



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Vaccine Target Discovery

Li C Chong, Universiti Putra Malaysia, Serdang, Malaysia

Asif M Khan, Perdana University, Selangor, Malaysia and John Hopkins University, Baltimore, MD, United States

© 2019 Elsevier Inc. All rights reserved.

Introduction

Vaccination is one of the most efficacious medical interventions that have decreased human morbidity and mortality in all regions of the world. It not only has dramatically reduced the incidence of numerous diseases (Hilleman, 1985; Andre *et al.*, 2008), such as measles, diphtheria, mumps, rubella, tetanus, yellow fever, pertussis, and poliomyelitis, but also eradicated the dreaded smallpox viral infection (Strassburg, 1980). Efforts are currently also focused on developing vaccines for treatment of other diseases (Ada, 2003). This triumph over infectious diseases has been achieved by using either killed or attenuated conventional vaccines. A conventional or prophylactic vaccination is based on deliberate exposure to non-virulent form of the pathogen to establish immunity from subsequent exposure to the virulent form of the pathogen (Lee *et al.*, 2012). This approach requires little knowledge of the molecular nature of the individual pathogen antigens or the immune responses they elicit. To date, this approach has undoubtedly been the most successful.

There is, however, an ongoing trend towards emerging infectious diseases (Fauci, 2001; Fauci *et al.*, 2005) against humankind in different parts of the world. These diseases are caused by the emergence of new pathogens, resurgence of old ones, constantly mutating pathogens, drug-resistant pathogens, and even include pathogens used as agents of bioterrorism. Previously undescribed pathogens, such as severe acute respiratory syndrome (SARS), bird flu (avian influenza), mad cow, and HIV/AIDS, are appearing at an increasing frequency. On the other hand, old diseases such as ebola, hanta, dengue and cholera, among others, are invading populations from which they have disappeared or cross species barrier to invade new species. Occasionally, pathogens mutate (or recombine) and then they adapt and show up as new variants that evade the host's acquired immunity. The annual influenza epidemics are generally due to genetically drifting strains of influenza that differ slightly from previous strains (Bouvier and Palese, 2008). The development of drug-resistance pathogens, such as malaria, pneumococci, enterococci, and tuberculosis, has increased over the years, partly due to the widespread and inappropriate use of drugs (Knobler *et al.*, 2003). Moreover, the pathogens responsible for the emerging infectious diseases are also of potential use as bioterrorist weapon (Ryan, 2008). For these reasons, emerging infectious diseases continue to pose threats to public health (Morens and Fauci, 2013). Therefore, there is a need for more effective vaccines to help reduce human morbidity and mortality from emerging infectious diseases.

The last decade has seen significant advances in new technologies for the development of new vaccines. These technologies, combined with our understanding of host response to foreign antigens, have laid the foundation for rapid advances in vaccinology. Few potential candidate approaches for this new family of vaccines (Arnon and Ben-yedidia, 2003; Minichiello, 2002; Babiuk, 1999; Nandy and Basak, 2016) are subunit vaccines, genetically engineered live vaccines, and polynucleotide (DNA or genetic) vaccines. All these new directions in vaccine development have something in common; their most important challenge is the discovery of key antigens from the array of proteins encoded by the pathogen genome that are able to elicit protective immune response against these pathogens and are effective for majority of the human population. The presence of genetic variation in the genes of the host immune system across the human population and the genomes of the pathogen variants make this a multi-dimensional and a combinatorial problem.

Since 1980s, the focus of vaccine design has been on the pathogen's variable domains, mutants, multivalent coverage and not so much on conserved regions (Plotkin, 2005). However, a large body of data is building up in the field of vaccine design and epitope prediction that points to the neglected aspect of conserved epitopes as important targets of vaccine development. In fact, evidence is accumulating that while many variable regions are highly antigenic, sufficiently large number of them are actually non-immunoprotective and have been exploited by viruses and other pathogens for immune escape and may lead to immunopathology (Haydon and Woolhouse, 1998). Conserved epitopes were thought to be un-important due to their lack of immunogenicity (Bona *et al.*, 1998; Li *et al.*, 2011). However, the immunogenicity of such conserved epitopes especially those which are immune-protective and conserved amongst many variants and mutant strains, can be boosted using adjuvants, which are chemicals or approved drug molecules. Moreover, conserved epitopes also help address the issues of ethnicity ("pan-haplotype responsive") compared to variable epitopes (Khan *et al.*, 2008). In addition, the conserved epitopes can circumvent vaccine design issues of escape mutants and the need of using the latest strains. In particular, if a set of conserved, immunogenic, and immunoprotective epitopes are suitably boosted, they will not elicit mutations in the pathogen, and thus can be reasonably predicted to remain unchanged in the next season of infection. The possibility of extended efficacy of vaccines would be a tremendous advance for the vaccine industry and will potentially serve the global population with protection against infections (Sylvester-Hvid *et al.*, 2002; De Groot *et al.*, 2004; Sette *et al.*, 2001; Raman *et al.*, 2014; Khan *et al.*, 2017).

Vaccine informatics, a fledgling sub-field of reverse vaccinology, has the potential to develop effective vaccines (Hegde *et al.*, 2017; Khan *et al.*, 2017). With the rapid expansion of vaccine related data (host and pathogen) stemming from both classical and high-throughput genomic/proteomic approaches, identifying conserved, robust, immunogenic, and immunoprotective epitopes manually from this large data pool is inefficient. Vaccine informatics is a practical science for designing new vaccines with a focus

on bioinformatics-driven acquisition, manipulation and analysis of data related to the immune system and disease agents (Raman *et al.*, 2014). It provides a means for systematic study of big data, pre-screening of targets, and facilitates experimental design for validation by a small number of key experiments. The bioinformatics support can be divided into two, the standard bioinformatics support and the more specialised immunoinformatics support (Petrovsky *et al.*, 2003). The standard support includes basic bioinformatics functions, such as sequence comparison and alignment, database searching, hunting for patterns and profiles, 3D-structure analysis and modeling, and data annotation (reviewed in Brusica and Petrovsky (2003)). Immunoinformatics is a more targeted bioinformatics support with an emphasis on data-warehousing and mining of immunological data, such as prediction of immunogenicity (Soria-Guerra *et al.*, 2015; Brusica and Petrovsky, 2003). Vaccine researchers are taking advantage of these bioinformatics approaches, in combination with experimental validation, to discover and facilitate better understanding of the components of the human immunome, which then aid in the design of new vaccines. The immunome can be defined as the complete set of genes and proteins of the immune system. Highly accurate target predictions can diminish discovery cost by 10–20 folds (De Groot *et al.*, 2002; Kast *et al.*, 1994).

Background

In the human and higher vertebrate host, the major functions of the immune system are the maintenance of homeostasis, surveillance and tolerance to self-structures, and defence followed by immunity against pathogens (Yatim and Lakkis, 2015). The immune system is widely distributed in the body and comprises of immune organs, tissues, and cells, connected as a complex, but tightly regulated network (Jerne, 1993; NIH, 2003; Nicholson, 2016). In general, the processes that take place at the molecular level and cellular level largely initiate and regulate the function of the immune system. A healthy immune system will discriminate 'non-self' or foreign antigenic proteins from those that are normally present ('self') in an organism and will raise appropriate responses. The number of self-structures is large, but finite; the number of non-self structures is practically infinite. In the cell, both self and non-self proteins are digested by the proteasomes into short peptide fragments, which are then bound by major histocompatibility complex (MHC) molecules to form peptide/MHC complexes and are displayed on the surface of host cells (Vigneron and Van den Eynde, 2014). These peptides are recognition labels, which display the contents of host cells to T cells of the immune system. The presence of non-self peptides is a prerequisite for the initiation of immune responses. Peptides produced by degradation of intracellular proteins bind MHC class I molecules and are recognised by CD8⁺ T cell receptor (Shastri *et al.*, 2002; Chowell *et al.*, 2015; Blum *et al.*, 2013). MHC class II molecules present peptides, produced by degradation of proteins of extracellular origin, on the surface of antigen-presenting cells to CD4⁺ T cells (reviewed in Lennon-Duménil *et al.* (2002)). A major function of CD8⁺ T cells is to recognize and destroy cells infected by pathogens (Nicholson, 2016). Peptides displayed by the MHC class II molecules mainly serve to regulate immune responses; they are crucial for the initiation, enhancement and suppression of immune responses.

Driven by methods of molecular and cell biology, significant advances in the understanding of immunological processes have been made during the last two decades. This progress over the years resulted in the continuous accumulation of huge amount of immunological data obtained experimentally. This growing number of immunological data and the high complexity of the functional and structural foundation of the immune processes created a need for improved data management to enable advance data analysis. The need to manage and analyse this growing amount of complex data has led to the development of a number of immunological databases (Brusica *et al.*, 2000), such as SYFPEITHI and MCHPEP, IMGT and FIMM, and complex computational models (Petrovsky and Brusica, 2002). The purpose of immunological databases is to facilitate the collection of, access to, and use of immunologically relevant data. One of the applications of the databases in immunology is for vaccine research and development by using complex computational models in combination with experimental approaches to help precise our understanding of antigen presentation and recognition by the immune system.

The combined effort between experimental and computational immunology provided us with a new perspective to designing vaccines that will be effective across demographic boundaries. Previously, the extreme degree of polymorphism observed in the MHC posed limitations for the development of such a vaccine because the ability to trigger an effective T-cell response is partly determined by the MHC phenotype of the individual and different individuals have different MHC allele (Macdonald *et al.*, 2001; Marrack *et al.*, 2017). The MHC genes are the most polymorphic of all human genes, with more than 10,000 alleles known (as of Feb 2018), and are important in increasing the range of responses that different individuals can mount. In humans, this polymorphism results from concentrated amino acid substitutions in the peptide-binding groove of human leukocyte antigen (HLA, the human MHC system) molecules that produce variability in peptide binding and presentation to T cells (MacDonald *et al.*, 2000). Over the years, few groups have investigated the possibility of a functional classification of HLA polymorphism based on peptide-binding specificities. It was found that majority of HLA alleles (both class I and II) could be grouped into 18 or more different 'supertypes', purely on the basis of similarities in their peptide binding specificity (Lund *et al.*, 2004). It has been suggested that the majority of all major human populations can be covered by only few HLA supertypes, where the different members of each supertype bind similar peptides ('promiscuous peptides') for presentation to T-cell (Sette *et al.*, 1999). Further, latest developments show evidence for presence of "immunological hot-spots" (Srinivasan *et al.*, 2004) in antigens. Immunological hot-spots are defined as antigenic regions possessing multiple promiscuous peptides that are supertype specific. This area of research raises the prospect of identifying key pathogenic antigens that possess immunological hot-spots as best candidates for vaccine design as it will provide protection at the population level, irrespective of ethnicity.

The humoral response involves antibodies produced by B-cells, which recognize both linear and conformational B-cell epitopes on the surface of the pathogen. Conformational neutralizing epitopes are the primary focus of various vaccine research for protective humoral responses. However, unlike linear B-cell epitopes and T-cell epitopes, reliable computational tools for prediction of conformational epitopes are limited (Kulkarni-Kale *et al.*, 2005; Zhang *et al.*, 2011).

As for the pathogens, the last decade witnessed the rapid expansion of sequence data at our disposal, stemming from both genomic and proteomic approaches, enabling analysis to map key antigens that are potential targets for protective immune responses. The genomic sequences of a large number of pathogens that threaten public health have been completed or are impending completion. For example, the whole genome of over 7,475 viruses have been sequenced and deposited in the major public database Entrez Genome (see "Relevant Website section"). The data derived from the genome/proteome sequencing and related projects for a particular species of pathogen, such as West Nile Virus, is the missing gap towards the development of vaccines effective against the majority of the variants currently known within that pathogenic species and probably against novel ones that are yet to emerge (Koo *et al.*, 2009). Such vaccines will be much superior to the current generation of vaccines, which are solely based on a single or few antigens providing protection only against certain variants of a pathogen and therefore might elicit too narrow a breadth of response to provide protection from the remaining diverse variants of the same pathogen species (Doolan, 2003).

We have made significant progress in understanding the processes in the host that are involved in mounting an immune response. However, we are still far from having a good understanding of the natural complexity of the pathogens. A good starting point would be by utilizing the pathogens genomic/proteomic data to study their sequence diversity, and identify antigens containing conserved and variable immunological hot-spots. To fully realize the promise of the available datasets for such study, we would require the development of appropriate technologies for systematically converting genomic/proteomic data into protective vaccines. Recently, systematic genome-wide approach to identify the key antigens of a pathogenic species from the numerous variant sequences of the same pathogen have been reported (Dhanda *et al.*, 2017; Rizwan *et al.*, 2017; Goodswen *et al.*, 2014; Vivona *et al.*, 2006; Del Tordello *et al.*, 2016; Maria *et al.*, 2017; Doolan, 2003; Raman *et al.*, 2014; Koo *et al.*, 2009; Khan *et al.*, 2017). The selected key antigens will represent the minimal representative sets of target sequences required to provide immunity against the majority of the existing variants of a particular pathogen species.

Bioinformatics is an inter-disciplinary field that is essential for the analysis and interpretation of complex and large quantity of biological data generated by functional studies and high throughput technologies. It is used to propose the next sets of experiments and, most importantly, to derive better understanding of biological processes. As stated, the number of pathogen sequence data in public databases is increasing rapidly, however, experimental approaches to study this large data pool for the development of immune interventions are time-consuming, costly and almost impractical. Through combination of bioinformatics and experimental approaches, it is possible to select key experiments and help optimize experimental design. Computer algorithms are increasingly used to speed-up the process of knowledge discovery by helping to identify critical experiments for testing hypothesis built upon the result of computational screening. A number of successful examples for application of computer models to study immunological problems have been described in Brusica *et al.* (2005). Such examples illustrate the power of computational approach to complex problems involving potentially vast datasets with potential biases, errors and discrepancies.

A Computational Framework for Vaccine Target Discovery

Reverse vaccinology, a bottom-up genomic approach, has been successfully applied to the development of vaccines against pathogens that were previously not suited to such development (Vernikos, 2008; Rappuoli and Covacci, 2003; Rappuoli, 2001; Del Tordello *et al.*, 2016). The pre-requisite for this approach is the sequence data of the target pathogen, which acts as input to various bioinformatics algorithms for prediction of putative antigens that are likely to be successful vaccine targets. These candidates can then be validated by a small number of key experiments in the lab. The approach has been successfully applied to the development of universal vaccines against group B Streptococcus (Maione *et al.*, 2005) and vaccine candidates against MenB (Pizza *et al.*, 2000), among others (Rappuoli and Covacci, 2003). Reverse vaccinology is a promising method for the high-throughput discovery of candidate vaccine targets that have the potential to mirror the dynamics and antigenic diversity of the target pathogen population, which includes the diversity of the interacting partner, the immune system. However, a big challenge to this end is the need to understand how vaccine developers can cover antigenic diversity and develop a systematic approach to rationally screen pathogen data to select candidate vaccine targets that cover the diversity.

Over the years, a number of bioinformatics pipelines have been designed that predict vaccine candidates both rapidly and efficiently. VacSol (Rizwan *et al.*, 2017), NERVE (Vivona *et al.*, 2006), and Vacceed (Goodswen *et al.*, 2014) are examples of such pipelines for proteomes of bacterial or eukaryotic pathogens and these pipelines are highly configurable and scalable (Zaharieva *et al.*, 2017). They include multiple steps and integrate various algorithms for analysis and comparison. Shortlisted candidate vaccine targets are ranked for prioritization towards experimental validation. These pipelines are expected to improve the vaccine target discovery process.

The general characteristics desired for a candidate vaccine target are (i) highly conserved; (ii) pathogen-specific; (iii) important for structure/function; (iv) immune-relevant; and (v) antigenically similar to circulating strains. Highly conserved targets are less likely to mutate and escape immune recognition. This is particularly so if they are important for structure and function, suggesting a robust historical conservation. High conservation also reduces the possibility of altered-peptide ligand (APL) effect from variant epitopes of the same pathogen species (Sloan-Lancaster and Allen, 1996; Evavold *et al.*, 1993; Rothman, 2004). Variants may also

originate from other pathogens, in particular those that co-circulate or co-infect with the pathogen of interest and if they belong to the same family. In vaccine design, epitopes common to other pathogens could either be useful by inducing cross-protection, or detrimental by inducing altered-ligand effect. Thus, potential vaccine targets should be analyzed for specificity to the target pathogen. The definition of virus species-specific vaccine targets can be further expanded to exclude those with one amino acid mismatch to human sequences in order to avoid possibility of molecular mimicry. Antigenic mismatch between a vaccine and circulating strains has been shown to increase the risk of disease outbreak by 1000-fold compared to immunization using identical strains (Park *et al.*, 2004). Thus, it is important to assess the extent of antigenic identity between the candidate vaccine targets and the circulating strains.

Typically, a generic semi-automated computational framework comprises of three key components: data collection, data processing, and data analysis (Khan, 2005, 2009; Khan *et al.*, 2006, 2017). Fig. 1 illustrates the workflow of the framework and Fig. 2 provides a non-comprehensive list of commonly used tools. Data collection would involve the user providing a set of sequences (aligned or unaligned) of the pathogen of interest for analysis. The sequences could be a single protein dataset, multiple or the complete proteome. Additionally, comparative analyses of the sequences can be performed between subtypes/groups of the pathogen. The sequences can be retrieved from public repositories, primary or specialist databases, such as the NCBI Entrez Protein database (NCBI Resource Coordinators, 2017) or Influenza Research Database (IRD) (Zhang *et al.*, 2017), respectively. User can also provide sequences derived from their experimental work, not available in public databases. Typically, downloaded data would comprise full-length or partial sequences, and the corresponding metadata. Data processing would involve removal of duplicate sequences from the dataset, and for comparative analysis, the merging of the input sequences will be required prior to the alignment step. Multiple sequence alignment will be carried out using an existing tool that is robust in dealing with large sequence data, such as (Sievers *et al.*, 2011; Katoh *et al.*, 2017; Edgar, 2004; Do *et al.*, 2005) Clustal Omega, MAFFT, MUSCLE or PROBCONS. The output alignment quality will be manually inspected for any errors and/or misalignments, which are common when dealing with partial sequences. Henceforth, the data would be ready for analyses, which can involve performing a diversity analysis, such as by measuring entropy values (Heiny *et al.*, 2007; Hu *et al.*, 2013; Khan *et al.*, 2008; Koo *et al.*, 2009) and quantifying variant motifs for each, user-defined, k-mer positions in the alignment. The results of these analyses will be plotted as an output for the user, providing a holistic view of the diversity, including variant distribution. The user can define a preferred conservation threshold for selection of highly conserved sequences. These selected sequences are then analyzed for distribution of variants in nature (Khan *et al.*, 2008; Koo *et al.*, 2009), including matches to human proteins, enabling the identification of pathogen specific, highly conserved sequences. The robust nature of the historical conservation can be assessed by performing functional and structural analysis (Sprenger *et al.*, 2008; Hung and Link, 2011; Wizemann *et al.*, 1999; Sachdeva *et al.*, 2005; Monterrubio-López *et al.*, 2015; He, 2014; Tang *et al.*, 2014; Maria *et al.*, 2017). The relevance of the identified potential candidate vaccine targets for use against current circulating strains can be assessed by measuring the incidence of the candidate targets in the corresponding sequences of recent strains of the virus of interest (Khan *et al.*, 2017). The antigenicity/immunogenicity of the candidate sequences are predicted to assess their immune relevance. Experimental validation includes matching the predicted epitopes with reported epitopes in public databases, such as the Immune Epitope Database (IEDB) or SYFPEITHI, or performing quantitative measurements of the pre-selected candidate peptides by generating synthetic constructs and testing for their immunogenicity, such as by use of HLA transgenic animal models that express specific HLA alleles (Rosloniec *et al.*, 1997; Lefranc *et al.*, 2009; Gourlay *et al.*, 2017; Khan, 2005).

Immunoinformatics

Bioinformatics tools can facilitate the process of epitope mapping by identifying peptides that can potentially elicit T-cell responses. Binding of epitopes to HLA antigens is highly allele-specific; core peptide-binding motifs (usually between 8- and 11-mers, most often 9-mers) have been defined experimentally for a number of HLA class I and class II alleles and incorporated into computational algorithms, allowing to predict candidate HLA-binding epitopes *in silico* from protein sequences (Parker *et al.*, 1994; Rammensee *et al.*, 1999; Sturniolo *et al.*, 1999; Zhang *et al.*, 2005; Nielsen *et al.*, 2010; Karosiene *et al.*, 2013; Paul *et al.*, 2013, 2015a,b; Andreatta *et al.*, 2015; Andreatta and Nielsen, 2015; Pro *et al.*, 2015; Trolle *et al.*, 2015; Abelin *et al.*, 2017; Jurtz *et al.*, 2017; Fleri *et al.*, 2017). More recently, quantifiable predictive features of TCR $\alpha\beta$ binding to HLA/epitope complexes have been also described (Birnbbaum *et al.*, 2014; Dash *et al.*, 2017; Glanville *et al.*, 2017; Gee *et al.*, 2017).

Two main categories of specialized immunoinformatics tools are available for prediction of MHC binding peptides – methods based on identifying patterns in sequences of binding peptides, and those that employ three-dimensional (3D) structures to model peptide/MHC interactions (Tong *et al.*, 2007; Liljeroos *et al.*, 2015; Gourlay *et al.*, 2017). Pattern-based methods includes binding motifs, quantitative matrices, decision trees, artificial neural networks (ANNs), hidden Markov models (HMMs) and support vector machines (SVMs), among others. In contrast, the structure based methods are theoretically rooted and include homology modeling, docking and 3D threading techniques (Dominguez *et al.*, 2003; De Vries *et al.*, 2010; Agostino *et al.*, 2016; Khan and Ranganathan, 2010; Liljeroos *et al.*, 2015; Gourlay *et al.*, 2017). Although less accurate, pattern based approaches are over-represented in the literature due to higher complexity in development and longer computational time of the more accurate structure-based approaches (Ranganathan and Tong, 2007; Maria *et al.*, 2017), including the sheer difference in the availability of linear versus structural data.

For a given sequence, typically all possible overlapping 9-mer (and later, 8- to 11- mer) peptide sequences are extracted. Epitope prediction for HLA class I and class II alleles are performed using benchmarked prediction models, including, for HLA

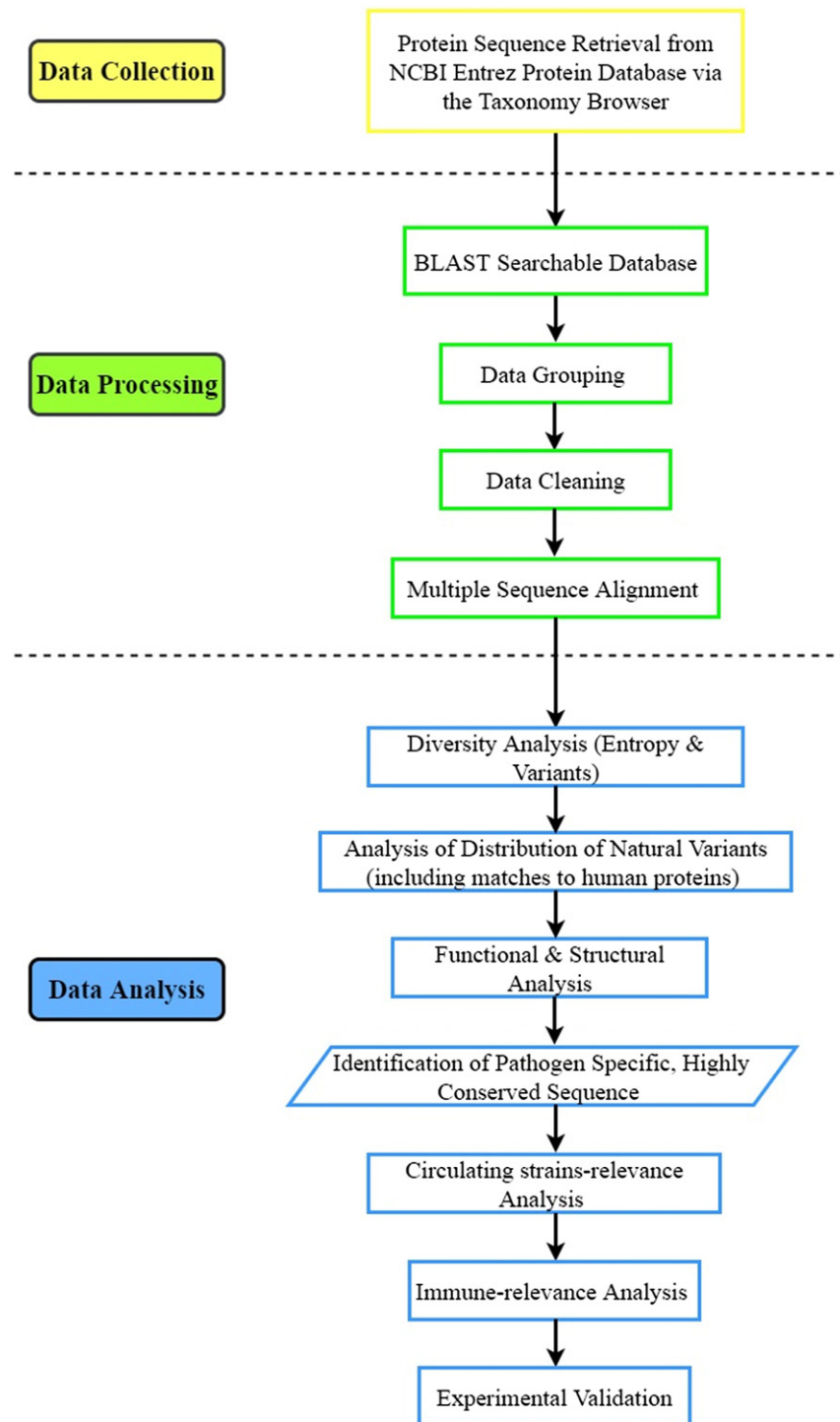


Fig. 1 Computational framework for vaccine target discovery.

class I epitopes, the artificial neural network (ANN)-trained NetMHCpan (version 4.0) (Andreatta *et al.*, 2015; Trolle *et al.*, 2015; Jurtz *et al.*, 2017), and for HLA class II epitopes, NetMHCIIpan (Karosiene *et al.*, 2013; Andreatta and Nielsen, 2015), and the allele-specific consensus percentile ranks of all algorithms queried by the Immune Epitope Database and Analysis Resource (IEDB) tools (combination of NN-align, SMM-align, and ComLib/Sturniolo) (Paul *et al.*, 2015a,b). Additionally, proteasome cleavage (Hakenberg *et al.*, 2003; Nussbaum *et al.*, 2001; Nielsen *et al.*, 2005) and TAP binding predictions (Zhang *et al.*, 2006; Bhasin *et al.*, 2007; Bhasin and Raghava, 2004) will be performed for priority ranking. Several tools, such as NetCTL (Larsen *et al.*, 2005), integrate these various predictions into one and predict for HLA supertypes, which are groups of HLA alleles with similar peptide

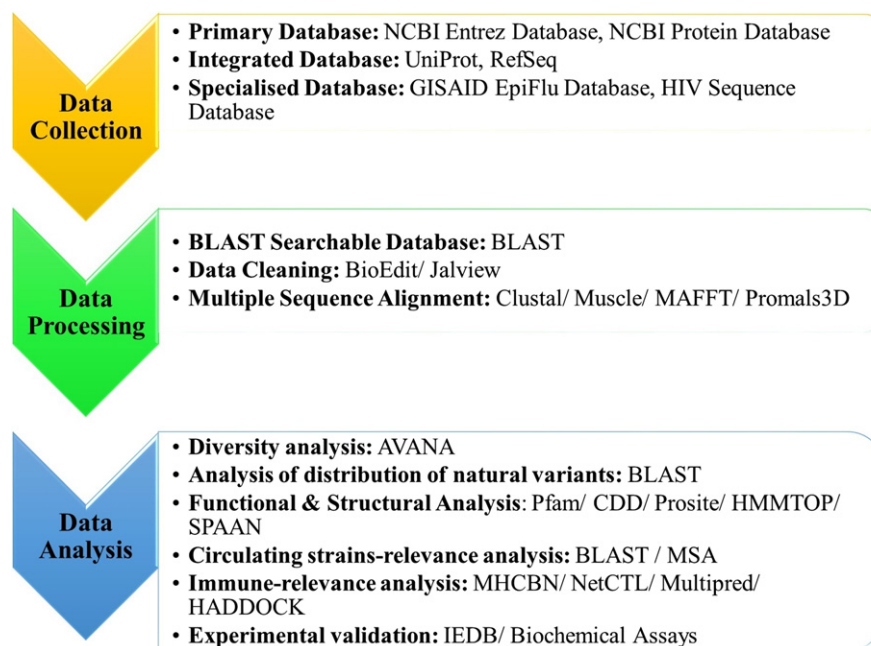


Fig. 2 An example of commonly used tools and databases for vaccine target discovery.

binding specificity. Potentially cross-reactive self epitopes may be searched within the human proteome using BLAST/BLAT alignment tools (Kent, 2002; Altschul *et al.*, 1990).

Case Studies: Vaccine on Infectious Diseases and Non-Infectious Diseases

Infectious Diseases

Reverse vaccinology immunoinformatics approaches are widely applied for viral vaccine design, such as for influenza virus, chikungunya virus, zika virus and others (Gupta *et al.*, 2016; María *et al.*, 2017), including parasites (Damfo *et al.*, 2017) and bacteria (Mistry and Flower, 2017; Zahroh *et al.*, 2016; Rappuoli, 2001). Khan *et al.* developed a bioinformatics pipeline for DENV, which proved generic as it was successfully applied to several viruses, such as WNV (Koo *et al.*, 2009), a close relative of DENV (Khan *et al.*, 2008), and a number of other viruses, such as HIV-1 (Hu *et al.*, 2013), among others. It provides a novel and generalized approach to the formulation of peptide-based vaccines targeting a broad diversity of pathogens and applicable to the human population at large. This methodology is a significant contribution to the field of reverse vaccinology as it enables the systematic screening and analyses of pathogen data which would otherwise be impossible to carry out experimentally, due to too many pathogen sequences (high viral diversity) and variations in immune system among individuals (extensive polymorphism of HLA). This approach therefore significantly reduces the efforts and cost of experimentation, while providing for systematic screening and analyses of pathogen proteomes (Raman *et al.*, 2014).

Khan *et al.* (2008), Koo *et al.* (2009) and Hu *et al.* (2013) analyzed a large number of dengue (DENV), West Nile virus (WNV) and clade B HIV-1, sequences, respectively, retrieved from the NCBI Entrez protein database (Table 1). The sequences were aligned and the overlapping nonamer amino acid positions of the viral proteome, each a possible core binding domain for human leukocyte antigen molecules and T-cell receptors, were quantitatively analyzed. The mean entropy of DENV nonamer sequences was low, with a range of 0.2–1.0 for within and 1.6–2.6 for between serotypes. This was even lower for WNV, ranging from 0.2–0.5, with the highest for HIV-1 clade B subtype, 1.9–4.2. Entropy is a general measure of diversity, and the data provided a holistic overview viral diversity across the proteome. Accordingly, the incidence of variants to the most prevalent, index sequence at the aligned nonamer positions was the lowest for WNV (intra: $\leq 10\%$) and the highest for HIV-1 clade B (intra: $\sim 80\%$ – 99%); the variants incidence within each DENV serotype was comparable to WNV, but between serotypes ($\sim 60\%$ – 80%) was closer to HIV-1 clade B subtype. Forty-four (44) sequences (pan-DENV sequences) identical in 80% or more of all recorded DENV sequences represented 15% of the DENV polyprotein length. The proportion (34%) was much higher for WNV and at complete conservation (100% incidence). Notably, at similar incidence level ($> = 80\%$ incidence) to DENV, although pan-clade, $\sim 35\%$ of the intra HIV-1 clade B proteome was highly conserved. The proportion of these conserved sequences that were immune-relevant showed an inverse relationship: DENV (59% matched 9aa or more of 45 class I and II reported epitopes), WNV (50% matched 9aa or more of 57 class I and II reported epitopes), and HIV-1 clade B (37% matched 9aa or more of 73 class I and II reported epitopes). Khan *et al.* (2008) highlighted that conservation analysis should go beyond the species of interest, extending to all those

Table 1 Key summary results of vaccine target discovery study on Dengue virus (DENV), West Nile virus (WNV) and clade B HIV-1.

Category	DENV (Khan <i>et al.</i> , 2008)	WNV (Koo <i>et al.</i> , 2009)	HIV-1 clade B (Hu <i>et al.</i> , 2013)
Mean nonamer entropy	Intra: 0.2–1.0 Inter: 1.6–2.6	Intra: 0.23–0.51	Intra: 1.9–4.2
General variants incidence trend	Inter: ~60%–80%	Intra: <10%	Intra: ~80%–99%
% of the proteome covered by conserved sequences	Pan-clade: 15% (>=80% incidence)	Intra: 34% (=100% incidence)	Intra: 35% (>=80% incidence)
Immune relevance	(26/44) 59% matched 9aa or more of 45 class I and II reported epitopes	(44/88) 50% matched 9aa or more of 57 class I and II reported epitopes	(29/78) 37% matched 9aa or more of 73 class I and II reported epitopes
No. of other viral species matched by the conserved sequences	(27/44) 61% matched 64 flaviviruses	(67/88) 76% of the sequences matched 68 flaviviruses	(74/78) 95% matched 9 deltaviruses
No. of pathogen specific conserved sequences	17	21	4

other species that are evolutionarily related as they may act as variants to the conserved epitopes identified. This step is necessary to identify conserved epitope sequences that are pathogen specific, with none or minimal number of variant sequences within or across other pathogen species. Variant epitopes are hypothesized to cause deleterious immune responses. Many of the conserved sequences matched nine consecutive amino acids of many (flaviviruses; family of WNV and DENV) to few (*deltaviruses*; genus of HIV-1) other related viruses, leaving only 17, 21 and 4 pathogen specific conserved sequences for DENV, WNV and HIV-1 clade B subtype, respectively.

Non-Infectious Diseases

According to World Health Organization (WHO), non-infectious diseases, especially chronic diseases will lead the disability by 2020. Hence, diseases such as cancers, obesity, neurodegenerative disease addictions and others have become a recent focus of vaccine development (Barrett, 2016). Cancer is the most common non-infectious disease that leads death world-wide. Due to the advanced technology and success of *in silico* methods in infectious disease vaccine design, computational approaches have been applied in study of cancer vaccine design. For example, VacImm was developed as a bioinformatics approach to simulate peptide vaccination in cancer therapy (von Eichborn *et al.*, 2013). In addition, modeling approaches using computational biology, such as Sim Triplex and MetastaSim model are important to understand the molecular interactions at the cellular and molecular level (Pappalardo *et al.*, 2013; Sankar *et al.*, 2013). Adekiya *et al.* (2017) reports a recent example of a study that included bioinformatics analysis in cancer vaccine development.

Conclusion: Vaccine Informatics and Future Vaccines

Future vaccines will be minimalistic in approach by focusing on key parts of the pathogen, such as regions containing epitopes that cover antigenic diversity and, thus, will target immunologically similar subgroups of the human population and multiple pathogen variants. This is evident from the trend observed in evolution of vaccine strategies, which has seen a shift from whole organisms to recombinant proteins, and further towards the ultimate in minimalist vaccinology, the peptide/epitope/multi-epitope based vaccines. The minimalist approach is also expected to cover the safety concerns that are associated with the traditional vaccine approach of using whole organism (Sette and Fikes, 2003; Dertzbaugh, 1998). Vected vaccines, suitable for 'combination immunization' that are produced by recombinant DNA technology and contain multivalent minimal antigens to protect against multiple infections, are considered to be the future of vaccinology (Kutzler and Weiner, 2008). The future will bring increased integration of vaccine research with advances in immunology, molecular biology, genomics, proteomics, informatics, and high-throughput instrumentation, collectively defined as the emerging field of "vaccinomics", which is hailed to be responsible for the next 'golden age' in vaccinology (Poland *et al.*, 2008). Awareness of the novel technological possibilities in vaccine research is also expected to grow. Future vaccinology will be based on detailed understanding of immune function, optimal stimulation of immune responses (using adjuvants) and precise mapping and rational selection of immune targets (Brusic *et al.*, 2005). To achieve this, vaccine development will routinely be conducted through large-scale functional studies supported by genomics, proteomics, and informatics techniques prior to clinical trials. This will provide an increased range of immune targets for vaccine design. The author expects the emergence of new generation of vaccines to be personalised to both the genetic make-up of the human population and of the disease agents. In summary, vaccinology will experience rapid progress and will eventually deliver benefits to patients from improved diagnosis, treatment and prevention of diseases.

See also: Computational Pipelines and Workflows in Bioinformatics. Extraction of Immune Epitope Information. Host-Pathogen Interactions. Natural Language Processing Approaches in Bioinformatics

References

- Abelin, J.G., *et al.*, 2017. Mass spectrometry profiling of HLA-associated peptidomes in mono-allelic cells enables more accurate epitope prediction. *Immunity* 46 (2). 315–326.
- Ada, G., 2003. Progress towards achieving new vaccine and vaccination goals. *Internal Medicine Journal* 33, 297–304.
- Adekiya, T., *et al.*, 2017. Structural analysis and epitope prediction of MHC class-I-chain related protein-a for cancer vaccine development. *Vaccines* 6 (1). 1.
- Agostino, M., *et al.*, 2016. Optimization of protein-protein docking for predicting Fc-protein interactions. *Journal of Molecular Recognition* 29 (11). 555–568.
- Altschul, S.F., *et al.*, 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215 (3). 403–410.
- Andre, F.E., *et al.*, 2008. Vaccination greatly reduces disease, disability, death and inequity worldwide. *Bulletin of the World Health Organization* 86, 140–146.
- Andreatta, M., *et al.*, 2015. Accurate pan-specific prediction of peptide-MHC class II binding affinity with improved binding core identification. *Immunogenetics* 67 (11–12). 641–650.
- Andreatta, M., Nielsen, M., 2015. Gapped sequence alignment using artificial neural networks: Application to the MHC class I system. *Bioinformatics* 32 (4). 511–517.
- Arnon, R., Ben-yedidia, T., 2003. Old and new vaccine approaches. *International Immunopharmacology* 3, 1195–1204.
- Babiuk, L.A., 1999. Broadening the approaches to developing more effective vaccines. *Vaccine* 17, 1587–1595.
- Barrett, A.D.T., 2016. Vaccinology in the twenty-first century. *npj Vaccines* 1 (1). 16009.
- Bhasin, M., Lata, S., Raghava, G.P.S., 2007. TAPPred prediction of TAP-binding peptides in antigens. *Methods in Molecular Biology* 409, 381–386.
- Bhasin, M., Raghava, G.P.S., 2004. SVM based method for predicting HLA-DRB1*0401 binding peptides in an antigen sequence. *Bioinformatics* 20 (3). 421–423.
- Birnbaum, M.E., *et al.*, 2014. Deconstructing the peptide-MHC specificity of t cell recognition. *Cell* 157 (5). 1073–1087.
- Blum, J.S., Wearsch, P.A., Cresswell, P., 2013. Pathways of antigen processing. *Annual Review of Immunology* 31, 443–473.
- Bona, C.A., Casares, S., Brumeanu, T.D., 1998. Towards development of T-cell vaccines. *Immunology Today* 19 (3). 126–133.
- Bouvier, N., Palese, P., 2008. The biology of influenza viruses. *Vaccine* 26, D49–D53.
- Brusic, V., August, J.T., Petrovsky, N., 2005. Information technologies for vaccine research. *Expert Review of Vaccines* 4 (3). 407–417.
- Brusic, V., Petrovsky, N., 2003. Immunoinformatics—The new kid in town. *Novartis Foundation Symposium* 254, 3–22. 98–101, 250–252.
- Brusic, V., Zeleznikow, J., Petrovsky, N., 2000. Molecular immunology databases and data repositories. *Journal of Immunological Methods* 238, 17–28.
- Chowell, D., *et al.*, 2015. TCR contact residue hydrophobicity is a hallmark of immunogenic CD8⁺ T cell epitopes. *Proceedings of the National Academy of Sciences of the United States of America* 112 (14). E1754–E1762.
- Damfo, S.A., *et al.*, 2017. In silico design of knowledge-based Plasmodium falciparum epitope ensemble vaccines. *Journal of Molecular Graphics and Modelling* 78, 195–205.
- Dash, P., *et al.*, 2017. Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature* 547 (7661). 89–93.
- Dertzbaugh, M.T., 1998. Genetically engineered vaccines: An overview. *Plasmid* 39 (2). 100–113.
- Dhanda, S.K., *et al.*, 2017. Novel in silico tools for designing peptide-based subunit vaccines and immunotherapeutics. *Briefings in Bioinformatics* 18 (3). 467–478.
- Do, C.B., Mahabhashyam, M.S., Brudno, M., Batzoglou, S., 2005. ProbCons: Probabilistic consistency-based multiple sequence alignment. *Genome Res.* 15 (2). 330–340. PubMed PMID: 15687296; PubMed Central PMCID: PMC546535.
- Dominguez, C., Boelens, R., Bonvin, A.M.J.J., 2003. HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *Journal of the American Chemical Society* 125 (7). 1731–1737.
- Doolan, D.L., 2003. Utilization of genomic sequence information to develop malaria vaccines. *Journal of Experimental Biology* 206 (21). 3789–3802.
- Edgar, R.C., 2004. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*. 5, 113. PubMed PMID: 15318951; PubMed Central PMCID: PMC517706.
- von Eichborn, J., *et al.*, 2013. VacImm: Simulating peptide vaccination in cancer therapy. *BMC Bioinformatics* 14.
- Evavold, B.D., Sloan-Lancaster, J., Allen, P.M., 1993. Tickling the TCR: Selective T-cell functions stimulated by altered peptide ligands. *Immunology Today* 14 (12). 602–609.
- Fauci, A.S., 2001. Infectious diseases: Considerations for the 21st Century. *Clinical Infectious Diseases* 32, 675–685.
- Fauci, A.S., Touchette, N.A., Folkers, G.K., 2005. Emerging infectious diseases: A 10-year perspective from the National Institute of Allergy and Infectious Diseases. *Emerging Infectious Diseases* 11 (4). 519–525.
- Fleri, W., *et al.*, 2017. The immune epitope database: How data are entered and retrieved. *Journal of Immunology Research* 2017.
- Gee, M.H., *et al.*, 2017. Antigen identification for orphan T cell receptors expressed on tumor-infiltrating lymphocytes. *Cell*.
- Glanville, J., *et al.*, 2017. Identifying specificity groups in the T cell receptor repertoire. *Nature* 547 (7661). 94–98.
- Goodswen, S.J., Kennedy, P.J., Ellis, J.T., 2014. Vacceed: A high-throughput in silico vaccine candidate discovery pipeline for eukaryotic pathogens based on reverse vaccinology. *Bioinformatics* 30 (16). 2381–2383.
- Gourlay, L., *et al.*, 2017. Structure and computation in immunoreagent design: From diagnostics to vaccines. *Trends in Biotechnology* 35 (12). 1208–1220.
- De Groot, A.S., *et al.*, 2002. Immuno-informatics: Mining genomes for vaccine components. *Immunology and Cell Biology* 80 (3). 255–269.
- De Groot, A.S., *et al.*, 2004. Engineering immunogenic consensus T helper epitopes for a cross-clade HIV vaccine. *Methods* 34, 476–487.
- Gupta, A.K., *et al.*, 2016. ZikaVR: An integrated Zika virus resource for genomics, proteomics, phylogenetic and therapeutic analysis. *Scientific Reports* 6.
- Hakenberg, J., *et al.*, 2003. MAPPP: MHC class I antigenic peptide processing prediction. *Applied Bioinformatics* 2 (3). 155–158.
- Haydon, D.T., Woolhouse, M.E., 1998. Immune avoidance strategies in RNA viruses: Fitness continuums arising from trade-offs between immunogenicity and antigenic variability. *J. Theor. Biol.* 193, 601–612.
- He, Y., 2014. Bacterial whole-genome determination and applications. *Molecular Medical Microbiology: Second Edition*. 357–368.
- Hegde, N.R., *et al.*, 2017. The use of databases, data mining and immunoinformatics in vaccinology: Where are we? *Expert Opinion on Drug Discovery* 13 (2). 117–130.
- Heiny, A., *et al.*, 2007. Evolutionarily conserved protein sequences of influenza A viruses, avian and human, as vaccine targets. *PLOS ONE* 2 (11). e1190. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18030326>.
- Hilleman, M.R., 1985. Newer directions in vaccine development and utilization. *Journal of Infectious Diseases* 151 (3). 407–419.
- Hu, Y., *et al.*, 2013. Dissecting the dynamics of HIV-1 protein sequence diversity. *PLOS ONE* 8 (4).
- Hung, M.-C., Link, W., 2011. Protein localization in disease and therapy. *Journal of Cell Science* 124 (20). 3381–3392.
- Jerne, N.K., 1993. The generative grammar of the immune system. *Scandinavian Journal of Immunology* 38 (1). 2–8.
- Jurtz, V., *et al.*, 2017. NetMHCpan-4.0: Improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *The Journal of Immunology* 199 (9). 3360–3368.
- Katoh, K., Rozewicki, J., Yamada, K.D., 2017. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform.* doi:10.1093/bib/bbx108. [Epub ahead of print] PubMed PMID: 28968734.

- Karosiene, E., *et al.*, 2013. NetMHCIIpan-3.0, a common pan-specific MHC class II prediction method including all three human MHC class II isotypes, HLA-DR, HLA-DP and HLA-DQ. *Immunogenetics* 65 (10). 711–724.
- Kast, W.M., *et al.*, 1994. Role of HLA-A motifs in identification of potential CTL epitopes in human papillomavirus type 16 E6 and E7 proteins. *Journal of Immunology* 152 (8). 3904–3912.
- Kent, W.J., 2002. BLAT – The BLAST-like alignment tool. *Genome Research* 12, 656–664.
- Khan, A.M., 2005. Mapping targets of immune responses in complete dengue viral genomes. MSc Thesis, National University of Singapore. Available at: [http://www.scholarbank.nus.edu.sg/bitstream/handle/10635/15081/MohammadAsifKhan\(KHAN_AM_Thesis.PDF\).pdf?sequence=1](http://www.scholarbank.nus.edu.sg/bitstream/handle/10635/15081/MohammadAsifKhan(KHAN_AM_Thesis.PDF).pdf?sequence=1).
- Khan, A.M., 2009. Antigenic diversity of dengue virus: Implications for vaccine design. PhD Thesis, National University of Singapore. Available at: <http://scholarbank.nus.edu.sg/handle/10635/17142>.
- Khan, A.M., *et al.*, 2006. A systematic bioinformatics approach for selection of epitope-based vaccine targets. *Cellular Immunology* 244 (2). 141–147. Available at: <http://www.sciencedirect.com/science/article/B6WCF-4NH6NCY-2/2/081b67baf4af7b5e9da5c0dd3ea404f>.
- Khan, A.M., *et al.*, 2008. Conservation and variability of dengue virus proteins: Implications for vaccine design. *PLOS Neglected Tropical Diseases* 2 (8). e272. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18698358>.
- Khan, A.M., *et al.*, 2017. Analysis of viral diversity for vaccine target discovery. *BMC Medical Genomics* 10 (Suppl. 4). S78.
- Khan, J.M., Ranganathan, S., 2010. PDOCK: A new technique for rapid and accurate docking of peptide ligands to major histocompatibility complexes. *Immunome Research* 6 (1). S2. doi: 10.1186/1745-7580-6-S1-S2.
- Knobler, S.L., *et al.*, 2003. The resistance phenomenon in microbes and infectious disease vectors: Implications for human health and strategies for containment. *National Academies Press* 336 (9). 309.
- Koo, Q.Y., *et al.*, 2009. Conservation and variability of West Nile virus proteins. *PLOS ONE* 4 (4). e5352.
- Kulkarni-Kale, U., Bhosle, S., Kolaskar, A.S., 2005. CEP: A conformational epitope prediction server. *Nucleic Acids Research* 33 (Suppl. 2).
- Kutzler, M.A., Weiner, D.B., 2008. DNA vaccines: Ready for prime time? *Nature Reviews Genetics* 9 (10). 776–788.
- Larsen, M.V., *et al.*, 2005. An integrative approach to CTL epitope prediction: A combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions. *European Journal of Immunology* 35 (8). 2295–2303.
- Lee, N.H., *et al.*, 2012. A review of vaccine development and research for industrially animals in Korea. *Clinical and Experimental Vaccine Research* 1, 18–34.
- Lefranc, M.P., *et al.*, 2009. IMGT[®], the international ImmunoGeneTics information system[®]. *Nucleic Acids Research* 37 (Suppl. 1).
- Lennon-Dumenil, A.-M., *et al.*, 2002. Analysis of protease activity in live antigen-presenting cells shows regulation of the phagosomal proteolytic contents during dendritic cell activation. *The Journal of Experimental Medicine* 196 (4). 529–540.
- Li, F., Finnefrock, A.C., Dubey, S.A., *et al.*, 2011. Mapping HIV-1 vaccine induced T-cell responses: Bias towards less-conserved regions and potential impact on vaccine efficacy in the Step study. *PLoS One*. 6 (6). e20479. doi:10.1371/journal.pone.0020479. Epub 2011 Jun 10. PubMed PMID: 21695251; PubMed Central PMCID: PMC3112144.
- Liljeroos, L., *et al.*, 2015. Structural and computational biology in the design of immunogenic vaccine antigens. *Journal of Immunology Research* 2015.
- Lund, O., *et al.*, 2004. Definition of supertypes for HLA molecules using clustering of specificity matrices. *Immunogenetics* 55 (12). 797–810.
- MacDonald, K.S., *et al.*, 2000. Influence of HLA supertypes on susceptibility and resistance to human immunodeficiency virus type 1 infection. *The Journal of Infectious Diseases* 181 (5). 1581–1589.
- MacDonald, K.S., *et al.*, 2001. Human leucocyte antigen supertypes and immune susceptibility to HIV-1, implications for vaccine design. *DNA Sequence* 9, 151–157.
- Maione, D., *et al.*, 2005. Immunology: Identification of a universal group B streptococcus vaccine by multiple genome screen. *Science* 309 (5731). 148–150.
- María, R.R., *et al.*, 2017. The impact of bioinformatics on vaccine design and development. In: Afrin, F. (Ed.), *Vaccines*. Place: IntechOpen, pp. 123–145. doi:10.5772/intechopen.69273. Available from: <https://www.intechopen.com/books/vaccines/the-impact-of-bioinformatics-on-vaccine-design-and-development>.
- Marrack, P., *et al.*, 2017. The somatically generated portion of T cell receptor CDR3 α contributes to the MHC allele specificity of the T cell receptor. *eLife* 6.
- Minichiello, V., 2002. New vaccine technology—what do you need to know? *Journal of the American Academy of Nurse Practitioners* 14 (2). 73–81.
- Mistry, J., Flower, D.R., 2017. Designing epitope ensemble vaccines against TB by selection: Prioritizing antigens using predicted immunogenicity. *Bioinformation* 13 (7). 220–223.
- Monterrubio-López, G.P., González-Y-Merchand, J.A., Ribas-Aparicio, R.M., 2015. Identification of novel potential vaccine candidates against tuberculosis based on reverse vaccinology. *BioMed Research International* 2015.
- Morens, D.M., Fauci, A.S., 2013. Emerging infectious diseases: Threats to human health and global stability. *PLOS Pathogens* 9 (7). e1003467.
- Nandy, A., Basak, S.C., 2016. A brief review of computer-assisted approaches to rational design of peptide vaccines. *International Journal of Molecular Sciences* 17 (5). 2017. Database resources of the national center for biotechnology information. *Nucleic Acids Research* 45 (D1). D12–D17.
- Nicholson, L.B., 2016. The immune system. *Essays in Biochemistry* 60 (3). 275–301.
- Nielsen, M., *et al.*, 2005. The role of the proteasome in generating cytotoxic T-cell epitopes: Insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics* 57 (1–2). 33–41.
- Nielsen, M., *et al.*, 2010. MHC Class II epitope predictive algorithms. *Immunology* 130 (3). 319–328.
- NIH, 2003. *Understanding the Immune System How It Works*. NIH. pp. 1–7.
- Nussbaum, A.K., *et al.*, 2001. PAPProC: A prediction algorithm for proteasomal cleavages available on the WWW. *Immunogenetics* 53 (2). 87–94.
- Pappalardo, F., Chiacchio, F., Motta, S., 2013. Cancer vaccines: State of the art of the computational modeling approaches. *BioMed Research International* 2013, 106407.
- Park, A.W., *et al.*, 2004. The effects of strain heterology on the epidemiology of equine influenza in a vaccinated population. *Proceedings of the Royal Society B: Biological Sciences* 271 (1548). 1547–1555.
- Parker, K.C., Bednarek, M.A., Coligan, J.E., 1994. Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. *Journal of Immunology* 152 (1). 163–175.
- Paul, S., Dillon, M.B.C., *et al.*, 2015a. A population response analysis approach to assign class II HLA-Epitope restrictions. *The Journal of Immunology* 194 (12). 6164–6176.
- Paul, S., Lindestam Arlehamn, C.S., *et al.*, 2015b. Development and validation of a broad scheme for prediction of HLA class II restricted T cell epitopes. *Journal of Immunological Methods* 422, 28–34.
- Paul, S., *et al.*, 2013. HLA class I alleles are associated with peptide-binding repertoires of different size, affinity, and immunogenicity. *The Journal of Immunology* 191 (12). 5831–5839.
- Petrovsky, N., Brusic, V., 2002. Computational immunology: The coming of age. *Immunology and Cell Biology* 80 (3). 248–254.
- Petrovsky, N., Schönbach, C., Brusic, V., 2003. Bioinformatic strategies for better understanding of immune function. *In Silico Biology* 3 (4). 411–416.
- Pizza, M., *et al.*, 2000. Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science* 287 (5459). 1816–1820.
- Plotkin, S.A., 2005. Vaccines: Past, present and future. *Nature Medicine* 11 (4 Suppl.). S5–S11.
- Poland, G.A., Ovsyannikova, I.G., Jacobson, R.M., 2008. Personalized vaccines: The emerging field of vaccinomics. *Expert Opinion on Biological Therapy* 8 (11). 1659–1667.
- Pro, S.C., *et al.*, 2015. Automatic generation of validated specific epitope sets. *Journal of Immunology Research* 2015, 763461.
- Raman, H., *et al.*, 2014. Bioinformatics for vaccine target discovery. *Asia Pacific Biotech News* 18 (9). 25–53.
- Rammensee, H.-G., *et al.*, 1999. SYFPEITHI: Database for MHC ligands and peptide motifs. *Immunogenetics* 50 (3–4). 213–219.

- Ranganathan, S., Tong, J.C., 2007. A practical guide to structure-based prediction of MHC-binding peptides. *Methods in Molecular Biology* 409, 301–308.
- Rappuoli, R., 2001. Reverse vaccinology, a genome-based approach to vaccine development. *Vaccine* 19 (17–19). 2688–2691.
- Rappuoli, R., Covacci, A., 2003. Reverse vaccinology and genomics. *Science* 302 (5645). 602.
- Rizwan, M., *et al.*, 2017. VacSol: A high throughput in silico pipeline to predict potential therapeutic targets in prokaryotic pathogens using subtractive reverse vaccinology. *BMC Bioinformatics* 18 (1).
- Rosloniec, E.F., *et al.*, 1997. An HLA-DR1 transgene confers susceptibility to collagen-induced arthritis elicited with human type II collagen. *The Journal of Experimental Medicine* 185 (6). 1113–1122.
- Rothman, A.L., 2004. Dengue: Defining protective versus pathologic immunity. *Journal of Clinical Investigation* 113 (7). 946–951.
- Ryan, C.P., 2008. Zoonoses likely to be used in bioterrorism. *Public Health Reports* 123 (3). 276–281.
- Sachdeva, G., *et al.*, 2005. SPAAN: A software program for prediction of adhesins and adhesin-like proteins using neural networks. *Bioinformatics* 21 (4). 483–491.
- Sankar, S., Nayanar, S., Balasubramanian, S., 2013. Current trends in cancer vaccines – a bioinformatics perspective. *Asian Pacific Journal of Cancer Prevention* 14 (7). 4041–4047.
- Sette, A., *et al.*, 2001. The development of multi-epitope vaccines: Epitope identification, vaccine design and clinical evaluation. *Biologicals*. 271–276.
- Sette, A., Fikes, J., 2003. Epitope-based vaccines: An update on epitope identification, vaccine design and delivery. *Current Opinion in Immunology* 15 (4). 461–470.
- Sette, A., Sidney, J., 1999. Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. *Immunogenetics* 50, 201–212.
- Shastry, N., Schwab, S., Serwold, T., 2002. Producing nature's gene-chips: The generation of peptides for display by MHC class I molecules. *Annual Review of Immunology* 20 (1). 463–493. Available at: <http://arjournals.annualreviews.org/doi/abs/10.1146/annurev.immunol.20.100301.064819>.
- Sievers, F., Wilm, A., Dineen, D., *et al.*, 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7, 539. doi:10.1038/msb.2011.75. PubMed PMID: 21988835; PubMed Central PMCID: PMC3261699.
- Sloan-Lancaster, J., Allen, P.M., 1996. Altered peptide ligand-induced partial T cell activation: Molecular mechanisms and role in T cell biology. *Annual Review of Immunology* 14, 1–27.
- Soria-Guerra, R.E., *et al.*, 2015. An overview of bioinformatics tools for epitope prediction: Implications on vaccine development. *Journal of Biomedical Informatics* 53, 405–414.
- Sprenger, J., *et al.*, 2008. LOCATE: A mammalian protein subcellular localization database. *Nucleic Acids Research* 36 (Suppl. 1). D230–D233.
- Srinivasan, K.N., *et al.*, 2004. Prediction of class I T-cell epitopes: Evidence of presence of immunological hot spots inside antigens. *Bioinformatics* 20 (Suppl. 1).
- Strassburg, M.A., 1980. The global eradication of smallpox. *American Journal of Infection Control* 10 (2). 53–59.
- Sturniolo, T., *et al.*, 1999. Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. *Nature Biotechnology* 17 (6). 555–561.
- Sylvester-Hvid, C., *et al.*, 2002. Establishment of a quantitative ELISA capable of determining peptide – MHC class I interaction. *Tissue Antigens* 59, 251–258.
- Tang, Y.W., Sussman, M., Liu, D., Poxton, I., Schwartzman, J. (Eds.), 2014. *Molecular Medical Microbiology*, second edn. Academic Press.
- Tong, J.C., Tan, T.W., Ranganathan, S., 2007. Methods and protocols for prediction of immunogenic epitopes. *Briefings in Bioinformatics* 8 (2). 96–108.
- Del Tordello, E., Rappuoli, R., Delany, I., 2016. Reverse vaccinology: Exploiting genomes for vaccine design. *Human Vaccines: Emerging Technologies in Design and Development*. 65–86.
- Trolle, T., *et al.*, 2015. Automated benchmarking of peptide-MHC class I binding predictions. *Bioinformatics* 31 (13). 2174–2181.
- Vernikos, G.S., 2008. Genome watch: Overtake in reverse gear. *Nature Reviews Microbiology* 6 (5). 334–335.
- Vigneron, N., Van den Eynde, B.J., 2014. Proteasome subtypes and regulators in the processing of antigenic peptides presented by class I molecules of the major histocompatibility complex. *Biomolecules* 4 (4). 994–1025.
- Vivona, S., Bernante, F., Filippini, F., 2006. NERVE: New enhanced reverse vaccinology environment. *BMC Biotechnology* 6, 35.
- De Vries, S.J., Van Dijk, M., Bonvin, A.M.J.J., 2010. The HADDOCK web server for data-driven biomolecular docking. *Nature Protocols* 5 (5). 883–897.
- Wizemann, T.M., Adamou, J.E., Langermann, S., 1999. Adhesins as targets for vaccine development. *Emerging Infectious Diseases* 5 (3). 395–403.
- Yatim, K.M., Lakkis, F.G., 2015. A brief journey through the immune system. *Clinical Journal of the American Society of Nephrology* 10 (7). 1274–1281.
- Zaharieva, N., *et al.*, 2017. Immunogenicity prediction by VaxiJen: A ten year overview. *Journal of Proteomics & Bioinformatics* 10 (11). 298–310.
- Zahroh, H., *et al.*, 2016. Immunoinformatics approach in designing epitopebased vaccine against meningitis-inducing bacteria (*Streptococcus pneumoniae*, *Neisseria meningitidis*, and *Haemophilus influenzae* type b). *Drug Target Insights* 10, 19–29.
- Zhang, G., *et al.*, 2005. MULTIPRED: A computational system for prediction of promiscuous HLA binding peptides. *Nucleic Acids Res* 33 (Web Server issue). W172–W179. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15980449>.
- Zhang, G.L., *et al.*, 2006. PRED(TAP): A system for prediction of peptide binding to the human transporter associated with antigen processing. *Immune Research* 2 (1). 3.
- Zhang, W., *et al.*, 2011. Prediction of conformational B-cell epitopes from 3D structures by random forests with a distance-based feature. *BMC Bioinformatics* 12.
- Zhang, Y., *et al.*, 2017. Influenza research database: An integrated bioinformatics resource for influenza virus research. *Nucleic Acids Research* 45 (D1). D466–D474.

Further Reading

- Aparicio, R., *et al.*, 2017. World's largest science, technology & medicine open access book publisher. The impact of bioinformatics on vaccine design and development.
- Barrett, A.D.T., 2016. Vaccinology in the twenty-first century. *NPJ Vaccines* 1 (1). 16009.
- Brusic, V., August, J.T., Petrovsky, N., 2005. Information technologies for vaccine research. *Expert Review of Vaccines* 4 (3). 407–417.
- Brusic, V., Petrovsky, N., 2003. Immunoinformatics – The new kid in town. *Novartis Foundation Symposium* 254, 13–22. 98–101, 250–252.
- Brusic, V., Zeleznikow, J., Petrovsky, N., 2000. Molecular immunology databases and data repositories. *Journal of Immunological Methods* 238, 17–28.
- Dhanda, S.K., *et al.*, 2017. Novel in silico tools for designing peptide-based subunit vaccines and immunotherapeutics. *Briefings in Bioinformatics* 18 (3). 467–478.
- Gourlay, L., *et al.*, 2017. Structure and computation in immunoreagent design: From diagnostics to vaccines. *Trends in Biotechnology* 35 (12). 1208–1220.
- Hegde, N.R., *et al.*, 2017. The use of databases, data mining and immunoinformatics in vaccinology: Where are we? *Expert Opinion on Drug Discovery* 13 (2). 117–130.
- Khan, A.M., *et al.*, 2017. Analysis of viral diversity for vaccine target discovery. *BMC Medical Genomics* 10 (Suppl. 4). S78.
- Mistry, J., Flower, D.R., 2017. Designing epitope ensemble vaccines against TB by selection: Prioritizing antigens using predicted immunogenicity. *Bioinformatics* 13 (7). 220–223.
- Tong, J.C., Ranganathan, S., 2013. *Computer-Aided Vaccine Design*, Woodhead Publishing Series In Biomedicine No. 23. Cambridge, UK: Woodhead, pp. 1–164.

Relevant Website

www.ncbi.nlm.nih.gov/genomes/VIRUSES/viruses.html
Viral Genomes.

Biographical Sketch



Chong Li Chuin is a Biomedical Science undergraduate student from Faculty of Medicine and Health Sciences, UPM. As part of her undergraduate studies, she completed an internship at Perdana University Centre for Bioinformatics (PU-CBi), Malaysia. She continued to be involved with the Centre to build on her passion for bioinformatics.



Associate Professor Dr. Mohammad Asif Khan is currently the Dean of the Perdana University School of Data Sciences, the Director of Perdana University Centre for Bioinformatics (PU-CBi), Malaysia, and a Visiting Scientist at Johns Hopkins University School of Medicine (JHUSOM), USA. He obtained his PhD in Bioinformatics from the National University of Singapore (NUS), and did his postdoctoral fellowships at NUS and JHUSOM. Dr. Khan's research interests are in the area of biological data warehousing and applications of bioinformatics to the study of immune responses, vaccines, venom toxins, drug design, and disease biomarkers.