vaccines

MDPI

Article

Identification of *Mycobacterium tuberculosis* Antigens with Vaccine Potential Using a Machine Learning-Based Reverse Vaccinology Approach

Blaine Teahan [1], Edison Ong [2] and Zhenhua Yang [1,*]

[1] Epidemiology Department, School of Public Health, University of Michigan, Ann Arbor, MI 48109, USA; bteahan@umich.edu

[2] Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA; edong@umich.edu

* Correspondence: zhenhua@umich.edu

Abstract: Tuberculosis (TB) is the leading cause of death of any single infectious agent, having led to 1.4 million deaths in 2019 alone. Moreover, an estimated one-quarter of the global population is latently infected with *Mycobacterium tuberculosis* (MTB), presenting a huge pool of potential future disease. Nonetheless, the only currently licensed TB vaccine fails to prevent the activation of latent TB infections (LTBI). These facts together illustrate the desperate need for a more effective TB vaccine strategy that can prevent both primary infection and the activation of LTBI. In this study, we employed a machine learning-based reverse vaccinology approach to predict the likelihood that each protein within the proteome of MTB laboratory reference strain H37Rv would be a protective antigen (PAg). The proteins predicted most likely to be a PAg were assessed for their belonging to a protein family of previously established PAgs, the relevance of their biological processes to MTB virulence and latency, and finally the immunogenic potential that they may provide in terms of the number of promiscuous epitopes within each. This study led to the identification of 16 proteins with the greatest vaccine potential for further in vitro and in vivo studies. It also demonstrates the value of computational methods in vaccine development.

Keywords: tuberculosis; vaccines; protective antigens; machine learning; reverse vaccinology

check for updates

Citation: Teahan, B.; Ong, E.; Yang, Z. Identification of *Mycobacterium tuberculosis* Antigens with Vaccine Potential Using a Machine Learning-Based Reverse Vaccinology Approach. *Vaccines* **2021**, *9*, 1098. https://doi.org/10.3390/vaccines9101098

Academic Editor: Sunhee Lee

Received: 13 August 2021
Accepted: 22 September 2021
Published: 28 September 2021

1. Introduction

Despite being an ancient disease, tuberculosis (TB) persists as one of the ten leading causes of death globally and the leading cause of death of any single infectious agent with 10 million new cases and 1.4 million deaths in 2019 alone [1,2]. The emergence of HIV and multidrug-resistant TB in the past several decades has further complicated global TB control [3,4].

In light of the significant impact of TB on global health, the World Health Organization (WHO) has prioritized the control of the TB epidemic, launching the End TB Strategy, which aspires to reduce the number of TB deaths by 95% and TB incidence by 90% by 2035, as compared to the 2015 figures [5,6]. Despite this ambition, the current pace of progress suggests that neither goal will be met [7]. Compounding the inadequacy of the current progress, the coronavirus disease 2019 (COVID-19) pandemic has impaired TB surveillance, which could increase TB mortality by 13%, undoing the last five years' progress [7].

Perhaps the greatest barrier to meeting these goals is the absence of a universally effective vaccine against all types of TB. This is in part because the unique natural history of MTB infection makes it a difficult vaccine target. MTB is able to survive many years within host immune cells, suppressing the intracellular attacks of the macrophage [8]. In order to contain these persistent bacteria, the host forms a granuloma—an aggregate of host immune cells encasing the site of infection [9,10]. This state of persistent asymptomatic infection is referred to as latent TB infection (LTBI). If the host immune system can no longer contain the bacteria, LTBI may activate to post-primary TB, including highly contagious pulmonary TB, which accounts for a majority of TB cases and deaths [1,11]. The WHO estimates that one-quarter of the global population has an LTBI, representing a huge pool of potential future disease [1]. Unfortunately, there are currently no licensed vaccines that prevent the activation of LTBI. Despite its efficacy in preventing miliary and meningeal TB in infants, the only licensed TB vaccine, Bacillus Calmette-Guérin (BCG), which is derived from an attenuated strain of *Mycobacterium bovis*, has widely variable efficacy against pulmonary TB in adult populations, ranging from 0 to 80%, being least efficacious in tropical climates, which is perhaps due to immunological sensitization to environmental mycobacteria [12,13]. Thus, in order to End TB by 2035, a new vaccine strategy that prevents both primary and post-primary TB in all age groups must be developed.

One promising approach that may compensate for the inadequacies of BCG in sensitized populations is the use of subunit vaccines given in conjunction or succession with BCG [12]. The first attempt at this strategy was the development of MVA85A [14]. Despite great promise, subsequent clinical trials showed that it was no more effective than BCG alone [15]. The candidate vaccine M72 has shown more success, offering 54% protection against pulmonary TB [16]; however, the need for an even more protective vaccine persists.

In light of the observation that MTB modifies the expression of many genes in the hypoxic and nutrient-starved conditions of the granuloma, more recent efforts have employed a subunit vaccine strategy that includes antigens relevant to both TB virulence and latency [17]. It has been hypothesized that such a vaccine can prevent both primary and post-primary TB [17]. Examples of such multistage vaccines include H56, which in early clinical trials has demonstrated better containment of late-stage TB than BCG alone, and ID93, which has been shown in animal models to protect against TB and in Phase I trials has elicited humoral and cell-mediated responses in humans [17–20]. The protective efficacy of these vaccines remains to be seen in later-stage clinical trials.

Despite the encouraging progress made with these recent candidate vaccines, it is unlikely that a single vaccine will be protective against all forms of TB in all age groups, and so diversification of candidate vaccines is important [21]. Thus, a broader search for MTB protective antigens (PAgs) is anticipated to provide a variety of new candidates and inform the design of new and effective multistage TB vaccines. While previous studies have applied bioinformatic strategies to the selection of PAgs for vaccine candidates, machine learning (ML) has not been previously used for this task. Additionally, reverse vaccinology (RV) allows for the identification of PAgs that may otherwise be impossible to identify or isolate using conventional methods [22]. Furthermore, while traditional RV methods primarily consider surface-exposed proteins, ML-based RV methods allow for the identification of non-surface-exposed proteins. This may be especially important for the development of vaccines against intracellular pathogens, such as MTB, because these non-surface-exposed proteins may induce cell-mediated immunity, which is critical to the control and clearance of intracellular infection. To contribute to the growing knowledge of PAgs in MTB, we conducted the present study employing a recently developed ML-based RV model, Vaxign-ML [23].

2. Materials and Methods

Vaxign-ML was applied to the MTB H37Rv proteome (Uniprot Proteome UP000001584) to compute the protegenicity score of each protein [24]. This score predicts the likelihood that a given protein will be a PAg. As described previously by Ong et al., Vaxign-ML used 397 bacterial PAgs with at least one experimental evidence of protection (e.g., in an animal challenge assay) to train an extreme gradient boosting model [23]. With a recommended protegenicity score threshold, Vaxign-ML achieved the highest performance with 0.96 weighted F1-score in a nested five-fold cross-validation and outperformed other existing web-based RV tools [23,25].

We then selected proteins with the previously recommended Vaxign-ML protegenicity score threshold [23]. Of these proteins, those that had previously been established as PAgs according to Protegen were excluded from selection so that all selected prospective PAgs were novel. Protegen is a web-based database that compiles PAgs of several pathogens, including MTB, which are curated from peer-reviewed articles [26]. The remaining proteins were designated as novel Vaxign-ML-predicted PAgs. This list of predicted PAgs was further refined using two independent criteria.

The first criterion considered whether each novel Vaxign-ML-predicted PAg belonged to the family of an established MTB PAg. The rationale for this selection criterion was that novel Vaxign-ML-predicted PAgs that belong to the protein family of an established PAg were likely to be similar in structure, function, and amino acid sequence to the established PAgs and thus were more likely to be PAgs themselves.

To do this, we compiled a list of the protein families of the established MTB PAgs using Protegen and the UniProt Knowledgebase, which is a manually annotated database of protein sequence data combined with summaries of experimentally verified or computationally predicted functional information about each protein [27]. The protein family for each novel Vaxign-ML-predicted PAg for which it was available was also identified using the UniProt Knowledgebase. Novel Vaxign-ML-predicted PAgs that belonged to the protein family of an established PAg were selected.

The second criterion considered whether each novel Vaxign-ML-predicted PAg was involved in biological processes related to either MTB virulence or LTBI. The rationale for the selection of proteins involved in MTB virulence was that PAgs are likely to come from virulence factors. The rationale for the selection of latency-related proteins was their importance in inducing a cell-mediated immune response in the latency of TB. As it has been hypothesized, we were operating under the assumption that a more effective vaccine preventing activation of LTBI could be developed by combining PAgs involved in virulence with those expressed in the latent stage of the disease [17].

The Gene Ontology (GO) biological processes of each novel Vaxign-ML-predicted PAg for which they were available were gathered via the UniProt Knowledgebase [27]. Gene Ontology is a project to systematically categorize the function of proteins in terms of molecular function, cellular component, and biological process. We then selected by literature review eleven categories of GO biological processes that were relevant to the unique pathophysiology of MTB in either latency or virulence, including cell envelope biogenesis and maintenance [28,29], DNA repair [30], interaction with host immune system [8], fatty acid beta-oxidation [31], growth in host [32], protein folding [33], response to antibiotic [34], response to acidic pH [35], response to hypoxia [36], response to nitrosative or oxidative stress [32], and response to starvation [37]. Novel Vaxign-ML-predicted PAgs whose biological processes belonged to at least one of these categories were selected.

Upon first selecting the proteins with sufficiently high Vaxign-ML protegenicity scores and then refining this list using the two selection criteria mentioned above, we finally selected the proteins with the greatest number of promiscuous MHC-I and MHC-II epitopes using T-cell epitope prediction. The rationale for this final selection was the necessity that vaccine candidates provide broad population coverage. The presence of promiscuous epitopes is important in the prediction of T-cell epitope candidates because of the highly polymorphic nature of HLA alleles [38].

For MHC-I binding prediction, the IEDB-recommended NetMHCpan-4.1 prediction method was used. This method employs an ML strategy trained on both binding affinity and mass spectrometry-eluted ligands [39,40]. Binding predictions were made for 9-mer epitopes with a reference set of 27 frequently occurring HLA alleles which together cover >97% of the global population [41]. The selection threshold for MHC-I binding prediction was percentile rankings less than 1%, as has previously been recommended [42,43].

For MHC-II binding prediction, the IEDB-recommended prediction method was used [40]. This method employs the consensus approach, which combines the NN-align, SMM-align, CombLib, and Sturniolo methods if a corresponding predictor is available for the given molecule, and if not, NetMHCIIpan is used [38,39,44–47]. Binding predictions were made for 15-mer epitopes with a reference set of 27 frequently occurring HLA alleles, which together cover >99% of the global population [48]. The selection threshold for MHC-II binding prediction was adjusted percentile rankings less than 10%, as has previously been recommended [49].

For both MHC-I and MHC-II binding prediction, epitopes that were predicted to bind with high affinity (i.e., binding percentile rank meeting the threshold) to at least four different HLA alleles within the reference set were considered promiscuous epitopes.

The novel Vaxign-ML-predicted PAgs having been selected by the two criteria mentioned above were ranked by the number of promiscuous epitopes they contained for both MHC-I and MHC-II separately. The top-ranked 11 proteins in each ranking were selected.

3. Results

3.1. Novel Protective Antigens Predicted by Vaxign-ML

Vaxign-ML was used to predict the protegenicity of the entire H37Rv proteome. This step yielded 238 proteins with protegenicity scores meeting the recommended threshold. Of these 238 proteins, 24 had previously been established as PAgs according to Protegen and thus were excluded from further selection. The remaining 214 novel Vaxign-ML-predicted PAgs (Figure 1, Table S1) were subjected to further selection as described in the following sections.

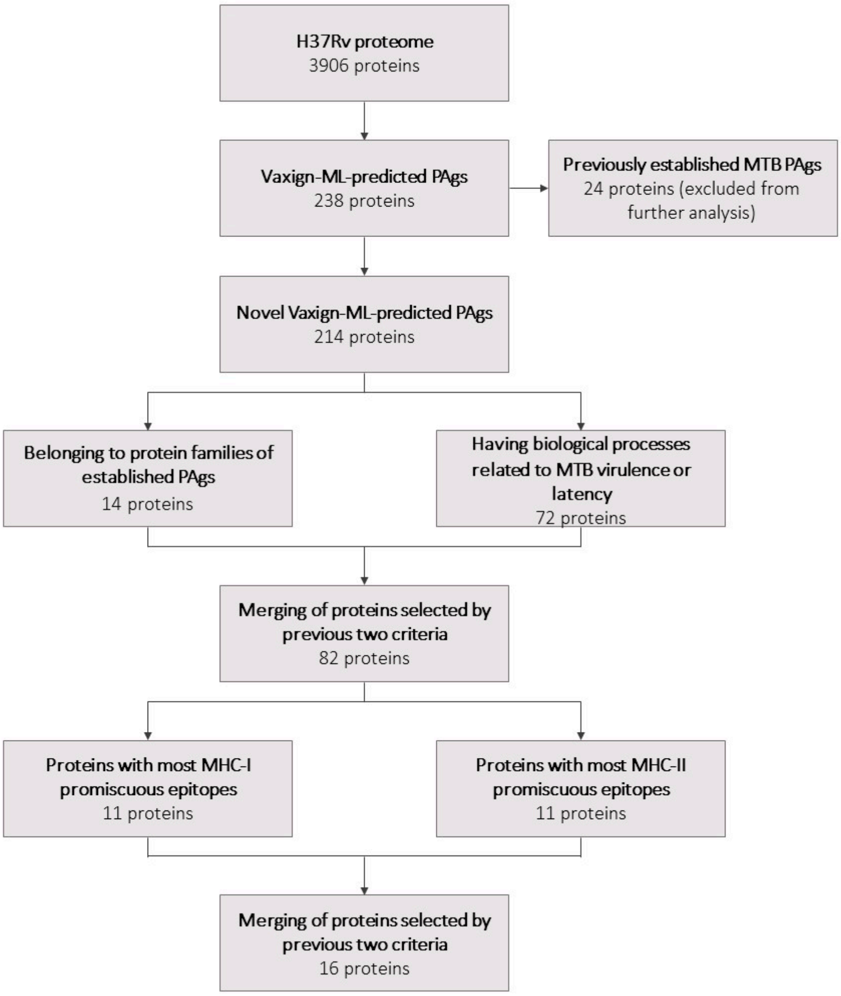
H37Rv proteome
3906 proteins
Vaxign-ML-predicted PAgs
238 proteins
Previously established MTB PAgs
24 proteins (excluded from further analysis)
Novel Vaxign-ML-predicted PAgs
214 proteins
Belonging to protein families of established PAgs
14 proteins
Having biological processes related to MTB virulence or latency
72 proteins
Merging of proteins selected by previous two criteria
82 proteins
Proteins with most MHC-I promiscuous epitopes
11 proteins
Proteins with most MHC-II promiscuous epitopes
11 proteins
Merging of proteins selected by previous two criteria
16 proteins

Figure 1. A flowchart summarizing the selection of potential vaccine candidates and the number of proteins selected at each step.

3.2. Antigens Belonging to Protein Families of Previously Established MTB PAgs

A list of all previously established PAgs was gathered using Protegen (Table S2). The protein family of each for which it was available was determined using UniProt. These established PAgs belonged to 11 unique protein families (Table S2). The protein family of each novel Vaxign-ML-predicted PAg for which it was available was also determined using UniProt. The 214 novel Vaxign-ML-predicted PAgs belonged to 116 unique protein families (Table S1). There were 14 novel Vaxign-ML-predicted PAgs that together belonged to seven unique protein families of established PAgs (Figure 1; Table 1).

Table 1. Novel Vaxign-ML-predicted PAgs that belong to protein families of established MTB PAgs.

Protein Family	Protein	Tuberculist ID
mycobacterial A85 antigen	FbpC (Ag85C)	Rv0129c
mycobacterial PE	PE_PGRS11 PE4 PE26	Rv0754 Rv0160c Rv2519
mycobacterial PPE	Hypothetical protein Rv3822 PPE28 PPE8 PPE12 PPE30	Rv3822 Rv1800 Rv0355c Rv0755c Rv1802
peptidase S1C	HtrA PepD	Rv1223 Rv0983
PstS	PstS2	Rv0932c
RsiV	Hypothetical protein Rv3036c	Rv3036c
WXG100	EsxI	Rv1037c

3.3. Antigens Having Biological Processes Associated with MTB Virulence or LTBI

The GO biological processes of each novel Vaxign-ML-predicted PAg for which they were available were identified using UniProt. These 214 proteins were involved in 226 unique biological processes, of which the most common were pathogenesis and growth of symbiont in host (Figure 2).

In summary, there were 72 novel Vaxign-ML-predicted PAgs that had GO biological processes that belonged to one or more of these biological process categories (Figure 1, Table 2). The 72 proteins selected here had biological processes that most commonly belonged to the growth in host, interaction with the host immune system, and cell envelope biogenesis and maintenance categories (Figure 3).

3.4. Antigens with the Greatest Number of Promiscuous MHC-I and MHC-II Epitopes

Fourteen proteins were selected by the first criterion on the basis of belonging to a protein family of an established PAg. Seventy-two proteins were selected by the second criterion on the basis of having a biological process related to the virulence or latency of MTB. Upon merging the proteins selected by these criteria, 82 unique proteins remained.

The binding affinity of each of these 82 proteins to the reference sets of MHC-I and MHC-II alleles was predicted using IEDB binding prediction tools. The 82 selected proteins were ranked by the number of promiscuous epitopes for both MHC-I and MHC-II (Tables S3 and S4, respectively). To further refine our selection, we took the 11 proteins with the most promiscuous epitopes for both the MHC-I and MHC-II reference set of alleles, yielding 22 proteins. Because six of these proteins were found in both the MHC-I and MHC-II rankings, a total of 16 unique proteins remained in this final selection (Figure 1; Table 3).

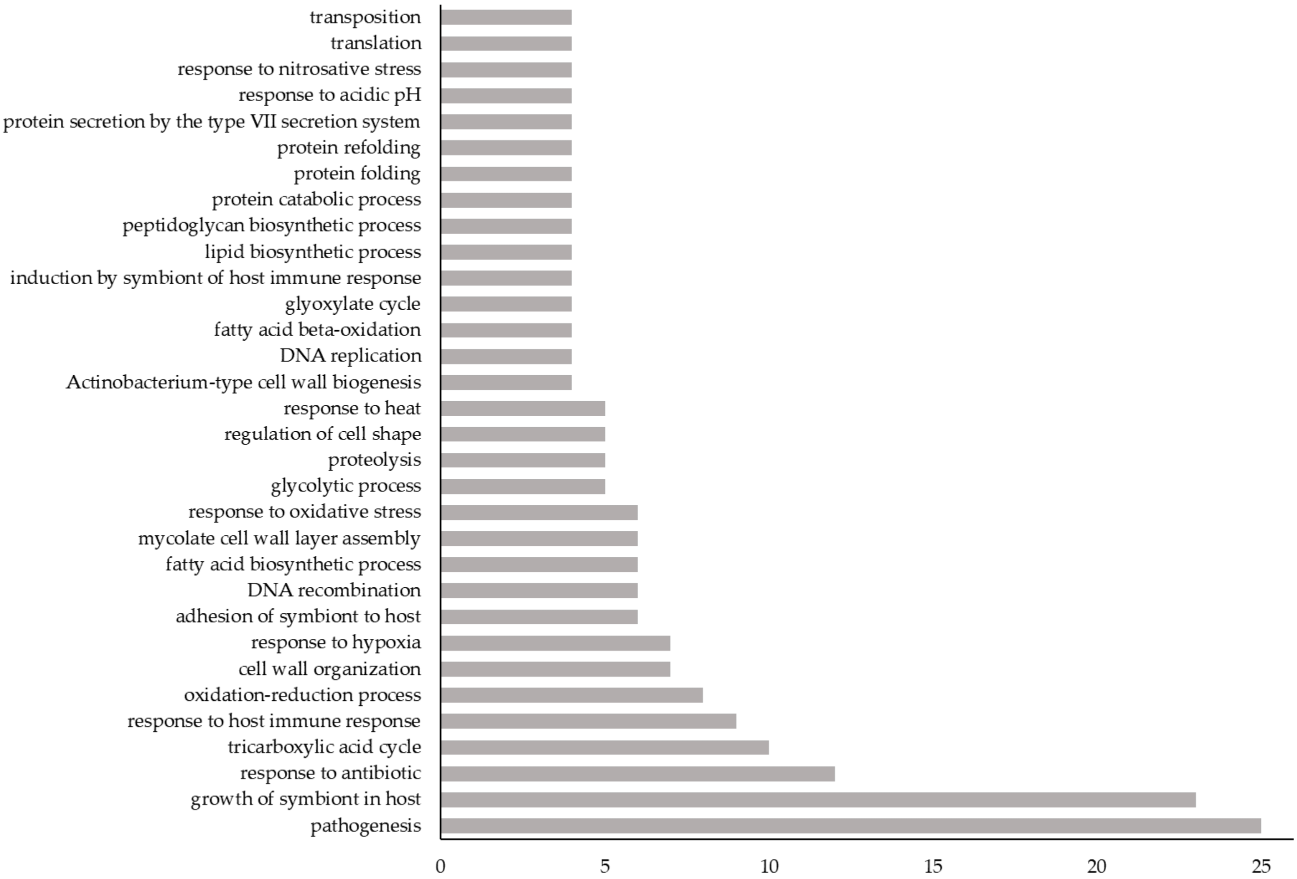

Figure 2. The frequency of GO biological processes among the 214 novel Vaxign-ML-predicted PAgs. *X*-axis: number of novel Vaxign-ML-predicted PAgs having the given GO biological process; *Y*-axis: GO biological processes to which at least 4 novel Vaxign-ML-predicted PAgs belonged. For a complete list of GO biological processes belonging to these 214 proteins, see Table S1.

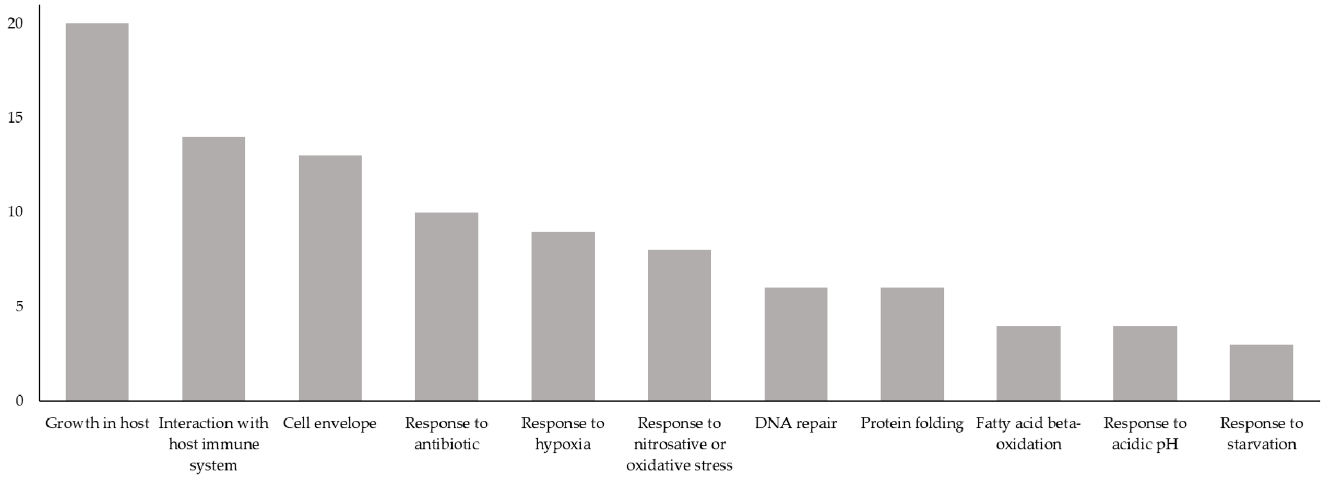

Figure 3. Biological process categories of the 72 proteins having biological processes related to the virulence or latency of MTB. *X*-axis: biological process categories relevant to MTB virulence or latency; *Y*-axis: number of novel Vaxign-ML-predicted PAgs having a GO biological process belonging to a given biological process category. Total exceeds 72 because several novel Vaxign-ML-predicted PAgs have GO biological processes belonging to more than one biological process category.

Table 2. Novel Vaxign-ML-predicted PAgs having biological processes related to MTB virulence or latency. More than 72 proteins are listed here due to redundancy between categories.

Biological Process Category	Proteins (Tuberculist IDs)
Cell envelope biogenesis and maintenance	PonA2 (Rv3682), PonA1 (Rv0050), FadD15 (Rv2187), LdtB (Rv2518c), PbpB (Rv2163c), FadD30 (Rv0404), FbpC (Ag85C) (Rv0129c), AccD4 (Rv3799c), FadD32 (Rv3801c), hypothetical protein Rv3811 (Rv3811), LprQ (Rv0483), PbpA (Rv0016c), FadD19 (Rv3515c)
DNA Repair	RecA (Rv2737c), HtpG (Rv2299c), UvrA (Rv1638), LigD (Rv0938), RecG (Rv2973c), UvrB (Rv1633)
Interaction with host immune system	FadE5 (Rv0244c), Mce1A (Rv0169), probable aldehyde dehydrogenase (Rv0458), CaeA (Rv2224c), LprA (Rv1270c), FadD30 (Rv0404), EccCa1 (Rv3870), Icl1 (Rv0467), PknH (Rv1266c), MmpL12 (Rv1522c), UvrB (Rv1633), EccB1 (Rv3869), halimadienyl diphosphate synthase (Rv3377c), FadD19 (Rv3515c)
Fatty acid beta-oxidation	FadB (Rv0860), FadA3 (Rv1074c), Ltp1 (Rv2790c), probable nonspecific lipid-transfer protein (Rv1627c)
Growth in host	FadD13 (Rv3089), Mce2C (Rv0591), Mce4A (Rv3499c), Mce1A (Rv0169), Mce1F (Rv0174), Tgs4 (Rv3088), Mce3C (Rv1968), Mce1C (Rv0171), Mce3D (Rv1969), Mce3A (Rv1966), EccCa (Rv3870), Mce4C (Rv3497c), Mce2F (Rv0594), Mce4D (Rv3496c), Mce2A (Rv0589), EccA1 (Rv3868), Mce2D (Rv0592), Mce1D (Rv0172), FadA (Rv0243), Mce4F (Rv3494c)
Protein folding	GroEL2 (Rv0440), Mpa (Rv2115c), GroEL1 (Rv3417c), ClpX (Rv2457c), HtpG (Rv2299c), ClpB (Rv0384c)
Response to antibiotic	PonA2 (Rv3682), GyrB (Rv0005), RecA (Rv2737c), RpoB (Rv0667), PonA1 (Rv0050), FbpC (Ag85C) (Rv0129c), possible penicillin-binding lipoprotein (Rv2864c), PepD (Rv0983), IleS (Rv1536), LprG (Rv1411c)
Response to acidic pH	FadD13 (Rv3089), Tgs4 (Rv3088), Icl1 (Rv0467), AccD4 (Rv3799c)
Response to hypoxia	GroEL2 (Rv0440), PonA1 (Rv0050), Tgs4 (Rv3088), Icl1 (Rv0467), AccD4 (Rv3799c), PE_PGRS11 (Rv0754), Tuf (Rv0685), SdhA (Rv3318), probable succinate dehydrogenase (Rv0248c)
Response to nitrosative or oxidative stress	Mpa (Rv2115c), FtsH (Rv3610c), HtpG (Rv2299c), Tgs4 (Rv3088), Mpt53 (Rv2878c), AccD4 (Rv3799c), UvrB (Rv1633), CysN (Rv1286)
Response to starvation	halimadienyl diphosphate synthase (Rv3377c), PknD (Rv0931c), CysN (Rv1286)

Table 3. Final selection of prospective novel MTB vaccine candidates with sufficiently high Vaxign-ML protegenicity scores, belonging to the protein family of an established PAg and/or having a biological process related to the virulence or latency of MTB, and having the greatest number of MHC-I and/or MHC-II promiscuous epitopes.

Protein	Tuberculist ID	GO Biological Process	Subcellular Location	MHC-I Promiscuous Epitopes	MHC-II Promiscuous Epitopes
Having highest numbers of both MHC-I and MHC-II promiscuous epitopes					
PPE8	Rv0355c	not available	not available	104	194
IleS	Rv1536	isoleucyl-tRNA aminoacylation, response to antibiotics	cytoplasm	92	156
MmpL12	Rv1522c	response to host immune response	cell membrane, multi-pass membrane protein	86	263
UvrA	Rv1638	cellular response to DNA damage stimulus, negative regulation of strand invasion, nucleotide-excision repair, SOS response	cytoplasm	73	116
RpoB	Rv0667	response to antibiotic, DNA-templated transcription	cell wall, cytosol, plasma membrane	72	109
ClpB	Rv0384c	protein refolding, response to heat	cytoplasm	62	104

Table 3. *Cont.*

Protein	Tuberculist ID	GO Biological Process	Subcellular Location	MHC-I Promiscuous Epitopes	MHC-II Promiscuous Epitopes
		Having highest number of MHC-I promiscuous epitopes only			
PonA2	Rv3682	peptidoglycan biosynthetic process, response to antibiotic	not available	60	88
FadE5	Rv0244c	response to host immune response	extracellular region, plasma membrane	57	86
Mce2D	Rv0592	growth of symbiont in host, growth of symbiont in host vacuole	cell wall	57	84
FadD30	Rv0404	Actinobacterium-type cell wall biogenesis, fatty acid biosynthetic process, induction by symbiont of host immune response, lipid biosynthetic process	not available	56	85
EccCa1	Rv3870	evasion of host immune response, growth of symbiont in host, pathogenesis, protein secretion by the type VII secretion system	cell inner membrane, multi-pass membrane protein	56	79
		Having highest number of MHC-II promiscuous epitopes only			
PPE28	Rv1800	not available	not available	54	176
PE4	Rv0160c	not available	not available	48	100
FadD15	Rv2187	Actinobacterium-type cell wall biogenesis, fatty acid biosynthetic process, lipid biosynthetic process, long-chain fatty acid metabolic process	cell wall, plasma membrane	48	113
PknD	Rv0931c	cellular response to phosphate starvation, negative regulation of catalytic activity, negative regulation of fatty acid biosynthetic process, negative regulation of protein binding, pathogenesis, positive regulation of catalytic activity	cell membrane, single-pass membrane protein	42	105
Mce2A	Rv0589	growth of symbiont in host, growth of symbiont in host vacuole	integral component of membrane	36	100

4. Discussion

This study aimed to identify novel MTB PAgs to assist in the creation of a multistage TB vaccine strategy that will overcome the inadequacy of the BCG vaccine and confer broad immunity against both primary and post-primary TB in all populations. Upon filtering the H37Rv genome through Vaxign-ML, we selected 82 novel predicted PAgs that either belonged to protein families of previously established PAgs or were relevant to the virulence or latency of MTB. From this group, we then identified 16 predicted PAgs with the broadest immunogenic potential as indicated by the number of promiscuous MHC-I and MHC-II epitopes.

Although none of the identified 16 prospective MTB PAgs have previously been studied as vaccine candidates, several have been characterized to various degrees [17,50–53]. PE4 (Rv0160c) has previously been shown to exhibit elevated expression during cellular stress, including in the persistent stage of the MTB infection and has been described as an immunodominant antigen that elicits a strong humoral response in patients [50]. ClpB (Rv0384c) has been demonstrated to be required for the persistence of MTB bacilli within macrophages and under the stressful conditions of latency [51]. EccCa1 (Rv3870) has been shown to be essential for the secretion of ESAT-6 and CFP-10, proteins involved in MTB pathogenesis [52]. FadE5 (Rv0244c) was shown to be expressed to a similar degree in early- and late-stage TB and to be involved in the stress response [17]. Finally, PknD (Rv0931c) is believed to be essential for MTB infections of the central nervous system [53]. As none of these 16 proteins have been studied as vaccine candidates and many have not been previously characterized at all, they may be good candidates for future laboratory-based studies.

Of the 11 biological process categories that were identified as relevant to MTB virulence or LTBI, seven were represented among the 16 selected proteins, including interaction with the host immune system, cell envelope biogenesis and maintenance, growth in host, response to antibiotic, DNA repair, protein folding, and response to starvation (Table 3). Four of the 16 selected proteins, MmpL12 (Rv1522c), FadE5 (Rv0244c), FadD30 (Rv0404), and EccCa1 (Rv3870), are involved in interaction with the host immune system; these processes allow MTB to evade, modulate, or suppress a host's immune system via mechanisms such as attenuation of macrophage antigen presentation to T-helper cells or downregulation of MHC-II gene expression, which can enable continued latency [8]. Three of the 16, PonA2 (Rv3682), FadD30 (Rv0404), and FadD15 (Rv2187), are involved in cell envelope biogenesis and maintenance; these processes are responsible for the unique cell envelope of MTB, which is critical to its slow intracellular growth, virulence, and innate impermeability to many drugs and antibiotics [28,29]. Three of the 16, Mce2D (Rv0592), EccCa1 (Rv3870), and Mce2A (Rv0589), are involved in growth in the host, a process self-evidently relevant to MTB's virulence; many of the proteins involved in this process are considered virulence factors [32]. Three of the 16, IleS (Rv1536), RpoB (Rv0667), and PonA2 (Rv3682), are involved in the response to antibiotics; though proteins involved in antibiotic resistance are not themselves virulence factors, they can be instrumental in the persistence of disease [34]. One of the 16, UvrA (Rv1638), is involved in DNA repair, which is critical for the persistence of MTB in the hostile, oxidative environment of the macrophage [30]. One of the 16, ClpB (Rv0384c), is involved in protein folding; proteins that ensure correct protein folding, such as chaperonins, are also essential to MTB survival under stressful environments such as in the macrophage [33]. Finally, one of the 16, PknD (Rv0931c), is involved in response to starvation; because MTB faces both energy and nutrient starvation within the granuloma, proteins that enable persistence of the bacilli in spite of these low-nutrient conditions are also necessary for latency [37].

Among the 16 proteins selected on the basis of promiscuous epitopes, six are among those having the most promiscuous epitopes for both MHC-I and MHC-II alleles (Table 3). These six may provide the greatest immunogenic potential within this set of 16 proteins. However, it is worth noting that we exclusively consider the proteome of the H37Rv reference strain of MTB in this study. Taking into account the genetic diversity of MTB globally, it is important that proteins selected as vaccine candidates are highly conserved. Future studies of the genetic diversity of selected antigens using MTB clinical strains representing different genetic lineages are, therefore, warranted.

A major limitation of our study is that only the peptide sequence of each antigen was considered in our analyses. Our prediction did not consider higher-level protein structures or vaccine formulation, both of which may affect the interaction between the antigen and the host immune system, thereby impacting the induced immune response. A second major limitation is the absence of experimental validation of the protective antigenicity of the identified proteins with vaccine potential.

In light of these limitations, it is important to note that RV methods, such as Vaxign-ML, are not intended to replace laboratory-based immunological studies, but rather they are to serve in a complementary fashion. For example, in a lab setting it is not feasible to conduct a genome-wide search for protective antigens; RV methods, however, can quickly narrow the list of potential protective antigens, thereby informing the prioritization of targets for laboratory investigation. This narrowed list can then be tested and validated in a lab setting. Additionally, Vaxign-ML has previously been validated by demonstrating that all of the PAgs included in five recent MTB vaccines in clinical trials received Vaxign-ML protegenicity scores that met the threshold used in this study [23]. Similarly, it is worth noting that of the 25 proteins that have been established as MTB PAgs according to Protegen, 24 (recall = 0.96) of these received Vaxign-ML protegenicity scores that met the threshold for selection within this study.

To summarize, the bioinformatic approach applied in this study has allowed for the identification of 16 prospective novel MTB PAgs that may have been difficult to identify using traditional vaccinology techniques and that may be used in future subunit vaccines. For the selected antigens, all computational measures of immunogenicity and epitope promiscuity should, however, be further validated in vitro and in vivo. This integration of traditional and computational vaccine development tools may be the best approach in developing a broadly effective TB vaccine strategy that prevents both primary and post-primary TB in all populations and takes us closer to our ambition to End TB.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/vaccines9101098/s1, Table S1: Novel Vaxign-ML-predicted protective antigens, Table S2: Previously established MTB protective antigens as according to Protegen, Table S3: Proteins selected on the basis of belonging to the protein family of a previously established protective antigen or on the basis of having a GO biological process related to the virulence or latency of MTB, ranked by number of promiscuous MHC-I epitopes, Table S4: Proteins selected on the basis of belonging to the protein family of a previously established protective antigen or on the basis of having a GO biological process related to the virulence or latency of MTB ranked, by number of promiscuous MHC-II epitopes.

Author Contributions: Conceptualization, Z.Y., E.O. and B.T.; methodology, E.O. and B.T.; validation, B.T., E.O. and Z.Y.; formal analysis, B.T.; investigation, B.T.; resources, E.O. and Z.Y.; data curation, B.T.; writing—original draft preparation, B.T.; writing—review and editing, Z.Y., E.O. and B.T.; visualization, B.T.; supervision, Z.Y.; project administration, Z.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All the data generated by the present study are provided either in the main body of the article or in its supplementary material.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. *Global Tuberculosis Report 2020*; World Health Organization, 2020. Available online: https://apps.who.int/iris/bitstream/handle/10665/336069/9789240013131-eng.pdf (accessed on 1 February 2021).
2. Daniel, T.M.; Bates, J.H.; Downes, K.A. History of Tuberculosis. In *Tuberculosis*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 1994; pp. 13–24. [CrossRef]
3. Seung, K.J.; Keshavjee, S.; Rich, M.L. Multidrug-Resistant Tuberculosis and Extensively Drug-Resistant Tuberculosis. *Cold Spring Harb. Perspect. Med.* **2015**, *5*, a017863. [CrossRef] [PubMed]
4. Bell, L.C.K.; Noursadeghi, M. Pathogenesis of HIV-1 and Mycobacterium tuberculosis co-infection. *Nat. Rev. Genet.* **2017**, *16*, 80–90. [CrossRef] [PubMed]
5. Uplekar, M.; Weil, D.; Lonnroth, K.; Jaramillo, E.; Lienhardt, C.; Dias, H.M.; Falzon, D.; Floyd, K.; Gargioni, G.; Getahun, H.; et al. WHO's new End TB Strategy. *Lancet* **2015**, *385*, 1799–1801. [CrossRef]
6. The End TB Strategy. World Health Organization, 2015. Available online: https://www.who.int/tb/End_TB_brochure.pdf (accessed on 7 February 2021).
7. *The Sustainable Development Goals Report 2020*; United Nations, 2020. Available online: https://unstats.un.org/sdgs/report/2020/The-Sustainable-Development-Goals-Report-2020.pdf (accessed on 9 February 2021).
8. Hmama, Z.; Peña-Díaz, S.; Joseph, S.; Av-Gay, Y. Immunoevasion and immunosuppression of the macrophage byMycobacterium tuberculosis. *Immunol. Rev.* **2015**, *264*, 220–232. [CrossRef]
9. Ramakrishnan, L. Revisiting the role of the granuloma in tuberculosis. *Nat. Rev. Immunol.* **2012**, *12*, 352–366. [CrossRef]
10. Cadena, A.M.; Fortune, S.M.; Flynn, J.L. Heterogeneity in tuberculosis. *Nat. Rev. Immunol.* **2017**, *17*, 691–702. [CrossRef]
11. Flynn, J.L.; Chan, J. Tuberculosis: Latency and Reactivation. *Infect. Immun.* **2001**, *69*, 4195–4201. [CrossRef]
12. Andersen, P.; Scriba, T.J. Moving tuberculosis vaccines from theory to practice. *Nat. Rev. Immunol.* **2019**, *19*, 550–562. [CrossRef]
13. Ong, E.; He, Y.; Yang, Z. Epitope promiscuity and population coverage of Mycobacterium tuberculosis protein antigens in current subunit vaccines under development. *Infect. Genet. Evol.* **2020**, *80*, 104186. [CrossRef]

14. McShane, H.; Pathan, A.A.; Sander, C.R.; Goonetilleke, N.P.; Fletcher, H.A.; Hill, A.V. Boosting BCG with MVA85A: The first candidate subunit vaccine for tuberculosis in clinical trials. *Tuberculosis* **2005**, *85*, 47–52. [CrossRef]
15. Tameris, M.D.; Hatherill, M.; Landry, B.S.; Scriba, T.; Snowden, M.A.; Lockhart, S.; Shea, J.E.; McClain, J.B.; Hussey, G.D.; Hanekom, W.A.; et al. Safety and efficacy of MVA85A, a new tuberculosis vaccine, in infants previously vaccinated with BCG: A randomised, placebo-controlled phase 2b trial. *Lancet* **2013**, *381*, 1021–1028. [CrossRef]
16. Van Der Meeren, O.; Hatherill, M.; Nduba, V.; Wilkinson, R.; Muyoyeta, M.; Van Brakel, E.; Ayles, H.M.; Henostroza, G.; Thienemann, F.; Scriba, T.; et al. Phase 2b Controlled Trial of M72/AS01EVaccine to Prevent Tuberculosis. *N. Engl. J. Med.* **2018**, *379*, 1621–1634. [CrossRef]
17. Aagaard, C.; Hoang, T.; Dietrich, J.; Cardona, P.-J.; Izzo, A.; Dolganov, G.; Schoolnik, G.K.; Cassidy, J.P.; Billeskov, R.; Andersen, P. A multistage tuberculosis vaccine that confers efficient protection before and after exposure. *Nat. Med.* **2011**, *17*, 189–194. [CrossRef]
18. Luabeya, A.K.K.; Kagina, B.; Tameris, M.D.; Geldenhuys, H.; Hoff, S.T.; Shi, Z.; Kromann, I.; Hatherill, M.; Mahomed, H.; Hanekom, W.A.; et al. First-in-human trial of the post-exposure tuberculosis vaccine H56:IC31 in Mycobacterium tuberculosis infected and non-infected healthy adults. *Vaccine* **2015**, *33*, 4130–4140. [CrossRef]
19. Bertholet, S.; Ireton, G.C.; Ordway, D.; Windish, H.P.; Pine, S.O.; Kahn, M.; Phan, T.; Orme, I.M.; Vedvick, T.S.; Baldwin, S.L.; et al. A Defined Tuberculosis Vaccine Candidate Boosts BCG and Protects against Multidrug-Resistant Mycobacterium tuberculosis. *Sci. Transl. Med.* **2010**, *2*, 53ra74. [CrossRef]
20. Coler, R.N.; Day, T.A.; Ellis, R.; Piazza, F.M.; Beckmann, A.M.; Vergara, J.; Rolf, T.; Lu, L.; Alter, G.; Hokey, D.; et al. The TLR-4 agonist adjuvant, GLA-SE, improves magnitude and quality of immune responses elicited by the ID93 tuberculosis vaccine: First-in-human trial. *NPJ Vaccines* **2018**, *3*, 1–9. [CrossRef] [PubMed]
21. Kaufmann, S.H.E.; Dockrell, H.M.; Drager, N.; Ho, M.M.; McShane, H.; Neyrolles, O.; Ottenhoff, T.H.M.; Brij, B.; Roordink, D.; Spertini, F.; et al. TBVAC2020: Advancing Tuberculosis Vaccines from Discovery to Clinical Development. *Front. Immunol.* **2017**, *8*. [CrossRef] [PubMed]
22. Rappuoli, R. Reverse vaccinology, a genome-based approach to vaccine development. *Vaccine* **2001**, *19*, 2688–2691. [CrossRef]
23. Ong, E.; Wang, H.; Wong, M.U.; Seetharaman, M.; Valdez, N.; He, Y. Vaxign-ML: Supervised machine learning reverse vaccinology model for improved prediction of bacterial protective antigens. *Bioinformatics* **2020**, *36*, 3185–3191. [CrossRef] [PubMed]
24. Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Eiglmeier, K.; Gas, S.; Barry, C.E.; et al. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature* **1998**, *393*, 537–544. [CrossRef]
25. Ong, E.; Cooke, M.F.; Huffman, A.; Xiang, Z.; Wong, M.U.; Wang, H.; Seetharaman, M.; Valdez, N.; He, Y. Vaxign2: The second generation of the first Web-based vaccine design program using reverse vaccinology and machine learning. *Nucleic Acids Res.* **2021**, *49*, W671–W678. [CrossRef]
26. Yang, B.; Sayers, S.; Xiang, Z.; He, Y. Protegen: A web-based protective antigen database and analysis system. *Nucleic Acids Res.* **2010**, *39*, D1073–D1078. [CrossRef]
27. The UniProt Consortium; Bateman, A.; Martin, M.-J.; Orchard, S.; Magrane, M.; Agivetova, R.; Ahmad, S.; Alpi, E.; Bowler-Barnett, E.H.; Britto, R.; et al. UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **2020**, *49*, D480–D489. [CrossRef]
28. Jarlier, V.; Nikaido, H. Mycobacterial cell wall: Structure and role in natural resistance to antibiotics. *FEMS Microbiol. Lett.* **1994**, *123*, 11–18. [CrossRef]
29. Brennan, P.J.; Nikaido, H. The Envelope of Mycobacteria. *Annu. Rev. Biochem.* **1995**, *64*, 29–63. [CrossRef]
30. Dos Vultos, T.; Mestre, O.; Tonjum, T.; Gicquel, B. DNA repair inMycobacterium tuberculosisrevisited. *FEMS Microbiol. Rev.* **2009**, *33*, 471–487. [CrossRef]
31. Bishai, W.R. Lipid lunch for persistent pathogen. *Nature* **2000**, *406*, 683–684. [CrossRef]
32. Forrellad, M.A.; Klepp, L.I.; Gioffré, A.; García, J.S.Y.; Morbidoni, H.R.; Santangelo, M.D.L.P.; Cataldi, A.A.; Bigi, F. Virulence factors of theMycobacterium tuberculosiscomplex. *Virulence* **2013**, *4*, 3–66. [CrossRef] [PubMed]
33. Qamra, R.; Mande, S.C.; Coates, A.R.; Henderson, B. The unusual chaperonins of Mycobacterium tuberculosis. *Tuberculosis* **2005**, *85*, 385–394. [CrossRef] [PubMed]
34. Nguyen, L. Antibiotic resistance mechanisms in M. tuberculosis: An update. *Arch. Toxicol.* **2016**, *90*, 1585–1604. [CrossRef] [PubMed]
35. Huynh, K.K.; Grinstein, S. Regulation of Vacuolar pH and Its Modulation by Some Microbial Species. *Microbiol. Mol. Biol. Rev.* **2007**, *71*, 452–462. [CrossRef] [PubMed]
36. Rustad, T.R.; Sherrid, A.M.; Minch, K.J.; Sherman, D.R. Hypoxia: A window intoMycobacterium tuberculosislatency. *Cell. Microbiol.* **2009**, *11*, 1151–1159. [CrossRef] [PubMed]
37. Betts, J.C.; Lukey, P.T.; Robb, L.C.; McAdam, R.A.; Duncan, K. Evaluation of a nutrient starvation model ofMycobacterium tuberculosispersistence by gene and protein expression profiling. *Mol. Microbiol.* **2002**, *43*, 717–731. [CrossRef] [PubMed]
38. Sturniolo, T.; Bono, E.; Ding, J.; Raddrizzani, L.; Tuereci, O.; Sahin, U.; Braxenthaler, M.; Gallazzi, F.; Protti, M.P.; Sinigaglia, F.; et al. Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. *Nat. Biotechnol.* **1999**, *17*, 555–561. [CrossRef] [PubMed]

39. Reynisson, B.; Alvarez, B.; Paul, S.; Peters, B.; Nielsen, M. NetMHCpan-4.1 and NetMHCIIpan-4.0: Improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res.* **2020**, *48*, W449–W454. [CrossRef] [PubMed]
40. Dhanda, S.K.; Mahajan, S.; Paul, S.; Yan, Z.; Kim, H.; Jespersen, M.C.; Jurtz, V.; Andreatta, M.; Greenbaum, J.A.; Marcatili, P.; et al. IEDB-AR: Immune epitope database—analysis resource in 2019. *Nucleic Acids Res.* **2019**, *47*, W502–W506. [CrossRef]
41. Weiskopf, D.; Angelo, M.A.; de Azeredo, E.L.; Sidney, J.; Greenbaum, J.A.; Fernando, A.N.; Broadwater, A.; Kolla, R.V.; De Silva, A.D.; de Silva, A.M.; et al. Comprehensive analysis of dengue virus-specific responses supports an HLA-linked protective role for CD8+ T cells. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, E2046–E2053. [CrossRef]
42. Moutaftsi, M.; Peters, B.; Pasquetto, V.; Tscharke, D.; Sidney, J.; Bui, H.-H.; Grey, H.M.; Sette, A. A consensus epitope prediction approach identifies the breadth of murine TCD8+-cell responses to vaccinia virus. *Nat. Biotechnol.* **2006**, *24*, 817–819. [CrossRef]
43. Kotturi, M.F.; Peters, B.; Buendia-Laysa, F.; Sidney, J.; Oseroff, C.; Botten, J.; Grey, H.; Buchmeier, M.J.; Sette, A. The CD8 + T-Cell Response to Lymphocytic Choriomeningitis Virus Involves the L Antigen: Uncovering New Tricks for an Old Virus. *J. Virol.* **2007**, *81*, 4928–4940. [CrossRef] [PubMed]
44. Wang, P.; Sidney, J.; Kim, Y.; Sette, A.; Lund, O.; Nielsen, M.; Peters, B. Peptide binding predictions for HLA DR, DP and DQ molecules. *BMC Bioinform.* **2010**, *11*, 568. [CrossRef]
45. Jensen, K.K.; Andreatta, M.; Marcatili, P.; Buus, S.; Greenbaum, J.A.; Yan, Z.; Sette, A.; Peters, B.; Nielsen, M. Improved methods for predicting peptide binding affinity to MHC class II molecules. *Immunology* **2018**, *154*, 394–406. [CrossRef]
46. Nielsen, M.; Lundegaard, C.; Lund, O. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. *BMC Bioinform.* **2007**, *8*, 238. [CrossRef]
47. Sidney, J.; Assarsson, E.; Moore, C.; Ngo, S.; Pinilla, C.; Sette, A.; Peters, B. Quantitative peptide binding motifs for 19 human and mouse MHC class I molecules derived using positional scanning combinatorial peptide libraries. *Immunome Res.* **2008**, *4*, 2. [CrossRef]
48. Greenbaum, J.; Sidney, J.; Chung, J.; Brander, C.; Peters, B.; Sette, A. Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes. *Immunogenetics* **2011**, *63*, 325–335. [CrossRef] [PubMed]
49. Southwood, S.; Sidney, J.; Kondo, A.; Del Guercio, M.F.; Appella, E.; Hoffman, S.; Kubo, R.T.; Chesnut, R.W.; Grey, H.M.; Sette, A. Several common HLA-DR types share largely overlapping peptide binding repertoires. *J. Immunol.* **1998**, *160*.
50. Singh, S.K.; Tripathi, D.K.; Singh, P.K.; Sharma, S.; Srivastava, K.K. Protective and survival efficacies of Rv0160c protein in murine model of Mycobacterium tuberculosis. *Appl. Microbiol. Biotechnol.* **2012**, *97*, 5825–5837. [CrossRef] [PubMed]
51. Tripathi, P.; Singh, L.K.; Kumari, S.; Hakiem, O.R.; Batra, J.K. ClpB is an essential stress regulator of Mycobacterium tuberculosis and endows survival advantage to dormant bacilli. *Int. J. Med Microbiol.* **2020**, *310*, 151402. [CrossRef] [PubMed]
52. Behr, M.A.; Sherman, D.R. Mycobacterial virulence and specialized secretion: Same story, different ending. *Nat. Med.* **2007**, *13*, 286–287. [CrossRef] [PubMed]
53. Be, N.A.; Lamichhane, G.; Grosset, J.; Tyagi, S.; Cheng, Q.; Kim, K.S.; Bishai, W.R.; Jain, S.K. Murine Model to Study the Invasion and Survival of Mycobacterium tuberculosis in the Central Nervous System. *J. Infect. Dis.* **2008**, *198*, 1520–1528. [CrossRef]