NO PICO

PAPER ID: PMC3854287

TITLE:
Coping with genetic diversity: the contribution of pathogen and human genomics to modern vaccinology

ABSTRACT:
Vaccine development faces major difficulties partly because of genetic variation in both infectious organisms and humans. This causes antigenic variation in infectious agents and a high interindividual variability in the human response to the vaccine. The exponential growth of genome sequence information has induced a shift from conventional culture-based to genome-based vaccinology, and allows the tackling of challenges in vaccine development due to pathogen genetic variability. Additionally, recent advances in immunogenetics and genomics should help in the understanding of the influence of genetic factors on the interindividual and interpopulation variations in immune responses to vaccines, and could be useful for developing new vaccine strategies. Accumulating results provide evidence for the existence of a number of genes involved in protective immune responses that are induced either by natural infections or vaccines. Variation in immune responses could be viewed as the result of a perturbation of gene networks; this should help in understanding how a particular polymorphism or a combination thereof could affect protective immune responses. Here we will present: i) the first genome-based vaccines that served as proof of concept, and that provided new critical insights into vaccine development strategies; ii) an overview of genetic predisposition in infectious diseases and genetic control in responses to vaccines; iii) population genetic differences that are a rationale behind group-targeted vaccines; iv) an outlook for genetic control in infectious diseases, with special emphasis on the concept of molecular networks that will provide a structure to the huge amount of genomic data.

Conventional versus reverse vaccinology ::: Introduction:
The first vaccines resulted from empirical studies and were mainly composed of whole organisms (live attenuated or killed) or inactivated toxins. The majority of vaccine studies developed later also used conventional approaches to select for appropriate antigens. Antigen selection was a key step in vaccine development in the pre-genomic era but was time-consuming and had many limitations. The properties aimed for in these molecules were: 1) being accessible to the immune system and inducing protective immunity and 2) being conserved within the species and not containing regions of high variability (3). The selection of good candidate vaccine antigens was primarily based on the capacity of induction of protective antibodies, which probably explains why effective vaccines currently available mainly induce a protective humoral immune response (1).
More recently, molecular biology methods have been applied to the development of new vaccines, primarily to enable large-scale production of subunit or peptide-based vaccine antigens. Nevertheless, only a small number of available vaccines are based on recombinant antigens.
The strategies of conventional vaccinology pose limitations for the development of effective vaccines against a variety of infectious diseases. The development of vaccines against pathogens that show important antigenic variation or that cannot be cultivated in the laboratory remains a great challenge. Currently, new approaches using high-throughput technologies are being used for the identification of new vaccine candidate molecules (4).
The publication of the Haemophilus influenzae genome, the first pathogen to have its complete genome sequence published as a result of an approach to genome analysis using new technologies of high-throughput sequencing (5), has opened the mind of scientists to a range of new possible approaches to the study of microorganisms and has marked the beginning of a new era in vaccine development: the identification of pathogen candidate antigens based on the knowledge of the genome of the pathogen and on the understanding of microbial biology and host-pathogen interactions, an approach called reverse vaccinology (6).
This approach has been favored in recent years by the use of high-throughput technologies in the fields of genomics, proteomics and immunomics, among other “omics”, which has expanded enormously. The number of available viral and parasitic genome sequences is growing continuously (2), and important advances have been made to make new biostatistics and bioinformatics tools available for candidate gene prospection. The current “genomic era” is

marked by a major impact on the ability to find new candidate vaccine antigens and to develop effective vaccines.

One of the approaches that may be used to analyze the DNA data sequences in order to identify vaccine candidate molecules is the in silico analysis of complete genomes. It does not require cultivation of the pathogen and it uses a variety of computer resources to process DNA sequence data and the available information about protein functions, cell transport and localization, antigen processing, and immunogenicity. Computational analysis may then enable the identification of which DNA sequences encode the proteins that have some relevant features of potential vaccine targets (7).

The impact of in silico analysis on vaccine development can be exemplified by the first study published using this approach (8). A capsular polysaccharide-based vaccine against Neisseria meningitidis serogroups A, C, W135, and Y is available, but not against serogroup B as its structure is identical to the polysialic acid structure present on human cells (3). The entire genome sequence of a virulent N. meningitidis serogroup B strain, which accounts for the majority of meningococcal disease burden in the United States and Europe (9), was analyzed to identify candidates for a new vaccine. The bacterial genome was screened using bioinformatics tools to identify relatively conserved, surface-exposed outer membrane protein antigens: 570 of 2158 open reading frames (ORFs) were thus selected. Immunoproteomics methods were then used to select the most immunogenic meningococcal antigens that would elicit a protective immunity. Five of 28 genome-derived candidate antigens are components of a rationally designed vaccine against N. meningitidis serogroup B (10) and a formulation of four of these antigens (4CMenB, Novartis, Switzerland) has induced high titers of bactericidal antibodies in a phase I/II clinical trial (11). The multivalent vaccine formulation has been proposed to overcome the issue of antigenic diversity between strains of N. meningitidis, and it would be expected to stimulate immune responses capable of recognizing multiple antigenic variants. The promising results of this new vaccine candidate may be the critical “proof of concept” stage for this novel approach that has significantly shifted the paradigm of vaccine development from microbiological to sequence-based approaches.

Genome-wide screening using functional selection methods such as DNA microarrays, gene knockout libraries, or screening of cDNA expression libraries using patients' sera has been employed in the prospection of vaccine candidates (reviewed in Ref. 12). These approaches allow the evaluation of the possible immunogenic and immuno-protective roles of all molecules (either surface expressed or secreted by pathogens), thereby enabling the identification of candidate antigens that were not identified by the methodologies applied in conventional vaccinology. This means that even molecules that are expressed at very low levels or only under certain conditions can be tested for their potential as candidate vaccines. In brief, genomics coupled with other new “omics” fields of study such as transcriptomics, proteomics, and immunomics enables the identification of vaccine candidate molecules based on pathogen genome analysis.

High-throughput sequencing technologies have also been used to understand the molecular biology of several viruses and parasites. Recently, genomic-based approaches have allowed the identification of potential novel viral and parasitic targets, which will be included in vaccines against infectious diseases, such as severe acute respiratory syndrome (SARS; 13), AIDS (14), trypanosomiasis (15), and malaria (16).

Pathogens have developed multiple strategies to evade the immune response, and one of the most commonly used is antigenic variation of immune response targets. The genetic variability observed in most microorganisms, especially those that are pathogenic, accounts for the capacity of microbial pathogens to adapt to genetically distinct hosts, different antibiotics, or immune system defense mechanisms (12). The high intra-species variability is an important limiting factor to the development of vaccines against several diseases such as HIV, hepatitis C virus (HCV), Plasmodium, and Meningococcus. Genotypic differences among infecting Mycobacterium tuberculosis strains have also been pointed out as potential factors influencing the efficacy of the Bacillus Calmette-Guérin (BCG) vaccine against tuberculosis (17). This vaccine has been extensively used and studied, and shows wide variation in efficacy among populations and geographic regions worldwide (18).

Entire microorganism genome sequences can be obtained relatively inexpensively and quickly, a fact that makes it feasible to have whole genome data available from multiple isolates of a single pathogen. This technological advance has provided the basis for the exciting, newly proposed concept of Pangenome (19). According to this concept, a bacterial species could be characterized by its global gene repertoire, which contains genes shared by all the strains (core genome) and genes shared by some strains but not all (dispensable genome). Additionally, genetic variants in genes that are found in all strains can be characterized by DNA sequence analysis. This concept has been used to characterize intraspecies diversity in a novel approach for the discovery of new vaccine candidate molecules: Pangenomics reverse vaccinology (20). Thus, the development of effective, multivalent and broad coverage genome-based vaccines will probably result from approaches that include genomes from multiple isolates from each microbial pathogen. Yet, underrepresented strains not targeted by the multivalent vaccine may become a matter of concern after broad vaccine coverage is established. Serotype replacement, a phenomenon by which minor bacterial serotypes become predominant over time, has been documented after the introduction of the heptavalent anti-pneumococcal conjugate vaccine. It can be attributed to the fact that strains not targeted by the vaccine are able to escape vaccine selective pressure (21).

The ongoing efforts of data mining for all the information in pathogen genomes already made available should effectively help to identify new potential vaccine targets for most infectious diseases (Figure 1). However, these new approaches using powerful technologies (genomic and pangenomic reverse vaccinology; functional and structural genomics) have advantages and limitations that should be considered in the rational design of a vaccine development project.

In general, the methodologies used for identification and analysis of vaccine candidate antigens are based on the assumption that surface and secreted proteins will induce a protective humoral immune response. That is true for almost all extracellular pathogens, but the presence of antibodies is not always associated with protection against intracellular pathogens. Then, eliciting both antibody and T-cell immune responses can be an important property of the candidate antigens to be evaluated.

In fact, despite the amount of available data about the genome of pathogens and the human genome, the genes, molecules, and mechanisms involved in pathogen virulence or in host defense are still not clear. Understanding the mechanisms of microbial infections and the immune response against them remains a major challenge in vaccine development. This could explain why relatively slow progress has been made in the development of new efficient vaccines in spite of the emergence of high-throughput sequencing technologies and advances in bioinformatics. Much research is still needed for novel vaccine development: the contribution of each microbial gene expressed to the disease and to the activation of a protective immune response should be understood. The current data show that there is a wide variety of new molecules that can potentially constitute new vaccine targets; the choice of the selection methods for these new vaccine antigens is, thereby, a critical step and depends on the accumulated knowledge in the fields of bioinformatics, genomics, immunology, pathology, microbiology, parasitology, and virology.

Group-targeted vaccines? ::: Introduction:

Host-pathogen interactions are also influenced by the host background. Thus, different animal models or human populations, age, specific nutrient deficiencies, co-infection/co-morbidities, among other host conditions can influence the immune responses to a given pathogen or vaccine.

There is accumulating evidence for a genetic control of infectious diseases. Rare primary immunodeficiencies have been shown to cause susceptibility to either multiple infectious agents or to a single type of infectious agent; these are very rare and Mendelian inherited. Twin studies have provided further evidence of a genetic control, and complex segregation analyses that explicitly model the effect of genes and the environment have revealed either multigenic predisposition or a genetic predisposition based on major genes having a big influence on the phenotypes related to the disease (22). Several candidate genes have been associated with infectious diseases (23). The candidate gene approach further supports the role of human genetics in resistance or susceptibility to infectious diseases. More recently, RNA interference has been used to exploit human cell models of infection with various agents to elucidate pathogen virulence mechanisms capable of subverting host pathways involved in immune regulation and pathogen killing response (24), an approach that can be used to discover and associate differences in preferential immunoregulatory networks involved in host-pathogen interactions in susceptible versus resistant hosts. These approaches are, however, restricted to well-known genes.

Several genome-wide linkage analyses have been performed to map new resistance or susceptibility genes. Major genes have been mapped to chromosomes 5q31-q33 (25) and 6q23 (26) for schistosomiasis, chromosomes 8q12 (27) and 15q/Xq (27,28) for tuberculosis, chromosomes 10p13 (29,30) and 6q25 (30) for leprosy, and chromosome 22q12 for visceral leishmaniasis (31). Blood infection loads of Plasmodium falciparum or mild malaria have been genetically linked to chromosomes 5q31-q33 (32), 5p15 (33), 6p21-23 (34), and 13q12-22 (35). The identification of genes underlying complex diseases remains a challenging task, and few genes have been identified. Nevertheless, IL13 (chromosome 5q31-q33), TNF and NCR3 (chromosome 6p21-p23), PARK2 and PACRG (chromosome 6q25) polymorphisms have been associated with schistosomiasis (36), mild malaria (37,38), and leprosy (39), respectively.

The first genome-wide association study in infectious disease was published by Fellay et al. (40), who found an association of HLA-B and HLA-C with HIV virus load, and of RNF39 and ZNRD1 with disease progression. More recently, several genome-wide association studies have been published for malaria, hepatitis B, and hepatitis C. The best signals of association were found at the HBB locus for severe malaria (35), at the HLA-DP locus for hepatitis B (41), and at the IL28B locus for hepatitis C (42,43). All the studies were based on single nucleotide polymorphism (SNP) arrays: 500,000 SNPs were genotyped in the HIV, malaria, and hepatitis B studies, while 900,000 SNPs were analyzed in the hepatitis C study.

Polymorphisms affect vaccine responses ::: Introduction:

Twin studies have recently established that genetic variation influences the response to hepatitis A and B, diphtheria, tetanus, measles, mumps, rubella, polio, H. influenzae type b, pertussis, and BCG vaccines (44). Such studies have led to estimate the heritability, defined as the ratio of genetic variance to total variance, of immune responses (14). Early papers have reported a heritability of 61 and 77% for the antibody response to the hepatitis B surface antigen vaccine in adults (45) and young children (46), respectively; it should be stressed that high levels of antibody against hepatitis B surface antigen correlate with protection against infection and persistent carriage (47). The heritability of vaccine responses is generally high (44), reaching 89% for the antibody response against measles vaccine (16). It should be stressed that the immune mechanisms responsible for protection are largely unknown for a variety of infectious diseases. Also immunological correlates of protection are needed to monitor the efficacy of vaccines and to study the genetics of protective responses to vaccines (48).

The influence of the genetic background has been further supported by twin-based and population-based association studies, which have revealed the association of HLA and non-HLA candidate genes with the response to hepatitis A and B, influenza, measles, rubella, and BCG vaccines (44,49). However, few candidates have been simultaneously tested by each research group, except for the study by Hennig et al. (47), who analyzed 133 candidate genes. It is likely that there are genes involved in both resistance to a particular infectious disease and response to the corresponding candidate vaccines. Therefore, genes that have been associated with resistance or susceptibility to infection or disease can be considered candidates potentially involved in the response to the vaccine. The genes that will be identified through genome-wide approaches in infectious diseases will be of particular interest; conversely, the genes that affect the response to vaccination may be analyzed for their potential role in resistance or susceptibility to the corresponding disease. It may be the time to move from the candidate gene approach to genome-wide linkage and/or association approaches in order to identify genes controlling the response to vaccination (Figure 1).

It is very likely that populations with different genetic backgrounds will differ in their genetic control of infectious diseases: for a particular infectious disease, the involvement of some loci may be evidenced in a population living in Africa and may not be detected in another one living in Asia. Ethnic groups living in the same area can also differ in their ability to control infection or disease. For example, the Fulani, whose genetic background is clearly different from that of other ethnic groups in West Africa, have been compared to sympatric ethnic groups for several phenotypes related to malaria resistance. The Fulani produce higher antibody levels than the Mossi and the Rimaibe, and have lower parasite densities and fewer and milder malaria attacks, indicating that the Fulani are more resistant against malaria than the Mossi and the Rimaibe (50). Similarly, it is known that populations differ in their response to vaccination, suggesting that populations differ in the frequency of the protective alleles. Poland et al. (51) have reported that native (Innu and Inuit) and Caucasian schoolchildren differed in their production of antibodies against measles antigens after vaccination in Canada. Additionally, Malawi and UK populations strikingly differed in the level of protection following BCG vaccination (i.e., 0% in Malawi and 80%

in the UK), and in their IFN-γ production in response to mycobacterium antigens (52). Furthermore, native children (Navajo, White Mountain Apache and Alaska) and children in the general US population differed in the incidence of disease due to H. influenzae type b both before and after vaccination, indicating that native children were more susceptible than other children in the US, and that their response to vaccination was also weaker (53).

Further understanding of the genetic basis of variation in vaccine response may lead investigators to consider different vaccine strategies for groups having different genetic backgrounds. The combination of high throughput methodologies and multifactorial analyses may help in the identification of genetic markers that affect immune and physiological responses leading to serious adverse effects, thus allowing the identification of groups at risk of vaccine-induced adverse events, which would be entitled to preventive therapies or alternative vaccine formulations (54). In the same way, the exposure to the infectious agents, current immune status (e.g., immunocompromised subjects), gender, and age could be taken into account to minimize the rate of vaccine failure or vaccine adverse events. Such a targeted-group vaccination approach requires more information about the human groups of interest. Efforts are particularly needed to study various populations in Africa, where genetic diversity is the most important in the world, and data are lacking. In these populations, the international scientific and medical community would be able to screen for millions of polymorphisms within each potential target group.

It has been also suggested that profiles based on individual information, such as genetic information, age, and gender, would allow prediction of the need to receive a given vaccine (i.e., being susceptible to infection), the number and amount of doses needed, the likelihood of response, the likelihood of adverse events, and so on, thereby making it possible to develop a personalized vaccination approach. At this stage, developing a health system in which each individual would be genotyped at the genome scale would become a political and economic choice. Before coming to the era of personalized vaccination, however, we should develop methods that will eventually allow us to predict the individual response to vaccination on the basis of individual information, including genome-wide genotypes. The challenge is less technological than conceptual.

Molecular networks and genetic control ::: Introduction:

The emergence of new technologies, such as the microarray technology and more recently the next generation sequencing technology, gives the illusion that genes controlling a complex phenotype would be easily identified. This is due to the fact that we can obtain a huge amount of biological data for one sample. For example, the whole-genome transcriptome can be analyzed by using one array for an individual, and more than 500,000 SNPs can be simultaneously genotyped by using SNP arrays for the same individual (the last generation of Affymetrix chips allows the analysis of 1,8 million SNPs). The next-generation sequencers (solexa/illumina and solid/ABI) allow the analysis of the transcriptome, of small RNAs, of DNA methylation, of SNPs, and of copy number variation at the genome scale (55). All these technologies can produce a huge amount of data that need to be stored, transferred, carefully analyzed and interpreted. The statistical analysis is a significant challenge because a number of statistical tests are performed; to tackle this multi-testing problem, new approaches have been proposed (55,56). Such approaches have been used in genome-wide studies and have provided significant results from the statistical point of view. However, it is clear that most biological information is missing. This lack is highlighted by the results obtained in genome-wide linkage and association in infectious diseases. For instance, Fellay et al. (40) estimated that the polymorphisms identified explained 15% of the virus load variance in the study population, suggesting that the effect of other genes had not been detected. Strikingly, a genome-wide association study only detected the association between protection against severe malaria and hemoglobin S after correcting for multi-tests (35), while several other genes and genetic variants have been associated with severe malaria in independent populations (57).

It is clear that the biological interpretation of such data needs an integration of various biological data stored in different databases, and statistical approaches that make use of these data. The central concept that can drive the development of new approaches is the molecular network. Figure 2 represents this concept in the context of infectious diseases.

This concept has stimulated the development of gene prioritization tools to propose candidate genes lying within a chromosomal region linked to the disease. The approaches used imply that most genes involved in a disease share functional properties and/or can be mapped in the same gene or protein network. The most frequent approach is to rank candidates on the basis of their similarity to a set of training genes that have been demonstrated to be involved in the disease (reviewed in Ref. 58). The biological concept of the gene network implies the existence of gene-gene interactions that are generally not taken into account in genome-wide linkage and association analyses. Gene-gene interaction can be explicitly included in the models used in linkage and association analyses. Statistical approaches have been proposed to detect such interactions also in studies of vaccine responses (59). Furthermore, the influence of TLR2 and TIRAP polymorphisms on clinical disease caused by M. tuberculosis has been found to depend on pathogen genotypes, suggesting the existence of bacterial gene-human gene interactions (60). Obviously, the gene-gene interaction studies at the genome scale pose additional multitest problems that must be tackled. In addition, such approaches are generally limited to low order interactions, and efforts are now needed to develop statistical approaches that make use of the whole biological network. The analysis of genetic variation and gene expression levels could also provide help in understanding how polymorphisms could perturb a transcriptional network in infectious diseases (61). Microarray studies that are based on mice infected with either virulent or nonvirulent pathogens have identified gene expression patterns associated with some infectious diseases, indicating the existence of at least one transcriptional network perturbed by a virulent strain. Moreover, gene expression profiles have been shown to discriminate between genetically cerebral malaria-resistant versus genetically cerebral malaria-susceptible mice at early and late stages of infection with P. berghei ANKA (62,63). This further indicates that genetic variants direct at least one transcriptional network towards a state conducting to cerebral malaria and death in susceptible mice after infection. In humans, blood cell gene expression profiles have been associated with the clinical status of patients infected with different pathogens (64), and have been used to monitor the response to BCG vaccination in children (65).

Although very few genomic studies have been performed in the field of vaccination it has been anticipated that the biological network concept would be useful. In this way, Poland et al. (44) proposed “the immune-response network theory”, whereby the response to the vaccine can be viewed as the cumulative result of gene interactions. The findings in naturally infected individuals and in mouse models of infectious diseases support this hypothesis. This urges the effort to perform genome-wide studies in vaccinated mice and humans, and to compare the results with those obtained in natural infections. The same authors have suggested that the network combined with individual genomic profiles may lead to predicting the response to vaccination. This is a big challenge because it will require biological networks that capture all the relevant information, including the effect of genetic and other biological information (age, gender, etc.). This further implies that one would be able to predict the effect of single or combined perturbations such as polymorphisms on molecular phenotypes (Figure 2). For this purpose, regulatory network modeling methods would be useful: models constructed on available data to predict molecular and cellular phenotypes will provide working hypotheses that can be tested, leading to the improvement of the initial model. Interestingly, immune response system-level tools have been made available (66). Thus, Systems Biology is a new rational approach that gives a more comprehensive understanding of networks or interacting components in the immune response to pathogen antigens. This concept has been recently reviewed (67). It can be applied to identify signatures of protection that would be evaluated in the design of new vaccines or predict vaccine efficacy (Systems Vaccinomics). This emergent research strategy in vaccine development was already applied to identify signatures of protection to influenza (68) and yellow fever vaccines (69). Nevertheless, the current tools can manage small biological networks; the challenge is to develop new approaches that allow the modeling of large regulatory networks.

In conclusion, genomics of infectious agents will accelerate the identification of new relevant vaccine candidate antigens, including antigens that are conserved between strains or collectively cover the diversity of strains and that are the targets of protective immune responses. The genome-based approaches should help in developing vaccines against old, but so far uncontrolled, infectious diseases that are expected in developing countries, and may play a critical role in the design of vaccines against emerging infectious diseases as well as non-infectious diseases. The identification of relevant molecular targets will require further research on protective immune responses and the development of immunological tools. Recent investigations that have systematically searched for viral protein-human protein interactions open new avenues for understanding host-pathogen interactions and therapeutic strategies. Further research on the mechanisms of action of adjuvants and the use and restraints of relevant animal models will be mandatory. The human genetic variation should also be taken into account when analyzing the response to vaccination. Vaccine strategies may have to be adapted to target groups in order to optimize vaccination efficacy and reduce the risk of adverse events. Similarly, the possibility of a personalized vaccination strategy (appropriate dose, schedule and even molecular components of the vaccine?) has been suggested, which would be based on individual information, including genome-wide genetic information. This implies the ability to predict an individual response to the vaccine. The challenge is to develop new tools that are necessary to accurately construct and model molecular networks, the perturbation of which causes either the absence of response or adverse events.