MixMHC2pred

MixMHC2pred is a pan-allele predictor of MHC class II ligands and epitopes. It is described in:
Racle, J., et al., Accurate predictions of MHC-II specificities (in prep.).

and

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

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

Installation

1. Download MixMHC2pred-2.0.zip file and move it to a directory of your choice, where you have writing permissions.
2. Unzip MixMHC2pred-2.0.zip in this directory.
3. To test your installation, make sure you are in *MixMHC2pred-2.0* directory and run the following command, depending on your operating system:
    - Mac OS: `./MixMHC2pred -i test/testData.txt -o test/out.txt -a DRB1_15_01 DRB5_01_01 DPA1_02_01__DPB1_01_01 DQA1_01_02__DQB1_05_01 DQA1_01_02__DQB1_06_02`
    - Unix: `./MixMHC2pred_unix -i test/testData.txt -o test/out.txt -a DRB1_15_01 DRB5_01_01 DPA1_02_01__DPB1_01_01 DQA1_01_02__DQB1_05_01 DQA1_01_02__DQB1_06_02`
    - Windows: `MixMHC2pred.exe -i test/testData.txt -o test/out.txt -a DRB1_15_01 DRB5_01_01 DPA1_02_01__DPB1_01_01 DQA1_01_02__DQB1_05_01 DQA1_01_02__DQB1_06_02`

    Your file *test/out.txt* should be the same as *test/out_compare.txt*. Running the software takes few seconds or more when testing lots of peptides and alleles.

    The *testData.txt* file corresponds to a subset of the HLA-II peptidomics data obtained from the cell line *DOHH2* in Dheilly et al., *Cancer Cell* (2020), containing some peptides bound to the HLA in the reverse direction.
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.

If using a non-standard OS, it is possible to compile MixMHC2pred using the Makefile found in the *bin* folder.

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 <file>` or `--input <file>`):
  File listing all the peptides. It should contain two columns: the first one is the sequence of the peptide, and 2nd column is the sequence of the context of the peptides (it should include the 3 AAs before the start of the peptide, followed by the 3 first AAs of the peptide, followed by the 3 last AAs of the peptide, followed by the 3 AAs just after the peptide). When the peptide lies near the begin or end of a protein, the corresponding context AAs should be written as "-", i.e. for the protein *ACDEFG*... if the peptide is *CDEFG*... the first 6 AA encoding its context should be written as *--ACDE* (and these 6 AAs should be directly followed by the 6 AAs describing the context near the C-terminal of the peptide). Also, if some AAs from the context of a peptide are not known, the unknown AAs should be written with the letter *X*. See *test/testData.txt* for an example of input file. When using the `no_context` option (see below), then this input file should only contain the list of the peptides, without any 2nd column of the context (an example input file without in such a case is available at *test/testData_noContext.txt*).

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

- Alleles (command: `-a <alleles>` or `--alleles <alleles>`):
  List of HLA-II alleles to test. Use 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_context`:
  In principle, MixMHC2pred includes the peptide context for its predictions (i.e. corresponding to a sequence of 12 AAs in total, including AAs just before and just after the peptide as explained above). It is nevertheless possible to decide to not consider any context information at all, when using this option. It is generally advised to include the context, in order to search for best candidate epitopes. But if analyzing a posteriori pre-cleaved peptide sequences (e.g. in experiments testing specific peptides directly, that therefore did not need to be cleaved by the cell), it may be a good to not consider the context encoding (often multiple overlapping epitopes are observed, so the peptide tested may not correspond to the best peptide based on context but it could still be recognized by the same T cells when given directly).

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 best score, i.e. the most likely allele by which the peptide would be presented).
- The two first columns of the output file give the *Peptide* and *Context* sequence of the peptide, which were given in the input file. When the option `--no_context` is used, the column *Context* is kept but it is empty.
- 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 presented by this allele than the given peptide.
- 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 alleles this can be obtained from the *CoreP1* as indicated above).
- The *subSpec_...* columns tell in which sub-specificity the given peptide is likely bound toward the given allele. The value *1* corresponds to the main sub-specificity (the only one for most alleles). But for example for *DRB1*08:01* allele a 2nd sub-specificity exists and is indicated by the value *2*. For alleles accomodating reverse binding, a value of *-1* indicates that the given peptide is bound in the reverse orientation.
- Peptides shorter than 12 amino acids, longer than 21 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 IPD-IMGT/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 2.0) 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., Accurate predictions of MHC-II specificities (in prep.).

and

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