The identification of tuberculosis biomarkers in human urine samples

Brandy L. Young[1], Zandile Mlamla[2], Putuma P. Gqamana[3], Salome Smit[4], Teri Roberts[5], Jonathan Peter[6], Grant Theron[6], Ureshnie Govender[6], Keertan Dheda[1,6,7] and Jonathan Blackburn[1,2,7]

Affiliations: [1]Institute of Infectious Disease and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Cape Town, [2]Division of Medical Biochemistry, Dept of Clinical Laboratory Sciences, University of Cape Town, Cape Town, [3]University of the Western Cape, Cape Town, [4]Proteomics Laboratory, Central Analytical Facility, University of Stellenbosh, Cape Town, and [6]Lung Infection and Immunity Unit, Division of Pulmonology and UCT Lung Institute, Dept of Medicine, University of Cape Town, South Africa. [5]Médecins Sans Frontières, Geneva, Switzerland. [7]Both authors contributed equally.

Correspondence: Jonathan Blackburn, Institute of Infectious Disease and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Anzio Road, Observatory, Cape Town 7925, South Africa. E-mail: jonathan.blackburn@uct.ac.za

ABSTRACT   We aimed to determine whether shotgun proteomic approaches could be used to identify tuberculosis (TB)-specific biomarkers in the urine of well-characterised patients with active TB *versus* no TB.

Patients with suspected TB (n=63) were classified as: definite TB (*Mycobacterium tuberculosis* positive culture, n=21); presumed latent-TB infection (LTBI) (*M. tuberculosis* negative culture, no radiological features of active TB, a positive QuantiFERON-TB Gold In-Tube (QFT-IT) test and a positive T-SPOT.TB test, n=24); and presumed non-TB/non-LTBI (*M. tuberculosis* negative culture, no radiological features of active TB, a negative QFT-IT test and a negative T-SPOT.TB test, n=18). Urine proteins, in the range of 3–50 kDa, were collected, separated by a one-dimensional SDS-PAGE gel and digested using trypsin, after which high-performance liquid chromatography-tandem mass spectrometry was used to identify the urinary proteome.

10 mycobacterial proteins were observed exclusively in the urine of definite TB patients, while six mycobacterial proteins were found exclusively in the urine of presumed LTBI patients. In addition, a gene ontology enrichment analysis identified a panel of 20 human proteins that were significant discriminators (p<0.05) for TB disease compared to no TB disease. Furthermore, seven common human proteins were differentially over- or under-expressed in the TB *versus* the non-TB group.

These biomarkers hold promise for the development of new point-of-care diagnostics for TB.

@ERSpublications

The application of proteomics for the identification of novel urinary tuberculosis biomarkers http://ow.ly/tN0Rm

Introduction

Tuberculosis (TB) is a major global health problem and to improve case detection the World Health Organization recently endorsed the Xpert MTB/RIF (Cepheid, Sunnyvale, CA, USA) assay as a frontline diagnostic test in TB endemic settings [1]. Although Xpert MTB/RIF is a significant step forward for improved diagnosis, it has several drawbacks. Xpert MTB/RIF requires patients to produce a representative sputum sample, which excludes many HIV positive individuals and children, it has a sensitivity of ~70% in smear-negative TB patients, it cannot directly diagnose extrapulmonary TB, and is relatively expensive [2, 3]. Other diagnostic tests such as smear microscopy have a sensitivity of only ~50%. The interferon-γ release assay (IGRA) is not useful for the diagnosis of active TB [4], and the Determine TB LAM antigen detection test (Alere Inc., Waltham, MA, USA) is only useful in HIV-infected persons [5–9]. Hence, new TB diagnostic tests, designed for use at the point of care in resource-poor settings, represent a major unmet global health priority. However, the lack of suitable biomarkers that distinguish active TB disease from non-TB constitutes a major impasse in the development of new TB diagnostic tests.

One way to address this is to use proteomics, a comprehensive protein identification methodology, to discover novel diagnostic biomarkers for TB [10]. Shotgun proteomics, using liquid chromatography (LC) coupled with tandem mass spectrometry (MS/MS), has gained much traction in recent years. Some of the initial TB proteomic studies involved analysing *Mycobacterium tuberculosis* (MTB) proteins from different infectious strains [11]. This provided insights into the proteome of the bacteria and served as a useful map for protein annotation. AGRANOFF *et al.* [12] used proteomic fingerprinting of serum to look for diagnostic biomarkers that could differentiate between active TB, latent tuberculosis infection (LTBI) and no TB in HIV positive and negative patients. In another study, KASHINO *et al.* [13] used one-dimensional LC combined with MS/MS to identify four *M. tuberculosis*-derived proteins in the urine of patients with pulmonary TB. A more recent follow-up urinary proteomics study found that a putative molybdopterin biosynthesis protein (Rv1681) was detected in 11 out of 21 TB patients [14].

Here, we used state-of-the-art proteomic approaches to discover urinary diagnostic biomarkers associated with pulmonary TB [15, 16]. We report 10 *M. tuberculosis* and 27 human proteins that were selectively identified in the urine of patients with active pulmonary TB disease, as well as six *M. tuberculosis* proteins that were selectively identified in the urine of patients with presumed LTBI.

Methods

Urine collection and clinical case definition

In this study, urine samples were collected from patients attending South African clinics who were suspected of having active pulmonary TB (fig. 1). Informed consent was obtained from all participants and the study was approved by the Faculty of Health Sciences Human Research Ethics Committee (University of Cape Town, Cape Town, South Africa) (approval number HREC421/2006). Patients with HIV co-infection and other major comorbidities (*e.g.* diabetes and chronic organ failure) were excluded from this study. Definite TB patients were classified based on having clinical presentation compatible with TB infection, a positive *M. tuberculosis* culture and a positive smear. Patients with presumed LTBI were defined as having: a negative *M. tuberculosis* culture, a negative smear, two positive IGRA test results, the QuantiFERON-TB Gold In-Tube (Qiagen Inc., Valencia, CA, USA) and the T-SPOT.TB (Oxford Immunotec Global PLC, Abingdon, UK), and chest radiography that was not compatible with active TB. Finally, the presumed no TB (non-TB/non-LTBI) group had: negative *M. tuberculosis* cultures, negative smears, two negative IGRAs, and chest radiography that was not compatible with active TB. The patient demographic information is shown in table 1.

Sample preparation

A total of 63 urine samples were collected. The urine samples were grouped according to the clinical cohort (TB, presumed LTBI and presumed non-TB/non-LTBI). Protein readings were obtained on the individual urine samples using a Bradford assay. Equal amounts of protein were obtained from each patient and then pooled based on the patient group (pooled n=6 patients, ~10 mL of urine in each pooled sample). For the definite TB and non-TB/non-LTBI groups, three biological replicate pools were prepared (*i.e.* six patient urine samples per group with three biological replicates; n=18 patients per group). For the LTBI group four biological replicate pools were prepared. The pooled urine sample protein concentrations were then compared and normalised across all groups and biological replicates. In addition, a set of three individual definite TB patient urine samples were similarly prepared and normalised.

All pooled or individual samples were filtered through a 50 kDa molecular weight cut-off (MWCO) filter to remove abundant, higher molecular weight human proteins that would otherwise prevent identification of lower abundance proteins. The flow-through was then concentrated on a 15 mL 3 kDa MWCO filter
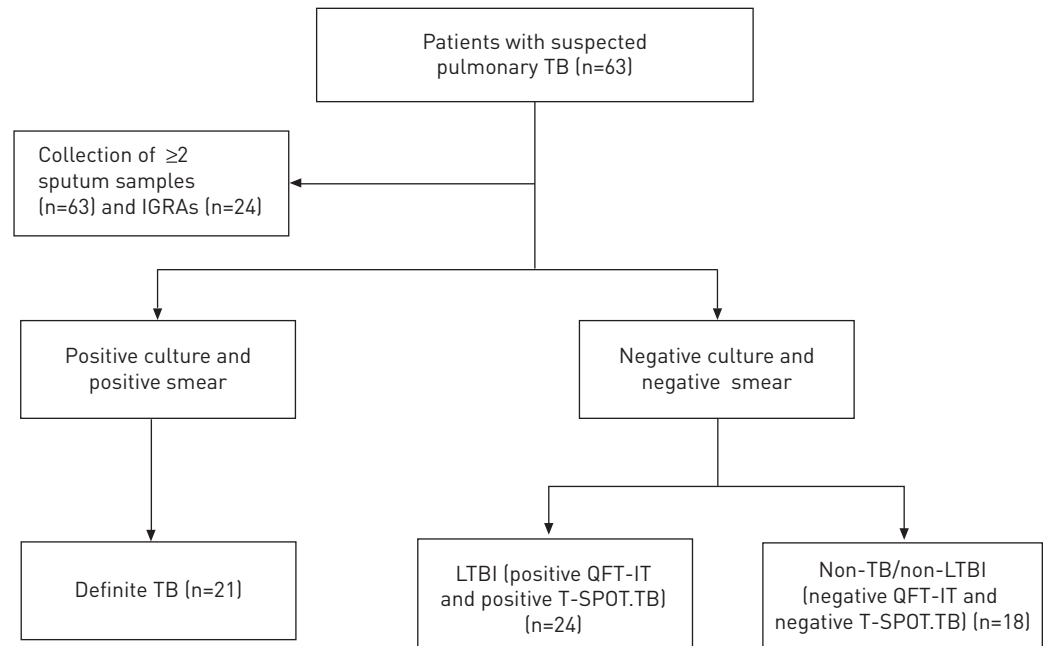
FIGURE 1 Flow chart detailing the patient's diagnostic outcomes stratified by final tuberculosis (TB) diagnosis category. IGRA: interferon-γ release assay; LTBI: latent TB infection; QFT-IT: QuantiFERON-TB Gold In-Tube (Qiagen Inc., Valencia, CA, USA).

followed by a 0.5 mL 3 kDa MWCO filter to reduce the volume. Proteins in the 3 kDa to 50 kDa range were collected, separated on a 10% SDS-PAGE gel and digested using trypsin (fig. 2).

High-performance liquid chromatography
The chromatographic separation was conducted using a Proxeon Easy n-LC II (Proxeon Biosystems, Odense, Denmark). A gradient chromatography separation was performed using a Proxeon Easy C18 trap column (2 cm, ID 100 μm, 5 μm, C18) and analytical column (10 cm, ID 75 μm, 3 μm, C18). Chromatography was performed at room temperature with a binary mobile phase system (A: water with 0.1% formic acid; and B: acetonitrile with 0.1% formic acid) at a flow rate of 300 nL·min^{-1}. Tryptic peptides were eluted from the column with a gradient of: 5% B at 0.00 min; 5% B at 5.00 min; and 80% B at 120 min.

Mass spectrometry
The shotgun proteomic analysis was performed using an LTQ-Orbitrap Velos Pro mass spectrometer (Thermo Fisher, San Jose, CA, USA) with sample introduction *via* flow injection analysis from the Proxeon high-performance liquid chromatography system. Positive ion electrospray ionisation was used with the following conditions: an ion spray voltage of 1500 V and a source temperature of 200°C. For the collision induced dissociation analysis, helium was used as the collision gas with a collision gas pressure of 1.2 mTorr and variable collision energy.

Results
In this analysis a multi-algorithm database search strategy was used. The search summary can be found in table 2, which shows the average number of peptide spectral matches, peptides and protein groups identified across the three patient groups (presumed non-TB/non-LTBI, presumed LTBI and TB). The average number of peptide spectral matches was higher for X! Tandem (X! TANDEM Spectrum Modeler, www.thegpm.org/tandem/) than for Omssa (Open Mass Spectrometry Search Algorithm, ftp://ftp.ncbi.nih. gov/pub/lewisg/omssa/CURRENT/) and accordingly X! Tandem identified 1332 peptides and 524 proteins compared to 830 peptides and 412 proteins identified by Omssa (see online supplementary material). The final data used in this analysis was the combined nonredundant subsumed protein list from the two algorithms for each patient group [17]. Figure 3a shows the Venn diagrams that highlight the overlapping and unique human and *M. tuberculosis* proteins identified across the three patient groups. An average of 560 proteins was found across the three cohorts. The data from each patient group showed that 319 (57%) out of 560 of the proteins identified were seen in all three groups. With this dataset the following questions were

TABLE 1 Patient demographic information

	Group 1 definite TB	Group 2 non-TB/LTBI	Group 3 non-TB/non-LTBI
Subjects	21	24	18
Age years	43 (42)	43 (41)	35 (31)
Male	15	20	14
Previous TB	10	8	3
Current Smoker	13	17	11
Mean weight kg	62	66	63
Ethnicity			
Black	15	12	12
Mixed ancestry	6	12	3
White	0	0	3
Diagnostic test			
Culture			
Positive	21	0	0
Negative	0	24	18
Smear			
Positive	17	0	0
Negative	4	24	18
T-SPOT.TB			
Positive	13	24	0
Negative	2	0	18
No result	6		
QFT-IT			
Positive	13	24	0
Negative	1	0	18
No result	7		

Data are presented as n or mean (median), unless otherwise stated. TB: tuberculosis; LTBI: latent TB infection; QFT-IT: QuantiFERON-TB Gold In-Tube.
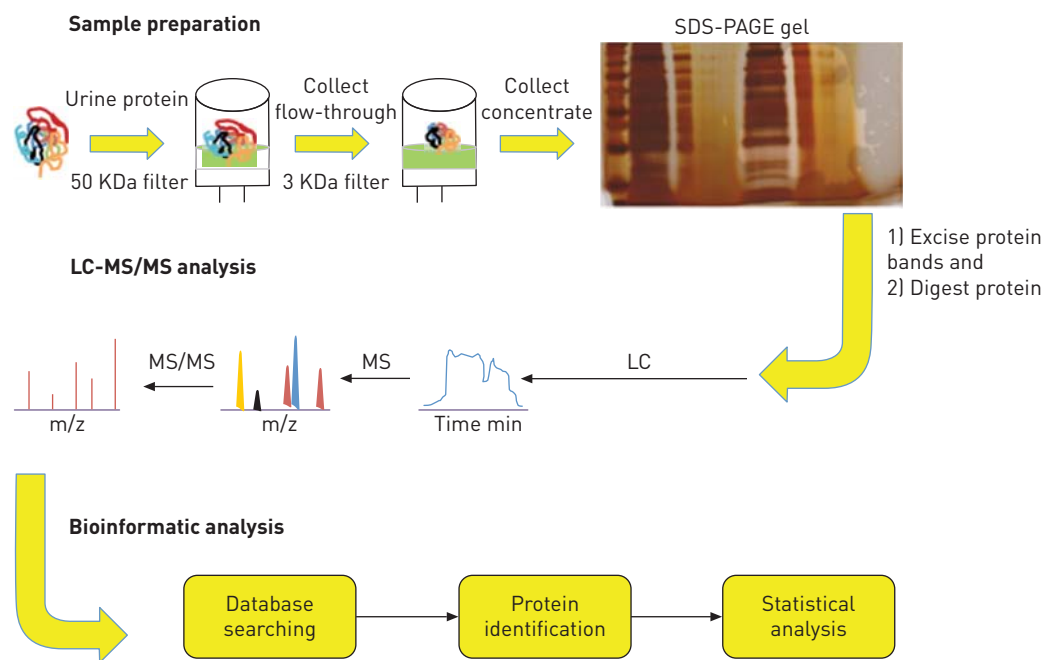

FIGURE 2 Schematic workflow detailing the proteomic sample preparation process, liquid chromatography (LC) tandem mass spectrometry (MS/MS) analysis and bioinformatics analysis.

TABLE 2 The database search parameters used and search summary for the bioinformatics algorithms

Parameter	Algorithm	
	X! Tandem	Omssa
Precursor charge state	+2 to +4	+2 to +4
Enzyme	Trypsin	Trypsin
Missed cleavages n	2	2
Precursor mass error ppm	25	25
Peptide mass error tolerance Da	0.5	0.5
Fixed modifications	Carbamidomethylation on cysteine	Carbamidomethylation on cysteine
Variable modifications	Asparagine/glutamine deamidation	Asparagine/glutamine deamidation
	Methionine oxidation	Methionine oxidation
FDR[#] %	1	1
Average PSMs[¶] n	7877	6642
Average peptides[¶] n	1332	830
Average proteins[¶] n	524	412

FDR: false discovery rate; PSM: peptide spectral matches. [#]: for each algorithm the PSM were filtered based on a score threshold of 1% FDR, which was established using a decoy database search; [¶]: the average parameters shown are average combined lists taken from the merged X! Tandem and merged Omssa data.

addressed. 1) Can exogenous *M. tuberculosis*-specific proteins be identified in urine from patients infected with pulmonary TB? 2) Can a panel of human proteins be identified that are associated with TB infection? 3) Can significant changes in common human urinary proteins be observed that are associated with TB disease?

Exogenous M. tuberculosis *proteins can be identified in patient urine samples*
A total of 26 *M. tuberculosis* proteins were identified across the various patient urine samples (fig. 3b). These *M. tuberculosis* proteins do not appear to have any functions in common and were primarily found to be membrane-associated or localised to the extracellular space (table 3). Surprisingly, we identified *M. tuberculosis* proteins in all the suspected TB patients. Given this result, a BLAST (Basic Local Alignment Search Tool, http://blast.ncbi.nlm.nih.gov/Blast.cgi) analysis was carried out using DBToolkit [18] to map the identified *M. tuberculosis* peptides to human or bacterial proteins in the UniProt database (http://www.uniprot.org/). This analysis revealed that six out of 26 of the identified *M. tuberculosis* peptides could be mapped to proteins from non-TB mycobacteria and/or other actinobacteria, in addition to proteins from
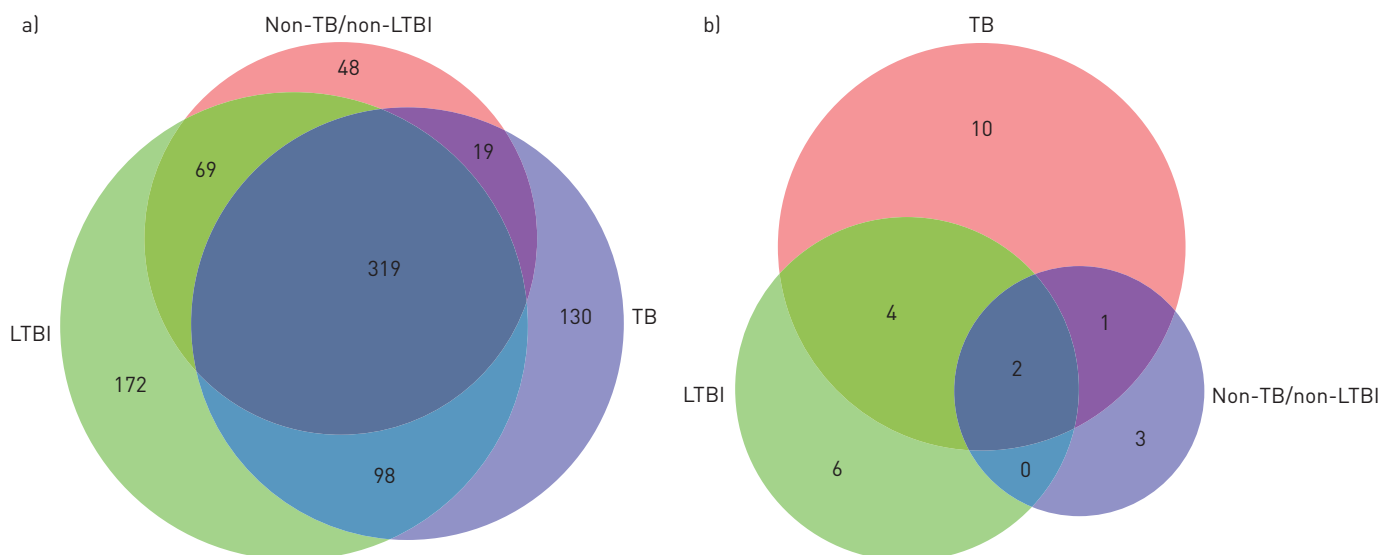

FIGURE 3 a) The total urinary protein repertoire and b) the unique and overlapping *Mycobacterium tuberculosis* proteins observed in suspected tuberculosis (TB) patients classified as active TB, presumed latent TB infection (LTBI) and presumed non-TB/non-LTBI.

organisms within the *M. tuberculosis* complex (table 3). Those peptides that mapped to other species outside of *M. tuberculosis* or the *M. tuberculosis* complex are identified as *Mycobacterium*-related peptides in table 3. For the active-TB patient group (including both the triplicate biological replicate pooled samples and the three individual samples), eight previously unreported *M. tuberculosis*-specific proteins and two *Mycobacterium*-related proteins were identified in the urine samples. Interestingly, several isoforms of the PE-PGRS family, which are glycan rich cell wall proteins, were identified in the TB and/or the presumed LTBI groups (table 3).

Functional classification of proteins found in patient urine samples
To classify the human urinary proteins based on their cellular compartmentalisation (CC), molecular function (MF) and biological processes (BP) a gene ontology (GO) enrichment analysis was performed (see online supplementary information for details) [19, 20]. Figure 4 shows the enriched GO terms obtained when the proteome of the presumed non-TB/non-LTBI group was set as the background gene product, which includes 10 GO-BP terms plus three GO-CC terms for the TB group and two GO-BP terms plus two GO-CC terms for the LTBI group. Interestingly, no GO-MF terms were significantly enriched for either group. For the TB and LTBI groups, the most significant GO-BP terms were "Response to Stimulus"

TABLE 3 *Mycobacterium tuberculosis* specific proteins and *Mycobacterium*-related proteins identified in human urine samples from active TB, presumed LTBI and presumed non-TB/non-LTBI patients

Cohort	Accession number	Protein name	*M. tuberculosis* specific[#]	*Mycobacterium*-related[¶]
TB only	Rv1475c	Aconitate hydratase		×
	Rv1977[+]	Conserved protein		×
	Rv0014c	Serine/threonine protein kinase	×	
	Rv2748c	DNA translocase FtsK	×	
	Rv1664	Polyketide synthase	×	
	Rv1161	Nitrate reductase α-subunit	×	
	Rv2280	Uncharacterised FAD-linked oxidoreductase	×	
	Rv2694c	Conserved hypothetical protein	×	
	Rv2490c[§]	PE-PGRS family protein (1661 amino acids)	×	
	Rv0578c	PE-PGRS family protein (1307 amino acids)	×	
TB and LTBI	Rv1450c	PE-PGRS family protein	×	
	Rv1522c	Putative membrane protein mmpL12	×	
	Rv2737c	RecA		×
	Rv2981c	D-alanine–D-alanine ligase		×
TB and non-TB/non-LTBI	Rv0765c	Probable oxidoreductase	×	
LTBI only	Rv2126c	PE-PGRS family protein	×	
	Rv1759c	WAG22 antigen	×	
	Rv1464[ƒ]	Probable cysteine desulfurase		×
	Rv3202c[##]	Possible ATP-dependent DNA helicase	×	
	Rv0668	DNA-directed RNA polymerase subunit	×	
	Rv3345c	PE-PGRS family protein	×	
LTBI and non-TB/non-LTBI	None			
Non-TB/non-LTBI only	Rv3775	Probable lipase LipE	×	
	Rv3736	Transcriptional regulator, AraC family	×	
	Rv2455c	2-oxoglutarate oxidoreductase subunit KorA	×	
TB, LTBI and non-TB/non-LTBI	Rv3885c	Possible membrane protein		×
	Rv1235	Probable sugar-binding lipoprotein LpqY	×	

Online supplementary tables S1 and S2 contain detailed peptide statistics and homology matching information. TB: tuberculosis; LTBI: latent TB infection; FAD: flavin adenine dinucleotide. [#]: proteins were classified as *M. tuberculosis* specific if the identifying peptides only mapped to proteins found in the *M. tuberculosis* complex; [¶]: proteins were classified as *Mycobacterium*-related if the identifying peptides mapped to both *M. tuberculosis* specific proteins as well as to other non-infectious mycobacteria and/or other proteins found in a combined UniProt human and bacteria database; [+]: this protein was identified by a peptide that mapped to both *M. tuberculosis* and *Mycobacterium marinum*; [§]: this protein was identified based on three peptides, all of which were *M. tuberculosis* specific; [ƒ]: this protein was identified by a peptide that that mapped to both *M. tuberculosis* and *Mycobacterium parascrofulaceum*; [##]: this protein was identified by two different peptides, one of which was *M. tuberculosis* specific and the other was *Mycobacterium* related.

FIGURE 4 Gene ontology (GO) enrichment analysis of the proteins identified from the suspected tuberculosis (TB) patients with the presumed non-TB/non-latent TB infection (LTBI) patient group set as the background gene product. The following terms were obtained *via* David Pathway Analysis with a cut-off of p ⩽ 0.05. GO-BP: terms that represent biological operations or sets of molecular events; GO-CC: terms that represent part of the cell or extracellular space.
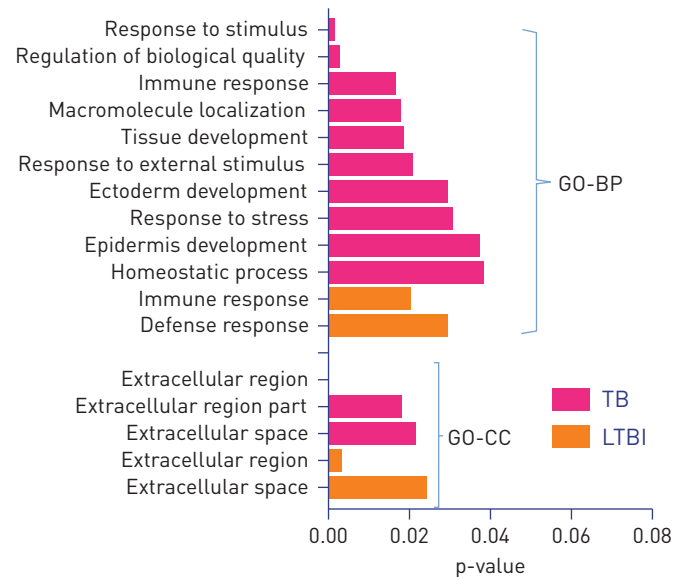
(GO:0050896) and "Immune Response" (GO:0006955), respectively. While the most significant GO-CC term for both groups was "Extracellular Region" (GO:0005576).

In total, these 17 enriched GO terms encompassed a total of 1230 redundant proteins. To reduce this list of potential biomarkers towards a smaller panel of proteins that could potentially predict TB disease, we did the following. First, the proteins that comprised the most significant GO-BP term for the TB group, "response to stimulus" (GO:0050896), were clustered. Secondly, the proteins that comprised the two GO-BP terms for the LTBI group, "immune response" (GO:0006955) and "defense response" (GO:0006952), were clustered. Finally, all the GO-CC terms and the corresponding proteins from both the TB and LTBI groups were clustered together since there was significant overlap in the GO-CC terms and proteins here. Figure 5 shows a Venn diagram of these clustered proteins and reveals that 20 of the proteins in the "response to stimulus" (GO:0050896) group were unique to this term in the TB group (table 4). The complete list of GO terms and proteins can be found in online supplementary table S3.

Immune-related proteins are differentially expressed in urine samples of patients with active-TB
The normalised spectral abundance factor is a label-free spectral counting method that enables the relative quantitation of protein abundances across samples. The spectral abundance is enumerated by summing the number of identified peptide spectra for each protein, normalising this sum to the protein length and then further normalising to the sum of all protein abundances in the given sample. Here, we used normalised spectral abundance factor values to determine the level of differential protein expression in the common human urinary proteins observed in the TB-suspected patients' urine by comparing the active TB and, separately, the LTBI patients to the presumed non-TB/non-LTBI group. An ANOVA test showed that when using either of the two bioinformatics algorithms, the TB group had upregulated levels of immunoglobulin kappa chain C (IGKC; accession number: P01834), retinol binding protein 4 (RBP4; accession number: P02753), α-1-acid glycoprotein 1 (ORM1; accession number: P02763) and immunoglobulin lambda-2 chain C (IGLC2; P0CG05, Omssa database only) that were statically significant (p ⩽ 0.05) compared to the non-TB/non-LTBI group (fig. 6). A statistically significant difference (p ⩽ 0.05) was also observed between the LTBI group compared to the non-TB/non-LTBI group for IGKC and ORM1 (X! Tandem database only). Although the difference in ORM1 abundance between the LTBI group and the non-TB/non-LTBI group found when using the Omssa database failed to reach statistical significance, we note that the trend is similar to that observed when using the X! Tandem database.

By contrast, in the X! Tandem dataset, the proteins prostaglandin-H2 D-isomerase (PTGDS; accession number: P41222), secreted and transmembrane protein 1 (SECTM1; accession number: Q8WVN6) and α-1-microglobulin/bikunin precursor (AMBP; accession number: P02760) were significantly (p ⩽ 0.05) downregulated in the TB group compared to the non-TB/non-LTBI group (fig. 6a). A similar pattern of relative expression values was observed for PTGDS in the Omssa dataset (fig. 6b).
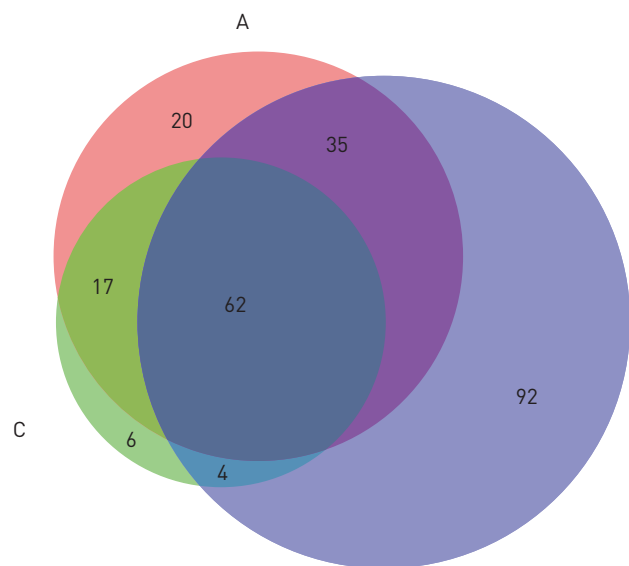
FIGURE 5 Gene ontology (GO) enrichment analysis where A: is the clustered GO-biological processes (GO-BP) proteins from the GO term "response to stimulus" (GO:0050896) enriched in the tuberculosis (TB) group; B: is the clustered GO-BP proteins from the GO terms "immune response" (GO:0006955) and "defense response" (GO:0006952) enriched in the presumed latent TB infection (LTBI) group; and C: is the clustered GO-cellular compartmentalisation (GO-CC) proteins from the terms in the TB and presumed LTBI group.

Discussion

Our initial urinary proteomics efforts showed that the abundant, high molecular weight human proteins, such as uromodulin and albumin (reported to be 60% and 20% of the total urinary proteome, respectively [21]), masked the signal of the lower abundant proteins; resulting in ∼50 total protein identifications (data not shown). However, noting that the average molecular weight of *M. tuberculosis* proteins is only ∼20 kDa (online supplementary fig. S1), we found that simple use of MWCO filters to remove the abundant, higher molecular weight human proteins enabled us to identify and quantify on average 560 proteins across the three patient groups. This figure compares well to the benchmark of ∼600 protein identifications set by NAGARAJ *et al.* [22] for label-free, quantitative urinary proteomic analyses, although it is some way short of a nonquantitative proteomic analysis that reported ∼1500 proteins in a human urine sample [23].

Using this experimental approach we identified 26 *M. tuberculosis* proteins across the various patient urine samples, of which 10 were unique to the TB group, six were unique to the LTBI group and three were unique to the non-TB/non-LTBI group (table 3). Given the continuing uncertainty about the metabolic status of *M. tuberculosis* bacilli within infected macrophages during a latent infection [24], the observation of unique *M. tuberculosis* peptides in urine from LTBI patients is not unreasonable. However, we note that the number of such peptides uniquely identified in LTBI patients urine is lower than that found uniquely in active-TB patients' urine, in accord with expectation. In addition, different isoforms of the PE-PGRS protein family were uniquely identified in the TB and LTBI patient groups, with one isoform shared between the two groups. Importantly, since several peptides in this protein family have been identified in both active-TB and LTBI patient urine samples in the present work, they may represent further interesting biomarker candidates, in addition to the 10 *M. tuberculosis* proteins that are unique to the active TB group.

The identification of *M. tuberculosis* proteins in the presumed non-TB/non-LTBI group was unexpected, but it is important to note that there is currently no reliable test for predicting LTBI. For example, RANGAKA *et al.* [4] reported that IGRAs are not good at predicting active TB disease let alone LTBI. It is, therefore, possible that some patients in the non-TB/non-LTBI group did in fact have LTBI; further work will be needed to evaluate this possibility. Nonetheless, the *M. tuberculosis* proteins identified selectively in the active TB and presumed LTBI patient urine samples represent important observations and warrant further investigation as candidate urinary biomarkers for active TB infection. For example, it is striking that we observed the same peptide derived from the *M. tuberculosis* protein Rv2981c in all four LTBI pools, as well as in two out of the six TB samples, but in none of the non-TB/non-LTBI samples, suggesting that this peptide might represent a more direct measure for LTBI than IGRAs (online supplementary table S1). In all cases, further validation studies are clearly needed to determine the frequency with which the 26 *M. tuberculosis* proteins identified here are also observed in individual urine samples from a larger cohort of patients with suspected TB. However, it is encouraging that our data already hints strongly at high frequencies for at least three of these proteins: Rv0765c, Rv1235 and Rv2981c.

A GO enrichment analysis [19, 20] aims to cluster proteins based on three defined ontology terms: GO-BP, GO-CC and GO-MF. Here, we observed 17 significantly enriched GO terms that together contained 1230

TABLE 4 The 20 unique human proteins identified from the enriched gene ontology (GO)-biological processes term ''response to stimulus'' (GO:0050896)

Accession number	Protein Name
P08758	Annexin A5
P10523	S-arrestin
P12830	Cadherin-1
P05937	Calbindin
Q9UBG3	Cornulin
O60494	Cubilin
P15924	Desmoplakin
P54764	Ephrin type-A receptor 4
P05413	Fatty acid-binding protein
P07359	Platelet glycoprotein Ib α chain
Q9HCN6	Platelet glycoprotein VI
P07476	Involucrin
Q9Y5Y7	Lymphatic vessel endothelial hyaluronic acid receptor
P04156	Major prion protein
P17900	Ganglioside GM2 activator
P18827	Syndecan-1
Q8TF72	Protein Shroom 3
Q9UBC9	Small proline-rich protein
P08138	Tumour necrosis factor
P30530	Tyrosine-protein kinase receptor

redundant proteins in our dataset (fig. 5). With this dataset, a set of 20 human proteins was mined from the most significant GO enrichment term ''response to stimulus'' (GO:0050896) (table 4). It is plausible that discrimination of TB disease status might be achieved using this panel of 20 human proteins; however, further studies will be required to determine the true discriminatory power of this panel.

Our analysis further identified seven immune-related human proteins (IGKC, RBP4, PTGDS, ORM1, IGCL2, AMBP and SECTM1) of which six were differentially expressed in active TB urine samples (fig. 6). A number of these have been previously observed to be associated with TB infection. For example, two related studies showed that RBP4 was a marker for TB infection, with elevated levels of RBP4 being found in the serum of cattle infected with *Mycobacterium bovis* [25] and with differential expression of RBP4 being
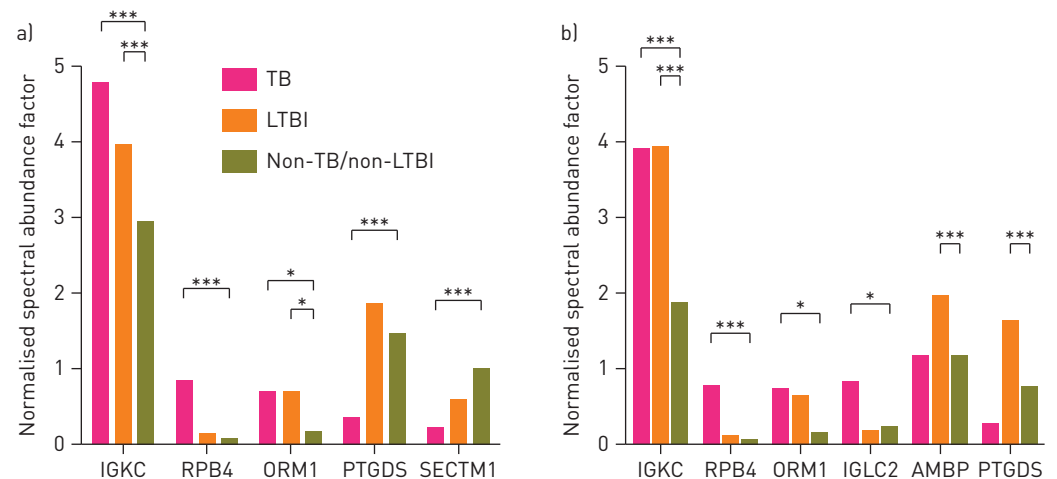

FIGURE 6 The relative quantitative analysis of the common human urine proteins from tuberculosis (TB) patients with active TB, presumed latent TB infection (LTBI) and presumed non-TB/non-LTBI observed from a) X! Tandem and b) Omssa. Significance was evaluated *via* two-way ANOVA with a Bonferroni correction by comparing the presumed LTBI and TB groups to the presumed non-TB/non-LTBI group. IGKC: immunoglobulin kappa chain C; RBP4: retinol binding protein 4; ORM1: α-1-acid glycoprotein 1; PTGDS: prostaglandin-H2 D-isomerase; SECTM1: secreted and transmembrane protein 1; IGLC2: immunoglobulin lambda-2 chain C; AMBP: α-1-microglobulin/bikunin precursor. *: $p < 0.05$; ***: $p < 0.001$.

observed in human whole blood supernatant [26]. In addition, SECTM1 has been shown to be a cognate of CD7, a protein found on T-cells, natural killer cells and pre-B cells and, while the function of SECTM1 is not well established [27], the SECTM1 cognate has been reported to stimulate the upregulation of CD25, CD54 and CD69 on human natural killer cells *in vitro* in TB infection. Furthermore, serum levels of AMBP have also been observed to be elevated levels in cattle with sub-clinical *M. bovis* infections [25]. Reduced levels of PTGDS have been reported in cerebrospinal fluid samples from patients with bacterial meningitis [28]; however, to our knowledge there has been no previous correlation between PTGDS and TB.

Based on our analyses, we hypothesise that the immune-related proteins IGKC, RBP4, PTGDS, AMBP, ORM1, IGCL2 and SECTM1 might be collectively used to diagnose TB disease in patient urine samples. For example, our observation of significantly increased abundance of RBP4 in active TB patients compared to the other two patient groups may provide the basis for a novel "rule-in" test for TB, whilst the significantly decreased abundance of PTGDS in active TB patients compared to LTBI patients may provide the basis for a "rule-out" test for TB (fig. 5). However, we note that it will be important to first determine whether such biomarkers can discriminate between TB and other respiratory diseases such as asthma, emphysema, pneumonia or sarcoidosis.

Proteomics offers tremendous opportunities for the discovery of diagnostic and/or prognostic biomarkers associated with localised or systemic disease, thus facilitating global policy for TB elimination [29]. Urine is a particularly good biological fluid for this type of analysis because, as an ultra-filtrate of blood, it is potentially representative of molecules from all organs of the body, including the lungs. Through the proteomic analysis of pooled and single urine samples collected from patients with definite pulmonary TB, presumed LTBI and presumed non-TB/non-LTBI, we have identified 10 exogenous *M. tuberculosis* proteins, as well as various panels of human proteins and seven common human proteins that appear to be indicative of TB infection. Validation studies are underway to confirm the frequency of these candidate biomarkers of TB disease.

Acknowledgements

References

1   World Health Organization. Rapid Implementation of the Xpert MTB/RIF Diagnostic Test: Technical and Operational 'How-to' Practical Considerations. Geneva, WHO Press, 2011.

2   Theron G, Peter J, van Zyl-Smit R, *et al*. Evaluation of the Xpert MTB/RIF assay for the diagnosis of pulmonary tuberculosis in a high HIV prevalence setting. *Am J Respir Crit Care Med* 2011; 184: 132–140.

3   Weyer K, Mirzayev F, Migliori GB, *et al*. Rapid molecular TB diagnosis: evidence, policy making and global implementation of Xpert MTB/RIF. *Eur Respir J* 2013; 42: 252–271.

4   Rangaka MX, Wilkinson KA, Glynn JR, *et al*. Predictive value of interferon-γ release assays for incident active tuberculosis: a systematic review and meta-analysis. *Lancet Infect Dis* 2012; 12: 45–55.

5   Peter J, Green C, Hoelscher M, *et al*. Urine for the diagnosis of tuberculosis: current approaches, clinical applicability, and new developments. *Curr Opin Pulm Med* 2010; 16: 262–270.

6   Pai M, Minion J, Steingart K, *et al*. New and improved tuberculosis diagnostics: evidence, policy, practice, and impact. *Curr Opin Pulm Med* 2010; 16: 271–284.

7   Dheda K, Ruhwald M, Theron G, *et al*. Point-of-care diagnosis of tuberculosis: past, present and future. *Respirology* 2013; 18: 217–232.

8   Minion J, Leung E, Talbot E, *et al*. Diagnosing tuberculosis with urine lipoarabinomannan: systematic review and meta-analysis. *Eur Respir J* 2011; 38: 1398–1405.

9   Peter JG, Theron G, van Zyl-Smit R, *et al*. Diagnostic accuracy of a urine lipoarabinomannan strip-test for TB detection in HIV-infected hospitalised patients. *Eur Respir J* 2012; 40: 1211–1220.

10   Wheelock CE, Goss VM, Balgoma D, *et al*. Application of 'omics technologies to biomarker discovery in inflammatory lung diseases. *Eur Respir J* 2013; 42: 802–825.

11   Rosenkrands I, Weldingh K, Jacobsen S, *et al*. Mapping and identification of *Mycobacterium tuberculosis* proteins by two-dimensional gel electrophoresis, microsequencing and immunodetection. *Electrophoresis* 2000; 21: 935–948.

12   Agranoff D, Fernandez-Reyes D, Papadopoulos MC, *et al*. Identification of diagnostic markers for tuberculosis by proteomic fingerprinting of serum. *Lancet* 2006; 368: 1012–1021.

13   Kashino SS, Pollock N, Napolitano DR, *et al*. Identification and characterization of *Mycobacterium tuberculosis* antigens in urine of patients with active pulmonary tuberculosis: an innovative and alternative approach of antigen discovery of useful microbial molecules. *Clin Exp Immunol* 2008; 153: 56–62.

14   Pollock NR, Macovei L, Kanunfre K, *et al*. Validation of *Mycobacterium tuberculosis* Rv1681 protein as a diagnostic marker of active pulmonary tuberculosis. *J Clin Microbiol* 2013; 5: 1367–1373.

15   Mehaffy MC, Kruh-Garcia NA, Dobos KM. Prospective on *Mycobacterium tuberculosis* proteomics. *J Proteome Res* 2012; 11: 17–25.

16   Zhang L, Lowrie D, Zhou H. Proteomic progress in studying tuberculosis from 2010 to 2011. *J Biophys Chem* 2011; 2: 395–400.

17   Nesvizhskii AI, Aebersold R. Interpretation of shotgun proteomic data: the protein inference problem. *Mol Cell Proteomics* 2005; 4: 1419–1440.

18   Martens L, Vandekerckhove J, Gevaert K. DBToolkit: processing protein databases for peptide-centric proteomics. *Bioinformatics* 2005; 21: 3584–3585.

19   Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 2009; 37: 1–13.

20   Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009; 4: 44–57.

21   Klahr S, ed. Atlas of Diseases of the Kidney Volume Four. Systemic Diseases and the Kidney. Hoboken, Wiley, 1999.

22   Nagaraj N, Mann M. Quantitative analysis of the intra- and inter-individual variability of the normal urinary proteome. *J Proteome Res* 2011; 10: 637–645.

23   Adachi J, Kumar C, Zhang Y, *et al.* The human urinary proteome contains more than 1500 proteins, including a large proportion of membrane proteins. *Genome Biol* 2006; 7: R80.

24   Baena A, Porcelli SA. Evasion and subversion of antigen presentation by *Mycobacterium tuberculosis. Tissue Antigens* 2009; 74: 189–204.

25   Seth M, Lamont EA, Janagama HK, *et al.* Biomarker discovery in subclinical mycobacterial infections of cattle. *PLoS One* 2009; 4: e5478.

26   Tanaka T, Sakurada S, Kano K, *et al.* Identification of tuberculosis-associated proteins in whole blood supernatant. *BMC Infect Dis* 2011; 11: 71.

27   Lyman SD, Escobar S, Rousseau a M, *et al.* Identification of CD7 as a cognate of the human K12 (SECTM1) protein. *J Biol Chem* 2000; 275: 3431–3437.

28   Tumani H, Reiber H, Nau R, *et al.* Beta-trace protein concentration in cerebrospinal fluid is decreased in patients with bacterial meningitis. *Neurosci Lett* 1998; 242: 5–8.

29   Diel R, Loddenkemper R, Zellweger JP, *et al.* Old ideas to innovate tuberculosis control: preventive treatment to achieve elimination. *Eur Respir J* 2013; 42: 785–801.