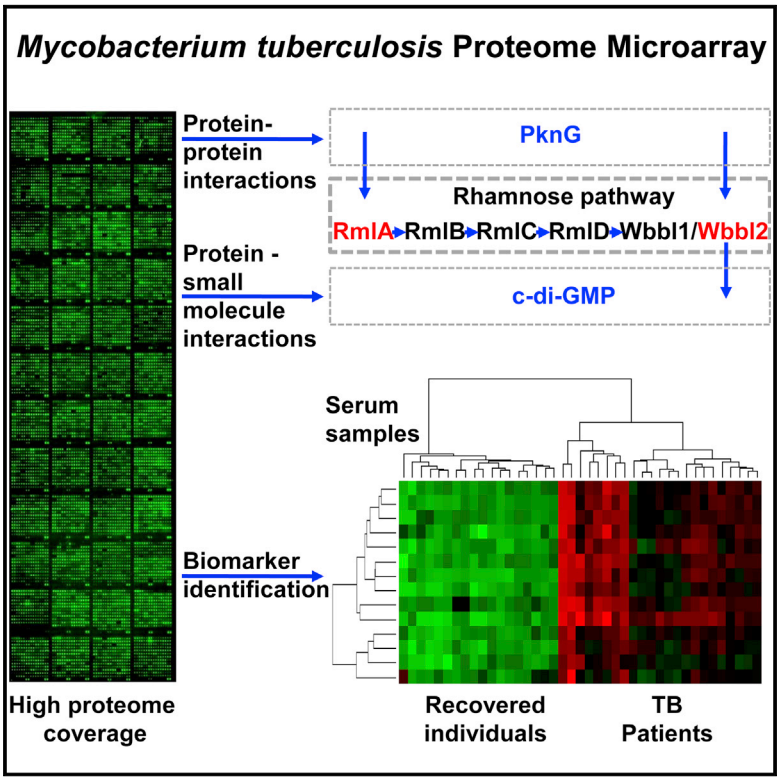
Cell Reports

Resource

Mycobacterium Tuberculosis Proteome Microarray for Global Studies of Protein Function and Immunogenicity

Graphical Abstract



Highlights

- An MTB functional proteome microarray covering most of the proteome is presented
- Applications include global protein-biomolecule interactions and biomarker discovery
- The MTB ORFome library is ready for protein expression
- The MTB rhamnose pathway is likely regulated by both PknG and c-di-GMP

Authors

Jiaoyu Deng, Lijun Bi, ..., Sheng-ce Tao, Xian-En Zhang

Correspondence

taosc@sjtu.edu.cn (S.-c.T.), zhangxe@sun5.ibp.ac.cn (X.-E.Z.)

In Brief

Deng et al. present a functional *Mycobacterium tuberculosis* proteome microarray covering most of the proteome and an ORFome library ready for protein expression and demonstrate its broad applicability for rapidly generating insightful proteome-wide information on this slow-growing pathogen by investigating global PknG and c-di-GMP interactions and identifying serum biomarkers.

Deng et al., 2014, Cell Reports 9, 2317–2329
December 24, 2014 ©2014 The Authors
http://dx.doi.org/10.1016/j.celrep.2014.11.023

Cell Reports
Resource

Mycobacterium Tuberculosis Proteome Microarray for Global Studies of Protein Function and Immunogenicity

Jiaoyu Deng,[3,8] Lijun Bi,[1,7,8] Lin Zhou,[4,8] Shu-juan Guo,[2] Joy Fleming,[1] He-wei Jiang,[2] Ying Zhou,[1] Jia Gu,[3] Qiu Zhong,[4] Zong-xiu Wang,[2] Zhonghui Liu,[1] Rui-ping Deng,[2] Jing Gao,[3] Tao Chen,[4] Wenjuan Li,[1] Jing-fang Wang,[2] Xude Wang,[3] Haicheng Li,[4] Feng Ge,[5] Guofeng Zhu,[6] Hai-nan Zhang,[2] Jing Gu,[3] Fan-lin Wu,[2] Zhiping Zhang,[3] Dianbing Wang,[1] Haiying Hang,[1,7] Yang Li,[2] Li Cheng,[2] Xiang He,[2] Sheng-ce Tao,[2,*] and Xian-En Zhang[1,*]

[1]National Key Laboratory of Biomacromolecules, Key Laboratory of Non-Coding RNA and Key Laboratory of Protein and Peptide Pharmaceuticals, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China
[2]Shanghai Center for Systems Biomedicine, Key Laboratory of Systems Biomedicine (Ministry of Education); State Key Laboratory of Oncogenes and Related Genes; and School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
[3]State Key Laboratory of Virology, Wuhan Institute of Virology, Chinese Academy of Sciences, Wuhan 430071, China
[4]Center for Tuberculosis Control of Guangdong Province, Guangzhou 510630, China
[5]Key Laboratory of Algal Biology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430071, China
[6]Shanghai Municipal Center for Disease Control and Prevention, Shanghai 200336, China
[7]TB Healthcare Biotechnology Co., Ltd., Foshan, Guangdong 528000, China
[8]Co-first author
*Correspondence: taosc@sjtu.edu.cn (S.-c.T.), zhangxe@sun5.ibp.ac.cn (X.-E.Z.)
http://dx.doi.org/10.1016/j.celrep.2014.11.023

SUMMARY

Poor understanding of the basic biology of *Mycobacterium tuberculosis* (MTB), the etiological agent of tuberculosis, hampers development of much-needed drugs, vaccines, and diagnostic tests. Better experimental tools are needed to expedite investigations of this pathogen at the systems level. Here, we present a functional MTB proteome microarray covering most of the proteome and an ORFome library. We demonstrate the broad applicability of the microarray by investigating global protein-protein interactions, small-molecule-protein binding, and serum biomarker discovery, identifying 59 PknG-interacting proteins, 30 bis-(3′-5′)-cyclic dimeric guanosine monophosphate (c-di-GMP) binding proteins, and 14 MTB proteins that together differentiate between tuberculosis (TB) patients with active disease and recovered individuals. Results suggest that the MTB rhamnose pathway is likely regulated by both the serine/threonine kinase PknG and c-di-GMP. This resource has the potential to generate a greater understanding of key biological processes in the pathogenesis of tuberculosis, possibly leading to more effective therapies for the treatment of this ancient disease.

INTRODUCTION

Mycobacterium tuberculosis (MTB), the etiological agent of tuberculosis and one of the most successful human pathogens, caused 8.6 million incident cases of tuberculosis (TB) and claimed 1.3 million lives in 2012 (WHO, 2013) in spite of century-long efforts to combat it. More effective drugs, vaccines, and diagnostic tests for TB are clearly needed, but their development is hampered by poor understanding of the basic biology of this pathogen (Galagan et al., 2013). Global studies on MTB at the systems level rather than traditional one-gene or one-protein approaches should lead to breakthroughs (Boshoff and Lun, 2010), but better tools with which to investigate the basic biology of MTB and its interactions with the host at the systems level are urgently needed (McFadden et al., 2013).

High-throughput "omics" techniques have been applied to the study of MTB biology in recent years to address questions at the systems level; genomic studies are increasing our understanding of MTB evolution (Comas et al., 2013) and the development of drug resistance (Farhat et al., 2013; Zhang et al., 2013), while proteomic studies, which have identified ~80% of all annotated MTB proteins (Schubert et al., 2013), the protein constituents of different cellular locations (Gu et al., 2003), and the in vivo MTB proteome (Albrethsen et al., 2013), and metabolomic studies are opening up our understanding of the real-time physiological status of bacilli in the host (Shin et al., 2011).

Proteome microarrays, usually composed of thousands of proteins from one species that are affinity purified and functionally active, are powerful highly parallel, high-throughput platforms for globally profiling thousands of molecular interactions in a single experiment (Chen et al., 2008; Zhu et al., 2001). Their use in the discovery of serum biomarkers for various diseases (Gnjatic et al., 2010) and global investigations of protein interactions with other proteins (PPI) (Chen et al., 2013), with DNA (Lin et al., 2009), with RNA (Zhu et al., 2007), with lipids (Lu et al., 2012), and with a range of small molecules (Huang et al., 2004) demonstrate the power of this approach. Furthermore, they

can also be used to study post-translational modifications (PTMs), including phosphorylation (Newman et al., 2013), acetylation (Lin et al., 2009), glycosylation (Kung et al., 2009), and nitrosylation (Lee et al., 2014), providing a window into dynamic cellular responses to changes in the environment. To date, the only MTB proteome-scale microarray available is one developed using an *E. coli*-based cell-free transcription/translation system where translated proteins were printed directly onto nitrocellulose-coated slides without further purification (Kunnath-Velayudhan et al., 2010). While representing a significant breakthrough toward defining the TB immunoproteome by facilitating proteome-scale analysis of serum samples, proteins on an array developed using this approach lack the purity necessary for broader functional studies. Difficulties in developing a suitable system for high-throughput protein expression and purification have, however, impeded the development of a functional MTB proteome microarray, as MTB proteins expressed in the *E. coli* system are frequently insoluble or trapped in inclusion bodies (Liu et al., 2014).

Here, we present a functional MTB proteome microarray developed using a yeast expression system and demonstrate its broad utility by using it to investigate protein-protein interactions and small-molecule-protein binding and identify serum biomarkers. We show that PknG, a Ser/Thr kinase (Prisic et al., 2010), and bis-(3′-5′)-cyclic dimeric guanosine monophosphate (c-di-GMP), a ubiquitous bacterial second messenger (Römling et al., 2013), are both involved in the regulation of the rhamnose biosynthesis pathway, a key pathway in cell wall synthesis (Brennan, 2003). In addition, we identified 14 serum biomarkers that can together differentiate between TB patients with active disease and recovered individuals and thus have potential as an index for monitoring treatment outcome. Our findings demonstrate the potential of the MTB proteome microarray for greatly increasing our understanding of the basic biology of TB and identifying additional molecular targets in the fight against this disease.

RESULTS

Construction of the MTB Proteome Microarray

To develop a suitable system for protein expression and purification, we first compared the expression of randomly selected MTB proteins in an *E. coli* system (Studier and Moffatt, 1986) with that in a *Saccharomyces cerevisiae* system that has previously been used to construct yeast (Zhu et al., 2001) and human proteome microarrays (Jeong et al., 2012). Consistent with previous reports that MTB proteins are difficult to express in *E. coli* (Liu et al., 2014), possibly due to the high GC content of MTB genes and unusual codon usage, less than 50% of 94 randomly selected proteins were expressed successfully in *E. coli*, and only 8 of these proteins were soluble (data not shown). Although a separate study showed that codon optimization significantly increases soluble expression in *E. coli* (13 of 16 tested proteins were solubly expressed at significantly higher levels after codon optimization; data not shown), the cost of using a codon optimization and whole-gene synthesis strategy to construct a proteome microarray for MTB was considered prohibitive. Expressing MTB proteins using the *S. cerevisiae* system was found to provide an attractive and cost-efficient alternative; 95 of 100 randomly selected proteins were successfully expressed in soluble form (data not shown). We thus chose the *S. cerevisiae* system to construct the MTB proteome microarray.

To express and purify MTB proteins for the microarray, we first shuttled 3,841 open reading frames (ORFs) in Gateway Entry vectors obtained from the Pathogen Functional Genomics Research Center (PFGRC), along with 794 additional full-length and sequence-verified ORFs generated in our lab, into a yeast expression vector (pEGH-A) that allows galactose-dependent overexpression of N-terminal glutathione S-transferase (GST)-tagged recombinant proteins as previously described (Jeong et al., 2012). The final expression library included 3,829 H37Rv clones and 433 CDC1551 clones (Figure 1A). After high-throughput expression and purification, performed as described previously (Jeong et al., 2012) with minor modifications (Figure S1), we determined the protein quality of a random selection of recombinant proteins using GST immunoblotting (249 proteins) (Figure 1B; Data S1) and capillary electrophoresis (305 proteins) (Figure 1C; Data S2). Many of the GST-fused MTB proteins (222/249) showed a clear band of the expected size (±10 kDa) and good purity, indicating that ~90% (222/249 = 89.28%) of the MTB proteins were successfully expressed in soluble format using this system. Most of the purified proteins had a concentration suitable for protein microarray fabrication (0.1 to 5 μg of protein was purified from 12 ml cultures). To further test the efficacy of the yeast expression system, we independently expressed 184 membrane proteins randomly selected from the 487 membrane proteins included on the microarray (MTB H37Rv has a total of 567 membrane proteins) and checked their expression by anti-GST western blotting. Results show that 84.23% (155/184) were solubly expressed, a slightly lower percentage than that of the 249 randomly selected proteins (Data S1).

We next spotted the full set of purified recombinant proteins, in duplicate, along with a set of controls (elution buffer, GST, BSA, and histones), on a single glass slide and evaluated the quality of the resulting array by probing with an anti-GST antibody (Figures 1D and S2A). The final protein set on the proteome microarray, consisting of 3,829 of the 4,016 ORFs of the MTB H37Rv proteome, was composed of 32% hypothetical proteins (Figure S2B), 10% membrane proteins, and 5% secreted proteins (Figure S2C), consistent with the known overall composition of the MTB proteome (Schubert et al., 2013) (proteins that we were unable to clone or express, and thus are not present on the current MTB proteome microarray, are listed in Table S1). This high coverage and close similarity in composition between our proteome microarray and that of the MTB proteome indicates that it should be suitable for a variety of global functional studies.

The distributions of foreground and background signal intensity exhibited typical bell-shaped curves, and the curves for foreground intensity (Figure S2D) and background intensity were almost completely separate, indicating that the vast majority of the printed spots on the microarray contained substantial levels of recombinant protein. Signal intensity was not significantly correlated with gene GC content, molecular weight (MW), or isoelectric point (Figures S2E–S2G), indicating that these variables

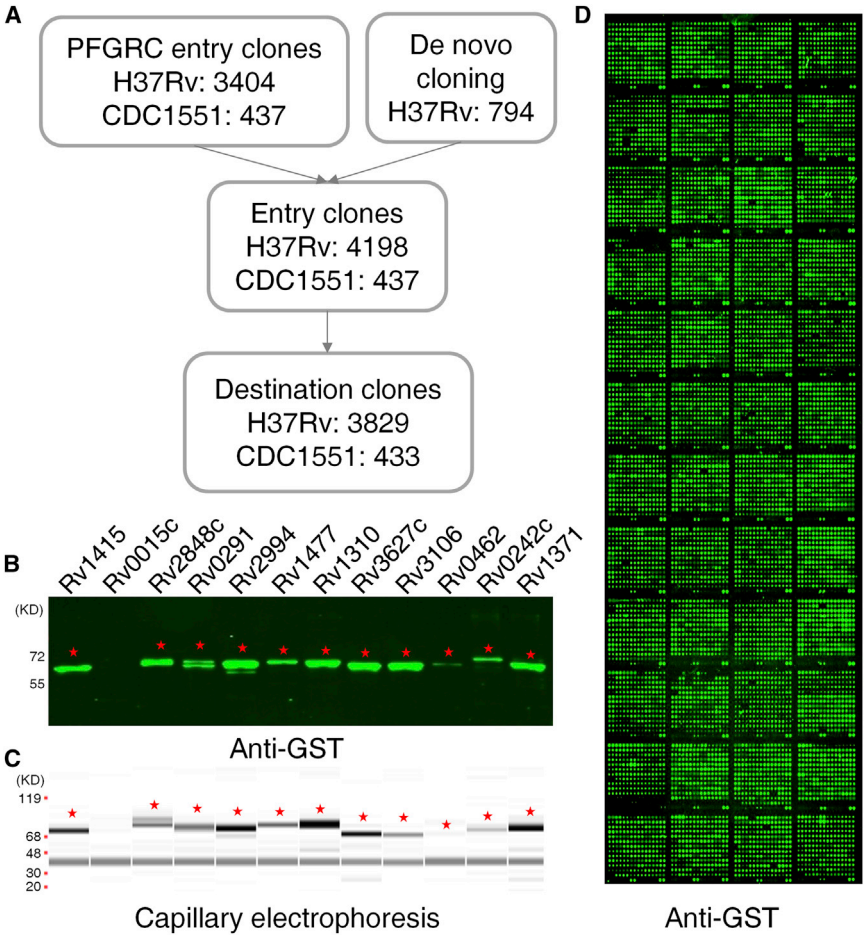


Figure 1. Construction of the MTB Proteome Microarray

(A) Protocol for high-throughput expression and purification of *M. tuberculosis* proteins.

(B and C) Quantitation and quality determination of purified *M. tuberculosis* proteins. Stars mark bands that are of the expected size for the given protein. (B) Protein quality was determined using an anti-GST antibody. (C) Capillary electrophoresis determination of protein quantity and size.

(D) A total of 3,829 and 433 proteins from MTB strains H37Rv and CDC1551, respectively, were spotted in duplicate on polymer slides (PolymerSlide H, CapitalBio). The slides were probed with an anti-GST antibody.

See also Figures S1 and S2 and Data S1 and S2.

did not affect the success of protein purification. To estimate the false-positive rate, we adopted a statistical model developed by Wang et al. for DNA microarray analysis (Wang et al., 2003). Given that the discordance rate of two replicated microarrays in proteome microarray experiments can easily be kept under 10% (the discordance rates for the PknG and c-di-GMP microarray experiments described below were 4.9% and 5.8%, respectively), the false-positive rate is $(10\%)^2 /(1 - 2 \times 10\% + (10\%)^2) = 1.23\%$.

The above analyses indicate that the yeast system is suitable for high-throughput expression and purification of MTB proteins. This microarray represents the most comprehensive MTB proteome microarray composed of individually purified proteins reported to date. Since affinity purification of proteins is protein-friendly, it is likely that the proteins on the microarray, like those on widely used human and yeast proteome microarrays prepared using the same protocols (Jeong et al., 2012; Zhu et al., 2001), are functionally active.

We demonstrate the broad utility of the MTB proteome microarray for functional studies in the experiments that follow by using it in three typical applications: investigations of protein-protein interactions and small-molecule-protein binding and identification of serum biomarkers.

Global Kinase-Protein Interactions

Protein serine/threonine kinases (PSTKs) play important roles in antibiotic resistance, host-pathogen interactions, and the survival of MTB in macrophages (Tiwari et al., 2009). MTB has 11 PSTKs, indicating that Ser/Thr phosphorylation is likely an important mechanism for signal transduction in this organism (Prisic et al., 2010). Of these, PknG is of particular interest, as it is secreted into the host cytosol (Walburger et al., 2004) and plays a key role in infection (Székely et al., 2008). It may thus have significant potential as a drug target (Scherr et al., 2007); however, its regulation and the mechanism by which it regulates the function of other proteins is still poorly understood.

To date, the only MTB protein known to bind to PknG is GarA (Rv1827), a key component of the glutamate metabolic pathway (O'Hare et al., 2008). To identify additional PknG-interacting proteins, we purified N-terminal V5-tagged PknG and used it to probe the MTB proteome microarray (Figure 2). Using the stringent criteria described in Experimental Procedures, we identified 59 candidate PknG-interacting proteins (Table S2), 20 of which are known to be secreted (Målen et al., 2007). The six proteins with the strongest interactions are shown in Figure 2B.

To understand the biological relevance of the 59 candidate PknG interactors (Table S2), we examined their enrichment in specific protein classes using PANTHER (Mi et al., 2013). The PknG interactors classified into 14 protein classes (Figure 2C), the top three of which were hydrolases (17.8%), transferases (17.8%), and oxidoreductases (13.3%). While previous studies have largely focused on the role of PknG in glutamate and glutamine metabolism (O'Hare et al., 2008), our results suggest that PknG's activity and substrate specificity may be regulated by a variety of proteins involved in widespread biological processes. Gene ontology (GO) analysis based on DAVID (Huang et al., 2009) showed that the 59 candidate PknG interactors are

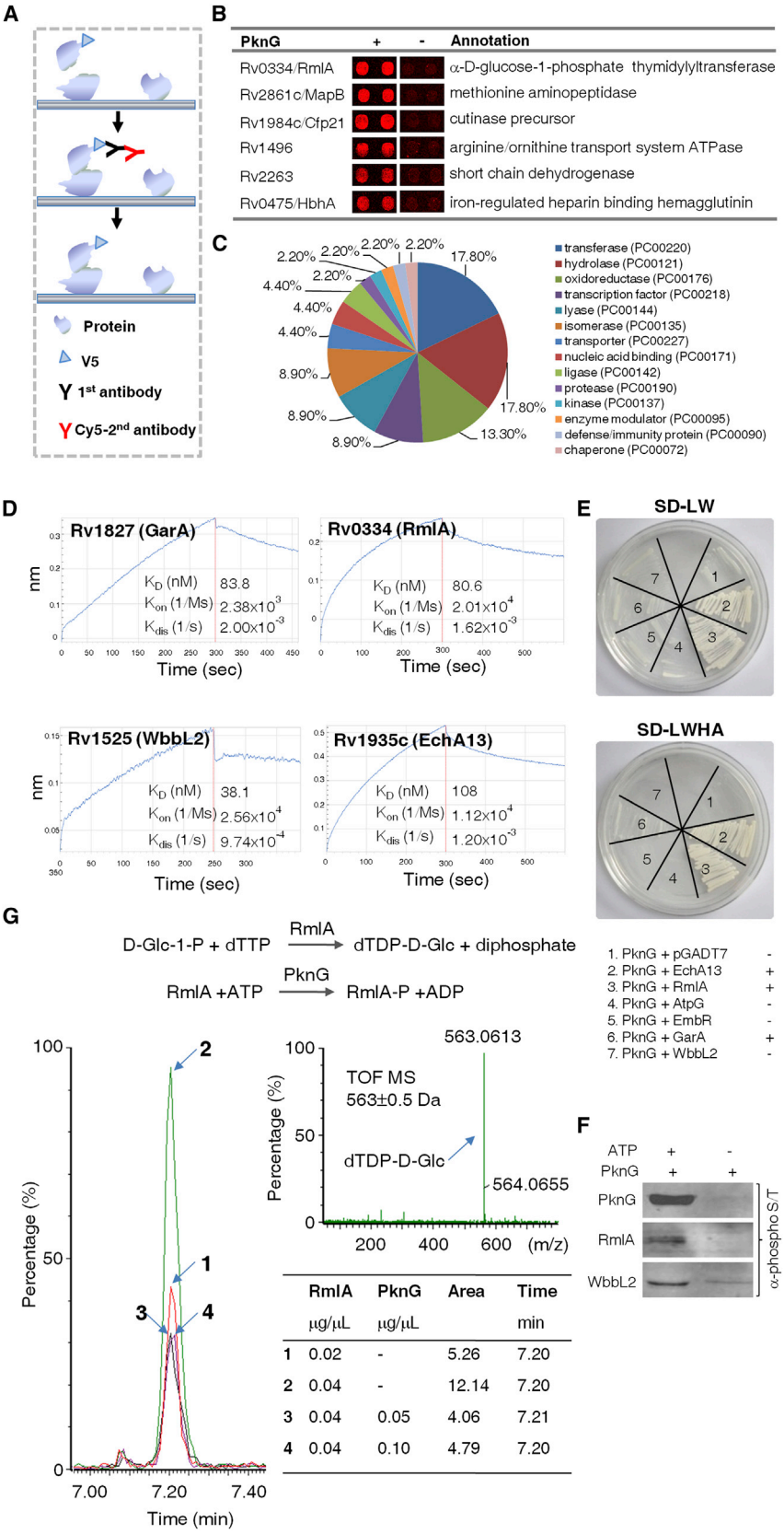


Figure 2. Identification of PknG Interactors and Interaction Kinetics

(A) Schematic of the procedure for detection of binding events. A V5-specific monoclonal antibody was applied to the microarray followed by incubation with a Cy5-labeled secondary antibody. GarA (Rv1827) and BSA were used as positive and negative controls, respectively.

(B) Representative PknG-interacting proteins. Cy5 signals were measured at 635 nm. Two replicate results are shown. +, in the presence of PknG; −, in the absence of PknG.

(C) PANTHER classification of PknG interactors.

(D) Bilayer interferometry data for the binding of PknG to selected proteins identified as interactors on the MTB proteome microarray and their interaction kinetics. PknG was immobilized on streptavidin-coated biosensors and exposed to binding partners in SD buffer. Binding was measured by coincident changes in the interference pattern. Results presented are representative of at least three experiments.

(E) Y2H validation of representative PknG-protein interactions. *S. cerevisiae* strain AH109 competent cells were transformed with the plasmid combinations indicated (bait plus prey) to assay for interactions between PknG and the selected proteins. PknG was used as the bait protein and Rv0334 (RmlA), Rv1525 (Wbbl2), and Rv1935 (EchA13) as prey. Rv1827 (GarA) and an empty pGADT7 were included as positive and negative controls, respectively.

(F) PknG phosphorylates RmlA and Wbbl2. PknG was incubated with RmlA or Wbbl2 for 1 hr at 30°C. RmlA/Wbbl2 phosphorylation was subsequently detected by western blotting with a phosphoserine/threonine antibody.

(G) RmlA activity is inhibited by PknG. An RmlA activity assay was performed using D-Glc-1-P and dTTP as substrates. The major product of this reaction, dTDP-d-Glc, was measured by ultraperformance liquid chromatography mass spectrometry.

See also Data S3.

significantly enriched in transferase activity, with 10 proteins being annotated as transferases ($p = 0.014$), and are slightly enriched in membrane proteins (Table S2).

Putative PknG-protein interactions were validated by Bio-Layer Interferometry (BLI) (Abdiche et al., 2009) and yeast-two-hybrid analysis, using Rv1827 (GarA), a known PknG binder, as a positive control. Sufficient quantities of protein for BLI analysis were obtained for 47 of the 59 putative interactors, and 72.3% (34/47) of these showed strong interactions with PknG (Data S3), indicating that these interactions are of relatively high affinity. PknG and BSA showed no interactions under the same conditions. For example, the equilibrium dissociation constants (K_D) for the interactions between PknG and Rv1827 (GarA), Rv0334 (RmlA), Rv1525 (Wbbl2), and Rv1935c (EchA13) were determined to be 83.8 nM, 80.6 nM, 38.1 nM, and 108 nM, respectively (Figure 2D). A few representative putative PknG interactors were selected for further validation by yeast two-hybrid analysis. Results confirmed that RmlA and EchA13 are genuine interactions (Figure 2E). Together, these results strongly suggest that most of the PknG interactors are authentic and that PknG may thus either have broad substrate specificity or be regulated by a variety of other proteins.

Interestingly, we found that PknG interacts with two key components of the rhamnose pathway that are essential for the growth of MTB: (1) RmlA (Rv0334), α-D-Glc-1-P-thymidylyltransferase, which catalyzes the formation of dTDP-Rha from thymidine triphosphate and glucose-1-phosphate, the first of four steps in the synthesis of L-rhamnose, and (2) Wbbl2 (Rv1525), a predicted rhamnosyl transferase possibly involved in cell wall arabinogalactan linker formation that inserts a rhamnose residue into the cell wall (Griffin et al., 2011). Rhamnose plays an essential structural role in the linker disaccharide that joins the peptidoglycan and arabinogalactan components of the mycobacterial cell wall (Brennan, 2003). As such, the rhamnose synthesis pathway has been viewed as a potential target for the design of inhibitors of bacterial growth (Grzegorzewicz et al., 2008). Having verified the interactions between PknG and RmlA and Wbbl2, we performed in vitro kinase assays (Prisic et al., 2010) to test whether PknG phosphorylates RmlA and Wbbl2. Results demonstrate clearly that both RmlA and Wbbl2 are phosphorylated by PknG (Figure 2F). We next measured the activity of RmlA when PknG and ATP were included in the reaction directly to test whether the interaction of RmlA with PknG alters its activity, (Figure 2G). Results indicated that PknG sharply inhibits the activity of RmlA in vitro, decreasing the dTDP-d-Glc reaction product peak area from 12.14 to 4.06 (mV*min), suggesting that RmlA activity may be regulated by its phosphorylation by PknG. Further experimental work is required to investigate the ultimate impact of the RmlA and PknG interaction on rhamnose and cell wall biosynthesis.

Global Profiling of c-di-GMP-Protein Interactions

c-di-GMP is a ubiquitous second messenger in bacteria that regulates a wide range of cellular processes such as biofilm formation, virulence, motility, and differentiation (Hengge, 2009). Its functional diversity can be attributed to its diverse upstream ligands and downstream effectors (Römling et al., 2013). Although genomic, biochemical, and gene expression profiling studies have clearly demonstrated that c-di-GMP signaling is active and may play important roles in *M. tuberculosis* (Li and He, 2012), our understanding of downstream effectors in *M. tuberculosis*, i.e., the proteins that c-di-GMP binds to, is very limited. The physiological roles of c-di-GMP in *M. tuberculosis* are thus still elusive.

To identify c-di-GMP-interacting proteins, we probed the MTB proteome microarray with a biotinylated c-di-GMP, then incubated it with Cy3-labeled streptavidin to detect binding events (Figure 3A). The top six of the 30 potential c-di-GMP-interacting proteins identified are shown in Figure 3B (Table S3).

Classification of these 30 candidate c-di-GMP binders using PANTHER (Table S3) showed that they are represented in 12 protein classes (Figure 3C), the top four of which are transferases (28.6%), hydrolases (21.4%), transporters (14.3%), and kinases (14.3%). These data suggest that c-di-GMP may bind to a wide range of proteins in diverse protein classes, consistent with reports that c-di-GMP signaling is involved in the regulation of many biological processes in other bacterial species (Römling et al., 2013). GO analysis based on DAVID (Huang et al., 2009) showed that the 30 c-di-GMP binders are significantly enriched in various functions including ABC-2 type transporters ($p = 0.002$), biphenyl-2,3-diol 1,2-dioxygenase activity ($p = 0.006$), and N-terminal protein amino acid acetylation ($p = 0.006$; Table S3).

We next examined the 30 candidate c-di-GMP binding proteins for consensus sequences that may represent important binding sites, performing a Dilimot motif search for enriched motifs (Neduva et al., 2005) on the primary amino acid sequences. The most significant motif identified, AxxxxAxV ($p = 1.35 \times 10^{-9}$), was present in 12 of the 30 binders. The well-known GGDEF, EAL, and PilZ domains, commonly found in c-di-GMP biosynthetic enzymes and some other binding proteins (Navarro et al., 2009), were not found in any of the 30 candidate c-di-GMP binding proteins identified here. Indeed, none of the H37Rv proteins present in the Pfam database (Finn et al., 2010) carry PilZ domains, only one protein (Rv1354c) carries a GGDEF domain, and two proteins (Rv1354c and Rv1357c) carry EAL domains. In view of the limited research to date on the role of c-di-GMP in MTB, it is not surprising that we have identified motifs among the 30 binders. Further detailed experimental investigation is necessary to understand the functional or regulatory role these structural motifs may play.

To validate these candidate c-di-GMP-protein interactions, we purified seven randomly selected proteins and incubated them with biotinylated c-di-GMP before performing western blotting with streptavidin to detect interactions (Figure 3D). Results clearly showed that c-di-GMP strongly interacts with all seven proteins but bound relatively weakly to the positive control Rv0767c, a known c-di-GMP binder (Li and He, 2012), and not at all to BSA.

BLI analysis of the interactions and kinetic parameters of putative c-di-GMP interactors Rv1525 (Wbbl2), Rv3756c (ProZ), and Rv1478 showed they had strong interactions with c-di-GMP and that there was no interaction between c-di-GMP and BSA under the same conditions. The equilibrium dissociation constants (K_D) for the interactions between c-di-GMP and Rv1525, Rv3756c, and Rv1478 were 16.9 nM, 102 nM, and 138 nM, respectively

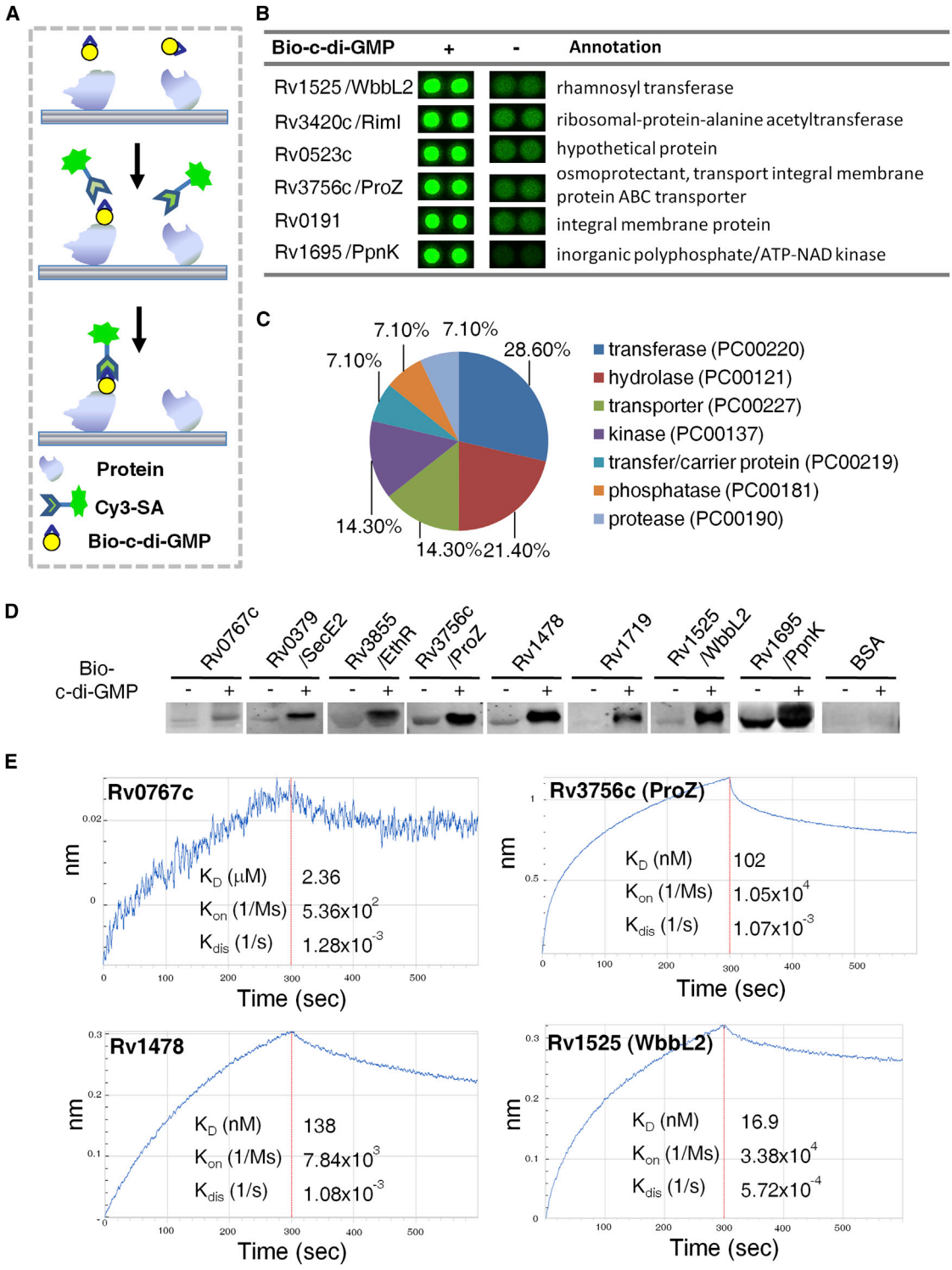


Figure 3. Identification of c-di-GMP Interactors and Interaction Kinetics

(A) Schematic of the procedure for detecting binding events. After probing the MTB proteome microarray with a biotinylated c-di-GMP, binding events were detected with Cy3-labeled streptavidin.

(B) Representative c-di-GMP-interacting proteins. Cy3 signals were measured at 532 nm. Two replicate results are shown. +, in the presence of biotinylated c-di-GMP; −, in the presence of biotin alone.

(C) PANTHER classification of c-di-GMP interactors.

(legend continued on next page)

(Figure 3E), indicating, as expected, that these interactions are of relatively high affinity. The positive control, Rv0767c, also had a significant interaction with c-di-GMP, with a K_D of 2.36 μM.

As the c-di-GMP-protein interactions validated here were selected at random, it seems reasonable to expect that the interactions of c-di-GMP with most of the proteins in Table S3 are authentic. Thus, as in other bacterial species, c-di-GMP appears to have a broad regulatory role in MTB.

Interestingly, like PknG, c-di-GMP also bound to Wbbl2, a predicted rhamnose transferase in the rhamnose synthesis pathway, as mentioned above. Further functional studies are needed to test whether c-di-GMP binding alters the activity of Wbbl2, thus affecting MTB cell wall synthesis.

Serum Biomarker Identification

Additional biomarkers for the diagnosis of TB and monitoring of treatment are urgently needed to supplement the three well known T cell antigens, ESAT-6, CFP-10, and Ag85, which are currently used in T cell based assays such as ELISPOT (Hill et al., 2007) but still lack sufficient diagnostic power to reliably distinguish patients with active disease from healthy individuals or provide treatment monitoring (Sester et al., 2011). B cell antigens, such as those used in the diagnosis of Hepatitis B (Jaroszewicz et al., 2010), can be easily applied via simple serum-based ELISA assays or Colloidal gold test strips and could potentially be used in routine TB diagnosis and treatment monitoring.

Here, we used the MTB proteome microarray to identify serum biomarkers for assessing therapy outcome, by comparing antibodies present in sera from 189 TB patients with active disease sampled before the onset of treatment with those from 150 'recovered' TB patients; individuals who were smear-positive before treatment and became smear-negative after completing six months of antibiotic treatment (Figure 4). There was no significant difference in the distribution of age and sex between the two groups (Figure 4B). Samples were divided into an initial training set (21 from the "patient" and 20 from the "recovered" groups) that was screened to identify disease-state-specific protein antigens (data were analyzed using the Bioconductor package in R 3.0.) and two test sets (64 and 104 samples from the "patient" and 43 and 87 samples from the "recovered" groups in set 1 and set 2, respectively) that were used to validate disease-state-specific antigens identified using the training set. Sample sizes in the training and test sets were optimized to maximize the likelihood of detecting all reactive proteins bearing in mind the total number of samples available (339) and the ~5% margin of error (0.95 confidence level) associated with this population size (Data S4).

Disease-state-specific antigens showed hierarchical clustering and were able to provide discrimination between the TB patient and recovered groups in the training set (Figure 4C). The serum antibodies against 13 MTB proteins were at significantly higher levels in TB patients than that in the recovered group, while that for GroEL (Rv0440) was significantly lower ($p < 0.05$; Table 1). Both rounds of validation confirmed that this panel of candidate biomarkers likely has potential for monitoring therapy outcome (Figure 4D). The diagnostic accuracy of the 14 candidate biomarkers identified here (Table 1; Data S4), assessed by receiver operating characteristic (ROC) curves using the Daim package in R 3.0.1, reached 78.4%, with a sensitivity of 80.2% and specificity of 79.0% (Figure 4E), indicating that this set of biomarkers has good potential as an index for monitoring treatment outcome.

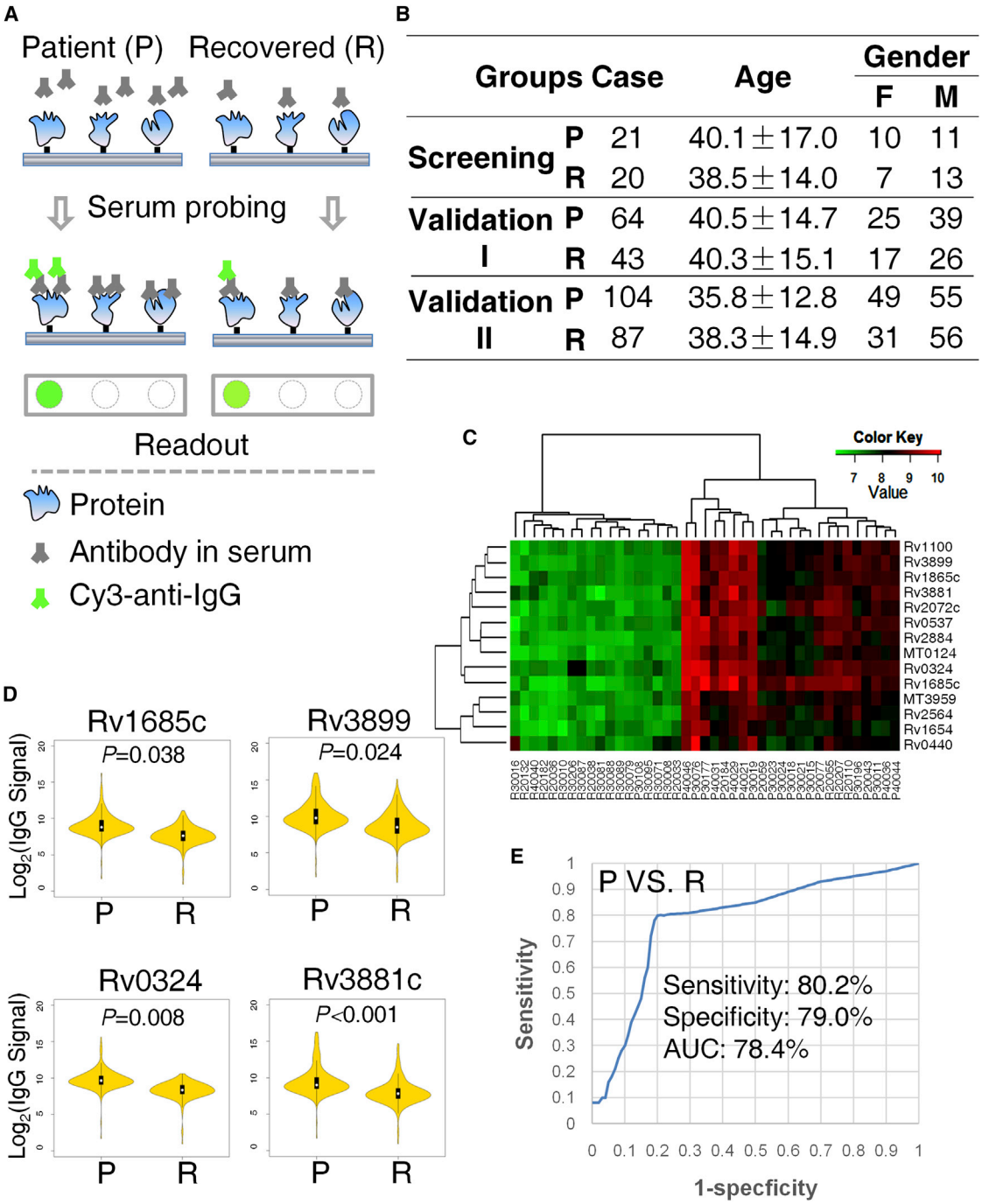
DISCUSSION

Systems biology approaches are beginning to unlock many aspects of the basic biology of MTB and its host-interactions, which have hitherto been bottlenecks in the development of more effective therapeutic approaches for the treatment of tuberculosis (McFadden et al., 2013). However, we do not yet have a global picture of intracellular and MTB-host interactions at the protein level due to a lack of suitable proteome-scale tools. The proteome microarray described here for MTB contains affinity-purified proteins and provides a convenient and versatile platform for proteome-scale functional studies. The MTB ORFome library constructed here (with >95% coverage of the MTB proteome) will also facilitate functional studies, as it provides a solution for readily expressing and purifying MTB proteins in soluble form. We have demonstrated the broad utility of the microarray by investigating global protein interactions with PknG, a Ser/Thr kinase and potential drug target, and c-di-GMP, a ubiquitous prokaryotic second messenger, identifying 59 candidate PknG-interacting proteins and 30 candidate c-di-GMP binders. Of particular interest, and demonstrating the power of unbiased global studies to generate unexpected and insightful results, these two independent investigations unexpectedly intersected, both bringing regulation of the rhamnose biosynthesis pathway enzymes RmlA and Wbbl2 into the spotlight. In addition, we identified 14 serum biomarkers that have potential as an index for monitoring treatment outcome; used together, they can discriminate between patients with active disease and those who have recovered from TB.

Our goal was to design a convenient experimental platform to expedite systems-level functional studies of this slow-growing and experimentally intransigent pathogen. To ensure the microarray was of high quality (high coverage of the MTB proteome with proteins of high purity), good reproducibility, and broad utility, we first constructed a complete MTB ORFome library, then expressed proteins using a yeast system. Proteins were affinity-purified via GST tags and spotted onto glass slides. This strategy maximizes protein purity and increases the likelihood

(D) Streptavidin blotting of selected c-di-GMP interactors. Purified proteins were coincubated with 10 μM biotin-c-di-GMP. After bring electrophoresized on 10% SDS-PAGE, proteins were transferred to a polyvinylidene fluoride membrane that was incubated with an IRDye 800 CW streptavidin. Rv0767c, a known c-di-GMP binder, and BSA were included as positive and negative controls, respectively.

(E) BLI data for the binding of c-di-GMP to selected binding partners identified on the MTB proteome microarray and their interaction kinetics. c-di-GMP was immobilized on streptavidin-coated biosensors and exposed to binding partners in buffer. Binding was measured by coincident changes in the interference pattern. Results presented are representative of at least three experiments.



Groups		Case	Age	Gender	
				F	M
Screening	P	21	40.1±17.0	10	11
	R	20	38.5±14.0	7	13
Validation I	P	64	40.5±14.7	25	39
	R	43	40.3±15.1	17	26
Validation II	P	104	35.8±12.8	49	55
	R	87	38.3±14.9	31	56

Figure 4. Identification of Serum Biomarkers for Recovery from TB

(A) Schematic of the experimental procedure for determining differential expression. Microarrays were probed with serum samples from patients with active TB disease or individuals who had recovered from TB and bound serum antibodies were detected by incubation with cy3-anti-immunoglobulin G.

(B) Characteristics of study population.

(C) Differentially expressed proteins showed hierarchical clustering, successfully discriminating between the TB patient and recovered groups.

(D) Violin plots of proteins expressed at significantly higher levels in TB patients than in recovered individuals. P, serum samples from TB patients; R, serum samples from recovered individuals. The median (white horizontal line within central black box), interquartile range (black box), and minimum and maximum values (black whiskers) are indicated. The yellow shaded area represents a kernel density plot. The significance of comparisons was tested using the Mann-Whitney *U* test (two tailed).

(E) Receiver operating characteristic curve.

Table 1. Candidate Serum Biomarkers Identified with the MTB Proteome Microarray

	Rv Number	Annotation
1	Rv0324[a]	sulfurtransferase
2	Rv0537c[b]	probable integral membrane protein
3	Rv1685c	transcriptional regulator, TetR family
4	Rv2072c[c]	precorrin-6Y C(5,15)-methyltransferase
5	Rv3899c[d]	uncharacterized protein
6	Rv1100[a]	conserved protein
7	Rv1865c	oxidoreductase, short-chain dehydrogenase/reductase family
8	Rv3881c[a,d,e,f]	ESX-1 secretion-associated protein EspB
9	MT0124	putative uncharacterized protein
10	Rv2884	probable transcriptional regulator protein
11	MT3959	putative uncharacterized protein
12	Rv2564[b]	uncharacterized ABC transporter ATP-binding protein
13	Rv1654[a]	acetylglutamate kinase
14	Rv0440[a,f]	60 kDa chaperonin 2

Annotations are based on the Sanger Institute database (http://tuberculist.epfl.ch/). See also Data S4.
[a]Secreted protein identified by Albrethsen et al. (Albrethsen et al., 2013).
[b]Annotated as a membrane protein.
[c]Located in the RD12 region of the genome (Behr et al., 1999).
[d]Secreted protein identified by Målen et al. (Målen et al., 2007).
[e]Identified previously as a serum biomarker (Kunnath-Velayudhan et al., 2010).
[f]Constituent of PPD (Prasad et al., 2013).

of correct protein folding, making the microarray suitable for functional studies. Although the current MTB proteome microarray already covers most of the MTB H37Rv proteome, our aim is to provide a microarray with 100% coverage. We are currently working to overcome problems encountered with the expression of the remaining 228 H37Rv proteins (Table S2), including several proteins of particular interest, i.e., GyrB, IniA, PknG, and KatG. As the absence of these proteins may hinder some applications of the proteome microarray, especially in studies related to these proteins, we will add these proteins when we update the microarray. While significant time, cost, and effort have been invested in the construction of this platform, the resulting microarrays are easy to use, versatile, and give reliable and reproducible results, without the need associated with other microarray platforms for performing in vitro translation experiments before use (Kunnath-Velayudhan et al., 2010), with their concomitant challenges in terms of protein purity and reproducibility.

Our screens for PknG and c-di-GMP binding proteins identified a significant number of binders (59 for PknG and 30 for c-di-GMP; Tables S2 and S3). BLI screening (Figures 2 and 3; Data S3) validated ~72% of the PknG interactions, confirming the reliability of the microarray. Based on this rate of validation, if we assume that all the validated interactors are functionally active for protein-protein interactions, we estimate that the array can potentially detect 65% (0.9 × 0.72 × 100) of all possible interactions of the proteins on the microarray. Bioinformatic analyses of the binders suggest the involvement of PknG and c-di-GMP in a broader range of biological processes than previously appreciated (Tables S2 and S3). PknG is known to be secreted and plays a role in host-pathogen interactions; Walburger et al. showed that secretion of PknG enhances mycobacterial survival within macrophages by inhibiting phagosome-lysosome fusion (Walburger et al., 2004). It is therefore of interest that 20 of the interactors identified here are also reported to be secreted (Albrethsen et al., 2013; Målen et al., 2007). The remaining binding partners are cytosolic and thus indicate that PknG may also play important intracellular regulatory roles. Further detailed functional studies are required to determine the biological significance of the interactions identified here.

Having a comprehensive picture of the proteins that are regulated by c-di-GMP or that regulate c-di-GMP intracellular levels should greatly expand our understanding of MTB pathophysiology. In other bacterial species, c-di-GMP levels are known to regulate a variety of biological functions. The high level of confidence associated with the c-di-GMP binding proteins identified here opens up the possibility of exploring these functions in MTB. In addition, c-di-GMP is also known as a regulator of immune responses to pathogen infection (Burdette et al., 2011). Investigation of the regulation of c-di-GMP itself, especially its levels and polarization in the cell, will likely provide important insights into host immune responses.

Unexpectedly, our results showed that both PknG and c-di-GMP interact with enzymes of the rhamnose synthesis pathway, a pathway that is crucial in cell wall synthesis. As drugs that inhibit cell wall biosynthesis are among the most highly effective antibiotics in use against TB, these findings may point toward additional strategies for drug design. Rhamnose is a glycan not found in humans and plays a key structural role in the mycobacterial cell wall (Brennan, 2003). Both the rhamnose biosynthesis pathway enzymes RmlA (interacts with PknG; Figure 2) and Wbbl2 (interacts with both PknG and c-di-GMP; Figures 2 and 3) are essential for the survival of *M. tuberculosis* (Griffin et al., 2011) and thus may be attractive drug targets. In addition, Wbbl2 is specific to the mycobacteria. Our investigation provides insight on the rhamnose pathway, showing that both PknG and c-di-GMP are likely involved in its regulation, suggesting two further avenues of approach for developing drugs that will inhibit cell wall synthesis.

Early proteomic studies on MTB have largely been descriptive, yielding information on protein abundance, localization, and differential expression under various conditions (Mehaffy et al., 2012). As the field moves away from a focus on simply cataloging proteins to unraveling their cellular and pathophysiological functions, two recent studies that address the lack of accurate and specific antigen markers for active TB disease using high-throughput approaches (Kunnath-Velayudhan et al., 2010; Liu et al., 2014) deserve particular comment and comparison with the approach taken here. Kunnath-Velayudhan et al. developed an in vitro transcription and translation-based MTB proteome microarray containing 3,988 unpurified proteins from the H37Rv strain and probed it with human serum samples, identifying 494 antigenic proteins. While this microarray has led to significant progress in defining the MTB immunoproteome (Kunnath-Velayudhan and Porcelli, 2013), the low purity of the

proteins on the microarray limits its use in both analytical and functional applications. Recognizing this limitation, Liu et al. chose a different strategy, successfully performing high-throughput cloning, expression, and purification of 1,250 MTB proteins (Liu et al., 2014). Despite the restricted coverage of the proteome, subsequent screening using semiquantitative western blotting and ELISPOT assays identified 29 antigenic proteins involved in humoral immune reactions and 34 in cellular immune reactions. The MTB proteome microarray described here, with its affinity-purified proteins, high-coverage of the proteome, and convenient and versatile experimental platform, offers an alternative platform that can be used for broader functional and analytical studies as well as the identification of protein antigens in serum samples.

Studies on biomarkers, such as those described above, have generally focused on identifying markers that can distinguish healthy individuals from patients with active TB disease. Here, we turned our attention to identifying biomarkers for monitoring treatment outcome. The length of anti-TB therapy (6–8 months for drug-sensitive TB and 2 years for drug-resistant TB) has negative implications for patient adherence, making clinical management of TB problematic (Lienhardt et al., 2012). As a shortened course of antibiotics may be sufficient for individuals responding early to treatment (Yew et al., 2011), early classification of TB patients into risk groups requiring different lengths of treatment is desirable to improve adherence and thus treatment outcome. Inexpensive, reliable, and rapid tests such as simple blood tests using serum biomarkers are needed to replace sputum microscopy, the current gold standard for monitoring the outcome of chemotherapy, because it lacks sensitivity and a substantial proportion of baseline sputum samples from new TB patients test negative (Steingart et al., 2006). The 14 biomarkers identified here (Table 1; Data S4) had a diagnostic accuracy of 78.4%, sensitivity of 80.2%, and specificity of 79.0% (Figure 4E), indicating that together they should be an effective index for monitoring treatment outcome.

Our proteome microarray has many potential applications. First, used to identify human proteins involved in host-pathogen interactions, it could facilitate identification of the host processes and interactions that are key in determining whether MTB infection leads to development of active disease, asymptomatic latent infection, or total clearance of bacteria. Second, it provides a versatile platform for investigating many aspects of the basic biology of MTB at the systems level, including global protein interactions with DNA, RNA, lipids, and global studies of posttranslational modifications, such as the profiling of kinase substrates demonstrated here. As such, this microarray represents a significant advance over traditional methodologies. Third, it is a powerful platform for the direct global identification of drug-targeted proteins, both for new drug candidates and for compounds that are effective but whose mechanism is poorly understood. The microarray could thus greatly accelerate mode of action studies for TB drugs/compounds. Investigations utilizing treatment monitoring biomarkers could also play a key role in the validation of drug candidates, accelerating drug development by reducing the length of clinical trials. In addition, investigations of the interactions between small molecules, their targets, and the system as a whole using this platform could lead to much-needed strategies for overriding the propensity of MTB to develop drug resistance. Fourthly, the proteome microarray could be used to identify potential biomarkers and immunogens and to expand our understanding of basic TB immunology. Examining serological differences among different groups of people, such as healthy individuals, those with latent TB, patients with active disease, and those recovered from TB, should lead to identification of useful diagnostic and therapeutic biomarkers.

Results reported here demonstrate that a systems-level approach utilizing our MTB proteome microarray can rapidly provide insightful proteome-wide information that opens up further avenues of research on the basic biology of MTB. This versatile proteome microarray with its many applications may greatly expedite research on this experimentally intransigent pathogen, ultimately leading to much-needed more effective therapeutic strategies and interventions for the treatment of TB.

EXPERIMENTAL PROCEDURES

MTB Proteome Microarray Construction

Cloning of MTB ORFs and protein purification was performed using published methods (Newman et al., 2013; Zhu et al., 2001). The entry clone collection, consisting of 3,404 H37Rv and 437 CDC1551 fully sequenced ORF clones, was a kind gift from the PFGRC (Rockville, MD). We designed primers for full-length Gateway-compatible cloning of 794 H37Rv ORFs that were either missing or had frameshift mutations/were truncated in the PFGRC collection using Primer Premier 5.0 from PREMIER Biosoft, following the same criteria used in the construction of the PFGRC collection. Primers were synthesized by Sangon Biotechnology. PCR amplification was carried out using a Phusion High-Fidelity PCR Kit (New England Biolabs), and the purity and identity of the PCR products was verified by gel electrophoresis. Verified PCR products were purified using PCR cleaning kit DP204 (Tiangen Biotech) and then subjected to the Gateway BP reaction using donor vector pDONR221 and Gateway BP mix (Life Technologies) according to the manufacturer's instructions. All sequence-verified entry clones were subjected to the Gateway LR reaction using pEGH-A (Lin et al., 2009) as the destination vector, then transformed into competent cells of *S. cerevisiae* strain Y258. Approximately 4,000 proteins, representing most of the proteome of MTB strain H37Rv, were then expressed and affinity-purified using GST tags and glutathione agarose beads using the protocol described by Zhu et al. (Zhu et al., 2001). After elution and quality control by western blotting with an anti-GST antibody and capillary electrophoresis, each protein was spotted twice on polymer slides (Polymer-Slide H; CapitalBio) using a SmartArrayer 48 microarrayer (CapitalBio). The resulting microarrays were stored at −80°C prior to use.

Global Profiling of PknG-Protein Interactions

MTB proteome microarrays were blocked for 1 hr at room temperature with shaking in blocking buffer (1xPBS, 3% BSA, 0.1% Tween 20 [pH 7.4]). After blocking, 200 μl purified V5-tagged PknG at a final concentration of 0.05 μg/μl in blocking buffer was loaded onto the microarray, which was then covered using a Liftslip (Grace Bio-Labs) during incubation. Arrays were incubated in a humidified chamber for 1 hr at 4°C without shaking, then washed three times in Tris-buffered saline solution containing 0.1% Tween 20 detergent (TBST).

Arrays were probed with 1 ml mouse anti-V5 antibody (1 μg/ml in blocking buffer; Sigma-Aldrich) and incubated in a humidified chamber for 1 hr at room temperature. After washing three times in TBST, arrays were probed with 1 ml cy5-donkey-anti-mouse antibody (1 μg/ml in blocking buffer; Jackson ImmunoResearch Laboratories) and incubated in a humidified chamber for 1 hr at room temperature. After washing three times in TBST, arrays were dried in a SlideWasher (CapitalBio) and then scanned with a GenePix 4200A microarray scanner (Molecular Devices). Data were analyzed with GenePix Pro 6.0 (Molecular Devices).

Global Profiling of c-di-GMP-Protein Interactions

MTB proteome microarrays were blocked for 1 hr at room temperature with shaking in blocking buffer (3% BSA, 1xTBST, and 0.1% Tween 20 [pH 7.4]). The 1 mM biotin-c-di-GMP (BioLog Life Science Institute) and 1 mM biotin (Sigma-Aldrich) solutions were prepared in 3 ml PBS and added respectively to deep-well plates containing a blocked microarray, ensuring that the microarray was covered during probing. After incubating overnight at 4°C with shaking, arrays were washed three times with shaking in 1xTBST and then probed with 3 ml Cy3-strepavadin (1 μg/mL in 1xTBST buffer; Sigma-Aldrich) for 1 hr at room temperature with shaking. After washing three times in 1xTBST and once in ddH_2O, arrays were dried in a SlideWash (Capital Bio) and then scanned in a GenePix 4200A slider scanner (Molecular Devices). Data were analyzed with GenePix Pro 6.0 (Molecular Devices).

Protein Microarray Data Analysis of PknG and c-di-GMP Interactions

Data were extracted from the microarray images by GenePix Pro 6.0 (Molecular Devices) and then processed using an R package developed in our lab (http://www.protein--microarray.com). To generate candidate lists, the signal to noise ratio (SNR) was defined as F635 median/B635 median. The SNR of each protein was averaged for the two duplicated spots on each microarray incubated with (SNR+) or without (SNR−) PknG/c-di-GMP. Proteins with positive signals were identified on each array based on the assumption that background values across the array followed a normal distribution. The median SNR of all proteins was used to estimate the mean of the normal distribution, and the median absolute deviation (MAD) was used to estimate the SD of the normal distribution. Proteins with p values < 0.01 were considered as positive interactors. To identify which proteins had significantly higher SNRs in the microarray incubated with PknG/c-di-GMP than in the control microarray, calling score was defined as SNR+/SNR− and calculated for all the proteins, assuming that the calling scores of proteins whose signal was not significantly different between the control microarray and the microarray incubated with PknG/c-di-GMP followed a normal distribution. Cutoff values were set as calling score >2 and SNR >3 for PknG-interacting proteins and Calling score >3 and SNR >3 for c-di-GMP-interacting proteins.

Candidate PknG and c-di-GMP binding proteins were classified using the PANTHER classification system (Mi et al., 2013) with default settings. Gene symbols were used as input.

Candidate c-di-GMP binding proteins were examined for motif enrichment using Dilimot, a linear motif-predicting tool (Neduva et al., 2005). Default settings were used, except that the maximum length for motifs was set at 8.

Serum Biomarker Identification

Serum samples from TB patients and recovered individuals who met the eligibility criteria were collected at Guangdong Center for Tuberculosis Control between 2011 and 2012. The study was approved by the Ethics Committee of the Guangdong Center for Tuberculosis Control and informed consent was obtained from each individual participating in the study before sampling. Samples were grouped according to the disease status of the donor into two phases of MTB infection, i.e., patients diagnosed with TB before treatment for active disease who were smear positive, and recovered individuals who were smear positive before treatment and became smear negative after successful completion of antibiotic treatment.

Each serum sample (3 ml, diluted 1:100 in TBST) was overlaid on a protein microarray and incubated at room temperature for 1 hr. After washing three times with TBST, bound serum antibodies were detected by incubation for 1 hr at room temperature with Alexa Fluor 532 goat anti-human immunoglobulin G (H^+L) (Jackson ImmunoResearch Laboratories) diluted 1:1,000 in TBST. After washing with TBST and drying in a SlideWasher (CapitalBio), arrays were scanned in a GenePix 4200A (Molecular Devices). The median foreground and background fluorescence intensity were obtained for each spot in the protein microarrays using GenePix 6.0 (Molecular Devices). The raw intensity of each spot was defined as the ratio of the foreground to the background median intensity. To remove the negative effects of nonspecific binding and spatial heterogeneity across the protein array, the background processing method "normexp" (Ritchie et al., 2007) was used for background correction. The NormalizeBetweenArrays package (Bolstad et al., 2003) in "limma" was employed for quantile normalization.

Twenty-one samples from the "patient" and 20 samples from the "recovered" groups were used as an initial training set to identify differentially expressed proteins using the Bioconductor package in R 3.0. Normalized protein expression values were transformed to log2 values, and all proteins on the protein microarrays were ordered by their fold-change values. To validate the differentially expressed proteins identified using the training set, the above analysis was repeated on two test sets (64 and 104 samples from the "patient" and 43 and 87 samples from the "recovered" groups in set 1 and set 2, respectively). Fisher's test was employed to calculate p values for each protein candidate, which were further adjusted for multiple testing using the Benjamini and Hochberg method to control the false discovery rate. The p value cutoff was set to 0.05. The "vioplot" and "Daim" packages in R 3.1.0 were used to generate Violin plots and to calculate the sensitivity, specificity, and area under the curve for the differentially expressed proteins.

Sample size was defined as $ss = Z2p(1 - p)/c2$, where c is the margin of error (also called the confidence interval) that can be tolerated, p is the percentage of samples chosen, and Z is a constant that is dependent on the confidence interval (e.g., Z = 1.96 for a 95% confidence interval).

Details of other experimental procedures are described in the Supplemental Experimental Procedures.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, two figures, three tables, and four data sets and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2014.11.023.

AUTHOR CONTRIBUTIONS

X.-E.Z., S.-c.T., J.D., and L.B. conceived and designed the project; L.Z., Q.Z., T.C., and H.L. provided key materials; S.-j.G., R.-p.D., Z.-x.W., Jia Gu, Z.L., H.-w.J., Jing Gu, W.L., L.C., J. Gao, Z.Z., X.W., F.G., Y.L., G.Z., D.W., and H.H. constructed the microarray; H.-n.Z., F.-l.W., Y.Z., X.H., and H.-w.J. performed the experiments; S.-c.T., X.-E.Z., J.D., L.B., J.-f.W., and J.F. analyzed the data; S.-c.T., J.F., X.-E.Z., J.D., and L.B. wrote and revised the manuscript.

ACKNOWLEDGMENTS

We thank the PFGRC for kindly providing the MTB entry clone collection used in the construction of the proteome microarray and Dr. Lei Feng of the Instrumental Analysis Center of Shanghai Jiao Tong University for her kind help with LC-MS analysis. We also thank Yu Xue for help with motif analysis and Yiming Zhou for help with microarray data analysis. We are grateful to Dan Czajkowsky for helpful discussions on the manuscript. This work was supported by grants from the Key Project Specialized for Infectious Diseases of the Chinese Ministry of Health (2013ZX10003006 and 2012ZX10003002) to L.B., the Chinese Academy of Sciences (KJZD-EW-TZ-L04) to X.-E.Z., the National High Technology Research and Development Program of China (2012AA020103 and 2012AA020203), and the National Natural Science Foundation of China (31370813) to S.-c.T. We are grateful for financial support from the Guangdong Province Program for the Introduction of Innovative R&D teams (2013S024) and the Science and Technology Innovation Fund Project of Foshan City (2013HT100103).

Received: July 3, 2014
Revised: September 17, 2014
Accepted: November 17, 2014
Published: December 11, 2014

REFERENCES

Abdiche, Y.N., Malashock, D.S., Pinkerton, A., and Pons, J. (2009). Exploring blocking assays using Octet, ProteOn, and Biacore biosensors. Anal. Biochem. *386*, 172–180.

Albrethsen, J., Agner, J., Piersma, S.R., Højrup, P., Pham, T.V., Weldingh, K., Jimenez, C.R., Andersen, P., and Rosenkrands, I. (2013). Proteomic profiling

of Mycobacterium tuberculosis identifies nutrient-starvation-responsive toxin-antitoxin systems. Mol. Cell. Proteomics *12*, 1180–1191.

Behr, M.A., Wilson, M.A., Gill, W.P., Salamon, H., Schoolnik, G.K., Rane, S., and Small, P.M. (1999). Comparative genomics of BCG vaccines by whole-genome DNA microarray. Science *284*, 1520–1523.

Bolstad, B.M., Irizarry, R.A., Astrand, M., and Speed, T.P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics *19*, 185–193.

Boshoff, H.I., and Lun, D.S. (2010). Systems biology approaches to understanding mycobacterial survival mechanisms. Drug Discov. Today Dis. Mech. *7*, e75–e82.

Brennan, P.J. (2003). Structure, function, and biogenesis of the cell wall of Mycobacterium tuberculosis. Tuberculosis (Edinb.) *83*, 91–97.

Burdette, D.L., Monroe, K.M., Sotelo-Troha, K., Iwig, J.S., Eckert, B., Hyodo, M., Hayakawa, Y., and Vance, R.E. (2011). STING is a direct innate immune sensor of cyclic di-GMP. Nature *478*, 515–518.

Chen, C.S., Korobkova, E., Chen, H., Zhu, J., Jian, X., Tao, S.C., He, C., and Zhu, H. (2008). A proteome chip approach reveals new DNA damage recognition activities in Escherichia coli. Nat. Methods *5*, 69–74.

Chen, Y., Yang, L.N., Cheng, L., Tu, S., Guo, S.J., Le, H.Y., Xiong, Q., Mo, R., Li, C.Y., Jeong, J.S., et al. (2013). Bcl2-associated athanogene 3 interactome analysis reveals a new role in modulating proteasome activity. Mol. Cell. Proteomics *12*, 2804–2819.

Comas, I., Coscolla, M., Luo, T., Borrell, S., Holt, K.E., Kato-Maeda, M., Parkhill, J., Malla, B., Berg, S., Thwaites, G., et al. (2013). Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. Nat. Genet. *45*, 1176–1182.

Farhat, M.R., Shapiro, B.J., Kieser, K.J., Sultana, R., Jacobson, K.R., Victor, T.C., Warren, R.M., Streicher, E.M., Calver, A., Sloutsky, A., et al. (2013). Genomic analysis identifies targets of convergent positive selection in drug-resistant Mycobacterium tuberculosis. Nat. Genet. *45*, 1183–1189.

Finn, R.D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J.E., Gavin, O.L., Gunasekaran, P., Ceric, G., Forslund, K., et al. (2010). The Pfam protein families database. Nucleic Acids Res. *38* (Database issue), D211–D222.

Galagan, J.E., Minch, K., Peterson, M., Lyubetskaya, A., Azizi, E., Sweet, L., Gomes, A., Rustad, T., Dolganov, G., Glotova, I., et al. (2013). The Mycobacterium tuberculosis regulatory network and hypoxia. Nature *499*, 178–183.

Gnjatic, S., Ritter, E., Büchler, M.W., Giese, N.A., Brors, B., Frei, C., Murray, A., Halama, N., Zörnig, I., Chen, Y.T., et al. (2010). Seromic profiling of ovarian and pancreatic cancer. Proc. Natl. Acad. Sci. USA *107*, 5088–5093.

Griffin, J.E., Gawronski, J.D., Dejesus, M.A., Ioerger, T.R., Akerley, B.J., and Sassetti, C.M. (2011). High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. PLoS Pathog. *7*, e1002251.

Grzegorzewicz, A.E., Ma, Y., Jones, V., Crick, D., Liav, A., and McNeil, M.R. (2008). Development of a microtitre plate-based assay for lipid-linked glycosyltransferase products using the mycobacterial cell wall rhamnosyltransferase WbbL. Microbiology *154*, 3724–3730.

Gu, S., Chen, J., Dobos, K.M., Bradbury, E.M., Belisle, J.T., and Chen, X. (2003). Comprehensive proteomic profiling of the membrane constituents of a Mycobacterium tuberculosis strain. Mol. Cell. Proteomics *2*, 1284–1296.

Hengge, R. (2009). Principles of c-di-GMP signalling in bacteria. Nat. Rev. Microbiol. *7*, 263–273.

Hill, P.C., Brookes, R.H., Fox, A., Jackson-Sillah, D., Jeffries, D.J., Lugos, M.D., Donkor, S.A., Adetifa, I.M., de Jong, B.C., Aiken, A.M., et al. (2007). Longitudinal assessment of an ELISPOT test for Mycobacterium tuberculosis infection. PLoS Med. *4*, e192.

Huang, J., Zhu, H., Haggarty, S.J., Spring, D.R., Hwang, H., Jin, F., Snyder, M., and Schreiber, S.L. (2004). Finding new components of the target of rapamycin (TOR) signaling network through chemical genetics and proteome chips. Proc. Natl. Acad. Sci. USA *101*, 16594–16599.

Huang, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

Jaroszewicz, J., Calle Serrano, B., Wursthorn, K., Deterding, K., Schlue, J., Raupach, R., Flisiak, R., Bock, C.T., Manns, M.P., Wedemeyer, H., and Cornberg, M. (2010). Hepatitis B surface antigen (HBsAg) levels in the natural history of hepatitis B virus (HBV)-infection: a European perspective. J. Hepatol. *52*, 514–522.

Jeong, J.S., Jiang, L., Albino, E., Marrero, J., Rho, H.S., Hu, J., Hu, S., Vera, C., Bayron-Poueymiroy, D., Rivera-Pacheco, Z.A., et al. (2012). Rapid identification of monospecific monoclonal antibodies using a human proteome microarray. Mol. Cell. Proteomics *11*, 016253.

Kung, L.A., Tao, S.C., Qian, J., Smith, M.G., Snyder, M., and Zhu, H. (2009). Global analysis of the glycoproteome in Saccharomyces cerevisiae reveals new roles for protein glycosylation in eukaryotes. Mol. Syst. Biol. *5*, 308.

Kunnath-Velayudhan, S., and Porcelli, S.A. (2013). Recent advances in defining the immunoproteome of mycobacterium tuberculosis. Front Immunol *4*, 335.

Kunnath-Velayudhan, S., Salamon, H., Wang, H.Y., Davidow, A.L., Molina, D.M., Huynh, V.T., Cirillo, D.M., Michel, G., Talbot, E.A., Perkins, M.D., et al. (2010). Dynamic antibody responses to the Mycobacterium tuberculosis proteome. Proc. Natl. Acad. Sci. USA *107*, 14703–14708.

Lee, Y.I., Giovinazzo, D., Kang, H.C., Lee, Y., Jeong, J.S., Doulias, P.T., Xie, Z., Hu, J., Ghasemi, M., Ischiropoulos, H., et al. (2014). Protein microarray characterization of the S-nitrosoproteome. Mol. Cell. Proteomics *13*, 63–72.

Li, W., and He, Z.G. (2012). LtmA, a novel cyclic di-GMP-responsive activator, broadly regulates the expression of lipid transport and metabolism genes in Mycobacterium smegmatis. Nucleic Acids Res. *40*, 11292–11307.

Lienhardt, C., Glaziou, P., Uplekar, M., Lönnroth, K., Getahun, H., and Raviglione, M. (2012). Global tuberculosis control: lessons learnt and future prospects. Nat. Rev. Microbiol. *10*, 407–416.

Lin, Y.Y., Lu, J.Y., Zhang, J., Walter, W., Dang, W., Wan, J., Tao, S.C., Qian, J., Zhao, Y., Boeke, J.D., et al. (2009). Protein acetylation microarray reveals that NuA4 controls key metabolic target regulating gluconeogenesis. Cell *136*, 1073–1084.

Liu, L., Zhang, W.J., Zheng, J., Fu, H., Chen, Q., Zhang, Z., Chen, X., Zhou, B., Feng, L., Liu, H., and Jin, Q. (2014). Exploration of novel cellular and serological antigen biomarkers in the ORFeome of Mycobacterium tuberculosis. Mol. Cell. Proteomics *13*, 897–906.

Lu, K.Y., Tao, S.C., Yang, T.C., Ho, Y.H., Lee, C.H., Lin, C.C., Juan, H.F., Huang, H.C., Yang, C.Y., Chen, M.S., et al. (2012). Profiling lipid-protein interactions using nonquenched fluorescent liposomal nanovesicles and proteome microarrays. Mol. Cell. Proteomics *11*, 1177–1190.

Målen, H., Berven, F.S., Fladmark, K.E., and Wiker, H.G. (2007). Comprehensive analysis of exported proteins from Mycobacterium tuberculosis H37Rv. Proteomics *7*, 1702–1718.

McFadden, J., Beste, D.J.V., and Kierzek, A.M. (2013). Systems Biology of Tuberculosis (New York: Springer).

Mehaffy, M.C., Kruh-Garcia, N.A., and Dobos, K.M. (2012). Prospective on Mycobacterium tuberculosis proteomics. J. Proteome Res. *11*, 17–25.

Mi, H., Muruganujan, A., and Thomas, P.D. (2013). PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Res. *41* (Database issue), D377–D386.

Navarro, M.V., De, N., Bae, N., Wang, Q., and Sondermann, H. (2009). Structural analysis of the GGDEF-EAL domain-containing c-di-GMP receptor FimX. Structure *17*, 1104–1116.

Neduva, V., Linding, R., Su-Angrand, I., Stark, A., de Masi, F., Gibson, T.J., Lewis, J., Serrano, L., and Russell, R.B. (2005). Systematic discovery of new recognition peptides mediating protein interaction networks. PLoS Biol. *3*, e405.

Newman, R.H., Hu, J., Rho, H.-S., Xie, Z., Woodard, C., Neiswinger, J., Cooper, C., Shirley, M., Clark, H.M., Hu, S., et al. (2013). Construction of human activity-based phosphorylation networks. Mol. Syst. Biol. *9*, 655.

O'Hare, H.M., Durán, R., Cerveñansky, C., Bellinzoni, M., Wehenkel, A.M., Pritsch, O., Obal, G., Baumgartner, J., Vialaret, J., Johnsson, K., and Alzari, P.M. (2008). Regulation of glutamate metabolism by protein kinases in mycobacteria. Mol. Microbiol. *70*, 1408–1423.

Prasad, T.S., Verma, R., Kumar, S., Nirujogi, R.S., Sathe, G.J., Madugundu, A.K., Sharma, J., Puttamallesh, V.N., Ganjiwale, A., Myneedu, V.P., et al. (2013). Proteomic analysis of purified protein derivative of Mycobacterium tuberculosis. Clin. Proteomics *10*, 8.

Prisic, S., Dankwa, S., Schwartz, D., Chou, M.F., Locasale, J.W., Kang, C.M., Bemis, G., Church, G.M., Steen, H., and Husson, R.N. (2010). Extensive phosphorylation with overlapping specificity by Mycobacterium tuberculosis serine/threonine protein kinases. Proc. Natl. Acad. Sci. USA *107*, 7521–7526.

Ritchie, M.E., Silver, J., Oshlack, A., Holmes, M., Diyagama, D., Holloway, A., and Smyth, G.K. (2007). A comparison of background correction methods for two-colour microarrays. Bioinformatics *23*, 2700–2707.

Römling, U., Galperin, M.Y., and Gomelsky, M. (2013). Cyclic di-GMP: the first 25 years of a universal bacterial second messenger. Microbiol. Mol. Biol. Rev. *77*, 1–52.

Scherr, N., Honnappa, S., Kunz, G., Mueller, P., Jayachandran, R., Winkler, F., Pieters, J., and Steinmetz, M.O. (2007). Structural basis for the specific inhibition of protein kinase G, a virulence factor of Mycobacterium tuberculosis. Proc. Natl. Acad. Sci. USA *104*, 12151–12156.

Schubert, O.T., Mouritsen, J., Ludwig, C., Röst, H.L., Rosenberger, G., Arthur, P.K., Claassen, M., Campbell, D.S., Sun, Z., Farrah, T., et al. (2013). The Mtb proteome library: a resource of assays to quantify the complete proteome of Mycobacterium tuberculosis. Cell Host Microbe *13*, 602–612.

Sester, M., Sotgiu, G., Lange, C., Giehl, C., Girardi, E., Migliori, G.B., Bossink, A., Dheda, K., Diel, R., Dominguez, J., et al. (2011). Interferon-γ release assays for the diagnosis of active tuberculosis: a systematic review and meta-analysis. Eur. Respir. J. *37*, 100–111.

Shin, J.H., Yang, J.Y., Jeon, B.Y., Yoon, Y.J., Cho, S.N., Kang, Y.H., Ryu, H., and Hwang, G.S. (2011). (1)H NMR-based metabolomic profiling in mice infected with Mycobacterium tuberculosis. J. Proteome Res. *10*, 2238–2247.

Steingart, K.R., Ng, V., Henry, M., Hopewell, P.C., Ramsay, A., Cunningham, J., Urbanczik, R., Perkins, M.D., Aziz, M.A., and Pai, M. (2006). Sputum processing methods to improve the sensitivity of smear microscopy for tuberculosis: a systematic review. Lancet Infect. Dis. *6*, 664–674.

Studier, F.W., and Moffatt, B.A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol. *189*, 113–130.

Székely, R., Wáczek, F., Szabadkai, I., Németh, G., Hegymegi-Barakonyi, B., Eros, D., Szokol, B., Pató, J., Hafenbradl, D., Satchell, J., et al. (2008). A novel drug discovery concept for tuberculosis: inhibition of bacterial and host cell signalling. Immunol. Lett. *116*, 225–231.

Tiwari, D., Singh, R.K., Goswami, K., Verma, S.K., Prakash, B., and Nandicoori, V.K. (2009). Key residues in Mycobacterium tuberculosis protein kinase G play a role in regulating kinase activity and survival in the host. J. Biol. Chem. *284*, 27467–27479.

Walburger, A., Koul, A., Ferrari, G., Nguyen, L., Prescianotto-Baschong, C., Huygen, K., Klebl, B., Thompson, C., Bacher, G., and Pieters, J. (2004). Protein kinase G from pathogenic mycobacteria promotes survival within macrophages. Science *304*, 1800–1804.

Wang, X., Hessner, M.J., Wu, Y., Pati, N., and Ghosh, S. (2003). Quantitative quality control in microarray experiments and the application in data filtering, normalization and false positive rate prediction. Bioinformatics *19*, 1341–1347.

WHO (World Health Organization) (2013). Global Tuberculosis Report 2013 (Geneva: WHO Press).

Yew, W.W., Lange, C., and Leung, C.C. (2011). Treatment of tuberculosis: update 2010. Eur. Respir. J. *37*, 441–462.

Zhang, H., Li, D., Zhao, L., Fleming, J., Lin, N., Wang, T., Liu, Z., Li, C., Galwey, N., Deng, J., et al. (2013). Genome sequencing of 161 Mycobacterium tuberculosis isolates from China identifies genes and intergenic regions associated with drug resistance. Nat. Genet. *45*, 1255–1260.

Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., Houfek, T., et al. (2001). Global analysis of protein activities using proteome chips. Science *293*, 2101–2105.

Zhu, J., Gopinath, K., Murali, A., Yi, G., Hayward, S.D., Zhu, H., and Kao, C. (2007). RNA-binding proteins that inhibit RNA virus infection. Proc. Natl. Acad. Sci. USA *104*, 3129–3134.