A Systems Biology Approach for Diagnostic and Vaccine Antigen Discovery in Tropical Infectious Diseases

Li Liang and **Philip L. Felgner**

Department of Medicine, Division of Infectious Diseases, University of California, Irvine, CA 92697, USA

Abstract

Purpose of review: There is a need for improved diagnosis and for more rapidly assessing the presence, prevalence and spread of newly emerging or re-emerging infectious diseases. An approach to the pathogen-detection strategy is based on analyzing host immune response to the infection. This review focuses on a protein microarray approach for this purpose.

Recent findings: Here we take a protein microarray approach to profile the humoral immune response to numerous infectious agents, and to identify the complete antibody repertoire associated with each disease. The results of these studies lead to the identification of diagnostic markers and potential subunit vaccine candidates. These results from over 30 different organisms can also provide information about common trends in the humoral immune response.

Summary: Systems biology approach to identify the antibody repertoire associated with infectious diseases challenge using protein microarray has become a powerful method in identifying diagnostic markers and potential subunit vaccine candidates, and moreover, in providing information on proteomic feature (functional and physically properties) of seroreactive and serodiagnostic antigens. Combining the detection of the pathogen with a comprehensive assessment of the host immune response will provide a new understanding of the correlations between specific causative agents, the host response, and the clinical manifestations of the disease.

Keywords

Introduction

A major component of the adaptive immune response to infection is the generation of protective and long-lasting humoral immunity. Analyses of antibody responses against different infectious agents are critical for diagnosing infectious diseases, understanding pathogenic mechanisms, and the development and evaluation of vaccines. Protein microarrays are well suited to identify, quantify and compare individual antigenic responses

Corresponding author: Philip Felgner, Department of Medicine, Division of Infectious Diseases, University of California, Irvine, CA 92697, USA. pfelgner@uci.edu; Phone: 949-824-1407; Fax: 949-824-9675.

following exposure to infectious agents. It can now evaluate antibody responses to hundreds, or even thousands, of recombinant antigens at one time. These large-scale studies have uncovered new antigenic targets, provided new insights into vaccine research and yielded an overview of immunoreactivity against almost the entire proteome of certain pathogens. This technology can be applied to the development of improved serodiagnostic tests, discovery of subunit vaccine antigen candidates, epidemiologic research and vaccine development, as well as providing novel insights into infectious disease and the immune system. In this review, we will discuss the use of protein microarrays as a powerful tool to define the humoral immune response to bacteria and viruses.

Factors governing selection of the particular antigens recognized are unclear [1,2]. It is not uncommon for viruses encoding a small number of proteins to generate antibodies against each encoded protein. But for infectious agents containing hundreds or thousands of proteins only a subset of the proteome is recognized and little is known about the extent or the characterisitics of this subset of antigens. Methods for making a complete empirical accounting of the immunoproteome have limitations, particularly when the genome of the organism is large. The Protein Microarray Laboratory at UC Irvine has developed a highly efficient method to determine the humoral immune response to microbial antigens. We have applied this approach to more than 30 medically important infectious microorganisms [3–33] including *M. tuberculosis*[33], *Plasmodium falciparum*[5,8,24], *Plasmodium vivax*, *Brucella melitensis[*14], *Chlamydia trachomatis*[3,25], *Francisella tularensis*[11,23], *Burkholderia pseudomallei*[6,19], *Coxiella burnetii*[7,26], *Borrelia burgdorferi*[10], *Salmonella enterica* typhi, *Rickettsia prowazekii, Rickettsia rickettsii, Orientia tsutsugamushi, Bartonella henselae*[17], *Leptospira interrogans, Toxoplasma gondii*[27], *Candida albicans*[28], *Schistosoma mansoni* [4] and viruses including vaccinia[9,29–31], monkeypox, Herpes 1 & 2, Varicella zoster, HPV[32], HIV, Dengue, influenza, West Nile and Chikungunya. Since launching this project 10 years ago we have made more than 40,000 plasmids, printed the encoded proteins on 25,000 microarrays and probed the arrays with 15,000 serum specimens in order to determine disease associated antibody profiles in people infected with each agent. These chips can be probed with sera from infected subjects to determine the immunodominant antigens for each agent and the methodology is amenable to the screening of sera from very large cohorts numbering in the thousands. When seroreactive and serodiagnostic antigen subsets from different infectious agents are printed onto the same array, the chip can discriminate between subjects infected with different agents and also identify individuals with co-infections or multiple infections. We have shown that the individual proteins printed on these arrays capture antibodies present in serum from infected individuals and the amount of captured antibody can be quantified using fluorescent secondary antibody. In this way a comprehensive profile of antibodies that result after infection or exposure can be determined that is characteristic of the type of infection and the stage of disease [9,10,31].

Here we summarize the approximate seroreactive and serodiagnostic antigens that were identified and published in 30 different organisms, and discuss the antibody response predictions from classification of reactive antigens based on functional and physical properties.

Protein Microarray Production, probing and analysis

Genes were amplified and cloned using a high-throughput PCR and recombination method[29]. ORFs from genomic DNA or cDNA were identified and amplified using gene specific primers containing about 20 bp nucleotide extension complementary to ends of linearized pXT7 vector, which allows homologous recombination between the PCR product and pXT7 vector in competent *E.coli* cells. The resulting fusion proteins also harbored a hemagglutinin epitope at 3′ end and polyhistidine at the 5′ end. Plasmids were expressed at 24°C in a 16 hour- *in vitro* transcription/translation *E. coli* system (expressway kits from Invitrogen). For no DNA controls, no plasmid DNA was added to the same amount of reagent from *in vitro* transcription/translation *E.coli* system to test *E.coli* background reactivity. For microarrays, 10 μl of reaction was mixed with 3.3 μl 0.2% Tween 20 to give a final concentration of 0.05% Tween 20, and printed onto nitrocellulose coated glass FAST slides (Whatman) using an Omni Grid 100 microarray printer (Genomic Solutions). Sera samples were diluted in *E. coli* lysate (Mclab). Slides were incubated in biotin-conjugated secondary antibody (Jackson ImmunoResearch) and detected by incubation with streptavidin-conjugated SureLight® P-3 (Columbia Biosciences). Microarray slides were scanned and analyzed using a Perkin Elmer ScanArray Express HT or Genepix microarray scanner. Intensities were quantified. All signal intensities were corrected for spot-specific background. All foreground values were transformed and normalized using robust linear model (RLM) or nonlinear variance stabilizing normalization (VSN) to remove systematic effects [24,34,35](Figure 1).

Percent of seroreactive antigens

Discovery of novel antigens associated with infectious diseases is fundamental to the development of serodiagnostic tests and protein subunit vaccines against existing and emerging pathogens. Through over ten years of effort, we have identified over 1000 antigens associated with infections or vaccinations in 30 different organisms (Table 1)[3–6,9–17,23,25,31–33,36–47], accounting for around 2–5% of bacterial genome; 20–57% of viral genome; and 10–45% of parasite genomes. Antigens differentially reactive among infected and healthy controls comprise even smaller percentage of the genome size: from 0.3% to 3% for bacteria; 16–40% for viruses and 2–18% for parasites. *Borrelia burgdorferi*, however, generate higher antibody responses against ~15% of polypeptides during natural infection, of which half are differentially reactive between naturally infected and uninfected individuals [10].

Antigens were classified as 'seroreactive' with mean reactivity greater than 2–3 standard deviations above the mean of the negative controls in most organisms; and differentially reactive antigens are classified by BH adjusted p value smaller than 0.05 by comparing the negative group with infected or vaccinated individuals.

Full proteome microarrays were constructed for only a limited number of bacterial species (ref), however, other data were published using partial arrays containing only partial proteome, and may overrepresent the percentages of seroreactive and serodiagnostic

antigens in the full proteome because the subset of proteins on the array were selected based on antigenic features seen previously.

Enrichment analysis reveals physical properties and cellular functions associated with immunogenicity

Another application for this empirical data is to train an algorithm to predict reactive antigens *in silico*, and several papers from our group apply enrichment analyses to identify proteomic features that tend to be seen more frequently in the seroreactive and serodiagnostic antigen sets [12,17,23].

Efforts to predict antigenicity have relied on a few computational algorithms predicting signal peptide sequences (signalP), transmembrane domains, or subcellular localization (Psort). The current database from this protein microarray approach contains quantitative antibody reactivity data against 40,000 proteins derived from 30 infectious microorganisms and more than 30 million data points derived from 15,000 patient sera. Interrogation of these data sets has revealed more than 10 proteomic features that are associated with antigenicity allowing an *in silico* protein sequence and functional annotation based approach to triage the least likely antigenic proteins from those that are more likely to be antigenic.

These proteomic enrichment features (Table 2) are: i) functionally annotated COGs (U, M, N and O) or Gene ontology (GO) function and process; ii) computationally predicted features (TMHMM, Signal peptide, pSort Outermembrane, pSort Periplasmic, and pI<5 for bacteria and pI 7–9 for parasites), and iii) abundance of expression. This approach applied to *Brucella melitensis* predicts 37% of the bacterial proteome containing 91% of the antigens empirically identified by probing proteome microarrays [12]. In *Salmonella enterica* and Leptospira.

Parasite *toxoplasma gondii* proteins were assigned by GO functions. Proteins involved in protein binding, catalytic activity, transporter activity, transferase activity were significantly enriched [13]. Proteins with enzymatic activity other than kinase activity were enriched at 2.0 fold, and enzyme regulator activity, structural molecule activity and ion channel activity were also highly enriched. Proteins with GO null functions, or involved in nucleotide and nucleic acid binding were underrepresented [13].

Proteins were also assigned by GO process classification. Proteins involved in ATP biosynthetic process were enriched. Several proteins involved in transport were also significantly enriched: ion transport, protein transport, vesicle mediated transport, and other transport functions. Proteins involved in metabolic process, proteolysis, and signal peptide processing were also enriched. Conversely, proteins not assigned with GO process categories were significantly underrepresented (0.5 fold; p value 3.301E-21) [13].

An examination with the Pf proteins on the microarray based on Gene ontological analysis revealed that ~40% of the immunogenic proteins are expressed in the membrane of the parasite or host erythrocyte and that they are overrepresented in the biological process

categories of "pathogenesis," "cytoadherence to microvasculature," "antigenic variation," and "rosetting" [5].

The data set of Vaccinia viral proteins also allowed us to identify properties of viral proteins that were associated with immunogenicity (). We found that membrane and core proteins, proteins with late or early/late temporal expression, and proteins with transmembrane domains were overrepresented in the immunoreactive antigen set relative to the whole proteome. These predictors are strongest in MVA profiles, since the antibody profile to MVA is more heavily skewed toward structural proteins. In contrast, early proteins were underrepresented relative to the whole proteome, and there was negligible influence of molecular weight, isoelectric point, or the presence of a signal sequence on immunogenicity. vaccinia antigens are either abundant components of MV particles, such as A10 and L4 [48], or are expressed at high levels in infected cells, such as I1 and WR148 [49,50]. Their abundance may contribute to immunogenicity once released from infected cells, particularly if, like D13 [51], such pro-teins have a propensity for self-assembly into macro-molecular structures.

Analysis was also done for the HSV-1 antibody profile based on GO component classifiers according to the database at www.uniprot.org. The percentage of the total number of genes assigned to each GO component present in the proteome and in the seroreactive antigens was determined, and the ratio was used to determine the fold enrichment. The analysis revealed 12 proteins on the array that were assigned the GO component virion membrane, of which 9 were seroreactive. Tegument proteins were not enriched in the seroreactive antigen set [37].

Overall, our data show that the antibody profile is not a random assortment of specificities, but strongly biased to-wards the recognition of certain proteomic features. Why we don't observe antibodies to all intracellular proteins expressed from infected cells remains unclear. It is also interesting to note that the rules that determine immunogenicity might be different from those that define protection.

Naïve bayes classification

Individual proteomic features provide some information about the likelihood of a protein being seroreactive; however, using all of these features together leads to a better segregation of the hits from the rest of the proteome. To analyze the relationship between all of these features and the seroreactivity of the proteins in a rigorous manner, we used a naïve Bayes formulation [52].

We applied a naïve Bayes classification approach to assign a relative numerical score to each antigen in the *Brucella melitensis* (Bm) proteome. This score reflects the relative likelihood that a protein will be reactive based on its functionally annotated or computationally predicted features. Our analyses indicates that 91% of serodiagnostic antigens are predictable from the top 20% of the genome ranked by this naïve Bayes classification approach, and the antigens with enriched features in the top 20% of the genome account for 100% of serodiagnostic antigens with these features. Without this naïve bayes classification approach, we would have to clone 37% of the genome with enriching

features to obtain 91% of serodiagnostic antigens. This analysis greatly enhances the predictive efficiency compared to previous studies, will provide a basis for targeted screens of entire proteomes based on likelihood of seroreactivity, and help determine trends in the humoral immune response to gram-negative bacteria. The same approach has been applied to *Salmonella enterica* and revealed that we would need to screen only 25% of the genome to be able to identify 72% of serodiagnostic antigens (Table 3).

Conclusions

The development of protein arrays for profiling the antibody response generated upon exposure to an infectious agent has allowed for new insight into the humoral immune response and the identification of potential subunit vaccine candidates and new diagnostics. No other existing approach can provide such a thorough perspective of the humoral immune response to infection. Moreover, it provides a systematic foundation formation on proteomic features (functional and physically properties) of seroreactive and serodiagnostic antigens. The information presented here will allow future protein microarray screening to focus efforts on portions of the proteome that most likely contain seroreactive proteins, and may also be useful for understanding the antibody responses to bacteria, viruses and parasites.

Acknowledgments

References

1. Mayers C, Duffield M, Rowe S, Miller J, Lingard B, Hayward S, Titball RW. Analysis of known bacterial protein vaccine antigens reveals biased physical properties and amino acid composition. Comp Funct Genomics. 2003; 4:468–478. [PubMed: 18629010]

2. Rappuoli R. Reverse vaccinology, a genome-based approach to vaccine development. Vaccine. 2001; 19:2688–2691. [PubMed: 11257410]

3. Molina DM, Pal S, Kayala MA, Teng A, Kim PJ, Baldi P, Felgner PL, Liang X, de la Maza LM. Identification of immunodominant antigens of Chlamydia trachomatis using proteome microarrays. Vaccine. 2010; 28:3014–3024. [PubMed: 20044059]

4. Driguez P, Doolan DL, Loukas A, Felgner PL, McManus DP. Schistosomiasis vaccine discovery using immunomics. Parasit Vectors. 2010; 3:4. [PubMed: 20181031]

5. Crompton PD, Kayala MA, Traore B, Kayentao K, Ongoiba A, Weiss GE, Molina DM, Burk CR, Waisberg M, Jasinskas A, et al. A prospective analysis of the Ab response to Plasmodium falciparum before and after a malaria season by protein microarray. Proc Natl Acad Sci U S A. 2010; 107:6958–6963. [PubMed: 20351286]

6. Felgner PL, Kayala MA, Vigil A, Burk C, Nakajima-Sasaki R, Pablo J, Molina DM, Hirst S, Chew JS, Wang D, et al. A Burkholderia pseudomallei protein microarray reveals serodiagnostic and cross-reactive antigens. Proc Natl Acad Sci U S A. 2009; 106:13499–13504. [PubMed: 19666533]

7. Chen C, Bouman TJ, Beare PA, Mertens K, Zhang GQ, Russell-Lodrigue KE, Hogaboam JP, Peters B, Felgner PL, Brown WC, et al. A systematic approach to evaluate humoral and cellular immune responses to Coxiella burnetii immunoreactive antigens. Clin Microbiol Infect. 2009; 15(Suppl 2): 156–157. [PubMed: 19281461]

8. Doolan DL, Mu Y, Unal B, Sundaresh S, Hirst S, Valdez C, Randall A, Molina D, Liang X, Freilich DA, et al. Profiling humoral immune responses to P. falciparum infection with protein microarrays. Proteomics. 2008; 8:4680–4694. [PubMed: 18937256]

9. Davies DH, Wyatt LS, Newman FK, Earl PL, Chun S, Hernandez JE, Molina DM, Hirst S, Moss B, Frey SE, et al. Antibody profiling by proteome microarray reveals the immunogenicity of the attenuated smallpox vaccine modified vaccinia virus ankara is comparable to that of Dryvax. J Virol. 2008; 82:652–663. [PubMed: 17977963]

10. Barbour AG, Jasinskas A, Kayala MA, Davies DH, Steere AC, Baldi P, Felgner PL. A genome-wide proteome array reveals a limited set of immunogens in natural infections of humans and white-footed mice with Borrelia burgdorferi. Infect Immun. 2008; 76:3374–3389. [PubMed: 18474646]

11. Sundaresh S, Randall A, Unal B, Petersen JM, Belisle JT, Gill Hartley M, Duffield M, Titball RW, Davies DH, Felgner PL, et al. From protein microarrays to diagnostic antigen discovery: a study of the pathogen Francisella tularensis. Bioinformatics. 2007; 23:i508–i518. [PubMed: 17646338]

12. Liang L, Tan X, Juarez S, Villaverde H, Pablo J, Nakajima-Sasaki R, Gotuzzo E, Saito M, Hermanson G, Molina D, et al. Systems Biology Approach Predicts Antibody Signature Associated with Brucella melitensis Infection in Humans. J Proteome Res. 2011; 10:4813–4824. [PubMed: 21863892]

13. Liang L, Doskaya M, Juarez S, Caner A, Jasinskas A, Tan X, Hajagos BE, Bradley PJ, Korkmaz M, Guruz Y, et al. Identification of potential serodiagnostic and subunit vaccine antigens by antibody profiling of toxoplasmosis cases in Turkey. Mol Cell Proteomics. 2011; 10:M110.006916.

14. Liang L, Leng D, Burk C, Nakajima-Sasaki R, Kayala MA, Atluri VL, Pablo J, Unal B, Ficht TA, Gotuzzo E, et al. Large scale immune profiling of infected humans and goats reveals differential recognition of Brucella melitensis antigens. PLoS Negl Trop Dis. 2010; 4:e673. [PubMed: 20454614]

15. Vigil A, Chen C, Jain A, Nakajima-Sasaki R, Jasinskas A, Pablo J, Hendrix LR, Samuel JE, Felgner PL. Profiling the humoral immune response of acute and chronic Q fever by protein microarray. Mol Cell Proteomics. 2011; 10:M110 006304. [PubMed: 21817167]

16. Kalantari-Dehaghi M, Molina DM, Farhadieh M, Morrow WJ, Liang X, Felgner PL, Grando SA. New targets of pemphigus vulgaris antibodies identified by protein array technology. Exp Dermatol. 2011; 20:154–156. [PubMed: 21255096]

17. Vigil A, Ortega R, Jain A, Nakajima-Sasaki R, Tan X, Chomel BB, Kasten RW, Koehler JE, Felgner PL. Identification of the feline humoral immune response to Bartonella henselae infection by protein microarray. PLoS One. 2010; 5:e11447. [PubMed: 20625509]

18. Yang ZR, Lertmemongkolchai G, Tan G, Felgner PL, Titball R. A genetic programming approach for Burkholderia pseudomallei diagnostic pattern discovery. Bioinformatics. 2009; 25:2256–2262. [PubMed: 19561021]

19. Tippayawat P, Saenwongsa W, Mahawantung J, Suwannasaen D, Chetchotisakd P, Limmathurotsakul D, Peacock SJ, Felgner PL, Atkins HS, Titball RW, et al. Phenotypic and functional characterization of human memory T cell responses to Burkholderia pseudomallei. PLoS Negl Trop Dis. 2009; 3:e407. [PubMed: 19352426]

20. Cannella AP, Lin JC, Liang L, Atluri V, Gotuzzo E, Felgner PL, Tsolis RM, Vinetz JM. Serial Kinetics of the Antibody Response Against the Complete Brucella melitensis ORFeome in Focal Vertebral Brucellosis. J Clin Microbiol. 2012

21. Tan X, Traore B, Kayentao K, Ongoiba A, Doumbo S, Waisberg M, Doumbo OK, Felgner PL, Fairhurst RM, Crompton PD. Hemoglobin S and C heterozygosity enhances neither the magnitude nor breadth of antibody responses to a diverse array of Plasmodium falciparum antigens. J Infect Dis. 2011; 204:1750–1761. [PubMed: 21998476]

22. Barry AE, Trieu A, Fowkes FJ, Pablo J, Kalantari-Dehaghi M, Jasinskas A, Tan X, Kayala MA, Tavul L, Siba PM, et al. The stability and complexity of antibody responses to the major surface antigen of Plasmodium falciparum are associated with age in a malaria endemic area. Mol Cell Proteomics. 2011; 10:M111 008326. [PubMed: 21825279]

23. Eyles JE, Unal B, Hartley MG, Newstead SL, Flick-Smith H, Prior JL, Oyston PC, Randall A, Mu Y, Hirst S, et al. Immunodominant Francisella tularensis antigens identified using proteome microarray. Proteomics. 2007; 7:2172–2183. [PubMed: 17533643]

24. Sundaresh S, Doolan DL, Hirst S, Mu Y, Unal B, Davies DH, Felgner PL, Baldi P. Identification of humoral immune responses in protein microarrays using DNA microarray data analysis techniques. Bioinformatics. 2006; 22:1760–1766. [PubMed: 16644788]

25**. Patton DL, Teng A, Randall A, Liang X, Felgner PL, de la Maza LM. Whole genome identification of C. trachomatis immunodominant antigens after genital tract infections and effect of antibiotic treatment of pigtailed macaques. J Proteomics. 2014; 108:99–109. This is the first time that Chlamydia trachomatis immunodominant and potential vaccine antigens have been identified in nonhuman primates following infection. [PubMed: 24862987]

26. Beare PA, Chen C, Bouman T, Pablo J, Unal B, Cockrell DC, Brown WC, Barbian KD, Porcella SF, Samuel JE, et al. Candidate antigens for Q fever serodiagnosis revealed by immunoscreening of a Coxiella burnetii protein microarray. Clin Vaccine Immunol. 2008; 15:1771–1779. [PubMed: 18845831]

27. Doskaya M, Kalantari-Dehaghi M, Walsh CM, Hiszczynska-Sawicka E, Davies DH, Felgner PL, Larsen LS, Lathrop RH, Hatfield GW, Schulz JR, et al. GRA1 protein vaccine confers better immune response compared to codon-optimized GRA1 DNA vaccine. Vaccine. 2007; 25:1824–1837. [PubMed: 17234306]

28. Mochon AB, Jin Y, Kayala MA, Wingard JR, Clancy CJ, Nguyen MH, Felgner P, Baldi P, Liu H. Serological profiling of a Candida albicans protein microarray reveals permanent host-pathogen interplay and stage-specific responses during candidemia. PLoS Pathog. 2010; 6:e1000827. [PubMed: 20361054]

29. Davies DH, Liang X, Hernandez JE, Randall A, Hirst S, Mu Y, Romero KM, Nguyen TT, Kalantari-Dehaghi M, Crotty S, et al. Profiling the humoral immune response to infection by using proteome microarrays: high-throughput vaccine and diagnostic antigen discovery. Proc Natl Acad Sci U S A. 2005; 102:547–552. [PubMed: 15647345]

30. Davies DH, McCausland MM, Valdez C, Huynh D, Hernandez JE, Mu Y, Hirst S, Villarreal L, Felgner PL, Crotty S. Vaccinia virus H3L envelope protein is a major target of neutralizing antibodies in humans and elicits protection against lethal challenge in mice. J Virol. 2005; 79:11724–11733. [PubMed: 16140750]

31. Davies DH, Molina DM, Wrammert J, Miller J, Hirst S, Mu Y, Pablo J, Unal B, Nakajima-Sasaki R, Liang X, et al. Proteome-wide analysis of the serological response to vaccinia and smallpox. Proteomics. 2007; 7:1678–1686. [PubMed: 17443847]

32. Luevano M, Bernard HU, Barrera-Saldana HA, Trevino V, Garcia-Carranca A, Villa LL, Monk BJ, Tan X, Davies DH, Felgner PL, et al. High-throughput profiling of the humoral immune responses against thirteen human papillomavirus types by proteome microarrays. Virology. 2010; 405:31–40. [PubMed: 20554302]

33. Kunnath-Velayudhan S, Salamon H, Wang HY, Davidow AL, Molina DM, Huynh VT, Cirillo DM, Michel G, Talbot EA, Perkins MD, et al. Dynamic antibody responses to the Mycobacterium tuberculosis proteome. Proc Natl Acad Sci U S A. 2010; 107:14703–14708. [PubMed: 20668240]

34. Sboner A, Karpikov A, Chen G, Smith M, Mattoon D, Freeman-Cook L, Schweitzer B, Gerstein MB. Robust-linear-model normalization to reduce technical variability in functional protein microarrays. J Proteome Res. 2009; 8:5451–5464. [PubMed: 19817483]

35. Huber W, von Heydebreck A, Sultmann H, Poustka A, Vingron M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics. 2002; 18(Suppl 1):S96–104. [PubMed: 12169536]

36*. Gerns Storey HL, Richardson BA, Singa B, Naulikha J, Prindle VC, Diaz-Ochoa VE, Felgner PL, Camerini D, Horton H, John-Stewart G, et al. Use of principal components analysis and protein microarray to explore the association of HIV-1-specific IgG responses with disease progression. AIDS Res Hum Retroviruses. 2014; 30:37–44. A collection of HIV-specific antibody responses that together were associated with reduced disease progression, by PCA and microarray analyses. [PubMed: 24134221]

37. Kalantari-Dehaghi M, Chun S, Chentoufi AA, Pablo J, Liang L, Dasgupta G, Molina DM, Jasinskas A, Nakajima-Sasaki R, Felgner J, et al. Discovery of potential diagnostic and vaccine

antigens in herpes simplex virus 1 and 2 by proteome-wide antibody profiling. J Virol. 2012; 86:4328–4339. [PubMed: 22318154]

38. Vigil A, Ortega R, Nakajima-Sasaki R, Pablo J, Molina DM, Chao CC, Chen HW, Ching WM, Felgner PL. Genome-wide profiling of humoral immune response to Coxiella burnetii infection by protein microarray. Proteomics. 2010; 10:2259–2269. [PubMed: 20391532]

39. Lessa-Aquino C, Borges Rodrigues C, Pablo J, Sasaki R, Jasinskas A, Liang L, Wunder EA Jr, Ribeiro GS, Vigil A, Galler R, et al. Identification of seroreactive proteins of Leptospira interrogans serovar copenhageni using a high-density protein microarray approach. PLoS Negl Trop Dis. 2013; 7:e2499. [PubMed: 24147173]

40**. Lessa-Aquino C, Wunder EA Jr, Lindow JC, Rodrigues CB, Pablo J, Nakajima R, Jasinskas A, Liang L, Reis MG, Ko AI, et al. Proteomic features predict seroreactivity against leptospiral antigens in leptospirosis patients. J Proteome Res. 2015; 14:549–556. This is the first full proteome Leptospira interrogans microarray that was constructed for identifying serodiagnostic antigens, and it also provides an empirical basis for predicting antigenicity from Gram negative bacteria. [PubMed: 25358092]

41. Liang L, Juarez S, Nga TV, Dunstan S, Nakajima-Sasaki R, Davies DH, McSorley S, Baker S, Felgner PL. Immune profiling with a Salmonella Typhi antigen microarray identifies new diagnostic biomarkers of human typhoid. Sci Rep. 2013; 3:1043. [PubMed: 23304434]

42. Baum E, Badu K, Molina DM, Liang X, Felgner PL, Yan G. Protein microarray analysis of antibody responses to Plasmodium falciparum in western Kenyan highland sites with differing transmission levels. PLoS One. 2013; 8:e82246. [PubMed: 24312649]

43*. Campo JJ, Aponte JJ, Skinner J, Nakajima R, Molina DM, Liang L, Sacarlal J, Alonso PL, Crompton PD, Felgner PL, et al. RTS,S vaccination is associated with serologic evidence of decreased exposure to Plasmodium falciparum liver- and blood-stage parasites. Mol Cell Proteomics. 2015; 14:519–531. These microarray data provide insight into the mechanism by which RTS,S vaccine protects from Malaria. [PubMed: 25547414]

44. Molina DM, Finney OC, Arevalo-Herrera M, Herrera S, Felgner PL, Gardner MJ, Liang X, Wang R. Plasmodium vivax pre-erythrocytic-stage antigen discovery: exploiting naturally acquired humoral responses. Am J Trop Med Hyg. 2012; 87:460–469. [PubMed: 22826492]

45*. Baum E, Sattabongkot J, Sirichaisinthop J, Kiattibutr K, Davies DH, Jain A, Lo E, Lee MC, Randall AZ, Molina DM, et al. Submicroscopic and asymptomatic Plasmodium falciparum and Plasmodium vivax infections are common in western Thailand – molecular and serological evidence. Malar J. 2015; 14:95. These microarray data are empirical data for understanding prevalance of Plasmodium in western Thailand. [PubMed: 25849211]

46*. Gaze S, Driguez P, Pearson MS, Mendes T, Doolan DL, Trieu A, McManus DP, Gobert GN, Periago MV, Correa Oliveira R, et al. An immunomics approach to schistosome antigen discovery: antibody signatures of naturally resistant and chronically infected individuals from endemic areas. PLoS Pathog. 2014; 10:e1004033. Several potentially protective and safe schistosomiasis vaccine antigens were identified. [PubMed: 24675823]

47**. Tang YT, Gao X, Rosa BA, Abubucker S, Hallsworth-Pepin K, Martin J, Tyagi R, Heizer E, Zhang X, Bhonagiri-Palsikar V, et al. Genome of the human hookworm Necator americanus. Nat Genet. 2014; 46:261–269. The authors report sequencing and assembly of the N. americanus genome, and provide an invaluable resource to boost ongoing efforts toward fundamental and applied postgenomic research. [PubMed: 24441737]

48. Chung CS, Chen CH, Ho MY, Huang CY, Liao CL, Chang W. Vaccinia virus proteome: identification of proteins in vaccinia virus intracellular mature virion particles. J Virol. 2006; 80:2127–2140. [PubMed: 16474121]

49. Patel DD, Pickup DJ, Joklik WK. Isolation of cowpox virus A-type inclusions and characterization of their major protein component. Virology. 1986; 149:174–189. [PubMed: 3456179]

50. Liu X, Kremer M, Broyles SS. A natural vaccinia virus promoter with exceptional capacity to direct protein synthesis. J Virol Methods. 2004; 122:141–145. [PubMed: 15542137]

51. Szajner P, Weisberg AS, Lebowitz J, Heuser J, Moss B. External scaffold of spherical immature poxvirus particles is made of protein trimers, forming a honeycomb lattice. J Cell Biol. 2005; 170:971–981. [PubMed: 16144903]

52. Witten, IH.; Frank, E. Data mining: practical machine learning tools and technique. 2nd. San Francisco Morgan Kaufman: Amsterdam; 2005.

Highlights

Protein microarray is a powerful tool in identifying diagnostic markers and potential subunit vaccine candidates.

Full proteome microarrays have provided valuable information on antigen predictions.

Certain proteomic feature (functional and physically properties) have been identified to be associated with seroreactive and serodiagnostic antigens

Naive Bayes classification further improves the sensitivity and specificity of this in silico predictive algorithm.
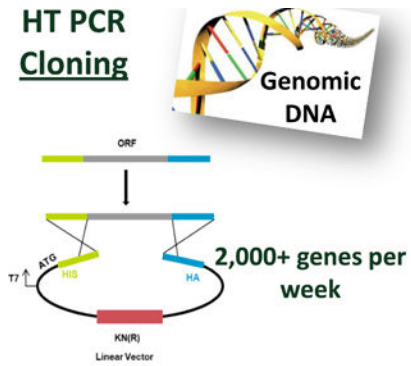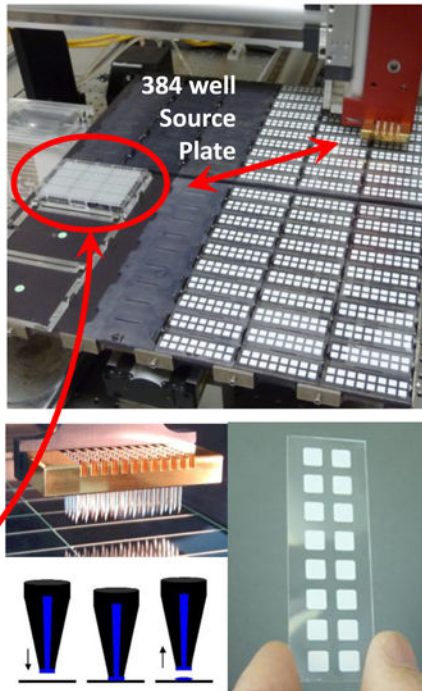
Figure 1. Microarray production, printing and analysis
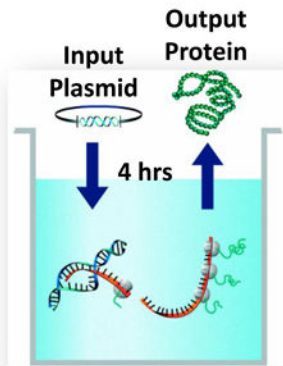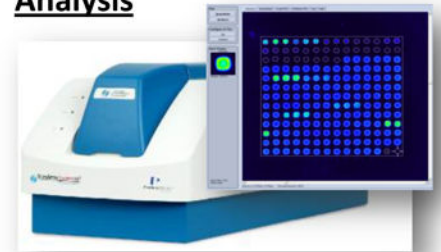Thousands of genes of interest were cloned using highthrough put method, *in vitro* expressed in *E.coli* cell free system. Protein microarrays were then produced, probed and data was analyzed as described in the text.

Table 1

An overview of seroreactive antigens and differentially reactive antigens identified in different organisms.

		Category	Total # proteins	# completed [*]	% completed	%Reactive Ags	% Differentially Reactive Ags	References
Viruses	Retroviruses: HIV 1&2 (5 subtype, 4 clades)		83	74	89%	28%		[36]
	HPV viruses (13 types)		104	104	100%	57%	16%	[32]
	Orthopoxviruses: 3 types		260	260	100%	20–22%	–	[9,31]
	Herpes viruses: HSV 1&2, VZV, EBV		300	270	90%	28.5%-42%	25.9%-41.7%	[37]
	Flaviviruses: WNV, Dengue, YF, SLE, JE	A	50	50	100%	NP	NP	
	Alphaviruses: Chikungunya	C	10	10	100%	NP	NP	
Bacteria	Brucella melitensis	B	3,198	3,046	95%	4%	1.10%	[12]
	Chlamydia trachomatis		911	894	98%	2.23%	0.56%	[3,25]
	Chlamydia muridarum		911	900	99%	NP	NP	
	Mycobacterium tuberculosis	C	3,990	3,899	98%	9.13%	0.26%	[33]
	Francisella tularensis	A	1,933	1,741	90%	2.76%		[11,23]
	Coxiella burnetii	B	2,065	2,000	97%	1–2.5% (IgG); 13.1% (IgM)	0.6–3.3%(G); 5.1%(M)	[15,38]
	Borrelia burgdorferi		1,640	1,293	79%	15.50%	7.97%	[10]
	Burkholderia pseudomallei	B	5,728	1,205	21%	8.96%	4.06%	[6]
	Leptospira interrogans [**]		3,667	3,359	92%	5.69%	1.07%	[39,40]
	Salmonella enterica Typhi [**]	B	4,318	4000	93%	7.31%(GM)	3.4%(GM)	[41]
	Orientia tsutsugamushi	C	1,400	1,400	100%	NP	NP	
	Rickettsia rickettsii	B	900	900	100%	NP	NP	
	Bartonella henselae	B	1,493	1,433	96%	7.33%	3.63%	[17]
	Enteric toxogenic E. coli					NP	NP	
Parasites	Clostridium difficile		4,000	3297	82%	NP	NP	
	Plasmodium falciparum		5,268	4,320	82%	13–21%	2.1–3.9%	[5,42,43]
	Plasmodium vivax		5,300	2,200	42%	8.04%	–	[44,45]
	Schistosoma mansoni / japonicum [****]		9,000	300	3%	9.8-44.7%	–	[4,46]

	Category	Total # proteins	# completed *	% completed	%Reactive Ags	% Differentially Reactive Ags	References
Toxoplasma gondii		8,155	1,015	12%		**18%**	[13]
Necator americanus		12,000	564	5%	**NP**	**3.90%**	[47]
Trypanosoma cruzi		15,099	240	2%	**NP**	**NP**	
Trypanosoma brucei	Human	8,529	214	3%	**NP**	**NP**	
Human Autoimmune array ****		21,000	1,800	9%		**2%**	[16]
Total		121,312	>40,000	33%			

*
 : non-published data included

**
 : both IgG and IgM were included in computation.

***
 : IgG and IgE subtype responses were measured

****
 : pemphigus vulgaris antibodies were measured

seroreactive cutoff usually 2-3SD no DNA controls; differentially reactive in distinguishing infected from healthy controls with a BH adjusted p value of 0.05 or less.

Table 2

Physical properties and cellular functions associated with immunogenity in bacteria, virus and parasites

Enriching features	Bacteria	Viruses	Parasites
Predicted functions	COG U Intracellular trafficking and secretion COG M Cell envelope biogenesis, outer memberane COG N Cell motility and secretion COG O Posttranslational modification, protein turnover, chaperones COG S	Virion membrane membrane proteins, core proteins	GO function protein binding, catalytic, transporter, transferase activities GO function enzymatic, regulator GO function structural molecule and ion channel activity GO ATP biosynthetic process, proteolysis; metabolic GO transport process
Predicted Computationally	TMHMM=1 Signal P>=0.7 pSortb Outer Membrane pSortb Periplasmic pI 0–5	TMHMM	TMHMM=1 to 10 Signal peptide pSort Outer Memberane pI 7–9
Expression	Expression evidence by Mass Spectrometry	abundance of expression	abundance of expression

Table 3

Naïve bayses classification of BM and SE.

top ranked antigens	Brucella melitensis		Salmonella enterica	
	% of total Seroreactive	% of total Serodiagnostic	% of total Seroreactive	% of total Serodiagnostic
1%	6%	15%	4%	5%
2%	11%	21%	4%	9%
5%	27%	45%	9%	25%
10%	44%	82%	19%	39%
20%	63%	91%	38%	64%
25%	72%	94%	45%	72%
50%	88%	100%	72%	86%
75%	94%	100%	100%	96%
100%	100%	100%	100%	100%