OncoPept*VAC* : A machine learning based approach for candidate vaccine identification and their validation using cell based assays

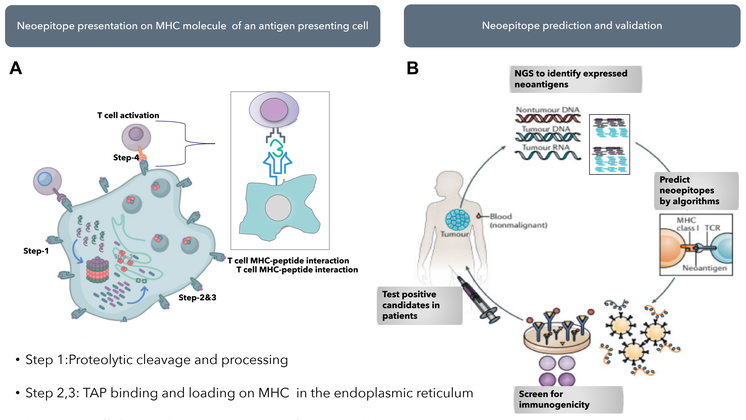
Ankita Das[1], Priyanka Shah[2], Xiaoshan "Shirley" Shi[1], Vasumathi Kode[1], Kayla Lee[1], Ravi Gupta[2], Amit Chaudhuri[1,2] and Papia Chakraborty[1]

1) MedGenome Inc, Foster City, CA, 2) MedGenome Labs, Bangalore, India; corresponding author : ankita.d@medgenome.com

Abstract

- T cell immunity provides significant therapeutic benefit to cancer patients treated with checkpoint inhibitors. Most tumors harbor a repertoire of somatic mutations, a fraction of which is capable of initiating potent T cell mediated anti-tumor activity. However, accurate identification of relevant immunogenic neoantigens remain a major challenge in therapeutic cancer vaccine research.
- Current *in silico* methods to predict immunogenic neoantigens suffer from lack of sensitivity and specificity because they rely heavily on features associated with antigen presentation alone, without considering features required for T cell receptor (TCR) binding.
- Here we report OncoPept*VAC,* an algorithm based on ensemble voting-based machine learning approach to identify immunogenic peptides from patient's somatic mutations. Our method combines physicochemical properties of amino acids favorable for TCR binding with features relevant for antigen presentation and processing.
- In alignment with crystal structure of MHCpeptide-TCR complex our model reveals enrichment of helix/turn features at TCR contact residues along with hydrophobicity features enriched at the HLA-binding anchor residues.
- Cell-based immune co-culture assays downstream of OncoPept*VAC* shows both monoclonal and polyclonal TCR responses. In addition, a considerable fraction of peptides displayed an inhibitory effect on interferon production by CD8 T cells.
- OncoPept*VAC* is trained on MHC Class I HLA-A*02:01 restricted 9-mer peptides available in IEDB data. The validation of the model on unseen peptides provided sensitivity and specificity of 81% and 78% respectively.

Origin of neoepitopes and their use as cancer vaccine



- Step 1:Proteolytic cleavage and processing
- Step 2,3: TAP binding and loading on MHC in the endoplasmic reticulum
- Step 2,3: Trafficking and presentation on surface on APC
- Step 4: TCR recognition and activation
- Inset Turn in the MHC-bound peptide (green) is critical for TCR binding

Modified from Berger et al, Nature reviews, Clinical Oncology, June 2018

Figure 1: A. Shows the biological steps of antigen presentation by human leukocyte antigen (HLA) class I molecules. Proteins are processed by either proteasome or other proteases to generate small peptides. Transporter associated with antigen processing (TAP), a protein that spans the membrane of the endoplasmic reticulum (ER), transports the peptides into the ER. HLA I binds with the peptides and transports them to the cell surface. Stable interaction of the HLA I complexed peptides with TCRs on CD8 T cells induces T cell activation. B. Highlights the key steps involved in personalized cancer vaccine identification. NGS is performed on patients tumor to identify expressed mutations. Patient blood samples are sequenced in parallel to parse out the somatic mutations in the patients tumor by subtraction analysis. The expressed mutations are funneled through in silico predication algorithms to prioritize immunogenic peptide candidates wherein 90% of the peptide candidates are eliminated. Cell-based assays are then performed on the prioritized candidates to obtain the final list of validated peptides. These assays involve measurement of CD8 T cell activation in presence of peptides loaded onto antigen presenting cells.

Workflow of OncoPept*VAC* and OncoPept*SCRN* to predict and validate immunogenic neoantigens

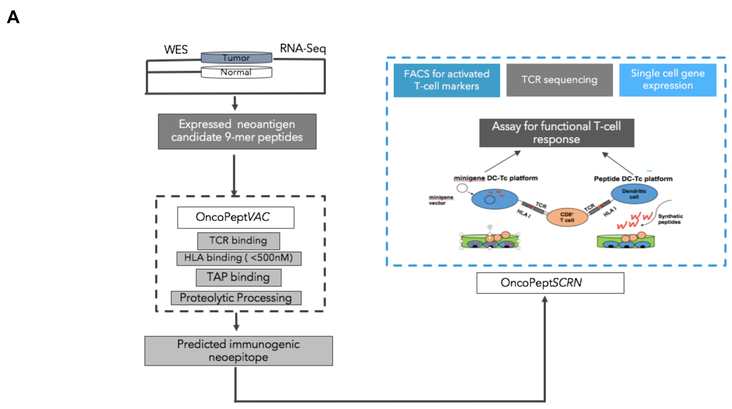


Figure 2: A. Workflow of OncoPept*VAC* pipeline. Whole exome sequencing and RNA-seq are performed using the genomic DNA from blood and RNA from the patient's tumor. All mutant peptides are passed through the OncoPept*VAC* pipeline. The potential immunogenic peptides are predicted based on a series of prioritization steps including peptide processing, peptide TAP binding, HLA binding and TCR binding. Finalized peptides are then subjected to OncoPept*SCRN* wherein peptides are tested for CD8 T cell activation assay in presence of antigen loaded APCs. Functional TCR cell response is assessed by IFN-γ production by flow cytometry, TCR repertoire sequencing and single cell gene expression analysis.

OncoPept*VAC* pipeline design and concordance of peptide features with structure of HLA-peptide-TCR complex

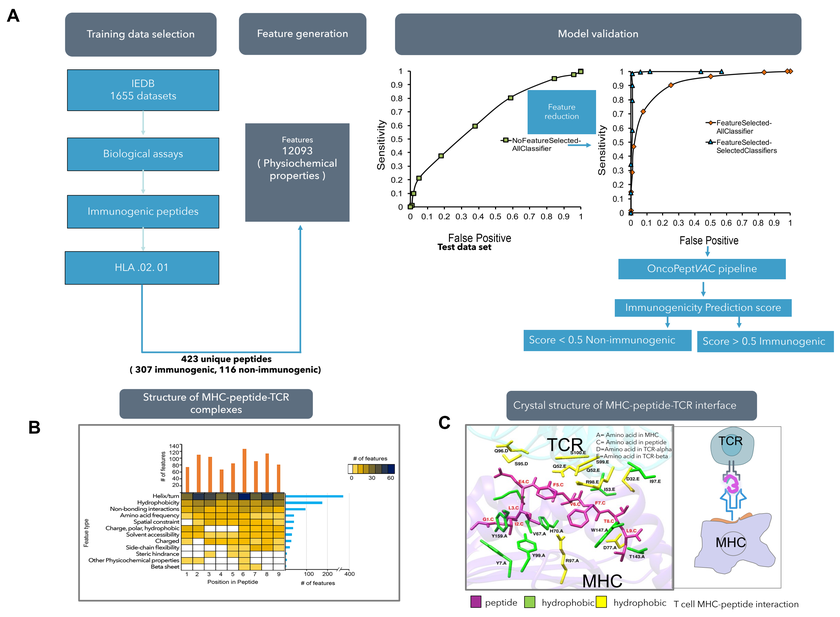
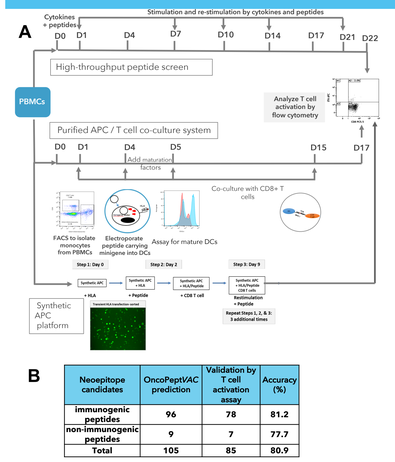


Figure 3: A. The methodology used for development of OncoPept*VAC*:12,093 features for a 9-mer peptide were identified from IEDB based on their physicochemical, peptide processing and HLA-binding properties. The data was subsampled into 500 training sets with a dataset of 100 data points per subsampling (balanced by immunogenic vs non-immunogenic). A feature reduction step was initially used followed by a decision tree based classifier to generate an ensemble voting score for each peptide. Peptides with score > 0.5 were labeled as immunogenic. B. Heatmap showing the selected features by the better performing classifiers. The most frequency feature type includes Helix/turn, hydrophobicity, and non-bonding interactions. C. 9mer peptide (shown in purple) and the MHC TCR interface. The hydrophobic residue in the MHC molecule is shown in green.

Engagement of TCR with MHC-peptide complex is not always coupled with INFγ production



B

Neoepitope candidates	OncoPept*VAC* prediction	Validation by T cell activation assay	Accuracy (%)
immunogenic peptides	96	78	81.2
non-immunogenic peptides	9	7	77.7
Total	105	85	80.9

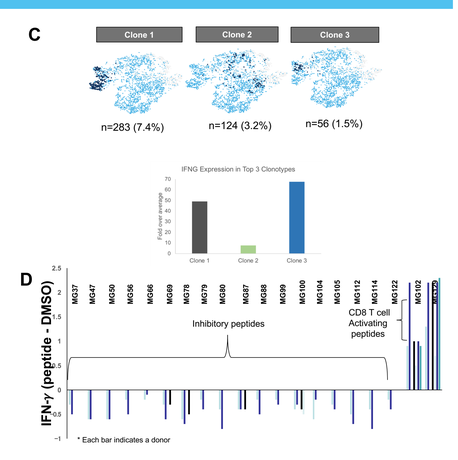
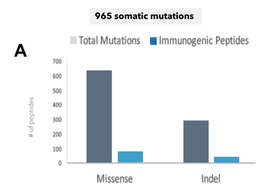


Figure 4*:* A. Cell based immunogenicity validation by OncoPept*SCRN*, and outcomes: A. Shows three different assay formats, for CD8 T cell activating antigen screens using monocyte derived dendritic cells or genetically engineered APCs or crude PBMCs. B. Inset table shows prediction accuracy of OncoPept*VAC*. C. TCR repertoire analysis coupled to single cell gene expression studies reveal that antigen induced TCR expansion is not always coupled to IFN-γ production. C. Cell based immune assays confirm the presence of CD8 T cell inhibitory peptides as evidenced by significant reduction in IFN-γ levels below control in more than one donor. Cell based validations are necessary for candidate vaccine identification.

Applying OncoPept*VAC* and OncoPept*SCRN* to Lynch syndrome



B

Gene	HLA	Affinity(nM) WT	Affinity(nM) MUT	TCR Prediction
MSH6	HLA-A02:06	71.02	8.61	High
PIGO	HLA-A02:06	32787.64	10.3	High
AXIN2	HLA-A02:06	12860.25	425.67	High

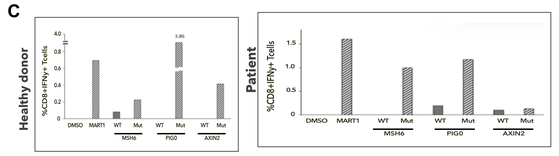


Figure 5: A. Clinical translatability of OncoPept*VAC* and OncoPept*SCRN*. A. OncoPept*VAC* was utilized to prioritize candidate peptide vaccines for a MSI(High) Lynch Syndrome patient. 965 somatic mutations were screened to prioritize ~150 missense and indels. B. Top candidates were further chosen for cell based assays based on functional pathways dominant in Lynch Syndrome etiology. C. All three prioritized peptides elicit a CD8 activation in healthy donor, where only two of the three candidates (MSH6 and PIGO) elicit CD8 T cell activation in the patient indication tolerance mechanisms may be in place.

Conclusions

- Personalized cancer vaccines derived from somatic mutations of a patients tumor is a new therapeutic modality. Machine learning algorithms can predict immunogenic features of peptides/vaccine candidates that interact with the MHC Class I and TCR interphase. Integration of MHC-peptide complex with TCR binding yields higher accuracy of cancer vaccine prioritization.
- *In vitro* cell-based assays can to used to validate cancer vaccine candidates. Antigen induced CD8 T cell expansion is not always coupled to *IFN-γ* production. MHC Class I peptides can elicit both an activating or inhibiting CD8 T cell response.
- OncoPept*VAC* and OncoPept*SCRN* can be used to predict and validate a cocktail of cancer vaccine for personalized cancer vaccine therapy.