Research Overview

Omics-Based Systems Vaccinology for Vaccine Target Identification

Yongqun He*

Unit for Laboratory Animal Medicine, Department of Microbiology and Immunology, Center for Computational Medicine and Bioinformatics, Comprehensive Cancer Center, University of Michigan Medical School, 018 Animal Research Facility, 1150 W. Medical Dr., Ann Arbor, MI 48109, USA

Strategy, Management and Health Policy

| Enabling Technology, Genomics, Proteomics | Preclinical Research | Preclinical Development Toxicology, Formulation Drug Delivery, Pharmacokinetics | Clinical Development Phases I-III Regulatory, Quality, Manufacturing | Postmarketing Phase IV |

ABSTRACT Omics technologies include genomics, transcriptomics, proteomics, metabolomics, and immunomics. These technologies have been used in vaccine research, which can be summarized using the term “vaccinomics.” These omics technologies combined with advanced bioinformatics analysis form the core of “systems vaccinology.” Omics technologies provide powerful methods in vaccine target identification. The genomics-based reverse vaccinology starts with predicting vaccine protein candidates through in silico bioinformatics analysis of genome sequences. The VIOLIN Vaxign vaccine design program (http://www.violinet.org/vaxign) is the first web-based vaccine target prediction software based on the reverse vaccinology strategy. Systematic transcriptomics and proteomics analyses facilitate rational vaccine target identification by detecting genome-wide gene expression profiles. Immunomics is the study of the set of antigens recognized by host immune systems and has also been used for efficient vaccine target prediction. With the large amount of omics data available, it is necessary to integrate various vaccine data using ontologies, including the Gene Ontology (GO) and Vaccine Ontology (VO), for more efficient vaccine target prediction and assessment. All these omics technologies combined with advanced bioinformatics analysis methods for a systems biology-based vaccine target prediction strategy. This article reviews the various omics technologies and how they can be used in vaccine target identification. Drug Dev Res 73: 559–568, 2012. © 2012 Wiley Periodicals, Inc.

Key words: omics; vaccine; microarray; genomics; transcriptomics; proteomics

INTRODUCTION Vaccination is one of the most effective tools to prevent against infectious diseases, cancer, allergy, and autoimmune diseases. Infectious diseases are a major source of mortality, contributing to 26% of global human mortality in 2001 [Becker et al., 2006]. Cancer, allergy, and autoimmune diseases also cause significant mortality and morbidity in human and animal victims. Vaccine immunization induces strong host immune responses to the administrated antigen and provides a
long-term protection against a disease. However, an
effective and safe vaccine for many deadly diseases,
including acquired immunodeficiency syndrome,
tuberculosis, and malaria, still does not exist. New
approaches toward efficient vaccine target identifica-
tion and vaccine development are desired.

A new era of vaccine research began in 1995 when
the complete genome of *Haemophilus influenzae* (a
pathogenic bacterium) was published [Fleischmann
et al., 1995]. Since then, thousands of pathogen
genomes have been sequenced. Various host (such as
human and mouse) genomes are also available. With
the availability of large amounts of genome sequences,
high-throughput-omics technologies—genomics,
transcriptomics, proteomics, metabolomics, immunomics,
and other omics approaches—have been invented.
Advance bioinformatics approaches have also been
developed to support the analysis of large amounts of
omics data at differing levels, ranging from gene annota-
tion, data normalization, significant gene expression
detection, function enrichment, to pathway analysis.
These omic and bioinformatic technologies enable the
testing and screening of millions of possible vaccine
candidates and vaccine-induced host immune targets
in real time.

This article reviews how omics technologies com-
combined with advanced bioinformatics data analyses have
been used in vaccine target identifications.

**GENOMICS-BASED REVERSE VACCINOLOGY FOR
VACCINE TARGET IDENTIFICATION**

Reverse vaccinology is an emerging and revolu-
tionary vaccine development approach that starts with
the prediction of vaccine targets by bioinformatics
analysis of genome sequences. Predicted proteins are
selected based on desirable attributes. Reverse vaccin-
ology was first applied to the development of a vaccine
against serogroup B *Neisseria meningitidis* (MenB), the
major cause of sepsis and meningitis in children and
young adults [Pizza et al., 2000]. The complete MenB
genome was screened using bioinformatics algorithms
for open reading frames coding for putative surface-
exposed or secreted proteins, which are susceptible to
antibody recognition and therefore are the most suitable
vaccine candidates. Out of approximately 600 novel
vaccine candidates, 350 were expressed in *Escherichia
coli*, and 28 were found to elicit protective antibody
response. It took less than 18 months to identify more
vaccine candidates in MenB than had been discovered
over the previous 40 years by conventional methods
[Pizza et al., 2000]. Derived from the first reverse vacci-

cession attempt, Bexsero, a multicomponent, broad-
coverage MenB vaccine, was developed. Following the
generation of comprehensive clinical and epidemiologi-
dal data on Bexsero, Novartis submitted a Marketing
Authorization Application in late 2010 to the European
Medicines Agency for the use of Bexsero in humans
[Althoff and Gassenbach, 2010]. This milestone
occurred approximately 10 years after the first reverse
vaccinology publication, representing a huge success in
vaccine research and development (R&D).

Besides identifying secreted or outer membrane
proteins, many more reverse vaccinology criteria have
been developed since its first application report in
2000 described above. For example, when an outer
membrane protein contains more than one transmem-
brane helix, the recombinant protein is often difficult
to clone and purify [Pizza et al., 2000]. Therefore, the
number of transmembrane domains of a protein can
be used as a filtering criterion. Another criterion is the
selection of bacterial adhesins that are responsible for
adherence, colonization, and invasion of microbes to
host cells [Sachdeva et al., 2005]. With the availability
of multiple genomes sequenced for pathogens, it is
also possible to run comparative genomics analyses to
identify vaccine targets shared by many pathogenic
organisms. These conserved proteins can be used to
induce protection against multiple pathogenic strains.
It is also important to compare sequence similarities
between vaccine protein candidates and host pro-
teome. A pathogen protein homologous to a host
protein may induce an autoimmune disease or
immune tolerance. The traditional immunoinformatics
approaches for prediction immune epitopes can also
be used for screening protective antigens [He et al.,
2010b].

Vaxign (http://www.violinet.org/vaxign) is the first
web-based vaccine design program utilizing the reverse
vaccinology strategy [Xiang and He, 2009a; He et al.,
2010b]. Predicted features in the Vaxign pipeline
include protein subcellular location, transmembrane
helices, adhesin probability, conservation among patho-
gen genomes, conservation to human and/or mouse
proteins, sequence exclusion from genome(s) of non-
pathogenic strain(s), and epitope binding to major his-
tocompatibility complex (MHC) class I and class II.
Vaxign has been demonstrated to successfully predict
vaccine targets for *Brucella* spp. [Xiang and He, 2009a;
He and Xiang, 2010], uropathogenic *E. coli* [He et al.,
2010b], and human herpesvirus 1 virus [He and Xiang,
2011]. Currently, more than 200 genomes have been
precomputed using the Vaxign pipeline and available
for query in the Vaxign website. Vaxign also performs
dynamic vaccine target prediction based on input
sequences.

The concept of reverse vaccinology has been suc-
cessfully applied to many other pathogens (Table 1).
| Pathogen Genomics | DNA microarray | Proteinomics | Immunomics |
|-------------------|----------------|--------------|------------|
| **Gram-negative bacteria** | | | |
| 1. *Bordetella pertussis* (pertussis) | [Boo et al., 2010] | | [Vidasovic et al., 2007; Altindis et al., 2009; Zhu et al., 2010] |
| 2. *B. burgdorferi* spp. (lyme disease) | [Krastel et al., 2008; Xiang and He, 2009; He and Xiang, 2010] | [Tudos et al., 2009] | |
| 3. *Brucella* spp. (brucellosis) | [Chong et al., 2006] | | |
| 4. *Chlamydia pneumoniae* (pneumonia) | [Montigiani et al., 2002] | [Vandahl et al., 2001; Montigiani et al., 2002] | |
| 5. *Chlamydophila* spp. (cervicitis) | [Heinz et al., 2009] | [Sharma et al., 2006; Molina et al., 2010; Cruz-Fisher et al., 2011] | [Vieser et al., 2010] |
| 6. *Clostridium difficile* enterotoxigenic | [Sommer et al., 2010] | | |
| 7. *Escherichia coli* uropathogenic | [Welch et al., 2002; Durant et al., 2007] | [Snyder et al., 2004] | [Alteri and Mobley, 2007; Walters and Mobley, 2009] |
| 8. *Helicobacter pylori* (ulcer) | [Chakravarti et al., 2000; Dutta et al., 2006] | | [Chakravarti et al., 2000; Kornilovs’ka et al., 2002; Meinke et al., 2010] |
| 9. *Legionella pneumophila* (pneumonia) | [D’Auria et al., 2010] | [Galka et al., 2008; Khemiri et al., 2008] | |
| 10. *Neisseria meningitidis* serogroup B | [Pizza et al., 2000; Giuliani et al., 2006] | [Grifantini et al., 2002] | [Bernardini et al., 2007] |
| 11. *Porphyromonas gingivalis* | | | |
| 12. *Salmonella* spp. (typhoid fever) | | | |
| 13. *Shigella dysenteriae* (dysentery) | | | |
| 14. *Shigella flexneri* (dysentery) | | | |
| 15. *Vibrio cholerae* | | | |
| 16. *Yersinia pestis* (plague) | | | |
| **Gram-positive bacteria** | | | |
| 1. *Bacillus anthracis* | [Ariel et al., 2002; Read et al., 2003] | [Bergman et al., 2007] | [Chitgari et al., 2004] |
| 2. *Listeria monocytogenes* (listeriosis) | [Reis et al., 2010] | [Trost et al., 2005] | |
| 3. *Mycobacterium tuberculosis* | [Cockle et al., 2002] | [Betts, 2002; Stewart et al., 2002] | [Betts, 2002; Sartain et al., 2006; Giri et al., 2010; Li et al., 2010] |
| 4. *Neisseria meningitidis* serogroup B | [Pizza et al., 2000; Tettelin et al., 2000; Giuliani et al., 2006] | [Steller et al., 2005; Bernardini et al., 2007] | |
| 5. *Staphylococcus aureus* (bacteremia) | [Stranger-Jones et al., 2006; McCarthy and Lindsay, 2010] | | |
| 6. *Streptococcus agalactiae* (GAS) | [Beres et al., 2004] | | |
| 7. *Streptococcus pyogenes* (GAS) | | | |
| 8. *Streptococcus suis* (meningitis) | | | |
| **Viruses** | | | |
| 1. *Dengue virus* (Dengue fever) | [Kanlaya et al., 2009] | [Sanchez-Burgos et al., 2010] | |
| 2. *HIV* (AIDS) | [Kong et al., 2010; Wu et al., 2010] | [Wu et al., 2010] | |
| 3. *Measles-mumps-rubella* | | [Ovsyannikova et al., 2004, 2007] | |
| 4. *RSV* (respiratory tract infections) | | [Anderson et al., 2010] | |
| 5. *SARS virus* (SARS) | | | |
| 6. *West Nile virus* (West Nile fever) | | | |
| **Parasites** | | | |
| 1. *Plasmodium falciparum* (malaria) | [Kariyala et al., 2009] | | [Sanchez-Burgos et al., 2010] |
| 2. *Trypanosoma brucei* (sleeping sickness) | | | |
| 3. *Trypanosoma cruzi* (Chagas disease) | | | |
| 4. *Trypanosoma lewisi* (Old World trypanosomiasis) | | | |
| 5. *Trypanosoma rhodesiense* (African trypanosomiasis) | | | |
| 6. *Trypanosoma brucei gambiense* (African trypanosomiasis) | | | |
| **AIDS**, acquired immunodeficiency syndrome; GAS, group A *Streptococcus*; GBS, group B *Streptococcus*; HIV, human immunodeficiency virus; RSV, respiratory syncytial virus; RV, reverse vaccinology; SARS, severe acute respiratory syndrome.
Other features may also be considered for vaccine target prediction, for example, the application of possible three-dimensional (3D) structure in epitope prediction and antigen discovery [Mora et al., 2006; Serruto and Rappuoli, 2006]. It is also possible to predict vaccine targets based on other omics technologies, which will be introduced below.

**TRANSCRIPTOMICS-BASED DATA ANALYSIS FOR VACCINE TARGET IDENTIFICATION**

High-throughput transcriptomics analysis of gene expression using DNA microarray or RNA-seq next-generation sequencing technologies has revolutionized the way of studying genes that are involved in microbial pathogenesis. These assay systems are able to measure the expression pattern of thousands of genes in parallel, permitting the generation of large amounts of gene expression data. DNA microarrays can be hybridized with complementary DNA (cDNA) prepared from messenger RNA isolated from microorganisms grown in vitro or in vivo under different growth conditions. RNA-seq, also called "Whole Transcriptome Shotgun Sequencing," is a technique that uses high-throughput next-generation sequencing technologies to sequence cDNA in order to get information about a sample’s RNA contents. The next-generation sequencing has deep coverage and base-level resolution and does not require the prior knowledge of the genome sequence [Pinto et al., 2011]. The genes that are differentially transcribed in response to alternation in environmental variables in wild type or gene mutant microbes can be measured. It is important to find out what genes are expressed during host infection. Those genes that are expressed during disease represent most likely protective vaccine targets.

Many examples exist. For instance, the sexual stages of malarial parasites are essential for transmission of malaria by the mosquito and can be targeted for rational development of malaria vaccines. To better understand how genes participate in the sexual development process, Young et al. utilized microarrays to profile the transcriptomes of *Plasmodium falciparum* gametocytes at sexual stages [Young et al., 2005]. A 246-gene cluster associated with sexual development was identified using an ontology-based pattern identification algorithm. Some of the genes in the cluster can be potentially used for malaria vaccine development. More examples can be found in Table 1.

One challenge in vaccine target identification is the difficulty in experimental verification. The gold standard in vaccine target validation is a vaccination-challenge experiment, which tests if a vaccine immunization is able to induce protection against challenge of an infection with virulent pathogen in vivo. However, such experiments are often expensive and difficult to perform, especially for deadly pathogens (e.g., human immunodeficiency virus [HIV] and *Mycobacterium tuberculosis*) that do not have ideal small animal models that mimic human pathogenesis mechanisms. Therefore, it is important to identify an immune response that correlates well with protection. For many diseases, an ideal immune response that correlates with protection has not been found. Omics approaches can be used to identify host gene signatures that correlate and even predict protective immunity. For example, to identify early gene signatures induced in humans vaccinated with the attenuated yellow fever vaccine YF17D, two studies have examined total peripheral-blood mononuclear cells from human volunteers at different time points following vaccination with YF17D [Gaucher et al., 2008; Querec et al., 2009]. Microarrays were used to determine early effects (3 and 7 days postvaccination) of the vaccination on gene expression. A group of transcription factors, including IRF7, STAT2, and ETS2, functions as key regulators of the early immune response to YF17D vaccination [Gaucher et al., 2008]. A list of gene signatures (e.g., EIF2AK4 and TNFRSF17), which correlate with the magnitude of antigen-specific CD8+ T-cell responses and antibody titers, has also been identified and verified [Querec et al., 2009]. These gene signature profiles may serve as correlates of protection and be used in high-throughput screening of vaccine targets.

**PROTEOMICS-BASED DATA ANALYSIS FOR VACCINE TARGET IDENTIFICATION**

The availability of complete genome sequences allows the identification of all possible protein products. The advances in protein separation and mass spectrometry technologies make it possible to identify total protein components of a given cellular population or a subset of proteins from a particular cell compartment (e.g., outer membrane) under any specific growth condition. The combination of proteomics with serological analysis forms a new valuable approach called serological proteome analysis [Serruto and Rappuoli, 2006]. These proteomics approaches have been used in vaccine target predictions (Table 1).

Bioinformatics analysis of high-throughput gene expression results is a key to make novel discoveries. Gene expression is the process by which information from a gene is used in the synthesis of a functional gene product. The gene expression process occurs in the transcription level as well as the protein expression level. Therefore, omics gene expression data analysis covers the transcriptional and proteomic data analy-
The bioinformatics methods used for both transcriptional and proteomic gene expression are similar and include the following steps: (i) data preprocessing, including data quality controls and normalization; (ii) identification of significantly regulated genes using statistics methods; (iii) clustering, classification, and pattern discovery analyses; and (iv) inference of biological pathways and networks [Liang and Kelemen, 2006; Hendrickson et al., 2008]. The detailed methods have been surveyed and described in previous report [He et al., 2010a].

**IMMUNOMICS-BASED DATA ANALYSIS FOR VACCINE TARGET IDENTIFICATION**

Immunomics is the study of the set of antigens, especially T- and B-cell epitopes, which are recognized by host immune systems including human or animal hosts.

As comprehensively reviewed in a previous article [He et al., 2010a], many immunoinformatics algorithms have been invented to predict T- and B-cell immune epitopes. T-cell epitopes are bound in a linear form to MHC class I or class II molecules. T-cell epitopes can be predicted with high accuracy [He et al., 2010a]. B-cell epitopes can be linear or nonlinear (or called conformational). It remains a huge challenge to computationally predict B-cell immune epitopes. Currently, the best accuracy of predicting linear B-cell epitopes is approximately 60–70%. Over 90% of B-cell epitopes are nonlinear and require the knowledge of native 3D protein structure. There has not been proper approach achieving high performance in nonlinear B-cell epitope prediction.

Many examples of proteomics-based vaccine target identification studies are summarized in Table 1. It is noted that computational epitope prediction-based immunomics methods have often been integrated with other omics technologies. For example, using DNA microarray data, Sturniolo et al., [1999] developed a matrix-based computational algorithm to successfully predict a list of immunogenic epitope peptides uniquely associated with colon cancer. These peptides are likely vaccine targets for development of a colon cancer vaccine.

**ONTOLOGY-BASED INTEGRATED OMICS DATA ANALYSIS FOR VACCINE TARGET IDENTIFICATION**

Different omics technologies, in combination with advanced bioinformatics analyses, can be integrated to more effectively support vaccine target identification. Without effective integration of various omics data, it is difficult to provide comprehensive prediction of vaccine targets. The data integration problem can be addressed by the use of a biomedical formal ontology, i.e., a consensus-based controlled vocabulary of terms and relations, with associated definitions that are logically formulated in such a way as to promote automated reasoning. The Gene Ontology (GO) database (http://www.geneontology.org) has been widely used [Ashburner et al., 2000] in three categories: biological process, molecular function, and cellular component. GO has been widely used in various data integration and omics data analysis studies. As of July, 2012, over 4000 papers in PubMed cited the GO.

The community-based Vaccine Ontology (VO; http://www.violinet.org/vaccineontology) was developed to support vaccine data standardization, integration, and computer-assisted reasoning. VO contains more than 3000 terms, including more than 2000 vaccines that are licensed, in clinical trial, or proven effective in animals. These vaccines are targeted for over 20 animal species (e.g., human, cattle, and fish) against over 100 pathogens. Each vaccine is classified in VO using a logically defined, structured ontological hierarchy. As a vaccine knowledge base, VO also stores terms related to other vaccine-associated information, including different vaccine components (e.g., protective antigens and vaccine adjuvants) and vaccine-induced immune responses. Semantic relations between these terms are also included in VO to represent existing knowledge. These representations, written in the Web Ontology Language, can be parseable and readable by computer programs. VO has been used to model the meta-analysis of vaccine protection investigation [Brinkman et al., 2010; Todd et al., 2012]. VO can also be used to improve the indexing and literature mining of vaccine articles for analysis of *Brucella* gene-vaccine interaction networks [Xiang and He, 2009b; Hur et al., 2010] and discovery of IFN-γ and vaccine-associated gene networks [Ozgur et al., 2010]. The VO-based literature mining of all PubMed literature provides a novel method to predict vaccine targets [Hur et al., 2011].

**DISCUSSION**

Omnomics technologies combined with bioinformatics data analysis form the core parts of the “systems vaccinology,” a term derived from “systems biology” that is applied to the field of vaccinology. Systems vaccinology studies scientific questions in the field of vaccine and vaccination in a systems biology way. These omics and bioinformatics-based systems vaccinology methods have greatly supported the prediction and identification of vaccine targets.
Besides the omics technologies introduced above, many other vaccine target identification methods have been developed. For example, analysis of manually curated vaccine target data available in existing databases provides a powerful way for vaccine target prediction. As part of the VIOLIN vaccine database and analysis system [Xiang et al, 2008], Protegen is a web-based database that contains over 600 protective antigen information [Yang et al, 2011]. These antigens are also collected in the VO. Protective antigens are targeted by host acquired immunity and able to induce protection against infectious diseases. To identify features enriched in protective protein antigens, 201 protective protein antigens from Gram-negative bacteria and 69 protective protein antigens from Gram-positive bacteria collected in Protegen were analyzed [He and Xiang, 2012]. Of the protective antigens in Gram-negative bacteria, 64% are extracellular or cell wall proteins and 45% of protective antigens in Gram-negative bacteria belong to extracellular or outer membrane proteins. Over 40% protective bacterial antigens are adhesins or adhesin-like proteins. Many conserved motifs, including Autotransporter and TonB domains, are enriched in protective bacterial antigens. A predictive method based on the support vector machine (SVM) algorithm has a performance of 92% true positive rate of sequence-based protection. However, this SVM-based method has poor performance in differentiating true negative from false negative results [He and Xiang, 2012]. Overall, this study is pioneer in identifying specific patterns in protective antigens and computationally predicting protective antigens.

Many challenges also exist in the area of omics-based data analysis for vaccine target identification. Particularly, new vaccines are still needed to fight against infections with many deadly pathogens, such as HIV, M. tuberculosis, and F. tularensis. Although huge amounts of financial support have been invested and intensive research has been conducted, there still have not been safe and effective vaccines available for tackling these problems. Novel methods and ideas derived from the area of omics-based systems biology may provide hope toward better identification of new vaccine targets to support vaccine development.

REFERENCES

Alteri CJ, Mobley HL. 2007. Quantitative profile of the uropathogenic Escherichia coli outer membrane proteome during growth in human urine. Infect Immun 75:2679–2688.

Althoff E, Gassenbach N. 2010. Novartis submits Bexsero®, a multi-component meningococcal B vaccine, for regulatory review in Europe. Novartis Media releases December 23. http://www.novartis.com/newsroom/media-releases/en/2010/1475236.shtml Accessed 7 September 2012.

Altidis E, Tefon BE, Yıldırım V, Ozçengiz E, Becher D, Hecker M, Ozçengiz G. 2009. Immuno proteomic analysis of Bordetella pertussis and identification of new immunogenic proteins. Vaccine 27:542–548.

Anderson R, Huang Y, Langley JM. 2010. Prospects for defined epitope vaccines for respiratory syncytial virus. Future Microbiol 5:585–602.

Ariel N, Zvi A, Grosfeld H, Gat O, Inbar Y, Velan B, Cohen S, Shafferman A. 2002. Search for potential vaccine candidate open reading frames in the Bacillus anthracis virulence plasmid pX01: in silico and in vitro screening. Infect Immun 70:6817–6827.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, et al. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25:25–29.

Asif AR, Oellerich M, Amstrong VW, Riemenschneider B, Monod M, Reichard U. 2006. Proteome of conidial surface associated proteins of Aspergillus fumigatus reflecting potential vaccine candidates and allergens. J Proteome Res 5:954–962.

Bart MJ, van Gent M, van der Heide HG, Boekhorst J, Hermans P, Parkhill J, Mooi FR. 2010. Comparative genomics of prevaccination and modern Bordetella pertussis strains. BMC Genomics 11:627.

Bartlam M, Xu Y, Rao Z. 2007. Structural proteomics of the SARS coronavirus: a model response to emerging infectious diseases. J Struct Funct Genomics 8:85–97.

Baums CG, Valentín-Weigand P. 2009. Surface-associated and secreted factors of Streptococcus suis in epidemiology, pathogenesis and vaccine development. Anim Health Res Rev 10:65–83.

Becker K, Hu Y, Biller-Andorno N. 2006. Infectious diseases—a global challenge. Int J Med Microbiol 296:179–185.

Benes SB, Sylva GL, Sturdevant DE, Granville CN, Liu M, Ricklefs SM, et al. 2004. Genome-wide molecular dissection of serotype M3 group A Streptococcus strains causing two epidemics of invasive infections. Proc Natl Acad Sci USA 101:11833–11838.

Bergman NH, Anderson EC, Swenson EE, Janes BK, Fisher N, Niemeyer MM, Miyoshi AD, Hanna PC. 2007. Transcriptional profiling of Bacillus anthracis during infection of host macrophages. Infect Immun 75:3434–3444.

Bernardini G, Bracconi D, Martelli P, Santucci A. 2007. Postgenomics of Neisseria meningitidis for vaccines development. Expert Rev Proteomics 4:667–677.

Betts JC. 2002. Transcriptomics and proteomics: tools for the identification of novel drug targets and vaccine candidates for tuberculosis. IUBMB Life 53:239–242.

Bombaci M, Grifantini R, Mora M, Reguzzi V, Petracca R, Meoni E, et al. 2009. Protein array profiling of tic patient sera reveals a broad range and enhanced immune response against group A Streptococcus antigens. PLoS One 4:e6332.

Brinkman RR, Courtot M, Derom D, Fostel J, He Y, Lord P, et al. 2010. Modeling biomedical experimental processes with OBI. J Biomed Semantics 2(1 Suppl 1):S7.

Chakravarti DN, Fiske MJ, Fletcher LD, Zagusky RJ. 2000. Application of genomics and proteomics for identification of bacterial gene products as potential vaccine candidates. Vaccine 19:601–612.

Charles RC, Harris JB, Chase MR, Lebrun LM, Sheikh A, LaRocque RC, et al. 2009. Comparative proteomic analysis of the PhoP
Heinz E, Tischler P, Rattei T, Myers G, Wagner M, Horn M. 2009. Comprehensive in silico prediction and analysis of chlamydial outer membrane proteins reflects evolution and style of the Chlamydiae. BMC Genomics 10:634.

Hendrickson EL, Lamont RJ, Hackett M. 2008. Tools for interpreting large-scale protein profiling in microbiology. J Dent Res 87:1004–1015.

Hernaez ML, Ximenez-Embun P, Martinez-Gomariz M, Gutierrez-Blazquez MD, Nombela C, Gil C. 2010. Identification of Candida albicans exposed surface proteins in vivo by a rapid proteomic approach. J Proteomics 73:1404–1409.

Hiller NL, Janto B, Hogg JS, Boissy R, Yu S, Powell E, et al. 2007. Comparative genomic analyses of seventeen Streptococcus pneumoniae strains: insights into the pneumococcal supragenome. J Bacteriol 189:8186–8195.

Hurst J, Xiang Z, Feldman E, He Y. 2010. Ontology-based vaccine literature mining and indexing. Bio-Ontologies 2010: Semantic Applications in Life Sciences. 18th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB).

Hurst J, Xiang Z, Feldman EL, He Y. 2011. Ontology-based Brucella vaccine literature indexing and systematic analysis of gene-vaccine association network. BMC Immunol 12:49.

Jabbour RE, Wade MM, Deshpande SV, Stanford MF, Wick CH, Zulich AW, et al. 2010. Identification of Yersinia pestis and Escherichia coli strains by whole cell and outer membrane protein extracts with mass spectrometry-based proteomics. J Proteome Res 9:3647–3655.

Kanlaya R, Pattanakitsakul SN, Sichaikul S, Chen ST, Thongboonkerd V. 2009. Alterations in actin cytoskeletal assembly and junctional protein complexes in human endothelial cells induced by dengue virus infection and mimicry of leukocyte transendothelial migration. J Proteome Res 8:2551–2562.

Khemiri A, Galland A, Vaudry D, Chan Tchi Song P, Vaudry H, Kanlaya R, Pattanakitsakul SN, Sinchaikul S, Chen ST, Thongboonkerd V. 2007. Association network. BMC Immunol 12:49.

Kornilovs'ka I, Nilsson I, Utt M, Ljungh A, Wadstrom T. 2002. Comparative genomic analyses of seventeen Streptococcus pneumoniae strains: insights into the pneumococcal supragenome. J Bacteriol 189:8186–8195.

Kong L, Sheppard NC, Stewart-Jones GB, Robson CL, Chen H, Xu X, et al. 2010. Expression-system-dependent modulation of HIV-1 envelope protein antigenicity and immunogenicity. J Mol Biol 403:131–147.

Kornilovs’ka I, Nilsson I, Utt M, Ljungh A, Wadstrom T. 2002. Comparative genomic analyses of seventeen Streptococcus pneumoniae strains: insights into the pneumococcal supragenome. J Bacteriol 189:8186–8195.

Kong L, Sheppard NC, Stewart-Jones GB, Robson CL, Chen H, Xu X, et al. 2010. Expression-system-dependent modulation of HIV-1 envelope protein antigenicity and immunogenicity. J Mol Biol 403:131–147.

Kornilovs’ka I, Nilsson I, Utt M, Ljungh A, Wadstrom T. 2002. Comparative genomic analyses of seventeen Streptococcus pneumoniae strains: insights into the pneumococcal supragenome. J Bacteriol 189:8186–8195.

Kornilovs’ka I, Nilsson I, Utt M, Ljungh A, Wadstrom T. 2002. Comparative genomic analyses of seventeen Streptococcus pneumoniae strains: insights into the pneumococcal supragenome. J Bacteriol 189:8186–8195.
Scaria J, Boucoum-Fesnou, Peyrefitte CN, Belghazi M, Fusi T, Rogier C, et al. 2009. Identification of cellular proteome modifications in response to West Nile virus infection. Mol Cell Proteomics 8:1623–1637.

Peng X, Ye X, Wang S. 2004. Identification of novel immunogenic proteins of *Shigella flexneri* 2a by proteomic methodologies. Vaccine 22:2750–2756.

Fieper R, Zhang Q, Parmar FP, Huang ST, Clark DJ, Alhami H, et al. 2009. *The Shigella dysenteriae* serotype 1 proteome, profiled in the host intestinal environment, reveals major metabolic modifications and increased expression of invasive proteins. Proteomics 9:5029–5045.

Pinto AC, Melo-Barbosa HP, Miyoshi A, Silva A, Azevedo V. 2011. Application of RNA-seq to reveal the transcript profile in bacteria. Genet Mol Res 10:1707–1718.

Pizza M, Scarlato V, Magistri G, Giuliani MM, Arico B, Comanducci M, et al. 2000. Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. Science 287:1816–1820.

Querec TD, Akondy RS, Lee EK, Cao W, Nakaya HI, Teuwen D, Pieper R, Zhang Q, Parmar PP, Huang ST, Clark DJ, Alhami H, et al. 2011. The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. Nature 432:81–86.

Ross BC, Czaikowski L, Hocking D, Valiente E, Corton C, et al. 2009. Systems biology approach predicts immunogenicity of the yellow fever vaccine in humans. Nat Immunol 10:116–125.

Read TD, Peterson SN, Tousasse N, Baillie LW, Paulsen IT, Nelson KE, et al. 2003. The genome sequence of *Bacillus thuringiensis* subsp. amus. Nature 428:81–86.

Reis O, Sousa S, Camejo A, Villiers V, Gouin E, Cossart P, Cabanes D. 2010. LapB, a novel *Listeria monocytogenes* LPXTG surface adhesin, required for entry into eukaryotic cells and virulence. J Infect Dis 202:551–562.

Rodriguez-Ortega MJ, Norais N, Bensi G, Liberatori S, Capo S, Mora M, et al. 2006. Characterization and identification of vaccine candidate proteins through analysis of the group A *Streptococcus* proteome. Nat Biotechnol 24:191–197.

Ross BC, Czaikowski L, Hocking D, Margetts M, Webb E, Rothel L, et al. 2001. Identification of vaccine candidate antigens from a genomic analysis of *Porphyromonas gingivalis*. Vaccine 19:4135–4142.

Sachdeva G, Kumar K, Jain P, Ramachandran S. 2005. SPAN: a software program for prediction of adhesins and adhesin-like proteins using neural networks. Bioinformatics 21:483–489.

Sanchez-Burgos G, Peyrefitte CN, Belghazi M, Fusi T, Rogier C, et al. 2009. Identification of cellular proteome modifications in response to West Nile virus infection. Mol Cell Proteomics 8:1623–1637.

Peng X, Ye X, Wang S. 2004. Identification of novel immunogenic proteins of *Shigella flexneri* 2a by proteomic methodologies. Vaccine 22:2750–2756.

Fieper R, Zhang Q, Parmar FP, Huang ST, Clark DJ, Alhami H, et al. 2009. *The Shigella dysenteriae* serotype 1 proteome, profiled in the host intestinal environment, reveals major metabolic modifications and increased expression of invasive proteins. Proteomics 9:5029–5045.

Pinto AC, Melo-Barbosa HP, Miyoshi A, Silva A, Azevedo V. 2011. Application of RNA-seq to reveal the transcript profile in bacteria. Genet Mol Res 10:1707–1718.

Pizza M, Scarlato V, Magistri G, Giuliani MM, Arico B, Comanducci M, et al. 2000. Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. Science 287:1816–1820.

Querec TD, Akondy RS, Lee EK, Cao W, Nakaya HI, Teuwen D, Pieper R, Zhang Q, Parmar PP, Huang ST, Clark DJ, Alhami H, et al. 2011. The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. Nature 432:81–86.

Ross BC, Czaikowski L, Hocking D, Margetts M, Webb E, Rothel L, et al. 2001. Identification of vaccine candidate antigens from a genomic analysis of *Porphyromonas gingivalis*. Vaccine 19:4135–4142.

Sachdeva G, Kumar K, Jain P, Ramachandran S. 2005. SPAN: a software program for prediction of adhesins and adhesin-like proteins using neural networks. Bioinformatics 21:483–489.

Sanchez-Burgos G, Peyrefitte CN, Belghazi M, Fusi T, Rogier C, et al. 2009. Identification of cellular proteome modifications in response to West Nile virus infection. Mol Cell Proteomics 8:1623–1637.

Peng X, Ye X, Wang S. 2004. Identification of novel immunogenic proteins of *Shigella flexneri* 2a by proteomic methodologies. Vaccine 22:2750–2756.

Fieper R, Zhang Q, Parmar FP, Huang ST, Clark DJ, Alhami H, et al. 2009. *The Shigella dysenteriae* serotype 1 proteome, profiled in the host intestinal environment, reveals major metabolic modifications and increased expression of invasive proteins. Proteomics 9:5029–5045.

Pinto AC, Melo-Barbosa HP, Miyoshi A, Silva A, Azevedo V. 2011. Application of RNA-seq to reveal the transcript profile in bacteria. Genet Mol Res 10:1707–1718.

Pizza M, Scarlato V, Magistri G, Giuliani MM, Arico B, Comanducci M, et al. 2000. Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. Science 287:1816–1820.

Querec TD, Akondy RS, Lee EK, Cao W, Nakaya HI, Teuwen D, et al. 2009. Systems biology approach predicts immunogenicity of the yellow fever vaccine in humans. Nat Immunol 10:116–125.

Read TD, Peterson SN, Tousasse N, Baillie LW, Paulsen IT, Nelson KE, et al. 2003. The genome sequence of *Bacillus thuringiensis* subsp. amus. Nature 428:81–86.

Reis O, Sousa S, Camejo A, Villiers V, Gouin E, Cossart P, Cabanes D. 2010. LapB, a novel *Listeria monocytogenes* LPXTG surface adhesin, required for entry into eukaryotic cells and virulence. J Infect Dis 202:551–562.

Rodriguez-Ortega MJ, Norais N, Bensi G, Liberatori S, Capo S, Mora M, et al. 2006. Characterization and identification of vaccine candidate proteins through analysis of the group A *Streptococcus* proteome. Nat Biotechnol 24:191–197.

Ross BC, Czaikowski L, Hocking D, Margetts M, Webb E, Rothel L, et al. 2001. Identification of vaccine candidate antigens from a genomic analysis of *Porphyromonas gingivalis*. Vaccine 19:4135–4142.

Sachdeva G, Kumar K, Jain P, Ramachandran S. 2005. SPAN: a software program for prediction of adhesins and adhesin-like proteins using neural networks. Bioinformatics 21:483–489.

Sanchez-Burgos G, Peyrefitte CN, Belghazi M, Fusi T, Rogier C, et al. 2009. Identification of cellular proteome modifications in response to West Nile virus infection. Mol Cell Proteomics 8:1623–1637.

Peng X, Ye X, Wang S. 2004. Identification of novel immunogenic proteins of *Shigella flexneri* 2a by proteomic methodologies. Vaccine 22:2750–2756.

Fieper R, Zhang Q, Parmar FP, Huang ST, Clark DJ, Alhami H, et al. 2009. *The Shigella dysenteriae* serotype 1 proteome, profiled in the host intestinal environment, reveals major metabolic modifications and increased expression of invasive proteins. Proteomics 9:5029–5045.
Chlamydia pneumoniae elementary body. Electrophoresis 22:1204–1223.

Viadas C, Rodriguez MC, García-Lobo JM, Sangari FJ, López-Goni I. 2009. Construction and evaluation of an ORFeome-based Brucella whole-genome DNA microarray. Microb Pathog 47:189–195.

Vidakovics ML, Paba J, Lamberti Y, Ricart CA, de Sousa MV, Rodríguez ME. 2007. Profiling the Bordetella pertussis proteome during iron starvation. J Proteome Res 6:2518–2528.

Vytvytska O, Nagy E, Bluggel M, Meyer HE, Kurzbauer R, Huber LA, Klade CS. 2002. Identification of vaccine candidate antigens of Staphylococcus aureus by serological proteome analysis. Proteomics 2:580–590.

Walters MS, Mobley HL. 2009. Identification of uropathogenic Escherichia coli surface proteins by shotgun proteomics. J Microbiol Methods 78:131–135.

Welch RA, Burland V, Plunkett G 3rd, Redford P, Roesch P, Rasko D, et al. 2002. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic Escherichia coli. Proc Natl Acad Sci USA 99:17020–17024.

Wieser A, Romann E, Magistro G, Hoffmann C, Norenberg D, Weinert K, et al. 2010. A multiepitope subunit vaccine conveys protection against extraintestinal pathogenic Escherichia coli in mice. Infect Immun 78:3432–3442.

Wizemann TM, Heinrichs JH, Adamou JE, Erwin AL, Kunsch C, Choi GH, et al. 2001. Use of a whole genome approach to identify vaccine molecules affording protection against Streptococcus pneumoniae infection. Infect Immun 69:1593–1598.

Yang XM, Li N, Chen JM, Ou YZ, Jin H, Lu HJ, et al. 2006. Comparative proteomic analysis between the invasive and commensal strains of Staphylococcus epidermidis. FEMS Microbiol Lett 261:32–40.

Wang Y, Sturdevant DE, Otto M. 2005. Genomewide analysis of gene expression in Staphylococcus epidermidis biofilms: insights into the pathophysiology of S epidermidis biofilms and the role of phenol-soluble modulins in formation of biofilms. J Infect Dis 191:289–298.

Wen YT, Bhattacharya M, Stathopoulos C. 2008. Genome-wide in silico mapping of the secretome in pathogenic Yersinia pestis KIM. FEMS Microbiol Lett 279:56–63.

Ying T, Wang H, Li M, Wang J, Shi Z, Feng E, et al. 2005a. Immunoproteomics of outer membrane proteins and extracellular proteins of Shigella flexneri 2a 2457T. Proteomics 5:4777–4793.

Ying TY, Wang JJ, Wang HL, Feng EL, Wei KH, Huang LY, et al. 2005b. Immunoproteomics of membrane proteins of Shigella flexneri 2a 2457T. World J Gastroenterol 11:6880–6883.

Young JA, Fivelman QL, Blair PL, de la Vega P, Le Roch KG, Zhou Y, et al. 2005. The Plasmodium falciparum sexual development transcriptome: a microarray analysis using ontology-based pattern identification. Mol Biochem Parasitol 143:67–79.

Zhang L, Zhang ZP, Zhang XE, Lin FS, Ge F. 2010. Quantitative proteomics analysis reveals BAG3 as a potential target to suppress severe acute respiratory syndrome coronavirus replication. J Virol 84:6050–6059.

Zhong W, Skwarczynski M, Toth I. 2009. Development of conformational mimetics of conserved Streptococcus pyogenes minimal epitope as vaccine candidates. Curr Drug Deliv 6:520–527.

Zhu YZ, Cai CS, Zhang W, Guo HX, Zhang JP, Ji YY, et al. 2010. Immunoproteomic analysis of human serological antibody responses to vaccination with whole-cell pertussis vaccine (WCV). PLoS One 5:e13915.