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-1.1.zip file and move it to a directory of your choice, where you have writing permissions.
2. Unzip MixMHC2pred-1.1.zip in this directory.
3. To test your installation, make sure you are in *MixMHC2pred-1.1* 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`
    - 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`
    - 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`

    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 (lines starting with ">" are skipped). 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 use 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 order from the input file. Peptides shorter than 12 amino acids or containing non-standard amino acids are kept but with a score of "nan".

- 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`). Only the score from the best allele for each peptide is returned.

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`: 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 with this switch (the binding score is then summed over all possible offsets).

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 of list of peptides, likely of sizes 12-21 amino acid long (shorter peptides return *nan* scores for the moment, while there is only a very low probability for longer peptides to be presented on HLA-II). At the moment, 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, and the maximal score is used to determine the most likely allele (column *BestAllele* in output file).

- The score returned (column *%Rank*) corresponds to a percentile rank (best score is 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. 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_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.
- The *BestCore* and *Best_s* returned correspond to the most likely binding core and offset for the given peptide towards its best allele.
- 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.

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 1.1) 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).