Predicted Cellular Immunity Population Coverage Gaps for SARS-CoV-2 Subunit Vaccines and their Augmentation by Compact Joint Sets

Ge Liu[1,2], Brandon Carter[1,2], and David K. Gifford[1,2,3,*]

[1]MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA, USA
[2]MIT Electrical Engineering and Computer Science, Cambridge, MA, USA
[3]MIT Biological Engineering, Cambridge, MA, USA
[*]Gifford@mit.edu

August 4, 2020

Abstract

Subunit vaccines induce immunity to a pathogen by presenting a component of the pathogen and thus inherently limit the representation of pathogen peptides for cellular immunity based memory. We find that SARS-CoV-2 subunit peptides may not be robustly displayed by the Major Histocompatibility Complex (MHC) molecules in certain individuals. We introduce an augmentation strategy for subunit vaccines that adds a small number of peptides to a vaccine to improve the population coverage of pathogen peptide display. We augment a subunit vaccine by selecting additional pathogen peptides to maximize the total number of vaccine peptide hits against the distribution of MHC haplotypes in a population. For each subunit we design independent MHC class I and MHC class II peptide sets for augmentation, and alternatively design a combined set of peptides for MHC class I and class II display. We evaluate the population coverage of 9 different subunits of SARS-CoV-2, including 5 functional domains and 4 full proteins, and augment each of them to fill a predicted coverage gap. We predict that a SARS-CoV-2 receptor binding domain subunit vaccine will have fewer than six peptide-HLA hits with $\leq$ 50 nM binding affinity per individual in 51.31% (class I) and 32.99% (class II) of the population, and with augmentation, the uncovered population is predicted to be reduced to 0.54% (class I) and 1.46% (class II). We find that a joint set of pathogen peptides for MHC class I and class II display is predicted to produce a more compact vaccine design than using independent sets for MHC class I and class II. We provide an open source implementation of our design methods (OptiVax), vaccine evaluation tool (EvalVax), as well as the data used in our design efforts here: https://github.com/gifford-lab/optivax/tree/master/augmentation.

1 Introduction

Subunit vaccines seek to reduce the safety risks of attenuated or inactivated pathogen vaccines by optimizing the portion of a pathogen that is necessary to produce durable immune memory (Moyle and Toth, 2013). Early vaccines that consisted of attenuated or inactivated pathogens provided robust protection to disease, but came with risks associated with the potential reversion of inactivation, harmful components, and the potential for pathogenic infection of attenuated vaccines in immunocompromised individuals. Thus, only including the non-infective components of a pathogen necessary to induce durable immunity is a natural goal. Such subunit vaccines have been enabled by our ability to engineer and express pathogen surface components that retain their three-dimensional structure. The retention of three-dimensional structure is important to induce neutralizing antibodies and a corresponding B cell memory. Neutralizing antibodies bind to a specific epitope on a pathogen and disable it, and choosing such an epitope as a subunit vaccine can productively focus an antibody-based immune response. However, the production of durable immune memory rests in part upon help from T cells, which get their cues from peptides displayed by Human Leukocyte Antigen (HLA) molecules encoded by the Major Histocompatibility Complex (MHC) of genes.

Subunit vaccines for coronaviruses have been proposed that utilize a diverse set of coronavirus proteins and domains. Suggested components include the spike (S) protein, the receptor binding domain (RBD), the S1 domain, the S2 domain, the nucleocapsid (N), the membrane (M), the envelope (E), the N-terminal domain (NTD), and the fusion domain (FD) (Wang et al., 2020; Yu et al., 2020). A universal RBD subunit design that can be adapted to COVID-19, MERS, and SARS has been proposed showing the versatility of a subunit approach (Dai et al., 2020). Here we examine how subsetting a pathogen for vaccination may influence the induction of cellular immunity.

Since a subunit vaccine does not fully represent a pathogen, vaccine excluded pathogen peptides will not be observed during vaccination by an individual's T cells. Subunit vaccine-based simulation of a T cell response is further limited because each

individual's HLA molecules have specific preferences for the peptides they will bind and display. Cytotoxic $CD8^+$ T cells observe peptides displayed by molecules encoded by an individual's classical class I loci (HLA-A, HLA-B, and HLA-C), and helper $CD4^+$ T cells observe peptides displayed by molecules encoded by an individual's classical class II loci (HLA-DR, HLA-DQ, and HLA-DP). Computational methods can be used to predict which peptides will be displayed by a given set of HLA alleles. Furthermore, given a set of peptides and the distribution of HLA alleles present in a population, the number of expected peptide-HLA *hits* that each individual in the population will display can be computed (Liu et al., 2020). A peptide-HLA hit is a predicted binding instance of a peptide to an HLA allele, and thus a given peptide can have multiple hits in a given individual. Since HLA loci exhibit linkage disequilibrium, we use the frequencies of population haplotypes in our coverage computations. Each haplotype describes the joint appearance of HLA alleles. It is important to note that a peptide's display does not guarantee that it will be immunogenic and stimulate a T cell response. However, it has been observed that most displayed peptides are immunogenic in a mouse model, with individual to individual variation in peptide immunogenicity (Croft et al., 2019).

One proposed solution to improve the response of helper T cells to subunit vaccines is including universal $CD4^+$ peptide epitopes that are pathogen independent (Alexander et al., 2000). Universal peptides are selected to bind a large fraction of human HLA-DR alleles, and can be derived from example pathogens (Falugi et al., 2001) or be synthetic sequences (Alexander et al., 2000). While these approaches have proven to be useful in promoting $CD4^+$ T cell responses at vaccination time, $CD4^+$ memory T cells created at vaccination time will not be activated upon pathogen infection because the peptides they recognize will not be present. This is an issue because pathogen specific memory $CD4^+$ T cells are thought to be important for durable immunity (Gasper et al., 2014).

We evaluate the population coverage of proposed subunit vaccines for SARS-CoV-2 based upon the predicted display of their component peptides across different populations of HLA haplotype frequencies. We use a conservative 50 nM threshold for peptide binding to an HLA molecule. Thus, it is possible in practice that the gaps in population coverage will be smaller than our predictions. However, our population coverage calculations assume that every peptide displayed is immunogenic. This is the best approximation we are able to make given that present methods for predicting immunogenicity of a peptide are inaccurate and limited to a few HLA alleles (Calis et al., 2013; Ogishi and Yotsuyanagi, 2019; Riley et al., 2019). Thus, we compensate for our inability to model immunogenicity by predicting the number of expected peptide-HLA hits in each individual. This is based upon our assumption that the probability of inducing cellular immune memory in an individual will in general increase with the number of viral peptides their immune system observes.

We present vaccine design methods that produce either two sets of peptides or a single set of peptides. In the *separate design* method we design MHC class I and MHC class II pathogen peptide sets that are predicted to be displayed to meet population coverage criteria (Figure 1A). In the *joint design* method we design a single set of peptides that are designed to meet both MHC class I and class II population coverage criteria (Figure 1B). The use of separate vs. joint peptide sets may depend upon the vaccine delivery platform utilized.

Using our analysis and design methods we present methods for augmenting subunit vaccines to increase their predicted population coverage and specific augmentation designs for SARS-CoV-2. The augmentation methods we introduce compute the relationship between augmentation peptide count and predicted increased population coverage. This permits a flexible selection of augmentation peptide count based upon their delivery. We provide an analysis of the population coverage gaps for SARS-CoV-2 subunit vaccines, showing the relationship between augmentation peptide count and population coverage for all subunits being considered for vaccines. Our predictions suggest that certain subunit vaccines, such as RBD, will have gaps in population coverage that can be addressed by vaccine augmentation with supplemental peptides. For example, we predict that an RBD subunit vaccine will have fewer than six peptides displayed in 51.31% (class I) and 32.99% (class II) of the population, and with augmentation the uncovered population is predicted to be reduced to 0.54% (class I) and 1.46% (class II).

2 Methods

2.1 SARS-CoV-2 proteome and candidate peptides

The SARS-CoV-2 proteome is comprised of four structural proteins (E, M, N, and S) and open reading frames (ORFs) encoding nonstructural proteins (Srinivasan et al., 2020). We obtained the SARS-CoV-2 viral proteome from GISAID (Elbe and Buckland-Merrett, 2017) sequence entry Wuhan/IPBCAMS-WH-01/2019, the first documented case, as provided by Liu et al. (2020). Nextstrain (Hadfield et al., 2018) was used to identify ORFs and translate the sequence. We use sliding windows to extract all peptides of length 8–10 (MHC class I) and 13–25 (MHC class II) inclusive from the SARS-CoV-2 proteome, resulting in 29,403 peptides for MHC class I and 125,593 peptides for MHC class II. For vaccine augmentation, we use the same filtered candidate peptide set as Liu et al. (2020), in which peptides with mutation rate > 0.001 or non-zero glycosylation probability predicted by NetNGlyc (Gupta et al., 2004) are filtered.

Candidate peptides are scored by their predicted binding affinity (IC50) to MHC molecules. For MHC class I, we use an ensemble that outputs the mean predicted binding affinity of NetMHCpan-4.0 (Jurtz et al., 2017) and MHCflurry 1.6.0 (O'Donnell et al., 2020, 2018). For MHC class II, we use NetMHCIIpan-4.0 (Reynisson et al., 2020b). We conservatively

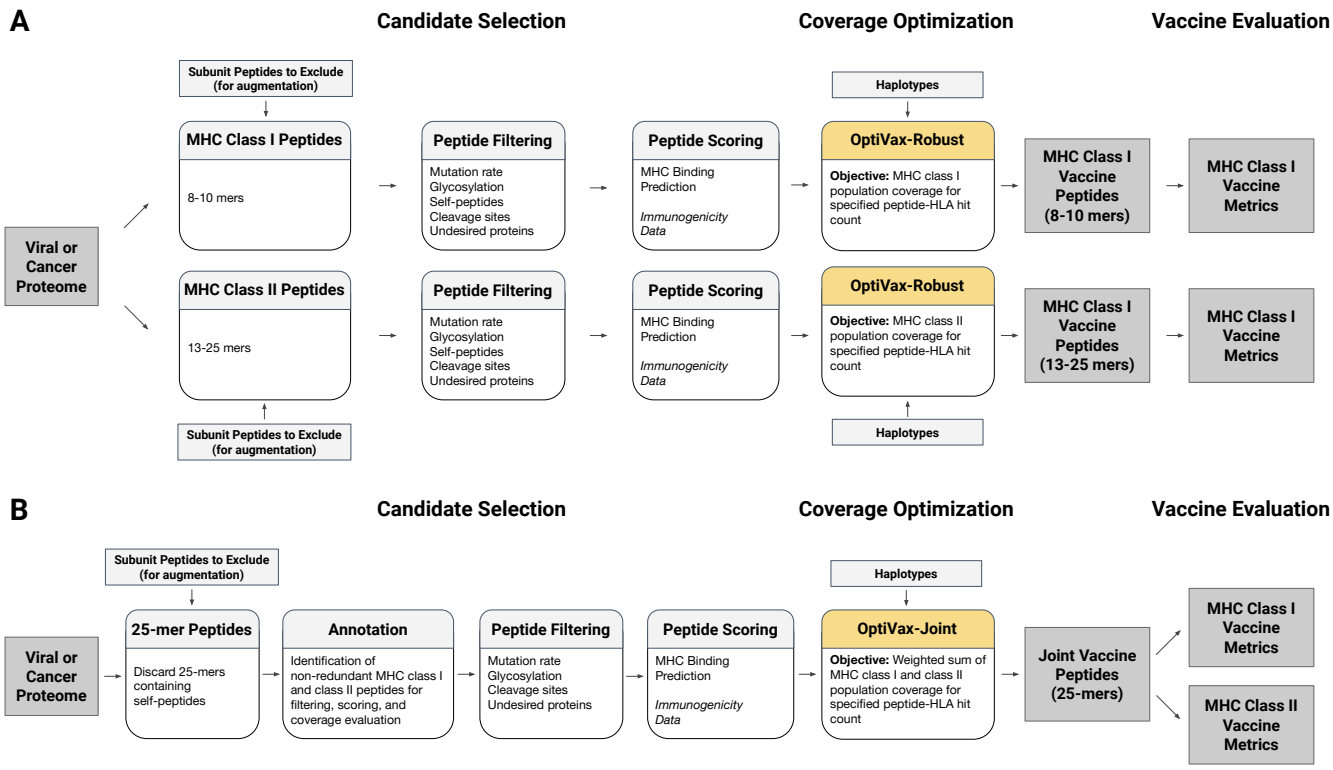

Figure 1. The separate and joint design methods for peptide vaccines. A) In the separate method, windowed pathogen proteomes are filtered for acceptable peptides and MHC class I and class II vaccine designs are chosen to optimize population coverage at specified levels of peptide-HLA hits. B) In the joint method, 25-mer pathogen peptides are annotated with their MHC class I and class II peptides, which are filtered, scored, evaluated for population coverage, and used to optimize the selection of their parent 25-mers into a joint vaccine.

use a 50 nM threshold of predicted binding affinity to classify peptides as binders, which has been shown to improve precision (Liu et al., 2020).

We use HLA class I and class II haplotype frequencies provided by Liu et al. (2020). For the HLA class I locus, this dataset contains 2,138 distinct haplotypes spanning 230 HLA-A, HLA-B, and HLA-C alleles. For HLA class II, this dataset contains 1,711 distinct haplotypes spanning 280 HLA-DP, HLA-DQ, and HLA-DR alleles. Population frequencies are provided for three populations self-reporting as having White, Black, or Asian ancestry.

We consider nine subunit vaccines: the full envelope (E), membrane (M), nucleocapsid (N), and spike (S) proteins as well as the S1, S2, receptor binding domain (RBD), N-terminal domain (NTD), and fusion peptide (FP) domains from S. The amino acid positions for each of the S protein subunits are shown in Figure 2. When evaluating these subunit vaccines we include all peptides of length 8–10 (MHC class I) and 13–25 (MHC class II) spanning the corresponding regions of the proteome.

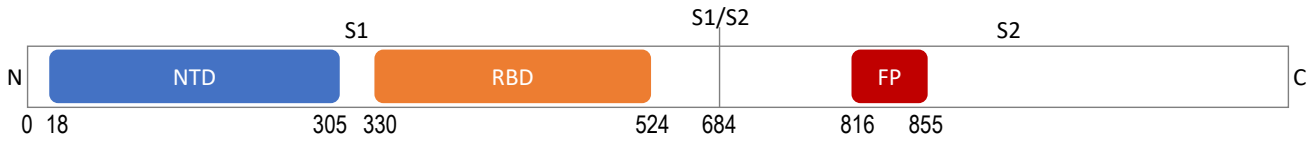

Figure 2. Illustration of functional domains on SARS-CoV-2 S protein.

2.2 EvalVax subunit vaccine evaluation

We evaluate population coverage of SARS-CoV-2 subunit vaccines using EvalVax-Robust (Liu et al., 2020). EvalVax-Robust computes population coverage of a given peptide set using the HLA haplotype frequencies in each population of individuals self-reporting as having Black, Asian, or White ancestry (Section 2.1). Population coverage $P(n)$ is defined as the fraction of individuals predicted to have $\geq n$ peptide-HLA binding hits with ≤ 50 nM predicted binding affinity. EvalVax-Robust computes

the frequency of diploid HLA genotypes, and accounts for both homozygous and heterozygous HLA loci. We compute the average population coverage as an unweighted average of population coverage over the three populations. Insufficient coverage of $\leq n$ hits is defined as $100\% - P(n+1)$.

Our subunit population coverage estimates are not lowered by discarding subunit peptides as unsuitable. We consider all peptides that result from a windowing of the subunit proteome, and include the redundant peptides caused by using varying window sizes at the same proteome start position. In addition, we do not filter peptides for mutation rate or glycosylation during evaluation.

2.3 Design of separate MHC class I and II peptide sets to augment subunit vaccine population coverage

In the separate design method we use OptiVax-Robust (Liu et al., 2020) to augment subunit vaccines with additional peptides to produce separate sets of peptides for class I and class II augmentation (Figure 1A). These additional peptides are selected from the remaining SARS-CoV-2 proteome (all peptides except those spanning the subunit), excluding peptides that are likely to mutate (have mutation rate > 0.001) or have non-zero predicted probability of glycosylation (Section 2.1). All candidate peptides considered during augmentation have ≤ 50 nM predicted binding affinity (Section 2.1).

The augmentation algorithm uses a starting peptide set which is extracted from the subunit vaccine to maximize the coverage of the subunit while removing redundant peptides resulting from overlapping sliding windows using the redundancy elimination algorithm found in Liu et al. (2020). Using a non-redundant starting peptide set ensures that augmentation does not depend upon redundant peptides for population coverage support. OptiVax-Robust performs vaccine augmentation by adding peptides to this starting set to improve the population coverage at each peptide-HLA hits cutoff n. At each iteration redundant peptides are removed from consideration, and redundancy is defined with an edit distance metric (Liu et al., 2020). OptiVax-Robust uses a beam search algorithm that iteratively expands the solution by one peptide and gradually optimizes population coverage from $n = 1$ to the targeting level of per-individual peptide-HLA hits (Liu et al., 2020). We use a beam size of 5 for the augmentation of subunit vaccines.

For each desired budget of augmentation peptides, OptiVax produces an augmentation set. Larger augmentation sets are not necessarily supersets of smaller augmentation sets, as the underlying combinatorial optimization problem is complex. A vaccine designer can evaluate how many peptides they wish to use to realize a predicted population coverage. For the augmentation sets in Table 1 we targeted 99% coverage at $n = 7$ for MHC class I augmentation (except for S which we target at 99.5% coverage) and 97% coverage at $n = 7$ for MHC class II augmentation. These population coverage targets are achieved during augmentation of the non-redundant subunit. For the evaluation of original and augmented subunit vaccines in Table 1, we provide results for all window derived subunit peptides and the non-redundant set of subunit peptides. All window peptides can include the same HLA binding epitope multiple times from its sampling by multiple windows, and thus serves as the predicted lower bound on population insufficient coverage. The non-redundant results are the predicted upper bound of population insufficient coverage.

2.4 Design of a single set of peptides to maximize MHC class I and II population coverage

We developed the OptiVax-Joint method to produce a minimal set of 25-mer peptides to reach a target population coverage probability at a threshold of n predicted hits for each individual for both MHC class I and class II (Figure 1B). The 25-mer candidate peptides are produced by windowing the pathogen proteome that is not part of a selected subunit, using a window step size of 8 amino acids between candidate peptides. Each of the candidate 25-mer peptides is annotated with its non-redundant peptides of length 8–10 (MHC class I) and 13–25 (MHC class II) and the HLA alleles they bind with ≤ 50 nM predicted binding affinity. Peptide redundancy is defined with an edit distance metric for the elimination of overlapping peptides (Liu et al., 2020).

OptiVax-Joint begins with the non-redundant set of peptides from a selected subunit or an empty set, and performs vaccine augmentation by adding candidate 25-mer peptides to this starting set to improve both MHC class I and class II population coverage at a target number of peptide-HLA hits n. When OptiVax-Joint is started with an empty set of peptides it produces a *de novo* peptide vaccine design without an associated subunit component. Each 25-mer is scored based on its contained annotated class I and class II peptides for its improvement in the number of per-individual peptide-HLA hits (Liu et al., 2020) over the haplotypes of the target population. Contained peptides are not counted towards population coverage if they have an observed mutation rate > 0.001 or have a non-zero predicted probability of glycosylation. OptiVax-Joint uses a beam search algorithm that iteratively expands the solution by one 25-mer peptide and gradually optimizes population coverage from $n = 1$ peptide hit to the targeted level of per-individual peptide-HLA hits for both MHC class I and class II (Liu et al., 2020). We use a beam size of 5 for the augmentation of subunit vaccines.

For each desired budget of augmentation peptides, OptiVax-Joint produces an augmentation set. Larger augmentation sets are not necessarily supersets of smaller augmentation sets, as the underlying combinatorial optimization problem is complex. A vaccine designer can evaluate how many peptides they wish to use to realize a predicted population coverage. For the joint

augmentation sets in Table 1, we targeted 99% coverage at $n = 7$ for MHC class I augmentation and 97% coverage at $n = 7$ for MHC class II augmentation jointly.

3 Results

3.1 SARS-CoV-2 subunit population coverage analysis

We first computed the predicted number of peptide-HLA hits that would result from an infection by the SARS-CoV-2 virus as a baseline. With a redundant sampling of the SARS-CoV-2 proteome we predict SARS-CoV-2 will have 342 (White), 330 (Black), and 426 (Asian) peptide-HLA hits for MHC class I on average in the respective self-reporting populations. For an MHC class II redundant sampling we predict SARS-CoV-2 will have 5108 (White), 3822 (Black), and 2041 (Asian) peptide-MHC hits. Thus the average number of predicted SARS-CoV-2 peptide-HLA hits for MHC class I is 366 and for MHC class II 3557.

We found that all subunits of SARS-CoV-2 have gaps in their predicted population coverage for robust peptide MHC display, either by having no of peptide-HLA hits or fewer than six peptide-HLA hits. We computed the predicted uncovered percentage of S subunit variants as a function of the minimum required predicted number of peptide-HLA hits displayed by an individual (Figure 3). Results for other subunits are shown in Figure S1. As shown in Table 1, the predicted percentage of the uncovered population increases for smaller subunits. The two subunits with the least coverage are the fusion peptide (FP) domain which comprises 40 amino acids and the E protein which comprises 75 amino acids. The receptor binding domain (RBD) subunit is predicted to have no MHC class II peptides displayed in 23.68% of the population (averaged across Asian, Black, and White self-reporting individuals) with ≤ 50 nM affinity. We note that the uncovered population of RBD with no predicted display of MHC class II peptides ranges from 5.81% for the population self-reporting as White, to a high of 51.54% for the population self-reporting as Asian. Thus, clinical trials need to carefully consider ancestry in their study designs to ensure that efficacy is measured across an appropriate population. For the RBD subunit, 32.99% of the population had fewer than 6 MHC class II peptide-HLA hits. For RBD MHC class I, the coverage gap is 1.19% for no hits and 51.31% for fewer than six hits. EvalVax predicted that on average for an S subunit vaccine the uncovered population would be 0.004% (class I) and 1.24% (class II) for no display, and 1.002% (class I) and 4.22% (class II) for fewer than 6 peptide-HLA hits.

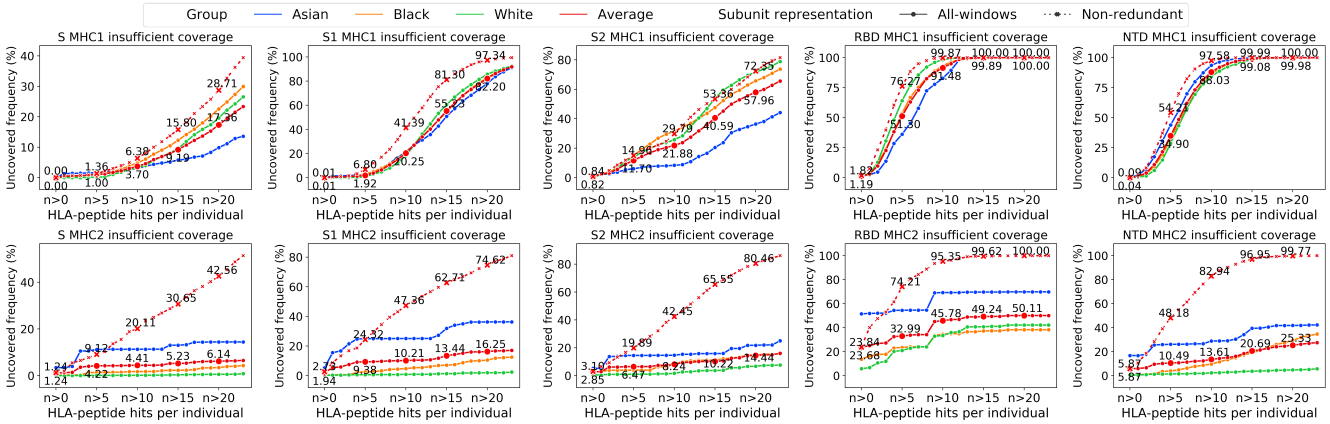

Figure 3. Predicted uncovered percentage of populations as a function of the minimum number of peptide-HLA hits in an individual. Annotated percentages are the average across populations self-reporting as Asian, Black, and White. A redundant sampling of peptides is depicted by solid lines for populations self-reporting as Asian, Black, and White as well their average. A non-redundant sampling of peptides is depicted by dotted lines.

3.2 SARS-CoV-2 subunit augmentation with separate peptide sets for MHC class I and II

We found augmentation peptides to improve the population coverage of MHC class I and class II peptide display for all of the SARS-CoV-2 subunit vaccines that we considered (Table 1; peptides given in Table S4). In the first design, we produced separate sets of peptides for display by MHC class I and class II molecules. We predicted the uncovered fraction of the population as a function of the additional peptides added to the vaccine subunits (Figure 4). We chose the augmentation set with minimal number of peptides that achieves the targeting criteria specified in Section 2.3. With the augmentation peptides, EvalVax predicted that on average for an S subunit vaccine the percentage of the population not displaying any vaccine peptides is reduced to 0.00% (class I) and 0.33% (class II), and for an RBD subunit vaccine it is reduced to 0.00% (class I) and 0.33% (class II). We then predicted the ability of the augmented vaccines to achieve more than five peptide-HLA hits and found the S subunit population coverage gap is reduced to 0.26% (class I) and 0.68% (class II), and 0.45% (class I) and 1.39% (class II) for

Subunit	Average				Asian				Black				White			
	All-windows	Non-redundant	Augmented (Separate)	Augmented (Joint)	All-windows	Non-redundant	Augmented (Separate)	Augmented (Joint)	All-windows	Non-redundant	Augmented (Separate)	Augmented (Joint)	All-windows	Non-redundant	Augmented (Separate)	Augmented (Joint)
Insufficient coverage (num. peptide-HLA hits = 0, 50 nM affinity)																
— MHC Class I —																
FP	47.510%	47.510%	0.000%	0.000%	24.241%	24.241%	0.000%	0.000%	73.577%	73.577%	0.000%	0.000%	44.713%	44.713%	0.000%	0.000%
E	17.866%	18.633%	0.000%	0.000%	12.732%	13.110%	0.000%	0.000%	17.993%	18.195%	0.000%	0.000%	22.873%	24.594%	0.000%	0.000%
N	1.500%	1.882%	0.000%	0.000%	2.353%	3.066%	0.000%	0.000%	1.599%	1.683%	0.000%	0.001%	0.547%	0.898%	0.000%	0.000%
RBD	1.188%	1.825%	0.000%	0.000%	1.822%	1.822%	0.000%	0.000%	1.400%	1.818%	0.000%	0.000%	0.342%	1.834%	0.000%	0.000%
S2	0.817%	0.844%	0.000%	0.000%	1.601%	1.617%	0.000%	0.000%	0.751%	0.811%	0.000%	0.000%	0.099%	0.103%	0.000%	0.000%
M	0.780%	1.438%	0.000%	0.001%	1.464%	1.510%	0.000%	0.000%	0.648%	1.329%	0.000%	0.002%	0.228%	1.474%	0.000%	0.000%
NTD	0.036%	0.086%	0.000%	0.001%	0.032%	0.073%	0.000%	0.000%	0.050%	0.147%	0.000%	0.002%	0.025%	0.037%	0.000%	0.000%
S1	0.008%	0.008%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.021%	0.023%	0.000%	0.001%	0.002%	0.002%	0.000%	0.000%
S	0.004%	0.004%	0.000%	0.004%	0.000%	0.000%	0.000%	0.000%	0.013%	0.013%	0.000%	0.012%	0.000%	0.000%	0.000%	0.000%
— MHC Class II —																
FP	82.289%	82.289%	0.334%	0.334%	96.918%	96.918%	0.926%	0.926%	75.902%	75.902%	0.058%	0.058%	74.049%	74.049%	0.018%	0.018%
E	71.600%	71.600%	0.334%	0.334%	77.335%	77.335%	0.926%	0.926%	79.977%	79.977%	0.058%	0.058%	57.488%	57.488%	0.018%	0.018%
RBD	23.682%	23.839%	0.334%	0.334%	51.537%	51.875%	0.926%	0.926%	13.698%	13.698%	0.058%	0.058%	5.811%	5.943%	0.018%	0.018%
N	13.893%	13.958%	0.334%	0.334%	32.505%	32.608%	0.926%	0.926%	4.769%	4.839%	0.058%	0.058%	4.405%	4.428%	0.018%	0.018%
NTD	5.869%	5.869%	0.334%	0.390%	16.760%	16.760%	0.926%	1.070%	0.422%	0.424%	0.058%	0.070%	0.424%	0.424%	0.018%	0.030%
M	4.247%	4.247%	0.047%	0.047%	6.751%	6.751%	0.123%	0.123%	3.482%	3.482%	0.017%	0.017%	2.507%	2.507%	0.003%	0.003%
S2	2.849%	3.104%	0.332%	0.332%	5.235%	5.290%	0.920%	0.920%	2.791%	3.494%	0.058%	0.058%	0.521%	0.529%	0.018%	0.018%
S1	1.942%	2.732%	0.334%	0.334%	5.257%	7.512%	0.926%	0.926%	0.313%	0.377%	0.058%	0.058%	0.256%	0.307%	0.018%	0.018%
S	1.241%	1.243%	0.332%	0.336%	3.439%	3.439%	0.920%	0.932%	0.257%	0.259%	0.058%	0.058%	0.026%	0.031%	0.018%	0.019%
Insufficient coverage (num. peptide-HLA hits $\leq$ 5, 50 nM affinity)																
— MHC Class I —																
FP	100.000%	100.000%	0.528%	0.518%	100.000%	100.000%	1.405%	1.293%	100.000%	100.000%	0.136%	0.212%	100.000%	100.000%	0.041%	0.049%
E	55.313%	87.668%	0.497%	0.199%	53.986%	82.507%	1.326%	0.013%	63.782%	89.893%	0.134%	0.487%	48.173%	90.603%	0.032%	0.096%
RBD	51.305%	76.267%	0.454%	0.544%	36.192%	67.913%	1.312%	1.330%	53.611%	73.194%	0.037%	0.258%	64.112%	87.692%	0.014%	0.043%
N	35.174%	51.885%	0.080%	0.192%	24.880%	44.617%	0.052%	0.013%	47.132%	59.576%	0.140%	0.492%	33.512%	51.461%	0.047%	0.071%
NTD	34.904%	54.215%	0.071%	0.176%	43.894%	61.144%	0.027%	0.018%	32.288%	52.560%	0.147%	0.442%	28.531%	48.941%	0.039%	0.070%
S2	11.704%	14.959%	0.117%	0.100%	6.047%	6.923%	0.026%	0.001%	15.071%	20.243%	0.276%	0.267%	13.994%	17.712%	0.048%	0.031%
M	10.105%	26.688%	0.068%	0.256%	3.577%	21.529%	0.001%	0.002%	14.449%	30.471%	0.184%	0.684%	12.287%	28.064%	0.020%	0.083%
S1	1.921%	6.802%	0.087%	0.528%	2.267%	10.712%	0.000%	1.347%	2.782%	5.721%	0.244%	0.225%	0.714%	3.973%	0.017%	0.012%
S	1.002%	1.364%	0.026%	0.128%	1.654%	1.692%	0.000%	0.001%	1.180%	1.462%	0.076%	0.367%	0.173%	0.938%	0.000%	0.016%
— MHC Class II —																
E	93.928%	99.044%	0.956%	1.839%	99.582%	99.950%	2.454%	4.925%	97.704%	99.783%	0.342%	0.475%	84.497%	97.399%	0.072%	0.117%
FP	89.868%	100.000%	1.308%	1.477%	97.567%	100.000%	3.493%	3.869%	82.084%	100.000%	0.317%	0.449%	89.953%	100.000%	0.112%	0.114%
RBD	32.991%	74.210%	1.386%	1.455%	54.476%	94.296%	3.487%	3.842%	23.506%	67.590%	0.584%	0.446%	20.992%	60.743%	0.086%	0.076%
N	23.763%	49.258%	0.959%	1.845%	50.429%	81.979%	2.444%	4.925%	8.029%	33.128%	0.358%	0.475%	12.830%	32.668%	0.076%	0.135%
NTD	10.494%	48.181%	1.116%	1.636%	26.124%	76.425%	2.630%	4.171%	3.925%	48.634%	0.537%	0.398%	1.433%	19.484%	0.179%	0.340%
S1	9.384%	24.317%	0.919%	1.431%	24.919%	50.111%	2.405%	3.861%	2.613%	16.525%	0.288%	0.292%	0.620%	6.315%	0.063%	0.141%
M	7.988%	27.749%	1.031%	1.191%	15.082%	48.539%	2.332%	3.104%	4.957%	19.686%	0.705%	0.405%	3.925%	15.021%	0.055%	0.064%
S2	6.475%	19.886%	0.913%	1.233%	14.499%	32.665%	1.747%	3.196%	3.946%	15.993%	0.930%	0.443%	0.980%	10.999%	0.062%	0.059%
S	4.216%	9.119%	0.678%	1.182%	11.194%	20.693%	1.722%	3.205%	1.345%	5.312%	0.268%	0.287%	0.109%	1.353%	0.042%	0.055%

Table 1. Percentage of a population that is insufficiently covered by subunit vaccines and the improvement after adding MHC class I and MHC class II augmentation peptides. Results are shown for both separate and joint designs of augmentation peptides. The list is sorted by decreasing insufficient coverage of unaugmented subunits.

RBD. The augmentation algorithm is also able to reduce the percentage of the population not displaying any vaccine peptides to 0.00% (class I) and less than 0.33% (class II), and reduce the population percentage displaying fewer than 6 vaccine peptides to less than 0.53% (class I) and less than 1.39% (class II) for all subunits by adding 16–42 additional peptides (Tables 1 and 2).

We used stringent criteria for predicting peptide-MHC hits (Section 2.1) to produce conservative metrics for population coverage, and thus alternative binding criteria may show smaller gaps in population coverage. Additional evaluation of subunits and augmentations using NetMHCpan-4.1 (Reynisson et al., 2020a) (MHC class I) and NetMHCIIpan-4.0 (Reynisson et al., 2020b) (MHC class II) predicted eluted ligand (EL) percentile ranking for selecting binders are shown in Table S3.

3.3 SARS-CoV-2 subunit augmentation with a joint peptide set for MHC class I and II

We used Optivax-Joint to compute a joint set of SARS-CoV-2 peptides to maximize the predicted population coverage for a target number of MHC class I and class II peptide hits in every individual. We computed joint peptide sets to augment 9 SARS-CoV-2 subunits and predicted the uncovered fraction of the population as a function of joint peptide set size (Table 1, Figure 5). We chose the set with the minimal number of peptides that achieves the targeting criteria specified in Section 2.4 for Table 1. We show that joint design simultaneously increases the predicted population coverage for both MHC class I and class II display (Figure 5).

3.4 A peptide-only SARS-CoV-2 vaccine design with a joint peptide set for MHC class I and II

We designed a standalone peptide vaccine that did not assume an associated subunit component and found that it was predicted to have greater than 94% population coverage with $n \geq 5$ hits with 9 25-mer peptides (Table 3). We explored the predicted decrease of the uncovered population as a function of peptide count and found that joint vaccine designs are predicted to simultaneously produce a large number of predicted hits for both MHC class I and class II display (Figure 5). The predicted

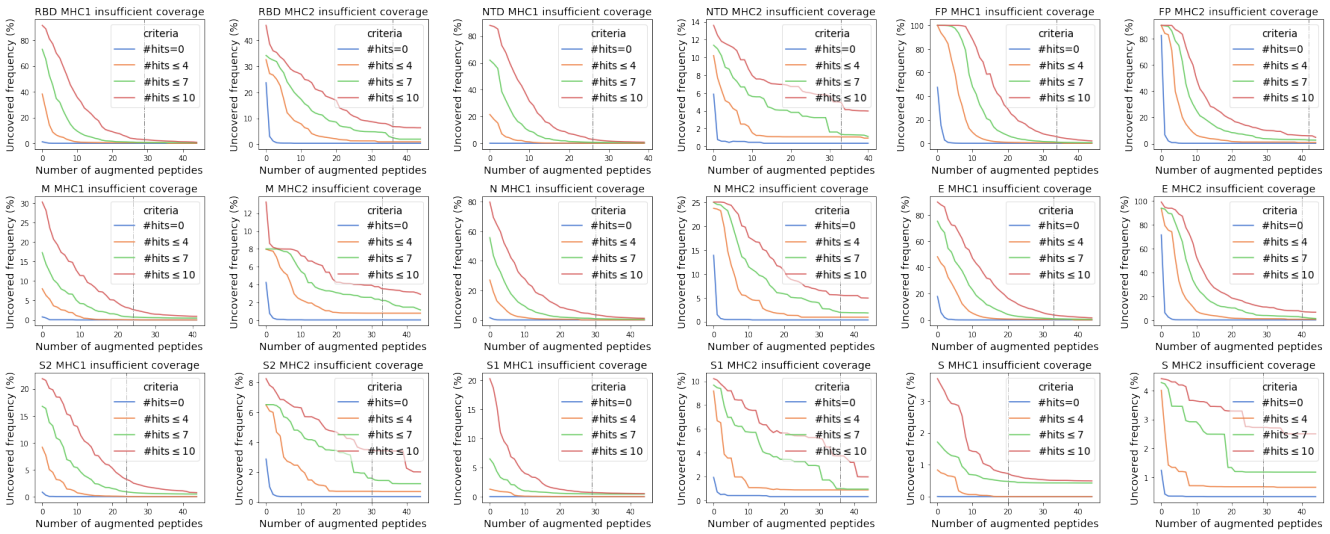

Figure 4. Predicted uncovered percentage of the population for MHC peptide display using separate sets for MHC class I and class II as a function of the number of augmentation peptides at different predicted peptide-HLA hit thresholds in each individual. The dotted vertical line shows the peptide count used in Table 1.

		S	S1	RBD	NTD	S2	FP	M	N	E
Joint design	Num. 25-mer peptides	18	22	31	25	28	34	27	30	33
	Num. amino acids in construct	620	760	1075	865	970	1180	935	1040	1145
Separate design	Num. MHC1 peptides	20	16	29	26	24	34	24	30	33
	Num. MHC2 peptides	24	31	36	30	26	42	33	33	40
	Num. amino acids in construct	1017	1144	1523	1300	1132	1765	1348	1434	1702

Table 2. Number of peptides used in augmentation set listed in Table 1 and total number of amino acids required to build the full construct, including 10 amino acid linkers between every peptide.

population coverage of a 24 25-mer peptide design exhibits a diverse display of peptides across populations self-reporting as Black, White, and Asian (Figure 6). Peptide-only designs have been found to be effective (Herst et al., 2020), and our joint design of a peptide-only vaccine is more compact than a separate design at equivalent levels of population coverage (Figure 7).

3.5 SARS-CoV-2 joint designs are more compact than separate designs

To compare the efficiency of joint optimization and separate optimization, we set a series of population coverage goals for MHC class I and MHC class II coverage with more than 7 peptide-HLA hits per individual. We considered 25 evenly spaced coverage levels between the unannotated subunit coverage and 100% coverage. We computed the total number of peptides needed to reach each set of MHC class I and MHC class II coverage goals simultaneously, where the number of MHC class I and MHC class II peptides are summed for separately designed peptide sets. We computed the total number of amino acids needed for a construct with a typical mRNA delivery platform with 10 amino acid linkers (Sahin et al., 2017) (Table 2, Section 4). A single mRNA delivery construct has been demonstrated to work for both MHC class I and class II peptides (Kreiter et al., 2008; Sahin et al., 2017). As shown in Figure 7 and Figure 8, the joint designs reduce the total number of required peptides and amino acids to achieve each level of population coverage.

4 Discussion

We augment subunit vaccines with a compact set of peptides to improve the display of a vaccine on HLA class I and II molecules across a population of people. Subunit vaccines offer safety advantages over inactivated or attenuated pathogen vaccines, but their ability to fully mimic a pathogenic infection to train cellular immunity is limited. Immunity to a pathogen may rest in part upon T cell based adaptive immunity and corresponding T memory cells. We expect that a vaccine that provides a diverse display of a pathogen's peptides will create reservoirs of $CD4^+$ and $CD8^+$ memory cells that will assist in establishing immunity to the pathogen.

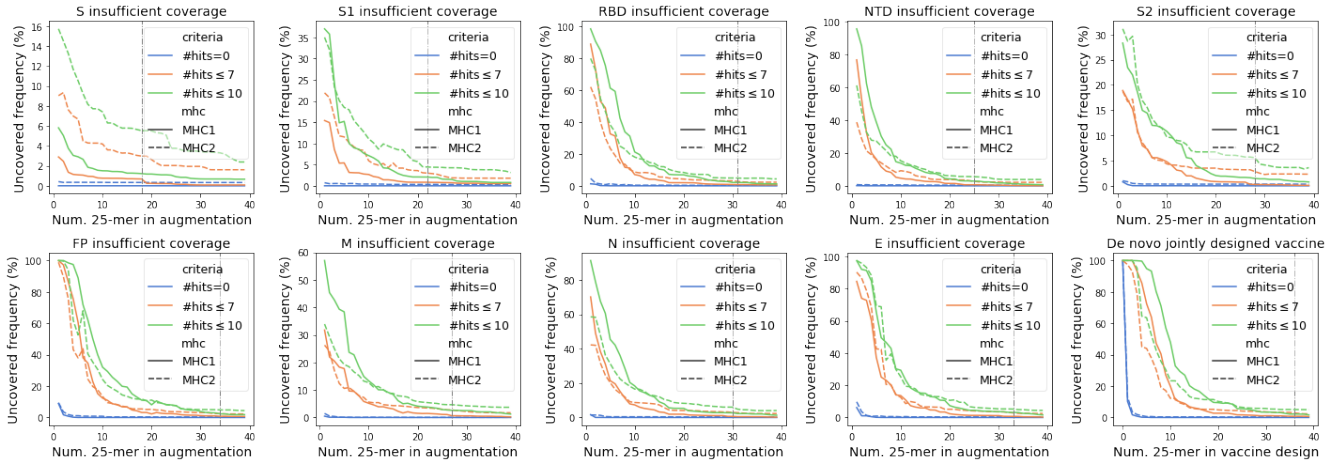

Figure 5. Predicted uncovered percentage of the population for MHC peptide display using a joint set as a function of the number of augmentation peptides at different predicted peptide-HLA hit thresholds in each individual. The dotted vertical line shows the peptide count used in Table 1.

25-mer Peptide Count	MHC Class	Extracted non-redundant peptides	EvalVax-Robust ($n \geq 1$)	EvalVax-Robust ($n \geq 5$)	EvalVax-Robust ($n \geq 10$)	Exp #peptide-HLA hits (Asian)	Exp #peptide-HLA hits (Black)	Exp #peptide-HLA hits (White)	Protein origin
4	MHC1	16	99.824%	60.413%	1.954%	5.934	4.387	5.115	ORF1b, M, S2
	MHC2	15	98.849%	71.326%	43.112%	5.330	9.351	10.515	
9	MHC1	35	99.999%	94.925%	52.390%	11.411	9.112	9.882	ORF1b, ORF1a, M, S2, ORF3a
	MHC2	31	99.597%	94.462%	75.454%	11.647	15.144	18.652	
14	MHC1	56	99.999%	98.777%	85.005%	17.131	14.268	15.067	ORF1a, ORF1b, M, S2, ORF3a
	MHC2	46	99.610%	97.148%	89.794%	15.732	19.821	24.356	
19	MHC1	72	100.000%	99.368%	93.087%	20.641	17.015	18.081	ORF1a, ORF1b, M, N, S2, ORF3a
	MHC2	58	99.610%	98.346%	91.452%	19.198	23.647	29.260	
24	MHC1	84	100.000%	99.461%	96.275%	24.914	19.629	21.125	ORF1a, ORF1b, M, N, S2, ORF3a
	MHC2	74	99.666%	99.013%	94.255%	21.559	28.946	34.721	

Table 3. Predicted population coverage of a peptide-only vaccine jointly optimized for MHC class I and class II coverage with 4, 9, 14, 19, and 24 25-mer peptides.

We found that for SARS-CoV-2 the joint optimization of predicted MHC class I and class II pathogen peptide display achieves population coverage criteria with a more compact vaccine design than designing separate peptide sets for MHC class I and class II. Using a simpler design with shorter constructs may contribute to the effectiveness of a vaccine by providing an equivalent diversity of peptide display in a population with a less complex mixture of vaccine peptides.

Augmentation peptides can be delivered using the same vehicle as their associated subunit vaccine or they can be delivered separately. Nucleic acid based vaccines can incorporate RNA or DNA sequences that encode class I and class II augmentation peptides with desired signal sequences, linkers, and protease cleavage sites (Kreiter et al., 2008; Sahin et al., 2017) (examples in Supplementary Information, Tables S1 and S2). The peptides can be expressed as part of the subunit or separately, and can be encoded on the same or different molecules as the primary subunit. When augmentation peptides are added as a new subunit domain a vaccine designer can trade-off domain complexity for additional coverage using Figures 7 and 8. Nucleic acid constructs carrying augmentation peptides can be delivered by injection in lipid nanoparticle particle carriers or directly (Dowdy, 2017; Wolff et al., 1990). Protein based vaccines can include independent augmentation peptides into the vaccine formulation. The delivery of independent augmentation peptides can be accomplished using nanoparticles (Herst et al., 2020).

Our computational objective function encodes the two key goals of our augmentation strategy: population coverage and the display of a highly diverse set of peptides in each individual. Our population coverage goal is ensured by optimizing predicted display coverage over population haplotype frequencies. The display of a diverse set of peptides is established by setting augmentation design goals for the number of peptides that need to be displayed by each individual.

Early results from clinical studies of subunit vaccines for SARS-CoV-2 show that some vaccine recipients did not develop

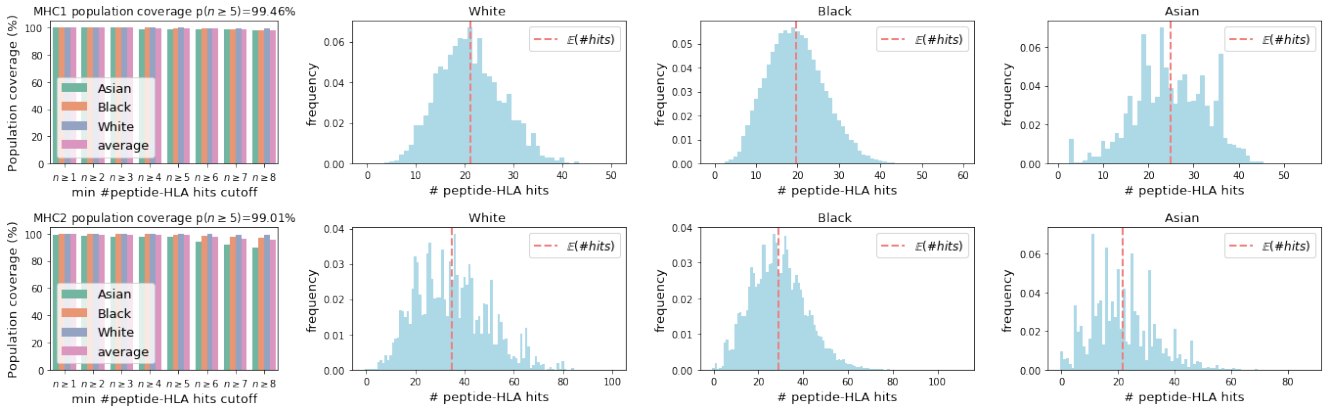

Figure 6. Predicted coverage in populations self-reporting as White, Black, and Asian with a peptide-only vaccine comprising 24 25-mer peptides jointly optimized for MHC class I and MHC class II coverage. The red dotted vertical line shows the expected number of hits.

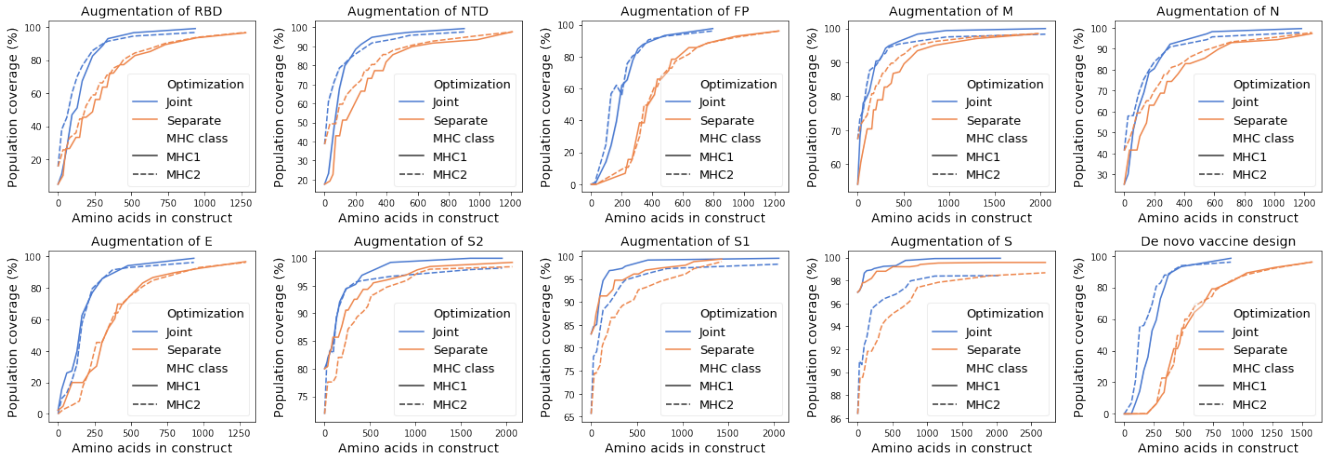

Figure 7. Comparison of the number of amino acids used by separate and joint designs and their respective predicted population coverage with more than 7 peptide-HLA hits per individual. The predicted population coverage is shown for the augmentation of nine subunits and for a peptide-only *de novo* vaccine design.

positive T cell responses. It is difficult to fully evaluate these results because the HLA types of study participants are not provided by these early studies. Thus these study populations may not be reflective of HLA types in the general world population. The BNT162b1 RBD subunit vaccine produced a less robust $CD8^+$ response than $CD4^+$ response (Sahin et al., 2020), and this was also noted in the mRNA-1273 S subunit vaccine results (Jackson et al., 2020). Further clinical data is required to fully assess the the T cell immunogenicity of various subunits and delivery methods. Clinical trials should select their participants to have representative HLA type distributions to test for population coverage. We note that in a study of SARS-CoV-2 subunit vaccines in non-human primates the level of neutralizing antibodies was observed to correlate with SARS-CoV-2 protection shortly after vaccination, but T cell response was not observed to correlate with protection (Yu et al., 2020). In this study rhesus macaques were vaccinated with an S subunit vaccine at week 0 and 3, T cell activity was measured at week 5, animals were challenged with SARS-CoV-2 at week 6, and then it was observed that week 5 T cell activity was uncorrelated with viral load post challenge. Future studies will need to examine the durability of immunity in individuals with minimal T cell response.

By simultaneously achieving the twin goals of coverage and diversity with peptides derived from a pathogen, we effectively compress the cellular immunologic fingerprint of a pathogen into a vaccine. To produce an antibody response, the subunit component of a vaccine can encode a three-dimensional epitope to stimulate neutralizing antibody production by B cells. Taken together, these two designed components, a pathogen subunit and its augmentation, will provide both B cell and T cell epitopes of a pathogen while permitting epitope selection to mitigate deleterious effects and improve population coverage.

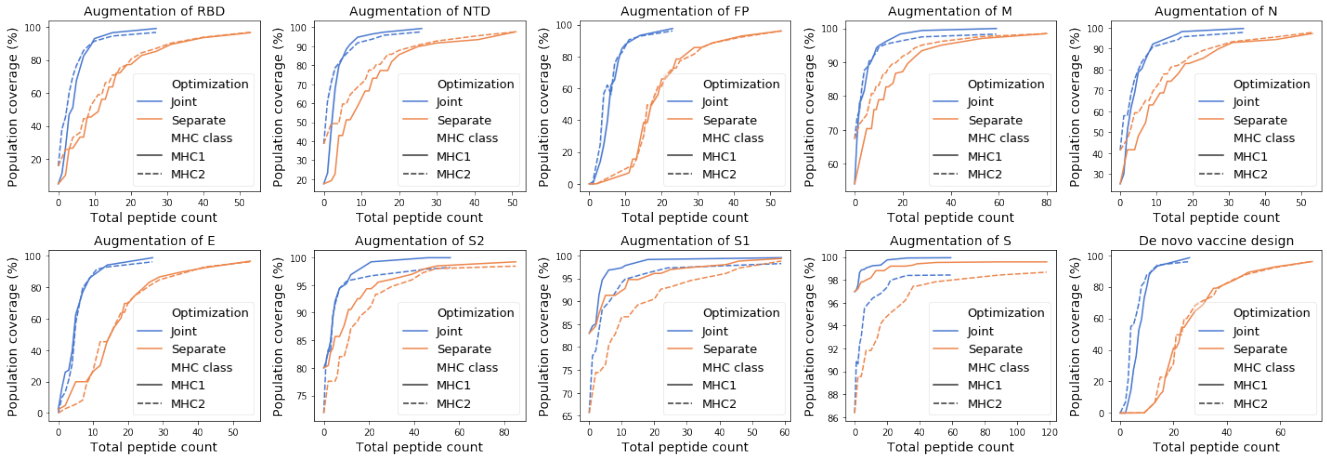

Figure 8. Comparison of the number of peptides used by separate and joint designs and their respective predicted population coverage with more than 7 peptide-HLA hits per individual. The predicted population coverage is shown for the augmentation of nine subunits and for a peptide-only *de novo* vaccine design.

All of our software and data are freely available as open source to allow others to use and extend our methods.

Acknowledgements

This work was supported in part by Schmidt Futures, a Google Cloud Platform grant, and NIH grant R01CA218094 to D.K.G. We benefited from thoughtful comments from Michael Birnbaum and Brooke Huisman.

References

Alexander, J., del Guercio, M.F., Maewal, A., Qiao, L., Fikes, J., Chesnut, R.W., Paulson, J., Bundle, D.R., DeFrees, S., Sette, A. (2000). Linear PADRE T helper epitope and carbohydrate B cell epitope conjugates induce specific high titer IgG antibody responses. The Journal of Immunology 164, 1625–1633.

Calis, J.J., Maybeno, M., Greenbaum, J.A., Weiskopf, D., De Silva, A.D., Sette, A., Keşmir, C., Peters, B. (2013). Properties of MHC class I presented peptides that enhance immunogenicity. PLoS computational biology 9.

Croft, N.P., Smith, S.A., Pickering, J., Sidney, J., Peters, B., Faridi, P., Witney, M.J., Sebastian, P., Flesch, I.E., Heading, S.L., et al. (2019). Most viral peptides displayed by class I MHC on infected cells are immunogenic. Proceedings of the National Academy of Sciences 116, 3112–3117.

Dai, L., Zheng, T., Xu, K., Han, Y., Xu, L., Huang, E., An, Y., Cheng, Y., Li, S., Liu, M., et al. (2020). A universal design of betacoronavirus vaccines against COVID-19, MERS and SARS. Cell https://doi.org/10.1016/j.cell.2020.06.035.

Dowdy, S.F. (2017). Overcoming cellular barriers for RNA therapeutics. Nature Biotechnology 35, 222–229.

Elbe, S., Buckland-Merrett, G. (2017). Data, disease and diplomacy: GISAID's innovative contribution to global health. Global Challenges 1, 33–46.

Falugi, F., Petracca, R., Mariani, M., Luzzi, E., Mancianti, S., Carinci, V., Melli, M.L., Finco, O., Wack, A., Tommaso, A.D., et al. (2001). Rationally designed strings of promiscuous CD4+ T cell epitopes provide help to haemophilus influenzae type b oligosaccharide: a model for new conjugate vaccines. European Journal of Immunology 31, 3816–3824.

Gasper, D., Tejera, M., Suresh, M. (2014). CD4 T-cell memory generation and maintenance. Critical Reviews in Immunology 34, 121–146.

Gupta, R., Jung, E., Brunak, S. (2004). Prediction of N-glycosylation sites in human proteins. In preparation URL: http://www.cbs.dtu.dk/services/NetNGlyc/.

Hadfield, J., Megill, C., Bell, S.M., Huddleston, J., Potter, B., Callender, C., Sagulenko, P., Bedford, T., Neher, R.A. (2018). Nextstrain: real-time tracking of pathogen evolution. Bioinformatics 34, 4121–4123.

Herst, C.V., Burkholz, S., Sidney, J., Sette, A., Harris, P.E., Massey, S., Brasel, T., Cunha-Neto, E., Rosa, D.S., Chao, W.C.H., et al. (2020). An effective CTL peptide vaccine for ebola zaire based on survivors' CD8+ targeting of a particular nucleocapsid protein epitope with potential implications for COVID-19 vaccine design. Vaccine 38, 4464–4475.

Jackson, L.A., Anderson, E.J., Rouphael, N.G., Roberts, P.C., Makhene, M., Coler, R.N., McCullough, M.P., Chappell, J.D., Denison, M.R., Stevens, L.J., et al. (2020). An mRNA vaccine against SARS-CoV-2 —— preliminary report. New England Journal of Medicine https://doi.org/10.1056/NEJMoa2022483.

Jurtz, V., Paul, S., Andreatta, M., Marcatili, P., Peters, B., Nielsen, M. (2017). NetMHCpan-4.0: improved peptide–MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. The Journal of Immunology 199, 3360–3368.

Kreiter, S., Selmi, A., Diken, M., Sebastian, M., Osterloh, P., Schild, H., Huber, C., Türeci, Ö., Sahin, U. (2008). Increased antigen presentation efficiency by coupling antigens to MHC class I trafficking signals. The Journal of Immunology 180, 309–318.

Liu, G., Carter, B., Bricken, T., Jain, S., Viard, M., Carrington, M., Gifford, D.K. (2020). Computationally optimized SARS-CoV-2 MHC class I and II vaccine formulations predicted to target human haplotype distributions. Cell Systems https://doi.org/10.1016/j.cels.2020.06.009.

Moyle, P.M., Toth, I. (2013). Modern subunit vaccines: development, components, and research opportunities. ChemMedChem 8, 360–376.

O'Donnell, T., Rubinsteyn, A., Laserson, U. (2020). A model of antigen processing improves prediction of MHC I-presented peptides. bioRxiv, https://doi.org/10.1101/2020.03.28.013714.

O'Donnell, T.J., Rubinsteyn, A., Bonsack, M., Riemer, A.B., Laserson, U., Hammerbacher, J. (2018). MHCflurry: open-source class I MHC binding affinity prediction. Cell Systems 7, 129–132.

Ogishi, M., Yotsuyanagi, H. (2019). Quantitative prediction of the landscape of T cell epitope immunogenicity in sequence space. Frontiers in immunology 10, 827.

Reynisson, B., Alvarez, B., Paul, S., Peters, B., Nielsen, M. (2020a). 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 Research https://doi.org/10.1093/nar/gkaa379.

Reynisson, B., Barra, C., Kaabinejadian, S., Hildebrand, W.H., Peters, B., Nielsen, M. (2020b). Improved prediction of MHC II antigen presentation through integration and motif deconvolution of mass spectrometry MHC eluted ligand data. Journal of Proteome Research 19, 2304–2315.

Riley, T.P., Keller, G.L., Smith, A., Devlin, J.R., Davancaze, L.M., Arbuiso, A.A., Baker, B.M. (2019). Structure based prediction of neoantigen immunogenicity. Frontiers in immunology 10, 2047.

Sahin, U., Derhovanessian, E., Miller, M., Kloke, B.P., Simon, P., Löwer, M., Bukur, V., Tadmor, A.D., Luxemburger, U., Schrörs, B., et al. (2017). Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. Nature 547, 222–226.

Sahin, U., Muik, A., Derhovanessian, E., Vogler, I., Kranz, L.M., Vormehr, M., Baum, A., Pascal, K., Quandt, J., Maurus, D., Brachtendorf, S., Loerks, V.L., Sikorski, J., Hilker, R., Becker, D., Eller, A.K., Gruetzner, J., Boesler, C., Rosenbaum, C., Kuehnle, M.C., Luxemburger, U., Kemmer-Brueck, A., Langer, D., Bexon, M., Bolte, S., Kariko, K., Palanche, T., Fischer, B., Schultz, A., Shi, P.Y., Fontes-Garfias, C., Perez, J.L., Swanson, K.A., Loschko, J., Scully, I.L., Cutler, M., Kalina, W., Kyratsous, C.A., Cooper, D., Dormitzer, P.R., Jansen, K.U., Tuereci, O. (2020). Concurrent human antibody and TH1 type T-cell responses elicited by a COVID-19 RNA vaccine. medRxiv, https://doi.org/10.1101/2020.07.17.20140533.

Srinivasan, S., Cui, H., Gao, Z., Liu, M., Lu, S., Mkandawire, W., Narykov, O., Sun, M., Korkin, D. (2020). Structural genomics of SARS-CoV-2 indicates evolutionary conserved functional regions of viral proteins. Viruses 12, 360.

Wang, N., Shang, J., Jiang, S., Du, L. (2020). Subunit vaccines against emerging pathogenic human coronaviruses. Frontiers in Microbiology 11, 298.

Wolff, J.A., Malone, R.W., Williams, P., Chong, W., Acsadi, G., Jani, A., Felgner, P.L. (1990). Direct gene transfer into mouse muscle in vivo. Science 247, 1465–1468.

Yu, J., Tostanoski, L.H., Peter, L., Mercado, N.B., McMahan, K., Mahrokhian, S.H., Nkolola, J.P., Liu, J., Li, Z., Chandrashekar, A., et al. (2020). DNA vaccine protection against SARS-CoV-2 in rhesus macaques. Science https://doi.org/10.1126/science.abc6284.

Supplementary Information

Sequence Construct Examples

Here we provide example construct designs optimized to augment RBD subunit vaccines. The constructs in Table S1 contain 29 peptides and 36 peptides independently optimized for either MHC class I and MHC class II, respectively (Table 1). Table S2 contains a construct with 31 25-mer peptides jointly optimized for both MHC class I and MHC class II. Peptides are prepended with a secretion signal sequence at the N-terminus and followed by an MHC class I trafficking signal (MITD) (Kreiter et al., 2008; Sahin et al., 2017). The MITD has been shown to route antigens to pathways for HLA class I and class II presentation (Kreiter et al., 2008). Here we combine all peptides of each MHC class into a single construct using a non-immunogenic glycine/serine linkers from Sahin et al. (2017), though it is also plausible to construct individual constructs containing single peptides with the same secretion and MITD signals as demonstrated by Kreiter et al. (2008).

Design	Sequence
MHC Class I RBD Augmentation Construct (OptiVax-Robust)	MRVTAPRTLILLLSGALALTETWAGSGGSGGGGSGG**ETISLAGSY**GGSGGGGSGG**FLAHIQWMV**GGSGGGGSGG **FPRGQGVPI**GGSGGGGSGG**FRNARNGVL**GGSGGGGSGG**FRYMNSQGL**GGSGGGGSGG**FSAVGNICY**GGSGGGGS GG**FTISVTTEI**GGSGGGGSGG**FTSDYYQLY**GGSGGGGSGG**FVNEFYAYL**GGSGGGGSGG**HSYFTSDYY**GGSGG GGSGG**KLFDRYFKY**GGSGGGGSGG**LPVNVAFEL**GGSGGGGSGG**LRAKHYVYI**GGSGGGGSGG**LSYFIASFR**GG SGGGGSGG**LTDEMIAQY**GGSGGGGSGG**MADQAMTQM**GGSGGGGSGG**MPASWVMRI**GGSGGGGSGG**MPYFFTLLL** GGSGGGGSGG**NYMPYFFTL**GGSGGGGSGG**RIMASLVLAR**GGSGGGGSGG**RVVRSIFSR**GGSGGGGSGG **SEFSSLPSY**GGSGGGGSGG**STFNVPMEK**GGSGGGGSGG**TRFFYVLGL**GGSGGGGSGG**TVIEVQGYK**GGSGGGGSGG **VLWAHGFEL**GGSGGGGSGG**VYMPASWVM**GGSGGGGSGG**YADVFHLYL**GGSGGGGSGG**YLQPRTFLL**GGSLGGGGSG IVGIVAGLAVLAVVVIGAVVATVMCRRKSSGGKGGSYSQAASSDSAQGSDVSLTA
MHC Class II RBD Augmentation Construct (OptiVax-Robust)	MRVTAPRTLILLLSGALALTETWAGSGGSGGGGSGG**AMQTMLFTMLRKLDN**GGSGGGGSGG**AYLRKHFSMMILSDD**GGSGG GGSGG**CTFEYVSQPFLMDLE**GGSGGGGSGG**FKSVLYYQNNVFMSEAKCWT**GGSGGGGSGG**GTITSGWTFGAGAALQIPFA**G GSGGGGSGG**GWTFGAGAALQIPFAM**GGSGGGGSGG**IPTITQMNLKYAISAKNRAR**GGSGGGGSGG**IQQVVDADSKIVQLS** GGSGGGGSGG**ITQMNLKYAISAKNRARTVA**GGSGGGGSGG**KHFYWFFSNYLKRRVVFNGV**GGSGGGGSGG **KVTSAMQTMLFTMLRKL**GGSGGGGSGG**LDISASIVAGGIVAIV**GGSGGGGSGG**LFVAAIFYLITPVHV**GGSGGGGSGG **LNRNYVFTGYRVTKNSKVQI**GGSGGGGSGG**LVASIKNFKSVLYYQNNVFM**GGSGGGGSGG**MNLKYAISAKNRARTVAGVS** GGSGGGGSGG**NFKSVLYYQNNVFMS**GGSGGGGSGG**NIFGTVYEKLKPVLD**GGSGGGGSGG**NNLVVMAYITGGVVQLT**G GSGGGGSGG**QQLIRAAEIRASANL**GGSGGGGSGG**QWLTNIFGTVYEKLK**GGSGGGGSGG**SATLVSDIDITFLKKD**GGS GGGGSGG**SFRLFARTRSMWSFN**GGSGGGGSGG**SPFVMMSAPPAQYEL**GGSGGGGSGG**STQEFRYMNSQGLLP**GGSGGG GSGG**SYAAFATAQEAYEQA**GGSGGGGSGG**TFCAGSTFISDEVARDLSLQFKRPI**GGSGGGGSGG**TKHFYWFFSNYLKRR** GGSGGGGSGG**TRYVLMDGSIIQFPN**GGSGGGGSGG**TYRRLISMMGFKMNY**GGSGGGGSGG**VAYESLRPDTRYVLM** GGSGGGGSGG**VAYESLRPDTRYVLMDGSIIQFPN**GGSGGGGSGG**VDTDFVNEFYAYLRKHFSMM**GGSGGGGSGG **VRKIFVDGVPFVVST**GGSGGGGSGG**YIICISTKHFYWFFSNYLKRRVVFN**GGSGGGGSGG**YVTQQLIRAAEIRAS**GGSLG GGGSGIVGIVAGLAVLAVVVIGAVVATVMCRRKSSGGKGGSYSQAASSDSAQGSDVSLTA

Table S1. Example protein constructs for augmentations to RBD subunit vaccines optimized for either MHC class I or MHC class II by OptiVax-Robust. Constructs contain a secretion signal sequence (red), peptides (bold) joined by non-immunogenic glycine/serine linkers, and an MHC class I trafficking signal (blue). The augmentation peptides encoded are the same as those evaluated in Table 1.

Design	Sequence
Joint MHC Class I and Class II RBD Augmentation Construct (OptiVax-Joint)	MRVTAPRTLILLLSGALALTETWAGSGGSGGGGSGG**ASEFSSLPSYAAFATAQEAYEQAVA**GGSGGGGSGG**DTDFVNEFYA** **YLRKHFSMMILSDDA**GGSGGGGSGG**EKVNINIVGDFKLNEEIAIILASFS**GGSGGGGSGG**FGLVAEWFLAYILFTRFFYV** **LGLAA**GGSGGGGSGG**FKIYSKHTPINLVRDLPQGFSALEP**GGSGGGGSGG**GIATVREVLSDRELHLSWEVGKPRP**GGSGG GGSGG**GLMWLSYFIASFRLFARTRSMWSFN**GGSGGGGSGG**GVYDYLVSTQEFRYMNSQGLLPPKN**GGSGGGGSGG**IICIS** **TKHFYWFFSNYLKRRVVFNG**GGSGGGGSGG**IPKDMTYRRLISMMGFKMNYQVNGY**GGSGGGGSGG**KFPRGQGVPINTNSS** **PDDQIGYYRR**GGSGGGGSGG**LDISASIVAGGIVAIVVTCLAYYFM**GGSGGGGSGG**LQSLQTYVTQQLIRAAEIRASANLA** GGSGGGGSGG**MPNMLRIMASLVLARKHTTCCSLSH**GGSGGGGSGG**MPYFFTLLLQLCTFTRSTNSRIKAS**GGSGGGGSGG **NNLVVMAYITGGVVQLTSQWLTNIF**GGSGGGGSGG**PLIQPIGALDISASIVAGGIVAIVV**GGSGGGGSGG**SASTSAFVET** **VKGLDYKAFKQIVES**GGSGGGGSGG**SIKNFKSVLYYQNNVFMSEAKCWTE**GGSGGGGSGG**SSANNCTFEYVSQPFLMDLE** **GKQGN**GGSGGGGSGG**TDTPKGPKVKYLYFIKGLNNLNRGM**GGSGGGGSGG**TFCAGSTFISDEVARDLSLQFKRPI**GGSGG GGSGG**TITQMNLKYAISAKNRARTVAGVSI**GGSGGGGSGG**TITSGWTFGAGAALQIPFAMQMAYR**GGSGGGGSGG**TSQWL** **TNIFGTVYEKLKPVLDWLEE**GGSGGGGSGG**VRKIFVDGVPFVVSTGYHFRELGVV**GGSGGGGSGG**VTSAMQTMLFTMLRK** **LDNDALNNII**GGSGGGGSGG**VVFVLWAHGFELTSMKYFVKIGPER**GGSGGGGSGG**VVLHSYFTSDYYQLYSTQLSTDTGV** GGSGGGGSGG**YESLRPDTRYVLMDGSIIQFPNTYL**GGSGGGGSGG**YFNMVYMPASWVMRIMTWLDMVDTS**GGSLGGGGSG IVGIVAGLAVLAVVVIGAVVATVMCRRKSSGGKGGSYSQAASSDSAQGSDVSLTA

Table S2. Example protein construct for augmentations to RBD subunit vaccines jointly optimized for both MHC class I and class II by OptiVax-Joint. Constructs contain a secretion signal sequence (red), 31 25-mer peptides (bold) joined by non-immunogenic glycine/serine linkers, and an MHC class I trafficking signal (blue). The augmentation peptides encoded are the same as those evaluated in Table 1.

Subunit	Average				Asian				Black				White			
	All-windows	Non-redundant	Augmented (Separate)	Augmented (Joint)	All-windows	Non-redundant	Augmented (Separate)	Augmented (Joint)	All-windows	Non-redundant	Augmented (Separate)	Augmented (Joint)	All-windows	Non-redundant	Augmented (Separate)	Augmented (Joint)
Insufficient coverage (num. peptide-HLA hits = 0)																
— MHC Class I, EL rank ≤ 0.5% —																
FP	14.573%	34.093%	0.000%	0.000%	21.136%	23.464%	0.000%	0.000%	18.332%	52.420%	0.000%	0.000%	4.252%	26.395%	0.000%	0.000%
E	0.024%	0.057%	0.000%	0.000%	0.000%	0.085%	0.000%	0.000%	0.048%	0.051%	0.000%	0.000%	0.023%	0.036%	0.000%	0.000%
M	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
N	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
NTD	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
RBD	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
S	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
S1	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
S2	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
— MHC Class II, EL rank ≤ 2% —																
FP	47.976%	80.428%	0.000%	0.000%	54.727%	83.763%	0.000%	0.000%	47.335%	72.975%	0.000%	0.001%	41.865%	84.547%	0.000%	0.000%
E	43.986%	84.768%	0.000%	0.000%	60.733%	96.098%	0.000%	0.000%	32.480%	86.090%	0.000%	0.001%	38.744%	72.117%	0.000%	0.000%
M	0.020%	7.409%	0.000%	0.000%	0.005%	5.952%	0.000%	0.000%	0.011%	7.719%	0.000%	0.000%	0.046%	8.557%	0.000%	0.000%
RBD	0.000%	5.255%	0.000%	0.000%	0.000%	4.538%	0.000%	0.000%	0.000%	6.612%	0.000%	0.000%	0.000%	4.616%	0.000%	0.000%
NTD	0.000%	1.689%	0.000%	0.000%	0.000%	4.434%	0.000%	0.000%	0.000%	0.352%	0.000%	0.000%	0.000%	0.281%	0.000%	0.000%
N	0.000%	0.102%	0.000%	0.000%	0.000%	0.298%	0.000%	0.000%	0.000%	0.002%	0.000%	0.000%	0.000%	0.005%	0.000%	0.000%
S	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
S1	0.000%	0.005%	0.000%	0.000%	0.000%	0.001%	0.000%	0.000%	0.000%	0.014%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
S2	0.000%	0.002%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.005%	0.000%	0.000%	0.000%	0.002%	0.000%	0.000%
Insufficient coverage (num. peptide-HLA hits ≤ 5)																
— MHC Class I, EL rank ≤ 0.5% —																
FP	98.232%	100.000%	0.000%	0.000%	99.962%	100.000%	0.000%	0.000%	99.520%	100.000%	0.000%	0.000%	95.214%	100.000%	0.000%	0.000%
E	15.378%	34.870%	0.000%	0.000%	12.606%	32.072%	0.000%	0.000%	13.573%	30.229%	0.000%	0.000%	19.956%	42.307%	0.000%	0.000%
RBD	0.010%	7.167%	0.000%	0.000%	0.000%	3.178%	0.000%	0.000%	0.028%	9.045%	0.000%	0.000%	0.000%	9.279%	0.000%	0.000%
M	0.000%	0.192%	0.000%	0.000%	0.000%	0.342%	0.000%	0.000%	0.001%	0.108%	0.000%	0.000%	0.000%	0.125%	0.000%	0.000%
N	0.000%	0.691%	0.000%	0.000%	0.000%	0.298%	0.000%	0.000%	0.000%	0.578%	0.000%	0.000%	0.000%	1.196%	0.000%	0.000%
NTD	0.000%	0.041%	0.000%	0.000%	0.000%	0.111%	0.000%	0.000%	0.000%	0.002%	0.000%	0.000%	0.000%	0.009%	0.000%	0.000%
S	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
S1	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
S2	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
— MHC Class II, EL rank ≤ 2% —																
E	99.667%	100.000%	0.207%	0.527%	99.933%	100.000%	0.302%	0.308%	99.467%	100.000%	0.174%	0.512%	99.602%	100.000%	0.145%	0.760%
FP	79.739%	100.000%	0.052%	0.723%	82.602%	100.000%	0.026%	0.365%	72.837%	100.000%	0.029%	0.561%	83.779%	100.000%	0.100%	1.245%
M	1.370%	90.327%	0.009%	0.019%	1.614%	89.267%	0.006%	0.009%	0.987%	95.383%	0.006%	0.008%	1.509%	86.331%	0.016%	0.039%
RBD	0.000%	73.558%	0.000%	0.000%	0.000%	76.126%	0.000%	0.000%	0.000%	72.454%	0.000%	0.000%	0.000%	72.095%	0.000%	0.000%
NTD	0.000%	47.560%	0.000%	0.000%	0.000%	61.940%	0.000%	0.000%	0.000%	36.596%	0.000%	0.000%	0.000%	44.143%	0.000%	0.000%
N	0.000%	25.495%	0.000%	0.000%	0.000%	44.091%	0.000%	0.000%	0.000%	12.893%	0.000%	0.000%	0.000%	19.499%	0.000%	0.000%
S	0.000%	0.043%	0.000%	0.000%	0.000%	0.099%	0.000%	0.000%	0.000%	0.006%	0.000%	0.000%	0.000%	0.023%	0.000%	0.000%
S1	0.000%	6.383%	0.000%	0.000%	0.000%	11.198%	0.000%	0.000%	0.000%	3.505%	0.000%	0.000%	0.000%	4.446%	0.000%	0.000%
S2	0.000%	10.522%	0.000%	0.000%	0.000%	18.725%	0.000%	0.000%	0.000%	8.328%	0.000%	0.000%	0.000%	4.513%	0.000%	0.000%
Insufficient coverage (num. peptide-HLA hits ≤ 20)																
— MHC Class I, EL rank ≤ 0.5% —																
FP	100.000%	100.000%	2.413%	2.898%	100.000%	100.000%	4.132%	5.013%	100.000%	100.000%	1.318%	1.786%	100.000%	100.000%	1.789%	1.896%
E	99.910%	100.000%	1.000%	0.660%	99.737%	100.000%	2.293%	0.777%	99.992%	100.000%	0.259%	0.498%	99.999%	100.000%	0.449%	0.706%
RBD	30.340%	100.000%	0.156%	0.183%	20.896%	99.999%	0.099%	0.063%	36.584%	100.000%	0.079%	0.178%	33.541%	100.000%	0.289%	0.307%
M	5.855%	84.753%	0.083%	0.011%	6.576%	83.200%	0.006%	0.002%	4.381%	80.926%	0.039%	0.011%	6.609%	90.134%	0.204%	0.021%
N	3.494%	97.957%	0.000%	0.026%	6.024%	98.429%	0.000%	0.021%	1.704%	96.458%	0.000%	0.035%	2.753%	98.984%	0.000%	0.022%
NTD	3.199%	68.625%	0.026%	0.022%	5.177%	69.526%	0.023%	0.023%	1.581%	60.434%	0.001%	0.023%	2.840%	75.916%	0.055%	0.019%
S1	0.025%	1.545%	0.000%	0.000%	0.001%	2.296%	0.000%	0.000%	0.022%	0.766%	0.000%	0.000%	0.052%	1.572%	0.000%	0.000%
S2	0.011%	8.674%	0.000%	0.000%	0.000%	2.572%	0.000%	0.000%	0.016%	11.724%	0.000%	0.000%	0.016%	11.726%	0.000%	0.000%
S	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
— MHC Class II, EL rank ≤ 2% —																
E	100.000%	100.000%	39.626%	70.239%	100.000%	100.000%	53.265%	77.060%	100.000%	100.000%	29.533%	66.596%	100.000%	100.000%	36.079%	67.060%
FP	99.696%	100.000%	21.719%	56.706%	99.597%	100.000%	27.033%	60.829%	99.801%	100.000%	14.931%	54.146%	99.690%	100.000%	23.193%	55.143%
M	4.479%	100.000%	1.392%	2.150%	5.911%	100.000%	1.851%	2.596%	4.679%	100.000%	0.941%	2.012%	2.847%	100.000%	1.383%	1.841%
RBD	0.234%	100.000%	0.071%	0.089%	0.503%	100.000%	0.127%	0.194%	0.111%	100.000%	0.043%	0.039%	0.087%	100.000%	0.043%	0.036%
NTD	0.007%	100.000%	0.002%	0.003%	0.000%	100.000%	0.000%	0.000%	0.013%	100.000%	0.006%	0.008%	0.008%	100.000%	0.000%	0.000%
S1	0.000%	94.225%	0.000%	0.000%	0.000%	97.555%	0.000%	0.000%	0.000%	94.082%	0.000%	0.000%	0.000%	91.037%	0.000%	0.000%
S2	0.000%	92.789%	0.000%	0.000%	0.000%	95.722%	0.000%	0.000%	0.000%	92.695%	0.000%	0.000%	0.000%	89.950%	0.000%	0.000%
N	0.000%	100.000%	0.000%	0.000%	0.000%	100.000%	0.000%	0.000%	0.000%	100.000%	0.000%	0.000%	0.000%	100.000%	0.000%	0.000%
S	0.000%	19.778%	0.000%	0.000%	0.000%	36.836%	0.000%	0.000%	0.000%	12.954%	0.000%	0.000%	0.000%	9.543%	0.000%	0.000%

Table S3. Percentage of population that is insufficiently covered by subunit vaccines and the improvement after adding augmented peptides evaluated using eluted ligand (EL) ranking. Augmented peptides and the non-redundant subunit peptide sets are the same as in Table 1. The list is sorted by decreasing insufficient coverage of unaugmented subunits. Here, for evaluation, peptides are considered binders using recommended EL percentile rank strong binder thresholds by NetMHCpan-4.1 (≤ 0.5%, MHC class I) and NetMHCIIpan-4.0 (≤ 2%, MHC class II). The list is sorted by decreasing insufficient coverage of unaugmented subunits.

Table S4. Detailed augmentation designs for optimized peptide sets in Table 1. See `AugmentationPeptides.xlsx`.

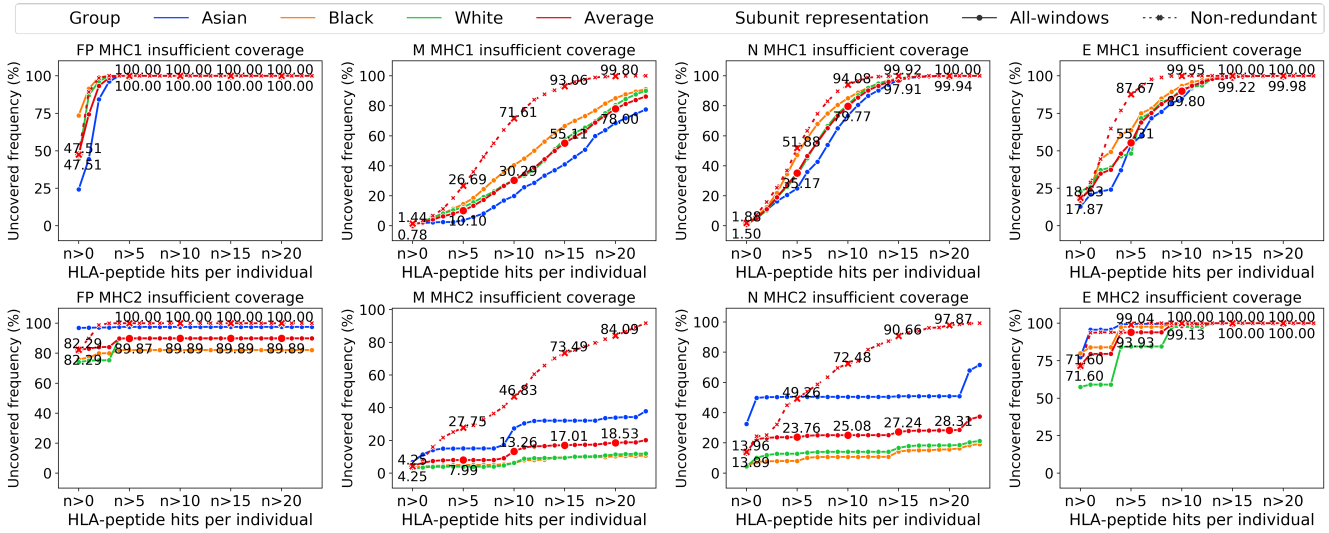

Figure S1. Predicted uncovered percentage of populations as a function of the minimum number of peptide-HLA hits in an individual for E, M, N protein and fusion peptide (FP). Annotated percentages are the average across populations self-reporting as Asian, Black, and White. A redundant sampling of peptides is depicted by solid lines for populations self-reporting as Asian, Black, and White as well their average. A non-redundant sampling of peptides is depicted by dotted lines.