Dynamic antibody responses to the *Mycobacterium tuberculosis* proteome

Shajo Kunnath-Velayudhan[a,1], Hugh Salamon[b,1], Hui-Yun Wang[a,1,2], Amy L. Davidow[c], Douglas M. Molina[d], Vu T. Huynh[d], Daniela M. Cirillo[e], Gerd Michel[f], Elizabeth A. Talbot[f,g], Mark D. Perkins[f], Philip L. Felgner[d,h,3], Xiaowu Liang[d], and Maria L. Gennaro[a,4]

[a]Public Health Research Institute, New Jersey Medical School, Newark, NJ 07103; [b]AbaSci LLC, Albany, CA 94706; [c]Department of Preventive Medicine & Community Health, New Jersey Medical School, Newark, NJ 07101; [d]Antigen Discovery Inc., Irvine, CA 92618; [e]Emerging Bacterial Pathogens Unit, San Raffaele Scientific Institute, 20132 Milan, Italy; [f]Foundation for Innovative New Diagnostics, 1202 Geneva, Switzerland; [g]Infectious Diseases and International Health, Dartmouth Hitchcock Medical Center, Lebanon, NH 03756; and [h]Department of Medicine, University of California, Irvine, CA 92697

Considerable effort has been directed toward controlling tuberculosis, which kills almost two million people yearly. High on the research agenda is the discovery of biomarkers of active tuberculosis (TB) for diagnosis and for monitoring treatment outcome. Rational biomarker discovery requires understanding host–pathogen interactions leading to biomarker expression. Here we report a systems immunology approach integrating clinical data and bacterial metabolic and regulatory information with high-throughput detection in human serum of antibodies to the entire *Mycobacterium tuberculosis* proteome. Sera from worldwide TB suspects recognized approximately 10% of the bacterial proteome. This result defines the *M. tuberculosis* immunoproteome, which is rich in membrane-associated and extracellular proteins. Additional analyses revealed that during active tuberculosis (*i*) antibody responses focused on an approximately 0.5% of the proteome enriched for extracellular proteins, (*ii*) relative target preference varied among patients, and (*iii*) responses correlated with bacillary burden. These results indicate that the B cell response tracks the evolution of infection and the pathogen burden and replicative state and suggest functions associated with B cell-rich foci seen in tuberculous lung granulomas. Our integrated proteome-scale approach is applicable to other chronic infections characterized by diverse antibody target recognition.

biomarkers | humoral immunity | protein microarray | granuloma

Genome-scale research has tremendously expanded our knowledge of infectious diseases (1). The availability of microbial genome sequences has made possible probing the evolution of microorganisms, identifying ecological niches, and interpreting infection outbreaks. Functional genomics has advanced structure-to-function investigation of pathogens and helped characterize virulence mechanisms. Moreover, transcriptomics has revealed effects of infection on host and pathogen gene expression, thereby elucidating host–pathogen interactions. In contrast, proteomics has been more a tool for large-scale identification of proteins than for investigating proteome biology. The study of the antibody response can advance how proteomics is used, because it investigates antibody–antigen interactions and also evaluates the effects on antibody responses of pathogen and host characteristics.

Antibody responses are typically investigated in infectious diseases where antibody production strongly affects pathogenesis and outcome (2). But antibodies are also produced in infections, such as those caused by intracellular bacteria and some fungi (3), where protective immunity is primarily elicited by the cell-mediated response. In the case of tuberculosis (TB), a global infectious disease caused by the intracellular bacterium *Mycobacterium tuberculosis* (http://www.who.int), detection of circulating antibody in patients with active TB dates back to 1898 (4). Although an estimated approximately 90% of TB patients produce antibody to *M. tuberculosis* proteins (5), little is known about the correlation between antibody production, antibody specificity, and disease process. It appears that, in addition to canonical lymphoid organs, B cells can

encounter antigen in ectopic B cell aggregates associated with tuberculous granulomas (6, 7). Therefore, although the correlation between the specificity of B cell receptors expressed in the infected lung and the circulating antibodies is still undefined, conditions exist for the antibody response to closely track the evolution of disease.

A vast literature has been generated on circulating antibodies in TB patients (8–10), with the goal of evaluating them as biomarkers of active disease. Thus, many antibody targets are known. However, because antibody profiles vary from one TB patient to another (5, 11, 12), as seen also in other chronic infections such as those caused by *Helicobacter pylori* (13) and *Pseudomonas aeruginosa* (14), the investigation of antibody targets during TB has been open-ended and largely unproductive. Moreover, host-associated variables affecting antibody responses have been only sporadically explored in small-scale studies (15, 16). Thus, despite much effort, we still do not know how much of the *M. tuberculosis* proteome is targeted by the human antibody response or how host characteristics and disease parameters affect target recognition. Given the person-to-person variability, satisfactory answers to these fundamental questions, which are critical for effective biomarker discovery, require interrogating the entire proteome of *M. tuberculosis* with large numbers of sera.

Here we report the results of a systems-level analysis of the antibody response to the entire *M. tuberculosis* proteome in diseased humans. We integrated proteome-scale antibody measurements obtained with a high-throughput protein microarray platform (17) and more than 500 TB suspects' sera collected at various sites worldwide with epidemiological and clinical parameters and bacterial protein class information. We found that, during active TB, the humoral immune response (*i*) shifts focus from membrane-associated to extracellular proteins of the tubercle bacillus and (*ii*) correlates with bacillary burden. These

IMMUNOLOGY

results link antibody levels with the evolution of the infection and the replicative state of the pathogen.

Results

***M. Tuberculosis* Proteome Microarrays.** To assess antibody responses to the *M. tuberculosis* proteome, we used a high-throughput proteome microarray technology (17). The microarrays carried 4,099 *M. tuberculosis*-protein spots, which corresponded to more than 99% of the ORFs in *M. tuberculosis* H$_{37}$Rv DNA (http://genolist.pasteur.fr/TuberculList/). Full-length *M. tuberculosis* protein was detected in more than 95% of the spots, as assessed by monoclonal antibody reactivity to epitopes fused to recombinant protein termini (*SI Appendix,* Fig. S1). The amount of protein varied among the spots (*SI Appendix,* Fig. S1); this was taken into account by our analytical strategy.

When proteome microarrays were tested for antibody binding, signal intensity at positive control spots varied with antigen and antibody concentration (*SI Appendix,* Fig. S2) as expected for immunoassays. A similar conclusion was reached when microarrays were probed with sera from infected mice (*SI Appendix,* Fig. S3). We also determined that measurements were highly reproducible (the coefficient of variation was less than 5% for 99% of the spots; *SI Appendix,* Fig. S4). Thus, our proteome microarrays detect antibody binding reliably and can be used as a discovery platform.

Data-Driven Analytical Approach. Proteome microarrays were probed with sera from more than 500 patients with TB disease or non-TB disease (NTBD) diagnosis who were enrolled as TB suspects under a uniform study protocol at various sites worldwide (*Materials and Methods,* *SI Appendix,* Fig. S5, and *SI Appendix,* Tables S1 and S2). We chose this population because (*i*) rigorous criteria were used to exclude active TB in the NTBD group; (*ii*) the multiple study sites ensured geographic diversity; and (*iii*) the NTBD group exhibited various characteristics associated with *M. tuberculosis* exposure, including origin from TB-endemic areas, latent *M. tuberculosis* infection, and past history of TB. Arrays were also probed with negative control sera collected in a nonendemic setting (Italy) from 64 healthy persons who tested negative for latent *M. tuberculosis* infection (LTBI) to generate a negative control distribution for each protein. When intensity distributions for each *M. tuberculosis*-protein spot were analyzed, the following observations were made. First, the range of signal intensities differed among spots (*SI Appendix,* Fig. S6*A*), which is presumably due, at least in part, to different amounts of protein per spot. Second, most proteins showed only subtle differences between intensity distributions obtained with active TB sera and those obtained with NTBD sera (*SI Appendix,* Fig. S6*B*). Third, some proteins exhibited a high-tailed intensity distribution

in active TB sera relative to NTBD (Fig. 1*A*). Together, the above observations led to an analytical approach whereby each protein spot was considered to represent an independent assay; moreover, tests that assess high-tail intensities, such as Z statistics, were favored over those assessing central tendency. Thus, serum reactivity was assigned to each protein by calculating Z-scores relative to the negative control distribution for the same protein. The associated *P* values were corrected for multiple testing by calculating false discovery rate (fdr) (Fig. 1*B*). A reactivity call was made when the fdr was ≤0.01.

Immunoproteome. With the analytical criteria described above, 484 proteins were recognized by serum from at least one patient (Fig. 2*A* and listed in *SI Appendix,* Table S3). Among the reactive proteins, a subset was frequently recognized, whereas much of the response was rare [286 proteins (59%) reacted to single serum samples]. The latter finding prompted two additional analyses. First, we tested whether the discovery of rarely reactive proteins was spurious. When we probed arrays with an additional, independent set of approximately 300 active TB sera, we found that the chance of rediscovering singly-reacting proteins was very high ($P = 4.7 \times 10^{-8}$), indicating that these were legitimate reactivities. Second, we evaluated whether the number of serum samples tested was sufficient to discover all reactive proteins. We used Monte Carlo simulations to assess the effect of sample size on the number of proteins discovered. We found that the proteome was not saturated by the number of sera used for array probing (*SI Appendix,* Fig. S7). However, the number of new discoveries per serum decreased as the number of sera increased (*SI Appendix,* Fig. S7). We calculated that the chance of failing to identify a protein reactive to 1% of sera was lower than half a percent [$(1.0 - 0.01) ^ 534 \sim 0.005$]. Thus, our proteome screen identified all of the frequently recognized targets on the platform.

We next asked how the presence of active TB affected antibody responses. When we stratified protein reactivity by final TB diagnosis (active TB vs. NTBD), we observed that the most reactive proteins were predominantly recognized by active TB sera (Fig. 2*B*). A significant association with active TB by odds ratio calculation was found for only 13 of these proteins ("TB-associated proteins," Table 1). Of these 13 proteins, which have been identified by proteomic studies as components of purified protein derivative and/or *M. tuberculosis* extracts, most have been reported as B cell and/or T cell antigens (Table 1).

Collectively, the above results, which are summarized in Fig. 3, show that approximately one-tenth of the proteome was recognized by sera from TB suspects, irrespective of TB diagnosis. This result defines the immunoproteome of *M. tuberculosis*. Although additional proteins could be discovered by testing a larger number
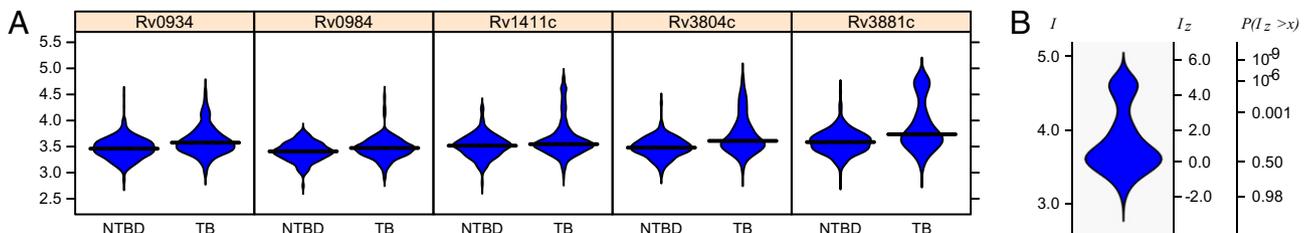

Fig. 1. Estimation of significant serum reactivity. The *M. tuberculosis* proteome arrays were probed with sera from 561 TB suspects [TB (*n* = 254) and non-TB disease (NTBD; *n* = 307)]. (A) Examples of proteins showing a second, high-tail distribution in the active TB sera. The distributions of protein spot intensities obtained with each protein are shown as violin plots. One plot represents one protein. In these plots, the *y* axis represents the log-transformed measurements of microarray spot intensity; the *x* axis represents the number of observations; the width of the plot is determined by a kernel density estimation (www.itl.nist.gov); the median of each distribution is shown as a horizontal black bar. All violin plot areas are equal. (B) The method used to define a reactivity call, i.e., the assignment of significant reactivity of one serum to one protein, is depicted using a hypothetical violin plot. I = Log-transformed intensity measurements; I_z = Z-scores (i.e., the distance from the mean of the reference intensity distribution, in units of SD); P (I_z > x) = *P* value associated with Z-score. As described in the text, *P* values were corrected for multiple testing by calculating false discovery rates (fdr) for each serum.
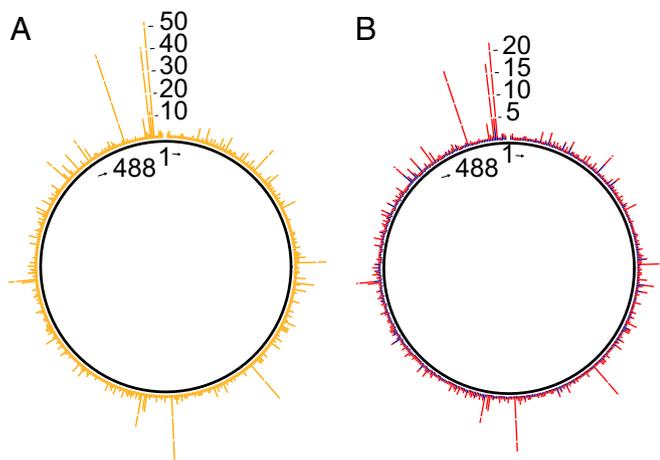
Kunnath-Velayudhan et al.

Fig. 2. The immunoproteome of *M. tuberculosis*. Each of the 484 reactive proteins is mapped onto a circular plot with the corresponding Rv (ID) numbers increasing clockwise, starting at position 1. One bar represents one protein. The bar location on the circular plot closely approximates that of the corresponding gene on the genome. (*A*) Number of reactive TB suspects' sera per protein; (*B*) The same as *A* but stratified as percent of active TB sera (red) and NTBD sera (blue) reacting with each protein.

of sera, our analyses indicate that it is exceedingly unlikely that frequently reactive proteins were not discovered. The results also show that, during active TB, the immune response focuses on a much smaller number of targets with increased frequency. Target focusing in active TB was also detected with other analytical methods, although the number of target antigens varied somewhat with the method used.

To understand why particular proteins were targeted by antibody in TB suspects' sera and why the response was focused on a subset of targets during active TB, we interrogated the clinical and epidemiological characteristics of the study population and assessed the nature of the *M. tuberculosis* proteins targeted by the immune response. The results of these analyses are reported below.

Table 1. Proteins associated with active TB

Protein ID*	Annotation‡
Rv3881c	Secreted antigen EspB
Rv3804c†	Secreted antigen 85A/mycolyltransferase
Rv3874†	Secreted antigen Cfp10/EsxB/MTSA-10
Rv1860†	Secreted glycoprotein 45–47 kDa antigen/MPT32
Rv1411c†	Lipoprotein LprG
Rv2031c†	16kDa antigen/alpha-crystallin
Rv0934	Glycolipoprotein 38kDa antigen/PstS1
Rv3616c	Conserved hypothetical protein
Rv3864	Conserved hypothetical protein
Rv1980c†	Secreted antigen MPT64
Rv0632c†	Enoyl-CoA hydratase/isomerase superfamily
Rv1984c†	Secreted antigen Cfp21
Rv2873†	Surface lipoprotein antigen MPT83

*Rv3874 is encoded in the RD1 region (i.e., it is absent in all bacillus Calmette–Guérin strains), whereas Rv1980c and Rv1984c are found in the RD2 region (i.e., they are absent in some bacillus Calmette–Guérin strains) (38).

†Identifies proteins found in the purified protein derivative of *Mycobacterium bovis* (39). All these proteins were identified by proteomics (http://web.mpiib-berlin.mpg.de/cgi-bin/pdbs/2d-page/extern/index.cgi and refs. 40–44). All proteins in the table, except Rv3864, were identified as B cell and/or T cell antigens (10, 36, 45–48).

‡Annotations were adapted from the Sanger Institute database (http://www.sanger.ac.uk/Projects/M_tuberculosis/Gene_list/).

Antibody Response and Characteristics of the NTBD Patients. We investigated the potential sources of antibody reactivity in NTBD patients' sera. One possibility was that NTBD seroreactivity was caused by exposure to nontuberculous mycobacteria (NTM), which produce proteins cross-reactive with *M. tuberculosis*. If so, one would expect that *M. tuberculosis* proteins reactive with NTBD sera would be more enriched than those reactive with TB sera for proteins sharing high similarity with NTM proteins. We tested this possibility (*SI Appendix, SI Materials and Methods*) and found no such enrichment (*P* = 0.90; *SI Appendix*, Fig. S8). Thus, our data do not support the possibility that NTBD reactivity is predominantly caused by NTM exposure, a covariate that is exceedingly difficult to assess directly due to the ubiquitous presence of these bacteria in the environment (18).

We next investigated associations between antibody reactivity and demographic and clinical characteristics of NTBD patients. We used hurdle regression, which models both presence of reactivity to any protein (logistic regression component) and number of reactive proteins (Poisson regression component). Univariate analysis identified chest X-ray status (indicating radiographic lung abnormalities, such as cavitation) and geographic region of origin as variables to be included in a multivariate model (*SI Appendix*, Table S4). The results showed that chest X-ray status was associated with antibody responses in NTBD subjects (*SI Appendix*, Table S5). An abnormal chest X-ray in the NTBD group may be a surrogate for past TB because NTBD subjects with past TB history were more likely to have an abnormal than a normal chest X-ray (odds ratio = 3.9, *P* < 0.001). History of past TB and abnormal chest X-ray together explained 35% of NTBD reactors, as per attributable risk calculation. We also found that 8 of the 10 NTBD patients reacting to the 13 TB-associated proteins had past history of TB (whereas the overall frequency of past TB in NTBD subjects was only 16%) (models could not be estimated because of the small number of reactors). Together, the above results point to history of past TB as a key host covariate of antibody responses in NTBD patients.

Antibody Response and Characteristics of the Active TB Patients. We assessed the relationship between antibody response and clinical and demographic characteristics of active TB patients (*SI Appendix*, Tables S6 and S7 for univariate analysis). The multivariate model included sputum smear grade (indicating bacillary counts in sputum samples) and chest X-ray status. The presence of a cavitary chest X-ray made active TB patients approximately twice as likely to respond to the immunoproteome or to the 13 TB-associated proteins (Table 2). Smear grade was only modestly associated with the response to the immunoproteome, but it showed a strong association with the response to the 13 TB-associated proteins (Table 2). The association became stronger with increasing smear grade (test for trend *P* = 0.03) (Table 2). These results indicate that increasing bacillary burden had little effect on the response to the immunoproteome as a whole, but it greatly increased the response to the 13 TB-associated proteins. We next interrogated individual patients' sera to assess the relative preference for each of these 13 proteins. We observed diverse antibody patterns (Fig. 4). Thus, within the pool of TB-associated proteins, target preference varied among patients.

Characteristics of Reactive Proteins. We investigated the subcellular localization of proteins in the immunoproteome. This is a key property of bacterial antigen because it affects both accessibility by the immune response and antibody function (8). We found that the immunoproteome as a whole was enriched for extracellular proteins (*P* < 0.001) and membrane proteins (*P* < 0.01) but not for cytoplasmic proteins (*SI Appendix, SI Materials and Methods*). Moreover, when only proteins reacting to more than one serum were analyzed, the enrichment for extracellular proteins (*P* < 0.001) was maintained but that for membrane proteins was lost
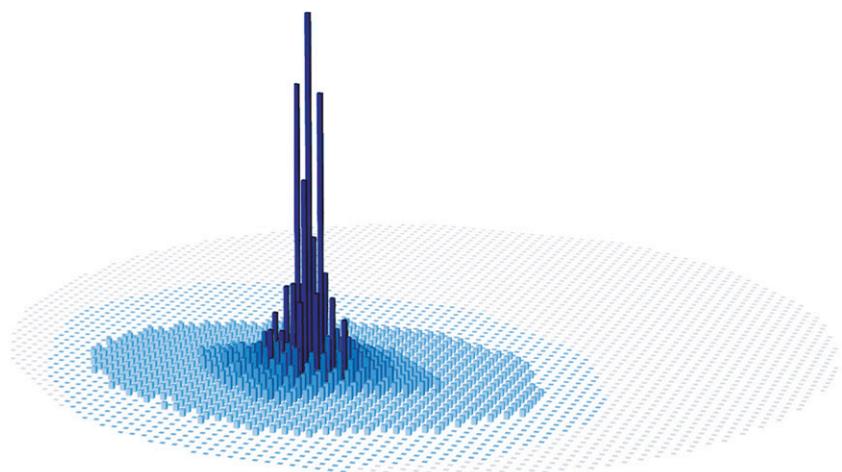
IMMUNOLOGY

Fig. 3. Reactivity of the immunoproteome. Each bar represents one of the 4,000 proteins of *M. tuberculosis*. The bar height and blue color gradient reflect the number of reactive sera to each protein. The lightest, outer portion of the proteome (flat white bars) represents non-reactive proteins (approximately 90% of the proteome). The raised bars of any hue of blue represent almost 500 reactive proteins, which define the immunoproteome (approximately 10% of the proteome). The flat, light-blue bars located in between represent an arbitrary number of rarely reactive proteins that may have been missed in the screen. Within the immunoproteome, the dark blue, tallest bars are the proteins showing statistical association with active TB status (approximately 0.5% of the proteome).

($P = 0.08$). Thus, the more frequently reacting proteins were found predominantly in the extracellular fraction of the tubercle bacillus, whereas rare reactivity predominantly pertained to membrane proteins. Accordingly, the subset of 13 TB-associated proteins was rich in proteins annotated as secreted (Table 1).

To further explore the characteristics of proteins preferentially reacting with active TB sera, we adopted protein class analysis. This method, which is typically used with gene-expression data-sets (19), increases statistical power by using data gathered for entire protein classes rather than for single proteins. When the serological data were analyzed against 258 metabolic pathways in the BioCyc database, the 16 proteins in "TCA cycle variation 1" (but not those in "TCA cycle") showed significant association with TB status (Fig. 5). The variant differs from the canonical TCA cycle by the inclusion of the glyoxylate bypass and the α-ketoglutarate-to-succinate pathway (20), which are found in bacteria but not in humans. Further analysis showed that only the glyoxylate bypass proteins were significantly associated with active TB (Fig. 5). Increased reactivity to the glyoxylate bypass proteins was also visible in a standard heat map (*SI Appendix*, Fig. S9). One protein in this pathway, malate synthase, was previously reported as antibody target during preclinical and clinical tuberculosis (21–

23). The same protein was found associated to the cell-surface and in bacterial culture supernatants (24). Thus, the observed reactivity of other proteins in the glyoxylate bypass suggests that also these proteins may be surface-associated, as has been shown for other metabolic enzymes in Gram-positive bacteria (25).

When classifications based on gene regulation were used, we found associations between antibody responses in active TB and proteins encoded by FurB(Zur)-regulated genes (26) (Fig. 5). This regulon is enriched for genes encoding proteins in the ESAT-6 (ESX-1) family and its dedicated secretion system (27). Together with the result that two of the TB-associated antigens are Rv3881c (ESX-1 secretion system), which is also secreted (28), and Rv3784 (ESX-1 homolog) (Table 1), the data point to the ESX protein family as antibody targets during active TB. Collectively, results of the protein class analysis support preferential antibody recognition of extracellular proteins in active TB.

Discussion

The reactivity of more than 500 sera from TB suspects worldwide divides the proteome of *M. tuberculosis* in three groups. The first, small group, which comprises roughly 0.5% of the proteome, in-

Table 2. Multivariate analysis of reactivity in active TB patients

Predictors*	Immunoproteome			TB-associated proteins		
	Odds ratio	P values	P value for trend	Odds ratio	P values	P value for trend
Sputum smear grade						
Negative[†]	1			1		
1+	0.99	0.97	0.06	1.83	0.12	0.03
2+	1.05	0.91		2.60	0.03	
3+/4+	1.46	0.38		3.36	< 0.01	
Chest X-ray						
Noncavitary[†‡]	1			1		
Cavitary	2.01	0.06	N.A.[§]	1.92	0.04	N.A.

Shown are the results of the logistic component of the multivariate analysis. The Poisson component showed no associations between antibody responses and sputum smear grade or chest X-ray and is not shown.
*All effects were adjusted for age.
[†]Reference group.
[‡]Includes abnormal/noncavitary ($n = 122$) and normal ($n = 8$).
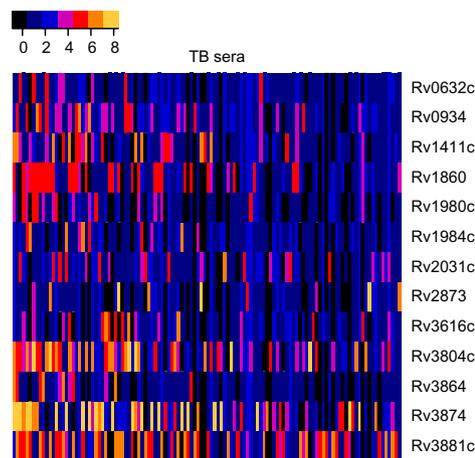[§]N.A., not applicable.

Fig. 4. Patterns of serum reactivity to 13 proteins in active TB. The heat map shows reactivity of sera from active TB patients to each of 13 TB-associated proteins. Shown are only active TB sera reacting to at least one of the 13 proteins ($n = 118$). Each column represents one serum and each row represents one TB-associated protein. Normalized signal intensities (Z-scores), are visualized as a color spectrum, as indicated.
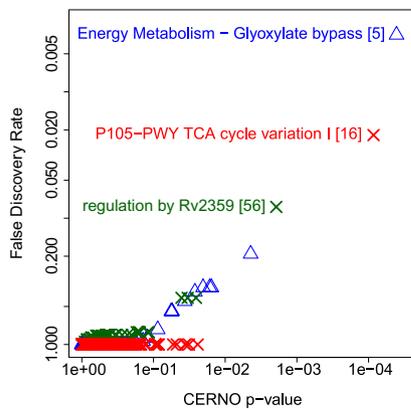
Kunnath-Velayudhan et al.

Fig. 5. Protein classes associated with serum reactivity in active TB. Results of protein class analysis are shown. False discovery rate correction (fdr) (*y* axis) is plotted as a function of CERNO *P* values (*x* axis) for each protein class tested. Protein classes are shown as per the BioCyc ontology (red symbols), Sanger Institute ontology (blue symbols), and the MtbNet regulon classification (green symbols). Annotations are shown only for protein classes having fdr < 0.1. Numbers in parentheses indicate the number of proteins in each class (listed in *SI Appendix, Appendix 1*).

cludes antigens that are recognized at high frequency only during active TB. This group is enriched for extracellular proteins. A second, larger group includes additional, frequently recognized antigens as well as proteins that are scored rarely. Rare targets are rich in membrane-associated proteins and are recognized by both active TB and NTBD sera. Together, the first two groups of proteins define the immunoproteome, which is approximately one-tenth the size of the total proteome. The remaining, largest group of the proteome contains proteins that do not react with human sera. It may also comprise, albeit at vanishingly small probability, additional rare reactivities that were missed by our screen. Thus, the present, proteome-wide serological interrogation defines the immunoproteome of *M. tuberculosis*. We surmise that additional, frequent targets of the antibody response might only be found if they correspond to particular proteins or antibodies missing from our screen [due to technical limitations, see below], or, perhaps, if the *M. tuberculosis* proteome was interrogated with large numbers of sera representing a very specific population or geographic region.

Integration of proteome-wide antibody data with host characteristics and protein class information showed that the antibody response tracks the load and metabolic activity of tubercle bacilli during infection. Within the immunoproteome, the rare targets are predominantly membrane-associated proteins, whereas the frequently recognized proteins tend to be extracellular. We propose that, given an overall preference of the immune response for membrane-bound over soluble antigen, membrane-associated proteins (which might derive from low numbers of live bacilli, dead bacilli, or be found in macrophage-secreted exosomes) (29) are occasionally targeted during latent infection or paucibacillary disease. In either condition, the extracellular proteins are underrepresented, either because dormant bacilli do not secrete (latent infection) or because the numbers of metabolically active (and secreting) mycobacteria are low (paucibacillary disease). As bacillary burden increases with disease, metabolically active bacilli secrete proteins, which become favored targets. This scenario is supported by the association seen in active TB between the response to the TB-associated proteins (but not to the immunoproteome as a whole) and sputum smear grade, which is a surrogate of bacillary burden. Thus, the composition of the immunoproteome and its reactivity during active TB show that the antibody responses are dynamic and well reflect progression of the infection. Physical evidence of the proximity between

granulomas, B cell maturation foci, and bloodstream in the tuberculous lung of humans and mice (6, 7) supports this view.

Why does relative target preference within the immunodominant antigen subset vary from one patient to another? The varied antibody profiles seen in TB patients make it difficult to exclude host characteristics. However, it is likely that the correlation between antigen load and serum reactivity heavily marks each patient's antibody profile, with variability being introduced by bacillary load and metabolic state (30) at the time of testing, relative antigen expression of the infecting bacterial strain (31), and relative immunodominance of each protein. Thus, the relative frequency by which the antibody response "samples" each immunodominant antigen will vary from one patient to another. Moreover, depending on relative antibody avidity, the effect of antigen load on frequency of sampling will be greater for some antigens than for others. Because the above events occur over considerable periods of time, due to the chronic nature of tuberculosis, cross-sectional analyses of those events may accentuate variation. Indeed, varied target recognition by antibody is typically seen in chronic infections (13, 14, 32, 33). It is also noted that, given the focused target selection during active TB and the relative high frequency at which some antigens may be recognized when bacillary burden is high, it is not surprising that antibody responses to tuberculosis may have been seen as homogeneous (34) under particular patient selection and testing conditions.

Our definition of the *M. tuberculosis* immunoproteome requires qualification. First, a small proportion of targets is not represented because the overall proteome coverage in our platform is lower than 100% (>95%). One example is the known B cell antigen ESAT-6 (Rv3875) (35), which was not detected. Second, antibodies directed against some cross-reactive epitopes present in *M. tuberculosis* and *Escherichia coli* may have been removed by the preabsorption of sera with *E. coli* lysates, which was required to minimize spurious reactivity to *E. coli* extracts in the microarray spots. Third, some conformational epitopes or epitopes associated with posttranslation modifications are missing from the screen due to the use of recombinant proteins associated with the high-throughput method. However, none of the above limitations should substantially affect the composition of the immunoproteome. In particular, highly reactive proteins typically contain multiple epitopes (36); thus, it is unlikely that the absence of some epitopes (e.g., conformational epitopes or glycosylated residues) would fully abrogate protein immunoreactivity.

In conclusion, our systems approach to the study of the antibody response in tuberculosis defines the boundaries of the immunologically reactive proteome, reveals the dynamics of the response associated with manifestation of disease, reconciles apparently divergent views over varied antibody profiles in tuberculosis patients, and provides directions for integrating aspects of cellular and humoral responses. The tenets of the approach (large-scale study and integration of host and pathogen information) should be applicable to other infections characterized by varied antibody responses where the genome of the infecting microorganism has been sequenced. Unraveling the biology of the humoral response to *M. tuberculosis* infection and disease will establish principles needed to develop effective immunodiagnostics. These principles include showing that the pool of antigens recognized during active TB is relatively small, strongly suggesting that universal biomarkers exist (the hurdle analysis found no association between geographic region and antibody reactivity of active TB patients), and establishing past TB history as a potential confounding factor.

Materials and Methods

Further details on patient characteristics, experimental methods, and data analysis are in *SI Appendix, SI Materials and Methods*.

Patient Characteristics. TB suspects were defined as patients having persistent cough (more than 3 wk) and at least another clinical feature suggestive

of active TB. Active TB was diagnosed when sputum was positive for *M. tuberculosis* culture. Diagnosis of active TB was excluded based on negative *M. tuberculosis* culture results at presentation and on mandatory, 2-mo follow-up. Negative control sera were collected in a low-endemicity setting (Italy) from healthy persons without latent *M. tuberculosis* infection, as indicated by negative results to two tests, T-SPOT.TB and QuantiFERON, and also to tuberculin skin test, when performed. All subjects were negative for HIV infection.

Microarray Fabrication and Hybridization. The protocols used for the fabrication of high-throughput protein microarrays are published (17).

Analytical Methods. Visualization and statistical analysis were performed with R statistical software (http://www.r-project.org/). Serum reactivity to a protein (reactivity call) was estimated based on Z-statistics, and multiple testing corrections were performed by fdr calculations. Hurdle regression analysis was used to identify associations between host parameters and reactivity calls. Fisher's exact test was used to identify associations between protein reactivity and subcellular localization, as predicted with the LocateP algorithm. The Coincident Extreme Ranks in Numerical Observations (CERNO)

method was used to identify associations between proteins classes and antibody responses in active TB (37).

1. Yang X, Yang H, Zhou G, Zhao GP (2008) Infectious disease in the genomic era. *Annu Rev Genomics Hum Genet* 9:21–48.
2. Janeway CA, Travers P, Walport M, Shlomchik M (2001) *Immunobiology* (Garland Publishing, New York).
3. Casadevall A, Pirofski LA (2003) Antibody-mediated regulation of cellular immunity and the inflammatory response. *Trends Immunol* 24:474–478.
4. Arloing S (1898) Agglutination de bacille de la tuberculose vraie. *Les Comptes Rendus de l'Academie des Sciences* 126:1398–1400.
5. Lyashchenko K, et al. (1998) Heterogeneous antibody responses in tuberculosis. *Infect Immun* 66:3936–3940.
6. Ulrichs T, et al. (2004) Human tuberculous granulomas induce peripheral lymphoid follicle-like structures to orchestrate local host defence in the lung. *J Pathol* 204:217–228.
7. Maglione PJ, Xu J, Chan J (2007) B cells moderate inflammatory progression and enhance bacterial containment upon pulmonary challenge with Mycobacterium tuberculosis. *J Immunol* 178:7222–7234.
8. Bothamley GH, Gennaro ML (2008) in *Handbook of Tuberculosis: Immunology and Cell Biology*, eds Kaufmann SHE, Britton W. (Wiley-VCH, Weinheim), pp 227–244.
9. Steingart KR, et al. (2007) A systematic review of commercial serological antibody detection tests for the diagnosis of extrapulmonary tuberculosis. *Postgrad Med J* 83:705–712.
10. Steingart KR, et al. (2009) Performance of purified antigens for serodiagnosis of pulmonary tuberculosis: A meta-analysis. *Clin Vaccine Immunol* 16:260–276.
11. Wu X, et al. (2010) Humoral immune responses against M. tuberculosis 38-kDa, MTB48 and CFP-10/ESAT-6 antigens in tuberculosis. *Clin Vaccine Immunol* 17:372–375.
12. Julián E, Matas L, Alcaide J, Luquin M (2004) Comparison of antibody responses to a potential combination of specific glycolipids and proteins for test sensitivity improvement in tuberculosis serodiagnosis. *Clin Diagn Lab Immunol* 11:70–76.
13. Haas G, et al. (2002) Immunoproteomics of Helicobacter pylori infection and relation to gastric disease. *Proteomics* 2:313–324.
14. Wehmhöner D, et al. (2003) Inter- and intraclonal diversity of the Pseudomonas aeruginosa proteome manifests within the secretome. *J Bacteriol* 185:5807–5814.
15. Davidow A, et al. (2005) Antibody profiles characteristic of Mycobacterium tuberculosis infection state. *Infect Immun* 73:6846–6851.
16. Silva VM, Kanaujia G, Gennaro ML, Menzies D (2003) Factors associated with humoral response to ESAT-6, 38 kDa and 14 kDa in patients with a spectrum of tuberculosis. *Int J Tuberc Lung Dis* 7:478–484.
17. Davies DH, et al. (2005) Profiling the humoral immune response to infection by using proteome microarrays: High-throughput vaccine and diagnostic antigen discovery. *Proc Natl Acad Sci USA* 102:547–552.
18. Petrini B (2006) Non-tuberculous mycobacterial infections. *Scand J Infect Dis* 38:246–255.
19. Curtis RK, Oresic M, Vidal-Puig A (2005) Pathways to the analysis of microarray data. *Trends Biotechnol* 23:429–435.
20. Tian J, Bryk R, Itoh M, Suematsu M, Nathan C (2005) Variant tricarboxylic acid cycle in Mycobacterium tuberculosis: Identification of alpha-ketoglutarate decarboxylase. *Proc Natl Acad Sci USA* 102:10670–10675.
21. Laal S, et al. (1997) Surrogate marker of preclinical tuberculosis in human immunodeficiency virus infection: Antibodies to an 88-kDa secreted antigen of Mycobacterium tuberculosis. *J Infect Dis* 176:133–143.
22. Gennaro ML, et al. (2007) Antibody markers of incident tuberculosis among HIV-infected adults in the USA: A historical prospective study. *Int J Tuberc Lung Dis* 11:624–631.
23. Hendrickson RC, et al. (2000) Mass spectrometric identification of mtb81, a novel serological marker for tuberculosis. *J Clin Microbiol* 38:2354–2361.
24. Kinhikar AG, et al. (2006) Mycobacterium tuberculosis malate synthase is a laminin-binding adhesin. *Mol Microbiol* 60:999–1013.

25. Pancholi V, Fischetti VA (1992) A major surface protein on group A streptococci is a glyceraldehyde-3-phosphate-dehydrogenase with multiple binding activity. *J Exp Med* 176:415–426.
26. Maciag A, et al. (2007) Global analysis of the Mycobacterium tuberculosis Zur (FurB) regulon. *J Bacteriol* 189:730–740.
27. Porcelli SA (2008) Tuberculosis: Shrewd survival strategy. *Nature* 454:702–703.
28. McLaughlin B, et al. (2007) A mycobacterium ESX-1-secreted virulence factor with unique requirements for export. *PLoS Pathog* 3:e105.
29. Russell DG (2007) Who puts the tubercle in tuberculosis? *Nat Rev Microbiol* 5:39–47.
30. Garton NJ, et al. (2008) Cytological and transcript analyses reveal fat and lazy persister-like bacilli in tuberculous sputum. *PLoS Med* 5:e75.
31. Pheiffer C, Betts J, Lukey P, van Helden P (2002) Protein expression in Mycobacterium tuberculosis differs with growth stage and strain type. *Clin Chem Lab Med* 40:869–875.
32. Eberhardt C, et al. (2009) Proteomic analysis of the bacterial pathogen Bartonella henselae and identification of immunogenic proteins for serodiagnosis. *Proteomics* 9:1967–1981.
33. Beare PA, et al. (2008) Candidate antigens for Q fever serodiagnosis revealed by immunoscreening of a Coxiella burnetii protein microarray. *Clin Vaccine Immunol* 15:1771–1779.
34. Samanich K, Belisle JT, Laal S (2001) Homogeneity of antibody responses in tuberculosis patients. *Infect Immun* 69:4600–4609.
35. Brusasca PN, et al. (2001) Immunological characterization of antigens encoded by the RD1 region of the Mycobacterium tuberculosis genome. *Scand J Immunol* 54:448–452.
36. Wiker HG, et al. (1998) Immunochemical characterization of the MPB70/80 and MPB83 proteins of Mycobacterium bovis. *Infect Immun* 66:1445–1452.
37. Yamaguchi KD, et al. (2008) IFN-beta-regulated genes show abnormal expression in therapy-naïve relapsing-remitting MS mononuclear cells: Gene expression analysis employing all reported protein-protein interactions. *J Neuroimmunol* 195:116–120.
38. Behr MA, et al. (1999) Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* 284:1520–1523.
39. Borsuk S, Newcombe J, Mendum TA, Dellagostin OA, McFadden J (2009) Identification of proteins from tuberculin purified protein derivative (PPD) by LC-MS/MS. *Tuberculosis (Edinb)* 89:423–430.
40. Mawuenyega KG, et al. (2005) Mycobacterium tuberculosis functional network analysis by global subcellular protein profiling. *Mol Biol Cell* 16:396–404.
41. Sinha S, et al. (2002) Proteome analysis of the plasma membrane of Mycobacterium tuberculosis. *Comp Funct Genomics* 3:470–483.
42. Fortune SM, et al. (2005) Mutually dependent secretion of proteins required for mycobacterial virulence. *Proc Natl Acad Sci USA* 102:10676–10681.
43. Desouza GA, et al. (2010) Using a label-free proteomic method to identify differentially abundant proteins in closely related hypo- and hyper-virulent clinical Mycobacterium tuberculosis Beijing isolates. *Mol Cell Proteomics* (in press).
44. Målen H, Berven FS, Fladmark KE, Wiker HG (2007) Comprehensive analysis of exported proteins from Mycobacterium tuberculosis H37Rv. *Proteomics* 7:1702–1718.
45. Sartain MJ, Slayden RA, Singh KK, Laal S, Belisle JT (2006) Disease state differentiation and identification of tuberculosis biomarkers via native antigen array profiling. *Mol Cell Proteomics* 5:2102–2113.
46. Bigi F, et al. (1997) A novel 27 kDa lipoprotein antigen from Mycobacterium bovis. *Microbiology* 143:3599–3605.
47. Mustafa AS, et al. (2006) Immunogenicity of Mycobacterium tuberculosis antigens in Mycobacterium bovis BCG-vaccinated and M. bovis-infected cattle. *Infect Immun* 74:4566–4572.
48. Wang BL, et al. (2005) Antibody response to four secretory proteins from Mycobacterium tuberculosis and their complex antigen in TB patients. *Int J Tuberc Lung Dis* 9:1327–1334.

Kunnath-Velayudhan et al.