**OnionMHC: a deep learning model for peptide - HLA-A\*02:01 binding predictions using both structure and sequence** **feature sets**

Shikhar Saxena1, Sambhavi Animesh2, Melissa J. Fullwood1,2, and Yuguang Mu\*1

1  School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore, 637551

2 Cancer Science Institute, National University of Singapore, 14 Medical Drive, Singapore, 117599.

\*Corresponding author: ygmu@ntu.edu.sg

**Supporting Information**

**Hyperparameter Tuning**

The hyperparameters for the OnionMHC were optimized and evaluated using the five-fold cross-validation with the area under the receiver operating curve (AUC) and Spearman Rank correlation coefficient (SRCC) used as the metric for comparison (Fig. S1). The structural module was kept almost the same as OnionNet, however, we optimized the hyperparameters in the sequence module such as the number of layers, number of neurons. OnionMHC with the sequence module containing 2 lstm layers achieved the best performance as compared to the module containing 1 or 3 lstm layers. Also in terms of the number of neurons, the lstm layer with 1000 neurons achieved the best performance on the five-fold cross-validation. We also optimized the batch size used for training the model. Batch sizes of 32 and 64 achieved almost comparable performance on the five-fold cross-validation with batch-size of 64 performing slightly better.

Figure S: Performance evaluation of OnionMHC on different hyperparameters a) number of lstm layers in sequence module b) number of neurons in lstm layer in sequence module c) batch size used for training