB-EVs)

Five grams of total protein of each CF were concentrated using an Amicon centrifugal filter unit with a molecular weight cut-off (MWCO) of 3 KDa (EMD Millipore) to 2 ml. The concentrated CF was processed through a column with the resin CaptoCore (CaptoCore 700, GE Healthcare Life Sciences). CaptoCore is resin of beads cross-linked agarose of 90 µm diameter with holes with a MWCO of 700 KDa. The inner core of the beads is functionalized with octylamine. When the sample pass-through CaptoCore soluble proteins (smaller than 700 KDa) are capture in the inside of the beads while the MTB-EVs are collected in the flow-through. The column was washed twice with 1 ml of PBS 1X and the collected material (about 3 ml) was concentrated using a Amicon of 100 KDa MWCO (EMD Millipore) to 500 µl. The total protein concentration was measured using BCA.

Production of MTB-EVs-depleted CF (CF-D)

Five grams of total protein of each CF were concentrated using an Amicon centrifugal filter unit with a molecular weight cut-off (MWCO) of 3 KDa (EMD Millipore) to 2 ml. The

concentrated CF was further filtered using an Amicon of 300 KDa MWCO, to elute most of the soluble proteins in the CF but capturing the MTB-EVs in the membrane. The membrane was washed twice with 1ml of PBS to maximize the elution of the soluble proteins. the eluted material (about 3 ml) was concentrated using a Amicon of 3 KDa MWCO (EMD Millipore) to 500 µl. The total protein concentration was measured using microBCA.

The concentration and size distribution of the vesicles in MTB-EVs and CF-D, were evaluated by nanoparticle tracking analysis (NTA), using the NanoSight NS300 (Malvern Instruments). One µg of each sample was evaluated.

Proteomic analysis of MTB-EVs and CF-D and statistical analysis

Five micrograms of CF-D and $9x10^9$ MTB-EVs were processed for in gel digestion with trypsin as described in chapter 2 section 2.2.8.1. Digested peptides for each sample was injected using an EASY nanoLC-II system (Thermo Scientific, San Jose, CA). Peptides were purified and concentrated using an on-line enrichment column (EASY-Column, 100 µm ID × 2 cm ReproSil-Pur C18). Subsequent chromatographic separation was performed on a reverse phase nanospray column (EASY-Column, 3 µm, 75 µm ID × 100 mm ReproSil-Pur C18) using a 90 minutes linear gradient from 5–45% solvent B (100% ACN, 0.1% formic acid) at a flow rate of 400 nanoliters/min. Peptides were eluted directly into the mass spectrometer (Thermo Scientific Orbitrap Velos). The instrument was operated in Orbitrap-LTQ mode where precursor measurements were acquired in the Orbitrap (60,000 resolution) and MS/MS spectra (top 20) were acquired in the LTQ ion trap with a normalized collision energy of 35%. Mass spectra were collected over a m/z range of 400–2000 Da using a dynamic exclusion limit of 2 MS/MS spectra of a given peptide mass for 30 s (exclusion duration of 90 s). Compound lists of the resulting spectra were generated using Xcalibur 2.2 software (Thermo Scientific) with an S/N threshold of

1.5 and 1 scan/group. Tandem mass spectra were extracted, charge state deconvoluted and deisotoped by ProteoWizard (MSConvert version 3.0). Raw data files were converted to mzXML format and submitted to the Sorcerer2 integrated data analysis platform (Sage-N Research, version 5.0.1); subsequent MS/MS analysis was performed using SEQUEST (Sage-N Research, Milpitas, CA, USA; version v. 3.5). SEQUEST was set up to search the *M. tuberculosis* database TBv3_reverse_042110 database (042110, 181470 entries) assuming the enzymatic digestion with trypsin (after Arg or Lys). SEQUEST was searched with a fragment ion mass tolerance of 1.00 Da and a parent ion tolerance of 50 PPM. Scaffold (version Scaffold_4.5.1, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identification thresholds were set such that a peptide FDR of 1% and a peptide confidence threshold of 95% was achieved based on hits to the reverse database (6). Protein identifications were accepted if they could be established at greater than 95.0% probability to achieve an FDR less than 1.0% and contained at least 2 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm (7). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principle of parsimony. Differences in protein abundances between MTB-EV and CF-D were evaluated by t-test, using the normalized spectral abundance factor (NSAF) (8). P values < 0.05 were accepted as statistically significant.

Results

We found nine proteins that were significantly more abundant in MTB-EVs compared to the fraction of the CF that was depleted of MTB-EVs. The MTB-EVs were enriched with lipoproteins as it has been reported in previous publications (2, 4)

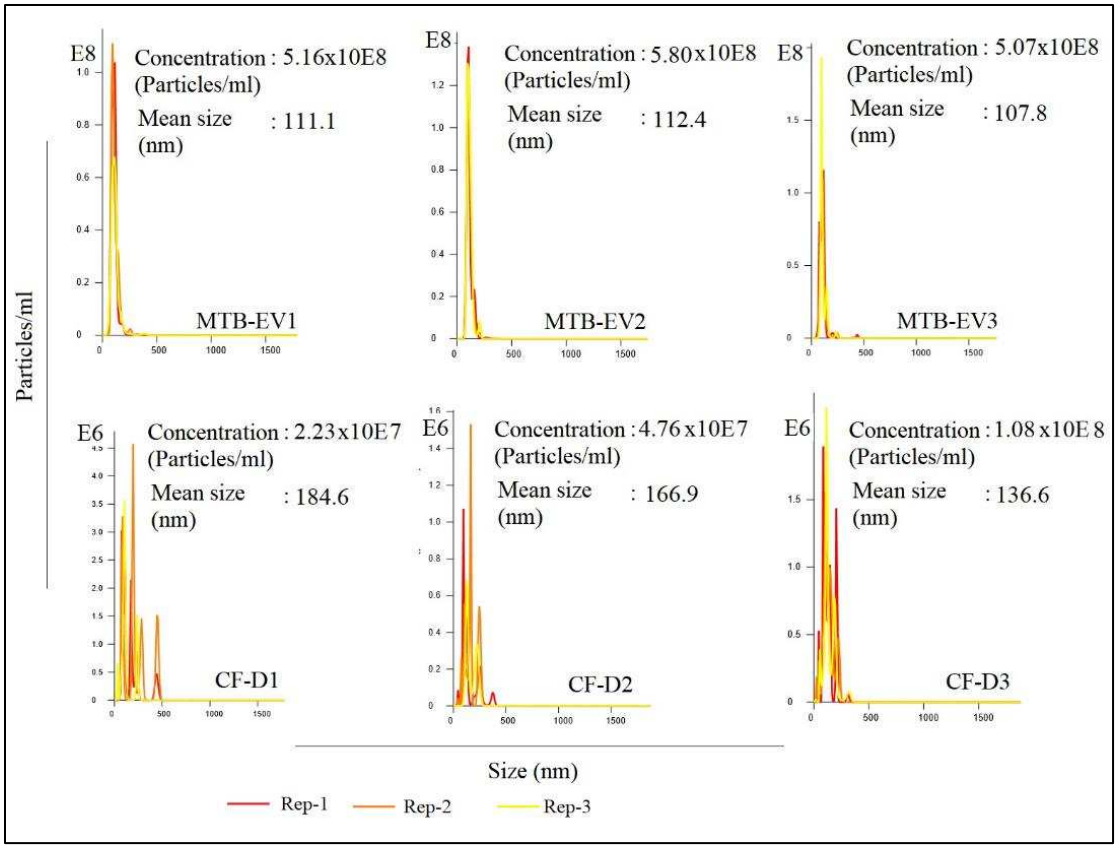

Figure 1 Light scattering analysis of MTB-EV and CF-D samples. Concentration and size distribution of nanovesicles present in MTB-EV and CF-D samples. Due to the low concentration of protein in MTB-EV samples the analysis was based on dilution factor. MTB-EV were processed at 1:2,000 dilution and CF-D at 1:100 dilution.

Table 1 Proteins significantly higher in MTB-EVs compared to CF-D.

Identified Proteins	T-test p-value	CF-D1	CF-D2	CF-D3	MTB-EV1	MTB-EV2	MTB-EV3
Lipoprotein LpqH	0.0029	9	34	8	55	123	81
Superoxide dismutase [Cu-Zn] SodC	0.015	0	0	1	18	21	53
Probable conserved lipoprotein LppZ	< 0.00010	5	5	3	4	11	10
Lipoprotein LprA	0.011	1	3	3	7	16	8
Putative phthiocerol dimycocerosate transporter LppX	0.013	0	0	0	6	17	7
Putative diacylated glycolipid transporter LprF	0.00017	0	0	0	3	9	7
Lipoarabinomannan carrier protein LprG	0.043	0	0	0	1	9	5
Putative integration host factor MihF*	0.14	0	0	0	2	3	0
Probable conserved lipoprotein DsbF	0.024	0	1	0	2	6	2
50S ribosomal protein L28-2	0.14	0	0	0	0	5	3

The statistical analysis was done using the normalized spectral abundance factor while the information presented in the table is the total spectral count. *this protein was not statistically different but was exclusively detected in MTB-EVs.

Table 2 Comprehensive list of Proteins found in MTB-EVs and CF-D.

Identified Proteins	MW	T-Test p-value	Profile	Accession Number	CF-D1	CF-D2	CF-D3	MTB-EV1	MTB-EV2	MTB-EV3
Lipoprotein LpqH GN=lpqH	15 kDa	0.0029	MTB-EV high	LPQH	9	34	8	55	123	81
Superoxide dismutase [Cu-Zn] GN=sodC	24 kDa	0.015	MTB-EV high	SODC	0	0	1	18	21	53
Probable conserved lipoprotein LppZ GN=lppZ	39 kDa	< 0.0001	MTB-EV high	I6Y293	5	5	3	4	11	10
Lipoprotein LprA GN=lprA	25 kDa	0.011	MTB-EV high	LPRA	1	3	3	7	16	8
Putative phthiocerol dimycocerosate transporter LppX GN=lppX	24 kDa	0.013	MTB-EV high	LPPX	0	0	0	6	17	7
Putative diacylated glycolipid transporter LprF GN=lprF	27 kDa	0.00017	MTB-EV high	LPRF	0	0	0	3	9	7
Lipoarabinomannan carrier protein LprG GN=lprG	25 kDa	0.043	MTB-EV high	LPRG	0	0	0	1	9	5
Probable conserved lipoprotein DsbF GN=dsbF	19 kDa	0.024	MTB-EV high	I6XYM2	0	1	0	2	6	2
Immunogenic protein MPT64 GN=mpt64	25 kDa	0.0016	CF-D high	MP64	134	118	141	4	7	15
Alanine and proline-rich secreted protein Apa GN=apa	33 kDa	0.0026	CF-D high	APA	76	136	118	0	0	0
10 kDa chaperonin GN=groS	11 kDa	0.037	CF-D high	CH10	89	74	105	4	10	18
Catalase-peroxidase GN=katG	81 kDa	0.0064	CF-D high	KATG	74	54	50	0	0	5
FHA domain-containing protein FhaA GN=fhaA	57 kDa	0.021	CF-D high	FHAA	27	36	65	0	0	0
Immunogenic protein MPT63 GN=mpt63	17 kDa	0.0019	CF-D high	MP63	50	41	64	0	0	0
ESAT-6-like protein EsxB GN=esxB	11 kDa	0.004	CF-D high	ESXB	27	42	28	0	2	1
Diacylglycerol acyltransferase/mycolyltransferase Ag85B GN=fbpB	35 kDa	0.0095	CF-D high	A85B	30	27	51	0	0	0
Dihydrolipoyl dehydrogenase GN=lpdC	49 kDa	0.011	CF-D high	DLDH	40	31	28	0	0	4
ESX-1 secretion-associated protein EspA GN=espA	40 kDa	0.0031	CF-D high	ESPA	32	25	19	0	0	0
Ferritin BfrB GN=bfrB	20 kDa	0.0094	CF-D high	BFRB	24	22	18	1	0	2
Succinate-semialdehyde dehydrogenase [NADP(+)] 1 GN=gabD1	49 kDa	0.00013	CF-D high	GABD1	22	25	19	0	0	0
Ferredoxin GN=fdxC	12 kDa	0.00037	CF-D high	O50433	24	24	18	0	0	0
Alpha-crystallin GN=hspX	16 kDa	0.046	CF-D high	ACR	12	30	25	0	1	3
27 kDa antigen Cfp30B GN=cfp30B	27 kDa	0.00072	CF-D high	CF30	24	24	17	0	0	0
Immunogenic protein MPT70 GN=mpt70	19 kDa	0.003	CF-D high	MP70	20	25	34	0	0	0
Soluble secreted antigen MPT53 GN=mpt53	18 kDa	0.001	CF-D high	MPT53	15	21	23	0	0	0
Thioredoxin GN=trxA	13 kDa	0.0088	CF-D high	THIO	27	14	18	0	0	0

Aminopeptidase N GN=pepN	94 kDa	0.0077	CF-D high	L7N655	11	6	12	0	0	0
60 kDa chaperonin 2 GN=groEL2	57 kDa	0.0055	CF-D high	CH602	4	7	4	0	0	0
Transcription elongation factor GreA GN=greA	18 kDa	0.013	CF-D high	GREA	12	11	5	0	0	0
Peptidyl-prolyl cis-trans isomerase A GN=ppiA	19 kDa	< 0.0001	CF-D high	PPIA	13	15	15	0	0	0
Pilin GN=mtp	11 kDa	0.0039	CF-D high	PILIN	13	8	9	0	0	0
50S ribosomal protein L7/L12 GN=rplL	13 kDa	0.0006	CF-D high	RL7	12	11	15	0	0	0
Transaldolase GN=tal	41 kDa	0.013	CF-D high	TAL	6	4	9	0	0	0
ESAT-6-like protein EsxL GN=esxL	10 kDa	0.048	CF-D high	ESXL	4	5	1	0	0	0
Sulfite reductase [ferredoxin] GN=sir	62 kDa	0.0011	CF-D high	SIR	6	7	9	0	0	0
Adenylate kinase GN=adk	20 kDa	0.013	CF-D high	KAD	11	5	11	0	0	0
Dehydrogenase GN=htdY	30 kDa	0.015	CF-D high	I6YBZ8	5	4	2	0	0	0
Penicillin-binding protein GN=ponA2	85 kDa	0.0052	CF-D high	I6YGX2	8	10	5	0	0	0
Proteasome subunit beta GN=prcB	30 kDa	0.039	CF-D high	PSB	4	5	11	0	0	0
Conserved protein GN=Rv1906c PE=1 SV=3	16 kDa	0.0073	CF-D high	O07726	10	10	5	0	0	0
Prokaryotic ubiquitin-like protein Pup GN=pup	7 kDa	0.0024	CF-D high	PUP	6	6	9	0	0	0
Membrane protein GN=Rv1887 PE=1 SV=3	40 kDa	0.033	CF-D high	O07745	2	6	3	0	0	0
Protein Rv2204c GN=Rv2204c	13 kDa	0.0006	CF-D high	Y2204	5	7	5	0	0	0
4-aminobutyrate aminotransferase GN=gabT	47 kDa	0.042	CF-D high	GABT	6	3	2	0	0	0
Probable cutinase Rv1984c GN=Rv1984c	22 kDa	0.017	CF-D high	CUT1	2	2	4	0	0	0
Probable cold shock protein A GN=cspA	7 kDa	0.022	CF-D high	CSPA	5	3	2	0	0	0
Uncharacterized protein Rv2302 GN=Rv2302	9 kDa	0.0095	CF-D high	Y2302	6	4	3	0	0	0
Peptidase S1 GN=pepD	46 kDa	0.0053	CF-D high	O53896	3	4	2	0	0	0
Low molecular weight antigen MTB12 GN=mtb12	17 kDa	0.11		MTB12	98	102	80	0	11	23
ESAT-6-like protein EsxK GN=esxK PE=3 SV=1	11 kDa	0.34		ESXK	25	57	42	7	10	17
Meromycolate extension acyl carrier protein GN=acpM	13 kDa	0.44		ACPM	49	59	57	3	13	17
Glutamine synthetase 1 GN=glnA1	54 kDa	0.07		GLNA1	44	43	33	1	0	9
Malate synthase G GN=glcB	80 kDa	0.17		MASZ	38	29	48	0	0	11
Chaperone protein DnaK GN=dnaK	67 kDa	0.075		DNAK	7	25	43	0	0	0
DNA topoisomerase 1 GN=topA	102 kDa	0.16		TOP1	0	0	0	4	7	42
Phosphate-binding protein PstS 1 GN=pstS1	38 kDa	0.18		PSTS1	20	12	18	7	4	33

Single-stranded DNA-binding protein GN=ssb	17 kDa	0.24		SSB	3	9	5	5	48	2
Superoxide dismutase [Fe] GN=sodB	23 kDa	0.31		SODF	25	23	17	0	0	8
Acetyl-CoA acetyltransferase GN=fadA3	43 kDa	0.73		O53422	16	9	13	3	0	0
6 kDa early secretory antigenic target GN=esxA	10 kDa	0.7		ESXA	19	17	13	2	6	0
Low molecular weight T-cell antigen GN=TB8.4	11 kDa	0.052		O50430	19	14	13	0	3	1
Isocitrate dehydrogenase GN=icd2	83 kDa	0.12		O53611	0	7	9	0	0	0
Adenosylhomocysteinase GN=ahcY	54 kDa	0.57		SAHH	5	13	7	0	0	4
Glycogen accumulation regulator GarA GN=garA	17 kDa	0.83		GARA	8	10	9	1	0	4
Putative lipoprotein LppO GN=lppO	17 kDa	0.1		LPPO	15	16	6	1	0	0
Phosphoenolpyruvate carboxykinase [GTP] GN=pckG	67 kDa	0.19		PCKG	10	0	4	0	0	0
Siderophore exporter MmpL4 GN=mmpL4	105 kDa	0.37		MMPL4	0	2	0	0	0	0
Phosphoserine aminotransferase GN=serC	40 kDa	0.74		SERC	11	3	5	0	1	3
Bacterioferritin GN=bfr	18 kDa	0.13		BFR	5	12	1	0	0	0
Conserved protein GN=Rv1211 PE=1 SV=3	8 kDa	0.062		O05312	13	7	3	0	0	0
MPT51/MPB51 antigen GN=mpt51	31 kDa	0.37		MPT51	0	0	7	0	0	0
Transcriptional regulatory protein KdpE GN=kdpE	25 kDa	0.37		KDPE	0	0	0	4	0	0
Uncharacterized protein Rv2557 GN=Rv2557	24 kDa	0.062		Y2557	6	1	6	0	0	0
Glucose-6-phosphate isomerase GN=pgi	60 kDa	0.26		G6PI	14	0	3	0	0	0
Putative thiosulfate sulfurtransferase GN=cysA1	31 kDa	0.12		THTR	0	5	6	0	0	0
Transketolase GN=tkt	76 kDa	0.37		TKT	0	0	4	0	0	0
Uncharacterized oxidoreductase Rv2971 GN=Rv2971	30 kDa	0.061		Y2971	1	3	5	0	0	0
Putative cystathionine beta-synthase Rv1077 GN=cbs	49 kDa	0.12		Y1077	3	3	0	0	0	0
Uncharacterized protein GN=Rv0333	13 kDa	0.077		O33273	11	6	2	0	0	0
Putative integration host factor MihF GN=mihF	21 kDa	0.14		P71658	0	0	0	2	3	0
Cell surface lipoprotein MPT83 GN=mpt83	22 kDa	0.099		MP83	0	3	0	2	1	8
Conserved protein GN=Rv1498A	8 kDa	0.76		I6XY36	7	4	1	0	0	5
O-succinylhomoserine sulfhydrylase GN=metZ	43 kDa	0.11		METZ	2	6	1	0	0	0
UPF0603 protein Rv2345 GN=Rv2345	70 kDa	0.15		Y2345	0	4	2	0	0	0
ESX-1 secretion-associated protein EspB GN=espB	48 kDa	0.37		ESPB	0	0	6	0	0	0
DNA polymerase I GN=polA	98 kDa	0.37		DPO1	0	0	0	0	0	4

Ribonuclease VapC21 GN=vapC21	16 kDa	0.39		VPC21	0	1	0	3	0	0
Probable conserved membrane protein GN=Rv3587c PE=1 SV=3	27 kDa	0.22		O53572	2	0	7	0	0	0
Probable conserved lipoprotein LpqN GN=lpqN	24 kDa	0.064		O53780	5	1	3	0	0	0
50S ribosomal protein L28-2 GN=rpmB2 PE=3 SV=1	9 kDa	0.14		RL28B	0	0	0	0	5	3
Serine hydroxymethyltransferase 1 GN=glyA1 PE=1 SV=2	46 kDa	0.13		GLYA1	0	2	3	0	0	0
Acid phosphatase GN=sapM	32 kDa	0.059		O53361	1	2	4	0	0	0
Phosphate-binding protein PstS 2 GN=pstS2	38 kDa	0.21		PSTS2	0	2	6	0	0	0
30S ribosomal protein S7 GN=rpsG	18 kDa	0.37		RS7	0	0	0	0	3	0
Alanine dehydrogenase GN=ald	39 kDa	0.37		DHA	0	0	4	0	0	0
UPF0098 protein Rv2140c GN=Rv2140c	19 kDa	0.37		Y2140	0	0	3	0	0	0
Resuscitation-promoting factor RpfA GN=rpfA	40 kDa	0.37		RPFA	0	3	0	0	0	0

The statistical analysis was done using the normalized spectral abundance factor while the information presented in the table is the total spectral count.

Conclusions

The results presented here suggest that CaptoCore is suitable strategy to purify MTB-EVs. The purity of MTB-EV is evaluated based of the proportion vesicle: protein. In this study, the total concentration of protein of MTB-EV samples was undetectable by Micro-BCA which has a lower limit of detection of 0.5 µg/ml. It is important to underline that the MicroBCA assay was run diluting the sample 1:15, to avoid the use of the whole collected material only for protein quantification. The light scattering analysis showed a very high and pure (mono dispersion figure 1) concentration of MTB-EVs. In average, we found $5x10^8$ vesicles in an estimated amount of protein lower than 0.5 µg. The concentration of nanovesicles in the CF-D is the normal readout of background material which is characterized by poly-dispersion and very low concentration of vesicles (Figure1).

We found that MTB-EVs were enriched in lipoproteins which is in line with previous reports (2). In this preliminary experiment, *for the first time,* is described the presence of Cu, Zn super oxide

dismutase SodC in MTB-EV, in fact, this protein was almost exclusively present in the vesicles. In 2001, was demonstrated that SodC is used by *M. tuberculosis* to resist against superoxide toxicity and macrophages producing oxidative burst (9). It is possible that one the function of MTB-EVs during intracellular infection is the protection from oxidative burst.

References

1. Marsollier L, Brodin P, Jackson M, Korduláková J, Tafelmeyer P, Carbonnelle E, et al. Impact of *Mycobacterium ulcerans* biofilm on transmissibility to ecological niches and Buruli ulcer pathogenesis. PLoS Pathog. 2007;3(5):e62.
2. Prados-Rosales R, Baena A, Martinez LR, Luque-Garcia J, Kalscheuer R, Veeraraghavan U, et al. Mycobacteria release active membrane vesicles that modulate immune responses in a TLR2-dependent manner in mice. J Clin Invest. 2011;121(4):1471-83.
3. Prados-Rosales R, Weinrick BC, Piqué DG, Jacobs WR, Casadevall A, Rodriguez GM. Role for *Mycobacterium tuberculosis* membrane vesicles in iron acquisition. J Bacteriol. 2014;196(6):1250-6.
4. Rath P, Huang C, Wang T, Li H, Prados-Rosales R, Elemento O, et al. Genetic regulation of vesiculogenesis and immunomodulation in *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A. 2013;110(49):E4790-7.
5. Prados-Rosales R, Brown L, Casadevall A, Montalvo-Quirós S, Luque-Garcia JL. Isolation and identification of membrane vesicle-associated proteins in Gram-positive bacteria and mycobacteria. MethodsX. 2014;1:124-9.
6. Käll L, Storey JD, MacCoss MJ, Noble WS. Assigning significance to peptides identified by tandem mass spectrometry using decoy databases. J Proteome Res. 2008;7(1):29-34.
7. Nesvizhskii AI, Keller A, Kolker E, Aebersold R. A statistical model for identifying proteins by tandem mass spectrometry. Anal Chem. 2003;75(17):4646-58.
8. Zhang Y, Wen Z, Washburn MP, Florens L. Refinements to label free proteome quantitation: how to deal with peptides shared by multiple proteins. Anal Chem. 2010;82(6):2272-81.
9. Piddington DL, Fang FC, Laessig T, Cooper AM, Orme IM, Buchmeier NA. Cu,Zn superoxide dismutase of *Mycobacterium tuberculosis* contributes to survival in activated macrophages that are generating an oxidative burst. Infect Immun. 2001;69(8):4980-7.

List of Abbreviations

AUC-ROC	Area under the curve of the receiver operating characteristic curve
CF-D	Culture filtrate depleted
DDA	Data dependent acquisition
DIA	Data independent acquisition
ELISA	Enzyme-linked immunosorbent assay
FA	Formic acid
HRM	Hyper reaction monitoring
HRPR	High ph reversed-phase chromatography
LAM	Lipoarabinomannan
LASSO	Least absolute shrinkage and selection operator
LC-MS/MS	Liquid chromatography-tandem mass spectrometry
MRM	Multiple reaction monitoring
MTB-EV	*Mycobacterium tuberculosis* extracellular vesicle
MΦ	Macrophage
NSAF	Normalized spectral abundance factor
PEG	Polyethylene glycol
PIM	Phosphatidylinositol mannoside
RT	Retention time
SRM	Selected reaction monitoring
SWATH-MS	Sequential windowed acquisition of all theoretical fragment ions
TB	Tuberculosis