MixMHC2pred

MixMHC2pred is a predictor of HLA class II ligands and epitopes. It is described in the publication (available here):
Racle, J., et al. Robust prediction of HLA class II epitopes by deep motif deconvolution of immunopeptidomes. *Nat. Biotechnol.* 37, 1283–1286 (2019).

MixMHC2pred is also available as a web application: http://mixmhc2pred.gfellerlab.org.

Installation

1. Download MixMHC2pred-1.2.zip file and move it to a directory of your choice, where you have writing permissions.

2. Unzip MixMHC2pred-1.2.zip in this directory.

3. To test your installation, make sure you are in *MixMHC2pred-1.2* directory and run the following command, depending on your operating system:

   - Mac OS:

     - `./MixMHC2pred -i test/testData.txt -o test/out.txt -a DRB1_11_01 DRB3_02_02 DPA1_01_03__DPB1_04_01 DQA1_05_05__DQB1_03_01`

     or to test phosphorylated ligands

     - `./MixMHC2pred -i test/testData_phospho.txt -o test/out_phospho.txt -a DRB1_01_01 DPA1_01_03__DPB1_04_01 DPA1_01_03__DPB1_06_01`

   - Unix:

     - `./MixMHC2pred_unix -i test/testData.txt -o test/out.txt -a DRB1_11_01 DRB3_02_02 DPA1_01_03__DPB1_04_01 DQA1_05_05__DQB1_03_01`

     or to test phosphorylated ligands

     - `./MixMHC2pred_unix -i test/testData_phospho.txt -o test/out_phospho.txt -a DRB1_01_01 DPA1_01_03__DPB1_04_01 DPA1_01_03__DPB1_06_01`

   - Windows:

     - `MixMHC2pred.exe -i test/testData.txt -o test/out.txt -a DRB1_11_01 DRB3_02_02 DPA1_01_03__DPB1_04_01 DQA1_05_05__DQB1_03_01`

     or to test phosphorylated ligands

     - `MixMHC2pred.exe -i test/testData_phospho.txt -o test/out_phospho.txt -a DRB1_01_01 DPA1_01_03__DPB1_04_01 DPA1_01_03__DPB1_06_01`

Your file *test/out.txt* should be the same as *test/out_compare.txt*. Running the software should take only few seconds.

The *testData.txt* file corresponds to HLA-II peptidomics data obtained in our study from the cell line *CD165* (it contains 8715 unique peptides).

4. (Optional) To run MixMHC2pred from anywhere on your computer, make an alias of MixMHC2pred executable (see above for which one depending on operating system) or add it in your path.

Running

Command

```
MixMHC2pred -i input_file -o output_file -a allele1 allele2 [additional
options]
```

- Depending on your operating system, use MixMHC2pred, MixMHC2pred_unix or MixMHC2pred.exe as indicated in the installation instructions.
- Do not use spaces in your file or directory names.
- Do not use other special characters (e.g., *, ?, %, &,...) in file or directory names.

Required arguments

- Input file (command `-i` or `--input`): File listing all the peptides with one peptide per line. It can also be a fasta file (note that the peptide description given by the lines with ">" are not kept). Please note that even in fasta format, the input should consist in a list of peptides: MixMHC2pred is not cutting inputted proteins into shorter fragments that could be presented but it uses the input sequences as given in the file directly.

- Output file (command `-o` or `--output`): The name of the output file (including the directory). Peptides are kept in the same order than in the input file.

- Alleles (command: `-a` or `--alleles`): List of HLA-II alleles to test. Use for example the nomenclature *DRB1_03_01* for HLA-DRB1*03:01 and *DPA1_01_03__DPB1_04_01* for HLA-DPA1*01:03-DPB1*04:01. The full list of alleles available and corresponding nomenclature is given in the file *Alleles_list.txt*.
  If you want to make predictions with multiple alleles, list the different alleles separated by a space (e.g. `-a DRB1_11_01 DRB3_02_02`).

Optional arguments

- `--no_Nterm` and/or `--no_Cterm`: When these switches are used (not recommended), the N- and/or C-terminal motifs are not included in the computations of the score from each peptide.

- `--flat_ws` or `--best_s`: MixMHC2pred uses binding core offset preferences when computing the score from each peptide. It is nevertheless possible (but not recommended) to turn this feature off by using one of these two switches:

  - `--flat_ws`: the binding score is the non-weighted sum over the scores from each offset.

- ○ `--best_s`: only the score that was highest among all the offsets is kept, without giving weights to the offsets.

- `-x` or `--with_unspec_aa`: By default the peptide sequences can only contain the 20 standard amino acids and any peptide containing a non-standard aa will return a *NA* score. But when this switch is used, the sequences can also include *unspecified amino acids* (should use either "-" or the letter "X" to represent such an unspecified amino acid).

Results returned and additional information

- MixMHC2pred is meant for scoring different peptides and prioritising the most likely HLA-II ligands and epitopes. As it is trained on naturally presented peptides, it does not output a predicted affinity value, simply a score.

- Input should consist in a list of peptides, not proteins. Currently, MixMHC2pred is not cutting longer peptides/proteins into shorter fragments but use the peptides given in input as is.

- The score is computed for each allele provided in input. Results are returned for each allele in separate columns and additional columns give the results from the best allele for each peptide (columns *BestAllele* and ..._*best* in the output file, determined by the allele that had the maximal score, i.e. the most likely allele with which the peptide would be bound).

- The scores returned (columns *%Rank*) correspond to a percentile rank (best score is about 0, worst score is 100). This tells among random peptides, the percent of peptides expected to be better binders to this allele than the given peptide (among peptides of sizes 12-25 amino acids). This score is computed such that the top 1% best random peptides will have a length distribution following the one observed in naturally presented peptides.

- The *%Rank_best_perL* is similar but computed only between peptides having the same length. This score thus doesn't follow the length distribution observed in naturally presented ligands (this value is only returned for the best allele).

- The *CoreP1_*... columns tell what is the most likely binding core position for the given peptide towards the allele (this tells the position of the first amino acid from the binding core (which has a size of 9 aa in the predictions), starting at a value of 1 (i.e. if binding core corresponds to the 9 first amino acids from the peptide, this *CoreP1 = 1*)).

- For conveniance, the binding core sequence is also indicated for the best allele per peptide (column *Core_best*, for the other allleles this can be obtained from the *CoreP1* as indicated above).

- Peptides shorter than 12 amino acids, longer than 25 amino acids or containing non-standard amino acids are kept but with a score of "NA".

- The list of alleles available is provided in *Alleles_list.txt* showing the HLA-nomenclature and the corresponding nomenclature to use when running MixMHC2pred.

Latest version

Latest version of MixMHC2pred is available at https://github.com/GfellerLab/MixMHC2pred.

Check the file *NEWS* to see the main changes of the given version.

Web application

MixMHC2pred is also available as a web application at http://mixmhc2pred.gfellerlab.org.

License

MixMHC2pred can be used freely by academic groups for non-commercial purposes (see license). The product is provided free of charge, and, therefore, on an "as is" basis, without warranty of any kind.

FOR-PROFIT USERS: If you plan to use MixMHC2pred (version 1.2) or any data provided with the script in any for-profit application, you are required to obtain a separate license. To do so, please contact eauffarth@licr.org at the Ludwig Institute for Cancer Research Ltd.

Contact information

For scientific questions, please contact Julien Racle (julien.racle@unil.ch) or David Gfeller (david.gfeller@unil.ch).

For license-related questions, please contact Ece Auffarth (eauffarth@licr.org).

How to cite

To cite MixMHC2pred, please refer to:

Racle, J., et al. Robust prediction of HLA class II epitopes by deep motif deconvolution of immunopeptidomes. *Nat. Biotechnol.* 37, 1283–1286 (2019).