How Data are Retrieved, Entered, and Organized

Randi Vita, M.D. **Lead Ontology and Quality Manager**

IEDB Scope

- Infectious diseases, Allergy, Autoimmunity, Transplant
  - HIV, cancer, etc only curated as structural data or when presented with above subjects as per NIH/NIAID
- Experimentally confirmed, no predictions, no reviews
- Negative data and supplemental data included
- Binding of an adaptive immune receptor to an epitope (T cell, B cell, MHC binding, MHC ligand elution)
- Must be epitope specific
  - All tested antigens are curated for epitope specific receptors

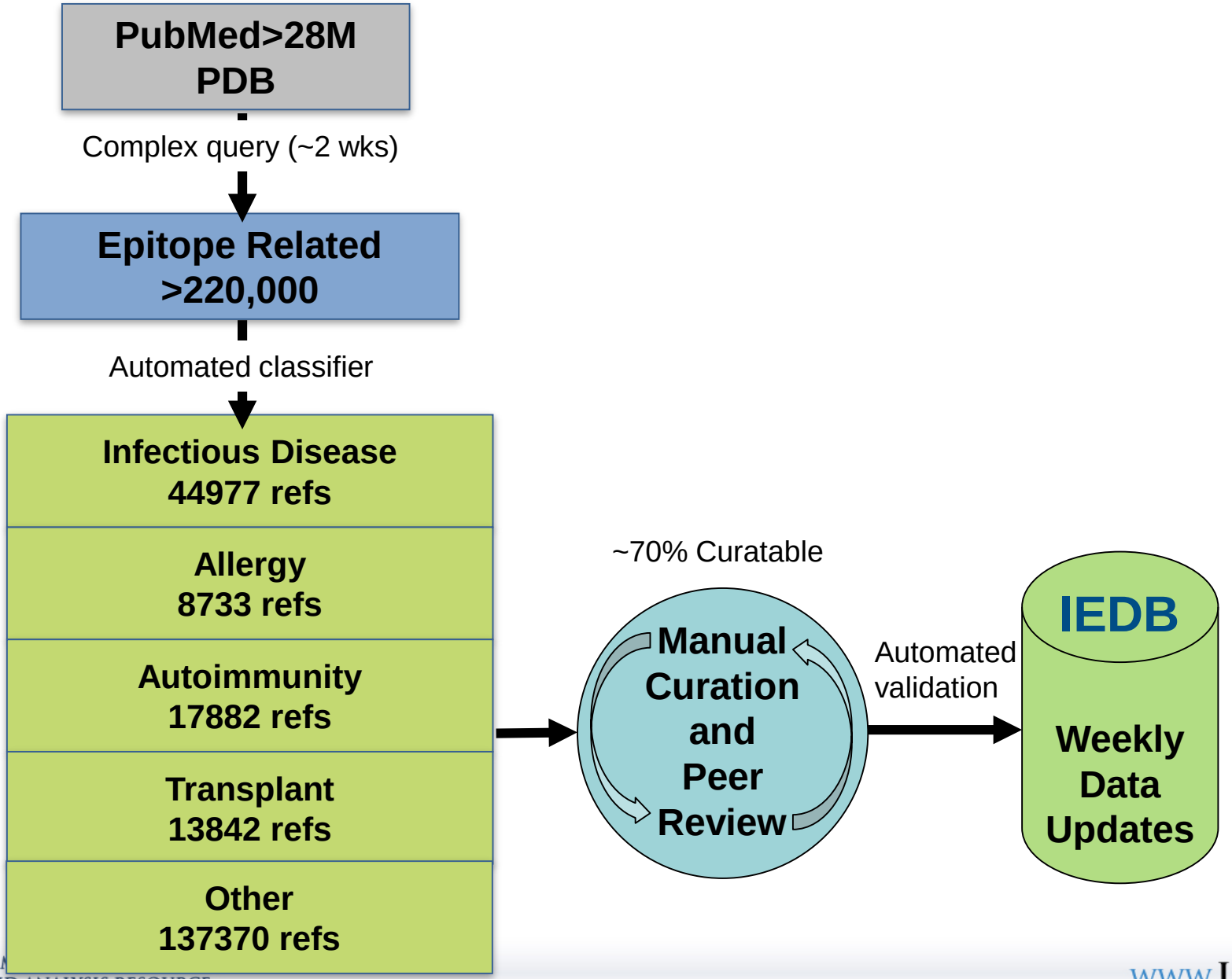
Minimal Criteria for Epitope Inclusion

- Linear peptide ≤50 amino acids in length
- Tested as an immunogen or an antigen
- Discontinuous residues shown to be important in recognition
- Non-peptidic epitopes ≤5000 Daltons
- Minimal information required (sequence, outcome, host, etc)

Data Sources

- IEDB does not create data. We capture experimentally generated data
- Published literature ~19,500
- Direct submissions 330

Literature Curation Process



Consistency and quality control measures

- Manually curated by a team of 8 PhD level scientists at LJI and Chemical Entities of Biological Interest (ChEBI) with specific expertise
- Formal curation guidelines and peer review
- External immunological experts
- Built in validation in the curation application

Submission Community

- Primarily Epitope Discovery Contract Holders
- P. vivax, P. falciparum, Dengue, JEV, RVFV, Oropouche virus, Influenza, Ricin, Ebola, HCV, S. aureus (MRSA), HHV-6, Lassa, Vaccinia, VZV, common allergens
- David Fremont, James Crowe, Nicholas Mantis, Ben Doranz, Gregg Silverman, Alex Sette, Lawrence Stern, Michael Oldstone, David Koelle, William Kwok

Submission Process

Submit files via IEDB data submission tool

Automated validation

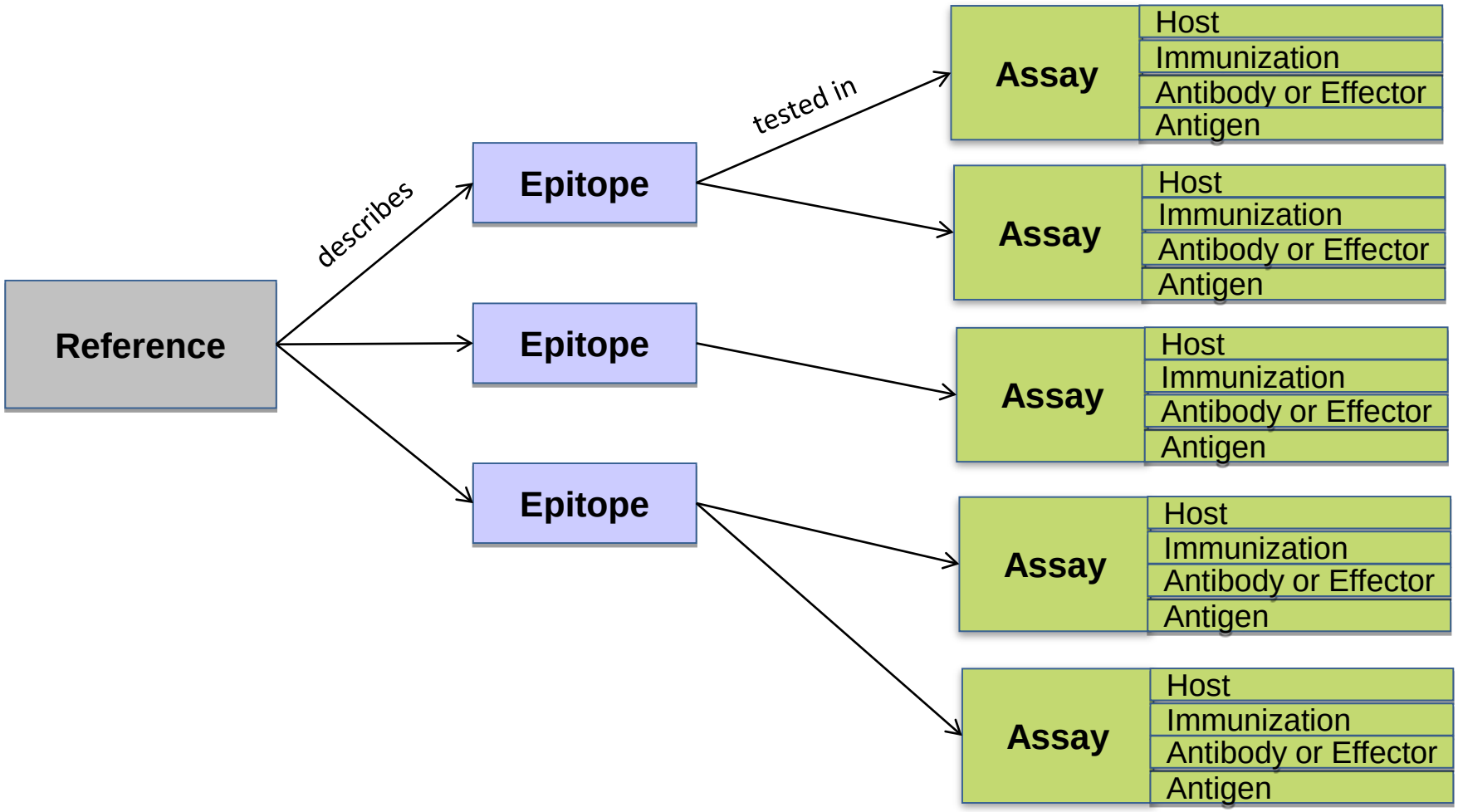
Curator review and submitter approval

Automated validation

Release date set by submitter

Data Structure

A database of experiments



Collaborations with ontologies

Provides standardized nomenclature, definitions, synonyms, and hierarchical relationships

Ensures consistency and accuracy

Finds errors

Makes curation easier

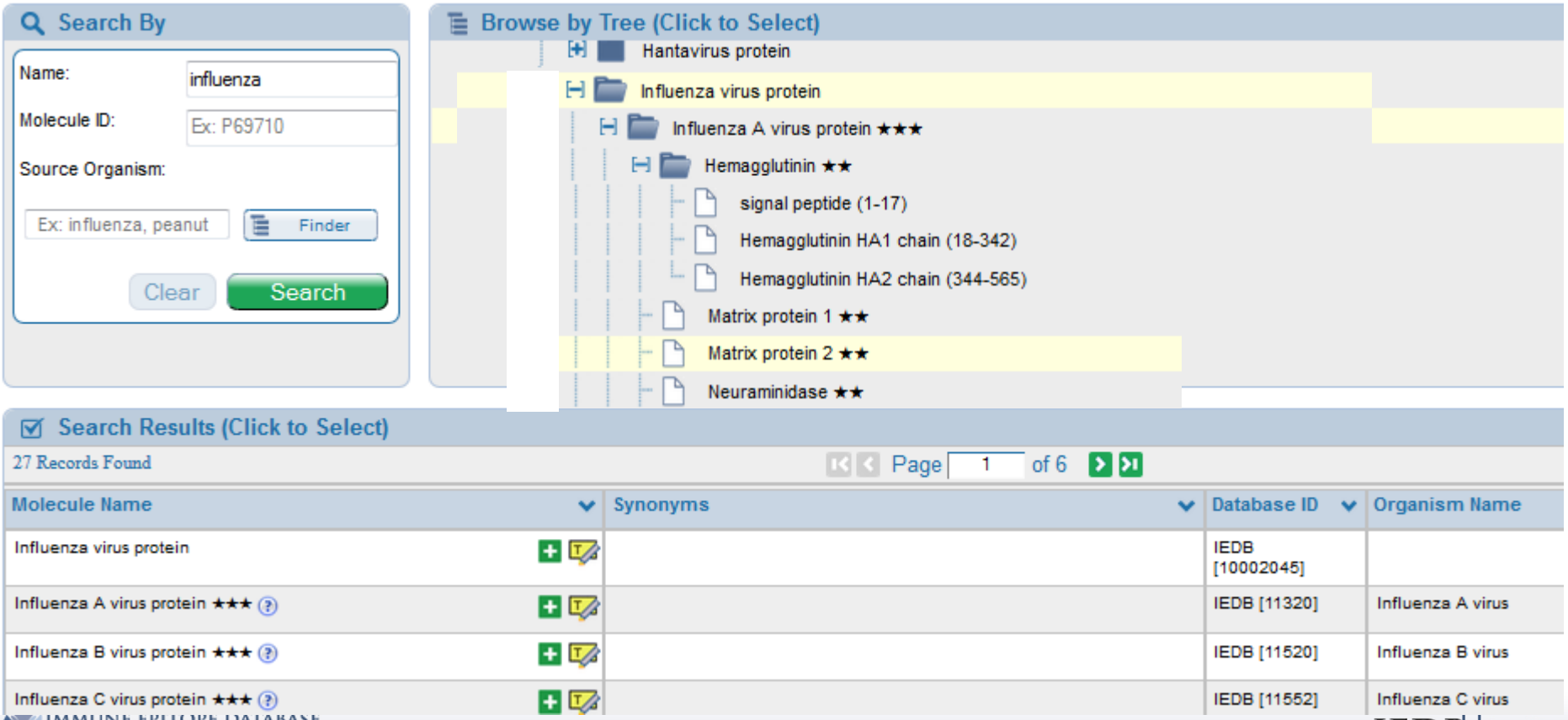
Enhances user experience

Facilitates interoperability

External Resources and Ontologies

Peptidic Epitope	Amino acid sequence
	Protein source
	Organism source

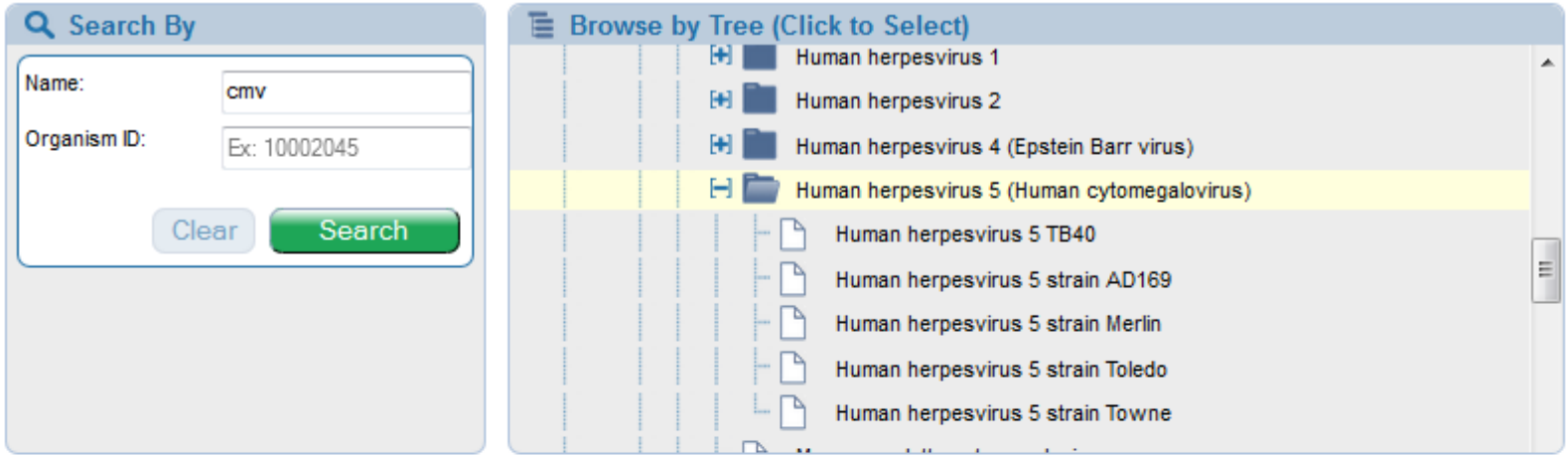
PKYVKQNTLKLAT
hemagglutinin HA1
GenBank/UniProt
GI:AAL62329.1—UniProt:Q8V285



External Resources and Ontologies

Peptidic Epitope	Amino acid sequence
	Protein source
	Organism source

Human herpesvirus 5 (HHV-5)
NCBI
taxon:10359



Search Results (Click to Select)

5 Records Found Page 1 of 1 5 Per Page

Organism Name	Synonyms	Organism ID
Cucumber mosaic virus (cucumber mosaic cucumovirus)	Cucumber mosaic virus, cucumber mosaic cucumovirus, cucumber mosaic cucumovirus CMV, cucumber mosaic virus, CMV, cucumber mosaic virus CMV, CMV	12305
Lymphocytic choriomeningitis mammarenavirus	Lymphocytic choriomeningitis mammarenavirus, Lymphocytic choriomeningitis virus, lymphocytic choriomeningitis virus LCMV, LCMV	11623
Human herpesvirus 5 (Human cytomegalovirus)	Human cytomegalovirus, Human herpesvirus 5, herpes virus 5,CMV,HHV5,HSV-5,HSV5, human herpesvirus type 5, human cytomegalovirus CMV, Human cytomegalovirus HCMV, HHV-5	10359

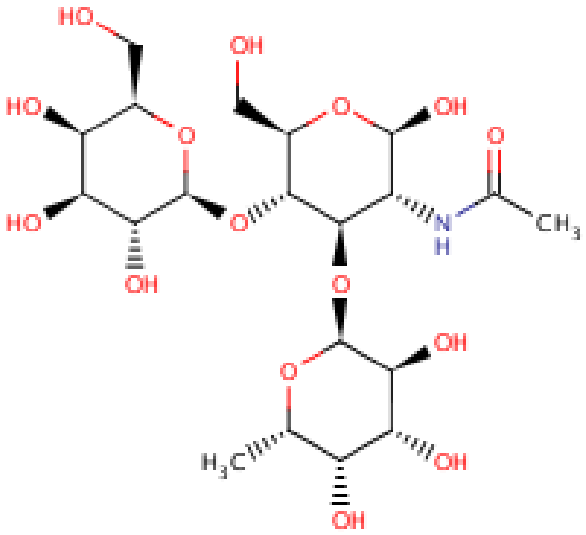
External Resources and Ontologies

Non-peptidic Epitope	
	Structure name
	Source structure
	Organism source

α-L-Fucp-(1→3)-[β-D-Galp-(1→4)]-β-D-GlcpNAc
CHEBI:59294

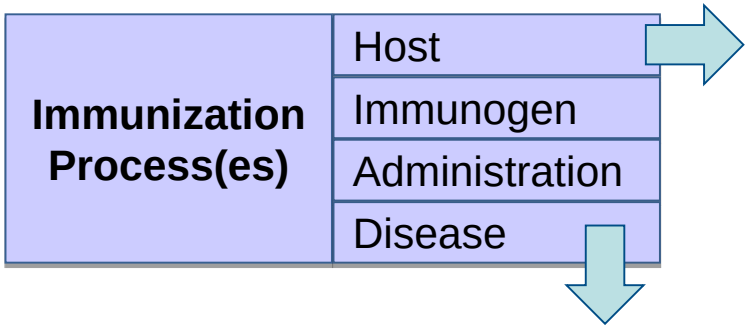
Lipopolysaccharide
CHEBI:16412

NCBI taxon:210
Helicobacter pylori



ChEBI Name	**α-L-Fuc*p*-(1→3)-[β-D-Gal*p*-(1→4)]-β-D-Glc*p*NAc**
ChEBI ID	**CHEBI:59294**
ChEBI ASCII Name	alpha-L-Fucp-(1->3)-[beta-D-Galp-(1->4)]-beta-D-GlcpNAc
Definition	An α-L-Fuc*p*-(1→3)-[β-D-Gal*p*-(1→4)]-D-Glc*p*NAc where the glucosamine at the reducing end has β-configuration at its anomeric centre. Commonly known as Lewis x trisaccharide or Le^x.
Stars	⭐⭐⭐ This entity has been manually annotated by the ChEBI Team.

External Resources and Ontologies



Homo sapiens (human)
NCBI
taxon:9606

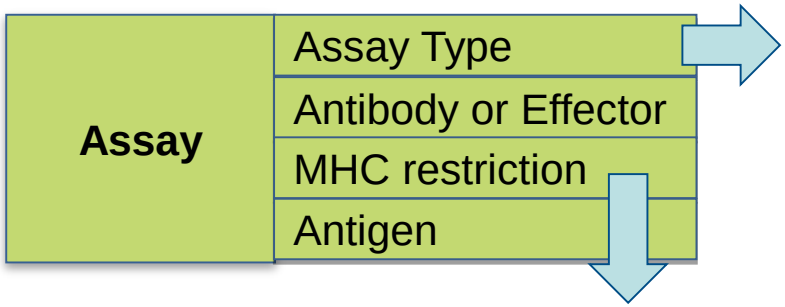
Peanut allergy
DO
DOID:4378

DISEASE ONTOLOGY

Metadata Visualize

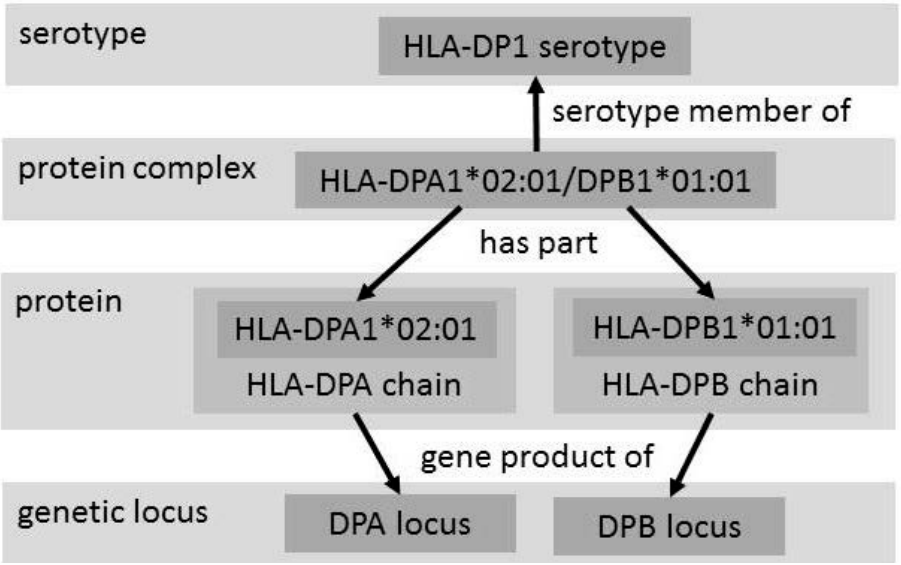
DOID	DOID:4378
Name	peanut allergic reaction
Definition	A food allergy that is an allergy or hypersensitivity to dietary substances from peanuts causing an overreaction of the immune system which in a small percentage of people may lead to severe physical symptoms. http://en.wikipedia.org/wiki/Allergy#Foods
Synonyms	Allergy to peanuts (disorder) [EXACT] Peanut allergy [EXACT]
Xrefs	MSH:D021183 SNOMEDCT_US_2015_03_01:213021008 SNOMEDCT_US_2015_03_01:91935009 UMLS_CUI:C0559470
Relationships	is_a food allergy

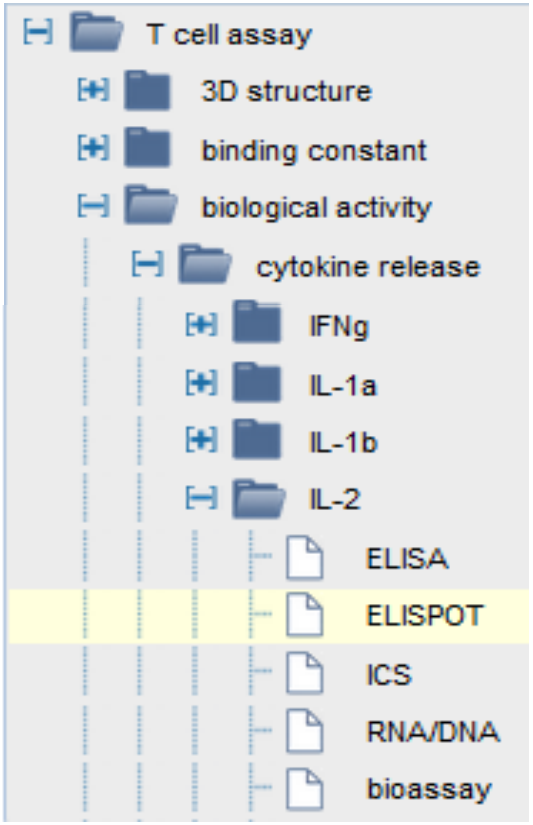
External Resources and Ontologies



ELISA IL-2 release
OBI
OBI:1110152

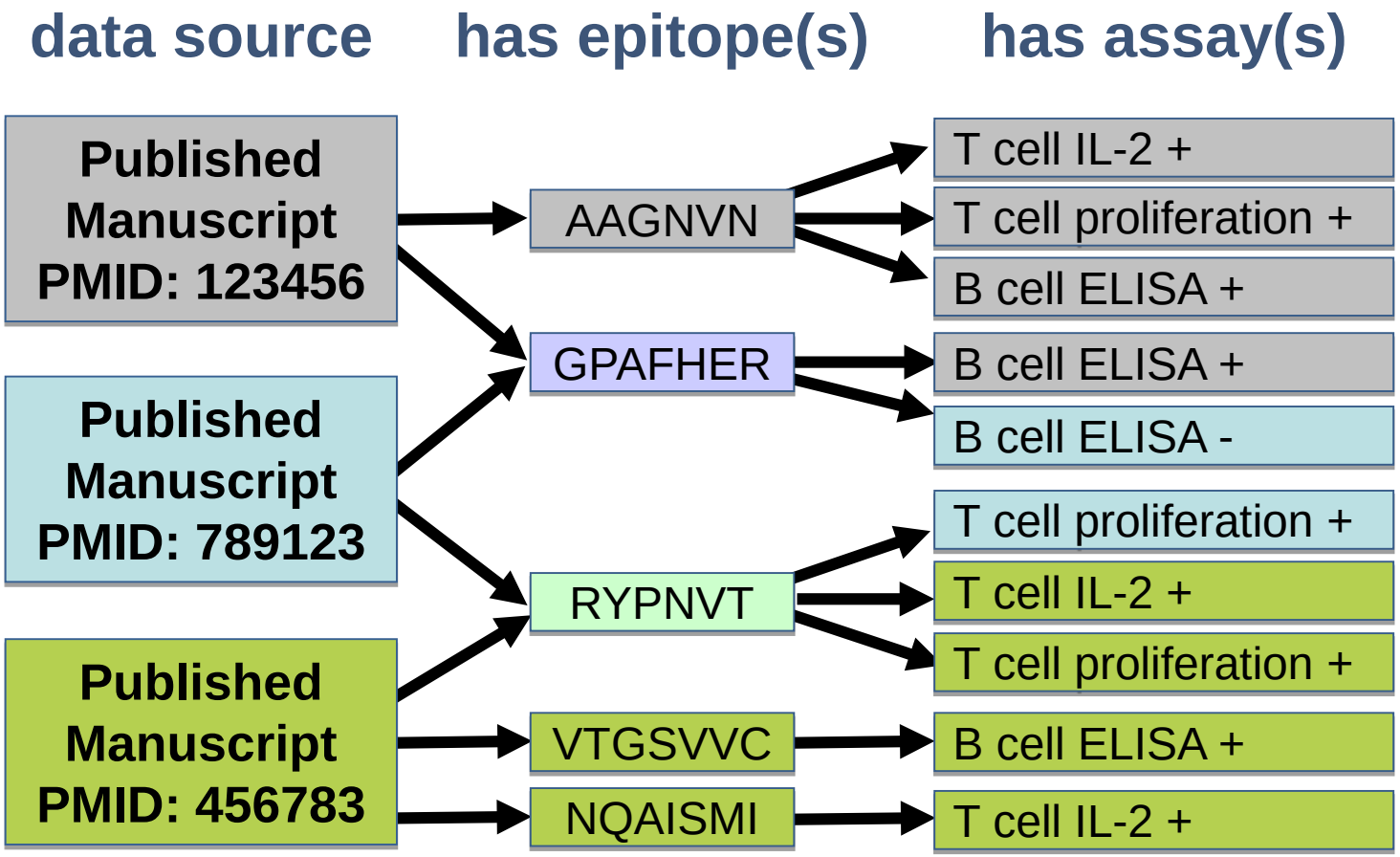
MHC Restriction Ontology (MRO)



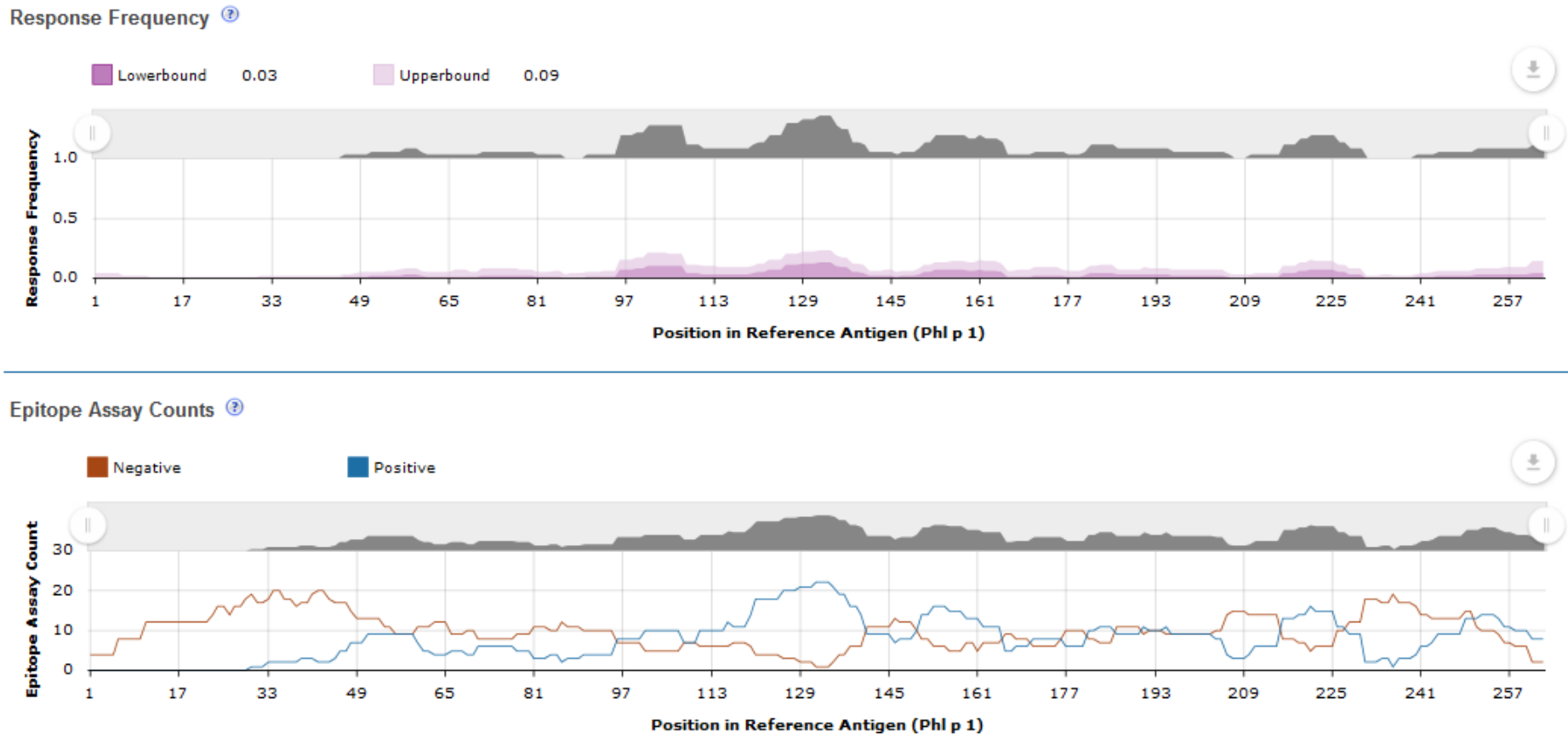


Data Aggregation

A database of experiments

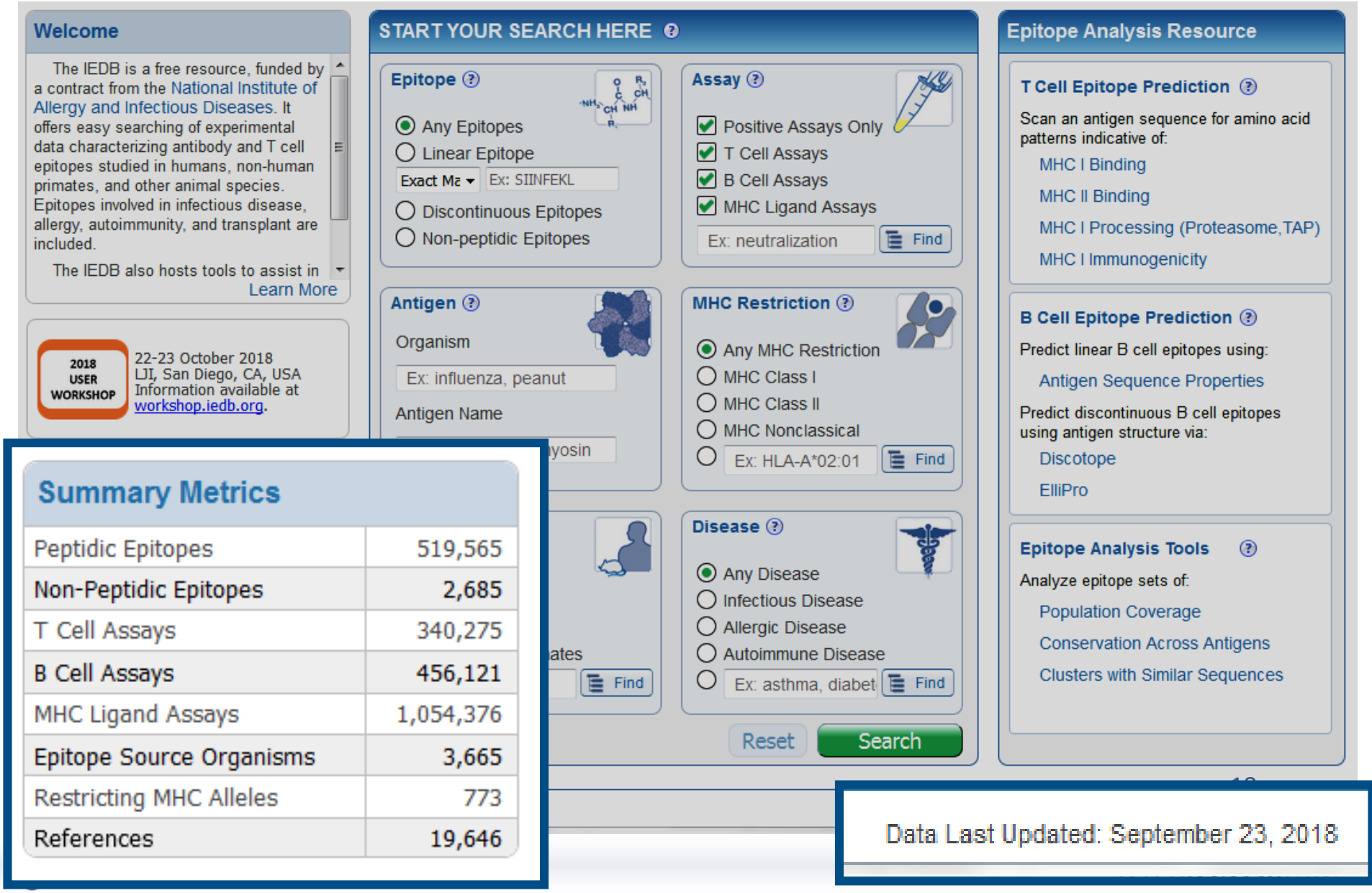


Data Aggregation



The human T cell response to Timothy grass Phl p 1 protein

Cumulative Data

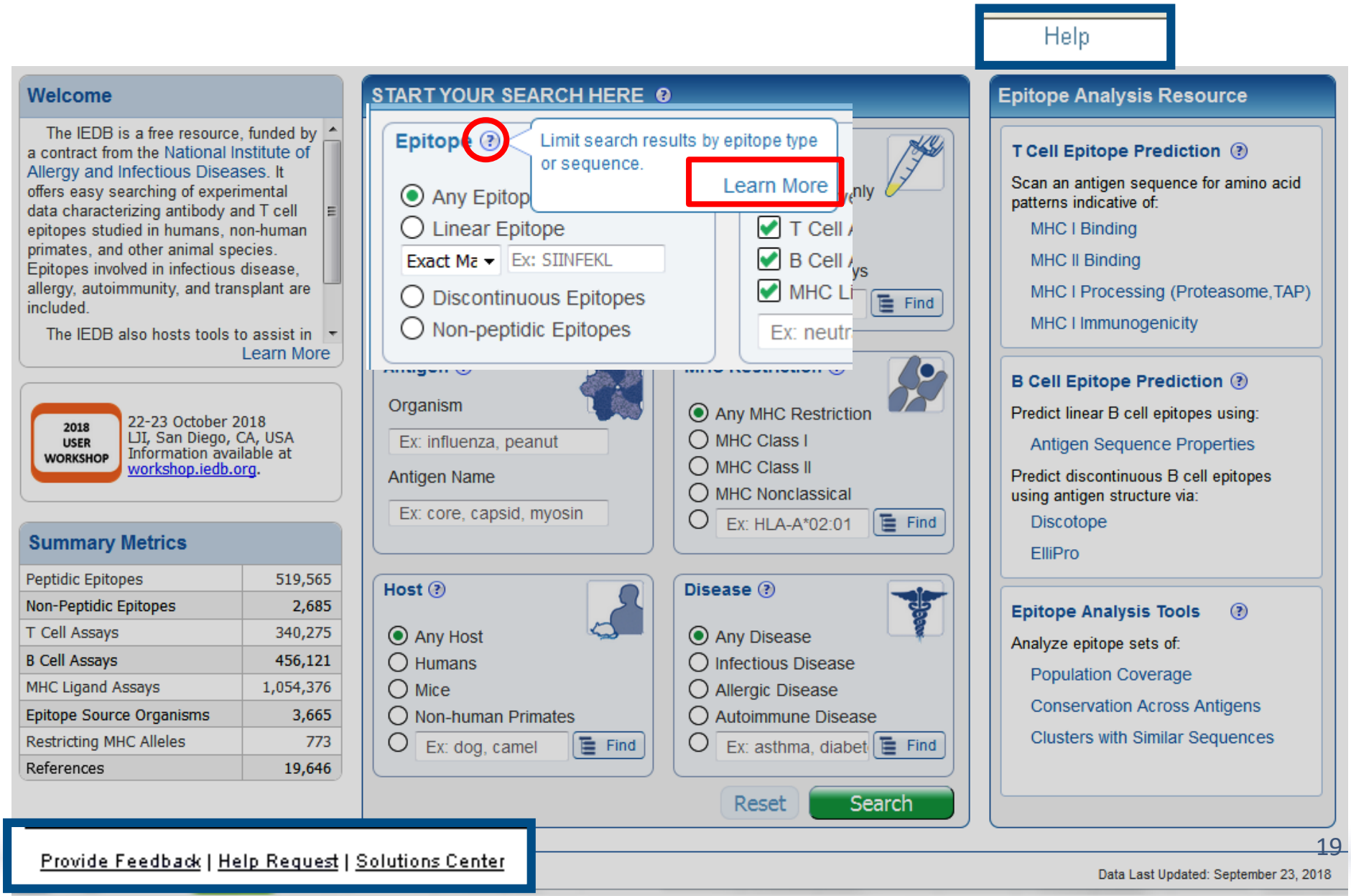


Summary Metrics

Peptidic Epitopes	519,565
Non-Peptidic Epitopes	2,685
T Cell Assays	340,275
B Cell Assays	456,121
MHC Ligand Assays	1,054,376
Epitope Source Organisms	3,665
Restricting MHC Alleles	773
References	19,646

Data Last Updated: September 23, 2018

Help and Feedback



Help and Feedback

IEDB Epitopes

Randi Vita

11 months ago · Updated

Follow

In order to be consistent across all publications, the IEDB defines an epitope as peptides less than or equal to 50 amino acids in size and non-peptidic structures less than or equal to 5000 Daltons. These structures must be experimentally tested for binding to an adaptive immune receptor (T cell receptor (TCR), antibody or B cell receptor (BCR), or major histocompatibility complex (MHC)) or the receptor must be known and stated to be epitope specific in order to be included in the IEDB.

Please see our full inclusion criteria at this link:

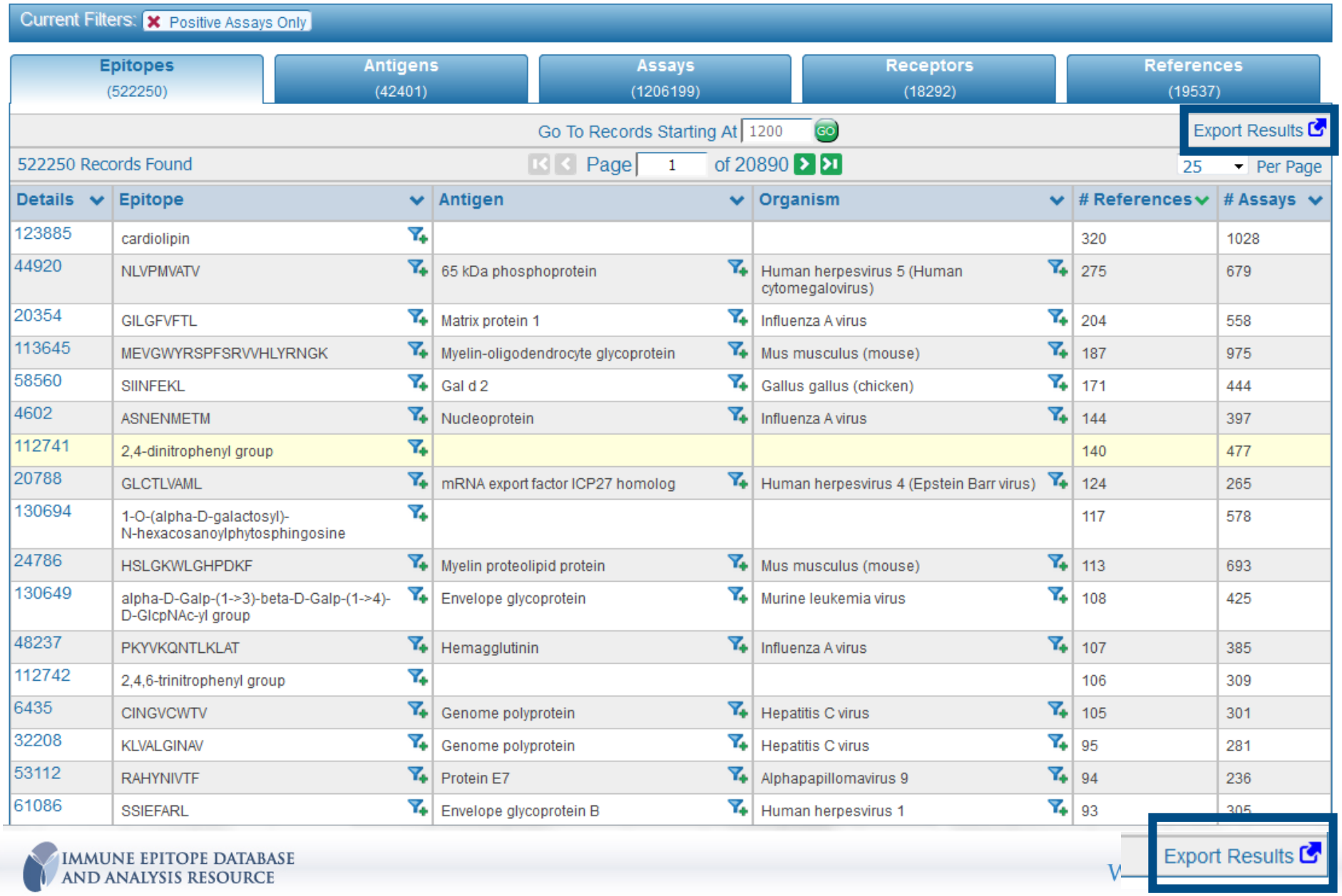
http://curationwiki.iedb.org/wiki/index.php/IEDB_Inclusion_Criteria

Users can easily narrow search results to include only linear peptidic epitopes, discontinuous peptidic epitopes, or non-peptidic epitopes. Peptidic epitopes may also be searched upon using any specific amino acid sequence, with the additional feature of being able to search for epitopes homologous to the input sequence at 70%, 80%, or 90% identity or using a substring match. Non-peptidic epitopes may be searched upon by textual name or via the Molecule Finder which presents all non-peptidic epitopes in a

Help and Feedback

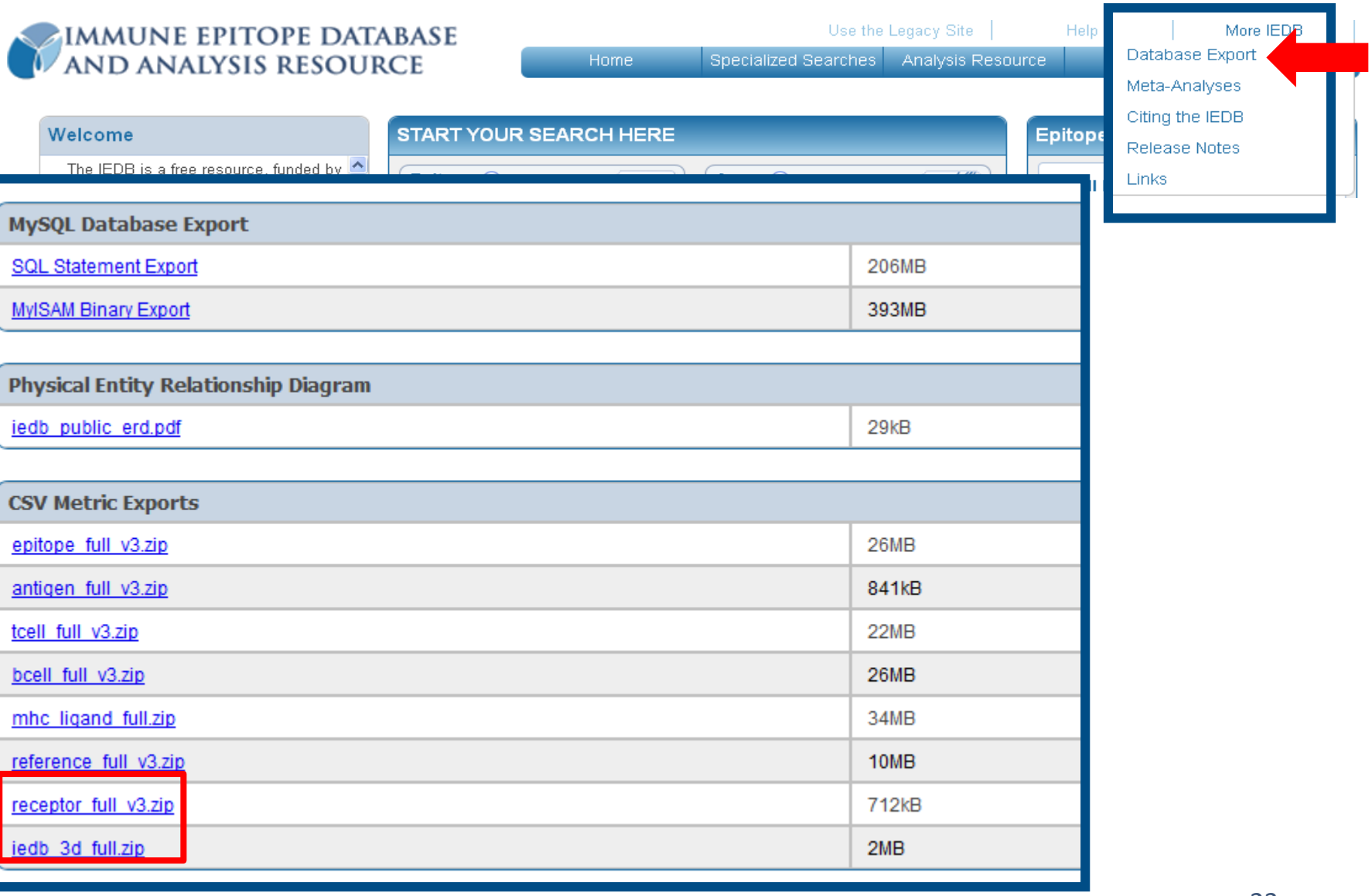
- Respond to all help requests within 1 business day and resolve issue within 2 days for 85% of requests
  - Submit by email to help@iedb.org
- Knowledgebase available at https://help.iedb.org
- User Workshop sessions will be recorded and converted into video tutorials

Exports



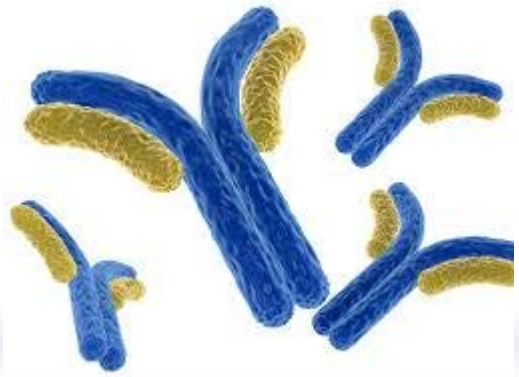
Details	Epitope	Antigen	Organism	# References	# Assays
123885	cardiolipin			320	1028
44920	NLVPMVATV	65 kDa phosphoprotein	Human herpesvirus 5 (Human cytomegalovirus)	275	679
20354	GILGFVFTL	Matrix protein 1	Influenza A virus	204	558
113645	MEVGWYRSPFSRVVHLYRNGK	Myelin-oligodendrocyte glycoprotein	Mus musculus (mouse)	187	975
58560	SIINFEKL	Gal d 2	Gallus gallus (chicken)	171	444
4602	ASNENMETM	Nucleoprotein	Influenza A virus	144	397
112741	2,4-dinitrophenyl group			140	477
20788	GLCTLVAML	mRNA export factor ICP27 homolog	Human herpesvirus 4 (Epstein Barr virus)	124	265
130694	1-O-(alpha-D-galactosyl)-N-hexacosanoylphytosphingosine			117	578
24786	HSLGKWLGHPDKF	Myelin proteolipid protein	Mus musculus (mouse)	113	693
130649	alpha-D-Galp-(1->3)-beta-D-Galp-(1->4)-D-GlcpNAc-yl group	Envelope glycoprotein	Murine leukemia virus	108	425
48237	PKYVKQNTLKLAT	Hemagglutinin	Influenza A virus	107	385
112742	2,4,6-trinitrophenyl group			106	309
6435	CINGVCWTV	Genome polyprotein	Hepatitis C virus	105	301
32208	KLVALGINAV	Genome polyprotein	Hepatitis C virus	95	281
53112	RAHYNIVTF	Protein E7	Alphapapillomavirus 9	94	236
61086	SSIEFARL	Envelope glycoprotein B	Human herpesvirus 1	93	305

Exports


IMMUNE EPITOPE DATABASE AND ANALYSIS RESOURCE
Use the Legacy Site
Help
Home
Specialized Searches
Analysis Resource
More IEDB
Database Export
Meta-Analyses
Citing the IEDB
Release Notes
Links
Welcome
START YOUR SEARCH HERE
MySQL Database Export
SQL Statement Export 206MB
MyISAM Binary Export 393MB
Physical Entity Relationship Diagram
iedb_public_erd.pdf 29kB
CSV Metric Exports
epitope_full_v3.zip 26MB
antigen_full_v3.zip 841kB
tcell_full_v3.zip 22MB
bcell_full_v3.zip 26MB
mhc_ligand_full.zip 34MB
reference_full_v3.zip 10MB
receptor_full_v3.zip 712kB
iedb_3d_full.zip 2MB

Receptor Export

- receptor_full contains exports of all antibody and T cell sequences in the IEDB
- Provides nucleotide and protein full length sequences
- Includes CDR1, 2, and 3 sequences
- Includes gene usage
- Listed by each epitope that they were shown to recognize



3D Export

- iedb_3d_full contains exports of all 3D structures in the IEDB
- Antibody, MHC, and T cell are in separate spreadsheets
- Provides structural details on receptor-Ag interactions (inter-molecular contacts)
- Include PDB IDs and resolutions of complexes, PDB chain IDs of antibody, TCR, MHC, and Ag chains
- Includes full length receptor and antigen sequences
- Can be combined with receptor export to get information on CDRs and VDJ gene usage