



Category: Bioinformatics

# An accurate machine learning approach to predict immunogenic peptides in human

Priyanka Shah<sup>1</sup>, Anand Kumar Maurya<sup>1</sup>, Rohit Gupta<sup>1</sup>, Amit Chaudhuri<sup>2</sup>  
and Ravi Gupta<sup>1\*</sup>

<sup>1</sup>MedGenome Labs Pvt. Ltd., Bangalore, INDIA

<sup>2</sup>MedGenome Inc., Foster City, USA

\*Corresponding author: [ravig@medgenome.com](mailto:ravig@medgenome.com)

## Abstract

Cancer immunotherapy provides durable response to a small subset of treated patients. A variety of approaches are being developed to increase the long term benefit of checkpoint blockade. These include radiation and cytotoxic therapies and use of cancer vaccines among others. Preclinical and clinical studies have demonstrated that cancer vaccines evoke strong anti-tumor immune response by mobilizing CD8<sup>+</sup> T-cells. A challenge in the field of cancer vaccines is identifying mutations that are T-cell activating (neopeptides). Advances in next generation sequencing permit accurate detection of cancer mutations, even when present at a low frequency. However, neopeptide prediction involves a large number of steps many of which cannot be accurately modeled. In humans, class-I peptides, 9-11-mer in length are presented by HLA – A, B and C alleles and activate CD8<sup>+</sup> T-cells. Class-II peptides are 14-17-mer, presented by DPA, DPB, DQA, DQB, DRA and DRB alleles and activate CD4<sup>+</sup> T-cells. The current method to identify epitopes (peptides) depends primarily on HLA binding prediction algorithm. Our analysis of 9mer peptides from IEDB database showed that there is no difference in binding affinity of peptide that can activate (immunogenic) and that cannot activate (non-immunogenic) the T-cells. The specificity to identify immunogenic peptide using HLA binding based method is only 27.59%. In this study, we present a novel approach using machine learning technique that can predict whether the peptide will be immunogenic or not. We have generated the model using features generated from amino-acid composition, HLA-binding, structural features, peptide processing and peptide transport. Our model achieved an overall accuracy of 77.30% with a specificity of 92.24% on unseen dataset.

**Citation:** Shah, P., Maurya, A.K., Gupta, R., Chaudhuri, A. and Gupta, R. An accurate machine learning approach to predict immunogenic peptides in human [Abstract]. In: Abstracts of the NGBT conference; Oct 02-04, 2017; Bhubaneswar, Odisha, India: *Can J Biotech*, Volume 1, Special Issue (Supplement), Page 224. <https://doi.org/10.24870/cjb.2017-a209>