La Jolla Institute FOR ALLERGY AND IMMUNOLOGY | Life Without Disease.

The Immune Epitope Database Analysis Resource:

MHC class I peptide binding predictions

Sandeep Kumar Dhanda
sdhanda@lji.org

IEDB user workshop 2018

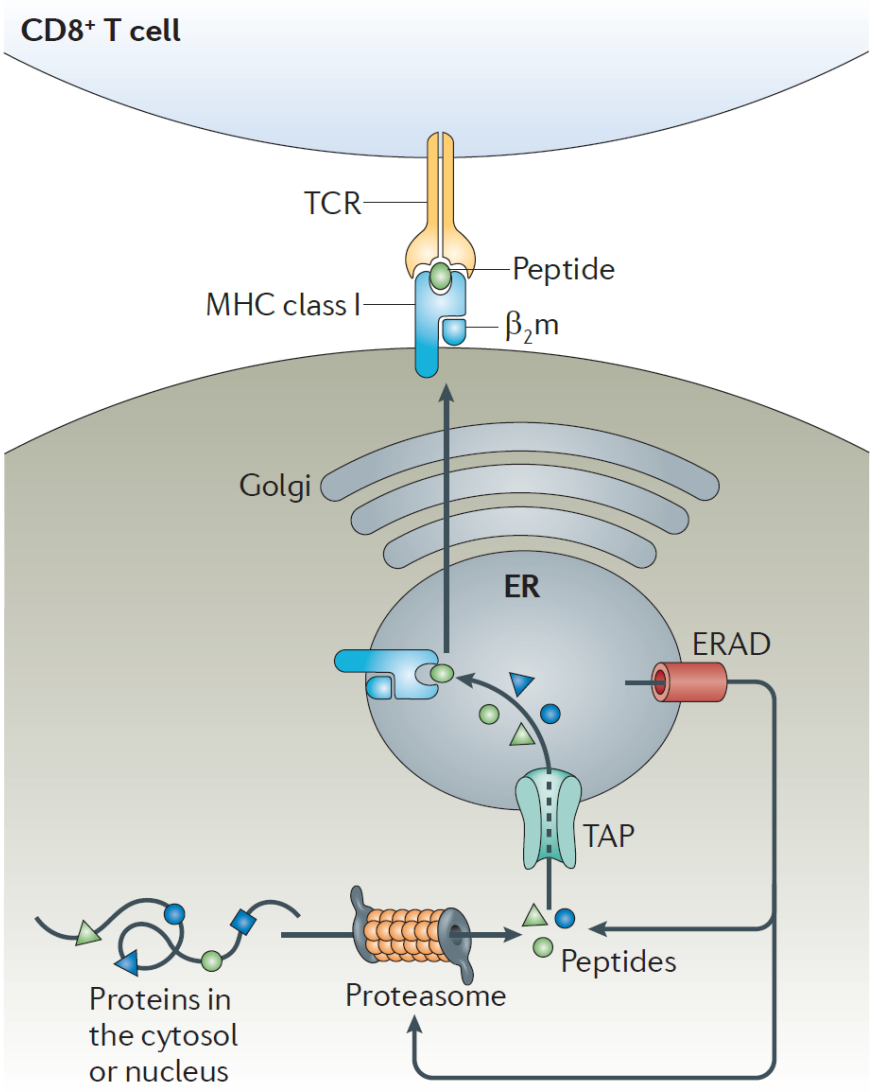
Oct 22-23, 2018

Outline

- Introduction
- Class I binding prediction tool – Web version
- IEDB recommendations & guidelines
- Exercise

Introduction

Endogenous antigen processing pathway (class I)



- Antigens generated within the cell
  - Viral particles
  - Self proteins
  - DRiPs (Defective Ribosomal Particles)

Nature Reviews Immunology **11**, 823-836 (December 2011) | doi:10.1038/nri3084

Towards a systems understanding of MHC class I and MHC class II antigen presentation

Jacques Neefjes[1], Marlieke L. M. Jongsma[1], Petra Paul[1] & Oddmund Bakke[2]

Class I MHC molecule

- Expressed by almost all nucleated cells.
- Presents antigen to **CD8⁺ T cells** (Cytotoxic T cells).
- One MHC encoded polymorphic chain (α) (2nd chain – β_2-microglobulin).
- The binding groove is **closed** at both ends and can accommodate peptides of **8-15 AA.**
- Only **α chain** impacts binding.

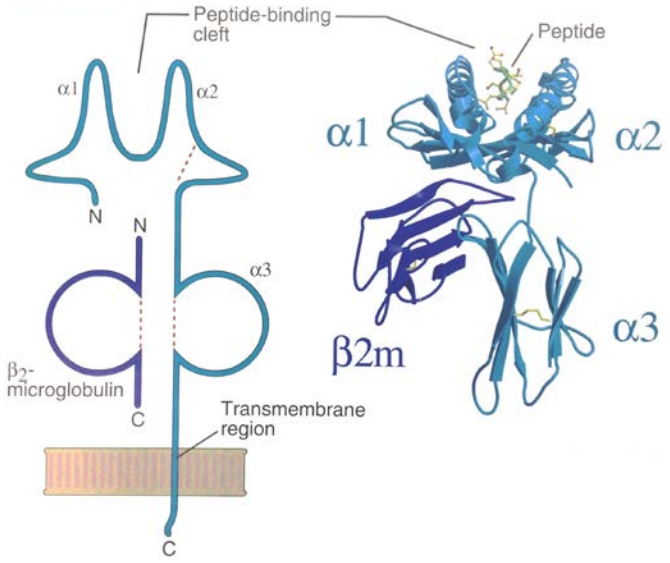


Figure source: Cellular & Molecular Immunology, 5th Ed by Abbas and Lichtman

MHC binding predictions

- MHC molecules are **highly polymorphic** – thousands of different variants exist
- MHC-peptide binding is **promiscuous** in nature
- Experimental characterisation of peptide–MHC interactions is highly **cost-intensive**
- **Prediction methods** facilitate selection of potential **epitopes** from a pool of peptides

Peptide binding data HLA-A*01:01

Peptide	IC_{50}(nM)
ASFCGSPY	51.4
LTDFGLSK	739.3
FTSFFYRY	1285.0
KSVFNSLY	1466.0
RDWAHNSL	1804.6
FSSCPVAY	1939.4
RNWAHSSL	2201.7
LSCAASGF	2830.1
LASIDLKY	3464.0
RAKFKQLL	5000.0

+

Machine learning algorithms

Binding data for MHC class I

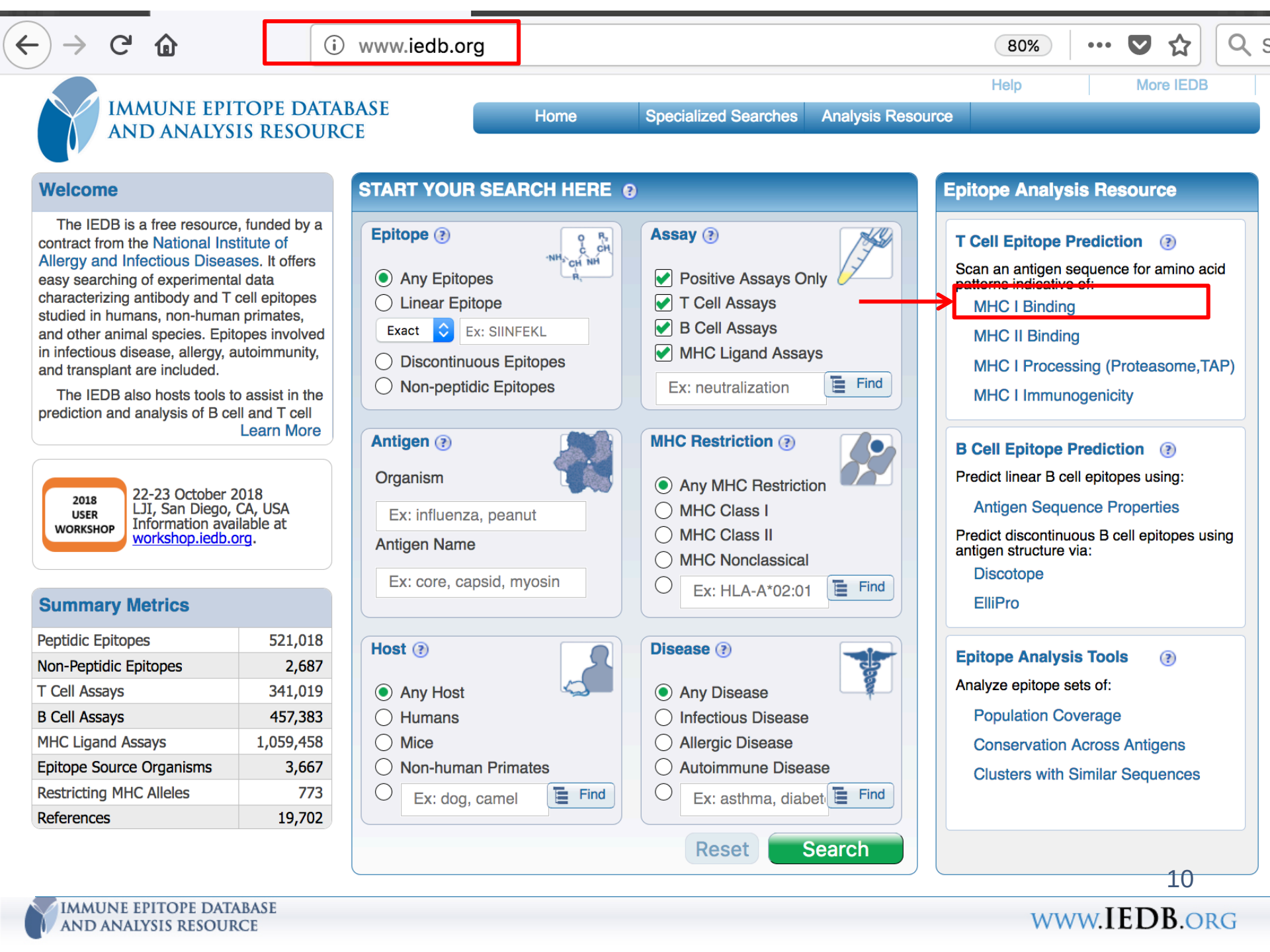
- **172 MHC molecules:**
  - 119 Human
  - 19 Macaque
  - 11 Chimpanzee
  - 8 Mouse
  - 7 Cattle
  - 4 Pig
  - 2 Rat
  - 1 Horse
  - 1 Gorilla

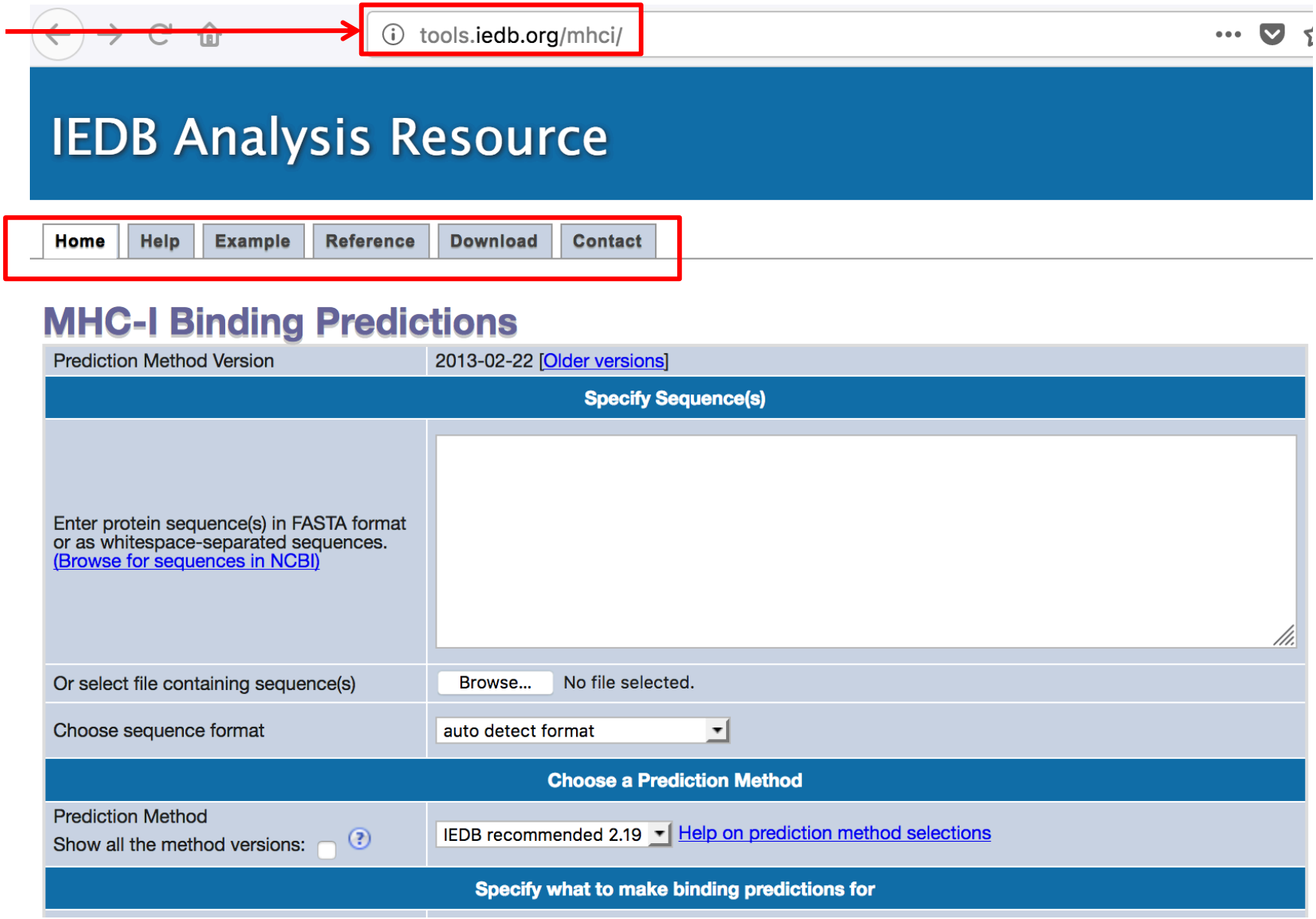
- Data sets available at **http://tools.iedb.org/main/datasets/**

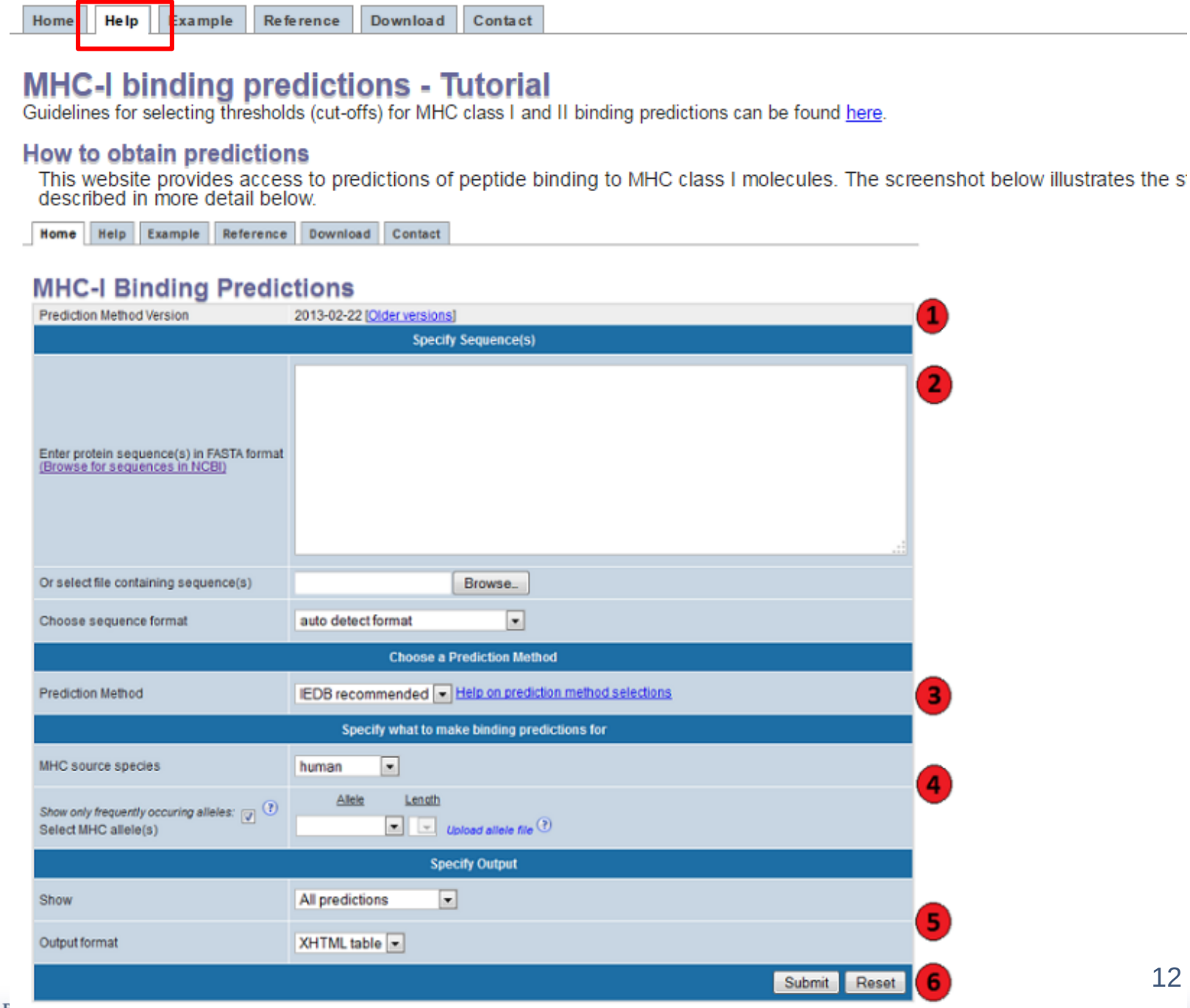
MHC class I binding prediction methods available

Method	Reference	Performance Reported
Consensus	Moutaftsi et al., 2006	
NetMHCpan-4.0	Jurts et al., 2017	0.960 AUC (average)
NetMHCpan-3.0	Nielsen & Andreatta, 2016	0.890 AUC (average)
ANN (NetMHC-4.0)	Andreatta & Nielsen, 2016	0.887 AUC (average)
SMM with Peptide:MHC Binding Energy Covariance matrix (SMMPMBEC)	Kim et al., 2009	0.894 AUC (average)
Stabilized matrix method (SMM)	Peters & Sette, 2005	0.887 AUC (average) (Kim et. al., 2009)
Combinatorial library (CombLib)	Sidney et al., 2008	0.909 AUC (HLA-A*0201)
PickPocket-1.1	Zhang et al., 2009	0.895 AUC (average)
NetMHCcons-1.1	Karosiene et al., 2012	0.729 PCC (average)
NetMHCstabpan-1.0	Rasmussen et al., 2016	0.980 AUC (average)

MHC-I peptide binding prediction tool
Web interface


www.iedb.org
80%
IMMUNE EPITOPE DATABASE AND ANALYSIS RESOURCE
Help
More IEDB
Home
Specialized Searches
Analysis Resource
Welcome
The IEDB is a free resource, funded by a contract from the National Institute of Allergy and Infectious Diseases. It offers easy searching of experimental data characterizing antibody and T cell epitopes studied in humans, non-human primates, and other animal species. Epitopes involved in infectious disease, allergy, autoimmunity, and transplant are included.
The IEDB also hosts tools to assist in the prediction and analysis of B cell and T cell
Learn More
2018 USER WORKSHOP
22-23 October 2018
LJI, San Diego, CA, USA
Information available at workshop.iedb.org.
Summary Metrics
Peptidic Epitopes 521,018
Non-Peptidic Epitopes 2,687
T Cell Assays 341,019
B Cell Assays 457,383
MHC Ligand Assays 1,059,458
Epitope Source Organisms 3,667
Restricting MHC Alleles 773
References 19,702
START YOUR SEARCH HERE
Epitope
Any Epitopes
Linear Epitope
Exact
Ex: SIINFEKL
Discontinuous Epitopes
Non-peptidic Epitopes
Assay
Positive Assays Only
T Cell Assays
B Cell Assays
MHC Ligand Assays
Ex: neutralization
Find
Antigen
Organism
Ex: influenza, peanut
Antigen Name
Ex: core, capsid, myosin
MHC Restriction
Any MHC Restriction
MHC Class I
MHC Class II
MHC Nonclassical
Ex: HLA-A*02:01
Find
Host
Any Host
Humans
Mice
Non-human Primates
Ex: dog, camel
Find
Disease
Any Disease
Infectious Disease
Allergic Disease
Autoimmune Disease
Ex: asthma, diabet
Find
Reset
Search
Epitope Analysis Resource
T Cell Epitope Prediction
Scan an antigen sequence for amino acid patterns indicative of:
MHC I Binding
MHC II Binding
MHC I Processing (Proteasome,TAP)
MHC I Immunogenicity
B Cell Epitope Prediction
Predict linear B cell epitopes using:
Antigen Sequence Properties
Predict discontinuous B cell epitopes using antigen structure via:
Discotope
ElliPro
Epitope Analysis Tools
Analyze epitope sets of:
Population Coverage
Conservation Across Antigens
Clusters with Similar Sequences


tools.iedb.org/mhci/
IEDB Analysis Resource
Home
Help
Example
Reference
Download
Contact
MHC-I Binding Predictions
Prediction Method Version
2013-02-22 [Older versions]
Specify Sequence(s)
Enter protein sequence(s) in FASTA format or as whitespace-separated sequences. (Browse for sequences in NCBI)
Or select file containing sequence(s)
Browse...
No file selected.
Choose sequence format
auto detect format
Choose a Prediction Method
Prediction Method
Show all the method versions:
IEDB recommended 2.19
Help on prediction method selections
Specify what to make binding predictions for



1. Specify prediction method version

Home | Help | Example | Reference | Download | Contact

MHC-I binding predictions - Example data

Choose one of the radio buttons below to select protein sequence(s) containing MHC class I epitopes d
These sequences will be transferred to the MHC class I binding predictions when clicking the "Submit" b
by no means considered equivalent to a formal performance evaluation.

GP and NP protein of LCMV virus strain armstrong

Peptide	Length	MHC restriction
FQPQNGQFI	9	H-2-Db
KAVYNFATC	9	H-2-Db
ISHNFCNL	8	H-2-Kb
YTVKYPNL	8	H-2-Kb

SARS spike protein

Peptide	Length	MHC restriction
FIAGLIAIV	9	HLA-A*02:01
LITGRLQSL	9	HLA-A*02:01
RLNEVAKNL	9	HLA-A*02:01

Submit

References

ANN:

Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, and Nielsen M. 2008. NetMHC-3.0: Accurate web accessible for peptides of length 8-11. *NAR* **36**:W509-512.
PMID: 18463140

Lundegaard C, Nielsen M, Lund O. 2006. The validity of predicted T-cell epitopes. *Trends Biotechnol* **24**:537-538.
PMID: 17045685

Lundegaard C, Lund O, and Nielsen M. 2008. Accurate approximation method for prediction of class I MHC affinities for 9mers. *Bioinformatics* **24**:1397-1398.
PMID: 18413329

Nielsen M, Lundegaard C, Worning P, Lauemøller SL, Lamberth K, Buus S, Brunak S, Lund O. 2003. Reliable prediction representations. *Protein Sci* **12**:1007-1017.
PMID: 12717023

Buus S, Lauemøller SL, Worning P, Kesmir C, Frimurer T, Corbet S, Fomsgaard A, Hilden J, Holm A, Brunak S. 2003. S 'Query by Committee' artificial neural network approach. *Tissue Antigens* **62**:378-384.
PMID: 14617044

SMM:

Peters B, Sette A. 2005. Generating quantitative models describing the sequence specificity of biological processes witl
PMID: 15927070

SMMPMBEC:

Kim Y, Sidney J, Pinilla C, Sette A, Peters B. 2009. Derivation of an amino acid similarity matrix for peptide:MHC binding **10**:394.
PMID: 19948066

CombLib:

Sidney J, Assarsson E, Moore C, Ngo S, Pinilla C, Sette A, Peters B. 2008. Quantitative peptide binding motifs for 19 hu scanning combinatorial peptide libraries. *Immunome Res* **4**:2.
PMID: 18221540

Home | Help | Example | Reference | Download | Contact

Scoring matrices of SMM and SMMPMBEC - Download

To download the dataset in tar.gz format: Download

Dataset used for retraining the IEDB class I binding prediction tools.

- Description of the dataset: The dataset is largely identical to that of Kim et al (2014), described above, but includes additional c
- Date of the dataset generation: 2013
- Details on the dataset generation: The dataset was compiled from three sources: the IEDB, the Sette lab, and the Buus lab. If measurement among the three sources, its geometric mean was taken.
- Data format: Compressed text file containing binding data.
- Dataset availability: binding_data_2013.zip

MHC-I binding predictions - Download

The MHC_I binding tool contains a collection of following peptide binding prediction methods for Major Histocompatibility Complex pythons scripts and linux 32-bit environment specific binaries.

- ann
- smm
- smmpmbec
- comblib_sidney2008
- consensus
- netmhcpan
- pickpocket
- netmhccons

License Agreements

By downloading the standalone tool, you are consenting to be bound by and become a party as the "Licensee" for the use of NetMHC 3.0".
Also you are consenting the terms and conditions of the Non-Profit Open Software License ("Non-Profit OSL") version 3.0
Please read folowing two agreements before proceeding. If you do not agree to all of the terms of these two agreements, you must not click the install the product nor use the product, and you do not become a LICENSEE under these agreements.

```
ACADEMIC SOFTWARE LICENSE AGREEMENT FOR END-USERS AT PUBLICLY FUNDED
ACADEMIC, EDUCATION OR RESEARCH INSTITUTIONS FOR THE USE OF NetMHC 3.0
By downloading the Software you are consenting to be bound by and become
a party to this agreement as the "Licensee". If you do not agree to all
of the terms of this agreement, you must not click the Acceptance
```

```
This Non-Profit Open Software License ("Non-Profit OSL") version 3.0 (the
"Licen
se") applies to any original work of authorship (the "Original Work") whose
owne
r (the "Licensor") has placed the following licensing notice adjacent to
```

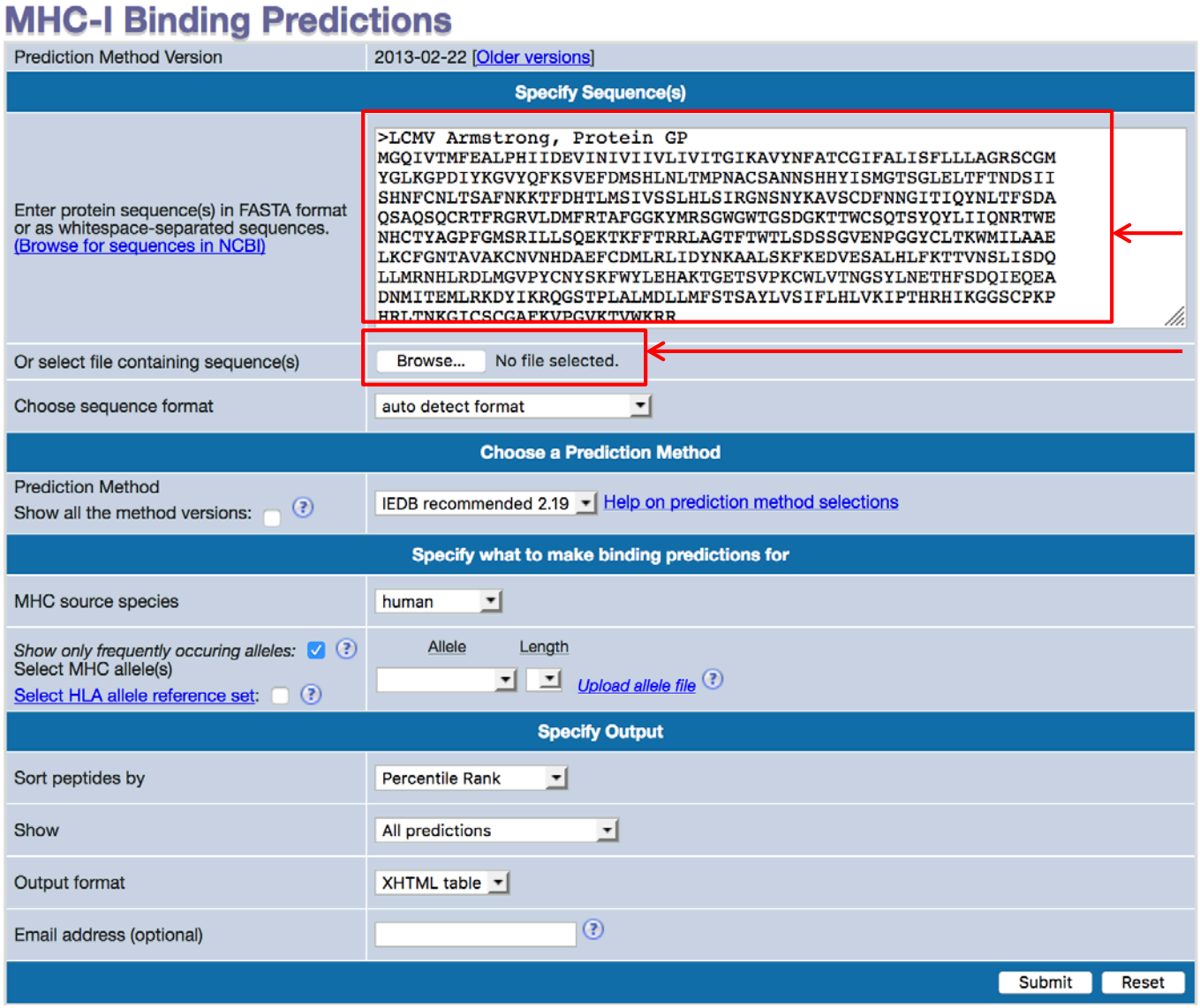
To download the tools in tar.gz format: Agree and Download
To return to the main page: Decline

Contact the developers

Should you require assistance beyond the help provided for each individual tool, pl
help@iedb.org

Using the tool

Sequences



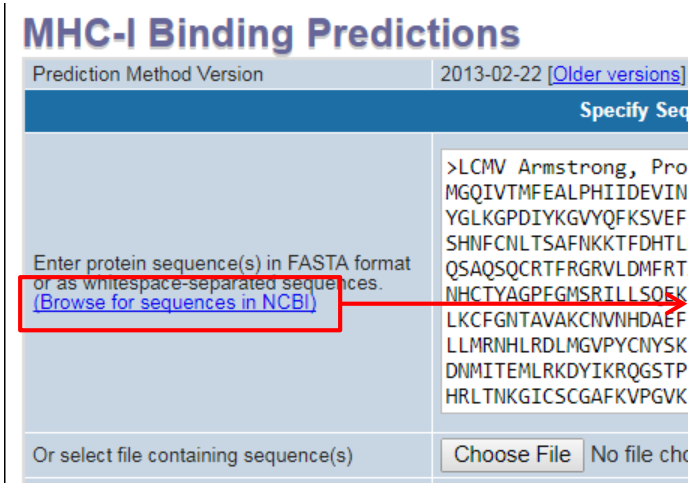
Copy sequences

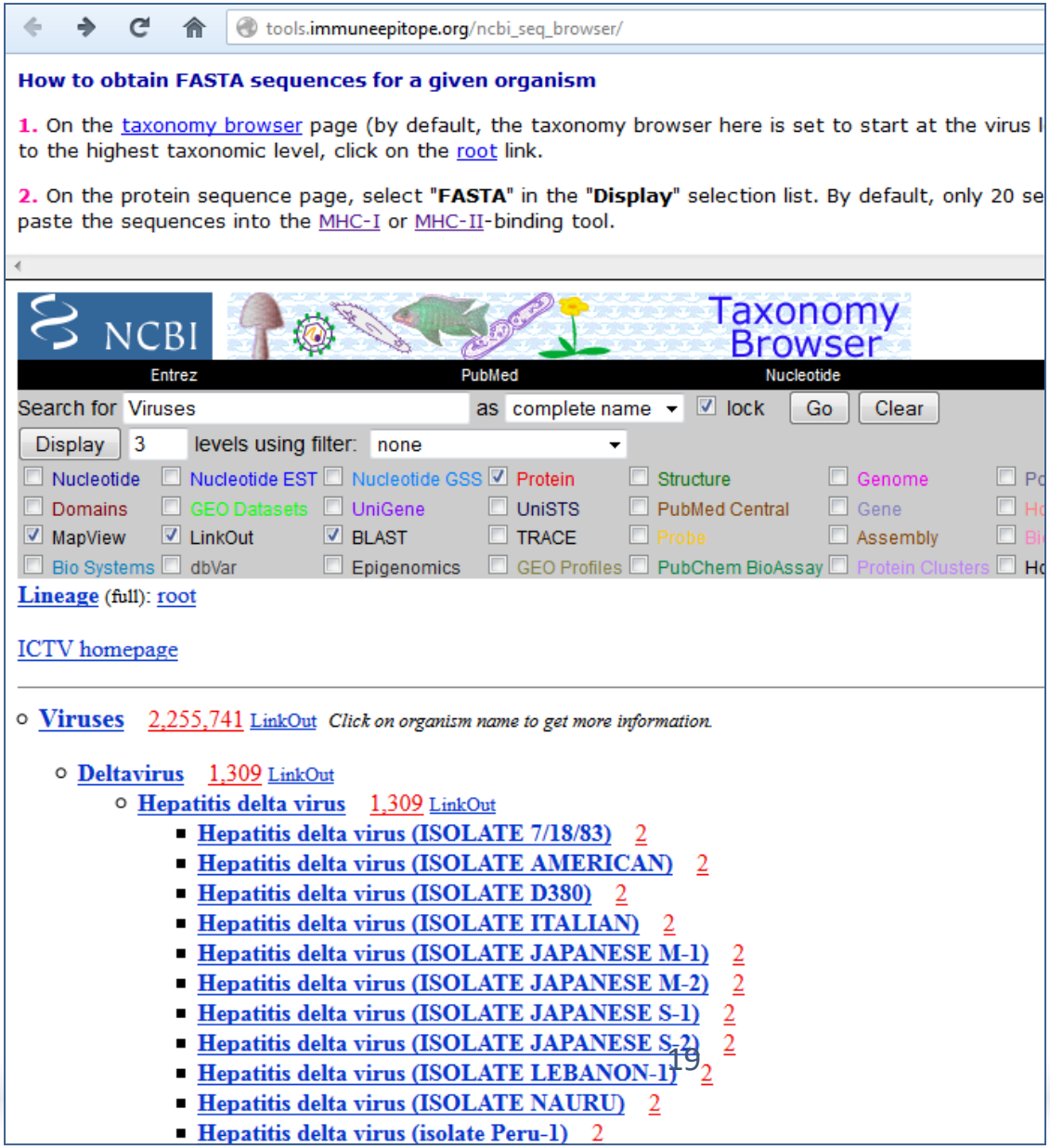
Or upload sequences as plain text file

Sequence formats:

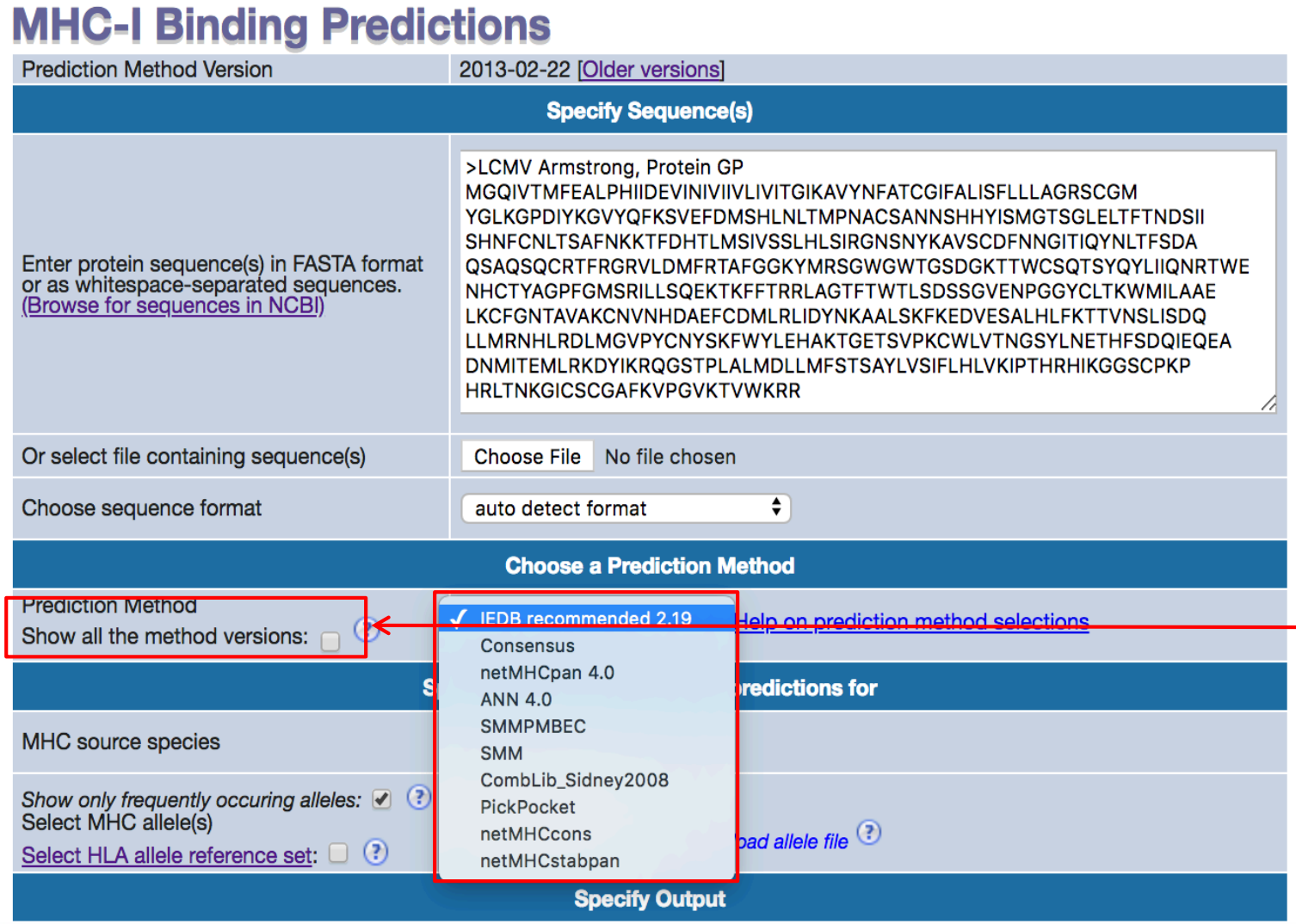
- FASTA format
- Space separated sequences
- One continuous sequence

NCBI sequence browser



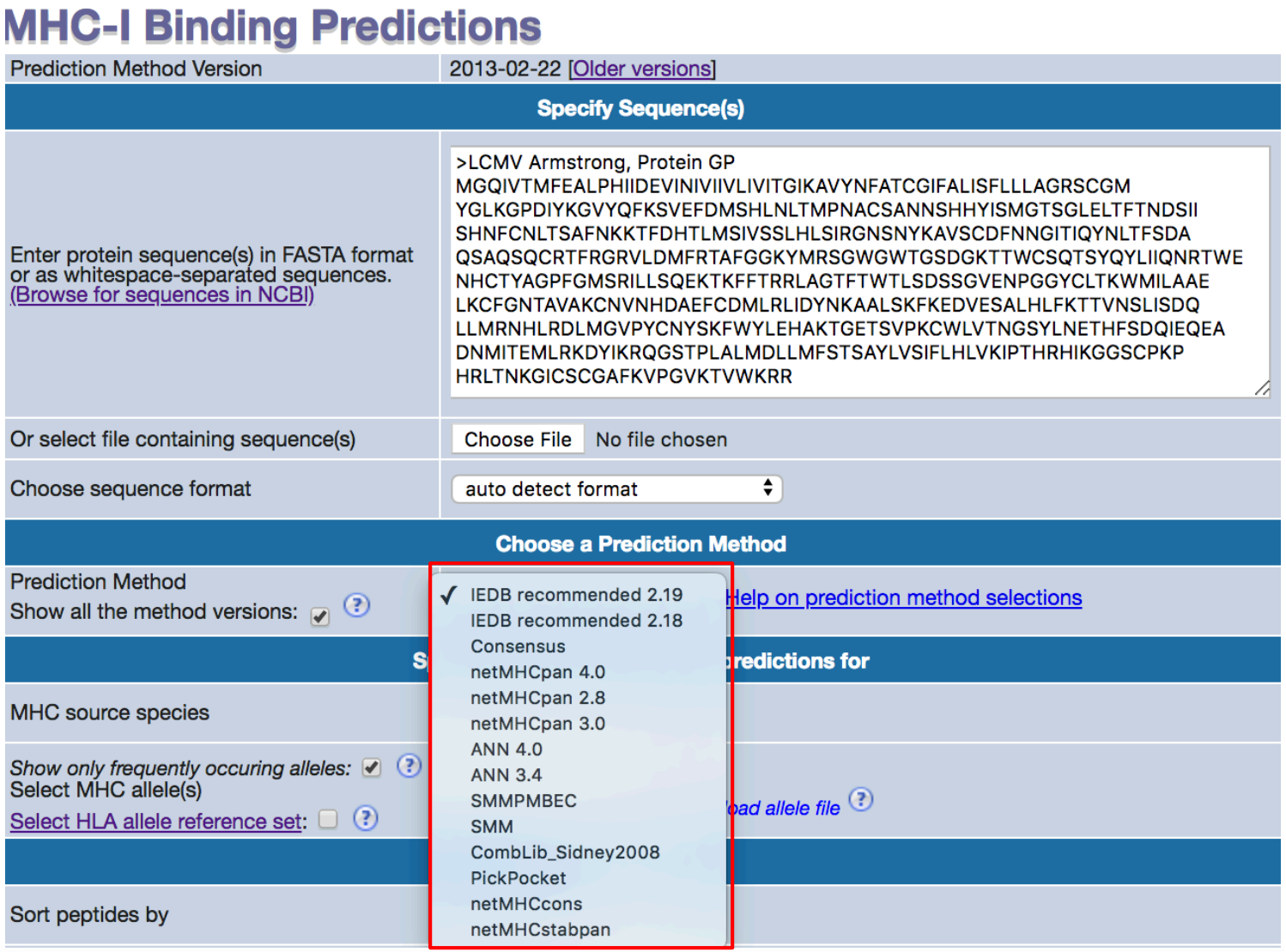


Prediction method



Other Versions

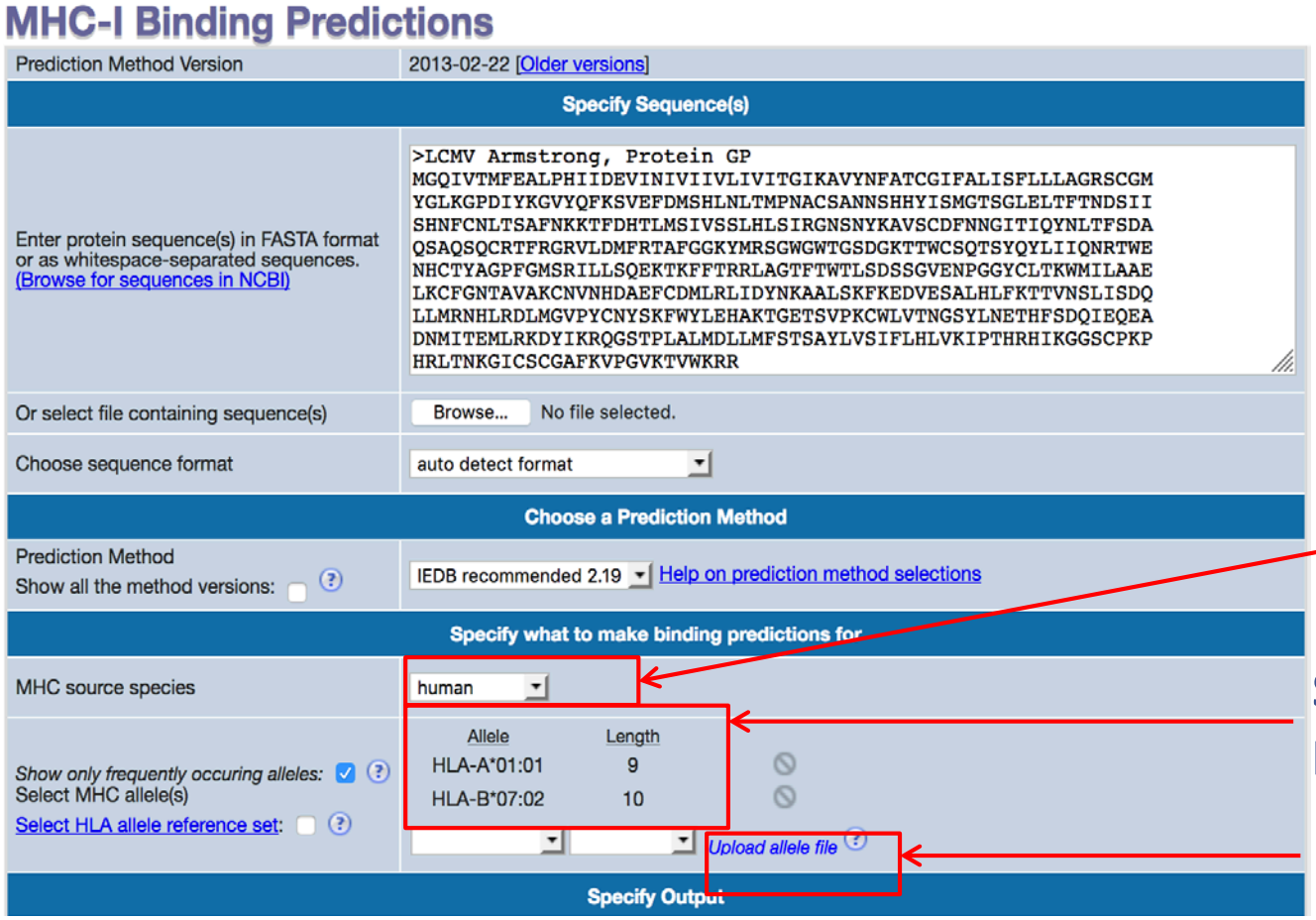
Prediction method



Guidelines: Choosing the prediction method

- Method to use: **IEDB recommended method** - employs Consensus (Combination of ANN, SMM & CombLib) or NetMHCpan depending on the allele.
- Advantages:
  - Best available methods.
  - Provides a consensus percentile rank.
  - Provides binding affinity & percentile rank for each method separately as well.
- Recommendation may change with the new benchmark studies

Allele selection



Choose species

Specify alleles & peptide length

Format:

HLA-A*02:01,9

HLA-B*15:01,9

HLA-A*02:06,10

Natural length distribution in epitope prediction

- Alleles differ in their preference for lengths on binding and presentation of peptides

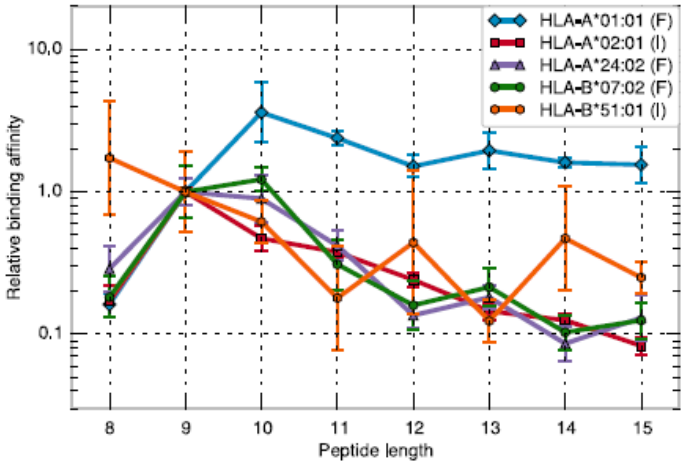


FIGURE 1. Peptide binding-length preference for five common HLA alleles. The length preference for each HLA was determined by measuring

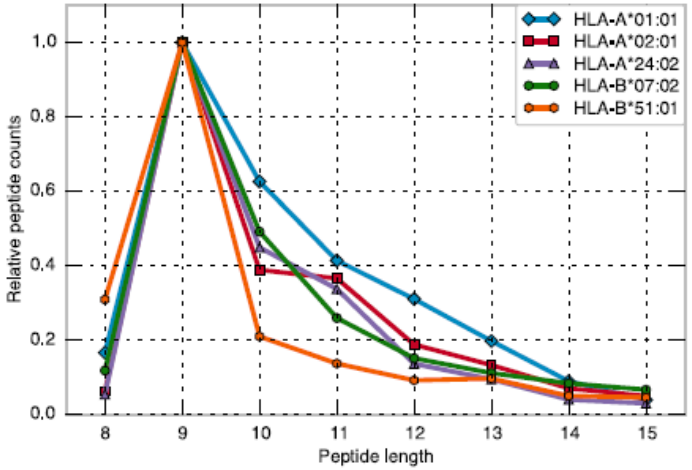
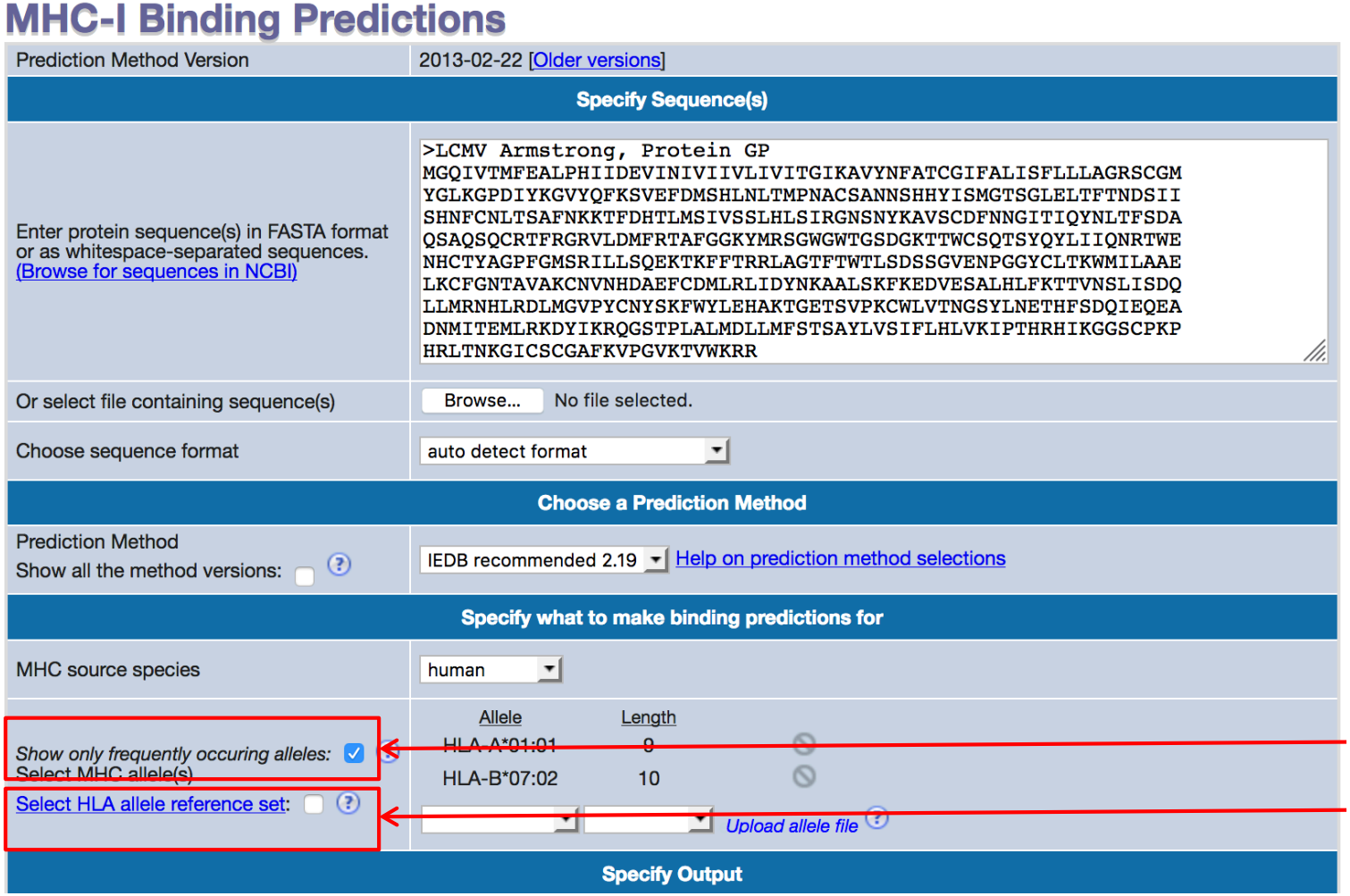


FIGURE 2. Length profiles of naturally presented peptides for five HLA molecules. Large datasets of HLA-I ligands were determined by the elu-

- New model developed – will be incorporated into prediction
- Trolle et al. (2016) Journal of Immunology

Allele selection – complete list of alleles



More alleles

Reference alleles

Allele selection – Reference set for global coverage

- Reference set of **27** alleles has been assembled – covers > **97%** of population
- http://iedb.zendesk.com/entries/25054538-HLA-allele-frequencies

HLA-A	Frequency	HLA-B	Frequency
A*01:01	16.2	B*07:02	13.3
A*02:01	25.2	B*08:01	11.5
A*02:03	3.3	B*15:01	5.2
A*02:06	4.9	B*35:01	6.5
A*03:01	15.4	B*40:01	10.3
A*11:01	12.9	B*44:02	9.2
A*23:01	6.4	B*44:03	7.6
A*24:02	16.8	B*51:01	5.5
A*26:01	4.7	B*53:01	5.4
A*30:01	5.1	B*57:01	3.2
A*30:02	5.0	B*58:01	3.6
A*31:01	4.7		
A*32:01	5.7		
A*33:01	3.2		
A*68:01	4.6		
A*68:02	3.3		

Allele selection

MHC-I Binding Predictions

Prediction Method Version	2013-02-22 [Older versions]
Specify Sequence(s)	
Enter protein sequence(s) in FASTA format or as whitespace-separated sequences. (Browse for sequences in NCBI)	>LCMV Armstrong, Protein GP MGQIVTMFEALPHIIDEVINIVIIVLIVITGIKAVYNFATCGIFALISFLLLAGRSCGM YGLKGPDIYKGVYQFKSVEFDMSHLNLTMPNACSANNSHHYISMGTSGLELTFTNDSII SHNFCNLTSAFNKKTFDHTLMSIVSSLHLSIRGNSNYKAVSCDFNNGITIQYNLTFSDA QSAQSQCRTFRGRVLDMFRTAFGGKYMRSGWGWTGSDGKTTWCSQTSYQYLIIQNRTWE NHCTYAGPFGMSRILLSQEKTKFFTRRLAGTFTWTLSDSSGVENPGGYCLTKWMILAAE LKCFGNTAVAKCNVNHDAEFCDMLRLIDYNKAALSKFKEDVESALHLFKTTVNSLISDQ LLMRNHLRDLMGVPYCNYSKFWYLEHAKTGETSVPKCWLVTNGSYLNETHFSDQIEQEA DNMITEMLRKDYIKRQGSTPLALMDLLMFSTSAYLVSIFLHLVKIPTHRHIKGGSCPKP HRLTNKGICSCGAFKVPGVKTVWKRR
Or select file containing sequence(s)	Browse... No file selected.
Choose sequence format	auto detect format
Choose a Prediction Method	
Prediction Method Show all the method versions:	IEDB recommended 2.19 Help on prediction method selections
Specify what to make binding predictions for	
MHC source species	human
Show only frequently occuring alleles: Select MHC allele(s) Select HLA allele reference set:	Allele: HLA-A*01:01, Length: 9 Allele: HLA-B*07:02, Length: 10 *Upload allele file*
Specify Output	

Allele	Length
HLA-A*01:01	9
HLA-A*01:01	10
HLA-A*02:01	9
HLA-A*02:01	10
HLA-A*02:03	9
HLA-A*02:03	10
HLA-A*02:06	9
HLA-A*02:06	10
HLA-A*03:01	9
HLA-A*03:01	10
HLA-A*11:01	9
HLA-A*11:01	10
HLA-A*23:01	9
HLA-A*23:01	10
HLA-A*24:02	9
HLA-A*24:02	10
HLA-A*26:01	9
HLA-A*26:01	10
HLA-A*30:01	9
HLA-A*30:01	10
HLA-A*30:02	9
HLA-A*30:02	10
HLA-A*31:01	9
HLA-A*31:01	10
HLA-A*32:01	9
HLA-A*32:01	10
HLA-A*33:01	9
HLA-A*33:01	10
HLA-A*68:01	9
HLA-A*68:01	10
HLA-A*68:02	9
HLA-A*68:02	10
HLA-B*07:02	9
HLA-B*07:02	10
HLA-B*08:01	9
HLA-B*08:01	10
HLA-B*15:01	9
HLA-B*15:01	10
HLA-B*35:01	9
HLA-B*35:01	10
HLA-B*40:01	9
HLA-B*40:01	10
HLA-B*44:02	9
HLA-B*44:02	10
HLA-B*44:03	9
HLA-B*44:03	10
HLA-B*51:01	9
HLA-B*51:01	10
HLA-B*53:01	9
HLA-B*53:01	10
HLA-B*57:01	9
HLA-B*57:01	10
HLA-B*58:01	9
HLA-B*58:01	10

Allele selection – upload file

MHC-I Binding Predictions

Prediction Method Version	2013-02-22 [Older versions]
Specify Sequence(s)	
Enter protein sequence(s) in FASTA format or as whitespace-separated sequences. (Browse for sequences in NCBI)	>LCMV Armstrong, Protein GP MGQIVTMFEALPHIIDEVINIVIIVLIVITGIKAVYNFATCGIFALISFLLLAGRSCGM YGLKGPDIYKGVYQFKSVEFDMSHLNLTMPNACSANNSHHYISMGTSGLELTFTNDSII SHNFCNLTSAFNKKTFDHTLMSIVSSLHLSIRGNSNYKAVSCDFNNGITIQYNLTFSDA QSAQSQCRTFRGRVLDMFRTAFGGKYMRSGWGWTGSDGKTTWCSQTSYQYLIIQNRTWE NHCTYAGPFGMSRILLSQEKTKFFTRRLAGTFTWTLSDSSGVENPGGYCLTKWMILAAE LKCFGNTAVAKCNVNHDAEFCDMLRLIDYNKAALSKFKEDVESALHLFKTTVNSLISDQ LLMRNHLRDLMGVPYCNYSKFWYLEHAKTGETSVPKCWLVTNGSYLNETHFSDQIEQEA DNMITEMLRKDYIKRQGSTPLALMDLLMFSTSAYLVSIFLHLVKIPTHRHIKGGSCPKP HRLTNKGICSCGAFKVPGVKTVWKRR

- Only available alleles
- No allele sequence

Help page

- ***Format for the upload allele file:***

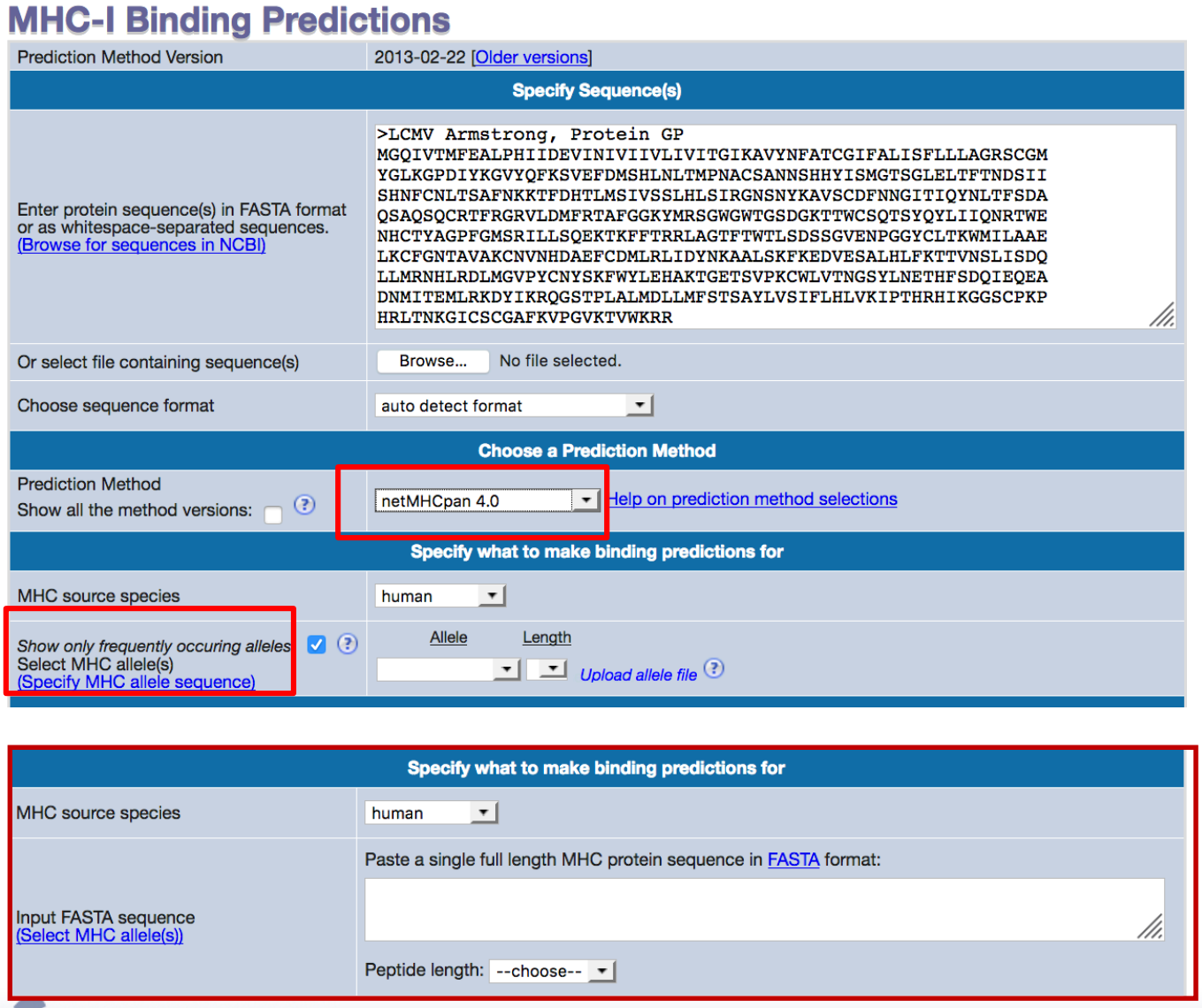
File should be in simple text format containing comma separated values, where each allele is separated from it's length by a comma followed by a new lir pair(example given below). However, you may also choose allele(s) and their length(s) from the drop-down selection in together with your uploaded file.
Example:

```
HLA-A*02:01,9
HLA-B*15:01,9
HLA-A*02:06,10
```

Additional information regarding HLA allele frequencies and nomenclature are also provided.
Note: for NetMHCpan method, there is an option to paste a single full length MHC protein sequence in FASTA format, instead of selecting alleles from the

Specify what to make binding predictions for	
MHC source species	human
Show only frequently occuring alleles: Select MHC allele(s) Select HLA allele reference set:	Allele: HLA-A*02:01, Length: 9 Allele: HLA-B*07:02, Length: 10 Upload allele file
Specify Output	

Allele selection – Specify by sequence



MHC-I Binding Predictions

Prediction Method Version	2013-02-22 [Older versions]
Specify Sequence(s)	
Enter protein sequence(s) in FASTA format or as whitespace-separated sequences. (Browse for sequences in NCBI)	>LCMV Armstrong, Protein GP MGQIVTMFEALPHIIDEVINIVIIVLIVITGIKAVYNFATCGIFALISFLLLAGRSCGM YGLKGPDIYKGVYQFKSVEFDMSHLNLTMPNACSANNSHHYISMGTSGLELTFTNDSII SHNFCNLTSAFNKKTFDHTLMSIVSSLHLSIRGNSNYKAVSCDFNNGITIQYNLTFSDA QSAQSQCRTFRGRVLDMFRTAFGGKYMRSGWGWTGSDGKTTWCSQTSYQYLIIQNRTWE NHCTYAGPFGMSRILLSQEKTKFFTRRLAGTFTWTLSDSSGVENPGGYCLTKWMILAAE LKCFGNTAVAKCNVNHDAEFCDMLRLIDYNKAALSKFKEDVESALHLFKTTVNSLISDQ LLMRNHLRDLMGVPYCNYSKFWYLEHAKTGETSVPKCWLVTNGSYLNETHFSDQIEQEA DNMITEMLRKDYIKRQGSTPLALMDLLMFSTSAYLVSIFLHLVKIPTHRHIKGGSCPKP HRLTNKGICSCGAFKVPGVKTVWKRR
Or select file containing sequence(s)	Browse... No file selected.
Choose sequence format	auto detect format
Choose a Prediction Method	
Prediction Method Show all the method versions:	netMHCpan 4.0 Help on prediction method selections
Specify what to make binding predictions for	
MHC source species	human
Show only frequently occuring alleles Select MHC allele(s) (Specify MHC allele sequence)	Allele Length *Upload allele file*

Specify what to make binding predictions for	
MHC source species	human
Input FASTA sequence (Select MHC allele(s))	Paste a single full length MHC protein sequence in FASTA format: Peptide length: --choose--

Output sorting order

MHC-I Binding Predictions

Prediction Method Version	2013-02-22 [Older versions]
Specify Sequence(s)	
Enter protein sequence(s) in FASTA format or as whitespace-separated sequences. (Browse for sequences in NCBI)	>LCMV Armstrong, Protein GP MGQIVTMFEALPHIIDEVINIVIIVLIVITGIKAVYNFATCGIFALISFLLLAGRSCGM YGLKGPDIYKGVYQFKSVEFDMSHLNLTMPNACSANNSHHYISMGTSGLELTFTNDSII SHNFCNLTSAFNKKTFDHTLMSIVSSLHLSIRGNSNYKAVSCDFNNGITIQYNLTFSDA QSAQSQCRTFRGRVLDMFRTAFGGKYMRSGWGWTGSDGKTTWCSQTSYQYLIIQNRTWE NHCTYAGPFGMSRILLSQEKTKFFTRRLAGTFTWTLSDSSGVENPGGYCLTKWMILAAE LKCFGNTAVAKCNVNHDAEFCDMLRLIDYNKAALSKFKEDVESALHLFKTTVNSLISDQ LLMRNHLRDLMGVPYCNYSKFWYLEHAKTGETSVPKCWLVTNGSYLNETHFSDQIEQEA DNMITEMLRKDYIKRQGSTPLALMDLLMFSTSAYLVSIFLHLVKIPTHRHIKGGSCPKP HRLTNKGICSCGAFKVPGVKTVWKRR
Or select file containing sequence(s)	Browse... No file selected.
Choose sequence format	auto detect format
Choose a Prediction Method	
Prediction Method Show all the method versions:	IEDB recommended 2.19 Help on prediction method selections
Specify what to make binding predictions for	
MHC source species	human
Show only frequently occuring alleles: Select MHC allele(s) Select HLA allele reference set:	Allele Length Upload allele file
Specify Output	
Sort peptides by	✓ Percentile Rank Position in sequence
Show	All predictions
Output format	XHTML table
Email address (optional)	

Optional filtering of prediction results

MHC-I Binding Predictions

Prediction Method Version	2013-02-22 [Older versions]
Specify Sequence(s)	
Enter protein sequence(s) in FASTA format or as whitespace-separated sequences. (Browse for sequences in NCBI)	>LCMV Armstrong, Protein GP MGQIVTMFEALPHIIDEVINIVIIVLIVITGIKAVYNFATCGIFALISFLLLAGRSCGM YGLKGPDIYKGVYQFKSVEFDMSHLNLTMPNACSANNSHHYISMGTSGLELTFTNDSII SHNFCNLTSAFNKKTFDHTLMSIVSSLHLSIRGNSNYKAVSCDFNNGITIQYNLTFSDA QSAQSQCRTFRGRVLDMFRTAFGGKYMRSGWGWTGSDGKTTWCSQTSYQYLIIQNRTWE NHCTYAGPFGMSRILLSQEKTKFFTRRLAGTFTWTLSDSSGVENPGGYCLTKWMILAAE LKCFGNTAVAKCNVNHDAEFCDMLRLIDYNKAALSKFKEDVESALHLFKTTVNSLISDQ LLMRNHLRDLMGVPYCNYSKFWYLEHAKTGETSVPKCWLVTNGSYLNETHFSDQIEQEA DNMITEMLRKDYIKRQGSTPLALMDLLMFSTSAYLVSIFLHLVKIPTHRHIKGGSCPKP HRLTNKGICSCGAFKVPGVKTVWKRR
Or select file containing sequence(s)	Browse... No file selected.
Choose sequence format	auto detect format
Choose a Prediction Method	
Prediction Method Show all the method versions:	IEDB recommended 2.19 Help on prediction method selections
Specify what to make binding predictions for	
MHC source species	human
Show only frequently occuring alleles: Select MHC allele(s) Select HLA allele reference set:	Allele Length Upload allele file
Specify Output	
Sort peptides by	All predictions IC50 below [cutoff] nM ✓ Percent rank below [cutoff]
Show	off 5
Output format	XHTML table
Email address (optional)	

Optional selection of output format

MHC-I Binding Predictions

Prediction Method Version	2013-02-22 [Older versions]
Specify Sequence(s)	
Enter protein sequence(s) in FASTA format or as whitespace-separated sequences. (Browse for sequences in NCBI)	>LCMV Armstrong, Protein GP MGQIVTMFEALPHIIDEVINIVIIVLIVITGIKAVYNFATCGIFALISFLLLAGRSCGM YGLKGPDIYKGVYQFKSVEFDMSHLNLTMPNACSANNSHHYISMGTSGLELTFTNDSII SHNFCNLTSAFNKKTFDHTLMSIVSSLHLSIRGNSNYKAVSCDFNNGITIQYNLTFSDA QSAQSQCRTFRGRVLDMFRTAFGGKYMRSGWGWTGSDGKTTWCSQTSYQYLIIQNRTWE NHCTYAGPFGMSRILLSQEKTKFFTRRLAGTFTWTLSDSSGVENPGGYCLTKWMILAAE LKCFGNTAVAKCNVNHDAEFCDMLRLIDYNKAALSKFKEDVESALHLFKTTVNSLISDQ LLMRNHLRDLMGVPYCNYSKFWYLEHAKTGETSVPKCWLVTNGSYLNETHFSDQIEQEA DNMITEMLRKDYIKRQGSTPLALMDLLMFSTSAYLVSIFLHLVKIPTHRHIKGGSCPKP HRLTNKGICSCGAFKVPGVKTVWKRR
Or select file containing sequence(s)	Browse... No file selected.
Choose sequence format	auto detect format
Choose a Prediction Method	
Prediction Method Show all the method versions:	IEDB recommended 2.19 Help on prediction method selections
Specify what to make binding predictions for	
MHC source species	human
Show only frequently occuring alleles: Select MHC allele(s) Select HLA allele reference set:	Allele: HLA-A*01:01, Length: 9 Allele: HLA-B*07:02, Length: 10 *Upload allele file*
Specify Output	
Sort peptides by	Percentile Rank
Show	All predictions
Output format	✓ XHTML table Text file
Email address (optional)	

Email address for sending results

MHC-I Binding Predictions

Prediction Method Version	2013-02-22 [Older versions]
Specify Sequence(s)	
Enter protein sequence(s) in FASTA format or as whitespace-separated sequences. (Browse for sequences in NCBI)	>LCMV Armstrong, Protein GP MGQIVTMFEALPHIIDEVINIVIIVLIVITGIKAVYNFATCGIFALISFLLLAGRSCGM YGLKGPDIYKGVYQFKSVEFDMSHLNLTMPNACSANNSHHYISMGTSGLELTFTNDSII SHNFCNLTSAFNKKTFDHTLMSIVSSLHLSIRGNSNYKAVSCDFNNGITIQYNLTFSDA QSAQSQCRTFRGRVLDMFRTAFGGKYMRSGWGWTGSDGKTTWCSQTSYQYLIIQNRTWE NHCTYAGPFGMSRILLSQEKTKFFTRRLAGTFTWTLSDSSGVENPGGYCLTKWMILAAE LKCFGNTAVAKCNVNHDAEFCDMLRLIDYNKAALSKFKEDVESALHLFKTTVNSLISDQ LLMRNHLRDLMGVPYCNYSKFWYLEHAKTGETSVPKCWLVTNGSYLNETHFSDQIEQEA DNMITEMLRKDYIKRQGSTPLALMDLLMFSTSAYLVSIFLHLVKIPTHRHIKGGSCPKP HRLTNKGICSCGAFKVPGVKTVWKRR
Or select file containing sequence(s)	Browse... No file selected.
Choose sequence format	auto detect format
Choose a Prediction Method	
Prediction Method Show all the method versions:	IEDB recommended 2.19 Help on prediction method selections
Specify what to make binding predictions for	
MHC source species	human
Show only frequently occuring alleles: Select MHC allele(s) Select HLA allele reference set:	Allele: HLA-A*01:01, Length: 9 Allele: HLA-B*07:02, Length: 10 *Upload allele file*
Specify Output	
Sort peptides by	Percentile Rank
Show	All predictions
Output format	XHTML table
Email address (optional)	sdhanda@lji.org
	Submit Reset

How the tool works

1. Breaks the sequence into all possible peptides (of chosen length).
2. Predicts the binding affinity for each peptide based on the method.
3. Compares the predicted affinity to that of a large set of randomly selected peptides.
4. Assigns a percentile rank depending on individual predicted affinity.
5. Consensus picks the median rank of the methods used.

Prediction results

MHC-I Binding Prediction Results

Input Sequences

#	Name	Sequence
1	LCMV Armstrong, Protein GP	MGQIVTMFEALPHIIDEVINIVIIVLIVITGIKAVYNFATCGIFALISFL LLAGRSCGMYGLKGPDIYKGVYQFKSVEFDMSHLNLTMPNACSANNSHHY ISMGTSGLELTFTNDSIISHNFCNLTSAFNKKTFDHTLMSIVSSLHLSIR GNSNYKAVSCDFNNGITIQYNLTFSDAQSAQSQCRTFRGRVLDMFRTAFG GKYMRSGWGWTGSDGKTTWCSQTSYQYLIIQNRTWENHCTYAGPFGMSRI LLSQEKTKFFTRRLAGTFTWTLSDSSGVENPGGYCLTKWMILAAELKCFG NTAVAKCNVNHDAEFCDMLRLIDYNKAALSKFKEDVESALHLFKTTVNSL ISDQLLMRNHLRDLMGVPYCNYSKFWYLEHAKTGETSVPKCWLVTNGSYL NETHFSDQIEQEADNMITEMLRKDYIKRQGSTPLALMDLLMFSTSAYLVS IFLHLVKIPTHRHIKGGSCPKPHRLTNKGICSCGAFKVPGVKTVWKRR

Prediction method: IEDB recommended | Low percentile_rank = good binders

Download result

Citations

Check to expanded the result:

Allele	#	Start	End	Length	Peptide	Method used	Percentile_rank
HLA-B*07:02	1	471	480	10	KPHRLTNKGI	Consensus (ann/smm)	0.3
HLA-A*02:01	1	137	145	9	TLMSIVSSL	Consensus (ann/comblib_sidney2008/smm)	0.4
HLA-A*02:01	1	447	455	9	YLVSIFLHL	Consensus (ann/comblib_sidney2008/smm)	0.4
HLA-A*02:01	1	6	14	9	TMFEALPHI	Consensus (ann/comblib_sidney2008/smm)	0.5
HLA-A*02:01	1	45	53	9	ALISFLLLA	Consensus (ann/comblib_sidney2008/smm)	0.5
HLA-A*02:01	1	440	448	9	LMFSTSAYL	Consensus (ann/comblib_sidney2008/smm)	0.5
HLA-A*02:01	1	435	443	9	ALMDLLMFS	Consensus (ann/comblib_sidney2008/smm)	0.7
HLA-A*02:01	1	452	460	9	FLHLVKIPT	Consensus (ann/comblib_sidney2008/smm)	0.7
HLA-A*02:01	1	10	18	9	ALPHIIDEV	Consensus (ann/comblib_sidney2008/smm)	0.8
HLA-B*07:02	1	243	252	10	GPFGMSRILL	Consensus (ann/smm)	1.1
HLA-B*07:02	1	425	434	10	YIKRQGSTPL	Consensus (ann/smm)	1.2
HLA-A*02:01	1	14	22	9	IIDEVINIV	Consensus (ann/comblib_sidney2008/smm)	1.4
HLA-B*07:02	1	320	329	10	RLIDYNKAAL	Consensus (ann/smm)	1.6

Prediction results – Downloaded

	A	B	C	D	E	F	G	H	I	J
1	allele	seq_num	start	end	length	peptide	method	percentile_	ann_ic50	ann_ra
2	HLA-B*07:02	1	471	480	10	KPHRLTNKGI	Consensus (ann/smm)	0.3	46.84	
3	HLA-A*02:01	1	137	145	9	TLMSIVSSL	Consensus (ann/comblib_sidney2008/smm)	0.4	7.27	
4	HLA-A*02:01	1	447	455	9	YLVSIFLHL	Consensus (ann/comblib_sidney2008/smm)	0.4	11.33	
5	HLA-A*02:01	1	6	14	9	TMFEALPHI	Consensus (ann/comblib_sidney2008/smm)	0.5	4.38	
6	HLA-A*02:01	1	45	53	9	ALISFLLLA	Consensus (ann/comblib_sidney2008/smm)	0.5	18.97	
7	HLA-A*02:01	1	440	448	9	LMFSTSAYL	Consensus (ann/comblib_sidney2008/smm)	0.5	8.36	
8	HLA-A*02:01	1	435	443	9	ALMDLLMFS	Consensus (ann/comblib_sidney2008/smm)	0.7	10.53	
9	HLA-A*02:01	1	452	460	9	FLHLVKIPT	Consensus (ann/comblib_sidney2008/smm)	0.7	303.68	
10	HLA-A*02:01	1	10	18	9	ALPHIIDEV	Consensus (ann/comblib_sidney2008/smm)	0.8	19.58	
11	HLA-B*07:02	1	243	252	10	GPFGMSRILL	Consensus (ann/smm)	1.1	418.14	
12	HLA-B*07:02	1	425	434	10	YIKRQGSTPL	Consensus (ann/smm)	1.2	59.83	
13	HLA-A*02:01	1	14	22	9	IIDEVINIV	Consensus (ann/comblib_sidney2008/smm)	1.4	51.87	
14	HLA-B*07:02	1	320	329	10	RLIDYNKAAL	Consensus (ann/smm)	1.6	1113.26	
15	HLA-A*02:01	1	448	456	9	LVSIFLHLV	Consensus (ann/comblib_sidney2008/smm)	1.6	168.83	
16	HLA-A*02:01	1	42	50	9	GIFALISFL	Consensus (ann/comblib_sidney2008/smm)	1.8	38.8	
17	HLA-B*07:02	1	190	199	10	RVLDMFRTAF	Consensus (ann/smm)	2	567.7	
18	HLA-B*07:02	1	469	478	10	CPKPHRLTNK	Consensus (ann/smm)	2	8001.41	
19	HLA-A*02:01	1	320	328	9	RLIDYNKAA	Consensus (ann/comblib_sidney2008/smm)	2	197.79	
20	HLA-A*02:01	1	436	444	9	LMDLLMFST	Consensus (ann/comblib_sidney2008/smm)	2	80.28	
21	HLA-B*07:02	1	432	441	10	TPLALMDLLM	Consensus (ann/smm)	2.1	767.22	
22	HLA-A*02:01	1	38	46	9	FATCGIFAL	Consensus (ann/comblib_sidney2008/smm)	2.1	53.54	
23	HLA-A*02:01	1	285	293	9	CLTKWMILA	Consensus (ann/comblib_sidney2008/smm)	2.1	334.32	
24	HLA-A*02:01	1	17	25	9	EVINIVIIV	Consensus (ann/comblib_sidney2008/smm)	2.2	1774.67	
25	HLA-B*07:02	1	263	272	10	RLAGTFTWTL	Consensus (ann/smm)	2.6	1327.49	
26	HLA-B*07:02	1	53	62	10	AGRSCGMYGL	Consensus (ann/smm)	2.7	4778.57	
27	HLA-A*02:01	1	405	413	9	FSDQIEQEA	Consensus (ann/comblib_sidney2008/smm)	2.7	6848.72	
28	HLA-A*02:01	1	439	447	9	LLMFSTSAY	Consensus (ann/comblib_sidney2008/smm)	2.7	6789.84	
29	HLA-A*02:01	1	100	108	9	YISMGTSGL	Consensus (ann/comblib_sidney2008/smm)	2.8	102.98	
30	HLA-A*02:01	1	349	357	9	SLISDQLLM	Consensus (ann/comblib_sidney2008/smm)	2.8	89.58	
31	HLA-B*07:02	1	101	110	10	ISMGTSGLEL	Consensus (ann/smm)	2.85	4882.84	

Prediction results – citations

HLA-A*02:01	1	51	59	9	LLAGRSCGM	Consensus (ann/comblib_sidney2008/smm)	4.6
HLA-A*02:01	1	270	278	9	WTLSDSSGV	Consensus (ann/comblib_sidney2008/smm)	4.6
HLA-A*02:01	1	433	441	9	PLALMDLLM	Consensus (ann/comblib_sidney2008/smm)	4.7
HLA-B*07:02	1	242	251	10	AGPFGMSRIL	Consensus (ann/smm)	4.85
HLA-A*02:01	1	328	336	9	ALSKFKEDV	Consensus (ann/comblib_sidney2008/smm)	4.9
HLA-A*02:01	1	399	407	9	YLNETHFSD	Consensus (ann/comblib_sidney2008/smm)	5.0
HLA-A*02:01	1	445	453	9	SAYLVSIFL	Consensus (ann/comblib_sidney2008/smm)	5.0

Download result

Citations:

If you use these predictions in a manuscript, please include the following in the method section:
The MHCI binding predictions were made on 9/22/2017 using the IEDB analysis resource Consensus tool [1] which combines predictions from ANN aka NetMHC (4.0) [2][3][4], SMM [5] and Comblib [6].

1. Kim Y, Ponomarenko J, Zhu Z, Tamang D, Wang P, Greenbaum J, Lundegaard C, Sette A, Lund O, Bourne PE, Nielsen M, Peters B. 2012. Immune epitope database analysis resource. NAR.
2. Nielsen M, Lundegaard C, Worning P, Lauemøller SL, Lamberth K, Buus S, Brunak S, Lund O. 2003. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. Protein Sci 12:1007-1017.
3. Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, and Nielsen M. 2008. NetMHC-3.0: Accurate web accessible predictions of Human, Mouse, and Monkey MHC class I affinities for peptides of length 8-11. NAR 36:W509-512.
4. Andreatta M. and Nielsen M. 2016. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. Bioinformatics 32:511-7.
5. Peters B, Sette A. 2005. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. BMC Bioinformatics 6:132.
6. Sidney J, Assarsson E, Moore C, Ngo S, Pinilla C, Sette A, Peters B. 2008. Quantitative peptide binding motifs for 19 human and mouse MHC class I molecules derived using positional scanning combinatorial peptide libraries. Immunome Res 4:2.

Prediction results – email

IEDB Tools MHC class I prediction result (2017-09-22 15:57:53) Inbox x

IEDB Tools <Prediction-results-noreply@tools.iedb.org> 3:57 PM (7 minutes ago)

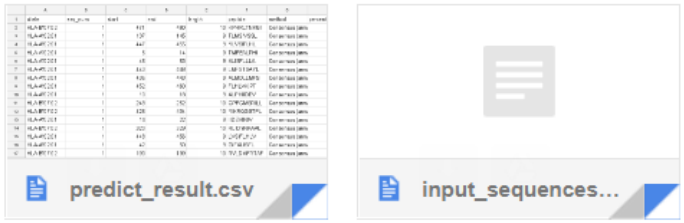
to me

Your MHC class I prediction completed on the IEDB servers (http://tools.iedb.org/mhci/) and the result is attached in csv format.

Input parameters
Method: recommended
Number of sequences: 1
Input sequences: attached
Alleles: HLA-A*02:01, HLA-B*07:02
Lengths: 9, 10

Job parameters
Submission date: 2017-09-22 15:57:53
Completion date: 2017-09-22 15:58:02
Total walltime since submission: 9 seconds

2 Attachments



Prediction results

MHC-I Binding Prediction Results

Input Sequences

#	Name	Sequence
1	LCMV Armstrong, Protein GP	MGQIVTMFEALPHIIDEVINIVIIVLIVITGIKAVYNFATCGIFALISFL LLAGRSCGMYGLKGPDIYKGVYQFKSVEFDMSHLNLTMPNACSANNSHHY ISMGTSGLELTFTNDSIISHNFCNLTSAFNKKTFDHTLMSIVSSLHLSIR GNSNYKAVSCDFNNGITIQYNLTFSDAQSAQSQCRTFRGRVLDMFRTAFG GKYMRSGWGWTGSDGKTTWCSQTSYQYLIIQNRTWENHCTYAGPFGMSRI LLSQEKTKFFTRRLAGTFTWTLSDSSGVENPGGYCLTKWMILAAELKCFG NTAVAKCNVNHDAEFCDMLRLIDYNKAALSKFKEDVESALHLFKTTVNSL ISDQLLMRNHLRDLMGVPYCNYSKFWYLEHAKTGETSVPKCWLVTNGSYL NETHFSDQIEQEADNMITEMLRKDYIKRQGSTPLALMDLLMFSTSAYLVS IFLHLVKIPTHRHIKGGSCPKPHRLTNKGICSCGAFKVPGVKTVWKRR

Prediction method: IEDB recommended | **Low percentile_rank = good binders**

Download result

Citations

Check to expanded the result:

Allele	#	Start	End	Length	Peptide	Method used	Percentile_rank
HLA-B*07:02	1	471	480	10	KPHRLTNKGI	Consensus (ann/smm)	0.3
HLA-A*02:01	1	137	145	9	TLMSIVSSL	Consensus (ann/comblib_sidney2008/smm)	0.4
HLA-A*02:01	1	447	455	9	YLVSIFLHL	Consensus (ann/comblib_sidney2008/smm)	0.4
HLA-A*02:01	1	6	14	9	TMFEALPHI	Consensus (ann/comblib_sidney2008/smm)	0.5
HLA-A*02:01	1	45	53	9	ALISFLLLA	Consensus (ann/comblib_sidney2008/smm)	0.5
HLA-A*02:01	1	440	448	9	LMFSTSAYL	Consensus (ann/comblib_sidney2008/smm)	0.5
HLA-A*02:01	1	435	443	9	ALMDLLMFS	Consensus (ann/comblib_sidney2008/smm)	0.7
HLA-A*02:01	1	452	460	9	FLHLVKIPT	Consensus (ann/comblib_sidney2008/smm)	0.7
HLA-A*02:01	1	10	18	9	ALPHIIDEV	Consensus (ann/comblib_sidney2008/smm)	0.8
HLA-B*07:02	1	243	252	10	GPFGMSRILL	Consensus (ann/smm)	1.1
HLA-B*07:02	1	425	434	10	YIKRQGSTPL	Consensus (ann/smm)	1.2
HLA-A*02:01	1	14	22	9	IIDEVINIV	Consensus (ann/comblib_sidney2008/smm)	1.4
HLA-B*07:02	1	320	329	10	RLIDYNKAAL	Consensus (ann/smm)	1.6

A **percentile rank** for a peptide is the percentage of randomly sampled peptides scoring better than the peptide.

Prediction results – expanded view

Home | Help | Example | Reference | Download | Contact

MHC-I Binding Prediction Results

Input Sequences

#	Name	Sequence
1	LCMV Armstrong, Protein GP	MGQIVTMFEALPHIIDEVINIVIIVLIVITGIKAVYNFATCGIFALISFL LLAGRSCGMYGLKGPDIYKGVYQFKSVEFDMSHLNLTMPNACSANNSHHY ISMGTSGLELTFTNDSIISHNFCNLTSAFNKKTFDHTLMSIVSSLHLSIR GNSNYKAVSCDFNNGITIQYNLTFSDAQSAQSQCRTFRGRVLDMFRTAFG GKYMRSGWGWTGSDGKTTWCSQTSYQYLIIQNRTWENHCTYAGPFGMSRI LLSQEKTKFFTRRLAGTFTWTLSDSSGVENPGGYCLTKWMILAAELKCFG NTAVAKCNVNHDAEFCDMLRLIDYNKAALSKFKEDVESALHLFKTTVNSL ISDQLLMRNHLRDLMGVPYCNYSKFWYLEHAKTGETSVPKCWLVTNGSYL NETHFSDQIEQEADNMITEMLRKDYIKRQGSTPLALMDLLMFSTSAYLVS IFLHLVKIPTHRHIKGGSCPKPHRLTNKGICSCGAFKVPGVKTVWKRR

Prediction method: IEDB recommended | Low percentile_rank = good binders

Download result

Citations

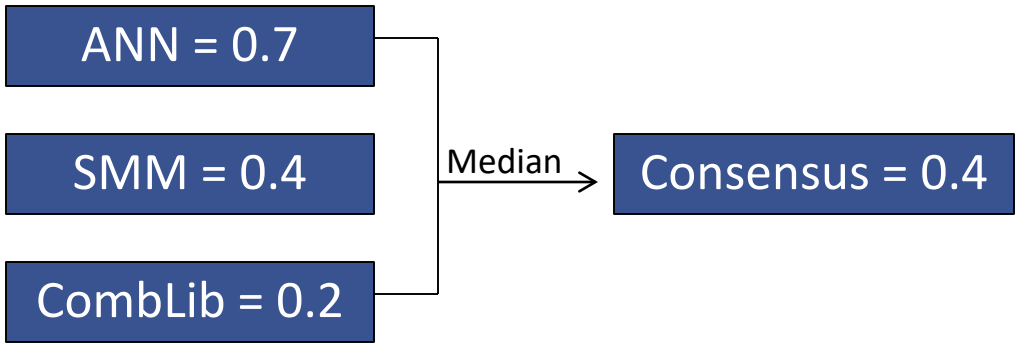
Check to expanded the result:

Individual scores for different methods

Allele	#	Start	End	Length	Peptide	Method used	Percentile_rank	ANN IC50(nM)	ANN rank	SMM IC50(nM)	SMM rank	Comblib_Sidney2008 score	Comblib_Sidney2008 rank
HLA-A*02:01	1	447	455	9	YLVSIFLHL	Consensus (ann/comblib_sidney2008/smm)	0.4	9	0.4	15.04	0.4	0.000195	12
HLA-B*07:02	1	471	480	10	KPHRLTNKGI	Consensus (ann/smm)	0.5	90	0.5	112.67	0.5	-	-
HLA-A*02:01	1	6	14	9	TMFEALPHI	Consensus (ann/comblib_sidney2008/smm)	0.5	5	0.3	26.07	0.5	0.000111	6.2
HLA-A*02:01	1	137	145	9	TLMSIVSSL	Consensus (ann/comblib_sidney2008/smm)	0.5	10	0.5	27.36	0.5	6.11e-06	0.4
HLA-A*02:01	1	440	448	9	LMFSTSAYL	Consensus (ann/comblib_sidney2008/smm)	0.5	9	0.4	21.24	0.5	0.000413	20
HLA-A*02:01	1	45	53	9	ALISFLLLA	Consensus (ann/comblib_sidney2008/smm)	0.7	17	0.7	28.32	0.5	4.47e-05	2.6
HLA-A*02:01	1	435	443	9	ALMDLLMFS	Consensus (ann/comblib_sidney2008/smm)	0.7	14	0.7	37.16	0.7	1.05e-05	0.7
HLA-A*02:01	1	10	18	9	ALPHIIDEV	Consensus (ann/comblib_sidney2008/smm)	0.8	19	0.8	43.26	0.8	7.66e-05	4.4
HLA-B*07:02	1	243	252	10	GPFGMSRILL	Consensus (ann/smm)	1.3	421	1.1	351.41	1.5	-	-
HLA-A*02:01	1	14	22	9	IIDEVINIV	Consensus (ann/comblib_sidney2008/smm)	1.4	33	1.2	78.18	1.4	0.000142	7.7
HLA-A*02:01	1	42	50	9	GIFALISFL	Consensus (ann/comblib_sidney2008/smm)	1.8	33	1.2	108.17	1.8	5.42e-05	3.2
HLA-A*02:01	1	448	456	9	LVSIFLHLV	Consensus (ann/comblib_sidney2008/smm)	1.8	57	1.8	95.96	1.6	0.0006	25
HLA-B*07:02	1	320	329	10	RLIDYNKAAL	Consensus (ann/smm)	2.0	905	1.7	595.42	2.3	-	-
HLA-B*07:02	1	425	434	10	YIKRQGSTPL	Consensus (ann/smm)	2.05	952	1.8	575.20	2.3	-	-
HLA-A*02:01	1	38	46	9	FATCGIFAL	Consensus (ann/comblib_sidney2008/smm)	2.1	48	1.6	139.99	2.1	0.000256	14
HLA-A*02:01	1	320	328	9	RLIDYNKAA	Consensus (ann/comblib_sidney2008/smm)	2.3	106	2.3	129.45	2	0.000124	6.8
HLA-A*02:01	1	436	444	9	LMDLLMFST	Consensus (ann/comblib_sidney2008/smm)	2.3	106	2.3	126.80	2	0.000834	31
HLA-B*07:02	1	190	199	10	RVLDMFRTAF	Consensus (ann/smm)	2.45	766	1.6	816.24	3.3	-	-
HLA-B*07:02	1	432	441	10	TPLALMDLLM	Consensus (ann/smm)	2.6	975	1.8	854.71	3.4	-	-
HLA-A*02:01	1	100	108	9	YISMGTSGL	Consensus (ann/comblib_sidney2008/smm)	2.8	115	2.4	209.46	2.8	0.00108	36

Consensus

- Combines **ANN**, **SMM** and **CombLib**
- Requires that all methods give predictions on the same scale – **percentile ranks** are used
- Moutaftsi M et al. (2006) PMID: 16767078



Allele	#	Start	End	Length	Peptide	Method used	Percentile rank	ANN IC50(nM)	ANN rank	SMM IC50(nM)	SMM rank	Comblib_Sidney2008 score	Comblib_Sidney2008 rank
HLA-A*02:01	1	769	777	9	FLYFVIFFV	Consensus (ann/comblib_sidney2008/smm)	0.1	2	0.1	0.30	0.1	1.51e-07	0.1
HLA-A*02:01	1	285	293	9	MLAAQMFIV	Consensus (ann/comblib_sidney2008/smm)	0.2	4	0.2	5.90	0.2	1.51e-05	1
HLA-A*02:01	1	322	330	9	MMMNWSPTA	Consensus (ann/comblib_sidney2008/smm)	0.2	3	0.2	7.30	0.2	7.59e-05	4.4
HLA-A*02:01	1	705	713	9	FMYGLSPAL	Consensus (ann/comblib_sidney2008/smm)	0.2	3	0.2	6.43	0.2	3.47e-05	2
HLA-A*02:01	1	846	854	9	FLWWLCYLL	Consensus (ann/comblib_sidney2008/smm)	0.2	3	0.2	0.72	0.1	3.98e-06	0.3
HLA-A*02:01	1	611	619	9	CLIDYPYRL	Consensus (ann/comblib_sidney2008/smm)	0.3	5	0.3	10.00	0.3	3.22e-05	1.9
HLA-A*02:01	1	842	850	9	LLSRFLWWL	Consensus (ann/comblib_sidney2008/smm)	0.3	6	0.3	13.46	0.3	0.000321	16
HLA-A*02:01	1	132	140	9	DLMGYIPVV	Consensus (ann/comblib_sidney2008/smm)	0.4	14	0.7	16.72	0.4	1.49e-06	0.2
HLA-A*02:01	1	177	185	9	FLLALLSCI	Consensus (ann/comblib_sidney2008/smm)	0.4	10	0.5	10.52	0.3	4.68e-06	0.4
HLA-A*02:01	1	181	189	9	LLSCITTPV	Consensus (ann/comblib_sidney2008/smm)	0.4	5	0.3	14.19	0.4	1.09e-05	0.8
HLA-A*02:01	1	618	626	9	RLWHYPCTV	Consensus (ann/comblib_sidney2008/smm)	0.4	9	0.4	6.04	0.2	0.000116	6.4
HLA-A*02:01	1	727	735	9	FLLLADARV	Consensus (ann/comblib_sidney2008/smm)	0.4	17	0.7	18.37	0.4	3.47e-06	0.3

How to choose “binders”

Selection of “binders”

1. Pick peptides below **percentile rank 1.0** (reported by the tool, not percentage of total peptides)

2. Pick peptides below predicted binding affinity of **500 nM**
    - IC50 < 50 nM - high affinity
    - IC50 < 500 nM - intermediate affinity
    - IC50 < 5000 nM - low affinity
    - Sette et al. 1994, J. Immunology (PMID: 7527444)
    - Ensures that all peptides have reasonable affinity (at least intermediate affinity).
3. Pick **top 1% of peptides** for each allele/length combination to cover most of immune responses
    - Moutaftsi et al. 2006 (PMID: 16767078), Kotturi et al. 2007 (PMID: 17329346)
    - Ensures equal number of peptides per allele
4. Select based on **allele specific binding affinity** threshold

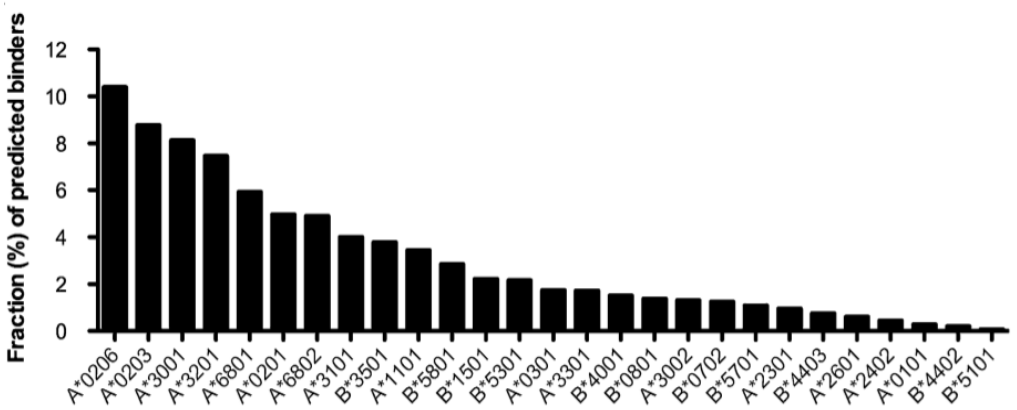
Different peptide-binding repertoires

The size of the peptide repertoire binding at a given affinity varies between alleles.

All peptides
Binders

HLA-A*02:01

HLA-A*01:01



Allele-specific affinity cutoffs

Paul et al. 2013, Journal of Immunology

Allele-specific thresholds

IEDB Analysis Resource

Home | Help | Example | Reference | Download | Contact

MHC-I binding predictions - Tutorial

Guidelines for selecting thresholds (cut-offs) for MHC class I and II binding predictions can be found here.

Selecting thresholds (cut-offs) for MHC class I and II binding predictions

Ward Fleri
posted this on May 21, 2013 04:33 PM

MHC class I

For MHC class I T cell epitope predictions, selection of predicted binders can be done based on the percentile rank or MHC binding affinity. The IEDB currently recommends making selections based on a percentile rank of <= 1% for each (MHC allele, length) combination to cover most of the immune responses.[1, 2] Alternatively, a binding affinity (IC50) threshold of 500 nM identifies peptide binders recognized by T cells and this threshold can be used to select peptides.[3] Recently, a paper from our group showed that absolute binding affinity threshold correlates better with immunogenicity and also that, for even better correlation, MHC-specific thresholds should be used.[4] The tables below show the allele-specific thresholds for the 38 most common HLA-A and HLA-B alleles, representative of the nine major supertypes. The tables can also be downloaded as an RTF file (see attached file).

Alleles sorted by population frequency		
Allele	Population frequency of allele	Allele specific affinity cutoff (IC50 nM)
A*0201	25.2	255
A*2402	16.8	849
A*0101	16.2	884
A*0301	15.4	602
B*0702	13.3	687
A*1101	12.9	382
B*0801	11.5	663
B*4001	10.3	639
B*4402	9.2	904
B*4403	7.6	780
B*3501	6.5	348
A*2301	6.4	740
A*3201	5.7	131
B*5101	5.5	939
B*5301	5.4	538
B*1501	5.2	528
A*3001	5.1	109
A*3002	5	674

Alleles sorted by name		
Allele	Population frequency of allele	Allele specific affinity cutoff (IC50 nM)
A*0101	16.2	884
A*0201	25.2	255
A*0203	3.3	92
A*0206	4.9	60
A*0301	15.4	602
A*1101	12.9	382
A*2301	6.4	740
A*2402	16.8	849
A*2501	2.5	795
A*2601	4.7	815
A*2902	2.9	641
A*3001	5.1	109
A*3002	5	674
A*3101	4.7	329
A*3201	5.7	131
A*3301	3.2	606
A*6801	4.6	197
A*6802	3.3	259

Recommendations

- Both approaches (**affinity and ranking**) are reasonable, and have been applied in numerous studies
- Cut-offs can be **combined** (peptides in top 1% and IC50 < 500 nM)
- Current studies suggest that **allele specific thresholds** can be derived.

Exercises

Exercise 1:

- Find the best epitope candidate of length 9 for HLA-A*02:01 from SARS spike glycoprotein (GenBank accession no: ABD72984.1)

Solution:

- Collect sequence from GenBank (NCBI Protein) - https://www.ncbi.nlm.nih.gov/protein/
- Copy sequence into the prediction tool
- Select prediction method as “IEDB recommended”
- Select species as “Human”
- Select the allele as HLA-A*02:01 & length as 9
- Submit

MHC-I Binding Predictions

Prediction Method Version	2013-02-22 [Older versions]
Specify Sequence(s)	
Enter protein sequence(s) in FASTA format or as whitespace-separated sequences. (Browse for sequences in NCBI)	>ABD72984.1 spike glycoprotein [SARS coronavirus] MFIFILFLTLTSGSDLDRCTTFDDVQAPNYTQHTSSMRGVYYPDEIFRSDTLYLTQDLFLPFYSNVTGFH TINHTFGNPVIPFKDGIYFAATEKSNVVRGWVFGSTMNNKSQSVIIINNSTNVVIRACNFELCDNPFFAV SKPMGTQTHTMIFDNAFNCTFEYISDAFSLDVSEKSGNFKHLREFVFKNKDGFLYVYKGYQPIDVVRDLP SGFNTLKPIFKLPLGINITNFRAILTAFSPAQDIWGTSAAAYFVGYLKPTTFMLKYDENGTITDAVDCSQ NPLAELKCSVKSFEIDKGIYQTSNFRVVPSGDVVRFPNITNLCPFGEVFNATKFPSVYAWERKKISNCVA DYSVLYNSTFFSTFKCYGVSATKLNDLCFSNVYADSFVVKGDDVRQIAPGQTGVIADYNYKLPDDFMGCV LAWNTRNIDATSTGNYNYKYRYLRHGKLRPFERDISNVPFSPDGKPCTPPALNCYWPLNDYGFYTTTGIG YQPYRVVVLSFELLNAPATVCGPKLSTDLIKNQCVNFNFNGLTGTGVLTPSSKRFQPFQQFGRDVSDFTD SVRDPKTSEILDISPCSFGGVSVITPGTNASSEVAVLYQDVNCTDVSTAIHADQLTPAWRIYSTGNNVFQ TOAGCLIGAEHVDTSYECDIPIGAGICASYHTVSLLRSTSOKSIVAYTMSLGADSSIAYSNNTIAIPTNF
Or select file containing sequence(s)	Browse... No file selected.
Choose sequence format	auto detect format
Choose a Prediction Method	
Prediction Method Show all the method versions:	IEDB recommended 2.19 Help on prediction method selections
Specify what to make binding predictions for	
MHC source species	human
Show only frequently occuring alleles: Select MHC allele(s) Select HLA allele reference set:	Allele: HLA-A*02:01 Length: 9 Upload allele file
Specify Output	
Sort peptides by	Percentile Rank
Show	All predictions
Output format	XHTML table
Email address (optional)	
	Submit Reset

MHC-I Binding Prediction Results

Input Sequences

#	Name	Sequence
1	ABD72984.1 spike glycoprotein [SARS coronavirus]	MFIFILFLTLTSGSDLDRCTTFDDVQAPNYTQHTSSMRGVYYPDEIFRSD TLYLTQDLFLPFYSNVTGFHTINHTFGNPVIPFKDGIYFAATEKSNVVRG WVFGSTMNNKSQSVIIINNSTNVVIRACNFELCDNPFFAVSKPMGTQTHT MIFDNAFNCTFEYISDAFSLDVSEKSGNFKHLREFVFKNKDGFLYVYKGY QPIDVVRDLPSGFNTLKPIFKLPLGINITNFRAILTAFSPAQDIWGTSAA AYFVGYLKPTTFMLKYDENGTITDAVDCSQNPLAELKCSVKSFEIDKGIY QTSNFRVVPSGDVVRFPNITNLCPFGEVFNATKFPSVYAWERKKISNCVA DYSVLYNSTFFSTFKCYGVSATKLNDLCFSNVYADSFVVKGDDVRQIAPG QTGVIADYNYKLPDDFMGCVLAWNTRNIDATSTGNYNYKYRYLRHGKLRP FERDISNVPFSPDGKPCTPPALNCYWPLNDYGFYTTTGIGYQPYRVVVLS FELLNAPATVCGPKLSTDLIKNQCVNFNFNGLTGTGVLTPSSKRFQPFQQ FGRDVSDFTDSVRDPKTSEILDISPCSFGGVSVITPGTNASSEVAVLYQD VNCTDVSTAIHADQLTPAWRIYSTGNNVFQTQAGCLIGAEHVDTSYECDI PIGAGICASYHTVSLLRSTSQKSIVAYTMSLGADSSIAYSNNTIAIPTNF SISITTEVMPVSMAKTSVDCNMYICGDSTECANLLLQYGSFCTQLNRALS GIAAEQDRNTREVFAQVKQMYKTPTLKYFGGFNFSQILPDPLKPTKRSFI EDLLFNKVTLADAGFMKQYGECLGDINARDLICAQKFNGLTVLPPLLTDD MIAAYTAALVSGTATAGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYE NQKQIANQFNKAISQIQESLTTTSTALGKLQDVVNQNAQALNTLVKQLSS NFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEI RASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQAAPHGVVFLHVTYV PSQERNFTTAPAICHEGKAYFPREGVFVFNGTSWFITQRNFFSPQIITTD NTFVSGNCDVVIGIINNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGD ISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYVWL GFIAGLIAIVMVTILLCCMTSCCSCLKGACSCGSCCKFDEDDSEPVLKGV KLHYT

Prediction method: IEDB recommended 2.19 | Low percentile_rank = good binders

Download result

Citations

Check to expand the result:

Allele	#	Start	End	Length	Peptide	Method used	Percentile_rank
HLA-A*02:01	1	700	708	9	FSISITTEV	Consensus (ann/comblib_sidney2008/smm)	0.34
HLA-A*02:01	1	1202	1210	9	FIAGLIAIV	Consensus (ann/comblib_sidney2008/smm)	0.4
HLA-A*02:01	1	982	990	9	RLQSLQTYV	Consensus (ann/comblib_sidney2008/smm)	0.7
HLA-A*02:01	1	354	362	9	VLYNSTFFS	Consensus (ann/comblib_sidney2008/smm)	0.73

Exercise 2

- Finding malaria minimal epitope from AMA1 15-mer peptide

Background

- Apical membrane antigen-1 (AMA1) is a protein expressed in the membrane of *P. falciparum* sporozoite liver and blood stages. In clinical trials AMA1 gives both $CD4^+$ & $CD8^+$ responses and is considered a good multi-antigen malaria vaccine candidate.

Methods

- Five volunteers were immunized with a vaccine containing full length of AMA1. A peptide pool of 15-mers overlapping by 11 amino acids in the AMA1 sequence was constructed. ELISpot responses of the peptides from the peptide pool were tested among the volunteers. HLA typing was done for each volunteer.

Reference

- *Sedegah M. et al. 2010, Malaria Journal* (PMID: 20735847)

Exercise 2

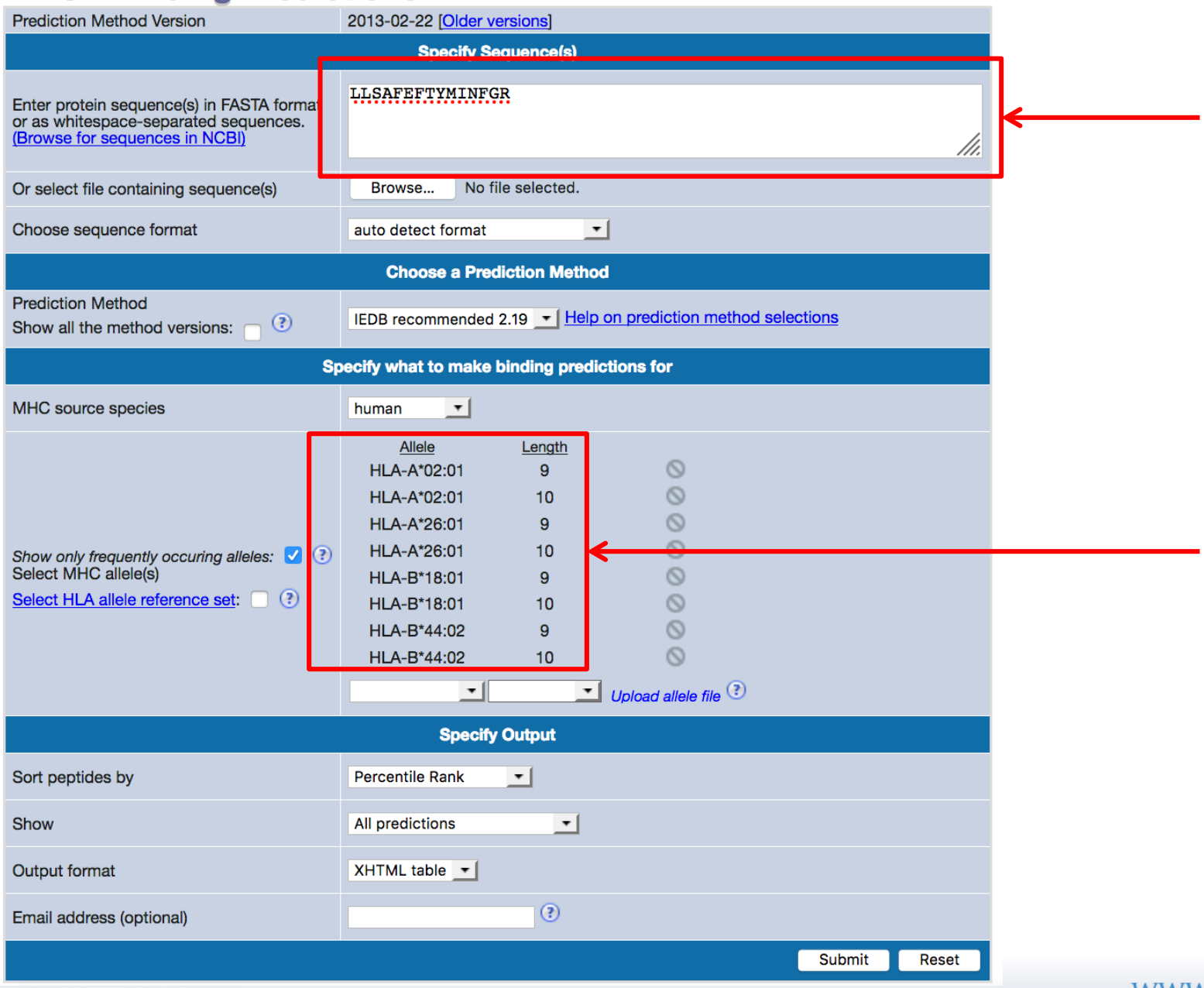
Problem statement

- Use IEDB prediction tools to determine the minimal (length 9-10) epitope within the 15-mer **LLSAFEFTYMINFGR**
- Volunteer's HLA set: HLA-A*02:01, HLA-A*26:01, HLA-B*18:01, HLA-B*44:02.

Solution:

- Copy the 15-mer peptide into the prediction tool
- Select prediction method as "IEDB recommended"
- Select species as "Human"
- Select the volunteer's allele sets
- Select lengths 9 & 10
- Submit

MHC-I Binding Predictions



MHC-I Binding Prediction Results

Input Sequences

#	Name	Sequence
1	ws-separated-0	LLSAFEFTYMINFGR

Prediction method: IEDB recommended 2.19 | Low percentile_rank = good binders

Download result

Citations

Check to expand the result:

Allele	#	Start	End	Length	Peptide	Method used	Percentile_rank
HLA-B*18:01	1	5	13	9	FEFTYMINF	Consensus (ann/smm)	0.12
HLA-B*44:02	1	5	13	9	FEFTYMINF	Consensus (ann/smm)	0.57
HLA-A*02:01	1	1	10	10	LLSAFEFTYM	Consensus (ann/smm)	0.63
HLA-B*44:02	1	4	13	10	AFEFTYMINF	Consensus (ann/smm)	0.68
HLA-A*02:01	1	3	11	9	SAFEFTYMI	Consensus (ann/comblib_sidney2008/smm)	2.3
HLA-B*18:01	1	5	14	10	FEFTYMINFG	Consensus (ann/smm)	2.95
HLA-A*26:01	1	2	10	9	LSAFEFTYM	Consensus (ann/smm)	3.1
HLA-A*02:01	1	2	11	10	LSAFEFTYMI	Consensus (ann/smm)	4.55
HLA-B*18:01	1	1	9	9	LLSAFEFTY	Consensus (ann/smm)	4.8
HLA-B*44:02	1	3	11	9	SAFEFTYMI	Consensus (ann/smm)	5.4

La Jolla Institute FOR ALLERGY AND IMMUNOLOGY | Life Without Disease.