2011

Harnessing the power of genomics and immunoinformatics to produce improved vaccines

Leonard Moise
University of Rhode Island, lmoise@uri.edu

Leslie Cousens

See next page for additional authors

Citation/Publisher Attribution
Moise, L., Cousens, L., Fueyo, J., & De Groot, A. S. (2011). Harnessing the power of genomics and immunoinformatics to produce improved vaccines. *Expert Opinion on Drug Discovery, 6*(1), 9-15. doi: 10.1517/17460441.2011.534454
Available at: https://doi.org/10.1517/17460441.2011.534454

Published in final edited form as:
Expert Opin Drug Discov. 2011 January ; 6(1): 9–15. doi:10.1517/17460441.2011.534454.

Harnessing the Power of Genomics and Immunoinformatics To Produce Improved Vaccines

Leonard Moise[1,2,*], **Leslie Cousens**[2], **Joanna Fueyo**[3], and **Anne S. De Groot**[1,2,4]

[1]Institute for Immunology and Informatics, University of Rhode Island, 80 Washington Street, Providence, Rhode Island 02903, USA

[2]EpiVax, Inc., 146 Clifford Street, Providence, Rhode Island 02903, USA

[3]Department of Biomedical and Pharmaceutical Sciences, INBRE Program, College of Pharmacy, 41 Lower College Road, University of Rhode Island, Kingston, Rhode Island 02871, USA

[4]Department of Pediatric Infectious Disease, Brown University Warren Alpert Medical School, Providence, Rhode Island 02912, USA

Abstract

The role of cellular immunity as a mediator of protection against disease is gaining recognition, particularly in regard to the many pathogens for which we presently lack effective vaccines. As a result, there is an ever-increasing need to understand the T cell populations induced by vaccination and therefore T cell epitopes responsible for triggering their activation. Although the characterization and harnessing of cellular immunity for vaccine development is an active area of research interest, the field still needs to rigorously define T cell epitope specificities, above all, on a genomic level. New immunoinformatic epitope mapping tools now make it possible to identify pathogen epitopes and perform comparisons against human and microbial genomic datasets. Such information will help determine whether adaptive immune responses elicited by a vaccine are both pathogen-specific and protective, but not cross-reactive against host or host-associated sequences that could jeopardize self-tolerance and/or human microbiome-host homeostasis. Here, we discuss advances in genomics and vaccine design and their relevance to the development of safer, more effective vaccines.

Keywords

genomics; microbiome; immunoinformatics; tolerance; vaccine

1. Introduction

Since antiquity, the story of “human against microbe” has been evolving: humans have long battled against infectious disease, sometimes prevailing, sometimes succumbing and, at

[*]Corresponding author: Leonard Moise, PhD, EpiVax, Inc., 146 Clifford Street, Providence, RI 02903, 401.272.2123 (v), 401.272.7562 (f), lmoise@epivax.com.

Declaration of Interest: L Moise, L Cousens and A de Groot are all employees of EpiVax Inc with A de Groot a Senior Officer and shareholder of the company. Additionally, J Fueyo has conflicts of interest.

other times, arriving at a stalemate with unfriendly pathogens. The advent of vaccination, whereby the immune system is trained to recognize and fight infection without requiring exposure to the pathogen, gave humans a powerful tool for survival in the evolutionary tussle between human and microbe. Vaccines save millions of lives, serve as a cost-effective means of fighting disease by prevention rather than by treatment, and ideally protect against re-infection. Overall, vaccines represent one of the greatest contributions to public health, yet there remains a great need to slow or stop the charge of many pathogens that have not yielded to traditional vaccine development approaches. Here, we discuss the promise of a genomes-to-vaccine strategy and the importance of emerging genomic data and immunoinformatics tools to the development of safer, more effective vaccines.

2. Genomes-to-Vaccines

A genomes-to-vaccine strategy for rational vaccine design rests on the premises that (i) a minimal set of immunogens capable of inducing a robust and sustained immune response to a pathogen can be discovered using immunoinformatics, and (ii) administration of these immunogens together with adjuvant in a suitable delivery vehicle will result in protection from disease. This diverges from conventional approaches where whole organism vaccines (live, live attenuated, killed/inactivated) provide more information than the immune system needs for protection, thereby presenting unnecessary safety risks for activation of undesired immune reactions. The genomes-to-vaccine approach aims to identify the minimal, essential information derived from pathogen genomic sequences that is needed to achieve protection while avoiding unintended and dangerous effects. Using bioinformatic/computational methods, it is possible to tease out this information, while also taking into account the sequence data representative of humans, commensal microbial species and infectious pathogens. Adherents to this approach believe that the minimum essential data is encoded, at least in part, by T cell epitopes, short peptide sequences that bind major histocompatibility complex (MHC) proteins and are displayed by antigen presenting cells to T cells, critical mediators of early processes in adaptive immunity. CD8 and CD4 T cells play an important role in containing infections and may be critical correlates of protection after re-exposure. CD4 T cell responses are critical for robust CD8 T cell proliferation and function and for their differentiation into memory cells [1,2]. Moreover, CD4 T cell responses provide required help to B cells to produce antibodies that are the mediators of protection in all currently approved human vaccines [3]. While B cell epitopes undeniably also contribute to protective immunity, our focus here is on T cell epitope identification and assessment through computational screening of genomic sequences.

There are four major steps that comprise the genomes-to-vaccine strategy, which can be thought of as a funneling process (Figure 1) [e.g. 4,5]: (i) Genomes are mined using computational tools to identify genes that encode proteins with promising antigenic properties such as secretion, up-regulated expression, reported immunogenicity and virulence [e.g. 6,7,8]. Alternatively, with no bias introduced by knowledge of protein function, expression kinetics or localization, complete open reading frame datasets are analyzed for the reason that the immune system is omnivorous and may present sequences from any protein antigen to stimulate T cell responses. (ii) Immunoinformatics tools are then used to discover, within protein sequences, short, linear, putative T cell epitopes. As the

number of sequenced genomes increases, it becomes increasingly impractical to identify T cell epitope vaccine candidates by experimental means, even in high-throughput screens. Robust computational algorithms and hardware, by comparison, can handle the data surge and expedite the process. (iii) Next, selected sequences are synthesized as peptides and evaluated for human leukocyte antigen (HLA) binding and antigenicity in survivors of infection or vaccinees. (iv) Finally, prototype epitope-based vaccines are evaluated for immunogenicity and efficacy in humanized mice. Epitopes may be formulated as multiple peptides or as concatamers in the form of DNA or protein vaccines, which elicit CD4 and CD8 T cell responses [e.g. 9,10,11]. In addition to antigen identification, clinical development of genome-derived, epitope-based vaccines requires co-formulation of antigens with adjuvant in a suitable delivery vehicle. As adjuvant and delivery technologies mature in the broader field of vaccinology, they, too, provide important lessons that will help epitope-based vaccines achieve requirements for clinical approval.

Having performed this process a number of times [12,13,14,15] our strategy is to require careful selection of epitopes at the computational phase to maximize the likelihood of success in bench-based experiments. To enable this computational approach, it is essential to consider not only the genomic sequence data of the pathogen of interest, but also that of human and human microbiome genomes.

3. Microbial Genomes

The Institute for Genomics Research (TIGR), now part of the J. Craig Venter Institute, sparked the field of genome sequencing in 1995 when it published the first complete sequence of a living organism, *Haemophilus influenzae*. That it was possible to overcome the computational enormity of assembling a shotgunned 1.8 Mb genome represented a major milestone that heralded the completion of new microbial genome projects in the years that followed. In recent years, whole genome sequences became available in public databases, which enabled genomic and proteomic comparisons for the study of large-scale patterns in pathogenic microbial sequence data that aid in understanding microbial ecology and host-pathogen interactions. In the process, these studies led to the development of algorithms and computational methods that have become essential to harnessing sequence data for understanding, preventing and treating infection (see Darren Flower's excellent text, *Bioinformatics for Vaccinology*, for a full review of the field [16]). Furthermore, they have yielded important new concepts of direct import to the field of vaccines that are discussed below.

Microbial species are comprised of both a set of core and dispensable genes, with the sum of these two components forming the species pan-genome [17]. The core genome comprises genes found in all strains of a species. These genes may contain immunogenic elements, or epitopes, that are highly conserved between pathogens. Conservation may be advantageous, expanding the breadth of coverage by a vaccine, or disadvantageous, impinging on the survival of beneficial commensals that possess similar sequences. The dispensable genome represents genes unique to single strains and those found in two or more, but not all, strains within a species. These genes can be of great interest to vaccine developers, especially if they are represented in a large proportion of circulating strains. A comparative genomic

analysis recently depicted that, because the number of unique genes is so large, the pan-genome of a bacterial species can be significantly larger than any single member genome [18]. Thus, careful, rational selection of vaccine targets becomes one of the most important first steps in the genome-to-vaccine process.

The breadth of genomes, and their potential for variation over time, remains a significant concern for genome-to-vaccine developers. Future studies will investigate whether core genomes are static or dynamic. There are not enough computational data sets for this determination to be made at present, although initial observations may be made for select cases. For instance, an analysis of 96 genome sequences derived from two closely related sympatric sister species of pathogenic bacteria - *C. coli* and *C. jejuni* - showed that the core genome is thought to be resistant to interspecies recombination between the two sister species, which supports the concept of a finite, species-specific core genome [17]. A finite genome in this context means that there are unique and cohesive features to each genome analyzed; these features define genomic identity, and may be of great importance for the development of species-specific vaccines.

Examination of genomic content including laterally transferred sequences within and between species, and core vs. dispensable genome data sets, will help to elucidate further detail regarding genome plasticity, as well as characteristics and features of each species. Because evolution of species-specific core genes possibly reflects different adaptive strategies of that species [17], these analyses could reveal adaptive strategies a species uses to inhabit, and possibly cause disease in its specific niche. Furthermore, laterally transferred sequences may impact on immune responses to commensals or other pathogens, if the sequences contain epitopes highly conserved between species.

Thus, the definition of genomic identity is essential for selection of species-specific antigens that can provide protection expressly against the pathogen of interest, leading to more focused vaccine designs and, potentially, improved safety and efficacy. Moreover, characterization of core genome plasticity over time and geography provides important information that directly impacts the likelihood that a vaccine will offer protective efficacy as the pathogen evolves. Antigens selected for T cell epitope identification will ideally be chosen from stable elements of a genome.

4. Human Genome

As important as it is to know which epitopes to include in a vaccine, it is just as important to know which epitopes to exclude (Figure 2). The potential for T cell cross-reaction between pathogen and human sequences may cause autoimmunity in genetically susceptible individuals, which is precisely the opposite intent of vaccination. For example, in individuals with treatment-resistant Lyme arthritis, cross-reactivity between human LFA-1 protein and the *Borrelia burgdorferi* OspA protein was shown to be the result of similar MHC-bound peptides derived from these antigens [19]. Interestingly, the recombinant OspA Lyme disease vaccine, LYMErix®, was withdrawn from the market following patient complaints about autoimmune side effects, although no direct connection was rigorously

demonstrated. Nonetheless, this incident sends a cautionary signal to vaccine developers to consider factors that may lead to autoimmune pathology.

We have implemented a straightforward approach to limit the possibility of cross-reactivity and autoimmunity in epitope-based vaccines [20]. Essentially, a BLAST search is performed with candidate epitope sequences against the human genome. Epitope sequences which are homologous to components of the human genome are excluded from the epitope data set to be considered for use in vaccine development, while the remaining “foreign” epitopes can be safely included in vaccine formulations. Recently, novel human sequences absent from the existing reference genome have been discovered in different individual genomes, illustrating that the human genome is also composed of individual-specific and core sequences [21]. To effectively screen candidate vaccine epitopes for potential cross-reactivity in all potential vaccinees, coding sequences in the human pan-genome need to be considered, when they become available.

5. Human microbiome genomes

Microorganisms that live symbiotically in humans are collectively known as the human microbiome. Although it is estimated that these microbes outnumber human cells by a factor of 10 to 1, little is known about the effect these complex human microbial communities have on human health. The NIH has prioritized the understanding of the human microbiome as part of its Roadmap for Medical Research and has funded a significant effort to map and sequence the human microbiome, thus enabling an analysis of its role in human health and disease [22]. Analysis of the first 178 reference microbial genomes was published this year by members of the Human Microbiome Project Jumpstart Reference Strains Consortium [23], not only illustrating, by sheer volume of genomic sequences, the distance we have come since the first bacterial genome was published, but also facilitating the expedited development of vaccines. The data also suggest that the impact of vaccination on the human microbiome may be significant because of the possibility that sequence similarities between vaccine antigens and the human microbiome can potentially activate T cell-mediated mechanisms that perturb natural tolerance to commensal organisms and disrupt the natural symbiosis between human and microbial flora. Identification of such homologous sequences is important to enable rational design of vaccines that focus on pathogen-specific epitopes without disrupting the host-microbiome interface (Figure 2).

6. Conclusion

Identification of core and dispensable genes in microbial genomes will not only help to elucidate the genomic roots of biological patterns inherent in pathogenesis, commensalism and genome evolution, but also aid in the careful selection of antigens to include and exclude from vaccines that elicit effective immunity against pathogenic organisms. Consideration of these concepts in vaccine design will make it possible to preserve the balance between human and commensal microbial communities, as well as control of regulatory mechanisms that prevent onset of autoimmune disease. This rigorous evaluation process will lead to a new generation of vaccines that will be better poised to control and even eradicate disease.

7. Expert Opinion

Vaccines are an effective means of preventing infectious disease. However, certain vaccines have been associated, either rigorously or anecdotally, with a number of adverse events and autoimmune sequelae. Detrimental effects associated with vaccines have raised concerns with the general public, dissuading some from participating in vaccination programs. The burden is on the scientific community to continue to conduct research on purported vaccine-associated complications, and to provide the public with reasonable assurances that recommended vaccines are as safe as possible.

A key issue that requires study is the role of T cells in adverse reactions to vaccines. This is because there is convincing evidence consistently building over time that demonstrates a significant role for T cell responses in vaccines that show clinical efficacy through humoral immunity. Moreover, recognition is growing among vaccinologists for the challenges involved in defining the correlates of vaccine-mediated protection against pathogens that may require a combination of humoral and cellular immunity to develop an effective host defense. T cells are key mediators of inflammatory processes that can be activated by vaccine epitopes.

Although rare, some vaccine epitopes may cross-react with human/microbiome epitopes, causing beneficial or detrimental effects. This phenomenon is known as heterologous immunity. In regards to vaccine development, heterologous immunity can be exploited to provide broad-based immunity to a family of related organisms. It can also describe epitopes included in vaccines with the potential to induce unintended immune responses against epitopes of human or human-associated (i.e. within the human microbiome) origin. The identification of relevant pathogen epitopes and their comparison against other pathogen, human and microbial genomic datasets are necessary steps to ensure that adaptive immune responses elicited by a vaccine are both pathogen-specific and protective, but not cross-reactive against host or commensal epitopes. Identification of cross-reactivities in vaccine design will lead to safer and more effective vaccines.

Heterologous immunity is a growing area of research that investigates more than the impact of T cell-dependent cross-reactive immunity between species. In a larger sense, it considers how past infections in individuals impact the shape immune responses take in later, unrelated infections. For example, patterns of $CD8^+$ T cell cross-reactivity between epitopes encoded by Epstein Barr virus and influenza A virus in humans, and between vaccinia virus and lymphocytic choriomeningitis virus in mice, are associated with varied disease states observed in infection [24,25]. Interestingly, similar differences may also be influenced by immunization. It was shown that vaccination with the bacillus Calmette-Guérin (BCG) strain of *Mycobacterium bovis*, which is used to immunize against *Mycobacterium tuberculosis*, induces $CD4^+$ T cell-mediated resistance to vaccinia infection [26]. This is significant because it implies that vaccine designers and clinicians should consider not only the standard issues of safety and efficacy upon immunization in the present but also the impact of immune memory on future immune responses to vaccination and natural infection. Taken altogether, we believe it is important to document and evaluate the immunological significance of epitope conservation and cross-reactivities between common vaccines and

commensal and pathogenic microbes to understand the nexus between microbe and host. It will require an understanding of immune responses using complementary experimental and computational approaches, involving the latest tools for single cell analysis of cross-reactive T cells [27] and immune system modeling [28]. Together, they will transform the current qualitative understanding of heterologous immunity into a quantitative analysis that can be used to predict beneficial and detrimental effects of immunization in individuals.

Acknowledgments

All authors were supported by an NIH grant (NIH 1U19AI082642)

References

1. Bevan MJ. Helping the CD8+ T-cell response. Nat Rev Immunol. 2004; 4:595–602. [PubMed: 15286726]
2. Bachmann MF, Wolint P, Schwarz K, Oxenius A. Recall proliferation potential of memory CD8+ T cells and antiviral protection. J Immunol. 2005; 175:4677–85. [PubMed: 16177115]
3. Edghill-Smith Y, Golding H, Manischewitz J, et al. Smallpox vaccine-induced antibodies are necessary and sufficient for protection against monkeypox virus. Nat Med. 2005; 11:740–7. [PubMed: 15951823]
4. McMurry JA, Gregory SH, Moise L, et al. Diversity of Francisella tularensis Schu4 antigens recognized by T lymphocytes after natural infections in humans: identification of candidate epitopes for inclusion in a rationally designed tularemia vaccine. Vaccine. 2007; 25:3179–91. [PubMed: 17291638]
5•. Gregory SH, Mott S, Phung J, et al. Epitope-based vaccination against pneumonic tularemia. Vaccine. 2009; 27:5299–306. Remarkable protection elicited by 14 HLA Class II epitopes against lethal challenge with a bacterial pathogen. [PubMed: 19616492]
6. De Groot AS, Jesdale B, Martin W, et al. Mapping cross-clade HIV-1 vaccine epitopes using a bioinformatics approach. Vaccine. 2003; 21:4486–504. [PubMed: 14505932]
7. Rappuoli R, Covacci A. Reverse vaccinology and genomics. Science. 2003; 302:602. [PubMed: 14576423]
8. He Y, Xiang Z, Mobley HL. Vaxign: the first web-based vaccine design program for reverse vaccinology and applications for vaccine development. J Biomed Biotechnol. 2010; 2010:297505. [PubMed: 20671958]
9. Hanke T, McMichael AJ, Mwau M, et al. Development of a DNA-MVA/HIVA vaccine for Kenya. Vaccine. 2002; 20:1995–8. [PubMed: 11983261]
10. Bazhan SI, Karpenko LI, Lebedev LR, et al. A synergistic effect of a combined bivalent DNA-protein anti-HIV-1 vaccine containing multiple T- and B-cell epitopes of HIV-1 proteins. Mol Immunol. 2008; 45:661–9. [PubMed: 17869341]
11. Cai Q, Peng G, Bu L, et al. Immunogenicity and in vitro protective efficacy of a polyepitope Plasmodium falciparum candidate vaccine constructed by epitope shuffling. Vaccine. 2007; 25:5155–65. [PubMed: 17548134]
12. Moise L, McMurry JA, Buus S, et al. In silico-accelerated identification of conserved and immunogenic variola/vaccinia T cell epitopes. Vaccine. 2009; 27:6471–9. [PubMed: 19559119]
13. Moise L, McMurry JA, Pappo J, et al. Identification of genome-derived vaccine candidates conserved between human and mouse-adapted strains of H. pylori. Hum Vaccin. 2008; 4(3):219–23. [PubMed: 18376134]
14. McMurry JA, Kimball S, Lee JH, et al. Epitope-driven TB vaccine development: a streamlined approach using immuno-informatics, ELISpot assays, and HLA transgenic mice. Curr Mol Med. 2007; 7:351–68. [PubMed: 17584075]
15•. De Groot AS, Marcon L, Bishop EA, et al. HIV vaccine development by computer assisted design: the GAIA vaccine. Vaccine. 2005; 23:2136–48. Clever vaccine design using overlapping immunogenic and conserved epitopes. [PubMed: 15755584]

16•. Flower, DR. Bioinformatics for Vaccinology. Wiley; West Sussex: 2008. A comprehensive presentation of vaccine design tools available in the age of informatics

17. Lefébure T, Bitar PD, Suzuki H, Stanhope MJ. Evolutionary dynamics of complete Campylobacter pan-genomes and the bacterial species concept. Genome Biol Evol. 2010; 2:646–55. [PubMed: 20688752]

18. Medini D, Donati C, Tettelin H, Masignani V, Rappuoli R. The microbial pan-genome. Curr Opin Genet Dev. 2005; 15:589–94. [PubMed: 16185861]

19. Willett TA, Meyer AL, Brown EL, Huber BT. An effective second-generation outer surface protein A-derived Lyme vaccine that eliminates a potentially autoreactive T cell epitope. Proc Natl Acad Sci U S A. 2004; 101:1303–8. [PubMed: 14742868]

20. De Groot AS, Martin W. From immunome to vaccine: epitope mapping and vaccine design tools. Novartis Found Symp. 2003; 254:57–72. [PubMed: 14712932]

21•. Li R, Li Y, Zheng H, et al. Building the sequence map of the human pan-genome. Nat Biotechnol. 2010; 28:57–63. Novel human DNA sequences discovered. [PubMed: 19997067]

22. Turnbaugh PJ, Ley RE, Hamady M, et al. The human microbiome project. Nature. 2007; 449:804–10. [PubMed: 17943116]

23. Nelson KE, Weinstock GM, Highlander SK, et al. Human Microbiome Jumpstart Reference Strains Consortium: A catalog of reference genomes from the human microbiome. Science. 2010; 328:994–9. [PubMed: 20489017]

24•. Kim SK, Cornberg M, Wang XZ, et al. Private specificities of CD8 T cell responses control patterns of heterologous immunity. J Exp Med. 2005; 201:523–33. Diversity of T cell specificities in individuals accounts for variation in pathogenesis of infection. [PubMed: 15710651]

25. Cornberg M, Chen AT, Wilkinson LA, et al. Narrowed TCR repertoire and viral escape as a consequence of heterologous immunity. J Clin Invest. 2006; 116:1443–56. [PubMed: 16614754]

26. Mathurin KS, Martens GW, Kornfeld H, Welsh RM. CD4 T cell-mediated heterologous immunity between mycobacteria and poxviruses. J Virol. 2009; 83:3528–39. [PubMed: 19193795]

27. Shen ZT, Brehm MA, Daniels KA, et al. BI-specific MHC heterodimers for characterization of cross-reactive. T cells J Biol Chem. 2010 in press.

28. Pappalardo F, Pennisi M, Castiglione F, Motta S. Vaccine protocols optimization: in silico experiences. Biotechnol Adv. 2010; 28:82–93. [PubMed: 19833190]

29. McMurry JA, Moise L, Gregory SH, De Groot AS. Tularemia vaccines – an overview. Med Health R I. 2007; 90:311–4. [PubMed: 18019188]

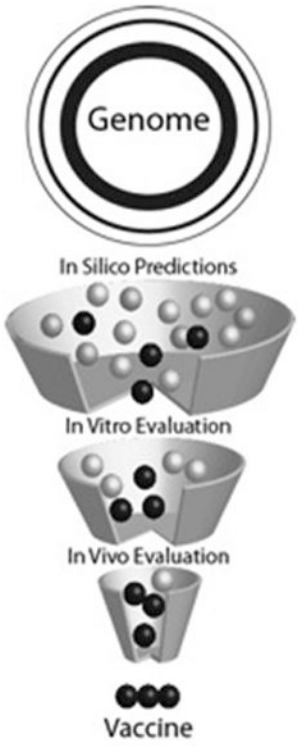

Figure 1.

Genomes-to-Vaccine Strategy. Genomes are mined using computational and experimental tools to identify genes that encode proteins with promising vaccine antigen properties such as secretion, up-regulated expression and virulence. *In silico*. Immunoinformatics tools are then used to map protein sequences for short, linear putative T cell epitopes. *In vitro*. Candidates are synthesized as peptides and evaluated for MHC binding and antigenicity. *In vivo*. Prototype epitope-based vaccines are evaluated for immunogenicity and protection in humanized mice. Used with kind permission from Medicine and Health Rhode Island [29].

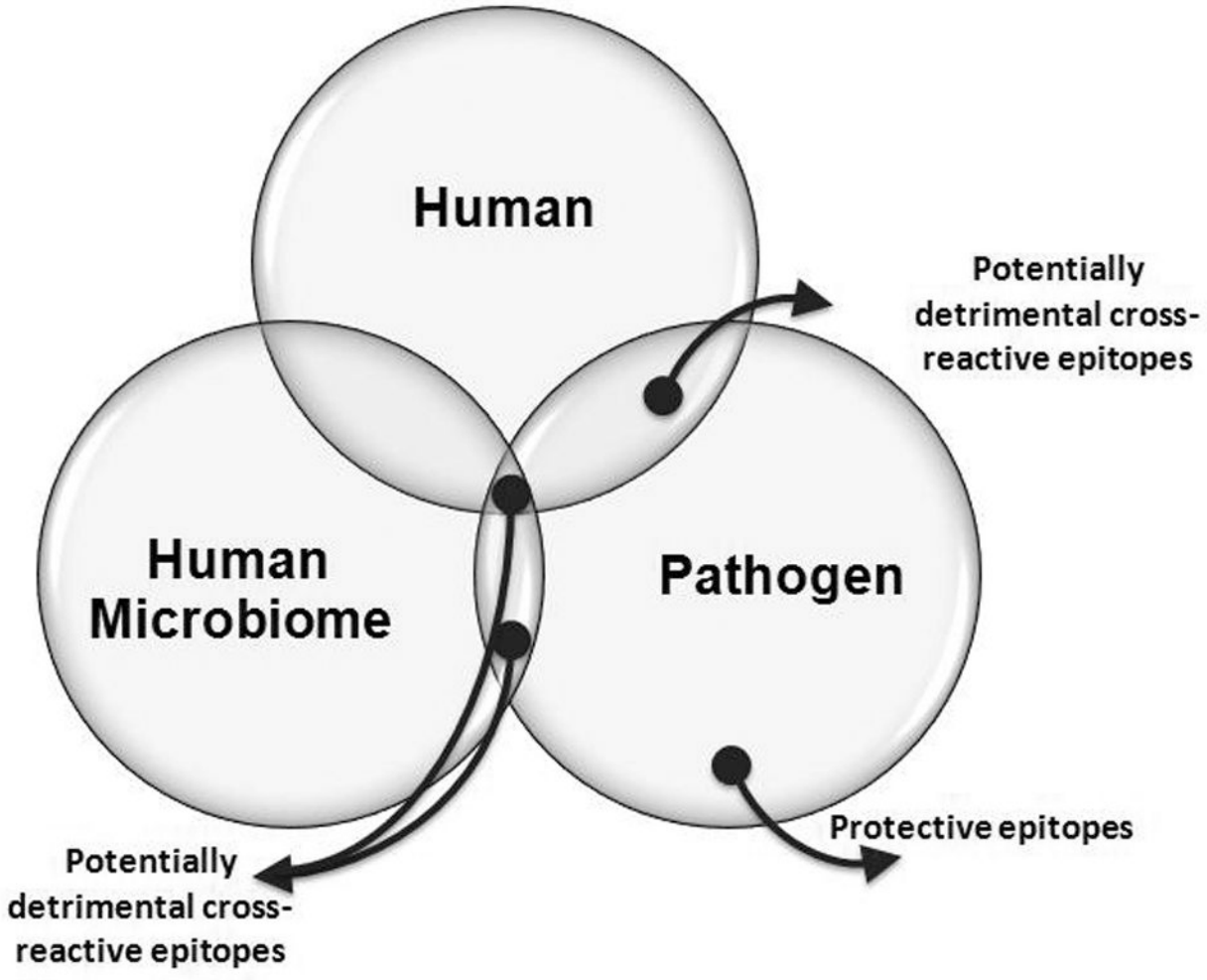

Figure 2.

Potential cross-reactive relationships among humans, commensal microbiota, and pathogens. In the process of designing a vaccine the importance of choosing epitopes for inclusion is balanced by the importance of choosing those to exclude. The definition of pathogen-specific epitopes for consideration in designing a new vaccine is a first step. To provide further assurances of safety to the vaccinee, screening of pathogen-derived epitopes against epitopes from the human genome and the microbiome is required. This secondary screening of epitopes will lead to the exclusion of those pathogen epitopes that have the potential to elicit immune reactions cross-reactive against the host or its commensal microbiota. These types of cross-reactivities could lead to initiation of autoimmune syndromes or to disruption of beneficial gut flora. Recent breakthroughs in high-throughput genomic analyses and epitope prediction algorithms make these types of comparisons possible.