22145. Immunological Bioinformatics written exam. January 2021.

Read carefully the guidelines and the questions:

- **Upload a pdf** with the answers before 15.00 to DTU inside.
- Answer each question in **no more than 5 lines** (min. font size 10).
- Remember it is an **individual exam** and that you have signed a code-of-honor Copying or sharing responses will be severely punished.

1. Regarding the exercise "one-gene-one-pathogen is obviously not possible". How is it possible for our immune system to mount an immune response to the many hundred thousand different pathogens encountered during a lifetime?

2. Arrange in chronological order the different processes regarding the humoral immune response:

a. TCR Recognition
b. BCR Recognition
c. Antigen processing
d. Peptide presentation on MHC
e. Antibody production
f. VDJ recombination of the TCR and BCR

3. You have located a partly completed document with the calculations for constructing a weight matrix (PSSM) for an HLA molecule with the information for the C terminal position included below

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
p()	0.060	0.038	0.033	0.024	0.017	0.022	0.028	0.029	0.012	0.109	0.165	0.035	0.060	0.037	0.018	0.045	0.087	0.006	0.020	0.155
q()	0.074	0.052	0.045	0.054	0.025	0.034	0.054	0.074	0.026	0.068	0.099	0.058	0.025	0.047	0.039	0.057	0.051	0.013	0.032	0.073

Given this information, which of the 3 peptides below is most likely to bind the receptor (justify your answer)

a) SIFPETAKL
b) SIFPETAKM
c) SIFPETAKR

4. Mention the three methods used to improve the accuracy and predictive power of a weight matrix introduced in the course and explain briefly when each method is important.

5. Does the artificial neural network below capture the XOR function (explain you answer)?

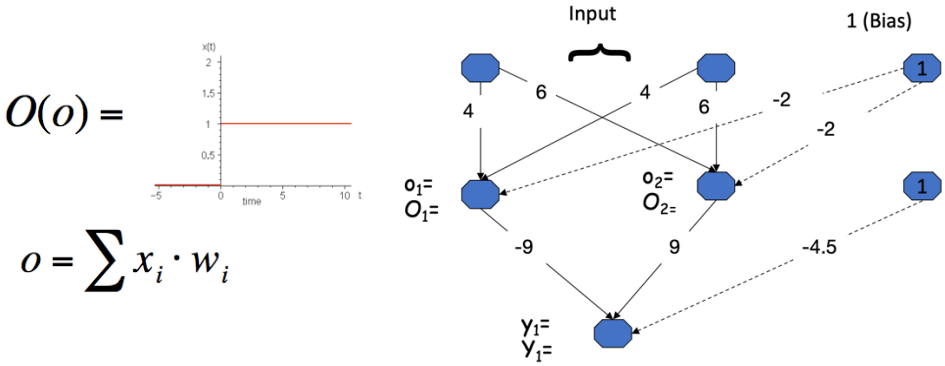

Figure ANN. An artificial neural network. The artificial neural network uses a step function with transition value at 0 to calculate the output from the neurons. The step function is given by O(o) = 1 if o>0 otherwise 0.

6. Answer in one-line the following questions

1) What type of machine learning method is used to train the NetMHCpan and NetMHCIIpan predictors?
2) What type of data are NetMHCpan-4.1 and NetMHCIIpan-4.0 trained on?
3) What do the methods predict?
4) Do the predictions include information related to HLA antigen processing (justify you answer)?

7. Does the inclusion of proteasomal cleavage and TAP transport predictions improve the accuracy for prediction of HLA ligands and T cell epitopes? Explain why.

8. Mention at least two prediction methods introduced in the course that could be used together with NetMHCpan (or NetMHC) to improve the accuracy of CD8 epitope identification. Justify.

9. As seen in the course, predictions of HLA class I antigen presentation are highly accurate. However only a small proportion of the predicted ligands turn out to be epitopes. Mention three reasons for this. Briefly justify.

a)
b)
c)

10. Say you have used the methods introduced in the course to identify a set of 10 potential epitopes. When constructing a CD8 epitope polytope vaccine from these, pick the most important feature from the following. Justify your answer.

a) Ensure that the epitopes maintain binding to HLA when included in the polytope
b) Ensure that the epitopes are processed out of the polytope
c) Ensure that novel neo-epitopes are not formed in the junction of two epitopes
d) Ensure that the polytope contains a strong B cell epitope

11. Name one example of a supervised and an unsupervised machine learning method utilized during the course and comment which could be the advantage of utilizing each in the context of immunopeptidomics.

12. You have a reasearch project idea to create a T-cell based vaccine against Malaria. How could you use the IEDB database to guide your initial steps of that project?

13. When performing sequence based protein drug deimmunization our goal is to introduce mutations that reduce immune recognition. This is commonly done by mutationally targetting CD4+ epitopes. Why do we target CD4+ epitopes and not CD8+ or B Cell epitopes?

14. Deimmunization of protein drugs is a field of protein engineering where protein drugs are made safer by reducing Anti-Drug Antibodies. This can be done by mutating the protein sequence and educated guesses can be made with in silico models. What are the two major aspects/concerns/factors you must balance when performing sequence based deimmunization?

15. Why do we believe that single-cell data is key to developing a model for prediction of T-cell epitopes?

16. What are the three broad categories of methods used for comparing TCR repertoires? Can you name one algorithm example for each of them?

17. What is the downside of comparing TCR sequences using alignment and substitution matrix scoring (i.e. BLOSUM62, PAM10, etc)?

18. What is a CDR3 and how did we process the sequences in the exercise such that they were represented in the same format? Exercise: Single cell technologies for T-cell epitopes – Prepare and analyze the data.

19. Describe the steps of antibody modeling, including canonical structures, the maximum number of templates that can be used, and for which regions such templates are used

20. What is antibody humanization and how is it performed?

21. These are peptides derived from the humanization and deimmunization process of a CAR construct. The **murine** antibody contains the peptide AVQLDNGRS, which is predicted to have a rank score of 0.04 for a given MHC class I allele of interest. The corresponding peptide in the **human** framework used for the humanization is ANQIDEGRN, with a predicted rank score for the same allele of 0.3. The peptide in the proposed **humanized** sequence is AVQIDEGRN, with a predicted rank score of 0.03. Answer to the following questions:

    a) Which of these 3 peptides is a strong binder for the allele of interest?
    b) Which is the backmutation in the humanized peptide?
    c) Does this backmutation introduce a high, low, or negligible risk of immunogenicity, and why?