

**Invitation to M.Tech. Thesis Defense of Sumeet Patiyal: May 30, 2022 (Monday): 11:00 AM-11.45 AM IST**

In Partial Fulfilment of the Requirements for the Degree of

**M.Tech. CB**

**Sumeet Patiyal (PhD17204)**

Will defend his thesis

**Title: “AI based prediction of HLA-DRB1\*04:01 binder for designing subunit vaccines”**

IIIT-D Faculty and Students are invited

**Date: May 30, 2022 (Monday)
Time:** **11.00 AM-11.45 AM IST**

**Venue: Room No: A-320 (Meeting Room), R & D Building**

**Examiner: Internal:   Jaspreet Kaur Dhanjal**

**External/~~Internal~~: Manish Kumar (University of Delhi)**

**Advisor: G P S Raghava**

**Co-Advisor NA**

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Abstract**

HLA gene complex is a highly polymorphic region in the human genome and mutations associated with these regions can lead to many deadly disorders such as bare lymphocyte syndrome, whereas presence of few HLA-class II alleles makes an individual more prone to some diseases. One of these class-II alleles named HLA-DRB1\*04:01 is associated with many autoimmune disorders such as multiple sclerosis, rheumatoid arthritis, type 1 diabetes, Lyme disease, etc. Moreover, a particular variant of HLA-DRB1\*04:01 gene is found to be abundant in the asymptomatic carriers of SARS-CoV-2. Hence, it is the need of the hour to develop a more accurate method with the ability to classify HLA-DRB1\*04:01 binding peptides. We have developed a systematic approach to predict, scan, and design the binders of class-II HLA allele HLA-DRB1\*04:01 and provided as a webserver. It is an updated version HLADR4Pred developed in year 2004. In this study, we have compiled the positive (HLA-DRB1\*04:01 binder) and negative dataset (HLA-DRB1\*04:01 non-binder) from IEDB. We have a total of 12676 peptides in the positive and 86300 peptides in the negative dataset. At first, we generated composition and binary profile based features using the Pfeature standalone package. After that we have implemented various machine learning techniques to develop prediction models by using different types of features. Secondly, we have segregated the complete dataset into training and validation dataset, where training dataset comprises 80% of the complete dataset and the remaining 20% was assigned as validation dataset. We have trained the models on the training dataset by applying a five-fold cross validation technique and performed external validation by evaluating our models on the validation dataset. Number of performance measures have been calculated to assess the performance of each model developed on different features. We observed that the extra tree based classifier using dipeptide composition outperformed other classifiers and achieved maximum AUROC of 0.96 on both training and validation dataset. After that, we have combined similarity search using BLAST with our best performing model to develop the hybrid method, which attains the highest performance i.e. AUROC of 0.98 and 0.99 on training and validation dataset, respectively. Finally, we have incorporated the hybrid model in our webserver named HLADR4Pred2 available at https://webs.iiitd.edu.in/raghava/hladr4pred2/. Along with that we have also provided the python- and Perl based standalone package which is available at webserver (https://webs.iiitd.edu.in/raghava/hladr4pred2/standalone.php) and at GitHub (https://github.com/raghavagps/hladr4pred2).