Identification of *Mycobacterium tuberculosis* peptides in serum extracellular vesicles from persons with latent tuberculosis infection

Carolina Mehaffy\textsuperscript{a,b}, Nicole A. Kruh-Garcia\textsuperscript{a}, Barbara Graham\textsuperscript{a}, Leah G. Jarlsberg\textsuperscript{c}, Charis E. Willyerd\textsuperscript{d}, Andrey Borisov\textsuperscript{e}, Timothy R. Sterling\textsuperscript{f}, Payam Nahid\textsuperscript{c}, Karen M. Dobos\textsuperscript{a#}

\textsuperscript{a}Department of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins, CO, USA.

\textsuperscript{b}Proteomic and Metabolomics Facility, Colorado State University, Fort Collins, Colorado, USA.

\textsuperscript{c}Division of Pulmonary and Critical Care Medicine, University of California San Francisco, Zuckerberg San Francisco General, San Francisco, California, USA.

\textsuperscript{d}Clinical Data Science Associates, LLC, Fort Collins, CO, USA.

\textsuperscript{e}Division of Tuberculosis Elimination, Centers for Disease Control and Prevention; Atlanta, Georgia, USA.

\textsuperscript{f}Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, USA.

Running title: Vesicle-associated *M. tuberculosis* peptides in LTBI

#Address correspondence to Karen M. Dobos, Karen.Dobos@colostate.edu
Abstract

Identification of biomarkers for latent *Mycobacterium tuberculosis* infection and risk of progression to tuberculosis (TB) disease are needed to better identify individuals to target for preventive therapy, predict disease risk, and potentially predict preventive therapy efficacy. Our group developed Multiple Reaction Monitoring Mass Spectrometry (MRM-MS) assays that detected *M. tuberculosis* (*Mtb*) peptides in serum extracellular vesicles from TB patients. We subsequently optimized this MRM-MS assay to selectively identify 40 *M. tuberculosis* peptides from 19 proteins that most commonly co-purify with serum vesicles of patients with TB. Here, we used this technology to evaluate if *Mtb* peptides can also be detected in individuals with latent TB infection (LTBI). Serum extracellular vesicles from 74 individuals presumed to have latent *M. tuberculosis* infection (LTBI) based on close contact with a household member with TB or a recent tuberculin skin test (TST) conversion were included in this study. Twenty-nine samples from individuals with no evidence of TB infection by TST and no known exposure to TB were used as controls to establish a threshold to account for non-specific/background signal. We identified at least one of the 40 *M. tuberculosis* peptides in 70 (95%) individuals with LTBI. A single peptide from the Glutamine synthetase (GlnA1) enzyme was identified in 61/74 (82%) individuals with LTBI, suggesting peptides from *M. tuberculosis* proteins involved in nitrogen metabolism as candidates for pathogen specific biomarkers for detection of LTBI. The detection of *M. tuberculosis* peptides in serum extracellular vesicles from persons with LTBI represents a potential advance in the diagnosis of LTBI.
Introduction

Tuberculosis (TB) is the leading cause of death due to single infectious agent (1). Public health control of the disease is difficult for several reasons, including lack of an effective vaccine, diagnostic and treatment delays, and the emergence of multi-drug resistant *Mycobacterium tuberculosis* (*Mtbc*) strains. About one quarter of the global population (1.7 billion people) is estimated to have latent TB infection (LTBI) (1, 2). Approximately 5-10% of these individuals will progress to TB disease (3). Individuals with LTBI harbor the bacteria in their lungs but do not present any clinical symptoms and are not contagious. However, they are at risk of developing TB, including the classic symptoms of the disease such as productive chronic cough, night sweats and weight loss. In contrast to LTBI, individuals with TB usually have *M. tuberculosis* in their sputum, resulting in spread of the disease. Treatment of individuals with LTBI has been effective in preventing TB, thus reducing the risk of transmission (4-6). However, some conditions, such as malnutrition, HIV infection and age < 5 years old increase the risk of progression to TB disease. In the United States, most TB cases are attributed to reactivation of LTBI (7, 8). Biomarkers of LTBI could be used to identify individuals in need of preventive treatment, and decrease TB risk.

LTBI is currently diagnosed by the tuberculin skin test (TST) or interferon gamma release assays (IGRA) (9-14). Both tests rely on an adequate host immune response and do not differentiate between LTBI and TB (10, 11, 15). Thus, novel diagnostic methods are needed to clearly identify individuals with LTBI, especially in high-risk groups in which the TST and IGRA may have low sensitivity, such as those living with HIV, cancer, diabetes or other immunosuppressed conditions. Identifying biomarkers of infection may lead to faster and more reliable diagnostic...
methods for LTBI. Several host biomarkers have been proposed to differentiate between LTBI and TB (16-21) as well as to indicate risk of disease progression (22, 23). Mycobacterial biomarkers of infection have been less studied, but evidence from our group and others indicate that they could also be exploited for diagnostics purposes (24-26).

Extracellular vesicles (EVs) include exosomes, microvesicles and apoptotic bodies. They are released from all nucleated cells and can be purified from different biofluids such as blood plasma, blood serum, and urine. Previous studies have demonstrated that *Mycobacterium* cell components, including RNA, lipids and proteins are found in EVs that are released from infected cells (27-34). In a proof-of principle study, our group demonstrated that *Mtb* proteins can be detected in serum EVs isolated from individuals with LTBI and active TB (24). While we used both shotgun proteomics as well as MRM-MS in that initial study, there were several factors that hindered our conclusions. First, individuals with LTBI were recruited when they presented to the hospital with symptoms of TB but did not have laboratory evidence of TB. Thus, it is possible that factors associated with other diseases may have been at play and could have contributed to the EVs content in these subjects. Second, the MRM-MS assays developed in our first study did not include synthetic labeled peptide standards, and thus the detection of each peptide was not validated nor quantitative. Our first study however, provided the platform needed to evaluate exosomes as potential carriers of *Mtb* biomarkers. In our subsequent study, we optimized the MRM-assays to include additional targets as well as to allow for monitoring of synthetic labeled standards. We then proceeded to validate these assays in a well-defined cohort of TB patients from geographical regions of high TB incidence as well as healthy controls.
We demonstrated that MRM-MS technology detected femtomolar concentrations of *Mtb* peptides in a complex sample such as serum EVs isolated from TB patients (25).

In the present study, we used the same technology and methodology to identify the *Mtb* peptide combinations and frequency of detection in serum EVs isolated from individuals with LTBI before treatment, establishing the groundwork for future studies to determine if these biomarkers may be used as part of prognostic tools to predict LTBI treatment success and/or progression to TB disease.

**Materials and Methods**

**Ethics statement**

This project was reviewed by the Institutional Review Board at Colorado State University and deemed human subject research exempt (IRB ID # 015-10H). The manipulation of human serum samples was approved by the Institutional Biosafety Committee at CSU (IBC PARF # 11-060B).

Samples from persons with LTBI were obtained from participants in TB Trials Consortium Study 26, which was approved by the CDC IRB (Protocol #3041) and by IRBs at each of the participating clinical sites. Healthy control samples were kindly donated by Dr. David Lewinsohn at Oregon Health and Science University (IRB Protocol e0186).

**Samples and study design**

LTBI: Serum samples (n=74) from individuals selected for LTBI treatment were obtained from the Centers for Disease Control and Prevention's TB Trials Consortium PREVENT TB Study (TBTC Study 26) (5). All participants in this sub-study were enrolled from one of five study sites within
the U.S: Site 14, Maryland (n=5); Site 20, North Texas, (n=24); Site 22, Colorado (n=14); Site 63, South Texas (n=19), Site 70, Tennessee (n=12). All participants were diagnosed with LTBI based on a positive tuberculin skin test, and were either a household contact of a TB patient or had a documented skin test conversion from negative to positive within the previous two years.

Based on these factors, participants were considered high-risk for progression to TB (5). All samples were collected prior to treatment initiation for LTBI. None of the individuals in this sub-study progressed to TB in the 33 month follow-up period. All samples were stored at -80°C until processed for exosome purification.

Healthy controls: Banked serum samples from individuals with negative TST and/or negative in-house IGRA were kindly donated by Dr. David Lewinsohn (n=29; Oregon Health and Sciences University) and were collected using the identical protocol as the TBTC Study 26 samples. All samples were stored at -80°C until processed for exosome purification, which was performed as previously described (25). See below for details.

Sample processing

Personal identifiers were removed from all serum samples, samples were assigned an unsorted code, and randomly selected prior to processing. EVs were isolated as previously described (24, 25). One hundred µL of serum samples were centrifuged at 3,000 x g for 15 minutes to remove whole cells and cellular debris. Twenty five µL of Exoquick (System Biosciences, Palo Alto, CA) were added to the cleared serum, mixed well and incubated overnight at 4°C. EVs were pelleted by centrifugation at 1.500 x g for 30 minutes. The EV’s depleted supernatant was discarded and the pellet was resuspended in 100 µL PBS (Corning, Corning, NY). EVs protein content was
quantified by micro bicinchoninic acid assay (Thermo Scientific Pierce, Rockford, IL). Thirty µg of each purified EV sample was loaded into a NuPAGE 4-20% gradient Tris-Glycine gel (Invitrogen, Carlsbad, CA) and resolved for 5 minutes at 200 V to remove residual Exoquick polymer from the protein. The region of the gel containing EVs was excised, cut into 1 mm pieces and digested with trypsin as previously described (24, 30). The extracted peptides were dried and resuspended in 5% acetonitrile, 0.1% formic acid, 94.9% water. Isotopic-labelled standards for each peptide (10 nM) (New England Peptides, Gardner, MA) listed in Table 1 and iRT (Biognosys, Schlieren, Switzerland) retention time standards were spiked into each digest sample. The final concentration of the digested EVs in each sample was roughly 1 µg/µL.

MRM-MS methods 1 and 2 were carried out as previously described (25). See Table 1 for list of targets.

**MRM-MS data analysis**

Data acquired using the Xevo TQ-S mass spectrometer (Waters Corporation, Millford, MA) were exported into Skyline Daily (64 bit) (38). Skyline files were visually inspected for retention time and peak boundary consistency. If needed, peak boundaries of the heavy labeled standard peptides were manually corrected such that each peptide was consistent among all samples. The peak areas for each peptide/transition ion, total area ratio, and library dot-product (DotP) value for each sample were exported into Excel. Base SAS v9.4 was utilized to automate analysis and processing of the Skyline output data (Supplemental Material 1). The output reports the normalized total peak average (nTPA) for each peptide with a delta DotP value equal to or less than 0.1. Higher delta DotP values were associated with significant deviation in fragmentation...
pattern from the isotopic standard and therefore the native peaks were disqualified and reported with a zero value. As isotopic standards are requisite for confident identification of the native peak, any peptide without a confident standard peak was removed from the analysis. The threshold of positivity was calculated as the 95th percentile of the nTPA for each peptide signal from healthy controls using GraphPad Prism (v.6 and v.8) (Supplementary Material 2). This threshold was then applied to the nTPA signal from subjects with LTBI to determine if the peptide was detected or not. If detected, the nTPA numeric value was used for further statistical analysis.

**Statistical analysis**

Descriptive statistics for demographics of LTBI individuals and MRM data were analyzed in R (version 3.5.1 (2018-07-02)). Signal from the heavy labeled spiked internal standard was not observed for some peptides in some samples. Since we relied on the detection of heavy peptides to confidently identify the native (light) counterparts, as well as to normalize the total peak average (TPA) of each peptide, cases in which the internal standard was not detected were labeled as “unknowns”. Number of peptides detected and number of “unknowns” were modeled separately with linear models to determine whether or not any demographic traits might have a significant association with the number of peptides detected. Traits that might be similar within a group were evaluated using classification trees with the complete dataset. Because classification trees often fit well to specific datasets, but are not readily applied to new data, these results were reported as descriptive statistics, and no claim
regarding statistical significance was made. All traits were individually tested for independence from categories (i.e. presence of peptides and presence of unknown) with a Fisher’s Exact Test.

A correlation matrix was calculated for the 31 of the 40 peptides that had at least one detectable value. Scatter plots of peptide concentrations were drawn to systematically review:

1) concentration distributions for each peptide, and 2) relationships between paired peptides.

Missing values were imputed as the concentration median, and a non-parametric Spearman method was used to calculate correlation coefficients and p values for each paired peptide comparison. Peptide pairs in the correlation matrix were ordered based on distance statistics using the hierarchical clustering method in R corrplot package (39) and the ordered correlations were tested to identify clusters. Further cluster analysis was done by calculating a distance matrix of dissimilarity equal to 1 minus the correlation coefficient. The hierarchical clustering method was applied to the dissimilarity matrix and the resulting dendrogram of peptide clusters was presented.

Statistical significance was defined as $p < 0.05$. All statistics were calculated using R version 3.5.1 (R Foundation for Statistical Computing, Vienna, Austria). No adjustment was made for multiple comparisons.

Data availability

MRM-MS data are publicly available through Panorama Web

(https://panoramaweb.org/project/Colorado%20State%20U%20Dobos%20Lab/S26%20LTBI/begin.view)

Results
Participants

The majority of the 74 participants with LTBI were white (n=44, 60%); 25 (34%) were black, 3 (4%) were Asian/Pacific Islander, one individual identified as American Indian/Alaskan Native (1.35%) and one individual as other (1.35%) (Table 2).

The study cohort was 45% female and 55% male, however sex was not equally represented in the two major races (Table 2). Forty-one (55.4%) of the study participants were selected for LTBI treatment because they had LTBI and were close contacts of a TB case, while 31 (44.6%) had a documented conversion of the TST (Table 2). The majority of the participants had a history of alcohol use (n=48, 64.9%) but only a few had a history of drug use (Table 2).

Additional demographic information can be found in Table 2. Patient demographics did not associate with MRM-MS results.

Detection of *Mtb* peptides in serum EVs from individuals with LTBI

Samples from healthy controls were used to establish a cut-off value for confident detection of *Mtb* peptides (Supplementary Material 2). The majority of individuals with LTBI had at least one, two or three peptides detected above the established cut off value (24%, 24% and 22% respectively). Nearly one quarter (24%) of the participants had 4 or more peptides detected, with the maximum number of peptides (n=12) identified in two individuals (Fig 1). Only 4 (5%) individuals with LTBI did not have any detectable peptides. No demographic trait had a statistically significant association with number of peptides in a linear model (data not shown).

Correlation analysis identified three clusters of peptides
Four peptides (IAL, NDP-A, SLA, and TVG) showed perfect correlation among one another as they were detected in only one individual with LTBI (coefficient: 1.0, p<0.001, Fig 2A). Most peptides had positive or no correlation with one another; IHV and SVF were the only peptides to have a significant inverse relationship as shown by a negative correlation coefficient (coefficient: −0.25, p=0.04; Fig 2A).

Peptides in the Spearman correlation matrix were ordered by hierarchical clustering to look for patterns. Three clusters were identified: Cluster 1: FLE-A, LAA, ITQ, VIQ, QEL NDP-A, IAL, TVG, and SLA (Fig 2A: green square); Cluster 2: TTP, HTI, and IPD (Fig 2A: purple square); Cluster 3: FLE-C, GVT, EAL, SLE, LYA, AGA, SVF, PGL, TAV and DVL (Fig 2A: blue square). These clusters were also observed when we calculated a distance matrix (where dissimilarity was equal to 1 minus the correlation coefficient) and derived a dendrogram of clustered peptides (Fig 2B).

*Mtb* peptides detected in individuals with LTBI are involved in Glutamine metabolism (GlnA and GarA), protein stability (GroES, DnaK) and fatty acid synthesis (AcpM).

Thirty one of the 40 evaluated peptides were detected in at least one subject (Table 3, Fig 3). Peptides belonging to GlnA (SVF), GroES (DVL), DnaK (TTP), AcpM (IPD) and GarA (FLL) were the peptides most commonly detected, with all of them identified in at least 12 different individuals with LTBI (Table 3). Peptide SVF from GlnA1 was the most commonly identified peptide, with detection in 61 of the 74 participants (82%). While the remaining top 5 peptides were not identified nearly as frequently as SVF, they were detected in 17 (DVL), 14 (TTP), 12 (FLL) and 12 (IPD) of the individuals with LTBI. All participants in which peptide IPD (AcpM) was identified...
also had either TTP (DnaK) or DVL (GroES) detected in their serum EVs. When DVL, TTP and FLL were combined, these peptides accounted for detection in 30 of the 74 participants (41%).

Proteins involved in nitrogen metabolism are among the top 5 proteins identified in EVs of individuals with LTBI

Two of the 5 peptides most frequently detected in individuals with LTBI belonged to proteins GlnA1 and GarA, both of which are related to nitrogen metabolism by regulating the ratio of glutamate/glutamine in the cell. The peptide identified in the largest number of individuals with LTBI was SVF (n=61, 82.4%) which belongs to the Glutamine synthetase, GlnA1 (Rv2220). We also observed GarA (Rv1827) as one of the most frequently detected proteins, with peptide FLL identified in 12 of the individuals with LTBI (16.2%).

Two heat shock chaperonins are among the top five proteins detected in LTBI serum exosomes

Along with GlnA1 and GarA, GroES and DnaK were among the five most commonly detected proteins in individuals with LTBI. Peptides DVL and RIP from GroES were identified in 17 (22.97%) and 5 (6.76%) of the LTBI participants, respectively; while peptides TTP and ITQ of DnaK were identified in 14 (18.92%) and 2 (2.70%) LTBI participants, respectively (Table 3, Fig 3). DnaK (Rv0350) is an essential chaperone (40-42), and together with GroES (among others) was identified in immunodominant fractions with T-cell activity (43), corroborating early observations of Mtb antigens that elicit T-cell responses (44).

Discussion
In this study, we demonstrate that in our tested cohort, at least one Mtb peptide was identified in the serum EV’s of 95% of persons infected with Mtb (Fig 1). This result was achieved by applying an optimized data analysis that accounts for false positive signals (present in complex matrices) and calculating a threshold of positivity for each peptide to increase the confidence of peptide identification in the LTBI samples (See Supplementary Material 2).

Of the 40 monitored peptides, a single peptide from GlnA1 was identified in 82% of the study participants. In addition to GlnA1, other proteins, including GarA, chaperones GroES and DnaK, and the acid fatty synthase, AcpM, were identified in at least 16% of the LTBI participants. All of these proteins have been shown to have some role during dormancy or key physiological processes of Mtb (45-50) and our findings indicate they may be good candidates as biomarkers of LTBI.

GlnA1 and GarA are involved in nitrogen metabolism (reviewed in (48)). GlnA1 is essential for both *in vitro* (51) and *in vivo* growth (52). It is involved in the synthesis of glutamine from L-glutamate and ammonium, serving not only as a means for nitrogen assimilation (53) but also as a building block for the synthesis of cell wall poly-α-L-glutamine (46, 54, 55). GlnA1 activity is regulated by adenylation via GlnE which senses the ratio of 2-oxoglutarate (2-OG) and glutamine, two products that are directly related to availability of nitrogen in the cell (45, 48).

Likewise, GarA, named as glycogen accumulation regulator, is involved in glutamate storage and metabolism by regulating the interconversion of glutamate and 2-OG (48, 56). GarA inhibits the activities of glutamate dehydrogenase (Gdh) and ketoglutarate decarboxylase (KGD) while stimulating the activity of the glutamine oxoglutarate aminotransferases (GltB and GltD), effectively promoting glutamate synthesis and decreasing entry of 2-OG into the Tricarboxylic
While the exact role of GlnA1 and GarA in nitrogen metabolism during *in vivo* conditions has not been explored, some studies point to the relevance of these two proteins during infection. For instance, GlnA1 has been shown to be involved in cell wall resistance and pathogenicity in *M. bovis* (46), and phosphorylation of GarA by PknG has been implicated in increasing survival of latent *Mtb* (57).

At the host level, GlnA1 induces cellular immune responses and has been proposed as a novel vaccine candidate component (47). Even though GlnA1 is not an actively secreted protein, its high extracellular stability as well as high expression (51), result in large quantities being found in the extracellular milieu of *in vitro* cultures (58). On the contrary, when *Mtb* is grown in pellicle form, GlnA1 is found to be highly abundant in the bacterial membrane when compared to shaken cultures, suggesting that expression and/or localization of GlnA1 may be dependent on growth conditions (59). This is relevant to both host exosome localization, as well as its significance in LTBI. For instance, biofilm growth (such as growth in surface pellicles) has been implicated as a possible source of aerosol transmission during TB disease (59), but it has also been proposed as a mechanism of persistence during dormancy (60). It is possible that higher expression of GlnA1 during LTBI may be a result of biofilm-related *Mtb* persistence, and thus why its peptide is readily detected in LTBI, and less so in active TB. Although *Mtb* proteins circulating in exosomes arise mostly from intracellular infection, phagocytosis of extracellular debris (such as released proteins from biofilm formation) can also be a source of exosome content (30). Additional studies are needed in order to determine if GlnA1 and/or GarA can be exploited as a biomarkers of LTBI as well as their potential role (if any) during LTBI. Our preliminary analysis comparing our results for GlnA1 (SVF peptide) to a subset of Active TB
patients from our previous study (25) indicates this peptide may be a LTBI biomarker candidate for further validation studies (Supplementary Material 3).

GroES and DnaK are both chaperones involved in the cellular response to stresses and AcpM is an acyl carrier protein involved in mycolic acid biosynthesis, a key step in the synthesis of cell wall (49, 61). The function of GroES (Rv3418c) is annotated as a suppressor of ATPase activity of GroEL1, another chaperonin that prevents misfolding of proteins and promotes proper protein folding during stress conditions. GroES is also referred to as the 10 KDa antigen because of its ability to elicit T-cell responses (62, 63). GroES appears to be highly abundant in the cytosol of Mtb grown under anaerobic conditions (64), but it is also one of the most abundant proteins under regular growth conditions (65). GroES was initially identified in human serum exosomes of individuals with TB infection, although its presence was only observed in a few individuals with pulmonary and extra pulmonary TB (24). However, GroES association with EVs in the context of Mtb infection was later confirmed in the guinea pig model in which GroES was identified in urine, serum and bronchoalveolar lavage EVs of guinea pigs infected with Mtb (37). GroES was also demonstrated to be ubiquitinated as a requirement for exosome transport (66).

GroES and AcpM are also among the most abundant proteins produced by Mtb both during exponential growth as well as during dormancy (65). It is thus possible that the success of detecting these proteins in host EVs may be due to their high expression during in vivo infection. While this hypothesis remains to be tested, there is evidence that these antigens are presented by the host during infection. GroES, DnaK and AcpM have been shown to elicit IFN-γ responses in individuals with LTBI. These proteins elicited IFN-γ in more than 80% of household
contacts but in only 58, 23 and 45% of pulmonary TB patients, respectively, indicating that these antigens may be more readily recognized by the host during LTBI (67).

The set of proteins identified in serum EVs from study participants with LTBI differ from those previously identified in patients with TB disease even though the same MRM and processing methods were used (25). The most commonly detected peptides in TB patients in the study by Mehaffy et al (2017) belonged to proteins Cfp2 and Mpt32, while peptides from GlnA, GroES, DnaK, AcpM and GarA were only detected in a small number of the study participants (25).

Indeed, the SVF peptide, found in most persons from this LTBI study, was rarely detected in TB patients (Supplementary Material 3). Peptides most frequently detected in the current study also differ from our initial exploratory study which included individuals with TB and with LTBI (24). For instance, AcpM, GroES and DnaK were among the proteins which our group originally identified as discriminatory of TB vs. LTBI or uninfected hospitalized patients (24). It is possible that these differences may be due to our improved methodologies (including the use of synthetic internal standards), different cohort, and/or demographic factors such as birthplace or geographical location, rather than TB vs. LTBI. However, it is also plausible that the physiologic state of Mtb varies across the spectrum of latency to clinical disease, and the differences in proteins are a reflection of this variability. Our traditional definitions of LTBI also likely represent a dynamic spectrum with different stages, from recent infection to reactivation leading to TB disease (68-77). It is possible then, that Mtb proteins, such as GlnA, GroES, DnaK, and GarA, may fluctuate in abundance between physiologic states, and thus could be evaluated as prognostic reagents for detection of progression from asymptomatic infection to disease.
Our study has limitations. First, the MRM methodology used in the present study was initially developed using a set of proteins optimized for the detection in serum EVs in TB patients (25).

The observations from this study, namely that Mtb peptides are abundantly detected in LTBI samples, along with further support for the potential for different peptide profiles to distinguish between LTBI and TB, was unexpected. Thus, we expect that additional proteins/peptides may be discovered specifically in individuals with LTBI. Additionally, our study was limited to a qualitative analysis (presence versus absence) of Mtb peptides. Future studies will focus on assessing if quantitative differences in peptide abundance also provide further significance when comparing TB vs. LTBI; or LTBI before and after treatment. Lastly, our analyses are based on convenience samples collected as part of a clinical trial, and as such provide exploratory evidence of Mtb peptide detection in LTBI samples, which needs to be confirmed in samples collected from well-characterized diagnostic cohorts.

Conclusion

We have shown that Mtb peptides can be detected in serum extracellular vesicles of individuals with LTBI. Furthermore, one of the peptides (SVF from GlnA1) was found in 82% of study participants. While our findings need to be validated in a larger and more geographically diverse cohort, the peptide SVF along with other peptides detected here may be good candidates for the development of Point of Care (POC) tests for the diagnosis of LTBI.

Acknowledgements

We would like to thank Bryan McNair (Colorado School of Public Health) for guidance on selecting the methodology for defining the healthy cut-off values, and Dr. David Lewinsohn and...
Lynne Swarbrick from Oregon Health and Science University for their assistance with these samples. We also would like to thank Kaitlyn Nagel for her technical contributions in sample processing.

**Disclaimer**

The findings and conclusions in this article are those of the authors and do not necessarily represent the official position of the U.S. Centers for Disease Control and Prevention (CDC).
Table 1. List of peptides targeted for identification in serum extracellular vesicles from individuals with latent tuberculosis infection (LTBI).

<table>
<thead>
<tr>
<th>Protein (Species)</th>
<th>Method 1</th>
<th>Method 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Protein</strong></td>
<td><strong>Peptide</strong></td>
<td><strong>Protein</strong></td>
</tr>
<tr>
<td>Ag85c (Rv0129c)</td>
<td>VQQGQGGPHAYLGLDGLR</td>
<td>DnaK (Rv0350)</td>
</tr>
<tr>
<td></td>
<td>NDPMVQIPR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FLEGLTLR</td>
<td></td>
</tr>
<tr>
<td>GclB (Rv1837c)</td>
<td>VVADLPQONQLRNAR.D</td>
<td>Ag85B (Rv1886c)</td>
</tr>
<tr>
<td></td>
<td>FALNAANAR</td>
<td>Cfp10 (Rv3874)</td>
</tr>
<tr>
<td></td>
<td>NYTAPGGQFTLPG</td>
<td></td>
</tr>
<tr>
<td>Mpt32 (Rv1860)</td>
<td>TGDPPFPQPPPVANDT</td>
<td>EsxA (Rv3875)</td>
</tr>
<tr>
<td></td>
<td>LYSAEATDSK</td>
<td></td>
</tr>
<tr>
<td>Mpt64 (Rv1980c)</td>
<td>SLENYIAQTR</td>
<td>GlnA1 (Rv2220)</td>
</tr>
<tr>
<td></td>
<td>FLSAATSTTPR</td>
<td></td>
</tr>
<tr>
<td>HspX (Rv2031c)</td>
<td>AEIPEGVDPDK</td>
<td>MrsA (Rv3441c)</td>
</tr>
<tr>
<td></td>
<td>TVSLPVGADDIK</td>
<td></td>
</tr>
<tr>
<td>Cfp2 (Rv2376c)</td>
<td>GSLVEGGIGITEAR</td>
<td>PpiA (Rv0009)</td>
</tr>
<tr>
<td></td>
<td>SLADPNVSFAK</td>
<td></td>
</tr>
<tr>
<td>SahH (Rv3248c)</td>
<td>GVTEETTGVLR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IHVEALGGHLTK</td>
<td>AcpM</td>
</tr>
</tbody>
</table>
*The first three letters of each peptide (bold and underlined) are used for reference throughout the manuscript.

### Table 2. Summary statistics of attributes measured in individuals with LTBI

<table>
<thead>
<tr>
<th></th>
<th>Female, n=33</th>
<th>Male, n=41</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age: median (IQR)</td>
<td>36 (26-46)</td>
<td>41 (30-49)</td>
</tr>
<tr>
<td>BMI: median (IQR)</td>
<td>30 (25-35)</td>
<td>28 (24-33)</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>24 (73%)</td>
<td>20 (49%)</td>
</tr>
<tr>
<td>Black</td>
<td>7 (21%)</td>
<td>18 (44%)</td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>2 (6%)</td>
<td>1 (2%)</td>
</tr>
<tr>
<td>American Indian/ Alaska Native</td>
<td>0</td>
<td>1 (2%)</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
<td>1 (2%)</td>
</tr>
</tbody>
</table>

History of alcohol use/abuse 21 (64%) 27 (66%)
Hepatitis comorbidity present 1 (3%) 4 (10%)
Close contact with TB patient 17 (52%) 26 (63%)
TST conversion¥ 16 (49%) 15 (37%)
History of drug use 1 (3%) 4 (10%)
Unknown HIV status 6 (18%) 10 (24%)
Homeless 1 (3%) 7 (17%)
WHO birthplace outside of the Americas 2 (6%) 3 (7%)
History of smoking in past 5 years (not including current smokers) 2 (6%) 3 (7%)
Current smoker 10 (30%) 16 (39%)

"¥TST: Tuberculin Skin Test"

Table 3. Peptides detected in individuals with LTBI

<table>
<thead>
<tr>
<th>Protein</th>
<th>Peptide</th>
<th>Detected (Number of participants)</th>
<th>Detected (Number of participants)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GlnA</td>
<td>SVF</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>GroES</td>
<td>DVL</td>
<td>17</td>
<td>SahH GVT 2</td>
</tr>
<tr>
<td></td>
<td>RIP</td>
<td>5</td>
<td>IHV 2</td>
</tr>
<tr>
<td>DnaK</td>
<td>TTP</td>
<td>14</td>
<td>HspX* AEL 3</td>
</tr>
</tbody>
</table>

21
<table>
<thead>
<tr>
<th>Protein</th>
<th>Peptide</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITQ</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>BfrB</td>
<td>EAL</td>
<td>3</td>
</tr>
<tr>
<td>AcpM*</td>
<td>IPD</td>
<td>12</td>
</tr>
<tr>
<td>AGA</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>TVG</td>
<td>1</td>
<td>MrsA</td>
</tr>
<tr>
<td>TAV</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>GarA</td>
<td>FLL</td>
<td>12</td>
</tr>
<tr>
<td>YVL</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>LVF</td>
<td>1</td>
<td>Mpt64</td>
</tr>
<tr>
<td>SLE</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Ag85A</td>
<td>NDP-A</td>
<td>1</td>
</tr>
<tr>
<td>Mpt32</td>
<td>TTG</td>
<td>4</td>
</tr>
<tr>
<td>FLE-A</td>
<td>3</td>
<td>LYA</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PpiA</td>
<td>IAL</td>
<td>1</td>
</tr>
<tr>
<td>GlcB*</td>
<td>NYT</td>
<td>1</td>
</tr>
<tr>
<td>VIQ</td>
<td>4</td>
<td>Cfp2*</td>
</tr>
<tr>
<td>SLA</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>HTI</td>
<td>6</td>
<td>Esat6*</td>
</tr>
<tr>
<td>LAA</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Ag85C*</td>
<td>FLE-C</td>
<td>1</td>
</tr>
<tr>
<td>Ag85B</td>
<td>PGL</td>
<td>9</td>
</tr>
<tr>
<td>NDP-C</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

*Protein with additional monitored peptides. Additional peptides resulted in either “Unknowns” or negative. The five peptides highlighted in the text are: SVF, FLL, IPD, TTP and DVL.*
REFERENCES


589
FIGURE LEGENDS

Figure 1. Number of peptides detected in individuals with LTBI. The y axis displays the percentage of individuals with LTBI with 0-12 detected peptides (x axis) in individuals (n).

Figure 2. Results of cluster analysis of peptide concentrations. The three clusters identified are cluster 1 (green), cluster 2 (purple), and cluster 3 (blue). Panel A) Correlogram shows Spearman correlation coefficients (R) where p < 0.05: Upper triangle symbols are blue to indicate positive and orange to indicate negative R, and area of the circle corresponds with R absolute value; lower triangle shows corresponding numeric R. Panel B) Dendrogram shows result of hierarchical clustering of a distance matrix calculated as dissimilarity = 1 – R.

Figure 3. Distribution of absent, unknown and detected peptides in the 74 study participants with LTBI. The y axis represents the frequency in which each individual peptide was detected. Orange represent absent (undetected) peptides, blue represents present (detected peptides) and red represents “unknowns” or peptides in which a measurement was not possible due to matrix effects.