

RESEARCH ARTICLE

# A comprehensive *in silico* analysis for identification of therapeutic epitopes in HPV16, 18, 31 and 45 oncoproteins

Heidar Ali Panahi<sup>1</sup>, Azam Bolhassani<sup>2\*</sup>, Gholamreza Javadi<sup>1</sup>, Zahra Noormohammadi<sup>1</sup>

**1** Department of Biology, School of Basic Sciences, Science and Research Branch, Islamic Azad University, Tehran, Iran, **2** Department of Hepatitis and AIDS, Pasteur Institute of Iran, Tehran, Iran

\* [azam.bolhassani@yahoo.com](mailto:azam.bolhassani@yahoo.com), [A\\_bolhasani@pasteur.ac.ir](mailto:A_bolhasani@pasteur.ac.ir)



## Abstract

Human papillomaviruses (HPVs) are a group of circular double-stranded DNA viruses, showing severe tropism to mucosal tissues. A subset of HPVs, especially HPV16 and 18, are the primary etiological cause for several epithelial cell malignancies, causing about 5.2% of all cancers worldwide. Due to the high prevalence and mortality, HPV-associated cancers have remained as a significant health problem in human society, making an urgent need to develop an effective therapeutic vaccine against them. Achieving this goal is primarily dependent on the identification of efficient tumor-associated epitopes, inducing a robust cell-mediated immune response. Previous information has shown that E5, E6, and E7 early proteins are responsible for the induction and maintenance of HPV-associated cancers. Therefore, the prediction of major histocompatibility complex (MHC) class I T cell epitopes of HPV16, 18, 31 and 45 oncoproteins was targeted in this study. For this purpose, a two-step plan was designed to identify the most probable CD8+ T cell epitopes. In the first step, MHC-I and II binding, MHC-I processing, MHC-I population coverage and MHC-I immunogenicity prediction analyses, and in the second step, MHC-I and II protein-peptide docking, epitope conservation, and cross-reactivity with host antigens' analyses were carried out successively by different tools. Finally, we introduced five probable CD8+ T cell epitopes for each oncoprotein of the HPV genotypes (60 epitopes in total), which obtained better scores by an integrated approach. These predicted epitopes are valuable candidates for *in vitro* or *in vivo* therapeutic vaccine studies against the HPV-associated cancers. Additionally, this two-step plan that each step includes several analyses to find appropriate epitopes provides a rational basis for DNA- or peptide-based vaccine development.

## OPEN ACCESS

**Citation:** Panahi HA, Bolhassani A, Javadi G, Noormohammadi Z (2018) A comprehensive *in silico* analysis for identification of therapeutic epitopes in HPV16, 18, 31 and 45 oncoproteins. PLoS ONE 13(10): e0205933. <https://doi.org/10.1371/journal.pone.0205933>

**Editor:** Craig Meyers, Penn State University School of Medicine, UNITED STATES

**Received:** June 28, 2018

**Accepted:** September 11, 2018

**Published:** October 24, 2018

**Copyright:** © 2018 Panahi et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** The authors received no specific funding for this work.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

HPVs are a large branch of the *Papillomaviridae* family, grouped in different genera (Alpha-, Nu-/Mu-, Beta- and Gamma-papillomaviruses), with more than 200 genotypes [1–4]. The classification of Papillomaviruses (PVs) has been based on L1 gene sequence. They are clinically divided into two groups: low-risk HPVs, like HPV 6 and 11, which cause benign lesions (warts and benign papillomas), and high-risk HPVs (hrHPVs), like HPV16 and 18, which are

carcinogenic to humans [5–7]. The global ratio of all the malignant diseases attributable to HPV infection is estimated to be 5.2% [8–10]. Almost all the cervical carcinomas and a significant part of anogenital and oropharyngeal malignancies are associated with HPV infections [11].

Currently, It is proven that all the oncogenic HPVs are genetically related, Although, they vary greatly in the prevalence and risk of triggering malignant lesions [12, 13]. According to the International Agency for Research on Cancer evaluation (IARC), twelve HPV types (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, and 59) are known as hrHPV. All hrHPVs belong to the alpha genus in *Papillomaviridae* family. Oncogenicity of some types that classified as probably carcinogenic (HPV 68) or possible carcinogenic (HPV 34, 73, 26, 69, 82, 30, 53, 66, 70, 85, 97, 67, 5 and 8) is still needed to be clarified [14].

The relatively simple genome of HPV contains three regions: the upstream regulatory region (URR), the early region, and the late region. The early and late regions encode six early genes (E1, E2, E4, E5, E6, and E7) and two late genes (L1 and L2), respectively. Among these early proteins, E5, E6, and E7 play a pivotal role in the cell transformation. They can interfere in several cell cycle pathways, especially the alteration of EGFR signaling pathways [15, 16], degradation of p53 [17] and degradation of pRB [18], respectively. These effects result in triggering several cascade events, which cause cell transformation, immune evasion and cancer progression [6, 19–26]. E6 and E7 oncoproteins are known as Ideal targets for the immunotherapy of HPV-associated cancers [27–31] since they are consistently expressed in almost all cervical cancer cells, but not in healthy cells, and are essential for the generation and maintenance of malignancy. Additionally, E5, E6, and E7 oncoproteins are structurally different from human cell proteome. Therefore, their side effects on healthy tissues are expected to be negligible [8].

Currently, there are three commercially available HPV prophylactic vaccines [32]. However, none of them showed an effective therapeutic effect on pre-existing HPV infection or its associated cancers [33–35]. Due to the high prevalence and mortality, there is an urgent need to develop an effective therapeutic HPV vaccine for clearance of these infections/cancers. So far, different therapeutic vaccines have been developed [27–31, 36–45]. However, they have induced inadequate immune responses, and thus further studies are needed to develop an effective therapeutic vaccine.

Among various therapeutic vaccines, peptide-based vaccines have appeared as attractive candidates to treat cervical and other HPV-associated cancers. Peptide-based vaccines have some advantages such as easy production and transportation, high selectivity, multivalency capability and epitope accessibility. With the development of genome sequencing techniques, the prediction of potential B and T cell epitopes has opened a promising view to developing peptide-based vaccines against infectious diseases and cancers. Currently, several therapeutic peptide-based HPV vaccines are in different phases of clinical trials [46].

Host genetic polymorphisms influence the immune response to a pathogen in the target population. HLA genes are the most polymorphic genes in the human genome. The vast HLA polymorphism and restriction phenomenon, result in serious problems in vaccine design and population coverage [47–50] because each allele binds to a particular group of peptides. However, many of HLA-I alleles can be classified by their similar peptide-binding properties into groups, covering over 80% of HLA-A and B alleles. Each HLA-I supertype (HLA-A\*01:01, HLA-A\*02:01, HLA-A\*03:01, HLA-A\*24:02, HLA-A\*26:01, HLA-B\*07:02, HLA-B\*08:01, HLA-B\*15:01, HLA-B\*27:05, HLA-B\*39:01, HLA-B\*40:01 and HLA-B\*58:01) represent a group of HLA molecules which bind to a similar set of peptides [51].

The previous studies have shown that the presence of high immunogenic CD8+ cytotoxic T lymphocytes (CTLs) epitopes in vaccine formulation is essential for inducing a robust immune

response. However, the addition of CD4+ T cell epitopes can significantly augment its strength and duration [49, 52, 53]. CD8+ CTLs commonly recognize intracellular-originated peptides presented by MHC-I molecules. They accommodate peptides with 8–11 residues; the ideal length is 9 residues. While CD4+ Helper T Lymphocytes (HTLs) commonly recognize extra-cellular-originated peptides presented by MHC-II molecules. They accommodate peptides with 10–30 residues or even more; the ideal length is 12–16. The strength of the interaction between a T cell receptor and a peptide-MHC complex (pMHC), depends on the presented peptide and the MHC structure [49, 54]. The binding of a peptide to MHC-I molecule is the most selective stage in the way of peptide presentation [55].

Bioinformatics tools can predict the potential immunogenic epitopes from thousands of epitopes in a short time [56]. Generally, the algorithms of these tools range from ones programmed to determine peptide- MHC molecule binding data to those based on structural similarity, molecular modeling, and molecular docking [57]. Peptides that bind to a specific MHC molecule have sequence similarity. Therefore, peptide sequence patterns have been used to predict their binding to MHC molecules [58]. In recent years, the accuracy of these methods has increased strikingly, and more than 90% of natural epitopes have been recognized at a high specificity of 98% [59]. This improvement in performance was achieved by the expanding experimental binding data, available in the *immune epitope database (IEDB) and analysis resource* (<http://www.iedb.org/>), and by the improvement of machine-learning algorithms [60].

Regarding the fundamental importance of epitope prediction in vaccine development, we investigated the best potential CD8+ T cell epitopes from the E5, E6, and E7 oncoproteins of four prevalent hrHPV genotypes (16, 18, 31 and 45) in the world and Iran [61], as shown in Fig 1.

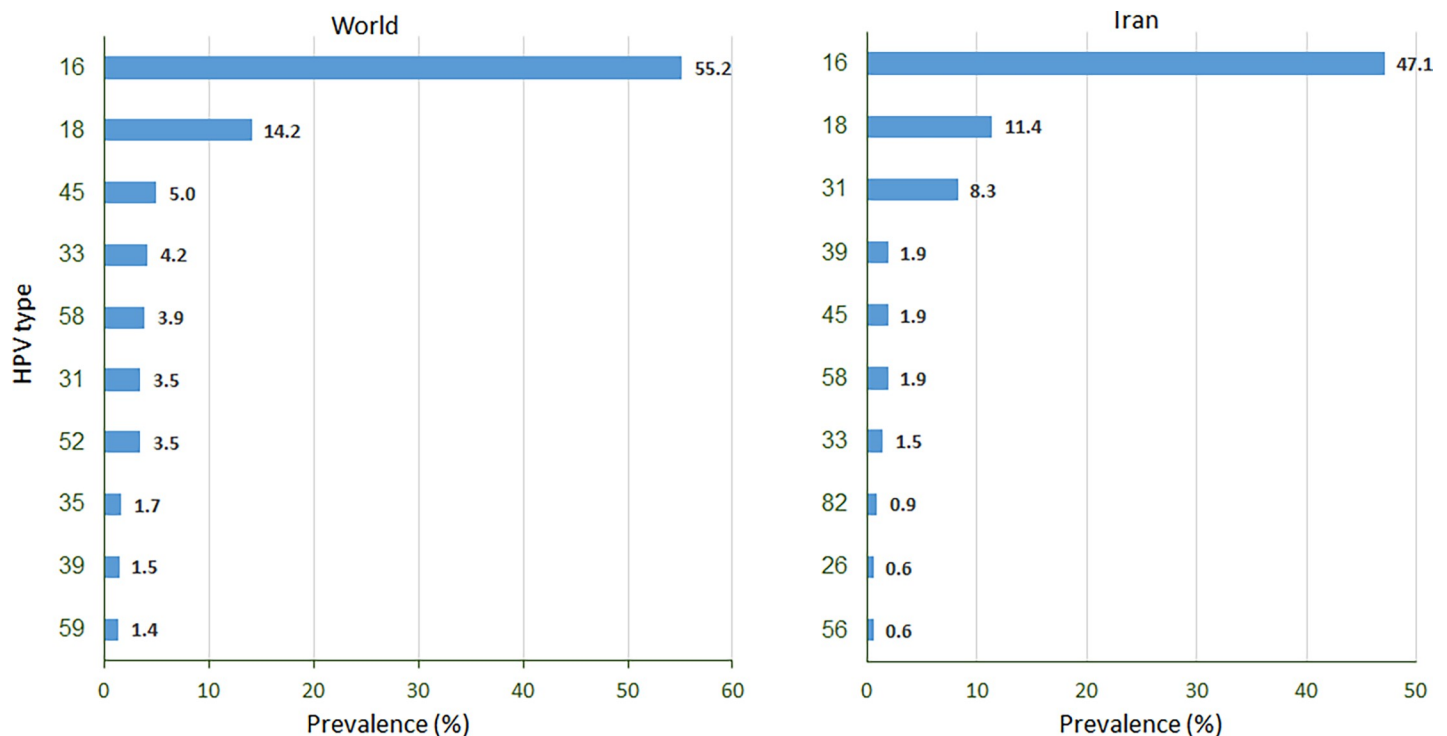


Fig 1. The most prevalent oncogenic HPV types among women with cervical cancer in the world and Iran, 2017 (<http://www.hpvcentre.net/datastatistics.php>).

<https://doi.org/10.1371/journal.pone.0205933.g001>

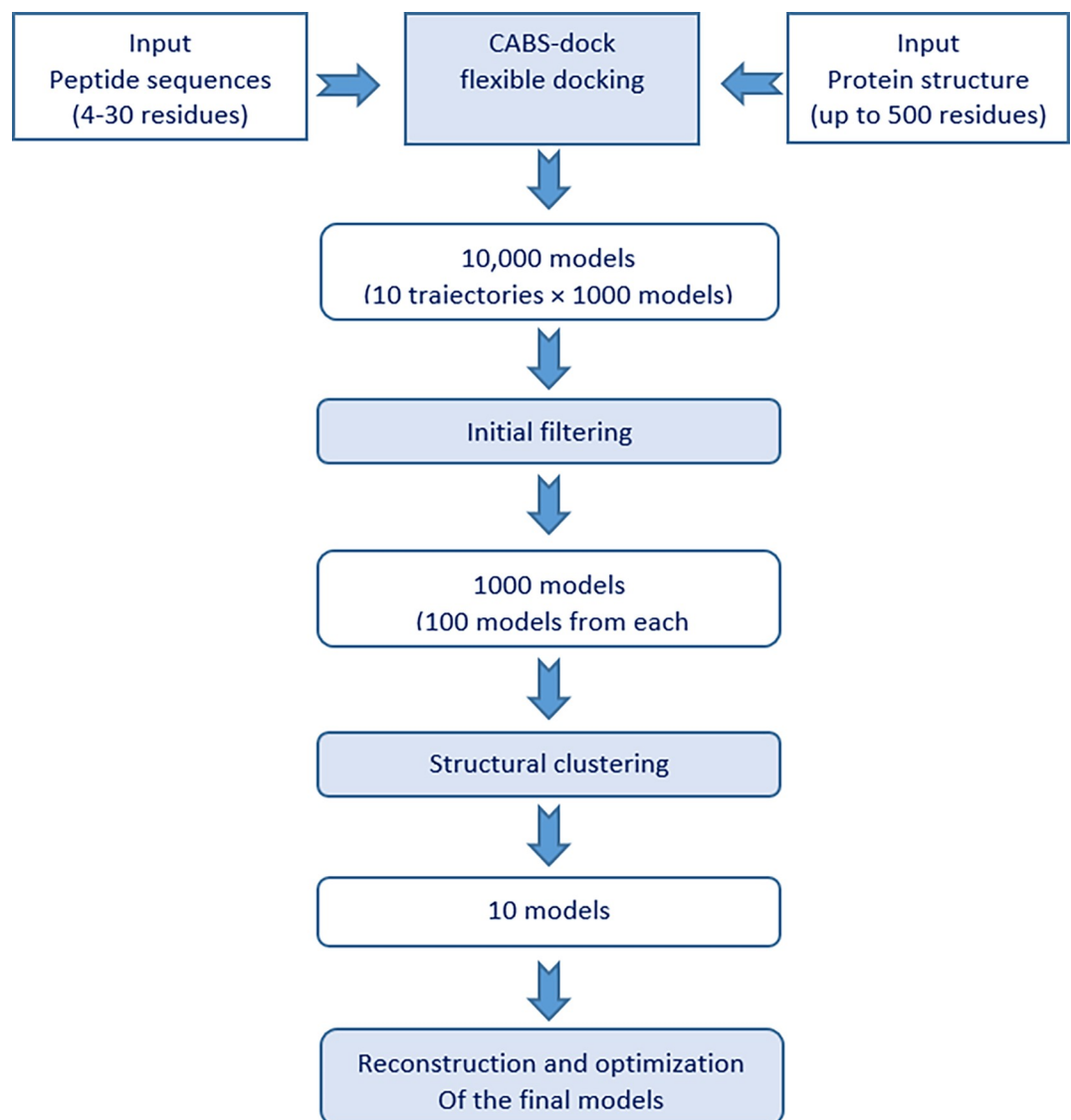
## Materials and methods

### Plan of the study

A two-step plan was designed to identify the most probable CD8+ T cell epitopes (Fig 2). For the first step, MHC-I and II binding, MHC-I processing, MHC-I population coverage and MHC-I immunogenicity prediction analyses, and for the second step, MHC-I and II protein-peptide docking, epitope conservation, and cross-reactivity with host antigens analyses were considered. The second step analyses were performed only for the selected peptides in the first step.

### Protein sequences

In Jan 2018, in order of priority, the RefSeq, reviewed or unreviewed sequences of hrHPV oncoproteins (E5, E6, and E7) were retrieved from the National Center for Biotechnology Information database (NCBI) (<http://www.ncbi.nlm.nih.gov/>) and UniProtKB/Swiss-Prot



**Fig 2. The flow chart of the study:** It represents the two-step epitope selection plan implemented to identify the most probable epitopes of hrHPV oncoproteins.

<https://doi.org/10.1371/journal.pone.0205933.g002>

database (<http://www.uniprot.org/>). The isoform sequences of HPV16, 18, 31, and 45 oncoproteins were retrieved from HPV T cell Antigen Database (<http://cvc.dfci.harvard.edu/hpv/HTML/search.php>). All the sequences are accessible in supporting information (S1 File).

### MHC-I binding prediction

Binding of epitopes to MHC-I molecules is an essential step for antigen presentation to CTLs. Herein, it was predicted by four online servers, as illustrated in Table 1. The HLA supertypes and frequently occurring HLA-I alleles provided by the servers were included in the analysis. However, when an allele (e.g., HLA-B\*14:02) was not provided, but its allele group (i.e., HLA-B\*14) was available, we used the allele group instead of the allele. The used human and mouse alleles, or allele groups are provided in supporting information (S1 Table).

### IEDB MHC-I binding prediction

Currently, eight prediction methods are available in the IEDB MHC-I binding prediction tool, i.e., *IEDB recommended* [62], *Consensus* [69], *NetMHCpan3* [59, 70], *artificial neural network* (ANN) [71, 72], *SMM with a peptide-MHC binding energy covariance matrix (SMMPMBEC)* [73], *stabilized matrix method (SMM)* [74], *CombLib\_Sidney2008* [75], *PickPocket* [76], *netMHCcons* [77] and *netMHCstabpan* [78]. The *IEDB-recommended* and *consensus* are not Independent methods; they use ANN, SMM and *CombLib\_Sidney2008* methods to generate a representative index for each predicted pMHC; The median of percentile ranks (PRs) or binding scores obtained from the used methods is reported as a representative PR or *consensus score* in the *IEDB-recommended* or *consensus* method respectively. The PR is calculated by comparing the *half maximal inhibitory concentration* (IC<sub>50</sub>) of subjected peptide against a group of random peptides from Swiss-Prot database. The IC<sub>50</sub> value, expressed as nanomolar, shows binding affinity. The lower IC<sub>50</sub> or PR means higher binding affinity. As a rough guideline, peptides with IC<sub>50</sub> values <50<sub>nM</sub> are considered as high affinity, 50-500<sub>nM</sub> intermediate affinity and more than 500-5000<sub>nM</sub> low affinity. No known T cell epitope has got an IC<sub>50</sub> value >5000<sub>nM</sub> to date [60].

In this study, *IEDB recommended* method was used. The outputs for each pMHC in this method consisted of a median PR, a method-specific IC<sub>50</sub>, and a method-specific PR. Predictions were made against 76 frequently occurring human MHC-I alleles (including 12 HLA supertypes) and 6 MHC-I mouse alleles. Epitope length was set on 8, 9, 10, and 11mer. Peptides with median PR <2.0 are applied for the analysis.

**NetMHCpan4 MHC-I binding prediction.** NetMHCpan4 server predicts binding of peptides to the known MHC molecules using ANNs method. It is trained on a combination of naturally eluted ligands (55 human and mouse MHC-I alleles) and binding affinity data (172 MHC molecules from human, mouse, cattle, primates, and swine). Besides, the user can perform a prediction against any custom MHC-I molecule by uploading its full-length sequence [66].

In this study, predictions were performed for 8, 9, 10, and 11mer peptides against 76 frequently occurring human MHC-I alleles and 8 MHC-I mouse alleles. PR thresholds for strong

**Table 1. Predictor servers used for MHC-I epitope binding prediction.**

Server name	Link	Prediction method	References
IEDB MHC-I binding (IEDB recommended)	<a href="http://tools.iedb.org/mhci/">http://tools.iedb.org/mhci/</a>	ANN, SMM and CombLib_Sidney2008	[60, 62–65]
NetMHCpan4	<a href="http://www.cbs.dtu.dk/services/NetMHCpan/">http://www.cbs.dtu.dk/services/NetMHCpan/</a>	ANN	[66]
Rankpep	<a href="http://imed.med.ucm.es/Tools/rankpep.html">http://imed.med.ucm.es/Tools/rankpep.html</a>	PSSM	[67, 68]
SYFPEITHI	<a href="http://www.syfpeithi.de/0-Home.htm">http://www.syfpeithi.de/0-Home.htm</a>	Published motifs	[54]

<https://doi.org/10.1371/journal.pone.0205933.t001>

and weak binders were set on 0.5 and 2.0, respectively. Peptides with PR <2.0 were applied for the analysis.

**Rankpep MHC-I binding prediction.** Rankpep predicts binder peptides of a given protein sequence or sequence alignments to MHC-I and II molecules. The algorithm of Rankpep based on the comparison of sequence similarities, using *position-specific scoring matrices* (PSSMs) method. It employs [profiles](#) of a group of aligned peptides recognized to bind to a specific MHC molecule and creates a consensus sequence by determining the most common residue for each position. Then, it allocates an optimal score to the consensus sequence, compares the score of the subjected peptide with the optimal score, and gives the peptide a percentile optimal value for comparison. Finally, it highlights strong binders in red [67, 68].

Herein, the prediction was made against 31 frequently occurring HLA-I and 7 H2-I alleles. The server did not provide all common lengths of epitopes for all the MHC alleles. Thus, the used alleles and their provided epitope lengths are shown together, as given in supporting information (S1 Table).

**SYFPEITHI MHC-I binding prediction.** SYFPEITHI (<http://www.syfpeithi.de/0-Home.htm/>) is a database of over 7000 published and verified peptide sequences of human, mouse, and other organisms, known as natural binders of MHC-I and II molecules. When SYFPEITHI analyzes a peptide for binding prediction against a specific MHC-I allele, its scoring system evaluates every residue of the query and gives it an arbitrary value between 1 and 15, according to whether it is an anchor, auxiliary anchor, or preferred residue. It allocates the value 1 to those residues which slightly preferred in that particular position, 15 to the Ideal anchor residues, and -1 to -3 to those residues which exhibit an adverse effect on the binding ability. The sum of these values is the score of the peptide. The maximal score could vary between different MHC alleles [54, 79].

Herein, the prediction was made against 26 frequently occurring HLA-I alleles and 5 H2-I alleles. Epitope length was set on 8, 9, 10, and 11mer. Every predicted pMHC which got a score less than 70% of the reference sequence score was excluded from the analysis. The allele-specific reference sequence was selected from Rankpep's *consensus sequence* [68], or our SYFPEITHI predicted epitopes, whichever got the highest score in SYFPEITHI server. The reference sequences, their sources, and their scores are given in supporting information (S2 Table).

## MHC-II binding prediction

Recognition of high immunogenic CD8+ T cell epitopes was the primary aim of this study. Therefore, all predictions were primarily made against epitopes with 8–11 residue length. However, it was valuable to determine that which 9mer MHC-I epitope is the core peptide of the MHC-II epitope(s) too. The core peptide lies on the MHC-II molecule grooves, and play the central role in constructing pMHC. With this strategy, the short minimal predicted epitopes could be used in designing of synthetic long peptides (SLPs), resulting in peptide loading to both MHC-I and II molecules.

**IEDB MHC-II binding prediction.** In this study, the MHC-II binding prediction was made by IEDB MHC-II binding predictor (<http://tools.iedb.org/mhcii/>) [60, 63, 64]. IEDB possess seven prediction methods for MHC-II binding prediction: *IEDB-recommended*, *consensus* [63], *NetMHCIIpan* [80], *NN-align* [81], *SMM-align* [82], *Combinatorial Libraries* [75] and *Sturniolo's method* [83]. Herein, the *IEDB-recommended* method was used, and all peptides with PR<2.0 were selected for the analysis.

The prediction was made against 35 human alleles (IEDB reference set) and three mouse alleles, given in supporting information (S3 Table). The server has fundamentally set the



epitope length on 15mer. Each *IEDB-recommended* method participated in the prediction process offered a core Peptide (9mer) for each predicted epitope (15mer). We associated the 9mer MHC-II core peptides with the 9mer MHC-I predicted epitopes to determine that which MHC-I epitope is the core peptide of the MHC-II epitope(s) too.

### MHC-I processing prediction

MHC-I T cell epitope processing predictions of E5, E6, and E7 oncoproteins are made by the *IEDB combined predictor* (<http://tools.iedb.org/processing/>). This tool combines predictors of three main steps of MHC-I antigen presentation pathway (proteasomal processing, transporter associated with antigen processing (TAP) transport, and MHC-I binding) and calculates a total processing score for each predicted epitope. It allows the user to choose a method from *ANN*, *SMM*, *SMMPMBEC*, *Comblib\_Sidney2008*, *NetMHCpan*, *NetMHCcons* and *PickPocket* methods for the binding prediction. In the current update (2018), the IEDB team has changed the choice of the recommended prediction method for the processing tool to be *NetMHCpan 3.0* rather than a *consensus*, since the processing tools requiring an  $IC_{50}$  value, which the *consensus* method does not provide. Furthermore, *NetMHCpan 3.0* has provided all MHC alleles and has performed the predictions very well in recent comparisons [65].

There are two types of proteasomes, the housekeeping types which are expressed instinctively, and immuno types which are provoked by IFN- $\gamma$  secretion. The immunoproteasomes are believed to improve the efficiency of antigen presentation [62, 65]. In this study, the immunoproteasome option was selected.

The program outputs for every predicted epitope consisted of proteasome score, TAP score, MHC score, processing score (proteasome + TAP score), total score (Proteasome + TAP + MHC score), and MHC-I  $IC_{50}$ . The TAP scoring system calculates a  $-\log(IC_{50})$  value for the binding of a peptide (or N-terminal of its precursors) to the TAP molecules. The higher TAP score, the higher transport rate. [62, 65, 84].

Herein, the analysis was made against the human and mouse MHC-I alleles used later in the IEDB binding prediction, with the *IEDB-recommended* method and other default settings of the program. Epitopes with  $IC_{50} < 1000$   $_{nM}$  for HLA-I alleles and  $< 5000$   $_{nM}$  for H2-I alleles were included in the analysis.

### MHC-I immunogenicity prediction

Several factors could clarify the difference between epitope and non-epitope peptides; An essential factor is epitope immunogenicity, *i.e.*, it could be recognized by T cells. Some amino acids, particularly those with large and aromatic side chains (especially tryptophan, phenylalanine, and Isoleucine), are associated with immunogenicity. Moreover, the positions P4–6 of a peptide are more critical for immunogenicity [85].

In this study, the MHC-I immunogenicity of all predicted epitopes was determined by the IEDB web server (<http://tools.iedb.org/immunogenicity/>) [85]. This tool uses the properties of amino acids and their locations to predict the immunogenicity of a pMHC. The default option was selected to specify which positions of the query peptide to be masked from the analysis, because it masked positions which are also suggested for the most frequent human MHC-I allele, HLA-A\*02:01.

### Population coverage prediction

IEDB population coverage prediction tool (<http://tools.iedb.org/population/>) [86] is used to predict the HLA-I population coverage of all 8–11mer predicted epitopes in the first step. This tool can accept a target population by two query levels: 1) area-country-ethnicity and 2) ethnicity alone. It can integrate allele frequency information retrieved from the *Allele Frequency*

*Net Database* (AFND) (<http://www.allelefreqencies.net/default.asp>) [87]. IEDB also accepts custom populations with allele frequencies defined by users. Since, HLA-I and HLA-II T cell epitopes elicit immune responses from two different T cell populations (CTL and HTL, respectively), the server provided three different population coverage modes: 1) HLA-I lonely, 2) HLA-II lonely, and 3) HLA-I and HLA-II together.

Herein, the MHC-I promiscuous predicted epitopes and their binding HLA-I alleles ( $IC_{50} < 500_{nM}$  or  $PR < 2.0$ ) were entered as inputs for the analysis against the world population.

### Molecular docking analysis

The primary aim of molecular docking is the prediction of the binding site of a ligand at a protein receptor surface, and then docking and modeling the ligand into the recognized site. In this study, the binding ability of the first step selected peptides to human and mouse MHC molecules, was analyzed by CABS-dock (<http://biocomp.chem.uw.edu.pl/CABSdock/>) server. The server uses a multistage procedure that involves multiple programs, with the C $\alpha$ -C $\beta$ -side chain (CABS) model at its heart. The detailed information about these stages is given in supporting information (S2 File) [88, 89]. Also, Fig 3 shows the pipeline of CABS-dock protocol [88].

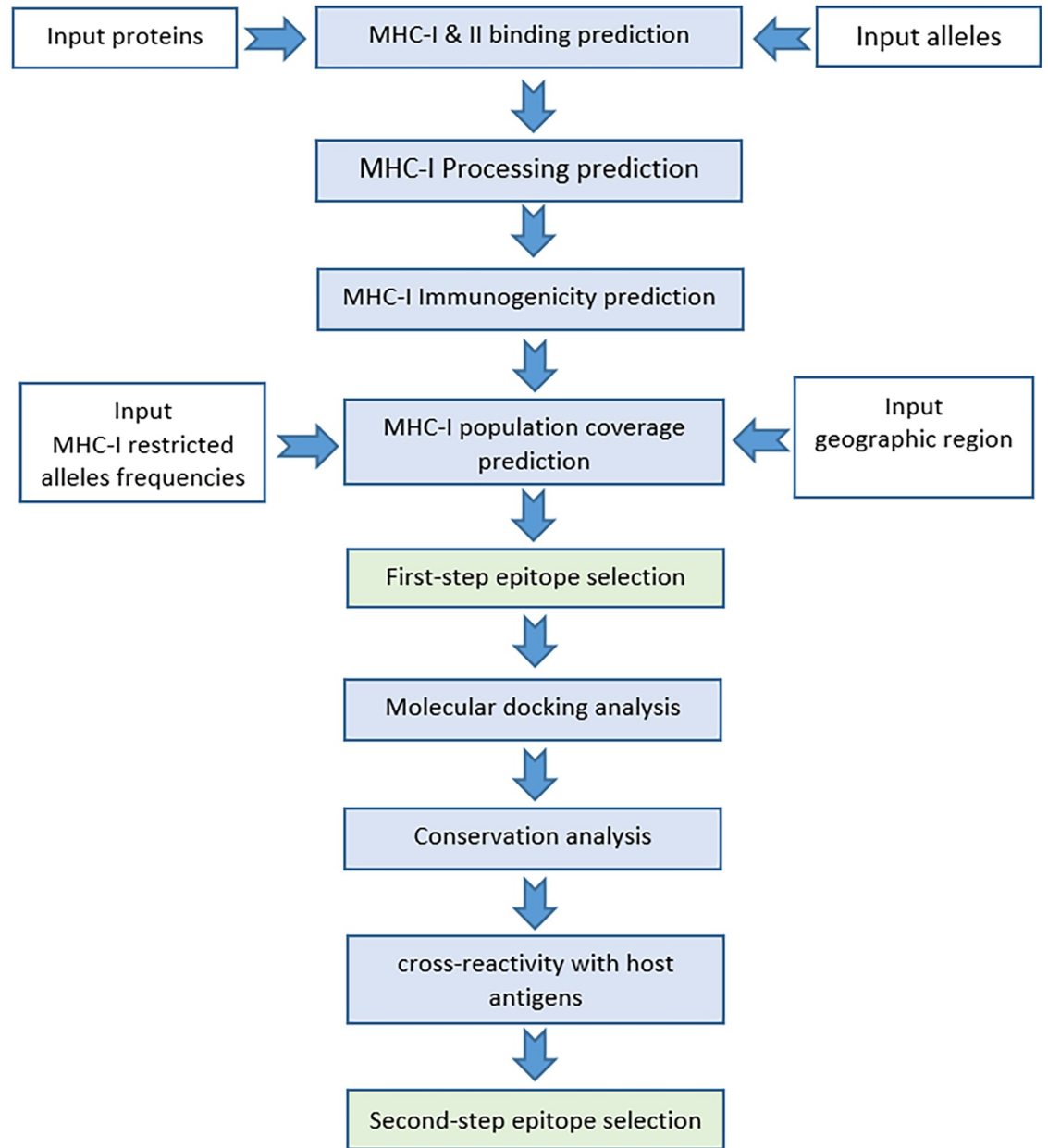
CABS-dock gets the 3D structure of the receptor and the sequence of the peptide as obligatory inputs. Furthermore, there are some non-obligatory inputs as recommendations which could improve outputs. In this study, duplicate dockings for each peptide (6240 dockings in total) were done against the most significant human/mouse MHC-I and II molecules which had at least one well-structured protein data bank (PDB) file in the RCSB Protein Data Bank (<https://www.rcsb.org/>), as shown in Table 2. These PDB files are in the complex with their peptidic ligand and some X-ray crystallography solution molecules (heteroatoms). Thus, these excess molecules, as well as redundant MHC molecules were removed before executing docking process. Since, the binding site of epitopes on the MHC molecules was well-known previously, the unlikely regions to bind masked before the analysis.

CABS-dock returns ten representative models (medoids) as the best-simulated models and ranks them by cluster density (CD). Cluster density is equal to the number of elements in a cluster divided by their average ligand *root mean square deviation* (RMSD). The higher CD value implies greater accuracy. Ligand RMSD value shows the differentiation measure between cluster elements. As a guideline;  $RMSD < 3.0 \text{ \AA}$  means high accuracy;  $RMSD \geq 3.0$  and  $\leq 5.5 \text{ \AA}$  means medium accuracy and  $RMSD > 5.5 \text{ \AA}$  means low accuracy [88]. Herein, the RMSD and CD of the best-simulated models were selected for the analysis. The best model, which has the highest CD value, is not necessarily the top-ranked model, because, in some cases, peptides were not attached to their binding site properly. Thus, these malformed models were excluded from the analysis. It is important to note that, due to the different frequency of MHC alleles in human populations, the equal CD value of different MHC alleles, don't have equal value regarding population coverage. Thus, to involve the effect of population coverage, the CD value of every model was multiplied by its allele population coverage (divided by hundreds for more facility) to obtain a weighted index. Then, the sum of all HLA-I or II weighted indexes of each peptide was calculated to get a total docking score (TDS), used as a score to compare the candidate peptides. It is the first time that the TDS has been formulated and used for this purpose. This formula is also applicable to the similar docking scores obtained from other servers.

### Epitope conservancy analysis

The use of highly conserved epitopes in a vaccine formulation reduces the risk of tumor immune escape and provides broader protection against different virus strains or genotypes. Thus, the conserved areas are preferred to use in therapeutic vaccines, if they are appropriate





**Fig 3. The pipeline of CABS-dock protocol:** The fully automated CABS-dock procedure contains four main stages, shown in the blue boxes.

<https://doi.org/10.1371/journal.pone.0205933.g003>

epitopes. Herein, the epitope conservancy analyses for the first step selected peptides were done in three levels:

1. Inter-isoform conservancy: the percent of conservancy between all isoforms of each E5, E6, or E7 oncoprotein.
2. Inter-type conservancy: the percent of conservancy between HPV16 and 31 (alpha-Papillomavirus 9), as well as between HPV18 and 45 (alpha- Papillomavirus 7).
3. Inter-hrHPV conservancy: complete (100%) conservancy between all hrHPVs (HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, and 59).

**Table 2. MHC alleles used for molecular docking analysis against the selected peptides in the first step.**

Human MHC-I		Human MHC-II		Mouse MHC-I		Mouse MHC-II	
Allele	PDB ID	Allele	PDB ID	Allele	PDB ID	Allele	PDB ID
HLA-A*02:01	4UQ3	HLA-DRB1*01:01	4AH2	H-2-Db	1JUF	H-2-IAb	4P23
HLA-A*24:02	5HGA	HLA-DRB1*03:01	2Q6W	H-2-Dd	5IVX	H-2-IAc	2IAD
HLA-A*01:01	4NQV	HLA-DRB1*04:01	5LAX	H-2-Kb	4PV9	H-2-Ag7	1ESO
HLA-A*03:01	3RL2	HLA-DRB1*11:01	6CPL	H-2-Kd	5GSV	H-2-IEk	1FNG
HLA-A*11:01	1X7Q	HLA-DRB1*15:01	5V4M	H-2-Kk	1ZTV		
HLA-B*07:02	5EO1	HLA-DRB5*01:01	1FV1	H-2-Ld	1LDP		
HLA-B*08:01	3SPV						
HLA-B*27:05	1OGT						
HLA-B*35:01	3LKN						
HLA-B*27:05	4UQ3						

<https://doi.org/10.1371/journal.pone.0205933.t002>

The selected peptides in the first step were analyzed to find inter-strains and inter-types conservancy percentages by IEDB tool, *conservation across antigens*, (<http://tools.iedb.org/conservancy/>). The inter-hrHPVs conservancy analysis was done by the IEDB and ExpASY ClustalW servers (<https://embnet.vital-it.ch/software/ClustalW.html>).

### Cross-reactivity with host antigens

Cross-reactivity with host antigens can cause adverse immune responses. Therefore, the selected peptides in the first step were checked for similarities with the mouse and human proteomes by the NCBI BLASTp tool (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

### Results

Regarding the studies, different peptides usually get different scores/ranks in different analyses. This inconsistency indicates that these results needed to be analyzed with an integrated approach. Indeed, integrated approach is more practical and efficient in such conditions in comparison with *analysis by analysis filtering* approach, in which those epitopes are chosen for the next analysis that have gotten an acceptable score in the previous analysis. Herein, the integrated approach was applied in both steps of epitope selection.

Since the ultimate goal of the discovery of therapeutic epitopes is to use them in human vaccines, only the scores/ranks of human alleles were used to rank epitopes in some studies. However, investigators usually test therapeutic vaccines on mouse species in preclinical trials, thus in the current study, the binding status of the predicted epitopes to mouse MHC-I alleles was also studied by several binding predictors and molecular docking, as well.

As stated above, CTL-mediated responses play a crucial role in killing the malignant cells. Besides, the binding of epitopes to MHC-I molecules is the most selective step for antigen presentation to CTLs. Therefore, in the first step, the selection was made primarily by the comparison of obtained MHC-I binding, processing and immunogenicity scores/ranks, and population coverage percentages. However, the MHC-II binding ranks were actually of secondary importance to the selection process as an added advantage. Additionally, the population coverage has a dual application. First, it determines the coverage of a given peptide in the target population. Second, it is the best index for summarizing and evaluating of the HLA-I binding predictions too, since it is calculated from the results of HLA-I binding prediction analyses.

**Table 3. The predicted epitopes from E5 oncoproteins in the first step selection.**

Protein	Epitope	Location (length)	Immunogenicity Score	Population Coverage	Proteasomal Cleavage score	TAP score	Processing score
HPV16-E5	FLIHTHARF	72–80 (9)	0.21	97.75%	1.31	1.12	2.43
	SAFRCFIVY	55–63 (9)	0.28	70.69%	1.55	1.43	2.98
	STYTSLLIL	37–45 (9)	0.06	99.70%	1.67	0.53	2.20
	YTSLLILVL	39–47 (9)	0.20	99.93%	1.57	0.40	1.97
	FLLCFCVLL	15–23 (9)	0.03	62.23%	1.47	0.39	1.86
	YIIFVYIPL	63–71 (9)	0.30	76.67%	1.18	0.51	1.70
	LSVSTYTSL	34–42 (9)	-0.18	93.47%	1.52	0.49	2.01
	FVYIPLFLI	66–74 (9)	0.20	89.59%	1.14	0.34	1.48
	FIVYIIFVY	60–68 (9)	0.39	53.36%	1.36	1.32	2.68
	IIFVYIPLF	64–72 (9)	0.19	99.99%	1.28	1.19	2.47
HPV18-E5	ATAFTVYVF	47–55 (9)	0.23	100.0%	1.50	1.10	2.59
	CAYAWVLVF	29–37 (9)	0.31	78.10%	1.45	1.20	2.65
	SPATAFTVY	45–53 (9)	0.26	82.84%	1.46	1.17	2.63
	MCAYAWVLVF	28–37 (10)	0.32	96.23%	1.45	1.13	2.58
	YAWVLVYVF	31–39 (9)	0.26	62.06%	1.29	1.43	2.71
	LHIHAILSL	64–72 (9)	0.13	84.77%	1.51	0.47	1.98
	FTVYVFCFL	50–58 (9)	0.18	79.20%	1.36	0.47	1.83
	LLLHIHAIL	62–70 (9)	0.30	57.57%	1.57	0.45	2.02
	TSPATAFTVY	44–53 (10)	0.28	90.18%	1.46	1.28	2.74
	MLLHIHAI	61–69 (9)	0.19	85.90%	1.05	0.21	1.25
HPV31-E5	FVIHTHASF	72–80 (9)	0.08	100.0%	1.11	1.17	2.28
	LSVSVYATL	34–42 (9)	-0.06	100.0%	1.48	0.49	1.97
	SVYATLLLL	37–45 (9)	0.05	99.80%	1.28	0.60	1.88
	VVFIYIPLF	64–72 (9)	0.28	99.59%	1.41	1.21	2.62
	VSVYATLLL	36–44 (9)	0.07	100.0%	1.43	0.52	1.95
	VYATLLLLI	38–46 (9)	0.02	99.96%	1.29	0.40	1.70
	YVVFYIPL	63–71 (9)	0.36	77.99%	1.13	0.51	1.64
	LIHTHARFL	73–81 (9)	0.05	100.0%	1.50	0.47	1.97
	FLLCFCVLL	15–23 (9)	0.03	57.05%	1.47	0.39	1.86
	FIYIPLFVI	66–74 (9)	0.23	70.90%	1.35	0.35	1.70
HPV45-E5	CAFAWLLVF	29–37 (9)	0.30	69.65%	1.35	1.17	2.53
	YVCAFAWLL	27–35 (9)	0.35	52.04%	1.45	0.49	1.94
	VYVCAFAWLL	26–35 (10)	0.33	99.98%	1.45	0.62	2.06
	LHMHALHTL	64–72 (9)	0.05	99.56%	1.58	0.47	2.05
	VITSPLTAF	42–50 (9)	-0.12	98.75%	1.41	1.19	2.59
	FLLCFSVCL	6–14 (9)	-0.10	56.13%	1.68	0.41	2.09
	FAWLLVFLF	31–39 (9)	0.18	59.90%	1.12	1.22	2.34
	VYVCAFAWL	26–34 (9)	0.27	99.86%	1.16	0.62	1.77
	SPLTAFVY	45–53 (9)	0.24	57.26%	1.32	1.17	2.49
	MFVLHMHAL	61–69 (9)	-0.08	92.98%	1.64	0.48	2.12

<https://doi.org/10.1371/journal.pone.0205933.t003>

In the first step, ten peptides (Tables 3–5) from each HPV genotype oncoprotein (120 peptides in total), which got better results in the first step analyses were selected for the second step analyses, including protein-peptide molecular docking, epitope conservation, and cross-reactivity with host antigens. The individual detailed results of the MHC-I and II binding ([S3 File](#)), MHC-I immunogenicity ([S4 File](#)) and MHC-I population coverage ([S5 File](#)) predictions,

Table 4. The predicted epitopes from E6 oncoproteins in the first step selection.

Protein	Epitope	Location (length)	Immunogenicity Score	Population Coverage	Proteasomal Cleavage score	TAP score	Processing score
HPV16-E6	FAFRDLCIVY	52–61 (10)	0.19	83.26%	1.94	1.24	3.18
	IVYRDGNPY	59–67 (9)	0.09	86.66%	1.13	1.42	2.55
	CYSLYGTTL	87–95 (9)	0.02	99.94%	1.58	0.49	2.06
	VYDFAFRDL	49–57 (9)	0.33	99.99%	1.35	0.41	1.75
	KFYSKISEY	75–83 (9)	-0.33	99.57%	1.39	1.49	2.88
	SEYRHYCYSL	81–90 (10)	-0.07	99.96%	1.59	0.47	2.25
	KLPQLCTEL	18–26 (9)	-0.09	99.98%	1.41	0.52	1.93
	ISEYRHYCY	80–88 (9)	0.08	24.94%	1.57	1.28	2.85
	EYRHYCYSL	82–90 (9)	-0.11	99.91%	1.59	0.49	2.08
	YCYSLYGTTL	86–95 (10)	-0.10	99.99%	1.58	0.49	2.06
HPV18-E6	KLPDLCTEL	13–21 (9)	0.05	99.99%	1.65	0.44	2.09
	TVLELTEVF	37–45 (9)	0.23	99.27%	1.43	1.20	2.63
	FAFKDLFVVY	47–56 (10)	0.01	86.98%	1.90	1.23	3.12
	LELTEVFEE	39–47 (9)	0.33	49.80%	1.40	1.07	2.47
	DFYSRIREL	70–78 (9)	0.11	87.51%	1.64	0.44	2.08
	FEFAFKDLF	45–53 (9)	0.01	75.51%	1.15	1.07	2.22
	FAFKDLFVV	47–55 (9)	-0.05	96.60%	1.02	0.09	1.11
	AFKDLFVVY	48–56 (9)	0.11	95.40%	1.90	1.36	3.25
	LQDIEITCVY	25–34 (10)	0.38	51.48%	1.31	1.25	2.56
	SVYGDITLEK	84–92 (9)	0.14	58.01%	0.94	0.31	1.24
HPV31-E6	FAFTDLTIVY	45–54 (10)	0.26	93.1%	1.87	1.24	3.11
	RYSVYGTTL	80–88 (9)	0.07	100.0%	1.48	0.61	2.09
	KVSEFRWYRY	72–81 (10)	0.45	68.2%	1.43	1.40	2.83
	FRWYRYSVY	76–84 (9)	0.00	66.2%	1.45	1.32	2.77
	YRYSVYGTTL	79–88 (10)	-0.04	100.0%	1.48	0.55	2.03
	LSSALEIPY	15–23 (9)	0.18	61.3%	1.00	1.21	2.22
	KVSEFRWYR	72–80 (9)	0.40	99.9%	0.90	0.77	1.67
	AFTDLTIVY	46–54 (9)	0.20	97.6%	1.87	1.34	3.21
	KLHELSSAL	11–19 (9)	-0.17	99.96%	1.39	0.54	1.93
	FAFTDLTIV	45–53 (9)	0.20	95.45%	1.03	0.11	1.13
HPV45-E6	IVYRDCIAY	54–62 (9)	0.16	81.72%	1.55	1.42	2.97
	KLPDLCTEL	13–21 (9)	0.05	99.99%	1.65	0.44	2.09
	YSRIRELRY	72–80 (9)	0.32	85.96%	1.46	1.25	2.71
	ATLERTEVY	37–45 (9)	0.29	97.51%	1.40	1.37	2.79
	FAFKDLCIVY	47–56 (10)	-0.08	83.30%	1.90	1.28	3.15
	DFYSRIREL	70–78 (9)	0.11	98.61%	1.56	0.44	1.99
	YQFAFKDL	45–52 (8)	0.01	80.70%	1.44	0.51	1.95
	NPAEKRRHL	113–121 (9)	0.02	45.86%	1.62	0.33	1.95
	RTEVYQFAF	41–49 (9)	0.08	99.99%	1.49	1.06	2.55
	YSRIRELRY	72–81 (10)	0.33	80.75%	1.30	1.25	2.56

<https://doi.org/10.1371/journal.pone.0205933.t004>

as well as, MHC-I and II molecular docking (S6 File) and epitope conservation (S7 File) analyses are given in supporting information, as 15 Excel files. Indeed, CABS-dock returns ten representative medoids as the best-simulated models and ranks them by cluster density (CD). Cluster density is derived from two factors (the number of elements in a cluster and their average ligand RMSD) that is an advantage for this server.

Table 5. The predicted epitopes from E7 oncoproteins in the first step selection.

Protein	Epitope	Location (length)	Immunogenicity Score	Population Coverage	Proteasomal Cleavage score	TAP score	Processing score
HPV16-E7	RAHYNIVTF	49–57 (9)	0.18	99.98%	1.48	1.18	2.66
	STHVDIRTL	71–79 (9)	0.27	99.35%	1.79	0.42	2.21
	LEDLLMGTL	79–87 (9)	-0.13	51.13%	1.70	0.29	1.98
	LQPETTDLY	15–23 (9)	0.18	81.48%	1.16	1.24	2.40
	TLHEYMLDL	7–15 (9)	-0.05	96.26%	1.17	0.37	1.53
	LLMGTLGIV	82–90 (9)	0.11	40.60%	0.97	0.11	1.08
	DRAHYNIVTF	48–57 (10)	0.22	80.89%	1.48	1.00	2.48
	GQAEPDRAHY	43–52 (10)	0.23	49.47%	1.57	1.30	2.87
	QAEPDRAHY	44–52 (9)	0.14	57.01%	1.57	1.22	2.79
	GTLGIVCPI	85–93 (9)	0.15	42.66%	0.76	0.17	0.93
HPV18-E7	SSADDLRAF	78–86 (9)	0.11	99.69%	1.40	1.07	2.47
	FQQLFNLTL	86–94 (9)	0.07	95.89%	1.71	0.46	2.17
	QLFLNLSF	88–96 (9)	-0.05	88.83%	1.30	1.14	2.44
	TLQDIVLHL	7–15 (9)	0.16	99.98%	1.57	0.47	2.04
	LRAFQQLFL	83–91 (9)	-0.03	62.34%	1.74	0.45	2.20
	RAFQQLFL	84–91 (8)	-0.10	99.98%	1.74	0.62	2.36
	QQLFLNLSF	87–96 (10)	0.03	39.85%	1.30	1.10	2.40
	RAEPQRHTM	53–61 (9)	0.01	99.95%	0.97	0.26	1.22
	SADDLRAF	79–86 (8)	0.10	74.99%	1.40	1.09	2.49
	RAFQQLFLNLT	84–94 (11)	-0.08	64.77%	1.70	0.62	2.32
HPV31-E7	TSNYNIVTF	49–57 (9)	0.17	99.95%	1.51	1.01	2.51
	TLQDYVLDL	7–15 (9)	0.02	99.77%	1.30	0.34	1.64
	GQAEPDTSNY	43–52 (10)	0.02	44.68%	1.48	1.30	2.78
	QAEPDTSNY	44–52 (9)	-0.06	56.83%	1.48	1.22	2.70
	QPEATDLHCY	16–25 (10)	0.12	42.18%	1.47	1.10	2.57
	TPTLQDYVL	5–13 (9)	-0.07	45.76%	1.63	0.19	1.81
	LLMGSGIV	82–90 (9)	0.03	46.52%	1.02	0.10	1.12
	YVLDLQPEA	11–19 (9)	-0.05	79.18%	1.42	-0.23	1.18
	IRILQELLM	76–84 (9)	0.00	62.34%	0.96	0.20	1.16
	VDIRILQEL	74–82 (9)	0.18	27.17%	1.43	0.32	1.75
HPV45-E7	TLQEIVLHL	7–15 (9)	0.24	99.96%	1.45	0.33	1.77
	NELDPVDLL	19–27 (9)	0.06	51.21%	1.44	0.36	1.80
	LQQLFLSTL	87–95 (9)	-0.06	37.63%	1.59	0.44	2.03
	QLFLSTLSF	89–97 (9)	-0.20	99.33%	1.30	1.14	2.44
	SSAEDLRTL	79–87 (9)	0.19	99.03%	1.56	0.39	1.95
	ELDPVDLLCY	20–29 (10)	0.01	57.92%	1.45	1.17	2.63
	LRTLQQLFL	84–92 (9)	-0.16	59.86%	1.80	0.43	2.23
	RETLQEIVL	5–13 (9)	0.12	46.68%	1.76	0.38	2.15
	LHLEPQNEL	13–21 (9)	0.03	74.50%	1.36	0.45	1.81
	QQLFLSTLSF	88–97 (10)	-0.13	37.17%	1.30	1.17	2.46

<https://doi.org/10.1371/journal.pone.0205933.t005>

In the second step, five peptides out of ten selected peptides in the first step (Tables 6–8), which got better results in all analyses of both steps, were selected as the final-predicted epitopes. None of the final predicted epitopes showed more than 90% sequence similarity with mouse and human proteomes.

**Table 6. The predicted epitopes from E5 oncoproteins in the second step selection.**

Protein	Epitope	Docking		Conservancy		
		TDS (HLA-I)	TDS (HLA-II)	Inter-isoform (% identity)	Inter-type* (% identity)	Inter-hrHPV (100% identical)
HPV16-E5	FLIHTHARF	110	74	87.5	77.8	None
	SAFRCFIVY	111	73	79.2	55.6	None
	STYTSILIL	131	62	16.7	44.4	None
	YTSLILVL	112	67	16.7	33.3	None
	FLLCFCVLL	122	67	91.7	100.0	HPV16, 31 and 35
HPV18-E5	ATAFTVYVF	114	80	88.9	55.6	None
	CAYAWVLVF	135	70	100.0	77.8	None
	SPATAFTVY	117	61	88.9	77.8	None
	MCAYAWVLVF	130	71	100.0	70.0	None
	YAWVLVYVY	137	68	100.0	55.6	None
HPV31-E5	FVIHTHASF	143	71	14.3	77.8	None
	LSVSVYATL	149	67	100.0	66.7	None
	SVYATLLL	139	72	100.0	44.4	None
	VVFIYIPLF	130	78	71.4	66.7	None
	VSVYATLLL	135	70	100.0	44.4	None
HPV45-E5	CAFAWLLVF	127	78	100.0	77.8	None
	YVCAFAWLL	109	62	100.0	55.6	None
	VYVCAFAWLL	127	56	100.0	60.0	None
	LHMHALHTL	111	82	66.7	55.6	None
	VITSPLTAF	125	69	100.0	88.9	None

\*Between HPV16 and 31 (alpha- Papillomavirus 9), and between HPV18 and 45 (alpha- Papillomavirus 7)

<https://doi.org/10.1371/journal.pone.0205933.t006>

## Discussion

High prevalence and mortality of oncogenic infectious pathogens such as HPV and *Helicobacter pylori* have caused serious problems for humans. Currently, people who are infected with hrHPVs but show normal cytology or precancerous lesions do not have any treatment option, causing the disease progress toward invasive carcinoma in some cases. Unfortunately, no FDA-approved immunotherapy exists for pre-existing HPV infections or their related cancers to date. Immunotherapy of HPV-associated cancers by DNA or peptide-based vaccines, depends on the recognition of highly immunogenic epitopes, inducing robust and specific immune responses, particularly cell-mediated responses against the malignant cells.

The primary aim of this study was the prediction of CD8+ T cell epitope from the E5, E6 and E7 oncoproteins, using a comprehensive two-step selection plan. These proteins chose because they play a pivotal role in the cell transformation, immune evasion, and maintenance of malignancy, as well as, their permanent expression (E6 and E7) by the malignant cells [24–26]. Expression of E5 oncoprotein occurs in the early phase of HPV infection. Evidence indicates that E5 play a prominent role in the genesis of HPV-associated cancers, but is not essential for cancer progression [90], since when HPV genome integrates into the host genome, it usually results in the disruption of E1, E2, and E5 genes. Therefore, targeting E5 protein provides an opportunity for treatment of HPV infections and preventing the precancerous lesions from the progression to established carcinomas [20, 91]. Some genotypes of hrHPVs are more involved in the genesis of epithelial tissue malignancies [61]. Thus, in this study, hrHPV16, 18, 31 and 45 were targeted due to their high prevalence in the HPV-associated cancers, especially cervical carcinoma.



Table 7. The predicted epitopes from E6 oncoproteins in the second step selection.

Protein	Epitope	Docking		Inter-isoform (% identity)	Conservancy	
		TDS (HLA-I)	TDS (HLA-II)		Inter-type* (% identity)	Inter-hrHPV (100% identical)
HPV16-E6	FAFRDLCIVY	110	66	96.6	80.0	None
	IVYRDGNPY	111	62	94.9	66.7	None
	CYSLYGTTL	131	73	63.6	77.8	None
	VYDFAFRDL	112	61	95.8	77.8	None
	KFYSKISEY	122	63	90.7	66.7	HPV16 and 35
HPV18-E6	KLPDLCTEL	114	62	87.5	100.0	HPV18 and 45
	TVLELTEVF	135	83	75.0	55.6	None
	FAFKDLFVVY	117	78	87.5	80.0	None
	LELTEVFEF	130	70	75.0	66.7	None
	DFYSRIREL	137	53	87.5	100.0	HPV18 and 45
HPV31-E6	FAFTDLTIVY	143	68	86.7	80.0	None
	RYSVYGTTL	149	70	100.0	77.8	None
	KVSEFRWYRY	139	51	100.0	60.0	None
	FRWYRYSVY	130	76	100.0	55.6	None
	YRYSVYGTTL	135	76	100.0	80.0	None
HPV45-E6	IVYRDCIAY	127	69	100.0	55.6	None
	KLPDLCTEL	109	65	56.3	100.0	HPV45 and 18
	YSRIRELRY	127	57	100.0	88.9	None
	ATLERTEVY	111	61	81.3	55.6	None
	FAFKDLCIVY	125	83	62.5	80.0	None

\*Between HPV16 and 31 (alpha- Papillomavirus 9), and between HPV18 and 45 (alpha- Papillomavirus 7)

<https://doi.org/10.1371/journal.pone.0205933.t007>

There are several limitations for epitope prediction: 1) The major drawback of peptide-based vaccines is low immunogenicity [92, 93]. Many studies have focused on enhancing immunogenicity using immune stimulating agents or adjuvants to avoid this problem. Another solution is the use of agonist epitopes [94]. Epitope immunogenicity is a crucial factor in vaccine development. However, many of known natural epitopes when are analyzed *in silico* by *IEDB MHC-I immunogenicity predictor*, do not obtain a high score. Therefore, in this study, epitope selection was based on the integrated approach, in which one analysis does not play an important role alone. 2) There are certain drawbacks associated with the function of each method invented for the MHC-peptide binding prediction [95]. For this reason, several predictors and a molecular docking program were used to augment the prediction accuracy. 3) Some web tools have been developed for MHC-II epitope prediction. Since MHC-II groove can bind to peptides with variable lengths, and different peptides have the different number of residues between their N-terminus and first anchor [54], the exact assignment of MHC-II core peptide would be a difficult problem which reduces the success rate of these prediction tools. Therefore, most MHC-II prediction tools did not usually make epitope predictions as accurately noted for MHC-I molecules [64, 96]. In cancer immunotherapy, the CTL-mediated responses play the central role in eradication of malignant cells, and the binding of epitopes to MHC-I molecules is an essential step for antigen presentation to CTLs. Thus, in this study, predicted epitopes were primarily selected by their MHC-I binding and processing scores. However, the MHC-II binding scores were actually of secondary importance to the epitope selection process as an extra advantage. Additionally, there are several other essential determinants which significantly affect the outcomes, such as antigen processing, immunogenicity,

**Table 8. The predicted epitopes from E7 oncoproteins in the second step selection.**

Protein	Epitope	Docking		Inter-isoform (% identity)	Conservancy	
		TDS (HLA-I)	TDS (HLA-II)		Inter-type* (% identity)	Inter-hrHPV (100% identical)
HPV16-E7	RAHYNIVTF	120	61	100.0	66.7	None
	STHVDIRTL	107	57	76.5	77.8	None
	LEDLLMGTL	102	58	100.0	55.6	None
	LQPETTDLY	107	65	82.4	77.8	None
	TLHEYMLDL	90	55	94.1	66.7	None
HPV18-E7	SSADDLRAF	90	52	90.9	66.7	None
	FQQLFNNTL	105	67	45.5	77.8	None
	QLFLNTLSF	127	69	45.5	81.8	None
	TLQDIVLHL	94	59	90.9	88.9	None
	LRAFQQLFL	108	76	90.9	77.8	None
HPV31-E7	TSNYNIVTF	116	60	93.3	66.7	None
	TLQDYVLDL	111	64	100.0	66.7	HPV18 and 35
	GQAEPDTSNY	109	50	40.0	70.0	None
	QAEPDTSNY	102	59	40.0	66.7	None
	QPEATDLHCY	110	54	66.7	80.0	None
HPV45-E7	TLQEIVLHL	85	57	100.0	88.9	None
	NELDPVDLL	98	68	100.0	55.6	None
	LQQLFLSTL	100	60	100.0	77.8	None
	QLFLSTLSF	108	66	100.0	88.9	None
	SSAEDLRTL	84	55	26.7	66.7	None

\*Between HPV16 and 31 (alpha- Papillomavirus 9), and between HPV18 and 45 (alpha- Papillomavirus)

<https://doi.org/10.1371/journal.pone.0205933.t008>

population coverage, conservancy and cross-reactivity with host antigens. Vaccine development requires a comprehensive approach to cover all these effectual elements, covered in this study.

The primary aim of molecular docking is the recognition of binding site of a ligand at a protein receptor surface, and docking and modeling the ligand into this recognized site. In this study, CABS-dock server was used for molecular docking analyses. CABS-dock has several main advantages: 1) The method does not require any data about the peptide structure and its binding site. 2) During docking process, peptide conformation is entirely flexible. 3) It is possible to apply dynamic conformational changes in the receptor structure and 4) to exclude some receptor regions from the docking search, leading to the more efficient search in the vicinity of the binding site at a sensible time. [88, 89].

In comparison with protein-ligand (small molecules) docking, Protein-peptide docking analysis is more problematic, since significant conformational changes occur during the process. As a general rule, how much the length of the query peptide to be longer, there are more torsions and conformational flexibilities. Additionally, in comparison to Protein-Protein interactions, Protein-peptide dockings are more transient, and their binding affinities are notably weaker [88]. These factors make structural predictions of long peptides very challenging. Therefore, in this study, 9mer peptides were preferred for selection compared to other possible lengths. They are also preferred by all MHC-I molecules as epitope and by MHC-II molecules as the core peptide of epitopes. Moreover, expansion of 9 or 10mer CTL epitopes to longer peptides may create a practical alternative, containing both CD4+ HTL and CD8+ CTL epitopes; Especially, when CD4+ HTL epitopes, covering CTL epitopes, are not recognized [97].

A large number of previous studies have used *in silico* analyses for epitope prediction against different pathogens [94, 96, 98–103]. However, the prediction of T cell epitopes inducing strong responses has remained a big challenge. For therapeutic HPV vaccines, many candidates have been designed to trigger the activation of CTLs or HTLs, mostly by targeting two major HPV oncoproteins, E6 or E7 [104], and in a few studies, E5 oncoprotein [98, 99]. As well as, several clinical trials have been launched for immunotherapy of HPV-associated cancers [46], although, they have not been so immunogenic, to induce a sufficient cellular immunity and eradicate malignant cells completely. Some studies have suggested that the use of E6 and E7 SLPs, containing both CD4+ HTL and CD8+ CTL epitopes, led to more potency and durability of CD8+ T cell reactivity *in vivo*, in comparison with the minimal CTL epitopes [97, 105].

In 1993, As pioneers in HPV epitope studies, Feltkamp *et al.* recognized the HPV16-E7 sequence RAHYNIVTF as an MHC-I epitope that can provoke CTL-mediated responses and eradicates established HPV 16-induced tumor cells in mice [106, 107]. This sequence is the first HPV16-E7 predicted epitope in our study as well.

In 2015, Kumar *et al.* studied HPV16-E5 oncoprotein to predict the candidate T-cell and B-cell epitopes [98]. They have screened 11 potent epitopes for MHC-I molecules according to PR and the immunogenicity score, using IEDB MHC-I binding and immunogenicity predictors. They found a 14mer potent epitope, SAFRCFIVYIIFVY, having the lowest PR and the highest immunogenicity score, *i.e.*, 0.5 and 0.70, respectively. Notably, our second HPV16-E5 predicted epitope, SAFRCFIVY, is the N-terminal part of SAFRCFIVYIIFVY, and our first predicted epitope, FLIHTHARF, is the C-terminal part of the third epitope of their study, VYIPLFLIHTHARF.

In 2017, Tsang *et al.* scanned the HPV16-E6 and E7 oncoproteins for the match peptides with the consensus motif of HLA-A2 binding peptides [94]. The BIMAS algorithm [108] was employed to rank probable binding peptides according to the predicted *one-half-time dissociation* of pMHCs. Three potential CTL predicted epitopes of the E6 protein (KLPQLCTEL, KISEYRHYC, and QQYNKPLCDL) and three of the E7 protein (YMLDLQPET, TLHEYMLDL, and RTLEDLLMGT) were selected. They showed the immunogenicity of these peptides was enhanced when their agonist epitopes were used. The KLPQLCTEL and TLHEYMLDL sequences are the seventh and the fifth predicted epitopes of HPV16-E6 and HPV16-E7 in our study, respectively.

Experimental evidences about hrHPV-derived epitopes in literatures are mostly limited to E6 and E7 oncoproteins of HPV 16 and 18. Among our first-step predicted epitopes: FLLCFCVLL and YIIFVYIPL from the E5-derived epitopes [109], FAFRDLICIVY [110], CYSLYGTTL [111], VYDFAFRDL [111, 112], KFYSKISEY [113], KLPQLCTEL [114–116], ISEYRHYCY [117], EYRHYCYSL [111], KLPDLCTEL [116, 118–120], FAFKDLFVV [119, 120] and KLPDLCTEL [116, 118–120] from the E6-derived epitopes, RAHYNIVTF [121], LEDLLMGT [122], TLHEYMLDL [115, 122–124], LLMGTGIV [115, 116, 125, 126], QAEPDRAHY [117], GTLGIVCPI [115, 126], FQQLFLNTL [127] and TLQDIVLHL [119] from the E7-derived epitopes were reported as T-cell epitopes experimentally. Besides, IVYRDGNPY, CYSLYGTTL, KLPQLCTEL and ISEYRHYCY from the E6-derived epitopes, and RAHYNIVTF and GTLGIVCPI from the E7-derived epitopes were also reported as HLA ligands [128]. Others are novel epitopes that they also require experimental studies for validation.

As far as we know, this is the first time that in a laborious *in silico* study for epitope prediction, E5, E6 and E7 oncoproteins of hrHPV16, 18, 31 and 45 have been investigated altogether. Moreover, in previous studies, usually only one predictor tool was used for making epitope prediction, or if several tools were used, no integrated approach was employed to make the conclusion. We believed that our predicted epitopes are valuable candidates for further *in vitro*

and *in vivo* therapeutic vaccine studies. Additionally, the introduction of the ten epitopes for each HPV genotype oncoprotein in the first step of the study shows which region of each oncoprotein is rich of the epitope, and thus, is more suitable for use in the design of SLPs. Notably, the previous *in vivo* studies have been conducted using SLPs of hrHPV-E6 and/or-E7 oncoproteins, in particular HPV16 oncoproteins [92, 129–133]. Furthermore, the two-step plan of this *in silico* study, which each step includes several analyses to find proper epitopes by an integrated approach, would provide a basis for rational epitope prediction. However, it could be more efficient by adding other useful analyses. Further studies are recommended on the peptide binding assays, the design of polyepitope constructions including E5, E6 and E7 epitopes, the expansion of the minimal CTL epitopes to longer peptides (SLPs), the use of various adjuvants, involvement of delivery routes, mouse immunization with the designed constructs, evaluation of immune responses such as cytokines, antibodies, CTLs and tumor growth for finding the best construct for clinical trials. It is important that improper vaccine design and immunosuppressive microenvironment were known as the main reasons of the failure in cancer immunotherapy by therapeutic cancer vaccines [134].

## Supporting information

### **S1 File. HPV16, 18, 31 and 45 oncoprotein sequences.**

(ZIP)

### **S2 File. CABS-dock procedure.**

(ZIP)

### **S3 File. MHC-I and II binding predictions.**

(ZIP)

### **S4 File. MHC-I immunogenicity predictions.**

(ZIP)

### **S5 File. MHC-I population coverages predictions.**

(ZIP)

### **S6 File. MHC-I and II molecular docking analyses.**

(ZIP)

### **S7 File. Epitope conservation analyses.**

(ZIP)

### **S1 Table. MHC-I binding prediction alleles.**

(ZIP)

### **S2 Table. Syfpeithi MHC-I binding prediction reference sequences.**

(ZIP)

### **S3 Table. MHC II binding predictions alleles.**

(ZIP)

## Acknowledgments

The authors sincerely thank Dr. Ali Namvar and Miss Elnaz Agi for their valuable guidance and comments during preparation of the paper.

## Author Contributions

**Conceptualization:** Heidar Ali Panahi, Azam Bolhassani, Gholamreza Javadi, Zahra Noormohammadi.

**Investigation:** Heidar Ali Panahi, Azam Bolhassani.

**Methodology:** Heidar Ali Panahi, Azam Bolhassani.

**Software:** Heidar Ali Panahi.

**Writing – original draft:** Heidar Ali Panahi.

**Writing – review & editing:** Heidar Ali Panahi, Azam Bolhassani, Zahra Noormohammadi.

## References

1. de Villiers E-M. Cross-roads in the classification of papillomaviruses. *Virology*. 2013; 445(1):2–10.
2. Kumar S, Biswas M, Jose T. HPV vaccine: Current status and future directions. *Medical journal, Armed Forces India*. 2015; 71(2):171. <https://doi.org/10.1016/j.mjafi.2015.02.006> PMID: 25859081
3. de Sanjosé S, Brotons M, Pavón MA. The natural history of human papillomavirus infection. *Best Practice & Research Clinical Obstetrics & Gynaecology*. 2017.
4. Bernard HU, Burk RD, Chen Z, Doorslaer K, Zur Hausen H, Villiers EM. Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. *Virology*. 2010; 401.
5. Forman D, de Martel C, Lacey CJ, Soerjomataram I, Lortet-Tieulent J, Bruni L, et al. Global burden of human papillomavirus and related diseases. *Vaccine*. 2012; 30:F12–F23. <https://doi.org/10.1016/j.vaccine.2012.07.055> PMID: 23199955
6. Schiffman M, Doorbar J, Wentzensen N, De Sanjosé S, Fakhry C, Monk BJ, et al. Carcinogenic human papillomavirus infection. *Nature Reviews Disease Primers*. 2016; 2:16086. <https://doi.org/10.1038/nrdp.2016.86> PMID: 27905473
7. Doorbar J, Egawa N, Griffin H, Kranjec C, Murakami I. Human papillomavirus molecular biology and disease association. *Rev Med Virol*. 2015; 25(S1):2–23.
8. Stanley M. Tumour virus vaccines: hepatitis B virus and human papillomavirus. *Phil Trans R Soc B*. 2017; 372(1732):20160268. <https://doi.org/10.1098/rstb.2016.0268> PMID: 28893935
9. Martel C, Ferlay J, Franceschi S, Vignat J, Bray F, Forman D. Global burden of cancers attributable to infections in 2008: a review and synthetic analysis. *The Lancet Oncology*. 2012; 13.
10. de Martel C, Plummer M, Vignat J, Franceschi S. Worldwide burden of cancer attributable to HPV by site, country and HPV type. *Int J Cancer*. 2017.
11. Bosch FX, Broker TR, Forman D, Moscicki A-B, Gillison ML, Doorbar J, et al. Comprehensive control of human papillomavirus infections and related diseases. *Vaccine*. 2013; 31:H1–H31. <https://doi.org/10.1016/j.vaccine.2013.10.003> PMID: 24332295
12. Burk RD, Harari A, Chen Z. Human papillomavirus genome variants. *Virology*. 2013; 445(1):232–43.
13. Guan P, Howell-Jones R, Li N, Bruni L, Sanjosé S, Franceschi S. Human papillomavirus types in 115,789 HPV-positive women: a meta-analysis from cervical infection to cancer. *International journal of cancer*. 2012; 131.
14. Humans IWGotEoCRt. Human papillomaviruses. IARC Monographs on the evaluation of carcinogenic risks to humans. 2007; 90:1. PMID: 18354839
15. Ashrafi GH, Haghshenas M, Marchetti B, Campo MS. E5 protein of human papillomavirus 16 downregulates HLA class I and interacts with the heavy chain via its first hydrophobic domain. *Int J Cancer*. 2006; 119(9):2105–12. <https://doi.org/10.1002/ijc.22089> PMID: 16823848
16. Marchetti B, Ashrafi GH, Tsirimonaki E, O'Brien PM, Campo MS. The bovine papillomavirus oncoprotein E5 retains MHC class I molecules in the Golgi apparatus and prevents their transport to the cell surface. *Oncogene*. 2002; 21(51):7808. <https://doi.org/10.1038/sj.onc.1205885> PMID: 12420217
17. Scheffner M, Werness BA, Huibregtse JM, Levine AJ, Howley PM. The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. *Cell*. 1990; 63(6):1129–36. PMID: 2175676
18. Gonzalez SL, Strelau M, He X, Basile JR, Mürnger K. Degradation of the retinoblastoma tumor suppressor by the human papillomavirus type 16 E7 oncoprotein is important for functional inactivation and is separable from proteasomal degradation of E7. *Journal of virology*. 2001; 75(16):7583–91. <https://doi.org/10.1128/JVI.75.16.7583-7591.2001> PMID: 11462030

19. Doorbar J, Quint W, Banks L, Bravo IG, Stoler M, Broker TR, et al. The biology and life-cycle of human papillomaviruses. *Vaccine*. 2012; 30:F55–F70. <https://doi.org/10.1016/j.vaccine.2012.06.083> PMID: 23199966
20. McBride AA, Warburton A. The role of integration in oncogenic progression of HPV-associated cancers. *PLoS Path*. 2017; 13(4):e1006211.
21. Stanley M. Immune responses to human papillomavirus. *Vaccine*. 2006; 24, Supplement 1:S16–S22.
22. Kawana K, Adachi K, Kojima S, Kozuma S, Fujii T. Therapeutic Human Papillomavirus (HPV) Vaccines: A Novel Approach. *Open Virol J*. 2012; 6:264–9. <https://doi.org/10.2174/1874357901206010264> PMID: 23341862
23. O'Brien PM, Campo MS. Evasion of host immunity directed by papillomavirus-encoded proteins. *Virus Res*. 2002; 88(1–2):103–17. PMID: 12297330
24. Howie HL, Katzenellenbogen RA, Galloway DA. Papillomavirus E6 proteins. *Virology*. 2009; 384.
25. DiMaio D, Petti LM. The E5 proteins. *Virology*. 2013; 445(1):99–114.
26. Roman A, Munger K. The papillomavirus E7 proteins. *Virology*. 2013; 445(1):138–68.
27. Li J, Chen S, Ge J, Lu F, Ren S, Zhao Z, et al. A novel therapeutic vaccine composed of a rearranged human papillomavirus type 16 E6/E7 fusion protein and Fms-like tyrosine kinase-3 ligand induces CD8+ T cell responses and antitumor effect. *Vaccine*. 2017.
28. Einstein MH, Kadish AS, Burk RD, Kim MY, Wadler S, Streicher H, et al. Heat shock fusion protein-based immunotherapy for treatment of cervical intraepithelial neoplasia III. *Gynecologic oncology*. 2007; 106(3):453–60. <https://doi.org/10.1016/j.ygyno.2007.04.038> PMID: 17586030
29. eltkamp MC, Smits HL, Vierboom MP, Minnaar RP, De Jongh BM, Drijfhout JW, et al. Vaccination with cytotoxic T lymphocyte epitope-containing peptide protects against a tumor induced by human papillomavirus type 16-transformed cells. *Eur J Immunol*. 1993; 23(9):2242–9. <https://doi.org/10.1002/eji.1830230929> PMID: 7690326
30. Manuri PR, Nehete B, Nehete PN, Reisenauer R, Wardell S, Courtney AN, et al. Intranasal immunization with synthetic peptides corresponding to the E6 and E7 oncoproteins of human papillomavirus type 16 induces systemic and mucosal cellular immune responses and tumor protection. *Vaccine*. 2007; 25(17):3302–10. <https://doi.org/10.1016/j.vaccine.2007.01.010> PMID: 17291642
31. Peng S, Trimble C, Wu L, Pardoll D, Roden R, Hung C-F, et al. HLA-DQB1\* 02–restricted HPV-16 E7 peptide–specific CD4+ T-cell immune responses correlate with regression of HPV-16–associated high-grade squamous intraepithelial lesions. *Clinical cancer research*. 2007; 13(8):2479–87. <https://doi.org/10.1158/1078-0432.CCR-06-2916> PMID: 17438108
32. Wang JW, Roden RB. L2, the minor capsid protein of papillomavirus. *Virology*. 2013; 445(1):175–86.
33. Schiller JT, Castellsagué X, Garland SM. A review of clinical trials of human papillomavirus prophylactic vaccines. *Vaccine*. 2012; 30:F123–F38. <https://doi.org/10.1016/j.vaccine.2012.04.108> PMID: 23199956
34. Joura EA, Giuliano AR, Iversen O-E, Bouchard C, Mao C, Mehlsen J, et al. A 9-valent HPV vaccine against infection and intraepithelial neoplasia in women. *New Engl J Med*. 2015; 372(8):711–23. <https://doi.org/10.1056/NEJMoa1405044> PMID: 25693011
35. Schiller JT, Castellsagué X, Villa LL, Hildesheim A. An update of prophylactic human papillomavirus L1 virus-like particle vaccine clinical trial results. *Vaccine*. 2008; 26:K53–K61. <https://doi.org/10.1016/j.vaccine.2008.06.002> PMID: 18847557
36. Gomez-Gutierrez JG, Elpek KG, de Oca-Luna RM, Shirwan H, Zhou HS, McMasters KM. Vaccination with an adenoviral vector expressing calreticulin-human papillomavirus 16 E7 fusion protein eradicates E7 expressing established tumors in mice. *Cancer Immunology, Immunotherapy*. 2007; 56(7):997–1007. <https://doi.org/10.1007/s00262-006-0247-2> PMID: 17146630
37. Cassetti MC, McElhiney SP, Shahabi V, Pullen JK, Le Poole IC, Eiben GL, et al. Antitumor efficacy of Venezuelan equine encephalitis virus replicon particles encoding mutated HPV16 E6 and E7 genes. *Vaccine*. 2004; 22(3):520–7.
38. Cheng W-F, Hung C-F, Pai SI, Hsu K-F, He L, Ling M, et al. Repeated DNA vaccinations elicited qualitatively different cytotoxic T lymphocytes and improved protective antitumor effects. *J Biomed Sci*. 2002; 9(6):675–87.
39. Lin C-T, Tsai Y-C, He L, Calizo R, Chou H-H, Chang T-C, et al. A DNA vaccine encoding a codon-optimized human papillomavirus type 16 E6 gene enhances CTL response and anti-tumor activity. *J Biomed Sci*. 2006; 13(4):481–8. <https://doi.org/10.1007/s11373-006-9086-6> PMID: 16649071
40. Peng S, Trimble C, Alvarez RD, Huh WK, Lin Z, Monie A, et al. Cluster intradermal DNA vaccination rapidly induces E7-specific CD8+ T-cell immune responses leading to therapeutic antitumor effects. *Gene Ther*. 2008; 15(16):1156–66. <https://doi.org/10.1038/gt.2008.53> PMID: 18401437



41. Chandy AG, Nurkkala M, Josefsson A, Eriksson K. Therapeutic dendritic cell vaccination with Ag coupled to cholera toxin in combination with intratumoural CpG injection leads to complete tumour eradication in mice bearing HPV 16 expressing tumours. *Vaccine*. 2007; 25(32):6037–46. <https://doi.org/10.1016/j.vaccine.2007.05.041> PMID: 17629599
42. Reinis M, Stepanek I, Simova J, Bieblova J, Pribylova H, Indrova M, et al. Induction of protective immunity against MHC class I-deficient, HPV16-associated tumours with peptide and dendritic cell-based vaccines. *Int J Oncol*. 2010; 36(3):545–51. PMID: 20126973
43. Da Silva DM, Schiller JT, Kast WM. Heterologous boosting increases immunogenicity of chimeric papillomavirus virus-like particle vaccines. *Vaccine*. 2003; 21(23):3219–27. PMID: 12804851
44. Kaufmann AM, Nieland JD, Jochmus I, Baur S, Friese K, Gabelsberger J, et al. Vaccination trial with HPV16 L1E7 chimeric virus-like particles in women suffering from high grade cervical intraepithelial neoplasia (CIN 2/3). *Int J Cancer*. 2007; 121(12):2794–800. <https://doi.org/10.1002/ijc.23022> PMID: 17721997
45. Warrino DE, Olson WC, Scarrow MI, D'Ambrosio-Brennan LJ, Guido RS, Da Silva DM, et al. Human papillomavirus L1L2-E7 virus-like particles partially mature human dendritic cells and elicit E7-specific T-helper responses from patients with cervical intraepithelial neoplasia or cervical cancer in vitro. *Human immunology*. 2005; 66(7):762–72. <https://doi.org/10.1016/j.humimm.2005.04.006> PMID: 16112023
46. Vici P, Pizzuti L, Mariani L, Zampa G, Santini D, Di Lauro L, et al. Targeting immune response with therapeutic vaccines in premalignant lesions and cervical cancer: hope or reality from clinical studies. Expert review of vaccines. 2016; 15(10):1327–36. <https://doi.org/10.1080/14760584.2016.1176533> PMID: 27063030
47. Paris R, Bejrachandra S, Thongcharoen P, Nitayaphan S, Pitisuttithum P, Sambor A, et al. HLA class II restriction of HIV-1 clade-specific neutralizing antibody responses in ethnic Thai recipients of the RV144 prime-boost vaccine combination of ALVAC-HIV and AIDSVAX B/E. *Vaccine*. 2012; 30(5):832–6. <https://doi.org/10.1016/j.vaccine.2011.11.002> PMID: 22085554
48. Singh SP, Mishra BN. Major histocompatibility complex linked databases and prediction tools for designing vaccines. *Human Immunology*. 2016; 77(3):295–306. <https://doi.org/10.1016/j.humimm.2015.11.012> PMID: 26585361
49. Abbas AK, Lichtman AH, Pillai S. Cellular and Molecular Immunology. Eighth ed: Elsevier Health Sciences; 2014.
50. Bui H-H, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinformatics*. 2006; 7(1):153.
51. Sidney J, Peters B, Frahm N, Brander C, Sette A. HLA class I supertypes: a revised and updated classification. *BMC Immunol*. 2008; 9(1):1.
52. Rosa DS, Ribeiro SP, Cunha-Neto E. CD4+ T cell epitope discovery and rational vaccine design. *Archivum immunologiae et therapiae experimentalis*. 2010; 58(2):121–30. <https://doi.org/10.1007/s00005-010-0067-0> PMID: 20155490
53. Ribeiro SP, Rosa DS, Fonseca SG, Mairena EC, Postol E, Oliveira SC, et al. A vaccine encoding conserved promiscuous HIV CD4 epitopes induces broad T cell responses in mice transgenic to multiple common HLA class II molecules. *PLoS One*. 2010; 5(6):e11072. <https://doi.org/10.1371/journal.pone.0011072> PMID: 20552033
54. Rammensee H-G, Bachmann J, Emmerich NPN, Bachor OA, Stevanović S. SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics*. 1999; 50(3–4):213–9. PMID: 10602881
55. Jurtz VI, Paul S, Andreatta M, Marcatili P, Peters B, Nielsen M. NetMHCpan 4.0: Improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *bioRxiv*. 2017:149518.
56. Sirskiy D, Diaz-Mitoma F, Golshani A, Kumar A, Azizi A. Innovative bioinformatic approaches for developing peptide-based vaccines against hypervariable viruses. *Immunol Cell Biol*. 2011; 89(1):81. <https://doi.org/10.1038/icb.2010.65> PMID: 20458336
57. Tsurui H, Takahashi T. Prediction of T-cell epitope. *Journal of pharmacological sciences*. 2007; 105(4):299–316. PMID: 18094522
58. Ruppert J, Sidney J, Celis E, Kubo RT, Grey HM, Sette A. Prominent role of secondary anchor residues in peptide binding to HLA-A2. 1 molecules. *Cell*. 1993; 74(5):929–37. PMID: 8104103
59. Nielsen M, Andreatta M. NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets. *Genome Med*. 2016; 8(1):33. <https://doi.org/10.1186/s13073-016-0288-x> PMID: 27029192
60. Vita R, Overton JA, Greenbaum JA, Ponomarenko J, Clark JD, Cantrell JR, et al. The immune epitope database (IEDB) 3.0. *Nucleic Acids Res*. 2015; 43(Database issue):D405–12. <https://doi.org/10.1093/nar/gku938> PMID: 25300482

61. Bruni L, Barrionuevo-Rosas L, Albero G, Aldea M, Serrano B, Mena M, et al. Human Papillomavirus and Related Diseases in the World. Summary Report 27 July 2017. ICO/IARC Information Centre on HPV and Cancer (HPV Information Centre). 2017:1–325.
62. Tenzer S, Peters B, Bulik S, Schoor O, Lemmel C, Schatz M, et al. Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. *Cell Mol Life Sci.* 2005; 62(9):1025–37. <https://doi.org/10.1007/s00018-005-4528-2> PMID: 15868101
63. Wang P, Sidney J, Kim Y, Sette A, Lund O, Nielsen M, et al. Peptide binding predictions for HLA DR, DP and DQ molecules. *BMC Bioinformatics.* 2010; 11:568. <https://doi.org/10.1186/1471-2105-11-568> PMID: 21092157
64. Wang P, Sidney J, Dow C, Mothe B, Sette A, Peters B. A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. *PLoS Comp Biol.* 2008; 4(4): e1000048.
65. Immune Epitope Database and analysis resource (IEDB) 3.0. MHC-I processing predictions—Tutorial; National Institute of Allergy and Infectious Diseases; 2018; [updated: January 07, 2018]. Available from: <http://tools.iedb.org/processing/help/>.
66. Jurtz V, Paul S, Andreatta M, Marcatili P, Peters B, Nielsen M. NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. *J Immunol.* 2017; 199(9):3360–8. <https://doi.org/10.4049/jimmunol.1700893> PMID: 28978689
67. Reche PA, Reinherz EL. Prediction of peptide-MHC binding using profiles. *Immunoinformatics: Predicting Immunogenicity In Silico.* 2007:185–200.
68. Reche PA, Glutting J-P, Reinherz EL. Prediction of MHC class I binding peptides using profile motifs. *Human immunology.* 2002; 63(9):701–9. PMID: 12175724
69. Moutaftsi M, Peters B, Pasquetto V, Tschärke DC, Sidney J, Bui HH, et al. A consensus epitope prediction approach identifies the breadth of murine T(CD8+)-cell responses to vaccinia virus. *Nat Biotechnol.* 2006; 24(7):817–9. <https://doi.org/10.1038/nbt1215> PMID: 16767078
70. Hoof I, Peters B, Sidney J, Pedersen LE, Sette A, Lund O, et al. NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics.* 2009; 61(1):1–13. <https://doi.org/10.1007/s00251-008-0341-z> PMID: 19002680
71. Andreatta M, Nielsen M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics.* 2016; 32(4):511–7. <https://doi.org/10.1093/bioinformatics/btv639> PMID: 26515819
72. Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, Nielsen M. NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8–11. *Nucleic acids research.* 2008; 36(suppl\_2):W509–W12.
73. Kim Y, Sidney J, Pinilla C, Sette A, Peters B. Derivation of an amino acid similarity matrix for peptide: MHC binding and its application as a Bayesian prior. *BMC Bioinformatics.* 2009; 10:394. <https://doi.org/10.1186/1471-2105-10-394> PMID: 19948066
74. Peters B, Sette A. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinformatics.* 2005; 6:132. <https://doi.org/10.1186/1471-2105-6-132> PMID: 15927070
75. Sidney J, Assarsson E, Moore C, Ngo S, Pinilla C, Sette A, et al. 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. <https://doi.org/10.1186/1745-7580-4-2> PMID: 18221540
76. Zhang H, Lund O, Nielsen M. The PickPocket method for predicting binding specificities for receptors based on receptor pocket similarities: application to MHC-peptide binding. *Bioinformatics.* 2009; 25(10):1293–9. <https://doi.org/10.1093/bioinformatics/btp137> PMID: 19297351
77. Karosiene E, Lundegaard C, Lund O, Nielsen M. NetMHCcons: a consensus method for the major histocompatibility complex class I predictions. *Immunogenetics.* 2012; 64(3):177–86. <https://doi.org/10.1007/s00251-011-0579-8> PMID: 22009319
78. Rasmussen M, Fenoy E, Harndahl M, Kristensen AB, Nielsen IK, Nielsen M, et al. Pan-Specific Prediction of Peptide–MHC Class I Complex Stability, a Correlate of T Cell Immunogenicity. *The Journal of Immunology.* 2016; 197(4):1517–24. <https://doi.org/10.4049/jimmunol.1600582> PMID: 27402703
79. BMI (Biomedical Informatics)-Heidelberg. Information on SYFPEITHI. institute for cell biology-department of immunology-Heidelberg; 2012; [updated: 27 Aug 2017]. Available from: <http://www.syfpeithi.de/bin/MHCServer.dll/Info.htm>.
80. Andreatta M, Karosiene E, Rasmussen M, Stryhn A, Buus S, Nielsen M. Accurate pan-specific prediction of peptide-MHC class II binding affinity with improved binding core identification. *Immunogenetics.* 2015; 67(11–12):641–50. <https://doi.org/10.1007/s00251-015-0873-y> PMID: 26416257

81. Nielsen M, Lund O. NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. *BMC Bioinformatics*. 2009; 10(1):296.
82. Nielsen M, Lundegaard C, Lund O. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. *BMC Bioinformatics*. 2007; 8(1):238.
83. Sturniolo T, Bono E, Ding J, Raddrizzani L, Tuereci O, Sahin U, et al. Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. *Nat Biotechnol*. 1999; 17(6):555. <https://doi.org/10.1038/9858> PMID: 10385319
84. Peters B, Bulik S, Tampe R, Van Ender PM, Holzhütter H-G. Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors. *The Journal of Immunology*. 2003; 171(4):1741–9. PMID: 12902473
85. Calis JJ, Maybeno M, Greenbaum JA, Weiskopf D, De Silva AD, Sette A, et al. Properties of MHC class I presented peptides that enhance immunogenicity. *PLoS Comput Biol*. 2013; 9(10):e1003266. <https://doi.org/10.1371/journal.pcbi.1003266> PMID: 24204222
86. Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinformatics*. 2006; 7:153. <https://doi.org/10.1186/1471-2105-7-153> PMID: 16545123
87. González-Galarza FF, Takeshita LY, Santos EJ, Kempson F, Maia MHT, Silva ALSd, et al. Allele frequency net 2015 update: new features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations. *Nucleic acids research*. 2014; 43(D1):D784–D8.
88. Blaszczyk M, Kurcinski M, Kouza M, Wieteska L, Debinski A, Kolinski A, et al. Modeling of protein–peptide interactions using the CABS-dock web server for binding site search and flexible docking. *Methods*. 2016; 93:72–83. <https://doi.org/10.1016/j.ymeth.2015.07.004> PMID: 26165956
89. Kurcinski M, Jamroz M, Blaszczyk M, Kolinski A, Kmiecik S. CABS-dock web server for the flexible docking of peptides to proteins without prior knowledge of the binding site. *Nucleic Acids Research*. 2015; 43(W1):W419–W24. <https://doi.org/10.1093/nar/gkv456> PMID: 25943545
90. Kim S-W, Yang J-S. Human papillomavirus type 16 E5 protein as a therapeutic target. *Yonsei Med J*. 2006; 47(1):1–14. <https://doi.org/10.3349/ymj.2006.47.1.1> PMID: 16502480
91. Liu D-W, Tsao Y-P, Kung JT, Ding Y-A, Sytwu H-K, Xiao X, et al. Recombinant adeno-associated virus expressing human papillomavirus type 16 E7 peptide DNA fused with heat shock protein DNA as a potential vaccine for cervical cancer. *Journal of virology*. 2000; 74(6):2888–94. PMID: 10684306
92. Melief CJ, Van Der Burg SH. Immunotherapy of established (pre) malignant disease by synthetic long peptide vaccines. *Nature Reviews Cancer*. 2008; 8(5):351–60. <https://doi.org/10.1038/nrc2373> PMID: 18418403
93. Purcell AW, McCluskey J, Rossjohn J. More than one reason to rethink the use of peptides in vaccine design. *Nature reviews Drug discovery*. 2007; 6(5):404–14. <https://doi.org/10.1038/nrd2224> PMID: 17473845
94. Tsang KY, Fantini M, Fernando RI, Palena C, David JM, Hodge JW, et al. Identification and characterization of enhancer agonist human cytotoxic T-cell epitopes of the human papillomavirus type 16 (HPV16) E6/E7. *Vaccine*. 2017; 35(19):2605–11. <https://doi.org/10.1016/j.vaccine.2017.03.025> PMID: 28389098
95. Lafuente EM, Reche PA. Prediction of MHC-peptide binding: a systematic and comprehensive overview. *Curr Pharm Des*. 2009; 15(28):3209–20. PMID: 19860671
96. Srivastava PN, Jain R, Dubey SD, Bhatnagar S, Ahmad N. Prediction of epitope-based peptides for vaccine development from coat proteins GP2 and VP24 of Ebola virus using immunoinformatics. *International Journal of Peptide Research and Therapeutics*. 2016; 22(1):119–33.
97. Bijker MS, van den Eeden SJ, Franken KL, Melief CJ, Offringa R, van der Burg SH. CD8+ CTL priming by exact peptide epitopes in incomplete Freund's adjuvant induces a vanishing CTL response, whereas long peptides induce sustained CTL reactivity. *The Journal of Immunology*. 2007; 179(8):5033–40. PMID: 17911588
98. Kumar A, Yadav IS, Hussain S, Das BC, Bharadwaj M. Identification of immunotherapeutic epitope of E5 protein of human papillomavirus-16: An in silico approach. *Biologicals*. 2015; 43(5):344–8. <https://doi.org/10.1016/j.biologicals.2015.07.002> PMID: 26212000
99. Lissabet JB. Computational prediction of linear B-cell epitopes in the E5 oncoprotein of the human papillomavirus type 16 using several bioinformatics tools. *Vacunas*. 2016; 17(1):18–26.
100. Abu-haraz AH, Abd-elrahman KA, Ibrahim MS, Hussien WH, Mohammed MS, Badawi MM, et al. Multi Epitope Peptide Vaccine Prediction against Sudan Ebola Virus Using Immuno-Informatics Approaches. *Adv Tech Biol Med*. 2017; 5(203):2379–1764.1000203.
101. Khan AM, Miotto O, Heiny AT, Salmon J, Srinivasan KN, Nascimento EJM, et al. A systematic bioinformatics approach for selection of epitope-based vaccine targets. *Cell Immunol*. 2006; 244(2):141–7. <https://doi.org/10.1016/j.cellimm.2007.02.005> PMID: 17434154

102. Nezafat N, Ghasemi Y, Javadi G, Khoshnoud MJ, Omidinia E. A novel multi-epitope peptide vaccine against cancer: An in silico approach. *J Theor Biol.* 2014; 349(Supplement C):121–34.
103. Shi J, Zhang J, Li S, Sun J, Teng Y, Wu M, et al. Epitope-based vaccine target screening against highly pathogenic MERS-CoV: an in silico approach applied to emerging infectious diseases. *PLoS one.* 2015; 10(12):e0144475. <https://doi.org/10.1371/journal.pone.0144475> PMID: 26641892
104. Liu TY, Hussein WM, Toth I, Skwarczynski M. Advances in peptide-based human papillomavirus therapeutic vaccines. *Curr Top Med Chem.* 2012; 12(14):1581–92. PMID: 22827526
105. Vambutas A, DeVoti J, Nouri M, Drijfhout J, Lipford G, Bonagura V, et al. Therapeutic vaccination with papillomavirus E6 and E7 long peptides results in the control of both established virus-induced lesions and latently infected sites in a pre-clinical cottontail rabbit papillomavirus model. *Vaccine.* 2005; 23(45):5271–80. <https://doi.org/10.1016/j.vaccine.2005.04.049> PMID: 16054734
106. Feltkamp MC, Smits HL, Vierboom MP, Minnaar RP, De Jongh BM, Drijfhout JW, et al. Vaccination with cytotoxic T lymphocyte epitope-containing peptide protects against a tumor induced by human papillomavirus type 16-transformed cells. *Eur J Immunol.* 1993; 23(9):2242–9. <https://doi.org/10.1002/eji.1830230929> PMID: 7690326
107. Feltkamp MC, Vreugdenhil GR, Vierboom MP, Ras E, van der Burg SH, Schegget JT, et al. Cytotoxic T lymphocytes raised against a subdominant epitope offered as a synthetic peptide eradicate human papillomavirus type 16-induced tumors. *Eur J Immunol.* 1995; 25(9):2638–42. <https://doi.org/10.1002/eji.1830250935> PMID: 7589138
108. Parker KC, Bednarek MA, Coligan JE. Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. *The Journal of Immunology.* 1994; 152(1):163–75. PMID: 8254189
109. Liu DW, Yang YC, Lin HF, Lin MF, Cheng YW, Chu CC, et al. Cytotoxic T-lymphocyte responses to human papillomavirus type 16 E5 and E7 proteins and HLA-A\*0201-restricted T-cell peptides in cervical cancer patients. *J Virol.* 2007; 81(6):2869–79. <https://doi.org/10.1128/JVI.02256-06> PMID: 17202211
110. Nakagawa M, Kim KH, Gillam TM, Moscicki AB. HLA class I binding promiscuity of the CD8 T-cell epitopes of human papillomavirus type 16 E6 protein. *J Virol.* 2007; 81(3):1412–23. <https://doi.org/10.1128/JVI.01768-06> PMID: 17108051
111. Mizuuchi M, Hirohashi Y, Torigoe T, Kuroda T, Yasuda K, Shimizu Y, et al. Novel oligomannose liposome-DNA complex DNA vaccination efficiently evokes anti-HPV E6 and E7 CTL responses. *Exp Mol Pathol.* 2012; 92(1):185–90. <https://doi.org/10.1016/j.yexmp.2011.10.002> PMID: 22032938
112. Morishima S, Akatsuka Y, Nawa A, Kondo E, Kiyono T, Torikai H, et al. Identification of an HLA-A24-restricted cytotoxic T lymphocyte epitope from human papillomavirus type-16 E6: the combined effects of bortezomib and interferon-gamma on the presentation of a cryptic epitope. *International journal of cancer.* 2007; 120(3):594–604. <https://doi.org/10.1002/ijc.22312> PMID: 17096336
113. Hara M, Matsueda S, Tamura M, Takedatsu H, Tanaka M, Kawano K, et al. Identification of human papillomavirus 16-E6 protein-derived peptides with the potential to generate cytotoxic T-lymphocytes toward human leukocyte antigen-A24+ cervical cancer. *Int J Oncol.* 2005; 27(5):1371–9. PMID: 16211234
114. Zehbe I, Kaufmann AM, Schmidt M, Hohn H, Maeurer MJ. Human papillomavirus 16 E6-specific CD45RA+ CCR7+ high avidity CD8+ T cells fail to control tumor growth despite interferon-gamma production in patients with cervical cancer. *J Immunother.* 2007; 30(5):523–32. <https://doi.org/10.1097/CJI.0b013e31803240fa> PMID: 17589293
115. Riemer AB, Keskin DB, Zhang G, Handley M, Anderson KS, Brusic V, et al. A conserved E7-derived cytotoxic T lymphocyte epitope expressed on human papillomavirus 16-transformed HLA-A2+ epithelial cancers. *J Biol Chem.* 2010; 285(38):29608–22. <https://doi.org/10.1074/jbc.M110.126722> PMID: 20615877
116. Matijevic M, Hedley ML, Urban RG, Chicz RM, Lajoie C, Luby TM. Immunization with a poly (lactide co-glycolide) encapsulated plasmid DNA expressing antigenic regions of HPV 16 and 18 results in an increase in the precursor frequency of T cells that respond to epitopes from HPV 16, 18, 6 and 11. *Cell Immunol.* 2011; 270(1):62–9. <https://doi.org/10.1016/j.cellimm.2011.04.005> PMID: 21550027
117. Bourgault Villada I, Beneton N, Bony C, Connan F, Monsonego J, Bianchi A, et al. Identification in humans of HPV-16 E6 and E7 protein epitopes recognized by cytolytic T lymphocytes in association with HLA-B18 and determination of the HLA-B18-specific binding motif. *Eur J Immunol.* 2000; 30(8):2281–9. [https://doi.org/10.1002/1521-4141\(2000\)30:8<2281::AID-IMMU2281>3.0.CO;2-N](https://doi.org/10.1002/1521-4141(2000)30:8<2281::AID-IMMU2281>3.0.CO;2-N) PMID: 10940919
118. Mora-Garcia Mde L, Duenas-Gonzalez A, Hernandez-Montes J, De la Cruz-Hernandez E, Perez-Cardenas E, Weiss-Steider B, et al. Up-regulation of HLA class-I antigen expression and antigen-specific CTL response in cervical cancer cells by the demethylating agent hydralazine and the histone

- deacetylase inhibitor valproic acid. *J Transl Med.* 2006; 4:55. <https://doi.org/10.1186/1479-5876-4-55> PMID: 17192185
119. Rudolf MP, Man S, Melief CJ, Sette A, Kast WM. Human T-cell responses to HLA-A-restricted high binding affinity peptides of human papillomavirus type 18 proteins E6 and E7. *Clin Cancer Res.* 2001; 7(3 Suppl):788s–95s.
  120. Yoon H, Chung MK, Min SS, Lee HG, Yoo WD, Chung KT, et al. Synthetic peptides of human papillomavirus type 18 E6 harboring HLA-A2.1 motif can induce peptide-specific cytotoxic T-cells from peripheral blood mononuclear cells of healthy donors. *Virus research.* 1998; 54(1):23–9. PMID: 9660068
  121. Eiben GL, Velders MP, Schreiber H, Cassetti MC, Pullen JK, Smith LR, et al. Establishment of an HLA-A\*0201 human papillomavirus type 16 tumor model to determine the efficacy of vaccination strategies in HLA-A\*0201 transgenic mice. *Cancer Res.* 2002; 62(20):5792–9. PMID: 12384540
  122. Nakagawa M, Kim KH, Moscicki AB. Different methods of identifying new antigenic epitopes of human papillomavirus type 16 E6 and E7 proteins. *Clin Diagn Lab Immunol.* 2004; 11(5):889–96. <https://doi.org/10.1128/CDLI.11.5.889-896.2004> PMID: 15358648
  123. Oerke S, Hohn H, Zehbe I, Pilch H, Schicketanz KH, Hitzler WE, et al. Naturally processed and HLA-B8-presented HPV16 E7 epitope recognized by T cells from patients with cervical cancer. *International journal of cancer.* 2005; 114(5):766–78. <https://doi.org/10.1002/ijc.20794> PMID: 15609316
  124. Rensing ME, Sette A, Brandt R, Ruppert J, Wentworth PA, Hartman M, et al. Human CTL epitopes encoded by human papillomavirus type 16 E6 and E7 identified through in vivo and in vitro immunogenicity studies of HLA-A\*0201-binding peptides. *The Journal of Immunology.* 1995; 154(11):5934–43. PMID: 7538538
  125. Ferrara A, Nonn M, Sehr P, Schreckenberger C, Pawlita M, Dürst M. Dendritic cell-based tumor vaccine for cervical cancer II: results of a clinical pilot study in 15 individual patients. *J Cancer Res Clin Oncol.* 2003; 129.
  126. Rensing ME, Sette A, Brandt RM, Ruppert J, Wentworth PA, Hartman M, et al. Human CTL epitopes encoded by human papillomavirus type 16 E6 and E7 identified through in vivo and in vitro immunogenicity studies of HLA-A\*0201-binding peptides. *J Immunol.* 1995; 154(11):5934–43. PMID: 7538538
  127. Kather A, Ferrara A, Nonn M, Schinz M, Nieland J, Schneider A, et al. Identification of a naturally processed HLA-A\*0201 HPV18 E7 T cell epitope by tumor cell mediated in vitro vaccination. *International journal of cancer.* 2003; 104(3):345–53. <https://doi.org/10.1002/ijc.10940> PMID: 12569558
  128. Kast WM, Brandt R, Sidney J, Drijfhout J-W, Kubo RT, Grey HM, et al. Role of HLA-A motifs in identification of potential CTL epitopes in human papillomavirus type 16 E6 and E7 proteins. *The Journal of Immunology.* 1994; 152(8):3904–12. PMID: 7511661
  129. Arens R, van Hall T, van der Burg SH, Ossendorp F, Melief CJ, editors. *Prospects of combinatorial synthetic peptide vaccine-based immunotherapy against cancer.* Semin Immunol; 2013: Elsevier.
  130. Quakkelaar ED, Melief CJ. Experience with synthetic vaccines for cancer and persistent virus infections in nonhuman primates and patients. *Adv Immunol.* 114: Elsevier; 2012. p. 77–106. <https://doi.org/10.1016/B978-0-12-396548-6.00004-4> PMID: 22449779
  131. van Hall T, van der Burg SH. Mechanisms of peptide vaccination in mouse models: tolerance, immunity, and hyperreactivity. *Adv Immunol.* 114: Elsevier; 2012. p. 51–76. <https://doi.org/10.1016/B978-0-12-396548-6.00003-2> PMID: 22449778
  132. Van Der Burg SH, Melief CJ. Therapeutic vaccination against human papilloma virus induced malignancies. 2011; 23(2):252–7. <https://doi.org/10.1016/j.coi.2010.12.010> PMID: 21237632
  133. Vici P, Mariani L, Pizzuti L, Sergi D, Di Lauro L, Vizza E, et al. Immunologic treatments for precancerous lesions and uterine cervical cancer. 2014; 33(1):29.
  134. Melief CJM, van Hall T, Arens R, Ossendorp F, van der Burg SH. Therapeutic cancer vaccines. *The Journal of Clinical Investigation.* 2015; 125(9):3401–12. <https://doi.org/10.1172/JCI80009> PMID: 26214521