

User guide for MARIA
#####

#MARIA is a neural network model predicting peptide ligands displayed by the cell HLA-II complexes. MARIA learned from MS-identified naturally presented HLA-II ligands and was not intended to predict *in vitro* protein-peptide binding affinities. This work is a product of Ash Alizadeh lab and collaborators at Stanford University. **Any users of MARIA should read and cite our recent [publication](#) on *Nature Biotechnology*.**

MARIA is available to academic groups for non-commercial purposes. The product is provided free of charge, and therefore on an “as is” basis without warranty of any kind. Please see the legal agreement at the end of this document.

FOR-PROFIT USERS

If you plan to use MARIA or any relevant files within the package, please obtain a separate license from Stanford Office of Technology Licensing with the reference number [S19-020](#).

For scientific queries, please contact Ash Alizadeh (arasha@stanford.edu) and Binbin Chen (bchen45@stanford.edu).

Copyright (2019) Binbin Chen and Stanford University

```
#####  
INSTALLATION  
#####
```

1) Download MARIA_package.zip file and unzip the zip file.

2) MARIA is built with python 2.7 and requires the following packages. We highly recommended users to install them with miniconda or anaconda (<https://conda.io/en/latest/miniconda.html>)

```
Keras 2.0.3  
Numpy  
Scipy  
Pandas  
Python-Levenshtein  
Tensorflow or Theano  
mkl #required if using theano
```

3) conda decency installation guide

#To create an environment specifically for MARIA (Optional)

```
conda create --name maria
```

source activate maria #do this every time before wanting to use this environment

#install dependencies

```
conda install pandas numpy scipy mkl tensorflow
```

```
pip install keras==2.0.3
```

```
pip install python-Levenshtein
```

#If tensorflow (recommended) does not work for your computer, you can install Theano.

```
conda install theano pygpu
```

```
#####  
RUNNING  
#####
```

```
python maria.py inputfile
```

optional arguments:

```
-h, --help          show this help message and exit
```

```
-scan SCAN
```

1 = Analyze each sequence with 15mer sliding windows
and use the best score, mandatory for sequences > 25
AAs; Default = 0.

```
-hladq2 HLADQ2
```

1 = Analyze HLA-DQ2.2 or HLA-DQ2.5; default 0 =
analyze HLA-DR presentation

```
-custom_tpm CUSTOM_TPM
```

1 = Use user input TPMs for analysis; default 0 = use reference RNA-Seq values

-tpm_reference TPM_REFERENCE
Gene expression of tissue/Cancer type to be used as reference gene expression profiles, default MCL.
Ignored when -custom_tpm 1

-cut_off CUT_OFF Percentile score cut-off to indicate whether a peptide will be presented; default 95.

-print_out PRINT_OUT 1 = Print the results; default 0 = save result it to a .output.txt file

#Input file:

We recommend users to modify provided input templates.

The input file is a plain tab-delimited text file with a header and 5 required columns. Column 1 and 2 are HLA-DR or DQ alleles of the cells (see Supported Alleles for details). Column 3 is the gene symbol (e.g. CTSK) of genes encoding the peptide of interest. Column 4 is peptide sequences in single letter format (all capitalized, no space). Column 5 is **optional** gene expression values if users want to provide specific gene expression values for this antigen gene (in TPM). Otherwise gene expression values will be estimated from external RNA-Seq references (e.g. TCGA) and genes with unknown gene expression will be assigned with a TPM of 5.

#Output file

Output file will be default be {**input_file**}.output.txt or be printed (if -print 1)

The output file contains all the same information in the input_file plus three or four additional columns:

1. **TPM estimated** (if the user did not provide custom values)
2. **MARIA raw scores** (0-1)
3. **MARIA percentile scores** (0-100%, raw scores normalized by comparing the scores to 20,000 random human peptides)
4. **15mer core** (String, fragments of the original sequence with the highest MARIA scor)
5. **Positive presenters** (0 or 1, indicating >= positive cut_off)

```
#####  
Example files  
#####
```

Three example inputs and outputs are provided. Check you can obtain expected outputs before running MARIA for your tasks.

#example 1

```
python maria.py Example1_K562_ligands.txt -tpm_reference K562 -cut_off 90
```

#example 2

```
python maria.py Example3_wheat_dq2.2_ligands.txt -cut_off 90 -hladq2 1
```

#example 3

```
python maria.py Example2_ott_et_al_vaccine_study.txt -custom_tpm 1 -scan 1 -cut_off  
95
```


Legal Agreement
#####

The Board of Trustees of the Leland Stanford Junior University (“Stanford”) provides MARIA website/script features and services (“Service”) free of charge for non-commercial use only. Use of the Service by any commercial entity for any purpose, including research, is prohibited.

By using the Service, you agree to be bound by the terms of this Agreement. Please read it carefully.

You agree not to use the Service for commercial advantage, or in the course of for-profit activities. You agree not to use the Service on behalf of any organization that is not a non-profit organization. **Commercial entities wishing to use this Service should contact Stanford University’s Office of Technology Licensing and reference docket S13-020.**

You agree not to use the Service for diagnosis, treatment, cure, prevention, or mitigation of disease or other conditions in man or other animals. You acknowledge that the Service and any information obtained therefrom is not intended to substitute for care by a licensed healthcare professional.

THE SERVICE IS OFFERED “AS IS”, AND, TO THE EXTENT PERMITTED BY LAW, STANFORD MAKES NO REPRESENTATIONS AND EXTENDS NO WARRANTIES OF ANY KIND, EITHER EXPRESS OR IMPLIED. STANFORD SHALL NOT BE LIABLE FOR ANY CLAIMS OR DAMAGES WITH RESPECT TO ANY LOSS OR OTHER CLAIM BY YOU OR ANY THIRD PARTY ON ACCOUNT OF, OR ARISING FROM THE USE OF THE SERVICE. YOU HEREBY AGREE TO DEFEND AND INDEMNIFY STANFORD, ITS TRUSTEES, EMPLOYEES, OFFICERS, STUDENTS, AGENTS, FACULTY, REPRESENTATIVES, AND VOLUNTEERS (“STANFORD INDEMNITEES”) FROM ANY LOSS OR CLAIM ASSERTED AGAINST STANFORD INDEMNITEES ARISING FROM YOUR USE OF THE SERVICE.

All rights not expressly granted to you in this Agreement are reserved and retained by Stanford or its licensors or content providers. This Agreement provides no license under any patent.

You agree that this Agreement and any dispute arising under it is governed by the laws of the State of California, United States of America, applicable to agreements negotiated, executed, and performed within California.

Subject to your compliance with the terms and conditions set forth in this Agreement, Stanford grants you a revocable, non-exclusive, non-transferable right to access and make use of the Service.