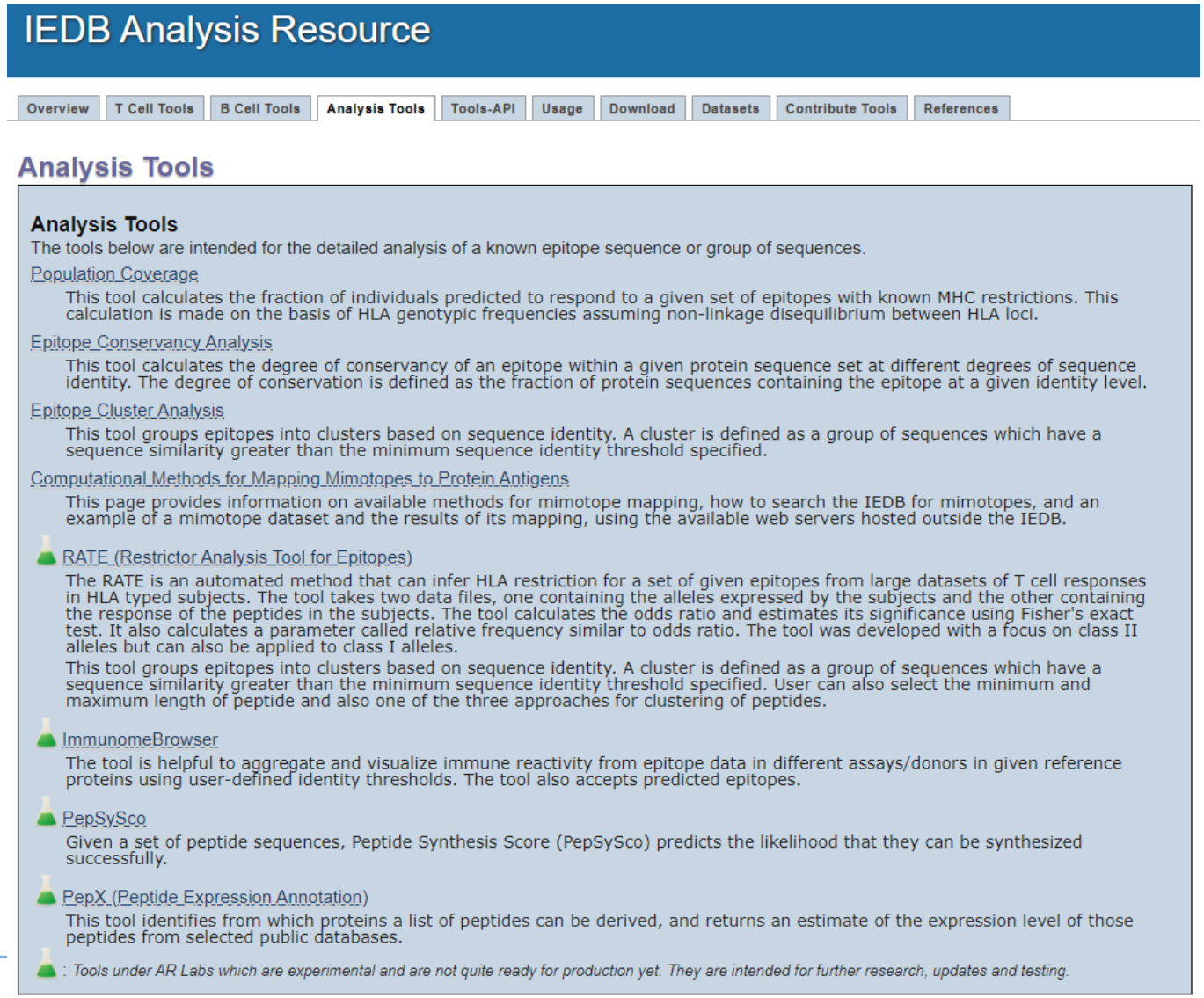
Analysis Tools

tools.iedb.org

Presented by: Alessandro Sette, IEDB Principal Investigator

Analysis tools with broad applications

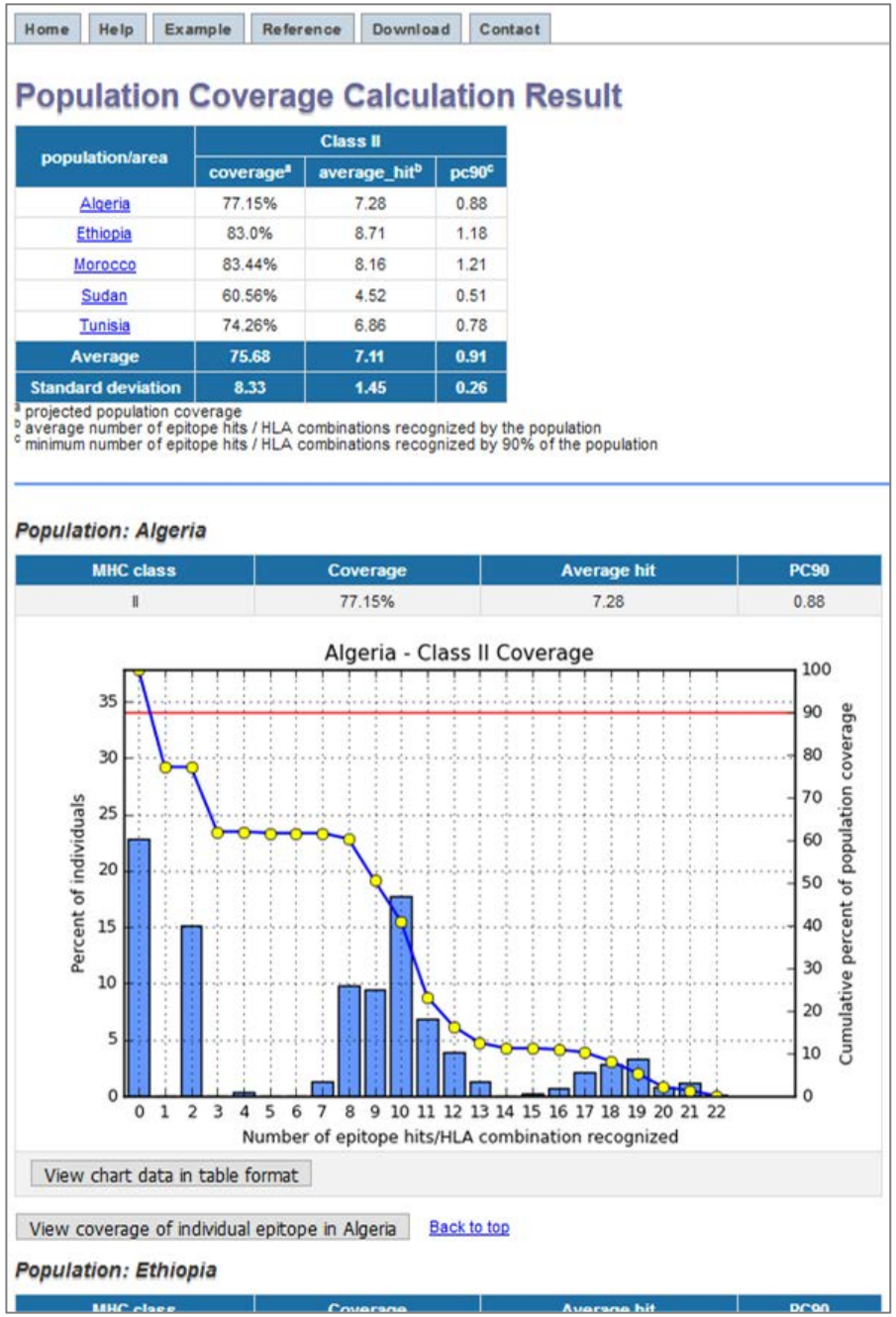
http://tools.iedb.org/main/analysis-tools/



Population Coverage

http://tools.iedb.org/population/

- Calculates the fraction of individuals projected to bind and/or respond to a given set of epitopes with defined reactivity
- Based on
  - Epitope known HLA binding/restrictions
  - HLA genotypic frequencies
- HLA genotypic frequencies vary in different ethnicities
  - http://allelefrequencies.net
  - Focused more on worldwide migration patterns/based on ancestral studies



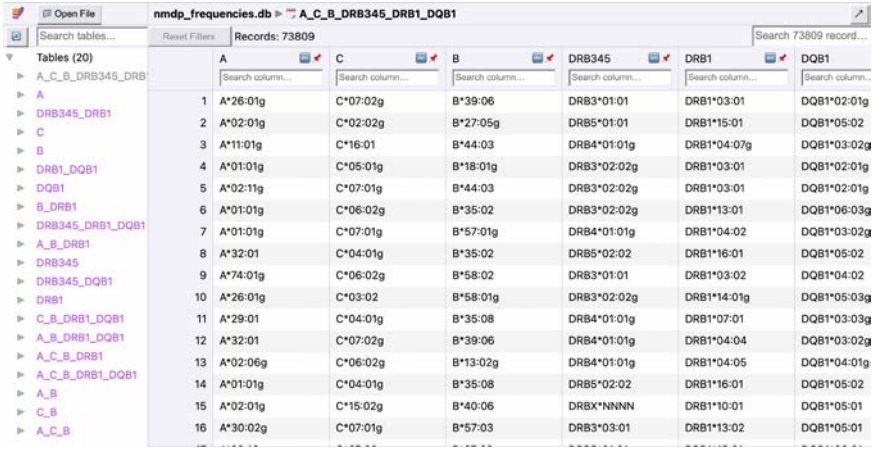
Updated Population Coverage Tool

New Data Source (In Progress)

- National Marrow Donor Program (NMDP) Database

BE THE MATCH®

High Resolution HLA Alleles and Haplotypes in the US Population



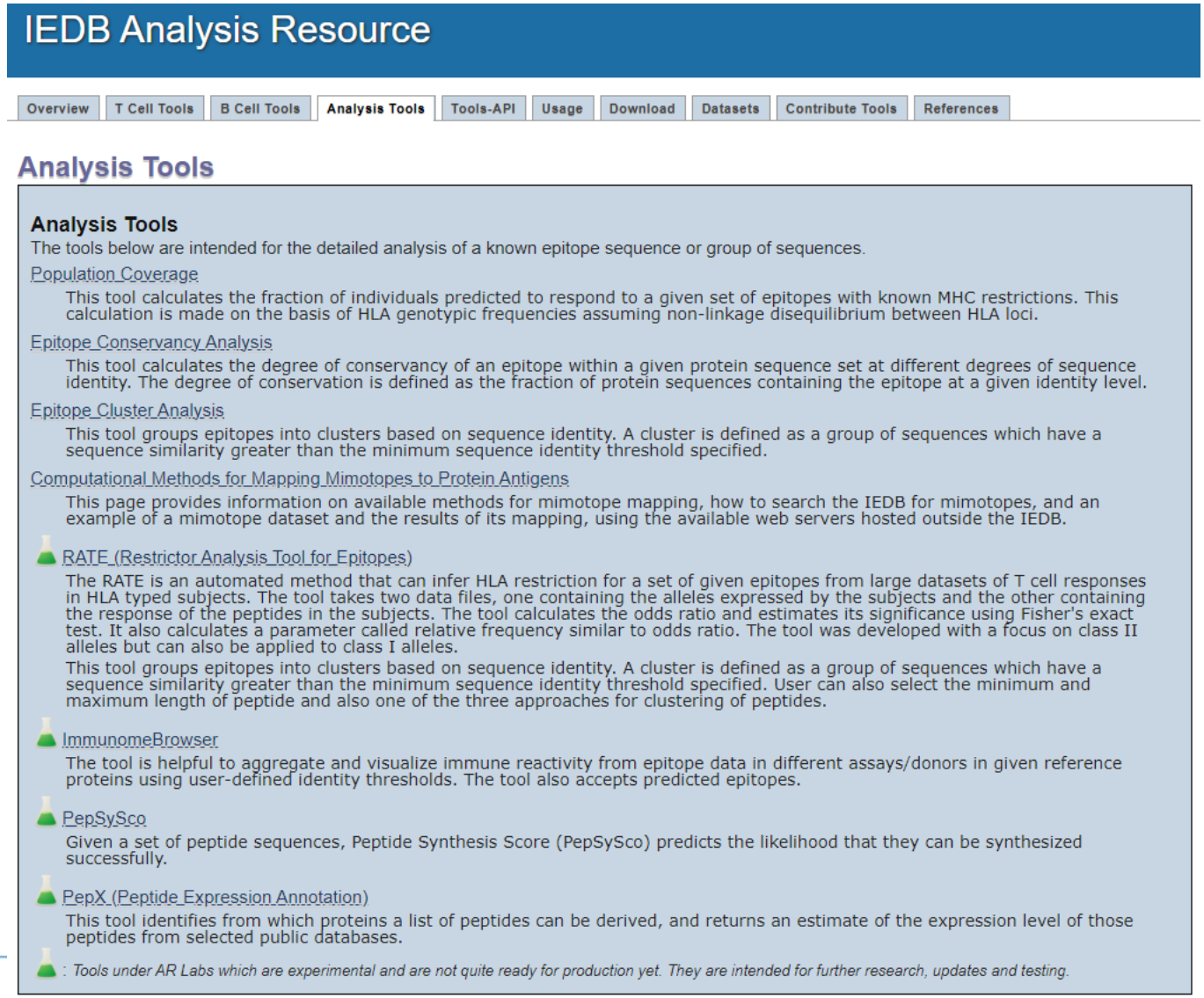
Gragert, L., Madbouly, A., Freeman, J., & Maiers, M. (2013). Six-locus high resolution HLA haplotype frequencies derived from mixed-resolution DNA typing for the entire US donor registry. Human Immunology, 74(10), 1313–1320. http://dx.doi.org/10.1016/j.humimm.2013.06.025.

Race Code	Detailed Race/Ethnic Description	Broad Race Group
AAFA	African American	AFA
AFB	African	AFA
AINDI	South Asian Indian	API
AISC	American Indian - South or Central Am.	NAM
ALANAM	Alaska Native or Aleut	NAM
AMIND	North American Indian	NAM
CARB	Caribbean Black	AFA
CARHIS	Caribbean Hispanic	HIS
CARIBI	Caribbean Indian	NAM
EURCAU	European Caucasian	CAU
FILII	Filipino	API
HAWI	Hawaiian or other Pacific Islander	API
JAPI	Japanese	API
KORI	Korean	API
MENAFC	Middle Eastern or N. Coast of Africa	CAU
MSWHIS	Mexican or Chicano	HIS
NCHI	Chinese	API
SCAHIS	Hispanic – South or Central American	HIS
SCAMB	Black - South or Central American	AFA
SCSEAI	Southeast Asian	API
VIET	Vietnamese	API

- Newer data (includes DRB345)
- Uniform, high resolution HLA typing
- Better reflection of HLA allele diversity/linkage patterns
- Populations all screened within US

Analysis tools with broad applications

http://tools.iedb.org/main/analysis-tools/

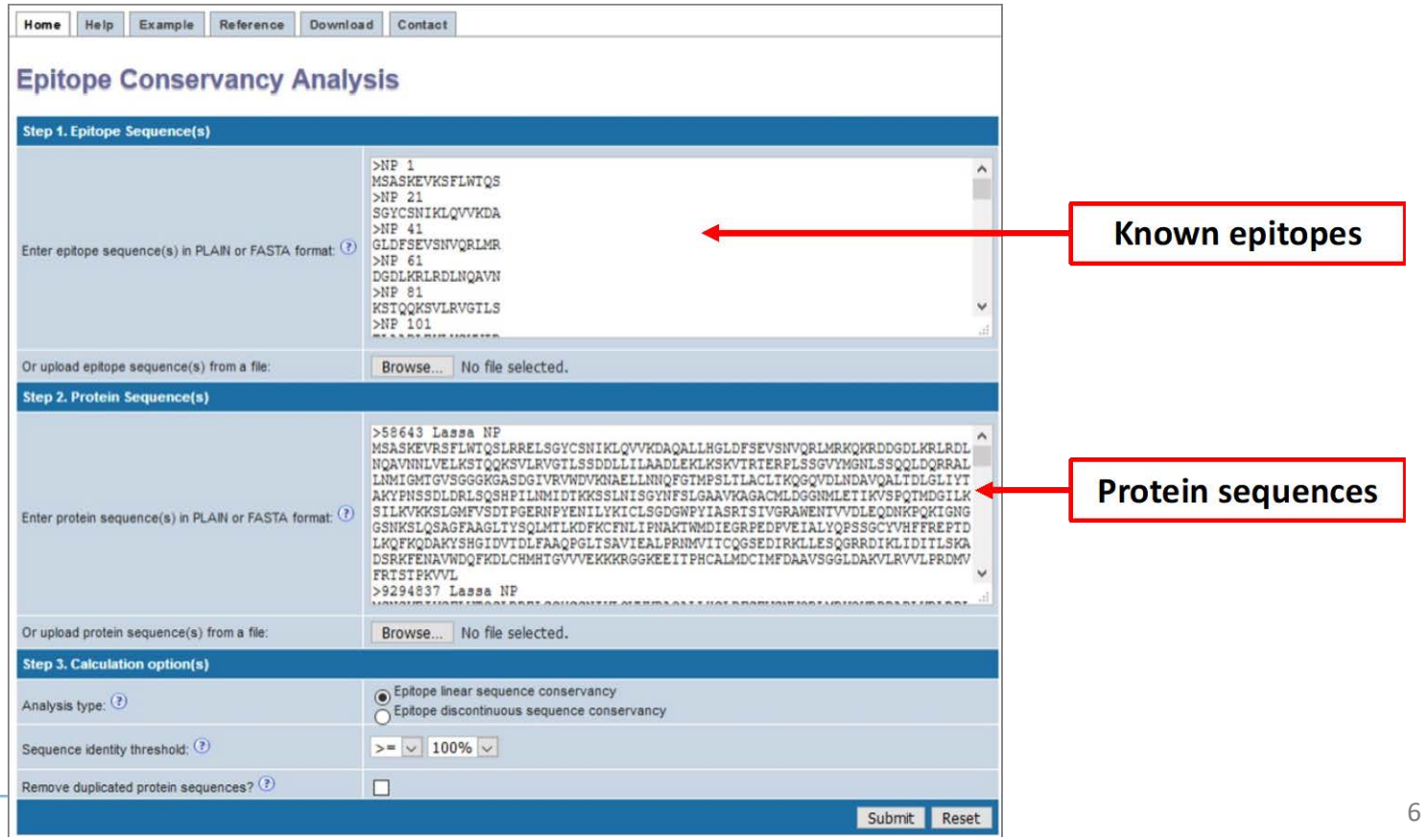


Epitope Conservancy Analysis

http://tools.iedb.org/conservancy/

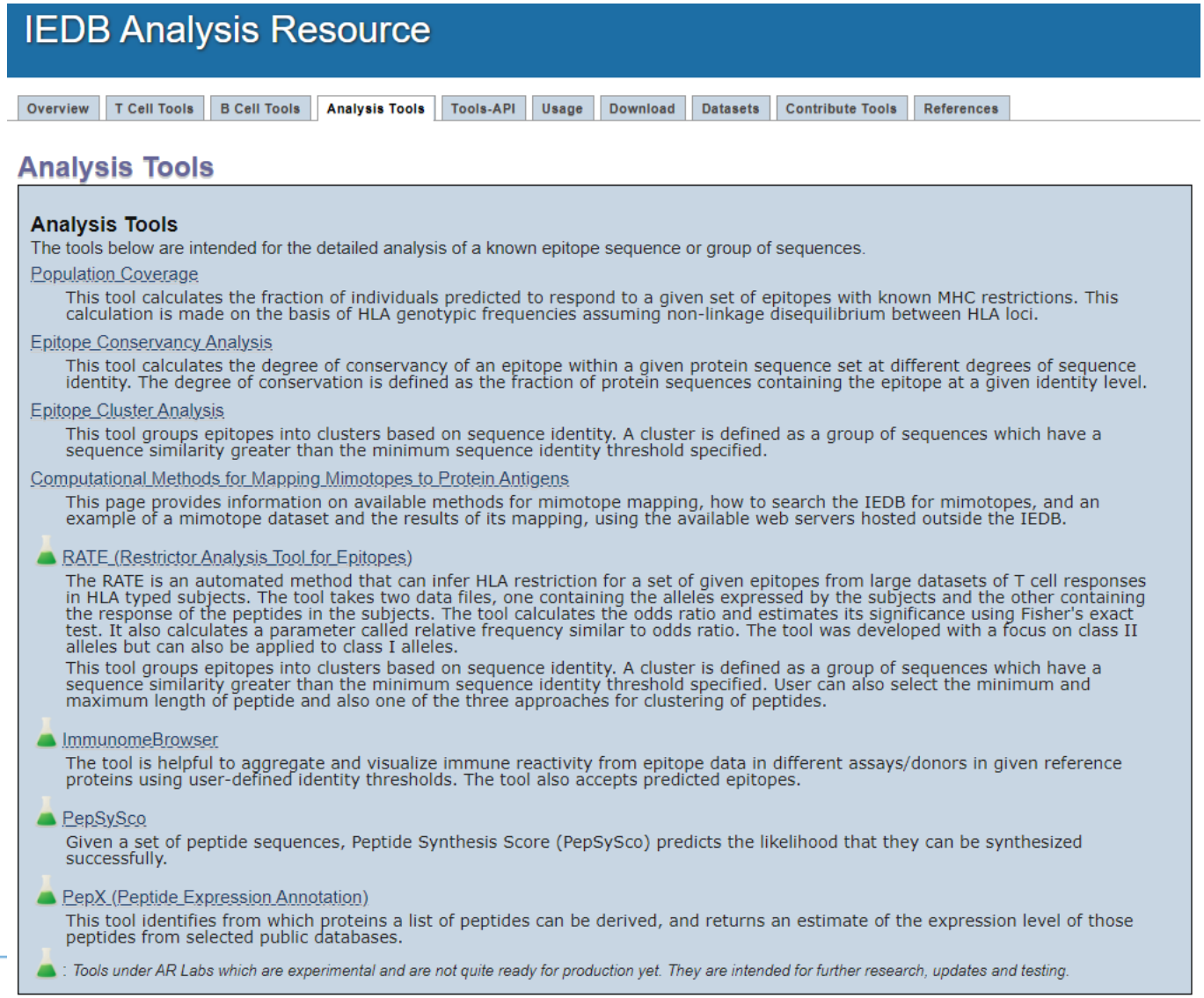
- Calculates the degrees of conservancy of one or more epitopes, within a given set of protein sequences
- Adjustable sequence identity threshold

"Degree of conservation" = the fraction of protein sequences containing the epitope at a given identity level



Analysis tools with broad applications

http://tools.iedb.org/main/analysis-tools/



Epitope Cluster Analysis Tool

http://tools.iedb.org/cluster/

- Analyzes how many epitopes in a set have significant sequence homology
- Groups epitopes into clusters based on having sequence identity greater than a specified threshold
- Three different clustering approaches are implemented
- Enables diverse applications such as generating epitope pools, and understanding cross-reactivity

Epitope Cluster Analysis Tool

Clustering Tool Updated! This is an updated version of the IEDB clustering tool that is more robust and has new functionality described in more detail here.

Epitope Cluster Analysis

step 1/3 Specify input peptides

Specify Sequence(s)	
Enter epitope sequence(s) in PLAIN or FASTA format	
Or upload epitope sequence(s) from a file	Choose File No file chosen

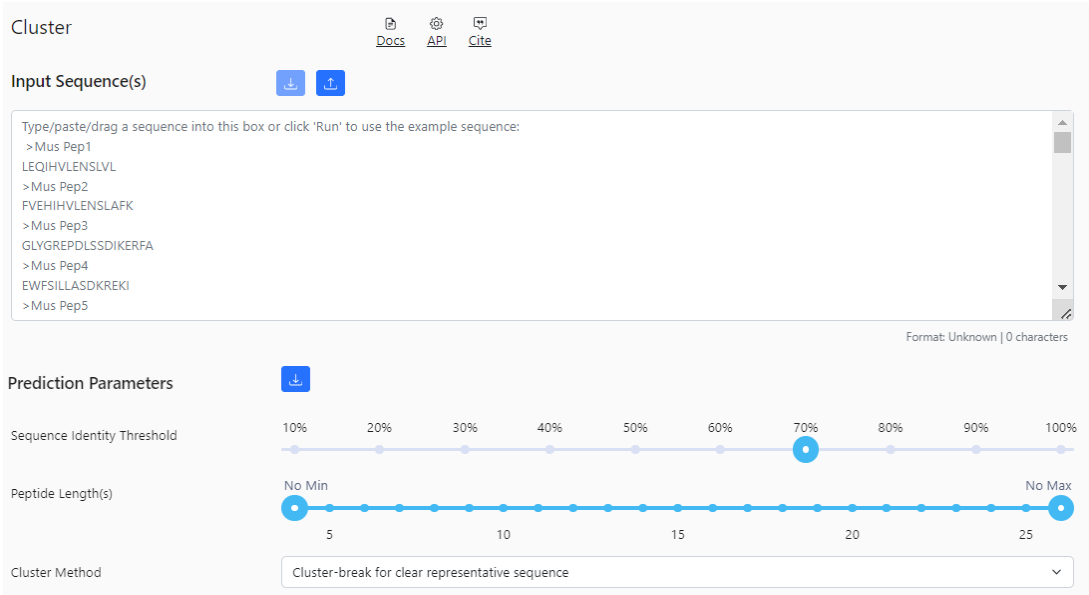
Go to Step 2 | Reset

http://tools.iedb.org/cluster/

https://nextgen-tools.iedb.org/clustering

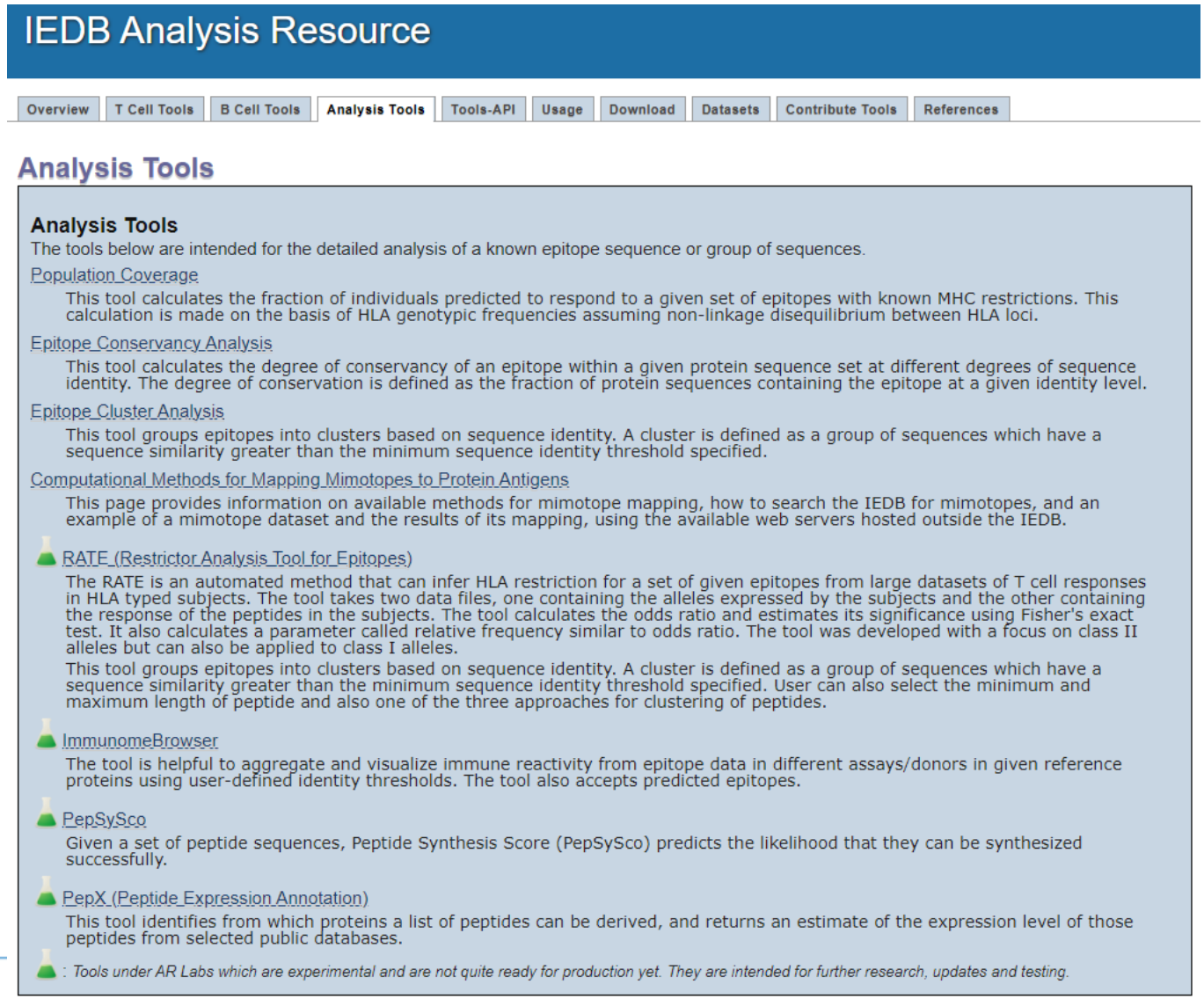
Learn more about the new Cluster Tool **tomorrow**!

Thursday, November 2, 9am PT
Next Generation Tool Pipelines - Cluster & PepMatch



Analysis tools with broad applications

http://tools.iedb.org/main/analysis-tools/



Restrictor Analysis Tool for Epitopes (RATE)

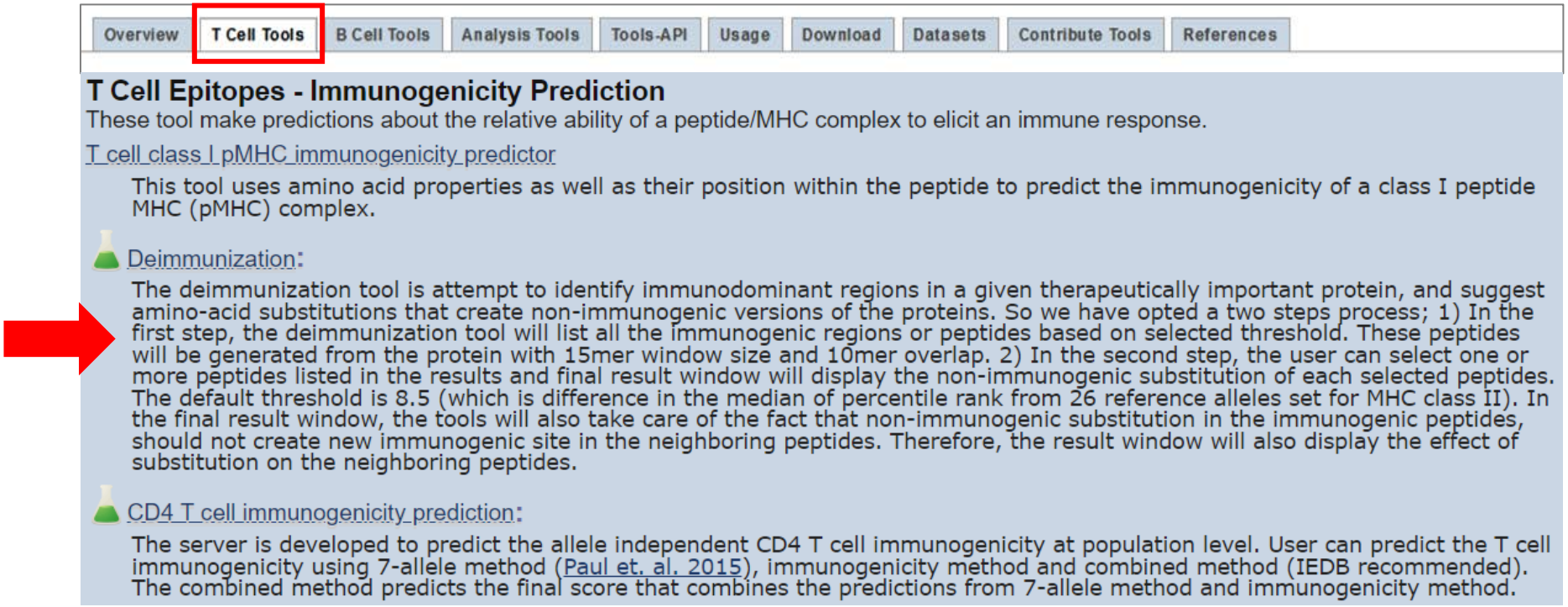
http://tools.iedb.org/rate/

- Automated method to infer HLA restriction of a given epitope set, from immune response data of HLA typed subjects
- Based on computing the frequency of alleles expressed in in donors who had an immune response to a given epitope
- Compare those frequencies in donors that did not have a response

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	**Peptide#**	**Peptide_id**	**peptide_seq**	**Allele#**	**Allele**	**A+R+**	**A-R+**	**A+R-**	**A-R-**	**No_of_Donors**	**Response_n/a**	**Relative_freq**	**Odds_ratio**	**P-value**
2	1	3531.0365	GINTIPIAINEAEYV	84	DRB3*02:02	6	0	18	26	50	0	2.08333	inf	0.00847
3	2	3531.0367	AAFQAAHARFVAAAA	54	DRB1*01:01	4	8	0	38	50	0	4.16667	inf	0.00215
4	2	3531.0367	AAFQAAHARFVAAAA	65	DRB1*07:01	5	7	2	36	50	0	2.97619	12.85714	0.00593
5	3	3531.0514	AAVVRFQEAANKQKQ	89	DRB5*01:01	7	2	8	30	47	3	2.43704	13.125	0.0025
6	5	3531.037	AAGTYVAADAAAAST	54	DRB1*01:01	4	2	0	44	50	0	8.33333	inf	7.00E-05

Deimmunization

http://tools.iedb.org/deimmunization/



Deimmunization

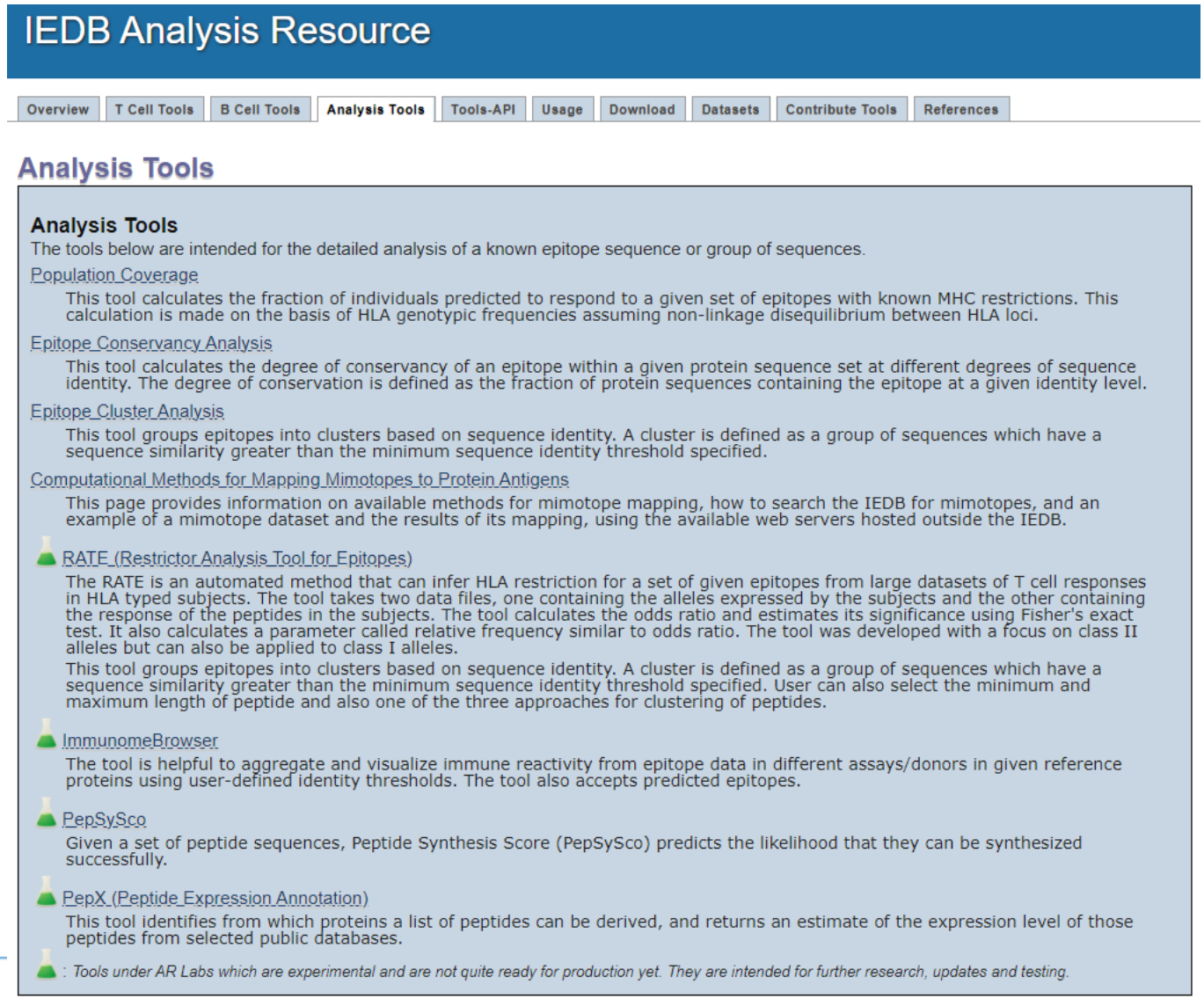
http://tools.iedb.org/deimmunization/

- Generate overlapping peptides from protein sequence
- Predict HLA class II binding binding regions
- Suggest amino acid substitutions that are predicted to decrease binding
  - Also consider the effect of substitutions on neighboring peptides

	A	B	C	D	E	F	G	H	I	J	K	L
1	Protein Number	Peptide	Peptide ID	Start Position	End Position	Median Percentile Rank	Median Difference	C terminal Neighbor 1 (Median)	C terminal Neighbor 2 (Median)	N terminal Neighbor 1 (Median)	N terminal Neighbor 2 (Median)	Deimmunization Score
2	1	DTFRKLFRVYSNFLR	wild	136	150	10.255	0	14.5	49.75	47	44.25	NA
3	1	DTFRKLFRVYSNFDR	L149D	136	150	32	21.745	41	68.5	NA	NA	3
4	1	DTFRKLFRVYSNFGR	L149G	136	150	27.5	17.245	31.5	68.75	NA	NA	3
5	1	DTFRKEFRVYSNFLR	L141E	136	150	26.5	16.245	26.5	NA	60.75	NA	3
6	1	DTFRKPFRVYSNFLR	L141P	136	150	26.35	16.095	26.5	NA	52.5	NA	3
7	1	DTFRKQFRVYSNFLR	L141Q	136	150	26	15.745	25.5	NA	53.25	NA	3
8	1	DTFRKLFRVYSNFNR	L149N	136	150	26	15.745	21.75	61.75	NA	NA	3
9	1	DTFRKLGRVYSNFLR	F142G	136	150	25.075	14.82	31	NA	50.75	NA	3
10	1	DTFRKLFDVYSNFLR	R143D	136	150	24.75	14.495	35.5	NA	49.25	NA	3
11	1	DTFRKKFRVYSNFLR	L141K	136	150	24.525	14.27	23	NA	59	NA	3
12	1	DTFRKLFRVYSNFCR	L149C	136	150	24.25	13.995	36	64	NA	NA	3
13	1	DTFRKCFRVYSNFLR	L141C	136	150	23.85	13.595	23.5	NA	73	NA	3
14	1	DTFRKLFRVYSCFLR	N147C	136	150	23.75	13.495	39.75	50	NA	NA	3
15	1	DTFRKDFRVYSNFLR	L141D	136	150	23.5	13.245	27.5	NA	62.5	NA	3
16	1	DTFRKGFRVYSNFLR	L141G	136	150	23.5	13.245	27	NA	57.5	NA	3
17	1	DTFRKLFRVYSNFKR	L149K	136	150	23.5	13.245	28	59	NA	NA	3
18	1	DTFRKLFRVYSNFER	L149E	136	150	22.75	12.495	34	62	NA	NA	3
19	1	DTFRKLERVYSNFLR	F142E	136	150	22.7	12.445	31	NA	48	NA	3
20	1	DTFRKLCRVYSNFLR	F142C	136	150	22.4	12.145	31	NA	50.25	NA	3
21	1	DTFRKLTRVYSNFLR	F142T	136	150	22.175	11.92	30	NA	48.75	NA	3

Analysis tools with broad applications

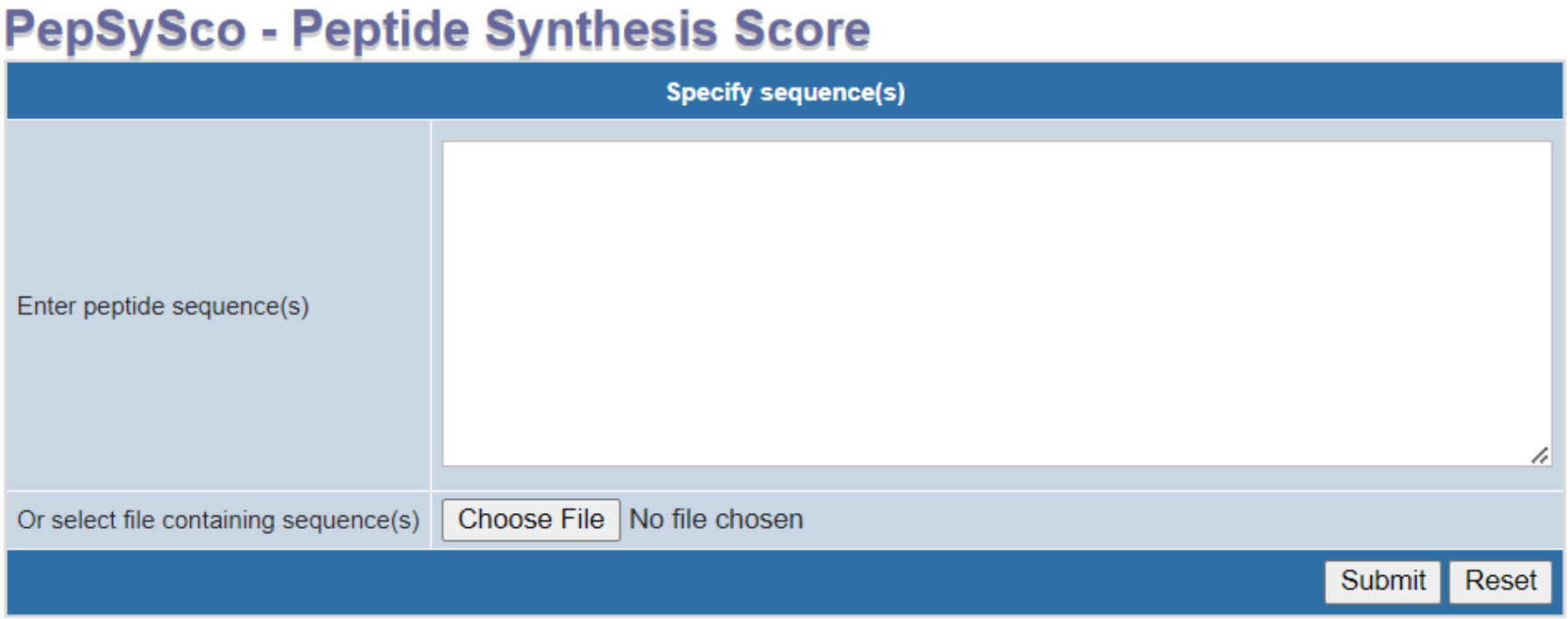
http://tools.iedb.org/main/analysis-tools/



Peptide Synthesis Score (PepSySco)

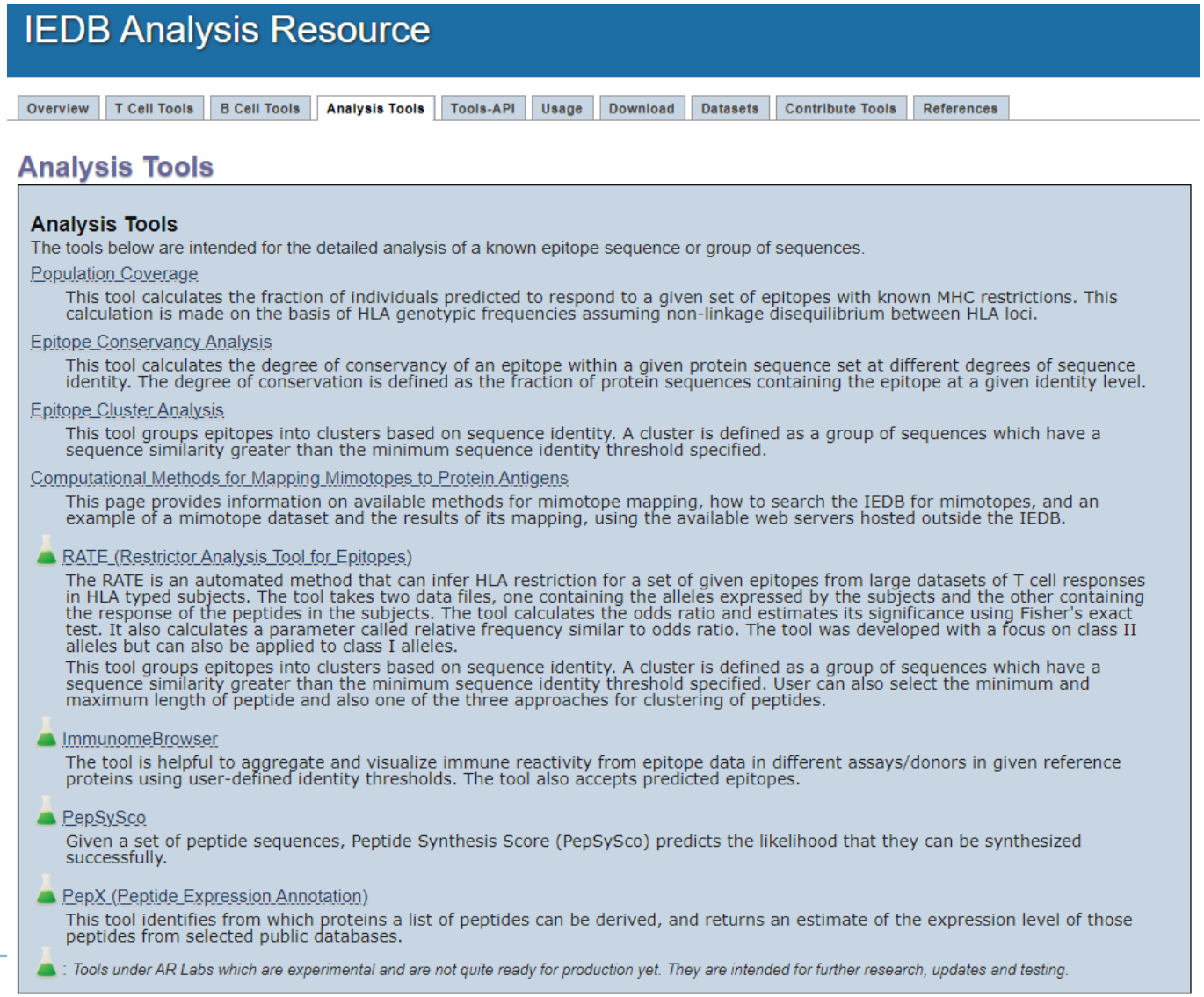
http://tools.iedb.org/pepsysco/

- Given a set of peptide sequences, PepSySco predicts the likelihood that they can be synthesized successfully.



Analysis tools with broad applications

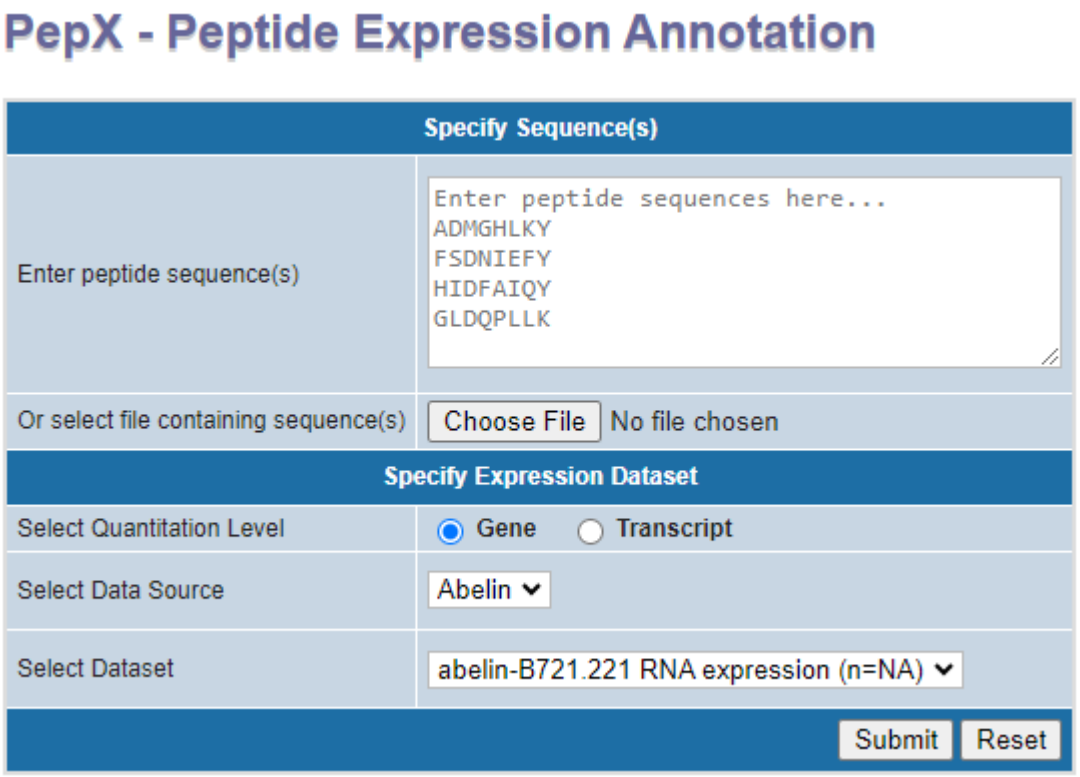
http://tools.iedb.org/main/analysis-tools/



Peptide Expression Annotation (PepX)

http://tools.iedb.org/pepx/

- PepX identifies from which proteins a list of peptides can be derived
- Returns an estimate of the expression level of those peptides from selected public databases



Analysis Tools Recap

- Help to examine existing sets of epitopes and gain new knowledge across a broad array of applications

Population Coverage	Analyze T cell epitopes with known HLA restriction that are recognized in a population based on HLA frequencies
Conservancy	Investigate epitope conservancy across different protein sequences.
Cluster	Cluster epitopes on the basis of homology
RATE	Infer HLA restrictions for epitopes of T cell response frequency in HLA typed subjects
Deimmunization	Identify immunodominant regions in a given protein, and suggest amino-acid substitutions that create non-immunogenic versions of the protein
PepSySco	Predict the likelihood that a set of peptides can be synthesized successfully
PepX	Identify peptides that can be derived from which proteins, and estimate the expression level of those peptides from selected public databases