

New Approaches to Vaccinology Made Possible by Advances in Next Generation Sequencing, Bioinformatics and Protein modeling

Amelia R. Woolums* and Cypriana Swiderski

Department of Pathobiology and Population Medicine, Mississippi State University, Mississippi State, MS 39759 USA.

*Corresponding author: amelia.woolums@msstate.edu

DOI: <https://doi.org/10.21775/cimb.042.605>

Abstract

Vaccines can be powerful tools, but for some diseases, safe and effective vaccines have been elusive. New developments in nucleic acid sequencing, bioinformatics, and protein modeling are facilitating the discovery of previously unknown antigens through reverse vaccinology approaches. Sequencing the complementarity-determining region of antibodies and T cell receptors allows detailed assessment of the immune repertoire and identification of paratopes shared by many individuals, supporting the selection of antigens that may be broadly protective. Systems vaccinology approaches to assess the global host response to vaccination by evaluation of differentially expressed genes in blood, cellular or tissue transcriptomes can reveal previously unknown pathways and interactions related to protective immunity. While it is important to remember that discoveries made through reverse vaccinology and systems vaccinology must still be confirmed with traditional challenge models and clinical trials, these approaches can provide new perspectives that may help solve longstanding problems in veterinary vaccinology.

Introduction

At its best, vaccination has substantial beneficial impact. The history of vaccination includes notable successes, such as the eradication of smallpox in humans and rinderpest in cattle (Greenwood, 2014). In ongoing campaigns, vaccines are used to prevent deaths due to pathogens such as rabies virus (Fisher and Schnell, 2018), and morbidity associated with agents such as foot-and-mouth disease virus (Diaz-San Segundo *et al.*, 2017). It is indisputable that vaccines can be powerful tools in the promotion of animal and human health. However, it has been difficult to develop reliable vaccines for some diseases, and vaccination has even rarely enhanced disease (Kapikian *et al.*, 1969). Clearly, information required to guide development of safe and effective vaccines is sometimes lacking.

In situations where knowledge gaps impede vaccine development, approaches leveraging new technologies that allow relatively rapid and inexpensive determination of nucleic acid sequences or protein structure may provide the missing insights. Here we summarize recent research using such technologies to identify new antigens, to characterize more holistically protective host responses to antigens and adjuvants, and to investigate adverse vaccine reactions. While much of the available information to date comes from investigations of human vaccines, the examples demonstrate the possibilities for answering a variety of questions relevant to veterinary vaccinology.

Vaccine development: classical versus next generation

Classical vaccinology is a hypothesis-driven approach to laboratory testing of pathogens that identifies antigenic components capable of eliciting protective immunity (Moxon *et al.*, 2019). Classical vaccinology employs a pipeline that begins by isolating the pathogen, followed by some manner of attenuation that retains immunogenicity, stimulates memory, and prevents pathogenicity. Subsequent iterations of this approach have applied biochemical, serological and microbiological methods to purify antigenic factors from both attenuated organisms and from organisms grown in culture—a laborious approach that yields few immunogens over very protracted time frames (Rappuoli, 2000). Clearly, classical vaccinology approaches have been informed by insight garnered from natural infections, such as the use of immune sera for screening candidate antigens, and recognition that farmers previously infected with cowpox did not develop lesions following smallpox

variolization, nearly 28 years prior to Jenner's application of this principle in the first vaccination (Boylston, 2013). Nonetheless, classical vaccinology relies heavily on injection into disease surrogates (most commonly laboratory animals) in order to identify the antigenicity of vaccine candidates and their ability to elicit a protective immune response.

Contrasting classical vaccinology, scientific and technological advances have ushered in a new generation of vaccinology in which candidate vaccine antigens are computationally predicted from the DNA sequences of pathogen genomes (Capecchi *et al.*, 2004). Moreover, the ability of peptides from these proteins to bind MHC molecules and become efficient T-cell epitopes; or to act as linear, conformational, or immunoglobulin class-specific (Saravanan and Gautham, 2018; Gupta *et al.*, 2013) B-cell epitopes, is being predicted *in silico* (reviewed in Dhanda *et al.*, 2017). Quite significantly, antigen-specific sequences of B-cell and T-cell receptors (ie paratopes) derived from individuals with protective immune responses are being characterized to identify their complementary antigenic epitopes. Paralleling these technological advances in vaccine antigen development are co-evolving technologies that are increasing knowledge and understanding of the mechanistic basis of protective immunity in the context of specific pathogens, as will be discussed later in this paper.

By refining candidate vaccine antigen prediction and the discernment of immune effector pathways that confer protective immunity, next generation vaccinology improves the efficiency with which animals are used to screen vaccine candidates for immunogenicity and protective immune responses. Significantly, the *in silico* predictive approaches, such as those used for candidate vaccine antigen prediction, reflect algorithms that are constantly evolving. Subsequent algorithms predict antigens that are both conserved, and novel relative to prior algorithms, while also failing to identify previously predicted antigens (Dalsass *et al.*, 2019). These discrepancies, which are inherent in the approach, highlight 1) a need for redundancy in the computational predictions employed to identify candidate vaccine antigens, and 2) the importance of comprehensive validation of the ability of predicted antigens to elicit protective immune responses in pre-clinical models and ultimately in the host species for the pathogen.

Reverse Vaccinology

Classical vaccinology successes are readily demonstrated for pathogens that lack antigenic variability or for which immunologic memory prevents reinfection (Rappuoli, 2007). However, most pathogens do not possess these characteristics. Classical vaccinology has proven inadequate for the development of vaccines against pathogens demonstrating antigenic diversity, pathogens that cannot be grown in vitro or lack adequate animal models of infection, and pathogens for which cell-mediated immune responses are protective (Rappuoli, 2000). Certain vaccines have also demonstrated safety issues tracing to the potential for recombination and immunologic events that result in enhanced disease (Agnew-Crumpton *et al.*, 2016, Kapikian *et al.*, 1969).

Contrasting the need for pathogen isolation with classical vaccinology, the advent of massively parallel DNA sequencing, also termed next generation sequencing (NGS), enabled rapid, cost-effective determination of a pathogen's genome sequence. Using a methodological approach coined *reverse vaccinology* by Rappuoli (2000), the full repertoire of protein coding sequences (the pathogen's proteome) is identified from the genome sequence of the pathogen. For bacteria and more complex pathogens, this step yields thousands of proteins, the majority of which are not relevant as vaccine antigens. To narrow the predicted proteins to those most relevant as candidate vaccine antigens, an expanding number of computational algorithms have been developed that model the encoded proteins from their DNA sequences, in order to predict characteristics of the proteins such as their 3-dimensional structure and predicted expression sites in the pathogen. Using this approach, for instance, proteins predicted to be secreted or exposed on the surface of the pathogen in particular conformations, and therefore subject to immune surveillance, can be identified and prioritized for analysis as vaccine candidates. Following this *in silico* prediction, DNA sequences of these candidate vaccine antigens are compared to genomic sequences of the host and homologous sequences with the potential to serve as autoantigens are eliminated.

Reverse vaccinology was first applied to Group B *Neisseria meningitidis*, an important cause of bacterial meningitis in humans. Whereas vaccines for other groups of *Neisseria meningitidis* employ the bacteria's capsular polysaccharide as an antigen, this approach is complicated by conserved polysialic acid residues in the

bacteria's capsular polysaccharide that are shared with human tissues (Serruto *et al.*, 2012). Employing the genome sequence of Group B *Neisseria meningitidis* (strain MC58), more than 2000 predicted proteins were analyzed *in silico* to identify proteins predicted to be surface exposed or excreted (Pizza *et al.*, 2000). Of the 570 predicted proteins, 350 were successfully expressed, purified, and used to immunize mice. Immune sera were screened by Western blot analysis against bacterial extracts to verify protein expression, while surface expression of the protein was confirmed by enzyme-linked immunosorbent assay (ELISA) and flow cytometry using intact, whole bacteria. Finally, complement-mediated killing activity of the antibodies (using human complement) was evaluated because it is an accepted correlate for *in vivo* protection in clinical trials of human meningococcal vaccines (Borrow *et al.*, 2006). Of the 91 proteins found to be positive in at least one of the first three assays, 28 induced antibodies with bactericidal activity (Serruto *et al.*, 2012). Five proteins were included in the final vaccine based on their ability to induce protection against diverse *N. meningitidis* strains. This protection was assessed by determining if specific antibodies to each protein antigen conferred passive protection in infant rodent models, or by identifying serum bactericidal antibodies following vaccination (Giuliani *et al.*, 2006). The final vaccine (4CMenB, Bexsero®) was released in Europe in 2013 (and the United States in 2015), 13 years following publication of the pathogen's genome sequence and initiation of the vaccine effort (Tettelin *et al.*, 2000).

While reverse vaccinology presents distinct advantages resulting from identification of the complete protein repertoire of a pathogen, dependence on protein coding sequences prevents identification of non-protein antigens such as polysaccharides, which have been important components of many successful vaccines, and CD1-restricted glycolipids which are promising vaccine candidates (Rappuoli, 2000). Also noteworthy is the finding that certain proteins can be restrictive in terms of their pathogen recognition and poorly antigenic (Sundling *et al.*, 2013), challenges that are circumvented by the use of adjuvants and construction of multi-epitope vaccines (Burton, 2017).

***In Silico* Prediction of B-cell Receptor/Antibody Epitopes**

The large number of candidate protein antigens identified by reverse vaccinology presents a significant bottleneck in the screening of these antigens to determine their

immunogenicity. Purified, full-length recombinant proteins used to elicit antibodies are often difficult to prepare in quantities sufficient for immunization. This is exemplified by the aforementioned screening of 570 predicted candidate vaccine proteins, 350 of which could be expressed in developing the Group B *N. meningitidis* vaccine, 4CMenB. To decrease the numbers of proteins subjected to *in vivo* screening, downstream technologies have been developed to increase the likelihood that predicted proteins from the genome sequence of a pathogen will be immunogenic. These analyses to predict immunogenic epitopes reflect the field of immunoinformatics that utilizes bioinformatics approaches to understand and interpreting immunological data.

B-cells are well recognized as the source of antibodies that provide protection from pathogens and cancerous cells. Antibodies recognize their target antigen by binding to portions of the antigen that are termed antigenic determinants or epitopes. Antigens generally possess many sites that function as epitopes. Regions of the antibody to which epitopes bind are termed paratopes, which are composed of six complementarity-determining regions (CDR) that confer antigen-specificity in epitope to paratope binding. Immunologic dogma holds that an antibody binds to a single epitope on an antigen, but expanding evidence indicates that some antibodies bind to more than one epitope (Van Regenmortel, 2014). A B-cell (and its clones) secretes, and also expresses on its surface, antibodies to the same epitope. Thus, antibodies expressed on the B-cell surface function as receptors (B-cell receptors) for the same epitope as the secreted antibodies. B-cell receptor stimulation by B-cell epitopes contributes to both the development of immunologic memory and antibody secretion. Accordingly, B-cell epitope prediction is an important goal to improve the likelihood that candidate vaccine proteins in reverse vaccinology are immunogenic.

B-cell epitopes can be linear (continuous) or conformational (discontinuous) (Stave and Lindpainter, 2013). Linear B-cell epitopes are formed by sequential amino acids in the protein. Conformational B-cell epitopes are formed by amino acids that are not sequential in the protein, but instead come into close contact to form an epitope as a result of the three-dimensional shape of the folded protein. Tools for *in silico* B-cell epitope have been reviewed elsewhere (Dhanda *et al.*, 2017). A primary challenge for *in silico* B-cell epitope prediction is that the majority of B-cell epitopes have been shown to be conformational (Greenbaum *et al.*, 2007; Kringelum *et al.*, 2013;

Ferdous *et al.*, 2019;). Illustrating the complexity of this challenge, Stave and Lindpaintner (2013) evaluated crystal structures from 111 antigen-antibody structures derived from proteins 22-442 amino acids long. They identified discontinuous epitopes that were formed from peptides ranging in length from 20 to 442 amino acids (average 50-79 amino acids) in which 20-101 (median=37) residues were in contact between the epitope and paratope, in clusters of 2-12 amino acids. Indeed, *in silico* B-cell epitope prediction tools that are based upon protein sequence attributes including calculations of hydrophilicity, flexibility, beta-turns, surface accessibility, amino acid composition and amino acid cooperativeness have been poorly predictive (Kringelum *et al.*, 2012; Jespersen *et al.*, 2017). Given that the theoretical B-cell receptor repertoire is nearly unlimited (10^{74} possible sequences, Saada *et al.*, 2007), the likelihood that any surface accessible region of a protein will have a complementary antibody conformation is rather high. Accordingly, the poor predictive value of *in silico* B-cell epitope predictions is likely an inherent characteristic of tools that are designed to comprehensively identify epitopes that are 'potentially' antigenic (Jespersen *et al.*, 2019). Supporting this assertion, conformational B-cell epitope predictions employing various artificial intelligence platforms that are trained in a more supervised manner, using crystal epitope structures that correspond to well characterized antibody paratopes, provide improved epitope predictions (Dhanda *et al.*, 2017; Jespersen *et al.*, 2017). However, these tools employ resolved crystal structures of target antigens that are not generally available. Addressing and perhaps solving this conundrum, Rahman *et al.* (2016) demonstrated, using a B-cell epitope prediction tool that characterizes sequence-based protein disorder tendency (IUpred-L), that epitope prediction based on short ≤ 11 -aa peptides falsely classifies B-cell epitopes as non-epitopes (53% accuracy). This is because short peptides of B-cell epitopes bind poorly. By modeling peptides of moderate length (16-30 amino acids) epitope prediction achieved 86% accuracy, indicating that training sets composed of appropriately sized longer peptides are necessary for accurate *in silico* B-cell epitope prediction.

***In Silico* Prediction of T-cell Receptor Epitopes**

Cell-mediated immunity is mediated by T-cells that have cytotoxic activity, ie cytotoxic T-cells, or by helper T-cells that provide co-stimulatory signals for B-cells and cytotoxic T-cells. Both helper T-cells and cytotoxic T-cells recognize epitopes in a highly specific manner, due to the antigen-specificity of paratopes in their T cell

receptors (TCRs). However, unlike B-cell receptors, TCRs recognize epitopes that are bound to Major Histocompatibility (MHC) receptors— specifically MHC I for cytotoxic T-cells, and MHC II for helper T-cells. MHC I receptors are expressed on the surface of all nucleated cells whereas MHC II receptors are found on antigen presenting cells. MHC genomic loci are among the most variable genes in mammals, enabling great diversity in MHC molecules. In humans, where MHC expression has been well characterized, there are at least 10,000 MHC I alleles (Robinson *et al.*, 2017) and 3,000 MHC II alleles (Rock *et al.*, 2016). Individuals express 3-6 different MHC I alleles (3 from each parent) and 3-12 different MHC II receptor alleles (Rock *et al.*, 2016). Each MHC allele is estimated to be able to bind 20 million (MHC I) to 200 billion (MHC II) epitopes (Rock *et al.*, 2016). Interaction between the T-cell receptor paratope and peptide bound to an MHC receptor confers antigen-specific recognition. A primary challenge for *in silico* TCR epitope prediction is evident in the possible combinations that are generated when MHC diversity is considered with the potential diversity in the antigen recognition regions of TCRs (ie TCR paratopes).

Approaches to *in silico* T-cell epitope prediction can be direct or indirect. Direct prediction targets identification of epitopes that bind to TCRs, while indirect prediction targets identification of epitopes that bind to MHC receptors. Indirect T-cell receptor epitope prediction is based on peptide binding to a groove within both MHC I and MHC II molecules in which peptides are maintained predominantly in an extended conformation (Madden, 1995). Early direct prediction algorithms were based upon structural analyses of helper T-cell epitopes that indicated these epitopes were amphipathic with helical turns (DeLisi and Berzofsky, 1985; Margalit *et al.*, 1987; Stille *et al.*, 1987). Subsequently, these findings were extended to cytotoxic T-cell epitopes (Reyes *et al.*, 1988). However, additional investigations of the MHC II peptide binding groove demonstrated that it cannot accommodate helices, but instead binds to linear peptides (Stern *et al.*, 1994).

MHC I molecules load peptides generated by proteasomal proteolysis, binding to the peptides in the endoplasmic reticulum (ER), after the peptides are translocated from the cytosol (Blum, *et al.*, 2013). The peptide-binding groove of MHC I molecules is closed at both ends and accommodates peptides that are typically 8 to 11 amino acids long (Rammensee *et al.*, 1995). N- and C-terminal ends of the peptide form hydrogen bonds with amino acids that are conserved across MHC I molecules, and

the peptide-binding groove contains deep binding pockets with tight physicochemical preferences (Madden, 1995). Collectively, these features enable binding predictions. MHC II bound epitopes classically result from cleavage of pathogen proteins to peptides in the endolysosome of antigen presenting cells, followed by association of peptide epitopes with MHC II molecules exiting the Golgi. Peptides bind to an MHC II binding groove through a series of hydrogen bonds in sites that are highly conserved within the MHC II molecule, despite differences in peptide epitope sequences and MHC proteins (Stern *et al.*, 1994). This confers a stereotyped but complex mode of binding across the spectrum of peptide-MHC II interactions that informs *in silico* prediction algorithms for MHC II binding epitopes. An important distinction between MHC I and MHC II molecules is that the binding groove for MHC class II is open at both ends (Painter and Stern, 2012). The option for bound peptide to protrude from the MHC molecule makes *in silico* MHC II binding prediction more difficult (Lundegaard *et al.*, 2010). As a consequence of the extension beyond the groove, peptides that bind MHC class II molecules tend to be of variable length but are typically 12 to 25 amino acids long (Jardetzky *et al.*, 1996).

Computational tools that predict peptide binding to MCH molecules are predicated on the availability of high-quality DNA sequence data for MHC alleles that are highly prevalent in the population that will be vaccinated. In humans >22,000 human MHC allele sequences have been deposited into the Immuo Polymorphism MHC Database (www.ebi.ac.uk/ipd/mhc/). In contrast, there are significantly fewer classical MHC alleles in this database for the horse, sheep, dog, pig and cow: currently 60, 247, 384, 450 and 678, respectively. MHC allele expression is vastly different across human ethnic groups (Terasaki, 2007), limiting world-wide application of T-cell epitope-based vaccination. This is relevant for veterinary medicine where the diversity and prevalence of MHC alleles within each species tends to be poorly characterized. Vaccines derived from MHC-peptide binding predictions are efficacious in individuals that express the MHC allele used for the epitope prediction. In pigs and cattle, where MHC allele expression is more likely to be defined within large portions of the population, these prediction methods are being applied. In this regard, MHC alleles have been identified in humans and cattle that share a high degree of epitope specificity (Lund *et al.*, 2004; Pandya *et al.*, 2015). These MCH supertypes enable T-cell epitope vaccines to address large populations with diverse MHC expression based upon expression of at least one

allele in the supertype. Accordingly, prioritizing the identification of MHC supertypes in species with highly polymorphic MHC molecules has particular advantages for vaccine design.

Structure based methods are available for predicting peptide binding to MHC complexes but these methods have low predictive performance (Nielsen *et al.*, 2018). In contrast, matrix-based predictions and machine learning technologies that employ protein sequences as input and are trained on known MHC-peptide interactions have improved predictive performance (Mei *et al.*, 2019; Nielsen *et al.*, 2018). Few of these tools have been validated on veterinary species. However, NetMHCpan employs artificial neural networks that have been trained through multiple iterations on peptide-MHC I affinity measurements from human, mouse, primates, cattle, and swine, as well as ligands and their complementary MHC I alleles from these species (Hoof *et al.*, 2009; Carrasco Pro *et al.*, 2014; Jurtz *et al.*, 2017; Hansen *et al.*, 2014). NetMHCpan can predict peptide-MHC class I binding for any allele of known sequence. Nine of 19 MHC I binding epitopes predicted by this tool for foot and mouth disease virus (FMDV, >1400 possible peptides) bound to pig MHC I (i.e. SLA) (Pedersen *et al.*, 2013). A web interface of NetMHCpan that is specifically trained for identifying MHC I restricted peptides in cattle is available (Nielsen *et al.*, 2018). This tool has been used to identify MHC I-peptide binding for FMDV (Pandya *et al.*, 2015), bovine herpes virus, and *Theileria parva*, the causative agent of East Coast Fever (Nielsen *et al.*, 2018; Svitek *et al.*, 2014). An alternative proprietary prediction tool for MHC I and MHC II peptide binding, PigMatrix, is based on the pocket method of MHC I-peptide binding, and has been used to identify T-cell vaccine epitopes to for influenza virus (Gutiérrez *et al.*, 2015; Hewitt *et al.*, 2019).

Additional Antigen Selection Criteria Employed in Reverse Vaccinology

RNA-sequencing (RNAseq) is a technology that both identifies and quantifies RNA sequences that are expressed within a biological sample. In one embodiment of its utility for candidate vaccine antigen identification, RNAseq has been used to quantify the level to which *in silico* predicted vaccine candidates are expressed (as mRNA) by the pathogen. This can be done for instance during conditions of natural infection, or during surrogates of natural infection such as low pH or osmotic stress, as would be anticipated in the phagosome. These results are then informative in ranking the *in silico* predicted vaccine candidates based upon the relevance of their expression

levels. Similarly, proteomic analysis, which identifies the proteins that are expressed and their magnitude of expression can also be employed to quantify expression of putative vaccine peptide epitopes in order to assist in ranking their relevance.

Reverse Vaccinology for Veterinary Pathogens

Employing a pathogen's genome sequence to predict vaccine epitopes using reverse vaccinology presents distinct advantages over classical approaches. This approach has enabled vaccine design in for problematic pathogens that cannot be cultivated. Such is the case for *Pajaroellobacter abortibovis*, the causative agent of foothills abortion in cattle, which can only be cultivated in SCID mice (Blanchard *et al.*, 2010). Using a reverse vaccinology approach, Welly *et al.* (2017) first determined the genome sequence of the pathogen by sequencing DNA from infected SCID mice. Computational methods were used to subtract the well-characterized genomic sequences originating from the SCID mice, leaving the remaining genomic regions belonging to *P. abortibovis*. DNA regions belonging to *P. abortibovis* were then arranged into a complete *de novo* genome sequence from which the authors identified 10 putative vaccine candidates using *in silico* prediction. Seven of these predicted proteins, when expressed as fusion proteins, were recognized by serum from *P. abortibovis* infected SCID mice, indicating that the proteins could be relevant to protective immunity.

In other research, a reverse vaccinology approach was recently used to identify cat flea (*Ctenocephalides felis*) surface antigens predicted to be immunogenic. When the antigens were incorporated into experimental vaccines given to cats, flea eggs collected from vaccinated cats were significantly less likely to hatch ($P < 0.05$), as compared to flea eggs collected from cats receiving adjuvant only (Contreras *et al.*, 2018). Tomazic *et al.* (2018) used reverse vaccinology to identify antigens of *Cryptosporidium parvum* that reacted with antibodies in serum collected from *Cryptosporidium*-infected calves in the first month of life, indicating that the antigens could be associated with protection at the time when calves are most vulnerable to diarrhea resulting from cryptosporidiosis. In both of these examples reverse vaccinology allowed rapid screening of a large number of candidate proteins to identify antigens which might have been overlooked in classical approaches using whole pathogens or their components to induce immunity with experimental challenge.

Reverse vaccinology also presents advantages when dealing with pathogens that exhibit antigenic diversity. For instance, the DNA from multiple strains of a given bacterial species, termed a pangenome, can be simultaneously analyzed to allow *in silico* prediction of shared epitopes. This approach was used to identify T cell epitopes in antigens shared among 30 *Acinetobacter baumannii* strains, with a goal to developing effective vaccines for this bacterial pathogen that demonstrates widespread resistance to antimicrobials (Hassan A *et al.*, 2016). Also relevant is the case of many RNA viruses where host selection pressure drives gene mutations that contribute to antigenic diversity. For example, sequencing of variable regions in nonstructural proteins 1 and 2 (nsp1 and nsp2) and the products of open reading frame 3 and 5 (ORF3 and ORF5) in pigs infected with porcine reproductive and respiratory syndrome virus (PRRSV) revealed genetic changes related to virus rebound at 42 days post infection (Chen *et al.*, 2016). Improved understanding of the mechanisms by which PRRSV escapes host immunity could facilitate development of vaccines that more effectively prevent infection.

Epitope-Focused Vaccine Design

Whereas massively parallel DNA sequencing provided a gateway to reverse vaccinology, it is now possible to identify the breadth of the immune repertoire by sequencing DNA encoding antibody and T-cell receptors (TCR) from hundreds of thousands of individual immune cells (single cell sequencing) (Davis and Boyd, 2019). Further, sequencing and quantification of RNA sequences that encode antibody and TCRs from large populations of immune cells can be used to identify changes in immune repertoire expression during responses to vaccination (DeKosky *et al.*, 2013; Waickman *et al.*, 2019) or infection (Sant *et al.*, 2018; Mitsunaga and Synder, 2020). Epitope-focused vaccine design (subsequently coined Reverse Vaccinology 2.0 (Rappuoli *et al.*, 2016), refers to the design of vaccine epitopes that bind to paratopes on an antibody or on a T-cell receptor that are known to confer protection during infection (Correia *et al.*, 2014; Sela-Culang *et al.*, 2015). Prediction of complementarity between an antigen epitope and an antibody paratope differs significantly from predicting B-cell epitopes. This is because prediction of epitope/paratope complementarity is constrained by stringent epitope/paratope interactions, while only surface accessibility constrains B-cell epitope prediction and accordingly yields a large number of protein sequences (Novotný *et al.*, 1986).

Epitope focused vaccine design can be viewed in four stages, with different methodological approaches being available to address each stage. First, individuals are identified that have protective immune responses to a pathogen, such as broadly neutralizing antibodies, bactericidal antibodies, or antigen-specific helper T-cell or cytotoxic T-cell responses. Second, B-cells or T-cells with paratopes that confer the desired immunologic outcome are isolated and the paratope sequences are determined. Third, structural features of the cloned paratopes are computationally modeled in order to predict the structure of the corresponding antigen epitope. Fourth, immunological testing and validation of the predicted antigen epitope is undertaken. In one embodiment of this approach, antibodies providing the majority of virus neutralizing activity for human respiratory syncytial virus were critical to the identification of an antigenic but transient fusion protein conformation. By inserting point mutations that stabilized the fusion protein structure, the antigen has been advanced as a vaccine candidate that has been effective in clinical trials (Magro *et al.*, 2012; Gilman *et al.*, 2016; Crank *et al.*, 2019). Alternatively, *in silico* co-modeling of CD4+T-cell receptor and MHC II sequences from human patients with latent tuberculosis, who were vaccinated with *M. tuberculosis* peptides, identified vaccine epitopes that triggered CD4+ T-cell responses (Dash *et al.*, 2017). While this approach has theoretical strength for identifying key protective antigens, its primary drawback is the resolution of only a portion of the full repertoire of epitope-specific antibodies and T-cell receptors, which limits comprehensive characterization of antigen-specific immune effector mechanisms.

Assessment of the host response: “systems vaccinology”

An effective immune response is the result of the choreographed interaction of thousands of molecular or cellular reactions which may be separated in the host by the distance of a meter or more. As the number of recognized components participating in the immune response has grown, it has become obvious that historical approaches to measuring immunity, which often focused only on serum antibody or a few cellular reactions, missed much of the picture. While for some applications one or a few signature responses may provide enough information, in cases where the nature or mechanics of a protective response are unclear, a more detailed portrait may be informative. This is the premise of “systems vaccinology”:

development and assessment of vaccines with more of the immune response in view.

Systems vaccinology can be considered a branch of systems biology, in which interdisciplinary approaches are used to characterize the complex networks within a biological system, in order to better understand and predict the behavior of the system (Pulendran *et al.*, 2010). A systems biology assessment evaluates all parts of a system responding to some perturbation, analyzing the results together in order to develop a model that more accurately represents the nature of the whole. Systems vaccinology approaches can reveal the involvement of formerly unknown genes and pathways related to the response to a vaccine; thus, these methods are being applied to identify correlates that predict protective immunity, to reveal mechanisms of vaccine or adjuvant efficacy, and to uncover factors contributing to adverse vaccine reactions.

In spite of the potential of systems vaccinology, the use of this approach may be limited by the high cost of RNAseq or other required technologies, which may substantially limit the number of individuals that can be evaluated, thus weakening study power. The complications of analyzing massive data sets is also a formidable hurdle. It is also important to remember that the findings of a systems vaccinology investigation must still be confirmed with prospective investigations requiring use of more classical methods. Therefore, systems vaccinology should be considered a complement to traditional vaccinology, and not a replacement.

Predicting protective immunity induced by vaccination

While some vaccines are reliably protective, many are not. Moreover, the correlates of protective immunity are known and can be measured for some vaccines, but they are not clear for others. And, at times, immune responses that are assumed to be important are not identified in vaccinated patients that are nevertheless protected. In cases such as these, a global assessment of the immune response may help to reveal unknown but important mechanisms. In most cases to date this global systems vaccinology assessment has been made by assessing gene expression in blood or other tissue, by measuring hundreds or thousands of transcripts, most commonly using microarrays or, more recently, high throughput RNAseq. In humans and livestock, blood or peripheral blood mononuclear cell (PBMC) transcriptomes

have been assessed, because blood is usually easy to collect, and also because lymphocytes and antigen presenting cells traffic through blood during their response to vaccination at various sites. Increased gene expression over baseline in whole blood or PBMC can sometimes be seen within 24 hours of vaccination (Vahey *et al.*, 2010), with expression at 24 hours sometimes significantly associated with outcomes weeks later (Matthijs *et al.*, 2019). Transcriptomes in tissues collected from vaccinated animals subjected to necropsy at various times after vaccination or challenge have also been used to investigate vaccine responses (Luo *et al.*, 2014; Li *et al.*, 2020).

An early application of systems vaccinology was used to characterize the response of humans vaccinated with the yellow fever vaccine YF-17D (Querec *et al.*, 2009). While YF-17D has been administered for decades and is highly effective, the reasons for its efficacy are still not clear. To improve understanding of the immunologic mechanisms elicited by YF-17D vaccination, Querec *et al.* investigated post-vaccination changes in plasma cytokines, chemokines, blood antigen presenting cells, and expression of 20,000 genes in peripheral blood mononuclear cells (PBMC) by microarray over the first 21 days after vaccination in 15 individuals. These early events were compared to correlates of later protective immunity: SN titers and CD8+ (cytotoxic T cell) activation at days 15 and 60 after vaccination. Following vaccination, early events including increases in CXCL10 (previously IP-10), IL-1 α , and the percent of circulating CD86+ antigen presenting cells were identified, but these events were not correlated with CD8+ responses at 60 days post vaccination. In contrast, when 839 genes that correlated to the magnitude of day 60 CD8+ responses were subjected to unsupervised principal component analysis (PCA) followed by discriminant analysis using mixed integer programming (DAMIP), predictive relationships between PBMC gene expression and day 60 CD8+ responses were identified. PCA is a statistical procedure commonly employed in the analysis of gene expression datasets which is sensitive to strong patterns of expression that are shared across samples. PCA uses an orthogonal transformation to convert expression data into a set of values of linearly uncorrelated variables called "principal components". DAMIP is a predictive modeling framework that uses a supervised-learning classification approach used to predict biomedical phenomena. When the investigators repeated this analytic approach in a second group of vaccinated individuals, the results predicted day 60 CD8+ responses with 87%

accuracy. Similarly, the investigators were able to predict SN responses in the second group of vaccinates with 80% accuracy.

In a greatly expanded effort, this group went on to evaluate the relationship between the PBMC transcriptome in the week after vaccination and protective immunity induced by each of five vaccines: two meningococcal vaccines, YF-17D, a trivalent inactivated influenza vaccine, or a live-attenuated influenza vaccine (Li *et al.*, 2014). For each vaccine, established correlates of immunity (such as serum antibody titers) were measured in vaccinated human subjects, and correlations with genes that were differentially expressed between baseline and the week after vaccination were assessed. For subjects vaccinated with one of the two meningococcal vaccines, 1,150 genes were differentially expressed between the day of vaccination and day 7 postvaccination. Importantly, the research showed that evaluation of the correlation between *individual* differentially expressed genes (DEG) and protective antibody titers led to few significant and meaningful relationships. However, when DEG were grouped into blood transcription modules (BTM), based on shared functions and previously described interactions, BTM were significantly associated with protective responses to all five of the vaccines evaluated. Notably, some BTM significantly associated with protective responses to viral vaccines were different from the BTM significantly associated with protective responses to bacterial vaccines. It seems likely that assessment of gene expression by evaluating genes together in functional groups or pathways, rather than individually, may provide a more accurate picture of the host response to vaccination. Also, evaluating the response of groups of genes with coordinated expression may reveal significant changes when change in expression of any one gene may be too small to detect (Haining 2014).

The BTM developed by Li *et al.* were later used to evaluate early immunologic events in pigs responding to five different *Mycoplasma hyopneumoniae* vaccines (Matthijs *et al.*, 2019). These investigators identified an early upregulation of modules associated with innate immunity, including monocyte and neutrophil function, as well as inflammation and pathogen sensing within 24 hours of vaccination, which correlated to adaptive immune responses to *M. hyopneumoniae* at later time points. Significantly, pigs demonstrating increased expression of genes in these innate immune modules more than 24 hours following vaccination had weaker adaptive responses than early responding pigs. Collectively these findings

indicate that the best adaptive immune response to *M. hyopneumoniae* vaccination was associated with increased gene expression that supports innate immunity in the first 24 hours after vaccination, followed immediately by decreased expression.

Researchers have subsequently used similar methods to characterize the response to other veterinary vaccines, for example, in German Landrace pigs and Pietrain pigs vaccinated against porcine respiratory and reproductive syndrome virus (PRRSV; Islam *et al.* 2019), in carp vaccinated orally with a DNA vaccine against *Vibrio mimicus* (Li *et al.*, 2020), and in horses vaccinated against African horsesickness virus (Pretorius *et al.*, 2016). However, to date relatively few groups have used blood transcriptome assessment developed in one cohort to predict vaccine efficacy in another cohort. Accurate and rapid prediction of vaccine efficacy in populations, without the need to wait to assess response to disease challenge, is a major goal of systems vaccinology. Ideally, findings initially made in relatively small groups with expensive and complicated transcriptomics could be adapted to a rapid, inexpensive test, perhaps based on PCR or sequencing of a small number of transcripts, that could be applied to larger populations.

More recently, whole blood and PBMC transcriptome data has been used to identify gene expression signatures that predict antibody responses in young (<35 years) and aged (>60 years) human cohorts following influenza vaccination in different regions of the United States over multiple years (HIPC-CHI Signatures Project Team and HIPC-I Consortium, 2017). These investigators identified high and low antibody responders based on changes in their serum antibody titers (hemagglutination inhibition or SN) following vaccination. By evaluating changes in the expression of more than 32,000 genes in response to vaccination, a “response score” based on expression levels of 9 differentially expressed genes (DEG) or 3 DEG modules (similar to the BTM of Li *et al.*, 2014) was identified which differentiated high versus low antibody responders in young but not aged cohorts. Significantly, the response score accurately predicted individuals that were high or low antibody responders when applied to a separate young vaccination cohort. While these investigators successfully used blood transcriptome data from one cohort of vaccinates to predict protective responses in a second cohort, their work also illuminated limitations of the systems vaccinology approach: 1) multiple cohorts were needed to identify the DEG that predicted SN antibody responses because there was too much variability in any

one cohort to find this relationship; and 2) the validity of the findings was dependent on the population structure of the discovery cohort. Specifically, the DEG that predicted SN antibody responses were valid for individuals under 35 years of age, but not for individuals over 60 years of age. In fact, there was an inverse relationship between gene expression associated with SN responses in young individuals versus older individuals, indicating that gene expression signatures that were *positively* associated with SN antibody titers following vaccination in young individuals were *negatively* associated with SN antibody titers following vaccination in old individuals. While this is likely an important finding, the results indicate that it may not be possible to extrapolate gene expression profiles from one type of population to another. Moreover, it may be necessary to assess blood transcriptomes in large numbers of individuals, or in multiple smaller cohorts, to identify meaningful predictive signatures. Though achieving populations of this magnitude may limit similar applications in veterinary research, advances in technology that have decreased the costs of assessing a transcriptome are likely to make these applications increasingly accessible.

Mechanisms of vaccine and adjuvant efficacy

In addition to identifying genes, gene modules, or pathways of molecular interactions that predict vaccine efficacy, systems vaccinology approaches can reveal previously unknown components or mechanisms related to vaccine or adjuvant efficacy, or immunity more generally. When differential expression of dozens or hundreds of blood or PBMC transcripts is associated with a favorable response to a vaccine, some of these transcripts may be from genes that are poorly characterized, or perhaps unexpected. Further investigation of these genes in association with more classical methods to assess immunity and disease resistance can lead to discovery of relevant new mechanisms (Hagen and Pulendran, 2019). For example, the finding that expression of Toll-like receptor 5 (TLR5, a pathogen recognition receptor that binds bacterial flagellin) was associated with influenza antibody titers a month after vaccination (Nakaya *et al.*, 2011) directed further research leading to the discovery that flagellin activates production of B-cell growth factors by macrophages. Moreover, treatment of mice with antimicrobials to decrease gut bacteria (a large source of flagellin exposure for local immune cells) led to decreased antibody production following vaccination with inactivated influenza vaccine, but not modified live vaccine (Oh *et al.*, 2014). Collectively, these findings indicate that the normal

intestinal flora may act as a sort of “endogenous adjuvant” for individuals vaccinated with certain vaccines. Further research in this area may lead to ways to improve response to vaccines by modifying the normal intestinal flora, or perhaps by co-treatment with probiotics. These findings also illustrate how investigation of an unexpected molecular interaction identified by a systems vaccinology approach led to discovery of a new concept in immune system regulation that is directly relevant to vaccine efficacy.

Though adjuvants have been used for over a century to improve response to vaccines, the exact immune mechanisms by which adjuvants exert their effects is surprisingly ill-defined. Systems vaccinology approaches present an opportunity to identify the genes and molecular pathways responsible for adjuvant effects. This approach has been described for adjuvants containing different cationic lipids or Toll-like receptor ligands in experimental *M. hyopneumoniae* vaccines (Mattijis *et al.*, 2019). In these experiments, cationic lipid adjuvants, which were associated with the highest serum antibody titers post vaccination, induced rapid but transient upregulation (generally limited to the first day after vaccination) in blood transcription modules related to myeloid cell activation in vaccinated pigs.

Investigation of adverse vaccine reactions

Just as systems vaccinology approaches can be used to reveal genes or pathways associated with the response to a vaccine or adjuvant, they can also be used to dissect adverse vaccine responses, in order to prevent them (Gonzalez-Dias *et al.*, 2020). To investigate the mechanistic basis of bovine neonatal pancytopenia, an adverse reaction linked to cattle vaccinated with inactivated bovine viral diarrhea virus (BVDV) vaccines, Demasius *et al.* (2013) analyzed whole blood transcriptome of vaccinated cattle. Perhaps surprisingly, the analysis identified evidence of a coordinated response to double stranded RNA, but not, as expected, to alloantigens. The investigation also led to the discovery of a gene for an apparently novel cytokine, which was strongly upregulated in vaccinated cattle. Together these findings provided evidence of unexpected aspects of the pathogenesis of neonatal pancytopenia, although further research will be necessary to establish the significance of these findings.

Systems vaccinology, summary

The immune response to a vaccine is the result of coordinated interaction of thousands of events. Given this situation, it is perhaps not surprising that research focused on a handful of molecules or cells sometimes fails to explain why a vaccine induces—or fails to induce—protection. Global assessment of gene activation in blood or other tissues through assessment of the transcriptome can provide new insights regarding mechanisms of vaccine efficacy. While reports to date have rarely combined assessment of the transcriptome with protein and metabolite expression (the “proteome” and the “metabolome”), as technology progresses and costs decrease, the simultaneous assessment of these components may provide yet more useful information. Currently such multi-modal approaches are limited by the cost or technical demands that limit the number of individuals that can be studied, and thus limit the power and generalizability of the research. However, given that the price of nucleic acid sequencing has steadily decreased since the discovery of DNA, it is likely that the limits due to cost will eventually be lifted. Continued advances in the field of bioinformatics should also improve the feasibility of systems vaccinology approaches. Used in tandem with classical methods to confirm new discoveries, systems vaccinology has clear potential to support the development and delivery of safer and more effective vaccines.

References

- Agnew-Crumpton, R., Vaz, P.K., Devlin, J.M., O'Rourke, D., Blacker-Smith, H.P., Konsak-Ilievski, B., Hartley, C.A., and Noormohammadi, A.H. (2016). Spread of the newly emerging infectious laryngotracheitis viruses in Australia. *Infect Genet Evol.* 43, 67-73. DOI: 10.1016/j.meegid.2016.05.023.
- Blanchard, M.T., Chen, C.I., Anderson, M., Hall, M.R., Barthold, S.W., and Stott, J.L. (2010). *Vet Microbiol* 144, 177-182. DOI: 10.1016/j.vetmic.2010.01.002.
- Blum, J.S., Wearsch, P.A., and Cresswell, P. (2013). Pathways of antigen processing. *Annu Rev Immunol* 31, 443-473.
- Borrow, R., Carlone, G.M., Rosenstein, N., Blake, M., Feavers, I., Martin, D., Zollinger, W., Robbins, J., Aaberge, I., Granoff, D.M., Miller, E., Plikaytis, B., van Alphen, L., Poolman, J., Rappuoli, R., Danzig, L., Hackell, J., Danve, B., Caulfield, M., Lambert, S., and Stephens, D. (2006). *Neisseria meningitidis* group B correlates of protection and assay standardization—international meeting report

- Emory University, Atlanta, Georgia, United States, 16-17 March 2005. *Vaccine* **24**, 5093-5107.
- Boylston, A. (2013). The origins of vaccination: myths and reality. *J R Soc Med* **106**, 351-354. DOI: 10.1177/0141076813499292.
- Burton, D.R. (2017). What Are the Most Powerful Immunogen Design Vaccine Strategies? Reverse Vaccinology 2.0 Shows Great Promise. *Cold Spring Harb Perspect Biol* **9**, pii: a030262. DOI: 10.1101/cshperspect.a030262.
- Capecchi, B., Serruto, D., Adu-Bobie, J., Rappuoli, R., and Pizza, M. (2004). The genome revolution in vaccine research. *Curr Issues Mol Biol* **6**, 17-27.
- Chen, N., Triple, B.R., Kerrigan, M.A., Tian, K., Rowland, R.R.R. (2016). ORF5 of porcine reproductive and respiratory syndrome virus (PRRSV) is a target of diversifying selection as infection progresses from acute infection to virus rebound. *Infect Genet Evol* **40**, 167-175.
- Carrasco Pro, S., Zimic, M., and Nielsen, M. (2014). Improved pan-specific MHC class I peptide-binding predictions using a novel representation of the MHC-binding cleft environment. *Tissue Antigens* **83**, 94-100. DOI: 10.1111/tan.12292.
- Contreras, M., Villar, M., Artigas-Jerónimo, S., Kornieieva, L., Mytروفanov, S., and de la Fuente, J. (2018). A reverse vaccinology approach to the identification and characterization of *Ctenocephalides felis* candidate protective antigens for the control of cat flea infestations. *Parasit Vectors* **11**, 43. DOI: 10.1186/s13071-018-2618-x.
- Correia, B.E., Bates, J.T., Loomis, R.J., Baneyx, G., Carrico, C., Jardine, J.G., Rupert, P., Correnti, C., Kalyuzhnyi, O., Vittal, V., Connell, M.J., Stevens, E., Schroeter, A., Chen, M., Macpherson, S., Serra, A.M., Adachi, Y., Holmes, M.A., Li, Y., Kleivit, R.E., Graham, B.S., Wyatt, R.T., Baker, D., Strong, R.K., Crowe, J.E. Jr., Johnson, P.R., and Schief, W.R. (2014). Proof of principle for epitope-focused vaccine design. *Nature* **507**, 201-2016. DOI: 10.1038/nature12966.
- Crank, M.C., Ruckwardt, T.J., Chen, M., Morabito, K.M., Phung, E., Costner, P.J., Holman, L.A., Hickman, S.P., Berkowitz, N.M., Gordon, I.J., Yamshchikov, G.V., Gaudinski, M.R., Kumar, A., Chang, L.A., Moin, S.M., Hill, J.P., DiPiazza, A.T., Schwartz, R.M., Kueltzo, L., Cooper, J.W., Chen, P., Stein, J.A., Carlton, K., Gall, J.G., Nason, M.C., Kwong, P.D., Chen, G.L., Mascola, J.R., McLellan, J.S., Ledgerwood, J.E., Graham, B.S., and VRC 317 Study Team. (2019). A proof of concept for structure-based vaccine design targeting RSV in humans. *Science*. **365**, 505-509. DOI: 10.1126/science.aav9033.

- Dalsass, M., Brozzi, A., Medini, D., and Rappuoli, R. (2019). Comparison of Open-Source Reverse Vaccinology Programs for Bacterial Vaccine Antigen Discovery. *Front Immunol* *14*, 113. DOI: 10.3389/fimmu.2019.00113.
- Dash, P., Fiore-Gartland, A.J., Hertz, T., Wang, G.C., Sharma, S., Souquette, A., Crawford, J.C., Clemens, E.B., Nguyen, T.H.O., Kedzierska, K., La Gruta, N.L., Bradley, P., and Thomas, P.G. (2017). Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature*. *547*, 89-93. DOI: 10.1038/nature22383.
- Davis, M.M., and Boyd, S.D. (2019). Recent progress in the analysis of $\alpha\beta$ T cell and B cell receptor repertoires. *Curr Opin Immunol* *59*, 109-114.
- DeKosky, B.J., Ippolito, G.C., Deschner, R.P., Lavinder, J.J., Wine, Y., Rawlings, B.M., Varadarajan, N., Giesecke, C., Dörner, T., Andrews, S.F., Wilson, P.C., Hunnicke-Smith, S.P., Willson, C.G., Ellington, A.D., and Georgiou, G. (2013). High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. *Nat Biotechnol* *31*, 166-169.
- DeLisi, C., and Berzofsky, J.A. (1985). T-cell antigenic sites tend to be amphipathic structures. *Proc Natl Acad Sci U.S.A.* *82*, 7048-7052.
- Demasius, W., Weikard, R., Hadlich, F., Müller, K.E., and Kühn, C. (2013). Monitoring the immune response to vaccination with an inactivated vaccine associated to bovine neonatal pancytopenia by deep sequencing transcriptome analysis in cattle. *Vet Res* *44*, 93. DOI: 10.1186/1297-9716-44-93.
- Dhanda, S.K., Usmani, S.S., Agrawal, P., Nagpal, G., Gautam, A., and Raghava, G.P.S. (2017). Novel in silico tools for designing peptide-based subunit vaccines and immunotherapeutics. *Brief Bioinform* *18*, 467-478. DOI: 10.1093/bib/bbw025.
- Diaz-San Segundo, F., Medina, G.N., Stenfeldt, C., Arzt, J., and de Los Santos, T. (2017). Food-and-mouth disease vaccines. *Vet Microbiol* *206*, 102-112. DOI: 10.1016/j.vetmic.2016.12.018.
- Ferdous, S., Kelm, S., Baker, T.S., Shi, J., and Martin, A.C.R. (2019). B-cell epitopes: Discontinuity and conformational analysis. *Mol Immunol* *114*, 643-650. DOI: 10.1016/j.molimm.2019.09.014.
- Fisher, C.R., and Schnell, M.J. (2018). New developments in rabies vaccination. *Rev Sci Technology* *37*, 657-672. DOI: 10.20506/rst.37.2.2831.
- Gilman, M.S., Castellanos, C.A., Chen, M., Ngwuta, J.O., Goodwin, E., Moin, S.M., Mas, V., Melero, J.A., Wright, P.F., Graham, B.S., McLellan, J.S., and Walker, L.M. (2016). Rapid profiling of RSV antibody repertoires from the memory B cells of

- naturally infected adult donors. *Sci Immunol* *1*, pii: eaaj1879. doi: 10.1126/sciimmunol.aaj1879.
- Giuliani, M.M., Adu-Bobie, J., Comanducci, M., Aricò, B., Savino, S., Santini, L., Brunelli, B., Bambini, S., Biolchi, A., Capecchi, B., Cartocci, E., Ciocchi, L., Di Marcello, F., Ferlicca, F., Galli, B., Luzzi, E., Massignani, V., Serruto, D., Veggi, D., Contorni, M., Morandi, M., Bartalesi, A., Cinotti, V., Mannucci, D., Titta, F., Ovidi, E., Welsch, J.A., Granoff, D., Rappuoli, R., and Pizza, M. (2006). A universal vaccine for serogroup B meningococcus. *Proc Natl Acad Sci USA* *103*, 10834-10839.
- Gonzalez-Dias, P., Lee, E.K., Sorgi, S., de Lima, D.S., Urbanski, A.H., Silveira, E.L., and Nakaya, H.I. (2020). Methods for predicting vaccine immunogenicity and reactogenicity. *Hum Vaccin Immunother* *16*, 269-276. DOI: 10.1080/21645515.2019.1697110
- Greenbaum, J.A., Andersen, P.H., Blythe, M., Bui, H.H., Cachau, R.E., Crowe, J., Davies, M., Kolaskar, A.S., Lund, O., Morrison, S., Mumey, B., Ofran, Y., Pellequer, J.L., Pinilla, C., Ponomarenko, J.V., Raghava, G.P., van Regenmortel, M.H., Roggen, E.L., Sette, A., Schlessinger, A., Sollner, J., Zand, M., and Peters, B. (2007). Towards a consensus on datasets and evaluation metrics for developing B-cell epitope prediction tools. *J Mol Recognit* *20*, 75-82.
- Greenwood B. (2014). The contribution of vaccination to global health: past, present and future. *Philos Trans R Soc Lond B Biol Sci* *369*, 20130433. DOI: 10.1098/rstb.2013.0433.
- Gupta, S., Ansari, H.R., Gautam, A., Open Source Drug Discovery Consortium, and Raghava, G.P. (2013). Identification of B-cell epitopes in an antigen for inducing specific class of antibodies. *Biol Direct* *8*, 27. DOI: 10.1186/1745-6150-8-27.
- Gutiérrez, A.H., Martin, W.D., Bailey-Kellogg, C., Terry, F., Moise, L., and DeGroot, A.S. (2015). Development and validation of an epitope prediction tool for swine (PigMatrix) based on the pocket profile method. *BMC Bioinformatics* *16*, 290. DOI: 10.1186/s12859-015-0724-8.
- Hagen, T., and Pulendran, B. (2019). Will systems biology deliver its promise and contribute to the development of new or improved vaccines? *Cold Spring Harb Perspect Biol* *10*, DOI:10.1101/cshperspect.a028894.
- Haining, W.N. (2014). Strength in numbers: comparing vaccine signatures the modular way. *Nat Immunol* *15*, 139-141.

- Hansen, A.M., Rasmussen, M., Svitek, N., Harndahl, M., Golde, W.T., Barlow, J., Vishvanath, N., Buus, S., and Nielsen, M. (2014). *Immunogenetics* 66, 705-718. DOI: 10.1007/s00251-014-0802-5.
- Hassan, A., Naz, A., Obaid A., Paracha, R.Z., Naz, K., Awan, F.M., Muhmmad, S.A., Janjua, H.A., Ahmad, J., and Ali, A. (2016). Pangenome and immuno-proteomics analysis of *Acinetobacter baumannii* strains revealed the core peptide vaccine targets. *BMC Genomics* 17, 732. DOI: 10.1186/s12864-016-2951-4.
- Hewitt, J.S., Karupppannan, A.K., Tan, S., Gauger, P., Halbur, P.G., Gerber, P.F., DeGroot, A.S., Moise, L., and Opriessnig, T. (2019). A prime-boost concept using a T-cell epitope-driven DNA vaccine followed by a whole virus vaccine effectively protected pigs in the pandemic H1N1 pig challenge model. *Vaccine* 37, 4302-4309. DOI: 10.1016/j.vaccine.2019.06.044.
- HIPC-CHI Signatures Project Team and HIPC-I Consortium. (2017). Multicohort analysis reveals baseline transcriptional predictors of influenza vaccination responses. *Sci Immunol* 2, DOI:10.1126/sciimmunol.aal4656.
- Hoof, K., Peters, B., Sidney, J., Pedersen, L.E., Sette, A., Lund, O., Buus, S., and Nielsen, M. (2009). NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics* 61, 1-13. DOI: 10.1007/s00251-008-0341-z.
- Jardetzky, T.S., Brown, J.H., Gorga, J.C., Stern, L.J., Urban, R.G., Strominger, J.L., and Wiley, D.C. (1996). Crystallographic analysis of endogenous peptides associated with HLA-DR1 suggests a common, polyproline II-like conformation for bound peptides. *Proc Natl Acad Sci U.S.A.* 23, 734-738.
- Jespersen, M.C., Peters, B., Nielsen, M., and Marcatili, P. (2017). BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res* 45, W24-W29. DOI: 10.1093/nar/gkx346.
- Jespersen, M.C., Mahajan, S., Peters, B., Nielsen, M., and Marcatili, P. (2019). Antibody Specific B-Cell Epitope Predictions: Leveraging Information From Antibody-Antigen Protein Complexes. *Front Immunol* 10, 298. DOI: 10.3389/fimmu.2019.00298.
- Jurtz, V., Paul, S., Andreatta, M., Marcatilli, P., Peters, B., and Nielsen, M. (2017). NetMHCpan-4.0: improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J Immunol* 199, 3360-3368. DOI: 10.4049/jimmunol.1700893.
- Kapikian, A.Z., Mitchell, R.H., Chanock, R.M., Shvedoff, R.A., and Stewart, C.E. (1969). An epidemiologic study of altered clinical reactivity to respiratory syncytial

- (RS) virus infection in children previously vaccinated with an inactivated RS virus vaccine. *Am J Epidemiol* **89**, 405-421.
- Kringelum, J.V., Lundegaard, C., Lund, O., and Nielsen, M. (2012). Reliable B cell epitope predictions: impacts of method development and improved benchmarking. *PLoS Comput Biol* **8**, e1002829. DOI: 10.1371/journal.pcbi.1002829.
- Kringelum, J.V., Nielsen, M., Padkjær, S.B., and Lund, O. (2013). Structural analysis of B-cell epitopes in antibody:protein complexes. *Mol Immunol* **53**, 24-34. DOI: 10.1016/j.molimm.2012.06.001.
- Li, S., Roupheal, N., Duraisingham, S., Romero-Steiner, S., Presnell, S., Davis, C., Schmidt, D.S., Johnson, S.E., Milton, A., Rajam, G., Kasturi, S., Carlone, G.M., Quinn, C., Chaussabel, D., Palucka, A.K., Mulligan, M.J., Ahmed, R., Stephens, D.S., Nakaya, H.I., and Pulendran, B. (2014). Molecular signatures of antibody responses derived from a systems biology study of five human vaccines. *Nat Immunol* **15**, 195-204. DOI: 10.1038/ni.2789.
- Li, J.N., Zhao, Y.T., Cao, S.L., Wang, H., and Zhang, J.J. (2020). Integrated transcriptomic and proteomic analyses of grass carp intestines after vaccination with a double-targeted DNA vaccine of *Vibrio mimicus*. *Fish Shellfish Immunol* **98**, 641-652. DOI: 10.1016/j.fsi.2019.10.045.
- Lund, O., Nielsen, M., Kesmir, C., Petersen, A.G., Lundegaard, C., Worning, P., Sylvester-Hvid, C., Lamberth, K., Røder, G., Justesen, S., Buus, S., and Brunak, S. (2004). Definition of supertypes for HLA molecules using clustering of specificity matrices. *Immunogenetics* **55**, 797-810. Doi: 10.1007/s00251-004-0647-4.
- Lundegaard, C., Hoof, I., Lund, O., and Nielsen, M. (2010). State of the art and challenges in sequence based T-cell epitope prediction. *Immunome Res* **6**, Suppl 2:S3. DOI: 10.1186/1745-7580-6-S2-S3.
- Luo, J., Carrillo, J.A., Menendez, K.R., Tablante, N.L., and Song, J. (2014). Transcriptome analysis reveals an activation of major histocompatibility complex 1 and 2 pathways in chicken trachea immunized with infectious laryngotracheitis virus vaccine. *Poult Sci* **93**, 848-855. DOI: 10.3382/ps.2013-03624.
- Magro, M., Mas, V., Chappell, K., Vázquez, M., Cano, O., Luque, D., Terrón, M.C., Melero, J.A., and Palomo, C. (2012). Neutralizing antibodies against the preactive form of respiratory syncytial virus fusion protein offer unique possibilities for clinical intervention. *Proc Natl Acad Sci U.S.A.* **109**, 3089-3094. DOI: 10.1073/pnas.1115941109.

- Madden, D.R., (1995) The three-dimensional structure of peptide-MHC complexes. *Annu Rev Immunol* *13*, 587-622.
- Margalit, H., Spouge, J.L., Cornette, J.L., Cease, K.B., Delisi, C., and Berzofsky, J.A.. (1987). Prediction of immunodominant helper T cell antigenic sites from the primary sequence. *J Immunol* *138*, 2213-2229.
- Matthijs, A.M.F., Auray, G., Jakob, V., García-Nicolás, O., Braun, R.O., Keller, I., Bruggman, R., Devriendt, B., Boyen, F., Guzman, C.A., Michiels, A., Haesebrouck, F., Collin, N., Barnier-Quer, C., Maes, D., and Summerfield, A. (2019). Systems Immunology Characterization of Novel Vaccine Formulations for *Mycoplasma hyopneumoniae* Bacterins. *Front Immunol* *10*, 1087. DOI: 10.3389/fimmu.2019.01087.
- Mei, S., Li, F., Leier, A., Marquez-Lago, T.T., Giam, K., Croft, N.P., Akutsu, T., Smith, A.I., Li, J., Rossjohn, J., Purcell, A.W., and Song, J. (2019). A comprehensive review and performance evaluation of bioinformatics tools for HLA class I peptide-binding prediction. *Brief Bioinform* June 14 bbz051. DOI: 10.1093/bib/bbz051.
- Mitsunaga, E.M., and Snyder, M.P. (2020). Deep Characterization of the Human Antibody Response to Natural Infection Using Longitudinal Immune Repertoire Sequencing. *Mol Cell Proteomics* *19*, 278-293. DOI: 10.1074/mcp.RA119.001633.
- Moxon, R., Reche, P.A., and Rappuoli, R. (2019). Reverse vaccinology. *Front Immunol* *3*, 2776. DOI: 10.3389/fimmu.2019.02776.
- Nakaya, H.I., Wrammert, J., Lee, E.K., Racioppi, L., Marie-Kunze, S., Haining, W.N., Means, A.R., Kasturi, S.P., Khan, N., Li, G.M., McCausland, M., Kanchan, V., Kokko, K.E., Li, S., Elbein, R., Mehta, A.K., Aderem, A., Subbarao, K., Ahmed, R., and Pulendran, B. (2011). Systems biology of vaccination for seasonal influenza in humans. *Nat Immunol* *12*, 786-795. DOI: 10.1038/ni.2067.
- Nielsen, M., Connelley, T., and Ternette, N. (2018). Improved prediction of bovine leucocyte antigens (BoLA) presented ligands by use of mass-spectrometry-determined ligand and in vitro binding data. *J Proteome Res* *17*, 559-567. Doi: 10.1021/acs.jproteome.7b00675.
- Novotný, J., Handschumacher, M., Haber, E., Bruccoleri, R.E., Carlson, W.B., Fanning, D.W., Smith, J.A., and Rose, G.D. (1986). Antigenic determinants in proteins coincide with surface regions accessible to large probes (antibody domains). *Proc Natl Acad Sci U.S.A.* *83*, 226-230.
- Oh, J.Z., Ravindran, R., Chassaing, B., Carvalho, F.A., Maddur, M.S., Bower, M., Hakimpour, P., Gill, K.P., Nakaya, H.I., Yarovinsky, F., Sartor, R.B., Gewirtz, A.T.,

- and Pulendran, B. (2014). TLR5-mediated sensing of gut microbiota is necessary for antibody responses to seasonal influenza vaccination. *Immunity* *41*, 478-492. DOI: 10.1016/j.immuni.2014.08.009.
- Painter, C.A., and Stern, L.J. (2012). Conformational variation in structures of classical and non-classical MHCII proteins and functional implications. *Immunol Rev* *250*, 144-157.
- Pandya, M., Rasmussen, M., Hansen, A., Nielsen, M., Buus, S., Golde, W., and Barlow, J. (2015). A modern approach for epitope prediction: identification of food-and-mouth disease virus peptides binding bovine leukocyte antigen (BoLA) class I molecules. *Immunogenetics* *67*, 691-703. Doi: 10.1007/s00251-015-0877-7.
- Pedersen, L.E., Harndahl, M., Nielsen, M., Patch, J.R., Jungersen, G., Buus, S., and Golde, W.T. (2013). Identification of peptides from foot-and-mouth disease virus structural proteins bound by class I swine leukocyte antigen (SLA) alleles, SLA-1*0401 and SLA-2*0401. *Anim Genet* *44*, 251-258. DOI: 10.1111/j.1365-2052.2012.02400.x.
- Pizza, M., Scarlato, V., Masignani, V., Giuliani, M.M., Aricò, B., Comanducci, M., Jennings, G.T., Baldi, L., Bartolini, E., Capecchi, B., Galeotti, C.L., Luzzi, E., Manetti, R., Marchetti, E., Mora, M., Nuti, S., Ratti, G., Santini, L., Savino, S., Scarselli, M., Storni, E., Zuo, P., Broecker, M., Hundt, E., Knapp, B., Blair, E., Mason, T., Tettelin, H., Hood, D.W., Jeffries, A.C., Saunders, N.J., Granoff, D.M., Venter, J.C., Moxon, E.R., Grandi, G., and Rappuoli, R. (2000). Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science* *10*, 1816-1820.
- Pretorius, A., Faber, F.E., and van Kleef, M. (2016). Immune gene expression profiling of PBMC isolated from horses vaccinated with attenuated African horsesickness virus serotype 4. *Immunobiology* *221*, 236-244. DOI: 10.1016/j.imbio.2015.09.002
- Pulendran, B., Li, S., and Nakaya, H.I. (2010). Systems vaccinology. *Immunity* *33*, 516-529. DOI: 10.1016/j.immuni.2010.10.006.
- Querec, T.D., Akondy, R.S., Lee, E.K., Cao, W., Nakaya, H.I., Teuwen, D., Pirani, A., Gernert, K., Deng, J., Marzolf, B., Kennedy, K., Wu, H., Bennouna, S., Oluoch, H., Miller, J., Vencio, R.Z., Mulligan, M., Aderem, A., Ahmed, R., and Pulendran, B. (2009) Systems biology approach predicts immunogenicity of the yellow fever vaccine in humans. *Nat Immunol* *10*, 116-125. DOI: 10.1038/ni.1688.

- Rahman, Kh.S., Chowdhury, E.U., Sachse, K, and Kaltenboeck, B. (2016). Inadequate Reference Datasets Biased toward Short Non-epitopes Confound B-cell Epitope Prediction. *J Biol Chem* 291, 14585-14599. DOI: 10.1074/jbc.M116.729020.
- Rammensee, H.G., Friede, T., and Stevanoviic, S. (1995). MHC ligands and peptide motifs: first listing. *Immunogenetics* 41, 178-228.
- Rappuoli, R. (2000). Reverse vaccinology. *Curr Opin Microbiol* 3, 445-450.
- Rappuoli, R. (2007). Bridging the knowledge gaps in vaccine design. *Nat Biotechnol.* 25, 1361-1366.
- Rappuoli, R., Bottomley, M.J., D'Oro, U., Finco, O., and De Gregorio, E. (2016). Reverse vaccinology 2.0: Human immunology instructs vaccine antigen design. *J Exp Med* 213, 469-481. DOI: 10.1084/jem.20151960.
- Reyes, V.E., Chin, L.T., and Humphreys, R.E. (1988). Selection of class I MHC-restricted peptides with the strip-of-helix hydrophobicity algorithm. *Mol Immunol* 25, 867-871.
- Robinson, J., Guethlein, L.A., Cereb, N., Yang, S.Y., Norman, P.J., Marsh, S.G.E., and Parham, P. (2017). Distinguishing functional polymorphism from random variation in the sequences of >10,000 HLA-A, -B and -C alleles. *PLoS Genet* 13, e1006862. DOI: 10.1371/journal.pgen.1006862.
- Rock, K.L., Reits, E., and Neefjes, J. (2016). Present Yourself! By MHC Class I and MHC Class II Molecules. *Trends Immunol* 37, 724-737. DOI: 10.1016/j.it.2016.08.010.
- Saada, R., Weinberger, M., Shahaf, G., and Mehr, R. (2007). Models for antigen receptor gene rearrangement: CDR3 length. *Immunol Cell Biol* 85, 323-332.
- Sant, S., Grzelak, L., Wang, Z., Pizzolla, A., Koutsakos, M., Crowe, J., Loudovaris, T., Mannering, S.I., Westall, G.P., Wakim, L.M., Rossjohn, J., Gras, S., Richards, M., Xu, J., Thomas, P.G., Loh, L., Nguyen, T.H.O., and Kedzierska, K. (2018). Single-Cell Approach to Influenza-Specific CD8+ T Cell Receptor Repertoires Across Different Age Groups, Tissues, and Following Influenza Virus Infection. *Front Immunol* 9, 1453. DOI: 10.3389/fimmu.2018.01453.
- Saravanan, V. and Gautham, N. (2018). BCIGEPRED-a dual-layer approach for predicting linear IgE epitopes. *Mol Biol (Mosk)* 52, 333-343. DOI: 10.7868/S0026898418020180.

- Sela-Culang, I., Ofan, Y., and Peters, B. (2015). Antibody specific epitope prediction-emergence of a new paradigm. *Curr Opin Virol* *11*, 98-102. DOI: 10.1016/j.coviro.2015.03.012.
- Serruto, D., Bottomley, M.J., Ram, S., Giuliani, M.M., and Rappuoli, R. (2012). The new multicomponent vaccine against meningococcal serogroup B, 4CMenB: immunological, functional and structural characterization of the antigens. *Vaccine* *30*, 30. DOI: 10.1016/j.vaccine.2012.01.033.
- Stave, J.W. and Lindpainter, K. (2013). Antibody and antigen contact residues define epitope and paratope size and structure. *J Immunol* *191*, 1428-1435. DOI: 10.4049/jimmunol.1203198.
- Stern, L.J., Brown, J.H., Jardetzky, T.S., Gorga, J.C., Urban, R.G., Strominger, J.L., and Wiley, D.C. (1994). Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. *Nature* *368*, 215-221.
- Stille, C.J., Thomas, L.J., Reyes, V.E., and Humphreys, R.E. (1987). Hydrophobic strip-of-helix algorithm for selection of T cell-presented peptides. *Mol Immunol* *24*, 1021-1027.
- Sundling, C., Martinez, P., Soldemo, M., Spångberg, M., Bengtsson, K.L., Stertman, L., Forsell, M.N., and Karlsson Hedestam, G.B. (2013). Immunization of macaques with soluble HIV type 1 and influenza virus envelope glycoproteins results in a similarly rapid contraction of peripheral B-cell responses after boosting. *J Infect Dis* *207*, 426-431. DOI: 10.1093/infdis/jis696.
- Svitek, N., Hansen, A.M., Steinaa, L., Saya, R., Awino, E., Nielsen, M., Buus, S., and Nene, V. (2014). Use of "one-pot, mix-and-read" peptide-MHC class I tetramers and predictive algorithms to improve detection of cytotoxic T lymphocyte response in cattle. *Vet Res* *45*, 50. DOI: 10.1186/1297-9716-45-50.
- Terasaki, P.I. (2007). A brief history of HLA. *Immunol Res* *38*, 139-148. DOI: 10.1007/s12026-007-0020-4
- Tettelin, H., Saunders, N.J., Heidelberg, J., Jeffries, A.C., Nelson, K.E., Eisen, J.A., Ketchum, K.A., Hood, D.W., Peden, J.F., Dodson, R.J., Nelson, W.C., Gwinn, M.L., DeBoy, R., Peterson, J.D., Hickey, E.K., Haft, D.H., Salzberg, S.L., White, O., Fleischmann, R.D., Dougherty, B.A., Mason, T., Ciecko, A., Parksey, D.S., Blair, E., Citterone, H., Clark, E.B., Cotton, M.D., Utterback, T.R., Khouri, H., Qin, H., Vamathevan, J., Gill, J., Scarlato, V., Massignani, V., Pizza, M., Grandi, G., Sun, L., Smith, H.O., Fraser, C.M., Moxon, E.R., Rappuoli, R., and Venter, J.C. (2000).

- Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. *Science* *287*, 1809-1815.
- Tomazic, M.L., Rodriguez, A.E., Lombardelli, J., Poklepovich, T., Garro, C., Galarza, R., Tiranti, K., Florin-Christensen, M., and Schnittger, L. (2018). Identification of novel vaccine candidates against cryptosporidiosis of neonatal bovines by reverse vaccinology. *Vet Parasitol* *264*, 74-78. DOI: 10.1016/j.vetpar.2018.11.007.
- Vahey, M.T., Wang, Z., Kester, K.E., Cummings, J., Heppner, D.G. Jr, Nau, M.E., Ofori-Anyinam, O., Cohen, J., Coche, T., Ballou, W.R., and Ockenhouse, C.F. (2010). Expression of genes associated with immunoproteasome processing of major histocompatibility complex peptides is indicative of protection with adjuvanted RTS,S malaria vaccine. *J Infect Dis* *201*, 580-589. DOI: 10.1086/650310.
- Van Regenmortel, M.H. (2014). Specificity, polyspecificity, and heterospecificity of antibody-antigen recognition. *J Mol Recognit* *27*, 627-639. DOI: 10.1002/jmr.2394.
- Waickman, A.T., Victor, K., Li, T., Hatch, K., Rutvisuttinunt, W., Medin, C., Gabriel, B., Jarman, R.G., Friberg, H., and Currier, J.R. (2019). Dissecting the heterogeneity of DENV vaccine-elicited cellular immunity using single-cell RNA sequencing and metabolic profiling. *Nat Commun* *10*, 3666. DOI: 10.1038/s41467-019-11634-7.
- Welly, B.T., Miller, M.R., Stott, J.L., Blanchard, M.T., Islas-Trejo, A.D., O'Rourke, S.M., Young, A.E., Medrano, J.F., and Van Eenennaam, A.L. (2017). Genome Report: Identification and Validation of Antigenic Proteins from *Pajaroellobacter abortibovis* Using De Novo Genome Sequence Assembly and Reverse Vaccinology. *G3*. *7*, 321-331. DOI: 10.1534/g3.116.036673.