1. Purpose

This document describes the process by which new CTL epitopes are generated for Influenza virus sequences stored in BioHealthBase.

2. Method Description

BioHealthBase NetCTL data production pipeline is composed of three JAVA scripts to semi-automate the submission of sequences to the **NetCTL** web server at the Technical Danish University (TDU) center for CTL epitope predictions. The NetCTL method integrates prediction of peptide MHC class I binding, proteasomal C terminal cleavage and TAP transport efficiency. The server allows for predictions of CTL epitopes restricted to 12 MHC class I supertype. MHC class I binding and proteasomal cleavage is performed using artificial neural networks. TAP transport efficiency is predicted using weight matrix.

3. Input Data Preparation

Influenza sequences in BioHealthBase are partitioned into group of files each containing 100 FASTA sequences in it. This is because of the limit set by TDU server for a maximum of 100 sequences in each file submitted. These files were then submitted to the NetCTL web server sing a Java script. This Java script allows one to submit a controlled number of input files to TDU server because of the limit of 50 submitted files per day. Each submission with PostData is specific to one epitope supertype. BioHealthBase currently provides data for five supertypes: A2, A3, A24, B44 and B7

4. Output Data, Processing and Display

All the data returned from TDU server are grouped by MHC supertypes. These data are then post-processed to insure the data conformity for database loading. The NetCTL-predicted epitopes are displayed in both Gene Details page and the genome browser on BioHealthBase website for the public use. A database search tool called “Epitope Search” is provided for users to retrieve epitopes corresponding to specified genes.

Here is an example of NetCTL Epitope Summary table in Gene Details page:

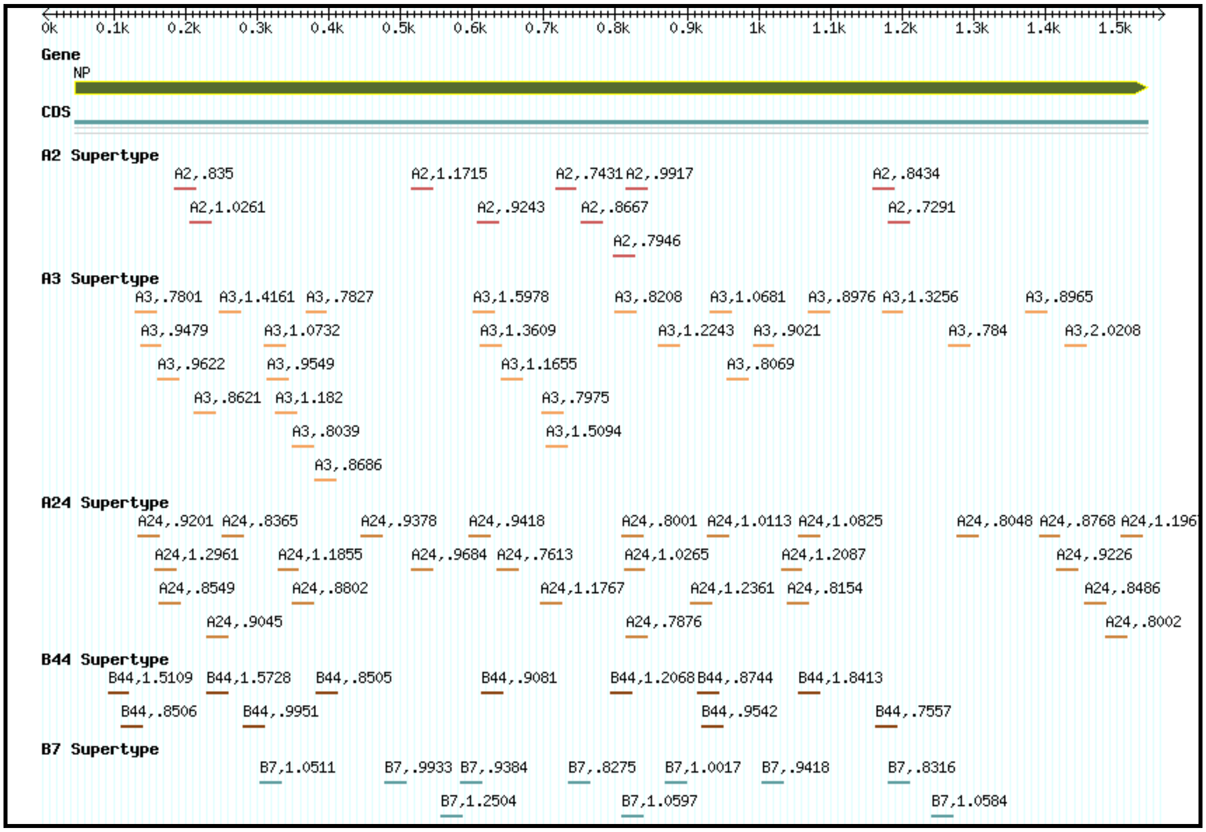
Prediction Details

MHC Supertype	# Predictions
A2	10
A24	14
A3	10
B44	10
B7	10
Total	54

Users can get access to Epitope Details page by clicking the link “Prediction Details”. As shown below is a portion of NetCTL predictions in Epitope Details page:

MHC Supertype	Peptide Sequence	AA Start	AA Stop	Score
A2	LLTEVETYV	3	11	1.3144
A2	ALMEWLKTR	41	49	0.7443
A2	ILSPLTKGI	51	59	1.0338
A2	GILGFVFTL	58	66	1.034
A2	LIYNRMGTV	130	138	0.7381
A2	RMGTVTTEV	134	142	0.956
A2	QMATTTNPL	164	172	0.8398

NetCTL data are also displayed in the genome browser, with each supertype per track, grouped into a “Predicted MHC Epitopes” category.



5. References

1. **An integrative approach to CTL epitope prediction. A combined algorithm integrating MHC-I binding, TAP transport efficiency, and proteasomal cleavage predictions.**
   Larsen M.V., Lundegaard C., Kasper Lamberth, Buus S,. Brunak S., Lund O., and Nielsen M.
   European Journal of Immunology. 35(8): 2295-303. 2005

2. **Reliable prediction of T-cell epitopes using neural networks with novel sequence representations.**
   Nielsen M, Lundegaard C, Worning P, Lauemoller SL, Lamberth K, Buus S, Brunak S, Lund O.
   Protein Sci., 12:1007-17, 2003.
3. **Sensitive quantitative predictions of peptide-MHC binding by a 'Query by Committee' artificial neural network approach.**
   Buus S, Lauemoller SL, Worning P, Kesmir C, Frimurer T, Corbet S, Fomsgaard A, Hilden J, Holm A, Brunak S.
   Tissue Antigens., 62(5):378-84, 2003.
4. **Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors**
   Peters, B., Bulik, S., Tampe, R., Endert, P. M. V. and Holzhutter, H. G.
   J. Immunol. 171: 1741-1749, 2003.