OncoPeptVAC™ accurately predicts neoantigens from tumors using a novel HLA-peptide-TCR-binding algorithm

Background

Cancer immunotherapy has established itself as a major therapeutic modality, and 70% of all cancer patients are pre-dicted to receive an immunotherapy drug by 2025. The remarkable efficacy of checkpoint inhibitors has opened new avenues for treating cancer, not available earlier. However, as more and more patients are treated with these drugs, the phenomenon of resistance, lack of responsiveness to treatment, poor durability of response and immune toxicity are emerging as key areas that need to be tackled to expand the benefit of the drugs to a larger patient population. Most tumors express T cell-specific neoantigens derived from intracellular proteolytic processing of somatic mutations in protein coding genes in the tumor (tumor-mutanome). These mutated peptides are expressed only by the tumor and not by the normal tissues, making them targets of cytotoxic T cells (CTLs) and therefore good vaccine candidates for eliminating tumor cells. There are three classes of tumor antigens that can be used as vaccines – tumor-specific anti-gens (TSA): genetically altered proteins perceived as foreign by the immune system; tumor associated antigens (TAA): proteins expressed at a higher level in tumor cells but are also expressed by normal cells at low levels and are not seen as foreign by the immune system; cancer testis antigens (CTA): have highly restricted tissue expression, expressed in tumors, but not in normal tissues. In humans, class-I peptides (9-11–mer in length) are presented by HLA – A, B and C alleles and activate CD8+ T cells. Class-II peptides are 14-17-mer, presented by DPA, DPB, DQA, DQB, DRA and DRB alleles and activate CD4+ T-cells. The process leading to a productive engagement of the HLA bound peptide to the CD8+ T cells is complex and has multiple steps. These include processing of the neoantigen by the proteasomal machinery to generate peptides, successful entry of the peptides into the endoplasmic reticulum where binding to the HLA class I occurs, followed by presentation of the HLA-peptide complex to the surface of the antigen presenting cells or tumor cells, where it engages with the T cell receptors expressed by CD8+ T cells. Most currently used in-silico can-cer vaccine prediction pipelines have low sensitivity and specificity because they rely on features associated with pre-sentation of the antigen on the surface of cells without considering features required for T cell receptor (TCR) binding. At MedGenome, we have developed OncoPeptVAC - a machine learning-based approach to identify features that favor TCR binding. These features were further confirmed by analyzing crystal structures of TCR and HLA-peptide complex present in the Protein Data bank.

Using exome and RNA sequencing data from tumor/normal pair, OncoPeptVAC identifies immunogenic peptides from neoantigens. The predicted immunogenic peptides are tested in an ex vivo dendritic cell – CD8+ T cell activation assay to assess their ability to generate T cells with cytolytic phenotype – by expression of IFNγ, for example.

In this white-paper we show a) data to demonstrate the superior performance of the OncoPeptVAC pipe-line compared to other neoantigen prediction pipelines that use HLA-binding to predict immunogenic peptides b) features of the 9-mer peptide that favors TCR-binding over those that do not and c) validation data of the accuracy of prediction using an in-house developed dendritic-cell - T cell activation assay.

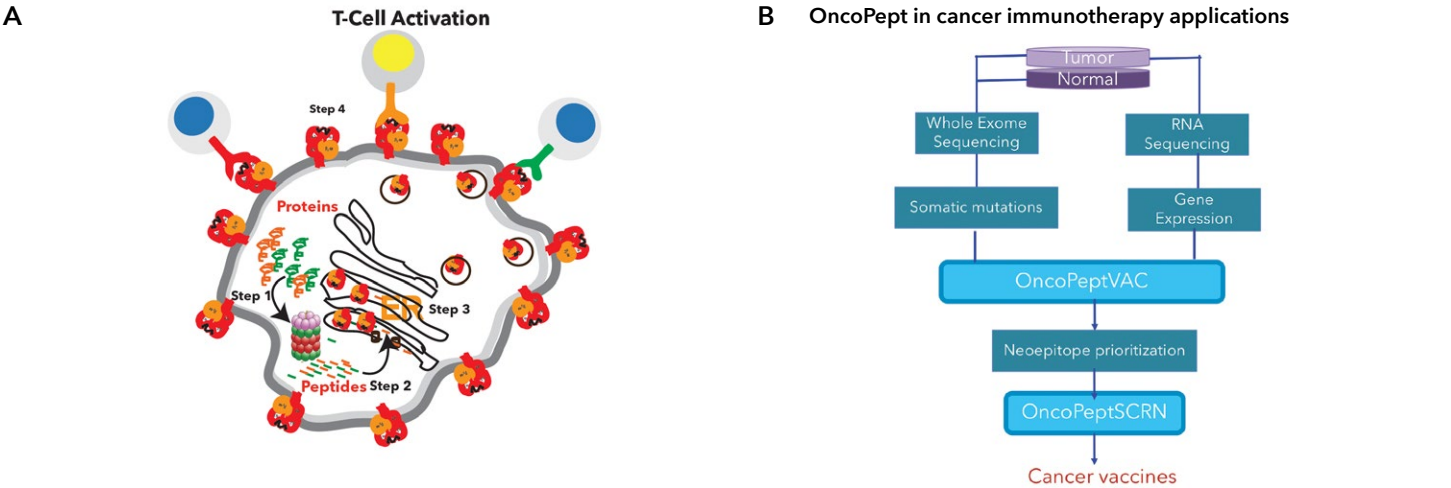


Figure 1A. *Antigen processing and cell surface presentation of the HLA-peptide complex is a multistep process. Regular turnover of proteins by proteasomal processing generates peptides that enter the endoplasmic reticulum (ER) by binding to the TAP transporter. Peptides bind HLA protein in the ER with a range of binding affinities) and the HLA-peptide complex is transported to the surface of the cells where they bind to TCRs expressed on CD8+ T-cell. B. Workflow of OncoPeptVAC for prediction and validation of immunogenic peptides from exome and RNA-seq data.*

Limitations in the current methods of neoepitope discovery

discoveryDeveloping a pipeline that assigns accurate prediction scores to each and every step in the neoantigen processing and presentation pathway is impossible [1], given that a very large number of peptides are competing at every step and the process of selection at these steps are poorly characterized. Therefore, most pipelines prioritize neoepitopes based on the expression of mutant proteins and their HLA binding affinity [2, 3]. Expression is usually predicted from mRNA levels of the wild-type and the mutant alleles and the binding affinity of each peptide is determined by NetMHCpan or NetMHCcons algorithms [4, 5]. Since productive engagement of HLA Class-I-bound peptide with the T cell receptor and activation of the CD8+ T-cells is the true hallmark of immunogenicity, we decided to build a pipeline that incorporates the TCR binding as a feature in predicting the neoantigens that would be immunogenic.

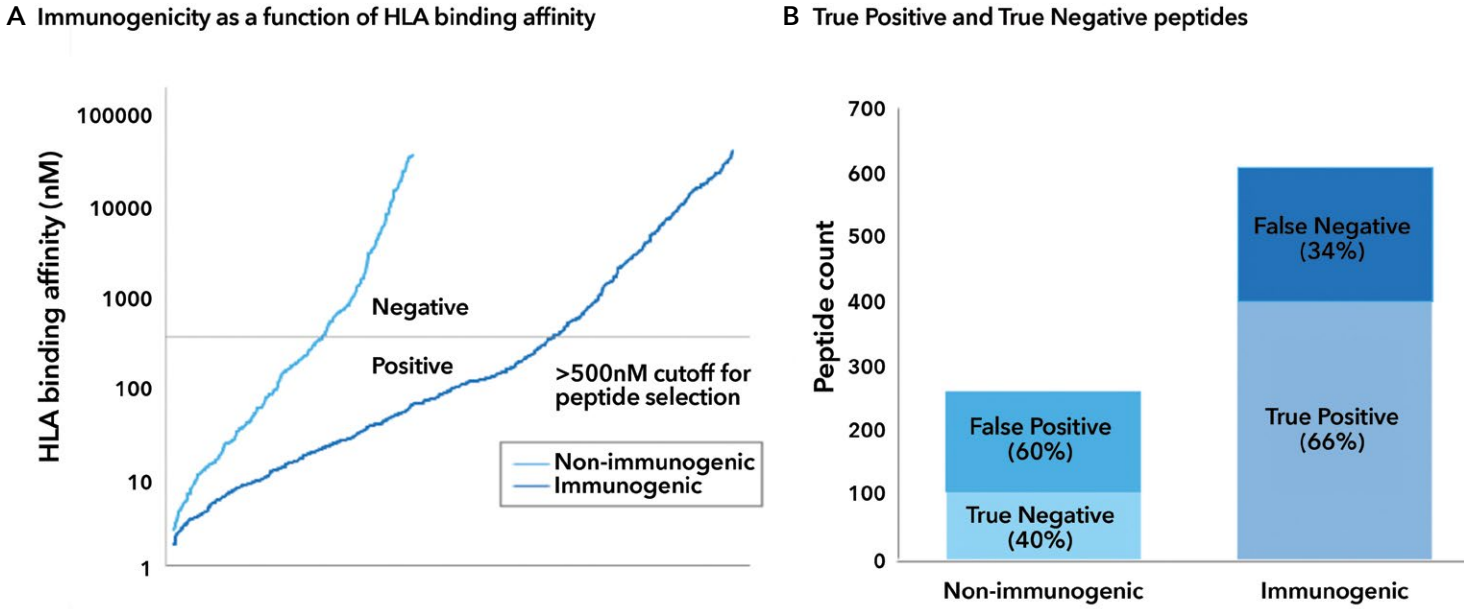
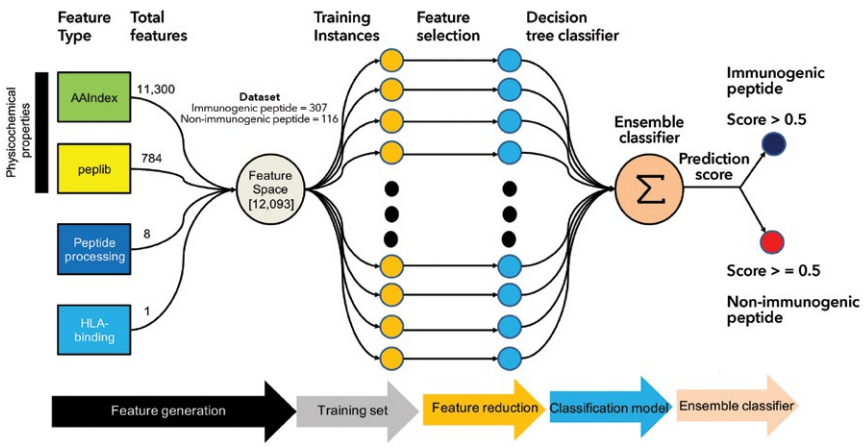


Figure 2: *Prediction of immunogenicity of peptides based on HLA binding : A. Shows the HLA binding affinity distribution of immunogenic and non-immunogenic 9-mer peptides. The HLA binding score is generated using NetMHCcons from IEDB. The grey line indicates the 500nM binding affinity. Affinity ≤500nM is used to define immunogenic peptides. B. Distribution of immunogenic & non-immunogenic peptides based on HLA binding shows 60% false positive and 30% false negative prediction.*

A



B

Performance metric	NetMHCcons	OncoPept*VAC*
Sensitivity (%)	74.50	90.23
Specificity (%)	27.59	99.14
Accuracy (%)	61.61	92.67

Figure 3: *A. Schematic of the workflow for separating immunogenic from the non-immunogenic peptides. Physicochemical properties taken from (AAIndex, Peplib), peptide processing features and HLA-binding affinity of the peptides generated 12,093 features for each 9-mer peptide. The dataset was sub-sampled and 500 training instances were generated with a balanced number (~100 in number) of immunogenic and non-immunogenic peptides. Feature reduction step was performed to reduce total features that will avoid overtraining by discarding correlated features. Decision tree-based classifier was used on the reduced features and the prediction from all classifiers was aggregated to generate an ensemble voting score for each peptide. Peptides with score > 0.5 were labeled as immunogenic. B. Table shows performance metrics of the HLA-binding and the ensemble classifiers.*

Building the OncoPeptVAC pipeline

The OncoPeptVAC pipeline combines TCR-binding with other features of the peptides associated with efficacy of processing and HLA binding. The TCR-binding algorithm was built on a set of non-immunogenic and immunogenic peptides curated from the IEDB data base. The database contains immunogenicity information of 2,752 unique 9-mer peptides from humans. We divided these peptides into self and foreign based on their matches to the reference human proteome. The peptides derived from mutated proteins or those derived from bacterial/viral proteins were considered as foreign, while the ones that matched the human proteome was considered self. Over 85% of the peptides belonged to the foreign category. We selected 116 non-immunogenic and 340 immunogenic peptides to build our algorithm. Next, to generate features that would be utilized for predicting the immunogenicity of the neoantigens, amino acid physiochemical properties, HLA binding, peptide processing features present in a 9-mer peptide were extracted to obtain a set of 12,093 features. To overcome the bias associated with unbalanced datasets of non-immunogenic to immunogenic peptides, we generated 500 different instances of the complete dataset by sub-sampling the overall set and created a balanced number of non-immunogenic and immunogenic peptides. Each balanced training set consisted of 80% of total non-immunogenic peptides and 30% of the total immunogenic peptides. The remaining datasets were kept as part of the test for evaluating the performance of the classifier.

Evaluation of the performance of the classification algorithm

The performance of the classifiers was evaluated by aggregating the individual classifier prediction for each peptide. A simple voting-based approach was used, where equal weight was given to each classifier output. If more than 50% of the classifiers predicted that the peptide is immunogenic, then we called it as immunogenic. In our initial evaluation of the classifier, the classifier built using all the features performed poorly on the unseen dataset, prompting us to do feature reduction using CfsSubsetEval method available in the Weka machine learning tool. Through a large number of iterations of the process, we narrowed the number of features to a total of 120 features enriched among the classifiers. By taking this approach we identified position-specific enrichment of features that favored TCR binding. The TCR-binding algorithm is integrated in the OncoPeptVAC cancer vaccine prediction pipeline along with other attributes required for antigen presentation (Figure 4).

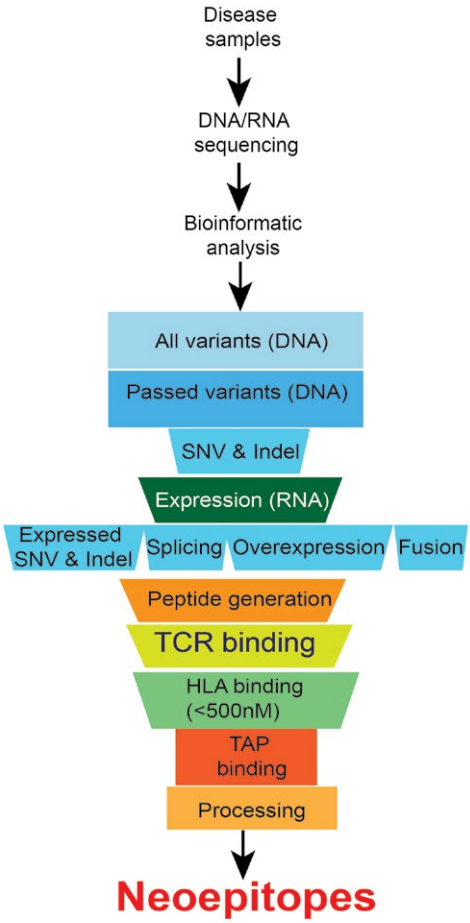
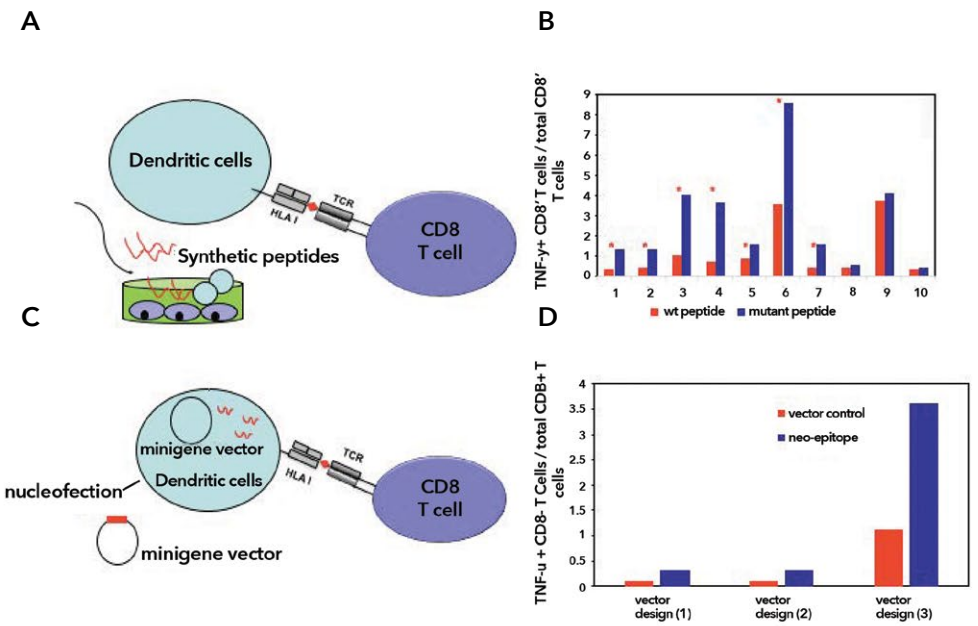


Figure 4: *Schematic representation of the OncoPEPTVAC : Workflow of OncoPeptVAC pipeline. Whole exome sequencing and RNA-seq are performed using the genomic DNA and RNA from the patient's tumor. All mutant peptides are passed through the OncoPeptVAC pipeline. The potential immunogenic peptides are predicted based on a series of prioritization steps including peptide processing, peptide TAP binding, HLA binding and TCR binding.*



E

Neoepitope candidates	Onco-PeptVAC prediction	Validation by T cell activation assay	Accuracy (%)
Immunogenic peptides	96	78	81.2
Non-immuno-genic peptides	9	7	77.7
Total	105	85	80.9

Figure 5: *Screening for immunogenicity of the vaccine candidates using the CTL assays on two different platforms : A) Schematic representation of the presentation of HLA- bound peptides for CD8+ T cell activation. B)Interferon gamma response in activated CD8+ T cells upon addition to wildtype or mutant peptides. C) Schematic representation of the presentation of HLA-bound peptides in dendritic cells that have been nucleofected with minigene vectors carrying neoepitopes and no-inserts for CD8+ T cell activation. D) Interferon gamma response in activated CD8+ T cells in different configurations of the minigenes introduced in the dendritic cells.*

Development of a mini-gene platform to screen for vaccine candidates identified using the OncopeptVAC pipeline

To screen vaccine candidates predicted to be immunogenic by the OncoPeptVAC pipeline, we have built two cell-based assays to examine antigen-induced T cell activation. The predicted antigenic peptides can be added from outside into a dendritic cell – T cell co-culture assay [6] (Figure 5A) or expressed as minigenes in the dendritic cells [7] (Figure 5B). The assay is initiated up by differentiating monocytes isolated human peripheral blood mononuclear cells (PBMCs) into mature dendritic cells (DCs). Mature DCs are incubated with synthetic peptides and then co-cultured with autol-ogous CD8+ T cells isolated from the same human PBMC. (Figure 5A). After incubation with the peptides activated T cells are quantitated by their expression of IFNγ by flow cytometry. We tested 10 predicted immunogenic peptides in this assay using their wildtype counterparts as controls and observed 7 of the 10 peptides to induce T cell activation (Figure 5B). While the peptide assay allowed us to validate the positive peptides for their immunogenicity, caveats of the peptide assay are that a) these candidates bypass the intracellular processing steps and may get loaded onto the MHC molecules even with weaker binding affinities b) peptides with secondary modifications that impart their immu-nogenicity can be missed. c) wild-type peptides may non-specifically stimulate T cells when added at non-phys-iological concentrations. To overcome some of these limitation, we developed a minigene platform that would allow testing the immunogenicity of a peptide under more physiological conditions. By utilizing this approach, we demon-strate antigen-specific T cell activation (Figure 5B), suggesting that the minigene assay can be utilized for screening neoantigen peptide candidates for development of vaccines for cancer immunotherapy.

References

1. Gubin, M.M., et al., Tumor neoantigens: building a framework for personalized cancer immunotherapy. J Clin Invest, 2015. 125(9): p. 3413-21.
2. McCarthy, M.K. and J.B. Weinberg, The immunoproteasome and viral infection: a complex regulator of inflammation. Front Microbiol, 2015.6: p. 21.
3. Hundal, J., et al., pVAC-Seq: A genome-guided in silico approach to identifying tumor neoantigens. Genome Med, 2016. 8(1): p. 11.
4. Nielsen, M. and M. Andreatta, NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiplereceptor and peptide length datasets. Genome Med, 2016. 8(1): p. 33.
5. Liepe, J., et al., A large fraction of HLA class I ligands are proteasome-generated spliced peptides. Science, 2016. 354(6310): p. 354-358.
6. Wolfl and Greenberg. 2014. Antigen-specific activation and cytokine-facilitated expansion of naive, human CD8+ T cells. Nat Protoc. 9(4): 950–966.
7. Aurisicchio et al. 2014. A novel minigene scaffold for therapeutic cancer vaccines. OncoImmunology. 195:3724–3733.

MEDGENOME INC.
348 Hatch Drive,
Foster City,
CA 94404, USA
(888) 440-0954
research@medgenome.com
www.medgenome.com