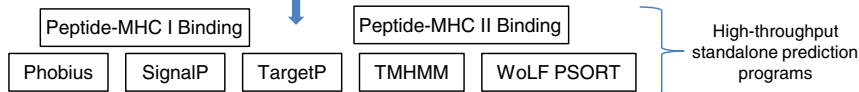


Protein Sequences

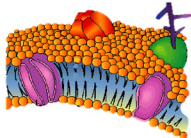
MAAQLLSMSEIEGPEENEN _ _ _

Input



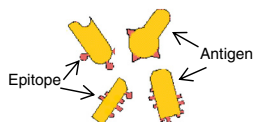
Clues for protein characteristics

Clue for membrane protein



Clue for secreted protein

Clue for epitopes



Create evidence profile for each protein

1=ID,2=Phobius_TM,3=Phobius_SP,4=SignalP,5=TargetP_SP,6=TargetP_loc,7=TargetP_RC,
 8=TMHMM_AA,9=TMHMM_First60,10=TMHMM_TM,11=WoLF_PSORT,
 12=WoLF_PSORT_annotation,13=MHCI,14=MHCII,15=Expected classification
 Q9UB12,0,Y,0.495,0.708,S,4,13.40,9.13,0,23.0,Secreted,0.535,0.335,YES
 B6KC27,8,N,0.141,0.040,M,4,155.01,1.86,7,26.0,Membrane,0.672,0.861,YES
 B6KDT4,1,N,0.116,0.026,M,5,38.65,0.51,1,0,NOT_secreted_or_membrane,0.614,0.331,YES
 B6KHJ3,0,N,0.108,0.021,U,4,1.97,0.00,0,5.0,Membrane,0.840,0.571,NO
 B6KPE4,1,N,0.101,0.050,U,1,14.26,0.00,0,0,NOT_secreted_or_membrane,0.533,0.205,NO

A collection of evidence profiles where each value in the profile is a clue extracted from a prediction program output

Input

Support vector machines

Naïve Bayes

K-Nearest neighbour

Adaptive boosting

Neural networks

Random forest

A pool of machine learning algorithms for classification

Vaccine candidate ?

NO

YES