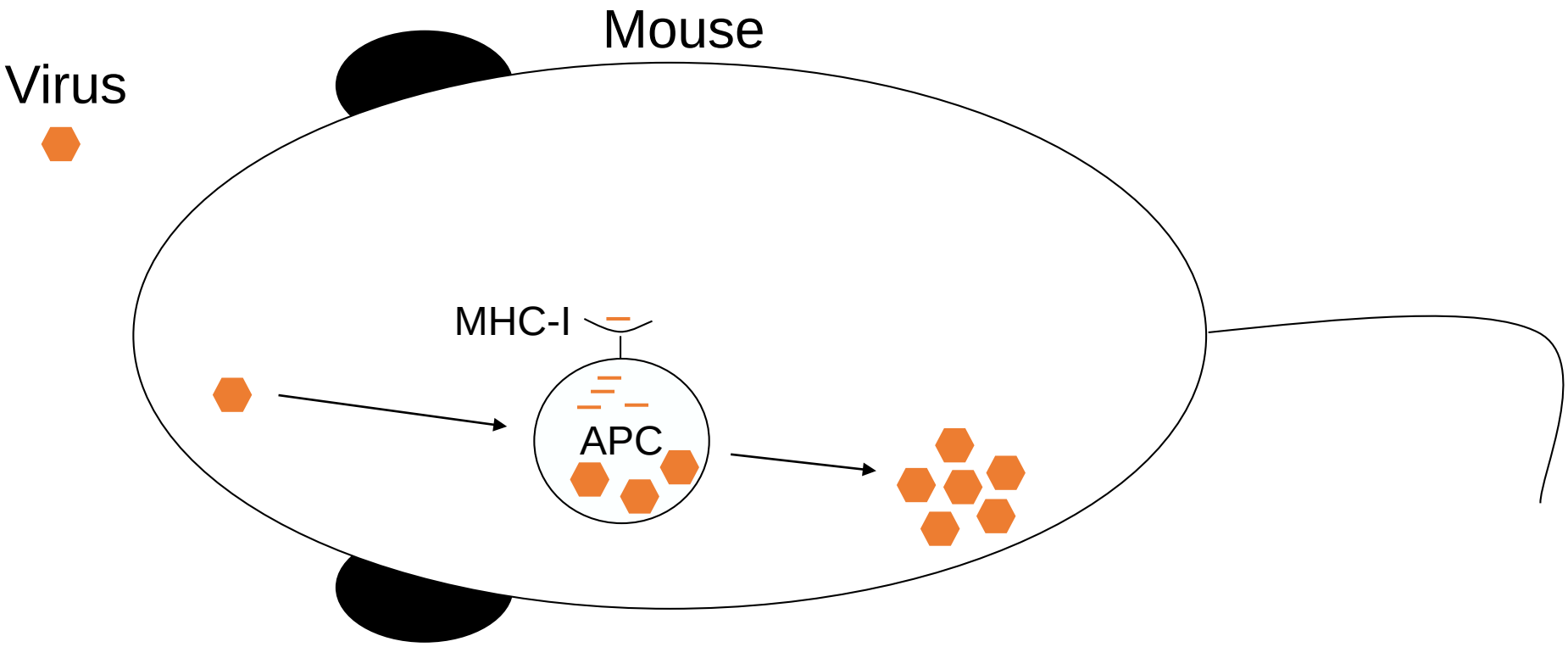
T Cell processing & immunogenicity predictions

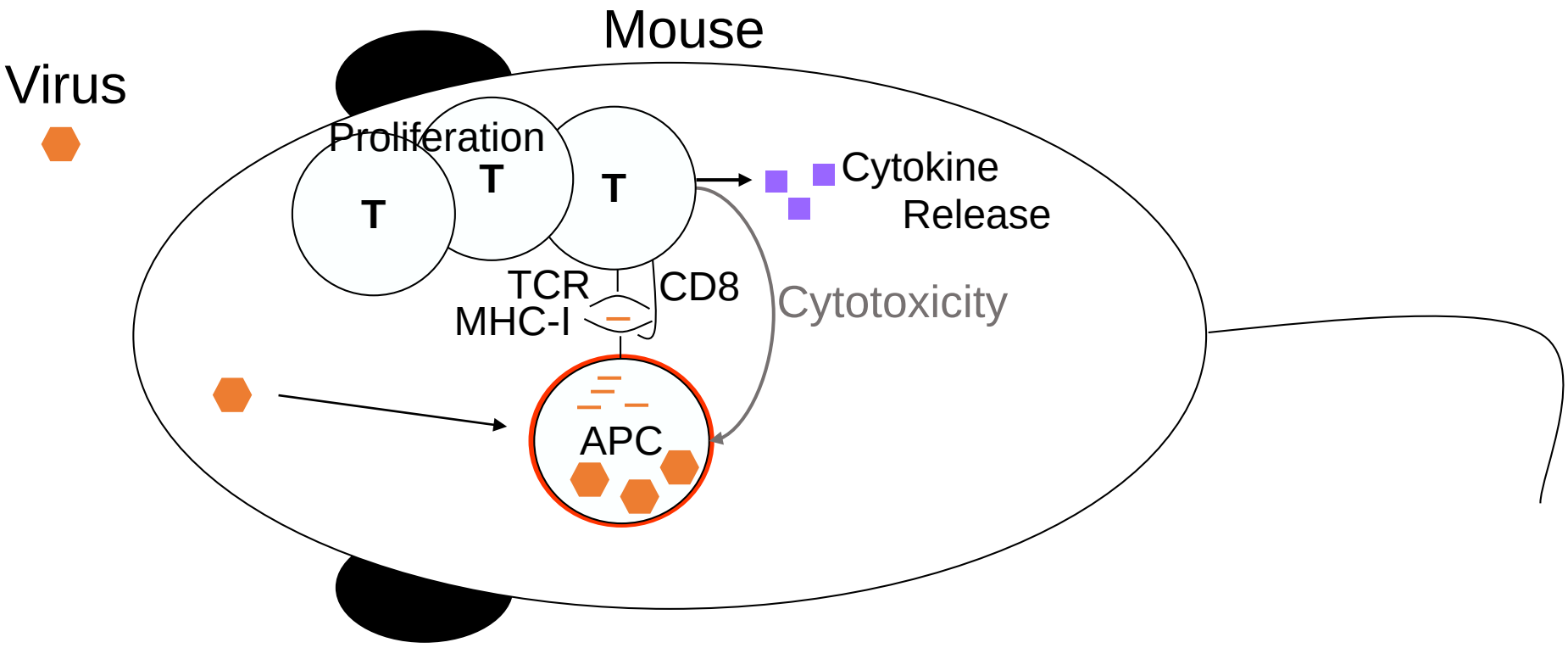
tools.iedb.org

Presented by: Bjoern Peters, PI

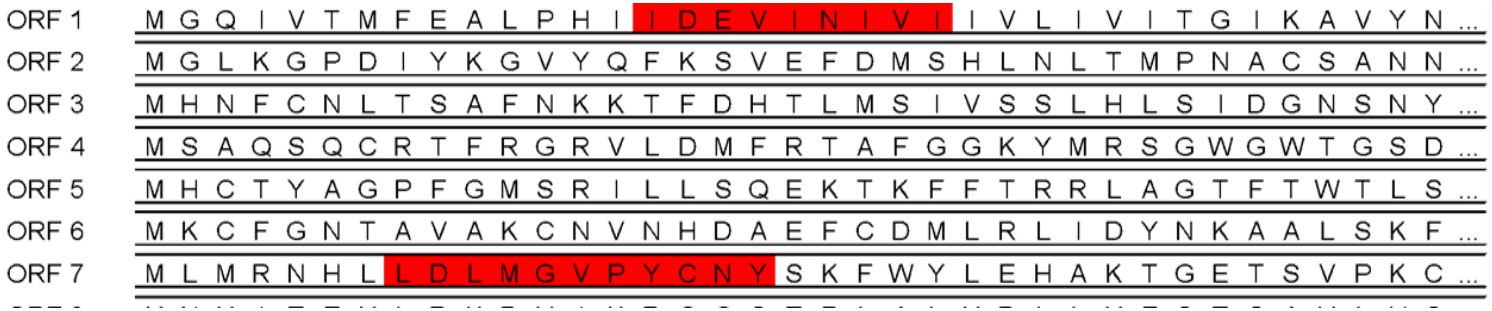
CD8+ T cell epitopes in viral infection



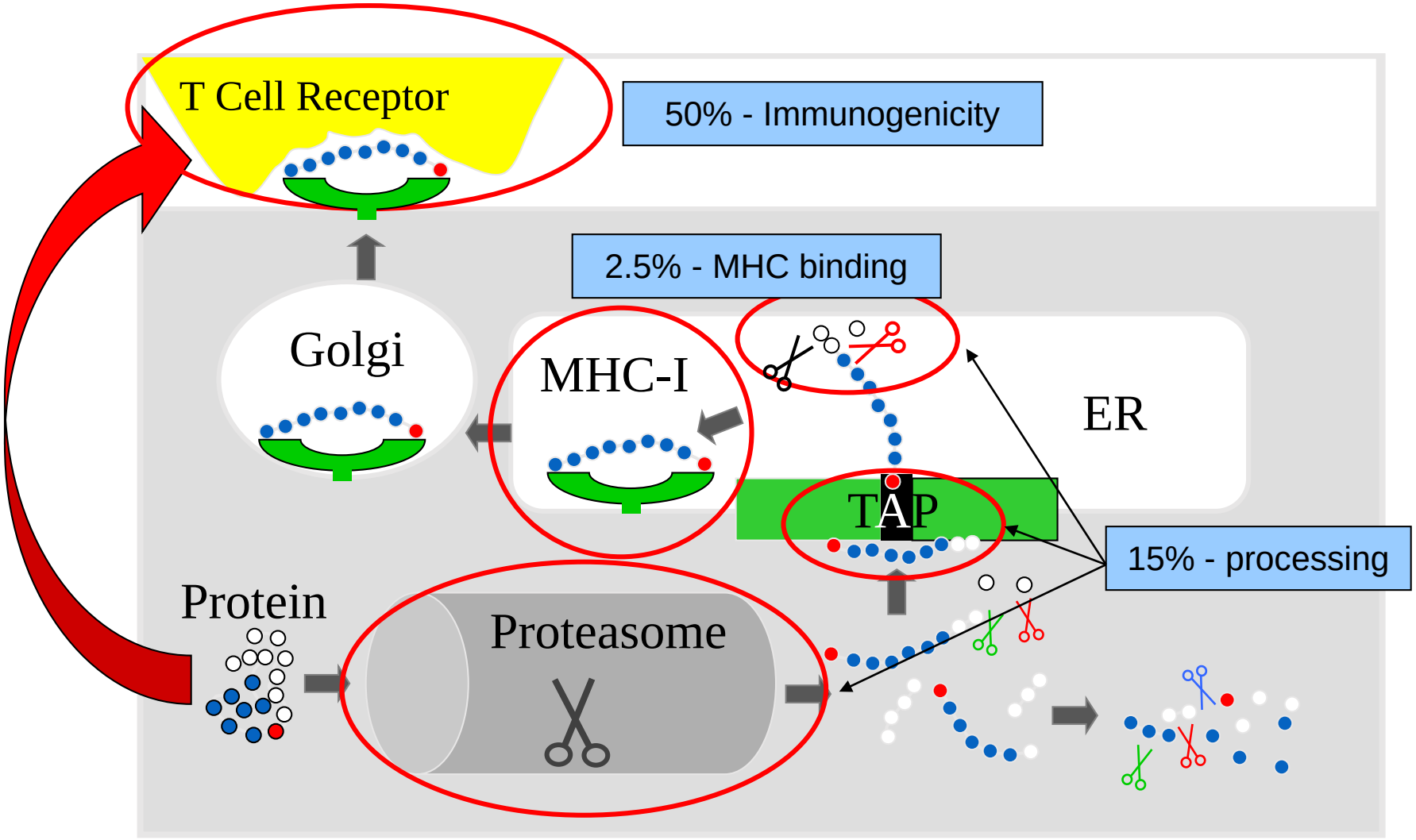
CD8+ T cell epitopes in viral infection



T cell epitope mapping



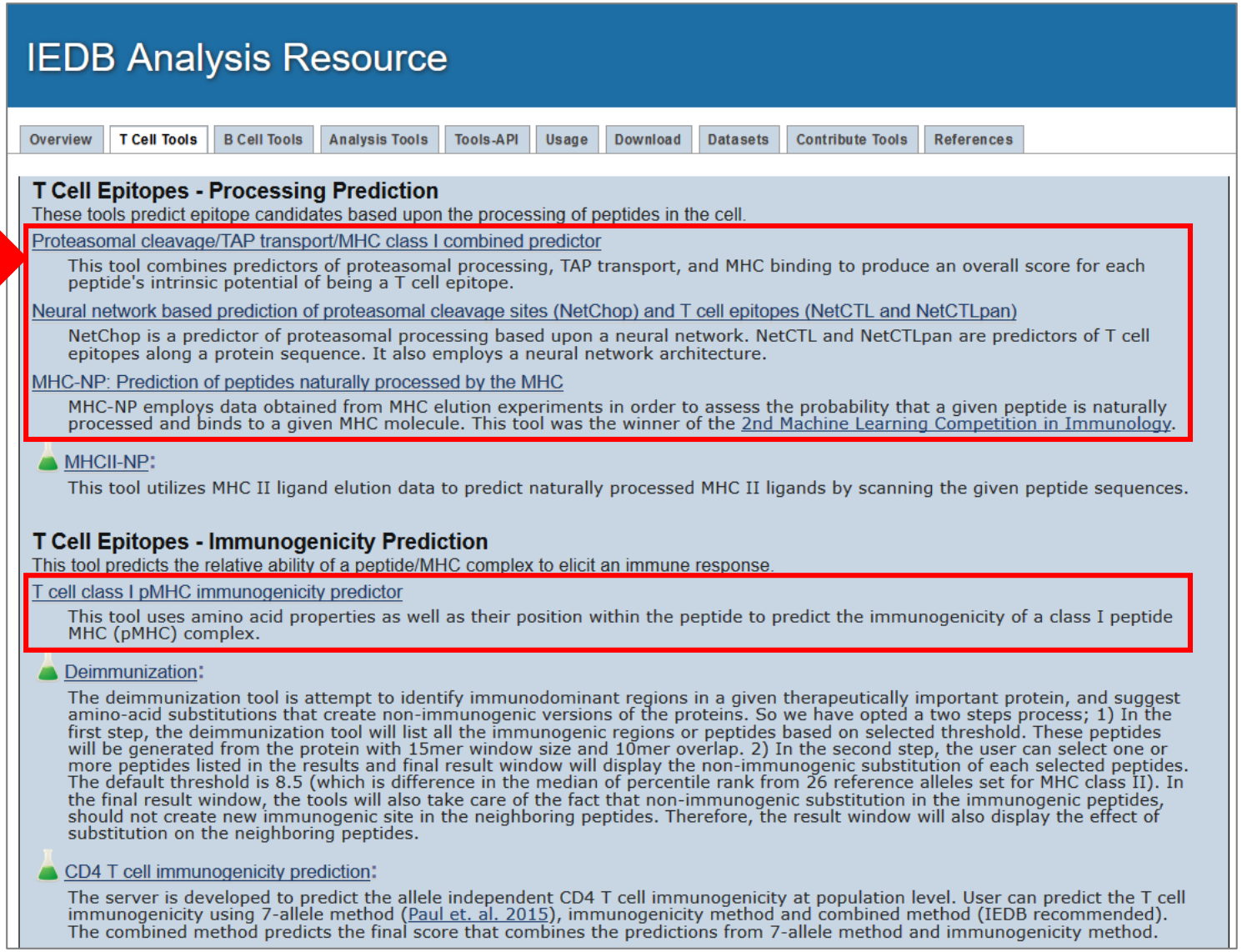
MHC I - Antigen processing and presentation pathway



Peters et al, J Mol Biol 2002, Bioinformatics 2003, J Immunol. 2003; CMLS 2005; Assarson, J Immunol 2007

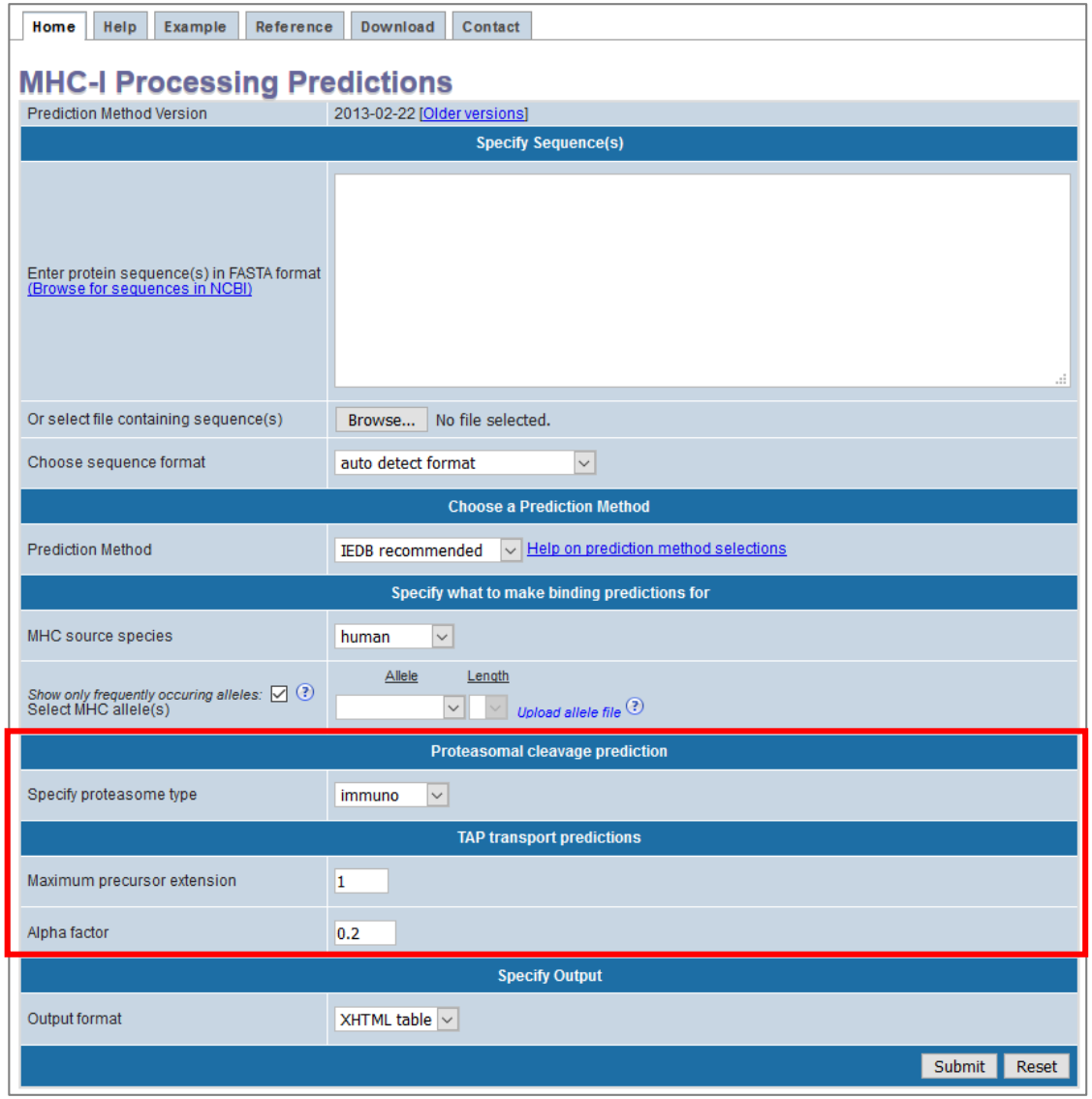
Class I Processing + immunogenicity tools available in the IEDB

http://tools.iedb.org/main/tcell/



Class I 'combined predictor'

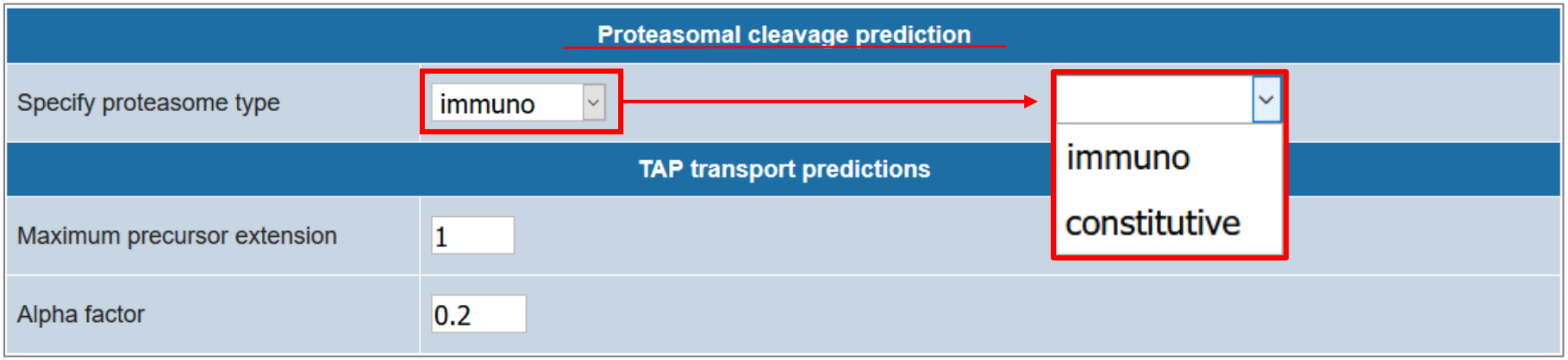
http://tools.iedb.org/processing/



- Combines predictions for:
  - proteasomal cleavage
  - TAP transport
- Trained on specific in vitro datasets

Proteasomal cleavage

http://tools.iedb.org/processing/



- Proteasomes create the C-terminal end of peptides
- Prediction looks for sequence motive up and downstream of potential cleavage site
- Cleavage sequence motif determined based on in vitro protein digests by proteasomes

TAP Transport

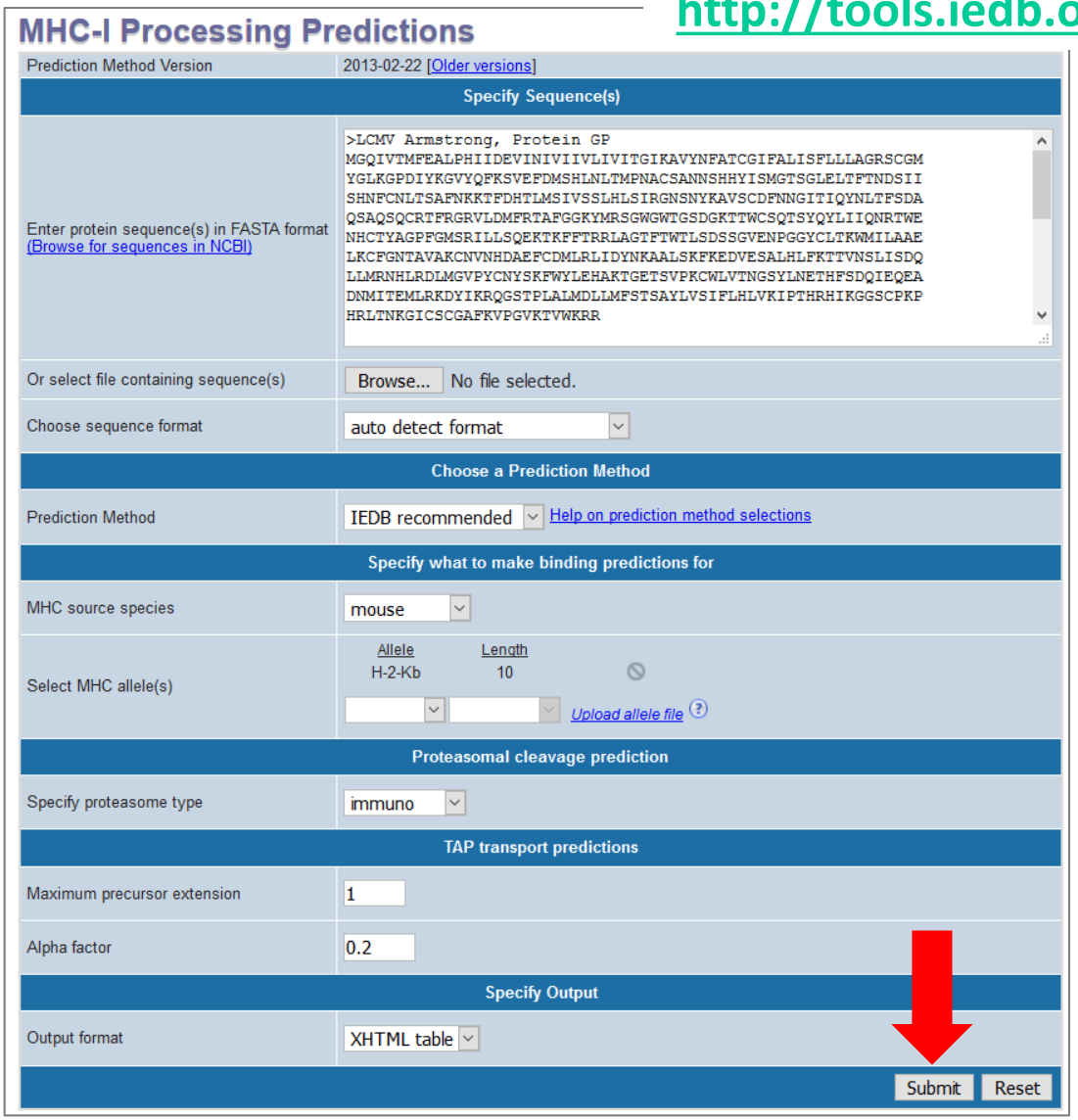
http://tools.iedb.org/processing/

Proteasomal cleavage prediction	
Specify proteasome type	immuno
TAP transport predictions	
Maximum precursor extension	1
Alpha factor	0.2

- TAP transport efficiency of peptides is sequence dependent; motif derived based on in vitro assays
- Overall TAP transport efficiency of a presented MHC ligand can be result of a collection of precursors
- Unless paper specifically read and details about the precursor length distribution are known, **keep parameters unchanged**

Class I 'combined predictor' - example

http://tools.iedb.org/processing/



Class I 'combined predictor' - example

MHC-I Processing Prediction Results

http://tools.iedb.org/processing/

Input Sequences

#	Name	Sequence
1	LCMV Armstrong, Protein GP	MGQIVTMFEALPHIIDEVINIVIIVLIVITGIKAVYNFATCGIFALISFLLLAGRS CGMYGLKGPDIYKGVYQFKSVEFDMSHLNLTMPNACSANNSHHYISMGTSGLELTF TNDSIISHNFCNLTSAFNKKTFDHTLMSIVSSLHLSIRGNSNYKAVSCDFNNGITI QYNLTFSDAQSAQSQCRTFRGRVLDMFRTAFGGKYMRSGWGWTGSDGKTTWCSQTS YQYLIIQNRTWENHCTYAGPFGMSRILLSQEKTKFFTRRLAGTFTWTLSDSSGVEN PGGYCLTKWMILAAELKCFGNTAVAKCNVNHDAEFCDMLRLIDYNKAALSKFKEDV ESALHLFKTTVNSLISDQLLMRNHLRDLMGVPYCNYSKFWYLEHAKTGETSVPKCW LVTNGSYLNETHFSDQIEQEADNMITEMLRKDYIKRQGSTPLALMDLLMFSTSAYL VSIFLHLVKIPTHRHIKGGSCPKPHRLTNKGICSCGAFKVPGVKTVWKRR
2	LCMV Armstrong, Protein NP	MSLSKEVKSFQWTQALRRELQSFTSDVKAAVIKDATNLLNGLDFSEVSNVQRIMRK EKRDDKDLQRLRSLNQTVHSLVDLKSTSKKNVLKVGRLSAEELMSLAADLEKLKAK IMRSERPQASGVYMGNLTTQQLDQRSQILQIVGMRKPQQGASGVVRVWDVKDSSLL NNQFGTMPSLTMACMAKQSQTPLNDVVQALTDLGLLYTVKYPNLNDLERLKDKHPV LGVITEQQSSINISGYNFSLGAAVKAGAALLDGGNMLESILIKPSNSEDLLKAVLG AKRKLNMFVSDQVGDRNPYENILYKVCLSGEGWPYIACRTSIVGRAWENTTIDLTS EKPAVNSPRPAPGAAGPPQVGLSYSQTMLLKDLMGGIDPNAPTWIDIEGRFNDPVE IAIFQPQNGQFIHFYREPVDQKQFKQDSKYSHGMDLADLFNAQPGLTSSVIGALPQ GMVLSCQGSDDIRKLLDSQNRKDIKLIDVEMTREASREYEDKVWDKYGWLCKMHTG IVRDKKKKEITPHCALMDCIIFESASKARLPDLKTVHNILPHDLIFRGPNVVTL

Prediction method: recommended | High Score = high efficiency

Download result

Citations

Allele	#	Start	End	Peptide Length	Peptide	Proteasome Score	TAP Score	MHC Score	Processing Score	Total Score	MHC IC50[nM]
H-2-Kb	2	203	212	10	LLYTVKYPNL	1.79	0.50	-2.01	2.28	0.27	103.5
H-2-Kb	1	116	125	10	SIISHNFCNL	1.51	0.46	-1.90	1.97	0.08	78.6
H-2-Kb	2	7	16	10	VKSFQWTQAL	1.42	0.39	-2.01	1.81	-0.21	102.9
H-2-Kb	2	235	244	10	INISGYNFSL	1.53	0.41	-2.18	1.95	-0.23	149.7
H-2-Kb	1	35	44	10	VYNFATCGIF	1.42	1.27	-3.12	2.68	-0.44	1327
H-2-Kb	1	75	84	10	KSVEFDMSHL	1.50	0.47	-2.45	1.96	-0.49	283.6
H-2-Kb	1	369	378	10	YCNYSKFWYL	1.45	0.31	-2.32	1.76	-0.56	206.8
H-2-Kb	1	449	458	10	VSIFLHLVKI	1.33	0.33	-2.25	1.66	-0.59	178.2

Class I 'combined predictor' - example

http://tools.iedb.org/processing/

Allele	#	Start	End	Peptide Length	Peptide	Proteasome Score	TAP Score	MHC Score	Processing Score	Total Score	MHC IC50[nM]
H-2-Kb	2	203	212	10	LLYTVKYPNL	1.79	0.50	-2.01	2.28	0.27	103.5
H-2-Kb	1	116	125	10	SIISHNFCNL	1.51	0.46	-1.90	1.97	0.08	78.6
H-2-Kb	2	7	16	10	VKSFQWTQAL	1.42	0.39	-2.01	1.81	-0.21	102.9
H-2-Kb	2	235	244	10	INISGYNFSL	1.53	0.41	-2.18	1.95	-0.23	149.7
H-2-Kb	1	35	44	10	VYNFATCGIF	1.42	1.27	-3.12	2.68	-0.44	1327
H-2-Kb	1	75	84	10	KSVEFDMSHL	1.50	0.47	-2.45	1.96	-0.49	283.6
H-2-Kb	1	369	378	10	YCNYSKFWYL	1.45	0.31	-2.32	1.76	-0.56	206.8
H-2-Kb	1	449	458	10	VSIFLHLVKI	1.33	0.33	-2.25	1.66	-0.59	178.2

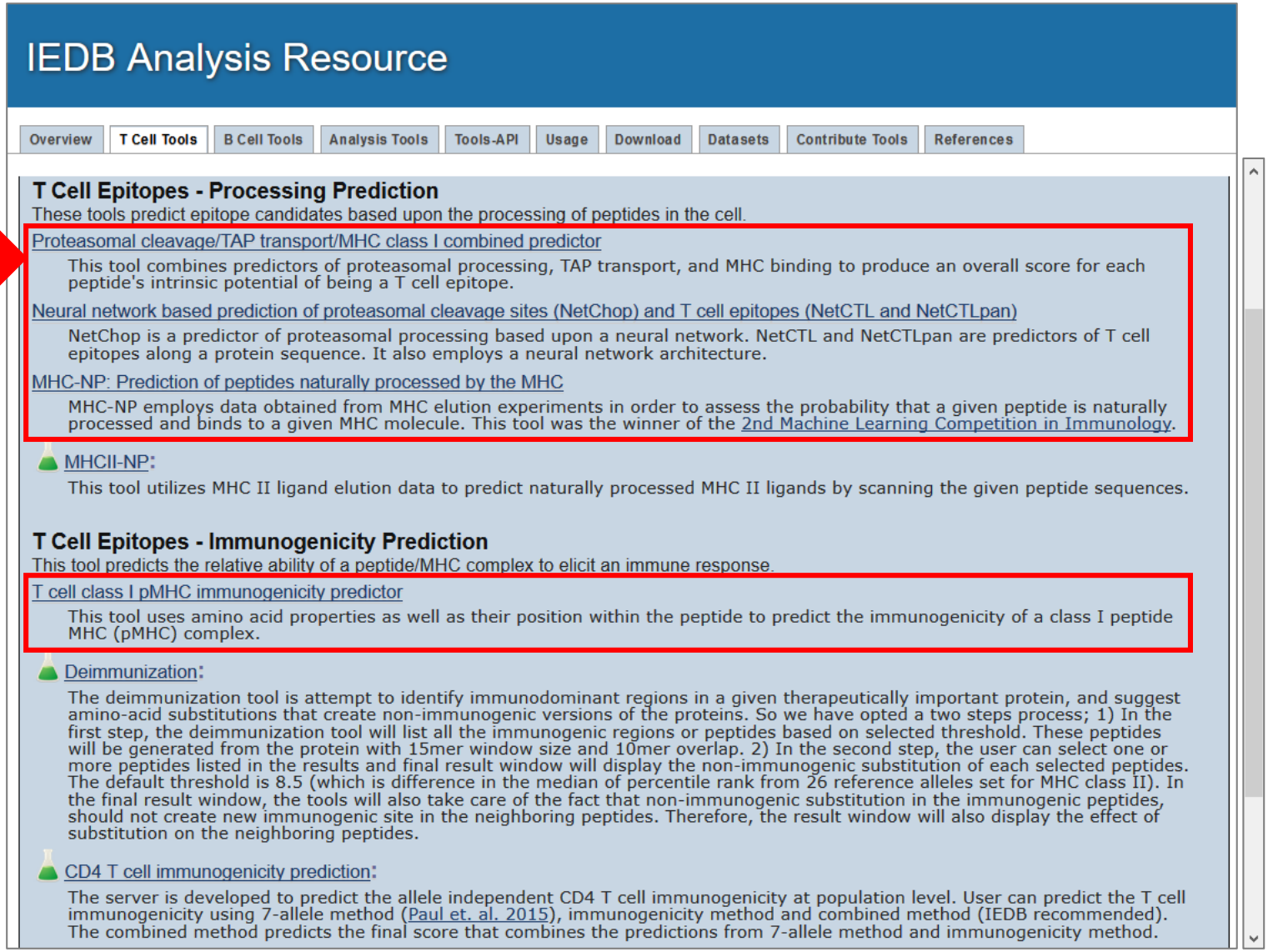
- Higher scores = higher efficiency for MHC-I presentation
- MHC binding score = –log10(IC50) (sign change)
- Combined scores are additive
  - Processing = proteasome + TAP
  - Total = proteasome + TAP + MHC
- Different variance in scores reflects different selectivity

Caveats / performance of processing predictions

- Processing predictions beat MHC binding predictions when predicting **eluted peptides**
- No clear evidence that processing predictions are better at predicting **epitopes**
- Eluted peptides may over represent ‘best possible’ ligands, and the difference in processing may not be relevant in practice
- Co-evolution of MHC molecules to bind peptides with motifs that are generated by proteasome and TAP means that most high affinity MHC binders are also efficiently processed
- Recommendation: Use MHC binding predictions alone by default
  - If resources require limiting the number of peptides considered, use total score of processing predictions as additional filter

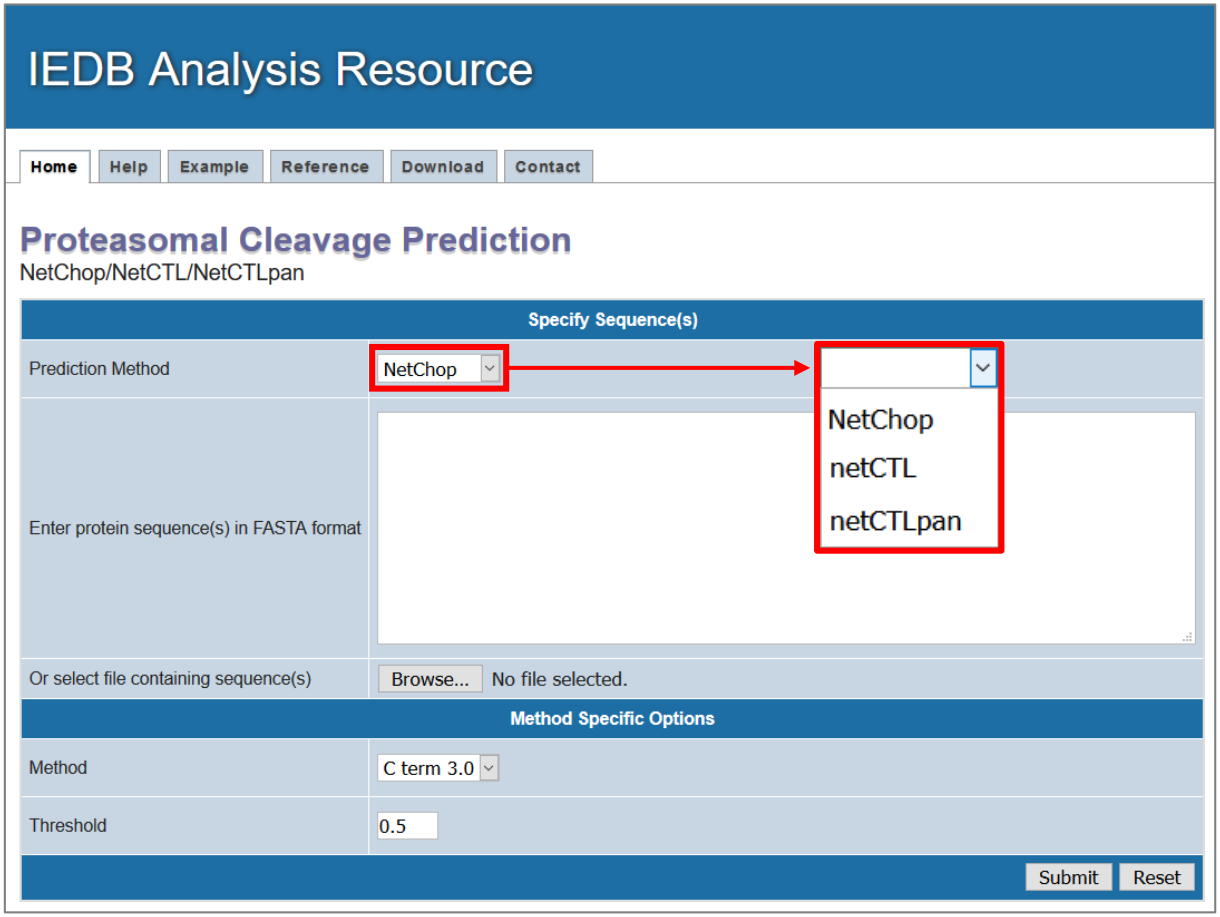
Class I Processing + immunogenicity tools available in the IEDB

http://tools.iedb.org/main/tcell/



Neural Network based predictors

http://tools.iedb.org/netchop/

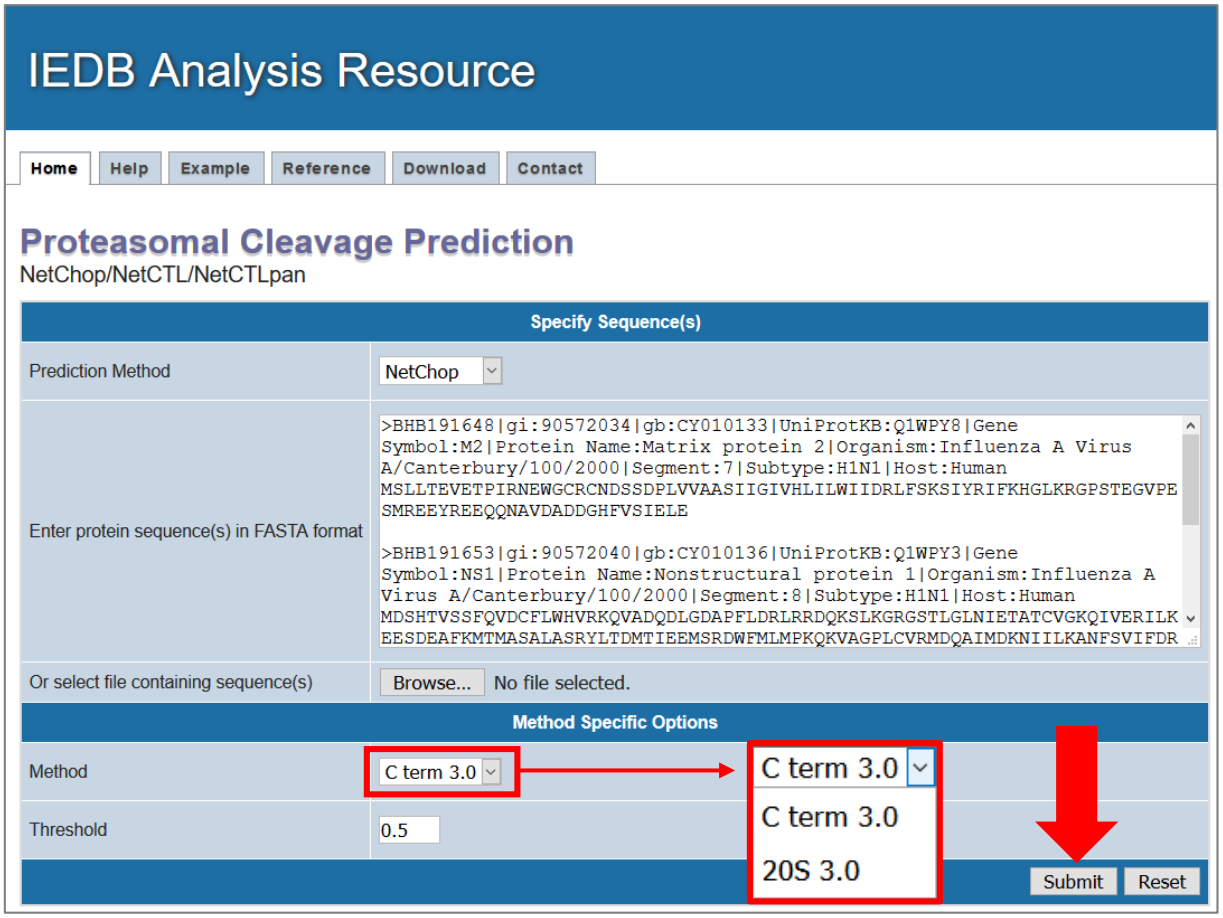


- **NetChop**: proteasomal cleavage
- **NetCTL:** combines NetChop, TAP transport, NetMHC
- **NetCTLpan**: combines NetChop, TAP transport, NetMHCpan

Key difference is the use of NetChop

NetChop -example

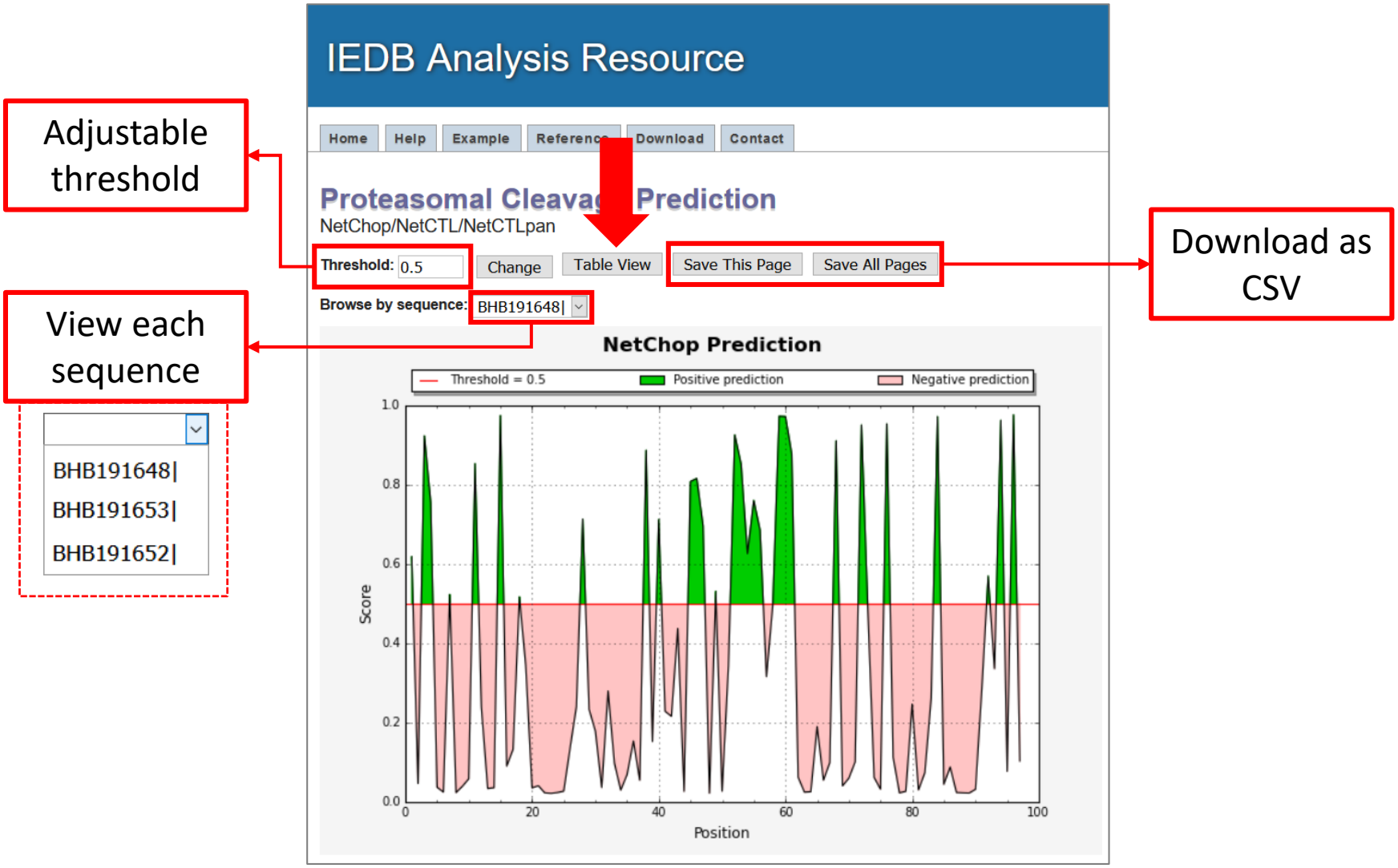
http://tools.iedb.org/netchop/



- Predicts C-terminal cleavage based on two approaches
  - **C-term 3.0**: C-terminal residues found for MHC ligands
  - **20S 3.0**: Cleavage sites based on in vitro protein digests
- C-term 3.0 is not truly a proteasome predictor but performs better
- NetCTL and NetCTLPan use C-term 3.0 by default

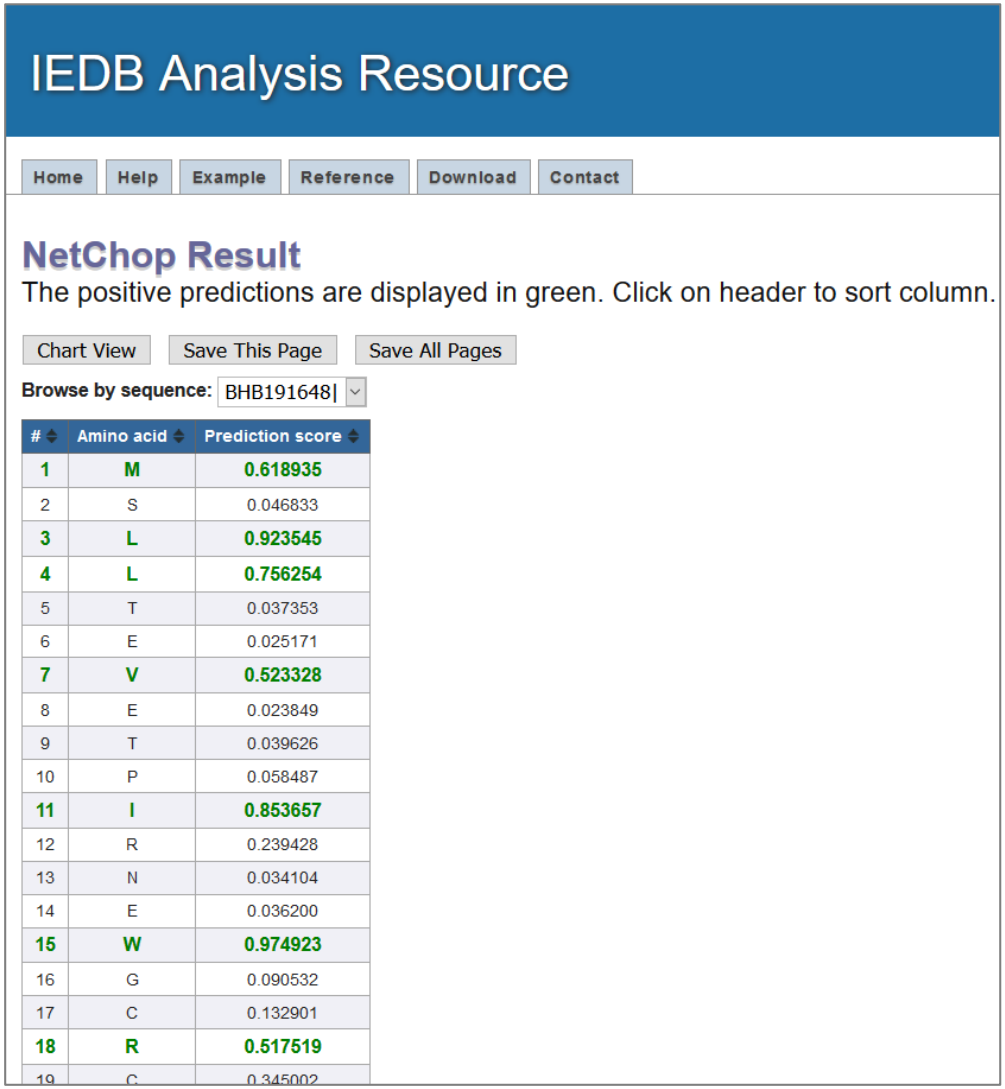
NetChop -example

http://tools.iedb.org/netchop/



NetChop -example

http://tools.iedb.org/netchop/



IEDB Analysis Resource

Home | Help | Example | Reference | Download | Contact

NetChop Result

The positive predictions are displayed in green. Click on header to sort column.

Chart View | Save This Page | Save All Pages

Browse by sequence: BHB191648|

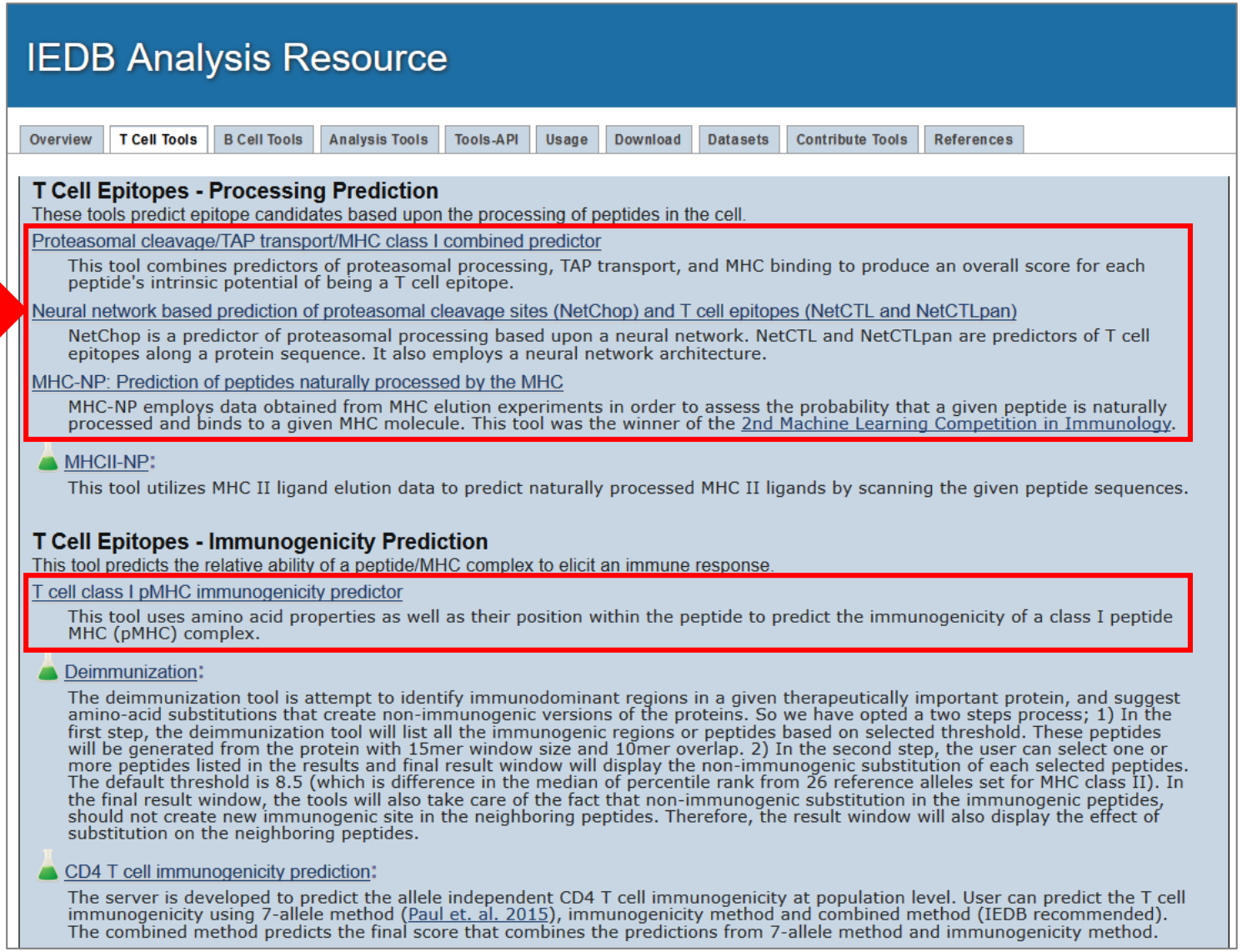
#	Amino acid	Prediction score
1	**M**	**0.618935**
2	S	0.046833
3	**L**	**0.923545**
4	**L**	**0.756254**
5	T	0.037353
6	E	0.025171
7	**V**	**0.523328**
8	E	0.023849
9	T	0.039626
10	P	0.058487
11	**I**	**0.853657**
12	R	0.239428
13	N	0.034104
14	E	0.036200
15	**W**	**0.974923**
16	G	0.090532
17	C	0.132901
18	**R**	**0.517519**
19	C	0.345002

References

- Peters et al, JMB 2002 (proteasome)
- Peters et al, J Immunol 2003 (TAP)
- Tenzer et al, CMLS, 2005 (combined)
- Nielsen, Immunogenetics, 2005 (NetChop)
- Larsen, BMC Bioinformatics, 2007 (NetCTL)
- Stranzl, Immunogenetics, 2010 (NetCTLPan)

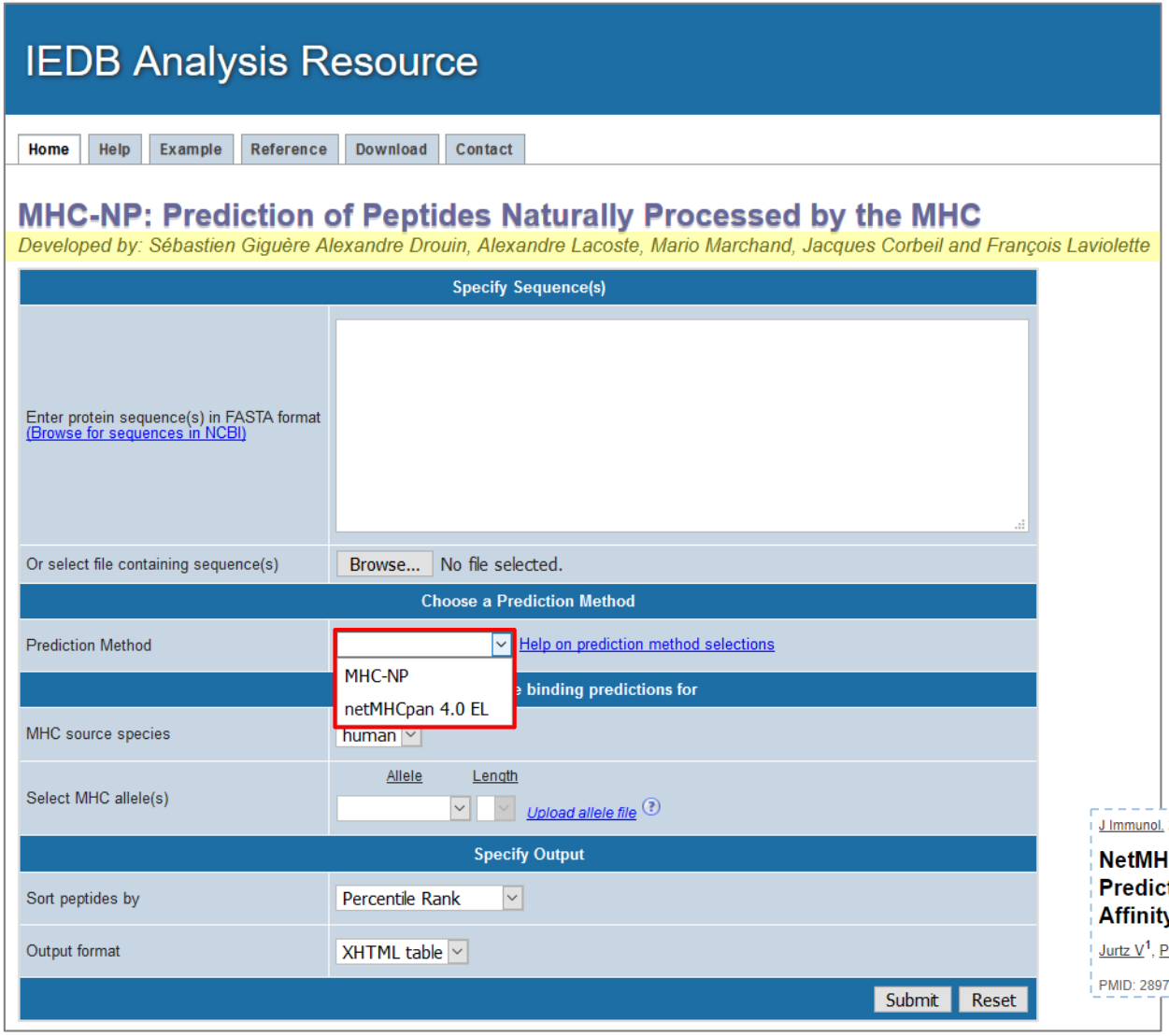
Class I Processing + immunogenicity tools available in the IEDB

http://tools.iedb.org/main/tcell/



MHC-NP: Prediction of peptides naturally processed by the MHC

http://tools.iedb.org/mhcnp/



Pan-predictions trained on both binding + eluted ligand data now available!

J Immunol. 2017 Nov 1;199(9):3360-3368. doi: 10.4049/jimmunol.1700893. Epub 2017 Oct 4.

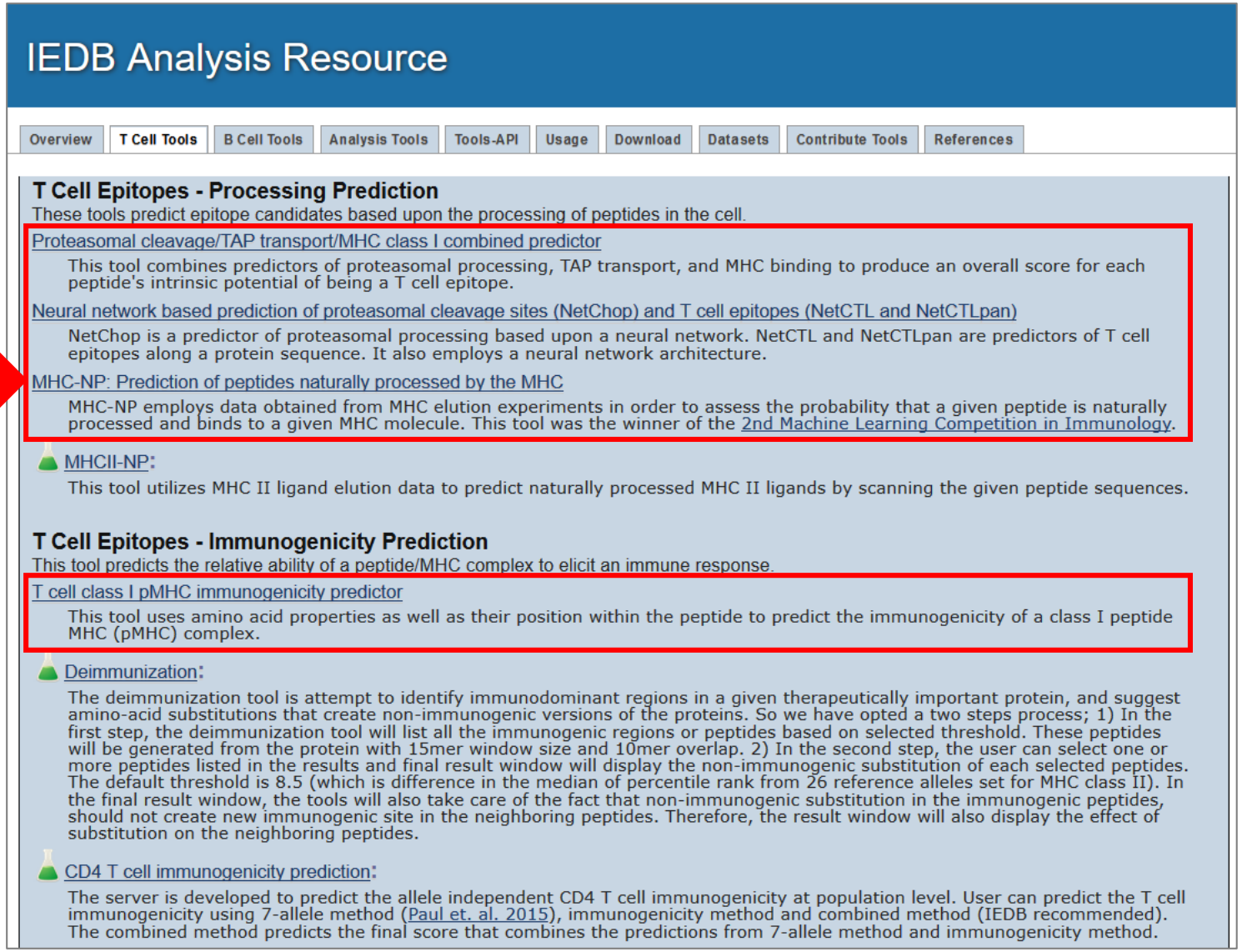
NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data.

Jurtz V[1], Paul S[2], Andreatta M[3], Marcatili P[1], Peters B[2], Nielsen M[4,3].

PMID: 28978689 PMCID: PMC5679736 DOI: 10.4049/jimmunol.1700893

Class I Processing + immunogenicity tools available in the IEDB

http://tools.iedb.org/main/tcell/

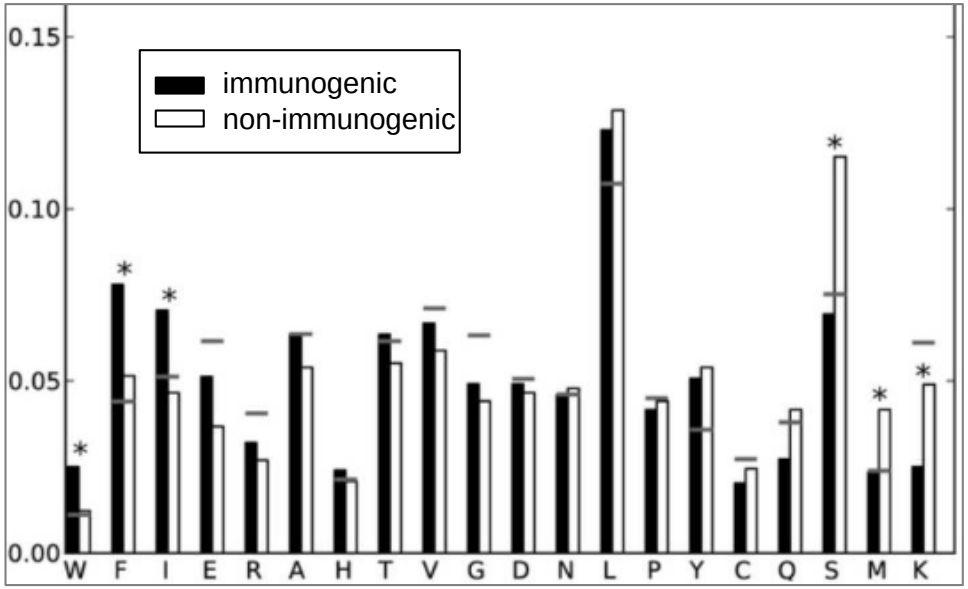


Class I immunogenicity prediction

- Approach: Assemble two datasets of peptides with similar MHC binding affinity, that are (i) recognized or (ii) not recognized by T cells

- Enrichment of W,F,I and depletion of S,M,K in immunogenic peptides

- Use enrichments to calculate propensity scores



PLoS Comput Biol. 2013 Oct;9(10):e1003266. doi: 10.1371/journal.pcbi.1003266. Epub 2013 Oct 24.

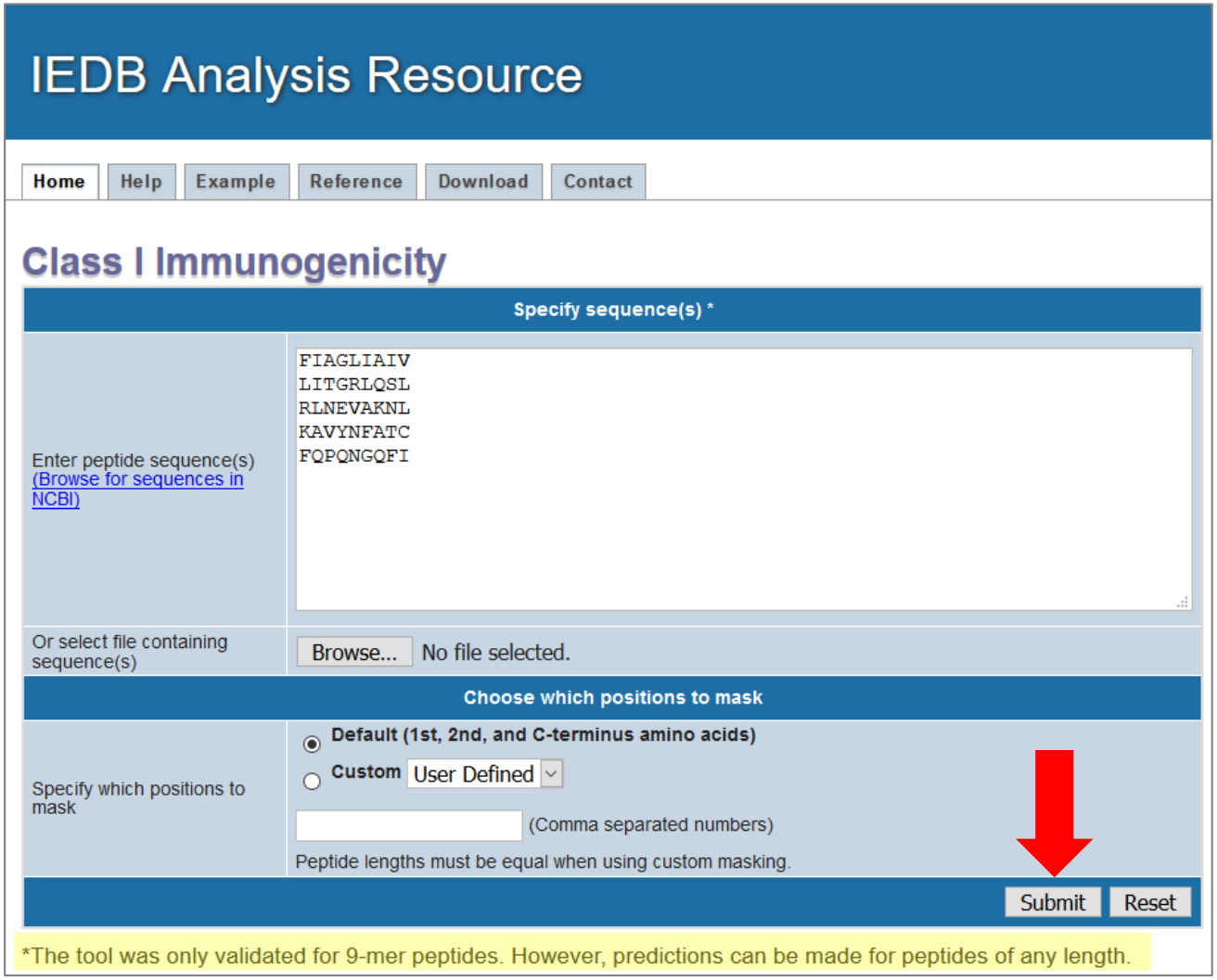
Properties of MHC class I presented peptides that enhance immunogenicity.

Calis JJ[1], Maybeno M, Greenbaum JA, Weiskopf D, De Silva AD, Sette A, Keşmir C, Peters B.

PMID: 24204222 PMCID: PMC3808449 DOI: 10.1371/journal.pcbi.1003266

Class I immunogenicity prediction -example

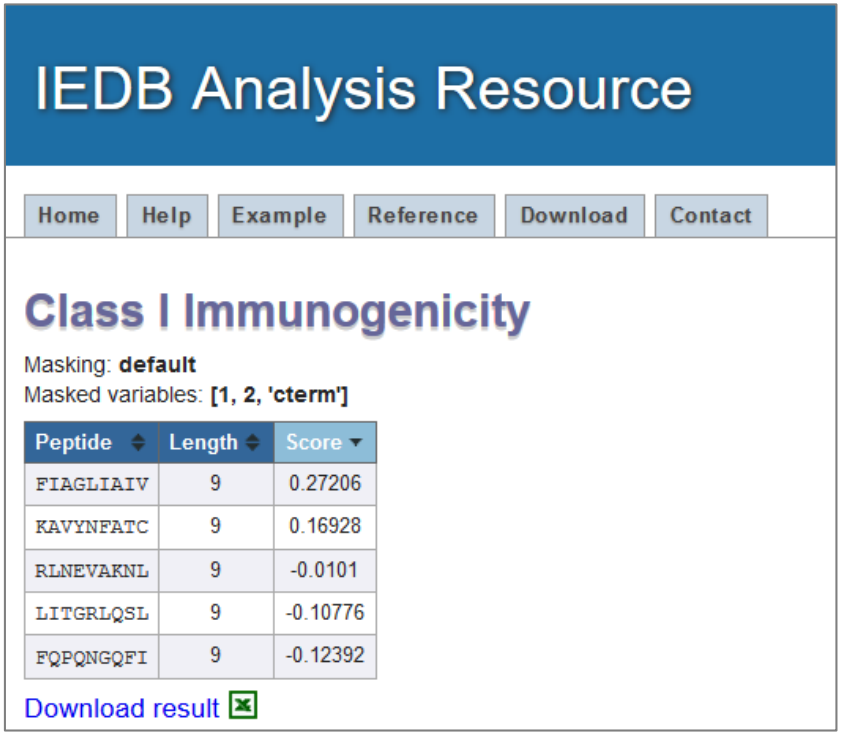
http://tools.iedb.org/immunogenicity/



Mask positions that are MHC anchors

Class I immunogenicity prediction -example

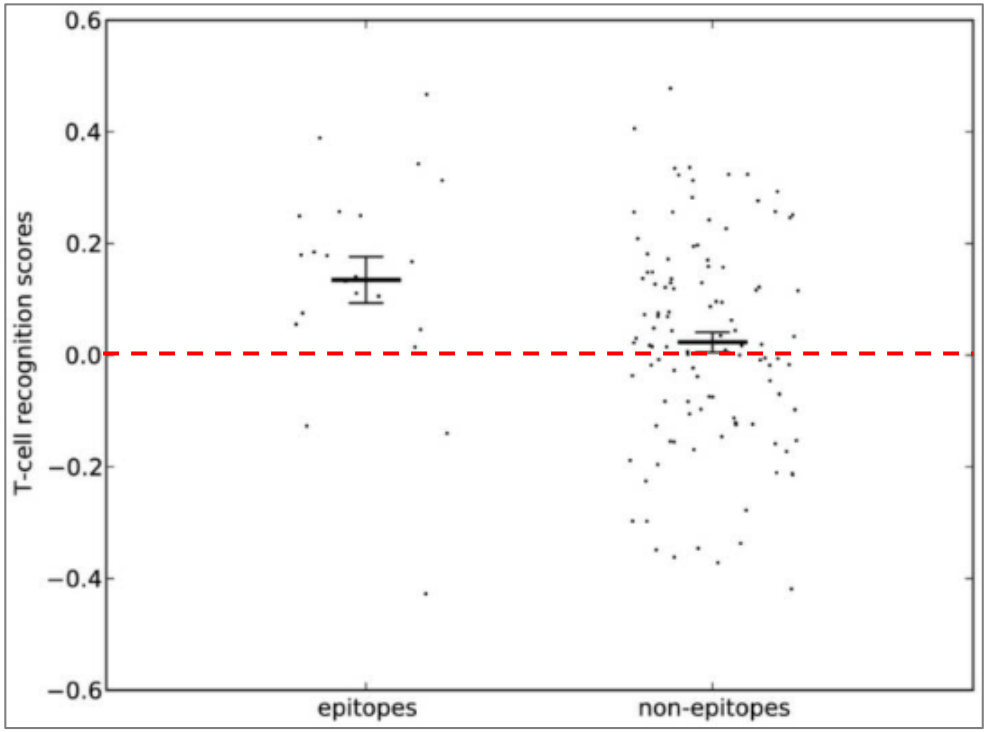
http://tools.iedb.org/immunogenicity/



- Scores are sums of propensity scores at all unmasked predictions
- High scores = peptide is more likely to be immunogenic

Class I immunogenicity prediction caveats / performance

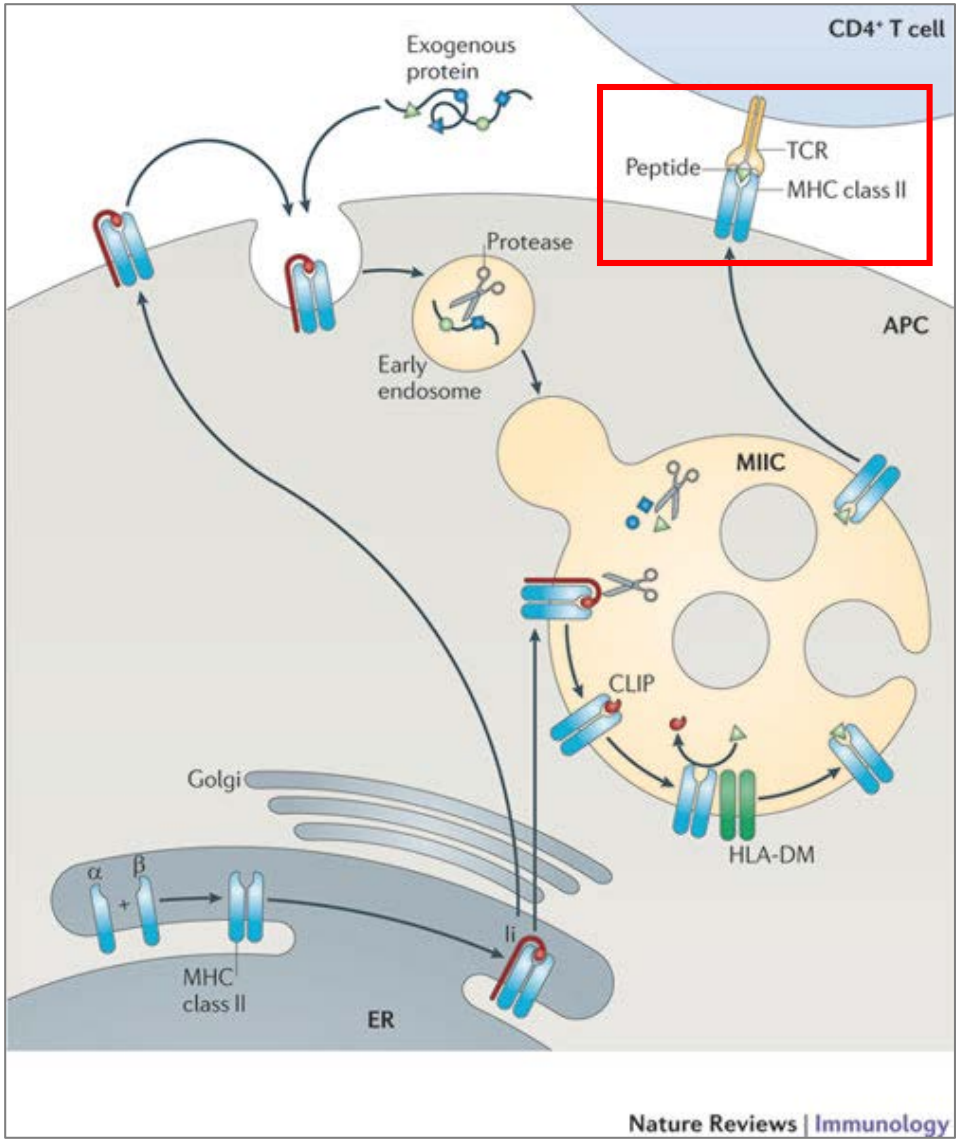
- Experimentally, many MHC binding peptides can be immunogenic (~50%)
- Cross validation gave AUC values ~ 0.65. Test on independent blind set gave AUC = 0.69
- Recommendation: Use as filter (cutoff 0) if high specificity is desired. Suggested cutoff is 0



Class I Summary

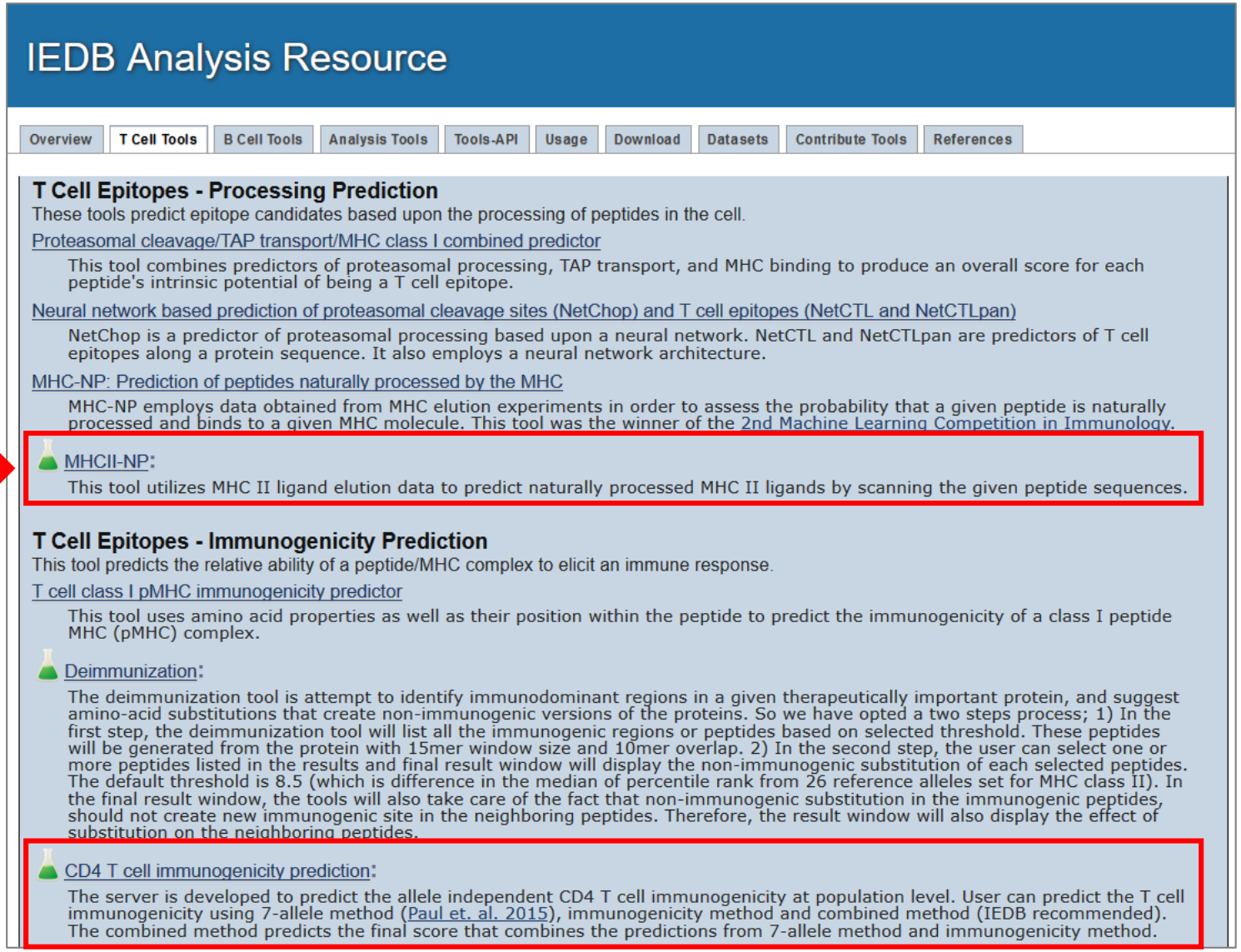
- Processing predictions are better at identifying naturally processed ligands, but have not been shown to be superior in identifying epitopes compared to MHC binding predictions
- Specific processing and immunogenicity predictions are good additional filters if the only goal is to select high likelihood T cell epitopes
- NetMHCPan 4.0 EL scores, which are trained on both MHC binding and ligand elution data are a straightforward replacement of MHC binding predictions, and show some enhanced performance
  → Use these, and consider pairing with immunogenicity scores, when predicting epitope candidates

CD4 T cell epitopes (MHC class II)



Class II Processing + immunogenicity tools available in the IEDB

http://tools.iedb.org/main/tcell/

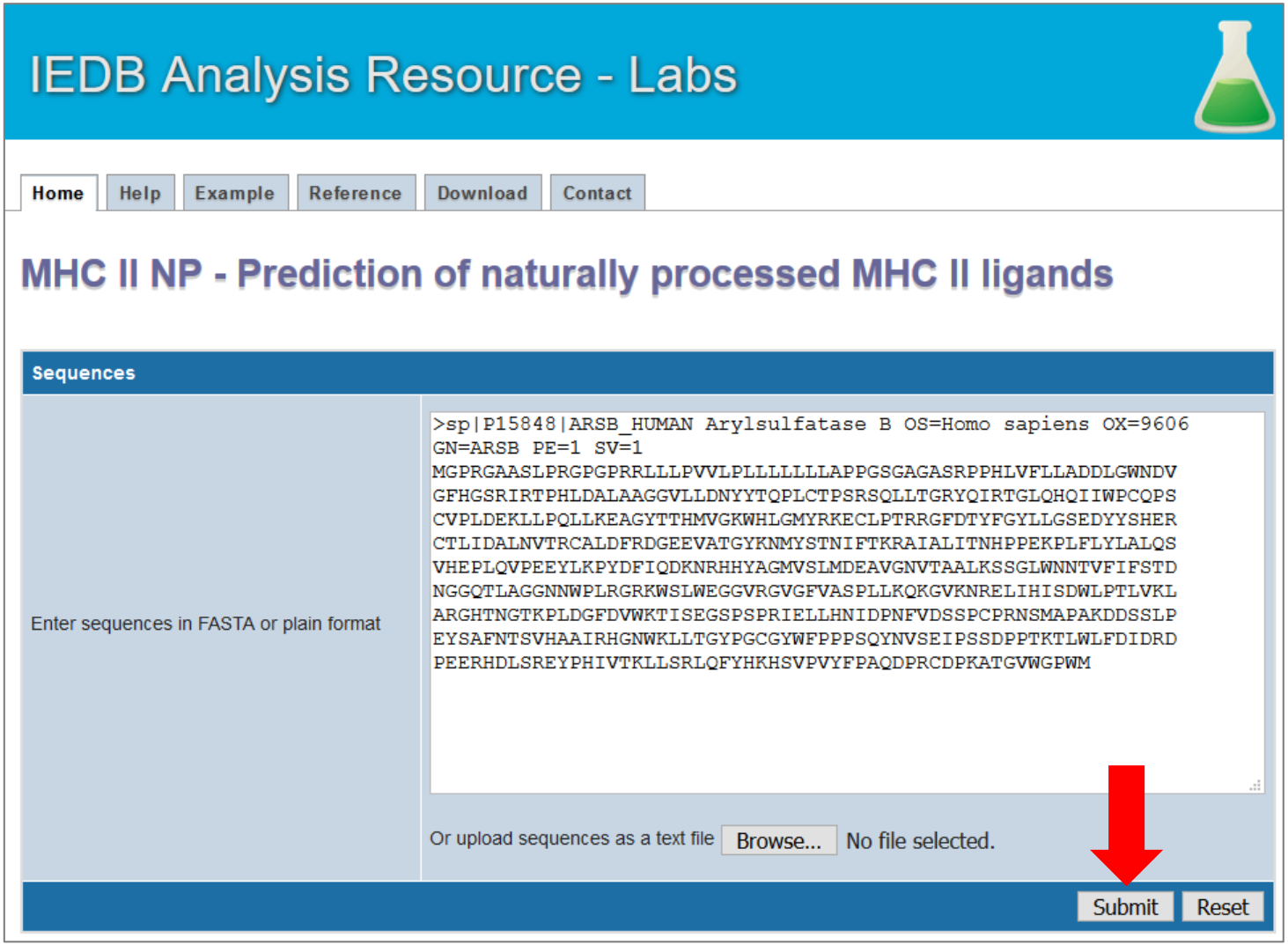


MHCII-NP

- Predicting the naturally processed peptides for MHC class II
- Based on
  - Cleavage motif analysis at C and N terminal of peptides
  - Ligand elution data derived from IEDB
- Ligand predictions is improved markedly when combining the binding and cleavage motifs
- T cell epitope prediction is not significantly improved

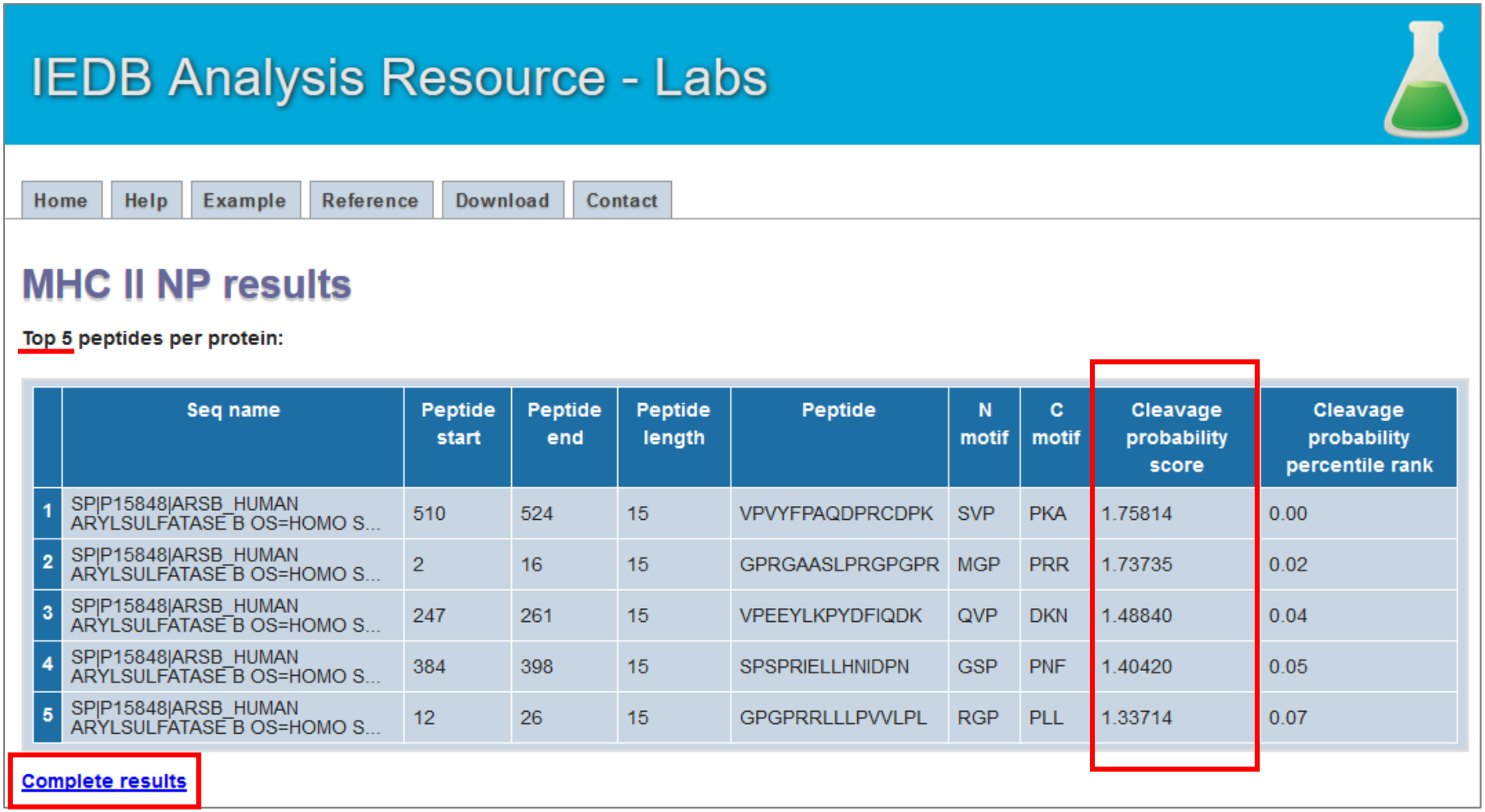
MHCII-NP -example

http://tools.iedb.org/mhciinp/



MHCII-NP -example

http://tools.iedb.org/mhciinp/



IEDB Analysis Resource - Labs

Home | Help | Example | Reference | Download | Contact

MHC II NP results

Top 5 peptides per protein:

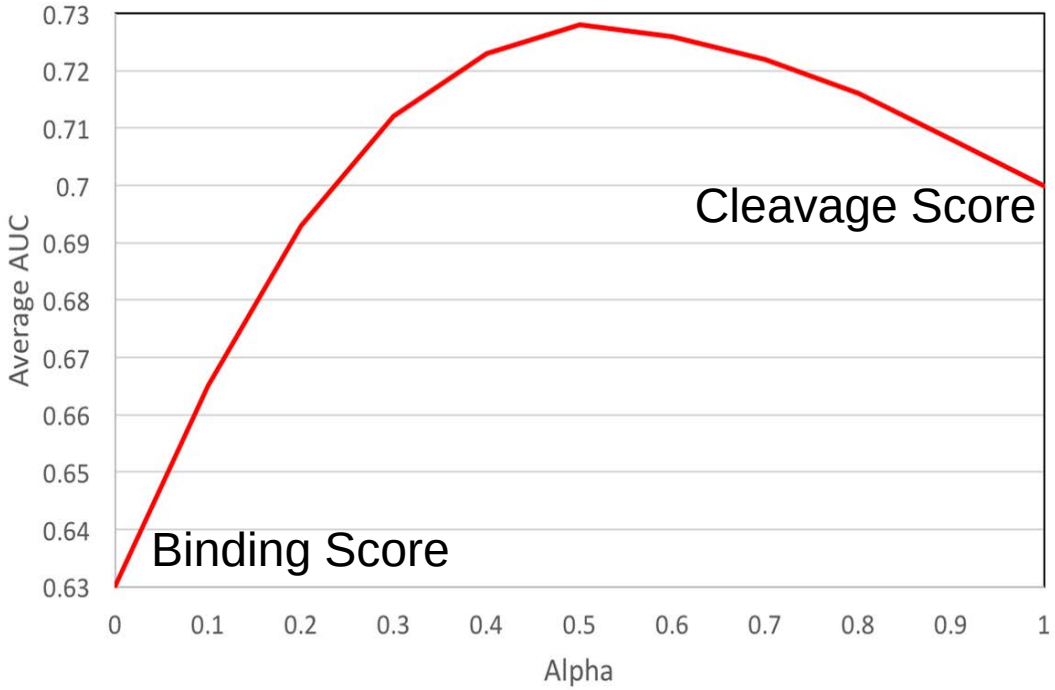
	Seq name	Peptide start	Peptide end	Peptide length	Peptide	N motif	C motif	Cleavage probability score	Cleavage probability percentile rank
1	SP|P15848|ARSB_HUMAN ARYLSULFATASE B OS=HOMO S...	510	524	15	VPVYFPAQDPRCDPK	SVP	PKA	1.75814	0.00
2	SP|P15848|ARSB_HUMAN ARYLSULFATASE B OS=HOMO S...	2	16	15	GPRGAASLPRGPGPR	MGP	PRR	1.73735	0.02
3	SP|P15848|ARSB_HUMAN ARYLSULFATASE B OS=HOMO S...	247	261	15	VPEEYLKPYDFIQDK	QVP	DKN	1.48840	0.04
4	SP|P15848|ARSB_HUMAN ARYLSULFATASE B OS=HOMO S...	384	398	15	SPSPRIELLHNIDPN	GSP	PNF	1.40420	0.05
5	SP|P15848|ARSB_HUMAN ARYLSULFATASE B OS=HOMO S...	12	26	15	GPGPRRLLLPVVLPL	RGP	PLL	1.33714	0.07

Complete results

MHCII-NP scores

- Cleavage Score: Derived from the cleavage motif analysis in ligand elution data
- Binding Score: Derived from HLA binding affinity using 7-allele method (Paul et. al. 2015).

$$Combined\ score = \alpha \times cleavage\ \ probability\ score + (1 - \alpha) \times binding\ \ score$$



Barra et al. Genome Medicine (2018) 10:84
https://doi.org/10.1186/s13073-018-0594-6

Genome Medicine

RESEARCH Open Access

Footprints of antigen processing boost MHC class II natural ligand predictions

Carolina Barra[1†], Bruno Alvarez[1†], Sinu Paul[2], Alessandro Sette[2], Bjoern Peters[2], Massimo Andreatta[1], Søren Buus[3] and Morten Nielsen[1,4*]

Front Immunol. 2018 Aug 6;9:1795. doi: 10.3389/fimmu.2018.01795. eCollection 2018.

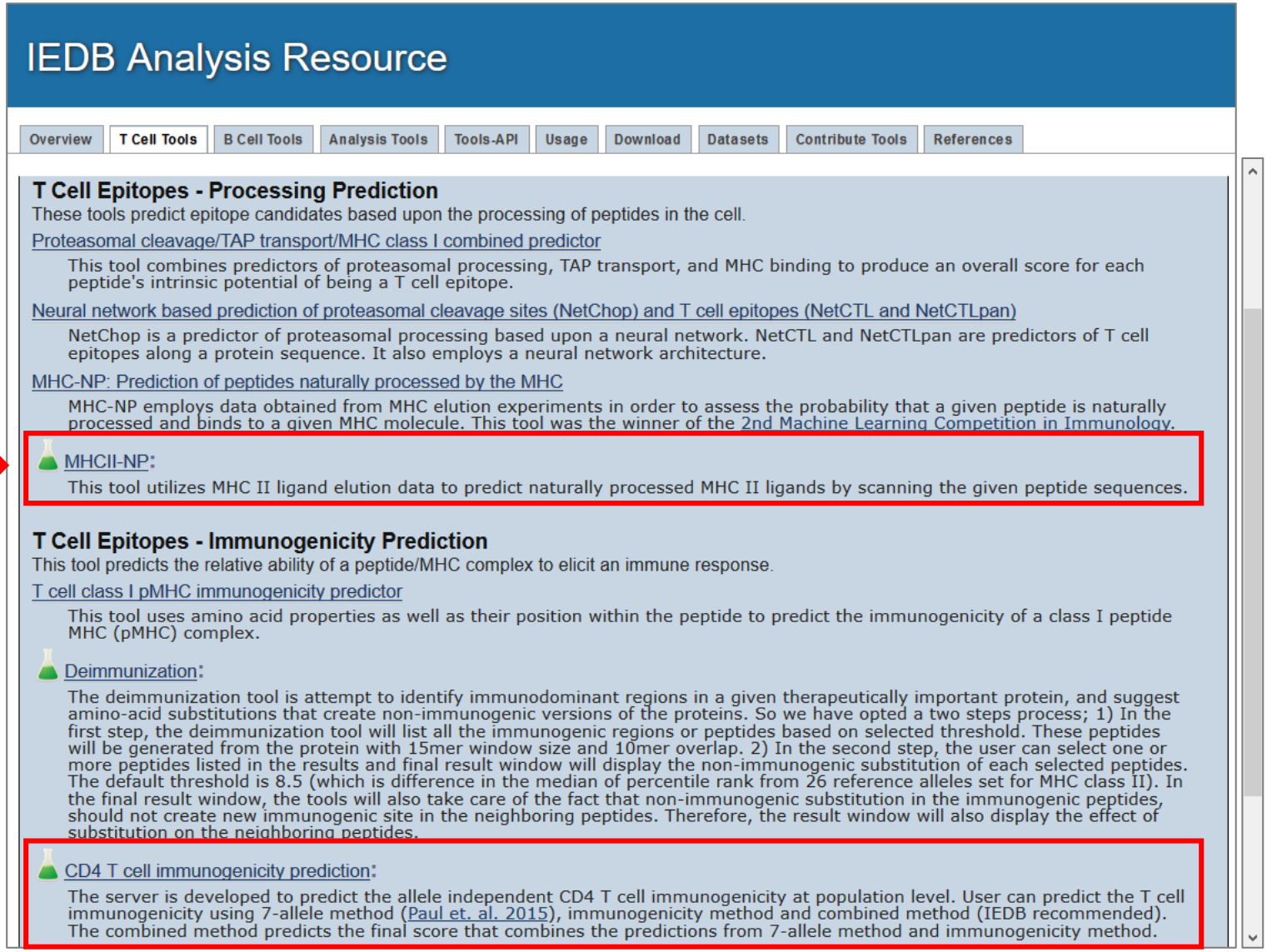
Determination of a Predictive Cleavage Motif for Eluted Major Histocompatibility Complex Class II Ligands.

Paul S[1], Karosiene E[1], Dhanda SK[1], Jurtz V[2], Edwards L[1], Nielsen M[2,3], Sette A[1,4], Peters B[1,4].

PMID: 30127785 PMCID: PMC6087742 DOI: 10.3389/fimmu.2018.01795

Class II Processing + immunogenicity tools available in the IEDB

http://tools.iedb.org/main/tcell/



MHC-II restricted immunogenicity prediction

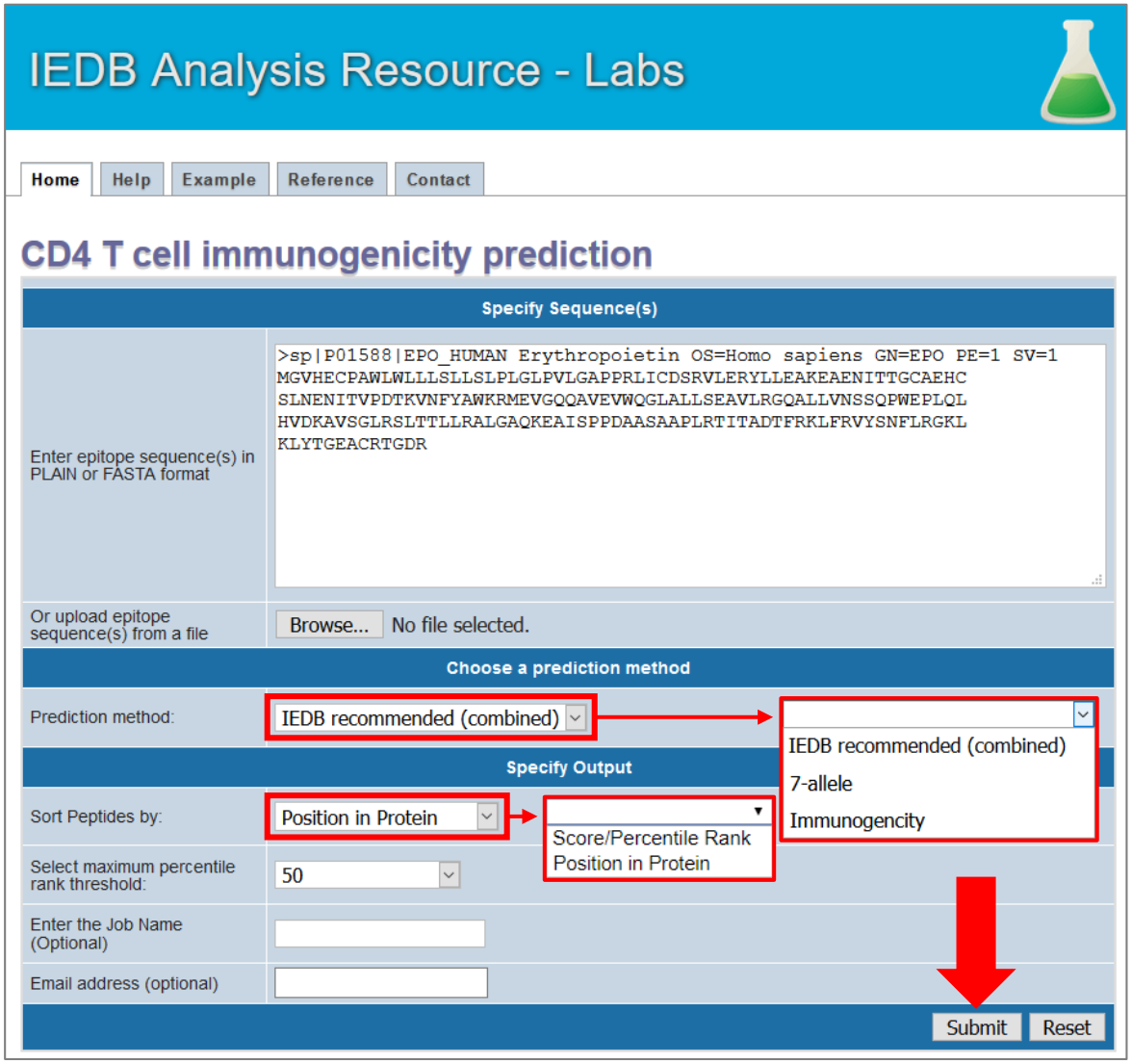
- Extracted datasets of proteins from the IEDB for which overlapping peptides were tested for immunogenicity
- Utilized these datasets to train a Neural Network to learn ‘motifs’ associated with immunogenicity independent of specific MHC alleles expressed
- Resulting score can be combined with ‘7 allele method’ quantifying MHC binding across alleles to predict overall immunogenicity

Class II immunogenicity prediction

- Based on Neural network model trained on
  - In house dataset for different antigens tested on different population cohorts
  - Tetramer dataset- derived from IEDB
- Validated on 57 independent studies from different groups across the world
- Implemented three approaches
  - 7-allele method (*Paul et. al. 2015*)
  - Immunogenicity predictions
  - Hybrid approach

Class II immunogenicity prediction -example

http://tools.iedb.org/CD4episcore/



Class II immunogenicity prediction -example

http://tools.iedb.org/CD4episcore/

IEDB Analysis Resource - Labs

Home | Help | Example | Reference | Contact

CD4 Immunogenicity prediction results

Number of proteins: 1

Number of 15mer (overlapping 10mer): 37

Threshold : 50.0%

Method : combined

Download result

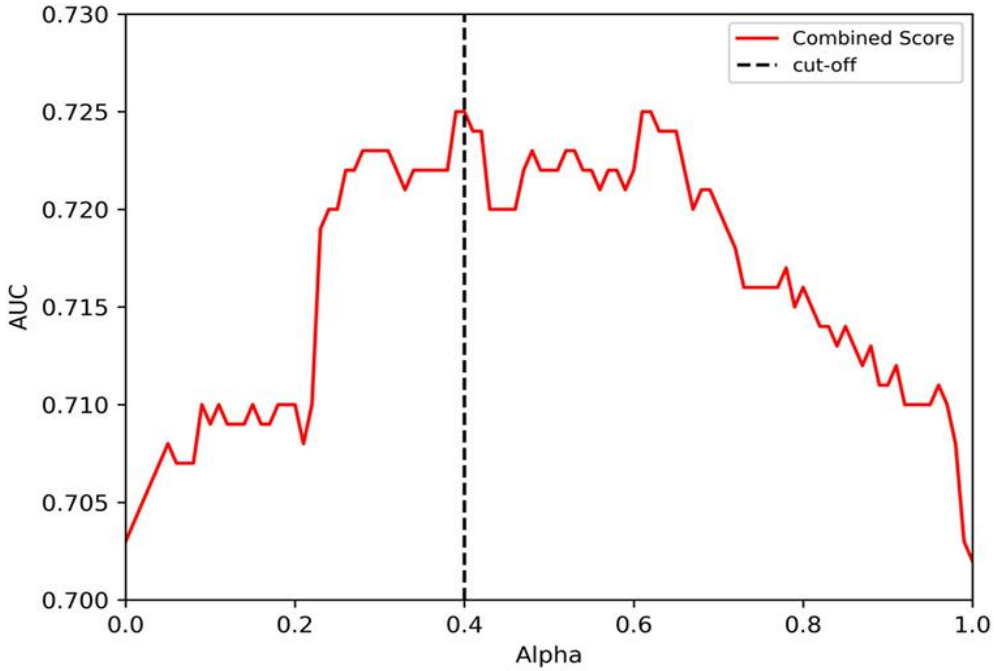
Citations

Protein Number	Protein Description	Peptide	Start	End	Combined Score	Immunogenicity Score	Peptide core	Median Percentile Rank (7-allele)	HLA-DRB1:03:01	HLA-DRB1:07:01	HLA-DRB1:15:01	HLA-DRB3:01:01	HLA-DRB3:02:02	HLA-DRB4:01:01	HLA-DRB5:01:01
1	sp|P01588|EPO_HUMAN Erythropoietin OS=Homo sapiens GN=EPO PE=1 SV=1	WLLLSLLSLPLGLPV	11	25	42.16452	95.0613	LLSLLSLPL	6.9	25.0	3.2	3.6	73.0	33.0	6.9	6.5
1	sp|P01588|EPO_HUMAN Erythropoietin OS=Homo sapiens GN=EPO PE=1 SV=1	TKVNFYAWKRMEVGQ	71	85	47.39488	67.4872	TKVNFYAWK	34.0	52.0	22.0	15.0	71.0	30.0	65.0	34.0
1	sp|P01588|EPO_HUMAN Erythropoietin OS=Homo sapiens GN=EPO PE=1 SV=1	EPLQLHVDKAVSGLR	116	130	32.55636	43.8909	LHVDKAVSG	25.0	5.4	59.0	40.0	22.0	7.0	38.0	25.0
1	sp|P01588|EPO_HUMAN Erythropoietin OS=Homo sapiens GN=EPO PE=1 SV=1	VSGLRSLTTLLRALG	126	140	44.95964	86.8991	LTTLLRALG	17.0	12.0	17.0	9.3	70.0	20.0	20.0	1.3
1	sp|P01588|EPO_HUMAN Erythropoietin OS=Homo sapiens GN=EPO PE=1 SV=1	SLTTLLRALGAQKEA	131	145	42.78744	69.4686	LLRALGAQK	25.0	47.0	46.0	21.0	89.0	25.0	14.0	1.6
1	sp|P01588|EPO_HUMAN Erythropoietin OS=Homo sapiens GN=EPO PE=1 SV=1	PLRTITADTFRKLFR	156	170	46.18064	85.4516	LRTITADTF	20.0	6.1	47.0	40.0	8.0	20.0	20.0	20.0
1	sp|P01588|EPO_HUMAN Erythropoietin OS=Homo sapiens GN=EPO PE=1 SV=1	TADTFRKLFRVYSNF	161	175	46.66984	44.6746	FRKLFRVYS	48.0	63.0	58.0	23.0	48.0	33.0	53.0	24.0
1	sp|P01588|EPO_HUMAN Erythropoietin OS=Homo sapiens GN=EPO PE=1 SV=1	RKLFRVYSNFLRGKL	166	180	13.8966	22.4415	FRVYSNFLR	8.2	53.0	8.2	0.12	27.0	4.8	26.0	1.5

Class II immunogenicity prediction scores

- Immunogenicity Score: Derived from the neural network model trained on Immunogenicity data
- HLA Score: Derived from HLA binding affinity using 7-allele method (Paul et. al. 2015).

$$\text{Combined score} = \alpha \times \text{Imm score} + (1 - \alpha) \times \text{HLA score}.$$



Front Immunol. 2018 Jun 14;9:1369. doi: 10.3389/fimmu.2018.01369. eCollection 2018.

Predicting HLA CD4 Immunogenicity in Human Populations.

Dhanda SK[1], Karosiene E[1], Edwards L[1], Grifoni A[1], Paul S[1], Andreatta M[2], Weiskopf D[1], Sidney J[1], Nielsen M[2,3], Peters B[1,4], Sette A[1,4].

PMID: 29963059 PMCID: PMC6010533 DOI: 10.3389/fimmu.2018.01369

Class II Summary

- Similar to MHC class I, enhancement of epitope prediction efficacy is minor compared to using MHC binding predictions alone
- Prediction of naturally eluted ligands is greatly improved with processing predictions
- As of now, recommendation is to stick to allele specific MHC binding predictions (NetMHCPanII), or the 7-allele method for broad populations

→ Both class I and II epitope predictions are constantly being re-evaluated, and these recommendations are subject to change